# Optimizing Motion Compensated Prediction for Error Resilient Video Coding

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

***Abstract*—This paper is concerned with optimization of the motion compensated prediction framework to improve the error resilience of video coding for transmission over lossy networks. First, accurate end-to-end distortion estimation is employed to optimize both motion estimation and prediction within an overall rate-distortion framework. Low complexity practical variants are proposed: a method to approximate the optimal motion via simple distortion and source coding rate models, and a source-channel prediction method that uses the expected decoder reference frame for prediction. Second, reference frame generation is revisited as a problem of filter design to optimize the error resilience versus coding efficiency tradeoff. The special cases of leaky prediction and weighted prediction (i.e., finite impulse response filtering), are analyzed. A novel reference frame generation approach, called "generalized source-channel prediction", is proposed, which involves infinite impulse response filtering. Experimental results show significant performance gains and substantiate the effectiveness of the proposed encoder optimization approaches.**

***Index Terms*—Error resilience, motion compensation, prediction, rate-distortion, source-channel prediction, weighted prediction.**

## I. Introduction

A critical concern in the design of video-over-network systems is how to effectively account for and mitigate the impact of packet loss on the overall video quality. An important aspect of the problem is that of redesigning the various video coding components such that they are optimized for both the source and network parameters. Of particular interest to us here is motion compensated prediction (MCP). Video coding standards generally adopt the classical predictive quantization framework, which uses past *encoder-reconstructed* frames for prediction. This conventional framework was primarily designed to improve source coding efficiency. However, in the case of lossy communications, encoder and decoder mismatch is inevitable, and a revised paradigm is needed. This paper considers the fundamental problem of achieving optimal MCP for video transmission over lossy networks.

Manuscript received May 12, 2008; revised August 28, 2009. First published September 22, 2009; current version published December 16, 2009. 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. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Antonio Ortega.

H. Yang is with Thomson Corporate Research, Princeton, NJ 08540 USA (e-mail: hua.yang2@thomson.net).

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/TIP.2009.2032895

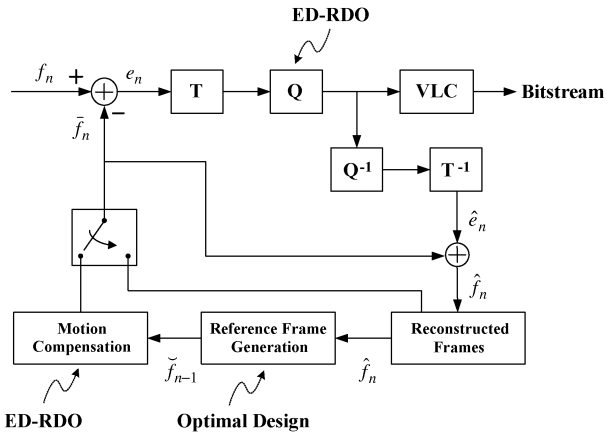



Fig. 1. Video encoder models optimized by the proposed method.

In video networking, system performance is determined by the ultimate playback video quality at the decoder, which is quantitatively measured as the overall end-to-end distortion. The fundamental performance tradeoff is of distortion versus rate, or the rate-distortion (RD) problem. Hence, the general framework employed for error-resilient video networking is that of *end-to-end distortion-based rate-distortion optimization* (ED-RDO). In the scenario of precompressed video streaming (e.g., video-on-demand) this framework can be applied to the video transport module to optimize the packetization scheme [1] or the packet delivery policy [2], [3]. In this paper, we focus on the scenario of live video streaming (e.g., video conferencing/telephony) where ED-RDO can further be applied to optimize the source coding modules and options. For example, at the macro-block (MB) level, optimization may involve various coding modes or parameters including: intra/intramode [4], [5], quantization step size [4], prediction reference frame [6], [7], coding modes associated with layered coding [8], [9] or multiple description coding [10], [11], etc.

Fig. 1 depicts a standard video encoder and identifies the modules targeted for improved optimization by this work. Herein, $f_n$, $\bar{f}_n$, and $\hat{f}_n$ denote the original input, the prediction, and the reconstruction of frame $n$, respectively. The prediction residue and its reconstruction are denoted by $e_n$ and $\hat{e}_n$, respectively. The prediction reference frame $\breve{f}_{n-1}$ is used for predicting frame $n$. Note that, for simplicity, a single reference frame is shown. The first part of this paper extends the applicability of the ED-RDO framework to optimize motion estimation (ME) or motion compensation, and prediction, and proposes new methods within this framework. The second part revisits the design of reference frame generation (RFG) so as to directly optimize the error resilience versus coding efficiency tradeoff.

### A. Rate-Distortion Optimized Motion Estimation

Conventional ME techniques improve performance in terms of source coding efficiency without explicit accounting for error robustness considerations in video streaming applications, e.g., various rate-constrained ME techniques [12]–[14], or RD-optimized ME [7], [15], [16]. Since motion vectors (MVs) are directly and critically involved in the error propagation mechanism encountered in transmission over lossy packet networks, ME may have a considerable impact on error resilience. Consequently, one objective of this work is to explore the potential of MV optimization for improving error resilience. Relevant earlier work appeared in [17] and [18], where the emphasis was mainly on error resilience advantages of using multiple frames in MCP. In [17], an error resilient variant of rate constrained ME was proposed, which incorporates a heuristic and rough estimate of the decoder prediction error within the Lagrangian cost. The current paper subsumes our early work [21], which was the first to propose a basic approach to high accuracy ED-RDO based on extensions to the recursive optimal per-pixel estimate (ROPE) method of [4]. Similar schemes employing ROPE were later proposed in [19] and [20] (the latter also included a higher accuracy rate model).

We hence propose ED-RDO-based ME using exhaustive coding to achieve the operational RD bound, and a low complexity variant based on simple quadratic RD modeling. This provides useful lower and upper bounds on the gains achievable by such a technique. To simplify and clarify both the derivation and the evaluation, we apply ROPE in its ideal setting, in particular assuming no modeling mismatches due to pixel averaging operations such as in sub-pixel motion compensation. Effective solutions to extend ROPE to mitigate mismatch issues are provided in [37] (and references therein), but are not incorporated here so as to avoid diluting the focus. In contrast, [17] uses heuristic distortion estimation while [20] applies ROPE in a practical setting of sub-pixel prediction and in-loop filtering but without mitigating the modeling mismatches, both of which significantly compromise the distortion estimation accuracy. We re-emphasize that our objective here is to bound the range of complexity-precision tradeoff by providing the two extremes of accurate RD modeling via exhaustive encoding, versus low complexity quadratic RD modeling. The approach in [20] uses an advanced heuristic rate model, which is expected to yield performance somewhere within our benchmarked performance range.

### B. Rate-Distortion Optimized Prediction

The second contribution of this work concerns the problem of prediction optimization. So far, many techniques have been proposed to modify the prediction mechanism with an eye towards error resilience, such as independent slice coding, video redundancy coding [22], multiple frame motion compensation [7], [17], [18], and reference picture selection [6], [23]. A common feature these methods share with conventional techniques is that prediction is based on past *encoder-reconstructed* frames. This predictive framework was originally designed for error-free transmission scenarios, where closed-loop prediction ensures no mismatch between the encoder and the decoder. However, in practical networking applications, encoder and decoder mismatch (or drift) is largely inevitable, motivating a reformulation of the optimal prediction problem for the setting of lossy transmission.

Here, too, we appeal to the general ED-RDO framework and hence define optimal prediction as the one that yields the best overall end-to-end RD tradeoff. An effective hybrid search algorithm is developed to solve the problem in the context of H.264. To reduce the optimization complexity, we also propose an efficient scheme source-channel prediction (SCP), which employs the *expected decoder reference frame* for prediction. Our preliminary SCP approach appeared first in [24]. A later publication proposing a similar SCP scheme appeared in [25], apparently independently of [24].

We note in passing that the problem of prediction optimization has a long history already within the pure source coding setting, motivated by the fact that optimal prediction and quantization are effectively inseparable. In [26], the problem was formulated as "optimal quantization" and was investigated for improving the coding efficiency of H.263, where a trellis-based search algorithm was developed to find the globally optimal solution. Trellis-based quantization is now available in the JM reference encoder of H.264/AVC. Nevertheless, such techniques implicitly assume error-free transmission. In this paper, we focus on the optimization problem in the context of error resilience and lossy transmission.

