Client Driven System of Depth Image Based Rendering

Norishige Fukushima[1] and **Yutaka Ishibashi**[2], Non-members

ABSTRACT

In this paper, we propose systems which can render free viewpoint images by using depth image based rendering for live video communications. Experimental results show that an image and depth transmission system is more suitable than an all multi view transmission and an only free viewpoint image transmission system. Especially in the image and depth transmission system, transmitting two images and one depth map, and then predicting the opposite depth map is the best for live communication.

Keywords: Free Viewpoint Image Synthesis, Multi Camera Array, Network Environment, Depth Image Based Rendering, Multi-View Video Coding, Video Live Transmission

1. INTRODUCTION

Recently, stereoscopic 3DTV comes under the spotlight and is going to commercial-produced stage. 3DTV can transmit more reality than the conventional 2DTV, but a limitation of current 3DTV exists. This is because the viewpoint is fixed. Users cannot change their viewpoints freely although we can see everywhere in a real environment. To make it possible, free viewpoint image rendering becomes one solution. In the decade researches, technologies of free viewpoint image generation and 3D reconstruction are upgrading. To generate the free viewpoint images, there are several approaches; model based rendering (MBR), which generates 3D mesh and texture, and image based rendering (IBR), which generates a novel view by modifying input images. The IBR can render more photo-realistic images than the MBR, and has a high compatibility of the conventional video processing technologies. In this paper, we focus on the IBR.

Rendering free viewpoint images with the IBR, ray space [1] and light field [2] are basic methods. These methods can render high quality images without computational complexity, while these require tremendous views to keep a light field sampling theory [3]. Depth image based rendering (DIBR) [4] is a medial method between the MBR and the IBR. This method requires a few views and depth maps, which is depth information image for the captured view. A limitation of this method is that the rendering image quality tends to be degraded if an input viewpoint is far from captured views. Recently, the moving picture experts group (MPEG) has started a new exploration experiment specifically targeted towards free viewpoint television (FTV) [5] applications. The FTV is one of systems for free viewpoint image rendering, transmission and compression. M-view + N-depth framework [5] is adopted in the FTV and the framework cover the shortcoming of the DIBR, that is the quality of image generated by the framework does not highly depend on viewpoint. Whereas the previous MPEG/JVT standardization activities of multi view coding (MVC) [6, 7] focused on improvement of compression efficiency for generic multi view videos, so that video compression and image rendering are completely separated. The MPEG-FTV project currently focuses on depth estimation, rendering and compression. To render high quality videos, a target of this project is storage media. Thus, contributors use cost-consuming algorithms for free viewpoint image generation. For live free viewpoint communication, however, additional aspects are required.

Under a network environment, multimedia communication requires not only media quality itself but also interactivity for user inputs, because some kind of latency is inevitable when the multimedia is transmitted through the network. Considering the free viewpoint image communication, media quality and interactivity, such as changing viewpoint, are also important. At the aspects of these points, how to render the images and how to compress and transmit the data will be hot-spot. In a networked rendering system, there are two extreme approaches. First one is to transmit compressed multi view data from the server and to render a free viewpoint image at the client. Second one is to render a free viewpoint image at the server side, and then to transmit only the free viewpoint image. The former requires large network band, but rendering quality and interactivity are independent of network condition. The Latter requires only normal network band, but the interactivity is highly affected by network condition. In this case, a timing of viewpoint changing will be delayed because of network latency. To realize a medium format between the multi view and the only free viewpoint image transmission system, a shared I field and a selec-

Manuscript received on February 28, 2011 ; revised on August 11, 2011.

[1,2] The authors are with Department of Scientific and Engineering Simulation, Graduate School of Engineering, Nagoya Institute of Technology Gokiso-cho, Showa-ku, Nagoya 466-8555, Japan, E-mail: fukushima@nitech.ac.th and ishibasi@nitech.ac.jp

tive B frame [8, 9] are proposed. With these methods, multi view images are well compressed but a range of viewpoint change is limited. By using these methods, we have the upper bound of viewpoint change; thus we cannot change viewpoint freely. If we want to change viewpoint anywhere, we should remove the limit, and then both methods are nearly equal to the multi view transmission system method.

To overcome the above problems, a DIBR system implemented by the server-client model is suitable, and we propose this type of system in this paper. With this system, selected depth maps are generated from input images and these media are compressed / transmitted at the server side. With this DIBR approach, limited depth and image data around the interested viewpoint are required. The client decodes the data and synthesizes a novel view from the data. The data size is smaller than the size of multi view data, and if a process of viewpoint requesting and a process of view generation are asynchronously paralleled, the interactivity becomes higher than the generated free viewpoint image streaming mode. A weak point of this system is that the rendering quality may be degraded when interest view is quite different between the server and the client because of network latency. This method is an extension of the method in [10], which does not have capability of selective processing over an IP network. The system always transmits one view and one depth on a fixed viewpoint.

