Intra-Adaptive Motion-Compensated Lifted Wavelets for Video Coding

Oscar Divorra Escoda, Markus Flierl and Pierre Vandergheynst
Ecole Polytechnique Fédérale de Lausanne (EPFL)
Signal Processing Institute (ITS)
CH-1015 Lausanne, Switzerland
Technical Report No. 27.2004

Abstract

In this work, we study the effect of inserting spatially local temporal adaptivity to motion compensated frame adaptive transforms for video coding. Motion compensation aligns the temporal wavelet decomposition along motion trajectories. However, valid trajectories for an efficient multi-scale filtering have a finite duration in time. This is due to well known effects like occlusions or inaccurate motion estimation. Hence, the signal encountered by the temporal wavelet transform can be seen as a piecewise-smooth signal. The breakpoints of this signal, may generate many wavelet coefficients when transforms with a fixed number of subbands are used. Theoretical results indicate that adaptive transformations can do better with this kinds of signals than simple wavelet transforms. In this paper we discuss the usage of a motion compensated lifting scheme that adapts the number of decomposition levels depending on the spatial location by means of a R-D criteria (use of a *Best Basis*). This adaptation is done by introducing the use of Intra Macroblocks in the lifting steps. This allows to virtually adapt the GOP length used for the wavelet decomposition in a local fashion. A detailed analysis of the benefits in terms of R-D and visual quality corroborates the expected improvement suggested by theory on the compression of piecewise smooth signals.

Index Terms

Sparse Approximations, Lifting Scheme, Motion Compensation, Temporal Adaptivity, Video Coding, Intra Macroblock, Wavelets, Piecewise Smooth Signals

Contents

Web page: http://lts2www.epfl.ch
The work of Oscar Divorra Escoda is sponsored by the IM.2 NCCR

I. Introduction

For still image compression, wavelet coding schemes are widely favored today, as they combine excellent compression efficiency with the possibility of an embedded representation. Temporal wavelet coding for video sequences, however, has not had such success. This can be attributed to the difficulty of incorporating motion compensation into the temporal subband decomposition. The discovery of motion-compensated lifted wavelet transforms had lead to renewed interest and the hope that temporal subband coding might ultimately outperform predictive hybrid coding, predominant in all current video coding standards [1].

Recent studies propose more and more flexible schemes able to adapt as much as possible to the particular nature video that signals present at every spatial and temporal location. This has been tackled, among many other approaches, by means of motion compensated lifting schemes [2], [3], the use of frame-adaptive lifting schemes [4] and lately the whole advanced motion compensation framework developed for the standard H.264 [5], [6].

In this work, we review the theoretical background behind non-linear approximations of piecewise-smooth signals based on wavelet transforms. This has the purpose of explaining why temporal adaptivity in wavelet transforms is important for an optimal R-D performance in video coding. An analogy between motion compensated video signal coding and coding of piecewise-smooth signals is presented. Non-stationary signals require adapted decompositions for being efficiently encoded. Indeed, smooth signal parts and discontinuities can not be treated in the same way. On this basis, we borrow the well known results of the R-D behavior for the class of 1D piecewise-smooth signals when coded by means of wavelet transforms, and compare them to those of non-linear, interval adaptive, based coding [7], [8], [9] which encodes separately signal breakpoints and smooth areas. In practice, however, this adaptivity is not new. This can be introduced in the MC lifting scheme by means of using Intra Macroblocks (MBs). Up to now the use of Intra MBs had just reported to have a better performance when MC was not able to efficiently predict the signal. In this work we situate the use of Intra MBs within the lifting scheme as an implementation of separate coding of breakpoints in piecewise-smooth signals (as described above). In this work, we investigate the extension to Intra-adaptive of MC lifting schemes. This report compares both approaches: Intra-adaptive and non temporally adaptive MC lifting schemes to balance on the overall benefits of using a locally temporally adapted wavelet transform.

The present work is structured as follows. Sec. II establishes the relation between MC lifting based video coding with non-linear approximations of piecewise-smooth signals based on wavelet transforms. Latter, in Sec. III, a recall of frame-adaptive lifting schemes is performed and the lifting step necessary for Intra-adaptivity is discussed. Intra-adaptivity is widely evaluated with several tests and results in Sec. **??**. Finally, conclusions are drawn in Sec. V.

II. Adapting Wavelet Expansions: Approximating Non-stationary Signals

The main topic in this work is adaptive video coding by means of an adaptive MC lifting scheme. In this section, we briefly review from a 1D point of view why wavelets are interesting in an approximation sense to represent and code the temporal dimension of video signals. In here, we assume that, as far as there is motion to track, the temporal motion-compensated wavelet transform is perfectly aligned with this, even if it may be seen as an ideal situation. In such a case, fixed length temporal subband representations, as used for video coding in [1], are suboptimal. This is due to the fact that abrupt transitions in the signal (e.g. sharp termination of a motion trajectory) may generate many wavelet coefficients. In the following, we introduce the concept of a *piecewise-smooth* signal model, its interaction with wavelet representations [10] and the relation with motion compensated temporal filtering (MCTF) based video representation. Based on this model, we discuss the need for local adaptivity in the number of subbands used for multi-scale subband representation for an optimal coding in Rate-Distortion terms.

A. The Piecewise-Smooth Model

The use of motion compensation within the wavelet temporal representation based on the lifting scheme has the purpose of performing the lifting filtering in the direction of motion. This motion oriented filtering drastically reduces the number of significant wavelet coefficients generated in the transform. Indeed, in this way, 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. The effect of including motion compensation from the transformation point of view, is to smooth out the signal that is going to be transformed with the temporal wavelet transform. 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 can not 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 wavelet representation of a step function needs a significant quantity of high amplitude wavelet coefficients. In Sec. II-B we review the rate distortion consequences of that.

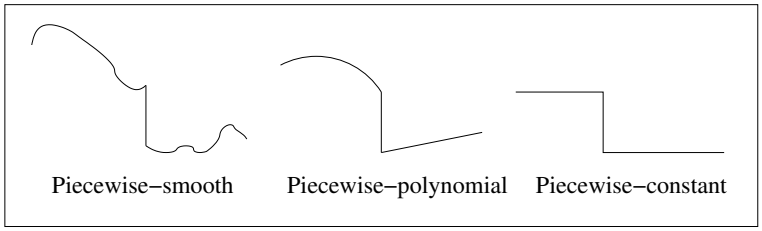

Fig. 1. Example of 1D piecewise-smooth, piecewise-polynomial, piecewise-constant signals.

Often, signals are modeled by stationary jointly Gaussian stochastic models. However, real signals can have a quite different behavior from the stationary Gaussian model. This is commonly the case for natural images, and video sequences [11], [7], [10]. Indeed, these are more commonly associated to a deterministic signal model that allows a better analysis of the R-D properties of using wavelet transforms for their coding. This deterministic model that fits the behavior of signals and video sequences is the so called *piecewise-smooth* signal model [10] (see Fig. 1). As can be inferred intuitively from the description above, from a 1D point of view, the temporal behavior of all the connected pixels by means of the motion vectors corresponds to a piecewise-smooth signal. Notice, however, that if no MC was applied, the suitable model for the signal transformed by the wavelet, in a real world sequence, would be, still, the piecewise-smooth model. The difference would be that the number of discontinuities in the signal to be transformed would significantly increase.

