

## ANÁLISE DE UM CODIFICADOR DE VÍDEO BASEADO EM TRANSFORMADA TRIDIMENSIONAL

MATEUS MELO; JONES GOEBEL; IGOR ROSLER;  
DANIEL PALOMINO; MARCELO PORTO; BRUNO ZATT

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

### 1. INTRODUÇÃO

A popularização de vídeos digitais em plataformas de compartilhamento, videoconferência e redes sociais têm tornado as tarefas de codificação e decodificação de vídeo cada vez mais frequentes, em especial em dispositivos móveis alimentados por bateria. Mesmo com o avanço tecnológico ocorrido nos últimos anos, seja por velocidade de transmissão na internet, capacidade de processamento, ainda não é viável a manipulação e armazenamento de vídeos sem compressão, devido ao elevado volume de informações. Deste modo, codificadores de vídeos são essenciais para o sucesso de aplicações e dispositivos capazes de manipular e capturar vídeos digitais.

Os codificadores de vídeos utilizam diversas ferramentas com o intuito de identificar e remover redundâncias presentes em vídeos digitais de modo a elevar sua taxa de compressão. Dentre as ferramentas utilizadas estão: 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 predição do domínio espacial 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 por gerar o *bitstream* do vídeo, por meio do processamento os coeficientes quantizados com base em sua probabilidade em utilização.

Os codificadores de vídeo atuais (WIEGAND et al, 2013) dividem o vídeo (sinal tridimensional: altura  $\times$  largura  $\times$  tempo) em quadros (sinal bidimensional: altura  $\times$  largura), e cada quadro em blocos de tamanhos menores, para aplicarem as etapas acima descritas. Entretanto, a utilização de transformadas tridimensionais, como a DCT-3D, surgem como alternativa à Estimação de Movimento, etapa de maior esforço computacional presente nos codificadores atuais (CORREA et al., 2012), presente na predição Interquadros, em caso de movimentos não translacionais ou regiões de baixa movimentação (LEE; CHAN; ADJEROH, 1997). Assim, visando avaliar o uso de transformada tridimensional, no proposto codificador, o vídeo é dividido sequências de oito quadros (sinal tridimensional: altura  $\times$  largura  $\times$  tempo) e posteriormente em cubos de  $8 \times 8 \times 8$  amostras, então processados pelas diferentes configurações suportadas.

### 2. METODOLOGIA

O codificador baseado em transformada tridimensional foi descrito utilizando a Linguagem C/C++ e possui suporte às etapas acima mencionadas. As etapas suportadas neste codificador são descritas a seguir. A Predição Intraquadro é aplicada somente ao primeiro quadro do vídeo, utilizando o software de referência do padrão H.264 em sua versão 19.0 (JM19). Este quadro então é utilizado como referência para os quadros seguintes, que serão processados pela predição

Interquadros. Para a predição Interquadros, três opções de estimação de movimento estão disponíveis: i) predição diferencial, onde seus resíduos são gerados a partir da diferença entre o quadro a ser codificado e o de referência; ii) *Full-Search* (FS): é realizada a busca pelo bloco de maior similaridade presente no quadro de referência, de modo cada profundidade do cubo é comparada com um mesmo bloco no quadro de referência, e então é utilizado um vetor de movimento por cubo; iii) *Full-Search+* (FS+) para cada profundidade do cubo é feita a busca pelo bloco de maior similaridade presente no quadro de referência, gerando assim oito vetores de movimento por cubo. Para as etapas de Transformada e Quantização foram utilizadas a DCT-3D e o volume de quantização proveniente do estudo de LEE (1997). Para os diferentes pontos de operação dos codificadores, existem parâmetros de quantização (QP – do inglês *Quantization Parameters*) que parametrizam os codificadores para diferentes qualidades e taxa de *bits*. Para os codificadores baseados em transformada tridimensional o volume de quantização foi multiplicado pelos fatores {0,1; 0,3; 0,5; 1,5}. Como codificador de Entropia foi utilizado o *Context-Adaptive Variable-Length Coding* (CAVLC) e suas tabelas de contexto presente do JM19.

Com as ferramentas acima descritas, o codificador apresenta seis configurações de codificação, entretanto para facilitar a nomenclatura, cada configuração será chamada de codificador: i) Espacial (Esp): compota pelas etapas de Transformada, Quantização e Entropia; ii) Residual (Res): composta pelas etapas de Predição Intraquadro, Predição Interquadros utilizando predição diferencial, Transformada, Quantização e Codificação de Entropia; iii) FS: mesmas etapas de Res, entretanto utilizando algoritmo FS na Predição Interquadros; iv) FS+: mesmas etapas de FS e Res, entretanto com algoritmo FS+ na Predição Interquadros. v) RD-All: seleciona a configuração menor custo na relação taxa-distorção dentre as acima mencionadas; vi) RD-Fast: mesma ideia do RD-All, entretanto podendo selecionar apenas entre Esp e Res.

