Fast Mode Decision using Global Disparity Vector for Multiview Video Coding

Dong-Hoon Han, and Yung-Lyul Lee
Digital Media System Lab., Department of Computer Engineering, Sejong University, 98, Kunja-Dong, Kwangjin-Gu, Seoul, Korea
dhhan@dms.sejong.ac.kr, yllee@sejong.ac.kr

Abstract

Since multiview video coding (MVC) based on H.264/AVC uses a prediction scheme exploiting inter-view correlation among multiview video, MVC encoder compresses multiple views more efficiently than simulcast H.264/AVC encoder. However, in case that the number of views to be encoded increases in MVC, the total encoding time will be greatly increased. To reduce computational complexity in MVC, a fast mode decision using both MB (Macroblock)-based region segmentation information and global disparity vector among views is proposed to reduce encoding time. The proposed method achieves on average 40% reduction of total encoding time with the PSNR (Peak Signal to Noise Ratio) degradation of about 0.05 dB.

1. Introduction

Multi-view video coding (MVC), which is being standardized in the joint video team (JVT) of the ITU-T video coding experts group (VCEG) and ISO/IEC moving picture experts group (MPEG), is expected to become a new video coding standard for the realization of future video applications such as 3D-TV and free viewpoint video [1]. The MVC group in the JVT has selected the H.264/AVC [2]-based MVC method that was proposed by [3] as the MVC reference model, since this method showed better coding efficiency than H.264/AVC simulcast coding and the other methods that were presented in response to the call for proposals made by MPEG [4].

The main difference between conventional video coding and MVC is the availability of multiple camera views on the same scene. As coding efficiency of hybrid video coding depends on temporal prediction method, more coding gain in MVC can be achieved by inter-view (spatial) and temporal predictions.

For improving coding efficiency, H.264/AVC-based MVC uses a Rate-Distortion Optimization (RDO) technique in H.264/AVC to find the efficient coding mode for each macroblock (MB).

Figure 1 shows the variable block-size MB modes in H.264/AVC and Figure 2 shows the computation process of Rate-Distortion cost (RD cost) for variable block-size Inter MB mode, Intra MB modes of 4×4, 8×8, and 16×16, SKIP, and Direct mode in H.264/AVC.

In Figure 2, the H.264/AVC encoder calculates the RD cost of all MB modes in H.264/AVC and selects the best MB mode having the minimum RD cost, and this process is repeatedly carried out for each MB. Therefore, the computational complexity of H.264/AVC is increased, compared with the conventional video coding standard such as MPEG-2 video, H.263, or MPEG4 video.

Furthermore, MVC taking into account both temporal prediction in H.264/AVC and inter-view prediction among
views for improving coding efficiency needs much higher computational complexity than H.264/AVC.

In this paper, a fast mode decision scheme is proposed for MVC encoder to reduce computational complexity.

The paper is organized as follows. In Section 2, multiview video coding scheme is introduced briefly. In Section 3, the detailed algorithm of the proposed fast mode decision method is described. In Section 4, the performance of the proposed method is compared to that of the MVC reference model by analyzing the various experimental results. Finally, the conclusions are given in Section 5.

2. Feature of Multiview Video Coding

2.1. Hierarchical B picture and prediction structure

The increased flexibility of H.264/AVC in comparison to the previous video coding standards provides its improved coding efficiency. In contrast to the previous video coding standards, the coding and display order of pictures in H.264/AVC is completely decoupled. These features allow the selection of arbitrary coding/prediction structures, which are not supported by the previous standards. Hierarchical B picture structure [5] which is a typical example using structural flexibility is introduced through scalable video coding (SVC) [6]. This structure obtains higher coding efficiency than the existing IBBP structure. A typical hierarchical prediction structure with 4 dyadic hierarchy stages is depicted in Figure 3.

![Figure 3. Hierarchical coding structure with 4 temporal levels.](image)

The H.264/AVC-based MVC method that is chosen as a reference model [7] for the standardization of MVC has shown significant coding efficiency. Figure 4 depicts an example of the H.264/AVC-based MVC structure, in which there are eight parallel views. As shown in Figure 4, this structure utilizes the hierarchical B pictures, which not only improves the coding efficiency, but also provides temporal scalability. This structure can be divided into three kinds of picture sets, i.e., the picture set predicted by the inter-view pictures on the view axis, the picture set predicted by the temporal pictures on the temporal axis, and the picture set predicted by the view-temporal (spatio-temporal) pictures on the view and temporal axes. In this structure, the I_k and P_k pictures are only used at random access points. The B_k pictures are used between the I and P pictures that are random access points and are also used in all other positions except at these random access points, where the subscript k means the temporal decomposition level. The B_1 pictures on the temporal axis T0 are predicted spatially (inter-view), the B_k pictures on the view axes V0, V2, V4, and V6 are predicted temporally, and the B_k pictures on the view axes V1, V3, V5, and V7 are predicted temporally and spatially. To allow synchronization, I-frames start each GOP (S0/T0, S0/T8, etc.).

