Blocking-Effect Reduction in a Reversed-Complexity Video Codec Based on a Mixed-Quality Framework

Edson M. Hung
Universidade de Brasilia
Campus Gama
Brasilia, Brazil
Email: mintsu@image.unb.br

Ricardo L. de Queiroz
Departamento de Ciencia da Computacao
Universidade de Brasilia
Brasilia, Brazil
Email: queiroz@ieee.org

***Abstract*—This paper presents a blocking-effect reduction in a method for reversed-complexity video codec. We used intra predicted frames encoded at different quality (distortion) targets. We propose a technique to improve the enhancement layer of a mixed-quality encoded sequence, using information from the high-quality (key) frames to enhance the low-quality (non-key) ones at the decoder. The results show that it is possible to subjectively reduce the blocking effect by using overlapped motion compensation at the enhancement layer. It is shown that the subjective quality is improved and the objective quality is incremented at the non-key frames.**

***Index Terms*—Video codec, reversity-complexity, blocking-effect, overlapped motion compensation.**

I. Introduction

In a previous work [1], we proposed an architecture of mixed quality video codec, i.e., frames with time-varying quality targets. Then, we try to enhance the higher-distortion (non-key) frames using the information contained in the lower-distortion (key) frames. This enhancement method has its roots on block-based motion estimation and motion compensation, which typically generates blocking artifacts [2]. This paper proposes a method to reduce these artifacts by applying overlapping blocks in motion compensation. The H.264 deblocking filter [3],[4] is an adaptive filter based on several parameters to filter the pixels in spatial domain. We propose to use a non-adaptive filter that is applied only to the enhancement layer.

This architecture intends to yield a reversed complexity codec, but without using any additional corretaled information from another source (that do not communicate to each other during encoding but are jointly decoded) that would characterize the proposed method as distributed source coding [5]. One possibility to turn the proposed method into a distributed video codec is to add a Wyner-Ziv layer [6],[7] or any other separately encoded enhancement layer that improves the visual quality [8]. At these cases, the decoder is more complex than the encoder. For typical digital video coding standards [2],[3], encoding is more complex than decoding due to operations such as the transform and the intra- and inter-frame (motion estimation and motion compensation) predictions. These predictions are chosen based on the minimization of a rate-distortion function cost. The prediction mode is encoded, along with residual information, if necessary. By using a mixed quality approach, we can reduce the encoded video bit rate, while, at the decoder, exploring the temporal redundancy, we enhance the low-quality frames based on the high-quality ones.

II. Mixed-Quality Framework

As illustrated Figure 1, we have two types of frames with different quantization (Q): the key frames with a better quality (Q_{key}) and the non-key frames with a reduced quality ($Q_{non-key} > Q_{key}$). The arrangement of key frames and non-key frames is defined as a GOP (group of pictures). The bit stream is decoded using any ordinary decoder. An optional post-process enhancement layer can be generated and applied to the non-key frames.

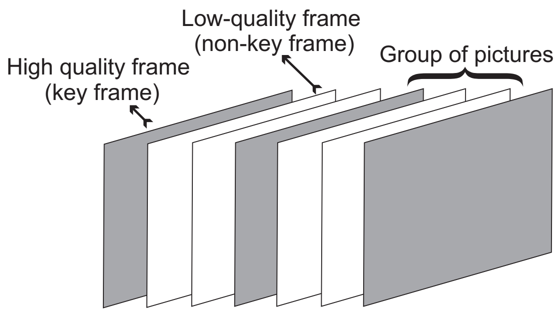


Fig. 1. Decoder side enhancement for a mixed quality video.

The enhancement method is inspired by a previous work [9], where a semi-super resolution is applied to a mixed resolution video codec [10]. In this work, instead of using non-key frames at a lower resolution, we apply a higher quantization parameter to the non-key frames compared to the key ones. Then, we try to enhance the higher-distortion non-key frames using the information contained in the lower-distortion key frames. The scheme is also similar to the techniques presented by Segall *et al*[11] and is illustrated in Figure 2.

