

Fine grain scalable video coding using 3D wavelets and active meshes

Nathalie Cammas^a, Stéphane Pateux^b

^aFrance Telecom RD,4 rue du Clos Courtel , Cesson-Sévigné, France

^bIRISA, Campus de Beaulieu, Rennes, France

ABSTRACT

This article introduces a novel approach for scalable video coding based on an analysis-synthesis scheme. Active meshes are used to represent motion model, this permits to exploit temporal redundancy along motion trajectories in a video sequence using temporal wavelet transform. The use of 3D wavelets in the coding strategy provides natural scalability functionalities to the video coder. Furthermore, the analysis-synthesis scheme allows to decouple motion and texture and to code these informations separately. Motion can then be lossy coded, bitrates gain can be reported to texture coding. Because motion is lossy coded, a new quality criterion measured in the texture domain is then proposed.

Finally, the proposed analysis-synthesis video coding scheme overcomes some of the limitations of existing video coding schemes using 3D wavelets, limitations due for the most part to the use of block-based motion model. Our video coding scheme performs as well as fully optimized H26Lv8, while providing a scalable bitstream.

Keywords: analysis-synthesis, scalability, active meshes, 3D wavelet, video coding

1. INTRODUCTION

Video coding is applied in various applications that require different kinds of resources and transmission over network with variable bandwidth. In order to meet these requirements, video coding must supply scalability functionalities. The scalability offers several hierarchical levels of representation of the information, it enables a compressed video bitstream to be decoded at successive progressive bitrates and with successive increasing qualities.

Wavelet is a good tool for scalability, it provides a hierarchical representation of the information. Furthermore, the resulted subbands are orthogonal and are well suited for rate-distortion (RD) optimization and progressive transmission. The idea is to extend 2D wavelet image coding as used in JPEG-2000, EZW or SPIHT to the 3D case of video sequences. To this extent, wavelet transform has to be performed on the three axes of the video frames: temporal, horizontal, and vertical, see figure 1.

Wavelets efficiency relies on the correlation of the signal, the more correlated the signal is, the more efficient the wavelet decorrelation is. As the video signal is almost constant along motion trajectories, temporal wavelet transform must exploit motion compensation in order to take advantage of temporal correlation.

In this context, we propose a novel approach for scalable video coding scheme based on an analysis-synthesis approach. This approach is achieved by decoupling motion and texture information. Decoupling is done by using active meshes as motion compensation tool. With this method, we overcome some of the limitations of previously proposed schemes using 3D wavelets. In section 2 of the article, we discuss 3D wavelet video coding schemes that have already been proposed and their limitations, then section 3 explains our analysis-synthesis scheme. Experimental results are presented in section 4 and section 5 concludes the article.

Further author information: (Send correspondence to Nathalie Cammas)

Nathalie Cammas: E-mail: ncammas@irisa.fr, Telephone: +33 (0)2 99 84 74 24

Stéphane Pateux: E-mail: spateux@irisa.fr, Telephone: +33 (0)2 99 84 73 60

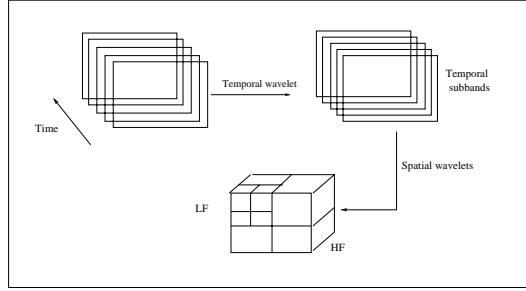


Figure 1. 2D+t wavelet transform

2. STATE OF THE ART IN 2D+T WAVELET CODING

Various video coding schemes using 3D wavelet have already been proposed. Taubman and Zackor¹ proposed a 3D subband video coding scheme performing 3D wavelet transform on spatially pre-aligned video frames, but motion model used was a global motion field.

Block-based temporal wavelet transform have been proposed by Ohm,² and by Choi and Woods.³ In these schemes, a 3D separable wavelet transform is performed on displaced blocks from the video frames, see figure 2. The use of block-based motion implies the appearance of disconnected and multiple connected pixels at blocks' boundaries, see figure 2. The handling of these particular pixels limits the wavelet filter to be a Haar filter in both schemes.

