FGS Coding Using Cycle-Based Leaky Prediction Through Multiple Leaky Factors

Xiangyang Ji, Yanyan Zheng, Debin Zhao, Feng Wu, *Senior Member, IEEE*, and Wen Gao, *Senior Member, IEEE*

***Abstract*—This paper proposes a fine granularity scalable (FGS) coding using cycle-based leaky prediction, in which the multiple leaky factors are used to yield enhancement layer prediction to make a good compromise between coding efficiency and drift error. In this proposed method, first, the error propagation for leaky prediction with two leaky factors is theoretically analyzed in case only the base-layer bitstream and part of the enhancement-layer bitstream are available at the decoder. Based on this analysis, in this paper, we investigate how to effectively introduce enhancement-layer information into the prediction loop for enhancement-layer coding by the proper leaky factors to constrain drift error while keeping high coding efficiency. Furthermore, a coefficient scaling approach in the transform domain is proposed to address the decoding complexity issue for multiple reconstructions of partial enhancement layers at different quality levels. Finally, an encoder optimization approach is presented to further control drift error for multiple FGS layers coding. The experimental results show that compared to AR-FGS in JSVM, the proposed method can significantly improve the coding performance over a wide range of bitrates.**

***Index Terms*—Drift error, fine granularity scalability (FGS), leaky prediction, video coding.**

I. INTRODUCTION

A SERIES of nonscalable hybrid video coders such as MPEG-x and H.26x has been successfully developed. The latest video coding standard H.264/AVC [1] jointly developed by ISO and ITU provides more than 50% bitrate saving against MPEG-2 video coding standard [2] at the same quality and also contains many network-friendly features. However, the generated nonscalable bitstream cannot adapt to video transmission well on time-varying networks including wireless and cable channels or terminal devices with different processing capabilities regarding available memory, computation power and accessible bandwidth etc. Scalable video coding is more effective to address these situations since its generated bitstream is decodable at different bit-rates which may have different quality levels, temporal or spatial resolutions etc.

In the early stage of scalable video coding, to meet a bandwidth-varied network, layer coding approaches such as in MPEG-2 [2] and H.263 [3] are used to generate SNR scalable bitstream. However, it is only allowed to be decoded at several prefixed bitrate points. Later, the bitplane-based FGS coding is introduced to MPEG-4 [4], [5]. Compressed video bitstream yielded by MPEG-4 FGS can be truncated at any bitrate point on top of the base layer on-the-fly and thus effectively adapt to the fluctuation of network bandwidth. However, due to a lack of enhancement-layer prediction of the adjacent frames, this technique usually leads to a big loss of coding efficiency for enhancement layer coding against single layer coding [6].

To improve the coding efficiency of enhancement layer, a progressive fine granularity scalable (PFGS) framework was presented in [7]. In PFGS coding, motion compensation loop is still used for base layer with the nonscalable coding approach like MPEG-4 FGS. However, an additional motion-compensation loop for enhancement-layer coding is introduced, in which partial enhancement layer can be introduced to enhancement layer prediction loop. But when the received bits of enhancement layer are not enough to reconstruct the desired one used as a reference into enhancement layer prediction loop at decoder, the prediction signal mismatch between encoder and decoder will occur and result in drift error. A macroblock-based PFGS (MBPFGS) coding is proposed [8] to further effectively control drift error. In addition, for adaptive motion-compensated fine granularity scalability (AMC-FGS) proposed in [9], a partial enhancement layer can be directly introduced to the base-layer prediction loop to effectively remove the temporal redundancy for the base layer coding. If the desired bitstream for reconstructing the partial enhancement layer is not completely received, severe drift error will also occur at base layer due to the use of different references at encoder and decoder. Furthermore, a leaky prediction-based robust FGS coding (RFGS) is proposed in [10], in which the reference of enhancement layer is generated by the weighted combination of previous partial enhancement layer and base layer with a given leaky factor. By properly selecting the leaky factor, the errors in the decoded bitstream at the enhancement layer will be quickly attenuated after several iterations in a few subsequent frames. An improved leaky prediction-based FGS coding with an adaptive leaky factor is

---

Manuscript received April 17, 2007; revised September 15, 2007. First published April 30, 2008; current version published October 8, 2008. This work was supported by the National Science Foundation of China under Grant 60736043. This paper was recommended by Associate Editor J. F. Arnold.

X. Ji was with the Institute of Computing Technology, Chinese Academy of Sciences, and the Graduate School of Chinese Academy of Science, Beijing 100080, China. He is now with the Broadband Networks and Digital Media Laboratory, Automation Department, Tsinghua University, Beijing 100084, China (e-mail: xyji@jdl.ac.cn; xyji@mail.tsinghua.edu.cn).

Y. Zheng is with the Institute of Computing Technology, Chinese Academy of Sciences, and the Graduate School of Chinese Academy of Science, Beijing 100080, China (e-mail: yyzheng@jdl.ac.cn).

D. Zhao is with the Department of Computer Science, Harbin Institute of Technology, Harbin 150001, China (e-mail: dbzhao@vilab.hit.edu.cn).

F. Wu is with Microsoft Research Asia, Beijing 100080, China (e-mail: fengwu@microsoft.com).

W. Gao is with the School of Electronic Engineering and Computer Science, Peking University, Beijing 100080, China (e-mail: wgao@jdl.ac.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCSVT.2008.924104

also proposed to more effectively control the drift error by flexibly determining different leaky factors for different bitplanes or bitrates [11]. However, the scaled enhancement-layer information introduced into the motion-compensation prediction loop usually degrades the prediction signal quality of the enhancement layer, and thus the coding efficiency improvement is still unsatisfactory.

Compared with MPEG-4 FGS coding, the aforementioned approaches usually degrade coding efficiency at the bitrate points close to the base layer. This is due to the drift error from the severe mismatch of enhancement-layer predictions at the encoder and decoder and inefficiently reusing the reconstructed residues from the base layer. In fact, it is more effective to directly predict those coefficients at the enhancement layer from the base layer when they have nonzero coefficients at the spatially corresponding positions of the base layer [12]. Based on this observation, an improved FGS coding with adaptive reference (AR-FGS) [13] was proposed in the Scalable Extension of H.264/AVC standard [14] for the low delay applications. It should be pointed out that FGS coding approach has been removed from the SVC amendment that has been finalized in July 2007 due to its complexity issue and unclear application requirements. It may be included in SVC phase II.

To make a better compromise between coding efficiency and drift error according to the theoretical error propagation analysis, this paper proposes an efficient FGS coding using cycle based leaky prediction through multiple leaky factors. Some of our preliminary works on the leaky prediction with two leaky factors were first reported in [15] and [16]. In this paper, the enhancement-layer prediction is further yielded by combining the base layer and multiple segments of the enhancement layer with proper leaky factors. Each segment corresponds to one part of the enhancement-layer bitstream. Furthermore, to address the decoding complexity issue for multiple reconstructions of partial enhancement layers at different quality levels of enhancement layer, the cycle-based enhancement-layer coefficients scaling in the transform domain is proposed. In addition, an encoder optimization approach to improve drift error control capability is also presented for multiple FGS layers coding.

The remainder of this paper is organized as follows. Section II gives a brief overview of FGS layer concept and AR-FGS in the Scalable Extension of H.264/AVC standard. Section III depicts the cycle-based leaky prediction through multiple leaky factors for FGS coding in detail, which is based on the error propagation analysis for leaky prediction when only the base-layer bitstream and part of the enhancement-layer bitstream are available at the decoder due to truncation operation. Section IV provides the experimental results in terms of rate distortion performance of the proposed method over a wide range of bitrates. Finally, some conclusions are given in Section V.