To realize the free viewpoint live communication, we need to compute all process on real-time, especially depth estimation and compression. These two processes are usually computationally expensive processes. Thus we use more convenient method for depth estimation and compression than the MPEG-FTV approaches and MVC. We have developed the system and evaluated relationships among an image quality of generated view, dependency of compressed quality factor and the system latency. As a result, we will show that the rendering image quality is enough and the interactivity is highly kept.

The rest of this paper is organized as follows. Section 2 describes the free viewpoint rendering algorithm and the server-client model systems. Section 3 explains the DIBR modes. Experimental results are presented in Section 4. Section 5 concludes this paper.

2. FREE VIEWPOINT RENDERING ALGORITHM AND SYSTEM

2.1 System overview

Requirements of networked free viewpoint server-client system are a) high frame rate, b) high rendering image quality, and c) high interactivity of user input. Several system configurations are supposable for these rendering processes. Now, we consider three cases of the system. The first is multi view transmission system, in which the server sends all multi view images to the client. The second is free viewpoint transmission system, in which the server generates a free viewpoint image and sends only the image. In the last system, the server transmits images and depth maps. Details of each system are described in this section.

Table 1: *Compressed images.*

	Image for Depth Estimation	Depth for View Synthesis	Image for View Synthesis	Free viewpoint Image
Local	RAW	RAW	RAW	RAW
Multi View	Compressed	RAW*	Compressed	RAW
Free Viewpoint	RAW	RAW	RAW	Compressed
Depth & Image	RAW	Compressed	Compressed	RAW

*depth map (multi view) is computed from compressed images.

To generate the free viewpoint image over an IP network, we need five steps; 1) capturing and correction images, 2) depth map estimation, 3) image coding/decoding, 4) data transmission, and 5) free viewpoint view synthesis. The order of these processes chain depends on the case of a system model. Main differences of each system are what types of data are transmitted, in which computer, server or client, performs the depth estimation process, view synthesis process, and what images are compressed. The flowchart of each system is depicted in Fig. 1 and what views are compressed are listed in Table 1, where local is a method of closed one computer system for comparison. To concentrate on the system evaluation, we do not deal with how to estimate the depth map and how to compress the images. To keep high frame rate and image quality, we use the semi-global matching [15] as a depth estimation algorithm. This method has high performance of depth accuracy, while keeping computational cost low. Comparing to the top performance algorithms [11, 12], such as belief propagation [14] and graph cut [13] algorithms, this method has an advantage for computational performance, and the quality of depth map is the second-best.

2.2 Multi view transmission

This system transmits only compressed views from a server to a client. The client decodes the views and then generates a novel view using the nearest 2 views form an input of viewpoint. The system requires high computational performance at the client side and also requires large bandwidth of network. If network condition and computational performance are enough, the system can render free viewpoint image with high quality and low latency. Note that the accuracy of the generating depth map highly depends on the compression rate of the input multi view images

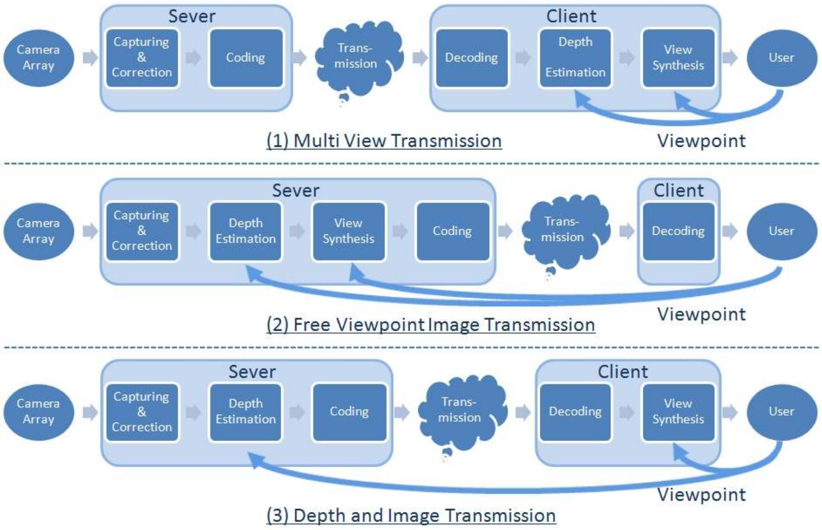


Fig.1: *Systems overview.*

2.3 Free viewpoint transmission

This system transmits a free viewpoint image from a server to a client. The free viewpoint image is generated by the server, and viewpoint information for controlling the rendering viewpoint is input by the client over a network. Thus the timing of viewpoint changing tends to be delayed by network latency and some system computational latencies. Note that we cannot control viewpoint changing at the server side. To keep this changing timing with high interactivity, intra based compression is suitable.

