

ISOLAMENTO DE OPERANDOS BASEADO EM APRENDIZADO DE MÁQUINA PARA PROJETO DE HARDWARE ENERGETICAMENTE EFICIENTE DOS FILTROS DE INTERPOLAÇÃO DA FME/MC DO AV1

LEONARDO MÜLLER¹; DAIANE FREITAS²; PATRICK ROSA³; DANIEL PALOMINO⁴; GUILHERME CORREA⁵

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

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

³Universidade Federal de Pelotas – psdrosa@inf.ufpel.edu.br

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

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

1. INTRODUÇÃO

O consumo de vídeo corresponde atualmente à maior parcela do tráfego da internet, sendo cerca de 65% do total (DATAREPORTAL, 2022), impulsionado principalmente pelo uso de dispositivos móveis. Esse cenário reforça a necessidade de codificadores de vídeo mais eficientes, capazes de reduzir a taxa de bits mantendo a qualidade visual, de forma a economizar largura de banda e energia.

Entre as soluções modernas, o AOMedia Video 1 (AV1), lançado em 2018 pela *Alliance for Open Media*, destaca-se por oferecer ganhos de compressão entre 30% e 40% em relação ao VP9, seu antecessor, e não possuir custos de licenciamento, diferente dos padrões como HEVC (ALLIANCE FOR OPEN MEDIA, 2020). Para alcançar esse desempenho, o AV1 introduz diversas novas ferramentas, entre elas o esquema de filtragem adaptativa por interpolação, utilizado nas etapas de *Motion Estimation* (ME) e *Motion Compensation* (MC).

A Fractional Motion Estimation (FME), em particular, depende da geração de pixels fracionários a partir de até 75 filtros de interpolação do tipo FIR, organizados em três famílias principais: Regular (6 taps), Sharp (8 taps) e Smooth (6 taps). O AV1 também adota um mecanismo de filtro duplo, permitindo a seleção independente de filtros para as direções horizontal e vertical, ampliando a flexibilidade na predição (ALLIANCE FOR OPEN MEDIA AV1, 2018).

No software de referência (libaom), a escolha da combinação ideal de filtros é feita por meio de um processo de busca que considera nove possíveis combinações de filtros horizontais e verticais para o mesmo vetor de movimento. Este processo avalia a combinação que resulta no menor custo taxa-distorção (RD), garantindo maior qualidade. Contudo, essa abordagem, embora eficaz, aumenta a complexidade computacional e o consumo de energia. Esse desafio tem motivado o desenvolvimento de soluções em hardware, voltadas à otimização da filtragem de interpolação do AV1. No entanto, as arquiteturas existentes ainda dependem de controle externo para decidir qual filtro aplicar, sem explorar mecanismos autônomos de seleção.

Neste contexto, este trabalho apresenta o DASH-AV1 (Decision-tree-based Architecture for Sub-pel Hardware in AV1), um acelerador de hardware que incorpora um módulo de decisão baseado em árvores de decisão de aprendizado de máquina. Essa abordagem permite a seleção dinâmica e eficiente dos filtros de interpolação, reduzindo significativamente a dissipação de energia, sem comprometer o desempenho de compressão ou a vazão do sistema.

2. DASH-AV1: ARQUITETURA DE HARDWARE BASEADA EM ÁRVORE DE DECISÃO PARA SUB-PIXEL NO AV1

A arquitetura proposta, denominada DASH-AV1 (Decision-tree-based Architecture for Sub-pel Hardware in AV1), é um acelerador de hardware eficiente em energia voltado à filtros de interpolação no codificador AV1. Seu principal diferencial está no uso de modelos de árvore de decisão para controlar dinamicamente a técnica de *operand isolation*, ativando apenas os filtros necessários de acordo com as características do conteúdo de vídeo.

