HSLU Hochschule Luzern

TRANSMIXR

AI-based Volumetric Content Creation for Immersive XR Experiences and Production Workflows

Prof. Dr. Aljosa Smolic

Dept. Computer Science, Immersive Realities Center
21. April 2025

FH Zentralschweiz

HSLU Hochschule Luzern

VOLOGRAMS

Trinity College Dublin
The University of Dublin
V-SENSE

HEAT

FAVORE
FAST VOLUMETRIC RECONSTRUCTION

TRANSMIXR

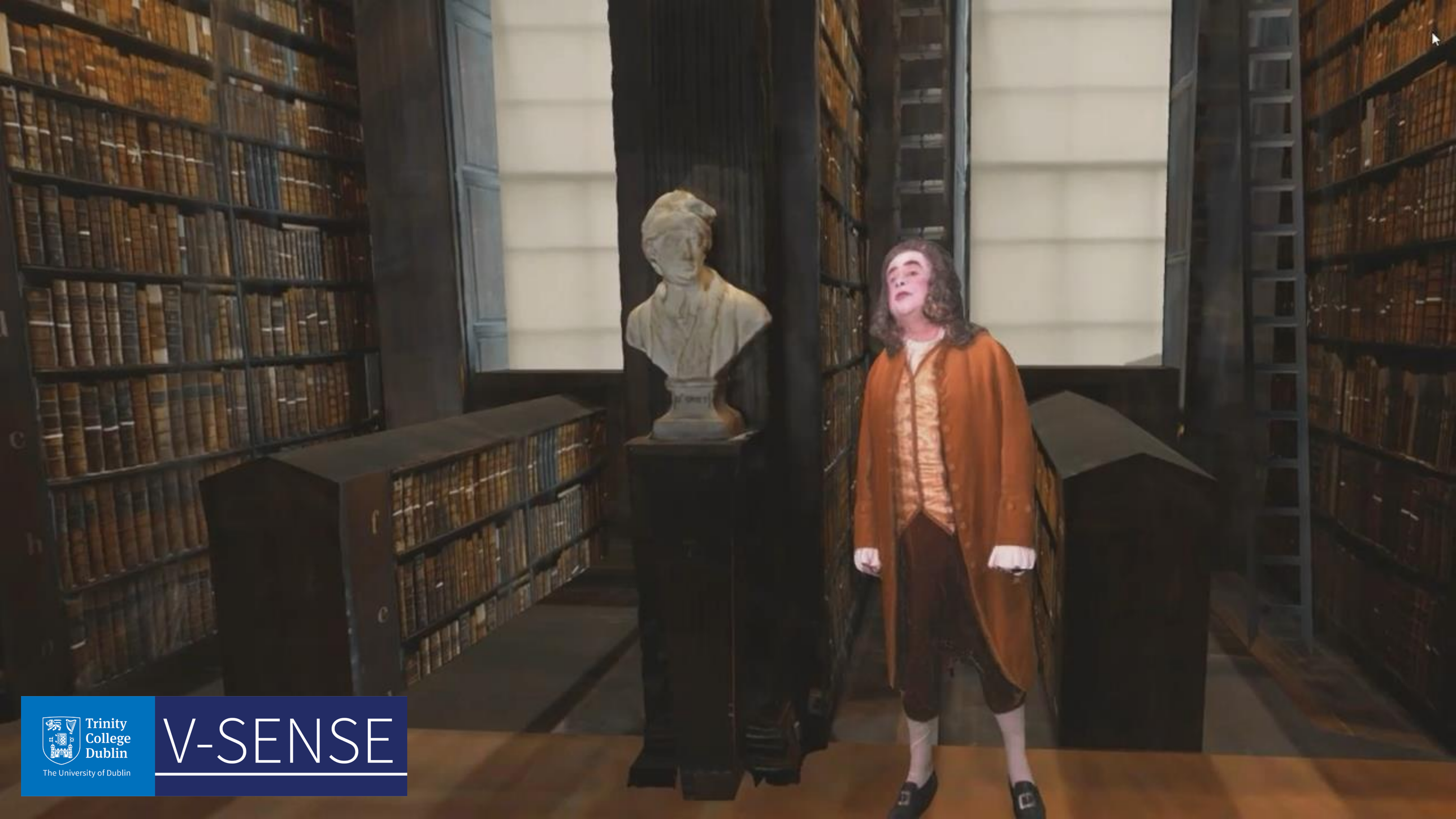
AI-based Volumetric Content Creation for Immersive XR Experiences and Production Workflows

Prof. Dr. Aljosa Smolic

Dept. Computer Science, Immersive Realities Center

21. April 2025

FH Zentralschweiz

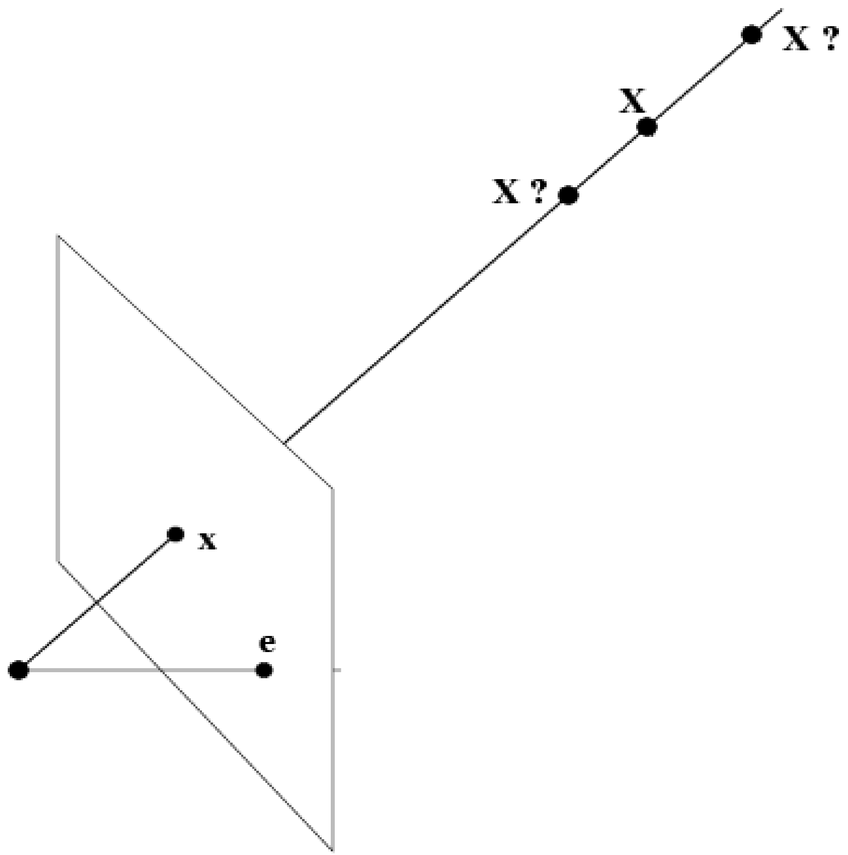

Trinity College Dublin
The University of Dublin
V-SENSE


HSLU_Planica_GS_VR - BtsScene - Windows, Mac, Linux - Unity 2022.3.10f1 <DX12>
File Edit Assets GameObject Component Services Jobs Tools Volograms Window Help
Experimental Packages In Use
Layers
Layout
Hierarchy
Scene
Game
SolidTwoSide
Dissolve
Inspector
Lighting
BtsScene
Directional Light
GaussianSplats
XR Interaction Manager
Globals
Billboard
Billboard (1)
Billboard (2)
Collision
Teleport Area
Cube (1)
Cube (2)
Cube (3)
Cube (4)
Cube (5)
TriggerBoxes
PlayBox
PlayBox (1)
PlayBox (2)
XR Origin (XR Rig)
ExitButton Variant
Vologram
EventSystem
Gaze Snap Volume
DontDestroyOnLoad
Persp
Project
Console
Favorites
All Modified
All Conflicts
All Excluded
All Materials
All Models
All Prefabs
Assets
GaussianAssets
Materials
Balloon
Meshes
Prefabs
Samples
Scenes
BTS
Immersive
Inspect
Scripts
Shaders
Snow_mountain
Textures
TutorialInfo
Video
Assets > Scenes > BTS
BtsScene

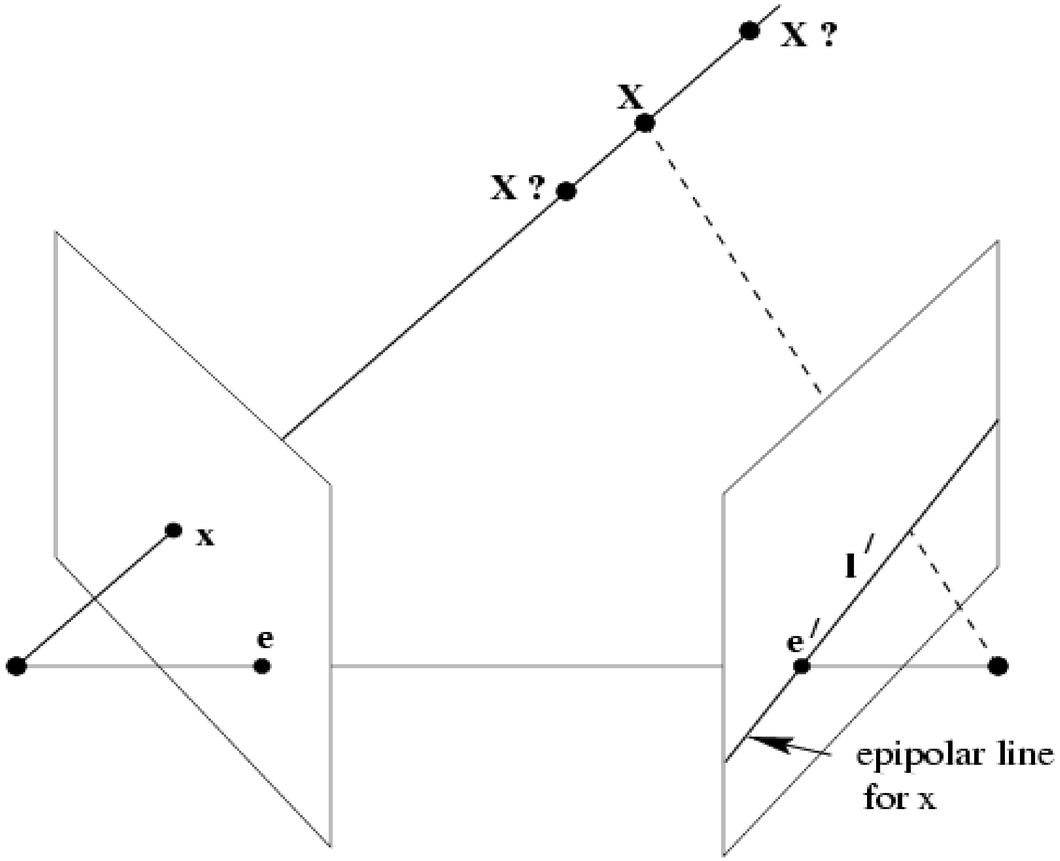
TRANSMIXR

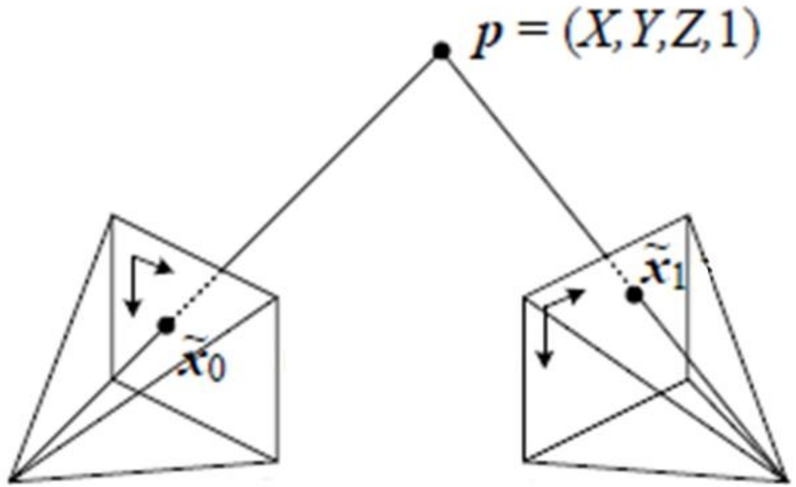
Traditional Volumetric Content Creation Static

