SOLUÇÃO BASEADA EM APRENDIZADO DE MÁQUINA PARA REDUÇÃO DA COMPLEXIDADE COMPUTACIONAL DO PROCESSO DE CODIFICAÇÃO INTRA QUADRO DO PADRÃO VVC

ANNA OLIVEIRA[1]; ADSON DUARTE[1]; DANIEL PALOMINO[1]

[1] Universidade Federal de Pelotas
{agmoliveira, airduarte, dpalomino}@inf.ufpel.edu.br

1. INTRODUÇÃO

Segundo a pesquisa da TIC Domicílios, realizada pelo Centro Regional para o Desenvolvimento de Estudos sobre a Sociedade da Informação (CETIC, 2019), 74% dos usuários de internet assistiram a programas, filmes, vídeos ou séries no Brasil em 2019. Assim, com o constante aumento tanto na demanda por este tipo de conteúdo quanto em sua resolução, torna-se imprescindível o estudo de técnicas capazes de realizar processos de compressão em um vídeo sem que perdas significativas na qualidade visual sejam inseridas. Desta forma, o padrão de codificação de vídeo *Versatile Video Coding* (VVC) foi desenvolvido com o objetivo de obter taxas de compressão maiores para uma mesma qualidade visual quando comparado ao seu antecessor, sendo este o *High Efficiency Video Coding Standard* (HEVC), resultando assim em uma eficiência de codificação maior.

Este aumento na eficiência de codificação se dá pelo desenvolvimento de várias ferramentas de codificação novas para o VVC, como por exemplo os 32 novos modos angulares para a predição intra, seleções de transformadas adicionais e novas estruturas de particionamento de bloco. Ainda que seja observado um aumento na eficiência de codificação, estas ferramentas também possuem um impacto negativo no custo computacional do processo de decisão de modo da predição intra-quadro, dado o número alto de combinação entre modos, transformadas e tamanhos de bloco que precisam ser avaliadas para a tomada de decisão. Sendo assim, torna-se importante o estudo de medidas capazes de reduzir o custo computacional do processo de decisão de modo intra do VVC.

Existem trabalhos que utilizam análise estatística e aprendizado de máquina supervisionado como estratégia para diminuir a quantidade de modos intra avaliados pelo processo de decisão de modo do VVC, a exemplo os trabalhos de Zhang (2020) e Saldanha; Sanchez; Marcon; Agostini (2021). Com a utilização de modelos de aprendizado de máquina supervisionado, torna-se possível a predição de quais modos intra são mais prováveis de serem escolhidos para a codificação de determinado bloco de uma sequência de vídeo. Entretanto, obter estes modelos não é uma tarefa trivial, uma vez que se estes lograrem uma acurácia baixa na tarefa de predizer a classe de modos intra mais provável para um bloco de uma sequência de vídeo, obter-se-ia uma situação não favorável entre redução no custo computacional e perda em eficiência de codificação. Uma vez que, a diminuição de um, resultaria em aumentos em outro.

Neste trabalho, o objetivo é utilizar um modelo de aprendizado de máquina supervisionado baseado em árvore de decisão como estratégia para acelerar o processo de decisão de modo intra no VVC. Primeiro, os modos intra do VVC foram agrupados em três classes, as quais são preditas pelo modelo. Então, a partir da predição do modelo para cada bloco do vídeo, são avaliados somente os

modos intra pertencentes a classe predita no processo de decisão de modo do VVC, o que reduz o custo computacional.

2. METODOLOGIA

Primeiramente, os modos de predição intra do VVC foram agrupados em três classes: Não angulares, Angulares e MIP. A classe dos Não Angulares agrupa os modos de predição intra Planar, DC, e *Intra Subpartition Prediction* (ISP) associado aos modos Planar e DC. A classe dos Angulares agrupa os 65 modos de predição intra angulares contidos no VVC, além de apresentar o modo de predição intra ISP associado a estes 65 modos de predição. Por último, a classe MIP agrupa os modos de predição intra *Matrix-based Intra Prediction* (MIP). O objetivo final do modelo é predizer, dentre estas três classes, a classe que contém o melhor modo intra para o bloco atual em uma sequência de vídeo.

