INTL JOURNAL OF ELECTRONICS AND TELECOMMUNICATIONS, 2018, VOL. 64, NO. 3, PP. 293–298
Manuscript received May 25, 2018; revised July, 2018. DOI: 10.24425/123521

Segmentation-based Method of Increasing The Depth Maps Temporal Consistency

Dawid Mieloch and Adam Grzelka

***Abstract*—In this paper, a modification of the graph-based depth estimation is presented. The purpose of proposed modification is to increase the quality of estimated depth maps, reduce the time of the estimation, and increase the temporal consistency of depth maps. The modification is based on the image segmentation using superpixels, therefore in the first step of the proposed modification a segmentation of previous frames is used in the currently processed frame in order to reduce the overall time of the depth estimation. In the next step, a depth map from the previous frame is used in the depth map optimization as the initial values of a depth map estimated for the current frame. It results in the better representation of silhouettes of objects in depth maps and in the reduced computational complexity of the depth estimation process. In order to evaluate the performance of the proposed modification the authors performed the experiment for a set of multiview test sequences that varied in their content and an arrangement of cameras. The results of the experiments confirmed the increase of the depth maps quality — the quality of depth maps calculated with the proposed modification is higher than for the unmodified depth estimation method, apart from the number of the performed optimization cycles. Therefore, use of the proposed modification allows to estimate a depth of the better quality with almost 40% reduction of the estimation time. Moreover, the temporal consistency, measured through the reduction of the bitrate of encoded virtual views, was also considerably increased.**

***Keywords*—depth maps, depth map estimation, temporal consistency, image segmentation, free-viewpoint television, virtual view synthesis**

I. Introduction

DEPTH maps are one of 3D scene representations [1] and are widely used in the free-viewpoint television systems [2], [3], [4], [5] for the virtual view synthesis purposes [6]), in a 3D scene modeling [7] and machine vision applications [8], [9]. In this paper we focus on a software depth estimation based on a stereoscopic correspondence. This type of the depth estimation is characterized by the high computational complexity [10], [11]. The recent introduction of new virtual reality systems and multiview camera systems such as lightfields shows the increased number of views and the increased resolution of used cameras in multiview systems [12]. Similar trend can be seen in free-viewpoint television systems [13], therefore, further increase of the depth estimation computational complexity is expected.

In order to acquire a depth of the scene the depth sensors can be used [14]. Many depth sensors can acquire a depth of the objects of the scene in the real time, however, the limitations of depth sensors, resulting from interferences from infrared illumination sources or even from other depth sensors [15], significantly reduce the usability of depth cameras.

In order to reduce the time of computations in the depth estimation process, the authors propose the modification of the graph-based estimation of depth maps. The proposed method is based on the utilization of the temporal information during the depth estimation and reduces the time of estimation, increases the temporal consistency of depth maps, and simultaneously increases the quality of estimated depth maps. The proposal was already described in [16], here we present it together with more comprehensive results, with the emphasis on the experimental testing of the temporal consistency.

The proposed modification was tested in the depth estimation framework [17] — the depth estimation that is based on the optimization of depth maps using the graph cut method [18], [19]. The method [17] is based on the superpixel segmentation [20] and provides the possibility of controlling the trade off between the quality of depth maps and the time of the estimation, with simultaneous preservation of the original resolution of depth maps. That depth estimation method was designed to work with any multiview system with arbitrarily located cameras.

II. Related Works

Methods for the increasing the temporal consistency of depth maps usually introduce an additional step during the depth estimation process. The method [21] performs an additional preprocessing in order to remove a noise from input views, what was shown to increase temporal consistency of estimated depth maps, while method [22] utilizes a preprocessing of depth maps with the temporal filter in order to smoothen depth maps. These methods increase the quality and the temporal consistency of depth maps, nevertheless, the overall processing time of the depth map estimation process is also increased. On the other hand, abovementioned methods, because of being performed independently from the main depth estimation process, can be used with all available depth estimation methods and do not state any assumptions about the number, the arrangement and type of used cameras.

Unfortunately, other methods are often adapted to work only for specific type of the depth estimation. For example, the method [23], based on a spatio temporal video segmentation, requires to be used with static scenes only. The other method [24] assumes that a depth sensor is used to estimate the depth maps. The method [25] allows to estimate a depth in a time close to the real time, but can be only used for a stereo-pair.

