UM MODO DE DECISÃO RÁPIDO PARA A PREDIÇÃO INTRA DO VVC BASEADO EM APRENDIZADO DE MÁQUINA

ADSON DUARTE[1]; BRUNO ZATT[2]; GUILHERME CORREA[3]; DANIEL PALOMINO[4]

[1]*Universidade Federal de Pelotas – airduarte@inf.ufpel.edu.br*
[2]*Universidade Federal de Pelotas – zatt@inf.ufpel.edu.br*
[3]*Universidade Federal de Pelotas – gcorrea@inf.ufpel.edu.br*
[4]*Universidade Federal de Pelotas – dpalomino@inf.ufpel.edu.br*

1. INTRODUÇÃO

Em apenas um minuto de internet, cerca de um milhão de horas de vídeo são transmitidas (DIXON, 2023). Entretanto, vídeos sem compressão alguma exigem altas taxas de armazenamento e transmissão. Assim, o uso de codificadores de vídeo como o *Versatile Video Coding* (VVC) (BROSS et al., 2020) é indispensável para que aplicações sejam capazes de manipular vídeos em tempo real.

Com o objetivo de atender as crescentes demandas por vídeos de alta-resolução, a predição intra do VVC conta com tamanhos de bloco quadrados e retangulares, 65 modos Angulares e 60 modos *Matrix-Based Intra Prediction* (MIP). Entretanto, O VVC possui um processo de codificação até 34 vezes mais lento que o *High Efficiency Video Coding* (MERCAT et al., 2021), devido a maior quantidade de tamanhos de bloco e modos intra a serem considerados pelo modo de decisão. Desta forma, é importante desenvolver soluções capazes de reduzir o esforço computacional exigido pelo modo de decisão intra do VVC.

Liu et al. (2023) usa uma rede neural convolucional que recebe como entrada a imagem sendo processada pelo codificador e retorna como saída a probabilidade para cada modo intra, onde somente os modos com maior probabilidade são avaliados. Duas árvores de decisão treinadas com *features* do processo de codificação são utilizadas em Saldanha et al. (2021), com o objetivo de predizer quando a avaliação dos modos angulares e MIPs podem ser evitadas.

Neste trabalho, um modo de decisão rápido para a predição intra do VVC é proposto. Os modos intra do VVC são agrupados em três classes: Planar/DC, Angular (65 modos) e MIP (60 modos). Após, uma árvore de decisão é treinada através de *features* do processo de codificação para predizer a classe mais provável para cada bloco processado pelo codificador. A partir desta predição, a avaliação dos modos contidos na classe menos provável é evitada, reduzindo o esforço computacional exigido pelo VVC. Este trabalho se diferencia das propostas de Liu et al. (2023) e de Saldanha et al. (2020) por utilizar um modelo mais simples e por utilizar apenas uma árvore de decisão ao invés de duas, respectivamente.

2. METODOLOGIA

A Figura 1 ilustra o modo de decisão rápido para a predição intra do VVC proposto. Os retângulos amarelos representam as etapas da nossa solução. No início, o *Rough Mode* Decision (RMD) é executado para obter a RD-List, um subconjunto de modos a serem avaliados no *Rate-Distortion Optimization Process* (RDO). O RMD gera informações úteis, como custos rápidos e modos mais prováveis com base nos blocos vizinhos. Na RD-List, os modos a serem avaliados são ordenados com base em seus custos, o que também é uma informação útil.

Assim, as *features* de entrada para a árvore de decisão são derivadas tanto do RMD quanto da RD-List. A árvore de decisão classifica em três classes: Planar/DC, Angular e MIP. Se a árvore prevê a classe Planar/DC, modos Angulares ou MIPs não são avaliados pelo RDO. Da mesma forma, se a árvore prevê a classe Angular ou MIP, modos MIPs e modos Angulares não são avaliados no RDO, respectivamente. Os modos Planar e DC são sempre avaliados pois existe uma alta probabilidade de estes serem os melhores modos (ZOUIDI et al., 2023).

