Distributed Coding Scheme for Multi-view Video through Efficient Side Information Generation

Jihwan Yoo*, Min Soo Ko**, Soon Chul Kwon***, Young-Ho Seo§,
Dong-Wook Kim§§ and Jisang Yoo†

Abstract – In this paper, a distributed image coding scheme for multi-view video through an efficient generation of side information is proposed. A distributed video coding technique corrects the errors in the side information, which is generated with the original image, by using the channel coding technique at the decoder. Therefore, the more correct the generated side information is, the better the performance of distributed video coding. The proposed technique is to apply the distributed video coding schemes to the image coding for multi-view video. It generates side information by selectively and efficiently using both 3-dimensional warping based on the depth map with spatially adjacent frames and motion-compensated temporal interpolation with temporally adjacent frames. In this scheme the difference between the adjacent frames, the sizes of the motion vectors for the adjacent blocks, and the edge information are used as the selection criteria. From the experiments, it was observed that the quality of the side information generated by the proposed technique was improved by the average peak signal-to-noise ratio of 0.97dB than the one by motion-compensated temporal interpolation or 3-dimensional warping. The result from analyzing the rate-distortion curves revealed that the proposed scheme could reduce the bit-rate by 8.01% on average at the same peak signal-to-noise ratio value, compared to previous work.

Keywords: Distributed video coding, Multi-view video coding, Side information, Depth map, 3D warping

1. Introduction

In the last few decades, broadcasting technologies have grown from black and white TV to HDTV. Especially in recent years, almost all countries in the world have been replacing traditional analog broadcasting with digital. Despite such progress, TV viewers have been demanding more realistic broadcasting and accordingly research institutes (R&D centers) and companies in the world have been developing 3-dimensional television (3DTV) and ultra high definition television (UHDTV) as the solution to broadcast more realistic image content. Under ISO / IEC, the Moving Picture Experts Group (MPEG) has worked on standardization for a 3D image coding scheme by the name of 3DAV from 2001 [1]. In particular, this group standardized the multi-view video coding (MVC) in July 2008 [2]. As the MVC refers to both spatially

and temporally adjacent images, it has a drawback that the encoding is very complex. A distributed video coding (DVC) encodes a video by dividing its frames into two kinds: key frames and Wyner-Ziv frames. The key frames are encoded with an intra coding technique, while for the Wyner-Ziv frames only the parity bits are transmitted. Therefore, a DVC scheme can reduce the encoding complexity quite a lot. Meanwhile, the decoder generates side information similar to the Wyner-Ziv frames from the received key frames. The error between the original Wyner-Ziv image and the generated side information is corrected by a channel coding technique with the transmitted parity bits. The research on DVC has been mainly in progress by Girod at Stanford University [3-5], Ramchandran at Berkeley University [6], and a collaborative research group of DISCOVER in Europe [7]. The distributed multi-view video coding (DMVC) [8-10] generates side information by a Motion-compensated Temporal Interpolation (MCTI) technique with temporally adjacent images or from spatial prediction with spatially adjacent view images at the same instant. Also, a technique that mixes the above two schemes has been proposed in which the differences in pixel values and the magnitudes of motion vectors are used to obtain more accurate side information [8].

However, the method in [8] failed to show better performance than when only MCTI was used. The one in

† Corresponding Author: Dept. of Electronic Engineering, Kwangwoon University, Korea. (jsyoo@kw.ac.kr)

* LG Electronics Inc. Seoul, Korea. (uzhwan@naver.com)

** Dept. of Electronic Engineering, Kwangwoon University, Korea. (kmsqwet@kw.ac.kr)

*** Dept. of Information & Contents, Kwangwoon University, Korea. (ksc0226@kw.ac.kr)

§ College of Liberal Arts, Kwangwoon University, Korea. (yhseo@kw.ac.kr)

§§ Dept. of Electronic Materials Engineering, Kwangwoon University, Korea. (dwkim@kw.ac.kr)

Received: September 23, 2013; Accepted: April 28, 2014

[9] employed a Homography-compensated Inter-view Interpolation (HCII) technique that uses the homography of images to generate side information from the adjacent view images. But because of the inaccuracy in the corresponding view points in the process of homography estimation and the limitation of interpolation performance in warping, it showed better image quality of only about 0.2~0.5dB in peak signal-to-noise ratio (PSNR) than MCTI. Also [10] used the reference image(s) in various ways such as left, right, and average values in using HCII, but it could not overcome the inherent limitation of HCII.

