

PROJETO DE HARDWARE PARA A ESTIMAÇÃO DE MOVIMENTO DO PADRÃO H.264/AVC COM MÚLTIPLOS QUADROS DE REFERÊNCIA E COM REDUÇÃO NOS ACESSOS À MEMÓRIA

GRELLERT, Mateus; SAMPAIO, Felipe M.; HECKTHEUER, Bruno B.
Grupo de Arquiteturas e Circuitos Integrados - GACI
Universidade Federal de Pelotas – UFPel

AGOSTINI, Luciano; MATTOS, Julio C. B
Grupo de Arquiteturas e Circuitos Integrados - GACI
Universidade Federal de Pelotas – UFPel

1 INTRODUÇÃO

O H.264/AVC é o mais novo padrão de codificação de vídeo. Desenvolvido por especialistas da ITU-T e ISO-IEC, esse padrão é capaz de dobrar as taxas de compressão quando comparado a padrões anteriores, como o MPEG-2. O ganho em compressão é proveniente de novas funcionalidades introduzidas pelo H.264/AVC, principalmente no que diz respeito à Estimação de Movimento (ME – *Motion Estimation*) (RICHARDSON, 2003). No entanto, a adição dessas funcionalidades causa um aumento considerável na complexidade computacional, o que dificulta extremamente o desenvolvimento de aplicações em software capazes de atingir processamento em tempo real de vídeos em alta resolução. Nesse contexto, soluções em hardware representam um papel importante na codificação de vídeo, especialmente quando se trata de processamento em tempo real (24 a 30 quadros por segundos) de vídeos em alta definição (HD – *High Definition*).

A Estimação de Movimento é a etapa da codificação que consiste, basicamente, em encontrar o bloco mais semelhante de um quadro previamente codificado (na terminologia de codificação de vídeo, quadro de referência) a um bloco do quadro atual (bloco atual). Para tanto, a ME se vale de uma área de busca, de um algoritmo que define como essa área será percorrida e de um critério de similaridade para avaliar o nível de semelhança entre blocos. A ME é um importante módulo do codificador H.264/AVC e contribui muito para os ganhos em compressão desse padrão quando relacionado a padrões anteriores, mas é também um dos grandes responsáveis pelo aumento na complexidade do codificador. A complexidade da ME, por sua vez, é ditada principalmente pelo algoritmo de busca utilizado.

Dentre os algoritmos de busca mais conhecidos, o *Full Search* (FS) usa um método de busca exaustivo e é compatível com todos os padrões de codificação de vídeo, incluindo o H.264/AVC. A comparação entre os diferentes blocos é realizada a partir de uma métrica de distorção, e a área de busca é percorrida pixel a pixel de forma linear.

Para cada bloco do quadro atual, a ME produz um vetor de movimento, que é interpretado como o deslocamento, dentro da área de busca, do bloco escolhido para representar o bloco atual. Quando múltiplos quadros são utilizados, a ME percorre múltiplas áreas de busca, aumentando consideravelmente a complexidade computacional desse procedimento. A principal justificativa para o uso de vários quadros de referência encontra-se no

fato de que o melhor casamento para um bloco atual não se encontra necessariamente no quadro previamente codificado; ele pode estar em quadros temporalmente anteriores. Além disso, o melhor casamento pode estar em quadros futuros, considerando a ordem de exibição. Isso é possível porque o H.264/AVC suporta processamento de quadros fora de ordem (RICHARDSON, 2003).

Esse trabalho apresenta uma arquitetura para a Estimação de Movimento com múltiplos quadros de referência (MRF-ME), utilizando o FS como algoritmo de busca, quatro quadros de referência e uma estratégia para o reuso de dados. A arquitetura processa quatro blocos atuais em paralelo e utiliza um esquema de reuso de dados que diminui os acessos à memória principal. Essa arquitetura foca o padrão H.264/AVC convencional, mas está sendo utilizada para o desenvolvimento de uma solução compatível com a extensão desse padrão para múltiplas vistas (MVC – *Multiview Video Coding*) (J. V. TEAM, 2009).

