Super Resolution of Video Using Key Frames

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

[1]Electrical Engineering Dept.
Universidade de Brasília
Brasília, Brazil
fernanda,queiroz@image.unb.br

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

Abstract — **In many video compression systems, the frames are down-sampled before transmission. Also, in many scalable systems, the residual after down- and up-sampling is encoded and transmitted. Sometimes, a few frames are encoded at normal resolution (key frames) while the other frames are encoded at reduced resolution. Super resolution can be used to enhance the up-sampling process, using motion information to improve traditional interpolation. In this paper, we propose to use a super resolution method to up-sample the non-key frames using the key frames as reference. We build dictionaries on-the-fly using the key frames instead of the traditional off-line training images. The high-frequency data of matching blocks are added to the low-resolution blocks. Since the key frames are very similar to the non-key frames, the method is robust enough to allow successful super resolution of highly compressed (severely degraded) sequences. Results are presented for many predefined block sizes, key frames frequencies, and compression parameters.**

I. INTRODUCTION

Image and video compression is all about the trade off between image quality and bit rate. One effective way to achieve high compression rates with reasonable quality is to down-sample the image (or images in the case of video sequences) before compression and transmission. Typically, one increases compression ratio by applying more aggressive quantization. Nevertheless, there is usually a compression ratio breakpoint above which it becomes more advantageous to increase compression through down-sampling. In order to try to get the best of both worlds, many compression schemes are scalable, i.e. they are based on multiple resolutions of the images (or frames, in video). They can increase compression either through down-sampling or through quantization. In effect, typical scalable video schemes can achieve both spatial and temporal scalability [1],[2].

Scalable coders generally work as follows. The video is down-sampled (either spatially or temporally) and interpolated back. The down-sampled version is transmitted along with the residue (error) between the original and interpolated images. A few scalable coders make use of key frames. Key frames are frames which are not down-sampled or removed (in case of temporal down-sampling). They are compressed at full resolution and are used to estimate the missing information. The scheme is depicted in Fig. 1, wherein non-key frames are spatially down-sampled or removed, for temporal down-sampling.

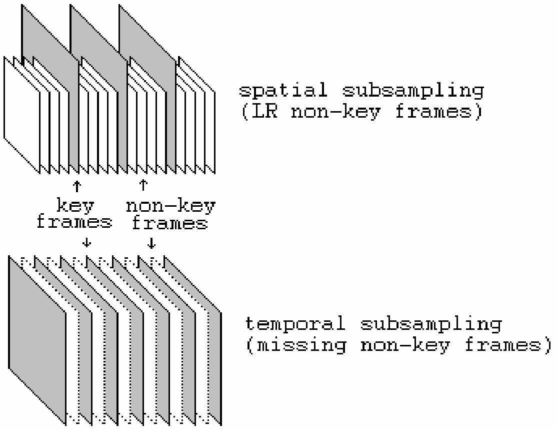


Figure 1. Illustration of key frames in scalable video. Spatial and temporal down-sampling demands posterior spatial or temporal interpolation.

Despite the focus on compression so far, this paper is not about compression. We are solely interested in the image up-sampling process, used to reconstruct the non-key frames in scalable coding.

II. SUPER RESOLUTION AND THE KEY FRAMES

Super resolution (SR) is the method of enhancing resolution obtaining high resolution (HR) frames from a set of low resolution (LR) ones [3]. Usually the SR methods involve Bayesian frameworks [4].

In this paper, we present a SR method with a distinct approach which does not need training nor implies probability distributions. We assume that key frames at a high resolution are available to help us to super-resolve the video frames. In this sense, we say our method is semi super resolution (SSR), i.e. we achieve higher resolution with the aid of other high resolution images. Our SSR generation scheme can be used in applications like video coders with spatial scalability and even in some cases for temporal scalability. For example, in

This work was partially supported by HP Brasil and by the Conselho Nacional de Desenvolvimento Científico e Tecnológico.

distributed video coding (DVC) [5] certain frameworks use low resolution coding for complexity reduction at the encoder side [6]. Our approach can improve the decoded interpolated frame. Since this is a crucial process in DVC [7], the SSR can lead to a better performance of the codec. Also, a similar approach for SR generation can be used to improve the zoom tool in a video sequence. In fact, our SSR scheme could be used in any situation where one or more high resolution images are available for further reference so the interpolation of the low resolution images can be improved.

III. Proposed Semi Super Resolution Method

Our method is example-based [8], where we seek to restore the high-frequency information of an interpolated block through searching in a database for a similar block, and by adding the high-frequency of the chosen block to the interpolated one. The database is typically populated with training images. However, in our case, we have access to HR frames and we do not need to construct unrelated training databases. Hence, we use the high-frequency of key HR frames to increase the quality of non-key restored frames. Our approach can be better explained using Figs. 2 and 3.

