ENCODING OF PREDICTIVE ERROR FRAMES IN RATE SCALABLE VIDEO CODECS USING WAVELET SHRINKAGE

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.

ABSTRACT

Rate scalable video compression is appealing for low bit rate applications, such as video telephony and wireless communication, where bandwidth available to an application cannot be guaranteed. In this paper, we investigate a set of strategies to increase the performance of *SAMCoW*, a rate scalable encoder [1, 2]. These techniques are based on based on wavelet decomposition, spatial orientation trees, and motion compensation.

1. INTRODUCTION

Most of the research in wavelet-based image and video compression has been directed towards optimizing performance for encoding of natural scenes [3, 4, 5]. Predictive error frames (PEFs), used in many video compression techniques, present a challenge for many codecs in that they are not "natural." In [6], an algorithm for space-frequency adaptive coding of PEFs is presented. A study of the optimal bit allocation between PEFs and motion vector fields is presented in [7].

In this paper we investigate new techniques for the coding of PEFs. Our approach is based on preprocessing a PEF before encoding it. This preprocessing step uses wavelet shrinkage [8, 9] to reduce the number of relatively insignificant wavelet coefficients before zerotree encoding. An approach to encoding the wavelet coefficients in predictive error frames based on *Color Embedded Zerotree Wavelet* (*CEZW*) [1, 10, 11] is described in Section 3. The techniques described above are integrated into a rate scalable video codec, using a dynamic bit allocation strategy for predictive-coded (P) frames. This codec is an extension of the *Scalable Adaptive Motion Compensated Wavelet* (*SAMCoW*) video compression technique presented in [1, 2]. In this paper we shall refer to this extension as *SAMCoW+*. Experimental results are shown in Section 4.

2. *SAMCOW*

Rate scalable video codecs have received considerable attention due to the growing importance of video delivery over heterogeneous data networks. Current video coding standards such as MPEG-2 [12], MPEG-4 [13], and H.263+ [14] provide layered temporal, spatial, and SNR scalability. *SAMCoW* [1, 2] uses embedded coding such that the data rate can be dynamically changed on a frame-by-frame basis, and does not require the use of separate layers for scalability.

The main features of *SAMCoW* are: i) a modified zerotree wavelet image compression scheme known as *CEZW* [1, 10, 11] used for coding intracoded and predictive error frames; and ii) adaptive block-based motion compensation [15, 16] used in the spatial domain to reduce temporal redundancy. A complete description of *SAMCoW* is provided in [1, 2].

2.1. *CEZW*: Embedded Coding of Color Images

CEZW uses a unique spatial orientation tree (SOT) in the YUV color space. It exploits the interdependence between color components to achieve a higher degree of compression by observing that at spatial locations where chrominance components have large transitions, the luminance component also has large transitions [1, 11]. Therefore, each node in the SOT of the luminance component also has descendants in the chrominance components at the same spatial location. The luminance component is scanned first. When a luminance coefficient and all its descendants in both the

---

This work was supported by a grant from Texas Instruments. Address all correspondence to E. J. Delp, `ace@ecn.purdue.edu`, `http://www.ece.purdue.edu/~ace`, or +1 765 494 1740.

0-7803-5467-2/99/ $ 10.00 © 1999 IEEE

luminance and chrominance components are insignificant, a zerotree symbol is assigned. Otherwise, a positive significant, negative significant, or isolated zero symbol is assigned. The chrominance components are scanned after the luminance component. *SAMCoW* uses *CEZW* for coding intracoded (I) and predictive error frames. A variation of *CEZW*, described below, is used for coding the PEFs in *SAMCoW+*.

3. *SAMCOW+*

In this section we introduce *SAMCoW+*. In *SAMCoW+*, *CEZW* is used for coding I frames. A modified *CEZW* algorithm is used for PEFs, as shown in Figure 1. The PEF is preprocessed by using feature emphasis techniques and the elimination of information that is not visually significant. The modified *CEZW* algorithm uses wavelet shrinkage to selectively encode spatial orientation trees.

