MISMATCH IMPACT ON PER-PIXEL END-TO-END DISTORTION ESTIMATION AND CODING MODE SELECTION

Hua Yang

Corporate Research, Thomson Inc.
Princeton, NJ 08540, USA
hua.yang2@thomson.net

Kenneth Rose *

Dept. of Electrical and Computer Engineering
Univ. of California Santa Barbara, CA 93106, USA
rose@ece.ucsb.edu

ABSTRACT

The recursive optimal per-pixel estimate (ROPE) is an effective end-to-end distortion estimation scheme. Most existing ROPE-based applications assume that: (i) the encoder knows exactly the actual packet loss rate and (ii) the decoder error concealment scheme; (iii) no deblocking in-loop filtering is employed. However, in practice, these assumptions may not all be valid. In this paper, we investigate the impact of mismatch between assumed and actual conditions on the performance of ROPE and corresponding rate-distortion optimized coding mode selection. Useful conclusions are drawn from extensive experimental results.

1. INTRODUCTION

The recursive optimal per-pixel estimate (ROPE) [1] is an effective means to accurately estimate end-to-end distortion (EED) in live video streaming applications, e.g., video telephony and conferencing. When compared with other EED estimation schemes (most notably block-based approaches [2]), ROPE is provably optimal and in practice achieves superior estimation performance. Hence, it has been applied in a large variety of rate-EED (REED) optimization techniques to improve the error resilience of video coding.

While the original ROPE was mainly proposed for full-pixel motion compensation [1], several schemes have since been proposed to circumvent the computational difficulties associated with estimating cross-correlation terms, so as to extend ROPE to accommodate sub-pixel prediction [3] [4] [5] [6]. In [5], we proposed low complexity and effective solutions to not only estimate cross-correlation, but also compensate for the impact of rounding error on EED estimation (a largely overlooked issue in the literatures). With these advances, ROPE was extended to EED estimation while accounting for pixel-averaging (or filtering) operations including, in particular, sub-pixel prediction, weighted prediction, Intra prediction, overlapped block motion compensation, linear transforms, and a large variety of error concealment schemes.

Most existing ROPE-based methods make three common simplifying assumptions: (i) the encoder knows the exact packet loss rate (PLR) of the channel; (ii) and the decoder error concealment (EC) scheme; (iii) no deblocking in-loop filtering (DIF) is employed. The latter is because in H.264/AVC DIF, one has to threshold absolute difference of block boundary pixels to determine the actual pixel-filtering operation to be conducted [7]. How to accurately account for this within ROPE is still an open issue. However, in practice, some or all of the above assumptions may not hold. Firstly, due to encoding delay, decoder feedback delay, and loss rate estimation error, the encoder may not know the exact current PLR when the frame is encoded. Secondly, there exist complicated EC schemes that cannot easily be accommodated in ROPE, e.g. various iterative EC schemes. Thirdly, DIF not only reduces blockiness artifact, improves coding efficiency, but also inherently reduces error propagation from packet loss, due to its involved pixel-averaging operations. Besides, DIF is already a normative feature in H.264/AVC. Therefore, more often than not, it will be employed in practical video coding systems.

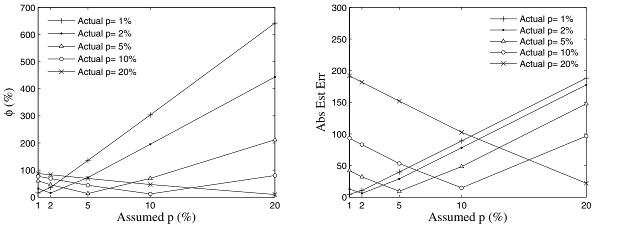
As such, in this work, we investigate the impact of PLR, EC, and DIF mismatch on the estimation performance of ROPE and, furthermore, the performance of ROPE-based REED coding mode selection. To the best of our knowledge, the only prior work touching on this question is [8], which considers the impact of PLR mismatch on their proposed variant of ROPE-based REED mode selection.

It is important to emphasize that this work focuses on the impact of mismatch on ROPE simply because ROPE is a leading EED estimation technique. However, it is obvious that mismatch is expected to generally compromise any EED estimation technique.

