MOTION COMPENSATED IN-BAND PREDICTION FOR WAVELET-BASED SPATIALLY SCALABLE VIDEO CODING

Claudia Mayer

Institute of Communications Engineering
Aachen University (RWTH)
D-52056 Aachen, Germany

ABSTRACT

In this paper, a wavelet-based spatially scalable video coding scheme with in-band prediction is investigated in combination with JPEG 2000 quantization and arithmetic coding. The wavelet decomposition of each frame of the video data is succeeded by the computation of a blockwise motion compensated prediction for the individual subbands. To achieve half-pixel motion vector accuracy the shift-invariant overcomplete wavelet transform is exploited. Higher accuracy is obtained by additional interpolation of the overcomplete subbands to maintain spatial scalability. A blockwise decision is made between coding of the motion compensated prediction error (inter mode) or the wavelet coefficients (intra mode). The concept of in-band prediction can be easily combined with a conventional intra-frame wavelet coder. In this approach, JPEG 2000 arithmetic coding is applied to the resulting prediction error.

1. INTRODUCTION

Due to the increasing demand for multimedia applications such as video streaming on the internet or the support of decoders with differing complexity a high level of scalability is required for the encoded data. Full spatial, temporal, and SNR scalability should, therefore, be provided by future video coding schemes. To achieve spatial scalability the discrete wavelet transform (DWT) has been successfully employed in image and video compression. It is used in the still image coding standard JPEG 2000 [1], and numerous approaches have been proposed for its application to video compression.

In this paper, we investigate a spatially scalable wavelet video coding scheme with blockwise motion compensated in-band prediction and combine it with JPEG 2000 quantization and arithmetic coding. As in [2, 3, 4], the shift-invariance of the overcomplete discrete wavelet transform (ODWT) is exploited to achieve half-pixel motion vector accuracy. Higher accuracy is obtained by additional interpolation of the overcomplete subbands. The video coding scheme is extended by a blockwise decision between inter and intra coding for the subbands of higher frequency, since in this case the coding of a motion compensated prediction and the corresponding motion vector data is not necessarily superior to the coding of the original wavelet coefficients. An important feature of the in-band prediction scheme is that it can be easily combined with a conventional intra-frame wavelet coder by merely inserting the computation of a motion compensated prediction as an additional processing step between wavelet decomposition and coefficient coding. In our approach, the still image coder JPEG 2000 is modified such that only quantization and arithmetic coding are applied to the prediction error that results from the motion compensated in-band coder.

This paper is organized as follows. In section 2 the in-band video coding scheme is presented, focusing on the application of the ODWT for motion compensation and the employed motion estimation procedure as well as the inter/intra decision and JPEG 2000 coding. Section 3 presents the experimental results, followed by some concluding remarks in section 4.

2. SPATIALLY SCALABLE VIDEO CODING WITH IN-BAND PREDICTION

Figure 1 shows the structure of the spatially scalable video coding scheme for a transform depth $d = 2$. Each video frame is decomposed by a DWT with d transform levels, employing the 9/7 linear phase FIR lowpass (L) and highpass (H) filters of [5]. At each level i filtering and subsequent downsampling leads to the three subbands LH^i, HL^i, HH^i, and the lowpass band LL^i which is further decomposed until LL^d is obtained. For each individual subband a blockwise motion compensated prediction is computed. The resulting prediction error is quantized and coded using JPEG 2000 lossy compression. In Figure 1 Q denotes the quantization performed by JPEG 2000 at a certain quality layer. Each subband is reconstructed at the encoder and

Email: mayer@ient.rwth-aachen.de

stored as reference for the prediction of the subbands of the following video frame. To maintain spatial scalability, no information of the subbands at a higher spatial decomposition level may be used for the subband prediction at a certain level i. On the other hand, all information up to the transform level i can be exploited to improve the prediction at this level.

