Computationally Efficient MCTF for MC-EZBC Scalable Video Coding Framework

A.K. Karunakar and M.M. Manohara Pai

Department of Information and Communication Technology,
Manipal Institute of Technology
Manipal 576 104 India
{karunakar.ak,mmm.pai}@manipal.edu

Abstract. The discrete wavelet transforms (DWTs) applied temporally under motion compensation (i.e. Motion Compensation Temporal Filtering (MCTF)) has recently become a very powerful tool in scalable video compression, especially when implemented through lifting. The major bottleneck for speed of the encoder is the computational complexity of the bidirectional motion estimation in MCTF. This paper proposes a novel predictive technique to reduce the computational complexity of MCTF. In the proposed technique the temporal filtering is done without motion compensation. The resultant high frequency frames are used to predict the blocks under motion. Motion estimation is carried out only for the predicted blocks under motion. This significantly reduces the number of blocks that undergoes motion estimation and hence the computationally complexity of MCTF is reduced by 44% to 92% over variety of standard test sequences without compromising the quality of the decoded video. The proposed algorithm is implemented in MC-EZBC, a 3D-subband scalable video coding system.

Keywords: Motion Estimation, Motion Compensated Temporal Filtering, Temporal Filtering, MC-EZBC.

1 Introduction

The Scalable Video Coding (SVC) is one of the most important features of modern video communication system. For a truly scalable coding, the encoder needs to operate independently from the decoder, while in predictive schemes the encoder has to keep track and use certain information from the decoder's side (typically, target bit-rate), in order to operate properly.

The 3D sub-band video coding has appeared recently as a promising alternative to hybrid DPCM video coding techniques; it provides high energy compaction, scalable bit-stream for network and user adaptation and resilience to transmission errors. While early attempts to apply separable 3D wavelet transform directly to the video data didn't produce high coding gains, it was soon realized that, in order to fully exploit inter-frame redundancy, the temporal part of the transform must compensate for motion between frames. In one of the first attempts to incorporate motion into 3D wavelet video coding, Taubman and

A. Ghosh, R.K. De, and S.K. Pal (Eds.): PReMI 2007, LNCS 4815, pp. 666–673, 2007.

Zakhor [1] pre-distorted the input video sequence by translating frames relative to one another before the wavelet transform so as to compensate for camera pan. Wang et. al. [2] used mosaicing to warp each video frame into a common coordinate system and applied a shape-adaptive 3D wavelet transform on the warped video. Both of these schemes adopt a global motion model that is inadequate for enhancing the temporal correlation among video frames in many sequences with local motion. To overcome this limitation, Ohm [3] proposed local block-based motion, similar to that used in standard video coders, while paying special attention to covered/uncovered and "connected/unconnected" regions. Failing to achieve perfect reconstruction with motion alignment at 1/2-pixel resolution, Ohm's scheme showed no significant performance improvement. Only recently it has been generalized to sub-pixel accuracies. This paper proposes a technique that will apply motion estimation only to those blocks which has undergone some motion and hence increase the speed of the scalable encoder significantly.

The rest of the paper is organized as follows. In section II, MCTF is explained. The section III discusses the proposed technique. Section IV and V is simulation results and conclusion respectively.

2 Motion Compensated Temporal Filtering

The idea of using motion-compensated temporal DWT (MCTF) was introduced by Ohm [3] and developed by Choi and Woods [4]. A motion compensated lifting framework for TDWT is proposed by several researchers [5],[6],[7],[8],[9],[10]. Temporal decomposition using any desired motion model and any desired wavelet kernel with finite support is possible using lifting framework for MC TDWT. The results reported in [11,12]indicate superior performance with the bi-orthogonal 5/3 wavelet kernel, compared to conventional Haar wavelet transform. As discussed in LIMAT framework [10], MC TDWT is accomplished through a sequence of temporal lifting steps and the motion compensation is performed inside each lifting steps. Let $\mathrm{M}_{k1\to k2}(f_{k1})$ denote a motion-compensated mapping of frame k1 onto the coordinate system of frame k2. Using this notation we can implement motion compensated lifting steps for 5/3 analysis as below.

