

TÉCNICA PARA COMPRESSÃO INTRA-QUADRO DO PADRÃO DE CODIFICAÇÃO DE VÍDEO HEVC QUE EXPLORA DISTORÇÕES GERADAS PELA PROJEÇÃO ERP EM VÍDEOS 360º.

BERNARDO BELING¹; IAGO STORCH²; DANIEL PALOMINO¹

¹*Universidade Federal de Pelotas – Video Technology Research Group (ViTech)
{berbeling, dpalomino}@inf.ufpel.edu.br*

²*Universidade Federal do Rio Grande do Sul – icstorch@inf.ufrgs.edu.br*

1. INTRODUÇÃO

Vídeos omnidirecionais, também conhecidos como vídeos 360º, possibilitam que o usuário tenha o controle do ângulo de visão da cena sendo apresentada, causando uma sensação de imersão ao vídeo. Segundo a (Quastmedia, 2018) vídeos 360º possuem 42% de chance a mais de serem visualizados e compartilhados comparado a vídeos convencionais, servindo como uma grande ferramenta para fins de *marketing online*. A fim de oferecer uma sensação mais imersiva, vídeos 360º geralmente são representados em resoluções 4K (3840x2160) e 8K (8192x4096), oferecendo uma grande qualidade visual. No entanto, um vídeo digital em seu formato original requer uma grande quantidade de dados para ser transmitido ou armazenado: um vídeo de resolução 4K apresentado a 60 quadros por segundo e com duração de 2 minutos, possuindo três canais de cores por pixel onde cada canal de cor é representado com 1 byte, possui tamanho de 179,15 GB (3840x2160x60x120x3). Em vista disso, o processo de compressão de vídeo que busca reduzir a quantidade de dados necessários para representar um vídeo digital com o mínimo de perdas de qualidade possível, se faz extremamente necessário, pois fazer o uso desses vídeos em seu formato original se torna inviável.

Tendo em vista que um vídeo é uma sequência de imagens (quadros) apresentados à uma determinada taxa por segundo para apresentar continuidade visual, pixels e quadros vizinhos tendem a ser muito semelhantes. Com isso, a compressão de vídeo é realizada utilizando padrões de codificação que buscam explorar redundâncias no sinal de vídeo, sejam essas redundâncias entre pixels (espaciais) ou entre quadros (temporais), a fim de representar o conteúdo do vídeo digital com menos dados. Dentre os padrões de codificação consolidados, o *High Efficiency Video Coding* (HEVC) possui uma grande performance comparado ao seu antecessor H.264, sendo capaz de representar um vídeo com mesma qualidade que o H.264 porém com 35% menos dados (CIANET, 2015).

Uma das técnicas de compressão do padrão HEVC é a compressão Intra-quadro, que busca redundâncias entre pixels e regiões vizinhas em um mesmo quadro. Para realizar essa compressão, o codificador divide um quadro em blocos quadrados denominados *Coding Tree Units* (CTUs), que é a unidade de codificação básica do HEVC. Normalmente, CTUs possuem tamanho de 64x64 pixels e são particionadas através de uma estrutura *quadtree* em sub-blocos denominados *Coding Units* (CUs). Inicialmente, uma CU possui o mesmo tamanho que sua CTU, e então é particionada em tamanhos 32x32, 16x16 e 8x8 pixels até incorporarem *Prediction Units* (PUs) que podem possuir tamanhos até 4x4 pixels. O tamanho final de uma PU é selecionado conforme a complexidade da textura que a sua CTU está englobando, e é escolhido através de um cálculo com alto custo computacional que testa todos os tamanhos de PU até encontrar o tamanho mais promissor.

Para a compressão de vídeos 360º, o padrão HEVC não possui um suporte nativo para vídeos esféricos, portanto, para realizar a compressão de um vídeo esférico utilizando o padrão HEVC, é necessário convertê-lo em um vídeo 2D para então codificá-lo normalmente. Em vista disso, a (JVET, 2018) propôs uma biblioteca para padrões de codificação de vídeo contendo diversas projeções para vídeos 360º. Dentre elas, a *Equirectangular Projection* (ERP) é a mais utilizada, pois consiste em converter coordenadas polares de uma esfera em coordenadas de latitude e longitude em um plano, demandando pouco esforço computacional. Porém, essa projeção acaba gerando distorções de imagem em regiões distantes do equador, pois é necessário que haja um esticamento de pixels para preencher o plano em regiões que o raio da esfera é menor que o comprimento do plano. Além disso, essa distorção se torna cada vez mais visível a medida que o ângulo da latitude aumenta. Em consequência disso, foi demonstrado em (STORCH, 2019) que na codificação Intra-quadro de vídeos 360º ERP utilizando o padrão HEVC, a taxa de ocorrência de PUs de tamanho 64x64 e 32x32 pixels é muito mais elevada em regiões polares, comparado aos outros tamanhos de PU. Isso se dá, pois o HEVC tende a selecionar tamanhos de bloco maiores para regiões com texturas homogêneas, e devido as distorções geradas pela ERP, regiões homogêneas são criadas especialmente na parte superior e inferior de um vídeo 360º, que correspondem às coordenadas polares.

