

Super-resolução de vídeo usando quadros chave

Karen F. de Oliveira, Fernanda Brandi, Edson M. Hung, Ricardo L. de Queiroz e Debargha Mukherjee

Resumo—Diferentes sistemas de codificação de vídeo com escalabilidade espacial codificam os quadros em distintas resoluções para que possam ser obtidas diferentes camadas de decodificação. Alguns desses sistemas também se utilizam da decimação de quadros para reduzir complexidade, permitindo o uso de camadas de realce. Nesse contexto, métodos de super-resolução de quadros associados a processos eficientes de interpolação podem ajudar a melhorar a qualidade dos quadros codificados em baixa resolução espacial. Este trabalho apresenta uma técnica de super-resolução a partir de quadros chave. O objetivo é recuperar a alta frequência dos quadros decimados a partir de quadros chaves codificados em alta resolução. Os quadros resultantes do processo de super-resolução podem ser utilizados em codificadores com escalabilidade espacial ou com qualidade variável, além de serem utilizados na geração de informação lateral para codificadores distribuídos. Os resultados mostram uma melhoria significativa em relação a trabalhos anteriores.

Palavras-Chave— Super-resolução, Resolução mista, Estimação de movimento, Interpolação, Decimação.

Abstract—Many scalable video coding systems use variable resolution frames to enable different decoding layers. Some of these systems also use frame down-sampling to reduce complexity, allowing the use of enhancement layers. In order to do that, super-resolution methods associated with efficient interpolation processes may help increasing low-resolution frames quality. This work presents a super-resolution technique based on key frames. The goal is to restore the high-frequency information of down-sampled frames using high-resolution frames as references. The super-resolved frames can be used on scalable video coders or in variable quality coders, and also in side information generation for distributed coders. Results indicate substantial improvements over previous works.

Keywords— Super-resolution, Mixed Resolution, Motion Estimation, Down sampling, Up sampling.

I. INTRODUÇÃO

Entende-se por super-resolução o conjunto de processos capaz de gerar imagens de alta resolução a partir de imagens de baixa resolução [1]. Uma imagem de baixa resolução é aquela em que se tem baixa densidade de *pixels*, isto é, baixa quantidade de *pixels* por unidade de área, oferecendo, portanto, poucos detalhes. Por contraposição, imagens de alta resolução são aquelas em que a densidade de *pixels* é elevada, assim como a quantidade de detalhes. Dessa maneira, o processo de super-resolução tem como entrada um conjunto de imagens de baixa resolução e deve processá-las de modo a se obter uma imagem em maior resolução. Na prática, as imagens de baixa resolução podem ser quadros de uma sequência de vídeo ou, até mesmo, diferentes vistas de uma única cena. Um método tradicionalmente utilizado na super-resolução de imagens é o

método Bayesiano [2], em que todas as imagens de entrada possuem a mesma resolução. Uma proposta diferente é a de se utilizar um algoritmo baseado em exemplos extraídos de outras imagens para aumentar a resolução da imagem de interesse [3].

Em compressão de vídeo, é necessária uma solução de compromisso entre a qualidade e a taxa necessárias à representação, isto é, quanto melhor a qualidade desejada, maior deve ser o gasto em *bits*. Geralmente, essa relação entre taxa e distorção é controlada principalmente por meio do parâmetro de quantização. Têm sido propostos esquemas alternativos de codificação, em que apenas alguns quadros da sequência são subamostrados, com o intuito de permitir modos de codificação escalonável [4], ou de reduzir a complexidade na codificação [5]. Nesse contexto, uma sequência de vídeo pode ser formada por quadros de diferentes resoluções, constituindo o que se chama de sequência com resolução mista [6], [7]. Pode-se, portanto, identificar dois tipos de quadros em uma sequência de resolução mista: os quadros que permanecem em alta resolução, chamados quadros chave (do inglês, *Key Frames*-KF); e aqueles com resolução reduzida, os quadros não-chave (do inglês, *Non-Key Frames*-NKF). Os KFs podem ser codificados como quadros do tipo intra (I), predito (P) ou bi-predito (B), e servem como referência para a codificação dos NKFs.

