ENHANCED MOTION ESTIMATION FOR INTERFRAME WAVELET VIDEO CODING

*Chia-Yang Tsai, Han-Kuang Hsu, Hsiang-Cheh Huang, Hsueh-Ming Hang and Guo-Zua Wu**

National Chiao Tung University, Hsinchu, Taiwan, R.O.C.
*OES, Industrial Technology Research Institute, Hsinchu, Taiwan, R.O.C.
hchuang@mail.nctu.edu.tw, hmhang@mail.nctu.edu.tw

ABSTRACT†

An enhanced motion estimation scheme is incorporated into the interframe wavelet coding architecture in this paper. Interframe wavelet coding has the advantage of SNR, temporal, and spatial scalability, and is a potential candidate for the on-going MPEG-21 scalable video coding standard. Motion-compensated temporal filtering (MCTF) is one of its essential components. Therefore, motion estimation plays an important role in deciding the coding performance. In this paper, we modified the motion estimation syntax/scheme originally specified in the MPEG Advanced Video Coding (AVC) and use it in the interframe wavelet structure. Besides, the techniques of I-block, bi-directional motion estimation, λ-value adjustment and motion information partitioning are employed. Simulation results show very promising performance particularly on subjective quality.

1. INTRODUCTION

Video compression is an essential element in multimedia applications. Conventional video coding systems, including MPEG-1, MPEG-2, H.261 and H.263 international standards, employ the so-called *hybrid coding* structure. In these schemes, the reconstructed previous frame is used to predict the current frame after motion compensation.

Different from the aforementioned schemes, Ohm proposed a motion-compensated t+2D frequency coding structure [1]. The major difference between the hybrid coding and the t+2-D coding is that in the latter case, it does not contain the closed DPCM loop. In addition, the t+2-D coding is suitable for scalable video coding. One of the successful example of this concept is the interframe wavelet video coder proposed by Woods and his co-workers [2][3][4]. This scheme is called Motion Compensated Temporal Filtering – Embedded Zero Block Coding (MCTF-EZBC or MC-EZBC). The architecture of the interframe wavelet video coder is shown in Figure 1.

In this paper, we focus on improving the motion estimation scheme in the described interframe wavelet coding structure. At the end, we compare the performance between our proposed scheme and that in [3], and show the effectiveness of the proposed scheme.‡

This paper is organized as follows. In Section 2, we outline the motion estimation syntax/scheme in AVC. In Section 3, we incorporate the motion estimation scheme described in Section 2 into the interframe wavelet coding structure. Various techniques have been used to improve the subjective image quality such as I-block, bi-directional motion estimation and λ-value adjustment. Motion information partitioning is described in Section 4. Simulation results are shown in Section 5, in which we compare the results with the existing scheme.

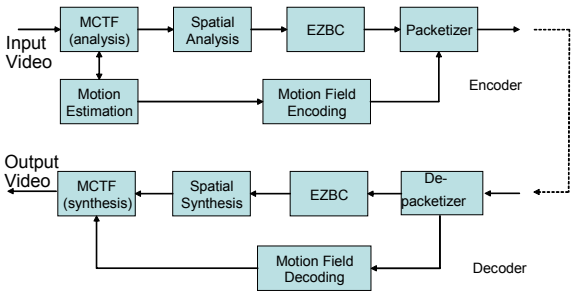

Figure 1. The interframe wavelet video coder.

2. MOTION ESTIMATION IN AVC

Motion estimation plays an important role in interframe wavelet coding. In this paper, we replace the hierarchical motion estimation algorithm in [3] by a modified version of the motion estimation scheme in the advanced video coding (AVC) standard [5].

† This work is partially supported by National Science Council (Taiwan, ROC) under Grant NSC 92-2219-E-009-008 and OES, Industrial Technology Research Institute (Taiwan, ROC) under Grant C92144.

‡ We would like to thank Woods and Chen for providing us the source codes of their MC-EZBC algorithm. The experiments in this paper are based on our modified version of these codes.

0-7803-8554-3/04/$20.00 ©2004 IEEE.

The motivation is to improve the motion compensated filter in [3]. This is because in the interframe wavelet coding structure, the error between the original and reconstruction frames accumulate due to inaccurate motion estimation. In addition, the motion-compensated temporal filtered frames are reference pictures in temporal scalability. They are decoded pictures shown in the temporally down-sampled playback.

