ESTIMAÇÃO DE MOVIMENTO RÁPIDA PARA O CODIFICADOR DE VÍDEO HEVC BASEADA EM APRENDIZADO DE MÁQUINA

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

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

1. INTRODUÇÃO

Frente à crescente demanda por avanços nas tecnologias de vídeos digitais, principalmente devido às diversas aplicações de vídeos de alta resolução em dispositivos portáteis, como *smartphones* e *tablets*, e serviços de *streaming*, que tornaram o consumo desta 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. 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 (OHM, 2012). 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), 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, pois é responsável por buscar o bloco mais similar ao bloco que está sendo codificado no quadro atual em um conjunto de quadros 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 esquema de terminação antecipada para o TZS, baseado em um extensivo processo de mineração de dados e árvores de decisão.

2. METODOLOGIA

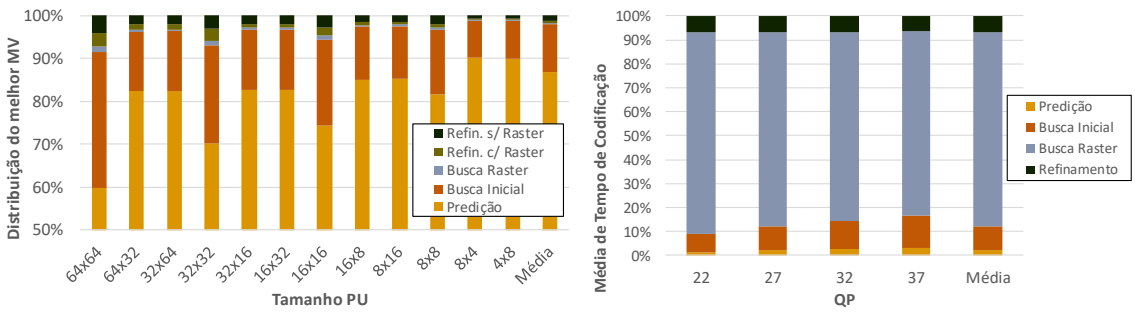
Inicialmente, foi realizado uma análise do comportamento do algoritmo do TZS no processo de codificação de alguns vídeos. Para isto, foram utilizadas seis sequências de vídeo: *BQTerrace*, *BasketballDrive*, *Traffic*, *NebutaFestival*, *HoneyBee* e *ReadySetGo*, todas com resoluções de alta definição. Também foram utilizados valores de Parâmetro de Quantização (QP): 22, 27, 32 e 37.

Uma primeira análise foi realizada com foco na distribuição das etapas do TZS em relação à decisão de quais etapas geram com mais frequência o melhor vetor de movimento (MV). Figura 1(a) mostra a distribuição dos diferentes tamanhos de PUs. Considerando o resultado médio (coluna mais à direita), 87% dos melhores MVs são obtidos após a etapa de Predição. Da mesma forma, 11% são obtidos após a etapa de Busca Inicial. As etapas de Busca Raster e Refinamento são responsáveis por obter, respectivamente, 0,4% e 1,6% dos melhores MVs.

Após analisar a distribuição dos melhores MVs, foi realizado uma análise de tempo para cada etapa do TZS, como mostrado na Figura 1(b). É possível perceber que a etapa da Busca Raster é a mais complexa dentre todas, responsável em média por 81% do tempo total de execução, apesar de apenas 2% dos MVs encontrados são originados desta etapa em diante. Entretanto, as etapas de Predição e Busca Inicial representam somente 2% e 10% do tempo total do TZS respectivamente, apesar de serem responsáveis por encontrar 87% e 11% dos melhores MVs.

Esta análise mostra que, na maioria dos casos, o custo computacional para processar algumas etapas do TZS é desperdiçado, sendo que a maior parte dos melhores MVs são encontrados nas duas primeiras etapas, especialmente para PUs de tamanhos maiores. Desta forma, identificando esses casos podemos ignorar as últimas etapas (especialmente a Busca Raster, a mais complexa), com baixas perdas em eficiência de compressão.