Use of motion vectors in temporal enhancement methods [26] can simultaneously increase the temporal consistency of depth maps and shorten the overall time of the estimation. Nevertheless, fast motion vector estimation methods are usually based on the block matching, therefore, in case of the fast

The presented work has been funded by the Polish Ministry of Science and Higher Education within the status activity task "Theory and algorithms of multidimensional signal processing" (DS) in 2018.

The authors are with the Chair of Multimedia Telecommunications and Microelectronics, Poznań University of Technology, Poznań, Poland (email: dawid.mieloch@put.poznan.pl; adam.grzelka@put.poznan.pl).

motion of objects it can fail to find a true motion of objects, what can decrease the reliability of depth maps estimated on the basis of such motion vectors.

Another type of methods require a modification of the depth maps estimation by adding of temporal information to the optimization step of the depth map estimation [27]. Such modification significantly increases the quality of estimated depth maps, but the processing time of the estimation is also increased, even by 25%. The method [28] also estimates a depth maps with the temporal consistency enforced during the optimization, but can be used only for a moving camera rig.

III. Proposed Method

The proposed modification is implemented as a part of the aforementioned method of the depth estimation [17]. The modification includes the utilization of the temporal information both in the used superpixel segmentation method [20] and in the depth map optimization process.

In the used segmentation method [20], the iterative process of segmentation is initialized with the user-defined number of square segments. In following iterations the segments change their shape on the basis of the spatial and color distance of points in the neighborhood of each segment. The process is stopped when none change in points affiliation is made.

In proposal we modified the segmentation method in order to use the initial segmentation obtained from the previous frame. It decreases the number of required iterations, therefore, the overall time of computations in the depth estimation is decreased. Moreover, the segmentation becomes less prone to a noise, so the quality of the segmentation is also increased. It is especially important for the estimation of depth maps, because the representation of objects silhouettes in depth maps has significant influence on the quality of virtual views in free-viewpoint television systems [29].

In second part of the proposed modification, the segmentation is used in order to define which regions of a depth map estimated in the previous frame can be used as initial depth for the actually processed depth map. The region represented by the segment s is marked as unchanged in comparison with the previous frame, if the mean luminance $Y(s, f)$ of a segment s in the current frame f did not significantly change in comparison with the luminance $Y(s', f-1)$ of the collocated segment s' in the previous frame. For segments that were marked as unchanged (in comparison with the previous frame) the initial depth $d(s, f)$ is equal to the depth from the previous frame:

$$d(s,f) = \begin{cases} d(s', f-1) & \text{if } |Y(s', f-1) - Y(s,f)| \leq Y_t, \\ 0 & \text{if } |Y(s', f-1) - Y(s,f)| > Y_t, \end{cases}$$

where Y_t is the threshold of the luminance difference and was set to 20. The presented initialization of the depth estimation increases the temporal consistency of depth maps and simultaneously decreases the time of depth optimization.

IV. Experimental Results

The presented modification of the graph-based depth estimation method improves both the quality and the temporal consistency of estimated depth maps. Following Sections IV-A and IV-B present an overview of the experiments. The respective results are presented in Section IV-C.

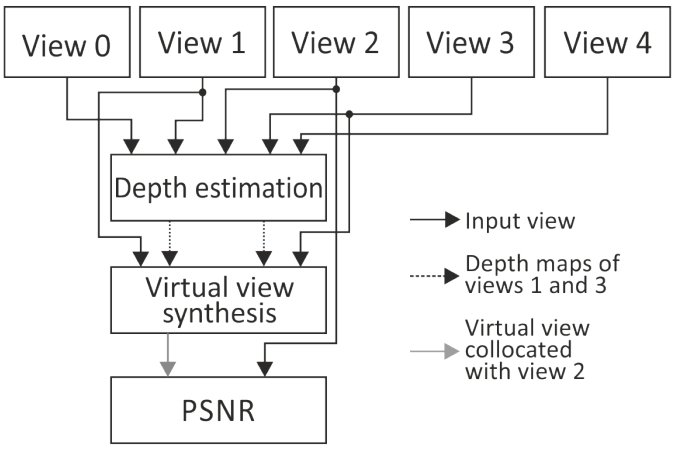

Fig. 1. The scheme of the PSNR calculation for the virtual view synthesized using depth maps estimated in the experiment.

