

# Codificação de Vídeo com Complexidade Reversa utilizando Qualidade Mista

Edson M. Hung e Ricardo L. de Queiroz

**Resumo**—Este artigo apresenta uma proposta de codificação com complexidade reversa diferente das arquiteturas clássicas, que em geral utilizam codificação em camadas ou um paradigma de Codificação de Vídeo Distribuída. Neste caso, utilizamos uma codificação de vídeo com qualidade mista, no qual propomos uma técnica de realçar a qualidade (objetiva e subjetiva) do vídeo, no decodificador, utilizando informações dos quadros de alta qualidade (quadros chave) para melhorarmos a qualidade dos quadros de baixa qualidade (quadros não-chave). Os resultados mostram que é possível realçar a qualidade objetiva do vídeo dos quadros-chave em mais de 1dB.

**Palavras-Chave**—Codificador de vídeo, H.264, Máximo a Posteriori, Wyner-Ziv, Complexidade Reversa.

**Abstract**—This article presents a reversed complexity video coding scheme that is not aligned with the classical Distributed Video Coding approach. In this case, we use a mixed quality video coding and we propose a technique to improve objectively and subjectively the video quality, using information of the high quality key frames to enhance the low quality frames (non-key frames) at the decoder. The results show that it is possible to improve the non-key-frames quality by up to 1dB.

**Keywords**—Video coding, H.264, Maximum a Posteriori, Wyner-Ziv, Reversed Complexity.

## I. INTRODUÇÃO

Os padrões de codificação de vídeo convencionais, como o H.264/AVC [1], exigem uma grande complexidade computacional (capacidade de processamento) no codificador se compararmos com o processo de decodificação [2]. Isto se deve às operações de predições *intra* e *inter* (estimação de movimento e compensação de movimento) [2]. Para que o H.264 atinja um bom desempenho de compressão, estes modos de predição são escolhidos a partir da minimização de uma função de custo que relacionam a taxa e a distorção do vídeo. No decodificador as predições possuem simplesmente seus dados interpretados, de forma a reconstruir de maneira aproximada (nos casos de codificação com perdas) o mesmo vídeo que foi codificado. Por outro lado, a academia e a indústria tem pesquisado intensamente em Codificação de Vídeo Distribuída [3], conduzindo a um novo paradigma de codificação onde a complexidade é reversa, em outras palavras, a complexidade da codificação é reduzida por meio da eliminação ou redução da estimação de movimento e no decodificador a estatística da fonte é explorada de modo a melhorar o desempenho (em termos de taxa-distorção) do *codec* [4] [5]. Neste artigo também estamos interessados

em transferir parte da complexidade do codificador para o decodificador, entretanto sem nenhuma outra informação lateral que o caracterize como um codificador de vídeo com fonte distribuída. Uma das formas para que a técnica proposta se torne um codificador de vídeo distribuído seria adicionar uma camada que pode ser Wyner-Ziv [6] ou quaisquer outros tipos de camada de realce [7] que pudesse melhorar ainda mais desempenho do *codec*, mas com o custo de gerar uma codificação que seja diferente do padrão H.264. Ou melhor, ao adicionar uma camada de realce, várias modificações devem ser feitas no processo de decodificação do H.264 para que resulte em um vídeo de melhor qualidade e por conseqüência maior taxa ao adicionar códigos para correção ou melhoria do vídeo.

## II. ARQUITETURA PROPOSTA

Neste trabalho propomos uma arquitetura de codificação e decodificação de vídeo com qualidade mista. Que apesar de ser apresentada para o H.264/AVC esta arquitetura pode ser aplicada em outros padrões de codificação de vídeo. Para codificar um vídeo com qualidade mista, vide Figura 1, basta determinar um tamanho para o grupo de quadros (comumente denotado por GOP na literatura [1]) que é o menor conjunto de quadros que represente a codificação total da seqüência. Neste caso, teremos dois tipos de quadros ao variarmos o parâmetro de quantização (QP): um de melhor qualidade ( $QP_{chave}$ ) e outro de qualidade reduzida ( $QP_{não-chave}$ ), ou seja,  $QP_{chave} < QP_{não-chave}$ .

