A Rate-Distortion Approach to Wavelet-Based Encoding of Predictive Error Frames

by E Asbun, P Salama, E Delp
Center for Education and Research
Information Assurance and Security
Purdue University, West Lafayette, IN 47907-2086

A RATE-DISTORTION APPROACH TO WAVELET-BASED ENCODING OF PREDICTIVE ERROR FRAMES

Eduardo Asbun†, Paul Salama‡, and Edward J. Delp†

† Video and Image Processing Laboratory (*VIPER*)
School of Electrical and Computer Engineering
Purdue University
West Lafayette, Indiana 47907-1285
U.S.A.

‡ Department of Electrical and Computer Engineering
Purdue School of Engineering and Technology
Indiana Univ. - Purdue Univ., Indianapolis (IUPUI)
Indianapolis, Indiana 46202-5132
U.S.A.

ABSTRACT

In this paper, we develop a framework for efficiently encoding predictive error frames (PEF) as part of a rate scalable, wavelet-based video compression algorithm. We investigate the use of rate-distortion analysis to determine the significance of coefficients in the wavelet decomposition. Based on this analysis, we allocate the bit budget assigned to a PEF to the coefficients that yield the largest reduction in distortion, while maintaining the embedded and rate scalable properties of our video compression algorithm.

1. INTRODUCTION

SAMCoW (*Scalable Adaptive Motion Compensated Wavelet*) [1, 2] is a wavelet-based, rate scalable video compression technique. Its two main features are: (i) a modified zerotree wavelet image compression scheme, known as *Color Embedded Zerotree Wavelet* (*CEZW*) [3, 4], used for coding intracoded (I) frames and predictive error frames (PEF); and (ii) adaptive block-based motion compensation [5] to reduce temporal redundancy. *CEZW* is a technique that uses a combination of a unique spatial orientation tree and color transform to exploit redundancy across color components. The use of advanced coding techniques in *SAMCoW*, including unrestricted motion vectors, half-pixel accuracy, and MPEG-like bidirectionally predicted (B) frames, was investigated in [6]. B frames provide a second dimension of scalability, that is, temporal scalability, to *SAMCoW*. In [7, 8], we presented preprocessing and postprocessing techniques to improve the performance of *SAMCoW* at low data rates. This extension to *SAMCoW* is known as *SAMCoW+*.

Motion estimation is a technique used to reduce temporal redundancy of video sequences. A frame is selected as a reference, and other temporally close frames are predicted from it. At the decoder, a set of motion vectors along with a PEF are used to reconstruct that frame based on the reference frame. Typically, PEFs have low energy content.
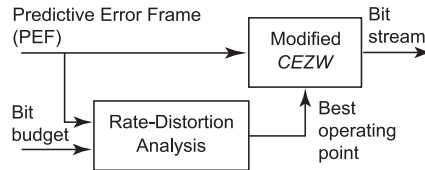

Figure 1: Block diagram of our proposed approach to encoding PEFs in *SAMCoW+*.

In this paper, we develop a framework for encoding PEFs using a modified version of *CEZW*, as shown in Figure 1. We perform a rate-distortion analysis on the wavelet decomposition of the PEF in order to determine the significance of the coefficients. We allocate the bit budget to those coefficients that yield the largest reduction in distortion, while maintaining the embedded and rate scalable properties of *SAMCoW+*. Rate-distortion analysis has been used for coding natural images [9, 10, 11]. We exploit the properties of "non natural" images, such as PEFs, to optimize its encoding using *CEZW*, although our work is not restricted to PEFs. A distinction between our work and that of [9] is that we preserve the embedded property of our algorithm. In embedded coders, the bit stream can be truncated at any point during decoding. In [10], a rate-distortion optimized embedded coder for grayscale images is presented. Our interest is to use our approach exploit the correlation between color components with *CEZW* on PEFs. Rate-distortion optimization is also used in the new JPEG2000 image compression standard [11].

2. DESCRIPTION OF THE PROBLEM

Video coding standards such as H.263+ [12] make efficient use of block-based transform coding on PEFs. Macroblocks (16×16 pixels) may be skipped in regions of the PEF where the energy content is below a threshold. Coding gains are obtained by assigning very short codewords to this SKIP mode. This strategy results in allocating more bits to areas with more activity.

