The Potential of AI for 3D Scene Digitization

–

Recent Learning-Based Innovations on Handling Challenging Scenarios

Michael Weinmann
Delft University of Technology, Netherlands

TU Delft

The Potential of AI for 3D Scene Digitization

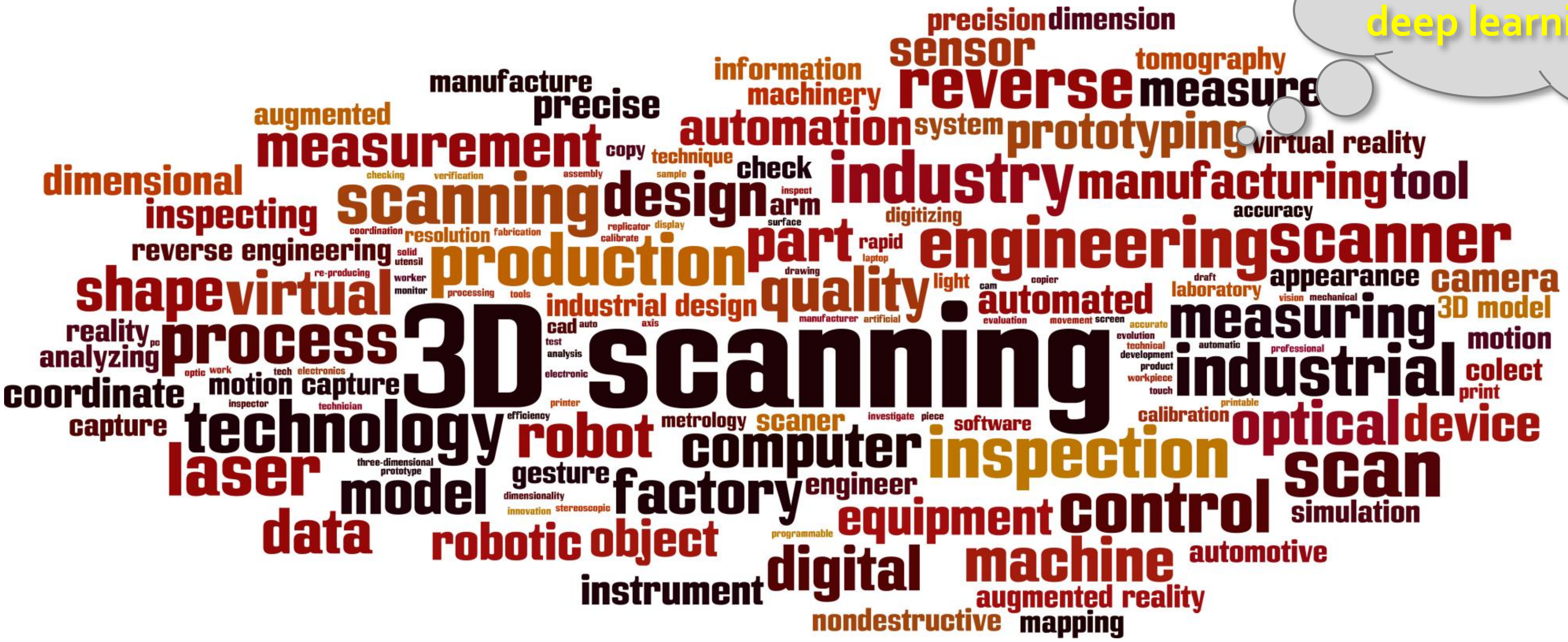
AI?

machine learning?

neural networks?

deep learning?

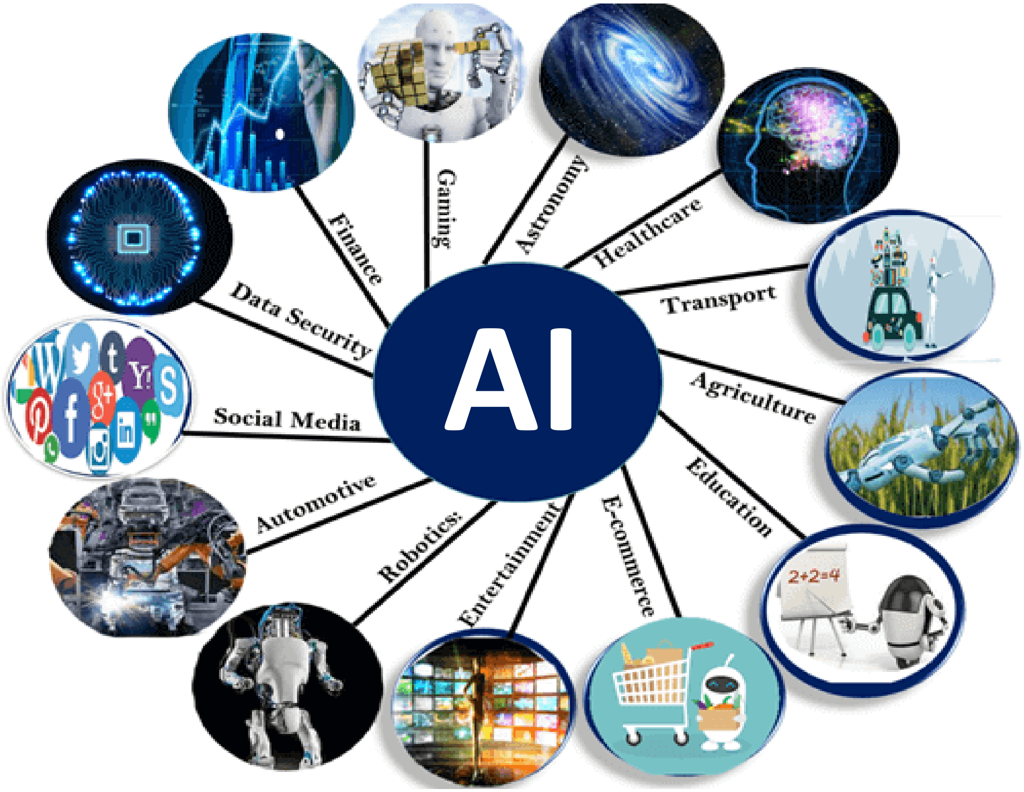
...



The Potential of AI for 3D Scene Digitization

Can machines think? – Alan Turing, 1950

- Replication/simulation of human intelligence in machines
  - *"Ability to learn and perform suitable techniques to solve problems and achieve goals appropriate to the context in an uncertain, ever-varying world"*
    – Manning, 2020



https://www.javatpoint.com/application-of-ai

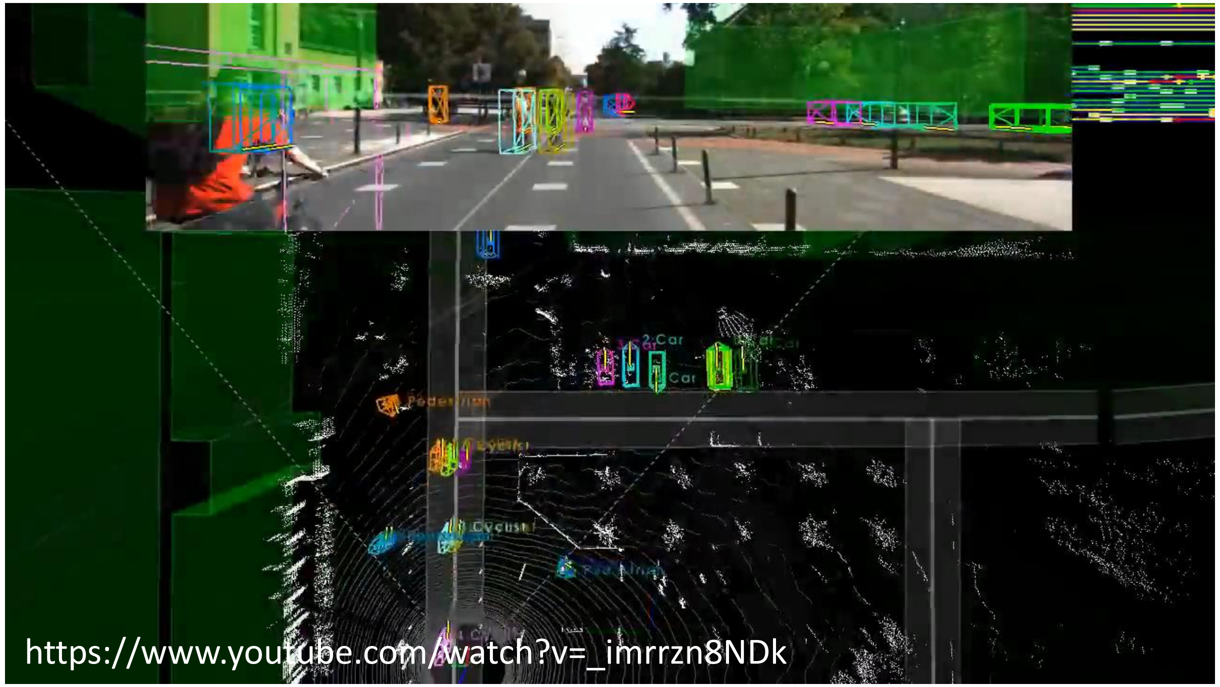
Why do we need digital scene models?

From (visual) navigation (e.g. Google Maps/Earth) …

Why do we need digital scene models?

... to scene understanding ...

- Based on scene geometry
- Based on spatial layout or arrangement of objects



autonomous driving

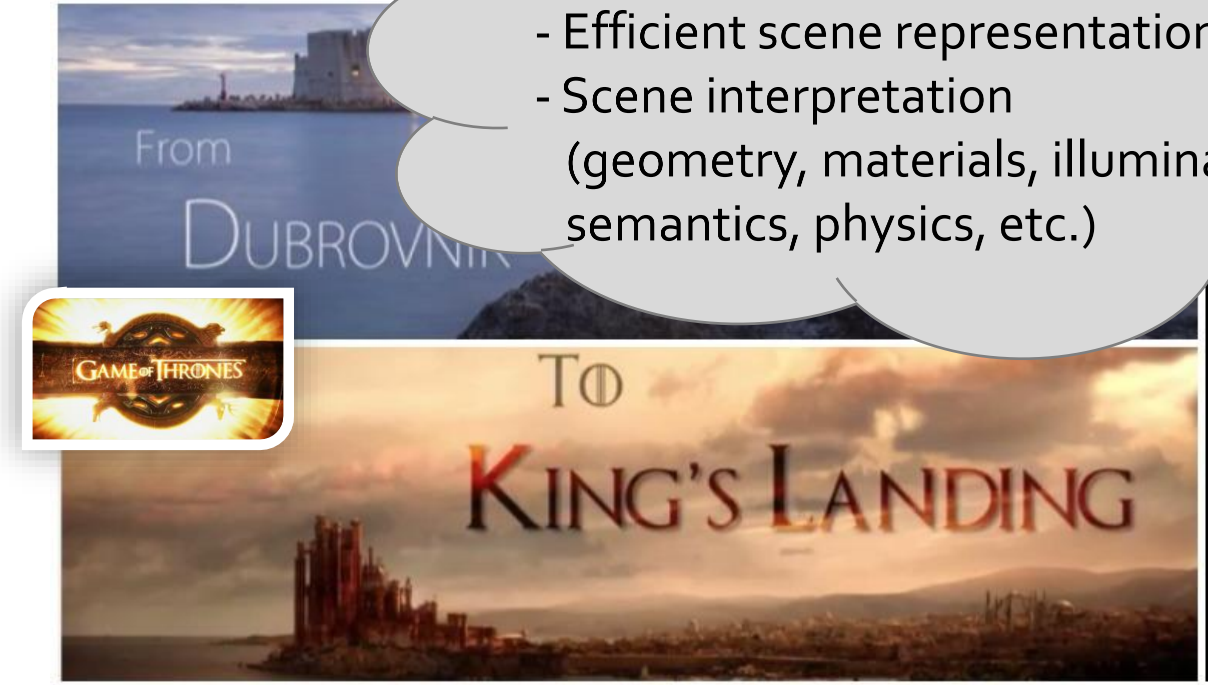
flood simulation

Why do we need digital scene models?

… and scene manipulati

- Modificati

Challenges:
- Accurate reconstruction/modeling/visualization
- Efficient scene representations
- Scene interpretation
  (geometry, materials, illumination, semantics, physics, etc.)



https://www.thedubrovniktimes.com/news/dubrovnik/item/1765-amazing-video-from-dubrovnik-to-king-s-landing?fb_comment_id=1072838662825020_1123759387732947

Where to place the object (ground inference)?
How does lighting influence object appearance?

Why do we need digital scene models?

“Holodeck” experience

- Telepresence: „Subjective experience of being in an environment that may differ from the user’s actual local physical surrounding”
    - Ideally multi-modal immersive experience
    - Beyond standard displays

Images partially taken from presentermedia.com

https://www.youtube.com/watch?v=kzNVkc4gB6U

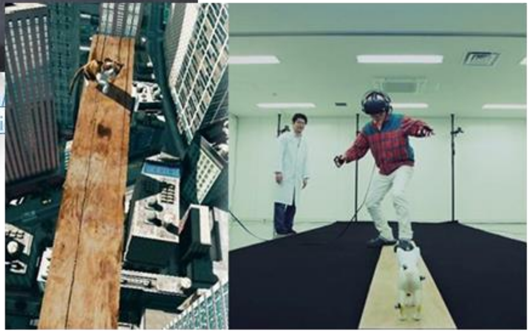
Why do we need digital scene models?

https://xrsweek.com/2017/08/vrs-focus-vr-ar-design/

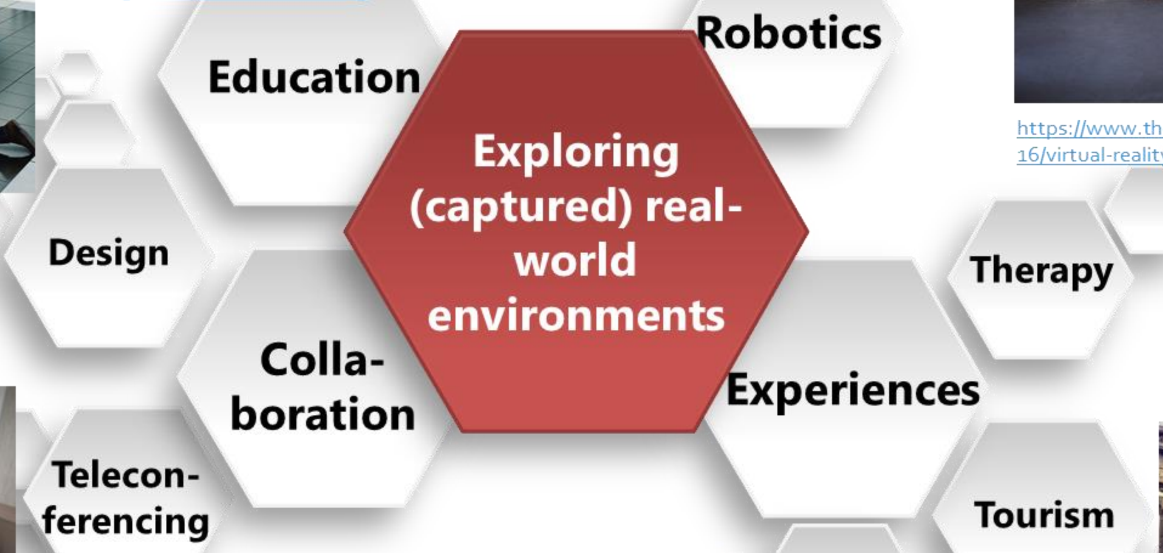
https://visualise.com/2017/11/education-vr-5-examples-bending-reality-enhance-learning

http://news.mit.edu/2017/mit-csail-new-sy -teleoperating-robots-virtual-reality-1009

https://www.theguardian.com/ 16/virtual-reality-phobias-publi



https://grapee.jp/en/60459



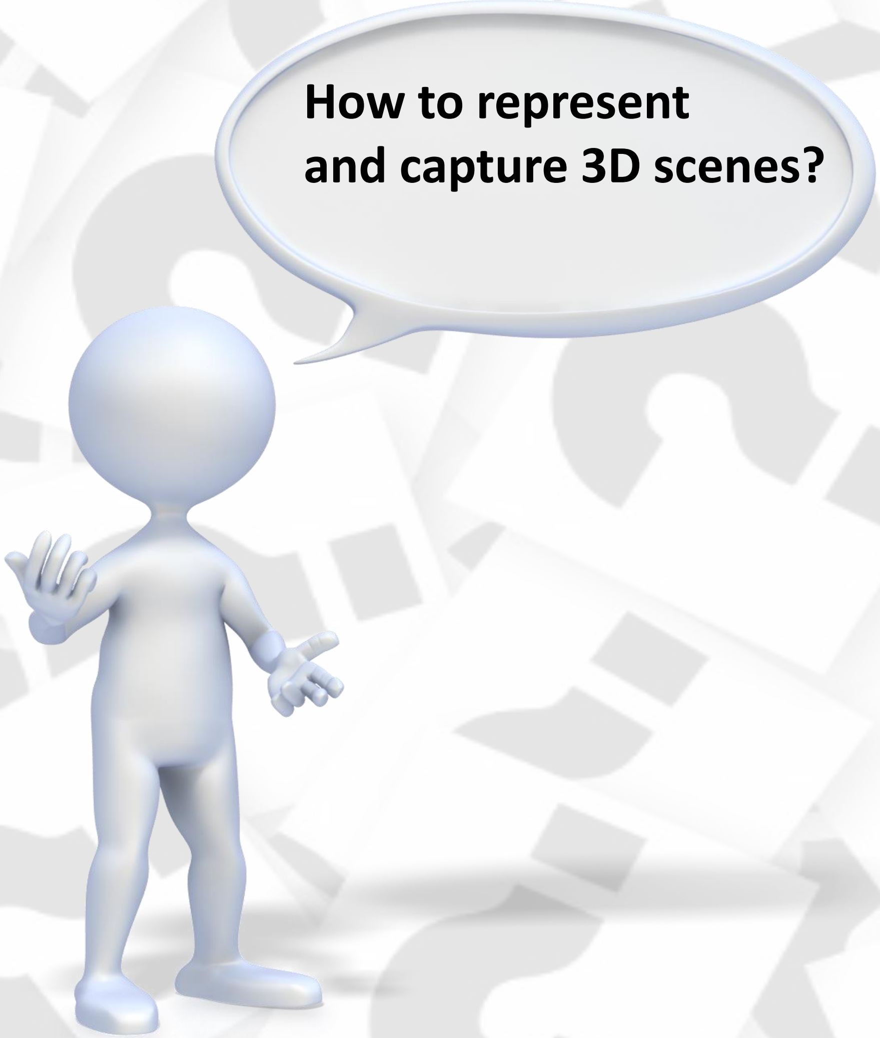
Holoportation [Orts-Escolano et al. 2016]

https://medium.com/arjs/why-web-apps-are-the-future-of-augmented-reality-c503e796a0c5

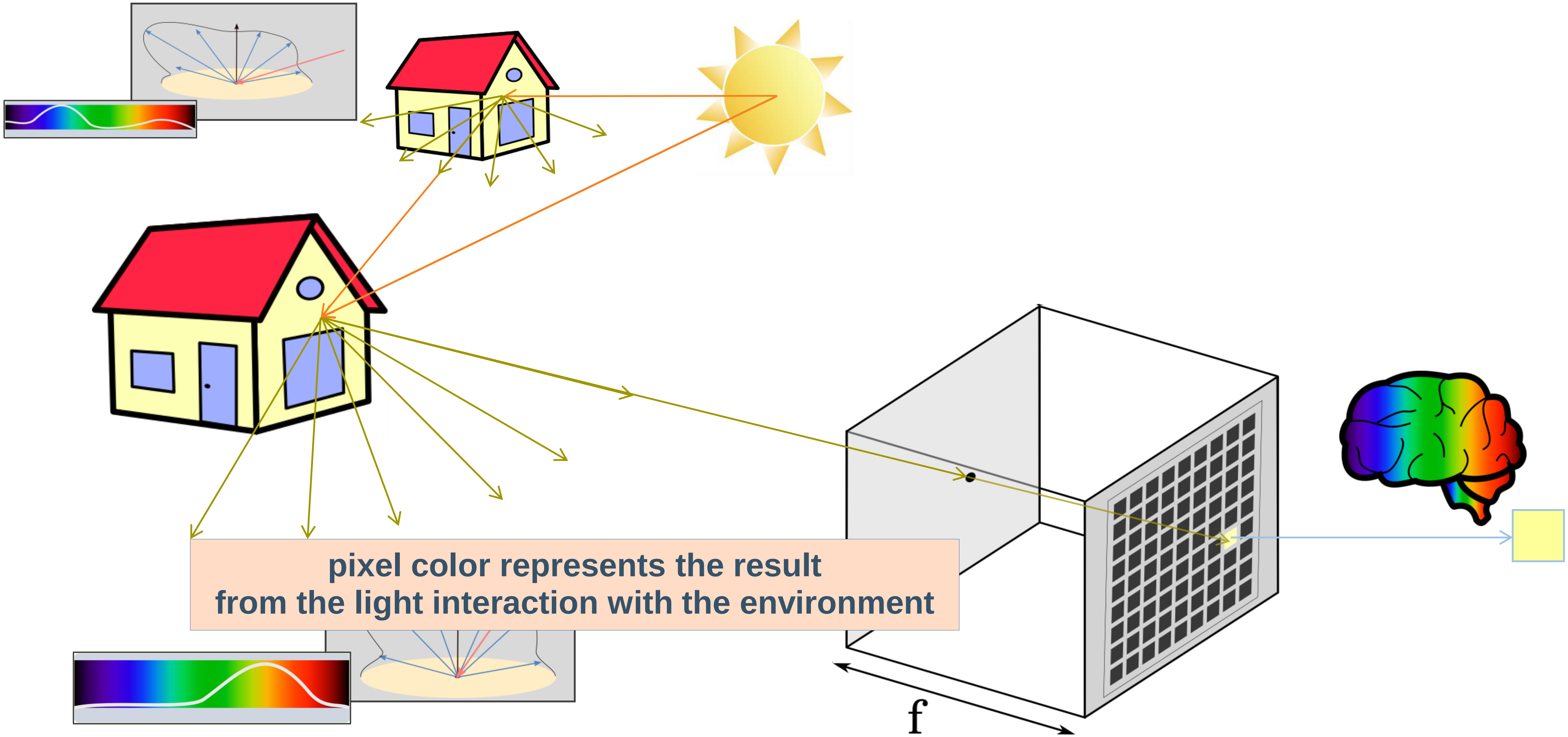
https://scooterise.com/modern-way-exploring-ancient-monuments/

https://medium.com/frulix/virtual-reality-emerging-backbone-of-tourism-industry-7b6000a526c0

How to (digitally) represent scenes?



What do we see?



pixel color represents the result
from the light interaction with the environment

What do we see?

Color snapshot

- Intensity of light
  - Seen from a single view point
  - At a single time
  - As a function of wavelength

Figure by Leonard McMillan

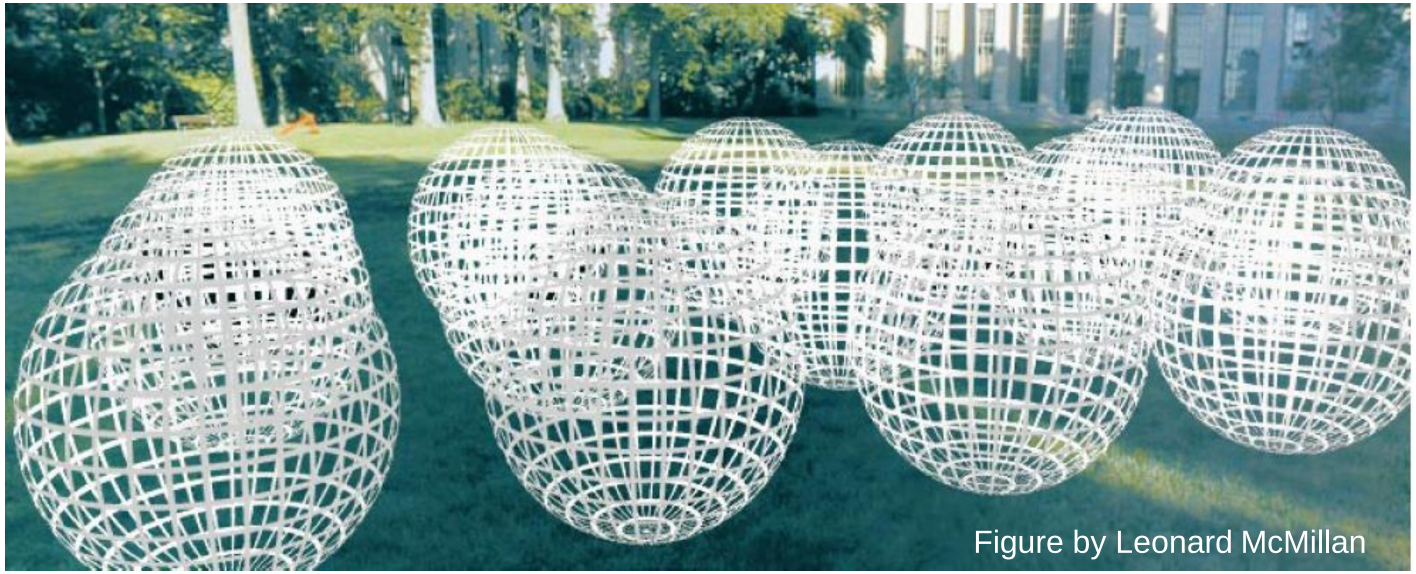
Slide credit: A. Efros

What do we see?

The plenoptic function

- Reconstruction of scene appearance under every possible view, at every moment, from every position, at every wavelength

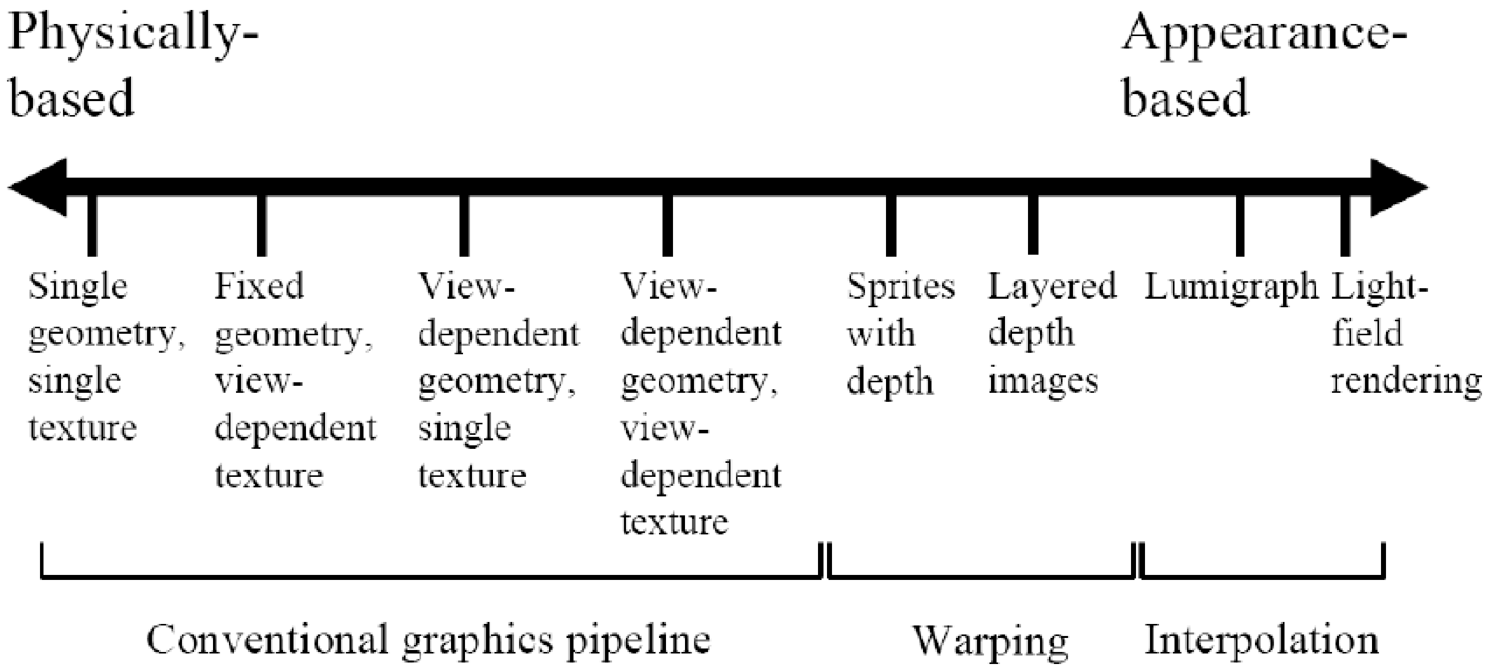
→ It completely captures our visual reality!



Slide credit: A. Efros

Model geometry or just capture images?

