

UMA AVALIAÇÃO DO COMPORTAMENTO DA PREDIÇÃO INTRA-QUADRO DO PADRÃO HEVC QUANDO APLICADA A VÍDEOS 360°

IAGO STORCH¹; LUCIANO AGOSTINI¹; BRUNO ZATT¹; DANIEL PALOMINO¹; LUIS CRUZ²

¹Universidade Federal de Pelotas – Video Technology Research Group (ViTech)

²Universidade de Coimbra – Instituto de Telecomunicações

{icstorch; agostini; zatt; dpalomino}@inf.ufpel.edu.br; lcruz@deec.uc.pt

1. INTRODUÇÃO

Vídeos digitais estão cada vez mais comuns no cotidiano da população, e com o objetivo de melhorar a experiência dos usuários, a indústria tem investido em abordagens interativas, como os vídeos 360°, também chamados de vídeos esféricos. A captura destes vídeos é realizada utilizando-se câmeras com múltiplas lentes em orientações distintas, e em seguida os vídeos provenientes destas múltiplas lentes são costurados para formar um vídeo esférico.

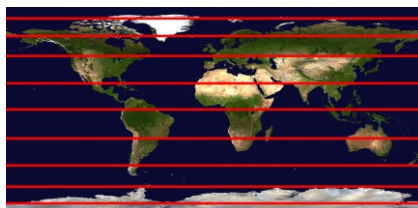
Como os vídeos 360° representam toda a cena ao redor da câmera, os vídeos 360° demandam uma quantidade de dados maior do que vídeos convencionais para serem representados com a mesma qualidade, sendo assim, eles também precisam ser comprimidos.

A codificação de vídeos convencionais é um assunto que já é estudado há muitos anos e atualmente existe uma infraestrutura robusta para a codificação dos mesmos. Sendo assim, ao invés de criar um novo padrão de codificação de vídeo específico para vídeos 360°, estes são pré-processados de modo a serem representados numa forma planar, e em seguida, são codificados pelos padrões convencionais com o auxílio de bibliotecas auxiliares, como a 360Lib (YE, 2017).

O processo responsável por transformar os vídeos esféricos em planares é chamado de projeção. Existem vários métodos para realizar a projeção, sendo que a projeção mais comumente usada é a equirretangular (ERP) (YE, 2017). Na projeção ERP, cada paralelo da esfera é transformado numa linha horizontal do retângulo. A Figura 1 (a) e (b) representa uma face da Terra em sua forma esférica e na projeção ERP, respectivamente, onde as linhas vermelhas representam alguns paralelos hipotéticos. É importante ressaltar que embora a Figura 1 (a) represente apenas uma face da esfera, a projeção da Figura 1 (b) representa a superfície completa da esfera. Ao comparar estas figuras, é visível que o espaçamento entre os paralelos não é alterado da esfera para o retângulo, sendo assim, a projeção ERP não causa distorção vertical. No entanto, como a largura da esfera é reduzida conforme aproxima-se dos polos, os paralelos próximos ao equador têm uma circunferência maior do que os paralelos próximos aos polos, e com isso, os paralelos das regiões polares precisam ser esticados para preencherem toda a largura do retângulo, criando uma distorção horizontal. A Figura 1 (c) mostra o primeiro quadro do vídeo *AerialCity* na projeção ERP.



(a) Terra esférica



(b) Terra na projeção ERP



(c) Quadro 1 de *AerialCity*

Figura 1. Funcionamento e exemplo de projeção ERP

Ao analisar a Figura 1 (c) é possível notar o esticamento causado nas regiões polares, característico da projeção ERP. Esse esticamento causa um aumento na redundância espacial, visto que são criadas novas amostras a partir de amostras vizinhas. Esse aumento na redundância espacial pode alterar o comportamento da predição intra-quadro, visto que esta é responsável pela exploração das redundâncias espaciais no vídeo. Sendo assim, o objetivo deste trabalho é estudar o efeito causado pela projeção ERP na predição intra-quadro do padrão de codificação de vídeo *High Efficiency Video Coding* (HEVC) (SULLIVAN, 2012), com o auxílio da biblioteca 360Lib (YE, 2017).

