Robust Video Coding for Packet Networks with Feedback

Rui Zhang, Shankar L. Regunathan and Kenneth Rose
Department of Electrical and Computer Engineering
University of California
Santa Barbara, CA 93106

Abstract

Robustness to packet loss is a critical requirement for video communication over packet switched networks. Intra-coding is an important tool to mitigate the effects of packet loss by limiting error propagation. This work proposes an algorithm for optimal intra/inter coding mode selection, while utilizing the feedback channel that carries acknowledgement information about received packets. The overall distortion in frame reconstruction at the decoder due to quantization, error concealment (after packet loss) and error propagation is estimated at *pixel-level* precision, and is dynamically refined based on the feedback information. The estimate is then integrated into a rate-distortion (RD) framework for optimal selection of coding mode for each macroblock. Simulation results demonstrate that precise distortion estimation enables the coder to achieve substantial and consistent gains in PSNR over known state-of-the-art feedback-based mode selection methods.

I. Introduction

Video coder design is facing major new challenges now that packet-switched networks, such as the Internet, have become overwhelmingly important. These networks can currently provide limited or no end-to-end QoS guarantees. It is also anticipated that wireless extensions to the wired backbone will result in additional packet losses. Thus, robustness to packet loss has become a critical requirement for video compression algorithms.

The standard video coding scheme employs inter-frame prediction (with motion compensation) to remove temporal redundancies. Although this prediction loop is known to achieve good compression efficiency, it is more sensitive to channel errors

This work was supported in part by the University of California MICRO program, Cisco Systems, Inc., Conexant Systems, Inc., Dialogic Corp., Fujitsu Laboratories of America, Inc., General Electric Co., Hughes Network Systems, Lernout & Hauspie Speech Products, Lockheed Martin, Lucent Technologies, Inc., Qualcomm, Inc., and Texas Instruments, Inc.

as it promotes error propagation in both the spatial and temporal domains. In many applications, a feedback channel from the receiver to the transmitter is available. The acknowledgement information provided by the feedback channel can be used to provide robustness to packet loss via techniques such as Automatic Repeat on reQuest (ARQ) [1] [2]. Refinements of ARQ schemes include retransmission of highly compressed video [3] or retransmission of multiple copies of the lost packet in a single retransmission trial [4]. However, retransmission schemes introduce additional delays that are undesirable in real-time applications. Other feedback based techniques for robustness include rate control [5], error tracking [6], error confinement [7] and reference picture selection [8] [9]. Refer to [10] for a review of such techniques.

In this work, we focus on optimal switching between intra-mode and inter-mode coding for each macroblock (MB), depending on the feedback information, to provide resilience to packet loss. By switching off the inter-frame prediction loop for certain MBs, the reproduced blocks are no longer dependent on past frames and error propagation is stopped. This is a very straightforward and standard-compatible approach, which incurs no additional delay. However, intra-coding typically requires higher bit rate than inter-coding and too many intra-coded MBs will significantly degrade the compression efficiency. Thus, the problem of coding mode selection, to achieve the right balance between compression and robustness, is very important and has been studied earlier in the context of networks with feedback. In [6] [11], mode selection strategies were proposed to intra-code those MBs that have been severely distorted due to packet loss. The main drawback of these methods is that they lack the ability to accurately estimate, at the encoder, the overall distortion of the decoder frame reconstruction. Further, they rely on heuristic thresholds to make their mode decisions.

The contribution of this paper is two fold. First, we develop an algorithm for *recursive* computation of the overall distortion of the decoder frame reconstruction at *pixel level precision.* This algorithm was first proposed in [12] in the context of channels without feedback. In this work, the estimate is dynamically refined based on the feedback information received by the encoder and accurately accounts for the effects of quantization, error concealment and error propagation. We then incorporate this estimate within a rate-distortion (RD) framework to select the coding mode such that the trade-off between compression efficiency and robustness is optimized.

Simulation results demonstrate substantial and consistent gains in PSNR over known state-of-the-art feedback-based mode selection methods. Refer to [13] for a more comprehensive description of the algorithm.

