# ADVANCES IN RECURSIVE PER-PIXEL ESTIMATION OF END-TO-END DISTORTION FOR APPLICATION IN H.264

*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

This paper is concerned with open questions and modifications to expand the applicability of the recursive optimal per-pixel estimate (ROPE) of end-to-end distortion, which are particularly relevant to H.264. One open question involves the emergence of cross-correlation terms in the case of sub-pixel prediction or other pixel filtering operations. A new and improved low-complexity approximation is proposed, which accounts for the inter-pixel distance. Another open question involves the commonly ignored effects of rounding and clipping. The cumulative impact of rounding on the distortion estimate is shown to be extensive at low to medium packet-loss rates. Two schemes are proposed for effective rounding and clipping error compensation. Simulation results for H.264 with 1/4-pel prediction show that the revised ROPE maintains low complexity and achieves estimation accuracy that closely matches that of ROPE in the simpler case of full-pixel prediction, where it is optimal.

## 1. INTRODUCTION

It is widely recognized that end-to-end distortion estimation, coupled with rate-distortion (RD) optimization, offers an efficient means to achieve error resilience in applications involving video streaming over networks. The performance, however, critically depends on the estimation accuracy. In the case of live video streaming, one may exploit modifications to the source coding module, which typically involve block or pixel-based estimation schemes [1] [2].

Of particular interest here is the recursive optimal per-pixel estimate (ROPE), which achieves superior distortion estimation accuracy [2]. We explore open questions regarding the applicability of ROPE in general. One question involves the estimation of cross-correlation terms, which arise in ROPE due to various standard pixel-filtering operations [2]. Some preliminary work on cross-correlation approximation (CCA) appeared in [3]. Here we revisit this problem and propose a more effective scheme which explicitly accounts for the inter-pixel distance.

We identify another largely overlooked issue, namely the rounding error, whose impact has long been considered insignificant, and hence neglected in end-to-end estimation. We show that, although negligible in terms of impact on reconstruction quality, rounding errors may greatly impact the *estimation accuracy* as they accumulate through the prediction loop. We hence propose two approaches for rounding error compensation (REC).

Finally, it is important to emphasize that although we focus attention on the sub-pixel prediction setting, both CCA and REC are concerned with problems that are inherent to "per-pixel" end-to-end distortion estimation in general. In fact, we believe that CCA and REC as proposed here effectively open the door to practical utilization of ROPE in the context of H.264, and offer the benefits of ROPE at the cost of only modest increase in complexity.

## 2. CROSS-CORRELATION APPROXIMATION

The accuracy of ROPE in end-to-end distortion estimation is attributed to its ability to calculate the 1st and 2nd moments of decoder reconstructed pixels, while accounting for all relevant factors including quantization, packet loss, error propagation and error concealment [2]. However, sub-pixel prediction involves interpolation of neighboring pixels [4], which gives rise to cross-correlation terms in the 2nd moment calculation. In the worst case this may require calculating and tracking cross-correlation for all pixel pairs in the frame, and incur impractical complexity. In [3] we proposed to perform CCA, using only the readily available marginal moments, thereby maintaining the low complexity of basic ROPE. we refer to these simple models as (see [3] for details): **Model 0**: no correlation; **Model I**: maximum correlation; and **Model II**: linear model with additive noise.

In this paper we propose an improved CCA model whose motivation stems from the realization that CCA must benefit from exploiting knowledge of the *inter-pixel distance*. Intuitively, one expects the correlation to decay with the dis-

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.

tance between pixels, and it is worthwhile to account for this within the model. In the specific case of H.264, the inter-pixel distance may range from $1/2$ to $5$, due to 6-tap filtering [4]. We note that the spatial random field of a source image has been modeled with the isotropic exponentially decaying autocorrelation function [5]. In a similar fashion, we propose a distance-adaptive CCA model as follows.

**Model III**: distance-adaptive correlation

$$\rho_{XY} = \exp(-\alpha \cdot d_{XY}), \tag{1}$$

where, $d_{XY}$ is the Euclidian distance between two decoder reconstructed pixels $X$ and $Y$, and $\alpha$ is a constant, whose value can be experimentally obtained from training data (typically $0.04$-$0.06$). Note that (1) models the correlation coefficient $\rho_{XY}$. With the 1st and 2nd moments of $X$, $Y$ readily available through ROPE, the cross-correlation $E[XY]$ is trivially obtained. Finally, $E[XY]$ is further bounded by Schwartz inequality as was proposed in [3].

