

Post-Processing for Decoding without Update Step in Motion-Compensated Lifted Wavelet Video Coding

Aditya Mavlankar, Markus Flierl and Bernd Girod

Max Planck Center for Visual Computing and Communication
Stanford University, Stanford, CA 94305

ABSTRACT

Motion-compensated temporal filtering is an open-loop coding technique, which generally employs a motion-compensated update step. Although the update step is essential at the encoder for good rate-distortion efficiency, it might be skipped at the decoder for benefits like lower complexity and lower playout delay. Previous investigations showed that skipping the update step at the decoder results in some quality degradation at high rates. In this paper we analyze how this degradation arises and also propose a simple method to reduce this degradation. The proposed solution can also be implemented as a post-processing procedure after conventional decoding without the update step. Experimental results show that the degradation in quality is reduced by half. For our t+2D wavelet coder, this gives a gain of $\sim 1.0 - 1.5$ dB at high bit-rates.

1. INTRODUCTION

The Joint Video Team (JVT) currently employs closed-loop coding with hierarchical B pictures for the scalable extension of H.264/AVC.¹ One of the goals of scalable video coding is to provide SNR scalability. This implies that the bit-rate of the coded representation can be downscaled by truncating portions of the bitstream. Depending on the exact prediction structure employed for the closed-loop coding this might introduce severe to hardly noticeable drift. The drift control mechanisms employed can be quite complex as described in Refs. 2 and 3. On the other hand, the open-loop coding approach controls distortion propagation by employing a temporal transform instead of a prediction loop. Motion-compensated lifted wavelet coding is an open-loop coding technique, which generally employs a motion-compensated update step.

Before adopting the closed-loop encoding approach, the Joint Video Team (JVT) has previously investigated the open-loop encoding approach for the scalable extension of H.264/AVC. The update step employed in the open-loop approach entails higher complexity. Besides this drawback, performing the update step at the decoder leads to increased playout delay. The playout delay is increased because the update step at the decoder implies that the reconstruction of certain frames depends on the reception of more subbands compared to the scenario where the decoder skips the update step. At the encoder, however, the update step is required for improving the coding efficiency with the open-loop encoding approach.^{4,5} For some representative video sequences, Ref. 6 presents a comparison of rate-distortion efficiency of three encoding approaches, viz. closed-loop coding with hierarchical B pictures; open-loop coding with update step at both the encoder and the decoder; and open-loop coding with update step at the encoder but *not* at the decoder. It was observed that for the *City* video sequence, the open-loop approaches with motion-compensated temporal filtering outperform closed-loop coding with hierarchical B pictures for the test conditions noted in Ref. 6. For the widely used 5/3 wavelet, it was observed that performing the update step at the encoder but *not* at the decoder entails some coding efficiency loss at high rates. Nevertheless, this loss vanishes as the rate is lowered. Li et al. reported this phenomenon in Ref. 7 for other video sequences and also theoretically analyzed it for the update step specified in Ref. 8. We have observed similar coding efficiency loss at high rates with our t+2D wavelet coder. In this paper we focus on the open-loop encoding approach followed by most t+2D wavelet coders, which provides a flexible scalable video representation. Furthermore, we assume that the encoder performs the update step but the decoder does

This work was supported by the Max Planck Center for Visual Computing and Communication.

Further author information: E-mail: {maditya, mflierl, bgirod}@stanford.edu, Telephone: +1 650 723 3476, Fax: +1 650 724 3648.

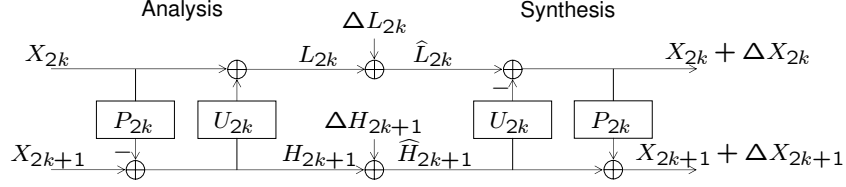


Figure 1. Motion-compensated lifted Haar wavelet with update step performed at the encoder as well as the decoder.

