

# Codificação de Macrobloco Duplo-Inter no Padrão H.264

Edson T. M. Manoel, Marcos Moecke e Rui Seara

**Resumo**—Este artigo apresenta uma proposta de extensão (aprimoramento) do padrão de codificação de vídeo H.264 visando melhorar a taxa de compressão para alguns cenários. Tal aprimoramento explora o fato de que em alguns macroblocos (segmentos do sinal de vídeo) existem regiões com características que não são codificadas eficientemente pelo referido padrão. A avaliação do modo de codificação proposto é realizada alterando o modelo de referência (JM), incluindo o novo modo tanto no codificador quanto no decodificador. Os resultados obtidos indicam ganhos de desempenho de até 0,2 dB para algumas seqüências de vídeo comuns.

**Palavras-chave**—Estimação e Compensação de movimento, Codificação de vídeo, H.264, Televisão digital.

**Abstract**—This paper presents an extension to the H.264 video coding standard, aiming to improve its compression performance for some scenes. Such an extension exploits the fact that in some macroblock types (segments of the video signal) there exist regions with characteristics that are not coded efficiently by the referred video standard. The H.264 test model code (JM) has been modified (including changes in both coding and decoding modules) in order to assess the new coding modes. The obtained results indicate a performance gain of up to 0.2 dB for some common video sequences.

**Keywords**—Motion and estimation compensation, Video coding, H.264, Digital television.

## I. INTRODUÇÃO

A compressão de vídeo permite reduzir a quantidade de dados necessária para representar vídeos digitais, diminuindo os custos de transmissão e armazenamento. Nos dias de hoje, aplicações de vídeo digital estão bastante difundidas, seja para armazenamento e distribuição como também para difusão. Como aplicações específicas, destacamos a TV digital e a videoconferência. Como meio de propagação podemos incluir a radiodifusão, redes Internet, telefonia celular; para o armazenamento, temos as mídias magnéticas, mídias ópticas (CD, DVD, HD-DVD e Blu-ray), dentre outras.

O fator limitante das tecnologias e aplicações de vídeo digital é o custo de armazenamento e transmissão. Os sinais

de vídeo digital na forma bruta demandam alta capacidade de armazenamento e elevada taxa de bits para a transmissão. Dessa forma, para viabilizar uma grande maioria de aplicações, faz-se necessário o uso de técnicas eficientes de compressão, mantendo aceitável a qualidade de imagem dos vídeos digitais.

Os codificadores de vídeo baseados no padrão H.264 [1] têm se mostrado bastante eficientes para compressão de vídeo, apresentando diversas vantagens em relação a seus antecessores, tais como os padrões MPEG-2 [2], H.263 [3] e MPEG-4 ASP [4], nas etapas de predição espacial e temporal, no uso da transformada inteira no lugar da transformada do cosseno discreta (DCT) e também no emprego de diversos modos de codificação.

No entanto, em alguns cenários especiais, os modos de codificação do padrão H.264 não têm mostrado ser tão eficientes, resultando em macroblocos (MBs) que apresentam uma distorção muito superior à média de outros MBs, implicando, assim, em uma maior taxa de codificação do erro de predição. Diversos trabalhos de pesquisa, que buscam explorar outras características desses MBs, através de novas formas de predição espacial e temporal, têm sido apresentados visando melhorar o desempenho em taxa-distorção (R-D) do H.264. Em [5], um novo particionamento de bloco *inter*, denominado *sliced blocks*, foi proposto, conseguindo reduzir em até 5% a taxa de bits para algumas conhecidas seqüências de vídeo da literatura. Em [6], um outro tipo de particionamento de bloco, semelhante ao proposto em [5], é considerado. Tal estratégia divide os blocos usando retas inclinadas em qualquer ângulo, obtendo ganhos de até 0,3 dB em relação ao padrão H.264.

Neste trabalho é proposta uma modificação na codificação do tipo *inter* para o codificador H.264 com o objetivo de melhorar o desempenho R-D em situações nas quais apenas uma pequena parte do MB exibe um movimento distinto do restante do MB. A modificação em questão consiste em definir um novo tipo de particionamento de MB, no qual um bloco flutuante (com predição *inter*) é inserido no MB predito de modo *inter*. Tal modo de codificação é aqui denominado *duplo-inter*.