We use a regular decoder that separates key-frames from non-key frames, as shown in Figure 2. Let a given non-key frame be denoted as $F_{non-key}$. Let this frame be enhanced bidirectionally by two key-frames $\left\{F_{key,(1)}, F_{key,(2)}\right\}$. Then,

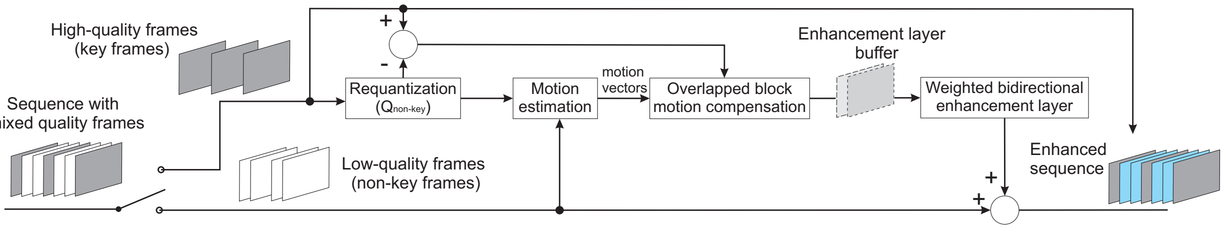


Fig. 2. The proposed architecture for enhancement at the decoder.

a requantization operation (with $Q_{non-key}$) is applied to the key frames resulting in a new pair of "low-quality" key frames: $\left\{F_{LQ\,key,\,(1)}, F_{LQ\,key,\,(2)}\right\}$. The layer $\widetilde{L}_k = F_{key,\,(k)} - F_{LQ\,key,\,(k)}$ represents the information lost through requantizing the k-th key frame, where $k \in \{1,2\}$. $\widetilde{L}_k$ is subject to motion compensation before applying it to enhance a non-key frame, due to temporal disparity. In this work, we use windowed overlapped block motion compensation (OBMC) [12] [13] in order to reduce the blocking artifacts. In order to illustrate the efficcacy of the proposed method, a subjective comparison using raw (uncompressed) sequence is shown at Figure 3. Motion estimation (M_E) is performed at the decoder between the frames $F_{LQ\,key}$ and $F_{non-key}$ using variable block size (16 $\times$ 16-pixels partitioned down to 8 $\times$ 8). The actual frame is divided into blocks. For each one, we look for the best-match block within a displacement window at the reference frame. The criteria may be the minimization of the SAD (sum of absolute differences) or SSD (sum of squared differences).

Note that the set of candidates that minimize the difference between the current (non-key) frame and the low quality key frame in the motion estimation using 16×16-pixels macroblocks is a sub-set of the partitioned blocks of 8×8-pixels. That could induce us to choose a smaller block size to perform the motion estimation. However, we empirically verified that the 16×16-pixel blocks yield better overall results. Differently from the motion estimation during encoding process, we are not only interested in the minimization of the prediction error, but also in the detection of scene objects that need enhancement. Thus, in larger block sizes the object content information is more easily identified than in partitioned blocks, even though they still may have larger prediction error. Hence, we suggest a penalty factor (with an empirical value of two) to be applied to the partitioned block prediction error.

$\widetilde{L}_k$ is a motion compensated layer using motion vectors between $F_{non-key}$ and $F_{LQ\,key,\,(k)}$ in order to find a contribution layer L_k such that

$$L_k = M_C\left(F_{key,\,(k)} - F_{LQ\,key,\,(k)}, \mathbf{V}\right), \tag{1}$$

where $M_C\left(\cdot\right)$ is the motion compensation operation and $\mathbf{V}$ is the set of motion vectors resulting from the $M_E\left(F_{non-key},\right.$ $\left.F_{LQ\,key,\,(k)}\right)$ operation. The enhanced non-key frame is then given by:

$$\hat{F}_{non-key} = F_{non-key} + p_{cf}\hat{L} \tag{2}$$

where $\hat{L}$ is a function of all $\{L_k\}$ and p_{cf} is a confidence factor.

