ADAPTIVE PREDICT BASED ON FADING COMPENSATION FOR LIFTING-BAED MOTION COMPENSATED TEMPORAL FILTERING

Li Song[1], Hongkai Xiong[1], JiZhen Xu[2], Feng Wu[2], Hui Su[1]

[1] Institute of Image Communication, Shanghai JiaoTong University, 200030, Shanghai
[2] Microsoft Research Asia, 100080, Beijing

ABSTRACT

Lifting implementation of the discrete wavelet transform applied along motion trajectories has recently gained a lot of attention in video community as strong candidates to incoming scalable video coders. In this paper, we generalize the coding scheme for classical lifting-based motion compensation temporal filtering and permit codec to adaptively choose between original reference frames and new fading-compensated reference frames to predict residuals while maintaining the invertibility of the inter-frame transform. Experimental results show that the proposed algorithm not only significantly improves subjective visual quality of the temporal low-pass frames but also have the 0.15~0.3db gain in PSNR performance compared with the normal (5, 3) lifting schemes.

1. INTRODUCTION

The lifting structure of wavelet transform is widely employed in the temporal decomposition of the 3D sub-band video coding because it is well-known that perfect reconstruction is an inherent property of the lifting structure. With such a power tool, the coding performance of motion compensated temporal filtering (MCTF) is improved significantly. The recent results have shown that the MCTF scheme with scalability support can achieve comparable coding performance with the state-of-the-art non-scalable H.264 standard, and even sometimes it outperforms H.264 [1][2].

The predict step in lifting-based MCTF scheme is very similar to motion estimation/compensation in classical hybrid coder, which usually plays a key role for improving coding performance, since it greatly decrease the temporal correlation and only few residual values remains to be coded later. Except block-based motion models that are extensively used in the MCTF scheme, many advance techniques such as overlapped block motion compensation, deformable mesh motion model have been investigated further to reduce residual in the high-pass frames[1][2]. Although these techniques improve the precision of motions estimation, all of them work under the hypothesis of constant illumination in scene. However, such hypothesis seldom holds in real image sequences. For example, if every luminance value in the current frame has changed relative to the previous frame, the motion-compensated algorithm will suffer from finding a good predictor in that frame. The absence of good predictors implies that the entire frame will be encoded as an intra-frame if codec support intra coding. Thus, fading causes a significant loss in compression efficiency. For MCTF, it also increases artifacts of low-pass frames if update steps is involved [3]. In addition, natural illumination changes, and artificial transitioning effects such as blending, cross-fades, and morphing also reduce the effectiveness of straightforward motion compensation.

In this paper, we propose an adaptive predict scheme, which integrates global illumination changes model into lifting-based MCTF scheme. By estimating the fading parameters based on the current video frame relative to the original reference video frames, our new motion estimation adaptively refer to either the original reference video frames or the new fading-compensated reference frames. The fading parameters and maps of reference frame are signaled to the decoder. This fading compensation process improves the overall performance of lifting-based MCTF on sequences with fading, or other global illumination changes.

The rest of this paper is organized as follows. Section 2 overviews the motion compensated lifting implementation with the (5, 3) filter. In section 3, we investigate the impact of global illumination changes on the predict step in video sequences and propose our adaptive predict scheme based on fading compensation scheme. Experimental results are given in section 4. Finally, Section 5 concludes this paper.

2. LIFTING BASED MOTION COMPENSATED TEMPORAL DECOMPOSTION

This work has been done while the author is with Microsoft Research Asia.

This section overviews the lifting-based motion compensated temporal decomposition with the (5, 3) filter. Assume that a video sequence, $\boldsymbol{I_0, I_1,\ldots, I_{2n-1}}$ are to be processed with temporal transform. *Figure 1* shows how the lifting based temporal transform is performed with the bi-orthogonal 5/3 wavelet filter.

