N92-23420
edited by E. Arikan
Elsevier Science Publishers B.V., 1990

A ROBUST LOW-RATE CODING SCHEME FOR PACKET VIDEO

Y.C. Chen, K. Sayood and D. J. Nelson
Department of Electrical Engineering
Center for Communication and Information Science
University of Nebraska-Lincoln

I. INTRODUCTION

Due to the rapidly evolving field of image processing and networking, video information promises to be an important part of tomorrow's telecommunication system. Up to now, video transmission has been mainly transported over circuit-switched networks. It is quite likely that packet-switched networks will dominate the communications world in the near future. Asynchronous transfer mode (ATM) techniques in broadband-ISDN can provide a flexible, independent and high performance environment for video communication. Therefore, it is necessary to develop techniques for video transmission over such networks.

The recent literature contains a number of proposed packet video schemes. Verbiest and Pinnoo proposed a DPCM-based system which is comprised of an intrafield/interframe predictor, a nonlinear quantizer, and a variable length coder[1]. Ghanbari has simulated a two-layer conditional replenishment codec with a first layer based on hybrid DCT-DPCM and second layer using DPCM[2]. Darragh and Baker presented a sub-band codec which attains a user-prescribed fidelity by allowing the encoder's compression rate to vary[3]. Kishino et al. describe a layered coding technique using discrete cosine transform coding, which is suitable for packet loss compensation[4]. Karlsson and Vetterli presented a sub-band coder using DPCM with a nonuniform quantizer followed by run-length coding for baseband and PCM with run-length coding for nonbaseband[5]. In this paper, a different coding scheme called MBCPT is investigated. Unlike the methods mentioned above, MBCPT doesn't use decimation and interpolation filters to separate the signals into sub-bands. However it does have the attractive property of dealing separately with high frequency and low frequency information. This separation is obtained by the use of variable blocksize transform coding[6].

To deliver packets in a limited time and provide a real time service is a difficult resource allocation and control problem, especially when the source generates a high and greatly varying rate. In packet-switching networks, packet losses are inevitable, but use of a packet switching network yields a better utilization of channel capacity. The video coder will require different channel capacity over time but the network will provide a channel whose capacity changes depending on the traffic in the network. Therefore, the interactions between the coder and the network have to be considered and be incorporated into the requirements for the coder. These requirements include:

1. Adaptability: The video source has a varying information rate. The encoder should therefore generate different bit rates corresponding to the varying information rate.
2. Insensitivity to error: The coding scheme has to be robust to packet loss so as to preserve the image quality. Recall that retransmission is impossible because of tight timing requirements.
3. Control of coding rate: Sensing the heavy traffic in the network, the coding scheme is required to adjust the coding rate. In the case of a congested network, the coder could be switched to another mode which generates fewer bits with a minimal degradation of image quality.
4. Parallel architecture: The coder should preferably be implemented in parallel. This would allow the coding procedure to be run at a lower rate in many parallel streams.

In the next section, we investigate the proposed coding scheme to see how well it satisfies the above requirements.

II. MIXTURE BLOCK CODING WITH PROGRESSIVE TRANSMISSION

Mixture Block Coding (MBC) is a variable-blocksize transform coding algorithm which codes the image with different blocksizes depending upon the complexity of that block area.[6] When using MBC, the image is first divided into maximum blocksize blocks. After coding, the distortion between the reconstructed and original block is calculated. The block being processed is subdivided into smaller blocks if that distortion fails to meet the predetermined threshold. The coding-testing procedure continues until the distortion is small enough or the smallest blocksize is reached.

MBCPT is a multipass scheme in which each pass deals with different blocksizes. The first pass codes the image with maximum blocksize and transmits it immediately. Only those blocks which fail to meet the distortion threshold go down to the second pass which processes the difference image block, coming from the original and coded image obtained in the first pass, with smaller blocks. The difference image coding scheme continues until the final pass which deals with the minimum size block. At the receiving end, a crude image is obtained from the first pass in a short time and the data from following passes serve to enhance it. The coding structure is similar to a quad tree structure proposed by Dreizen[7], and Vaisey and Gersho[8]. In the quad tree coding structure used in this paper, a 16x16 block is coded and the distortion of the block is calculated. If the distortion is greater than the predetermined threshold for 16x16 blocks, the block is divided into four 8x8 blocks for additional coding. This coding-checking procedure is continued until the only image blocks not meeting the threshold are those of size 2x2.

