

TRANSCODIFICAÇÃO DE VÍDEO ACELERADA DE VP9 PARA AV1 USANDO ÁRVORE DE DECISÃO E ESTRUTURAS DE PARTICIONAMENTO

ALEX BORGES¹; BRUNO ZATT²; MARCELO PORTO³; GUILHERME CORREA⁴

¹Universidade Federal de Pelotas – amborges@inf.ufpel.edu.br

²Universidade Federal de Pelotas – zatt@inf.ufpel.edu.br

³Universidade Federal de Pelotas – porto@inf.ufpel.edu.br

⁴Universidade Federal de Pelotas – gcorrea@inf.ufpel.edu.br

1. INTRODUÇÃO

Segundo BITMOVIN (2021), o formato de codificação de vídeo AOMedia Video 1 (AV1) já está presente em 15% do mercado mundial de codificação de vídeo, com uma previsão de crescimento de 7% ainda em 2022. O AV1 é sucessor do VP9, superando-o em 30% em termos de eficiência de codificação (HAN et al., 2021). Dessa forma, existe a necessidade de se realizar uma migração dos vídeos codificados no formato VP9 para o novo formato AV1.

Para possibilitar uma rápida conversão de formatos entre os codificadores, é preciso reaproveitar as decisões que podem ser obtidas do decodificador para serem utilizadas pelo novo codificador, como bem mostrou a revisão bibliográfica ainda não publicada de BORGES (2022). A árvore de particionamento de blocos, elemento de codificação que divide o vídeo em blocos de tamanho $M \times N$ píxeis, e um desses elementos decisórios que pode ser reaproveitado. Uma das decisões existentes nessa árvore de particionamento é denominada SPLIT, que informa ao codificador se a árvore deve permanecer no atual nível de particionamento ou se deve subdividir o bloco atual em quatro novas sub-árvores. Esse tipo de reaproveitamento já pôde ser visto em BORGES (2021) na transcodificação de VP9 para AV1, no entanto, sem fazer qualquer uso de aprendizado de máquina no processo. BORGES (2021) apresenta uma redução média do tempo de transcodificação de 28.16% a um custo de reduzir a eficiência de codificação do AV1 em 4.34%.

Assim sendo, este trabalho visa apresentar uma proposta de transcodificação de vídeo do formato VP9 para o formato AV1, acelerando-se a transcodificação através do reaproveitamento de estruturas de particionamento com auxílio de aprendizado de máquina. A Figura 1 resume a árvore de particionamento do VP9 e do AV1, onde é possível concluir que as combinações de blocos do AV1 cresceram exponencialmente em relação ao VP9. Portanto, o custo computacional para que o AV1 encontre a melhor árvore de particionamento é consideravelmente superior ao observado no VP9, justificando a necessidade de se reduzir esse custo.

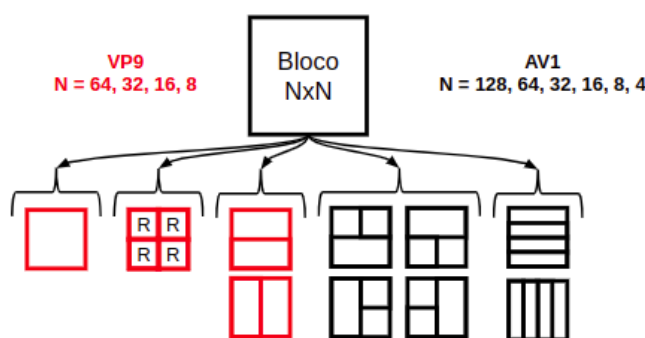


Figura 1. Visão geral da estrutura de particionamento do VP9 (em vermelho) e do AV1. N indica o tamanho do bloco, R indica a subdivisão do bloco.

Em BORGES (2022), o autor conclui que o algoritmo de aprendizado de máquina mais utilizado nos trabalhos de transcodificação de vídeo é a árvore de decisão C4.5/J48, estando presente em 34% dos artigos publicados na literatura científica entre os anos 2011 e 2021 que tratam sobre transcodificação de vídeo com auxílio de aprendizado de máquina. Desta forma, pretende-se fazer uso de um algoritmo de árvore de decisão similar ao C4.5/J48, para auxiliar na decisão do codificador de vídeo baseado no formato AV1 para determinar se a árvore de particionamento, de uma determinada região do vídeo, deverá permanecer no atual nível de particionamento ou se deverá ser subparticionado em quatro novos ramos.

