

PROJETO DE HARDWARE PARA O INTERPOLADOR DA FME DO PADRÃO DE CODIFICAÇÃO DE VÍDEO VP9

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

E meados de 2011, a Google começou o desenvolvimento de um codificador livre de royalty, este codificador viria a ser chamado de VP9 (Mukherjee *et al.*, 2015). O desenvolvimento do VP9 tinha como objetivo em proporcionar para Google um codificador para ser utilizado nas suas plataformas, como por exemplo no YouTube. O codificador VP9 foi finalizado em 2013, no mesmo ano que o codificador HEVC (High Efficiency Video Coding) (Sullivan *et al.*, 2012) seria normalizado, que é um codificador não é livre de royalty como o VP9.

O Objetivo da codificação de vídeo consiste em reduzir a quantidade de dado necessário para representar o vídeo e sempre que possível procurando evitar a degradação da qualidade do vídeo. 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.

Dentre todos os módulos presentes em um codificador a predição Interquadros é a que proporciona maiores ganhos na codificação. Ela é constituída por uma etapa de Estimção de movimento (ME), sendo este módulo um dos mais complexos nos codificadores atuais. Após a aplicação da ME é aplicada o módulo da Estimção de Movimento Fracionário (FME), que tem o objetivo de calcular possíveis movimentos fracionários. E por último existe a Compensação de Movimento (MC), que é responsável por reconstruir o quadro predito através das informações geradas pelos módulos da ME e da FME. A FME do VP9 introduziu uma maior complexidade na etapa de predição Interquadros, justamente por definir a utilização de quadro famílias de filtros para a Interpolação e os filtros definidos podem atingir a precisão de 1/8 para as amostras de luminância e 1/16 para as amostras de croma (Paim *et al.*, 2016).

Como pode ser notado o projeto de hardware para o Interpolador da FME do padrão VP9 é um desafio para o projeto de hardware, devido principalmente a complexidade empregada nesta etapa de codificação. Esta complexidade se deve a alta precisão utilizada nos filtros e o alto número de filtros suportados pelo padrão (28 filtros distribuídos em quatro famílias). Contudo o projeto de hardware para os codificadores é indispensável devido ao aumento das resoluções dos vídeos digitais e pelo proibitivo número de bits necessário para representá-lo em caso de nenhuma codificação. Deste modo este trabalho vem abordar um projeto de hardware com o foco no Interpolador da FME do padrão VP9, diante dos desafios elucidados anteriormente esta arquitetura desenvolvida deve ser eficiente em recursos de área e em potência dissipada.

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 geração dos pixels fracionários através da interpolação dos pixels do quadro atual (bloco a ser codificado e apontado pela ME). A FME também é responsável por realizar uma comparação com os pixels do bloco já codificados (que foram utilizados também na ME) e as amostras fracionárias geradas por ele. Esta comparação terá como resultado a Soma das Diferenças Absolutas (SAD) e os vetores de movimento fracionários para o melhor bloco candidato.

O processo de interpolação definido no VP9 é baseado em filtros FIR (Finite Impulse Response) e bi-lineares. Como mencionado anteriormente a FME atinge a máxima precisão de 1/8 para as amostras de luminância e 1/16 para as amostras de croma. Para a aplicação da interpolação também são definidos pelo padrão quatro famílias de filtros sendo todos de 8 tap (amostras inteiras de pixels): *Lagrange-Based*, *DCT-Based Sharp*, *Low-Pass Smooth* e *Bi-linear*. Na Tabela 1 é possível observar todos os filtros e com seus respectivos pesos para cada precisão fracionária.

A arquitetura desenvolvida por este trabalho aborda o Interpolador da FME do padrão VP9. Para simplificar o projeto de hardware foi realizada uma avaliação para determinar os tamanhos de blocos mais selecionados na decodificação. Para a realização desta análise foram utilizados as seguintes sequências de vídeos: *Traffic* e *PeopleOnStreet* (2560x1600 pixels); *Kimono*, *ParkScene*, *Cactus*, *BQTerrace* e *BasketballDrive* (1920x1080 pixels); *FourPeople*, *Johnny*, *KristenAndSara*, *SlideEditing* e *SlideShow* (1280x720 pixels). Esta primeira análise indicou que os quatro maiores blocos quadrados correspondem ao total de 89% de todas as seleções realizadas (64x64 representa 35%, 32x32 representa 27%, 16x16 representa 21% e 8x8 representa 6%), sendo que o codificador VP9 tem suporte para 13 tamanhos de blocos. A partir destes resultados foi realizada uma segunda análise que tinha o objetivo de limitar o codificador a estes quatro tamanhos de blocos. Nesta segunda análise podemos constatar que o tempo médio de codificação reduziu 26% e o BD-rate apresentou um acréscimo de 2,94%.

