A DISCUSSION OF LEAKY PREDICTION BASED SCALABLE CODING

Yuxin Liu, *Zhen Li,* *Paul Salama,***†and Edward J. Delp**

*Video and Image Processing Laboratory (*VIPER*)
School of Electrical and Computer Engineering
Purdue University, West Lafayette, IN, 47907, U.S.A.

**Department of Electrical and Computer Engineering
Indiana University – Purdue University Indianapolis
Indianapolis, IN, 46202 U.S.A.

ABSTRACT

In this paper, we focus on the leaky prediction based scalable coding (LPSC) structure and present a general framework for LPSC. We demonstrate the similarity between LPSC and a motion compensation based multiple description coding scheme. We show that since the information contained in the enhancement layer in LPSC is actually a mismatch between two descriptions for each frame, it cannot be guaranteed that the enhancement layer always achieves superior reconstruction quality beyond that achieved by the base layer. We derive three reconstructions for each frame under the LPSC framework, and propose a maximum-likelihood (ML) estimation scheme for LPSC video reconstruction at the decoder. This generally achieves superior decoded video quality than both the enhancement layer and the base layer.

1. INTRODUCTION

Layered scalable coding has a structure that is nested whereby different levels of the bit stream are decoded in a fixed sequential order. This property is sometimes referred to as "nested scalability." Fine granularity scalability (FGS) is a specific layered scalable coding structure, which possesses fully rate (or SNR) scalability over a wide range of data rates [1,2]. Scalable coding is desired for error resilient video transmission over heterogeneous networks with changing bandwidth because the inherent structure allows one to protect parts of the bit stream differently, i.e. the use of unequal error protection. 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. For example, the base layer bit stream could be protected by Forward Error Correction (FEC) coding, or transmitted using an error-recovery capable network protocol such as TCP. The enhancement layer of course still remains vulnerable to errors. Consequently, traditional layered scalable coding schemes usually do not incorporate the enhancement layer into the motion compensation loop at the encoder to prevent error drift at the decoder. To circumvent this leaky prediction based scalable coding (LPSC) [3,4,5] includes an incomplete version of the enhancement layer within the motion compensation loop to improve coding efficiency while maintaining graceful error resilience performance.

2. GENERAL FRAMEWORK FOR LPSC

2.1 Overview of LPSC

Suppose that the nth frame in a video sequence is being encoded using a layered technique. Let $\hat{F}_B(n-1)$ be the reconstructed image from the base layer for the previous frame, and $\hat{F}_E(n-1)$ the reconstructed image from the enhancement layer. In LPSC, $\hat{F}_B(n-1)+\alpha \cdot (\hat{F}_E(n-1)- \hat{F}_B(n-1))$ is stored as the reference frame of the enhancement layer, instead of $\hat{F}_E(n-1)$. Let $\tilde{F}_B(n-1)$ and $\tilde{F}_E(n-1)$ denote the motion compensated base layer and enhancement layer pictures, then $e_B(n)= F(n)-\tilde{F}_B(n-1)$ is the predicted error frame (PEF) of the base layer. A quantized version $\hat{e}_B(n)$ is transmitted. The enhancement layer PEF is given by $e_E(n)= F(n)-\alpha\tilde{F}_E(n-1)-(1-\alpha)\tilde{F}_B(n-1)$, and a quantized version of the mismatch between the two PEFs,

$$\psi(n)=e_E(n)-\hat{e}_B(n)$$
$$= F(n)-\alpha\tilde{F}_E(n-1)-(1-\alpha)\tilde{F}_B(n-1)-\hat{e}_B(n) \qquad (1)$$

is transmitted. Thus, the reconstructed pictures at the decoder for the current frame from the base layer and the enhancement layer are:

$$\hat{F}^{(1)} = \hat{F}_B(n) = \tilde{F}_B(n-1)+\hat{e}_B(n) \qquad (2)$$
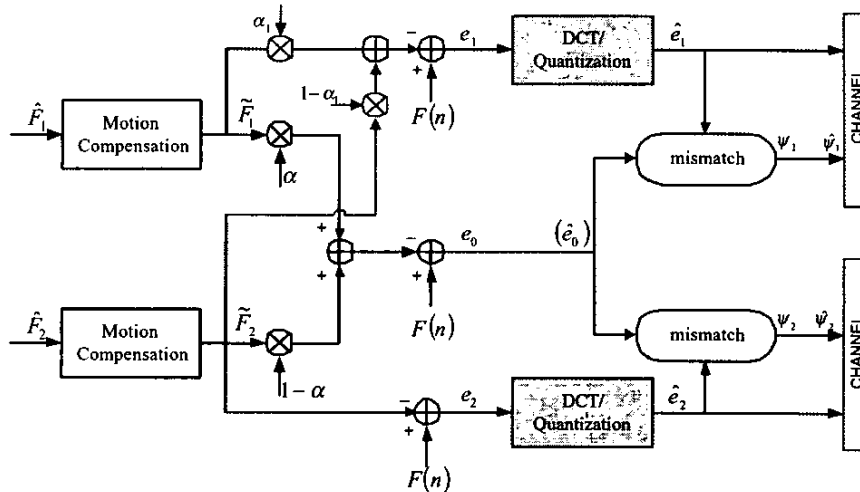
Figure 1 General framework for leaky prediction based scalable coding (LPSC)

$$\hat{F}^{(2)} = \hat{F}_E(n) = \alpha \tilde{F}_E(n-1) + (1-\alpha)\tilde{F}_B(n-1) + \hat{e}_B(n) + \hat{\psi}(n) \quad (3)$$

Due to α the enhancement layer will degrade exponentially with the time [3]. A significant advantage of LPSC is that the trade-off between coding efficiency and error resilience is balanced and controlled by α. When $\alpha=1$, all of the enhancement layer is used. This maximizes coding efficiency but minimizes error robustness. The opposite performance is obtained when $\alpha=0$, and the traditional layered scalable coding structure attained.

2.2 General Framework for LPSC

We propose a general framework for LPSC (see Figure 1). In this framework, if we let $\hat{F}_2 = \hat{F}_B(n-1)$ and $\hat{F}_1 = \hat{F}_E(n-1)$, the bottom loop becomes the base layer motion compensation loop in LPSC, and the middle loop becomes the enhancement layer motion compensation loop in LPSC. We add another motion compensation loop on top to incorporate a second leaky prediction based enhancement layer. Notice that the top two layers represent the essential framework for the high quality base layer scheme proposed in [3], when $\alpha_1 < \alpha$.

Based on the general framework in Figure 1, we can show that LPSC is similar to the multiple description coding (MDC) scheme MDMC proposed in [6]. If we let $\hat{F}_1 = \hat{F}(n-1)$ and $\hat{F}_2 = \hat{F}(n-2)$, where $\hat{F}(n-1)$ and $\hat{F}(n-2)$ denote the reconstructed pictures for the previous two frames, and let $\alpha_1 = 1$, then the framework given is exactly MDMC. The middle loop corresponds to the central loop in MDMC while the top and bottom loops correspond to the two side loops.

From the encoder's point of view, the major difference between LPSC and MDMC is that in LPSC $\hat{F}_1$ and $\hat{F}_2$ are the two descriptions of a frame and contain overlapping

information. From the decoder's point of view, the two descriptions created by MDMC, each carrying the PEF generated by the central loop and the mismatch between its own PEF and the central PEF, are transmitted over two channels. The decoder thus has to worry about three scenarios, whether description I or description II or both are available. LPSC is only concerned with whether description I (the base layer description) or both are available, since the enhancement layer information becomes useless without the base layer due to the nested scalability structure. Based on this similarity between LPSC and MDMC, we investigate whether, for LPSC, the information transmitted by the enhancement layer using $\psi(n) = e_E(n) - \hat{e}_B(n)$ will always "enhance" the reconstructed video at the decoder. We know that in MDMC, a coarse quantized mismatch $\hat{\psi}(n)$ is conveyed by the side loop in order to reduce redundancy. When both descriptions are available, the decoder always favors the reconstructed signal of the central loop without using $\psi(n)$. Similarly, a coarse quantized version of $\psi(n)$ might also result in an inferior reconstructed signal in LPSC.