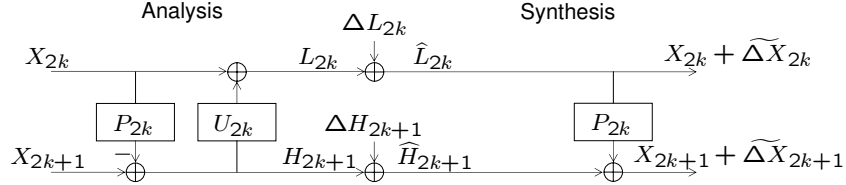


Figure 2. Motion-compensated lifted Haar wavelet with update step performed at the encoder but *not* at the decoder.

not. For the Haar wavelet as well as the 5/3 wavelet, we propose low-complexity post-processing, which helps to mitigate the loss in coding efficiency.

For both considered wavelets, we exploit the fact that when the decoder skips the update step, there appears a spurious signal component, i.e., an artifact in the reconstruction, which is close to a scaled version of the original highband. The proposed method is simple because it involves subtracting a scaled version of the received highband. This not only gets rid of the spurious signal component but also reduces the impact of the quantization error added to the original highband during its encoding.

This paper is organized as follows: Section 2 discusses the proposed post-processing method for the Haar wavelet and section 3 for the 5/3 wavelet. Each of these sections analyzes the quantization error propagation for the respective wavelet before describing the proposed post-processing. Each of these sections also proposes an alternative decoding procedure, which achieves the same distortion as the conventional decoding followed by the proposed post-processing, and is yet less computationally expensive. Section 4 shows the experimental results.

2. RECONSTRUCTION PROCEDURE FOR HAAR WAVELET

2.1. Analysis of Quantization Error Propagation

As shown in **Fig. 1**, the even and odd frames of a video sequence are represented by column vectors X_{2k} and X_{2k+1} respectively. Prediction and update operators can then be represented by square matrices P_{2k} and U_{2k} respectively. The lowpass subband and the highpass subband, obtained after linear operations in the analysis stage, are represented by column vectors L_{2k} and H_{2k+1} respectively. The errors introduced, due to quantization, in the two subbands L_{2k} and H_{2k+1} are represented by column vectors ΔL_{2k} and ΔH_{2k+1} respectively. Using this notation, we analyze how the distortion propagates from the temporal subbands to the reconstructed frames. Then we point out how the distortion propagation changes when the decoder skips the update step. This observation leads us to the new decoding procedure, which entails lower distortion. The proposed solution could also be implemented as a post-processing procedure.

When the update step is performed at the decoder, as shown in **Fig. 1**, the errors propagated to the reconstructed frames are denoted by ΔX_{2k} and ΔX_{2k+1} . When the update step is not performed at the decoder, as shown in **Fig. 2**, these errors are denoted by $\widetilde{\Delta X}_{2k}$ and $\widetilde{\Delta X}_{2k+1}$. These error vectors are given by

$$\Delta X_{2k} = \Delta L_{2k} - U_{2k} \Delta H_{2k+1}, \quad (1)$$

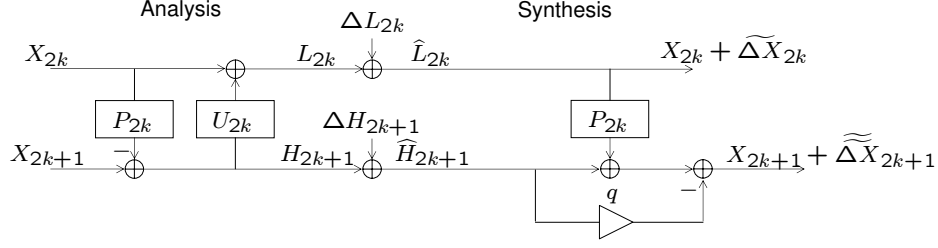


Figure 3. Motion-compensated lifted Haar wavelet with post-processing for decoding without the update step.

