



ARQUITETURA DE HARDWARE DEDICADA PARA O MODO DMM-1 DA PREDIÇÃO INTRA-QUADRO DO 3D-HEVC

VINICIUS BORGES¹; MURILO PERLEBERG²; VLADIMIR AFONSO³;
MARCELO PORTO⁴; LUCIANO AGOSTINI⁵

¹Universidade Federal de Pelotas – vdaborges@inf.ufpel.edu.br

²Universidade Federal de Pelotas – mperleberg@inf.ufpel.edu.br

³Instituto Federal Sul-rio-grandense – vafonso@inf.ufpel.edu.br

⁴Universidade Federal de Pelotas – porto@inf.ufpel.edu.br

⁵Universidade Federal de Pelotas – agostini@inf.ufpel.edu.br

1. INTRODUÇÃO

Atualmente a evolução tecnológica nas mais diversas áreas ocorre muito rapidamente, e estas tecnologias se tornam cada vez mais importantes e mais presentes na vida da população. Dentre os diversos tipos de tecnologia presentes no dia a dia das pessoas estão os vídeos digitais, sejam eles filmes, vídeos feitos no celular ou até mesmo transmissões ao vivo. Dentre os vídeos ainda existem os chamados vídeos 3D, que são vídeos que utilizam técnicas que simulam um ambiente de três dimensões para o espectador, o que permite ao espectador a percepção da profundidade dos objetos que compõem a cena.

Mas os vídeos, principalmente os vídeos 3D, demandam uma quantidade de dados gigantesca se utilizados sem nenhum método de compressão, de forma que seria inviável armazenar um filme na nuvem ou em um computador convencional. Por este motivo, a utilização de métodos de compressão de vídeos é necessária para tornar estes vídeos viáveis de serem armazenados e ao mesmo tempo não tenham perdas significativas na qualidade de sua imagem.

Um dos padrões em compressão de vídeos 3D é o 3D-HEVC. O 3D-HEVC, na verdade, é uma extensão do padrão HEVC (*High Efficiency Video Coding*). Esta extensão suporta o formato MVD (*Multi-View plus Depth*) que foi desenvolvido para a codificação de múltiplas vistas (representam a cena utilizando várias câmeras sob diferentes perspectivas). Neste formato, cada uma destas vistas é dividida em um canal de textura e um canal de profundidade, onde o canal de textura representa as informações de cores e luminância da imagem, enquanto que o canal de profundidade representa a distância entre os objetos e a câmera.

O formato MVD também permite ao 3D-HEVC a síntese de vistas intermediárias baseada nas vistas originais capturadas com o objetivo de criar mais vistas (vistas virtuais) para a cena. A geração destas vistas pelo 3D-HEVC usa um processo síntese chamado DIBR (*Depth-Image-Based Rendering*) que gera vistas intermediárias sintéticas (TECH, 2016).

Para codificar um vídeo, cada quadro é dividido em blocos chamados de CTUs (*Coding Tree Units*). Essas CTUs tem o tamanho padrão de 64x64. Após, essas CTUs são divididas em blocos menores chamados de PUs (*Prediction Units*). As PUs possuem diversos tamanhos diferentes, porém a predição intra-quadro trabalha somente com PUs quadráticas, como 4x4, 8x8, 16x16, 32x32 e 64x64.

O 3D-HEVC possui todos os modos de predição intra do HEVC, e também alguns novos exclusivos para o conteúdo dos mapas de profundidade, que são o DMM-1, DMM-4 (*Depth Modeling Modes 1 e 4*), e o DIS (*Depth Intra Skip*). O modo DIS foi projetado para vídeos com áreas mais homogêneas evitando o

processo de codificação dos resíduos dessas áreas. Por sua vez, modo DMM-1 divide o bloco que vai ser predito em duas partições, essa divisão é feita por uma reta que corta o bloco formando essas duas partições. Cada partição é representada usando um CPV (*Constant Partition Value*), que é a média dos valores das amostras da partição específica. Embora o modo DMM-1 não seja aplicado sobre PUs de tamanho 64x64, ele possui uma alta complexidade visto que ele possui até 1908 retas diferentes para avaliação. E por fim, o modo DMM-4 que apresenta similaridades ao DMM-1, porém este utiliza um valor de limiar para calcular o contorno dos objetos do bloco predito. Este valor de limiar é obtido de uma média com amostras específicas do bloco de textura colocalizado.

2. METODOLOGIA

Uma arquitetura visando diminuir a complexidade da ferramenta DMM-1 foi implementada. A arquitetura desenvolvida avalia apenas seis retas das 1908 retas possíveis. Estas seis retas cortam as bordas do bloco, sendo apenas uma reta para cada par formado por bordas diferentes. Conforme pode ser visto na Figura 1, as possibilidades de divisão de um bloco 4x4 são superior-esquerda (SE), superior-inferior (SI), superior-direita (SD), esquerda-direita (ED), inferior-esquerda (IE), e inferior-direita (ID), respectivamente. A posição inicial e final de cada uma das retas é definida através do gradiente de cada uma das bordas do bloco, e a reta é traçada por meio do algoritmo de Bresenham (FLANAGAN, [SD]), que define quais amostras pertencem a cada uma das regiões.