TABLE I
Test Sequences Used in Experiments

Test sequence	Resolution	Used views	Sequence source
Ballet Breakdancers	1024×768	0 to 7	Microsoft Research [31]
BBB Butterfly BBB Rabbit	1024×768	6, 12, 19, 26, 32, 38, 45, 52	Holografika [32]
Poznan Blocks Poznan Blocks2 Poznan Fencing2 Poznan Service2	1920×1080	0 to 7	Poznań University of Technology [33] [34]

A. Assessment of the Quality of Depth Maps

The quality of depth maps is measured with the use of the virtual view synthesis (Fig 1). The depth estimation is performed for 5 views of a multiview test sequence. After the depth estimation, Views 1 and 3 and their depth maps are used to synthesize a virtual view placed in the same position as real View 2. At the end, we measure the PSNR between real View 2 and the collocated virtual view. The virtual view synthesis is performed using the VSRS method (View Synthesis Reference Software [30]). Used test sequences are presented in Table I.

A change in the quality of estimated depth maps influences the quality of synthesized virtual views. Therefore, the presented scheme of measurement of depth maps quality is a good determinant of the performance of the depth estimation method.

The depth estimation is performed for the unmodified method [35] and for the estimation with the proposed modification. Used depth estimation is based on the graph cut optimization, therefore the result of the estimation is dependent on the number of performed optimization cycles. Therefore, the estimation is done for 1, 2 and 3 cycles. The configuration of the depth estimation software is as follows: the estimation of 50 depth maps for 50 frames, 5 input views, 10 000 segments for each view, the correspondence matching performed in 3×3 blocks, the estimation for 250 levels of a depth and the smoothing coefficient equal to 1.

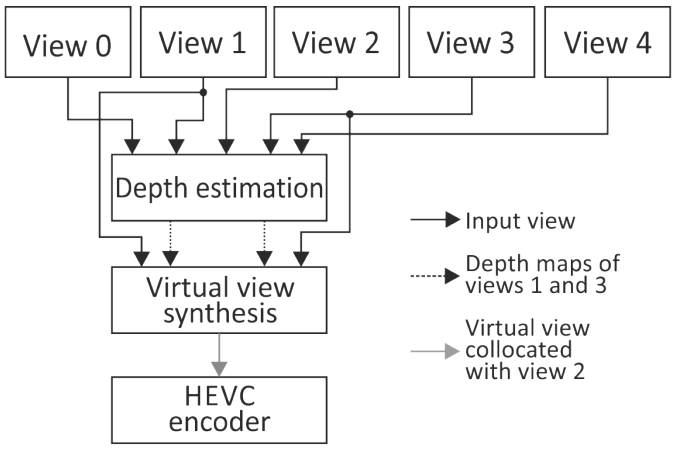

Fig. 2. The scheme of calculating PSNR of the virtual view synthesized using depth maps estimated in the experiment.

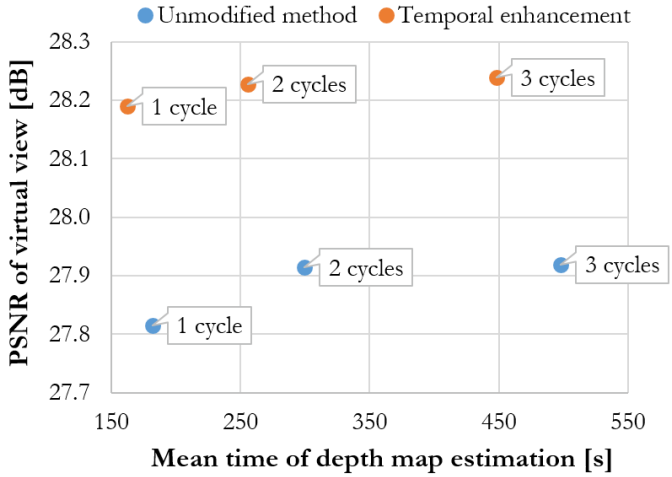

Fig. 3. Comparison of the mean quality of virtual views and the mean time of depth estimation for unmodified and proposed method.

B. Assessment of the Temporal Consistency of Depth Maps

The temporal consistency of depth maps can be measured indirectly, e.g. through the encoding of the depth [21]. Video sequences that are more consistent increase the performance of the inter-frame prediction in an encoder, what results in the decreased bitrate of the encoded sequence, while maintaining the same quality. In this paper, in order to ensure the continuity of performed experiments, instead of the encoding of estimated depth maps, the virtual views that were synthesized using scheme presented in Section IV-A are encoded (Fig 2). The improvement of the temporal consistency was expressed using the Bjontegaard metric [36].

