Advances in Recursive Per-Pixel End-to-End Distortion Estimation for Robust Video Coding in H.264/AVC

Hua Yang, *Member, IEEE*, and Kenneth Rose, *Fellow, IEEE*

***Abstract*—This paper is focused on expanding the applicability of the recursive optimal per-pixel estimate (ROPE) of end-to-end distortion, to the H.264/AVC standard. One open question involves the complexity of cross-correlation terms that appear in the case of subpixel prediction and other pixel filtering operations. Several efficient model-based solutions are proposed. Another open question involves the largely ignored effects of rounding operations, whose cumulative impact may seriously degrade the estimate accuracy. Two approaches are proposed for rounding error compensation: one appeals to the maximum entropy principle, while a low complexity alternative is based on quantization theoretic approximations. The former effectively estimates the distribution of the decoder reconstruction, thereby significantly broadening the applicability of ROPE to all additive distortion measures. Simulation results for H.264/AVC with 1/4-pel prediction show that the proposed ROPE extensions achieve fairly high estimation accuracy, while maintaining low complexity. Another set of results demonstrates the level of overall coding gains achievable by exploiting such improved end-to-end distortion estimation.**

***Index Terms*—Coding mode selection, end-to-end distortion, error resilience, video coding.**

I. Introduction

END-TO-END distortion (EED) estimation is a central component in many techniques that employ rate-distortion (RD) optimization for error resilience in video networking applications. In live video streaming applications (e.g., video telephony/conferencing), error robustness is typically achieved via encoding decision optimization involving various coding parameters or options [1]–[4]. While it is straightforward to find the exact bit rate cost of various encoding decisions, EED is more elusive as it depends on various factors (e.g., packet loss events) that are not known at the encoder. In fact, the accuracy of EED estimation has a critical impact on the overall RD optimization performance and the resulting error resilience.

Most existing EED estimation techniques for robust video coding may be roughly categorized as either "block-based" or "pixel-based" methods. A block-based approach generates and recursively updates a block-level distortion map for each frame [5]–[7]. However, since inter-frame displacements involve subblock motion vectors, a motion compensated block may inherit errors propagated from multiple blocks in prior frames. Hence, block-based techniques must involve a possibly rough approximation (for example, weighted averaging of propagated block distortion [5], [6] or motion vector approximation [7]), whose errors may build up to significantly degrade estimation accuracy. In contrast, pixel-based approaches track the distortion estimate per pixel and have the potential to provide high accuracy. The obvious question is that of complexity. One extreme approach was proposed in [8] where the distortion per pixel is calculated by exhaustive simulation of the decoding procedure and averaging over many packet loss patterns. Another pixel based approach was proposed in [2], where only the two most likely loss events are considered. However, it turns out that low complexity can be maintained without sacrificing optimality as has been demonstrated by the recursive optimal per-pixel estimate (ROPE) [1]. ROPE recursively calculates the first and second moments of the decoder reconstruction of each pixel, while accurately taking into account all relevant factors, including error propagation and error concealment. ROPE has been applied for EED estimation in numerous RD optimization based coding techniques, including: intra-/inter-mode selection [1], [9] and extension thereof to layered coding [10], [11], multiple description coding [12], [13], prediction reference frame and/or motion vector selection [3], [4], joint video coding and transport optimization [14]–[17], etc. Variants of ROPE have since been applied in the transform domain to estimate the EED of DCT coefficients [18], [19]. Beside distortion estimation, other applications of ROPE in robust video coding include source-channel prediction [20] and error resilient rate control [21]. Recently, ROPE has been proposed for video quality monitoring and assessment in video streaming over lossy networks [22].

However, despite the interest and extensive work on ROPE applications, there exist unsolved open problems that significantly restrict its application in practical video coding and streaming systems. The main objective of this work is to analyze these limitations and to propose effective solutions, so as to expand the general applicability of ROPE in practice. Much emphasis is given to issues relevant to the H.264/AVC standard.

A. Important Open Issues and Limitations

An important open question concerns the emergence of cross-correlation terms in the estimate due to pixel filtering (or averaging) operations. Various forms of pixel filtering operations

Manuscript received July 30, 2006; revised November 11, 2006. This work was supported in part by the National Science Foundation under Grant EIA-0080134, in part by the University of California MICRO Program, in part by Applied Signal Technology, Inc., in part by Dolby Laboratories, Inc., and in part by Qualcomm, Inc.

H. Yang was with the Department of Electrical and Computer Engineering, University of California, Santa Barbara, CA 93106 USA. He is now with Thomson Corporate Research, Princeton, NJ 08540 USA.

K. Rose is with the Department of Electrical and Computer Engineering, University of California, Santa Barbara, CA 93106 USA (e-mail: rose@ece.ucsb.edu).

Digital Object Identifier 10.1109/TCSVT.2007.897116

are performed by standard encoders, e.g., subpixel motion compensation, intra-prediction, weighted prediction, deblocking filtering, overlapped block motion compensation, etc. [23], [24]. Note that the discrete cosine transform itself can be regarded as a certain form of pixel filtering/averaging. Pixel-averaging operations may also be performed by the decoder, e.g., for error concealment [25], [26]. Within the exact ROPE procedure, such pixel filtering operations may, in the worst case, require computation and storage of cross-correlation values for all pixel pairs in the frame, which is of impractical complexity. This difficulty was recognized in the original ROPE paper, where cross-correlation calculation in 1/2-pel prediction was circumvented by assuming full-pixel prediction for ROPE calculation purposes, although the encoder did allow subpixel motion [1]. While such "naive" ROPE provided substantial gains over competing methods, this approximation did compromise its accuracy. Hence, effective and low-complexity cross-correlation approximation (CCA) is highly desirable.

An alternative approach was proposed in [27], where cross-correlation terms are computed but only within a predefined inter-pixel distance. However, the technique required substantial computation and storage complexity. Moreover, when cross-correlation outside the subset is needed, the method defaulted to assuming that the pixels were uncorrelated, which compromises the estimation performance. Another method of lower complexity was proposed in [28], where only cross-correlation of adjacent pixels were recursively calculated for each frame. Whenever needed, cross-correlation of two non-adjacent pixels were derived from that of the adjacent pixels, along the shortest path subject to a linear model assumption. However, maintaining a cross-correlation map of all adjacent pixel pairs still entails substantial computation and storage cost, especially when applying this approach in H.264/AVC, which involves 6-tap 1/2-pel and 1/4-pel filtering. A very low complexity solution was proposed in [3], where cross-correlation is simply approximated by an upper-bound obtained from the Schwarz inequality. A similar cross-correlation estimation problem was also addressed in [29] in the context of frame-level EED estimation, where various multiplication factors are adopted to account for the impact of different pixel averaging operations. In this paper, we propose several efficient pixel-level CCA models and demonstrate that they do enable ROPE to achieve high estimation accuracy while maintaining low complexity in a general setting of pixel averaging operations.

We consider the largely overlooked issue of rounding errors, whose impact has long been considered insignificant and has hence been neglected in EED estimation. We show that rounding errors may, in fact, greatly impact EED estimation accuracy, as they accumulate through the prediction loop. We propose two approaches for rounding error compensation (REC): a method that appeals to the maximum entropy principle (MEP), and a low complexity method that employs basic approximations from quantization theory. Rather than estimate the first two moments directly, MEP effectively estimates the distribution of the decoder reconstruction, which greatly expands ROPE's applicability to a large variety of other problems of interest. For example, it enables ROPE to accurately compensate for clipping errors, estimate any higher order moments, or estimate EED for a broad class of additive distortion measures beyond the classical mean squared error (MSE) distortion.

B. Relevance to H.264/AVC

The H.264/AVC video coding standard offers significant improvement of coding efficiency. The coding performance gains are largely due to the greater flexibility afforded by a variety of new features. However, an unintended side effect of such flexibility is its impact on EED estimation. From this perspective, H.264/AVC presents a more challenging scenario than its predecessors [26]. For example, H.264/AVC uses a 6-tap filter for 1/2-pel interpolation, rather than bilinear filtering as in H.263, which results in rapid growth of cross-correlation terms in EED estimation. The situation is exacerbated by the additional 1/4-pel prediction, which not only increases the number of cross-correlation terms, but also introduces more rounding operations. Other features that involve pixel filtering operations include: weighted prediction, intra-prediction and in-loop deblocking filtering [23]. It should be appreciated that these options represent major challenges for accurate EED estimation. On the other hand, H.264/AVC also represents an enormous potential to improve error resilience, as detailed in [26] and [30]. Therefore, if one could accurately estimate the EED, then the wealth of coding options and features will offer much greater flexibility in optimizing error resilience.