2.1. Motion Compensation and the Overcomplete Discrete Wavelet Transform

While the reconstructed reference lowpass band $\widetilde{LL^d}$ of the previously coded image has to be interpolated to achieve sub-pixel accuracy for motion compensation, the overcomplete discrete wavelet transform is employed for the prediction of the subbands of higher frequency. For a level i an inverse discrete wavelet transform (IDWT) is applied to the reconstructed subbands $\widetilde{LH^i}$, $\widetilde{HL^i}$, $\widetilde{HH^i}$, and the corresponding synthesized lowpass band $\widetilde{LL^i}$ at this level, resulting in the recovered lowpass band $\widetilde{LL^{i-1}}$. This lowpass band is again decomposed using the ODWT, i.e. the downsampling of the coefficients by two is ommitted. This shift-invariant transform results in the three overcomplete subbands $\widetilde{OLH^i}$, $\widetilde{OHL^i}$, $\widetilde{OHH^i}$ which now serve as reference subbands with double resolution for motion compensation, as shown in figure 2. The IDWT-ODWT procedure can thus be interpreted as the optimum interpolation of half-pixel positions. An efficient algorithm for this conversion from complete to overcomplete subbands has been proposed in [2]. If higher motion vector accuracy is required an interpolation of the overcomplete subbands is necessary to maintain spatial scalability. In our video coding scheme, we apply bilinear interpolation to the overcomplete subbands. For the lowpass band at spatial transform level d higher order interpolation filters are employed.

2.2. Motion Estimation

There are various possibilities to perform motion estimation. Motion vectors can be determined separately either for each individual subband or for each spatial transform level. To save motion vector costs we choose the second approach in which a new motion vector field is computed for each transform level i, i.e. the same motion vectors are used for the prediction of the three subbands LH^i, HL^i, HH^i. For level d the lowpass band LL^d is treated separately. In our video coding scheme, motion estimation is performed on the original video data using a block-matching algorithm. The same motion vector field is thus employed for compensation at all quantization layers which allows for later extension of the in band-video coding scheme to SNR scalability. Alternatively, motion estimation can be performed in the coding loop between the current video frame and the recon-

