Error Concealment Aware Rate Shaping for Wireless Video Transport¹

Trista Pei-chun Chen and Tsuhan Chen² Department of Electrical and Computer Engineering, Carnegie Mellon University, PA 15213 {peichun, tsuhan}@andrew.cmu.edu

Abstract—Streaming of video, which is both source- and channel- coded, over wireless networks faces challenges of the time-varying packet loss rate and fluctuating bandwidth. Rate shaping (RS) has been proposed to reduce the bit rate of a precoded video bitstream to adapt to the real-time bandwidth variation. In our earlier work, rate shaping was extended to consider not only the bandwidth but also the packet loss rate variations. Rate-distortion optimized rate adaptation is performed on the precoded video that is a scalable coded bitstream protected by forward error correction codes. However, none prior work in rate shaping takes into account that the decoder may perform error concealment (EC) if any video data is lost during transmission. In this paper, we propose a novel rate shaping scheme that is aware of the EC method used at the decoder. We refer to this scheme as EC aware RS (ECARS). Given any precoded video, ECARS first evaluates the distortion optimization for rate adaptation. Furthermore, if the precoding process is also aware of the EC method used at the decoder, it can take advantage of the distortion measure based on the EC method, and ECARS can directly use the same distortion measure for the rate-distortion optimization. We present an example EC aware precoding process by means of macroblock prioritization. Experiment results of ECARS together with EC aware precoding are shown to have excellent performance.

I. INTRODUCTION

Due to the rapid growth of wireless communications, video over wireless networks has gained a lot of attention. Challenges as to cope with the time-varying error rate and fluctuating bandwidth bring out the need of error resilient video transport.

Joint source-channel coding techniques [1][2] are often applied to achieve error resilient video transport with online coding. However, joint source-channel coding techniques are not suitable for streaming *precoded* video. The precoded video is both source- and channel- coded prior to transmission. The network conditions are not known at the time of coding. "Rate shaping", which was called dynamic rate shaping (DRS) in [3][4], was proposed to "shape", that is, to reduce, the bit rate of the single-layered pre source-coded (pre-compressed) video, to meet the real time bandwidth requirement.

To protect the video from transmission errors in the wireless networks, source-coded video bitstream is often protected by forward error correction (FEC) codes [5]. Redundant information, known as parity bits, is added to the original source-coded bits. Conventional DRS did not consider shaping for the parity bits in addition to the source-coding bits. In our earlier work, we extended rate shaping for transporting the precoded video that is both pre source- and channel- coded [6], which we refer to as "baseline rate shaping (BRS)". The source coding in particular refers to scalable video coding as used by H.263 [7] and MPEG-4 [8]. By means of discrete rate-distortion (R-D) combination, BRS drops part of the precoded video to achieve the best video

¹ Work supported in part by Industrial Technology Research Institute.

² Corresponding author: Prof. Tsuhan Chen, Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA 15213, U.S.A. Tel: (412) 268-7536 Fax (412) 268-3890, E-mail: tsuhan@cmu.edu.

quality. The part being dropped can consist of bits from the scalable coding or the parity bits from the FEC coding.

None prior work in rate shaping, including DRS and BRS, takes into account that the decoder may perform error concealment (EC) if any video data is lost during transmission. In this paper, we propose a novel rate shaping scheme that is aware of the EC method used at the decoder. We refer to this scheme as "error concealment aware rate shaping (ECARS)". Given any precoded video, ECARS first evaluates the distortion measure considering a particular EC method used at the decoder. The distortion of not sending some part of the precoded video is large if the EC method used at the decoder cannot represent this part very well. Different EC methods will result in different distortion measures. ECARS then performs a two-stage R-D optimization for rate adaptation. The proposed two-stage R-D optimization aims for both efficiency and optimality by using model-based hyper-plane and hill-climbing based refinement.