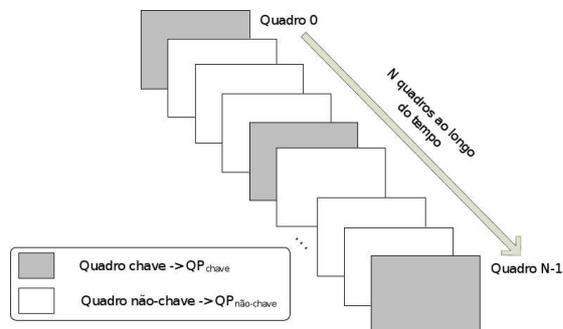


Fig. 1. Codificador de vídeo com qualidade mista.

A decodificação é feita da maneira tradicional (com o decodificador padrão), e em seguida são adicionadas aos quadros não-chave as informações de realce (na grande maioria de alta freqüência) que se encontram nos quadros chave, como mostra a Figura 2. Observe que são usados os quadros de um

Edson M. Hung e Ricardo L. de Queiroz, Departamento de Engenharia Elétrica, Universidade de Brasília, CP 04591, 70910-900 Brasília, DF, Brasil. E-mails: mintsu@image.unb.br, queiroz@ieee.org. Este trabalho foi parcialmente financiado pela HP Brasil e pelo CNPq (47.3696/2007-0).

GOP adicionado ao quadro chave posterior. Assim realce bi-direcional é realizado ao utilizarmos dois quadros chaves (um anterior e um posterior) para cada quadro não-chave.

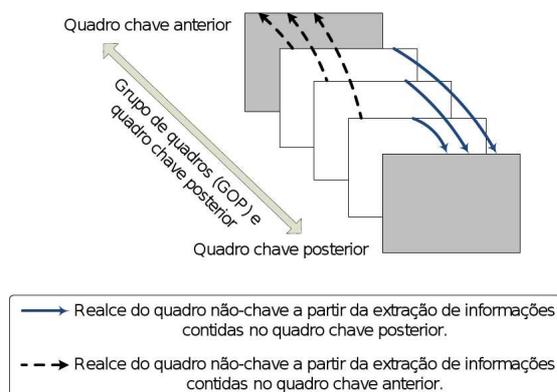


Fig. 2. Realce da seqüência de vídeo para decodificação de vídeo com qualidade mista.

Na seção seguinte apresentaremos um método para estimação de informações perdidas (quantizadas com  $QP_{\text{não-chave}}$ ) dos quadros não-chave que estão contidos nos quadros chave (que mantém uma qualidade maior, pois foram quantizadas com  $QP_{\text{chave}}$ ). A Figura 3 mostra um esquemático simplificado da aplicação do processo de realce na arquitetura proposta.

### III. MÉTODO DE REALCE PROPOSTO

O método de realce proposto é baseado no trabalho de Brandi et al. [8], onde é feita uma semi-super resolução de uma seqüência de vídeo baseado em quadros chave. Neste trabalho, ao invés de serem utilizados quadros não-chaves com resolução reduzida, os quadros não-chaves passam por uma quantização maior (que a dos quadros chave), o que implica em uma redução de qualidade. A princípio o processo de decodificação de um vídeo comprimido com qualidade mista pode ser feito com um decodificador padrão. E o processo de realce poderá ser adicionado ao processo de decodificação de forma a aumentar a qualidade dos quadros não-chave (vide Figura 3). A Figura 4 mostra um diagrama de blocos simplificado do processo de realce proposto, e nas subseções seguintes iremos discutir os detalhes de cada etapa.

#### A. Estimação de movimento com tamanho de blocos variáveis

A técnica proposta utiliza a estimação de movimento no decodificador, mas para que os quadros de referência (quadros chaves) tenham as mesmas características que os quadros atuais (quadros não-chave), deve-se gerar quadros chaves tão degradados quanto o quadro atual, ou seja, o processo de quantização com o  $QP_{\text{não-chave}}$  é aplicado aos quadros chave decodificados, obtendo quadros chaves degradados. Considera-se a informação perdida pelo processo quantização com o  $QP_{\text{não-chave}}$  como sendo a diferença entre o quadro chave e o quadro chave degradado, baseado em Segall et. al [9]. Entretanto essa informações de realce oriunda dos quadros chave devem ser adaptadas aos quadros não-chave, que neste

caso, será feito por meio da estimação e da compensação de movimento. Para cada bloco do quadro atual (não-chave) uma procura do bloco mais similar é feita em uma janela de busca de tamanho fixa no quadro de referência (quadro chave degradado), gerando assim um fluxo óptico (um conjunto de vetores de movimento). Este processo é realizado para tamanhos de blocos variáveis (*quadtree*) no qual a busca de um bloco de tamanho  $16 \times 16$  pixels pode ser substituída pelas buscas de quatro blocos de tamanho  $8 \times 8$ , conforme mostra a Figura 5.