2.4 Depth and image transmission

This system transmits images and depth maps which is computed at a server side to a client. At the client side, the free viewpoint image is generated by views and the associated depth maps. Note that the transmission data is only a few images and depth maps which are a kind of images, so that the system does not require large band width. In addition, this method have high interactivity for the viewpoint changing, because the viewpoint input occurs at the client side and the novel view is synthesized rapidly. However, a generated view is highly affected by depth map quality so that compression of depth map is noticeable. Compression of depth map has different characteristics of usual image compression.

3. DEPTH IMAGE BASED RENDERING MODES

In this section, we introduce a method of free viewpoint view generation via reference images and depth maps. A basic DIBR method of free viewpoint image rendering [16] is as follows. Given two reference left and right images with depth maps, firstly, we warp left and right images to target viewpoint via depth map information. Secondly, we weightily blend warped left and warped right images depending on a viewpoint. Finally we fill holes with depth based or normal image inpainting [17]. A flowchart of free viewpoint image generation based on the DIBR is depicted in Fig. 2. This DIBR flow is called M-view + N-depth architecture adopted in MPEG-FTV. In this architecture, we use the nearest image and depth data set from the viewpoint; thus the parameters are M=N=2.

In our system, we deal 3 modes of the method of the DIBR; a) 1 view + 1 depth map (1V1D), a) 2 views + 2 depth maps (2V2D), c) 2 views+ 1 depth map and prediction (2V1DP). In the 2V2D mode, we use 2 views and 2 depth maps on the views to synthesize a free viewpoint image. This mode is the basic mode of M-view + N-depth frame work. The 2V2D mode has the highest image quality, but rendering time is slowest among the three modes. This is because the 2V2D mode should perform twice depth map estimation process which is the most computationally heavy method. The merit of 2 depth maps estimation is that we can perform LR consistency check [11] which can improve depth accuracy with

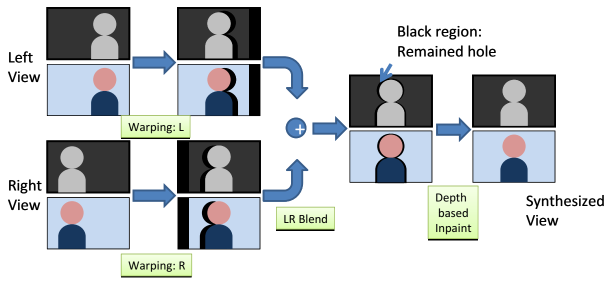
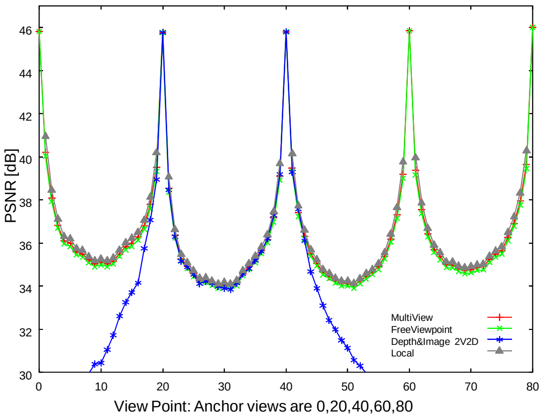
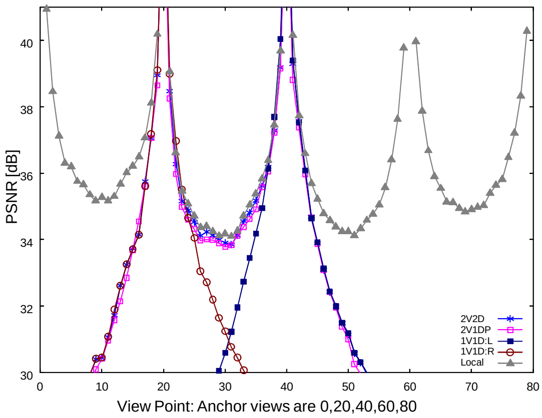


Fig.2: *Flowchart of free viewpoint image rendering with various systems.*

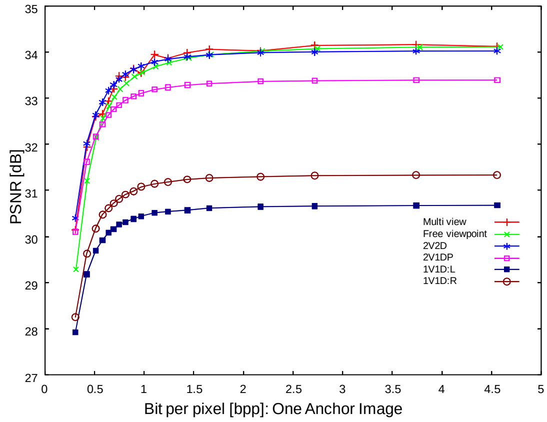


(a) Multi view, free viewpoint and depth&image.

