Rate Distortion Analysis of Layered Video Coding by Leaky Prediction

Yuxin Liu[a], Paul Salama[b], Gregory W. Cook[c], and Edward J. Delp[a]

[a]Video and Image Processing Laboratory (*VIPER*)
School of Electrical and Computer Engineering
Purdue University, West Lafayette, Indiana

[b]Department of Electrical and Computer Engineering
Indiana University - Purdue University Indianapolis
Indianapolis, Indiana

[c]Department of Medicine
Division of Nephrology and Indiana Center for Biological Microscopy
Indiana University School of Medicine
Indianapolis, Indiana

ABSTRACT

Leaky prediction layered video coding (LPLC) partially includes the enhancement layer in the motion compensated prediction loop, by using a leaky factor between 0 and 1, to balance the coding efficiency and error resilience performance. In this paper, rate distortion functions are derived for LPLC from rate distortion theory. Closed form expressions are obtained for two scenarios of LPLC, one where the enhancement layer stays intact and the other where the enhancement layer suffers from data rate truncation. The rate distortion performance of LPLC is then evaluated with respect to different choices of the leaky factor, demonstrating that the theoretical analysis well conforms with the operational results.

Keywords: rate distortion theory, layered video coding, leaky prediction, error resilience

1. INTRODUCTION

Layered video coding* has a nested structure whereby different levels of the bitstream are decoded in a fixed sequential order. Fine granularity scalability (FGS) is a specific layered scalable coding structure, which possesses full rate (or SNR) scalability over a wide range of data rates [1–3]. Layered coding is desired for error resilient video streaming over heterogeneous networks with changing bandwidth mainly because: (1) It can be adapted to varying channel bandwidth by simply discarding the higher layer(s), or, by truncating the bitstream when using FGS; (2) It allows one to protect parts of the bitstream differently, i.e., the use of unequal error protection (UEP).

For error resilient video transmission in an error-prone environment, error protection can be used for the base layer since it carries more significant information. This achieves a trade-off between coding efficiency and robustness. The enhancement layer however, still remains vulnerable to errors. Due to the potential incompleteness or destruction of the enhancement layer, traditional layered coding schemes usually do not incorporate the enhancement layer into the motion compensated prediction (MCP) loop at the encoder to prevent drift at the decoder. This results in poor coding efficiency, when compared to non-scalable coding, since the high-quality reconstruction offered by both the enhancement layer and the base layer is not exploited by the motion compensation operation.

This work was supported by a grant from the Indiana 21st Century Research and Technology Fund. Address all correspondence to E. J. Delp at ace@ecn.purdue.edu, telephone: +1 765 494 1740.

*In this paper, we are specifically interested in the SNR layered video coding approach, and refer to it as layered video coding.

Visual Communications and Image Processing 2004, edited by Sethuraman Panchanathan, Bhaskaran Vasudev, Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 5308

To circumvent this coding inefficiency, leaky prediction layered coding (LPLC) [4–6] includes an incomplete version of the enhancement layer within the MCP loop to improve the coding efficiency while maintaining graceful error resilience performance. LPLC has attracted much attention in the literature recently due to its performance in handling the trade-off between coding efficiency and drift. It provides a flexible coding structure, by utilizing a leaky factor which is between 0 and 1, to down-scale the enhancement layer before it is incorporated into the MCP loop. When the leaky factor α is 0, the enhancement layer is completely excluded from the MCP loop, resulting in a codec that has the least coding efficiency and best error resilience performance. If, however, $\alpha = 1$, then the codec has the best coding efficiency and least error resilience. For intermediate values of α, the codec has intermediate coding efficiency and error resilience performance.

A deficiency inherent in the LPLC structure was described in [7], namely that LPLC cannot guarantee that the decoded video quality obtained from both the enhancement layer and the base layer will always be superior to that offered by the base layer alone. Larger leaky factors, especially when α is close to 1, might yield worse coding efficiency than smaller leaky factors. This deficiency was analytically and experimentally demonstrated, and confirmed by addressing the similarity between LPLC and a specific multiple description coding (MDC) scheme, namely MDMC [8]. A general framework, which applies to both LPLC and MDMC, was further established. Based on this framework, maximum-likelihood (ML) estimation, which was originally used in MDC [9], was utilized, and a new approach, referred to as ML-LPLC, was proposed to ameliorate the specified deficiency. It was emphasized in [7] that the leaky factor is critical for the LPLC approach, which determines three types of the performance of LPLC: the coding efficiency, the prediction drift (or in another sense, the error resilience performance), as well as the superiority of the enhancement layer as opposed to the base layer. In [10], the dual-leaky prediction framework was proposed, which introduces nested scalability into each description of the MDC stream, and multiple description scalable coding (MDSC) was achieved.

The analysis of the MCP based video coding, derived from rate distortion theory, was presented in [11]. Essentially, the analysis characterized the properties of the MCP operation by a stochastic filter, which fulfills the operations of time delay, motion compensation, and spatial filtering, combined with the optimum forward channel that yields the rate distortion (in the mean-square-error (MSE) sense) for Gaussian stationary signals [12]. It was shown that the power spectral density (PSD) of the MCP error signal is related with the PSD of the input signal as well as the distribution of the estimated motion vector errors.

The rate distortion analysis of conventional layered scalable video coding, which excludes the enhancement layer from the MCP loop, is described in [13]. The MCP rate is first defined, which denotes the data rate (in bits) that is incorporated into the MCP loop at the encoder. The work in [13] is advanced by first developing the rate distortion analysis for the MSE optimal layered image codec, as well as the cascaded optimal image codec. The rate distortion analysis of layered video coding is then presented, with two scenarios derived in closed form, where the bitstream is decoded *above* and *below* the MCP rate.

When a scalable coded bitstream is decoded *above* the MCP rate, the base layer at the decoder is consistent with that at the encoder and no drift occurs. In this case, the well-accepted fact, that layered scalable coding always demands more or at least as much data rate as required by the non-scalable coding approach to achieve the same distortion, was theoretically confirmed. When the bitstream is decoded *below* the MCP rate, the base layer is truncated upon decoding and thus drift occurs. It was theoretically demonstrated that the distortion in this case steeply increased with the decrease of the decoding data rate. The theoretical analysis in [13] is shown to agree with operational rate distortion results published in the literature.

In this paper, we present the rate distortion analysis of LPLC by extending the work in [13]. First, we describe a block diagram that features the leaky prediction but is amenable to theoretical analysis. We derive the rate distortion functions for LPLC in closed form for one scenario where the enhancement layer is intact and the other where it has drift. We demonstrate that the leaky factor is critical in the performance of coding efficiency, and validate that with the partial or full inclusion of the enhancement layer in the MCP loop, LPLC does improve the coding efficiency as opposed to the conventional layered scalable coding. We also show that the leaky factor is critical in the error resilience performance when the enhancement layer in LPLC suffers from data rate truncation.

