SYNTHESIS ERROR COMPENSATED MULTIVIEW VIDEO PLUS DEPTH FOR REPRESENTATION OF MULTIVIEW VIDEO

Mehrdad Panahpour Tehrani[1], Akio Ishikawa[1], Makoto Okui[1], Naomi Inoue[1]
Keita Takahashi[2], Toshiaki Fujii[2]

[1]NICT-National Institute of Information and Communications Technology, Japan
[2]Nagoya University, Japan

ABSTRACT

SECOND-MVD (Synthesis Error COmpeNsateD Multiview Video plus Depth) is an alternative 3D format that we introduce for representation of multiview video. In this data format, images at some viewpoint remain original, and the others are converted to a novel format. Residual based representation, such as layered depth video and free-viewpoint TV data unit were proposed. We propose hybrid image that not only consists of residual but also remainder pixels. Generation and reconstruction process of a hybrid image uses virtual image synthesized by the images that remained original in SECOND-MVD. In this paper, we investigate the compression performance of SECOND-MVD using hybrid image. Experiments demonstrate reduction in bit rate using hybrid image against residual image in SECOND-MVD framework.

Index Terms—Synthesis error, multiview video plus depth, remainder, residual, 3D format

1. INTRODUCTION

In the recent years, 3D video attracted many attentions and applications such as 3DTV [16][17] and FTV (Free-viewpoint TV) [18][19]. These new mediums introduce different 3D depth impression. Current 3D display can only provide very narrow view angle at a moment. However, SMV (Super Multiview Video) displays, with wide motion parallax are emerging widely, anticipated as the next generation of auto-stereoscopic display. Several displays are currently being developed, such as 3DTV by MERL[2], REI (Ray Emergent Imaging) by NICT [3], Holovizio by Holographica [4], SMV256 by Tokyo University of Agriculture and Technology [5], Light Field 3D Display by Samsung[6], IP (Integral Photography) by NHK [7] and 3D VIVANT[8], Projection type IP by Seoul National University display [9], 360 degree LFD by USC [10], Seelinder by Nagoya University and The University of Tokyo [11], Holo Table by Holy Mine [12], fVisiOn by NICT [13], Walk-through by KDDI[14], etc. Fig. 1 shows REI [3] system and snapshots of the displayed images. Since these super multiview 3D displays require a huge number of multiview images in realtime, a new coding standard is essential to realize their services and products in the market [15]. In 105th ISO/IEC JTC1/SC29/WG11 MPEG (Motion Picture Expert Group) meeting a new AHG (Ad Hoc Group) on FTV was established [1] to address these new use cases, and their coding standard.

Previously, MVV (Multi-View Video) coding has been investigated by MPEG since 2004. MVC (Multiview Video Coding) [20][21] was standardized in 2007 and has been commercialized since 2009 for stereoscopic displays. MVC exploits time domain motion compensation in H.264/MPEG-AVC (Advanced Video Coding) [22] for interview prediction. Later, for autostereoscopic displays, MPEG defined Multiview Video plus Depth (MVD) and targeted the standardization of 3DV (3D-Video) coding, which is adopted in 2008. MPEG is currently working on compression standard of MVD. MVD requires less number of data for transmission, and capturing devices.

MVD can be initial candidate data format for SMV contents [2-13]; however the data size for compression is still huge. Alternative formats such as LDI (Layered Depth Image) [23], LDV (Layered Depth Video) [24] and (FDU) (FTV data unit) [25] were also proposed based on MVD. They consist of residual values, where one or more views remain as original the rest are generated from the difference between the synthesized warped original view and neighbor views. The warped images are generated by using depth map, i.e. DIBR (Depth Image Based Rendering) method [26][34]. Depth maps are mainly generated by using graph-cut energy minimization [32][33]. Depth maps representations also can be considered [27][28]. Methods in [27][28] generate two types of global depth maps with two different approaches. In both, the global depth map represents two or more depth maps in a single depth map. The global view in [27][28] are the residual based data format. However, compression of the data formats based on residual has loss of synthesis error, in area with small residual values. Therefore a better compensation method than only using residual is needed.

