CONCATENATED MULTIPLE DESCRIPTION CODING OF FRAME-RATE SCALABLE VIDEO

Ivan V. Bajić and John W. Woods *

Center for Next Generation Video and ECSE Department
Rensselaer Polytechnic Institute, Troy, NY 12180
ivanb@cipr.rpi.edu, woods@ecse.rpi.edu

ABSTRACT

In this paper we present a method for concatenated multiple description (MD) coding of frame-rate scalable video. The proposed method combines domain-based MD coding and FEC-based MD coding. We find that the combined system benefits from both of its components and is significantly better than either of them at higher packet loss rates.

1. INTRODUCTION

MD coding has recently emerged as a powerful framework for robust encoding and transmission of visual information over packet networks. This type of coding offers a high degree of resilience to losses and delays, the two most common channel impairments of packet based wired networks. Several methods for MD coding of images and video have been introduced in the past several years [1] - [5]. In this paper we investigate the combined use of domain-based MD coding [4] and FEC-based MD coding [5].

Domain-based MD coding is a combination of dispersive packetization and error concealment. Data is partitioned so that any missing samples are surrounded with many available neighboring samples, which enables easy error concealment. This method is bandwidth-efficient since no extra redundancy is included. It performs well for low packet loss rates. However, at higher packet loss rates, error concealment cannot recover the missing data with sufficient accuracy. FEC-based MD coding is a form of unequal error protection of embedded bitstreams by a set of Reed-Solomon codes. When properly designed, it offers very good performance for a wide range of loss rates. But as with most FEC schemes, its design relies heavily on the exact knowledge of the channel state.

In this study we have found that the two forms of MD coding discussed above can be joined in a successful symbiosis. The combined system significantly outperforms each of its components at higher packet loss rates, and is fairly insensitive to channel mismatch.

2. CONCATENATED MD CODING

The two forms of MD coding are used in a concatenated fashion, in a manner similar to the concatenated channel coding [6]. In concatenated coding, the idea is to combine two codes with different error correction capabilities, so that the outer code corrects the errors which the inner cannot correct. In our case we use domain-based MD coding as the outer code, and FEC-based MD coding as the inner code, as illustrated in Fig. 1. At the receiver, data which cannot be recovered by FEC is estimated by error concealment.

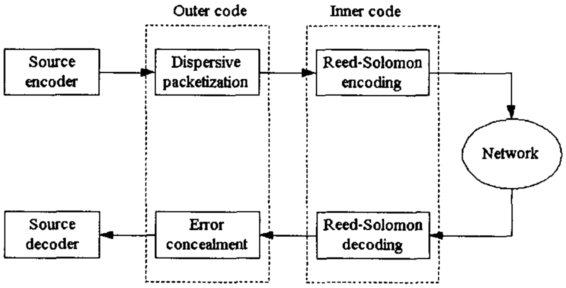

Fig. 1. Concatenated MD coding

The process of creating concatenated multiple descriptions is illustrated in Fig. 2. Video data is first dispersively packetized as described in [4]. This procedure creates a stream of packets organized into temporal scalability layers (top part of Fig. 2). The first layer provides the video at the lowest frame-rate, and other layers provide the enhancements for higher frame rates. The figure shows three temporal scalability layers, each of them packetized into three packets (descriptions). The packet stream is then partitioned into several sections, and each section is encoded by a suitable RS code. Since lower layers tend to be more important,

---

*This work was supported in part by the ARO grant DAAD 19-00-1-0559

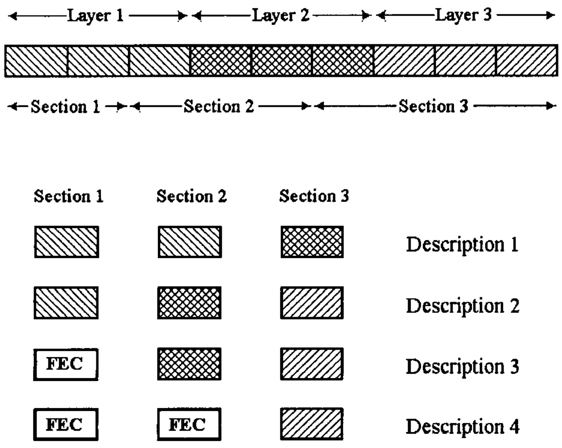

Fig. 2. FEC coding of dispersively packetized video.