The paper is organized as follows: Section II states the problem of optimal mode selection with feedback information. The Recursive Optimal per-Pixel Estimate (ROPE) of the overall distortion at the decoder, and its incorporation within an RD framework (ROPE-RD) are introduced in section III. Simulation results demonstrating the performance of the ROPE-RD method are given in section IV.

II. Mode Selection Based on Feedback Information

The feedback information is usually not part of the video syntax, but transmitted in a different layer of the protocol stack where control information is exchanged. For example, in the H.324 low bit rate multimedia communication system [14], the H.245 control protocol [15] allows reporting of the temporal and spatial location of MBs that could not be decoded successfully. In the H.323-based packet video systems [16], RTCP [17] provides control information through a backward channel [2]. Compared to the video bit rate in the forward channel, the bit rate required for sending feedback is modest. Since, in general, the control information is transmitted using a retransmission protocol, error-free reception of feedback information is normally guaranteed. In the sequel we will assume reliable transmission of feedback information.

Note that the feedback information arrives at the encoder with some delay. Let the round trip delay be specified by the number of frames, d, encoded during its duration. After the encoder finishes encoding the n-th frame, it receives the feedback information for the $(n-d)$-th frame. In this case, the encoder has access to the exact, albeit delayed, status of the decoder. In other words, the encoder knows the loss status of each MB at, and prior to, frame $(n-d)$. Consequently, it can exactly compute the decoder reconstruction at, and prior to, frame $(n-d)$, as long as we may assume that the error concealment method employed by the decoder is known. However, the packet loss history from $(n-d+1)$ to frame n remains unknown to the encoder at this point. The decoder reconstruction of frames $(n-d+1), (n-d+2), ..., n$ must be treated by the encoder as a random signal.

The effect of packet loss propagates in the spatial direction beyond MB boundaries due to motion compensation. Thus, only by computing the distortion per each individual pixel can we accurately account for error propagation. Further, note that the distortion due to quantization and the distortion due to concealment are not simply additive. Instead, they are combined in a highly complex fashion to produce the overall distortion. In the next section, we derive an algorithm to recursively compute the overall decoder distortion at pixel level precision, for the given packet loss rate and error concealment technique in use at the decoder. The proposed estimate is dynamically refined to account for new feedback information. This estimate is integrated within an RD framework to select coding modes that optimize the tradeoff between the overall rate and overall distortion.

III. Optimal Mode Selection for Robustness

A. Preliminaries

We form a group of blocks (GOB) from all the MBs in a particular row (slice), and assume that each GOB is carried in a separate packet. In this setting, the loss rate of a pixel equals the packet loss rate p. We assume that the packet loss rate, p, is available at the encoder. This can be either specified as part of the initial negotiations, or adaptively calculated from information provided by the transmission protocol such as RTCP [17].

If a packet is lost, the decoder performs error concealment. Here we use the temporal replacement method as follows: The motion vector of a lost MB is estimated as the median of the motion vectors of the nearest three MBs in the previous GOB (above). When the previous GOB is also lost, the estimated motion vector is set to zero. The missing pixels are replaced by the corresponding pixels in the previous frame.

B. Recursive Optimal per-Pixel Estimate of the Distortion

Let f_n^i denote the original value of pixel i in frame n, and let $\hat{f}_n^i$ denote its *encoder* reconstruction. The reconstructed value at the *decoder*, possibly after error concealment, is denoted by $\tilde{f}_n^i$. For the encoder, $\tilde{f}_n^i$ is a random variable. Using the mean square error as distortion metric, the overall expected distortion for this pixel is

$$d_n^i = E\{(f_n^i - \tilde{f}_n^i)^2\} = (f_n^i)^2 - 2f_n^i E\{\tilde{f}_n^i\} + E\{(\tilde{f}_n^i)^2\}. \tag{1}$$

The computation of d_n^i requires the first and second moments of each random variable in the sequence $\tilde{f}_n^i$. We develop recursion functions to sequentially compute these two moments. For the recursion step, we consider two cases depending on whether the pixel belongs to an intra-coded MB or an inter-coded MB.