Another approach of wavelet coding introduces the lifting wavelet transform. Secker and Taubman⁴ perform temporal transform concomitantly with motion compensation. The wavelet transform is performed using the 5-3 lifting implementation, lifting thereby enables wavelet transform to be inverted without loss. But invertibility implies that both backward and forward motion fields are needed, involving high motion cost, see figure 3. Another scheme⁵ proposed to use a truncated 5-3 lifting filter to perform temporal transform on a group of nine frames. The use of truncated 5-3 allows to reduce motion costs, as frames are predicted using bi-directional compensation, see figure 4. The scheme turns out to be an extended IPB scheme but works in an open loop rather than in a traditional closed loop.

In,⁶ long temporal filters are used in lifting scheme performed on a group of frames. Adaptive filters are used when motion estimation failed to find texture trajectories, in the case of block-based motion compensation, this happens when unconnected or multiple connected pixels appear. Temporal wavelet filter coefficients depends on the trajectories estimated. For example, in the case of multiple trajectories, a criteria is minimized to find the best trajectory. At boundaries frames, subbands calculated at previous level decomposition are used from neighbors frames outside the GOF.

In,⁷ a comparison study on these approaches was presented. This study shows that when using rate-constrained block-based motion field, short temporal wavelet filters performed better than long filter. The use of forward and backward block-based motion field was a solution to the problem of unconnected or multiple connected pixels appearing in² or.³ When motion is not rate-constrained, long filters, like 5-3 or 9-7 filters, better compact energy in high temporal frequencies subbands than short filters, like Haar filters. But, when taking into account motion coding cost, the situation is completely different. Motion coding cost has a strong impact on overall performances, energy compaction is better using short filters than long filters. The bi-directional prediction scheme using truncated 5-3⁵ appears to be the better trade-off between energy compaction and motion coding cost.

3. PROPOSED SCHEME

3.1. General overview

In order to lower limitations, we then proposed an analysis-synthesis video coding scheme, see figure 5. The analysis stage permits to decouple motion and texture informations. Texture information can then be decorrelated using wavelet temporal transform and 2D spatial wavelet transform. Spatio-temporal subbands are then coded

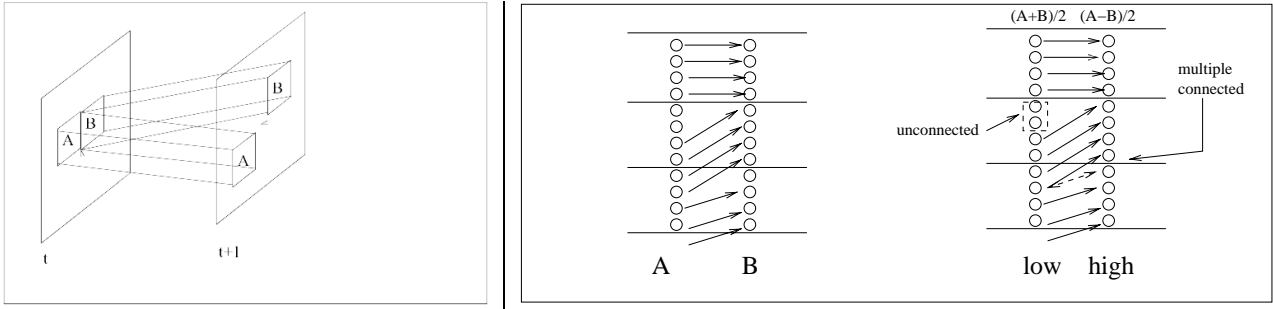


Figure 2. Block-based temporal wavelet transform

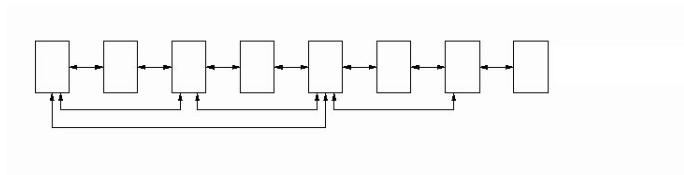


Figure 3. Forward and backward motion compensation using 9/7 lifting filter

using codecs providing scalable and progressive bitstream. The synthesis stage reconstructs the video sequence using motion and texture information. An analogy can be made with model-based approaches. A 3D active model of the scene and textures images to be mapped are used to reconstruct the scene. In our case, the model is represented by active meshes. Textures to be mapped are the texture informations given by the analysis stage.