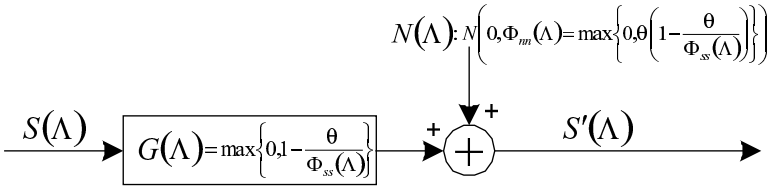

Figure 1. Optimum forward channel that yields the Gaussian MSE rate distortion function

2. RATE DISTORTION FUNCTIONS FOR TWO-DIMENSIONAL IMAGE CODING

As background knowledge to this paper, we first summarize the fundamental results of the analysis of two-dimensional (2D) image coding derived from rate distortion theory.

Theorem 1. [12][11]: The parametric representation of the MSE rate distortion functions for a 2D stationary Gaussian source $\{s\}$ is given by

$$D^{\theta} = \frac{1}{4\pi^2} \iint_{\Lambda} \min\{\theta, \Phi_{ss}(\Lambda)\} d\Lambda, \tag{1}$$

$$R^{\theta} = \frac{1}{8\pi^2} \iint_{\Lambda} \max\left\{0, \log_2\left(\frac{\Phi_{ss}(\Lambda)}{\theta}\right)\right\} d\Lambda, \tag{2}$$

where $\Phi_{ss}(\Lambda)$ denotes the PSD of the signal $\{s\}$ at the spatial frequencies $\Lambda = (\omega_x, \omega_y)$, and θ denotes the parameter ranging over $[0, \text{ess sup}\,\Phi(\Lambda)]$, with $\text{ess sup}\,\Phi_{ss}(\Lambda)$ indicating the essential supremum of $\Phi_{ss}(\Lambda)$.

The following proposition makes eases of further analysis of the rate distortion performance of image and video signals.

Proposition 1. [12][11] The optimum forward channel yielding the MSE parametric rate distortion functions for a Gaussian process $\{s\}$, as given by (1) and (2), is composed of a non-ideal bandlimit filter over $\{\Lambda : \Phi_{ss}(\Lambda) > \theta\}$, plus the addition of an independent non-white bandlimit Gaussian noise over $\{\Lambda : \Phi_{ss}(\Lambda) > \theta\}$, as shown in Figure 1.

It is easy to show Proposition 1 with the help of the following lemma:

Lemma 1. [14] Assume $\{s\}$ and $\{s'\}$ are jointly 2D stationary Gaussian processes, with $\Phi_{ss}(\Lambda)$, $\Phi_{s's'}(\Lambda)$, and $\Phi_{ss'}(\Lambda)$ denoting the PSD of $\{s\}$, the PSD of $\{s'\}$, and the cross spectral density between $\{s\}$ and $\{s'\}$ respectively. The mutual information rate between $\{s\}$ and $\{s'\}$ is then given by

$$I(S; S') = -\frac{1}{8\pi^2} \iint_{\Lambda} \log\left[1 - \frac{|\Phi_{ss'}(\Lambda)|^2}{\Phi_{ss}(\Lambda)\Phi_{s's'}(\Lambda)}\right] d\Lambda, \tag{3}$$

where each integral extends over $[-\pi, \pi]$ for discrete time processes, and $[-\infty, \infty]$ for continuous time processes.

Two scenarios of 2D image coding are further analyzed in [13]:

Scenario I for 2D image coding: The rate distortion functions by both layers in the scalable image coding structure, as described in Figure 2 where the base layer is parametric with θ and the enhancement layer with $\tilde{\theta}$, are

$$D_e^{I,\theta,\tilde{\theta}} = \frac{1}{4\pi^2} \iint_{\Lambda} \min\{\min\{\theta, \tilde{\theta}\}, \Phi_{ss}(\Lambda)\} d\Lambda, \tag{4}$$

$$R_e^{I,\theta,\tilde{\theta}} = \frac{1}{8\pi^2} \iint_{\Lambda} \max\left\{0, \log_2\left(\frac{\Phi_{ss}(\Lambda)}{\min\{\theta, \tilde{\theta}\}}\right)\right\} d\Lambda. \tag{5}$$

It is straightforward to show (4) by noting that $s - s'_e = \tilde{s}_b - \tilde{s}'_b$, and to show (5) by summing the mutual information between $\{s\}$ and $\{s'_b\}$ plus the mutual information between $\{\tilde{s}_b\}$ and $\{\tilde{s}'_b\}$.

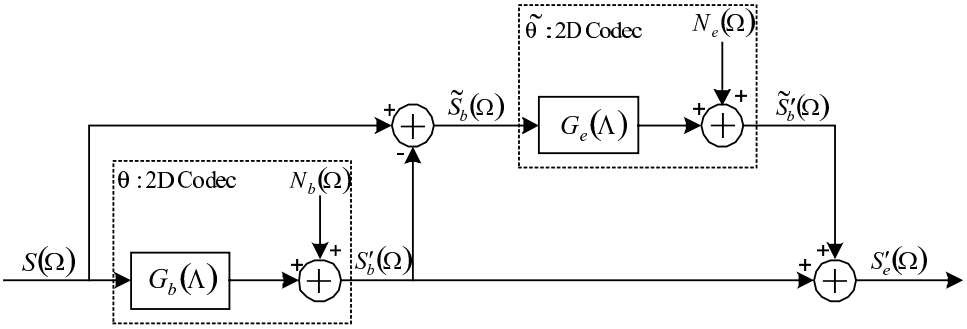

Figure 2. Block diagram of a scalable image codec

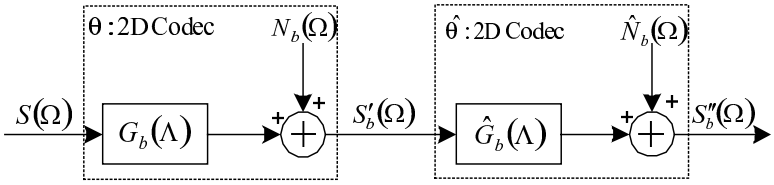

Figure 3. Block diagram of a cascaded image codec

Scenario II for 2D image coding: The rate distortion functions of the cascaded non-scalable image codec, as described in Figure 3, where the first codec is parametric with θ and the second codec with $\hat{\theta}$, and let $\tilde{\theta} = \theta + \hat{\theta}$, are

$$D^{II,\theta,\tilde{\theta}} = \frac{1}{4\pi^2} \iint_{\Lambda} \min\{\tilde{\theta}, \Phi_{ss}(\Lambda)\} d\Lambda, \tag{6}$$

$$R^{II,\theta,\tilde{\theta}} = \frac{1}{8\pi^2} \iint_{\Lambda} \max\left\{0, \log_2\left(\frac{\Phi_{ss}(\Lambda)}{\tilde{\theta}}\right)\right\} d\Lambda. \tag{7}$$

Note that (6) can be derived by noticing that $\{\tilde{s}_b \triangleq s - s'_b\}$ and $\{\tilde{s}'_b \triangleq s'_b - s''_b\}$ are uncorrelated [13], and (7) can be derived using the following proposition:

Proposition 2. The cascaded Gaussian MSE optimum forward channel is still optimal in the rate distortion sense. Furthermore, the parameter of the equivalent optimum forward channel is the sum of the parameters featuring each of the cascaded channels.