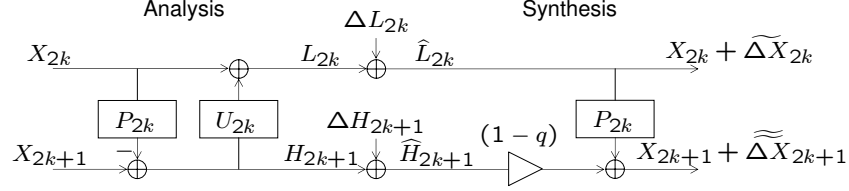


Figure 4. Motion-compensated lifted Haar wavelet with proposed procedure for decoding without the update step.

$$\widetilde{\Delta X}_{2k} = \Delta L_{2k} + U_{2k} H_{2k+1}, \quad (2)$$

$$\Delta X_{2k+1} = P_{2k} \Delta L_{2k} + \Delta H_{2k+1} - P_{2k} U_{2k} \Delta H_{2k+1}, \quad (3)$$

$$\widetilde{\Delta X}_{2k+1} = P_{2k} \Delta L_{2k} + \Delta H_{2k+1} + P_{2k} U_{2k} H_{2k+1}. \quad (4)$$

2.2. Proposed Post-Processing Procedure

We make the following approximation: $P_{2k} U_{2k} \approx \frac{1}{2} I$. This is because for the Haar wavelet, the update operator scales the signal by $\frac{1}{2}$ after performing motion compensation with motion vectors that reverse the direction used by the prediction operator. However, this approximation ignores the effect that if a temporal subband is subjected to the operator $P_{2k} U_{2k}$, then the high spatial frequencies might be attenuated due to spatial smoothing resulting from sub-pel motion compensation in the prediction and update operators. In general, the attenuation of the high and mid spatial frequencies depends on the interpolation filter used for sub-pel motion compensated prediction and update operators. With bilinear interpolation used for the motion compensation and the conventional update step, it was observed that most temporal subbands looked approximately like they were scaled by $\frac{1}{2}$ when subjected to the operator $P_{2k} U_{2k}$.

The proposed post-processing comprises of subtracting a scaled version of the received highband, i.e., $q \hat{H}_{2k+1}$, from the reconstructed odd frame. This is shown in **Fig. 3**. The reconstruction error then becomes

$$\begin{aligned} \widetilde{\Delta X}_{2k+1} &= \widetilde{\Delta X}_{2k+1} - q \hat{H}_{2k+1} \\ &= P_{2k} \Delta L_{2k} + (1 - q) \Delta H_{2k+1} + (P_{2k} U_{2k} - qI) H_{2k+1}. \end{aligned} \quad (5)$$

If q is set to $\frac{1}{2}$ then a comparison of the reconstruction errors in (4) and (5) shows that the contribution due to H_{2k+1} is nearly completely canceled and the contribution due to ΔH_{2k+1} is reduced. Although $q > \frac{1}{2}$ would reduce the contribution due to ΔH_{2k+1} even further, it might entail more net distortion since H_{2k+1} is likely to have more energy than ΔH_{2k+1} . The advantage of low playout delay due to skipping the update step is preserved. The proposed post-processing has low complexity. However, if the decoding procedure itself is slightly modified, as explained in the next subsection, then the complexity is even lower.

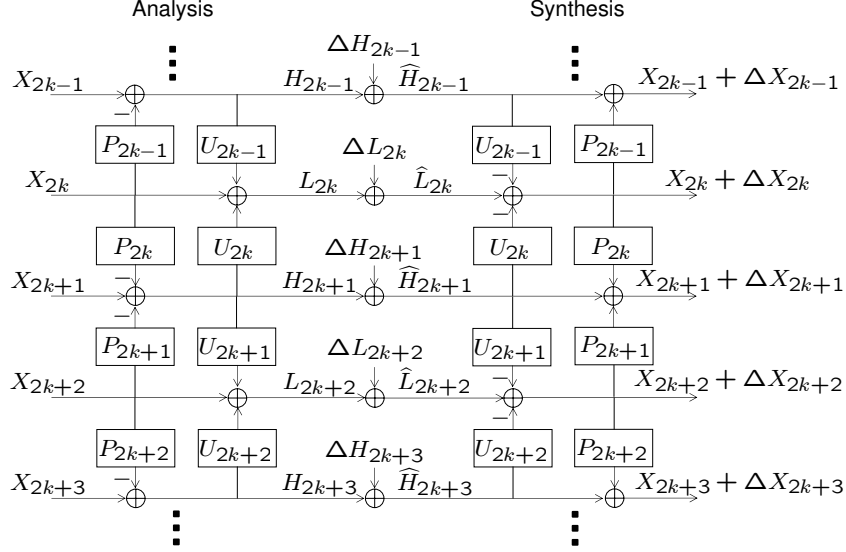


Figure 5. Motion-compensated lifted 5/3 wavelet with update step performed at the encoder as well as the decoder.

