RATE-DISTORTION OPTIMIZED REGION-BASED VIDEO CODER

Stéphane Pateux and Claude Labit

IRISA, Temics project
Campus Universitaire de Beaulieu
35042 Rennes Cedex, FRANCE
e-mail: spateux,labit@irisa.fr

ABSTRACT

In this paper, we present an efficient region-based video coding scheme. The region partioning is optimized according to a rate-distortion criterion in order to achieve constant bitrate or constant quality. This optimization is led through a labeling for each image of an initial segmentation map using regions merging, regions' label relaxation and quantization variation. Results obtained show the performance of this coding scheme compared to MPEG and H263 codecs.

1. INTRODUCTION

When dealing with region-based video coding as proposed in [2], the hard point is to find a segmentation which is suitable for coding. Early papers proposed various techniques to define this region partitioning (spatial segmentation [2], motion-based segmentation [3], or spatio-temporal segmentation [4]), but these techniques were not taking into account any rate-distortion consideration, thus not leading to efficient region-based video coding. In more recent studies [5, 6], rate-distortion consideration has been introduced in order to define a suitable level of segmentation between a coarse segmentation (low cost, poor prediction) and a fine segmentation (high cost, efficient prediction) respectively for quad-tree segmentation and segmentation based on partition-tree. However since the rate-distortion optimization was led over a predefined hierarchical segmentation, results are not ensured to be optimal.

In [1], we proposed another technique based on the Minimum Description Length (MDL) formalism in order to optimize the segmentation; results obtained were very good compared to the one observed in [5, 6]. But the MDL formalism did not really took into account the distortion introduced and thus did not necessary led to an optimal rate distortion solution (cf. figure 3). On the other hand, the optimization was not led on a hierarchical segmentation but on a classical segmentation.

We then propose in this paper to adapt our proposed technique in order to achieve better results in the sense of rate-distortion.

2. RATE-DISTORTION OPTIMIZATION OF THE REGION PARTITIONING

2.1. Overview of the coding scheme

The region-based coding scheme we use is schematized on figure 1; this coding scheme is similar to the predictive coding scheme used by MPEG and H26x codecs, except that motion compensation is performed on regions rather than blocks, using affine motion model instead of simple translation. Advantages of this approach is to be able to take into account complex motion such as zoom, rotation and skewing; moreover region segmentation prevents from having areas containing objects with different motion. In order to code the residual error ε^t, we use a DCT coding technique with Huffman tables used in H263 codec.

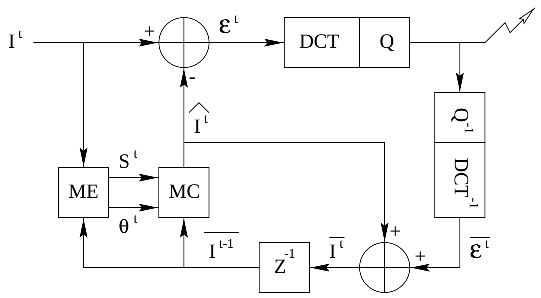


Figure 1: Region-based predictive coding scheme. ME and MC state respectively as Motion Estimation and Motion Compensation.

2.2. Rate distortion optimization of the segmentation

In [7, 1], in order to optimize the segmentation, we used a technique labeling an initial over-segmented partition; regions were grouped in the aim to reduce the global coding cost (cost of residual error, motion parameters and segmentation map). When performing this labeling, we distinguished two operations: merging of regions and label relaxation of a region using the label of an adjacent region (see figure 2). All these operations were sorted according to a rate-distortion criterion. For each possible operation i is computed its variation in cost ΔDL_i and its variation in distortion ΔMSE_i. The operation minimizing $\frac{\Delta MSE_i}{|\Delta DL_i|}$ under the constraint that $\Delta DL_i < 0$ is then chosen. Compared to [5, 6], our approach does not explicitly use Lagrangian multipliers, however, it could be interpreted in a Lagrangian minimization scheme. When using Lagrangian multipliers the aim is to minimize a functional such as $R + \lambda.D$ for a given λ and then to make vary λ in order to respect given constraints (over rate or distortion). When minimizing $R+\lambda.D$ with regions merging and region relaxation, the ratios $\frac{\Delta MSE_i}{|\Delta DL_i|}$ will have to be lower than $\frac{1}{\lambda}$. Thus sorting operations as we do is equivalent to perform a search on λ starting from a high λ and reducing it progressively.