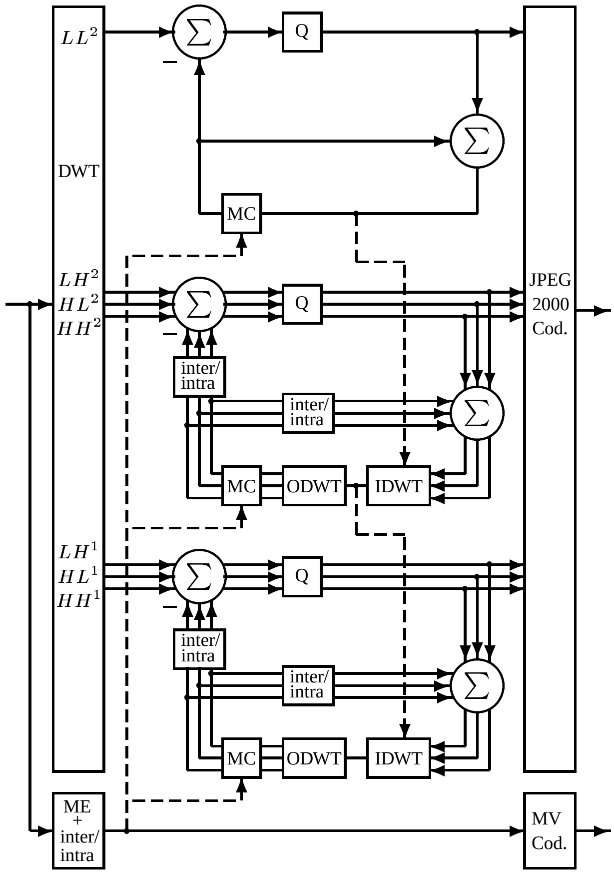

Fig. 1. Spatially scalable video coding scheme for two transform levels

structed previously coded frame at a given base layer quality. The estimation of motion on the individual subbands LH^i, HL^i, and HH^i does not work satisfactory, since minimizing the sum of absolute differences (SAD) does not result in the estimation of true motion and a homogeneous vector field [6]. This is mainly due to the lack of the lowpass component. Therefore, for a transform level i, motion estimation is performed on the lowpass band LL_n^{i-1} of level $i-1$ of the current video frame, as shown in figure 2. The corresponding lowpass band LL_{n-1}^{i-1} of the previous video frame serves as reference image. For each block of size $2N \times 2N$ an optimum motion vector with full-pixel accuracy is computed by minimizing the SAD. A motion compensated prediction with half-pixel accuracy is then computed for the corresponding three subblocks of size $N \times N$ of the subbands LH_n^i, HL_n^i, and HH_n^i by applying this vector to the overcomplete subbands $\widetilde{OLH_{n-1}^i}$, $\widetilde{OHL_{n-1}^i}$, $\widetilde{OHH_{n-1}^i}$ of the previously coded frame. To achieve higher motion vector accuracy the lowpass band $\widetilde{LL_{n-1}^{i-1}}$ and the reference subbands $\widetilde{OLH_{n-1}^i}$, $\widetilde{OHL_{n-1}^i}$, $\widetilde{OHH_{n-1}^i}$ have to

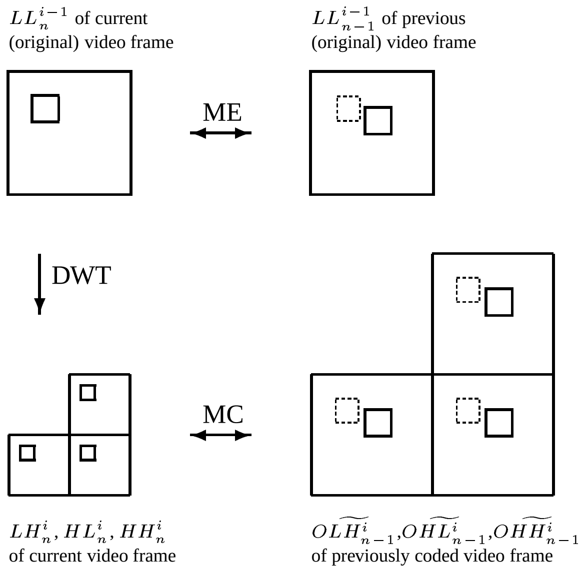

Fig. 2. Motion estimation (ME) and compensation (MC)

be interpolated accordingly. The same interpolation filters are employed as for motion compensation. In this case the motion vector search is hierarchically refined to determine the optimum sub-pixel position.

2.3. Inter/Intra Decision

The wavelet coefficients of the subbands of higher frequency are small compared to those of the lowpass band, and aliasing effects are induced by the application of filter banks with non-ideal lowpass and highpass filters. Further, no correspondence may exist between the actual coefficient and its prediction due to occlusion effects and wrong motion estimates, since only translational motion is considered. For these reasons, the coding of a motion compensated prediction and the additional motion vector data does not necessarily outperform the coding of the original wavelet coefficients. Therefore, a blockwise decision is made for the subbands of higher frequency whether to code the motion compensated prediction error (inter mode) or the wavelet coefficients (intra mode). For the lowpass band inter coding mode is set as default.

The decision between inter and intra coding is made during motion estimation, i.e. on the original video data. The mode selection can, alternatively, be performed in the coding loop at a given quality layer. The sum of the SAD of the prediction error of three $N \times N$ subblocks is compared with the sum of the absolute wavelet coefficients of the three corresponding subblocks. In addition, motion vector costs can be taken into account to favor intra coding. The inter/intra mode selection has to be signaled. This corresponds to an additional information of 1 bit for three $N \times N$ subblocks.

