

TÉCNICA DE PRÉ-SELEÇÃO DE MODOS PARA A ACELERAÇÃO DA PREDIÇÃO INTRA-QUADRO DE VÍDEOS OMNIDIRECIONAIS

IAGO STORCH, LUCIANO AGOSTINI, BRUNO ZATT, DANIEL PALOMINO

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

1. INTRODUÇÃO

Vídeos digitais estão muito comuns no dia a dia da população, e o advento de novas tecnologias tem feito a indústria investir em abordagens imersivas, como vídeos omnidirecionais, para aumentar a experiência do usuário.

Vídeos omnidirecionais são vídeos que representam todo o arredor do usuário, na forma de uma superfície esférica, na qual o usuário se sente inserido. Como estes vídeos representam todo o arredor do usuário, este pode olhar ao redor da cena durante a reprodução. Embora vídeos omnidirecionais representem uma superfície esférica, toda a infraestrutura de distribuição de conteúdo atual foi desenvolvida para lidar com vídeos convencionais, que representam uma superfície retangular. Sendo assim, para explorar a infraestrutura já existente, vídeos omnidirecionais são mapeados para uma superfície retangular através de projeções. Existem diversas projeções capazes de realizar essa tarefa, mas a projeção equirretangular (ERP) é a mais utilizada. A Figura 1 apresenta um vídeo na projeção ERP e suas divisões, explicadas adiante neste trabalho.

Quando vídeos digitais estão em seu formato original, estes demandam um volume excessivo de dados que torna sua transmissão/reprodução impraticáveis, e nesse contexto, padrões de codificação de vídeo são empregados para reduzir tal volume de dados. Como vídeos omnidirecionais apresentam resoluções significativamente maiores do que vídeos convencionais devido ao maior campo de visão, vídeos omnidirecionais também precisam ser codificados.

Os padrões de codificação de vídeo atuais foram desenvolvidos para lidar com vídeos convencionais, e embora vídeos omnidirecionais sejam representados através de projeções retangulares, estes apresentam características diferentes de vídeos convencionais. No caso da projeção ERP, o esticamento na região superior/inferior e a continuidade entre as laterais do vídeo são exemplos disso.

Visto que o processo de codificação de vídeo gera uma alta carga computacional, técnicas capazes de reduzir tal carga são alvos de investigação constante. Além disso, como vídeos omnidirecionais têm características espaciais distintas, é possível explorar estas características para desenvolver técnicas especializadas para a codificação de vídeos omnidirecionais.

Trabalhos como (WANG, 2017) reduzem o tempo de codificação de vídeos omnidirecionais. Entretanto, a maior parte dos ganhos obtidos se deve à exploração de características genéricas de vídeos digitais, aplicáveis a vídeos de qualquer natureza. Sendo assim, este trabalho propõe uma técnica de redução de complexidade que explora as características próprias de vídeos omnidirecionais.

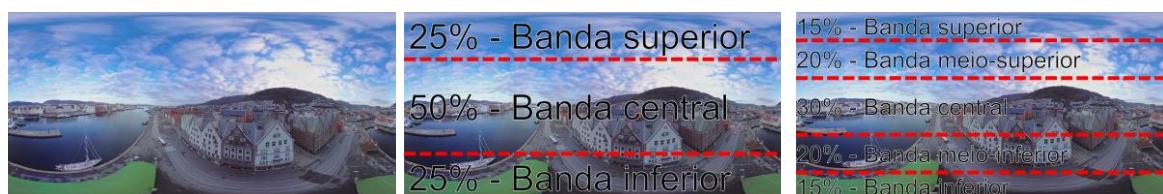


Figura 1. Quadro na projeção ERP e respectivas divisões em três e cinco bandas

2. METODOLOGIA

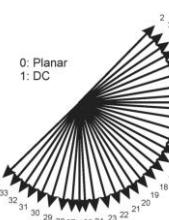
Como a principal peculiaridade de vídeos na projeção ERP está na distorção espacial, que se manifesta pelo esticamento horizontal, é possível que a predição intra-quadro possa ser otimizada para obter ganhos de velocidade durante a codificação destes vídeos. A predição intra-quadro busca explorar redundâncias espaciais no vídeo, isto é, redundâncias dentro do mesmo quadro, e para isso ela representa algumas regiões do quadro com base em regiões já codificadas.

O padrão *High Efficiency Video Coding* (HEVC) é o estado da arte em codificação de vídeo, logo, este foi o padrão adotado neste trabalho. A predição intra do HEVC é composta por 35 modos de predição, apresentados na Figura 2 (a). Os modos direcionais (2-34) representam um bloco a partir de blocos nas posições referentes as setas, enquanto que os modos não-direcionais (0-1) são modos especiais. A predição intra é realizada em blocos denominados *Prediction Units* (PUs) com dimensões desde 4×4 até 64×64 amostras (ITU-T, 2019).