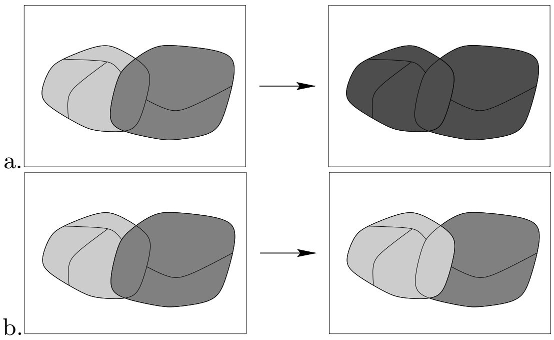


Figure 2: Operations made for labeling of the initial spatial segmentation: (a) merging of regions, (b) relaxation of frontier regions.

The main advantage of this technique is to permit to refine the motion estimation by selecting the best motion parameters among several candidates when merging regions. Generally people use motion estimation based on minimization of the error prediction, but this motion is not guaranteed to be the best one in terms of coding efficiency. For instance, in the case of occlusion, the motion minimizing the prediction error is not always the real motion of the moving areas. For this reason, when merging regions, we introduce this selection of motion parameters; parameters tested are the parameters from regions merged (2 sets), mean parameters of the two regions, best parameters approximating the previous dense motion field, and refined version of these sets.

Figure 3 shows typical rate-distortion evolutions through the different steps of the labeling algorithm for different quantization factors (evolution goes from high bit-rates to low bit-rates). We can see on this figure that at the beginning of the optimization process, there is a distortion decrease even though the bit-rate decrease. This unusual aspect is due to our motion selection technique which helps to find more suitable motion parameters for coding, especially when regions are small. Since labeling is performed as long as there is a reduction in cost, we can also see on this figure, that we could get non-optimal results, and that it would have been better to change the quantization factor.

Thus to get an algorithm that is efficient in a rate-distortion sense, we introduce the possibility to change the quantization factor during the optimization process. To this

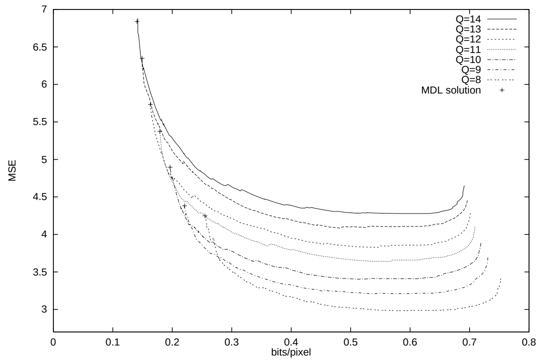


Figure 3: Non rate-distortion optimality of the MDL segmentation technique proposed in [1]. Each curves correspond to a given quantization factor, cross points give the final result obtained with the MDL formalism.

extent, we associate a variation in rate and distortion as was done for region merging and region relaxation. Finally this operation is treated as the other ones and sorted according to the same rate-distortion criterion.

In order to stop the optimization process, a target in bit-rate or quality is given. This target is updated on each frame with a simple control law in the aim of obtaining a given bit-rate or quality:

$$T_{t+1} = T_t + (T^{\star} - O_t)$$

where T_t is the target defined for time t, $T^{\star}$ is the global target to reach and O_t is the obtained result at time t.

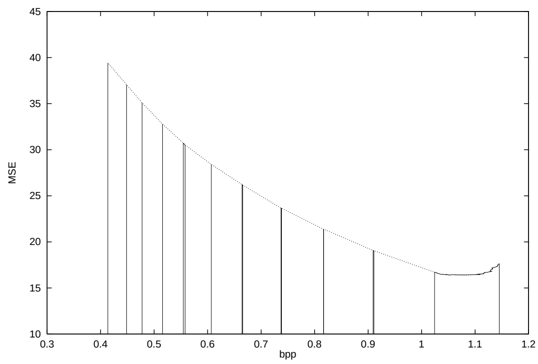
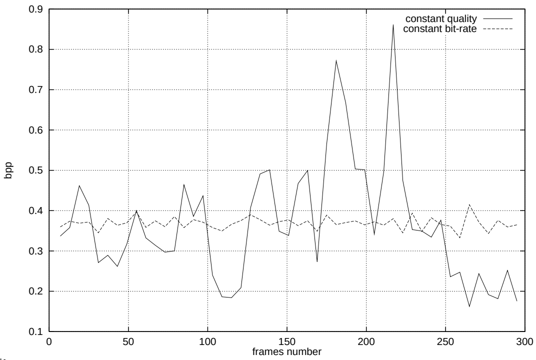


Figure 4: Rate-distortion evolution through mergings and quantization changes; dashed lines and impulses show instant where quantization step changes, solid lines show variations due to merging and relaxations. (frame 24 of *Foreman* sequence)

