

ARQUITETURA DE HARDWARE DEDICADA PARA A FME DO PADRÃO DE CODIFICAÇÃO DE VÍDEO VP9/VP10

JONES GÖEBEL; LUCAS AGOSTINI; LUCIANO AGOSTINI; BRUNO ZATT;
MARCELO PORTO

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

1. INTRODUÇÃO

O avanço tecnológico vem trazendo para as pessoas nos dias atuais, mais lazer e praticidade nas tarefas diárias. Para que a tecnologia continue trazendo avanços indispensáveis para a nossa sociedade, ela deve se encarregar de solucionar os problemas que vem surgindo. Alguns destes desafios presentes nos dias atuais que merecem destaque são projetos de hardwares para a codificação de vídeo. Os vídeos digitais estão cada vez mais presentes em nosso meio e podem ser encontrados em inúmeros dispositivos, como smartphones, televisores, *Blu-ray players*, etc. Para que estes dispositivos possam manipular os vídeos digitais, eles precisam de codecs (codificadores/decodificadores) de vídeo, mas estes codecs devem garantir também uma codificação em tempo real e ao mesmo tempo, que possibilitem um armazenar/transmissão viável do vídeo.

O aumento das resoluções dos vídeos em conjunto com o aumento da taxa de reprodução vem se tornando um dos principais desafios na área de pesquisa em codificação de vídeo, pois eles implicam na necessidade em aumentar a taxa de processamento dos vídeos digitais. Devido ao volume extremamente alto de dados que precisam ser processados em tempo real a codificação é uma tarefa impraticável para os processadores de propósito geral. Para solucionar este problema, umas das soluções adotadas é a implementação do codificador em uma arquitetura (hardware) dedicada para o processamento de vídeo.

A codificação de vídeo tem como objetivo em reduzir o tamanho do vídeo cru, o que facilita o armazenamento e a transmissão do mesmo. Desta forma, a sociedade científica vem formulando novos codificadores com o intuito de aumentar a eficiência na codificação de vídeo. Tendo na mente isso, vários padrões de codificadores de vídeo foram criados e vem sendo criados, para suprir esta necessidade, como exemplos, podemos citar os padrões MPEG-2, H.264, HEVC, VP9, AVS2, etc. Cada padrão teve seu desenvolvimento para atender a demanda e a necessidade da sua época de desenvolvimento. Com o intuito de atender a necessidade da Google, em 2011 teve inicio o desenvolvimento do padrão VP9 (MUKHERJEE et al., 2015a), tendo como o foco em vídeos em alta resolução, este desenvolvimento ocorreu quase que em paralelo com o desenvolvimento do HEVC (*High Efficiency Video Coding*) (SULLIVAN et al., 2012). O objetivo em desenvolver outro padrão independente (VP9 em relação ao HEVC) tem como a motivação ter um padrão livre de *royalty*, o que não é o caso do HEVC. Com o intuito de melhorar ainda mais a eficiência do codificador a Google lançou o software de referência para o VP10. Esta nova versão traz novas ferramenta para que o codificador possa proporcionar um ganho na ordem dos 10% na eficiência na compressão do VP9 (MUKHERJEE et al., 2015b).

O princípio básico da compressão de vídeos digitais consiste na exploração das informações redundantes. As redundâncias que podem ser encontradas nos vídeos são: espacial, temporal ou entrópica. A redundância espacial ocorre pela similaridade de uma amostra com seus vizinhos dentro de um mesmo quadro do vídeo, sendo ela explorada pela predição Intraquadro. A redundância temporal é explorada pela predição Interquadros que é formada a partir de informações replicadas entre os quadros vizinhos que compõem o vídeo (AGOSTINI, 2007). A

redundância entrópica esta relacionando a quantidade de dados necessária para representar de cada símbolo do vídeo.

No codificador a predição Interquadros é o módulo que mais proporciona ganhos na codificação de vídeo. Ela é constituída por uma etapa de Estimação de movimento (ME), sendo este módulo um dos mais complexos na codificação. Após a aplicação da ME é aplicada o módulo da Estimação de Movimento Fracionário (FME), que tem o objetivo de calcular possíveis movimentos fracionários. E por ultimo existe a Compensação de Movimento (MC), que é responsável por reconstruir o quadro que predito com as informações dos módulos da ME e da FME.

Este trabalho tem como objetivo em abordar um projeto de hardware dedicado para a FME do padrão VP9/VP10, onde esta arquitetura tem a capacidade de processar vídeos com a resolução 4K com uma taxa de reprodução de 30 frames.

2. METODOLOGIA

O Bloco da ME é o primeiro bloco da predição Interquadros, ele é responsável em encontrar o melhor casamento dos blocos do quadro atual (quadro a ser codificado) em relação com os blocos dos quadros já codificados. Este casamento gera um vetor de movimento para os melhores blocos dos quadros codificados, mas em alguns casos este vetor de movimento pode ser fracionário. Para o cálculo dos vetores fracionários é utilizada a FME, este bloco é responsável em realizar a interpolação ou a geração dos pixels fracionários a partir dos pixels do quadro atual (bloco a ser codificado e apontado pela ME) e realizar uma comparação com os pixels do bloco já codificados (que foram utilizados também na ME). A FME irá retornar como resultado a Soma das Diferenças Absolutas (SAD) e os vetores de movimento fracionários.