Geometry-image continuum …

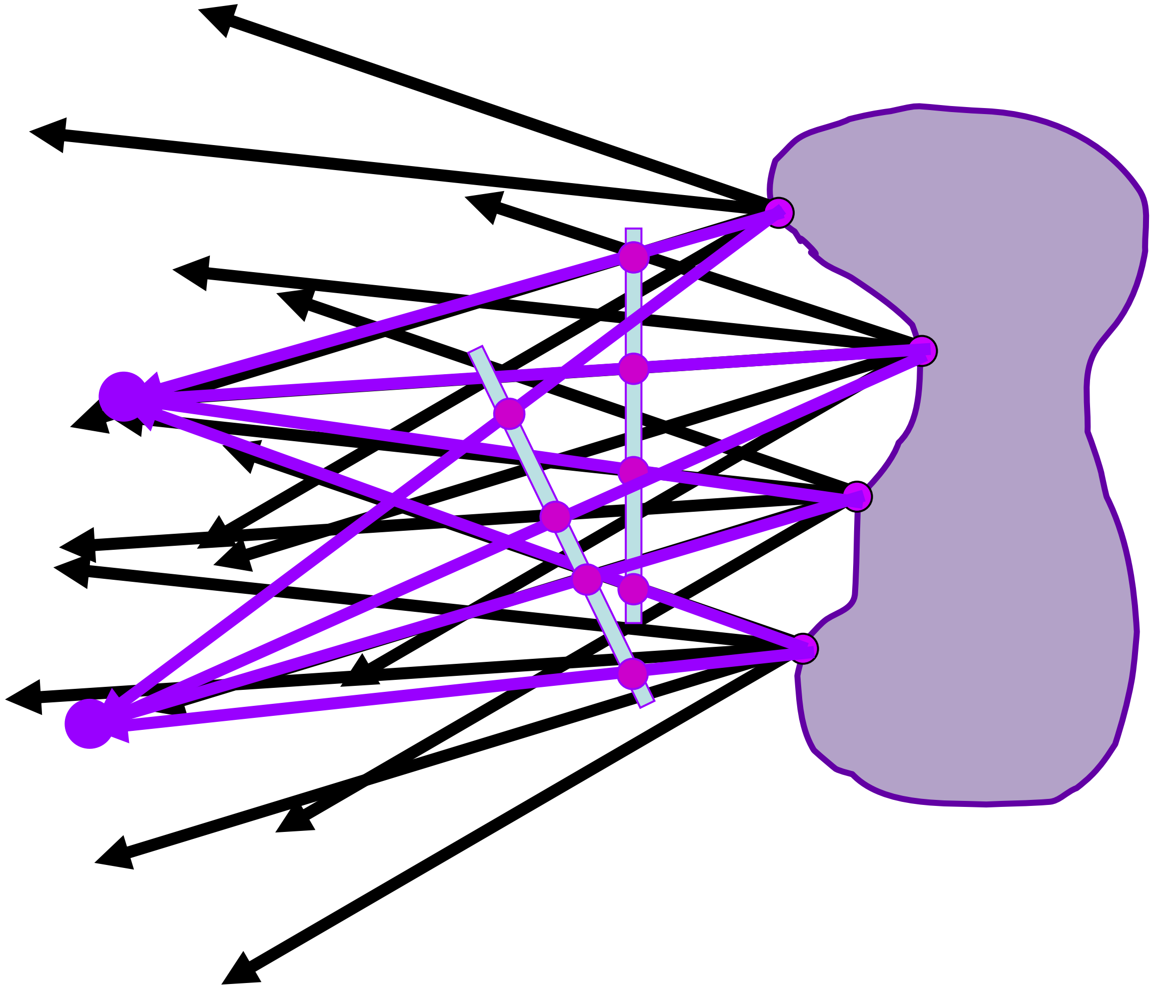


Model geometry or just capture images?

Reduce complexity?

- Assumption:
  known surfaces

→ (traditional)
computer graphics



Slide by Rick Szeliski and Michael Cohen

Traditional approach: Geometry + Reflectance

Background: 2D image synthesis (via rendering)

- Underlying principle: ray tracing (standard technique in graphics)
- Idea:
  - Measure light that arrives in the camera image
  - „Shooting“ rays from a viewpoint through an image grid into the 3D scene
  - Intersection with geometric structures (e.g., ray-triangle intersection for meshes)
  - Color assignment (per pixel) based on evaluation of reflectance model

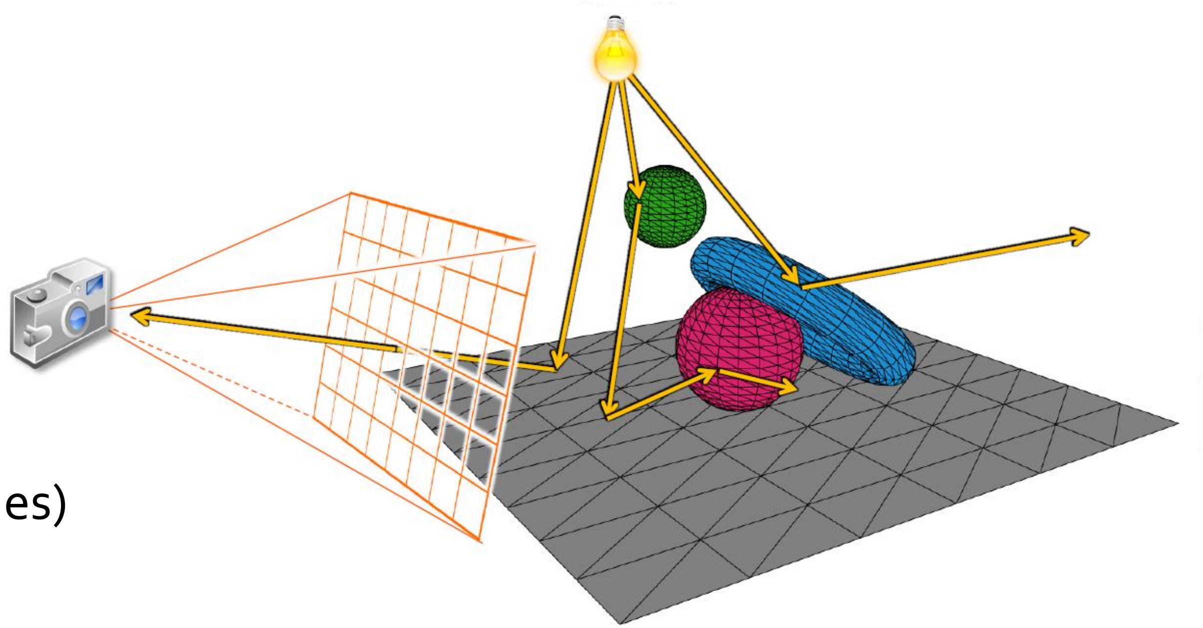
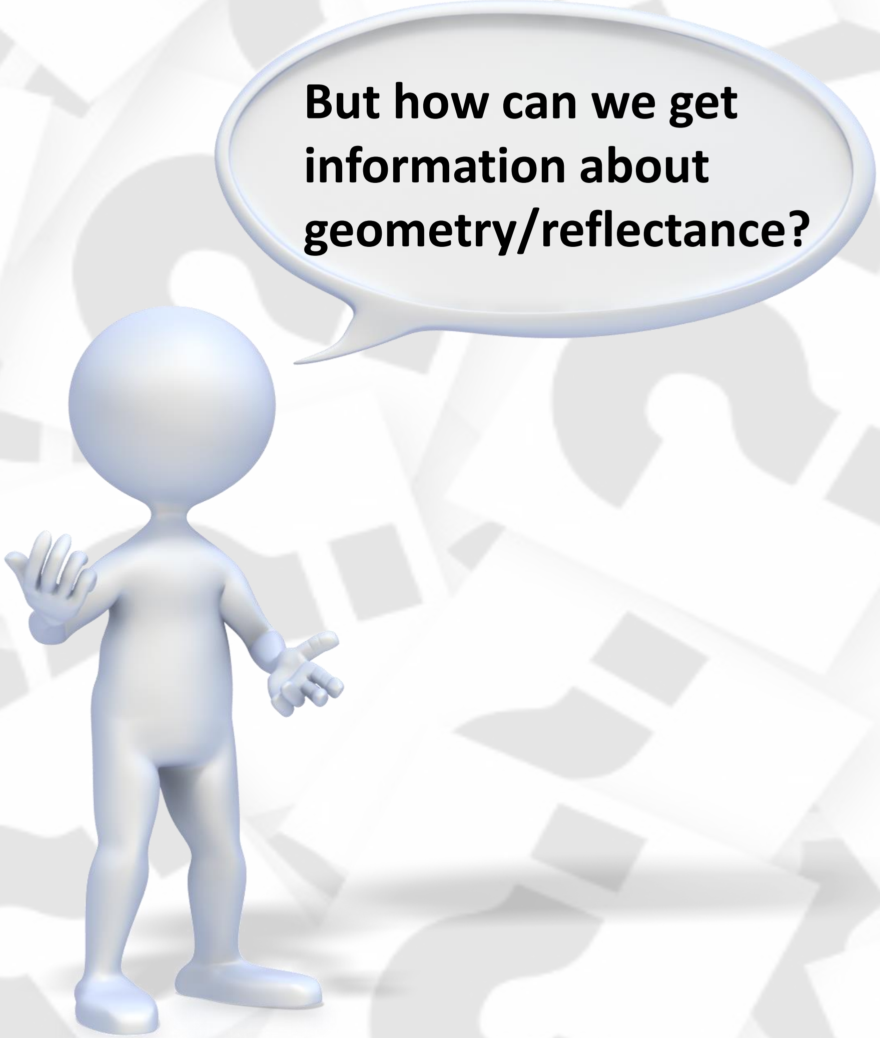


Image credits: C. Dachsbacher

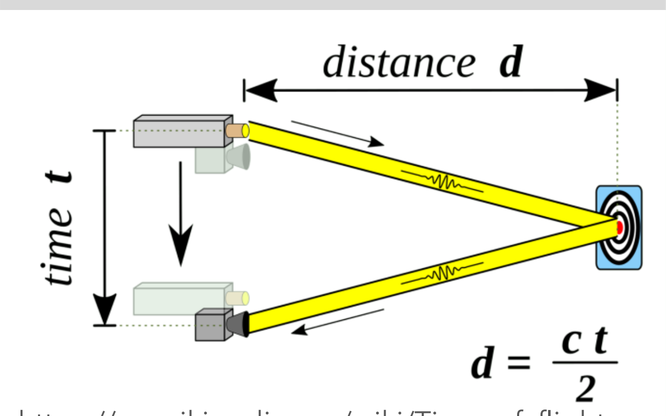
How should we represent complex 3D scenes?



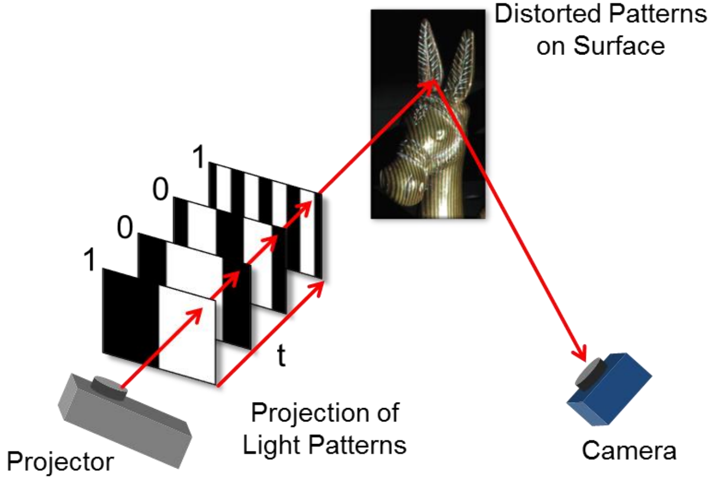
Conventional 3D Scanning

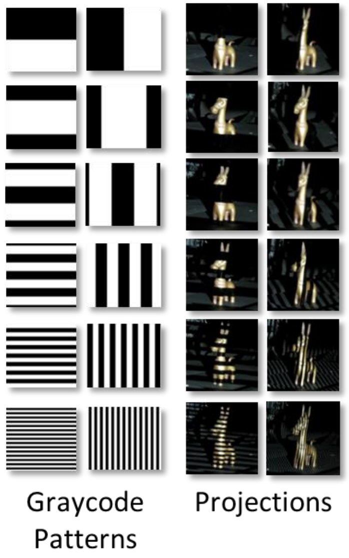
Separate 3D scanning with classical techniques:

- Structured light systems (active)
- Laser scanners (active)
- Multi-view stereo (passive)

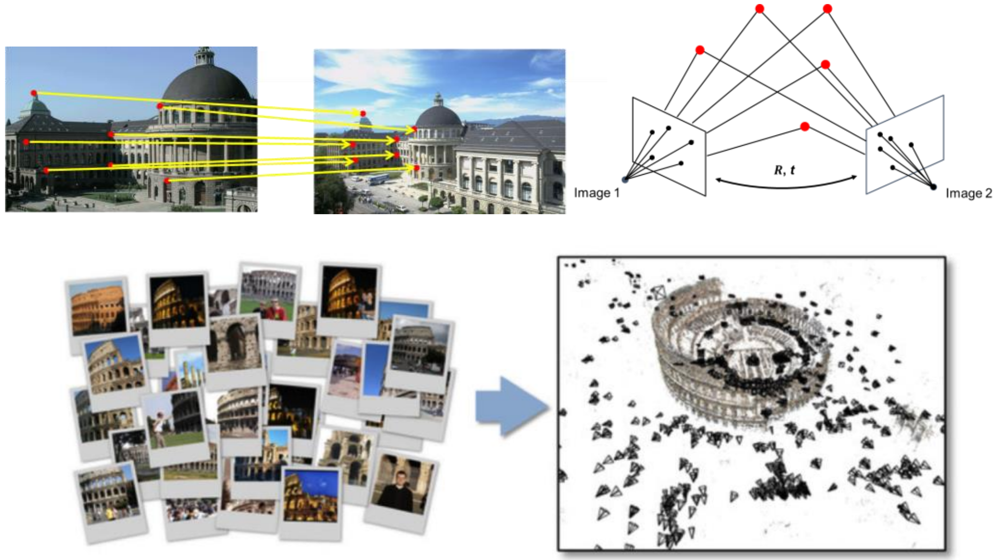


https://en.wikipedia.org/wiki/Time_of_flight





M. Weinmann, C. Schwartz, R. Ruiters, and R. Klein. *A Multi-Camera, Multi-Projector Super-Resolution Framework for Structured Light.* 3DIMPVT, 2011

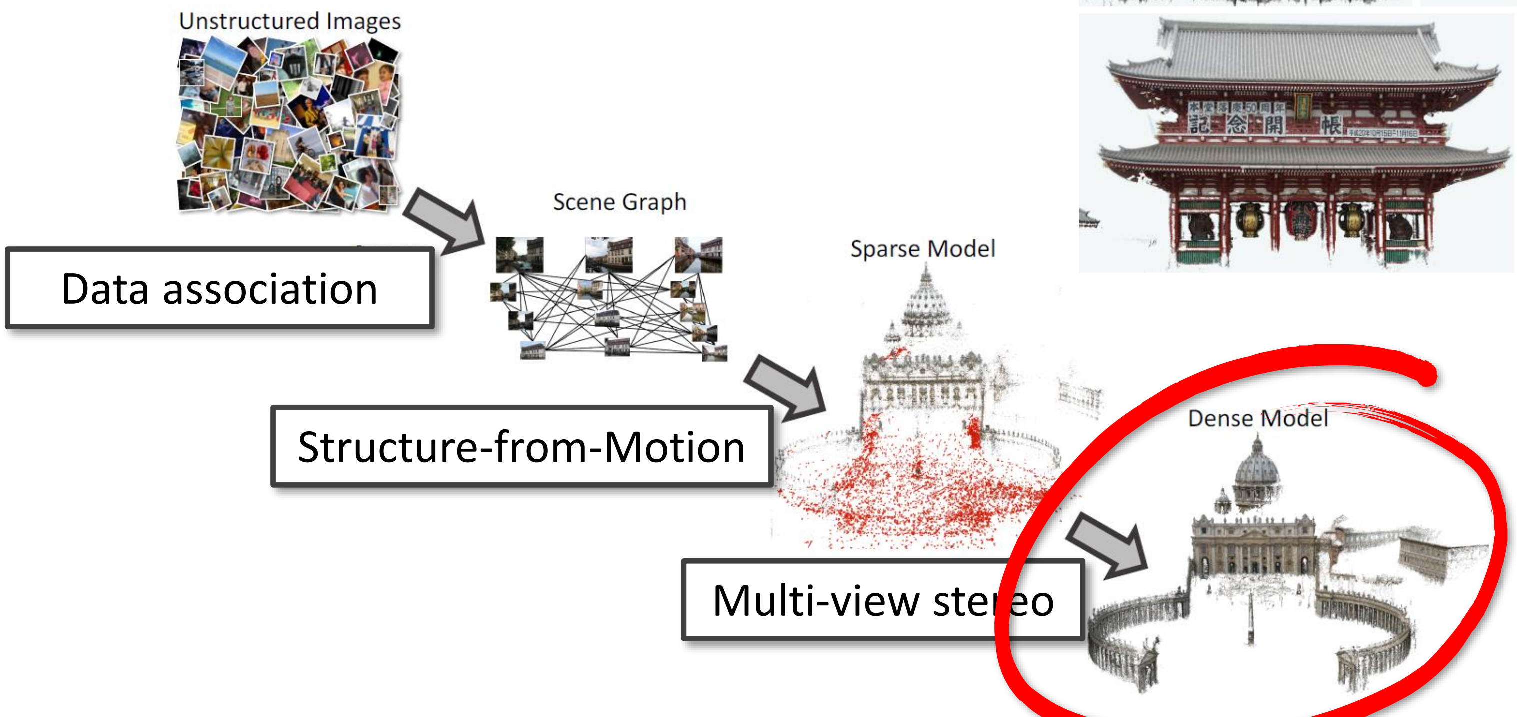


McCann, 3D Reconstruction from Multiple Images

Conventional 3D Scanning

Example:

- SfM + MVS
  (offline reconstruction)

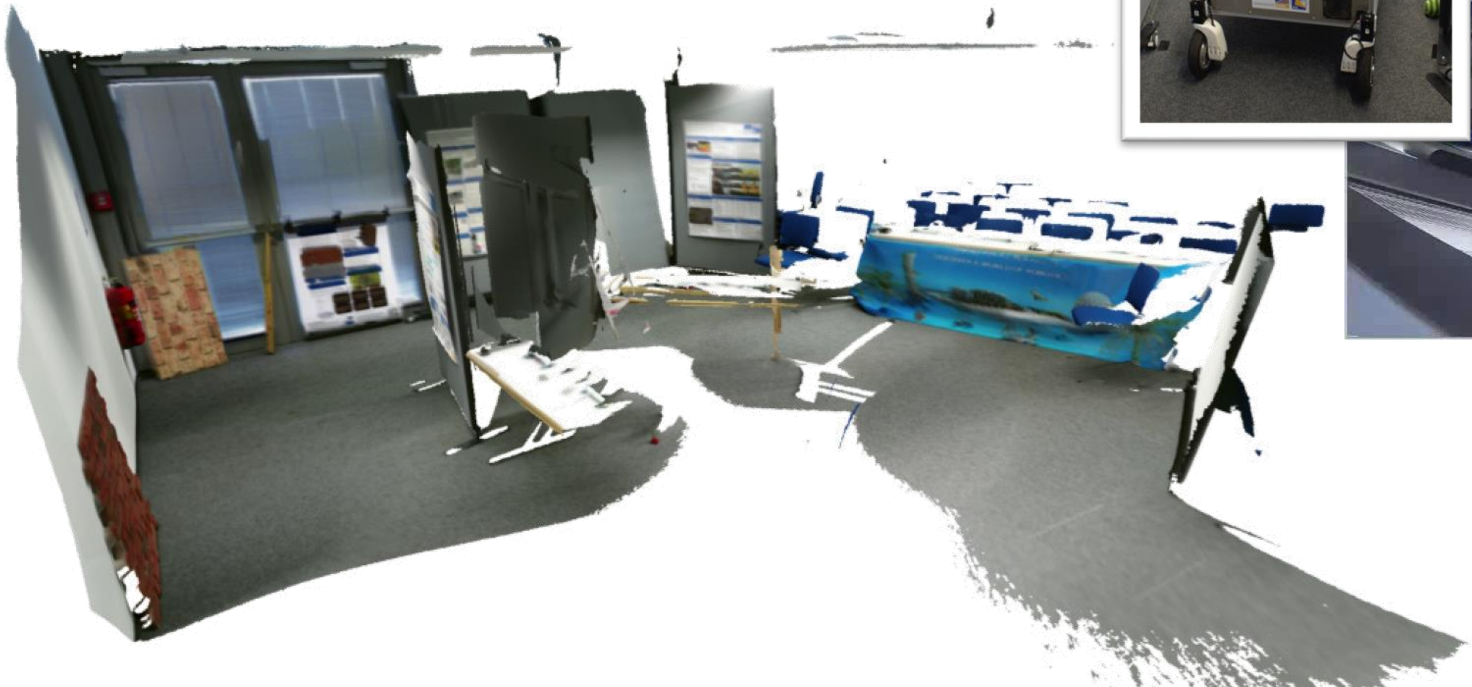


Frahm et al. (2016): Large-Scale 3D Modeling from Crowdsourced Data

Conventional 3D Scanning

Example:

- Visual SLAM
  (online/real-time reconstruction)

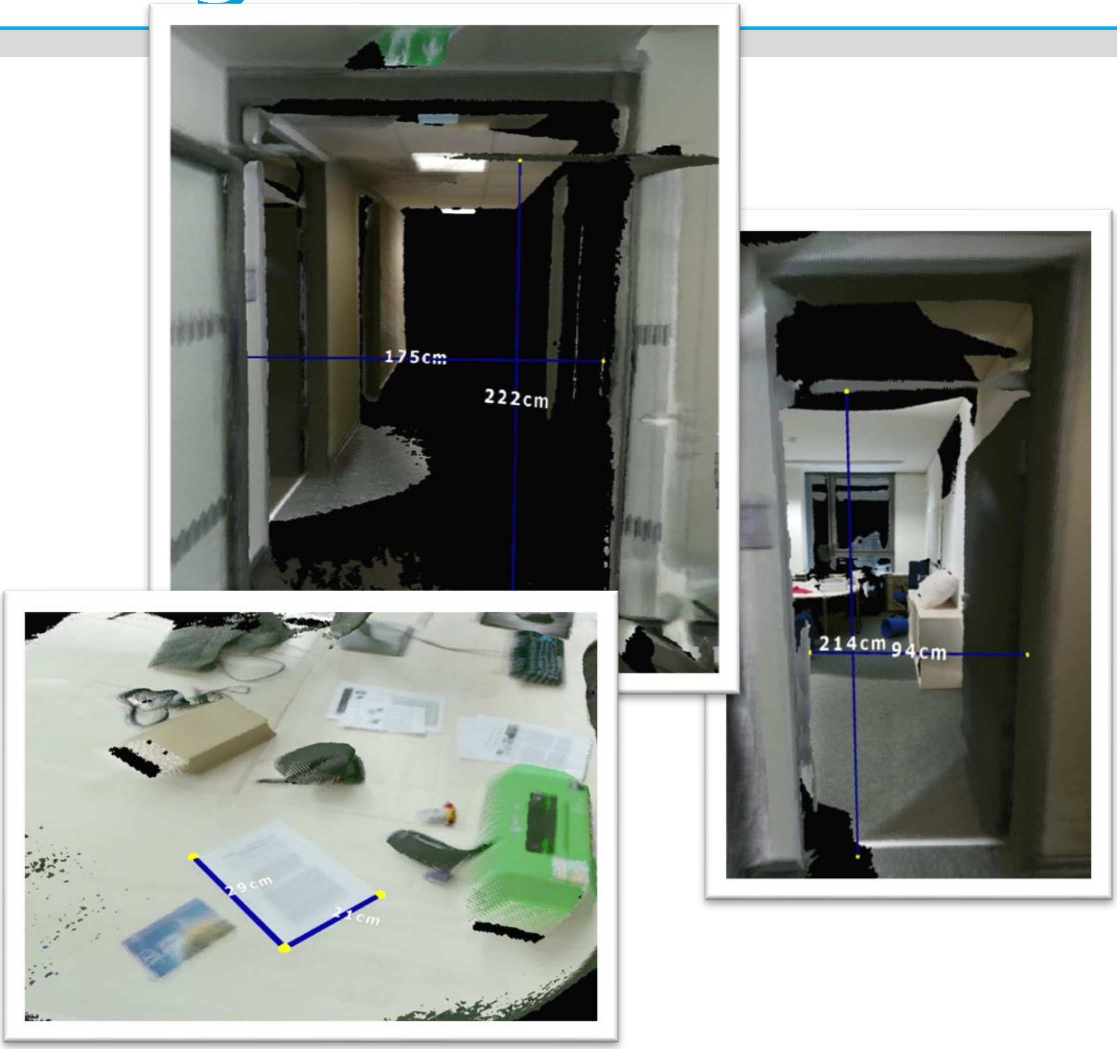


VR-based telepresence/teleoperation [TVCG'19, ISMAR'19, 3DV'19, CVPRW'19, IROS'19, ICCVW'23]

Conventional 3D Scanning

Example:

- Visual SLAM
  (online/real-time reconstruction)



VR-based telepresence/teleoperation [TVCG'19, ISMAR'19, 3DV'19, CVPRW'19, IROS'19, ICCVW'23]

Classical 3D Scanning Techniques

Separate 3D scanning with classical techniques

https://www.youtube.com/watch?v=A5c1ouRSEkg

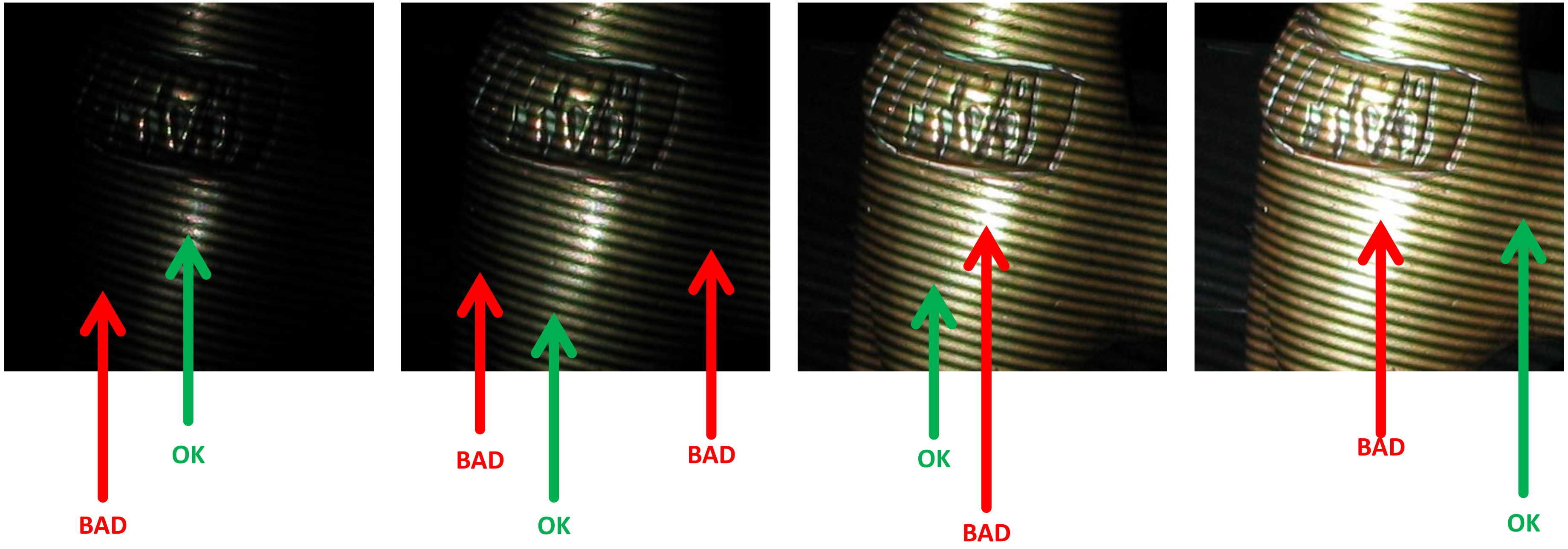
https://www.youtube.com/watch?v=XDuY9Qe4WPs

https://www.youtube.com/watch?v=sQmIxPVtOGk

Conventional 3D Scanning Techniques

Challenges:

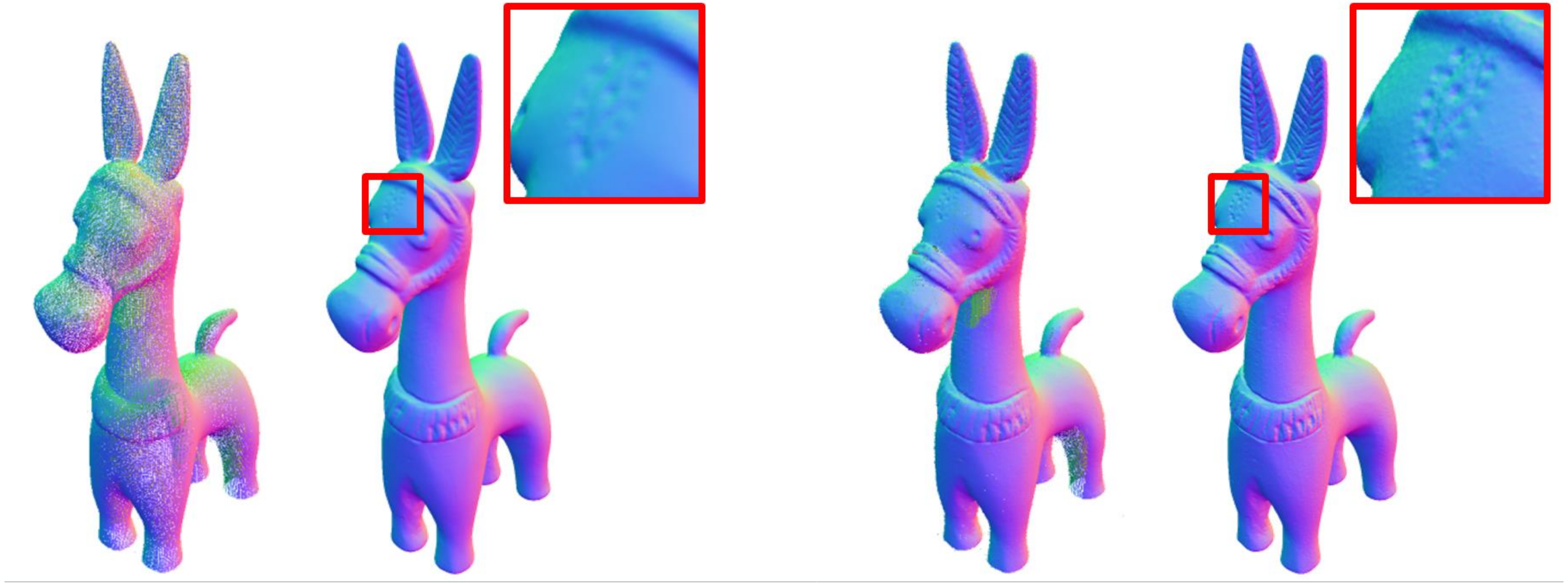
- Dynamic range



Conventional 3D Scanning Techniques

Challenges:

- Dynamic range
- Device resolution



Conventional 3D Scanning Techniques

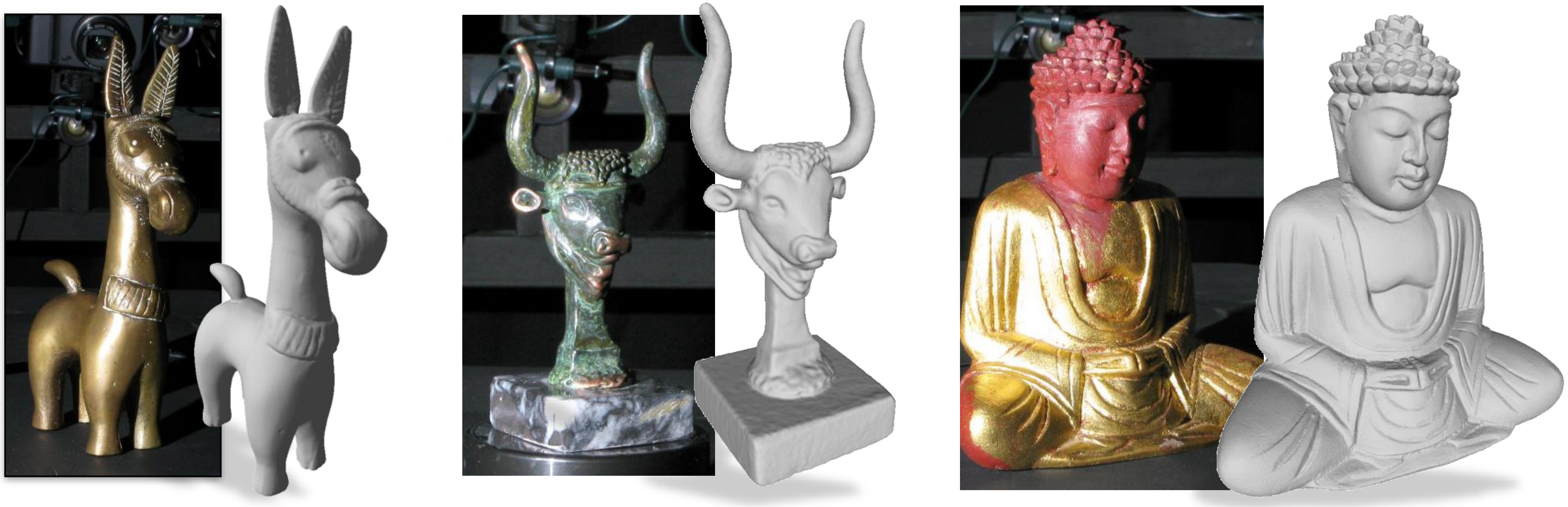
Challenges:

- Dynamic range
- Device resolution
- Optically complicated materials

Advanced 3D Scanning Techniques

Overcome challenges of classical techniques based on:

- HDR Scanning

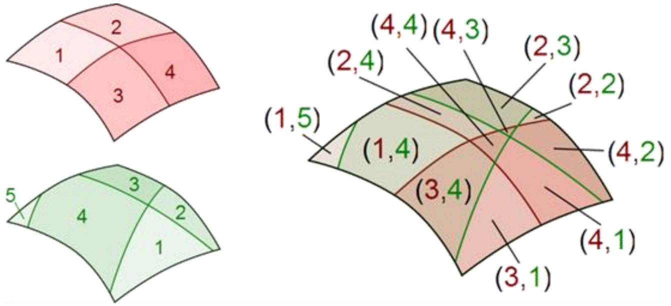


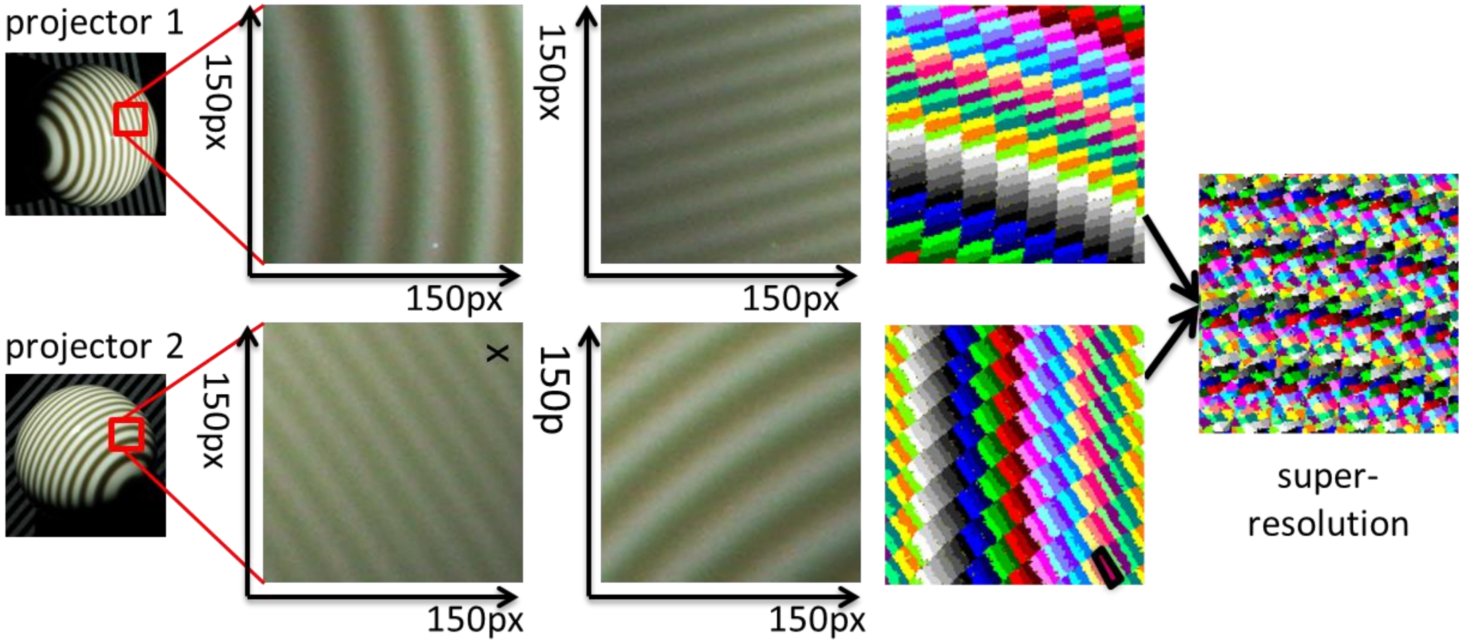
M. Weinmann, R. Ruiters, A. Osep, C. Schwartz, and R. Klein, A Multi-Camera, Multi-Projector Super-Resolution Framework for Structured Light, 3DIMPVT 2011

Advanced 3D Scanning Techniques

Overcome challenges of classical techniques based on:

- HDR Scanning
- Superresolution



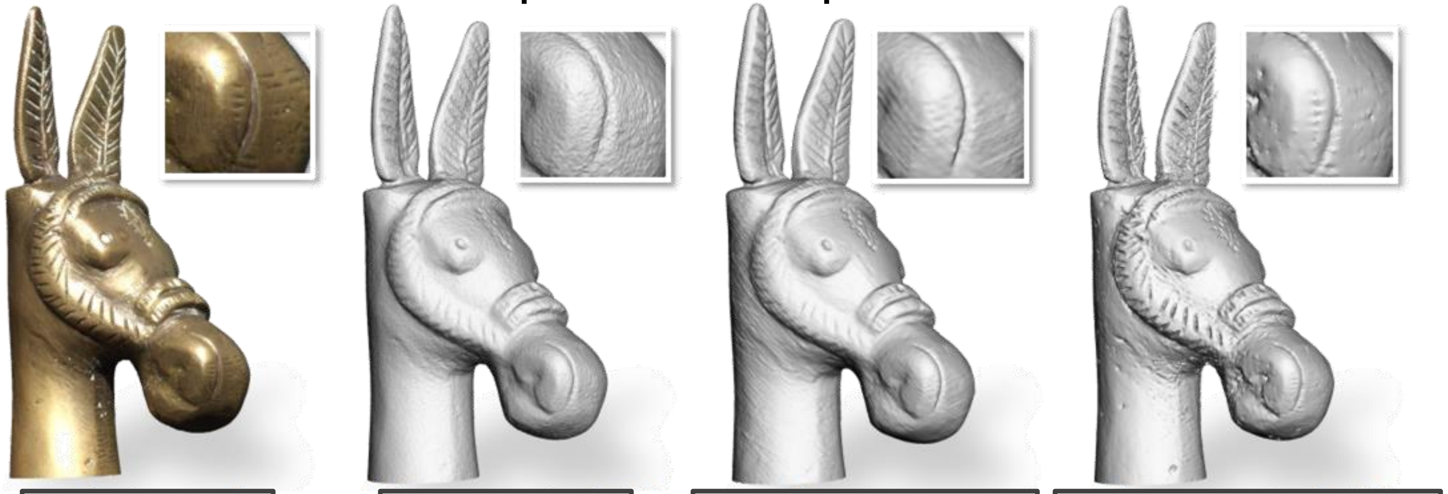


M. Weinmann, R. Ruiters, A. Osep, C. Schwartz, and R. Klein, A Multi-Camera, Multi-Projector Super-Resolution Framework for Structured Light, 3DIMPVT 2011

Advanced 3D Scanning Techniques

Overcome challenges of classical techniques based on:

- HDR Scanning
- Superresolution
- Combination of different techniques to compensate for their individual limitations



photograph | laser scan | super-resolution (Weinmann et al. 2011) | structured light + normals (Weinmann et al. 2012)

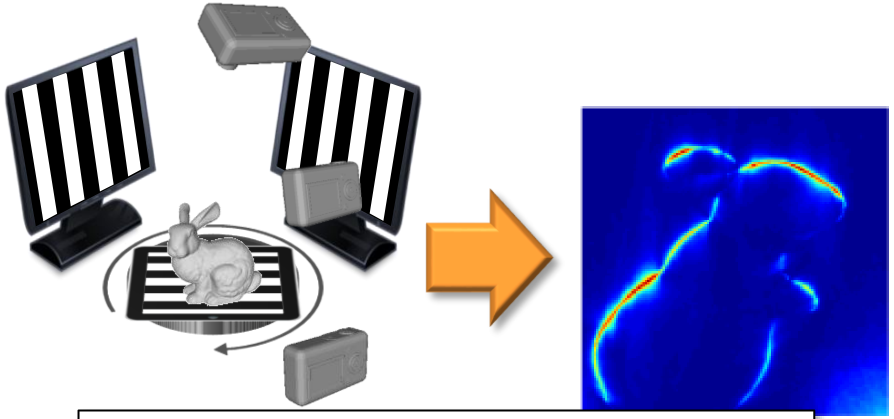
M. Weinmann, R. Ruiters, A. Osep, C. Schwartz, and R. Klein, Fusing Structured Light Consistency and Helmholtz Normals for 3D Reconstruction, BMVC 2012

BMVC 2012
Surrey, September 3rd - 7th

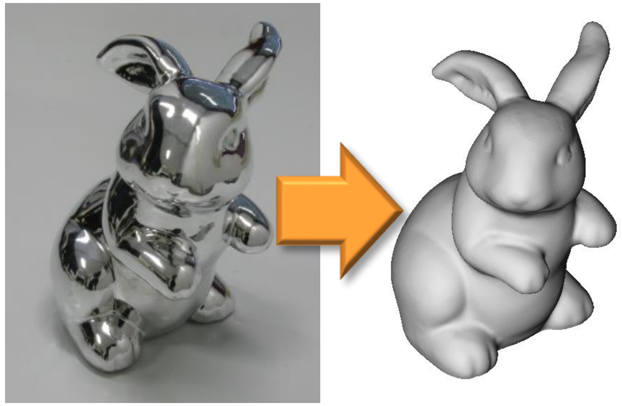
Advanced 3D Scanning Techniques

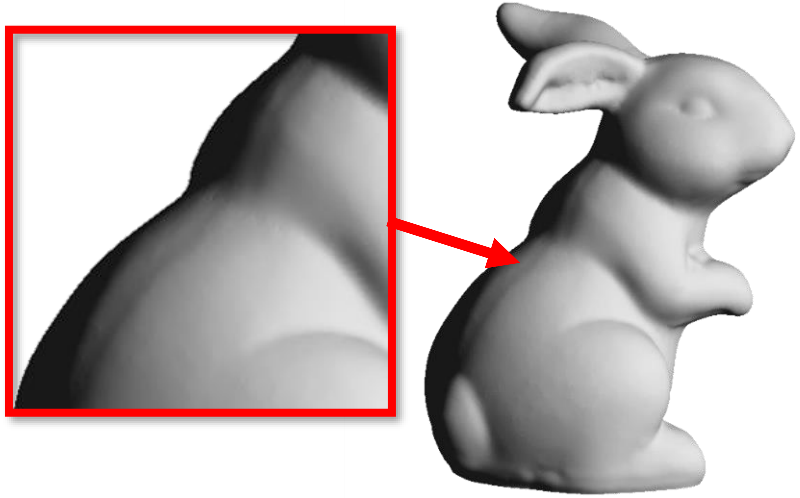
Overcome challenges of classical techniques based on:

- HDR Scanning
- Superresolution
- Combination of different techniques to compensate for their individual limitations
- Techniques tailored to complicated objects/surfaces



observations → surface consistency



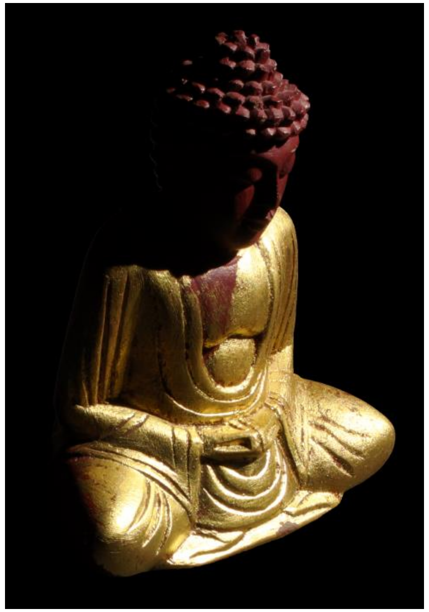


M. Weinmann, A. Osep, R. Ruiters and Reinhard Klein. Multi-View Normal Field Integration for 3D Reconstruction of Mirroring Objects, ICCV 2013

How should we represent complex 3D scenes?

Why do we need reflectance?

Important visual cue regarding **„how do things feel“**



Photo

Texture

BTF

How to model light exchange at the surface?

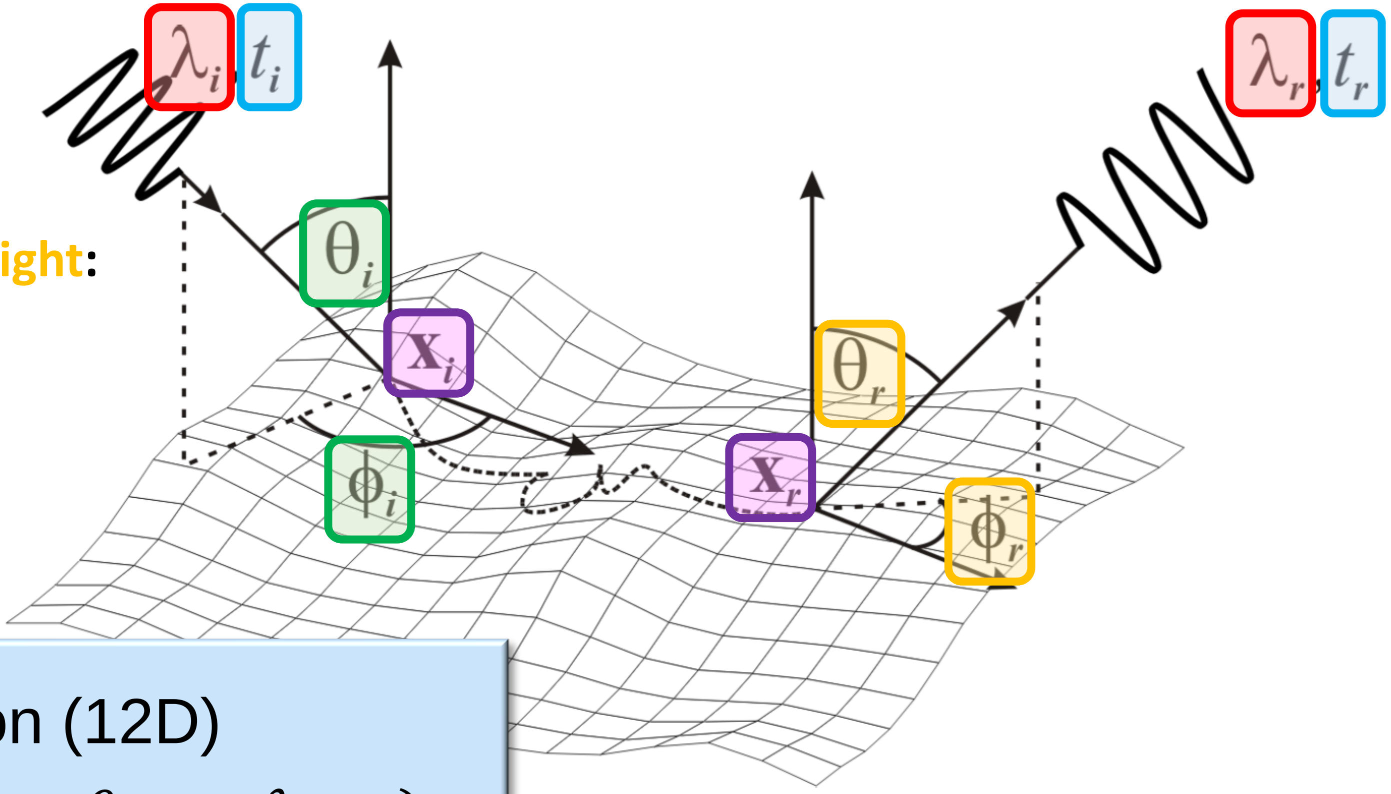
Describe the material appearance decoupled from the environment, lighting and observer characteristics

Timesteps: t_i, t_r

Directions of incoming/reflected light:
(θ_i, ϕ_i), (θ_r, ϕ_r)

Surface points: x_i, x_r

Wavelengths: λ_i, λ_r



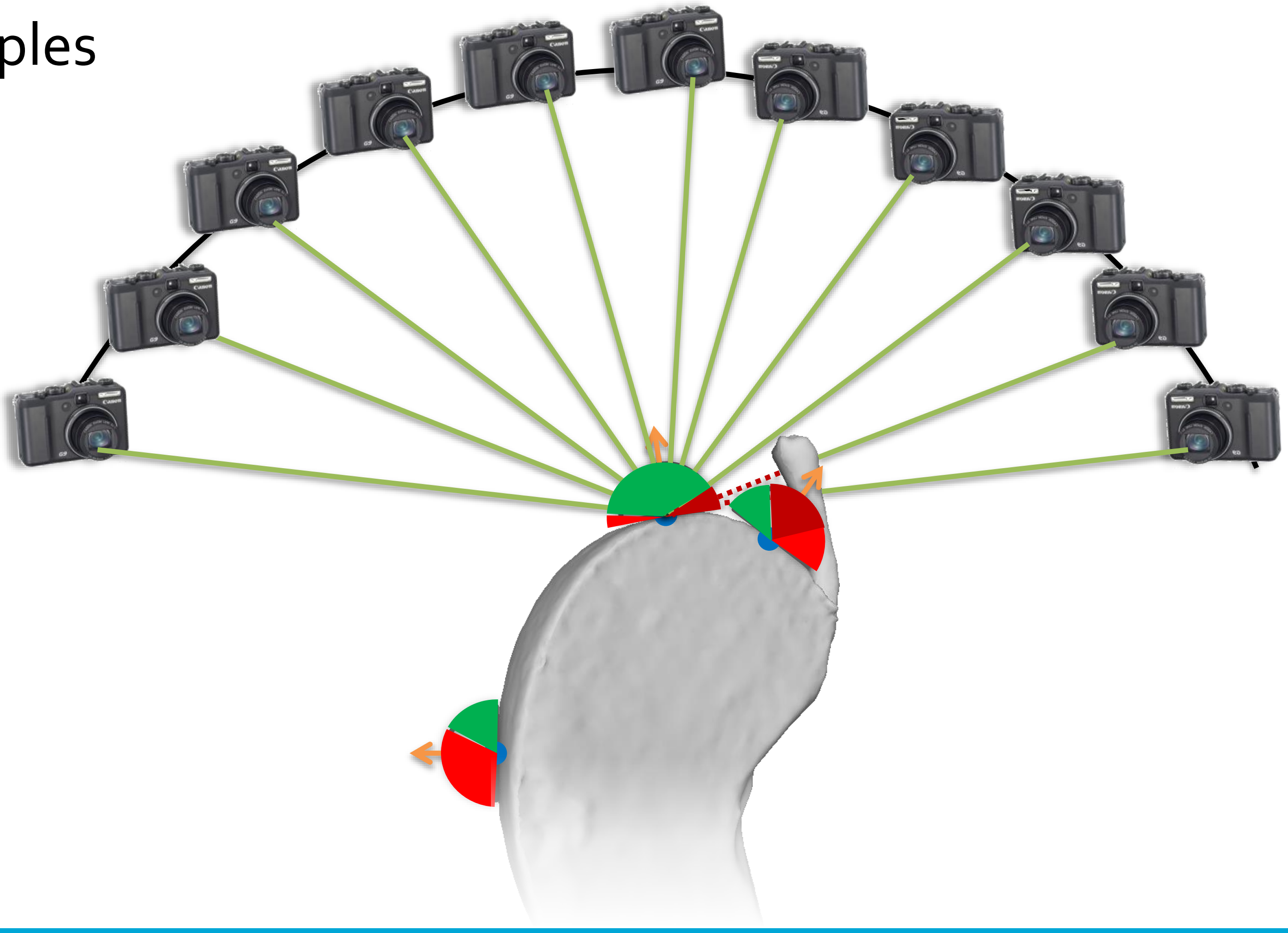
general function (12D)

$$\rho(x_i, y_i, \theta_i, \varphi_i, \lambda_i, t_i, x_r, y_r, \theta_r, \varphi_r, \lambda_r, t_r)$$

Conventional Appearance Capture

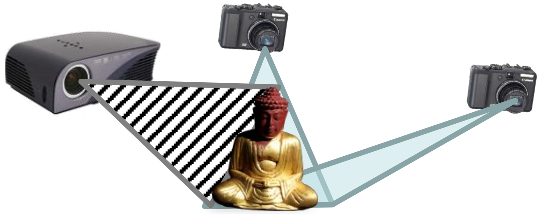
How to capture appearance characteristics?

