

ESQUEMA DE COMPRESSÃO EFICIENTE DE LIGHT FIELDS BASEADO EM CODIFICAÇÃO DE VÍDEO

RUHAN CONCEIÇÃO; MARCELO PORTO; BRUNO ZATT; LUCIANO AGOSTINI

Universidade Federal de Pelotas – Video Technology Research Group (ViTech)
{radconceicao,porto,zatt,agostini}@inf.ufpel.edu.br

1. INTRODUÇÃO

Após um breve período de popularidade e forte apelo comercial, a televisão 3D (*3D Television* – 3DTV) sofreu um forte e abrupto declínio, consolidado pela decisão das principais fabricantes em não fabricar mais *displays* para esta tecnologia (KATZMAIER, 2016). Diversos fatores contribuíram para este insucesso, tais como: (1) poucos conteúdos 3D relevantes que justificassem colocar óculos 3D; (2) a disponibilidade de conteúdo 3D associou-se basicamente às mídias de *Blu-ray*, enquanto serviços de *streaming*, como o Netflix, tornavam-se cada vez mais populares; e (3) muitos usuários sentiam desconforto visual e dores de cabeça enquanto assistiam conteúdo 3D (CASS, 2014; LESWING, 2017).

Este último revés foi devido principalmente ao conflito entre vergência e acomodação (*Vergence-Accommodation Conflict* – VAC). Basicamente, vergência está relacionado à rotação do globo ocular realizada pelo sistema visual humano para convergir ambos olhos em um mesmo objeto próximo ou afastado do espectador. Acomodação visual refere-se ao esforço realizado pelo olho para “focar” em uma dada distância. Note que, conforme ilustrado na Figura 1(a), a acomodação visual acompanha a vergência realizada pelo sistema visual. Entretanto, quando assistimos a um vídeo 3D, embora a vergência varie, ambos olhos do espectador seguem focando na tela 3D, conforme ilustrado na Figura 1(b). Por conta deste comportamento insólito do sistema visual humano, os usuários tendem a sentirem-se desconfortáveis ao assistirem vídeos 3D.

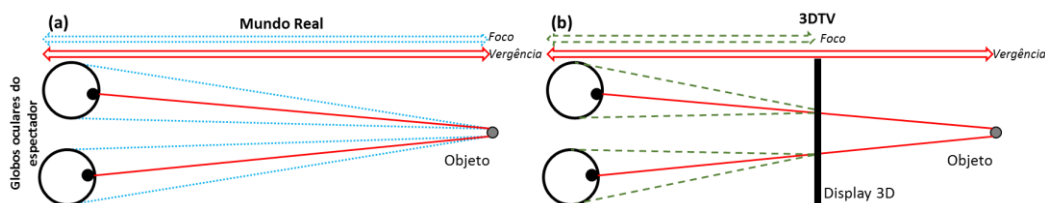


Figura 1. Relação vergência-acomodação (a) no mundo real e (b) na televisão 3D

Tendo em vista estes fatos, evidencia-se a necessidade por novas tecnologias de exibição que superem estas questões que resultaram no insucesso da 3DTV. Neste cenário surgem os *light fields* (IHRKE; RESTREPO; MIGNARD-DEBISE, 2016), os quais, baseados na função plenóptica, visam expandir a representação da informação visual em uma cena. Diferentemente das tecnologias de exibição atuais, *light fields* representam não somente a posição no plano de observação em que os raios de luz provenientes da cena incidem, mas também o ângulo em que estes atingem o plano. Para que isto seja possível, câmeras *light field*, como a *Lytro Illum* (LYTRO, 2016), empregam uma matriz de microlentes entre a lente principal da câmera e o sensor luminoso. Cada microlente gera uma microimagem (MI) que compreende os diferentes ângulos em que os raios luminosos incidem.

A Figura 2 apresenta duas formas de visualizarmos *light fields*. A primeira, ilustrada em (a), mostra cada MI do *light field* lado-a-lado, evidenciando uma MI. Nesta representação, é possível ter uma visão aproximada da cena capturada. Na Figura 2(b), temos uma outra forma de representar *light fields*, onde cada imagem (também denominada vista) representa um ângulo capturado da cena. É importante destacar que cada vista é gerada agrupando um pixel de uma determinada posição de cada MI. Note que a Figura 2(b) também evidencia a vista central do *light field*, a qual é composta pelo agrupamento do pixel central de cada MI.