Let the original sequence of frames at a high resolution have both key frames *A* and non-key frames *B*. All frames can be compressed with codecs such as H.264 or H.263, for example. We do not want to transmit the non-key frames *B* at full resolution so we down-sample them at the encoder, yielding LR frames B_L (Fig. 2a). Frames *A* and B_L are encoded and sent.

At the decoder (Fig. 2b), we receive the video sequence with *A* and B_L. First, we decode them and store the key frames *A* since they already are at the desired resolution. Then, we subsample the key frames (in the same process applied to the *B* frames at the encoder) yielding A_L. At this point we have both key and non-key frames at LR, so we interpolate all these frames to the original resolution (A_H and B_H).

The next step is to filter B_H with a high-pass filter yielding $F(B_H)$. High energy blocks of $F(B_H)$ contain high-frequency components and are candidates to be improved. We use a threshold parameter *T* to decide whether a given block *k* contains sufficient high-frequency energy or not, i.e. we test whether

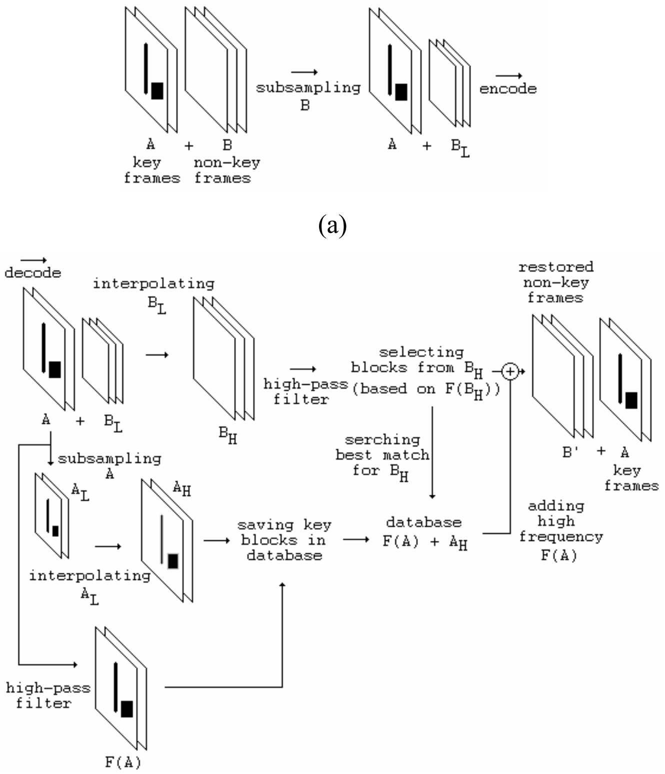
$$\| F(B_H)^{(k)} \| > T \qquad (1)$$

where $F(B_H)^{(k)}$ denotes the *k*-th block of $F(B_H)$. If the block passes the above test it is super-resolved. It is important to note that no block of $F(B_H)$ is actually used in the reconstruction, just in the decision process.

In order to carry our SSR example-based method, we need to build a database. For that, we also high-pass filter the key frames *A*, yielding *F(A)*. If

$$\| F(A)^{(k)} \| > T \qquad (2)$$

then the k-th block of *F(A)* is selected to be part of the database. The database is constructed in pairs, i.e. along with the selected $F(A)^{(k)}$ we also save in the database the corresponding block of A_H.



(a)

(b)

Figure 2. Frame processing at (a) encoder and (b) decoder.

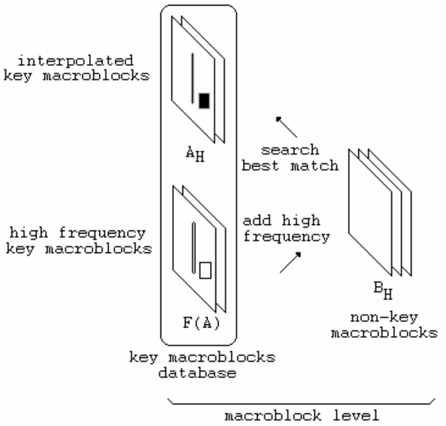


Figure 3. After searching for a best match in the database, we add the corresponding high-frequency to the block to be super-resolved.

The database is updated with all blocks of interest in *F(A)* and their respective interpolated versions A_H are saved for every incoming key-frame. In summary, the database has the selected block pairs ($F(A)^{(k)}$, $A_H^{(k)}$). It can have a given size so that if the limit is reached, older block pairs are excluded to give way for those of a newer key frame.

