Optimization of camera positions for free-navigation applications

Marek Domański, Adrian Dziembowski, Adam Grzelka, Dawid Mieloch
Poznań University of Technology,
Chair of Multimedia Telecommunications and Microelectronics
Poznań, Poland

Abstract — **In the article we deal with the problem of camera positioning in sparse multiview systems with applications to free navigation. The limited number of the cameras, though makes the system relatively practical, implies problems with proper depth estimation and virtual view synthesis, due to increased amount of the occluded areas. We present experimental results for the optimal positioning of the cameras, depending on two factors – characteristics of an acquired scene and the multi-camera system (linear or circular camera setup). The results show the correlation between the number of occlusions in the scene and a gain from using camera pairs instead of uniformly distributed cameras.**

Keywords — multiview video acquisition; free-viewpoint television; free navigation; depth estimation; view synthesis.

I. Introduction

In this paper, we deal with future interactive video services that will provide a user an ability to navigate through natural dynamic scenes, or at least to virtually navigate around such scenes. The virtual navigation means that a viewer, in any time instant, is able to watch the scene as a video taken from a virtual viewpoint that is changing according to the motion of a viewer fingers on a touch pad of a laptop or on a touch screen of a smartphone or tablet, or even by the motion of the viewer eyes registered by an eye-tracking device. The systems that will provide such a functionality are often referred as free-viewpoint television (FTV) [1], where the audiovisual content is acquired by multiple synchronized cameras and microphones located around a scene. The forthcoming practical applications, where free navigation will play a vital role, have been already discussed in several papers, e.g. [1,2,3]. In this paper, we have in mind mainly such potential applications like interactive broadcasts of sports (in particular volleyball, judo, wrestling, sumo, dance etc.) and performances (theater, circus), as well as interactive courses (medical, cosmetics, dance etc.), interactive manuals and interactive school teaching materials.

Let us restrict our considerations to multiview video content thus leaving the audio processing beyond the scope of this paper. Thus we consider multiview video that has to be corrected (lens aberrations, individual color characteristics of cameras etc.) and calibrated (precise calculation of the intrinsic and extrinsic camera parameters) [4]. Such corrected and calibrated video may be used to create the visual representation of a natural dynamic scene [5]. The current research and development activities use mostly the multiview video plus depth (MVD) representation [6] but other representations may also be considered for free navigation applications, like object-based [7,8] or ray-space oriented [1,9]. Any such representation should be rich enough to provide the ability to calculate, at any time, a virtual view from an arbitrary position.

In this paper, we choose the multiview plus depth representation of a scene, as most works on free navigation have been done in the context of this representation. Estimation of such a representation requires estimation of depth for consecutive frames of multiview video taken from several cameras. For video, the multiview depth estimation is very demanding in a sense of the computational power required but its accuracy is crucial for the quality of the output synthetic video. The study of the depth estimation techniques [10] must be left beyond the scope of this paper. Let just assume application of the efficient state-of-the-art technique. For the experiments we use the technique [5] that improves the popular state-of-the-art technique implemented in the MPEG reference software [11]. This technique is based on the graph-cut algorithm [12].

The general study of the solutions for the free navigation systems must be left beyond the scope of the paper but it can be found elsewhere, e.g. in [1,5,8,13,14]. Here, we focus on the multicamera video acquisition system that should be designed and implemented in such a way that the quality of the output video is as good as possible for an arbitrary virtual viewpoint. We are going to study the relation between the camera arrangement and the quality of the rendered virtual views corresponding to the points on an arbitrary trajectory of free virtual navigation. Moreover, we aim at practical systems featured by limited investment and operational costs. This important practical assumption yields that very limited number of cameras may be used. Thus the cameras are sparsely distributed around a scene and the question is how to choose their locations maximizing the quality of the virtual views. Further in this paper we are going to deal with this problem.