3D Information is lost by imaging

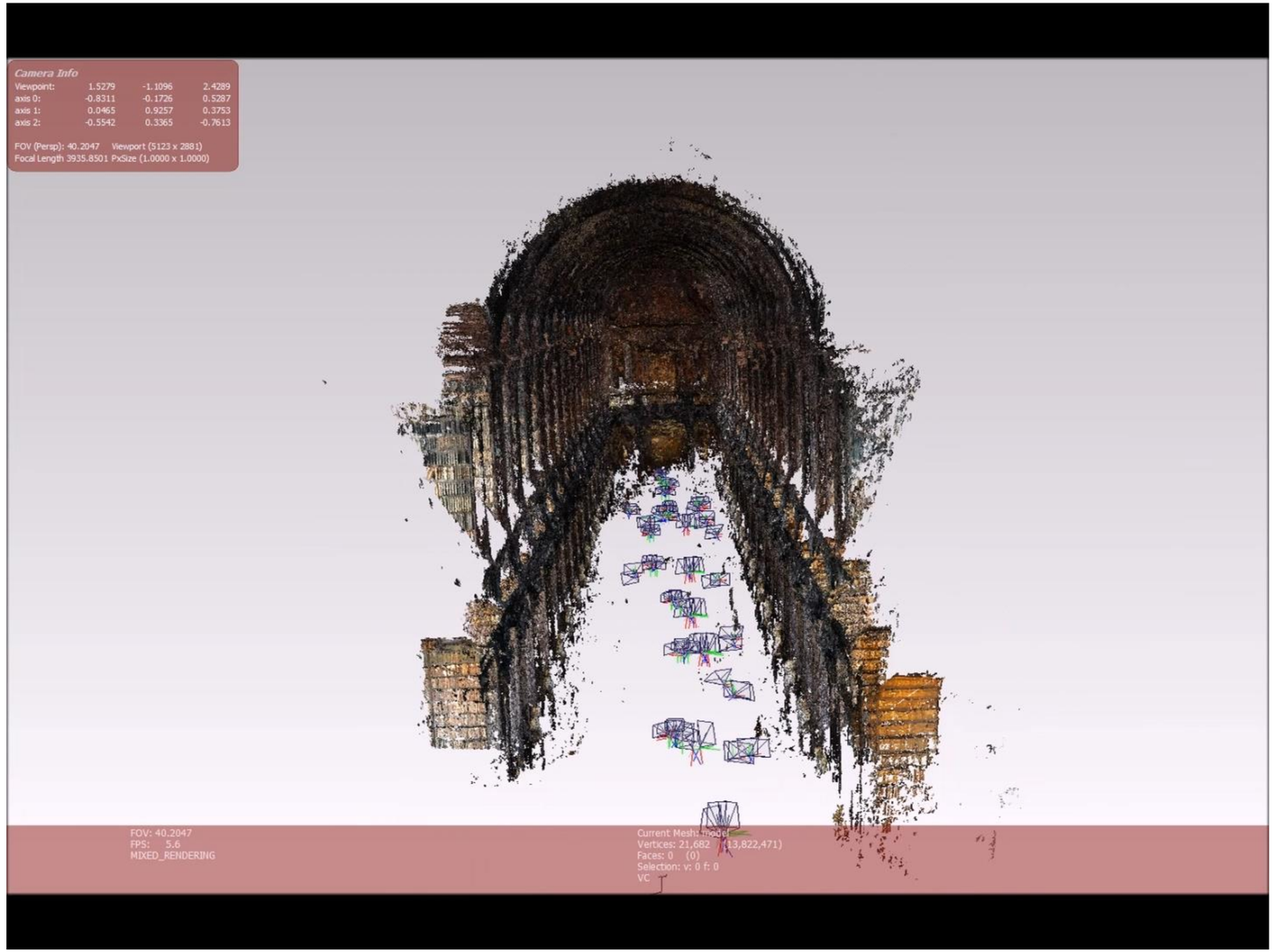


A second view is needed to recover 3D info
Multiview stereo, Shape from stereo

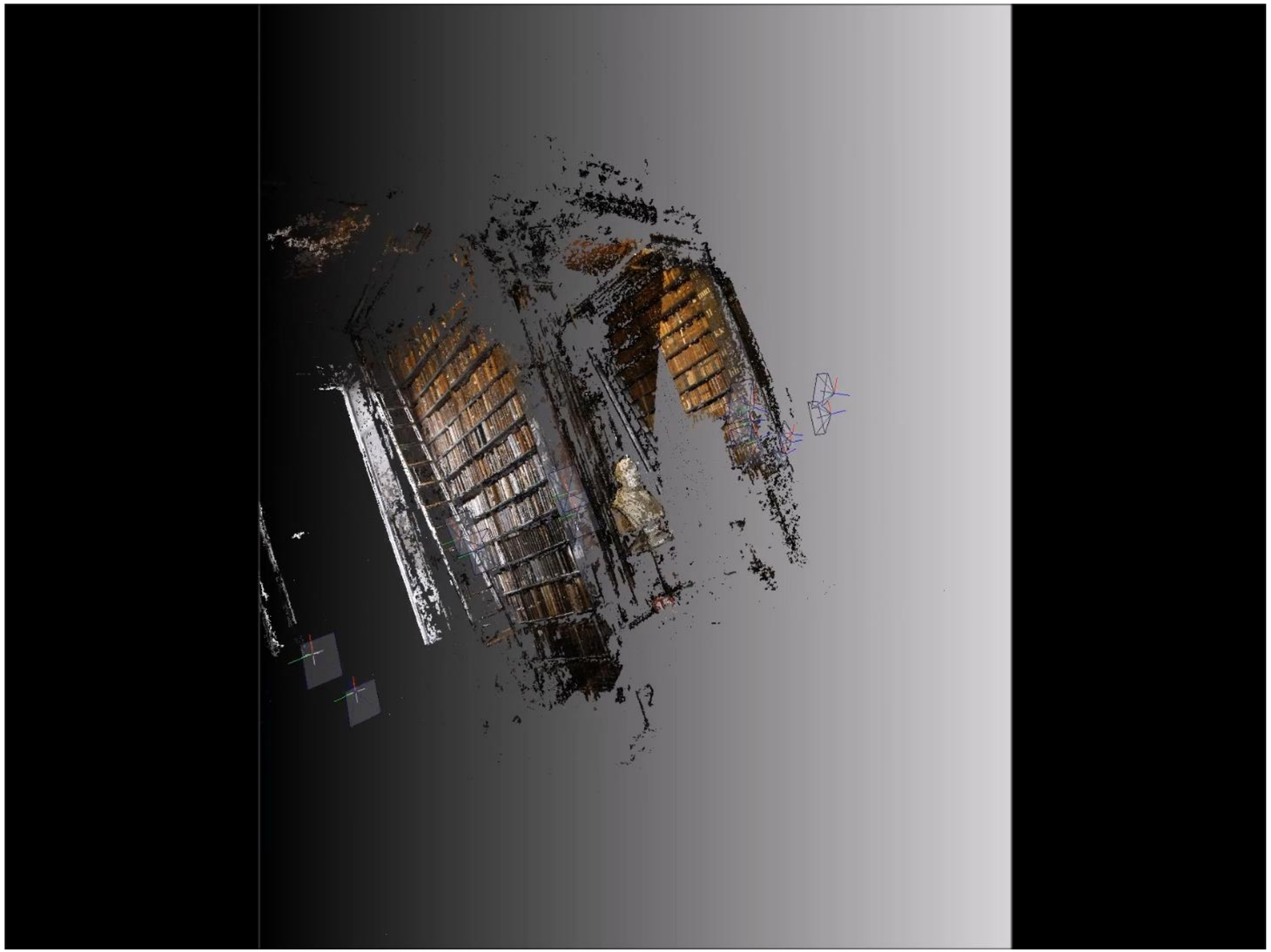




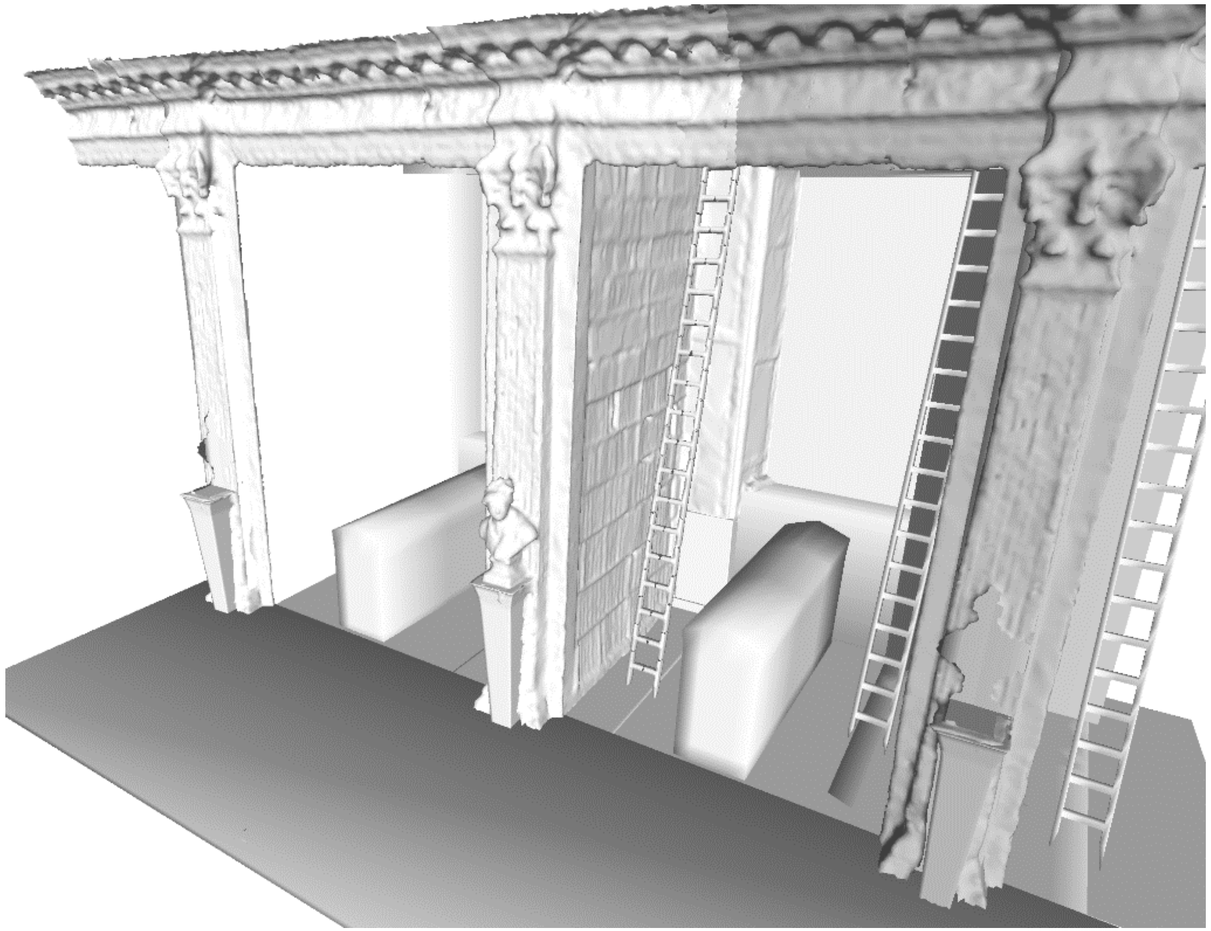
SfM for AR/VR Content Creation



SfM for AR/VR Content Creation



Jonathan Swift in VR and AR



SfM – Examples: Hintze Hall

Natural History Museum, London

Hintze Hall
3D Model

artfletch PRO
FOLLOW

1.7k 10.6k 258

Download 3D Model + Add To Embed Share Report

Triangles: 2M Vertices: 1.4M More model information

The Hintze Hall is the central hall and largest space at the Natural History Museum, London.

https://www.nhm.ac.uk/visit/galleries-and-museum-map/hintze-hall.html
https://historicengland.org.uk/listing/the-list/list-entry/1080675

3284 photos taken as a tourist in March 2023 with a Sony a7R III and processed in Reality Capture. Model heavily manually cleaned up in places to improve quality. Gorilla in north east corner hasn't processed properly and I will try to inmprove that area in time.

https://sketchfab.com/3d-models/hintze-hall-45f5e56887f44075bbf283977c99541f

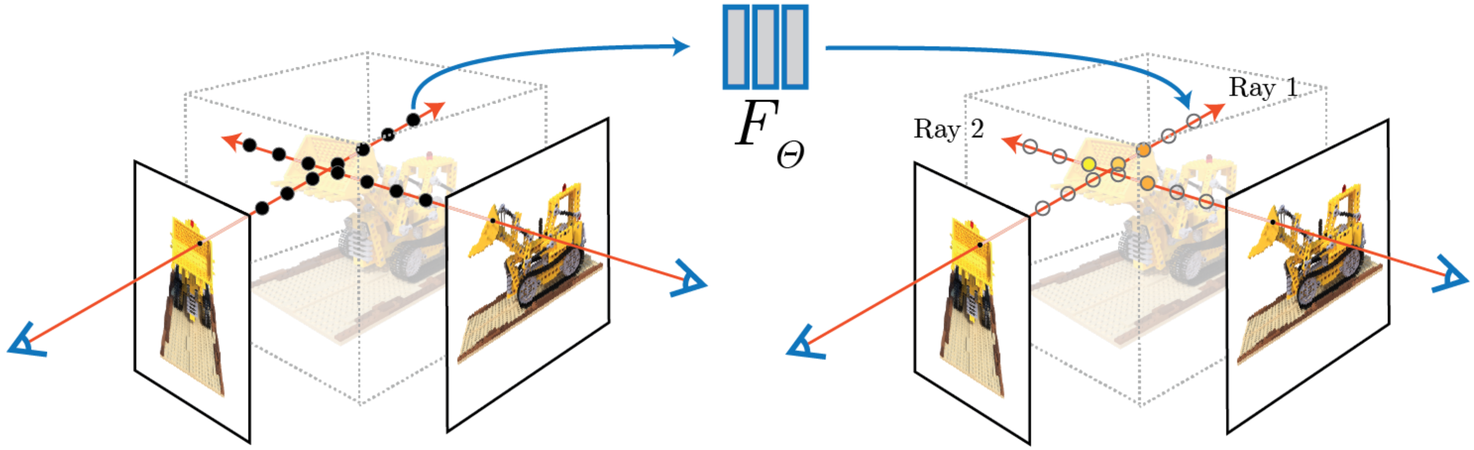
AI-based Volumetric Content Creation Static

NeRF: Neural Radiance Fields

NeRF: Neural Radiance Fields

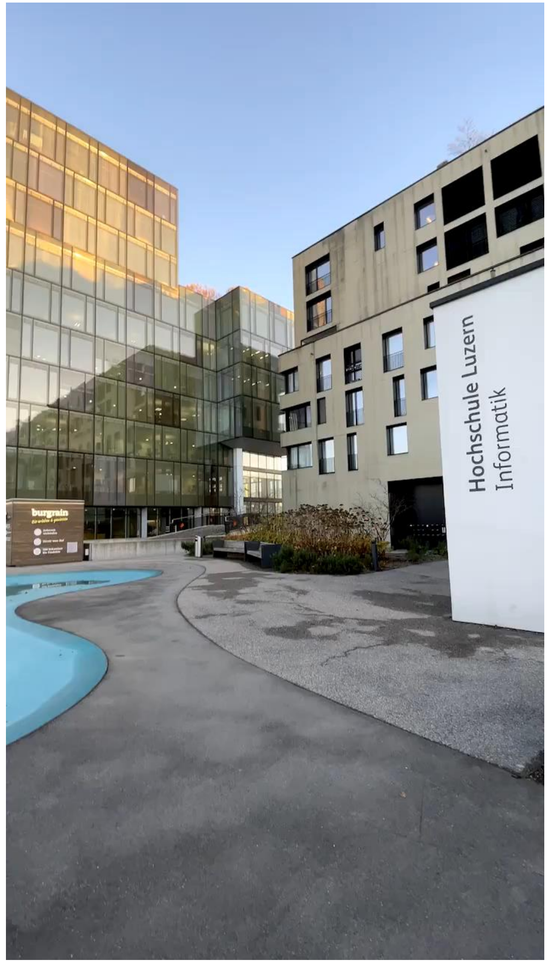
NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis
Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, Ren Ng

ECCV 2020 (oral).



AI-based 3D Reconstruction

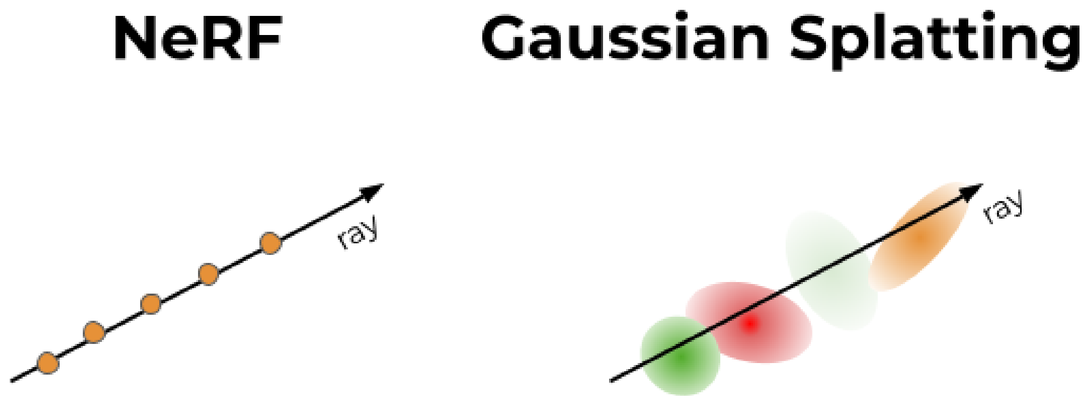
Neural Radience Fields (NeRF)
Gaussian Splatting



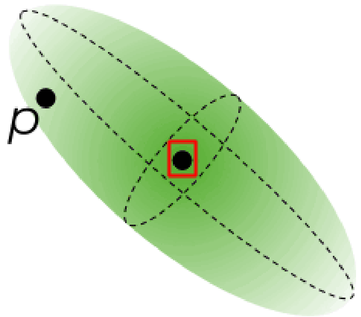
Sunny2

Fountain (Ufschötti in Luzern)

3D Gaussian Splatting



$$f_i(p) = \sigma(\alpha_i)\exp(-\frac{1}{2}(p - \boxed{\mu_i})\Sigma_i^{-1}(p - \boxed{\mu_i}))$$

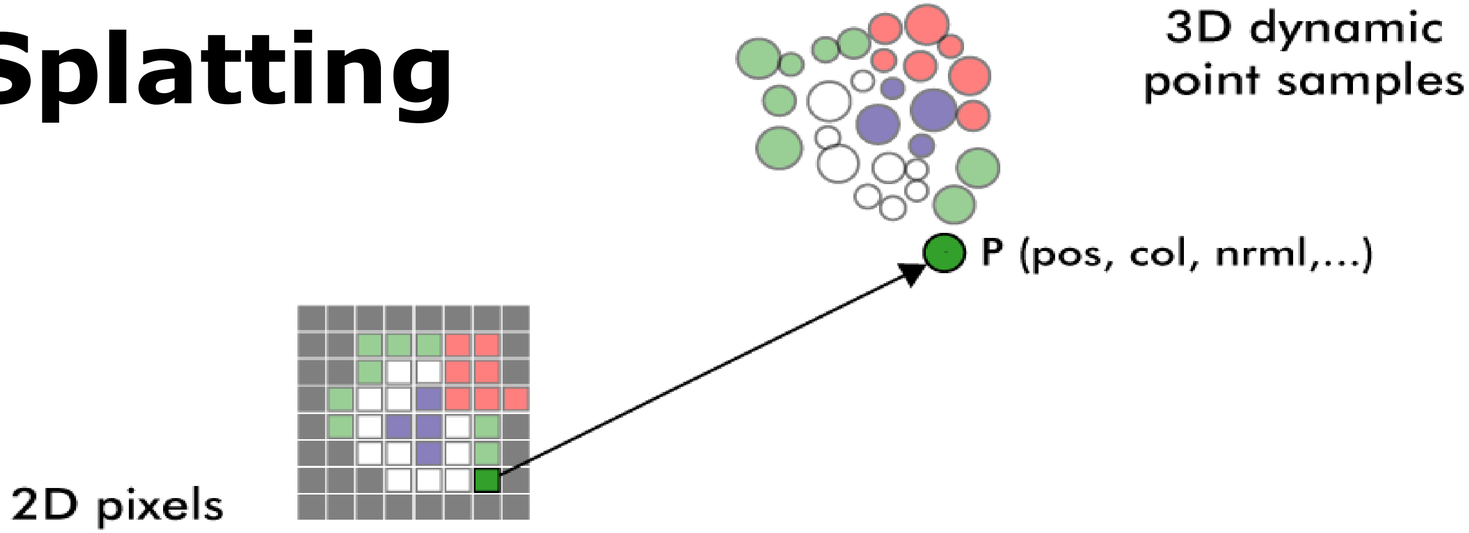


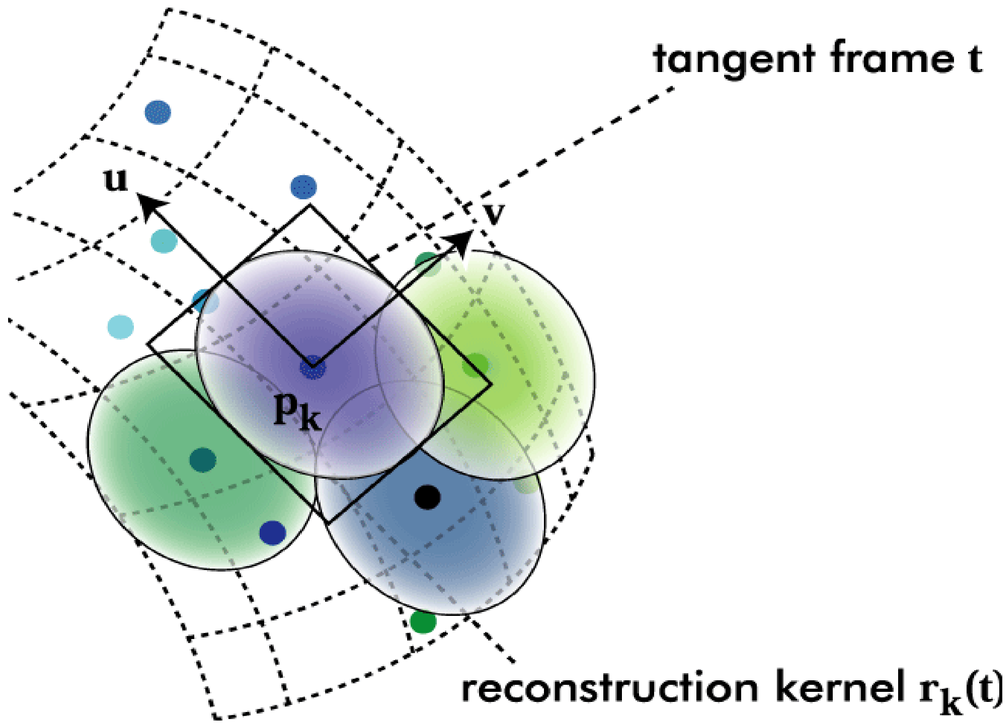
Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, George Drettakis, “3D Gaussian Splatting for Real-Time Radiance Field Rendering” (ACM Transactions on Graphics).

