An Efficient Mode Selection Exploiting Property of Region in Multi-view Video Coding

Kwangmu Shin, Sunghwan Chun, Kidong Chung

School of Computer Science & Engineering, Pusan National University
Busan, South of Korea
sin@pusan.ac.kr; { rainingman, kdchungm }@melon.cs.pusan.ac.kr
doi: 10.4156/ijipm.vol2.issue3.6

Abstract

A multi-view video coding exploits inter-view correlation between cameras to improve the coding efficiency. However, multi-view video coding has a problem of having higher complexity compared to the single-view video coding due to the above-mentioned property. In this paper, we propose a fast mode decision using region analysis to reduce the computational complexity of the encoding process. Experimental results show that computational time of about 15% ~ 22% could be saved compared to the JMVC 3.0. And it has only negligible coding efficiency losses.

Keywords: Multi-view Video Coding, Mode Selection, Property of Region

1. Introduction

A multi-view video is composed of captured videos by multiple cameras of the same scene. The multi-view video could be used in several video applications such as free-viewpoint television [1] and 3D-TV. But video coding method needs to be efficient because data is increased in proportion to the number of cameras.

The multi-view video coding (MVC) has encoding / decoding scheme with inter-view prediction. So the MVC uses not only various methods of removing data redundancy to exploit spatial, temporal and statistical correlations, but also methods of removing data redundancy to exploit inter-view correlation. As a result, the coding efficiency of MVC is heavily increased, however simultaneously computational complexity is very increased too. So several researches have been done to decrease the computational complexity in MVC [2][3][4][5][6][7][8].

In the paper, we propose a restricted inter-view prediction and a fast mode prediction using region segmentation to decrease computational complexity in MVC. This paper are organized as follows. Section 2 reviews simply multi-view video coding scheme, global disparity and regional disparity. Our proposed method is presented in Section 3. Experimental results are presented in Section 4. Finally, concluding remarks are described in Section 5.

2. Related Works

2.1. The encoding method for multi-view video

Currently most researches of MVC are progressing based on scalable hybrid prediction encoding. Disparity estimation about reference views can be processed in scalable hybrid prediction encoding by adding inter-view prediction. So it is possible to encode efficiently multi-view video only than using inter-frame and intra prediction. MVC prediction scheme could be influenced by coding efficiency, computational complexity and random access time etc.

"This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education, Science and Technology(No. 2009-0087838)"
2.2. Global disparity

The distance between cameras exists in multi-view video because of arrangement structure of several cameras. A disparity vector can be found by referencing a picture of other views of same time instances. It is called an inter-view prediction. The global disparity has global difference value between a picture of current view and a picture of another with same time instance.

A variety of methods to calculate global disparity were proposed. The method used in motion skip mode is representative. A motion skip mode is motivated by the idea that there is a high temporal correlation between views in the multi-view video [9]. The motion skip mode skips the motion information of the current view in encoding, and infers motion data using the motion data of the adjacent views in decoding.

![Figure 1. Calculating Global Disparity Vector (Anchor Picture)](image1)

Figure 1 describes how to calculate global disparity vector (GDV) in anchor picture. According to Figure 1, img0 and img1 are two images for calculating GDV. R represents an overlapped area.

2.3. Regional disparity

A regional disparity means each disparity of regions with different depth values. The regional disparity was proposed because the global disparity does not to predict effectively corresponding macroblock in adjacent reference picture.

![Figure 2. Example about Region Segmentation (Ballroom Sequence)](image2)

Figure 2 describes an example to segment into regions properly when view 3 references view 4 in Ballroom sequence.

3. Proposed Fast Mode Prediction

In this paper, we propose two methods using region analysis of image to decrease computational complexity. First, we propose a method to select efficiently a candidate mode using motion degree of each region in image. The image is segmented into four region using disparity vector. Second, we also propose method that does not use the inter-view prediction when macroblocks are occluded.
3.1. Fast mode prediction using region segmentation

3.1.1. Image segmentation using disparity vector

Generally a region with large motion has high possibility to be disparity compensated prediction mode. And the more the region is close to foreground, the more the motion is large. To check this condition, we need to distinguish between foreground and background regions in picture. So we segmented into four regions using corresponding disparity vectors of reference picture in non-anchor picture.