$$h_k[x] = f_{2k+1}[x] - \frac{1}{2}(f_{2k}[M_{2k\to 2k+1}(x)] + f_{2k+2}[M_{2k+2\to 2k+1}(x)]) \tag{1}$$

$$l_k[x] = f_{2k}[x] + \frac{1}{4}(h_{k-1}[M_{2k-1\to 2k}(x)] + h_k[M_{2k+1\to 2k}(x)]) \tag{2}$$

The motion compensation of lifting steps effectively causes the temporal subband analysis filters to be applied along the motion trajectories induced by motion compensation operators, M. These temporal lifting steps are shown in Fig. 1. Equation (1) is commonly known as `prediction` step it produces the high pass temporal frames h_k, as the residual left after bi-directional motion compensation of the odd indexed frames based on the even indexed frames. In region where the motion model captures the actual motion, the energy in the high pass frames will be close to zero. Motion model failure, however causes multiple edges and increased energy in the high pass temporal frames. Equation (2) is commonly known as the `update` step. Its interpretation is not as immediate as that of the prediction step, but it servers to ensure that frame l_k corresponds to a low pass filtering of the input frame sequence along the motion trajectories using the transforms five-tap low-pass analysis filter. Regardless of the motion model used for the M operators, the temporal transform can be trivially inverted by reversing the order of the lifting steps and replacing addition with subtraction as follows.

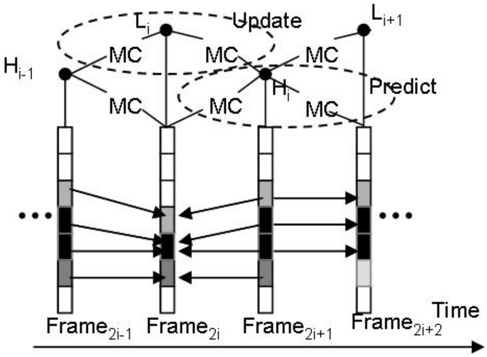


Fig. 1. One level lifting based MCTF using bi-orthogonal 5/3 filter

$$f_{2k}[x] = l_k[x] - \frac{1}{4}(h_{k-1}[M_{2k-1\to 2k}(x)] + h_k[M_{2k+1\to 2k}(x)]) \qquad (3)$$

$$f_{2k+1}[x] = h_k[x] + \frac{1}{2}(f_{2k}[M_{2k\to 2k+1}(x)] + f_{2k+2}[M_{2k+2\to 2k+1}(x)]) \qquad (4)$$

3 Computationally Efficient MCTF

The temporal decomposition of a video sequence using any wavelet filter without motion compensation causes blurriness in the region wherever there is a motion, as shown in Fig. 2(a). In order to remove the blurriness in the low frequency frames temporal filtering is done along the motion trajectory (i.e. MCTF) as shown in Fig. 2(b) and the energy in the high frequency frame is also reduced, thus supports for compression efficiency. The MCTF framework consumes dominant portion of the encoding time due to its bi-directional motion estimation (four times motion estimation is to be performed in order to decompose a pair of frames using 5/3 filters)[10], in which the motion estimation is performed for all the blocks in the frame irrespective of motion presence.

In slow motion videos (e.g. Akiyo, News, etc.,) and the video with fixed background (e.g. Claire) most of the blocks in a frame remains stationary. These observations on several test video sequences motivates us for proposing a novel technique to identify zero motion blocks (ZMB), without actually carrying out the motion estimation and hence contributes to reduce the computationally complexity of MCTF. The temporal filtering is applied along the corresponding blocks of the frames without motion compensation.

