Video Coding with Lifted Wavelet Transforms and Frame-Adaptive Motion Compensation

Markus Flierl

Signal Processing Institute, Swiss Federal Institute of Technology
1015 Lausanne, Switzerland
markus.flierl@epfl.ch

Abstract. This paper investigates video coding with wavelet transforms applied in the temporal direction of a video sequence. The wavelets are implemented with the lifting scheme in order to permit motion compensation between successive pictures. We generalize the coding scheme and permit motion compensation from any even picture in the GOP by maintaining the invertibility of the inter-frame transform. We show experimentally, that frame-adaptive motion compensation improves the compression efficiency of the Haar and 5/3 wavelet.

1 Introduction

Applying a linear transform in temporal direction of a video sequence may not be very efficient if significant motion is prevalent. Motion compensation between two frames is necessary to deal with the motion in a sequence. Consequently, a combination of linear transform and motion compensation seems promising for efficient compression. For wavelet transforms, the so called *Lifting Scheme* [1] can be used to construct the kernels. A two-channel decomposition can be achieved with a sequence of prediction and update steps that form a ladder structure. The advantage is that this lifting structure is able to map integers to integers without requiring invertible lifting steps. Further, motion compensation can be incorporated into the prediction and update steps as proposed in [2]. The fact that the lifting structure is invertible without requiring invertible lifting steps makes this approach feasible. We cannot count on invertible lifting steps as, in general, motion compensation is not invertible.

Today's video coding schemes are predictive schemes that utilize motion-compensated prediction. In such schemes, the current frame is predicted by one motion-compensated reference frame. This concept can be extended to multi-frame motion compensation which permits more than one reference frame for prediction [3]. The advantage is that multiple reference frames enhance the compression efficiency of predictive video coding schemes [4].

The motion-compensated lifting scheme in [2] assumes a fix reference frame structure. We extend this structure and permit frame-adaptive motion compensation within a group of pictures (GOP). The extension is such that it does not affect the invertibility of the inter-frame transform. In Section 2, we discuss

frame-adaptive motion compensation for the Haar wavelet and provide experimental results for a varying number of reference frames. Section 3 extends the discussion to the bi-orthogonal 5/3 wavelet. In addition, a comparison to the performance of the Haar wavelet is given.

2 Haar Wavelet and Frame-Adaptive Motion Compensation

In this section, the video coding scheme is based on the motion-compensated lifted Haar wavelet as investigated in [5]. We process the video sequence in groups of $K = 32$ pictures. First, we decompose each GOP in temporal direction with the motion-compensated lifted Haar wavelet. The temporal transform provides $K = 32$ output pictures. Second, these K output pictures are intra-frame encoded. For simplicity, we utilize a 8×8 DCT with run-length coding.

2.1 Motion Compensation and Lifting Scheme

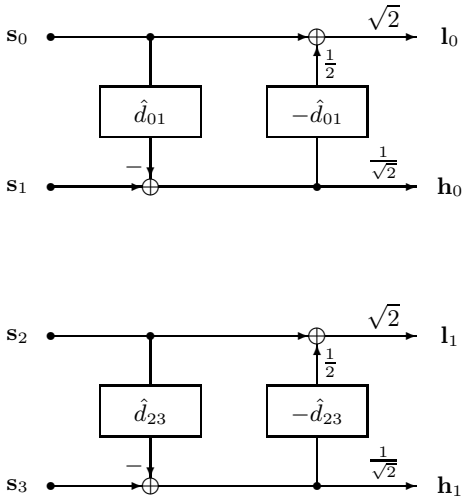

Fig. 1. First decomposition level of the Haar transform with motion-compensated lifting steps. A fix reference frame structure is used.

Fig. 1 explains the Haar transform with motion-compensated lifting steps in more detail. The even frames of the video sequence $\mathbf{s}_{2\kappa}$ are displaced by the estimated value $\hat{d}_{2\kappa,2\kappa+1}$ to predict its odd frames $\mathbf{s}_{2\kappa+1}$. The prediction step is followed by an update step with the displacement $-\hat{d}_{2\kappa,2\kappa+1}$. We use a block-size of 16×16 and half-pel accurate motion compensation with bi-linear interpolation in the prediction step and select the motion vectors such that they minimize a

cost function based on the energy in the high-band $\mathbf{h}_\kappa$. In general, the block-motion field is not invertible but we still utilize the negative motion vectors for the update step. If the motion field is invertible, this motion-compensated lifting scheme permits a linear transform along the motion trajectories in a video sequence. Additional scaling factors in low- and high-bands are necessary to normalize the transform.