## II. CODIFICAÇÃO DE MACROBLOCOS NO H.264

A principal unidade de codificação no padrão H.264 é o MB, que consiste de um bloco de  $16 \times 16$ , o qual é composto de  $16 \times 16$  pixels de luminância e de dois blocos de  $8 \times 8$  pixels de crominância, quando usada uma subamostragem de

Edson T. M. Manoel, Rui Seara, LINSE - Laboratório de Circuitos e Processamento de Sinais, Depto. de Eng. Elétrica, Universidade Federal de Santa Catarina, Florianópolis, SC, E-mail: {tadeu, seara}@linse.ufsc.br.

Marcos Moecke, Centro Federal de Educação Tecnológica de Santa Catarina - CEFET/SC, Curso de Telecomunicações, São José, SC. E-mail: moecke@sj.cefetsc.edu.br.

Este trabalho foi parcialmente financiado pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

crominância 4:2:0. Cada MB é codificado em um dos modos *intra*, *predito*, *bi-predito*, ou *intra chaveado*. A Fig. 1 mostra os principais processos envolvidos na codificação de um macrobloco no padrão H.264. Como resultado da codificação, um fluxo de bits com reduzida taxa é produzido no processo. A função dos principais processos envolvidos na codificação de MB, ilustrados na Fig. 1, são descritos a seguir.

- **EM (Estimação de Movimento).** Consiste da busca de um bloco de imagem, com máxima semelhança, na região em torno do bloco em quadros anteriores. A informação de movimento é indicada no fluxo de bits pelo vetor de movimento (VM) e pela referência temporal.
- **CM (Compensação de Movimento).** É o processo que utiliza os VMs do MB corrente e as referências temporais do fluxo de bits para obter o sinal predito.
- **Predição *Intra*.** É uma das grandes novidades do H.264, consistindo na predição utilizando amostras das bordas de blocos vizinhos previamente decodificados.
- **T (Transformada do bloco).** No padrão H.264, é utilizada a transformada DCT inteira 4×4 [7].
- **Q (Quantização dos coeficientes).** É o processo que efetivamente permuta a qualidade pela taxa.
- **$Q^{-1}$  (Quantização inversa).**
- **$T^{-1}$  (Transformada inversa).**
- **Codificação de Entropia.** Consiste no processo que, dados os elementos de sintaxe (coeficientes, VMs, etc.), resulta no fluxo de bits comprimido. O H.264 possui dois tipos de codificação de entropia: o CABAC (*context adaptive binary arithmetic coding*), baseado na codificação aritmética e o CAVLC (*context adaptive variable length coding*), baseado em códigos de comprimento variável e o utilizado neste trabalho.

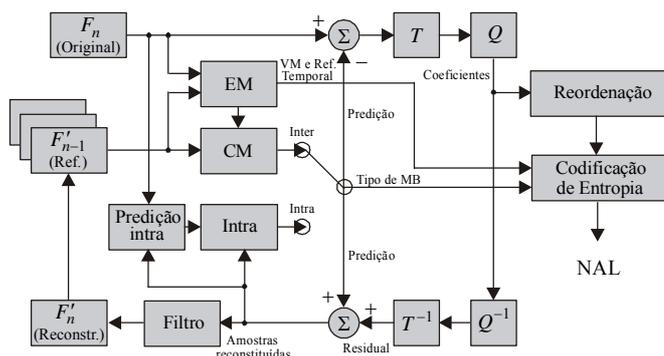


Fig. 1. Diagrama de blocos do codificador do padrão H.264.

O modo *predito* é também denominado *inter*, pois explora a redundância temporal existente entre quadros através dos processos de EM e CM. Nesse modo de codificação, o particionamento do MB pode ser realizado conforme ilustrado na Fig. 2. Os processos de EM/CM são executados de forma independente em cada partição, sendo associado um vetor de movimento a cada uma.

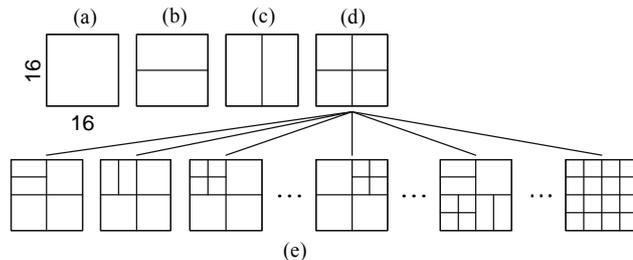


Fig. 2. Possibilidades de partição dos MBs nos modos *Inter*. (a) Partição 16×16. (b) Partições 16×8. (c) Partições 8×16. (d) Partições 8×8. (e) Subpartições do 8×8 em blocos menores de até 4×4 pixels.