For the purpose of the study, we assume that the video processing chain is implemented according to the state-of-the-art technology as described e.g. in [5]. In particular, we assume that the virtual view synthesis is implemented according the algorithms corresponding to the state-of-the-art view synthesis reference software VSRS used as reference by MPEG [15].

Reseach project was supported by The National Centre for Research and Development, Poland. Project no. TANGO1/266710/NCBR/2015

II. THE PROBLEM

In this paper, we study the optimization of the camera positions around a scene. The goal is to obtain the best synthesized-video quality assuming a constant number of cameras.

For the sake of simplicity, we assume that the cameras located around a scene are all placed on an arc. Nevertheless, we do not assume that the cameras are uniformly spaced on an arc around a scene. Contrary to the [1,2], where the experimental FTV and free navigation systems are described, we do not limit our considerations to a uniform distribution of the cameras on an arc. Even more, we ask a question if maybe a non-uniform distribution of cameras would be beneficiary for the quality of the synthesized virtual views.

We are going to answer the abovementioned question in the context of the assumptions discussed in the previous section of the paper. Here, we stress only that the number of cameras is limited, therefore, they are sparsely distributed on an arc.

For the purpose of the further considerations, the quality of a synthesized view will be measured with the reference to the view acquired by an additional reference camera located at the current virtual viewpoint. For the sake of simplicity, we will use the PSNR measure and we will skip the subjective tests. The respective subjective tests will be left for the future research on the topic.

III. PROBLEM ANALYSIS

In the free navigation and FTV systems with sparse camera arrangement, the depth-image-based rendering [16] is used for virtual synthesis purposes. Therefore, to analyze how the camera positions affect the quality of free navigation, the interdependence between the camera arrangement and the depth accuracy has to be examined.

Let us consider a camera pair (the focal length is f) with the base b. The depth z of a recorded object is [4]:

$$z = \frac{fb}{d} \quad , \tag{1}$$

where d is the disparity between object positions in two views. The minimum recordable depth difference $|z_1\text{-}z_2|$ may be estimated as

$$|z_1 - z_2| \geq \frac{z_1 z_2}{fb}\Delta \quad , \tag{2}$$

where z_1, z_2 are two depth values that correspond to the disparity Δ. In order to obtain fine depth resolution we need large values of the base b as the value of Δ has its minimum value implied both by the pixel pitch of the sensor, as well as the resolution of the lenses. Therefore, a system with sparse camera arrangement theoretically provides fine resolution of depth. Unfortunately, by increasing the base value b the areas of occlusion also grow up. In the occluded regions a point is visible from one camera only, so depth estimation is impossible. Therefore, the choice of the base b is the trade-off between depth resolution and occlusions. This relation will be studied experimentally further in the paper.

IV. EXPERIMENTS

The goal of the experiments is to check if a non-uniform distribution of cameras may yield better quality of the synthesized video as compared to the uniform distribution. The camera positions mostly affect the quality of the depth maps, and the quality of the depth maps significantly influences the quality of the synthesized views.

The experiment is organized as follows. We optimize the positions of 4 cameras located on an arc. The positions of the two outer cameras (Positions 0 and 30 in Fig. 1) are fixed in such a way that their convergent optical axes form an angle of 15 degrees between them. We vary the positions of the remaining two inner cameras in order to maximize the average quality of all the views synthesized at the locations within 15 degrees of arc with the step of 0.5 degree. We test the positions of the remaining two cameras at 10 positions within 15 degrees with the same step of 0.5 degree (Fig. 1). Therefore, we test 10 camera arrangements. In Arrangement A1 cameras are uniformly distributed, i.e. the inner cameras are at Positions 10 and 20. The remaining Arrangements 2-10 are formed by 2 camera pairs with variable base distance of each camera pair. In Arrangement 2, the inner cameras are at Positions 9 and 21. In Arrangement 3, they are at Positions 8 and 22, and so on. In A10, the inner cameras are at Positions 1 and 29 (Fig. 1).