To solve those problems, we proposed a DMVC scheme in this paper. It selectively uses both 3D warping [11] and MCTI according to the characteristics of the target and adjacent images. In selecting the techniques, it considers the intensity difference between the previous and the next time-adjacent frames, the magnitudes of the motion vectors of the current block and adjacent blocks, edge information resulting from the depth map, and the intensity of the residual signals obtained by motion compensation. Here, we assumed that the depth map (or image) is given for the two multi-view videos titled Breakdancers and Ballet provided by MPEG [12], which we take as our test multi-view videos. A depth map is an image or image channel that contains information relating to the distance of the surfaces of scene objects from a viewpoint [13]. Also many depth cameras are available in public these days [14] that provide depth information. With that information, the proposed method finds the characteristics of the block to be reconstructed and selects an appropriate reconstruction technique to generate more accurate side information.

First, the existing methods of multi-view video coding are shortly introduced in the next section. Then, the proposed method to solve the problems that the existing methods have is explained in Section 3. The performance of the proposed method is experimentally proved in Section 4 and this paper is concluded in Section 5 on the basis of the experimental results.

Fig. 1 shows the current reference structure for MVC.

means the predicted images obtained by an intra-prediction or inter-prediction, which is accessible at both the encoding and decoding sides. Meanwhile in a distributed video coding, the correlation is removed by the decoder. In this case, differently from the case of (a), the side information generated by the decoder is only accessible in the decoder. Note that the performance of a distributed video coding technique could be the same as that of the conventional coding technique, which is based on the two theories from [17] and [18].

In [18], Slepian-Wolf showed that removing the correlation at the decoder side can have the same performance as removing the encoder. Fig. 3 shows the possible bit-rate regions for Slepian-Wolf compression coding. In the figure, R_X and R_Y are the bit-rate of X and Y , respectively, and $H(X)$ and $H(Y)$ are the quantitative entropies of X and Y , respectively. If X and Y are statistically independent, the achievable bit-rates have the relationship of (1) and they correspond to Region A.

Fig. 3. Achievable bit-rate regions for Slepian-Wolf distributed compression

$$\begin{aligned} R_X &\geq H(X) \\ R_Y &\geq H(Y) \end{aligned} \quad (1)$$

But if the two pieces of data are statistically correlated rather than independent, the relationship of the bit-rates is changed as in (2). In this case, the bit-rate region is extended to region B.

$$\begin{aligned} R_X &\geq H(X|Y) \\ R_Y &\geq H(Y|X) \\ R_X + R_Y &\geq H(X, Y) \end{aligned} \quad (2)$$

That is, as the correlation between X and Y increases, the amount of the occurring bits decreases, which means the compression ratio increases.

Wyner-Ziv extended the Slepian-Wolf's coding theory to the lossy compression cases by adding a quantization step [18]. It is general that the relationship between the rate-distortion (R-D) function of the conventional coding $R_{X|Y}(d)$ and that of the distributed coding $R^*(d)$ is $R_{X|Y}(d) < R^*(d)$. $R_{X|Y}(d)$ and $R^*(d)$ represents the rate-distortion functions of the conventional video coding and the distributed video coding, respectively. d is defined as bit rate. But it was proved that if the data has some special characteristics such that they have Gaussian distributions the relationship could be $R_{X|Y}(d) = R^*(d)$.

Fig. 4 shows a usual structure of distributed coding by the Wyner-Ziv method. First, the original image frames are divided into two kinds: key frames and Wyner-Ziv frames. At this time a group of pictures (GOP) consists of several key frames and more than one Wyner-Ziv frame [3]. A key frame is coded by one of the existing intra coding techniques such as JPEG or H.264/AVC. But a Wyner-Ziv image is coded with a channel coding technique such as turbo code [19] or Low Density Parity Check (LDPC) code [20] after quantization. Here, only the parity bits are stored

Fig. 4. Structure of distributed video coding

in the buffer for the Wyner-Ziv image. Before quantization a transformation step to transform to a frequency domain can be accomplished. In general, inserting a transformation step removes spatial redundancy quite a lot and the coding efficiency increases in turn.