In this paper, we focus on H.264/AVC with 1/4-pel prediction. However, the general applicability of the proposed CCA and REC schemes covers as well the cases of weighted prediction, intra-prediction, overlapped block motion compensation, linear transforms, etc. Simulation results will show that the revised ROPE achieves superior estimation accuracy, while maintaining low complexity. Furthermore, replacing the classical RD framework with a rate-EED (REED) framework, e.g., REED coding mode selection, we will observe that the accuracy improvements offered by the revised ROPE translate into overall end-to-end coding performance gains. In experiments, we also investigate the robustness to mismatched encoder assumptions regarding various parameters such as packet loss rate, error concealment method, and deblocking filtering. Overall, we demonstrate that the proposed extensions to ROPE represent a powerful tool toward realization of the error-resilience potential of H.264/AVC as well as future standards.

The rest of the paper is organized as follows. Section II provides the necessary preliminaries on the basic ROPE approach, which serves as the starting point for this work. The various proposed ROPE extensions are described and discussed in Sections III-A–C. EED estimation results are given in Section IV-A. Section IV-B presents simulation results to measure the overall gains achieved by employing the revised estimate for REED coding mode selection.

II. Preliminaries: The Basic ROPE Approach

The ROPE method was proposed in [1] as an efficient tool to achieve accurate estimation of overall end-to-end MSE distortion. We first review its main principles which form the starting point for the contributions of this paper. Let f_n^i denote the original value of pixel i in frame n, and let $\hat{f}_n^i$ and $\tilde{f}_n^i$ denote its *encoder* and *decoder* reconstruction, respectively. Due to possible packet loss in the channel, $\tilde{f}_n^i$ must be considered a random variable by the encoder. The overall expected MSE distortion of a pixel is

$$\begin{aligned} E\left\{d_n^i\right\} &= E\left\{\left(f_n^i - \tilde{f}_n^i\right)^2\right\} \\ &= \left(f_n^i\right)^2 - 2f_n^i E\left\{\tilde{f}_n^i\right\} + E\left\{\left(\tilde{f}_n^i\right)^2\right\}. \end{aligned} \quad (1)$$

Based on the observation that MSE is completely determined by the first and second moments of the decoder reconstruction, ROPE was derived as an optimal recursive algorithm to calculate these moments per pixel in the frame, while accounting for all relevant factors, including quantization, packet loss, error propagation, and error concealment at the decoder.

Let us assume that packets are lost independently, and that the packet loss rate (PLR), denoted by p, is available at the encoder. We further assume that the data of one frame are transmitted in one packet. In this case, the pixel loss rate equals the packet loss rate. Also, throughout this paper, unless otherwise noted, we assume that whenever there is a packet loss, the decoder simply uses the previous frame reconstruction for error concealment. The respective recursion formulae of ROPE are as follows.

- Pixel in an intra-coded MB

$$E\left\{\tilde{f}_n^i\right\}(I) = (1-p)\left(\hat{f}_n^i\right) + pE\left\{\tilde{f}_{n-1}^i\right\} \quad (2)$$

$$E\left\{\left(\tilde{f}_n^i\right)^2\right\}(I) = (1-p)\left(\hat{f}_n^i\right)^2 + pE\left\{\left(\tilde{f}_{n-1}^i\right)^2\right\} \quad (3)$$

- Pixel in an inter-coded MB

$$\begin{aligned} E\left\{\tilde{f}_n^i\right\}(P) &= (1-p)\left(\hat{e}_n^i + E\left\{\tilde{f}_{n-1}^j\right\}\right) \\ &\quad + pE\left\{\tilde{f}_{n-1}^i\right\} \end{aligned} \quad (4)$$

$$\begin{aligned} E\left\{\left(\tilde{f}_n^i\right)^2\right\}(P) &= (1-p)E\left\{\left(\hat{e}_n^i + \tilde{f}_{n-1}^j\right)^2\right\} \\ &\quad + pE\left\{\left(\tilde{f}_{n-1}^i\right)^2\right\} \\ &= (1-p)\left(\left(\hat{e}_n^i\right)^2 + 2\hat{e}_n^i E\left\{\tilde{f}_{n-1}^j\right\}\right. \\ &\qquad \left. + E\left\{\left(\tilde{f}_{n-1}^j\right)^2\right\}\right) \\ &\quad + pE\left\{\left(\tilde{f}_{n-1}^i\right)^2\right\}. \end{aligned} \quad (5)$$

The inter-coding notation above assumes that pixel i is predicted from pixel j in the previous frame. The prediction error e_n^i is quantized to the value $\hat{e}_n^i$.

Next, we discuss assumptions made and practical limitations. Firstly, assumptions of independence and time-invariance of packet loss were made strictly for simplicity of exposition, and the ROPE method itself is extendible to more complex loss models. For example, see [13] and [9] for ROPE implementations that handle bursty packet loss, and bit error channels, respectively. Secondly, ROPE itself has no special restrictions on packetization. Besides the simple one-frame-per-packet packetization scheme, other more complicated and/or more practical schemes can be accomodated as well, e.g., the scheme of one independent group-of-blocks per packet addressed in [1] and the fixed packet length schemes proposed in [9]. Thirdly, although simple (but fairly common) temporal error concealment (EC) techniques were used in this paper and in [1], more sophisticated EC techniques can be accomodated, especially with the extensions proposed herein.

We should also note that despite its expanded generality, there still exist practical limitations on ROPE that may compromise its estimation performance or applicability, as well as open issues. In practice, due to PLR estimation error and decoder feedback delay [9], occasional mismatch is inevitable between the encoder's assumed PLR and the actual PLR. There exist complicated EC schemes that cannot be easily accommodated in ROPE, e.g., various iterative EC schemes [25]. Even with the proposed CCA and REC extensions, the revised ROPE does not fully account for deblocking in-loop filtering (DIF). Note that in H.264/AVC DIF, one has to threshold the absolute difference of block boundary pixels to determine the actual pixel-filtering operation to be conducted [23]. How to accurately account for such condition checking within ROPE is still an open issue. In Section IV-B2, we provide some experimental results on the impact of PLR, EC, and DIF mismatch on REED mode selection performance.

III. Critical ROPE Extensions

A. Cross-Correlation Approximation (CCA)

It is well known that subpixel motion compensated prediction considerably improves coding efficiency, and is widely adopted in video coding standards, such as H.263, H.264/AVC and MPEG4. However, as it involves interpolation of pixel values, cross-correlation terms will arise in ROPE's second moment calculation. For illustration, let us consider a simple linear interpolation example: $Z = (X + Y)/2$, where random variables X and Y denote reconstructed pixels, and Z an interpolated pixel, all *at the decoder*. Given the first and second moments of X and Y, we have

$$E\{Z\} = \frac{1}{2}\left(E\{X\} + E\{Y\}\right) \quad (6)$$

$$E\{Z^2\} = \frac{1}{4}\left(E\{X^2\} + E\{Y^2\} + 2 \cdot E\{XY\}\right). \quad (7)$$

It is evident that $E\{Z\}$ can be calculated directly from $E\{X\}$ and $E\{Y\}$, which are made available by ROPE. However, a new cross-correlation term $E\{XY\}$ appears in (7) and is needed to calculate $E\{Z^2\}$. In fact, as discussed in Section I-A, cross-correlation terms appear in all pixel-filtering situations, whose accurate estimation via the exact ROPE recursion, in the worst case, requires computation and storage of cross-correlation values for all possible pixel pairs in a frame. Such complexity has been considered a significant practical limitation on the applicability of ROPE.

Several prior approaches to address this issue have been discussed in the Introduction. In this paper, we analyze the CCA problem from the correlation coefficient perspective. The correlation coefficient of X and Y is

$$\rho_{XY} = \frac{E\{XY\} - E\{X\}E\{Y\}}{\sigma_X \sigma_Y}, \quad \rho_{XY} \in [-1, 1]. \quad (8)$$

Here, σ_X and σ_Y denote the respective standard deviation (trivially computable from the available first and second marginal moments). Obviously, CCA is equivalent to the problem of correlation coefficient estimation.

