VERY LOW BIT RATE WAVELET-BASED SCALABLE VIDEO COMPRESSION

Eduardo Asbun, Paul Salama, Ke Shen, and Edward J. Delp

Video and Image Processing Laboratory (*VIPER*)
School of Electrical and Computer Engineering
Purdue University
West Lafayette, Indiana 47907-1285
USA

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. A technique known as Scalable Adaptive Motion Compensated Wavelet (*SAMCoW*) compression is described in this paper. *SAMCoW* is well suited for data rates less than 32 kbps. Several techniques to improve the subjective picture quality of *SAMCoW* are investigated. These include the use of B frames, half-pixel accuracy, and unrestricted motion vectors.

1. INTRODUCTION

The transmission of video over heterogeneous networks for multimedia applications, such as video telephony and wireless communication, has recently become an area of active research. The advent of the MPEG standards [1, 2, 3] has fueled the interest in this area due to their rich set of algorithms and tools for compression of various video formats and sizes at different data rates, for content-based manipulation, and for interactive access to audio-visual data.

Meeting bandwidth requirements and maintaining acceptable image quality simultaneously is a challenge. Continuous rate scalable applications can prove valuable in scenarios where the channel is unable to provide a constant bandwidth to the application. Rather than terminating the session, a decoder can adjust the bit rate to use the limited resources, yet produce video of acceptable quality. Such decoders are particularly attractive because of their flexibility. Only one sequence needs to be stored in the database, avoiding the overhead of maintaining several coded sequences at different data rates.

Rate scalability is the capability of decoding a compressed sequence at different data rates. Several approaches have been proposed to achieve rate scalability in still image compression. Shapiro's zerotree [4] exploits the interdependence between the subbands of a wavelet decomposed image. Said and Pearlman [5] investigated different tree structures that improved the quality of the decomposition in some applications. These approaches have been extended to video compression. The separable 3-D multiresolution coding algorithm of Taubman and Zakhor [6, 7] employs several levels of the wavelet transform for the temporal dimension. It requires that multiple frames be processed at the same time, resulting in a delay. Wang and Ghanbari [8] used a scheme where motion compensation was done in the wavelet transform domain. However, motion tracking errors in high frequency bands occur due to phase shifting produced by spatial shifting. The *Scalable Adaptive Motion Compensated Wavelet* (*SAMCoW*) video compression technique described in [9, 10, 11, 12] allows the data rate to be dynamically changed during decoding, with performance comparable to that of MPEG-1 [1] and H.263 [13] . Results in [9, 12] show that image quality in *SAMCoW* is perceptibly lower than H.263 at data rates less than 32 kbps in sequences with high motion.

In this paper we describe improvements to the *SAMCoW* algorithm, including the use of MPEG-type B frames, half-pixel accuracy, and unrestricted motion vectors, adapted to the particular scalability requirements of *SAMCoW*. The performance of *SAMCoW* is be compared to that of H.263+ [14].

This work was supported by grants from the AT&T and the Rockwell Foundations. Address all correspondence to E. J. Delp, `ace@ecn.purdue.edu`, `http://www.ece.purdue.edu/~ace`, or +1 765 494 1740.

2. *SAMCOW*

SAMCoW uses an embedded rate scalable coding strategy to obtain continuous rate scalability. Its main features are adaptive block-based motion compensation in the spatial domain to reduce temporal redundancy and improve image quality at low data rates, and a modified zerotree wavelet image compression scheme used on the intracoded and predictive difference frames that exploits the interdependence between the color components.

2.1. Adaptive Motion Compensation

In a scalable codec, the decoded frames have different distortions at different data rates, making it impossible for the encoder to generate the exact reference frames as in the decoder for all the possible data rates. One solution is to have the encoder lock on to a fixed data rate (usually the highest data rate) and let the decoder run freely. The codec will work exactly as the non-scalable codec when decoding at the highest data rate. However, when decoding at a low data rate, the quality of the decoded reference frames will deviate from that at the encoder [9, 10, 12]. Hence, both the motion prediction and the decoding of the predicted error frames contribute to the increase in distortion of the decoded video sequence. This distortion also propagates from one frame to the next within a group of pictures (GOP). If the size of a GOP is large, the increase in distortion can be unacceptable.