II. AR-FGS IN BRIEF

In FGS coding like the aforementioned PFGS [7] and RFGS [10], the predicted residual e_t^E of the enhancement layer in the frame F_t at time t is formed by

$$e_t^E = F_t - P_t^E \tag{1}$$

where P_t^E is the prediction signal for the enhancement layer. In RFGS, P_t^E can be calculated by

$$\begin{aligned} P_t^E &= \alpha \left(\tilde{F}_{t-1}^P\right)_{MC} + (1-\alpha)\left(\tilde{F}_{t-1}^B\right)_{MC} + \tilde{e}_t^B \\ &= \alpha \left(\tilde{F}_{t-1}^P - \tilde{F}_{t-1}^B\right)_{MC} + \tilde{F}_t^B \\ &= \alpha \left(\tilde{F}_{t-1}^{P-B}\right)_{MC} + \tilde{F}_t^B. \end{aligned} \tag{2}$$

Here, the subscript MC denotes motion compensation operation and the hat “~” represents the reconstructed signal at encoder. $\tilde{e}_t^B$ represents the reconstructed residual signal at the base layer. $\tilde{F}_{t-1}^B$ is the reconstructed base layer and $\tilde{F}_{t-1}^P$ is the reconstructed frame from the base-layer bitstream and partial enhancement-layer bitstream at time $t-1$. The differential signal $\tilde{F}_{t-1}^{P-B}$ is the partial data of enhancement layer at time $t-1$. α is a leaky factor $(0 \le \alpha \le 1)$ used to scale this differential signal $\tilde{F}_{t-1}^{E-B}$ so that the potential drift error caused by the partial enhancement-layer information loss can be attuned iteratively in the subsequent frames. Therefore, when α is zero, this equation represents FGS in MPEG-4.

In the AR-FGS coding [13], the whole reconstructed enhancement layer $\tilde{F}_{t-1}^{E-B}$ will be introduced into enhancement layer prediction loop. P_t^E can be calculated by

$$\begin{aligned} P_t^E &= \alpha \left(\tilde{F}_{t-1}^E - \tilde{F}_{t-1}^B\right)_{MC'} + \tilde{F}_t^B \\ &= \alpha \left(\tilde{F}_{t-1}^{E-B}\right)_{MC'} + \tilde{F}_t^B \end{aligned} \tag{3}$$

where MC' represents the general motion compensation operation. For each current enhancement-layer inter-block, MC' corresponds to the normal motion-compensation operation MC if all transform coefficients in the collocated base-layer block are zero. Otherwise, every transform coefficient of the motion compensated prediction block from a previously reconstructed enhancement layer $\tilde{F}_{t-1}^{E-B}$ will be forced to zero if the reconstructed residual transform coefficient at its collocated base-layer position is not equal to zero, and then the inverse transform is performed on these adjusted transform coefficients to form a new motion-compensated prediction block. It should be noted that, in (3), different values of the leaky factor α could be used for the enhancement-layer inter-block coding according to whether it has nonzero coefficients at the collocated base layer block. Please refer to [13] for more details. For simplicity of our following discussions, we assume they are the same.

On the other hand, fine granular SNR scalability is supported by the utilization of progressive refinement quantization and the associated entropy coding in the Scalable Extension of H.264/AVC standard. Each FGS layer is a refinement signal relative to the preceding layer, and this refinement signal corresponds to a bisection of the quantization step size, namely decrease of 6 in terms of quantization parameter. Consequently, fine granular SNR scalability can be naturally supported by discarding the refinement information at the highest quality layer of bitstream [14]. In this paper, the reconstructed enhancement layer is comprised of all FGS layers. The partial enhancement layer represents the partially reconstructed enhancement layer

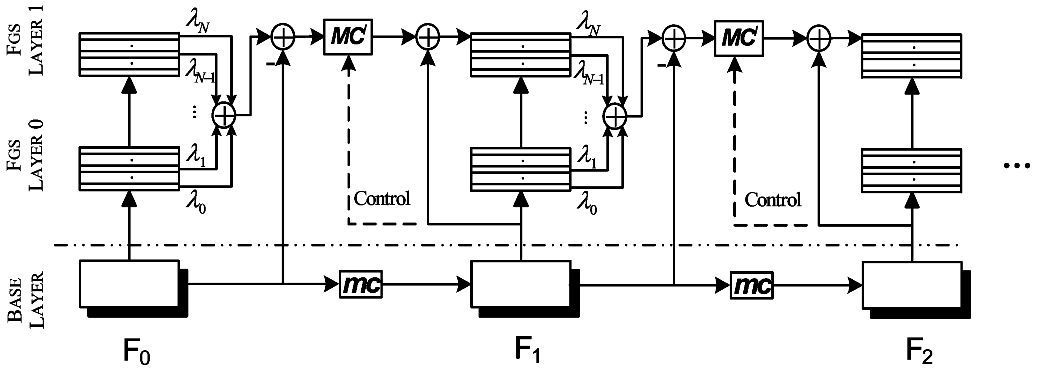


Fig. 1. Block diagram of the leaky prediction utilizing the weighted combination of multiple reconstructed frame signals from the base-layer bitstream and part or all of the enhancement-layer bitstream when two FGS layers are appended on top of the base layer.

at a certain quality level of enhancement layer, which corresponds to the first coded enhancement-layer bitstream with a certain amount of bits.

III. CYCLE-BASED LEAKY PREDICTION FOR FGS CODING

A. *Multiple Leaky Factor-Based Leaky Prediction*

In AR-FGS coding, the determination of leaky factor α value, which is used to scale the reconstructed enhancement layer $\tilde{F}^{E-B}$, should be able to efficiently make a compromise between coding efficiency and drift error. By introducing more information of $\tilde{F}^{E-B}$ with a relatively large value of leaky factor α, better enhancement-layer prediction can be achieved. However, the decoded video usually suffers from more serious drift error when only part of enhancement layer bitstream is received. Therefore, a proper leaky factor value should be selected to yield a good rate distortion performance compromise over all potential truncated bitrates. As suggested in [7], [9], and [10], another effective method to further control drift error is to only introduce the partial enhancement layer $\tilde{F}^{E-B}$, as used in (2), into an enhancement-layer prediction loop. However, such an approach usually degrades the enhancement-layer prediction quality and results in poor coding efficiency.

To get out of this dilemma, the weighted combination of multiple different quality-level frame signals from the base layer bitstream and part or all of the enhancement-layer bitstream can be introduced to form enhancement layer prediction as follows:

$$\begin{aligned} P_t^E &= \alpha \left(\sum_{n=0}^{N} \lambda_n \tilde{F}_{t-1}^{P_n} - \tilde{F}_{t-1}^{B} \right)_{MC'} + \tilde{F}_t^B \\ &= \alpha \left(\sum_{n=0}^{N} \lambda_n \left(\tilde{F}_{t-1}^{P_n} - \tilde{F}_{t-1}^{B} \right) \right)_{MC'} + \tilde{F}_t^B \\ &= \alpha \left(\sum_{n=0}^{N} \lambda_n \tilde{F}_{t-1}^{P_n - B} \right)_{MC'} + \tilde{F}_t^B \end{aligned} \tag{4}$$

with

$$\lambda_n \leq 1(\lambda_0 + \lambda_1 + \cdots + \lambda_N) = 1.$$

Here, there are $N+1$ different quality-level reconstruction frame signals from the base-layer bitstream and part of enhancement-layer bitstream and $\tilde{F}^{P_n}$ represents its nth reconstructed signal. Especially, $\tilde{F}^{P_n}$ corresponds to the fully reconstructed signal. Fig. 1 gives the corresponding block diagram. Therefore, this procedure introduces multiple partial enhancement layers $\tilde{F}_{t-1}^{P_n - B}$ into the enhancement-layer prediction loop with proper weighting factors, where the multiple partial enhancement layers are obtained by multiple reconstructions at different quality levels of the enhancement layer. Furthermore, (4) can be rewritten as

$$\begin{aligned} P_t^E &= \alpha \left(\sum_{n=0}^{N} \lambda_n \left(\tilde{F}_{t-1}^{P_n} - \tilde{F}_{t-1}^{B} \right) \right)_{MC'} + \tilde{F}_t^B \\ &= \alpha \left(P_{t-1}^E - \tilde{F}_{t-1}^B + \sum_{n=0}^{N} \beta_n \tilde{e}_{t-1}^{D_n} \right)_{MC'} + \tilde{F}_t^B \end{aligned} \tag{5}$$

with $\beta_n = (\lambda_n + \lambda_{n+1} + \cdots + \lambda_N), \beta_0 = (\lambda_0 + \lambda_1 + \cdots + \lambda_N) = 1$, $\tilde{e}^{D_n} = \tilde{F}^{P_n} - \tilde{F}^{P_{n-1}}$, and $\tilde{e}^{D_0} = \tilde{F}^{P_0} - \tilde{F}^B$.