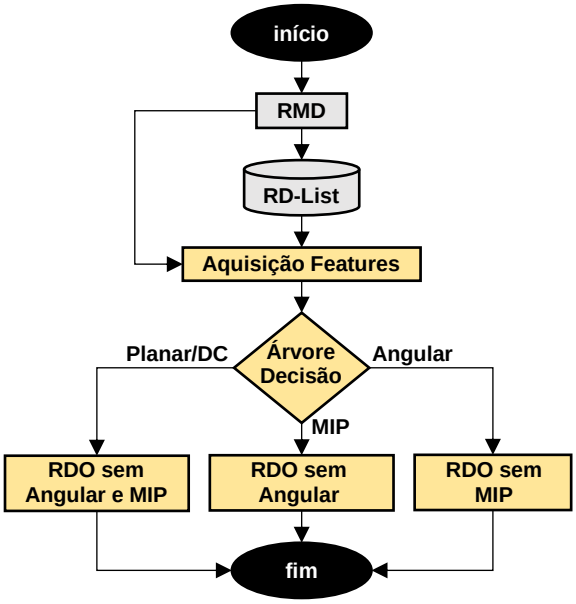

Figura 1: Modo de decisão rápido proposto.

Um total de 51 *features* são adquiridas: 35 do RMD, 8 da RD-List e 8 de informações do bloco (largura, altura, posições x e y, parâmetro de quantização (QP), particionamento). Do RMD, extraiu-se: soma das diferenças absolutas, soma das diferenças absolutas transformadas, bits estimados e custo taxa-distorção para os modos Planar, DC, Angular e MIP. Também foi obtida a linha de referência usada nos modos DC e Angular, o modo Angular e MIP com menor custo, e se o modo MIP é transposto. A partir da lista de modos mais prováveis pelo RMD, obtemos 14 *features*, incluindo os 5 modos mais prováveis e indicadores booleanos dos melhores modos nos blocos vizinhos. Além disso, extraímos posições da primeira ocorrência de um modo Planar, DC, Angular e MIP, o número dos primeiros modos Angular e MIP, e a quantidade de modos Angulares e MIPs em sequência nas primeiras posições da RD-List. Para obter o conjunto de dados, a mesma metodologia de Duarte et al. (2023) foi seguida, e cerca de 800 mil exemplos balanceados por vídeo, tamanho de bloco, QP e classe foram extraídos.

O treinamento da árvore de decisão foi realizado com uma busca de hiperparâmetros dividida em duas etapas: *Random Search* (RS) e *Grid Search* (GS). Na etapa de RS, os hiperparâmetros considerados foram: *criterion*, *min samples split*, *min samples leaf*, *max features*, *max depth* e *max leaf nodes*, e o objetivo foi encontrar os dois hiperparâmetros com maior correlação no aumento da F1 do modelo. Na etapa de GS, os dois hiperparâmetros com a maior correlação, sendo estes *max features* e *max leaf nodes*, passaram por um refinamento em uma busca exaustiva. As etapas de RS e de GS avaliaram as combinações através de *cross-validation* com k = 5. O modelo final obteve acurácias de 71%, 75% e 70% para as classes Planar/DC, Angular e MIP, respectivamente.

3. RESULTADOS E DISCUSSÃO

A solução foi implementada no VTM 18.0 e avaliada conforme as Condições Comuns de Teste (CTC) do VVC (BOSSEN et al., 2020), onde 22 vídeos foram codificados com os valores de QP 22, 27, 32 e 37, tanto no VTM original quanto no VTM com nossa solução. Os 22 vídeos são distribuídos em seis classes de resolução: seis em 4K (classes A1 e A2), cinco em Full HD (classe B), três em HD (classe E), quatro em 480p (classe C) e quatro em 240p (classe D). O desempenho da solução foi medido em termos de redução no tempo de codificação (RT) ao comparar os tempos de codificação do VTM original e do VTM com a nossa solução, e em termos de eficiência de compressão através da métrica *Bjontegaard Delta Bit Rate* (BD-BR) (BJONTEGAARD, 2001), onde valores positivos indicam uma perda em eficiência de compressão e valores negativos indicam o oposto. **Nenhum** dos vídeos da CTC do VVC foram vistos pelo modelo no treinamento.

Na Tabela 1 são apresentados os resultados de RT e de BD-BR para cada classe de vídeos da CTC da VVC. A solução obtém uma média de RT de 17,11% com apenas 0,84% de perda em eficiência de compressão. Para diferentes classes de vídeos, a RT se mantém semelhante, com um mínimo de 16,18% para a classe A2 e um máximo de 18,15% para a classe B, o que demonstra a efetividade da solução proposta em reduzir o esforço computacional do VVC no modo de decisão.

Tabela 1: Resultados de Redução no Tempo de Codificação e de BD-BR.

