Omnidirectional View Synthesis and Test Images

Krzysztof Wegner
Chair of Multimedia Telecommunications and Microelectronics
Poznan University of Technology
Poznań, Poland
kwegner@multimedia.edu.pl

Dominika Łosiewicz
Chair of Multimedia Telecommunications and Microelectronics
Poznan University of Technology
Poznań, Poland

Tomasz Grajek
Chair of Multimedia Telecommunications and Microelectronics
Poznan University of Technology
Poznań, Poland
tgrajek@multimedia.edu.pl

Olgierd Stankiewicz
Chair of Multimedia Telecommunications and Microelectronics
Poznan University of Technology
Poznań, Poland
ostank@multimedia.edu.pl

Adrian Dziembowski
Chair of Multimedia Telecommunications and Microelectronics
Poznan University of Technology
Poznań, Poland
adziembowski@multimedia.edu.pl

Marek Domański
Chair of Multimedia Telecommunications and Microelectronics
Poznan University of Technology
Poznań, Poland
marek.domanski@put.poznan.pl

***Abstract*—In this paper we introduce an algorithm for omnidirectional virtual view synthesis based on Omnidirectional Video plus Depth (OVD) format. The implementation is done on the basis of View Synthesis Reference Software (VSRS) developed by MPEG of ISO/IEC. Also, we address the problem of lack of benchmark data disallowing objective quality assessment. We present a method for generating test images in OVD representation along with example omnidirectional images and omnidirectional depths, called "*Poznan Hall 360*".**

Keywords—omnidirectional view synthesis, OVD, DIBR, immersive, MPEG-I, 6 DoF

I. Introduction

One of the challenges for the new generation of video systems is immersiveness. Classically, a scene is watched on a rectangular screen or monitor, which places a user in a position external to the scene. This is true even for most of 3D stereoscopic system. In the envisioned immersive systems, a user is expected to be inside a scene and possibly interact with it. One of the considered solutions is panoramic or omnidirectional video, displayed on a Head Mounted Display (HMD).

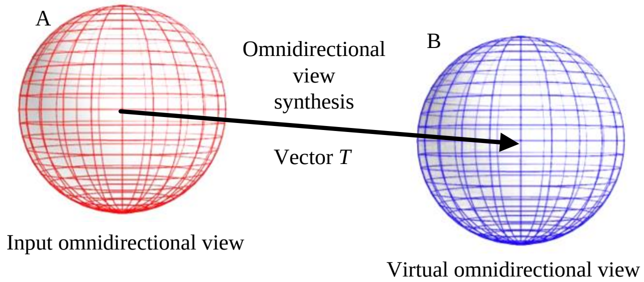

Fig. 1. Omnidirectional virtual view synthesis.

One of the problems related to omnidirectional video formats is that natively they provide capability of changing direction of viewing but not the position of viewing. Although in many applications a user is not supposed to move significantly, e.g. walk, the inability to shift position of the head in the virtual scene even slightly (and observe perspective change in HMD) causes unpleasant symptoms similar to motion sickness, called virtual reality sickness [8]. This problem can be solved by allowing a user to move head freely and to reflect those movements in the presented content. Therefore, there is a need for a synthesis of omnidirectional view as seen from virtual position, different than the original position of acquisition (Fig. 1).

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" in 2018.

In this paper we consider a novel 3D scene representation format that allows a user to look around and change position. It is composed of omnidirectional video accompanied by omnidirectional depth, called Omnidirectional Video plus Depth (OVD). As it has been shown in our previous works [1], [2], OVD representation can be generated from stereoscopic omnidirectional video by means of depth estimation. Stereoscopic omnidirectional panoramas are commonly captured with e.g. circular cameras (Fig. 2), and thus OVD representation is feasible for natural content.

Fig. 2. Commercially available 360 degree camera.

Up to now, several custom applications for omnidirectional rendering have been demonstrated [4], [5], [6], [7] but none of them works with OVD format. Other representation formats, which support comparable feature set, like point-clouds or 3D-mesh-based representations, are not image-based and are beyond the scope of this work.

In this paper, we introduce an algorithm for omnidirectional virtual view synthesis based on OVD format. The implementation is done on the basis of View Synthesis Reference Software (VSRS) developed by Motion Picture Experts Group (MPEG) of International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC), during the works on 3D and free viewpoint television (FTV) [9]. VSRS is commonly used as a reference technique for classical, rectangular-image virtual view synthesis algorithms evaluation and with the

© 2018 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

DOI: 10.1109/ICSES.2018.8507329

enhancement proposed in this paper would also be used in the context of omnidirectional video.

