IMPROVED MOTION COMPENSATION IN THE ENHANCEMENT LAYER FOR SPATIALLY SCALABLE VIDEO CODING

Rong Zhang and Mary L. Comer

School of Electrical and Computer Engineering, Purdue University,
West Lafayette, IN, USA, 47907

ABSTRACT

This paper describes an efficient inter-layer motion compensation approach for enhancement layer prediction in spatially scalable video coding. The proposed approach provides improved performance compared to the current motion compensation scheme in the SVC standard. The improvement in performance is achieved by adding a new mode for enhancement layer motion compensation. Our experimental results show that the proposed method achieves an improvement up to about 1 dB in a two-layer spatially scalable framework compared to current SVC enhancement layer prediction.

Index Terms— Video Coding, Spatial Scalability, Motion Compensation, SVC.

1. INTRODUCTION

Spatial scalability has become an attractive feature of a video bitstream in many applications due to varying network conditions and customer device capabilities. The coded video bitstream is desired to be partitioned in such a way that the base layer is decoded independently to form a lower resolution and the enhancement layers which contain additional data can be decoded as needed to provide higher resolutions.

Much research has been done on scalable video coding using either a subband decomposition coding framework or the hybrid coding framework [1, 2]. Hybrid video coding has been a dominant video coding technology for years and the state-of-the-art standard H.264/AVC [3] provides significant improvement for single layer video coding compared to other existing standards [4]. There is currently also an effort within the Joint Video Team (JVT) to develop a scalability extension (SVC) [2] to the H.264/AVC standard.

In spatially scalable coding, as the input of consecutive layers are actually from the same original video but with different resolutions, both the motion vectors and the motion-compensated residual frames are highly correlated between layers. Hence, instead of encoding each layer separately, inter-layer redundancies should be explored to achieve efficient compression. The MPEG-2 standard [5] uses the interpolated decoded base layer frame as one of the references, in addition to the decoded higher layer frames for motion compensating higher layer blocks. Another idea called inter-layer intra prediction in SVC uses the interpolated decoded base layer blocks as the higher layer prediction for those intra-predicted blocks, which usually reduces rate-distortion cost [2].

The inter-layer prediction mechanism in SVC spatial scalability includes the inter-layer intra prediction, inter-layer motion vector prediction and a relatively simple method for enhancement layer motion compensation named inter-layer residual prediction [2]. The inter-layer residual prediction chooses to either predict the high resolution pixels from the high resolution references, or predict the high resolution residue from the interpolated low resolution residue, which we call pyramid motion compensation in this paper. However, the method proposed in this paper also includes another approach of motion compensation prediction, the subband method, which, when adaptively combined with the current SVC motion compensation, provides better coding performance in certain cases.

The subband and pyramid motion compensation methods differ from each other in the way they use the base layer data to encode the enhancement layer [6, 7]. Each method is more efficient than the other for different coding parameters and video input. This paper represents an extension of our previous work in that we combine the subband and pyramid techniques and adaptively choose between the two using rate distortion optimization. In addition, we also describe a new method, extended edge prediction, to reduce the block artifacts of the enhancement layer prediction in the subband method. Our experimental results will show that the proposed method achieves a coding improvement of about $0.2 \sim 1$ dB in a two-layer spatially scalable coding framework for different sequences. One limitation of SVC motion compensation is that it does not provide significant improvement compared to inter-layer intra prediction when the base layer is encoded at high quality. But the proposed method works well in this case, providing an especially efficient spatial scalability solution for applications requiring high quality video.

2. PYRAMID AND SUBBAND MOTION COMPENSATION

The pyramid and subband motion compensation methods are motivated from frequency scalable video coding [8]. Our de-

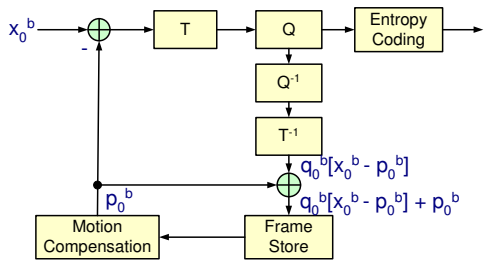

Fig. 1. Base layer encoding

tailed descriptions on the extension work to spatial scalability can be found in [6, 7]. Considering a two-layer spatially scalable encoder, the base layer is encoded by a non-scalable hybrid video coder as shown in Fig. 1. The prediction error $x_0^b - p_0^b$, after encoding and decoding, results in the base layer reconstructed prediction error $q_0^b[x_0^b - p_0^b]$, where x_0^b is the low resolution sequence and p_0^b is the base layer prediction.

