

IDENTIFICANDO OPORTUNIDADES DE ECONOMIA DE ENERGIA PARA CODIFICAÇÃO DE VÍDEO AV1 EM SERVIÇOS DE STREAMING

CAROLINE SOUZA CAMARGO; ALEX BORGES; GUILHERME CORREA

Universidade Federal de Pelotas (UFPEL) – {caroline.sc, amborges, gcorrea}@inf.ufpel.edu.br

1. INTRODUÇÃO

O consumo de *streaming* de vídeo cresceu 26% no ano de 2020 e já possui o dobro de assinantes da TV a cabo, conforme pode ser observado no relatório publicado pela MOTION PICTURE ASSOCIATION (2020). O funcionamento dessas plataformas só é possível graças à existência de eficientes formatos de compressão de vídeo, que garantem a compatibilidade com diversos dispositivos multimídia e permitem a variabilidade da largura de banda dos usuários (KEMPF et al., 2021). De forma a prover a compressão de vídeos, os serviços de *streaming* recorrem a codificadores de vídeo. Lançado em 2015, o AOMedia Video 1 (HAN et al., 2021) é um formato de codificação de vídeo que oferece um aumento na eficiência de compressão em torno de 30% em relação ao formato VP9 (HAN et al., 2021). Embora a primeira versão do *software* de referência AV1 (libaom) tenha sido lançada apenas em julho de 2018, sua adoção é significativa, estando presente atualmente em 15% do mercado mundial (FRANCIS, 2022).

O AV1 requer um elevado custo computacional para atingir altas taxas de compressão, o que leva a um significativo consumo de energia. Segundo HAN et al. (2021), o aumento da complexidade do AV1 é evidente quando comparado ao seu antecessor, o VP9, principalmente devido ao número de ferramentas implementadas pelo AV1. Sendo assim, esse problema é destacado por HAN et al. (2021), mostrando que o AV1 requer um tempo de codificação 100 vezes maior que o libvpx, o *software* de referência do VP9. Vale ressaltar, que os servidores das empresas de *streaming* armazenam um único vídeo em diversas opções de *bitstreams*, que são codificados sob diferentes condições, como resolução de vídeo e níveis de qualidade. Assim, devido aos diferentes níveis de qualidade que os vídeos precisam ser codificados e pelo aumento no conjunto de ferramentas de testes do AV1, há um consumo de energia muito alto no processo de codificação.

Como o AV1 é um formato desenvolvido a partir do formato VP9, muitas características de codificação são compartilhadas, em especial a estrutura de particionamento de blocos. Ambos os formatos AV1 e VP9 iniciam o processo de particionamento partindo do Superbloco (SB), cujo tamanho inicial é de 128×128 e 64×64, respectivamente, e assumem particionamentos de tamanho mínimo de 4×4 e 8×8, em um processo de particionamento recursivo em forma de árvore (HAN et al., 2021). A Figura 1 ilustra todos tipos de particionamentos que podem ocorrer durante o processo de codificação do AV1, enquanto que no formato VP9, apenas os tipos NONE, HORZ, VERT e SPLIT podem ser utilizados. Vale enfatizar que o tipo de particionamento SPLIT indica ao codificador que ocorreu um avanço no nível de profundidade da árvore de particionamentos, devendo-se criar, então, quatro novas subárvores. Ademais, destaca-se que os vídeos presentes nos servidores de *streaming* precisam ser codificados em diferentes níveis de quantização (CQ), para assim possibilitar a adequação do tamanho do arquivo para a banda disponível do usuário.

O CQ pode assumir um valor inteiro entre zero e 63 (HAN et al., 2021). Quanto maior o valor do CQ, menor será a taxa de *bits* resultante e,

consequentemente, pior será a qualidade da imagem, portanto, menor será o tempo de codificação. Dessa forma, este trabalho tem como objetivo analisar dados estatísticos acerca das ocorrências das profundidades dos blocos ao longo do processo de codificação sob diferentes níveis de quantização entre os formatos AV1 e VP9. A ideia é observar as distribuições dos particionamentos e elaborar regras que possam ser empregadas no desenvolvimento de heurísticas para reduzir o consumo de energia dos vídeos nos servidores de *streaming*.