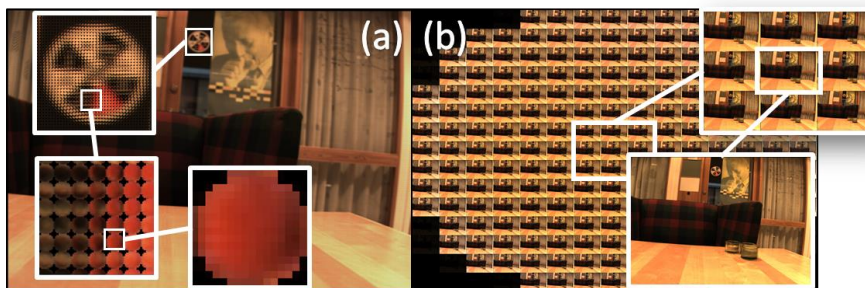


Figura 2. *Light field* agrupando (a) MIs e (b) vistas lado-a-lado

Embora propiciem um aumento expressivo na representação de uma cena, *light fields* requerem uma quantidade elevada de dados para serem representados. Por exemplo, uma cena capturada com a câmera *Lytro Illum*, a qual possui 625x434 microlentes de resolução 15x15 pixels cada, necessitaria de aproximadamente 174,61MB para ser armazenado. Além disso, um vídeo *light field* que possua 30 quadros por segundo e duração de duas horas, requisitaria um espaço em disco de 35,9TB. Estes fatos evidenciam a necessidade por técnicas eficientes de compressão de *light fields*. Tendo isto em vista, este trabalho propõe utilizar o padrão estado-da-arte de codificação de vídeos para comprimir *light fields*, visando reduzir significativamente o espaço em disco necessário para representá-los.

2. METODOLOGIA

O processo empregado neste trabalho para comprimir *light fields* utilizando codificadores de vídeo é ilustrado na Figura 3. Primeiramente, o *light field* é convertido para uma sequência de vídeo utilizando um compilador $LF \rightarrow \text{Video}$. Juntamente com a sequência de vídeo, o compilador também gera informações adicionais que relacionam cada vista do *light field* com os frames do vídeo, possibilitando sua reconstrução. Uma vez gerada a sequência de vídeo, esta é codificada utilizando o padrão *High Efficiency Video Coding* (HEVC) (ITU-T, 2014). O *bitstream* gerado juntamente com as informações adicionais poderão ser transmitidas e/ou armazenadas. Visando realizar a conversão do *bitstream* para o *light field* novamente, primeiramente o vídeo é decodificado, e então compilado novamente para o formato de *light fields*.

O compilador $LF \rightarrow \text{Video}$ pode decompor de diversas maneiras as vistas do *light field* em uma sequência de vídeo. Neste trabalho, é proposta uma forma dinâmica baseada na lista dos quadros de referência utilizada pela estimação de movimento do HEVC (SZE; BUDAGAVI; SULLIVAN, 2014). Em linhas gerais, o compilador $LF \rightarrow \text{Video}$ cria a sequência de vídeo de uma forma que toda vez em que um determinado quadro está sendo processado pelo HEVC – utilizando a predição inter-quadros – haverá um candidato similar na lista de quadros de

referência. Desta forma, o custo de codificação será reduzido, aumentando a eficiência do codificador.

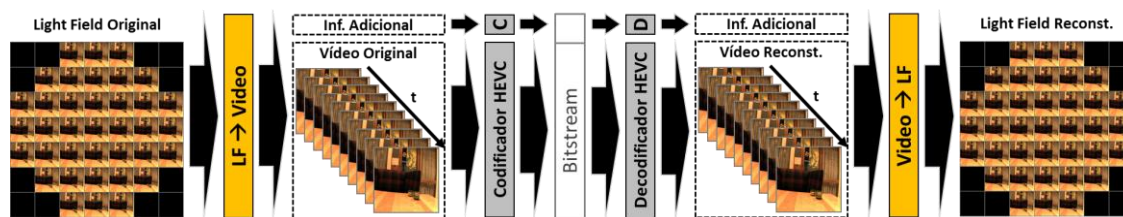


Figura 3. Processo de compressão de *light fields*

3. RESULTADOS E DISCUSSÃO

Para a realização deste trabalho, utilizou-se o software de referência do HEVC versão 16.15 (ROSEWARNE et al., 2015) e configuração *Low Delay B*. Além disso, quatro parâmetros de codificação (*Quantization Parameter* – QP) foram empregados: 22, 27, 32 e 37. Por fim, o dataset empregado foi o conjunto de 13 *light fields* disponibilizados em (PAUDYAL et al., 2016).

Para mensurar a eficiência do método empregado, implementou-se outras quatro formas de decompor o *light field* em uma sequência de vídeo, conforme ilustrados na Figura 4. É importante destacar que esses formatos são estáticos, sendo propostos em trabalhos relacionados (HARIHARAN; LANGE; HERFET, 2017; VIEIRA et al., 2015; ZHAO et al., 2016).