https://repo-sam.inria.fr/fungraph/3d-gaussian-splatting/

https://towardsdatascience.com/a-comprehensive-overview-of-gaussian-splatting-e7d570081362

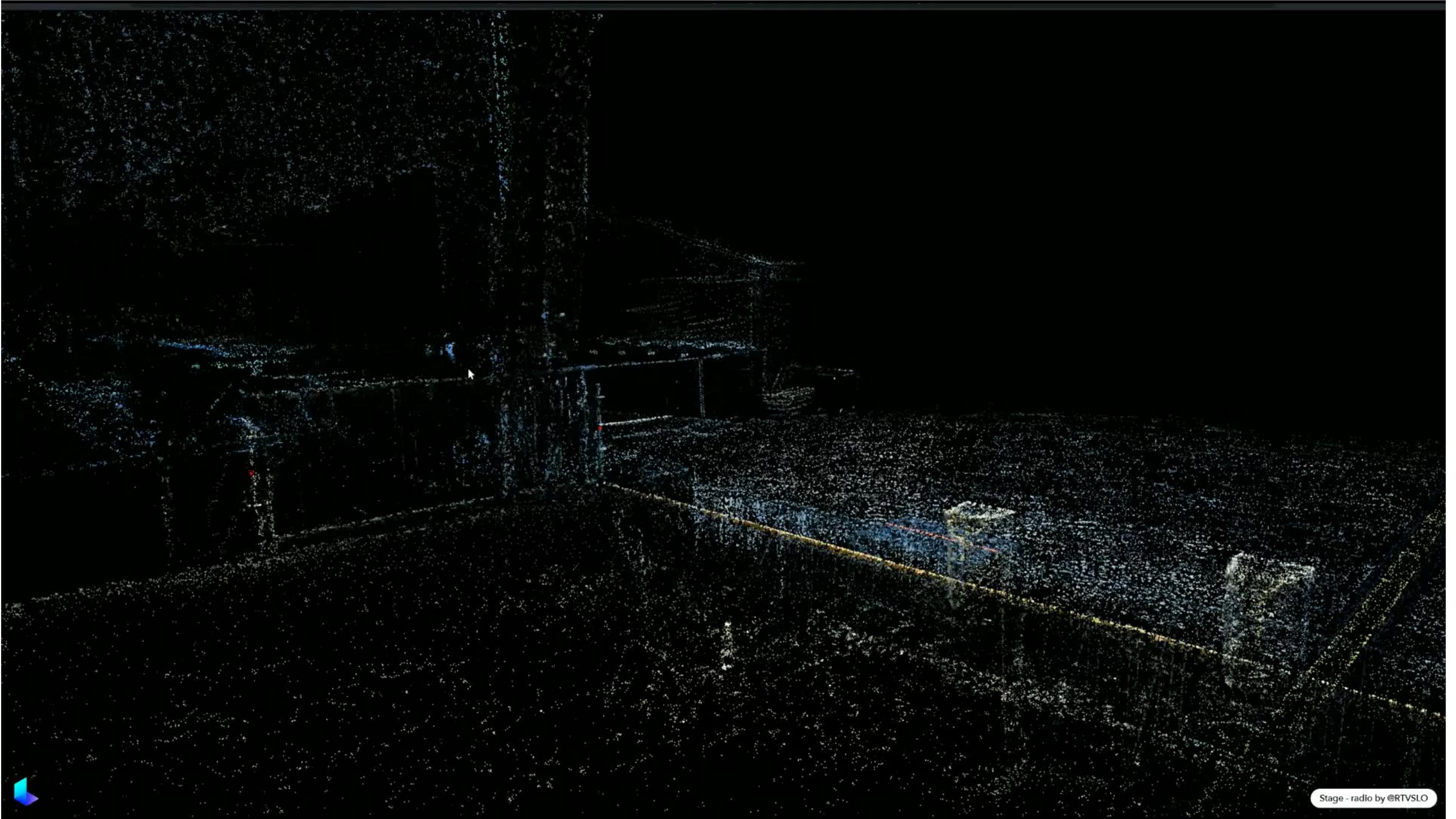
3D Gaussian Splatting





Matthias Zwicker, Hanspeter Pfister, Jeroen van Baar, and Markus Gross, “EWA Splatting”, IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS, VOL. 8, NO. 3, JULY-SEPTEMBER 2002.

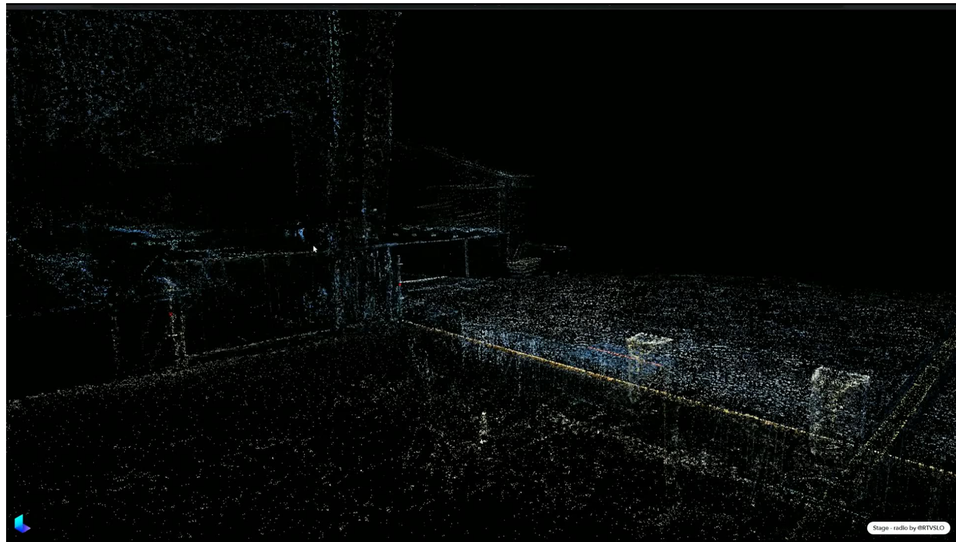
Stephan Würmlin, Edouard Lamboray, Markus Gross, “3D video fragments: dynamic point samples for real-time free-viewpoint video”, Computers & Graphics, Volume 28, Issue 1, February 2004, Pages 3-14.


Stage - radio by @RTVSLO

AI-based Content Creation Workflow

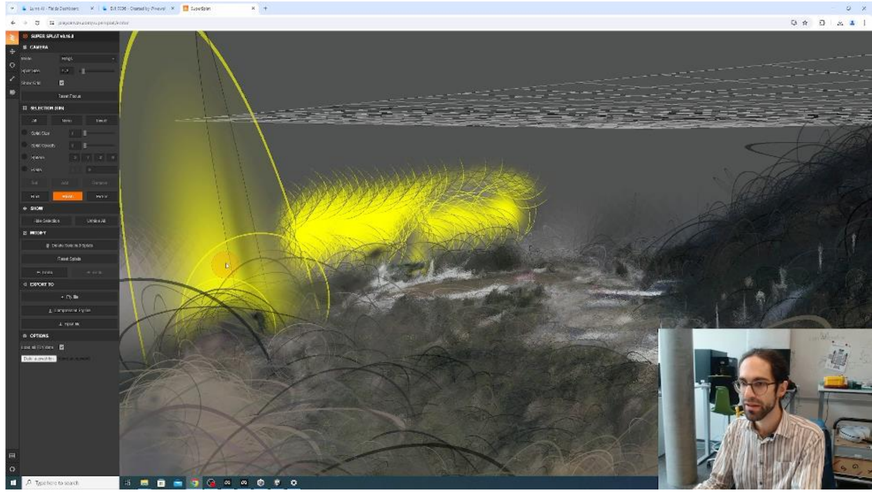
- Recent AI-based algorithms for 3D reconstruction (NeRF, Gaussian Splatting) provide high quality 3D assets from simple captures, e.g. a single video that anyone can produce
- Enable 3D digitization of real world objects and scenes that can be used in digital media, such as XR applications
- However, integration into professional production workflows and systems is not yet resolved

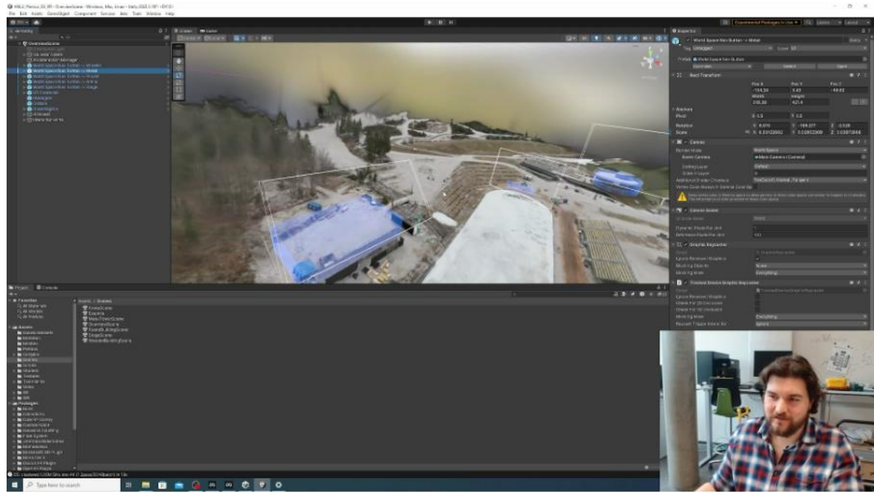
Objective:
Developing and showcasing the related AI-based content creation workflow



AI-based Content Creation Workflow

- **Capture**
  - A person (e.g. a journalist in a location) captures a video of an object/scene to be digitized
- **3D Reconstruction**
  - The video is processed by a tool like LumaAI or Postshot
- **Editing**
  - The resulting 3D asset is cleaned up and further manipulated as needed
- **Integration**
  - All assets are combined into a target application, e.g. a VR app by using a game engine like Unity
- **Experience**
  - Audiences can experience the final product





AI-based Content Creation Workflow

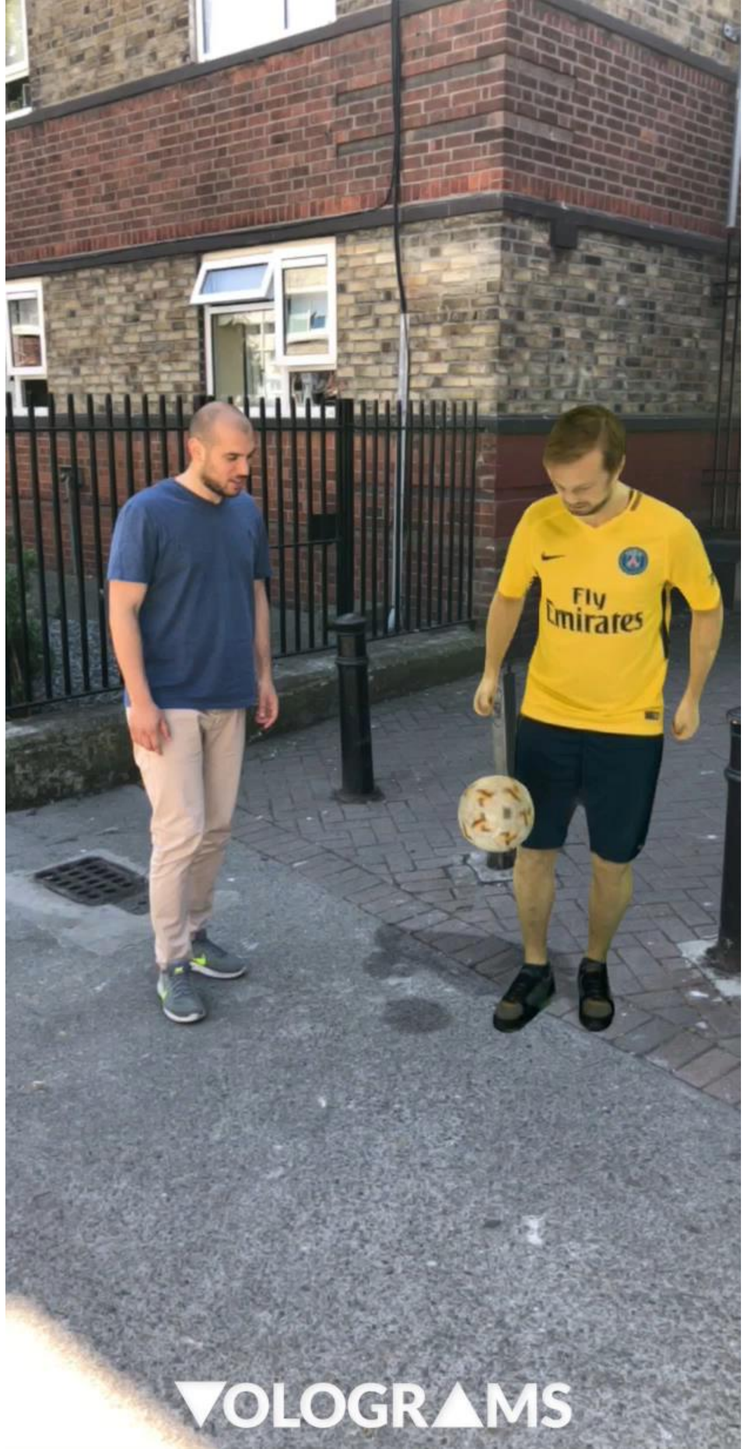
TRANSMIXR

VR Planica Experience


HSLU Hochschule Luzern

Traditional Volumetric Content Creation Dynamic


Fly
Emirates


Fly
Emirates
VOLOGRAMS


HSLU Hochschule Luzern
TRANSMIXR

Volumetric Video for XR Content Creation

Augmented Reality

Virtual Reality

VV by Shape from Silhouette (SfS)

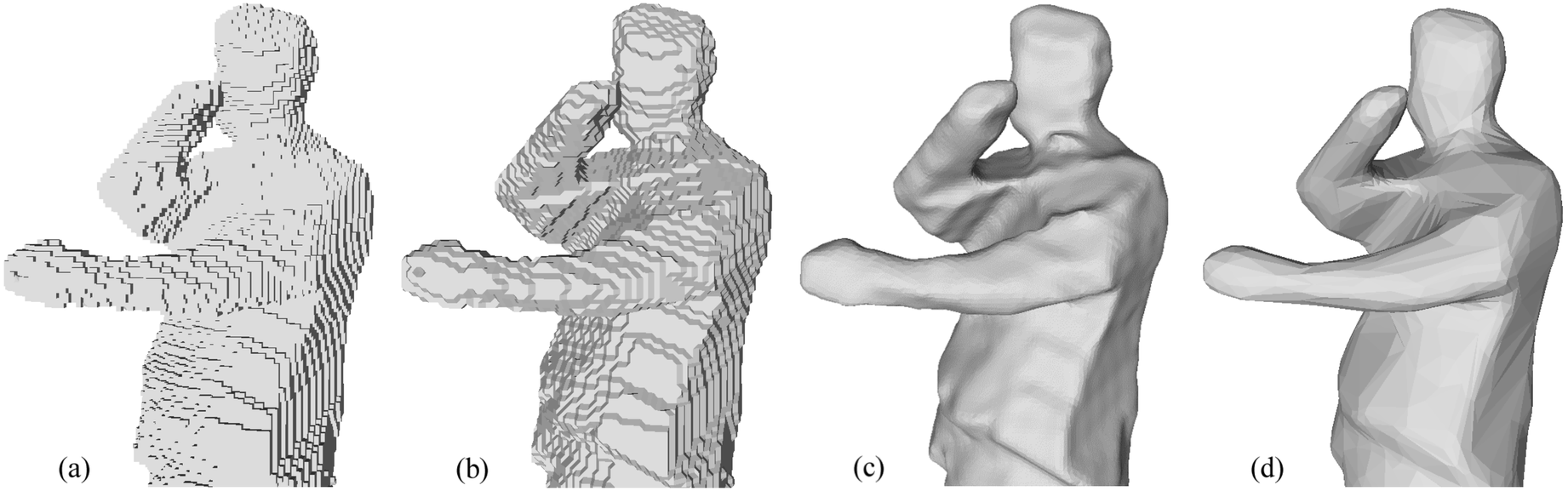
Shape-from-Silhouette-Approach

a) Hierarchical Voxel Modeling
b) Marching-Cubes-Approach
c) Surface Smoothing
d) Mesh Reduction

T. Lewiner *et al.*: Efficient Implementation of Marching Cubes Cases with Topological Guarantee", Journal of Graphics Tools, Vol. 8, pp. 1-15, 2003.

G. Taubin: Curve and Surface Smoothing Without Shrinkage, Proc. IEEE International Conference on Computer Vision (ICCV 95), 1995.

