SUPER-RESOLUTION OF VIDEO USING KEY FRAMES AND MOTION ESTIMATION

Fernanda Brandi[1], Ricardo de Queiroz[1] and Debargha Mukherjee[2]

[1]Departamento de Engenharia Elétrica
Universidade de Brasília
Brasília, Brazil
fernanda,queiroz@image.unb.br

[2]Hewlett Packard Laboratories
Palo Alto, CA, USA
debargha.mukherjee@hp.com

ABSTRACT

Many scalable video coding systems use frame down-sampling in order to reduce complexity and to enable enhancement layers. Super-resolution (SR) can be used to help the up-sampling and recovering processes of those frames. We are interested in reversed-complexity (distributed) coding methods, wherein few key frames are encoded at normal resolution, while the rest are down-sampled and encoded at reduced resolution along with the enhancement layers. We are only interested on the decoder side, wherein we carry motion estimation of the down-sampled frames using the key frames as references. When a match is made, the high-frequency components of the key frames (KF) are used to super-resolve the non-key frames (NKF). The motion estimation process is performed using blocks of band-pass versions of the frames, rather than low-pass ones. Results indicate the improved performance of the proposed super-resolution algorithm.

Index Terms— Super-Resolution, Reversed-Complexity Video Coding, Key Frames, Motion Estimation

1. INTRODUCTION

Video compression usually involves quantization parameters to control the balance between image quality and bit rate. Further control can be achieved if one down-samples some frames of a video sequence at the encoder to achieve higher compression rates (see Fig. 1). Such multiple resolution approaches are being used in many scalable video schemes [1],[2]. Two main advantages of proceeding in this way are the reduction of both complexity at the encoder and the information to be compressed. The collateral effect is a decrease in image quality since information is lost in the quantization and sub-sampling processes.

The sub-sampled frames have to be up-sampled at the decoder so they can achieve full resolution. However, since some frame information is discarded at the encoder, there are losses after interpolation. We use super-resolution methods for increasing the reconstructed image quality.

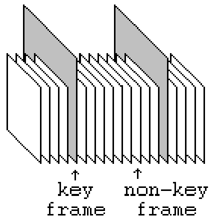


Figure 1. Illustration of key frames in scalable video with spatial down-sampling. Key frames are at full resolution and non-key frames are at a reduced resolution.

Depending on the encoding scheme or limitations, a distinct super-resolution technique can be used. In distributed video coding (DVC) [3] certain frameworks use low resolution coding for complexity reduction at the encoder side [4],[5]. For the case of reversed-complexity video coding, the usual approaches based on Bayesian frameworks [6],[7] may not be suitable because the complexity, in this method, is concentrated at the encoder. As an alternative, a super-resolution technique that works at the decoder side is preferred.

In this paper, a new method of super-resolution for a reversed-complexity video coding schemes is presented. Our method does not need any training, databases [8],[9], nor assumes probability distributions. Super-resolution using KF takes advantage of the fact that one might have enough information to recover NKF (encoded at low resolution) with the aid of few KF (encoded at full resolution), since they are largely correlated. In this way, the method reduces complexity at the encoder, increases compression rates, and improves the quality of decoded frames.

2. PROPOSED SUPER RESOLUTION METHOD

The framework can be summarized in Fig. 2. At the encoder side (Fig. 2a), some frames are transmitted at full resolution (KF) and others are sub-sampled before encoding (NKF). At the decoder side, the NKF are up-sampled to full resolution and super-resolved with the aid of the KF.

This work was supported by HP Brasil and by Conselho Nacional de Desenvolvimento Científico e Tecnológico, CNPq, Brasil, under grant 473696/2007-0.

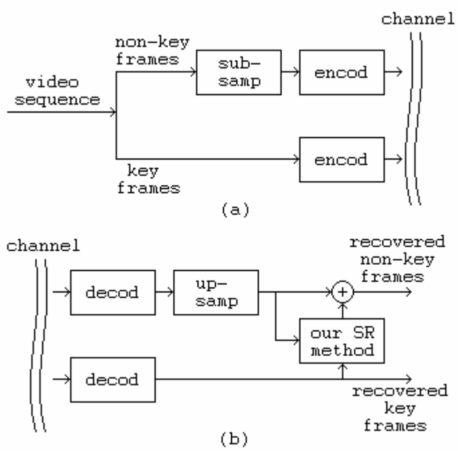


Figure 2. Simplified scheme of (a) encoding and (b) SR-based decoding.