The DISCOVER [7] side information generation method uses equal weights for the forward and backward predictions. Here, we can use multiple predictions with different weights. Let $\hat{L}(i,j)$ be the fused enhancement layer of a block, at the spatial position (i,j), and let $L_k(i,j)$ be a enhancement block prediction in the (i,j) position at the k-th reference key frame. Also, let $D_k(i,j)$ be the smallest SSD distance for a block positioned at (i,j,k). As shown in [1] the predicted enhancement layer is a fusion of the key-frames information based on maximum a posteriori:

$$\hat{L}(i,j) = \frac{D_2(i,j)}{D_1(i,j)+D_2(i,j)}L_1(i,j) + \frac{D_1(i,j)}{D_1(i,j)+D_2(i,j)}L_2(i,j). \tag{3}$$

Where the sub-indexes 1 and 2 represent the enhancement layer prediction from the previous and the next key frame, respectively.

The motion estimation method always picks a prediction block to enhance a non-key frame block. However, at sudden scene changes, the enhancement layer may decrease the objective and the subjective quality of a non-key frame. In order to reduce this problem, we only apply a percentage (p_{cf}) of the fused enhancement layer ($\hat{L}$) to the non-key frames ($F_{non-key}$). That percentage is interactively obtained by finding

$$\min_{p_{cf}}\left(\sum_{k=1}^{n} MSE\left(F_{non-key} + p_{cf}\hat{L}\,, F_{key\,(k)}\right)\right). \tag{4}$$

Finally, we add the enhancement layer to the low quality key frame as in (2).

III. Overlapped motion compensation

As mentioned before, we also use variable block size in motion estimation/compensation process. In order to allow different block sizes in the OBMC we make a virtual re-partition [13] of the blocks until the smaller size permitted to the quadtree partition is achieved. That enables a transformation into a fixed block size scheme shown in Figure 4.

The OBMC not only avoids blocking artifacts, but can also increase the prediction accuracy. When using OBMC, blocks are typically twice the size in each dimension and overlap quadrant-wise with all neighbour blocks. Thus, each pixel belongs to four blocks. In such a scheme, there are four predictions for each pixel which are summed up to a weighted mean. For this purpose, blocks are associated with a normalized window function such that the sum of overlapped windows is unitary everywhere [12]. Studies show that the

(a) (b) (c)

Fig. 3. (a) Regular block motion compensation. (b) Original frame. (c) Overlapped block motion compensation.

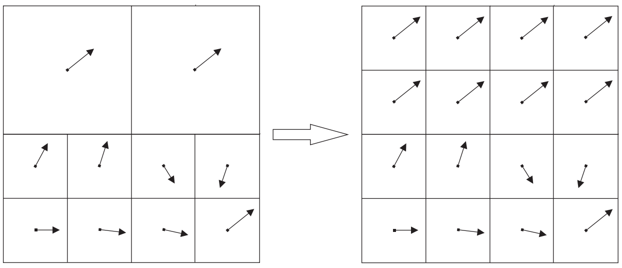


Fig. 4. Virtual block re-partitioning.

diagonally-adjacent block has the lowest contribution to the window function. That is, we can reduce the number of overlaps from three blocks rather than four. That leads to a substantial complexity reduction without a significant quality penalty. Such scheme is found in the H.263 Annex F. Figure 5, illustrates the reduction in overlap. In addition it enables a very good approximation for regular motion estimation methods.

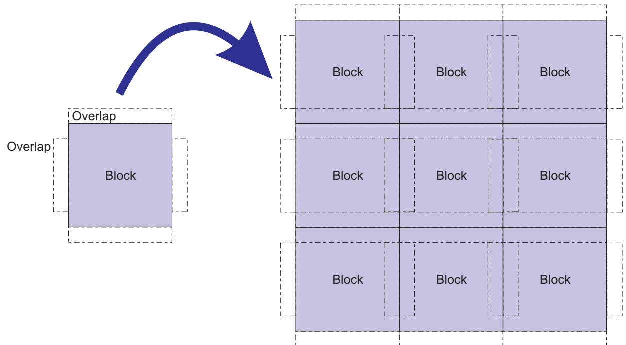


Fig. 5. The implemented overlapped block window.

