LETTER

Improvement of Inter-Layer Motion Prediction in Scalable Video Coding

Tae Meon BAE†, Truong Cong THANG†, Nonmembers, and Yong Man RO†(a), Member

SUMMARY In this letter, we propose an enhanced method for inter-layer motion prediction in scalable video coding (SVC). For inter-layer motion prediction, the use of refined motion data in the Fine Granular Scalability (FGS) layer is proposed instead of the conventional use of motion data in the base quality layer to reduce the inter-layer redundancy efficiently. Experimental results show that the proposed method enhances coding efficiency without increasing the computational complexity of the decoder.

Key words: MPEG-4 AVC, H.264, scalable video coding, inter-layer prediction, fine granular scalability, FGS motion refinement

1. Introduction

Currently, ISO/IEC MPEG and ITU-T VCEG are jointly making a scalable video codec (SVC) standard that is based on the H.264/AVC [1]. The objective of SVC is to generate temporal, spatial, and SNR-scalable coded stream, which provides users with QoS (Quality of Service) guaranteed streaming independent of video consuming devices in the heterogeneous network environments.

However, SVC sacrifices coding efficiency to enable scalability. One of the reasons for lowering coding efficiency is the layered coding structure of spatial scalability, which results in inter-layer redundancy. In order to reduce the redundancy due to the layered structure, inter-layer predictions for motion and texture/residual are used in SVC. But encoding with a multi-layer configuration still shows less coding efficiency than that with a single layer configuration [2], which means that there still exists inter-layer redundancy even with the current inter-layer coding scheme.

Recently, FGS motion refinement that performs motion estimation/compensation in the FGS layer has been proposed to enhance the performance of FGS codec [3], [4], and it is adopted as an optional encoding scheme in SVC. If FGS motion refinement is enabled, motion based inter-frame coding is performed with newly estimated motion data in the FGS codec. In SVC, Inter-layer prediction of texture and residual uses the data in FGS layers, but inter-layer motion prediction does not consider the motion in the FGS layer, because there was no available motion information in FGS layer. With FGS motion refinement, FGS layer has its own motion information, and the motion information could be a new candidate for the inter-layer motion prediction in addition to the motion information in the base quality layer. Because FGS layer is encoded with smaller value of QP (Quantization parameter) than the value of QP in the base layer, the estimated motion vectors in FGS layer will be more accurate than those estimated in the base layer. Therefore, if the motion vectors in FGS layer are used for inter-layer prediction, the coding efficiency will be increased.

Thus, in this letter, we exploit the structure of the codec and inter-layer motion prediction method when the motion information in the FGS layer of lower spatial resolution is used to predict the motion vectors of higher spatial resolution when motion refinement is enabled.

The letter is organized as follows. In Sect. 2, we will see the effect of QP on the motion estimation, which gives a hint to improve inter-layer motion prediction. The proposed inter-layer motion prediction using refined motion in FGS layer is explained in detail in Sect. 3. And experiments and their results to verify the efficiency of the proposed scheme are represented in Sect. 4. Finally, some concluding remarks on our work are given in Sect. 5.

2. Motion Prediction and Estimation in SVC

Motion estimation in H.264 and SVC is preformed based on the rate-distortion optimization [5], [6], and the rate-distortion cost for each macroblock mode is represented using the Lagrangian function as follows:

\[ J = D + \lambda_{\text{MOTION}} \times R(m - p), \]

(1)

where \( \lambda_{\text{MOTION}} \) is Lagrange multiplier, \( m \) is the motion vector, \( p \) is the prediction for the motion vector, and \( R(m - p) \) is the amount of bits required to encode the predicted motion vector error information. \( D \) is distortion measured by SSD (Sum of the Squared Difference) or SAD (Sum of the Absolute Difference).

In H.264/AVC, \( p \) is predicted from the motion vectors of the previously decoded neighboring macroblocks. The motion vectors may be similar to neighboring motion vectors with high probability; the motion prediction decreases the amount of motion information to encode.

The distortion \( D \) measured by SSD is represented by

\[ D = \sum_{(x,y) \in B} \left(f(x,y) - \hat{f}(x+m_x,y+m_y)\right)^2, \]

(2)

where \( f(x,y) \) is the luminance value at location \((x,y)\) in the
Thus if the QP value is increased, it is quantized and smoothed by in-loop deblocking filtering. A vector, and smoothed, which results in estimating degraded motion macroblock (or sub-macroblock) \( B \) of the current frame, and \( f(x+m_x,y+m_y) \) is the luminance value at location \((x+m_x,y+m_y)\) in the reference frame.