Fig. 1 depicts an example of a piecewise-smooth 1D signal, as well as two more examples that represents two particular cases of the piecewise-smooth class of signals. These particular cases are the piecewise-polynomial case, and the piecewise-constant case. As depicted in the picture, these are a kind of signals composed, respectively, by polynomial and constant signal intervals separated by singularities. In this class of signals, wavelets have shown to be quite successful because they are specially suited for representing them. Thanks to the locality of wavelets, these capture well abrupt changes in the signal. Moreover, smooth or stationary parts are efficiently represented by the coarse approximations obtained from the scaling functions of the wavelet basis. Nevertheless, wavelet transforms do not exploit the interrelation among wavelet coefficients from different subbands generated by an edge (see Fig. 2). Thus, even though the wavelet transform is well suited for representing discontinuities, as discussed in II-B it reveals to be suboptimal in terms of R-D.

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

Wavelet based signal coding involves non-linear approximation of the signal being coded [9]. This means that after generating the wavelet representation of a signal, coefficients are quantized, some are dismissed and finally the position and amplitude of the remaining quantized coefficients is encoded. The use of 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.

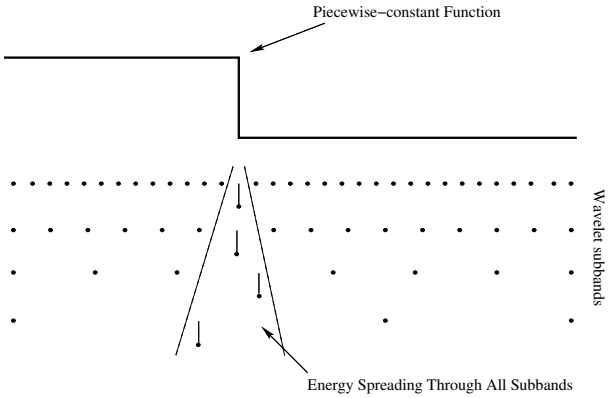

Fig. 2. Spreading of coefficients through the wavelet subbands of a 1D piecewise-constant signal representation.

In Fig. 2, one can see a scheme that illustrates the effect of a wavelet transform on a step. 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 generate wavelet coefficients are the discontinuities. Taking into account coding costs, the coding of the transformed signal implies coding the amplitude of each one of the coefficients as well as their position.

On the other hand, Fig. 3 shows a non-linear approach widely discussed in approximation theory [10], [7], [8] that intends a more efficient representation of piecewise signals in general. The approach is based on the assumption of 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. 3, 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 one of the intervals will not generate additional information to code, i.e. consider a Haar wavelet is used in each one of the intervals, 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.

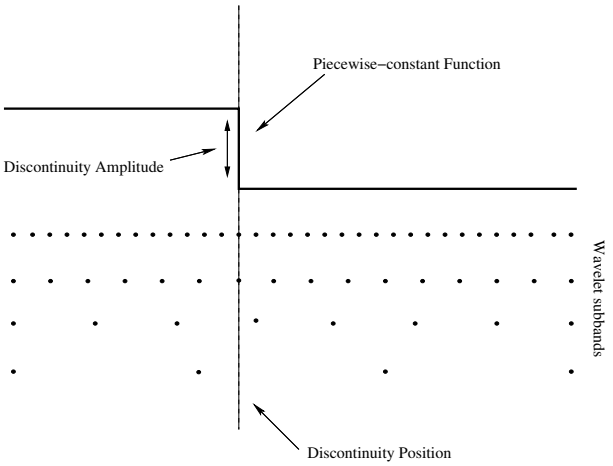

Fig. 3. There are no wavelet coefficients to code from a 1D piecewise-constant signal representation when using an oracle to code the discontinuities, i.e. position and amplitude.

Based on this intuition, *Prandoni* proved in [7] how can be characterized the behavior of oracle based coding of piecewise-polynomial signals and, more generally, *Cohen* extended it to the piecewise-smooth case in [10]. 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):

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

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

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

where A is a positive constant. Unlike in (1), distortion decreases as a power of $\sqrt{R}$ 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 [7], [8], and to understand the coding efficiency of different approximation approaches (e.g. [7], [11]). In the work of *Prandoni*, the R-D optimized coding of the polynomial parts of the signal is performed by means of dynamic programing and the use of local Legendre polynomial expansions. In this work, we consider the use of frame-adaptive lifting scheme proposed in [4] (see Sec. III). The length of wavelet transforms (i.e. the number of decomposition subbands) is adapted to cover smooth areas while avoiding wavelet kernels to cross edges. The approach we study is in some sense like the wavelet *Best Basis* in double trees for image coding studied by *Ramchandran* [12], [13]. Indeed, our purpose is to find a R-D adapted decomposition of the video signal. However, unlike in the double trees case, we are only concerned by the spatially local adaptation of the number of decomposition levels of the 1D temporal transform. In our work, no further decomposition in packets of the dyadic wavelet subbands is considered.

In the case of motion compensated lifted wavelet transform for video coding, this adaptivity can be implemented, as seen in Sec. III by means of inserting the so called *Intra* macroblocks as additional coding mode in the lifted coding scheme. As analyzed in the remaining of the paper, this introduces the necessary local spatio-temporal adaptivity to reduce the distortion of the compressed video and enhance its visual quality.

C. Examples: To join or not to join?

In the following, some simple examples illustrate the need for adapting the number of levels of a wavelet decomposition for an optimal R-D relation for coding purposes. Signal decompositions in terms of wavelet coefficients present an optimal R-D behavior if these are used to approximate polynomial signals when wavelets with enough vanishing moments are used. However, singularities between polynomial pieces generate many wavelet coefficients that are costly in terms of rate for coding applications. As exposed in this section, in the next points we show the coding improvements for some simple examples when the number of decomposition levels can be selected. Indeed, only one level of dyadic wavelet decomposition is considered. The comparison is performed by coding the spatio-temporal representation of two consecutive images coded as if they where a sequence. In all cases, a dyadic wavelet spatial decomposition is performed in all images by means of the Daubechies 9/7 filters [14]. We compare between a Haar transform in the temporal dimension (i.e. a one level Haar transform) and no temporal transformation. Uniform dead-zone quantization is applied to the wavelet coefficients and the coding cost is measured by means of their Shannon entropy.

Temporally aligned video pixels that, due to the effect of some motion, belong to different objects, have often no relation among them. If these are modeled as being a set of IID Gaussian random variables, then there would be no change in the coding efficiency with or without applying an orthonormal transform. However, this is not the case for the image sequences presented in here. They have a spatio-temporal structure and temporal changes follow a model that does not correspond to a set of IID Gaussian random variables. Although signals are often modeled by jointly Gaussian stochastic models, real images and video have a quite different behavior. In fact, deterministic piecewise-constant models [7] suit better our purposes.

As proved theoretically from a 1D point of view in [7], if a temporal pixel has an edge (important gray level change in the temporal dimension) the transform will spread the energy of this throughout the subbands and coding will be inefficient in terms of R-D. On the other hand, if a temporal pixel stays more or less unchanged, energy will be compacted and the pixel will be coded efficiently. For this experiment, the decision is made on an image level.

Fig. 4. Lenna (left) and Barbara (right) 256x256 pictures.

