

# DESENVOLVIMENTO DE HEURÍSTICA BASEADA EM HOMOGENEIDADE E DISTORÇÃO PARA A DECISÃO DE MODO DE CODIFICAÇÃO DA PREDIÇÃO INTRA DO PADRÃO H.264/AVC

PALOMINO, Daniel; AGOSTINI, Luciano  
GACI – Grupo de Arquiteturas e Circuitos Integrados  
Universidade Federal de Pelotas

CORRÊA, Guilherme  
Universidade Federal do Rio Grande do Sul

## 1 INTRODUÇÃO

A codificação de vídeo é de extrema importância nos dias atuais, tendo em vista a grande quantidade de informação presente em vídeos de alta resolução e a quantidade de aparelhos eletrônicos que manipulam este tipo de mídia, tais como computador pessoal e portátil, aparelho celular, televisão digital de alta resolução, *Blue-ray players*, entre outros.

O padrão H.264/AVC é o estado de arte em compressão de vídeo e foi desenvolvido por especialistas da ITU-T e ISO-IEC (JVT, 2003), com o objetivo de dobrar as taxas de compressão quando comparado com os padrões anteriores, como MPEG-2 (RICHARDSON, 2003).

Para atingir tal objetivo, o padrão H.264/AVC reuniu várias ferramentas eficientes de compressão. Entre estas novas ferramentas, é possível destacar a inovação nos módulos de predição presente em codificadores H.264/AVC. O macrobloco é a unidade de trabalho utilizada em codificadores H.264/AVC, e é formado por 16x16 amostras de luminância, 8x8 amostras de crominância azul e 8x8 amostras de crominância vermelha.

A predição intra-quadro, que é uma inovação do padrão H.264/AVC, é responsável por reduzir a redundância espacial presente em um mesmo quadro, explorando a similaridade entre amostras vizinhas. A predição sobre um bloco é realizada através de uma cópia dos pixels vizinhos, previamente codificados, para o bloco que está sendo codificado. Considerando as amostras de luminância, a predição pode ser realizada sobre todo o bloco 16x16 (116MB) ou sobre os 16 blocos 4x4 (14MB) que compõem o bloco 16x16. Existem 13 modos de cópia possíveis quando a predição intra-quadro é considerada: quatro quando a predição é 116MB e nove para 14MB. Além disso, são utilizados os mesmos quatro modos que são usados para os blocos 16x16 de luminância, para realizar a predição sobre os blocos 8x8 de crominância.

A predição inter-quadros é responsável por reduzir a redundância temporal presente em vídeos digitais. A predição sobre um bloco é gerada a partir de informação de blocos presentes em quadros previamente codificados. A grande inovação no módulo de predição inter-quadros é a possibilidade de usar vários tamanhos de bloco (16x16, 16x8, 8x16, 8x8, 8x4, 4x8 e 4x4) para realizar a predição sobre um bloco, além da possibilidade de utilizar o *skip-mode*, quando a informação do bloco anterior é utilizada para o bloco atual, não sendo necessário nenhum tipo de informação extra.

O módulo de decisão de modo é o responsável por escolher, dentre todos os modos de predição intra-quadro e inter-quadros, o que melhor se ajusta

ao tipo de bloco que está sendo codificado. Entretanto, devido ao grande número de possibilidades de predição sobre um bloco, a tarefa de escolher o melhor modo não é trivial. O RDO (*Rate-Distortion Optimization*) (SULLIVAN, 1998) é a técnica que realiza a escolha ótima para a relação entre taxa de bits do vídeo codificado e distorção. A equação (1) mostra como é calculado o custo RDO.

$$J = D + \lambda \cdot R \quad (1)$$

O  $D$  é o valor relacionado à distorção do bloco codificado. O  $R$  é a quantidade de bits necessária para representar o bloco após todo o processo de codificação. O parâmetro  $\lambda$  é chamado de fator de Lagrange e serve para relacionar a quantidade de bits com a distorção.  $J$  é o custo distorção-quantidade de bits.

Essa técnica possui uma elevada complexidade computacional, visto que é necessário realizar todo o processo de compressão (transformadas diretas, quantização direta e codificação de entropia) sobre todos os blocos considerando todos os modos, antes de realizar a decisão.