978-1-4799-2893-4/14/$31.00 ©2014 IEEE

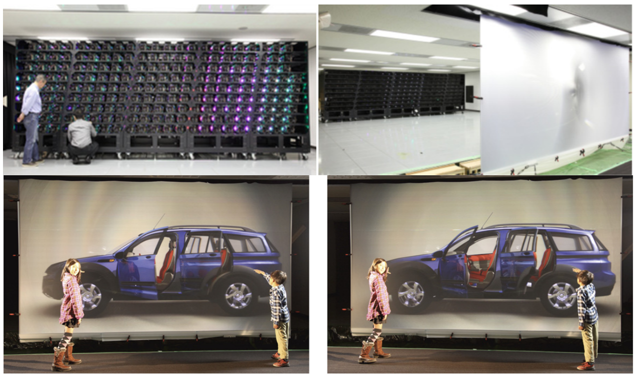

Fig. 1 (top) REI, the NICT SMV display, projector array and screen. (bottom) Two views at a frame from two different angles show horizontal motion parallax.

To address this problem, we propose SECOND-MVD as an alternative data format for SMV. In this format, images at some viewpoint remain as original, and the others are converted to a novel format, namely hybrid image. Hybrid image [29] not only consists of residual but also remainder [30] pixels. We investigate the compression performance of SECOND-MVD using hybrid image, against using residual in SECOND-MVD framework.

The rest of the paper is organized as follows. Section 2 explains the SECOND-MVD as a new 3D format, its generation and reconstruction methods. Experimental results are shown in section 3. We conclude the paper in section 4.

2. SECOND-MVD

SECOND-MVD consists of original and converted images. Converted images are generated and reconstructed using the neighbor images, i.e. left and right side reference views that are remained unchanged (*VL/DL* and *VR/DR*). The target image (*VT*) is converted to a new format image, i.e. hybrid image. In SECOND-MVD, the depth maps remain unchanged. Fig. 2 shows MVD and SECOND-MVD.

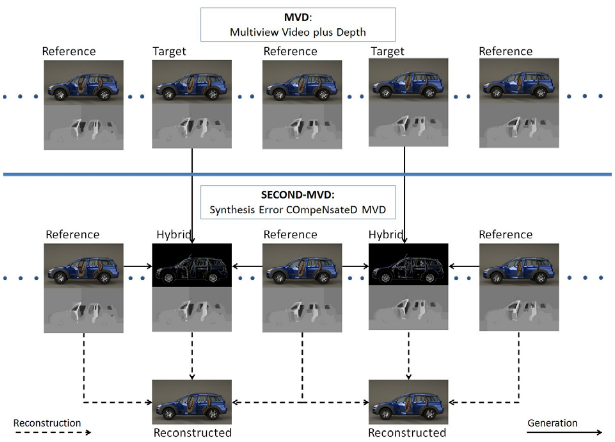

Fig. 2 MVD and SECOND-MVD data format configurations. Data conversion flowchart for SECOND-MVD from MVD.

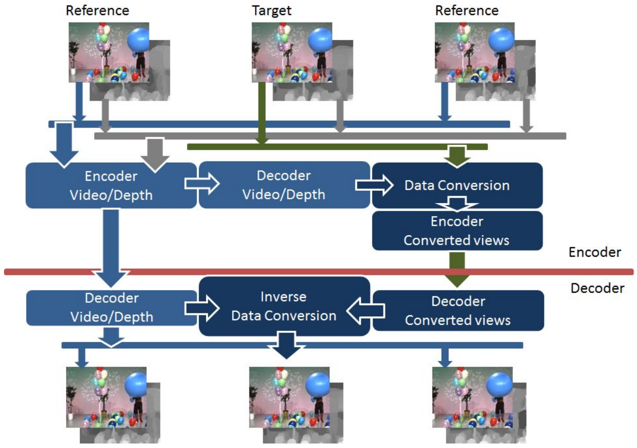

Fig. 3 Configuration for Compression of SECOND-MVD.

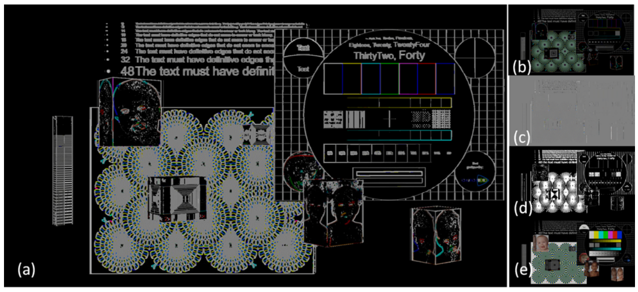

Fig. 4 (a) Hybrid image (residual & remainder values), (b) remainder image, (c) residual image, (d) estimated error mask, and (e) Original image. For visualization, remainder image is scaled, and contrast in residual is increased.

