

Projeto de datasets de light fields sintéticos

Harllon Oliveira da Paz

Instituto Militar de Engenharia
Rio de Janeiro, RJ, Brasil, 22290-270
Email: harllon.paz@gmail.com

Luiz Velho

Instituto de Matemática Pura e Aplicada
Rio de Janeiro, RJ, Brasil, 22460-320
Email: lvelho@impa.br

Carla L. Pagliari

Instituto Militar de Engenharia
Rio de Janeiro, RJ, Brasil, 22290-270
Email: carla@ime.eb.br

Resumo—A crescente quantidade de artigos relacionados a “light field” indicam a relevância que esse assunto tem tomado nos dias atuais. Nesse contexto, percebe-se a necessidade de mais trabalhos com uma visão para criação de vídeos “light field”. Dessa forma, o presente trabalho objetiva suprir uma carência presente na comunidade científica, permitindo o acesso ao início de um projeto de datasets de “lightfield” de vídeos sintéticos. A partir de cenários e animações criadas no Blender, foi possível criar alguns vídeos “light field”, com os quais é possível realizar estudos aprofundados sobre as particularidades deste tipo de vídeo.

Abstract—The growing number of articles related to light fields indicates the relevance that this subject has taken on today. In this context, there is a need for more work with a vision for creating light field videos. Thus, the present work aims to supply a lack in the scientific community, allowing access to the beginning of a project of synthetic datasets of light field videos. From scenarios and animations created in Blender, it was possible to create light field videos, with which it is possible to explore the particularities of this type of video.

I. INTRODUÇÃO

Os “light fields” capturam todos os raios de luz que passam por um determinado volume de espaço. Comparado aos sistemas de imagem 2D tradicionais que capturam a intensidade espacial no plano da imagem, os “light fields” também contêm a direção angular dos raios de luz. Ou seja, “light fields” contêm informações espaciais e angulares dos raios de luz em um única exposição. Essas informações adicionais permitem múltiplas aplicações em diferentes áreas de pesquisa, como processamento de imagem, visão computacional e computação gráfica, incluindo (mas não se limitando a) a reconstrução da geometria 3D de uma cena, criando novas imagens do ponto de vista virtual, ou alterar o foco de uma imagem depois de capturada, sendo um tópico de interesse crescente na comunidade Virtual Reality/Augmented Reality (VR/AR). Esta nova modalidade de imagem despertou o interesse de comitês de normatização, tais como Joint Photographic Experts Group (JPEG) [1] e o Moving Picture Experts Group (MPEG) [2]. Diante desta demanda do MPEG por representação de vídeo “light fields”, a proposta deste artigo é a geração de datasets sintéticos de “light fields” que contenham informações espaciais e angulares de exposições sucessivas (variando ao longo do tempo). Uma das finalidades é a de atender às chamadas para propor técnicas de compressão de vídeos “light fields” [2]. Os “light fields” sintéticos oferecem flexibilidade na composição de valores de espaçamento entre as vistas, no grau de textura dos objetos e em suas diferentes posições na cena, criando desafios para as técnicas de compressão. Outra vantagem é a de prover o *ground truth* dos mapas de

disparidade, necessários para reconstrução 3D e para métodos de compressão baseados na profundidade. Pesquisadores da Google têm realizado trabalhos que se destacam no ambiente científico reportados nos artigos [3], e o recente sistema de captura de “light fields” [4].

Este artigo apresenta uma breve revisão bibliográfica na Seção II, seguida da Seção III em que o método é detalhado. Os resultados são descritos na Seção IV e a Seção V apresenta as conclusões.

II. REVISÃO BIBLIOGRÁFICA

“Light Fields” são uma função vetorial que descreve a quantidade de luz que atravessa o espaço em todos os pontos e em todas as direções. O espaço de todas as possibilidades de raios de luz é dado pela função plenóptica 5D [5]. Ou seja, a função plenóptica é a representação matemática de um “light field”. No entanto, cabe aqui ressaltar que a função plenóptica, na verdade, é uma função 7D que modela um ambiente 3D dinâmico, gravando os raios de luz em qualquer local (x,y,z) sob qualquer direção (θ, ϕ) , com qualquer faixa de comprimento de onda (λ) e a qualquer instante de tempo (t) [6].