The superior performance achieved by the enhancement layer as opposed to that by the base layer in LPSC is largely determined by α. From Equation (1) and (2),

$$\psi(n) = e_E(n) - \hat{e}_B(n) = (e_E(n) - e_B(n)) + (e_B(n) - \hat{e}_B(n))$$
$$= (F(n) - \alpha \tilde{F}_E(n-1) - (1-\alpha)\tilde{F}_B(n-1)) - (F(n) - \tilde{F}_B(n-1)) + (F(n) - \hat{F}_B(n))$$
$$= \alpha(\tilde{F}_B(n-1) - \tilde{F}_E(n-1)) + (F(n) - \hat{F}_B(n))$$
$$= \alpha \cdot MC\{\hat{F}_B(n-1) - \hat{F}_E(n-1)\} + (F(n) - \hat{F}_B(n)).$$

Therefore, if $\alpha = 0$, $\hat{\psi}(n)$ is the quantized version of the mismatch between the reconstructed picture of the base layer and the original signal. Thus, any approximation of $\psi(n)$ has more information about the original signal than

knowledge of just the base layer. If $\alpha > 0$, $\hat{\psi}(n)$ includes another portion that is the motion compensated version of the difference between the two reconstructed pictures of the previous frame multiplied by the leaky factor. The larger the leaky factor α, the larger the difference between the previous two reconstructions. In other words, in LPSC, the leaky factor not only balances the trade-off between coding efficiency and error resilience, but it also is related to the superior performance of the enhancement layer. In summary, three factors cause inferior performance: the quantization step for $\psi(n)$, the leaky factor α, and the difference between the two constructions by the base layer and the enhancement layer from the reference frame.

3. LPSC FACILITATED BY ML ESTIMATION

3.1 Three Reconstructions by LPSC

Using the similarity between LPSC and MDMC we obtain three reconstructions for each frame using the traditional LPSC framework. The first two reconstructions are the ones obtained by the base layer as indicated by Equation (2) and the enhancement layer in Equation (3). Motivated by the scheme in [6], if only one description is available at the decoder, the other description is estimated from the correlation between the central loop and the side loop, we derive a third reconstruction for each frame by exploiting the information from the next frame via backward motion compensation.

From Equations (2) and (3), we have

$$F(n+1) = \alpha\tilde{F}_L(n) + (1-\alpha)\tilde{F}_B(n) + \hat{e}_B(n+1) + \hat{\psi}(n+1) + (\psi(n+1) - \hat{\psi}(n+1))$$

$$= \tilde{F}_B(n) + \hat{e}_B(n+1) + (e_B(n+1) - \hat{e}_B(n+1)), \text{ then}$$

$$\tilde{F}_E(n) = \tilde{F}_B(n) - \frac{\psi(n+1)}{\alpha} + \frac{(e_B(n+1) - \hat{e}_B(n+1)) - (\psi(n+1) - \hat{\psi}(n+1))}{\alpha}$$

Given $\hat{F}_B(n)$ and $\hat{\psi}(n+1)$, the optimal estimation for $\tilde{F}_E(n)$ in the mean square error (MSE) sense is:

$$\tilde{F}_{mse}(n) = E[\tilde{F}_E(n)|\hat{F}_B(n), \hat{\psi}(n+1)] = \tilde{F}_B(n) - \frac{\psi(n+1)}{\alpha} +$$

$$\frac{E[(e_B(n+1) - \hat{e}_B(n+1)) - (\psi(n+1) - \hat{\psi}(n+1))|\hat{F}_B(n), \hat{\psi}(n+1)]}{\alpha}$$