In *CEZW*, a wavelet decomposition is performed on the PEF. Then, sets of coefficients from different bands are clustered in a unique spatial orientation tree (SOT) in the YUV color space. A complete description of *CEZW* is provided in [3, 4]. A special symbol, *zerotree root*, is used to indicate that all the coefficients in the SOT are below a threshold, that is, insignificant. For PEFs, the use of this special symbol is similar in concept to that of the SKIP mode in H.263+. However, a penalty is paid when representing insignificant regions that are not well aligned with the SOT. The effect is manifested as a lack of localization in the reconstruction of the PEF. The bit budget is effectively allocated to coefficients in the PEF that are significant, regardless of the contribution they make to the overall quality of the reconstruction. The resulting effect is that at low data rates, the performance of *CEZW* on PEFs can be improved.

3. FORMULATION OF THE PROBLEM

In [7], we described a modification of *CEZW* aimed at allocating the largest number of bits to the "significant trees". A tree is said to be "significant" when the root at the coarsest scale is positive or negative significant in the first dominant pass. In the remaining dominant passes, until the bit budget is exhausted, only coefficients that belong to the "significant trees" are examined.

In this paper, we develop a framework for improving the identification of the "significant trees" based on a rate-distortion analysis [13] of wavelet coefficients in the decomposition of the PEF. Our overall objective can be stated as follows: "Given a target data rate, minimize the overall distortion of the reconstruction of the coefficients in the wavelet domain, while preserving the embedded and rate scalable properties of the encoded bit stream."

In our framework, each PEF gets allocated a specific number of bits based on a rate control method. Given the target data rate allocated to each PEF, we perform the rate-distortion analysis, such that we optimize the allocation of the bit budget to the PEF. Let W be the coefficients of the wavelet decomposition of the PEF in subbands LL, LH, HL, and HH. Initially, we consider a spatial orientation tree of W as a coding unit. It is to be noted that each tree is treated as an independent coding unit, as done in EZW [14] and *CEZW* [1, 4]. Therefore, the problem can be formulated as follows:

$$\min_{\forall m_j(t_i):t_i \in T} \sum_{\forall t_i \in T} d(m_j(t_i))$$

$$\text{subject to} \quad \sum_{\forall t_i \in T} r(m_j(t_i)) \leq R_{MAX} \qquad (1)$$

where T is the collection of all the SOTs of the wavelet decomposition W of the PEF, t_i is a SOT in T, $m_j(t_i)$ is an encoding of SOT t_i, $d(m_j(t_i))$ is the distortion incurred when encoding SOT t_i with mapping $m_j(t_i)$ at the data rate $r(m_j(t_i))$, and R_{MAX} is the maximum allowable rate at which the PEF can be encoded. The distortion metric used in our work is the squared error between the coefficient and its quantized value.

It can be shown that the optimization problem in Equation (1) can be solved using Lagrangian optimization as follows [15]:

$$\min_{\forall m_j(t_i):t_i \in T} \left\{ \sum_{\forall t_i \in T} \{d(m_j(t_i)) + \lambda r(m_j(t_i))\} \right\} \qquad (2)$$

We treat each SOT as an independent coding unit, therefore Equation (2) can be rewritten as:

$$\sum_{\forall t_i \in T} \min \{d(m_j(t_i)) + \lambda r(m_j(t_i))\} \qquad (3)$$

The optimal operating points for all coding units are obtained by varying λ and minimizing Equation (3) such that the sum of the resulting rates $r(m_j(t_i))$, $\forall t_i \in T$, equals the target data rate.

To obtain finer granularity in the solutions to Equation (3), we further subdivide the SOT into "subtrees", whose roots are coefficients in the original SOT that are in the LH, HL, and HH bands, respectively. The coefficients in the LL band are treated separately. Thus, a subtree is our basic coding unit. The rate-distortion data is gathered experimentally, and the optimal solution for Equation (3) is found using a bilinear search on λ. The best operating points are then used by a modified version of *CEZW* to encode the PEF. In contrast to [9], our approach can be seen as a "tree growth" instead of a "tree pruning", because we first consider an "empty" subtree, and gradually add coefficients to it until we achieve the best operating point. Tree growth is done as in *CEZW*, until the best operating point is reached. Then, a "STOP" symbol is produced to indicate that no more symbols are to be used for this subtree, and the coefficients already found significant are removed from the significant list.

