RATE-DISTORTION OPTIMIZED MOTION ESTIMATION FOR ERROR RESILIENT VIDEO CODING

Hua Yang and Kenneth Rose

Department of Electrical and Computer Engineering
University of California Santa Barbara, CA, 93106-9560, USA
hua, rose@ece.ucsb.edu

ABSTRACT

Rate-distortion optimized motion compensation was originally proposed to improve the source coding efficiency of low bit-rate video coding. This paper investigates the potential benefits of extensions to explicitly handle transmission over lossy networks. The important impact of appropriate motion compensation on error resilience is identified and exploited to enhance performance. The expected end-to-end distortion, rather than the source coding distortion, is employed in the rate-distortion optimization criterion. The accuracy of the distortion estimate is crucial when optimizing over the motion vectors (as compared to the less sensitive Inter/Intra mode decision), and is ensured here by adopting the recursive optimal per-pixel estimate (ROPE) method. The main objective is to assess the gains achievable by such source-channel motion compensation, and this is given priority over complexity considerations. Significant performance gains in simulation demonstrate the potential of rate-distortion optimized motion compensation to improve the error resilience of video coding.

1. INTRODUCTION

Motion compensated prediction is a widely adopted technique in current video coding standards to efficiently remove temporal redundancy between successive frames. Conventional motion estimation selects motion vectors so as to minimize the (mean squared or mean absolute) prediction error, while ignoring the eventual cost in motion vector bit-rate consumption. However, in the case of low bit-rate video coding, this may lead to poor coding efficiency, as motion information represents a significant portion of the total bit budget.

In order to optimize the bit allocation tradeoff between the motion vectors and quantizers, rate-distortion (RD) optimized motion estimation was proposed (see [1][2]). In [1] and [3], motion vector selection and quantizer selection are jointly optimized via minimizing the Lagrangian RD cost, which involves exactly the overall *source coding* rate and distortion. To reduce computational complexity, other schemes perform sequential RD optimization of motion estimation and quantization, where the Lagrangian cost applied to motion estimation balances the *prediction error* against the motion vector bit-rate [4]. A theoretical analysis of this technique is provided in [2].

The emergence of video-over-network and wireless video applications have triggered a change in paradigm and research emphasis. Due to the inevitable packet loss over current Internet or wireless networks, coding efficiency must be expanded to encompass error resilience – a challenging task in the case of video coding. More specifically, in terms of RD optimization, the distortion of primary concern is no longer the source coding distortion (as for lossless transmission) but, rather, the overall end-to-end distortion accounting for the impact of packet loss. Consequently, end-to-end distortion based RD optimization has been widely recognized as an efficient means to introduce error resilience for lossy video transmission, and several recently proposed schemes focus on RD optimized coding mode selection, wherein the coding mode refers to Intra/Inter mode and/or quantization scale [5][6]. The underlying philosophy is, roughly, that appropriate Intra/Inter decisions combat error propagation, and quantizer scaling enhances source coding efficiency. However, not much attention was given to optimizing motion vectors within such an end-to-end RD framework.

It is intuitively obvious that since motion vectors have a central role in the prediction loop, they are directly and critically involved in the error propagation mechanism encountered in transmission over lossy packet networks. Hence, motion compensation may have a considerable impact on error resilience and on the ultimate overall system performance. As a specific example, consider the fact that predicting the current macroblock (MB) from an Intra MB in the previous frame is more robust choice than predicting it from an Inter MB that may entail propagated error. More

This work is 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.

generally, it is potentially beneficial to explicitly account for error resilience considerations while selecting motion vectors. Consequently, the objective of this work is to explore the potential of motion vector optimization for improving error resilience of video transmission. Existing work addressing similar objectives can be found in [7] and [8]. In contrast, our focus here is on the crucial need for precise end-to-end estimation and exact and explicit tradeoff optimization, in order to maximize the gains.

In summary, this work considers the scenario of video transmission over lossy packet networks. Specifically, the expected end-to-end distortion is calculated and employed in the proposed RD optimized motion compensation. To obtain an upperbound on the expected gains we give priority to performance over complexity considerations, and jointly optimize motion vector and quantization step size selection. It is also worthwhile noting that end-to-end distortion estimation itself is an underlying crucial problem. Herein, we adopt and build on the recursive optimal per-pixel estimate (ROPE) method [6], which guarantees an accurate distortion estimate. Significant performance gains achieved in simulation provide experimental evidence in support of the potential of RD optimized motion compensation to enhance error resilience of video coding.