Proposition 2 can be proved in a similar way as the proof of Proposition 1, which is derived from Lemma 1.

3. RATE DISTORTION FUNCTIONS FOR CONVENTIONAL LAYERED CODING

The rate distortion analysis of conventional layered video coding is presented in [13] by extending the work on the rate distortion analysis of the non-scalable MCP video coding derived in [11]. We summarize the major results [13].

Scenario I for conventional layered video coding: As shown in Figure 4, two optimum forward channels are included, yielding the rate distortion optimized 2D signal compression of the base layer and the enhancement layer respectively. Referring to the rate distortion analysis in *Scenario I for 2D image coding*, we have the rate distortion functions for both layers of the conventional layered video coding as follows when the parameters

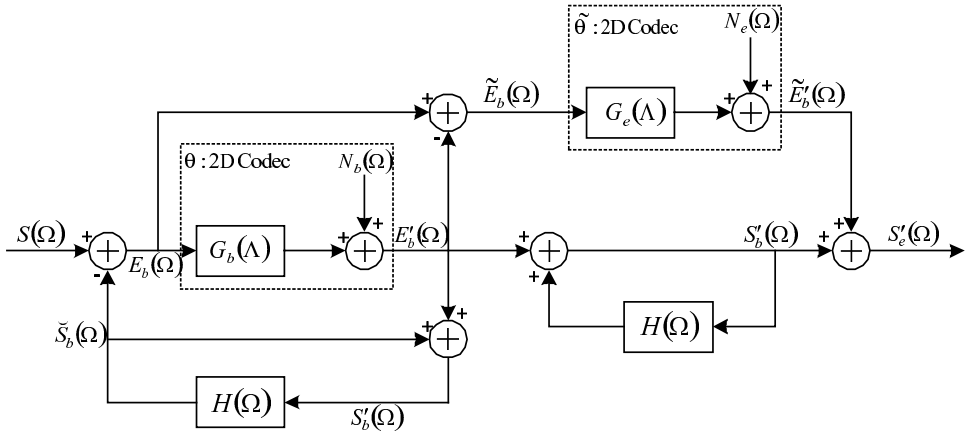

Figure 4. Block diagram of a conventional layered video codec when the base layer is decoded above the MCP rate

satisfy $\tilde{\theta} \leq \theta$:

$$D_e^{I,\theta,\tilde{\theta}} = \frac{1}{4\pi^2} \iint_\Lambda \min\{\tilde{\theta}, \Phi^\theta_{e_b e_b}(\Lambda)\} d\Lambda, \tag{8}$$

$$R_e^{I,\theta,\tilde{\theta}} = \frac{1}{8\pi^2} \iint_\Lambda \max\left\{0, \log_2\left(\frac{\Phi^\theta_{e_b e_b}(\Lambda)}{\tilde{\theta}}\right)\right\} d\Lambda. \tag{9}$$

Similar to the analysis in [11], the PSD of the MCP error signal in the base layer, $\Phi^\theta_{e_b e_b}(\Lambda)$, can be approximated as

$$\Phi^\theta_{e_b e_b}(\Lambda) \approx \Phi^{appr,\theta}_{e_b e_b}(\Lambda) = \begin{cases} \Phi_{ss}(\Lambda) & \Lambda : \Phi_{ss}(\Lambda) \leq \theta \\ \max\{\theta, \Phi^{I,\theta}_{e_b e_b}(\Lambda)\} & \Lambda : \Phi_{ss}(\Lambda) > \theta \end{cases}, \tag{10}$$

where

$$\Phi^{I,\theta}_{e_b e_b}(\Lambda) = \Phi_{ss}(\Lambda)[1 - 2\mathrm{Re}\{F(\Lambda)P(\Lambda)\} + |F(\Lambda)|^2] + \theta|F(\Lambda)|^2, \tag{11}$$

where $P(\Lambda)$ denotes the characteristic function of the probability density function (*p.d.f.*) of the estimated motion vector error. It is discussed in [11] and [13] that the spatial filter $F(\Lambda)$ can be chosen as 0 for intra-frame coding, or 1 for inter-frame coding without spatial filtering. The optimal $F(\Lambda)$ that minimizes $\Phi^{I,\theta}_{e_b e_b}(\Lambda)$ in (11) is

$$F_{opt}(\Lambda) = P^*(\Lambda)\frac{\Phi_{ss}(\Lambda)}{\Phi_{ss}(\Lambda) + \theta}, \qquad \text{for } \Lambda : \Phi_{ss}(\Lambda) > \theta, \tag{12}$$

which results

$$\Phi^{I,\theta}_{e_b e_b}(\Lambda) = \Phi_{ss}(\Lambda)\left(1 - \frac{|P(\Lambda)|^2 \Phi_{ss}(\Lambda)}{\Phi_{ss}(\Lambda) + \theta}\right). \tag{13}$$

Moreover, by (10), the distortion function in (8) can be simplified to

$$D_e^{I,\theta,\tilde{\theta}} = \frac{1}{4\pi^2} \iint_\Lambda \min\{\tilde{\theta}, \Phi_{ss}(\Lambda)\} d\Lambda. \tag{14}$$

Note that the PSD of the input signal to the enhancement layer forward channel, $\{\tilde{e}_b\}$, satisfies that

$$\Phi_{\tilde{e}_b \tilde{e}_b}(\Lambda) = \min\{\theta, \Phi^\theta_{e_b e_b}(\Lambda)\} = \min\{\theta, \Phi_{ss}(\Lambda)\} \leq \theta. \tag{15}$$

Thus when $\tilde{\theta} > \theta$, the filter in the enhancement layer in Figure 4, $G_e(\Lambda)$, becomes zero. As a result the data rate consumed by the enhancement layer is zero, and no distortion is further caused beyond that by the base layer.