Here, P_{t-1}^E is the prediction signal of the enhancement layer at time $t-1$ and $\tilde{F}_{t-1}^B$ is the reconstructed base layer at time $t-1$. Each differential signal $\tilde{e}^{D_n}$ is a segment of the enhancement layer, which corresponds to one part of the enhancement-layer bitstream S_n. All of these bitstreams $\sum_{n=0}^{N} S_n$ form the enhancement-layer bitstream. It can be noted that, for the aforementioned special case, that only the partially reconstructed enhancement layer is introduced into the prediction loop for enhancement-layer coding, in (5), only $\tilde{F}^{P_0}$ is introduced into the enhancement-layer prediction loop, and all $\{\lambda_n\}_{n>0}$ are set to zero. On the other hand, compared with the leaky prediction with adaptive leaky factors in [11], each differential signal $\tilde{e}^{D_n}$ should be scaled by the leaky factors α and β_n together. Obviously, the leaky factor $\alpha\beta_n$ should be not more than α.

B. *Error Propagation Analysis*

First, we assume that only two leaky factors are used for leaky prediction when one FGS layer is coded on top of the base layer.

Equation (5) can be expressed by recursive motion compensation procedure as follows:

$$\begin{aligned} P_t^E &= \alpha\left(P_{t-1}^E - \tilde{F}_{t-1}^B + \beta\tilde{e}_{t-1}^{D_1} + \tilde{e}_{t-1}^{D_0}\right)_{MC'} + \tilde{F}_t^B \\ &= \alpha\Bigg\{\alpha\Bigg(P_{t-2}^E - \tilde{F}_{t-2}^B + \beta\tilde{e}_{t-2}^{D_1} + \tilde{e}_{t-2}^{D_0}\Bigg)_{MC'} \\ &\quad + \beta\tilde{e}_{t-1}^{D_1} + \tilde{e}_{t-1}^{D_0}\Bigg\}_{MC'} + \tilde{F}_t^B \\ &= \alpha\left\{\alpha\left\{\alpha\left(P_{t-3}^E - \tilde{F}_{t-3}^B + \beta\tilde{e}_{t-3}^{D_1} + \tilde{e}_{t-3}^{D_0}\right)_{MC'}\right.\right. \\ &\quad \left.\left. + \beta\tilde{e}_{t-2}^{D_1} + \tilde{e}_{t-2}^{D_0}\right\}_{MC'} + \beta\tilde{e}_{t-1}^{D_1} + \tilde{e}_{t-1}^{D_0}\right\}_{MC'} + \tilde{F}_t^B \\ &= \cdots \end{aligned} \tag{6}$$

Therefore, if only the base-layer bitstream and part of the enhancement-layer bitstream are available for the frame F_{t-3} at the decoder due to bitstream truncation during transmission, the enhancement-layer prediction signal for the frame F_t can be calculated by

$$\begin{aligned} P_t^E &= \alpha\left\{\left[\alpha\left\{\left[\alpha(P_{t-3}^E - \tilde{F}_{t-3}^B + \beta\tilde{e}_{t-3}^{D_1} + \tilde{e}_{t-3}^{D_0} - \hat{e})_{MC'}\right]\right.\right.\right. \\ &\quad \left.\left.\left. + \beta\tilde{e}_{t-2}^{D_1} + \tilde{e}_{t-2}^{D_0}\right\}_{MC'}\right] + \beta\tilde{e}_{t-1}^{D_1} + \tilde{e}_{t-1}^{D_0}\right\}_{MC'} + \tilde{F}_t^B \\ &= \ldots \end{aligned} \tag{7}$$

with

$$\hat{e} = \begin{cases} \beta\Delta e, & \Delta e \in \tilde{e}_{t-3}^{D_1} \\ \beta\tilde{e}_{t-3}^{D_1} + (\Delta e - \tilde{e}_{t-3}^{D_1}), & \Delta e \in \tilde{e}_{t-3}^{D_0} \end{cases} \tag{8}$$

Here, the hat “ˆ” represents the reconstructed signal at decoder and Δe represents the reconstruction error between the fully reconstructed frame $\tilde{F}_{t-3}^E$ at encoder and the actually decoded frame $\hat{F}_{t-3}^E$ at decoder. Hence, the resultant error in the frame F_t is $\alpha^3\hat{e}$. For the special case with $\beta = 1$, namely using only one leaky factor, the resultant error in the frame F_t is $\alpha^3\Delta e$. As a result, if there is an error Δe in the frame F_t, its resultant overall mean square error (MSE) distortion for all affected subsequent frames can be calculated as

$$\begin{aligned} \mathrm{MSE}_{\mathrm{Drift}}^1 &= (1 + \alpha^2 + \cdots + \alpha^{2K})\zeta(\Delta e) \\ &= (1 + f(\alpha))\,\zeta(\Delta e) \end{aligned} \tag{9}$$

with

$$\zeta(\Delta e) = \frac{1}{UV}\sum_{u=0}^{U-1}\sum_{v=0}^{V-1}\left(\tilde{F}_t^E(u,v) - \hat{F}_t^E(u,v)\right)^2 \tag{10}$$

$$f(\alpha) = \frac{\alpha^2 - (\alpha^2)^{K+1}}{(1 - \alpha^2)}\alpha \in [0, 1). \tag{11}$$

Here, U and V represent the height and width of the image frame, respectively. The superscript of MSE represents the number of leaky factors. $f(\alpha)$ is a monotonically increasing function and K is the number of all inclusive subsequent frames affected by error propagation in prediction loop. Correspondingly, when the leaky prediction with two leaky factors are used for the enhancement-layer coding and there is an error Δe in a frame F_t, its resultant overall MSE distortion for itself and all affected subsequent frames can be calculated as shown in (12) at the bottom of the page.

For the bitplane coding used in MPEG-4 FGS, every bit within each bitplane has the same impact on distortion since it is coded as a binary number for a certain value. With this type of coding approach, it is reasonable to think the relationship between the number of bits and the corresponding distortion is linear within each bitplane [17]. It is also further verified in [18] that the rate-distortion $(R - D)$ characteristic of RFGS within each bitplane is more accurate to be modeled as a linear function than an exponential function. Since the enhancement layer with progressive refinement quantization in the Scalable Extension of H.264/AVC standard is coded similar to bitplane coding, the $R - D$ characteristic of each FGS layer can also be modeled as a linear function. If all bitstream to code the predicted residual signal e_t^E in one FGS layer can be divided into N segments with the same bitrate, according to the linear relationship of $R - D$ characteristic, MSE distortion $\zeta(\Delta e_{seg})$ caused by the loss of each segment also should be equivalent, and thus we have

$$N\zeta(\Delta e_{\mathrm{seg}}) = \frac{1}{UV}\sum_{u=0}^{U-1}\sum_{v=0}^{V-1}\left(\tilde{F}^E(u,v) - \tilde{F}^B(u,v)\right)^2. \tag{13}$$

If each of $N + 1$ potential truncation points, including a base layer point, has the truncation probability p_n, the average MSE distortion caused by the different truncation operations of bitstream for a certain frame is given by (14), shown at the bottom of the next page. Here, κ corresponds to the partial reconstruction point of the enhancement layer in (5) when the leaky prediction with two leaky factors is used. Therefore, according to (12) and (14), overall MSE distortion for all affected frames from $N + 1$ different truncation points in a certain frame is

$$\mathrm{MSE}_{F\_\mathrm{Drift}}^2 = (1 + \beta^2 f(\alpha))\mathrm{MSE}_F^{D_1} + (1 + f(\alpha))\mathrm{MSE}_F^{D_0}. \tag{15}$$

As a result, we are faced with the dilemma that smaller leaky factor β value for a given α in (15) provides better drift error control but usually degrades enhancement-layer coding

---

$$\mathrm{MSE}_{\mathrm{Drift}}^2 = \begin{cases} (1 + \beta^2 f(\alpha))\zeta(\Delta e), & \Delta e \in \tilde{e}_t^{D_1} \\ (1 + \beta^2 f(\alpha))\,\zeta(\tilde{e}_t^{D_1}) + (1 + f(\alpha))\,\zeta(\Delta e - \tilde{e}_t^{D_1}), & \Delta e \in \tilde{e}_t^{D_0} \end{cases} \tag{12}$$

efficiency. In Section II-C, to tackle this dilemma, how to effectively generate the reconstructed signal from part of the enhancement-layer bitstream and properly select a pair of leaky factors (α, β) is investigated to make a good compromise between coding efficiency and drift error.