Meanwhile, the decoder generates the side information similar to the Wyner-Ziv image, which is the target image to reconstruct. At this time the decoder tries to correct the error between the side information and the Wyner-Ziv image using a strong channel coding technique with the parity bits by regarding the error as a virtual channel noise. The decoder continuously requests additional parity information to the encoder until the reconstruction is succeeded. Accordingly the performance of a distributed coding technique totally depends on correctness of the side information and the performance of the channel coding technique.

2.3 Distributed multi-view video coding

In the environments of the multi-view video it is possible to generate the side information more correctly because the spatially adjacent images (the images at the different viewpoints at the same time) as well as the temporally adjacent images (the images at the different time at the same view point) can be used. The general techniques to generate side information in a Distributed Multi-view Video Coding (DMVC) are as follows.

The first one is an MCTI technique that uses the motion vectors between temporally-adjacent images. That is, it generates side information of the k^{th} frame with the previous $(k-1)^{\text{th}}$ and the next $(k+1)^{\text{th}}$ images, as shown in Fig. 5(a). For this, the motion vector between the $(k-1)^{\text{th}}$ and $(k+1)^{\text{th}}$ frame is obtained by motion estimation. Then the side information is interpolated from the adjacent frames with half of the motion vector. But because there might be some empty or overlapped regions between the interpolated blocks, the motion vector in the point A is shifted to the point B. After that the final motion vector is found by motion estimations from the $(k-1)^{\text{th}}$ and $(k+1)^{\text{th}}$ frames as shown in Fig. 5(b). Finally, each pixel value in the side information is interpolated with the average value of the corresponding pixels in the $(k-1)^{\text{th}}$ and $(k+1)^{\text{th}}$ frames.

Another method to generate side information is Homography Compensated Inter-view Interpolation (HCII) that uses spatially adjacent images (adjacent view images) with the same time stamp. It extracts the corresponding disparities from the view-adjacent images to generate the side information as shown in Fig. 6.

Finally, there is a side information generation method that mixes the above two methods, MCTI and HCII. Because the motion vector for a region having a large motion is possibly incorrect, the corresponding interpolated block by MCTI might show a large error. Therefore a region whose motion vector is larger than a pre-defined value is interpolated by HCII rather than MCTI.

Fig. 5. Side information generation by MCTI: (a) selection of motion vectors, (b) bidirectional motion estimation

Fig. 6. Side information generation by HCII

In a conventional multi-view video coding technique, the temporally adjacent images are more likely to be selected as reference images than the spatially adjacent images because the temporal correlation is higher than the spatial correlation. In a DMVC technique, however, the distance between the key frames that are used to predict the motions might be so large that the temporal correlation could be too small to be used. To get more accurate side information, a mixed or fused method to compensate these problems is highly required.

3. The proposed Distributed Multi-view Coding Method

The DMVC method to be proposed in this paper uses both the 3D warping and MCTI appropriately according to the characteristics of the target and the adjacent image blocks. It assumes that the necessary depth map and the camera parameters are transmitted and available at the

decoder side. Camera matrix is used to denote a projective mapping from world coordinates to pixel coordinates in 3D warping. Generally two kinds of matrices are required in 3D warping, which are the intrinsic and extrinsic matrices. The intrinsic matrix contains 5 intrinsic parameters for focal length, and image format. The extrinsic parameters define the position of the camera center and the camera's heading in world coordinates [21]. Of course transmitting depth information increases the bit-rate. But it is more effective than extracting depth information by a stereo matching technique at the decoder side. Thus, MPEG is processing standardization for encoding a depth map also. So we also assume that the necessary depth information is provided, not generated at either the encoder or decoder,

Fig. 7. Block diagram of the proposed distributed multi-view video coding method

because it is quite common to provide depth information and is not hard to obtain depth information with a publically available depth camera.

Fig. 7 shows a brief block diagram of the DMVC method to be proposed in this paper. As can see in the figure, it uses correlations for both time-adjacent images (MCTI) and space-adjacent images (3D warping) to generate side information on the basis of the usual DMVC structure.