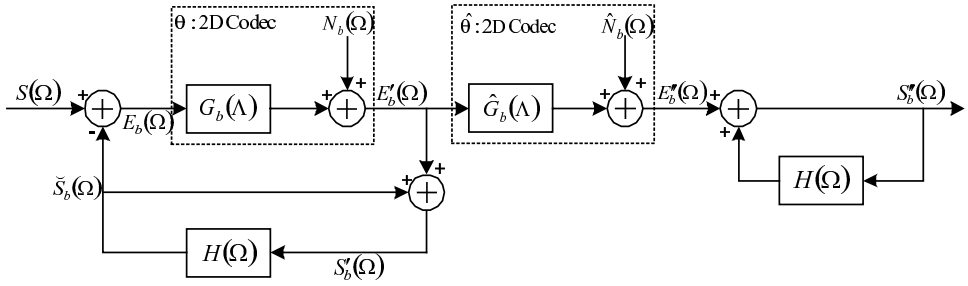

Figure 5. Block diagram of a conventional layered video codec when the base layer is decoded below the MCP rate

The rate distortion function of the MCP layered codec described in Figure 4 is then fixed at one point that is specified by the base layer, when θ is fixed and $\tilde{\theta}$ varies between θ and infinity:

$$D_b^{\theta} = \frac{1}{4\pi^2}\iint_{\Lambda}\min\{\theta, \Phi_{ss}(\Lambda)\}d\Lambda, \tag{16}$$

$$R_b^{\theta} = \frac{1}{8\pi^2}\iint_{\Lambda}\max\left\{0, \log_2\left(\frac{\Phi^{\theta}_{e_be_b}(\Lambda)}{\theta}\right)\right\}d\Lambda. \tag{17}$$

Scenario II for conventional layered video coding: The block diagram of the layered video coding when the base layer is decoded below the MCP rate is shown in Figure 5.

As discussed in Section 2, the cascaded optimum forward channel linking $\{e_b\}$ and $\{e''_b\}$ is equivalent to one optimum forward channel with the parameter of $\theta + \hat{\theta}$. The rate distortion functions are then given by [13]:

$$D^{II,\theta,\tilde{\theta}} = \frac{1}{4\pi^2}\iint_{\Lambda}\min\{\theta, \Phi_{ss}(\Lambda)\} + \frac{1}{1-|F(\Lambda)|^2}\min\{\tilde{\theta}-\theta, \max\{0, \Phi^{\theta}_{e_be_b}(\Lambda)-\theta\}\}d\Lambda, \tag{18}$$

$$R^{II,\theta,\tilde{\theta}} = \frac{1}{8\pi^2}\iint_{\Lambda}\max\left\{0, \log_2\left(\frac{\Phi^{\theta}_{e_be_b}(\Lambda)}{\tilde{\theta}}\right)\right\}d\Lambda, \tag{19}$$

where $\tilde{\theta} = \theta + \hat{\theta} \geq \theta$.

4. RATE DISTORTION FUNCTIONS FOR LEAKY PREDICTION LAYERED CODING

Unlike the conventional layered coding structure described in Figure 4, LPLC introduces a second MCP loop in the enhancement layer that uses the same motion vectors as the base layer, and buffers $\alpha(s'_e(t)-s'_b(t))+s'_b(t)$ as the reference for the encoding of the video signal sampled at time $t+\triangle t$. The mismatch signal, $\{\psi\}$, which is the difference between the MCP error signal in the enhancement layer, $\{e_e\}$, and the encoded MCP error signal in the base layer, $\{e'_b\}$, is coded and carried by the enhancement layer. Equivalently, a linear combination of the two reconstructed signals, $\{s'_e\}$ yielded by both layers and $\{s'_b\}$ by the base layer alone, namely $\alpha s'_e(t)+(1-\alpha)s'_b(t)$, is utilized as the reference signal in the MCP loop of the enhancement layer. The framework of LPLC is described in Figure 6.

It is shown in [11] that since $\Phi^{I,\theta}_{e_be_b}(\Lambda)$ represents the value of $\Phi^{\theta}_{e_be_b}(\Lambda)$ for $\Lambda : \Phi^{\theta}_{e_be_b}(\Lambda) \gg \theta$, the optimal spatial filter given in (12) can be approximated by

$$F_{opt}(\Lambda) \approx P^*(\Lambda). \tag{20}$$

If the same spatial filter is used in both MCP loops in Figure 6, the 3D filters combining spatial filtering and motion compensation in the base layer and the enhancement layer, $H_b(\Omega)$ and $H_e(\Omega)$, become identical, which

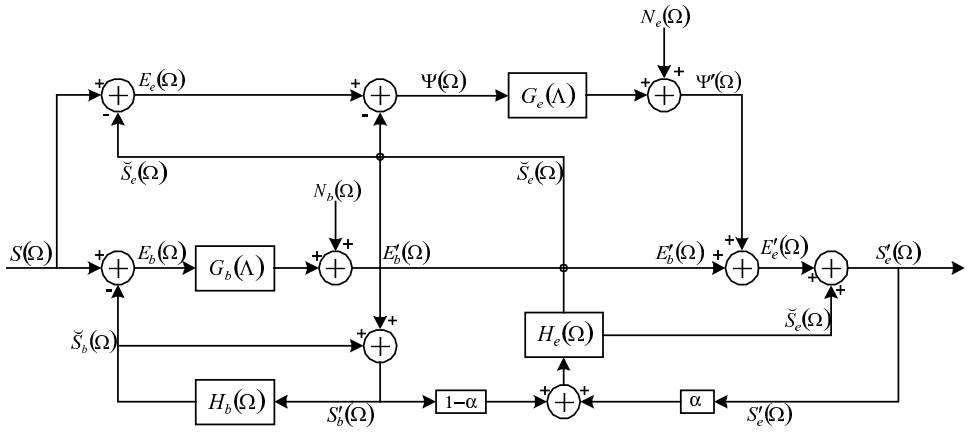

Figure 6. Block diagram of an leaky prediction layered video codec (LPLC)

is referred to as $H(\Omega)$ hereafter. We then have

$$\begin{aligned} S'_e(\Omega) &= E'_e(\Omega) + \check{S}_e(\Omega) = E'_b(\Omega) + \Psi'(\Omega) + \check{S}_e(\Omega) \\ &= E'_b(\Omega) + \Psi'(\Omega) + H(\Omega)[(1-\alpha)S'_b(\Omega) + \alpha S'_e(\Omega)], \end{aligned}$$

which derives

$$S'_e(\Omega) = \frac{E'_b(\Omega) + H(\Omega)(1-\alpha)S'_b(\Omega) + \Psi'(\Omega)}{1-\alpha H(\Omega)} = S'_b(\Omega) + \frac{1}{1-\alpha H(\Omega)}\Psi'(\Omega). \tag{21}$$

Moreover, we have

$$\Psi(\Omega) = E_e(\Omega) - E'_b(\Omega) = S(\Omega) - \check{S}_e(\Omega) - E'_b(\Omega) = (S(\Omega) - S'_b(\Omega)) - \frac{\alpha H(\Omega)}{1-\alpha H(\Omega)}\Psi'(\Omega).$$