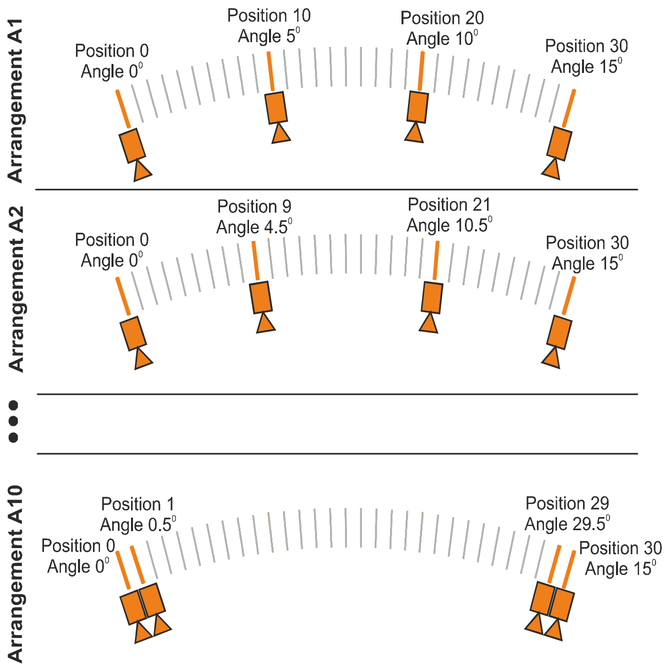

Figure 1. Camera positions in the experiment.

In that way, we test the uniform camera distribution with the cameras located at each 5 degrees of arc (Arrangement A1). We also test 9 non-uniform camera distributions with the cameras grouped into pairs, and the pairs are located on an arc each 15 degrees. In Arrangement A1, the angle between the axes of the 2 cameras of a pair is 4.5 degrees, in Arrangement A2, it is 4.0 degrees etc. In last Arrangement A10, angle between the axes of the 2 cameras of a pair is 0.5 degrees.

Obviously, the above described camera arrangements are not the only possible. Nevertheless, they constitute reasonable set arrangements with the parameter value being reasonable for free navigation, and for the already mentioned trade-off between the depth resolution limitations and occlusions.

For the purpose of measuring the quality of the synthetic views, we use the multiview test sequences acquired from 90 positions placed each 0.5 degrees on an arc [17]. For such sequences the reference views are provided at all positions of the synthetic views in the tests designed as described above. Unfortunately, there are very few available test sequences with circular camera arrangement and the number of views exceeding 31. Therefore, we additionally use also some test sequences with linear camera arrangements. Unfortunately, such available test sequences have various camera parameters and various bases. Therefore, the views for sequences with linear camera arrangement were chosen to ensure similar maximum disparity between the first and the last camera. For cameras located on a line, the camera arrangements are defined in the same way as for circular arrangements.

Syntheses of all virtual views from 1 to 29, were performed using depth maps and texture from only two marginal cameras (0 and 30) to ensure the usage of the same views for virtual view synthesis purposes for all tested arrangements. Thereby, we evaluated the impact of different camera arrangements on the quality of estimated depth maps (thus synthesis), not the influence of angle between views used for synthesis.

For all sequences, the depth estimation parameters were the same, with the exception of smoothing parameter, which had to be chosen individually for each sequence, depending on its depth dynamic range. At the end, the mean PSNR value for all synthesized virtual views was estimated.

For the tests, the sequences approved by MPEG for testing the compression algorithms for the free-viewpoint television and the free navigation were used [2]. The list of the sequences used, together with information about its source, used views numbers, and camera arrangement is featured in Table I.

TABLE I. TEST SEQUENCES.