The pyramid and subband motion compensation methods for enhancement layer are illustrated in Figs. 2 and 3. As shown in the figures, the base layer decoded prediction error $q_0^b[x_0^b - p_0^b]$, after interpolation, is used in both methods to predict enhancement layer prediction error. In certain cases, this prediction is effective and makes the prediction error difference have a lower entropy. But if the prediction error in the higher layer mainly contains high frequency components, it may not work well because these high frequency components cannot be predicted from the lower resolution base layer data. Therefore, encoding the enhancement layer prediction residual with no prediction from the base layer is also included as a candidate besides pyramid and subband methods in the proposed approach, as it is in SVC. Additionally, as shown in Fig. 3, the prediction used in subband method consists of the high resolution portion of the enhancement layer prediction $p_{1,1}$ and the interpolated base layer prediction p_0.

The analysis of the transformed prediction residue X_c in [7] has shown that whether pyramid or subband method is better depends on the corresponding value of $Q_0[X_0 - P_0]$ (the interpolated version of $q_0^b[x_0^b - p_0^b]$ in the transform domain). For those coefficients X_c with $Q_0[X_0 - P_0] \neq 0$, we have

$$X_{c,p} = X_{1,1} - P_{1,1} - E_0 + P_0 - P_{1,0}, \quad (1)$$

$$X_{c,s} = X_{1,1} - P_{1,1} - E_0, \quad (2)$$

where E_0 is the interpolated base layer quantization error

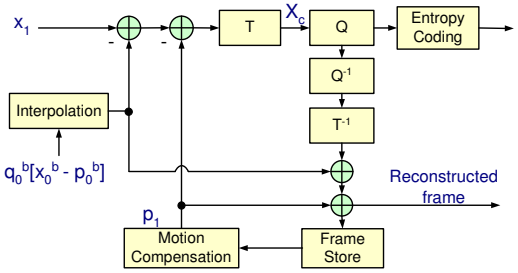

Fig. 2. Enh. layer encoding of pyramid method

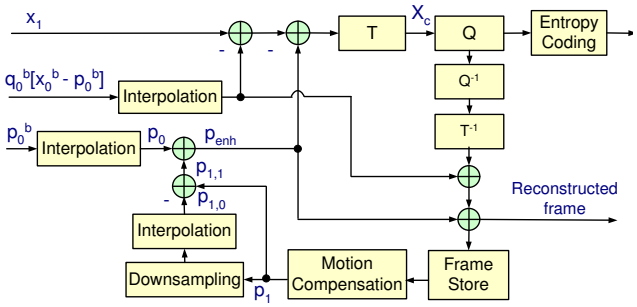

Fig. 3. Enh. layer encoding of subband method

which satisfies $Q_0[X_0 - P_0] = X_0 - P_0 + E_0$. Assuming the refinement of the low resolution data $P_0 - P_{1,0}$ is uncorrelated with E_0, the pyramid value $X_{c,p}$ is expected to have a higher entropy than the subband value $X_{c,s}$, which indicates the pyramid value will consume more bits than the subband value when $Q_0[X_0 - P_0] \neq 0$.

On the other hand, if $Q_0[X_0 - P_0] = 0$, we have

$$X_{c,p} = X_1 - P_1, \quad (3)$$

$$X_{c,s} = X_1 - (P_0 + P_{1,1}). \quad (4)$$

For this case, $X_{c,p}$ is predicted using the higher quality prediction P_1, whereas $X_{c,s}$ is predicted using the higher quality prediction for the high resolution part ($P_{1,1}$) and the lower quality prediction for the low resolution part (P_0). Therefore, the pyramid approach is expected to outperform the subband approach since higher quality prediction usually predicts better than lower quality prediction [8].

