ITU – Telecommunications Standardization Sector Document Q15-G-09
Study Group 16 Filename: q15g09.ps
Video Coding Experts Group (Question 15) Generated: 08 Feb '99

Seventh Meeting: Monterey, California, 16-19 February, 1999

Question: Q.15/SG16
Source: Vinay Vaishampayan, Amy R. Reibman, and Barry G. Haskell
AT&T Laboratories – Research
180 Park Ave.
Florham Park, NJ 08932
Tel: +1-973-360-8434
Fax: +1-973-360-8178
email: vinay@research.att.com
Title: Packet Loss/Frame Erasure Resilient Video Coding: A case for Balanced Multiple Descriptions in H.26L
Purpose: Proposal

1. INTRODUCTION

Packet video services are expected to become more widespread in coming years in both wired and wireless environments. A significant problem in both environments is the degradation in video quality due to channel imperfections. Layered video coders have been developed in order to address such issues in networks where strong QOS guarantees can be made. For networks where such guarantees are not possible, we propose balanced multiple description video compression as a candidate for future video compression standards. In this document, we describe balanced multiple descriptions and make the argument both intuitively and through simulated results.

This document is organized as follows. In Section 2, we outline some problems with standard packet video coders in lossy environments. Multiple description data compression is introduced in Section 3 and an overview of the proposed coder is presented in Section 4. In Section 5, we summarize the experiment that was conducted in support of the argument of this document.

2. SINGLE DESCRIPTION COMPRESSION AND PACKET LOSS CHANNELS

In the standard data compression problem, the objective is to construct a rate constrained binary description of the source sequence, while minimizing a certain distortion measure. There are no requirements about being able to decode something meaningful from subsets of the binary description. We refer to such coders as single description video coders. In packet video applications, in wired and wireless environments, the channel is expected to be less than perfect, resulting in losses in the binary description. Reconstructed video from single description coders can become unacceptable as a result.

The problem of packet loss is significant in *interframe* video coders, where the loss of information associated with a single video frame has undesirable consequences for later frames as well, because the encoder and the decoder states become unsynchronized. Here the encoder makes a prediction for encoding the next frame based on information which the decoder does not have, causing the decoder to apply encoded frame difference information to the wrong frame. Thus lost data corresponding to a particular video frame has the potential to degrade future frames. Many methods have been proposed for improving the performance of single description coders over such channels, notably (i) error concealment techniques in the decoder that rely on residual correlation in order to "cover up" the missing data (ii) data partitioning techniques which are useful in making the task of error concealment easier in the decoder (iii) methods for reducing the mistracking problem by forcing the encoder into the intra-frame mode more often in order to resynchronize the encoder and the decoder states. We feel that in addition to such methods, there is a need to build in redundancy that is specifically tailored to the loss process. This is where multiple description coding comes in.

3. MULTIPLE DESCRIPTION DATA COMPRESSION

In order to better protect against lost packets, it is obvious that some form of redundancy that is matched to the channel degradation would be advantageous, thus allowing the decoder to reconstruct meaningful video even when there are "holes"

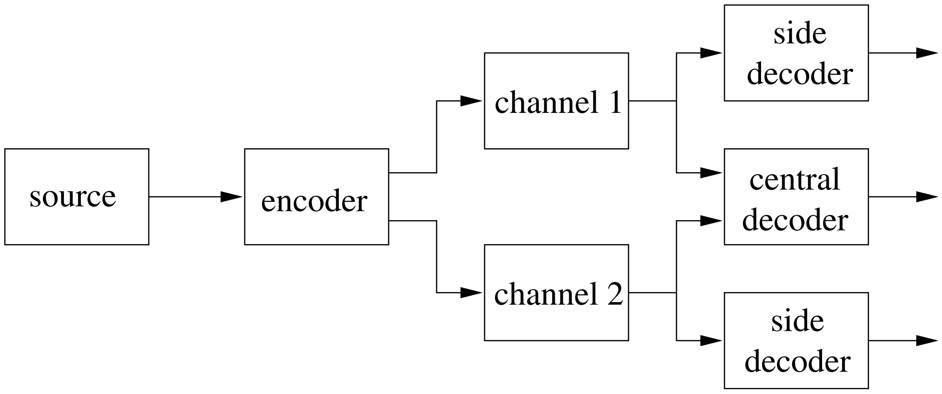

Figure 1: The multiple description setup.

in the binary description. For example, it would be useful if information about each frame could be placed into two packets instead of just one, so both packets would need to be lost before the video quality would be seriously degraded. At the same time, it would also be desirable to keep the corresponding rate overheads low. This is the topic of multiple description data compression, and we give a brief overview now.