IV. Experiments

In order to compare the performance of the enhancement technique using regular block motion compensation and OBMC, we processed two video sequences at CIF resolution (352×288 pixels) encoded with H.264 Intra with GOP length of 4 (that is, for each key frame, there are three non-key frames). We use a $Q_{non-key}$ parameter that corresponds to twice the quantization step of Q_{key}, i.e. $Q_{non-key} = Q_{key}+6$. In the enhancement method we also use a motion estimation window of 32×32 pixels for full macroblocks and partitioned blocks. We use two key-frame references (the closest forward and backward key frames) to generate a fused enhancement layer.

Figures 6 and 8 show the performance of fixed-QP intra-only H.264 compression compared to the mixed-QP H.264 intra with the enhancement technique using different motion compensation techniques. Figures 7 and 9 are the differential version of Figure 6 and 8, respectively. The fixed QP rate-distortion (RD) curve was used as reference.

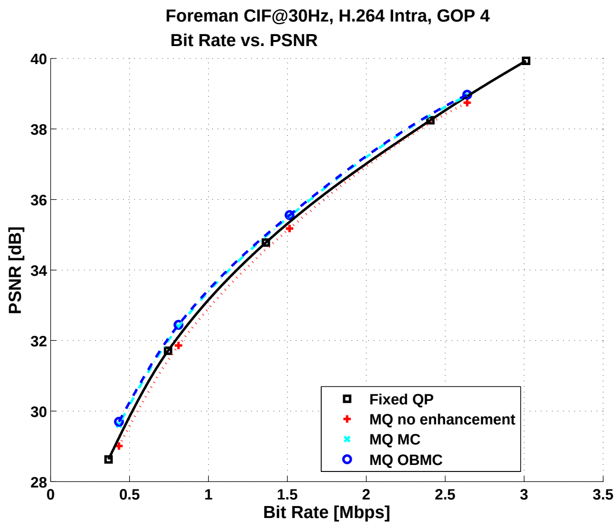
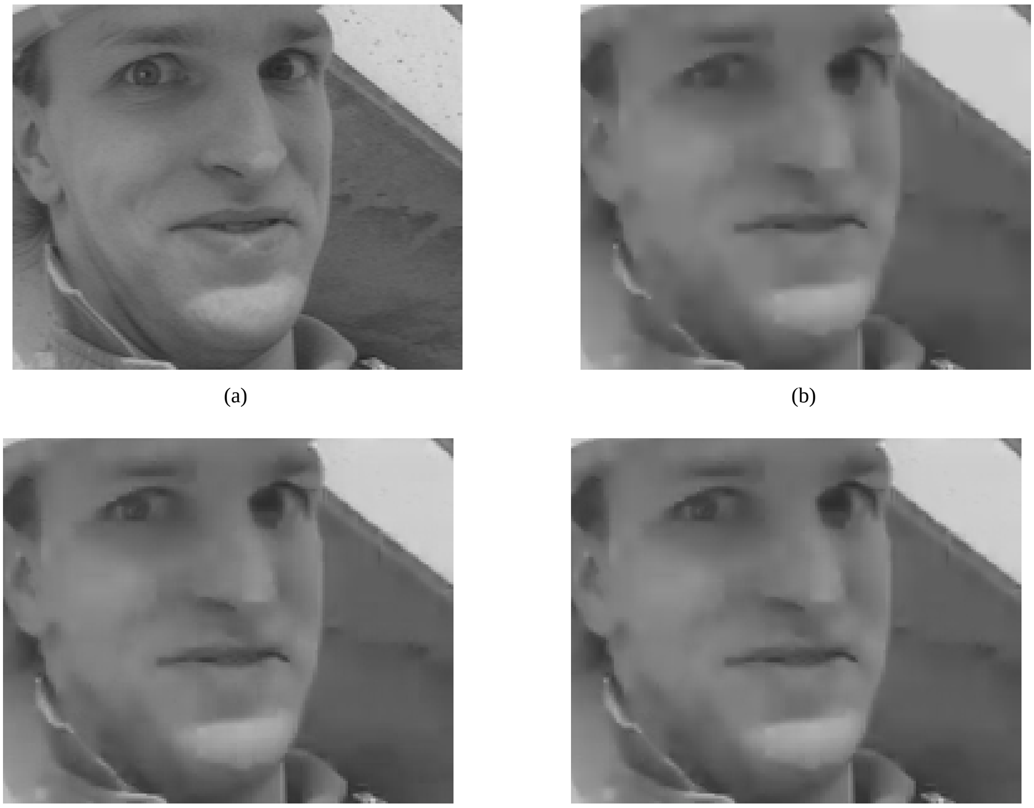


