

ARQUITETURA EM HARDWARE PARA INTERPOLAÇÃO E ESTIMAÇÃO DE MOVIMENTO PARA HALF-PIXELS SEGUNDO O PADRÃO H.264/AVC

CORRÊA, Marcel; SCHOENKNECHT, Mateus

Grupo de Arquiteturas e Circuitos Integrados
Universidade Federal de Pelotas

AGOSTINI, Luciano

Universidade Federal de Pelotas

1 – INTRODUÇÃO

A compressão de vídeos é uma técnica utilizada para tornar possível o armazenamento e a transmissão de vídeos digitais de alta qualidade os quais, sem este processo, utilizariam uma grande quantidade de informação para serem representados. O padrão H.264/AVC (ITU-T, 2003) é o mais novo padrão de compressão de vídeos digitais e foi definido por especialistas da ITU-T e ISO/IEC com o intuito de obter taxas de compressão duas vezes maiores que os padrões já existentes, como o MPEG-2.

Esse padrão também é o que possui maior complexidade computacional entre todos os padrões existentes (RICHARDSON, 2003). Devido a esta alta complexidade, soluções para codificadores H.264/AVC que atinjam processamento de tempo real (30 quadros por segundo) para vídeos de alta resolução se tornam inviáveis por meio de software, incentivando o desenvolvimento de soluções em hardware.

O H.264/AVC é o padrão adotado pelo Sistema Brasileiro de TV Digital (SBTVD) e, para sua implementação, diversos pesquisadores brasileiros estudam soluções arquiteturais para codificadores e decodificadores H.264/AVC.

A ME é o módulo responsável por explorar a redundância temporal de um vídeo, ou seja, reduzir a necessidade de se armazenar informações semelhantes entre quadros próximos no tempo. Conforme ilustrado na Figura 1, a ME utiliza partições do quadro atual chamadas macroblocos (16x16 pixels), e para cada um destes macroblocos, é realizada uma busca dentro de uma área limitada do quadro de referência pelo bloco que mais se assemelha ao bloco atual. Uma vez que este melhor casamento (*match*) é encontrado, um vetor de movimento é gerado para representar o deslocamento do bloco no quadro de referência (RICHARDSON, 2003). O melhor casamento é calculado de acordo com um critério de similaridade previamente definido, neste trabalho será usado o SAD (*Sum of Absolute Differences*).

Uma característica que aumenta drasticamente as taxas de compressão atingidas pela ME é a possibilidade de geração de vetores de movimento fracionários, ou seja, vetores que apontam para posições não inteiras de pixel, que não estão disponíveis diretamente nos quadros de referência. Os blocos apontados por estes vetores fracionários deverão ter suas amostras obtidas por um processo de interpolação de sub-pixels. O objetivo deste trabalho é o desenvolvimento de uma solução arquitetural para a interpolação e a estimação de movimento (ME) para half-pixels (meio pixel).

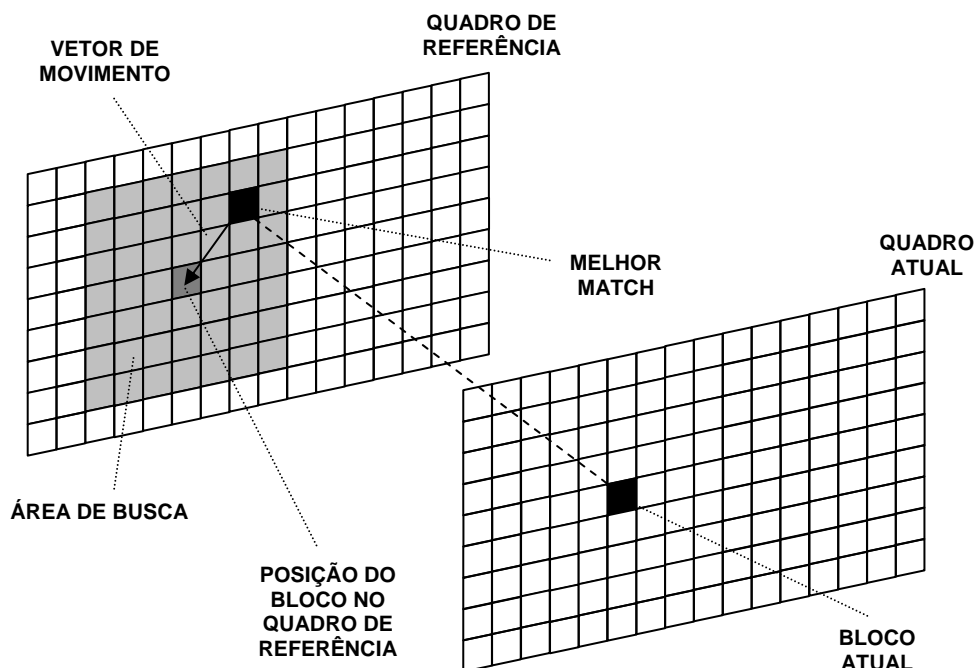


Figura 1 – Processo de Estimação de Movimento.

