SMOOTHED REFERENCE PREDICTION FOR IMPROVING SINGLE-LOOP DECODING PERFORMANCE OF H.264/AVC SCALABLE EXTENSION

So-Young Kim and Woo-Jin Han
DM R&D Center, Samsung Electronics 416, Maetan3-dong, Youngtong-gu, Youngtong-dong
Kyounggi-do 442-742, Korea

Keywords: Video coding, video signal processing, smoothing methods, digital filters and prediction methods.

Abstract: It is well-known that multi-layer extension of H.264/AVC shows good spatial scalability performance mainly due to its efficient inter-layer prediction techniques. Although single-loop decoding is a kind of technique to reduce the decoder-side computational complexity by performing only one motion compensation to decode multi-layer data, its limited use of inter-layer prediction sometimes degrades the performance especially for fast-motion sequences. In this paper, smoothed reference prediction technique is proposed to improve the single-loop decoding performance by replacing base-layer information with current-layer information and simple block-based smoothing function. Experimental results show that the proposed method can improve the coding efficiency with all benefits of single-loop decoding mode. In addition, the proposed method was adopted to scalable extension of H.264/AVC standard Working Draft.

1 INTRODUCTION

Scalable video coding has received considerable attention for many multimedia applications in terms of both coding algorithms and standard activites. Among many technologies used in scalable video coders, scalable extension of H.264/AVC (Schwarz and Hinz, 2004., Schwarz and Marpe, 2004., Schwarz and Marpe, 2005., Reichel and Schwarz, 2005) has been considered as one of the best compromise between the coding efficiency and the excellent scalability features. The scalable extension of H.264/AVC is based on a multi-layer structure to allow the spatial, temporal and SNR scalabilities.

Generally, it is well-known that the multi-layer has some penalty on the coding efficiency compared to the single-layer structure mainly due to the redundant representation of information (Schwarz and Marpe, 2005). To minimize the multi-layer overhead, several kinds of inter-layer prediction techniques have been exploited in the H.264/AVC scalable extension. More specifically, motion, texture, and residual data can be predicted from already coded layers. With the inter-layer prediction techniques, it was reported that the H.264/AVC scalable extension can provide comparable coding efficiency compared to the current state-of-the-art video coder (Schwarz and Marpe, 2005).

A major drawback of the multi-layer structure lies in its heavy complexity requirements. Especially, when the more the number of layers are, the larger total complexity requirements is applying to the real world applications. Single-loop decoding technique (Schwarz and Hinz, 2005) was proposed to reduce the decoding complexity of multi-layer structure by allowing the inter-layer texture prediction only when the corresponding base-layer macroblock is one of the intra-type, which does not need motion compensation. In other words, the decoder of H.264/AVC scalable extension performs motion compensation only once in the top-most layer even in the bit-stream with a multiple number of layers. Although the decoding complexity can be reduced significantly, the restriction of the inter-layer texture prediction sometimes shows a non-negligible degradation up to 0.7 dB especially in the fast-motion sequences. Figure 1 shows a general coder structure for scalable extension of H.264/AVC with two spatial layers. The redundant information such as motion and texture between consecutive layers is used for inter-layer prediction. Inter-layer texture prediction and residual prediction are explited for effective inter-layer prediction.

Kim S. and Han W. (2007).
SMOOTHED REFERENCE PREDICTION FOR IMPROVING SINGLE-LOOP DECODING PERFORMANCE OF H.264/AVC SCALABLE EXTENSION.
In *Proceedings of the Second International Conference on Signal Processing and Multimedia Applications*, pages 67-71
DOI: 10.5220/0002135800670071

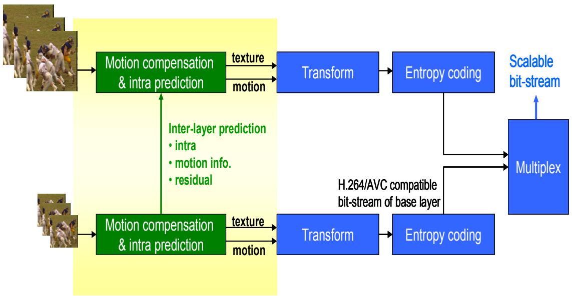


Figure 1: A coder structure for the scalable extension of H.264/AVC.