The frame structure to apply the proposed method is shown in Fig. 8, which consists of two kinds of frames: a key frame (I) that is intra-coded and a Wyner-Ziv frame (WZ) that is channel-coded. The key frames and the Wyner-Ziv frames are alternatively placed in both the time axis and the viewpoint axis. Thus, a Wyner-Ziv frame can be reconstructed with the four adjacent key frames. In this paper, we propose a method to mix or selectively use the MCTI technique and the 3D warping technique. Also in Fig. 9, the proposed selection procedure of the two

Fig. 8. Frame structure to apply the proposed method

Fig. 9. The technique selection scheme to generate side information

Fig. 10. Generation of side information by 3D warping

techniques to generate the side information (Wyner-Ziv frame) is shown. Each condition of the following process is explained in the following section.

3.1 Generating side information by 3D warping

The procedure to generate side information by 3D warping with the space-adjacent frames is shown in Fig. 10. First, a Sobel mask is applied to the depth map of the right image (refer to Fig. 11(a)) to extract the edge information (Fig. 11(b)). In general the edge information extracted with this method does not exactly correspond to the actual edge information of the image. Thus, it is highly possible that the side information by 3D warping has errors in the edge regions. To avoid this error, we remove the regions of the original image corresponding to the edge regions.

The next step is to convert the 2D-coordinate of the 2D image into the corresponding 3D-coordinate with the depth map and the camera parameters. (3) shows the relationship between the 2D-coordinate and the corresponding 3D-coordinate, which is obtained by the geometric structure of a pin-hole camera.

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = K [R | T] \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (3)$$

Here, (x, y) is the 2D-coordinate in the image, while (X, Y, Z) is the 3D-coordinate in the real world. K is the intrinsic parameters of the camera consisting of a 3×3 matrix. R is the rotation parameters and is expressed by a 3×3 matrix, while T is the translation parameters with a 3×1 matrix. $[R|T]$ stand for the concatenation matrix of R and T . Depth map stores depth information as 8-bit gray values with the gray level 0 specifying the furthest value and the gray level 255 defining the nearest value [22, 23]. The real depth value $Z(i, j)$ which corresponds to the pixel (x, y) is

Fig. 11. Example of side information generation by 3D warping; (a) depth map, (b) edge of depth map, (c) projected image, (d) generated side information

transformed into the 8-bit gray value $P(i, j)$ as (4).

$$P(i, j) = \left\lfloor 255 - \frac{255 \cdot (Z(i, j) - \text{Min}Z)}{\text{Max}Z - \text{Min}Z} + 0.5 \right\rfloor \quad (4)$$

Where, $P(i, j)$ is the depth value at (i, j) in the depth map and $\text{Min}Z$ and $\text{Max}Z$ are the minimum and maximum depth values, respectively. The symbol $\lfloor \alpha \rfloor$ means the largest integer smaller than or equal to α . The real depth value $Z(i, j)$ which corresponds to the pixel (x, y) is expressed as (5) from (4).

$$Z(i, j) = 1 / \left(\frac{P(i, j)}{255} \times \left(\frac{1}{\text{Min}Z} - \frac{1}{\text{Max}Z} \right) + \frac{1}{\text{Max}Z} \right) \quad (5)$$

The converted 3D coordinate is again projected into the 2D plane of the target Wyner-Ziv image by using (3) as Fig. 11(c). The camera parameters at this time are the ones belonging to the Wyner-Ziv image. But as can be seen in Fig. 11 (c), the projected image has some holes and occluded regions. The holes occur during the coordinate conversion process and the occluded regions are the ones occluded before the view point change and appearing after. The holes can be filled by interpolating the adjacent images and the occluded regions can be filled with the data in the left image. Finally, the resulting image from filling the holes and the occluded regions is filtered with a median filter to get the final side information, as in Fig. 11 (d).

3.2 Generating of side information with low error

Because the previous multi-view video coding techniques could access both the time-adjacent frames and the space-

Fig. 12. Errors in the generated side information; (a) boundary regions by 3D warping, (b) regions with large motions by MCTI

adjacent frames to use as reference images, they calculate the rate-distortion (R-D) value for each prediction mode for each reference image and select the one with the minimum value. However, it is impossible for a Wyner-Ziv frame in coding by DMVC because only the parity bits are sent without the image information. Therefore, our method generates side information with the difference between the previous and next frames to the current frame, the edge information of the depth map, and the intensity of the residual signal generated with the motion vector and the compensated block.