C. Cycle-Based Partial Enhancement-Layer Reconstruction

For FGS coding like in MPEG-4 [5], the coefficients for the enhancement layer are coded macroblock by macroblock in progressive scan order and the coefficients of one macroblock have to be totally coded before coding the coefficients of next macroblock within each FGS layer. As a result, the quality of the reconstructed signal from the truncated bitstream is not uniform in the whole frame [17]. Fig. 2(a) shows the average MSE distortion of the predicted residual signal for enhancement layer in each frame, which is achieved by the selection of different leaky factor β values for each given leaky factor α when only half of one FGS layer bitstream is used to yield $\tilde{e}^{D_0}$. Here, the leaky factors α and β belong to $\{0, 1/32, \ldots, 31/32, 32/32\}$. It can be observed that the MSE distortion of the predicted residual signal for the enhancement layer is monotonically decreasing with β for a given α, namely, the leaky factor β with larger value is always able to yield better quality prediction for the enhancement-layer coding. However, this is contrary to the statement that a smaller leaky factor β is able to provide better drift error control capability. Consequently, it is difficult to make a good compromise between coding efficiency and drift error.

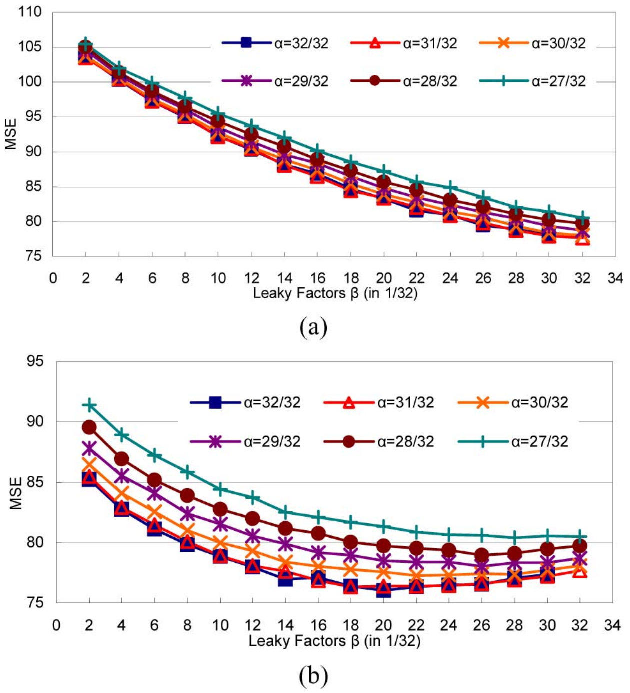


Fig. 2. MSE distortion of the predicted residual for the enhancement layer when (a) MPEG-4 like FGS coding and (b) cyclic block FGS coding are used for *Mobile* sequence in CIF@15 Hz.

As we know, the bitstream to form $\tilde{e}^{D_0}$ for leaky prediction corresponds to the first coded one of the enhancement-layer bitstream with a certain amount of bits and, thus, has higher priority to be correctly received than the remaining bitstream. Consequently, it usually results in drift error with less possibility. Considering that high-frequency information usually plays a lesser role in improving the prediction signal quality than low-frequency information does, an ideal solution is that this part of the bitstream should contain relatively more low-frequency information. Therefore, high-frequency coefficients should be later coded within each FGS layer. This desire can be well fulfilled by employing the cyclical block FGS coding [14], [19] in the Scalable Extension of H.264/AVC standard. Cyclical block FGS coding involves a number of coding "cycles." In each cycle, every block in a frame is processed sequentially. For a given block, a group of coefficients scanned in zigzag order, which contains all zero coefficients and the first encountered nonzero coefficient, is encoded in each cycle and, thus, usually low-frequency transform coefficients among all coefficients of one block are coded with the high priority in each cycle [19]. Using cyclical block FGS coding, for each different leaky factor α value, the MSE distortion of the predicted residual signal is no longer monotonically decreasing with β when only half of one FGS layer bitstream is used to yield $\tilde{e}^{D_0}$ for leaky prediction. As shown in Fig. 2(b), for a given leaky factor α, the leaky factor β with relatively less value is able to yield comparable or even better prediction quality over that with larger value for the currently encoding enhancement layer. Hence, it is possible to offer better drift error control capability

$$\begin{aligned}
\mathrm{MSE}_F &= \sum_{n=0}^{N} p_n\left[(N-n)\zeta(\Delta e_{\mathrm{seg}})\right] \\
&= \sum_{n=\kappa}^{N} \left(p_n\left[(N-n)\zeta(\Delta e_{\mathrm{seg}})\right]\right) + \sum_{n=0}^{\kappa-1} \left(p_n\left[(N-n)\zeta(\Delta e_{\mathrm{seg}})\right]\right) \\
&= \underbrace{\sum_{n=\kappa}^{N} \left(p_n\left[(N-n)\zeta(\Delta e_{\mathrm{seg}})\right]\right) + \sum_{n=0}^{\kappa-1} \left(p_n\left[(N-\kappa)\zeta(\Delta e_{\mathrm{seg}})\right]\right)}_{\Delta e \in \tilde{e}_t^{D_1}} + \underbrace{\sum_{n=0}^{\kappa-1} \left(p_n\left[(\kappa-n)\zeta(\Delta e_{\mathrm{seg}})\right]\right)}_{\Delta e \in \tilde{e}_t^{D_0}} \\
&= \mathrm{MSE}_F^{D_1} + \mathrm{MSE}_F^{D_0}
\end{aligned} \tag{14}$$

while still keeping high coding efficiency for the enhancement layer.

Usually, it is difficult to verify whether the leaky prediction with two leaky factors can actually provide better drift error control capability than that with only one leaky factor. It is ideal to compare the drift error control capability if the same bitrate is needed for the segments $\Delta e_{\mathrm{seg}}^1$ and $\Delta e_{\mathrm{seg}}^2$ from these two methods and the resultant MSE distortions $\zeta\left(\Delta e_{\mathrm{seg}}^1\right)$ and $\zeta\left(\Delta e_{\mathrm{seg}}^2\right)$ due to their losses are also equivalent within a certain FGS layer. According to the linear $R-D$ characteristic within each FGS layer, it can be satisfied when there are the same reconstruction MSE distortions calculated by (13), namely $N\zeta\left(\Delta e_{\mathrm{seg}}^1\right) = N\zeta\left(\Delta e_{\mathrm{seg}}^2\right)$ and the same bitrates of the enhancement layer for one FGS layer coding of the predicted residual signals $e_{t,1}^E$ and $e_{t,2}^E$, which are achieved by one leaky factor and two leaky factors, respectively. For the cycle-based partial enhancement-layer reconstruction, Fig. 3(a) and (b) shows that the bitrates of the enhancement layers and their reconstructed MSE distortions frame by frame for the leaky predictions with one leaky factor ($\alpha = 30/32$ when $\beta = 32/32$) and two leaky factors ($\alpha = 30/32$ and $\beta = 18/32$), respectively for test sequence *Mobile* in CIF@15 Hz. In this testing, only one FGS layer is appended on top of base layer and approximately half of one FGS layer bitstream is used to yield $\tilde{e}^{D_0}$. It can be observed that when they have almost the same MSE distortion for the reconstructed enhancement layers in a frame, their bitrates are also roughly the same. Consequently, to more effectively control drift error by the leaky prediction with two leaky factors, according to (14) and (15), the following inequality should be satisfied:

$$\begin{aligned}\mathrm{MSE}_{F\_\mathrm{Drift}}^1 - \mathrm{MSE}_{F\_\mathrm{Drift}}^2 =&(1+f(\alpha_1))\mathrm{MSE}_F^1\\&-(1+f(\alpha_2))MSE_F^{D_0,2}\\&-(1+\beta^2 f(\alpha_2))\mathrm{MSE}_F^{D_1,2}\\&>0.\end{aligned}\tag{16}$$

Thus, we have (17), shown at the bottom of the page.