B.1 Pixel in an Intra-coded MB:

When the packet containing the intra-coded MB to which the pixel i belongs is received correctly, we have $\tilde{f}_n^i = \hat{f}_n^i$, and the probability of this event is $1-p$. If the packet is lost, the estimated motion vector is used to associate pixel i in the current frame with pixel k in the previous frame when the previous GOB is available, or pixel i in the previous frame otherwise. We thus have $\tilde{f}_n^i = \tilde{f}_{n-1}^k$ with probability of $p(1-p)$, and $\tilde{f}_n^i = \tilde{f}_{n-1}^i$ with probability p^2. Thus:

$$\begin{aligned} E\{\tilde{f}_n^i\} &= (1-p)(\hat{f}_n^i) \\ &+ p(1-p)E\{\tilde{f}_{n-1}^k\} + p^2 E\{\tilde{f}_{n-1}^i\}, \quad (2) \\ E\{(\tilde{f}_n^i)^2\} &= (1-p)(\hat{f}_n^i)^2 \\ &+ p(1-p)E\{(\tilde{f}_{n-1}^k)^2\} \\ &+ p^2 E\{(\tilde{f}_{n-1}^i)^2\}. \quad (3) \end{aligned}$$

B.2 Pixel in an Inter-coded MB:

Let the true motion vector of the MB be such that pixel i is predicted from pixel j in the previous frame, and let the quantized residue denoted by $\hat{e}_n^i$. Thus we have $\hat{e}_n^i = \hat{f}_n^i - \hat{f}_{n-1}^j$. But even when the residue is received correctly, the decoder reconstruction of pixel i is still given by $\tilde{f}_n^i = \hat{e}_n^i + \tilde{f}_{n-1}^j$. This explains how the error propagates even if current packets are correctly received. If the packet is lost, the decoder performs error concealment in a manner identical to that of an intra-coded MB. Thus:

$$\begin{aligned} E\{\tilde{f}_n^i\} &= (1-p)(\hat{e}_n^i + E\{\tilde{f}_{n-1}^j\}) \\ &+ p(1-p)E\{\tilde{f}_{n-1}^k\} + p^2 E\{\tilde{f}_{n-1}^i\}, \\ E\{(\tilde{f}_n^i)^2\} &= (1-p)E\{(\hat{e}_n^i + \tilde{f}_{n-1}^j)^2\} \\ &+ p(1-p)E\{(\tilde{f}_{n-1}^k)^2\} \\ &+ p^2 E\{(\tilde{f}_{n-1}^i)^2\}. \quad (4) \end{aligned}$$

We reemphasize that these recursions are performed at the *encoder* in order to calculate the expected distortion at the *decoder* precisely per pixel. The estimate is precise for integer-pixel motion estimation. In the half-pixel case, the use of bilinear interpolation makes the exact computation of the second moment highly complex. However, we found the estimate is well approximated by the simpler recursion of integer-pixel motion compensation and substantial gains are nevertheless maintained.

C. ROPE Refined by Feedback Information

We now extend the ROPE estimator to the feedback case. After getting the acknowledgement with delay d, the encoder first computes exactly the $(n-d)$-th frame of decoder reconstruction, by employing error concealment wherever a block was lost. Then this reconstructed frame is used to initialize the recursion formulas to compute the first and second moments of the decoder reconstruction of frames $(n-d+1), (n-d+2), ..., n$. Thus, the information obtained via feedback is fully utilized in refining the estimate of the overall distortion at the decoder precisely at the pixel level. Possible losses from frame $(n-d+1)$ to frame n are taken into account in the expectation. This refined estimate is then incorporated into the rate-distortion Lagrangian cost and used to optimize mode selection as is explained in the next subsection. Moreover, besides tracking the decoder status, the encoder can also adapt to variations in packet loss rate p, according to the available feedback information. This helps to track the network condition and decreases the possibility of mismatch in packet loss rate.

D. RD-based Mode Selection Algorithm

We incorporate the estimated overall distortion, as computed by the ROPE model, within the rate-distortion framework in order to automatically choose the *number* and the *location* of the intra-coded MBs. This enables minimization of the overall distortion (including compression and concealment) for the given packet loss and bit rate. We refer to the resulting technique as ROPE-RD.

