

Otimização taxa-distorção para quadros-B no padrão H.264/AVC

Diego F. de Souza, Eduardo A. B. da Silva, Alessandro J. S. Dutra,
Nuno M. M. Rodrigues, Sergio M. M. de Faria e Vitor M. M. da Silva

Resumo— O padrão de codificação de vídeo H.264/AVC, estado da arte em compressão de vídeo e recentemente adotado no Sistema Brasileiro de TV Digital (SBTVD), apresenta desempenho muito superior a seus antecessores, como o amplamente utilizado MPEG-2, devido à utilização intensiva e combinada de diversas técnicas de codificação como, por exemplo, estimação/compensação de movimento, predição bidirecional etc. Apesar da elevada complexidade, esta combinação de métodos é a principal responsável pelos excelentes resultados obtidos. Neste artigo propomos a adição de um modo, compatível com a norma H.264/AVC, nas rotinas de otimização taxa-distorção empregadas na codificação de quadros-B. Este modo permite ao codificador descartar os coeficientes da transformada do resíduo de predição sempre que o custo de codificação associado for menor que o dos modos anteriormente testados (p.ex. predição por “lista 0”). Diversos testes de codificação de seqüências padrão da literatura são apresentados, confirmando o potencial do método proposto.

Palavras-Chave— vídeo, codificação, H.264/AVC, quadros-B

Abstract— The state-of-the-art H.264/AVC video codec, recently adopted as part of the Brazilian Digital TV System, presents a much improved performance when compared to its predecessors (e.g. MPEG-2), due in large part to the combined and intensive use of several coding techniques, such as motion estimation/compensation, bi-directional prediction etc, albeit at a much larger computational complexity. In this article, we propose the inclusion of an additional, standard compliant, mode, in the rate-distortion optimization loop for the encoding of B-frames. This mode will allow the encoder to selectively discard transform coefficients of the prediction residue whenever the associated $D + \lambda R$ cost is smaller than all the other tested modes (e.g. list0 prediction). Several encoding tests, performed on standard video sequences, are presented that confirm the potential of the proposed method.

Keywords— vídeo, coding, H.264/AVC, B-frames

I. INTRODUÇÃO

O padrão de codificação de vídeo H.264/AVC [2], estado da arte em compressão de vídeo, e recentemente adotado no Sistema Brasileiro de TV Digital (SBTVD), apresenta desempenho muito superior a seus antecessores, como o amplamente utilizado MPEG-2, devido à utilização intensiva e combinada de diversas técnicas de codificação como, por exemplo, estimação/compensação de movimento, predição bidirecional etc.

Diego F. de Souza, Eduardo A. B. da Silva, Alessandro J. S. Dutra, Laboratório de Processamento de Sinais, DEL/COPPE, Universidade Federal do Rio de Janeiro, E-mails: diego.fekilis@gmail.com, eabsilva@ieee.org, a.dutra@ieee.org, Nuno M. M. Rodrigues, Sergio M. M. de Faria e Vitor M. M. da Silva, Instituto de Telecomunicações, Universidade de Coimbra—Pólo II, Portugal E-mails: nuno.rodrigues@co.it.pt, sergio.faria@co.it.pt, vitor.silva@co.it.pt. Este trabalho foi parcialmente financiado pela FINEP.

O emprego combinado de métodos de codificação já presentes em padrões anteriores e de propostas apresentadas especificamente para o novo padrão é a principal razão para excelentes resultados obtidos.

Entretanto, esta associação de métodos traz, como subproduto, um aumento considerável de complexidade computacional, complexidade esta que pode ser controlada através da habilitação ou não das diversas opções de codificação disponíveis. O padrão H.264/AVC, ao permitir tal flexibilidade, aumenta consideravelmente suas possibilidades de adoção nos mais diversos sistemas, tanto para transmissão de TV aberta - caso do SBTVD - como em sistemas proprietários - caso do sistema de transmissão por satélite *DirectTV*.

