HOCHSCHULE FURTWANGEN UNIVERSITY HFU

Free3D: Free-viewpoint 3D Video Creation

Felix Brunn, Jan Christmeier, Nick Philipp Häcker, Patrick Kaserer,
Laura Christin Stempfle, Lukas Willmann, Simon Zakowski, and
Prof. Dr. Uwe Hahne

Furtwangen University, Germany - 2024

Teaser




Inspiration from Arts&Science

Woodkid at ZDF Magazin Royale:

[Video source: ZDF Magazin Royale]

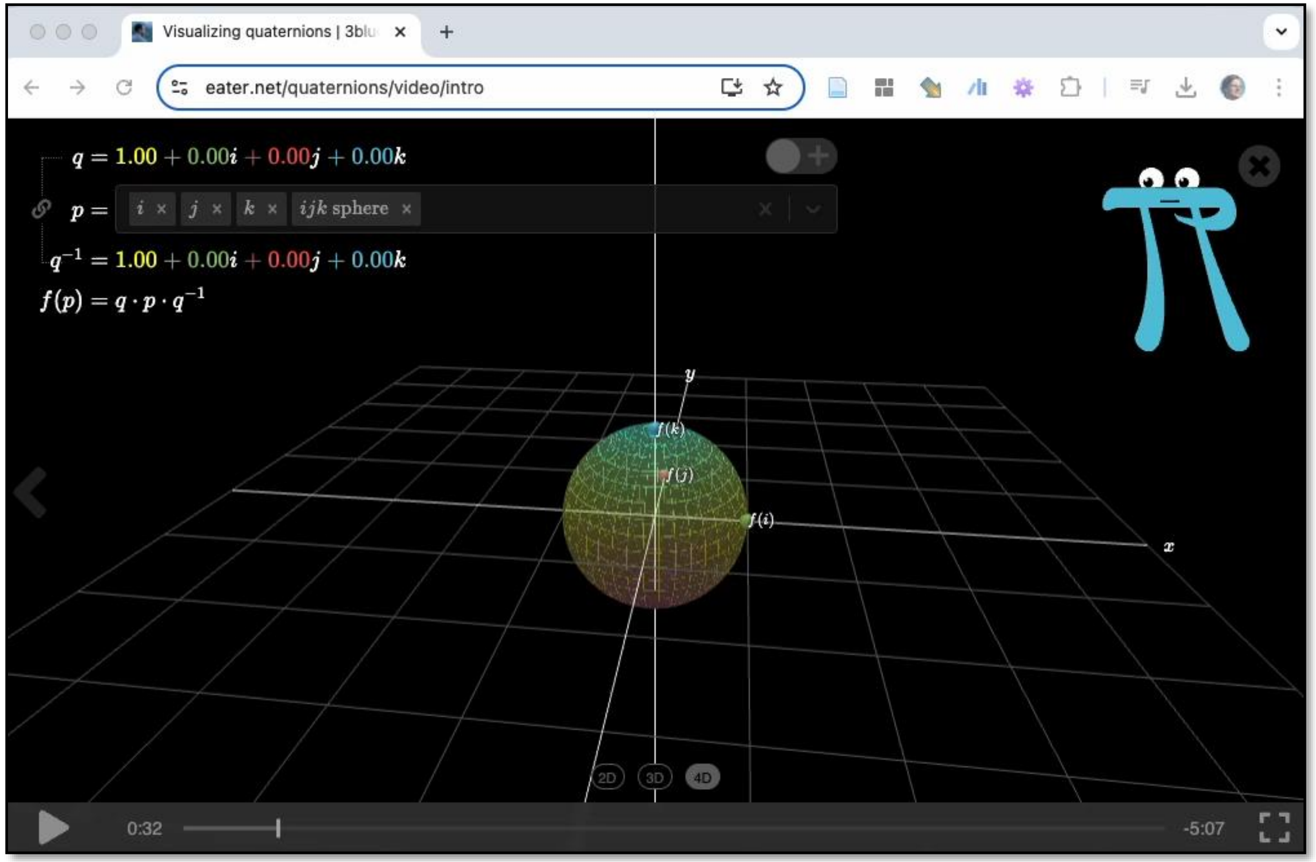
Dynamic 3D Gaussians:

[Video source: Jonathon Luiten]

Inspiration from Education

Visualizing quaternions: An explorable video series

Lessons by Grant Sanderson, Technology by Ben Eater



[Video source: Ben Eater]

Introduction

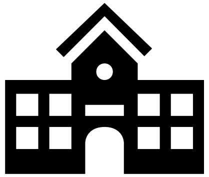
- The **Free3D** project focuses on creating 3D videos using just **three Azure Kinect cameras.**
- Utilizes innovative techniques such as **NeRF (Neural Radiance Fields)** and **3D Gaussian Splatting**.
- The goal is to create high-quality 3D images and videos from any viewpoint, while minimizing setup complexity.