Dentre todos os modos de codificação disponível, o codificador deve escolher aquele que resulte no melhor desempenho em R-D. A decisão da escolha do modo de codificação (*inter* ou *intra*) normalmente é feita através de um processo de otimização em taxa-distorção (RDO), que determina os melhores parâmetros de codificação para cada MB (tipo de particionamento, VMs, quadros de referência, direção de predição) [8]. Nesse processo, é necessário codificar cada MB com diversos conjuntos de parâmetros, visando eleger aquele que minimize a medida de R-D considerada, a qual é geralmente o custo Lagrangiano [9].

### III. CODIFICAÇÃO PROPOSTA (MACROBLOCO DUPLO-INTER)

#### A. Motivação

Em alguns MBs, a predição *inter* (realizada de acordo com o padrão H.264) torna-se ineficiente devido a uma pequena parte do MB necessitar de uma predição diferente do restante do MB. Isso ocorre, por exemplo, quando um pequeno objeto em movimento (ou sua borda) adentra um MB, que está representando um outro objeto com movimento em outra direção. Nessa situação, o codificador H.264 geralmente realiza a codificação do MB através de uma das seguintes alternativas:

- O MB não é particionado, sendo utilizado apenas um VM. Isso resulta em um erro residual elevado no bloco que representa o novo objeto, resultando em que a predição *inter* seja ineficiente para esse bloco. O custo de codificação dependerá dos coeficientes do bloco e do parâmetro de quantização considerado. Esse esquema de codificação é ilustrado na Fig. 3(a).
- O MB é particionado em 4 subMBs 8×8 e um deles é dividido em blocos 4×4. O bloco correspondente ao novo objeto apresentará um VM diferente dos demais. Nessa hipótese, o erro residual é reduzido, mas a quantidade de bits necessária para a codificação dos VMs é maior. Essa codificação é ilustrada pela Fig. 3(b).

O novo modo de codificação *duplo-inter* visa justamente melhorar o desempenho em R-D nas supramencionadas situações, subdividindo o MB em apenas duas partições com dois VMs distintos, conforme mostrado na Fig. 3(c).

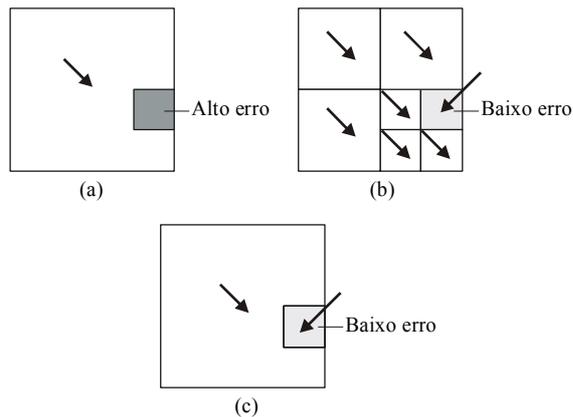


Fig. 3. Modos de particionamento do MB. (a) Modo *inter* utilizando apenas um VM. (b) Modo *inter* utilizando vários VMs. (c) Modo *duplo-inter* que usa 2 VMs.

### B. Particionamento do Macrobloco Duplo-Inter

No modo *duplo-inter*, o MB é dividido em uma partição menor (com  $4 \times 4$  pixels) e uma partição maior (de  $16 \times 16$  pixels), da qual são excluídas as amostras da partição menor, conforme ilustra a Fig. 4. Nesse caso, além das informações que normalmente são codificadas em um MB do tipo *inter* (particionamento, VMs, quadros de referência, erro de predição), são necessárias ainda a codificação de informações relativas ao tamanho e posição da partição menor. Para reduzir essas informações, permite-se apenas tamanhos de partições de  $4 \times 4$  com sua posição alinhada em múltiplos de 4 pixels, sendo suficiente o acréscimo de apenas 4 bits para a indicação da posição do bloco.

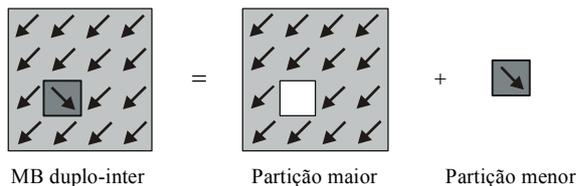


Fig. 4. Particionamento do MB *duplo-inter*.