3.2. Motion estimation

The proposed scheme is based on an analysis-synthesis approach. The analysis stage works on group of N frames (GOP) and performs motion estimation between frames using active meshes^{8,9} see figure 6 for an example of motion estimation between successive frames using active meshes.

Active meshes are the best tools for motion compensation, they provide long term continuous tracking of the texture which justify the use of wavelet transform along motion trajectories performed later. Motion estimation is performed as in.⁹

After motion estimation, frames are then mapped on reference grids, like in.¹ This step allows to separate motion and texture informations.

3.3. Temporal transformation

Using active meshes as motion model justifies the use of wavelet transform along motion trajectories. The use of 3D wavelet provides natural scalability functionalities to our video coding scheme. To apply 3D wavelet transform, a temporal transform is first performed along motion trajectories. The use of reference grids to represent texture permits to use textures independently from motion during the temporal transform. Motion

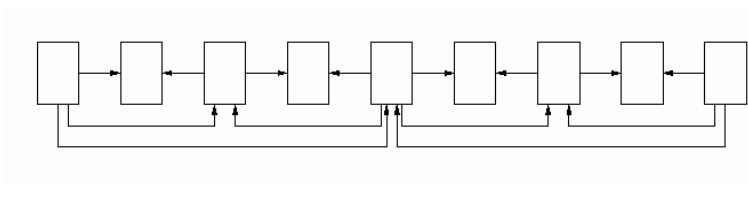


Figure 4. Bi-directional motion compensation using truncated 5/3 lifting filter

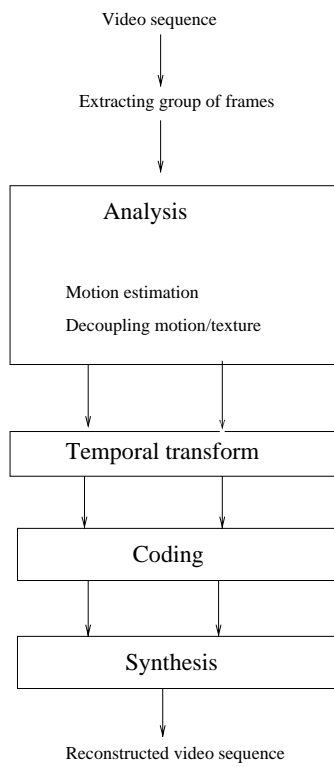


Figure 5. General view of the analysis-synthesis video coding scheme

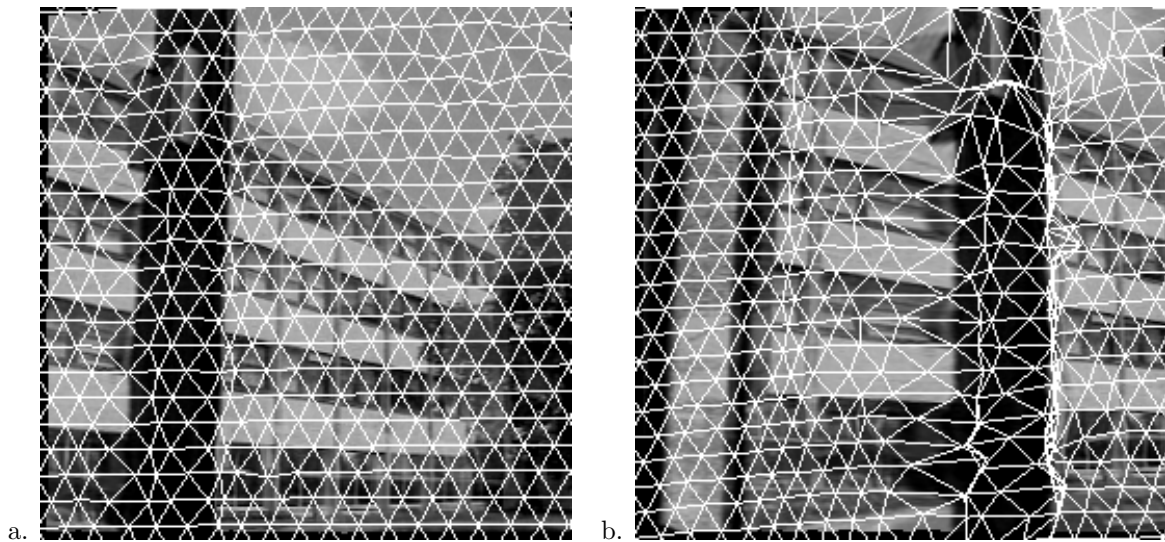


Figure 6. Motion estimation between successive frames

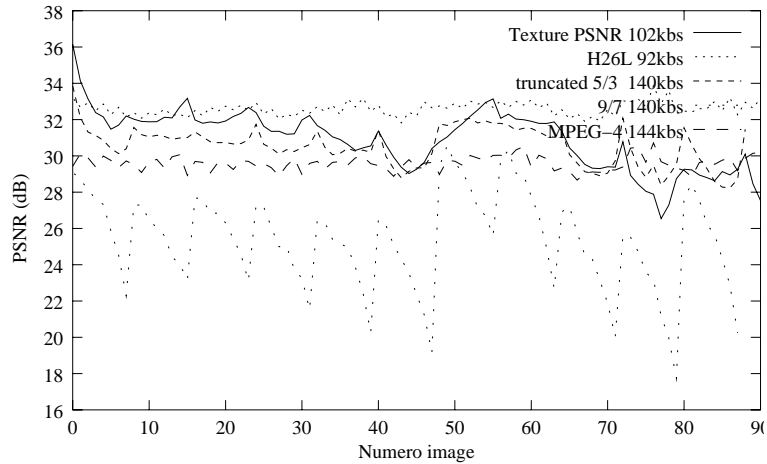


Figure 8. Comparison on Foreman sequence, 15Hz

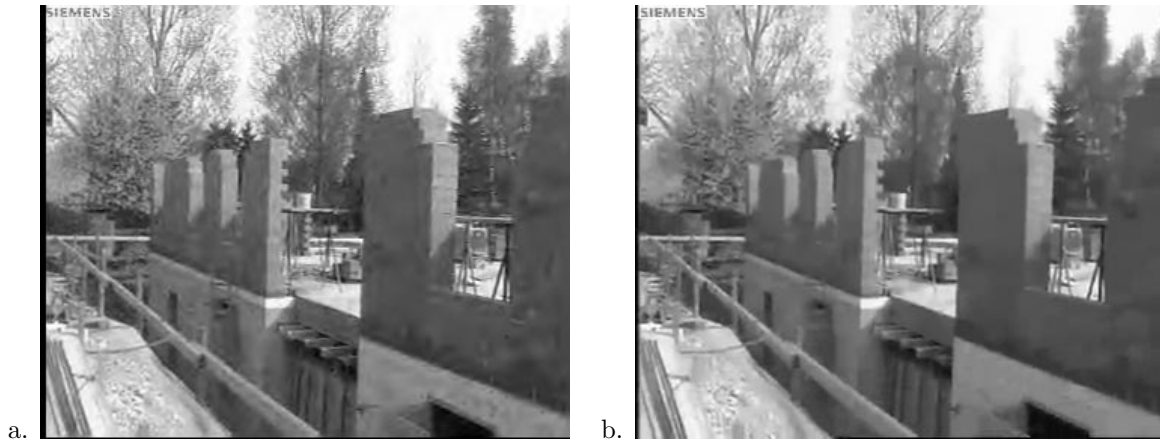


Figure 9. Foreman, frame 117, a: analysis-synthesis, 102kbs and b: H26L, 92kbs

4. RESULTS

4.1. Experimental conditions

The proposed scheme has been tested on various sequences and at different bitrates. We present results obtained on sequence Foreman Cif, YUV420, 15Hz and sequence Tempete Cif, YUV420, 30Hz. Figure 8 compares results obtained by our video coding scheme on Foreman sequence for various codecs: H26Lv8 (with fully optimized profile, IPBB frame structure, QP=31, RD optimization enabled and arithmetic coding), MPEG-4 Momusys, and two coders based on 3D wavelets.⁷

