ULB

ECOLE
POLYTECHNIQUE
DE BRUXELLES

Photo-realistic depth image-based view synthesis with multi-input modality

Thesis presented by Sarah FACHADA
in fulfilment of the requirements of the PhD Degree in Engineering and Technology ("Docteur en Sciences de l'Ingénieur et Technologie")
Année académique 2022-2023

Supervisor: Professor Gauthier LAFRUIT
Co-supervisor: Professor Mehrdad TERATANI
Laboratory of Image Synthesis and Analysis

Thesis jury :

Olivier DEBEIR (Université libre de Bruxelles, Chair)
Dragomir MILOJEVIC (Université libre de Bruxelles, Secretary)
Maja Krivokuća (InterDigital, France)
Adrian MUNTEANU (Vrije Universiteit Brussel)
Patrice RONDAO ALFACE (Nokia)
Lu YU (Zhejiang University, China)

fnrs
LA LIBERTÉ DE CHERCHER

“Il est essentiel de ne jamais cesser de découvrir.
En voyageant ou en restant immobile,
en échangeant ou en se taisant,
en réfléchissant ou en innovant. ”

Pierre Bottero

Abstract

Who has never immersed themselves in the memories of a photo album or the universe rendered by a movie? Cameras have been invented to record a part of the world at a precise moment for recollection, information or entertainment purposes. However, beside the ever-increasing quality of the captured images, only one viewpoint, holding in a flat image, is captured. In order to increase the immersion, new methods have emerged to reconstruct and render the three dimensions of a scene, allowing the user to better understand, measure or simply live the captured moment.

Due to the public and industry's interest for immersive representations of the real-world (3D cinema, metaverse, virtual reality,...), the MPEG-I group (subgroup of MPEG for Immersive video coding) launched a vast work for standardization of immersive video coding. Research on view synthesis, the technique to create new viewpoints from a given set of input images, has been tackled by companies such as Google and NVIDIA, obtaining incredibly photo-realistic results. However, these methods are not yet accessible at large, due to the acquisition setups, the reconstruction process, the hardware cost or the processing time needed to render new viewpoints.

In this manuscript, we present a view synthesis method relying on the principle of depth image-based rendering: given a set of real-world images representing a scene, along with some geometry information represented by depth maps, we reconstruct new images taken from any location in the scene. We aim to reach photo-realistic results, wide navigation range, real-time processing on midrange material and multi-input modality.

We demonstrate the quality of the image synthesized by this method compared to other state-of-the-art view synthesis approaches. Thanks to a GPU implementation, we reach real-time results and high visual quality with less than ten input images. Thus, we prove that our method is suitable not only for real-time virtual reality applications, but

also for holography and 3D displays.

We also address the challenge of rendering non-Lambertian objects (e.g. mirrors, glasses, transparent liquids...). While omnipresent, they are left aside or barely mentioned in most of the view synthesis methods. Indeed, their particular interactions with light make them violate the Lambertian assumption, which is the implicit basis of most of the reconstruction and rendering methods. However, to reach a faithful rendering of the scene, and not only a plausible approximation of the scene, we extended the depth image-based rendering paradigm to reproduce their changing appearance.

Eventually, we studied the particular case of plenoptic cameras. Those devices, mimicking multi-faceted insect eyes, capture in one shot several viewpoints of the scene: exactly what is needed to overcome the gap between a flat photograph and the 3D scene in which we can immerse. However, the view synthesis with plenoptic camera, and all the related methods to make view synthesis possible (calibration, geometry estimation), are still at their first faltering steps, compared to analogous methods with regular imaging devices. As an invitation to take a journey to the plenoptic world, we explored the calibration of the plenoptic cameras, for single and multi-view datasets. We contributed to the development of view synthesis of new images within a small range. And finally, we extended our software to free viewpoint view synthesis using plenoptic cameras with calibrated parameters.

This thesis contributes to the development of photo-realistic, multi-input view synthesis. It extends the understanding of depth image-based rendering by pushing its limits beyond non-Lambertian scenes and plenoptic cameras. Furthermore, it makes the method more accessible at large through the release of open-source datasets and software, the low requirements of the proposed approaches (pattern-free calibration, portable methods) and the extensive comparisons with other state-of-the-art methods. In this sense, it opens doors to fascinating new challenges and applications in the domain of view synthesis.

Remerciements

De nombreuses personnes ont contribué à l'aboutissement de cette thèse durant les cinq dernières années et je tiens à les remercier ici.

En premier lieu, merci à mon promoteur, Pr. Gauthier Lafruit et mon co-promoteur, Pr. Mehrdad Teratani, pour leur encadrement, leur soutien, leurs conseils et leurs encouragements. Merci d'avoir cru en moi dès le début.

Merci à Pr. Olivier Debeir, Pr. Dragomir Milojevic, Pr. Maja Krivokuća, Pr. Adrian Munteanu, Dr. Patrice Randao Alface et Pr. Lu Yu qui m'ont fait l'honneur de faire partie du jury, pour leur lecture de ce manuscrit et leurs retours.

Merci au FNRS pour sa confiance en moi, merci de m'avoir permis de chercher les solutions à des problèmes qui me fascinaient depuis si longtemps. Merci au projet européen HoviTron pour son support aux projets menés à bien pendant cette thèse. Merci à Innoviris de m'avoir permis de rejoindre le projet 3DLicorneA.