Neste artigo propomos a adição de um modo nas rotinas de otimização taxa-distorção empregadas na codificação de quadros-B, com o objetivo de verificar se a transmissão dos coeficientes da transformada do resíduo de predição de um macrobloco (e suas partições) é realmente necessária ou se o descarte seletivo de tal informação produziria melhores resultados finais de codificação.

É importante notar que a incorporação de tal opção ao sistema não contraria as especificações do padrão H.264/AVC - que tratam unicamente da sintaxe da *bitstream* a ser decodificada - sendo uma modificação que afeta exclusivamente o comportamento do codificador. Portanto, o método proposto é totalmente compatível com o padrão atual.

Na seção II apresentamos uma breve revisão dos conceitos envolvidos na codificação de quadros-B, segundo a norma que regula o padrão H.264/AVC. A seção III mostra resultados preliminares de codificação de quadros-B onde todo o resíduo de codificação é descartado (não transmitido), sem quaisquer considerações relativas ao comportamento taxa-distorção do codificador [4]. Estes resultados preliminares indicam que, em certos casos, pode ser vantajoso não transmitir esta informação adicional (sobre os resíduos) e empregar os bits “economizados” para outros fins. A seguir, na seção IV, propomos um método de otimização taxa-distorção que considera a possibilidade de transmissão ou não da informação adicional, comparando os custos obtidos a partir da análise de um operador de Lagrange, do tipo $D + \lambda R$. Os resultados obtidos na codificação de diversas seqüências de vídeo padrão utilizadas na literatura são apresentados na seção V e comparados tanto com os resultados originais obtidos pelo codificador H.264/AVC quanto com o codificador modificado para descartar completamente a informação referente ao resíduo de predição. Por fim, na seção VI apresentamos as conclusões e propostas de trabalho futuro.

II. CODIFICAÇÃO DE QUADROS-B NO PADRÃO H.264

O padrão H.264/AVC [5] permite que, durante o processo de estimação/compensação de movimento, um macrobloco de luminância (16×16 pixels) seja particionado em sub-macroblocos de dimensões 16×16 , 16×8 , 8×16 e 8×8 . Cada sub-macrobloco 8×8 pode ainda ser dividido em partições de tamanho 8×4 , 4×8 e 4×4 . Cada partição possui um vetor de movimento próprio e um índice apontando para o quadro de referência utilizado. Assim, para cada macrobloco, a quantidade de vetores de movimento varia entre 1 (para uma partição 16×16) e 16 (quando partições 4×4 são utilizadas), onde cada vetor de movimento pode ser associado a diferentes quadros de referências.

Na codificação de quadros-B, o processo de predição inter utiliza duas listas, denominadas L0 e L1, para armazenar quadros de referência e os macroblocos do tipo-B (e suas partições e subpartições) podem ser preditos utilizando-se qualquer combinação das mesmas, isto é, $L0 + L1$, $L0 + L0$ ou $L1 + L1$ [3]. A cada partição de um macrobloco pode-se associar um diferente modo de predição. Entretanto, todas as subpartições de um bloco 8×8 terão modos de predição idênticos.

Na predição bidirecional, cada vetor de movimento pode ser associado à lista L0 – ordenada pelo passado, i.e. o menor índice da lista é associado ao quadro passado mais próximo na ordem de apresentação – ou à lista L1 – ordenada pelo futuro. O valor predito para o pixel (i, j) é definido da seguinte forma:

$$\text{pred}(i, j) = \frac{\text{pred0}(i, j) + \text{pred1}(i, j) + 1}{2} \quad (1)$$

Pode-se ainda utilizar uma bipredição ponderada (*weighted prediction*), onde se aplicam os fatores multiplicativos w_0 e w_1 a cada uma das referências.

Existem, também, duas opções de modo direto – espacial e temporal – para a predição bidirecional, onde os vetores de movimento para um determinado bloco não são transmitidos, mas calculados a partir dos macroblocos de mesma posição espacial nos quadros de menor índice das listas L0 e L1 [3]. Blocos em quadros-B codificados como *skipped* são decodificados de acordo com as regras do modo direto.