1) Coding Very Different Pictures: Consider the two images of Fig. 4 as if they were a sequence. Such situation can be seen as example of an extreme case for a scene cut in a sequence. These are very different pictures that barely offer any significant redundancy among them to be exploited. Moreover, taking the piecewise-constant model perspective, it seems clear that there are so many temporal edges that coding both pictures independently may be better than coding them jointly in terms of R-D. This is indeed the case for the *Lenna-Barbara* example. In Fig. 5 it can be appreciated

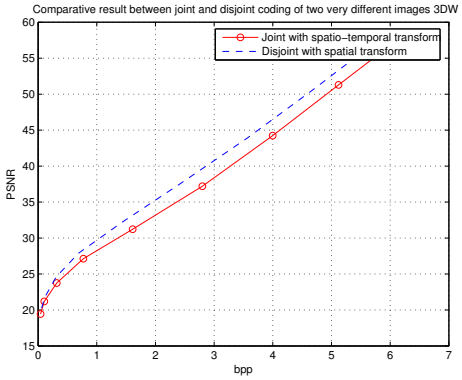

Fig. 5. Demonstration of an artificial scene cut: Coding the images Barbara and Lena jointly with a Haar transform or independently.

how coding without orthonormal Haar transform is more efficient in terms of R-D. The two upper curves correspond

to the case where spatial transformation is also applied (5 levels dyadic wavelet transform with Daubechies 9/7 filters). The pair of lower curves depict the behavior when only a temporal transformation is used. It can be easily seen how, in the case of having such different images, a join temporal implies a significant loss in coding efficiency. This behavior is observed in a large range of coding bit-rates. Nevertheless, at very low bit-rates (very high distortion) the difference becomes negligible due to high amount of quantization noise introduced.

2) Coding a Change of Scene: Let us take now a more common situation with regard to a video sequence: the table tennis sequence (Fig. 6). Two very similar frames can be efficiently decorrelated by means of a wavelet transform (let us forget for a while about the additional benefits of using in addition motion information).

Fig. 6. Table tennis sequence frames 129, 130 and 131.

This is underlined by the left graphic in Fig. 7 where results of joint coding of frames 129 and 130 by means of a Haar transform is shown. This is more R-D efficient when compared to the case where no temporal transformation is performed. In the right plot of Fig. 7, another particular example is analyzed. This time a change of scene is taking place (frames 130 to 131) and this gives rise to a similar situation as in the example of *Lenna* and *Barbara* (Sec. II-C.1). Again, as can be drawn from the curves, the joint coding of both pictures is not R-D efficient.

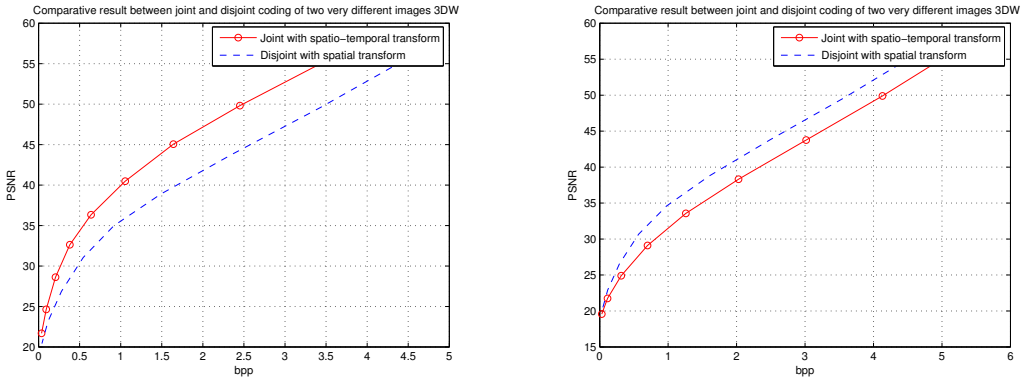

Fig. 7. R-D efficiency of one level of temporal wavelet decomposition for two different scene events from the sequence table tennis. Left: No change of scene (frames 129 and 130). Right: Change of scene (frames 130 and 131).

In order to be optimally adaptive in the temporal wavelet decomposition depth, the number of wavelet subband should be adaptively chosen locally in space. In this way, only when local unpredictable changes appear in a sequence of images, the wavelet decomposition used to jointly coding them can be optimally adapted for that particular scene region and still profit from a high number of wavelet decomposition levels in the rest of scene spatial locations. Of course, to find the best decomposition of the video signal, all possible spatial and temporal divisions of the spatio-temporal transform should be tested in order to retrieve the optimal (in terms of distortion) description of discontinuities, smooth areas and quantization for a given rate. This is a very costly approach that, finally, depending on the size of the data may be practically infeasible due to complexity. In the following sections an Intra-adaptive lifting scheme is studied as a solution to find a locally adaptive temporal wavelet decomposition of the video signal.

III. Motion-Compensated Lifted Wavelet Transforms for Video Coding

Our analysis and results of Sec. **??** are based on the multi-hypothesis, frame-adaptive motion compensated lifting scheme proposed in [4], [15]. 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 (Haar or 5/3) for a minimum distortion at a given rate. Nevertheless, it still forces a fixed number of multi-scale subbands in the MC temporal wavelet decomposition of the signal and limits temporal adaptation dealing with temporal discontinuities and motion misalignments as discussed in Sec. II. In the following we review the scheme presented in [4]

and [15], we discuss its implicit temporally adaptive properties and we comment on the inclusion of Intra macroblocks as an additional mode in the lifting scheme to allow further flexibility in the number of subbands considered in the temporal wavelet transform.

A. Motion-Compensated Lifted Wavelet

Most efforts of present video coding research for efficient compression are being done toward the promising combination of linear transforms and motion compensation. A scheme that has appeared to be successful exploiting temporal redundancy is based on motion compensated temporal wavelets [16], [17]. This has been made possible thanks to the use of the flexible lifting scheme, that allows the inclusion of non-linear, non-invertible, operations into its ladder structure such as quantization (e.g. integer to integer transforms) or motion compensation, being still the whole scheme invertible. In the following we briefly review the well known lifting schemes used to generate one decomposition level of the motion compensated Haar and 5/3-wavelet transforms.

1) Motion-Compensated Lifted Haar Wavelet: At this point we recall the basic structure of a MC lifted Haar step widely used for video coding (e.g. [2], [3]). Fig. 8 depicts the ladder scheme that carries out the Haar transform in the video signal along motion trajectories. As can be seen, the prediction/update steps of this structure are performed using the samples that motion vectors connect. [4], [15], uses for the update step the negative version of the motion vector retrieved in the prediction step.

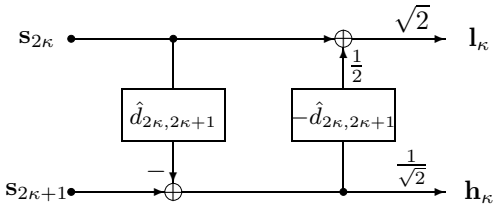

Fig. 8. Haar transform with motion-compensated lifting steps. Both steps, prediction (with motion vector -MV- $\hat{d}_{2\kappa,2\kappa+1}$) and update (with MV $-\hat{d}_{2\kappa,2\kappa+1}$), utilize block-based motion compensation. The update steps use the negative motion vectors of the corresponding prediction steps.

2) Motion-Compensated Lifted 5/3 Wavelet: At this point, the MC 5/3 wavelet lifting scheme is reviewed [2]. Akin to the Haar scheme, motion vectors are decided during the prediction stage and reused in its negative form for the update step. This scheme uses a multi-hypothesis MC prediction step that looks for the two most optimal vectors that achieve the most suitable linear combination for the prediction step. This scheme performs better than the Haar one, i.e. the associated wavelet has one additional vanishing moment.