We search the database of A_H blocks for a best match for each selected block from B_H, for example using the sum of absolute differences (SAD), as illustrated in Fig. 3. When the best match is found between blocks of B_H and A_H, the associated high-frequency block of *A*, *F(A),* is added to B_H. In other words, the original high-frequency component of *A* is added to the interpolated non-key block B_H yielding the super-resolved block *B'*. The process is repeated until the last selected non-key block is super-resolved. In the end, we have the key frames *A* at full original resolution and non-key super-resolved frames *B'*.

By adding high-frequency information instead of replacing the whole block, we assure that the natural characteristics of the interpolated frame are somewhat preserved. Within this SSR framework we can adjust the high-frequency energy threshold parameter and the block size. We could also change the frequency of key frames and the compression factor at the encoder. These choices can be made depending on the application and on system limitations.

IV. Experimental Results

To test how robust our scheme is and to evaluate the process we did apply our SR method on different lossy encoded CIF sequences. It was used the bilinear down-sampling process – both at the encoder and decoder – decreasing the resolution at a ratio of 2 in each direction yielding QCIF frames. To up-sample the frames back to CIF resolution we carried bi-cubic interpolation.

The high-frequency band of a given frame is the result of the convolution of that frame (original or interpolated) with a simple high-pass kernel M_h, i.e, for a given image *I*:

$$F(I) = I * M_h = I * \begin{pmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{pmatrix} \qquad (3)$$

The high-frequency threshold was T = 0.002 and the energy is the square of the L_2-norm of the block. In case there is not enough high-frequency, the block is not super-resolved (if it is a non-key-frame) or is not included in the database (if it is a key frame). The threshold not only reduces complexity, but also avoids adding high-frequency where originally there was none.

We present objective SR results in Figs. 4 through 6. In Fig. 4 we super-resolved the "Foreman" sequence fixing the density of key frames (the proportion of key frames to the total number of frames) in 1/2 and varying the H.264 quantization parameter and block sizes. In Figs. 5 and 6 we varied block sizes (16x16, 8x8, 4x4) and tested different densities of key frames (1/30, 1/10, 1/5 and ½).

We also present a subjective comparison among interpolated and enhanced frames of "Coastguard" and "Mother and Daugher" sequences. In Fig. 7 it is shown enlarged portions of those frames both interpolated and super-resolved. In these tests, it was used a 1:1 proportion of key frames to non-key frames, and frames were encoded using H.264 intra coding with QP set to 28. The block sizes to super-resolve "Mobile" and "Mother and Daughter" were both 16x16.

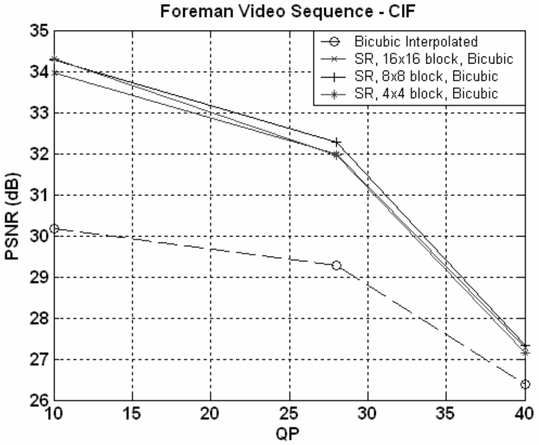
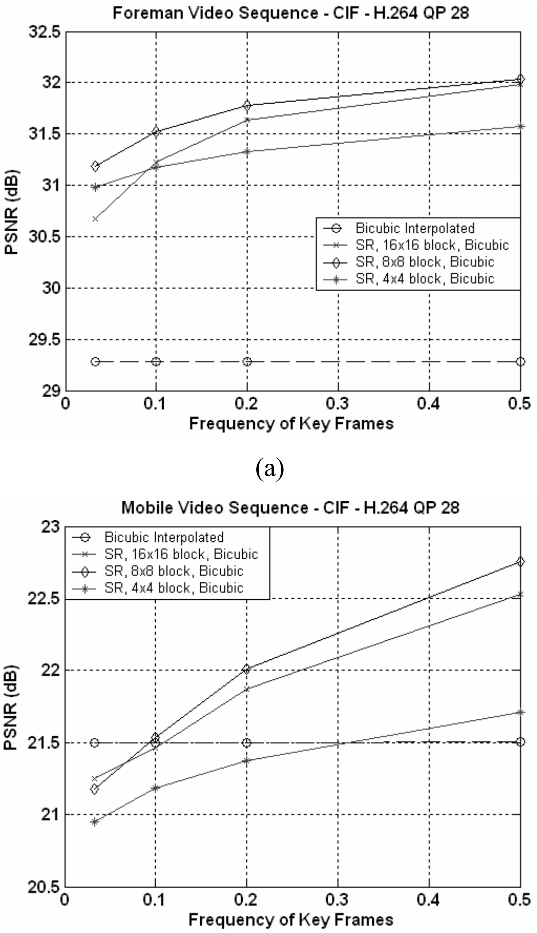