Com isso, é possível explorar essa característica da projeção ERP de modo a implementar uma técnica no codificador do padrão HEVC que favoreça a seleção de tamanhos de bloco maiores em regiões muito distorcidas, buscando evitar testes custosos de tamanhos menores de bloco reduzindo a complexidade de codificação com o mínimo de perdas de qualidade possível.

2. METODOLOGIA

Tendo em vista que este trabalho procura elaborar uma técnica que explore a seleção de blocos maiores em regiões muito distorcidas de vídeos 360º ERP utilizando a compressão Intra-quadro do padrão HEVC, a seguinte técnica é proposta: um valor que representa a quantidade de distorção na região em que uma CTU se encontra é comparado com um valor que representa a tendência de profundidade daquela CTU, dessa forma, a CTU só será particionada em tamanhos menores caso o valor de distorção não seja tão significativo, dificultando a ocorrência de CTUs menores em regiões muito distorcidas.

Buscando encontrar um valor que quantifique a distorção em uma região do quadro codificado, o fator de escala da projeção ERP, que mapeia a quantidade de distorção em função da latitude (SNYDER, 1987) foi utilizado. Esse valor é dado por $k=\sec(\theta)$, onde θ é a latitude. Além disso, como o valor k para a latitude 0º ($\sec(0^\circ)=1$) representa o valor de escala real da projeção, pois não há distorções nessa região, caso o valor de escala real seja subtraído de qualquer valor k calculado, resultará no esticamento (E) que um pixel projetado sofreu. Por exemplo: no equador $E=\sec(0^\circ)-1=0$ mostrando que um pixel nessa latitude será esticado em 0 unidades, já em 60º temos $E=\sec(60^\circ)-1=1$ mostrando que na latitude 60º um pixel é esticado em 1 unidade. Sendo assim, o valor $E=k-1$ será utilizado como indicador de distorção da latitude em que uma CTU se encontra.

A seguir, a fim de obter um valor que represente a tendência de profundidade de uma CTU, foi criada uma estrutura que agrupa um número de quadros visando realizar uma média de profundidade (M) de CTUs com mesmas coordenadas, a medida que os quadros agrupados são codificados. Essa estrutura foi implementada no software de referência *HEVC Test Model 16.16*

(HM-16.16) que utiliza a linguagem de programação C++, nele, valores de profundidade de PU são representados por 0, 1, 2, 3 e 4 para tamanhos 64×64 , 32×32 , 16×16 , 8×8 e 4×4 pixels respectivamente. A partir disso, foi adotado que a profundidade máxima de uma CTU condiz com o menor tamanho de PU dentro daquela CTU. Com isso, cada CTU terá um valor M calculado a partir da profundidade máxima dessa mesma CTU dentro do conjunto de quadros agrupados. Por exemplo: caso a primeira CTU do quadro 1 tenha profundidade máxima 3, e no segundo quadro essa mesma CTU tenha profundidade máxima 2, no quadro 3 essa CTU terá uma média de profundidade $M = (3+2)/2 = 2,5$. Tendo em vista que a parte inteira de M (M_{INT}), representa números de profundidades utilizados pelo HM-16.16, e que a parte decimal de M (M_{DEC}), é um número que varia conforme profundidades anteriores daquela CTU, M_{DEC} será utilizado como indicador de tendência de profundidade de uma CTU.

Por fim, após calcular os valores E e M da CTU codificada, a comparação entre esses dois valores irá determinar a profundidade máxima permitida durante o particionamento daquela CTU. Essa profundidade será igual a M_{INT} caso $E \geq M_{DEC}$, ou $M_{INT} + 1$ caso $E < M_{DEC}$. Porém, como E assume valores entre $[0, +\infty)$, e M_{DEC} assume valores entre $[0,1]$, há uma disparidade entre a grandeza desses valores. Com isso, valores $E > 1$ foram normalizados como 1 utilizando a função $\min(\sec(\emptyset) - 1, 1)$, fazendo com que E assuma valores entre $[0,1]$. Além disso, utilizando essa abordagem é possível escalar E de tal forma que seu intervalo varie até um valor T , permitindo uma flexibilidade na comparação de valores E e M_{DEC} . Sendo assim, a função esticamento implementada neste trabalho ficou definida como: $E = [\min(\sec(\emptyset) - 1, 1)] \times T$, sendo $T \leq 1$.

3. RESULTADOS E DISCUSSÃO

Aplicando o cálculo de E e M descritos na metodologia no software de referência HM-16.16, foram realizadas codificações experimentais de 4 vídeos 360º testando valores distintos de T , a fim de observar o valor que resulta numa melhor eficiência de codificação. Sendo assim, a Tabela 1 apresenta o resultado de redução de tempo de codificação (RT), Bjøntegaard Delta Rate (BD-Rate) que corresponde a uma métrica que avalia perda de qualidade em um vídeo, e a razão entre Redução de Tempo e BD-Rate (RT/BD) para três valores T distintos.