2. METODOLOGIA

Os softwares de referências para os formatos VP9 e AV1 foram utilizados para a execução do experimento, respectivamente os softwares libvpx (versão 1.10, hash 52b3a0) e libaom (versão 3.0, hash 9b1809). Para possibilitar a transcodificação, o vídeo original foi codificado no formato VP9 com o nível de quantização (CQ) igual à 20, depois decodificado para gerar o vídeo intermediário a ser transcodificado. Juntamente com o vídeo decodificado, a distribuição dos tamanhos de blocos de cada quadro foi exportado para um arquivo. No codificador libaom, os dados extraídos do VP9 são reinterpretados como uma matriz, de forma a se obter a profundidade de cada área de 4x4 píxeis. Uma segunda matriz é utilizada para armazenar as profundidades das regiões já processadas pelo libaom. Essas duas matrizes são atualizadas a cada novo superbloco em processamento e são utilizadas em conjunto para gerar os dados a serem enviados para o modelo de aprendizado de máquina, que deverá decidir se o bloco em processamento atual deve ser subparticionado ou não. A Figura 2 mostra as regiões (A, B, C, D e E) onde serão capturados os dados para serem submetidos ao modelo. De cada uma das cinco regiões, calcula-se a média da profundidade observada. Há um modelo treinado para cada combinação de nível de quantização (CQ = 20, 32, 43 e 55) e de níveis de profundidades (N = 128, 64, 32, 16 e 8), totalizando 20 modelos disponíveis para uso. Para reduzir o erro causado pelo modelo, configurou-se ele para ter uma confiabilidade mínima de 90% na predição, senão sua resposta não será utilizada. Importante observar que as regiões B, C, D e E podem não existir nas bordas dos vídeos, portanto, nestes casos o modelo também não será utilizado e, desta forma, o libaom será executado sem interferências.

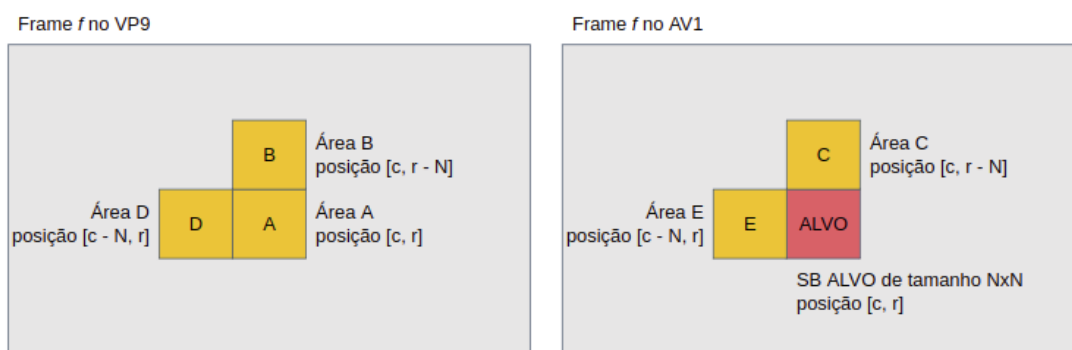


Figura 2. Identificação das regiões a serem obtidos os dados para treinamento e predição do modelo de aprendizado de máquina.

O algoritmo de aprendizado de máquina escolhido para esse trabalho é o Classification and Regression Trees (CART) disponível no pacote SciKit-Learn do

Python. CART é uma árvore binária similar ao C4.5/J48, todavia, com suporte a decisões numéricas ao invés de decisões categóricas. Esse algoritmo permite a manipulação de alguns parâmetros e, neste trabalho, esses parâmetros estão configurados conforme descrito a seguir: `criterion='gini'`, `splitter='best'`, `max_depth=None`, `min_samples_split=6`, `min_samples_leaf=3`, `min_weight_fraction_leaf=0`, `max_features='sqrt'`, `random_state=42`, `max_leaf_nodes=None`, `min_impurity_decrease=0`, `class_weight=None`, `ccp_alpha=0.2`. Importante ressaltar durante a fase de treinamento, os dados foram aleatoriamente selecionados para haver o mesmo número de ocorrências de decisão positiva e negativa e, posteriormente, os dados remanescentes foram reordenados randomicamente.

Para a realização dos experimentos, foram considerados os primeiros 60 quadros de quatro vídeos de treinamento, três vídeos de teste e três vídeos de predição, todos com a resolução 1280x720. Para possibilitar a inferência sobre redução de tempo e o percentual da perda da eficiência de codificação, dado um mesmo nível de qualidade da imagem (medido pela métrica Bjøntegaard Delta-Rate - BD-rate), todos os vídeos foram transcódificados sem a interferências de aprendizado de máquina e com o auxílio de aprendizado de máquina.