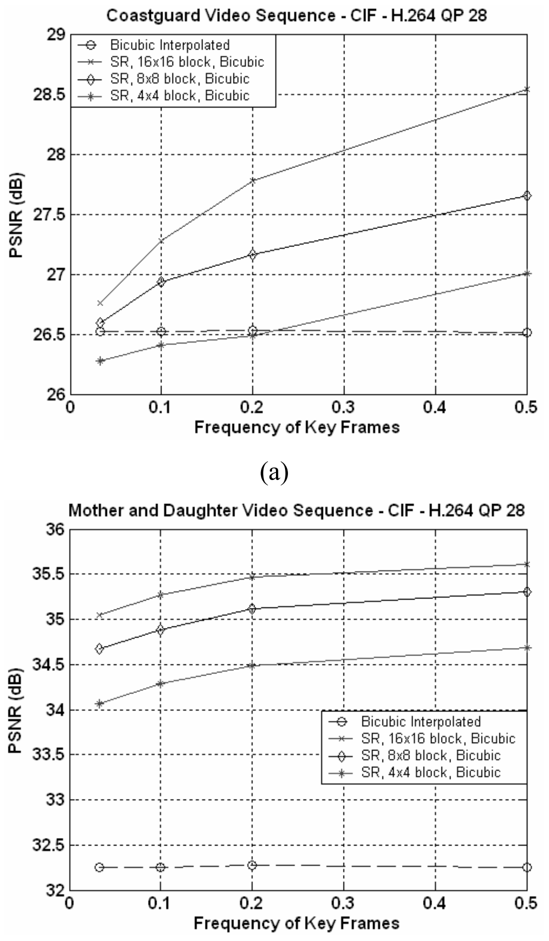
Figure 4. Results varying H.264 quantization parameter for a density of key frames equals to 1/2.



(a)

(b)

Figure 5. PSNR plots comparing interpolation and SSR for different block sizes and density of key frames. Resultas for sequences (a) Foreman and (b) Mobile.



(a)

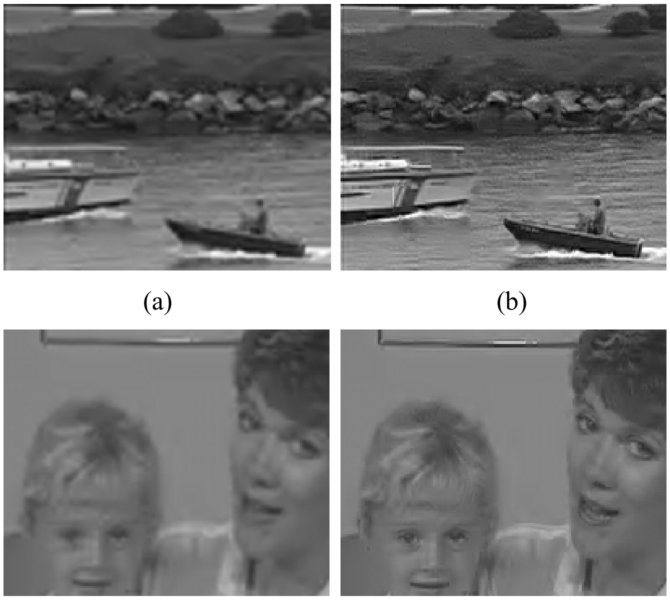
(b)

Figure 6. PSNR plots comparing interpolation and SSR for different block sizes and density of key frames. Resultas for sequences (a) Coastguard and (b) Mother and Daughter.

Based on the curves in Figs. 5 and 6 it is natural to think that an adaptive block size approach could be rather efficient. This approach could decide which block size to use for the whole frame. We can also analyze frame regions and make the block size adaptive within a frame.

V. CONCLUSIONS

Our semi super resolution scheme has been shown to be effective in real compression scenarios, wherein there are higher resolution frames among lower resolution ones. Many tests were carried with variable compression and configuration parameters. The method does not need training nor implies data distributions. Since the key frames are extremely correlated with the non-key frames, the database fits well to the super-resolved frames. The complexity is concentrated at the decoder side, which is suitable to reversed-complexity video coders [2]. Still, there is room for improvements such as adaptive block sizes and to develop a high-pass filter that better suits each image or video sequence. Other improvements could be to take advantage of super-resolved frames to help super-resolving others non-key frames and the use of motion estimation instead of building a database.



(a) (b)

(c) (d)

Figure 7. Subjective results for Coastguard's frame 20: (a) interpolated and (b) super-resolved frame; and for Mother and Daughter's frame 20: (c) interpolated and (d) super-resolved frame.

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

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.