PERFORMANCE OPTIMIZATION FOR MOTION COMPENSATED 2D WAVELET VIDEO COMPRESSION TECHNIQUES

Zhen Li[1], Feng Wu[2], Shipeng Li[2], and Edward Delp[1]*

[1]Video and Image Processing Laboratory (VIPER)
School of Electrical and Computer Engineering, Purdue Univ., West Lafayette, Indiana, USA
[2]Microsoft Research Asia, Beijing, China

ABSTRACT

In this paper we present two performance optimization methods for a motion compensated (MC) 2D wavelet video coding technique, which is based on two of the current state-of-the-art codecs: H.26L TML9.4 and JPEG2000 VM7.2. First, a new metric for motion vector selection is proposed to take both edge and texture complexity into account in motion prediction. Second, a frame level rate allocation algorithm, which is an extension of JPEG2000 PCRD (Post Compression Rate Distortion) optimization, is proposed. Experimental results demonstrate the significant performance improvements by these two techniques.

1. INTRODUCTION

Motion compensated 2D wavelet video coding structures have been investigated since Shapiro's pioneer work on the embedded wavelet coding technique [1]. Due to its implementation of an embedded data stream, 2D wavelet coding is widely used to achieve rate scalability, which is identified as desirable in the latest video coding standards [2][3]recently.

Interestingly, although the idea of using MC 2D rate scalable wavelet coding [4] dates back even earlier than its DCT counterpart, the Fine Granularity Scalability (FGS) profiles in MPEG-4 [5], the use of wavelet suffers from performance and complexity issues. One of the reasons is that most DCT-domain FGS codecs have used a lot of the recent advances from current motion prediction and transform coding research, including variable block size, quarter-pixel (or even finer) motion search, 4×4 DCT coding, more complicated entropy coding and rate distortion optimizations. In this paper, we first present an MC 2D wavelet video compression technique based on H.26L TML9.4 [6] and JPEG2000 VM7.2 [7]. We call it as MC-EBCOT hereafter, where EBCOT (Embedded Block Coding with Optimized Truncation) [8] is the coding kernel adopted in JPEG2000. In this codec, we incorporate many techniques mentioned above. However, we note that even with the direct use of these latest coding techniques, the performance of this wavelet codec in non-scalable (single layer) case is still inferior to that of H.26L. Hence, we focus the rest of our paper on improving the single layer coding efficiency rather than exploring rate scalability as done in most other papers on this topic. Our argument is based on the fact that the performance of a MC 2D wavelet rate scalable video codec depends heavily on its reference quality. Hence the improvement in single layer will be fundamental for the overall performance.

The rest of this paper is organized as follows. In Section 2 we introduce the structure of MC-EBCOT. We also present an analysis for sources of potential loss in traditional MC 2D wavelet codecs here. Based on these analyses a new metric for motion vector selection is proposed in Section 3. The generalized JPEG2000 PCRD algorithm to frame level rate allocation is discussed in Section 4. The experimental results are given in Section 5. Section 6 concludes our work with a brief remark on future work.

2. MC-EBCOT

The diagram of MC-EBCOT codec is presented in Fig. 1. The motion prediction part adopts some newly developed techniques in H.26L and JVT [9], such as variable block size and quarter-pixel motion search, which contribute approximately 1dB performance gain compared to H.263+ [10]. The predictive error frame (PEF) obtained from motion prediction is sent to a JPEG2000 codec and encoded at data rate R_{enc}. The data stream for residue frames is thereafter sent with motion vectors generated during motion prediction with data rate R_{mv}. Meanwhile the data stream is decoded at a data rate R_{ref} at both the encoder and decoder. Generally $R_{mv} \le R_{ref} \le R_{enc}$ holds and R_{ref} is the base layer data rate that all decoders can guarantee to achieve. In this way both the encoder and decoder are using the same reference frame and hence avoid drifting problems. And due to the inherent embedded nature of JPEG2000 data stream, MC-EBCOT can also achieve rate scalability. However, since we are only interested in the single layer performance here, we assume $R_{ref} = R_{enc}$ throughout the rest of this paper unless otherwise specified.