No decodificador, serão reconstruídos quadros de alta resolução, intercalados com quadros de baixa resolução, como mostrado na Figura 1. Pretende-se, portanto, se utilizar dos KFs vizinhos para melhorar a qualidade dos NKFs. Essa solução é desejável em qualquer sistema de codificação em que se tenha um conjunto de quadros menos degradados intercalados por quadros mais degradados, como, por exemplo, em sistemas de codificação com qualidade variável [8]. Além disso, o método proposto também é relevante no processo de geração de informação lateral em certas arquiteturas de codificação distribuída [9].

O problema de melhoramento por super-resolução de quadros não-chave é bastante genérico, abrangendo diferentes cenários de codificação de vídeo. O foco desse trabalho é apresentar uma técnica eficiente para melhoramento de quadros decodificados à baixa resolução, baseada na extração de informações de mais alta qualidade de quadros em resolução integral.

II. SUPER-RESOLUÇÃO USANDO QUADROS CHAVE

O método de super-resolução adotado baseia-se na codificação de quadros em resolução mista [10], como ilustrado na Figura 2. No codificador, nota-se a separação entre os KF e NKF. Os NKF são filtrados e subamostrados, reduzindo sua resolução, enquanto os KFs serão codificados na resolução completa. Em seguida, a sequência codificada

Trabalho financiado pela HP Brasil. K. F. de Oliveira, E. M. Hung, F. Brandi e R. L. de Queiroz são do Departamento de Engenharia Elétrica, Universidade de Brasília, Brasília, Brasil, E-mails: {karen, mintsu, fernanda}@image.unb.br, queiroz@ieee.org. D. Mukherjee é do Hewlett Packard Laboratories, Palo Alto, California, USA, E-mail: debargha.mukherjee@hp.com.

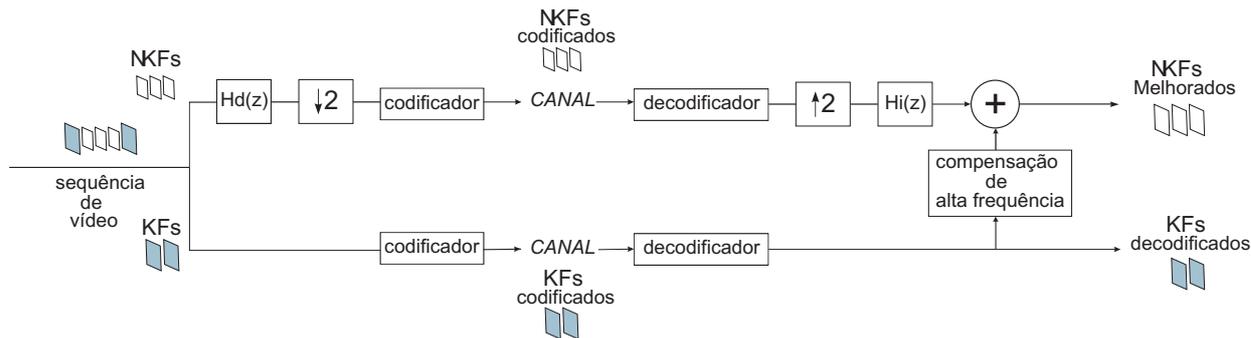


Fig. 2. Esquemático geral do método de super-resolução adotado. Do lado do codificador, os quadros chave (KFs) são codificados separadamente e em resolução integral, enquanto os quadros não-chave (NKFs) são filtrados e decimados para que fiquem com resolução reduzida. Já no lado do decodificador, tanto os KFs quanto os NKFs são decodificados. Os NKFs são submetidos ao processo de melhoramento por super-resolução, que se inicia com o processo de interpolação seguido de filtragem, recuperando-se o tamanho original dos quadros. A esses quadros interpolados será adicionada uma informação de alta frequência, resultante da ponderação das altas frequências dos blocos de melhor casamento dos KFs vizinhos.

