

ACELERAÇÃO DO PROCESSO DE SELEÇÃO DE FILTROS DE INTERPOLAÇÃO DO CODIFICADOR AV1 POR MACHINE LEARNING

LEONARDO LUÍS MÜLLER*; DAIANE FREITAS*;
PATRICK ROSA*; GUILHERME CORREA*

*Universidade Federal de Pelotas – {llmuller, dffreitas, psdrosa, gcorrea}@inf.ufpel.edu.br

1. INTRODUÇÃO

O consumo de mídias em vídeo está se tornando cada vez mais comum no cotidiano de todo o mundo, segundo uma pesquisa da Nokia cerca 69% do consumo de mídia em dispositivos móveis serão vídeos até 2026 (NOKIA, 2023). Para conseguir transmitir toda essa informação se faz necessário desenvolvimento de codificadores mais eficientes para tornar possível a transmissão desse volume tão grande de informações.

Um dos codificadores mais modernos atualmente é o AV1, desenvolvido por um grupo de empresas em código livre. Como todo codificador moderno, o AV1 é baseado em blocos, eles possuem tamanhos que variam de 4×4 até 128×128 pixels. Dentro desse processo, a Estimativa de Movimento (ME) e a Compensação de Movimento (MC) são responsáveis por reduzir a redundância entre blocos de quadros distintos. Essas etapas utilizam filtros de interpolação para gerar amostras de pixels fracionários, otimizando a precisão dos vetores de movimento (MV) e melhorando a previsão de movimento.

No AV1, há um esquema de filtragem adaptativa que inclui 75 filtros de interpolação do tipo FIR, divididos em três famílias diferentes: regular, sharp e smooth. Esses filtros são usados para suportar MVs com precisão de até $1/8$ e $1/16$ de um pixel para dados de luminância e crominância, respectivamente.

A escolha dos filtros ocorre em um processo de três estágios, onde as combinações de filtros para as direções horizontal e vertical são testadas, sempre buscando a opção com o menor custo de taxa-distorção (*RD cost*) (AOMEDIACODEC). Esse processo de interpolação adaptativa permite que o AV1 selecione filtros adequados para diferentes características de uma imagem, melhorando tanto a qualidade quanto a eficiência da codificação.

Como motivação, algumas análises iniciais foram realizadas. Inicialmente, analisamos quais filtros são escolhidos, através da codificação de oito vídeos, usando os parâmetros de modo de acesso aleatório, obtidas das condições comuns de teste da AOM (ZHAO, 2021), e com as *Constrained Quality* (CQ) recomendadas para o AV1, de 20, 32, 43 e 55 (Chen, 2020). Como resultado desta análise, entendemos que existe uma larga preferência pelo filtro regular, tanto na horizontal quanto na vertical, sendo na horizontal variando de 56% a 99% de uso e na vertical de 70% a também 99% de uso. Esses resultados levaram a uma nova investigação, para saber qual seria o desempenho da codificação utilizando apenas o filtro regular em ambas as direções. Para isso então foi desenvolvido uma versão modificada do AV1 que utiliza apenas este filtro. Com esta versão modificada, foram codificados seis vídeos, seguindo as mesmas condições comuns de teste e os CQs recomendados. Para medir a qualidade da nova codificação foi utilizada a métrica de BD-Rate, uma métrica baseada na Bjøntegaard Difference (BJONTEGAARD, 2001), além do tempo de codificação. Com isso, percebeu-se um BD-Rate relativamente alto, de 0,79%, e uma redução de tempo de 12,49%.

Com base nas análises realizadas, o **objetivo deste trabalho** foi definido: propor um algoritmo inteligente de decisão para os modos de interpolação, que seja capaz de escolher os filtros corretamente, dependendo do contexto atual, de forma a minimizar a perda em eficiência de compressão, mas ainda assim evitando testar todos os modos, para reduzir o tempo de codificação.

2. METODOLOGIA

A fim de gerar um modelo mais preciso, ao invés de desenvolvermos um único modelo que faz a decisão entre três filtros, foram desenvolvidos quatro modelos, cada um decidindo se a busca de filtros deve parar naquele estágio ou prosseguir.