### C. Mudanças nos Elementos de Sintaxe

Para indicar o modo *duplo-inter* é necessário especificar um novo valor para o elemento `mb_type`, que especifica o tipo de MB. O novo valor é aqui denominado `P_MIX` e sua inclusão altera a tabela de códigos do `mb_type`, de acordo com a Tabela I.

Na versão modificada (a coluna “modo *duplo-inter*” da Tabela I), todos os valores maiores do que 2 estão deslocados para um valor acima. Devido a essa mudança, alguns códigos de modos utilizados no padrão H.264 têm seu tamanho aumentado. Em termos de número de bits, o único tipo de MB com uma mudança significativa é o *inter*  $8 \times 16$ , cujo código passou de 3 para 5 bits (de 011 para 00100). Alguns tipos de *intra*  $16 \times 16$  assim como o tipo PCM também requerem 2 bits a mais na codificação do `mb_type`. No entanto, nesses

casos, o custo relativo na taxa de bits final é insignificante, pois além desses tipos de MB serem usados com uma menor frequência de ocorrência, eles já requerem uma grande quantidade de bits.

TABELA I  
TABELA DE CÓDIGOS PARA O ELEMENTO `mb_type`. VERSÃO PADRÃO H.264 E  
VERSÃO ALTERADA PARA INCLUIR O MODO *DUPLO-INTER*

mb_type	Padrão		Modo <i>duplo-inter</i>	
	Valor	Código	Valor	Código
P_L0_16x16	0	1	0	1
P_L0_L0_16x8	1	010	1	010
P_MIX			2	011
P_L0_L0_8x16	2	011	3	00100
P_8x8	3	00100	4	00101
P_8x8ref0	4	00101	5	00110
I_4x4	5	00110	6	00111
I_16x16_0_0_0	6	00111	7	0001000
I_16x16_0_0_1	7	0001000	8	0001001
...	...	...	...	...
I_16x16_3_2_1	29	000011110	30	000011111
I_PCM	30	000011111	31	00000100000

### D. Alterações na Estimação de Movimento

Para implementar o modo *duplo-inter* e manter a codificação otimizada em R-D, são ainda necessárias as alterações descritas a seguir.

Primeiramente, a EM é realizada 16 vezes, uma para cada possível posição da partição menor, sendo efetuada uma busca do melhor VM para a partição maior e em seguida para a partição menor. Tal procedimento é necessário porque o VM da partição menor pode utilizar em seu processo de predição o VM da partição maior. Deve ser considerado também o fato de que a partição menor pode ter uma referência temporal diferente daquela da partição maior.

O cálculo da SAD (*sum of absolute differences*) e da SATD (*sum of absolute transformed differences*) da partição maior deve ser modificado para que sejam desconsideradas as amostras da partição menor.

Finalmente, é necessário ainda adaptar as estruturas utilizadas no processo de RDO para armazenar temporariamente os códigos gerados e os parâmetros (tais como VMs e outros) para cada tipo de MB e particionamento considerado.

### E. Complexidade Computacional

A introdução de um novo modo de codificação no padrão H.264, a princípio, implica um pequeno aumento da complexidade computacional do processo de RDO. No entanto, utilizando a abordagem ICC proposta em [10], o novo modo de codificação apenas é avaliado quando os modos de menor custo de codificação se mostrarem ineficientes. Dessa forma, o aumento de complexidade é mínimo, tornando perfeitamente viável a utilização desse novo modo de codificação.

## IV. TESTES

## A. Condições de Teste

Para avaliar o modo de codificação proposto, diversas modificações foram incluídas no codificador de referência JM 7.3 [11], conforme descrito na Seção III. Para a realização dos testes, foram utilizadas as seqüências *Hall*, *Stefan*, *Tennis*, *Foreman*, *Coastguard*, *Mother & Daughter* e *Mobile & Calendar*, todas considerando uma taxa de 15 quadros por segundo.

O parâmetro de quantização (QP) é variado entre 18 e 40, permitindo medir a influência da inclusão do novo modo em diferentes qualidades de imagem e taxas de bits.