2. METODOLOGIA

A predição intra-quadro do padrão HEVC é composta por 35 modos de predição, sendo 33 modos angulares responsáveis por explorar texturas direcionais e 2 modos não-angulares responsáveis por explorar texturas não direcionais. A Figura 2 apresenta os modos de predição intra do padrão HEVC, onde as setas indicam as direções exploradas pelos modos angulares. Caso uma região do quadro possua uma textura orientada na horizontal, por exemplo, é provável que esta região seja codificada com o modo 10 (modo horizontal) ou seus vizinhos. A predição é realizada em blocos chamados *Prediction Units* (PUs), que delimitam regiões quadradas de dimensões que podem variar desde 4×4 até 64×64 pixels. Durante a predição, todos os 35 modos são avaliados para uma PU 64×64, em seguida essa PU é subdividida em 4 PUs 32×32 e todos os modos são avaliados para cada PU 32×32, e esse processo é repetido até que todos os modos tenham sido avaliados para as PUs de tamanho 4×4. Por fim, a combinação de tamanhos de PU e modos de predição que resulta na melhor relação taxa-distorção é selecionada para a codificação final (SULLIVAN, 2012).

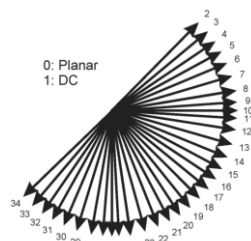


Figura 2. Modos da predição intra-quadro do padrão HEVC

Para avaliar o impacto da projeção ERP na predição intra, um conjunto de 4 vídeos foi codificado e o modo de predição escolhido para cada PU foi extraído. No entanto, como a maioria dos vídeos 360° possui conteúdos semelhantes na região superior e inferior, como o céu e o chão, é possível que o resultado das avaliações seja tendencioso ao conteúdo dos vídeos e não avalie adequadamente a distorção causada pela projeção ERP. Sendo assim, os vídeos foram rotacionados desde 0° até 330°, em múltiplos de 30°, nos eixos X, Y e Z antes de realizar a projeção dos mesmos. Com isso, todos os conteúdos do vídeo foram codificados em diversas regiões do quadro projetado, eliminando o efeito de conteúdos semelhantes em regiões co-localizadas dos vídeos. A Figura 3 mostra



(a) Original



(b) Rotacionado 60° em Y



(c) Rotacionado 60° em Z

Figura 3. Frame 1 de *AerialCity* na posição original e rotacionado nos eixos Y e Z

o primeiro quadro do vídeo *AerialCity*, onde (a) é o vídeo na posição original, (b) é o vídeo rotacionado 60° no eixo Y, e (c) é o vídeo rotacionado 60° no eixo Z.

3. RESULTADOS E DISCUSSÃO

Após a extração dos modos para cada PU, foi realizada uma avaliação espacial da predição intra com o objetivo de analisar o comportamento da mesma ao longo do quadro. Para isso, foi calculado o modo de predição médio selecionado em cada pixel do vídeo. Como os modos não-angulares não exploram texturas com orientações específicas, eles foram desconsiderados durante essa avaliação. A Figura 4 mostra os resultados desta avaliação, onde o azul mais escuro corresponde ao modo 2 e o vermelho mais escuro ao modo 34.

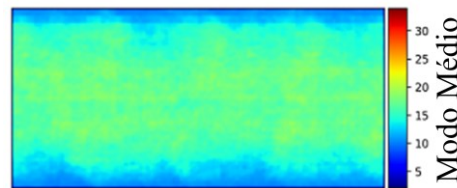


Figura 4. Modo de predição médio por pixel do vídeo

Ao analisar a Figura 4, é possível perceber que a região central possui um tom azul/amarelo, ou seja, a média dos modos selecionados nessa região está entre 15 e 20. Como este valor está próximo do centro da escala, isso pode ser evidência de que na região central todos os modos tendem a ocorrer com a mesma frequência. Já ao analisar as regiões superior e inferior, é visível que o modo médio é representado por um tom azul mais escuro, próximo do modo 10. Como este modo está afastado do centro da escala, é possível que nessas regiões haja uma tendência à ocorrência de alguns modos ao invés de outros.

Com o objetivo de analisar o comportamento dos vídeos nestas diferentes regiões de forma mais precisa, os vídeos foram divididos em três regiões distintas: a *banda superior* corresponde à 25% dos pixels mais ao topo do quadro, a *banda central* corresponde à 50% dos pixels no centro do quadro, e a *banda inferior* corresponde à 25% dos pixels mais abaixo no quadro. As bandas superior e inferior também podem ser designadas como bandas polares. Com essa divisão, foi avaliada a taxa de ocorrência para todos os 35 modos intra de acordo com a banda do quadro e o tamanho da PU. Os resultados dessa avaliação para as bandas inferior, central e superior são apresentados na Figura 5 (a), (b) e (c), respectivamente. Para facilitar a visualização, os modos não-angulares foram representados com as cores azul e vermelho, o modo horizontal (10) foi representado com verde e o modo vertical (26) foi representado com roxo.