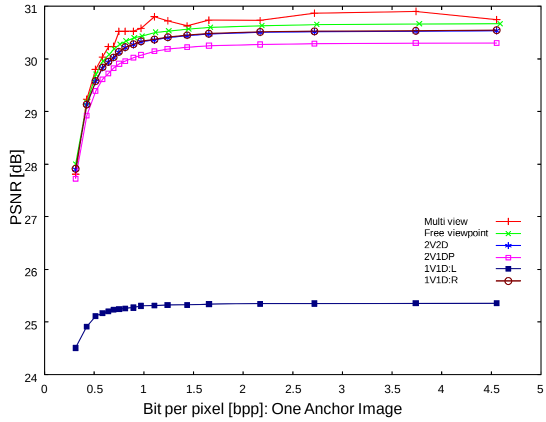


(b) Various depth&image modes.

Fig.3: *PSNR versus viewpoint with various systems.*



(a) Viewpoint is 30, depth is compressed at 0.53 bpp per image.



(b) Viewpoint is 10, depth is compressed at 0.53 bpp per image.

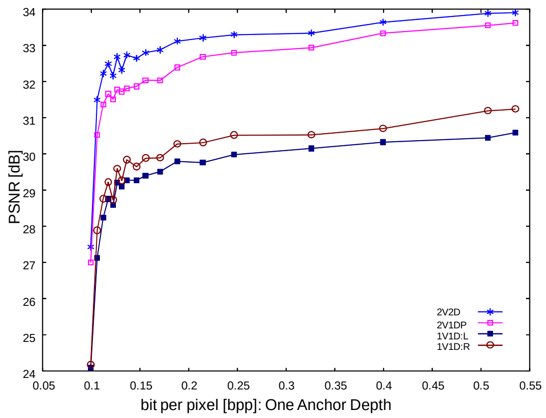
Fig.4: *PSNR versus image compression rate.*

low computational cost. In the 1V1D mode, we use 1 view and a depth map on the view to synthesize a free viewpoint image. Using this mode, a view synthesis flow is slightly different from the flow of the 2V2D mode. Warping second image and blending process are omitted. The 1V1D is the fastest view synthesis mode, but the rendering image quality is the lowest method. The last mode of 2V1DP is similar to the 2V2D mode, but this mode only requires 1 depth map. In this mode, the opposite depth map is predicted by another depth map. The prediction method is the same as view synthesis module without image warping. After prediction, view synthesis process is the same as the 2V2D mode. Note that LR consistency check does not work well, because the prediction of the opposite depth map does not increase geometrical information. In addition, the depth prediction quality highly depends on the depth map compression rate.

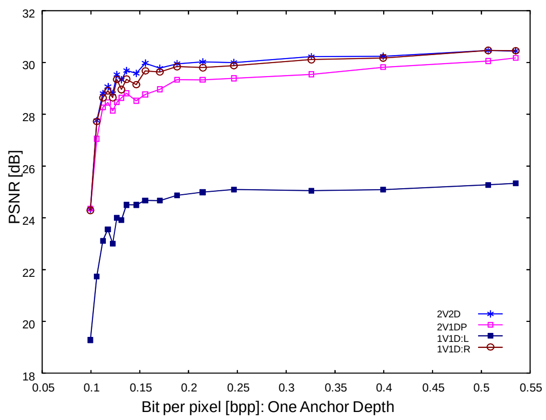
4. EXPERIMENTAL RESULTS

In this section, we evaluate the system models and the DIBR algorithms. We use a multi view sequence which is captured by one camera and machine controlled stage. The sequence contains 81 views and the image resolution is 640 × 480 -24 bit color. Camera setup is 1-D parallel and all views are horizontally rectified by the light field rectification algorithm [18]. We uploaded this sequence on the site in [19]. In our experiments, we use selected cameras, which are camera number 0, 20, 40, 60 and 80, for depth estimation and view synthesis as anchor views. Other cameras are used for only evaluation. We use Peak Signal to Noise Ratio (PSNR) of Y signal in YUV color space as an objective measure.

In the first experiment, we compare the system models. Figure 3(a) shows PSNR of rendering views at the various viewpoints. 4 system models which are multi view, free viewpoint, depth and image (only 2V2D), and local are depicted. The local means that all the processes are computed in one computer and is used as reference. Images are compressed by JPEG and the compressed image has 1.54 bit per pixel (bpp) per one view. The compressed depth map has 0.53 bpp per one depth map. Thus the transmitting data size of multi view system has 5 times larger than that of the free viewpoint system, and the depth and image



(a) Viewpoint is 30, image is compressed at 1.54 bpp per image.



(b) Viewpoint is 10, image is compressed at 1.54 bpp per image.