2.3. Proposed Decoding Procedure

The above post-processing can be replaced with an equivalent modified decoding procedure as follows. The received highband can be scaled by $(1 - q)$ before using it for the synthesis. This is shown in **Fig. 4**. This approach preserves the benefit of low playout delay and entails lower complexity than using the above post-processing following conventional decoding without the update step.

2.4. Reduction of Quality Fluctuation

It has been observed^{9,10} that the distortion in the temporal subbands propagates in an uneven manner to the reconstructed frames. Typically, the odd frame suffers higher distortion than the even frame. The following modification in the proposed post-processing allows even further reduction of the distortion in the odd frame. If a scaled version of the update step, i.e., $\widehat{U}_{2k} = \beta U_{2k}$, $1 \leq \beta \leq 2$, is used then more noise cancelation in the odd frame can be achieved, i.e.,

$$\begin{aligned} \widetilde{\Delta X}_{2k+1} &= \widetilde{\Delta X}_{2k+1} - \beta q \widehat{H}_{2k+1} \\ &= P_{2k} \Delta L_{2k} + (1 - \beta q) \Delta H_{2k+1} + (\beta P_{2k} U_{2k} - \beta q I) H_{2k+1}. \end{aligned} \quad (6)$$

However, this is achieved at the cost of increased distortion in the even frame, given by

$$\widetilde{\Delta X}_{2k} = \Delta L_{2k} + \beta U_{2k} H_{2k+1}. \quad (7)$$

The value of β can be chosen by the encoder to minimize total distortion or achieve equal distortion in the odd and even reconstructed frames.

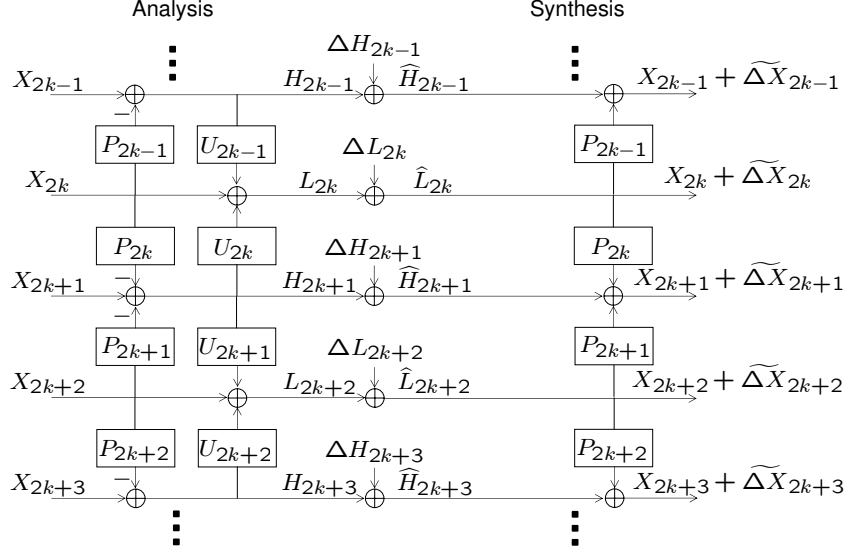


Figure 6. Motion-compensated lifted 5/3 wavelet with update step performed at the encoder but *not* at the decoder.