Furthermore, if the precoding process is also aware of the EC method used at the decoder, it can take advantage of the distortion measure based on the EC method, and ECARS can directly use the same distortion measure for the rate-distortion optimization. We present an example EC aware precoding process by means of macroblock (MB) prioritization. Each MB is ranked according to how well this MB can be reconstructed by the EC method used at the decoder.

This paper is organized as follows. In Section II, we introduce baseline rate shaping (BRS) and error concealment (EC) as the background. In Section III, "error concealment aware rate shaping (ECARS)" is proposed. Given any precoded video, ECARS first evaluates the distortion measure considering a particular EC method used at the decoder followed by a two-stage R-D optimization for rate adaptation. In addition, we also introduce EC aware precoding where a MB prioritization scheme is presented. In Section IV, experiment results of ECARS together with EC aware precoding are shown. Concluding remarks are given in Section V.

II. BACKGROUND

We will give brief descriptions of baseline rate shaping (BRS) and error concealment (EC) in this section.

A. Baseline Rate Shaping (BRS)

There are three stages for transmitting the video from the sender to the receiver: (i) precoding, (ii) streaming with BRS, and (iii) decoding, as shown from Figure 1 to Figure 3.

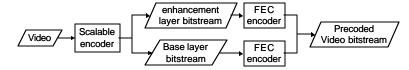


Figure 1. System diagram of the precoding process: scalable encoding followed by FEC encoding

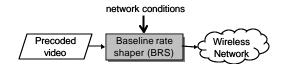


Figure 2. Transport of the precoded video with BRS

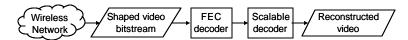


Figure 3. System diagram of the decoding process: FEC decoding followed by scalable decoding

BRS reduces the bit rate of each *decision unit* of the precoded video before it sends the precoded video to the wireless network. A decision unit can be a frame, a macroblock, etc., depending on the granularity of the decision. We use a frame as the decision unit herein. Let us consider the case in which the video sequence is scalable coded into two layers: one base layer and one enhancement layer. These two layers are FEC coded with unequal packet loss protection (UPP) capabilities. Therefore, there are four segments in the precoded video. The first segment consists of the bits of the base layer video bitstream (upper left segment of Figure 4 (a)). The second segment consists of the bits of the enhancement layer video bitstream (upper right segment of Figure 4 (a)). The third segment consists of the parity bits for the base layer video bitstream (lower left segment of Figure 4 (a)). The fourth segment consists of the parity bits for the enhancement layer video bitstream (lower right segment of Figure 4 (a)). BRS decides a subset of the four segments to send. When the channel has abundant bandwidth, BRS will send with the configuration shown in Figure 4 (a). When the bandwidth is reduced, the second configuration shown in Figure 4 (b) is chosen. When the bandwidth is reduced even more, either Figure 4 (c) or Figure 4 (d) will be chosen depending on the wireless network condition. A rule of thumb is to choose parity bits to send instead of bits of the enhancement layer when the packet loss rate is high. Interested readers can read more from [6], which consists of BRS by mode decision that we just describe and the discrete R-D combination.

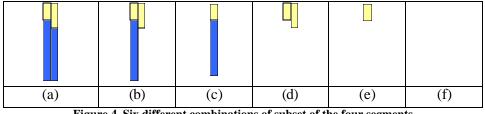


Figure 4. Six different combinations of subset of the four segments

B. Error Concealment (EC)

Error concealment relies on some a priori to reconstruct the lost video content. Such a prior can come from spatial or temporal neighbors. For example, we can assume that the pixel values are smooth across the boundary of the lost and retained regions. To recover lost data with the smoothness assumption, interpolation or optimization based on certain objective functions are often used. Figure 5 and Figure 6 show corrupted frames and the corresponding reconstructed frames. The black regions in Figure 5 (a) and Figure 6 (a) indicate losses of the video data. Figure 5 shows an error concealment method using spatial interpolation from the neighboring pixels. Figure 6 shows an error concealment method using temporal interpolation. That is, if some pixel values are lost, the decoder copies the pixel values from the previous frame at the corresponding locations to the current frame. The error concealment method using temporal interpolation can be extended to copying the pixel values from the previous frame at the motion-compensated locations. The motion vectors used for motion compensation either are assumed error-free or can be estimated at the decoder [10][11].

