

Amostragem Esparsa e Codificação VVC: Uma proposta para Compressão de Dados de *Light Fields*

João Matheus Dalmolin Montanha
Universidade Federal do Pampa
Alegrete, RS - Brasil
joaomontanha.aluno@unipampa.edu.br

Gerardo Cachina Gaetjens
Universidade Federal do Pampa
Alegrete, RS - Brasil
gerardogaetjens.aluno@unipampa.edu.br

Bruno Boessio Vizzotto
Universidade Federal do Pampa
Alegrete, RS - Brasil
brunovizzotto@unipampa.edu.br

ABSTRACT

Light field imaging has increasingly gained attention in the research community for its ability to provide a richer and more complete representation of visual scenes by capturing both spatial and angular information. However, the massive data volume generated by light field images poses significant challenges in terms of storage and transmission, particularly in bandwidth-limited or real-time environments. This work proposes a novel approach for compressing light field data by integrating Sparse Sampling techniques with the Versatile Video Coding (VVC) standard. The proposed method leverages the Discrete Cosine Transform (DCT) to generate a sparse representation of light field sub-aperture images, followed by random sparse sampling to reduce the dimensionality of the data. A reconstruction process using orthogonal matching pursuit ensures high-fidelity recovery of the light field prior to VVC encoding. The performance of the proposed solution is evaluated using samples of the HCI Dataset, and results demonstrate significant reductions in data size while maintaining visual quality in up to 38db (PSNR) and 0,95 SSIM. The integration of sparse sampling and VVC coding demonstrates promising results for light field compression, enabling its deployment in high-speed communication networks.

KEYWORDS

Ligh Fields, Sparse sampling, Data Compression

1 INTRODUÇÃO

A representação e compressão de *Light Fields* têm se tornado um tópico de crescente interesse na comunidade científica devido ao seu potencial em aplicações como realidade virtual (VR), telepresença 3D, streaming holográfico e computação visual avançada. A representação do campo de luz oferece uma descrição rica e multidimensional de uma cena, capturando não apenas as informações espaciais das imagens tradicionais, mas também suas direções angulares [1]. Essa capacidade é alcançada por meio da aquisição de subimagens de diferentes perspectivas, formando um conjunto de dados volumoso que pode alcançar centenas de *gigabytes* [2].

Contudo, o alto volume de dados gerado pelos *Light Fields* representa um grande desafio para sistemas de armazenamento, processamento e, principalmente, transmissão em redes de alta velocidade. Métodos tradicionais de compressão de imagens e vídeos, como JPEG e HEVC, não são diretamente eficientes para lidar com a redundância angular e espacial intrínseca dos *Light Fields* [3]. Para lidar com essa complexidade, técnicas modernas

de compressão esparsa e codificação adaptativa têm sido exploradas. Trabalhos recentes destacam avanços significativos na codificação de *Light Fields* com o uso de 4D-DCT [4] e arquiteturas hardware otimizadas, aplicados ao padrão JPEG Pleno e HEVC [5].

No contexto de redes de alta velocidade, como 5G e 6G, a transmissão eficiente de dados de *Light Fields* irá exigir a integração de algoritmos de compressão de dados com codificadores eficientes computacionalmente e energeticamente. O padrão VVC, desenvolvido como sucessor do HEVC, oferece ferramentas avançadas para codificação adaptativa, resultando em taxas de compressão com ganhos superiores a 50% [6]. Contudo, mesmo o VVC enfrenta desafios para lidar com a alta redundância dos *Light Fields* sem um pré-processamento eficiente.

Uma abordagem promissora é a aplicação de Amostragem Esparsa (CS - *Compressed Sensing*) combinada com DCT (Transformada de Cosseno Discreta). A CS explora a esparsidade presente na representação dos *Light Fields* no domínio da DCT, permitindo reduzir significativamente o volume de dados a serem codificados [7]. Estudos preliminares indicam que a reconstrução eficiente de subimagens, a partir de amostras comprimidas, pode ser realizada com algoritmos como *Orthogonal Matching Pursuit* (OMP) [8]. A integração dessa técnica com a codificação VVC promete uma redução de dados expressiva, sem perdas significativas na qualidade visual.