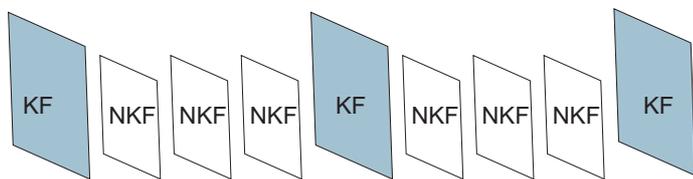


Fig. 1. Exemplo de sequência de quadros em um decodificador com resolução espacial mista. Observa-se uma sequência de quadros chave (KF), em alta resolução, intercalados por quadros não-chave (NKF), em resolução reduzida.

será transmitida por um canal. Já no decodificador, após todos os quadros serem decodificados, o método de super-resolução proposto será aplicado para o melhoramento da qualidade dos NKFs. Esse método é constituído basicamente pelas seguintes etapas:

- Interpolação dos NKFs, para que tenham o mesmo tamanho dos KFs;
- Degradação dos KFs, resultante dos procedimentos de decimação e interpolação, com posterior estimação de movimento entre os blocos dos NKFs e os blocos do quadro chave anterior (KFb, do inglês, *Key-Frame backward*), e do quadro chave posterior (Kff, do inglês, *Key-Frame forward*) ao quadro atual;
- Extração da informação de alta frequência dos blocos dos NKFs;
- Ponderação das altas frequências dos blocos de melhor casamento do KFb e do Kff, de acordo com o grau de confiabilidade associado, que é inversamente proporcional às SSDs;
- Adição da alta frequência ponderada aos blocos interpolados dos NKFs, processo denominado compensação de alta frequência.

Como indicado na Figura 2, identifica-se a presença de filtros para decimação ($H_d(z)$) e interpolação ($H_i(z)$), com a função de evitar, respectivamente, efeitos de *aliasing* e de *imaging* [11]. Para tal, foi utilizado o filtro Lanczos3, resultante do janelamento da função *sinc* com a preservação de três lobos, já implementado em funções do *software* MatLab®. Em seguida, serão descritos pormenorizadamente cada um dos processos

realizados após a decodificação para o melhoramento dos quadros de baixa resolução, que constitui o foco do trabalho.

A. Compensação de alta frequência

Na decodificação de uma sequência de vídeo em resolução mista, os NKFs devem ser interpolados após serem decodificados. No entanto, nesse processo, os quadros interpolados perdem informação de alta frequência, já que foram codificados à baixa resolução. Assim, baseando-se no fato de que, em geral, os quadros temporalmente vizinhos muito se assemelham [12], propôs-se um método de estimação de movimento nos Kfb e no Kff, para se extrair a alta frequência do bloco mais parecido dos dois KFs, que será somada ao NKF atual de baixa resolução. Naquele primeiro trabalho [10], a estimação já era realizada tanto no Kfb quanto no Kff, mas apenas a alta frequência do melhor casamento de um deles era usada na compensação. Nesse trabalho propõe-se ponderar a informação de alta-frequência dos melhores casamentos do Kfb e do Kff por uma grandeza associada à confiança que se tem em cada e, por fim, adicionar a informação composta ao bloco interpolado.

Uma questão central do trabalho é a de como promover um eficiente casamento dos blocos interpolados dos NKFs com os blocos dos KFs, de forma a se obter a informação de alta frequência mais próxima da que se tinha originalmente. Dessa forma, os KFs são também submetidos ao mesmo processo de degradação experimentado pelos NKFs, isto é, serão também decimados e interpolados, como procedimento preliminar ao processo de estimação de movimento. Além do mais, foi observado que realizar a busca na alta frequência dos quadros interpolados levaram a resultados melhores do que simplesmente realizar a busca nos quadros interpolados, especialmente nos casos em que se tem uma alta densidade de KFs. É interessante notar que, adotar esse procedimento de busca, ilustrado na Figura 3, equivale a utilizar a alta frequência da componente de baixa frequência para prever as componentes de alta frequência em uma decomposição em sub-bandas. Esse é o mesmo conceito hierárquico explorado por codificadores de imagem baseados em *zerotrees* [13],[14]. Nesses codificadores, há predição de coeficientes através das