We propose hybrid image that consist of both residual and remainder pixels. To decide areas corresponding to residual or remainder a mask shall be identically estimated at both, data conversion inverse conversion steps.

2.1. Data conversion

The steps toward generating a hybrid image are shown in following subsections. Fig. 3 shows that before data conversion, the reference views are encoded/decoded so that the coding error can also be compensated by the hybrid image. Fig. 4 shows example of hybrid, residual, remainder, estimated error mask images at the location of the hybrid image, and the original image.

2.1.1. View synthesis

Intermediate virtual view *VT(virtual)* at *VT* location is synthesized using reference views, *VL/DL* and *VR/DR*. The view synthesis is 3D warping [31][26][34] with depth maps.

2.1.2. Residual image

Residual image, *VT(residual)*, is the subtraction of a virtually generated and the original target images. It may follow by 1-bit reduction in bit-plane.

2.1.3. Remainder image

Symbol (M) is the remainder of division of dividend (I) by divisor (D), i.e. $I = q \mathrm{x} D + M$, $I \geq 0$, d $D > 0$, and q is the quotient. Remainder image, *VT(remainder),* is the output of modulo operation "$M = VT \bmod D$" on each pixel value of target image, given a divisor (D) [30] value. D is chosen from a look up table (LUT), given a gradient value of the same location. Gradient values are pixel values in gradient-like image, *VT*(*gradient*) that is generated from the synthesized virtual image at the target view location. Gradient-like image is generated by applying (i) Gaussian filter; (ii) Canny edge detector; (iii) Series of morphological operations consist of dilation-erosion-erosion-dilation, and again (iv) Gaussian filter, on *VT(virtual)*.

2.1.4. Estimation of synthesis error mask

Error mask is required to distinguish corresponding areas in hybrid image as residual or remainder. It is important that the error mask is estimated by only reference views, i.e. *VL/DL* and *VR/DR* to be independent of using original image of *VT*. By estimating the error mask, we can make sure that the same error mask can be re-estimated at reconstruction of a hybrid image. Fig. 5 summarizes the processes for estimation of synthesis error mask. Note that the procedure is not simply thresholding the residual image.

2.1.5. Hybrid image generation

Due to loss of synthesis error with small value in residual image, they are replaced with remainder to generate the hybrid images as follows. Therefore, given the estimated error mask, the areas without mask are the area with low error value, so the hybrid image in these areas is inherited from remainder values. The areas with mask correspond to high synthesized error; therefore they are filled by residual values in the hybrid image.

2.2. Inverse of data conversion

The algorithm for inverse data conversion performs view synthesis and estimation of synthesis error mask (as in 2.1.4). Using the estimated error mask, we distinguish the areas corresponding to remainder and residual areas. Finally, the inverse of generation processes for residual and remainder parts [29][30] are performed, respectively.

Reconstruction of the residual parts is the summation of and residual and *VT(virtual).* To reconstruct the remainder parts, an inverse modulo operation [30] is performed as follows. At first *VT*(*gradient*) is generated. Then, the reconstruction candidates "$C(q)$" for remainder "M" are listed "$C(q) = q \mathrm{x} D + M \leq N$ where $q \in \{0,1,2,\ldots, N\}$". For each pixel in the remainder image with value "M", the inverse operation finds a "$C(q)$" in the list that has the minimum difference with the same pixel location in the *VT*(*virtual*). N is the maximum intensity value in an image.

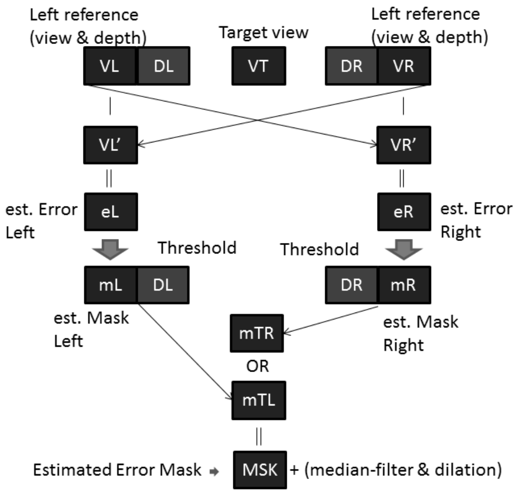

Fig. 5 Procedure to estimate synthesis error mask, which is identical at encoder and decoder sides.

