

UTILIZANDO APRENDIZADO DE MÁQUINA SUPERVISIONADO PARA PREDIZER UM SUBCONJUNTO DE MODOS INTRA DO VERSATILE VIDEO CODING (VVC)

ADSON DUARTE¹; GUILHERME CORREA¹; BRUNO ZATT¹; DANIEL
PALOMINO¹

¹Universidade Federal de Pelotas – Programa de Pós-Graduação em Computação
Video Technology Research Group (ViTech)
{airduarte, gcorrea, zatt, dpalomino}@inf.ufpel.edu.br

1. INTRODUÇÃO

Um estudo realizado pela CISCO apresentou que em 2016 o fluxo de dados na internet relacionados a transmissão de vídeo correspondeu a 73% do tráfego total e que até 2022 existe uma estimativa deste valor subir para 82% (CISCO, 2020). Desta forma, a área de codificação de vídeo desempenha um papel importante ao estudar e desenvolver técnicas que realizam o processo de compressão de um vídeo com baixo impacto em sua qualidade, com o objetivo de reduzir a quantidade de dados que precisa ser armazenada e/ou transmitida.

Existem diversos padrões de codificação de vídeo. Neste trabalho, será utilizado o *Versatile Video Coding* (VVC), um padrão que está sendo desenvolvido pelo grupo *Joint Video Experts Team* com o objetivo de aumentar as taxas de compressão para uma mesma qualidade visual quando comparado ao padrão anterior *High Efficiency Video Coding* (HEVC).

Para atingir tal objetivo, o VVC incorpora novas ferramentas e algoritmos de codificação. Somente na predição intra, o número de modos de predição direcionais que eram trinta e três e o número de tamanhos de bloco que eram cinco no HEVC, passaram para sessenta e cinco e dezessete no VVC, respectivamente (CHEN et al., 2019). Isto contribui para que se tenha uma eficiência de codificação maior em relação ao HEVC. Entretanto, um número maior de modos intra e de tamanhos de bloco disponíveis aumenta a complexidade do codificador, pois existem mais possibilidades a serem avaliadas pelo processo de decisão de modo.

O processo de decisão de modo é responsável por escolher qual o melhor particionamento e o melhor modo de predição para um determinado bloco do vídeo com base em um custo. Este custo é obtido através do método *Rate-Distortion Optimization* (RDO) (SULLIVAN; WIEGAND, 1998), um processo complexo que precisa realizar todas as etapas de codificação para obter o custo de cada modo de predição. Como seria inviável aplicar o RDO para cada uma das combinações possíveis entre particionamentos e modos disponíveis, apenas um subconjunto de modos é selecionado para ser testado pelo RDO.

Existem trabalhos que buscam reduzir o número de modos testados pelo RDO no HEVC através da relação entre características extraídas dos blocos a serem preditos e os modos intra. As características que aparecem nestes trabalhos são estimativas de gradiente, como Sobel, Prewitt ou Roberts (JAMALI et al., 2015) (ZHANG et al., 2016) (WANG; XUE, 2016), e medidas estatísticas, como variância ou desvio padrão (SUN et al., 2019) (LIU et al., 2017). Todos trabalhos obtiveram reduções no tempo de execução com baixo acréscimo na métrica BDBR (BJONTEGAARD, 2001). Entretanto, os trabalhos feitos com o HEVC não servem para o VVC já que o número de particionamentos e modos intra aumentou. Além disso, nenhum dos trabalhos realizou uma avaliação de

todas as características juntas ou utilizou aprendizado de máquina como estratégia para reduzir o número de modos testados pela decisão de modo, e aplicar esta estratégia pode vir a trazer melhores resultados. Por fim, devido à alta complexidade gerada pelo aumento de particionamentos e de modos intra no VVC, existe a necessidade de se desenvolver modos de decisão rápidos para a predição intra do mesmo. Desta forma, o objetivo geral deste trabalho é desenvolver um modo de decisão rápido para a predição intra do VVC considerando os modos direcionais. Este modo de decisão rápido será baseado em Random Forests, uma técnica de aprendizado de máquina supervisionado que é leve e costuma gerar bons resultados.

