

AVALIAÇÃO DO USO DE DIFERENTES UNIDADES DE CODIFICAÇÃO EM PREDIÇÃO INTER QUADROS NO CODIFICADOR VVC

RAMIRO VIANA¹; FERNANDO SANGRILLO²; MARTA LOOSE³;
GUSTAVO SANCHEZ⁴; GUILHERME CORRÊA⁵; LUCIANO AGOSTINI⁶

¹Universidade Federal de Pelotas (UFPEL) – rgsviana@inf.ufpel.edu.br

²Universidade Federal do Pampa (UNIPAMPA) – fernandosagrilo.aluno@unipampa.edu.br

³Universidade Federal de Pelotas (UFPEL) – marta.breunig@inf.ufpel.edu.br

⁴Instituto Federal do Sertão Pernambucano (IFSertãoPE) – gustavo.sanchez@ifsertao-pe.edu.br

⁵Universidade Federal de Pelotas (UFPEL) – gcorrea@inf.ufpel.edu.br

⁶Universidade Federal de Pelotas (UFPEL) – agostini@inf.ufpel.edu.br

1. INTRODUÇÃO

Nos últimos anos, a demanda por acesso a vídeos digitais tem se intensificado muito, principalmente devido ao constante aumento do uso da Internet. Outro fator para esta demanda crescente foi a pandemia de Covid-19 que impulsionou o trabalho e estudo remoto, além das diversas plataformas de streaming disponíveis para entretenimento. Por essas razões, maneiras eficientes de codificar vídeos são desenvolvidas para apoiar este aumento na demanda de transmissão de vídeos (TAKAMURA, 2019).

O padrão de última geração para codificação de vídeo é o *Versatile Video Coding* (VVC) (ZHANG, 2018). Apesar de apresentar muito esforço computacional necessário, o VVC fornece uma alta proporção de compressão (SALDANHA, 2020), sendo superior a outros codificadores disponíveis no mercado em termos de eficiência de compressão vídeo (TAKAMURA, 2019). No VVC, o vídeo a ser codificado é primeiramente particionado em blocos chamados Unidades de Árvore de Codificação, do inglês *Coding Tree Units* (CTU), e cada CTU pode ter de 32x32 a 128x128 amostras de luminância, seguidas das respectivas amostras de crominância (Cb e Cr). Cada CTU ainda pode ser dividida em blocos chamados Unidades de Codificação, do inglês *Coding Units*. As Unidades de Codificação são utilizadas para melhor adaptação o processo de codificação às características da imagem, além de serem os blocos de quadros que são, de fato, codificados (JUNG, 2021).

O VVC é um codificador híbrido, o que significa que cada Unidade de Codificação é prevista por meio de uma predição Inter Quadros ou Intra Quadros e os resíduos das predições são então transformados e quantizados e, por fim, a informação quantizada é processada pelo codificador de entropia para finalizar o processo de codificação. A predição inter quadros explora a redundância temporal em uma sequência de vídeo prevendo as Unidades de Codificação de quadro atuais usando outros quadros previamente codificados do vídeo (SALDANHA, 2020). O passo mais importante dentro da predição inter quadros é a Estimção de Movimento, do inglês *Motion Estimation* (ME), foco deste trabalho. A ME tenta encontrar, dentro dos referenciais, qual bloco é o melhor para ser usado para prever a Unidade de Codificação atual, chamado de "melhor casamento". Naturalmente, este processo requer um alto nível de esforço computacional. A estimção de movimento do VVC é composta por três ferramentas principais: Unidirecional, Bidirecional e Affine (GONÇALVES, 2021).

Este trabalho apresenta uma investigação sobre os diferentes tamanhos de Unidade de Codificação que são usados durante essas ferramentas de predição Inter Quadros, incluindo o número de amostras codificadas com cada tamanho de

Unidade de Codificação para cada uma dessas três ferramentas de codificação e também o tempo gasto para codificar cada tamanho de Unidade de Codificação.