* This work was performed while the author was with Microsoft Research Asia.

0-7803-7761-3/03/$17.00 ©2003 IEEE

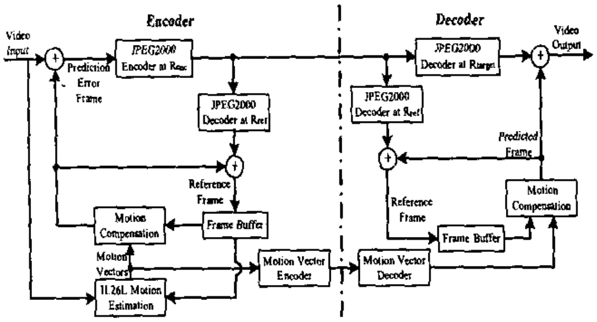

Fig. 1 The diagram of MC-EBCOT encoder and decoder

Although both H.26L and JPEG2000 are among the best available coding standards in the sense of coding efficiency, direct use of them does not immediately make MC-EBCOT comparable to H.26L in single layer case. There is typically 0.5-1 dB loss compared to H.26L as shown in the experimental results in Section 5.

By investigating the performance gain in H.26L [10], we summarize the source of potential losses in our structure:

- **Inaccurate motion vector selection and mode decision**

In H.26L, only texture complexity is considered in motion vector selection. This does not affect the performance of H.26L based on block DCT. But the lack of considering edge complexity between blocks hurts MC-EBCOT since the residue frame is encoded by JPEG2000, which is basically based on global wavelet transform coding.

- **Inefficient frame level rate allocation**

In H.26L, the rate of each residue frame is controlled by a quantization parameter QP, which roughly represents the complexity of that frame. However, in embedded wavelet codec such as JPEG2000, all coefficients are encoded to the finest level first and hence there is no such control parameter as QP. In our original MC-EBCOT, we assign constant data rate to each P frame, which is obviously not fair considering the complexity fluctuation across the sequences.

- **Intra prediction**

In H.26L it is shown in [10] that intra prediction yields approximately 0.5 dB gains on average. On the other hand, since JPEG2000 employs global transform coding and is therefore incapable of exploiting such local properties.

- **Loop filtering**

It was noted in [10] that the loop filtering contribution is 0.1-0.2 dB in H.26L. In our MC-EBCOT, we do not use loop filtering yet for the sake of simplicity.

- **Inefficient wavelet transform**

Although JPEG2000 is one of the best still image codecs, we find that JPEG2000 may not be adequate for residue frames, where a lot of edges and discontinuities exist. Hence we have developed a more efficient wavelet transform, which is described in [11].

- **In-Frame rate distortion optimization**

TML9.4 includes some sophisticated modes for R-D optimization. However, this is not critical in the comparison since the R-D optimization modes are turned off in our experiments.

We note that although the analyses above are targeted at the MC-EBCOT codec, we believe most of them apply to other MC 2D wavelet codecs due to the inherent similarity in coding structure. Hence here we present two general performance optimization techniques in this paper, as discussed in Section 3 and Section 4 respectively, to improve the single layer performance of MC-EBCOT.

3. MOTION VECTOR SELECTION

In H.26L, the following Lagrange Multiplier defines the motion vector search criterion

$$L = D + \lambda R \tag{1}$$

where R is the bit cost of a macroblock (MB) and D is the corresponding distortion. The distortion is evaluated by the SAD (sum of absolute difference) distortion

$$D = \sum_{i \in MB} | a_i - \tilde{a}_i | \tag{2}$$

where a_i and $\tilde{a}_i$ are the original and reconstructed coefficients of the macroblock. Obviously, the SAD metric in (2) approximates the complexity of texture inside each macroblock.