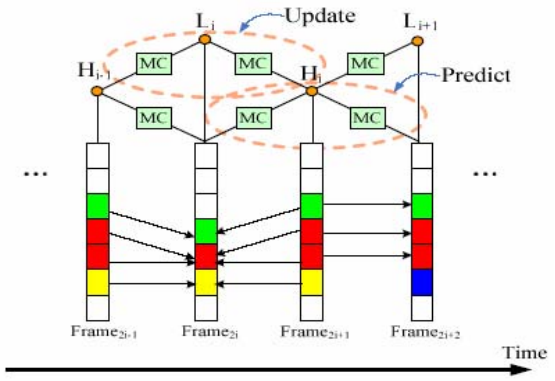


Figure 1: One level temporal transform in lifting based motion compensated video coding

The first step is prediction to calculate the high-pass frame, which predicts the odd frame from consecutive even frames as follows,

$$H_i = I_{2i+1} - P(I_{2i}, I_{2i+2}) \qquad (1)$$

where

$$P(I_{2i}, I_{2i+2}) = \frac{1}{2}(MC(I_{2i}, MV_{2i+1->2i}) + MC(I_{2i+2}, MV_{2i+1->2i+2}))$$

H_i is the high-pass frame generated in the predict step. $MV_{2i+1->2i}$ mean motion vectors from the frame $2i+1$ to the frame $2i$. So do $MV_{2i+1->2i+2}$. And MC () means motion compensation process that generates the current frame's prediction from its consecutive frame.

The update step follows the predict step to complete one level 5/3 sub-band transform which generates the low-pass frame

$$L_i = I_{2i} + U(H_{i-1}, H_i) \qquad (2)$$

where

$$U(H_{i-1}, H_i) = \frac{1}{4}((MC(H_{i-1}, MV_{2i->2i-1}) + MC(H_i, MV_{2i->2i+1}))$$

Since the predict step attempts to minimize the bit-rate required to encode the high-pass frame along with motion vectors used for prediction, H_i is essentially the residue from bi-directional motion compensated prediction of the relevant odd indexed input video frames I_{2i+1}. Then the "original" even indexed frame I_{2i} is updated with the predicted residues as the low-pass frame. Therefore, the precision of motion prediction is direct related with coding efficiency. If the motion prediction is inaccurate, it would not only increase the energy of high frequency but also introduce ghosts to low-pass frame. Obviously, it has a negative effect both on the coding performance and on subjective quality.

For nature image sequences, every video frame encodes two types of more or less distinct information about a scene. One is geometric cues which is based on the projection geometry, constrain the position of points in the scene in terms of the coordinates of their projections onto the image plane. Another one is radiometric cues which are tied to a large number of scene properties, including illumination condition, medium properties, sensor spectral response characteristics, as well as shape, position, and reflectance characteristics of the scene surfaces. Regrettably, the current standard motion-compensation techniques partially omit the second type of information by assuming there are nearly no global illumination changes in one shot. However, many successive video frames show artificial transitioning effects such as fade-to-black, fade-from-black and dissolves or natural luminance change even for static scene. In this situation, standard MCTF techniques are ineffective and require relatively large amounts of bits to encode.

To deal with this problem, the state-of-art H.264/AVC introduces weighted prediction to allow explicit or implicit control of the relative contributions of reference picture to the motion compensation prediction process [4]. However, it does not specify the method of illumination change detect and estimation of weighting values and this tool works in the unit of slice. In the next section, we discuss a novel block-based adaptive predict method to improve the coding performance on video frames with global illumination changes.

3. ADPATIVE PREDICT BASED ON FADING COMPENSATION

As discussed in Section 2, the classical predict step may have a negative effect in terms of PSNR and visual quality when omitting effect of illumination changes. To solve the problem, a better predict scheme should compensate such negative effect. In classical (5,3) lifting based MCTF, motion estimation is firstly performed on each macroblock of odd picture to obtain its forward and backward motion parameters. Our new scheme extends the number of reference frames by adding two new reference frames which are compensated with global illumination changes as show in *Figure* 2. The two fading compensated reference frames (FCRFs) are estimated according to the current frame and original two reference frames.

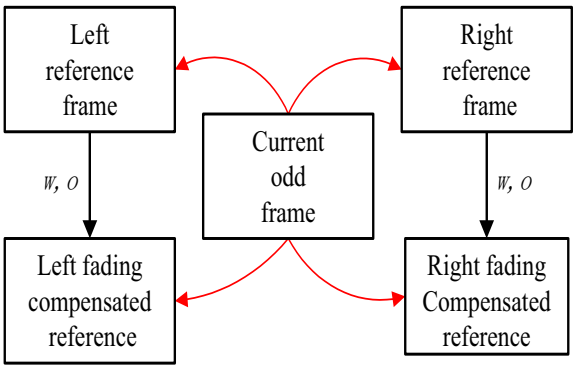


Figure 2: motion estimation with fading compensation.