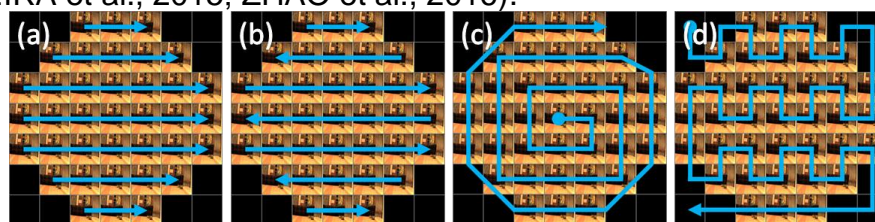


Figura 4. Formas de decompor *light fields* em uma sequência de vídeo: (a) *line*, (b) *raster*, (c) *circular* e (d) *u-shape*.

A comparação entre o trabalho proposto nesse resumo com os relacionados foi dado utilizando a métrica *BD-rate* (BJONTEGAARD, 2001). Em linhas gerais, observou-se qual a porcentagem média de bits a mais que os trabalhos relacionados necessitam para atingir a mesma qualidade de imagem comparado com o proposto neste resumo. Note que quantos menos bits para uma dada qualidade de imagem, mais eficiente é a solução de codificação. As formas *line*, *raster*, *circular* e *u-shape* de decompor *light fields* em sequência de vídeo necessitam em média de 10,97%, 10,71%, 13,08% e 19,78% mais bits para atingir a mesma qualidade de imagem da solução proposta neste trabalho. Desta forma, observa-se que a solução proposta é mais eficiente que as demais apresentadas na literatura.

4. CONCLUSÕES

Tendo em vista as restrições impostas pela 3DTV que culminaram no declínio de sua popularidade, emerge a necessidade por novas tecnologias de imagem que venham a superar todos os seus contrapontos. Neste cenário surgem os *light fields*, os quais baseados na função plenóptica, oferecem uma maior representação das cenas. Entretanto, esta nova tecnologia de imagem necessita de um espaço proibitivo em disco para ser armazenado quando não

comprimido, havendo a necessidade por estratégias eficientes de codificação. Tendo isto em vista, este trabalho propôs uma forma de decompor *light fields* em uma sequência de vídeo baseada na construção das listas de referência da estimação de movimento do HEVC. Este formato dinâmico possibilitou ganhos em termos de eficiência de codificação sobre trabalhos relacionados, que adotaram formas estáticas de converter *light fields* em sequências de vídeo.

5. REFERÊNCIAS BIBLIOGRÁFICAS

BJONTEGAARD, G. **Calculation of Average PSNR Differences Between RD Curves (VCEG-M33)**. Austin, USA: [s.n.].

CASS, S. **3-D TV is Officially Dead (For Now) and This is Why it Failed - IEEE Spectrum**. Disponível em: <<https://spectrum.ieee.org/tech-talk/consumer-electronics/audiovideo/3d-tv-is-officially-dead-for-now-and-this-is-why-it-failed>>. Acesso em: 15 set. 2017.

HARIHARAN, H. P.; LANGE, T.; HERFET, T. **Low complexity light field compression based on pseudo-temporal circular sequencing**. 2017 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB). **Anais...2017**

IHRKE, I.; RESTREPO, J.; MIGNARD-DEBISE, L. Principles of Light Field Imaging: Briefly revisiting 25 years of research. **IEEE Signal Processing Magazine**, v. 33, n. 5, p. 59–69, 2016.

ITU-T. **High efficiency video coding: Recommendation ITU-T H.265**. [s.l.: s.n.]. Disponível em: <<https://www.itu.int/rec/T-REC-H.265-201612-I/en>>.

KATZMAIER, D. **With a bullet to the head from Samsung, 3D TV is now deader than ever**. Disponível em: <<https://www.cnet.com/news/3d-tv-is-now-more-dead-than-ever/>>. Acesso em: 14 set. 2017.

LESWING, K. **3D TV is dead - Business Insider**. Disponível em: <<http://www.businessinsider.com/3d-tv-is-dead-2017-1>>. Acesso em: 14 set. 2017.

LYTRO. **Lytro Illum Technical Specifications**. [s.l.: s.n.]. Disponível em: <<https://illum.lytro.com/illum/specs>>. Acesso em: 7 jun. 2017.

PAUDYAL, P. et al. SMART: a Light Field image quality dataset Pradip. **Proceedings of the International Conference on Multimedia Systems - MMSys '16**, p. 1–6, 2016.

ROSEWARNE, C. et al. **High Efficiency Video Coding (HEVC) Test Model 16 (HM 16): Improved Encoder Description**. [s.l.: s.n.].

SZE, V.; BUDAGAVI, M.; SULLIVAN, G. **High Efficiency Video Coding (HEVC): Algorithms and Architectures**. [s.l.] Springer International Publishing Switzerland, 2014.

VIEIRA, A. et al. Data formats for high efficiency coding of Lytro-Illum light fields. **5th International Conference on Image Processing, Theory, Tools and Applications 2015, IPTA 2015**, p. 494–497, nov. 2015.

ZHAO, S. et al. **Light field image coding with hybrid scan order**. 2016 Visual Communications and Image Processing (VCIP). **Anais...nov. 2016**