In this paper, we compare the characteristics of the inter-layer texture and residual prediction techniques and propose a modified version of inter-layer residual prediction to compensate the coding penalty from the restriction of the inter-layer texture prediction. This smoothed reference prediction technique was adopted to Working Draft in scalable extension of H.264/AVC standardIt will be shown that the major coding penalty of the single-loop decoding results from the blocking artifacts due to the mismatch between two different layers and adaptive usage of simple smoothing function can help to reduce this effect.

Section 2 describes the basic concept of the inter-layer texture and residual predictions, Section 3 presents the proposed smoothed reference prediction technique. Experimental results and our conclusions arising from our work are depicted in Section 4 and section 5.

2 INTER-LAYER TEXTURE AND RESIDUAL PREDICTIONS

Inter-layer texture prediction exploits the fact that the decoded texture of the lower layer is generally similar to the corresponding current texture. It uses the decoded macroblock of the lower layer as a predictor of the current macroblock. Only the difference between them is coded and transmitted as

$$R_T = O_c - U[O_b] \quad (1)$$

where R_T is the residual signal using the inter-layer texture prediction. O_c and O_b are original signals of the current and lower layers, respectively. The upsampling filter, $U(\cdot)$, is optionally applied when the spatial resolutions of two layers are different.

On the contrary, inter-layer residual prediction exploits redundancy of the residual signals between two layers. Thus, the decoded residual signal of the lower-layer is used as a predictor of the current residual signal.

(a) (b)

Figure 2: Visual example of (a) single-loop decoding and (b) multi-loop decoding.

The rationale behind the inter-layer residual prediction lies in the similarity of the motion fields including both reference frames and motion vectors in multi-layer structure. In that case, the resultant residual signals of two layers are also similar thus the inter-layer residual prediction can improve the coding efficiency. The residual signal of the inter-layer residual prediction, R_R can be defined as

$$R_R = R_c - U[R_b] \quad (2)$$

where Rc is the residual signal of the current layer without inter-layer prediction and Rb is the reconstructed residual signal of the lower layer, respectively. It should be noted that the actual implementation of the upsampling filter U is different to (1) and the same notation is used for the simplicity's sake.

Although the inter-layer texture and residual predictions seem to be very different in nature, the following section shows some similarities between two inter-layer prediction techniques and the inter-layer residual prediction can be modified to improve the performance in the regions where the inter-layer texture prediction is not allowed.

3 SMOOTHED REFERENCE PREDICTION

In single-loop decoding mode, the inter-layer texture prediction is not allowed if the lower-layer macroblock is coded using the motion compensation process, that is the intra mode macroblocks of lower layers are reconstructed only at a decoder. In this case, inter-layer residual prediction or inter prediction not considering lower layer information is used instead of inter-layer texture prediction, which sometimes makes strong high-pass artifacts as shown in Figure 2. Figure 3 depicts well how many macroblocks in enhancement layer are coded by

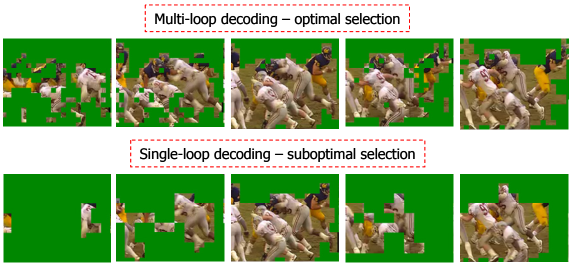


Figure 3: The intra and inter coded macroblocks in multi-loop decoding and single-loop decoding mode(green parts are the inter coded macroblocks).

normal intra prediction in H.264 and inter-layer texture prediction, and the green means the macroblocks coded by inter mode. It is shown that the restriction not to use inter coded macroblock in lower layer for inter-layer texture prediction in single-loop decoding reduces the ratio of intra coded macroblocks even where intra mode can give the best R-D performance.