Sequence name	**Sequence source**	**Used views**	**Camera arrangement**
BBB[a] Butterfly (1)	Holografika [17]	6,7... 36	Arc
BBB Flowers (1)	Holografika [17]	6,7... 36	Arc
BBB Rabbit (1)	Holografika [17]	6,7... 36	Arc
BBB Butterfly (2)	Holografika [17]	30,32... 90	Linear
BBB Flowers (2)	Holografika [17]	30,32... 90	Linear
BBB Rabbit (2)	Holografika [17]	0,3... 90	Linear
Bee	NICT [18]	20,23... 110	Linear
Champagne	Nagoya Univ.[19]	30,31... 60	Linear
Dog	Nagoya Univ.[19]	0,2... 60	Linear
Pantomime	Nagoya Univ.[19]	0,2... 60	Linear
San Miguel	Hasselt Univ. [20]	60,61... 90	Linear

a. BBB – Big Buck Bunny

V. EXPERIMENTAL RESULTS

A. The impact of camera arrangement on synthesis quality

The results of described experiments, performed for first frames of each sequence, were presented in Table II. For each test sequence the three highest values of PSNR were marked red. The highest value was additionally bolded. The mean quality of virtual view synthesis for 10 different camera arrangements is presented too.

The mean quality of the synthesized views noticeably differs for different sequences. For example, in the BBB Butterfly (1) sequence, the mean PSNR for uniform arrangement of cameras is bigger than 32dB. On the other hand, objective quality of BBB Flowers (1) and Champagne sequences is less than 20dB. The reason of low quality of view synthesis for both sequences result from very different characteristics of presented scene. Champagne sequence contains many translucent and shiny objects; in BBB Flowers the depth dynamic range is very high, which causes many objects to be occluded in neighboring views. It affects both the depth estimation and view synthesis quality.

TABLE II. SYNTHESIS QUALITY FOR INDIVIDUAL CAMERA ARRANGEMENTS.

Sequence name	**Average luminance PSNR for all synthetic views [dB]**									
	A10	***A9***	***A8***	***A7***	***A6***	***A5***	***A4***	***A3***	***A2***	***A1***
BBB Butterfly (1)	26.69	27.52	30.12	31.10	30.95	31.20	31.55	31.94	31.83	**32.15**
BBB Flowers (1)	20.03	21.03	**21.33**	21.23	20.95	20.33	20.55	19.46	19.33	19.11
BBB Rabbit (1)	26.81	26.78	26.93	27.10	27.44	27.74	27.90	28.09	28.28	**28.43**
BBB Butterfly (2)	22.51	25.71	27.68	27.92	28.36	28.81	29.28	29.40	**29.48**	29.31
BBB Flowers (2)	21.41	23.06	24.02	24.42	24.32	24.43	**24.45**	24.28	24.01	23.69
BBB Rabbit (2)	21.55	22.66	23.29	23.46	23.61	23.75	**24.31**	23.57	23.92	23.65
Bee	19.16	20.68	20.95	**21.15**	21.14	21.01	20.82	20.67	20.43	20.04
Champagne	17.43	18.91	18.85	**19.21**	18.71	19.02	18.40	17.80	18.14	17.65
Dog	24.44	24.58	25.01	25.48	25.52	25.67	**26.12**	25.60	25.37	25.77
Pantomime	19.85	22.13	24.17	26.03	26.53	26.35	26.58	26.41	**26.79**	26.74
San Miguel	23.74	24.47	24.74	**25.14**	24.83	25.09	24.83	24.40	24.39	24.10
Mean quality	**22.15**	**23.41**	**24.28**	**24.75**	**24.76**	**24.85**	**24.98**	**24.69**	**24.72**	**24.60**

The highest mean quality of the synthesized views was obtained for position A4 – Locations 0, 7, 23 and 30. Nevertheless, the difference of virtual view synthesis quality between A4 arrangement and uniform camera positioning (A1) is only 0.4dB.

Conducted tests suggest that (on average) the uniform camera positioning in multiview system provides mean quality of synthesis lower than in other arrangements. Obviously, this statement is not true for every multiview sequence (e.g. for BBB Butterfly (1) Arrangement A4 is 0.6dB worse than A1), so non-uniform camera positioning is not always the best solution. A gain from positioning cameras non-uniformly and scene characteristics will be discussed in the next section.