2. METODOLOGIA

Os experimentos deste trabalho foram feitos usando o software de referência VVC, chamado VVC Test Model (VTM), versão 14.0, utilizando a configuração Random-Access com Parâmetros de Quantização, do inglês *Quantization Parameters* (QPs), sendo 22, 27, 32 e 37, conforme definido por as Condições Comuns de Teste (CTCs) (BOSSSEN, 2020). Quinze vídeos com três resoluções diferentes foram usados e foram considerados os primeiros 32 frames desses vídeos. Esses vídeos foram divididos em três resoluções: HD (1280x720), Full HD (1920x1080) e 4K Ultra HD (3840x2160). As sequências em HD foram *Dark*, *Netflix DrivingPOV*, *Vidyo4*, *Netflix DinnerScene* e *KristenAndSara*. As sequências em Full HD foram *Netflix TunnelFlags*, *Jockey*, *Beauty*, *Touchdown Pass* e *Rush Field Cuts*. Por fim, as sequências Ultra HD 4K foram *ToddlerFountain*, *SunBath*, *Lips*, *BuildingHall2* e *Netflix Dancers*. Todos os vídeos foram codificados quatro vezes, uma vez para cada valor de QP.

O VTM foi modificado para permitir a extração das informações necessárias para a avaliação apresentada neste trabalho. A informação foi extraída para os quatro estágios da ME: Unidirecional, Bidirecional, Affine de quatro parâmetros e de seis parâmetros. Os resultados para Unidirecional e Bidirecional foram agrupados e analisados em conjunto, uma vez que essas ferramentas têm comportamentos semelhantes e estão presentes em outros codificadores. Os dois modos de Affine, com quatro e seis parâmetros, também foram agrupados e analisados em conjunto, uma vez que seus comportamentos são semelhantes.

Para tornar a avaliação mais compreensível, os tamanhos das Unidades de Codificação foram agrupados de acordo com o número de amostras que possuem, como mostrado na Tabela 1, onde a coluna "Peso" foi definida como o número de amostras dentro de cada Unidade de Codificação, dividido por 16, apenas para simplificar a análise apresentada na próxima seção deste trabalho.

Tabela 1: Agrupamento dos Tamanhos das Unidades de Codificação

Grupo	Tamanhos das Unidades de Codificação	Peso
1	4x8 8x4	2
2	4x16 8x8 16x4	4
3	4x32 8x16 16x8 32x4	8
4	4x64 8x32 16x16 32x8 64x4	16
5	8x64 16x32 32x16 64x8	32
6	16x64 32x32 64x16	64
7	32x64 64x32	128
8	64x64	256
9	64x128 128x64	512
10	128x128	1024

Após a codificação de todas as sequências com os quatro QPs, as características foram extraídas e os resultados finais foram alcançados seguindo os passos abaixo:

- As médias de uso como a melhor correspondência e do tempo gasto para codificar todos os tamanhos de Unidades de Codificação para cada vídeo e

para os quatro QPs foram calculados, considerando os dois grupos de ferramentas: (I) Unidirecional e Bidirecional e (II) Affine;

- Em seguida, essas médias de uso e tempo foram agrupadas de acordo com as resoluções dos vídeos e novas médias foram calculadas, novamente considerando os dois grupos de ferramentas ME;
- Depois, foram somadas as médias de cada tamanho de Unidade de Codificação de acordo com ao seu peso, considerando os 10 grupos definidos na Tabela 1, para cada resolução e para cada grupo de ferramentas ME;
- Por fim, o número de amostras dentro de cada Unidade de Codificação foi multiplicado pelo uso médio de cada grupo de Unidades de Codificação, para gerar o número médio de amostras codificadas com esse grupo de Unidades de Codificação.

