JOINTLY OPTIMIZED MODE DECISIONS IN REDUNDANT VIDEO STREAMING

J. Christian Schmidt and Kenneth Rose

Department of Electrical and Computer Engineering
University of California
Santa Barbara, California 93106–9560
Email: {schmidt,rose}@ece.ucsb.edu

ABSTRACT

This paper revisits the problem of source-channel coding for error-resilient video streaming, using redundant encoding. We propose a new method to jointly optimize the macroblock (MB) mode in both the primary and the secondary, redundant encoding. Encoding decisions are based on end-to-end distortion using the ROPE estimate, and are adaptive at the MB level. Further, we present a reduced complexity approach to redundant encoding. The proposed methods are general in nature, and could be implemented on top of any (hybrid) video codec. An example implementation employs the redundant slice mechanism of H.264 (JM 13.2). Simulation results show significant performance gains over conventional methods such as fixed redundant encoding schemes or non-redundant optimal MB mode selection.

Index Terms— video streaming, error resilience, source-channel coding, redundant slices, H.264

1. INTRODUCTION

Packet video streaming remains a challenging problem due to the inherent best-effort nature of the underlying network and lack of guaranteed end-to-end quality of service (QoS). This motivates ongoing research efforts into error resilience mechanisms to mitigate the impact of packet loss. An error-resilient encoder may adjust its macroblock (MB) coding mode decisions, e.g. using Intra refresh to stop error propagation instead of temporal prediction. Error propagation can be minimized by appropriately selecting motion parameters such as the motion vector (MV) and the reference frame. At the transport level, channel coding tools such as forward error correction (FEC) or automatic re-transmission requests (ARQ) can be used to protect data packets. Finally, an error-resilient decoder can minimize the impact of lost packets by performing suitable error concealment.

In live streaming, source-channel coding algorithms are used to optimize the rate-distortion (RD) tradeoff between source compression, robustness to packet loss and (expected) distortion at the receiver. Such optimization depends critically on accurate end-to-end distortion estimation. We resort to the “recursive optimal per-pixel estimate” (ROPE) [1, 2], which enables end-to-end distortion estimation by tracking, per pixel, the first and second moments of the decoder reconstruction. ROPE accounts for all sources of distortion such as quantization, packet loss, error propagation, and error concealment at the decoder. It has been successfully applied to MB coding mode selection [1], error-resilient motion estimation/compensation [3] and reference picture selection [4], multiple description video coding [5], and joint mode and quality of service (QoS) selection [6]. For more details on ROPE, the reader may refer to [1, 2].

Fundamentally, an error-resilient encoder must balance the conflicting objectives of mitigating channel loss and stopping error propagation versus compression efficiency. Rather than allocate the entire bit budget for source coding, some rate may be designated for channel protection, trading some source coding fidelity for a decrease in the effective packet-loss rate (PLR). Channel coding mechanisms such as FEC or ARQ are subject to practical drawbacks. ARQ is dependent on feedback, requiring a longer buffering period due to feedback delay. FEC can be applied per packet (in bit-error channels) or across packets (packet-erasure channels). In either case, FEC rate allocation is performed at the packet level, *after* encoding. Due to uneven packet sizes and padding, it is difficult to estimate the effective rate at encode time. In [6], the authors proposed a Trellis-based algorithm to address this problem, albeit at the cost of delay and complexity. In practice, channel coding mechanisms lower the effective PLR as experienced by the source coder, leaving open the traditional issue of error propagation. Recently, redundant encoding has been proposed [9], e.g. as enabled by the redundant slice mechanism in H.264 [7].

Conventional redundant encoding algorithms, e.g. [9], employ a fixed scheme, simply retransmitting each frame (or parts) at lower rate with the same MB modes for simplified mode decision. Not accounting for secondary RD cost, the mode is chosen by optimizing for the primary MB.

This work was supported in part by the University of California MICRO program, Applied Signal Technology Inc., Cisco Systems Inc., Dolby Laboratories Inc., Qualcomm Inc., and Sony-Ericsson Inc.