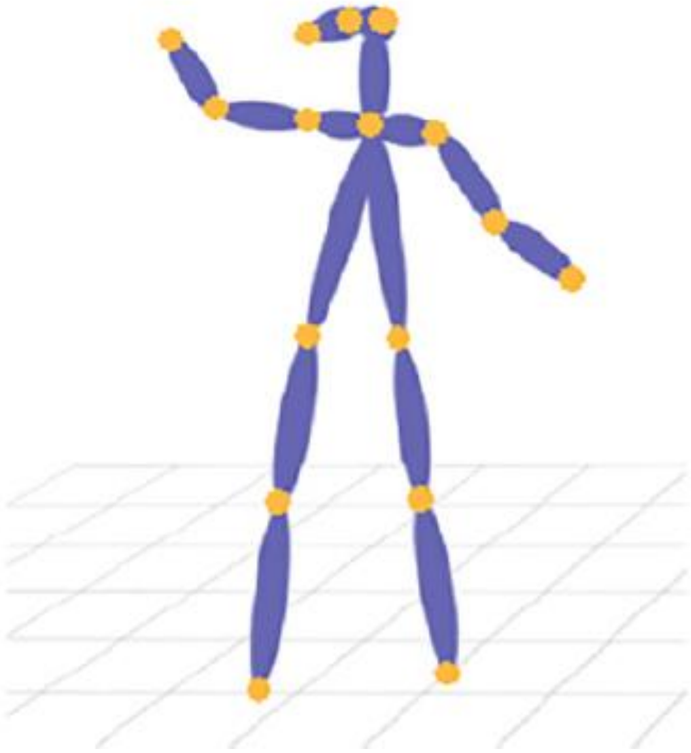
H. Hoppe, Efficient Implementation of Progressive Meshes, Computers & Graphics, Vol. 22, No. 2, pp. 27-36, 1998 / Tool in DirectX



VV by Structure from Motion (SfM), Multi-View Stereo (MVS)

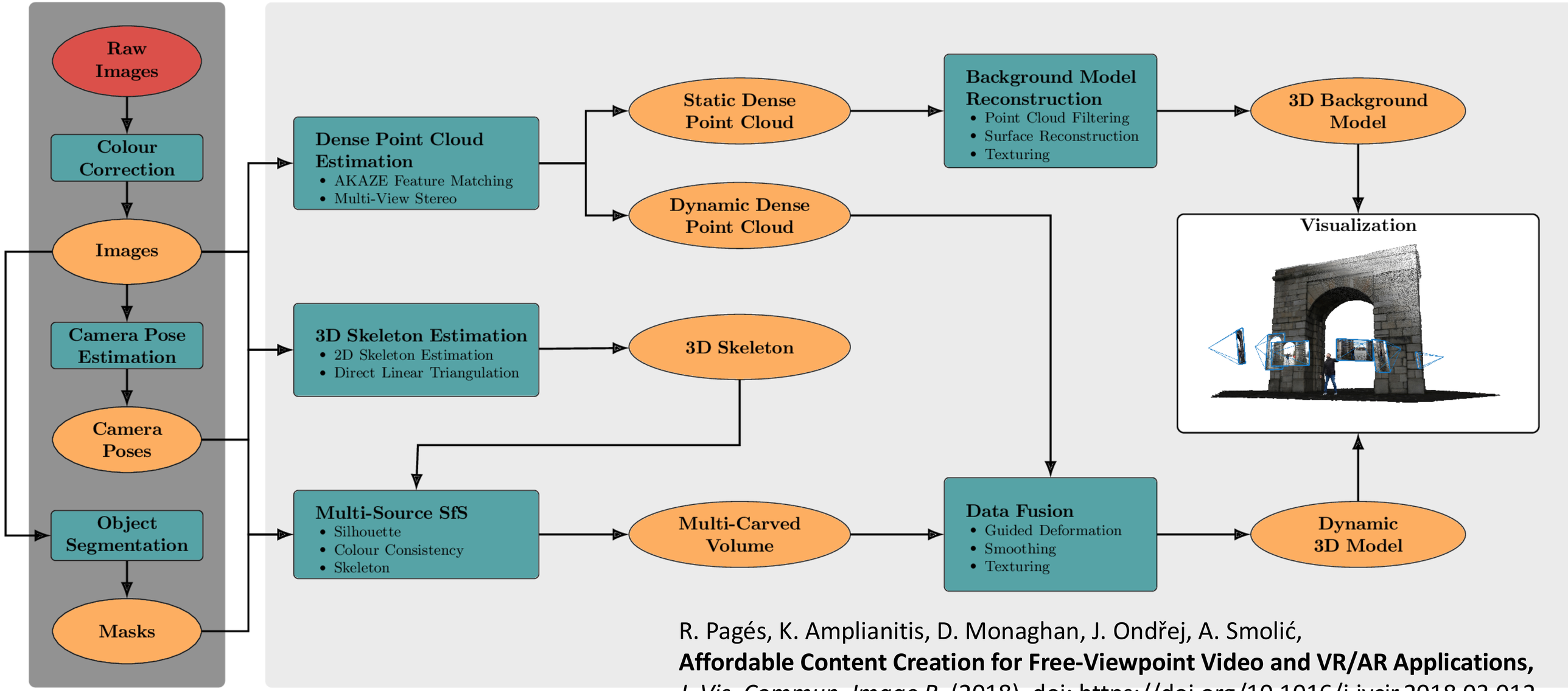
Moynihan, Matthew; Pagés, Rafael; Smolic, Aljosa
A Self-regulating Spatio-Temporal Filter for Volumetric Video Point
1182 , pp. 391-408, Springer International Publishing, 2020, ISBN: 978-3-030-41590-7.

3D Pose Estimation by Multiview Triangulation



Hybrid VV Content Creation

2019 JVCI Best Paper Award



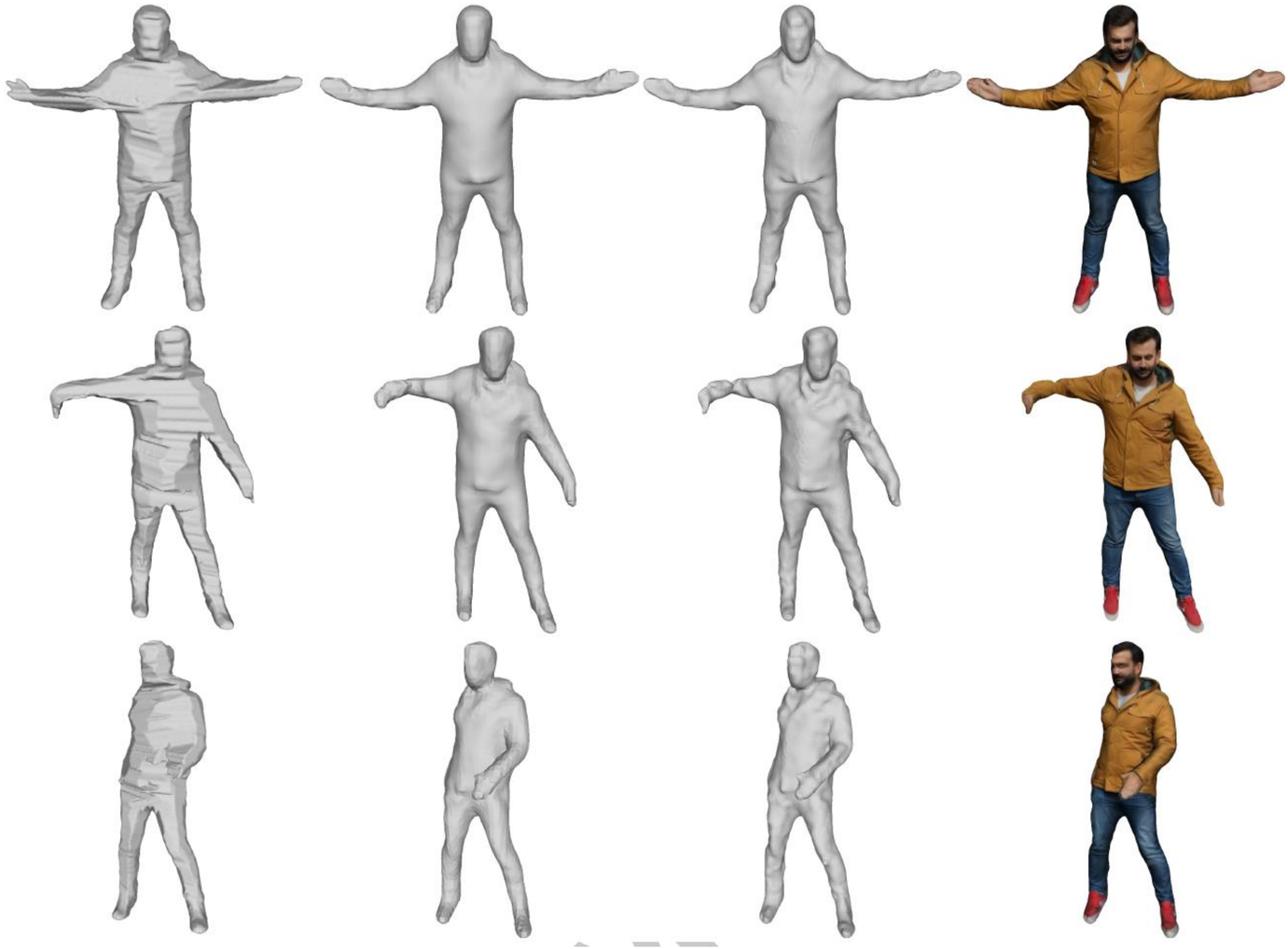
R. Pagés, K. Amplianitis, D. Monaghan, J. Ondřej, A. Smolić,
Affordable Content Creation for Free-Viewpoint Video and VR/AR Applications,
J. Vis. Commun. Image R. (2018), doi: https://doi.org/10.1016/j.jvcir.2018.03.012

VV for VR/AR Content Creation

http://authors.elsevier.com/sd/article/S1047320318300683

R. Pagés, K. Amplianitis, D. Monaghan, J. Ondřej, A. Smolić,
Affordable Content Creation for Free-Viewpoint Video and VR/AR Applications,
J. Vis. Commun. Image R. (2018),
doi: https://doi.org/10.1016/j.jvcir.2018.03.012

2019 JVCI Best Paper Award



VOLOGRAMS

Trinity College Dublin
The University of Dublin

V-SENSE

Making Volumetric Video Mainstream

VOLOGRAMS

VOLOGRAM™ STUDIO

ACCESSIBLE 3D HUMAN CREATION

60 DSLR fixed studio

12 camera setup

5 DSLR temporary studio

16 camera facial capture studio

84 camera high-end studio

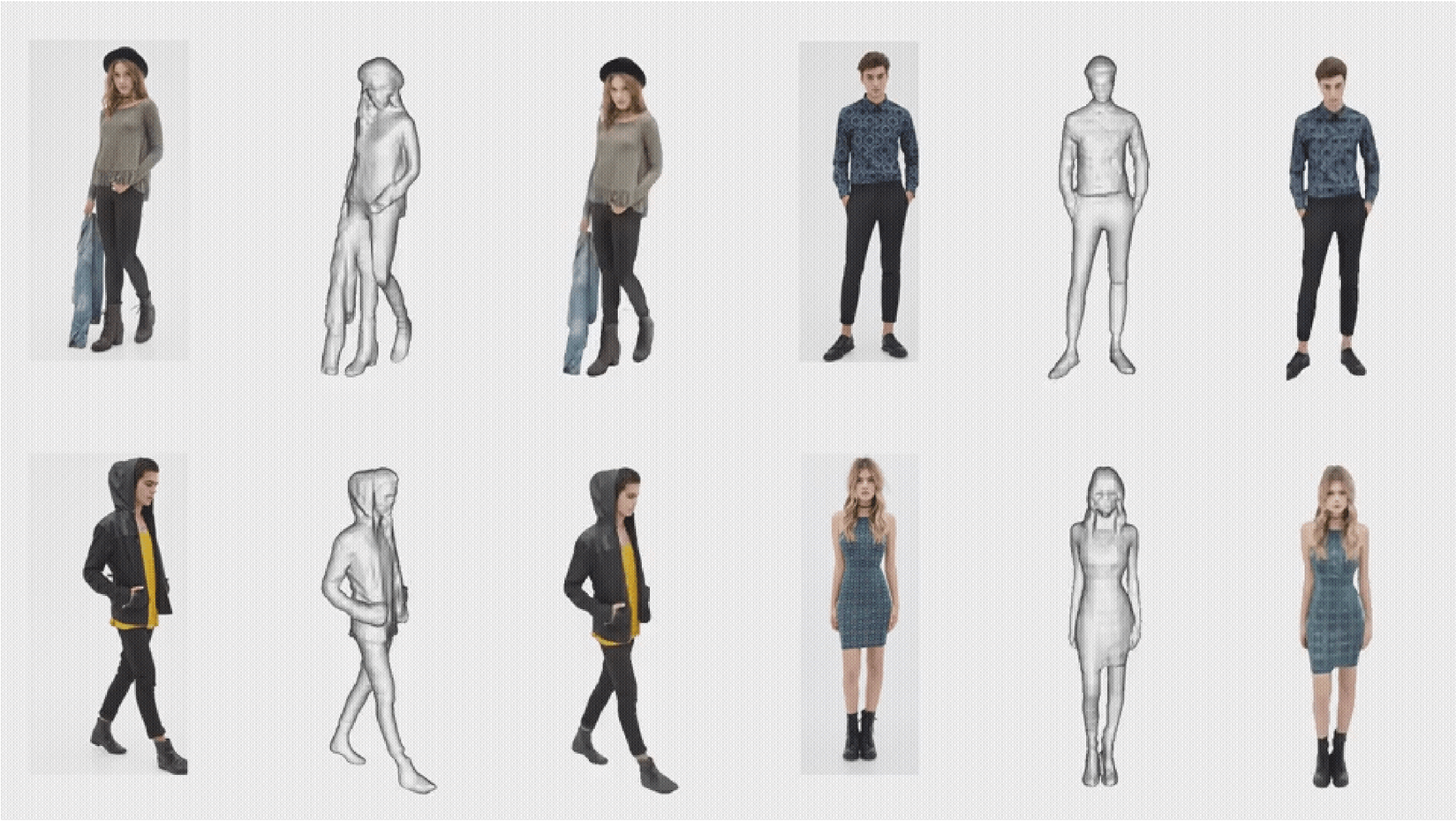
24 smartphone temporary studio

68 DSLR high-end studio

Alvaro Collet, Ming Chuang, Patrick Sweeney, Don Gillett, Dennis Evseev, David Calabrese, Hugues Hoppe, Adam Kirk, Steve Sullivan. "High-quality streamable free-viewpoint video." ACM Transactions on Graphics (SIGGRAPH), 34(4), 2015.

AI-based Volumetric Content Creation Dynamic

3D Deep Learning



Saito, S., Simon, T., Saragih, J. and Joo, H.
PIFuHD: Multi-Level Pixel Aligned Implicit Function for High-Resolution 3D Human Digitization, CVPR 2020

3D Deep Learning

Saito, Shunsuke, et al. "Pifuhd: Multi-level pixel-aligned implicit function for high-resolution 3d human digitization." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020.

Bringing reality capture closer to everyone. Volograms. (n.d.). Retrieved April 23, 2022, from https://www.volograms.com/

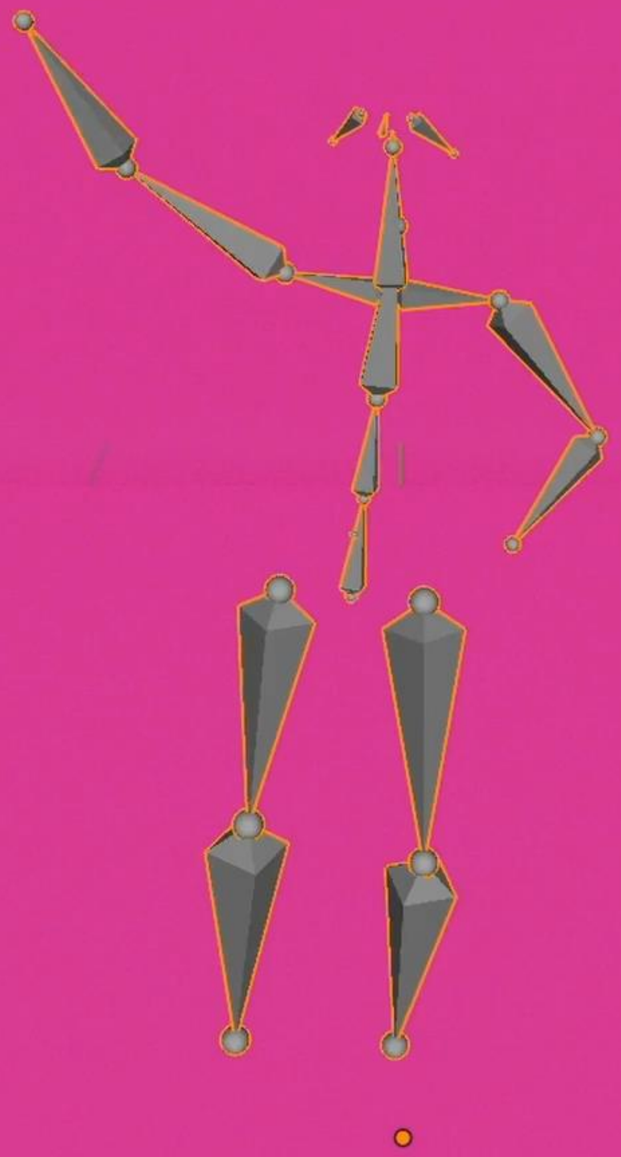
POWERED BY VOLOGRAMS AI

SEMANTIC SEGMENTATION

MONOCULAR 3D RECONSTRUCTION

3D SKELETON ESTIMATION

AND MORE!



