EFFICIENT SCALABLE MOTION CODING FOR WIDE-RANGE SCALABLE VIDEO COMPRESSION

Guillaume Boisson [1], Edouard François [1] and Christine Guillemot [2]

[1]-THOMSON multimedia R&D France,
1 Avenue de Belle Fontaine, CS17616, 35576 Cesson-Sévigné, France (Europe)
phone: +33 2 99 27 38 28, fax: +33 2 99 27 30 15, email : guillaume.boisson@thomson.net
[2]-IRISA, Campus Universitaire de Beaulieu, 35042 Rennes Cédex, France (Europe)
phone: +33 2 99 84 71 00, fax: +33 1 99 84 71 71, email : christine.guillemot@irisa.fr

ABSTRACT

For a scalable video coder to cover a wide range of bit-rates, e.g. from mobile video streaming applications to TV broadcasting, it is essential that motion information presents some form of scalability. In this paper we propose a new accuracy-scalable motion codec in a wavelet-based spatio-SNR-temporal-scalable framework. It allows to decode a reduced amount of motion information at sub-resolutions, taking advantage that motion compensation requires less and less accuracy at lower spatial resolutions. This new motion codec proves its efficiency in our full-scalable framework, by improving significantly video quality at sub-resolutions without inducing any noticeable penalty at high bit-rates.

1. INTRODUCTION

During its 66th meeting in Brisbane, Australia, the MPEG community published a Call for Proposals (CfP) on Scalable Video Coding, which can be seen as the starting point of a standardization process for scalable video coding [1]. Evidence had indeed been proven that scalable coding technologies can match single-layer coding performances, while addressing several applicative requirements that can not be easily fulfilled by non-scalable technologies (Cf. [2]). The main test of this Call for Proposal consists in a single encoding of HD video material into a single embedded bit-stream, then in its decoding at various resolutions, frame-rates and bit-rates, from 6Mbps (high-quality TV) down to QCIF 64Kbps (mobile video streaming)

For a scalable video coder to remain efficient over such a wide range of bit-rates and resolutions, some form of scalability must exist in the motion information. Since prior Call for Evidence (CfE) on scalable video coding, several solutions have been proposed.

Responding to CfE, Hang et al. proposed in [3] a scalable motion coder coupled with famous (2D+t)WT scheme MC-EZBC. Each motion field was divided into a base layer (16x16 blocks and above) and an enhancement layer (smaller blocks), both having approximately the same cost. Although the adequate number of enhancement layers was determined manually and empirically for each bit-rate, Hang et al. proved that their scalable motion codec can significantly improve MC-EZBC performances at low bit-rates.

Another solution to introduce natural scalability within motion information is to perform the spatial transform first, then estimate the displacement in each sub-band independently, before processing e.g. wavelet-domain motion-compensated temporal filtering [5], wavelet-domain prediction, or sophisticated contextual entropy coding [4]. In [5], Andreopoulos et al. proposed an in-band MCTF scheme, based on overcomplete wavelet transform, that outperformed spatial-domain MCTF (with full-pel accurate ME/MC and fixed block size).

In [6], Taubman and Secker showed that the video distortion introduced by quantizing motion information is roughly additive to the distortion due to texture quantization. Using JPEG2000-like techniques (reversible wavelet transform and fractional bit-plane coding), the authors built a rate-scalable motion bit-stream and determined empirically an optimal balance between motion and texture bit-budgets. Nevertheless, performances of this solution may be limited by the amount of motion in such a 5/3 MCTF scheme with triangular mesh model.

Our proposed method is far less complex than previous approaches and give promising results. This paper is organized as follows. The global scalable video coding framework is described in Section 2. In Section 3 we investigate the scalable motion coding issue, and propose a new layered motion representation according to spatial resolution. Experimental results in Section 4 show that the over-cost introduced by scalability is negligible at high bit-rates. Furthermore, at lower spatial resolution, significant quality improvement is perceived in comparison with non-scalable coding of motion.