The rate-distortion problem can be recast as an unconstrained Lagrange minimization ($J = D + \lambda R$), where λ is the Lagrange multiplier. Since individual MB contributions to this cost are additive, we choose the optimal encoding mode for each MB independently by a simple minimization:

$$\min_{\text{mode}}(J_{\text{MB}}) = \min_{\text{mode}}(D_{\text{MB}} + \lambda R_{\text{MB}}) \tag{5}$$

where the distortion of the MB is the sum of the distortion contributions of the individual pixels:

$$D_{\mathrm{MB}} = \sum_{i \in \mathrm{MB}} d_n^i. \tag{6}$$

The ROPE model is used to calculate the distortion *per pixel*, and then decide on the coding mode *per MB* via (5).

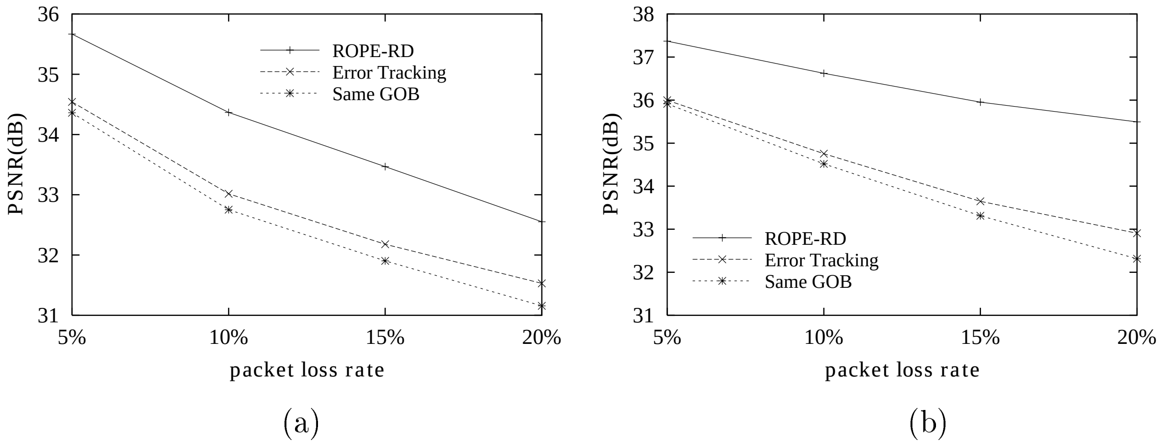
For each MB, the *mode* and the *quantization step size* are selected to minimize the rate-distortion Lagrangian.

IV. Simulation Results

We implemented the ROPE-RD mode selection strategy by appropriately modifying the Telenor H.263 codec [18]. We assume the RTP payload format for packetizing the H.263 video stream [19], and that each packet contains only one GOB. A random packet loss generator is used to drop packets at a specified loss rate. The temporal-replacement method for error concealment stated in subsection III-A is used in all the competing methods, and the rate control scheme of section III-D is used in ROPE-RD. The peak signal to noise ratio (PSNR) of the reconstruction is computed for each frame and averaged over the whole sequence. We average the PSNR over 30 different channel realizations (with different packet loss patterns).

We compressed 200 frames from each of the two QCIF video sequences *carphone* and *grandma*. We compare the proposed ROPE-RD method with the "same GOB" method [6] [11] and "error tracking" method [6]. Simulation results are presented for various bit rates, packet loss rates and feedback delays.

In the "same GOB" method, a forced intra-mode refreshment is employed to the region where a loss had occurred. The method ignores spatial error propagation due to motion compensation during the round trip delay. The "error tracking" approach in [6] intra-updates MBs whose "error energy" is greater than a predefined threshold. The "error energy" of a MB is initialized as the sum of absolute differences between the original block and the reconstructed block whenever the MB is reported as lost, and updated through temporal and spatial propagation given the motion vectors. While the method takes into account spatial error propagation, the estimate is imprecise as the updates are at the MB level. The "error tracking" method also uses heuristic thresholds to make mode decisions. During our simulations, we used the threshold of 200. Further, both the "same GOB" and "error tracking" methods ignore the effects of possible additional packet losses from frame $(n-d+1)$ to frame n.



