INTRA-ADAPTIVE MOTION-COMPENSATED LIFTED WAVELETS FOR VIDEO CODING

Oscar Divorra Escoda, Markus Flierl and Pierre Vandergheynst

Signal Processing Institute
Ecole Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland
{oscar.divorra,markus.flierl,pierre.vandergheynst}@epfl.ch, http://lts2www.epfl.ch

ABSTRACT

This paper investigates intra-adaptive wavelets for video coding with frame-adaptive motion-compensated lifted wavelet transforms. With motion-compensated lifted wavelets, the temporal wavelet decomposition operates along motion trajectories. However, valid trajectories for efficient multi-scale filtering have a finite duration in time. This is due to well known effects like occlusions or inaccurate motion estimation. These discontinuities may generate many non-zero wavelet coefficients when a transform with a fixed dyadic structure is used. To investigate the advantage of an adaptive transform, we introduce intra macroblocks in the frame-adaptive lifting steps. Depending on the rate-distortion costs at a given macroblock location, we choose the number of wavelet decomposition levels locally. We discuss motion-compensated lifted wavelets that are frame- and intra-adaptive. We evaluate the efficiency of intra-adaptive wavelets when frame-adaptive motion-compensated wavelets are used. We observe that intra-adaptivity is rate-distortion efficient for discontinuities that cannot be handled by frame-adaptivity.

1. INTRODUCTION

Currently, motion-compensated lifted wavelet transforms [1] are studied extensively as they combine excellent compression efficiency with the possibility of an embedded representation. To improve compression efficiency, adaptive lifting schemes are investigated for motion-compensated temporal filtering (MCTF). In particular, frame-adaptive motion-compensated lifted wavelets [2] and multi-hypothesis motion-compensated lifted wavelets [3] have been proposed in the framework of a MCTF extension of H.263++ [4]. Further improvements have been accomplished with the MCTF extension of H.264/AVC [5].

In this work, we investigate intra-adaptive wavelets. We review the theoretical background of non-linear approximations of piecewise-smooth signals based on wavelet transforms. With this, we motivate the temporal adaptivity for wavelet transforms that improve the rate distortion efficiency. An analogy between motion compensated coding of video signals and coding of piecewise-smooth signals is presented. For non-stationary signals, adaptive decompositions are useful for efficient coding. Smooth signal parts and discontinuities should not be treated in the same way. We refer to well known rate distortion results for the class of 1D piecewise-polynomial signals [6, 7]. Interval adaptive coding techniques which permit separate encoding of smooth areas can outperform wavelet based transform coding.

For coding video signals, the motion-compensated lifting scheme can be extended by intra macroblocks. They allow separate encoding of intervals with smooth motion trajectories. Recall that frame-adaptive motion-compensated lifted wavelets [2] permit a flexible encoding of motion trajectories as long as sufficient candidate reference frames are available. The number of efficient reference frames decreases not only due to encoding constraints but also due to frequent scene changes. If frame-adaptive motion-compensated wavelets are not able to provide the desired flexibility, intra macroblocks will be used to permit separate encoding of intervals with smooth motion trajectories. We compare both intra-adaptive and frame-adaptive motion-compensated wavelets to study separate encoding of intervals with smooth motion trajectories.

The paper is structured as follows: Sec. 2 reviews non-linear approximations of piecewise-smooth signals based on wavelet transforms. Sec. 3 summarizes frame-adaptive motion-compensated lifted wavelets and discusses the intra-adaptive scheme. Experimental results are presented in Sec. 4. Finally, Sec. 5 concludes our investigation.

2. ADAPTING WAVELET EXPANSIONS: APPROXIMATING NON-STATIONARY SIGNALS

The use of motion compensation (MC) within the wavelet temporal representation has the purpose of performing filtering in the motion direction. This motion oriented filtering drastically reduces the number of significant wavelet coefficients generated in the temporal transform. Indeed, multi-scale redundancy can be exploited not only from those regions that remain unchanged in a period of time but also those objects subject to a motion through time. As long as the motion of the scene can be accurately estimated, the temporal signal encountered by the wavelet transform will be smooth or even constant if no local temporal illumination changes are present in the scene. When the motion cannot be estimated correctly, or simply, when there is an occlusion or an appearing object, the signal seen by the wavelet transform presents a step in amplitude. This step issues from the mismatch between the best signal sample candidate for prediction found by the MC and the signal sample being predicted. As one can expect, the representation of a step function needs a significant quantity of wavelet coefficients.