The analysis above shows that selecting the best method may improve coding efficiency and the choice can be made at the macroblock level or even more complicated at the coefficients level. In fact, when the quantization step (QStep) in the base layer is smaller, i.e., the base layer is encoded at a higher bitrate, many of the base layer coefficients $Q_0^b[X_0^b - P_0^b]$ would be quantized into nonzero value. In this case, subband method should outperform pyramid method. Intuitively, if the base layer is finely encoded and the enhancement layer has a relatively larger QStep, the interpolated base layer prediction p_0 may have a smaller sum of absolute difference (SAD) than the low resolution portion of the higher layer prediction $p_{1,0}$ because of the lower quality of the higher layer references. But when the QStep in the base layer is larger, many of the base layer coefficients $Q_0^b[X_0^b - P_0^b]$ are quantized to zero and pyramid method outperforms subband method. It should be noted that the downsampling/interpolation cascade shown in Fig. 3 could be combined into one filtering operation, reducing the computational complexity of the subband method.

3. THE PROPOSED MOTION COMPENSATION

The proposed inter-layer motion compensation approach selects the best method with the lowest rate-distortion cost at the macroblock level from the three methods, pyramid, subband

or encoding enhancement layer residuals with no base layer prediction as discussed in Sect. 2.

As mentioned in Sect. 1, for the subband motion compensation, the extended edge prediction is used to reduce the block artifacts of the predicted frame $p_{enh} = p_0 + p_{1,1}$. This technique is motivated from the extension work of reduced resolution update (RRU) video coding presented in [9]. In the subband method, for a given macroblock, the prediction from the higher resolution references p_1 is downsampled and then interpolated in order to get the high resolution portion $p_{1,1}$. During the interpolation process, the first and last rows and columns are actually extrapolated instead of interpolated because the outside block samples may not be available. This can contribute to severe blockiness, which makes the enhancement layer prediction p_{enh} not very accurate and therefore degrades the coding performance.

Hence, in the extended edge motion prediction, for each $n \times n$ block, an extra row and column at each side are fetched from reference frames to form a $(n+2) \times (n+2)$ block during motion prediction. However, the motion vectors are obtained based on the $n \times n$ block data in the motion estimation process. The extended edge $(n+2) \times (n+2)$ block prediction is also available at the decoder. Therefore, the $n \times n$ highpass portion $p_{1,1}$ produced at the decoder is exactly the same as at the encoder. Figure 4 gives an example of the *foreman* sequence, which includes the predicted frame p_{enh} with normal prediction and the predicted frame p_{enh} with extended edge prediction used in the subband method. We can see from the figure that p_{enh} obtained using the extended edge method has much fewer block artifacts than the one without using this particular technique.

(a) without ext. edge pred. (b) with ext. edge pred.

Fig. 4. Enhancement layer prediction $p_{enh} = p_0 + p_{1,1}$ comparison for *foreman* sequence (cropped frame No. 8) for subband method

Since in the proposed approach the encoder selects the best motion compensation method, the side information indicating which method is chosen in the encoder should be transmitted to the decoder to ensure correctly decoding. Extra bits $'0'$ for independently encoding, $'10'$ for pyramid method and $'11'$ for subband method are sent to the entropy coder. In general, the motion compensation type of a macroblock is highly correlated with its neighbors. Hence, this side information is context-based encoded in the proposed approach similar to encoding the intra prediction mode in H.264/AVC context-based adaptive binary arithmetic coding (CABAC) [3].

4. EXPERIMENTAL RESULTS

In this section, the performance of the proposed approach will be presented. All methods included were implemented in the H.264/AVC JM9.8 and applied to the sequences: 352×288 *foreman*, 352×240 *garden* and *football*. For simplicity, the two-layer spatial scalability is implemented and the horizontal and vertical spatial scale factor between layers is 2. Some coding parameters of interest are: GOP structure IBBPBBP with an I frame every 30 frames, search range ± 16 and 5 references. In the base layer, the set of quantization parameter $QPI = QPP = QPB - 2$ was used, and in the enhancement layer, the parameter was $QPI = QPP - 1 = QPB - 3$.