### C. Optimized Reference Frame Generation

So far, we have considered the problem of optimizing the encoder MCP decisions subject to the given (and fixed) decoder MCP mechanism. We next extend the scope, and attack the more general problem of optimizing the overall performance given the freedom to redesign the prediction mechanism at *both encoder and decoder*. In particular, we focus on the RFG module, see Fig. 1, where it is explicitly shown to be decoupled from motion compensation. Such RFG-MCP decoupling is important, as it enables consideration of infinite impulse response (IIR) filtering architectures for RFG. Specifically, we investigate the error resilience versus coding efficiency performance of candidate RFG architectures under various scenarios (e.g., with or without ED-RDO optimized intra-updates). We consider a broad "spectrum" of filter design, including the conventional "complete" prediction, leaky or "partial" prediction, weighted or finite impulse response (FIR) prediction, and the newly proposed "generalized source-channel" prediction (GSCP) which employs IIR filtering.

To the best of our knowledge, most if not all past research efforts on error-resilient prediction architectures focused on either leaky prediction or weighted prediction (also known as "multihypothesis" MCP [27]). Specifically, no prior work taps the potential of IIR filtering architectures. Moreover, leaky prediction efforts were largely focused on layered, or scalable, video coding [28]–[30] under the assumption of perfectly reliable base layer, while weighted prediction was applied within MCP [27], [31], [32]. Their applicability to RFG (i.e., decoupled from motion compensation) for basic son-scalable video coding has largely been ignored. Our work on optimal RFG in this paper attempts to fill these gaps.

### D. Additional Comments

We emphasize that accurate distortion estimation itself is a critical underlying problem that impacts the ultimate performance of the proposed error resilient ME and prediction schemes. Herein, we build on the ROPE method [4], which has been widely adopted in a variety of error resilient video coding schemes [7], [9], [10], [33]–[36]. Recent advances in ROPE further expand its capability to accommodate useful practical application scenarios, such as sub-pixel prediction [37], more complicated error concealment [37], bursty packet loss [34], or more practical packetization schemes [38]. In this paper we mainly focus on characterizing the performance gains achievable by the proposed approaches. To avoid dilution of the focus with (potentially significant but) "fuzzifying" considerations, we make some simplifying assumptions that ensure the accuracy of the adopted ROPE. For example, we employ H.264 with full-pixel prediction, assume one-frame-per-packet packetization, an independent random packet loss model, and frame-copy error concealment. All these assumptions can be removed by appealing to the above ROPE extensions, but doing so here would compromise the focus. Similarly, simulations assume no mismatch in packet loss rate and decoder error concealment, and disable in-loop filtering. We reported on a study on the impacts of such mismatch in [39], where it was noted that performance degradation was similar to that of existing conventional techniques, and that the gains were, therefore, largely maintained.

The rest of the paper is organized as follows. In Section II, preliminaries for ED-RDO and the basic ROPE method are provided to facilitate discussion of the proposed RD optimal ME and prediction schemes. Section III contains the detailed formulation and analysis of our ED-RDO ME, and its low complexity approximation. Optimal encoder-based prediction, its low complexity version of SCP and our proposed search algorithm are discussed in Section IV. We introduce the proposed GSCP framework along with the other RFG candidate designs in Section V. Section VI provides simulation results and analysis.

## II. Preliminaries and the Basic ROPE Approach

Practical RD optimization problems can be equivalently formulated as minimization of the appropriate Lagrangian cost

$$J = E\{D\} + \lambda R \tag{1}$$

where $\lambda$ is the Lagrangian multiplier, $E\{D\}$ and $R$ denote the expected distortion and coding rate cost, respectively. To accurately estimate the end-to-end distortion, we adopted the ROPE approach as proposed in [4], which is defined as follows. 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, $\tilde{f}_n^i$ is considered a random variable at the encoder. The overall expected mean-squared-error (MSE) distortion of a pixel is

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

which is clearly determined by the first and second moments of the decoder reconstruction. ROPE consists of an optimal recursive algorithm to accurately calculate these two moments for each pixel of a frame [4].

Let us assume for simplicity that packet loss events are independent, the packet loss rate $p$ is available at the encoder, each frame is transmitted in one packet, error concealment at the decoder copies reconstructed pixels from the previous frame, and prediction at the encoder only employs the previous reconstructed frame. Note that all these assumptions can be discarded (and, in particular, expressions are trivially extendible to cover multiple frame motion compensation). The recursion formulae of ROPE are as follows.

- Pixel in an intracoded MB

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

$$E\left\{\left(\tilde{f}_n^i\right)^2\right\} = (1-p)\left(\hat{f}_n^i\right)^2 + pE\left\{\left(\tilde{f}_{n-1}^i\right)^2\right\}. \tag{4}$$

- Pixel in an intercoded MB

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

$$\begin{aligned} E\left\{\left(\tilde{f}_n^i\right)^2\right\} &= (1-p)\left(\left(\hat{e}_n^i\right)^2 + 2\hat{e}_n^i E\left\{\tilde{f}_{n-1}^j\right\}\right. \\ &\quad \left. + E\left\{\left(\tilde{f}_{n-1}^j\right)^2\right\}\right) \\ &\quad + pE\left\{\left(\tilde{f}_{n-1}^i\right)^2\right\} \end{aligned} \tag{6}$$

where intercoded pixel $i$ is predicted from pixel $i + mv$ in the previous frame. The prediction error, $e_n^i$, is quantized to the value $\hat{e}_n^i$, which is conveyed together with the MVs to the decoder.

## III. RD Optimized Motion Compensation for Error Resilient Video Coding

### A. Motion Compensation for Error-Free Transmission

Motion compensation was generally studied in the context of error-free transmission, where the MV is selected so as to minimize a measure of the encoder prediction error, for example, MSE

$$\min_{mv} D_{\mathrm{DFD}} = \min_{mv} \sum_{i \in \mathrm{Blk}} \left(f_n^i - \hat{f}_{n-1}^{i+mv}\right)^2 \tag{7}$$

where $Blk$ denotes a block in the current frame. In H.264, the block size could be 16 × 16, 16 × 8, 8 × 16, 8 × 8, 8 × 4, 4 × 8, or 4 × 4 [40]. $D_{\mathrm{DFD}}$ is the (squared) displaced frame difference, and $mv$ is a particular MV candidate.

In the case of low bit-rate video coding, the MVs represent a significant portion of the total bit budget, leading to the proposal of rate constrained ME (RCME) [12]–[14]. Let $R_{mv}$ denote the MV coding rate. The problem is to minimize the Lagrangian

$$\min_{mv}\{J_{\mathrm{Blk}} = D_{\mathrm{DFD}} + \lambda R_{mv}\}. \tag{8}$$

Finally, the ultimate optimal ME solution (still assuming error-free transmission) is to minimize the overall source coding RD cost, as suggested in [15] and [16], where RD optimization was carried out independently for each block to avoid the complexity of joint optimization accounting for MV coding dependencies. It is also assumed that for each MB, the quantization step size is specified before ME, as is the case in the JM9.0 reference model of H.264. Hence, the source coding RD Lagrangian is

$$J_{\text{Blk}} = D + \lambda(R_{\text{res}} + R_{mv}) \tag{9}$$

where

$$D = \sum_{i \in \text{Blk}} \left(f_n^i - \hat{f}_n^i\right)^2 \tag{10}$$

and $R_{\text{res}}$ denotes the coding rate of the quantized prediction residue, which of course depends on $mv$. Note that we neglect the header bits, as they do not depend significantly on $mv$. Ignoring clipping effects, $D$ reduces to the quantization distortion $D_Q$

$$D_Q = \sum_{i \in \text{Blk}} \left(e_n^i - \hat{e}_n^i\right)^2. \tag{11}$$

### B. Proposed End-to-End RD Optimal Motion Compensation

We extend the above approach to the case of video transmission over lossy channels by replacing the source-coding distortion $D$ with the expectation $E\{D\}$, which accounts for the impact of packet loss

$$J_{\text{Blk}} = E\{D\} + \lambda(R_{\text{res}} + R_{mv}) \tag{12}$$

where

$$E\{D\} = \sum_{i \in \text{Blk}} E\left\{\left(f_n^i - \hat{f}_n^i + \hat{f}_n^i - \tilde{f}_n^i\right)^2\right\} \tag{13}$$

$$\simeq \sum_{i \in \text{Blk}} \left[\left(f_n^i - \hat{f}_n^i\right)^2 + E\left\{\left(\hat{f}_n^i - \tilde{f}_n^i\right)^2\right\}\right] \tag{14}$$

$$= D_Q + (1-p) \sum_{i \in \text{Blk}} E\left\{\left(\hat{f}_{n-1}^{i+mv} - \tilde{f}_{n-1}^{i+mv}\right)^2\right\} + p \sum_{i \in \text{Blk}} E\left\{\left(\hat{f}_n^i - \tilde{f}_{n-1}^i\right)^2\right\} \tag{15}$$

$$= D_Q + (1-p)D_{EP} + pD_{EC}. \tag{16}$$

The approximation in (14) assumes that decoder drift is zero-mean, and (15) is obtained by noting that both encoder and decoder reconstructions employ the same residual when the packet is received. We use intuitive names for the three distortion terms: the familiar quantization distortion $D_Q$, the error propagation distortion $D_{EP}$, and the error concealment distortion $D_{EC}$.

