OPTIMIZATION OF IMAGE SEQUENCES SCALABLE CODING

Erwan LAUNAY

THOMSON multimedia R&D France
1 avenue de belle fontaine, BP19, 35511 Cesson-Sévigné Cedex, France.
erwan.e.l.launay@tcetbs1.thomson.fr

ABSTRACT

In a previous article [1], we examined to what extent and in what conditions scalable coding of image sequences as defined in MPEG2 could be a useful tool. We thus showed that, at a constant coding quality, up to 35% of the base layer rate could be spared by scalable coding. This gain was achieved mostly on the luminance (on I and P frames particularly) and increased when the motion content of the sequences coded was difficult to handle.
This previous work, however, entirely relied on earlier results provided by literature. This contribution aims at further investigating on optimization of scalable coding based on our own previous observations. As will be seen our experiments confirm the choices made in MPEG2, but also give some interesting insights on the mechanisms of scalable coding, examining problems such as motion handling and optimal segmentation in a scalable scheme.

1. INTRODUCTION

With the multiplication and diversification of the applications that require transmitting or storing video data, the new challenge in the field of image sequences compression seems to lie more in the definition of a flexible norm providing functionalities and performances adapted to each specific service, rather than in improving the rough performances of already well optimized basic coding algorithms. This is the aim of MPEG4, indeed.
Scalable coding of image sequences belongs to this new area of research. It consists in transmitting video data in several distinct bitstreams corresponding to several transmission layers. These layers can be decoded together, thus yielding maximum quality, or only part of these layers can be decoded separately, yielding a lower quality or lower resolution. Scalable coding appears as a useful tool in specific applications needing graceful degradation of image quality with channel noise, but it also eases interworking of video services and compatibility with existing standards.
However if scalability has been extensively studied prior to the finalization of MPEG2 and the definition of a scalable coding toolbox in the "scalable extensions" of MPEG2, and even if MPEG4 is still concerned with scalability, there have been very few new studies on scalability since then. The main reason for this is that most people seem to consider that scalability such as defined by MPEG2 is now a well-optimized and finalized tool. And the few papers that recently dealt with scalable coding [1], [2], [3], only tried to assess the performances of scalability as defined in MPEG2.
Even if, as shown in the first part of this paper, the definition of an optimum scalable coding architecture does appear to be thoroughly addressed by literature and globally conforms to the MPEG2 requirements, when assessing the performances of spatial scalability based on such an architecture, we realized that scalable coding needed some further investigation for two reasons :

- Further optimize scalable techniques used in this architecture (for example in the more flexible framework of MPEG4).
- Have a better understanding of scalable coding and thus use it more efficiently (in the framework of MPEG2).

This article will be divided into three parts. A first part explains why we focus mainly on spatial scalability, and based on literature, describes and justifies the baseline coding scheme used here. The second part details some early results obtained when preparing the work presented in [1] and explains why it led us to deeper investigations on spatial scalability. Finally we describe the new experiments conducted and discuss simulation results.

2. BASELINE CODING SCHEME

Scalability as defined in the introduction could take several forms and be a useful tool for several applications. However scalability has a cost [3]. First a scalable decoder, intended for decoding all layers for a maximum transmission quality will be more complex than a "standalone" decoder, intended for decoding non scalable bitstreams and yielding the same quality. Secondly the global transmission rate will be higher in the scalable case where the bitstream is divided into several bitstreams than in the "standalone" case. When considering these extra costs of scalability, only some forms of scalability appear to be interesting, and only for some applications. We chose to focus on TV coding at

rates similar to those aimed at by MPEG2. Thus the only interesting scalabilities for us will be those chosen by the experts of ISO to be part of the MPEG2 toolbox [2].
However among the scalabilities allowed by the MPEG2 "scalable extensions" only one will interest us here, mainly because any results obtained on this complex technique can be extended to the other scalabilities, but also because it is the only one to imply some extra complexity of the decoder [3]. This scalability is called spatial scalable coding. Spatial scalable coding uses two transmission layers corresponding to two different spatial resolutions. The base layer represents the original sequence at a quarter resolution, and the enhancement layer contains the additional information necessary to reconstruct the full-resolution sequence. Ideally we should study HDTV/TV scalable coding but for implementation purposes, we had to restrict ourselves to TV/¼TV scalable coding. Most results obtained for this last scheme can be extended to the previous case anyway.
The question of spatial scalable coding was first addressed in [4], [5], and after trying to develop some customized coding algorithms for this purpose [4],[6], [7], it became obvious that the only valuable way to implement spatial scalable coding was to construct it as a simple extension of hybrid coding schemes. In [1] we described how the observation of the works of Bosveld [8], Mau [9] and Taubman [6] (among others) led us to define a baseline spatial scalable coding scheme for our experimentation. This scheme, is constructed as follows :