O objetivo deste trabalho é o desenvolvimento de uma heurística rápida baseada em distorção e homogeneidade para o módulo de decisão de modo do padrão H.264/AVC, focando na decisão entre blocos intra-quadro.

## 2 METODOLOGIA

Neste trabalho, o processo de decisão para a predição intra-quadro foi dividido nos seguintes passos:

- Decisão entre o melhor bloco 4x4 de luminância (I4MB);
- Decisão entre o melhor bloco 16x16 de luminância (I16MB);
- Decisão entre o tamanho de bloco que será utilizado (I4MB ou I16MB).

A decisão entre blocos de mesmo tamanho, ou seja, entre os nove modos para I4MB e entre os quatro modos para I16MB, foi realizada usando um simples critério de similaridade, conhecido como SAD (*Sum of Absolute Differences*), que avalia a diferença entre o bloco original e o bloco predito. O bloco candidato com o menor SAD é o escolhido. Existem outros critérios, como SATD (*Sum of Absolute Transformed Differences*) e SSD (*Sum of Squared Differences*), que não foram utilizados, pois o SAD possui uma complexidade computacional menor. A equação (2) mostra como é realizado o cálculo do SAD. O  $R$  é amostra na posição  $(i,j)$  do bloco reconstruído enquanto o  $O$  na posição  $(i,j)$  é a amostra do bloco original. Os valores de  $m$  e  $n$  são referentes à resolução do quadro.

$$SAD = \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} |R_{i,j} - O_{i,j}| \quad (2)$$

A decisão entre os dois tamanhos de bloco foi realizada utilizando uma métrica de homogeneidade do bloco 16x16, ou seja, para blocos mais homogêneos (amostras que possuem valores próximos) é interessante utilizar o modo I16MB, enquanto que para blocos que possuem mais detalhes, o modo I4MB é o mais recomendado. A aplicação de uma transformada bi-dimensional sobre os resíduos gerados pelas etapas de predição permite uma avaliação da homogeneidade de um bloco no domínio das frequências. Um bloco heterogêneo

tende a gerar coeficientes altos nas regiões de baixa frequência, enquanto um bloco homogêneo tende a gerar valores mais baixos. Sendo assim, foi utilizada uma transformada DCT (*Discrete Cosine Transform*) de tamanho 16x16 para gerar a métrica de homogeneidade. Os valores nas regiões de baixa frequência (borda superior e borda esquerda) foram acumulados para gerar a métrica que define o nível de homogeneidade ( $AC_{low}$ ). Blocos que geram um valor abaixo do valor de  $AC_{low}$  (definido em simulação) são classificados como blocos homogêneos e, então, preditos utilizando o modo I16MB, enquanto que blocos que geram um valor acima de  $AC_{low}$  são preditos utilizando o modo I4MB.

As equações (3) e (4) mostram como é realizado o cálculo da DCT, onde  $\mathbf{A}$  é a matriz da DCT,  $\mathbf{A}^T$  é a matriz transposta da DCT e  $\mathbf{X}$  é a matriz de entrada.

$$Y = AXA^T \tag{3}$$

$$A_{ij} = C_i \cos \frac{(2j+1)i\pi}{2N}, \quad \text{where} \tag{4}$$

$$C_i = \sqrt{\frac{1}{N}} \text{ if } (i=0) \quad \text{and} \quad C_i = \sqrt{\frac{2}{N}} \text{ if } (i > 0)$$

A matriz  $\mathbf{A}$  resultante possui 256 elementos diferentes; contudo, como a função cosseno se repete a cada  $2\pi$ , os valores da matriz  $\mathbf{A}$  foram reduzidos ao primeiro quadrante do círculo trigonométrico, gerando 16 valores diferentes. Além disso, com o objetivo de simplificar o cálculo da DCT, todos os valores gerados foram arredondados para que o cálculo da transformada DCT fosse realizado somente sobre números inteiros. A figura 1 apresenta o diagrama em blocos do módulo de decisão completo para a decisão intra-quadro.

