

AVALIAÇÃO EXPERIMENTAL DA COMPENSAÇÃO DE MOVIMENTO FRACIONÁRIA EM CODECS DE VÍDEO SEGUNDO O PADRÃO *VERSATILE VIDEO CODING*

LAIANE SOUZA¹; YASMIN CAMARGO¹; BRUNO ZATT¹;
SÉRGIO BAMPI²; FELIPE SAMPAIO³

¹ Universidade Federal de Pelotas (UFPEL) - {lssouza, yasmin.sc, zatt}@inf.ufpel.edu.br

² Universidade Federal do Rio Grande do Sul (UFRGS) - bampi@inf.ufrgs.br

³ Instituto Federal do Rio Grande do Sul (IFRS) - felipe.sampaio@farroupilha.firs.edu.br

1. INTRODUÇÃO

Etapas de compressão vêm sendo indispensáveis nas aplicações que manipulam vídeos digitais, visto que os custos de um vídeo não comprimido são impraticáveis quanto ao armazenamento e transmissão (AGOSTINI, 2007). São utilizados codecs, que consistem em aplicações de software e hardware para codificação e decodificação de vídeos e áudios (ALENCAR, 2023). O *Versatile Video Coding* (VVC) é o padrão estado-da-arte em termos de codificação de vídeo e alcançou um aumento de 50% nas taxas de compressão em comparação com o seu antecessor, o *High Efficiency Video Coding* (HEVC). Para tal, o VVC adota ferramentas de codificação inovadoras que aumentaram a complexidade, levando a restrições severas em termos de desempenho e consumo de energia, especialmente para dispositivos alimentados por bateria (PAKDAMAN, 2020). Embora a maioria desses problemas de complexidade estejam presentes no lado do codificador, os decodificadores de vídeo VVC ainda exigem um esforço computacional elevado e estão em todos os dispositivos finais capazes de reproduzir vídeos digitais. Estudos recentes mostram que os decodificadores VVC exigem até 3 vezes mais comunicação de memória em comparação com o HEVC (PAKDAMAN, 2020).

Uma ferramenta importante presente nos codecs de vídeo é a Compensação de Movimento (MC - *Motion Compensation*), responsável pela decodificação de blocos codificados pela predição inter-quadros. A MC destaca-se como uma das etapas de decodificação mais complexas nos decodificadores de vídeo, sendo responsável por 37-38% do esforço computacional total necessário. Assim, estratégias de otimização e aceleração para aplicações de codecs de vídeo têm um potencial global para aliviar os problemas de desempenho e energia das infraestruturas de processamento multimídia.

Assim, o principal objetivo deste artigo é analisar a correlação entre os parâmetros de decodificação de MC vizinhos para viabilizar estratégias de otimização que proporcionem um desempenho aprimorado para implementações especulativas de unidades de decodificação de MC executadas em plataformas baseadas em NDP.

2. COMPENSAÇÃO DE MOVIMENTO FRACIONÁRIA

A etapa de MC tem o objetivo de remontar os dados da predição intra-quadros. Para cada bloco inter-predito, os vetores de movimento (MV's) e os índices de quadros de referência, que são decodificados do fluxo de bits de entrada, são as

principais informações necessárias para que a MC reconstrua as amostras preditas. Assim, a MC acessa o buffer de imagens decodificadas para buscar os blocos de quadros de referência das listas de referências 0 (L0) ou 1 (L1) e, alternativamente, de ambas as listas quando a bi-predição é adotada.

Em geral, os MV's podem ser de dois tipos: inteiros e fracionários. No caso de um MV inteiro ser recebido para decodificar um bloco, o MC acessa diretamente os dados do quadro de referência para remontar o bloco predito. Por outro lado, MV's fracionários requerem o cálculo de filtros de interpolação para recriar as amostras fracionárias processadas pelo codificador.

3. METODOLOGIA

Para gerar os dados necessários nesse experimento, adotamos o VTM 23.0 como modelo de teste para avaliações VVC, considerando as configurações de *Random Access* (RA) e *Low Delay* (LD). Como conjunto de dados de entrada, selecionamos quinze sequências de vídeo definidas nas condições comuns de teste para VVC (JVET, 2019): *Campfire*, *FoodMarket4*, e *Tango2* (Classe A1 - 3840x2160); *CatRobot*, *DaylightRoad2* e *ParkRunning* (Classe A2 - 3840x2160); *Cactus*, *BasketballDrive*, *BQTerrace*, *MarketPlace* e *RitualDance* (Classe B - 1920x1080); *BasketballDrill*, *BQMall*, *PartyScene* e *RaceHorsesC* (Classe C - 832x480). Todos os experimentos foram executados para valores do QP (*Quantization Parameter*) de 22, 27, 32 e 37.