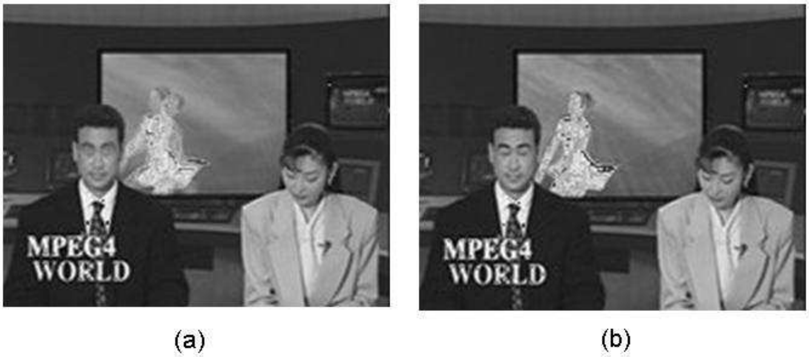


Fig. 2. (a)Blurred image with out MCTF Clear image after MCTF

In a region where there is no motion and pixel values also remain same, the temporal filtering results in zero energy in high frequency frame. If there is no motion, the pixel value does not vary much due to inherent nature of video, thus there will be very less energy in high frequency frame.

In the proposed technique the "predict" step of lifting is applied to the input frames without any motion compensation and the high frequency frame is obtained (i.e. residual energy). During this step sum of the pixel values for each block in high frequency frame is computed. The zero motion blocks are detected using sum of the pixel values of high frequency frame. If the sum is less than 512 (in case of block size 16 X 16), that block is considered as zero motion block otherwise as a block with motion. The threshold value for motion detection is taken as 512, since there are total 256 pixels in each block and the technique empirically (Table-1) decides average value of each pixel as two when there is no motion.

Our assumption is empirically proved by applying the assumption on all classes of test videos and the results is shown in Fig. 3, Fig. 4, Fig. 5 and Fig. 6. In all the test videos actual number of zero motion blocks and the estimated approximate number of zero motion blocks using our assumptions are coinciding with each other on almost all the frames. Hence our assumption to detect zero motion blocks is true and gives expected results in complexity reduction of MCTF. The technique computes the sum of pixel values of a residual block. In general for zero motion blocks this sum will be less than n^2 (where n X n is the block size) and for remaining block MCTF is done as usual.

Algorithm

```
Step 1. Apply "prediction" operation of lifting wavelet transform
on the input frames and obtain residual (or high frequency)
frame using following equation.
```

$$h_k[x] = f_{2k+1}[x] - \frac{1}{2}(f_{2k}(x)] + f_{2k+2}(x)]) \tag{5}$$

```
During this process calculate the sum of the pixel values of
the individual blocks.
Step 2. For each block
  If ( SUM > n X n )// when n X n is block size
  {
  Consider that as a block with motion do motion, estimation for
that block and obtain   motion vector. Again carry out temporal
filtering along the motion vector using Equ. (1) and update the
corresponding block in the residual frame obtained in step 1.
  }
Endif
Step 3. For each block apply following "update" step of lifting to
obtain low frequency frame.
If (a block undergoes motion)
          Equ. (2)
Else
```

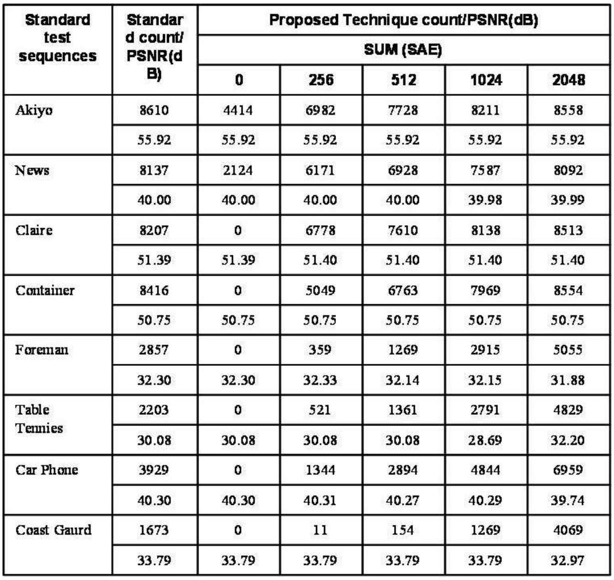
$$l_k[x] = f_{2k}[x] + \frac{1}{4}(h_{k-1}(x)] + h_k(x)]) \tag{6}$$

```
Endif
```

4 Simulation Results

During simulation of the proposed technique we have considered full search motion estimation with 1/8 pixel accuracy, 5/3 wavelet transform for temporal filtering, Debauchees 9/7 wavelet filter for spatial wavelet transform, window size 15 X 15 and block size 16 X 16. Standard test sequences like Akiyo, News, Foreman, etc., of QCIF resolution at 30 frames per second showing all varieties of motions are considered.