Ao analisar os resultados da Figura 5, fica visível que os modos não-angulares possuem uma alta taxa de ocorrência para todas as regiões do vídeo,

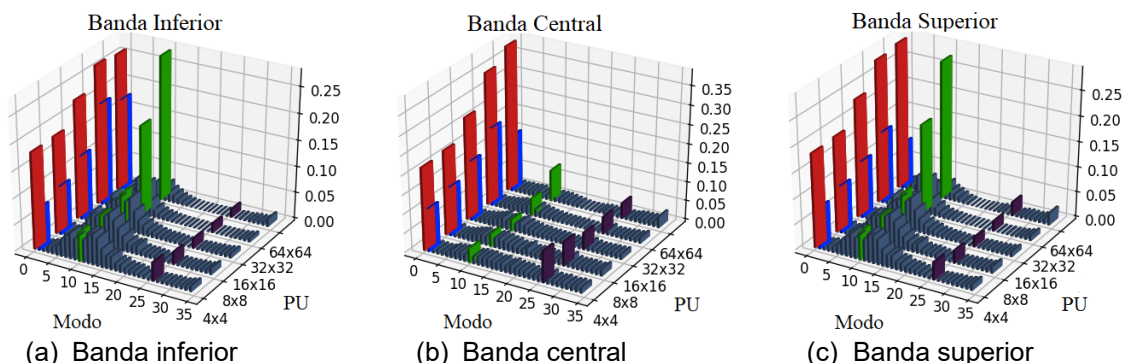


Figura 5. Taxa de ocorrência de modo intra por banda e tamanho de PU

no entanto, ao analisar os modos angulares é possível ver que estes possuem um comportamento diferente entre as regiões. Para a banda central, o modo horizontal (10) e vertical (26) se destacam entre os modos angulares, mas os demais modos apresentam uma taxa de ocorrência muito semelhante entre si. Já para as bandas polares, fica claro que há uma tendência em torno dos modos horizontais. Considerando apenas os modos angulares, o modo 10 costuma apresentar a maior taxa de ocorrência, seus modos vizinhos também apresentam altas taxas de ocorrência, e as taxas de ocorrência diminuem conforme os modos se afastam do modo 10. Observando os diferentes tamanhos de PU para as bandas polares, também é possível notar que a taxa de ocorrência para o modo 10 aumenta conforme o tamanho da PU aumenta, e chega a ultrapassar os modos não-angulares para as PUs maiores. Isso ocorre porque nas regiões mais próximas dos polos o vídeo sofre um esticamento ainda mais agressivo, o que gera maior redundância espacial e consequentemente mais PUs tendem a ser codificadas com tamanhos maiores (KHAN, 2013). Além disso, como esta redundância é criada devido ao esticamento horizontal gerado pela projeção ERP, a redundância nestas regiões costuma ser orientada na horizontal, e com isso, as PUs acabam sendo codificadas com o modo 10 (horizontal). Por fim, os modos verticais raramente ocorrem nas bandas polares, com exceção do modo 26.

Como a projeção ERP causa um esticamento horizontal nas regiões polares, essas regiões tendem a apresentar uma alta redundância espacial no sentido horizontal. Já nas regiões próximas ao equador, os vídeos apresentam nenhuma ou pouca distorção. Sendo assim, como a predição intra busca explorar redundâncias espaciais, as regiões polares (que apresentam maior esticamento horizontal) tendem a ser codificadas utilizando modos horizontais, próximos ao modo 10. Já na região central, como a distorção não é tão significativa e os conteúdos apresentam todo o tipo de textura, não há uma tendência para alguns modos em detrimento de outros.

4. CONCLUSÕES

Os resultados apresentados neste trabalho mostram que a codificação intra-quadro de vídeos 360° na projeção ERP apresenta um comportamento característico. Mais precisamente, é possível dizer que as regiões do vídeo mais próximas das extremidades superior e inferior do quadro costumam ser codificadas utilizando-se, na maior parte das vezes, os modos não-angulares e os modos angulares próximos ao modo horizontal. Sendo assim, durante a codificação de vídeos 360° na projeção ERP é possível avaliar apenas um subconjunto dos modos de predição de acordo com a região do vídeo, e com isso, atingir uma redução de complexidade e de consumo energético, com um pequeno impacto na eficiência de codificação

5. REFERÊNCIAS BIBLIOGRÁFICAS

- YE, Y. et al. "Algorithm descriptions of projection format conversion and video quality metrics in 360Lib", **JVET-E1003**, Geneva, 2017.
- SULLIVAN, G. J. et al. 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.
- KHAN, M. et al. An adaptive complexity reduction scheme with fast prediction unit decision for HEVC intra encoding. In: **IEEE International Conference on Image Processing**, Melbourne, 2013, p. 1578-1582.