

## AVALIAÇÃO DO USO DA DCT-3D EM CODIFICADORES DE VÍDEO

**MATEUS MELO; JONES GOEBEL; BRUNO ZATT; MARCELO PORTO**

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

### 1. INTRODUÇÃO

Mesmo com o avanço tecnológico ocorrido nos últimos anos, seja por velocidade de transmissão na internet, capacidade de processamento dos dispositivos, ainda não é possível viabilizar a manipulação e armazenamento de vídeos sem compressão, devido especialmente à largura de bits a serem transmitidos e ao gargalo na comunicação com a memória, referente a este grande volume de dados. Deste modo, codificadores de vídeos são essenciais para o sucesso de aplicações que capturam e manipulam vídeos digitais, de modo a reduzir o número de bits necessários para representá-los, ou seja, visam uma elevada compressão dos vídeos, entretanto preocupando-se com que não haja grandes perdas em termos de qualidade visual.

Um vídeo sem compressão possui diversos tipos de redundâncias de informações, na qual codificadores buscam explorá-las de forma a maximizar a eficiência de codificação. A Redundância Espacial é a similaridade presente em áreas homogêneas em um quadro, Redundância Temporal é similaridade entre quatro temporalmente vizinhos, e Redundância Entrópica é a presente na representação do vídeo na forma de bitstream, e relacionada com a probabilidade de ocorrência dos símbolos a serem codificados.

Os codificadores de vídeos utilizam diversas ferramentas de modo a elevar a taxa de compressão de dados, tais como: Predições Intraquadro e Interquadros, seguidos pela Transformada, Quantização e Codificação de Entropia. A Predição Intraquadro é utilizada para identificar redundâncias espaciais. A Predição Interquadros é responsável por identificar as redundâncias temporais. A Transformada é responsável por transformar os erros de predições para o domínio de frequência, após, a Quantização é utilizada para reduzir a amplitude destes resíduos transformados, considerando características do sistema visual humano, e facilitando a Codificação de Entropia, responsável pela geração do bitstream.

No fluxo de codificação, os codificadores dividem o vídeo sem compressão em quadros, e posteriormente, dividem cada quadro em blocos de tamanhos menores, para aplicarem as etapas de predição e codificação em cada um destes blocos (sinal 2D). Entretanto, a extensão da DCT-2D (Transformada Discreta dos Cossenos) – comumente utilizada na etapa de Transformada – para a DCT-3D pode trazer algumas vantagens, como por exemplo eliminar problemas causados pela Estimação de Movimento, presente na Predição Interquadros, em vídeos com alta movimentação de objetos (caso não utilizado uma predição mais sofisticada), e na presença de movimentos não-translacionais (rotação, zoom, entre outros) (LEE; CHAN; ADJEROH, 1997).

Deste modo, este estudo tem como objetivo avaliar o impacto da utilização da DCT-3D, ao invés da DCT-2D, em implementações simplificadas de codificadores. São consideradas três configurações de codificadores: duas de Codificador Residual (CR), compatível com os resíduos das etapas de predição para *i*) blocos (CR\_2D), e *ii*) cubos (CR\_3D), e uma implementação que não utiliza técnicas de predição, ou seja, codifica diretamente as amostras originais do vídeo sem compressão, a qual é uma estratégia utilizada por maioria dos trabalhos encontrados na literatura para o uso de DCT-3D, chamada neste estudo, de *iii*) CE\_3D (Codificador Espacial 3D).

## 2. METODOLOGIA

No fluxo de codificação dos codificadores atuais, o vídeo não comprimido é dividido em quadros, e cada quadro em blocos para a aplicação das etapas de Predição, Transformada, Quantização e Codificação de Entropia. Este fluxo, neste trabalho, é adotado na implementação CR\_2D, ou seja, cada quadro é processado a nível de blocos de tamanho  $n \times n$ . Diferentemente deste, os avaliados codificadores 3D, tanto CR\_3D como CE\_3D, dividem o vídeo a ser codificado em sequências de  $n$  quadros, e então, divide estes para o processamento a nível de cubos de tamanho  $n \times n \times n$ . Os codificadores foram desenvolvidos com o uso da Linguagem C/C++, e em sua execução são capazes de processar um único tamanho de bloco/cubo. Sendo assim, os Codificadores Residuais são compostos pelas etapas de: Predições Intraquadro e Interquadros, Transformada e Quantização; e suas inversas, além da codificação de Entropia. Já o Codificador Espacial, é composto das mesmas etapas, exceto as de Predição.