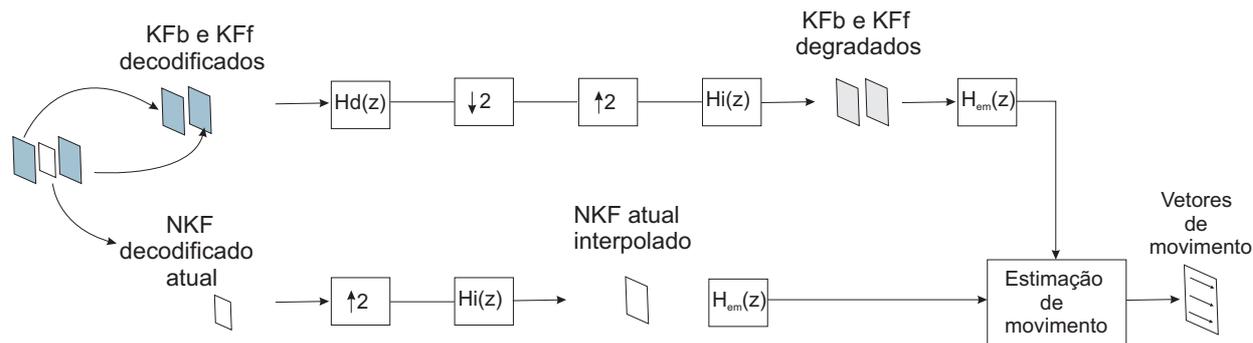


Fig. 3. Procedimentos realizados para viabilizar a busca dos blocos dos NKFs nos KFs e KFs. Os NKFs são apenas interpolados, enquanto os KFs são decimados e interpolados para sofrerem degradação similar à dos NKFs. Em seguida, todos os quadros são submetidos à filtragem do tipo passa-altas (filtro H_{em}), realizando-se a estimação de movimento na alta frequência dos quadros interpolados.

escalas, de forma que coeficientes nulos em determinada posição e escala sugerem que seus descendentes (coeficientes na mesma posição, mas em escala superior), também sejam nulos. Para se implementar esse processo de predição, tanto os KFs degradados quanto os NKFs serão submetidos a uma simples filtragem espacial passa-altas, apenas para fins de estimação de movimento, cuja máscara bidimensional é dada por:

$$H_{em} = \frac{1}{9} \begin{pmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{pmatrix}. \quad (1)$$

Ainda sobre o processo de estimação de movimento, ressalta-se que o algoritmo de busca foi implementado com tamanho de bloco variável. Dessa maneira, toda busca é iniciada com blocos de tamanho 16×16 , particionados até tamanho 8×8 , se a soma das SSDs desses sub-blocos for menor que a SSD do macrobloco.

É importante notar que não é necessário implementar um filtro passa-altas para obter a alta frequência dos KFs, que será somada aos NKFs. A alta frequência é simplesmente determinada fazendo-se a diferença entre o KF decodificado e sua versão decimada e interpolada. A Figura 4 ilustra esse procedimento de compensação descrito.

B. Determinação dos pesos de ponderação das altas frequências dos KFb e KfF

Como mencionado, no processo de compensação da alta frequência dos NKFs, as componentes alta frequência dos blocos de melhor casamento dos KFb e KfF serão ponderadas, usando-se pesos que reflitam a confiabilidade que se tem em cada uma delas. Para tal, propõe-se utilizar pesos inversamente proporcionais às SSDs (do inglês, *Sum of Squared Differences*), obtidas no processo de estimação de movimento. No entanto, deve-se, ainda, promover a normalização dos pesos para que não se altere a faixa de variação média. Obtêm-se assim as seguintes expressões:

$$p_b = \frac{SSD_f}{SSD_b + SSD_f} \quad (2)$$

$$p_f = \frac{SSD_b}{SSD_b + SSD_f}, \quad (3)$$

em que SSD_b e SSD_f são, respectivamente, as SSDs obtidas pela busca nos KFb e KfF. Uma ponderação similar também tem sido utilizada para o caso de se terem quadros com diferentes quantidades de distorção [8].