Dois modelos independentes foram treinados: Modelo X para filtros horizontais e Modelo Y para filtros verticais, que devem escolher entre uma das famílias de filtros. Ambos utilizam o algoritmo *DecisionTreeClassifier* da biblioteca Scikit-Learn, configurado com profundidade máxima de 6, no máximo 20 folhas e no mínimo 500 amostras por divisão, buscando equilibrar precisão e simplicidade para viabilizar a implementação em hardware. Os conjuntos de dados para treinamento foram obtidos a partir do software de referência *libaom*, modificado para extrair informações em nível de bloco de seis vídeos, em quatro níveis distintos de quantização (CQ 20, 32, 43 e 55), seguindo as Condições Comuns de Teste (CTC) (CHENG, 2020). O dataset inclui oito atributos principais, entre eles: tamanho e posição do bloco, modo de predição, estatísticas de blocos vizinhos e informações contextuais sobre o uso de filtros.

No hardware, a arquitetura é composta por dois módulos centrais: o Decision Tree Controller (DTC) e o núcleo de filtragem sub-pixel. O DTC, mostrado na Figura 1, integra os modelos X e Y diretamente em hardware, realizando previsões em tempo real para selecionar a família de filtros adequada em cada direção. Para coordenar essa decisão, o DTC utiliza uma unidade de contadores baseada em registradores que acompanha o progresso do processamento de cada bloco de 4×4 pixels, distribuído em 51 ciclos de clock: 12 para filtragem horizontal, 4 para vertical e 35 para a combinação das direções. O resultado da predição gera sinais de controle que, através de portas AND, ativam apenas os filtros necessários, implementando a técnica de isolamento de operandos e reduzindo significativamente a dissipação de potência dinâmica.

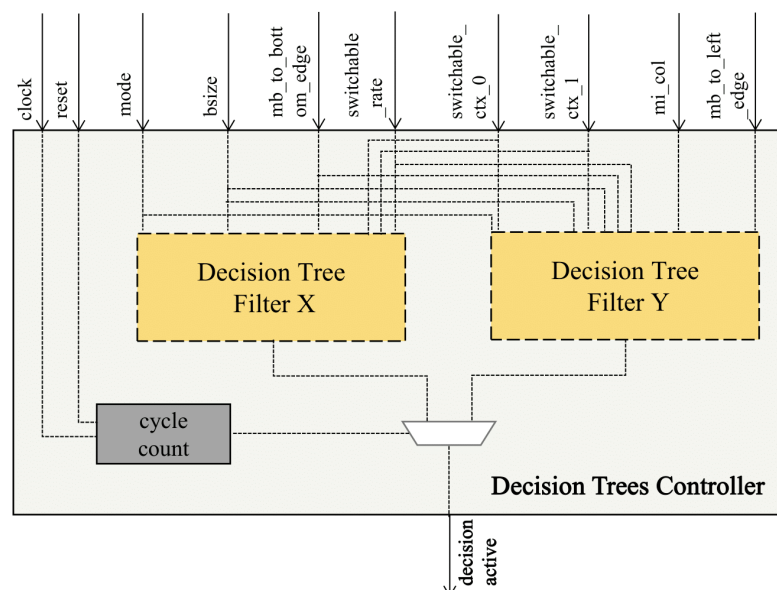


Figura 1: Controlador das Árvore de Decisão usado na arquitetura DASH-AV1.

O núcleo de filtragem sub-pixel reutiliza a arquitetura multiversão apresentada em trabalhos anteriores, capaz de realizar tanto versões precisas quanto aproximadas das três famílias de filtros do AV1. A operação é viabilizada por uma unidade compartilhada de *Multiple Constant Multiplication (MCM)* baseada no algoritmo HCub, seguida de árvores de somadores e unidades de deslocamento que realizam as operações aritméticas. O processamento resulta em 15 amostras sub-pixel por bloco, garantindo alta eficiência e adaptação ao conteúdo sem perda de throughput.

