

ACELERAÇÃO DO TEMPO DE CODIFICAÇÃO V-PCC USANDO MODELOS DE DECISÃO POR APRENDIZADO DE MÁQUINA

GUSTAVO REHBEIN; CRISTIANO SANTOS; GUILHERME CORREA; MARCELO PORTO

Universidade Federal de Pelotas – ghrehbein@inf.ufpel.edu.br, cfd santos@inf.ufpel.edu.br,
gcorrea@inf.ufpel.edu.br, porto@inf.ufpel.edu.br

1. INTRODUÇÃO

As nuvens de pontos constituem uma forma de mídia digital tridimensional composta por um conjunto de pontos distribuídos em um espaço 3D, aos quais podem ser associadas informações adicionais, como cor ou outros atributos. Esse tipo de representação é amplamente empregado em áreas como reconstrução de patrimônios históricos, transmissão de vídeos 3D imersivos, aplicações em realidade virtual e aumentada, navegação autônoma e também em imagens médicas. Quando organizadas em sequências temporais, formando as chamadas nuvens de pontos dinâmicas, tornam-se equivalentes a vídeos digitais 2D, mas com um volume de dados muito superior.

O padrão *Video-based Point Cloud Compression* (V-PCC) (GRAZIOSI et al., 2020), desenvolvido pelo MPEG, surgiu como solução para a compressão dessas sequências, explorando codificadores de vídeo 2D já consolidados. O codificador utilizado por padrão é o *High Efficiency Video Coding* (HEVC) (SULLIVAN et al., 2012), reconhecido por sua alta eficiência, mas que impõe elevado custo computacional. No V-PCC, o processo ocorre em duas fases: inicialmente, a nuvem de pontos é projetada em três subfluxos de vídeos 2D que descrevem a geometria, os atributos e o mapa de ocupação; em seguida, cada um desses fluxos é comprimido individualmente pelo HEVC. Como a etapa de codificação de vídeo é repetida três vezes para cada nuvem dinâmica, ela responde por aproximadamente 90% do custo computacional do V-PCC, o que limita seu uso em sistemas de tempo real e dispositivos com restrições energéticas.

Neste trabalho, propõe-se uma abordagem baseada em aprendizado de máquina para reduzir esse custo computacional, com foco na etapa de compressão dos subfluxos de vídeo do V-PCC.

2. METODOLOGIA

Para reduzir o elevado custo computacional da compressão de nuvens de pontos dinâmicas, especialmente na etapa de codificação de vídeo, este trabalho propõe uma abordagem baseada em aprendizado de máquina, com o objetivo de acelerar o processo sem comprometer a qualidade dos dados.

A solução desenvolvida envolve a criação e integração de modelos de decisão que atuam em diferentes níveis do particionamento dos blocos de vídeo. As *Coding Tree Units* (CTUs), maiores unidades de codificação, servem como ponto inicial da divisão hierárquica. Cada CTU é subdividida em *Coding Units* (CUs), que por sua vez podem ser divididas em uma ou mais *Prediction Units* (PUs), responsáveis por definir os modos e formas de predição (intra ou inter). Esses modelos utilizam informações extraídas durante a codificação para antecipar decisões e reduzir o processamento desnecessário, acelerando os subfluxos de geometria e atributos do V-PCC.

A primeira abordagem foca nas Prediction Units (PUs), que representam diferentes formas de dividir e prever os blocos dentro das CTUs. Os modelos de decisão avaliam diferentes tamanhos de PU (64×64 , 32×32 , 16×16 e 8×8), verificando, após a análise dos modos principais de particionamento, se é necessário explorar os modos restantes ou se o melhor resultado já pode ser selecionado. Essa estratégia permite interromper de forma inteligente o processo de busca por novos particionamentos, reduzindo significativamente o tempo de codificação.

Em seguida, a segunda abordagem atua sobre as CTUs, utilizando modelos de árvore de decisão para determinar se o bloco deve ser subdividido em níveis mais profundos. Para cada nível de profundidade (64×64 , 32×32 e 16×16), o modelo avalia características do bloco e decide se a subdivisão pode ser encerrada antecipadamente, evitando operações desnecessárias e acelerando a codificação.

Ao combinar essas duas abordagens, o método proposto atua de forma complementar nos diferentes níveis do particionamento, potencializando a redução do custo computacional do V-PCC sem comprometer a qualidade dos dados comprimidos. A Figura 1 ilustra o funcionamento da solução. Inicialmente, para cada CU, são testados os modos principais de particionamento de PUs (MSM e $2N \times 2N$). Em seguida, o modelo de decisão das PUs verifica se é necessário explorar os modos restantes; caso positivo, todos os modos são testados; caso contrário, o melhor resultado entre os já avaliados é escolhido. Por fim, o modelo de CTU decide se o bloco deve ser subdividido em níveis mais profundos ou se a avaliação pode ser encerrada naquele estágio. Esse fluxo permite antecipar decisões, reduzir o tempo de processamento e abrir caminho para aplicações do V-PCC em tempo real e em dispositivos com recursos limitados.