The coding technique used is the discrete cosine transform. For all blocksizes, only four coefficients of the transform, including the dc and three lowest order frequency coefficients, are coded and others are set to zero. The dc coefficient in the first pass is coded with an 8-bit uniform quantizer. In the remaining passes it has a laplacian distribution and a 5-bit optimal laplacian nonuniform quantizer is used. As an alternative, an LBG vector quantizer with a 512 codebook size is used to quantize the vector which comprises the three ac coefficients. The threshold of each pass has to be pre-selected and is readjustable during the operation according to the channel condition and quality required. Because only partial blocks which fail to meet the distortion threshold need to be coded, there must be some side information to instruct the receiver how to reconstruct the original image. One bit of overhead is needed for each block. If a block is to be divided, a 1 is assigned to be its overhead; if not, a 0 is assigned.

The interframe coder used in this paper is a differential scheme. This coder processes the difference image coming from the current frame and the previous frame which is locally decoded using data from the first three passes. Only the data from the first three passes is used because, while under conditions of no packet loss there is almost no difference between using three passes and using all four passes, there is substantially less degradation in the former approach when there is packet loss. This can be seen from the results in Fig. 1. In this paper, the Kronkite motion sequence with 16 frames is used as the simulation source. Every image consists of 256x256 pixels with graylevels ranging from 0 to 255. It is similar to a video conferencing type image which has neither rapid motion nor scenes changes. No motion detection or motion compensation techniques are used but could be implemented when broadcasting video.

From the datastream output listed in Table 1, we can see that the data in pass 4 represents 30-40% of the entire data. Pass 4 is primarily responsible for the clarity of the image and is usually labeled with the lowest priority in the network. We therefore call this the least significant pass(LSP). The packets containing this data have substantial possibility of being discarded due to low priority. As they are not used in the prediction process, their loss does not cause error propagation.

III. INTERACTION OF THE CODER AND THE NETWORK

The network simulator used for this study was a modified version of an existing simulator developed by Nelson et al[9]. Details of this simulator can be found in [10, 11]. When the video data is packed and sent into a nonideal network, some problems emerge.

A. Packetization

The task of the packetizer is to assemble video information, coding mode information, if it exists, and synchronization information into transmission cells. In order to prevent the propagation of the error resulting from the packet loss, no data from the same block or same frame is separated into different packets. As the segmentation process in the transport layer has no information regarding the video format, the packetization process has to be integrated with the encoder, which is in the presentation layer on the user's premise. Otherwise, some overhead has to be added into the datastream to guide the transport layer to perform the packetization in the desired manner.

Every packet must contain an absolute address which indicates the location of the first block it carries. Because every block in MBCPT has the same number of bits in each pass, there is no need to indicate the relative address of the following blocks contained in the same packet. Fixed length packetization is used in this paper for simplicity.

B. Error Recovery

There is no way to guarantee that packets won't get lost after being sent into the network. Packet loss can be mainly attributed to bit errors in the address field, leading the packets astray in the network, or due to congestion which exceeds the networks management ability causing packets to be discarded. Effects created by higher pass packet loss (like pass 4) in MBCPT coding will be masked by the basic passes and replaced with zeros. The distortion is almost invisible when viewing at video rates because the lost area is scattered spatially and over time. However, low pass packets loss (like pass 1), though rare due to high priority, will create an erasure effect which can be quite objectionable. Replacing lost data with reconstructed values from the corresponding area in the previous frame is not very effective. Motion detection and motion compensation could be used to find a best matched area for replacement in the previous frame.

Side information in the MBCPT decoding scheme is very important. To prevent this vital information from getting lost, error control coding can be applied in both directions along with and perpendicular to the packetization. The former is for bit error in the data field while the latter is for packet loss. The minimum distance that the error control coding should provide depends on the network's probability of packet loss, correlation of such loss and channel bit error rate. Also, as the output rate of side information and pass 1 and even pass 2 is quite steady, a fixed amount of channel capacity could be allocated to these outputs to ensure their timely arrival. That means circuit-switching can be used for important and steady data.