Como referencial de comparação de resultados, o JM19 foi configurado para ter suporte apenas as ferramentas semelhantes as suportadas pelo codificador desenvolvido, ou seja: utilizando o codificador de entropia CAVLC, algoritmo de busca *Full-Search*, ordem de codificação de baixa latência (do inglês, *Low Delay*), apenas blocos de tamanho 8x8 para as etapas de Predições e Transformada, desativadas as ferramentas de: bi-predição, estimação de movimento fracionária, filtros de suavização, entre outros. Assim o codificador JM, com estas configurações, será chamado neste trabalho de JM LD8.

Como resultados, foram coletadas informações sobre: qualidade visual utilizando a métrica de qualidade subjetiva PSNR (*Peak Signal-to-Noise Ratio*), o número de bits necessários para representar o vídeo codificado, e o tempo total de codificação. 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/acrescido em termos de taxa entre dois codificadores, na qual valores negativos são desejáveis, representando a redução média do número de *bits* do vídeo codificado para mesma qualidade de vídeo.

Para os resultados foram utilizados cinco vídeo de resolução 1920x1080p presentes Condições Comuns de Teste do padrão HEVC (BOSSÉN, 2013). Foram codificados: 40 quadros com o codificador Esp, e 41 quadros com os demais. Para fins de comparação, no codificador Esp o primeiro quadro do vídeo é ignorado, e nos demais, são descartadas as informações referentes ao primeiro quadro.

### 3. RESULTADOS E DISCUSSÃO

Os resultados obtidos em termos de eficiência e tempo de codificação pelo codificador desenvolvido e pelo JM estão apresentados na Figura 1. A Figura 1 (a) apresenta as curvas de taxa-distorção de cada codificador para quatro vídeos, as curvas referentes ao vídeo ParkScene foram suprimidas visto que apresentam comportamento similar ao vídeo Kimono1. Os pontos presentes nessas curvas referem-se ao QP utilizado pelo codificador, na qual os valores de QPs são inversamente proporcionais à qualidade e taxa de bits, ou seja, menores QPs estão associados a melhor qualidade visual e maior taxa de bits.

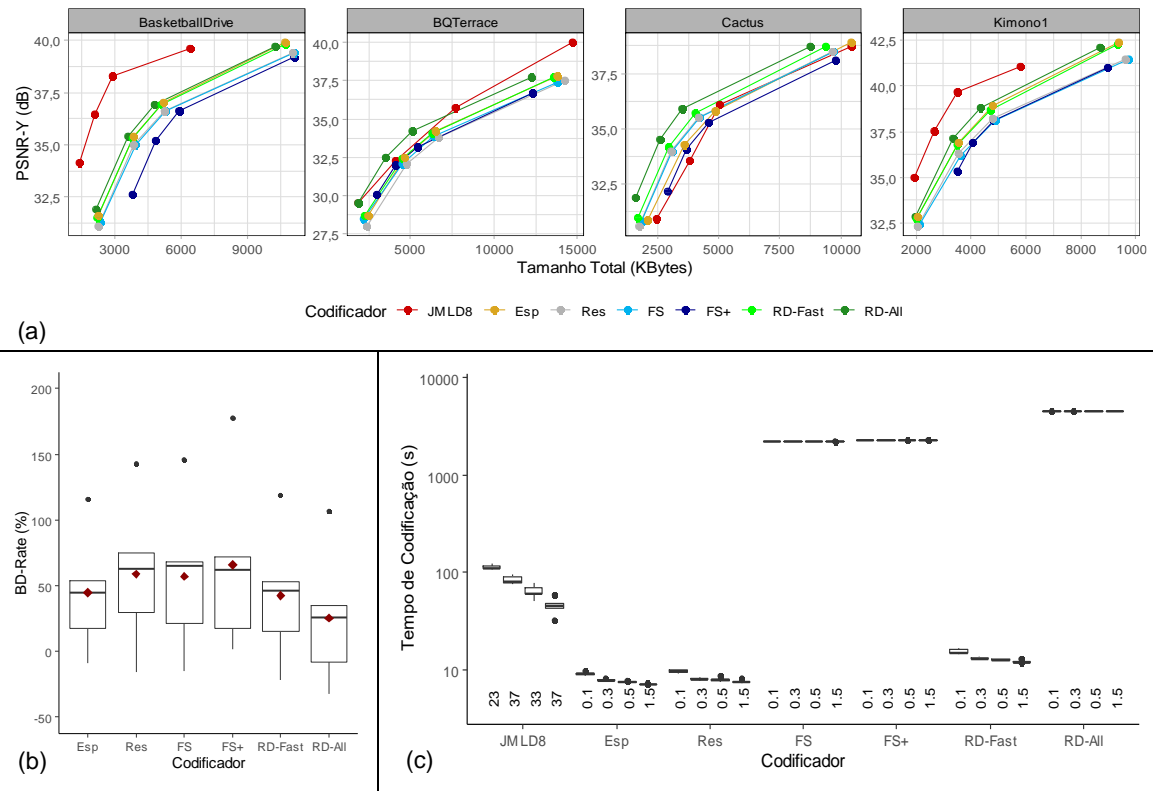


Figura 1: Resultados obtidos pelos codificadores: (a) Curvas de Taxa-Distorção (b) BD-Rate com o JM LD8 como referência e (c) Tempo de Codificação