However, as JPEG2000 is basically a global transform, the complexity of each frame depends not only on the texture complexity $D_{texture}$ but also the edge complexity D_{edge} among blocks inside the frame. Hence we propose a new metric for distortion

$$D = D_{texture} + D_{edge} \tag{3}$$

For the texture complexity, we need to first evaluate the wavelet transform coefficients of each macroblock. We then use the weighted mean square error (wMSE), as proposed in the R-D optimization in JPEG2000, of these coefficients as $D_{texture}$. Since it can be computationally expensive to get the accurate wavelet coefficients at each motion search operation, a simplest wavelet transform, Haar transform, is used here. The wMSE distortion is then defined as

$$D_{texture} = w_{b_i}^2 \sum_{m \in B_i} (s[m] - \tilde{s}[m])^2 \tag{4}$$

where $s[m]$ and $\tilde{s}[m]$ are the original and reconstructed coefficients of the block with Haar transform. w_{b_i} denotes the L^2-norm of the wavelet basis functions for the subband b_i to which code-block B_i belongs. Meanwhile, we use the conventional edge operator to get D_{edge}, i.e.,

$$D_{edge} = \sum_{m,n \in T_i} | b[m] - b[n] | \tag{5}$$

where T_i is the boundary area of the adjacent blocks. $b[m]$ and $b[n]$ are a pair of pixels in either horizontal or vertical edge areas.

4. FRAME LEVEL PCRD RATE ALLOCATION

In this section, we first give a brief introduction to the rate-distortion optimization algorithm in JPEG2000, i.e., the Post Compression Rate Distortion (PCRD) algorithm. We then generalize the idea of PCRD to frame level rate allocation.

4.1. PCRD Algorithm in JPEG2000

The rate distortion problem in JPEG2000 is basically formulated as follows.

Given a target data rate budget $R^{\max}$, truncate each of the independent code-block data stream such that the distortion is minimized subject to $R^{\max}$.

The algorithm to solve this is referred to as PCRD, since the R-D optimization is used after all data streams have been generated. The basic idea of PCRD is to collect both the bit cost $R_i^{n_i}$ and distortion $D_i^{n_i}$ at each truncating point n_i for each code-block. Then the R-D optimization is solved with the following Lagrange Multiplier

$$L = D(\lambda) + \lambda R(\lambda) = \sum_i (D_i^{n_i}(\lambda) + \lambda R_i^{n_i}(\lambda)) \quad (6)$$

It is obvious that in this operational model there exists some optimal λ in the sense that the distortion cannot be reduced without increasing $R^{\max}$. Hence, by collecting the bit cost for each corresponding λ, PCRD can quickly find the optimal λ such that $\sum_i R_i^{n_i} \le R^{\max}$ and the particular truncating point $n_i(\lambda)$ for each code-block.

4.2. Generalized Frame Level PCRD Algorithm

We generalize the idea of PCRD to frame level rate allocation by collecting the bit cost and distortion across the whole sequence (or part of the sequence) for each particular λ. We then use one fixed λ for the whole sequences to achieve the bit budget. This fixed λ algorithm is roughly comparable to the simple rate allocation in H.26L in which QP is fixed.

However, it should be noted that in JPEG2000, each code-block is independent, hence the change of λ in one code-block does not affect the distortion of others, i.e., $D_i^{n_i}$'s are independent. However, in the frame level rate allocation, the distortion of one frame will propagate to the next frames due to the use of motion compensation, which is known as the generalized drifting problem. Currently we are investigating the distortion influence of each coding pass for referenced frames to get a more accurate model. Despite the lack of considering dependency among the frames, our experimental results show that the preliminary algorithm described above already achieves significant visual quality improvement over the constant bit rate allocation scheme.

5. EXPERIMENTAL RESULTS

This section first presents the performance of MC-EBCOT without optimizations; then we verifies the performance improvement with the new motion vector selection criterion and frame level PCRD rate allocation algorithm. Two standard test sequences, Foreman QCIF and Coastguard CIF, are used here. The encoding frame rate is 30 frames/second. Only the first frame is encoded as an I frame while the rest are all encoded as P frames. The motion search range is ±16 pixels, seven variable block sizes, including 16×16, 16×8, 8×16, 8×8, 8×4, 4×8 and 4×4, are used with quarter-pixel motion search. For H.26L, the context-based adaptive binary arithmetic coding (CABAC) mode is used while R-D optimization mode is off. In MC-EBCOT, (9,7) Daubechies wavelet kernel is used with four level of decomposition.