![Figure 3. Generating Disparity Map using from View 1 (Ballroom Sequence)](image)

Figure 3 describes to segment into discontinuous four regions using disparity vectors in Ballroom sequence. The black block of right picture in Figure 3 represents the farthest regions from camera.

3.1.2. Image segmentation using motion information

To distinguish between temporal and inter-view correlation, we exploit the influence of motion intensity about segmented regions. This has an influence on selecting the correlation of temporal or inter-view correlation. It is needed to readjust region information using motion vector of corresponding block in non-anchor picture.

![Figure 4. Probability Distribution of Inter-View Prediction Mode as Applying about Segmented Region (Ballroom and Race1 Sequence)](image)

Figure 4 describes the probability distribution of inter-view prediction mode by analysing segmented regions in Ballroom and Race1 sequence. We use disparity vector with proper search range.

In Figure 4, we use four motion vector sizes (8, 16, 24 and 32). The R0 ~ R3 indicate segmented regions using disparity vector. The M0 ~ M3 indicate segment regions using motion vector.

The red solid line (0 < M0 < 8) indicate that motion vector is the smallest. And the distribution of inter-view prediction mode of four regions is the smallest.
The black solid line (24 < M0 < 32) indicate that motion vector is the largest. So inter-view prediction mode probability is over about 50%. We can interpret the green solid line (8 < M0 < 16) and the blue solid line (16 < M0 < 24) with similar methods.

So we can readjust segmentation of region as Table 1. In this Table 1, The “IV” (inter-view) is in case that inter-view prediction mode is over 50%. If not, it is “I” (inter).

<table>
<thead>
<tr>
<th>MV</th>
<th>M0</th>
<th>M1</th>
<th>M2</th>
<th>M3</th>
</tr>
</thead>
<tbody>
<tr>
<td>R0</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>IV</td>
</tr>
<tr>
<td>R1</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>IV</td>
</tr>
<tr>
<td>R2</td>
<td>I</td>
<td>I</td>
<td>IV</td>
<td>IV</td>
</tr>
<tr>
<td>R3</td>
<td>I</td>
<td>IV</td>
<td>IV</td>
<td>IV</td>
</tr>
</tbody>
</table>

3.1.3. Fast prediction mode selection

Generally a block of region with large motion has a high possibility of having disparity compensated prediction mode.

Figure 5 describes how to select temporal or inter-view candidate mode according to each region. As we process candidate mode of block X of view 1, if it is “I”, it means that temporal correlation is higher. So the block can be compensated using motion prediction and the mode of temporally corresponding block, block of adjacent left, left-upper, upper and right-upper block become candidate mode.

3.2. Restricted inter-view prediction structure [10]

3.2.1. Calculating the region using global disparity

Our proposed method can get the global disparity vector by calculating disparity vector per block between two frames, not per pixel. The reason is that both calculation methods have similar rate-distortion performance and this method improves encoding time compared to the motion skip’s method.
We briefly explain our scheme below. First, we have to get the global disparity vectors in the anchor pictures. In this paper, we get macroblock sized the displacement. And this disparity vector is divided into numbers which denote the number of macroblocks selected as 16x16 inter-view modes.

\[
(g_x, g_y)_L = \frac{1}{C} \sum (d_x, d_y)_{L}, \quad (dx, dy)_{L} \in 16 \times 16_{DV_L}
\]

(1)

\[
(g_x, g_y)_R = \frac{1}{C} \sum (d_x, d_y)_{R}, \quad (dx, dy)_{R} \in 16 \times 16_{DV_R}
\]

(2)

In Eq. (1) (2), \((g_x, g_y)\) represents global disparity vector. \(L, R\) is left reference frame and right reference frame respectively. In other words, in the current view, we derive global disparity vectors from the left reference frame and right reference frame. These global disparity vectors are used for current frame. \(C\) represents a number of macroblocks that are selected 16x16 inter-view.

### 3.2.2. Restricted inter-view prediction

The restricted inter-view prediction scheme proposed in this paper is shown in Figure 6. Figure 6 (a) represents the frame which is encoded in B-frame and (b) and (c) represents the frames which are referenced by the frame in Figure 6. The square region which has a white vertical line on the right side of the frame at (a) does not have corresponding region in the left frame at (b). Also, the square region which has a vertical line on the left side of the frame at (a) does not have corresponding region in the right frame at (c). Therefore, macroblocks of view 0 and view 2 belong to such regions at Figure 6 (a) can be predicted by referencing to the corresponding macroblocks of frames of only one view instead of referencing two frames.