2. METODOLOGIA

Inicialmente, foi necessário gerar os datasets a serem utilizados no treinamento das Random Forests. Para tanto, foram selecionados cinco vídeos de três resoluções diferentes, 720p, 1080 e 4k, totalizando quinze vídeos. Cada um destes vídeos passou pelo processo de codificação do VVC com quatro valores de QP (22, 27, 32 e 37) e com o codificador configurado para utilizar apenas predição intra. Durante o processo de codificação, as posições x e y, número do frame e a decisão de modo retornada pelo RDO foram salvas para cada bloco que era predito.

Após finalizar esta extração, algumas decisões foram tomadas antes de gerar os datasets. Primeiro, como o VVC possui dezessete tamanhos de bloco, decidiu-se agrupar tamanhos de bloco que possuem largura igual, o que totalizou cinco conjuntos de tamanhos de bloco. Realizar a predição de exatamente apenas um modo intra seria inviável, já que existem sessenta e cinco modos intra direcionais. Logo, estes também foram agrupados em classes de acordo com as direções. A Figura 1 mostra como este agrupamento foi feito de duas formas. Por fim, para cada um dos cinco conjuntos de tamanhos de bloco, foi gerado um dataset pseudoaleatório contendo cerca de oitocentos mil exemplos balanceados por vídeo, número do frame, tamanho do bloco, QP e classe.

Neste momento, os datasets contêm apenas as posições x e y dos blocos e a classe referente ao modo intra. Ainda são necessárias as características de direção de textura extraídas dos blocos. Sendo assim, foi implementado em Matlab um script que leu os datasets e de acordo com o vídeo ao qual o bloco pertence, número do frame e posições x e y aplicou os operadores Sobel, Prewitt, Roberts, Desvio Padrão e Variância sobre o bloco e salvou estes resultados no dataset.

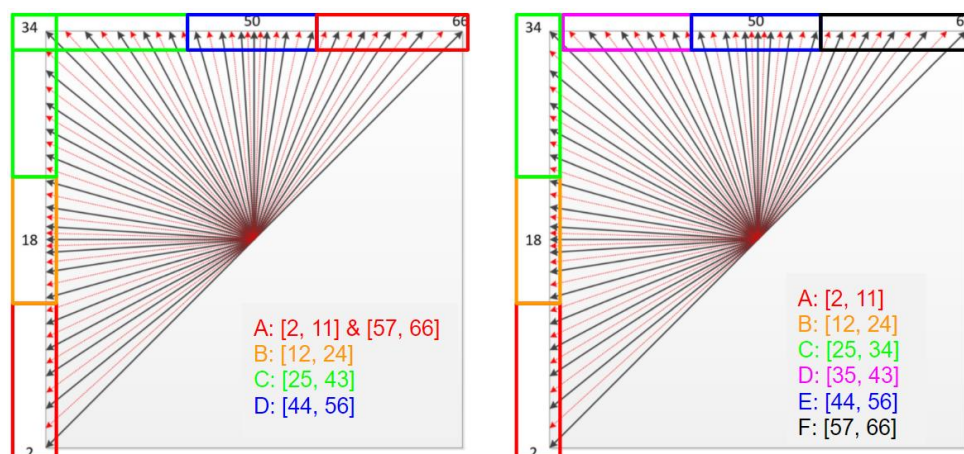


Figura 1: Modos Intra Direcionais Agrupados em Quatro e Seis Classes

Após calcular os operadores, foi feita uma seleção das características que possuíam maior ganho de informação. Dentre os operadores que estimam o gradiente, o Sobel foi selecionado já que obteve uma maior média de ganho de informação em relação ao Prewitt e ao Roberts, enquanto para as medidas estatísticas a Variância obteve uma maior média de ganho de informação em relação ao Desvio Padrão. Após esta seleção de características, o dataset ficou com doze características.

Com os datasets prontos, chega o momento de verificar a precisão dos modelos. Como a Random Forest possui diversos hiperparâmetros que são definidos pelo usuário e influenciam na precisão final do modelo, foi rodado o RandomSearch (RS) em cima de alguns destes hiperparâmetros para verificar quais possuem uma maior correlação com o aumento da F-Score. Depois, foi rodado o GridSearch (GS) apenas sobre os dois hiperparâmetros que possuem maior correlação com a F-Score. Tanto no RS quanto no GS o dataset foi dividido em 75% para treinamento e 25% para teste, sendo que nos 75% para treinamento foi realizado Cross Validation (CV) com 5 folds.