NO MORE MASSIVE CAPTURE STUDIOS

NO MORE

- ✕ Dozens of cameras
- ✕ Green screens
- ✕ Server farms
- ✕ Specialized staff

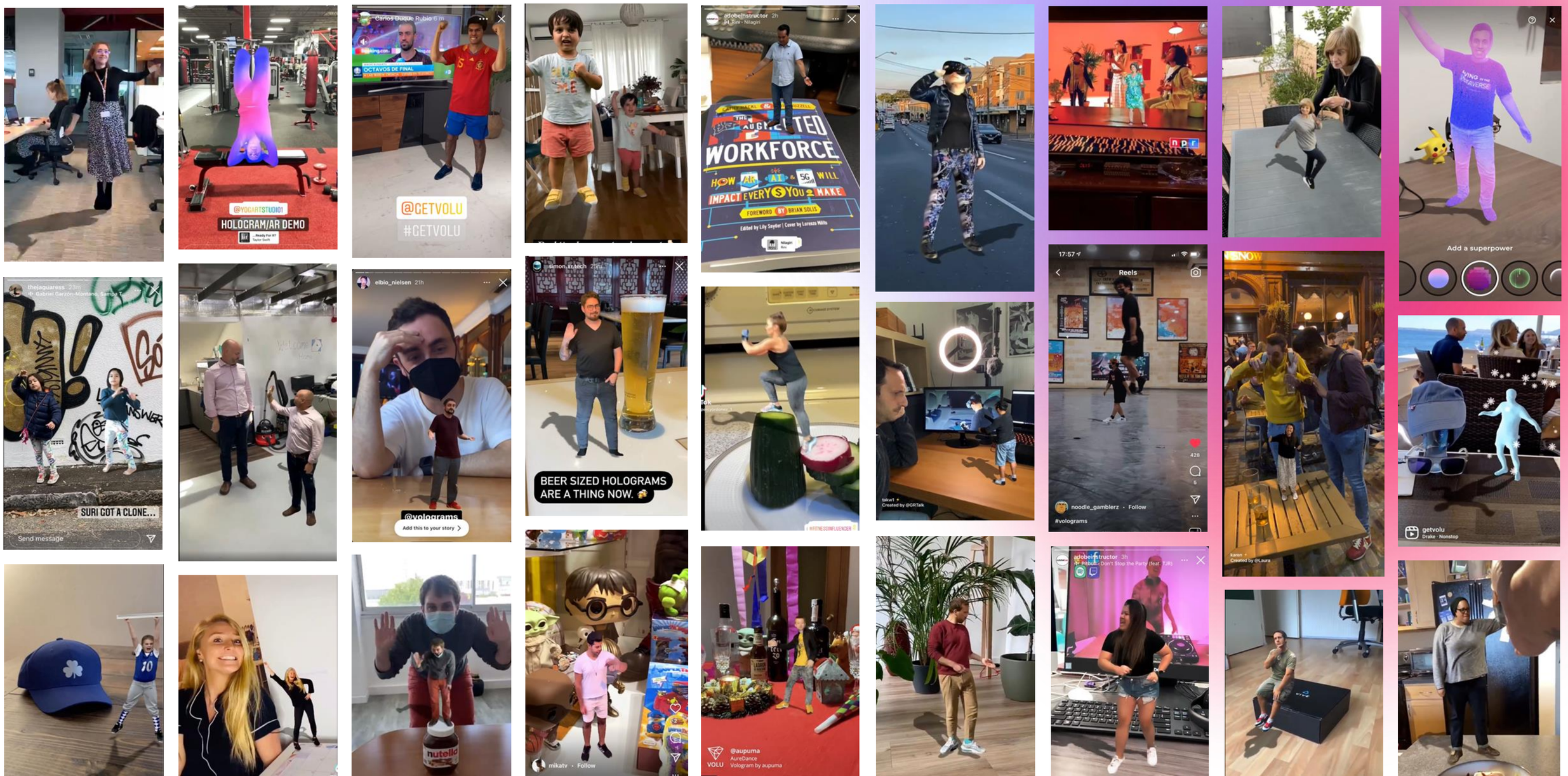
YOU ONLY NEED A SMARTPHONE

VOLU USERS

OUR CREATORS ARE HAVING FUN!

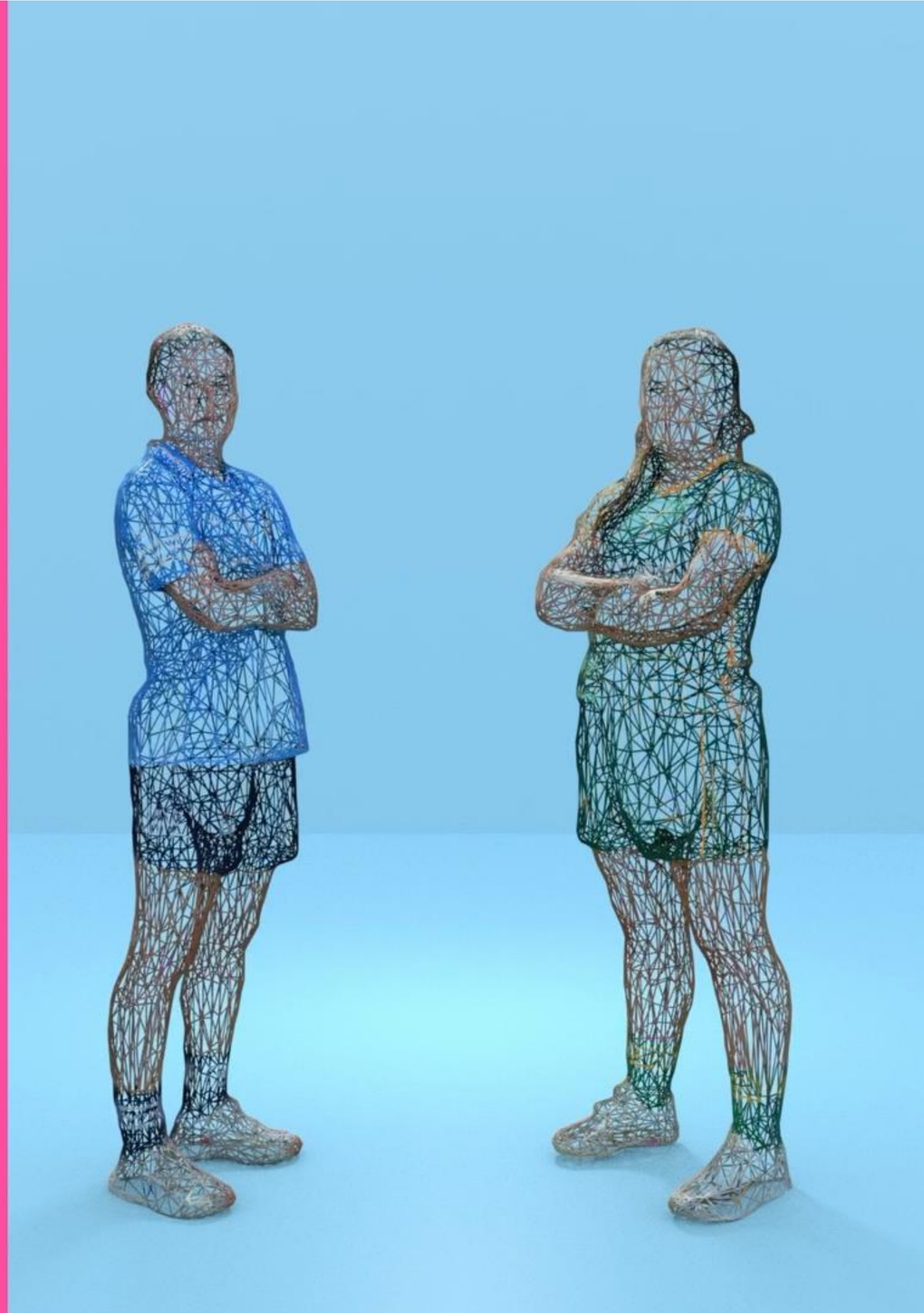
See it in action on Twitter, Twitter, LinkedIn, TikTok or Instagram

Download on the App Store



READY FOR YOUR PRODUCTION
LIMITLESS POSSIBILITIES

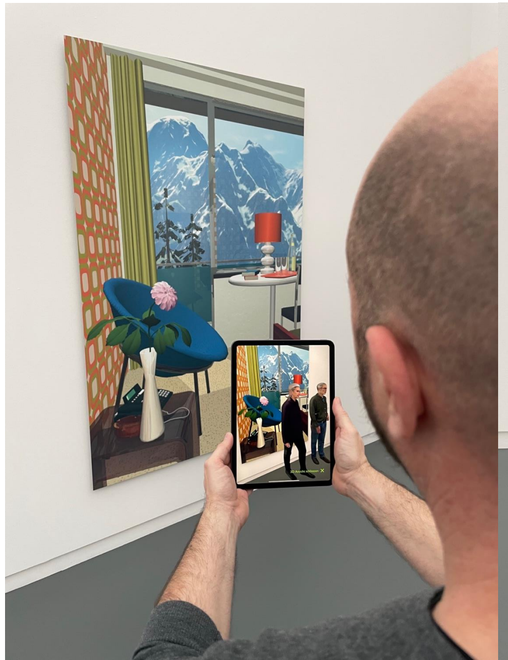
POWERED BY VOLOGRAMS AI



Volumetric Holograms in the Aargauer Kunsthaus

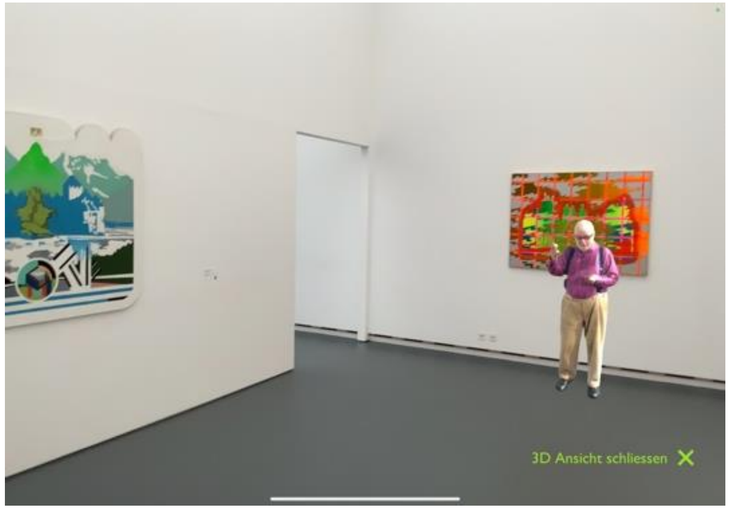
Innosuisse Innovation Cheque VEMAR

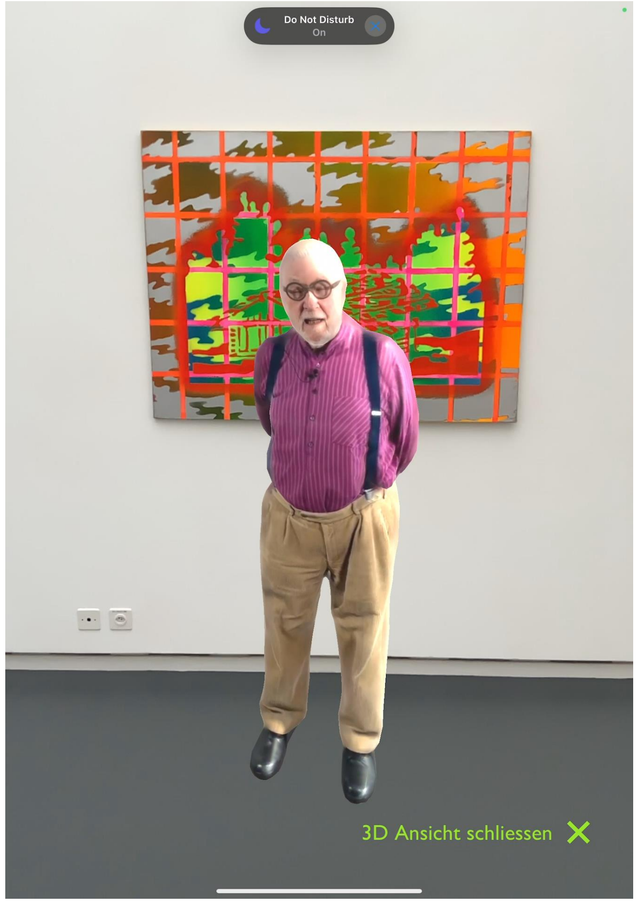
Volumetric Video for Enhanced Museum Experiences in AR



HSLU

VEMAR





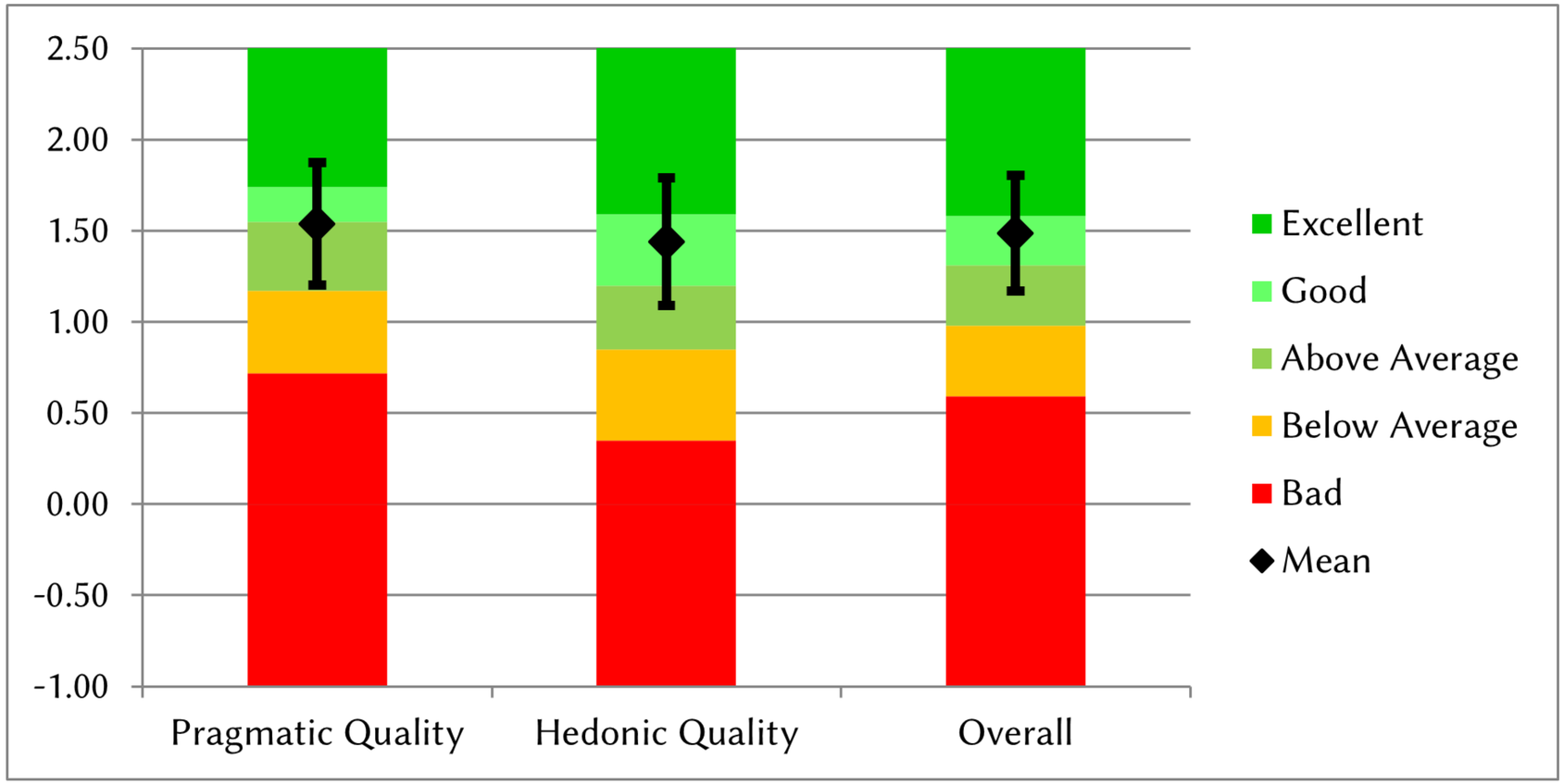
HSLU

Samuel Buri, Châlet Psychédélique, 1967

VEMAR Workflow

- Ideation, experience design
- Capture
- Post-production
- Game engine implementation
- Software integration
- User study

VEMAR UEQ-S



Anh Nguyen, Jan Lässig, Anna F. Kälin, Ariana Huwiler, Philipp Haslbauer, Gareth W. Young, Lam Kit Yung, Aljosa Smolic, “A Volumetric Video Application to Enhance Museum Experiences”, VRST '24: Proceedings of the 30th ACM Symposium on Virtual Reality Software and Technology, Article No.: 76, Pages 1 – 3, https://doi.org/10.1145/3641825.3689695

VV Streaming: VV-glTF

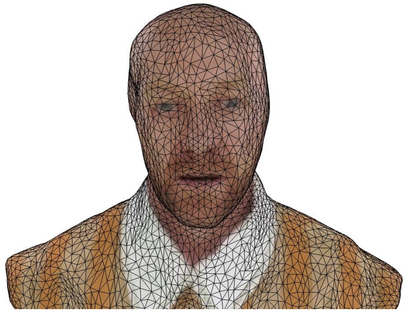
3D Mesh Resolution and Data Rates

50k
polys/frame
~1GB / minute

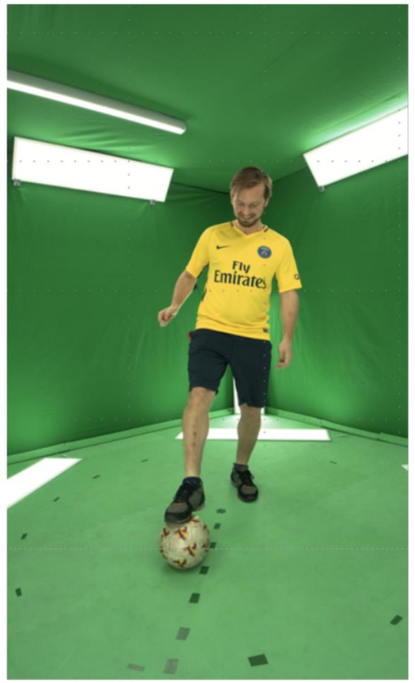
15k
polys/frame
~400MB /
minute

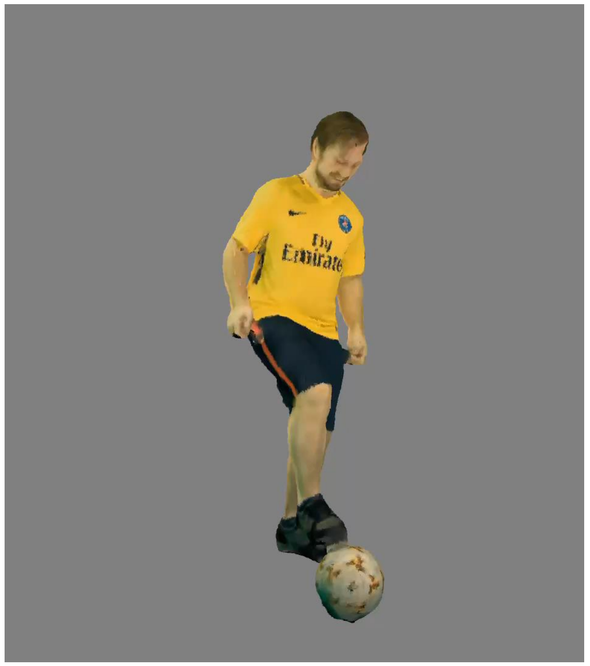
4k polys/frame
~150MB /
minute

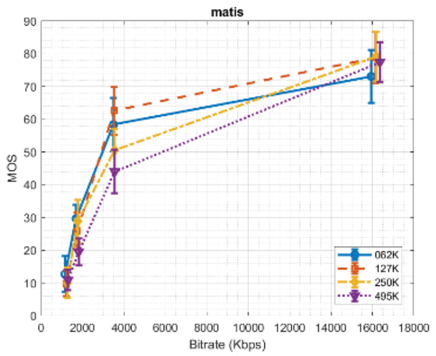
Content Delivery Pipeline for Volumetric Video