![Figure 4. Inter-view/temporal prediction structure using hierarchical B pictures.](image)

In the example above, a GOP-length of 8 is shown for coding scheme explanation, but for the coding experiments GOP-lengths of 12 and 15 were used.

2.2. Disparity Vector

Since there is redundancy information among each view in multiview video, disparity estimation (DE) is a key technology area to eliminate the view (spatial) redundancy between the views in MVC. Just like motion estimation (ME) to remove the temporal redundancy in a conventional video coding, DE can be used to eliminate the redundancy among the views in MVC. As an example, the views according to the camera positions are shown in ‘Exit’ and ‘Ballroom’ MVC sequences of Figure 5(a) and (b), respectively. It can be inferred that view 2 in Figure 5(b) and (d) is located right side of view 0 in Figure 5(a) and (c). The disparity information among the views can be estimated from DE process.
3. Fast Mode Decision using Global Disparity Vector

3.1. Region Partition

The H.264/AVC-based MVC performs ME/DE process based on the variable block-sizes shown in Figure 1. From the point of view of H.264/AVC codec, large block-based ME is usually performed on homogeneous areas and small block-based ME is performed on detailed areas containing edge information[8]. In addition to that, most video sequences contain moving objects with a stationary background, where object regions are mostly encoded in small block-size while stationary background regions are mostly encoded in large block-size like Inter16×16 block-size, Direct mode, or SKIP mode in inter frame coding. Table 1 shows the proposed segmentation of the background and objects block modes for fast mode decision in inter-view prediction.

Table 1. Segmentation of background and object block modes

<table>
<thead>
<tr>
<th>MB mode</th>
<th>Background mode</th>
<th>Object mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>P_Skip, B_Skip, Inter16x16</td>
<td>the other modes except background mode</td>
<td></td>
</tr>
</tbody>
</table>

In Table 1, an MB is decided as background block mode if a derived motion vector is smaller than 1/4 in integer pixel unit in case of Direct mode or Inter 16×16 mode, or the MB mode is P_SKIP or B_SKIP.

3.2. Motion Skip Mode and Global Disparity Vector

Global disparity vector (GDV) represents the difference of disparity among each view. Even though there are various ways to derive GDV, GDV for motion skip mode [9] that derivates GDV in MB-based unit (±16 integer pixel unit) from neighbor view already decoded is utilized in the proposed method as shown in Figure 7, in which full pixel-based disparity estimation is not necessary. Motion skip mode adopted in Joint Multiview Video coding Model (JMVM), which is the reference model in MVC, copies the motion information from that of the corresponding neighbor MB indicated by GDV in reference view (view 0) to that of the current MB in the current view (view 2).
Figure 7. Derivation process of global disparity vector in motion skip mode.

Figure 7 shows the derivation of GDV\textsubscript{cur}. The two GDVs, GDV\textsubscript{ahead} and GDV\textsubscript{behind}, of the anchor pictures are derived from two anchor pictures in an independent way, respectively, by using SAD (Sum of Absolute Difference) in an MB unit. GDV\textsubscript{cur} of the non-anchor picture between the anchor pictures is derived by Eq. (1), depending on POC\textsubscript{cur}, POC\textsubscript{ahead}, and POC\textsubscript{behind} where POC means picture order count (display order) on the time axis.

\[
GDV_{\text{cur}} = GDV_{\text{ahead}} \times \left( \frac{POC_{\text{cur}} - POC_{\text{behind}}}{POC_{\text{behind}} - POC_{\text{ahead}}} \right) \times (GDV_{\text{behind}} - GDV_{\text{ahead}})
\]  

(1)

3.3. Fast Mode Decision in View for Inter-view Prediction

Figure 8 shows the flowchart of the proposed method. Using region partitioning in Section 3.1, the region segmentation map separating object and background is generated from the pictures of the reference view. Regions of the views (e.g. V1, V2 in Figure 4) using inter-view prediction are estimated using MB-based GDV and region segmentation map of reference view V0.

Figure 9 shows an example of region segmentation estimation. Figure 9 (a) is the region segmentation map of the reference view where black blocks are decided to object regions. Figure 9 (b) is the region segmentation map which is estimated using GDV and the region segmentation map of the reference view in Figure 9 (a). Black blocks in Figure 9 (b) are estimated object region.