We use the simple temporal interpolation method in this paper. Future extension includes using motion-compensated temporal interpolation, or more sophisticated error concealment methods as mentioned in [9].



Figure 5. Error concealment example by spatial interpolation: (a) the corrupted frame without error concealment, and (b) the reconstructed frame with error concealment



Figure 6. Error concealment example by temporal interpolation: (a) the corrupted frame without error concealment, and (b) the reconstructed frame with error concealment

III. ERROR CONCEALMENT AWARE RATE SHAPING (ECARS)

In this section, we will start from describing the system of wireless video transport, including precoding, streaming with rate shaping, and decoding. We then introduce the new EC aware RS scheme (ECARS), which first evaluates the distortion measure considering a particular EC method used at the decoder then performs the two-stage R-D optimization. In addition, if the system allows for EC aware precoding, ECARS can take advantage of that. We will present an EC aware precoding process by means of MB prioritization.

There are three stages to transmit the video from the sender to the receiver in a wireless video transport system: (i) precoding, (ii) streaming with rate shaping, and (iii) decoding, as shown from Figure 7 to Figure 9.

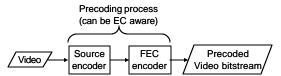


Figure 7. System diagram of the precoding process: MB prioritization followed by source encoding and FEC encoding

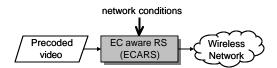


Figure 8. Transport of the precoded video with ECARS

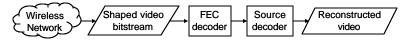


Figure 9. System diagram of the decoding process: FEC decoding followed by source decoding

A. R-D Optimization for ECARS

Given the precoded video, which is both source- and channel coded, ECARS will perform bandwidth adaptation for streaming. We start from a simple example as an extension to BRS then give a more general ECARS.

Let us consider that the precoded video consists of two layers of video bitstream, namely, the base layer and the enhancement layer. Each layer is protected by some parity bits from the FEC coding. The setting is shown earlier in Figure 4 (a). The rate shaper is extended to give a finer decision on how many symbols³ to send (or how many symbols to drop) for each layer, instead of deciding which segment(s) to drop. Since the rate shaper is aware of the EC method used at the decoder, it can evaluate how much distortion it will result in

³ "Symbols" are used instead of "bits" since the FEC codes use a symbol as the encoding/decoding unit. In this paper, we use 14 bits to form one symbol. The selection of symbol size in bits depends on the user.

if the rate shaper decides to send a certain amount of symbols for each layer. In other words, the rate shaper can evaluate how much gain it will get if t decides to send this certain amount of symbols for each layer. In general, the base layer can be reconstructed well with error concealment since the base layer consists of coarse information of the video that can be easily reconstructed. On the other hand, the enhancement layer, which consists of fine details of the video, cannot be easily reconstructed. The EC aware rate shaper may assign a higher gain on sending symbols in the enhancement layer than the symbols in the base layer.

Having understood how the gain of sending some part of the precoded video is determined considering the EC used at the decoder, we can now introduce a more general case of ECARS. Suppose ECARS is given the precoded video, which is composed of several *sublayers*. Each sublayer consists of symbols from source coding, which is shown as the upper portion of each stripe in Figure 10 (a), and symbols from channel coding, which is shown as the lower portion of each stripe in Figure 10 (a). Note that the construction of the sublayers does not necessarily consider the EC method of the decoder. That is, precoding and ECARS processes are not necessarily coupled. They are considered separately. ECARS simply performs rate shaping on the *given* precoded video. We will consider that the precoding process constructs the sublayers with the knowledge of the EC method of the decoder in the second part of this section III.B.

The darken bars in Figure 10 (b) represent the symbols to be sent by ECARS. We will see how ECARS is done.