2. RD OPTIMIZED MOTION COMPENSATION FOR LOSSY PACKET NETWORKS

Traditionally, motion estimation in video coding has been studied in the (explicit or implicit) context of lossless transmission. Most reference models of existing video coding standards select the motion vector so as to minimize some measure of the prediction error. For example, the minimum mean-squared prediction error is given by

$$\min_{mv} \sum_{i\in MB} (f_n^i - \hat{f}_{n-1}^{i+mv})^2 = \min_{mv} \sum_{i\in MB} (e_n^i)^2, \qquad (1)$$

where, f_n^i and $\hat{f}_n^i$ denote the original and encoder reconstruction values of pixel i in frame n, respectively, mv is searched over all the possible motion vector choices of a MB, and e_n^i is the prediction error before quantization.

In cases of very low bit rates, or where complex choices for motion vector selection were available (e.g., multi-frame motion compensation [7][8]), RD optimization of motion compensation was considered in order to better allocate bitrate among motion vector and prediction residue coding, and thereby improve the overall coding efficiency [1][3]. Herein, for simplicity, we assume that RD optimization is carried out independently for each MB (as global optimization taking into account motion vector coding dependency often incurs impractical complexity [3][4]). Hence, the source coding rate distortion problem is:

$$\min_{\{mv,QP\}} [D + \lambda \cdot (R_{res} + R_{mv} + R_{header})] \qquad (2)$$

where,

$$D = \sum_{i\in MB} (f_n^i - \hat{f}_n^i)^2. \qquad (3)$$

Here QP denotes the quantization step size, R_{res}, R_{mv}, and R_{header} denote the number of bits for coding the prediction residue, selected motion vector, and MB header, respectively. Ignoring the clipping effect, D reduces to the quantization distortion D_Q. Let $\hat{e}_n^i$ denote the quantized prediction error. Then D_Q is defined as

$$D_Q = \sum_{i\in MB} (e_n^i - \hat{e}_n^i)^2. \qquad (4)$$

Next, we extend the above RD approach to the case of video transmission over lossy channels. The only change in formulation is the replacement of source-coding distortion D with the expected end-to-end distortion $E\{D\}$, which also accounts for the impact of packet loss. For convenience, but without loss of generality, we assume the data of one frame is delivered in one packet, and error concealment of a lost frame is performed by copying the decoder reconstruction of the previous frame. Let $\tilde{f}_n^i$ denote the decoder reconstruction, and let p be the packet loss rate. The expected end-to-end distortion is:

$$\begin{aligned} E\{D\} &= \sum_{i\in MB} E\{(f_n^i - \tilde{f}_n^i)^2\} \qquad (5) \\ &= \sum_{i\in MB} [(1-p)\cdot E\{(f_n^i - \tilde{f}_{n-1}^{i+mv} - \hat{e}_n^i)^2\} \\ &\quad + p\cdot E\{(f_n^i - \tilde{f}_{n-1}^i)^2\}] \\ &\simeq \sum_{i\in MB} [(1-p)\cdot E\{(\hat{f}_{n-1}^{i+mv} - \tilde{f}_{n-1}^{i+mv})^2\} \\ &\quad + (1-p)\cdot(e_n^i - \hat{e}_n^i)^2 \\ &\quad + p\cdot E\{(f_n^i - \tilde{f}_{n-1}^i)^2\}] \qquad (6) \\ &= (1-p)\cdot D_{EP} + (1-p)\cdot D_Q + p\cdot D_{EC}. \qquad (7) \end{aligned}$$

Note that (6) is derived from (5) under the assumption that the effects of quantization and channel errors are uncorrelated. Then, in (7), in an intuitive fashion, we rename the three distortion terms of (6) as the error propagation distortion D_{EP}, the familiar quantization distortion D_Q, and the error concealment distortion D_{EC}, respectively.

We can see that in addition to R_{mv} and R_{res}, the impact of mv on $E\{D\}$ appears explicitly in D_{EP}, which quantifies the distortion propagated from the previous frame for a given motion vector choice. Note further how D_{EP} favors motion vectors that point to areas in the previous frame where there is less encoder-decoder mismatch, e.g., Intra coded MB's. We claim that, in lossy video transmission scenarios, motion compensation affects not only the trade-off between R_{mv} and R_{res} as in (2) (source coding efficiency), but also (and perhaps more importantly) the trade-off between coding efficiency and error resilience.