2 – METODOLOGIA

O desenvolvimento da arquitetura proposta foi dividido em dois passos: (1) o desenvolvimento da unidade de interpolação, e (2) o desenvolvimento de uma unidade de busca e comparação.

A unidade de interpolação de half-pixels é o módulo responsável pela obtenção das amostras de half-pixel através da interpolação das amostras de posições inteiras. O módulo desenvolvido recebe um bloco de 8x8 amostras de luminância, mais uma borda de 3 amostras ao redor deste bloco, e partir desta entrada são gerados 8 novos blocos 8x8 compostos por amostras de half-pixel.

O processo de interpolação é realizado através do deslocamento do bloco de entrada de forma a enviar os valores de uma linha inteira para as unidades básicas de processamento (UP). Cada uma destas UPs geram um half-pixel y com valor limitado entre 0 e 255 através das seis amostras mais próximas e alinhadas de posições inteiras (A-F). As equações otimizadas para implementação em hardware que descrevem o cálculo realizado pelas UPs estão definidas em (1), (2) e (3).

$$y_1 = (A + F) + 4[4(C + D) - (B + E)] + 4(C + D) - (B + E) \quad (1)$$

$$y = Clip_{0-255}[(y_1 + 16) \gg 5] \quad (2)$$

$$y = Clip_{0-255}[(y_1 + 512) \gg 10] \quad (3)$$

A unidade de busca e comparação é o módulo que compara os 8 novos blocos (possíveis casamentos) gerados pela unidade de interpolação com o bloco atual. O objetivo disso é verificar se existe entre estes blocos compostos por half-pixels um casamento melhor do que aquele encontrado pela ME para posições inteiras, e gerar um vetor de movimento fracionário caso um melhor casamento seja encontrado. Este processo é chamado refinamento da busca para half-pixel.

Este módulo é composto basicamente pela Árvore de SAD, que é responsável por calcular e quantificar a diferença entre o bloco candidato e o bloco atual, e por um Comparador de SAD, que é responsável por selecionar o bloco mais semelhante ao atual, ou seja, aquele com o menor SAD dentre os oito novos possíveis casamentos e então comparar este com o SAD do melhor casamento selecionado pela ME para posições inteiras.

A Figura 2 ilustra a arquitetura completa para a interpolação e busca de half-pixels. São mostrados os módulos explicados anteriormente e também os *buffers* que são utilizados para armazenar entradas e resultados gerados.

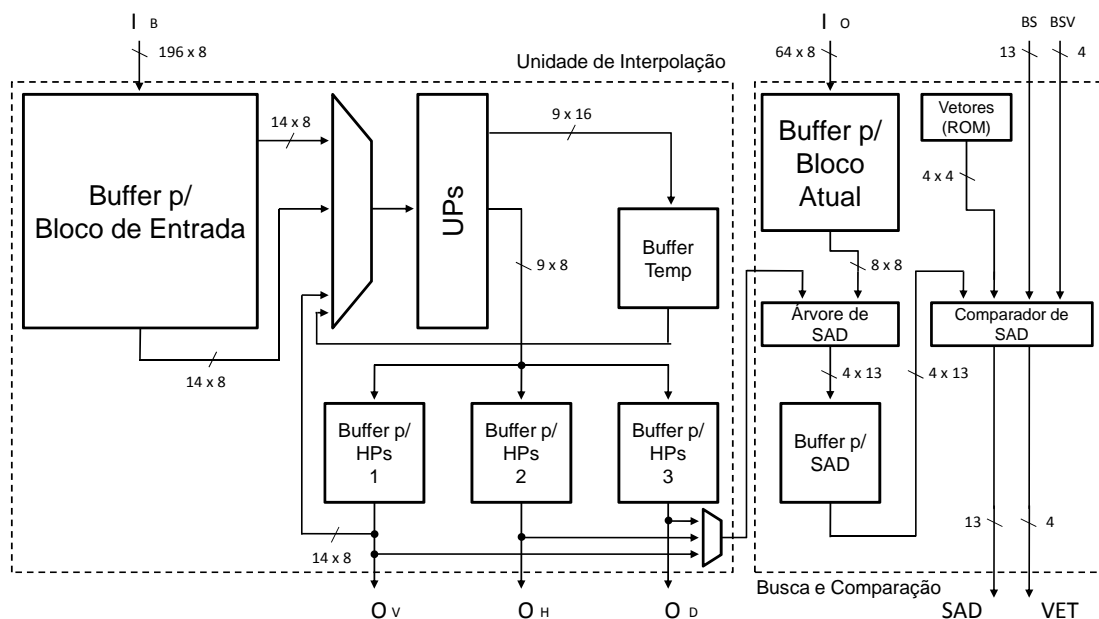


Figura 2 – Diagrama da Arquitetura Completa.