O sinal de predição obtido a partir do processo de estimação e compensação de movimento é subtraído da informação original, gerando o resíduo a ser codificado. As sinal residual aplica-se uma transformada para redução da redundância entre os coeficientes e os coeficientes da transformada são, então, codificados por entropia. Esta etapa é geralmente reconhecida como crucial para o desempenho de qualquer codificador de vídeo.

Os altos ganhos alcançados pelo processo de estimação e compensação de movimento do padrão H.264/AVC, aliado a um eficiente esquema de otimização taxa-distorção, leva-nos a crer que, ao menos em alguns casos, a compensação de movimento já é eficiente o bastante e pode ocasionalmente justificar a não transmissão da informação referente ao resíduo de predição, o que pode proporcionar razoável economia de bits no processo.

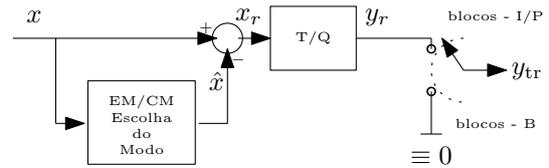


Fig. 1. Modo de codificação alternativo – resíduo de codificação nulo para quadros-B

III. DESCARTANDO O RESÍDUO DE CODIFICAÇÃO PARA QUADROS-B

A fim de verificar o desempenho obtido pelo codificador H.264/AVC ao descartar o resíduo de codificação em quadros-B e, ao mesmo tempo, manter a compatibilidade do *bitstream* com o padrão, optou-se por igualar a zero todos os coeficientes residuais antes do início do processo de codificação [4]. O codificador então comprime a informação residual através do parâmetro *Coded Block Pattern* (CBP), informando ao decodificador da existência de um bloco de dados nulos.

É interessante notar que no software de referência, em sua implementação atual, o controle do estágio de quantização é feito através do valor do *parâmetro de quantização* QP e, portanto, a possibilidade de igualar a zero os coeficientes do resíduo de predição existe desde que o valor de QP seja alto.

A Fig. 1 mostra os passos de codificação seguidos pelos blocos de uma imagem, destacando o caminho alternativo para blocos sujeitos à predição bidirecional. Neste caso, os dados a serem transmitidos são igualados a zero ($y_{tr} = 0$). Já para blocos codificados como Intra ou Inter-P, os valores transmitidos correspondem aos valores quantizados dos coeficientes da transformada do resíduo ($y_{tr} = y_r$).

Os demais passos no processo de codificação de quadros-B, tais como a otimização taxa-distorção e a escolha do modo de codificação, são mantidos inalterados, exceto pelo fato de processarem um bloco de dados nulos. Uma consequência deste fato é que os passos de transformada direta e inversa deixam de ser necessários, proporcionando uma redução na complexidade computacional do codec.

IV. OTIMIZANDO A CODIFICAÇÃO DE QUADROS-B

A análise dos resultados obtidos com o método de descarte dos resíduos de codificação para quadros-B apresentado na Seção III, que proporciona uma melhora de desempenho de codificação para determinadas sequências, indica que a inclusão de um método de otimização taxa-distorção que leve em conta a *possibilidade* de igualar a zero os coeficientes do resíduo de predição pode resultar em ganhos ainda maiores de desempenho.

O método aqui proposto altera o laço de otimização taxa-distorção para quadros-B, incluindo modos de descarte dos coeficientes do resíduo de predição para cada um dos modos de particionamento do macrobloco sendo codificado, conforme pode ser visto na Fig. 2.

O custo de codificação $D + \lambda R$ é calculado para cada um dos possíveis modos, com e sem a exclusão dos coeficientes, sendo o modo de menor custo o escolhido pelo codificador.

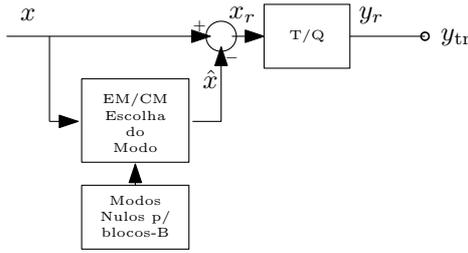


Fig. 2. Modo de codificação alternativo – modos adicionais para quadros-B

TABELA I
VALORES DE QP UTILIZADOS.