Virtual views are compressed using the HEVC technique (the HM 16.15 framework [37]) The encoder is set in the low delay mode, therefore only the first frame of a sequence is an intra encoded frame. MPEG Common Test Conditions and reference software configurations (both available in [37]) are used.

C. Results

The quality of virtual views and the mean time of the depth estimation for the unmodified method and for the method with the proposed modification, averaged for all test sequences, are presented in Fig. 3.

As it can be seen, use of more than 2 optimization cycles does not change the quality of estimated depth maps. However, when the proposed modification is used, higher quality of depth maps is achieved, regardless of the number of performed optimization cycles. Moreover, the time of the depth estimation is shortened. Table II presents the quality of virtual views for individual test sequences. For all sequences the quality of virtual views is increased when the proposed modification is used.

In Table III the average luma bitrate reductions of encoded virtual views are presented. Bitrate reductions are calculated in comparison to virtual views synthesized using depth maps for unmodified method and 1 optimization cycle

Using more than 1 optimization cycle provides only minor bitrate reduction of 2%. On the other hand, when the proposed modification is used in the depth estimation process, the average bitrate is reduced by 13%, showing the significant

TABLE II
COMPARISON OF THE QUALITY OF VIRTUAL VIEWS SYNTHESIZED USING DEPTH MAPS ESTIMATED WITH UNMODIFIED METHOD AND WITH THE PROPOSED MODIFICATION

Used method of depth estimation		Unmodified method [17]			Proposed modification		
Number of optimization cycles		1	2	3	1	2	3
		PSNR of virtual view [dB]					
Test sequence	Ballet	26.58	26.76	26.77	27.65	27.67	27.74
	Breakdancers	32.02	32.11	32.13	32.08	32.18	32.18
	BBB Butterfly	29.11	29.26	29.27	29.37	29.61	29.60
	BBB Rabbit	23.81	23.89	23.90	23.95	23.93	23.95
	Poznan Blocks	23.15	23.24	23.24	23.38	23.37	23.36
	Poznan Blocks2	29.03	29.18	29.20	29.60	29.63	29.65
	Poznan Fencing2	29.79	29.79	29.77	29.89	29.91	29.90
	Poznan Service2	26.33	26.40	26.41	26.52	26.60	26.60
Mean quality of a virtual view [dB]		27.48	27.58	27.59	27.81	27.86	27.87
Mean time of depth estimation [s]		175	290	481	155	249	438

improvement of the temporal consistency of estimated depth maps. The results of the encoding for individual sequences (that include bitrates and PSNR values for all QPs) are presented in Table IV (for the unmodified method) and in Table V (for the method with the proposed modification)

The visual comparison of fragments of synthesized views for 3 consecutive frames is presented in Fig. 4 (Poznan Fencing2 sequence reduction of bitrate around 30%) and Fig. 5 (Poznan Blocks2 one of smallest reductions of bitrate, around 5%). For both sequences the comparison shows both the better temporal consistency of virtual views and the higher similarity of the synthesized virtual view to the collocated reference real view.

V. CONCLUSIONS

In this paper, the segmentation-based modification of the depth estimation was presented. The proposed modification consists of two main parts. The first one is the initialization of a segmentation using the segmentation from previous frames

TABLE III
AVERAGE LUMA BITRATE REDUCTIONS OF ENCODED VIRTUAL VIEWS SYNTHESIZED USING DEPTH MAPS ESTIMATED WITH UNMODIFIED METHOD AND WITH THE PROPOSED MODIFICATION

Method of depth estimation		Unmodified method [17]		Proposed modification		
Number of optimization cycles		2	3	1	2	3
		Encoded virtual views bitrate reductions with respect to encoded virtual views synthesized using depth maps calculated in one cycle with unmodified method				
Test sequence	Ballet	-1.1%	-0.8%	-13.2%	-17.4%	-17.9%
	Breakdancers	-3.8%	-3.4%	-14.1%	-13.3%	-13.2%
	BBB Butterfly	-4.9%	-4.9%	-14.6%	-11.4%	-12.2%
	BBB Rabbit	-1.7%	-1.9%	-4.3%	-4.7%	-4.7%
	Poznan Blocks	-1.5%	-2.2%	-12.5%	-14.7%	-14.9%
	Poznan Blocks2	-0.1%	0.6%	-6.2%	-5.5%	-4.6%
	Poznan Fencing2	-1.9%	-2.1%	-29.6%	-30.4%	-30.6%
	Poznan Service2	-0.7%	-0.2%	-4.2%	-6.2%	-6.1%
Average:		-2.0%	-1.9%	-12.3%	-13.0%	-13.0%