Travailler au LISA, c'est aussi cotoyer des personnes extraordinaires, que ce soit face à un tableau et craie à la main, lors d'un repas, d'un séminaire ou d'un café. Merci à Pr. Christine Decaestecker et Pr. Olivier Debeir pour le soutien et leur enthousiasme. Merci à Arlette Grave pour son support administratif. Merci à Daniele Bonatto d'avoir ensoleillé mes journées à travers les nombreuses discussions que nous avons eues ensemble. Merci à Rudy Ercerk, Corentin Martens, Adrien Foucart, Thomas Vandamme, Laurie Van Bogaert, Hamed Razavi, Eline Soetens et Armand Losfeld et tant d'autres.

Merci aux professeurs qui m'ont guidée vers les sciences depuis le lycée, en particulier Madame Altschuh et Pr. Tatatiana Beliaeva.

Merci à mes amis, toujours présents malgré les kilomètres, Mercedes, Laura, Antoine, Tito, Clément, François et bien d'autres.

Merci à ma famille qui m'entoure d'amour : ma maman, Martine Dury, qui a toujours été une base arrière qui pousse en avant. Merci à Lili, ma petite *speur*, à mes enfants, Léonie et Raphaël, pour leur joyeuse contribution. Finalement, merci à mon compagnon, 'God', pour sa patience et son soutien.

Acknowledgements

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

This work was supported in part by the Fonds de la Recherche Scientifique - FNRS, Belgium, under Grant n° 33679514, ColibriH.

This work was supported in part by the HoviTron project that received funding from the European Union's Horizon 2020 research and innovation program under grant agreement N° 951989 (`https://www.hovitron.eu/`).

This work was supported in part by the FER 2021 project, Grant N° 1060H000066-FAISAN.

This work was supported in part by the Emile DEFAY 2021 project, Grant N° 4R00H000236-Emile DEFAY, Belgium.

Contents

List of Figures

List of Tables

List of Abbreviations

2D two-dimensional. 17, 34, 94–96, 100, 181

3D three-dimensional. 1, 2, 5–8, 15, 22, 26, 33–36, 39, 47, 54, 55, 60, 61, 83, 85–88, 94, 96, 97, 101, 113–115, 117, 130, 131, 144

4D four-dimensional. 15, 17, 21, 33, 93, 94, 97, 100

CPU central processing unit. 46, 55, 81

DERS Depth Estimation Reference Software. 45, 62, 72, 73, 77, 103, 108

DIBR depth image-based rendering. 5, 7, 8, 25, 26, 36–40, 46, 48, 54, 60, 72, 78–80, 83, 85, 91–93, 100, 103, 104, 108, 109, 144, 145

EPI epipolar plane image. 20, 21, 33, 34, 86, 88, 90, 93–95, 101, 104, 107, 144, 181

FPS frames per second. 55, 78, 81

GCD global color difference. 28

GLC general linear camera. 5, 7, 19, 20, 22, 56

GPU graphic processing unit. 39, 46, 55, 56, 63, 81, 108, 109, 143

HMD head-mounted display. 78

IBR image-based rendering. 6, 7, 35, 42, 61, 85, 87

IV-PSNR Immersive Video PSNR. 28, 63, 77, 174, 177

LLFF Local Light Field Fusion. 41, 62, 88, 104, 107, 108, 145

LLMV Lenslet to MultiView. 114, 116, 127, 128, 134–136, 141, 180

LPIPS Learned Perceptual Image Patch Similarity. 30, 63, 72, 104, 174, 177

MIV MPEG Immersive Video. 8, 36, 39, 144

MLA microlens array. 5, 22, 23, 114, 115, 120, 121, 124, 134, 135

MPEG Moving Picture Experts Group. 8, 121, 141

MPEG-I MPEG-Immersive. 8, 31, 36, 39, 43, 45, 46, 62, 81, 92, 119, 143, 144, 177

MPI multi-plane image. 5, 40–42, 72, 88, 90, 107, 108, 144, 145

MS-SSIM Multi-Scale SSIM. 29, 63, 72, 177

MSE mean square error. 27, 28

NeRF neural radiance field. 35, 42, 43, 62, 63, 72, 88, 90, 91, 103, 104, 107, 108

NGP neural graphics primitives. 43

PSNR Peak Signal to Noise Ratio. 27, 28, 63, 77, 104, 107, 177

RDE Reference Depth Estimation software. 45, 175, 177

RGB red green blue. 27, 61, 90

RGBD RGB+depth. 61, 79, 81

RLC Raytrix Lenslet Content Convertor. 8, 114, 117, 119, 127, 128, 133–136, 141, 180

RPVC Reference Plenoptic Virtual Camera Calibrator. 8, 179

RVS Reference View Synthesizer. 8, 31, 39, 46, 54–56, 62, 63, 72, 73, 77, 79–81, 83, 88, 103, 104, 107, 108, 114, 130, 135, 136, 141, 143–145, 177, 180

SfM structure-from-motion. 5, 6, 8, 34–36, 39, 46, 113, 121–123, 125, 133

SSIM Structural Similarity. 28, 29

SVD singular value decomposition. 98

TMIV Test Model of MPEG Immersive Video. 8, 36, 39, 62, 63, 72, 81, 177

VR virtual reality. 1, 7, 9, 32, 54

VSRS View Synthesis Reference Software. 8

WVS View Weighting Synthesizer. 39

YUV luminance+chrominance (blue and red). 27