B. The impact of occlusions on a gain from non-uniform arrangement

Let us consider Arrangement A7 – the non-uniform arrangement with one of the highest quality of synthesis. For 3 of 11 sequences the quality gain over uniform Arrangement A1 was the highest for Arrangement A7. Moreover, for other two sequences that gain was one of the three highest. We observed that there is a correlation between these 5 sequences. All of them (Bee, Champagne, two versions of BBB Flowers and San Miguel) are characterized by many occlusions in synthesized virtual views. On the other hand, in the remaining sequences we noticed relatively less occluded areas.

To estimate an amount of occlusions, we decided to use the following algorithm:

1. calculate depth for uniform camera arrangement,
2. synthesize virtual view without inpainting operations,
3. count the points which were not synthesized or synthesized from only one camera (Fig. 2),
4. divide number of counted points by image size,
5. repeat steps 2 – 4 for every considered viewpoints,
6. divide the obtained result by a number of viewpoints.

Figure 2. Occlusions in examplary virtual view (white areas): left – non-synthesized regions (from any camera), right – partially synthesized regions (only from one camera).

The result of the described occlusions estimation process is an average percentage value of occluded areas for each sequence (see Table III). Through analysis of the acquired results a set of sequences was distinguished into two subsets according to the amount of occlusions – the sets of sequences with and without noticeable occlusions. We decided to set a threshold at the value of 20%, which corresponds to the subjective notion of presence of the occlusions in the sequence.

TABLE III. PERCENTAGE OF OCCLUDED AREAS.

Sequence name	Occluded area [%]
BBB Butterfly (1)	9.05
BBB Butterfly (2)	16.53
BBB Flowers (1)	38.68
BBB Flowers (2)	29.18
BBB Rabbit (1)	4.93
BBB Rabbit (2)	15.41
Bee	35.57
Champagne	32.55
Dog	9.81
Pantomime	15.61
San Miguel	29.21

Table IV gathers differences between each of the considered camera arrangement and uniform A1 for sequences with noticeable occlusions. The optimal arrangement of cameras for these sequences is A7. This arrangement is a good compromise between depth maps dynamic range (the number of possible disparities to choose) and amount of occluded areas (which is proportional to the base of camera pairs). Mean gain over uniform camera arrangement for sequences with noticeable occlusions for all tested arrangements is shown in Fig. 3.

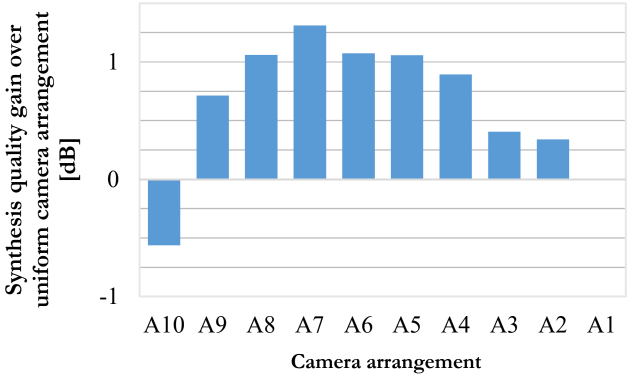

Figure 3. A gain over uniform camera arrangement for sequences with noticeable occlusions (> 20% of image area)

TABLE IV. A GAIN OVER UNIFORM CAMERA ARRANGEMENT FOR SEQUENCES WITH NOTICEABLE OCCLUSIONS.

