

DECISÃO RÁPIDA DE PARTICIONAMENTO NO CODIFICADOR VERSATILE VIDEO CODING COM USO DE RANDOM FORESTS

MATHEUS LINDINO¹; BRUNO ZATT¹; MATEUS GRELLERT²; GUILHERME CORREA¹

¹*Universidade Federal de Pelotas – Video Technology Research Group*

²*Universidade Federal de Santa Catarina – Embedded Computing Laboratory*

¹*{mclindino, zatt, gcorrea}@inf.ufpel.edu.br, ²mateus.grellert@ufsc.br*

1. INTRODUÇÃO

A demanda por serviços de *streaming* de vídeos cresceu significativamente no cenário mundial atual. Neste sentido, uma pesquisa levantada em (CISCO, 2017) estima que, em 2021, 82% do tráfego de dados da internet são provenientes desses serviços. A quantidade de dados necessários para representar mídias digitais vem crescendo em virtude do aumento da informação presente na tela. Isso se deu como surgimento de novas tecnologias, por exemplo, *Ultra High-Definition* (UHD) e *Virtual Reality* (VR).

O *High Efficiency Video Coding* (HEVC) (Sullivan et al., 2012) não consegue comprimir com eficácia esses tipos de mídia, apesar dele já ter sido considerado o mais adequado para codificação de vídeos. Em 2015, um novo padrão de codificação de vídeo passou a ser desenvolvido pelo *Moving Picture Experts Group* (MPEG), denominado *Versatile Video Coding* (VVC) (BROSS, 2018). Esse padrão apresentou novos algoritmos e ferramentas que permitem adaptação e alto grau de liberdade para o codificador realizar as suas operações com unidades de dados em diferentes formatos tamanhos.

O VVC mostrou-se mais competente em questão de eficiência de codificação que seu predecessor, principalmente em resoluções UHD. Todavia, ele acaba se tornando computacionalmente mais complexo. Uma análise realizada em (Pakdaman et al., 2020) demonstrou que a codificação VVC é 31 vezes mais complexa que a codificação HEVC na configuração *All-Intra* e cinco vezes na configuração *Low-Delay*. Logo, estratégias de otimização que possuem o objetivo de diminuir a complexidade sem alterar a eficiência de codificação são extremamente necessárias para o VVC. Desta forma, este trabalho intenta desenvolver modelos preditivos, utilizando técnicas de *machine learning*, a fim de acelerar o tempo de codificação do VVC sem alterar a qualidade de compressão.

2. REFERENCIAL TEÓRICO

O particionamento do VVC introduziu em seu processo as estruturas *Quadtrees* (QTs) e *Multi-type Trees* (MTs). Elas são utilizadas para dividir o quadro do vídeo em blocos menores, com o intuito de agrupar a máxima quantidade de pixels homogêneos. Cada quadro é fracionado em blocos de tamanhos iguais de 128x128 pixels, chamados de *Coding Tree Units* (CTUs). Na sequência, as CTUs são particionadas pela QT ou pela estrutura MT, gerando blocos de tamanhos variados, chamados de *Coding Units* (CUs). Caso a CTU seja particionada pela estrutura QT, o resultado são quatro blocos quadrados (64x64, 32x32, 16x16 ou 8x8). Se utilizada a estrutura MT, quatro tipos de particionamentos (Fig. 1) podem ser realizados: Binário Vertical (BV), Binário Horizontal (BH), Ternário Vertical (TV) e Ternário Horizontal (TH).

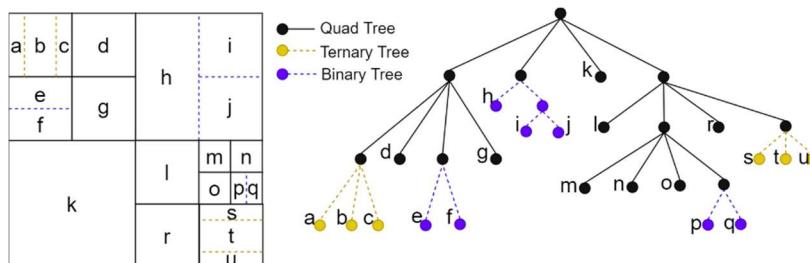


Figura 1 - Estrutura QT e MT: CTU (esq.) e árvore correspondente (dir.)