We observe that (12) does not account for error propagation to future frames, which compromises the ME optimality. To compensate for potential future error propagation, the proposed Lagrangian is redefined as

$$J_{\text{Blk}} = E\{D\} + \beta E\left\{\left(\hat{f}_n^i - \tilde{f}_n^i\right)^2\right\} + \lambda(R_{\text{res}} + R_{mv}) = E\{D\} + \beta(1-p)D_{EP} + \lambda(R_{\text{res}} + R_{mv}) \tag{17}$$

where $\beta$ is a positive constant. From (15), we see that $D_{EP}$ at frame $n$ can be expressed in terms of frame $n-1$ as $E\{(\hat{f}_{n-1}^{i+mv} - \tilde{f}_{n-1}^{i+mv})^2\}$. Likewise, the contribution of frame $n$ to future error propagation is captured by $E\{(\hat{f}_n^i - \tilde{f}_n^i)^2\}$. Hence, the weight $\beta$ is applied to this term only. Similar to the derivation of (16), this term can be expressed in terms of $D_{EP}$ and $D_{EC}$, and as $D_{EC}$ does not depend on $mv$, it is omitted leading to (17). As in [17], we treat $\beta$ as a free parameter whose value is exhaustively optimized and fixed per sequence. In [17], it was observed that $\beta$ could be determined adaptively per MB or even per pixel, see, e.g., [41]. We do not pursue this option here.

### C. Low Complexity Approximation

As presented, the above scheme assumes actual coding per MV candidate to calculate rate and distortion. To reduce complexity, a common practice is to e,ploy source coding RD models to efficiently predict $R_{\text{res}}$ and $D_Q$ in (17) [20]. Many RD models have been proposed, e.g., [42]–[46]. Here, we employ simple RD models, and derive an approximation to the optimal motion compensation scheme at very low computation complexity. The rationale is that by investigating the two extremes of the accuracy-complexity tradeoff, namely, exhaustive encoding versus simple quadratic RD modeling, we effectively bound the performance range. Clearly, any advanced and particularly more accurate RD models (e.g., see [20]) is expected to yield performance somewhere within the "benchmarked range" we provide.

A simple RD model is defined as follows:

$$R_{\text{res}} = \frac{K}{Q^2} D_{\text{DFD}} \tag{18}$$

$$D_Q = \frac{1}{12} Q^2. \tag{19}$$

Here, $Q$ denotes the quantization step size, and $K$ is a constant model parameter. The rate model of (18) is consistent with the simple intuition that a larger frame difference or smaller values of $Q$ generally yield a higher coding bit rate. This is in fact the basic observation underlying most existing source coding rate models [42]–[44]. Since we are concerned with intracoded blocks and the residual is approximately zero mean, then $D_{\text{DFD}}$ approximates the variance. The distortion model of (19) follows from the assumption the quantization error is uniformly distributed.

Combining (17) with (16) and then inserting (18) and (19), we obtain

$$J_{\text{Blk}} \simeq D_{\text{DFD}} + (1+\beta)(1-p)D_{EP} + \lambda R_{mv}. \tag{20}$$

where

$$\lambda = \frac{Q^2}{K}. \tag{21}$$

As shown in (16), $D_{EC}$ does not depend on $mv$ and has already been discarded. The optimal value of K has to be found experimentally. In [47], such experiments have been conducted to conclude that for the MSE metric, the choice $\lambda = 0.85 \cdot Q^2$ yields good coding performance, which we adopted here. Note that the Lagrangian of (20) is an error resilient extension of RCME of (8). An additional term of properly weighted error propagated distortion is introduced in (20), which captures the impact of packet loss and future error propagation.

## IV. RD Optimized Prediction for Error Resilient Video Coding

The conventional prediction mechanism uses past *encoder-reconstructed* frames for prediction. The typical predictor is given by

$$\bar{f}_n^i = \hat{f}_{n-1}^{i+mv} \tag{22}$$

where, $\bar{f}_n^i$ denotes the prediction for pixel $i$ of frame $n$. This predictive coding framework implicitly assumes lossless transmission, and is designed in "closed loop" to prevent encoder and decoder mismatch due to quantization loss [48]. However, in video networking applications, encoder and decoder prediction mismatch is generally unavoidable due to packet loss during transmission. This observation motivates us to revisit the problem of the optimal prediction scheme for this scenario.

Let us start with the expression for $E\{D\}$ in (16). Since we now focus on the impact of the prediction values, we omit explicit reference to $mv$ for notational simplicity and conciseness. Prediction only affects the error propagation term $D_{EP}$, which can be re-written in terms of a general predictor $\bar{f}_n^i$ as

$$\begin{aligned} D_{EP} &= \sum_{i\in \mathrm{Blk}} E\left\{\left(\bar{f}_n^i - \tilde{f}_{n-1}^{i+mv}\right)^2\right\} \\ &= \sum_{i\in \mathrm{Blk}} \left[\left(\bar{f}_n^i - E\left\{\tilde{f}_{n-1}^{i+mv}\right\}\right)^2 + \sigma^2\left(\tilde{f}_{n-1}^{i+mv}\right)\right] \\ &= \bar{D}_{EP} + \sigma_{EP}^2. \end{aligned} \tag{23}$$

Here, $\sigma^2$ denotes the variance of the designated decoder reconstruction. $\bar{D}_{EP}$ is the portion of $D_{EP}$ that is affected by the prediction $\bar{f}_n^i$. From (23), it is easy to see that the optimal prediction that minimizes the end-to-end distortion is determined by the *expected decoder reconstruction* in the previous frame. We call this prediction source-channel prediction (SCP), as it accounts for not only the source coding quantization loss, but also for loss in the channel. As will be shown by the simulation results, SCP proves to be a fairly good low complexity substitute for the more complicated overall RD optimal prediction. SCP is defined as

$$\bar{f}_{n,\mathrm{SCP}}^i = \arg\min_{\bar{f}_n^i} E\left\{d_n^i\right\} = E\left\{\tilde{f}_{n-1}^{i+mv}\right\}. \tag{24}$$

Next, we observe that the truly optimal prediction will be determined within the RD framework, as prediction affects not only the distortion, but also the rate cost for transmitting the prediction error (or residual), i.e., $R_{\mathrm{res}}$. In other words, in predictive coding, the efficiency of the prediction is ultimately measured by the RD cost incurred by the quantized prediction residual. The bottom line observation is: given a fixed *decoder prediction procedure*, the ultimate quantity that the encoder must optimize is *the value of quantized residue* to convey to the decoder. Therefore, in the sequel we adopt an overall quantization view of the problem and show that it leads to an effective optimization algorithm. Note that the search for the optimal quantized residual will be performed in the transform domain.

Employing the RDO formulation with $\beta$ as in (17), the optimal quantized values of the prediction residue are

$$\hat{\mathcal{E}}_{\mathrm{Blk}} = \left\{\hat{e}_n^i\right\}_{i\in \mathrm{Blk}} = \min_{\tilde{\mathcal{E}}_{\mathrm{Blk}}} \left[E\{D\} + \beta p D_{EC} + \lambda R_{\mathrm{res}}\right] \tag{25}$$

where $Blk$ refers to the basic transform coding block; In H.263 or MPEG4 it is an $8\times 8$ block [23], [49], but in H.264 the block size is $4\times 4$ [40]. Note that as the resulting $R_{\mathrm{res}}$ is determined by the entire block of quantized prediction residue (due to the run-length coding for the quantized coefficients), quantizations for all the transform coefficients should be *jointly* optimized as well and we hence consider the *vector of all the quantized coefficients*, denoted by $\hat{\mathcal{E}}_{\mathrm{Blk}}$. Similarly as in (17), we apply $\beta$ on $E\{(\hat{f}_n^i - \tilde{f}_n^i)^2\}$ to compensate for future error propagation impacts of frame $n$. Herein, optimizing quantization will only affect the $D_{EC}$ term, but not the $D_{EP}$ term. (Note that while error concealment does not depend on the quantized residual, the concealment distortion does). Therefore, $\beta(1-p)D_{EP}$ is omitted in (25).