The performance of MC-EBCOT without optimizations is shown in Fig. 2 and Fig. 3. We see that there is generally 0.5-1dB loss in MC-EBCOT compared to H.26L.

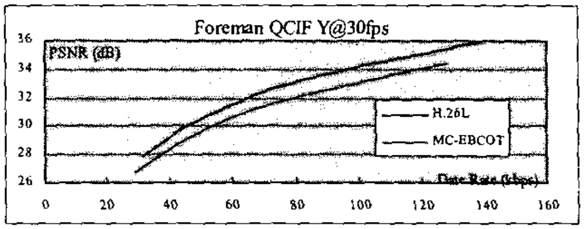

Fig. 2 Performance comparison of MC-EBCOT and H.26L.

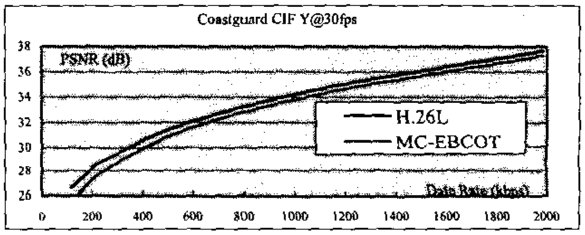

Fig. 3 Performance comparison of MC-EBCOT and H.26L.

The experimental results of the new motion vector selection criterion are shown in Fig. 4 and Fig. 5. The results indicated that the new criterion gains 0.5-0.8 over the simple SAD criterion. In addition, some sequences' performance is comparable to H.26L, as shown in Fig. 5.

The results of the new rate allocation algorithm are shown in Fig. 6 and Fig. 7. While this simple rate allocation algorithm does not contribute much coding gain on average (generally 0.1-0.2 dB), it dramatically reduces the quality variance across the sequences, which also help to improve the perceptual quality. In Fig. 6, it is found that the standard variance reduces from 0.8dB to 0.3dB. Fig. 8 presents rate allocation corresponding to Fig. 7. Not surprisingly, the new scheme introduces some rate fluctuations. We note that such small rate fluctuations can be smoothed by buffer control during packetization period and hence will not lead to big problems.

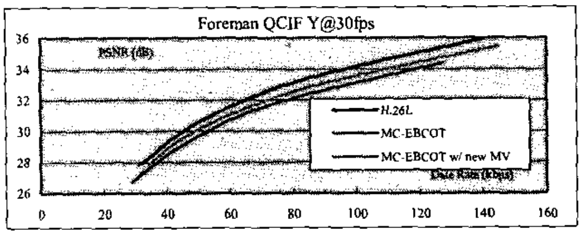

Fig. 4 Performance comparison of motion vector selection criterion.

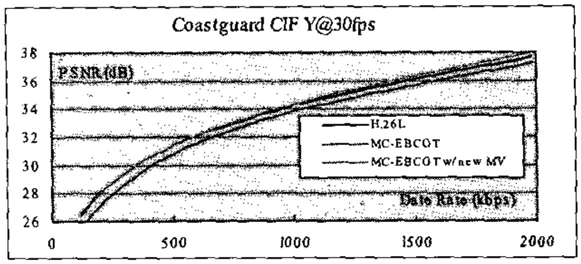

Fig. 5 Performance comparison of motion vector selection criterion.

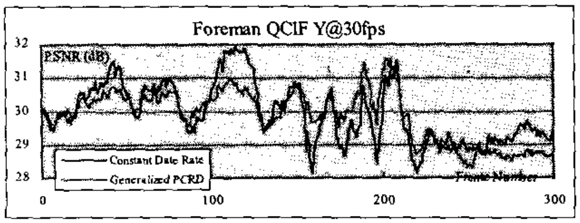

Fig. 6 Performance of new rate allocation scheme.