Ao observar a tabela, nota-se que o valor $T=0.6$ obteve o melhor valor RT/BD, isto é, obteve o melhor balanceamento entre redução de complexidade e perdas de qualidade. Com isso, este valor T foi utilizado para realizar simulações padronizadas a fim de possibilitar a comparação da técnica proposta com os trabalhos (WANG, 2017), (ZHANG, 2019) e (STORCH, 2019) que também realizam técnicas de redução de complexidade para vídeos 360 ERP. Vale ressaltar que como os trabalhos comparados não codificaram exatamente os mesmos conjuntos de vídeos, nossa técnica foi implementada com o mesmo conjunto de vídeos codificados em cada trabalho, a fim de realizar uma comparação justa.

Vídeo 360	T = 0.5			T = 0.6			T = 0.7		
	RT	BD-Rate	RT/BD	RT	BD-Rate	RT/BD	RT	BD-Rate	RT/BD
Aerial City	22%	0,08%	277,2	22%	0,09%	242,2	23%	0,11%	206
Pole Vault	16%	0,03%	496,7	17%	0,03%	628,4	18%	0,03%	549,1
Driving in City	27%	0,19%	145,2	27%	0,19%	141,6	27%	0,18%	151,8
Skateboard in Lot	30%	0,23%	129,1	30%	0,23%	133,8	31%	0,23%	131,6
Média	24%	0,13%	262	24%	0,13%	286,5	25%	0,14%	259,6

Tabela 1. Resultados experimentais para valores T distintos

Trabalho	Redução de Tempo	BD-Rate	RT/BD
Wang	24.5%	0.24%	102.1
Nosso	22.0%	0.09%	244.4
Zhang	53%	1.3%	40.8
Nosso	24%	0.11%	218.2
Storch (a)	11.2%	0.30%	37.3
Storch (b)	15.2%	0.40%	38.0
Nosso	25%	0.08%	320.1

Tabela 2. Comparação da técnica proposta com trabalhos relacionados

Os resultados de comparação são apresentados na Tabela 2, onde é possível observar que a técnica proposta neste trabalho obteve a melhor razão entre redução de tempo de codificação e perda de qualidade em relação aos trabalhos comparados.

4. CONCLUSÕES

Este trabalho apresentou uma técnica que explora as distorções geradas pela projeção ERP na codificação de vídeos 360º. Os resultados obtidos com a técnica demonstraram uma redução de tempo de codificação considerável ao custo de pouca perda de qualidade. Além disso, este trabalho foi o único entre os comparados anteriormente que utilizou de forma matemática características da projeção ERP. Devido a isso, a técnica possui uma abordagem inovadora e se mostrou superior a todos trabalhos relacionados comparados.

5. REFERÊNCIAS BIBLIOGRÁFICAS

- YE, Y.; ALSHINA, E.; BOYCE, J. JVET-E1003: Algorithm descriptions of projection format conversion and video quality metrics in 360Lib. In: **Joint Video Exploration Team ITU-T SG16 WP3 ISO/IECJTC 1/SC 29/WG 11 5th Meet**, Geneva, 2017.
- QUASTMEDIA, **The 360 Degree Video Revolution**. 8 jan. 2018. Acessado em 25 set. 2020. Online. Disponível em: <https://www.quastmedia.com/single-post/2018/01/08/The-360-Degree-Video-Revolution>.
- Cianet. **Saiba o que é o novo formato H.265**. Acessado em 25 set. 2020. Online. Disponível em: <https://www.cianet.com.br/blog/inovacao-e-tendencias/saiba-o-que-e-o-novo-formato-h-265/>
- STORCH, I.; CRUZ, L.; AGOSTINI, L.; ZATT, B.; PALOMINO, D. The Impacts of Equirectangular 360-degrees Videos in the Intra-Frame Prediction of HEVC. **Journal of Integrated Circuits and Systems**, Online, v.14, n.1, 2019
- SNYDER, J.P. **Map Projections: A working manual (U.S. Geological Survey Professional Paper 1395)**. Washington: United States Government Printing Office, 1987.
- WANG, Y.; LI, Y.; YANG, D.; CHEN, Z. A fast intra prediction algorithm for 360-degree equirectangular panoramic video. **IEEE Visual Communications and Image Processing (VCIP)**, St. Petersburg, FL, pp. 1-4, 2017.
- ZHANG, M.; DONG, X.; LIU Z. et al. Fast intra algorithm based on texture characteristics for 360 videos. **EURASIP Journal on Image and Video Processing**, 2019.
- STORCH, I.; ZATT, B.; AGOSTINI, L.; CRUZ, L.; PALOMINO, D. **FastIntra360: A Fast Intra-Prediction Technique for 360-Degrees Video Coding**. **Data Compression Conference (DCC)**, Snowbird, UT, USA, pp. 605-605, 2019.