Historicamente, sempre houve o interesse em conseguir visualizar uma foto por diferentes pontos de vista, mantendo a qualidade e sem gerar distorção na foto tirada. Para resolver esse problema, surge o sistema de renderização de imagem, o qual gera visões diferentes de um ambiente a partir de um conjunto de imagens pré-definidas. Em 1936, Arun Gershun, com seu artigo “The Light Field”, faz a primeira citação à teoria e à renderização de “light field”, mas os artigos que mais contribuíram para o início e para a propagação dessa teoria foram “The Lumigraph” [7] e “Light Field Rendering” [8]. Essencialmente, os dois artigos tratam de maneiras ligeiramente diferentes sobre a aquisição de imagens de “light field” a partir de imagens sintéticas e reais. Imagens Reais são imagens obtidas por uma câmera, enquanto imagens sintéticas são imagens criadas artificialmente pelo computador. A contribuição do artigo [8] é a parametrização da função plenóptica para construir novas perspectivas da cena. [8] “The Lumigraph”, por sua vez, contribui com a ideia de utilizar a geometria 3D para construir “light fields” a partir de entradas de imagens irregulares. [7] Esta função, que pode ser reduzida para 5 dimensões, pode ainda ser representada como uma função 4D $((L(u, v, s, t)))$, sem considerar a variação no tempo, conforme ilustrado pela Figura 1 e representadas como uma matriz de pontos de vista, em que as coordenadas (s, t) indicam a posição do ponto de vista e as coordenadas (u, v) , os valores

dos pixels dentro da vista. A representação 4D parametriza cada raio de luz pelas duas coordenadas 2D de sua interseção com os planos (s, t) e (u, v) . A Figura 2 exibe a mesma cena capturada a partir dos pontos de vista (s, t) . Os pixels de cada vista/imagem (s, t) são endereçados pelas coordenadas (u, v) . Desta forma, uma estrutura 4D consegue ser endereçada com um conjunto de estruturas 2D.

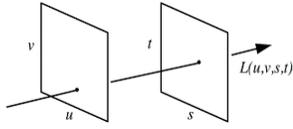


Figura 1. Parametrização (u, v, s, t) [8]

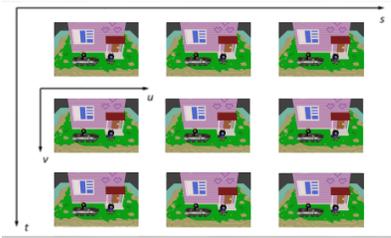


Figura 2. Exemplo de “light field” com parametrização (u, v, s, t) (“light field” Batman x Batman, Seção IV)

Uma câmera plenóptica é responsável pela criação de imagens de “light fields” a partir de cenas reais. Essa câmera captura informações sobre o campo de luz emanado da cena, ou seja, intensidade da luz e direção em que o raio de luz está trafegando no espaço [9]. Um exemplo de câmera plenóptica é a câmera que apresenta uma lente principal e uma matriz de microlentes posteriores. Um ponto específico do objeto é capturado (em diferentes ângulos) por estas várias microlentes, [10].

Há outros modos de aquisição de “light fields” usando matrizes de câmeras monoculares convencionais. Isso exige o emprego de mecânica fina e outros desafios, tais como calibrar um grande número de câmeras, necessárias para capturar [11] a mesma cena a partir de diferentes pontos de vista, além do engenhoso sistema apresentado em [3].

III. METODOLOGIA

A metodologia para realização deste trabalho foi baseada na geração de cenários por meio do software Blender [12], em que os elementos dos cenários foram obtidos no site sketchfab [13] e, então, modificados para serem adequados ao objetivo deste projeto. Além disso, a criação das imagens foi feita com câmeras virtuais criadas no artigo “A Dataset and Evaluation Methodology for Depth Estimation on 4D Light Fields” [14], base inicial deste projeto, que é um dataset de light fields sintéticos “estáticos”(que não variam ao longo do tempo), assim como os datasets gerados por [15].

O software de criação 3D com scripts em python, Blender [12], permite a modelagem de ambientes, cenários, objetos, além da criação de animações. É um software voltado para

designer que, em seu ambiente, permite a implementação de câmeras virtuais, as quais podem simular a aquisição de “light fields” e podem ser utilizadas para a geração de “light fields” sintéticos.

Um importante passo para a reconstrução de objetos 3D a partir de informações 2D é a disparidade. A disparidade pode ser horizontal ou vertical e, em linhas gerais, consiste na medida da localização de dois pontos (pixels) homólogos em duas vistas diferentes [16]. Quanto maior for a diferença, maior é o valor de disparidade e menor é o valor da profundidade deste ponto na cena, que estará mais perto do observador/câmera. A essa diferença de posição, dá-se o nome de disparidade. Por exemplo, o pixel na vista s_0, t_0 que aparece na coordenada u_x, v_y e é deslocado pelo software para a coordenada u_{x+d}, v_y na vista s_1, t_0 , onde o valor d é o valor da disparidade horizontal entre os pixels u_x, v_y e u_{x+d}, v_y .