As etapas de predição, tanto Intraquadro como Interquadros, são responsáveis por gerar os erros de predição (resíduos) à serem codificados pelas etapas seguintes no fluxo de codificação. Conforme mencionado anteriormente, as etapas de predição aqui descritas, são utilizadas somente no Codificador Residual (CR\_2D e CR\_3D). A Predição Intraquadro é aplicada apenas ao primeiro quadro do vídeo ( $Q_0$  – o qual é simplesmente descartado no Codificador Espacial), e então, utilizado como referência para o(s) próximo(s) quadros a serem processados pela predição Interquadros. A Predição Intraquadro foi aplicada utilizando o software de referência do padrão HEVC, o HM em sua versão 16.9 (KEN MCCANN et al., 2014), com suas configurações padrão e QP igual à 32.

A Predição Interquadros, utilizada neste estudo, é aplicada por meio da predição diferencial simples, a maneira mais simples de se aplicar este tipo de predição, e apresenta bons resultados em regiões estáticas (sem movimentação). Esta predição é realizada por meio da subtração entre o quadro a ser codificado e o quadro de referência, de modo a gerar os erros de predição, que posteriormente são processados pelas demais etapas no fluxo de codificação. Os resíduos são gerados considerando o último quadro codificado (que não necessariamente é o temporalmente anterior para o CR\_3D) como referência para o quadro a ser processado, ou seja, para a Predição Interquadros 3D de  $[Q_{c \times n+1}, Q_{(c+1) \times n}]$ , o quadro  $Q_{c \times n}$ , tal que  $c \in \mathbb{N}$  e múltiplo de  $n$  (para CR\_2D,  $n=1$ ).

A Transformada é responsável por transformar os resíduos/amostras do domínio espacial para o domínio de frequência, de modo a identificar os componentes de alta frequência – possuem menor grau de relevância para o sistema visual humano. De acordo com a propriedade da DCT, os sinais transformados concentram maior parte da informação em coeficientes de baixa-frequência, facilitando assim a Quantização de remover as informações de alta frequência. A propriedade de separabilidade (KHAYAM, 2003) da DCT, é uma abordagem comumente utilizada para a aplicação de transformada multidimensionais. De acordo com esta propriedade, uma transformada de um sinal  $n$ -dimensões pode ser computada em  $n$  passos de aplicações de DCT-1D, em cada direção do sinal.

A Quantização é a etapa responsável por reduzir a amplitude dos coeficientes transformados, considerando as características do sistema visual humano, bem como facilitar a codificação de entropia, em termos de redução do número de símbolos a serem codificados. Os atuais codificadores de vídeo utilizam Quantização 2D, pois utilizam Transformada 2D, entretanto para o uso de Transformada 3D (DCT-3D) é desejável a utilização de uma Quantização 3D.

Desse modo, a Quantização 3D utilizada neste trabalho foi a proposta por (LEE; CHAN; ADJEROH, 1997), a partir do estudo da distribuição de amostras transformadas. Os parâmetros utilizados ao longo deste trabalho são os mesmos do referido trabalho, entretanto aqui considerando precisão em ponto flutuante. Para manter a consistência dos resultados, a matriz de coeficientes utilizada para a Quantização 2D foi extraída da primeira profundidade do cubo de coeficientes referente à Quantização 3D.

A codificação de entropia é responsável por gerar o bitstream do vídeo, por meio do processamento os coeficientes quantizados com base na probabilidade de utilização de cada símbolo, ou seja, símbolos com maior probabilidade recebem um código de tamanho menor, atingindo maiores taxas de compressão, facilitando assim o armazenamento ou transmissão. Códigos *Exp-Golomb* (WIEGAND et al., 2003), são códigos de tamanho variável com estrutura regular, na qual o tamanho das palavras geradas cresce exponencialmente de acordo com o valor a ser codificado.

### 3. RESULTADOS E DISCUSSÃO

Os resultados aqui descritos foram simulados com vídeos presentes Condições Comuns de Teste (CTCs) do padrão HEVC (BOSSEN, 2013), considerando diferentes classes: *Classe B*: cinco vídeos de resolução 1920x1080, *Classe C*: quatro vídeos de resolução 832x480, e *Classe D*: quatro vídeos de resolução 416x240 amostras. Para as simulações, todos os codificadores consideraram  $n = 8$ .

Resultados de qualidade visual, são apresentados em termos da métrica de qualidade objetiva PSNR. O documento (BJONTEGAARD, 2001) descreve uma metodologia para calcular a diferença média entre duas condições de simulações. Desse modo, resultados em termos de BD-Rate expressam para a mesma qualidade objetiva o quanto foi salvo em termos de taxa entre duas métricas, na qual valores negativos são desejáveis, representando a redução de *bitrate* do vídeo codificado para mesma qualidade de vídeo.

Primeiramente, os coeficientes das matrizes de quantização 2D e 3D foram multiplicadas por constantes {0,1; 0,2; 0,3; 0,5; 1,0; 1,5}, de modo a selecionar quatro pontos que representam a maior área de interseção entre as curvas (duas diferentes simulações). Os vídeos simulados apresentam mesmo comportamento de curva (entre os fatores multiplicativos) conforme demonstrado pela Figura 1. As curvas cinza e amarela representam os pontos de operação para os codificadores: CR\_2D, CR\_3D, CE\_3D. Após, quatro pontos em cada curva (pontos pretos) foram selecionados para a extração de resultados em termos de BD-Rate. Os pontos selecionados foram os com fatores: {0,1; 0,2; 0,3; 0,5} para a matriz de Quantização 2D, e {0,2; 0,3; 0,5; 1,0} para a 3D.