The first 3D wavelet video coder uses block-based motion model and performs temporal wavelet transform with motion compensation using a 9-7 lifting filter. Temporal subbands are coded using a progressive scalable coder and motion cost is set to 45% of the total bitrate (this percentage has been revealed experimently to give best performances at low bitrates). Motion compensation model uses both forward and backward motion field, see figure 3.

The other 3D wavelet video coder also uses block-based motion model, but performs temporal wavelet transform using a truncated 5-3 lifting filter. Frames prediction is bi-directional, see figure 4.

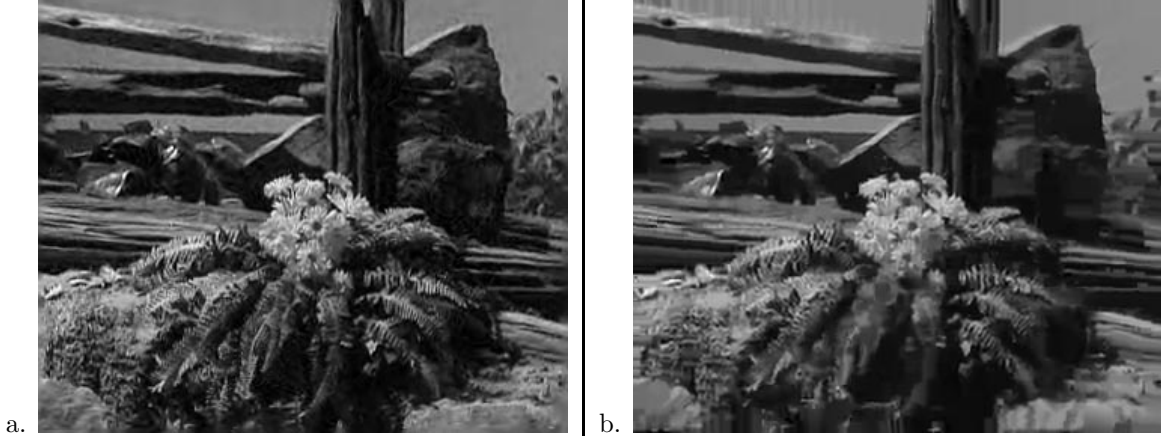


Figure 10. Tempete, frame 75, a: analysis-synthesis, 105kbs and b: H26L, 90kbs

| Sequence | H26L | 3D wavelet+meshes, a | 3D wavelet+meshes, b |
|-----------|----------------|----------------------|----------------------|
| Rue(25Hz) | 100kbs-27.46dB | 100kbs-28.85dB | 160kbs-29.16dB |

Table 1. PSNR H26L and 3D wavelets+meshes

4.2. Discussion

4.2.1. Comparison with block-based 3D wavelet video coders

Results (figure 8) show that block-based video coder using 3D wavelet performs better when using short temporal wavelet filter than long temporal wavelet filter. This is because block-based motion field is rate-constrained, and so it is not perfect. This implies discontinuities at block's boundaries during motion compensation, block's artifacts are visible in temporal frequencies subbands. These discontinuities are not well suited to a temporal wavelet transform; resulting high frequencies subbands present high energy and are difficult to code.⁷ That is why the truncated 5-3 filter performs better than the 9-7 filter (figure 8).

Active meshes used as motion compensation tool provide continuous tracking of texture frames even if motion is rate constrained and so are more suited for temporal wavelet decomposition. Active meshes permits to use longer temporal wavelet filter than block-based motion model. This is shown in figure 8, the analysis-synthesis scheme using active meshes and 5-3 lifting filter performs as well as the block-based video coder using truncated 5-3 lifting filter.

Furthermore, continuous tracking of texture by active meshes allows to code motion informations in a lossy way, which is not allowed with block-based motion model. Block-based motion model must have a good motion model to avoid block's artifacts in temporal frequencies subbands. In the case of forward and backward motion fields, active meshes have the advantage to be invertible, and so only one motion field is needed to have forward and backward motion fields.