To understand the situation further, (1) can be rewritten as

$$R_T = O_c - U \cdot D[P_b + R_b] \qquad (3)$$

where $D(\cdot)$ is a in-loop deblocking operator defined in H.264/AVC. Similarly, (2) can be rewritten as

$$R_R = O_c - (P_c + U[R_b]) \qquad (4)$$

Comparing (3) and (4) shows two differences: the first one is the temporal predictor. The inter-layer texture prediction uses the temporal predictor of the base-layer and the inter-layer residual prediction uses that of the current-layer. In addition, the inter-layer texture prediction exploits the in-loop deblocking operator as well as the upsampling operation on the reconstructed signal.

(a) inter prediction. (b) inter-layer residual prediction.

(c) inter-layer texture prediction. (d) smoothed reference prediction.

Figure 4: Four predictors using (a) inter prediction (b) inter-layer residual prediction (c) inter-layer texture prediction (d) smoothed reference prediction.

The first difference which uses the base-layer temporal prediction cannot be fixed due to the single-loop constraints. On the other hand, the second difference related to the deblocking and upsampling filters can be minimized if we apply some filters to (4). The main idea of the proposed method is to apply the suitable filter to the inter-layer residual prediction to give the inter-layer texture prediction benefits.

The most straightforward choice of the filter may be the upsampling and in-loop deblocking filters used in (3). It gives

$$R_S = O_c - U \cdot D \cdot M\left(P_c + U[R_b]\right) \qquad (5)$$

where R_S is a new residual signal and $M(\cdot)$ is a downsampling filter used for generating the lower resolution layer. However, the upsampling and downsampling operations generally require frame-based processing. The additional complexity due to the resampling and deblocking filters is also significant. Since the predictor used in the inter-layer texture prediction can be thought as the low pass approximate of the current layer, we use simple bi-linear smoothing filter instead of the three filter combinations, $U{\cdot}D{\cdot}M(\cdot)$, defined as

$$x'(n) = \{x(n-1) + 2x(n) + x(n+1) + 2\}/4 \qquad (6)$$

where $x(n)$ is n-th pixel value in the signal. (6) is applied to both horizontal and vertical direction for all samples inside the current macroblock. For the macroblock boundaries, only top and left macroblock boundaries are modified to allow the block-based processing.

We call this kind of prediction as the smoothed reference since it can be realized as the smoothed version of the inter-layer residual prediction. The final equation of the proposed method is given by

$$R_S = O_c - S\left(P_c + U[R_b]\right) \qquad (7)$$

where $S(\cdot)$ is a bi-linear smoothing filter. It should be noted that (6) requires only 3 additions and 2 shifts thus it can be implemented very efficiently. Furthermore, the actual complexity burden is much less since the new prediction is only used where single-loop constraint has significant penalty with respect to suitable rate-optimization criteria.

Figure 4 shows four different predictors generated by inter prediction, inter-layer residual prediction, inter-layer texture prediction and smoothed reference prediction, respectively. As shown in the figure, the inter prediction and the inter-layer residual prediction sometimes show very strong high-pass artifacts subjectively related to the blocking artifacts. The prediction signal by new smoothed reference prediction is remarkably similar to that of the inter-layer texture prediction, which justifies the reason to use the smoothing filter.

Table 1 shows a comparative result of the ratio of macroblock modes for single-loop, multi-loop and single-loop with smoothed reference prediction.