## 3. ROUNDING ERROR COMPENSATION

An important, yet largely neglected, issue in end-to-end estimation is that of rounding error. Rounding is typically employed whenever pixel filtering or averaging operations produce floating point outputs. In H.264, rounding operations are encountered in sub-pixel prediction, weighted prediction, in-loop filtering, etc.

Rounding can be viewed as a special case of *uniform quantization* with quantization step size of one unit, and where the quantized value is the nearest integer. The rounding error is $\Delta =< X > -X$, where $< \cdot >$ denotes the rounding/quantization operation and $X$ is the input random variable. If $X$ is a continuous random variable, then basic quantization theory states:

$$\sigma_X \to 0 \quad : \quad \Delta \to < E[X] > -E[X], \tag{2}$$

$$\sigma_X \gg 1 \quad : \quad E[\Delta] \simeq 0,\ E[\Delta^2] \simeq 1/12. \tag{3}$$

¿From (2), we see that in the case of small $\sigma_X$, the rounding error tends to some typically non-zero value that is determined by $E[X]$. Note that in video coding, this non-zero rounding error may be propagated via inter-frame prediction, and accumulate to seriously degrade end-to-end estimation.

Herein, we propose two approaches to rounding error compensation (REC). To maintain low complexity, we only use quantities made available by basic ROPE. We hence pose the general problem:

- Let $Y =< X >$, where $X$ is a random variable with known moments $E[X]$, $E[X^2]$. Estimate $E[Y]$, and $E[Y^2]$.

Our first approach appeals to the maximum entropy principle (MEP) [7]. Specifically, we estimate the distribution of $X$ as the one that maximizes the entropy while maintaining the given $E[X]$ and $E[X^2]$. It is then straightforward to calculate $E[Y]$ and $E[Y^2]$. Note that in H.264, due to the particular 6-tap (1/2 pel) and bilinear (1/4 pel) filtering, the input $X$, or the filter output, is not continuous but discrete and takes value in a 1/32-grid or 1/2-grid, respectively [4]. MEP directly yields the Gibbs distribution. However, since the 1/32-grid represents a fairly high resolution, we approximately treat 1/2-pel prediction $X$ as a continuous random variable, where MEP yields the Gaussian distribution. Thus, the **MEP-based REC approach** yields:

- for 1/2-pel prediction:

$$X \sim N(\mu_X, \sigma_X^2) \tag{4}$$

- for 1/4-pel prediction:

$$p(x_i) = \frac{1}{Z} \exp\left(-\frac{(x_i - \mu)^2}{2\sigma^2}\right). \tag{5}$$

In (4), $\mu_X$ and $\sigma_X^2$ are the known mean and variance of $X$. In (5), $Z$ is the normalization coefficient, while the $\mu$ and $\sigma^2$ parameters are chosen such that $p(x)$ satisfies the prescribed 1st and 2nd moment constraints.

The second approach originates in quantization theory (QT). Specifically, we have: $E\Delta = EY - EX$, and

$$\sigma_\Delta^2 = (\sigma_Y - \rho_{XY}\sigma_X)^2 + (1 - \rho_{XY}^2)\cdot\sigma_X^2. \tag{6}$$

For $\sigma_X^2 \geq \sigma_\Delta^2$, we deduce from (6) that:

$$|\rho_{XY}| \geq A,\ where\ A = \sqrt{1 - \frac{\sigma_\Delta^2}{\sigma_X^2}}. \tag{7}$$

For large $\sigma_X^2$ ($\sigma_X^2 \gg 1$), we can reasonably assume that: (i) $\Delta$ is uniformly distributed, and (ii) $\rho_{XY}$ is positive. Hence, we simply assume $\rho_{XY} \simeq A$. For small $\sigma_X^2$, we may round $E\{X\}$ directly. Thus, we have the **QT-based REC approach**:

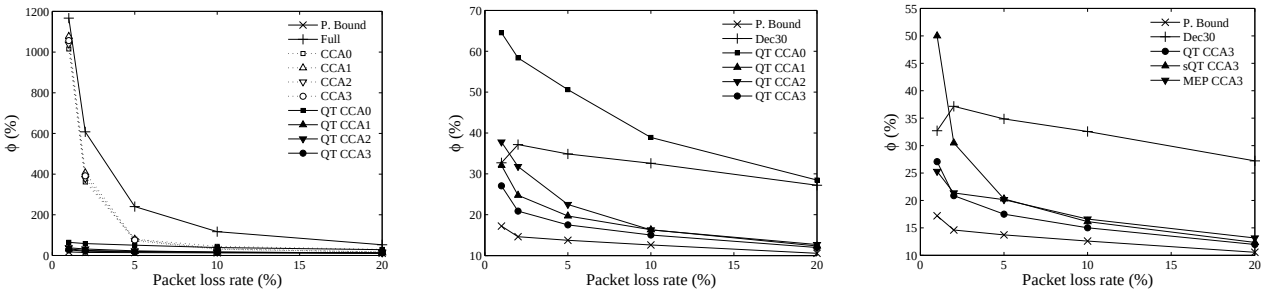
1. If $\sigma_X^2 > \beta$:

$$E[Y] = E[X] + E\Delta,\ \sigma_Y^2 \simeq \sigma_X^2 - \sigma_\Delta^2. \tag{8}$$

2. otherwise:

$$E[Y] \simeq < E[X] >,\ \sigma_Y^2 \simeq \sigma_X^2, \tag{9}$$

where $\beta$ is an experimentally defined threshold (typically $0.2$-$1.2$). Note that $\sigma_Y^2$ in (8) can be derived by plugging $\rho_{XY} \simeq A$ of (7) into (6). Finally, we determine $E\Delta$ and $\sigma_\Delta^2$ depending on the prediction cases: For 1/2-pel prediction: $E\Delta = 0$, $\sigma_\Delta^2 = 1/12$. For 1/4-pel prediction: $E\Delta = -1/4$, $\sigma_\Delta^2 = 1/16$.




(a) General comparison. (b) CCA comparison. (c) REC comparison.

**Fig. 1**. Estimation performance vs packet loss rates. (Carphone: QCIF, 15f/s, 100kb/s)

We note that clipping also introduces an error that may impact estimation similarly to rounding. The MEP-based approach is extendible in a straightforward manner to handle clipping error compensation (CEC). In simulations we observed that clipping does not cause as much damage as rounding and hence CEC appears less significant. Nevertheless, MEP-based REC/CEC is superior in principle, and its gains may become significant for other video sequences or operating points.

## 4. SIMULATION RESULTS

The purpose of this section is to evaluate the estimation accuracy offered by the proposed enhancements. For space constraints we restrict attention to estimation performance evaluation while the obvious applicability to improve the overall coding efficacy is not covered herein. Estimation performance will be measured by:

$$\phi = mean(|d^i_{n,Est} - d^i_{n,Dec}|)/mean(d^i_{n,Dec}). \qquad (10)$$

Here, $d^i_{n,Est}$ and $d^i_{n,Dec}$ denote respectively the distortion estimate at the encoder and actual decoder distortion (averaged over many channel simulations) at pixel $i$ of frame $n$.

We used the JM9.0 H.264 codec with 1/4-pel prediction. In the experiments, we adopted periodic Intra updating, which enforces Intra coding of fraction $p$ of the MB's in each frame (where $p$ denotes the packet loss rate.) 200 randomly generated packet loss patterns were applied to the coded bitstream, and average MSE distortion is computed for each pixel of each frame in order to obtain the performance measure $\phi$. We tested four CCA methods, denoted "CCA0"-"CCA3", which correspond to the four models enumerated in Section 2. For REC, we tested the proposed MEP-based ("MEP") method and QT-based method ("QT"). We also examined the performance of a simplified version of "QT" (denoted "sQT") which only employs (9). For benchmarking purposes: *i)* When the coder employs full-pel prediction ROPE is optimal (no CCA or REC issues) and this provides an estimation performance bound (denoted "P. Bound".) *ii)* A brute-force benchmark denoted "Dec30" [6], has the encoder calculate average distortion exhaustively via 30 runs of decoding simulation with different loss patterns. (Note that its estimation error is essentially due to the limit on decoding runs.) *iii)* For completeness we also provide the ROPE performance when it simplistically assumes full-pel prediction for distortion estimation, ignoring CCA and REC, a version that appeared in the original ROPE paper [2] (denoted "Full"),

In Fig. 1 (a), we see that, relative to "Full", CCA improves estimation accuracy and achieves good performance at medium to high packet loss rates (e.g., $p \geq 10\%$). However, its performance is poor at low packet loss rates. At $p \leq 2\%$, we note the catastrophic deterioration in performance of CCA-only ROPE methods. However, this severe problem is eliminated by handling REC (here by QT REC). This result clearly demonstrates the significance of the rounding error problem at low packet loss rates, and the efficiency of the proposed REC solution.