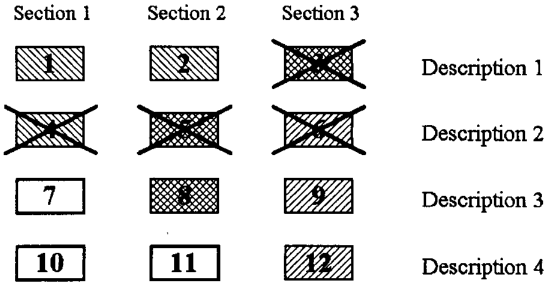

Fig. 3. Packets 3, 4, 5 and 6 are lost.

sections containing lower layers are assigned more protection. The figure shows the example where three sections are created, and packets from the three sections are organized into columns. The first section is protected by a $(4, 2)$ RS code, the second one is protected by a $(4, 1)$ RS code, while the last section is not protected. Each row of this matrix of packets represents a description, as in [5]. The order of transmission is description-by-description, i.e. row-by-row.

An example of a possible packet loss realization is shown in Fig. 3. For convenience, in this figure packets are labelled in the order of transmission. In this case four packets are lost. Packets 4 and 5 can be recovered exactly by the RS codes for sections 1 and 2. This means that all packets from layer 1 are available at the decoder, while one packet from layer 2 and one from layer 3 are missing. In contrast to the scheme in [5], in our case variable length code does not run across packet boundaries, so packets 9 and 12 from layer 3 are decodable. Data from the missing packets which cannot be recovered by FEC is then estimated through error concealment.

3. PROBLEM FORMULATION

Suppose that the frame-rate scalable video was encoded and dispersively packetized at a certain source coding rate R_s. Let $D_l^{(i)}(R_s)$ be the distortion of the video decoded from layers 1 through i, and let $N^{(i)}$ be the number of packets in layer i (determined by the rate of the data in layer i and the packet length L). Within each layer packets produced by dispersive packetization are equally important. We model this situation by considering each packet in layer i as carrying a certain amount of *reduction in distortion* given by

$$\widetilde{D}^{(i)}(R_s) = \left(D_l^{(i-1)}(R_s) - D_l^{(i)}(R_s)\right)/N^{(i)}.$$

Let N be the number of descriptions after FEC coding and M be the number of sections. Section j is composed of (N, K_j) RS codewords and it contains K_j FEC packets and $N_j = N - K_j$ data packets. We assume that the number of FEC packets decreases by 1 in each subsequent section, so $N_j = \min(N, N_1 + j - 1)$ and $K_j = \max(0, N - N_1 - j + 1)$. The total number of packets is MN and the total rate is $R_{tot} = MNL$. Reduction in distortion $\widetilde{D}_j$ in section j is the sum of reductions in distortion of all the data packets in that section.

If K_j or fewer packets are missing from section j, then all data packets from section j can be recovered by RS decoding. Let p be the packet loss probability. Assuming independent losses, the probability that K_j or fewer packets are missing from section j is given by

$$q_j = \sum_{k=0}^{K_j} \binom{N}{k} p^k (1-p)^{N-k}. \qquad (1)$$

This is the probability that section j will be recovered by FEC alone. Expected reduction in distortion in section j coming from FEC alone is $q_j \widetilde{D}_j$. If more than K_j packets are lost in section j (which happens with probability $(1-q_j)$) then FEC is not useful here, but a fraction of $(1-p)$ of data packets is still available. Expected reduction in distortion for this case is $(1-q_j)(1-p)\widetilde{D}_j$. The total expected reduction in distortion is given by

$$E[\widetilde{D}] = \sum_{j=1}^{M} \widetilde{D}_j \left(q_j + (1-q_j)(1-p)\right), \qquad (2)$$

so expected distortion at the receiver is $E[D] = D_{\max} - E[\widetilde{D}]$, where $D_{\max}$ is the maximal distortion (i.e. distortion at zero rate).

As seen in the analysis above, the expected distortion at the receiver depends on the source coding rate R_s, packet length L, number of descriptions N, number of sections M, and the number of data packets N_1 in the first section. To formulate the optimization problem, we assume that N is fixed, and optimize over the other parameters. The problem becomes

$$\begin{array}{ll} \text{minimize} & E[D], \\ \text{subject to} & R_{tot} \leq R_{\max}, \\ & L_{\min} \leq L \leq L_{\max}, \\ & 1 \leq M \leq N, \\ & 1 \leq N_1 \leq N, \end{array} \qquad (3)$$

where $R_{\max}$ is the maximal allowed rate, and $L_{\min}$ and $L_{\max}$ are, respectively, the lower and upper bound on the packet length.