Moreover, instead of trellis-based optimal search [26], we propose a novel hybrid search algorithm, which effectively combines exhaustive search (where computationally cost-effective) and heuristic search, and achieves close to optimal performance at reduced computation complexity.

The proposed hybrid search algorithm is as follows.

- Step 1. The initial quantization levels: $X_{\mathrm{SCP}} = \{x_i\}_{i=1}^{16} = Zigzag(Quant(DCT(\{e_{i,SCP}\}_{i\in \mathrm{Blk}})))$. Herein, $\{e_{i,SCP}\}_{i\in \mathrm{Blk}}$ is the resultant prediction residue from SCP.
  The total number of level combinations: $A = \prod_{i=1}^{16}(|x_i|+1)$.
- Step 2. If $A \le N$, exhaustively search over all the possible combinations to find the optimal solution, and then, stop. Otherwise, continue to Step 3.
- Step 3. Search over heuristically selected combination candidates. Find the optimum, and then, stop. The searching candidates are selected as follows.
  - The first candidate is $X_{\mathrm{SCP}}$. Then, sort all the nonzero elements of $X_{\mathrm{SCP}}$ in ascending order, and get $\{\mathrm{index}(i)\}_{i=1}^{N'}$. Herein, $x_{\mathrm{index}(1)} \le x_{\mathrm{index}(2)} \le \cdots \le x_{\mathrm{index}(N')}$, and $N'$ denotes the number of nonzero elements of $X_{\mathrm{SCP}}$.
  - Initialize $X'$: $X' = \{x_i'\}_{i=1}^{N'} = \{x_{\mathrm{index}(i)}\}_{i=1}^{N'}$.
  - Decrease the magnitude of each of the first $M$ consecutive elements in $X'$ by 1, and get a new $X'$. Here, the constant $M = \max\{1, \mathrm{Round}(\sum_{i=1}^{16}|x_i|/N)\}$. Then, update $X$ with $X'$ via $x_{\mathrm{index}(i)} = x_i'$ $(i = 1,\ldots,N')$. The new $X$ is taken as a new candidate.
  - Repeat the above step for the next $M$ consecutive elements in $X'$, get another new candidate, and so on. The selection of $M$ consecutive elements of $X'$ is conducted in a round-robin fashion, i.e., whenever it reaches the end of $X'$, the selection resumes from the beginning. Also, whenever $x_i'$ is reduced to zero, it is removed from $X'$, and $N' = N' - 1$.
  - This candidate searching process ends when $X$ is all-zero.

Basically, $X_{\mathrm{SCP}}$ and the "all-zero" $X$ represent two extreme RD points (i.e., minimizing $E\{D\}$ and $R_{\mathrm{res}}$ respectively), and our proposed algorithm actually sweeps over all the selected intermediate RD points in between. This explains why the total number of level combinations is $\prod_{i=1}^{16}(|x_i|+1)$. Here, $N$ is a threshold to switch between exhaustive search and heuristic

search. In experiments, we set $N = 10$ and found that in this case over 80% of total searching operations will be exhaustive search. Despite the fact that less than 20% of the blocks employ heuristic search, the computational savings are tremendous. As an illustrative example, let a block be represented by a coefficient magnitude vector of $\{7, 4, 2, 2, 3, 1, 3, 1, 2, 1, 0, \ldots, 0\}$, the sum of magnitudes is 26 (greater than $N = 10$). The computational savings due to using heuristic rather than exhaustive search here, with 10 versus 138240 operations, yielding a ratio of $4 \times 10^{-9}$. On the other hand, since optimal exhaustive search is conducted for more than 80% of the blocks, where it is not very costly, the resulting overall performance may still closely approach that of the optimal trellis search.

## V. Reference Frame Generation

So far, we derived MC and prediction schemes that optimize the MCP decisions at the encoder, given a standard MCP procedure employed by the decoder. Now, we expand the scope and consider re-design of the entire MCP mechanism at both encoder and decoder, so as to improve the overall system performance. Particularly, we focus on the RFG module of Fig. 1. As explained in Section I-C, the RFG-MCP decoupling results in "MC-free" RFG, which allows us to consider a broad range of filter paradigms and identify a better RFG architecture. Specifically, besides the conventional complete prediction, we also consider leaky prediction (i.e., partial prediction), FIR-based weighted prediction, and a proposed novel IIR-based prediction called GSCP.

The conventional practice of directly using reconstructed past frames as reference frames for predictive coding, may result in substantial error propagation due to packet loss. One strategy to reduce error propagation is to employ leaky prediction, which scales down reconstructed frames to generate reference frames that yields exponential decay of propagated errors

$$\check{f}_n^i = \alpha \cdot \hat{f}_n^i + (1 - \alpha) \cdot C \tag{26}$$

where $\check{f}_n^i$ is a reference frame pixel, $\hat{f}_n^i$ is the reconstructed pixel, $\alpha$ is the leak factor, and $C$ is an appropriate constant. Leaky prediction, or leaky integration, has a long history in general signal compression, as well as in video coding in particular. It has been most widely used for enhancement layer drift control in layered video coding [28]–[30], where a no drift (i.e., no error propagated) base-layer reconstruction of a frame is available, can be used as substitute for $C$ in (26) for improved coding efficiency. However, in the case of single layer coding, such an option is not available and one must default to constant $C$, whose value is typically the mid-range signal level of 128 as in [50], and at significant impact to coding efficiency.

An alternative prediction approach that offers error resilience advantages is weighted prediction, which is already part of the H.264/AVC standard [51]. In practice, weighted prediction is usually applied along with MC, where two versions of motion compensated predictions from two individual past coded frames are weighted and combined together to predict the current frame [27], [31], [32]. In contrast, in our RFG formulation, *no MC* is involved. Applying weighted prediction in this case, we get

$$\check{f}_n^i = \alpha \cdot \hat{f}_n^i + (1 - \alpha) \cdot \hat{f}_{n-1}^i. \tag{27}$$

In this paper, we propose a new IIR-based RFG scheme, called generalized source-channel prediction (GSCP) defined as

$$\check{f}_n^i = \alpha \cdot \hat{f}_n^i + (1 - \alpha) \cdot \check{f}_{n-1}^i. \tag{28}$$

Comparing (26), (27), and (28), it is easy to see that the only difference lies in their second term, weighted by $(1 - \alpha)$, for which $C$, $\hat{f}_{n-1}^i$, and $\check{f}_{n-1}^i$ are used, respectively. We emphasize that this difference is fundamental and impacts performance in terms of coding efficiency and error control.

First, considering leaky prediction, although it yields the fastest exponential error decay among the three, it also causes serious coding efficiency degradation. On the other hand, either weighted prediction or GSCP yields much better coding efficiency than leaky prediction, while still achieving effective error propagation control. (Discussion of error control effectiveness of weighted prediction can be found in [27].) As will be shown in Section VI-C, when compared with weighted prediction and GSCP, its coding efficiency loss, more often than not outweighs its error control gain, which, thus, leads to the worst overall performance among the three. Sometimes its performance may be even worse than that of conventional prediction.

Considering weighted prediction and GSCP, it is easy to see from (27) and (28) that, in essence, they represent FIR filtering and IIR filtering, respectively. Given the same $\alpha$, but with $\check{f}_{n-1}^i$ involved, GSCP implies stronger filtering than weighted prediction. Intuitively, reference frames generated with heavier filtering will be more robust to error propagation, but less correlated with the original frame, which generally impacts prediction and hence coding efficiency. Note that unlike their performance gains over leaky prediction, the performance comparison between weighted prediction and GSCP is much more subtle. As will be shown in the simulation results, weighted prediction achieves the best overall coding performance whenever highly efficient ED-RDO-based intra-update is employed, while GSCP outperforms all methods in conjunction with the "standard" random intra-update.

We note that (28) can be viewed as a generalization from the SCP scheme of (24) which subsumes the ROPE update of (5), and hence, the name of GSCP. In SCP, one uses $\alpha = (1 - p)$ (where $p$ is packet loss rate) so that the prediction becomes the expected reconstructed frame at the decoder. Note further that SCP employs this modified prediction only at the encoder. GSCP, on the other hand, offers a more flexible weighting of the two terms and modifies both the encoder and decoder. To investigate performance bounds, similarly as for $\beta$ in ED-RDO for ME and prediction, $\alpha$ is treated as a free parameter, optimized and fixed for each sequence.