Para construir uma base de dados com o fim de treinar o modelo, modificando uma versão do codificador AV1, foram retiradas informações importantes do processo de codificação de cada bloco. Com a versão modificada foram codificados 30 quadros de seis vídeos, todos seguindo os parâmetros das condições comuns de teste e os CQs recomendados. Além disso, para cada um dos quatro modelos, em sua base de informações para treinamento, foi criada uma nova coluna, que indica se a busca dos filtros deveria continuar ou não naquele ponto, ela se tornou o alvo a ser predito.

Os quatro modelos foram desenvolvidos em Python, utilizando o modelo de *DecisionTreeClassifier*, cada um deles foi treinado usando as informações que cada etapa tem disponível naquele momento. No treinamento foram utilizados os hiperparâmetros de random_state 42, max_depth 8, min_samples_split 60000 e max_leaf_nodes 60, essas características foram utilizadas após alguns testes visando diminuir o tamanho da árvore e o número de folhas para que assim ao serem implementadas no codificador elas possuísem um bom desempenho sem grande perda da precisão da decisão.

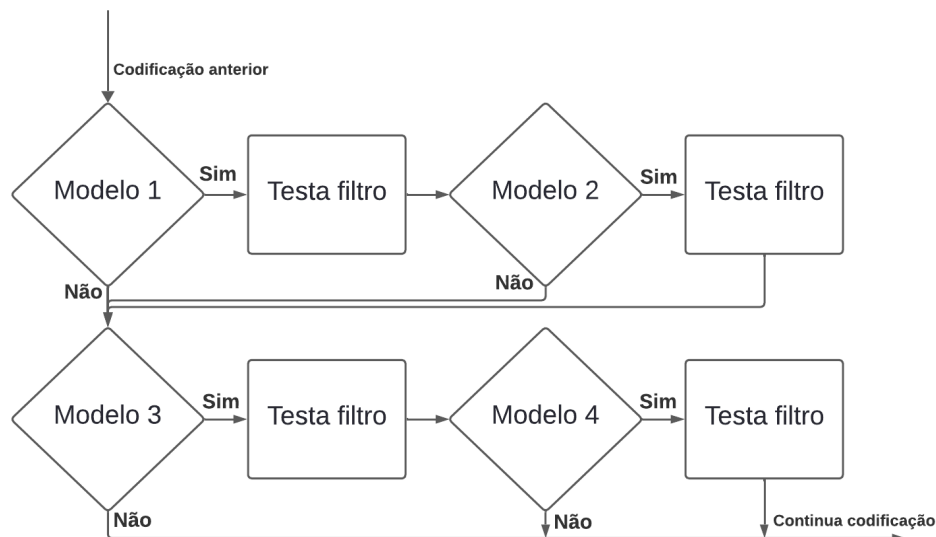


Figura 1: Diagrama da implementação dos modelos.

Com os modelos de predição prontos, eles foram integrados ao codificador AV1 na etapa de busca do filtro de interpolação. Esse processo foi pensado como um sistema de tomada de decisão em múltiplos estágios, então cada modelo determina se deve prosseguir com o teste do filtro atual ou parar e seguir em frente. Essa implementação visa otimizar o processo de busca do filtro, realizando

uma terminação antecipada. A Figura 1 ilustra o fluxo de tomada de decisão entre os modelos e os testes de filtro, onde cada modelo avalia sequencialmente se são necessários testes adicionais de filtros.

Para testar a precisão de cada modelo a base de dados não foi completamente usada para treinamento, 20% foi separada para atestar a precisão da previsão.

No intuito de testar essa nova ferramenta implementada em eficiência de codificação, novos testes foram realizados, utilizando sete novos vídeos, que não participaram do processo de treinamento, quatro com resolução HD e três com resolução FULL HD. Da mesma maneira que nas outras codificações, utilizamos os parâmetros das condições comuns de testes e os CQs recomendados. Por fim, a versão com do codificador modificada foi comparada com a versão original nas métricas de redução de tempo e BD-Rate.