Fig. 6. Objective RD comparison between the proposed enhancement with regular motion compensation and overlapped motion compensation both applied to the sequence Foreman.

Despite the modest objective video quality gains, we show in Figure 10 a visual improvement in the enhancement method and a very tiny improvement in the proposed method. In order to evaluate the gains, we compare the original $102-th$ frame of sequence Foreman with the non-key frame without post-processing and a non-key frame with regular and OBMC enhancement.



(a) (b)

(c) (d)

Fig. 10. (a) Original frame. (b) Non-key frame without enhancement. (c) Enhanced sequence with regular block motion compensation. (d) Enhanced sequence with overlapped block motion compensation. This Figure can be better visualized in the electronic version.

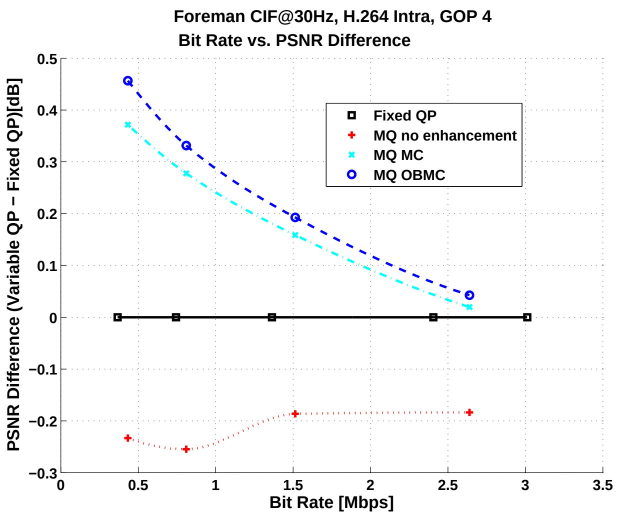


Fig. 7. Differential RD curves of the sequence Foreman comparing H.264 intra-only performance for regular fixed frame-quality, mixed frame-quality with regular motion compensation and overlapped motion compensation

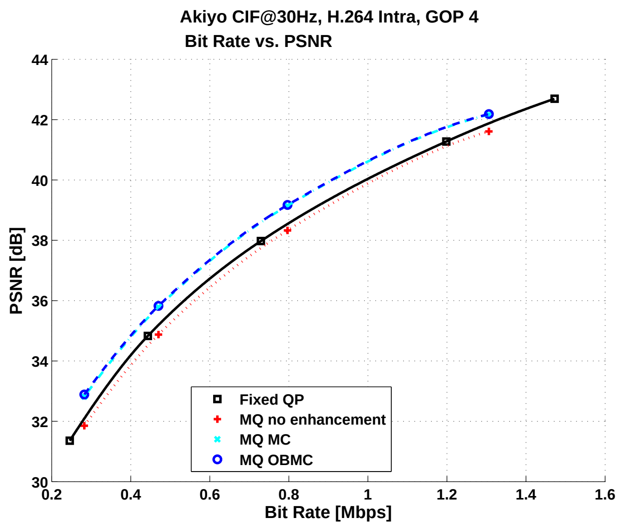


Fig. 8. Objective RD curve comparison between the proposed enhancement with regular motion compensation and overlapped motion compensation to the sequence Akiyo.

