RESILIENT COMPRESSION OF VIDEO FOR TRANSMISSION OVER THE INTERNET

Wai-tian Tan and Avideh Zakhor

Department of Electrical Engineering and Computer Sciences
University of California, Berkeley, CA 94720 USA
e-mail: {dtan avz}@eecs.berkeley.edu

ABSTRACT

We introduce a point to point video transmission scheme over the Internet combining a low-delay TCP-friendly transport protocol in conjunction with a novel compression method that is error resilient and bandwidth scalable. Compressed video is packetized into individually decodable packets that are of equal expected visual importance. As a result, relatively constant video quality can be achieved at the receiver under lossy conditions. The packets can be truncated to meet the time varying bandwidth imposed by the transport protocol. Actual Internet experiments together with simulations are used to evaluate the performance of the overall scheme.

1. INTRODUCTION

Supporting low latency video communication over the Internet is an important yet challenging task. A few possible applications include video conferencing, telemedicine, and interactive access to pre-recorded videos stored in remote databases. However, most video compression methods used for Internet communications are neither bandwidth-scalable nor error-resilient. This produces a constant volume of inter-dependent packets that are prone to error propagation.

Producing non-adaptive streams with constant rate has two disadvantages. First, it would lead to congestive collapse when the aggregate bandwidth of the video traffic exceeds network capacity. Second, it competes unfairly with other adaptive traffic, such as TCP, that reduces transmission rate in face of network congestion.

Current approaches to error resilient video communication include error control mechanisms at the transport level. This typically takes the form of retransmissions or forward error correction (FEC). Retransmission based error control methods often fail to be real-time, particularly when round-trip propagation delay is large. FEC schemes on the other hand, are often ineffective when losses are bursty [1].

To address the above problems of flow and error control, one possible solution is to re-engineer the network to provide the necessary quality of service (QOS) guarantees via reservations and admission control. Besides requiring changes in the current infra-structure, this approach also introduces call blocking when resources become temporarily over-subscribed. Furthermore, even when such QOS guarantees are widely available, it is likely to be more costly than the plain old best-effort service.

An attractive alternative is to use bandwidth scalable video with feedback rate control whereby transmission sources adjust their rates in response to changing network conditions. This would solve the flow control problem while the error control problem can be solved by further requiring the compression method to be resilient to packet losses. In this paper, we will introduce one such compression method based on 3D subband coding. Taubman and Zakhor [6] have described 3D subband coding schemes that allow decoding at many finely grained rates without sacrificing compression efficiency. Later, Tan, *et.al.* [4] have further demonstrated the feasibility of real-time implementations. In this paper, we propose new packetization schemes for the basic compression algorithm in [4] in order to make it error resilient. This is accomplished by imposing the requirement to produce individually decodable and equally important packets. We will show via actual Internet experiments and simulations that the resilient compression performs significantly better than its otherwise equivalent non-resilient scheme. Because the proposed compression method is finely scalable, it is compatible with virtually any flow control algorithms. However, in order to be fair to TCP traffic, we have chosen to use a TCP-friendly transport protocol [5] for our Internet experiments.

This work is supported by AFOSR grant F19620-91-1-0359, California State MICRO, Sun Microsystems, Philips, and Sharp Corporation.

0-7803-5148-7/98/$10.00©1998 IEEE

2. RESILIENT SCALABLE COMPRESSION

Traditionally, scalable video compression algorithms are designed with transport prioritization in mind and often produce packets that are inter-dependent. For example, earlier 3D subband based video compression algorithms [6, 4] produce packets that are *linearly* dependent, i.e., for every K frames N packets are produced so that if packet i is lost, error propagation would prevent decoding of packets $i+1,...,N$. This is undesirable for Internet transmissions where the lack of prioritized transport causes packet losses to appear random, resulting in large variability in received video quality. To see why linearly dependent packets yields large variability under random loss, consider independent packet reception rate of p. The probability that we can decode exactly i packets out of a total of N transmitted packets is then $(1-p)p^i$ for $i \neq N$, and p^N for $i = N$, a bimodal distribution that is geometric but with a tall spike at $i = N$. For example, for $N = 20$ and $p = 0.9$, over 70% of the time we can either decode all 20 or at most 6 packets.