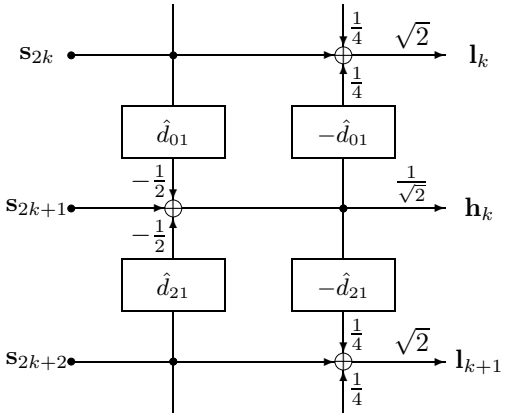

Fig. 9. Lifted 5/3 wavelet with motion compensation. Both steps, prediction (with MVs $\hat{d}_{01}$ and $\hat{d}_{21}$) and update (with MVs $-\hat{d}_{01}$ and $-\hat{d}_{21}$), utilize block-based motion compensation. The update steps use the negative motion vectors of the corresponding prediction steps.

B. Adaptive Motion-Compensated Lifted Wavelet Transforms

As previously seen, optimal R-D compression of piecewise-smooth signals requires temporal adaptivity in order to improve the R-D behavior that wavelets are capable to supply for them. In some sense, the multi-hypothesis, frame-adaptive variation on the lifting scheme proposed in [4], [15] gives the possibility to select the best prediction signal for a given lifting step without any constraint on the reference frame to be used. 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 levels for an optimal compression in a R-D sense. In the following, we briefly review the enhanced scheme proposed in [4], [15] and then we discuss how temporal adaptivity can be inserted in the lifting scheme by using Intra MBs.

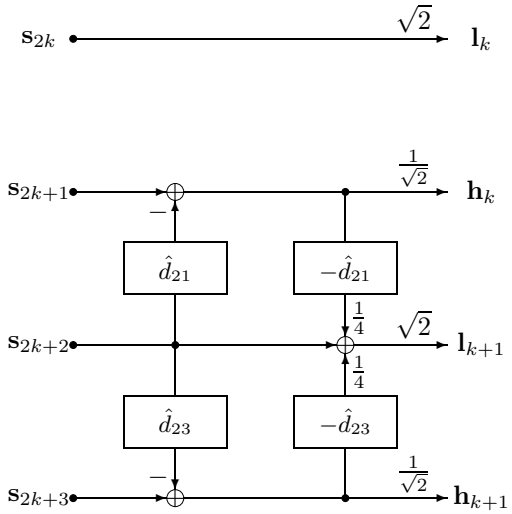

Fig. 10. Example of the first decomposition level of the Haar transform with frame-adaptive motion-compensated lifting steps. The frame $\mathbf{s}_{2k+2}$ is used to predict frame $\mathbf{s}_{2k+1}$

.

1) Frame-adaptive Motion-Compensated Lifting Scheme: Frame-adaptive MC Lifting schemes is a very flexible approach that allows a signal representation able to adapt, up to a certain degree, to the signal. This helps to overcome occlusion effects, changes of scene or to palliate the effects of deficient motion compensation at fine scales. Figs. 10 and 11 show the way frame adaptivity is implemented in the lifting scheme for the Haar and 5/3 wavelet cases. As can be seen, in this two particular examples, the fixed and classical structure of the lifting scheme is broken such that for every instance of this, every even frame in the GOP can be used to select the best prediction signals. Of course, the update step is realized in consequence and following the complementary scheme to the one determined in the prediction step.

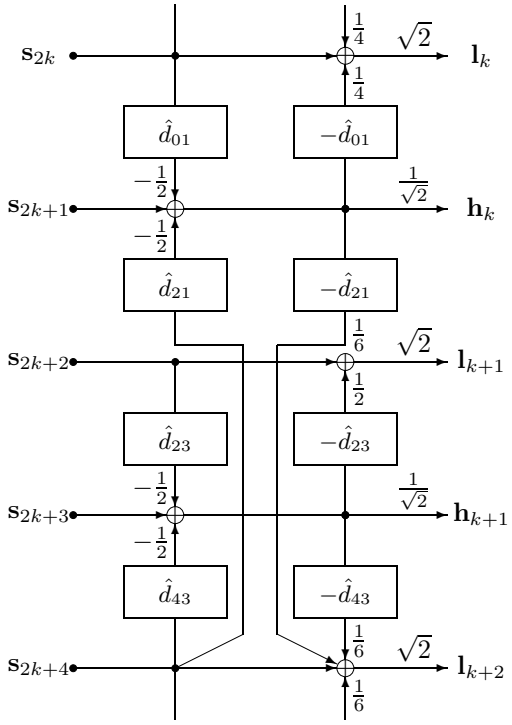

Fig. 11. Example of the first decomposition level of the 5/3 transform with frame-adaptive motion-compensated lifting steps. The frames $\mathbf{s}_{2k}$ and $\mathbf{s}_{2k+4}$ are used to predict frame $\mathbf{s}_{2k+1}$

2) Intra-Adaptive Motion-Compensated Lifting Scheme: Nevertheless, the "frame-adaptive" approach can be improved. Indeed, the frame-adaptive scheme does not variate the default number of wavelet decomposition levels nor

considers alternative methods for MC at a particular location. In order to tackle this problem, people have introduced in the MC lifted wavelet based video coding ([18], [5], [19]) the so called intra refresh in the classic predictive video coding approach [6]. For MC lifted wavelet based video coding corresponds to the adaptive insertion of void lifting steps. The void lifting steps do not perform further filtering on the signal and serve to implement the necessary breakpoints in the wavelet decomposition for an efficient R-D signal approximation (as discussed in Sec. II). A visual example of such a special "lifting" mode can be seen in Fig. 12. By properly selecting such a lifting mode, the so desired temporal wavelet decomposition with local adaptation of the number of decomposition levels can be obtained.

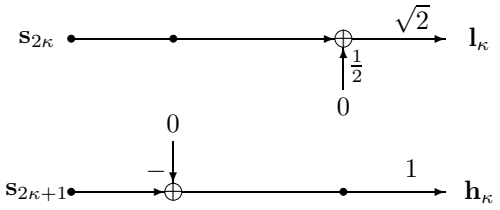

Fig. 12. Broken lifting step, neither prediction nor update is used. The use of the ladder scheme would reduce R-D performance due to the piecewise-smooth signal. Both, prediction and update, are inhibited on a macroblock level. Notice the change on the scaling factors to control the noise in the quantization stage.

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

IV. Results

In this section, we present a review of the effect of introducing local spatio-temporal adaptivity into the lifting scheme used for motion compensated temporal filtering. This adaptivity is introduced in practice by the insertion of Intra macroblocks in the lifting scheme, as described in Sec. III-B. We evaluate the benefits of using this temporal adaptivity and compare the improvements in terms of R-D supplied by the use of Intra MBs in the lifting scheme. The tests are performed using four different test sequences in order to supply different signal characteristics to the coder. These sequences are in QCIF format (176x144 at 30 Hz) and they are identified as *cnn*, *football*, *foreman* and *table tennis*. Furthermore, results are presented as well concerning when Intra Macro-Blocks are used and selected by the coding algorithm.