## B. Frequência de Utilização dos Modos de Codificação

Devido a RDO, o modo de codificação escolhido será sempre aquele que resultar no melhor desempenho em R-D, para um QP predefinido. Portanto, a frequência com que os diferentes modos (ver Tabela II) ocorrem indica a eficiência desses modos de codificação para uma dada taxa de bits. A Tabela III apresenta as frequências de ocorrência para seqüências padrão consideradas na literatura da área. Na Fig. 5, são também mostradas essas frequências para as seqüências *Hall*, *Stefan*, *Foreman* e *Mobile* considerando diferentes qualidades de imagem. Esses histogramas mostram que o modo *duplo-inter* substitui com vantagem o modo *inter* 16×16, apesar do primeiro ter um custo de codificação maior. Isso ocorre porque o emprego do modo *duplo-inter* resulta em uma menor distorção devido a uma melhor CM. O mesmo ocorre com o modo *inter* 8×8, que possui um custo de codificação maior. Nas seqüências avaliadas, a porcentagem de MBs codificados no modo *duplo-inter* é maior para altas taxas de bits, atingindo até 29%. No entanto, mesmo quando uma quantização grosseira é usada, o modo proposto continua sendo selecionado em até 16% dos MBs para o caso da seqüência *Mobile*. Esses resultados indicam claramente que a proposta é eficiente para situações que implicam movimentos diversificados dentro do MB.

A grande variação na escolha do modo de codificação em função da qualidade da imagem ocorre, principalmente, porque em baixas taxas de bits os modos de menor custo de codificação são mais adequados, sendo freqüentemente utilizado o modo *skip*, indicado usando menos do que um bit.

TABELA II  
MODOS DE CODIFICAÇÃO

Nº. do Modo	Modo
0	<i>skip</i>
1	<i>inter</i> 16×16
2	<i>inter</i> 16×8
3	<i>inter</i> 8×16
4	<i>inter</i> 8×8
5	<i>intra</i> 4×4
6	<i>intra</i> 16×16
7	<i>duplo-inter</i>

## C. Desempenho em Taxa-Distorção

A influência da introdução do modo *duplo-inter* também foi avaliada em termos de desempenho médio em R-D. A

Tabela IV mostra que os ganhos médios, obtidos variando o QP entre 18 e 30 (diferentes taxa de bits), não são tão significativos em relação aos obtidos pelo padrão H.264. Na Fig. 6, são mostradas as diferenças de PSNR para diversas taxas de bits.

Para a seqüência *Stefan*, existem taxas nas quais ganhos médios de até 0,21 dB são obtidos, enquanto em outras taxas há perdas de até 0,13 dB. Essa variação de ganho indica que o modo proposto não é sempre eficiente nesse tipo de seqüência, apesar de apresentar muitos MBs nos quais uma pequena parte do MB possui movimentação distinta do restante. O motivo da menor eficiência do modo *duplo-inter* nessa seqüência é a existência de muita movimentação de câmara com um fundo de imagem apresentando grande complexidade visual, situação na qual os modos *intra* 4×4 e *intra* 16×16 são mais adequados, resultando em uma alta porcentagem de uso desses modos, obtendo-se aproximadamente 25% e 8%, respectivamente.

Por outro lado, para a seqüência *Foreman* codificada com taxas de bits superiores a 60 kbps, para a seqüência *Hall*, com taxa superior a 40 kbps, e para a seqüência *Mobile* em todas as taxas de bits, ocorre um ganho em PSNR com a inclusão do modo *duplo-inter*. Isso ocorre porque essas seqüências possuem muitos MBs que apresentam regiões com movimentos em direções diversas do movimento geral do MB, lembrando que os modos *intra* não são eficientes nesses casos, praticamente não sendo usados.

TABELA III  
FREQUÊNCIA DE OCORRÊNCIA DO MODO *DUPLO-INTER* PARA DIFERENTES QPS

Seqüência	Formato	QP=18	QP=21	QP=24	QP=27	QP=30
<i>Hall</i>	QCIF	29,23%	13,20%	5,16%	2,90%	2,30%
<i>Stefan</i>	SIF	10,61%	9,59%	8,92%	8,15%	7,13%
<i>Tennis</i>	SIF	15,01%	10,28%	8,11%	6,62%	4,98%
<i>Foreman</i>	CIF	14,96%	11,78%	9,28%	6,82%	4,39%
<i>Coastguard</i>	CIF	15,81%	13,16%	11,40%	9,64%	7,54%
<i>Foreman</i>	QCIF	17,51%	14,80%	12,49%	9,47%	6,21%
<i>Coastguard</i>	QCIF	17,95%	15,06%	13,29%	11,49%	9,05%
<i>Mother</i>	QCIF	9,75%	8,10%	6,41%	4,75%	3,32%
<i>Mobile</i>	QCIF	22,39%	20,18%	19,65%	17,97%	15,59%
<b>Média</b>	-	17,02%	12,91%	10,52%	8,65%	6,72%