2.1 Piecewise-Smooth Signals: A Deterministic Model for Motion-Compensated Video Sequences

Often, signals are modeled by stationary jointly Gaussian stochastic models. However, real signals can have a quite different behavior. This is commonly the case for natural images and video sequences. Indeed, these are more often associated to a deterministic signal model that allows a better analysis of the R-D properties of different methods used to code them. A deterministic model that fits the signal behavior in video sequences is the so called *piecewise-smooth* signal model [6, 7].

The temporal behavior of all connected pixels by means of motion vectors may be seen as a piecewise-smooth signal. In this class of signals, wavelets have shown to be quite successful because they are specially suited for representing them. Thanks to their locality, these capture well abrupt changes in the signal. Moreover, smooth parts are efficiently approximated by the scaling functions of the wavelet basis. Nevertheless, wavelet transforms do not exploit the interrelation among coefficients from different subbands generated by an edge (see Fig. 1). Even though wavelet transforms are well suited for representing discontinuities, their R-D sub-optimality can be demonstrated as discussed in the following.

2.2 Optimal R-D Coding of Signal Discontinuities and the use of Wavelet Transforms

Wavelet based signal coding involves non-linear approximation [7] by means of partial reconstruction and coefficients quantization. The deterministic piecewise model serves to evaluate the R-D performance of wavelet based coding for this class of signals composed by the mixture of switching components and smooth parts.

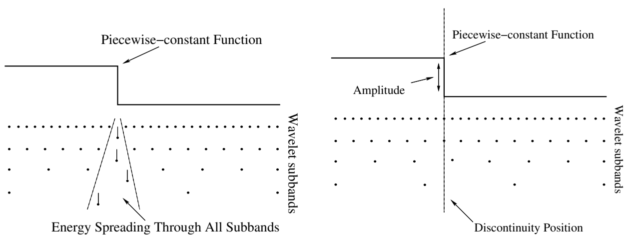


Figure 1: Left: Spreading of coefficients through the wavelet subbands of a 1D piecewise-constant signal representation. Right: There are no wavelet coefficients to code when using an oracle to code the discontinuities, i.e., position and amplitude.

Fig. 1 illustrates the response of a wavelet transform to a step function. If a wavelet with sufficient vanishing moments is used, all the polynomial areas can be represented by means of the coefficients of the low frequency bands (the scaling functions). In such case, the only part of the signal that generates wavelet coefficients are discontinuities. To accurately code the step using a wavelet transform, non-zero coefficients amplitudes and positions need to be coded. On the other hand, Fig. 1 shows a non-linear approach widely discussed in approximation theory [6, 7, 8]. This intends a more efficient representation of piecewise signals in general, assuming the existence of an oracle that tells where the switching points among smooth pieces are located. If this is the case, since very efficient approximations of the smooth intervals can be achieved, a better R-D behavior than in the case where only wavelets are used is possible. Indeed, it is more efficient to code separately discontinuities location and smooth parts. See in Fig. 1, in order to locate the edge and to set its size, it is just necessary to supply one position plus one amplitude. Moreover, the use of an independent representation in each of the intervals will not generate additional information to code, i.e. consider a Haar wavelet that is used in each one of these (where boundary effects and scaling functions coefficients are properly handled). Then, no non-zero coefficients will be generated. To the contrary, in the simple 1D wavelet case, the number of locations and amplitudes to code is proportional to the number of decomposition subbands.

In oracle based coding of piecewise-polynomial signals, the asymptotic behavior of distortion (D) at high rates is described as a function of rate (R) [6]:

$$D_O(R) \sim 2^{-B \cdot R}, \tag{1}$$

where B is a positive constant. In case of wavelet coding, the asymptotic behavior at high rates is worse:

$$D_W(R) \sim \sqrt{R} \cdot 2^{-A\sqrt{R}}, \tag{2}$$

where A is a positive constant. Unlike in (1), distortion decreases with exponent $\sqrt{R}$ which indicates a slower decay with the rate. Notice that even if the asymptotic R-D behavior is analyzed at high rate, it is sufficient to motivate the use of adaptive coding [6, 7], and to understand the coding efficiency of different approximation approaches.