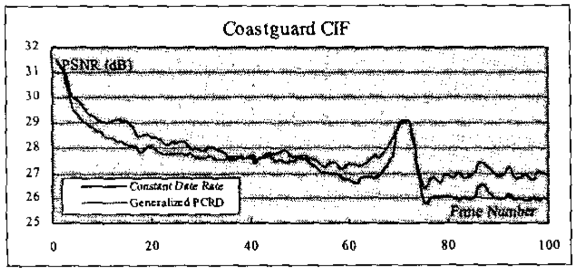

Fig. 7 Performance of new rate allocation scheme.

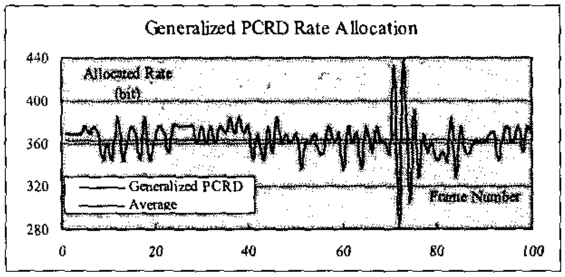

Fig. 8 Rate allocation by generalized PCRD for Coastguard CIF.

6. CONCLUSIONS AND FUTURE WORK

In this paper, we present an MC 2D wavelet video codec based on H.26L and JPEG2000. We discuss the potential performance loss for this traditional coding structure. We also develop two useful techniques, which can be used by the general structure. The experimental results confirm that our new motion vector selection and frame level rate allocation can significantly improve the visually quality both objectively and subjectively.

Currently we are developing frame level rate allocation for single layer-multiple passes wavelet codec by considering the drifting problem across subbands.

ACKNOWLEDGEMENT

The authors would like to take this opportunity to thank Mr. Jizheng Xu and other colleagues in Internet Media Group at Microsoft Research Asia for many beneficial discussions during the development of the work.

REFERENCES

[1] J. Shapiro, "Embedded Image Coding using Zerotrees of Wavelet Coefficients," *IEEE Trans. On Signal Processing*, vol.41, pp.3445-3462, Dec 1993.

[2] *Coding of Audio-Visual Objects, Part-2 Visual, Amendment 4: Streaming Video Profile*, ISO/IEC 14496-2 FPDAM4, July 2000.

[3] *JVT Pattaya 1st Meeting Draft Report*, Joint Video Team (JVT) of ISO/IEC JTC1/SC29/WG11 and ITU-T/SG16/VCEG (Q.6), Pattaya, Dec 2001.

[4] K. Shen and E. Delp, "Wavelet based Rate Scalable Video Compression," *IEEE Trans. On Circuits Syst. Video Technol.*, vol.9, pp.109-122, Feb 1999.

[5] W. Li, "Overview of Fine Granularity Scalability in MPEG-4 Video Standard," *IEEE Trans. On Circuits Syst. Video Technol.*, vol.11, pp.301-317, March 2001.

[6] *H.26L Test Model Long-Term Number 9*, ITU-T/SG16/VCEG (Q.6), VCEG N83d1, Pattaya, DEC 2001.

[7] *Coding of Still Pictures, JPEG2000 Part I, Final Committee Draft Version 1.0*, ISO/IEC JTC/SC29/WG1 FCD15444-1, March 2000.

[8] D. Taubman, "High Performance Scalable Image Compression with EBCOT," *IEEE Trans. On Image Processing*, vol. 9, pp.1158-1170, July 2000.

[9] *Text of Committee Draft of Joint Video Specification*, ITU-T Rec. H.264 | ISO/IEC 14496-10 AVC, MPEG02/N4810, Fairfax, VA, May 2002

[10] A. Joch, F. Kossentini and P. Nasiopoulos, "A Performance Analysis of ITU-T Draft H.26L Video Coding Standard", *Packet Video 2002*, Pittsburgh, PA, April 2002

[11] Z. Li, F. Wu, S. Li and E. Delp, "Video Coding via Context-based Adaptive Lifting Structure," *To be submitted to ICASSP2003*, Hong Kong, March 2003.