C. Flow Control

In order to shield the viewer from severe network congestion. Flow control schemes can be used. If the encoder is aware of congestion in the network, it can adjust its coding scheme to reduce the output rate. In the MBCPT coding scheme, if the output buffer exceeds a given threshold, the encoder can switch to a coarse quantizer with fewer steps or loosen the threshold to decrease its output rate. In this way, smooth quality degradation is obtainable. Of course, this also complicates the encoder design. It is also possible to use the congestion control of the network protocols to prevent the drastic quality change by assigning different priorities to packets from different passes. In the MBCPT coding scheme, side information and packets from pass 1 are assigned highest priority and higher pass packets are assigned decreasing priority.

D. Interaction with protocols

In the ISO model, physical, datalink and network layers comprise the lower layers which form a network node. The higher layers which consist of transport, session, presentation and application layers, typically reside in a customer's premises and perform all the functions of the packet video coder. The transport layer does the packetization and reassembly. The session layer supervises set-up and tear-down for sessions which have different types and quality. The quality of a set-up session can be determined by the threshold in the coding scheme and the priority assignment for transmission. Of course, the better the quality, the higher the cost. Fig. 2 shows the tradeoff between PSNR and video output rate by adjusting thresholds. The presentation layer does most of the signal processing, including separation and compression. Because it knows the video format exactly, if any error concealment is required, it will be performed here. The application layer works as a boundary between the user and the network and deals with all the analog-digital signal conversion.

IV. RESULTS FROM PACKET VIDEO SIMULATION

The results obtained in this packet video simulation show that a reasonable compression with good image quality can be obtained using the proposed scheme. The monochrome sequence used in this simulation corresponds to a bit rate of 15.3 Mbits/s, given a video rate of 30 frames/s. As Table 1 shows, the average data rates of our system is 1.539 Mbits/s. The compression rate is about 10 with a mean PSNR of 38.74 dB. Fig. 3 shows the data rate of the sequence frames with side information, 4 passes and total rate. It is clear that data rate of pass 1 is constant as long as the quantization mode is kept the same. Side information and data from pass 2, even pass 3, is quite steady. The data rate of pass 4 is bursty and highly-uncorrelated. Fig. 4 shows the PSNR for each frame in the sequence. The standard deviation is only 0.2 dB. In the simulation, the same threshold is used throughout the sequence. If constant visual quality is desired, a varying threshold can be used for different frames. That will generate a more variable bit rate and, of course, motion detection would be required.

From the difference images of this sequence, frames 1-8 seem quite motionless while frames 9-13 contain substantial motion. We adjusted the traffic condition of the network to force some of the packets to get lost so as to check the robustness of the coding scheme. Heavy traffic is set up in the motionless and motion period separately. The average packet loss percentage is 3.3% which is considered high for most networks. Fig. 5 shows an image which suffers packet loss from pass 4. As can be seen, the effect of lost packets is not at all severe, even if the lost packet rate is unrealistically high. Fig. 6 shows the case when packet loss occurs in pass 1. Clearly there are visible defects in the motion period. Apparently the replenishing scheme used here is not sufficient in areas with motion. It is believed that the performance can be improved with a motion compensator algorithm which would find the appropriate area for replenishment.

V. CONCLUSIONS

The network simulator was used only as a channel in this simulation. In fact, before the real-time processor is built, a lot of statistics can be collected from the network simulator to improve upon the coding scheme. These include transmission delays and losses from various passes under different network loads. For resynchronization, the delay jitter between received packets can also be estimated from this simulation. MBCPT has been investigated for use over packet networks and has been found to provide high compression rate with good visual performance, robustness to packet lost, tractable integration with network mechanics and simplicity in parallel implementation. For fast moving scenes, the differential MBCPT scheme seems insufficient. Motion compensation, error concealment or even attaching function commands into the coding scheme are believed to be useful tools to improve the performance and will be the direction of future research.

ACKNOWLEDGEMENT

This work was supported by a grant from the NASA Goddard Space Flight Center (NAG 5-916).

REFERENCES

[1] W. Verbiest and L. Pinnoo, "A Variable Bit Rate Codec for Asynchronous Transfer Mode Networks," *IEEE J. Select. Areas Commun.*, vol. 7, pp. 801–806, June 1989.

[2] M. Ghanbari, "Two-Layer Coding of Video Signals for VBR Networks," *IEEE J. Select. Areas Commun.*, vol. 7, pp. 801-806, June 1989.

[3] J. Darragh and R. Baker, "Fixed Distortion Subband Coding of Images for Packet-Switched Networks," *IEEE J. Select. Areas Commun.*, vol. 7, pp. 801–806, June 1989.

