

REDUÇÃO DO TEMPO DE CODIFICAÇÃO DOS MAPAS DE PROFUNDIDADE NO 3D-HEVC BASEADO EM APRENDIZADO DE MÁQUINA

MÁRIO SALDANHA¹; GUSTAVO SANCHEZ²; CÉSAR MARCON³;
LUCIANO AGOSTINI¹

¹*Universidade Federal de Pelotas (UFPel) – Video Technology Research Group (ViTech)*

²*Instituto Federal Farroupilha (IFFar)*

³*Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS)*

{mrdfsaldanha, agostini}@inf.ufpel.edu.br, gustavo.sanchez@iffarroupilha.edu.br,
cesar.marcon@pucrs.br

1. INTRODUÇÃO

O 3D-*High Efficiency Video Coding* (3D-HEVC) (TECH, 2016) foi desenvolvido pelos *experts* do *Joint Team Collaborative Team 3D Video Coding* (JCT-3V) como uma extensão do padrão *High Efficiency Video Coding* (HEVC) para a codificação de vídeos 3D. O 3D-HEVC foi desenvolvido visando alcançar alta eficiência de codificação de vídeos 3D, especialmente, em aplicações com sistemas autoestereoscópicos, que necessitam de uma grande densidade de vistas em sua aplicação. Um dos principais fatores que permitem o 3D-HEVC alcançar uma alta eficiência de codificação é a utilização do formato de dados *Multiview plus Depth* (MVD) (MULLER, 2013), onde cada quadro de textura está associado a um mapa de profundidade. Os mapas de profundidade fornecem informações geométricas da cena e permitem a geração de vistas virtuais (ou vistas sintetizadas) após o processo de decodificação do vídeo. A geração dessas vistas virtuais é realizada através de técnicas de síntese de vistas que possuem baixo custo computacional. Sendo assim, apenas um subconjunto das vistas de textura, juntamente com os seus mapas de profundidade associados, é capturado e manipulado nos processos de codificação e decodificação do 3D-HEVC, impactando em uma redução expressiva no tamanho final do vídeo.

No entanto, os sistemas de codificação e decodificação do 3D-HEVC realizam um alto esforço computacional devido à inserção dos mapas de profundidade. A codificação dos mapas de profundidade é realizada através de uma estrutura flexível de particionamento, conhecida como *quadtree*, onde cada quadro é particionado em *Coding Tree Units* (CTUs) e cada CTU pode ser particionada recursivamente em Unidades de Codificação (UCs) com tamanhos variando de 8x8 até 64x64 amostras. Cada UC pode conter uma, duas ou quatro Unidades de Predição (UPs). Além disso, para cada possível particionamento de bloco diversos modos de predição (intraquadro, interquadros e intervistas) são avaliados através do processo complexo *Rate-Distortion Optimization* (RDO), que avalia diversas combinações de particionamento de bloco e modos de predição visando alcançar a melhor possibilidade de codificação. Esse processo permite alcançar alta eficiência de codificação, mas aumenta significativamente o esforço computacional do codificador.

Este trabalho apresenta uma solução capaz de reduzir o tempo de codificação dos mapas de profundidade no 3D-HEVC utilizando técnicas de mineração de dados e aprendizado de máquina para extrair atributos relevantes do processo de codificação e construir árvores de decisão estáticas, que são capazes de definir quando uma UC deve ser dividida em UCs menores para evitar o processo complexo RDO. Resultados experimentais demonstraram que a solução proposta é capaz de reduzir o tempo de codificação do 3D-HEVC em até 50% mantendo a mesma eficiência de codificação.

2. METODOLOGIA

Análises experimentais foram realizadas para avaliar o impacto do esforço computacional utilizado na codificação dos mapas de profundidade em um codificador 3D-HEVC e a distribuição dos tamanhos de UCs selecionadas para a codificação dos mapas de profundidade em diferentes cenários. Os experimentos foram realizados utilizando o software de referência 3D-HTM 16.0, considerando as Condições Comuns de Testes (CCT) (MULLER, 2014) que definem oito sequências de vídeos 3D e quatro pares de parâmetros de quantização (*Quantization Parameter – QP*) para as avaliações.

A Figura 1(a) apresenta a distribuição do tempo de codificação no 3D-HEVC, considerando os resultados médios para as sequências de teste e os pares de QPs. A Figura 1(b) apresenta a distribuição dos tamanhos de UCs escolhidas durante a codificação dos mapas de profundidade.