We assume that the noise introduced by the quantization in both the base layer and the enhancement layer are independent of $\hat{F}_B(n)$ and $\hat{\psi}(n+1)$ and have zero means. Thus we obtain a third reconstructed picture for the current frame as follows

$$\hat{F}^{(3)} = \hat{F}_{mse}(n) = MC^{-1}\left\{\tilde{F}_B(n) - \frac{\hat{\psi}(n+1)}{\alpha}\right\}, \quad 0 < \alpha \le 1 \quad (4)$$

where $MC^{-1}\{\cdot\}$ denotes the backward motion compensation using the forward motion vector data. Notice that if we use $\tilde{F}_{mse}(n)$ to reconstruct the next frame, we obtain

$$\hat{F}_{mse}(n+1) = \alpha\left(\tilde{F}_B(n) - \frac{\hat{\psi}(n+1)}{\alpha}\right) + (1-\alpha)\tilde{F}_B(n) + \hat{e}_B(n+1) + \hat{\psi}(n+1)$$

$$= \tilde{F}_B(n) + \hat{e}_B(n+1) = \hat{F}_B(n+1),$$

which means the enhancement layer in one frame can help conceal the enhancement layer of previous frames, but cannot provide any superior reconstruction quality beyond that achieved by the base layer for itself and all the frames thereafter. Note that the third reconstruction is desirable especially for video transmission over error-prone networks, since the current enhancement layer may be destroyed or lost while the next enhancement layer information might be available. However, two disadvantages are associated with $\hat{F}^{(3)}(n)$: (1) $0 < \alpha < 1$ amplifies the quantization error introduced in $\hat{\psi}(n+1)$, and (2) the implementation of the backward motion compensation based on forward motion vectors for fractional motion vectors is still an open problem.

3.2 LPSC by Maximum-Likelihood (ML) Estimation

We use the maximum likelihood (ML) scheme originally proposed for MDC in [7]. We assume that the quantization noise in each pixel is an independent, identical distributed zero-mean Gaussian random variable. For a frame with size $M \times N$, let $\hat{F}_B = \{\hat{p}_B(x,y)\}_{x,y=(1,1)}^{(M,N)}$, $\hat{F}_E = \{\hat{p}_E(x,y)\}_{x,y=(1,1)}^{(M,N)}$, and we obtain the ML estimate for each pixel as

$$\hat{p}_{ML}(x,y) = \left([1 \quad 1]Corr^{-1}\begin{bmatrix}1\\1\end{bmatrix}\right)^{-1}[1 \quad 1]Corr^{-1}\begin{bmatrix}\hat{p}_B(x,y)\\\hat{p}_E(x,y)\end{bmatrix}$$

$$= [\pi \quad 1-\pi]\begin{bmatrix}\hat{p}_B(x,y)\\\hat{p}_E(x,y)\end{bmatrix} \quad (5)$$

where

$$Corr = E\left(\begin{bmatrix}p-p_B\\p-p_E\end{bmatrix} \cdot [p-p_B \quad p-p_E]\right)$$

$$= \begin{bmatrix}E(p-p_B)^2 & E(p-p_B)(p-p_E)\\E(p-p_E)(p-p_B) & E(p-p_E)^2\end{bmatrix} \triangleq \begin{bmatrix}a & b\\c & d\end{bmatrix}$$

As in [7], we use empirical averages to approximate expectation $E(p-p_B)^2 \cong \frac{1}{M \times N}\sum_{x,y=(1,1)}^{(M,N)}(p(x,y) - \hat{p}_B(x,y))^2$

Thus, we transmit an additional ML parameter associated with each frame $\pi = d - c/a + d - b - c$, and a fourth reconstructed picture can be thus achieved at the decoder as: $\hat{F}_{ML} = \{\hat{p}_{ML}(x,y)\}_{x,y=(1,1)}^{(M,N)}$ \quad (6)

4. EXPERIMENTAL RESULTS AND CONCLUSION

We use the *Foreman* sequence in our experiments. All frames are 4:2:0 YUV QCIF and 400 frames in length. We use PSNR as the distortion metric for each decoded frame. We modified the H.26L reference software version TML9.4 [8] to implement LPSC, and encoded both the base and enhancement layers using the UVLC mode with the same VLC table as the non-scalable coding structure. To determine the performance of the enhancement layer in LPSC as a function of the leaky factor and the quantization steps, we intra-coded the first frame of each sequence and inter-coded all successive frames.