2 METODOLOGIA

Inicialmente, foram estudadas diferentes arquiteturas capazes de processar a MRF-ME mantendo a melhor qualidade de vídeo possível. A partir das arquiteturas projetadas, diversas avaliações em software (desenvolvido na linguagem Ruby) foram realizadas, com o intuito de avaliar o número de cálculos, juntamente com a qualidade atingida. Para avaliar a qualidade, foi utilizada uma métrica chamada *Peak Signal-to-Noise Ratio* (PSNR), a qual é amplamente utilizada na área de codificação de vídeo. O PSNR é medido em decibéis (dB), e quanto maior for o seu valor, maior será a qualidade do vídeo (RICHARDSON, 2003). Outra consideração importante é que o acréscimo de um dB no valor de PSNR representa um aumento considerável na qualidade do vídeo.

Após os testes em software, iniciou-se o desenvolvimento da arquitetura na linguagem VHDL, uma linguagem de descrição de hardware amplamente utilizada no meio acadêmico e industrial. A arquitetura é baseada no núcleo ME da arquitetura proposta em (DORNELLES, 2010), a qual realiza a ME para um quadro de referência, além de oferecer suporte a outras funcionalidades importantes e integrar a Compensação de Movimento, outro módulo importante do codificador. A descrição VHDL da arquitetura desenvolvida foi sintetizada utilizando a ferramenta ISE 10.1 da Xilinx.

A arquitetura desenvolvida é composta principalmente por cinco módulos: (1) um banco de registradores local, (2) unidades de processamento, (3) buffers locais para dados dos blocos atuais assim como outros propósitos de armazenamento, (4) comparadores para decidir qual bloco representa o melhor casamento entre os candidatos e (5) uma unidade de controle que gerencia o controle de processamento com sinais de controle. A Figura 1 apresenta o diagrama de blocos para essa arquitetura.

Uma importante característica do algoritmo FS é a área de busca comum entre blocos atuais vizinhos. Essa área é determinada de acordo com a posição desse bloco no quadro, portanto blocos próximos possuem regiões comuns entre as suas áreas de busca. Com base nesta constatação, esse trabalho aplica um esquema de reuso de dados inspirado em (CHEN, 2006). Considerando quatro blocos atuais vizinhos (numerados 0,1, 2 e 3), as áreas de busca comuns entre esses blocos são ilustradas na Figura 2.

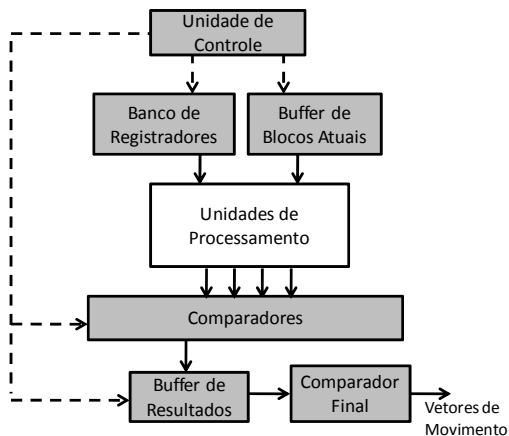


Figura 1. Diagrama de Blocos da arquitetura

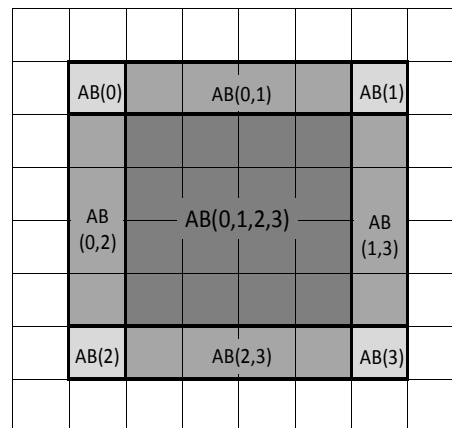


Figura 2. Áreas de Busca de blocos vizinhos.

3 RESULTADOS E DISCUSSÕES

Através dos testes em software, foi possível observar que, no melhor resultado, o suporte a múltiplos quadros gera um aumento de 2,1 dB, quando comparado à ME com um único quadro de referência. No pior, o ganho é de 1,1 dB, o qual também é significativo. Por conseguinte, o uso de vários quadros de referência gera ganhos importantes na qualidade do vídeo.

