

Redução de Custo Computacional da Codificação de Nuvens de Pontos Dinâmicas

Gustavo Rehbein

ghrehbein@inf.ufpel.edu.br
Universidade Federal de Pelotas
Pelotas, Brazil

Cristiano Santos

cfdsantos@inf.ufpel.edu.br
Universidade Federal de Pelotas
Pelotas, Brazil

ABSTRACT

This dissertation presents a machine learning-based solution to reduce the computational cost of compressing dynamic point clouds under the V-PCC standard. The proposed method employs decision models to accelerate the video encoding stage, which accounts for most of the processing. Results show an average 60% reduction in encoding time in Random Access mode, with minimal impact on coding efficiency (approximately 1.3% increase in BD-Rate), making the solution suitable for real-time applications and resource-constrained devices.

KEYWORDS

Compressão de Nuvens de Pontos, Aprendizado de Máquina, Redução de Custo Computacional

1 INTRODUÇÃO

Nuvens de pontos são um formato de mídia digital tridimensional, constituído por uma coleção de pontos em um espaço 3D, aos quais podem ser atribuídas informações como cor ou outras características. Esse formato é amplamente utilizado em diversas aplicações, como mapeamento de objetos históricos, vídeos 3D tele-imersivos, visualização em realidade virtual ou aumentada, navegação autônoma de veículos e imagens médicas. Nuvens de pontos dinâmicas, que representam uma sequência de nuvens de pontos ao longo do tempo, são análogas a vídeos 2D digitais e apresentam um grande volume de dados. O *Video-based Point Cloud Compression* (V-PCC) [3] é um padrão de compressão de nuvens de pontos dinâmicas criado pelo MPEG, que utiliza codificadores de vídeo 2D convencionais para comprimir os dados; por padrão, o codificador utilizado é o *High Efficiency Video Coding* (HEVC)[5], um padrão de compressão de vídeo com alta eficiência de codificação, porém que demanda recursos computacionais significativos. No V-PCC, o funcionamento ocorre em duas etapas principais: primeiro, as nuvens de pontos são convertidas em três subfluxos de vídeos 2D, que representam diferentes características da nuvem de pontos, sendo eles geometria, atributos e mapa de ocupação. Em seguida, cada um desses

In: VII Concurso de Teses e Dissertações (CTD 2025). Anais Estendidos do XXXI Simpósio Brasileiro de Sistemas Multimídia e Web (CTD'2025). Rio de Janeiro/RJ, Brasil. Porto Alegre: Brazilian Computer Society, 2025.
© 2025 SBC – Sociedade Brasileira de Computação.
ISSN 2596-1683

Guilherme Corrêa

gcorrea@inf.ufpel.edu.br
Universidade Federal de Pelotas
Pelotas, Brazil

Marcelo Porto

porto@inf.ufpel.edu.br
Universidade Federal de Pelotas
Pelotas, Brazil

subfluxos é codificado separadamente pelo HEVC. Esse processo faz com que a etapa de codificação de vídeo, repetida três vezes para cada nuvem de pontos dinâmica, represente cerca de 90% do custo computacional total, tornando o V-PCC especialmente desafiador para aplicações em tempo real e dispositivos com restrições de energia.

Esta dissertação apresenta uma solução de aprendizado de máquina para a redução do custo computacional do V-PCC, focada na etapa de codificação de vídeo dos subfluxos do V-PCC.

2 METODOLOGIA

Para enfrentar o desafio do alto custo computacional na compressão de nuvens de pontos dinâmicas, especialmente na etapa de codificação de vídeo, esta dissertação propõe uma solução baseada em aprendizado de máquina. O objetivo é tornar o processo mais rápido, sem comprometer a qualidade dos dados.

A abordagem desenvolvida consiste na criação e integração de modelos de decisão que atuam em diferentes níveis do particionamento dos blocos de vídeo: nas *Coding Tree Units* (CTUs), que são as unidades iniciais de codificação, e nas *Prediction Units* (PUs), que representam diferentes possibilidades de divisão e predição dentro das unidades de codificação. Esses modelos utilizam informações extraídas durante a codificação para antecipar decisões e evitar etapas desnecessárias, acelerando o processamento dos subfluxos de geometria e atributos do V-PCC.

A primeira abordagem utiliza modelos de árvore de decisão para auxiliar na escolha do particionamento das CTUs, que são as principais unidades de codificação do vídeo. Para cada nível de profundidade das CTUs (64×64 , 32×32 e 16×16), um modelo é responsável por analisar características do bloco e decidir se o processo de subdivisão pode ser interrompido antecipadamente, evitando etapas desnecessárias e acelerando a codificação. Já a segunda abordagem foca nas PUs, que representam diferentes formas de dividir e prever os blocos dentro das CTUs. Os modelos de decisão atuam sobre diferentes tamanhos de PU (64×64 , 32×32 , 16×16 e 8×8), avaliando, após os testes dos modos principais de particionamento, se é necessário continuar testando outros modos ou se já é possível escolher o melhor resultado entre os avaliados. Essa estratégia permite interromper o processo de busca por novos particionamentos de forma inteligente, reduzindo ainda mais o tempo de codificação.