Fig. 5. Estimação de movimento utilizando tamanho de bloco variável.

Note que o conjunto de possíveis candidatos a minimização da soma das diferenças absolutas (SAD) ou da soma das diferenças quadráticas (SSD) na estimação de movimento em blocos partidos (de tamanho  $8 \times 8$ ) contempla o caso do bloco de tamanho  $16 \times 16$  (não partido). O que poderia nos induzir a escolher sempre um bloco de tamanho menor para fazer a estimação de movimento, mas verificamos experimentalmente que tamanhos de blocos grandes (de  $16 \times 16$  pixels) tendem a ter melhores resultados. Logo, as partições (blocos menores) foram penalizadas, de forma a favorecer as partições de bloco de maior tamanho por um fator de dois. Repare que o objetivo da estimação de movimento neste caso seria buscar nos quadros não-chave um melhor casamento entre objetos nos quadros não-chave, de forma que consigamos realçá-la. Ou seja, diferente do caso onde a estimação de movimento é feita no codificador, o nosso objetivo não seria somente minimizar o erro de predição. A partir do fluxo óptico gerado pelo processo de estimação de movimento, a compensação de movimento é feita com a informação de realce do quadro chave, ou seja, com a diferença entre o quadro chave e o quadro chave degradado.

#### B. Utilizando uma estimativa Bayesiana para ponderação de informações de realce bidirecionais

O modelo que utilizaremos para descrever uma predição será:

$$X_i = X + \epsilon_i, \quad \epsilon_i \sim N(0, \sigma_i^2) \quad (1)$$

onde  $X_i$  é um bloco do  $i$ -ésimo quadro de referência, que é modelado como sendo o bloco da imagem original  $X$  corrompido por um ruído espacial  $\epsilon_i$ , i. i. d., e  $i = \{1, 2\}$  é um elemento do conjunto de quadros de referência, que neste caso será restrito aos quadros chave de anteriores e posteriores ao quadro não-chave atual. Temos ainda que  $E\{X_1\} = X$ ,  $E\{X_2\} = X$  e  $X_i \sim N(X, \sigma_i^2)$ . Como nenhum conhecimento *a priori* de  $X$  é dado, então utilizaremos  $\hat{X} = X_1$  como sendo a estimativa inicial, nos permitindo construir essa informação *a priori*:

$$\hat{X} \sim N(X_1, \sigma_1^2) \quad (2)$$

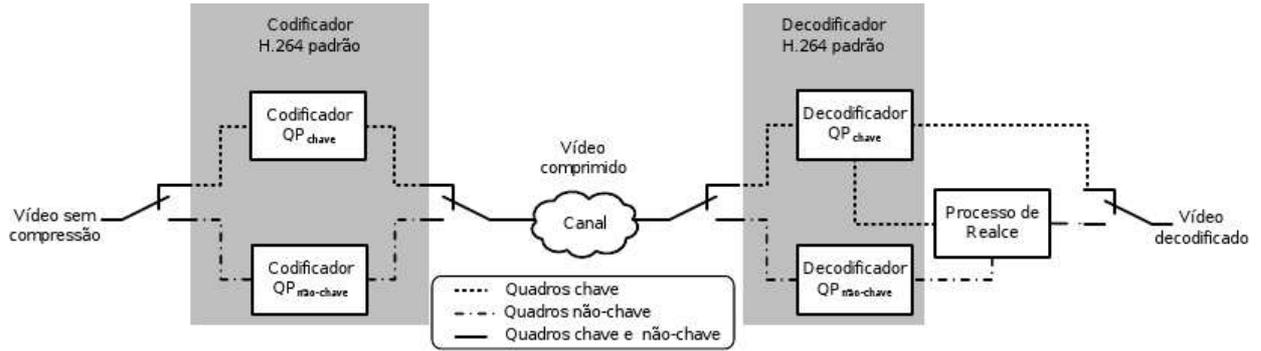


Fig. 3. Esquemático do funcionamento do processo de codificação e decodificação de vídeo com qualidade mista utilizando o realce proposto.