## VI. Simulation Results

Our simulation setting builds on the JM9.0 H.264 codec [52]. We used constrained intraprediction and CAVLC for entropy coding. We adopted rate control from the JM codec and set one common quantization scale to all the MBs of one row. For each sequence, only the first frame was coded as I-frame, and the rest were coded as P-frames. To simulate the channel, at each packet

TABLE I
PERFORMANCE OF OPTIMIZED MOTION COMPENSATION WITH OPTIMAL INTRA-UPDATING. CARPHONE, MOBILE: QCIF, 10 f/s, 200 kb/s. FOREMAN, TEMPETE: CIF, 30 f/s, 800 kb/s. WHEN VARYING PACKET LOSS RATE, NUMBER OF REFERENCE FRAMES IS 3. WHEN VARYING NUMBER OF REFERENCE FRAMES, PACKET LOSS RATE IS 5%

| PSNR or ΔPSNR (dB) | | Packet Loss Rate (%) | | | | Number of Reference Frames | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 5 | 10 | 1 | 3 | 6 | 10 |
| Carphone | Conv. | 39.22 | 38.12 | 36.07 | 33.95 | 35.98 | 36.07 | 36.08 | 36.08 |
| | EOMC | +0.26 | +0.20 | +0.11 | +0.06 | +0.01 | +0.11 | +0.20 | +0.25 |
| | LCMC | +0.12 | +0.13 | +0.06 | +0.06 | -0.01 | +0.06 | +0.12 | +0.12 |
| Mobile | Conv. | 29.80 | 29.06 | 27.69 | 26.29 | 27.26 | 27.69 | 27.85 | 27.87 |
| | EOMC | +0.17 | +0.16 | +0.16 | +0.13 | -0.01 | +0.16 | +0.28 | +0.46 |
| | LCMC | +0.08 | +0.13 | +0.13 | +0.11 | +0.00 | +0.13 | +0.26 | +0.37 |
| Foreman | Conv. | 35.45 | 34.76 | 33.47 | 32.05 | 33.13 | 33.47 | 33.69 | 33.86 |
| | EOMC | +0.12 | +0.11 | +0.13 | +0.18 | +0.05 | +0.13 | +0.25 | +0.29 |
| | LCMC | +0.08 | +0.10 | +0.12 | +0.16 | +0.04 | +0.12 | +0.23 | +0.26 |
| Tempete | Conv. | 28.27 | 27.98 | 27.31 | 26.48 | 26.35 | 27.31 | 27.94 | 28.27 |
| | EOMC | +0.09 | +0.09 | +0.12 | +0.21 | +0.03 | +0.12 | +0.23 | +0.27 |
| | LCMC | +0.05 | +0.04 | +0.11 | +0.10 | +0.00 | +0.11 | +0.18 | +0.19 |

loss rate, 300 packet loss patterns were randomly generated, and the average luminance PSNR at the decoder was computed to measure the system performance. Only the first 100 frames of each testing sequence are used for encoding. As discussed in Section III-B, this is to constrain the possible optimality loss due to sequence-wise free parameter optimization for $\beta$ in ED-RDO ME and prediction, and $\alpha$ in GSCP. The best $\beta$ value is exhaustively selected from $\{0, 0.5, 1.0, 1.5, 2.0\}$, the $\alpha$ value is selected from $\{0.9, 0.8, 0.6, 0.4\}$ for weighted prediction and GSCP, and from $\{0.95, 0.9, 0.8\}$ for leaky prediction. In practice, we found that $\alpha$ values below the above ranges lead to perceptually annoying coding artifacts.

The methods were tested under two extreme intra-updating scenarios: random intra-updating and optimal intra-updating. In random intra-updating, given packet loss rate $p$, a fraction $p$ of MBs in each frame are selected for intracoding. (The intra-MBs are selected according to the implementation in the JM9.0 encoder.) In optimal intra-updating, the coding mode is optimally selected per MB from all the available coding mode options, via the ED-RDO framework. The Lagrange multiplier is handled as in the JM codec implementation, and the distortion is estimated by ROPE of [4].

### A. Optimized Motion Vector Selection

We tested the proposed ED-RDO motion compensation schemes with RD values that are either exhaustively calculated via actual encoding (denoted "EOMC" for "exhaustively optimized motion compensation"), or calculated at low complexity via simple RD modeling (denoted as "LCMC"). We compare their performance with that of the conventional RCME (denoted as "Conv."), which is the scheme adopted by the JM9 reference codec.

The results for the cases of random and optimal intra-updating are shown in Fig. 2 and Table I, respectively. From Fig. 2, it is easy to see that in the case of random intra-updating, both EOMC and LCMC yield significant performance gains over the conventional scheme, e.g., up to 5.08 and 2.95 dB, respectively. This shows the effectiveness of the proposed schemes. Another observation is that a large performance gap may still exist between EOMC and LCMC, due to the simplicity of the assumed RD models. On the other hand, Table I shows that the performance gains are substantially reduced when optimal intra-updating is used. This result mainly speaks to the effectiveness of ED-RDO optimal intra-updating, which considerably mitigates the error propagation effect, and leaves a largely reduced scope for further performance enhancement. In this case, up to 0.46 and 0.37 dB gain can still be achieved by EOMC and LCMC, respectively, when a large number of reference frames (e.g., 10) is employed. Employing more reference frames for MCP tends to increase the gains due to the proposed optimization schemes. Also, note that in this case, LCMC closely approaches the performance of EOMC (with a performance gap less than 0.1 dB).

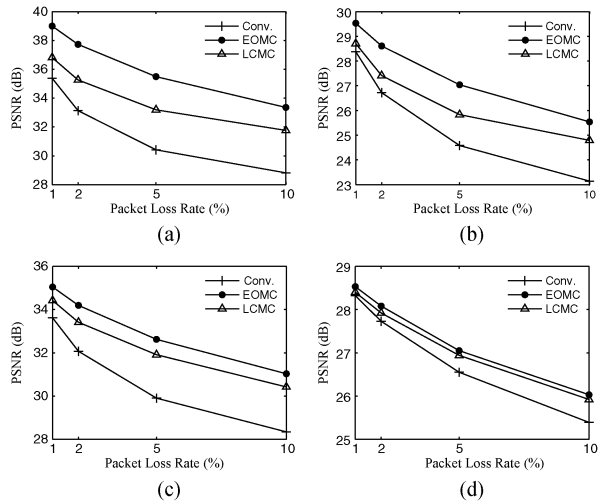



Fig. 2. Performance of optimized motion compensation with random intra-updating. Three reference frames. (a), (b) QCIF, 10 f/s, 200 kb/s. (c), (d) CIF, 30 f/s, 800 kb/s. (a) Carphone; (b) Mobile; (c) Foreman; (d) Tempete.

### B. Optimized Prediction

We then move on to examine the performance of optimized prediction. Note that RCME is always used here for MV selection, and competing schemes differ only in the way they determine the prediction values. The conventional encoder reconstruction-based prediction solution is denoted as "Conv.", while our proposed RD optimal prediction and SCP schemes are denoted as "OPred" and "SCP", respectively.

Fig. 3 and Table II show the result with random and optimal intra-updating, respectively. From Fig. 3, it is clearly seen that either "OPred" or "SCP" may achieve significant performance

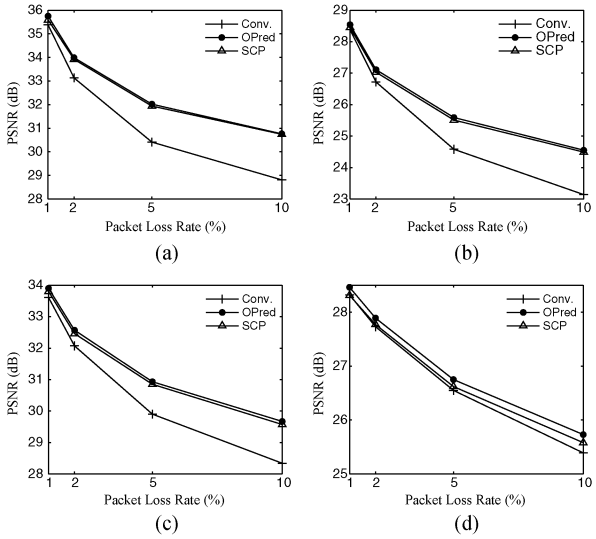



Fig. 3. Performance of optimized prediction with random intra-updating. Three reference frames. (a), (b) QCIF, 10 f/s, 200 kb/s. (c), (d) CIF, 30 f/s, 800 kb/s. (a) Carphone; (b) Mobile; (c) Foreman; (d) Tempete.