To eliminate error propagation, we need every packet to be individually decodable. One way to achieve this is to employ a forward decomposition of the source material into M *components* and then compress each component independently to form a packet. Each packet can then be decoded to a co-image where the sum of all co-images form the original images.

There are many such decompositions for still images. One example is the polyphase decomposition which takes every M consecutive pixels and distributes one pixel to every component. Each component then would clearly be individually decodable and approximately of equal importance. This scheme suffers from low compression efficiency. Another approach is to use block based coding in the pixel domain. However, when one packet contains all information about a spatial location, its loss will cause all information in that location to be lost. Yet another approach is to use subband decomposition to divide source into subbands that can be compressed independently. However, the DC-subband contains most of the energy for natural images. If each subband goes into a packet, this skewness would cause large variability in decoded picture quality under lossy conditions.

To overcome the problems of the above approaches, we use an alternative data partitioning scheme for subband based compression: instead of making each subband a component, we partition each subband into an equal number of coefficient blocks. Each coefficient block in a subband carries information about some localized region in the original frames. The components are then formed by grouping from each subband, equal number of coefficient blocks that correspond to different spatial regions of the source. As an example, Fig. 1 shows the formation of one component out of a total of nine. Since there are seven subbands, it would take at least seven packet losses to completely eradicate a particular spatial region.

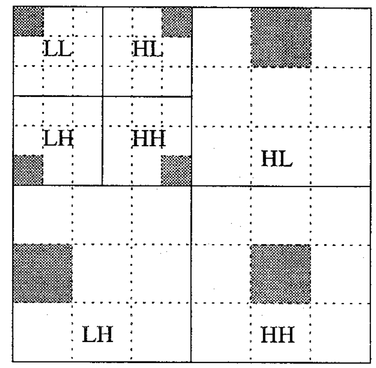

Figure 1: Grouping coefficient blocks from different subbands to form a component.

Each coefficient block is progressively quantized and compressed independent of other blocks using layered block coding [4]. However, subsequent quantizer outputs of the same block are compressed inter-dependently. To achieve error resilience, the compressed quantization layers are packed in a pre-determined order based on the relative importance of the subbands while preserving the dependency between quantization layers. Because the decoder will decode in the same order, the packet length can be truncated to suit any targeted transmission rate.

Fig. 2(a) shows original "Lena" image at 512×512. Five levels of spatial decomposition, using a 5/3-tap biorthogonal filter, are performed on the image to get 16 subbands. Each subband is then divided into 256 co-

(a) (b)

Figure 2: Original Lena (a) and Lena at 0.3 bits/pixel and 22% loss (b).

efficient blocks. The largest coefficient block is 16×16 while the smallest is 1×1. We form 256 components and compress each component using layered block coding method described in [4] to get 256 packets which are then subjected to a 22% random packet loss. The image reconstructed from the survived packets is shown in Fig. 2(b). No error concealment has been applied to the image. We see that errors are dispersed over a wide support and while the image is uniformly blurred and the total energy is diminished, all features of the original image are still visible. Furthermore, even though block based coding is employed, there are no sharp discontinuities because data partitioning is performed in the frequency domain instead of the pixel domain.

To extend the framework from still image to video, one possible way is to use 2D subband decomposition with motion compensation. However, since motion compensation does not perform well when required to produce finely scalable video, but also introduces error propagation, a scheme based on 3D subband coding is used. A Haar filter is used to generate temporal subbands. A component then is formed by getting coefficient blocks of different spatial locations from the set of spatio-temporal subbands.

3. PERFORMANCE

In this section, we evaluate the performance of the proposed error resilient compression scheme by considering (1) its compression efficiency when there is no packet loss, (2) its performance when subjected to packet losses, and (3) its performance compared to systems using FEC. We also experimented with using FEC with the proposed scheme.