(a) (b)

Fig. 1. PSNR vs. packet loss rate. Methods: ROPE-RD(proposed), Error Tracking [6], Same GOB [6] [11]. Bit rate=300kbps, frame rate=30fps, delay=500ms. Sequences: (a) *carphone*, (b) *grandma*.

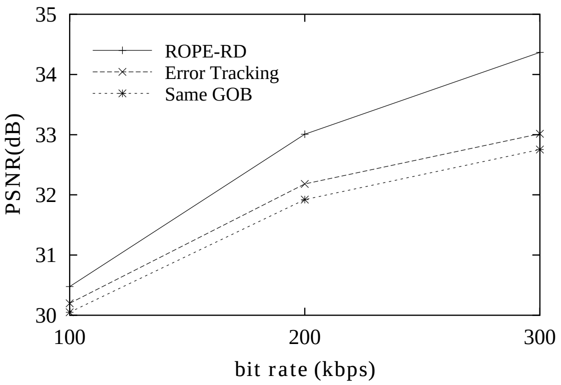
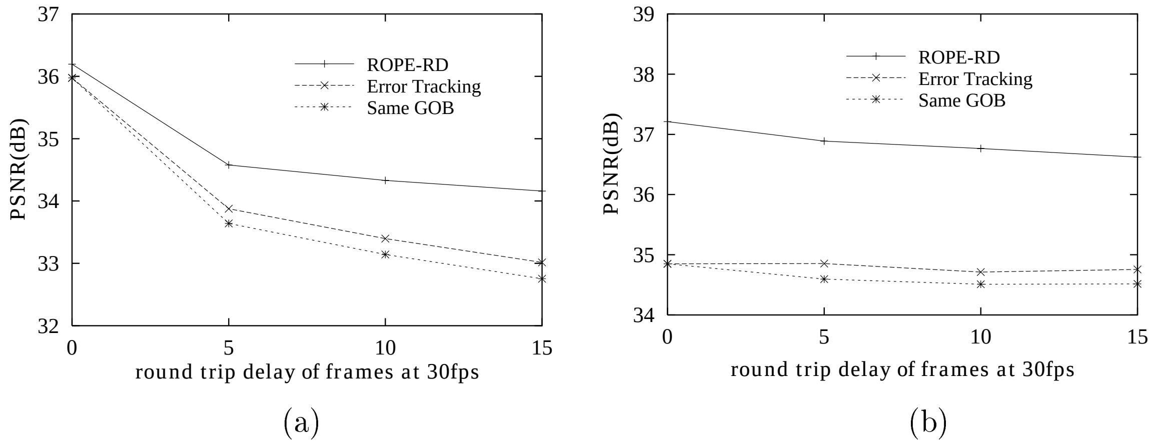


Fig. 2. PSNR vs. bit rate. Methods: ROPE-RD(proposed), Error Tracking [6], Same GOB [6] [11]. Packet loss rate =10%, frame rate=30fps, delay=500ms, Sequence: *carphone*.

Figure 1 shows the performance versus packet loss rate for a feedback channel with delay of 500ms. Figure 2 presents the performance versus bit rate for packet loss rate of 10% and delay of $500ms$. The figures provide ample evidence for the superiority of ROPE-RD, which outperforms the competing methods by 0.3∼2.6dB.

In Figure 3, we present the performance versus feedback delay at packet loss rate of 10%. The frame rate is fixed at 30fps, and the delay is expressed in terms of the number of frames: $d = 0$ implies that packet loss information for the n-th frame is received right after it is encoded.



(a) (b)

Fig. 3. PSNR vs. feedback delay. Methods: ROPE-RD(proposed), Error Tracking [6], Same GOB. [6] [11]. Bit rate=300kbps, packet loss rate =10%, frame rate=30fps. Sequences: (a) *carphone*, (b) *grandma.*

V. Conclusion

A method is proposed for rate distortion optimized mode selection, which fully exploits feedback information intelligently to enhance the robustness of video coders to packet loss. The encoder computes an optimal estimate of the overall distortion of decoder reconstruction for the given packet loss rate, feedback information and error concealment method. The algorithm dynamically initializes the estimate to account for new feedback information and recursively computes the overall distortion at pixel-level precision to accurately account for both temporal and spatial error propagation. We incorporate the estimate within an RD framework to optimally select the coding mode for each MB. Simulation results show substantial and consistent gains over state-of-the-art mode selection methods. The proposed method only requires modification of the encoder decisions, and is thus standard-compatible.