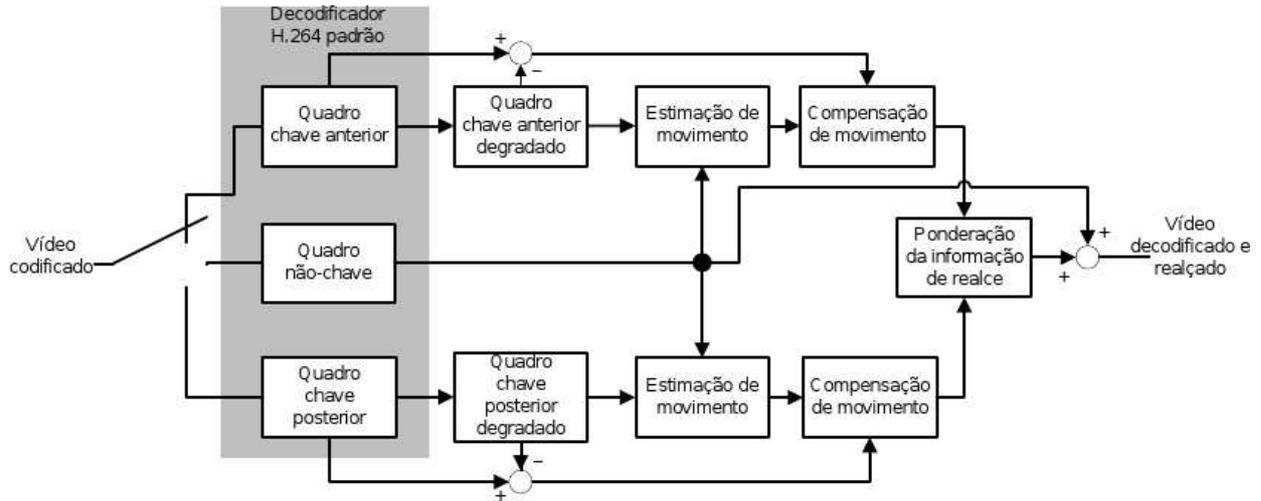


Fig. 4. Esquemático do funcionamento do realce proposto.

e

$$p_{X_1}(X) = \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp \left\{ -\frac{1}{2} \frac{(X - X_1)^2}{\sigma_1^2} \right\}. \quad (3)$$

Então,  $p_X(X) = p_{X_1}(X)$  entra como informação *a priori*. Temos ainda que

$$p(X_2|X) = \frac{1}{\sqrt{2\pi\sigma_2^2}} \exp \left\{ -\frac{1}{2} \frac{(X_2 - X)^2}{\sigma_2^2} \right\}. \quad (4)$$

Neste caso o estimador máximo *a posteriori* (MAP) Bayesiano é dado por:

$$\hat{X} = \max_X (p(X_2|X)p_X(X)) \quad (5)$$

e equivale a

$$\hat{X} = \max_X (\ln (p(X_2|X)p_X(X))). \quad (6)$$

Para calcular o valor máximo de X, temos que:

$$\frac{\partial (\ln (p(X_2|X)p_X(X)))}{\partial X} \Big|_{X=\hat{X}} = 0 \quad (7)$$

resultando em:

$$\frac{(X_1 - \hat{X})}{\sigma_1^2} + \frac{(X_2 - \hat{X})}{\sigma_2^2} = 0. \quad (8)$$

Portanto, temos como a estimativa Bayesiana para a fusão das informações de realce bidirecionais:

$$\hat{X} = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} X_1 + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} X_2. \quad (9)$$

Intuitivamente sabe-se que quanto maior o parâmetro de quantização (QP), maior a variância do erro entre o sinal original e o decodificado ( $\sigma^2$ ). De maneira similar, o QP também é proporcional à soma das diferenças absolutas entre o sinal original e o decodificado (SSD). Portanto é intuitivo pensar que podemos substituir  $\sigma^2$  pela SSD na Equação (9) que resulta em:

$$\hat{X} = \frac{SSD_2}{SSD_1 + SSD_2} X_1 + \frac{SSD_1}{SSD_1 + SSD_2} X_2, \quad (10)$$

onde os sub-índices 1 e 2 representam as predições dos macroblocos em quadros anteriores e posteriores ao quadro atual, respectivamente. Em outras palavras, temos que as informações de realce dos quadros chaves (anterior e posterior) são fundidos bloco a bloco a partir dos pesos dados pela Equação (10).