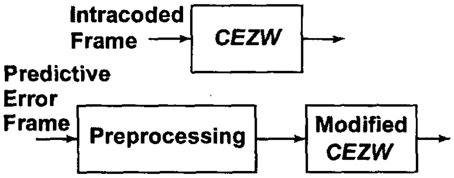


Figure 1: Coding of intracoded and predictive error frames in *SAMCoW+*.

3.1. Preprocessing and Wavelet Shrinkage

In the preprocessing stage, an adaptive gain (AG) function is used on the PEF. In this function, the areas where the predictive error is more significant are enhanced . The parameters of the AG function are set dynamically, therefore incorporating flexibility to adapt to the varying content of PEFs in a sequence. This AG function is similar to the GAG operator described in [17]. Figure 2 shows the AG function used in preprocessing the PEFs.

The AG function is defined as

$$H_{AG}(p) = \begin{cases} 0 & \text{, if } 0 \leq |p| < t_1, \\ p & \text{, if } t_1 \leq |p| < t_2, \\ p + K * (t_3 - p) & \text{, if } t_2 \leq |p| < t_3, \\ p & \text{, if } t_3 \leq |p| < max, \end{cases} \quad (1)$$

where t_1, t_2, and t_3 are thresholds that depend on the content of the PEF, K is constant that controls the feature enhancement, and max is the largest pixel magnitude in the PEF. The thresholds are chosen based on the statistics of the frame.

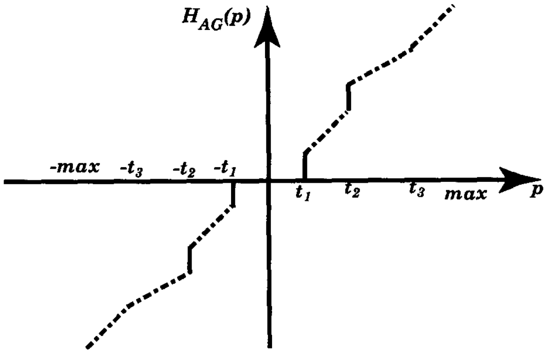
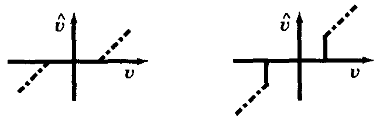


Figure 2: Adaptive gain (AG) function used to emphasize features in a PEF.

Soft- and hard-thresholding of wavelet coefficients has been used for signal and image denoising [8, 9, 17, 18]. Typical thresholding functions are shown in Figure 3. In [8], a uniform soft-threshold is used across scales of the decomposition, whereas in [17, 18] soft-thresholding is scale-dependent. The latter approach is consistent with the observation that the statistics of the coefficients change at each scale.



Soft-thresholding **Hard-thresholding**

Figure 3: Soft- and hard-thresholding of coefficient v

In this paper, we follow the procedure described in [8], using a scale adaptive threshold as in [17]. Let $f(m,n)$ be a PEF, and $v = W_j^d[f(m,n)]$ be a wavelet coefficient of $f(m,n)$ at level j $(1 \leq j \leq J)$ and spatial orientation d $(d \in \{HH, HL, LH, LL\})$. The new wavelet coefficient $\hat{v}$ is obtained as follows:

$$\hat{v} = sign(v)(|v| - t_j^d)_+ \quad (2)$$

where

$$sign(v) = \begin{cases} +1, & \text{if } v > 0, \\ 0, & \text{if } v = 0, \\ -1, & \text{if } v < 0, \end{cases} \quad (3)$$

$$(|v| - t_j^d)_+ = \begin{cases} |v| - t_j^d, & \text{if } |v| > t_j^d, \\ 0, & \text{otherwise}, \end{cases} \quad (4)$$

and t_j^d is some appropriately chosen threshold. The value of t_j^d depends on the statistics of the wavelet decomposition at level j and orientation d, and is obtained as follows:

$$t_j^d = \begin{cases} (T_{max} - \alpha(j-1))\sigma_j^d, & \text{if } T_{max} - \alpha(j-1) > T_{min} \\ T_{min}\sigma_j^d, & \text{otherwise} \end{cases} \quad (5)$$

Here, α is a decreasing factor between two consecutive levels, and T_{max} and T_{min} are maximum and minimum factors for σ_j^d, the empirical standard deviation of the wavelet decomposition at the corresponding level and orientation, respectively.