As shown in the table, in the single-loop decoding mode, the ratio of inter-layer texture prediction is significantly small whereas that of inter-layer residual prediction is large. All other macroblock types are used with similar relative ratios. It indicates that the inter-layer residual prediction is used for most macroblocks that violates the single-loop constraint instead of the inter-layer texture prediction. When the smoothed reference prediction is used, the relative distribution of macroblock types is very similar to the multi-loop case. This is because the smoothed reference prediction is used for the macroblocks suffering from single-loop constraint instead of the inter-layer residual prediction.

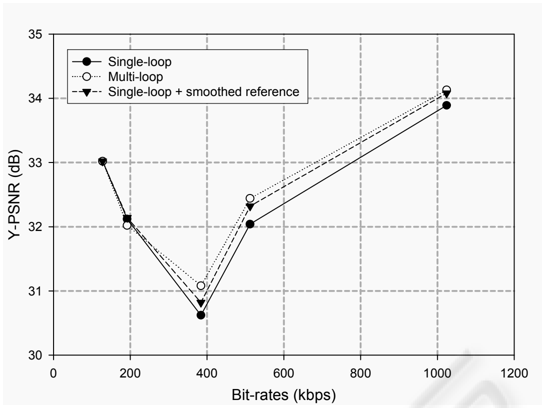
Table 1: Macroblock types for Football CIF.

Prediction type	Single-loop	Multi-loop	Smoothed reference
Inter-layer texture (+smoothed reference)	12%	29%	28%
Directional intra	1%	2%	1%
Inter-layer residual	42%	24%	26%
No inter-layer prediction	45%	45%	45%

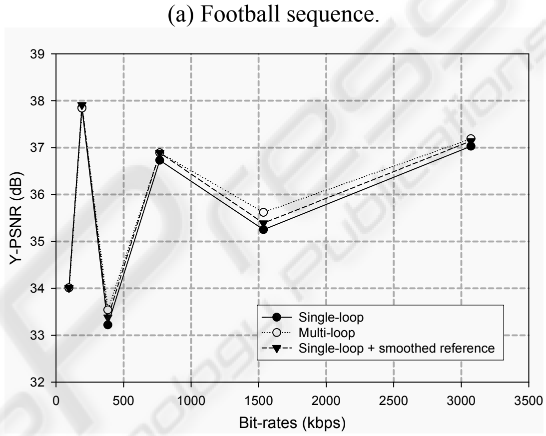
4 EXPERIMENTAL RESULTS

4.1 Experimental Condition

The proposed method was implemented in the reference software of H.264/AVC scalable extension developed in the MPEG/JVT as an extension of H.264/AVC. Football CIF and Soccer 4CIF sequences were used for the performance verification since these sequences have relatively large coding penalty when single-loop decoding is used. For the scalability test points, we use 5 layers for Football CIF and 6 layers for Soccer 4CIF as defined in Table 2.



(a) Football sequence.



(b) Soccer sequence.

Figure 5: PSNR curves for Football and Soccer sequence.

Table 2: Scalability test points used in the experiments sequence.

Football	Soccer
QCIF@7.5Hz, 128kbps	QCIF@15Hz, 96kbps
QCIF@15Hz, 192kbps	QCIF@15Hz, 192kbps
CIF@15Hz, 384kbps	CIF@30Hz, 384kbps
CIF@15Hz, 512kbps	CIF@30Hz, 768kbps
CIF@30Hz, 1024kbps	4CIF@30Hz, 1536kbps
	4CIF@60Hz, 3072kbps

4.2 PSNR Results

Figure 5 shows PSNR graphs for Football and Soccer sequences. In Football sequence, the maximum PSNR degradation from single-loop decoding constraint is about 0.5 dB. With the proposed method, the performance gap is reduced to 0.2 dB. The performance trend is similar in soccer sequence. The performance gap is reduced by a half with the proposed method. It indicates that there is further room to be improved in the single-loop decoding with smoothed reference prediction. The proposed method actually reduces the performance penalty due to the single-loop decoding constraint while still allowing the single-loop decoding concepts, which means much lower computational complexity compared to the multi-loop decoding case.