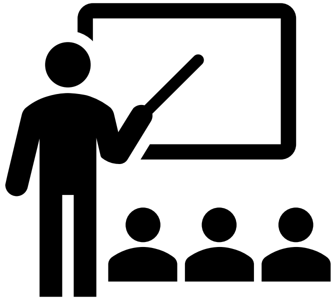
[Image source: Microsoft]

[Video source: Matthew Tancik]

Project Goal: Learn

- This was a **student research project**, aimed at giving students hands-on experience in cutting-edge 3D reconstruction and visualization.
- The main goal was to **learn about modern methods** like NeRF and 3D Gaussian Splatting and **explore their practical applications**.
- Students gain **experience in project work** by combining theory with real-world applications.





Motivation

Science & Technology

- How to implement an **easy-to-setup** capture system with limited resources?
- How to capture a **dynamic 3D scene**?
- How to **store, transfer and visualize** such a dynamic 3D scene?

⇆

Application

- An approach to enhance **online education** by making complex concepts more interactive and easier to learn.
- Explore the potentials to include **Augmented Reality** into classrooms.

Related work

- **VoluProf** - Volumetric professor for omnipresent and user-optimized teaching in mixed reality (BMBF project)

[Image source: HHI & 3it-berlin]

- State of the art in 3D video:
  - [Lin et al. 23] **Im4D**: Combines grid-based and image-based methods.
  - [Xu et al. 24] **4K4D**: Achieves real-time 3D synthesis at 4K resolution.
  - [Luiten et al. 24] **Dynamic 3D Gaussians**: Tracks dynamic scenes without needing correspondence or flow data.

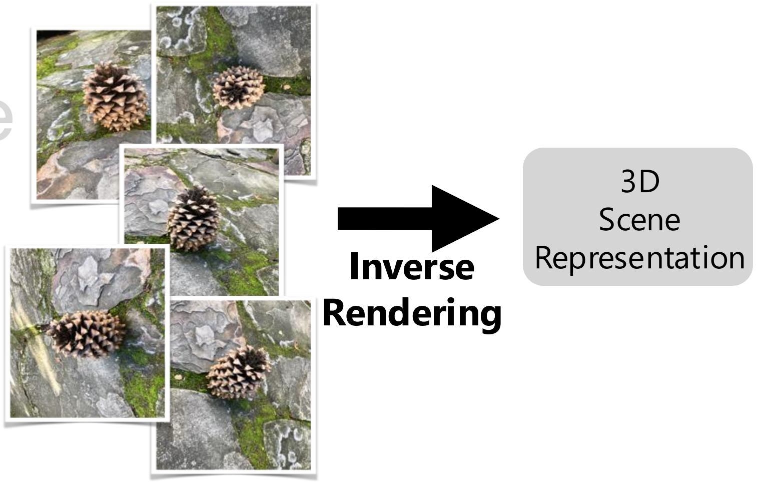
[Video source: Xu et al. 4K4D]

NeRF vs 3DGS

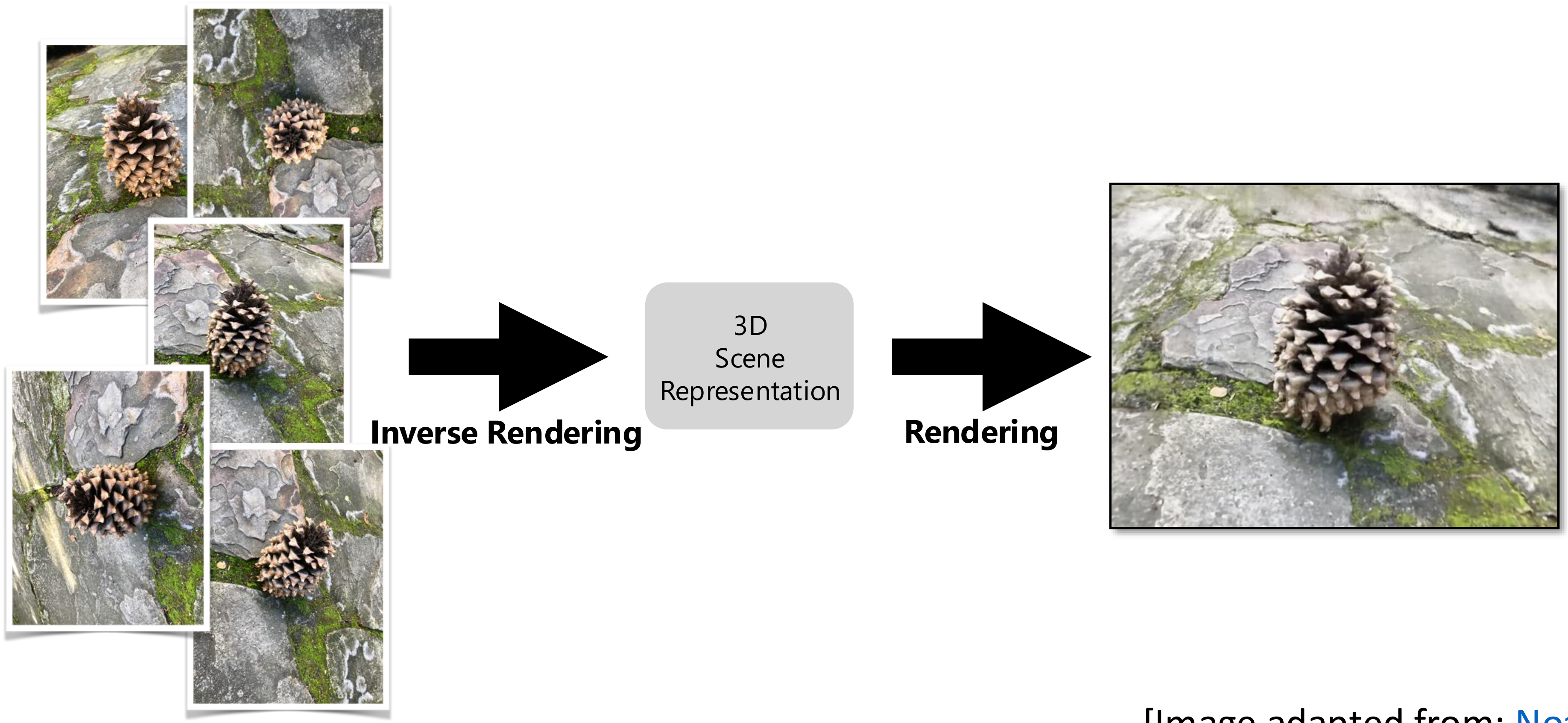
Name	Neural Radiance Fields	3D Gaussian Splatting
Paper title	*NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis*	*3D Gaussian Splatting for Real-Time Radiance Field Rendering*
Conference, Year	ECCV 2020	Siggraph 2023
First Author(s)	Ben Mildenhall, Pratul P. Srinivasan and Matthew Tancik	Bernhard Kerbl
Citation count at Google Scholar **per month** since official publication	**180**	**120**