3. EXPERIMENTAL EVALUATIONS

To evaluate our framework, we conducted the compression experiments. Results show the improvement in coding efficiency of SECOND-MVD. Parameters for hybrid image generation/reconstruction are set through a course of pre-experiments to maximize the performance. The compression ratio is adjusted by varying the QP.

For compression evaluation five test sequences are used. Two of them were generated for SMV displays, "Sphere" and "Blue-car", and three are from MPEG test sequences, "Balloons (1024x768, 30fps), "Kendo (1024x768, 30fps)" and "Champagne_tower (1280x960, 30fps)". For MPEG sequences 5 views are used with 5cm baseline, and depth maps are available at 1st, 3rd and 5th views, i.e. 10cm baseline. The computer generated test sequences contain 177 views, 1920x1080, single frame. A 45 downsized set of views (960x540) out of 177 views are selected with 4-baseline distance (4x2.28cm=9.12cm). Given the downsized set, depth maps for all 45 views are estimated [32][33]. Therefore, the input data is MVD, 45view+45depth.

PSNR in the graphs is the average of decoded and synthesized PSNRs. One synthesized image is generated in the middle of each pair of input images. Therefore, for MPEG sequences, we have (3 decoded + 2 synthesized), and for the SMV sequences (45 decoded + 44 synthesized). Bit rates are total bit rate for compressed data.

For all sequences we performed following experiments. Fig 6 shows RD results for computer generated test sequences "Sphere" and "Blue-car". In this experiment, color and depth QPs of reference views are c46d48, c40d46, c35d40, c25d35, while QP for converted image is constant, QP=36. MVC codec is used, with IPPP coding configuration, and GOP size of 12. Reference views are one set, and hybrid views are another set that are coded separately.

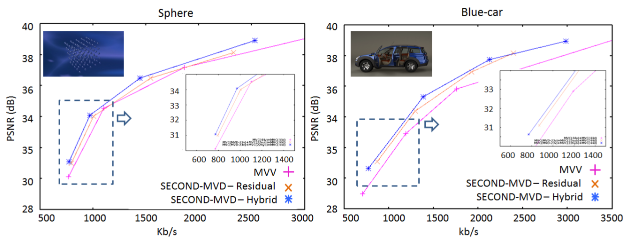

Fig. 6 Compression performance: Average PSNR vs. total bit rate of views and/or depth. RD curve for MVD, SECOND-MVD-Residual and SECOND-MVD-Hybrid encoded/decoded by MVC. QP for converted views is 36, QPs for color/depth for MVD and unconverted view in SECOND-MVD are variable, c46d48, c40d46, c35d40, c25d35. (left) Sphere; (right) Blue car.

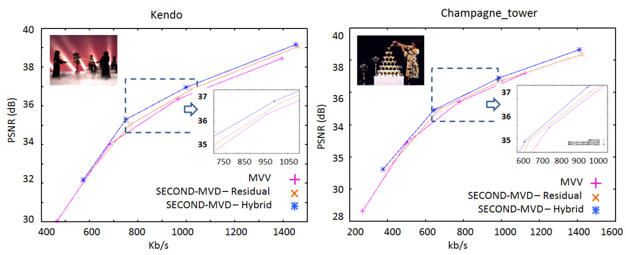

Fig. 7 Compression performance: Average PSNR vs. total bit rate of views and/or depth. RD curve for MVD, SECOND-MVD-Residual and SECOND-MVD-Hybrid encoded/decoded by MVC. QP for converted views is 36, QPs for color/depth for MVD and unconverted view in SECOND-MVD are variable. (left) Kendo, c48d48, c40d48, c36d42, c32d38; (right) Champagne tower, c48d50, c40d38, c36d34, c33d30.

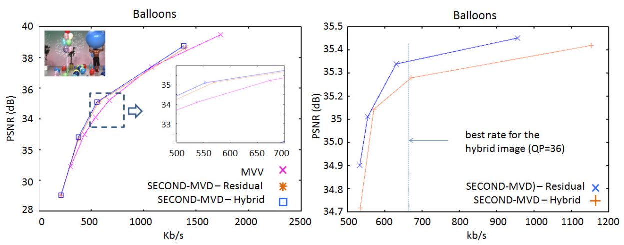

Fig. 8 Compression performance: Average PSNR vs. total bit rate of color and/or depth. RD curve for MVV, SECOND-MVD-Residual and SECOND-MVD-Hybrid encoded/decoded by MVC. (left) QP for converted views is 36, QPs for color/depth of MVV and unconverted view in SECOND-MVD are variable, c48d48, c40d48, c36d42, c28d38; (right) QPs for color/depth is c36d42, QP for converted view is variable, 24, 30, 36, 44.