To evaluate the coding performance, we compared the enhancement layer encoding of four schemes, simulcast, inter-layer intra prediction only, SVC inter-layer residual prediction plus inter-layer intra prediction, and the proposed method plus inter-layer intra prediction. Figures. 5-7 show the comparison of the average luma PSNRs versus the bitrates of the enhancement layer. Note that only the enhancement layer performance is compared since the base layer is encoded in the same way for all schemes. For each sequence, there are two graphs corresponding to two different base layer bitrates, a lower one and a higher one. Given a certain base layer bitrate, the enhancement layer is encoded several times under different QPs to obtain different bitrates and PSNR performances.

It can be seen from the figures that the proposed inter-layer motion compensation method improves coding performance by about $0.2 \sim 1$ dB for different sequences and coding parameters. For each sequence, the improvement increases as the base layer bitrate increases, which is because the higher quality the base layer has, the more base layer information can be used in the inter-layer motion compensation. Our proposed method provides the most significant improvement for high quality base layer coding, making it very useful for high-quality video applications requiring spatial scalability.

Note that we only compared with the spatial scalability of the SVC standard while SVC provides some other scalability functionalities. However, the proposed method could be implemented as a relatively minor extension to the current SVC. A final note is that we did not implement inter-layer motion vector prediction in any of our experiments, so the performance of the methods shown in Figs. 5-7 could be further improved compared to simulcast.

5. CONCLUSION

We proposed in this paper an efficient inter-layer motion compensation technique for enhancement layer in spatially scalable video coding, which chooses the optimal method at the macroblock level from the three candidates: pyramid method, subband method and no prediction from the base layer residual. Our experimental results showed that the proposed approach improved coding performance in the two-layer spa-

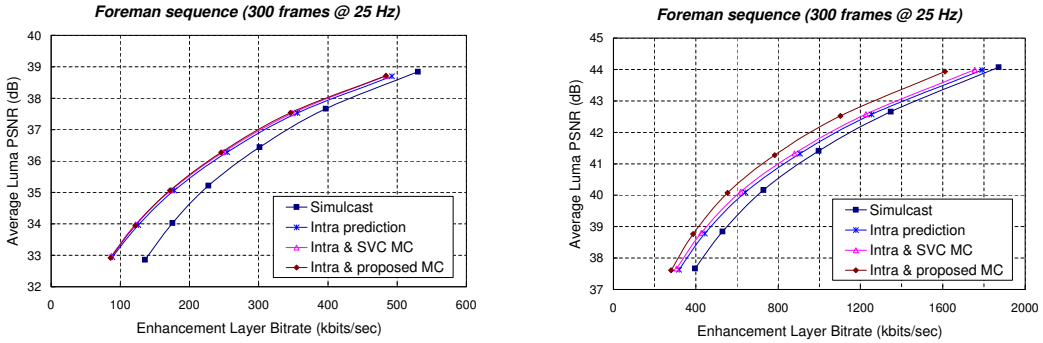

Fig. 5. *Foreman* (left: base layer 255.454 kbps (QPI=24); right: base layer 1314.502 kbps (QPI=12)