Com as três classes definidas, o conjunto de dados necessário para treinar o modelo foi obtido através dos dados previamente extraídos no trabalho de Duarte (2021). Esta base de dados contém, aproximadamente, 200.000 registros e conta com 38 features, as quais consideram informações tais como o menor custo de taxa distorção obtido após o processo *Rough Mode* Decision (RMD) para cada um dos modos, a posição na lista de modos a serem avaliados pela decisão de modo para cada tipo de modo intra, e o número dos primeiros modos Angular e MIP na lista de modos a serem avaliados pela decisão de modo.

Com o conjunto de dados obtido, o treinamento da árvore de decisão foi realizado na biblioteca *scikit-learn* (PEDREGOSA et al., 2011) através de duas etapas principais, sendo que a primeira etapa consistiu em um *Random Search* (RS) e a segunda etapa consistiu em um *Grid* Search (GS). A etapa de RS foi realizada com o objetivo de identificar quais hiperparâmetros possuem uma maior correlação com o aumento da F1-Score. Assim, nesta etapa foram utilizados largos espaços de busca nos hiperparâmetros do modelo da árvore de decisão, sendo estes: criterion, min samples split, min samples leaf, max features, max depth e max leaf nodes. A partir dos resultados obtidos no RS, o coeficiente de Pearson entre cada hiperparâmetro e a F1-Score foi calculado, observou-se que os hiperparâmetros *max features* e *max leaf nodes* foram os dois hiperparâmetros que obtiveram maior correlação com o aumento da F1-Score. Desta forma, estes dois hiperparâmetros foram selecionados para seguirem na etapa de GS, onde todas as combinações destes hiperparâmetros em um espaço de busca otimizado e reduzido foram avaliadas através de uma validação cruzada (cross-validation) de 5, sendo que ao final escolheu-se a combinação que resultou na melhor F1-score.

Para obter os resultados, o modelo final foi implementado no software de referência *VVC Test Model* (VTM) 14.0 (SO/IEC, 2019). Após, três sequências de vídeo das *Common Test Conditions* (CTC) do VVC (BOSSEN et al., 2018) foram codificadas com a configuração *all intra* e com os valores de *Quantization Parameter* (QP) 22, 27, 32 e 37. As sequências selecionadas foram *FourPeople* (720p), *BasketballDrive* (1080p) e *FoodMarket4* (4K) e as codificações foram realizadas tanto no VTM original quanto no VTM com o modelo implementado, a fim de comparar os resultados obtidos para redução no custo computacional e eficiência de codificação. A redução no custo computacional foi medida ao comparar os tempos totais de codificação obtidos para o VTM original e para o VTM modificado, e a eficiência de codificação foi obtida através da métrica *Bjontegaard Delta Bit Rate* (BD-BR) (BJONTEGAARD, G., 2001).

3. RESULTADOS E DISCUSSÃO

A Figura 1 mostra a matriz de confusão para o modelo proposto, avaliado em um conjunto de teste à parte do processo de treinamento e validação, com tamanho equivalente a 20% da base de dados original. O eixo das abscissas é destinado às predições do modelo para cada uma das três classes definidas, no das ordenadas foram atribuídos os rótulos reais dos exemplos testados pelo modelo. O modelo obtém êxito quando a classe dos rótulos reais é equivalente à classe dos rótulos preditos, e, o número representado mensura quantos exemplos foram classificados corretamente. Para todos os outros casos, o número representará quantos registros da classe predita eram, na verdade, da classe real.

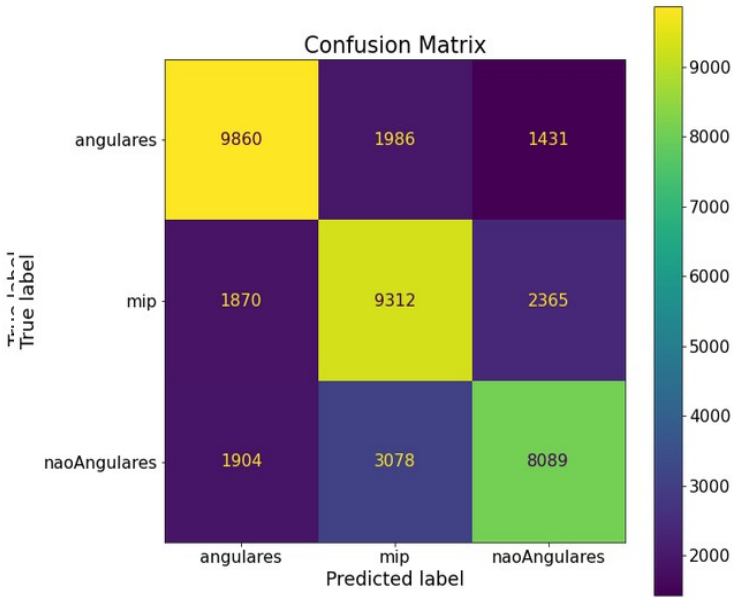

Figura 1: Matriz de confusão do modelo.