In general, side information by 3D warping has a high possibility of errors in the boundary regions as Fig. 12(a) and the one by MCTI in the regions with large motions as Fig. 12(b). Thus, it can result in more accurate side information to use MCTI in the boundary regions and 3D warping in the regions having large motions. In using MCTI in the boundary regions also, the characteristic of inaccuracy for the large motion is still applied. Thus, if the motion vectors of the neighboring macro-blocks are greater than a pre-defined threshold value (Th2, refer to Fig. 9) or the signal intensity of the residual block by motion compensation is greater than another pre-defined threshold value (Th3, refer to Fig. 9), it uses 3D warping.

However, the motion estimation itself is not always correct in that a motionless region such as a boundary region may have error resulting from MCTI, as can be seen in Fig. 13(a). Thus, the proposed method first compares the difference between the two time-adjacent blocks to the current frame with a pre-defined threshold Th1 (refer to Fig. 9). If the difference is less than Th1 the block is regarded as the boundary region without motion and the side information corresponding to the block is generated with the average value of the two blocks. An example is shown in Fig. 13, where the errors in by MCTI in (a) is corrected by the above process in (b).

When motion vector should be used the motion vectors of the neighboring macro-blocks are considered together. For example, Fig. 14(a) shows a case when the center block has a small motion even though the neighboring blocks have large ones. In this case it is highly possible that the estimated motion for the center block is not correct. It can result in inaccurate side information if the neighboring

Fig. 13. Processing background error; (a) before process, (b) after process

Fig. 14. Improvement for the inaccurate estimated motion region: (a) inaccurate motion region before process, (b) inaccurate motion region after process, (c) residual image before process, (d) residual image after process

blocks are not considered, as shown in Fig. 14(c). To reduce the amount of these kinds of errors, 3D warping rather than MCTI is used in such a case or vice versa, as shown in Fig. 14(b) (refer to Fig. 9). Fig. 14(d) shows the improved result after applying this scheme.

4. Experiments and Results

We have performed some experiments to figure out the performance of the proposed method. The multi-view video sequences used in the experiments were the “Breakdancers” and “Ballet” videos provided by Microsoft Research, “Poznan Street” video provided by Poznan University of Technology, and “Dancer” video provided by Nokia Inc. The sequences have 8/3 view-points and depth

maps. Each sequence consists of 100 frames for each view-point. The resolution and the frame rate of the former two sequences were XVGA (1024x768) and 15fps respectively, but the resolution and the frame rate of the latter two sequence were full HD (1,920x1,080) and 25fps respectively.

In the experiments, the transform unit for DCT was 8×8 and, for the quantization factors, the results from multiplying the scaling factors ($Q = 0.5, 1, 2, 4$) of the quantization table in JPEG [24] were used. As the channel code Low-Density Parity-Check Accumulate (LDPCA) code was used, which was proposed by Varodayan et al. [25]. LDPCA has the advantage of having an adaptive bit-rate, while Low-Density Parity Check (LDPC) code has a fixed bit-rate. After testing the coding efficiency of various videos with various thresholds for finding the best result, we empirically choose the threshold values Th1, Th2, and Th3 which have the best coding performance. All experiments were executed by the same threshold values which are not changed according to images.

Fig. 15 shows examples of the generated side information for Break-dancers ((a), (b), (c)) and Ballet ((d), (e), (f)) by MCTI ((a), (d)), by 3D warping ((b), (e)) and by the proposed method ((c), (f)). Also Fig. 16 shows the different (residual) images resulting from subtracting the generated image in Fig. 15 from the originals are shown with the same order. It is obvious that the regions with large motions in Fig. 15(a) or (d) show large errors, while in Fig. 15(b) and (e) large errors appear in the boundary

regions. As can see in Fig. 15(c) and (f) or Fig. 16(c) and (f), the proposed method greatly reduces those errors. The experimental results are summarized in Table 1, where the PSNR values are the averages for all the frames (30 frames

