Intraframe Prediction with Intraframe Update Step for Motion-Compensated Lifted Wavelet Video Coding

Aditya Mavlankar, Chuo-Ling Chang, and Bernd Girod *

Information Systems Laboratory,
Department of Electrical Engineering, Stanford University,
Stanford CA 94305, USA

Abstract. In motion-compensated lifted wavelet video coding, the calculation of the highpass frame involves interframe prediction. Due to occlusions, or change of intensity, or new objects in the video frame, interframe prediction is not always effective. Intraframe prediction provides an alternative and increases coding efficiency. However, previous analysis shows that these intraframe predicted pixels suffer higher distortion propagation than the rest of the pixels in the open-loop scheme. The main contribution of this paper is to mitigate this increased distortion through an intraframe update step. We propose a perfectly invertible sequence of operations. In addition, we present theoretical analysis of the reduction in distortion due to the inclusion of the proposed intraframe update step for various intraframe prediction modes. This is compared with practical measurements.

Index Terms lifted wavelet, intra prediction, update step

1 INTRODUCTION

Motion-compensated lifted wavelet transform for video compression [1] performs a multi-level wavelet transform along motion trajectories, which decomposes a set of frames into temporal subbands. Every temporal subband can be either encoded using wavelet-based image coding techniques like JPEG2000 [2] or by block-transform-based techniques as recently proposed in the MCTF extension of the H.264/AVC standard [3]. These schemes provide scalability of bit-rate, spatial resolution and temporal resolution.

The calculation of the highpass frame involves interframe prediction. Due to occlusions, or change of intensity, or new objects in the video frame, interframe prediction is not always effective. With more levels of temporal wavelet decomposition, this problem becomes even more serious due to growing temporal distance. In case interframe prediction is inefficient, the interframe update step following it introduces artifacts in the lowpass signal [4]. This affects the quality of temporally downscaled video. Intraframe prediction provides an alternative and also increases coding efficiency. However, the analysis in [4] shows that these intraframe predicted pixels suffer higher distortion propagation than the rest of the pixels in the open-loop scheme. We propose an intraframe update step to mitigate this problem. An improved mode decision based on a Lagrangian cost function and a better intraframe prediction technique for 4x4 blocks whose neighbors are unavailable for reference are among the other improvements presented. Section 2 describes the incorporation of the intraframe update step in the lifting framework. Theoretical analysis of the reduction in distortion due to its inclusion are also presented. Experimental results are analyzed in Section 3.

2 INTRAFRAME PREDICTION WITH INTRAFRAME UPDATE STEP

We propose a perfectly invertible sequence of operations, represented with the block diagram in Figure 1. For intraframe prediction, all 4x4, 8x8 and 16x16 blocks are checked for the best out of at the most 9 intraframe prediction modes. H.264/AVC [5] defines 9 modes for 4x4 blocks. As shown in Figure 2, pixels "a-p" are predicted using pixels "A-M" as reference. The same 9 modes are extended to 8x8 blocks. For 16x16 blocks only 3 modes, viz. Vertical, Horizontal and DC are used. For a given block under consideration, some modes are ruled out due to some neighboring pixels being themselves intraframe predicted and hence not considered available for reference. This is done to limit distortion propagation according to the analysis in [4]. To compensate for this dearth of reference pixels, the unavailable reference pixels can be replaced by surrogates (designated with numbers 1-3 in preferential order according to availability) as shown in Figure 2. Only

* This work was supported, in part, by the Max Planck Center for Visual Computing and Communication.