4.3 Visual Examples

Figure 7 and 8 show visual examples of the single-loop decoding mode and the single-loop decoding mode with the proposed method. Not only blocking artifact but high frequency component of prediction signals is removed. Comparing Figure 3 and 6, it is exploited that the ratio of intra coded macroblocks by R-D cost is improved almost similar to the ratio in multi-loop decoding mode.

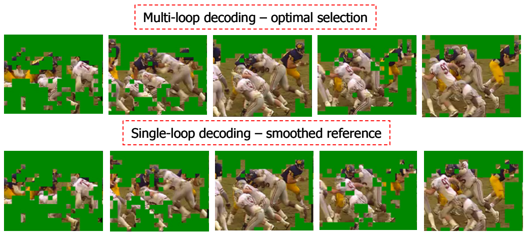


Figure 6: The intra and inter coded macroblocks in multi-loop decoding and single-loop decoding with smoothed reference prediction (green parts are the inter coded macroblocks.

5 CONCLUSION

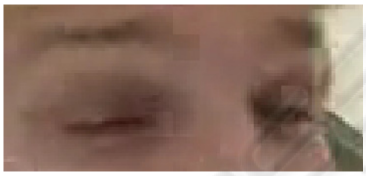
In this paper, we proposed a new prediction technique, which modified the inter-layer residual prediction to compensate the performance penalty of the single-loop decoding by adding block-based bi-linear smoothing function to the inter-layer residual prediction process.

(a) Single-loop decoding mode.

(b) Single-loop decoding with smoothed reference prediction.

Figure 7: Visual examples: (a) single-loop decoding mode (b) single-loop decoding with smoothed reference prediction for 7th frame of football.

From experimental results, it was shown that the performance penalty due to the single-loop decoding constraint could be reduced by the proposed method and the improvements of subjective quality was also meaningful. In addition, we expierimented several filters with three filter coefficients as a smoothing filter, however no filters were superior to bi-linear smoothing filter. Finally, the proposed technique was adopted the sclable extension of H.264/AVC standard Working Draft.



a) Single-loop decoding mode.

(b) Single-loop decoding with smoothed reference prediction.

Figure 8: Visual examples: (a) single-loop decoding mode (b) single-loop decoding with smoothed reference prediction for 36th frame of foreman.

REFERENCES

Schwarz, H., Hinz, T., Kirchhoffer, H., Marpe, D., Wiegand, T., Oct, 2004. Technical Description of the HHI proposal for SVC CE1. *ISO/IEC JTC1/SC29/WG11, M11244.*

Schwarz, H., Marpe, D., Wiegand, T., Dec, 2004. MCTF and Scalability Extension of H.264/AVC. *Proc. of PCS, San Francisco, CA, USA.*

Schwarz, H., Marpe, D., Schierl, T., Wiegand, T., July, 2005. Combined Scalability Support for the scalable Extensions of H.264/AVC. *Proc. of ICME*

Reichel, J., Schwarz, H., Wien, M(eds.)., Jan, 2005. Scalable Video Coding – Working Draft 1.

Joint Video Team (JVT), Doc. JVT-N020, Hong Kong, CN

Schwarz, H., Hinz, T., Marpe, D., Wiegand, T., Sep, 2005. Constrained Inter-Layer Prediction for Single-Loop Decoding in Spatial Scalability. *Proc. of ICIP, Genova, Italy.*

W. J. Han and S. Y. Kim, Jul, 2005. Smoothed reference prediction for single-loop decoding. *Joint Video Team (JVT), Doc. JVTO085, Poznan, Poland.*

W. J. Han and S. Y. Kim, Jan, 2006. Modified IntraBL design using smoothed reference. *Joint Video Team (JVT), Doc. JVTR091, Bangkok, Thailand.*