3. RESULTADOS E DISCUSSÃO

Os resultados alcançados de amostras codificadas usando cada grupo de tamanhos de Unidade de Codificação, considerando o método apresentado na seção anterior, estão mostradas na Figura 1, com um gráfico para cada grupo de sequências em cada uma das três resoluções focadas nesta investigação: HD, Full HD e Ultra HD. O eixo X desses gráficos apresenta o "Peso" definido na Tabela 1, indicando os grupos de tamanhos de Unidade de Codificação. O eixo Y apresenta o número médio de amostras codificadas em cada grupo de Unidades de Codificação para cada execução (considerando a média dos quatro QPs para cada sequência, e a média dos resultados para todas as sequências com cada resolução, conforme apresentado na seção anterior). Esses gráficos apresentam, em azul, os resultados para ME Unidirecional e Bidirecional e, em vermelho, os resultados para os dois modos ME Affine.



Figura 1: Número médio de amostras codificadas com cada grupo de tamanhos de Unidades de Codificação em vídeos HD, Full HD e Ultra HD

Outra análise importante está relacionada ao tempo gasto para codificar as Unidades de Codificação para os dois grupos de ferramentas para as três resoluções. A Figura 2 apresenta gráficos com os resultados médios de tempo de codificação gasto para o peso de cada grupo de Unidade de Codificação, de forma semelhante aos gráficos apresentados na Figura 1.



Figura 2: Tempo gasto para codificar os grupos de tamanhos de Unidades de Codificação em vídeos HD, Full HD e Ultra HD

4. CONCLUSÕES

Este trabalho apresentou uma investigação sobre as MEs Unidirecional, Bidirecional e Affine do padrão VVC, ao processar três resoluções de vídeo diferentes.

Através dos resultados e discussões apresentadas neste trabalho, é possível concluir que as MEs Unidirecional e Bidirecional são usadas para codificar muito mais amostras do que a ME Affine, mas requerem um esforço computacional menor, independentemente da resolução do vídeo. Os resultados apresentados também permitiram concluir que quanto maior a resolução do vídeo, menor é o uso relativo da ME Affine em comparação com a ME Unidirecional e Bidirecional. Por outro lado, o uso relativo de diferentes tamanhos de Unidade de Codificação não apresenta alterações significativas para diferentes resoluções e ferramentas de codificação.

Os resultados apresentados neste trabalho apontam claramente para novas oportunidades de pesquisa relacionadas com a alta complexidade da ME Affine. Então, novas soluções focadas em reduzir o esforço computacional para esta ferramenta de codificação tende a gerar impactos positivos no esforço computacional de todo o codificador.

5. REFERÊNCIAS BIBLIOGRÁFICAS

TAKAMURA, S. Versatile Video Coding: a Next-generation Video Coding Standard. **NTT Technical Review**, Japão, v.17, n.6, p.49-52, 2019.

ZHANG, T.; MAO, S. An Overview of Emerging Video Coding Standards. **GetMobile: Mobile Computing and Communications**, EUA, v.22, n.4, p.13-22, 2018.

SALDANHA, M.; CORRÊA, M.; PALOMINO, D.; PORTO, M.; ZATT, B.; AGOSTINI, L. An Overview of Dedicated Hardware Designs for State-of-the-Art AV1 and H.266/VVC Video Codecs. In: **2020 27th IEEE International Conference on Electronics, Circuits and Systems (ICECS)**, p.1-4.

JUNG, S.; JUN, D. Context-Based Inter Mode Decision Method for Fast Affine Prediction in Versatile Video Coding. **Electronics**, Suíça, v.10, n.11, p.1-12, 2021.

GONÇALVES, P.H.R. **Um Esquema Rápido Baseado em Aprendizado de Máquina para a Predição Interquadros do Codificador de Vídeo VVC**. 2021. 90f. Dissertação (Mestrado em Ciência da Computação) – Curso de Pós-graduação em Ciência da Computação, Universidade Federal de Pelotas.

BOSSSEN, F.; BOYCE, J.; SUEHRING, K.; LI, X.; VADIM, S. **VTM common test conditions and software reference configurations for SDR vídeo**. Document WG 05 MPEG Joint Video Coding Team(s) with ITU-T SG 16 JVET-T2010.ed. [S.l.: s.n.], 2020.