*C. Inserção das informações de realce dos quadros chave nos quadros não-chave*

Note que o método de estimação de movimento sempre seleciona uma predição para realçar o quadro não-chave.

Entretanto em mudanças abruptas na cena, a adição das informações de realce contidas nos quadros chave poderá degradar (quantitativa e qualitativamente) os quadros não-chave. Para contornar esta limitação, propôs-se uma adição gradativa (em porcentagem) desta informação até que o quadro não-chave realçado maximize a relação sinal ruído de pico (PSNR) entre os quadros chave.

IV. EXPERIMENTOS

Para avaliarmos o desempenho da técnica proposta, faremos o realce para as seqüências de vídeo com 300 quadros e tamanho QCIF (176 × 144 pixels) codificados com H.264 Intra com GOP quatro (ou seja, para cada quadro chave teremos três quadros não-chave). Neste caso, fazemos com que o  $QP_{\text{não-chave}}$  tenha o dobro do fator de quantização que o  $QP_{\text{chave}}$ , o que corresponde a  $QP_{\text{não-chave}} = QP_{\text{chave}} + 6$  [10]. Para o método de realce, realizamos a estimação de movimento no decodificador com a janela de busca de 16 × 16 pixels para o bloco sem partição, e 8 × 8 pixels para blocos com partição. Repare que a complexidade do codificador H.264 é reduzida ao utilizarmos apenas quadros do tipo intra, ou seja, não é feita nenhuma estimação/compensação de movimento. E com a diminuição da qualidade dos quadros não-chave conseguimos reduzir a quantidade de bits na compressão do vídeo. Este trabalho propõe a melhoria da qualidade dos quadros não-chave baseado em informações contidas nos quadros chave por meio da estimação de movimento no decodificador, revertendo assim a complexidade do codec de vídeo.

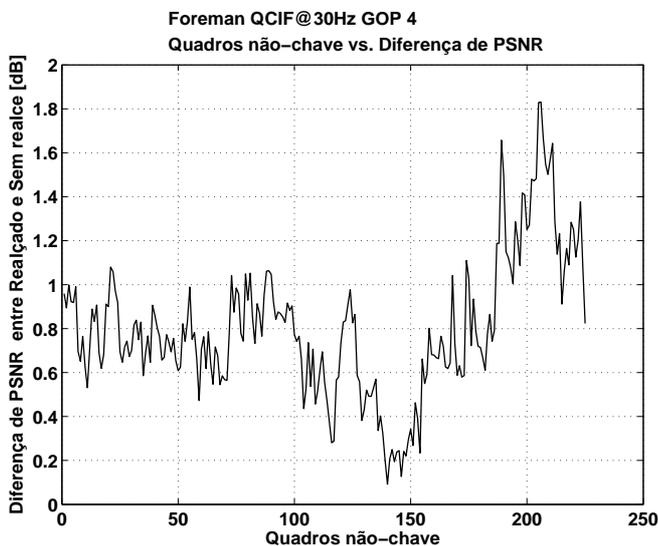


Fig. 6. Ganhos dos quadros não-chave utilizando a técnica de realce para a seqüência de vídeo Foreman codificado com  $QP_{\text{chave}} = 30$  e  $QP_{\text{não-chave}} = 36$ .