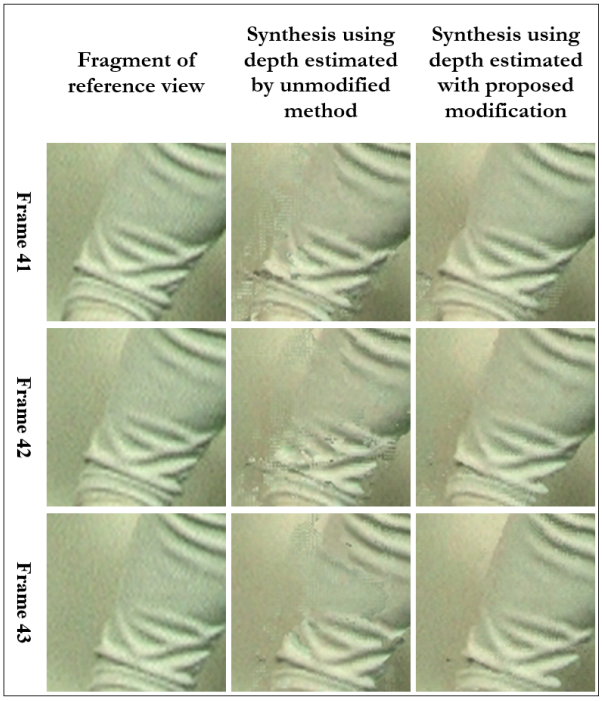

Fig. 4. Comparison of the virtual view synthesis for the sequence Poznan Fencing2.

that improves the representation of objects edges in estimated depth maps and decreases the overall time of the depth estimation. Further, the segmentation is also in the initialization of the depth optimization, where depth values from a previous frame are used as initial values of the currently estimated depth map.

The experiments demonstrate that the use of the proposed modification increases the quality of estimated depth maps and shorten the estimation time. The depth maps of the better quality than for the unmodified method can be estimated in 45% shorter time.

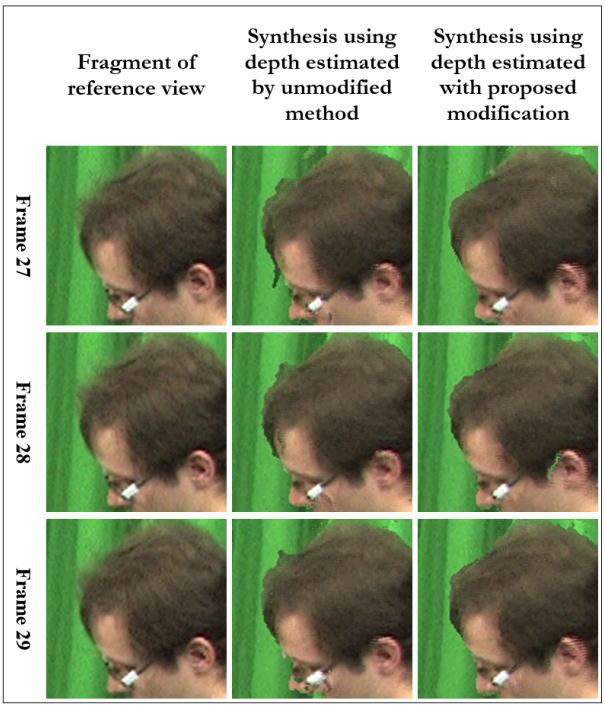

Fig. 5. Comparison of the virtual view synthesis for the sequence Poznan Blocks2.

The temporal consistency, crucial for depth maps used in the virtual view synthesis process in a free viewpoint television, was also tested. The direct estimation of the temporal consistency of depth maps is difficult. In presented paper, the temporal consistency was measured indirectly, through the compression of the virtual views synthesized using tested depth maps. As experiments show, when the proposed modification is used, the bitrate of compressed virtual views is even 13% smaller than for virtual views synthesized using depth maps estimated with the unmodified method. Such reduction of bitrate indicates significant increase in the temporal consistency of depth maps.

REFERENCES

[1] K. Muller, P. Merkle, and T. Wiegand, "3-D video representation using depth maps," *Proceedings of the IEEE*, vol. 99, no. 4, pp. 643–656, April 2011.

[2] O. Stankiewicz, M. Domański, A. Dziembowski, A. Grzelka, D. Mieloch, and J. Samelak, "A free-viewpoint television system for horizontal virtual navigation," *IEEE Transactions on Multimedia*, pp. 1–1, 2018.