3. RECONSTRUCTION PROCEDURE FOR 5/3 WAVELET

3.1. Analysis of Quantization Error Propagation

Using the same notation as above, the reconstruction error with and without the update step at the decoder, as shown in **Fig. 5** and **Fig. 6** respectively, can be written as

$$\Delta X_{2k} = \Delta L_{2k} - U_{2k} \Delta H_{2k+1} - U_{2k-1} \Delta H_{2k-1}, \quad (8)$$

$$\widetilde{\Delta X}_{2k} = \Delta L_{2k} + U_{2k} H_{2k+1} + U_{2k-1} H_{2k-1}, \quad (9)$$

$$\begin{aligned} \Delta X_{2k+1} = & P_{2k} \Delta L_{2k} + P_{2k+1} \Delta L_{2k+2} - P_{2k+1} U_{2k+2} \Delta H_{2k+3} - P_{2k} U_{2k-1} \Delta H_{2k-1} \\ & + \Delta H_{2k+1} - (P_{2k} U_{2k} + P_{2k+1} U_{2k+1}) \Delta H_{2k+1}, \end{aligned} \quad (10)$$

$$\begin{aligned} \widetilde{\Delta X}_{2k+1} = & P_{2k} \Delta L_{2k} + P_{2k+1} \Delta L_{2k+2} + P_{2k+1} U_{2k+2} H_{2k+3} + P_{2k} U_{2k-1} H_{2k-1} \\ & + \Delta H_{2k+1} + (P_{2k} U_{2k} + P_{2k+1} U_{2k+1}) H_{2k+1}. \end{aligned} \quad (11)$$

3.2. Proposed Post-Processing Procedure

For the 5/3 wavelet we make the following approximation $P_i U_j (i = j) \approx \frac{1}{8} I$. The reasoning is similar to that for the Haar wavelet with the difference that the scaling factors employed by both the prediction and update operators in the 5/3 wavelet are half of the factors employed in the Haar wavelet. We propose similar post-processing as for the Haar wavelet, however, with a different value of q . The diagram is not shown. If q is chosen to be $\frac{1}{4}$, then the contribution due to H_{2k+1} is nearly completely canceled. The contribution due to ΔH_{2k+1} is reduced. The reconstruction error then becomes

$$\begin{aligned} \widetilde{\widetilde{\Delta X}}_{2k+1} &= \widetilde{\Delta X}_{2k+1} - q \hat{H}_{2k+1} \\ &= P_{2k} \Delta L_{2k} + P_{2k+1} \Delta L_{2k+2} + P_{2k+1} U_{2k+2} H_{2k+3} + P_{2k} U_{2k-1} H_{2k-1} \\ &\quad + (1 - q) \Delta H_{2k+1} + (P_{2k} U_{2k} + P_{2k+1} U_{2k+1} - q I) H_{2k+1}. \end{aligned} \quad (12)$$

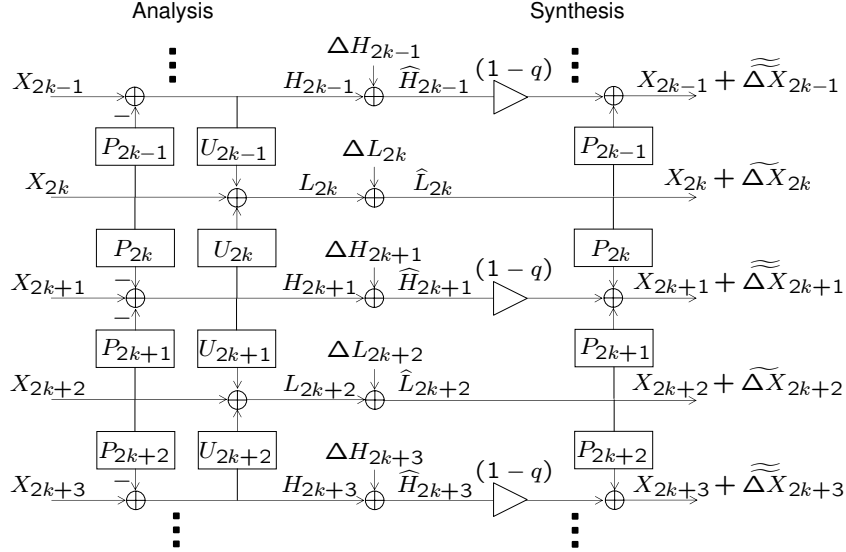


Figure 7. Motion-compensated lifted 5/3 wavelet with proposed procedure for decoding without the update step.