3. ADAPTIVE MOTION-COMPENSATED LIFTED WAVELET TRANSFORMS FOR VIDEO CODING

The length of wavelet transforms (i.e. the number of wavelet decomposition subbands) may be adapted to cover smooth areas while avoiding wavelet kernels to cross edges. Our purpose is to find a R-D adapted decomposition of the video signal. In the case of motion compensated lifted wavelet transform for video coding, this adaptivity can be implemented by means of inserting the so called *intra* macroblocks as an additional coding mode for the coding scheme.

Our analysis and results of Sec. 4 are based on the multi-hypothesis, frame-adaptive motion compensated lifting scheme proposed in [2, 3]. This scheme already introduces some temporal adaptivity allowing a free selection of reference frames for MC within a GOP. Moreover, it allows an adaptive selection of the most suitable lifting step (e.g. Haar or 5/3) to achieve the minimum rate-distortion costs. Due to the dyadic decomposition, a fixed number of multi-scale subbands is forced. This limits the adaptation to temporal discontinuities and to motion misalignments as discussed in Sec. 2.

3.1 Frame-Adaptive Motion-Compensated Lifted Wavelets

Current research aims to combine the advantages of linear temporal transforms and efficient motion compensation. A promising scheme for exploiting successfully temporal redundancy is based on motion-compensated temporal wavelets implemented in the lifting scheme. The lifting scheme allows the inclusion of non-linear, non-invertible, operations into its ladder structure. Non-reversible operations such as quantization or motion compensation are possible without affecting the reversibility of encoding schemes. Motion-compensated lifting steps are introduced in [1]. The prediction/update steps are performed using the samples that motion vectors connect. [2, 3] use for the update steps the negative versions of the motion vectors used in the prediction steps.

The multi-hypothesis, frame-adaptive variation on the lifting scheme [2, 3] provides the possibility to select the best reference signal for a given lifting step. This allows an interesting improvement in fine scale detail subbands. However, this does not allow to adaptively choose in space and time the number of desired decomposition subbands for an efficient R-D compression.

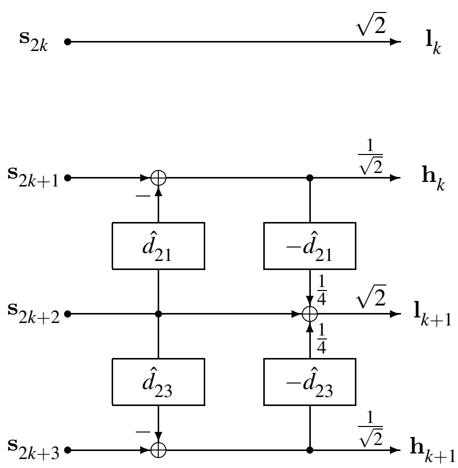


Figure 2: Example of the first decomposition level of the Haar transform with frame-adaptive motion-compensated lifting steps. The frame $\mathbf{s}_{2k+2}$ is used to predict frame $\mathbf{s}_{2k+1}$. $\hat{d}_{ij}$ denote the motion vectors.

Frame-adaptive MC lifted wavelets are a flexible approach permitting adaptive signal representations. They reduce occlusion effects, difficulties due to scene changes, and defects caused by inaccurate motion compensation at fine scales. Fig. 2 shows how frame adaptivity is implemented in the lifting scheme for the Haar wavelet case. The fix structure of the lifting scheme is broken such that any even frame in the GOP can be used to select the best prediction/update signals.

3.2 Intra-Adaptive Scheme

The frame-adaptive scheme does not change the default number of wavelet decomposition subbands nor considers alternative methods for MC at a particular location. In order to tackle this problem, the so called intra refresh may be introduced into MC lifted wavelet schemes. For example, recent work is also published in [5].