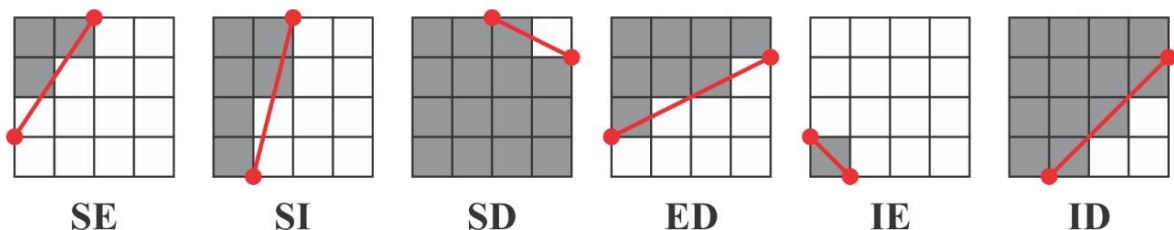


Figura 1: Exemplos das seis retas e as bordas que elas cortam. A parte em cinza é referente à partição A e a parte em branco é referente à partição B.

A arquitetura possui duas unidades definidas, a unidade de controle e a unidade de operação. A unidade de controle é responsável pelos sinais que controlam as máquinas da unidade de operação. Na unidade de controle, o sinal de início é enviado até a parte operativa e esta também espera por um sinal de resposta das máquinas. A unidade de operação é onde as máquinas fazem todo o processo de predição dos blocos. Esta unidade possui apenas duas máquinas, a máquina que realiza predição apenas com blocos 4x4, e a máquina que realiza predição dos outros tamanhos de bloco como o 8x8, 16x16, e 32x32. A divisão das máquinas de acordo com os tamanhos de bloco foi definida desta forma devido ao tempo de processamento de uma CTU para blocos 4x4, que é maior que o tamanho dos outros três tamanhos juntos.

Dentro de cada máquina ainda existem seis módulos, um módulo para cada reta, conforme pode ser visto na Figura 2-(b). O processamento de cada módulo é idêntico, e o processamento é feito em paralelo para otimizar o tempo de processamento, pois a predição de cada uma das retas é independente das outras.

Na Figura 2-(b) é possível ver o diagrama de cada módulo. Para realizar a predição de cada PU, primeiramente, cada módulo define o valor do gradiente das

bordas através de um calculador de gradiente, que armazena a posição da maior diferença entre duas amostras vizinhas da borda, conforme apresentado na Figura 2-(a). Estes valores de gradiente são utilizados para determinação dos pontos iniciais e finais das retas que cortarão os blocos preditos.

Após, um bloco *bitmap* é gerado, onde a partição A é preenchida com o valor 1, e a partição B é preenchida com o valor 0, conforme representado na Figura 2-(c). Baseado nesse *bitmap*, o calculador de bloco preenche o bloco predito, onde todas as posições de valor 1 são preenchidas com o valor médio da partição A, enquanto que as posições de valor 0 são preenchidas com o valor médio da partição B, como representado na Figura 2-(c). Por fim, este bloco com os valores já alterados passa por um calculador de SAD, que compara o bloco predito com o bloco original e, assim, gera o resíduo.