2. GLOBAL FRAMEWORK

This work is an improvement of TWAVIX (for WAvelet-based VIdeo Coder with Scalability), whose performances have been proven comparable to state-of-the-art scalable solutions (Cf [7]).

2.1 TWAVIX overall architecture : MCTF + JPEG2000

TWAVIX is a (2D+t)WT coding scheme, briefly described in [8]. Like MC-EZBC, it performs first temporal analysis at full resolution, then spatial analysis and finally entropy coding of both motion and texture (see Fig. 1).

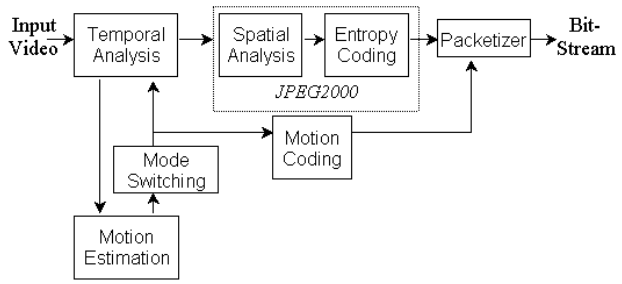

Figure 1: Overall architecture of TWAVIX

Motion is estimated thanks to a fast, eighth-pel accurate, hierarchical variable-size block matching algorithm, from 256x256 blocks down to 4x4 blocks. Multi-level motion field quad-trees are then pruned according to rate-distortion slopes of each node.

Depending on the video source, temporal analysis can consist in Haar MCTF, Backward or Forward Prediction, or Intra Coding. Regarding spatial analysis, texture coding and bit-stream layering, TWAVIX is coupled with JPEG2000 VM8.0 implementation.

Note that rather than adjust an average bit-rate over the entire encoded sequence, TWAVIX performs rate allocation independently for each GOF. This is far more realistic for applications involving varying bandwidth, such as *Video streaming over heterogeneous IP networks* and *Mobile streaming video* (Cf. [2]).

As regards motion coding, TWAVIX classically uses a non-scalable context-based adaptive arithmetic coding, inspired from [9]. Motion fields are computed and encoded at full spatial resolution into a single-layer bit-stream.

2.2 Motion compensation at lower spatial resolution

Let us stress that even if motion estimation is performed at original resolution, TWAVIX, unlike many scalable codecs in the literature, does not systematically reconstruct full resolution frames before performing temporal synthesis. All computations are processed at the real decoded resolution, by rescaling original motion field quad-tree structure and vector components. This choice is motivated by reality of applications (Cf. [2]) : we do not imagine a cellular phone or a PDA can afford to perform motion compensation at SD or HD resolution.

This means that motion compensation at decoder side will not systematically be processed at the same resolution as at encoder side. Consequently sub-pel interpolation demands a special care as regards filter size for each sub-resolution. Actually we use 8-tap FIR filters at original resolution and bilinear interpolation at lower resolutions.

3. SCALABLE MOTION CODING

In state-of-the-art coding schemes, motion parameters are usually coded losslessly as side-information. The tradeoff between the volume of information and the efficiency of motion compensation and energy compaction has been widely recognized. In non-scalable coders, various techniques have been used to optimize the number of bits spent on motion for a target bit-rate, but in scalable coding the target bit-rate is unknown.

3.1 Natural motion scalability : accuracy & block size

Optimally, at decoder side, rate-adapted motion subsets should be available to maximize video quality. However, in a (2D+t)WT scheme, temporal filtering has been performed once with full resolution motion field (Cf. Fig. 2). The point is therefore to deduce subsets from this original resolution motion field, that will allow the decoder to preserve a reasonable motion/texture ratio, without penalizing too much motion compensation quality.