4. EXPERIMENTAL RESULTS

In Figure 2, we show the result of applying our rate-distortion approach to a PEF from the *foreman* sequence. We compare it with baseline JPEG and the *CEZW*. The output from our proposed technique has a better PSNR than both *CEZW* and JPEG, plus more detail in the area of the face.
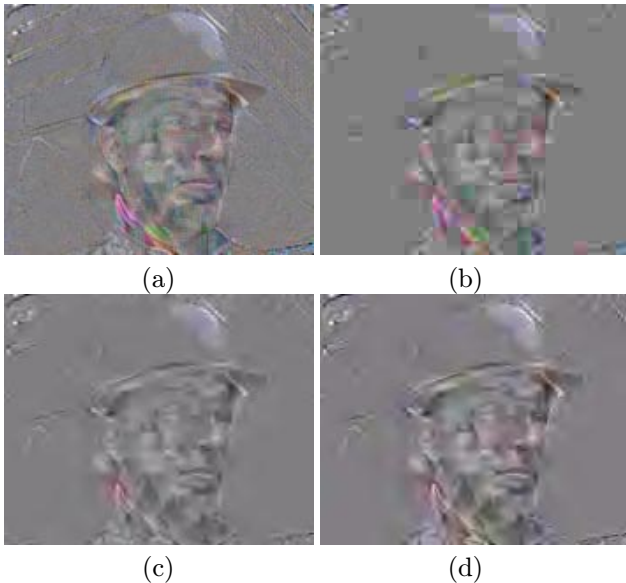

Figure 2: A predictive error frame from the *foreman* sequence. (a) Original PEF. Compressed at 0.25 bpp: (b) using JPEG (32.73 dB) (c) using *CEZW* (33.34 dB) (d) using the proposed approach (33.71 dB)

Figure 3 shows the result of using our technique on the first frame of the *foreman* sequence. The reconstruction is sharper when compared with *CEZW*.

Figure 3: The first frame from the *foreman* sequence. Compressed at 1.00 bpp: (a) using *CEZW* (35.69 dB) (b) using the proposed approach (36.59 dB)

In Figure 4, we present the results of using *SAM-CoW+* with our proposed technique, and compare its performance against H.263+. We used the *carphone*

coastguard, and *foreman* sequences, QCIF size (176x144 pixels), YUV 4:1:1, at 10 frames/s. We used Release 0.2 of the H.263+ software, obtained from UBC. Annexes D (Unrestricted Motion Vectors mode), E (Advanced Prediction mode), and F (Arithmetic Coding mode) were used, with TMN8 rate control and no frame skip.
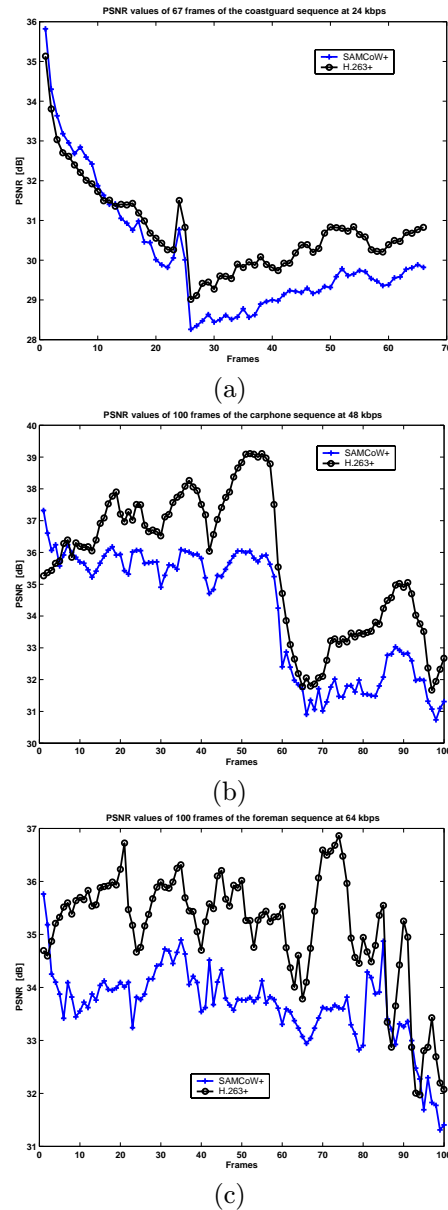

Figure 4: Comparison between *SAMCoW+* and H.263+. (a) *coastguard* at 24 kbps (b) *carphone* at 48 kbps (c) *foreman* at 64 kbps

A distinction between our implementation of *CEZW* and the one presented in [1], lies in the order in which the color components are encoded. In [1], all the bands

in the Y component are encoded first, followed by those of the U and V components. In our work, we order the color components on a per-band basis. That is, the *LL* band of the Y component is encoded first, followed by the *LL* band of U and V. Then, the *LH* band of the Y component is encoded, followed by the *LH* band of U and V, and so on. Therefore, we obtain a better balance among color components when decoding is terminated before the end of a pass – a very likely scenario.