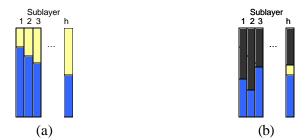


Figure 10. (a) Precoded video in sublayers and (b) ECARS decision on which symbols to send

Let us start with the problem formulation. The total gain is increased (or the total distortion is decreased) as more sublayers are correctly decoded. With Sublayer 1 correctly decoded, the total gain is increased by G_1 (accumulated gain is G_1); with Sublayer 2 correctly decoded, the total gain is increased further by G_2 (accumulated gain is $G_1 + G_2$); and so on. Note that G_i of Sublayer *i* is calculated given the EC method used at the decoder. For example, G_i can be the square of the residue between the true sublayer and its error-concealed version without sending this sublayer. G_i is different with different EC method. Therefore, G_i is EC aware. The expected accumulated gain is then:

$$G = \sum_{i=1}^{h} G_i v_i \tag{1}$$

if each sublayer can be decoded independently⁴. v_i is the recovery rate of Sublayer *i*. With Reed-Solomon codes used in this paper, Sublayer *i* is recoverable (or successfully decodable) if the number of erasures is no more than $r_i - k_i$. k_i is the message (symbols from the source coding) size in Sublayer *i* and r_i is the number of symbols selected to send in Sublayer *i*. Thus, the recovery rate v_i is the summation of the probabilities that no loss occur, one erasure occurs, and so on until $r_i - k_i$ erasures occur.

⁴ If Sublayer *i* can be decoded only if Sublayer *i*-1 is decoded correctly, (1) can be modified to $G = \sum_{i=1}^{n} G_{i} \prod_{j=1}^{i} v_{j}$. Other types of dependency in decoding will give more complex formulation of *G*.

$$v_{i} = \sum_{l=0}^{r_{i}-k_{i}} \left[\binom{r_{i}}{l} e_{sym} \right]^{j} \left(1 - e_{sym} \right)^{r_{i}-l} , \quad i = 1 \sim h$$
⁽²⁾

where *h* is the number of sublayers of this frame in total and e_{sym} is the symbol loss rate. If the packet loss rate is small, the symbol loss rate can be approximated by the packet loss rate divided by the number of symbols per packet $e_{sym} \approx e_p / (s/m)$, where *s* is the packet size and *m* is the symbol size in bits. By choosing different combinations of the number of symbols for each sublayer, the expected accumulated gain will be different. The rate shaping problem can be formulated as follows:

maximize
$$G = \sum_{i=1}^{h} G_i v_i$$

subject to $\sum_{i=1}^{h} r_i \le B$ (3)

To solve this problem, we propose a new two-stage R-D optimization approach. The two-stage R-D optimization first finds the near-optimal solution globally. The near-optimal global solution is then refined by a hill climbing approach.

Let us start from Stage 1. We can see from (1) and (2) that *G* is related to $\mathbf{r} = [r_1 \quad r_2 \quad \cdots \quad r_h]$ implicitly through the recovery rates $\mathbf{v} = [v_1 \quad v_2 \quad \cdots \quad v_h]$. We can instead find a model-based hyper-plane that explicitly relates \mathbf{r} and *G*. The model parameters can be trained from a set of (\mathbf{r}, G) values, where \mathbf{r} values are chosen by the user and *G* values can be computed by (1) and (2). The optimal solution is then the intersection of this hyper-plane and the bandwidth constraint as illustrated in Figure 11. The complexity of the model determines the preciseness of the model in finding the optimal solution.