Furthermore, when truncation at each bitrate point has the same probability, (14) can be rewritten as

$$\mathrm{MSE}_F = \underbrace{\frac{N^2}{8}\zeta(\Delta e_{\mathrm{seg}}) + \frac{N^2}{4}\zeta(\Delta e_{\mathrm{seg}})}_{\Delta e \in e_t^{D_1}} + \underbrace{\frac{N^2}{8}\zeta(\Delta e_{\mathrm{seg}})}_{\Delta e \in e_t^{D_1}} \tag{18}$$

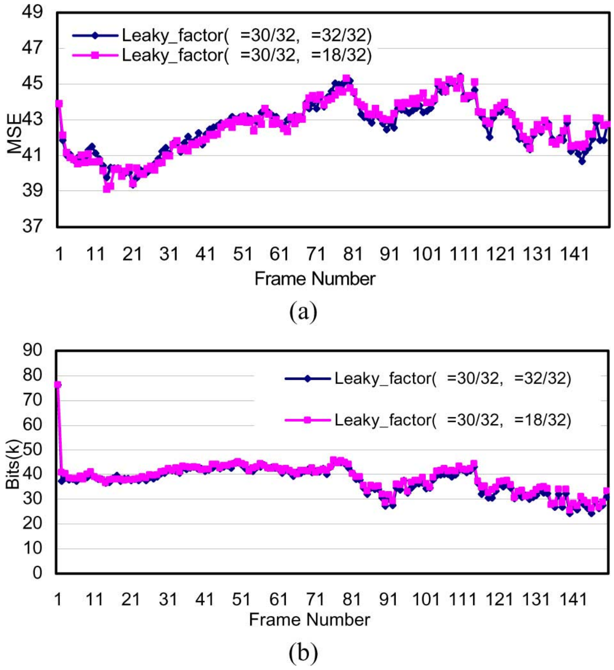


Fig. 3. (a) MSE distortion and (b) bitrate comparisons frame by frame when one leaky factor and two leaky factors are respectively used for leaky prediction for *Mobile* sequence in CIF@15 Hz.

and (17) can be simplified to

$$\begin{aligned}\beta &< \sqrt{\frac{\frac{N^2}{2}f(\alpha_1)\zeta(\Delta e_{\mathrm{seg}}) - \frac{N^2}{8}f(\alpha_2)\zeta(\Delta e_{\mathrm{seg}})}{\frac{3N^2}{8}f(\alpha_2)\zeta(\Delta e_{\mathrm{seg}})}}\\&=\sqrt{\frac{4\times f(\alpha_1) - f(\alpha_2)}{3\times f(\alpha_2)}}.\end{aligned}\tag{19}$$

Here, if α_2 is equal to α_1, the leaky prediction with two leaky factors definitely has better drift control capability than that with one leaky factor as long as leaky factor β is less than unity, when their predicted residual signals for enhancement layer coding have the same MSE distortion.

D. Coefficient Scaling in the Transform Domain

As discussed in the previous subsection, compared with the case using only one leaky factor, the leaky prediction with two leaky factors may provide better drift error control capability

$$\begin{aligned}\beta &< \sqrt{\frac{(1+f(\alpha_1))\mathrm{MSE}_F^1 - (1+f(\alpha_2))\mathrm{MSE}_F^{D_0,2} - \mathrm{MSE}_F^{D_1,2}}{f(\alpha_2)\mathrm{MSE}_F^{D_1,2}}}\\&=\sqrt{\frac{f(\alpha_1)\mathrm{MSE}_F^1 - f(\alpha_2)\mathrm{MSE}_F^{D_0,2}}{f(\alpha_2)\mathrm{MSE}_F^{D_1,2}}}\end{aligned}\tag{17}$$

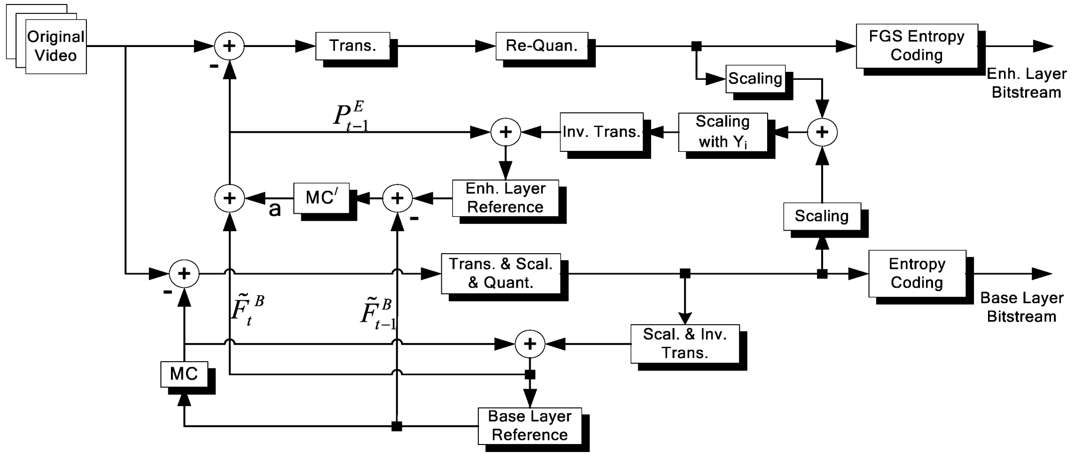


Fig. 4. Diagram of FGS encoder with coefficient scaling in the transform domain.

while keeping high coding efficiency. Furthermore, it is possible to more effectively control drift error by the leaky prediction with more leaky factors. However, the use of more leaky factors usually requires multiple partial enhancement-layer reconstructions. To avoid high complexity of multiple reconstruction operations, (5) can be implemented by coefficient scaling in transform domain as follows:

$$P_t^E = \alpha\left(P_{t-1}^E + \mathrm{I}\Gamma\left(\sum_{m=1}^{M}\gamma_{t-1}^m \otimes \tilde{\Psi}_{t-1}^m\right) - \tilde{F}_{t-1}^B\right)_{MC'} + \tilde{F}_t^B \tag{20}$$

Here, $\mathrm{I}\Gamma$ represents the inverse transform operation and the operator $\otimes$ denotes element-by-element multiplication instead of normal matrix multiplication. M is the total number of FGS layers appended on top of the base layer and the superscript m is used to represent the FGS layer number. $\tilde{\Psi}$ is the de-quantized transform coefficients matrix at encoder and γ is the leaky factors matrix, in which every leaky factor $\gamma(j,i)$ with the row and column indexes j, i in the same cycle has the same value. It should be noted that the leaky factor values of different cycles should be monotonically nondecreasing with cycle index within one FGS layer since the bitstream belonging to low cycles has higher priority to be correctly received and thus can make a good compromise between drift error and coding efficiency. However, different from the leaky factors selection in [11], the leaky factors for different FGS layers are no longer always monotonically nondecreasing with FGS layer level m in multiple FGS layers coding. To make a good compromise between effective drift error control and high coding efficiency, the leaky factors at low cycles usually including low-frequency information in one FGS layer should be allowed to take larger values than the ones for higher cycles usually including high-frequency information in lower-level FGS layer. Fig. 4 illustrates the diagram of the FGS encoder with coefficient scaling in transform domain. The enhancement layer prediction generation process as depicted in (20) can be clearly illustrated in Fig. 4.

At the decoder, the corresponding prediction signal for the enhancement layer is yielded by

$$\begin{aligned}\hat{P}_t^E &= \alpha\left(\hat{P}_{t-1}^E + \mathrm{I}\Gamma\left(\sum_{m=1}^{M}\gamma_{t-1}^m \otimes \hat{\Psi}_{t-1}^m\right) - \tilde{F}_{t-1}^B\right)_{MC'} \\ &\quad + \tilde{F}_t^B \\ &= \alpha\left(\hat{F}_{t-1}^{E'} - \tilde{F}_{t-1}^B\right)_{MC'} + \tilde{F}_t^B.\end{aligned} \tag{21}$$

Here, $\hat{\Psi}$ represents the transform coefficients decoded from the received enhancement-layer bitstream at decoder. According to (5), those leaky factors of the matrix γ, which are applied to the transform coefficients $\hat{\Psi}$ to yield the first reconstruction part $\tilde{F}^{P_0}$ of the enhancement layer, are all equal to unity and thus no transform coefficients are scaled. As a result, if the received FGS layer bitstream for a certain frame is not enough or is just enough to reconstruct $\tilde{F}^{P_0}$, $\hat{F}^{E'}$ can be directly used as the frame $\hat{F}^E$ to be decoded for displaying. Otherwise, for each specific inter-block whose bitstream exceeds that to reconstruct $\tilde{F}^{P_0}$, two de-quantization and inverse transform operations are required to form $\hat{F}^{E'}$ and $\hat{F}^E$. As stated in [20], compared with other decoding modules like motion compensation and deblocking filtering, all operations including the de-quantization and inverse transform operations only account for 13% overall decoding time in H.264/AVC baseline profile decoding. In addition, compared with single-layer decoding, twice motion-compensated prediction loops are required for enhancement-layer decoding in AR-FGS. Therefore, the increased decoding complexity from an extra inverse transform operation is trivial.