Multiple description coders produce multiple bit streams, or descriptions of the sequence being encoded and come in many flavors. The best known examples of such coders are the SNR scalable, (or layered, or hierarchical) video coders that produce two binary streams. The first stream, referred to as the base layer, contains essential information about the video sequence. The second stream contains refinement information only. It is possible to decode the base layer alone into a slightly degraded replica of the original and it is possible to decode the two layers jointly, with smaller degradation. However, the refinement information cannot be used by itself. The class of multiple description codes is quite large and layered coders may be thought of as staking out one end of this class. At the other end lies the class of balanced multiple description coders which we now describe.

The idea behind balanced multiple description video coding is to produce two bit streams that are of equal importance and yet are able to refine one another. It should be possible to decode either stream by itself, to yield acceptable video quality, while at the same time it should be possible to decode both streams together to yield excellent quality. Such coders have a diversity advantage, since either stream may be lost with acceptable levels of degradation.

The application areas of these two kinds of coders are quite different since they have very different requirements of the packet network that they connect to. Layered coders rely on the notion of priority classes in a packet network. Such networks make strong guarantees for arrival of high priority packets and these packets are then used to carry the base layer information. The refinement information is sent using lower priority packets. It is assumed that the network will drop base layer packets very rarely, thus a certain minimum fidelity is always guaranteed. Balanced multiple description coders, on the other hand, do not have an explicit base layer and hence are suitable for networks that do not have a priority structure. In other words, when strong QoS guarantees are not possible, e.g., when a system cannot be engineered end-to-end, balanced multiple description codes would be suitable. A good example is, of course, is the internet.

We now describe a more formal setup of the multiple description problem with the help of the block diagram of Fig. 1. The setup consists of a source, which is connected to the user via two channels. Either channel may be broken, this is known at the receiver but not at the transmitter. Conceptually, a multiple description coder consists of an encoder and three decoders. The three decoders are (i) the channel-one decoder which is used when channel-one alone works, (ii) the channel-two decoder which is used when channel-two alone works, and, (iii) the two-channel decoder which is used when both channels work. The aim of code design is to minimize the distortion when both channels work, while ensuring that the distortions when either channel fails, remain below certain preset levels. In the case of layered codes, the distortion contraint when channel two alone works is set to infinity, whereas, in the balanced case, the both distortion constraints are set to the same value, say, to some acceptable level. An important question is that of rate overhead which is measured relative to a single description coder for which the distortion equals the two-channel distortion of the multiple description coder. An example of the most rudimentary balanced multiple description coder is a repetition of the single-description bit stream, with a corresponding rate overhead of 100%. Fortunately methods exist with rate overheads between 10% and 30% and research to reduce the overheads even further is in progress.

4. INTERFRAME BALANCED MULTIPLE-DESCRIPTION VIDEO COMPRESSION

Multiple description data compression system design has been an active area of research in recent years. Methods for quantization and still-frame image compression have been well studied. However, the problem of designing interframe video coders has remained open for some time primarily because of the predictive nature of the coder.

An interframe coder is essentially a predictive coder. Single description predictive coders work by predicting the next sample, using past decoded samples. This is done in order to prevent the decoder and the encoder from mistracking. On the one hand are multiple description coders which allow no mistracking, and on the other are coders which carefully manage the mistracking so that it is not visually disturbing. It is not clear at this stage how important the mistracking problem is for multiple description coders, when the final distortion is measured over a packet loss channel. It seems realistic to expect the final solution will depend on the nature of the packet loss process itself. The first step, however, is to design various coders so that experiments can be carried out to assess the merits of each system.