Two simple and extreme cross-correlation models consist of either assuming that X and Y are uncorrelated, or that they are maximally correlated, which can be expressed in terms of ρ_{XY} as follows.

- Model 0: no correlation

$$\rho_{XY} = 0. \tag{9}$$

- Model I: maximum correlation

$$\rho_{XY} = min(1,\ \bar{\rho}_{XY}). \tag{10}$$

Here, $\bar{\rho}_{XY}$ is an upper bound on ρ_{XY} obtained from the Schwarz inequality $E\{XY\} \leq \sqrt{E\{X^2\}E\{Y^2\}}$. Note that this model is similar to the one proposed in [3], which employs the Schwarz bound directly to estimate $E\{XY\}$. The model of (10) is restated in terms of the correlation coefficient, and imposes the additional condition that it cannot exceed 1, thereby reducing the estimation error, as will be shown in Section IV-A.

We consider next a linear signal model.

- Model II: linear signal model

$$X = N + bY. \tag{11}$$

where b is a constant, N is a zero-mean noise random variable that is independent of Y. Given the moments of X and Y we can determine b and the variance of N, and finally obtain

$$\rho_{XY} = \min\left(\frac{E\{X\}\sigma_Y}{E\{Y\}\sigma_X},\quad \frac{E\{Y\}\sigma_X}{E\{X\}\sigma_Y},\quad 1\right). \tag{12}$$

It is easy to show that (12) implies satisfaction of the Schwarz upper-bound, which is hence not explicitly included. More complicated, non-linear models may be considered in a similar way, but the linearity property of Model II ensures that the cross-correlation can be calculated from the available first and second marginal moments without recourse to higher moments. This linear model was first proposed in [31]. It was later also applied in the CCA scheme of [28]. However, therein, the model was only used for correlation estimation of non-adjacent pixel pairs, while requiring calculation and storage of cross-correlation values for all adjacent pixel pairs. In contrast, our scheme directly applies the model to all pixel pairs, and estimates correlations only when they are needed by ROPE and only using the available first and second marginal moments. Hence, its computational complexity is considerably reduced.

All the above models are generally applicable to any pair of random variables, ignoring the obvious and important fact that we are concerned with correlation between the decoder reconstruction of *two pixels in a video frame*. Clearly, one expects the correlation between two pixel values to decay with the distance. In fact, the Euclidian inter-pixel distance has long been used in models for the autocorrelation function of spatial random fields that model images [32], [33]. Inspired by this line of reasoning and past modelling work, we propose an inter-pixel distance-based correlation model. Specifically, we employ exponentially decaying functions.

- Model III: pixel distance-based correlation model

$$\rho_{XY} = \min(e^{-\alpha d_{XY}},\ 1,\ \bar{\rho}_{XY}) \tag{13}$$

where d_{XY} is the Euclidian distance between the two pixels X and Y, and α is a constant. Note that other distance measures may be used where appropriate. Simulation results show that this model outperforms all other proposed CCA models.

B. Rounding Error Compensation (REC)

Rounding is typically employed whenever pixel filtering/averaging operations produce non-integer output values, as seen in H.264/AVC's subpixel prediction, intra-prediction, etc. Rounding, where a floating point value is quantized to the nearest integer, can be viewed as a special case of *uniform quantization* with quantization step size of one unit. The rounding error is

$$\Delta = X - Q(X) \tag{14}$$

where $Q(\cdot)$ denotes the rounding/quantization operation and X is a random variable, say, some filtered pixel at the decoder. From basic quantization theory we obtain the following properties:

$$\sigma_X^2 \to 0 : \Delta \to E\{X\} - Q\left(E\{X\}\right) \tag{15}$$

$$\sigma_X^2 \gg 1 : E\{\Delta\} \simeq 0,\ E\{\Delta^2\} \simeq 1/12. \tag{16}$$

From (15), we see that in the case of small variance, the rounding error approaches some typically small but non-zero value that depends on $E\{X\}$. In video coding, this virtually constant rounding error, although initially small, will be *propagated via inter-frame prediction*, resulting in accumulation that may seriously degrade the accuracy of end-to-end estimation, as will be shown in the simulation results. We emphasize that this is not a problem in pure source coding as both encoder and decoder perform rounding, and thus, yield the same reconstructions. The problem is in end-to-end estimation where the encoder does not know the exact value actually being rounded by the decoder. However, so far, rounding errors have been misleadingly viewed as insignificant, and completely ignored in all EED estimation techniques.

This work is primarily focused on H.264/AVC with subpixel prediction, where a 6-tap filter with coefficients [1/32, −5/32, 20/32, 20/32, −5/32, 1/32] and a 2-tap filter with coefficients [1/2, 1/2] are used for 1/2-pel and 1/4-pel interpolation, respectively. Since the interpolating pixels all take integer reconstruction values, the resultant interpolated 1/2-pels and 1/4-pels cannot take any arbitrary floating point values, but *only the corresponding 1/32-grid and 1/2-grid values*, respectively. Specifically, in our proposed REC schemes, we properly treat the 1/4-pel rounding input as 1/2-grid discrete random variables, while in the case of 1/2-pel rounding, we approximate the discrete input as *continuous* random variable, since the 1/32-grid represents relatively high resolution.

We propose two different REC solutions. Again, to maintain low complexity, we only use the quantities made available by basic ROPE, namely the first and second marginal moments. We hence pose the following problem.

- Given random variable X with known moments $E\{X\}$ and $E\{X^2\}$, define $Y = Q(X)$. Find $E\{Y\}$, and $E\{Y^2\}$.

1) Maximum Entropy Approach to REC: Our first approach appeals to the maximum entropy principle (MEP) [34], where given $E[X]$ and $E[X^2]$ we can find out the MEP optimal probability distribution of X, and then, trivially calculate $E[Y]$ and $E[Y^2]$. MEP states that, among all possible distributions satisfying the given constraints, one should select the one that maximizes the entropy. The rationale is that this choice maximizes the uncertainty and is hence the least restrictive while satisfying the constraints. In other words, any other choice reduces the uncertainty and therefore must make some implicit restrictive assumption. Despite some lingering controversy around the intuitive justification of the principle, MEP has been applied successfully in a remarkable variety of fields. In the context of ROPE, the available first and second moments provide natural constraints to derive the MEP-optimal probability distribution for the decoder reconstruction random variable. Moreover, the resulting distribution opens the door to attack many other important end-to-end estimation problems beside REC as will be explained later.

Specializing to H.264/AVC, in the 1/2-pel case, as X is assumed to be a continuous random variable (r.v.), the MEP optimal probability density function (pdf) given the first and second moments is a Gaussian distribution. In the 1/4-pel case, X is a 1/2-grid discrete r.v., whose optimal probability mass function (pmf) is a Gibbs distribution (see e.g., [35] for the straightforward derivation).

MEP-based REC approach:

- Generate MEP estimate for the distribution of X
  — for 1/2-pel X (approximately continuous r.v.)

$$X \sim N\left(\mu_X, \sigma_X^2\right) \tag{17}$$

  — for 1/4-pel X (1/2-grid discrete r.v.)

$$p(x) = \frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sum_x e^{-\frac{(x-\mu)^2}{2\sigma^2}}}. \tag{18}$$

- Extract the distribution of $Y = Q(X)$.
- Compute $E[Y]$ and $E[Y^2]$.

Note that μ_X and σ_X^2 in (17) are exactly the available mean and variance of X. In (18), parameters μ and σ^2 are chosen such that the moment constraints are satisfied. In practice, we employ the following iterative search algorithm to find μ and σ^2 that ensure that the Gibbs distribution moments match the known $E\{X\}$ and $E\{X^2\}$.

0) Set $\mu_0 = E\{X\}$, $\sigma_0^2 = \sigma_X^2$, and $n = 0$. Let the set $\mathcal{X} = \{x_0, \ldots, x_K\}$ contain all 1/2-grid values within $[E\{X\} - a \cdot \sigma_X, E\{X\} + a \cdot \sigma_X]$, where $a = 4.09$ (values beyond this range are neglected).

1) $n \leftarrow n + 1$.

2) Update the parameters so as to minimize the below error criterion Err_n.

— Fix $\mu_n = \mu_{n-1}$, exhaustively search for the best σ_n, with the search range $[0.5\sigma_{n-1}, 1.5\sigma_{n-1}]$ and search step size $0.1\sigma_{n-1}$.