TABELA IV  
GANHOS MÉDIO E MÁXIMO OBTIDOS PARA CADA SEQÜÊNCIA. VALORES REFERENTES À VARIAÇÃO DO PARÂMETRO QP ENTRE 18 E 30

Seqüência	Resolução	Ganho Médio	Ganho Máximo
<i>Hall</i>	QCIF	0,02 dB	0,09 dB
<i>Stefan</i>	SIF	0,01 dB	0,21 dB
<i>Tennis</i>	SIF	0,01 dB	0,03 dB
<i>Foreman</i>	CIF	0,01 dB	0,08 dB
<i>Coastguard</i>	CIF	0,00 dB	0,03 dB
<i>Foreman</i>	QCIF	0,01 dB	0,07 dB
<i>Coastguard</i>	QCIF	0,01 dB	0,04 dB
<i>Mother</i>	QCIF	0,00 dB	0,03 dB
<i>Mobile</i>	QCIF	0,04 dB	0,11 dB

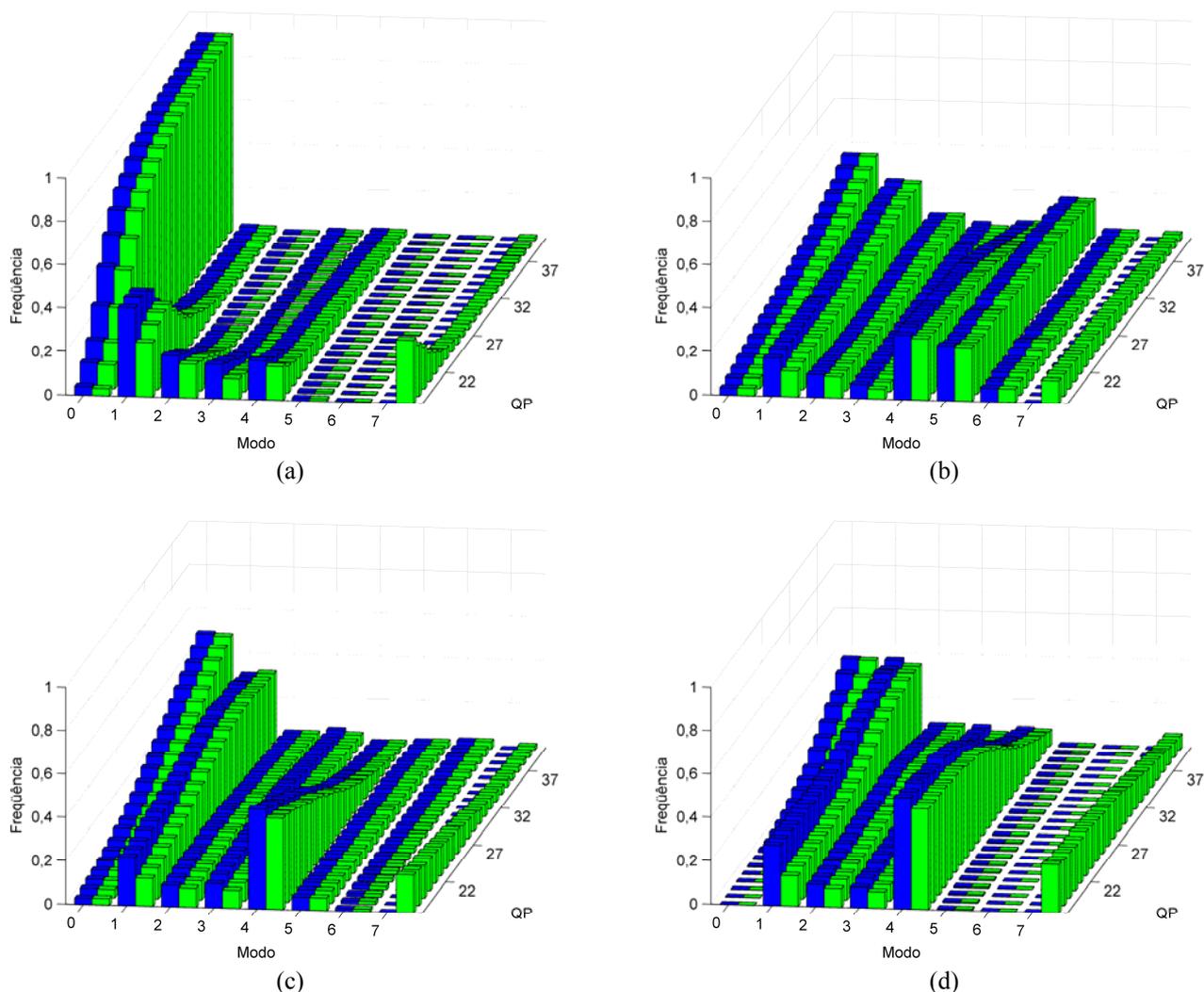


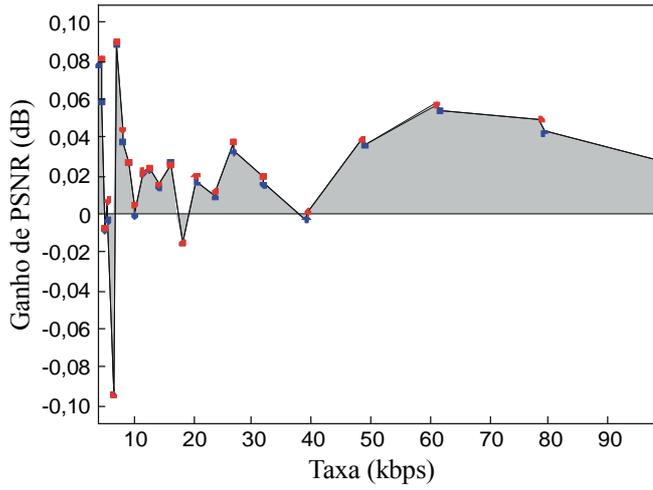