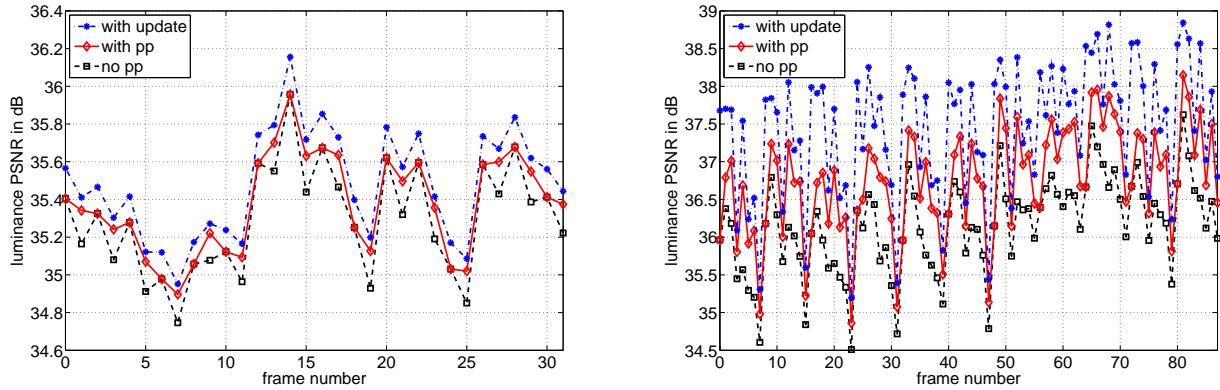


Figure 8. PSNR for a sequence of frames from the *City* (CIF) sequence. The gain due to the proposed post-processing (pp) can be seen. 1 level of Haar wavelet (*left*). 3 levels of Haar wavelet (*right*). The results shown here are representative of other frames in the video sequence. The bit-rates are 1800 kbps (*left*) and 1300 kbps (*right*).

3.3. Proposed Decoding Procedure

Similar to the case of the Haar wavelet, the above post-processing can be avoided by modifying the decoding procedure itself. This is done by scaling the received highband by $(1 - q)$ before using it for the synthesis. The implementation of the proposed decoding procedure is shown in Fig. 7.

4. EXPERIMENTAL RESULTS

For our experiments, we employ the conventional update step. We encode the temporal subbands using JPEG2000. The PSNR for a sequence of frames is illustrated for the Haar wavelet and the *City* (CIF) sequence in Fig. 8. The gain obtained by the proposed post-processing can be seen. For the more widely used 5/3 wavelet, the rate-distortion (R-D) performance, for various levels of temporal decomposition, is shown in Figures 9 through 11. The R-D curves on the *left* are for the *City* (CIF) sequence and those on the *right* are for the *Foreman* (CIF) sequence. The reduction in quality degradation at high rates can be seen from these

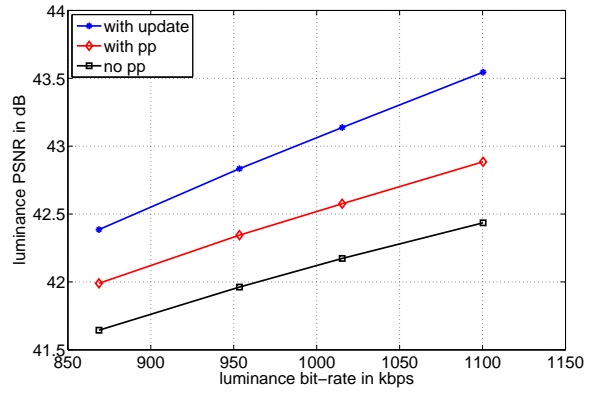
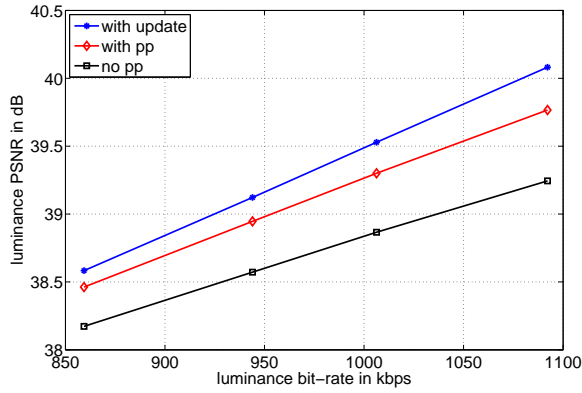