O modelo de câmera virtual do Blender inclui os parâmetros intrínsecos e extrínsecos [16] da câmera virtual, que representam dados geométricos e ópticos da câmera. Foram empregados os valores para as *baselines* (distância entre as câmeras) horizontais e verticais no valor de 50mm e com o mesmo valor de F-stop (abertura da lente) para todas as câmeras virtuais de todos os datasets. Os cenários foram criados utilizando o valor padrão de “frames” por segundo (fps) do software Blender que são vinte “frames”, logo, cada animação criada tem duração inferior a um segundo. Todos os datasets apresentam um total de 8 frames, exceto o cenário 4 (Return to Home), que apresenta 6 frames. Para manter os eixos ópticos paralelos, as câmeras virtuais são deslocadas via um *offset* [17]. Cada dataset (cenário) foi criado usando diferentes comprimentos focais e variando parâmetros intrínsecos e extrínsecos das câmeras virtuais para criar diferentes cenários. Por exemplo, o comprimento focal mede a distância, em geral em milímetros (mm), entre o centro óptico da lente e o sensor da câmera (que pode ter diferentes tamanhos) e é determinado com a câmera focada no infinito. O comprimento focal descreve o ângulo de visão de uma lente (*field-of-view-FoV*) [16]. O parâmetro distância focal também assume diferentes valores, em que a distância focal refere-se à distância que o objeto em foco está da câmera. Desta forma, foram gerados cenários (ainda que com pouca variação de textura) com diferentes parâmetros intrínsecos e extrínsecos com diferentes objetos em diferentes profundidades.

Por fim, após a geração das imagens, elas foram tratadas e editadas utilizando ferramentas como imagemagick [18] e Matlab [19] para criação dos mapas de disparidade, sendo, então, visualizadas em um visualizador obtido no artigo “Light Field Video Capture Using a Learning-Based Hybrid Imaging System” [20]. Este visualizador foi modificado de forma a possibilitar que as mudanças de foco do vídeo sejam realizadas automaticamente em um intervalo de tempo bem definido, deste modo, não há a necessidade de modificar manualmente o foco do vídeo, permitindo melhor visualização dos vários valores de disparidade da cena.

A Figura 3 ilustra como seria um dataset com 3×3 vistas ($s \times v$) capturado/gerado ao longo do tempo (representado pelo eixo *frames*).

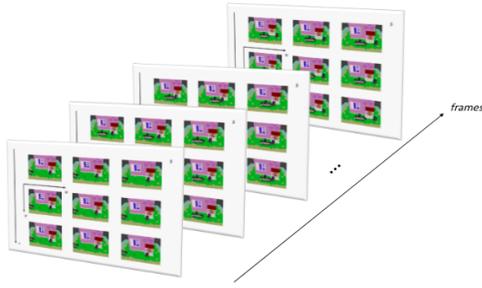


Figura 3. Exemplo de vídeo “light field” com parametrização (u, v, s, t) com variação ao longo do tempo (frames) (“light field” Batman x Batman, Seção IV)

IV. RESULTADOS

Para a realização dos cenários que serão apresentados abaixo, foram utilizados os seguintes projetos disponíveis na plataforma sketchfab [13]: Whale House de KattyLi, Lighthouse de Troublesome, Nurnberg de gfon296, Homework-Submarine de lucidvoo, Nortensky de Euvand, BTR 80 de Autem.mortale, Little Batman de agustinR, Steampunk Lighthouse de Akhikyan e An Island house de s20541. Estes modelos 3D, de acordo com os direitos de suas respectivas licenças, foram parcialmente modificados a fim de serem empregados neste projeto. Para isso, foram feitas as seguintes alterações: Junção com outros modelos 3D, retirada de elementos, alterações na textura, coloração e efeitos de animação.

Todos os parâmetros (intrínsecos e extrínsecos [16]) das câmeras virtuais e os vídeos “light field” apresentados podem ser acessados em [21].

A. Cenário 1: Tower of Gods

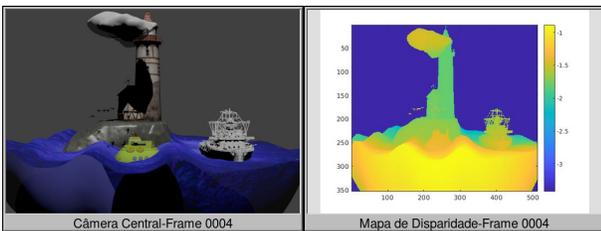


Figura 4. Visão Central e sua Disparidade - Frame 0004

Este cenário teve como objetivo apresentar objetos em diferentes profundidades, com movimentos temporais em direções diferentes (barco verde e a nuvem) e que estão ausentes em alguns *frames*, dificultando a estimação de disparidades (processo que pode estar ou não presente no esquema de compressão de “light fields”). A disparidade varia de -3.6 a -0.7 (Figura 4). Note que a precisão da disparidade é fracionária, impondo desafios para estimadores de disparidade.