[3] M. Tanimoto, "FTV standardization in MPEG," in *2014 3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON)*, July 2014, pp. 1–4.

[4] M. Tanimoto, M. P. Tehrani, T. Fujii, and T. Yendo, "FTV for 3-D spatial communication," *Proceedings of the IEEE*, vol. 100, no. 4, pp. 905–917, April 2012.

[5] M. Domański, A. Dziembowski, T. Grajek, A. Grzelka, K. Klimaszewski, D. Mieloch, R. Ratajczak, O. Stankiewicz, J. Siast, J. Stankowski, and K. Wegner, "Demonstration of a simple free viewpoint television system," in *2017 IEEE International Conference on Image Processing (ICIP)*, Sept 2017, pp. 4589–4591.

[6] A. Dziembowski, A. Grzelka, D. Mieloch, O. Stankiewicz, K. Wegner, and M. Domański, "Multiview synthesis - improved view synthesis for virtual navigation," in *2016 Picture Coding Symposium (PCS)*, Dec 2016, pp. 1–5.

[7] M. Camplani, T. Mantecon, and L. Salgado, "Depth-color fusion strategy for 3-D scene modeling with Kinect," *IEEE Transactions on Cybernetics*, vol. 43, no. 6, pp. 1560–1571, Dec 2013.

TABLE IV
THE BITRATE AND QUALITY OF ENCODED VIRTUAL VIEWS SYNTHESIZED USING DEPTH MAPS ESTIMATED WITH THE UNMODIFIED METHOD [17]

Test sequence	QP	Bitrate [Mbps]			PSNR [dB]		
		Number of optimization cycles					
		1	2	3	1	2	3
Ballet	22	7.6	7.7	7.7	40.7	40.7	40.7
	27	3.2	3.2	3.2	38.2	38.2	38.2
	32	1.4	1.4	1.4	35.6	35.6	35.6
	37	0.6	0.6	0.6	33.3	33.4	33.4
Break-dancers	22	45.1	44.6	44.7	41.4	41.5	41.5
	27	20.5	20.5	20.5	37.2	37.3	37.3
	32	8.9	8.6	8.7	33.8	33.9	33.9
	37	3.6	3.4	3.4	31.3	31.4	31.4
BBB Butterfly	22	21.4	22.0	22.0	40.7	40.6	40.6
	27	6.5	6.8	6.8	38.3	38.1	38.1
	32	2.2	2.3	2.3	36.3	36.1	36.2
	37	0.8	0.8	0.8	34.7	34.5	34.5
BBB Rabbit	22	7.7	7.6	7.6	40.7	40.7	40.7
	27	2.5	2.5	2.5	38.6	38.6	38.6
	32	1.1	1.1	1.0	36.9	36.9	36.9
	37	0.5	0.5	0.5	35.2	35.2	35.2
Poznan Blocks	22	9.6	9.6	10.3	44.3	44.3	44.0
	27	4.5	4.4	4.7	40.5	40.5	40.3
	32	2.1	2.0	2.2	37.3	37.3	37.1
	37	0.9	0.9	1.0	34.7	34.8	34.6
Poznan Blocks2	22	27.6	26.9	27.0	41.0	41.0	41.0
	27	11.6	11.3	11.4	38.2	38.2	38.2
	32	4.5	4.5	4.5	35.6	35.5	35.5
	37	1.6	1.6	1.6	33.5	33.4	33.4
Poznan Fencing2	22	41.3	40.9	41.0	39.5	39.5	39.6
	27	19.0	18.8	18.8	34.5	34.5	34.5
	32	7.2	7.1	7.1	30.7	30.7	30.8
	37	2.4	2.3	2.3	28.3	28.3	28.3
Poznan Service2	22	28.8	28.6	28.7	41.2	41.2	41.2
	27	12.2	12.2	12.2	38.0	38.0	38.0
	32	5.0	4.9	5.0	35.1	35.2	35.2
	37	1.8	1.8	1.8	32.8	32.8	32.8

TABLE V
THE BITRATE AND QUALITY OF ENCODED VIRTUAL VIEWS SYNTHESIZED USING DEPTH MAPS ESTIMATED WITH THE PROPOSED MODIFICATION

