INTERNATIONAL ORGANISATION FOR STANDARDISATION
ORGANISATION INTERNATIONALE DE NORMALISATION
ISO/IEC JTC1/SC29/WG11
CODING OF MOVING PICTURES AND AUDIO

ISO/IEC JTC1/SC29/WG11
MPEG2016/ m38248
May 2016, Geneva, Switzerland

Source: **Poznań University of Technology**

Status: **Contribution to the 115th MPEG Meeting, FTV AhG**

Title: **Study on nouniform distributions of cameras located on an arc**

Authors: **Marek Domański (domanski@et.put.poznan.pl),**
Adrian Dziembowski (adziembowski@multimedia.edu.pl)
Adam Grzelka (agrzelka@multimedia.edu.pl),
Dawid Mieloch(dmieloch@multimedia.edu.pl),

1. Introduction

In this contribution, we summarize the conclusions resulting from our research on nonuniform distribution of the cameras around a scene during multi-camera video acquisition for free-viewpoint television and free navigation. Somewhat wider description may be found in the paper [1]. Some experiments have been also reported in [8] showing the advantages of video acquisition using the pairs and triples of cameras.

Assume that we are going to install a fixed number of cameras around a scene. Our goal is to obtain the most accurate depth maps as possible. The question arises, if we should distribute the cameras uniformly around a scene, or maybe better results could be obtained when the cameras are nonuniformly distributed around a scene.

For the sake of simplicity, we assume that the cameras located around a scene are all placed on an arc. For the purpose of the further considerations, the quality be measured by the quality of the synthesized views, and 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.

For the experiments the modified versions of DERS [2] and VSRS [3] will be used for depth estimation and view synthesis, respectively.

2. Brief theoretical considerations

Consider a camera pair with the base b. The depth z of a recorded object is:

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

where f is the focal length, and 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.

3. Design of the experiment

The goal of the experiments is to compare a non-uniform distribution of cameras with the uniform distribution in the context of the quality of the synthesized video. The chosen 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.

We assume that even nonuniform camera distribution is periodical. For the sake of simplicity, we assume that a period consists of 3 cameras. Therefore, 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).

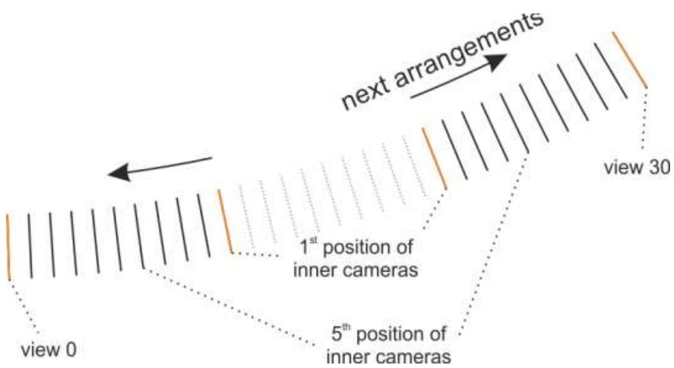

Figure 1. Camera positions in the experiment (the angles between possible locations are quantized with quantization step of 0.5 degrees).

In that way, we test the uniform camera distribution with the cameras located at each 5 degrees of arc (Arrangement 1). We also test 9 non-uniform camera distributions with the cameras grouped into pairs, and the pairs are located on an arc each 15 degress. In Arrangement 1, the angle between the axes of the 2 cameras of a pair is 4.5 degrees, in Arrangement 2, it is 4.0 degrees etc. In last Arrangement 10, 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 a reasonable set of 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 [4]. 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 virtual views 1 to 29, were performed using depth maps and texture from only two marginal cameras 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. 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 [4]	6,7... 36	Arc
BBB Flowers (1)	Holografika [4]	6,7... 36	Arc
BBB Rabbit (1)	Holografika [4]	6,7... 36	Arc
BBB Butterfly (2)	Holografika [4]	30,32... 90	Linear
BBB Flowers (2)	Holografika [4]	30,32... 90	Linear
BBB Rabbit (2)	Holografika [4]	0,3... 90	Linear
Bee	NICT [5]	20,23... 110	Linear
Champagne	Nagoya Univ.[6]	30,31... 60	Linear
Dog	Nagoya Univ.[6]	0,2... 60	Linear
Pantomime	Nagoya Univ.[6]	0,2... 60	Linear
San Miguel	Hasselt Univ. [7]	60,61... 90	Linear

a. BBB – Big Buck Bunny

3. Experimental results

3.1. The impact of camera arrangement on synthesis quality

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

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 [dB]:	**22.15**	**23.41**	**24.28**	**24.75**	**24.76**	**24.85**	**24.98**	**24.69**	**24.72**	**24.60**

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 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.