[4] F. Kishino, K. Manabe, Y. Hayashi and H. Yasuda, "Variable Bit-Rate Coding of Video Signals for ATM Networks," *IEEE J. Select. Areas Commun.*, vol. 7, pp.801–806, June 1989

[5] G. Karlsson and M. Vetterli, "Packet Video and Its Integration into Network Architecture," *IEEE J. Select. Areas Commun.*, vol. 7, pp. 739–751, June 1989.

[6] M. C. Rost, "Data Compression Using Adaptive Transform Coding," Ph.D. Dissertation, University of Nebraska-Lincoln, 1988.

[7] H. M. Dreizen, "Content-Driven Progressive Transmission of Grey-Scale Images," *IEEE Trans. Commun.*, vol. COM-35, pp. 289-296, March 1987.

[8] D. Vaisey and A. Gersho, "Variable Block-Size Coding," *Proc. ICASSP*, pp. 1051–1054, April 1987.

[9] D. Nelson, K. Sayood, G. Ankenman and H. Chang, "Pnetsim System Programmer's Manual," University of Nebraska-Lincoln, Final Report for ARMY Contract DAAB07-85-K-K535, Dec. 1986.

[10] Y. C. Chen, "Mixture Block Coding with Progressive Transmission in Packet Video," MS Thesis, University of Nebraska-Lincoln, 1989.

[11] Y. C. Chen, K. Sayood, and D. J. Nelson, "A Packet Video Scheme Based On Variable Block-Size Coding," submitted to IEEE Transactions on Communications.

Table 1

	OVERHEAD	PASS1	PASS2	PASS3	PASS4	TOTAL
MEAN	65.28	130.56	214.50	591.87	538.86	1539.36
DEVIATION	8.70	00.00	32.82	95.37	210.00	311.85
MAXIMUM	77.04	130.56	280.56	735.84	821.52	1990.80
MINIMUM	44.88	130.56	136.08	384.72	21.84	1042.08

Output bit rate for each and total pass calculated with 30 frames/sec video rate. The maximum and minimum values are the instantaneous rates, which correspond to the respective maximum and minimum number of bits needed to encode a particular frame in the sequence. The unit is kilobits.

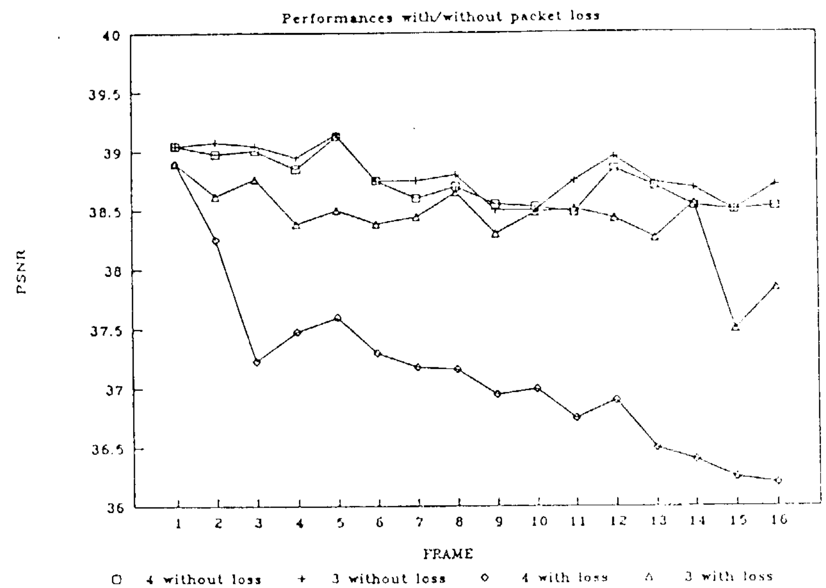

Figure 1 Performance with and without packet loss.

