## Poster Session ThursAM3: Multiview Video Coding

**Location:** Poster Room
**Time:** 10:50 - 12:00

### VIEW INTERPOLATION PREDICTION FOR MULTI-VIEW VIDEO CODING

Cheon Lee, Kwan-Jung Oh, Yo-Sung Ho, *Gwangju Institute of Science and Technology, South Korea*

**Abstract:** Since multi-view video is a collection of videos captured by a multiple camera array, the volume of data is huge. Various algorithms have been developed for multi-view video coding. In this paper, we propose two methods: efficient view interpolation and 'VIP P-picture' coding. The view interpolation method includes initial disparity estimation, variable block-based matching, and pixel-level disparity estimation operations. The 'VIP P-picture' coding method is an additional motion estimation process at the encoder. The proposed view interpolation method improves objective quality of generated images up to about 1 4 dB, and the 'VIP P-picture' coding method increases the average coding gain about 0.66 dB in well synthesized sequences.

### OPTIMAL SUBBAND BIT ALLOCATION FOR MULTI-VIEW IMAGE CODING WITH DISPARITY COMPENSATED WAVELET LIFTING TECHNIQUE

Wuttipong Kumwilaisak, Pongsak Lasang, *KMUTT, Thailand*

**Abstract:** This paper presents the optimal subband bit allocation based on a new rate distortion (R-D) model for multi-view image coding with disparity-compensated wavelet lifting. First, the distortion prediction of the reconstructed multi-view image with lifting scheme is presented. A new rate distortion model combining the exponential and power model is developed. Then, the analyzed prediction error and rate distortion model are used in the optimal bit allocation framework. The bit allocation framework allocates bits to all subbands with the goal to minimize distortion of the reconstructed multi-view images. Low-pass and high-pass subbands are compressed by SPIHT [5] with the optimal bit solution. We verify the proposed method with several test multi-view images. Simulation results show that the bit allocation based on the proposed method provides close results to the exhaustive search method in both allocated bits and PSNR. It also outperforms the uniform bit allocation over a wide range of target bit rate.

### DIAGONAL INTERVIEW PREDICTION FOR MULTIVIEW VIDEO CODING

Ping Yang, Yun He, *Tsinghua University, China*

**Abstract:** Multi-view video receives many attentions in these years, because it can support a wide range of applications, such as 3D video communication and free view point video. To improve the coding efficiency, multi-view video coding not only employs temporal predictions but also predictions between different views. This paper investigates the parallel computing, low delay and scalability problems induced from the temporal-spatial predictions, proposes a diagonal interview prediction (DIP) structure. It is shown in the test results that the DIP can exploit the correlations between different views largely, it can support parallel computing, low delay and scalable decoding features, and it obtains a high coding gain especially for the large motion cases.

# VIEW INTERPOLATION PREDICTION FOR MULTI-VIEW VIDEO CODING

*Cheon Lee, Kwan-Jung Oh, and Yo-Sung Ho*

Gwangju Institute of Science and Technology (GIST),
1 Oryong-dong, Buk-gu, Gwangju, 500-712, Republic of Korea
*{leecheon, kjoh81, hoyo}@gist.ac.kr*

## ABSTRACT

Since multi-view video is a collection of videos captured by a multiple camera array, the volume of data is huge. Various algorithms have been developed for multi-view video coding. In this paper, we propose two methods: efficient view interpolation and 'VIP P-picture' coding. The view interpolation method includes initial disparity estimation, variable block-based matching, and pixel-level disparity estimation operations. The 'VIP P-picture' coding method is an additional motion estimation process at the encoder. The proposed view interpolation method improves objective quality of generated images up to about 1~4 dB, and the 'VIP P-picture' coding method increases the average coding gain about 0.66 dB in well synthesized sequences.

***Index Terms*—** multi-view video coding, disparity, view interpolation

## 1. INTRODUCTION

Improvements of technologies and speed for transmitting data through internet enable diverse forms of multimedia contents. Recently, demands for interactive contents and realistic contents are growing rapidly. Multi-view video is a good alternative to satisfy these demands because it can provide arbitrary view-points of dynamic scenes. This can be used in various applications including free viewpoint video (FVV), free viewpoint TV (FTV), 3DTV, surveillance and home entertainment.

