

PADRÃO DE BUSCA PARA ESTIMAÇÃO DE MOVIMENTO RÁPIDA NO CODIFICADOR DE VÍDEO HEVC

PAULO GONÇALVES; MARCELO PORTO; BRUNO ZATT;
LUCIANO AGOSTINI; GUILHERME CORREA

Universidade Federal de Pelotas – Video Technology Research Group (ViTech)
{phrgoncalves, porto, zatt, agostini, gcorrea}@inf.ufpel.edu.br

1. INTRODUÇÃO

Frente a crescente demanda por avanços nas tecnologias de vídeos digitais, principalmente devido as diversas aplicações de vídeos de alta resolução em dispositivos portáteis, como smartphones e tablets, e serviços de streaming, que tornaram o consumo por esta mídia uma parte rotineira no cotidiano de grande parte das pessoas, é inviável transmitir e armazenar vídeos digitais quando esses estão em seu formato original, e nesse contexto o uso de compressores de vídeos digitais torna-se praticamente imprescindível.

A compressão de vídeo é realizada explorando redundâncias temporais e espaciais, podendo assim atingir altas taxas de compressão (SULLIVAN, 2012). Atualmente o padrão de codificação *High Efficiency Video Coding* (HEVC) (JCT-VC, 2013), é considerado o estado-da-arte para vídeos 2D, e apresenta uma eficiência de compressão 40-50% maior que seu antecessor, o padrão H.264/AVC. Em contrapartida, esse aumento de eficiência veio ao custo de um aumento significativo de complexidade, que dependendo da configuração utilizada pode ser de 500% (CORRÊA, 2012).

Uma das etapas presentes nos codificadores de vídeo é a Estimação de Movimento (ME, do inglês *Motion Estimation*), a qual consiste em reduzir as redundâncias temporais, ou seja, diminuir a quantidade de dados repetidos nos diferentes quadros de um vídeo. Esta é uma das etapas que mais demanda tempo em um codificador de vídeo, e no HEVC, um dos motivos para este aumento de complexidade é a utilização de 24 diferentes tamanhos de blocos na realização da ME.

O processo de particionamento e subparticionamento das unidades de codificação segue uma hierarquia, onde cada quadro de um vídeo é dividido em *Coding Tree Units* (CTUs), que são unidades de dimensão 64×64 (pixels). Estas, por sua vez, são divididas em *Coding Units* (CUs), cujos tamanhos podem variar entre 64×64, 32×32, 16×16 e 8×8. Estas CUs também são divididas em *Prediction Units* (PUs), que diferente das outras partições, podem apresentar formatos assimétricos. A etapa de ME é responsável por buscar o bloco mais similar ao bloco que está sendo codificado no quadro atual em um quadro de referência. Essa similaridade pode ser calculada através de diferentes métricas, sendo a Soma das Diferenças Absolutas (SAD) mais comumente utilizada.

Existem diversos algoritmos para a realização da etapa da ME, dentre eles o algoritmo *Test Zone Search* (TZS), o qual está implementado no software de referência do HEVC, o *HEVC Model* (HM). O algoritmo do TZS consiste em, basicamente, quatro etapas: A Predição, que busca apontar a região mais próxima ao possível melhor resultado; a etapa de Busca Inicial, que consiste em uma busca em expansão na região apontada pela etapa anterior até atingir o ponto de parada. Neste momento são comparadas as distâncias do melhor bloco candidato até o momento e do bloco encontrado após a etapa de Predição, se

essa distância for maior que cinco então executa-se a etapa de Busca Raster (que consiste em uma busca semelhante à busca exaustiva, mas de forma subamostrada), e a seguir, o Refinamento. Caso a distância for menor que cinco, então executa-se somente a etapa de Refinamento, a qual é semelhante a Busca Inicial, também realizando uma busca em expansão, porém a faz de forma iterativa, ou seja, a cada iteração, se um novo bloco é tido como melhor candidato, então na próxima iteração, este bloco é escolhido como ponto inicial.

Tendo conhecimento do funcionamento e das características do algoritmo do TZS, este trabalho propõe um novo padrão de busca octogonal para a etapa de Busca Raster do TZS.

2. METODOLOGIA

Uma análise de custo computacional realizada por (GONÇALVES, 2017), mostrou que a Busca Raster é de longe a etapa que mais demanda tempo no algoritmo TZS, alcançando em média 81% do tempo total de execução. Este trabalho aprofundou esta análise para todos os diferentes tamanhos de PUs, como mostrado na Figura 1. Nesta análise, foi utilizado o software de referência do HEVC, o codificador HM na versão 16.14. Ao todo, cinco vídeos foram analisados: *NebutaFestival*, *Traffic*, *Cactus*, *BQTerrace* e *ParkScene*, todos com resolução de alta definição. Também foram utilizados valores de Parâmetro de Quantização (QP): 22, 27, 32, 37.