Diante desse cenário, este trabalho propõe uma solução para compressão de *Light Fields*, integrando CS e codificação VVC. O método utiliza a DCT para transformação dos dados em uma base esparsa, aplica a técnica para redução da dimensionalidade e, por fim, utiliza o codificador VVC para compactação final dos dados reconstruídos. Os experimentos foram realizados utilizando amostras do HCI *Heidelberg Light Field Dataset* [9] como referência, avaliando o desempenho em termos de taxa de compressão e qualidade da reconstrução com as métricas *Peak Signal-to-Noise Ratio* (PSNR) e *Structural Similarity Index* (SSIM).

2 SOLUÇÃO PROPOSTA

Nesta seção, é apresentada a solução proposta para a compressão de dados *Light Field*, integrando técnicas de Amostragem Esparsa com o codificador VVC. A abordagem consiste em três etapas principais: pré-processamento com DCT e Amostragem Esparsa, reconstrução aproximada e codificação com VVC. O objetivo é reduzir significativamente o volume de dados a serem transmitidos, mantendo a qualidade visual das imagens

reconstruídas. A Figura 1 apresenta o fluxograma das ações de codificação nas etapas de Pré-processamento e Reconstrução.

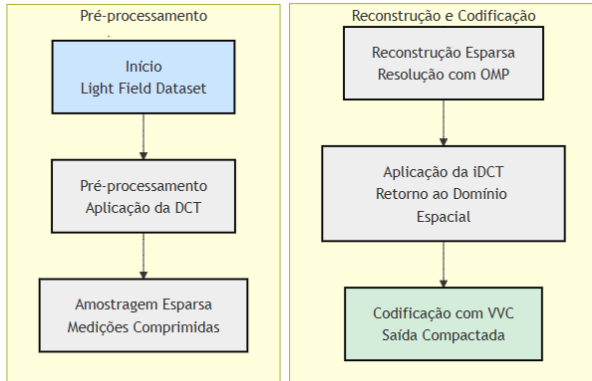


Figura 1 - Fluxograma das Etapas

1.1 Pré-processamento e CS

O pré-processamento tem como objetivo transformar os dados das subimagens do *Light Field* para uma representação mais compacta e esparsa no domínio da frequência, utilizando a DCT. Posteriormente, os dados esparsos são comprimidos, reduzindo significativamente a quantidade de informações.

1.1.1 Transformada Discreta do Cosseno:

A DCT é aplicada em cada subimagem I das amostras. A transformação é definida como:

$$DCT(I) = C \cdot I \cdot C^T$$

onde:

- I representa a subimagem original com dimensões $M \times N$.
- C é a matriz da DCT unidimensional ortonormal de tamanho M ou N .
- $DCT(I)$ é a representação esparsa da subimagem no domínio da frequência [7].

A aplicação da DCT é vantajosa, pois permite uma representação mais esparsa e compacta dos dados ao reduzir a quantidade de coeficientes necessários para a descrição da informação.

1.1.2 Amostragem Esparsa:

Após a DCT, os dados transformados são comprimidos utilizando uma matriz de amostragem aleatória Φ , conforme definido em esparsidade [8]. A CS é modelada como em [9]:

$$y = \Phi \cdot DCT(I)$$

- $\Phi \in R^{M \times N}$ uma matriz aleatória incoerente com base DCT.
- y é o vetor resultado das medições compactas com $m \ll n$.

Essa abordagem reduz drasticamente o volume de dados, garantindo que apenas informações consideradas relevantes sejam transmitidas para a etapa de reconstrução.

1.2 Reconstrução aproximada e codificação

A reconstrução dos dados consiste em recuperar a subimagem original a partir do vetor de medições comprimidas y . Isso é

realizado resolvendo um problema de otimização esparsa, que busca encontrar a representação esparsa no domínio da DCT.

1.2.1 Problema de Reconstrução Esparsa:

A reconstrução utiliza técnica de minimização da norma l_1 , no algoritmo OMP, para resolver:

$$\min_x ||x||_1 \text{ sujeito a } y = \Phi \cdot x$$

- x é o vetor esparsa no domínio da DCT que queremos recuperar.
- Φ é a matriz de amostragem esparsa.
- y é o vetor comprimido das medições [8].

A minimização da norma l_1 é utilizada porque favorece soluções esparsas, ou seja, soluções onde muitos dos coeficientes são zero ou próximos de zero.