3. RESULTADOS E DISCUSSÃO

Os modelos de árvore de decisão mostraram ganhos expressivos em relação à seleção aleatória de filtros. O Modelo X (horizontal) alcançou 68,3% de acurácia, enquanto o Modelo Y (vertical) obteve 71,96%, contra um baseline de 33,3%. Na avaliação de eficiência de codificação, a variação média de BD-Rate foi de apenas 0,83% na versão precisa da arquitetura, aumentando levemente para 0,92% e 1,21% para as versões aproximada 1 e aproximada 2, respectivamente.

Tabela 1: Resultado da Síntese do DASH-AV1.

Parâmetros	DASH-AV1		
	Precisa	App. 1	App. 2
Precisão			
Frequência Máxima (MHz)	833.3		
Resolução MC Suportada	4320p @50fps		
Vazão (Msm/s)	2,656.14		
Área Total(μm²)	778,323		
Contagem de Portas (K)	332.82		
Núcleo Regular	91.67		
Núcleo Sharp	127.01		
Núcleo Smooth	76.01		
Decision Tree Controller	1.6		
Potência Total (mW)	29.450	28.918	27.214
Potência de Fuga (mW)	0.797	0.788	0.778
Potência dinâmica total (mW)	28.653	28.129	26.437
Núcleo Regular	1.93	1.73	1.29
Núcleo Sharp	4.88	4.64	3.83
Núcleo Smooth	1.84	1.60	0.93
Decision Tree Controller	0.33		

A síntese em tecnologia TSMC 40nm demonstrou que o DTC (Decision Tree Controller) possui custo irrisório em área (1,6K portas) e consumo mínimo de 0,33 mW, validando sua viabilidade.

Quando comparado a arquiteturas mais recentes, como (DOMANSKI, R, 2021), a arquitetura destaca-se por integrar tanto filtros precisos quanto aproximados com seleção inteligente via árvores de decisão. Além disso, ao comparar com a versão anterior da arquitetura, sem o DTC, reduziu em 26,5% a dissipação de potência dinâmica, eliminando atividade de comutação desnecessária.

4. CONCLUSÕES

O DASH-AV1 apresenta-se como uma solução inovadora para a FME do AV1, integrando árvores de decisão em hardware para controlar dinamicamente a ativação de filtros. Essa abordagem permitiu manter alto throughput (2.656,14 Msamples/s), suficiente para vídeos 8K@50fps, com baixo consumo de energia (27,2–29,45 mW). Trata-se da única solução que adota aprendizado de máquina (árvores de decisão) para guiar a seleção de filtros, alcançando significativa redução de potência sem comprometer a eficiência do codec.

5. REFERÊNCIAS BIBLIOGRÁFICAS

DATAREPORTAL. Digital 2022: time spent using connected tech continues to rise. 2022. Online. Disponível em: <https://datareportal.com/>

ALLIANCE FOR OPEN MEDIA. Alliance for open media. 2020. Online. Disponível em: <https://aomedia.org/>

ALLIANCE FOR OPEN MEDIA. 2018. Online. Disponível em: <https://aomedia.org/av1-features/>

CHEN, Y.; MUKHERJEE, D.; HAN, J.; GRANGE, A.; XU, Y.; PARKER, S.; CHEN, C.; SU, H.; JOSHI, U.; CHIANG, C.-H.; WANG, Y.; WILKINS, P.; BANKOSKI, J.; TRUDEAU, L.; EGGE, N.; VALIN, J.-M.; DAVIES, T.; MIDTSKOGEN, S.; NORKIN, A.; RIVAZ, P. de; LIU, Z. An overview of coding tools in AV1: the first video codec from the alliance for open media. APSIPA Transactions on Signal and Information Processing, [S.l.], v.9, 2020.

DOMANSKI, R.; KOLODZIEJSKI, W.; CORREA, G.; PORTO, M.; ZATT, B.; AGOSTINI, L. Low-power and high-throughput approximated architecture for AV1 FME interpolation. In: IEEE INTERNATIONAL SYMPOSIUM ON CIRCUITS AND SYSTEMS (ISCAS), 2021, [S.l.]. Proceedings... IEEE, 2021. p.1-5.