It should be mentioned that the above end-to-end distortion formulation builds on and extends the formulation of the ROPE method [6], where the emphasis was on Intra/Inter mode selection. The important novelty of the current approach is in the focus on the potential of optimizing motion compensation for error resilience, and thereby further enhance the RD performance of Inter-coded MB's. Alternatively, it can be viewed as extending the existing definition of coding mode to incorporate the motion vector and solving the optimization problem.

Finally, we emphasize that accurate estimation of the expected end-to-end distortion is in itself fundamental and critical to the success of the proposed RD optimization. We borrow from ROPE [6], the realization that the problem reduces to that of calculating the first and second moments of the decoder reconstruction, and that this can be efficiently and optimally performed in a recursive manner. By building the solution on the basis of ROPE, we ensure accuracy of the end-to-end distortion estimate. Due to limited space, we refer readers to [6] for details on basic ROPE.

3. SIMULATION RESULTS

The objective of this work is to evaluate or upperbound the potential performance gains due to end-to-end distortion based RD optimization of motion compensation. Hence, we give higher priority to performance than complexity considerations, and jointly optimize motion vector and quantization step size selection to make the encoding decisions.

For simplicity, our simulation system is built on the UBC H.263+ codec. Note that the approach clearly is applicable to later standards such as H.264. Prior to decoding, 50 different randomly generated packet loss realizations are used. The test is performed on the first 200 frames of each sequence. Only the 1st frame is encoded as I-frame, and all the remaining are P-frames. System performance is measured by averaging the luminance PSNR. We compare the performance of our proposed RD optimized motion compensation ("RDMC"), with that of the conventional motion compensation ("cMC") that ignores channel loss. The two methods are tested with either random Intra updating ("rI") or optimal Intra updating ("oI"). Random Intra method arbitrarily assigns MB's to $1/p$ groups, and cycles through them updating one group per frame [5], while the optimal Intra method employs ROPE for Intra/Inter decisions [6].

Fig. 1, Fig. 2, and Tab. 1 show the results versus packet loss rate, total bit-rate, and for various QCIF sequences, respectively. An obvious observation is that in the setting of random Intra, the proposed RDMC method provides considerable gains over conventional motion compensation. The performance of RDMC-rI closely approaches and sometimes is even better than that of cMC-oI, e.g. for $p = 0.2$, 0.3 in Fig. 1 (a), 32kb/s in Fig. 2 (a), and "Table tennis" in Tab. 1. This result proves that RDMC offers efficient control of error propagation from previous frames and, in particular, explicitly accounts for and exploits the increased error robustness of Intra MBs in the previous frame, when optimizing motion vector selection. Thus, the overall error resilience performance is improved.

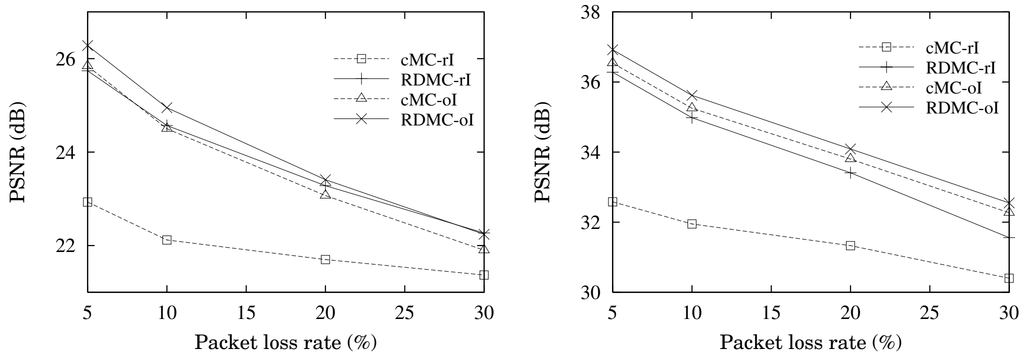
On the other hand, we note that the gains due to RD optimized Intra and RDMC are not additive, but instead there is considerable overlap. Nevertheless, combining the two offers some additional gains, as can be easily observed from the fact that RDMC-oI consistently outperforms both RDMC-rI and cMC-oI under all testing scenarios. Fig. 2 shows further that significant additional gains from RDMC-oI are mostly achieved at low bit-rates, e.g. 32kb/s or 48kb/s. Note that such rates effectively limit the degree to which Intra mode can be selected. It is in the more challenging low bit rates that RDMC offers needed additional gains above those already obtained via optimal Intra updating.