Os resultados de síntese para a arquitetura desenvolvida são apresentados na Tabela 1. A síntese foi direcionada para o dispositivo FPGA XC2VP30 da família VirtexII-Pro da Xilinx. Como esperado, o custo em uso de hardware é elevado, já que essa arquitetura foi projetada para codificar vídeos *Full HD* em tempo real, mas é importante salientar a diminuição considerável no número de acessos à memória, o que justifica o investimento em recursos de hardware. Com essa técnica, utilizando uma área de busca de 19x19 amostras, a solução tradicional acessaria 1.444 amostras da memória principal do codificador. No entanto, utilizando o esquema proposto, somente 529 amostras serão acessadas, gerando uma redução de 63,7% no número de acessos. Essa é a principal contribuição desta arquitetura, já que o acesso à memória é um ponto crítico em codificadores de vídeo.

Tabela 1. Resultados de Síntese para FPGA da Xilinx

Módulo	Frequência (MHz)	Número de ALUTs	Número de Slices	Número de Slice Flip-Flops
UP	371,61	450	286	292
Banco Reg.	591,17	2.944	1.472	1.472
Comparador	266,75	224	125	170
Buff. Blocos Atuais	1045,75	2	294	512
MRF-ME	261,4	12.177	7.126	9.785

A Tabela 2 apresenta a taxa de processamento atingida (em quadros por segundo - qps), bem como a frequência mínima para diferentes resoluções. Como pode ser observado na Tabela 2, a arquitetura desenvolvida é capaz de codificar 70 qps e 31 qps, considerando as resoluções 720p HD e 1080p HD (também chamada de *Full HD*) respectivamente. Mais que isso, é possível atingir tempo real de vídeos SD com uma frequência de somente 42 MHz.

Tabela 2. Taxa de processamento e frequência mínima

Resolução	Quadros/segundo	Frequência Mínima (MHz)
CIF (352x288)	644,6	12,2
VGA (640x480)	212,7	36,9
SD (720x480)	189,1	41,5
720p HD (1280x720)	70,9	110,6
1080p HD (1920x1080)	31,5	248,8

A elevada taxa de processamento alcançada por essa arquitetura é importante quando aplicações que demandam baixo consumo de energia e baixa resolução (por exemplo, para celulares) são levadas em consideração, já que é possível atingir tempo real com frequências baixas, consumindo, portanto, pouca energia. Ademais, é importante notar que a frequência mínima é independente da tecnologia utilizada na síntese, tornando esses dados importantes para futuras comparações com diferentes arquiteturas.

4 CONCLUSÕES E TRABALHOS FUTUROS

Esse trabalho apresentou uma arquitetura para a Estimação de Movimento utilizando quatro quadros de referência. Para reduzir acessos à memória, foi proposto um esquema de reuso de dados, permitindo uma redução de aproximadamente 63%, e essa é a principal contribuição desse trabalho. Para processar o volume de dados requerido por vídeos *Full HD*, o casamento de quatro blocos atuais é realizado em paralelo.

Em trabalhos futuros, pretende-se utilizar essa arquitetura em uma ME para a codificação multivisão (MVC), que é uma extensão do padrão H.264/AVC utilizada em aplicações como TV 3-D. Para tanto, a arquitetura apresentada necessita ainda de pequenos ajustes para ser totalmente compatível com esse padrão. A codificação multivisão é o objetivo final dessa investigação.

5 REFERÊNCIAS

RICHARDSON, Iain. **H.264 and MPEG-4 Video Compression – Video Coding for Next Generation Multimedia**. Chichester: John Wiley&Sons, 2003.

J. V. TEAM, **Draft Revised ITU-T Recommendation H.264 Advanced Video Coding for Generic Audiovisual Services**. Doc JVT-AA007, 2009.

CHEN, C.-H., HUANG, C.-T., CHEN, Y.-H., CHEN, L.-G.. + Data Reuse Scheme for Motion Estimation With Corresponding Coding Orders. **IEEE Trans. Circuits and Systems for Video Tech.**, Local de Edição, v. 16, n. 4, p. 553 - 558, 2006.

DORNELLES, Robson, SAMPAIO, Felipe, AGOSTINI, Luciano. Variable Block Size Motion Estimation Architecture with a Fast Bottom-Up Decision Mode and an Integrated Motion Compensation Targeting the H.264/AVC Video Coding Standard. In: **23rd SYMPOSIUM ON INTEGRATED CIRCUITS AND SYSTEM DESIGN**, São Paulo, setembro de 2010. (aceito para publicação)