SAMCoW uses a technique known as *adaptive motion compensation* (AMC) [10, 11], in which a feedback loop is added in the decoder, as shown in Figure 1. The decoded reference frames at both the encoder and the decoder are then locked to the same data rate. This makes the reference frames at the encoder and the decoder identical, hence eliminating error propagation.

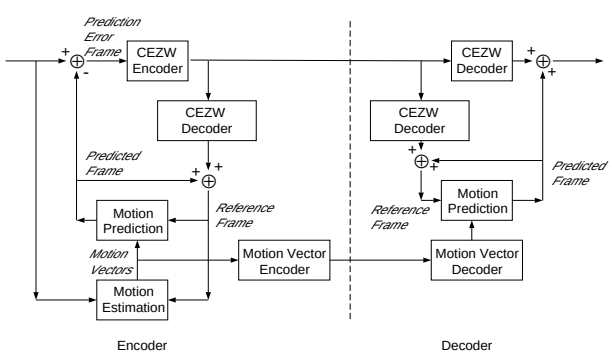

Figure 1: Block diagram of the codec for predictively coded frames. Adaptive motion compensation is used.

2.2. Embedded Coding of Color Images: *CEZW*

Interdependence between color components can be exploited to achieve a higher degree of compression. In [15, 9], a unique spatially oriented tree (SOT) is described for the color components in the YUV color space. This still image compression technique is known as the Color Embedded Zerotree Wavelet (*CEZW*). The luminance component is scanned first in a dominant pass, using the following symbols: positive significant, negative significant, isolated zero and zerotree. The two chrominance components are alternately scanned after the luminance component, in a so called subordinate pass, which is essentially the same as that in Shapiro's algorithm [15]. *SAMCoW* uses *CEZW* for coding the intracoded and predictive difference frames.

3. ADVANCED CODING OPTIONS

Results presented in [9, 12] indicate that perceived image quality in *SAMCoW* degrades rapidly at rates less than 32 kbps. Several reasons exist for this behavior. First, the wavelet transform is obtained for the entire image and the algorithm cannot allocate extra bits to areas with high activity. Second, *SAMCoW* uses only I and P frames. The MPEG standards have demonstrated that the use of B-frames can improve the performance of the coding scheme without significantly increasing the data rate. Furthermore, B frames can be discarded when needed because they are not used to predict future frames.

Several issues in the implementation of B frames in a wavelet-based codec need to be addressed. To produce lower data rates, MPEG uses coarse quantization on B frames, since they are not used as reference for other frames and, therefore, can accommodate higher distortion. In SAMCoW, a bit budget is statically allocated to each frame. As the coefficients are tested for significance using the CEZW algorithm, they are assigned the appropriate symbols during the dominant and subordinate passes. The process terminates when the budget is exhausted. Alternatively, statistical techniques for modeling wavelet coefficients [16, 17] may be incorporated to improve compression efficiency.

A form of rate control was implemented, where the intracoded (I) frames have preference over P and B frames during bit allocation. Therefore, a number of frames are skipped immediately after encoding an I frame to produce a bitstream of the desired data rate. When coding a B frame a choice is made between using forward, backward, or bidirectional prediction. The decision is based upon which motion vectors result in the best match for a particular macroblock given the amount of bits allocated for encoding that B frame.