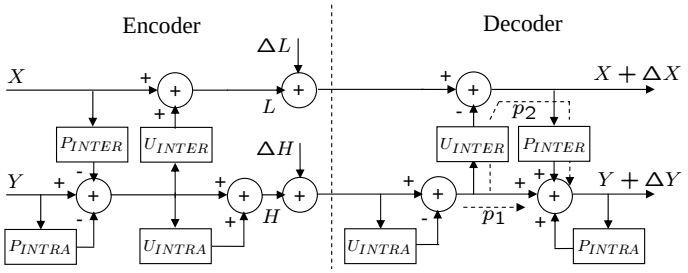

Fig. 1. X and Y are two input video frames. Prediction and Update operators are denoted by P and U respectively. The operators P_{INTER} and P_{INTRA} are picked such that the blocks which are interframe predicted are not intraframe predicted and vice-versa. Error introduced in the subbands is denoted by ΔL and ΔH. Error propagated to reconstructed frames is denoted by ΔX and ΔY.

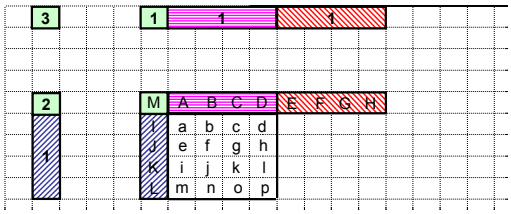

Fig. 2. A 4x4 block can be intraframe predicted from neighboring reference pixels A-M. The unavailable reference pixels can be replaced by surrogates designated with numbers 1-3 in preferential order according to availability. Only surrogate(s) shown using the same shaded pattern can replace a given reference pixel. In case a surrogate reference pixel is used, the encoder and decoder use the available surrogate labeled with the smallest number, i.e., one which is itself not intraframe predicted.

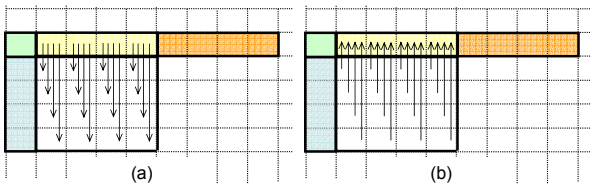

Fig. 3. (a) Intraframe prediction in mode 0, i.e. Vertical Prediction. Weight associated with each arrow is 1. (b) Intraframe update step for mode 0. Weight associated with each arrow is 0.2.

surrogate(s) shown using the same shaded pattern can replace a given reference pixel. Both encoder and decoder follow the same rule and hence no additional signaling is required.

2.1 Intraframe Update

We base our intraframe update step on the optimal update step in [6]. The exact mode-dependent operations (performed on a block-by-block basis) during intraframe update can be stated as follows: Let the pixels from an intraframe predicted block in frame Y be represented as a column vector Y_b. Let the reference pixels be represented as a column vector Y_r. Let P_b be the mode-dependent matrix which forms the intraframe prediction P_bY_r for the block. The highpass coefficients for the block are represented by

$$H_b = Y_b - P_bY_r. \quad (1)$$

The highpass coefficients at the reference pixel locations, represented by column vector $H_{r,INTER}$, receive the intraframe update given by

$$H_r = H_{r,INTER} + (I + P_b^T P_b)^{-1} P_b^T H_b. \quad (2)$$

The location of H_r and $H_{r,INTER}$, is the same and the notation emphasizes that reference pixels are interframe predicted. As an example, consider intraframe prediction in mode 0 as shown in Figure 3(a). The intraframe update step is as shown in Figure 3(b). Consider the reconstruction of pixel 'a' to see the reduced distortion propagation. The values in pixel locations 'A', 'a', 'e', 'i' and 'm' are changed according to the following assignment operations.

Update at decoder: $A \leftarrow A - 0.2(a + e + i + m)$
Predict at decoder: $a \leftarrow a + A$, $e \leftarrow e + A$
$i \leftarrow i + A$, $m \leftarrow m + A$

These operations are intraframe. The reconstruction of the interframe predicted reference pixels takes place between Update and Predict. Let ΔH_x denote the error in subband H at location 'x'. Let $\Delta Y_{A,NO\_UPDATE}$ denote the reconstruction error in pixel 'A' if the intraframe update step is skipped. The reconstruction error in pixel 'A', with the inclusion of the intraframe update step, denoted by