NeRF vs 3DGS

- Both are **inverse rendering** methods
- Both need the **extrinsic calibration** of the cameras
- Both use the gradient flow from **volume rendering** for optimization
- For NeRF images, each pixel is **computed from one ray**
  - Bottle neck is sampling empty space
- Gaussian Splatting computes the whole image (in tiles)
  - Using **3D Gaussians as primitives**
  - Bottle neck is sorting the primitives



Inverse rendering



[Image adapted from: Noah Snavely]

Inverse rendering

Inverse Rendering

Neural Radiance Field

or

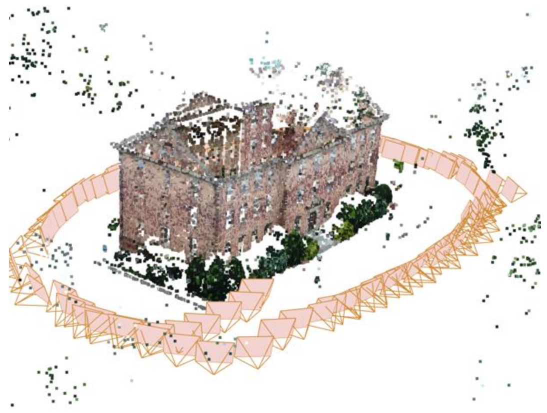
Set of 3D Gaussians

Rendering

[Image adapted from: [Noah Snavely]]

NeRF vs 3DGS

- Both are **inverse rendering** methods
- Both need the **extrinsic calibration** of the cameras
- Both use the gradient flow from **volume rendering** for optimization
- For NeRF images, each pixel is **computed from one ray**
  - Bottle neck is sampling empty space
- Gaussian Splatting computes the whole image (in tiles)
  - Using **3D Gaussians as primitives**
  - Bottle neck is sorting the primitives