A. The Coding Scheme

The coding scheme divides the frames into macroblocks of size 16x16. 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 R-D Lagrangian cost of the high band that issues from the ladder scheme. Haar-type, 5/3-type and void-type are used as candidates to encode each macroblock. Thanks to the flexibility of the frame adaptive scheme, one or two reference frames may be used by the algorithm to code a macroblock. For simplicity, and since the goal of this work is the experimental verification, there are no further subdivisions of the macroblocks for motion compensation purposes, hence motion vectors are associated to blocks (macroblocks) of size 16x16. These are obtained by block-based rate-constrained motion estimation jointly optimized with the lifting mode selection. The motion information is estimated in each decomposition level depending on the results of the lower level and an accuracy of half-pel motion compensation is used to align wavelet transform with motion. Half-pel positions are bi-linearly interpolated. All subband macroblocks generated by the temporal wavelet transform are encoded in a intra-frame fashion. For this purpose, since only the performance of temporal adaptation is analyzed, a 8x8 DCT with run-length coding is used for simplicity. All intra-frame coded subbands are quantized with the same quantization step-size. Finally, the whole generated information is encoded with Huffman codes, in particular, before the entropy coding, motion vectors are predicted from spatial neighbors and only the differences are encoded. GOPs of size 32 are used in our experiments (up to five decomposition levels), shorter length wavelet transforms are provided by the temporal adaptation introduced by Intra MBs.

B. Global R-D Performance of Local Temporal Wavelet Transform Length Adaptation

The usage of shorter instances of the lifted wavelet scheme than the maximum allowed GOP length of 32 commonly contributes locally to the areas where object trajectories are shorter than 32 frames. Hence, the benefit is going to be of local nature when the particular characteristics of the sequence require it. In Fig. 13 we show how the effect of using this additional coding mode in the MC lifting scheme, does introduce a moderate overall gain to the whole R-D performance of a coded sequence. Average improvements range from 0.2 to 0.5 dBs in middle and low motion sequences

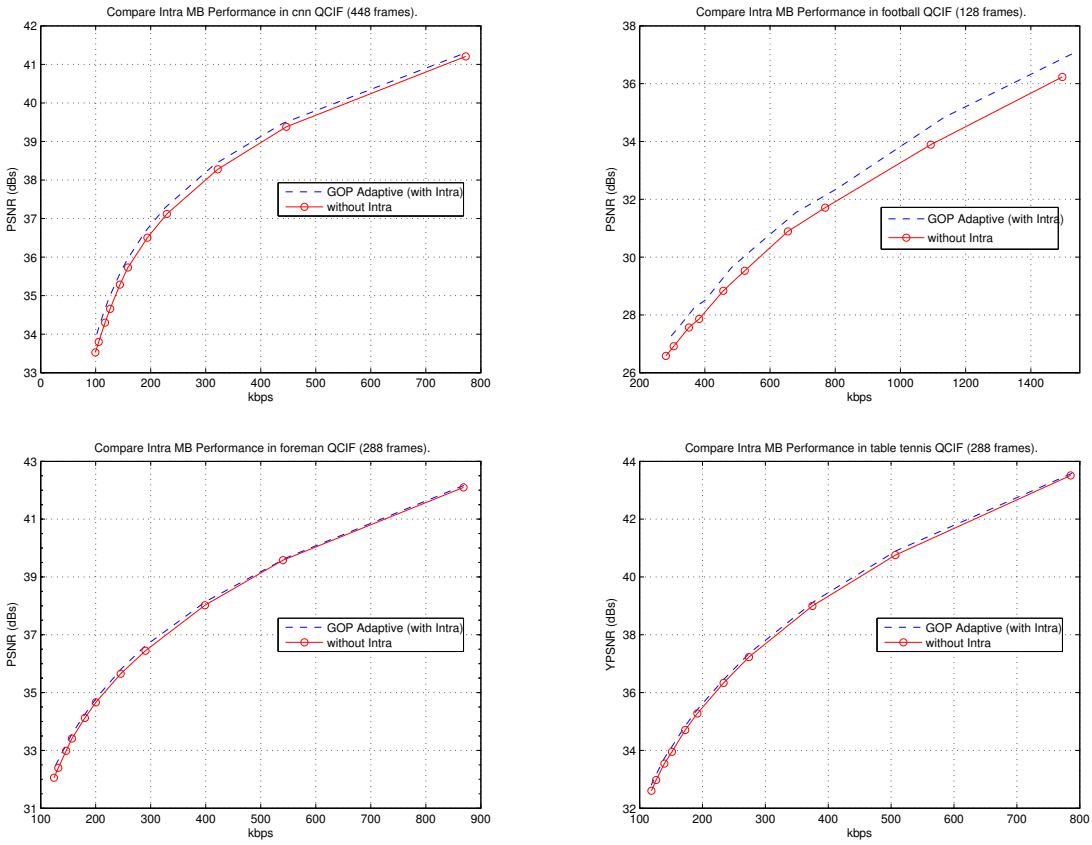

Fig. 13. Global sequence R-D comparison of the improvement due to GOP Adaptivity.

with some change of scenes and local fast motion. However, in highly moving sequences, like football, the improvement of introducing the Intra adaptation through the use of Intra MBs is of higher relevance as new information is efficiently coded.

C. R-D Performance of Local Temporal Wavelet Transform Length Adaptation Through Time

In Fig. 14, the PSNR of the adaptive wavelet decomposition length is depicted over time. For *cnn*, *table tennis* or even *foreman*, a very strong panning is present in a particular moment of the sequence. Due to the strong motion appearing in the sequence *football* a significant overall improvement of every frame can be observed in the upper right chart of Fig. 14.

When using exclusively one reference frame for the prediction/update steps, the use of the Intra-adaptive scheme contributes the achieve a slighly better global R-D improvement with respect to the non Intra-adaptive

In case only one reference frame is used for the prediction/update steps, Intra-adaptive R-D improvement is slightly more significant than when two reference frames are allowed. This is illustrated in Fig. 15 by means of the coding R-D performance of the sequence *cnn*.

D. R-D Performance and Length Adaptation on a Particular GOP

Let us take a GOP where relevant changes appear in the sequence signal and the MC lifting scheme can not efficiently represent them. R-D gains can be as high as 1.0 dB: see the upper left chart in Fig. 16 for sequence *cnn*, the upper right chart for *football* or the lower right for *table tennis*.

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 more significant, as depicted by Fig. 17.

E. Intra Macro-Blocks and Length Adaptation

Figs. 18 and 19 show the quantitative usage of Intra Macroblocks to split lifting steps in different contexts. In Fig. 18 can be observed the proportions of different kinds of prediction modes in the lifting steps. Each bar represents the totality of Macroblocks used in each GOP of 32 frames (in this count, we do not take into account the fixed