2.4. JPEG 2000 Arithmetic Coding

The in-band video coding scheme is combined with JPEG 2000 lossy compression, using the software included in [1]. Since no further wavelet decomposition needs to be performed, the JPEG 2000 codec is modified such that only the quantization and coding unit are employed. The prediction error is quantized and arithmetically coded, resulting in a JPEG 2000 bit stream partitioned into 12 quality layers. The JPEG 2000 decoder is modified accordingly.

2.5. Temporal and SNR scalability

The presented video coding scheme with in-band prediction can be extended with respect to temporal and SNR scalability. To achieve temporal scalability, the selection of reference frames for prediction has to be adapted. Frames not used as reference can then be skipped without inducing drift. Drift-free SNR scalability can for instance be achieved by combining motion compensated predictive coding at a defined base layer quality with additional intra coding of the resulting quantization error. On the other hand, different methods with motion compensation in the enhancement layer that also avoid drift or, alternatively, allow low drift can be employed [7].

3. SIMULATION RESULTS

Simulations have been performed with CIF sequences at 30 Hz frame rate. A wavelet decomposition with $d = 2$ transform levels is employed. The block size for motion compensation is chosen as 4×4 pixel for all spatial transform levels. For the lowpass band, a motion vector accuracy of 1/4-pixel is selected. For the subbands of higher frequency the application of both 1/2-pixel and 1/4-pixel accuracy is evaluated. The motion vectors are coded using a variable length code. Figure 3 shows the rate-distortion performance of the in-band video coder with different parameter settings in comparison to pure JPEG 2000 intra coding (INTRA) for the luminance components of the sequences ”Foreman”, ”Tempete”, and ”Paris”. Simulation results are given for pure motion compensated inter coding with 1/2-pixel and 1/4-pixel motion vector accuracy, respectively, for the subbands of higher frequency (MC-1/2 INTER, MC-1/4 INTER) as well as for additional inter/intra mode selection (MC-1/2 INTER/INTRA, MC-1/4 INTER/INTRA).

JPEG 2000 intra coding is outperformed by motion compensated inter coding, in particular for sequences with low motion activity or static background (”Paris”). For sequences with high motion activity (”Foreman”, ”Tempete”) the coding performance can be improved by the application of 1/4-pixel motion vector accuracy. The inter/intra mode selection leads to an additional performance gain.

On the other hand, for sequences with a static background, the coding performance is not altered by these variations. In this case, the performance is actually comparable to simple DPCM coding with an additional inter/intra selection.

4. CONCLUSIONS

A spatially scalable video coding scheme with in-band prediction combined with the intra-frame wavelet coder JPEG 2000 has been presented. The shift-invariant ODWT is exploited for motion compensation to achieve half-pixel vector accuracy. While the prediction can be improved by higher motion vector accuracy, the motion vector costs increase. A trade-off has to be made to optimize the overall coding performance. Our current work, therefore, includes the investigation of the impact of different block sizes and motion vector accuracy on the coding performance.

5. REFERENCES

[1] D. S. Taubman and M. W. Marcellin, *JPEG 2000 — Image Compression Fundamentals, Standards and Practice,* Kluwer Academic Publishers, USA, 2002.

[2] H. W. Park and H. S. Kim, "Motion estimation using low-band-shift method for wavelet-based moving-picture coding," *IEEE Trans. Image Processing*, vol. 9, no. 4, pp. 577–587, Apr. 2000.

[3] X. Li, L. Kerofski, and S. Lei, "All-phase motion compensated prediction in the wavelet domain for high performance video coding," in *Proc. IEEE Int. Conf. Image Processing ICIP '01*, Thessaloniki, GR, 2001, vol. 3, pp. 538–541.

[4] Y. Andreopoulos, A. Munteanu, G. Van der Auwera, P. Schelkens, and J. Cornelis, "Wavelet-based fully scalable video coding with in-band prediction," in *Benelux Signal Processing Symposium*, Leuven, BE, Mar. 2002, IEEE.