Through the estimated region information, RDO process is performed on only four modes as mentioned in Table 1 if the current MB is selected as background region. Otherwise, RDO is performed on all modes. In case of the former, RDO computation amount can be much reduced because RDO selects the best MB mode among only four background modes.

4. Experimental results

Four MVC test sequences are simulated, which has recommended for the MVC experiments to verify the performance of the proposed scheme [10]. The sequences, each of which has 2 GOPs in 3 views, are used for this experiment. The proposed method is compared with the MVC reference software JMVM 4.0. As shown in Figure 10 and Table 2, the proposed method reduces the encoding time approximately 40% compared with JMVM 4.0 in views using inter-view prediction (e.g. V1, V2 in Figure 4) with the average 0.05dB BD (Bjøntegaard delta)-PSNR[11] degradation.

<table>
<thead>
<tr>
<th>Sequence</th>
<th>QP</th>
<th>JMVM 4.0 Bitrate (kbps)</th>
<th>JMVM 4.0 PSNR (dB)</th>
<th>Proposed (P,B view) Bitrate (kbps)</th>
<th>Proposed (P,B view) PSNR (dB)</th>
<th>ΔBD-PSNR (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ballroom</td>
<td>22</td>
<td>1476.79</td>
<td>39.43</td>
<td>1494.27</td>
<td>39.41</td>
<td>-0.09</td>
</tr>
<tr>
<td></td>
<td>27</td>
<td>720.18</td>
<td>37.37</td>
<td>731.18</td>
<td>37.33</td>
<td>-0.07</td>
</tr>
<tr>
<td></td>
<td>32</td>
<td>368.40</td>
<td>34.85</td>
<td>374.67</td>
<td>34.79</td>
<td>-0.06</td>
</tr>
<tr>
<td></td>
<td>37</td>
<td>206.14</td>
<td>32.22</td>
<td>208.19</td>
<td>32.15</td>
<td>-0.04</td>
</tr>
<tr>
<td>Exit</td>
<td>22</td>
<td>804.92</td>
<td>40.35</td>
<td>807.70</td>
<td>40.33</td>
<td>-0.05</td>
</tr>
<tr>
<td></td>
<td>27</td>
<td>327.47</td>
<td>39.07</td>
<td>331.22</td>
<td>39.05</td>
<td>-0.04</td>
</tr>
<tr>
<td></td>
<td>32</td>
<td>166.72</td>
<td>37.32</td>
<td>168.50</td>
<td>37.28</td>
<td>-0.03</td>
</tr>
<tr>
<td></td>
<td>37</td>
<td>98.38</td>
<td>35.13</td>
<td>99.03</td>
<td>35.09</td>
<td>-0.02</td>
</tr>
<tr>
<td>Flamenco</td>
<td>22</td>
<td>1930.97</td>
<td>41.54</td>
<td>1935.90</td>
<td>41.52</td>
<td>-0.04</td>
</tr>
<tr>
<td></td>
<td>27</td>
<td>1028.76</td>
<td>38.69</td>
<td>1031.20</td>
<td>38.66</td>
<td>-0.03</td>
</tr>
<tr>
<td></td>
<td>32</td>
<td>538.03</td>
<td>35.64</td>
<td>537.27</td>
<td>35.59</td>
<td>-0.02</td>
</tr>
<tr>
<td></td>
<td>37</td>
<td>288.36</td>
<td>32.62</td>
<td>288.13</td>
<td>32.56</td>
<td>-0.01</td>
</tr>
<tr>
<td>Race1</td>
<td>22</td>
<td>1564.40</td>
<td>40.29</td>
<td>1566.12</td>
<td>40.28</td>
<td>-0.02</td>
</tr>
<tr>
<td></td>
<td>27</td>
<td>737.30</td>
<td>37.69</td>
<td>737.66</td>
<td>37.67</td>
<td>-0.01</td>
</tr>
<tr>
<td></td>
<td>32</td>
<td>369.63</td>
<td>35.02</td>
<td>369.81</td>
<td>34.99</td>
<td>-0.01</td>
</tr>
<tr>
<td></td>
<td>37</td>
<td>214.64</td>
<td>32.25</td>
<td>215.32</td>
<td>32.23</td>
<td>-0.01</td>
</tr>
<tr>
<td>Avg.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.05</td>
</tr>
</tbody>
</table>
5. Conclusions

This paper presents a fast mode decision method using both GDV and region partition in MVC. The fast mode decision is proposed to make use of correlation among the views for fast multi-view video coding. Experimental results show that the proposed scheme is able to achieve a reduction of about 40% encoding time on average with 0.05dB BD-PSNR degradation.

References