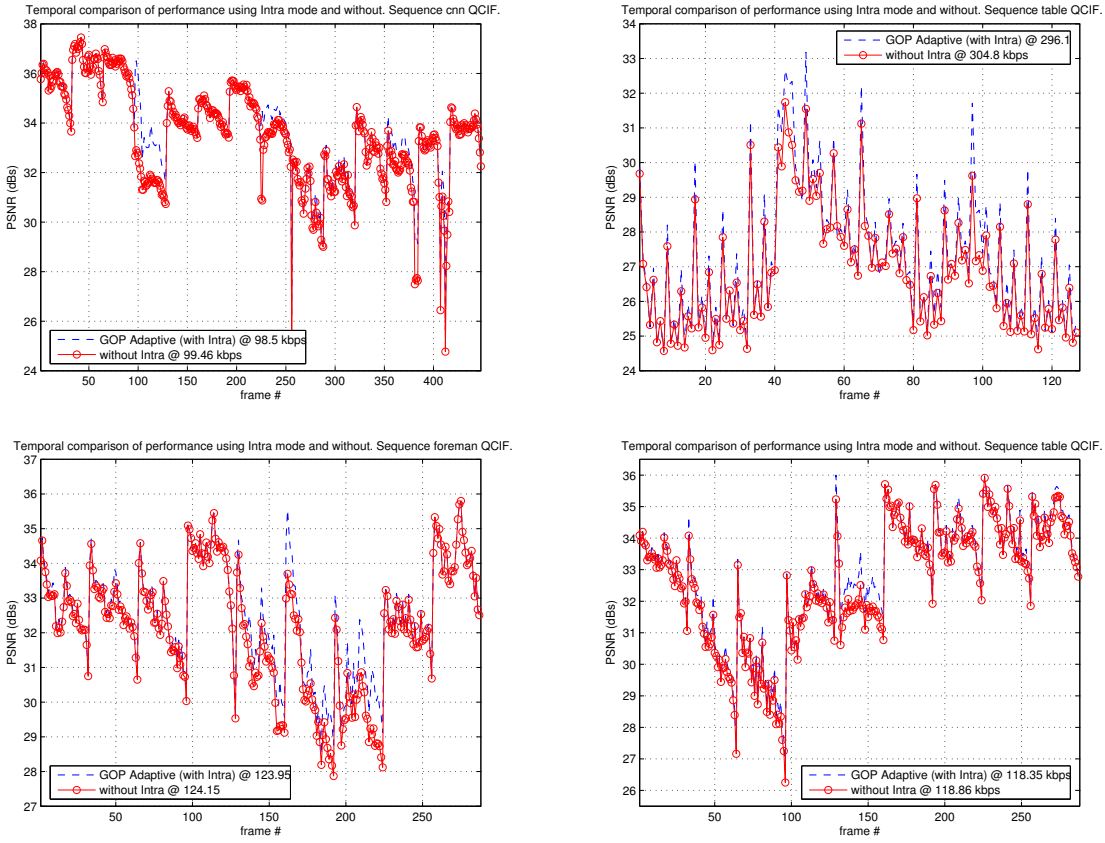

Fig. 14. Temporal evolution of the PSNR and comparison of the improvement due to GOP Adaptivity.

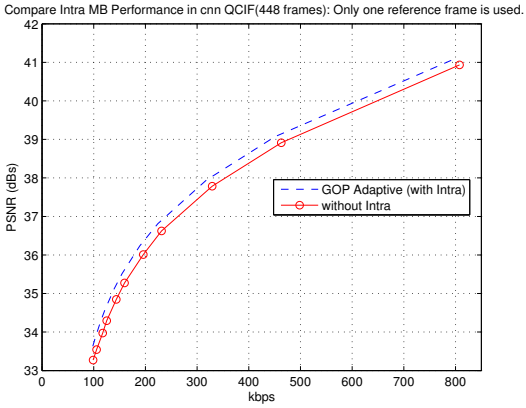

Fig. 15. Overall sequence R-D comparison of the improvement due to GOP Adaptivity using just one reference frame.

number of Intra coded MBs always present in a GOP and commonly used to code the lowest frequency band). In red we observe the percentage of 5/3 wavelet multi-reference prediction modes. In blue appear the usage of Haar wavelet single reference prediction mode. Finally in light green appear the proportion of Intra coded MBs used to spatially locally break particular lifting steps that are not interesting from a R-D point of view. As expected, the use of Intra MBs appears coherent with the temporal scene changes or very fast moving sequence periods.

Fig. 19 shows, on average, at which decomposition level Intra MBs are used more often. The index below the column represents the corresponding decomposition level of the wavelet scale according to 2^{-j} for $j \in \{1, 2, 3, 4, 5\}$. Indeed, for low frequency subbands, lifting steps have to be split for the case where middle length wavelet transforms are required (16 or 8) as well as when the smallest of the number of decomposition levels is required (0 or 1). Thus, Intra MBs are frequently allocated in big scale subbands.

However, since *football* is a highly moving sequence that requires a higher re-injection of the information than for

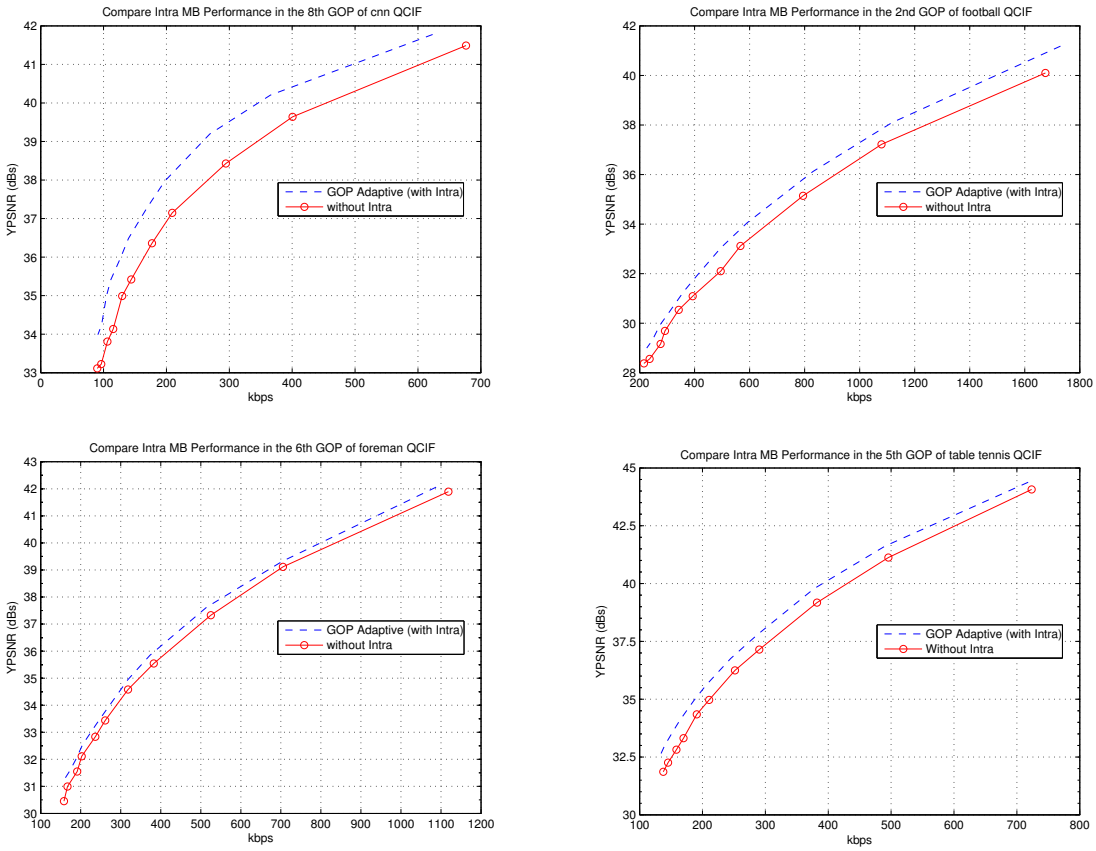

Fig. 16. Particular GOP R-D comparison of the improvement due to GOP Adaptivity.

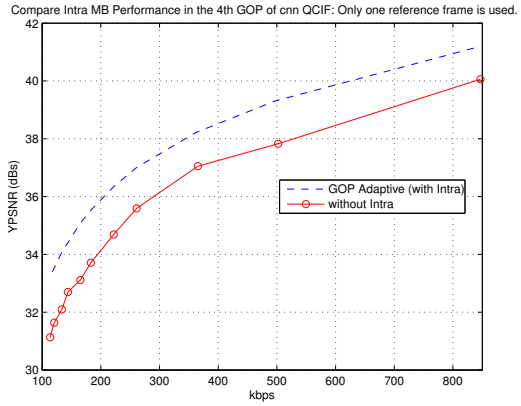

Fig. 17. 4th GOP R-D comparison of the improvement due to GOP Adaptivity using just one reference frame.

other sequences and, thus, it presents a very different kind of motion with respect to the other test sequences, we show in Fig. 19 the same statistic presented in Fig. 19 but without taking into account *football*. This shows that for slower sequences, Intra MBs are concentrated in lower subbands. Hence, longer wavelet transforms are used.