Arquitetura foi projetada para realizar o processamento de blocos que constituem 8x8 amostras, para realizar o processamento completo deste bloco é preciso 72 ciclos de clock. Para realizar a interpolação de um bloco 8x8 é necessária a entrada de um bloco de 15x15 amostras, sendo assim a arquitetura é alimentada com uma matriz de dimensões 15x15 amostras do quadro atual e outra matriz 8x8 de pixels do quadro já codificado.

Tabela 1. Quatro famílias de filtros suportados pelo padrão VP9.

| Y | Lagrange-Based | DCT-Based Sharp | Low-Pass Smooth | Bi-linear |
|-----|----------------------------|-----------------------------|---------------------------|----------------------|
| 1/8 | {-1,3,-10,122,18,-6,2,0} | {-2,5-13,125,17,-6,3,-1} | {-2,-2,29,63,41,2,-3,0} | {0,0,0,112,16,0,0,0} |
| 2/8 | {-1,4,-16,112,37,-11,4,-1} | {-4,9,-20,115,37,-13,6,-2} | {-2,-3,24,62,46,5,-4,0} | {0,0,0,96,32,0,0,0} |
| 3/8 | {-1,5,-19,97,58,-16,5,-1} | {-4,10,-24,100,59,-19,9,-3} | {-1,-4,18,59,51,9,-4,0} | {0,0,0,80,48,0,0,0} |
| 4/8 | {-1,6,-19,78,78,-19,6,-1} | {-4,11,-23,80,80,-23,11,-4} | {-1,-4,14,55,55,14,-4,-1} | {0,0,0,64,64,0,0,0} |

Na Figura 1.a é apresentada a estrutura interna do bloco de interpolação, este bloco é responsável por receber as 15 amostras de entrada e realizar a interpolação do mesmo. Como pode ser observado na Figura 1.a o primeiro modulo 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. A aplicação da interpolação faz com que a arquitetura tenha como saída 64 amostras dentre eles estão as oito interiras inteiros e os demais fracionários.

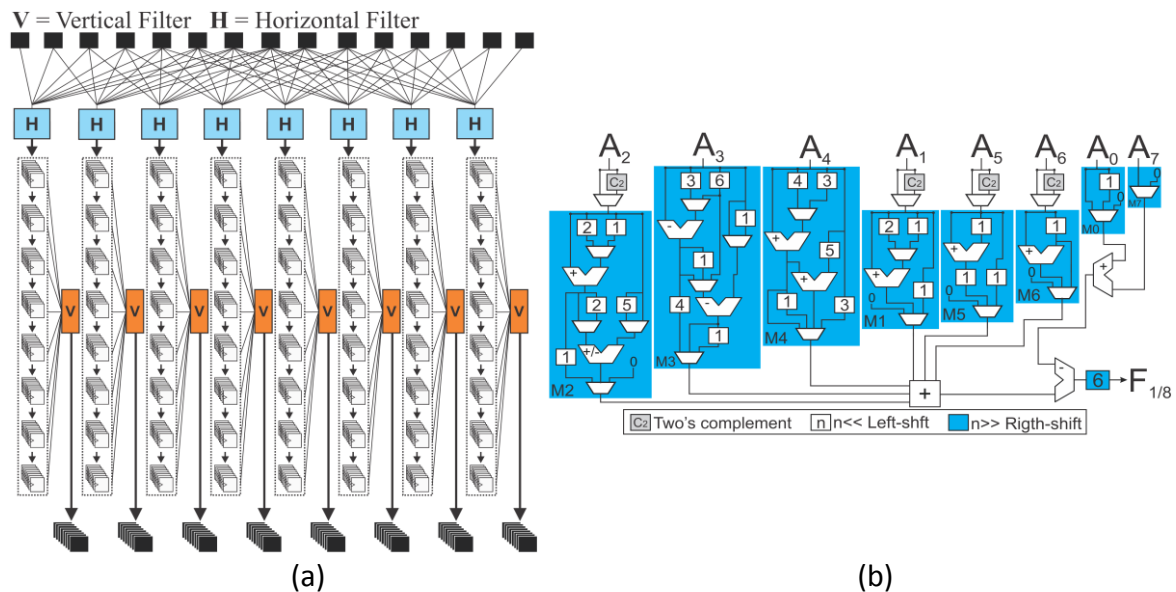


Figura 1: a) arquitetura do Interpolados desenvolvido, b) Um dos setes filtros otimizados desenvolvidos.