SAMCoW makes use of overlapped block motion compensation (OBMC) for the Y component to reduce the blockiness in the prediction error image. The motion vectors are obtained via half-pixel accuracy search techniques. In particular, a full-pixel search is initially carried out over a search region, $B \times B$ pixels in size. The parameter B is chosen by the user. The top left pixel of the best matching block is then used as the center of another search region that is 3×3 pixels in size. Sixteen half-pixel searches are then carried out within the second region. In this case half-pixel data is generated by averaging the values of adjacent pixels. In addition, the best motion vectors are not restricted to point within the confines of the boundaries of each reference image. Rather, each dimension of each reference image is increased by a particular number of rows/columns. When a pixel referenced by a motion vector lies outside the original reference image boundaries, the value used is that of the closest edge pixel. This mode is commonly referred to as the unrestricted motion vectors mode. Arithmetic coding is used to produce the encoded bitstream.

4. EXPERIMENTAL RESULTS

In Figures 2 and 3 the performance of *SAMCoW* and H.263+ on the *akiyo* and *foreman* sequences (QCIF, GOP size 30) is compared. The overall PSNRs of each frame in a GOP are shown. It was observed that the decoded image quality of *SAMCoW* to be comparable to that of H.263+. The use of B frames and unrestricted motion vectors improved the results for the sequence with high degree of activity (*foreman*). As indicated in Figure 4, the frames were encoded at different frame rates to achieve the target data rate. Each of these sequences present a different level of activity.

Decoded frames using *SAMCoW* and H.263+ at different data rates (64, 32, 20, and 16 kbps) are shown in Figure 4. The H.263+ software (version 3.1.2) was obtained from `ftp://dspftp.ece.ubc.ca/pub/tmn/`. The H.263+ options used in these experiments were: syntax-based arithmetic coding (-E), advanced prediction mode (-F) (that is, OBMC and four motion vectors per macroblock), and unrestricted motion vector mode (-D 3). This implementation currently does not support frame rate control when using temporal scalability (B frames), Also, for some combinations of the above options, it produced corrupted output. In these cases, the option causing this problem (-E) was not used.

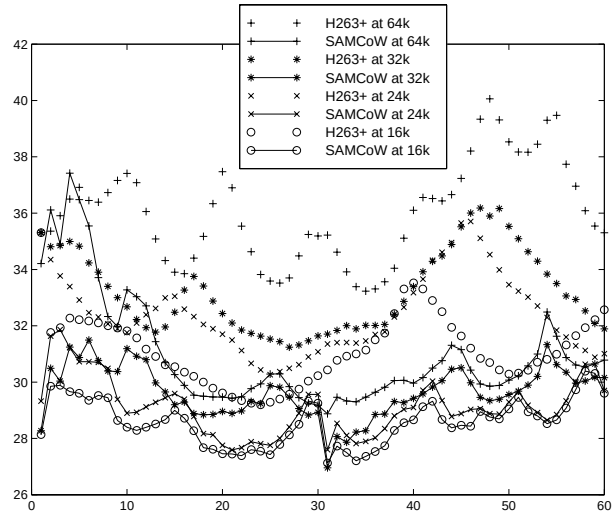

Figure 2: PSNR of each frame within a GOP of the *akiyo* sequence at different data and frame rates.

5. SUMMARY

In this paper we present the addition of techniques to the *SAMCoW* algorithm, targeted to increase the picture quality at bit rates less than 32 kbps. These include the use of MPEG-type B frames, half-pixel accuracy, and unrestricted motion vectors. Implementation issues in a wavelet-based codec were addressed and the results were compared with an implementation of H.263+. The use of the advanced coding options improved the performance of *SAMCoW* at low data rates. At higher data rates, B frames add to the range at which the sequence can be decoded, since they can be dropped as needed.

A form of rate control that allocates the available data rate to each encoded frame is implemented. The trade-off between visual quality and satisfying the constraints (data and frame rates) was incorporated into the rate control mechanism. One solution to this trade-off was to skip frames to reduce the decoded data rate. Bit allocation is done statically. However, dynamically allocating the bit budget during encoding, based on the activity of the scene, can help improve the overall quality of the sequence.

A postscript version of this paper is available via anonymous ftp to `skynet.ecn.purdue.edu` in the directory `/pub/dist/delp/icip98-samcow`.