3. RESULTADOS E DISCUSSÃO

A Tabela 1 apresenta a F-score tanto no treinamento quanto no teste para a melhor combinação de hiperparâmetros testada no GS. Ainda que as F-scores para os datasets de quatro classes sejam maiores, cabe destacar que em ambos os casos a melhora em relação a uma decisão totalmente aleatória foi a mesma, já que para o caso de quatro classes a melhora foi de cerca de 55% e para o caso de seis classes a melhora foi de cerca de 57%. Além disso, a quantidade de modos por classe é menor no caso de seis classes, e isto aumenta as chances de que mais modos sejam removidos do RDO. De qualquer forma, ambos os casos precisam ser implementados e testados no VVC para verificar os impactos de BDBR e de tempo de execução.

Tabela 1: F-Score para cada Dataset após o GS

Nº de Classes	Dataset	F-score (%) CV=5	F-score (%) Teste
4	4x4_4x32	75,70	78,36
	8x4_8x32	77,73	80,54
	16x4_16x32	77,95	80,75
	32x4_32x32	78,50	80,01
	64x64	83,07	84,49
6	4x4_4x32	67,57	70,87
	8x4_8x32	69,32	72,58
	16x4_16x32	69,31	72,45
	32x4_32x32	70,50	72,06
	64x64	81,59	81,71

4. CONCLUSÕES

Este trabalho apresentou uma metodologia para gerar um modelo de aprendizado de máquina capaz de prever um conjunto de modos intra direcionais relacionados a um bloco. Os próximos passos consistem em



implementar os modelos no codificador e verificar os impactos no tempo de execução e no BDBR.

5. REFERÊNCIAS BIBLIOGRÁFICAS

BJONTEGAARD, G. **Calculation of average PSNR differences between RD-curves**. 26 mar. 2001. Acessado em 26 set. 2020. Online. Disponível em: https://www.itu.int/wftp3/av-arch/video-site/0104_Aus/VCEG-M33.doc.

CHEN, J. et al. **Algorithm description for Versatile Video Coding and Test Model 5 (VTM 5)**. 11 jun. 2019. Acessado em 26 set. 2020. Online. Disponível em: http://phenix.it-sudparis.eu/jvet/doc_end_user/current_document.php?id=6641.

CISCO. **Cisco Annual Internet Report (2018-2023) White Paper**. 09 mar. 2020. Acessado em 26 set. 2020. Online. Disponível em: <https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html>.

JAMALI, M. et al. Fast HEVC intra mode decision based on edge detection and SATD costs classification. In: **2015 DATA COMPRESSION CONFERENCE**, 2015, Snowbird, 2015, **Anais...** IEEE, [s.n.], p.43-52.

LIU, X. et al. An Adaptive Mode Decision Algorithm Based on Video Texture Characteristics for HEVC Intra Prediction. **IEEE Transactions on Circuits and Systems for Video Technology**, [S.l.], v.27, n.8, p.1737-1748, 2017.

SULLIVAN, G. J.; WIEGAND, T. Rate-distortion optimization for video compression. **IEEE signal processing magazine**, [S.l.], v.15, n.6, p.74-90, 1998.

SUN, X. et al. Fast CU size and prediction mode decision algorithm for HEVC based on direction variance. **Journal of Real-Time Image Processing**, [S.l.], v.16, n.5, p.1731-1744, 2019.

WANG, X; XUE, Y. Fast HEVC intra coding algorithm based on Otsu's method and gradient. In: **2016 IEEE INTERNATIONAL SYMPOSIUM ON BROADBAND MULTIMEDIA SYSTEMS AND BROADCASTING (BMSB)**, 2016, Nara, 2016, **Anais...** IEEE, [s.n.], p.1-5.

ZHANG, T. et al. Fast intra-mode and CU size decision for HEVC. **IEEE Transactions on Circuits and Systems for Video Technology**, [S.l.], v.27, n.8, p.1714-1726, 2016.