As Figuras 6 e 7 mostram o desempenho do realce para cada quadro não-chave, que possuem um ganho médio de 1,07 dB e 1,85 dB respectivamente. Para seqüências de vídeo de grande movimento, como é o caso da seqüência Foreman o ganho em termos de PSNR do realce pode ser bem pequeno (em alguns casos) por causa da grande variação entre quadros. Já para a

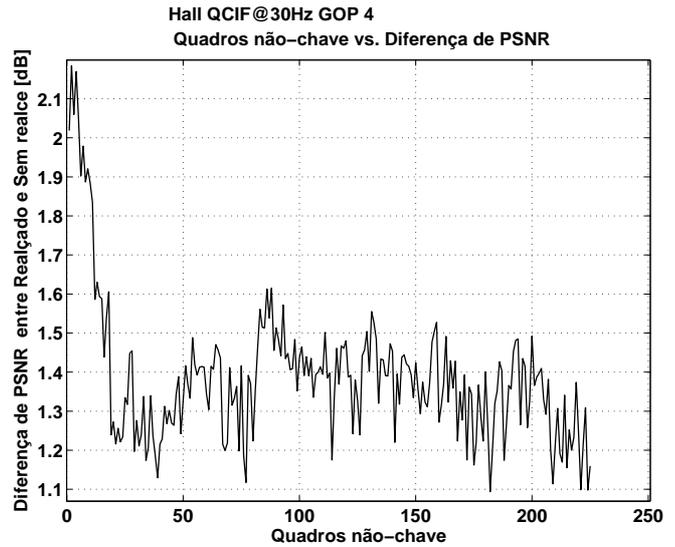


Fig. 7. Ganhos dos quadros não-chave utilizando a técnica de realce para a seqüência de vídeo Hall codificado com  $QP_{\text{chave}} = 30$  e  $QP_{\text{não-chave}} = 36$ .

seqüência Hall, cujo conteúdo possui um movimento pequeno, o método de realce gera ganhos mais expressivos.

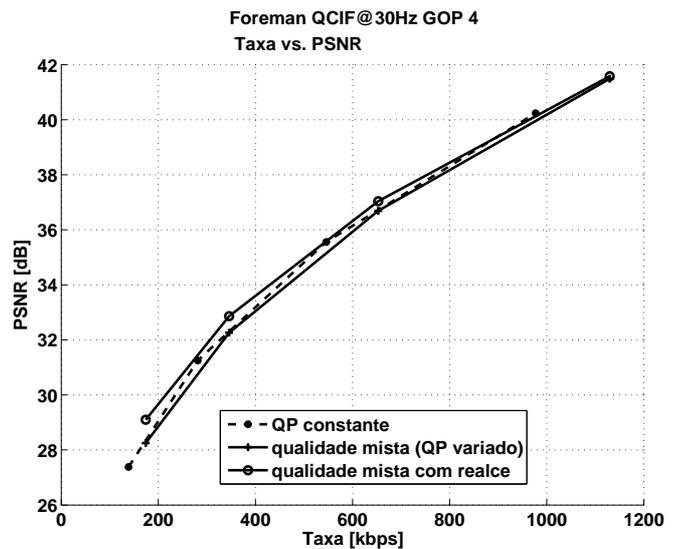


Fig. 8. Comparação em termos de taxa-distorção da seqüência de vídeo Foreman das técnicas com qualidade fixa, qualidade mista e qualidade mista com realce.

O desempenho do método de realce para diferentes fatores de quantização pode ser visto nas Figuras 8 e 9. Neste caso, calculamos o valor das PSNRs para todos os quadros da seqüência decodificada e comparamos com a codificação utilizando QP fixo, qualidade mista (QP variado) e qualidade mista com realce. Observe que ao fazermos o realce não mudamos o valor da taxa, ou seja, não utilizamos nenhuma informação adicional para a realização da mesma. Além disso, é importante ressaltar que os ganhos obtidos são relativos a seqüência de vídeo completa (quadros-chave e não-chave), ou melhor, os ganhos de realce nos quadros não-chave são “diluídos” em toda a seqüência, pois neste trabalho, nenhum



Fig. 10. Quadro não-chave (esquerda), Quadro original (centro), Quadro não-chave realçado (direita). A seqüência Foreman foi codificada com  $QP_{chave}=30$ ,  $QP_{não-chave}=36$  e  $GOP=4$ . As imagens são referentes ao terceiro quadro, note que esse quadro possui a “maior distância” entre quadros chaves.