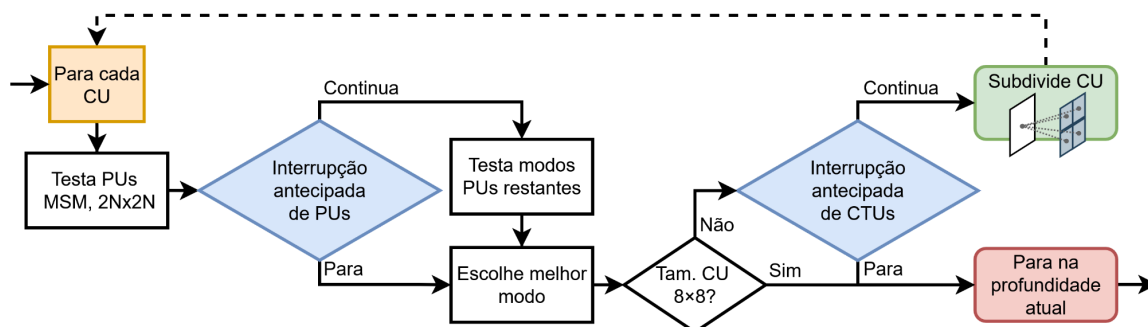


Figura 1 - Fluxograma do método proposto.

Para o treinamento dos modelos de decisão, foram utilizadas sequências de nuvens de pontos dinâmicas com precisão de 11 bits do conjunto de dados UVG-VPC (GAUTIER et al., 2023). Os dados foram extraídos da codificação dos subfluxos de vídeo de geometria e atributos no HEVC, considerando diferentes configurações de taxa de bits do V-PCC no modo *Random Access* (RA). Os modelos de aprendizado de máquina foram treinados separadamente para cada abordagem (PU e CTU), para cada subfluxo (geometria e atributos) e para cada tamanho de bloco do HEVC. O algoritmo de árvore de decisão foi empregado no treinamento, e uma busca de hiperparâmetros foi realizada para cada combinação de abordagem, subfluxo e tamanho de bloco, selecionando os modelos com melhor desempenho segundo as métricas de avaliação.

3. RESULTADOS E DISCUSSÃO

Para avaliar a proposta, os modelos desenvolvidos foram integrados ao software de referência do V-PCC. Os experimentos utilizaram as sequências de teste recomendadas pelas condições comuns de teste do V-PCC (MPEG, 2020) e as configurações temporais RA e All-Intra (AI).

Em cada experimento, foram coletadas métricas de qualidade das nuvens de pontos comprimidas, incluindo *Peak Signal-to-Noise Ratio* (PSNR) para geometria (D1 e D2) e para atributos de cor (Luma, Cb e Cr). O foco principal da avaliação, entretanto, foi o impacto dessas métricas na eficiência de codificação, medida pelo BD-Rate (BJONTEGAARD, 2001), que permite comparar diretamente a relação entre taxa de compressão e qualidade, mostrando se a solução proposta mantém o equilíbrio entre redução de tempo de codificação e preservação da qualidade dos dados.

No modo RA, a solução reduziu, em média, 60% do tempo total de codificação, com todas as sequências apresentando diminuição superior a 55% e algumas chegando a mais de 66%. O subfluxo de atributos apresentou reduções ainda maiores, atingindo 68%. Esses resultados demonstram que o método é eficiente em diferentes cenários e tipos de nuvem de pontos. Quanto à eficiência de codificação, o impacto foi pequeno: o aumento médio no BD-Rate foi de 1,3% para a geometria e 1,75% para o canal de luminância dos atributos, valores considerados aceitáveis diante da expressiva redução no tempo de processamento. Dessa forma, a solução mantém um bom equilíbrio entre velocidade e qualidade, sendo adequada para aplicações que exigem processamento rápido sem perda significativa da qualidade dos dados comprimidos.

Tabela 1 - Resultados do método proposto utilizando o modo RA.