2. MISMATCH IMPACT ON EED ESTIMATION BY ROPE

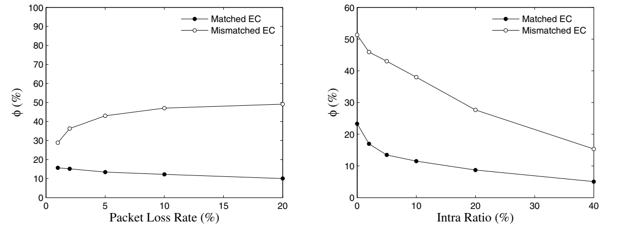
Consider the extended ROPE scheme of [5], where inter-pixel distance adaptive cross-correlation approximation (CCA) and

*This work was supported in part by the NSF under grant EIA-0080134, the University of California MICRO Program, Applied Signal Technology, Inc., Dolby Laboratories, Inc., and Qualcomm, Inc.



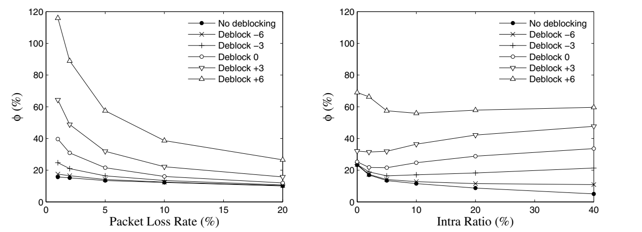
(a) (b)

Fig. 1. Distortion estimation performance with PLR mismatch. Carphone, $100kb/s$, Intra ratio= 5%, (a) in relative estimation error, (b) in absolute estimation error.



(a) (b)

Fig. 2. Distortion estimation performance with EC mismatch. Carphone, $100kb/s$, (a) Intra ratio= 5%, (b) $p = 5\%$.



(a) (b)

Fig. 3. Distortion estimation performance with DIF mismatch. Carphone, $100kb/s$, (a) Intra ratio= 5%, (b) $p = 5\%$.

quantization theoretic rounding error compensation (REC) are employed. CCA and REC enable ROPE to accurately account for any pixel-filtering operations. In the experiments, we only include 1/2-pel and 1/4-pel prediction as an example, and neither Intra-prediction nor weighted prediction is involved. We first investigate how the three types of mismatch affect the EED estimation performance of ROPE.

We used the JM9.0 H.264/AVC codec [9]. For each sequence, only the first frame is coded as I-frame, and all the remaining frames are coded as P-frames. The other encoding settings are: using single reference frame, all the H.264/AVC available macro-block (MB) coding modes of I- and P-frames enabled, using the JM9.0 none rate-distortion (RD) optimized mode decision option and the JM9.0 random Intra updating scheme. Each frame was packed into one packet for transmission. At the decoder, the default EC scheme is the simple copy-from-previous-frame (or so-called frame-copy) scheme.

500 randomly generated packet loss patterns were applied at each PLR, and the actual average distortion was computed for each pixel of each frame. Estimation performance is measured by the "distortion error ratio" defined as:

$$\phi = \frac{\sum_n \sum_i |d^i_{n,Est} - d^i_{n,Dec}|}{\sum_n \sum_i d^i_{n,Dec}}. \tag{1}$$

Here, $d^i_{n,Est}$ and $d^i_{n,Dec}$ denote, for pixel i of frame n, the distortion estimated at the encoder, and the actual decoder distortion averaged over all loss patterns, respectively. We will refer to the enforced Intra MB ratio per frame as "Intra Ratio", and denote the PLR by p. All our testing sequences are QCIF with $15f/s$, and only the first 150 frames of each sequence are coded.

2.1. PLR Mismatch

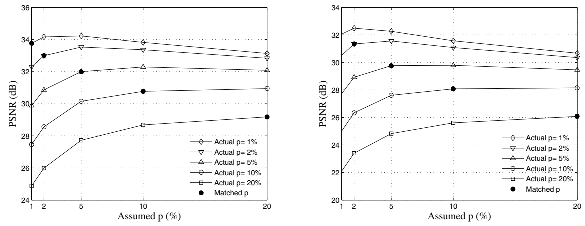
Fig. 1 gives the PLR mismatch results. From Fig. 1 (a), the most serious degradation in terms of relative EED estimation error happens at low PLR, e.g. $1 \sim 5\%$, which is overestimated. It should be noted that the estimation error itself behaves in a fairly symmetric fashion but when normalized with the actual decoder error, as per (1), its impact is seen to be more pronounced in the above circumstances. This is verified in Fig. 1 (b), where the un-normalized absolute estimation error is depicted. Therefore, overall, in terms of relative estimation error ϕ, more significant performance degradation is experienced at low PLRs.