Fig.5: *PSNR versus depth compression rate.*

system (2V2D) has 2 times larger images and two additional depth maps. As a result, Fig. 3 has peaks on the viewpoint of anchor views, because DIBR keeps the signal of original views. The methods of using all the anchor views, such as multi view, free viewpoint and local, repeat the shape of graph which is decreasing PSNR exponentially from the anchor view to the center of the anchor views and then increasing PSNR correspondingly. The highest quality method is of course the local system, but the free viewpoint and multi view systems keep almost the same quality. The differences between these two methods are transmitting data size and interactivity for viewpoint changing. Latency of viewpoint changing occurs in free viewpoint case, because of viewpoint information communications between the server and client. Only the depth and image system has the different shape. In this condition, we have only number 20 and 40 views. Thus when we get away from these anchor views, PSNR are decreasing. Especially extrapolation cases (view$< 20, 40 >$ view) are obvious. Figure 3(b) shows 4 DIBR modes and the local system as a reference. 2 view cases of the 2V2D and the 2V1DP keep high quality in range of view 20 to

Table 2: *PSNR of various methods at fixed viewpoint.*

	Local	Multi View	Free Viewpoint	2V2D	2V1DP	1V1D Cam. 20	1V1D Cam. 40
View 30	34.17	34.04	33.93	33.91	33.78	31.24	30.59
View 15	36.02	36.00	35.84	33.72	33.67	33.71	26.13
View 10	35.27	35.10	34.99	30.44	30.41	30.45	25.33

40, while in 1 view cases of the left view's (1V1D:L) and the right view's (1V1D:R) 1V1D mode, the image quality is rapidly decreasing. Comparing to the 2V2D and 2V1DP at the viewpoint 30, the difference of the image quality is 0.13 dB. PSNR values at some specified viewpoint are listed in Table 2.

In the second experiment, we evaluate DIBR method by changing a compression parameter. Avoiding burst of cycles of evaluation, we evaluate only representative viewpoint, which is view 30 as the center of the interpolation view and view 10 as an extrapolation case. In the 2V1DP case, we use the depth map of view 40 as an anchor. In the 1V1D:L case, we also use the depth map of view 40 as an anchor. In the case of 1V1D:R, the depth map of view 20 is used as a anchor. Figure 4(a) shows the image compression rate versus PSNR at view 30. The multi view and the free viewpoint transmission results are shown as references. Note that data size of each method is different, that is discussed later. All the modes have the same trend. In the high bit rate range (about 1.0 bpp $>$), the compression rate does not affect the rendering image equality. In the low bit rate range, PSNR is exponentially decreasing. The extrapolation case of view 10 which is depicted in Fig. 4(b) has the same tendency except for the left view 1V1D mode. This view is too far to render high quality images. Figure 5(a) shows the depth compression rate versus PSNR at view 30. Decreasing rate of 1 depth modes, such as the 2V1DP, and the 1V1D, is higher than requiring 2 depth maps mode, because dependency of distance from the reference view for depth prediction/warping and depth map correctness are high. In contrast with image compression case, all the modes depend on the depth compression rate even if the compressed depth quality is high. The extrapolation case of Fig. 5(b) has also the same trends.

Computational times of each step in the free viewpoint generation are;

- image correction: 1 ms
- image coding: 2.5 ms per image
- image decoding: 2 ms per image
- depth estimation: 30 ms per depth
- view synthesis: 4.3 ms
- depth prediction: 3 ms

We use Intel Core i7 920@3.2GHz as a CPU, and NVIDIA GeForce GTX260 as a GPU for computation. Some functions are parallelized by CUDA on the GPU. To sum up each cost obeying the processing chain of Fig.1, the computational times of the server and the client side are shown in Table 3. The multi view and the free viewpoint transmissions can generate any view synthesis method because we have all the views in these methods. Thus we show the case of rendering using one depth map and two depth maps[1].

To evaluate the system, we measure the data size per frame, time of network time cost, frame rate, time of optical signal latency and time of viewpoint change latency under the following conditions;

- image resolution is 640×480 pixels
- image is coded with 1.0 bpp
- depth map is coded with 0.2 bpp
- network band width ($=BW$) is 100 Mbps
- fixed network latency ($=L$) is 10 ms[2]

The example of the system performances is listed in Table 4.