3.2. Encoding of Significant Trees

After the features of the PEF are enhanced and the coefficients of the wavelet decomposition of the PEFs are "shrunk" using the technique described above, the resulting coefficients are then encoded. When using *CEZW* to encode the coefficients of a wavelet decomposition, several passes are made to refine the precision of the approximations. As the coefficients are examined, the symbols positive significant (POS), negative significant (NEG), isolated zero (IZ), and zerotree (ZTR) are assigned [10, 11]. A coefficient is assigned the symbol IZ when the coefficient is not significant but some of its descendants are significant with respect to a threshold. In this paper, we modify the *CEZW* algorithm as follows:

1. In the first dominant pass, we will identify the coefficients that are significant (positive and negative) at the coarsest scale. We refer to these coefficients as "significant tree roots", and their descendants are part of a "significant tree." The result of this step is that only a select number of trees are considered for further processing.
2. In the remaining dominant passes, until the bit rate is exhausted, only coefficients that belong to the "significant trees" are examined.

This strategy effectively skips certain trees in the wavelet decomposition. With this modification, we intend to select the most representative information in the decomposition. Therefore, we will use the bit budget for the PEFs as efficiently as possible, encoding the most significant information and disregarding coefficients whose contribution is not significant in terms of quality of the encoding.

3.3. Dynamic Bit Allocation

In *SAMCoW*, all PEFs are assigned an equal number of bits to be used for encoding [2]. However, this approach is not efficient considering that the quality of a motion compensated frame in a group of pictures (GOP) diverges from that of the original since predictive-coded (P) frames are used as reference for other P frames. This causes PEFs towards the end of a GOP to carry more information, especially in sequences with high degree of motion. In DCT-based video codecs such as MPEG-2 or H.263+, a macroblock can be skipped when all quantized coefficients within that macroblock are zero. In a wavelet-based encoder, the coefficients in the decomposition are examined and refined until the bit budget is exhausted. However, when PEFs such as those occurring near the beginning of a GOP do not carry as much information, bits will be used to encode information that is not visually relevant. The opposite will occur near the end of the GOP.

In *SAMCoW+*, a variable number of bits is allocated to the PEF based on the number of "significant trees" being examined. This allows the data rate to vary depending on the level of activity in the scene. Furthermore, certain frames are not encoded (skipped), that is, no bits are allocated to them. This is to avoid compromising the quality of the encoded frames.

4. RESULTS AND CONCLUSIONS

We used a four-level wavelet decomposition on the PEFs, and applied soft-thresholding to all four levels. A PEF towards the end of the GOP in the *akiyo* sequence is shown in Figure 4(a). The PEF after preprocessing, as described in Section 3.1, is shown in Figure 4(b). After preprocessing, the information that is most visually significant in Figure 4(b) is still preserved, but requires fewer bits to represent it.

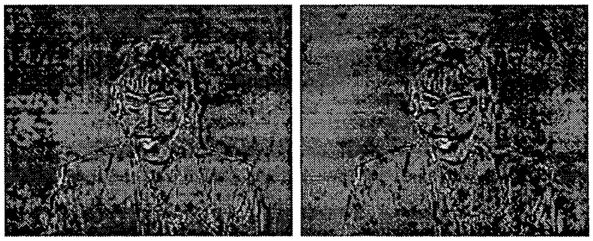

(a) (b)

Figure 4: A predictive error frame from the *akiyo* sequence. (a) Original PEF. (b) PEF after preprocessing.