2.2. EC Mismatch

Performance with EC mismatch is shown in Fig. 2. The encoder always assumes that the decoder employs frame-copy EC. We tested the performance where the decoder employs either frame-copy EC (as assumed by the encoder) or motion-copy EC (as an example for EC mismatch). Motion-copy EC replaces a lost frame while employing motion-vectors from collocated MBs in the previous frame to motion compensate and conceal MBs in the current frame [10]. The results show that EC mismatch also compromises estimation accuracy, although not as severely as PLR overestimation at low PLRs.

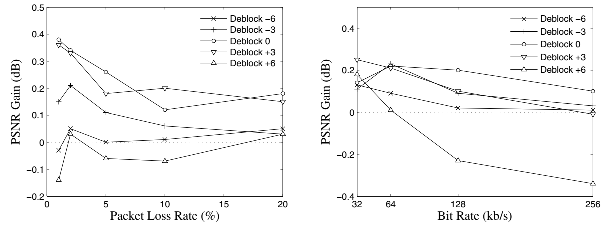
2.3. DIF Mismatch

Fig. 3 presents the DIF mismatch results. Here ROPE assumes (for simplicity) that no DIF is employed, while in reality DIF is employed at various levels of filtering strength. Deblocking filtering strength is given by the threshold table offset, ranging from -6 (lowest filtering strength) to $+6$ (highest filtering strength) [7]. Herein, "No deblocking" denotes the case of no DIF mismatch, while "Deblock -6"$\sim$"Deblock $+6$" denote the mismatch cases at different DIF strength lev-



(a) (b)

Fig. 4. REED mode selection performance with PLR mismatch. $100kb/s$, (a) Carphone, (b) Foreman.



(a) (b)

Fig. 5. REED mode selection performance with DIF mismatch. Carphone, (a) $100kb/s$, (b) $p = 5\%$.

els. The estimation performance is degraded with DIF mismatch, and the degradation increases with DIF filtering strength. Also, similarly to EC mismatch, the degradation is not as serious as in the case of PLR mismatch.

3. MISMATCH IMPACT ON ROPE-BASED CODING MODE SELECTION

Next, we investigate how the above types of mismatch will affect the performance of ROPE-based coding mode selection. As mentioned in Section 1, a major application of ROPE is in the general and effective REED framework, where selection of various encoding parameters (or options) may be optimized for the best REED trade-off, and thereby improve error resilience. These coding parameters may include MB coding mode, motion vectors, prediction reference frames, and quantization step sizes, etc. As an example, we only focus on MB coding mode selection. This problem is usually formulated as independently selecting the best coding mode for each MB/block to minimize the REED Lagrangian cost. In the experiments, the source coding distortion of the conventional "non-error-resilient" RD optimized mode selection scheme of JM9.0 was simply replaced with ROPE-estimated EED for REED optimization [9]. At the decoder side, average PSNR over 500 packet loss patterns for each PLR is calculated to measure the overall system performance. All the other settings are kept the same as in Section 2.

3.1. PLR Mismatch

Results with mismatched PLR are shown in Fig. 4. The first observation is that mismatched p does not necessarily yield worse performance. In fact, the highest PSNR is always achieved by slightly *overestimating* p. To understand this result, we emphasize that in most existing REED schemes (including in our experiment), coding parameters (in our case the MB coding mode) are optimized for each frame without considering future frames. However, due to motion compensated prediction, inter-frame dependency inherently exists in video coding, which implies that MB coding mode decisions in the current frame will affect the REED Lagrangian cost, and thus, MB coding mode decisions in the following frames. Hence, this "zero delay" REED is only *locally optimal*, but not *globally optimal*. Existing efforts on more globally optimal (delayed decision) coding parameter selection can be found in [11] for non-error-resilient RD coding, and [12] for REED coding methods, where coding parameters of a group of frames are jointly optimized. Intuitively, to account for inter-frame dependency, EED in one frame should carry more weight (than 1) in calculating the REED Lagrangian cost so as to account for its propagation impact in the following frames. This is somewhat equivalent to *overestimating* EED. On the other hand, overestimated p yield overestimation of EED. With EED "*properly*" overestimated, overall REED performance may be improved. One can also observe that performance will degrade with excessive overestimation of p.

3.2. EC Mismatch

To investigate the EC mismatch impact on REED mode selection, we again use motion-copy EC as the mismatched EC scheme at the decoder. The results are summarized in Table 1. We can see that except for Carphone, applying motion-copy EC at the decoder, although mismatched, always yields better performance than that of applying the matched frame-copy EC. In experiment, we also evaluated the effectiveness of the two different EC schemes for different sequences without REED, and found that motion-copy EC outperforms frame-copy EC for all the sequences except Carphone. The bottom line is clearly that the impact of mismatch here is much less significant that the relative effectiveness of the EC methods themselves.