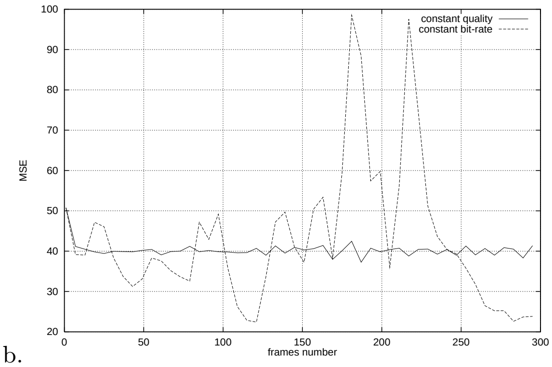
Figure 4 shows the typical evolution of rate and distortion during the optimizing steps. This figure shows that most of the merging and relaxation are performed at the beginning of the optimization while at the end there are nearly only quantization factor changes.

3. RESULTS

In order to evaluate performance of our approach, we have tested our coding scheme on various sequences. For each sequence, the first frame was coded using JPEG, the other frames were coded with the prediction scheme shown on figure 1 using only 'P-frames'. No further intra frame was introduced. For each frame an initial segmentation was given by the spatial algorithm presented in [7], with no tracking of the segmentation. In order to define the distortion, we use a distortion taking into account distortion in luminance and chrominances: $D = MSE_Y + \frac{MSE_U + MSE_V}{4}$; MSE of chrominances are divided by 4 in order to take into account the down-sampling of the 4:2:0 format.



a.



b.

Figure 5: Variation of rate (a) and distortion (b) for *Foreman* sequence coded at 5 Hz; rate is defined to 0.37 bpp and distortion to 40 (Distortion is defined as $D = MSE_Y + \frac{MSE_U+MSE_V}{4}$)

First, the *Foreman* sequence in QCIF format is treated, with a coding frame-rate of 5 Hz. We successively code it with a constant quality and a constant bit-rate. In the constant quality case, we choose a target distortion of 40, e.g. 32 dB. For the constant bit-rate case, we choose as target bit-rate the one obtained by the previous method, e.g. 0.37 bpp. On figure 5 are plotted the variation of the distortion and the bit-rate for the two treated case. We can see on this figure, that the aimed target given is well respected and that the two methods give results quite similar (MSE of 43.9 for the constant bit-rate approach). Compared to an H263 codec for this sequence, the gain is almost 1 dB for luminance component (32.6 dB on the Y

a

b

c

Figure 6: Results obtained of frame 96 of *Foreman* sequence: initial spatial segmentation (a), final segmentation obtained (b), decoded frame (c).

component, and almost 38.5 dB for the U-V components). Furthermore, the visual quality is enhanced compared to H263 codec since block artifacts have nearly disappeared (see decoded frame of *Foreman* sequence on figure 6.c). Improvements could still be done by taking into account only the luminance component in the distortion, or using different quantization factor for each component.

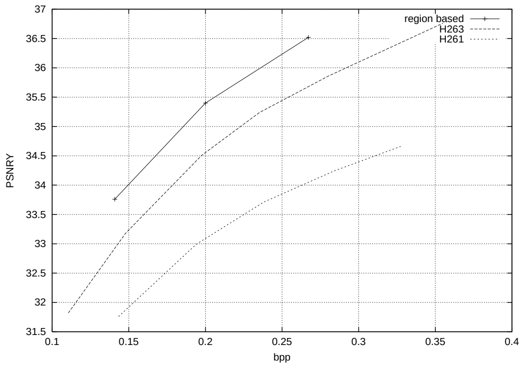
Figure 6 shows the segmentation obtained on one frame of the *Foreman* sequence. We can see on this figure that the various moving objects are well segmented. But it has to be noticed that this is not always the case (see figure 7). There could have frames where motion is not important leading to no segmentation, or even in the case of strong motion or new object apparition, since even with a good segmentation it is not possible to improve significantly the prediction quality.


a.

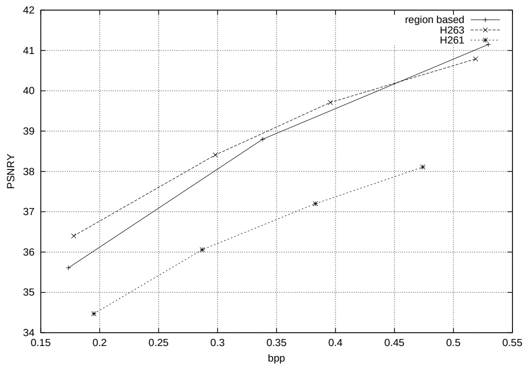

b.