QP combinations for "Balloons" and "Kendo" are c48d48, c40d48, c36d42, c32d38. QP combinations for "Champagne_tower" are c48d50, c40d38, c36d34, c33d30. For converted images QP is 36. Fig. 7 and Fig. 8(left) show the compression performance for "Kendo", "Champagne tower" and "Balloons". For these test sequences, MVC codec was used for compression of 48 frames, with Hierarchical prediction in time and view, GOP (Group of Picture) size 12.

Fig 6, Fig 7, and Fig. 8(left) show 10%~20% saving in bit rate, at about 35dB PSNR, using SECOND-MVD-hybrid, against MVV. At the same PSNR, SECOND-MVD using hybrid image has maximum 5% reduction of bit rate in comparison with SECOND-MVD using residual image only.

In Fig 8(right), we show different RD evaluation result, where the color/depth QPs are constant (c36d42) for reference view, and the QP for converted images varies, i.e. 24, 30, 36, 44. Fig 8(right) shows comparison of SECOND-MVD with the hybrid and residual images, when color and depth QPs are constant and QP for compression of converted image is varied. Such a test can be used to find a suitable bit rate for converted images in SECOND-MVD. In this case, QP=36 gives nearly the highest improvement in PSNR for a lowest possible bit rate.

4. CONCLUSIONS

In this paper, we introduced SECOND-MVD data format for multiview video representation. It has potential application in transmission of large number of views for super multiview displays. The new data format compensates the synthesis error at the viewpoint where the image is converted to the new format. In this scenario, we introduced the hybrid representation as a novel data format for multiview video. A hybrid image consists of residual and remainder values, while the state-of-arts use either residual or remainder. After compression, the residual values recovers the large synthesis error, whereas the remainder the small synthesis error. Experiments show better reconstruction quality by SECOND-MVD using hybrid image, than only using residual or remainder images. Furthermore, compression performance verified effectiveness of using hybrid image in SECOND-MVD by reducing bit rate against SECOND-MVD using residual image.

Optimizing the parameters and coding configuration leads to a better coding efficiency that is our future step.

5. REFERENCES

[1] "AHG on FTV (Free-viewpoint Television)", ISO/IEC JTC1/SC29/WG11/N13832, July-Aug. 2013.

[2] W. Matusik, H. Pfister, "3D TV: A scalable system for real-time acquisition, transmission and autostereoscopic display of dynamic scenes", ACM SIGGRAPH, 23(3): 814-824, Aug 2004.

[3] S, Iwasawa, M. Kawakita, and N. Inoue, "REI: an automultiscopic projection display," 3DSA 2013, Jun 2013.

[4] http://www.holografika.com/

[5] Y. Takaki and N. Nago, "Multi-projection of lenticular displays to construct a 256-view super multi-view display," Opt. Express 18, 8824-8835 (2010).

[6] J. Lee, J. Park, D. Nam, S. Young Choi, D. Park, C. Y. Kim, "Optimal projector configuration design for 300-mpixel light-field 3D display" SID Sym, Tech Paper, 44(1):400–403, Jun 2013.

[7] J. Arai, F. Okano, M. Kawakita, M. Okui, Y. Haino, M. Yoshimura, M. Furuya, and M. Sato, "Integral three-dimensional television using a 33-megapixel imaging system," Jr. of Display Tech., 6-10, 422-430, Oct. 2010.

[8] JCT3V-C0210: 3D holoscopic video test material, FP7 3D VIVANT Consortium, Jan. 2013.

[9] Hong J, Kim Y, Park SG, Hong JH, Min SW, Lee SD, Lee B., "3D/2D convertible projection-type integral imaging using concave half mirror array" Opt Exp. 27,18(20):20628-37, Sep2010.

[10] http://gl.ict.usc.edu/Research/3DDisplay/

[11] T. Yendo, T. Fujii, M. Tanimoto, M. Panahpour Tehrani, "The Seelinder: cylindrical 3D display viewable from 360 degrees", Jr of Vis. Com. and Image Rep, 21:5-6, 586-594, (2010)

[12] http://www.holymine3d.com/prod/prod03.html

[13] S. Yoshida, "fVisiOn: glasses-free tabletop 3D display to provide virtual 3D media naturally alongside real media," Proc. SPIE 8384, 3D Imag. Visu., and Disp. 838411 (1-9), Apr. 2012.