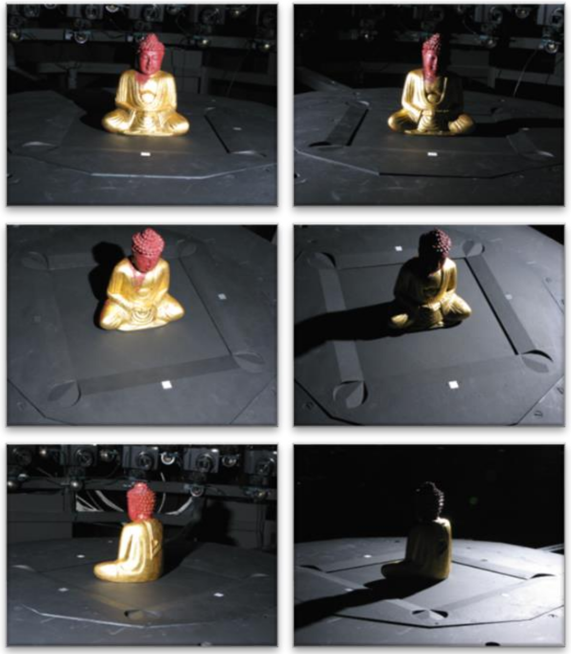
- Measurement of reflectance samples
  - **Irregular**
  - **Occlusion**
  - **Not measured**



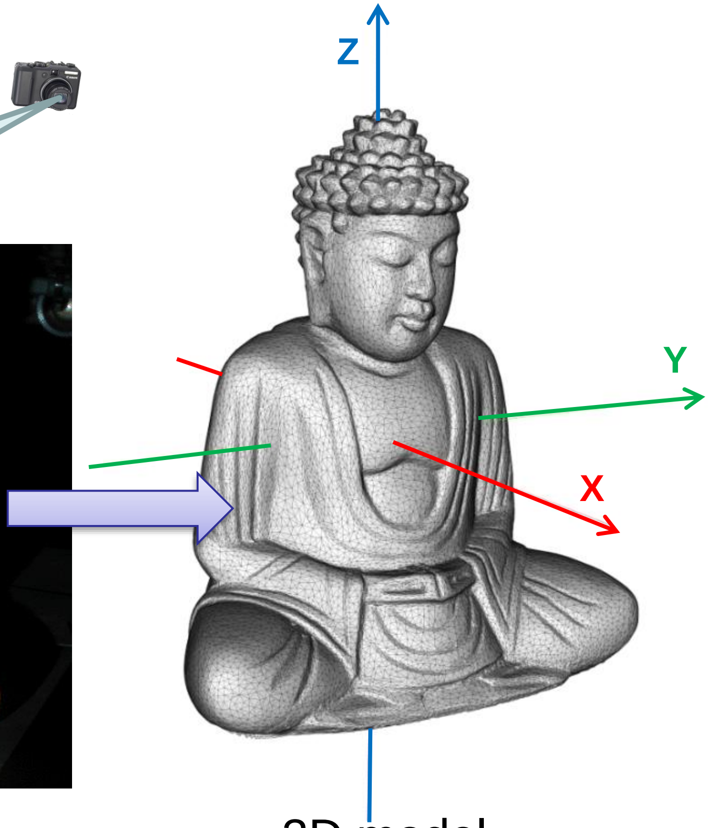
Conventional Appearance Capture

Separate geometry and reflectance reconstructions





Weinmann et al. 2011

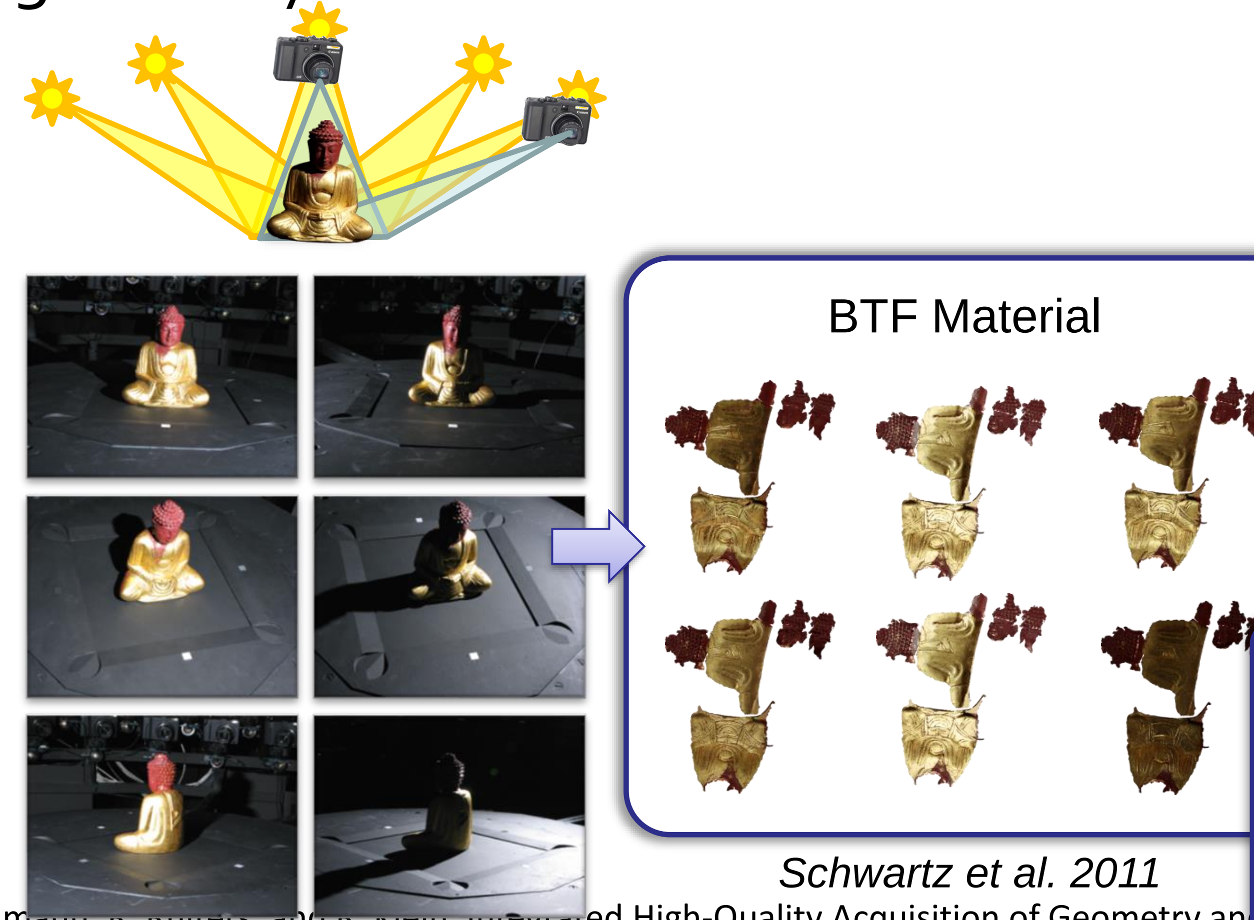


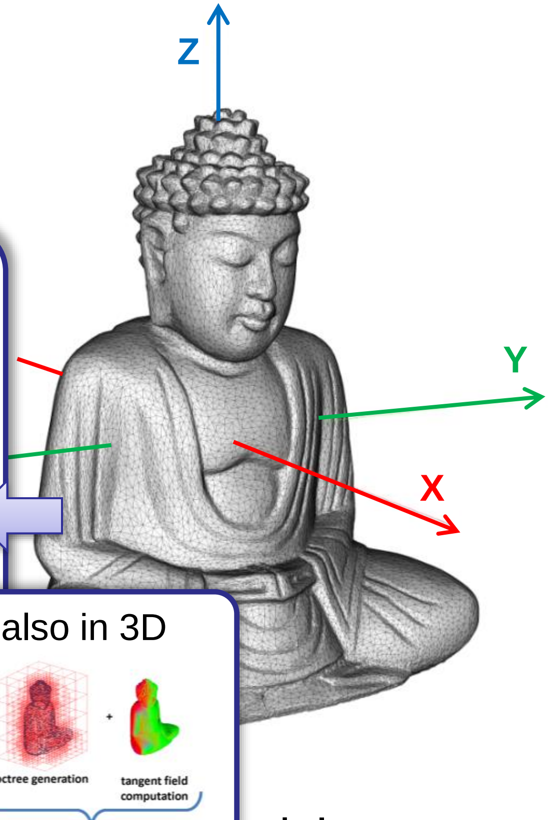
3D model

C. Schwartz, M. Weinmann, R. Ruiters, and R. Klein, Integrated High-Quality Acquisition of Geometry and Appearance for Cultural Heritage, VAST 2011

Conventional Appearance Capture

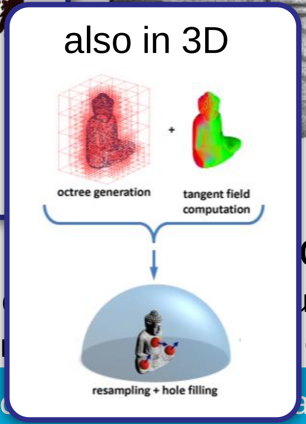
Separate geometry and reflectance reconstructions





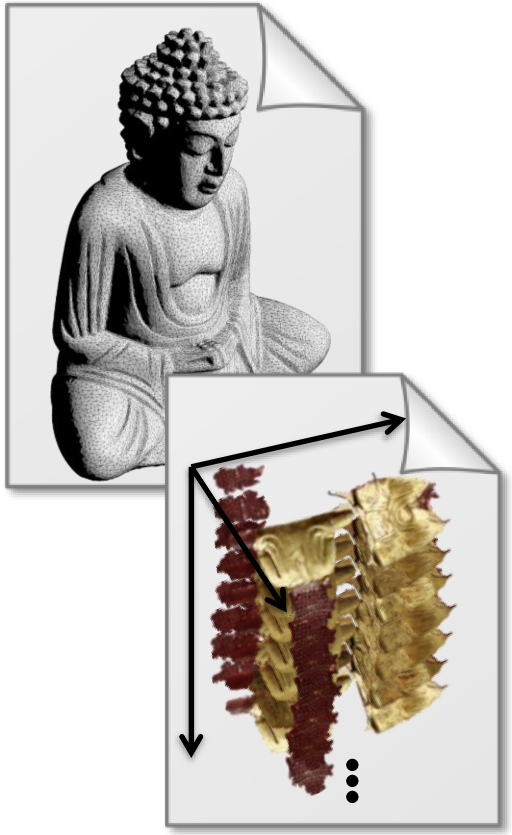
Schwartz et al. 2011

odel



C. Schwartz, M. Weinmann, R. Ruiters, and R. Klein, Integrated High-Quality Acquisition of Geometry an… …ultural Heritage, VAST 2011
S. Krumpen, M. Weinmann, and R. Klein, OctreeBTFs – A compact, seamless and distortion-free reflecta… …Computers & Graphics 2017

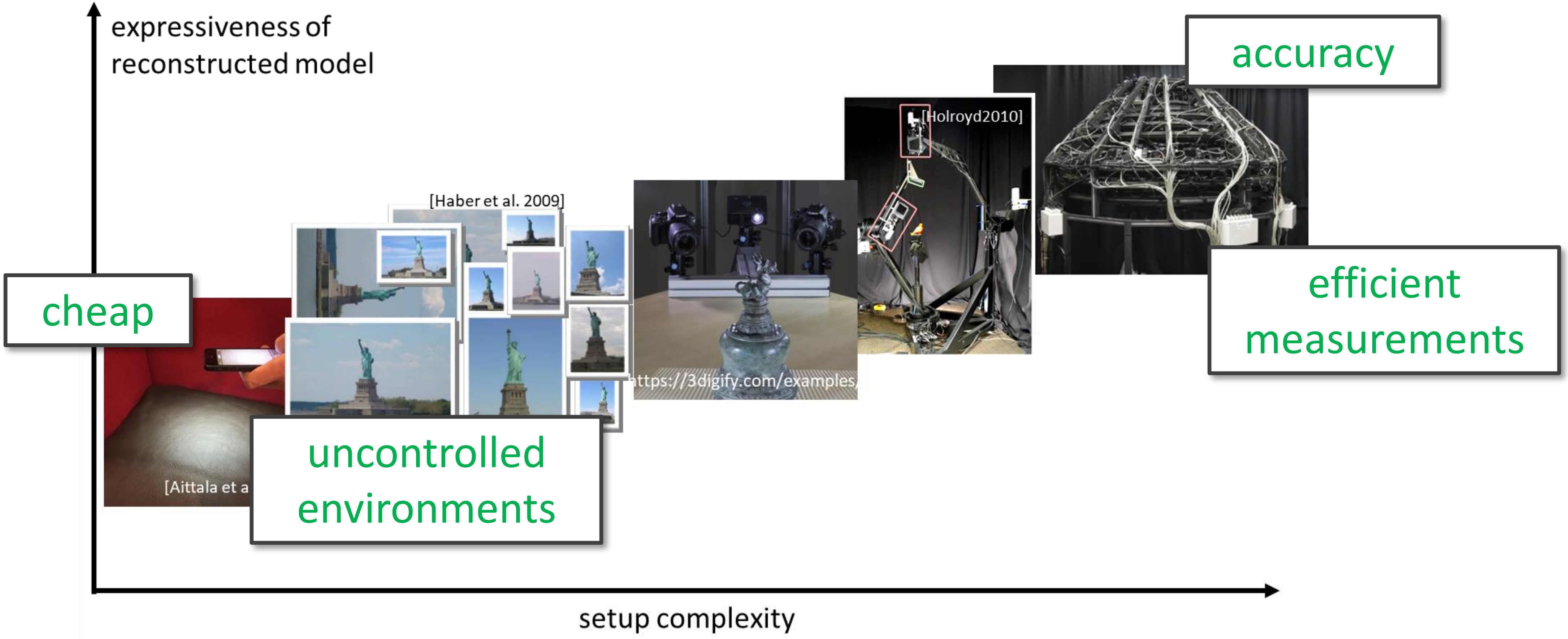
Conventional Appearance Capture



C. Schwartz, M. Weinmann, R. Ruiters, and R. Klein, Integrated High-Quality Acquisition of Geometry and Appearance for Cultural Heritage, VAST 2011

High-quality results, but …

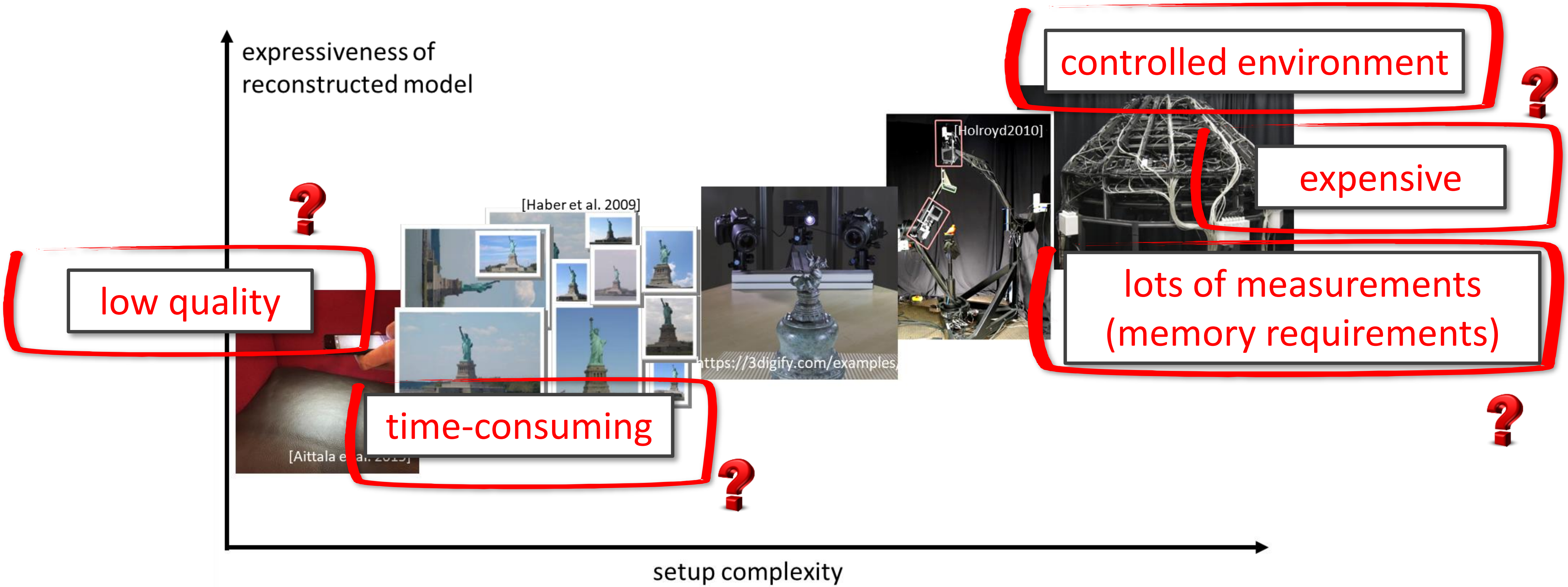
Trade-off: Acquisition technology vs. expressiveness of models:



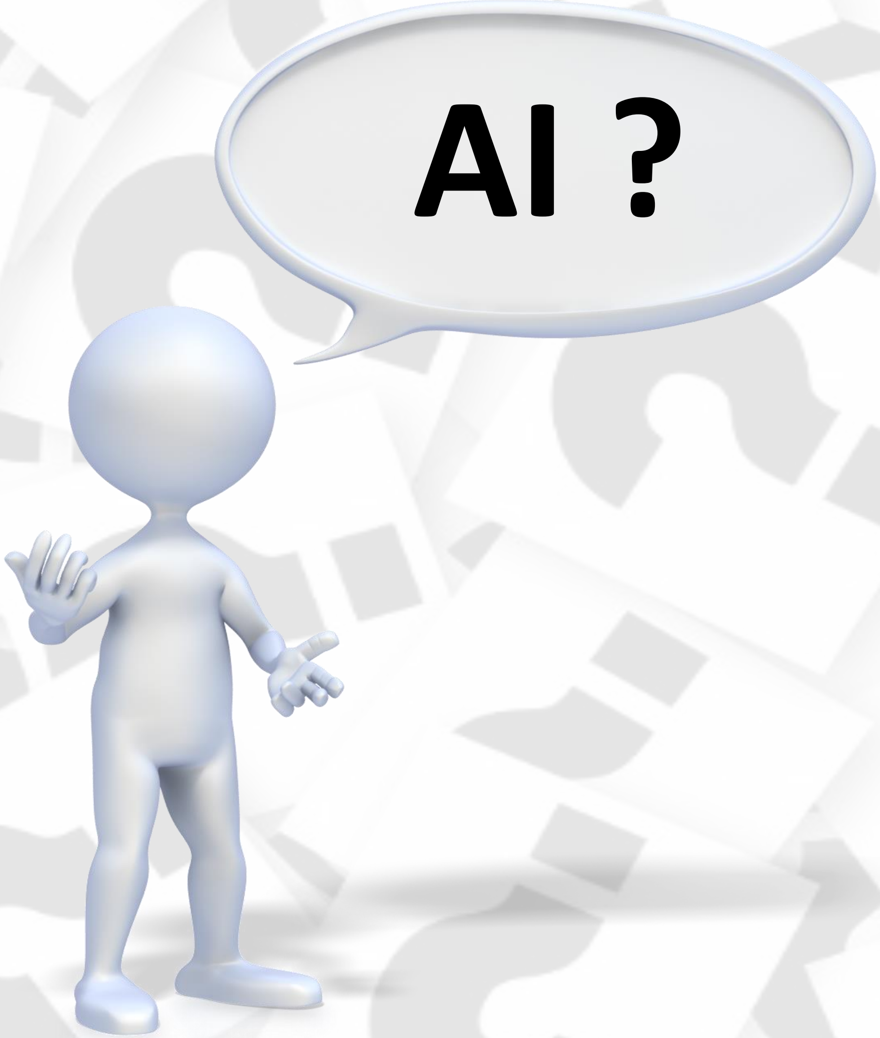
How should we represent complex 3D scenes?

Challenges

Trade-off: Acquisition technology vs. expressiveness of models:



How should we represent complex 3D scenes?



AI-based Scene Capture/Modeling

Traditional learning-based methods:

- Supervised learning

what we have to provide

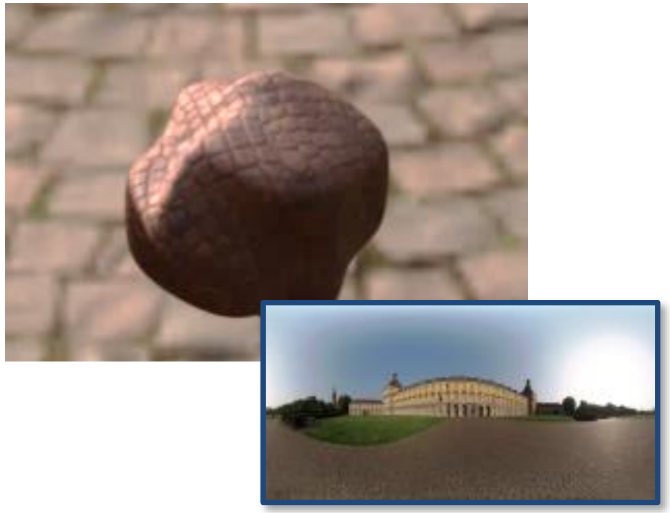
```
def train( train_data , train_labels):
  # build a model for images -> labels...
  return model
```

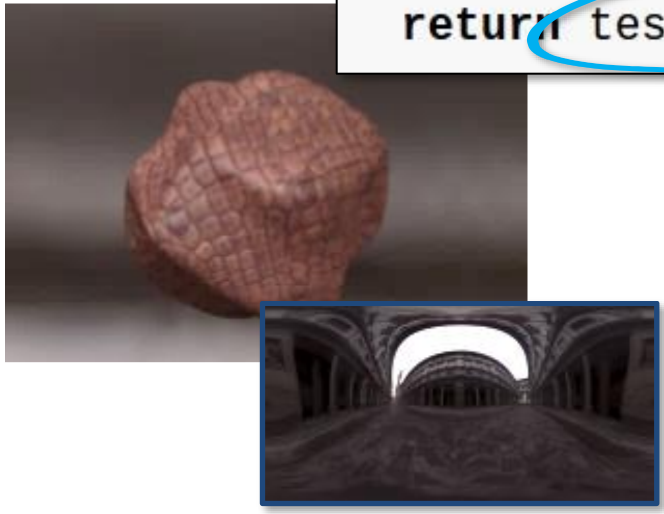
what we can get from our data

```
def predict(model, test_data ):
  # predict test_labels using the model...
  return test_labels
```

what we can do with our model

- Problem:



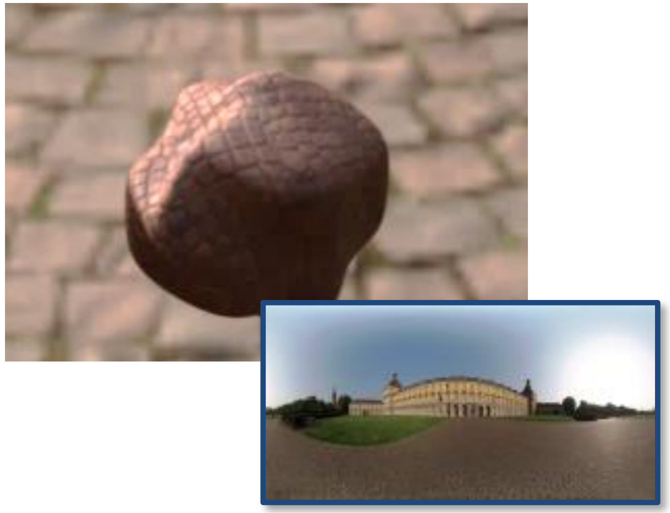


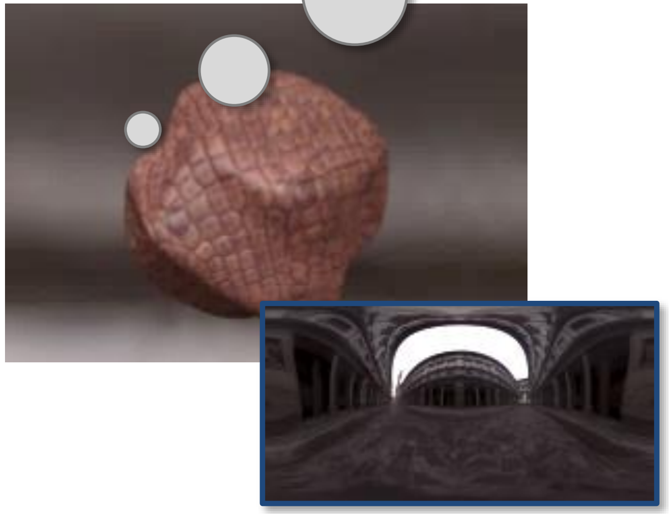
AI-based Scene Capture/Modeling

Traditional learning-based methods:

- Supervised learning

- Problem:



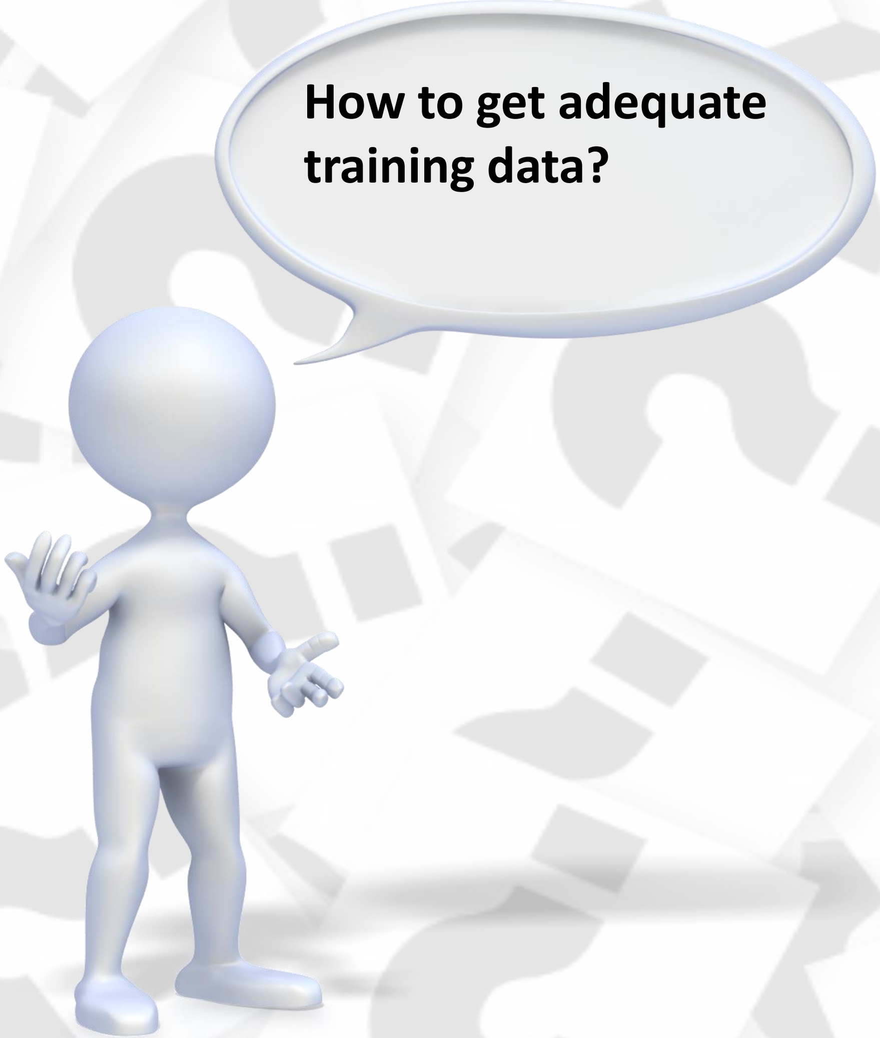


Challenges:

- Different objects/scenes
- Different materials
- Different view conditions
- Different lighting conditions
- …

→ **Large datasets!**

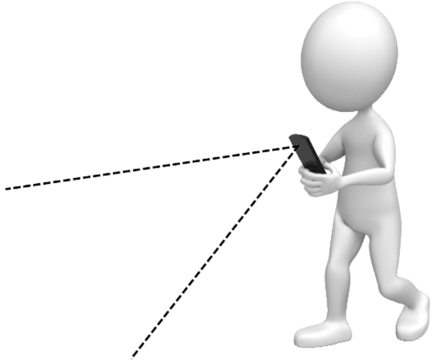
AI-based Scene Capture/Modeling



How to get adequate training data?

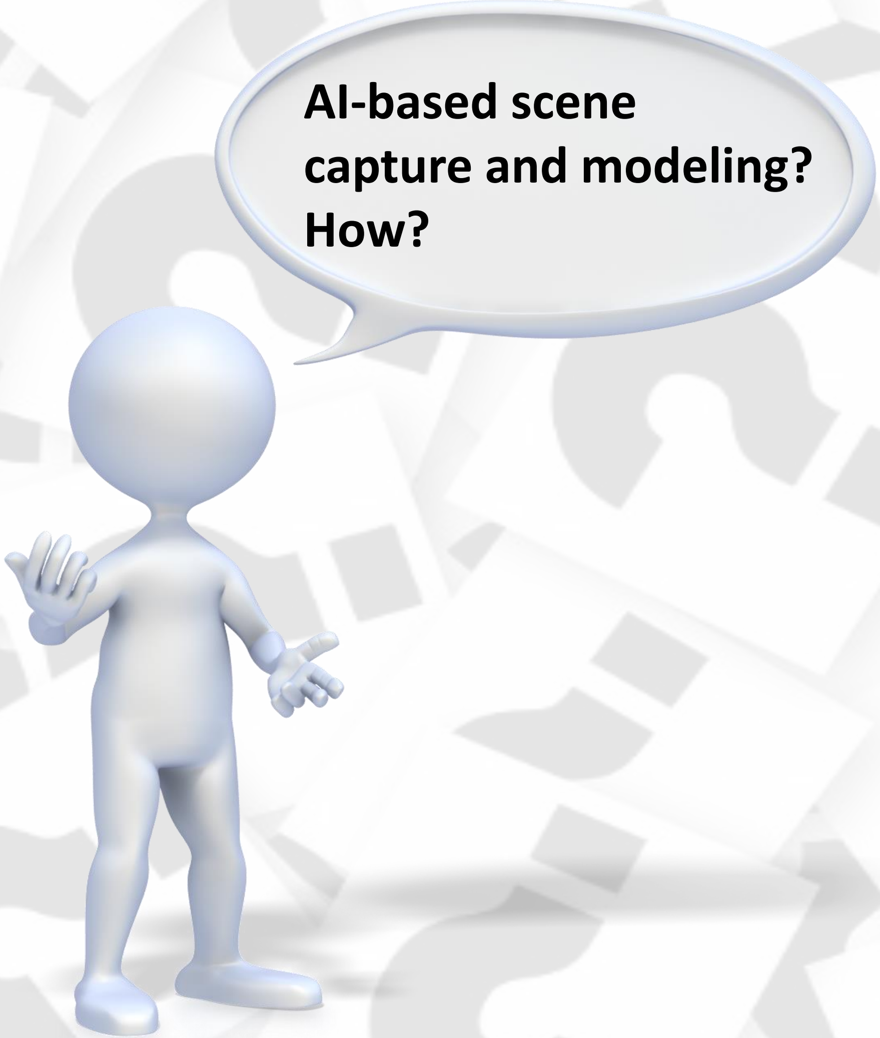
Approaches:

- Manual collection of labeled training data
  - Time-consuming (too many configurations)
  - Costly
- Use of virtual scenarios (e.g., Virtual KITTI 2 dataset, CARLA, etc.)
  - Easy to label



(https://europe.naverlabs.com/research/computer-vision/proxy-virtual-worlds-vkitti-2/)

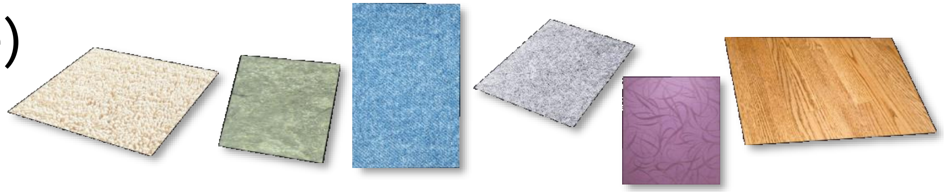
AI-based Scene Capture and Modeling



What can we do with pre-trained models?