— Fix σ_n, exhaustively search for the best μ_n, with the search range $[\mu_{n-1} - 1, \mu_{n-1} + 1]$ and search step size 0.1.

3) If $\mathrm{Err}_n < \mathrm{Err}_{\mathrm{th}}$ or $n > N$, set $\mu = \mu_n$ and $\sigma = \sigma_n$, and STOP. Otherwise, GOTO Step 1. (In our simulation, $\mathrm{Err}_{\mathrm{th}} = 0.0025\sigma_X^2$ and $N = 10$.)

The error criterion:

- Given μ_n, σ_n and support $\mathcal{X}$ produce the current Gibbs distribution from (18) and calculate moments $E\{X\}_n = \sum_k p_n(x_k)x_k$ and $E\{X^2\}_n = \sum_k p_n(x_k)x_k^2$.
- The error criterion measures the moment mismatch: $\mathrm{Err}_n = (E\{X\}_n - E\{X\})^2 + |E\{X^2\}_n - E\{X^2\}|$.

2) Quantization Theoretic Approximation in REC: In simulations we observe that, due to the parameter search procedure (usually 1∼2 iterations), REC by MEP may still incur non-trivial computational complexity. This motivates the proposal of an alternative REC scheme with a guaranteed low level of complexity. The approach has its roots in the quantization theoretic (QT) rounding error analysis of (15) and (16). Specifically, for Δ as defined in (14), we have

$$E\{\Delta\} = E\{X\} - E\{Y\} \tag{19}$$

$$E\{\Delta^2\} = E\{X^2\} + E\{Y^2\} - 2 \cdot E\{XY\}. \tag{20}$$

Hence

$$\begin{aligned}\sigma_\Delta^2 &= E\{\Delta^2\} - E\{\Delta\}^2 \\ &= \sigma_Y^2 + \sigma_X^2 + 2E\{X\}E\{Y\} - 2E\{XY\} \\ &= \sigma_Y^2 + \sigma_X^2 - 2\rho_{XY}\sigma_X\sigma_Y \\ &= (\sigma_Y - \rho_{XY}\sigma_X)^2 + (1 - \rho_{XY}^2)\sigma_X^2.\end{aligned} \tag{21}$$

Clearly

$$\sigma_\Delta^2 \geq \left(1 - \rho_{XY}^2\right) \cdot \sigma_X^2 \tag{22}$$

or equivalently

$$\rho_{XY}^2 \geq 1 - \frac{\sigma_\Delta^2}{\sigma_X^2}. \tag{23}$$

If $\sigma_X^2 \geq \sigma_\Delta^2$, we may rewrite this as

$$|\rho_{XY}| \in [A, 1], \quad \text{where } A = \sqrt{1 - \frac{\sigma_\Delta^2}{\sigma_X^2}}. \tag{24}$$

For large σ_X^2 (e.g., $\sigma_X^2 \gg 1$), we can reasonably assume that: 1) Δ is *uniformly* distributed and 2) ρ_{XY} is positive. Also, as usually $\sigma_\Delta^2 \ll 1$, A is very close to 1. Hence, we simply use

$$\rho_{XY} \simeq A. \tag{25}$$

Note that (19), (21), and (25) actually provide the means to compensate for rounding errors in the case of large σ_X^2. In the case of small variance, we simply round $E\{X\}$ directly. We summarize the complete scheme below.

- QT-based REC approach:
  — if $\sigma_X^2 > \gamma$

$$E\{Y\} = E\{X\} - E\{\Delta\}, \quad \sigma_Y^2 \simeq \sigma_X^2 - \sigma_\Delta^2. \tag{26}$$

— otherwise

$$E\{Y\} \simeq Q\left(E\{X\}\right), \quad \sigma_Y^2 \simeq \sigma_X^2. \tag{27}$$

The variance σ_Y^2 in (26) is obtained by plugging (25) and (24) into (21). The parameter γ is a heuristic threshold. Specializing to the subpixel prediction of H.264/AVC we obtain the following.

- For 1/2-pel X

$$E\{\Delta\} = 0, \quad \sigma_\Delta^2 = 1/12. \tag{28}$$

- For 1/4-pel X

$$E\{\Delta\} = 1/4, \quad \sigma_\Delta^2 = 1/16. \tag{29}$$

Note that (26)–(29) specify all the computation due to the QT-based approach, and reveals its extremely low complexity.

C. *ROPE Capabilities Extended by MEP*

As mentioned earlier, the proposed MEP approach essentially provides the means to approximate the distribution of the decoder reconstruction random variable. Given a distribution, it is possible to attack a much broader set of end-to-end estimation problems. First, besides rounding errors, it also gives a solution to compensate for clipping errors, which is another phenomenon that can safely be ignored in the context of source coding, but could have an impact on end-to-end estimation. Moreover, given the distribution, we can now estimate higher moments, whose potential use to improve system performance has been proposed in [36].

An important extension of ROPE, made possible by the MEP approximation, is in the new capability to estimate *any additive distortion criterion*, rather than the initial limitation to MSE distortion. By "additive distortion" we mean that the distortion can be written as the sum of individual contribution by pixels in a region of interest (e.g., a block). The end-to-end additive distortion can be expressed as

$$D_{n,\mathcal{B}} = \frac{1}{|\mathcal{B}|} \sum_{i \in \mathcal{B}} d_n^i \left(f_n^i, \tilde{f}_n^i\right) \tag{30}$$

where $\mathcal{B}$ denotes the region of interest. For example, the mean absolute error is an additive non-MSE distortion that is commonly used. Note that no restrictions are imposed on the component-wise distortion measures $d_n^i(f_n^i, \tilde{f}_n^i)$. The additive property implies that only marginal (and not joint) distributions are needed in order to calculate the distortion. There has been much interest in identifying distortion measures that better quantify the perceptual impact than MSE, e.g., [37]–[39]. The above opens the door to ROPE extensions that offer end-to-end estimation of a much broader class of distortion measures. Note that the same task is highly difficult, if not impossible, for block-based estimation approaches, as they heavily rely on second moment properties of MSE for the separation of overall EED into source coding induced distortion and channel loss induced distortion. Such properties, however, do not generally hold for other distortion metrics (e.g., mean absolute error).

IV. Simulation Results

A. *Estimation Accuracy*

We employed the JM9.0 H.264/AVC codec [40]. The performance of the proposed ROPE extensions is examined in the context of subpixel prediction, where 1/2-pel and 1/4-pel prediction were enabled. In our experiments, we only used single reference frame in motion estimation, and DIF was disabled. intra-prediction is only allowed from neighboring intra-coded MBs. The first frame in a sequence is coded as I-frame, and all remaining frames are coded as P-frames. There are no B-frames, and no weighted prediction. In this testing scenario, subpixel prediction is the only pixel-averaging operation to affect EED estimation.

We adopted the simple random intra-updating of the codec as implemented in the JM9.0 encoder [40], where for each frame a certain percentage of MBs (termed "intra-ratio") is forced to be intra-coded. At the encoder, data of one frame were packed into one packet for transmission. At the decoder, unless otherwise noted, when a packet was lost, the simple frame-copy scheme was used for concealment, which was also assumed at the encoder for distortion estimation.

A set of 500 randomly generated loss patterns were applied at each PLR, and the actual average distortion was computed for each pixel of a frame. Unless otherwise noted, the encoder assumed the correct value of PLR in its ROPE procedures. Estimation performance is measured by the "distortion difference ratio" defined as

$$\phi = \frac{\sum_n \sum_i \left| d_{n,\mathrm{Est}}^i - d_{n,\mathrm{Dec}}^i \right|}{\sum_n \sum_i d_{n,\mathrm{Dec}}^i}. \tag{31}$$

Here, $d_{n,\mathrm{Est}}^i$ and $d_{n,\mathrm{Dec}}^i$ denote, for pixel i of frame n, the distortion estimated at the encoder, and the actual decoder distortion averaged over all loss patterns, respectively.

All the testing sequences are QCIF sequences with 15 fps, and the first 150 frames of each sequence are coded. Due to limited space, figures are only given for the medium motion sequence Carphone ("Carphone") results. Results on many other sequences are summarized in the tables.

