An Efficient View Interpolation Scheme and Coding Method for Multi-view Video Coding

Cheon Lee, Kwan-Jung Oh, Seung-Hwan Kim, and Yo-Sung Ho
Gwangju Institute of Science and Technology (GIST)
1 Oryong-dong, Buk-gu, Gwangju, 500-712, Korea
Phone: (+82) 62-970 2258 Fax: (+82) 62-970 3164 E-mail: {leecheon, kjoh81, hoyo}@gist.ac.kr

Keywords: multi-view video coding, view interpolation, disparity estimation

Abstract - Multi-view video allows us to experience more realistic feelings with arbitrary view-points and a wide range of scenes, because those are captured by multiple cameras simultaneously but at different positions. Since the amount of data to be processed increases in proportion to the number of cameras, we need to develop efficient coding methods. In this paper, we propose an efficient view interpolation scheme and coding method for multi-view video using the correlation between views. After generating an intermediate view image, we use it at the encoder as an additional reference frame using the proposed 'VIP P-picture' coding method. The proposed view interpolation method improves the generated inter-view image about 1~4 dB and the 'VIP P-picture' coding method improves coding efficiency about 0.5 dB.

1. INTRODUCTION

As the interest for the realistic multimedia grows, various types of technologies have been researched such as stereoscopic display, holography, and multi-view video coding (MVC). Among them, the multi-view video is a collection of multiple videos obtained by capturing the same scene at different camera locations. It can provide arbitrary view-points of dynamic scenes and thus allows more realistic videos to users. This can be used in various applications including free viewpoint video (FVV), free viewpoint TV (FTV), 3DTV, surveillance, and home entertainment.

Because of the increased number of cameras, the multi-view video contains a large amount of data. It is serious limitation on data distributive applications, so we need to compress the multi-view sequences efficiently without significant sacrificing the visual quality [1-2]. Recently, 3DAV (3-D audio and visual) group on MPEG (moving picture experts group) and JVT (joint video team) group is working on the MVC standards. For this purpose, many algorithms have been proposed to improve the coding efficiency; illumination compensation, prediction structure, disparity vector coding, and view synthesis prediction. Especially, view synthesis prediction employs additional reference frames which are generated by using depth or disparity map. The method of using depth map exploits the epipolar geometry to match the corresponding lines. However, the view interpolation exploits disparity values for every pixel positions.

In this paper, we focus on the view interpolation method and describe both the previous view interpolation method and the proposed method. In addition, we propose 'VIP P-picture' coding method which is composed of five additional motion estimation modes and modified motion vector prediction method.

2. BACKGROUNDS OF VIEW INTERPOLATION

View interpolation method proposed by Chen can reconstruct arbitrary viewpoints using optical flow between two input images [3]. Droese proposed a view interpolation in disparity domain for multi-view video [4].

Disparity can be defined as a distance in horizontal coordinates of two corresponding pixels, which is described by Eq. (1).

$$I_L(x,y) = I_R(x+d,y) \tag{1}$$

where d stands for the disparity of a pixel between the left and right image, and it can be factorized into two disparities by α ($0 \le \alpha \le 1$), which is the position of the new view between two anchor images. The relationship for cameras is given as Eq. (2), where $\lfloor\ \rfloor$ is a rounding to integer values. $I_a(x,y)$ stands for intensity of intermediate view position.

$$\begin{aligned} I_\alpha(x,y) &= I_L(x+\lfloor \alpha d \rfloor, y) \\ &= I_R(x+\lfloor (1-\alpha) d \rfloor, y) \end{aligned} \tag{2}$$

Equation (3) represents the cost function which forms the relation between the similarity measure and the neighborhood regularization of the disparity map D. C_{sim} stands for a simple block-matching method using MAD (mean absolute difference), and Eq. (4) is an average disparity value of neighboring four disparities. λ controls the influence of two terms.

$$C(x,y,d) = C_{sim}(x,y,d) + \lambda \cdot C_{reg}(x,y,d) \tag{3}$$

$$C_{reg}(x,y,d) = \frac{1}{4}\left[\begin{array}{l} |D(x-1,y)-d| + |D(x-1,y-1)-d| + \\ |D(x,y-1)-d| + |D(x+1,y-1)-d| \end{array}\right] \tag{4}$$