Sequência	Geom.BD-GeomRate		Attr.BD-AttrRate			ΔTempo		
	D1	D2	Luma	Cb	Cr	ΔT Total	ΔT Geom	ΔT Atrib
<i>longdress</i>	0,90%	1,01%	2,65%	3,20%	3,57%	-55,24%	-54,43%	-56,86%
<i>loot</i>	1,03%	0,94%	1,83%	3,16%	2,94%	-61,85%	-56,82%	-68,84%
<i>queen</i>	1,36%	1,91%	1,86%	3,41%	2,35%	-66,63%	-66,88%	-66,75%
<i>redandblack</i>	0,77%	0,72%	2,05%	2,36%	1,71%	-56,50%	-53,92%	-60,01%
<i>soldier</i>	1,53%	2,00%	1,38%	0,67%	3,23%	-63,04%	-60,69%	-66,14%
<i>basketball</i>	1,39%	1,35%	1,12%	1,27%	5,22%	-59,37%	-56,56%	-62,90%
<i>dancer</i>	1,11%	1,21%	1,36%	1,14%	3,81%	-57,75%	-55,25%	-60,98%
Média	1,16%	1,31%	1,75%	2,17%	3,26%	-60,05%	-57,79%	-63,21%

Embora a solução tenha sido desenvolvida para o modo RA, também foi avaliada no modo AI do V-PCC, apresentando resultados competitivos. Devido à diferenças na estrutura dos quadros nesse modo, a redução média no tempo de codificação foi ligeiramente inferior à observada no RA, alcançando 42,7% no total, sendo 47% para o subfluxo de geometria e 38% para o de atributos. Ainda assim, a proposta manteve impacto mínimo na eficiência de codificação, com perdas praticamente nulas no BD-Rate. Esses resultados reforçam a robustez e a aplicabilidade da solução em diferentes cenários.

Tabela 2 - Resultados do método proposto utilizando o modo RA.

Sequência	Geom.BD-GeomRate		Attr.BD-AttrRate			Δ Tempo		
	D1	D2	Luma	Cb	Cr	Δ T Total	Δ T Geom	Δ T Atrib
<i>longdress</i>	-0,13%	0,04%	0,09%	0,15%	0,12%	-39,76%	-43,98%	-35,74%
<i>loot</i>	0,16%	0,12%	-0,01%	0,26%	-0,60%	-44,52%	-48,47%	-39,76%
<i>queen</i>	-0,29%	-0,23%	0,08%	-0,07%	0,52%	-46,03%	-53,02%	-37,84%
<i>redandblack</i>	-0,11%	-0,13%	0,04%	-0,07%	-0,03%	-41,66%	-44,87%	-38,02%
<i>soldier</i>	-0,06%	0,10%	-0,01%	-0,50%	0,06%	-41,70%	-47,56%	-35,12%
<i>basketball</i>	0,08%	0,14%	-0,01%	-0,85%	-0,35%	-42,68%	-45,16%	-39,74%
<i>dancer</i>	0,07%	0,28%	-0,04%	-0,12%	-0,06%	-42,74%	-45,85%	-38,91%
Média	-0,04%	0,05%	0,02%	-0,17%	-0,05%	-42,73%	-46,99%	-37,88%

4. CONCLUSÕES

Este trabalho investigou e propôs uma abordagem baseada em aprendizado de máquina para reduzir o custo computacional da compressão de nuvens de pontos dinâmicas utilizando o padrão V-PCC. Os resultados demonstram que a metodologia acelera significativamente a codificação, com impacto mínimo na eficiência de compressão e preservando a qualidade dos dados reconstruídos. A solução mostrou-se robusta e generalizável, sendo validada em diferentes condições de teste e configurações.

Como limitação, destaca-se o foco no modo de codificação RA. Trabalhos futuros podem estender a metodologia a outros modos, como AI, além de explorar técnicas de aprendizado de máquina mais avançadas para otimizações adicionais. A contribuição deste estudo viabiliza aplicações em tempo real e em dispositivos com restrições computacionais, ampliando as possibilidades de uso de nuvens de pontos dinâmicas em cenários práticos.

5. REFERÊNCIAS BIBLIOGRÁFICAS

- GRAZIOSI, D.; NAKAGAMI, O.; KUMA, S.; ZAGHETTO, A.; SUZUKI, T.; TABATABAI, A. *An overview of ongoing point cloud compression standardization activities: Video-based (V-PCC) and geometry-based (G-PCC)*. *APSIPA Transactions on Signal and Information Processing*, v. 9, e13, 2020.
- SULLIVAN, G. J.; OHM, J.-R.; HAN, W.-J.; WIEGAND, T. *Overview of the high efficiency video coding (HEVC) standard*. *IEEE Transactions on Circuits and Systems for Video Technology*, v. 22, n. 12, p. 1649–1668, 2012..
- GAUTIER, G.; MERCAT, A.; FRÉNEAU, L.; PITKÄNEN, M.; VANNE, J. *UVG-VPC: Voxelized Point Cloud Dataset for Visual Volumetric Video-based Coding*. In: *International Conference on Quality of Multimedia Experience (QoMEX)*, 15., 2023. Anais [...]. p. 244–247.
- MPEG. *Common Test Conditions for V3C and V-PCC*. ISO/IEC JTC 1/SC 29/WG 11, 2020.
- BJONTEGAARD, G. *Calculation of average PSNR differences between RD-curves*. ITU SG16 Doc. VCEG-M33, 2001.