Because \( f(x,y) \) is in the reconstructed reference frame, it is quantized and smoothed by in-loop deblocking filtering. Thus if the QP value is increased, \( f(x,y) \) is more quantized and smoothed, which results in estimating degraded motion vector.

Lagrange multiplier \( \lambda_{MOTION} \) also affects the estimation of motion vector because it constrains the bitrate to encode motion vectors. The increase of \( \lambda_{MOTION} \) results in the decrease of the bit amount for motion vector, which causes the degradation of motion vector [7]. When SSD is used for distortion measure, \( \lambda_{MOTION} \) is calculated as follows in SVC [5]:

\[
\lambda_{MOTION} = 0.85 \times 2^{(QP-12)/3}.
\]  

As shown in Fig. 1, \( \lambda_{MOTION} \) is increased with the increase in QP, thereby the bit amount for motion vector decreases. Thus the increase in QP causes the degradation of motion vector.

The inter-layer prediction of motion vector is represented as \( p = U(b) \), where \( b \) is the motion vector of the base layer and \( U(\cdot) \) is the upscaling operation to fit the resolution ratio between the base and the enhancement layer. The motion vector of the inter-layer prediction mode is represented as

\[
m = U(b) + q
\]  

where \( q \) is quarter-sample motion vector refinement (-1, 0, or +1 for each motion vector component).

If QP of the base layer is increased, \( b \) is more degraded due to degraded reference frame and reduction of bit amount for motion vector, which results in increasing the difference between the motion vectors estimated in the enhancement layer and inter-layer prediction.

Therefore, motion vectors in the FGS layer will provide a better prediction for the motion vectors in the enhancement layer than the motion vectors in the base quality layer, because the QP value of the FGS layer is smaller than that of the base layer.

3. Inter-Layer Motion Prediction Using FGS Motion

Figure 2 shows the structure of SVC with proposed inter-layer motion prediction when two spatial resolution videos is encoded and SNR scalability is supported by FGS layers in the base layer.

The input video is decimated to a lower resolution, and it is encoded by H.264/AVC codec in the base layer. In the spatial enhancement layer, the original input video is encoded by H.264/AVC scalable extension codec, which performs inter-layer prediction of motion and texture/residual data in addition to the coding scheme of H.264/AVC.

The texture/residual data of the base and those of each FGS layer can be used for the prediction. In Fig. 2, quality level represents the layer used for inter-layer prediction among the base and FGS layers. High quality level requires high bitrate of lower resolution layer but guarantees more accurate inter-layer texture/residual prediction. Quality level is an input parameter of the encoder in SVC.

The shaded block in Fig. 2 is inter-layer motion estimation that uses the proposed inter-layer motion prediction, which is depicted in more detail in Fig. 3. The bold lines in Fig. 2 indicate new information inputted to the inter-layer motion prediction for the proposed method. With the FGS motion refinement, each FGS layer has its own motion information. Thus the motion information in FGS layers is newly inputted to the inter-layer prediction. And to select layer that is used for inter-layer motion prediction, quality level is also inputted to the inter-layer prediction.

In Fig. 3, inter-layer motion estimation is performed with the proposed method. The motion information of the layer selected by quality level is used for the inter-layer prediction. The macroblock partition of the selected layer is scaled according to the resolution ratio between the base and the enhancement layer. The basic concept of scaling mac-

![Fig. 1](image1.jpg)  
**Fig. 1** The value of Lagrange multiplier \( \lambda_{MOTION} \) according to the QP.

![Fig. 2](image2.jpg)  
**Fig. 2** The structure of the SVC codec with the proposed inter-layer motion prediction.
roblock partition is resizing the partition size according to the resolution ratio. The reference indices indicate the reference frame for motion compensation, and those of selected layer are directly used in the enhancement layer. The motion vectors of the selected layer are upscaled according to the resolution ratio between the enhancement and the base layer, and then quarter-sample motion vector refinement is searched. In the motion coding module in Fig. 2, the cost of the inter-layer motion prediction is compared with those of the other modes, and the mode that has minimum cost is used for macroblock encoding.

Inter-layer motion prediction is also performed in the decoder, and as shown in Fig. 3, no additional computational processing is required in the decoder except selecting motion information among base and FGS layers in lower resolution layer.

4. Experiment

The proposed method was implemented in JSVM (Joint Scalable Video Model) 6 [8]. The following SVC test sequences are used for the experiment: “Bus,” “Soccer,” “Crew,” and “City.” We measured the bitrate savings of the bitstreams provided by our approach with respect to the bitstreams provided by JSVM 6. In the experiment, a two layer configuration—{QCIF, 15 fps (frame per second)}, {CIF, 30 fps}—is used for encoding the sequences, and each spatial layer has 3 FGS layers. The QP values of the base layer and the enhancement layer are set to 40, and GOP (Group of Picture) size is set to 16. And quality level is set to the highest FGS layer.