The super-resolution method presented here was inspired by [9],[10]. The sub-sampling by up-sampling process tends to preserve the low-frequency but reduces the high-frequency energy of the image. With that in mind, a super-resolution method is proposed to help recovering only the high-frequency loss since the low-frequency is still there and should not be discarded or replaced.

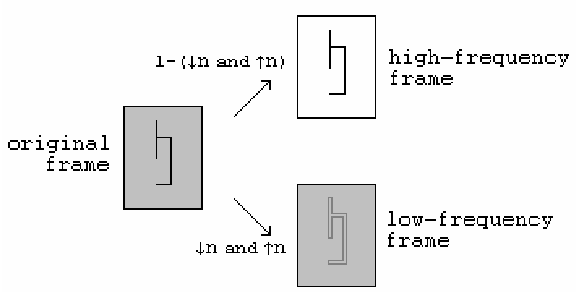


Figure 3. Illustration of the frequency bands division of one given frame. Original frame and its low-frequency (due to decimation and interpolation by a given factor *n*) and high-frequency (the complement) versions.

As in Fig. 3, we can separate a frame into two distinct frequency bands: low- and high-frequency. The low-frequency band is found by the down- and up-sampling processes to which a frame is subject. The high-frequency band is the complement of the frequency response of this low-pass filter. At the decoder side, the high-frequency of the NKF is unknown. We need to estimate it and add it to the low frequency band in order to recover the frame.

The KF have all the low- and high-frequency bands, while the NKF have just the low-frequency band. Our SR method proposes to take advantage of the high-frequency present in the KF to recover the NKF missing high-frequency band (Fig. 4). The proposed method uses motion estimation to find the best-match between a block from the NKF and the KF. When the best-match is found, we add the high-frequency of the KF block into the NKF block yielding a super-resolved version of this last one.

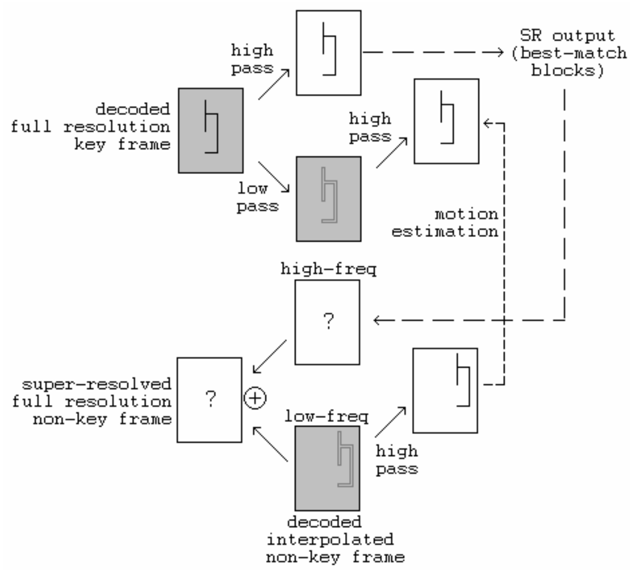


Figure 4. Super-resolution scheme using key frames and motion estimation at the decoder to find high-frequency best-matches between key and non-key frames. After the motion estimation, we have our SR output which is a set of best-match high-frequency key blocks to create an estimated high-frequency for the non-key frame.

At the decoder side (Fig. 2b), we interpolate the NKFs. We also sub- and up-sample the KFs to simulate the losses suffered by the NKFs (indicated in Fig. 4 as a "low-pass" filter). To perform the motion estimation, instead of operating on the low-pass versions of the KF and NKF, we do high-pass the low-pass version and then search for matches. In effect, this band-pass information better predicts the high-frequency bands (like cross-scale prediction in wavelet compression [11],[12]). When a best-match is found, the original key block high-frequency component is added to the non-key block (see Fig. 4).