Figure 9. Rate-distortion performance for 2 levels of the 5/3 wavelet. *City* (CIF) sequence (left). *Foreman* (CIF) sequence (right). The gain in rate-distortion performance due to the proposed post-processing (pp) can be seen.

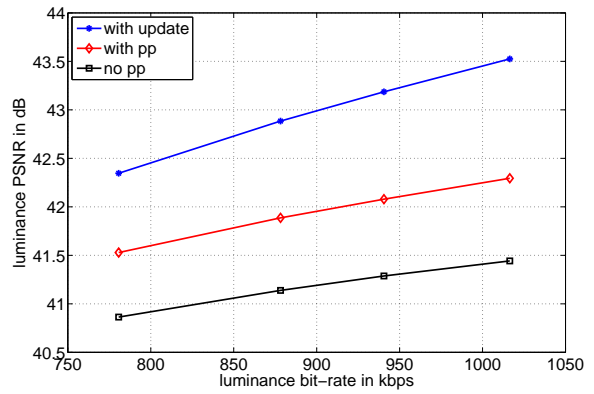
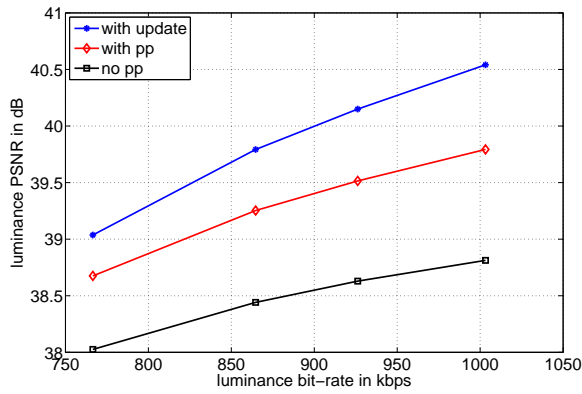


Figure 10. Rate-distortion performance for 3 levels of the 5/3 wavelet. *City* (CIF) sequence (left). *Foreman* (CIF) sequence (right). The gain in rate-distortion performance due to the proposed post-processing (pp) can be seen.

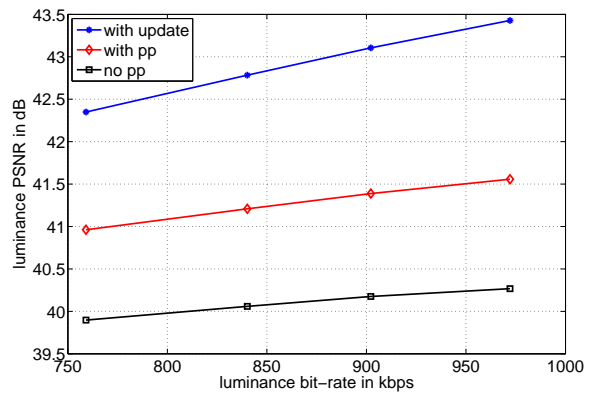
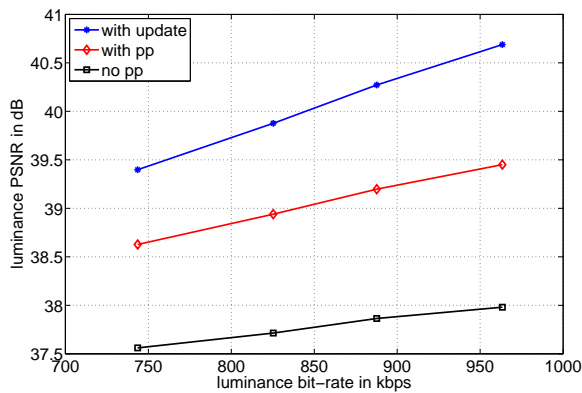


Figure 11. Rate-distortion performance for 4 levels of the 5/3 wavelet. *City* (CIF) sequence (left). *Foreman* (CIF) sequence (right). The gain in rate-distortion performance due to the proposed post-processing (pp) can be seen.

plots. These R-D curves are obtained after encoding 128 frames. We employ quarter-pel motion compensation with hierarchical variable size block matching. Note that in case of multi-level temporal decomposition, the post-processing is carried out for odd reconstructed temporal subbands (or frames) at each level.