Network t_n shows network time cost which means that the time of data flow from the server to the client if a network cable has no length. This is expressed by the following equation.

$$t_n = \text{size[bit]} \cdot 1000/BW\text{[bps]}. \quad (1)$$

The frame rate of each method depends on the time cost of the server t_s, the time cost of the client t_c and the time cost of network. If we can ideally make these process pipelined parallelization, the frame rate depends only on the maximum cost of the three costs.

$$\text{fps} = 1000/\max(t_s, t_c, t_n). \quad (2)$$

The optical latency, which means the delay of an optical light signal of real world, is as follows.

$$\text{O-latency} = t_s + t_c + t_n + L, \quad (3)$$

where L is fixed network latency which is a factor of physical length of a network cable. The latency of viewpoint change becomes view synthesis time (=4.3ms) except for the free viewpoint system. The free viewpoint system takes additional round trip cost.

$$\text{V-latency} = t_s + t_c + t_n + 2L. \quad (4)$$

Above system performances show that the multi view system uses large network band, and has large optical latency while the method has high interactivity of viewpoint change. The free viewpoint system is a low latency method of optical signal but the high latency of viewpoint change is inevitable. The DIBR system of the 1V1D and 2V1DP modes has the low optical latency while keeping the high interactivity of viewpoint change.

[1]In the PSNR computation experiments, both method are computed by two depth maps.

[2]The parameter assumes a domestic area communication in Japan.

Table 3: *Computational time of each step at server and client. Computational times are shown by (two depth maps / one depth map) in multi view and free viewpoint cases.*

	Multi View	Free Viewpoint	1V1D	2V2D	2V1DP
Server [ms]	13.5	40.8 / 67.8	36.0	71.0	38.5
Client [ms]	47.3 / 74.3	2.0	8.3	12.3	13.3

Table 4: *Example of system parameters. Image and depth map are coded by 1.0 bpp and 0.2 bpp respectively. Network bandwidth is 100 Mbps and network latency is 10 ms.*

	Multi View	Free Viewpoint	1V1D	2V2D	2V1DP
Data size [Mbit]	1.54	**0.3**	0.37	0.74	0.68
Network [ms]	15.4	**3.1**	3.7	7.4	6.8
Frame rate [fps]	21.1 / 13.5	24.5 / 14.8	**27.8**	14.1	26.0
O-latency [ms]	86.2 / 113.2	**55.9** / 82.9	58.0	100.7	68.4
V-latency [ms]	**4.3**	92.9	**4.3**	**4.3**	**4.3**

5. DISCUSSION

The characteristics of each transmission system are as follows;

- Multi view system: We can see any viewpoint with high visual quality and with low latency of the viewpoint change, while we need large network band, and the optical signal latency becomes large.
- Free viewpoint system: We can see any viewpoint with high visual quality with low latency of optical signal, and do not need large network band, while we need tremendous time to change viewpoint.
- DIBR system of 2V2D, 1V1D, and 2V1DP modes: We can see any viewpoint with low latency of the viewpoint change, and with low latency of optical signal, however the visual quality is not guaranteed. The quality has the peak around the anchor views. We need a middle capacity of network band.

For FTV applications, such as TV broadcast, TV conference, movies in a storage media and surveillance, requirements of FTV performance highly depend on the types of application. Thus a suitable transmission type also depends on the applications. For example, the storage media and the TV broadcast (except for live sports broadcasting) require high video quality and interactivity of viewpoint change while they are not strict for the latency of optical signal. Thus the multi view system with more effective compression method, such as multi view video coding, becomes a better solution. For surveillance application, the visual quality is important but the frame rate and interactivity of viewpoint change are not so important. Thus the free viewpoint system is a good solution. The selective DIBR system is suitable for live TV conferences. This application requires high interactivity of viewpoint change and low optical latency, but requirement for visual quality is lower than the other types of application. In the TV conferences, large viewpoint change rarely happens, thus the disadvantage of the proposed method that the quality is not guaranteed is not a big problem. The frame rates of DIBR system, especially the 1V1D and 2V1DP cases, are almost 30 fps with VGA resolution. Thus the application has enough performance except for data size. The coding method of this paper is the simplest method of JPEG intra coding so that the coding efficiency is not enough. The coding methods of extension of shared I filed and selective B field that are proposed by the context of multi view video compression or other effective approaches are required for realizing this application.

With the DIBR system of the 2V2D and 2V1DP cases, the acceptable range of viewpoint change would be ±3 cm, if we assume that the peak value in the PSNR curve of concave between anchor views (see in Fig. 3), whose distance is 4 cm in the experiments, is the lower bound of image quality. In the 1V1D case, ±1 cm around anchor view is an acceptable range. This means that we can render good quality images within the range of one-and-a-half time of baseline in the 2 views case, and within the range of half length of baseline in the 1 view case.

However, this results depend on the number of disparity range D_{range}, which is shown in [16]. D_{range} is express as follows;

$$D_{range} = D_{max} - D_{min} = fl(\frac{1}{1/z_{near}} - \frac{1}{1/z_{far}}), \quad (5)$$

where D_{max}, D_{min} are the max and min disparity values respectively, z_{near}, z_{far} are the depth values of the nearest object and the farthest object, f, l is the focal length of camera and the length of baseline of the nearest camera.