Although multi-view video allows more realistic scene to users, the volume of data is huge. It is a serious limitation for applying to distributive services. Therefore, an efficient coding technology for multi-view video is required without significant sacrificing the visual quality [1-2]. 3DAV (3-D audio and visual) group on MPEG (moving picture experts group) is working on the multi-view video coding (MVC) standards since Dec. 2001. Many algorithms have been proposed to improve coding efficiency such as prediction structure, illumination compensation, disparity vector coding, and view synthesis prediction.

The view synthesis prediction method employs an additional reference frame which is generated by using depth or disparity data. The view interpolation method generates an additional reference frame using disparities on stereoscopic images. The synthesized image can be added to reference frames.

In this paper, we describe both the previous and proposed view interpolation methods. In addition, we explain 'VIP P-picture' coding method that exploits the generated intermediate frame as an additional reference frame at the encoder.

## 2. PREVIOUS VIEW INTERPOLATION SCHEME

View interpolation method, which is proposed by Chen and Williams, can reconstruct arbitrary viewpoint images using disparity between two input images [3]. Therefore, multi-view video is useful for view interpolation because it has high correlation between views.

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

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

where $d$ ($d \geq 0$) stands for disparity of two matching pixels between stereoscopic images, and it can be factorized into two disparities by $\alpha$ ($0 \leq \alpha \leq 1$), which is a new viewpoint between two anchor images. The relationship for cameras is given as Eq. (2), where $\lfloor \ \rfloor$ is a rounding. *I(x,y)* stands for the intensity value of the sample at position *(x,y)*. $I_\alpha(x,y)$ indicates the intermediate view point image.

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

Droese proposed a view interpolation method in disparity domain for multi-view video coding [4]. He uses a simple block matching algorithm using WTA (winner-takes-all) strategy. Equation (3) represents the cost function which consists of the similarity term of textures and regularization term of disparities. $C_{sim}$ stands for MAD (mean absolute difference) and, $C_{reg}$ is an average difference between neighboring four disparities and $d$ as described in Eq. (4). $\lambda$ controls the influence of the second term. The range of $d$ is zero to the predefined maximum disparity value.

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

$$C_{reg}(x,y,d) = \frac{1}{4}\begin{bmatrix} |D(x-1,y)-d| + |D(x-1,y-1)-d| + \\ |D(x,y-1)-d| + |D(x+1,y-1)-d| \end{bmatrix} \tag{4}$$

## 3. PROPOSED VIEW INTERPOLATION SCHEME

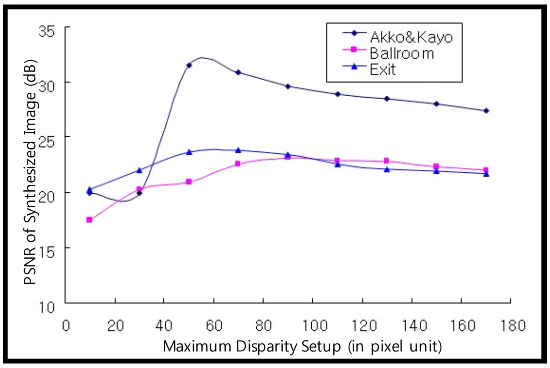



Fig. 1. Variation curves of PSNRs by Max. disparity setup

### 3.1 Disparity estimation problems

The matching block in disparity estimation moves according to the range of disparity. Therefore, a proper value of maximum disparity can contribute to the quality of the synthesized image. Figure 1 shows PSNR variation curves according to maximum disparity setup, where the size of test sequences is 640x480. To overcome this problem, we propose an initial disparity estimation method using region dividing. Another problem is disparity error in occlusion regions. It can be corrected by the proposed variable block matching method. Figure 2 shows the whole process of the proposed method.