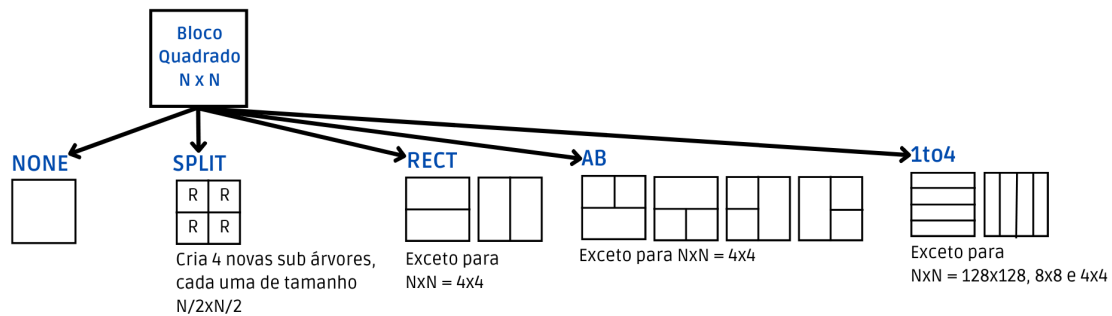


Figura 1. Tipos de particionamento de bloco no formato AV1

2. METODOLOGIA

É com base nos tamanhos de blocos que as demais etapas de codificação de vídeo são executadas e decisões são tomadas. Portanto, neste trabalho, será feito uma análise estatística sobre o comportamento das escolhas das profundidades ao longo do processo de codificação sob diferentes níveis de quantização e, assim, tentar buscar uma redução no custo computacional. Isso será feito comparando as ocorrências de níveis de profundidade da árvore de particionamento gerada por uma codificação sob um CQ mais elevado e comparado a uma codificação sob um CQ menos elevado. Ou seja, ao observar a profundidade da árvore de codificação do CQ maior, será possível entender a probabilidade do CQ menor escolher qualquer outra profundidade. Dessa forma, em futuras implementações será possível fornecer ao libaom essas informações e evitar a codificação de profundidades que são improváveis de serem escolhidas para certas regiões do vídeo. Para execução dos experimentos, foram usados cinco vídeos com resolução 1920x1080 disponíveis no conjunto de testes recomendados para o AV1 (HAN et al, 2021): *RushHour*, *NetflixCrosswalk*, *NetflixTunnelFlag*, *RushFieldCuts* e *TouchdownPass*.

Para realizar as codificações, foi utilizado a versão 3.3 do *software* de referência do AV1, o libaom. Os CQs utilizados durante a codificação seguiram as recomendações estabelecidas por HAN et al. (2021), no qual cada um dos vídeos foi codificado com os seguintes CQs: 20, 32, 34 e 55. A fim de simular um pedaço de vídeo a ser codificado em servidores de *streaming*, apenas os primeiros 60 quadros de cada vídeo foram codificados. Para a obtenção dos dados, o *software* libaom foi modificado para permitir a exportação de informações da árvore de particionamento. Para cada quadro visível que está sendo codificado, informações como posição do bloco (linha e coluna) e tamanho de bloco são exportados. Após a codificação de todos os vídeos, um algoritmo foi implementado de forma a interpretar os dados exportados de cada *frame* em uma matriz própria, onde cada célula dessa matriz representa a profundidade da árvore de particionamento de uma região de tamanho 4x4 do quadro.

3. RESULTADOS E DISCUSSÃO

Os resultados gerados pelos experimentos podem ser observados na Figura 2 deste artigo, que faz referência às relações 32-20 (Figura 2(a)), 43-32 (Figura 2(b)) e 55-43 (Figura 2(c)). As colunas das relações correspondem aos dados provenientes da codificação com CQ menor e as linhas, por sua vez, correspondem ao CQ maior. A relação estabelecida entre esses dois CQs refere-se à probabilidade de qual profundidade o CQ menor irá escolher dado que se observa na profundidade do CQ maior. Por exemplo, na relação 32-20 da Figura 2(a), quando olhamos uma determinada região do quadro que foi codificada com profundidade 2 no CQ 32 (linha 2) também será codificado com profundidade 2 no CQ 20 (coluna 2) em 61.4% das vezes.

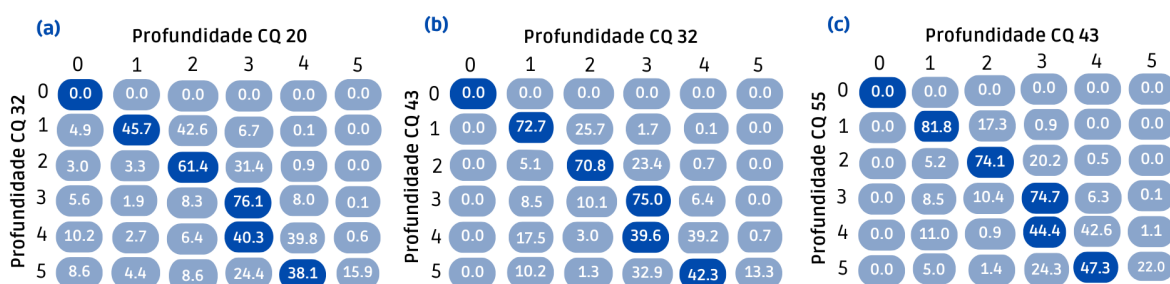


Figura 2. Resultados das análises estatísticas em porcentagem

Na Figura 2, destacou-se a maior probabilidade de ocorrência de cada linha, dessa forma, é possível identificar que a maior probabilidade de algum nível de profundidade ocorrer para qualquer relação tende a ser o mesmo nível de profundidade observado na codificação de CQ inferior. As exceções são para as profundidades 4 e 5, cuja maior probabilidade de ocorrência é o nível de profundidade anterior (ou seja, 3 e 4, respectivamente). A hipótese para essa exceção está na utilização de melhores escolhas de tipos de particionamento e/ou de modos preditivos. Observa-se, também, a ausência de probabilidades na profundidade 0. A hipótese, neste caso, é de que os vídeos escolhidos para o experimento não codificam com a profundidade 0 em nenhum caso.