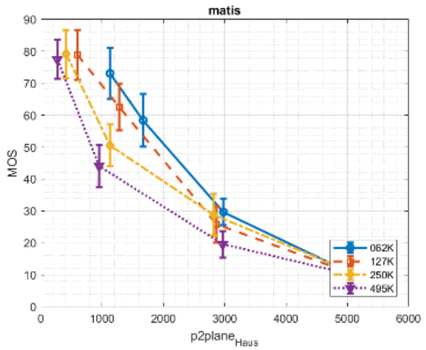


VV Acquisition → Processing → Compression → Streaming → Display → Quality Assessment





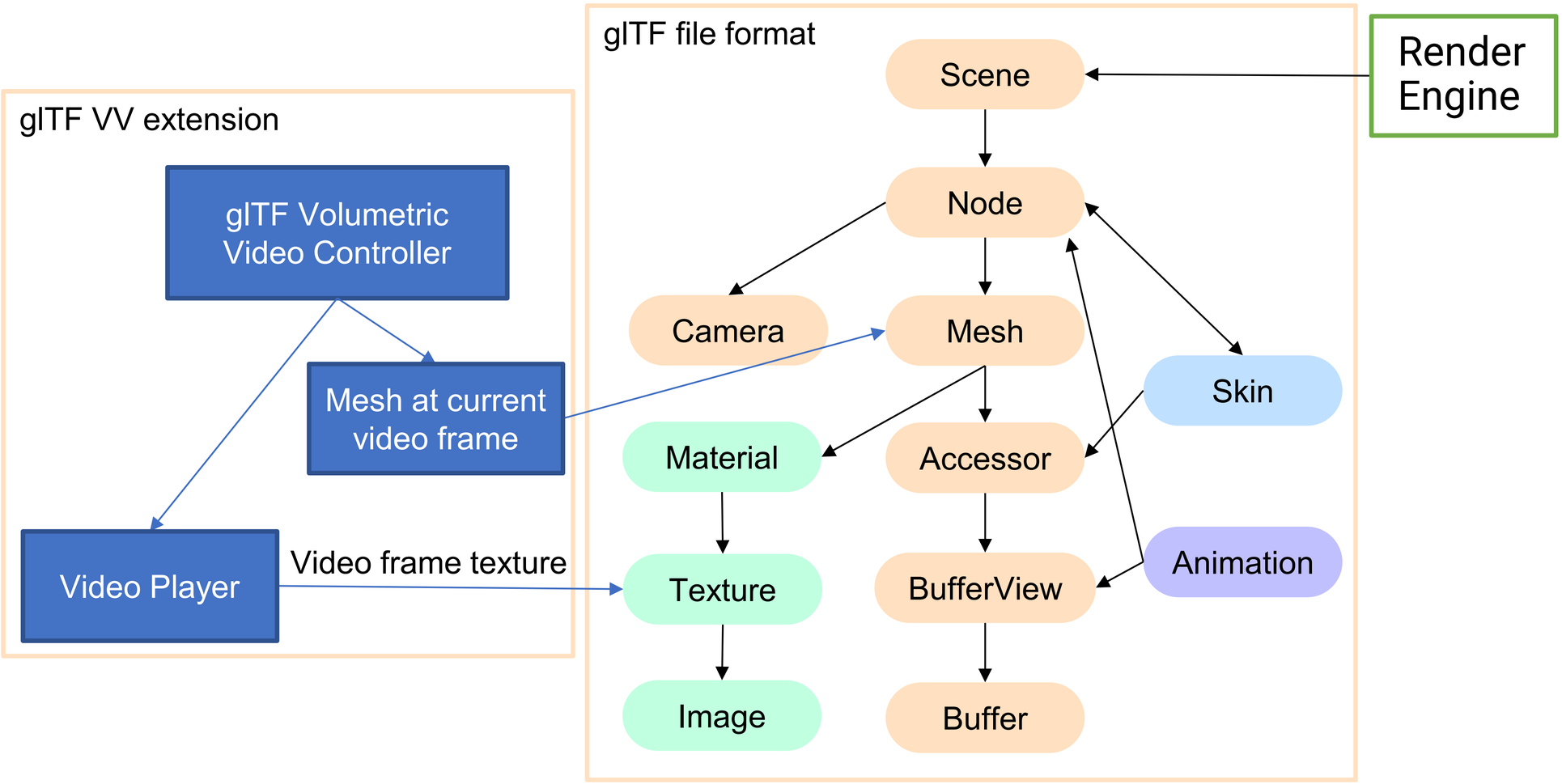




Volumetric video:

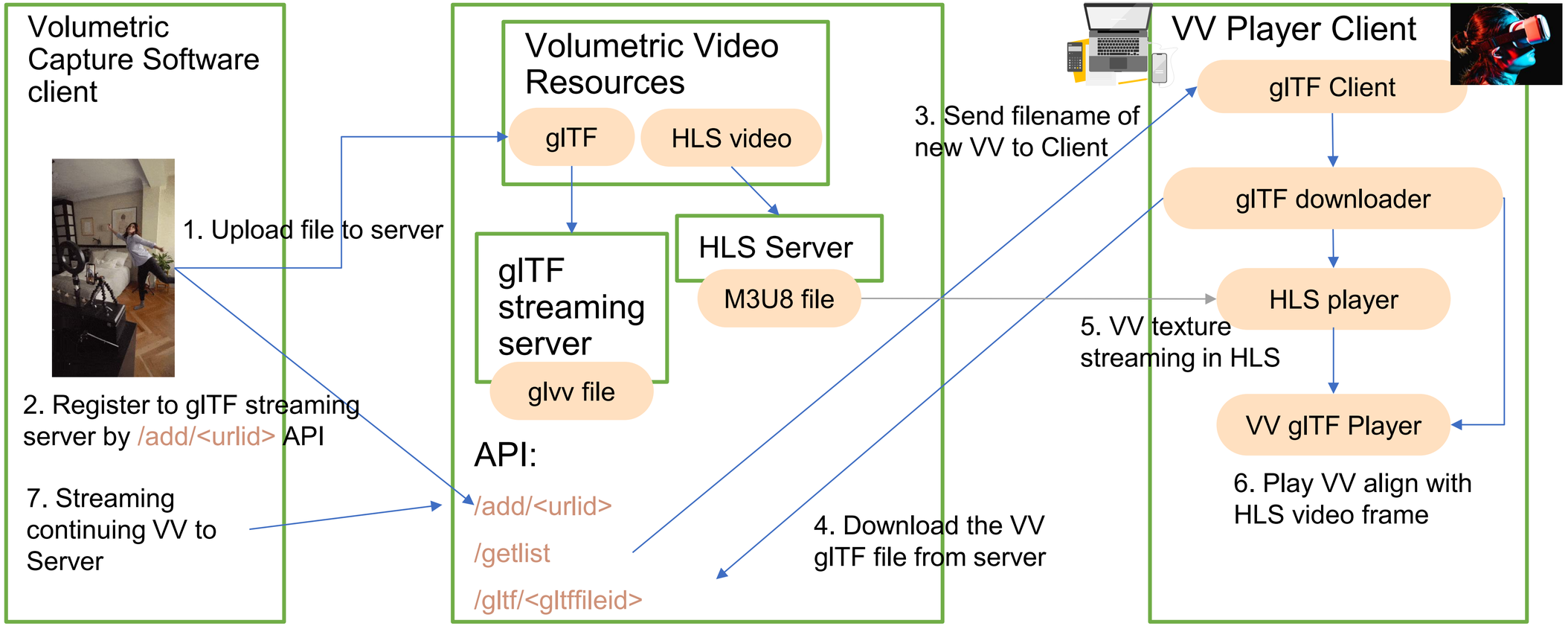
- Geometry + Texture
- Dynamic, changing over time

VV-glTF Extension



Kit Yung Lam, Simone Croci, Philipp Haslbauer, Aljosa Smolic, “VVGLTF: EFFICIENT STREAMING OF VOLUMETRIC VIDEO WITH GLTF”, Technical Paper, IBC – International Broadcast Convention, Amsterdam 2024.

VV Streaming with VV-glTF TRANSMIXR



VV-glTF Extension Submitted

https://github.com/KhronosGroup/glTF/pull/2438

Dynamic GS Reconstruction Pipeline

HEAT

FAVORE
FAST VOLUMETRIC RECONSTRUCTION

TRANSMIXR

Exploration of Reconstruction Methods

Exploration of 3DGS and recent inverse methods:

- Pixel-wise Splats (GPS Gaussian)
- Depth from Sensor / Stereo / Color
- Alignment of Monocular Depth to Sensor (Metric)
- Assembly of survey paper material (Dynamic Human Representations*)

* PCD, GS, Spherical Splats, Splats on top of mesh, Tracking

GPS Gaussian Teaser
Showcasing fidelity and real-time inference

Source: *Generalizable Pixel-wise 3D Gaussian Splatting for Real-time Human Novel View Synthesis*, Zheng et al.
Conference on Computer Vision and Pattern Recognition 2024

METHOD CANDIDATES GPS-Gaussian

High Quality
Delivers exceptional visual results

Real-Time Performance
Enables real-time reconstruction and rendering

Efficiency
Requires fewer cameras compared to other methods

Innovative Approach
Exemplifies cutting-edge methods on pixel-aligned 3DGS

***Generalizable Pixel-wise 3D Gaussian Splatting for Real-time Human Novel View Synthesis*, Zheng et al.**
Conference on Computer Vision and Pattern Recognition 2024

GPS-Gaussian

Dynamic Gaussian Splatting Pipeline for 3D Human Reconstruction

Capture

circular configuration
color + depth

→

Preprocessing

- color matching
- camera calibration
- instance segmentation
- depth refinement

→

Pixel Gaussians

partial reconstructions for each pair of neighbouring cameras

→

Merging

global reconstruction by merging the partial reconstructions

one set of 3D Gaussians per frame

- time consistency
- compression
- compatibility

→

Rendering

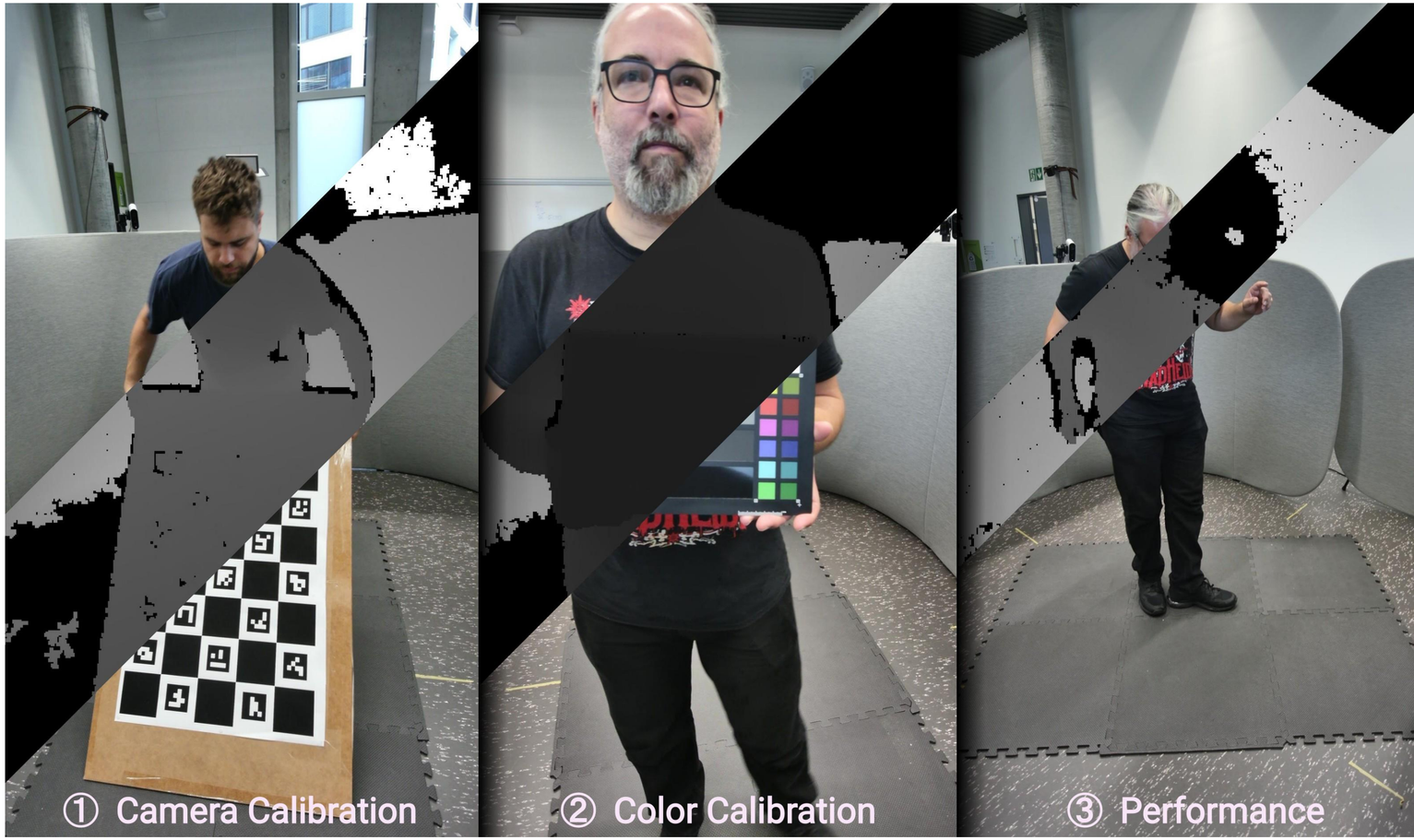
- Viewer
- Game Engine
- ...

First Round of Data Captures

September ‘24 Captures
- Setup consisted of 10×RGBD cameras
- 360° circular arrangement
- Minute-length, single-person performances

Calibration using Charuco / Davinci Boards

Processing of the data:
- Camera subset selection
- Depth filtering
- Reconstruction

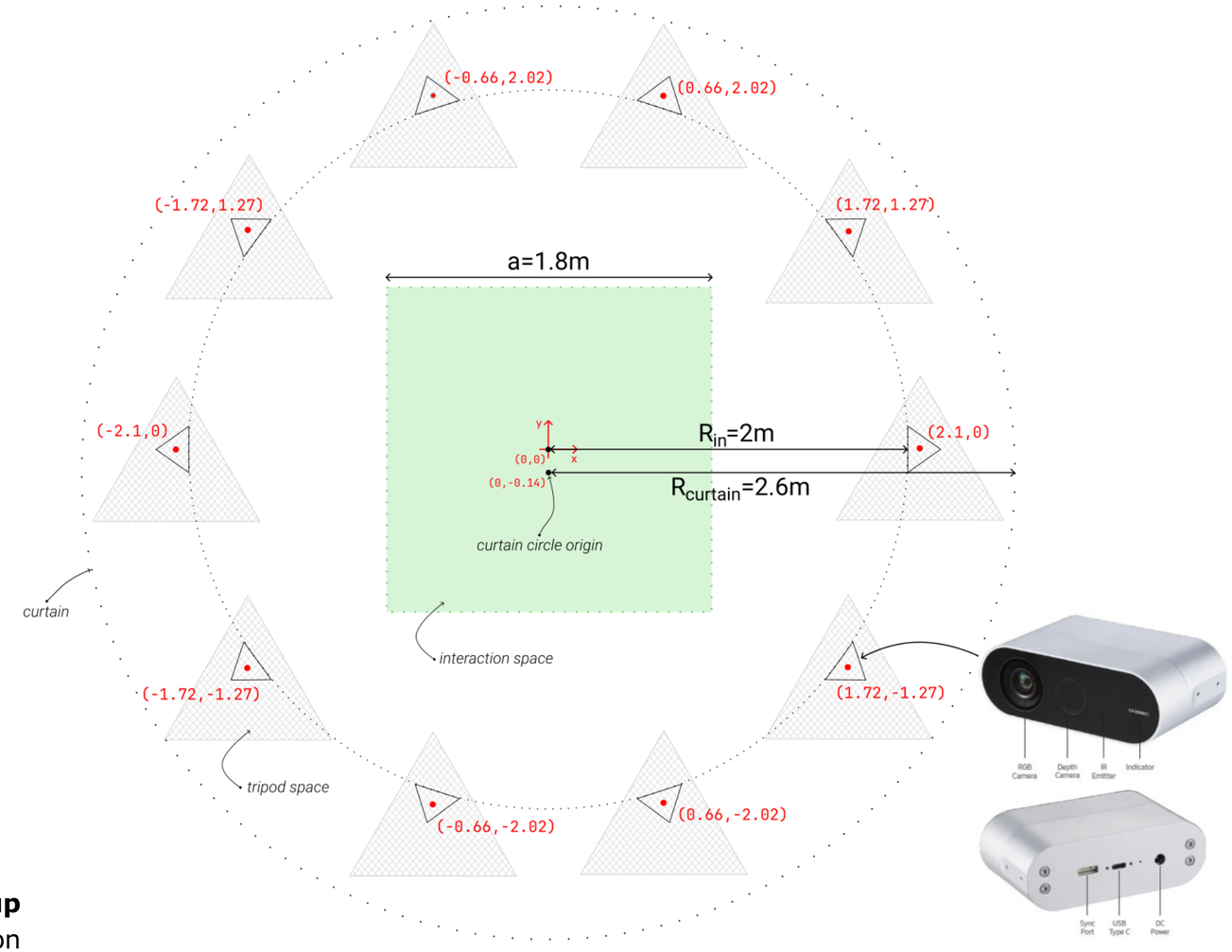


Capturing Steps
Calibration had to be re-done before every performance

First Round of Data Captures

September ‘24

- Setup consisted of 10×RGBD cameras
- 360° circular arrangement
- Minute-length, single-person performances