Classe	RT	BD-BR
A1	17,24%	0,86%
A2	16,18%	0,65%
B	18,15%	0,75%
C	16,55%	0,87%
D	17,65%	0,77%
E	16,88%	1,12%
Média	**17,11%**	**0,84%**

Nossa solução é comparada com os trabalhos relacionados na Tabela 2 ao considerar somente os vídeos com resultados em todas as soluções, com foco na versão do software, RT, BD-BR, e a razão RT/BD-BR, a qual indica para cada 1% de perda em eficiência de compressão o quanto a solução ganhou em RT. A comparação com Liu et al. (2023) mostra que nossa solução obtém uma RT superior, mantendo o mesmo BD-BR, mesmo com um modelo mais simples. O trabalho de Saldanha et al. (2021) utiliza duas árvores de decisão para decidir se os modos angulares e MIPs devem ser avaliados, o que às vezes resulta em perda de RT, já que ambas as árvores podem selecionar esses modos para o mesmo bloco. Em contraste, nossa solução utiliza apenas uma árvore de decisão, o que garante ganhos consistentes em RT, pois não existe a possibilidade de tanto os modos angulares quanto MIPs serem avaliados para um mesmo bloco, justificando assim a maior RT e o maior BD-BR em comparação com Saldanha et al. (2021).

Tabela 2: Comparação dos Resultados com Trabalhos Relacionados.

Solução	Software	RT	BD-BR	RT/BD-BR
Nossa	**VTM 18.0**	**17,20%**	**0,83%**	**20,72**
Liu et al. (2023)	VTM 14.0	16,72%	0,83%	20,14
Saldanha et al. (2021)	VTM 10.0	10,47%	0,29%	36,10

4. CONCLUSÕES

Um modo de decisão rápido para a predição intra do VVC foi proposto, onde uma árvore de decisão foi treinada com *features* do processo de codificação para prever a classe de modos mais provável para cada bloco codificado. Isso permitiu evitar a avaliação de modos menos prováveis, e os resultados demonstraram uma redução significativa no tempo de codificação do VVC. Em comparação com trabalhos relacionados, nossa solução alcança uma redução de tempo de codificação superior e é competitiva no *trade-off* entre redução do tempo de codificação e perda de eficiência de compressão.

5. REFERÊNCIAS BIBLIOGRÁFICAS

BJONTEGAARD, G. **Calculation of average PSNR differences between RD-curves**. VCEG Meeting, https://www.itu.int/wftp3/av-arch/video-site/0104_Aus/VCEG-M33.doc.

BOSSEN, F. et al. **VTM common test conditions and software reference configurations for SDR video**. JVET-T2010-v1, https://jvet-experts.org/doc_end_user/current_document.php?id=10545.

BROSS, B.; CHEN, J.; LIU, S.; WANG, Y.-K. **Versatile Video Coding Editorial Refinements on Draft 10**. JVET-T2001-v2, https://jvet-experts.org/doc_end_user/current_document.php?id=10540.

DIXON, S. J. **Media usage in an internet minute as of April 2022**. Disponível em: <https://www.statista.com/statistics/195140/new-user-generated-content-uploaded-by-users-per-minute/>.

DUARTE, A.; ZATT, B.; CORREA, G.; PALOMINO, D. Fast Intra Mode Decision Using Machine Learning for the Versatile Video Coding Standard. In: IEEE INTERNATIONAL SYMPOSIUM ON CIRCUITS AND SYSTEMS (ISCAS), 2023, Monterey, CA, USA. **Anais. . .** IEEE, 2023. p.1–5.

LIU, Z. et al. Deep Multi-task Learning based Fast Intra-mode Decision for Versatile Video Coding. **IEEE Transactions on Circuits and Systems for Video Technology**, p.1–1, 2023.

MERCAT, A. et al. Comparative Rate-Distortion-Complexity Analysis of VVC and HEVC Video Codecs. **IEEE Access**, v.9, p.67813–67828, 2021.

SALDANHA, M.; SANCHEZ, G.; MARCON, C.; AGOSTINI, L. Learning-Based Complexity Reduction Scheme for VVC Intra-Frame Prediction. In: INTERNATIONAL CONFERENCE ON VISUAL COMMUNICATIONS AND IMAGE PROCESSING (VCIP), 2021, Munich, Germany. **Anais. . .** IEEE, 2021. p.1-5.

ZOUIDI, N. et al. Complexity assessment of the intra prediction in Versatile Video Coding. **Multimedia Tools and Applications**, p.1–20, 2023.