3 – RESULTADOS E DISCUSSÕES

Um dos principais objetivos deste trabalho é aliar melhorias na qualidade de compressão de vídeos digitais com baixo custo em recursos de hardware, e para atingir este objetivo, algumas decisões importantes foram tomadas ao longo do projeto. Com base em avaliações realizadas em software, foi percebido que desenvolver uma arquitetura de ME fracionária que de suporte apenas a tamanho de bloco 8x8 ao invés de dar suporte a tamanho de blocos variados conforme previsto no padrão H.264/AVC (sete ao todo) (ITU-T, 2003) reduz a complexidade computacional e o custo para implementação em hardware drasticamente.

As avaliações foram feitas através da codificação de um total de cinco vídeos QCIF (176x144 pixels). Para isso foi utilizado o software de referência do padrão H.264/AVC e os resultados foram avaliados com base no *bit rate* (taxa de bits) para taxa de compressão e no PSNR (*Peak Signal-to-Noise Ratio*) para qualidade visual.

As Tabelas 1 e 2 mostram os resultados obtidos respectivamente para o caso ótimo (todos tamanhos de bloco) e para o caso escolhido (apenas tamanho 8x8). A configuração de codificação utilizada foi: algoritmo de busca *Full Search*, área de busca de tamanho 32x32, um único quadro de referência, QP com valor 28 e precisão de quarter-pixel. Quando comparado com o caso ótimo, o uso de tamanho de bloco fixado em 8x8 em média aumentou o *bit rate* em 3,84% e reduziu o PSNR

em 0.32 dB, e isso não é uma perda expressiva quando a redução no custo da arquitetura é levado em consideração.

Tabela 1 – Resultados de codificação para tamanhos de bloco variável.

Sequência	PSNR (dB)	Bit rate (Kbps)
Akiyo	40.82	114.5
Bridge Close	37.38	155.33
Coastguard	34.12	272.33
Foreman	36.47	207.87
Mobile	32.59	391.81
Média	36.28	228.37

Tabela 2 – Resultados de codificação para tamanho de bloco fixo (8x8).

Sequência	PSNR (dB)	Bit rate (Kbps)
Akiyo	40.47	116.53
Bridge Close	37.16	154.56
Coastguard	33.80	277.21
Foreman	36.01	225.24
Mobile	32.34	412.16
Média	35.96	237.14

A arquitetura proposta foi totalmente descrita em VHDL e sintetizada para um dispositivo FPGA Xilinx Virtex4 com o auxílio da ferramenta de síntese Xilinx ISE.

Em termos de desempenho, o *design* proposto utiliza 37 ciclos de *clock* para realizar todo o processo em um bloco 8x8, e quando sintetizado ao dispositivo FPGA alvo, atinge uma frequência máxima de operação de 140 MHz, podendo processar até 3,784 milhões de vetores de movimento por segundo, ou mais precisamente: 262 quadros HD (1280x720) por segundo, 116 quadros *Full HD* (1920x1080) por segundo ou 30 quadros QHDTV (3840x2048) por segundo.

Em termos de custo, a parte operativa da arquitetura utilizou os seguintes recursos do FPGA alvo: 6488 *slices*, 708 *slice flip-flops* e 4181 LUTs. Já os buffers utilizaram os seguintes elementos de hardware: 539 registradores de 8-bits, 10 registradores de 13-bits, 108 registradores de 16-bits, 386 multiplexadores 4:1 de 8-bits, 153 multiplexadores 2:1 de 8-bits, e 108 multiplexadores 2:1 de 16-bits.

4 – CONCLUSÕES

Este trabalho apresentou o desenvolvimento de uma arquitetura em hardware para a interpolação e o refinamento da busca realizada na estimação de movimento considerando uma precisão de half-pixel. Os resultados obtidos foram bastante satisfatórios tendo como objetivo uma boa relação entre custo e desempenho. Quando sintetizada para o FPGA alvo, a arquitetura se mostrou capaz de processar vídeos de altíssima resolução (QHDTV) em tempo real (30 quadros por segundo ou mais rápido).

Como próximos trabalhos, uma arquitetura para interpolação e busca com precisão de um quarto de pixel já está sendo desenvolvida pelo grupo.

5 – REFERÊNCIAS

- JOINT VIDEO TEAM OF ITU-T AND ISO/IEC JTC 1. Draft ITU-T Recommendation and Final Draft International Standard of Joint Video Specification (ITU-T Rec. H.264 or ISO/IEC 14496-10 AVC), 2003.
- RICHARDSON, I. **H.264 and MPEG-4 Video Compression: Video Coding for the Next-generation Multimedia**. Chichester: Wiley, 2003.