TABLE II
PERFORMANCE OF OPTIMIZED PREDICTION WITH OPTIMAL INTRA-UPDATING. CARPHONE, MOBILE: QCIF, 10 f/s, 200 kb/s. FOREMAN, TEMPETE: CIF, 30 f/s, 800 kb/s. WHEN VARYING PACKET LOSS RATE, NUMBER OF REFERENCE FRAMES IS 3

| PSNR or ΔPSNR (dB) | | Packet Loss Rate (%) | | | |
|---|---|---|---|---|---|
| | | 1 | 2 | 5 | 10 |
| Carphone | Conv. | 39.22 | 38.12 | 36.07 | 33.95 |
| | OPred | +0.12 | +0.09 | +0.01 | +0.00 |
| | SCP | +0.11 | +0.08 | -0.01 | -0.02 |
| Mobile | Conv. | 29.80 | 29.06 | 27.69 | 26.29 |
| | OPred | +0.17 | +0.11 | +0.10 | +0.03 |
| | SCP | +0.11 | +0.07 | +0.03 | -0.01 |
| Foreman | Conv. | 35.45 | 34.76 | 33.47 | 32.05 |
| | OPred | +0.30 | +0.14 | +0.03 | -0.01 |
| | SCP | +0.24 | +0.12 | -0.01 | -0.07 |
| Tempete | Conv. | 28.27 | 27.98 | 27.31 | 26.48 |
| | OPred | +0.47 | +0.35 | +0.12 | -0.06 |
| | SCP | +0.36 | +0.26 | +0.04 | -0.20 |

gain over "Conv." (i.e., up to 1.95 and 1.93 dB, respectively). Furthermore, the low complexity SCP scheme achieves performance that is quite similar to that of the more complex optimal prediction (with a performance gap less than 0.16 dB). In fact, it is observed in experiments that over 80% prediction optimization operations end up with SCP as their final optimal choice. This result shows that SCP can be used as a fairly good substitute for the overall RD optimal prediction scheme in practice. On the other hand, similarly as in the case of optimized motion compensation, in conjunction with optimal intra-updating, the gain from "OPred" and "SCP" also greatly decreases, as shown in Table II. In spite of that, up to 0.47 and 0.36 dB gain can still be achieved by "OPred" and "SCP", respectively, when the packet loss rate is low (e.g., less than 2%).

### C. Reference Frame Generation

We compare the performance of the conventional complete prediction ("Conv.") with our proposed GSCP ("GSCP"), weighted prediction ("Weighted") and leaky prediction ("Leaky"). The optimization of $\alpha$ is as specified in the beginning of this section. Note that ROPE is modified to accommodate nonconventional RFG architectures. We assume that when the current frame data is lost, $\hat{f}_n^i$ in (26), (27), and (28) will be simply concealed by $\breve{f}_{n-1}^i$. Also, cross correlation terms will be estimated using the linear signal model outlined in [37].

Let us first look at results obtained in conjunction with random intra-updating, Figs. 4 and 5. None of the RFG architectures consistently outperforms the others. For example, leaky prediction outperforms conventional complete prediction most of the time (e.g., for Carphone, Mobile, and Foreman), but not always (e.g., for Tempete). Similar observations are made for GSCP and weighted prediction. The above observations reinforce the understanding that these approaches represent

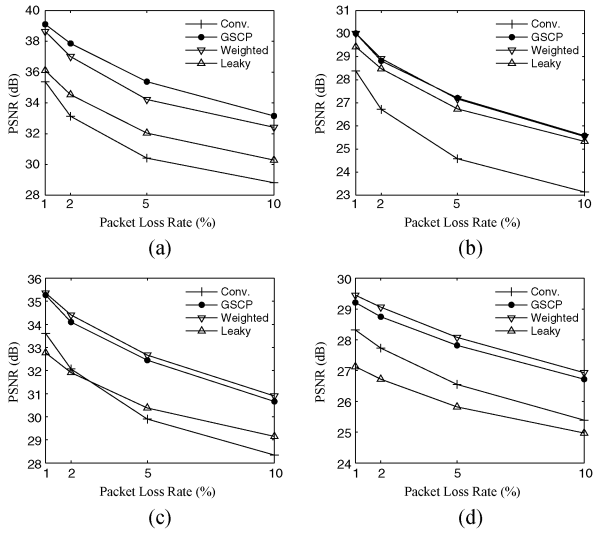



Fig. 4. PSNR versus packet loss rate performance of various reference frame generation schemes with random intra-updating. Three reference frames. (a), (b) QCIF, 10 f/s, 200 kb/s. (c), (d) CIF, 30 f/s, 800 kb/s. (a) Carphone; (b) Mobile; (c) Foreman; (d) Tempete.

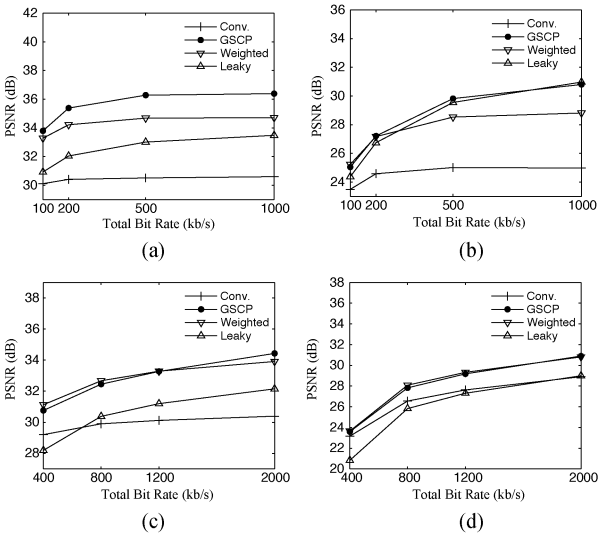



Fig. 5. PSNR versus total bit rate performance of various reference frame generation schemes with random intra-updating. Packet loss rate: 5%. 3 reference frames. (a), (b) QCIF, 10 f/s. (c), (d) CIF, 30 f/s. (a) Carphone; (b) Mobile; (c) Foreman; (d) Tempete.

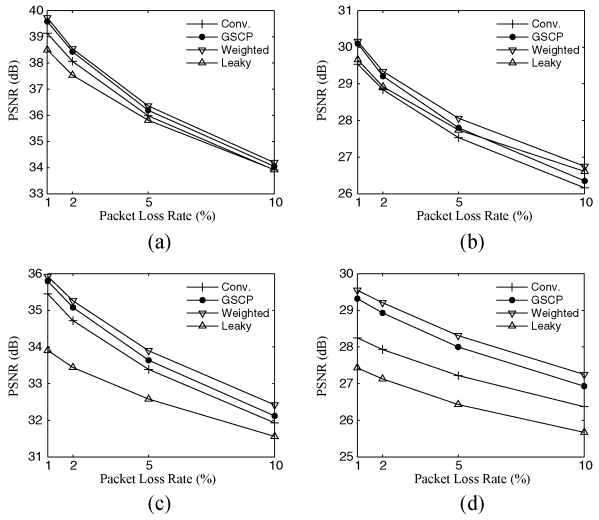



Fig. 6. Performance of various reference frame generation schemes with optimal intra-updating. Three reference frames. (a), (b) QCIF, 10 f/s, 200 kb/s. (c), (d) CIF, 30 f/s, 800 kb/s. (a) Carphone; (b) Mobile; (c) Foreman; (d) Tempete.

different levels of tradeoff between error control and coding efficiency. Hence, a different scheme may be the performance leader depending on the circumstances. Nevertheless, GSCP emerges as the overall better scheme in the case of random intra-updating. GSCP and weighted prediction always significantly outperform conventional prediction, while leaky prediction may perform significantly worse than the conventional scheme (e.g., for Foreman and Tempete). Even in the extreme case of 1 Mb/s in Fig. 5(b), its gain over GSCP is only 0.17 dB. We found that GSCP significantly outperforms weighted prediction (with up to 1.99 dB gain) for high motion video (e.g., for Carphone and News) or high bit rate coding [e.g., for 500 kb/s and 1 Mb/s in Fig. 4(b) and 2 Mb/s in Fig. 4(c)]. Otherwise, weighted prediction may only outperform GSCP with marginal gains (no more than 0.33 dB). This is mainly because in the case of random intra-updating, error control dominates coding efficiency in impact on the overall performance. As analyzed in Section V, in general, GSCP offers better error control but somewhat compromised coding efficiency relative to weighted prediction. In the case of high motion video or high bit rate coding, the error control advantage of GSCP outweighs its coding efficiency disadvantage, thereby leading to significantly improved overall performance. Therefore, GSCP is the best RFG scheme in the case of random intra-updating. Up to 6.60 dB gain can be achieved over the conventional scheme.