- We chose a "standalone" hybrid coding scheme which, except in some aspects (PRMF subband transform and quantization steps constrained to be power of two), conforms with MPEG2 syntax.
- A "simulcast" coding scheme is then built by spatially downsampling the original sequence and then coding it in the base layer using the "standalone" coder. The full resolution sequence is coded separately by another identical (though this one is independent with its own parameters and rate control) "standalone" coder in the enhancement layer.
- Spatial scalable coding is finally implemented as follows : based on the base layer low-resolution decoded images at time t, we elaborate an alternate prediction image (called spatial prediction thereafter) for the full resolution image coded by the enhancement layer at time t and chose the best prediction between this prediction and the prediction provided by motion compensation. This selection is done in an MSE sense and on a macroblock by macroblock basis. Additional side information specifying this choice is transmitted.

Our scalable coder finally appears as two independent hybrid coders linked by a simple "additional spatial prediction process".

3. EARLY OBSERVATIONS

We implemented several "prediction processes", most of them not conform with MPEG2 syntax :

- In the first one the spatial prediction is the base layer decoded image upsampled by simple linear filtering (as in MPEG2). We call it "MPEG2-like spatial scalability" in the following. The choice between temporal and spatial predictions performed in the enhancement layer takes place on a macroblock by macroblock basis in the image. But we also investigated what happened if, discarding the additional side information needed, we implemented this choice relaxing the constraint to have the same choice at the same location of luminance and chrominance fields, and if we relaxed the constraint to perform this choice on square blocks, thus processing each pixel individually. These constraints we relax are called thereafter "color constraint" and "block constraint".
- We then implemented "frequency scalability". The choice between spatial and temporal prediction is performed in the transform domain. The spatial prediction thus simply consists in the 16 decoded bands of the subband transformed low resolution image. They are used (in competition with the 16 low frequency bands of the full resolution temporal prediction image) to predict the 16 low frequency bands of the full resolution image in the enhancement layer. The remaining 48 high frequency bands are simply temporally predicted. The choice between both predictions is performed on the projection of a macroblock in the 16 low frequency bands, based on an MSE criterion. Associated side information is transmitted. As in the previous case we also examined what happened if relaxing the "block" and "color" constraints.
- Finally we investigated residual scalability. This scheme, suggested by [10] , is similar to frequency scalability but uses the coded residual of the base layer to predict the motion compensated residual of the enhancement layer. This is equivalent to using as spatial prediction the sum of the coded base layer residual and the 16 low frequencies of the enhancement layer temporal prediction. As in the previous case we chose between both predictions on a macroblock by macroblock basis and transmit additional information, but we also investigate what happens if "color constraint" and "block constraint" are relaxed.

The first aim of these simulations was to assess the performances of these spatial scalable coding schemes, computing what could be gained when comparing the "standalone" and scalable enhancement layers coding rates at the same coding qualities. The results obtained are detailed in [1]. But it also motivated several remarks that leaded to the work presented thereafter :

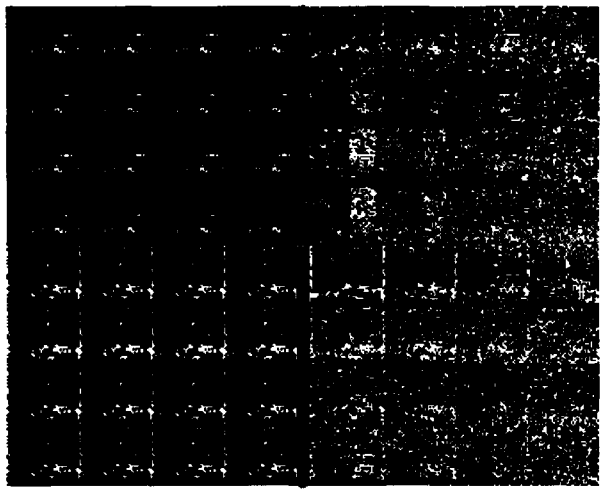

Figure 1 : Frequency scalability label images of MOBILE CALENDAR. Consecutive B (up) and P (down) images, with (left) and without (right) "block constraint".