Os resultados na Figura 1 mostram que, para todos os tamanhos de PUs, a etapa de Busca Raster é a que demanda maior esforço computacional. Para o maior tamanho de PU (64x64), a etapa de Busca Raster atingiu em média 90% do tempo de execução total do TZS. Entretanto, este esforço computacional diminui para blocos menores, mas mesmo em PUs de tamanho 8x4 e 4x8, a etapa de Busca Raster corresponde a mais de 55% do tempo do TZS. Além disso, outra análise de (GONÇALVES, 2017) também revelou que, apesar de sua alta complexidade, a Busca Raster é responsável por encontrar apenas 0,4% das melhores correspondências de blocos no TZS. No entanto, simplesmente eliminar esta etapa do algoritmo resultaria em uma perda de eficiência proibitiva. Desta forma, foi realizada uma análise do comportamento da Busca Raster, visando identificar a distribuição das melhores correspondências de blocos retornadas a partir desta etapa.

Além dos vídeos listados anteriormente, para esta análise foram adicionados quatro vídeos de ultra-alta definição (UHD, do inglês *Ultra High Definition*), são

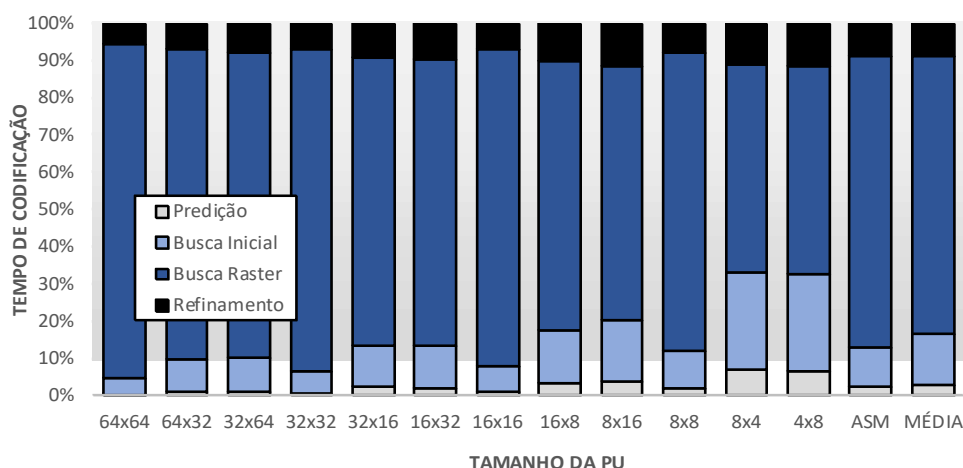


Figura 1: Distribuição do tempo de processamento do TZS em suas quatro etapas para diferentes tamanhos de PUs

eles: *Beauty*, *HoneyBee*, *Jockey*, *YachtRide*. Os 100 primeiros quadros foram codificados e as especificações foram os mesmos citados anteriormente.

A análise da distribuição das melhores correspondências de blocos consiste em coletar, ao longo de toda a execução do TZS, a posição da melhor correspondência dentro da área de busca. O tamanho da área de busca define o número de comparações que será realizado na etapa de Busca Raster, por exemplo, em uma área de busca de tamanho 512×512 posições, considerando uma subamostragem de 5 posições, como implementado no TZS, aproximadamente 10.500 posições são testadas durante uma execução desta etapa.

Os resultados foram traçados em mapas de calor, e são apresentados na Figura 2, onde as cores mais quentes indicam a área de maior ocorrência de melhor correspondência de bloco e as cores mais frias apresentam a área que foi raramente ou nunca escolhida. A Figura 2(a)-(c) apresenta o mapa de calor para três sequências de vídeo (*BQTerrace*, *YachtRide*, e *Cactus*, respectivamente), e a Figura 2(d) mostra a média obtida para todos os vídeos analisados. *BQTerrace* e *YachtRide* são os casos que apresentam a distribuição mais incomum entre todos os vídeos analisados. Em *BQTerrace*, uma grande concentração de correspondência de blocos é encontrada no eixo vertical central, isto acontece devido as características do vídeo, que apresenta um movimento constante na direção cima-direita. Em *YachtRide*, a distribuição é mais uniforme, contudo uma grande concentração de correspondência de blocos é encontrada acima do eixo horizontal. Novamente, esta concentração deve se as características do vídeo, que é composto por um barco em movimento da esquerda para a direita, com oscilações para cima. Em *Cactus*, podemos observar um padrão mais recorrente na maioria dos vídeos restantes, com uma clara concentração em torno do ponto central, muito parecida com a média encontrada na Figura 2(d).

Esta análise mostrou que o esforço computacional para calcular grande parte dos pontos dentro da Busca Raster é desperdiçado, sendo que a maioria dos pontos comparados fora do centro e dos eixos da área de busca dificilmente são escolhidos como melhores, sendo assim, é perceptível que esses pontos podem ser removidos, com baixas perdas em eficiência de codificação.

Para explorar as características observadas na análise anterior de forma eficiente, este trabalho propõe um novo padrão de busca octogonal, que prioriza a área central e os eixos na área de busca, conforme ilustrado na Figura 2(e). Este padrão foi desenvolvido para explorar somente 25% da área de busca original, que corresponde ao traçado em vermelho na Figura 2(e). Contudo, o padrão octogonal é capaz de cobrir em média 62,3% de todas as melhores correspondências de blocos encontradas na etapa de Busca Raster, considerando os mesmos vídeos analisados anteriormente.