Given adequate datasets, we can …

- … learn to recognize materials (e.g. to steer capture)



M. Weinmann, J. Gall, R. Klein. Material classification based on training data synthesized using a BTF database, ECCV, 2014

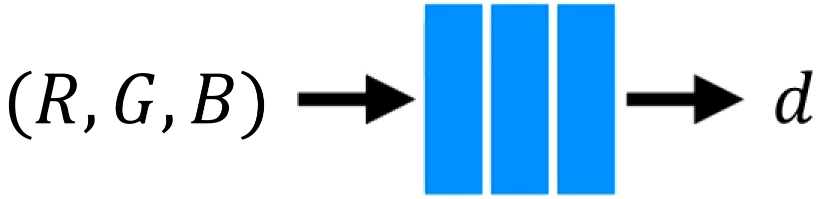
- … learn feature correspondences

P. Truong, M. Danelljan, L. Van Gool, R. Timofte. Learning Accurate Dense Correspondences and When to Trust Them, CVPR, 2021

What can we do with pre-trained models?

Given adequate datasets, we can …

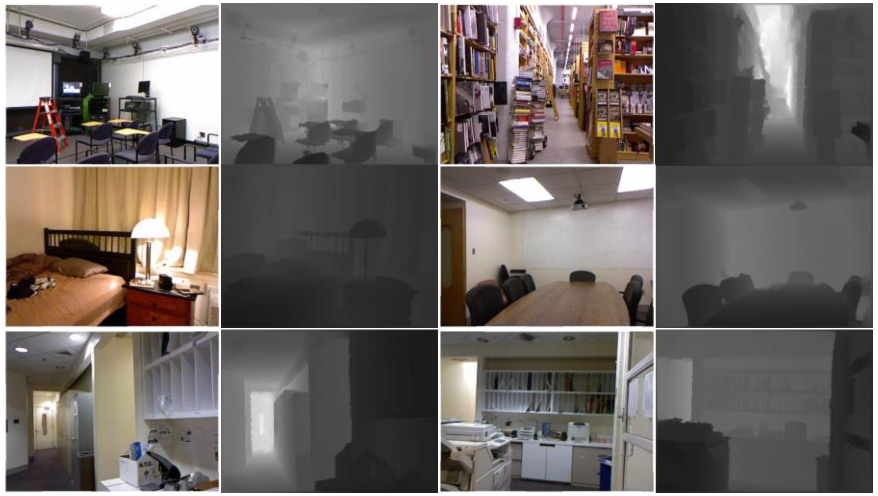
- … learn to estimate geometry from a single image (also used for SLAM …)



Loss components:

- Depth / depth relations
- Surface normal
- Gradient information

Training data: RGB + Depth pairs



NYU Depth V2 dataset

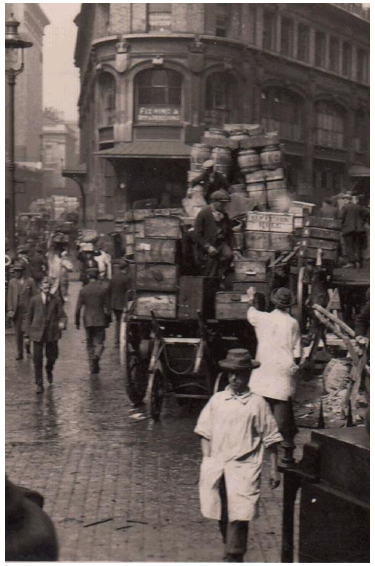
Predictions:

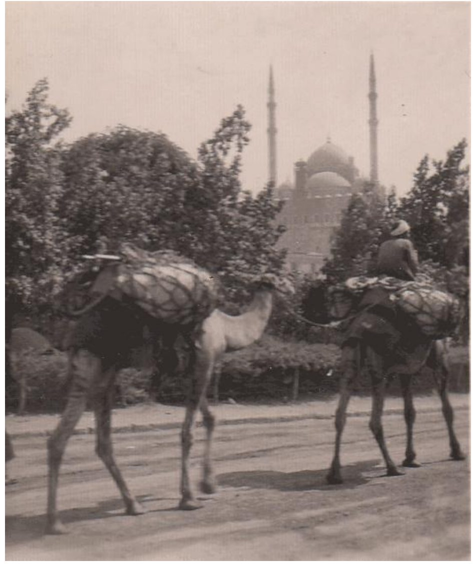
https://paperswithcode.com/task/monocular-depth-estimation

What can we do with pre-trained models?

Various applications:

- Turning pictures into 3D experiences
  (→ geometry estimation from a single image + warping + inpainting)





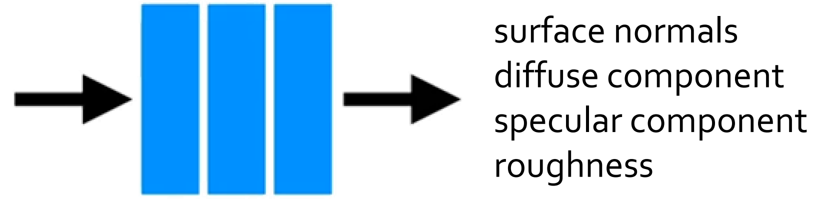
Joseph Redmon, CSE455: Computer Vision, University of Washington

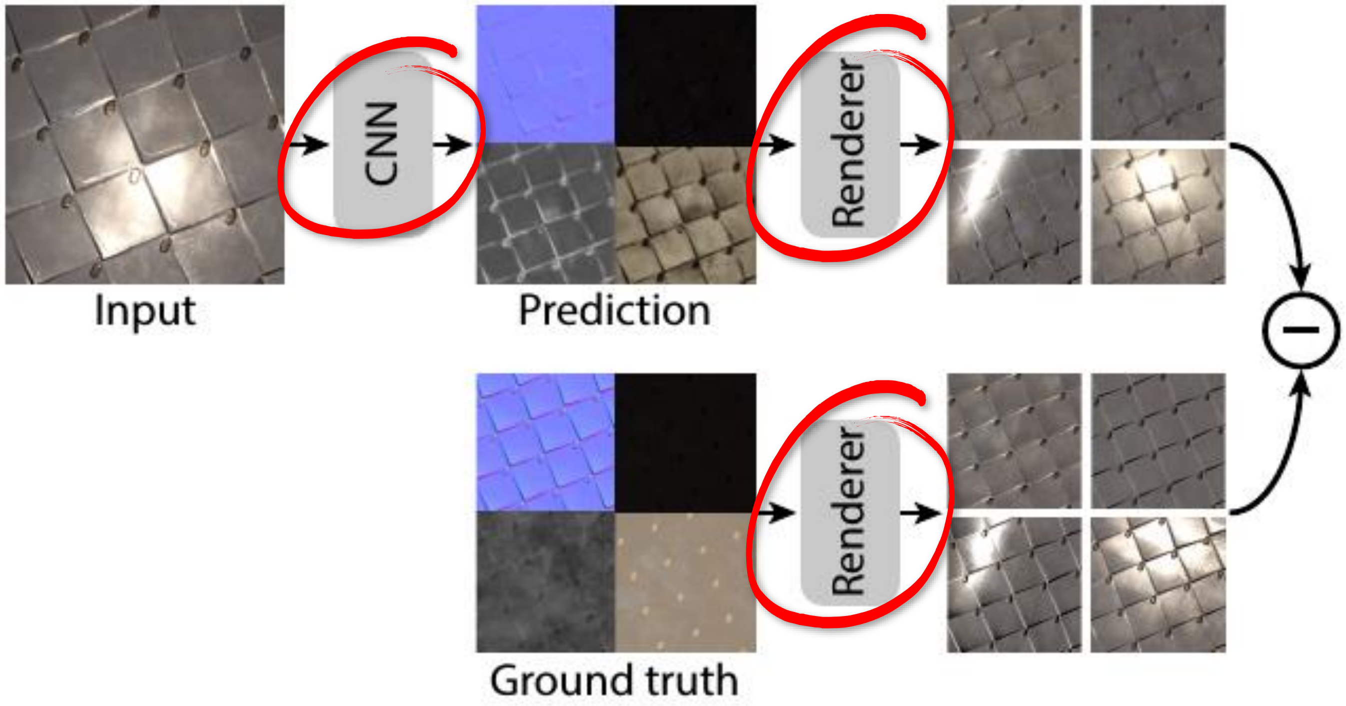
What can we do with pre-trained models?

Given adequate datasets, we can ...

- ... learn to estimate reflectance from a single image
  - Combine AI ...
  - ... and CG

(R, G, B)





V Deschaintre, M Aittala, F Durand, G Drettakis, A Bousseau. Single-image svbrdf capture with a rendering-aware deep network, ACM Transactions on Graphics, 2018

What can we do with pre-trained models?

Given adequate datasets, we can ...

- ... learn to synthesize novel views from 2 input views

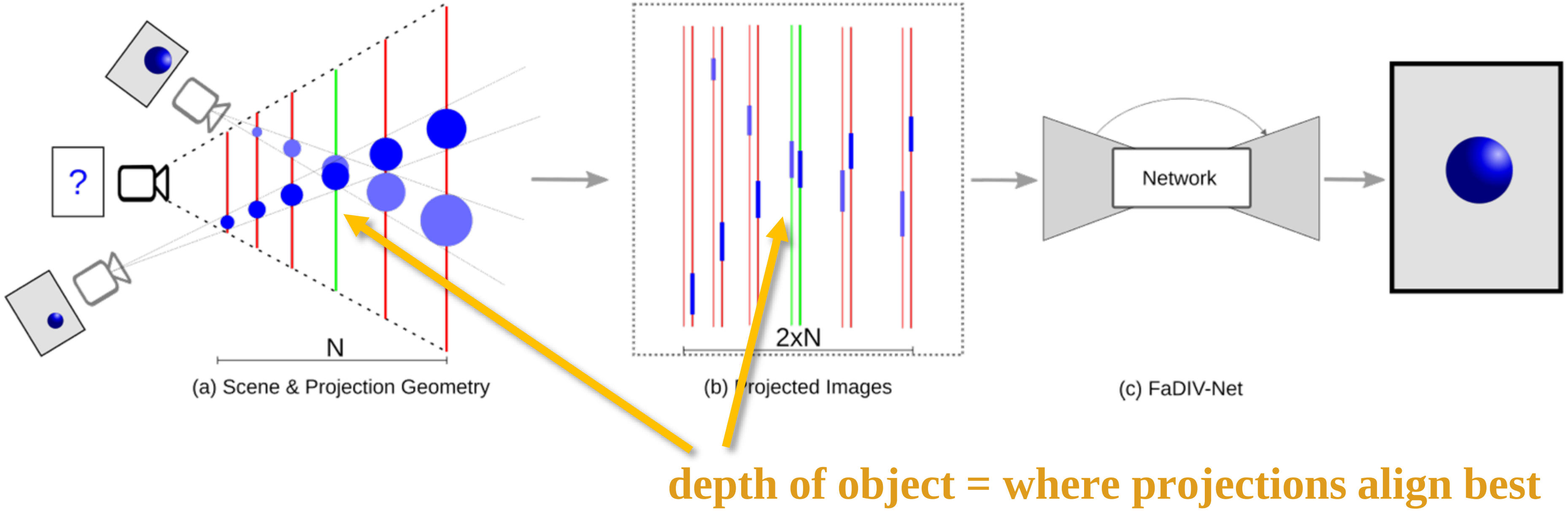
extremely sparse-view scenario ...

Rochow, Schwarz, Weinmann, Behnke. FaDIV-Syn: Fast Depth-Independent View Synthesis using Soft Masks and Implicit Blending, RSS, 2022

What can we do with pre-trained models?

Given adequate datasets, we can …

- … learn to synthesize novel views from 2 input views



Rochow, Schwarz, Weinmann, Behnke. FaDIV-Syn: Fast Depth-Independent View Synthesis using Soft Masks and Implicit Blending, RSS, 2022

What can we do with pre-trained models?

Given adequate datasets, we can …

- … learn to synthesize novel views from 2 input views

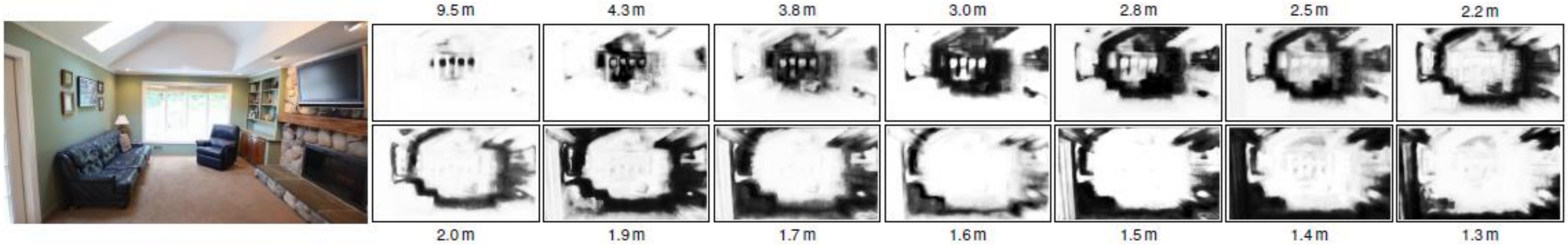


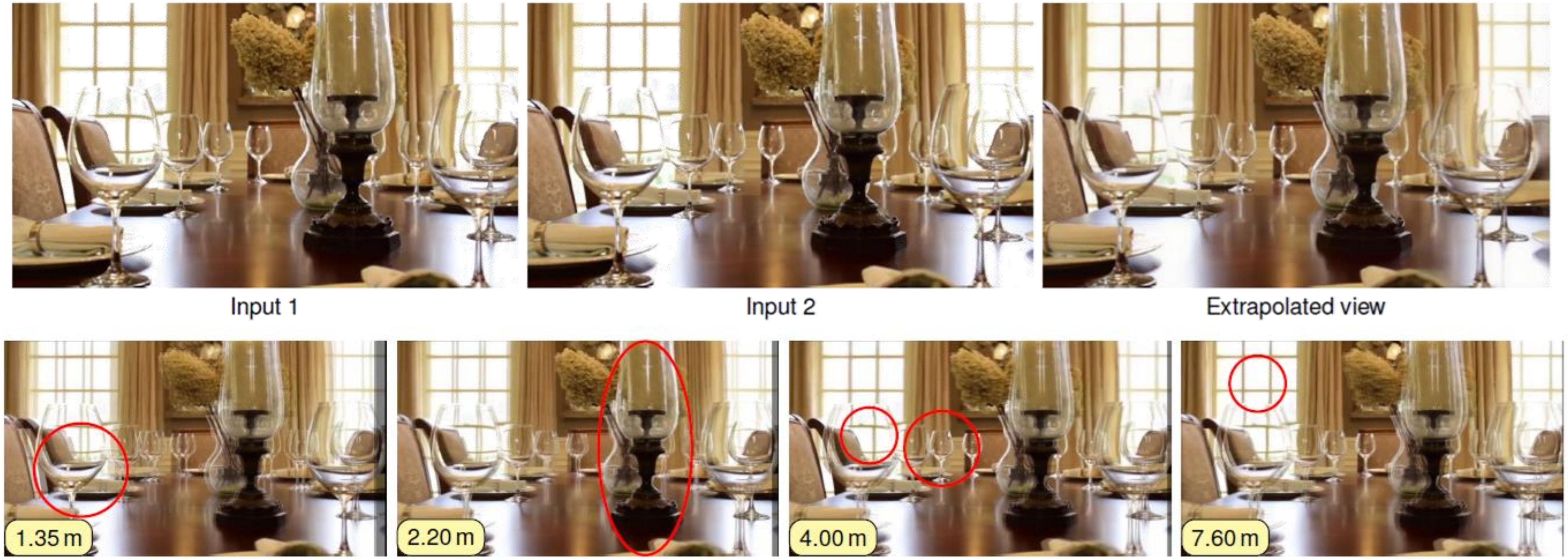
Fig. 6: Self-supervised soft masking. Predicted image (left) and learned mask activations (right). The masks are normalized (dark = high activation). The PSV layer depth is up to scale. Note how the learned masks correlate with depth in the scene.

Rochow, Schwarz, Weinmann, Behnke. FaDIV-Syn: Fast Depth-Independent View Synthesis using Soft Masks and Implicit Blending, RSS, 2022

What can we do with pre-trained models?

Given adequate datasets, we can …

- … learn to synthesize novel views from 2 input views

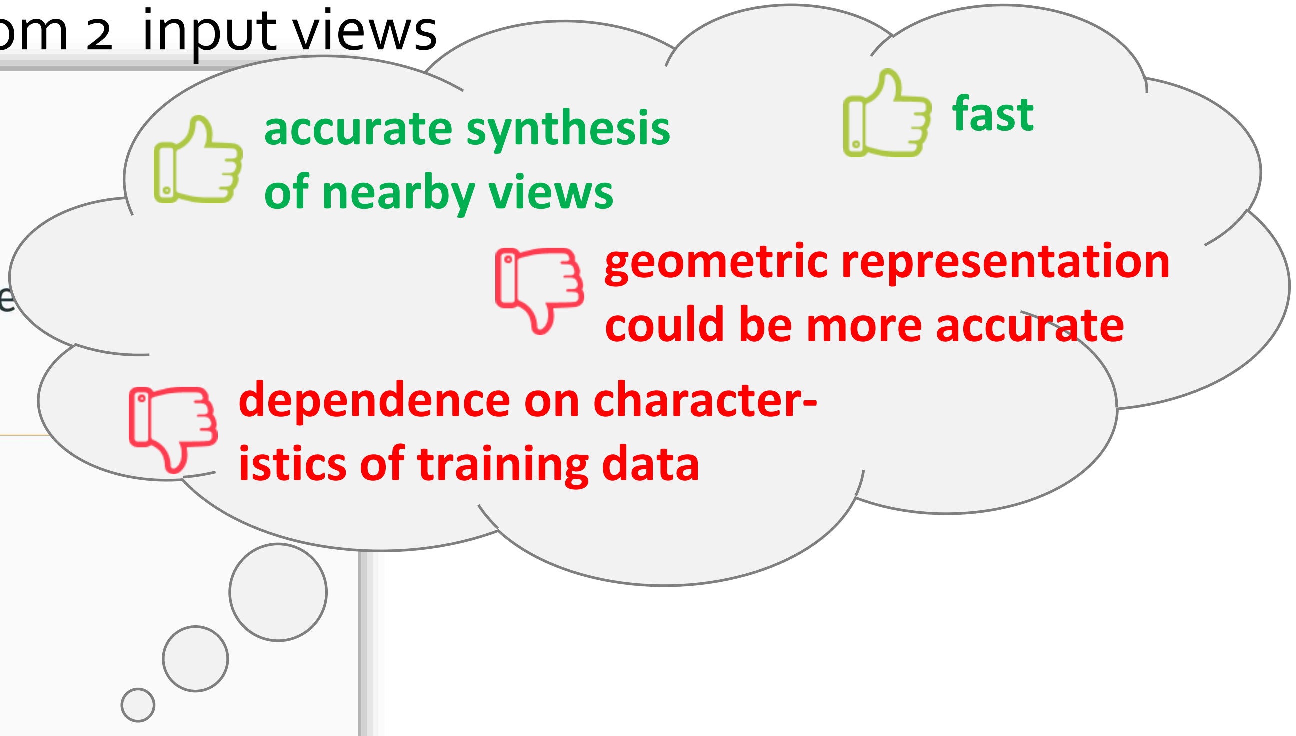


Rochow, Schwarz, Weinmann, Behnke. FaDIV-Syn: Fast Depth-Independent View Synthesis using Soft Masks and Implicit Blending, RSS, 2022

What can we do with pre-trained models?

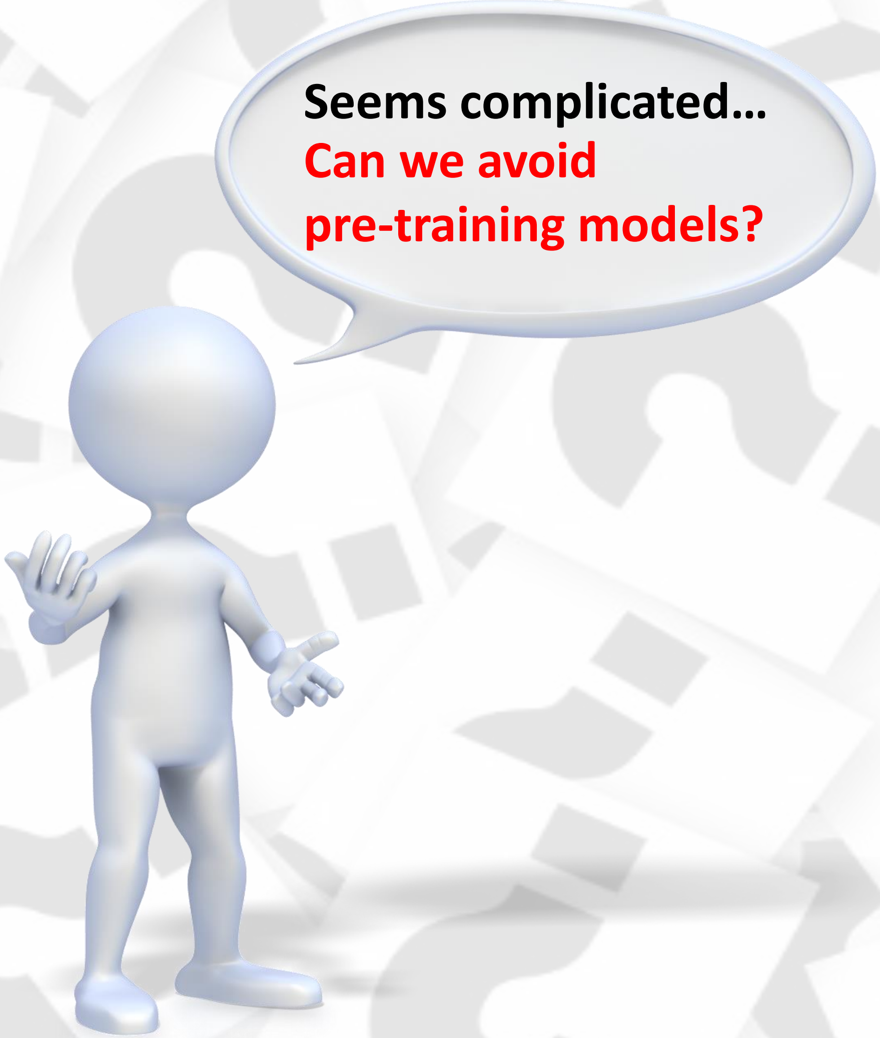
Given adequate datasets, we can ...

- ... learn to synthesize novel views from 2 input views



Rochow, Schwarz, Weinmann, Behnke. FaDIV-Syn: Fast Depth-Independent View Synthesis using Soft Masks and Implicit Blending, RSS, 2022

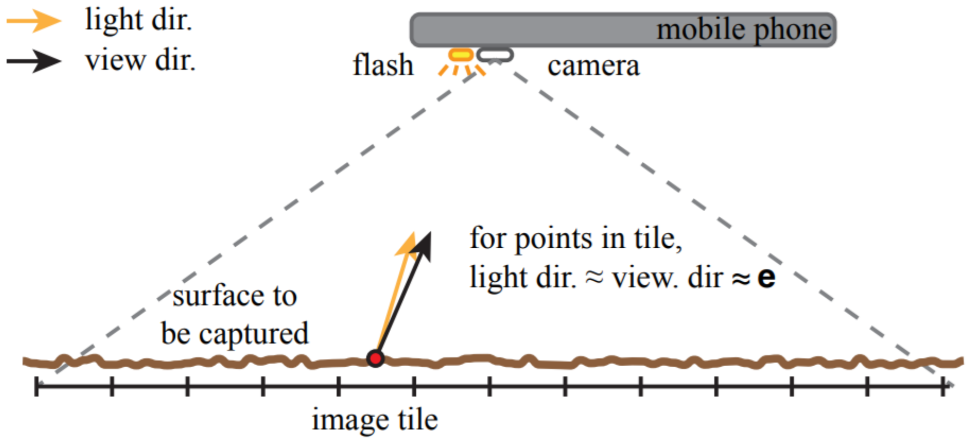
AI-based Scene Capture and Modeling

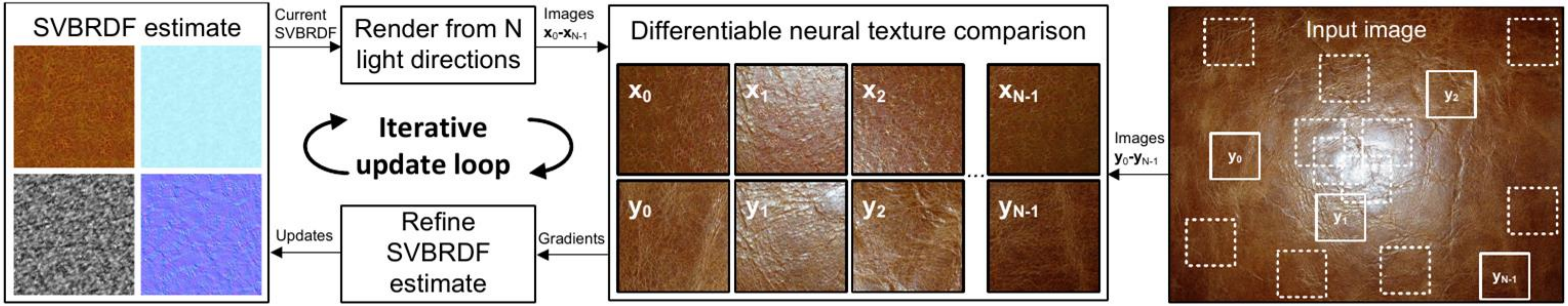


What can we represent based on AI?

Self-supervision for …

- Reflectance estimation from a single image
  - Combine AI …
  - … and CG



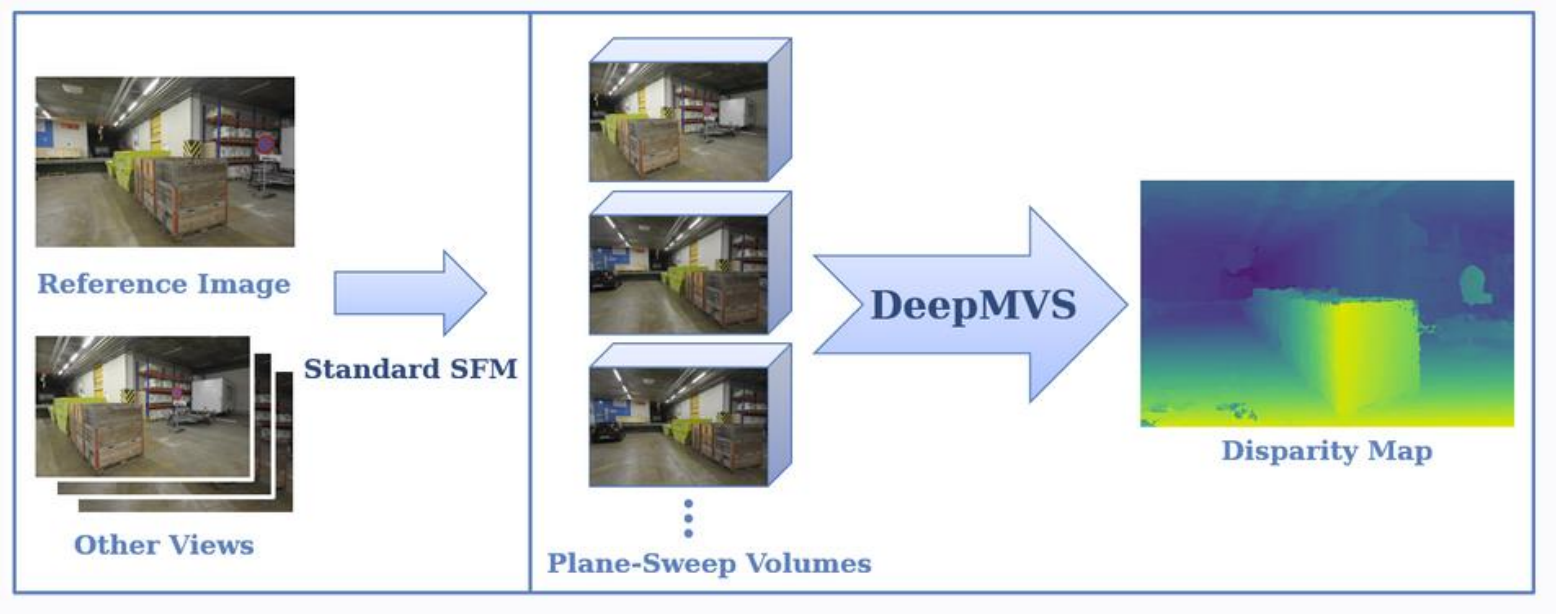


Aittala et al., Reflectance Modeling by Neural Texture Synthesis

What can we represent based on AI?