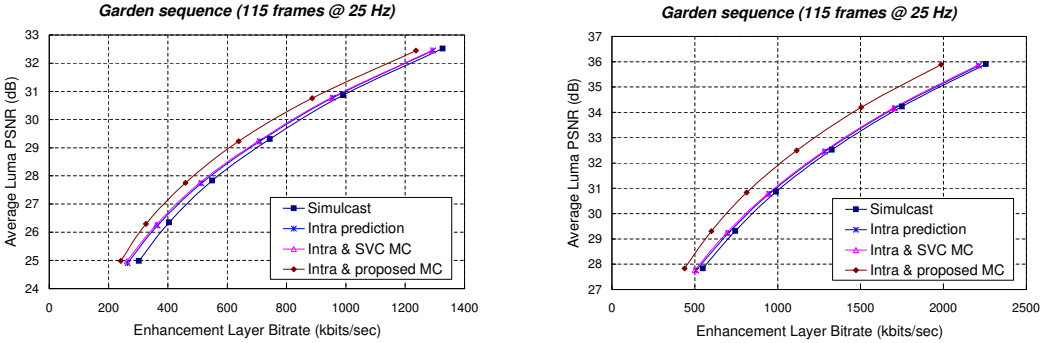

Fig. 6. *Garden* (left: base layer 703.743 kbps (QPI=26); right: base layer 2104.534 kbps (QPI=14)

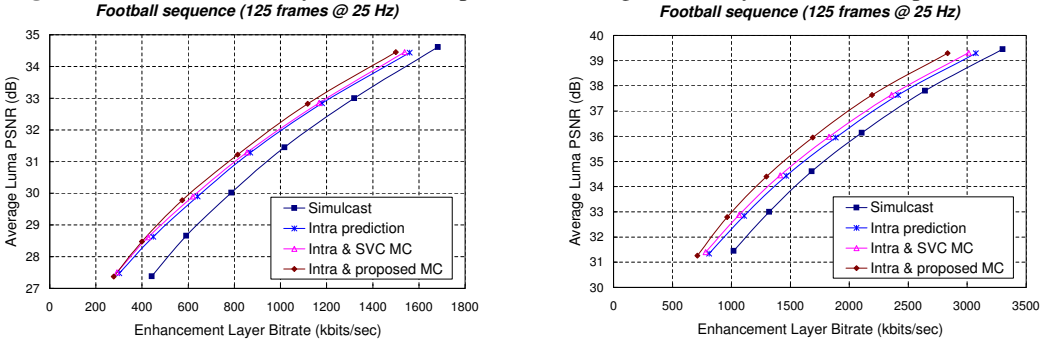

Fig. 7. *Football* (left: base layer 709.855 kbps (QPI=26); right: base layer 2356.01 kbps (QPI=12)

tially scalable coding compared to SVC. The future work will focus on developing inter-layer motion compensation method where the choice is made at the transform coefficient level.

6. REFERENCES

[1] R. Atta and M. Ghanbari, "Spatio-temporal scalability-based motion-compensated 3D subband/DCT video coding,"*IEEE Trans. on Circuits and Systems for Video Tech.*, vol. 16, pp. 43–55, Jan. 2006.

[2] "Scalable Video Coding - Working Draft 1," JVT of ITU-T VCEG and ISO/IEC MPEG JVT-N020, Jan. 2005.

[3] T. Weigand and G. Sullivan, "Draft text of final draft international standard (FDIS) of joint video specification (ITU-T Rec. H.264 — ISO/IEC 14496-10 AVC)," JVT of ISO/IEC JTC1/SC29/WG11 and ITU-T SG16/Q.6, Mar. 2003.

[4] T. Wiegand, G. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. on Circuits and Systems for Video Tech.*, vol. 13, pp. 560-576, July 2003.

[5] B. Haskell, A. Puri, and A. Netravali, *Digital Video: An Introduction to MPEG-2*. Chapman & Hall, 1996.

[6] M. L. Comer, "A new approach to motion compensation in spatially scalable video coding," in *Proc. of the SPIE Conf. on Visual Commun. and Image Proc.*, Jan. 2006.

[7] R. Zhang and M. L. Comer, "Subband motion compensation for spatially scalable video coding," in *Proc. of the SPIE Conf. on Visual Commun. and Image Proc.*, Jan. 2007.

[8] T. Tan, K. Pang, and K. Ngan, "A frequency scalable coding scheme employing pyramid and subband techniques," *IEEE Trans. on Circuits and Systems for Video Tech.*, vol. 4, no. 2, pp. 203-207, Apr. 1994.

[9] M. L. Comer, "Efficient reduction of block artifacts in reduced resolution update video coding," *IEEE Trans. on Circuits and Systems for Video Tech.*, vol. 16, pp. 386-395, Mar. 2006.