Ao combinar essas duas abordagens, o método proposto atua de forma complementar nos diferentes níveis do particionamento,

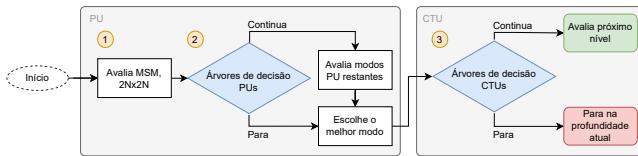


Figura 1: Fluxograma do método proposto.

potencializando a redução do custo computacional do V-PCC sem comprometer a qualidade dos dados comprimidos. A Figura 1 ilustra o funcionamento da solução proposta, destacando as principais etapas do processo. Inicialmente, em nível de PUs, são testados os modos principais de particionamento (MSM e $2N \times 2N$) (etapa 1). Em seguida, o modelo de decisão referente às PUs avalia se é necessário explorar os modos restantes de particionamento (etapa 2); caso positivo, todos os modos são testados; caso contrário, o melhor entre os já avaliados é escolhido. Por fim, o modelo de CTU decide se o bloco deve ser subdividido para níveis mais profundos ou se a avaliação de particionamento pode ser encerrada naquela etapa (etapa 3). Esse fluxo permite antecipar decisões e evitar etapas desnecessárias, reduzindo o tempo de processamento e abrindo caminho para futuras aplicações do V-PCC em cenários de tempo real e dispositivos com recursos limitados.

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 [2]. Os dados utilizados 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 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 utilizado para o treinamento dos modelos, e para cada combinação de abordagem, subfluxo e tamanho de bloco foi realizada uma busca de hiperparâmetros, selecionando os melhores modelos conforme o desempenho nas métricas de avaliação.

3 RESULTADOS EXPERIMENTAIS

Para avaliar a proposta, os modelos desenvolvidos foram integrados ao software de referência do V-PCC. Experimentos foram realizados utilizando as sequências de teste recomendadas pelas condições comuns de teste do V-PCC [4] e as configurações temporais Random Access (RA) e All-Intra (AI) do software de referência do V-PCC.

Para cada experimento, foram coletadas métricas de qualidade das nuvens de pontos comprimidas, como PSNR para geometria (usando D1 e D2) e para os atributos de cor (Luma, Cb e Cr). No entanto, o foco principal da avaliação foi o impacto dessas métricas na eficiência de codificação, avaliado pelo BD-Rate [1]. Essa métrica permite comparar diretamente a relação entre taxa de compressão e qualidade, mostrando se a solução proposta mantém o equilíbrio desejado entre redução de tempo de codificação e preservação da qualidade dos dados.

No modo RA, a solução proposta apresentou uma redução média de 60% no tempo total de codificação, com todas as sequências avaliadas registrando diminuição superior a 55%. Em algumas

sequências avaliadas, a redução chegou a mais de 66%. O subfluxo de atributos também se destacou, alcançando reduções de tempo ainda maiores, chegando a 68%. Esses resultados demonstram que o método é eficiente em diferentes cenários e tipos de nuvem de pontos. Em relação à eficiência de codificação, o impacto observado foi pequeno: o aumento médio no BD-Rate foi de apenas 1,3% para a geometria e 1,75% para o canal de luminância dos atributos. Esses valores são considerados aceitáveis, especialmente 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 de qualidade dos dados comprimidos.

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

Em comparação com trabalhos relacionados, a solução proposta apresentou a maior redução de tempo de codificação entre os métodos avaliados, especialmente no modo RA. Além disso, foi testada em todas as sequências e configurações recomendadas, demonstrando maior robustez e generalização. O impacto na eficiência de codificação permaneceu baixo, tornando o método competitivo para aplicações práticas.

4 CONCLUSÃO

Esta dissertação apresentou uma solução de aprendizado de máquina para reduzir o custo computacional da compressão de nuvens de pontos dinâmicas no padrão V-PCC. A proposta acelera o processamento com impacto mínimo na eficiência de codificação, mostrando-se robusta e generalizável. Como limitação, a otimização foi focada apenas no modo RA, e como perspectiva futura, sugere-se a adaptação para outros modos e a exploração de técnicas mais avançadas. A solução contribui para viabilizar aplicações em tempo real e em dispositivos com recursos limitados, ampliando o potencial de uso das nuvens de pontos dinâmicas.

REFERÊNCIAS

- [1] Gisle Bjontegaard. 2001. Calculation of average PSNR differences between RD-curves. *ITU SG16 Doc. VCEG-M33* (2001).
- [2] Guillaume Gautier, Alexandre Mercat, Louis Fréneau, Mikko Pitkänen, and Jarno Vanne. 2023. UVG-VPC: Voxelized Point Cloud Dataset for Visual Volumetric Video-based Coding. In *2023 15th International Conference on Quality of Multimedia Experience (QoMEX)*. 244–247.
- [3] Danillo Graziosi, Ohji Nakagami, Shinroku Kuma, Alexandre Zaghetto, Teruhiko Suzuki, and Ali Tabatabai. 2020. 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* 9 (2020), e13.
- [4] MPEG. 2020. Common Test Conditions for V3C and V-PCC. *ISO/IEC JTC 1/SC 29/WG 11* (2020).
- [5] Gary J Sullivan, Jens-Rainer Ohm, Woo-Jin Han, and Thomas Wiegand. 2012. Overview of the high efficiency video coding (HEVC) standard. *IEEE Transactions on circuits and systems for video technology* 22, 12 (2012), 1649–1668.