Operationally, the total bit rate is distributed between the primary and secondary transmission by adjusting the QP offset. Clearly, the rate and end-to-end distortion contributions from both the primary and secondary MBs impact RD performance. Hence, coding parameters (e.g. MB mode, QP, MVs) for both MBs should not be chosen independently, but rather jointly optimized. We have previously addressed this topic in [10], where we proposed selecting some MBs for identical retransmission. In this work, we present a new source-channel scheme for redundant encoding. Our algorithm jointly optimizes the MB coding modes in the primary and secondary encoding while accounting for their true rate and end-to-end distortion costs. Since the scheme depends on accurate estimation of end-to-end distortion and hence the importance of an MB, we extend the basic ROPE technique to the objective at hand.

This paper is structured as follows: Section 2 briefly recaps the ROPE estimate, and describes distortion estimation for redundant coding. We describe the joint redundant encoding algorithm, and also introduce a simplified method. Section 3 contains simulation results. We implement the algorithms on top of the JM 13.2 reference software [8] and compare its performance with other error-resilient coding methods. The paper concludes in Section 4 with a brief summary and future research directions.

2. REDUNDANT ENCODING

2.1. End-to-end distortion estimation

There are three possible channel outcomes in redundant transmission: (i) primary data received, (ii) primary data lost, secondary data received, and finally (iii) both transmissions lost and the affected region needs to be concealed. Assuming iid packet loss for simplicity, these outcomes have probabilities $1-p$, $p(1-p)$ and p^2, respectively. Let $\tilde{f}_n^{i,1}$, $(\tilde{f}_n^{i,1})^2$ and $\tilde{f}_n^{i,2}$, $(\tilde{f}_n^{i,2})^2$ denote the *successful* reconstruction (and its squared value) of the primary and secondary coded data, respectively. The resulting ROPE moments $E\{\tilde{f}_n^i\}$, $E\{(\tilde{f}_n^i)^2\}$ for end-to-end distortion estimation can be expressed as a weighted combination of these and the moments due to error concealment:

$$E\{\tilde{f}_n^i\} = (1-p)E\{\tilde{f}_n^{i,1}\} + p(1-p)E\{\tilde{f}_n^{i,2}\} + p^2 E\{\tilde{f}_{n-1}^i\} \quad (1)$$

$$E\{(\tilde{f}_n^i)^2\} = (1-p)E\{(\tilde{f}_n^{i,1})^2\} + p(1-p)E\{(\tilde{f}_n^{i,2})^2\} + p^2 E\{(\tilde{f}_{n-1}^i)^2\} \quad (2)$$

With f_n^i denoting source pixels, the expected end-to-end distortion is

$$E\{D\} = \sum_i (f_n^i)^2 - 2f_n^i E\{\tilde{f}_n^i\} + E\{(\tilde{f}_n^i)^2\} \quad (3)$$

Inserting (1) and (2) into (3), it is evident that the distortion depends on the primary and secondary MBs in a complex way.

2.2. Joint optimization

The equations above illustrate that the primary MB has a larger weight towards end-to-end distortion, but the ratio changes with the packet loss rate (PLR). Therefore, the redundant MB should be coded at a lower rate, i.e. using a higher QP. It is not intuitively obvious which choice of QP is optimal for the secondary data. A higher QP results in less rate from a smaller coded residual, while the rate for side information (e.g. mode, MV) is unchanged. Therefore, the primary MB mode may well not be the optimal choice for the secondary MB, even when both are only considered individually. Recall that the MB pair in primary and secondary transmission jointly impacts end-to-end performance. Hence, optimal performance can only be obtained when considering all possible MB coding combinations for the primary and secondary MB pair:

1. Determine the reconstruction values $\tilde{f}_{n,1}^i$, $(\tilde{f}_{n,1}^i)^2$ and rates of all available coding modes for the MB in the *primary* coded picture.
2. Determine the reconstruction values $\tilde{f}_{n,2}^i$, $(\tilde{f}_n^{i,2})^2$ and rates of all available coding modes for the MB in the *redundant* coded picture.
3. For each combination of primary and secondary MB coding mode, compute the combined rate and end-to-end distortion using the values from steps 1 and 2.
4. Pick the mode combination that achieves the best Lagrangian $J = D + \lambda R$.