Com a Figura 1 (a) é possível observar que, quando comparados ao JM LD8, os codificadores baseados em transformada tridimensional apresentaram os melhores resultados para o vídeo Cactus, resultados competitivos para o vídeo BQTerrace, e piores resultados para os vídeos BasketballDrive, Kimono1 e Parkscene. Conforme esperado, dentre as curvas dos codificadores baseados em transformada tridimensional, os melhores resultados foram obtidos pelos codificadores RD (*Fast* e *All*), visto que podem selecionar a melhor dentre as configurações de codificação por eles suportadas. Os codificadores com o algoritmo *Full-Search* apresentaram os piores resultados, em especial a configuração FS+, visto que estes contêm mais informações laterais – vetor(es) de movimento – associados a cada cubo. Com isso, analisando os codificadores Esp e RD-*Fast*, é possível constatar que estes, surgem como boas alternativas à Estimação de Movimento combinadas com transformada tridimensional em termos de eficiência de codificação.

A Figura (b) apresenta resultados de BD-Rate para cada codificador baseado em transformada tridimensional tomado como referência o codificador JM LD8. Cada *boxplot* contém o valor de BD-Rate dos cinco vídeos utilizados em cada um

dos codificadores, além disso, os losangos indicam o valor médio dos vídeos para cada configuração. Estes resultados facilitam visualizar a diferença média entre as curvas de taxa-distorção de cada vídeo apresentadas na Figura 1 (a). Estes resultados são diretamente relacionados à estas curvas. Vale destacar que os *outlier* são referentes ao vídeo BasketballDrive e os valores negativos ao vídeo Cactus, conforme é possível observar nas curvas de taxa-distorção.

A Figura 1 (c) apresenta o tempo de codificação dos distintos codificadores e seus QPs. Cada *boxplot* contém aos valores de tempo de codificação dos cinco vídeos utilizados. Os tempos de codificação são apresentados em escala logarítmica para facilitar a visualização, pois a diferença do tempo de codificação por cada codificador (FS, FS+ e RD-All comparado ao JM LD8, e JM LD8 comparado a Esp, Res e RD-Fast) é em torno de uma ordem de grandeza. Esta diferença de tempo entre os codificadores está relacionada ao algoritmo *Full-Search* utilizado nestes codificadores, visto esta é uma etapa de maior representatividade no tempo de processamento dos codificadores de vídeo (CORREA et al., 2012). Entretanto, serão realizadas análises mais sofisticadas em torno dos codificadores FS e FS+, de modo a investigar o motivo do acréscimo significativo no tempo de codificação mesmo quando comparados ao JM LD8, o qual também adota o algoritmo *Full-Search* de forma similar. Outro ponto a ser observado é a diferença de tempo de codificação em função do QP utilizado. A diferença de tempo é facilmente perceptível no JM LD8, entretanto isto não ocorre nos codificadores baseados em transformada tridimensional, ou seja, nestes codificadores mesmo que para configurações que requerem uma melhor qualidade visual não requerem um acréscimo significativo no tempo de codificação.

#### 4. CONCLUSÕES

Este trabalho apresentou uma avaliação do uso de transformada tridimensional no processo de codificação de vídeo. Com os resultados apresentados, em termos de eficiência e tempo de codificação juntos, o uso de transformada tridimensional aplicada sobre as amostras originais do vídeo (Esp), ou sobre os resíduos obtidos de predição simples (Res), ao invés de predições com maior esforço computacional adotadas na estimação de movimento (FS, FS+), demonstram-se competitivas. Em especial, competitivas para cenários com restrição energética, como dispositivos móveis alimentados por bateria, devido seu baixo tempo de codificação, mesmo com perdas em eficiências de codificação.

#### 5. REFERÊNCIAS BIBLIOGRÁFICAS

BJONTEGAARD, G. Calculation of average PSNR differences between RD-curves. **ITU-T Q. 6/SG16 VCEG**, 15th Meeting, Austin, Texas, USA, April, 2001.

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.

CORREA, Guilherme et al. Performance and computational complexity assessment of high-efficiency video encoders. **IEEE Trans. on Circuits and Sys. for Video Technology**, v. 22, n. 12, p. 1899-1909, 2012.

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.

WIEGAND, T. et al. Overview of the H.264/AVC video coding standard. **IEEE Trans. on Circuits and Syst. for Video Technology**, v13, n7, p 560–576, jul. 2003.