De acordo com a Figura 1, a classe que abrigou o maior número de acertos foi a dos Angulares. Além disso, a que abrigou o maior número de erros foi a dos não Angulares, com 4982 erros. O modelo tendeu a confundir as classes dos não Angulares e MIP. Obtendo, dessa forma, 68.59% na métrica f1-*score.*

A Tabela 1 mostra, para cada um dos vídeos, a economia de tempo, a perda em eficiência de codificação (BD-BR) e o tempo total de predição do modelo em relação ao tempo total de codificação. Estes valores foram obtidos ao comparar os tempos e eficiências de codificação obtidos no VTM original e no VTM com o modelo implementado.

Tabela 1: Resultados de tempo e eficiência de codificação.

Vídeo	Classe na CTC	Economia de Tempo	BD-BR	Tempo do modelo
FoodMarket4	A1	16,10%	1,24%	0,12%
BasketballDrive	B	18,19%	1,22%	0,18%
FourPeople	E	18,57%	1,49%	0,18%
Média		**17,62%**	**1,32%**	**0,16%**

Conforme a Tabela 1, a sequência de vídeo FourPeople obteve a melhor economia de tempo em 18,57%, como também a maior perda em eficiência de codificação em 1,49%. A pior economia de tempo atingida foi com a sequência FoodMarket4 com 16,10%, e a menor perda de eficiência de codificação foi da sequência BaskettballDrive em 1,22%. Além disso, observa-se que o modelo teve pouco impacto no tempo total de codificação, com resultados inferiores a 0.2%.

4. CONCLUSÕES

O presente trabalho mostra que há relativa vantagem ao considerar apenas os modos intra com maior probabilidade de ocorrência na codificação de vídeos pelo VVC. Sequencialmente a este trabalho, uma proposta similar está em fase de desenvolvimento, utilizando-se do mesmo processo com diferentes modelos de aprendizagem de máquina, sendo estes as florestas randômicas e XGBoost.

5. REFERÊNCIAS BIBLIOGRÁFICAS

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

BOSSEN. **JVET-J1010: JVET common test conditions and software reference configurations**, 2018. Acessado em 21 Jul. 2022. Online. Disponível em: https://tinyurl.com/2warwmua

CETIC. **TIC Domicílios,** 2019. Acessado em 05 Ago. 2022. Online. Disponível em: https://tinyurl.com/b3v8pmad

DUARTE, A. **Redução de Complexidade do Processo de Decisão de Modo da Predição Intra-Quadro do Codificador de Vídeo VVC utilizando Aprendizado de Máquina**, 2021. Dissertação (Mestrado em Ciência da Computação) - Curso de Pós-graduação em Ciência da Computação, Universidade Federal de Pelotas.

PEDREGOSA; VAROQUAUX; GA’’EL; GRAMFORT;MICHEL;THIRION;GRISEL; … outros. **Scikit-learn: Machine learning in Python.** ***Journal of Machine Learning Research***, 2011. p.2825–2830.

SALDANHA; SANCHEZ; MARCON; AGOSTINI. **Learning-Based Complexity Reduction Scheme for VVC Intra-Frame Prediction**, 2021. International Conference on Visual Communications and Image Processing (VCIP), 2021, Acessado em 10 Ago. 2022. Online. Disponível em: https://tinyurl.com/bdzmbphk

SO/IEC Moving Picture Experts Group; ITU-T Video Coding Experts Group. **VVC Test Model (VTM) Software**, 2019. Acessado em 10 Ago. 2022. Online. Disponível em: https://vcgit.hhi.fraunhofer.de/jvet/VVCSoftware_VTM

ZHANG, Y.; KWONG, S.; WANG, S. **Machine learning based video coding optimizations: A survey**. Information Sciences, China, v.506, p.395–423, 2020.