References

[1] M. Khansari, A. Jalali, E. Dubois and P. Mermelstein, “Robust low bit-rate video transmission over wireless access systems,” in *Proc. Int. Conf. Communication (ICC)*, New Orleans, May 1994, pp. 571-575.

[2] Q. F. Zhu and L. Kerofsky, “Joint source coding, transport processing and error concealment for H.323-based packet video,” *Proceedings of the SPIE, VCIP 99*, vol.3653, pp. 52-62, San Jose, CA, USA, Jan. 1999.

[3] P. Cherriman and L. Hanzo, “Programmable H.263-based video transceivers for interference-limited environments,” *IEEE Trans.Circuits Syst. Video Technol.*, vol.6, pp. 1-11, Feb. 1996.

[4] Y. Wang and Q. F. Zhu, “Error control and concealment for video communication: a review,” *Proc. of the IEEE*,vol.86, pp. 974-997, May 1998.

[5] C. Y. Hsu, A. Ortega, and M. Khansari, "Rate control for robust video transmission over burst-error wireless channels," *IEEE Journal on Selected Areas in Communications*, vol. 17, No. 5, May 1999, pp. 756-773.

[6] E. Steinbach, N. Farber and B. Girod, "Standard compatible extension of H.263 for robust video transmission in mobile environments," *IEEE Trans. on Circuits and Systems for Video Technology*, pp. 872-881, Vol.7, No.6, Dec. 1997.

[7] R. Talluri, "Error-resilent video coding in the MPEG-4 standard," *IEEE Commun. Mag.*, vol. 36, pp. 112-119, Jun. 1998.

[8] S. Fukunaga, T. Nakai, and H. Inoue, "Error resilient video coding by dynamic replacing of reference pictures," in *Proc. IEEE Global Telecommunications Conf. (GLOBECOM)*, London, U.K, Nov.1996, vol.3, pp. 1503-1508.

[9] Y. Tomita, T. Kimura, and T. Ichikawa, "Error resilient modified inter-frame coding system for limited reference picture memories," in *Proc. Int. Picture Coding Symp. (PCS)*, Berlin, Germany, Sept. 1997, pp.743-748.

[10] B. Girod and N. Farber, "Feedback-based error control for mobile video transmission," *Proc. of the IEEE*, vol.87, no.10, pp. 1707-1723, Oct. 1999.

[11] T. Turletti and C. Huitema, "Videoconferencing on the Internet,"*IEEE/ACM Transactions on Networking*, pp. 340-351, Vol.4, No.3, Jun. 1996.

[12] R. Zhang, S. L. Regunathan and K. Rose, "Optimal intra/inter mode switching for robust video communication over the Internet," *Thirty-third Asilomar Conference on Signals, Systems, and Computers*, CA, USA, Oct.24-27, 1999.

[13] R. Zhang, S. L. Regunathan and K. Rose, "Video Coding with Optimal Inter/Intra Mode Switching for Packet Loss Resilience," to appear on *IEEE Journal of Selected Areas in Communication.*

[14] ITU-T Recommendation H.324, "Terminal for low bitrate multimedia communication," 1995.

[15] ITU-T Recommendation H.245, "Control Protocol for Multimedia Communication," 1996.

[16] ITU-T Recommendation H.323, "Visual Telephone Systems and Equipment for Local Area Networks Which Provide a Non-Guaranteed Quality of Service," 1996.

[17] H. Schulzrinne, S. Casner, R. Frederick and V. Jacobson, "RTP: A transport protocol for real-time applications," RFC1889, Jan. 1996.

[18] Telenor H.263 codec, ftp://bonde.nta.no/pub/tmn/software.

[19] "RTP Payload Format for the 1998 Version of ITU-T Rec. H.263 Video (H.263+)" Internet Draft, RFC2429, ftp://ftp.isi.edu/in-notes/rfc2429.txt.