2.2 Frame-Adaptive Motion Compensation

For frame-adaptive motion compensation, we go one step further and permit at most M even frames $\mathbf{s}_{2\kappa}$ to be reference for predicting each odd frames. In the prediction step, we select for each 16×16 block one motion vector and one picture reference parameter. The picture reference parameter addresses one of the M even frames in the GOP and the update step modifies this selected frame. Both motion vector and picture reference parameter are transmitted to the decoder. As we use only even frames for reference, the inter-frame transform is still invertible.

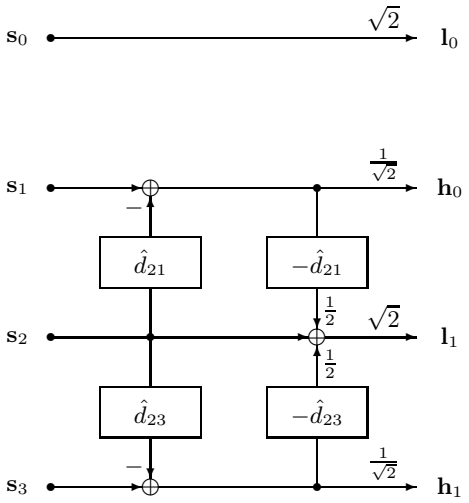

Fig. 2. Example of the first decomposition level of the Haar transform with frame-adaptive motion-compensated lifting steps. The frame $\mathbf{s}_2$ is used to predict frame $\mathbf{s}_1$.

Fig. 2 depicts the example where frame $\mathbf{s}_2$ is used to predict frame $\mathbf{s}_1$. If parts of an object in frame $\mathbf{s}_1$ are covered in frame $\mathbf{s}_0$ but not in frame $\mathbf{s}_2$, the selection of the later will avoid the occlusion problem and, therefore, will be more efficient. Note that for each block, the reference frame is chosen individually.

2.3 Experimental Results

Experimental results are obtained by using the following rate-distortion techniques: Block-based rate-constrained motion estimation is used to minimize the

Lagrangian costs of the blocks in the high-bands. The costs are determined by the energy of the block in the high-band and an additive bit-rate term that is weighted by the Lagrangian multiplier λ. The bit-rate term is the sum of the lengths of the codewords that are used to signal motion vector and picture reference parameter. The quantizer step-size Q is related to the Lagrangian multiplier λ such that $\lambda = 0.2Q^2$. Employing the Haar wavelet and setting the motion vectors to zero, the dyadic decomposition will be an orthonormal transform. Therefore, we select the same quantizer step-size for all K intra-frame encoder. The motion information that is required for the motion-compensated wavelet transform is estimated in each decomposition level depending on the results of the lower level.

For the experiments, we subdivide the test sequences *Foreman* and *Mobile & Calendar*, each with 288 frames, into groups of $K = 32$ pictures. We decompose the GOPs independent of each other and the Haar kernel causes no boundary problems.

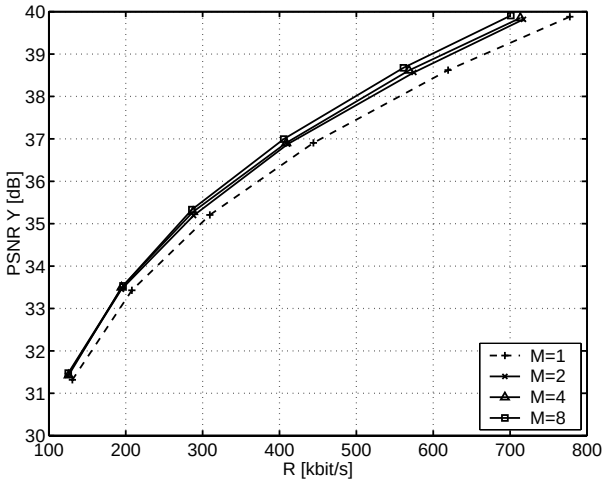

Fig. 3. Luminance PSNR vs. total bit-rate for the QCIF sequence *Foreman* at 30 fps. A dyadic decomposition is used to encode groups of $K = 32$ pictures with the frame-adaptive Haar kernel. Results for the non-adaptive Haar kernel ($M = 1$) are given for reference.