Tipo de quadro	QP						
B	45	40	35	30	25	20	15
I, P	43	38	33	28	23	18	13

V. RESULTADOS OBTIDOS

A implementação do método proposto foi baseada no software de referência do padrão H.264/AVC, versão *JM 15.0* [1].

O conjunto de testes realizado avaliou o desempenho das três variações do codificador H.264/AVC aqui discutidas – o codec original, o codificador com descarte automático do resíduo para quadros-B (denominado *B_noCoeffs*) e o modelo proposto, com descarte seletivo acoplado ao laço de otimização taxa-distorção (denominado *B_rdoCoeffs*).

Todas as simulações envolveram a codificação dos primeiros 99 quadros das sequências de teste *Akiyo* (formato QCIF), *Foreman* (CIF), *Flower Garden* (SIF) e *Mobile Calendar* (CIF), tendo as componentes de crominância sido subamostradas segundo o padrão 4:2:0.

A configuração utilizada, idêntica para os três codificadores, incluiu:

- perfil *high*
- codificação bipreditiva
- transformadas 4×4 e 8×8
- quadros Intra: somente o primeiro
- Lista 0: 5 quadros de referência
- Lista 1: 1 quadro de referência
- GOP (*group of pictures*): IBPBPBP...
- modo Intra habilitado para blocos de quadros Inter
- modo para utilizar quadros-B como referência: desabilitado

Um teste adicional foi realizado para a sequência *Foreman*, com a definição do GOP alterada para o seguinte padrão: IBPBPBPBP...

Diferentes taxas - em bits/frame ou kb/s - podem ser alcançadas variando o parâmetro de quantização (QP) para cada tipo de quadro. Os valores utilizados nas simulações aqui descritas são apresentados na Tabela I. É interessante mencionar que há uma relação direta entre os valores de QP e o multiplicador de Lagrange λ utilizado no processo de otimização taxa-distorção [1].

Em todas as simulações realizadas, o comportamento das curvas taxa-distorção para as componentes de crominância (U e V) foi muito similar ao de luminância (Y) e, portanto, a fim de permitir uma melhor visualização e evitar o excesso de

curvas em um mesmo gráfico, serão apresentados somente os resultados para esta última.

A primeira sequência analisada – *Akiyo*, no formato QCIF – diferencia-se das demais por ter quadros de pequena dimensão (176×144 pixels) e pouco movimento dos objetos apresentados. Em consequência, o método *B_noCoeffs*, onde não há envio dos dados referentes ao resíduo da transformada, apresenta a melhor relação custo-benefício dentre os três, principalmente em taxas mais altas de codificação, onde a PSNR média para quadros-B é bastante similar, apesar da quantidade de bits utilizada ter sido significativamente menor, conforme pode ser visto na Fig. 3.

As demais sequências, *Foreman*, *Mobile Calendar* e *Flower Garden*, possuem grande quantidade de movimento, inclusive com mudanças de fundo (*background*). Nestes casos, conforme apresentado nas Figs. 4-7, o método proposto apresenta desempenho superior aos demais testados, particularmente em taxas mais altas. Os resultados confirmam a utilidade das alterações feitas nas rotinas de otimização taxa-distorção, com a inclusão dos modos adicionais de descarte seletivo de coeficientes.

Na Fig. 5 pode-se também verificar que, mesmo com o uso de GOPs com múltiplos quadros-B consecutivos (IBBBP), o método proposto também apresentou melhores resultados que os demais. Convém salientar que os quadros-B não foram utilizados como referência para codificação.

VI. CONCLUSÕES

Neste trabalho apresentamos uma adaptação do codificador baseado no padrão H.264/AVC, incluindo novos modos de codificação que preveem o descarte seletivo dos coeficientes relativos à informação do resíduo de predição, a partir da análise do custo de codificação dado pelo operador $D + \lambda R$.