We tested the four CCA models of Section III-A, denoted by "CCA0"–"CCA3". For comparison, we also tested the simple Schwarz upper-bound only method proposed in [3], denoted by "sCCA1," to distinguish it from "CCA1" which combines the Schwarz inequality with the additional constraint $\rho_{XY} \leq 1$. For REC, we tested the MEP-based and QT-based methods, denoted by "MEP" and "QT," respectively. Unless otherwise noted, we set α in (13) of CCA3 to 0.10 and γ in (28) of QT REC to 0.5. For low-end benchmarking, we included the performance of ROPE where full-pixel prediction is assumed in the estimate as an approximation, despite the fact that the encoder actually employs subpixel motion compensation, thereby

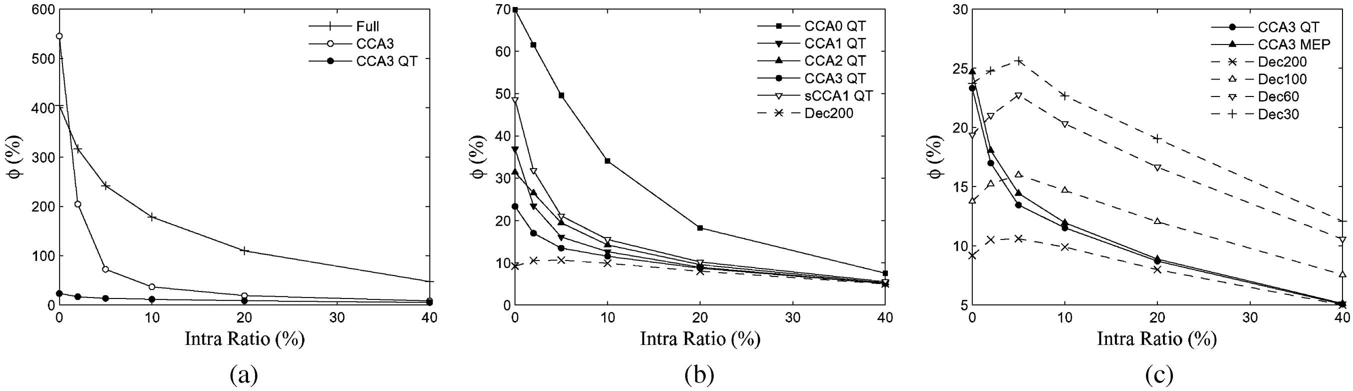


Fig. 1. Distortion estimation performance versus intra-ratio. Carphone, $p = 5\%$, 100 kb/s. (a) High level view. (b) Various CCA models. (c) Two REC schemes.

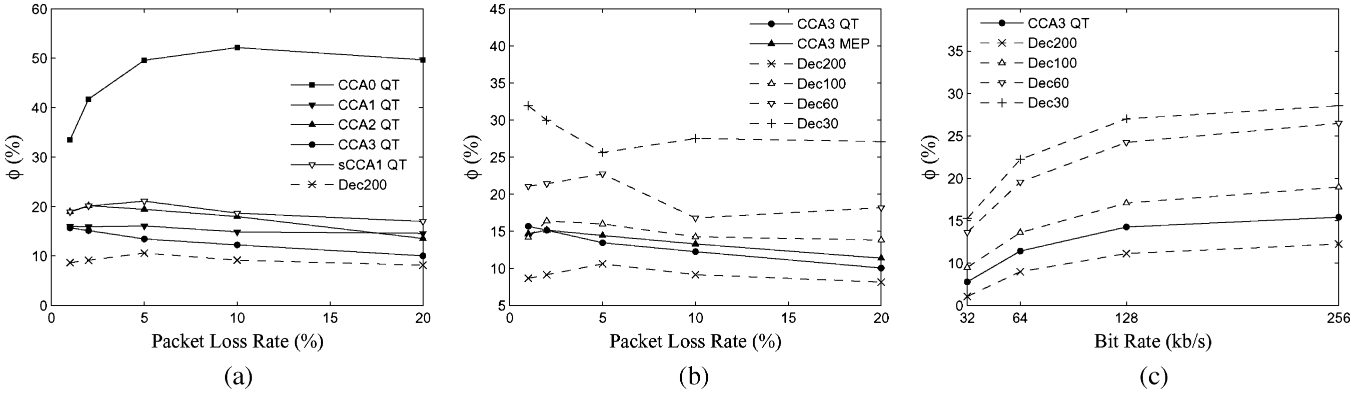


Fig. 2. Distortion estimation performance versus PLR (at coding rate of 100 kb/s), on Carphone with intra ratio $= 5\%$. (a) Various CCA models. (b) Two REC schemes. (c) Performance versus coding bit rate, with $p = 5\%$.

circumventing the need for CCA and REC ("Full") [1]. For high-end benchmarking, we included the results of the decoder simulation method proposed in [8] wherein the encoder simulates numerous packet loss patterns and the decoder reconstruction in order to produce an average EED estimate. For example, "Dec200" denotes applying 200 different packet loss patterns and decoding, while its ϕ actually represents the distortion measurement error between applying 200 and 500 loss patterns.

First, let us take a look at the estimation error versus intra-ratio results in Fig. 1. Fig. 1(a) gives a high level view. In comparison with "Full" which ignores both CCA and REC, it is evident that CCA generally improves estimation performance, and achieves high estimation accuracy at high intra-ratios, e.g., 20%. However, without REC, the performance dramatically degrades at low intra-ratios. At intra-ratio below 5%, the relative estimation error of CCA3 may exceed 100%, rendering the estimate unusable. On the other hand, the combination of both CCA and REC consistently yields high estimation accuracy at all intra-ratios. This result strongly demonstrates the importance of REC, especially at low intra-ratios.

Fig. 1(b) evaluates the various CCA models in conjunction with QT REC. We observe that all CCA models, except CCA0, achieve fairly high estimation accuracy, as their performance closely approaches the performance of Dec200. The somewhat surprising fact that the perhaps naive "maximum correlation" model of CCA1 consistently outperforms the linear signal model of CCA2, can be explained by the prevalence of very high correlation in the context of H.264/AVC subpixel prediction. It is further easy to see that overall CCA3 achieves the best performance. We also note from the comparison of CCA1 with the "Schwarz upper-bound only" approach sCCA1 [3], that further imposing $\rho_{XY} \leq 1$ improves estimation accuracy.

In Fig. 1(c), we compare the performance of the two proposed REC schemes combined with CCA3. For benchmark comparison, we also provide the accuracy of the "decoder simulation method" at various levels of complexity (number of packet loss patterns). Both MEP and QT generally provide considerable performance gains over Dec30, which is the de facto scheme in the JM11.0 reference encoder. Furthermore, we can see that with intra-ratio larger than 5%, the accuracy of the proposed schemes is always between those of Dec100 and Dec200, which signifies a fairly high estimation accuracy. In other words, to achieve a level of distortion estimation accuracy similar to the proposed approaches, one has to exhaustively simulate 100~200 decoding runs at the encoder, which represents considerably higher computational, storage and delay costs, as will be discussed further in Section IV-C-2. Comparing QT REC with MEP REC, we note that the overall performance of CCA3 QT is better than that of CCA3 MEP. Since, as discussed in Section III-B, QT REC is of very low complexity, we consider QT-based REC the preferred choice.

More results at various PLRs, coding bit rates, and over a variety of sequences are summarized in Fig. 2 and Table I, respectively. These show that the observations and conclusions we de-

TABLE I
DISTORTION ESTIMATION PERFORMANCE ON VARIOUS SEQUENCES. intra ratio $= 5\%$, $p = 5\%$, STEFAN AND FOOTBALL: 200 kb/s, OTHERS: 100 kb/s

ϕ (%)	Miss_am	Mobile	News	Carphone	Table	Foreman	Stefan	Football
CCA3	163.17	55.58	33.72	72.60	29.38	48.06	38.05	23.72
CCA0 QT	53.84	31.04	47.78	49.58	46.68	54.17	51.21	44.66
CCA1 QT	17.99	14.18	20.73	16.10	20.59	14.79	27.61	17.78
sCCA1 QT	23.38	18.29	25.85	21.09	25.93	18.87	33.04	20.96
CCA2 QT	24.24	16.29	16.72	19.46	17.11	19.40	17.34	14.04
CCA3 QT	**14.59**	**12.05**	**14.10**	**13.46**	**14.82**	**12.80**	**16.80**	**13.02**
CCA3 MEP	14.88	12.14	14.81	14.44	16.13	14.19	16.15	13.70
Dec100	**15.74**	**9.08**	**22.57**	**16.00**	**16.87**	**15.84**	**13.91**	**16.98**
Dec30	**20.72**	**15.24**	**35.92**	**25.63**	**30.69**	**26.67**	**24.22**	**30.44**

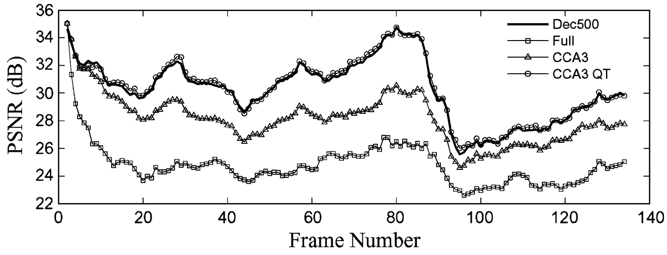


Fig. 3. Distortion frame-level estimation performance. Carphone, intra ratio $= 5\%$, $p = 5\%$, 100 kb/s.

rived from the previous figures largely hold across the broad test scenarios.