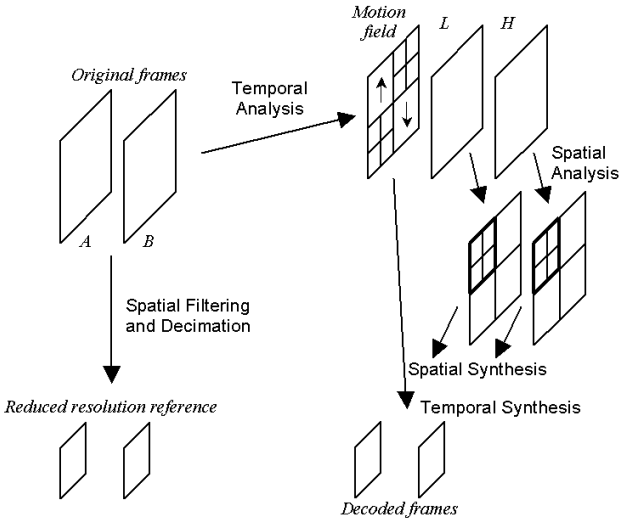

Figure 2: Spatial scalability in a (2D+t)WT framework

At low bit-rates, video is usually decoded at a reduced spatial resolution, so hand high-precision motion vectors are virtually useless. Besides, smallest blocks tend to vanish.

These statements could motivate to apply on motion information the same coding techniques as those used for texture samples (spatially-scalable transform and progressive coding), like in [6]. Let us first point out that unlike the triangular mesh motion model used by Taubman & Secker, our variable-size block-based motion description does not suit a spatial transform, but presents inherent sparseness properties thanks to pruning.

Let us moreover note that a three-resolution scenario (e.g. QCIF-CIF-SD) is not sufficient to take advantage of block-size scalability. Even smallest (4x4) blocks of SD resolution do not disappear at QCIF resolution, at least for luminance. In addition, discarding blocks that should still exist induces annoying visual artifacts. So in such a configuration, it seems relevant to rely on accuracy scalability.

3.2 Accuracy-scalable motion coding

Having investigated the impact of accuracy at decoder side, it appears that its utility decreases with spatial resolution. Once spatially filtered and decimated, temporal low and high frequencies do not benefit from the sub-pel accuracy that has been used at original resolution during temporal analysis.

This leads us to parting the bit-stream into accuracy layers. But unlike in [6] where each bit-plane of motion is di-

vided in several coding passes, we only introduce as many truncature points as decoded resolutions, in order to confine scalability over-cost.

For a three-level scenario, the optimal layering seems to consist of two enhancement layers of one-level accuracy, and a base layer of the corresponding approximate field. Figure 3 shows an example corresponding to ⅛ pixel-accurate motion estimation with SD video source.

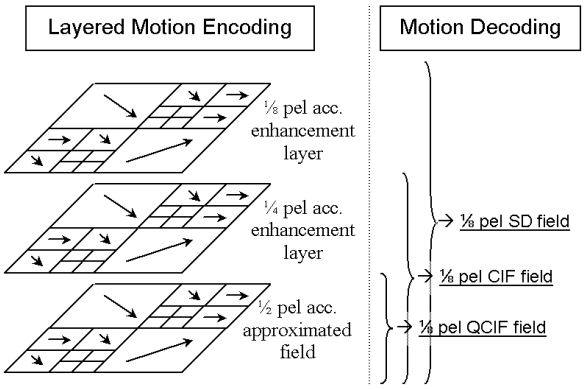

Figure 3: Example of three-level accuracy-scalable motion coding-decoding

After having encoded once the quad-tree structure, prediction residue of each vector of the base layer are encoded with our context-based adaptive arithmetic coder inspired from [9]. Then enhancement layers are successively similarly encoded, but without prediction coding since these layers can be assimilated with noise.

This simple and systematic technique allows to perform the same level of sub-pel interpolation through all decoding resolutions, while saving some bit-budget for lower resolutions. For full resolution, a certain overcost is observed in comparison with non-scalable coding. This may be legitimately interpreted as the cost of scalability. Indeed, scalability inevitably lowers prediction and entropy coding efficiency.