III. RESULTADOS

O método de super-resolução, descrito na Seção II, foi aplicado às seguintes sequências de vídeo no formato CIF: Foreman, Akyio e Silent. Nos testes, foram utilizados 30 quadros de cada sequência. As sequências foram codificadas utilizando o padrão H.264/AVC no modo Intra, isto é, utilizando somente predição espacial. Foram utilizados dois valores para o parâmetro de quantização (Qp), 20 e 28, e a quantidade de KFs foi uniformemente variada em cada sequência de quadros (1/30, 3/30, 6/30 e 15/30).

Durante o processo de super-resolução, para realizar a estimação de movimento nos quadros KFb e KfF, as buscas foram realizadas segundo o método de busca completa (*full search*) bipreditivo, limitando-se o tamanho da janela de busca a 32×32 pixels, para blocos de 16×16 pixels, e uma janela de busca de 16×16 pixels, para blocos de 8×8 pixels. Como ilustrado na Figura 4, as buscas foram realizadas na alta frequência dos quadros interpolados. Além disso, utilizou-se o critério de menor SSD para decidir o melhor casamento. As Figuras 5 a 7 ilustram os resultados obtidos, comparando-se o método de melhoramento por super-resolução proposto com o intuitivo método de simples interpolação dos NKFs (método interpolativo), utilizando o mesmo filtro Lanczos3.

A partir dos resultados apresentados, pode-se observar que o método de super-resolução sempre gera ganho, quando comparado ao método interpolativo, além de ser eficiente nos dois extremos de Qp utilizados. Deve-se também notar a efetiva melhoria promovida pelo método tanto para sequências de baixo movimento, representadas pela sequência Akyio, quanto para sequências de alto movimento, representadas pela sequência Foreman.

Comparando-se os resultados atuais com resultados de trabalhos anteriores [10], nota-se que foram obtidos ganhos da ordem de 4dB. Parte desse ganho deve ser atribuída à melhoria promovida nos quadros interpolados. Porém, podem ser apontados como fatores mais significativos de melhoria as novas

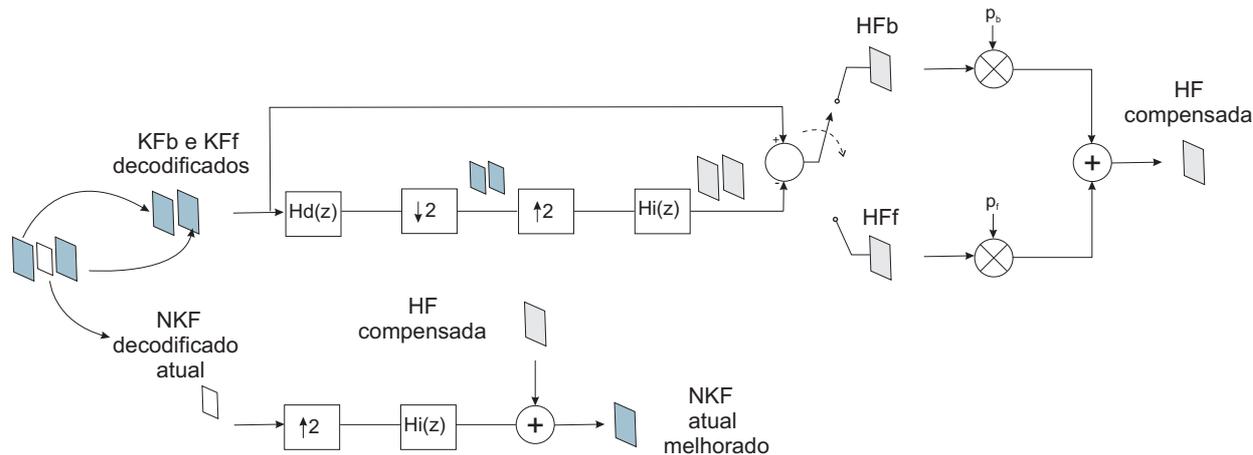


Fig. 4. Procedimentos de compensação da alta frequência dos NKF. A alta frequência dos KFs é obtida pela diferença entre os KFs decodificados e suas versões degradadas. A informação de alta frequência dos KFs e KfF são ponderadas e, por fim, adicionadas ao NKF interpolado, melhorando-o.