Fig. 3 depicts the time evolution of frame-level estimation performance results. We observe that CCA3 QT achieves highly accurate estimation as compared with Dec500, and that the absence of either REC or both CCA and REC always causes considerable performance degradation. interestingly, such degradation takes the form of *overestimation* of the distortion. Since pixel-averaging generally reduces error propagation, without CCA, an overestimated EED will be computed. REC yields an EED estimation procedure that better mimics or accounts for the actual encoding process, hence $E\{\tilde{f}_n^i\}$ is generally closer to f_n^i in (1). Moreover, REC generally decreases the variance of $\tilde{f}_n^i$ as seen from (26) and (27). Both of these factors tend to reduce the estimated EED in (1). Hence, excluding REC also results in distortion overestimation.

The experiments also evaluated the impact of the CCA3 model parameter α in (13). We found that for various sequences, both QCIF and CIF, at a variety of PLRs, intra-ratios, etc., the estimation performance was not highly sensitive to α, and the optimal value of α was always between 0.05 and 0.20. (Hence, we adopted $\alpha = 0.10$.) Due to limited space, we omit the detailed results herein.

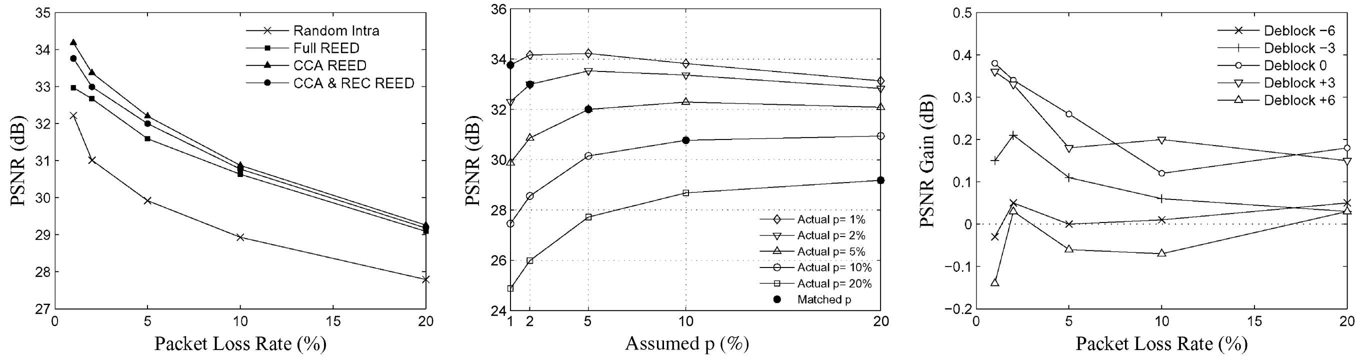
B. Impact on Coding Mode Selection

As discussed in Section I, ROPE can be applied in various ways to improve the error resilience of video coding. Here we demonstrate this in the context of REED coding mode selection. The mode selection problem is usually formulated as independently selecting the best coding mode for each MB/block to minimize a Lagrangian cost that weighs EED versus bit rate [1], [5].

The general experiment settings are identical to those of the previous subsection. System performance is measured by decoder average luminance PSNR over 500 loss patterns. We tested the performance of REED coding mode selection, where competitors differ in EED estimation: original ROPE with full-pixel approximation ("Full REED"), revised ROPE with CCA3 only ("CCA REED"), and revised ROPE with CCA3 and QT REC ("CCA & REC REED"). For comparison, we also tested the aforementioned random intra-updating scheme ("Random intra"), where the forced intra-ratio per frame exactly equals the PLR.

1) REED Performance Without Mismatch: Results on REED performance without mismatch are summarized in Fig. 4(a) and Table II. We can see that both "CCA REED" and "CCA & REC REED" significantly outperform "Full REED" and "Random intra," especially at low PLRs, e.g., $p \leq 5\%$. This proves that the EED estimation accuracy increase offered by the revised ROPE generally translates into significant overall system performance gains. From the results, a "surprising" observation is that more accurate distortion estimation does not always lead to better overall PSNR performance, as "CCA REED" consistently achieves higher average PSNR than "CCA & REC REED." The gain even reaches 0.72 dB for Mobile sequence in Table II. 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 [42] for non-error-resilient RD coding, and [43] for REED coding methods, where coding parameters of a group of frames are jointly optimized. Intuitively, accounting for propagation impact on future frames, which implies more distortion cost, is somewhat equivalent to *overestimating* EED in the current frame. On the other hand, as shown earlier in Fig. 3, both "CCA3" and "Full" 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 EED, as is the case of "Full."

2) REED Performance With Mismatch: So far, we assumed that both the PLR and EC scheme used by the encoder for ROPE computations correspond exactly to the actual PLR and



(a) (b) (c)

Fig. 4. REED mode selection performance. Carphone, 100 kb/s. (a) Without mismatch, 100 kb/s. (b) With PLR mismatch. (c) With DIF mismatch.

TABLE II
REED MODE SELECTION PERFORMANCE ON VARIOUS SEQUENCES. $p = 2\%$, STEFAN AND FOOTBALL: 200 kb/s, OTHERS: 100 kb/s

PSNR (dB)	Miss_am	Mobile	News	Carphone	Table	Foreman	Stefan	Football
Random Intra	38.04	25.77	33.03	31.01	27.60	27.56	24.04	27.09
Full REED	40.66	25.19	35.50	32.67	31.35	30.83	26.22	29.15
CCA REED	41.90	27.53	36.91	33.37	32.16	31.53	27.13	29.24
CCA & REC REED	41.70	26.81	36.90	32.99	32.07	31.34	27.07	29.22

TABLE III
REED MODE SELECTION PERFORMANCE WITH EC MISMATCH ON VARIOUS SEQUENCES. $p = 5\%$, STEFAN AND FOOTBALL: 200 kb/s, OTHERS: 100 kb/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**

EC scheme at the decoder. We also assumed that there is no DIF in both encoding and decoding. However, as discussed in Section II, mismatched PLR, EC, or DIF is of practical concern. Clearly, mismatch will compromise the EED estimation accuracy. Next, we investigate the impacts from such mismatch on the overall REED mode selection performance.

Results with mismatched PLR are shown in Fig. 4(b). 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. This result is very consistent with our earlier observation that the fact that ROPE neglects the impact on future frames may be roughly compensated for by overestimating EED.

Fig. 4(c) gives the DIF mismatch result. 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) [23]. Herein, "No deblocking" denotes the case of no DIF mismatch, while "Deblock -6" $\sim$ "Deblock $+6$" denote the mismatch cases at different DIF strength levels. In the figure, we directly give the average PSNR gains of the cases with DIF mismatch over those without DIF (and without mismatch). It is easy to see that, although DIF mismatch degrades the estimation performance, this does not necessarily lead to degraded overall REED performance. For deblocking strength between -3 and $+3$, DIF mismatch usually yields better performance than that without mismatch. Note that pixel-averaging operations in DIF generally have the effect of reducing error propagation from packet loss. Therefore, ROPE overestimates EED, and the overestimation argument suggests that the damage will not be excessive. Finally, it is clear that the benefits of employing DIF at mild strength, at least outweigh any damage due to mismatch.