Sequence name	Camera pairing gain [dB]									
	A10	*A9*	*A8*	*A7*	*A6*	*A5*	*A4*	*A3*	*A2*	*A1*
BBB Flowers (1)	0.93	1.93	**2.23**	2.12	1.85	1.23	1.44	0.35	0.23	0.00
BBB Flowers (2)	-2.28	-0.63	0.32	0.73	0.63	0.74	**0.76**	0.59	0.32	0.00
Bee	-0.88	0.64	0.91	**1.11**	1.10	0.96	0.78	0.63	0.39	0.00
Champagne	-0.22	1.25	1.20	**1.56**	1.06	1.36	0.75	0.15	0.48	0.00
San Miguel	-0.36	0.37	0.64	**1.04**	0.73	0.99	0.73	0.30	0.29	0.00

The results for the second subset of test sequences are presented in Table V. Inversely to Table IV, the maximum quality of synthesis for sequences without noticeable occlusions concentrates in the right side of the table, what corresponds to the uniform or almost uniform arrangements.

TABLE V. GAIN OVER UNIFORM CAMERA ARRANGEMENT FOR SEQUENCES WITHOUT NOTICEABLE OCCLUSIONS.

Sequence name	Camera pairing gain [dB]									
	A10	*A9*	*A8*	*A7*	*A6*	*A5*	*A4*	*A3*	*A2*	*A1*
BBB Butterfly (1)	-5.46	-4.64	-2.03	-1.05	-1.21	-0.95	-0.60	-0.21	-0.32	**0.00**
BBB Butterfly (2)	-6.79	-3.60	-1.63	-1.38	-0.94	-0.50	-0.02	0.09	**0.17**	0.00
BBB Rabbit (1)	-1.62	-1.65	-1.50	-1.33	-0.99	-0.69	-0.53	-0.34	-0.15	**0.00**
BBB Rabbit (2)	-2.10	-0.99	-0.36	-0.19	-0.04	0.09	**0.66**	-0.08	0.26	0.00
Dog	-1.33	-1.19	-0.76	-0.29	-0.25	-0.10	**0.34**	-0.17	-0.40	0.00
Pantomime	-6.90	-4.61	-2.57	-0.71	-0.22	-0.40	-0.17	-0.33	**0.05**	0.00

An average gain over uniform camera arrangement for sequences without noticeable occlusions is presented in Fig. 4. For every non-uniform position the gain was negative, what implies, that the best camera arrangement for sequences without noticeable occlusions is the uniform one.

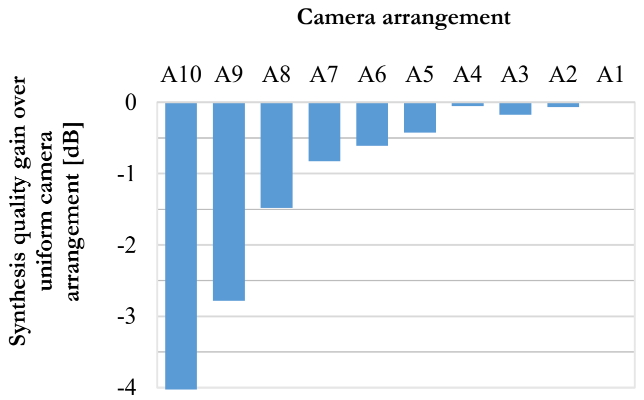

Figure 4. A gain over uniform camera arrangement for sequences without noticeable occlusions (< 20% of image area).

In Fig. 5 we presented a gain from non-uniform positioning of the cameras in the function of a percentage of occluded area. It clearly shows the correlation between both factors. Each dot represents one of the analyzed sequences. The dashed trend line was estimated to provide the lowest mean square error to 11 data value pairs.

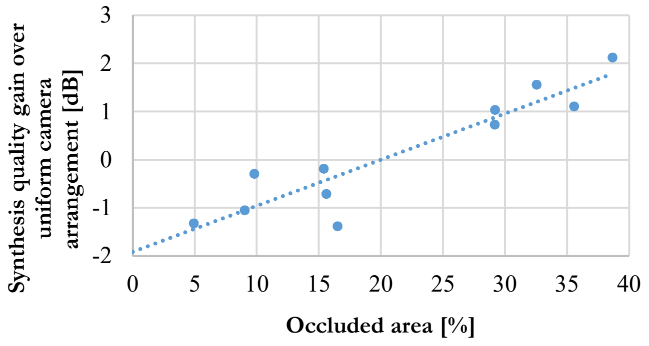

Figure 5. Correlation between quality gain over uniform camera arrangement and percentage of occluded area.