Fig. 6 shows the performance with optimal intra-updating. It is easy to see that unlike the results with random intra-updating, when applied together with optimal intra-updating, weighted prediction always performs best among all the candidate schemes. Its gain over the conventional scheme may reach 1.3 dB. Due to the effective error resilience clearly achieved by optimal intra-updating, the error control benefits of GSCP are not as important as in the case of random intra-updating. Overall, weighted prediction renders a better coding efficiency and error control tradeoff in this case than GSCP (with gains up to 0.40 dB).

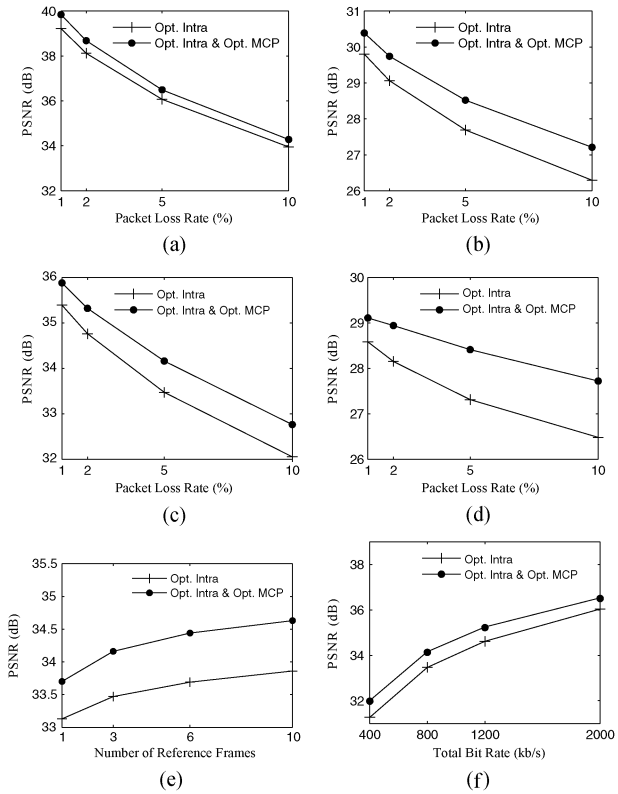



Fig. 7. Performance of the overall improved video codec with optimal intra-updating. (a), (b) QCIF, 10 f/s, 200 kb/s, three reference frames. (c)–(f) CIF, 30 f/s. (c), (d) 800 kb/s, three reference frames. (e) 800 kb/s, packet loss rate: 5%. (f) Three reference frames, packet loss rate: 5%. (a) Carphone; (b) Mobile; (c) Foreman; (d) Tempete; (e) Foreman; (f) Foreman.

### D. Overall Improved Video Codec

In fact, the proposed various encoder optimizations are complementary, and hence can be combined to maximize the overall performance improvement. Note that with the accurate distortion estimate available, optimal intra-updating will naturally be adopted in the video coding and streaming system. Therefore, in this subsection, we investigate the performance of the video codec resulting from the combination of the proposed various encoder optimizations in conjunction with optimal intra-updating. The conventional video codec coupled with ED-RDO optimal intra-updating is denoted as "Opt. Intra", while the overall improved video codec incorporating the proposed optimizations of motion compensation, source-channel prediction and reference frame generation (in this case, weighted prediction) is denoted as "Opt. Intra & Opt. MCP". The overall improved video codec represents a comprehensive solution for optimizing and modifying MCP for error resilience, while maintaining low complexity.

Fig. 7 summarizes the performance results. It is obvious that the improved codec with optimized MCP and intra-updating significantly outperforms the conventional codec coupled with optimal intra-updating in all the testing cases, i.e., for all the tested packet loss rates, numbers of reference frames, and coding bit rates with up to 1.24 dB performance gain achieved. This result substantiates the effectiveness of the proposed encoder optimizations.

Note that the overall improved codec involves both $\alpha$ and $\beta$ optimizations. In simulation, we tested whether a more involved two-pass search to optimize $\alpha$ and $\beta$ provides benefits. We found that a single pass produced results of similar overall coding performance (the performance gap always less than 0.1 dB). This suggests that one may simply optimize $\alpha$ and $\beta$ independently at considerably lower complexity, and at negligible sacrifice in performance relative to joint optimization.

### E. Complexity Issues

In order to characterize the best achievable performance of the proposed optimal motion compensation and prediction schemes, all RD data used in ED-RDO are calculated after performing actual encoding operations, despite the prohibitive complexity. The complexity is substantially reduced by the proposed low complexity variants, LCMC and SCP. In comparison with conventional RCME, we note that LCMC only involves an additional distortion term that accounts for error propagation, as is explicit in (20). Given the availability of first and second moments that are already calculated by ROPE, the computation of the additional term only incurs complexity similar to that of the squared difference calculation in RCME. The other low-complexity scheme, SCP, exploits the first moment provided by ROPE, and introduces no further computation complexity relative to conventional prediction.

RFG with the proposed GSCP involves separate construction of prediction reference frames. We note that a conventional codec needs to record the reconstruction of each frame so it can serve as prediction reference for the following frames. Compare this with the GSCP codec where, rather than directly store the reconstruction as reference, we first perform the additional calculation specified in (28) per pixel, to obtain the GSCP reference. This implies that no additional storage cost is incurred by GSCP. The additional computation of (28) consists of two multiplications and one addition per pixel. This clearly translates into very marginal computational complexity cost, relative to the computationally significant modules of a conventional coder such as ME and transformation. Similar analysis also holds true for weighted prediction-based RFG.

Experimental evaluation of the computational complexity of the overall improved ROPE method, has been performed on a computer with Pentium IV 3.0-GHz CPU and 504-MB RAM. It shows overall increase in total encoding time (relative to a conventional coder) by about 130%$\sim$160%. For a more detailed report see [37]. In terms of storage/memory consumption, for each pixel, the standard coded integer pixel costs 1 byte, while ROPE needs additional $4 \times 2 = 8$ bytes to store the first and second moments in floating point representation. Given the current trend in processing and storage capabilities and cost, the complexity required by ROPE is modest and poses no serious problems in practice.

## VII. Conclusion

This paper is concerned with the optimization and modification of the fundamental MCP framework to improve the error resilience of video coding. While most of the existing end-to-end RD-based error resilient video coding techniques focus on intra/intramode selection, a novelty of this work lies in its particular focus on the potential of optimizing and modifying the basic underlying MCP framework for error resilience, and thereby enhancing the overall RD performance from the *intercoded* MBs.

First, we extend the existing ED-RDO framework to optimize ME and prediction. The proposed schemes also account for the impact of future error propagation for enhanced optimality. A novel hybrid search algorithm is developed to solve the optimal prediction problem in the context of H.264. Low complexity practical variants are also proposed for optimal ME and prediction, respectively. Second, we investigate RFG design optimization in terms of error resilience and coding efficiency tradeoffs. For this purpose, we consider a broad spectrum of filter paradigms, including not only leaky prediction, FIR-based weighted prediction, but also the newly proposed IIR-based GSCP scheme. Note that the introduction of the IIR filter-based RFG design would not be possible without decoupling RFG from MCP. This decoupling formulation opens the door to further error resilience improvement. While weighted prediction can be applied with motion compensation in the "conventional" way, both weighted prediction and GSCP can be applied in conjunction with RFG to enhance error resilience and thereby the overall system performance. We conducted experiments to investigate the achievable performance gains due to the proposed encoder optimizations. The significant performance gains observed in simulations justify their effectiveness.

## References


[1] R. Puri, K. Ramchandran, and V. Bhaarghavan, "An integrated source transcoding and congestion control paradigm for video streaming in the Internet," *IEEE Trans. Multimedia*, vol. 3, no. 1, pp. 18–32, Jan. 2001.

[2] P. A. Chou and Z. Miao, "Rate-distortion optimized streaming of packetized media," *IEEE Trans. Multimedia*, 2001, submitted for publication.

[3] R. Zhang, S. L. Regunathan, and K. Rose, "End-to-end distortion estimation for RD-based robust delivery of pre-compressed video," in *Proc. 35th Asilomar Conf.*, 2001, vol. 1, pp. 210–214.