Although there are complicated illumination change models, we limited illumination change model in this paper to linear first order [5]. It consists of a multiplier and an offset field:

$$I(\mathbf{r}+\delta\mathbf{r}) = w(\mathbf{r})I(\mathbf{r}) + o(\mathbf{r}) \qquad (3)$$

where $\mathbf{r} = [x, y, t]^T$ denotes a space-time 3D vector, $w(\boldsymbol{r})$ and $o(\boldsymbol{r})$ are the weighting factor and the additive offset related with every point $\boldsymbol{r}$. It is certainly true that illumination changes between two images can be modelled pixel-wisely according to (3). However, it is difficulty to discriminate different physical causes of illumination changes. Therefore, we suppose that $w(\boldsymbol{r})$ and $o(\boldsymbol{r})$ to be const value for every frame, which is similar to H.264/AVC:

$$I_i^c = w\, I_i + o \qquad (4)$$

where I_i^c means FCRF of the original frame I_i and w, o are parameters for whole i-index frame. For fades that are uniformly applied across the entire picture, a single weighting factor and offset are sufficient to efficiently code all macroblocks in a picture that are predicted from the same reference picture. For fades that are non-uniformly applied spatially across an image, the intuitive solution is different macroblocks in the same picture use different weighting factors, however , it will add more bit rate of these side information. To solve this dilemma, we adaptively switch the reference frames in macroblock level by a model indicator map.

Based on the above analysis, the first thing we need to do is to estimate the parameters in equation (4). By replacing current frame and the original reference frame into left sides and right sides of equation (4), we use the least square method to get the estimation of w and o. Then we can get the new reference frame according to the parameter w and o. Now we will have four reference frames at hand before motion estimation is performed on each macroblock.

During block-based motion estimation, we search the best motion parameters of every macroblock in all of four combination pairs: (I_L, I_R), (I_L, I_R^c), (I_L^c, I_R) and (I_L^c, I_R^c). Adaptive block-size motion alignment (ABSMA) and correlated motion estimation (CME) modes with rate-distortion optimization proposed in the previous works [1] is used in macroblock layer. The best pair and motion parameters are chosen as reference macroblock. At same time, the map for reference frames is recorded for usage later, where a variable *mode* encodes reference frame pairs: 0 is (I_L, I_R), 1 is (I_L, I_R^c), and 2 is (I_L^c, I_R) and 3 is (I_L^c, I_R^c).

Therefore, the proposed adaptive predict step can be generalized as follows:

$$H_i = I_{2i+1} - P_{new}(I_{2i}, I_{2i+2}) \qquad (5)$$

P_{new} () means the proposed adaptive predict steps. Different from methods which change the weighting value of (5, 3) filter coefficients, our adaptive predict acts on the both original reference video frames and illumination compensated frames while keeping the standard (5, 3) filter coefficients untouched. Specifically, P_{new} is defined as follows:

$$P_{new} = \begin{cases} P(I_{2i}, I_{2i+2}) & \text{mode} = 0 \\ P(I_{2i}, I_{2i+2}^c) & \text{mode} = 1 \\ P(I_{2i}^c, I_{2i+2}) & \text{mode} = 2 \\ P(I_{2i}^c, I_{2i+2}^c) & \text{mode} = 3 \end{cases} \qquad (6)$$

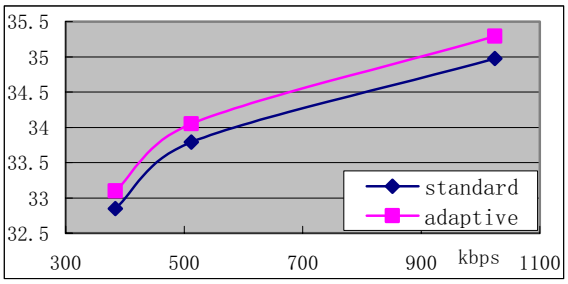
Where $P()$ has same definition as equation (1). Additional side information is not significant since only four parameters and a sparse map should be encoded and sent to decoder only for every high-pass frame, which can be reduced further by taking effective variable length coding. In decoder, the same map is decoded and indicates which reference video frames are used to motion compensation and are figured out by according parameters.