The Fig. (3) shows the actual blocks with zero motion(standard count) found after motion estimation and calculated number of blocks with zero motion (proposed technique count) without computing motion estimation. For various values of SAE the number of computed blocks with zero motion is shown for various types of videos. Hence we have chosen a threshold of 512 for the SAE, which compromises with complexity and quality of the decoded video.



Standard test sequences	Standar d count/ PSNR(d B)	Proposed Technique count/PSNR(dB)				
		SUM (SAE)				
		0	256	512	1024	2048
Akiyo	8610	4414	6982	7728	8211	8558
	55.92	55.92	55.92	55.92	55.92	55.92
News	8137	2124	6171	6928	7587	8092
	40.00	40.00	40.00	40.00	39.98	39.99
Claire	8207	0	6778	7610	8138	8513
	51.39	51.39	51.40	51.40	51.40	51.40
Container	8416	0	5049	6763	7969	8554
	50.75	50.75	50.75	50.75	50.75	50.75
Foreman	2857	0	359	1269	2915	5055
	32.30	32.30	32.33	32.14	32.15	31.88
Table Tennies	2203	0	521	1361	2791	4829
	30.08	30.08	30.08	30.08	28.69	32.20
Car Phone	3929	0	1344	2894	4844	6959
	40.30	40.30	40.31	40.27	40.29	39.74
Coast Gaurd	1673	0	11	154	1269	4069
	33.79	33.79	33.79	33.79	33.79	32.97

Fig. 3. The actual number of blocks with zero motion and identified blocks as zero motion blocks using proposed technique with qualtiy of the decoded video in terms of PSNR (dB)

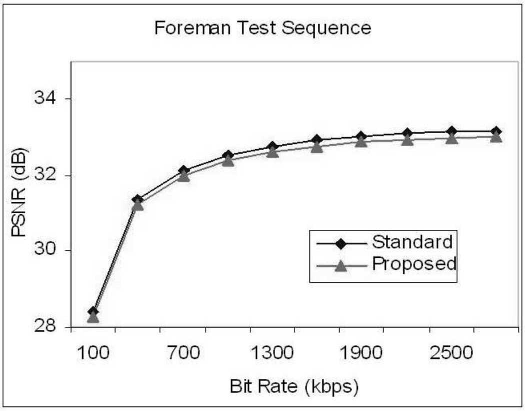


Fig. 4. Comparison of objective quality PSNR (dB) for Foreman sequence at various bit rates

The quality of the decoded video at various bit rate for standard and proposed techniques are shown in Fig. 3, Fig. 4, Fig. 5 and Fig. 6. The objective and subjective quality of the proposed technique is same as the standard techniques.

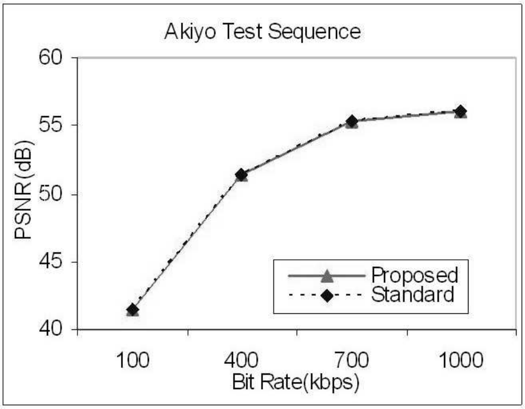


Fig. 5. Comparison of objective quality PSNR (dB) for Akiyo sequence at various bit rates