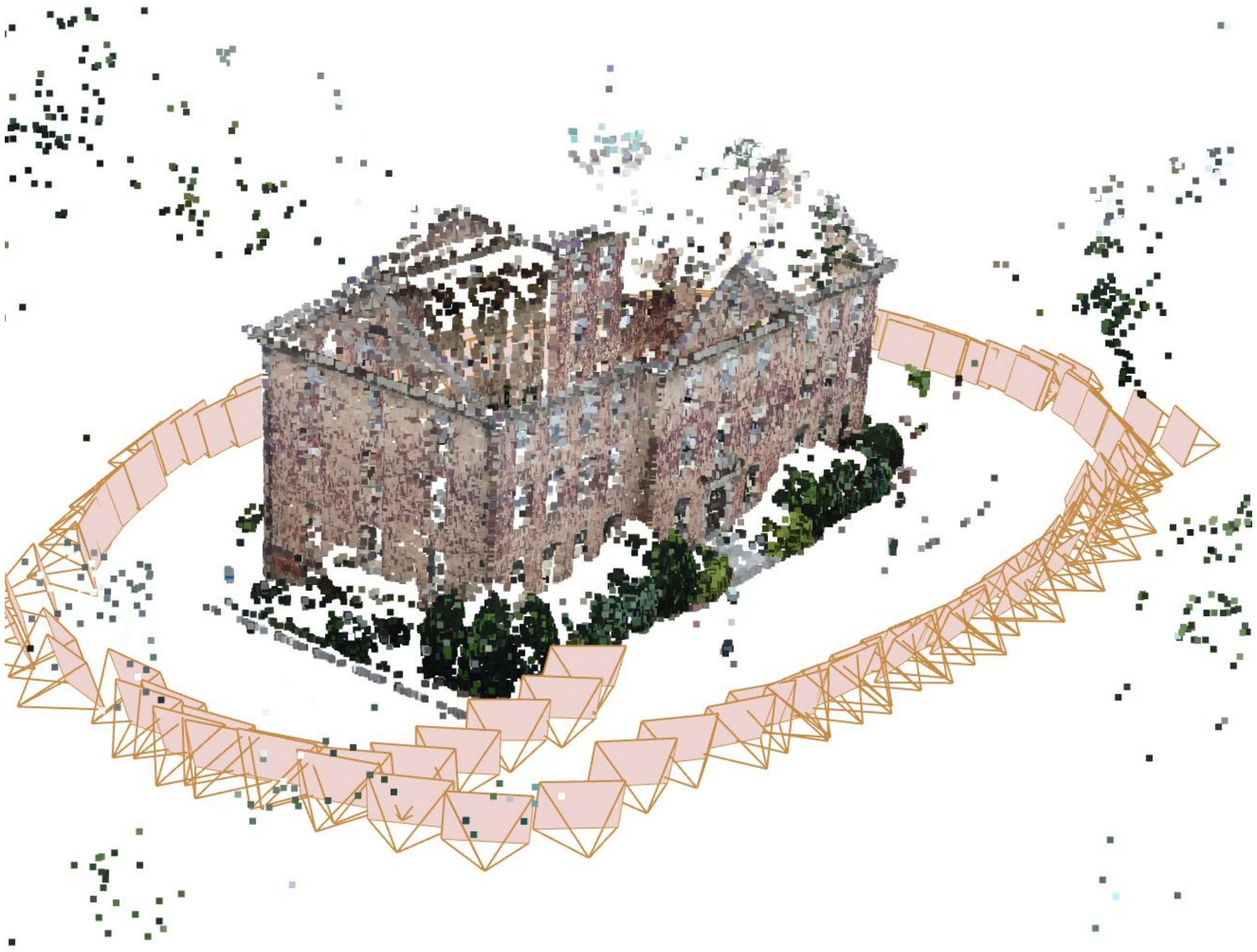


Extrinsic calibration

We need to know from where in space the images have been captured – **extrinsic calibration**.

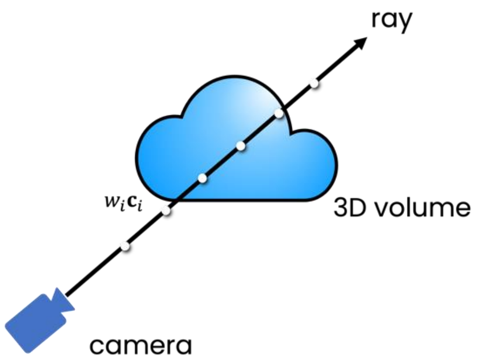
Typically extracted with **structure-from-motion** methods.

Most popular tool: **COLMAP**



NeRF vs 3DGS

- Both are **inverse rendering** methods
- Both need the **extrinsic calibration** of the cameras
- Both use the gradient flow from **volume rendering** for optimization
- For NeRF images, each pixel is **computed from one ray**
  - Bottle neck is sampling empty space
- Gaussian Splatting computes the whole image (in tiles)
  - Using **3D Gaussians as primitives**
  - Bottle neck is sorting the primitives