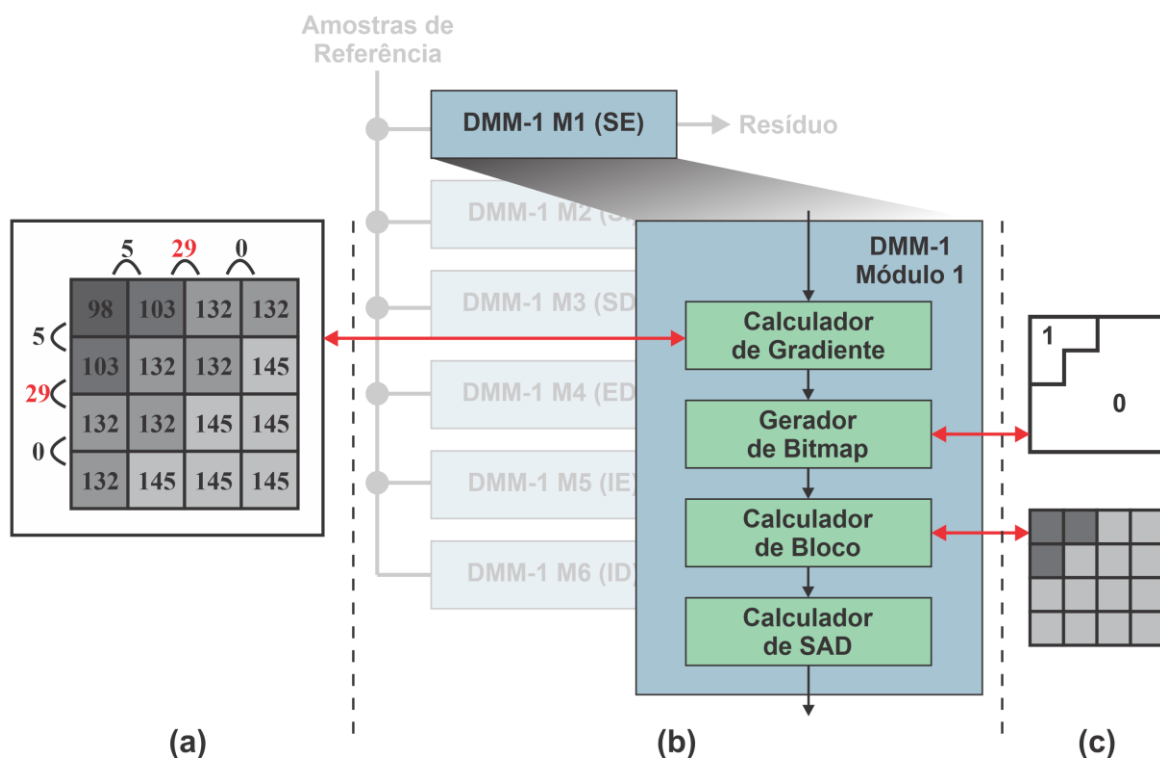


Figura 2: Processo de avaliação de cada reta. (a) Cálculo do gradiente. (b) Diagrama de blocos da máquina que processa PUs 4x4. (c) *Bitmap* e o bloco predito.

3. RESULTADOS E DISCUSSÃO

A arquitetura foi descrita em VHDL e foi sintetizada utilizando a tecnologia TSMC 40nm. Para realizar esta síntese, foi utilizada a ferramenta RTL Compiler da Cadence.

Conforme pode ser visto na Tabela 1, ambos os trabalhos relacionados possuem uma área significativamente menor, além de trabalharem com outros modos do 3D-HEVC. Porém, este trabalho não necessita de memória para armazenar as retas, visto que são apenas seis retas utilizadas, também é possível alcançar o processamento de maiores resoluções do que nestes trabalhos, de quase 2 vezes maior que a resolução alcançada por Amish (2019), e 11 vezes maior que a resolução alcançada por Sanchez (2017). Por fim, a potência

dissipada é menor, sendo aproximadamente 69,37% menor que a dissipação da arquitetura desenvolvida por Sanchez (2017).

Em comparação ao algoritmo original que avalia 1908 retas, o algoritmo proposto que avalia apenas 6 retas teve um aumento de 4,444% na métrica BD-rate considerando o perfil *all-intra*, e um aumento de 2,040% no BD-rate para o perfil *random-access*. O BD-rate define qual a porcentagem de bits que se necessita acrescentar para que não haja perda na qualidade da imagem.

Trabalhos	SANCHEZ (2017)	AMISH (2019)	Este trabalho
Modos Intra-quadro	DMM-1 e DMM-4	DIS, DMM-1, e DMM-4	DMM-1
Tecnologia	ASIC 65nm	FPGA Xilinx Virtex 6 (40nm)	TSMC 40 nm
Área	219,95 k* gates (2-entradas NAND)	330 k gates (2-entradas NAND)	1.283 k gates (2-entradas NAND)
Memória para retas	Sim (101,35 k bits*)	Sim (1,95 M bits)	Não
Resolução Máxima	1080p@30fps 1 vista*	1080p@30fps 6 vistas	1080p@30fps 11 vistas
Potência para HD1080p@30fps	166,5 mW* (1 vista)	Não avaliado	51 mW (1 vista)

*Valores considerados apenas para processamento de blocos 32x32.

Tabela 1: Resultados obtidos no estudo em comparação com outros trabalhos.

4. CONCLUSÕES

Este trabalho apresenta um hardware dedicado ao modo DMM-1 do 3D-HEVC que é capaz de realizar a predição de todos os tamanhos de bloco utilizados por este modo.

Apesar de requerer mais recursos computacionais que os trabalhos relacionados, esta arquitetura é capaz de processar maiores resoluções sem grandes perdas na qualidade da imagem, aumentando a eficiência temporal do modo e reduzindo consideravelmente o consumo energético.

5. REFERÊNCIAS BIBLIOGRÁFICAS

SANCHEZ, G. et al. "Real-time scalable hardware architecture for 3D-HEVC bipartition modes," **J. Real-Time Image Processing**, pp. 71-83. 2017.

AMISH, F. and E. Bourennane. "An efficient hardware solution for 3D-HEVC intra-prediction," **J. Real-Time Image Processing**, pp. 1559-1571. 2019.

TECH, G. et al. "Overview of the multiview and 3D extensions of high efficiency video coding," **IEEE Trans. Circuits Syst. Video Technol.**, vol. 26, n. 1, pp. 35-49, Jan. 2016.

FLANAGAN C. **The Bresenham Line-Drawing Algorithm**. SD. Acessado em 15 set. 2020. Online. Disponível em: <https://www.cs.helsinki.fi/group/goa/mallinnus/lines/bresenh.html>