Table 1. Average PSNR value comparison

Sequence	Q (scaling factor)	PSNR(dB)				
		MCTI	3D warping	HCII	MCTI + HCII	Proposed
Break dancers	Q=0.5	26.82	33.23	27.37	28.18	33.41
	Q=1	26.85	33.34	27.41	28.25	33.54
	Q=2	26.81	33.06	27.33	28.13	33.13
	Q=4	26.68	32.00	26.81	27.83	32.13
	Average	26.79	32.91	27.23	28.10	33.05(+0.14)
Ballet	Q=0.5	33.03	29.14	26.54	33.05	34.93
	Q=1	33.00	29.15	26.52	33.01	34.93
	Q=2	32.94	29.02	26.46	32.90	34.74
	Q=4	32.62	28.55	26.02	32.57	34.21
	Average	32.90	28.97	26.39	32.88	34.70(+1.80)
Poznan Street	Q=0.5	28.76	31.67	26.88	29.23	32.75
	Q=1	28.71	31.66	26.85	29.21	32.73
	Q=2	28.64	31.54	26.57	29.08	32.67
	Q=4	28.57	31.23	26.30	28.85	32.56
	Average	28.67	31.53	26.65	29.09	32.68(+1.15)
Dancer	Q=0.5	28.34	33.12	27.65	29.67	34.54
	Q=1	28.32	33.12	27.64	29.65	34.53
	Q=2	28.14	32.96	27.40	29.36	34.34
	Q=4	28.05	32.77	27.23	29.18	34.10
	Average	28.21	32.99	27.48	29.47	34.38(+1.39)

Fig. 15. Generated side information examples for Breakdancers by: (a) MCTI, (b) 3D warping, (c) proposed method, and for Ballet by: (d) MCTI, (e) 3D warping, (f) proposed method

Fig. 16. The difference images of Fig. 15 images to the original images. For Breakdancers by: (a) MCTI, (b) 3D warping, (c) proposed method, and for Ballet by: (d) MCTI, (e) 3D warping, (f) proposed method

Fig. 17. R-D curves for; (a) Breakdancers, (b) Ballet, (c) Poznan Street, (d) Dancer

per sequence) of all the view-points in the two test video sequences. As can be seen in the table, the proposed method that used both 3D warping and MCTI appropriately showed 0.14~1.8dB higher PSNR values than the cases using the single technique of 3D warping or MCTI. Here the PSNR values are quite different according to the kind of test sequences and the applied technique because the two have quite different image characteristics. The results from applying MCTI to the Breakdancers video sequence show relatively low quality because it has relatively large motions. Meanwhile, the Ballet sequence has the characteristics that the distance between the camera and the objects are small so that the occluded regions are relatively large. Thus Ballet shows relatively low quality when 3D warping is used.

Fig. 17 shows the R-D curves for the Breakdancers, Ballet, Poznan Street, and Dancer sequences resulting from applying various techniques to generate side information. If observing the graphs of Fig. 17, HCII generally has low performance, especially it has lower performance in the case of large motion. In case of small motion such as the Ballet sequence, 3D warping-based techniques show better performance, but they have worse performance in case of large motion.

After the average of the bitrates for each PSNR was calculated, the average bitrate of the proposed algorithm was compared with those of the other methods. In the case of the Breakdancers sequence in Fig. 17(a), each PSNR is 41.5dB, 40.1dB, 38.4dB, and 36dB and the ratios of bitrate reduction between the proposed and the others are 7.91% (MCTI), 49% (HCII), 36.71% (3D warping), and 7.08% (MCTI+HCII). In the case of the Ballet sequence in Fig. 17(b), each PSNR is 42.7dB, 41.1dB, 39dB, and 36.3dB and the ratios of bitrate reduction are 7.91% (MCTI), 49% (HCII), 36.71% (3D warping), and 7.08% (MCTI+HCII). In the case of the Poznan Street sequence in Fig. 17(c),

each PSNR is 41.3dB, 40.8dB, 38.7dB, and 36.3dB and the ratios of bitrate reduction are 34.25% (MCTI), 9% (HCII), 38.44% (3D warping), and 29.47% (MCTI+HCII). In the case of the Dancer sequence in Fig. 17(d), each PSNR is 42.1dB, 40.5dB, 38.3dB, and 36.2dB and the ratios of bitrate reduction are 33.13% (MCTI), 8.21% (HCII), 36.43% (3D warping), and 28.83% (MCTI+HCII).

Table 2 shows the average bitrates of the depth maps. The depth map was compressed under the same condition as Table 1. Since the resolution of Poznan Street and Dancer is full HD, they have somewhat higher bitrate than other sequences. The total average bitrate which includes the resultant bitrate of depth map is shown in Table 3. Because the resolution of Break Dancers and Ballet is X VGA (1,024x768) but the resolution of Poznan Street and Dancer is full HD (1,920x1,024), the amount of the average bitrate is different each other. When the average bitrates are compared between the proposed algorithm and the others, they are reduced to 27.64%, 24.24%, 27.89%, and 25.02%, respectively.