O esquema proposto de terminação antecipada para o algoritmo do TZS é baseado em um extensivo processo de mineração de dados e na construção de três modelos de árvores de decisão, chamados de Pr-ET, FS-ET e RS-ET, que correspondem respectivamente aos modelos aplicados após as etapas de Predição, Busca Inicial e Busca Raster, conforme mostrado na Figura 2. O objetivo de cada modelo é determinar se o melhor MV encontrado ao final da execução do TZS é oriundo da etapa anterior. Por ser o primeiro aplicado, o modelo Pr-ET tem maiores chances de tomar uma decisão errada entre interromper ou continuar a execução do TZS. Contudo, as maiores reduções de tempo são obtidas quando a interrupção ocorre nesta etapa. Ao contrário disto, o modelo RS-ET apresenta menores ganhos em redução de tempo, porém suas decisões impactam em pequenas perdas na eficiência de compressão.



(a) Escolha do melhor MV (b) Tempo médio de codificação por QP

Figura 1: Distribuição por etapas do TZS

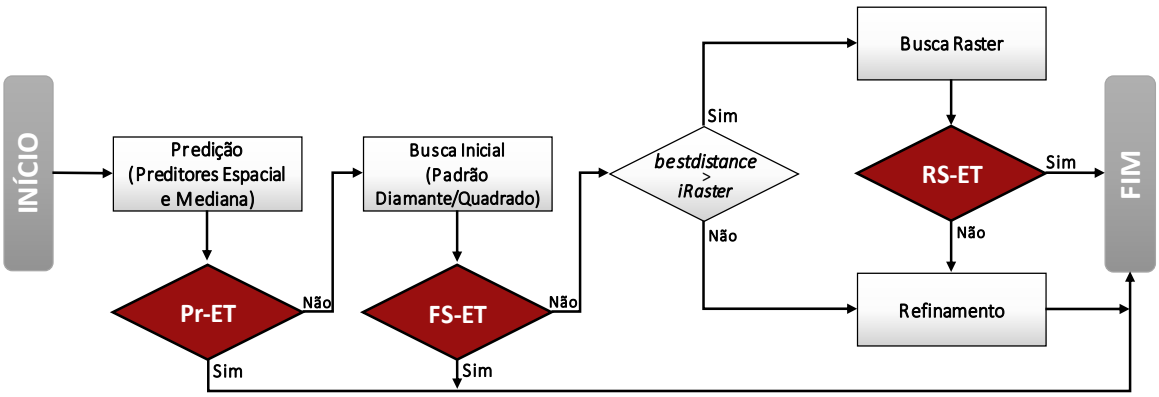


Figura 2: Fluxograma de execução do e-TZS

O processo de mineração de dados considerou mais de 40 parâmetros coletados em diversas execuções do codificador HM (versão 16.14) com as mesmas sequências e condições de teste mencionados anteriormente. Somente os primeiros 16 quadros de cada vídeo foram codificados, devido a grande quantidade de dados coletados. Para cada um dos modelos, um total de 600.000 instâncias foram coletadas (25.000 de cada combinação vídeo-QP), garantindo assim que os seis vídeos e os quatro valores de QP sejam igualmente representados no conjunto de dados.

O processo de mineração de dados foi realizado separadamente para cada modelo. Isto é necessário pois a quantidade de parâmetros disponíveis em cada etapa é diferente. Por exemplo, o SAD obtido do bloco escolhido após a etapa de Busca Inicial é obviamente disponível somente após a execução desta etapa. Sendo assim, este parâmetro não pode ser incluído no conjunto de dados para o modelo Pr-ET, porém pode ser incluído no conjunto de dados de ambos FS-ET e RS-ET. Para a escolha dos parâmetros utilizados em cada modelo, foi medida a relevância de cada um através do método de Avaliação de Atributos de Ganho de Informação (IGAE), implementado no *software* WEKA, uma ferramenta de mineração de dados gratuita e de código aberto (HALL, 2009). O treinamento foi realizado utilizando o algoritmo C4.5, implementado no WEKA como a ferramenta J48.