Em todas as etapas para a execução dos experimentos, os scripts foram gerados em Python. Como primeira etapa, foram extraídas as ocorrências dos MV's fracionários de arquivos de traço, extraídos a partir da execução do decodificador VTM, considerando os parâmetros de entrada definidos para a avaliação (mencionados acima), totalizando 120 execuções. A segunda etapa consistiu na construção de planilhas contendo a porcentagem de regiões fracionárias codificadas pela MC, com o objetivo de calcular os valores médios de representatividade dos MV's fracionários com relação à totalidade de regiões do vídeo decodificadas pela MC. Por fim, de forma a avaliar a distribuição das posições fracionárias dos MV's dentro regiões delimitadas do vídeo, a última etapa deste trabalho consistiu em gerar histogramas para demonstrar diferentes cenários. Para se adequar a janelas de interpolação definidas em trabalhos relacionados do grupo de pesquisa, foi definida esta região do tamanho de 2048x128 amostras, abrangendo 16 blocos do tamanho 128x128, sendo o maior tamanho de bloco suportado pelo padrão VVC.

4. RESULTADOS DA AVALIAÇÃO

A Figura 1, apresenta a porcentagem de regiões do vídeo codificadas com MVs fracionários, que deverão ser processadas pela MC. Durante a etapa de codificação, a predição inter-quadros realizou o processo de Estimação de Movimento e, para alguns blocos do vídeo, foi decidido que o uso de MVs fracionários traria melhores resultados em termos de eficiência de codificação do que o uso de MVs inteiros. Assim, esta primeira análise avalia, para cada vídeo, quais parâmetros e cenários de codificação resultaram em maior ou menor ocorrência de MVs fracionários.

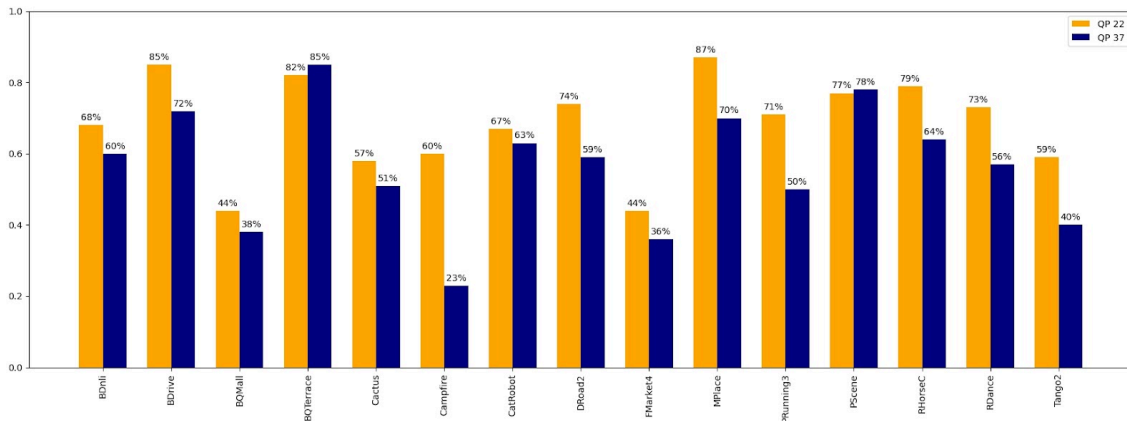


Figura 1. Regiões fracionárias decodificadas pela MC.

É possível notar que os resultados da Figura 1 variam consideravelmente de vídeo para vídeo, e a análise do gráfico reflete essa diferença: em alguns vídeos há mais ocorrências de vetores fracionários enquanto que, em outros, menos. A ocorrência de regiões codificadas com MVs fracionários varia de 23% (Campfire-QP22) e 87% (MarketPlace-QP22), demonstrando a grande variação que foi verificada nos experimentos realizados. Além disso, é possível notar também a influência do QP: menores valores de QP ocasionam um aumento na ocorrência de MVs fracionários durante a etapa de predição inter-quadros. Considerando os valores extremos de QP analisados, a ocorrência de MVs fracionários é de 68,4% na média para os experimentos com QP 37, tendo um aumento para 56,3% quando o valor de QP 22 é utilizado.

A posição do vetor de movimento fracionário é importante porque determina qual filtro será aplicado em seguida e se existe correlação entre os vetores fracionários. Nos experimentos realizados neste trabalho, adotamos a interpolação fracionária com precisão de até $\frac{1}{4}$ de pixel, resultando em 15 diferentes tipos de interpolação para cada posição inteira do quadro: três posições de $\frac{1}{2}$ pixel (H0, H1 e H2), e doze posições de $\frac{1}{4}$ de pixel (Q0, Q1, ..., Q11).