A arquitetura proposta por este trabalho é uma FME para o padrão VP9/VP10, esta arquitetura desenvolvida consiste de três blocos que podem ser observada na Figura 1.a: interpolador, árvore de SAD e o Best SAD. O primeiro bloco na nossa arquitetura é interpolador, sendo ele responsável por realizar o cálculo dos pixels fracionários a partir dos pixels inteiros. A segunda etapa é a árvore de SAD que consiste em realizar a subtração com os pixels originais já preditos com os pixels fracionários gerados pelo o interpolador, com o SAD é possível determinar qual das posições dos pixels fracionários apresentará o menor SAD. A terceira e a última etapa é responsável em determinar/armazenar os valores de SAD intermediário e no fim do processo determinar o menor SAD e seus respectivos vetores de movimento fracionário. Arquitetura foi projetada para realizar o processamento de blocos que constituem 8x8 pixels, para realizar o processamento completo deste bloco é preciso 71 ciclos de clock. Para realizar a interpolação de um bloco 8x8 é necessária a entrada de um bloco de 15x15 pixels, sendo assim a arquitetura é alimentada com uma matriz de dimensões 15x15 pixels do quadro atual e outra matriz 8x8 de pixels do quadro já codificado.

Na Figura 1.b é apresentada a estrutura interna do bloco de interpolação, este bloco é responsável por receber os 15 pixels de entrada e realizar a interpolação do mesmo. Como pode ser observado na Figura 1.b o primeiro módulo da arquitetura do interpolador, são os blocos de filtros H (horizontais), cada um destes blocos processa oito pixels e armazena os valores processados em registradores, para após este processamento, seja aplicado os filtros V (verticais), completando assim a etapa de interpolação. A arquitetura do interpolador utiliza oito filtros H e oito filtros V, sendo todos eles idênticos. A aplicação da interpolação faz com que a arquitetura tenha como saída 64 pixels dentre eles estão oito pixels inteiros e os demais fracionários.

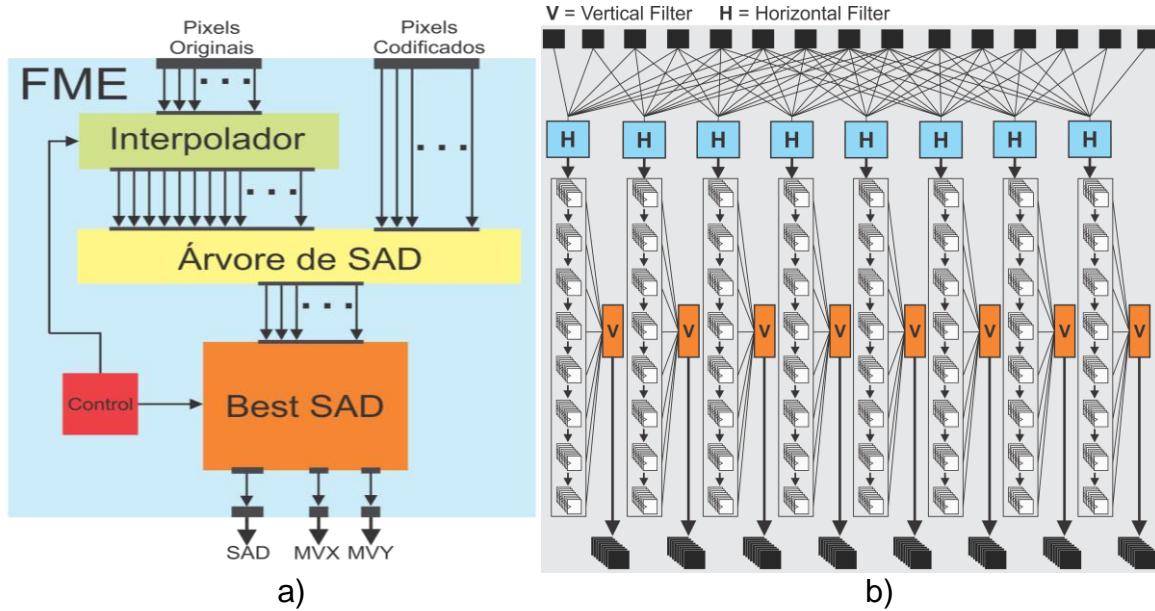


Figura 1: a) Blocos da arquitetura da FME, b) Arquitetura do Interpolador desenvolvida.