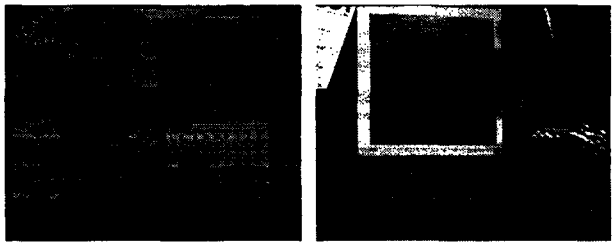

Figure 2 : Original image sequences. MOBILE CALENDAR (left) and RENATA_RAI (right)

- When relaxing either the "color constraint" or "block constraint", we observe a significant growth of the rate gain (discarding additional side information). This gain doubles or triples when discarding both constraints. This means that the simple segmentation (on a macroblock basis) used to determine what points should be spatially predicted fails to take full advantage of the potential gain of scalability.
- This improvement is observed for both frequency scalability and MPEG2-like spatial scalability, but it remains negligible for residual coding. This means that both spatial and frequential segmentation of spatially predicted points need to be improved, but also shows the influence of the mismatch between the base layer motion information and the enhancement layer motion information in the case of residual coding.

While the interest of scalable coding is obvious for I picture, the gain observed on B and P pictures when using scalability is much smaller. The main reason for this is that P and B full resolution pictures are already well predicted by the motion compensation loop of the enhancement layer and the spatial prediction hardly competes with this temporal prediction. Another reason for this lies in the structure of the coding cost of P and B full resolution pictures. For P pictures motion information

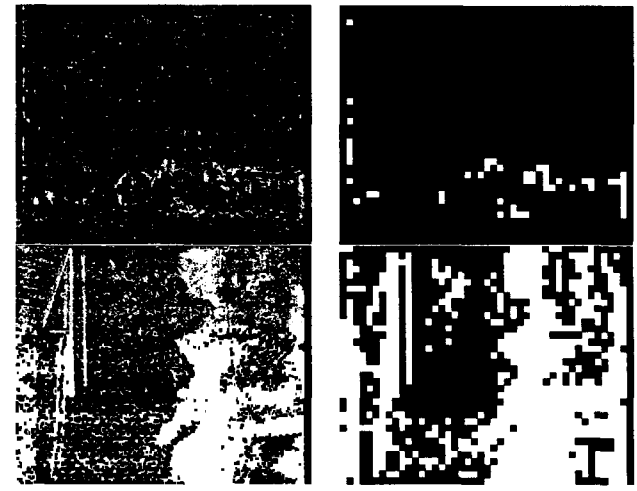

Figure 3 : P image MPEG2-like spatial scalability labels. With (left) and without (right) "block constraint" for MOBILE CALENDAR (up) and RENATA_RAI (down).

represents a great part of the bits transmitted (20%) and for B pictures it represents most of the information used (75%). But the only gain achieved by spatial scalability is based on improving the coding of transform coefficients and thus we fail to achieve any scalable gain on motion information. Since B and P pictures represent 80% of the coding cost for an entire sequence, it could be interesting to investigate scalable coding of the motion vectors.

4. EXPERIMENTS CONDUCTED

These observations led to several experiments :

- First we created label images to examine the distribution of the spatially predicted points in a scalable coding scheme. Frequency scalability results in labels in the transform domain and enables us to examine the frequency distribution of these points, while their spatial distribution is provided by MPEG2-like spatial scalability label images. We created such label images with and without "block constraint" and with and without "color[1] constraint" on 6 different video sequences. We focused on P and B pictures since no temporal prediction competes with spatial prediction for I pictures.

  The results showed here concern RENATA_RAI and MOBILE CALENDAR (labels of frequency scalability for the 16 Low frequency bands in Figure 1, labels of MPEG2-like spatial scalability in Figure 3, and originals in Figure 2,). We use white points for spatially predicted pixels and black points for temporally predicted pixels. The main conclusions drawn from this study are the following :

  ⇒ All label images without "color constraint" look very noisy. This means that chrominance and

[1]Label images without « color » constraint are not presented herein since the use of 9 labels resulted in color label images.

luminance scalable choices are very decorrelated and explains the very bad performances of spatial scalability based on macroblocks in chrominance fields [1]. Comparing the gains observed on chrominance without "color constraint", to those obtained on luminance shows that it won't be valuable to transmit extra side information to have a spatial coding scheme based on each block in each color component. The best solution is to systematically predict temporally P and B chrominance fields.