6. REFERENCES

[1] ISO/IEC 11172-2, *Coding of moving pictures and associated audio for digital storage media at up ot about 1.5 Mbit/s.* MPEG (Moving Pictures Expert

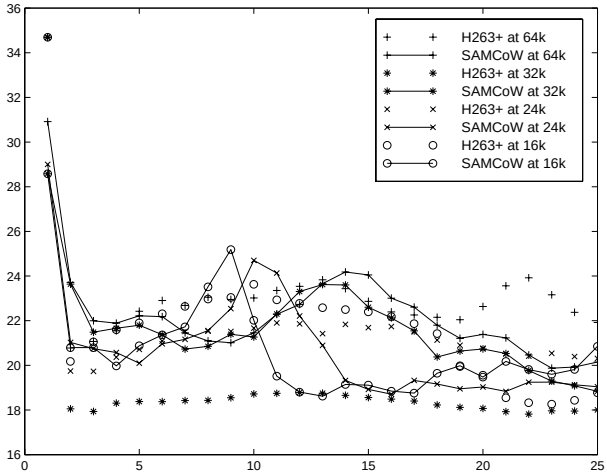

Figure 3: PSNR of each frame within a GOP of the *foreman* sequence at different data and frame rates.

Group), International Organization for Standardisation, 1993. (MPEG-1 Video).

[2] ISO/IEC 13818-2, *Generic coding of moving pictures and associated audio information.* MPEG (Moving Pictures Expert Group), International Organization for Standardisation, 1994. (MPEG-2 Video).

[3] ISO/IEC 14496-2, *Information Technology – Coding of Audio-Visual Objects: Video*, November 1997. (MPEG-4 Committee Draft).

[4] 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.

[5] 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.

[6] D. Taubman and A. Zakhor, "Multirate 3-D subband coding of video," *IEEE Transactions on Image Processing*, vol. 3, no. 5, pp. 572–588, September 1994.

[7] D. Taubman and A. Zakhor, "A common framework for rate and distortion based scaling of highly scalable compressed video," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 6, no. 4, pp. 329–354, August 1996.

[8] Q. Wang and M. Ghanbari, "Scalable coding of very high resolution video using the virtual zerotree," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 7, no. 5, pp. 719–727, October 1997.

[9] K. Shen and E. J. Delp, "Wavelet based rate scalable video compression," *Accepted for publication in the IEEE Transactions on Circuits and Systems for Video Technology*, 1998.

[10] 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, September 16–19 1996, Lausanne, Switzerland, pp. 69–72.

[11] 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*, March 3–6 1996, Belize City, Belize, pp. 162–163.

[12] K. Shen, *A Study of Real Time and Rate Scalable Image and Video Compression.* PhD thesis, School of Electrical and Computer Engineering, Purdue University, December 1997.

[13] ITU-T, *ITU-T Recommendation H.263: Video Coding for Low Bitrate Communication*, May 1996.

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

[15] 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, October 26–29 1997, Santa Barbara, California, pp. 34–37.

[16] S. LoPresto, K. Ramchandran, and M. Orchard, "Image coding based on mixture modeling of wavelet coefficients and a fast estimation-quantization framework," *Proceedings of 1997 Data Compression Conference*, March 25–27 1997, Snowbird, Utah, pp. 221–230.

[17] E. Ordentlich, M. Weinberger, and G. Seroussi, "A low-complexity modeling approach for embedded coding of wavelet coefficient," *Proceedings of 1998 Data Compression Conference*, March 30–April 1 1998, Snowbird, Utah.

Akiyo (SAMCoW) *Akiyo* (H.263+) *Foreman* (SAMCoW) *Foreman* (H.263+)

Figure 4: Frames of the *akiyo foreman* sequence, decoded at different data rates. All H.263+ frames are P frames, in contrast to *SAMCoW* which uses both P and B frames.