4. RESULTADOS E DISCUSSÃO

O modelo de aprendizado de máquina, conforme a configuração escolhida e descrita na seção anterior, obteve índices médios de 86.37% de acurácia, 73.98% de precisão, 74.55% de F1, 90.72% de Recall e 88.22% de área sob a curva ROC, considerando todos os 20 modelos testados (cinco níveis profundidades x quatro CQs). Considerando que este modelo deve retornar uma decisão binária, os resultados médios do modelo são bons, haja visto que o modelo prevê corretamente 90% de todos os particionamentos de árvore que deveriam ser feitos (recall). Uma observação possível ao analisar todos os modelos, é que há uma redução da precisão do modelo conforme o nível de profundidade da árvore aumenta, atingindo uma média de 50.98% de precisão nos blocos 8x8. Essa queda de precisão provavelmente é ocasionado pelo aumento expressivo de exemplos conflituosos para treinamento, já que há exponencialmente mais dados para treinamento em relação aos níveis superiores de profundidade da árvore.

Na Tabela 1 apresentamos os resultados obtidos com a transcodificação, onde podemos observar que os resultados de transcodificação de VP9 para AV1 são relativamente equivalentes ao alcançado pelo trabalho de BORGES (2021), já que enquanto aquele trabalho atingiu 28.16% de redução do tempo a um custo de 4.34% de BD-rate, o nosso transcodificador atinge 14.28% de redução de tempo a um custo de 2.22% de BD-rate. Caso fizermos uma relação entre BD-rate e redução de tempo, BORGES (2021) obteria 0.154 de BD-rate para cada percentual de tempo acelerado, enquanto que o nosso trabalho obtém 0.155. Fica evidente que o trabalho atual, mesmo que ainda em fase inicial, já apreseta resultados interessantes. Portanto, nos trabalhos futuros deverão ser considerados outras configurações deste algoritmo de aprendizado de máquina, de forma a possibilitar resultados mais interessantes, principalmente em termos de redução do tempo de transcodificação. Além disso, o trabalho final irá abordar a aceleração da transcodificação para vídeos 1920x1080, com um total de 11 vídeos para treinamento, 5 vídeos para teste e 31 vídeos para predição.

Tabela 1. Resultados da transcodificação

Sequência	Johnny	Netflix Driving POV	Netflix Roller Coaster	Média
BD-rate (%)	2.5077	1.6659	2.4866	2.2201
Redução de Tempo (%)	17.88	11.19	13.78	14.28

5. CONCLUSÕES

O AV1 vem crescendo como um dos principais formatos de codificação de vídeo. A migração de formatos de codificação antigos para outros mais novos é uma alternativa para redução de custos de armazenamento e de transmissão de dados, ao mesmo tempo que se melhora a qualidade da imagem. No entanto, realizar a transcodificação envolve um elevado custo computacional, por essa razão, é possível encontrar na literatura trabalhos de transcodificação acelerada. A transcodificação de VP9 para AV1 não é novidade na literatura científica, contudo, ainda não foram utilizadas técnicas de aprendizado de máquina nesse processo. Este trabalho visa realizar a transcodificação de vídeo do formato VP9 para AV1 usando árvore de decisão para inferir sobre o subparticionamento ou não da árvore de particionamento do AV1. Neste artigo, foram apresentadas as impressões iniciais desse trabalho, que ainda está em desenvolvimento. É possível concluir que a aplicação de aprendizado de máquina na decisão de particionamento é possível, apesar dos resultados de redução de tempo ainda não serem compatíveis com os existentes na literatura científica, apesar de já apresentarem perdas de BD-rate significativamente menores. Desta forma, trabalhos futuros deverão considerar outras possibilidades de configuração do modelo de aprendizado de máquina utilizado, de forma a encontrar uma combinação de hiperparâmetros que satisfaça as necessidades de uma transcodificação rápida e eficiente.

6. REFERÊNCIAS BIBLIOGRÁFICAS

BITMOVIN. **Bitmovin's 5th Annual Video Developer Report 2021**. Inc. Bitmovin, 14 jan. 2021. Acessado em 8 ago. 2022. Online. Disponível em: <https://go.bitmovin.com/video-developer-report>

BORGES, A.; PALOMINO, D.; B. ZATT, B.; PORTO, M.; G. CORREA, G. Fast VP9-to-AV1 Transcoding based on Block Partitioning Inheritance. **28th European Signal Processing Conference (EUSIPCO)**, p. 555-559, 2021. DOI: 10.23919/Eusipco47968.2020.9287453.

BORGES, A.; B. ZATT, B.; PORTO, M.; G. CORREA, G. A Systematic Literature Review on Video Transcoding Acceleration: Challenges, Solutions, and Trends. **Signal Processing: Image Communications**, 2022. (submetido à publicação)

HAN, J.; et al. A Technical Overview of AV1. **Proceedings of the IEEE**, 2021, p. 1-28. DOI: 10.1109/JPROC.2021.3058584.