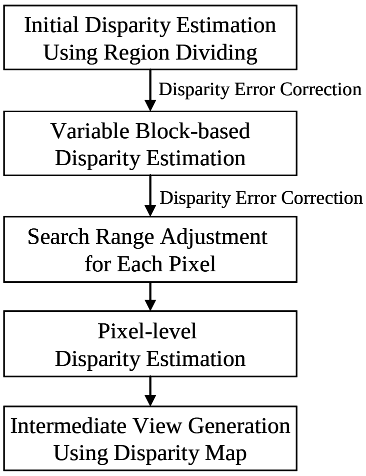



Fig. 2. Proposed view interpolation process

### 3.2 Initial disparity estimation

The objective of initial disparity estimation is to find approximate disparity values without pre-defined maximum disparity setup. Since stereo images follow the ordering constraint [5], we can reduce search range of disparity by using region dividing as shown in Fig. 3. If $Block_1$ have the most outstanding edges, it is efficient for disparity estimation. Then the estimator find a value for disparity with search range [0, width-1]. Then we divide the region into the left and right sides depending on the estimated disparity value. The next block to be estimated is $Block_2$ whose search range is $[x_1+d_1$, width-1] which is adjusted by considering the previous disparity value of $Block_1$. In the same manner, we can find a disparity of $Block_3$ with the search range $[0, x_1+d_1]$.

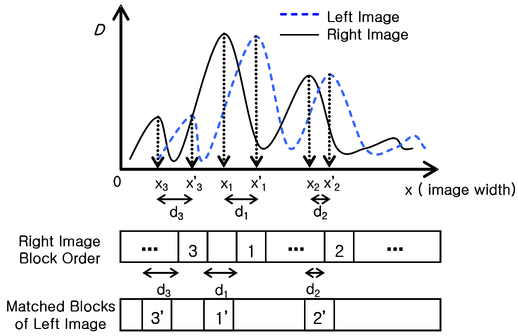



Fig. 3. Initial disparity estimation

Through the initial disparity estimation process, we use a gradient-based cost function as Eq. (5). It is more robust to changes in camera gain and bias.

$$C_{grad}(x,y,d) = \sum_{(i,j)\in N_x(x,y)} \left|\nabla_x I_L(x,y) - \nabla_x I_R(x+d,y)\right| + \sum_{(i,j)\in N_y(x,y)} \left|\nabla_y I_L(x,y) - \nabla_y I_R(x+d,y)\right| \quad (5)$$

where $N_x(x,y)$ is a 3x3 surrounding window without the rightmost column, $N_y(x,y)$ a surrounding window without the lowest row, $\nabla_x$ the forward gradient to the right and $\nabla_y$ the forward gradient to the bottom.

### 3.3 Variable block-based disparity estimation

Since the previous method uses a fixed block size for block matching, the estimator may detect wrong disparity when the reference block covers boundary region. The variable block-based disparity estimation method is described in Fig. 4. Once a disparity is determined for a basic block, estimator checks differences of costs between the larger block and sub-blocks. If a difference of costs is bigger than the pre-determined threshold $T$ between one of sub-blocks and the lager block, estimator find a disparity again for the sub-block with the same search range. The iteration continues until the block size becomes 2x2.

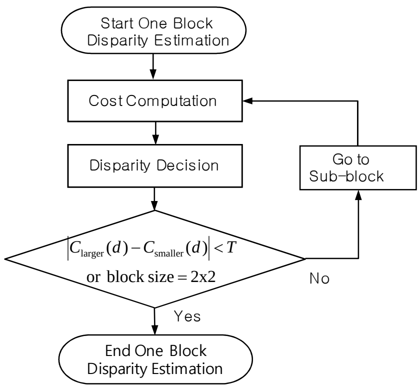



Fig. 4. Variable block-based disparity estimation

### 3.4 Pixel-level disparity estimation

The final step of disparity estimation is the pixel-level disparity estimation operation. The procedure of the pixel-level disparity estimation is similar to the previous method. It employs the cost function of Eq. (3). However, search range of disparity for each pixel is adjusted by Eq. (6), which considers the obtained disparity map *D(x,y)*. Since the search range does not need to be large to cover the maximum disparity value, we used smaller value than the basic block size in our experiments.

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

### 3.5 Disparity error correction

The proposed method exploits the disparity error correction method to increase accuracy and consistency of disparities. If an object contains many different disparity values, some of them may be erroneous disparity values. Those disparities can be corrected by checking cost values with neighboring disparities using lager block size.