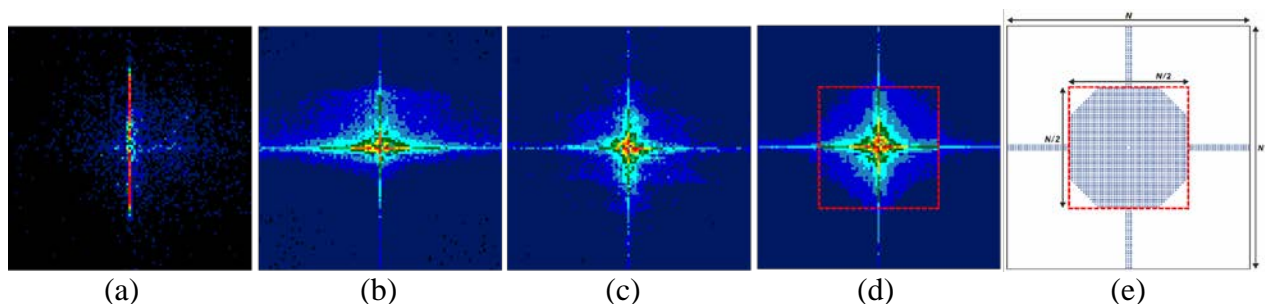


Figura 2: Distribuição de correspondência de bloco na área de busca 512×512 para os vídeos (a) *BQTerrace*, (b) *YachtRide*, e (c) *Cactus*; (d) Distribuição média para todos os vídeos analisados; (e) Padrão de busca octogonal proposto

3. RESULTADOS E DISCUSSÃO

Para validar a eficiência do esquema proposto, foram utilizadas oito sequências de vídeo de ultra-alta e alta definição. Todos os quadros de cada vídeo foram codificados com QPs 22, 27, 32 e 37. É importante enfatizar que todas as sequências são diferentes das utilizadas para as análises apresentadas neste trabalho, e diferem em taxa de quadros, profundidade de bits, resolução espacial e conteúdo de movimento/textura. A implementação do HM modificado foi comparada ao original para permitir comparações em termos de redução de tempo total de codificação e redução de tempo somente para o TZS. A eficiência de compressão foi calculada em termos de BD-rate (BJØNTEGAARD, 2011).

Os resultados do padrão proposto são mostrados na Tabela I. O esquema proposto atinge redução de tempo em média de 60,91% para o TZS e 21,57% para todo o processo de codificação, com um aumento de BD-rate de apenas 0,037% em relação ao software de referência, o que é uma perda insignificante quando leva em consideração as grandes reduções no tempo de codificação.

Tabela 1: Comparações entre o Padrão Proposto e o TZS original

Classe	Sequência de Vídeo	BD-rate (%)	Redução de Tempo TZS (%)	Redução de Tempo Total (%)
Classe A	<i>PeopleOnStreet</i>	-0,2519	60,65	18,99
	<i>SteamLocTrain</i>	0,0510	61,40	21,30
Classe B	<i>BasketballDrive</i>	-0,0291	59,55	20,93
	<i>Kimono</i>	-0,0721	53,19	13,58
Classe A1	<i>CampfireParty</i>	0,0955	68,16	30,04
	<i>ToddlerFountain</i>	-0,0305	59,96	16,08
Classe A2	<i>CatRobot</i>	0,0255	61,71	22,88
	<i>DaylightRoad</i>	0,5084	65,66	28,78
Média		0,0371	60,91	21,57

4. CONCLUSÕES

Este trabalho apresentou um novo padrão de busca octagonal na etapa de Busca Raster para redução de esforço computacional do algoritmo Test Zone Search. Com base nos resultados apresentados, é possível afirmar que o esquema obteve um ganho em redução de tempo considerável quando comparado a versão original, com uma variação de BD-rate desprezível.

5. REFERÊNCIAS BIBLIOGRÁFICAS

- SULLIVAN et al. Overview of the High Efficiency Video Coding (HEVC) Standard; **IEEE Transactions on Circuits and Systems for Video Technology**. [S.l:s.n], v. 22 n.12, p.1649-1668, 2012. JCT-VC. High Efficiency Video Coding text specification draft 10, doc. JCTVC-L1003, Genebra, Suíça, 2013.
- CORRÊA, G.; et al. Performance and Computational Complexity Assessment of High Efficiency Video Encoders. **IEEE Transactions on Circuits and Systems for Video Technology**, v. 22, n. 12, p. 1899-1909, 2012.
- GONÇALVES, P; et al. Multiple early-termination scheme for TZ search algorithm based on data mining and decision trees. **2017 IEEE 19th International Workshop on Multimedia Signal Processing (MMSP)**, p. 1-6, 2017.
- BJØNTEGAARD, G. **Calculation of Average PSNR Differences between RD-curves**. ITU Documents, Austin 2011. Acessado em 05 julho de 2017, Online. Disponível em: wftp3.itu.int/av-arch/video-site/0104_Aus/VCEG-M33.doc