3. PROPOSED VIEW INTERPOLATON METHOD

Above view interpolation method estimates disparities according to the range of pre-determined maximum disparity value. It is obvious that a proper selection of maximum disparity value can contribute to the accuracy of disparity estimation. However, if the pre-determined maximum value for disparity estimation is smaller than the actual maximum disparity value, estimator may decide wrong disparity value in large actual disparity value regions. To avoid this, we propose an initial disparity estimation method using region dividing. On the other

hand, some erroneous disparities occur by exploiting fixed block size, so we use variable block. In addition, we use disparity error correction process. Figure 1 shows the whole process of proposed method [5].

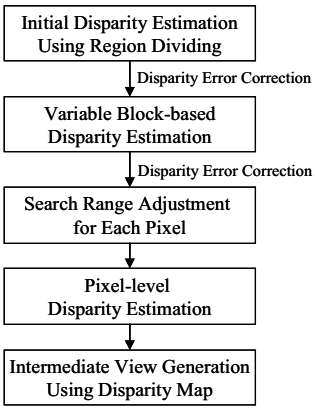


Fig. 1. Proposed view interpolation process

3.1 Initial Disparity Estimation

The objective of initial disparity estimation is to find approximate disparity values without pre-determined maximum disparity setting. A convenient property of stereo images captured by parallel cameras is ordering constraint [6]. It guides the positions of objects. For instance, when object A is located on the right side of object B in left image, the corresponding object A' is always located on the right side of the object B' in right image. We matched blocks using pixel difference values computed by Eq. (5), because it is less sensitive to the illumination mismatches between views.

$$D = I(x+1, y) - I(x, y) \tag{5}$$

Figure 2 describes the initial disparity estimation process. Block 1 has the most outstanding differences. It is efficient for finding the disparity value firstly. Therefore, estimator decides a disparity with search range [*0*, *width*]. The next block to be estimated is block 2, which is estimated the disparity using the search range (x_1+d_1, *width*] adjusted by considering the previous disparity value of block 1. The rest of the blocks are estimated in this manner according to the adjusted search ranges.

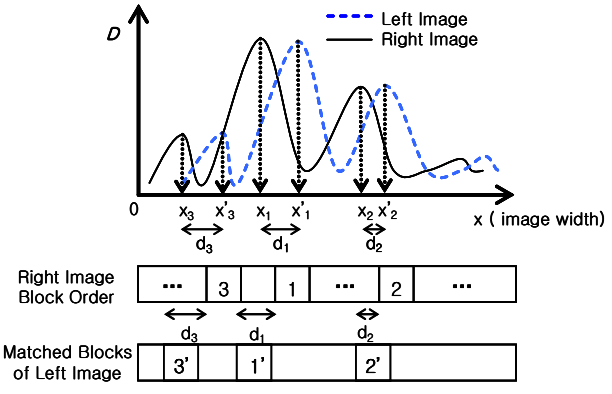


Fig. 2. Initial disparity estimation

3.2 Variable Block-based Disparity Estimation

Since the previous method uses invariant block for block matching, the estimator may detect wrong disparity value when a reference block covers boundary region. To avoid this, we propose a variable block-based disparity estimation process as shown in Fig. 3. Once a disparity is determined for a basic block, estimator checks the difference of costs between the larger block and the smaller block. If the difference is larger than pre-determined threshold *T*, it estimates disparity again for the smaller block. The minimum block size is 2x2. This method is effective for the boundary regions having large disparity discontinuity.

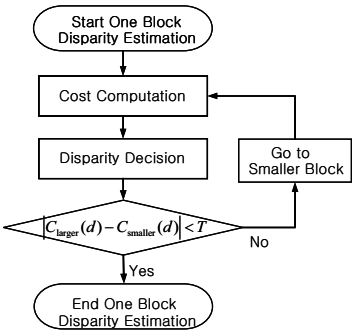


Fig. 3. Variable block-based disparity estimation

3.3 Pixel-level Disparity Estimation and Error Correction

The final step of disparity estimation is a pixel-level estimation. The search range can be adjusted by using Eq. (6) considering the obtained disparity map in previous steps. The search range does not need to be large enough to cover the maximum disparity value. The procedure of pixel-level disparity estimation is similar to the previous method but the cost function is modified by adding C_{StDev} term as described in Eq. (7). C_{StDev} term stands for the standard deviation of pixels in a block.

