Warsaw University of Technology

FACULTY OF ELECTRONICS AND INFORMATION TECHNOLOGY

PhD Thesis

in the discipline of Information and Communication Technology

Few-Shot Human Neural Rendering with Partial Information

Kacper Kania, M.Sc.

supervisor
Tomasz Trzciński, Prof. PhD DSc.

assistant supervisor
Marek Kowalski, PhD DSc.

WARSZAWA 2025

Acknowledgements

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

Abstract

This thesis is a series of publications that introduce novel methods for human neural rendering using limited information, focusing on Neural Radiance Fields (NeRFs) and 3D Gaussian Splatting (3DGS). It explores how these models construct 3D representations from 2D images and demonstrates ways to condition these representations for generating high-quality human renderings. We propose techniques that use simple, interpretable inputs derived from sparse training data and extends these methods to perform effectively in few-shot learning scenarios.

We begin by examining the field of neural radiance fields, addressing limitations in existing approaches and presenting contributions to controllable radiance fields. By incorporating partial and sparse data during training, it leverages the smoothness of neural networks to produce controllable, high-quality human images.

To tackle the reliance on extensive, high-quality data annotations from multi-view videos, we introduce a new method for training neural radiance fields in few-shot, multi-view settings. This approach learns internal deformation templates, which blend smoothly during inference, significantly improving image quality compared to existing baselines and enabling effective human rendering from limited input images.

The work also addresses the need for adaptable computational efficiency during inference. It proposes a fine-to-coarse learning strategy for 3D Gaussian Splatting, which upscales a latent 2D grid that stores Gaussian representations. This strategy achieves competitive results while allowing deployment on various computational devices with minimal quality loss.

In addition, we develop a novel model for controlling radiance fields through environmental lighting. By incorporating precomputed radiance transfer, this model enables physically plausible scene relighting and provides users with intuitive control over lighting in reconstructed scenes.

This research advances the state of the art in controllable neural radiance fields and expands their application to few-shot learning scenarios. These innovations enhance the possibilities for human rendering from limited information and open new directions for future research in the field.

Keywords: Neural Rendering, Neural Radiance Fields, Few-Shot Learning, Human Rendering, Partial Information, Gaussian Splatting

Streszczenie

To jest streszczenie. To jest trochę za krótkie, jako że powinno zająć całą stronę.

Słowa kluczowe: A, B, C

Lay Summary

ok

Publications in this thesis

Title	Authors	Venue	Status
CoNeRF: Controllable Neural Radiance Fields	**Kacper Kania**, Kwang Moo Yi, Marek Kowalski, Tomasz Trzciński, Andrea Tagliasacchi	CVPR 2022	Accepted
BlendFields: Few-Shot Example-Driven Facial Modeling	**Kacper Kania**, Stephan J. Garbin, Andrea Taglisacchi, Virginia Estellers, Kwang Moo Yi, Julien Valentin, Tomasz Trzciński, Marek Kowalski	CVPR 2023	Accepted
LumiGauss: High-Fidelity Outdoor Relighting with 2D Gaussian Splatting	Joanna Kaleta, **Kacper Kania**, Tomasz Trzciński, Marek Kowalski	WACV 2025	Accepted
CLoG: Leveraging UV Space for Continuous Levels of Detail	**Kacper Kania**, Rawal Khirodkar, Shunsuke Saito, Kwang Moo Yi, Julieta Martinez	CVPR 2025	Under Review

Contents

List of Abbreviations and Symbols

List of Figures

List of Tables

π	Stała matematyczna równa stosunkowi obwodu okręgu do jego średnicy
I	Natężenie prądu elektrycznego

Chapter 1

Introduction

With the advent of deep learning, research have been exploring varying ways to apply it to computer graphics. One of the most recent and promising approaches is neural rendering. Neural rendering is a field that combines deep learning and computer graphics to generate realistic images of 3D scenes. The neural radiance field (NeRF) is a popular neural rendering technique that represents a 3D scene as a continuous function that maps 3D coordinates to radiance values. NeRF has shown impressive results in generating photorealistic images of 3D scenes. However, NeRF has limitations in terms of memory and computational requirements, which makes it difficult to scale to large scenes.

To alleviate the problem, Kerbl *et al.* [8] proposed a new technique—3D Gaussian Splatting (3DGS). 3DGS is a neural rendering technique that represents a 3D scene as a set of 3D Gaussian that are splatted to an image space using algorithm proposed by Zwicker *et al.* [21]. In contrast to NeRF, 3DGS is more memory efficient and can be used to render large scenes. It can also render scenes with millions of points in real-time on a single GPU.

In this thesis, we focus on those two milestone techniques in neural rendering and address their fundamental problem—lack of controllability.