[14] A. Ishikawa, M. P. Tehrani, S. Naito, S. Sakazawa A. Koike, "Free viewpoint video generation for walk-through experience using image-based rendering", ACM-MM 1007-1008, (2008)

[15] M. P. Tehrani, T. Senoh, M. Okui, K. Yamamoto, N. Inoue, T. Fujii, H. Nakamura, "Proposal to consider a new work item and its use case - REI : an ultra-multiview 3D display -", ISO/IEC JTC1/SC29/WG11/m30022, July-Aug. 2013.

[16] A. Kubota, A. Smolic, M. Magnor, M. Tanimoto, T. Chen, C. Zhang, "Multiview imaging and 3DTV", SPM 24(6)10-21, 2007.

[17] A. Smolic, P. Kauff, S. Knorr, A. Hornung, M. Kunter, M. Muller, M. Lang, "Three-dimensional video postproduction and processing", Proc. of IEEE 99(4), 607-625, 2011.

[18] M. Tanimoto, M. P. Tehrani, T. Fujii, T. Yendo, "Free viewpoint TV," IEEE SPM 28(1), 67–76, 2011.

[19] M. Tanimoto, M. P. Tehrani, T. Fujii, T. Yendo, "FTV for 3D spatial communications," Proc. of IEEE, 100(4), 905-917, 2012.

[20] "Introduction to multi-view video coding", ISO/IEC JTC 1/SC 29/WG11, N7328, Jul. 2005.

[21] Merkle, A. Smolic, K. Müller, T. Wiegand, "Efficeint compression of multi-view video exploiting inter-view dependency based on H.264/MPEG-AVC," Proc. ICME ,1717-1720, 2006.

[22] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard", IEEE Trans. on CSVT, 13- 7, 560-576, (July 2003).

[23] J. Shade, S. J. Gortler, L. He, R. Szeliski, "Layered depth images," Proc. SIGGRAPH 98, 231-242, Jul. 1998.

[24] K. Müller, A. Smolic, K. Dix, P. Merkle, P. Kauff T. Wiegand, "Reliability-based generation and view synthesis in layered depth video", Proc. of IEEE Intl Conf on MMSP, 34-39, Oct.2008.

[25] M. Tanimoto, M. Wildeboer, "Frameworks for FTV coding", Proc. of PCS, 1-4, May 2009.

[26] C. Fehn, "Depth-image-based rendering (DIBR), compression and transmission for a new approach on 3D-TV", Proc. of SPIE Ster. Disp. and Virt.l Reality Systems, XI, 93-104, Jan.2004.

[27] T. Ishibashi, T. Yendo, M. Panahpour Tehrani, T. Fujii, M. Tanimoto, "3D space representation using epipolar plane depth image", Proc. of PCS, 1-4, Dec 2010.

[28] T. Senoh, K. Yamamoto, R. Oi, Y. Ichihashi, T. Kurita, "Simple multi-view coding with depth maps"3DSA 223-227, 2012.

[29] M. P. Tehrani, A. Ishikawa, M. Kawakita, N. Inoue, T. Fujii, "A hybrid represenation for multi-view image," 3DTV-CON 2012.

[30] M. P. Tehrani, T. Fujii, M. Tanimoto, "The adaptive distributed source coding of multi-view images in camera sensor networks", IEICE Trans, E88-A(10), 2005.

[31] Y. Mori, N. Fukushima, T. Yendo, T. Fujii M. Tanimoto, "View generation with 3D warping using depth information for FTV," Signal Pro Image Com., 24, 65– 72, Jan. 2009.

[32] R. Szeliski, R. Zabih, D. Scharstein, O. Veksler, V. Kolmogorov, A. Agarwala, M.Tappen and C. Rother, "A comparative study of energy minimization methods for Markov random fields with smoothness-based priors," PAMI, 30(6), 1068-1080, 2008.

[33] Y. Boykov, O. Veksler and R. Zabih, "Fast approximate energy minimization via graph cuts,"PAMI 1222–1239, Nov. 2001.

[34] L. Yang, T. Yendo, M. P. Tehrani, T. Fujii and M. Tanimoto, "Probabilistic reliability based view synthesis for FTV", Proc. of ICIP, 1785-1788, Sep. 2010.