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

The 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. However, presented data show a correlation between a gain from positioning cameras non-uniformly and scene characteristics, which will be discussed in the next section.

Taking into account the amount of occlusions in the individual test sequences, the following conclusion may be drawn from the abovementioned results:
- the uniform camera arrangement is close to optimum for low amounts of occlusions,
- for the content with many occlusions, it is better to locate the cameras in close pairs.

3.2. The gain resulting from camera pairing for highly occluded content

Table III 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 seems to be 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. 2.

TABLE III. 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

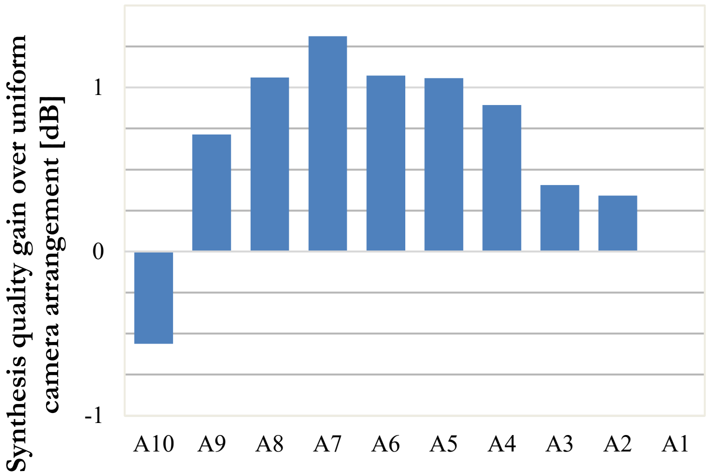

Figure 2. A gain over uniform camera arrangement for sequences with noticeable occlusions (> 20% of image area)

In Fig. 3 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.

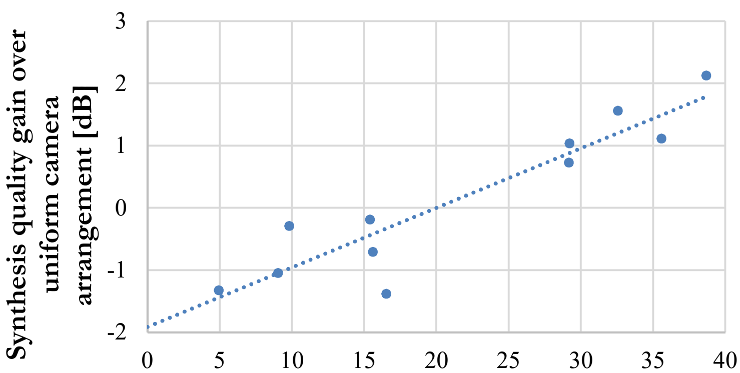

Figure 3. Correlation between quality gain over uniform camera arrangement and percentage of occluded area

4. 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 camera pairs with small base provides virtual view synthesis of low quality, because of small number of disparities possible to choose. On the other hand, the uniform cameras arrangement is optimal only for sequences with small percentage of occluded objects in neighboring cameras. 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.

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 gain from different camera positioning was not observed.

These results have encouraged the authors to produce natural video test sequences acquired by pairs of cameras [9].

ACKNOWLEDGEMENT

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

REFERENCES

[1] M. Domański, A. Dziembowski, A. Grzelka, D. Mieloch, Optimization of camera positionsfor free-navigation applications, to be published.

[2] 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.

[3] 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. 2013.21, pp. 533–541, 2010.

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

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

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

[7] 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.

[8] 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.

[9] 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. MPEG2016/ M38247, Geneva, Switzerland, May 2016.