If we change the camera array condition, such as the focal length, the image resolution and the length of camera baseline, the above parameters vary. In addition, when positions of objects in the cameras are changed, the parameters are also changed. Thus changing camera condition depending on the positions of the objects is important to keep the image quality high.

6. CONCLUSION

In this paper, we have remarked about networked free viewpoint image rendering which requires not only high image quality and high frame rate but also low latency of viewpoint changing. We have presented the depth image based rendering (DIBR) system and have dealt with 3 modes in DIBR; 1 view + 1 depth (1V1D), 2 views + 2 depths (2V2D), and 2 views + 1 depth + prediction (2V1DP). Comparing to other two extreme system models, such as the only free viewpoint image transmission system and the multi view transmission system, the DIBR system can render a free viewpoint image with low latency, and does not require high computational performance at a client side while the rendering image hardly lose the image quality. In addition, we have revealed that relationships between distance from reference views, the rendering image quality. To keep the quality of the image high when the viewpoint exists between the reference views, the 2V2D and the 2V1DP mode are better than the 1D1V mode. Considering the rendering frame rate, the 2V1DP is the best among the 3 modes, because the loss of image quality from the 2V2D is slight little. In addition, we have revealed the dependency of depth-image compression rate and the dependency of the distance from reference views. As a result, the image compression rate does not affect image qualities in high bit rate range. On the contrary, in the depth compression case, the quality of the synthesized image is deteriorated in all the ranges. At the aspect of 1 depth case, such as 2V1DP and 1V1D, the characteristics become more remarkable, thus applying high compression rate to depth map is not suitable. The network latency depends on an environment we do not evaluate this factor, but this DIBR method has no latency for viewpoint changing. Thus we will develop a novel view synthesis oriented compression algorithm. In addition, we will research about the dependency between compression and rendering.

7. ACKNOWLEDGEMENT

This work was partly supported by the Grand-In-Aid for Young Scientists (B) of Japan Society for the Promotion of Science under Grant 22700174 and The Hori Information Science Promotion Foundation.

References

[1] T. Fujii, T. Kimoto, and M. Tanimoto, “Ray Space Coding for 3D Visual Communication,” Proc. of PCS’96, pp. 447–451, Mar. 1996.

[2] M. Levoy, and P. Hanrahan, “Light Field Rendering,” Proc. of ACM SIGGRAPH’96, pp. 31–42, Aug. 1996.

[3] J. X. Chai, S. C. Chan, H. Y. Shum, and X. Tong, “Plenoptic Sampling,” Proc. of ACM SIGGRAPH’00, pp. 307–318, Aug. 2000.

[4] C. Fehn, “Depth-Image-Based Rendering (DIBR), Compression, and Transmission for a New Approach on 3D-TV,” Proc. of SPIE, Vol. 5291, pp. 93–104, Jan. 2004.

[5] M. Tanimoto, “Overview of Free Viewpoint Television,” Signal Processing: Image Communication, Vol. 21, Issue 6, pp. 454–461, July 2006.

[6] A. Vetro, S. Yea, M. Zwicker, W. Matusik, and H. Pfister, “Overview of Multiview Video Coding and Anti-Aliasing for 3D Displays,” Proc. of ICIP’07, Vol. 1, pp. 1–17, Sep. 2007.