On the other hand, at low bitrates, the coarse quantization of the transform coefficients possibly causes visually disturbing discontinuities at the block boundaries. Thus, the deblocking operation plays an important role on improving objective and visual qualities of the decoded video. In the proposed leaky prediction, if the received FGS layer bitstream for a certain frame is not enough to reconstruct $\tilde{F}^{P_0}$, the deblocking may be used for reconstructing $\hat{F}^E$. For such a bitstream decoding, no any additional decoding complexity is required compared with AR-FGS since $\hat{F}^{E\prime}$ is the same as $\hat{F}^E$. If the received FGS layer bitstream exceeds that to reconstruct $\tilde{F}^{P_0}$, the decoded frame $\hat{F}^{E}$ can no longer be used as the desired frame $\hat{F}^{E\prime}$ introduced into prediction loop and, thus, need an additional deblocking operation. However this deblocking operation has been turned into a post filter instead of an in-loop filter since it does not affect the prediction in (21). In this situation, the deblocking operation for $\hat{F}^{E}$ usually plays a lesser role since the bitstream is at relatively high bitrate, which is over that to reconstruct $\tilde{F}^{P_0}$. Thus, it usually can be disabled. As a result, during decoding enhancement layer bitstream coded by the proposed leaky prediction, twice deblocking operations are also usually enough like in AR-FGS. Of course, such a post filter also can be turned on at decoder at any time if needed.

E. Encoder Optimization for Multiple FGS Layer Coding

To effectively control drift error, a decoder-oriented two-loop structure [21] is proposed for multiple FGS-layer coding. It still keeps the two-loop motion compensation structure at the decoder while providing performance close to multiloop structure. At the encoder, different from (20), the prediction signal for each current FGS-layer coding is yielded from the weighted combination of the signal decoded from enhancement-layer bitstream not over the current-level FGS layer and the base layer at previous time by a given leaky factor to restrain drift error. Since each FGS-layer coding only uses the previously reconstructed base-layer and partial-enhancement layer not over the same-level FGS layer, its prediction signal is usually not sufficient, although it is able to effectively confine drift error within each FGS layer. In this paper, to make a better compromise between coding efficiency and drift error at the encoder, the prediction signal for each FGS layer can also use enhancement layer information at higher-level FGS layer by weighting combination as follows:

$$\begin{aligned} P_t^{E_n} =& \alpha \Bigg(\Bigg(\sum_{m=n}^{M} \varepsilon_m \Bigg(P_{t-1}^{E_m} + \sum_{j=1}^{m} \left(\gamma_{t-1}^{j} \otimes \tilde{\Psi}_{t-1}^{j} \right) \Bigg) \Bigg) \\ & -\tilde{F}_{t-1}^{B} \Big)_{MC'} + \tilde{F}_t^{E_{n-1}} \end{aligned} \tag{22}$$

with $\varepsilon_n + \cdots + \cdots \varepsilon_M = 1$,$\varepsilon_m \leq 1$, and $\tilde{F}_t^{E_0} = \tilde{F}_t^{B}$. Here, n and m are both used to represent the FGS-layer level numbers. P^{E_n} represents the prediction signal for the nth FGS-layer coding, and $\tilde{F}^{E_n}$ represents the reconstructed frame from the base-layer bitstream and the partial enhancement-layer bitstream not over the nth FGS layer. It should be noted that this approach only belongs to encoder optimization process, and at the decoder, when decoding the enhancement layer, (21) should still be used instead.

IV. EXPERIMENTAL RESULTS

To demonstrate the improvement of the proposed method in terms of rate-distortion performance, the proposed method was integrated into JSVM_7_2, which is the reference software of the Scalable Extension of H.264/AVC standard. Test sequences are comprised of *bus*, *mobile* in CIF@15 Hz and *city*, *harbour* in 4CIF@30 Hz. For each test sequence, the first frame is encoded as intra-frame, and all of the remaining frames are encoded as inter-P-frames. The base layer is coded at QP of 38 and CABAC entropy coder is used. Deblocking is enabled at both the base layer and the enhancement layer. The reference for motion estimation is yielded by the average weight of the fully reconstructed enhancement layer and the base layer. The rate distortion performances are evaluated for both single FGS layer and two FGS layers coding.

A. Rate-Distortion Performance for a Single FGS Layer

To demonstrate the advantage of the proposed leaky prediction, first the rate-distortion performance curves for different FGS coding approaches are compared when only one FGS layer is appended on top of the base layer. Here, Anchor is used to represent AR-FGS coding approach. Its first and second parameters α_1 and α_2 are two different values of the leaky factor α. As depicted in Section II, they are employed according to whether the reconstructed residual coefficients at the collocated base layer block of each enhancement layer inter-block are all zero or not, respectively. For Anchor, the optimum leaky factors setting has been provided in [22], in which α_2 is simply fixed to 18 since it usually has little effect on performance. For the proposed leaky prediction with coefficient scaling in transform domain (CSTD_FGS), the usage of first two parameters is the same as Anchor and the following parameters set represents leaky factors in scaling matrix γ as used in (20). For simplicity of denotation, the same leaky factor value in the successive cycles will be assembled in one item, for example, (m, n) means that the same leaky factor value m is used for n successive cycles. Usually, cycle based FGS decoding is high in terms of the computation and memory requirements since decoding of each cycle has to start after previous cycle coefficients in entire frame. A cycle-aligned fragments based FGS coding [23], [24] is proposed to reduce cycle-based FGS decoding complexity, in which frame-based cycle decoding can be converted to block-based decoding with low memory access. Its typical fragment pattern is that the first cycle, the next successive threes cycles and all of the reaming cycles are included into three different fragments, respectively [24].

To demonstrate the effectiveness of the cycle-based partial enhancement layer, as illustrated in Fig. 5, performance comparison of the different reconstructed enhancement layers when different numbers of cycles of enhancement-layer coefficients are introduced to the enhancement-layer prediction loop for *Mobile* and *Bus* in CIF@15 Hz. It can be observed that the first several cycles of enhancement-layer coefficients play a important role to improve enhancement-layer prediction quality. Therefore, the previously mentioned fragment pattern is used for test because it is reasonable in terms of both coding efficiency and easy practical application. The parameters for scaling matrix γ are fixed as (1,32),(3,18),(12,12) for test sequences. It should be noted that all given leaky factors are represented in 1/32 unit. These fixed leaky factors are able to provide good prediction capability with effective drift error control. In general, to still keep good prediction quality, leaky factors α_1 and α_2 in (20) should be assigned with larger values when compared with Anchor, since

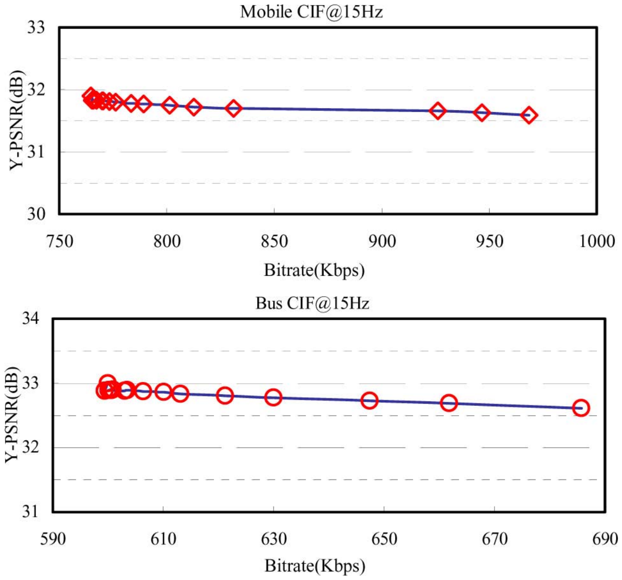


Fig. 5. Performance comparison of the reconstructed frames when different numbers of cycles of enhancement-layer coefficients are introduced to the enhancement-layer prediction loop (the cycle number is increased from the right to left) for *Mobile* and *Bus* in CIF@15 Hz.

the enhancement-layer information has been first scaled by γ. On the one hand, similar to AR_FGS, α_2 is fixed to 24 in our test; on the other hand, α_1 can be easily adjusted to yield the ideal prediction signal because it is not within motion-compensation operation.

For different test sequences, it can be observed that, in Fig. 6, CSTD_FGS obviously outperforms Anchor with optimum leaky factors over a wide range of bitrates. On the other hand, compared with Anchor—32,32 with no scaling operation, comparable coding efficiency can be achieved for the enhancement layer and significant performance gain can be observed at each truncated bitrate point. In addition, it also can be observed that CSTD_FGS without coefficients scaling, obviously suffers from severe drift error against the CSTD_FGS with γ (1,32),(3,18),(12,12) when they have the same values for the first two parameters.