Fig. 1 (b) provides a "close-up" look to compare the various CCA models, here in conjunction with QT REC. We see that all models except CCA0 perform well and approach the bound "P. Bound". Moreover, they significantly outperform Dec30. Note also that CCA3 is consistently the best performing model, which shows the importance of inter-pixel distance to correlation approximation. (As a side note, the superiority of ROPE itself can be clearly seen from the gains of "P. Bound" over Dec30.)

Adopting CCA3, we provide a comparison of various REC methods in Fig. 1 (c). We note that the most naive sQT REC offers substantial performance gains. We further see that by proper handling of large variance cases (QT CCA3) results in further improvement. Interestingly, the accuracy of QT CCA3 is often somewhat better than that of the MEP CCA3. (This suggests that the first and second moment constraints do not capture all the information available for the distribution estimation, hence hindering MEP). Given the

**Table 1**. Estimation performance with various sequences. (15f/s, 100kb/s, $p = 5\%$)

| $\phi$ (%) | Miss_am | Mthr_dotr | Salesman | Coastguard | Carphone | Foreman | Stefan |
|---|---|---|---|---|---|---|---|
| P. Bound | **15.18** | **10.66** | **8.86** | **12.02** | **13.73** | **12.70** | **13.00** |
| Dec30 | **30.03** | **25.89** | **24.86** | **24.06** | **34.86** | **36.61** | **28.34** |
| QT CCA0 | 52.73 | 42.52 | 34.66 | 41.62 | 50.60 | 54.46 | 41.36 |
| QT CCA1 | 24.04 | 19.39 | 15.23 | 21.50 | 19.70 | 17.80 | 19.73 |
| QT CCA2 | 26.22 | 20.66 | 15.20 | 17.33 | 22.51 | 21.40 | 16.24 |
| QT CCA3 | **20.03** | **17.06** | **13.04** | **16.26** | **17.50** | **15.46** | **15.46** |
| sQT CCA3 | 24.12 | 23.34 | 15.79 | 21.37 | 20.28 | 16.63 | 16.03 |
| MEP CCA3 | 21.63 | 18.98 | 13.89 | 21.32 | 20.14 | 16.78 | 15.92 |

extremely low complexity of QT REC, we believe that at this point it is the leading candidate for adoption.

Table 1 shows the performance with various sequences. Clearly, QT CCA3 outperforms all the other CCA REC combinations. Moreover, its performance is always considerably better than that of Dec30.

Finally, we emphasize that the improved ROPE method poses no serious complexity concerns in practice. Running with Pentium IV $3.0GHz$ CPU and 504MB RAM, we observed that the total encoding time of ROPE with QT CCA3 is 2.3-2.6 times that of the standard. In terms of storage/memory, the standard uses 1 byte per integer pixel, while ROPE additionally needs 8 bytes to store the 1st and 2nd moments in floating point.

## 5. CONCLUSIONS

In this paper we considered problems that pose practical obstacles on the general applicability of ROPE. One is the emergence of cross-correlation terms in the estimate. Other problems involve proper accounting, within the recursive estimate, for rounding and clipping operations and their cumulative impact. We propose low-complexity solutions to these problems and demonstrate by simulations (H.264 with 1/4-pel prediction) that while these problems are highly significant, they can be overcome in practice to substantially enhance estimation performance. The proposed modifications make ROPE a powerful tool to achieve the error resilience potential of H.264.

## 6. REFERENCES


[1] G. Côté, F. Kossentini, "Optimal intra coding of blocks for robust video communication over the Internet," *Image Commun.*, pp. 25-34, Sept. 1999.

[2] 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.

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

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

[5] J. R. Jain and A. K. Jain, "Displacement measurement and its application in interframe image coding," *IEEE Trans. Commun.*, vol. COM-29, pp. 1799-1804, Dec. 1981.

[6] T. Stockhammer, T. Wiegand, and S. Wenger, "Optimized transmission of H.26L/JVT coded video over packet-lossy networks,"*Proc. ICIP 2002*, Rochester, NY, 2002.

[7] E. T. Jaynes, "Information theory and statistical mechanics," *Papers on Probability, Statistics and Statistical Physics,* Reidel, Dordrecht, 1982.