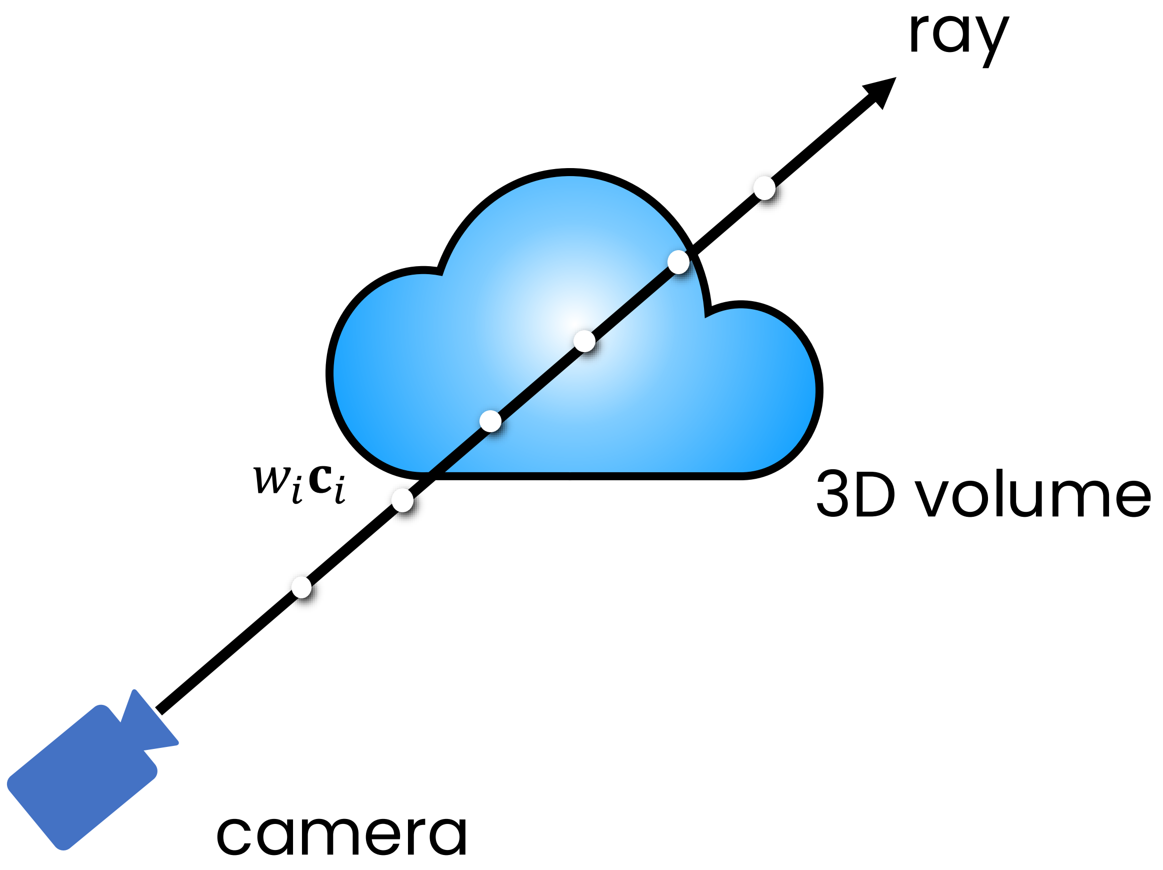
Volume rendering

For each pixel, shoot a ray through the volume that's meant to be visualized.

The resulting pixel color is the integral sum of weighted color values $(w_i \mathbf{c}_i)$.

The opacity (transparency) and visibility (transmittance) defines the weights (w_i).

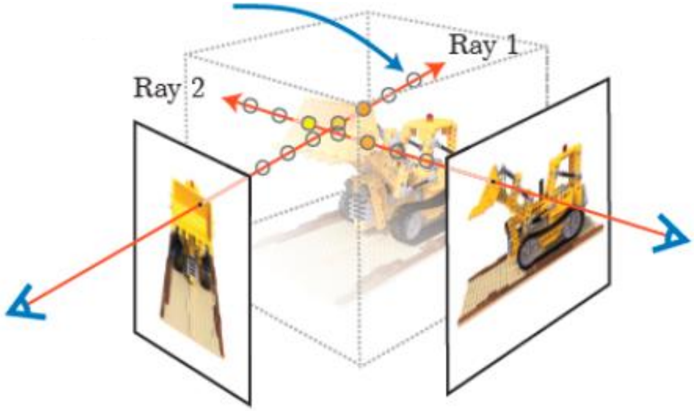
As it is an integral, it is easily differentiable and hence we can use gradient descent to optimize the result.



[Image adapted from: Pratul Srinivasan]