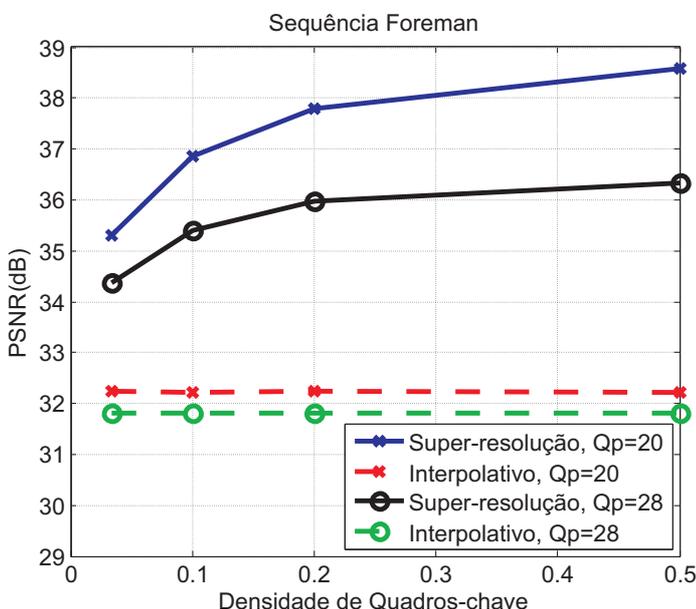


Fig. 5. Resultados do melhoramento por super-resolução para a sequência Foreman, comparados ao simples método interpolativo.

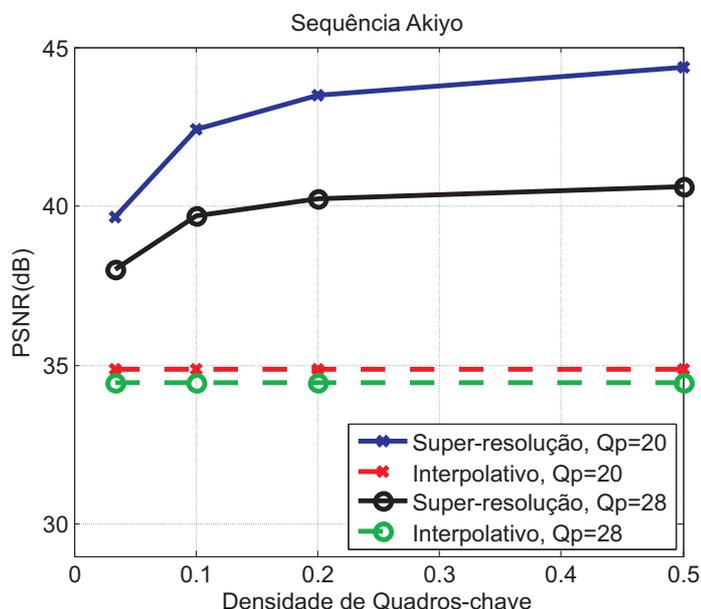


Fig. 6. Resultados do melhoramento por super-resolução para a sequência Akiyo, comparados ao simples método interpolativo.

técnicas de ponderação bipreditiva, com pesos determinados de maneira inversamente proporcional às SSDs; e a consideração de blocos de tamanho variável na estimação de movimento. Para que se possa ter idéia do incremento de ganho promovido por cada uma dessas novas contribuições, a Figura 8 ilustra, para a imagem Foreman e $Qp = 20$, a comparação entre o método de melhor desempenho (Método SR1), que considera busca com tamanho de bloco variável, com uma versão que usa apenas blocos fixos de 16×16 (Método SR2). Além disso, mostra-se também uma versão com bloco fixo que só considera como referência o KF de menor SSD (Método SR3), ao invés de ponderar as altas frequências dos dois KFs disponíveis. Por fim, pode-se também verificar uma queda no desempenho trocando-se os filtros Lanczos3 por subamostragem com média seguida de interpolação bicúbica (Método SR4).