Uma vez que a representação esparsa x é reconstruída, aplicamos a transformada inversa para retornar ao domínio espacial onde a subimagem é reconstruída. Este processo garante uma aproximação da subimagem original, com pequenas perdas introduzidas pela CS.

1.2.2 Codificação com o VVC:

A etapa final do pipeline proposto é a codificação com o padrão VVC, utilizando o codificador VVenC, uma implementação *open-source* do VVC desenvolvida pelo Fraunhofer HHI [6]. O VVC oferece ferramentas avançadas para compressão de vídeos e imagens, o *script* utilizado na codificação é representado na Figura 2.

Configuração do script VVenC para entradas de LFs

```

1. function CodVVenC (LightField[], QP, preset)
2.   foreach subimagem in LightField
3.     imagem_YUV ← Converter_YUV(subimagem)
4.     cmd ← "vvencapp -i " + imagem_YUV
5.     cmd ← cmd + " -o saida_vvc.bin"
6.     cmd ← cmd + " --preset " + preset
7.     cmd ← cmd + " --qp " + QP
8.     system(cmd)
9.   end foreach;
10. end function

```

Figura 2 – Pseudocódigo para codificação com VVenC.

O VVenC é configurado através de um conjunto de parâmetros ajustáveis, permitindo um trade-off entre taxa de compressão, qualidade visual e tempo de processamento. Os principais parâmetros relevantes ajustados no codificador incluem a Taxa de Quantização (QP), que controla o nível de compressão e qualidade da imagem, sendo que valores menores resultam em maior qualidade e menor compressão, enquanto valores maiores aumentam a compressão com perdas visuais. Além disso, o *preset* de velocidade define o equilíbrio entre eficiência e tempo de processamento, com opções como *faster*, *medium* e *slower*, onde *faster* prioriza velocidade em detrimento da eficiência, *medium* busca um compromisso entre os dois, e *slower* oferece maior

compressão com maior custo computacional. O formato de entrada utilizado é YUV 4:2:0, amplamente suportado pelo VVC, e o particionamento adaptativo dos blocos, viabilizado pelo *Multi-Type Tree*, otimiza a compressão ao explorar redundâncias presentes nos dados reconstruídos. Por fim, a saída do codificador é armazenada em um arquivo binário (.bin), que contém os dados comprimidos resultantes da codificação.

3 RESULTADOS PRÉVIOS

A análise dos resultados prévios deste estudo foi concentrada nas amostras *herbs* e *bicycle* do HCI Heidelberg Light Field Dataset [9], selecionadas por apresentarem características representativas de texturas complexas e estruturas simples, respectivamente. Essas amostras permitiram uma avaliação preliminar adequada do desempenho das métricas de qualidade (PSNR e SSIM) sob diferentes níveis de compressão.

A Figura 3 apresenta a câmera 38 da amostra *herbs* em sua reconstrução de diferentes níveis de coeficientes utilizados. Como esperado, quanto maior o número de coeficientes utilizados, mais acurácia apresenta a amostra reconstruída.

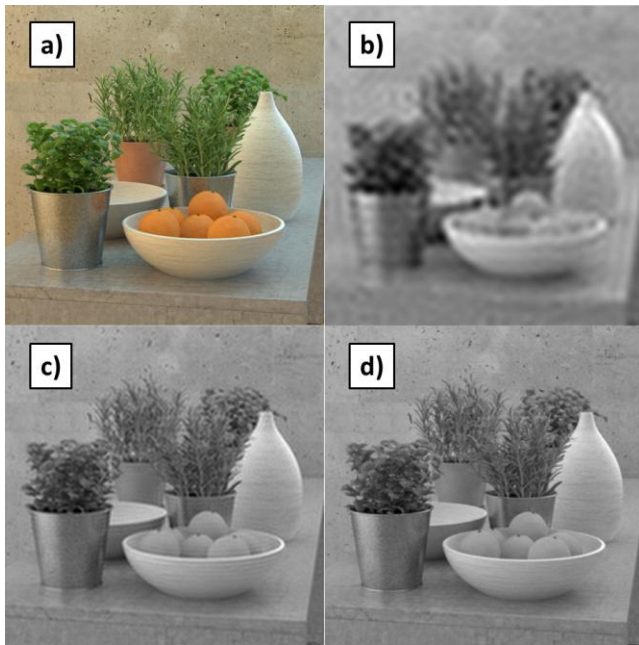


Figura 3 – *herbs* a) original; b) 10%; c) 30%; e d) 60%.