In the framework of MC lifted wavelets, intra refresh corresponds to the adaptive insertion of void lifting steps. These do not perform further temporal filtering on the signal. They implement the necessary breakpoints in the wavelet decomposition for an efficient R-D signal approximation (as discussed in Sec. 2). This special "lifting" mode is depicted in Fig. 3. With proper selection of this lifting mode, an approximation to the desired temporal wavelet decomposition with a local adaptation of the number of subbands is obtained.

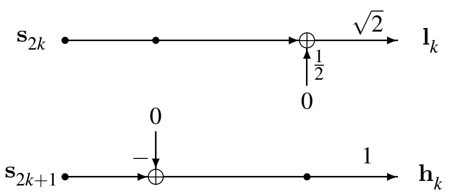


Figure 3: Broken lifting steps, neither prediction nor update are used. Both are inhibited on a macroblock level. Notice the change on the scaling factors to control the noise in the quantization stage.

Due to the absence of prediction and update steps, the scaling factors of the output signals have been modified. We adapt the output signals h_k to the fixed quantizer step size used for all subbands. Regarding the low band l_k, the scaling by $\sqrt{2}$ adapts the dynamic range of the signal to fit that of the next scale level for further decomposition.

Both frame- and intra-adaptivity are suitable to handle the discontinuities of motion-trajectories. If suitable reference frames are available, frame-adaptivity will be the best choice to reduce the energy in the detail subbands. But if suitable reference pictures are out of reach, the intra mode is required. Moreover, the final level of the dyadic decomposition offers only one reference picture. In that case, the intra mode is the sole alternative.

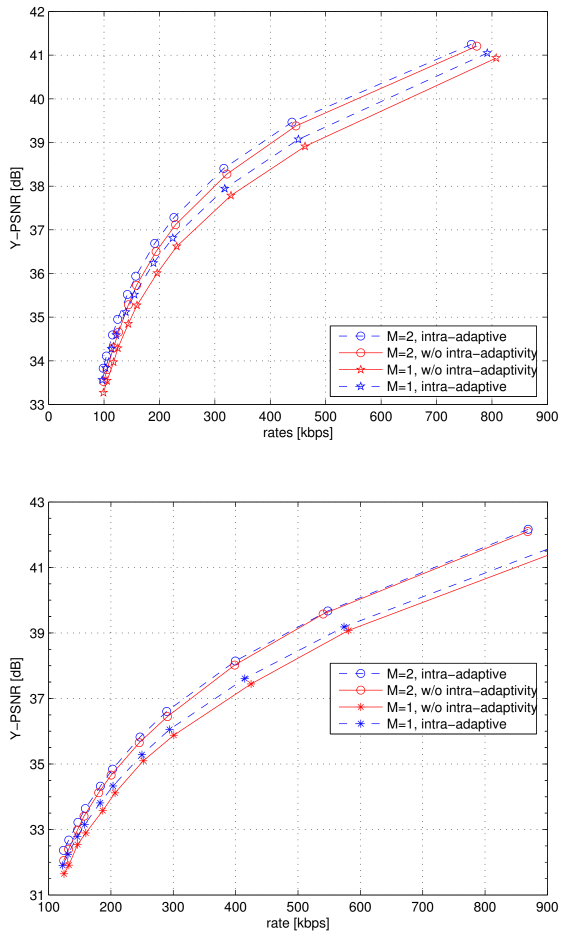


Figure 4: R-D comparison with intra-adaptivity for the entire sequence. Top: *CNN*. Bottom: *Foreman*.

4. RESULTS

We evaluate the benefits of temporal adaptivity and compare the improvements in R-D efficiency. Tests are performed with the QCIF sequences *CNN* and *Foreman* at 30 fps. Further results are given in the technical report [9].