$$\begin{aligned} \text{MinRange} &= D(x, y) - \text{SearchRange}/2 \\ \text{MaxRange} &= D(x, y) + \text{SearchRange}/2 \end{aligned} \tag{6}$$

$$\begin{aligned} C'(x, y, d) = C_{sim}(x, y, d) &+ \lambda \cdot C_{reg}(x, y) \\ &+ \gamma \cdot C_{StDev}(x, y, d) \end{aligned} \tag{7}$$

The proposed view interpolation method exploits disparity error correction methods to increase the accuracy and consistency of disparities. If an object has different disparity values, some of them may be erroneous disparity values. By checking cost values using the lager block, those disparities can be corrected by replacing neighboring disparity value. The median filtering is another useful tool for reducing noisy disparities.

3.4 View Synthesis using Disparity Map

An intermediate view image can be synthesized by Eq. (8), which assumes that the illumination changes linearly. $\hat{I}$ is interpolated value when $\alpha \cdot D(x, y)$ or $(\alpha - 1) \cdot D(x, y)$ is not an integer.

$$I_\alpha(x,y) = (1-\alpha)\cdot \hat{I}_L(x+\alpha\cdot D(x,y), y) + \alpha\cdot \hat{I}_R(x+(\alpha-1)\cdot D(x,y), y) \quad (8)$$

4. PROPOSED 'VIP P-picture' CODING

The basic coding structure of MVC is H.264/AVC, and it uses hierarchical B picture coding in temporal direction as shown in Fig. 4 [7]. In order to use adjacent views as reference frames, S1, S3, and S5 sequences are coded after encoding of surrounding views. View interpolation process can be applied to those inter-view sequences because those are available for disparity estimation when the surrounding views have been encoded.

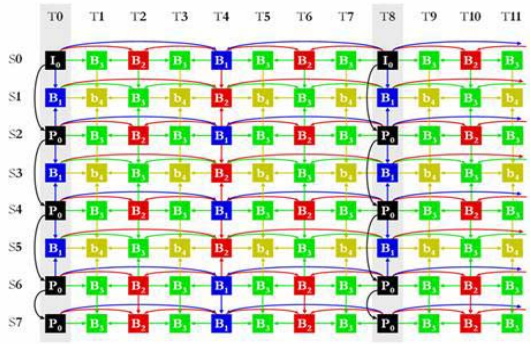


Fig. 4. Reference prediction structure of MVC

4.1 Additional Estimation Modes

View interpolation method can be applied to inter views such as S1, S3, and S5 in Fig. 4, and it is possible when encoding of adjacent views is finished. The interpolated images can be used as additional reference frames. One of advantages of the intermediate image is that it is mostly overlapped frame with the frame to be coded. Thus, we propose a view interpolated prediction P-picture (VIP P-picture) coding method which refers only the intermediate frame. 'VIP P-picture' coding is composed of five additional modes: 'VIP_SKIP', 'VIP_16x16', 'VIP_8x16', 'VIP_16x8', and 'VIP_P8x8'. VIP_SKIP mode refers the co-located block of the intermediate image and it does not need to encode motion vectors and residuals.

4.2 Modified Motion Vector Prediction

The H.264/AVC encodes the difference value between the real motion vector of the current macroblock and the predicted motion vector using neighboring macroblocks. The reason is that all reference frames have the similar correlation to the frame to be coded in H.264/AVC. Therefore, the proposed 'VIP P-picture' coding scheme classifies reference frames into temporal frame (T frame), spatial frame (V frame), and VIP frame. Since there are three types of reference frames in reference lists, the traditional motion vector prediction method is not efficient. Therefore, we propose a modified motion vector prediction method which predicts a motion vector more efficiently by classifying the types of frames. Figure 6 is a flowchart of the modified motion vector prediction methods. If the estimation mode is one of the traditional modes, motion vector is predicted only from neighboring blocks which refer to V/T frames. On the other hand, if the mode is one of VIP modes, motion predictor considers blocks referred VIP frame. This method can reduce the length of motion vector to be encoded.