![Figure 6. Restricted Inter-View Prediction (Exit Sequence)](image)

### 3.3. Final algorithm

Figure 7 describes final algorithm of this paper. First of all, we find global disparity to calculate regions used in restricted prediction method in anchor pictures. We segment occluded regions using global disparity in non-anchor pictures. And the occluded regions are segmented into four using disparity vector. We readjust to regions with the high temporal or inter-view correlation considering motion degree in segmented regions. The following, motion compensated prediction is
performed if the regions have high temporal correlation. If not, we refer to occluded regions in non-anchor picture. We perform disparity compensated prediction with uni-direction if the blocks belong to occluded regions. If not, we perform disparity compensated prediction with bi-direction.

**Figure 7. Final Algorithm**

4. Experimental Results

To evaluate our proposed method, experiments have been done considering both complexity and coding efficiency. The employed sequences are Ballroom and Exit. We carried out our experimentations based on JMVC 3.0. The JMVC is a reference software which is the standard of experiments.

**Figure 8. Encoding Time Comparsion**
Figure 8 describes encoding time result of proposed method compared with JMVC 3.0. This figure shows encoding time percentage of Ballroom and Exit sequences compared with JMVC. The encoding time of Ballroom and Exit sequences which are applied with proposed method, each is about 85%, 78% respectively compared with JMVC.

### Table 2. Ballroom Sequence

<table>
<thead>
<tr>
<th>View No.</th>
<th>JMVC PSNR (dB)</th>
<th>JMVC Bitrates (kbps)</th>
<th>Proposed PSNR (dB)</th>
<th>Proposed Bitrates (kbps)</th>
<th>Bit-saving (%)</th>
<th>PSNR Gain (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>34.537</td>
<td>209.72</td>
<td>34.508</td>
<td>204.4</td>
<td>-2.4</td>
<td>-0.02</td>
</tr>
<tr>
<td>3</td>
<td>34.346</td>
<td>200.34</td>
<td>34.309</td>
<td>194.09</td>
<td>-3</td>
<td>-0.04</td>
</tr>
<tr>
<td>5</td>
<td>34.981</td>
<td>208.13</td>
<td>34.964</td>
<td>204.42</td>
<td>-1.96</td>
<td>-0.01</td>
</tr>
<tr>
<td>Avg.</td>
<td>34.61</td>
<td>206.06</td>
<td>34.59</td>
<td>200.97</td>
<td>-2.53</td>
<td>-0.02</td>
</tr>
</tbody>
</table>

### Table 3. Exit Sequence

<table>
<thead>
<tr>
<th>View No.</th>
<th>JMVC PSNR (dB)</th>
<th>JMVC Bitrates (kbps)</th>
<th>Proposed PSNR (dB)</th>
<th>Proposed Bitrates (kbps)</th>
<th>Bit-saving (%)</th>
<th>PSNR Gain (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>36.89</td>
<td>85.06</td>
<td>36.85</td>
<td>82.73</td>
<td>-2.81%</td>
<td>-0.04</td>
</tr>
<tr>
<td>3</td>
<td>36.976</td>
<td>94.21</td>
<td>36.927</td>
<td>92.18</td>
<td>-2.2%</td>
<td>-0.04</td>
</tr>
<tr>
<td>5</td>
<td>36.66</td>
<td>107.5</td>
<td>36.635</td>
<td>105.64</td>
<td>-1.76%</td>
<td>-0.02</td>
</tr>
<tr>
<td>Avg.</td>
<td>36.842</td>
<td>95.59</td>
<td>36.804</td>
<td>93.516</td>
<td>-2.21%</td>
<td>-0.03</td>
</tr>
</tbody>
</table>

Table 2 and 3 describe the coding efficiency result of Ballroom and Exit sequence. The PSNR represents objective quality of sequences and the bitrate represents coding efficiency. As a result, about 15% ~ 22% of computational time can be saved compared to the JMVC 3.0 only with negligible coding efficiency losses.

### 5. Conclusion

In this paper, we propose a fast mode decision using region analysis to reduce the computational complexity of the encoding process. Experimental results shows that computational time of about 15% ~ 22% was saved compared to the JMVC 3.0. And it has only negligible coding efficiency losses.

### 6. References