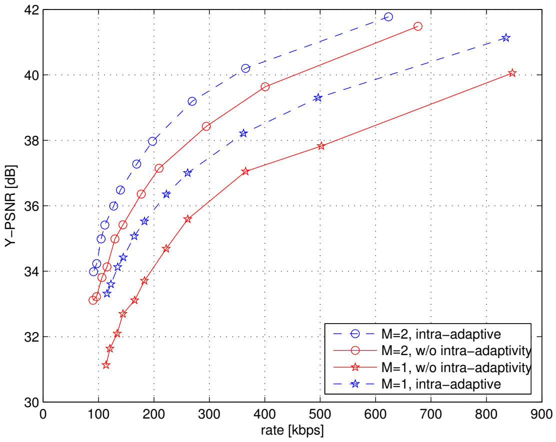
4.1 Coding Scheme

The coding scheme used to test the intra-adaptive approach is a MCTF extension of H.263++ [4, 2, 3]. The encoder chooses for each macroblock the best type of lifting scheme in a rate-distortion sense. This selection is carried out macroblock by macroblock minimizing the Lagrangian costs of the high band of the lifting scheme. Haar-, 5/3-, and void-type structures are used as candidates to encode each macroblock. Due to the flexibility of the frame adaptive scheme, $M = 1$ or $M = 2$ reference frames may be used by the algorithm to code each macroblock. Since the goal of this work is the experimental verification of the referenced theoretical concept, there are no further subdivisions of the macroblocks for motion compensation purposes. Motion vectors are obtained by block-based rate-constrained motion estimation jointly optimized with the lifting mode selection. The motion information is estimated at each decomposition level depending on the results of the lower level by using half-pel accurate motion compensation. All subband macroblocks generated by the temporal wavelet transform are encoded with the H.263 8x8 DCT codec. All intra-frame coded subbands are quantized with the same quantizer step size. Huffman codes are used for entropy coding and motion vectors are predicted from spatial neighbors. GOPs of size 32 are used in our experiments (up to five decomposition levels). Shorter wavelet transforms are provided by the intra macroblocks.

4.2 Experiments

Shorter instances of the lifted wavelet scheme (shorter than the maximum allowed GOP length of 32) contribute locally to the areas where motion trajectories are shorter than 32 frames. Hence, the benefit is going to be of local nature when particular characteristics of the sequence require it. Fig. 4 shows how the use of this additional coding mode introduces a moderate overall gain in R-D coding performance. Average improvements range from 0.2 to 0.5 dB for middle and low motion sequences with some scene changes and fast local motion. When using exclusively one reference frame for the prediction/update steps ($M = 1$ in Fig. 4), the intra-adaptive scheme provides slightly better global R-D improvements when compared to those obtained with two reference frames ($M = 2$).

Let us take a GOP where relevant changes appear in the image sequence and the MC lifting scheme cannot efficiently represent them. R-D gains can be as high as 1 dB: See the top chart in Fig. 5. In the same way as for the global R-D measure, when only one reference frame is used, intra-adaptive R-D improvement is slightly higher for $M = 1$ when compared to $M = 2$. The use of intra MBs appears coherent with the temporal scene changes or very fast mov-



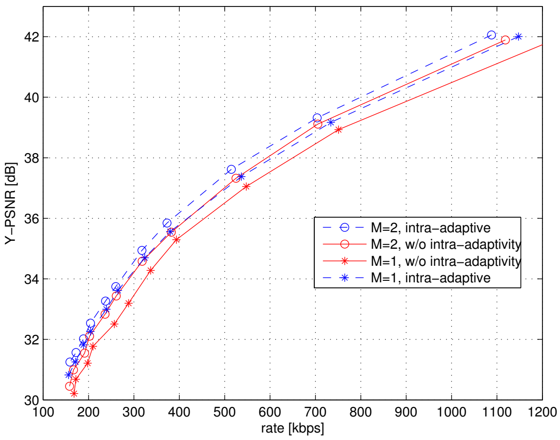


Figure 5: R-D comparison with intra-adaptivity for a particular GOP. Top: 8th GOP of *CNN*. Bottom: 6th GOP of *Foreman*.

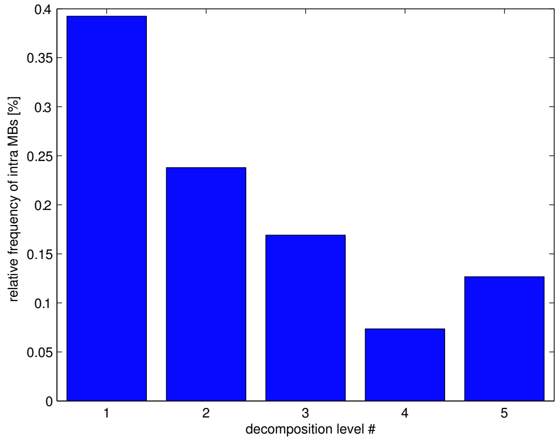


Figure 6: Average frequency of intra MBs depending on the decomposition level for a maximum GOP length of 32. The statistics have been collected for the sequences *CNN*, *Foreman*, and *Table Tennis* which have been coded at various bit rates. Up to $M = 2$ reference frames are utilized.

ing sequence periods.