TABLE IV
COMPUTATIONAL COMPLEXITY MEASURED BY CODING TIME INCREASE OVER STANDARD (NOT ERROR-RESILIENT) CODER, AVERAGED OVER ALL TESTING SEQUENCES AT $p = 5\%$ AND intra ratio $= 5\%$

ROPE Variant	Average Coding Time Increase (%)
Full	33.91
CCA3	123.13
CCA3 QT REC	149.15

To investigate the EC mismatch impact on REED mode selection, we use motion-copy EC as the mismatched EC scheme at the decoder. The encoder always assumes frame-copy EC. Recall that motion-copy EC employs motion-vectors from collocated MBs in the previous frame to conceal the MBs of the current frame via motion compensation [41]. The results are summarized in Table III. 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 the experiment,

TABLE V
COMPUTATIONAL COMPLEXITY COMPARISON BETWEEN REVISED ROPE AND THE DECODER SIMULATION METHOD

	Compuational Complexity Increase			Memory Increase
	$[+](A)$	$[\times](A)$	$[mem](A/256)$	(Bytes per pixel)
DecK	$(44-p)K$	$27K+32$	K	K
Dec100 ($p=0.1$)	4390	2732	100	100
Dec30 ($p=0.1$)	1317	842	30	30
Revised ROPE	418	618	4	8

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.

C. Complexity

1) Complexity of Proposed ROPE Variants: Simulations were run on Pentium IV 3.0 GHz with 504 MB RAM. Table IV summarizes the average increase in encoding time due to ROPE variants, in terms of a percentage of standard encoding time (i.e., without ROPE). We note that ROPE with CCA3 QT increases the encoding time by 149.15%. It should be mentioned that these results are for the existing ROPE implementation, which has not been optimized to reduce the run time. For example, for implementation expediency, we currently maintain two complete 1/4-pel resolution maps for the first and second moments respectively for each frame. However, not all calculated subpixel moments of the reference frame will be used in calculating the full-pixel moments of the current frame. Hence, computation and storage complexity may be greatly reduced, if subpixel quantities are only calculated whenever they are needed. In terms of storage/memory consumption, we note that for each pixel, the standard coded integer pixel costs 1 byte, while ROPE additionally needs $4\times 2=8$ bytes to store in floating point the first and second moments. As for the various proposed extensions, they only require some small amount of in-field local calculation memory, which incurs a modest memory cost increase. Overall, it appears that the revised ROPE offers significant performance benefits at complexity cost that would be acceptable in many applications.

2) Complexity Comparison With the Decoder-Simulation Method: We compare the complexity of ROPE (with CCA3 and QT REC) with that of the exhaustive decoder simulation method. DecK denotes decoder simulation with K decoding runs, and A denotes the total number of integer pixels in a frame. We assume that only the first frame is an I-frame and all the rest are P-frames. We consider the distortion estimation complexity for one single P-frame. Moreover, for simplicity but without loss of generality, we assume that all the MBs in a P-frame are coded with inter-16 × 16 mode. For each P-frame, we consider the computational complexity of estimating the complete 1/4-pel resolution EED map. Hence, one needs to first calculate the full-pixel resolution distortion map, then the 1/2-pel distortion map, and finally the 1/4-pel distortion map.

First, we take a look at full-pixel distortion map calculation. We denote the memory fetching, addition, and multiplication operation by $[\mathrm{mem}]$, $[+]$, and $[\times]$, respectively. For DecK, if in a certain decoding run, the current P-frame is lost, frame-copy EC will be applied, which only requires one time memory fetching operation per MB, i.e., $1[\mathrm{mem}]$. Otherwise, one needs to calculate the simulated decoder reconstruction with $1[\mathrm{mem}]$ per MB and one addition, i.e., $1[+]$ per full-pixel. For simplicity, we ignore the memory fetching operation complexity of frame-copy EC. Approximately, there are $(1-p)K$ no loss runs, and pK loss runs. After that, one needs to average over K decoder reconstructions, which takes $(K-1)[+]$ and $1[\times]$. Finally, calculating distortion takes $1[+]$ and $1[\times]$. Similarly, for the basic ROPE calculation described in (1), (4) and (5), for each full-pixel, it takes $2[+]$ and $2[\times]$ for first moment calculation, $3[+]$ and $5[\times]$ for second moment calculation, and $2[+]$ and $2[\times]$ for distortion calculation, while for each MB it takes $4[\mathrm{mem}]$. In summary, to calculate the full-pixel distortion map, DecK requires $((1-p)K+(K-1)+1)A[+]=(2-p)KA[+]$, $2A[\times]$, and $KA/256[\mathrm{mem}]$, while ROPE requires $7A[+]$, $9A[\times]$ and $4A/256[\mathrm{mem}]$.

As for 1/2-pel and 1/4-pel distortion map calculation, we assume fast table-look-up is used to get the exponential number in (13) (as overall there are only five different possible inter-pixel distance values in subpixel interpolations). Moreover, we approximate the complexity of one square root operation in the Schwarz upper-bound calculation to be the same as that of one time multiplication, i.e., $1[\times]$. Similarly, one can work out the overall complexity for 1/2-pel and 1/4-pel distortion map calculations.

The final complexity results are summarized in Table V. Regarding memory costs, it is easy to see that for each pixel, DecK additionally needs K bytes for the K decoding runs. From the table, it is obvious that both computational complexity and memory cost of DecK grow linearly with K, while our revised ROPE only requires a limited increase on computational and memory costs. As specific examples, Table V also gives the costs of the Dec100 and Dec30 methods, both of which considerably exceed that of revised ROPE. Recall that the EED estimation accuracy of revised ROPE is better than that of Dec100, as shown earlier. Therefore, overall, the revised ROPE approach is much more cost-effective than the exhaustive decoder simulation method.

V. CONCLUSION

In this paper we considered open problems that pose practical obstacles to the general applicability of ROPE. One is the emergence of cross-correlation terms in the estimate. Another problem involves proper accounting for rounding operations and their cumulative impact, within the recursive estimate. We propose low-complexity solutions to these problems and demonstrate by simulations (H.264/AVC with 1/4-pel prediction) that the problems are of significant impact, and that the solutions substantially enhance the estimation performance, as well as the overall system performance.

The proposed techniques are applicable to other modules where the underlying problem emerges including, in particular, weighted prediction, intra-prediction, overlapped block motion compensation, linear transforms, and a variety of advanced error concealment schemes. Moreover, a spin-off of this work was the introduction of a means to estimate the distribution of the decoder reconstructed pixel based on the inference principle of maximum entropy. This opens the door to broaden the scope of ROPE applications to include end-to-end estimation for any additive distortion measure, clipping error compensation, and more. The proposed modifications demonstrate the potential and applicability of ROPE and its variants within a broad spectrum of practical settings, and in particular that of H.264/AVC.

In this paper, we also investigated the impacts of PLR, EC and DIF mismatch on the overall REED optimization performance. We found that although such mismatch compromises distortion estimation, in practice, it may not significantly degrade the overall REED coding performance. It should be noted that our observations regarding the impact of EC mismatch, while encouraging, were obtained experimentally in the context of specific choices of EC and may not generalize to other settings. Clearly, it is important to examine the mismatch impact given any specific system settings.

ACKNOWLEDGMENT

The authors are grateful to C. Schmidt who was the first to notice the impact of rounding and clipping on ROPE's performance.

REFERENCES

[1] R. Zhang, S. L. Regunathan, and K. Rose, "Video coding with optimal inter/intra-mode switching for packet loss resilience," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 6, pp. 966–976, Jun. 2000.

[2] T. Wiegand, N. Farber, K. Stuhlmuller, and B. Girod, "Error-resilient video transmission using long-term memory motion-compensated prediction," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 6, pp. 1050–1062, Jun. 2000.

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

[4] H. Yang and K. Rose, "Rate-distortion optimized motion estimation for error resilient video coding," in *Proc. ICASSP*, 2005, pp. 187–190.

[5] G. Cote, S. Shirani, and F. Kossentini, "Optimal mode selection and synchronization for robust video communications over error-prone networks," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 6, pp. 25–34, Jun. 2000.

[6] Y. Zhang, W. Gao, H. Sun, Q. Huang, and Y. Lu, "Error resilience video coding in H.264 encoder with potential distortion tracking," in *Proc. ICIP 2004*, 2004, vol. 1, pp. 173–176.