Cabe ressaltar que após o início do uso da estrutura MT, não é permitido realizar um particionamento do tipo QT. Também, o codificador VVC deve determinar o modo de predição que será utilizado para uma determinada CU, podendo ser intra-quadro (que explora redundância espacial) ou inter-quadros (que explora redundância temporal). Logo, em um cenário em que se busca o particionamento ótimo para um determinado bloco, a decisão do melhor particionamento é complexa; uma vez que o codificador testará todas as combinações de QT e MT existentes de particionamento.

Uma abordagem possível para acelerar a decisão de particionamento do codificador VVC, é utilizar *machine learning* para identificar as necessidades de particionamento de blocos. Nesse contexto, as Random Forests (RF) são uma classe de algoritmos frequentemente utilizados. Em comparação com as árvores de decisão, RFs são menos propensas a sofrer *overfitting*, pois são compostas por várias árvores de decisão construídas sistematicamente pela seleção pseudoaleatória de subconjuntos de recursos (KAM HO, 1998). Desta forma, se um conjunto relevante de variáveis de entrada é usado no processo de treinamento da RFs, estas podem alcançar alta precisão ao custo de uma pequena sobrecarga em termos de uso de recursos computacionais.

Sobre VVC, os autores (YANG et al., 2019) propõem um algoritmo para diminuir a complexidade da estrutura QT e MT, no modo intra-quadro do codificador, utilizando árvores de decisão. (WIECKOWSKI et al., 2019) relata diversas regras e heurísticas para limitar a profundidade das árvores QT e MT, tanto na configuração *Random Access* (RA) quanto *All Intra* (AI). (FAN et al., 2020) utiliza o cálculo da variância dos blocos para decidir sua partição, acelerando o codificador. As técnicas desenvolvidas pelos autores (YANG et al., 2019) e (FAN et al., 2020), apesar de acelerarem a decisão de particionamento do bloco do codificador VVC, focam apenas na configuração intra do codificador, com versão em desuso do software de referência VTM. Também, em (WIECKOWSKI et al., 2019), observa-se que a criação de heurísticas apresenta regras genéricas importadas do HEVC, o que dificulta o seu uso em versões do VVC mais atuais.

Nesse contexto, e considerando que as estratégias presentes na literatura não são suficientes para lidar com a alta complexidade, este trabalho propõe desenvolver modelos preditivos baseados em *machine learning*.

3. METODOLOGIA

Para a execução deste trabalho, foram retirados vídeos da base de dados do *Ultra Video Group* (UVG) e calculado os valores de *Spatial Index* (SI) e *Temporal Index* (TI) (ITU-R, 2008), totalizando 15 vídeos selecionados. Eles foram codificados em quatro QPs (22, 27, 32, 37) com o software de referência VTM 9.0



(VTM). No momento da decisão de particionamento, foram extraídas as seguintes informações do bloco atual:

- QP: parâmetro de quantização do *slice* atual;
- Localização espacial do bloco;
- Profundidade das estruturas de particionamento (QT e MTT);
- Variância, média e gradiente horizontal e vertical dos pixels;
- Razão dos gradientes.

Também foram extraídas as seguintes informações dos sub-blocos que podem particionados:

- Variância, média e gradiente horizontal e vertical dos sub-blocos;
- Inconsistência Horizontal (IH) e Inconsistência Vertical (IV) dos possíveis sub-blocos;
- Diferença entre IH e IV.

Com isso, foi possível estabelecer um conjunto de dados para ser utilizado na etapa de treinamento dos modelos preditivos. Após essa coleta, os dados foram divididos em 7 *datasets*, baseados na área do bloco, como mostra a Tabela 1.

S0	S1	S2	S3	S4	S5	S6
128x128	128x64	64x64	64x32	32x32	32x16	16x16
	64x128	128x32	32x64	64x16	16x32	32x8
		32x128	128x16	16x64	64x8	8x32
			16x128	128x8	8x64	
				8x128		

Tabela 1. Divisão do *dataset*

Na sequência, foram definidos três modelos para cada conjunto de dados:

- **QT_SPLIT**: Modelo para definir se o bloco atual deve ser particionado em quatro sub-blocos;
- **HORZ_SPLIT**: Decide se o bloco atual deve ser particionado horizontalmente (binária ou ternária);
- **VERT_SPLIT**: Decide se o bloco atual deve ser particionado em binária ou ternária vertical.