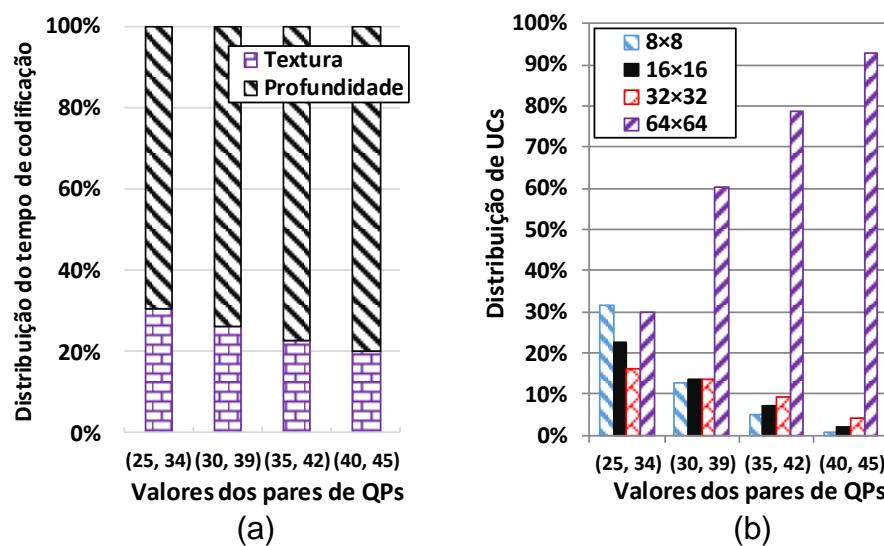


Figura 1 – (a) Distribuição do tempo de codificação do codificador 3D-HEVC e (b) distribuição dos tamanhos de UCs dos mapas de profundidade.

Analizando a Figura 1(a) é possível notar que os mapas de profundidade representam a maior parte do tempo de codificação no 3D-HEVC para todos os cenários avaliados. Considerando o par de QP (40, 45), 80% do tempo de codificação é utilizado para a codificação dos mapas de profundidade e 20% para a textura. Em média a codificação dos mapas de profundidade demanda um tempo 2,9 vezes maior do que a codificação dos dados de textura. Essa avaliação demonstra que os mapas de profundidade têm um grande impacto na complexidade total do codificador 3D-HEVC e soluções para reduzirem o esforço computacional são necessárias.

Analizando a Figura 1(b) é possível notar que o valor do QP afeta diretamente a distribuição dos tamanhos de UCs. Isso ocorre, pois, o QP define a taxa de compressão, influenciando na qualidade da imagem. Valores altos de QPs geram mais regiões homogêneas na imagem codificada, que são eficientemente codificadas utilizando UCs maiores. Contudo, para valores menores de QPs o codificador preserva os detalhes da imagem utilizando tamanhos de UCs menores para alcançar maior eficiência de codificação.

A partir das análises apresentadas é possível concluir que um algoritmo que seja capaz de decidir antecipadamente se o codificador deve evitar a avaliação de UCs menores para a codificação dos mapas de profundidade pode fornecer uma redução expressiva no tempo de codificação, sem gerar perdas na eficiência de

codificação. Além disso, outras análises realizadas durante esse estudo demonstraram que soluções especializadas para quadros I (utilizam apenas predição intraquadro) e soluções especializadas para quadros P e B (exploram interdependências) podem alcançar uma eficiência de codificação mais expressiva.

Então, este trabalho propõe utilizar mineração de dados e aprendizado de máquina para identificar atributos que possuem alta correlação com a decisão de partitionar a UC e, assim, realizar o treinamento de árvores de decisão. A utilização de mineração de dados e aprendizado de máquina permite descobrir correlações entre o contexto de codificação e seus atributos. Se essas correlações são suficientemente fortes, é possível definir árvores de decisão eficientes para reduzir o esforço computacional do codificador mantendo alta eficiência de codificação.

Uma grande quantidade de dados durante a codificação do vídeo foi coletada para realizar o treinamento das árvores de decisão. As árvores de decisão foram treinadas a partir dos dados extraídos da sequência *Kendo*. A solução foi definida com um conjunto de três árvores de decisão para os quadros I, e um conjunto de três árvores de decisão foram definidos para os quadros P e B. Essas árvores de decisão são capazes de decidir para tamanhos de UCs 16×16, 32×32 e 64×64 quando tamanhos menores devem ser avaliados ou não durante a codificação, assim evitando o processo custoso RDO.

3. RESULTADOS E DISCUSSÃO

As árvores de decisão propostas foram implementadas no 3D-HTM 16.0 e avaliadas seguindo as recomendações das CCTs. A Tabela 1 apresenta os resultados da solução em comparação ao codificador original em termos de *Bjontegaard Delta-bitrate* (BDBR) (BJONTEGAARD, 2001), considerando as vistas sintetizadas (vistas virtuais). O BDBR permite avaliar de forma objetiva a eficiência de compressão para uma mesma qualidade de vídeo, onde valores positivos indicam que são necessários mais bits para representar o vídeo em uma mesma qualidade visual. Além disso, também são apresentados os resultados de redução do tempo de codificação (RTC) do 3D-HTM. Os resultados médios não consideram a sequência de vídeo *Kendo* que foi utilizada para o treinamento das árvores de decisão.