A Figura 5 ilustra a propriedade de mudança de foco aplicada ao vídeo “light field” para este cenário.

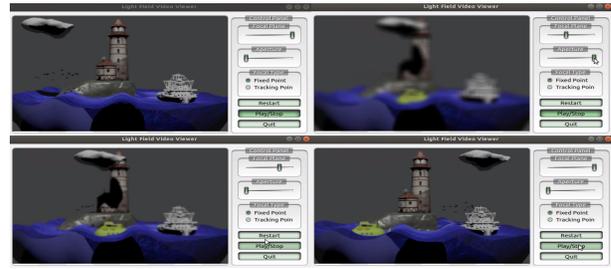


Figura 5. LightField - Variação de Foco do Vídeo

B. Cenário 2: Batman x Batman

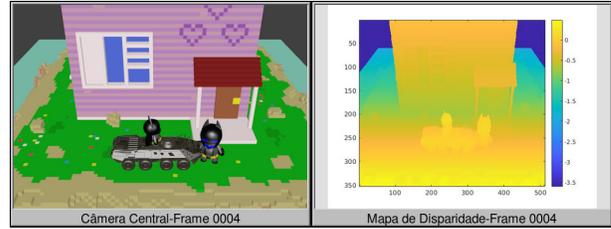


Figura 6. Visão Central e sua Disparidade - Frame 0004

A disparidade deste cenário (-3.8 a 0.6) varia significativamente na região do chão do cenário, onde o início do “gramado” (parte bege claro na Figura 6) está bem próximo à câmera virtual, enquanto a parte final está bem distante (parte ciano na Figura 6). Na casa, a disparidade varia lentamente e, do final do “gramado” para o fundo, há uma variação bastante abrupta. Somente os objetos “carro” e “mini-Batman” sofrem variação temporal.

A Figura 7 exibe a mudança de foco para o vídeo “light field” criado.



Figura 7. LightField - Variação de Foco do Vídeo

C. Cenário 3: The Airplane

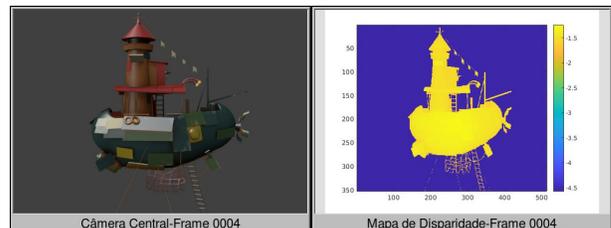


Figura 8. Visão Central e sua Disparidade - Frame 0004

Este cenário foi feito para testar qual o efeito na variação de disparidade quando se tem um único objeto com detalhes muito próximos. O avião estilizado apresenta várias regiões com disparidades (a diferentes profundidades). Apesar de se ter apenas um objeto em cena (o Airplane), esta variação de disparidade - ainda que suave e baixa - ao longo do mesmo objeto desafia a precisão de estimadores de diparidade e pode facilitar determinados esquemas de compressão de vídeo “light field”. Note que a variação mais abrupta de disparidade ocorre na transição do objeto “airplane” para o fundo, enquanto dentro do “airplane” varia lentamente. A Figura 8 ilustra a variação de disparidade (-4.7 a -1.1), em que os tons de amarelo mostram que a disparidade varia suavemente ao longo do “airplane”.

A Figura 9 mostra a mudança de foco para esta animação.



Figura 9. LightField - Variação de Foco do Vídeo

D. Cenário 4: Return to Home

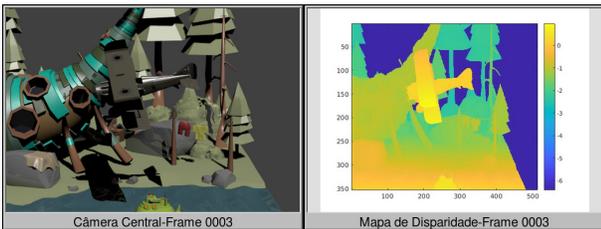


Figura 10. Visão Central e sua Disparidade - Frame 0003

O cenário 4 apresenta vários objetos com intervalo de disparidade, maior do que o do cenário 1, que varia de -6.6 a 1.1 (Figura 10), exibindo disparidades positivas e negativas. Este cenário impõe mais desafios tanto para estimadores de disparidade quanto para esquemas de compressão de “light fields”, devido à variação de disparidade (com pixels em diferentes profundidades) que gera mais áreas de oclusão, e por apresentar áreas não Lambertianas em alguns objetos.