Figure 5 shows the PSNR of the first 60 frames in the *akiyo* sequence decoded at 24 kbps using *SAMCoW+*, *SAMCoW*, and H.263+. The GOP size for *SAMCoW+* and *SAMCoW* was 20. Figure 6 shows the PSNR of frames 200-259 of the *foreman* sequence decoded at 64 kbps using *SAMCoW+*, *SAMCoW*, and H.263+. The GOP size for *SAMCoW+* and *SAMCoW* was 10. For both experiments, the target frame rate was 10 frames per second. In *SAMCoW+*, some frames are not encoded, that is, they are skipped. When this occurs, the decoder repeats the previously decoded frame. To obtain the PSNR values of skipped frames for Figures 5 and 6, we compared the repeated frame, with the frame in the original sequence that would correspond to the frame that was skipped. Therefore, the PSNR values for these frames are low.

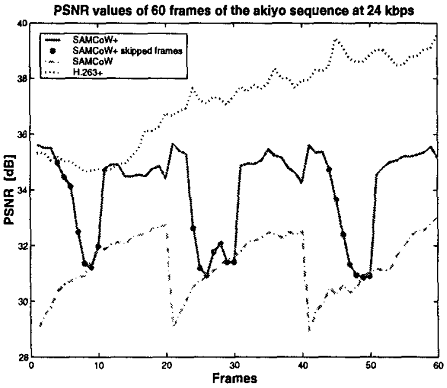


Figure 5: PSNR values of the *akiyo* sequence at 24 kbps.

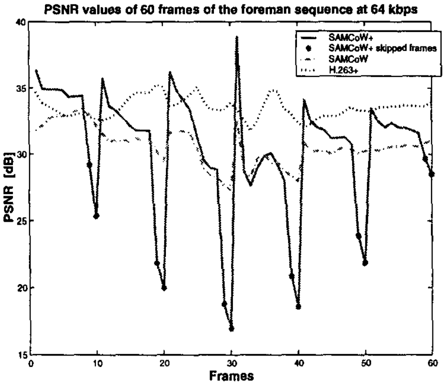
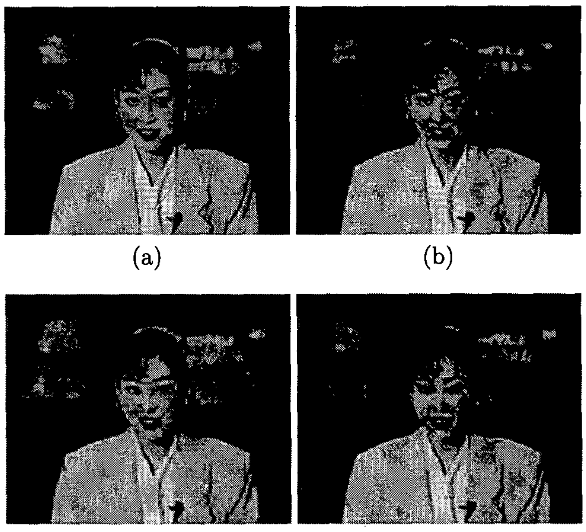


Figure 6: PSNR values of the *foreman* sequence at 64 kbps.

Figure 7 shows a frame of the decoded akiyo sequence (frame 11 in the decoded sequence, corresponding to frame 33 in the original sequence) at 24 kbps. Figure 8 shows a frame of the decoded foreman sequence (frame 13 in the decoded sequence, corresponding to frame 239 in the original sequence) at 64 kbps.

In this paper, we have presented new techniques for coding of PEFs. They include preprocessing the PEF to enhance its most important features, and soft-thresholding of coefficients of the wavelet decomposition. These techniques are integrated to *SAMCoW+*. A new bit allocation scheme is also used in *SAMCoW+*. The performance and visual quality of *SAMCoW* is improved for data rates between 24 and 64 kbps. Preprocessing has the advantage of enhancing the most visually important features of the PEFs. A disadvantage is that information about the PEF is being discarded. However, at low data rates, this information would not be encoded anyway due to the limited bit budget. Soft-thresholding has the effect of a low-pass filter on the wavelet decomposition. Therefore, a post-processing stage may be necessary to reduce this effect.