### 3.6 View synthesis and hole filling

An intermediate view image can be synthesized by Eq. (7). We assume that illumination changing is linear. However, miss disparity or occlusion area makes hole area as shown in Fig. 5(a). Hole area can be filled with corresponding texture values by replacing the neighboring disparity.

$$\begin{aligned} & I_\alpha(x+(1-\alpha)\cdot D(x,y), y) \\ & \quad = (1-\alpha)\cdot I_L(x+D(x,y),y) + \alpha \cdot I_R(x,y) \end{aligned} \tag{7}$$

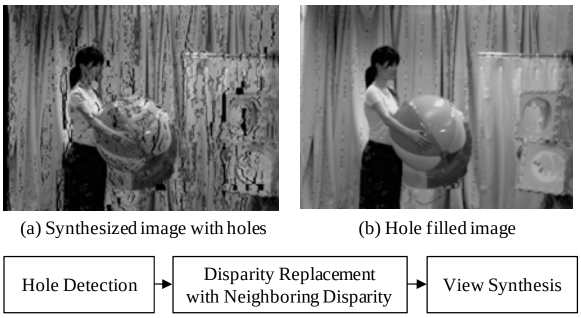



Fig. 5. Hole filling

## 4. 'VIP P-PICTURE' CODING

The basic coding structure of MVC is based on H.264/AVC coding standard, and it uses the hierarchical B picture coding method in temporal direction as shown in Fig. 6 [6]. S1, S3, and S5 views in Fig. 6 are coded right after encoding adjacent views. View interpolation process can also be applied to those inter-view sequences since it can refer to two reference views.

### 4.1 Additional estimation modes

The interpolated images can be used as additional reference frames at the encoder. Thus we propose the VIP (view interpolated prediction) P-picture coding method which refers to the added frame. 'VIP P-picture' coding consists of five additional modes: 'VIP_SKIP', 'VIP_16x16', 'VIP_8x16', 'VIP_16x8', and 'VIP_P8x8'. We call these modes 'VIP mode'. 'VIP_SKIP' refers to co-located block of the intermediate image. Therefore, it does not need to encode motion vectors and residuals except for the reference frame index and mode type.

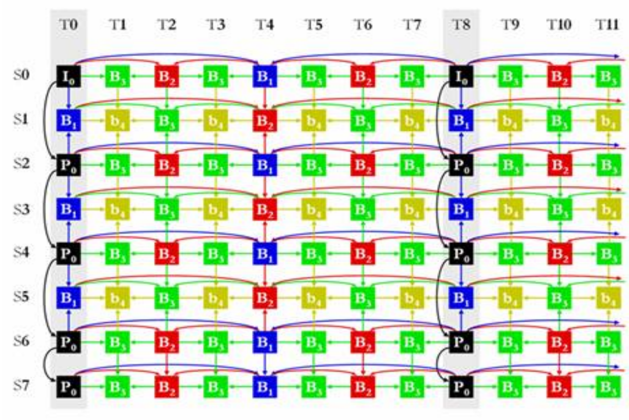



Fig. 6. Reference prediction structure of MVC

### 4.2 Modified motion vector prediction

The H.264/AVC encodes a difference value between the real motion vector and predicted motion vector. Because the generated frame is added to the reference frame, the accuracy of the motion vector prediction can be decreased. Therefore, we propose a modified motion vector prediction method. If estimating mode is one of traditional modes, motion vector predictor considers only MVs referred to traditional reference frames. On the other hand, if the mode is one of VIP modes, it considers only MVs referred to the interpolated frame.

## 5. EXPERIMENTAL RESULTS

In order to evaluate both the proposed view interpolation scheme and 'VIP P-picture' coding method, we compared generated intermediate images with original view images in terms of PSNR and bit-rates. We used the reference software JMVM 1.0 (joint multi-view video model) and test sequences 'Akko&Kayo', 'Rena', and 'Ballroom', which are provided by JVT.

We used 16x16 window size for block-based disparity estimation. The threshold value in Fig. 4 is 15, and search ranges of Eq. (6) are 5, 10, and 15 in pixel unit at the pixel-level disparity estimation. We used the reference coding structure of Fig. 6. The simulated QPs are 22, 27, 32, and 37. The search range is 96 for the traditional motion estimation modes and 48 for additional VIP modes. More detail is represented in [7].