Table 1. Performance comparison for different QCIF sequences. (48kb/s, 10f/s, $p = 0.1$)

Sequence	cMC-rI	RDMC-rI	cMC-oI	RDMC-oI
Miss America	31.95dB	34.98dB	35.25dB	35.62dB
Mother/Daughter	29.22dB	30.61dB	30.98dB	31.14dB
Carphone	27.12dB	28.85dB	29.17dB	29.38dB
Foreman	22.12dB	24.57dB	24.50dB	24.95dB
Table Tennis	21.92dB	24.49dB	24.13dB	24.54dB

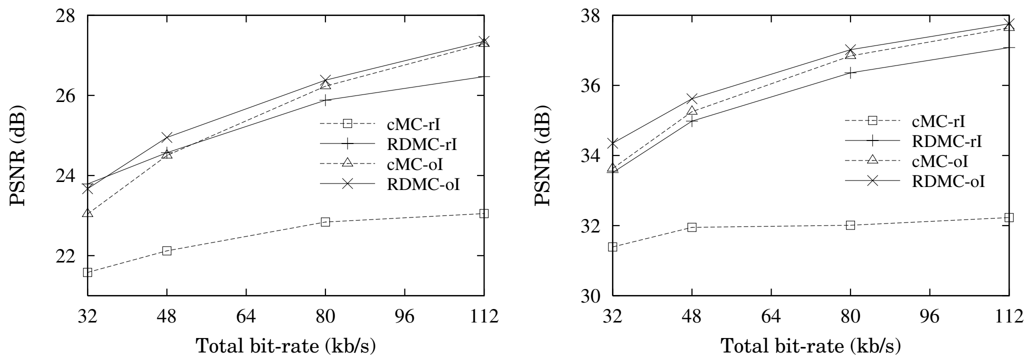
4. CONCLUSION

In the context of lossy video transmission, motion vectors adopted at the encoder may greatly affect the ultimate error propagation process at the decoder, as MBs in the previous frame may vary in their error history and accumulated propagated distortion, e.g., in the extreme case of Intra MBs all propagation was in fact stopped. This suggests that there is always scope to effectively control error propagation by judicious selection of motion vectors. This paper identifies this opportunity, and an RD optimized solution is proposed. Our scheme employs expected end-to-end distortion in its RD optimization, which is a natural extension of the existing RD-optimized motion compensation approach to encompass error resilience. Simulations investigate the upperbound performance of the proposed scheme. Results show



(a) Foreman (b) Miss_am

Fig. 1. Performance comparison for different packet loss rates. (QCIF, 10f/s, 48kb/s)



(a) Foreman (b) Miss_am

Fig. 2. Performance comparison for different total bit-rates. (QCIF, 10f/s, $p = 0.1$)

that, in conjunction with random Intra updating, our scheme consistently provides substantial performance improvement over conventional motion compensation, and in conjunction with RD-optimal Intra updating, system performance at low bit-rates is ameliorated.

5. REFERENCES

[1] W. C. Chung, F. Kossentini, and M. J. T. Smith, "An efficient motion estimation technique based on a rate-distortion criterion," *Proc. of IEEE ICASSP 1996*, vol. 4, p. 1926-29, May 1996.

[2] B. Girod, "Rate-constrained motion estimation," *Proc. of SPIE Visual Commun. Image Processing 1994*, vol. 2308, p. 1026-34, Nov. 1994.

[3] G. M. Schuster and A. K. Katsaggeslos, "A theory for the optimal bit allocation between displacement vector field and displaced frame difference," *IEEE J. Select. Areas Commun.*, vol. 15, p. 1739-51, Dec. 1997.

[4] M. C. Chen and A. N. Willson, "Rate-distortion optimal motion estimation algorithms for motion-compensated transform video coding," *IEEE Trans. on Circuits Sys. Video Tech.*, vol. 8, no. 2, p. 147-58, Apr. 1998.

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

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

[7] T. Wiegand, N. Farber, K. Stuhlmuller and B. Girod, "Error-resilient video transmission using long-term memory motion-compensated prediction," *IEEE J. Select. Areas Commun.*, vol. 18, no. 6, , p. 1050-62, June 2000.

[8] M. Budagavi and J. D. Gibson, "Multiframe video coding for improved performance over wireless channels," *IEEE Trans. Image Processing*, vol. 10, no. 2, p. 252-65, Feb. 2001.