The motion estimation scheme in AVC has three main parts: (i) tree structured motion compensation, (ii) sub-pel motion vectors, and (iii) motion vector prediction. They are outlined below.

2.1. Tree structured motion compensation

The basic unit in AVC motion estimation is the 16×16 macroblock structure. The luminance part of each macroblock can be divided into four types of sub-macroblocks, namely, 16×16, 16×8, 8×16, and 8×8.Besides, the 8×8 sub-macroblocks can further be partitioned into 8×8, 8×4, 4×8, and 4×4 blocks.

2.2. Sub-pel motion vectors

After completing motion search, the border of the reference picture is used for padding, and the full-pel motion estimation finds the best-matched motion vector and the mode with the least cost. After the full-pel search, the interpolated picture is used for $\frac{1}{2}$-pel and $\frac{1}{4}$-pel motion search.

2.3. Motion vector prediction

The motion vector of one block is highly correlated with those of its neighboring blocks. This phenomenon becomes more apparent when the block sizes get smaller. Thus, we can make use of the left, upper-left, upper, and upper-right blocks to reduce the correlation among near-by motion vectors.

3. ENHANCED MOTION ESTIMATION FOR MCTF

We adopt the afore-mentioned motion estimation scheme for the MCTF component [6][7] in MC-EZBC [3]. Temporal subband decomposition is achieved by applying high-pass and low-pass filtering along the temporal axis. Motion compensated techniques are necessary to produce better compression performance by effectively removing the temporal redundancy.

3.1 MCTF structure

The MC-EZBC coder processes one group of picture (GOP) at a time. Each GOP contains 2^n frames, where n equals to the levels of temporal subband decompositions in one GOP. The temporal subband decomposition process is performed by first constructing the motion vector map between two consecutive frames, and then the motion compensated temporal filtering (MCTF) is applied to these two frames to generate the temporal high- and low-pass frames. The temporal low-pass frames are grouped as another sub-set of GOP, and these frames are further temporally decomposed again. Decomposition process, illustrated in Figure 2 [8], is iterated until there is only one temporal low-pass frame, and a temporal filtering pyramid is thus constructed.

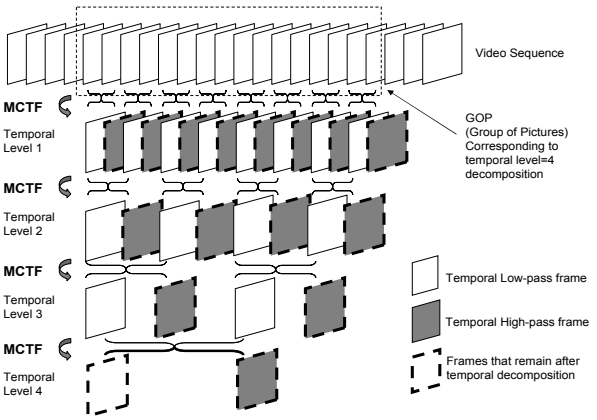

Figure 2. Temporal filtering pyramid

3.2 Lifting scheme temporal filtering

The temporal filtering operation in interframe wavelet coding is the so-called lifting scheme [9], which can achieve perfect reconstruction even when sub-pel motion estimation is used. Before temporal filtering, we find the connection relationship of pixels between the reference and the predicted frames. Then we follow the motion trajectory to generate temporal low-pass ($L[m,n]$) and high-pass ($H[m,n]$) frames. Figure 3 [3] shows the state of each pixel defined in MCTF as specified by Eqs.(1)-(5).

After the detection of the connection state of each pixel, Eq. (1) is used to generate the high-pass frame and Eq. (2) to generate the low-pass frame for connected pixels, and Eq. (3) is used for unconnected pixels [4].

$$H[m,n]=\left(A[m,n]-\widetilde{B}[m-d_m,n-d_n]\right)/\sqrt{2} \quad (1)$$

$$L[m,n]=\widetilde{H}[m+d_m,n+d_n]+\sqrt{2}B[m,n] \quad (2)$$

$$L[m,n]=\sqrt{2}B[m,n] \quad (3)$$

At the decoder, we can do the same interpolation on *H* and reconstruct *A* exactly by using Eq. (4) for connected pixels and the inverse process in Eq. (3) for unconnected ones if there is no quantization error.

$$B[m_m,n]=\left(L[m,n_n]-\widetilde{H}[m+d_m,n+d_n]\right)/\sqrt{2} \quad (4)$$

Then, *A* can be reconstructed by Eq. (5).

$$A[m,n]=\sqrt{2}H[m,n]+\widetilde{B}[m-d_m,n-d_n] \quad (5)$$

The interpolator we use is the one for generating $\frac{1}{4}$-pel accuracy reference frame in AVC motion estimation.