Tabela 1 – Avaliação da solução proposta com as CCTs.

Vídeo	Solução proposta		(MORA, 2014)		(LEI, 2016)	
	BDBR	RTC	BDBR	RTC	BDBR	RTC
<i>Balloons</i>	0,06%	47,7%	2,03%	48,2%	4,32%	39,7%
<i>Kendo</i>	-	-	1,65%	48,5%	1,20%	33,9%
<i>Newspaper</i>	0,05%	48,1%	1,40%	52,3%	3,43%	43,0%
<i>GT_Fly</i>	0,07%	52,2%	0,33%	50,3%	0,87%	45,1%
<i>Poznan_Hall2</i>	0,39%	59,4%	2,01%	58,6%	3,31%	43,3%
<i>Poznan_Street</i>	0,10%	57,0%	0,05%	56,5%	1,22%	45,4%
<i>Undo_Dancer</i>	0,55%	53,3%	0,37%	50,7%	1,67%	48,3%
<i>Shark</i>	0,01%	51,4%	0,35%	50,8%	0,48%	33,3%
Média	0,18%	52,7%	1,02%	52,0%	2,06%	41,5%
<i>Kendo</i>	0,03%	47,1%	-	-	-	-

A solução proposta alcançou uma redução no tempo total de codificação de 52.7% em média (variando entre 47.7% e 59.4%) com um impacto médio no BDBR de apenas 0.18% (variando entre 0.01% e 0.55%). Além disso, a solução

foi comparada com dois trabalhos, que também focaram em reduzir a complexidade da codificação dos mapas de profundidade no 3D-HEVC. O trabalho de (MORA, 2014) alcança uma redução no tempo de codificação de 52,0% com uma perda na eficiência de codificação de 1,02%. Em (LEI, 2016), uma redução no tempo de codificação de 41,5% foi alcançada, com um impacto de 2,06% no BDBR. Portanto, a solução proposta neste trabalho alcançou a maior redução do tempo de codificação com a menor perda na eficiência de codificação, quando comparado aos trabalhos relacionados.

4. CONCLUSÕES

Este trabalho apresentou uma solução rápida para a codificação dos mapas de profundidade no 3D-HEVC, que utiliza técnicas de mineração de dados e aprendizado de máquina para extrair correlações entre os atributos do codificador e para construir árvores de decisão capazes de definir o particionamento das UCs, ao invés de avaliar os particionamentos através do processo complexo RDO. Após a codificação de um determinado tamanho de UC, uma árvore de decisão é responsável por decidir se a UC deve ser dividida em tamanhos menores. As árvores de decisão foram divididas em dois grupos, sendo três árvores responsáveis por avaliar as UCs em quadros I e três árvores responsáveis por decidir o particionamento das UCs em quadros P e B, já que o codificador possui um comportamento diferente entre esses tipos de quadros, necessitando de árvores especializadas. Resultados experimentais demonstraram que a solução proposta pode reduzir pela metade o tempo de processamento do codificador 3D-HEVC, com baixo impacto na eficiência de codificação. Além disso, a solução proposta supera os trabalhos relacionados considerando redução no tempo e eficiência de codificação.

5. REFERÊNCIAS BIBLIOGRÁFICAS

- BJONTEGAARD, G. **Calculation of Average PSNR Differences between RD-Curves**. VCEG Meeting, pp. 1-5, 2001.
- LEI, J. et al. Fast Mode Decision Based on Grayscale Similarity and Inter-View Correlation for Depth Map Coding in 3D-HEVC. **IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)**, v. 28, n. 3, pp. 706-718, Mar. 2016.
- MORA, E., et al. Initialization, Limitation, and Predictive Coding of the Depth and Texture Quadtree in 3D-HEVC. **IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)**, v. 24, n. 9, pp. 1554-1565, Set. 2014.
- MULLER, K. et al. 3D High-Efficiency Video Coding for multi-view video and depth data. **IEEE Transactions on Image Processing (TIP)**, vol. 22, n. 9, pp. 3366-3378, Set. 2013.
- MULLER, K., VETRO, A. **Common Test Conditions of 3DV Core Experiments**, ISO/IEC JTC1/SC29/WG11 MPEG2011/N1274, Jan. 2014.
- TECH, G. et al. Overview of the Multiview and 3D extensions of High Efficiency Video Coding. **IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)**, v. 26, n. 1, pp. 35-49, Jan. 2016.