Since

$$S(\Omega) - S'_b(\Omega) = E_b(\Omega) - E'_b(\Omega) \triangleq \tilde{E}_b(\Omega), \tag{22}$$

then

$$\Psi(\Omega) = \tilde{E}_b(\Omega) - \frac{\alpha H(\Omega)}{1-\alpha H(\Omega)}(G_e(\Lambda)\Psi(\Omega) + N_e(\Omega)),$$

hence

$$\Psi(\Omega) = \frac{1-\alpha H(\Omega)}{1-\alpha H(\Omega) + G_e(\Lambda)(\alpha H(\Omega))}\tilde{E}_b(\Omega) - \frac{\alpha H(\Omega)}{1-\alpha H(\Omega) + G_e(\Lambda)(\alpha H(\Omega))}N_e(\Omega). \tag{23}$$

We observe that the Fourier transform of the mismatch signal $\{\psi\}$ in (23) has exactly the same form as that of the base layer MCP error signal $\{e_b\}$ that is related to $\{s\}$ and $\{n_b\}$ [11]. Here $\{\tilde{e}_b\}$ acts as the input signal to the MCP loop as opposed to $\{s\}$, $\{n_e\}$ serves as the additional Gaussian noise in the optimum forward channel as opposed to $\{n_b\}$, and $(\alpha H(\Omega))$ the MCP 3D filter that combines spatial filtering and motion compensation as opposed to $H(\Omega)$. Therefore, we obtain an alternative diagram for LPLC as in Figure 7.

Next we complete the rate distortion analysis of two scenarios for LPLC: with and without drift in the enhancement layer.

Scenario I for LPLC: Following a similar manner as the analysis developed for $\Phi^{\theta}_{e_b e_b}(\Lambda)$, as given by (10) and (11), we obtain an approximation of the PSD of the mismatch signal $\{\psi\}$ as

$$\Phi^{\tilde{\theta}}_{\psi\psi}(\Lambda) \approx \Phi^{appr,\tilde{\theta}}_{\psi\psi}(\Lambda) = \begin{cases} \Phi_{\tilde{e}_b\tilde{e}_b}(\Lambda) & \Lambda : \Phi_{\tilde{e}_b\tilde{e}_b}(\Lambda) \le \tilde{\theta} \\ \max\{\tilde{\theta}, \Phi^{I,\tilde{\theta}}_{\psi\psi}(\Lambda)\} & \Lambda : \Phi_{\tilde{e}_b\tilde{e}_b}(\Lambda) > \tilde{\theta} \end{cases}, \tag{24}$$

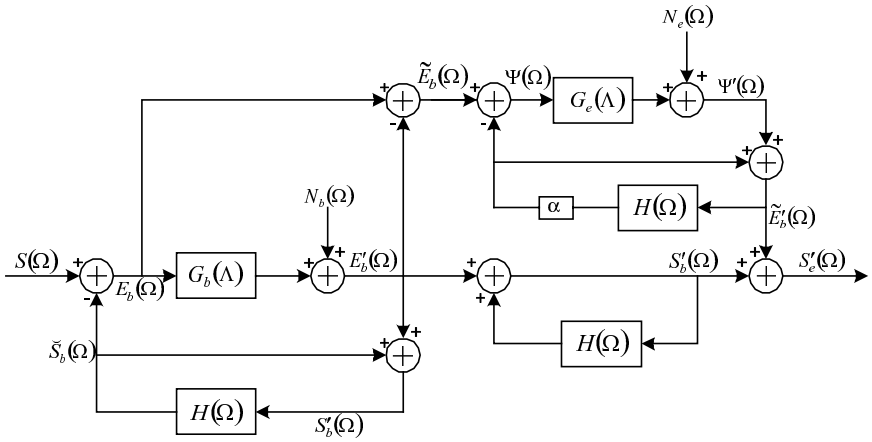

Figure 7. Alternative block diagram of the LPLC codec

where

$$\Phi_{\psi\psi}^{I,\tilde{\theta}}(\Lambda) = \Phi_{\tilde{e}_b\tilde{e}_b}(\Lambda)[1 - 2\alpha\mathrm{Re}\{F(\Lambda)P(\Lambda)\} + \alpha^2|F(\Lambda)|^2] + \tilde{\theta}\alpha^2|F(\Lambda)|^2, \tag{25}$$

and $\Phi_{\tilde{e}_b\tilde{e}_b}(\Lambda)$ is given by (15). When $F(\Lambda)$ is chosen as the optimum spatial filter for the base layer in (20), we have

$$\Phi_{\psi\psi}^{I,\tilde{\theta}}(\Lambda) = \Phi_{\tilde{e}_b\tilde{e}_b}(\Lambda)[1 - \alpha(2-\alpha)|P(\Lambda)|^2] + \tilde{\theta}\alpha^2|P(\Lambda)|^2. \tag{26}$$

Note that only when $\alpha = 1$, $F(\Lambda)\approx P^*(\Lambda)$ is also approximately optimized for the minimization of $\Phi_{\psi\psi}^{I,\tilde{\theta}}(\Lambda)$ for $\Lambda : \Phi_{\tilde{e}_b\tilde{e}_b}(\Lambda) > \tilde{\theta}$.

Moreover, according to Figure 7, we have

$$S(\Omega) - S'_e(\Omega) = S(\Omega) - S'_b(\Omega) - (S'_e(\Omega) - S'_b(\Omega)) = \tilde{E}_b(\Omega) - \tilde{E}'_b(\Omega) = \Psi(\Omega) - \Psi'(\Omega) \triangleq \tilde{\Psi}(\Omega). \tag{27}$$

Similar to the way we derived $\Phi_{\tilde{e}_b\tilde{e}_b}(\Lambda)$ in (15), we have

$$\Phi_{\tilde{\psi}\tilde{\psi}}(\Lambda) = \min\{\tilde{\theta}, \Phi_{\psi\psi}^{\tilde{\theta}}(\Lambda)\} = \min\{\tilde{\theta}, \Phi_{\tilde{e}_b\tilde{e}_b}(\Lambda)\}. \tag{28}$$

Similar to the discussion for *Scenario I for conventional layered video coding* in Section 3, we choose $\tilde{\theta} \leq \theta$ first. Thus according to (15), we simplify (28) as

$$\Phi_{\tilde{\psi}\tilde{\psi}}(\Lambda) = \min\{\tilde{\theta}, \Phi_{ss}(\Lambda)\}. \tag{29}$$

Therefore, the distortion in the MSE sense between the input video signal and the decoded signal by both layers in LPLC is

$$D_e^{I,\theta,\tilde{\theta}} = E\{(s - s'_e)^2\} = \frac{1}{4\pi^2}\iint_\Lambda \Phi_{\tilde{\psi}\tilde{\psi}}(\Lambda)d\Lambda = \frac{1}{4\pi^2}\iint_\Lambda \min\{\tilde{\theta}, \Phi_{ss}(\Lambda)\}d\Lambda. \tag{30}$$

Note that the distortion function given in (30) for LPLC has the same form as in (14) for the conventional layered video coding structure.