Compression Efficiency: Since the proposed compression method is required to produce independently decodable packets and be bandwidth scalable, it is necessarily more restrictive than compression schemes that do not. In particular, it lacks the motion models of MPEG and H.263 and does not exploit correlation between subbands, as is done in Shapiro's embedded zerotree scheme [3]. Furthermore, correlation between coefficient blocks within the same subband is also deliberately ignored. As a result, there is in general a decrease in compression efficiency when there is no packet loss.

We compare PSNR for two sequences: "Raider of the Lost Arc", and "Mother and Daughter", with 600 and 300 frames respectively. The *Raider* sequence is a high motion fighting scene whereas *Mother* is a low motion head and shoulder sequence. The MPEG results are generated using an MPEG-1 software [2] with 10 slices per frame, GOP size of 4, and exhaustive search. While our method can produce one embedded bit-stream which can be decoded at many different rates, MPEG requires a different bit-stream to be generated for each rate. The results are shown in Table. 1. We see that the two compression methods have comparable rate-distortion performance. Because temporal subband decomposition of our proposed scheme (P) is more restrictive than the block based motion compensation of MPEG (M), scheme (P) suffers a loss in compression efficiency at low bit rates such as 500 *kbps* and below. At high bit rates such as 3 *Mbps*, (P) typically outperforms (M) because the more efficient residue coding of (P) makes up for the inefficiency of the motion model.

Bit Rates (kbps)	500	1000	1500	3000
Raider (P)	32.0	34.8	37.1	41.0
Raider (M)	32.0	34.8	36.9	40.0
Mother (P)	35.4	38.7	40.9	43.0
Mother (M)	36.7	39.7	41.3	42.7

Table 1: Compression Performance for proposed scheme (P) and MPEG-1 (M).

Packet Loss Resilience: We next compare the performance of our proposed scheme (P) under 5% simulated packet loss. Besides scheme (M), two other schemes are considered: (S) 3D subband coding where every packet contains one subband and, (T) the scalable compression scheme of [5] which produces linearly dependent packet. The results are shown in Fig. 3. We see that only scheme (P) enjoys a uniform high quality of received video. Even though packets under scheme (S) are independent, the skewness in their energy causes large variability in received video quality. Schemes (T) and (M) suffer from error propagation and show even greater variability. In particular, as a consequence of motion compensation in MPEG, we can see from Fig. 3 that errors have longer "tail" with longer GOP pattern. For scheme (P), simple error concealment is performed on the DC-subband where every missing coefficient is estimated by the average of its surviving neighbors. Similar concealment techniques are not applicable to scheme (T) under which the all spatial locations in the DC subband are compressed and transmitted together.

We next compare the simultaneous Internet transmission of scalable video compressed under schemes (P) and (T). The *Raider* sequence is repeatedly transmitted from Hongkong to Berkeley at 12 *fps*. A TCP-friendly rate-based transport protocol (TFRP) [5] is used to carry the traffic. The packet loss rate is measured in 1/3 second intervals and is shown in the top graph of Fig. 4. There are moderate and bursty packet

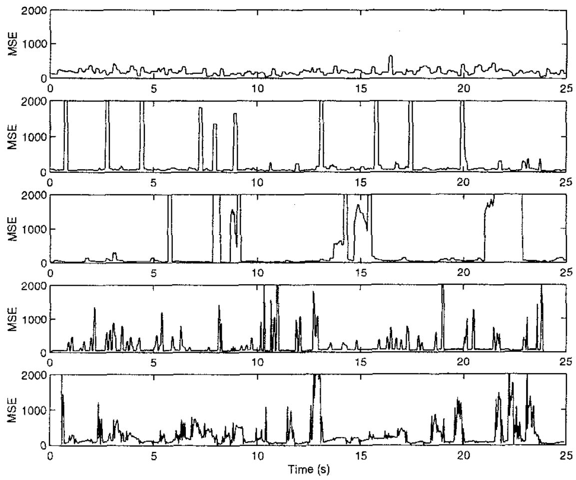

Figure 3: Variability of quality at 5% simulated random packet loss. From top to bottom: (P), (S), (T), (M) with GOP 2, (M) with GOP 15.