Fig. 5. Histograma dos tipos de MBs selecionados com o codificador H.264 (em azul/escuro) e usando o modo duplo-inter (em verde/claro). Sequências: (a) *Hall* – QCIF, (b) *Stefan* – SIF, (c) *Foreman* – QCIF e (d) *Mobile* – QCIF.

## V CONCLUSÕES

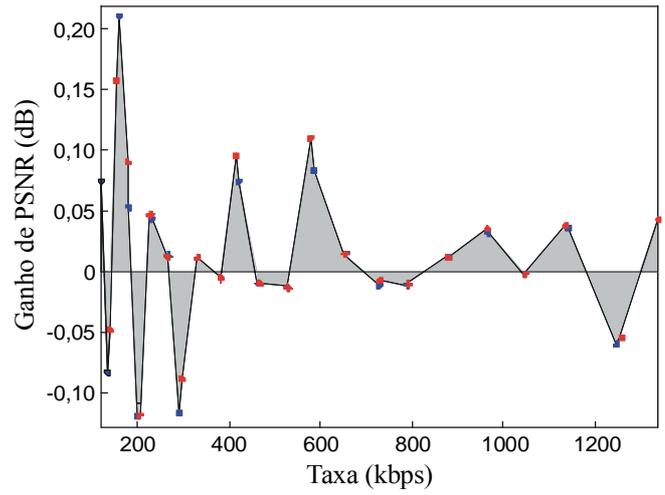
Este artigo apresentou uma proposta de modificação do padrão H.264, buscando melhorar a taxa de compressão para alguns cenários. Foram implementadas alterações no modelo de referência (JM), visando avaliar o desempenho da referida proposta. Os resultados obtidos indicam ganhos de desempenho de até 0,2 dB para algumas seqüências de vídeo comuns, tal como *Mobile*, que é bastante complexa em termos visuais e possui grande movimentação, e para outras menos dinâmicas, como a *Hall*.