The data rate in units of bits consumed by the LPLC framework described in Figure 7 is the sum of the mutual information between $\{e_b\}$ and $\{e'_b\}$ plus the mutual information between $\{\psi\}$ and $\{\psi'\}$. Therefore,

$$R_e^{I,\theta,\tilde{\theta}} = \frac{1}{8\pi^2}\iint_\Lambda \max\left\{0, \log_2\left(\frac{\Phi_{e_be_b}^{\theta}(\Lambda)}{\theta}\right)\right\} + \max\left\{0, \log_2\left(\frac{\Phi_{\psi\psi}^{\tilde{\theta}}(\Lambda)}{\tilde{\theta}}\right)\right\} d\Lambda, \tag{31}$$

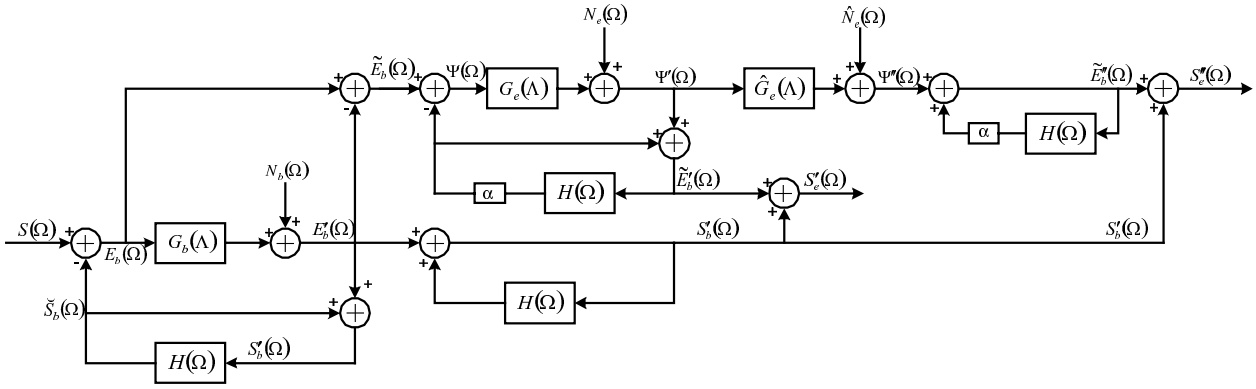

Figure 8. Block diagram of an LPLC codec when the enhancement layer is decoded below the MCP rate

where $\Phi^{\tilde{\theta}}_{\psi\psi}(\Lambda)$ can be approximately obtained by (24), and $\Phi^{\theta}_{e_be_b}(\Lambda)$ approximately by (10). Note that when $\alpha = 0$, LPLC reduces to the conventional layered coding structure. At this time, $\Phi^{\tilde{\theta}}_{\psi\psi}(\Lambda)$ equals to $\Phi_{\tilde{e}_b\tilde{e}_b}(\Lambda)$ according to (24) and (25), and the rate function given by (31) reduces to the same form as in (9).

When $\tilde{\theta} > \theta$, we have $\Phi^{\tilde{\theta}}_{\psi\psi}(\Lambda) = \Phi_{\tilde{e}_b\tilde{e}_b}(\Lambda) \leq \theta < \tilde{\theta}$. It is easy to show that the rate distortion functions for both layers of LPLC, $D_e^{I,\theta,\tilde{\theta}}$ and $R_e^{I,\theta,\tilde{\theta}}$ as given by (30) and (31), reduce to the functions of the base layer that are parametric with θ, and have exactly the same form as in (16) and (17).

Scenario II for LPLC: As we mentioned, LPLC is designed to balance coding efficiency and the error resilience performance. The use of the leaky factor, when less than 1, is targeted to ameliorate the effect of error propagation that is caused by the truncation or destruction of the enhancement layer in the MCP loop. The rate distortion functions are next derived for LPLC when drift occurs in the enhancement layer. Hence evaluate the error resilience performance of LPLC with respect to α.

Based on the ideas used in *Scenario II for conventional layered video coding* in Section 3 where drift occurs in the base layer in the MCP loop, we introduce a third optimum forward channel taking the coded mismatch signal $\{\psi'\}$ as its input. This channel models the scenario where the mismatch signal carried by the enhancement layer is decoded below the MCP rate, as described in Figure 8.

For the convenience of analysis, we change the parameter that yields the rate distortion functions between $\{\psi\}$ and $\{\psi'\}$ to $\breve{\theta}$, and let $\tilde{\theta} \triangleq \breve{\theta} + \hat{\theta}$. Since

$$S(\Omega) - S_e''(\Omega) = (\Psi(\Omega) - \Psi'(\Omega)) + \frac{1}{1-\alpha H(\Omega)}(\Psi'(\Omega) - \Psi''(\Omega)) \triangleq \tilde{\Psi}(\Omega) + \frac{1}{1-\alpha H(\Omega)}\hat{\Psi}(\Omega), \tag{32}$$

and the two terms in the right hand side of (32) are uncorrelated with each other[†], we have

$$D_e^{II,\theta,\tilde{\theta}} = \frac{1}{4\pi^2}\iint_{\Lambda} \Phi_{\tilde{\psi}\tilde{\psi}}(\Lambda) + E\left\{\left|\frac{1}{1-\alpha H(\Omega)}\right|^2\right\}\Phi_{\hat{\psi}\hat{\psi}}(\Lambda)d\Lambda, \tag{33}$$

where

$$E\left\{\left|\frac{1}{1-\alpha H(\Omega)}\right|^2\right\} = \frac{1}{1-\alpha^2|F(\Lambda)|^2}, \tag{34}$$

which is derived in a similar way as in [13],

$$\Phi_{\tilde{\psi}\tilde{\psi}}(\Lambda) = \min\{\breve{\theta}, \Phi^{\breve{\theta}}_{\psi\psi}(\Lambda)\}, \tag{35}$$

[†]This can be proved in a similar way as in [13].

and

$$\Phi_{\hat{\psi}\hat{\psi}}(\Lambda) = \min\{\hat{\theta}, \Phi_{\psi'\psi'}(\Lambda)\} = \min\{\tilde{\theta} - \breve{\theta}, \max\{0, \Phi^{\breve{\theta}}_{\psi\psi}(\Lambda) - \breve{\theta}\}\}. \tag{36}$$

When $F(\Lambda)$ is evaluated as $P^*(\Lambda)$, we have the distortion between the input video signal and the decoded signal, when the mismatch signal carried by the enhancement layer is decoded below the MCP rate, as

$$D_e^{II,\theta,\tilde{\theta}} = \frac{1}{4\pi^2} \iint_{\Lambda} \min\{\breve{\theta}, \Phi^{\breve{\theta}}_{\psi\psi}(\Lambda)\} + \frac{1}{1-\alpha^2|P(\Lambda)|^2} \min\{\tilde{\theta} - \breve{\theta}, \max\{0, \Phi^{\breve{\theta}}_{\psi\psi}(\Lambda) - \breve{\theta}\}\} d\Lambda. \tag{37}$$