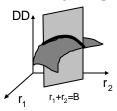


Figure 11. Intersection of the model-based hyper-plane and the bandwidth constraint, illustrated with h = 2

Stage 1 of the two-stage R-D optimization gives a near-optimal solution. The solution can be refined by a hill-climbing based approach (Figure 12). We perturb the solution from Stage 1 to yield a larger accumulated gain under the bandwidth constraint. The process can be iterated until the solution reaches a stopping criterion such as the convergence.

```
While (stop == false)
    z_i = r_i for all i=1~h
    For (j=1; j<=h; j++)
        For (k=1; k<=h; k++)
            z_k = z_k + delta for k==j //Increase sublayer j
             z_k = z_k - delta/(h-1) for k!=j //Decrease others
        End-for
        Evaluate G, by equations (1) and (2)
    End-for
    Find the j* with the largest G_{j*}.
    For (i=1; i<=h; i++)</pre>
        r_i = r_i + delta \text{ for } i==j^*
        r_i = r_i - delta/(h-1) for i!=j*
    End-for
    Calculate the stop criterion.
End-while
```

Figure 12. Pseudocodes of hill-climbing algorithm

Note that we can also consider an even more general case where the gain has effectiveness across frames. That is, the frames are inter-coded and thus the gains need to be evaluated jointly among frames. This paper only considers the case where all frames are intra-coded and can be extended in this direction.

B. Error Concealment Aware Precoding

If the precoding process is also aware of the EC method used at the decoder, it can take advantage of the distortion measure based on the EC method, and ECARS can directly use the same distortion measure for the rate-distortion optimization. We present an EC aware precoding process by means of macroblock (MB) prioritization. Each MB is ranked according to how well this MB can be reconstructed by the EC method used at the decoder. That is, we can use square sum of the pixel differences between the original MB and the reconstructed MB as the measure for priority. The larger the value, the higher the priority is. More sophisticated error concealment methods can be applied. The resulting priorities of the MB need to be modified according to different error concealment methods.

An observation to make is that the conventional video coding can be considered as a special case of the proposed EC aware MB prioritization. Let us consider the case where no motion vector is used in video coding. The MB with large residues is encoded and transmitted, while the MB with small residues does not need to be transmitted since the small residues will become zero after quantization. This case translates to the case of EC aware MB prioritization using temporal interpolation with zero motion vectors. Let us consider another case where motion vectors are included in video coding. This then translates to the case of EC aware MB prioritization using temporal interpolation vectors. We can see that the proposed EC aware MB prioritization is more general since it is not limited to any specific error concealment method.

The source-coded bitstream with EC aware MB prioritization can be added with parity bits from the FEC coding. First, the bits of the highest priority MB is placed followed by the bits of the second highest priority MB and so on, as shown in Figure 13 (a). The bits are further divided into sublayers as shown in Figure 13 (b). Sublayer i+1 is longer than Sublayer i since we want to achieve UPP for the sublayers when appended with the parity bits. For example, we can let Sublayer 1 consists of bits from the first 10 highest priority MB, Sublayer 2 consists of bits from the following 20 highest priority MB, and so on. Each sublayer is then appended with parity bits from the FEC coding as shown in Figure 13 (c).

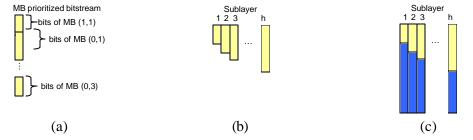


Figure 13. Precoded video: (a) MB prioritized bitstream, (b) MB prioritized bitstream in sublayers, and (c) FEC coded MB prioritized bitstream

Note again that ECARS can perform rate adaptation with or without EC aware precoding. To summarize, the proposed ECARS achieves the best performance with the two-stage R-D optimization. Compare to the convention video coding, the video encoder seeks, for example, the quantization level that achieves a certain bit rate (accordingly, some MB are not transmitted). No FEC codes are included in the rate adaptation process. Given the same the bandwidth requirement, ECARS may choose fewer MB to send and include some parity symbols to protect them, as opposed to conventional video coding that sends more MB but not any parity symbols.