Os blocos de filtro H e V destacados na Figura 1.a apresentam a mesma configuração interna, sendo a única diferença a largura de bits. Estes módulos têm como saída 8 amostras, sendo que a primeira saída é um bypass para a terceira amostra da entrada e as outras sete saídas estão conectadas aos seus respectivos filtros otimizados de acordo com a precisão (1/8, 1/4, 3/8, 1/2, 5/8, 3/4 e 7/8). Na Figura 1.b é apresentado o filtro de 1/8 de precisão que é um dos sete filtros internos que compõem os módulos H e V. Como pode ser percebido as multiplicações existentes foram substituídas por soma e deslocamento. Vale destacar também que foi realizada um compartilhamento de expressão para que estes módulos comportassem todos os filtros de cada família.

3. RESULTADOS E DISCUSSÃO

A arquitetura desenvolvida foi um Interpolador para o modulo da FME para o processamento das amostras de luminancia para o padrão VP9. A arquitetura foi projetada para operar sobre blocos com o tamanho 8x8. Ela foi descrita em verilog HDL e sintetizada para ASIC utilizando a ferramenta RTL compiler da Cadence e a biblioteca de Sandard-Cell utilizada na síntese foi Nangate 45 nm (nangate, 2016) com uma tensão de alimentação de 0,95V.

Na Tabela 2 é apresentado 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 398,1 MHz. Considerando esta frequência e que a arquitetura precisa processar os quatros tamanhos de quadrados mais selecionados e as quatros famílias de filtros, ela é

capaz de processar vídeos com a resolução 1024x720 pixels. Nos resultados de sínteses obtidos podemos constatar que a arquitetura usou aproximadamente 250,3 K gates (199755 μm^2) em área, tendo apresentado uma potência dissipada na faixa de 123,3 mW.

Tabela 2 Resultados de síntese ASIC para a arquitetura desenvolvida.

| Parametros | Arquitetura Desenvolvida |
|-----------------------------------|--------------------------|
| Ciclo para processar um Bloco 8x8 | 72 |
| Área | 250,3 K gates |
| Tecnologia | 45 nm |
| Frequência | 398,1 MHz |
| Potência total | 123,3 mW |

4. CONCLUSÕES

Este trabalho desenvolveu uma solução arquitetural para a interpolação do padrão VP9 de codificação de vídeo. A arquitetura de Interpolação desenvolvida tem suporte a 28 filtros, que estão distribuídos entre quatro famílias: *Lagrange-Based*, *DCT-Based Sharp*, *Low-Pass Smooth* e *Bi-linear*. Os filtros foram otimizados para permitir a utilização de soma e deslocamento em vez de multiplicadores. Com o intuito de reduzir o esforço computacional a arquitetura foi projetada para processar os quatro maiores tamanhos blocos que representam uma redução de 26% no tempo de codificação e uma perda em BD-rate de 2,94%.

O Interpolador desenvolvido deve processar os quatro tamanhos de bloco e as quatro famílias de filtros, desta forma ela tem a capacidade de processar vídeos com a resolução 1024x720 pixels dissipando a potência total de 123,3 mW.

5. REFERÊNCIAS BIBLIOGRÁFICAS

- Agostini, L. V. (2007) '**Desenvolvimento de Arquiteturas de Alto Desempenho Dedicadas à Compressão de Vídeo Segundo o Padrão H.264/AVC**', (Tese (Doutorado em Ciência da Computação) – Instituto de Informática, UFRGS, Porto Alegre), p. 172.
- Mukherjee, D. *et al.* (2015) '**A Technical Overview of VP9—The Latest Open-Source Video Codec**', *SMPTE Motion Imaging Journal*, pp. 44–54. doi: 10.5594/j18499.
- nangate (2016) *NanGate - The Standard Cell Library Optimization Company*. Available at: http://www.nangate.com/?page_id=22 (Accessed: 25 September 2017).
- Paim, G. *et al.* (2016) '**An efficient sub-sample interpolator hardware for VP9-10 standards**', in *2016 IEEE International Conference on Image Processing (ICIP)*, pp. 2167–2171. doi: 10.1109/ICIP.2016.7532742.
- Sullivan, G. J. *et al.* (2012) '**Overview of the high efficiency video coding (HEVC) standard**', *IEEE Transactions on Circuits and Systems for Video Technology*, 22(12), pp. 1649–1668. doi: 10.1109/TCSVT.2012.2221191.