Self-supervision for …

- Geometry reconstruction from multiple views



P.-H. Huang, K. Matzen, J. Kopf, N. Ahuja, J.-B. Huang. DeepMVS: Learning Multi-View Stereopsis, CVPR, 2018

What ca[…] represen[…] based[…]

2020

2021

2022

2023

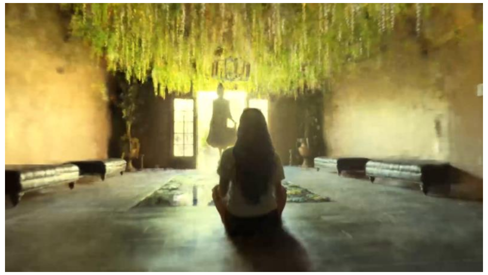
2024

2025

Neural Radiance Fields (NeRFs)

[…]üller, M. Weinmann, and R. Klein. *Unbiased Gradient Estimation for Differentiable Surface Splatting via Poisson Sampling*, ECCV 2022

Differentiable Splatting Techniques



[AI Artists with Instant NeRF | NVIDIA]

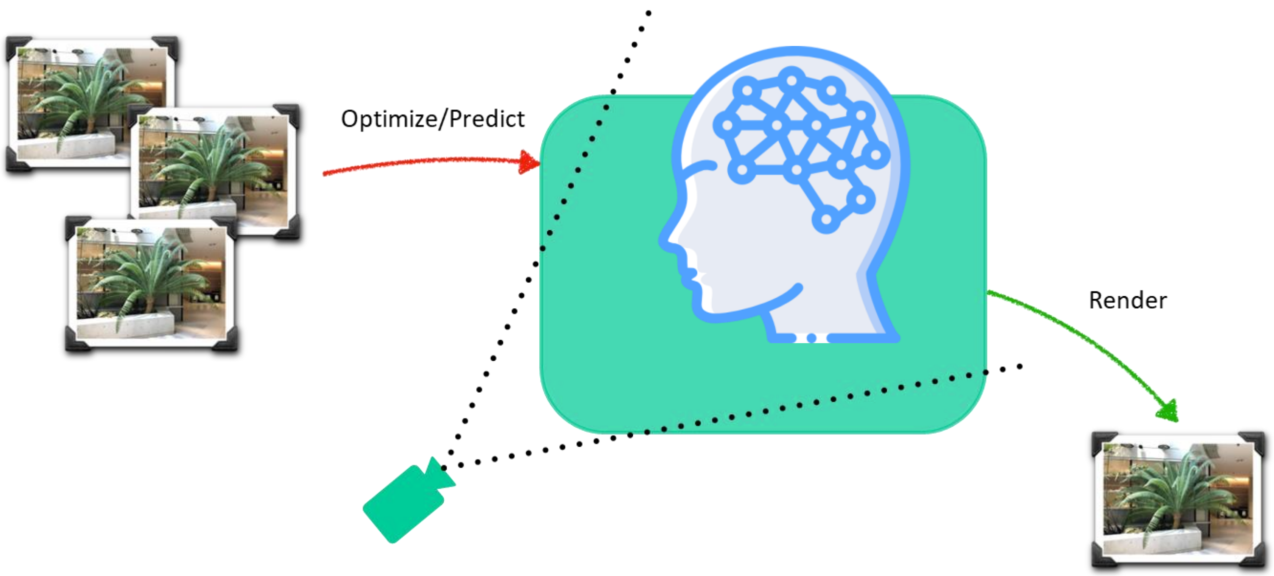
https://www.bizzlogic.com/insights/gaussian-splatting

Can we use AI to encode/represent whole scenes?

Self-supervision for ...

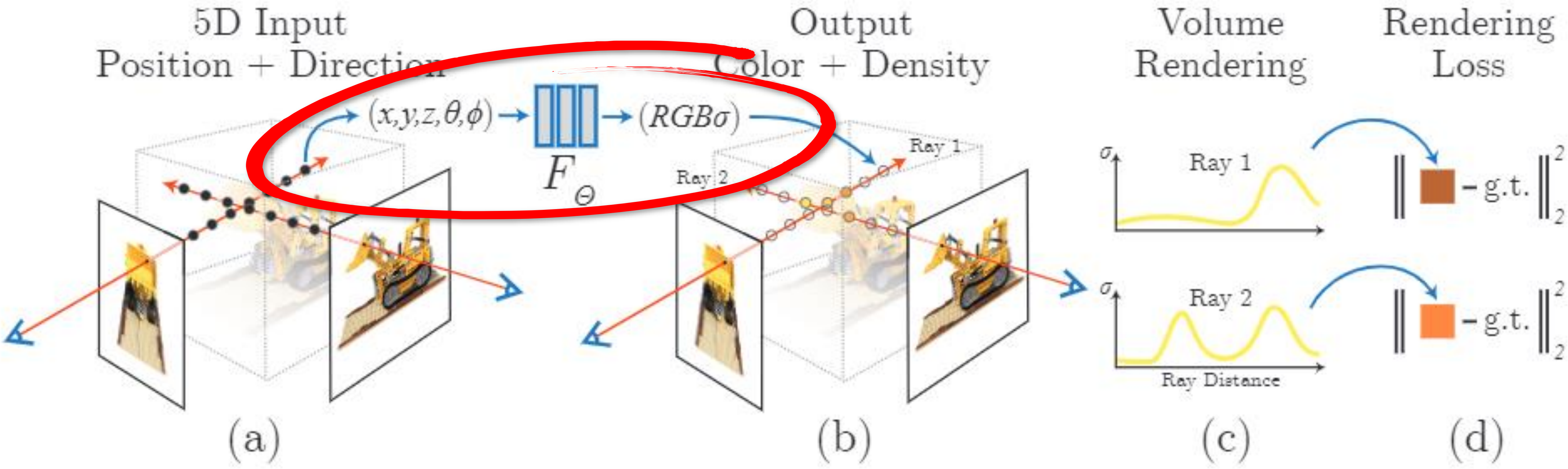
- Novel view synthesis (→ no need for additional datasets!)

- Idea:
  - Train a model which overfits to one object/scene!
  - Leverage inverse rendering setup



- Assumption:
  - Good results = expressive/accurate model

Neural Radiance Fields [Mildenhall et al. 2020]



Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R., Ng, R..
NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis, ECCV, 2020

AI-based Scene Representation & Visualization

Neural Radiance Fields [Mildenhall et al. 2020]:

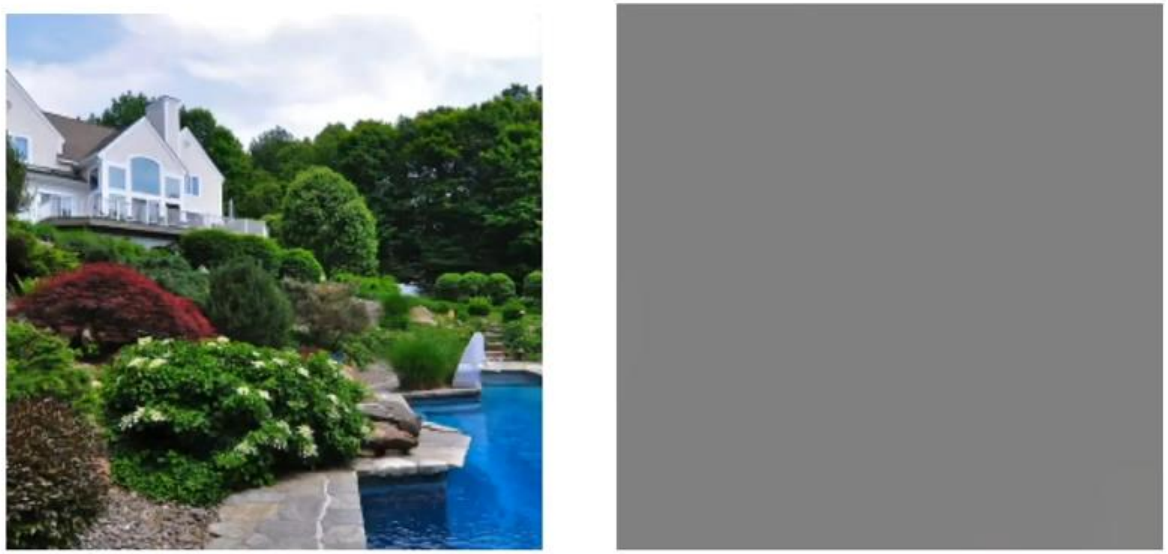
- Blurry results …

NeRF (Naive)

Neural Radiance Fields [Mildenhall et al. 2020]:

- How to get MLPs to represent high-frequency functions?

Ground truth image | Standard fully-connected net



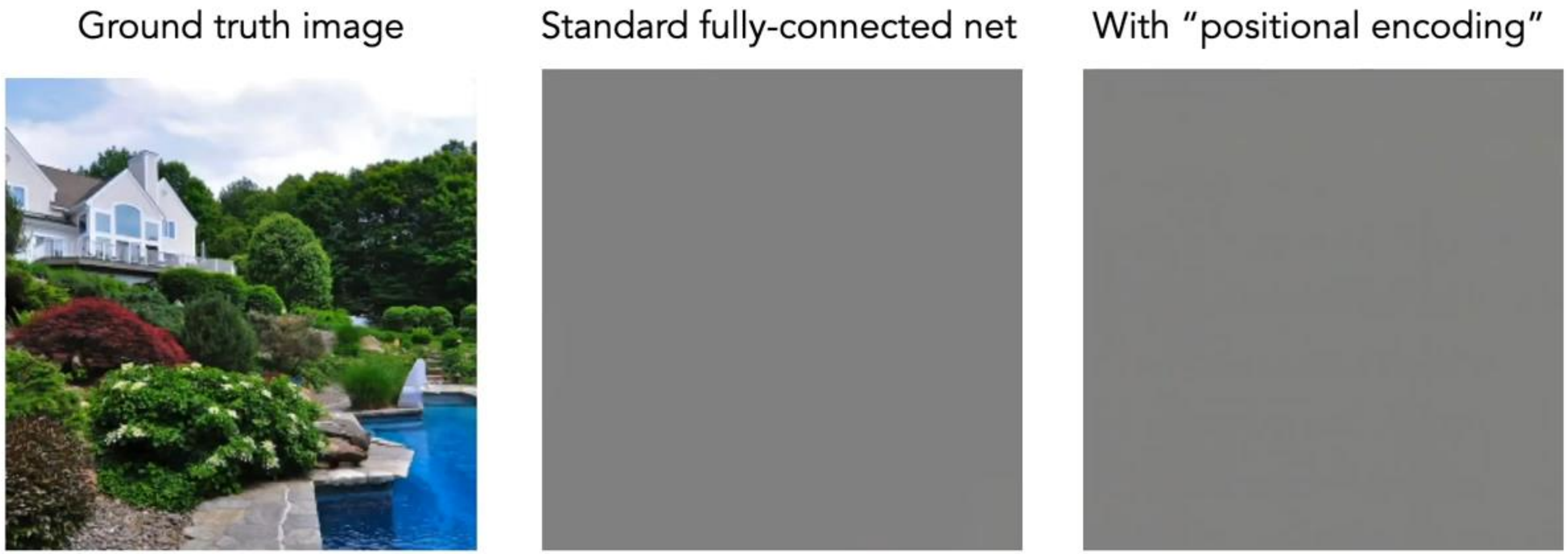
Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R., Ng, R..
NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis, ECCV, 2020

AI-based Scene Representation & Visualization

Neural Radiance Fields [Mildenhall et al. 2020]:

$(x, y) \rightarrow$  $\rightarrow (R, G, B)$

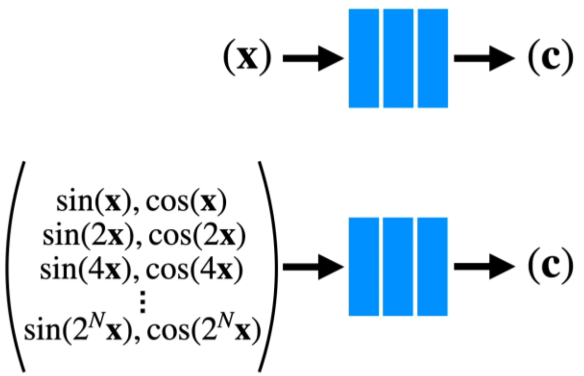
- How to get MLPs to represent high-frequency functions?



Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R., Ng, R..
NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis, ECCV, 2020

Neural Radiance Fields [Mildenhall et al. 2020]:

- Improvement for high-frequency embedding of input



NeRF (Naive)

NeRF (with positional encoding)

AI-based Scene Representation & Visualization

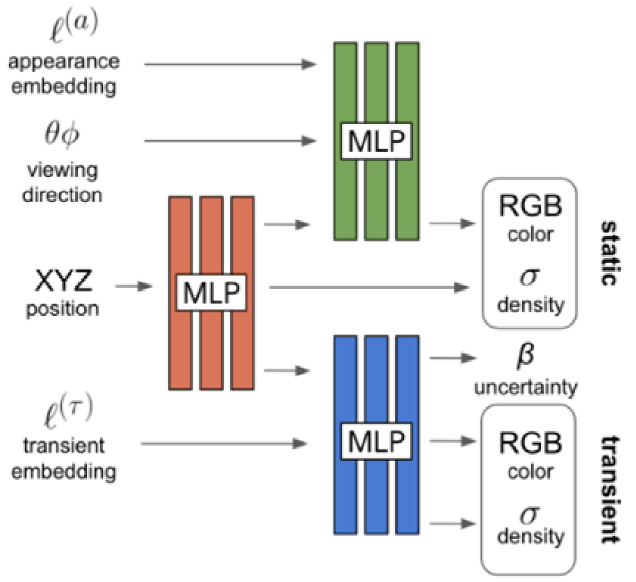
Results

32

Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R., Ng, R..
NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis, ECCV, 2020

AI-based Scene Representation & Visualization

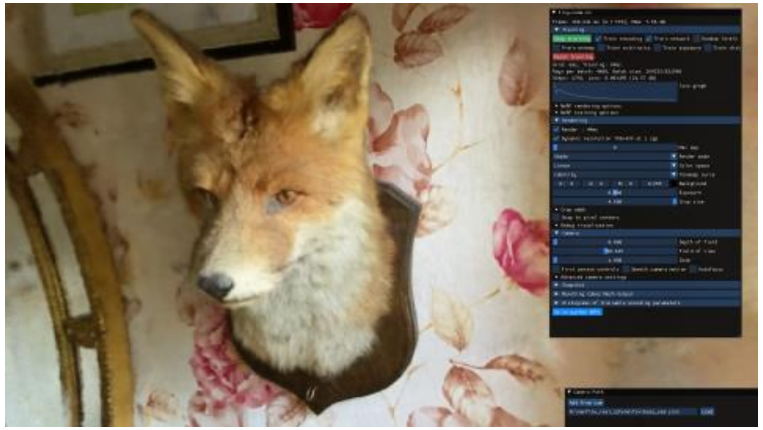
Extension to unconstrained photo collections …



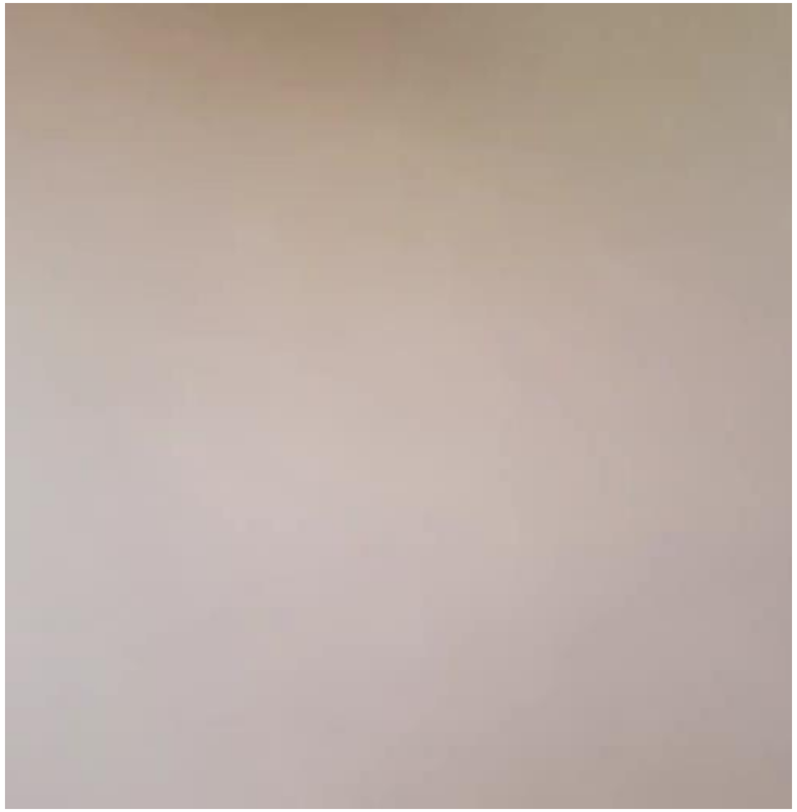
Martin-Brualla et al., NeRF in the Wild - Neural Radiance Fields for Unconstrained Photo Collections

AI-based Scene Representation & Visualization

... close to real-time ...



one of the input images



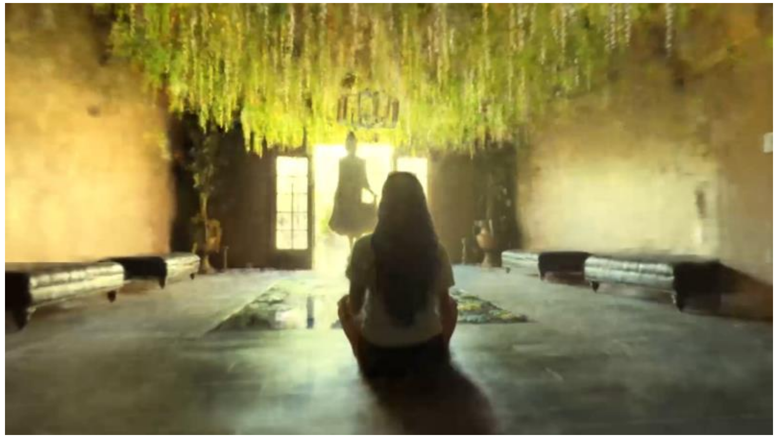
live training

Müller, T., Evans, A., Schied, C. and Keller, A., Instant Neural Graphics Primitives with a Multiresolution Hash Encoding, SIGGRAPH, 2022

AI-based Scene Representation & Visualization

... close to real-time ...

ongoing work on real-time NeRF

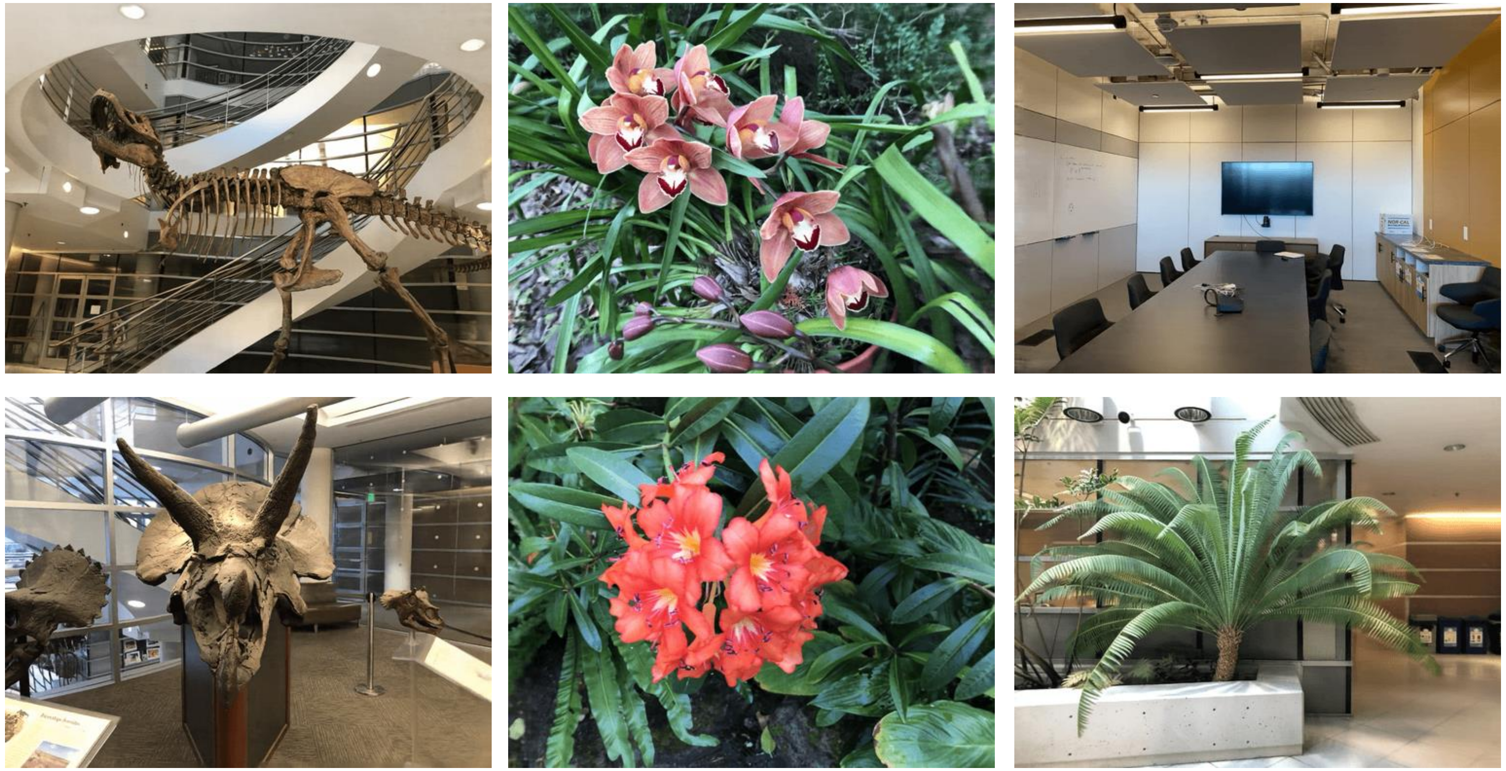


[AI Artists with Instant NeRF | NVIDIA]

Can we use AI to encode/represent whole scenes?

Results:

- Dense reconstruction (ideally surfaces …)
- Preservation of details
- Inspection under novel views



Images taken from *Alex Trevithick and Bo Yang. GRF: Learning a General Radiance Field for 3D Scene Representation and Rendering*

AI-based Scene Representation & Visualization

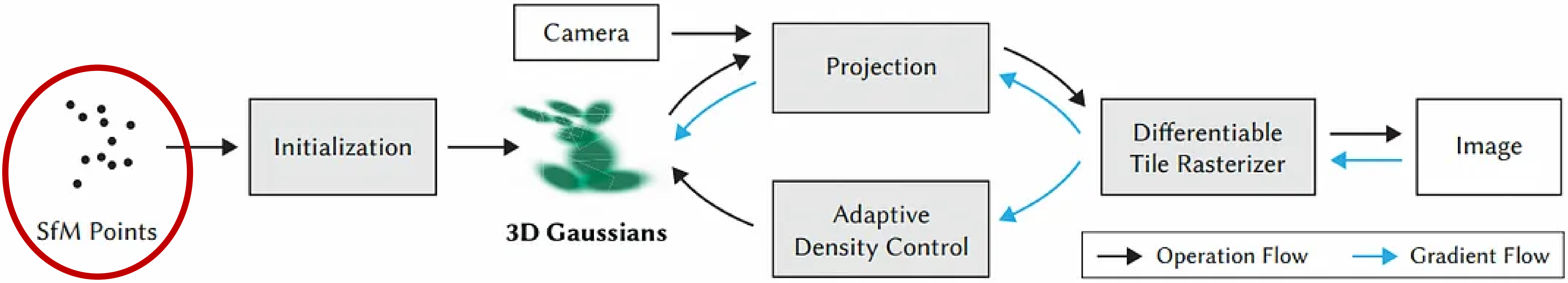
Current trend: Back to explicit 3D representations

- Keep optimizer, but replace neural network → 3D Gaussian Splatting

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

B. Kerbl, G. Kopanas, T. Leimkühler, G. Drettakis.
3D Gaussian Splatting for Real-Time Radiance Field Rendering, SIGGRAPH 2023

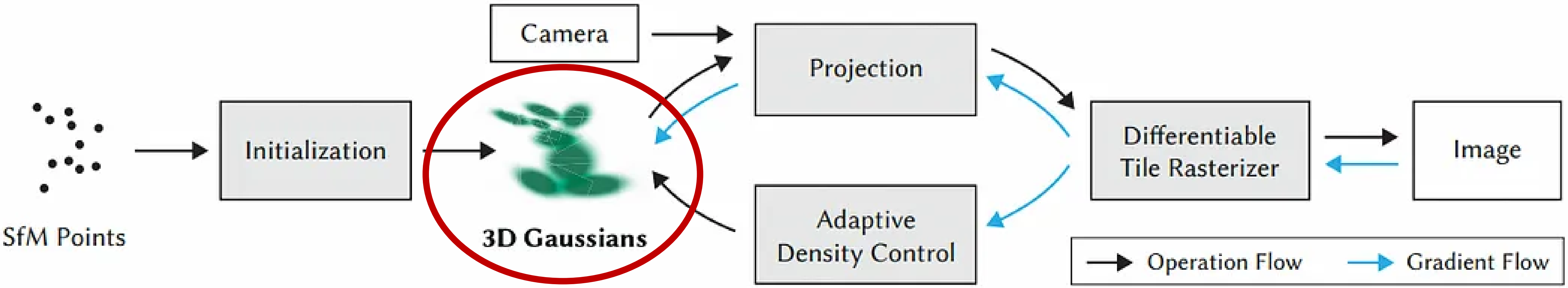
AI-based Scene Representation & Visualization



1) Initialization
(e.g., via SfM)

B. Kerbl, G. Kopanas, T. Leimkühler, G. Drettakis.
3D Gaussian Splatting for Real-Time Radiance Field Rendering, SIGGRAPH 2023

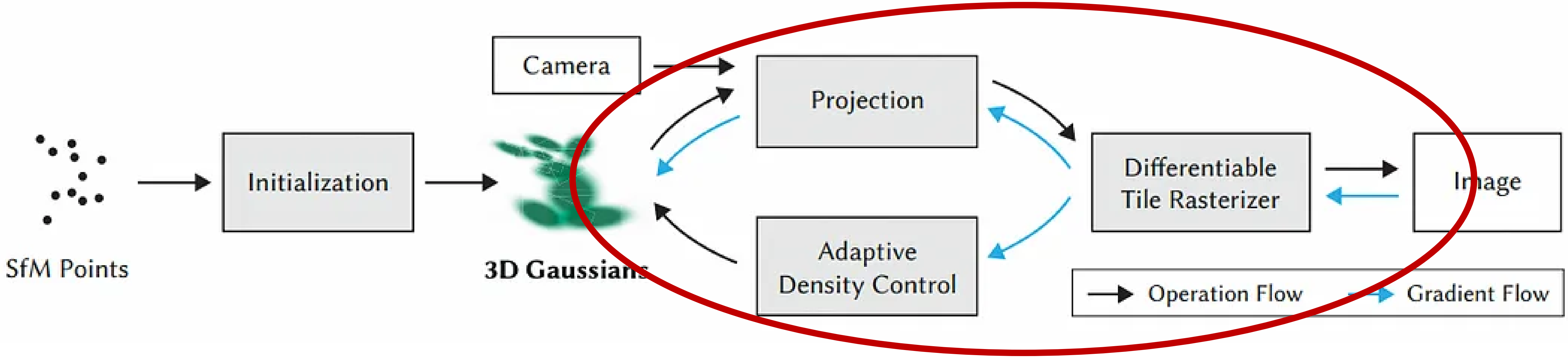
AI-based Scene Representation & Visualization



2) Create a set of Gaussians

B. Kerbl, G. Kopanas, T. Leimkühler, G. Drettakis.
3D Gaussian Splatting for Real-Time Radiance Field Rendering, SIGGRAPH 2023

AI-based Scene Representation & Visualization



3) Optimization and adaptive control of the density of this set of Gaussians (involves tile-based renderer)

B. Kerbl, G. Kopanas, T. Leimkühler, G. Drettakis.
3D Gaussian Splatting for Real-Time Radiance Field Rendering, SIGGRAPH 2023