## REFERÊNCIAS

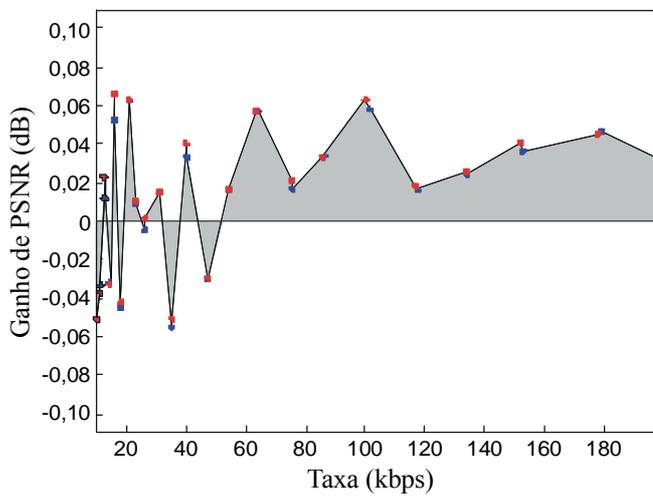
- [1] ITU-T Recommendation H.264 | ISO/IEC 14496-10:2005 (MPEG-4), *Advanced Video Coding for Generic Audiovisual Services*, Int'l Telecomm. Union | Int'l Organization for Standardization, Geneva, Switzerland, 2005.
- [2] ITU-T Recommendation H.262 | ISO/IEC 13818-2 (MPEG-2), *Information Technology – Generic Coding of Moving Pictures and Associated Audio Information – Part 2: Video*, Int'l Telecomm. Union | Int'l Organization for Standardization, Geneva, Switzerland, 1995.
- [3] ITU-T Recommendation H.263, *Video Coding for Low Bit Rate Communication*, Int'l Telecomm. Union, Geneva, Switzerland, 1998.
- [4] I. E. G. Richardson, *H.264 and MPEG-4 Video Compression: Video Coding for Next-Generation Multimedia*. John Wiley & Sons, West Sussex: 2003.
- [5] S. Kondo and H. Sasai, "A motion compensation technique using sliced blocks in hybrid video coding," in *Proc. IEEE Int. Conf. Image Processing (ICIP 05)*, Genova, Italy, vol. 2, Sept. 2005, pp. 305-308.
- [6] E. M. Hung, R. L. de Queiroz, and D. Mukherjee, "On macroblock partition for motion compensation," in *Proc. IEEE Int. Conf. Image Processing (ICIP 06)*, Atlanta, USA, Oct. 2006, pp. 1697-1700.
- [7] H. Malvar, A. Hallapuro, M. Karczewicz et al., "Low-complexity transform and quantization in H.264/AVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 637-644, Jul. 2003.
- [8] T. Wiegand, H. Schwarz, A. Joch et al., "Rate-constrained coder control and comparison of video coding standards," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 688-703, Jul. 2003.
- [9] A. Ortega and K. Ramchandran, "Rate-distortion methods for image and video compression," *IEEE Signal Processing Mag.*, vol. 15, no. 6, pp. 23–50, Nov. 1998.
- [10] M. Moecke and R. Seara, "Rate–distortion optimized video coding with stopping rules: Quality and complexity," in *Proc. IEEE Int. Conf. Image Processing (ICIP 04)*, Singapore, Oct. 2004, pp. 753-756.
- [11] ITU-T & ISO/IEC JVT, "H.264/AVC Reference Software Encoder", <http://iphome.hhi.de/suehring/tml/download/> (current Apr., 2007).



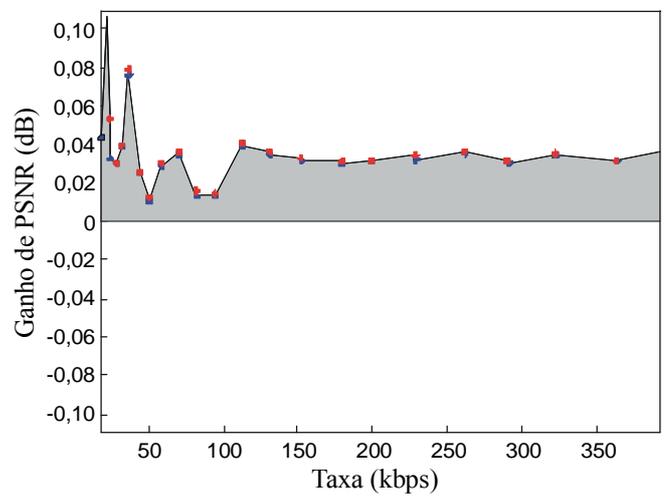
(a)



(b)



(c)



(d)

Fig. 6. Ganho em PSNR do codificador H.264 considerando o modo *duplo-inter* em relação ao codificador original. Sequências: (a) *Hall*, (b) *Stefan*, (c) *Foreman* e (d) *Mobile*.