One of the major problems with works on new immersive systems is lack of benchmark data allowing objective quality assessment. In this paper we address also this subject, by proposing a method for generation of test images in OVD representation along with some generated example omnidirectional images and omnidirectional depths.

II. Omnidirectional View Synthesis

The goal of the omnidirectional virtual view synthesis is to create omnidirectional virtual view from any position that is different from the position of input omnidirectional view, e.g. translated by vector T. The created omnidirectional virtual view is expected to look like directly captured with a 360 degree camera.

There are many ways to map omnidirectional video to rectangular video [13]. One of the most popular in the literature is equirectangular projection format (ERP).

The ERP format is the most widely used projection format for representing 360 degree video as a rectangular video. It is the default projection format used by Joint Video Exploration Team (JVET) [14] of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11 in works on compression of 360 degree materials.

In this work, we assume that both input omnidirectional views and the requested virtual viewpoint use equirectangular projection (Fig. 3).

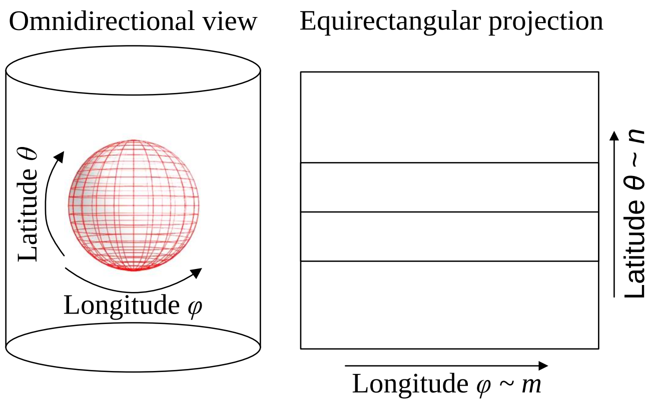

Fig. 3. Representation of 360-degree panorama video in equirectangular format.

In order to create virtual omnidirectional view, each pixel in each input omnidirectional view needs to be mapped to appropriate position in virtual viewpoint. At first, 3D location of each pixel of input omnidirectional view has to be obtained.

For every pixel (m, n) in the rectangular input video of size (W, H) we obtain appropriate direction (φ, θ) in 3D space through (1):

$$\begin{aligned} \varphi &= \left(\frac{m}{W} - 0.5\right) \cdot 2\pi \, , \\ \theta &= \left(0.5 - \frac{n}{H}\right) \cdot \pi \, . \end{aligned} \quad (1)$$

From the depth map associated with video, we get depth value for the processed point (m, n). Given the depth map format used [3] (e.g. normalized disparity) we can easily calculate distance z of the given point from the center of the omnidirectional viewpoint.

For direction (φ, θ) and the distance z, exact 3D position of the point can be calculated (2):

$$\begin{aligned} X &= \ \ z \cdot cos(\theta) \cdot cos(\varphi) \, , \\ Y &= \ \ z \cdot sin(\theta) \, , \\ Z &= -z \cdot cos(\theta) \cdot sin(\varphi) \, . \end{aligned} \quad (2)$$

Next, 3D point is projected onto a virtual viewpoint sphere to create an omnidirectional view of the requested viewpoint. Moving of the viewpoint is equivalent to moving of the coordinates of all points in 3D space. Thus each 3D point is shifted by translation of virtual view point from input viewpoint $T = [T_x \quad T_y \quad T_z]$ as expressed in following equations (3):

$$\begin{aligned} X' &= X + T_x \, , \\ Y' &= Y + T_y \, , \\ Z' &= Z + T_z \, . \end{aligned} \quad (3)$$

For the obtained 3D position (X', Y', Z'), with respect to virtual viewpoint coordinate system, we can calculate direction (φ', θ') relative to the virtual viewpoint center (4):

$$\begin{aligned} \varphi' &= tan^{-1}\left(-\frac{Z}{X}\right) \, , \\ \theta' &= sin^{-1}\left(\frac{Y}{\sqrt{X^2 + Y^2 + Z^2}}\right) . \end{aligned} \quad (4)$$

In the end, appropriate pixel position (m', n') in the output rectangular video (representing the omnidirectional virtual view) can be obtained from direction (φ', θ') in 3D space through equation (5):

$$\begin{aligned} m' &= \left(\frac{\varphi'}{2\pi} + 0.5\right) \cdot W \, , \\ n' &= \left(0.5 - \frac{\theta'}{\pi}\right) \cdot H \, . \end{aligned} \quad (5)$$

Based on a one or two input omnidirectional videos and omnidirectional depth in equirectangular projection format (ERP) new omnidirectional viewpoint can be rendered by means of Depth Image Based Rendering (DIBR) known from classical rectangular video.

III. Implementation

We have implemented the presented omnidirectional view synthesis method in View Synthesis Reference Software (VSRS) developed by MPEG [9].

The general design of VSRS is shown in Fig. 4. Synthesis takes place in two separate processing paths - one per each input view. In each path, at first the depth map of a rendered omnidirectional virtual view is created in depth synthesis block. This is performed independently based on both input depth maps. Forward depth warping of each pixel from input view is used. Next, virtual depth maps are filtered by the same set of processing filters already present in VSRS. After that, in each path, the texture of virtual view is created by backward warping in the view synthesis block. Based on rendered omnidirectional virtual depth map a texture of a processed input view is being sampled to create variants of virtual view texture. The process is ended by application of set of filters already present in VSRS.

Finally, the two variants of texture of virtual views are merged together at the view blending stage. All remaining holes and regions that could not be rendered due to occlusion within the scene are inpainted at the hole-filling and inpainting stage.

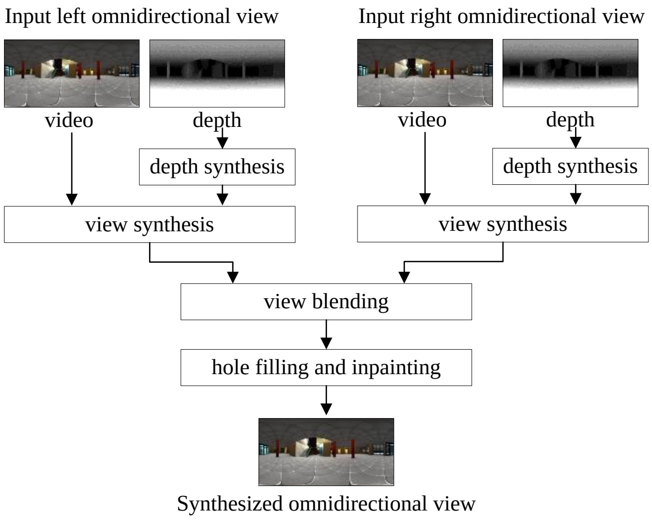

Fig. 4. General overview of the algorithm used in VSRS for virtual view rendering.

Thus, the only steps that has been modified in order to support omnidirectional view rendering are depth synthesis and view synthesis. The rest of the processing pipeline remains unaltered. This means that all quality improvement tools in VSRS, including modes for view blending and hole filling are not affected.

IV. Test Methodology

As mentioned in the introduction, the works on new immersive systems are hampered by the difficulties in objective quality assessment.

We propose to solve this problem with the usage of view-resynthesis-based approach. The input omnidirectional view at position A (Fig. 1) is used to synthesize a virtual omnidirectional view at position B. The quality of the attained result is evaluated by comparison with the original omnidirectional view in the same position B.

Obviously, for this we require at least two omnidirectional views of the same scene: captured at positions A and B. Capturing of such test material practically is difficult if we consider real scenes and cameras.

We propose to generate test images and reference ground truth depth maps from a 3D model. The ability to directly render reference viewpoint in any place in the 3D space allows evaluation of omnidirectional depth estimation and omnidirectional view synthesis.

Of course, various objective quality measures can be used. Here we mention commonly known Peak Signal-to-Noise Ratio (PSNR) and Weighted-Samples PSNR (WS-PSNR) [15].

V. Generation of Omnidirectional Test Image

In order to support the testing methodology presented above, we present a 3D model which can be used to render synthetic omnidirectional content. The model has been created in Blender [12]. This allows modifications and simple rendering of the required data.

The scene presents entrance hall of the building of Faculty of Electronics and Telecommunications of Poznan University of Technology (Fig. 5). The same hall has been used in "*Poznan Hall 2*" sequence, widely used in experiments related to Multiview Video plus Depth (MVD) compression and rendering [10]. Therefore, we call this new sequence "*Poznan Hall 360*".

The main omnidirectional camera is placed in the middle of the hall, allowing a user to look around. Additional omnidirectional camera is placed to the left and to the right of the main camera.

All omnidirectional cameras are configured to render images and depth maps in equirectangular projection (ERP) format. Exemplary rectangular view has been shown in Fig. 6.

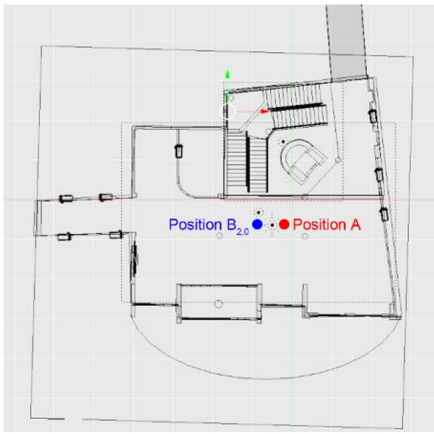

Fig. 5. Overview of the camera positioning in "*Poznan Hall 360*" scene.

Fig. 6. Example rectangular view frame of omnidirectional view of the proposed "*Poznan Hall 360*" scene (left) and original photo of the hall for comparison (right).

Instead of creating depth map in Blender's native z-distance format, we have applied a processing pipeline directly in Blender rendering graph which converts the rendered z-buffer from distance measurement Z to normalized disparity format v. Therefore, the attained depth format is the same as usually used in VSRS [11]:

$$v = \frac{\frac{1}{Z} - \frac{1}{Z_{far}}}{\frac{1}{Z_{near}} - \frac{1}{Z_{far}}} \cdot v_{max} \,. \qquad (6)$$

We provide (https://svn.multimedia.edu.pl/content) rendered images and depth maps in two resolutions: 2048×1024 and 8196×4096. Of course, images and depth maps with other resolution can be rendered out of the provided source Blender file.

VI. Results

We have used the proposed "*Poznan Hall 360*" as input data for the proposed omnidirectional-view synthesis method. Basing on omnidirectional view at position A, we have synthesized 4 omnidirectional views in different positions, in between of position A and position $B_{2.0}$ (Fig. 5) (index expressed in Blender units). At the same positions, also references from Blender have been rendered. The quality has been measured with PSNR and WS-PSNR [15]. The results are depicted in Table 1. As it can be seen (also c.f. Fig. 7) the quality of the synthesized omnidirectional view is quite high, which corresponds to PSNRs around 22 to 35dB.

a)

b)

c)

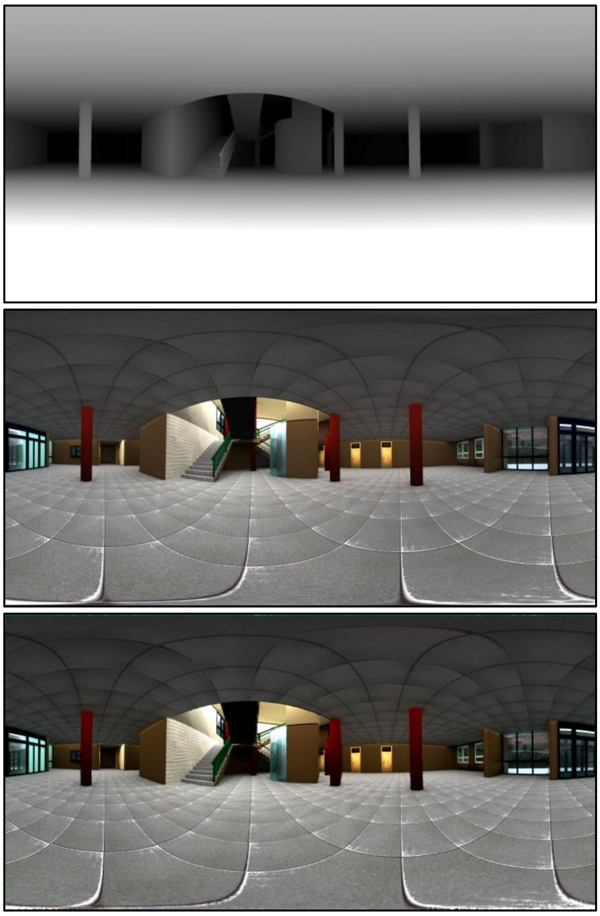

Fig. 7. Example frame of omnidirectional view of the proposed *"Poznan Hall 360"* scene rendered in Blender (b), corresponding rendered depth map (a) and omnidirectional view synthesized in the same position ($B_{2.0}$) with the use of proposed synthesis method (c).

TABLE I. OBJECTIVE QUALITY EVALUATION RESULTS IN TERMS OF LUMINANCE PSNR AND LUMINANCE WEIGHTED-SAMPLE PSNR

Position of synthesized virtual viewpoint	PSNR [dB]	WS-PSNR [dB]
$B_{0.5}$	35.32	34.79
$B_{1.0}$	31.94	31.39
$B_{1.5}$	25.03	24.50
$B_{2.0}$	22.93	22.39

VII. CONCLUSIONS

In this paper, we have presented a new omnidirectional view synthesis method based on novel Omnidirectional Video plus Depth (OVD) format. The method has been implemented in MPEG View Synthesis Reference Software (VSRS) [11] which enables using it as a reference technique in the future research.

We have also considered evaluation of the quality of virtual omnidirectional views. For this, we have proposed a methodology based on view re-synthesis and comparison to the original view, generated from 3D model. To support this methodology, we have provided a 3D model in Blender along with generated example views and depth maps, which may be used as benchmark test material. The model represents the entrance hall of a building of Poznan University of Technology – Chair of Multimedia Telecommunications and Microelectronics - and thus is called "Poznan Hall 360". We report PSNR and WS-PSNR quality metrics which vary from about 22dB to 35dB depending on the position of virtual omnidirectional viewpoint.

ACKNOWLEDGMENT

The test materials presented in the paper may only be used for the scientific purposes. Acknowledgement and reference to this paper is required in all documents that report any usage of the materials.

REFERENCES

[1] K. Wegner, O. Stankiewicz, T. Grajek, M. Domański, "Depth estimation from circular projection of 360 degree 3D video," ISO/IEC JTC1/SC29/WG11, MPEG2017/m40596, Hobart, Australia, April 2017.

[2] M. Domański, O. Stankiewicz, K. Wegner, T. Grajek, "Immersive visual media – MPEG-I: 360 video, virtual navigation and beyond," IEEE International Conference on Systems, Signals and Image Processing IWSSIP 2017, Poznań, Poland, 22-24 May 2017.

[3] T. Senoh, K. Yamamoto, R. Oi, T. Mishina, M. Okui, "Consideration of depth format," ISO/IEC JTC1/SC29/WG11 MPEG m15047, Antalya, Turkey, January 2007.

[4] B. Kroon, "Adding depth maps to fulfill MPEG-I phase 1b parallax requirement," ISO/IEC JTC1/SC29/WG11 m41824, Macau, China, October 2017.

[5] B. Wang, Y. Sun, L. Yu, "[MPEG-I-Visual] 360 video sequence with depth map," JTC1/SC29/WG11 m41863, Macau, China, October 2017.

[6] R. Dore, M.-L. Champel, "3DoF+ proof of concept," JTC1/SC29/WG11 m40862, Torino, Italy, July 2017.

[7] R. Doré, I. Orlac, G. Briand, M.-L. Champel, "User tests on interactive parallax and stereoscopy," ISO/IEC JTC1/SC29/WG11 MPEGM40863, Torino, Italy, July 2017.

[8] R. A. Ruddle, "The effect of environment characteristics and user interaction on levels of virtual environment sickness," Proceedings of IEEE Virtual Reality, vol. 11, pp. 141–148, 2004.

[9] K Wegner, O. Stankiewicz, M. Tanimoto, M. Domański, "Enhanced view synthesis reference software (VSRS) for free-viewpoint television," JCT-3V (ITU-T SG 16 WP 3 and ISO./IEC JTC 1/SC 29/WG 11), MPEG_M31520, Geneva, Switzerland, 25 October-1 November 2013.

[10] M. Domański, T. Grajek, K. Klimaszewski, M. Kurc, O. Stankiewicz, J. Stankowski, K. Wegner, "Poznań multiview video test sequences and camera parameters," ISO/IEC JTC1/SC29/WG11, MPEG/M17050, m17050, Xian, China, 26-30 October 2009.

[11] K. Wegner, O. Stankiewicz, T. Grajek, M. Domański, "Depth map formats used within MPEG 3D frameworks," ISO/IEC JTC1/SC29/WG11 MPEG2017, M40019, Geneva, Switzerland, 16-20 January 2017.

[12] Blender homepage, https://www.blender.org/ [online access 10.02.2018]

[13] Y. Ye, E. Alshina, J. Boyce, "Algorithm descriptions of projection format conversion and video quality metrics in 360Lib," Joint Video Exploration Team (JVET) of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11 6th Meeting: doc. JVET F-1003, Hobart, AU, 31 March – 7 April 2017.

[14] D. Rusanovskyy, F. Chen, L. Zhang, T. Suzuki, "3D-AVC test model 8," Joint Collaborative Team on 3D Video Coding Extensions of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, 6th Meeting: Geneva, Document: JCT3V-F1003, November 2013.

[15] S. Yule, A. Lu, Y. Lu, "WS-PSNR for 360 video objective quality evaluation," Joint Video Exploration Team of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, JVET-D0040, 4th Meeting, Chengdu, China, October 2016.