Table 1 shows the reduction of bitrate for each base quality and FGS layer in the CIF layer. Because the proposed method only affects inter-layer prediction, the bitrate of the QCIF layer is unchanged. In Table 1, we can see that there are always some improvements at the CIF base quality layer up to 9.78%. And the overall bitrate saving is from 0.53% to 3.62%. Because the proposed inter-layer motion prediction works in the base quality layer of the CIF layer, the enhancement of coding efficiency is also observed in that layer.

<table>
<thead>
<tr>
<th>Layer</th>
<th>Sequence</th>
<th>Bus</th>
<th>Soccer</th>
<th>Crew</th>
<th>City</th>
</tr>
</thead>
<tbody>
<tr>
<td>CIF Base quality</td>
<td>6.85</td>
<td>2.59</td>
<td>9.78</td>
<td>3.43</td>
<td></td>
</tr>
<tr>
<td>CIF FGS1</td>
<td>0.66</td>
<td>0.14</td>
<td>2.62</td>
<td>-0.02</td>
<td></td>
</tr>
<tr>
<td>CIF FGS2</td>
<td>3.30</td>
<td>0.72</td>
<td>0.55</td>
<td>1.21</td>
<td></td>
</tr>
<tr>
<td>CIF FGS3</td>
<td>5.22</td>
<td>2.07</td>
<td>-0.50</td>
<td>3.35</td>
<td></td>
</tr>
<tr>
<td>Total (QCIF+CIF)</td>
<td>3.62</td>
<td>1.35</td>
<td>0.53</td>
<td>2.07</td>
<td></td>
</tr>
</tbody>
</table>

Table 2 Bitrate reduction (%) of CIF base quality layer by the proposed approach according to the FGS layer index of the QCIF layer used for inter-layer motion prediction.

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Bus</th>
<th>Soccer</th>
<th>Crew</th>
<th>City</th>
</tr>
</thead>
<tbody>
<tr>
<td>quality level</td>
<td>1</td>
<td>1.02</td>
<td>0.28</td>
<td>1.01</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>4.86</td>
<td>1.10</td>
<td>7.11</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>6.85</td>
<td>2.59</td>
<td>9.78</td>
</tr>
</tbody>
</table>

Fig. 4 Ratio of macroblocks that are set to inter-layer motion prediction mode (%).

Fig. 5 Rate-Distortion curves of JSVM 6 and proposed method for "Crew" sequence.

Because the QP value of the higher FGS layer is smaller than that of the lower layer, the motion information of the higher FGS layer will provide a better prediction for the motion vectors in the enhancement layer. Table 2 shows the coding efficiency of CIF base quality layer according to the quality level for inter-layer motion prediction, where
quality level \( n \) represents \( n \)-th FGS layer in QCIF layer. And it shows the increase of the quality level increases the coding efficiency of CIF base quality layer by the enhancement of inter-layer motion prediction.

We observed the ratio of macroblocks that are encoded with inter-layer motion prediction mode in CIF base quality layer. As shown in Fig. 4, the proposed method shows higher ratio than JSVM 6 for all test sequences, which means that inter-layer motion prediction by the proposed method is better than that by JSVM 6.

R-D (Rate-Distortion) curve is used to compare coding efficiency of different encoding methods by measuring PSNRs in the same bitrate. The bitrate of encoded bitstream in JSVM is adjusted to the target bit-rate by cropping the data of FGS when the frame rate and the spatial resolution are given. Figure 5 and Fig. 6 show the R-D curves of the “Crew” and “Soccer” sequences of the CIF layer with 30 fps, where the improvements of PSNR are up to 0.3 dB and 0.1 dB, respectively. In addition, the improvement of PSNR is shown over the entire bitrate range.

5. Conclusion

We proposed a new inter-layer motion prediction using refined motion in the FGS layer. Because motion vectors estimated with small QP value are more accurate than those estimated with high QP value, refined motion vectors in FGS layer are more suitable for inter-layer prediction than those in the base quality layer. In the experiment, the ratio of macroblocks encoded as inter-layer motion prediction mode is increased due to the improvement of inter-layer motion prediction, and this results in the improvement of coding efficiency. The proposed method does not require computational operation but additional switching operation in the decoder, thus it enhances coding efficiency without increasing the computational complexity of the decoder.

Acknowledgments

The authors would especially like to thank Dong Soo Lee for his contributions to this study.

References