ΔY_A, is given by:

$$\begin{aligned}\Delta Y_A &= \Delta Y_{A,NO\_UPDATE} \\ &+ (-0.2)(\Delta H_a + \Delta H_e + \Delta H_i + \Delta H_m) \\ &+ f(\Delta H_a, \Delta H_e, \Delta H_i, \Delta H_m) \end{aligned} \quad (3)$$

where $(-0.2)(\Delta H_a + \Delta H_e + \Delta H_i + \Delta H_m)$ is the propagation of the errors $\Delta H_a, \Delta H_e, \Delta H_i, \Delta H_m$ along path p_1, and $f(\Delta H_a, \Delta H_e, \Delta H_i, \Delta H_m)$ is the propagation of the same errors along path p_2, as shown in Figure 1. If interframe prediction uses only full-pel MC, then $f(\Delta H_a, \Delta H_e, \Delta H_i, \Delta H_m) \approx (0.1)(\Delta H_a + \Delta H_e + \Delta H_i + \Delta H_m)$, since the interframe update operator scales the input by 0.5 and this is followed by a sign reversal along path p_2. If sub-pel MC is used, then due to the interpolation involved, we can neglect the propagation of terms like $(-0.2\Delta H_a)$ through the interframe update and prediction steps to obtain $f(\Delta H_a, \Delta H_e, \Delta H_i, \Delta H_m) \approx 0$. Simulation results confirm this assumption. The expected distortion at pixel 'a' is given by:

$$\begin{aligned} &E\{(\Delta Y_a)^2\} \approx E\{[(1-0.1)\Delta H_a]^2 \\ &\quad +(-0.1\Delta H_e)^2 + (-0.1\Delta H_i)^2 + (-0.1\Delta H_m)^2 \\ &\quad +(\Delta Y_{NO\_UPDATE})^2\} \quad (4) \\ &\approx 1.84 \text{ for full-pel MC and} \\ &E\{(\Delta Y_a)^2\} \approx E\{[(1-0.2)\Delta H_a]^2 \\ &\quad +(-0.2\Delta H_e)^2 + (-0.2\Delta H_i)^2 + (-0.2\Delta H_m)^2 \\ &\quad +(\Delta Y_{NO\_UPDATE})^2\} \quad (5) \\ &\approx 1.76 \text{ for sub-pel MC,} \end{aligned}$$

assuming that the errors are all zero-mean, uncorrelated and unit variance. If the intraframe update step is skipped then the expected distortion is given by:

$$\begin{aligned} E\{(\Delta Y_a)^2\} &\approx E\{(\Delta H_a)^2 \\ &\quad +(\Delta Y_{NO\_UPDATE})^2\} \quad (6) \\ &\approx 2 \text{ for both full-pel and sub-pel MC.} \end{aligned}$$

This translates to a gain of about 0.36 dB and 0.56 dB for full-pel and sub-pel MC respectively. For mode 0, the gain is the same for all pixels 'a', 'b', $\cdots$, 'p'. This relative gain is obtained at the expense of slightly increased distortion in the reference pixels 'A', 'B', 'C', 'D'. For the same example, this increase in the distortion at pixel 'A' is given by:

$$\begin{aligned} &E\{(-0.1\Delta H_a)^2 + (-0.1\Delta H_e)^2 + (-0.1\Delta H_i)^2 + \\ &\quad (-0.1\Delta H_m)^2\} \approx 0.04, \text{ for full-pel MC and} \\ &E\{(-0.2\Delta H_a)^2 + (-0.2\Delta H_e)^2 + (-0.2\Delta H_i)^2 + \\ &\quad (-0.2\Delta H_m)^2\} \approx 0.16, \text{ for sub-pel MC,} \end{aligned}$$

which is a loss of about 0.17 dB and 0.64 dB for full-pel and sub-pel MC respectively. However, the number of reference pixels is small compared to the number of intraframe predicted pixels. In the given example, 16 pixels were predicted from 4 reference pixels. Figure 1 indicates that some reference pixels in frame X also suffer slightly increased distortion, which comes along path p_2. For full-pel MC, this increase is exactly the same as that suffered by affected reference pixels in frame Y. For sub-pel MC, due to interpolation, the magnitude is reduced. The relative gain in dB for intraframe predicted pixels, due to the inclusion of the proposed intraframe update step, can be obtained using similar analysis for the other modes. The average relative gain per pixel is given in Table 1.