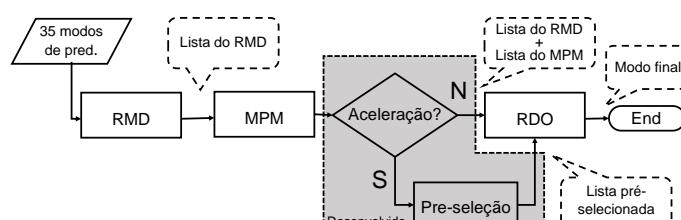
Para evitar avaliar todos os modos de predição com todos os tamanhos de PU numa análise exaustiva, o *HEVC Test Model* (HM), software de referência do padrão HEVC, implementa a predição intra numa abordagem de três etapas, como apresentado na Figura 2 (b), e considerando o caminho “N” no teste “Aceleração?”. Primeiro os 35 modos são avaliados de forma superficial (RMD) para definir os candidatos a melhor modo, e ou 3 modos (no caso de PUs 64×64 , 32×32 e 16×16) ou 8 modos (no caso de PUs 8×8 e 4×4) são selecionados. Em seguida, o MPM deriva os modos mais prováveis com base no modo de predição de PUs vizinhas, e até 3 modos são adicionados à lista do RMD. Por fim, essa lista é enviada ao RDO, que é a etapa mais custosa e realiza uma análise aprofundada para definir qual modo apresenta a melhor eficiência de codificação.

Visando entender o impacto das distorções da projeção ERP na predição intra, foi feita uma avaliação da taxa de ocorrência de cada modo de predição por tamanho de PU e região do vídeo. Foram consideradas divisões em três e em cinco regiões, como apresentado na Figura 1. As bandas superior e inferior são chamadas polares, enquanto que as meio-superior e meio-inferior são chamadas meio-polares. Essa abordagem busca dividir o vídeo em diferentes regiões de distorção: enquanto que nas bandas polares a distorção é excessiva, na banda central ela é mínima, e nas bandas meio-polares a distorção é moderada.

Os resultados da taxa de ocorrência de modos de predição intra por região do vídeo e tamanho de PU são apresentados na Figura 3 para a divisão em três bandas. Ao analisar estes dados, fica visível que na banda central os modos não-direcionais são muito frequentes independente do tamanho de PU, e os modos direcionais apresentam taxas de ocorrência muito semelhantes, com exceção dos modos 10 e 26 que se destacam ligeiramente. Nas bandas polares, os modos não-direcionais seguem apresentando altas taxas de ocorrência, já os modos direcionais apresentam um comportamento diferente: existe uma concentração de ocorrências em torno do modo horizontal (10), e essa concentração é reduzida conforme os modos se afastam do modo 10. Esse comportamento mostra que



(a) Modos de predição



(b) Fluxo de processamento da predição intra

Figura 2. Modos de predição intra e fluxo de processamento no HM

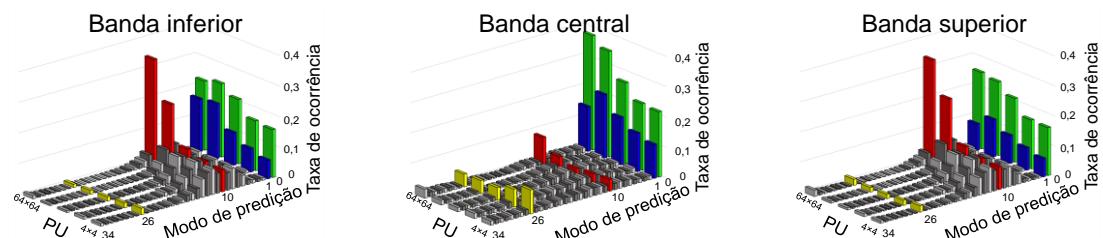


Figura 3. Taxa de ocorrência de modos de predição por região e tamanho de PU

nas regiões polares, onde há mais distorção, os modos de predição orientados na horizontal tendem a representar a distorção horizontal de forma mais eficiente.

Os gráficos da avaliação com cinco bandas não são apresentados neste trabalho, mas estes apontam que as bandas polares, por serem menores do que na divisão em três bandas, concentram PUs mais distorcidas e tem ocorrências de modos horizontais maiores do que na divisão em três bandas. Já nas bandas meio-polares, onde estão as PUs com distorção moderada, a ocorrência de modos horizontais é maior do que na banda central, mas menor do que nas bandas polares. Por fim, estas mesmas avaliações realizadas em vídeos convencionais demonstram que todo o quadro se comporta de forma similar à banda central de vídeos omnidirecionais, mostrando que a banda central de vídeos omnidirecionais tem um comportamento de vídeos convencionais.

Tendo em vista este comportamento, foi desenvolvida uma técnica de pré-seleção baseada em pontuações para eliminar a avaliação de modos de predição menos prováveis na etapa de RDO. Para determinar quais modos devem ou não ser enviados ao RDO, três parâmetros são considerados: taxa de ocorrência dos modos intra, decisões do RMD e decisões do MPM.