IV. CONCLUSÕES

Esse trabalho apresenta um método para o melhoramento de quadros deteriorados por redução da resolução, baseado na compensação da alta frequência composta por ponderação de informação obtida dos quadros anterior e posterior que tenham resolução integral. Esse método constitui uma continuação de um trabalho anterior [10], apresentando como principais modificações: (i) a inclusão da informação de alta frequência de quadros anteriores e posteriores ponderadas e, não apenas, a do quadro de melhor casamento; (ii) a consideração de tamanho de bloco variável no processo de busca; (iii) e a troca dos filtros de decimação e de interpolação. Notadamente, as modificações sugeridas provocam ganhos na qualidade da ordem de 4dBs e, além disso, deve-se verificar a homogeneidade de ganhos para sequências consideradas de baixo e de

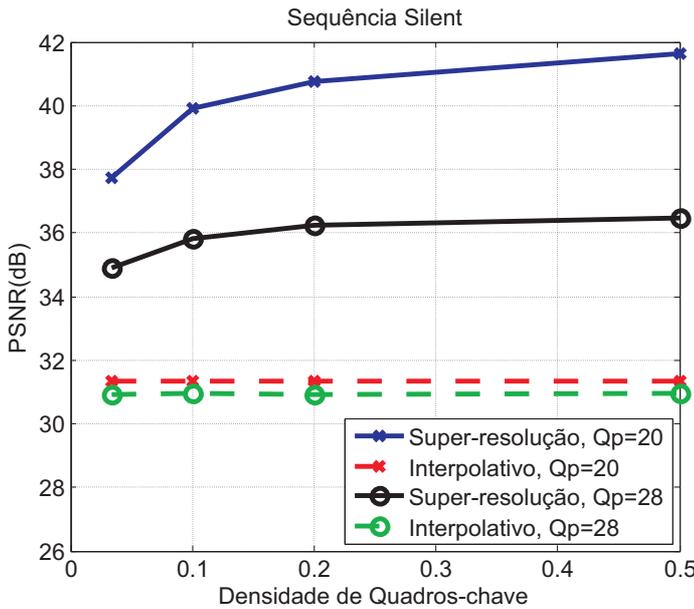


Fig. 7. Resultados do melhoramento por super-resolução para a sequência Silent, comparados ao simples método interpolativo.

alto movimento. Mesmo em situações de baixa densidade de quadros chave como, por exemplo, considerando-se a presença de apenas um KF em 30 quadros, é possível aumentar em mais de 2dBs a qualidade dos quadros puramente interpolados. Deve-se ressaltar a generalidade do método, apontando sua aplicação a qualquer sistema que apresente quadros com diferente resolução. De pronto, duas aplicações específicas podem ser indicadas: melhoria no processo de geração de informação lateral em codificadores distribuídos [5]; e em sistemas com qualidade variável [8].

REFERÊNCIAS

[1] A. K. Katsaggelos, R. Molina e J. Mateos, "Super Resolution of Images and Video". *Synthesis Lectures on Image, Video, and Multimedia Processing*. ISBN 978-1598290844, Morgan and Claypool Publishers, 2007.

[2] C. A. Segall, A. K. Katsaggelos, R. Molina e J. Mateos, "Bayesian Resolution Enhancement of Compressed Video". *IEEE Transactions on image processing*, vol. 13, no. 7, 2004.

[3] W. T. Freeman, T. R. Jones e E. C. Pasztor, "Example-based super-resolution", *IEEE Computer Graphics and Applications*, vol. 22, pp. 56-65, 2002.

[4] H. Schwarz, D. Marpe e T. Wiegand, "Overview of the scalable video coding extension of the H.264/AVC standard", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 9, pp. 1103-1120, 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. IS&T/SPIE Symp. on Electronic Imaging, Visual Communications and Image Processing*, vol. 6508, pp. 65081Y1-65081Y12, 2007.