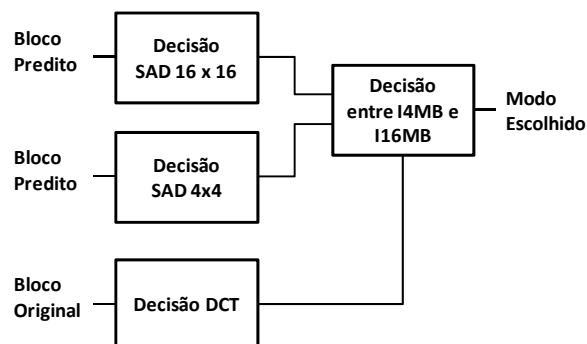


Figura 1 – Diagrama em blocos do módulo de decisão.

### 3 RESULTADOS E DISCUSSÕES

A heurística proposta neste trabalho foi desenvolvida e integrada com o software de referência do padrão H.264/AVC (VCEG, 2010), com o objetivo de avaliar os resultados de qualidade e quantidade de bits do vídeo codificado. Foram utilizadas várias sequências de vídeos CIF (352x288 pixels), para avaliar o desempenho da heurística proposta. A tabela 1 mostra os resultados destas avaliações para cada vídeo e compara com a técnica RDO. Os resultados são apresentados considerando quantidade de bits e qualidade, que é medida com o PSNR (*Peak Signal-to-Noise Ratio*).

Tabela 1 – Resultados da heurística proposta

Seqüência	Decisão RDO		Decisão SAD & DCT		RDO x SAD & DCT	
	Taxa de bits	PSNR	Taxa de bits	PSNR	+ Taxa de Bits (%)	+ PSNR (dB)
<i>BUS</i>	1.369.328	36,39	1.417.912	36,04	3,55	-0,35
<i>CITY</i>	1.193.560	36,24	1.241.008	35,87	3,98	-0,37
<i>CREW</i>	491.760	39,14	534.240	38,91	8,64	-0,23
<i>FOOTBALL</i>	1.380.760	36,22	1.426.184	35,85	3,29	-0,37
<i>FOREMAN</i>	652.568	38,11	700.704	37,92	7,38	-0,19
<i>ICE</i>	500.320	40,45	533.760	40,28	6,68	-0,17
<i>MOBILE</i>	2.264.904	36,27	2.332.496	35,88	2,98	-0,39
<i>SOCCER</i>	1.115.384	35,95	1.137.064	35,64	1,94	-0,31
<b>MÉDIA</b>	<b>1.121.073</b>	<b>37,35</b>	<b>1.165.421</b>	<b>37,05</b>	<b>3,96</b>	<b>-0,30</b>

Os resultados mostram que a utilização da heurística proposta neste trabalho, causa um pequeno aumento de 3,96% em média na quantidade de bits do vídeo codificado em relação ao uso da técnica RDO. O PSNR diminui em 0,3dBs em média quando comparado ao RDO. Entretanto, o decréscimo na complexidade computacional do módulo de decisão proposto neste trabalho é muito alto quando comparado à redução de qualidade e o aumento da taxa de bits. Segundo estimativas iniciais, a redução de complexidade com o uso da heurística é da ordem de 9 vezes se comparado ao RDO.

#### 4 CONCLUSÕES

Este trabalho apresentou o desenvolvimento de uma heurística baseada em distorção e homogeneidade para a decisão de modo de codificação do padrão H.264/AVC considerando a predição intra-quadro. Os resultados mostram um pequeno acréscimo na taxa de bits e uma pequena perda na qualidade (PSNR). Entretanto, a complexidade do módulo de decisão é drasticamente reduzida em comparação a técnica RDO, já que são utilizados o SAD como métrica para decisão entre mesmos tamanhos de blocos e uma transformada DCT 16x16 para decidir entre os dois tamanhos de bloco possíveis na decisão intra (I4MB e I16MB).

#### 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/AVC and MPEG-4 Video Compression – Video Coding for Next-Generation Multimedia**. Chichester :John Wiley and Sons, 2003.
- SULLIVAN, G.; WIEGAND, T. Rate Distortion Optimization for Video Compression. **IEEE Signal Processing Magazine**, v. 15, p. 74-90, 1998.
- VCEG. **JM Reference Software 16.2**. Disponível em <<http://iphome.hhi.de/suehring/tml/>>. Acesso em fev. 2010.