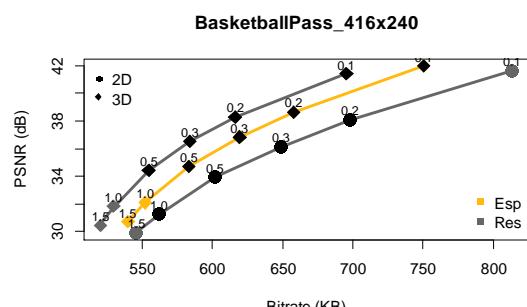


Figura 1. Curva RD para o vídeo BasketballPass

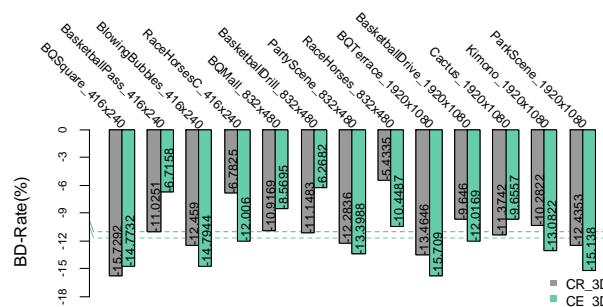


Figura 2. Resultados em termos de BD-Rate para as sequências avaliadas

De modo geral, conforme a Figura 2, os resultados demonstram que CR\_3D e CE\_3D, melhores resultados em termos de BD-Rate (menores que 0%) com referência ao CR\_2D, para todas as sequências de vídeo avaliadas. Os resultados podem ser justificados, considerando que as sequências de testes avaliadas são vídeos naturais que apresentam diversos coeficientes temporais (movimentação de cena) no domínio de frequência e ruídos de captação, assim, a Quantização 3D pode explorar eficientemente a codificação neste contexto, apresentando assim melhores resultados quando comparada à Quantização 2D – que explora esta categoria de coeficientes.

Ao comparar os codificadores 3D (residual e espacial), para maioria das das sequências de testes avaliadas, o CE\_3D (Codificador Espacial 3D) apresentou melhores resultados em termos de BD-Rate, exceto para os vídeos: *BQSquare*, *BasketballPass*, *BQMall*, *BasketballDrill* e *Cactus*. Como próximas etapas, será realizado um levantamento estatísticos dos dados ao longo do fluxo de execução, para identificar informações mais relevantes ao estudo e justificar os resultados de melhor qualidade no CR\_3D; além de investigar o uso de alguma técnica simplificada de Estimação de Movimento de modo a comparar com as atuais implementações.

#### 4. CONCLUSÕES

Este resumo, apresentou brevemente a avaliação do impacto da utilização da DCT-3D, ao invés da DCT-2D, em implementações simplificadas de codificadores, visito que esta estratégia em como objetivo contornar o não uso da técnica de Estimação de Movimento no habitual fluxo de codificação. Com os resultados apresentados anteriormente, é possível verificar que o uso da DCT-3D apresenta melhorias significativas em eficiência de codificação, tanto utilizando Codificador Espacial como Residual, quando comparados Codificador Residual utilizando DCT-2D, demonstrando que DCT-3D auxilia em uma melhor codificação dos coeficientes temporais, de acordo com a proposta do uso desta.

#### 5. REFERÊNCIAS BIBLIOGRÁFICAS

- BOSSEN, F. Common test conditions and software reference configurations, Document JCTVC-L1100, JCTVC of ITU-T SG16 WP3 and ISO. IEC JTC1/SC29/WG11, 2013.
- BJONTEGAARD, G. Calculation of average PSNR differences between RD-curves. ITU-T Q. 6/SG16 VCEG, 15th Meeting, Austin, Texas, USA, April, 2001.
- KHAYAM, Syed Ali. The discrete cosine transform (DCT): theory and application. Michigan State University, v. 114, 2003.
- LEE, M. C.; CHAN, R. K. W.; ADJEROH, D. A. Quantization of 3D-DCT Coefficients and Scan Order for Video Compression. **Journal of Visual Communication and Image Representation**, v. 8, n. 4, p. 405–422, 1997.
- MCCANN, Ken et al. High efficiency video coding (HEVC) test model 16 (HM 16) improved encoder description. Joint Collaborative Team on Video Coding, JCTVC-S1002, Strasbourg, FR, 2014.
- WIEGAND, T. et al. Overview of the H.264/AVC video coding standard. **IEEE Transactions on Circuits and Systems for Video Technology**, v. 13, n. 7, p. 560–576, jul. 2003.