In addition, if the deblocking filter is only used at the base layer and disabled at the enhancement layer in CSTD (CSTD_FGS—NoLoopFilter) at encoder and decoder, the rate-distortion performance will be obviously degraded. However, if we only disable the postfiltering operation for the enhancement-layer decoding (CSTD_FGS—NoPostFilter), the rate-distortion performance loss is slight since the in-loop filter for motion compensation is still used and the postfilter only affects the rate-distortion performance of the decoded signal of the received enhancement-layer bitstream over that to yield the first reconstruction signal $\tilde{F}^{P_0}$ for the leaky prediction in (5). We also compare FGS coding with single-layer coding in Fig. 6. It can be observed, for FGS coding, there still exists an obvious coding performance gap against single-layer coding in wide bitrate range because it usually also suffers from drift error from the truncated points and the motion field representation is not optimized at a given bitrate point.

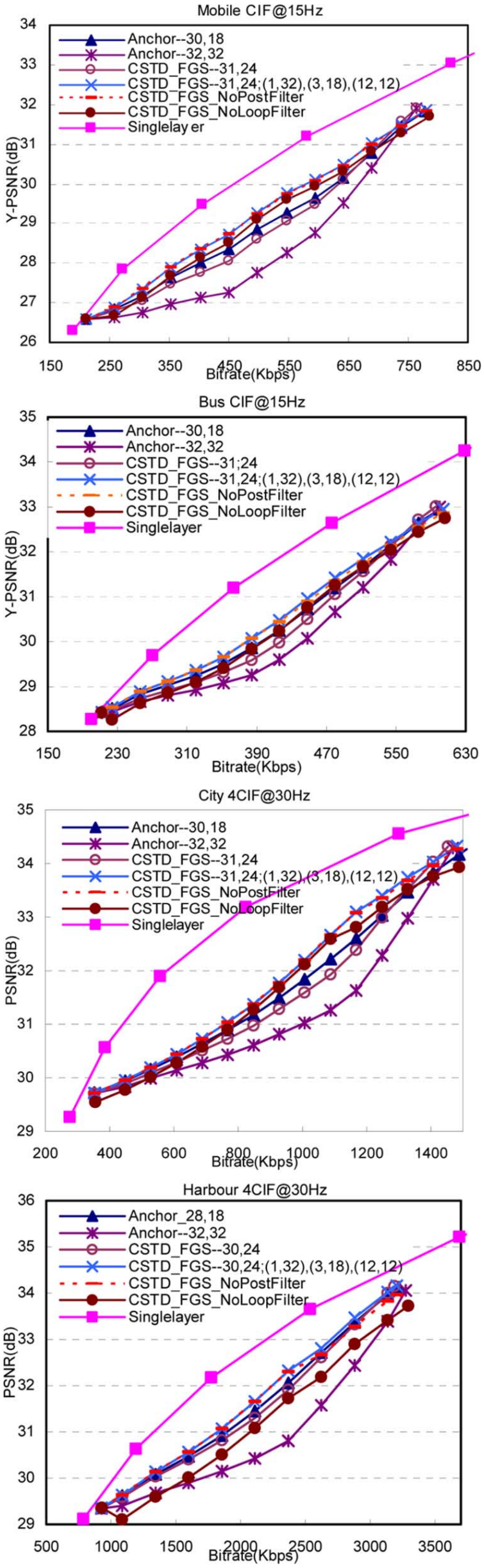


Fig. 6. Rate-distortion performance curves comparison between single layer coding, AR-FGS (Anchor) and the proposed method (CSTD_FGS) for different test sequences when only one FGS layer is appended on top of the base layer.

B. Rate-Distortion Performance for Two FGS Layers

To give the rate-distortion performance comparisons between AR-FGS and the proposed leaky prediction over a wider

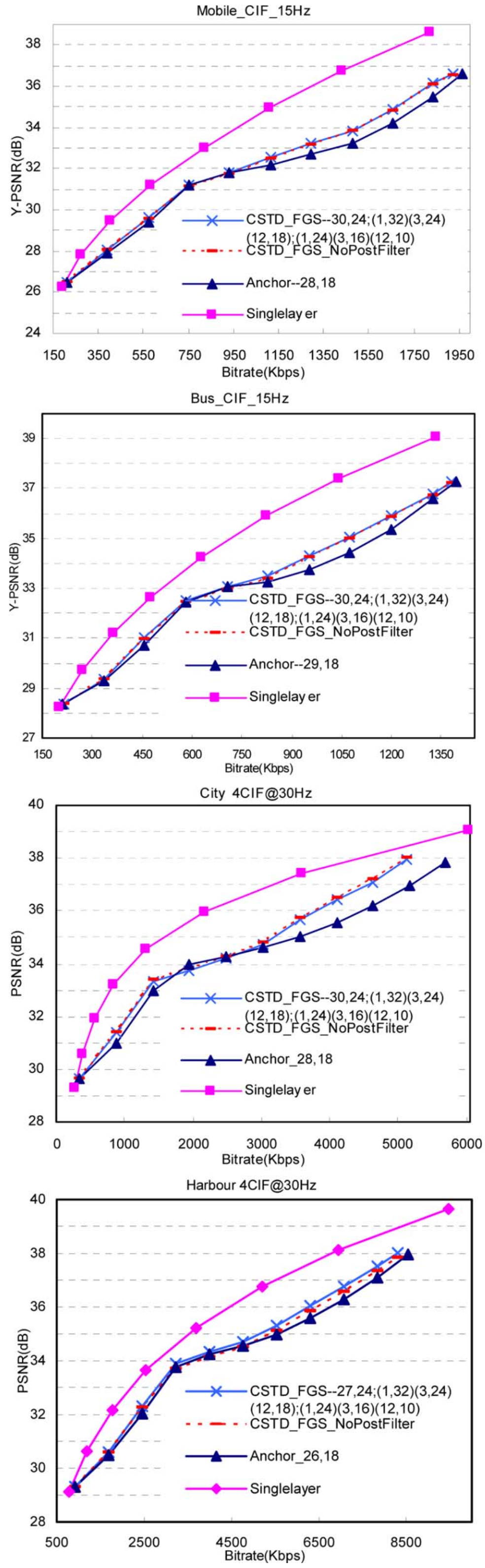


Fig. 7. Rate-distortion performance curves comparison between AR-FGS (Anchor), single-layer coding, and the proposed method (CSTD_FGS) with two FGS layers on top of base layer.

range of bitrates, Fig. 7 demonstrates their rate-distortion curves when two FGS layers are appended on top of the base layer. In this testing, scaling matrixes γ_0 and γ_1 for the first and second FGS layer are fixed to (1,32),(3,24),(12,18) and (1,24),(3,16),(12,10), respectively. Here, the leaky factor α_2 is also fixed to 24. As mentioned previously, α_1 can be easily adjusted to yield the ideal prediction signal because it is not within motion-compensation operation. Furthermore, encoder optimization approach for multiple FGS layers coding depicted in Section III-E is used to more effectively control drift error. The weighting value ε in (22) is set to 0.9 for the first FGS layer coding for all test sequences. As shown in Fig. 7 for all test sequences, the proposed leaky prediction can significantly improve the rate-distortion performance compared with AR-FGS over a wide range of bitrates not only at the first FGS layer but also at the second FGS layer. The maximum coding gain is even up to 1 dB for the *city* sequence. Similar to single FGS-layer coding, CSTD_FGS—NoPostFilter for two FGS layer also almost does not result in any rate-distortion performance loss when compared with CSTD_FGS with post filtering.

Compared with single-layer coding, CSTD_FGS with two FGS layers has bigger performance penalty against CSTD_FGS with a single FGS layer, because it usually suffers from more serious drift error for the truncated points and its motion field representation is more difficult to be optimized at a wider range of bitrates. Here, we just demonstrate the performance of FGS bitstream when only one or two FGS layers are appended on top of the base layer. In practical application, we may have a good and flexible tradeoff between coding efficiency and the allowed scalable range of FGS bitstream.