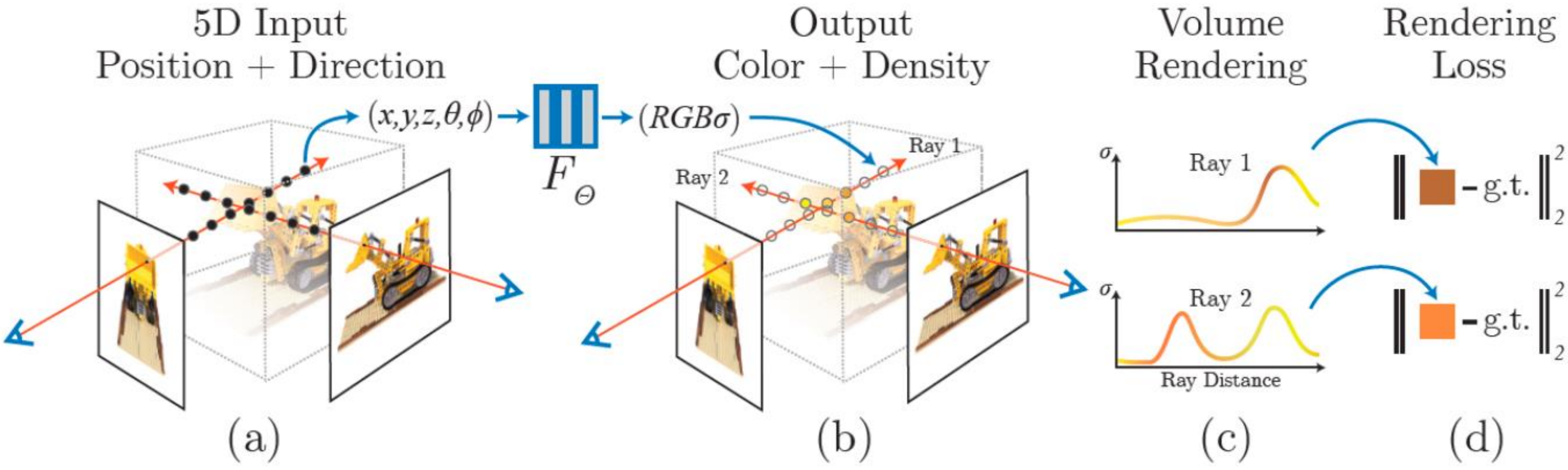
NeRF vs 3DGS

- Both are **inverse rendering** methods
- Both need the **extrinsic calibration** of the cameras
- Both use the gradient flow from **volume rendering** for optimization
- For NeRF images, each pixel is **computed from one ray**
  - Bottle neck is sampling empty space
- Gaussian Splatting computes the whole image (in tiles)
  - Using **3D Gaussians as primitives**
  - Bottle neck is sorting the primitives



Fundamentals: NeRF

- NeRF **samples along the ray** which typically goes through a lot of empty space.



[Image source: Mildenhall et al. 20]

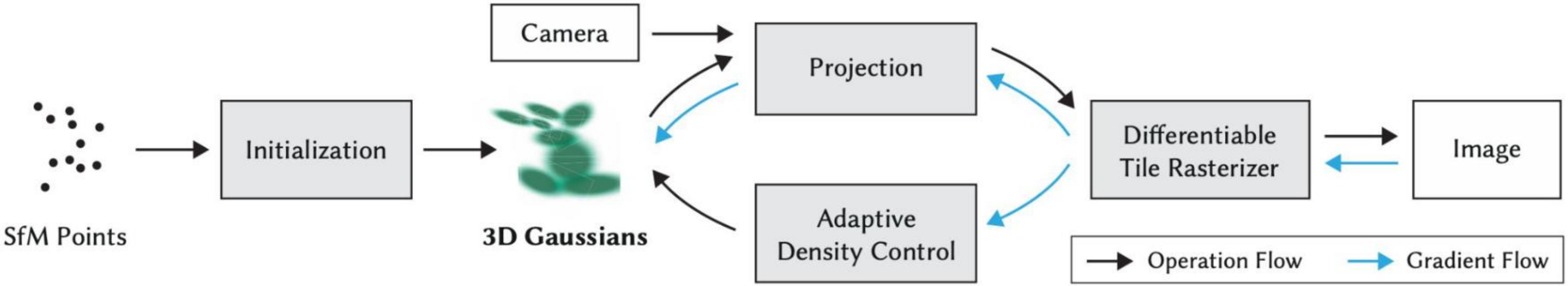
NeRF vs 3DGS

- Both are **inverse rendering** methods
- Both need the **extrinsic calibration** of the cameras
- Both use the gradient flow from **volume rendering** for optimization
- For NeRF images, each pixel is **computed from one ray**
  - Bottle neck is sampling empty space
- Gaussian Splatting computes the whole image (in tiles)
  - Using **3D Gaussians as primitives**
  - Bottle neck is sorting the primitives

3D Gaussians

Fundamentals: 3DGS

- Goal is to preserve continuous properties, but avoid computations in empty space
- Faster in rendering using tile rasterizer
- Explicit **primitives** (3D Gaussians) instead of a neural net



[Image source: Kerbl et al. 23]

NeRF vs 3DGS

NeRF

Gaussian Splatting

NeRF vs Gaussian Splatting

- Both try to **overfit the model** as much as possible

- Both are originally designed for **static scenes**

- Both have a **high computational demand**
  - Development supported by Google and/or NVIDIA

Our approach

- **Combine** NeRF or Gaussian Splatting with dynamic scenes captured by three Azure Kinect (depth) cameras

[Image source: Microsoft]

- Deal with **limited computational** resources to make the creation process accessible.

[Image source: NVIDIA]