[5] Marc Antonini, Michel Barlaud, Pierre Mathieu, and Ingrid Daubechies, "Image coding using wavelet transform," *IEEE Trans. Image Processing*, vol. 1, no. 2, pp. 205–220, Apr. 1992.

[6] Simon Albert, "Coding control for wavelet-based video compression," M.S. thesis, RWTH Aachen, Institute of Communications Engineering, 2002.

[7] Claudia Mayer, Holger Crysandt, and Jens-Rainer Ohm, "Bit plane quantization for scalable video coding," in *Proc. SPIE Visual Communications and Image Processing '02*, San Jose, CA, USA, Jan. 2002, pp. 1142–1152.

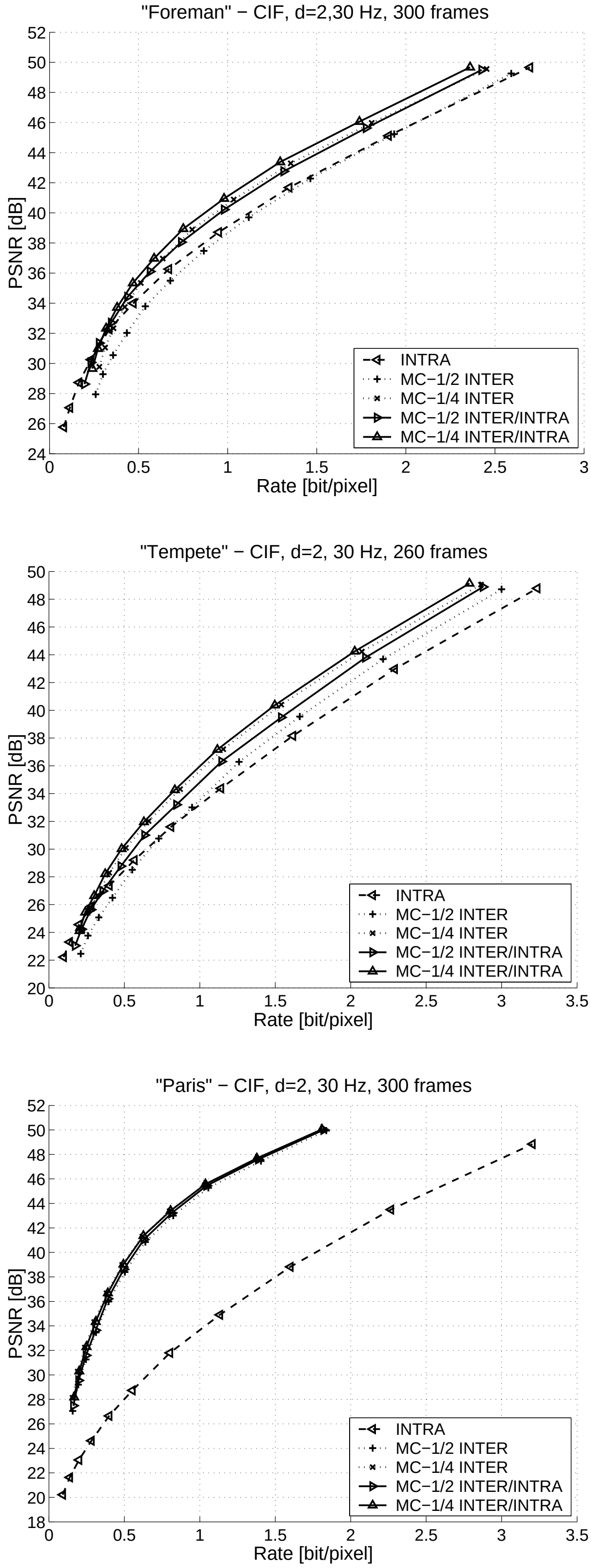

Fig. 3. Simulation results for the sequences "Foreman", "Tempete", and "Paris"