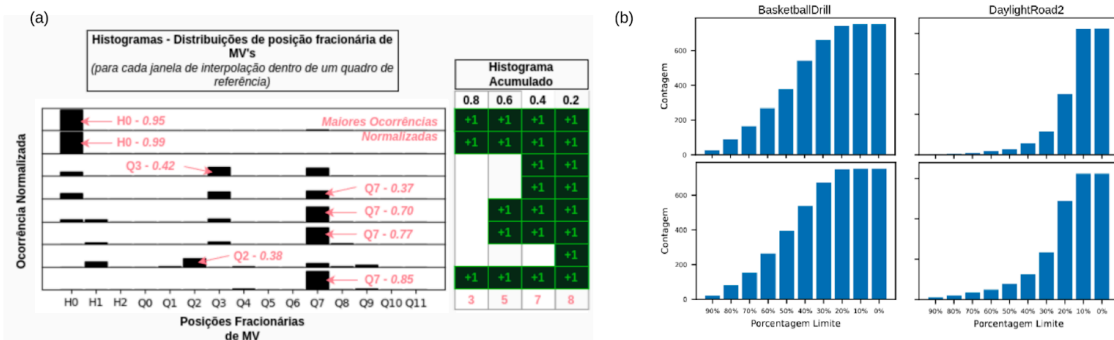


Figura 2. (a) Dinâmica dos histogramas acumulados para representar as distribuições das posições dos MVs fracionários para cada experimento; **(b)** histogramas para os vídeos BasketballDrill LD 22 e LD 32 e DaylightRoad2 RA 22 e RA 32.

Para capturar essa relação, utilizamos a metodologia apresentada na Figura 2a. Inicialmente, dividimos os quadros do vídeo em janelas de 2048x128 e, para cada uma delas, foram realizadas as contagens das regiões que foram

codificadas com cada uma das quinze posições fracionárias. Então, a posição com maior ocorrência normalizada é calculada. Na primeira janela avaliada na Figura 2a, por exemplo, o histograma aponta que a posição mais frequente (H0) aparece com um fator de 0,95. Este processo é então executado para todas as janelas de todos os quadros avaliados do vídeo. Para a construção dos histogramas acumulados, foram definidos níveis de 0 a 1, com passo de 0,1, seguindo a lógica de construção apresentada na Figura 2a.

Como resultados, foram gerados histogramas acumulados com o objetivo de obter uma caracterização dos perfis de distribuição das posições dos MVs fracionários para o conjunto de casos de teste analisados neste trabalho. Como demonstração destes resultados, foram escolhidos dois casos extremos que podem ser observados na Figura 2b: um vídeo com maiores índices de concentração das posições dos MVs fracionários (*BasketballDrill*) e outro com menores índices (*DaylightRoad2*). Nota-se que ao aplicar uma estratégia para explorar a correlação entre os vetores, há uma variação entre os vídeos, sendo mais expressiva em alguns do que em outros.

5. CONCLUSÕES

Na conclusão deste trabalho, observou-se a influência da compensação de movimento fracionária nos codecs de vídeo VVC. A análise mostrou que a ocorrência de vetores fracionários varia conforme os vídeos e configurações de QP, e que a distribuição das posições apresenta correlação em algumas sequências, indicando potencial de otimização. Em vídeos como *BasketballDrill*, há maior concentração de vetores fracionários, sugerindo oportunidades mais diretas de otimização. Já em *DaylightRoad2*, a distribuição é mais dispersa, exigindo estratégias mais complexas. Assim, explorar a correlação entre vetores pode trazer ganhos relevantes em desempenho para decodificadores VVC, abrindo caminho para pesquisas futuras em aceleração e otimização, com impacto na eficiência energética e de processamento.

6. REFERÊNCIAS BIBLIOGRÁFICAS

AGOSTINI, L. V. **Desenvolvimento de Arquiteturas de Alto Desempenho Dedicadas à Compressão de Vídeo Segundo o Padrão H.264/AVC**. 2007. Tese (Doutorado e Mestrado em Ciência da Computação) - Programa, UFPel.

ALENCAR, F. **O que são codecs e como funcionam?** Hardware.com.br, 10 fevereiro 2023. Especiais. Acessado em 11 set. 2024. Online. Disponível em: <https://www.hardware.com.br/artigos/o-que-sao-codecs-como-funcionam/>

PAKDAMAN, F. et al. **Complexity analysis of next-generation VVC encoding and decoding**. In: 2020 IEEE International Conference on Image Processing (ICIP). IEEE, 2020. p. 3134-3138.

JVET. **JVET common test conditions and software reference configurations**. San Diego: Joint Video Experts Team (JVET), 2018. Acessado em 11 set. 2024. Disponível em: <https://www.researchgate.net/publication/326506581>