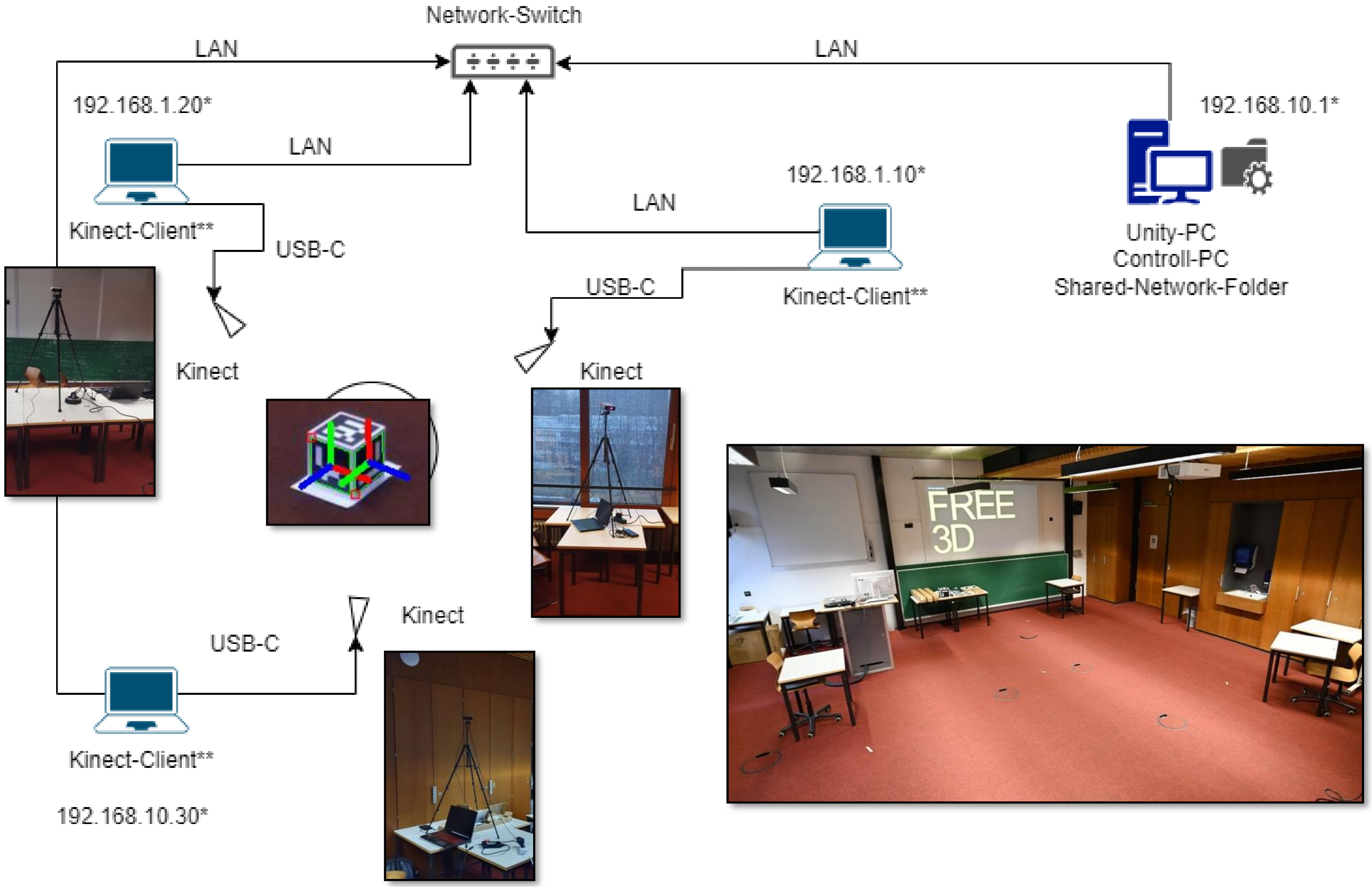
- Create visual **high-quality** results that illustrate the potential of 3D video

Our setup

Three Azure Kinect cameras on tripods in a triangular arrangement.

Camera **synchronization** with audio cables.

Extrinsic calibration is done with ArUco markers on a cube.



Naive approach

Reproducing results from
Dynamic 3D Gaussians
[Luiten et al. 24]

Naive approach

Reproducing results from **Dynamic 3D Gaussians** [Luiten et al. 24]

…but only use three cameras out of 29.