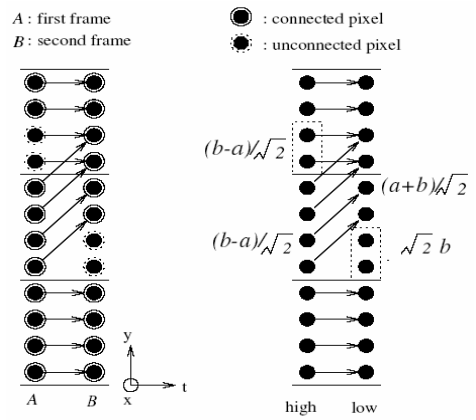

Figure 3. State of connection of each pixel

3.3 I-block and bi-directional motion estimation

The concepts of I-block and bi-directional motion estimation for MCTF were described in [3]. We adopted these concepts with some modifications in our MCTF scheme.

The temporal low-pass frame is generated by Eqs. (2) and (3) based on the state of connection of each pixel. Typically, motion compensation works well on the connected pixels. However, it is possible to have connected blocks with a poor match after motion estimation. These blocks tend to produce artifacts in the temporal low-pass frame, which lead to poor visual quality for temporal scalability. These blocks are forced to be unconnected as proposed in [3][4].

Our I-block size is 16x16. As shown in Figure 3, let $A[m,n]$ be the block with connected state at the location (m,n) of the A frame and $B[m-d_m,n-d_n]$ be the motion-compensated block with motion vector (d_m,d_n) in the B frame. We compute the variance of these two blocks, and choose the minimum as V_{min}. If the mean squared prediction error between these two blocks is larger the threshold $F*V_{min}$, this block is declared as an unconnected block, where F is an adjusting parameter. Based on our experiments, F is taken around 0.7. Figure 4 shows the subjective improvement (left upper corner, for sample) using the I-blocks.

Furthermore, an A frame block may find a better match (motion compensation) from the previous B frame. Thus, frame A has both forward and backward motion vectors. The use of bi-directional motion estimation reduces high-pass frame magnitude and thus increases coding efficiency.

3.4 Motion cost function adjustment

The rate-distortion cost function, $J=D+\lambda R$, is used to decide the best motion vectors in the AVC motion estimation, in which D is the frame difference, and R is the estimated motion vector coding bits. However, as the temporal level increases in MCTF, the energy of temporal low-pass frame is also increased. Therefore, the λ value should be increased to maintain a constant rate-distortion relation at the higher temporal levels. Therefore, the λ value is increased by a factor of $\sqrt{2}$ for each additional temporal level.

(a) (b)

Figure 4. The 2nd temporal low-pass frame: (a) without I-block (b) with I-block.

4. MOTION INFORMATION PARTITIONING

In [8], Hang and Tsai proposed the concept of motion information scalability for MC- EZBC. In this paper, we adopted this concept with some modifications to partition the motion information generated by the AVC inter-frame-prediction in MCTF. If the required bitrate is very slow, the extractor may fail to extract the bitstream because the motion information bits are larger than the specified bits. Also, at low rates, we may want to save some bits from motion information and use these bits for wavelet coefficients to achieve acceptable quality. Therefore, we partition motion information after motion estimation according to the steps below.

Step 1: Do 16x16 block size motion search with integer-pixel accuracy. The generated motion vectors are the *base layer* motion vectors.

Step 2: Do 16x16 and 8x8 block size motion search with 1/2-pixel accuracy. The difference between these motion vectors and the base-layer is the *first enhancement layer* motion vectors.

Step 3: Do all sub-block size motion search with 1/4-pixel accuracy. The difference between these motion vectors and the base-layer plus the first-enhancement-layer is the *second enhancement layer* motion vectors.

Step 4: Encode the above three layers motion information using CABAC separately.

If the required bitrate is too small, the extractor will drop one or two enhancement layers according to the given conditions. Furthermore, if one likes to extract the spatially down-sampled bitstream, the extractor can also drop proper the enhancement layers. When the codec scalability range is small, we can reduce the enhancement layers to one to save bits in encoding motion vectors.