V. CONCLUSION

This paper proposes a novel FGS coding with cycle-based leaky prediction, in which multiple leaky factors can be applied to form enhancement-layer prediction through multiple leaky factors. First, the advantage of the leaky prediction with two leaky factors is theoretically analyzed when compared with the leaky prediction only with one leaky factor. This gives a guidance of how to effectively generate a partial enhancement layer and select the proper leaky factors to control drift error while keeping high coding efficiency. In addition, the leaky prediction with coefficient scaling in the transform domain is proposed to avoid the high-decoding complexity issue on multiple reconstructions of a partial enhancement layer at different quality levels of the enhancement layer. Furthermore, an encoder optimization approach for multiple FGS layers coding is also introduced to more effectively restrain drift error. Experimental results show that compared to the latest AR-FGS coding in the Scalable Extension of H.264/AVC standard, the proposed method is able to significantly improve the rate distortion performance over a wide range of bitrates for both single FGS layer and two FGS layers coding. Therefore, the proposed leaky prediction is able to afford a more efficient FGS coding solution in practical application.

REFERENCES

[1] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, “Overview of the H.264/AVC video coding standard,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 560–576, Jul. 2003.

[2] *Generic Coding of Moving Pictures and Associated Audio Information—Part 2: Video*, , Nov. 1994, ITU-T Rec. H.262 and ISO/IEC 13818-2 (MPEG-2).

[3] Video Coding for Low Bitrate Communication 1995, ITU-T Recommendation H.263.

[4] *Coding of Audio-Visual Objects—Part 2: Visual*, , Jan. 1999, ISO/IEC 14496-2 (MPEG-4 Part 2).

[5] W. Li, "Overview of fine granularity scalability in MPEG-4 video standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 2, pp. 301–317, Mar. 2001.

[6] H. Radha, M. van der Schaar, and Y. Chen, "The MPEG-4 fine-grained scalable video coding method for multimedia streaming over IP," *IEEE Trans. Multimedia*, vol. 3, no. 3, pp. 53–68, Mar. 2001.

[7] F. Wu, Li Shipeng, and Y.-Q. Zhang, "A framework for efficient progressive fine granular scalable video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 3, pp. 332–344, May 2001.

[8] F. Wu, Li Shipeng, X. Sun, B. Zeng, and Y.-Q. Zhang, "Macroblock-based progressive fine granularity scalable video coding," *Int. J. Imaging Syst. Technol.*, vol. 13, no. 6, pp. 297–307, 2004.

[9] M. van der Schaar and H. Radha, "Adaptive motion-compensation fine-granular-scalability (AMC-FGS) for wireless video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 4, pp. 360–371, Jun. 2002.

[10] 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, 2002.

[11] Y. Gao and L. P. Chau, "Efficient fine granularity scalability using adaptive leaky factor," *IEEE Trans. Broadcasting*, vol. 51, no. 4, pp. 512–519, Dec. 2005.

[12] T. K. Tan, K. K. Pang, and K. N. Ngan, "A frequency scalable coding scheme employing pyramid and subband techniques," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 4, no. 2, pp. 203–207, Apr. 1994.

[13] Y. Bao, M. Karczewicz, X. Wang, and J. Ridge, "FGs coding with adaptive reference for low-delay application," in *Proc. Int. Conf. Image Processing (ICIP)*, Atlanta, GA, Oct. 2006, pp. 185–188.

[14] , T. Wieg, G. Sullivan, J. Reichel, H. Schwarz, and M. Wien, Eds., "Joint Draft 7 of SVC Amendment," in *Joint Video Team (JVT), Doc. JVT-T201*. Klagenfurt, Austria: , 2006.

[15] Y. Zheng, X. Ji, D. Zhao, and W. Gao, "Low-Complexity FGS coding for low-delay applications," in *Joint Video Team, Doc. JVT-U077*, Hangzhou, China, Oct. 2006.

[16] Y. Zheng, X. Ji, D. Zhao, and W. Gao, "Improvements to multiple FGS layers coding for low-delay applications," in *Joint Video Team, Doc. JVT-T075*, Klagenfurt, Austria, Jul. 2006.

[17] X. M. Zhang, A. Vetro, Y. Q. Shi, and H. Sun, "Constant quality constrained rate allocation for FGS-coded video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 2, pp. 121–130, Feb. 2003.

[18] J. Wu, J. Cai, and C. W. Chen, "Rate-distortion analysis of leaky prediction based FGS video for constant quality constrained rate adaptation," *J. Vis. Commun. Image Representation*, vol. 18/1, pp. 45–58, Feb. 2007.

[19] J. Ridge, Y. Bao, M. Karczewicz, and X. Wang, "Fine-grained scalability for H.264/AVC," in *Proc. 8th Int. Symp. Signal Process. Its Applications*, Sydney, Australia, Aug. 2005.

[20] M. Horowitz, A. Joch, F. Kossentini, and A. Hallapuro, "H.264AVC baseline profile decoder complexity analysis," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 704–716, Jul. 2003.

[21] X. Wang, M. Karczewicz, J. Ridge, and N. Ammar, "CE7 report, multiple FGS layer coding for low-delay applications," in *Joint Video Team, Doc. JVT-R077*, Bangkok, Thailand, Jan. 2006.

[22] S. Jeong, S. Kamp, X. Wang, and X. Ji, "CE 5: Improvement of AR PR slices," in *Joint Video Team, Doc. JVT-T305r1*, Klagenfurt, Austria, Jul. 2006.

[23] Y. Bao and Y. Ye, "FGS complexity reduction," in *Joint Video Team, Doc. JVT-T087*, Klagenfurt, Austria, Jul. 2006.

[24] Y. Bao and Y. Ye, "Report of core experiment on PR slice improvements (CE1)," in *Joint Video Team, Doc. JVT-U125*, Hangzhou, China, Oct. 2006.

Xiangyang Ji received the B.S. and M.S. degrees in computer science from Harbin Institute of Technology, Harbin, China, in 1999, and 2001, respectively. He is currently working toward the Ph.D. degree at the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China.

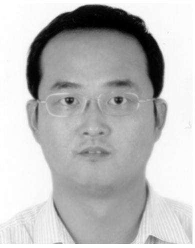
He has authored or coauthored over 30 conference and journal papers. His research interests include video/image coding, video streaming, and multimedia processing.

Yanyan Zheng, photograph and biography not available at the time of publication.

Debin Zhao received the B.S., M.S., and Ph.D. degrees from the Harbin Institute of Technology (HIT), Harbin, China, in 1985, 1988, and 1998, respectively, all in computer science.

He joined the Department of Computer Science, HIT, as an Associate Professor in 1993. He is currently a Professor with HIT and the Institute of Computing Technology, Chinese Academy of Sciences, Beijing. He has been a Research Fellow with the Department of Computer Science, City University of Hong Kong. His research interests include multimedia compression and its related applications. He has authored or coauthored over 100 publications.

Dr. Zhao was the recipient of three National Science and Technology Progress Awards of China (Second Prize) and the Excellent Teaching Award from the Baogang Foundation.



Feng Wu (M'99–SM'06) received the B.S. degree in electrical engineering from Xidian University, Shaanxi, China, in 1992 and the M.S. and Ph.D. degrees in computer science from Harbin Institute of Technology, Harbin, China, in 1996 and 1999, respectively.

He joined in Microsoft Research China as an Associate Researcher in 1999. He has been a Researcher with Microsoft Research Asia, Beijing, since 2001. His research interests include image and video representation, media compression and communication, and computer vision and graphics. He has been an active contributor to ISO/MPEG and ITU-T standards. Some techniques have been adopted by MPEG-4 FGS, H.264/MPEG-4 AVC and the coming H.264 SVC standard. He served as the chairman of China AVS video group in 2002–2004 and led the efforts on developing China AVS video standard 1.0. He has authored or coauthored over 100 conference and journal papers. He has approximately 30 U.S. patents granted or pending in video and image coding.

Wen Gao (M'99–SM'05) received the M.S. and Ph.D. degrees in computer science from Harbin Institute of Technology, Harbin, China, 1985 and 1988, respectively, and the Ph.D. degree in electronics engineering from University of Tokyo, Tokyo, Japan, in 1991.

He was a Research Fellow with the Institute of Medical Electronics Engineering, University of Tokyo, in 1992, and a Visiting Professor with the Robotics Institute, Carnegie-Mellon University, Pittsburgh, PA, in 1993. From 1994 to 1995, he was a Visiting Professor with Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge. Currently, he is a Professor with the School of Electronic Engineering and Computer Science, Peking University, Peking, China, and the External Fellow of International Computer Science Institute, University of California, Berkeley. He has published seven books and over 200 scientific papers. His research interests are in the areas of signal processing, image and video communication, computer vision and artificial intelligence. He is the Editor-in-Chief of the *Chinese Journal of Computers.* He chairs the Audio Video coding Standard (AVS) workgroup of China, and he is the head of Chinese National Delegation to MPEG working group (ISO/SC29/WG11).