F. Visual Comparison and Length Adaptation

Finally, a visual comparison between the normally coded sequences with the scheme proposed in [4], [15] and the wavelet decomposition depth adaptive is presented in Fig. 20. It is mainly relevant, in addition to improvements in the numerical figures shown in the previous results, the visual noise reduction evidenced in the pictures. Noise reduction is present in all the four sequences. This noise reduction is due to the fact that the quantization error introduced at big scale wavelet subbands is not spread all over the GOP when the wavelet decomposition is allowed to be split in less deeper decompositions. Indeed, Intra-adaptivity allows for a higher energy compaction in the signal approximation.

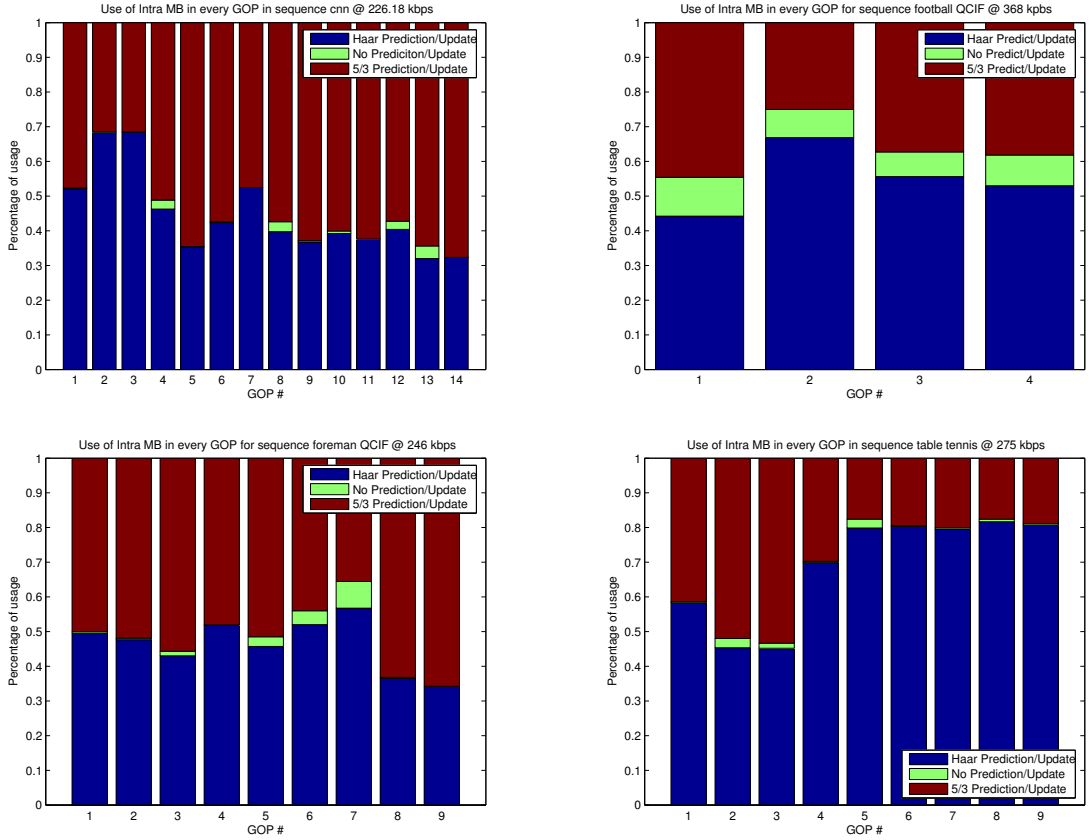

Fig. 18. Usage of Intra MBs through the different GOPs, GOP length Adaptivity is mainly present in highly moving scenes where MC performs bad and in scene shots.

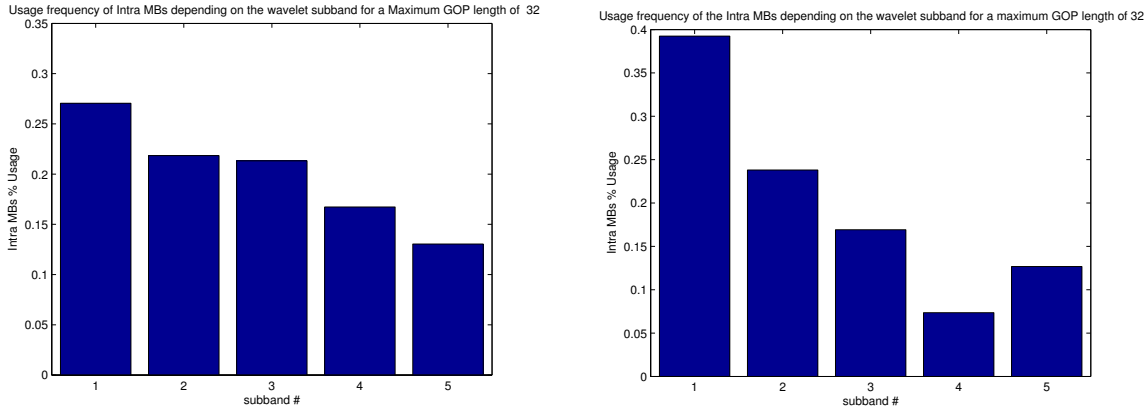

Fig. 19. Average frequency of usage of Intra MBs depending on the temporal wavelet subband for a maximum GOP length of 32. The lower the number of the wavelet subband, the bigger the scale of details that it represents. The graphic shows the descending tendency in the splitting frequency of the temporal lifting scheme. (Left) All four test sequences are considered to generate the statistic. (Right) This statistic does not contain the sequence *football*