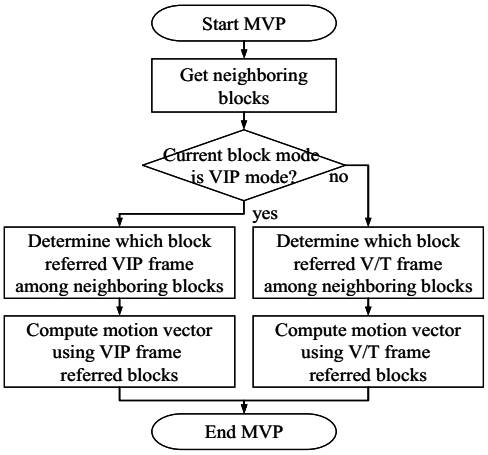


Fig. 5. Modified motion vector prediction

5. EXPERIMENTAL RESULTS

In order to evaluate the proposed view interpolation scheme and multi-view video coding method, we compared the generated intermediate images in terms of PSNR and coding efficiency of multi-view video coding. We used the reference software JMVM 1.0 (joint multi-view video model) provided by MPEG/JVT. 'Akko&Kayo', 'Rena', and 'Ballroom' sequences are used for interpolating an intermediate view image and for encoding of multi-view video.

We used 16x16 block size for block-based disparity estimation. The threshold of Fig. 3 is 15, and the search ranges are 5, 10, and 15 in pixel-level disparity estimation. The size of median filter is 10x10. The reference coding structure is based on Fig. 4, and it is represented in [8]. The simulated QPs are 27, 32, and 37. CABAC is used for entropy coding. The search range is 96 for the traditional motion estimation modes and 48 for additional VIP modes.

5.1 Results of View Interpolation Scheme

Table 1 shows the average PSNRs for each sequence. As a result, the proposed method have improved the quality of interpolated images about 1~4 dB. 'Akko&Kayo' and 'Rena' sequences were better than 'Ballroom' in terms of PSNR. It is because those two sequences have less occlusions and disparity levels than 'Ballroom' sequence.

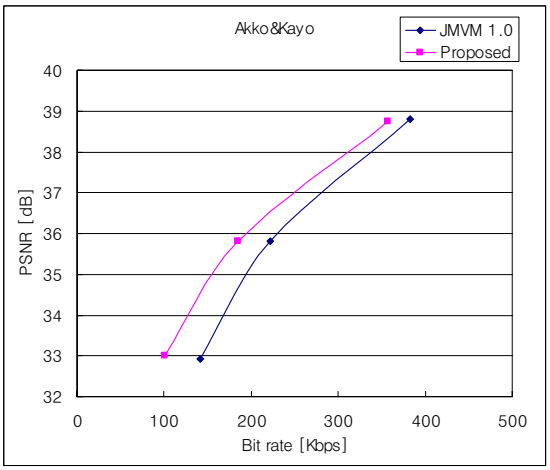
Table 1. View interpolation results: Avg. PSNR of 30 frames

Test sequences	Previous method			Proposed method		
	Max. search range			Search range		
	30	40	50	5	10	15
Akko&Kayo	27.8	31.5	30.4	33.0	32.7	32.3
Rena	28.4	27.5	26.4	32.6	32.7	32.8
Ballroom	20.7	21.0	21.4	25.3	25.3	25.3