Os resultados obtidos confirmam a utilidade que os modos adicionais de descarte apresentam em relação ao codificador convencional, uma vez que, para baixas taxas, o método de descarte automático descrito em [4] mostra-se superior, enquanto para taxas médias e altas o algoritmo aqui proposto produz o melhor desempenho.

O desempenho obtido pelo novo algoritmo indica também que a inclusão de um novo modo de codificação auxiliar, baseado no uso de dicionários adaptativos, pode vir a proporcionar ainda melhores resultados – o que será objeto de investigação futura.

REFERÊNCIAS

- [1] Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG. H.264/AVC Reference Software. <http://iphome.hhi.de/suehring/tml/>.
- [2] Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG. Draft of Version 4 of H.264/AVC (ITU-T Recommendation H.264 and ISO/IEC 14496-10 (MPEG-4 part 10) Advanced Video Coding, January 2005. JVT-N050d1.
- [3] I. Richardson. *H. 264 and MPEG-4 video compression: video coding for next-generation multimedia*. Wiley, 2003.
- [4] N. Rodrigues, E. da Silva, M. de Carvalho, S. de Faria, and V. da Silva. On Overriding H.264/AVC B-Slice Predicted Residue Coding. In *Picture Coding Symposium (PCS), Lisbon, Portugal, Nov, 2007*.
- [5] T. Wiegand, G. Sullivan, G. Bjontegaard, and A. Luthra. Overview of the H.264/AVC video coding standard. *Circuits and Systems for Video Technology, IEEE Transactions on*, 13(7):560–576, July 2003.

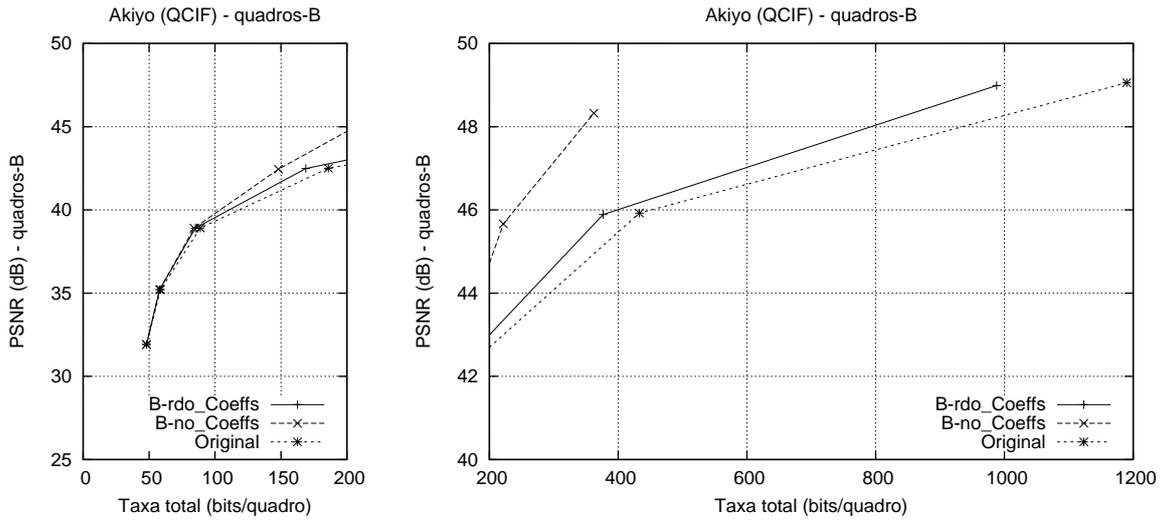


Fig. 3. PSNR (dB) média para quadros-B - Luminância - Sequência Akiyo - 99 quadros