4. EXPERIMENTAL RESULTS

Results provided in this section correspond to CfP scenario 1 (Cf. [1]). They are obtained by encoding once 704x576 60fps sequences CITY and ICE, then performing the decoding at the various bit-rates, frame-rates and resolutions described in Table 4.

width	height	frames/s	Kbit/s
176	144	15.0	64
176	144	15.0	128
352	288	15.0	192
352	288	30.0	384
352	288	30.0	750
704	576	30.0	1500
704	576	60.0	3000
704	576	60.0	6000

Table 4: CfP scenario 1 spatio-SNR-temporal scalability tests

Figure 5 compares motion budgets in percentage at each of these configurations for ICE sequence, using our accuracy-scalable motion coder and a single-layer motion coder.

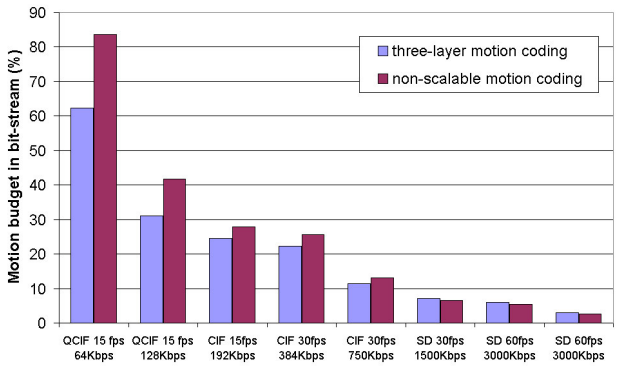

Figure 5: Motion bit-budget percentage in global bit-stream for ICE sequence

For reasons of brevity, we present here average PSNRs on luminance component, with, for clarity over the wide range of bit-rates, a logarithmic abscissa scale (see Figures 6 and 7 respectively for ICE and CITY).

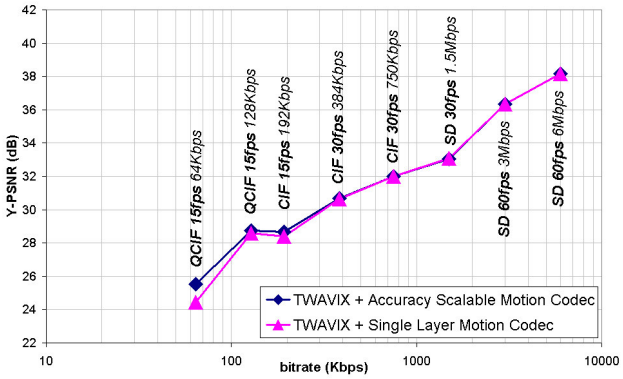

Figure 6: PSNR results for ICE sequence

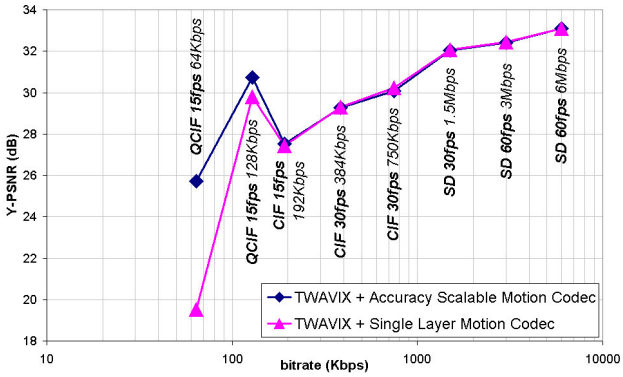

Figure 7: PSNR results for CITY sequence

As one shall notice, there are five different spatio-temporal configurations in Table 4, namely QCIF 15fps, CIF 15fps, CIF 30fps, SD 30fps, and SD 60fps. There are therefore five different reference sequences. These references have been defined in CfP procedure (see [1] Annex B). For lower spatial resolutions, these sequences are obtained by down-sampling using normative filters. For lower frame-rates, reference sequences are obtained by frame-skipping, keeping even frames and discarding odd ones.