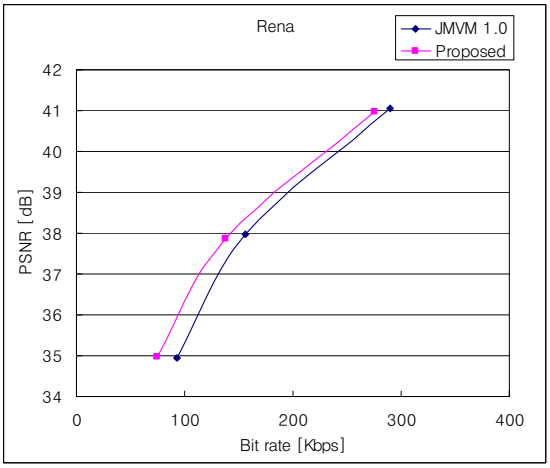
Unit: dB

5.2 Results of Multi-view Video Coding

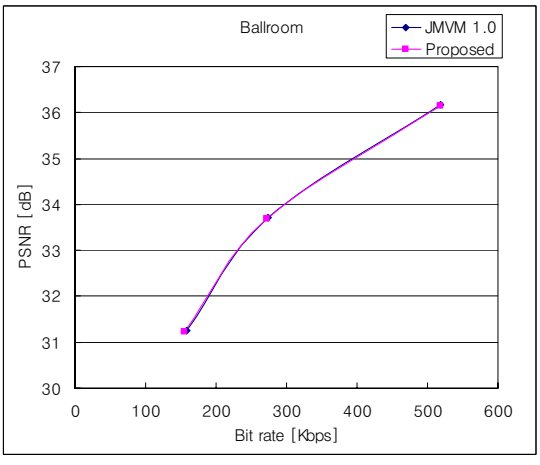
Because the reference model of MVC exploits the inter-view direction estimation, the proposed 'VIP P-picture' coding can be applied to several in-between sequences. 6 in 15 views of 'Akko&Kayo', 7 in 16 views of 'Rena', and 3 in 8 views of 'Ballroom' are adopted to the proposed coding method respectively. Figure 6 shows experimental results of MVC.



(a) Rate-distortion curves for Akko&Kayo



(b) Rate-distortion curves for Rena



(c) Rate-distortion curves for Ballroom

Fig. 6. Results of multi-view video coding

As shown in Fig. 6, PSNR values of 'Akko&Kayo' and 'Rena' sequences are improved over 0.5 dB, but PSNR values of 'Ballroom' are quite similar to the reference model. The reason is that the quality of generated images is not good enough to contribute to the coding efficiency. VIP frames of 'Ballroom' sequence are not good enough to contribute to coding efficiency in terms of PSNR. This results show that well interpolated image can contribute to the coding efficiency.

6. CONCLUSION

In this paper, we have described an efficient view interpolation scheme and multi-view video coding method using the interpolated image. The proposed view interpolation method consists of initial disparity estimation, variable block-based estimation, and disparity error correction. These can improve the quality of the generated view image compared to the previous method. Based on this method, we proposed the 'VIP P-picture' coding method which employs the intermediate image as an additional reference frame. It is composed of five additional modes referring to interpolated images and modified motion vector prediction scheme. As a result, the interpolated images with good quality can contribute to the coding efficiency over 0.5 dB on average.

ACKNOWLEDGEMENTS

This work was supported in part by MIC through RBRC at GIST, in part by MOE through the BK21 project, and in part by MCIE through the project of core technology development.

REFERENCES

[1] A. Smolic and P. Kauff, "Interactive 3D Video Representation and Coding Technologies," *Proceedings of the IEEE*, Spatial Issue on Advances in Video Coding and Delivery, Vol. 93, pp. 99-110, 2005.

[2] A. Smolic, K. Mueller, T. Rein, P. Eisert, and T. Wiegand, "Free Viewpoint Video Extraction, Representation, Coding, and Rendering," *Proc. of IEEE International Conference on Image Processing,* Vol. 5, pp. 3287-3290, 2004.

[3] S. Chen and L. Williams, "View Interpolation for Image Synthesis," *Computer Graphics (SIGGRAPH '93)*, pp. 279-288, 1993.

[4] M. Droese, T. Fujii, and M. Tanimoto, "Ray-Space Interpolation based on Filtering in Disparity Domain," *Proc. 3D Conference*, pp. 213-216, 2004.

[5] JVT of ISO/IEC MPEG & ITU-T VCEG, "View Interpolation for Multi-view Video Coding," U102, 2006.

[6] A. L. Yuille and T. Poggio, "A Generalized Ordering Constraint for Stereo Correspondence," *MIT A.I. Laboratory Memo 777*, 1984.

[7] ISO/IEC JTC1/SC29/WG11, "Description of Core Experiments in MVC," N8019, 2006.

[8] JVT of ISO/IEC MPEG & ITU-T VCEG, "Common Test Condition for Multiview Video Coding," U211, 2006.