### 5.1 Results of view interpolation scheme

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

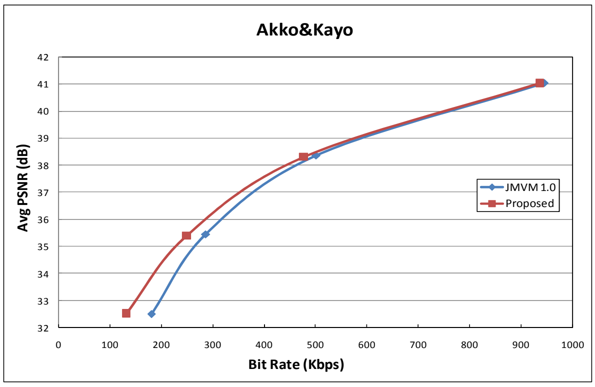
Table 1. View interpolation results: Avg. PSNR of 30 frames

| 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 |

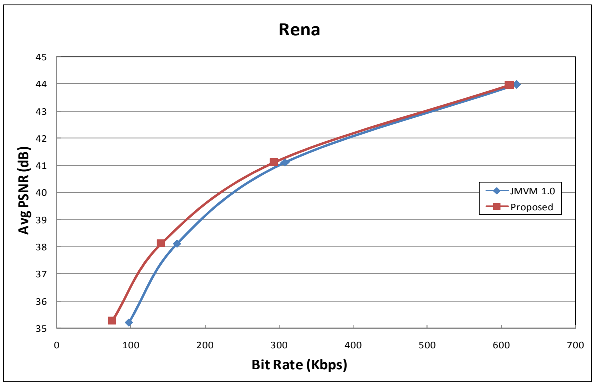
Unit: dB

### 5.2 Results of multi-view video coding

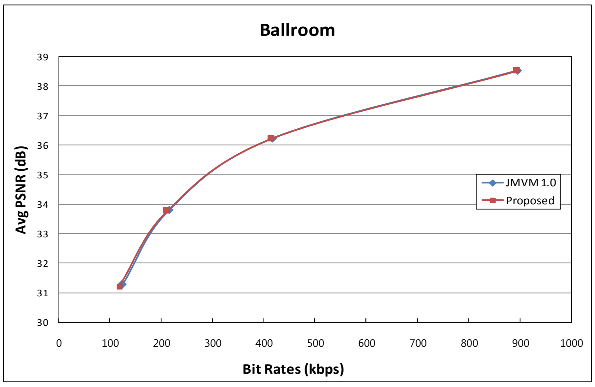
Since the reference prediction model of MVC exploits inter-view direction estimation which can refer to adjacent view images, the proposed 'VIP P-picture' coding can be applied to several in-between sequences. 6 in 15 views of 'Akko&Kayo', 7 in 16 views of 'Rena', and 3 in 8 views of 'Ballroom' are applied to the proposed coding method. Figure 7 shows multi-view video coding results of three test sequences.




(a) Rate-distortion curves for Akko&Kayo




(b) Rate-distortion curves for Rena




(c) Rate-distortion curves for Ballroom

Fig. 7. Results of multi-view video coding

As shown in Fig. 7, coding gains of 'Akko&Kayo' and 'Rena' sequences are improved about 0.74 dB and 0.57 dB in Bjontegaard measure [8]. One considerable point is that coding gains are increasing at low bit rates. Since the quality of the added frame is far from other reference frames, the interpolated frame is hardly chosen at high bit rates. However, the quality of the reconstructed frame and interpolated frame are quite similar at high QP. Thus VIP modes can be selected at low bit rates. In addition, the coding gain of 'Ballroom' is near to zero because the quality of generated images is not good enough to contribute to coding gain.

The proposed coding method employs view interpolation process at both the encoder and decoder. Therefore, it is quite complex. In fact, the view interpolation method occupies the encoding time about 13 %. However, it is meaningful approach because this method does not send any disparity information.

## 6. CONCLUSION

In this paper, we have described an efficient view interpolation scheme and a multi-view video coding method. The proposed view interpolation method consists of three disparity estimation processes. This framework can improve the quality of the intermediate image about 1~4 dB compared to the previous method. Based on this, we proposed the 'VIP P-picture' coding method which employs the intermediate image as an additional reference frame. It includes five additional modes referring to the interpolated image and the modified motion vector prediction method. As a result, our proposal improved the coding efficiency about 0.66 dB on average in well interpolated sequences.

## ACKNOWLEDGEMENTS

This work was supported in part by MIC through RBRC at GIST, and in part by MOE through the BK21 project.

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

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

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

[8] G. Bjontegaard, "Calculation of Average PSNR Differences between RD-curves," VCEG-M33, Austin TX, 2001.