Note that in terms of PSNR, these specifications do not favor (2D+t)WT solutions, which perform temporal filtering

instead of rough decimation, and spatial low-pass filtering does not match MPEG-4 filters.

Finally, Figure 8 illustrates the visual quality gain that can be obtained with scalable motion codec in comparison with non-scalable one, at resolution QCIF 15fps, 96Kbps.

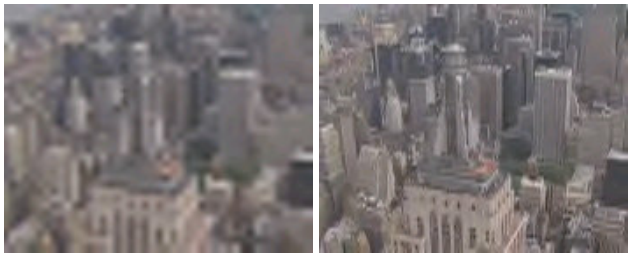

Figure 8: CITY QCIF 15fps 96Kbps, with non-scalable motion codec (left), and with accuracy-scalable motion codec (right)

5. CONCLUSION

A new scalable video coding scheme has been presented, that introduces scalable coding of motion in addition to full-scalable coding of texture. Motion codec principle consists of a layer partitioning according to accuracy in order to fit the level of spatial scalability. Although simple, this technique allows to cover a very wide range of bit-rates and improves significantly video quality at lower spatial resolutions without any noticeable penalty at high bit-rates and full resolution.

REFERENCES

[1] ISO/IEC JTC1/SC29/WG11, "Call For Proposals On New Tools For Scalable Video Coding Technology", *MPEG Document N5958*, Brisbane, October 2003.

[2] ISO/IEC JTC1/SC29/WG11, "Requirements and Applications for Scalable Video Coding", *MPEG Document N6052*, Gold Coast, October 2003.

[3] H.-M. Hang, S.S. Tsai, and T. Chiang, "Motion Information Scalability for MC-EZBC : Response to the Call for Evidence on Scalable Video Coding", ISO/IEC JTC1/SC29/WG11, *MPEG Document M9756*, Trondheim, July 2003.

[4] G. Boisson, E. François, D. Thoreau, and C. Guilemot, "Motion-Compensated Spatio-Temporal Context-Based Arithmetic Coding for Full Scalable Video Compression", *Picture Coding Symposium*, Saint Malo, France, April 2003.

[5] Y. Andreopoulos, M. Van der Schaar, A. Munteanu, J. Barbarien, P. Schelkens, and J. Cornelis, "Complete-To-Overcomplete Discrete Wavelet Transforms for Scalable Video Coding with MCTF", *Proc. SPIE/IEEE Visual Communication and Image Processing*, Lugano, Switzerland, July 2003.

[6] D. Taubman and A. Secker, "Highly Scalable Video Compression With Scalable Motion Coding", *Proc. IEEE Intenational Conference on Image Processing*, Barcelona, Spain, September 2003.

[7] ISO/IEC JTC1/SC29/WG11, "Report on Call for Evidence on Scalable Video Coding technology", *MPEG Document N5701*, Trondheim, July 2003.

[8] G. Marquant, J. Viéron, G. Boisson, P. Robert, E. François and C. Guillemot, "Response to the Call for Evidence on Scalable Video Coding Advances", ISO/IEC JTC1/SC29/WG11, *MPEG Document M9784*, Trondheim, July 2003.

[9] D. Marpe, G. Blättermann, G. Heising, and T. Wiegand, "Video Compression Using Context-Based Adaptive Arithmetic Coding", *Proc. IEEE Intenational Conference on Image Processing*, Thessaloniki, Greece, September 2001.