Figs. 3 and 4 show the luminance PSNR over the total bit-rate for the sequences *Foreman* and *Mobile & Calendar*, respectively. Groups of 32 pictures are encoded with the Haar kernel. We utilize sets of reference frames of size $M = 2$, 4, and 8 to capture the performance of the frame-adaptive motion-compensated transform as depicted in Fig. 2. For reference, the performance of the scheme with fix reference frames, as depicted in Fig. 1, is given ($M = 1$). We observe for both sequences, that the compression efficiency of the coding scheme improves by

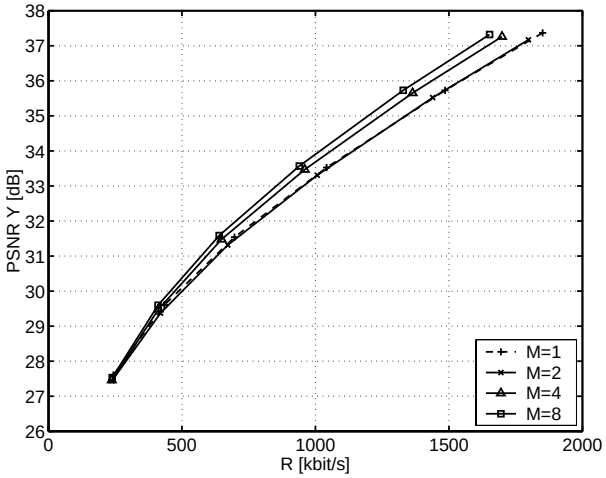

Fig. 4. Luminance PSNR vs. total bit-rate for the QCIF sequence *Mobile & Calendar* at 30 fps. A dyadic decomposition is used to encode groups of $K = 32$ pictures with the frame-adaptive Haar kernel. Results for the non-adaptive Haar kernel ($M = 1$) are given for reference.

increasing the number of possible reference frames M. For this test sequences, gains up to 0.8 dB can be observed at high bit-rates. At lower bit-rates, less reference frames are sufficient for a competitive rate-distortion performance.

3 5/3 Wavelet and Frame-Adaptive Motion Compensation

The Haar wavelet is a short filter and provides limited coding gain. We expect better coding efficiency with longer wavelet kernels. Therefore, we replace the Haar kernel with the 5/3 kernel and leave the remaining components of our coding scheme unchanged. We decompose the GOPs of size 32 independent of each other and use a cyclic extension to solve the boundary problem. When compared to symmetric extension, the cyclic extension is slightly beneficial for the investigated test sequences.

3.1 Motion Compensation and Lifting Scheme

Fig. 5 explains the 5/3 transform with motion-compensated lifting steps in more detail. For the 5/3 kernel, the odd frames are predicted by a linear combination of two displaced neighboring even frames. Again, we use a block-size of 16×16 and half-pel accurate motion compensation with bi-linear interpolation in the prediction steps and choose the motion vectors $\hat{d}_{2\kappa,2\kappa+1}$ and $\hat{d}_{2\kappa+2,2\kappa+1}$ such that they minimize a cost function based on the energy in the high-band $\mathbf{h}_\kappa$.

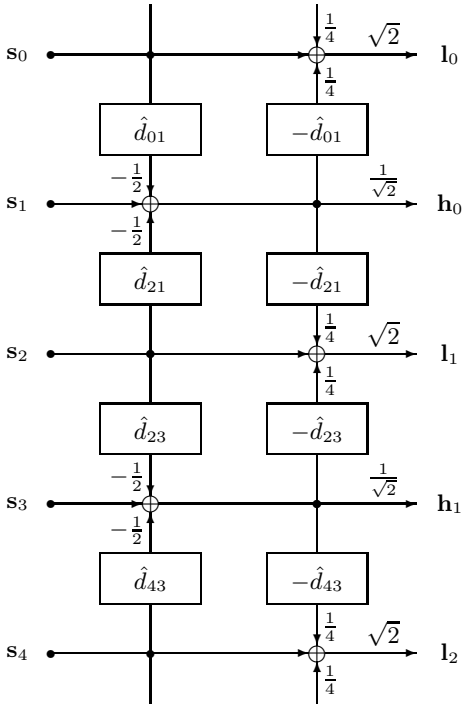

Fig. 5. First decomposition level of the 5/3 transform with motion-compensated lifting steps. A fix reference frame structure is used.