In practice, the decision space can be pruned by considering only secondary MB modes that incur a rate lower than the rate of the primary MB mode.

2.3. Simplified optimization

Despite the pruning of the decision space, full joint optimization remains computationally complex. A simplified encoding algorithm could indeed use the same mode for the primary and secondary MB, significantly reducing number possible MB mode combinations. If we still account correctly for the combined RD costs of the MB pair in the primary and secondary transmission, good performance should be possible at reduced computational complexity.

3. SIMULATION & RESULTS

We implemented the proposed algorithms for redundant encoding on top of the JM 13.2 reference software [8]. In the figures below, the full joint optimization and the simplified method are labelled "o 1+2" and "o 1=2", respectively. For comparison, we implemented a general redundant encoding scheme similar to the one proposed in [9]: source RD optimization for the primary MB, random Intra MB refresh and lower rate retransmission. It is denoted as "red+rI". We also

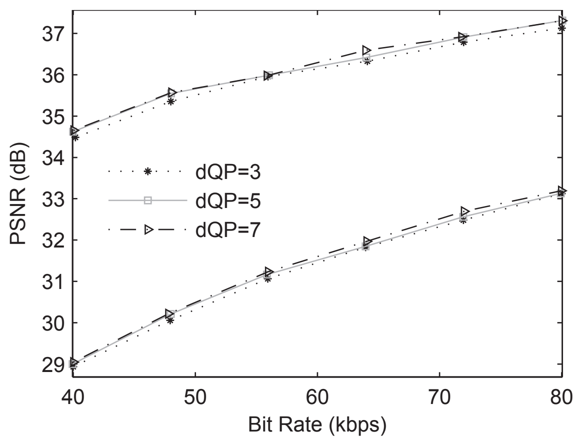


Fig. 1. Impact of secondary QP offset, PSNR vs. bit rate (hall monitor, qcif, 10 fps, 40–80 kbps, p=10%). Upper set of curves shows joint redundant encoding and lower set shows general fixed redundant encoding.

provide results of conventional non-redundant MB mode selection using ROPE ("opt I"). All sequences were encoded at 10 fps, QCIF resolution, packet size $\leq$ 512 bytes. The bitstreams were then simulated at different packet loss rates, and results averaged over 500 loss patterns at each PLR.

3.1. Choice of secondary QP offset

In the first experiment, we investigate how critical the choice of secondary QP is. We encoded the hall monitor sequence with QP offsets $dQP = 3, 5, 7$ (target bit rates 40–80 kbps). The upper plot lines in Figure 1 show the results for the proposed full joint optimization method, and the lower plot lines depict the basic general redundant coding. The QP offset has a small impact of 0.1–0.3 dB, with higher QPs achieving better performance. Hence, for the following results, we use a QP offset of $dQP = 7$.

3.2. Performance vs. bit rate

We also compared performance across a range of target bit rates (50–100 kbps, effective bit rates accurately denoted), and PLR fixed at 10%. Figure 2(a) shows the foreman sequence. Both proposed redundant encoding algorithms outperform basic redundant coding, which achieves small gains over non-redundant mode selection. At low bit rates, the proposed schemes achieve gains of 1–1.2 dB over basic redundant coding; the gain increases to 1.5–2dB at higher bit rates. Equivalently, the proposed methods can achieve a $\geq$40% bit rate reduction over to basic redundant coding.

For the coastguard sequence (Figure 2(b)), "red+rI" falls behind non-redundant mode selection. Our proposed simplified method achieves 0.4–0.5 dB over non-redundant coding, or bit rate savings of $\approx$20%. Joint optimization of primary and secondary MB mode achieves 0.7–1.2 dB, or rate savings of over 30%.