Desta forma, foram treinados os modelos baseados em *Random Forest* e validadas com a técnica de *cross-validation* com cinco *folds*, no qual avalia a capacidade de generalização de um modelo. Também, foi definido o número máximo de árvores de decisão (até 21) e o número máximo de profundidade de cada árvore (51), uma vez que esses dois parâmetros podem criar modelos complexos, caso não controlados.

4. RESULTADOS E DISCUSSÃO

Cada modelo desenvolvido foi implantado no codificador VTM 9.0, a fim de avaliar o impacto desses em *BD-Rate* (Métrica corresponde a qualidade de compressão) e *Time Saving* (Redução de complexidade do codificador). Para tanto, foi codificado um vídeo fora dos presentes no treinamento, para que os modelos não sofram *overfitting*. A Tabela 2 apresenta os resultados de cada subconjunto para cada estrutura de particionamento do VVC.

É possível afirmar que grande parte da redução de complexidade é na MTT, uma estrutura nova inserida no VVC. Além disso, vale ressaltar que o conjunto S2 foi o que obteve o melhor desempenho, uma vez que teve uma redução de tempo considerável (22,06%) para uma perda de eficiência de codificação mínima (0,2%).



	QT		MTT	
	BD-Rate (%)	TS (%)	BD-Rate (%)	TS (%)
s0	22,59%	21,88%	0,00%	4,18%
s1	-	-	0,00%	1,75%
s2	4,55%	3,93%	0,20%	22,06%
s3	-	-	0,37%	14,19%
s4	0,81%	-0,34%	11,59%	26,79%
s5	-	-	1,74%	16,25%
s6	-	-	0,23%	4,49%
Todos	23,40%	31,15%	16,08%	63%

Tabela 2. Resultados de cada dataset no VTM 9.0

5. CONCLUSÕES

O presente trabalho apresenta uma abordagem com *machine learning* para a redução de complexidade do codificador VVC. Na estrutura MTT, foi possível obter uma redução de tempo de codificação de 63%, para uma perda na eficiência de codificação de 16,08%. Isso mostra que há diversos testes realizados na etapa de particionamento que podem ser evitados, sem alterar o resultado final. Em trabalhos futuros pretende-se refinar os modelos, como utilizar a técnica de *GridSearch* para parametrizar as RFs. Além disso, explorar novas divisões de datasets, utilizando algoritmos de agrupamento para agrupar os blocos com *features* semelhantes.

6. REFERÊNCIAS BIBLIOGRÁFICAS

- CISCO. Cisco Visual Networking Index Predict Global Annual IP Traffic to Exceed Three Zettabyter by 2021. Disponível em: <<https://newsroom.cisco.com/press-release-content?articleId=1853168>>, 2021.
- Sullivan, G. J.; Ohm, J.; Han, W.; Wiegand, T. Overview of the High Efficiency VideoCoding (HEVC) Standard. *IEEE Transactions on Circuits and Systems for VideoTechnology*, [S.I.], v.22, n.12, p.1649–1668, 2012.
- BROSS, B. *Versatile Video Coding (Draft 1)*. JVET-J1001. 10th JVET Meeting.
- Pakdaman, F.; Adelimanesh, M. A.; Gabbouj, M.; Hashemi, M. R. Complexity AnalysisOf Next-Generation VVC Encoding And Decoding. In: *IEEE INTERNATIONAL CON-FERENCE ON IMAGE PROCESSING (ICIP)*, 2020.
- Tin Kam Ho, "The random subspace method for constructing decision forests," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 8, pp. 832-844, Aug. 1998, doi: 10.1109/34.709601.
- YANG, H. et al. Low-complexity CTU partition structure decision and fast intra mode decision for versatile video coding. *IEEE Transactions on Circuits and Systems for Video Technology*, [S.I.], v.30, n.6, p.1668–1682, 2019.
- WIECKOWSKI, A. et al. Fast partitioning decision strategies for the upcoming versatilevideo coding (VVC) standard. In: *IEEE INTERNATIONAL CONFERENCE ON IMAGEPROCESSING (ICIP)*, 2019., 2019. Anais. . [S.I.: s.n.], 2019. p.4130–4134.
- FAN, Y. et al. A fast QTMT partition decision strategy for VVC intra prediction. *IEEE Access*, [S.I.], v.8, p.107900–107911, 2020.
- Versatile Video Coding Test Model(VTM) 9.0: https://vcgit.hhi.fraunhofer.de/jvet/VVCSoftware_VTM/-/tree/VTM-9.0
- ITU-R P.910, "Subjective Video Quality Assessment Methods for Multi-Media Applications," (2008).