To eliminate any mistracking in a multiple description coder requires that the encoder and decoder frame memories track, when either description is available by itself and also when both descriptions are available. We have recently developed a multiple description video coder structure which solves the mistracking problem and experiments with it yield promising results. This balanced multiple description frame difference coder consists of two predictive loops, one for each channel. Each loop maintains its own prediction of the next sample and the tracking problem is solved by carefully addressing the quantization issues. Details of the coder as well as a more lucid description of the multiple description problem may be found in the attached papers.

Future work will include coders that carefully manage the mistracking instead of completely elimiinating it, thus allowing a more thorough exploration of the design space.

5. SIMULATION RESULTS

Encoded video was packetized and transmitted from machine-1 using the UDP protocol to machine-2, where packets were dropped randomly and independently and sent to machine-3. Video was decoded and displayed on machine-3. Packets were constructed so that information from only a single frame lies in each packet, i.e., packets do not cross frame boundaries. The video shows simulation results with a packet loss rate of 20% for both the standard (single-description) H.263 video coder and a balanced multiple-description video coder. Packet sequence numbers are assumed and used to detect lost packets. In both cases an error concealment method was used to conceal the effect of lost packets. If information corresponding to a given frame is completely lost, that portion of the current frame is replaced with the previous frame. The single description coder uses this strategy when any packet is lost, while the multiple description coder uses this strategy only when both packets associated with a given frame are lost. When only a single packet is lost, the corresponding (channel-one or channel-two) decoder is used. Signal-to-noise ratios are also presented, along with the simulation. The SNR's are presented in Table 1 and were obtained for the sequence "silent". There is a significant gap in performance.

Packet Loss Prob. (%)	SNR (Multiple Descr.)	SNR (Single Descr.)
0	42.3	42.4
5	40.4	39.9
10	38.72	38.32
15	37.10	36.56
20	35.91	33.67
25	34.67	32.53
30	33.54	31.57

Table 1: PSNR measured on the luminance component as a function of the packet loss probability for the multiple description and single description video coders.

6. SUMMARY

Experimental results suggest that the balanced multiple description approach has strong potential for lossy packet environments. We believe that the technology has advanced to a point where it is reasonable to think of this method as a candidate for a future, packet loss resilient, video compression standard.

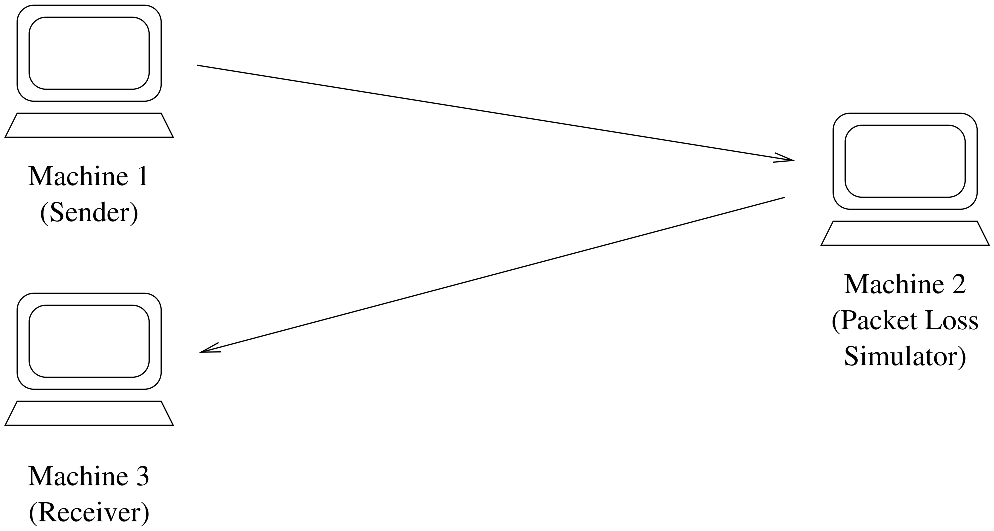

Figure 2: Packet video simulation setup.