(a) (b)

(c) (d)

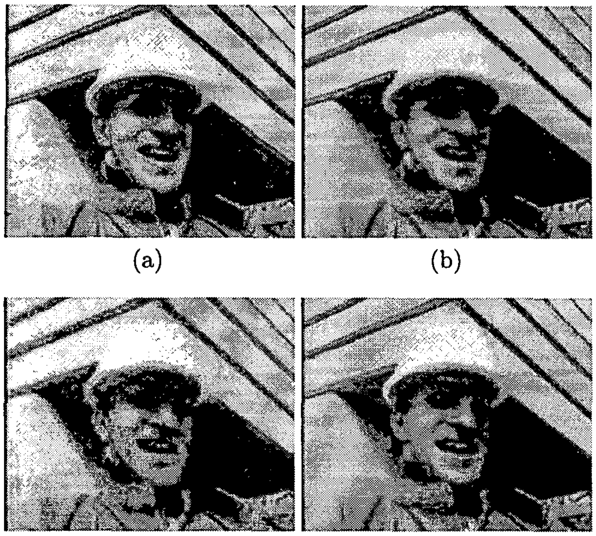
Figure 7: A frame in the *akiyo* sequence, decoded at 24 kbps. (a) Original, (b) *SAMCoW+* , (c) *SAMCoW*, and (d) H.263+.

5. 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. 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 1998.

[3] J. M. Shapiro, "Embedded image coding using zerotrees of wavelets coefficients," *IEEE Tran-*


(a) (b)

(c) (d)

Figure 8: A frame in the *foreman* sequence, decoded at 64 kbps. (a) Original, (b) *SAMCoW+* , (c) *SAMCoW*, and (d) H.263+.

sactions on Signal Processing, vol. 41, no. 12, pp. 3445–3462, December 1993.

[4] A. Said and W. A. Pearlman, "New, fast, and efficient image codec based on set partitioning in hierarchical trees," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 6, no. 3, pp. 243–250, June 1996.

[5] 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.

[6] M. Wien and W. Niehsen, "Space-frequency adaptive coding of motion compensated frame differences," *Proceedings of the 1999 Picture Coding Symposium*, pp. 65–68, Portland, Oregon, April 1999.

[7] G. M. Schuster and A. K. Katsaggelos, "A theory for the optimal bit allocation between displacement vector field and displaced frame difference," *IEEE Journal on Selected Areas in Communications*, vol. 15, no. 9, pp. 1739–1751, December 1997.

[8] D. L. Donoho, "De-noising by soft-thresholding," *IEEE Transactions on Information Theory*, vol. 41, no. 3, pp. 613–627, May 1995.

[9] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation via wavelet shrinkage," *Biometrika*, vol. 81, no. 3, pp. 425–455, 1994.

[10] 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.

[11] 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 1997.

[12] International Organization for Standardization, *ISO/IEC 13818-2, Information Technology – Generic coding of moving pictures and associated audio information*, 1994. (MPEG-2 Video).

[13] International Organization for Standardization, *ISO/IEC 14496-2, Information Technology – Coding of Audio-Visual Objects: Video*, October 1998. (MPEG-4 Version 1, Part 2: Final Draft International Standard).

[14] ITU-T, *Draft ITU-T Recommendation H.263 Version 2: Video Coding for Low Bitrate Communication*, September 1997. (H.263+).

[15] 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 1996.

[16] M. L. Comer, K. Shen, and E. J. Delp, "Rate-scalable video coding using a zerotree wavelet approach," *Proceedings of the Ninth Image and Multidimensional Digital Signal Processing Workshop*, pp. 162–163, Belize City, Belize, March 1996.

[17] X. Zong, A. Laine, and E. Geiser, "Speckle reduction and constrast enhancement of echocardiograms via multiscale nonlinear processing," *IEEE Transactions on Medical Imaging*, vol. 17, no. 4, pp. 532–540, August 1998.

[18] S. G. Chang and M. Vetterli, "Spatial adaptive wavelet thresholding for image denoising," *Proceedings of the IEEE International Conference on Image Processing*, vol. II, pp. 374–377, Santa Barbara, California, October 1997.