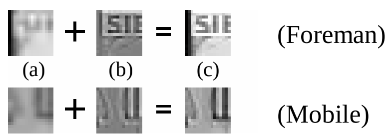
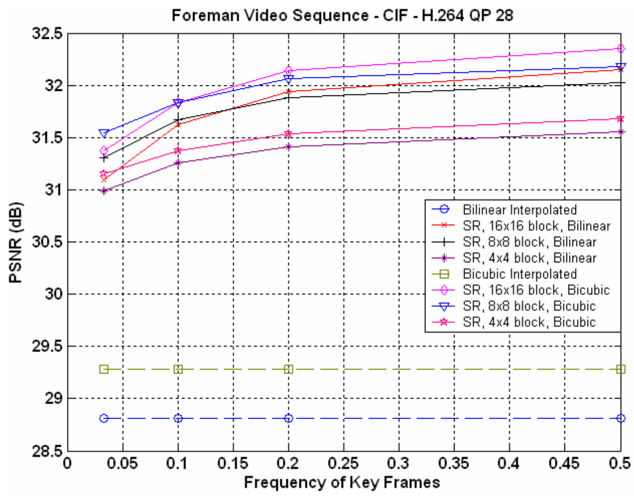


Figure 5. Magnified illustration of 16x16 pixels blocks for Foreman and Mobile sequences. Examples showing the sum of (a) low- and (b) high-frequency blocks yielding (c) super-resolved ones.

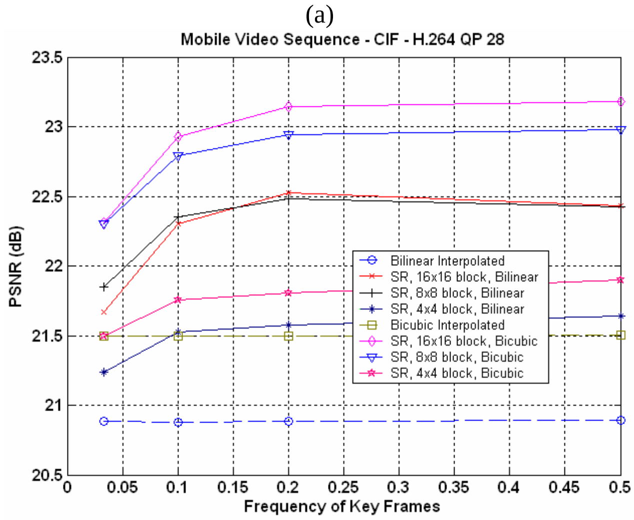
Thus, the NKF would be super-resolved (Fig. 5c) by integrating the matching KF high-frequency band (Fig. 5b) to the interpolated frame (Fig. 5a). Ideally, the high-frequency added into the NKF would be the complement of the low-frequency part (Fig. 3). However, since usually the best-match is not perfect, adding the complement of the low-frequency would certainly add noise to the NKF. Hence, instead of adding all the information, the high-frequency added is the result of the convolution of the original KF with a simple high-pass kernel (which is somehow a ponderation of the complement). The best filter in this case is still under investigation.

3. EXPERIMENTAL RESULTS

To evaluate the proposed method we ran two sets of tests. First we applied our method to CIF video sequences, reducing the non-key frames to QCIF (using bilinear filters). We did use H.264 intra coding with quantization parameter (QP) set to 28 to encode every frame (both full and reduced resolution). We also tested various block sizes (16x16, 8x8 and 4x4) and different densities of uniformly spaced KF (1/30, 1/10, 1/5 and 1/2, which is the proportion of KF to the total number of frames). Motion estimation was carried using bi-prediction with full search on a limited 32x32-pixel window. In order to estimate the best-matches we search within filtered blocks (see Fig. 4). We use the sum of squared differences (SSD) as best-match criterion. In Figs. 6 and 7 the results of super-resolving both bilinear and bicubic interpolated frames are presented.