[6] B. Macchiavello, D. Mukherjee e R. L. De Queiroz, "Iterative side-information generation in a mixed resolution Wyner-Ziv framework," a ser publicado, *IEEE Trans. Circuits and Systems for Video Technology*, 2009.

[7] D. Mukherjee, "A robust reversed-complexity Wyner-Ziv video coding mode based on a spatial reduction framework", *HP Labs Technical Report*, HPL-2006-80, 2006.

[8] E. M. Hung, R. L. de Queiroz e D. Mukherjee, "Codificação de vídeo com complexidade reversa utilizando qualidade mista", *submetido ao Simpósio Brasileiro de Telecomunicações (SBRT 2009)*.

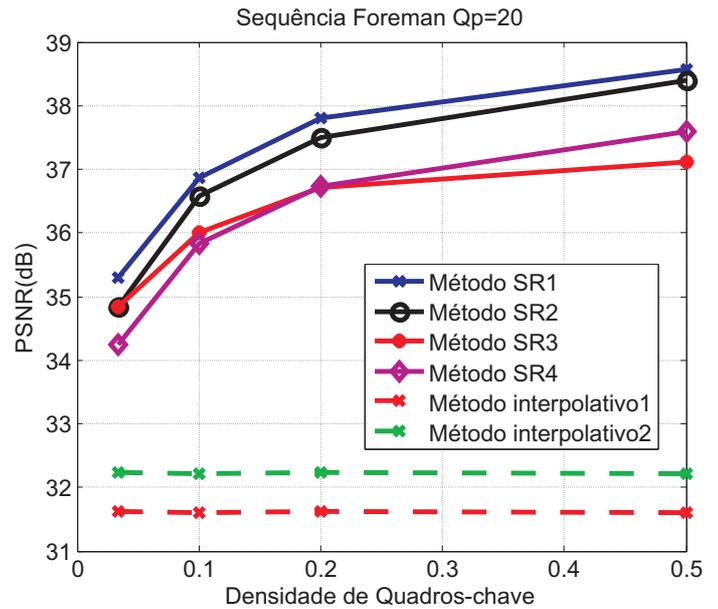


Fig. 8. Comparação entre os resultados alcançados pelas diferentes contribuições oferecidas. O método SR1 é o resultante da utilização de bloco variável, bipredição e filtros Lanczos3 para subamostragem e interpolação. O método SR2 é também bipreditivo e se utiliza dos filtros Lanczos3, mas contempla apenas blocos de tamanho fixo. O método SR3 utiliza bloco fixo e filtros Lanczos 3, mas é apenas preditivo. Já o método SR4 utiliza bloco fixo e é bipreditivo, porém se utiliza de subamostragem com média e interpolação bicúbica. O método interpolativo 1 usa subamostragem com média e interpolação bicúbica, enquanto o interpolativo 2 usa filtros Lanczos3.

[9] B. Macchiavello, F. Brandi, E. Peixoto, R. L. de Queiroz e D. Mukherjee, "Side-information generation for temporal and spatial scalable Wyner-Ziv codecs", *EURASIP Journal of Image and Video Processing*, vol. 2009, pp.1-11, 2009.

[10] F. Brandi, R. L. de Queiroz e D. Mukherjee, "Super resolution of video using key frames and motion estimation", *Proc. of International Conference on Image Processing*, 2008.

[11] P.P. Vaidyanayhan, "Multirate Systems and Filter Banks", Prentice-Hall Signal Processing Series, 2001.

[12] I. E. G. Richardson, "Video Codec Design: Developing Image and Video Compression Systems", Wiley, 2002.

[13] J. M. Shapiro, "Embedded image coding using zerotrees of wavelets coefficients", *IEEE Transactions on Signal Processing*, SP-41:3445-3462, 1993.

[14] A. Said e W. A. Pearlman, "A new fast and efficient coder based on set partitioning in hierarchical trees", *IEEE Transactions on Circuits and Systems for Video Technology*, pp. 243-250, 1996.