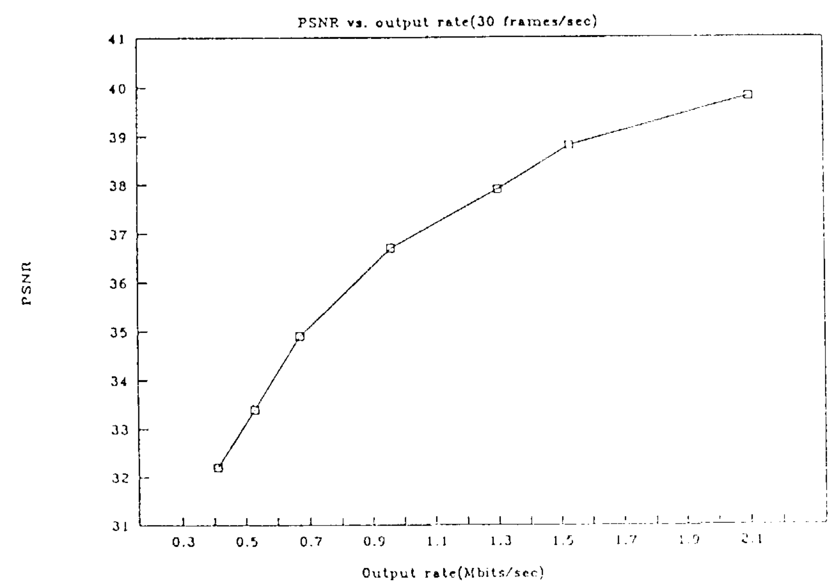

Figure 2 PSNR versus video output rate with 30 frames per second.

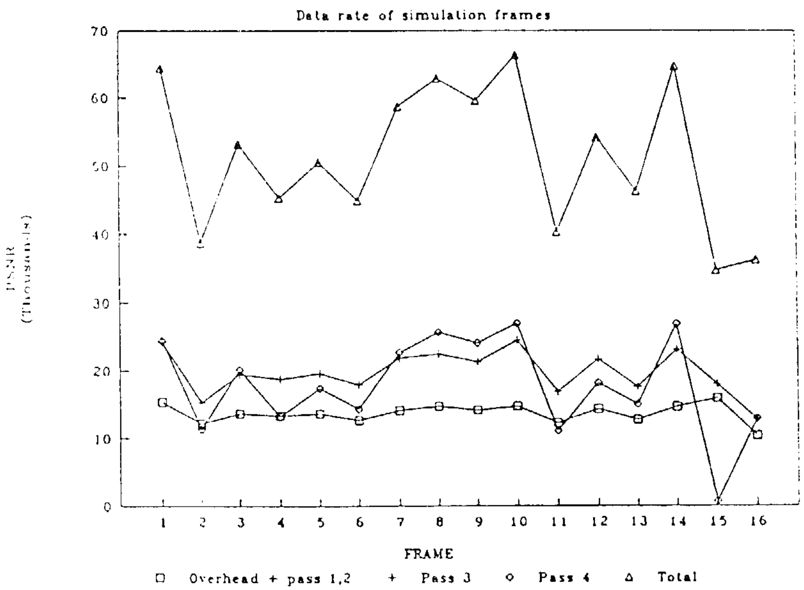

Figure 3 Data rate of simulation sequence frames.

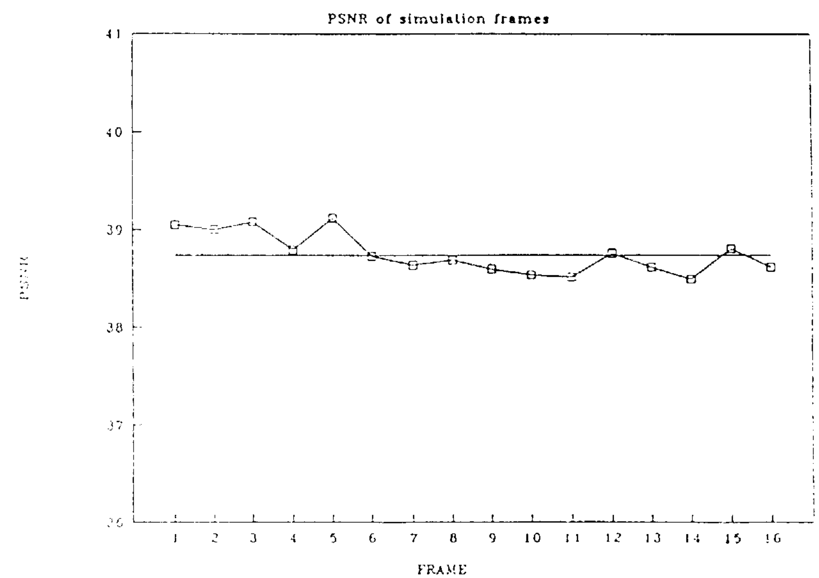

Figure 4 PSNR of simulation sequence frames.