Ainda na Figura 3, a imagem a) apresenta a referência, enquanto a imagem b) demonstra degradação severa devido ao borramento, indicando a baixa qualidade de reconstrução. Já, as imagens c) e d), mostram melhor preservação dos contornos e detalhes. Enquanto c) mantém nitidez, d) apresenta uma versão mais suave, adequada para eliminar ruídos, mas com leve perda de texturas finas.

Em termos comparativos, como apresentado na Tabela 1 a compressão afeta *bicycle* mais do que *herbs* em níveis baixos (10%). Já em níveis altos (60%), ambas apresentam PSNR e SSIM elevados, evidenciando boa preservação da qualidade. A

diferença de desempenho indica que texturas complexas (*herbs*) são mais resistentes à perda de qualidade sob compressão em comparação a objetos uniformes ou estruturas simples (*bicycle*).

Coef	<i>herbs</i>			<i>bicycle</i>		
	PSNR	SSIM	T(bits%)	PSNR	SSIM	T(bits%)
10%	31,49	0,600	12,7	29,34	0,441	12,2
30%	33,44	0,813	17,3	31,10	0,771	17,8
60%	38,08	0,958	26,8	35,40	0,957	27,2

Tabela 1 – Qualidade objetiva das amostras *herbs* e *bicycle*.

A redução para menos de 30% no número de bits da amostra comprimida é extremamente relevante, porém, considerando que a etapa de codificação sobre o VVC é mais intensiva, uma escolha mais conservadora na etapa de pré-processamento pode indicar um caminho interessante.

4 CONSIDERAÇÕES FINAIS

Este estudo apresentou uma abordagem promissora para compressão de *Light Fields*. A aplicação da DCT para representação esparsa, seguida da reconstrução com OMP, permitiu reduzir o volume de dados em até 73,2%, preservando a qualidade visual com resultados de até 38 dB (PSNR) e 95% de similaridade estrutural. A solução proposta é promissora na aplicação CS para *Light Fields* em cenários como realidade virtual e *streamings* holográfico, mostrando-se favorável para sistemas modernos de comunicação e processamento. Futuramente pretende-se estender a análise sobre todo o conjunto de dados HCI, além de averiguar a demanda computacional para aplicações em codificação e decodificação em tempo real.

REFERENCES

- [1] M. Levoy and P. Hanrahan, "Light field rendering," in *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*, ACM, 1996, pp. 31–42. DOI: 10.1145/237170.237199.
- [2] R. Ng, Digital Light Field Photography. Ph.D. Dissertation, Stanford University, 2006.
- [3] D. Palomino, I. Rosler, S. Bampi, and B. Zatt, "Novel Light Field Encoding Framework Based on Optical Flow and Phase Correlation," in *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, 2022.
- [4] G. Corrêa, L. Agostini, M. Porto, and B. Zatt, "4D-DCT Hardware Architecture for JPEG Pleno Light Field Coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 9, 2020, pp. 2940–2953. DOI: 10.1109/TCSVT.2020.2996464.
- [5] R. Conceicao, M. Porto, and B. Zatt, "LF-CAE: Context-Adaptive Encoding for Lenslet Light Fields Using HEVC," *IEEE Transactions on Multimedia*, vol. 20, no. 1, 2018, pp. 155–168. DOI: 10.1109/TMM.2017.2758402.
- [6] F. Bossen, J. Boyce, X. Li, V. Seregin, and K. Stühling, "VVC Coding Tools and Performance," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 10, 2021, pp. 3736–3754. DOI: 10.1109/TCSVT.2021.3082356.
- [7] E. J. Candes and M. B. Wakin, "An Introduction to Compressive Sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, 2008, pp. 21–30. DOI: 10.1109/MSP.2007.914731.
- [8] D. L. Donoho, "Compressed Sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, 2006, pp. 1289–1306. DOI: 10.1109/TIT.2006.871582.
- [9] K. Honauer, O. Johannsen, D. Kondermann, and B. Goldlücke, "A Dataset and Evaluation Methodology for Depth Estimation on 4D Light Fields", in *Proceedings of the Asian Conference on Computer Vision (ACCV)*, Springer, 2016, pp. 19–34. DOI: 10.1007/978-3-319-54187-2_2