⇒ Image labels for frequency scalability and MPEG2-type spatial scalability without "block constraint" still exhibit a noisy aspect, showing how difficult it would be to find, as initially proposed, an efficient alternate segmentation (spatial or frequential) for spatially predicted points, even if using a complex algorithm in an MPEG4 framework.

⇒ However on all these images we observe that the spatially predicted points seem to concentrate on areas where motion is difficult to predict : the train an the areas uncovered by camera panning in MOBILE CALENDAR, the uncovered areas and the shadow of the woman in RENATA_RAI. We also observe that spatially predicted points are much denser on P images than in B images, mainly because the temporal prediction is less efficient, especially on uncovered areas (no backward prediction). Thus if any method should be used to improve selection of spatially predicted points, it should base on the local performances of motion estimation.

- Then we investigated what happened if we calculated low resolution motion vectors using downsampled high resolution motion vectors, instead of using two independent motion estimation loops in a spatially scalable coding scheme.

  ⇒ This led to an increase of 2 to 10% of the entropy of the base layer residual but no significant additional scalable coding gain.

  ⇒ Hierarchically coding the motion vectors thus obtained (subtracting scaled low resolution motion vectors from high resolution motion vectors), led to no improvement compared to traditional independent DPCM coding of motion vectors in each layer.

Finally we can conclude from these experiments that (as expected) some correlation does exist between the performance of the motion estimators and the performance of spatial scalability, but unfortunately, it seems difficult to efficiently exploit it. We also conclude that forcing both motion estimations in a spatial scalable scheme to be similar for more correlation between both residuals and both motion vector fields does slightly improve the number of spatially coded points (in "unconstrained schemes") and improves the performance of motion vectors hierarchical coding. But this is not enough to be competitive with very optimized and more simple MPEG2 scalable schemes and no efficient "spatial scalable" processing of motion can be constructed.

5. CONCLUSION

The results observed here, if they do not provide us with all the gains expected, illustrate and confirm very well the conclusions of [1]. But our study also reveals that the weaknesses observed in scalable coding scheme optimization when achieving the work in [1] are difficult to correct in the framework of a hybrid coding scheme. The main reason for this is that hybrid coding was optimized for "standalone" coding. And scalable techniques, when implemented in a hybrid coding scheme, hardly compete with the simple but optimal techniques used in a "standalone" hybrid coding scheme. Our results seem to indicate that only a scheme entirely optimized for scalable coding could lead to a significant improvement over MPEG2 scalable extensions and since this possibility has already been explored, with no really interesting results (see introduction of this article), the only solution could be to place this problem in a different framework (using object oriented schemes as allowed in MPEG4).

6. REFERENCES

[1] E. Launay, "On scalable coding of image sequences", in *Proc. third Intern. Workshop on Image/Signal Processing*, pp. 249-252, 1996.

[2] J. Delameilleure, S. Pallavicini, "Scalability in MPEG2", in *Proc. European Workshop on Image Analysis and Coding for TV, HDTV and Multimedia Applications (EWIAC)* , pp.69-75, 1996.

[3] J. Delameilleure, S. Pallavicini, "A comparative study of simulcast and hierarchical coding", in *Proc. EWIAC*, pp. 59-65, 1996.

[4] M. Pécot, P. J. Tourtier, Y. Thomas, "Compatible coding of television images", *Signal Processing : Image Comm.*, Vol. 2, No. 3, pp. 259-268, 1990.

[5] B. Macq, L. Vandendorpe, "Optimum quality and progressive resolution of video signals", *Ann. Télécommun.*, Vol. 45, No. 9-10, pp. 487-502, 1990

[6] D. Taubman, A. Zakhor, "Multirate 3-D subband coding of video", *IEEE Trans. on Image Processing*, Vol. IP-3, No. 5, pp. 572-588, 1994.

[7] J. R. Ohm, "Three-dimensional subband coding with motion compensation", *IEEE Trans. on Image Processing*, Vol. IP-3, No. 5, pp. 559-571, 1994.

[8] F. Bosveld, R. L. Lagendijk, J. Biemond, "Hierarchical coding of HDTV", *Signal Processing : Image Comm.*, Vol. 4, No. 3, pp. 195-225, 1992.

[9] J. Mau "Perfect reconstruction modulated filter banks", in *Proc. ICASSP*, pp. III.225-III.228, 1993.

[10] J. F. Vial, "Multiresolution coding schemes with layered bitrate regulation", in *Proc. fourth Intern. Workshop on HDTV*, 1993.