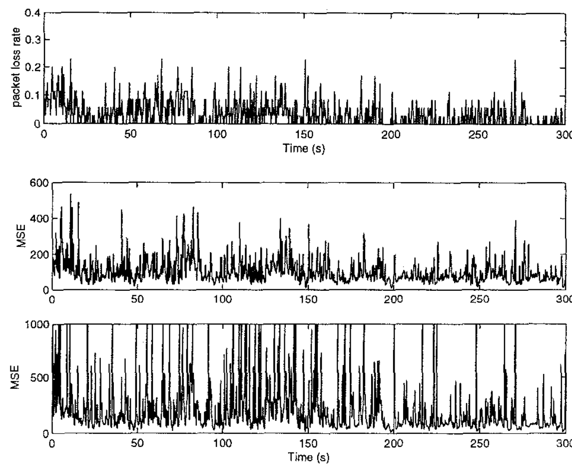

Figure 4: Simultaneous video transmission from Hong Kong to Berkeley with rate-control. (3 pm on 10/20/98). Top shows loss rate, middle and bottom correspond to MSE for schemes (P) and (T) respectively.

losses, with an average around 3.8% and variation between 0 to 23%. We see that scheme (P) out-performs (T) significantly both in having a lower average MSE of 105 versus 244, and in having smaller variability. Visually, packet losses under (P) appear as blurred patches in different locations in different frames versus the uniformly blurred frames of (T). However, the data dependency between packets in (T) causes high temporal variability in quality, resulting in oscillations between high quality and very poor quality pictures. Such wild oscillations do not exist for scheme (P) since the packets do not depend on each other. As a result, scheme (P) produces much more visually pleasing decoded video than scheme (T).

Comparison with FEC schemes: We will next consider the incorporation of FEC to non-resilient compression methods. We will restrict our discussion to the comparison between schemes (P) and (T) because both can be easily used in conjunction with rate controllers for Internet experiments, and are similar in that both employ hierarchical block coding in addition to 3D subband analysis.

Fig. 5 shows the distortion-rate characteristics when the *Raider* sequence under schemes (P) and (T) are subjected to different simulated losses and with no FEC added. We see that the two schemes have comparable performance when there is no loss. However, the performance of scheme (T) deteriorates much faster than that of scheme (P) as loss rate increases. If we were to know the packet loss rates *a priori*, one possible way to improve the quality of scheme (T) transmission is to omit the transmission of less important packets in favor of duplicates of more important one. Specifically, at any time, the transmission budget R is dictated by the flow control algorithm according to the estimated channel conditions. Assuming a total of N linearly dependent packets and that p_i is the probability that all packets up to packet i are received while packet $i+1$ is lost, then the expected distortion is given by:

$$\mathbf{E}D = \sum_{i=1}^{M} p_i D_i \tag{1}$$

where D_i is the expected distortion when the first i packets are decoded and is approximated by the 0 loss curve in Fig. 5. M is the least important packet that is being transmitted and is given by the bandwidth constraint:

$$\sum_{i=1}^{M} n_i R_i \leq R \tag{2}$$

where n_i is the number of copies that packet i is transmitted and R_i is the size of packet i. R_i is chosen to be 300 bytes to yield fine grained bandwidth-scalability. Decreasing R_i further will provide finer granularity at the expense of higher transmission overhead. p_i is calculated assuming independent, identically distributed packet losses according to a running estimate of packet loss rate that is updated 3 times every second. The n_i are pre-computed using exhaustive search for 16 fixed loss rates ranging from 0.25% to 20%.

Similarly for our proposed scheme (P), given a fix rate to code a component, it is possible to reduce that

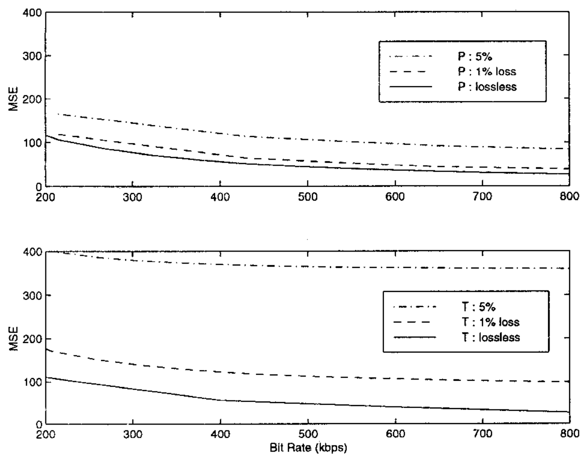

Figure 5: Distortion-rate characteristics of schemes (P) at 0, 1 and 5% packet loss.

rate to make room for piggy-backing low rate versions of other components. Following a formulation similar to that of Equations. 1 and 2, an optimal FEC scheme can be obtained for every assumed packet loss rate.