4. EXPERIMENTAL RESULTS

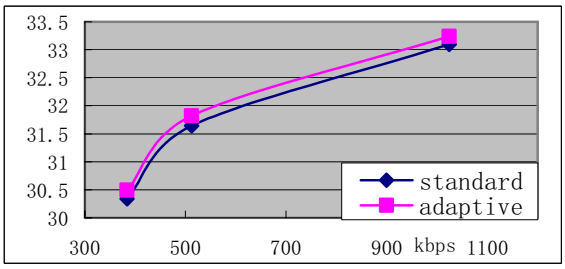
We have conducted extensive experiments to test the performance of our proposed predict steps. The motion threading (5,3) lifting-based MCTF scheme in [1] is selected as the test benchmark in this paper. Each sequence is temporally de-composed into four-layer and each temporal frame is further spatially decomposed by spatial wavelet transform. The resulted wavelet coefficients are coded and truncated to the target bit rate. The map unit is same as size of macroblock, 16x16.

In order to demonstrate the improvement of our techniques, comparison is done with the classical predict scheme. The coding performance comparison of classical predict steps and the proposed adaptive scheme is depicted in *Figure 3*(*crew-D1 and football-CIF*) at different bit rates. It shows that the proposed method has

average 0.3db gain for *crew* sequences and 0.15db gain for *football* sequences. *Figure 4* shows the improved quality of the reconstructed video frame #55 at 512kbps and at half frame rate (15 frames) for the *crew* sequence. Obviously the adaptive predict steps in the proposed scheme work better in improving the subject quality. The fading detection is implemented by comparing the DC of current and reference frame, only the difference surpasses the certain threshold, our scheme is turned on. In these frames, the complexity is nearly four times of classical one while other sequences where no obvious fading, our method keep the same performance and add little complexity.



(a) Coding performance comparison for *crew*



(b) Coding performance comparison for *football*

Figure 3: The performance evolutions of the proposed technique.

5. CONCLUSIONS

An adaptive predict based on global illumination compensation is proposed in the lifting based MCTF. This new technology effectively improves the coding performance for sequences with fading, or other global illumination changes. The experimental results validate the effectiveness of the proposed predict scheme. In this preliminary implementation, we assume the model of global illumination change the first-order linear model; we will investigate more complicated and effective illumination change models and more effective parameters estimation method to further improve the coding performance in lifting based MCTF.

6. ACKNOWLEDGMENTS

This paper was partially supported by China NSF grant No.04ZR14082.

7. REFERENCES

[1]. R. Xiong, F. Wu, S. Li, Z. Xiong and Y.-Q. Zhang, "Exploiting temporal correlation with adaptive block-size motion alignment for 3D wavelet coding", SPIE/IEEE Visual Communications and Image Processing (VCIP2004), San Jose, California, USA, Jan.2004.

[2]. A. Secker, and D. Taubman, "Highly scalable video compression using a lifting-based 3D wavelet transform with deformable mesh motion compensation", Proceedings of the IEEE Int. Conf. on Image Processing (ICIP2002), Rochester,Vol.3 ,pp.24-28, June 2002.

[3]. N. Mehrseresht, and D. Taubman, "Adaptively weighted update steps in motion compensated lifting based on scalable video compression" , Proceedings of the IEEE Int. Conf. on Image Processing (ICIP2003), Barcelona, vol.2, pp.771-774, September 2003.

[4]. JVT, "Joint Final Committee Draft (JFCD) of Joint Video Specification", document JVT-D157, Joint Video Team of ISO/IEC MPEG & ITU-T VCEG, Klagenfurt, July 2002.

[5]. S. Negahdaripour. "Revised definition of optical flow: integration of radiometric and geometric clues for dynamic scene analysis", IEEE Trans. PAMI, vol.20, pp. 961-979, September 1998.

(a) Normal predict

(b) Adaptive predict

Figure 4: visual quality comparison for crew.