O padrão de codificação VP9/VP10 possui quatro tipos de filtros que podem ser aplicados na FME para interpolação: *Lagrange-Based*, *DCT-Based Sharp*, *Low-Pass Smooth* e *Bi-linear*. Para simplificar o trabalho desenvolvido foi escolhido implementar apenas um conjunto de filtros, sendo o conjunto de filtros implementados o DCT-Based Sharp, e somente para as amostras de luminância e sendo que os filtros implementados possuem uma precisão de 8 taps (amostras inteiras de pixels) para realizar a interpolação. No total a DCT-Based Sharp utilizada oito tipos de filtros, que podem ser observados na Tabela 1, onde os F_x são os filtros implementados, e os A_x correspondem aos pesos que cada amostra de entrada é multiplicada para a geração dos pixels fracionários, estes filtros apresentados estão presentes tanto nos filtros V como nos filtros H e são aplicados em paralelo.

Tabela 1. Pesos multiplicativos dos filtros DCT do padrão VP9/VP10.

	A0	A1	A2	A3	A4	A5	A6	A7
F0	0	0	0	128	0	0	0	0
F1	-2	5	-13	125	17	-6	3	-1
F2	-4	9	-20	115	37	-13	6	-2
F3	-4	10	-24	100	59	-19	9	-3
F4	-4	11	-23	80	80	-23	11	-4
F5	-3	9	-19	59	100	-24	10	-4
F6	-2	6	-13	37	115	-20	9	-4
F7	-1	3	-6	17	125	-13	5	-2

3. RESULTADOS E DISCUSSÃO

A arquitetura desenvolvida foi uma FME para o processamento das amostras de luminância para o padrão de codificação de vídeo VP9/VP10 para os tamanhos de blocos 8x8, sendo que ela foi descrita em VHDL e sintetizada em ASIC. Na síntese ASIC a ferramenta utilizada foi RTL compiler da Cadence e a biblioteca de Standard-Cell utilizada na síntese foi Nangate 45 nm (NANGATE, 2016) com uma tensão de alimentação de 0,95V.

Na Tabela 2 é os principais resultados obtidos na síntese em conjunto com alguns dados provenientes da característica da arquitetura. Na síntese a arquitetura não atingiu uma frequência muito superior que 276,05 MHz, mas esta frequência já é o suficiente para atingir o processamento de vídeos 4K(3840x2160 pixels) a 30 quadros por segundo. Podemos observar que a arquitetura usou

aproximadamente 276,05 Kgates (160909 um^2) em área, tendo atingido uma potência dissipada na faixa de 116,88 mW. Destaca-se que a contagem das portas lógicas é realizada baseada na área da NAND de duas entradas ($0,798 \text{ } \mu\text{m}^2$).

Tabela 2 Resultados de síntese ASIC para a arquitetura desenvolvida.

Parametros	Arquitetura Desenvolvida
Ciclo para processar um Bloco 8x8	71
Área	201,64 Kgates
Tecnologia	45 nm
Frequência	276,05 MHz
Potência total	116,88 mW

Tendo como base na frequência de 276,05MHz, atingida pela arquitetura na síntese podemos destacar que ela é capaz de realizar o processamento de 30 quadros por segundo de vídeos com a resolução UHD 4K ou 120 quadros por segundo de um vídeo na resolução Full HD (1920x1080 pixels).

4. CONCLUSÕES

O trabalho desenvolvido visa encontrar uma solução arquitetural aplicada para a FME do padrão de codificação de vídeo VP9/VP10. O trabalho apresentou uma arquitetura que contem basicamente três blocos de operação (interpolador, árvore de SAD e o *Best SAD*) para completar a FME. Os resultados obtidos na síntese demonstram que a arquitetura pode realizar o processamento de vídeos com a resolução UHD 4K com uma taxa de reprodução de 30 quadros por segundo, dissipando no total uma potência de 116,88 mW.

Para trabalhos futuros o objetivo é implementar a arquitetura utilizando todos os 4 conjuntos de filtros suportados pelo padrão, em paralelo realizar a troca de multiplicadores por soma e deslocamento. Como também reestruturar a arquitetura para permitir um maior grau de paralelismo o que possibilitaria o processamento de outros tamanhos de blocos suportados pelo padrão (além do tamanho 8x8 abordado por este escopo).

5. REFERÊNCIAS BIBLIOGRÁFICAS

- AGOSTINI, L. V. **Desenvolvimento de Arquiteturas de Alto Desempenho Dedicadas à Compressão de Vídeo Segundo o Padrão H.264/AVC.** n. Tese (Doutorado em Ciência da Computação) – Instituto de Informática, UFRGS, Porto Alegre, p. 172, 2007.
- MUKHERJEE, D. et al. **An overview of new video coding tools under consideration for VP10: the successor to VP9.** SPIE Optical Engineering + Applications, v. 9599, p. 95991E, 2015a.
- MUKHERJEE, D. et al. **A Technical Overview of VP9—The Latest Open-Source Video Codec.** SMPTE Motion Imaging Journal, v. 124, n. 1, p. 44–54, 2015b.
- NANGATE. NanGate - The Standard Cell Library Optimization Company. Disponível em: <http://www.nangate.com/?page_id=22>. Acesso em: 25 set. 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.