IV. EXPERIMENT

In the experiment, we will show results of the proposed ECARS together with EC aware precoding, compared with the naïve rate shaping method "unequal error protection rate shaping (UPPRS)" described in Figure 14 (a). UPPRS will drop from the bottom if the bandwidth is not enough. In that, UPP can be achieved since more parity symbols are sent for Sublayer *i* than Sublayer i+1. Wireless networks are generally with time-varying packet loss rate and fluctuating bandwidth. The packet loss rate and bandwidth vary at each time interval. We simulate random bandwidth fluctuation and use a two-state Markov-chain [12] to simulate the bursty bit errors. Example traces of simulated bandwidth and packet loss rate are shown in Figure 14 (b)(c). Each interval in the axis of time index represents 0.033 sec. The test video sequences are "akiyo", "foreman", and "stefan" in common intermediate format (CIF) (Figure 15 (a)-(c)). We use H.263 [7] for video encoding. The frame rate is 30 frames/sec. All results in the following are shown for the luminance Y components only.

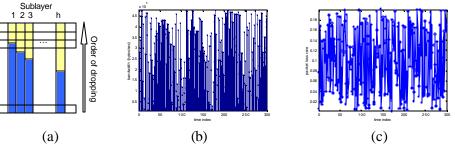


Figure 14. Experiment setup: (a) UPPRS, (b) Bandwidth, and (c) packet loss rate



Figure 15. Test video sequences in CIF: (a) akiyo, (b) foreman, and (c) stefan

Figure 16 shows the EC aware precoding by MB prioritization for Sequence "stefan". A MB is more important than the others are, if its square sum of the pixel differences between the original MB and the reconstructed MB is larger. The brighter the MB, the larger the square sum is, and hence the higher the MB priority is. The region that represents the tennis player is shown with a high priority.

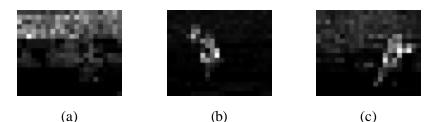


Figure 16. EC aware MB prioritization of Sequence "stefan" in (a) Frame 2, (b) Frame 32, and (c) Frame 122

PSNR result of Sequence "stefan" is shown in Figure 17. The overall PSNR performance for all three sequences is shown in Figure 18. We can see that the proposed ECARS performs better than UPPRS. The improvement of ECARS over UPPRS is the most significant in Sequence "stefan" followed by Sequence "foreman" and "akiyo". Sequence "stefan" is difficult to be reconstructed well by the error concealment if the

video data is lost during the transmission. It is more crucial to send the right combination of symbols that is aware of the EC at the decoder. Therefore, the improvement of ECARS over UPPRS is more prominent.

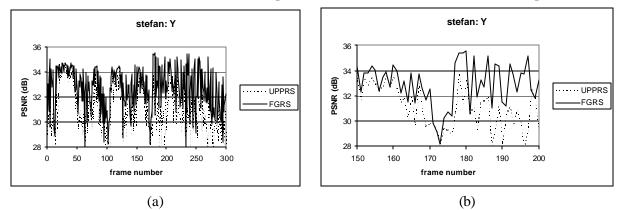


Figure 17. Frame by frame PSNR of UPPRS and ECARS with Sequence "stefan": (a) result from Frame 1 to Frame 300, (b) zoomed result from Frame 150 to Frame 200

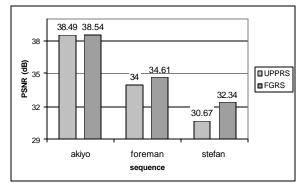


Figure 18. Overall PSNR of UPPRS and FGRS with sequences "akiyo", "foreman", and "stefan"

To examine how ECARS outperforms UPPRS, we look at the MB recovery rates of all the MB in Frame 2, Frame 32, and Frame 122, as shown in Figure 19 and Figure 20. The brighter the MB, the higher the chance it can be received without errors. We can see that Figure 20 resembles Figure 16 better than Figure 19 does. With ECARS, the MB that is with higher priority indeed gets higher recovery rate.