3. RESULTADOS E DISCUSSÃO

Os quatro modelos tiveram sua precisão testada utilizando a base de dados não utilizada no processo de treinamento, o primeiro modelo obteve 80,29% de precisão, o segundo 93,74%, o terceiro 92,75% e o quarto 93,3%.

Sequência	Resolução	BD-Rate (%)	Redução de Tempo Média (%)
<i>Vidyo3</i>	HD	0,268	2,24
<i>FourPeople</i>	HD	0,109	1,211
<i>ControlledBurn</i>	HD	0,004	1,936
<i>Shields_ter</i>	HD	0,497	1,975
<i>Touchdown_pass</i>	FULL HD	0,078	2,507
<i>CrowdRun</i>	FULL HD	0,133	1,867
<i>Boat</i>	FULL HD	0,161	2,055
Média	HD	0,22	1,841
Média	FULL HD	0,124	2,143
Média	HD/FULL HD	0,179	2,0712

Tabela 1: BD-Rate e média de redução de tempo para várias sequências

A Tabela 1 descreve o impacto na taxa de bits (BD-Rate) e a redução média no tempo de codificação, para todos os quatro valores de CQ, em comparação com a codificação original. Com base nesses resultados, observamos que a maior redução no tempo ocorreu na sequência "Touchdown pass", com uma diminuição de 2,5%, acompanhada por um pequeno aumento na BD-Rate. O menor aumento na BD-Rate ocorreu na sequência "ControlledBrun", mostrando um aumento de apenas 0,004%. No pior caso para a BD-Rate, a sequência "Shields ter" teve um aumento de 0,497%, enquanto para o tempo de codificação, a sequência "FourPeople" teve uma redução de tempo de apenas aproximadamente 1,21%. Em média, o FULL HD apresentou melhores resultados, com um aumento de 0,12% na BD-Rate e uma redução de 2,14% no tempo de codificação.

4. CONCLUSÕES

Este resumo apresentou o desenvolvimento de um modelo de predição inteligente para os filtros de interpolação do codificador AV1, motivado pela alta complexidade decorrente do teste de todos os filtros possíveis e, também, do desempenho ruim ao utilizar apenas o filtro de interpolação regular. O foco foi no desenvolvimento da base de dados e na ferramenta para predição, em conjunto da análise do seu impacto em eficiência de codificação, especificamente com as métricas de BD-Rate e tempo de codificação, além da análise da precisão dos modelos.

De maneira geral, cada modelo apresentou uma precisão alta e a implantação gerou um resultado parcialmente satisfatório, os modelos implementados apresentaram uma redução modesta de tempo, em média 2,07%, mas um baixo aumento em BD-Rate, em média 0,18%. Ao comparar com um experimento inicial, que utilizava apenas o filtro de interpolação regular, a implementação inteligente obteve um bom resultado na métrica de BD-Rate. No entanto, em quesito de tempo de execução, o uso de apenas um filtro reduziu o tempo de codificação em média 12,49%, bem acima do resultado obtido pela implementação dos modelos preditivos.

Embora com o uso do modelo preditivo tenhamos conseguido alcançar resultados interessantes em termos de BD-Rate, ainda existe espaço para investigações com foco na redução de tempo. Trabalhos futuros podem explorar a implementação de novos modelos em etapas diferentes da codificação, para obter resultados mais expressivos.

5. REFERÊNCIAS BIBLIOGRÁFICAS

BJONTEGAARD, G., Calculation of average PSNR differences between RD curves. *VCEG-M33*, 2001.

NOKIA. Global Network Traffic 2030 Report. Acessado em 09 de out. 2024. Online. Disponível em : <https://onestore.nokia.com/asset/213660>.

CHEN, Y. An Overview of Coding Tools in AV1: the First Video Codec from the Alliance for Open Media. **APSIPA Transactions on Signal and Information Processing**, v.9, 2020.

ZHAO, X. Alliance for Open Media, Codec Working Group Document: CWG-B075o - AOM Common Test Conditions v2.0. Alliance for Open Media, 2021.

AOMEDIACODEC. Sub-pel Interpolation and Interpolation Filter Search (Technical Appendices). Acessado em 09 de out. 2024. Online. Disponível em: <https://gitlab.com/AOMediaCodec/SVT-AV1>.