Table 1. Average relative gain per pixel for various modes defined in H.264/AVC [5].

Mode	Average relative gain in dB			
	Full-pel MC		Sub-pel MC	
	4x4 block	8x8 block	4x4 block	8x8 block
0	0.36	0.19	0.56	0.28
1	0.36	0.19	0.56	0.28
2	0.14	0.04	0.22	0.06
3	0.44	0.27	0.77	0.45
4	0.43	0.27	0.75	0.45
5	0.43	0.28	0.76	0.46
6	0.43	0.28	0.76	0.46
7	0.41	0.25	0.71	0.40
8	0.30	0.17	0.48	0.27

Figure 4 shows the results of an experiment where pairs of frames, with varying temporal distance between the two frames, undergo MC lifted Haar wavelet decomposition with intraframe prediction. One scheme performs intraframe update and the other scheme does not. For Figure 4(a), to compare the two schemes, exactly the same noise, depicting quantization noise, is added to the temporal subbands. The noise added in the two temporal subbands L and H is zero-mean with no cross-correlation but has the same variance and is chosen from a uniform distribution. For Figure 4(b), temporal subbands are encoded using JPEG2000 [2] and the same rate allocation among L and H subbands is used for every pair within each scheme. Luminance frames of the *Foreman* (CIF) sequence are used and every point is the average of 4 trials with different pairs. The vertical axis on the left shows the reconstruction PSNR of only the intraframe pre-

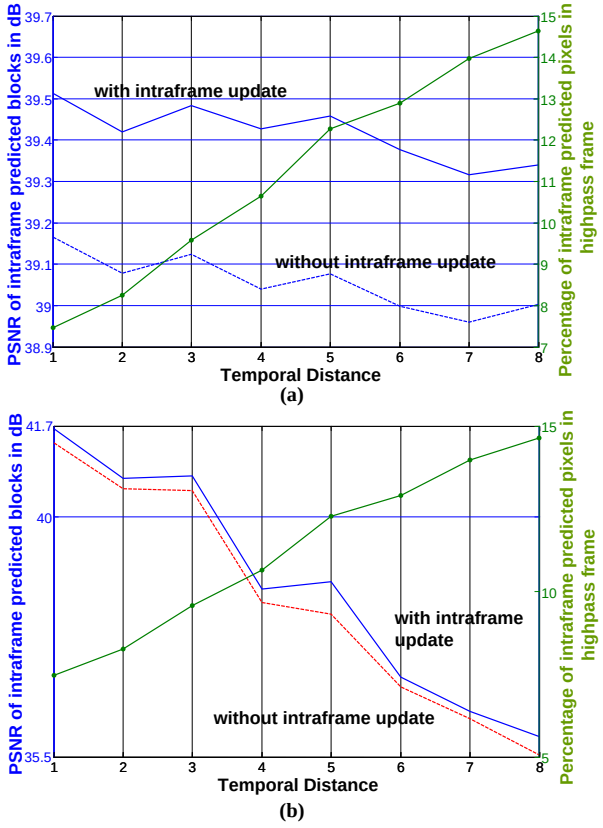

Fig. 4. PSNR of intraframe predicted blocks with and without the proposed intraframe update step. The vertical axis on the right shows the percentage of intraframe predicted pixels in the highpass frame. Quarter-pel MC is used. (a)Quantization noise is simulated as described. (b)Subbands are encoded using JPEG2000.

dicted blocks and the distortion reduction due to the proposed intraframe update step can be seen. For Figure 4(b), the reduction in the PSNR of reference pixels used for intraframe prediction is around 0.12 dB. The reduction in the PSNR of the interframe predicted pixels (which includes the reference pixels used for intraframe prediction) is negligible, around 0.01 dB. The mean PSNR was observed to increase slightly, by around 0.02 dB. Similar low gain in mean PSNR is observed for multi-level temporal transform, which leads to the conclusion that the main benefit of the proposed intraframe update step is to alleviate the distortion propagation suffered by the intraframe predicted blocks.