Similar to the Haar transform, the update steps use the negative motion vectors of the corresponding prediction steps. In general, the block-motion field is not invertible but we still utilize the negative motion vectors for the update step. Additional scaling factors in low- and high-bands are used.

3.2 Frame-Adaptive Motion Compensation

For frame-adaptive motion compensation, we permit at most M even frames $\mathbf{s}_{2\kappa}$ to be reference for predicting an odd frame. In the case of the 5/3 kernel, odd frames are predicted by a linear combination of two displaced frames that are selected from the set of M even frames. Each odd frame has its individual set of up to M reference frames. In the prediction step, we select for each 16×16 block two motion vectors and two picture reference parameters. The picture reference parameter address two of the M even frames in the GOP and the update step modifies the selected frames. All motion vectors and picture reference parameters are transmitted to the decoder. As we use only even frames for reference, the inter-frame transform based on the 5/3 kernel is still invertible.

Fig. 6 depicts the example where frames $\mathbf{s}_0$ and $\mathbf{s}_4$ are used to predict frame $\mathbf{s}_1$. This frame-adaptive scheme is also an approach to the occlusion problem.

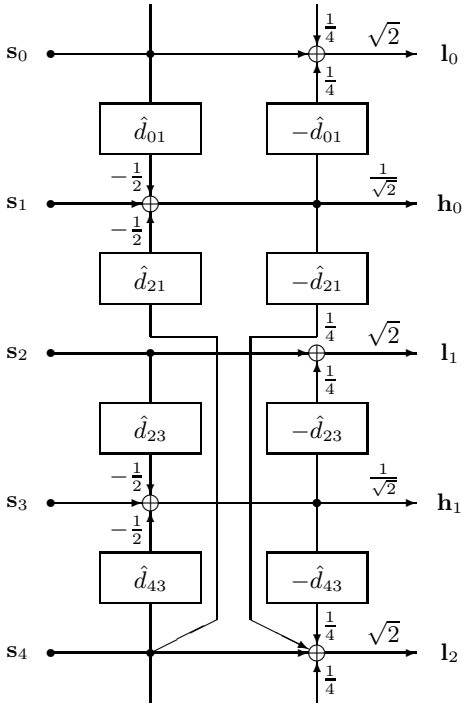

Fig. 6. Example of the first decomposition level of the 5/3 transform with frame-adaptive motion-compensated lifting steps. The frames $\mathbf{s}_0$ and $\mathbf{s}_4$ are used to predict frame $\mathbf{s}_1$.

Moreover, due to averaging of two prediction signals, noisy signal components can be suppressed efficiently. Note, that the encoder chooses the pair of reference frames individually for each block.

3.3 Experimental Results

To obtain results with the frame-adaptive 5/3 kernel, we employ the same rate-distortion techniques as outlined for the Haar kernel. Block-based rate-constrained motion estimation is also used to minimize the Lagrangian costs of the blocks in the high-bands. Here, the pairs of displacement parameters are estimated by an iterative algorithm such that the Lagrangian costs are minimized. The bit-rate term in the cost function is the sum of the lengths of the codewords that are used to signal two motion vectors and two picture reference parameters. That is, with the 5/3 kernel, we have macroblocks that have more side information when compared to macroblocks of the Haar kernel. But in an efficient rate-distortion framework, a Haar-type macroblock should also be considered for encoding. Therefore, we permit both Haar-type and 5/3-type to encode each macroblock. The encoder chooses for each macroblock the best type

in the rate-distortion sense. Further, the 5/3 wavelet is not orthonormal, even if the motion vectors are zero. For simplicity, we keep the same quantizer step-size for all K intra-frame encoder. Again, the motion information that is required for the motion-compensated wavelet transform is estimated in each decomposition level depending on the results of the lower level.