Test sequence	QP	Bitrate [Mbps]			PSNR [dB]		
		Number of optimization cycles					
		1	2	3	1	2	3
Ballet	22	7.2	7.1	7.1	40.8	40.8	40.8
	27	3.0	2.9	2.9	38.3	38.3	38.4
	32	1.3	1.2	1.2	35.8	35.9	35.9
	37	0.5	0.5	0.5	33.6	33.7	33.7
Break-dancers	22	40.5	40.8	40.8	41.5	41.5	41.5
	27	18.6	18.7	18.8	37.4	37.4	37.4
	32	7.8	7.9	7.9	34.0	34.0	34.0
	37	3.0	3.0	3.0	31.5	31.4	31.4
BBB Butterfly	22	22.3	21.4	21.3	40.6	40.7	40.7
	27	7.0	6.5	6.5	38.2	38.2	38.2
	32	2.5	2.2	2.2	36.1	36.2	36.2
	37	0.9	0.7	0.7	34.5	34.6	34.6
BBB Rabbit	22	7.5	7.5	7.5	40.7	40.7	40.7
	27	2.4	2.4	2.4	38.7	38.7	38.7
	32	1.0	1.0	1.0	36.9	36.9	36.9
	37	0.5	0.5	0.5	35.2	35.2	35.2
Poznan Blocks	22	10.6	10.4	9.6	44.0	44.0	44.3
	27	4.8	4.7	4.4	40.3	40.3	40.5
	32	2.2	2.2	2.0	37.1	37.1	37.3
	37	1.0	1.0	0.9	34.5	34.6	34.8
Poznan Blocks2	22	26.1	26.2	26.3	41.0	41.0	41.0
	27	10.8	10.9	10.9	38.3	38.3	38.2
	32	4.3	4.3	4.3	35.6	35.6	35.6
	37	1.5	1.5	1.6	33.5	33.5	33.5
Poznan Fencing2	22	30.0	29.5	29.5	39.6	39.7	39.7
	27	14.4	14.3	14.3	34.8	34.9	34.8
	32	5.7	5.7	5.6	31.1	31.1	31.1
	37	1.9	1.9	1.9	28.6	28.6	28.7
Poznan Service2	22	28.0	27.6	27.6	41.2	41.2	41.2
	27	11.8	11.7	11.7	38.0	38.1	38.1
	32	4.7	4.8	4.8	35.2	35.2	35.2
	37	1.8	1.8	1.8	32.8	32.9	32.9

[8] J. Hernandez-Aceituno, R. Arnay, J. Toledo, and L. Acosta, "Using Kinect on an autonomous vehicle for outdoors obstacle detection," *IEEE Sensors Journal*, vol. 16, no. 10, pp. 3603–3610, May 2016.

[9] X. Suau, J. Ruiz-Hidalgo, and J. R. Casas, "Real-time head and hand tracking based on 2.5D data," *IEEE Transactions on Multimedia*, vol. 14, no. 3, pp. 575–585, June 2012.

[10] L. Li, S. Zhang, X. Yu, and L. Zhang, "PMSC: PatchMatch-based superpixel cut for accurate stereo matching," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, no. 3, pp. 679–692, March 2018.

[11] O. Stankiewicz, K. Wegner, M. Tanimoto, and M. Domański, "Enhanced Depth Estimation Reference Software (DERS) for free-viewpoint television," ISO/IEC JTC1/SC29/WG11, Doc. MPEG M31518, Geneva, 2013.

[12] G. Lafruit, M. Domański, K. Wegner, T. Grajek, T. Senoh, J. Jung, P. Kovcs, P. Goorts, L. Jorissen, A. Munteanu, B. Ceulemans, P. Carballeira, S. Garca, and M. Tanimoto, "New visual coding exploration in MPEG: Super-MultiView and free navigation in free viewpoint TV," in *2016 Proceedings of the Electronic Imaging Conference: Stereoscopic Displays and Application*, February 2016, pp. 1–9.

[13] C.-C. Lee, A. Tabatabai, and K. Tashiro, "Free viewpoint video (FVV) survey and future research direction," vol. 4, 10 2015.

[14] Y. S. Kang and Y. S. Ho, "High-quality multi-view depth generation using multiple color and depth cameras," in *2010 IEEE International Conference on Multimedia and Expo*, July 2010, pp. 1405–1410.

[15] S. Xiang, L. Yu, Q. Liu, and Z. Xiong, "A gradient-based approach for interference cancelation in systems with multiple Kinect cameras," in *2013 IEEE International Symposium on Circuits and Systems (ISCAS2013)*, May 2013, pp. 13–16.