We carry out an actual Internet experiment again between Hongkong and Berkeley. The results are shown in Fig. 6. The average MSE for schemes (P) and (T) are 100.53 and 201 respectively. Both show an improvement over the case when FEC is not used. A comparison of Figs. 4 and 6 indicates that FEC reduces the variability in the decoded video for scheme (T). However, scheme (T) with FEC still shows higher average distortion and larger variability compared to scheme (P) with or without FEC. The better performance of the error resilience scheme compared to that of the non-resilient scheme with FEC is due to the mismatch between the assumed and actual packet loss rates. Since only delayed channel estimates are available for the Internet, such mismatch is unavoidable. However, FEC schemes are typically optimized only for a fixed assumed loss rate and are sensitive to such mismatch. The proposed error-resilient compression scheme however, is designed without assuming a packet loss rate and therefore tends to perform better when the channel states are not precisely known.

Complexity: For *Mother* sequence of size 320×224 on a 400 MHz Pentium machine, the encoding speeds are given by 35 to 21 frames per second in the range of 200 $kbps$ to 2 $Mbps$ respectively. The decoding speeds in the same range varies from 55 to 25 frames per second. The reported times exclude disk access and display time [4].

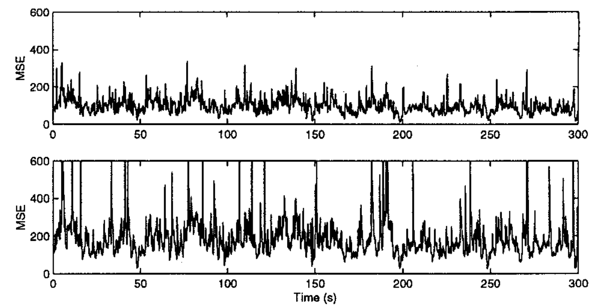

Figure 6: Simultaneous video transmission from Hongkong to Berkeley with rate control and FEC. Top is scheme (P) and bottom is (T).

4. CONCLUSION

In this paper, we devise a new packetization scheme for scalable video compression [4] that is well suited for low delay Internet communications. It offers a finely layered representation that facilitates the use of arbitrary flow control algorithms, and produces individually decodable packets so that error propagation is practically eliminated. We carried out simulation and actual Internet experiments, and found it to out-perform schemes that produce inter-dependent packets.

5. REFERENCES

[1] I.Cidon, A.Khamisy and M.Sidi. Analysis of Packet Loss Processes in High-speed Networks. *IEEE Trans. Info. Theory*, Vol. 39, No. 1, pp 98-108, January 1993.

[2] K.Patel, B.Smith, L.Rowe, Performance of a Software MPEG Video Decoder, *Proc. of the ACM Multimedia '93* pp 75-82.

[3] J.Shapiro. Embedded Image Coding using Zerotrees of Wavelet Coefficients. *IEEE Trans. Sig. Proc.*, Vol. 41, No. 12, pp 3445-3462, December 1993.

[4] W. Tan, E. Chang, and A. Zakhor. Real Time Software Implementation of Scalable Video Codec, *Proc. ICIP*, Vol. 1, pp 17-20, September 1996.

[5] W.Tan and A.Zakhor. Internet Video using Error Resilient Scalable Compression and Cooperative Transport Protocol, *Proc. ICIP*, Vol. 3, pp 458-462, October 1998.

[6] D.Taubman and A.Zakhor. Multirate 3-D Subband Coding of Video. *IEEE Trans. Image Proc.*, Vol. 3, No. 5, pp 572-88, September 1994.