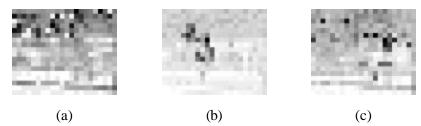


Figure 19. MB loss recovery rates in (a) Frame 2, (b) Frame 32, and (c) Frame 122 using UPPRS approach

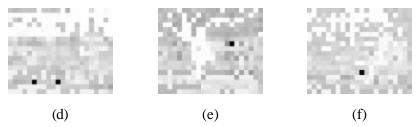


Figure 20. MB loss recovery rates in (a) Frame 2, (b) Frame 32, and (c) Frame 122 using ECARS

V. CONCLUSION

We proposed in this paper a novel error concealment aware rate shaping scheme (ECARS) for video transport over wireless networks. Given any precoded video, ECARS first evaluates the distortion measure considering a particular EC method used at the decoder. ECARS then performs a two-stage R-D optimization for rate adaptation. The proposed two-stage R-D optimization obtains the near-optimal solution by finding the intersection of the model-based hyper-plane and the bandwidth constraint, and then refines the solution by a hill-climbing based approach. The two-stage R-D optimization aims for both the efficiency and optimality. Furthermore, if the precoding process is also aware of the EC method used at the decoder, it can take advantage of the distortion measure based on the EC method, and ECARS can directly use the same distortion measure for the rate -distortion optimization. The proposed ECARS outperforms the naïve UPPRS approach in the experiment. Future work includes investigating how ECARS performs on the precoded video that is constructed unaware of the EC method used at the decoder, building more sophisticated model of expected gain that considers the dependency between individual gains G_i , and using more advanced EC methods.

VI. REFERENCE

- [1] G. Cheung and A. Zakhor, "Bit Allocation for Joint Source/Channel Coding of Scalable Video", *IEEE Transactions on Image Processing*, 9(3), March 2000.
- [2] L. P. Kondi, F. Ishtiaq, and A. K. Katsaggelos, "Joint Source-Channel Coding for Motion-Compensated DCT-based SNR Scalable Video", *IEEE Transactions on Image Processing*, 11(9), September 2002.
- [3] W. Zeng and B. Liu, "Rate Shaping by Block Dropping for Transmission of MPEG-precoded Video over Channels of Dynamic Bandwidth", *ACM Multimedia 96*, Boston, MA, U.S.A, 1996.
- [4] S. Jacobs and A. Eleftheriadis, "Streaming Video Using Dynamic Rate Shaping and TCP Congestion Control", *Journal of Visual Communication and Image Representation*, 9(3), 211-222, 1998.
- [5] S. Wicker, Error Control Systems for Digital Communication and Storage, Prentice-Hall, 1995.
- [6] T. P.-C. Chen and T. Chen, "Adaptive Joint Source-Channel Coding using Rate Shaping", *ICASSP 2002*, Orlando, FL, U.S.A., May 2002.
- [7] D. S. Turaga and T. Chen, "Fundamentals of Video Compression: H.263 as an Example", in Compressed Video over Networks, edited by M.-T. Sun and A. R. Reibman, Marcel Dekker, Inc., 2001.
- [8] Motion Pictures Experts Group, "Overview of the MPEG-4 Standard", ISO/IEC JTC1/SC29/WG11 N2459, 1998.
- [9] Trista Pei-chun Chen and Tsuhan Chen, "Second-Generation Error Concealment for Video Transport over Error Prone Channels", *Wireless Communications and Mobile Computing*, Special Issue on Multimedia over Mobile IP, October 2002.
- [10] W. M. Lam, A. Reibman, and B. Liu, "Recovery of Lost or Erroneously Received Motion Vectors", *ICASSP 1993*, vol. 5, pp. 417-420.
- [11] M. E. Al-Mualla, N. Canagarajah, D. R. Bull, "Multiple-reference temporal error concealment", *ISCAS 2001*, vol. 5, pp. 149-152.
- [12] F. Alajaji and T. Fuja, "A Communication Channel Modeled on Contagion", *IEEE Transactions on Information Theory*, 40(6), pp. 2035-2041, 1994.