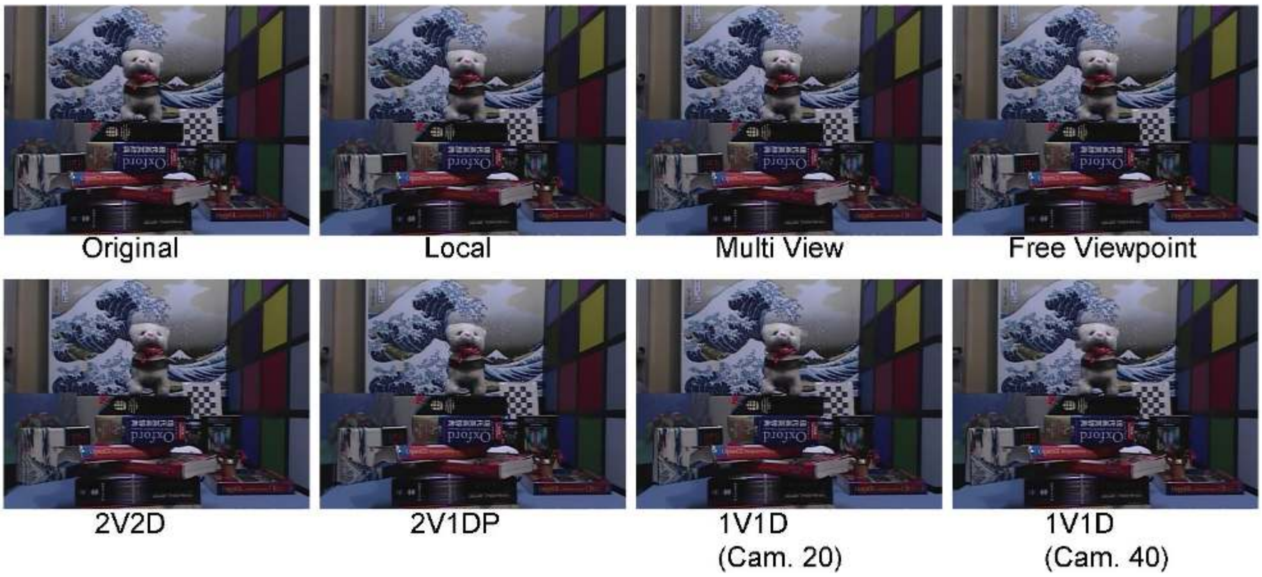


Fig.6: *Rendering results: viewpoint 30.*

[7] K. Yamamoto, M. Kitahara, H. Kimata, T. Yendo, T. Fujii, M. Tanimoto, S. Shimizu, K. Kamikura, and Y. Yashima, "Multiview Video Coding Using View Interpolation and Color Correction," IEEE Trans. on CSVT, Vol. 17, Issue 11, pp. 1436–1449, Nov. 2007

[8] Z. Han and Q. Dai, "A New Scalable Free Viewpoint Video Streaming System over Ip Network," Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing, pp. II-773–II-776, Apr. 2007.

[9] X. Cao,Y. Liu, and Q. Dai, "A Flexible Client-driven 3DTV System for Real-time Acquisition, Transmission, and Display of Dynamic Scenes," EURASIP Journal on Advances in Signal Processing, article ID 351452, doi:10.1155/2009/351452, 2009.

[10] C. Fehn, P. Kauff, M. Op De Beeck, F. Ernst, W. IJsselsteijn, M. Pollefeys, L. Van Gool, E. Ofek, and I. Sexton, "An Evolutionary and Optimised Approach on 3D-TV," in Proc. International Broadcast Conference, pp. 357–365, Sep. 2002.

[11] D. Scharstein and R. Szeliski, "A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms," Int. J. of Computer Vision, Vol. 47, Issue 1–3, pp. 7–42, Apr. –June 2002.

[12] D. Scharstein and R. Szeliski, Middlebury Benchmark. http://vision.middlebury.edu/stereo/.

[13] Y. Boykov, O. Veksler, and R. Zabih, "Fast Approximate Energy Minimization via Graph Cuts," IEEE Trans. PAMI, Vol. 23, No. 11, Nov. 2001.

[14] P. Felzenszwalb and D. Huttenlocher. "Efficient Belief Propagation for Early Vision," Int. J. of Computer Vision, Vol. 70, No. 1, pp. 41–54, Oct. 2006.

[15] H. Hirschmuller, "Stereo Processing by Semi-Global Matching and Mutual Information," IEEE Trans. on PAMI, Vol. 30, Issue 2, pp. 328–341, Feb. 2008.

[16] Y. Mori, N. Fukushima, T. Yendo, T. Fujii, and M. Tanimoto, "View Generation with 3D Warping Using Depth Information for FTV," Signal Processing: Image Communication, Vol. 24, No. 1–2, pp. 65–72, Jan. 2009.

[17] S. Zinger, L. Do, and P.H.N. de With, "Free-Viewpoint Depth Image Based Rendering, " J. of Visual Communication and Image Representation, Vol. 21 , Issue 5–6, pp. 533–541, July 2010.

[18] M. Ota, N. Fukushima, T. Yendo, M. Tanimoto, and T. Fujii, "Rectification of Pure Translation 2D Camera Array," Proc. of IWAIT'09, 0044, Jan. 2009.

[19] "2D Camera Array Multi View Sequence," http://nma.web. nitech.ac.jp/fukushima/multiview/2darray.html.

Norishige Fukushima received a B.E., M.E., and Ph.D. degree from Nagoya University, Japan, in 2004, 2006, and 2009, respectively. Since 2009, he has been an assistant professor at Graduate School of Engineering, Nagoya Institute of Technology, Japan. His research interests are multi view image capturing, calibration, processing, and coding.

Yutaka Ishibashi received the B.E., M.E., and Dr.E. degrees from Nagoya Institute of Technology, Nagoya, Japan, in 1981, 1983, and 1990, respectively. In 1983, he joined the Musashino Electrical Communication Laboratory of NTT. From 1993 to 2001, he served as an Associate Professor of Department of Electrical and Computer Engineering, Faculty of Engineering, Nagoya Institute of Technology. Currently, he is a Professor of Department of Scientific and Engineering Simulation, Graduate School of Engineering, Nagoya Institute of Technology. His research interests include networked multimedia, QoS (Quality of Service) control, and media synchronization. He is a fellow of IEICE and a member of IEEE, ACM, IPSJ, ITE, and VRSJ.