The proposed algorithm can provide an acceptable video quality at very low bit rates, especially for

high-motion cases. However, if not all the motion vectors are used in reconstruction, the "mismatch errors" would occur. That is, the residual image data calculated at the encoder are based on the complete set of motion vectors but only "partial" motion vectors are available at the decoder if they are truncated. This problem will be further studied.

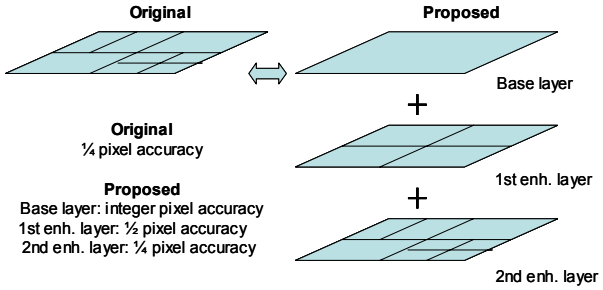

Figure 5. The base and enhancement layer motion vectors.

4. SIMULATION RESULTS

We perform two sets of simulations to show the effectiveness of the algorithm proposed in this paper. In our MCTF, we adopt both the motion vector (MV) syntax and the arithmetic coding for MV in AVC. We compare the results with those in [3]. We found improvements both objectively and subjectively but the subjective performance is more important because the final judgment of an image processing algorithm is the subjective picture quality.

In the MPEG scalable video coding call-for-proposal [10], the MPEG committee specifies three main test conditions. For the test sequence Bus_CIF, one test point in Test 2b is 30 frames per second (fps) at 512kbps [10]. The subjective quality of these two coding schemes at this test point is compared. AVC has a very complicated interframe prediction scheme, and the motion block size could be one of seven block types. Therefore, the motion estimation is very accurate. As we can see from Figure 6, the proposed coding scheme has a better subjective quality. But the PSNR value is not much changed.

6. CONCLUSION AND FUTURE WORK

In this paper, we propose an enhanced motion estimation scheme to improve the existing interframe wavelet coding algorithm (MC-EZBC). We modify the motion estimation syntax/scheme specified in AVC to fit into the motion compensated temporal filtering (MCTF) structure. Various additional techniques such as I-block, bi-directional motion estimation and λ-value adjustment are incorporated. Also, we propose the motion information partitioning technique for AVC interframe-prediction to improve coding performance at low rates. Preliminary simulation results indicate that this new motion estimation algorithm has improved the subjective image quality. Further parameter tuning should provide even better results.

(a) (b)

Figure 6. Subjective quality of the 2nd frame of Bus_CIF.yuv sequence at 512kbps with GOP=4: (a) proposed scheme, (b) MC_EZBC [3].

6. REFERENCES

[1] J.-R. Ohm, "Three-dimensional subband coding with motion compensation," *IEEE Trans. Image Processing*, vol. 3, no. 5, pp. 559–571, Sep. 1994.

[2] S.-T. Hsiang and J. W. Woods, "Embedded video coding using invertible motion compensated 3-D subband/wavelet filter bank," *Signal Processing: Image Communications*, vol. 16, pp. 705–724, May 2001.

[3] P. Chen, *Fully scalable subband/wavelet coding*, Ph.D. thesis, Rensselaer Polytechnic Institute, Troy, New York, May 2003.

[4] T. Rusert, et al., *Recent Improvements to MC-EZBC*, ISO/IEC/JTC1 SC29/WG11 doc. M9232, Dec. 2002.

[5] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 560–576, Jul. 2003.

[6] J.-R. Ohm, "Three-dimensional subband coding with motion compensation," *IEEE Trans. Image Processing*, vol. 3, no. 5, pp. 559–571, Sep. 1994.

[7] T. Kronander, "Motion compensated 3-dimensional wave-form image coding," *Int'l Conf. Acoustic, Speech, and Signal Processing*, vol. 3, pp1921–1924, 1989

[8] S. S. Tsai, *Motion information scalability for interframe wavelet video coding*, MS thesis, National Chiao Tung University, Hsinchu, Taiwan, R.O.C., Jun. 2003.

[9] B. Pesquet-Popescu, V. Bottreau, "Three-dimensional lifting schemes for motion compensated video compression," *Int'l Conf. Acoustic, Speech, and Signal Processing*, vol. 3, pp. 1793–1796, 2001.

[10] Call for proposals on scalable video coding technology, ISO/IEC JTC1/SC29/WG11 MPEG2003/N6193, Dec. 2003.