Initial Capturing Setup
10 Orbbec Fempto Bolt RGBD cameras in circular configuration

Outside

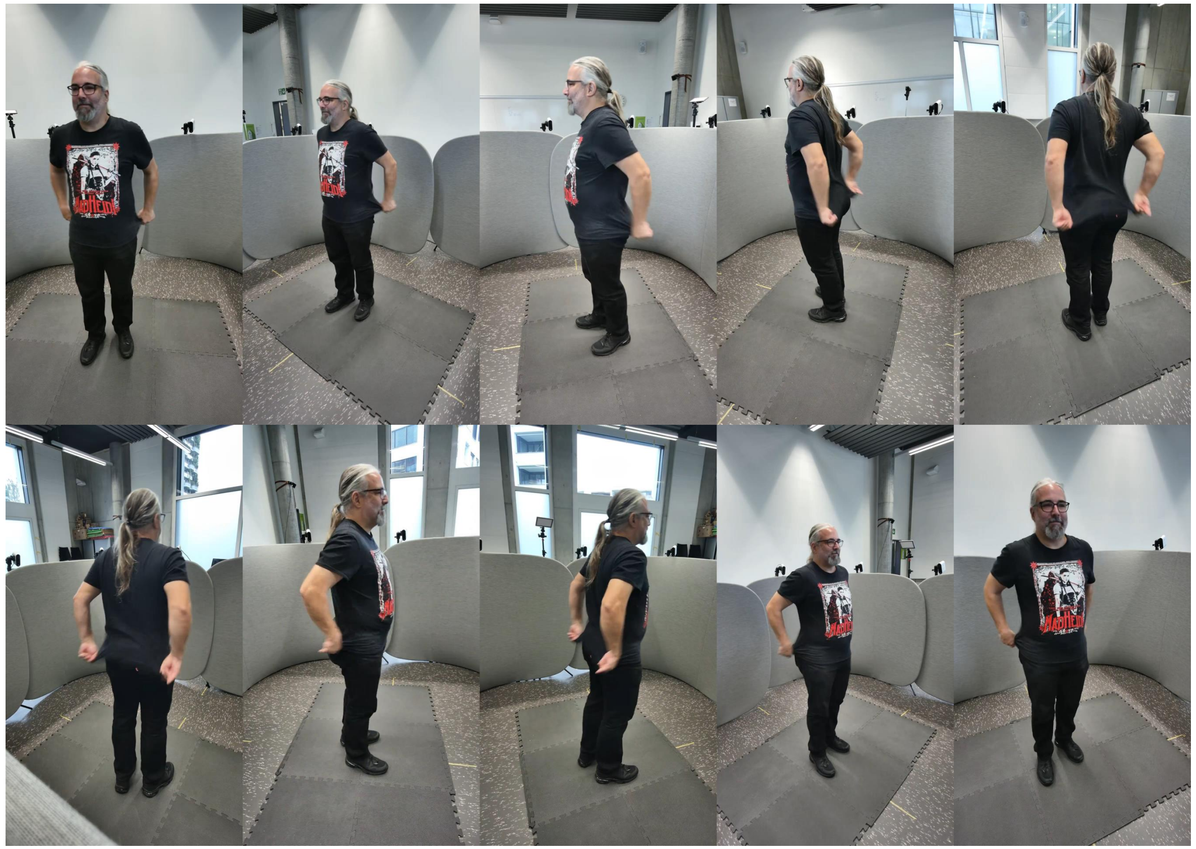
HSLU

RGB: 2K
Depth: 576p
(aligned)

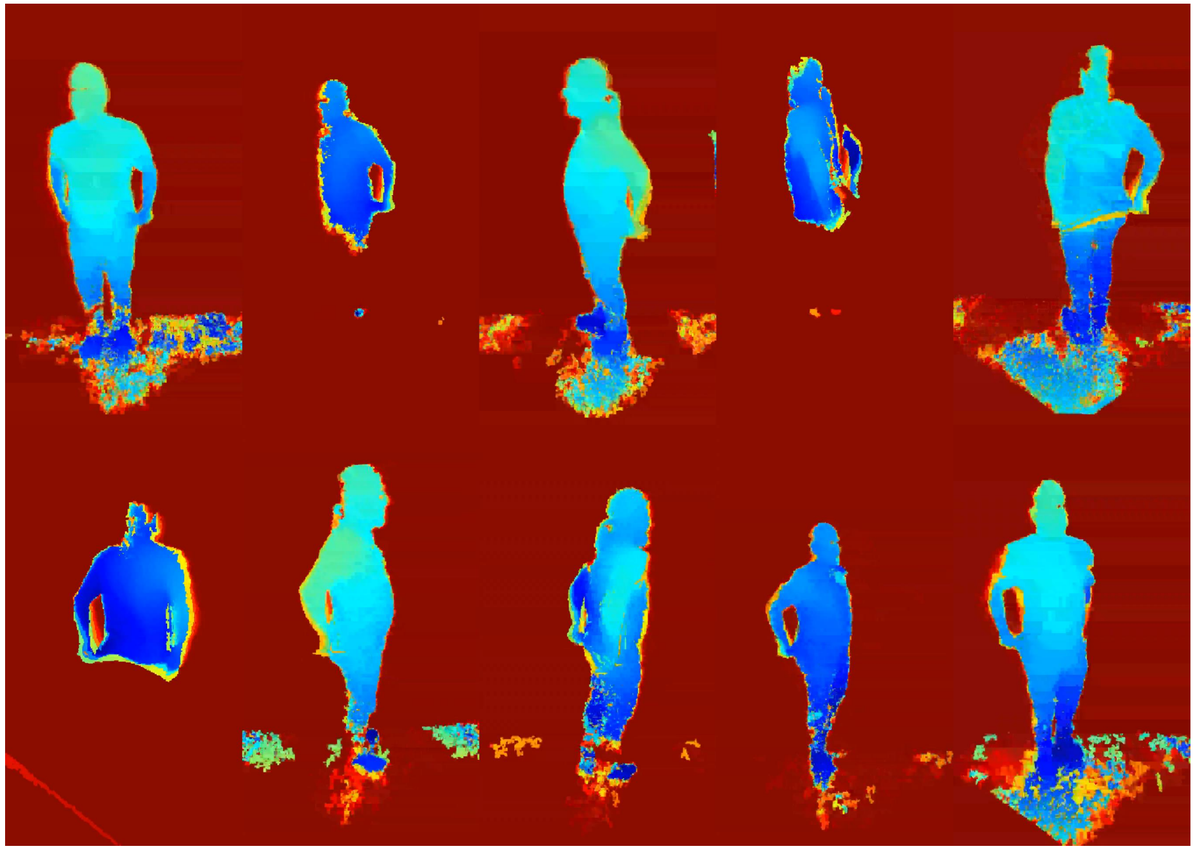
≈ 10 GBps
aggr. influx

SW PIPELINE

Raw Inputs



Multiview RGB Streams
HW/SW Synchronized



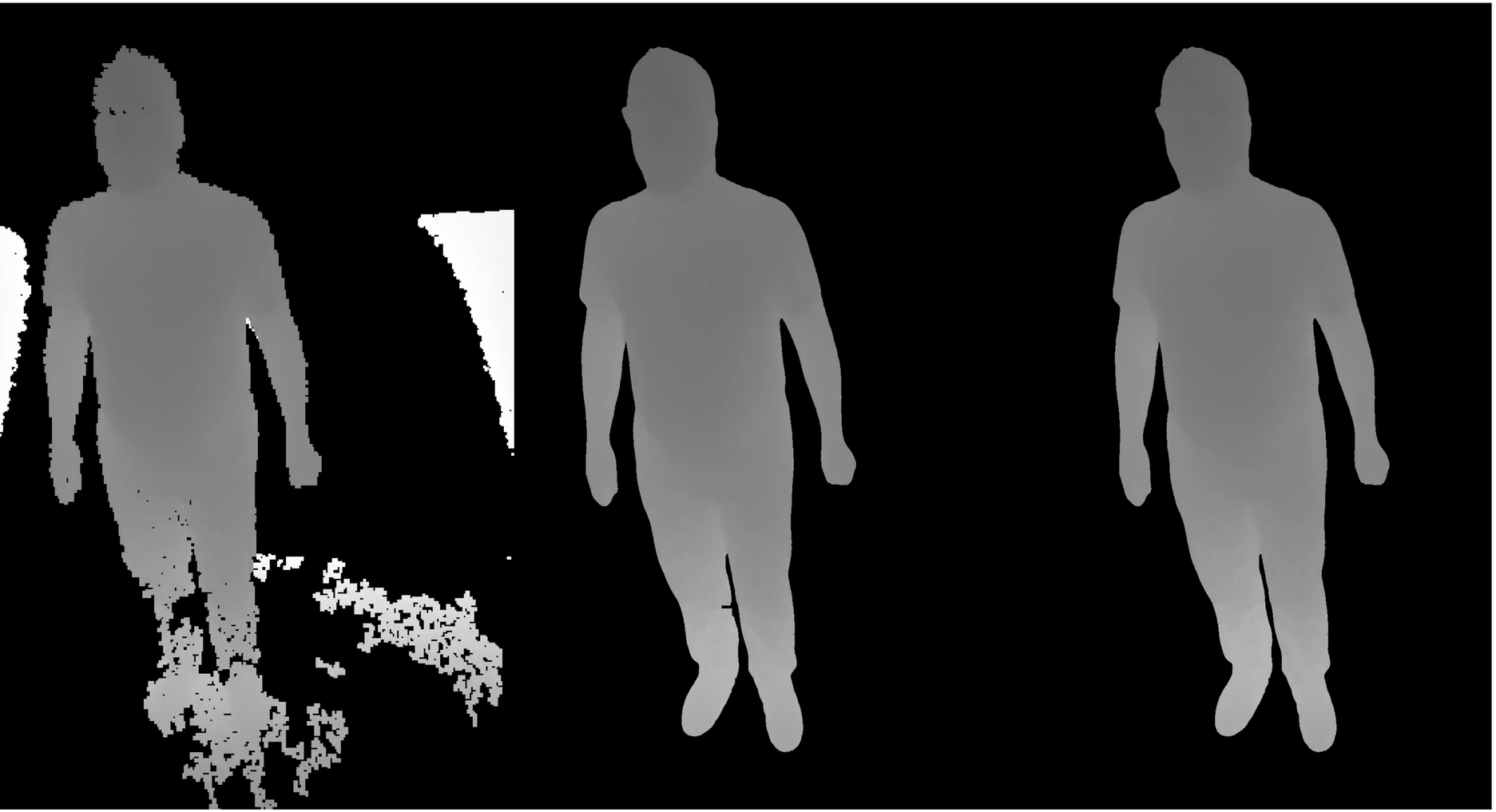
Multiview Depth Streams
Aligned to RGB, Filtered

Orbbec Femto Bolt – Improving Depth Signal

Depth flickering

Depth Filtering

Raw ***Color-guided Bilateral*** ***(Color & OF)-guided Bilateral***



Depth Refinement using Bilateral Filter
(left) Raw, (center) Spatial, (right) Spatiotemporal

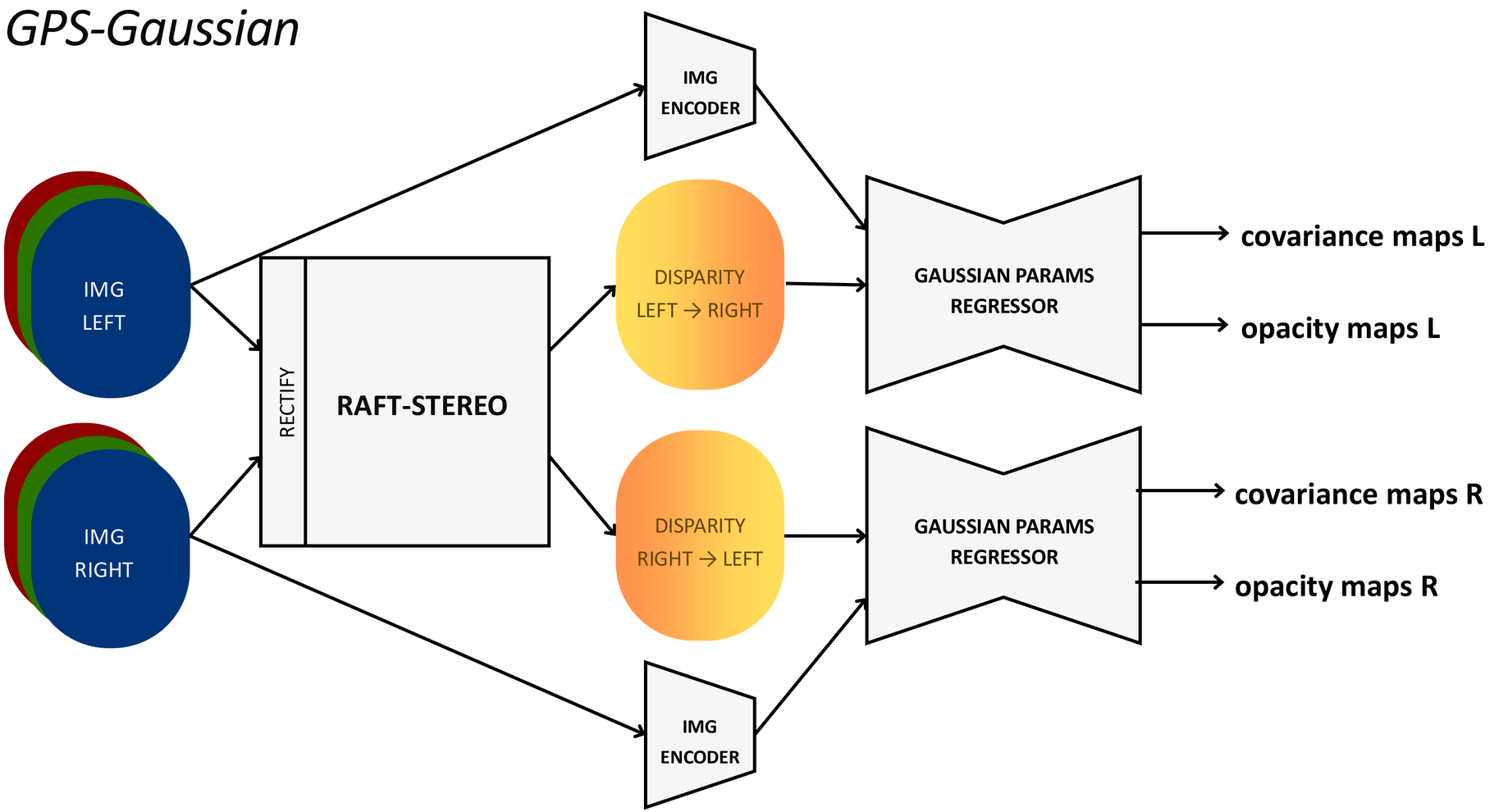
Reconstructions from Different Cameras

Cam 1 ***Cam 10*** ***Cam 10+1***

	PSNR (dB)	SSIM
CAM 1	29.06	0.941
CAM 10	29.52	0.961
CAM 10+1	27.23	0.932

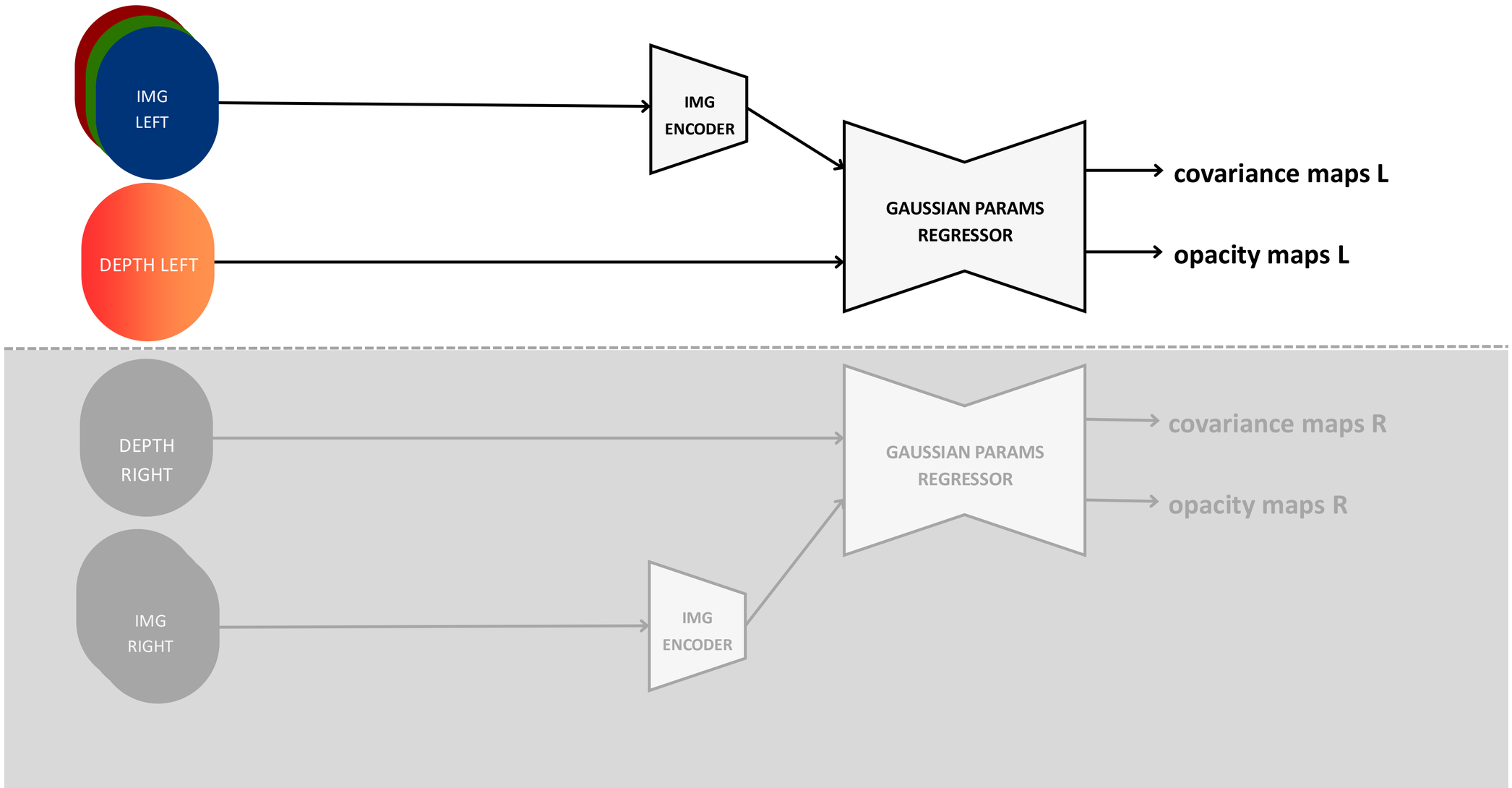
Cam 1, Cam 10, Cam 1+10
2 view reconstructions: single, and joint (right)