The rate consumed by the LPLC framework in Figure 8 is the sum of the mutual information between $\{e_b\}$ and $\{e'_b\}$ plus the mutual information between $\{\psi\}$ and $\{\psi''\}$, therefore,

$$R_e^{II,\theta,\tilde{\theta}} = \frac{1}{8\pi^2} \iint_{\Lambda} \max\left\{0, \log_2\left(\frac{\Phi^{\theta}_{e_b e_b}(\Lambda)}{\theta}\right)\right\} + \max\left\{0, \log_2\left(\frac{\Phi^{\breve{\theta}}_{\psi\psi}(\Lambda)}{\tilde{\theta}}\right)\right\} d\Lambda. \tag{38}$$

Note that if θ is fixed for the base layer optimum forward channel, and $\breve{\theta} > \theta$, then no information is carried by the enhancement layer. The rate distortion functions given by (37) and (38) reduce to a fixed point that is parametric with θ, which is the same as (16) and (17). The PSD of the output signal from the first channel in the enhancement layer, $\Phi_{\psi'\psi'}(\Lambda)$, is zero, and no drift occurs in the enhancement layer. If the encoded bitstream still suffers from data rate truncation, it will be the base layer that suffers from drift, which yields *Scenario II* in Section 3 where the base layer is decoded below the MCP rate. Therefore, we have $\breve{\theta} \leq \theta$ for modelling the case where the base layer is intact while the enhancement layer is likely to suffer from drift.

If $\breve{\theta}$ is fixed, namely $\breve{\theta}_0$ satisfying $\breve{\theta}_0 < \theta$, the parameter of the second optimum forward channel in the enhancement layer, $\hat{\theta}$, which models the drift that affects the enhancement layer, ranges between 0 and $\theta - \breve{\theta}_0$. When $\hat{\theta} > \theta - \breve{\theta}_0$, i.e., $\tilde{\theta} > \theta$, the rate distortion function specified by (37) and (38) also stays at a fixed point. Given the above and combining (24), (25), and (15), we have

$$\Phi^{\breve{\theta}_0}_{\psi\psi}(\Lambda) \leq \Phi_{\tilde{e}_b\tilde{e}_b}(\Lambda) \leq \theta < \tilde{\theta}, \tag{39}$$

thus the rate function in (38) reduces to the same form as (17), but the distortion function reduces to

$$D_e^{II,\theta,\tilde{\theta}} = \frac{1}{4\pi^2} \iint_{\Lambda} \min\{\breve{\theta}_0, \Phi^{\breve{\theta}_0}_{\psi\psi}(\Lambda)\} + \frac{1}{1-\alpha^2|P(\Lambda)|^2} \max\{0, \Phi^{\breve{\theta}_0}_{\psi\psi}(\Lambda) - \breve{\theta}_0\} d\Lambda,$$
$$for\ \ \breve{\theta}_0 < \theta\ \ and\ \ \tilde{\theta} > \theta. \tag{40}$$

Note that when $\alpha = 0$,

$$\min\{\breve{\theta}_0, \Phi^{\breve{\theta}_0}_{\psi\psi}(\Lambda)\} + \frac{1}{1-\alpha^2|P(\Lambda)|^2} \max\{0, \Phi^{\breve{\theta}_0}_{\psi\psi}(\Lambda) - \breve{\theta}_0\} = \Phi^{\breve{\theta}_0}_{\psi\psi}(\Lambda) = \Phi_{\tilde{e}_b\tilde{e}_b}(\Lambda),$$

and the distortion given by (40) reduces to that given by (16). When $\alpha = 1$, for $\Lambda : \Phi_{\tilde{e}_b\tilde{e}_b}(\Lambda) > \breve{\theta}_0$, we have $\Phi^{\breve{\theta}_0}_{\psi\psi}(\Lambda) \geq \breve{\theta}_0$, and

$$\frac{1}{1-|P(\Lambda)|^2} \max\{0, \Phi^{\breve{\theta}_0}_{\psi\psi}(\Lambda) - \breve{\theta}_0\} = \max\left\{0, \frac{\Phi_{\tilde{e}_b\tilde{e}_b}(\Lambda)(1-|P(\Lambda)|^2) + \breve{\theta}_0|P(\Lambda)|^2 - \breve{\theta}_0}{1-|P(\Lambda)|^2}\right\} = \Phi_{\tilde{e}_b\tilde{e}_b}(\Lambda) - \breve{\theta}_0.$$

Thus it is easy to show that the distortion given by (40) also reduces to that given by (16) for $\alpha = 1$. Due to the approximation we used in (24) for the evaluation of $\Phi^{\breve{\theta}}_{\psi\psi}(\Lambda)$, when $0 < \alpha < 1$, the distortion (40) is usually smaller than that for (16).

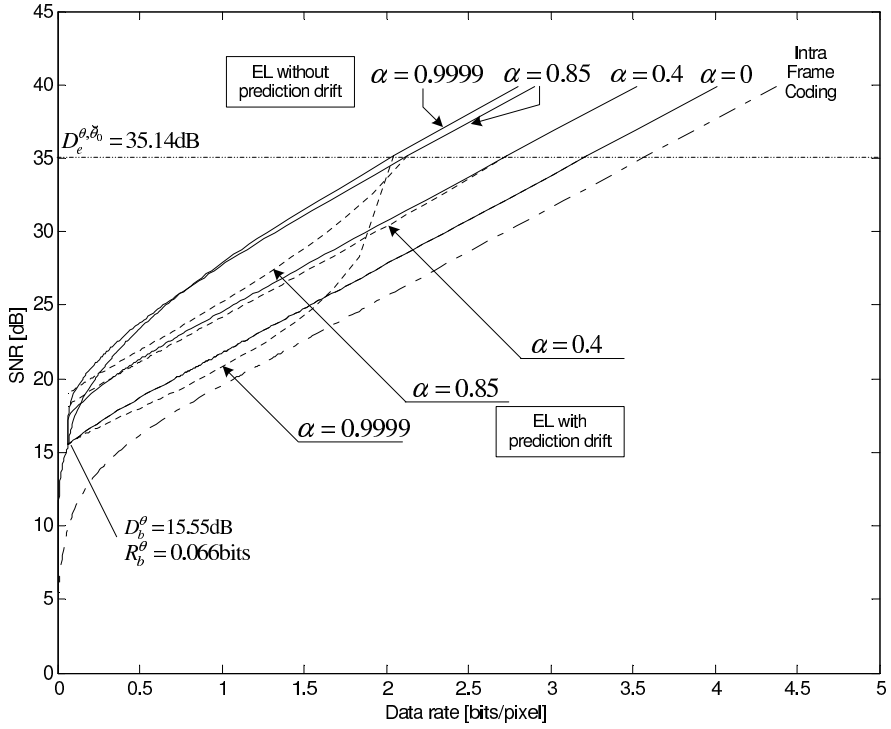

Figure 9. Rate distortion functions of LPLC for various leaky factors (α) ($\sigma^2_{\triangle d} = 0.04$ for $P(\Lambda)$ in (42))