2.2 Mode Selection

Let D and R denote prediction residual energy and rate, respectively. The criterion minimised during interframe motion estimation is $D_{inter} + \lambda_{inter} R_{inter}$. For intraframe prediction, let R_{intra} be the rate associated with the intra mode that minimizes D_{intra}. We define the following quantities:

$$\Delta D = D_{intra} - D_{inter}$$
$$\Delta R = R_{intra} - R_{inter}$$

For the 4x4, 8x8 or 16x16 block under consideration, we choose intraframe prediction mode if

$$J = \Delta D + \lambda_{intra} \Delta R < 0, \qquad (7)$$

or else we choose the best inter MV as found previously by minimizing $D_{inter} + \lambda_{inter} R_{inter}$. Increasing λ_{intra} helps to save more bits on MV with a corresponding trade-off in distortion. This has more use in the low bit-rate region. The implementation in [4] corresponds to $\lambda_{intra} = 0$ and selects intra mode only in case of reduction of prediction residual energy.

2.3 Conveying mode selection

The use of a separate field to convey inter/intra for every block is avoided as follows: For the scheme which does not use intraframe prediction at all, the code-tree shown in Fig. 5 (a) is used for both MV-x and MV-y. For every block, MV-x is immediately followed by MV-y. However, for the scheme which uses intraframe prediction, the code-tree shown in Fig. 5 (b) is used for MV-x. The code-tree for MV-y is not modified. The symbol "Intra" indicates that the block is intraframe predicted. The Huffman code, instead of MV-y, for the particular intra mode follows this. Compared to the scheme which uses only interframe prediction, we incur a penalty of 1 bit on a block only if MV-x=0, which is more efficient than spending 1 bit on every block. The "Intra" symbol can be inserted into any other branch with symbols less likely than "0" but this would also lengthen the "Intra" code. The probabilities of these two symbols were found to be comparable for most combinations of λ_{inter} and λ_{intra}, implying that the code-tree shown here is the optimal choice.

2.4 Interframe Update Step

During interframe update, linear operations are performed on certain pixels of the highpass frame, i.e. the residual of the interframe predicted blocks. At the encoder, interframe update is performed first because some of the pixels of the highpass frame

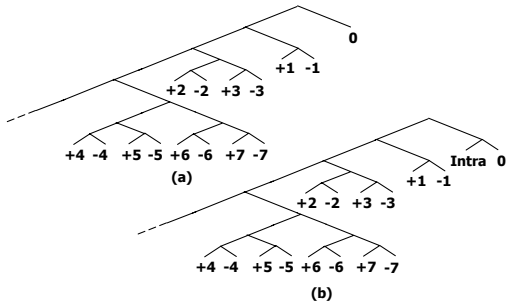

Fig. 5. (a) Exponential Golomb code tree for each MV component when intraframe prediction is turned off. (b) Exponential Golomb code tree for x-component of MV when intraframe prediction is turned on.

will be modified during intraframe update. Conventional update with sub-pel MVs involves interpolation of highpass coefficients but the intraframe predicted locations should not participate in interframe update at all. It might not be possible to invert sub-pel motion neighboring intraframe predicted locations. This problem is solved by using modified Barbell update [7] which is an inversion based on pixel-connections. With more intraframe predicted blocks there are less pixel locations in the lowpass frame receiving the interframe update. This makes the lowpass frame closer to a natural image. This results in better display at low frame-rates and improves compression of the low-band with a standard image coder. In addition, the motion between these lowpass frames can be tracked better during further temporal decomposition.