Inferring Pixel-wise Gaussians (no GT Depth)



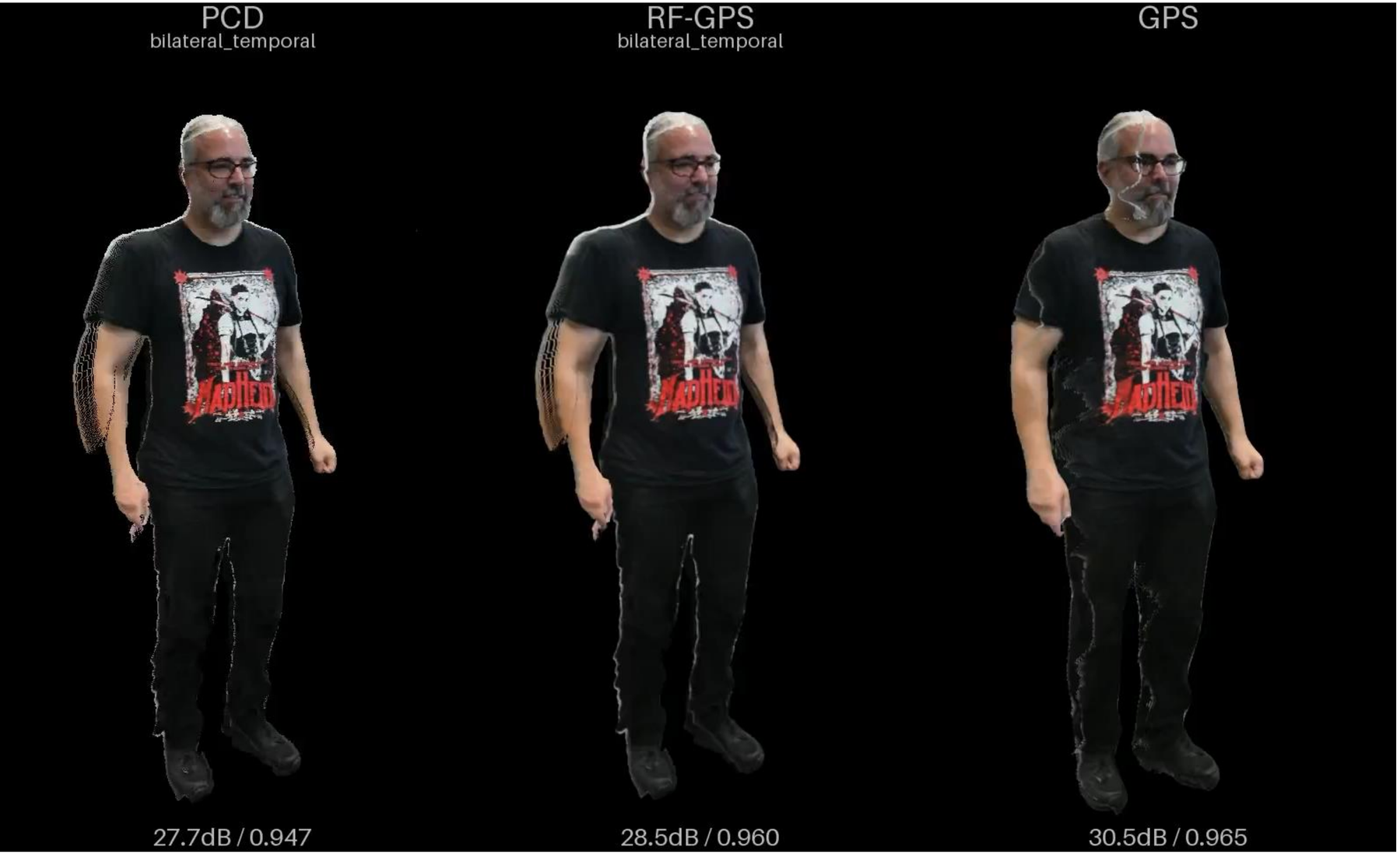
Inferring Pixel-wise Gaussians (w/ Depth)

RAFT-Free GPS-Gaussian (RF-GPS)



Reconstructions using PCs vs. Gaussians

POINT CLOUDS ***GAUSSIAN SPLATS***



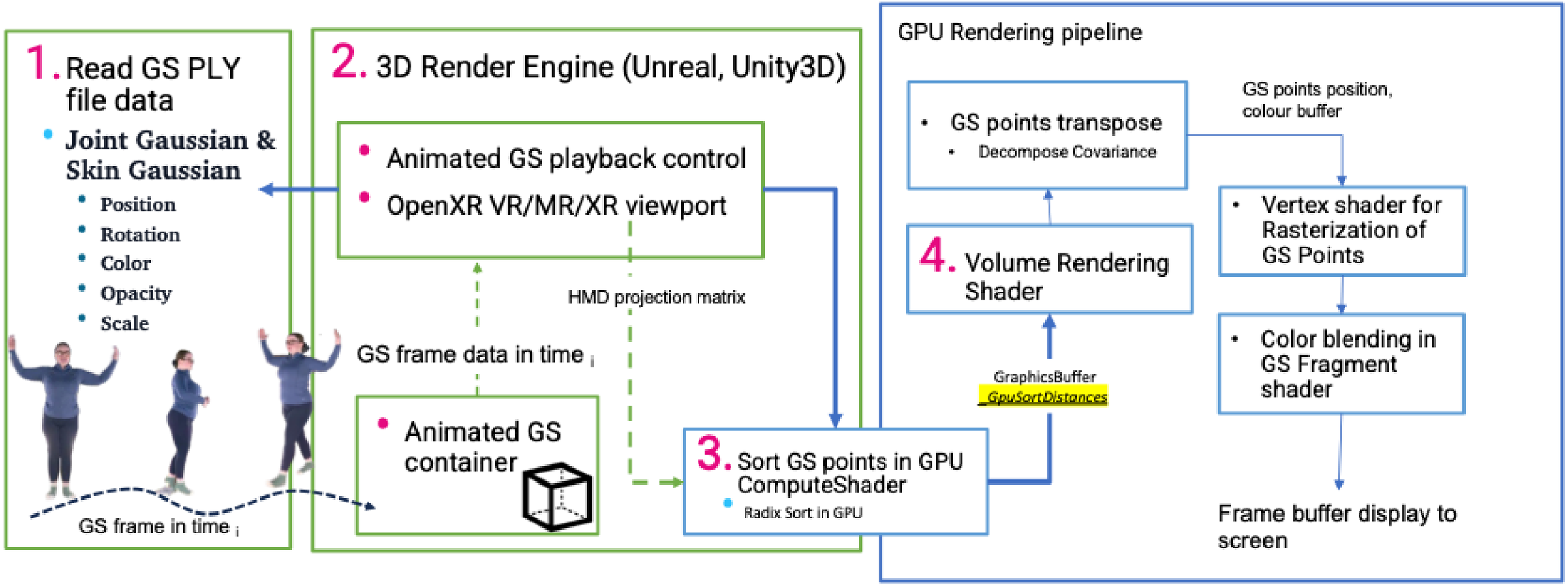
	PSNR (dB)	SSIM
PCD	27.23	0.932
RF-GPS	**32.30**	**0.975**
GPS	32.26	0.973

Different Representations from Cams 10+1
2 view reconstructions with PCs, and Gaussians

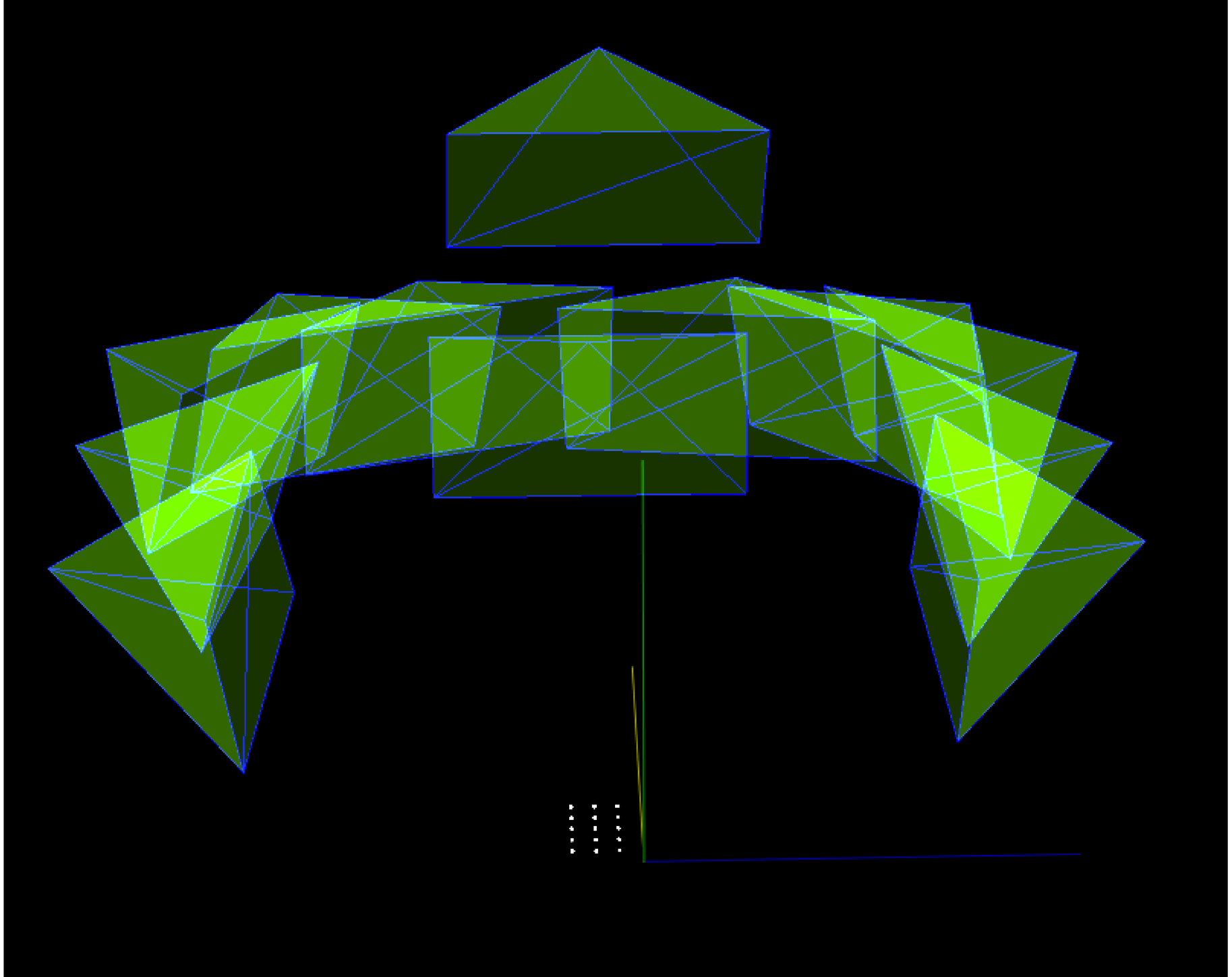
UNREAL
ENGINE

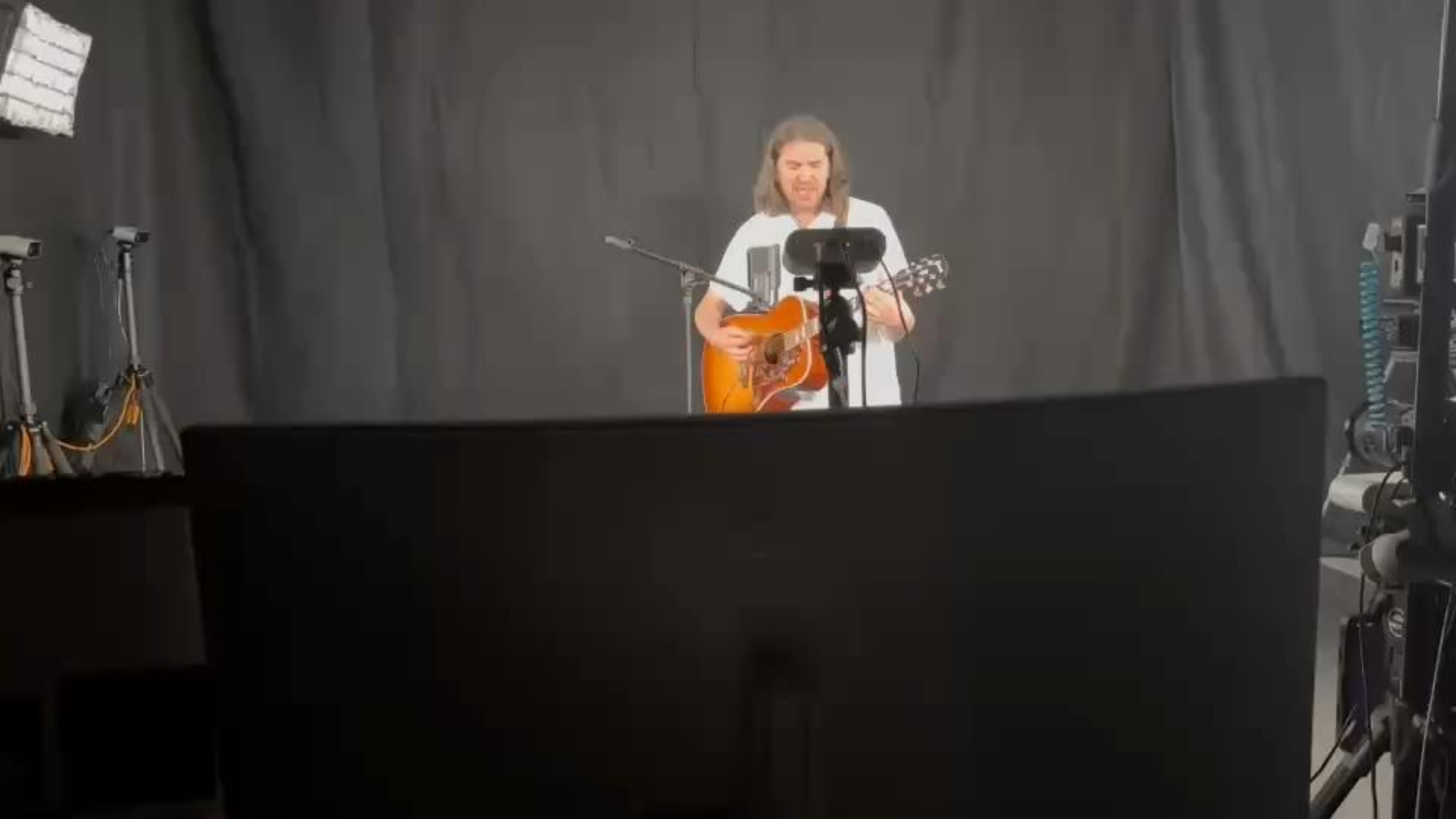
UNREAL
ENGINE

Dynamic Gaussian Spatting (GS) Rendering in Unreal/Unity




Rollei





HSLU

What’s Next?

- AI/ML has disrupted volumetric content creation and created significant new opportunities
- Further improvement/extension of reconstruction and visualization algorithms
- Integration with production pipelines and platforms (game engines)
- Editing, animation (format conversions, semantic Gaussians)
- Evaluation vintage vs. AI-based (e.g. photogrammetry vs. GS)
- Capture in-the-wild
- Formats and standards (efficiency and interoperability)
- End-to-end pipeline (coding, streaming, quality)
- Combination with Gen-AI and LLMs
- Development and evaluation of application scenarios
- …

HSLU Hochschule Luzern

VOLOGRAMS

Trinity College Dublin
The University of Dublin
V-SENSE

HEAT

FAVORE
FAST VOLUMETRIC RECONSTRUCTION

TRANSMIXR

Thank you!

aljosa.smolic@hslu.ch

Prof. Dr. Aljosa Smolic

Dept. Computer Science, Immersive Realities Center
21. April 2025