DEPTH IMAGE-BASED RENDERING OF NON-LAMBERTIAN CONTENT IN MPEG IMMERSIVE VIDEO

Sarah Fachada[1*], *Daniele Bonatto*[1], *Yupeng Xie*[1],
Patrice Rondao Alface[2], *Mehrdad Teratani*[1], *Gauthier Lafruit*[1]

[1]Laboratories of Image Synthesis and Analysis, Université Libre de Bruxelles, Brussels, Belgium
[2]Nokia Bell Labs, Antwerp, Belgium

(a) Cadillac (b) Mirror

(c) Magritte Mirror (d) Magritte Transparent

Fig. 1: The four test sequences used in the exploration experiment. (a) (b) *Courtesy of InterDigital*.

ABSTRACT

In the context of the development of MPEG-I standard for immersive video compression ISO/IEC 23090-12 (MIV), the need of handling scenes with non-Lambertian materials arose. This class of material is omnipresent in natural scenes, but violates all the assumptions on which depth image-based rendering (DIBR) is based. In this paper, we present a view-synthesizer designed to handle non-Lambertian objects with DIBR, replacing the classical depth maps by multi-coefficients non-Lambertian maps. We report the results of the exploration experiments on Future MIV designed to test this rendering method against the classical DIBR approaches, and demonstrate promising results on all the tested sequences.

***Index Terms*—** MPEG-I, standardization, view synthesis, DIBR, non-Lambertian

*Sarah Fachada is a Research Fellow of the Fonds de la Recherche Scientifique - FNRS, Belgium

1. INTRODUCTION

Depth image-based rendering (DIBR) [1] is a view synthesis method that recreates the parallax of a scene by warping the pixels of the input images according to their disparity, inversely proportional to their depth. This technique, based on the assumption that the scene is Lambertian, is the core basis of MPEG-I's 6DoF view synthesizer RVS [2, 3, 4, 5]. Since the creation of DIBR, many advances have been made to solve the problems specific to this technique, such as disocclusions handling [6, 7, 8, 9], cracks in the objects [10, 11], color correction [12] or ghosting [13]. However, advances focusing on non-Lambertian objects remain in the minority and require additional information such as geometry and normals or prior knowledge of the background [14, 15] or handle only planar specular reflection [16, 17].

A call for new test material including non-Lambertian objects [18] has brought to MPEG-I new challenging video [19, 20, 21, 22] that classical DIBR algorithms fail to render with high fidelity. In order to handle such scenes with a DIBR-based method, RVS has been extended to RVS 4.0 that replaces traditional disparity maps by non-Lambertian polynomial maps, which describe the non-linear displacement of features visible on non-Lambertian surfaces [23, 24]. To test the performance of the view synthesizer on this specific content, an exploration experiment has been designed to compare our results with single depth map DIBR [25, 26].

2. PROPOSED METHOD

RVS 4.0 is based on the observation that features observed on the surface of non-Lambertian objects move non-linearly with respect to a linear camera movement, contrary to diffuse objects, which parallax is inversely proportional to their distance to the camera. As described in [24], two polynomials of degree up to three model this displacement.

For a Lambertian object, the pixel (p_x, p_y) displacement in function of the camera (x, y) displacement in a camera array depends only on the object depth d and focal length f:

$$\frac{\partial p_x}{\partial x} = \frac{\partial p_y}{\partial y} = \frac{f}{d}$$
$$\frac{\partial p_x}{\partial y} = \frac{\partial p_y}{\partial x} = 0 \quad (1)$$

For non-Lambertian objects, however, this displacement also depends on the normal of the objects, the refraction index and the geometry of the surrounding scene, so a closed-form equation is impossible to formulate. Instead, we approximate the pixel displacement by a two-variables polynomial function of the camera position:

$$p_x = P_x(x, y)$$
$$p_y = P_y(x, y) \quad (2)$$

For a forward-backward camera movement (z direction), the ray is projected in the input camera plane and the pixel displacement approximated. In practice, these polynomials are of degree up to three (eg. eighteen coefficients, nine for P_x and nine for P_y) to model more complex non-Lambertian objects. In simple cases, such as planar reflections where the pixel movement is linear and corresponds to the depth of the reflected object, one coefficient is enough: $P_x(x, y) = a_0 x$ and $P_y(x, y) = a_0 y$. The choice of the degree is done manually, depending on the scene type. Based on the detected pixel displacements among the views, the non-Lambertian objects can be segmented [27], but it is not sufficient to choose the most suited polynomial degree. First, in the case of a low number of input views to compute the maps, low degree polynomials are needed to avoid overfitting (contrary to stereo depth estimation, polynomials are an approximation). Second, in the case of a weak accuracy of the matching (textureless areas, repetitive patterns), false positives can occur on objects that should be modeled by low degree polynomials (Lambertian objects and planar mirrors).

The polynomials are computed pixel-wise given the optical flow between input images (exactly like a depth map computation). We provide an ablation study on the required precision (number of bits) of the polynomials coefficients and their degree in [24].

Those coefficients do not bring information on the object geometry, but describe how their apparent features are displaced with the camera movement. Hence this method is designed for non-Lambertian objects where clear features are identifiable among the input images, such as mirrors, specularities and refracted objects. Consequently, failure cases include small multi-faceted objects and surfaces where no features are recognizable.

Additionally to the coefficients of the polynomial, we use a traditional depth map to render the diffuse objects, which are identified with a mask. Occlusions are handled with the diffuse depth maps as the polynomial maps do not contain geometry information. In this experiment, we used the ground truth depth map for occlusion handling and diffuse object rendering.

	Ground truth	Estimated	Polynomial (Proposed)
Cadillac	**28.35**	20.40	26.43
Mirror	25.47	**26.40**	25.54
Magritte T	30.66	25.18	**31.12**
Magritte M	31.30	25.97	**33.76**

Table 1: Average PSNR (dB) for each kind of depth map (ground truth, estimated and polynomial).

	Ground truth	Estimated	Polynomial (Proposed)
Cadillac	**34.16**	24.90	32.07
Mirror	30.11	**32.34**	31.40
Magritte T	36.92	31.68	**40.34**
Magritte M	36.74	32.81	**41.64**

Table 2: Average IV-PSNR (dB) for each kind of depth map (ground truth, estimated and polynomial).

3. DATASETS

In this paper, we focus on synthetic ray-traced datasets, where the geometry of the object is available through ground truth depth maps. The datasets are 17-frames video test sequences of MPEG-I arranged in camera arrays. They present various kinds of non-Lambertian objects: Cadillac [22] has a transparent showcase with reflections and a glossy car with a transparent windshield (Figure 1a), Mirror [19] shows a planar mirror and a small curved mirror (Figure 1b) and Magritte [21] presents two versions of a ball: one fully reflective (Figure 1c), the other fully refractive (Figure 1d). The resolution is 1920×1080 for Cadillac and Mirror and 2000×2000 for Magritte sequences.

4. EXPLORATION EXPERIMENT

We compare the datasets in three experimental conditions: ground truth depth map for every object; estimated depth maps with IVDE [28]; ground truth depth maps for Lambertian-objects and polynomial maps for non-Lambertian objects (proposed). IVDE is MPEG-I's new reference depth estimator, enabling view consistency thanks to micro-segment matching instead of pixel matching, improving over DERS - the former depth estimation reference software of MPEG-I, a multi-stereo depth estimator with graph-cut refinement [29]. The polynomial maps have one coefficient for Magritte and Cadillac, based on an IVDE estimation, due to the small number of available input images (higher degree polynomials would overfit the pixel displacement), and four coefficients for Magritte Mirror and eighteen coefficients for Magritte Transparent (based on preliminary results on the dataset [30]). For Mirror and Cadillac, we used four input im-

(a) Reference (b) Ground truth depth (c) Estimated depth [28] (d) Polynomial map [24]

Fig. 2: Zoomed details on a synthesized view of the dataset Cadillac.

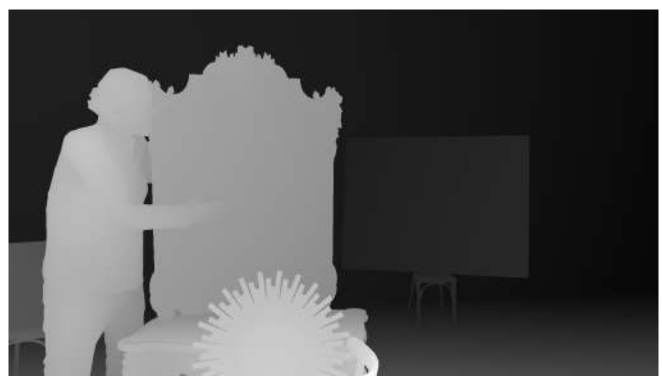
ages to synthesize the novel views. Those views correspond to the corner of the 5×3 camera array of the dataset. For Magritte sequences, we used nine input images evenly spaced in the 21×21 square camera array of the dataset.

The experiment consists in synthesizing the other views of the camera arrays using RVS 4.0. Objective metrics (PSNR and IV-PSNR [31]) are computed on the 17 frames of the synthetic cameras. IV-PSNR is an error metric designed for immersive video: instead of the pixel-wise error, it allows a window of displacement to find the best matching pixels within a user's unnoticeable shift range.

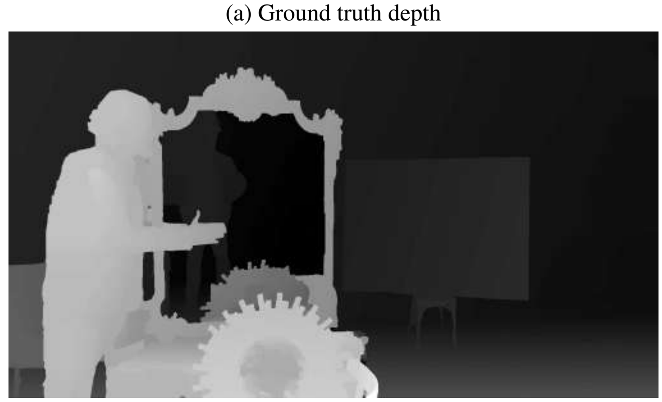
The objective results are reported in tables 1 (PNSR) and 2 (IV-PSNR). Depending on the sequence and nature of the non-Lambertian objects, the best performing type of depth map varies. We explain those differences using the visual results in the following paragraphs.

Cadillac As most of the non-Lambertian content of this sequence consists of semi-reflections (red car with superimposed reflections and inside of the shop with superimposed reflections), the ground truth depth maps reach the best objective results. Indeed, the general color is mostly correct in non-Lambertian objects. Moreover, the ground truth depth maps do not suffer from geometric errors as the estimated depth maps do. However, we can observe geometrically inconsistent reflections in the car hood, handled more correctly when using polynomial maps (Figure 2).

Mirror Reflections in planar mirrors behave as Lambertian objects placed behind the mirror instead of on the mirror sur-

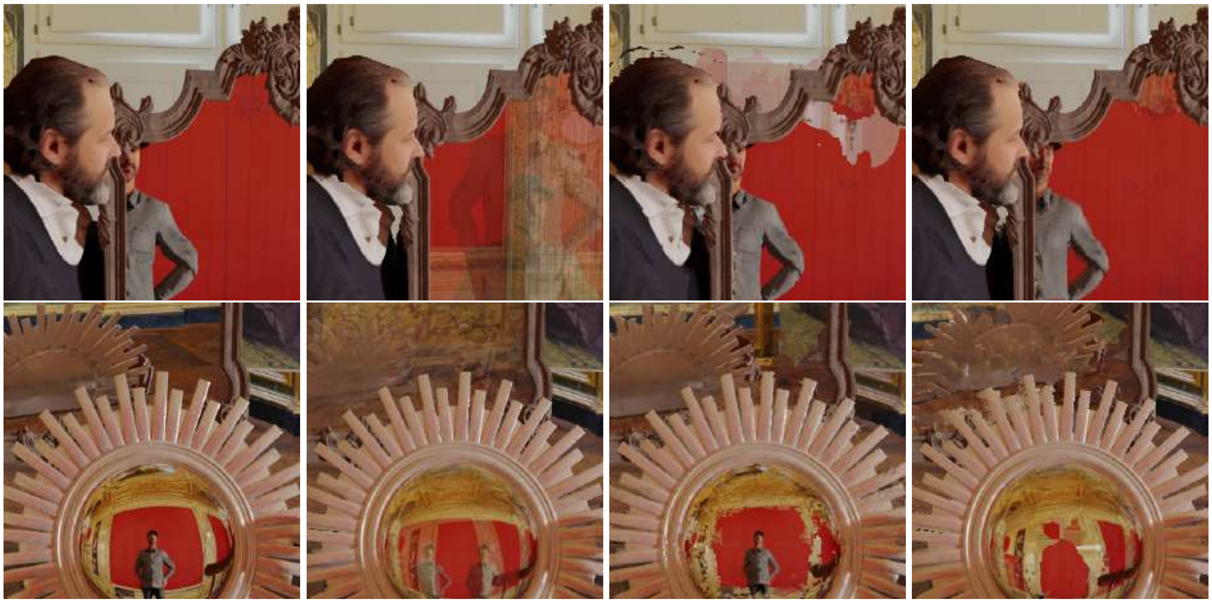


(a) Ground truth depth



(b) Estimated depth map

Fig. 3: Depth maps for mirror sequence. For non-Lambertian objects, the geometry (a) does not correspond to the perceived disparity (b).



(a) Reference (b) Ground truth depth (c) Estimated depth [28] (d) Polynomial map [24]

Fig. 4: Zoomed details on a synthesized view of the dataset Mirror.

face, as illustrated in Figure 3. Hence, the ground truth depth map, indicating the depth of the mirror surface, induces ghosting artifacts (Figure 4). Additionally, the reflected objects should be limited to the limits of the mirror frame, which is not the case with the estimated depth maps, causing bleeding artifacts around the mirror's frame. Using apparent disparity (polynomial map with one coefficient) with knowledge of the physical depth of the mirror solves this problem. Ghosting artifacts also appear in the small curved mirror for ground truth depth maps. Polynomial maps have failed on the small curved mirror.

Magritte sequences We observe ghosting when using the ground truth depth maps, because the apparent disparity is too different from the inverse of the depth (similarly to Figure 3). Using the estimated depth map gives acceptable results though blurrier than the polynomial depth, computed with four coefficients for the mirror version and eighteen for the transparent version.

5. CONCLUSION

We presented a comparison between three types of input maps for DIBR methods for non-Lambertian objects rendering: geometrical depth, estimated disparity and polynomial pixel displacement. Depending on the type of non-Lambertian object targeted for rendering, the best solution varies: for fully refractive or reflective non-planar objects, a polynomial approximation is best suited. For planar mirrors, classical DIBR with estimated depth maps leads to better results in terms of objective metrics. For semi-transparent and semi-refractive objects, it would be interesting to explore new solutions, able to segment the transparency layers [17] of the non-Lambertian objects.

6. ACKNOWLEDGEMENTS

This work was supported by the Fonds de la Recherche Scientifique - FNRS, Belgium, under Grant n° 33679514, ColibriH. The results presented in this paper have been obtained within the HoviTron activities. The project HoviTron has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement N° 951989.

7. REFERENCES

[1] Christoph Fehn. Depth-image-based rendering (DIBR), compression, and transmission for a new approach on 3D-TV. In *Stereoscopic Displays and Virtual Reality Systems XI*, volume 5291, pages 93–105. International Society for Optics and Photonics, May 2004.

[2] Sarah Fachada, Daniele Bonatto, Arnaud Schenkel, and Gauthier Lafruit. Depth Image Based View Synthesis With Multiple Reference Views For Virtual Reality. In *2018-3DTV-Conference: The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON)*, Stockholm, Sweden, 2018. IEEE.

[3] Bart Kroon. Reference View Synthesizer (RVS) manual [N18068]. *ISO/IEC JTC1/SC29/WG11*, October 2018.

Fig. 5: Zoomed details on a synthesized view of the dataset Magritte (top: Mirror, bottom: Transparent).

[4] Daniele Bonatto, Sarah Fachada, and Gauthier Lafruit. RaViS: Real-time accelerated View Synthesizer for immersive video 6DoF VR. In *Society for Imaging Science and Technology (IS&T) - Electronic Imaging*, Burlingame, USA, January 2020.

[5] Daniele Bonatto, Sarah Fachada, Ségolène Rogge, Adrian Munteanu, and Gauthier Lafruit. Real-Time Depth Video Based Rendering for 6-DoF HMD Navigation and Light Field Displays. *IEEE Access*, October 2021.

[6] Tomoyuki Tezuka, Mehrdad Panahpour Tehrani, Kazuyoshi Suzuki, Keita Takahashi, and Toshiaki Fujii. View synthesis using superpixel based inpainting capable of occlusion handling and hole filling. In *2015 Picture Coding Symposium (PCS)*, pages 124–128. IEEE, 2015.

[7] Shuai Li, Ce Zhu, and Ming-Ting Sun. Hole filling with multiple reference views in DIBR view synthesis. *IEEE Transactions on Multimedia*, 20(8):1948–1959, 2018. Publisher: IEEE.

[8] Theo Thonat, Abdelaziz Djelouah, Fredo Durand, and George Drettakis. Thin Structures in Image Based Rendering. *Eurographics Symposium on Rendering 2018*, page 13, 2018.

[9] Hui-Yu Huang and Shao-Yu Huang. Fast Hole Filling for View Synthesis in Free Viewpoint Video. *Electronics*, 9(6):906, May 2020.

[10] Yu Mao, Gene Cheung, and Yusheng Ji. Image interpolation for DIBR viewsynthesis using graph fourier transform. In *2014 3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON)*, Budapest, Hungary, July 2014. IEEE.

[11] Mehrdad Panahpour Tehrani, Tomoyuki Tezuka, Kazuyoshi Suzuki, Keita Takahashi, and Toshiaki Fujii. Free-viewpoint image synthesis using superpixel segmentation. *APSIPA Transactions on Signal and Information Processing*, 6, 2017.

[12] Adrian Dziembowski and Marek Domański. Adaptive Color Correction In Virtual View Synthesis. In *2018-3DTV-Conference: The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON)*, page 4, Stockholm, 2018. IEEE.

[13] Peter Hedman, Julien Philip, True Price, Jan-Michael Frahm, George Drettakis, and Gabriel Brostow. Deep blending for free-viewpoint image-based rendering. *ACM Transactions on Graphics*, 37(6), December 2018.

[14] Samuel Boivin and Andre Gagalowicz. Image-based rendering of diffuse, specular and glossy surfaces from a single image. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques - SIGGRAPH '01*, pages 107–116, Not Known, 2001. ACM Press.

[15] Gerrit Lochmann, Bernhard Reinert, Tobias Ritschel, Stefan Müller, and Hans-Peter Seidel. Real-time Reflec-

tive and Refractive Novel-view Synthesis. *VMV*, pages 9–16, 2014.

[16] Richard Szeliski. Image-Based Rendering for Scenes with Reflections. *ACM Transactions on Graphics (TOG)*, 31(4):10, 2012.

[17] Sven Wanner and Bastian Goldluecke. Reconstructing Reflective and Transparent Surfaces from Epipolar Plane Images. In Joachim Weickert, Matthias Hein, and Bernt Schiele, editors, *Pattern Recognition*, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.

[18] Gauthier Lafruit, Mehrdad Panahpour Tehrani, Renaud Doré, Gun Bang, Bart Kroon, and Joël Jung. Call for MPEG-I Visual Test Materials [N18788]. *ISO/IEC JTC1/SC29/WG11*, October 2019.

[19] Renaud Doré and Gerard Briand. [MPEG-I] Interdigital Mirror Content Proposal for advanced MIV investigations on reflection [m55710]. *ISO/IEC JTC1/SC29/WG11*, December 2020.

[20] Sarah Fachada, Bonatto, Yupeng Xie, Mehrdad Teratani, and Gauthier Lafruit. Two Non-Lambertian Test Materials and View Synthesis Performance [m56450]. *ISO/IEC JTC1/SC29/WG11*, April 2021.

[21] Sarah Fachada, Daniele Bonatto, Mehrdad Teratani, and Gauthier Lafruit. Magritte sphere video. `https://doi.org/10.5281/zenodo.5048270`, June 2021.

[22] Renaud Doré, Gerard Briand, and Franck Thudor. [MIV] New Cadillac content proposal for advanced MIV v2 investigations [m57186]. *ISO/IEC JTC1/SC29/WG11*, July 2021.

[23] Sarah Fachada, Daniele Bonatto, Mehrdad Teratani, and Gauthier Lafruit. Light field rendering for non-Lambertian objects. *Electronic Imaging*, pages 054–1–054–6, 2021.

[24] Sarah Fachada, Daniele Bonatto, Mehrdad Teratani, and Gauthier Lafruit. Polynomial Image-Based Rendering for non-Lambertian Objects. In *Visual Communication and Image Processing 2021*, page 5, Munich, Germany, December 2021.

[25] Yupeng Xie, Sarah Fachada, Daniele Bonatto, Mehrdad Teratani, and Gauthier Lafruit. ULB's Report on EE3: Coding and rendering of non-Lambertian content [m57836]. *ISO/IEC JTC1/SC29/WG11*, October 2021.

[26] Maria Santamaria, Patrice Rondao Alface, and Vinod Kumar Malamal Vadakital. Nokia Results for Future MIV Exploration Experiments [m57832]. *ISO/IEC JTC1/SC29/WG11*, October 2021.

[27] Yichao Xu, Hajime Nagahara, Atsushi Shimada, and Rin-ichiro Taniguchi. Transcut: Transparent object segmentation from a light-field image. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3442–3450, 2015.

[28] Dawid Mieloch and Adrian Dziembowski. Proposal of IVDE 3.0 [m55751]. *ISO/IEC JTC1/SC29/WG11*, December 2020.

[29] Segolene Rogge, Daniele Bonatto, Jaime Sancho, Ruben Salvador, Eduardo Juarez, Adrian Munteanu, and Gauthier Lafruit. MPEG-I Depth Estimation Reference Software. In *2019 International Conference on 3D Immersion (IC3D)*, Brussels, Belgium, December 2019. IEEE.

[30] Sarah Fachada, Bonatto, Yupeng Xie, Mehrdad Teratani, and Gauthier Lafruit. View Synthesis Experiment using Magritte Non-Lambertian Dataset [M57104]. *ISO/IEC JTC1/SC29/WG11*, June 2021.

[31] Adrian Dziembowski. Software manual of IV-PSNR for Immersive Video [N18709]. *ISO/IEC JTC1/SC29/WG11*, July 2019.