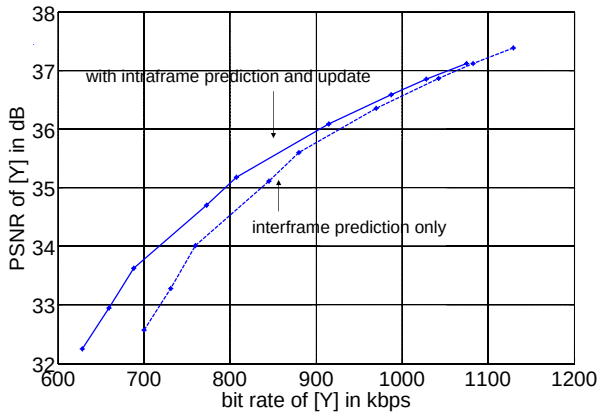

Fig. 6. R-D performance for 288 frames of *Foreman* (CIF) sequence with 3 levels of temporal Haar wavelet.

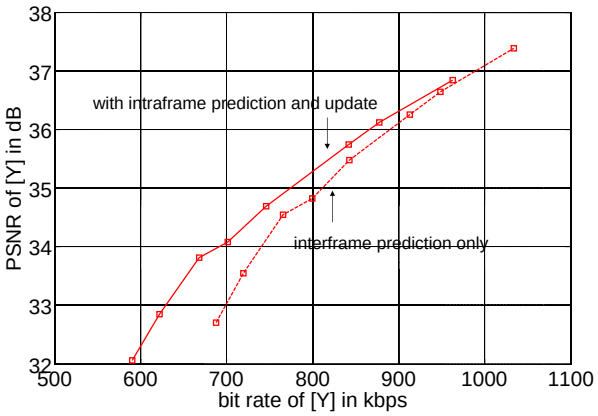

Fig. 7. R-D performance for 288 frames of *Foreman* (CIF) sequence with 4 levels of temporal Haar wavelet.

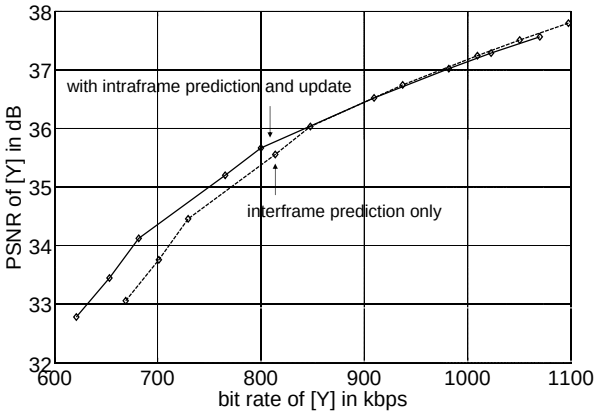

Fig. 8. R-D performance for 288 frames of *Foreman* (CIF) sequence with 3 levels of temporal 5/3 wavelet.

3 EXPERIMENTAL RESULTS

Figures 6, 7 and 8 show the R-D performance of various temporal wavelets with and without intraframe prediction for the *Foreman* (CIF) sequence. The mean PSNR with the inclusion of the intraframe update step is almost the same as the mean PSNR with intraframe prediction and no corresponding update step. Hence the performance with its inclusion is directly shown. Figures 9 and 10 show the R-D performance for the *Mobile and Calendar* (CIF) sequence. The gains at low bit-rates are substantial (∼1.5 dB). The gains increase as we perform more levels of temporal wavelet transform. This is because of the increased temporal distance and the increase in the number of MV fields per unit time. The implementation of the 5/3 temporal wavelet involves boundary extension for enhanced performance [8]. For any

curve shown with intraframe prediction, at the most two values of λ_{intra} are used, a larger value at low bit-rates and a smaller value at high bit-rates.