[7] S. Ekmekci and T. Sikora, "Recursive decoder distortion estimation based on AR(1) source modeling for video," in *Proc. ICIP*, Singapore, 2004, pp. 187–190.

[8] T. Stockhammer, T. Wiegand, and S. Wenger, "Optimized transmission of H.26L/JVT coded video over packet-lossy networks," in *Proc. ICIP*, Rochester, NY, 2002, vol. 2, pp. 173–176.

[9] Y. Shen, P. C. Cosman, and L. Milstein, "Video coding with fixed length packetization for a tandem channel," *IEEE Trans. Image Process.*, vol. 15, no. 2, pp. 273–288, Feb. 2006.

[10] A. R. Reibman, L. Bottou, and A. Basso, "DCT-based scalable video coding with drift," in *Proc. ICIP*, 2001, vol. 2, pp. 989–992.

[11] H. Yang, R. Zhang, and K. Rose, "Drift management and adaptive bit rate allocation in scalable video coding," in *Proc. ICIP*, 2002, vol. 2, pp. 49–52.

[12] A. R. Reibman, "Optimizing multiple description video coders in a packet loss environment," presented at the Packet Video Workshop, 2002.

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

[14] F. Zhai, C. E. Luna, Y. Eisenberg, T. N. Pappas, R. Berry, and A. K. Katsaggelos, "Joint source coding and packet classification for video streaming over differentiated services networks," *IEEE Trans. Multimedia*, vol. 7, no. 4, pp. 716–726, Aug. 2005.

[15] F. Zhai, Y. Eisenberg, T. N. Pappas, R. Berry, and A. K. Katsaggelos, "Rate-distortion optimized hybrid error control for packetized video communications," *IEEE Trans. Image Process.*, vol. 15, no. 1, pp. 40–53, Jan. 2005.

[16] F. Zhai, Y. Eisenberg, T. N. Pappas, R. Berry, and A. K. Katsaggelos, "Joint source-channel coding and power adaptation for energy efficient wireless video communications," *Signal Process.: Image Commun.*, vol. 20/4, pp. 371–387, Apr. 2005.

[17] E. Masala, H. Yang, and K. Rose, "Rate-distortion optimized slicing, packetization and coding for error-resilient video transmission," in *Proc. IEEE DCC*, 2004, pp. 182–191.

[18] A. Majumdar, J. Wang, and K. Ramchandran, "Drift reduction in predictive video transmission using a distributed source coded sidechannel," in *Proc. 12th Ann. ACM Int. Conf. Multimedia*, New York, 2004, pp. 404–407.

[19] M. Fumagalli, M. Tagliasacchi, and S. Tubaro, "Improved bit allocation in an error-resilient scheme based on distributed source coding," in *Proc. ICASSP*, May 2004, vol. 2, pp. 61–64.

[20] H. Yang and K. Rose, "Source-channel prediction in error resilient video coding," in *Proc. ICME*, 2003, vol. 2, pp. 233–236.

[21] H. Yang and L. Lu, "A novel source-channel constant distortion model and its application in error resilient frame-level bit allocation," in *Proc. ICASSP*, 2004, vol. 3, pp. 277–280.

[22] M. Fumagalli, R. Lancini, and S. Tubaro, "Video quality assessment from the perspective of a network service provider," in *Proc. Int. Workshop Multimedia Signal Process.*, BC, Canada, Oct. 2006, pp. 324–328.

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

[24] *Video Coding for Low Bitrate Communications*, ITU-T Rec. H.263 Version 2 (H.263+), Jan. 1998.

[25] Y. Wang and Q.-F. Zhu, "Error control and concealment for video communication: A review," *Proc. IEEE*, vol. 86, no. 5, pp. 974–997, May 1998.

[26] T. Stockhammer, M. M. Hannuksela, and T. Wiegand, "H.264/AVC in wireless environments," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 657–673, Jul. 2003.

[27] K. Stuhlmuller, *Modeling and Optimization of Video Transmission Systems*. Berlin, Germany: Shaker Verlag, 2000, pp. 45–55.

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

[29] Y. Wang, Z. Wu, J. Boyce, and X. Lu, "Modelling of distortion caused by packet losses in video transport," in *Proc. ICME*, Jul. 2005, pp. 1206–1209.

[30] S. Wenger, "H.264/AVC over IP," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 645–656, Jul. 2003.

[31] H. Yang and K. Rose, "Recursive end-to-end distortion estimation with model-based cross-correlation approximation," in *Proc. ICIP*, Sep. 2003, vol. 3, pp. 469–472.

[32] J. R. Jain and A. K. Jain, "Displacement measurement and its application in interframe image coding," *IEEE Trans. Commun.*, vol. COM-29, no. 12, pp. 1799–1804, Dec. 1981.

[33] B. Girod, "The efficiency of motion-compensating prediction for hybrid coding of video sequences," *IEEE J. Sel. Areas Commun.*, vol. SAC-5, no. 7, pp. 1140–1154, Aug. 1987.

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

[35] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991, pp. 266–279.

[36] Y. Eisenberg, F. Zhai, C. E. Luna, T. N. Pappas, R. Berry, and A. K. Katsaggelos, "Variance-aware distortion estimation for wireless video communications," in *Proc. ICIP*, 2003, vol. 1, pp. 89–92.

[37] A. Leontaris and A. R. Reibman, "Comparison of blocking and blurring metrics for video compression," in *Proc. ICASSP*, 2005, vol. 2, pp. 585–588.

[38] E.-P. Ong, X. Yang, W. Lin, Z. Lu, and S. Yao, "Perceptual quality metric for compressed videos," in *Proc. ICASSP*, 2005, vol. 2, pp. 581–584.

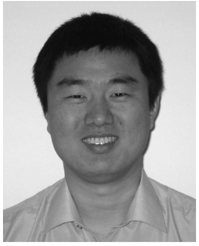
[39] I. Avcibas, B. Sankur, and K. Sayood, "Statistical evaluation of image quality measures," *J. Electron. Imag.*, vol. 11, no. 2, pp. 206–223, Apr. 2002.

[40] Fraunhofer Heinrich-Hertz-Institut, [Online]. Available: http://www.iphome.hhi.de/suehring/tml/download

[41] M. C. Hong, L. Kondi, H. Scwab, and A. K. Katsaggelos, "Error concealment algorithms for compressed video," *Signal Process.: Image Commun.*, vol. 14, pp. 437–492, 1999.

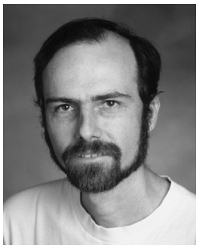
[42] K. Ramchandran, A. Ortega, and M. Vetterli, "Bit allocation for dependent quantization with applications to multiresolution and MPEG video coders," *IEEE Trans. Image Process.*, vol. 3, no. 5, pp. 533–545, Sep. 1994.

[43] R. Zhang, S. L. Regunathan, and K. Rose, "Prescient mode selection for robust video coding," in *Proc. ICIP*, Oct. 2001, vol. 1, pp. 974–977.



Hua Yang (S'02–M'06) received the B.S. and M.S. degrees in electrical engineering from Tsinghua University, Beijing, China, in 1997 and 2000, respectively, and the Ph.D. degree in electrical and computer engineering from University of California, Santa Barbara, in 2005.

Since 2005, he has been with Mobility Group at Thomson Corporate Research, Princeton, NJ. In 2003, he was a summer intern with Multimedia Technologies Group at IBM T. J. Watson Research Center, Yorktown, NY. His research interests include video coding, video transmission over networks, and perceptual video quality metrics and improvement.



Kenneth Rose (S'85–M'91–SM'01–F'03) received the Ph.D. degree in 1991 from the California Institute of Technology, Pasadena.

He is currently a Professor with the Department of Electrical and Computer Engineering, University of California at Santa Barbara. His main research activities are in the areas of information theory and signal processing, and include rate-distortion theory, source and source-channel coding, audio and video coding and networking, pattern recognition, and nonconvex optimization. He is also particularly interested in the relations between information theory, estimation theory, and statistical physics, and their potential impact on fundamental and practical problems in diverse disciplines.

He currently serves as Area Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS. He was co-recipient of the 1990 William R. Bennett Prize Paper Award of the IEEE Communications Society, and of the 2004 IEEE Signal Processing Society Best Paper Award (in the area of image and multidimensional signal processing).