1.1 Motivation and challenges

NeRF and 3DGS are both impressive techniques that can generate realistic images. However, a single scene representation needs to be trained on a high-end GPU for hours or even days just to render a novel view at the inference time. However, any type of controllability is difficult to achieve with those models. That includes changing the lighting conditions, subject's attributes or even the scene itself. We see imbuing those models with controllability as a an important step towards making them more useful in practice. Our proposed models are designed to address this issue.

One may ask why the controllability is a feat sought after to be researched. We see the inspiration in how human artists work. Imagine an artist working on 3D game where they

need an asset, like a 3D mesh, to be created. Such a mesh takes much effort since it includes modeling, creating a UV map which can then be textured. After the process is finish, the artist's supervisor may task him to change the model to some extent which requires the artist to redo all the effort again. Such a process is not limited to 3D assets as meshes and could be applied to 3DGS or NeRF. However, 3DGS and NeRFs are volumetric in nature. Our exploited and well-established practices no longer apply to them since volumetric representations do not have the underlying surface representation. For that reason, we see a couple of avenues which we explore in this thesis.

Firstly, Park *et al.* [15] proposed NeRFies, a model that creates a volumetric representation of a person from a self-captured sequence with a phone camera. Since the inception of NeRFs [12], it was among the first works the achieved such a high quality of reconstructions from a casual videos. In its primal form, NeRFies were unable to control the avatar in any other way than by a linear interpolation of latent embeddings that embedded the video's time dimension. The follow-up work, HyperNeRF [14] handles this issue by projecting the learnable embeddings with D onto a lower-dimensional space $\mathbb{R}^d$ where $d \ll D$. After the assumption that the $d=2$ is enough to explain the sequence variability, that projected embedding becomes a 2-dimensional space that can be traversed in an interpretable way. However, that space is not intuitive since the projection is a non-linear operation and one cannot predict how values affect the results. To mitigate that issue, we propose to leverage smoothness of Multilayer Perceptrons (MLPs) [15, 17] to constrain the projection via sparse supervision. We realize our approach as a weakly-supervised MLP that out of many images from the sequence (we assume at least 100 frames in our work) only a few are provided with a coarse annotation. Such annotations denote what values a chosen attribute takes and where its effect spans in the image space. We show that our method, which we dubbed CoNeRF [6] and published at the CVPR 2022 conference, imbues NeRFs with a flexible editability feature without the lose of the rendering quality.

Secondly, approaches such as CoNeRF [6], EditNeRF [10] or FigNeRF [18] focus solely on static elements of the scene, hence their controllability is limited to changing colors or textures in general. HyperNeRF [14] arises as a potential solution due to its ability to model object deformations. However, our initial experiments showed that those changes cannot handle motions that affect a subject globally, *e.g.*, jumping jacks performed by a person. To solve the issue, Fang *et al.* [2] proposes to model the deformation via a multi-scale voxel structure which works well in the synthetic setting, such as the one proposed by Pumarola *et al.* [16].

There exists a plethora of works that approach the problem from the another angle—instead of modeling the motion purely from data, they use a template model in the form of a 3D mesh to canonicalize deformed points [20]. Such methods rely on the accuracy of the *registration*, *i.e.*, fitting the template mesh to subject. Since the registration methods [3, 19] are imperfect estimators, they inherently contain registration errors. Those deviations are exacerbated by learnable radiance field models which assume a perfectly calibrated scene. The authors of those approaches usually mitigate the issue with additional latent space [5, 6, 11] that requires

thousands of video frames to learn an avatar of high-fidelity that reacts correctly to deformations such as wrinkles on the forehead. At the same time, performing the registration on the large scale is costly [1]. In this thesis, we seek a remedy for those obstacles. We propose a method that is data-efficient, easy to improve with a minimal user input and can model realistic deformation dependent changes in the subject. Inspired by classical methods in character texturing [13] and motion modeling [9], we propose BlendFields [7]. We build on VolTeMorph's [4] approach to point canonicalization to provide a data-efficient way to control the character. We further introduce a physically-based mixture of predefined, learned from data wrinkle templates that represent expression-dependent skin deformations. Our proposed was acknowledged by the reviewers and was accepted to the CVPR 2023 conference.

Thirdly, . ..

1.2 Research objectives

1.3 Contributions

1.4 Thesis outline

1.5 Publications not included in the thesis

Chapter 2

Background

2.1 Neural Rendering

2.2 Neural Radiance Field

2.3 3D Gaussian Splatting

Chapter 3

Final remarks and discussion

3.1 Conclusions

3.2 Future work

Bibliography

[1] Cao, C. *et al.*, „Authentic Volumetric Avatars from a Phone Scan," *ToG*, vol. 41, no. 4, pp. 1–19, 2022 (cit. on p. 5).