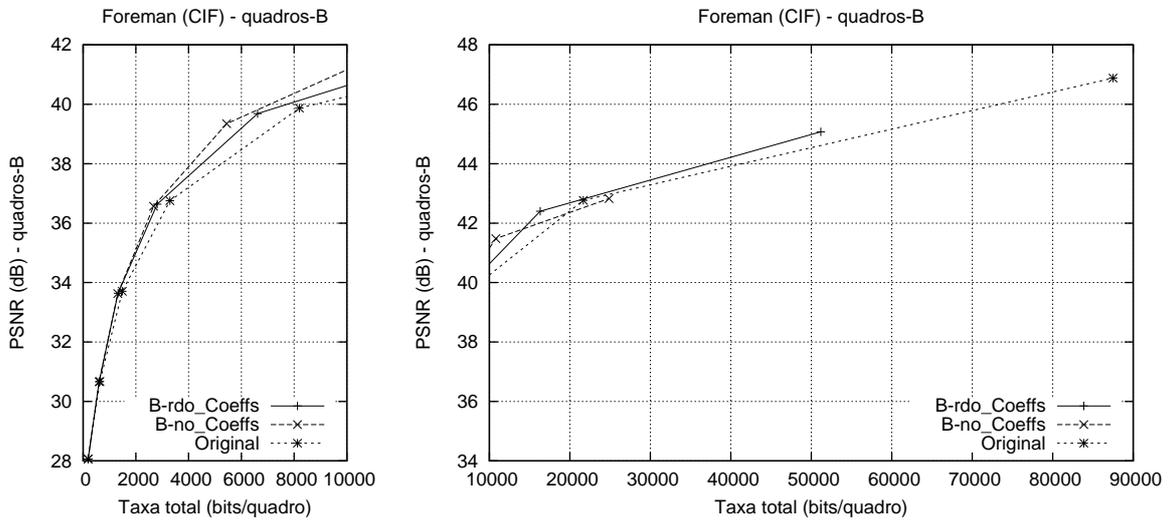


Fig. 4. PSNR (dB) média para quadros-B - Luminância - Sequência Foreman - 99 quadros

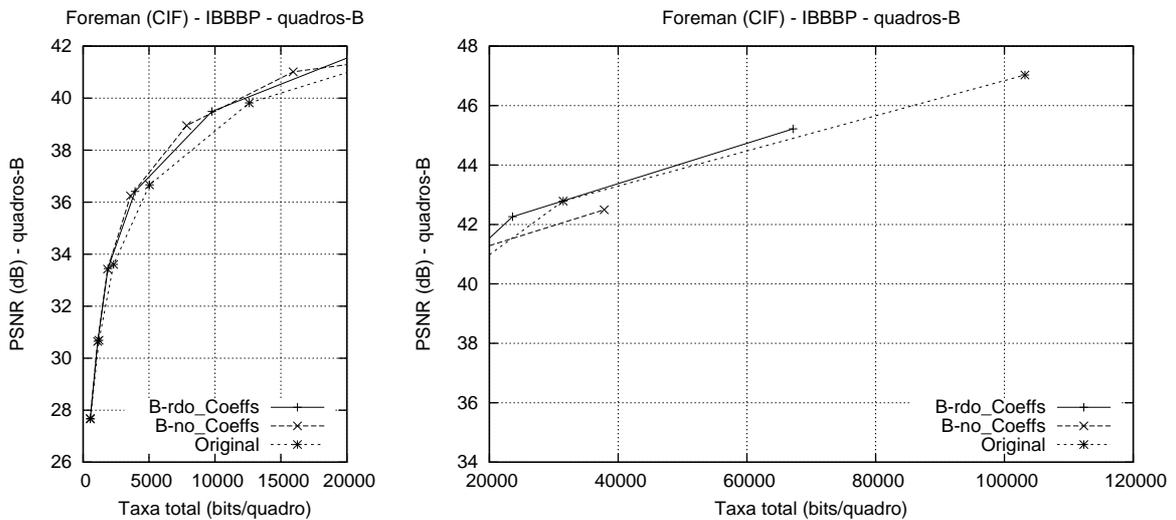


Fig. 5. PSNR (dB) média para quadros-B - Luminância - Sequência Foreman - IBBBP - 99 quadros