AI-based Scene Represent

accurate novel view synthesis

fast

huge memory footprint

lots of views (with camera parameters) required

Current trend: Back to explicit 3D

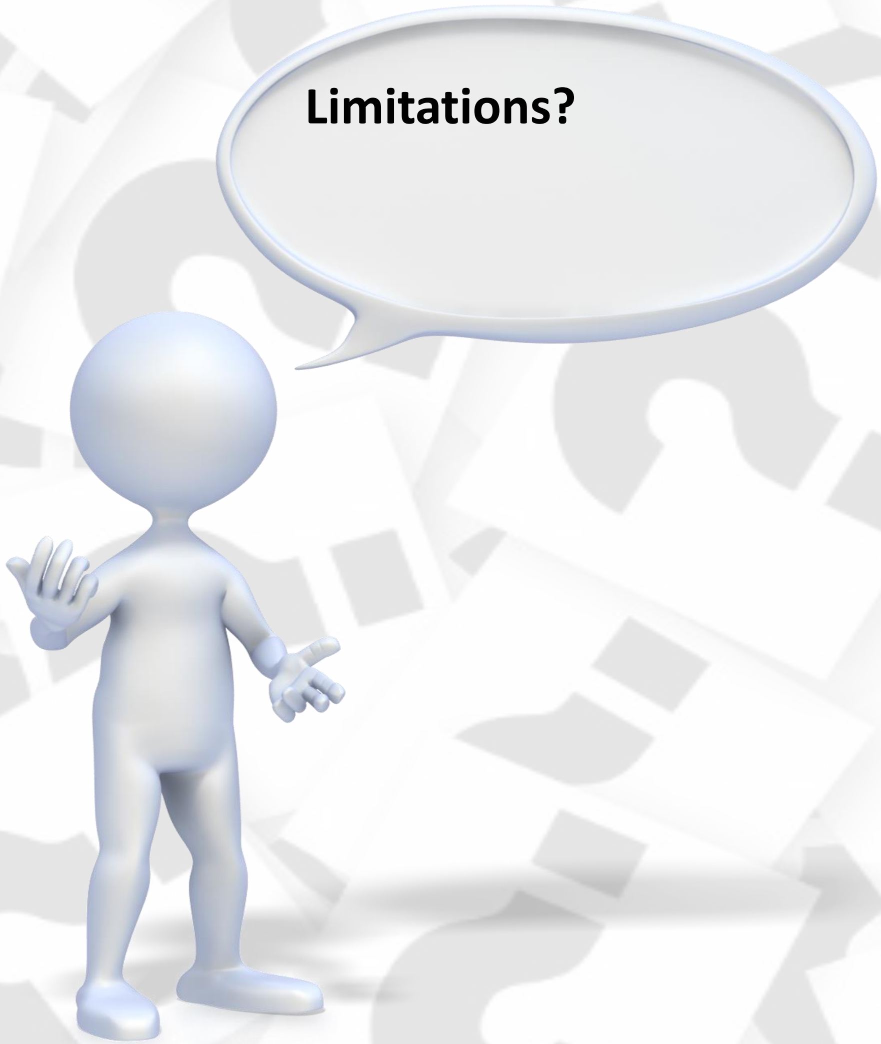
- Keep optimizer, but replace neura


https://www.bizzlogic.com/insights/gaussian-splatting


https://www.magnopus.com/blog/the-rise-of-3d-gaussian-splatting

B. Kerbl, G. Kopanas, T. Leimkühler, G. Drettakis. 3D Gaussian Splatting for Real-Time Radiance Field Rendering, SIGGRAPH 2023

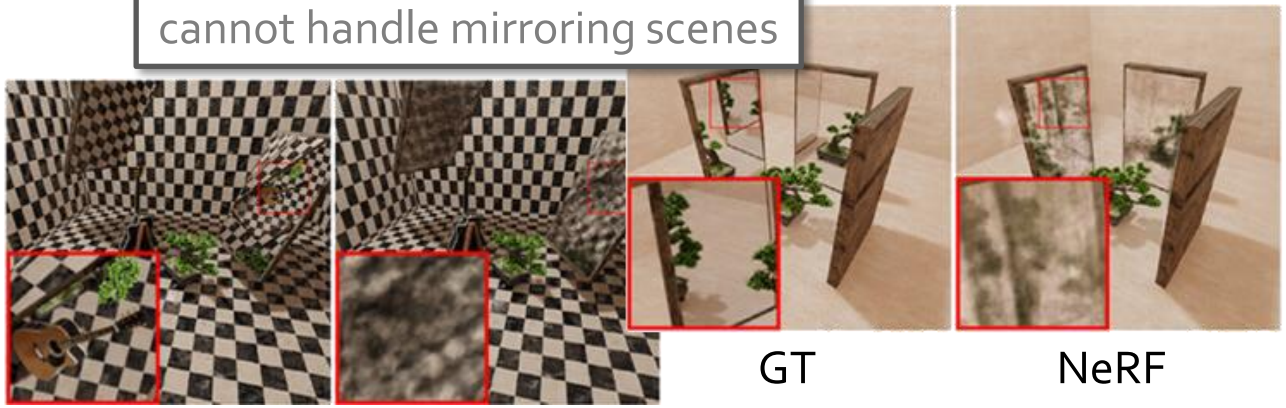
AI-based Scene Capture



Remaining Limitations

Challenges for NeRFs / Gaussian Splatting:

cannot handle mirroring scenes



GT NeRF GT NeRF

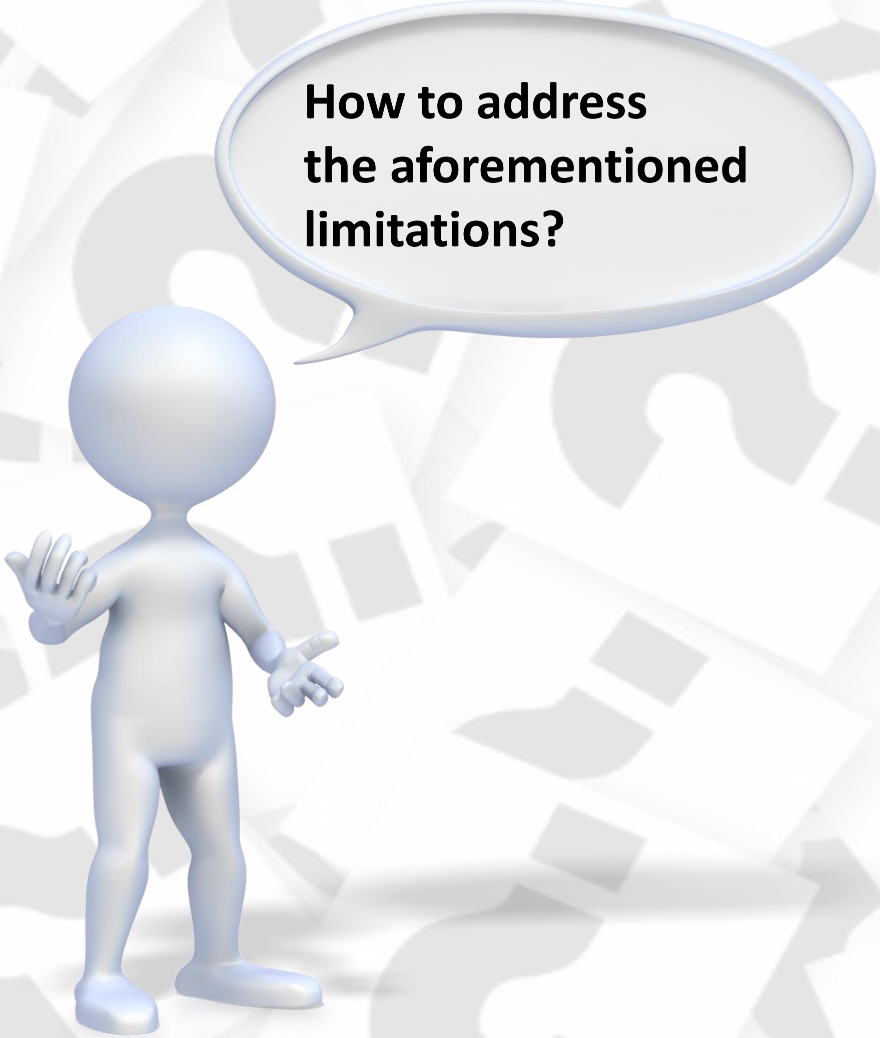
lacking robustness in sparse-view scenarios

lacking robustness to distractors

occluders noisy camera poses blurred frames

limitation to RGB-based scene representation

AI-based Scene Capture



Robustness to Distractors

Observation:

- NeRFs suffer from artifacts by distraction and occlusion

B. Buschmann, A. Dogaru, E. Eisemann, M. Weinmann, B. Egger,
RANRAC: Robust Neural Scene Representations via Random Ray Consensus, ECCV 2024

Robustness to Distractors

Idea:

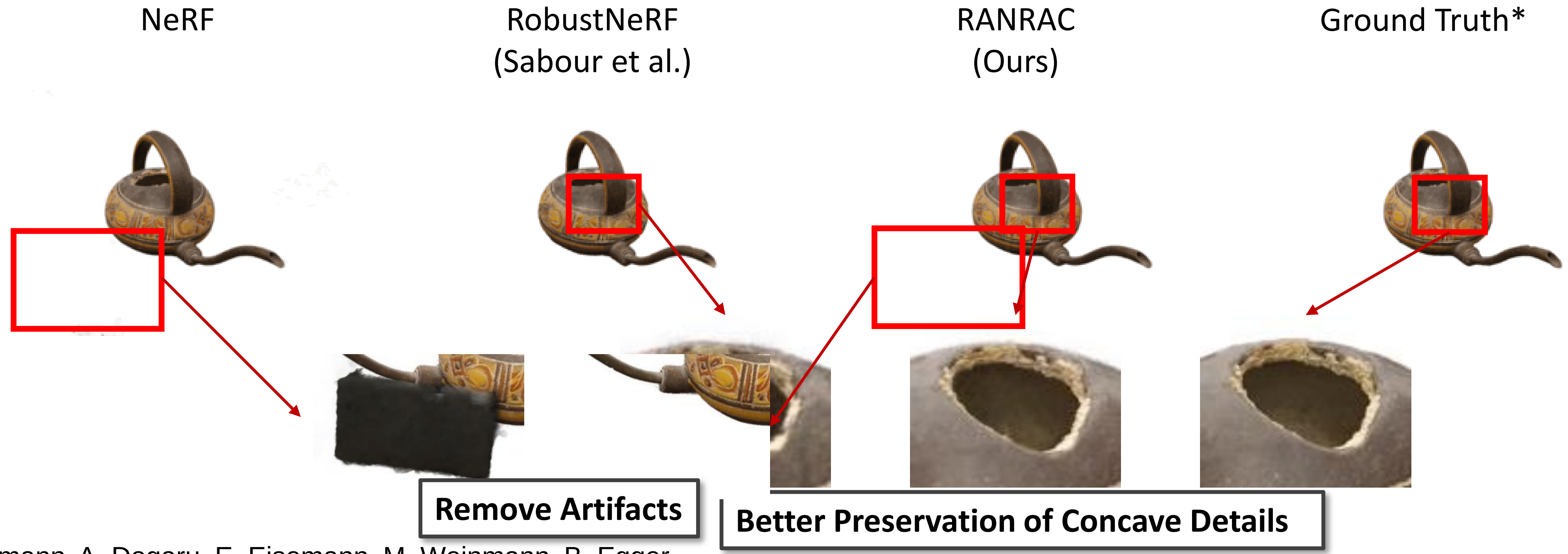
- Outlier filtering via RANSAC-like scheme
  Random **Ra**y **C**onsensus (iterative subset evaluation)
  - **S**ampling of observations from the images with camera poses
  - Fitting of hypothesis/model
  - Use obtained model to render images for unused input views
  - Hypothesis validation (→ inlier/outlier classification)
  - **Selection of model** with most inliers and re-estimation of model based on this consensus set

B. Buschmann, A. Dogaru, E. Eisemann, M. Weinmann, B. Egger,
RANRAC: Robust Neural Scene Representations via Random Ray Consensus, ECCV 2024

Robustness to Distractors

Idea:

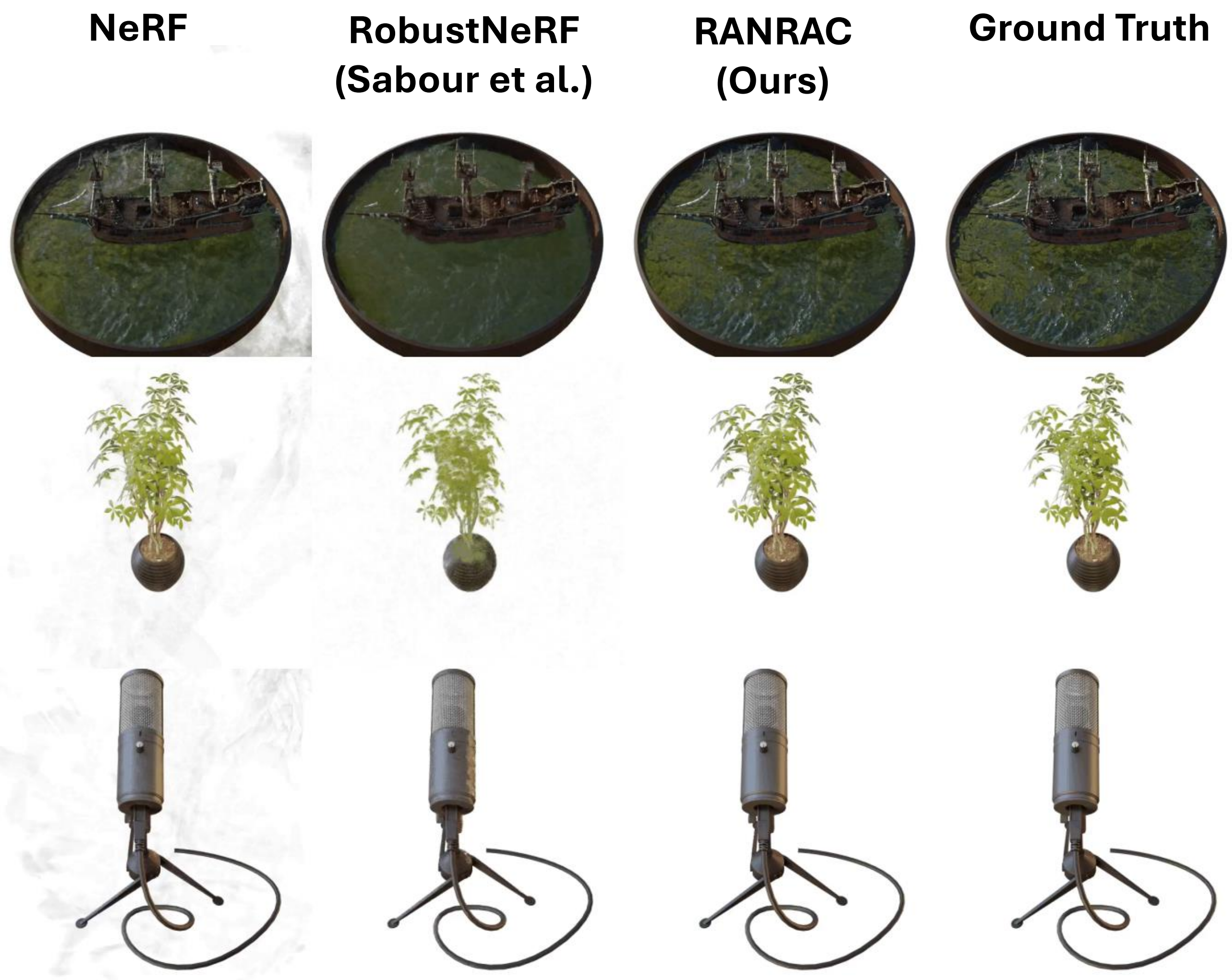
- Outlier filtering via **Ran**dom **Ra**y **C**onsensus (iterative subset evaluation)



B. Buschmann, A. Dogaru, E. Eisemann, M. Weinmann, B. Egger,
RANRAC: Robust Neural Scene Representations via Random Ray Consensus, ECCV 2024

Robustness to Distractors

Robustness to noisy poses



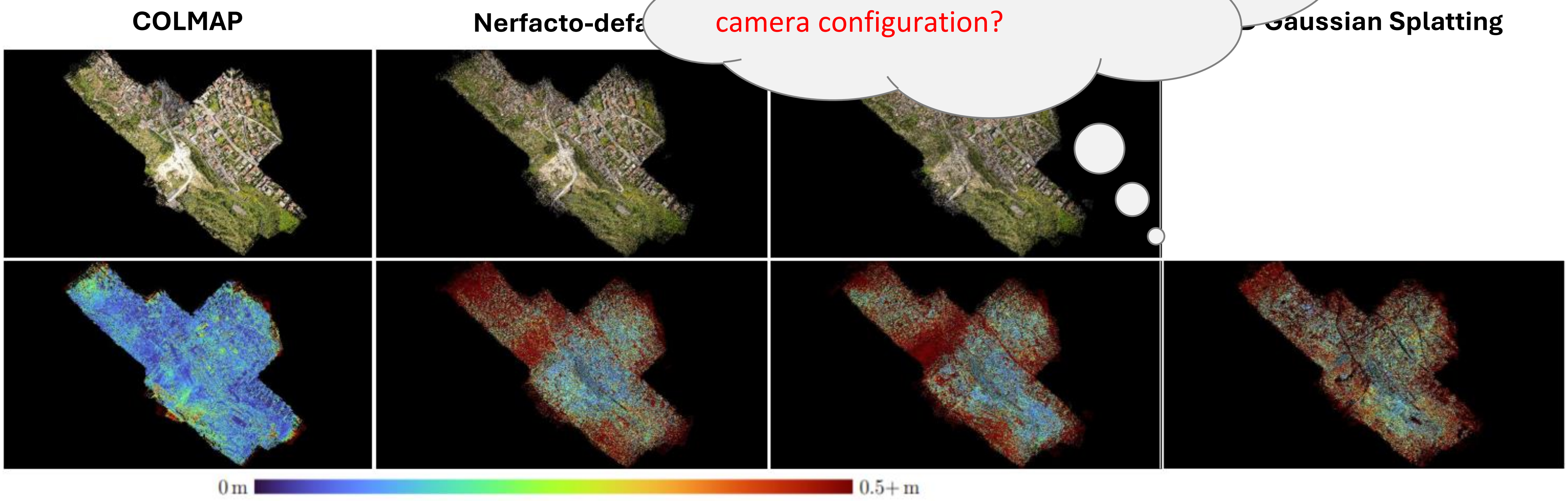
B. Buschmann, A. Dogaru, E. Eisemann, M. Weinmann, B. Egger,
RANRAC: Robust Neural Scene Representations via Random Ray Consensus, ECCV 2024

Sparse-view Scenarios

Why do NeRF and GS perform so bad?

too few views per scene part?

(almost) nadir views → too constrained camera configuration?

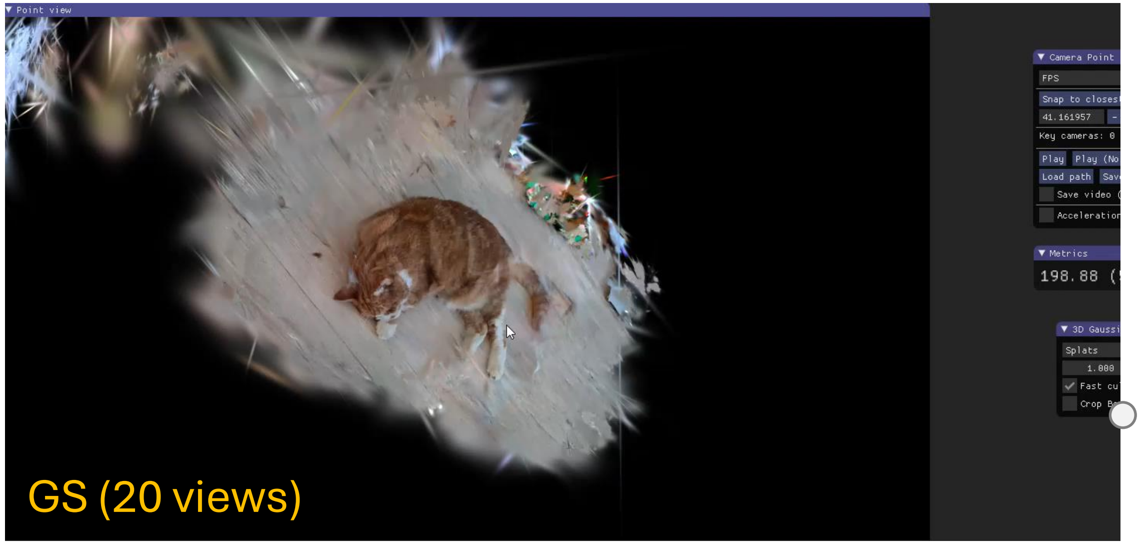


D. Haitz, M. Hermann, A. S. Roth, M. Weinmann, M. Weinmann. The Potential of Neural Radiance Fields and 3D Gaussian Splatting for 3D Reconstruction from Aerial Imagery, ISPRS Annals, 2024

Sparse-view Scenarios

Observation:

- NeRFs and Gaussian splatting produces a lot of noise for configurations with only a few views



Can we improve robustness to handle sparse-view scenarios?

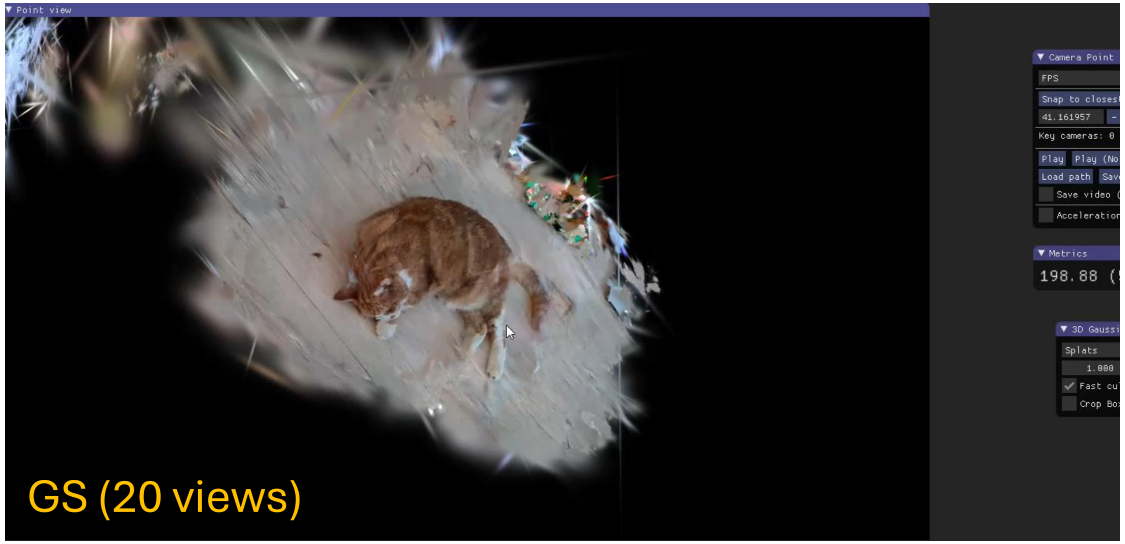
B. Buschmann, E. Eisemann, M. Weinmann.
ongoing work on robust sparse-view Gaussian splatting

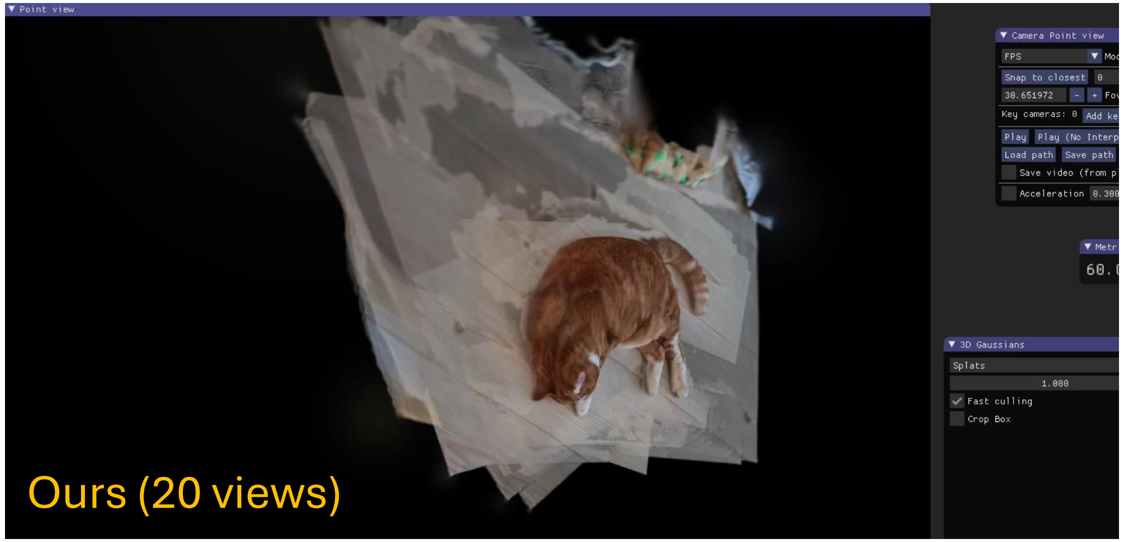
Wang et al., PriNeRF, Prior-constrained Neural Radiance Fields for Robust Novel View Synthesis of Urban Scenes with Fewer Views, ISPRS Journal on Photogrammetry and Remote Sensing, 2024

Handling Sparse-view Scenarios

Potential solution for few-shot capture

- Integration of neural inference of 2D-3D mappings, 3D Gaussian Splatting and camera refinement





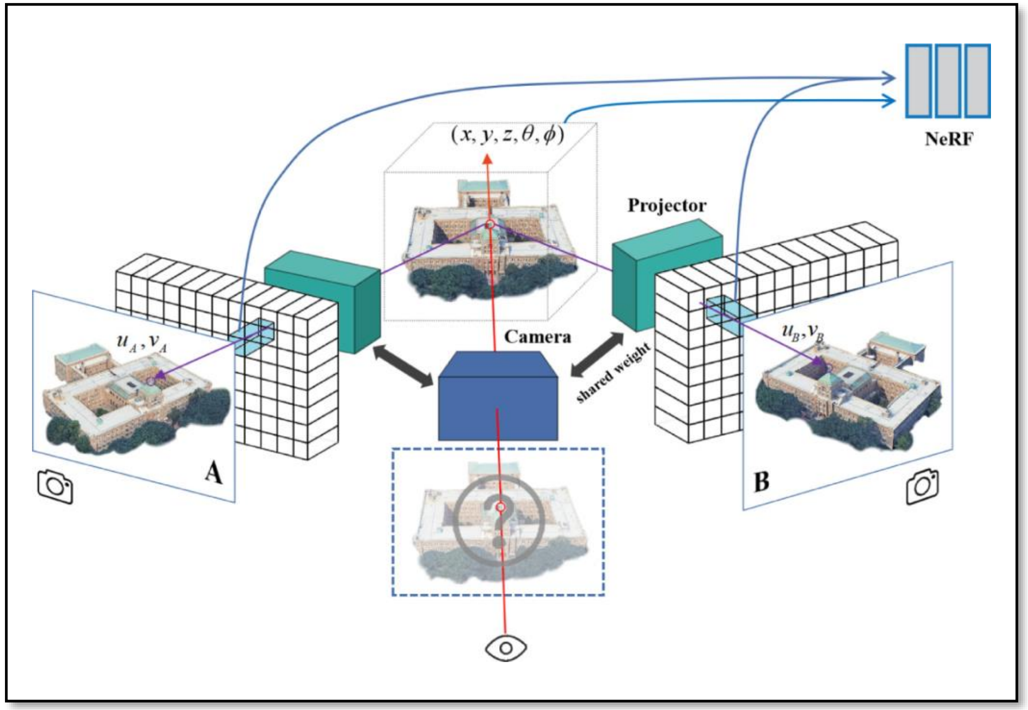
B. Buschmann, E. Eisemann, M. Weinmann.
ongoing work on robust sparse-view Gaussian splatting

Improved practical relevance!
(robustness, few-shot, less priors, …)

Handling Sparse-view Scenarios

Potential solution for few-shot capture

- Scene priors → feature space (beyond RGB)
  - Projection of sampling points on the ray emitted from the target view into the reference views
  - Obtain the features in the reference views by interpolation

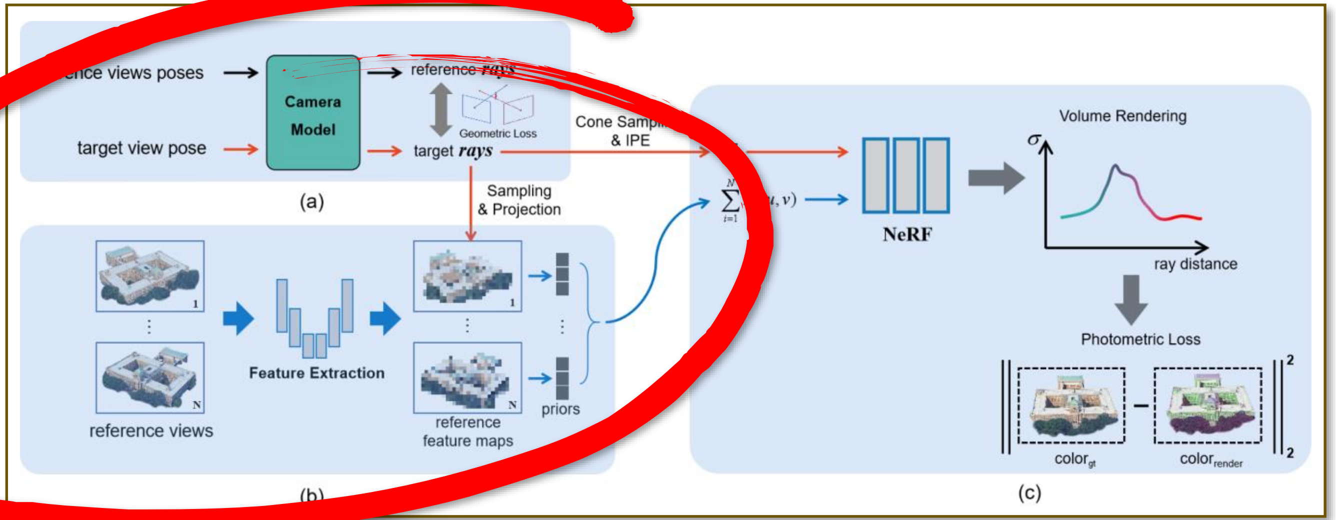


Wang et al., PriNeRF, Prior-constrained Neural Radiance Fields for Robust Novel View Synthesis of Urban Scenes with Fewer Views, ISPRS Journal on Photogrammetry and Remote Sensing, 2024