5. CONCLUSIONS AND FUTURE WORK

In this paper, we presented a rate-distortion approach used to improve the performance of *CEZW* on PEFs. Given the bit budget allocated to the PEF, and based on the rate-distortion data, we select the best operating point for every spatial orientation tree.

Other distortion metrics could be used in this study. Minimizing the overall distortion of the wavelet coefficients does not guarantee a better reconstruction of the pixel values of the image. Thus, a metric that relates the coefficients in the decomposition and the reconstructed pixel values would yield improved results.

Currently, we are investigating the use of *SAM-CoW+* for video streaming on the Internet, an application for which the use of a rate scalable codec is highly desirable.

6. REFERENCES

[1] K. Shen and E. J. Delp, "Wavelet based rate scalable video compression," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 9, no. 1, pp. 109–122, February 1999.

[2] E. J. Delp, P. Salama, E. Asbun, M. Saenz, and K. Shen, "Rate scalable image and video compression techniques," *Proceedings of the 42nd Midwest Symposium on Circuits and Systems*, Las Cruces, New Mexico, August 8–11, 1999.

[3] K. Shen and E. J. Delp, "Color image compression using an embedded rate scalable approach," *Proceedings of the IEEE International Conference on Image Processing*, vol. III, pp. 34–37, Santa Barbara, California, October 26–29, 1997.

[4] M. Saenz, P. Salama, K. Shen, and E. J. Delp, "An evaluation of color embedded wavelet image compression techniques," *SPIE Conference on Visual Communications and Image Processing '99*, pp. 282–293, San Jose, California, January 1999.

[5] K. Shen and E. J. Delp, "A control scheme for a data rate scalable video codec," *Proceedings of the IEEE International Conference on Image Processing*, vol. II, pp. 69–72, Lausanne, Switzerland, September 16–19, 1996.

[6] E. Asbun, P. Salama, K. Shen, and E. J. Delp, "Very low bit rate wavelet-based scalable video compression," *Proceedings of the IEEE International Conference on Image Processing*, pp. 948–952, Chicago, Illinois, October 4–7, 1998.

[7] E. Asbun, P. Salama, and E. J. Delp, "Encoding of predictive error frames in rate scalable video codecs using wavelet shrinkage," *Proceedings of the IEEE International Conference on Image Processing*, Kobe, Japan, October 1999.

[8] E. Asbun, P. Salama, and E. J. Delp, "Preprocessing and postprocessing techniques for encoding predictive error frames in rate scalable video codecs," *Proceedings of the 1999 International Workshop on Very Low Bitrate Video Coding*, pp. 148–151, Kyoto, Japan, October 1999.

[9] Z. Xiong, K. Ramchandran, and M. T. Orchard, "Space-frequency quantization for wavelet image coding," *IEEE Transactions on Image Processing*, vol. 6, no. 5, pp. 677–693, May 1997.

[10] J. Li and S. Lei, "An embedded still image coder with rate-distortion optimization," *IEEE Transactions on Image Processing*, vol. 8, no. 7, pp. 913–924, July 1999.

[11] D. Taubman, "High performance scalable image compression with EBCOT," *to appear in IEEE Transactions on Image Processing*, 2000.

[12] International Telecommunication Union (ITU-T), *ITU-T Recommendation H.263 Version 2: Video Coding for Low Bit Rate Communication*, February 1998. (H.263+).

[13] A. Ortega and K. Ramchandran, "Rate-distortion methods for image and video compression," *IEEE Signal Processing Magazine*, vol. 15, no. 6, pp. 23–50, November 1998.

[14] J. M. Shapiro, "Embedded image coding using zerotrees of wavelets coefficients," *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3445–3462, December 1993.

[15] H. Everett, "Generalized Lagrange multiplier method for solving problems of optimum allocation of resources," *Operations Research*, vol. 11, pp. 399–417, 1963.