5. Conclusion

In this paper we proposed a more efficient distributed multi-view video coding method based on depth information. That is, it proposed a method to generate more exact side information to increase the performance of distributed multi-view video coding. In its scheme, 3D warping and MCTI are selectively used with three threshold parameters according to the characteristics of the image. The three threshold parameters are to determine which technique is applied with the difference between the previous and the next time-adjacent frames, the edge information extracted from the depth map, the magnitude of the motion vector, and the residual signal value generated by motion compensation. Also, we have considered the motion vectors for the neighboring macro-blocks to reduce the errors resulting from incorrect motion estimation.

Acknowledgements

This work was supported by the IT R&D program of MSIP/KEIT. [10039199, A Study on Core Technologies of Perceptual Quality based Scalable 3D Video Codecs].

References

- [1] ISO/IEC JTC1/SC29/WG11, "Requirements for Standardization of 3D Video," m8107, Jeju Island, Korea, March 2002.
- [2] ISO/IEC JTC1/SC29/WG11, "Text of ISO/IEC 14496-10:200X/FDAM 1 Multi-view Video Coding," N9978,

Table 2. Average bitrate of depth map

Sequence	PSNR	Bitrate(kbps)
Break dancers	36.0	53,849
	38.4	80,837
	40.1	146,499
	41.5	307,212
	Average	147,099
Ballet	36.3	65,171
	39.0	87,845
	41.1	133,791
	42.7	249,446
	Average	134,063
Poznan Street	36.3	151,537
	38.7	228,599
	40.8	413,872
	41.3	868,764
	Average	415,693
Dancer	36.2	138,146
	38.8	218,624

- Hannover, Germany, July 2008.
- [3] B. Girod, A. Aaron, S. Rane and D. Rebollo Monedero, "Distributed video coding," in *Proc. IEEE*, vol. 93, pp.447-460, January. 2005.
 - [4] A. Aaron, S. Rane, E. Setton and B. Girod, "Transform-domain Wyner-Ziv codec for video," in: *SPIE Visual Communications and Image Processing Conference*, vol. 5308, pp. 520-528, San Jose, CA, 2004.
 - [5] A. Aaron, R. Zhang and B. Girod, "Wyner-Ziv coding of motion video," in: *Proceedings of Asilomar Conference on Signals and Systems*, Pacific Grove, CA, November 2002.
 - [6] R. Puri and K. Ramchandran, "PRISM: A new robust video coding architecture based on distributed compression principles," in *Proc. Allerton Conference on Communication, Control, and Computing*, Allerton, IL, October 2002.
 - [7] <http://www.discoverdvc.org>
 - [8] X. Guo, Y. Lu, F. Wu, W. Gao, and S. Li, "Distributed Multi-view Video Coding," *Visual Communications and Image Processing 2006*, San Jose, CA, January 2006.
 - [9] M. Ouaret, F. Dufaux, and T. Ebrahimi, "Fusion-based Multiview Distributed Video Coding," *4th ACM international workshop on video surveillance and sensor networks 2006*, Santa Barbara, CA, October 2006.
 - [10] F. Dufaux, M. Ouaret, and T. Ebrahimi, "Recent Advances in Multi-view Distributed Video Coding," *SPIE Mobile Multimedia/Image Processing for Military and Security Applications*, Orlando, FL, April 2007.
 - [11] C. L. Zitnick, S. B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski, "High-Quality Video View Interpolation Using a Layered Representation," *ACM SIGGRAPH and ACM Trans. On Graphics*, Los Angeles, CA, USA, Volume 23, Issue 3, pp. 600-608, Aug. 2004.
 - [12] <http://www.research.microsoft.com/ImageBasedRealities/3DVideoDownload/>
 - [13] http://en.wikipedia.org/wiki/Depth_map
 - [14] <http://www.mesa-imaging.ch/>
 - [15] ISO/IEC MPEG & ITU-T VCEG, "Joint Draft 1.0 on Multiview Video Coding," JVT-U209, Nov. 2006
 - [16] H.-S. Koo, Y.-J. Jeon, B.-M. Jeon, "MVC Motion Skip Mode," *ITU-T and ISO/IEC JTC1, JVT-W081*, San Jose, California, USA, April 2007.
 - [17] D. Slepian and J. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inform., Theory* 19 (1973) 471-480.
 - [18] A. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the receiver," *IEEE Trans. Inform., Theory* 22 (1976) 1-11.
 - [19] J. Garcia-Frias, "Compression of correlated binary sources using Turbo codes," *Communications Letters, IEEE*, Vol.5, Issue 10, October, 2001.
 - [20] A. Liveris, Z. Xiong, and C. Georgiades, "Compression of binary sources with side information at the decoder using LDPC codes," *IEEE Commun. Lett.*, vol.6, no.10, pp.440-442, October. 2002.
 - [21] http://en.wikipedia.org/wiki/Camera_resectioning
 - [22] Richard Hartley and Andrew Zisserman, "Multiple View Geometry," Cambridge University, pp.152~247, Second Edition 2003.
 - [23] Masayuki Tanimoto, Toshiaki Fujii, and Kazuyoshi Suzuki, Improvement of Depth Map Estimation and View Synthesis, ISO/IEC JTC1/SC29/WG11 M15090, Jan. 2008.
 - [24] ITU-T, I. JTC1, "Digital compression and coding of continuous-tone still images," ISO/IEC 10918-1-ITU-T Recommendation T.81 (JPEG).
 - [25] D. Varodayan, A. Aaron, B. Girod, "Rate-adaptive distributed source coding using low-density parity-check codes," *Signals Systems and Computers Conference Record of the Thirty-Ninth Asilomar Conference*, November 2005.