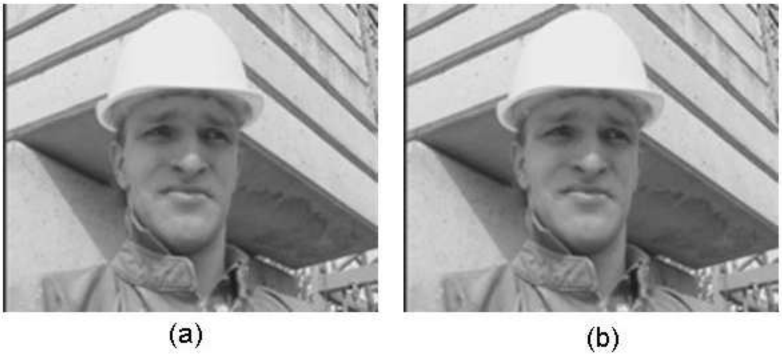


Fig. 6. Decoded 30th Foreman video frame at 2800kbps (a) Proposed Technique (b) Standard Technique

5 Conclusion

This paper proposed a novel idea to reduce the computational complexity of MCTF by effectively applying MCTF to the blocks having some motion. For the remaining blocks not having any motion, simply temporal filtering is applied. Hence unnecessary motion estimation for most of the blocks is avoided and complexity of the entire MCTF framework is reduced significantly. The results obtained from the MC-EZBC framework show that subjective and objective quality of the decoded video remains almost the same as that of the standard.

References

1. Taubman, D., Zakhor, A.: Multirate 3-D subband coding of video. IEEE Trans. Image Process. 3, 572–588 (1994)
2. Wang, A., Xiong, Z., Chou, P., Mehrotra, S.: Three-dimensional wavelet coding of video with global motion compensation. In: Proc. Data Compression Conference, pp. 404–413 (March 1999)
3. Ohm, J.: Three-dimensional subband coding with motion compensation. IEEE Trans. Image Process. 3, 559–571 (1994)
4. Choi, S., Woods, J.: Motion compensated 3d subband coding of video. IEEE Trans. Image Proc. 8, 155–167 (1999)
5. Pesquet-Popescu, B., Bottreau, V.: Three dimensional lifting schemes for motion compensated video compression. In: IEEE Int. Conf. Accoust. Speech and Signal Proc., pp. 1793–1796 (2001)
6. Bottreau, V., Benetiere, M., Felts, B., Pesquet-Popescu, B.: A fully SCalable 3d subband video codec. In: IEEE Int. Conf. Image Proc., pp. 1017–1020 (2001)
7. Luo, L., Li, J., Li, S., Zhuang, Z., Zhang, Y.-Q.: Motion compensated lifting wavelet and its application in video coding. In: IEEE, Int. Conf. on Multimedia and Expo, pp. 481–484 (2001)
8. Secker, A., Taubman, D.: Motion-compensated highly scalable video compression using an adaptive 3d wavelet transform based on lifting. In: IEEE Int. conf. Image Proc., pp. 1029–1032 (2001)
9. Secker, A., Taubman, D.: Highly scalable video compression using a lifting-based 3d wavelet transform with deformable mesh motion compensation. In: IEEE Int. conf. Image Proc., pp. 749–752 (2002)
10. Secker, A., Taubman, D.: Lifting based invertible motion adaptive transform (LIMAT) framework for highly scalable video compression. IEEE Trans. Image Proc. 12, 1530–1542 (2003)
11. Secker, A., Taubman, D.: Motion-compensated highly scalable video compression using an adaptive 3d wavelet transform based on lifting. In: IEEE Int. conf. Image Proc., pp. 1029–1032 (2001)
12. Secker, A., Taubman, D.: Highly scalable video compression using a lifting-based 3d wavelet transform with deformable mesh motion compensation. In: IEEE Int. conf. Image Proc., pp. 749–752 (2002)
13. Woods, et al.: Bi-Directional MC-EZBC with lifting implementation. IEEE Transaction of Circuits, Systems and Video Technology 14(10) (October 2004)
14. Choi, S., Woods, J.W.: Motion-compensated 3-D subband coding of video. IEEE Trans. Image Processing 8, 155–167 (1999)
15. Antonini, M., Barlaud, M., Mathieu, P., Daubechies, I.: Image coding using wavelet transform. IEEE Trans. Image Processing 1, 205–220 (1992)
16. Woods, et al.: Embedded image coding using zeroblocks of subband/wavelet coefficients and context modeling. Presented at the MPEG-4 Workshop and Exhibition at ISCAS 2000, Geneva, Switzerland (May 2000)