A técnica desenvolvida é executada após o MPM, e é caracterizada pela região acinzentada da Figura 2 (b), considerando o caminho “**S**” no teste “**Aceleração?**”. Primeiro os 35 modos de predição intra recebem uma pontuação fixa, de acordo com a taxa de ocorrência de cada modo. Essa pontuação está relacionada à altura das barras na Figura 3. Em seguida, os modos selecionados pelo RMD recebem pontuações de acordo com sua probabilidade de serem selecionados pelo RDO, isto é, o modo com melhor avaliação durante o RMD ganha uma pontuação, o segundo melhor modo ganha uma pontuação menor, e assim por diante para 3 ou 8 modos de acordo com o tamanho da PU. Em seguida, os 3 modos derivados pelo MPM também recebem uma pontuação. Sendo assim, cada modo de predição pontua em uma, duas, ou todas as etapas. Por fim, os modos de predição são ordenados de acordo com o somatório das pontuações obtidas, e os modos com melhor pontuação são enviados ao RDO.

Para reduzir complexidade, esta técnica reduz o número de modos enviados ao RDO de acordo com a Tabela 1. Na abordagem com 3 bandas, há uma redução moderada no número de modos para as bandas polares. Já ao utilizar 5 bandas, há uma redução significativa para as bandas polares, enquanto que para as bandas meio-polares há uma redução mais amena. Na banda central, onde há pouca distorção, a codificação é feita de acordo com o HM sem interferências.

Tabela 1. Número de modos enviado ao RDO para diferentes bandas

PU	Original Todo quadro	3 Bandas		5 Bandas	
		Polar	Meio-polar	Polar	Meio-polar
64x64	3 – 6	1		1	1
32x32	3 – 6	1		1	2
16x16	3 – 6	2		1	3
8x8	8 – 11	3		2	3
4x4	8 – 11	3		2	4

3. RESULTADOS E DISCUSSÃO

A técnica desenvolvida foi implementada dentro do HM versão 16.16 em conjunto com o pacote 360Lib-5.0, que possibilita codificar vídeos omnidirecionais no HM. Em seguida, 9 vídeos foram codificados de acordo com as Condições Comuns de Teste para a Codificação de Vídeos Omnidirecionais (BOYCE, 2017) para avaliar o desempenho da técnica desenvolvida em relação ao HM original.

Para medir a redução de complexidade, foi avaliado o tempo de processamento do codificador com a técnica desenvolvida em relação ao codificador sem modificações. Já para avaliar a eficiência de codificação, BD-BR foi utilizado. O BD-BR mede o acréscimo de *bitrate* para que uma técnica produza uma codificação com a mesma qualidade visual de uma referência, sendo assim, valores positivos de BD-BR indicam que uma técnica é menos eficiente em termos de compressão do que a referência. Os resultados de redução de tempo e BD-BR para a técnica desenvolvida são apresentados na Tabela 2 para a abordagem com três e cinco bandas, onde a referência é a codificação convencional do HM. Os resultados representam a média de todos os vídeos.

Ao avaliar os resultados da Tabela 2, fica claro que a técnica desenvolvida é capaz de atingir uma redução de tempo significativa com uma pequena redução na eficiência de codificação. Para a abordagem com três bandas, a técnica desenvolvida foi capaz de reduzir o tempo de codificação em 11,2%, em média, com um BD-BR médio de 0,3%. Já para a abordagem com cinco bandas, a técnica foi capaz de reduzir o tempo de processamento em 15,2%, em média, com um BD-BR de 0,4%. Estes resultados apontam que a utilização de mais bandas torna possível modelar de forma mais adequada a distorção dentro do quadro, desde que a regulagem destas seja feita corretamente.

Tabela 2. Desempenho da técnica desenvolvida

3 Bandas		5 Bandas	
Redução de Tempo	BD-BR	Redução de Tempo	BD-BR
11,2%	0,3%	15,2%	0,4%

4. CONCLUSÕES

Este trabalho apresentou uma técnica de redução de complexidade para a codificação de vídeos omnidirecionais. Esta técnica explora as distorções espaciais causadas pela projeção ERP para reduzir o número de modos de predição avaliados durante a predição intra-quadro destes vídeos. Resultados experimentais demonstram que a técnica desenvolvida é capaz de reduzir consideravelmente o tempo de codificação de vídeos omnidirecionais causando um pequeno impacto na eficiência de codificação, logo, ela pode ser uma solução para sistemas que precisam prover codificação em tempo real.

5. REFERÊNCIAS BIBLIOGRÁFICAS

- WANG, Y; et al. A fast intra prediction algorithm for 360-degree equirectangular panoramic video. In: 2017 IEEE VISUAL COMMUNICATIONS AND IMAGE PROCESSING (VCIP), St. Petersburg, 2017, p. 1-4.
- ITU-T, High Efficiency Video Coding, 2019.
- BOYCE, J; et al. JVET Common Test Conditions and Evaluation Procedures for 360° Video. Documento: JVET-G1030, Torino, 2017.