3.3. Performance vs. PLR p

The third experiment evaluate the algorithms across a PLR range of 1–25%. Figure 3(a) shows the foreman sequence (75 kbps). At low PLRs (p=1%) performance is similar, but it diverges at medium to high PLRs. The proposed algorithms achieve the best results, followed by the general redundant encoding scheme and non-redundant encoding using optimal MB mode selection. Full joint optimization outperforms the simplified method by up to 0.5 dB, albeit at the cost of increased complexity. The general redundant coding scheme improves upon non-redundant mode selection by up to 0.6 dB (it performs worse at p=25%), but lags behind both methods proposed in this paper by over 1 dB. Figure 3(b) depicts the results for the carphone sequence at 60 kbps. Joint redundant optimization gains 0.1–0.2 dB over the the simplified method, followed by "opt I" and "red+rI".

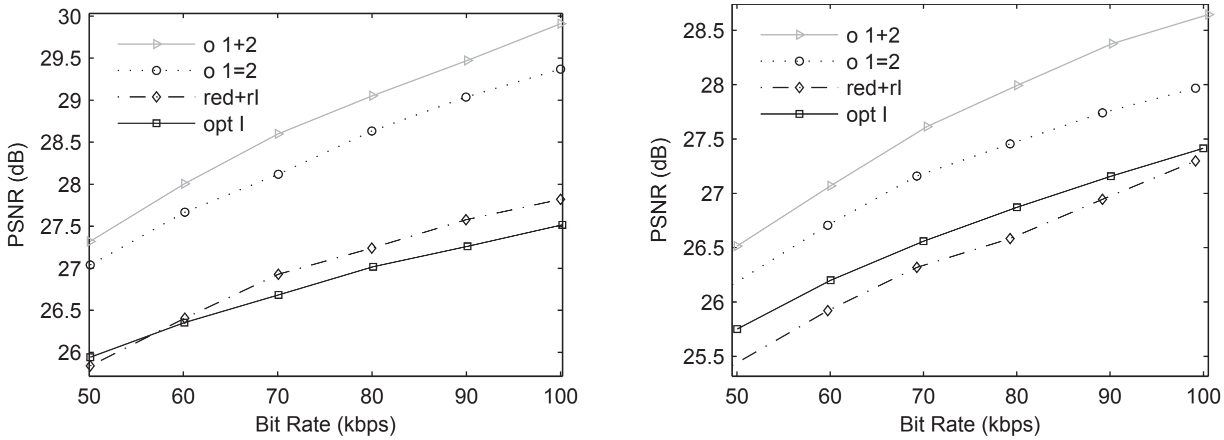
4. CONCLUSION & FUTURE WORK

We present a new error-resilient encoding algorithm that enables optimal redundant encoding by jointly optimizing the primary and secondary MB mode. Results show RD performance gains over naive redundant coding schemes and conventional ROPE-based optimal MB coding mode selection. A simplified encoding algorithm is presented that uses the same MB mode for both the primary and secondary representation (but considers true end-to-end RD costs), enabling a trade-off of encoding complexity for end-to-end performance.

The proposed scheme can be further enhanced beyond the presented results: Adaptively adjusting the rate for the redundant encoding, e.g. by controlling the QP or the coded residual, may enable additional gains. The algorithm could employ principles of multiple descriptions (e.g. prediction from different reference pictures) and enable improved reconstruction when both descriptions are received. Flexible macroblock ordering (FMO) [7] could enable improved concealment: the gains from optimal redundant encoding and improved concealment should be additive. is data partitioning. It separates the individual components contributing to overall reconstruction quality, e.g. coding mode, motion information and residual. Adaptive optimization for these components may enable further gains.