Figure 7: Examples of segmentations obtained that are not suited with the boundaries of the objects present in the scene. (a) *Mother&Daughter* sequence, the hand and the girl are not segmented and parts of the woman are merged with the background. (b) *Suzie* sequence, the regions obtained do not correspond to objects of the scene.

Figure 8 shows results obtained on *Suzie* and *Mother&Daughter* sequences coded at 5 Hz for H261 and H263 codecs and for our proposed coding scheme. On this figure, we can see that when there is significant motion, our coding scheme can get for the same bit-rate an increase of up to 1 dB compared to H263 codec and more than 2 dB with H261. But when there is not a lot of motion, such as for the *Mother&Daughter* sequence, our codec is close to H263 (even less effective for very low bit rate coding due to the cost of the region definition).



a.



b.

Figure 8: Coding results of the QCIF *Suzie* (a) and *Mother&Daughter* (b) sequences coded at 5 Hz. PSNR shown is the one of the luminance, bit-rate is the bit-rate of the color sequence (figures are averaged over the sequence).

Figure 9 shows that our region-based coding scheme outperforms also MPEG2 codec, with a gain of nearly 3 dB on the luminance component. As observed for QCIF sequences, visual quality is enhanced with less blocking artifacts and a more stable image over time (see figure 10).

4. CONCLUSION

To conclude, we have presented in this paper a new region-based video coder optimizing the segmentation used for motion compensation based on a rate-distortion criterion. Experimental results show that this coder outperforms normalized predictive coders, such as H261, H263 and MPEG2 codecs. Moreover due to the strong reduction of block artifacts, visual quality is significantly enhanced.

Several extensions to this work could be investigated. For instance, other prediction error coding techniques could be used or quantization could be different for each component.

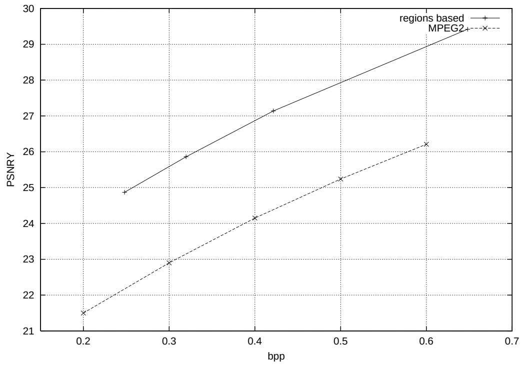


Figure 9: Coding results of the 20 first frames of the *Flower Garden* sequence (CIF format). Comparison between our proposed coding scheme and MPEG2. PSNR shown is the one of the luminance, bit-rate is the bit-rate of the color sequence.

a.
b.


Figure 10: Decoded Frame of the *Flower Garden* sequence (frame 20); proposed scheme (a), MPEG2 (b).

Improvements could also be done by applying the proposed optimization algorithm to bidirectional predictive coding (e.g. B-frame of MPEG2).

5. REFERENCES

[1] S. Pateux and C. Labit. Region-based video coder unsing the MDL formalism. In *ICIP'98*, volume 1, pages 304–308, Chicago, Illinois, USA, October 1998.

[2] P. Salembier, L. Torres, F. Meyer, and C. Gu. Region-based video coding using mathematical morphology. In IEEE, editor, *Proceedings of the IEEE*, volume 83, pages 843–857. IEEE, June 1995. Special Issue on Digital Television.

[3] F. Moscheni and S. Bhattacharjee. Robust region merging for spatio-temporal segmentation. In *ICIP'96*, volume 1, pages 501–504, Lausanne, Switzerland, September 1996.

[4] F. Morier, J. Benois-Pineau, D. Barba, and H. Sanson. Robust segmentation of moving image sequences. In *ICIP'97*, pages 719–722, Santa Barbara, California, October 97.

[5] E. Reusens, T. Ebrahimi, C. Le Buhan, R. Castagno, V. Vaerman, L. Piron, C. de Sola, S. Battacharjee, F. Bossen, and M. Kunt. Dynamic approach to visual data compression. *IEEE Trans. on Circuits and Systems for Video Technology*, 7(1):197–211, February 1997.

[6] P. Salembier, F. Marqués, M. Pardàs, R. Morros, I. Corset, S. Jeannin, L. Bouchard, F. Meyer, and B. Marcotegui. Segmentation-based video coding system allowing the manipulation of objects. *IEEE Transactions on Circuits and Systems for Video Technology*, 7(1):60–73, February 1997.

[7] L. Bonnaud, S. Pateux, and C. Labit. Multiple objects tracking for efficient motion-based segmentation coding using a temporal prediction. In *Proc. of PCS 97*, pages 125–128, Berlin, RFA, September 1997.