Fig. 6 shows the average frequency of intra MBs at each decomposition level. The index below the column indicates the decomposition level in concordance to the scale of its associated wavelet in the case that non-adaptive lifting steps are used. For dyadic wavelets, the basis function scale evolves depending on the level j as 2^{-j} for $j \in \{1,2,3,4,5\}$. Lifting steps are split when middle or short length wavelet transforms are efficient. Intra MBs are more frequently allocated at low decomposition levels (1, 2 or 3). At high decomposition levels, there is a higher probability that the frame-adaptive scheme finds good reference frames.

5. CONCLUSIONS

This article discusses intra-adaptive motion-compensated temporal transform coding of video signals. Although the major signal partition for coding is the GOP structure of K pictures, we do not consider a fixed number of wavelet decomposition levels to obtain a fixed number of decomposition subbands. Our approach is such that GOPs of K pictures are adaptively broken in smaller units at the macroblock level. With this, the number of wavelet subbands in the temporal transform is adapted in space and time. Local signal breakpoints are coded independently by using intra macroblocks and wavelet kernels are reserved for those signal areas where efficient coding is achievable with the MC lifting scheme. In this paper, we refer to related theoretical work and discuss local spatio-temporal adaptations of MCTF schemes for video coding. The approach improves R-D performance as well as visual quality. Intra macroblocks are chosen if they achieve the smallest rate-distortion costs. Sudden scene changes trigger this mode and disrupt temporal filtering.

REFERENCES

[1] B. Pesquet-Popescu and V. Bottreau, "Three-dimensional lifting schemes for motion compensated video compression," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Salt Lake City, UT, May 2001, pp. 1793 – 1796.

[2] M. Flierl, "Video coding with lifted wavelet transforms and frame-adaptive motion compensation," in *Visual Content Processing and Representation*, Madrid, Sept. 2003, vol. 2849 of *Lecture Notes in Computer Science*, pp. 243–251, Springer.

[3] M. Flierl, P. Vandergheynst, and B. Girod, "Video coding with lifted wavelet transforms and complementary motion-compensated signals," in *SPIE Conference on Visual Communications and Image Processing*, San Jose, CA, jan 2004, vol. 5308, pp. 497–508.

[4] M. Flierl and B. Girod, "Investigation of motion-compensated lifted wavelet transforms," in *Proceedings of the Picture Coding Symposium*, Saint-Malo, France, Apr. 2003, pp. 59–62.

[5] H. Schwarz, D. Marpe, and T. Wiegand, *Scalable Extension of H.264/AVC (Proposal) - MPEG04/M10569/S03*, ISO/IEC JTC1/SC29/WG11, Munich, mar 2004.

[6] P Prandoni and M. Vetterli, "Approximation and compression of piecewise smooth functions," *Phil. Trans. R. Soc. Lond.*, vol. 357, no. 1760, pp. 2573–2591, Sept. 1999.

[7] M Vetterli, "Wavelets, approximation and compression," *IEEE Signal Processing Magazine*, vol. 18, no. 5, pp. 59–73, Sept. 2001.

[8] A. Cohen, I. Daubechies, O. Guleryuz, and M. Orchard, "On the importance of combining wavelet-based non-linear approximation in coding strategies," *IEEE Transactions on Information Theory*, vol. 48, no. 7, pp. 1895–1921, July 2002.

[9] O. Divorra Escoda, M. Flierl, and P. Vandergheynst, "Intra-adaptive motion-compensated lifted wavelets for video coding," Tech. Rep. 27.2004, LTS2 - ITS - EPFL, http:// lts2www.epfl.ch/ publications.php, nov 2004.