VI. Conclusions

In this paper, the problem of camera position optimization for free-navigation applications was described. We focused on simple, practical free-viewpoint television systems with a limited number of cameras located around the scene. The presented research was conducted to test the influence of physical camera arrangement on virtual view synthesis quality, omitting the impact of algorithmic modifications.

The paper presents experimental results that show the correlation between the scene characteristics and the gain of non-uniform camera positioning. As it was observed, FTV system arranged as a set of camera pairs with very small base provides virtual view synthesis of low quality, because of small number of disparities possible to choose. The highest quality of the synthesis quality for sequences with noticeable occlusions can be achieved for FTV system arrangement with stereo pairs with medium base. Surprisingly, the uniform cameras arrangement (the most common in available FTV systems) is optimal only for sequences with small percentage of occluded objects in neighboring cameras.

Sequences from different sources that varied in camera arrangement and content were used during performed tests. The influence of sequences camera setup (linear or circular), origin (synthetic or recorded sequence) on the quality gain from optimizing of camera positioning was not observed.

Presented results can be used to easily improve the quality of free navigation for future FTV sequences. As it was shown, optimal camera positions (when considering quality of synthesized virtual views) highly depend on scene characteristics. Therefore, for scenes with possible occlusions (e.g. a boxing ring) it is recommended for cameras to be arranged as a set of stereo pairs with medium base, as experiments have shown that this arrangement provided the best trade-off between spatial depth resolution and occlusions.

MPEG, a working group of ISO/IEC which develops standards for audio, video and related data coding, announced in 2015 so-called Call for Evidence on Free-Viewpoint Television [21]. It is a search for technology which significantly outperforms current MPEG state-of-the-art technology for free navigation and super-multiview video. The results of this call [22] did not show sufficient increase in free navigation quality. However, authors submitted new test sequences [23] which were the result of the research on camera positions optimization. These sequences were described as first which enable to create a satisfying free navigation experience for a viewer [24].

References

[1] M. Tanimoto, M.P. Tehrani, T. Fujii, T. Yendo, “FTV for 3-D spatial communication”, Proceeding of the IEEE, vol. 100, no. 4, pp. 905-917, 2012.

[2] M. Tanimoto, T. Senoh, S. Naito, S. Shimizu, H. Horimai, M. Domański, A. Vetro, M. Preda, K. Mueller, Proposal on a new activity for the third phase of FTV, ISO/IEC JTC 1/SC 29/WG 11, Doc. M30229/M30232, Vienna, Austria, July/Aug. 2013.

[3] K. Kawamura, H. Sankoh, S. Naito, “Requirements and use cases of free-navigation for fly-through experience,” ISO/IEC JTC1/SC29/WG11, Doc. M31052, Geneva, CH, Oct. 2013.

[4] R. Hartley, A. Zisserman, “Multiple view geometry in computer vision” (2nd ed.), Cambridge Univ. Press, 2015.

[5] M. Domański, A. Dziembowski, D. Mieloch, A. Łuczak, O. Stankiewicz, K. Wegner, „A practical approach to acquisition and processing of free viewpoint video,” Picture Coding Symposium PCS, Cairns, pp. 10-14, 2015.

[6] K. Müller, P. Merkle, T. Wiegand, “3D Video Representation Using Depth Maps”, Proceedings of the IEEE, vol. 99, no. 4, pp. 643–656, Apr. 2011.

[7] G. Miller, J. Starck, A. Hilton, “Projective surface refinement for free-viewpoint video,” 3rd European Conf. Visual Media Production, CVMP, pp.153-162, 2006.

[8] A. Smolic, K. Müller, P. Merkle, M. Kautzner, T. Wiegand, “3D video objects for interactive applications.” 13th European Signal Processing Conference EUSIPCO, 2005.