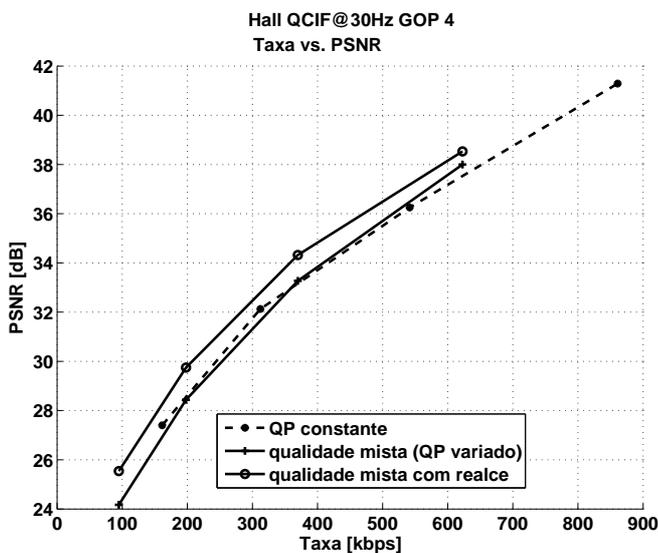


Fig. 9. Comparação em termos de taxa-distorção da seqüência de vídeo Hall das técnicas com qualidade fixa, qualidade mista e qualidade mista com realce.

tipo de realce para os quadros chave foi proposto.

Observe na Figura 10, em um exemplo da melhoria subjetiva da camada de realce, pode se observar que o quadro realçado (direita) possui um detalhamento um pouco mais acentuado, principalmente na bochecha, no queixo e na imagem de fundo se compararmos com o quadro não realçado (esquerda). Neste caso, o quadro original (centro) também é mostrada para efeitos comparativos.

## V. CONCLUSÕES

Neste artigo propomos um *codec* com uma nova arquitetura de complexidade reversa ao utilizarmos qualidade mista. Além disso, essa proposta pode ainda ser utilizada como uma técnica de geração de informação lateral para codificadores Wyner-Ziv com qualidade mista [5]. Os resultados mostram que é possível melhorar os quadros degradados de uma seqüência de vídeo, caso inserirmos quadros menos degradados, sem a adição de informação para o decodificador. Como trabalho futuro propomos estudos e testes para determinar a melhor relação entre  $QP_{chave}$  e  $QP_{não-chave}$ .

## REFERÊNCIAS

- [1] T. Wiegand, G. Sullivan, G. Bjontegaard e A. Luthra, *Overview of the H.264 video coding standard*. IEEE Transactions on Circuits and Systems for Video Technology, v. 13, pp. 560–576, Julho 2003.
- [2] K. R. Rao e J. J. Hwang, *Techniques and Standards for Image, Video and Audio Coding*. Prentice Hall, PTR, 1996.
- [3] B. Girod, A. Aaron, S. Rane e D. Rebollo-Monodero, *Distributed Video Coding*. Proceedings of the IEEE, v. 93, pp. 71–83, Janeiro 2005.
- [4] X. Artigas, J. Ascenso, M. Dalai, S. Klomp, D. Kubasov e M. Oualet, *The DISCOVER codec: Architecture, Techniques and Evaluation*. Picture Coding Symposium, Portugal, 2007.
- [5] D. Mukherjee, B. Macchiavello e R. L. de Queiroz, *A simple reversed complexity Wyner-Ziv video coding mode based on a spatial reduction framework*. Proc. SPIE Visual Communications and Image Processing, VCIP, Janeiro 2007.
- [6] A. Wyner e J. Ziv, *The rate-distortion function for source coding with side information at the decoder*. IEEE Transactions on Information Theory, v. 22, pp. 1–10, Janeiro 1976.
- [7] H. Schwarz, D. Marpe, e T. Wiegand, *The scalable H.264/MPEG-4 AVC extension: Technology and applications*. European Symposium on Mobile Media Delivery, Sardinia, Italy, Setembro 2006.
- [8] F. Brandi, R. L. de Queiroz e D. Mukherjee, *Super-resolution of video using key-frames and motion estimation*. Proc. IEEE Intl. Conf. on Image Processing, ICIP, San Diego, CA, USA, Outubro 2008.
- [9] C. A. Segall, R. Molina e A. K. Katsaggelos, *High-resolution images from low-resolution compressed video*. IEEE Signal Processing Magazine, v. 20, pp. 37–48, 2003.
- [10] I. E. Richardson, *H.264 and MPEG-4 Video Compression: Video Coding for Next Generation Multimedia*. Wiley, 2003.