3.3. DIF Mismatch

As for DIF mismatch, in Fig. 5, we directly give the average PSNR gains of the cases with DIF mismatch over those without DIF (and without mismatch). Obviously, DIF mismatch does not necessarily lead to degraded PSNR performance. For deblocking strength between -3 and $+3$, DIF mismatch usually yields better performance than that without mismatch. One reason for this result is that pixel-averaging operations

Table 1. REED mode selection performance with EC mismatch on various sequences. $p = 5\%$, Stefan and Football: $200kb/s$, others: $100kb/s$.

PSNR (dB)	Miss_am	Mobile	News	Carphone	Table	Foreman	Stefan	Football
Frame-copy (matched)	40.26	26.00	35.28	32.00	30.55	29.77	25.42	27.73
Motion-copy (mismatched)	40.65	27.44	35.35	30.76	31.23	30.28	26.03	27.79
Gain of Motion-copy	**+0.39**	**+1.44**	**+0.07**	**-1.24**	**+0.68**	**+0.51**	**+0.61**	**+0.06**

in DIF generally have the effect of reducing error propagation from packet loss, and thus, improves error resilience. Moreover, for the same reason, DIF mismatch usually yields EED overestimation. The earlier overestimation argument suggests that the resultant performance degradation will not be excessive. Overall, it is clear that the benefits of employing DIF at mild strength at least outweigh any damage due to mismatch.

4. CONCLUSIONS

In this paper, we examined the impact of PLR, EC and DIF mismatch on ROPE-based EED estimation, and found that all types of mismatches will compromise the estimation performance. As an application example, we then investigated the impact on REED coding mode selection. We observe that the mismatch does not necessarily translate into significant degradation of the overall REED coding performance. In particular, some overestimation of the PLR often improves overall performance, a phenomenon that we attribute to an effective compensation for the fact that coding decisions neglect the impact of loss on future frames.

An important direction for future study is developing efficient and effective techniques to directly and optimally mitigate the various mismatch impacts on distortion estimation, which will guarantee the reliability, and thus, applicability of ROPE in a broader spectrum of practical settings.

5. REFERENCES

[1] R. Zhang, S. L. Regunathan and K. Rose, "Video coding with optimal intra/inter-mode switching for packet loss resilience". *IEEE Journal Select. Areas Commun.*, vol. 18, no. 6, pp. 966-76, 2000.

[2] G. Cote and F. Kossentini, "Optimal intra coding of blocks for robust video communication over the Internet," *Sig. Processing: Image Commun.*, pp. 25-34, vol. 15, Sept. 1999.

[3] H. Yang and K. Rose, "Recursive end-to-end distortion estimation with model-based cross-correlation approximation," *Proc. of ICIP 2003*, vol. 3, pp. 469-72, Sept. 2003.

[4] A. Leontaris and 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. 885-97, July 2004.

[5] H. Yang and K. Rose, "Advances in recursive per-pixel estimation of end-to-end distortion for application in H.264," *Proc. ICIP 2005*, vol. 2, pp. 906-9, Sept. 2005.

[6] V. Bocca, M. Fumagalli, R. Lancini, and S. Tubaro, "Accurate estimate of the decoded video quality: Extension of ROPE algorithm to halfpixel precision," *Proc. of Picture Coding Symp.*, San Francisco, Dec. 2004.

[7] JVT of ISO/IEC MPEG and ITU-T VCEG, "ITU-T Rec. H.264, ISO/IEC 14496-10 AVC," Aug. 2002.

[8] Y. Shen, P. C. Cosman, and L. Milstein, "Video coding with fixed length packetization for a tandem channel," *IEEE Trans. on Image Processing*, vol. 15, no. 2, pp. 273-88, Feb. 2006.

[9] [Online]. http://iphome.hhi.de/suehring/tml/download.

[10] M. C. Hong, L. Kondi, H. Scwab, and A. K. Katsaggelos, "Error Concealment Algorithms for Compressed Video," *Sig. Processing: Image Commu.*, vol. 14, pp. 437-92, 1999.

[11] K. Ramchandran, A. Ortega, M. Vetterli, "Bit allocation for dependent quantization with applications to multiresolution and MPEG video coders," *IEEE Trans. on Image Processing*, vol. 3, no. 5, pp. 533-45, Sept. 1994.

[12] R. Zhang, S. L. Regunathan, and K. Rose, "Prescient mode selection for robust video coding," *Proc. ICIP 2001*, vol. 1, pp. 974-7, Oct. 2001.