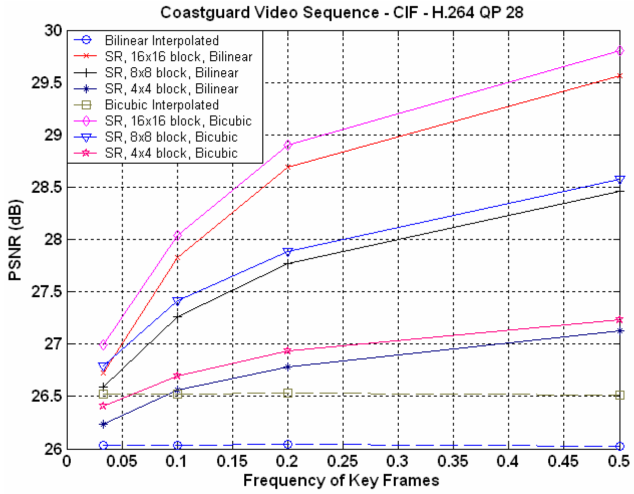


(a)

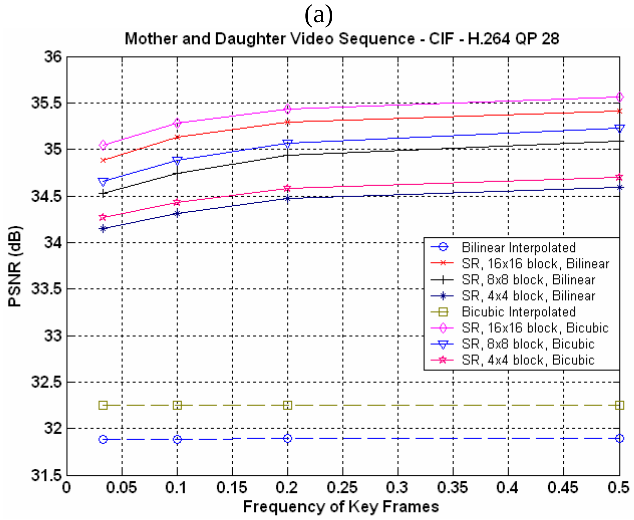


(b)

Figure 6. Super-resolution results for (a) Foreman, (b) Mobile.



(a)



(b)

Figure 7. Super-resolution results for (a) Costguard and (b) Mother and Daughter.

In a second test (see Fig. 8) we varied the QP at the encoder and fixed the density of key frames at 0.5, i.e. every other frame is a KF. We tested different block sizes and interpolation methods for sequence "Foreman".

A subjective comparison can be made with processed frames of the "Coastguard" and "Mobile" sequences. Enlarged portions of those sequences are presented in Fig. 9. In those tests, it was used a 1:1 proportion of KF to NKF, and frames were encoded using H.264 intra coding with QP set to 28.

Have in mind that our system is not supposed to stand on its own as a coding mechanism. The super-resolved images are subject to enhancement layers. Those layers and the overall enconding are not the focus of this paper.

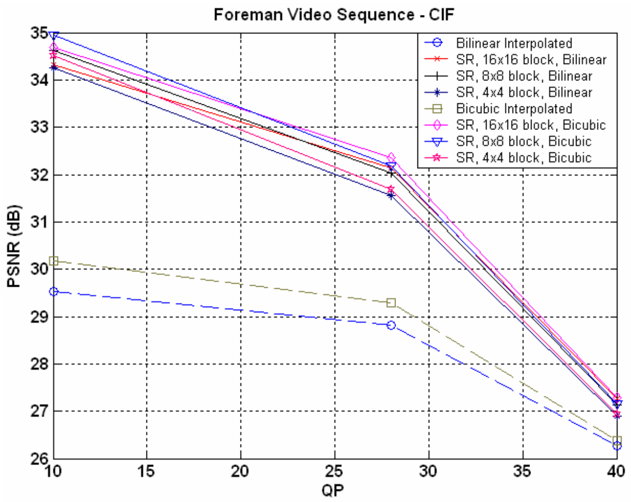


Figure 8. Results varying H.264 quantization parameter for a 1:1 proportion of KF to NKF.