V. Conclusion

In this work, we propose a deblocking scheme for a post-processing video enhancement architecture that explores temporal redundancy at the decoder by using a block based matching approach. This framework allows for a reversible complexity coding, by using a mixed quality approach, i.e. varying frame quality among frames. In addition, the framework can also be applied as a side information generation technique to Wyner-Ziv codecs that may use mixed quality. The results show that it is possible to enhance low quality

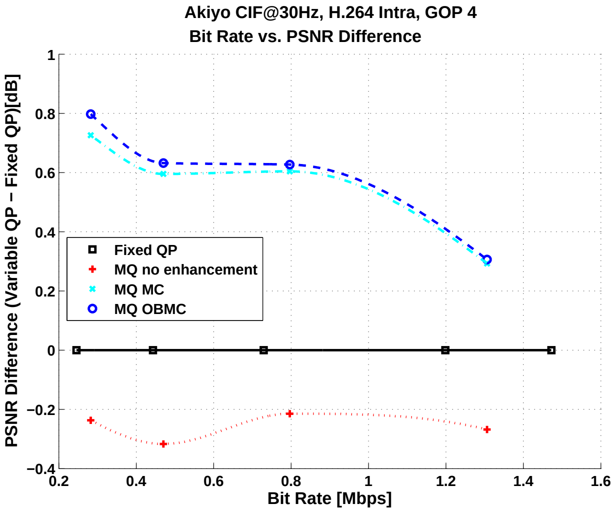


Fig. 9. Differential RD curves of the sequence Akiyo comparing H.264 intra-only performance for regular fixed frame-quality, mixed frame-quality with regular motion compensation and overlapped motion compensation

frames using high-frequency details from the key-frames, without any additional information being sent to the decoder. Improved performance occurs when we apply overlapped block motion compensation at the enhancement layer.

Acknowledgment

The authors would like to thank the support by HP Brasil and by CNPq under grant 47.3696/2007-0.

References

[1] E. M. Hung, and R. L. de Queiroz, "Codificação de vídeo com complexidade reversa utilizando qualidade mista," *Anais Simpósio Brasileiro de Telecomunicações*, Blumenau, PR, Brazil, Sep. 2009.

[2] K. R. Rao and J. J. Hwang, *Techniques and Standards for Image, Video and Audio Coding*, Prentice Hall, PTR, 1996.

[3] T. Wiegand, G. Sullivan, G. Bjoontegaard and A. Luthra, "Overview of the H.264 video coding standard," *IEEE Transactions on Circuits and Systems for Video Technology*, v. 13, pp. 560–576, Jul 2003.

[4] I. E. Richardson, *H.264 and MPEG-4 Video Compression: Video Coding for Next Generation Multimedia*. Wiley, 2003.

[5] B. Girod, A. Aaron, S. Rane and D. Rebollo-Monodero, "Distributed video coding," *Proceedings of the IEEE*, v. 93, pp. 71–83, Jan 2005.

[6] A. Wyner and 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, Jan 1976.

[7] X. Artigas, J. Ascenso, M. Dalai, S. Klomp, D. Kubasov and M. Ouaret, "The DISCOVER codec: Architecture, Techniques and Evaluation," *Picture Coding Symposium*, Portugal, 2007.

[8] H. Schwarz, D. Marpe, and T. Wiegand, "The scalable H.264/MPEG4-AVC extension: Technology and applications," *European Symposium on Mobile Media Delivery*, Sardinia, Italy, Sep 2006.

[9] F. Brandi, R. L. de Queiroz and D. Mukherjee, "Super-resolution of video using key-frames and motion estimation," *Proc. IEEE Intl. Conf. on Image Processing*, ICIP, San Diego, CA, USA, Oct 2008.

[10] D. Mukherjee, B. Macchiavello and 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, Jan 2007.

[11] C. A. Segall, R. Molina and A. K. Katsaggelos, "High-resolution images from low-resolution compressed video". *IEEE Signal Processing Magazine*, v. 20, pp. 37–48, 2003.

[12] S. Nogaki and M. Ohta, "An overlapped block motion compensation for high quality motion picture coding", *Proc. IEEE Int. Symp. Circuits Systems*, v. 1, pp. 184-187, 1992.

[13] J. Zhang, M. O. Ahmad and M. N. S. Swamy, "Overlapped variable size block motion compensation," *Proc. IEEE Intl. Conf. on Image Processing*, ICIP, Santa Barbara, CA, USA, Oct 1997.