A Figura 11, exibe a variação do foco da animação criada.



Figura 11. LightField - Variação de Foco do Vídeo

V. CONCLUSÕES

A finalidade deste artigo foi apresentar datasets sintéticos de vídeos de “light fields” com diferentes graus de disparidade e movimento. Projetos futuros visam a aumentar o número de frames dos datasets atuais, bem como de criar mais datasets, com variações de textura e graus de disparidade, para que possam ser usados na avaliação de métodos de compressão de “light fields” e de estimação de disparidade. Também há a possibilidade de utilizar plataformas de modelagem como Unity e Unreal Engine, ao invés do Blender, como pesquisadores da Google [4] fazem para VR/AR.

REFERÊNCIAS

- [1] T. Ebrahimi, S. Foessel, F. Pereira, and P. Schelkens, “Jpeg pleno: Toward an efficient representation of visual reality,” *IEEE Multimedia*, vol. 23, no. 4, pp. 14–20, 2016.
- [2] I. JTC1/SC29/WG11, “MPEG-I Visual activities on 6DoF and light fields,” 10 2017.
- [3] M. Broxton, J. Flynn, R. Overbeck, D. Erickson, P. Hedman, M. DuVall, J. Dourgarian, J. Busch, M. Whalen, and P. Debevec, “Immersive light field video with a layered mesh representation,” vol. 39, no. 4, pp. 86:1–86:15, 2020.
- [4] R. S. Overbeck, D. Erickson, D. Evangelakos, M. Pharr, and P. Debevec, “A system for acquiring, processing, and rendering panoramic light field stills for virtual reality,” *ACM Trans. Graph.*, vol. 37, no. 6, Dec. 2018. [Online]. Available: <https://doi.org/10.1145/3272127.3275031>
- [5] E. H. Adelson, J. R. Bergen *et al.*, *The plenoptic function and the elements of early vision*. Vision and Modeling Group, Media Laboratory, Massachusetts Institute of ..., 1991, vol. 2.
- [6] C. Zhang and T. Chen, “Light field sampling,” *Synthesis lectures on image, video, and multimedia processing*, vol. 2, no. 1, pp. 1–102, 2006.
- [7] S. J. Gortler, R. Grzeszczuk, R. Szeliski, and M. F. Cohen, “The lumigraph,” in *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, 1996, pp. 43–54.
- [8] M. Levoy and P. Hanrahan, “Light field rendering,” in *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, 1996, pp. 31–42.
- [9] E. H. Adelson and J. Y. Wang, “Single lens stereo with a plenoptic camera,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 14, no. 2, pp. 99–106, 1992.
- [10] J. BRADLEY. (2012) Review: Lytro light field camera. [Online]. Available: <https://www.wired.com/2012/02/lytro-camera-2/>
- [11] [Online]. Available: <http://graphics.stanford.edu/projects/lightfield>
- [12] [Online]. Available: <https://www.blender.org/>
- [13] [Online]. Available: <https://sketchfab.com/3d-models>
- [14] K. Honauer, O. Johannsen, D. Kondermann, and B. Goldluecke, “A dataset and evaluation methodology for depth estimation on 4d light fields,” in *Asian Conference on Computer Vision*. Springer, 2016, pp. 19–34.
- [15] C. G. Jinglei Shi, Xiaoran Jiang, “A framework for learning depth from a flexible subset of dense and sparse light field views,” *IEEE Transactions on Image Processing, Institute of Electrical and Electronics Engineers*, 2019.
- [16] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*. Cambridge university press, 2003.
- [17] [Online]. Available: <https://github.com/lightfield-analysis/blender-addon>
- [18] [Online]. Available: <https://imagemagick.org/index.php>
- [19] [Online]. Available: https://www.mathworks.com/?s_tid=gn_logo
- [20] T.-C. Wang, J.-Y. Zhu, N. K. Kalantari, A. A. Efros, and R. Ramamoorthi, “Light field video capture using a learning-based hybrid imaging system,” *ACM Transactions on Graphics (TOG)*, vol. 36, no. 4, pp. 1–13, 2017.
- [21] [Online]. Available: <https://harllon.github.io/LightField-Blender/>