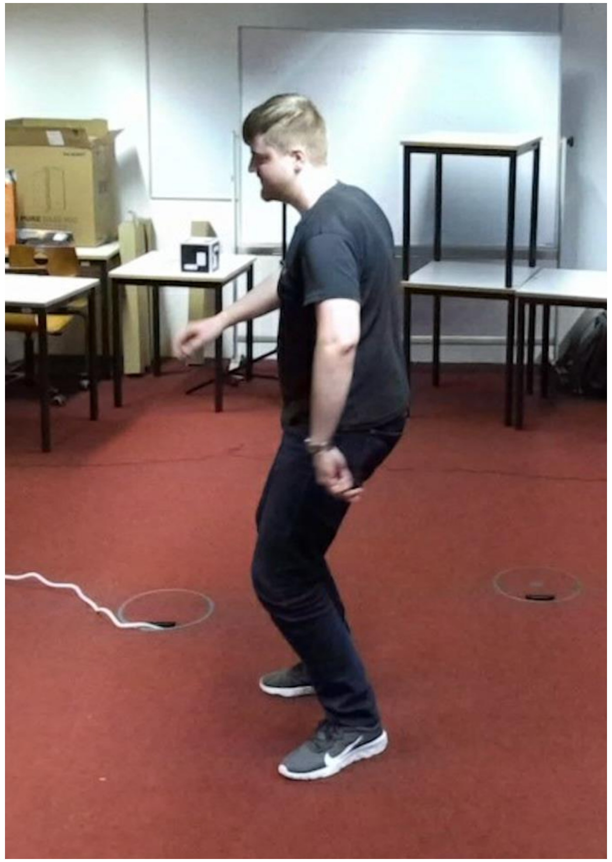
Our solution

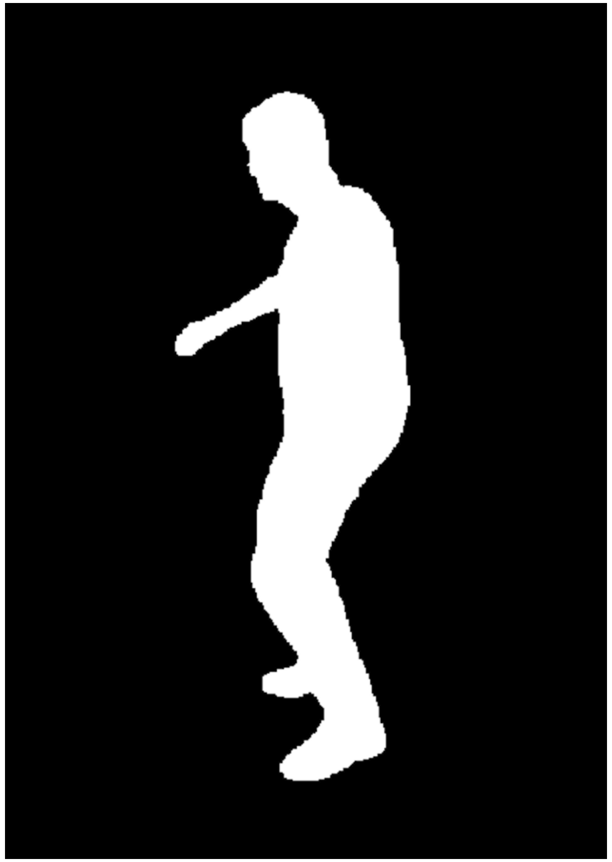
Dynamic Scenes
Fused real-time RGB-D point clouds from the three calibrated Azure Kinect cameras.

Our solution

Dynamic Scenes
Fused real-time RGB-D point clouds from the three calibrated Azure Kinect cameras.

Automated masking
The person is masked using the segmentation model SAM-HQ [Ke et al. 23]





Our solution

Dynamic Scenes

+

Automated masking

+

Static Backgrounds
Generated using NeRF or 3D Gaussian Splatting.

Resulting process

Setup: Place three Azure Kinect cameras on tripods in a triangular arrangement, connect the cables and run the calibration script (15 min).

Capture: The person needs to be in the middle between the cameras and act as intended while an operator starts and stops the recording (up to 30 s).

Processing: The data is transferred to a desktop PC to be processed. The person is masked out, the point clouds are merged and then blended into a pre-recorded NeRF or 3DGS scene (10 min).

Output: A camera path is rendered from the combined scene and the resulting video file can be sent to the user (2-5 min).

a) Setup b) Capture c) Processing d) Output

Final results

Conclusion

- We had a **lot of fun and learned** many things about the new technologies and the challenges of 3D video.
- The Free3D project provides a **lightweight, cost-effective** solution for 3D video creation.
- Combines novel techniques to offer a simplified setup that produces **visually impressive results**.
- A **promising tool for education** and other applications requiring immersive, dynamic visualizations.

No future work

- Students graduate and leave ☹

Future work

- Students do further research on the application of NeRF and 3D Gaussian Splatting in real world scenarios
- Focus on more complex and larger-scale scenes:

We try to reconstruct the whole campus at Furtwangen.

…and we a searching for partners and supporters!

References

- [Kerbl et al. 23] - Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. **3D Gaussian Splatting** for Real-Time Radiance Field Rendering, SIGGRAPH 2023 (ACM Transactions on Graphics)
- [Luiten et al. 24] - Jonathon Luiten, Georgios Kopanas, Bastian Leibe, and Deva Ramanan. **Dynamic 3d gaussians**: Tracking by persistent dynamic view synthesis - International Conference on 3D Vision (3DV) 2024
- [Lin et al. 23] - Haotong Lin, Sida Peng, Zhen Xu, Tao Xie, Xingyi He, Hujun Bao, and Xiaowei Zhou. **Im4D**: High-Fidelity and Real-Time Novel View Synthesis for Dynamic Scenes, SIGGRAPH Asia 2023
- [Mildenhall et al. 20] - Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. **NeRF**: Representing Scenes as Neural Radiance Fields for View Synthesis, ECCV 2020.
- [Xu et al. 24] - Zhen Xu, Sida Peng, Haotong Lin, Guangzhao He, Jiaming Sun, Yujun Shen, Hujun Bao, and Xiaowei Zhou. **4K4D**: Real-Time 4D View Synthesis at 4K Resolution, CVPR 2024.
- [Ke et al. 23] - Lei Ke, Mingqiao Ye, Martin Danelljan, Yifan Liu, Yu-Wing Tai, Chi-Keung Tang, Fisher Yu, **SAM-HQ**: Segment Anything in High Quality, NeurIPS 2023