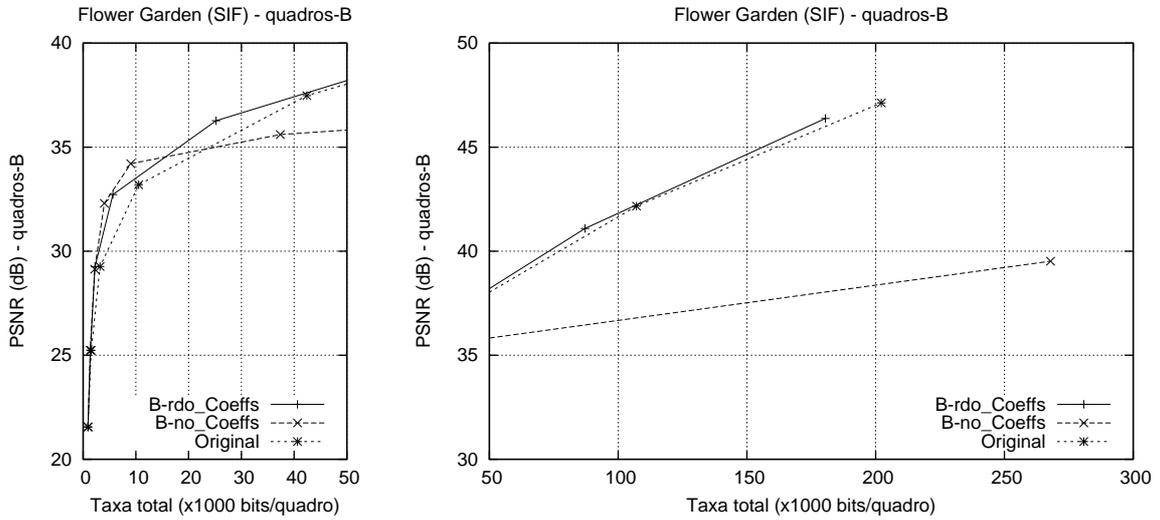


Fig. 6. PSNR (dB) média para quadros-B - Luminância - Sequência Flower Garden - 99 quadros

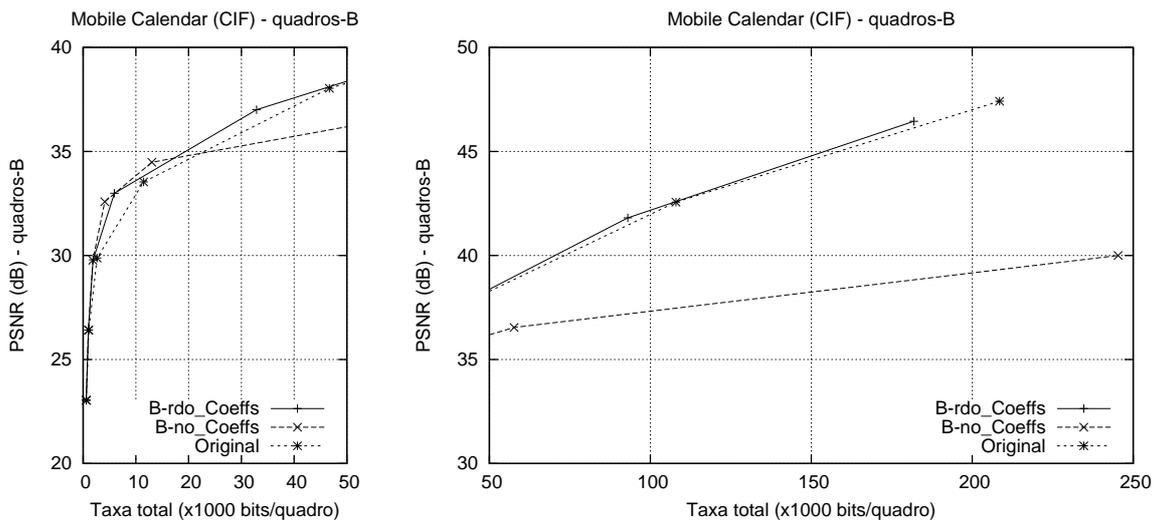


Fig. 7. PSNR (dB) média para quadros-B - Luminância - Sequência Mobile Calendar - 99 quadros