In solving this problem, R_s and L may be treated as continuous variables, however M and N_1 are necessarily integers. This makes it a mixed-integer programming problem. A simple strategy for solving such problems is to solve the nonlinear optimization problem in continuous variables for each feasible combination of integer variables and pick the one which gives the lowest value of the objective function. In our implementation we constrain the packet length to a certain small set Λ of values between $L_{\min}$ and $L_{\max}$, and further constrain M to be no greater than $R_{\max}/(N \cdot L_{\min})$, since for any M greater than this value the rate constraint would be violated. Then for each allowable L, M and N_1 we solve the one-dimensional optimization problem in R_s and pick the best solution.

4. RESULTS

Results in this section are based on the grayscale Football sequence (SIF resolution, 30 fps). The coder we used is based on the invertible motion-compensated 3-D subband video coder (IMC-3DSBC) from [7]. We considered the case of three frame-rate scalability layers corresponding to full, one-half, and one-quarter frame rate. The group-of-pictures (GOP) size was set to 4 frames. All results correspond to the video decoded at the full frame rate of 30 fps.

In the optimization problem we used the following parameters: $N = 16, R_{\max} = 1.2$ Mbps, $L_{\min} = 400$ bytes and $L_{\max} = 500$ bytes. This set of parameters was targeted for Internet transmission. We then solved the problem for several values of packet loss probability p and simulated video transmission over the network at these values of p. Network transmission was simulated using a Gilbert model as in [8]. Fig. 4 shows the PSNR performance of the combined system (curve labelled FEC+EC) compared to the performance of the component systems: FEC-based MD which relies exclusively on FEC and does not use error concealment, and domain-based MD which relies exclusively on error concealment (EC) and uses no FEC. We see that the combined system outperforms its component systems by up to 1.5 dB over the range of loss rates considered.

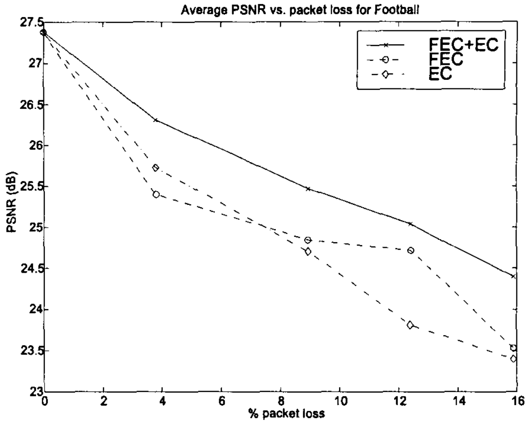

Fig. 4. Comparison of FEC-based, domain-based and concatenated MD coding on 96 frames of Football sequence.

In Fig. 5 we show what happens if there is a mismatch between the packet loss probability p used in the design and the actual packet loss rate. Here, FEC was designed for the assumed packet loss probability of 8.8%, while the actual packet loss rate in the simulations was varied. Again we see that combined system outperforms the component systems. Further, while the performance of FEC-based MD system alone drops by about 1 dB for loss rates higher than the one used in the design, the combined system suffers a loss of less than 0.2 dB.

As an illustration, Fig. 6 shows frame 13 from the Football sequence. This is a frame from the sequence encoded for 8.8% packet loss and transmitted with 12.4% average packet loss. The packet loss for this frame is 31.3%. In case of a FEC-only system, largest errors (black holes) correspond to the low temporal frequency data from packets which cannot be recovered by FEC, while other areas are relatively unaffected. In the system which relies only on error concealment, errors are dispersed evenly across the frame, but the quality of the resulting frame is low due to high loss. Combined system gives the best result.

5. CONCLUSIONS

In this paper we have described an MD coding system which uses domain-based MD coding and FEC-based MD coding in a concatenated manner. While each of the two MD coding methods has certain disadvantages, the concatenated system seems to be more robust and combines the advantages. It was shown that the combined system outperforms each of its two component systems and has low sensitivity to the assumptions about channel state.

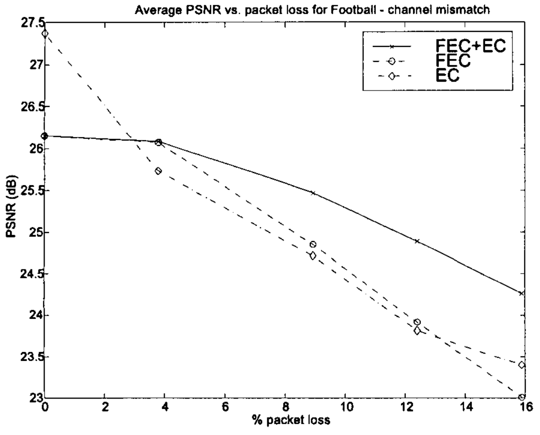

Fig. 5. PSNR performance of three MD schemes in case of channel state mismatch.