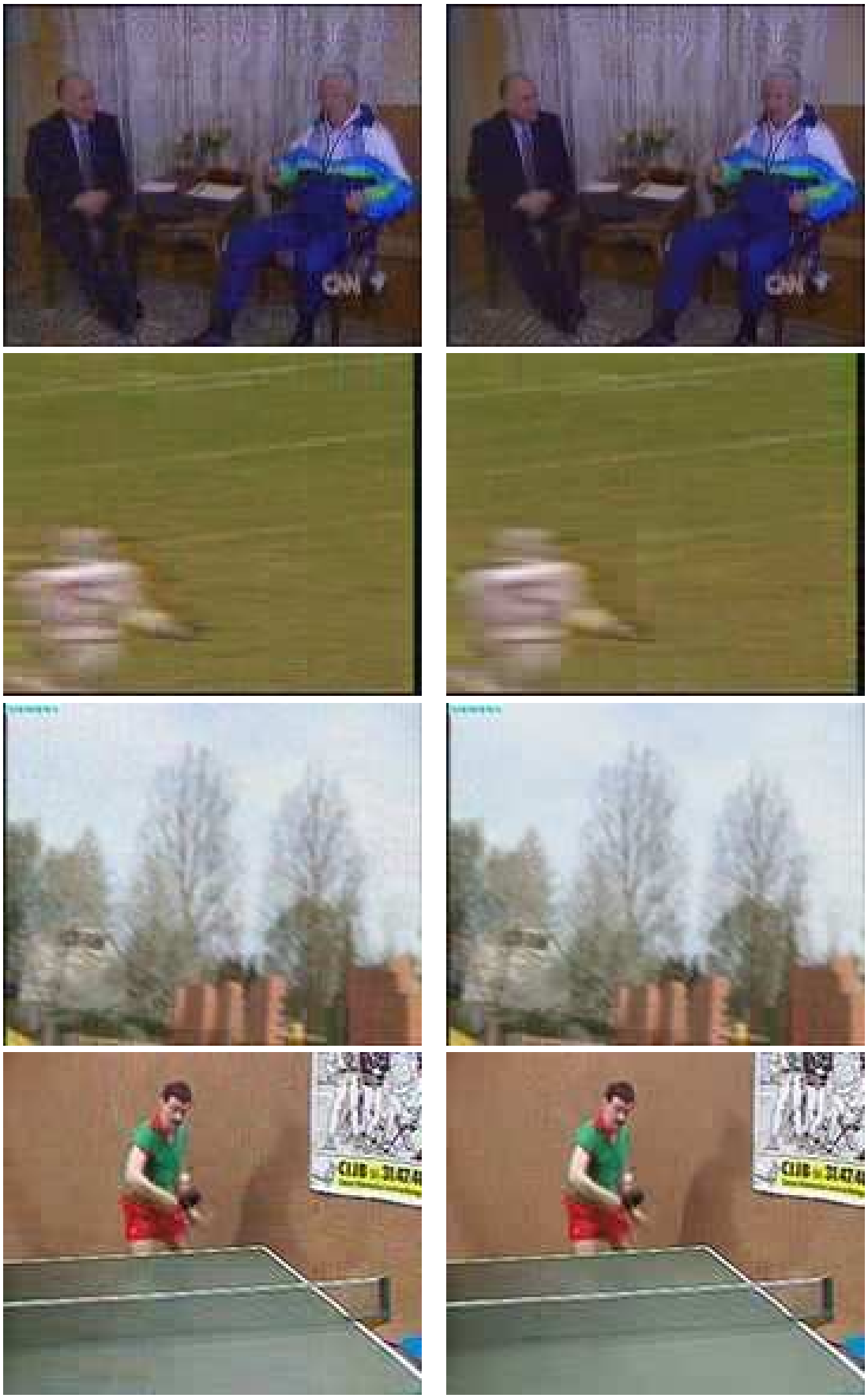

Fig. 20. Visual quality improvement for selected frame in the test sequences. Left: no GOP Adaptivity. Right: With GOP Adaptivity. Rows from up down: cnn, football, foreman, table tennis. The respective bit rates are the following (for each sequence first is indicated the rate without Intra-adaptivity): cnn 126.21 kbps and 124.637 kbps, football: 1093 kbps and 1007 kbps, foreman: both at 246 kbps, table: 151.38 kbps and 150.92 kbps

Hence, signal structures are represented by a fewer number of wavelet coefficients which reduces the introduction of quantization noise in relevant signal components. Moreover, the amount of rate spared thanks to the new lifting mode can be invested in other critical macroblocks. In some cases the noise reduction appears as a relevant increase of the reconstruction detail of the objects appearing in the sequence. Notice, for instance, the sharper and more clear appearance of *Yeltsin* in the *cnn* sequence, and the lower blocking effect in the player t-shirt of *football* sequence.

V. Conclusions

This article discusses Intra-adaptive motion-compensated three-dimensional transform coding for video signal. Although the main major signal division for its processing is the GOP of K pictures, we do not consider a fixed number of decomposition levels to obtain a fixed number of decomposition subbands. Our approach is such that GOPs of K pictures are adaptively broken in smaller ones at a macroblock level. Like this, the number of wavelet decomposition subbands in the temporal transform is adapted in space and time. Local signal breakpoints are coded independently by using Intra macroblocks and wavelets kernels are reserved for those areas of the signal where prediction can be made efficiently with the MC lifting scheme. In this work we underline the theoretical reasoning and discuss local spatio-temporal adaptation of MCTF schemes for video coding. We present a detailed analysis of the practical usage of Intra MBs in the Motion Compensated lifting scheme proposed in [4], [15] for video coding. The benefits of using adaptive length temporal wavelet decompositions are presented. The approach improves R-D performance as well as visual quality. Intra macroblocks are selected by R-D optimization. Sudden changes of scene trigger this mode and switch off term temporal filtering. We discuss the dependency on the video signal, the characteristics of adaptation and the usage of this Intra mode overall improvement of RD performance.

References

[1] M. Flierl and B. Girod, "Video coding with motion-compensated lifted wavelet transforms," *EURASIP Journal on Image Communication*, vol. 19, no. 7, pp. 561–575, August 2004, special Issue on Subband/Wavelet Interframe Video Coding.

[2] A. Secker and D. Taubman, "Motion-compensated highly scalable video compression using an adaptive 3d wavelet transform based on lifting," in *IEEE International Confernce on Image Processing (ICIP)*, vol. 2, Thessaloniki, Greece, 2001, pp. 1029–1032.

[3] 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*, vol. 3, May 2001, pp. 1793 – 1796.

[4] M. Flierl, "Video coding with lifted wavelet transforms and frame-adaptive motion compensation," in *EURASIP VLBV*, Madrid, September 2003.

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

[6] ITU-T Recommendation H.264 - ISO/IEC 14496-10 AVC: *Advanced Video Coding for Generic Audiovisual Services*, ITU-T and ISO/IEC JTC1, 2003.

[7] P. Prandoni and M. Vetterli, "Approximation and compression of piecewise smooth functions," *Phil. Trans. R. Soc. Lond.*, 1999.

[8] M. Vetterli, "Wavelets, approximation and compression," *IEEE Signal Processing Magazine*, 2001.

[9] 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.

[10] A. Cohen, W. Dahmen, I. Daubechies, and R. DeVore, "Tree approximation and optimal encoding," *Applied and Computational Harmonic Analysis*, vol. 11, no. 2, pp. 192–226, 2001.

[11] M. Do, P. Dragotti, R. Shukla, and M. Vetterli, "On the compression of two-dimensional piecewise smooth functions," in *IEEE International Conference on Image Processing (ICIP)*, Thessaloniki, Greece, October 2001.

[12] K. Ramchandran, Z. Xiong, K. Asai, and M. Vetterli, "Adaptive transforms for image coding using spatially varying wavelet packets," *IEEE Trans. on Image Processing*, vol. 5, no. 7, pp. 1197–1204, July 1996.

[13] K. Ramchandran and M. Vetterli, "Best wavelet packet bases in a rate-distortion sense," *IEEE Trans. on Image Processing*, vol. 2, no. 2, pp. 160–165, April 1993.

[14] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Images coding using the wavelet transform," *Ieee Trans. Image Proc.*, pp. 205–220, April 1992.

[15] 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, January 2004.

[16] S.-J. Choi and J. Woods, "Motion-compensated 3-d subband coding of video," *IEEE Transactions on Image Processing*, vol. 8, no. 2, pp. 155 – 167, February 1999.

[17] J.-R. Ohm, "Three-dimensional subband coding with motion compensation," *IEEE Transactions on Image Processing*, vol. 3, no. 5, pp. 559 – 571, 1994.

[18] C. Mayer, "Motion compensated in-band prediction for wavelet-based spatially scalable video coding," in *IEEE International Conference on Acoustics, Speech & Signal Processing (ICASSP)*, Hong Kong (cancelled), April 2003.

[19] D. S. Turaga and M. van der Schaar, "Content-adaptive filtering in the umctf framework," in *IEEE International Conference on Acoustics, Speech & Signal Processing (ICASSP)*, Hong Kong (cancelled), April 2003.