[9] M. Tanimoto, “Overview of free viewpoint television”, Signal Processing: Image Communication, vol. 21, no. 6, pp. 454-461, 2006.

[10] O. Stankiewicz, Stereoscopic depth map estimation and coding techniques for multiview video system, PhD dissertation, Poznań University of Technology, also available at: theses.eurasip.org.

[11] O. Stankiewicz, K. Wegner, M. Tanimoto, M. Domański, “Enhanced Depth Estimation Reference Software (DERS) for Free-viewpoint Television”, ISO/IEC JTC1/SC29/WG11 Doc. MPEG M31518, Geneva, CH, Oct.2013.

[12] V. Kolmogorov, R. Zabih, “What energy functions can be minimized via graph cuts?”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 26, no. 2, pp. 147-159, 2004.

[13] M. Domański. M. Bartkowiak, A. Dziembowski, T. Grajek, A. Grzelka, A. Łuczak, D. Mieloch, J. Samelak, O. Stankiewicz, J. Stankowski, K. Wegner, “New Results in Free-viewpoint Television Systems for Horizontal Virtual Navigation”, to be published in IEEE International Conference on Multimedia and Expo 2016 (ICME),Seattle 2016.

[14] G. Lafruit, M. Domański, K. Wegner, T. Grajek, T. Senoh, J. Jung, P. .Kovács, P. Goorts, L. Jorissen, A. Munteanu, B. Ceulemans. P. Carballeira. S. García, M. Tanimoto, “New visual coding exploration in MPEG: Super-MultiView and Free Navigation in Free viewpoint TV”, IST Electronic Imaging, San Francisco, 2016.

[15] O. Stankiewicz, K. Wegner, M. Tanimoto, M. Domański, “Enhanced view synthesis reference software (VSRS) for Free-viewpoint Television”, ISO/IEC JTC 1/SC 29/WG 11, Doc. M31520, Geneva, CH Oct. 2013.

[16] S. Zinger, Q. Do, P. de With, “Free-viewpoint depth image based rendering”, Journal Visual Communication and Image Representation, 21, pp. 533–541, 2010.

[17] P. Kovacs, “[FTV AHG] Big Buck Bunny light-field test sequences”. ISO/IEC JTC1/SC29/WG11, Doc. M35721, Geneva, 2015.

[18] T. Senoh, A. Ishikawa, M. Okui, K. Yamamoto, N. Inoue, "FTV AHG: EE1 and EE2 Results with Bee by NICT", IEC JTC1/SC29/WG11 Doc. MPEG2013/ M32995, Valencia, Spain, April 2014.

[19] M. Tanimoto, T. Fujii, N. Fukushima, „1D Parallel Test Sequences for MPEG-FTV”, ISO/IEC JTC1/SC29/WG11, Doc. M15378, Archamps, 2008.

[20] P. Goorts, M. Javadi, S. Rogmans, G. Lafruit, “San miguel test images with depth ground truth”, ISO/IEC JTC1/SC29/WG11, Doc. M33163, Valencia, 2014.

[21] “Call for Evidence on Free-Viewpoint Television: Super-Multiview and Free Navigation”, ISO/IEC JTC1/SC29/WG11 Doc. N15348, Warsaw, 2015.

[22] “Summary of the results of the Call for Evidence on Free-Viewpoint Television: Super-Multiview and Free Navigation” , ISO/IEC JTC1/SC29/WG11, Doc. W16318, Geneva, 2016

[23] M. Domański, A. Dziembowski, A. Grzelka, D. Mieloch, O. Stankiewicz, K. Wegner, “Multiview test video sequences for free navigation exploration obtained using pairs of cameras”, ISO/IEC JTC1/SC29/WG11, Doc. M38247, Geneva, 2016

[24] “Resolutions of the 115th Meeting” , ISO/IEC JTC1/SC29/WG11, Doc. N16152, p. 17, Geneva, 2016