[16] D. Mieloch, A. Dziembowski, A. Grzelka, O. Stankiewicz, and M. Domański, "Temporal enhancement of graph-based depth estimation method," in *2017 International Conference on Systems, Signals and Image Processing (IWSSIP)*, May 2017, pp. 1–4.

[17] D. Mieloch, A. Dziembowski, A. Grzelka, O. Stankiewicz, and M. Domański, "Graph-based multiview depth estimation using segmentation," in *2017 IEEE International Conference on Multimedia and Expo (ICME)*, July 2017, pp. 217–222.

[18] V. Kolmogorov and R. Zabin, "What energy functions can be minimized via graph cuts?" *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 2, pp. 147–159, Feb 2004.

[19] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 11, pp. 1222–1239, Nov 2001.

[20] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Ssstrunk, "Slic superpixels compared to state-of-the-art superpixel methods," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 11, pp. 2274–2282, Nov 2012.

[21] O. Stankiewicz, M. Domański, and K. Wegner, "Estimation of temporally-consistent depth maps from video with reduced noise," in *2015 3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON)*, July 2015, pp. 1–4.

[22] M. Kppel, M. B. Makhlouf, M. Mller, and P. Ndjiki-Nya, "Temporally consistent adaptive depth map preprocessing for view synthesis," in *2013 Visual Communications and Image Processing (VCIP)*, Nov 2013, pp. 1–6.

[23] H. Jiang, G. Zhang, H. Wang, and H. Bao, "Spatio-temporal video segmentation of static scenes and its applications," *IEEE Transactions on Multimedia*, vol. 17, no. 1, pp. 3–15, Jan 2015.

[24] L. Sheng, K. N. Ngan, C. L. Lim, and S. Li, "Online temporally consistent indoor depth video enhancement via static structure," *IEEE*

Transactions on Image Processing, vol. 24, no. 7, pp. 2197–2211, July 2015.

[25] N. Vretos and P. Daras, "Temporal and color consistent disparity estimation in stereo videos," in *2014 IEEE International Conference on Image Processing (ICIP)*, Oct 2014, pp. 3798–3802.

[26] M. Mueller, F. Zilly, C. Riechert, and P. Kauff, "Spatio-temporal consistent depth maps from multi-view video," in *2011 3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON)*, May 2011, pp. 1–4.

[27] J. Lei, J. Liu, H. Zhang, Z. Gu, N. Ling, and C. Hou, "Motion and structure information based adaptive weighted depth video estimation," *IEEE Transactions on Broadcasting*, vol. 61, no. 3, pp. 416–424, Sept 2015.

[28] M. Sizintsev and R. P. Wildes, "Spatiotemporal stereo and scene flow via stequel matching," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 6, pp. 1206–1219, June 2012.

[29] G. Nur, S. Dogan, H. K. Arachchi, and A. M. Kondoz, "Impact of depth map spatial resolution on 3D video quality and depth perception," in *2010 3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video*, June 2010, pp. 1–4.

[30] O. Stankiewicz, K. Wegner, M. Tanimoto, and M. Domański, "Enhanced view synthesis reference software (VSRS) for free-viewpoint television," ISO/IEC JTC1/SC29/WG11, Doc. MPEG M31520, Geneva, 2013.

[31] L. Zitnick, S. B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski, "High-quality video view interpolation using a layered representation," in *ACM SIGGRAPH*, vol. 23. Association for Computing Machinery, Inc., August 2004, p. 600608. [Online]. Available: https://www.microsoft.com/en-us/research/publication/high-quality-video-view-interpolation-using-a-layered-representation/

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

[33] M. Domański, A. Dziembowski, M. Kurc, A. Luczak, D. Mieloch, J. Siast, O. Stankiewicz, and K. Wegner, "Poznan University of Technology test multiview video sequences acquired with circular camera arrangement 'Poznan Team' and 'Poznan Blocks' sequences," ISO/IEC JTC1/SC29/WG11, Doc. MPEG M35846, Geneva, 2016.

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

[35] M. Domanski, O. Stankiewicz, K. Wegner, and T. Grajek, "Immersive visual media - MPEG-I: 360 video, virtual navigation and beyond," in *2017 International Conference on Systems, Signals and Image Processing (IWSSIP)*, May 2017, pp. 1–9.

[36] G. Bjøntegaard, "Calculation of average PSNR differences between RD986 curves," ISO/IEC JTC1/SC29/WG11, Doc. MPEG M15378, Austin, 2001.

[37] HEVC reference codec. [Online]. Available: https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware