Enhancing View Synthesis with Image and Depth Map Upsampling

Adrian Dziembowski, Adam Grzelka, Dawid Mieloch, Olgierd Stankiewicz, Marek Domański
Chair of Multimedia Telecommunications and Microelectronics,
Poznań University of Technology, Poland
{adziembowski, agrzelka, dmieloch, ostank}@multimedia.edu.pl

Abstract **– In the paper we present a method for increasing the quality of views synthesized with typical Depth-Image-Based Rendering (DIBR) view synthesis algorithms. In the proposed idea the resolution of input real views and corresponding depth maps is doubled before the view synthesis. After the synthesis, the resolution of a synthesized view is downsampled back to the original resolution. This approach is transparent for the view synthesis algorithms, thus can be used with any DIBR method. In the paper, tests for two synthesis algorithms (the state-of-the-art MPEG reference software and our view synthesis method) are presented. For both algorithms, the proposed upsampling improves objective and subjective quality of synthesized views.**

Keywords – **View Synthesis; Free-Viewpoint Television; Depth Map Upsampling**

I. Introduction

The main principle of the Free-Viewpoint Television systems [1,2] is to allow a user (a viewer) to freely navigate around a scene. In order to provide the possibility of smooth virtual movement, the views from the virtual viewpoints (between the viewpoints of the real cameras) have to be synthesized [3]. Moreover, the good quality virtual view synthesis allows reducing multiview video bitstreams [4-6]. Therefore, view synthesis is one of the most crucial parts in development of Free-Viewpoint Television systems.

In practical multiview systems, where the number of used cameras is limited [7], the virtual view synthesis quality is obviously lower than in experimental systems [1] due to larger physical distance between the cameras, more occluded areas and different lighting conditions in the neighboring cameras. Therefore, in order to create high-quality multiview content, the view synthesis algorithms should be enhanced.

The state-of-the-art method of virtual view synthesis, MPEG reference software – VSRS (View Synthesis Reference Software) was designed for simple, linear camera setups. In successive improvements [8], VSRS has been only adapted for multiview systems with arbitrary camera arrangements, but the core synthesis technique remained the same. Nevertheless, VSRS is still a reputable algorithm and many researchers commonly use it.

Of course, in order to provide better synthesis quality new, more sophisticated algorithms can be created. For instance, Poznań University of Technology team developed an algorithm much more suitable for practical multiview systems – MVS (Multiview Synthesis) [9]. Just like VSRS, it is a Depth-Image-Based Rendering [10] algorithm, using scene representation based on multiple views with corresponding depth maps.

In this paper, we present a method increasing the quality of synthesized virtual views, that can be applied for any DIBR algorithm. The proposed method is simple and has very low computational complexity. Moreover, as it is only an additional preprocessing step, it does not change view synthesis algorithms, so can be easily used by any researchers developing their own view synthesis algorithms.

II. The Idea

We propose the preprocessing of the input views and corresponding depth maps. In this additional step, the resolution of all the input images is increased twice in both directions, e.g. for FullHD images the view synthesis is performed on 4K views and depth maps. After the synthesis, the resolution of the virtual view is decreased back to the original value (Fig. 1).

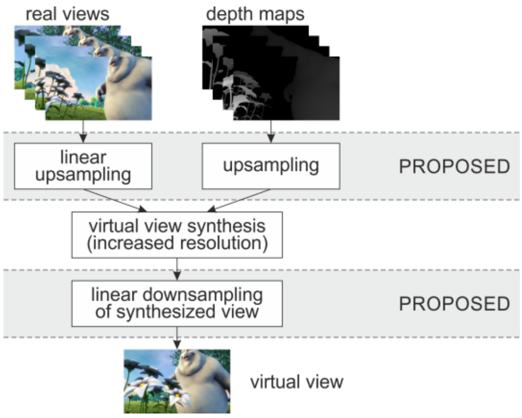


Figure 1. View synthesis process with upsampling

In the proposed method, the real views are upsampled linearly – remaining pixels are calculated by 1st order interpolation. In order to increase resolution of the depth maps, three simple upsampling methods were tested.

- 1st order (bi-linear) upsampling – analogous to texture upsampling, depth values are interpolated linearly,
- 0th order (nearest-neighbor) upsampling – remaining depth values are copied from left or upper neighbor,
- mixed upsampling, joining advantages of both the methods, as described below.

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

While for texture there are no contraindications for using simple linear interpolation of remaining pixels because sharp texture edges are not crucial for view synthesis [11], linear upsampling of the depth maps may raise doubts – the increased depth map may contain erroneous, nonexistent values at the edges.

The advantage of the 0^{th} order interpolation is that it does not generate nonexistent depth values. Copying depth values from the neighbors may only slightly displace the edges of the objects.

However, in the smooth areas the better method is to interpolate depth linearly. Therefore, we proposed third upsampling method – mixed upsampling. In the mixed upsampling, depth is interpolated linearly in the smooth areas, but it is copied from the neighbors at the edges. The threshold separating smooth and sharp areas of the depth maps was set experimentally to 1% of the depth dynamic range, e.g. for 8-bit depth maps all the areas where the difference between depth values is smaller than 3 are treated as smooth ones.

III. Experiments

In order to present, that proposed method allows to increase the quality of virtual views independently on the view synthesis algorithm, experiments were performed using two different methods: MPEG reference software VSRS [8] and our algorithm – MVS [9].

The test set contains 6 multiview sequences:

- Ballet and Breakdancers [12],
- Big Buck Bunny Flowers Arc and Butterfly Arc [13],
- Poznań Blocks2 and Poznań Fencing2 [14].

All the sequences were captured by multiview systems with cameras arranged along an arc. Two of them (both BBB sequences) are synthetic, 4 were captured by real-world multicamera systems.

Both the subjective and objective quality were measured. For the objective quality estimation, PSNR was chosen. We are aware about the disadvantages of such an approach [15], however, PSNR was chosen because of its simplicity and popularity. For objective quality measurement, the virtual views were synthesized in the position of the real views, which were treated as reference views for PSNR measurement.

In order to estimate the subjective quality of synthesized views the similarity judgment method [16,17] was chosen. It allows the direct comparison of two views synthesized in two ways: with and without upsampling.

For each test sequence one static frame was scored. In moving sequences, flickering characteristics of depth artifacts do not allow to detect the differences between views being rated.

The subjective tests were performed with help of 12 experts. Each of them compared 6 randomly ordered pairs of virtual views synthesized with and without upsampling. For each pair, the testers evaluated, which image has better quality and how big is the difference between them in a scale from 0 to 5, where 0 is no difference and 5 means, that the quality is significantly higher. Thus, the subjective quality gain from the upsampling is between -5 and 5.

IV. Experimental Results

In Tables I and II the calculated PSNR values for two tested view synthesis algorithms are presented. For each test sequence, the virtual views were synthesized in four ways: the typical approach (without upsampling) and proposed approach with upsampling with 3 different depth interpolation methods.

TABLE I. Objective Quality of Virtual Views Synthesized Using Reference Software (VSRS) with 4 Upscaling Methods

Test Sequence	**PSNR of synthesized views [dB]**			
	no upsampling	***0th order upsampling***	***1st order upsampling***	***mixed upsampling***
Ballet	31.05	31.14	31.20	31.13
Breakdancers	31.47	31.69	31.61	31.83
BBB[a] Flowers Arc	22.68	23.13	23.14	23.22
BBB Butterfly Arc	32.04	32.71	32.57	32.71
Poznań Blocks2	29.25	29.92	29.82	29.87
Poznań Fencing2	28.33	28.97	29.01	29.12
Avg. PSNR	**29.14**	**29.59**	**29.56**	**29.65**

a. Big Buck Bunny [13]

TABLE II. Objective Quality of Virtual Views Synthesized Using Our Method (MVS) with 4 Upscaling Methods

Test Sequence	**PSNR of synthesized views [dB]**			
	no upsampling	***0th order upsampling***	***1st order upsampling***	***mixed upsampling***
Ballet	31.74	32.14	32.17	32.29
Breakdancers	31.73	32.26	32.08	32.14
BBB[a] Flowers Arc	25.81	26.49	26.56	26.73
BBB Butterfly Arc	33.56	34.27	34.10	34.25
Poznań Blocks2	29.61	30.28	30.19	30.37
Poznań Fencing2	28.82	29.58	29.47	29.53
Avg. PSNR	**30.21**	**30.83**	**30.76**	**30.88**

a. Big Buck Bunny [13]

Independently on used upsampling method and synthesis algorithm, the upsampling allows to increase the quality of the virtual views by more than 0.4 dB (average for the entire test set). The mean PSNR values for both tested synthesis algorithms and all upsampling methods are presented in Fig. 2.

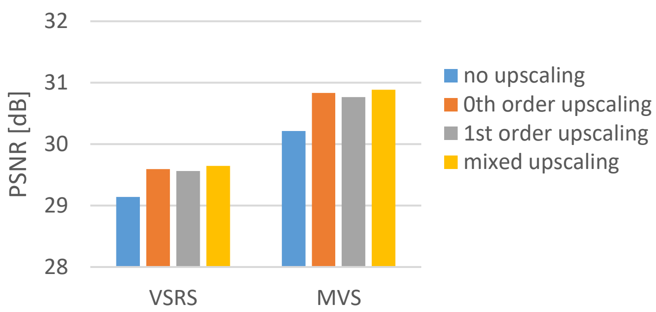


Figure 2. Objective quality of virtual views synthesized with and without upscaling for reference software (VSRS) and our method (MVS)

Presented results confirm, that linear, 1^{st} order interpolation is not the best choice for depth upsampling. On average, the best method of upsampling is the mixed one, which allows increasing the quality by additional 0.1 dB compared to 1^{st} order upsampling.

In subjective tests only MVS with one of the upsampling methods was tested and compared with the typical, non-upsampling approach. Mixed upsampling was chosen because of the highest gain in PSNR. In Fig. 3. the gain of the subjective quality (including confidence intervals) for all test sequences is presented.

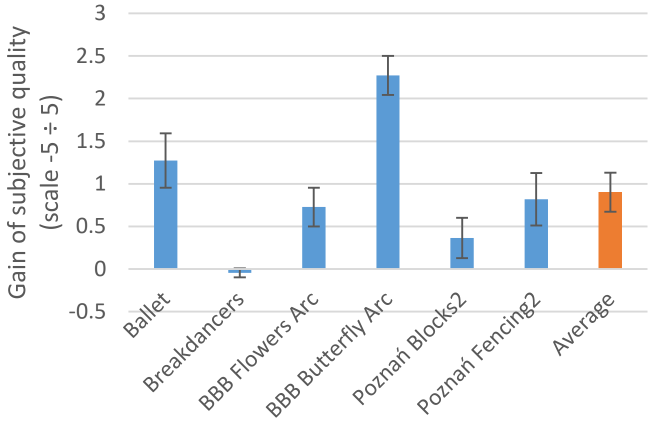
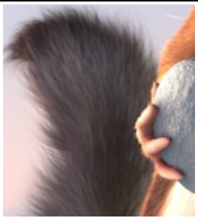
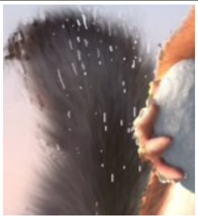
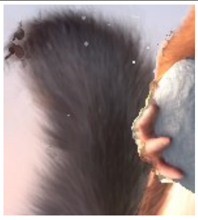
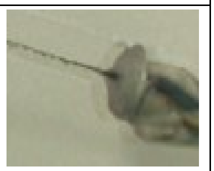
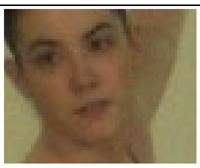
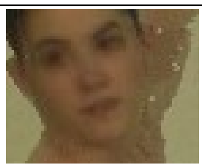
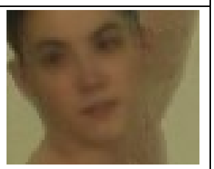
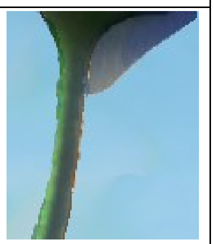


Figure 3. Gain of subjective quality from upscaling

On average, the quality of virtual views synthesized using upsampling was higher by 0.9 in a scale between -5 and 5. For almost all the test sequences the quality increase was statistically significant (significance level: 5%).

TABLE III. FRAGMENTS OF REFERENCE VIEWS AND VIEWS SYNTHESIZED USING MVS [9] WITH AND WITHOUT UPSCALING

	reference view	fragments of synthesized views	
		no upscaling	*upscaling*
a.			
b.			
c.			
d.			

The only sequence, for which there were no subjective quality gain was Breakdancers. The reason of that is uncommon characteristics of that sequence. Breakdancers is a dark, blurred sequence, with few sharp edges. Thus, any resolution increase provided by proposed upsampling does not increase the subjective quality of synthesized view.

In Table III the fragments of the virtual views synthesized without and with upsampling are presented. The left column of the Table contains fragments of the reference views. Upsampling allows to better preserve edges in the virtual view (Table III, fragments b. and d.). Views synthesized using upsampling are also more detailed and have less annoying artifacts (fragments a. and c.).

V. CONCLUSIONS

An efficient method for quality improvement of view synthesis was presented in this paper. It consist an additional simple preprocessing step before view synthesis, in which the input views and corresponding depth maps are upsampled. Three different upsampling methods were tested. As it was shown, independently of particular upsampling method, application of the proposed scheme results in increase the quality of the virtual views.

Probably, more sophisticated upsampling methods than three presented in the paper would increase the synthesis quality even more, but even the primitive methods of upsampling increase both the objective and subjective quality of the virtual views. Presented upsampling increases PSNR by more 0.5 dB and subjective quality by 0.9 in a scale between -5 to 5.

Our further works will focus on testing better upsampling methods in order to increase the view synthesis quality much more.

REFERENCES

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

[2] G. Lafruit, M. Domański, K. Wegner, T. Grajek, T. Senoh, J. Jung, P. Kovacs, P. Goorts, L. Jorissen, A. Munteanu, B. Ceulemans, P. Carballeira, S. Garcia, M. Tanimoto, "New visual coding exploration in MPEG: Super-MultiView and Free Navigation in Free viewpoint TV", IST Electronic Imaging, Stereoscopic Displays and Applications XXVII, San Francisco 2016.

[3] M. Domański, M. Gotfryd, K. Wegner, "View synthesis for multiview video transmission", The Int. Conf. Image Proc., Computer Vision, and Pattern Recognition, 2009.

[4] M. Domański, O. Stankiewicz, K. Wegner, M. Kurc, J. Konieczny, J. Siast, J. Stankowski, R. Ratajczak, T. Grajek, "High Efficiency 3D Video Coding Using New Tools Based on View Synthesis", IEEE Trans. Image Proc., vol. 22, 2013, pp. 3517-3527.

[5] K. Müller et al., "3D High-Efficiency Video Coding for multi-view video and depth data", IEEE Trans. Image Proc., vol. 22, 2013, pp. 3366-3378.

[6] J. Stankowski et al., "3D-HEVC extension for circular camera arrangements", The 3DTV Conference, 2015.

[7] 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, 2015.

[8] 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, MPEG M31520, 2013.

[9] A. Dziembowski, A. Grzelka, D. Mieloch, O. Stankiewicz, K. Wegner, M. Domański, "Multiview Synthesis – improved view synthesis for virtual navigation", Picture Coding Symposium, 2016.

[10] Z. Sun; Ch. Jung "Real-Time Depth-Image-Based Rendering on GPU," in 2015 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery, Xi'an, 2015, pp: 324 – 328.

[11] A. Dziembowski, A. Grzelka, D. Mieloch, O. Stankiewicz, M. Domański, "Depth map upsampling and refinement for FTV systems", International Conference on Signals and Electronic Systems, ICSES 2016.

[12] L. Zitnick, S.B. Kang, M. Uyttendaele, S. Winder, R. Szeliski, "High-quality video view interpolation using a layered representation", ACM SIGGRAPH, 2004.

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

[14] 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. MPEG M38247, May 2016.

[15] Z. Wang, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity", IEEE Transactions on Image Processing, vol. 13, no. 4, pp. 600-612, April 2004.

[16] R. Mantiuk, A. Tomaszewska, R Mantiuk, "Comparison of four subjective methods for image quality assessment", Computer Graphics Forum", vol. 0, no. 0, 1981, pp. 1 – 13.

[17] H. de Ridder, G. Majoor, "Numerical category scaling: an efficient method for assessing digital image coding impairments", Proc. of SPIE, vol. 1249, 1990, p. 65.