Dessa forma, analisando os resultados das tabelas sugere-se uma regra para elevar a probabilidade de acerto desse reaproveitamento de informação: caso a profundidade observada for menor ou igual à 3, permite-se que o codificador possa aplicar o mesmo nível de profundidade ou um nível a mais. Caso contrário (profundidades 4 e 5), evita-se que o codificador considere níveis de profundidade menores que 3. A primeira regra foi desenvolvida pela percepção de que há uma tendência de que a codificação com o CQ menor opte por particionar o nível atual, ou seja, ao verificar o nível de particionamento do CQ menor nota-se que existe uma propensão de ser o mesmo nível de profundidade do CQ maior. Além disso, a escolha do próximo nível de profundidade, a ser aplicado na primeira regra, justifica-se pela observação de que, em profundidades até o nível 3, percebe-se que a segunda maior probabilidade de ocorrência é precisamente o próximo nível da profundidade.

Com base nessa proposta de regra exposta, elaborou-se a Tabela 1, que apresenta a probabilidade de assertividade de se realizar a codificação comparando as ocorrências das profundidades de um CQ menor com um CQ maior. Observe que a regra sugerida iria decidir corretamente em 84.73%, 88.26% e 90.97% dos casos, respectivamente para as relações 32-20, 43-32 e 55-43.

Esta regra pode ser utilizada em um trabalho futuro como base para uma heurística de reutilização de estruturas de particionamento entre diferentes representações de CQ, visando a aceleração da codificação. Em outras palavras, evitará que o libaom dedique tempo e energia processando os níveis de profundidade estatisticamente pouco recomendados. Dessa forma, espera-se uma economia considerável no consumo energético, especialmente no contexto de servidores de serviços de *streaming*.

Relação CQs	Profundidade						Média Probabilidade de acerto
	0	1	2	3	4	5	
32-20	--	88.31%	92.85%	84.13%	80.02%	78.35%	84.73%
43-32	--	98.31%	94.20%	81.39%	78.82%	88.56%	88.26%
55-43	--	99.03%	94.25%	81.05%	86.93%	93.58%	90.97%

Tabela 1. Probabilidade de acerto nas escolhas das profundidades ao longo do processo de codificação de acordo com a regra proposta

4. CONCLUSÕES

A codificação dos vídeos nos servidores de *streaming* é realizada várias vezes para o mesmo vídeo, gerando fluxos de *bits* para várias resoluções e qualidades de imagem. Sendo assim, através de um estudo de análise estatística, este trabalho identificou oportunidades de reaproveitamento de decisões de tamanho de bloco ao longo da codificação sob diferentes níveis de quantização. Mais especificamente, a análise permitiu identificar que é possível reduzir a quantidade de teste de possibilidades de particionamento ao codificar vídeos de alta qualidade com base nas características observadas ao codificar vídeos de baixa qualidade. Mesmo evitando certas profundidades da árvore de codificação, ainda existe uma possibilidade superior a 84% de acertar que um bloco seja ou não particionado corretamente. Como trabalhos futuros, pretende-se implementar as análises realizadas nesse artigo, a fim de testar as hipóteses geradas e confirmá-las através do emprego do BD-Rate e de consumo energético estimado.

5. REFERÊNCIAS BIBLIOGRÁFICAS

FRANCIS, A. **Top video technology trends 2022: The future of streaming is about device reach**. 11 jan. 2020. Acessado em 25 jun. 2022. Online. Disponível em: <https://bitmovin.com/top-video-technology-trends/>

HAN, J et al. **A technical overview of av1**. Proceedings of the IEEE, vol. 109, n. 9, p. 1–28, 2021, doi: 10.1109/JPROC.2021.3058584.

KEMPF, A et al. **Consumo das plataformas de streaming antes e durante a pandemia da covid-19**. 2021. Acessado em 23 jun. 2022. Online. Disponível em: https://admpg.com.br/2021/anais/arquivos/09122021_230937_613eb539011b0.pdf

MOTION PICTURE ASSOCIATION. **Theme report: A comprehensive analysis and survey of the theatrical and home/mobile entertainment market environment for 2020**. Mar. 2020. Acessado em 23 jun. 2022. Online. Disponível em: <https://www.motionpictures.org/wp-content/uploads/2021/03/MPA-2020-THEME-Report.pdf>