6. REFERENCES

[1] S. D. Servetto, K. Ramchandran, V. A. Vaishampayan, and K. Nahrstedt, "Multiple Description Wavelet Based Image Coding," *IEEE Trans. Image Processing*, vol. 9, no. 5, pp. 813-826, May 2000.

[2] Y. Wang, M. T. Orchard, and A. R. Reibman, "Multiple description image coding for noisy channels by pairing transform coefficients," in *Proc. IEEE Workshop on Multimedia Signal Proc.*, pp. 419-424, June 1997.

[3] V. K. Goyal, J. Kovačević, R. Arean, and M. Vetterli, "Multiple description transform coding of images," in *Proc. IEEE ICIP'98*, pp. 674-678, Oct. 1998.

[4] I. V. Bajić and J. W. Woods, "Domain-based multiple description coding of images and video," in *Proc. SPIE 4671 (VCIP'02)*, pp. 124-135, Jan. 2002.

[5] R. Puri and K. Ramchandran, "Multiple description source coding through forward error correction codes," in *Proc. 33rd Asilomar Conf. Signals, Systems, and Computers*, Oct. 1999.

[6] J. G. Proakis, *Digital communications*, 3rd edition, McGraw-Hill, 1995.

[7] S-T. Hsiang and J. W. Woods, "Invertible three-dimensional analysis/synthesis system for video coding with half-pixel-accurate motion compensation," in *Proc. SPIE 3653 (VCIP'99)*, pp. 537-546, Jan. 1999.

[8] U. Horn, K. Stuhlmüller, M. Link, and B. Girod, "Robust Internet video transmission based on scalable coding and unequal error protection," *Image Commun.*, vol. 15, no. 1-2, pp. 77-94, Sept. 1999.

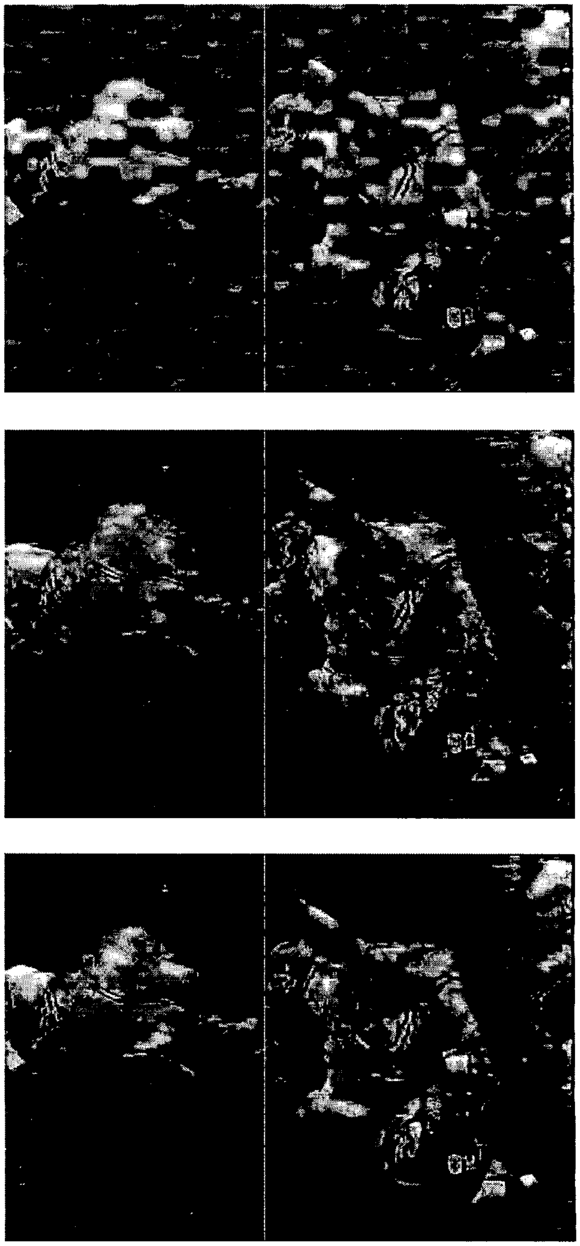

Fig. 6. Frame 13 of Football sequence. Top: FEC only (No error concealment is used here, the assumption being that FEC alone would be sufficient if the *correct* values of p were known prior to FEC design; this example, however, illustrates a real-world situation where channel state information can be incorrect.) (PSNR = 13.26 dB); middle: EC only (PSNR = 21.94 dB); bottom: FEC + EC (PSNR = 23.24 dB).