[2] Fang, J. *et al.*, „Fast Dynamic Radiance Fields with Time-Aware Neural Voxels," in *SIGGRAPH Asia 2022 Conference Papers*, 2022, pp. 1–9 (cit. on p. 4).

[3] Feng, Y., Feng, H., Black, M. J., and Bolkart, T., „Learning an Animatable Detailed 3D Face Model from In-The-Wild Images," *ToG*, vol. 40, no. 4, pp. 1–13, 2021 (cit. on p. 4).

[4] Garbin, S. J. *et al.*, „VolTeMorph: Real-time, Controllable and Generalizable Animation of Volumetric Representations," in *CGS*, Wiley Online Library, vol. 43, 2024, e15117 (cit. on p. 5).

[5] Grassal, P.-W., Prinzler, M., Leistner, T., Rother, C., Nießner, M., and Thies, J., „Neural Head Avatars From Monocular RGB Videos," in *CVPR*, 2022, pp. 18 653–18 664 (cit. on p. 4).

[6] Kania, K., Yi, K. M., Kowalski, M., Trzciński, T., and Tagliasacchi, A., „CoNeRF: Controllable Neural Radiance Fields," in *CVPR*, 2022 (cit. on p. 4).

[7] Kania, K. *et al.*, „BlendFields: Few-Shot Example-Driven Facial Modeling," in *CVPR*, 2023 (cit. on p. 5).

[8] Kerbl, B., Kopanas, G., Leimkühler, T., and Drettakis, G., „3D Gaussian Splatting for Real-Time Radiance Field Rendering," *TOG*, vol. 42, no. 4, pp. 139–1, 2023 (cit. on p. 3).

[9] Lewis, J. P., Anjyo, K., Rhee, T., Zhang, M., Pighin, F., and Deng, Z., „Practice and Theory of Blendshape Facial Models," in *Eurographics 2014 - State of the Art Reports*, Lefebvre, S. and Spagnuolo, M., Eds., The Eurographics Association, 2014 (cit. on p. 5).

[10] Liu, S., Zhang, X., Zhang, Z., Zhang, R., Zhu, J.-Y., and Russell, B., „Editing Conditional Radiance Fields," in *ICCV*, 2021, pp. 5773–5783 (cit. on p. 4).

[11] Martin-Brualla, R., Radwan, N., Sajjadi, M. S., Barron, J. T., Dosovitskiy, A., and Duckworth, D., „NeRF in the Wild: Neural Radiance Fields for Unconstrained Photo Collections," in *CVPR*, 2021, pp. 7210–7219 (cit. on p. 4).

[12] Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R., and Ng, R., „NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis," *Communications of the ACM*, vol. 65, no. 1, pp. 99–106, 2021 (cit. on p. 4).

[13] Oat, C., „Animated Wrinkle Maps," in *SIGGRAPH*, ser. SIGGRAPH '07, San Diego, California: Association for Computing Machinery, 2007, pp. 33–37, ISBN: 9781450318235 (cit. on p. 5).

[14] Park, K. *et al.*, „HyperNeRF: A Higher-Dimensional Representation for Topologically Varying Neural Radiance Fields," *ToG*, vol. 40, no. 6, 2021 (cit. on p. 4).

[15] Park, K. *et al.*, „Nerfies: Deformable Neural Radiance Fields," in *ICCV*, 2021, pp. 5865–5874 (cit. on p. 4).

[16] Pumarola, A., Corona, E., Pons-Moll, G., and Moreno-Noguer, F., „D-NeRF: Neural Radiance Fields for Dynamic Scenes," in *CVPR*, 2021, pp. 10 318–10 327 (cit. on p. 4).

[17] Tancik, M. *et al.*, „Fourier Features Let Networks Learn High Frequency Functions in Low Dimensional Domains," in *NeurIPS*, 2020 (cit. on p. 4).

[18] Xie, C., Park, K., Martin-Brualla, R., and Brown, M., „FiG-NeRF: Figure-Ground Neural Radiance Fields for 3D Object Category Modelling," 2021 (cit. on p. 4).

[19] Zielonka, W., Bolkart, T., and Thies, J., „Towards Metrical Reconstruction of Human Faces," in *ECCV*, Springer, 2022, pp. 250–269 (cit. on p. 4).

[20] Zielonka, W., Bolkart, T., and Thies, J., „Instant Volumetric Head Avatars," in *CVPR*, 2023, pp. 4574–4584 (cit. on p. 4).

[21] Zwicker, M., Pfister, H., Van Baar, J., and Gross, M., „EWA volume splatting," in *Proceedings Visualization, 2001. VIS'01.*, IEEE, 2001, pp. 29–538 (cit. on p. 3).