4.2.2. Quality comparison with H26L

Figures 9 and 10 show visual rendered quality of the reconstructed sequence with our method and with H26L for Foreman sequence and Tempete sequence. The analysis-synthesis scheme permits to render more details of the video sequence than H26L, which renders smoothed reconstructed video sequence, due to strong deblocking filters used to delete block's artifacts. Furthermore, the analysis-synthesis video coding scheme provides natural scalability functionalities to the compressed bitstream. Texture and motion can each be decoded separately at successive higher qualities associated with successive higher bitrates.

Table 1 show PSNR results on sequences Rue 25Hz for several bitrates for the analysis-synthesis scheme. The first number is the total bitrates of the compressed bitstream, the second number is the PSNR of the



Figure 11. Rue, a: analysis-synthesis, 100kbs, frame 64 and b: H26L, 100kbs, frame 64

sequence. We see that the analysis-synthesis scheme outperforms H26L on this sequence for the same bitrate. As our bitstream is scalable, we can improve video quality increasing decoded bitrate as shown for bitrate 160kbs, without re-encoding the video sequence.

5. CONCLUSION

This article introduced a novel approach for video coding scheme based on an analysis-synthesis scheme using 3D wavelet and active meshes. The analysis-synthesis scheme provides a novel representation of the informations to encode. The analysis stage separates motion informations from texture informations. These informations are then coded separately. The synthesis stage reconstructs a video sequence of visual quality as close as possible from the original sequence, while allowing to reconstruct the video sequence at different qualities. The use of wavelets decomposition permits to provide natural scalability to the video coding scheme. Wavelets coupled with active meshes motion compensation tool provide continuous tracking of texture, these tools permits to better exploit temporal redundancy in a video sequence.

The analysis-synthesis scheme exploits the hypothesis that human eye is more sensitive in a video sequence to texture than to motion, that is human eye is less sensitive to motion errors than to texture errors. With this hypothesis, motion can be lossy encoded and the coding gain may be reported to texture coding.

REFERENCES

1. D. Taubman and A. Zakhor, "Multirate 3-d subband coding of video," *IEEE Transactions on Image Processing* **3**, pp. 572–588, september 1994.
2. J. Ohm, "Three-dimensional subband coding with motion compensation," *IEEE Transactions on Image Processing* **3**, pp. 559–571, September 1994.
3. S.-J. Choi and J. Woods, "Motion-compensated 3-d subband coding of video," *IEEE Transactions on Image Processing* **8**, pp. 155–167, february 1999.
4. A. Secker and D. Taubman, "Motion-compensated highly scalable video compression using an adaptative 3d wavelet transform based on lifting," *IEEE*, 2001.
5. L.Luo, J.Li, S.Li, Z.Zhuang, and Y-Q.Zhang, "Motion-compensated lifting wavelet and its application in video coding," in *IEEE International Conference on Multimedia and Expo*, August 2001.
6. Y. Zhan, M. Picard, B. Pesquet-Popescu, and H. Heijmans, "Long temporal filters in lifting schemes for scalable video coding," tech. rep., ISO/IEC JTC1/SC29/WG11 MPEG02/M8680, July 2002.
7. J. Vieron, C. Guillemot, and S. Pateux, "Motion compensated 2d+t wavelet analysis for low rate fgs vido compression," in *International Thyrrhenian workshop on digital communications 2002 (invited paper)*, 2002.

8. G.Marquant, S. Pateux, and C.Labit, "Mesh and "crack lines": Application to object-based motion estimation and higher scalability," in *IEEE International Conference on Image Processing, ICIP'00, vancouver, Canada*, September 2000.
9. S. Pateux, G. Marquant, and D. Chavira-Martinez, "Object moosaicking via meshes and crack-lines technique. application to low bit-rate video coding," in *PCS*, 2001.
10. F. Galpin and L. Morin, "Computed 3d models for very low bitrate video coding," in *Proceedings of the IEEE conference on Visual Communications and Image Processing, VCIP'2001*, **4310**, p. ?, 2001.