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

[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. J. Liang, E. Setton, and B. Girod, "Channel-adaptive video streaming using packet path diversity and rate-distortion optimized reference picture selection," in *Proc. IEEE 5th Workshop on Multimedia Signal Processing*, Dec. 2002, pp. 420–423.

[7] 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.

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

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

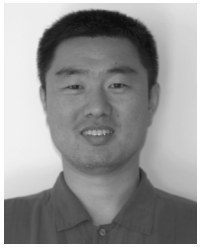
[10] A. R. Reibman, "Optimizing multiple description video coders in a packet loss environment," presented at the Packet Video Workshop, 2002.

[11] A. Reibman, H. Jafarkhani, Y. Wang, and M. Orchard, "Multiple description video using rate-distortion splitting," *Proc. IEEE Int. Conf. Image Processing*, vol. 1, pp. 978–981, 2001.

[12] G. J. Sullivan and T. Wiegand, "Rate-distortion optimization for video compression," *IEEE Signal Process. Mag.*, vol. 15, no. 6, pp. 74–90, Jun. 1998.

[13] M. C. Chen and A. N. Willson, "Rate-distortion optimal motion estimation algorithms for motion-compensated transform video coding," *IEEE Trans. Circuits Syst. Video Tech.*, vol. 8, no. 2, pp. 147–158, Apr. 1998.

[14] T. Wiegand, X. Zhang, and B. Girod, "Long-term memory motion-compensated prediction," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, no. 2, pp. 70–84, Feb. 1999.
[15] W. C. Chung, F. Kossentini, and M. J. T. Smith, "An efficient motion estimation technique based on a rate-distortion criterion," in *Proc. IEEE ICASSP*, May 1996, vol. 4, pp. 1926–1929.
[16] G. M. Schuster and A. K. Katsaggeslos, "A theory for the optimal bit allocation between displacement vector field and displaced frame difference," *IEEE J. Sel. Areas Commun.*, vol. 15, no. 12, pp. 1739–1751, Dec. 1997.
[17] 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.
[18] M. Budagavi and J. D. Gibson, "Multiframe video coding for improved performance over wireless channels," *IEEE Trans. Image Process.*, vol. 10, no. 2, pp. 252–265, Feb. 2001.
[19] O. Harmanci and A. M. Tekalp, "A stochastic framework for rate-distortion optimized video coding over error-prone networks," *IEEE Trans. Image Process.*, vol. 16, no. 3, pp. 684–697, Mar. 2007.
[20] S. Wan and E. Izquierdo, "Rate-distortion optimized motion-compensated prediction for packet loss resilient video coding," *IEEE Trans. Image Process.*, vol. 16, no. 3, pp. 1327–1338, May 2007.
[21] H. Yang and K. Rose, "Rate-distortion optimized motion estimation for error resilient video coding," in *Proc. IEEE ICASSP*, Philadelphia, PA, Mar. 2005, vol. 2, pp. 173–176.
[22] S. Wenger, "Video redundancy coding in H.263+," presented at the AVSPN, Aberdeen, U.K., Sep. 1997.
[23] ITU-T Recommendation H.263 Version 2 (H.263+), Video Coding for Low Bitrate Communications, Jan. 1998.
[24] H. Yang and K. Rose, "Source-channel prediction in error resilient video coding," in *Proc. IEEE ICME*, Baltimore, MD, Jul. 2003, vol. 2, pp. 233–236.
[25] O. Harmanci and A. M. Tekalp, "Stochastic frame buffers for rate distortion optimized loss resilient video communications," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2005, vol. 1, pp. 789–792.
[26] J. Wen, M. Luttrell, and J. Villasenor, "Trellis-based R-D optimal quantization in H.263+," *IEEE Trans. Image Process.*, vol. 9, no. 8, pp. 1431–1434, Aug. 2000.
[27] W.-Y. Kung, C.-S. Kim, and C.-C. J. Kuo, "Analysis of multihypothesis motion compensated prediction (MHMCP) for robust visual communication," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 1, pp. 146–153, Jan. 2006.
[28] H. C. Huang, C. N. Wang, and T. Chiang, "A robust fine granularity scalability using trellis-based predictive leak," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 6, pp. 372–385, Jun. 2002.
[29] S. Han and B. Girod, "Robust and efficient scalable video coding with leaky prediction," in *Proc. IEEE Int. Conf. Image Process.*, Rochester, NY, Sep. 2002, vol. 2, pp. 41–44.
[30] W.-H. Peng and Y.-K. Chen, "Error drifting reduction in enhanced fine granularity scalability," in *Proc. IEEE Int. Conf. Image Processing*, Rochester, NY, Sep. 2002, vol. 2, pp. 61–64.
[31] C.-S. Kim, R.-C. Kim, and S.-U. Lee, "Robust transmission of video sequence using double-vector motion compensation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 9, pp. 1011–1021, Sep. 2001.
[32] Y. Wang and S. Lin, "Error-resilient video coding using multiple description motion compensation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 6, pp. 438–452, Jun. 2002.
[33] 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.
[34] B. A. Heng, J. G. Apostolopoulos, and J. S. Lim, "End-to-end rate-distortion optimized mode selection for multiple description video coding," presented at the ICASSP 2005, 2005.
[35] 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*, 2005, to be published.
[36] E. Masala, H. Yang, K. Rose, and J. C. De Martin, "Rate-distortion optimized slicing, packetization and coding for error-resilient video transmission," in *Proc. IEEE DCC*, 2004, pp. 182–191.
[37] H. Yang and K. Rose, "Advances in recursive per-pixel end-to-end distortion estimation for robust video coding in H.264/AVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 7, pp. 845–856, Jul. 2007.
[38] 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.
[39] H. Yang and K. Rose, "Mismatch impact on per-pixel end-to-end distortion estimation and coding mode selection," in *Proc. IEEE ICME*, Jul. 2007, pp. 2178–2181.
[40] JVT of ISO/IEC MPEG and ITU-T VCEG, ITU-T Rec. H.264, Advanced Video Coding for Generic Audiovisual Services Nov. 2007.
[41] R. Zhang, S. L. Regunathan, and K. Rose, "Prescient mode selection for robust video coding," in *Proc. IEEE Int. Conf. Image Processing*, Oct. 2001, vol. 1, pp. 974–977.
[42] H.-M. Hang and J.-J. Chen, "Source model for transform video coder and its application—Part I: Fundamental theory," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, no. 2, pp. 287–297, Apr. 1997.
[43] J. R. Corbera and S. Lei, "Rate control in DCT video coding for low-delay communications," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, no. 1, pp. 172–185, Feb. 1999.
[44] K. H. Yang, A. Jacquin, and N. S. Jayant, "A normalized rate-distortion model for H.263-compatible codecs and its application to quantizer selection," in *Proc. Int. Conf. Image Processing*, Santa Barbara, CA, Oct. 1997, vol. II, pp. 41–44.
[45] L.-J. Lin and A. Ortega, "Bit-rate control using piecewise approximated rate-distortion characteristics," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 38, no. 1, pp. 82–93, Jan. 1990.
[46] Z. He and S. K. Mitra, "A unified rate-distortion analysis framework for transform coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 12, pp. 1221–1236, Dec. 2001.
[47] T. Wiegand and B. Girod, "Lagrangian multiplier selection in hybrid video coder control," presented at the Int. Conf. Image Processing, Thessaloniki, Greece, Oct. 2001.
[48] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*. Norwell, MA: Kluwer, 2001, pp. 206–214.
[49] Video Group, Text of ISO/IEC 14496-2 MPEG4 video VM-Version 8.0 ISO/IEC JTC1/SC29/WG11 Coding of Moving Pictures and Associated Audio MPEG 97/W1796, 1997.
[50] D. J. Connor, "Techniques for reducing the visibility of transmission errors in digitally encoded video signals," *IEEE Trans. Commun.*, vol. COM-21, no. 6, pp. 695–706, Jun. 1973.
[51] J. M. Boyce, "Weighted prediction in the H.264/MPEG AVC video coding standard," in *Proc. Int. Symp. Circuits and Systems*, May 2004, vol. 3, pp. 789–792.
[52] [Online]. Available: http://iphome.hhi.de/suehring/tml/download



**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 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 joined the Department of Electrical and Computer Engineering, University of California, Santa Barbara, where he is currently a Professor. 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 interested in the relations between information theory, estimation theory, and statistical physics, and their potential impact on fundamental and practical problems in diverse disciplines.

Dr. Rose was co-recipient of the 1990 William R. Bennett Prize Paper Award of the IEEE Communications Society, as well as the 2004 and 2007 IEEE Signal Processing Society Best Paper Awards.