5. RATE DISTORTION PERFORMANCE EVALUATION OF LPLC

Similar to [11] and [13], we model the PSD of the input video signal as

$$\Phi_{ss}(\Lambda) = \Phi_{ss}(\omega_x, \omega_y) = \begin{cases} \frac{2\pi}{\omega_0^2}\left(1 + \frac{\omega_x^2+\omega_y^2}{\omega_0^2}\right)^{-3/2} & |\omega_x| \le \pi f_{sx} \text{ and } |\omega_y| \le \pi f_{sy} \\ 0 & \text{otherwise} \end{cases}, \tag{41}$$

where f_{sx} and f_{sy} denote the sampling frequencies when $\{s\}$ is spatially sampled at the Nyquist rate, and $\omega_0 = \frac{\pi f_{sx}}{42.19} = \frac{\pi f_{sy}}{46.15}$. We model the *p.d.f.* of the estimated motion vector error with its characteristic function

$$P(\Lambda) = \exp\left[-\frac{\sigma^2_{\triangle d}}{2}\Lambda \cdot \Lambda\right] = \exp\left[-\frac{\sigma^2_{\triangle d}}{2}(\omega_x^2 + \omega_y^2)\right], \tag{42}$$

where $\sigma^2_{\triangle d}$ denotes the variance of the estimated motion vector error.

As described in Figure 9, we evaluate the rate distortion performance of LPLC with respect to the leaky factor, α, according to the closed forms we derived for the two scenarios for LPLC in Section 4.

Results of *Scenario I for LPLC* are shown in solid lines in Figure 9, where the enhancement layer (denoted as EL in the figure) does not suffer from drift in LPLC. When θ is fixed at the point that yields the rate distortion function at $D_b^\theta = 15.55$ dB in (16) and $R_b^\theta = 0.066$ bits in (17), we vary the parameter $\tilde{\theta}$ to obtain the rate distortion functions for different values of α according to (30) and (31). It is shown that when the data rate is sufficiently large, LPLC achieves better performance in the rate distortion sense, or in coding efficiency (at a fixed distortion), with increasing leaky factor. To obtain the same amount of distortion, a larger leaky factor requires less data rate than the smaller ones. For example, to obtain the distortion $D_e^{\theta,\tilde{\theta}} = 35.14$ dB as shown in

the figure, LPLC has a gain of approximately 0.12 bits/pixel in rate whenever α increases by 0.1. It is interesting to note that when the enhancement layer MCP rate is small, it might be possible that a larger leaky factor yields a less efficient codec, especially when the leaky factor is close to 1. In our derivations, this arises because the PSD in (26) can be rewritten as

$$\Phi_{\psi\psi}^{I,\tilde{\theta}}(\Lambda) = (\Phi_{\tilde{e}_b\tilde{e}_b}(\Lambda) - \tilde{\theta})\left[1 - \alpha(2-\alpha)|P(\Lambda)|^2\right] + \tilde{\theta}\left[1 - 2\alpha(1-\alpha)|P(\Lambda)|^2\right].$$

For a fixed Λ, the first term of $\Phi_{\psi\psi}^{I,\tilde{\theta}}(\Lambda)$ achieves its minimum with respect to α when $\alpha = 1$ while the second term achieves its minimum when $\alpha = \frac{1}{2}$. We believe this conforms with the operational results in [7].

Results of *Scenario II for LPLC* are shown by dotted lines in Figure 9, where the enhancement layer suffers from data rate truncation. We fix $\breve{\theta}_0$ so that the distortion in SNR is at 35.14dB, and vary $\tilde{\theta}$ between this value and θ in (37) and (38). It is observed that larger leaky factors yield a larger drop in the rate distortion performance when drift occurs in the enhancement layer, which conforms well with the published operational results. In our closed form expressions, the term $\frac{1}{1-\alpha^2|P(\Lambda)|^2}$ in (37) stands for the effect of error propagation when drift occurs. The larger α, the larger decrease in fidelity as a result of the amplification of the drift by this term, implying poor error resilience performance. Notice that a leaky factor of 0.5 is a good choice in balancing between error resilience performance and coding efficiency.

REFERENCES

1. W. Li, "Overview of fine granularity scalability in mpeg-4 video standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, pp. 301–317, Mar. 2001.
2. F. Wu, S. Li, and Y.-Q. Zhang, "A framework for efficient fine granularity scalable video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, pp. 332–344, Mar. 2001.
3. K. Shen and E. J. Delp, "Wavelet based rate scalable video compression," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, pp. 109–122, Feb. 1999.
4. H. Huang, C. Wang, and T. Chiang, "A robust fine granularity scalability using trellis-based predictive leak," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, pp. 372–385, June 2002.
5. S. Han and B. Girod, "Robust and efficient scalable video coding with leaky prediction," in *Proceedings of IEEE International Conference on Image Processing*, vol. 2, Rochester, NY, Sept. 22–25, 2002, pp. 41–44.
6. W.-H. Peng and Y.-K. Chen, "Error drifting reduction in enhanced fine granularity scalability," in *Proceedings of IEEE International Conference on Image Processing*, vol. 2, Rochester, NY, Sept. 22–25, 2002, pp. 61–64.
7. Y. Liu, Z. Li, P. Salama, and E. J. Delp, "A discussion of leaky prediction based scalable coding," in *Proceedings of IEEE International Conference on Multimedia and Expo*, vol. 2, Baltimore, MD, July 6–9, 2003, pp. 565–568.
8. Y. Wang and S. Lin, "Error-resilient video coding using multiple description motion compensation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, pp. 438–452, June 2002.
9. X. Tang and A. Zakhor, "Matching pursuits multiple description coding for wireless video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, pp. 566–575, June 2002.
10. Y. Liu, P. Salama, and E. J. Delp, "Multiple description scalable coding for error resilient video transmission over packet networks," in *Proceedings of the SPIE International Conference on Image and Video Communications and Processing*, vol. 5022, Santa Clara, CA, Jan.20–24, 2003, pp. 157–168.
11. B. Girod, "The efficiency of motion-compensation prediction for hybrid coding of video sequences," *IEEE J. Select. Areas Commun.*, vol. SAC-5, pp. 1140–1154, Aug. 1987.
12. T. Berger, *Rate distortion theory: a mathematical basis for data compression.* Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1971.
13. G. W. Cook, "A study of scalability in video compression: Rate-distortion analysis and parallel implementation," Ph.D. Thesis, School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN, Dec. 2002.
14. M. S. Pinsker, *Information and information stability of random variables and processes.* San Francisco: Holden-Day, 1964.