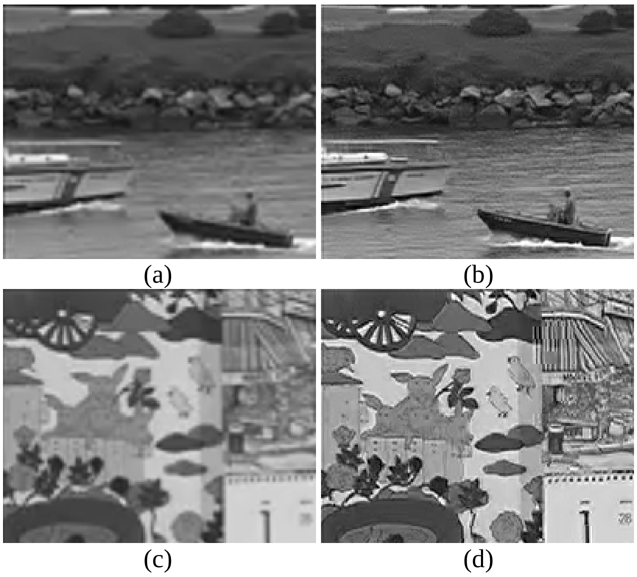


Figure 9. Subjective results for Coastguard's frame 20: (a) interpolated and (b) super-resolved frame; and for Mobile's frame 20: (c) interpolated and (d) super-resolved frame.

4. CONCLUSIONS

Our SR scheme does not need any training, databases nor assumes probability distributions. It is meant to super-resolve frames on DVC systems with multiple resolution frames since our SR method explores the correlation between key frames and non-key frames.

We could observe improvements of the super-resolved frames over interpolated ones showing that our SR method could be used in reversed-complexity video coders wherein there are higher resolution frames among lower resolution ones. The method is robust enough to improve even strongly degraded (quantized) frames or to work well even with few key frames. Future research will be directed at adaptively choosing the best block size and to optimize the high-pass filter applied prior to motion estimation.

5. REFERENCES

[1] H. Schwarz, D. Marpe, and 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, Sept. 2007.

[2] D. Mukherjee, B. Macchiavelo, and 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*, San Jose, CA, USA, SPIE Vol. 6508, pp. 65081Y1-65081Y12, Jan. 2007.

[3] B. Girod, A. Aaron, S. Rane and D. Rebollo-Monedero, "Distributed video coding," *Proceedings of the IEEE*, vol. 93, no 1, pp. 71-83, Jan. 2005.

[4] B. Macchiavelo, R. L. de Queiroz, and D. Mukherjee, "Motion based side information generation for a scalable Wyner-Ziv video coder," *Proc. IEEE Intl. Conf. on Image Processing, ICIP*, San Antonio, TX, USA, Sept. 2007.

[5] Z. Li and E. J. Delp "Wyner-Ziv video side estimator: conventional motion search methods revisited," *Proc. IEEE International Conf on Image Processing*, vol. 1, p. 825-828, 2005.

[6] C. A. Segall, R. Molina, and A. K. Katsaggelos, "High-resolution images from low-resolution compressed video," *IEEE Signal Process. Mag*., Vol 20, , pp. 37-48, 2003.

[7] A. K. Katsaggelos, R. Molina and J. Mateos, *Super-resolution of images and video*, San Rafael, CA, US, Morgan & Claypool Publishers, 1st ed., 2007.

[8] F. Brandi, R. De Queiroz and D. Mukherjee, "Super-resolution of video using key frames," *IEEE International Symposium on Circuits and Systems*, 2008

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

[10] E. Shechtman, Y. Caspi, and M. Irani, "Space-time super-resolution," *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 27, no. 4, pp.531–545, April 2005.

[11] J. M. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients", *IEEE Trans. on Signal Processing*, Vol. 41, no. 12, pp. 3445-3463, Dec. 1993.

[12] N. Adami, M. Brescianini, M. Dalai, R. Leonardi and A. Signoroni, "A fully scalable video coder with inter-scale wavelet prediction and morphological coding", *Visual Communication and Image Processing, VCIP*, SPIE vol. 5960, no. 58, July 2005.