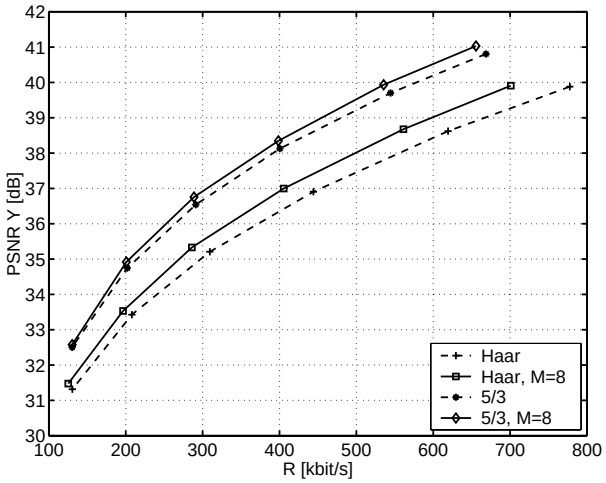

Fig. 7. Luminance PSNR vs. total bit-rate for the QCIF sequence *Foreman* at 30 fps. A dyadic decomposition is used to encode groups of $K = 32$ pictures with the frame-adaptive 5/3 kernel. Results for the non-adaptive 5/3 kernel as well as the Haar kernel are given for reference.

Figs. 7 and 8 show the luminance PSNR over the total bit-rate for the sequences *Foreman* and *Mobile & Calendar*, respectively. We subdivide the sequences, each with 288 frames, into groups of $K = 32$ pictures and encode them with the 5/3 kernel. We utilize a set of $M = 8$ reference frames to capture the performance of the frame-adaptive motion-compensated transform as depicted in Fig. 6. For reference, the performance of the scheme with fix reference frames, as depicted in Fig. 5, is given (5/3). The performance of the Haar kernel with fix reference frames (Haar) and frame-adaptive motion compensation (Haar, $M = 8$) is also plotted. We observe for both sequences, that the 5/3 kernel outperforms the Haar kernel and that frame-adaptive motion compensation improves the performance of both. In any case, the gain in compression efficiency grows with increasing bit-rate.

4 Conclusions

This paper investigates motion-compensated wavelets that are implemented by using the lifting scheme. This scheme permits not only efficient motion compen-

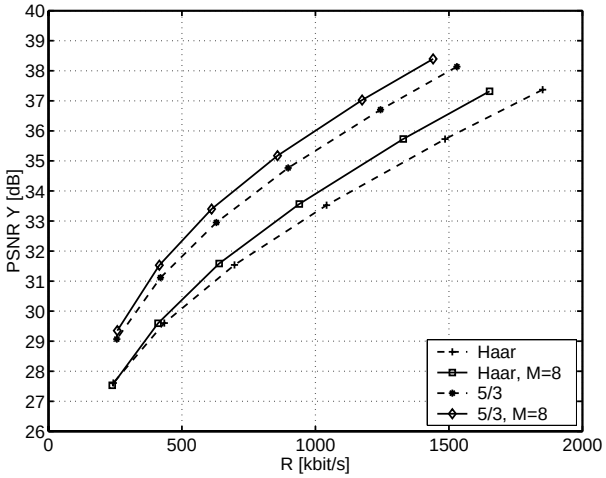

Fig. 8. Luminance PSNR vs. total bit-rate for the QCIF sequence *Mobile & Calendar* at 30 fps. A dyadic decomposition is used to encode groups of $K = 32$ pictures with the frame-adaptive 5/3 kernel. Results for the non-adaptive 5/3 kernel as well as the Haar kernel are given for reference.

sation between successive pictures, but also frame-adaptive motion compensation that utilizes a set of M even pictures in the GOP for reference. We investigate Haar and 5/3 kernels and show experimentally that both kernels can be improved by frame-adaptive motion compensation.

References

1. Sweldens, W.: The lifting scheme: A construction of second generation wavelets. SIAM Journal on Mathematical Analysis **29** (1998) 511–546
2. Secker, A., Taubman, D.: Motion-compensated highly scalable video compression using an adaptive 3D wavelet transform based on lifting. In: Proceedings of the IEEE International Conference on Image Processing. Volume 2., Thessaloniki, Greece (2001) 1029–1032
3. Budagavi, M., Gibson, J.: Multiframe block motion compensated video coding for wireless channels. In: Thirtieth Asilomar Conference on Signals, Systems and Computers. Volume 2. (1996) 953–957
4. Wiegand, T., Zhang, X., Girod, B.: Long-term memory motion-compensated prediction. IEEE Transactions on Circuits and Systems for Video Technology **9** (1999) 70–84
5. Flierl, M., Girod, B.: Investigation of motion-compensated lifted wavelet transforms. In: Proceedings of the Picture Coding Symposium, Saint-Malo, France (2003) 59–62