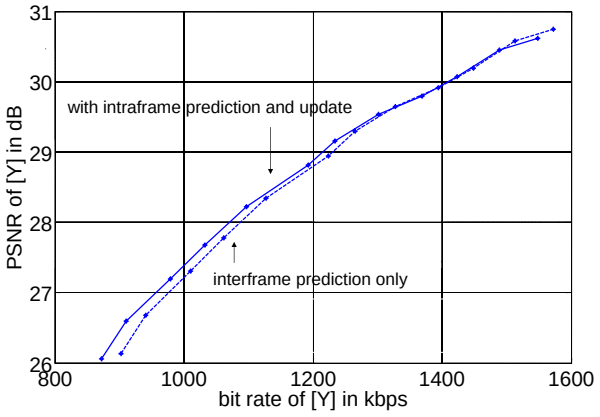

Fig. 9. R-D performance for 96 frames of *Mobile and Calendar* (CIF) sequence with 3 levels of temporal Haar wavelet.

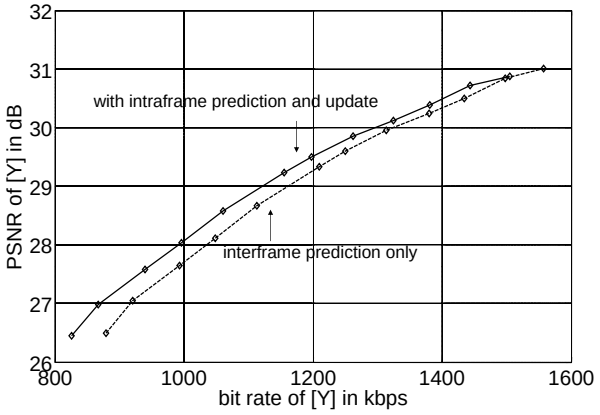

Fig. 10. R-D performance for 96 frames of *Mobile and Calendar* (CIF) sequence with 3 levels of temporal 5/3 wavelet.

4 CONCLUSION

The incorporation of intraframe prediction with the corresponding intraframe update step into the lifting framework is presented. The proposed sequence of operations can be perfectly inverted at the decoder to achieve perfect reconstruction when no quantization is performed. The Lagrangian mode decision helps to achieve gains in the R-D performance through better prediction and/or reduction of MV bit-rate. In this scheme, the proposed intraframe update step lessens the severity of quantization error propagation in the reconstruction of intraframe predicted blocks.

References

1. Ohm, J.R.: Motion-compensated wavelet lifting filters with flexible adaptation. Proceedings of the Intl. Workshop on Digital Communications (IWDC), Capri, Italy (2002) 113 –120
2. Taubman, D., Marcellin, M.W.: JPEG2000: Image Compression Fundamentals, Standards and Practice. (Kluwer International Series in Engineering and Computer Science, 642), Kluwer Academic Publishers (2001)
3. Schwarz, H., Marpe, D., Wiegand, T.: MCTF and scalable extension of H.264/AVC. Proceedings of the IEEE Intl. Picture Coding Symposium (PCS), San Francisco, U.S.A. (2004)
4. Wu, Y., Woods, J.W.: Directional spatial I-blocks for the MC-EZBC video coder. Proceedings of the IEEE Intl. Conference on Acoustics, Speech and Signal Processing (ICASSP), Montreal, Canada **3** (2004) 129 –132
5. Final Draft International Standard: Information Technology- Coding of Audio-Visual Objects- Part 10: Advanced Video Coding. ISO/IEC FDIS 14 496-10 (2003)
6. Girod, B., Han, S.E.: Optimal update step for motion-compensated lifting. IEEE Signal Processing Letters **12** (2005) 150 –153
7. Mavlankar, A.A., Han, S.E., Chang, C.L., Girod, B.: A new update step for reduction of PSNR fluctuations in motion-compensated lifted wavelet video coding. Proceedings of the IEEE Intl. Workshop on Multimedia Signal Processing (MMSP), Shanghai, China (2005)
8. Xu, J., Li, S., Xiong, Z., Zhang, Y.Q.: On boundary effects in 3-D wavelet video coding. Proceedings of the SPIE Intl. Symposium on Optical Science and Technology, San Diego, CA, U.S.A. (2000)