Jihwan Yoo He received B.S. and M.S. degrees in electronics engineering from Kwangwoon University in 2005 and 2010, respectively. He is currently a research engineer at LG Electronics. His research interests include video coding, free-view point system and GPGPU.

Min Soo Ko He received the B.S. and M.S. degrees in electronics engineering from Kwangwoon University, Seoul, Korea in 2010 and 2012, respectively. He is currently a PhD student at Kwangwoon University. His research interests include stereo matching and 3D image processing and image compression.

Soon Chul Kwon He received the B.S. degree in industrial engineering from Hanyang University, Seoul, Korea in 2002. He received the M.S. degree in digital contents in 2008 and Ph.D degree in information display in 2012 from Kwangwoon University. From september 2005 to may 2009, he worked as an assistant manager of network engineering at LG Telecom, Seoul, Korea. He is currently an assistant professor with the department of media contents,

graduation school of information and contents, Kwangwoon University, Seoul, Korea. His research interests are in 3D display and holography.

Young-Ho Seo He has received his M.S and Ph.D degree in 2000 and 2004 from Dept. of Electronic Materials Engineering of Kwangwoon University in Seoul, Korea. He was a researcher at Korea Electrotechnology Research Institute (KERI) in 2003 to 2004. He was a research professor of Dept. of Electronic and Information Engineering at Yuhan College in Buchon, Korea. He was an assistant professor of Dept. of Information and Communication Engineering at Hansung University in Seoul, Korea. He is now an associated professor of College of Liberal Arts at Kwangwoon University in Seoul, Korea and a director of research institute in TYJ Inc. His research interests include realistic media, digital holography, SoC design and contents security.

Dong-Wook Kim He has received his M.S degree in 1985 from Dept. of Electronic Engineering of Hangyang University in Seoul, Korea and his Ph.D degree in 1991 from Dept. of Electrical Engineering of Georgia institute of Technology in GA, U.S.A. He is a Professor of Dept. of Electronic Materials Engineering at Kwnagwoon University in Seoul, Korea.

Jisang Yoo He received the B.S. and M.S. degrees from Seoul national university, Seoul, Korea in 1985 and 1987, all in electronics engineering, and Ph.D. degree from Purdue university, West Lafayette, IN, USA, in electrical engineering in 1993, respectively. From september 1993 to august 1994, he worked as a senior research engineer in industrial electronics R&D center at Hyundai Electronics Industries Co., Ltd, ichon, Korea, in the area of image compression and HDTV. He is currently a professor with the department of electronics engineering, Kwangwoon university, Seoul, Korea. His research interests are in signal and image processing, nonlinear digital filtering, and computer vision. He is now leading 3DTV broadcast trial project in Korea. He received B.S degree in electrical engineering from Yonsei university. His research interests are robust control, filtering and signal processing.