Após o treinamento, Pr-ET apresentou uma precisão de 72,4%. Isso significa que o modelo é incapaz de decidir corretamente, após a etapa de predição, entre interromper ou continuar a execução do algoritmo do TZS em 27,5% dos casos. No entanto, é importante notar que as classificações incorretas ocorrem em apenas 17,2% das decisões imprecisas, quando o TZS é encerrado, porém deveria continuar executando. Os 10,4% das decisões incorretas restantes não causam uma classificação incorreta, uma vez que o e-TZS indica que a execução deve continuar, mas o melhor MV já foi encontrado. Estes casos apenas impactam na falta de redução de complexidade, entretanto sem perdas na eficiência de codificação. Desta forma, a precisão real pode ser representada como a soma da precisão do modelo e dos casos onde a decisão incorreta não causa perda em eficiência de codificação. Portanto, a precisão real do modelo Pr-ET é de 82,8%. O modelo FS-ET apresenta precisão real de 99,8%, enquanto que o modelo RS-ET chegou a uma precisão real de 100%.

Devido a essas características, é possível notar que através de uma exploração eficiente dos atributos extraídos do processo de mineração de dados, o esquema do e-TZS pode prever as decisões com uma precisão média de 94,2% e, através disto, acelerar a execução do algoritmo do TZS com perdas insignificantes em eficiência de codificação, como mostrado na próxima seção.

3. RESULTADOS E DISCUSSÃO

Para validar a eficiência do esquema proposto, foram utilizadas seis sequências de vídeo. São elas: *SteamLocomotiveTrain*, *PeopleOnStreet*, *Kimono*, *ParkScene*, *Beauty* e *YachtRide*. Os primeiros 150 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 o treinamento das árvores de decisão, e diferem em taxa de quadros, profundidade de bits, resolução espacial e conteúdo de movimento/textura. A implementação do e-TZS foi comparado a versão original para permitir resultados 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 e BD-PSNR (BJØNTEGAARD, 2011).

Os resultados do e-TZS são mostrados na Tabela 1. O esquema proposto atinge redução de tempo em média de 62,53% para o TZS e 13,53% para todo o processo de codificação, com um aumento de BD-rate de apenas 0,49% 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 e-TZS e o TZS original

Sequência de Vídeo	BD-Rate (%)	BD-PSNR (dB)	Red. Tempo Total (%)	Red. Tempo TZS (%)
Kimono	0.42	-0.01	15.47	67.76
ParkScene	0.35	-0.01	10.07	64.49
SteamLocomotiveTrain	0.33	-0.01	10.03	66.30
PeopleStreet	0.65	-0.03	10.69	52.56
Beauty	0.61	0.00	15.04	56.85
YachtRide	0.57	-0.02	19.85	67.21
Média	**0.49**	**-0.01**	**13.53**	**62.53**

4. CONCLUSÕES

Este trabalho apresentou um esquema de terminação antecipada, chamado e-TZS, baseado em mineração de dados e árvores de decisão para redução de complexidade do algoritmo *Test Zone Search* (TZS). Com base nos resultados apresentados, é possível afirmar que o e-TZS atingiu uma redução de tempo considerável (62,53%) quando comparado à versão original do TZS, com uma variação de BD-rate de apenas 0,49%.

5. REFERÊNCIAS BIBLIOGRÁFICAS

SULLIVAN et al. Overvew of the High Eficiency 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.

OHM et al. Comparison of the Coding Efficiency of Video Coding Standards – Including High Efficiency Video Coding (HEVC). **IEEE Transactions on Circuits and Systems for Video Technology**, v. 22, n. 12, p. 1669-1684, 2012.

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.

HALL, M.; et al. **The WEKA data mining software: an update**. SIGKDD Explor. Newsl., v. 11, p. 10-18, 2009.

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