For both the Haar and the 5/3 wavelets, we make two observations. Firstly, the gain obtained by the proposed post-processing increases with the number of levels of temporal decomposition. Secondly, all frames within a group of pictures (GOP), except one frame, benefit from the proposed post-processing. When multiple levels of the Haar or 5/3 wavelet decomposition are carried out, the reconstructed odd frame at a given level typically impacts the reconstruction of multiple frames in a GOP. The frame corresponding to the lowpass frame, which generally gets the least distortion⁹ within the GOP, is not affected by the proposed post-processing.

5. CONCLUSION

Motion-compensated lifted wavelet video coding is an open-loop coding approach in which the update step is essential at the encoder for higher rate-distortion efficiency. However, skipping the update step at the decoder has the benefits of lower complexity and lower playout delay. Hence, we propose a new reconstruction procedure which skips the update step at the decoder yet mitigates half of the quality degradation. This approach preserves the benefits of lower complexity and lower playout delay. This new reconstruction procedure can be implemented in two possible ways, i.e., either by slightly modifying the conventional decoding procedure without the update step or by implementing a post-processing procedure after the conventional decoding without the update step.

REFERENCES

1. "JSVM 7.0 Software," Document JVT-T203, Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG (ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6), Klagenfurt, Austria, July 2006. <http://ftp3.itu.ch/av-arch/jvt-site/>.
2. H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable extension of the H.264/MPEG-4 AVC Video Coding Standard," *IEEE Transactions on Circuits and Systems for Video Technology (CSVT)*, scheduled March 2007. Invited Paper.
3. H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable H.264/MPEG-4 AVC Extension," *Proceedings of the IEEE Intl. Conference on Image Processing (ICIP), Atlanta, Georgia, U.S.A.*, October 2006.
4. J. Reichel, D. S. Cruz, and F. Ziliani, "CE1-Improved residual picture buffer management," Document JVT-P105, Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG (ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6), Poznan, Poland, July 2005. <http://ftp3.itu.ch/av-arch/jvt-site/>.
5. A. Tabatabai, Z. Visharam, and T. Suzuki, "Study of effect of update step in MCTF," Document JVT-Q026, Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG (ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6), Nice, France, October 2005. <http://ftp3.itu.ch/av-arch/jvt-site/>.
6. H. Schwarz, D. Marpe, and T. Wiegand, "Comparison of MCTF and closed-loop hierarchical B pictures," Document JVT-P059, Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG (ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6), Poznan, Poland, July 2005. <http://ftp3.itu.ch/av-arch/jvt-site/>.
7. F. Li, N. Ling, and S. A. Chiappari, "The effect of MCTF update step on coding efficiency for H.264/AVC Scalable Extension: A theoretical approach," *Proceedings of the IEEE Intl. Picture Coding Symposium (PCS), Beijing, China*, April 2006.
8. "JSVM 2.0 Software," Document JVT-O203, Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG (ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6), Busan, Korea, April 2005. <http://ftp3.itu.ch/av-arch/jvt-site/>.
9. C.-L. Chang, A. Mavlankar, and B. Girod, "Analysis of quantization error propagation for motion-compensated lifted wavelet video coding," *Proceedings of the IEEE Intl. Workshop on Multimedia Signal Processing (MMSP), Shanghai, China*, November 2005.
10. A. Mavlankar and E. Steinbach, "Distortion prediction for motion-compensated lifted Haar wavelet transform and its application to rate allocation," *Proceedings of the IEEE Intl. Picture Coding Symposium (PCS), San Francisco, U.S.A.*, December 2004.