We chose the quantization parameter for the base layer in INTRA frames to be 10 and in INTER frames to be 24. We fixed the quantization step for the enhancement layer in INTRA frames to 6. We encoded all ML coefficients using 6-bits. We noticed that even when we use a coarse quantization step for the enhancement layer, such as 26, the enhancement layer comprised approximately more than 200 bits/frame. Therefore, 6-bits is a very acceptable redundancy associated with each frame. We observed that the ML coefficients do not change much from frame to frame, implying that a more efficient way to encode ML coefficients is possible.
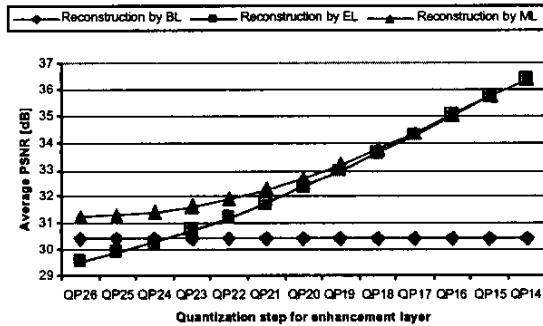

Figure 2 LPSC performance of Foreman with respect to the quantization step for the enhancement layer (leaky factor $\alpha = 0.95$)
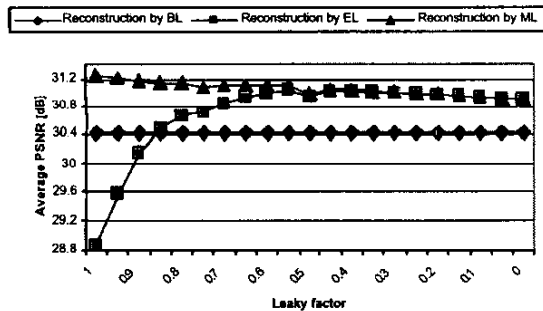

Figure 3: LPSC performance of Foreman with respect to the leaky factor (QP for enhancement layer equal to 26)

As shown in Figure 2, when the leaky factor is close to 1, a coarse quantization for the enhancement layer results in inferior performance at the enhancement layer as opposed to the base layer. Increasing the accuracy of $\psi(n)$ increases the performance of the enhancement layer beyond the video quality achieved by the base layer. On the other hand, as shown in Figure 3, when $\psi(n)$ is coarsely quantized, the performance of the enhancement layer is closely related to the value of the leaky factor. The larger the leaky factor, the worse the video quality the

enhancement layer provides. Both figures demonstrate that our scheme for LPSC facilitated by ML estimation achieves superior decoded video quality of up to 0.8 dB beyond the base layer and 1.5 dB beyond the enhancement layer when α is close to 1 and $\psi(n)$ is coarsely quantized. With the increase in α and the accuracy of $\psi(n)$, the performance achieved by ML estimation is close to that of the enhancement layer but never is inferior to both layers.

The contribution of this paper is the development of the similarity between Multiple Description Coding and leaky layered coding. We demonstrated a shortcoming of leaky layered coding and used the ideas for MDC to better coding performance. We found that the leaky factor is not only key in balancing the trade-off between coding efficiency and error resilience, but also validated the functionality of the enhancement layer. In the future we will: (1) investigate backward motion compensation based on the forward motion vectors, and (2) investigate the performance of ML estimation for erasure packet networks.

5. 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] H.C. Huang, C.N. Wang, and T. Chiang, "A robust fine granularity scalability using trellis-based predictive leak," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, pp. 372-385, Jun. 2002.

[4] S. Han and B. Girod, "Robust and efficient scalable video coding with leaky prediction," *Proceedings of IEEE International Conference on Image Processing*, vol. 2, pp. 41-44, Rochester, New York, Sep. 22-25, 2002.

[5] W.-H. Peng and Y.-K. Chen, "Error drifting reduction in enhanced fine granularity scalability," *Proceedings of IEEE International Conference on Image Processing*, vol. 2, pp. 61-64, Rochester, New York, Sep. 22-25, 2002.

[6] 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, Jun. 2002.

[7] X. Tang and A. Zakhor, "Matching pursuits multiple description coding for wireless video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, pp. 566-575, Jun. 2002.

[8] T. Wiegand, "H.26L test model long-term number 9 (TML-9)," ITU-T Q.6/SG 16, VCEG-N83d1, Pattaya, Thailand, Dec. 2001.