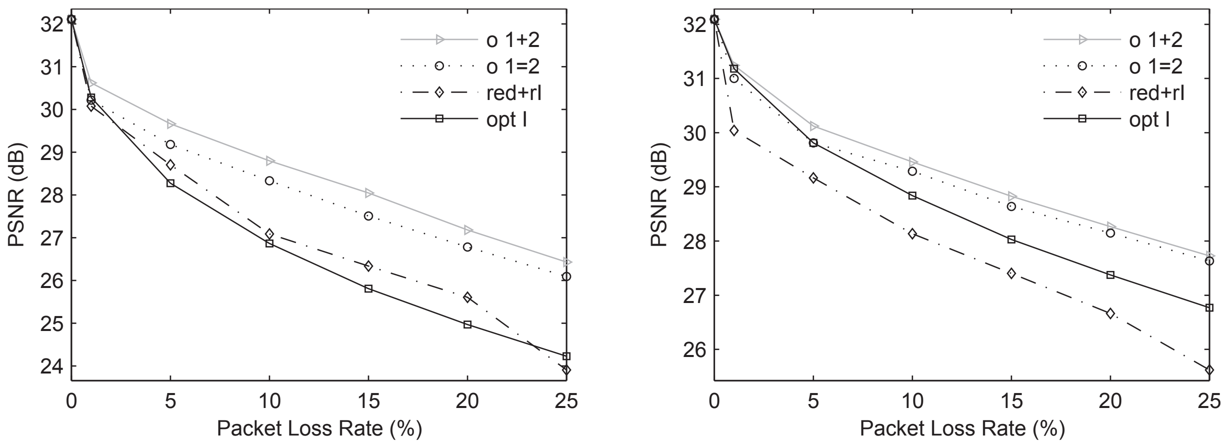
5. REFERENCES

[1] R. Zhang, S. L. Regunathan, K. Rose, "Video coding with optimal inter/intra-mode switching for packet loss resilience", *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 6, pp. 966–976, 2000.



(a) foreman (b) coastguard

Fig. 2. Delivery performance, PSNR vs. bit rate (qcif, 10 fps, 50-100 kbps, p=10%)



(a) foreman, 75 kbps (b) carphone, 60 kbps

Fig. 3. Delivery performance, PSNR vs. PLR p (qcif, 10 fps)

[2] H. Yang, K. Rose, "Advances in Recursive Per-Pixel End-to-End Distortion Estimation for Robust Video Coding in H.264/AVC", *IEEE Trans. Circuits and Systems for Video Technology*, vol. 17, no. 7, pp. 845–856, 2007.

[3] H. Yang, K. Rose, "Rate-Distortion Optimized Motion Estimation for Error Resilient Video Coding", *IEEE ICASSP*, vol. 2, pp. 173–176, 2005.

[4] A. Leontaris, P. C. Cosman, "Video compression for lossy packet networks with mode switching and a dual-frame buffer", *IEEE Trans. Image Processing*, vol. 13, no. 7, pp. 885–897, 2004.

[5] B. A. Heng, J. G. Apostolopoulos, J. S. Lim, "End-to-end rate-distortion optimized mode selection for multiple description video coding", *IEEE ICASSP*, vol. 5, pp. 905–908, 2005.

[6] E. Masala, H. Yang, K. Rose, J. C. De Martin, "Rate-distortion optimized slicing, packetization and coding for error-resilient video transmission", *DCC*, pp. 182–191, 2004.

[7] Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG, "Advanced video coding for generic audiovisual services", *ITU-T Recommendation H.264*, ISO/IEC 14496-10 AVC, March 2005.

[8] H.264/AVC Software Coordination: *http://iphome.hhi.de/suehring/tml/*

[9] P. Baccichet, S. Rane, A. Chimienti, B. Girod, "Robust Low-Delay Video Transmission using H.264/AVC Redundant Slices and Flexible Macroblock Ordering", *IEEE ICIP*, vol. 4, pp. 93–96, 2007.

[10] J. C. Schmidt, K. Rose, "Macroblock-based retransmission for error-resilient video streaming", to appear in *IEEE ICIP*, 2008.