Handling Sparse-view Scenarios

Potential solution for few-shot capture

- Scene priors → feature space (beyond RGB)



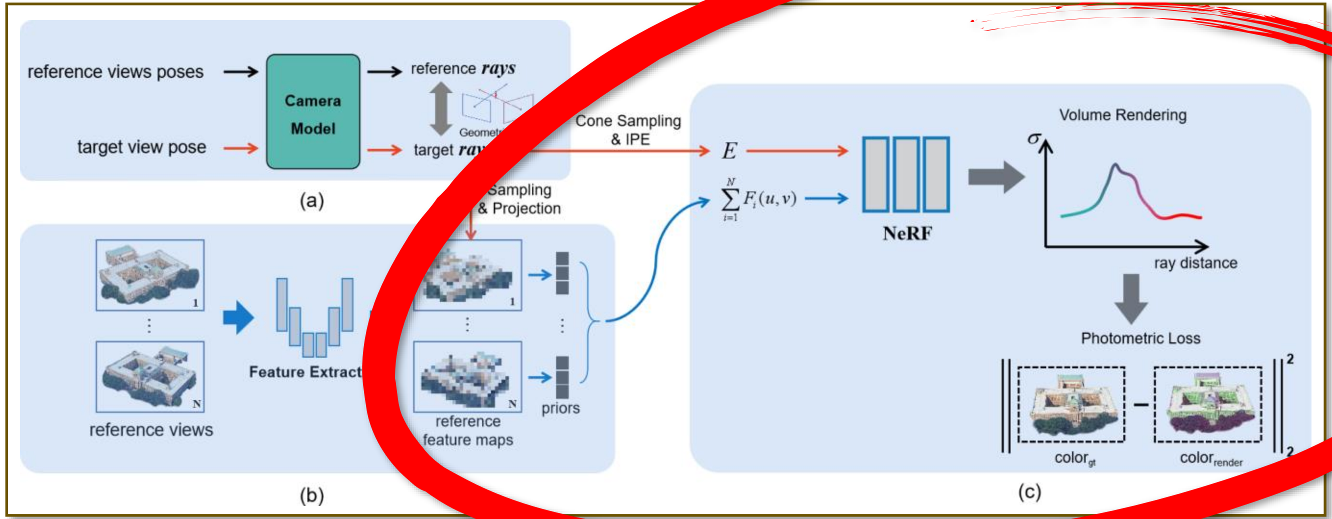
Feature extraction module:
- extracts deep features of the reference view image
- aggregates the features under different reference views as priors

Wang et al., PriNeRF, Prior-constrained Neural Radiance Fields for Robust Novel View Synthesis of Urban Scenes with Fewer Views, ISPRS Journal on Photogrammetry and Remote Sensing, 2024

Handling Sparse-view Scenarios

Potential solution for few-shot capture

- Scene priors → feature space (beyond RGB)



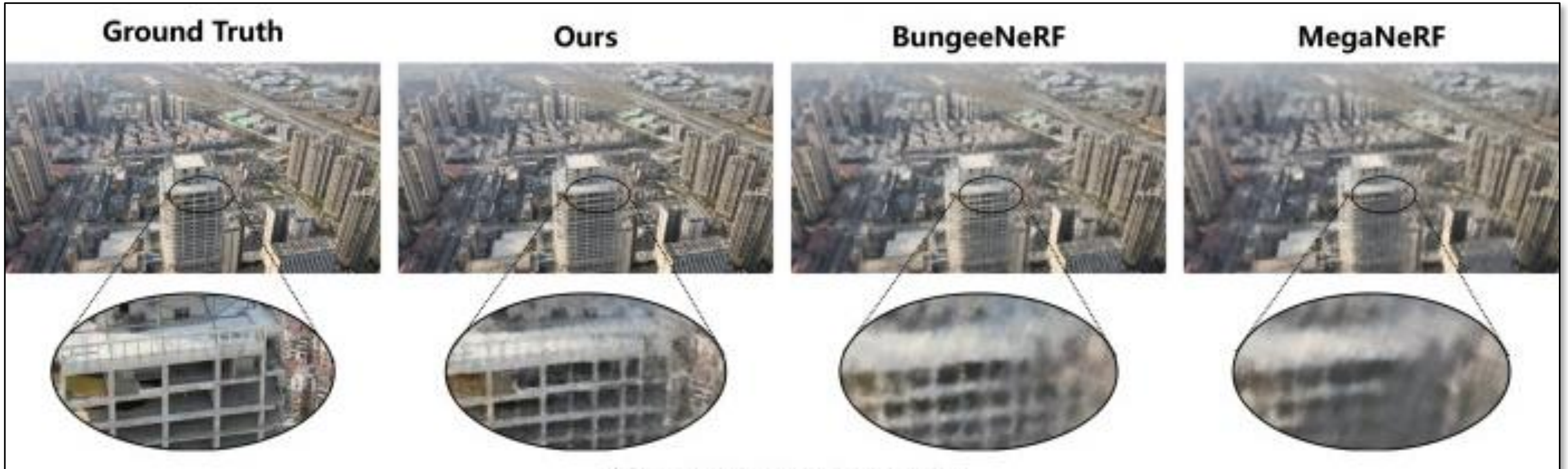
NeRF takes the scene priors along with position and orientation encodings to generate color and opacity

Wang et al., PriNeRF, Prior-constrained Neural Radiance Fields for Robust Novel View Synthesis of Urban Scenes with Fewer Views, ISPRS Journal on Photogrammetry and Remote Sensing, 2024

Handling Sparse-view Scenarios

Potential solution for few-shot capture

- Scene priors + camera refinement



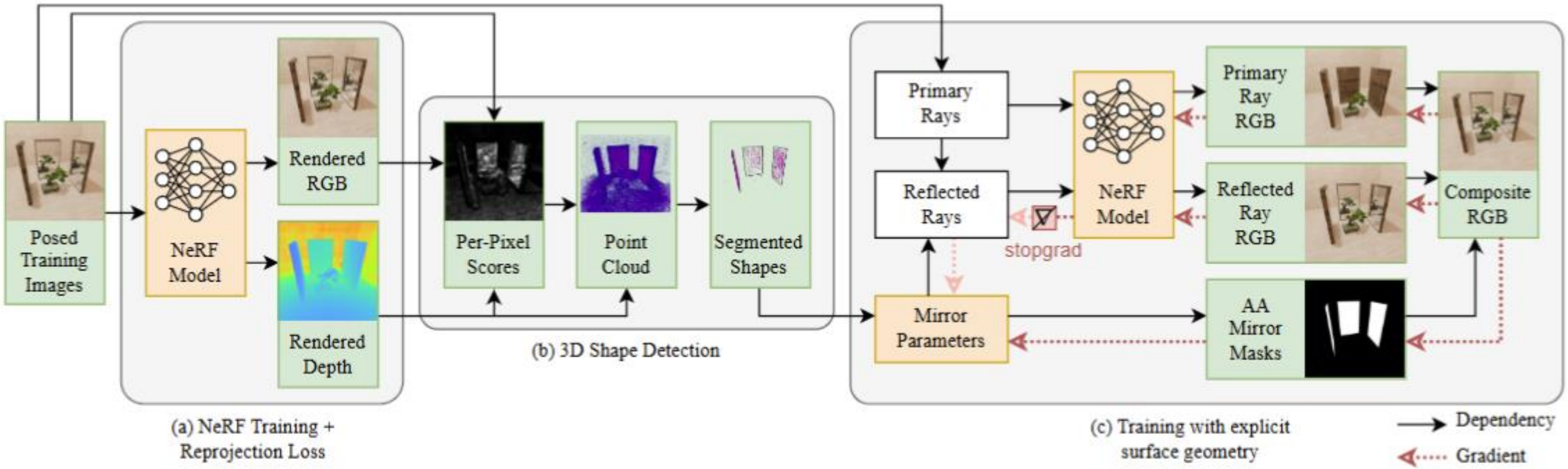
(d) Greenland Xindu Mall, Hefei.

Wang et al., PriNeRF, Prior-constrained Neural Radiance Fields for Robust Novel View Synthesis of Urban Scenes with Fewer Views, ISPRS Journal on Photogrammetry and Remote Sensing, 2024

Handling Scenes with Mirroring Surfaces

Idea: Leverage inconsistencies in observations for mirroring surfaces

- Allows detection and handling of mirrors

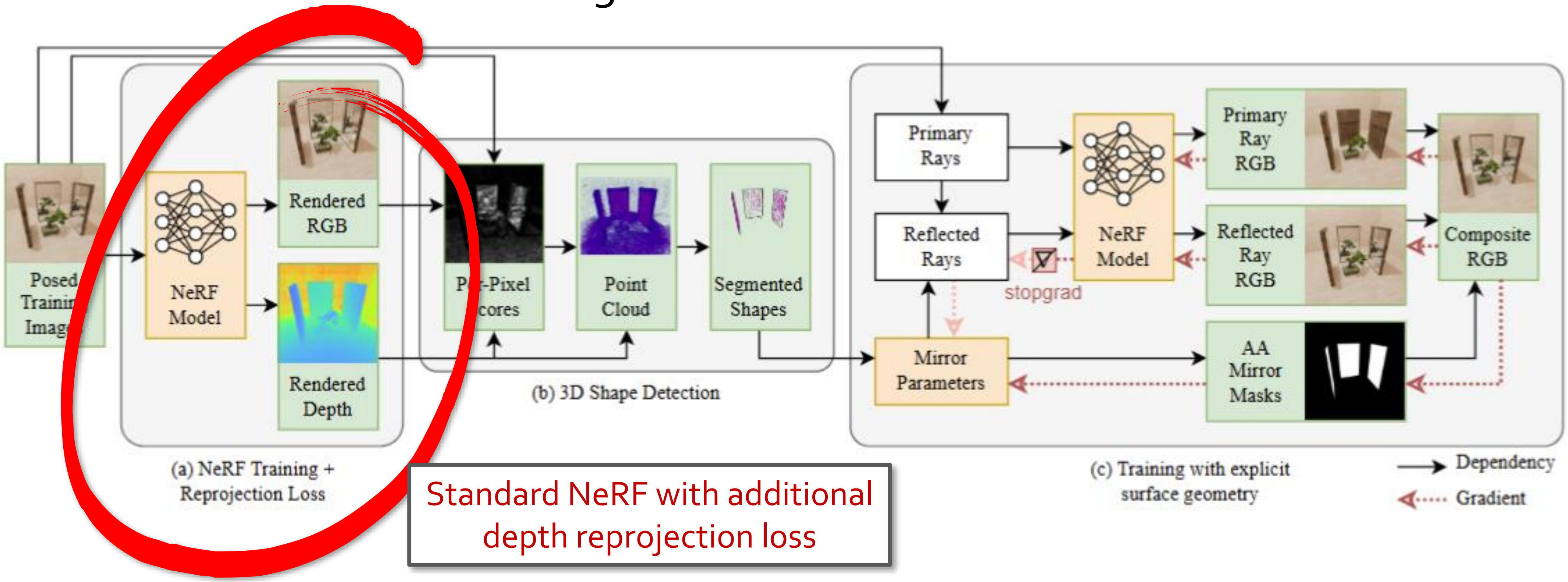


L. V. Holland, M. Weinmann, P. Stotko and R. Klein, NeRFs are Mirror Detectors: Using Structural Similarity for Multi-View Mirror Scene Reconstruction with 3D Surface Primitives, WACV 2025

Handling Scenes with Mirroring Surfaces

Idea: Leverage inconsistencies in observations for mirroring surfaces

- Allows detection and handling of mirrors

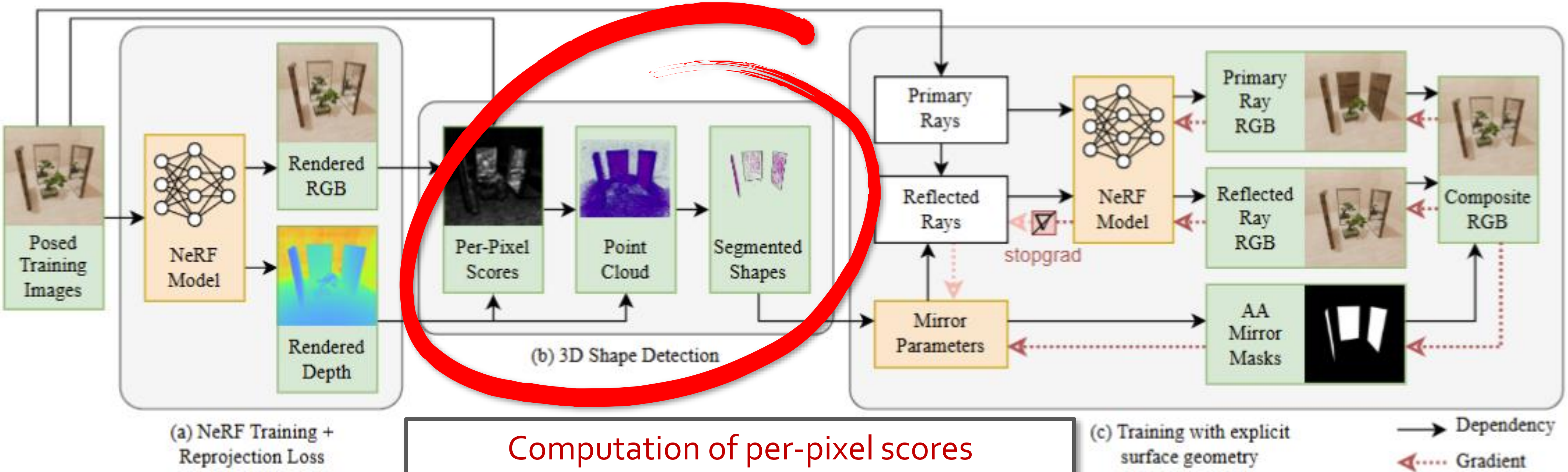


L. V. Holland, M. Weinmann, P. Stotko and R. Klein, NeRFs are Mirror Detectors: Using Structural Similarity for Multi-View Mirror Scene Reconstruction with 3D Surface Primitives, WACV 2025

Handling Scenes with Mirroring Surfaces

Idea: Leverage inconsistencies in observations for mirroring surfaces

- Allows detection and handling of mirrors



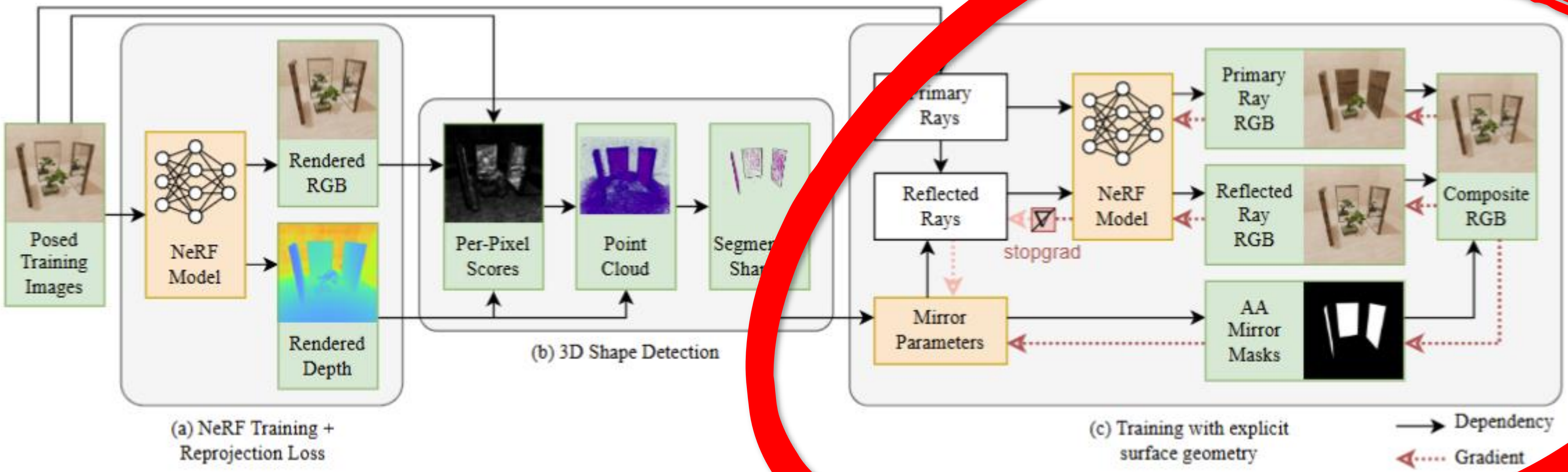
Computation of per-pixel scores
(using SSIM and depth variance values)

L. V. Holland, M. Weinmann, P. Stotko and R. Klein, NeRFs are Mirror Detectors: Using Structural Similarity for Multi-View Mirror Scene Reconstruction with 3D Surface Primitives, WACV 2025

Handling Scenes with Mirroring Surfaces

Idea: Leverage inconsistencies in observations for mirroring surfaces

- Allows detection and handling of mirrors



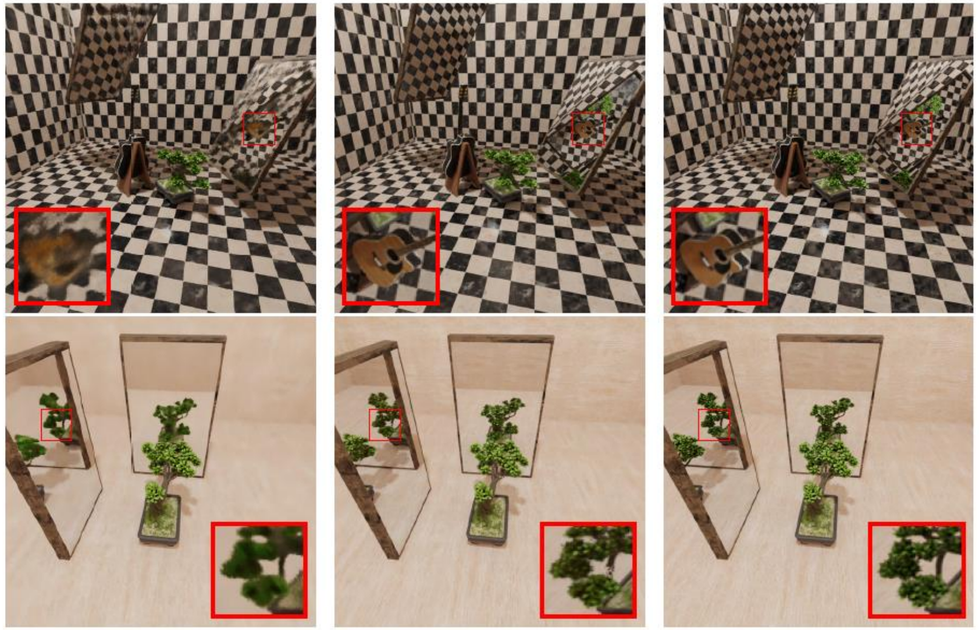
modified pipeline to jointly optimize NeRF and mirror parameters by blending primary and reflected images based on anti-aliased mirror masks that are generated in differentiable manner

L. V. Holland, M. Weinmann, P. Stotko and R. Kl
Using Structural Similarity for Multi-View Mirr

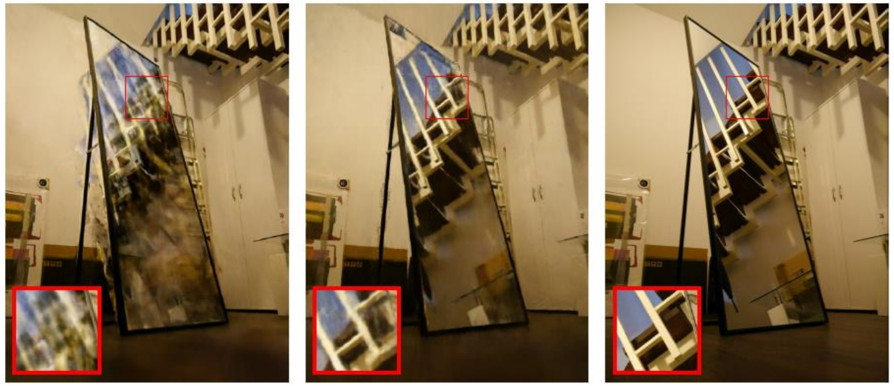
Handling Scenes with Mirroring Surfaces

Idea: Leverage inconsistencies in observations for mirroring surfaces

- Allows detection and handling of mirrors



Best Baseline MS-NeRF [112] | Ours | Ground Truth



Best Baseline Mip-NeRF 360 [3] | Ours | Ground Truth

MS-NeRF [112] PSNR: 33.55 dB | Ours PSNR: 35.27 dB | Ground Truth

L. V. Holland, M. Weinmann, P. Stotko and R. Klein, NeRFs are Mirror Detectors: Using Structural Similarity for Multi-View Mirror Scene Reconstruction with 3D Surface Primitives, WACV 2025

Semantics-informed Scene Representation

Idea:

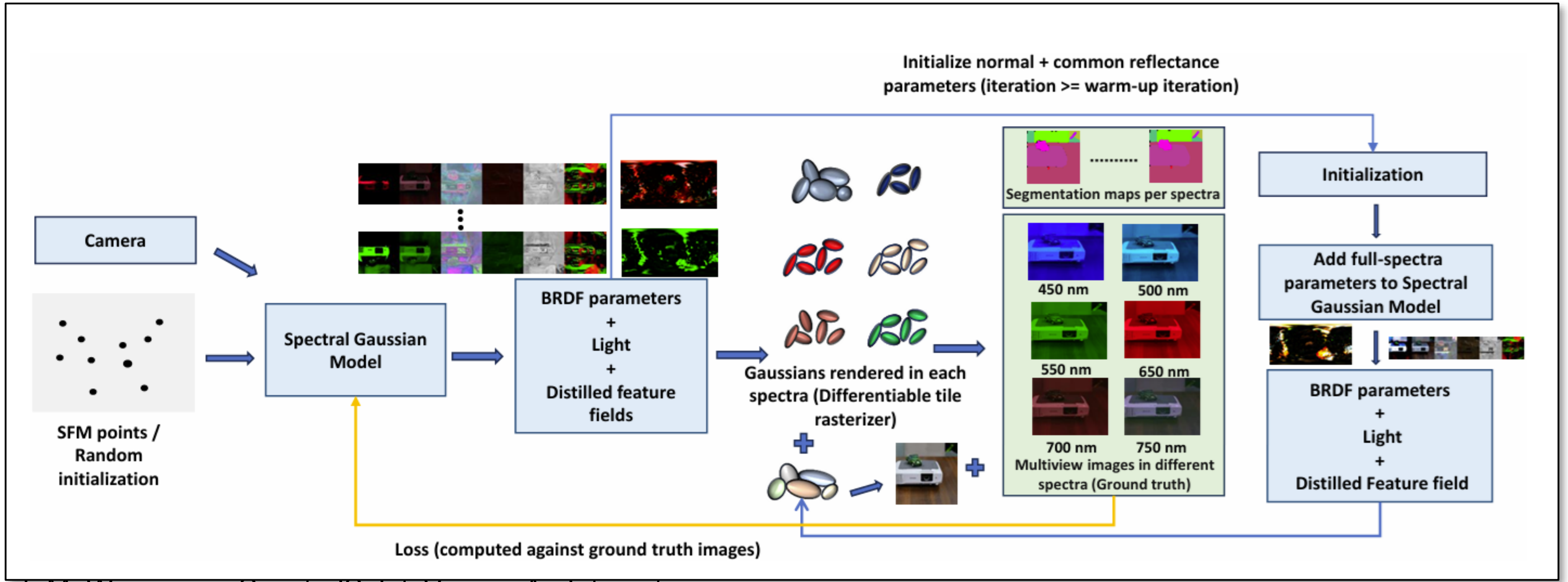
- Scene representation with semantics
  e.g. *Gaussian Splats* + *Segment Anything*

Ground truth (Multi-view input frames) | Rendered Gaussian Splats | Rendered Segmentation maps | Distilled feature fields (Object-Ids for scene editing)

Extension to Multi-Spectral Scene Representation

Idea:

- Representations per sub-band combined with material-wise shading model

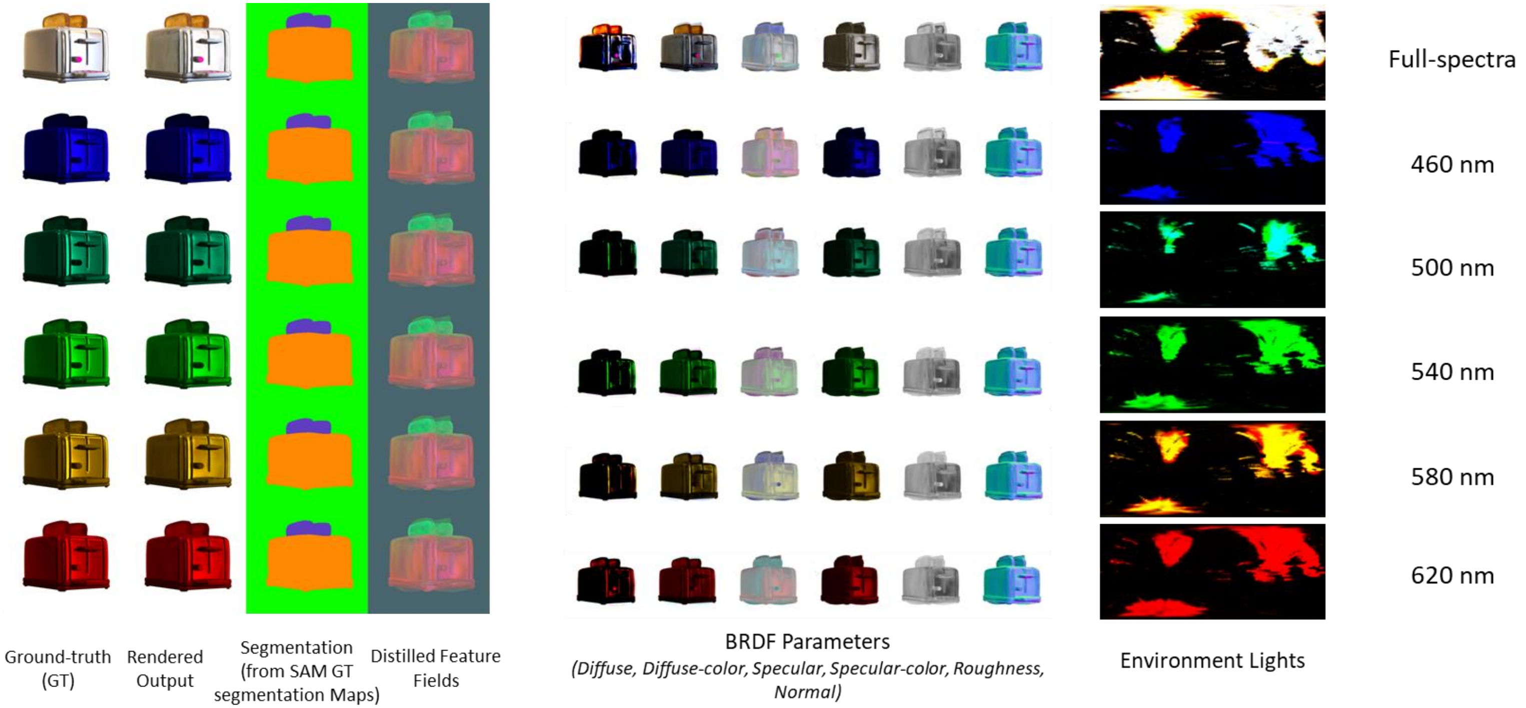


S. N. Sinha, J.Kühn, H. Graf, M. Weinmann. SpectralSplatsViewer: An Interactive Web-Based Tool for Visualizing Cross-Spectral Gaussian Splats, Web3D, 2024

S. N. Sinha, H. Graf, M. Weinmann. SpectralGaussians: A relightable spectral Gaussian splatting framework to generate photorealistic semantic Gaussian splats in different spectra, ongoing work

Idea:

- Representations per sub-band combined with material-wise shading model

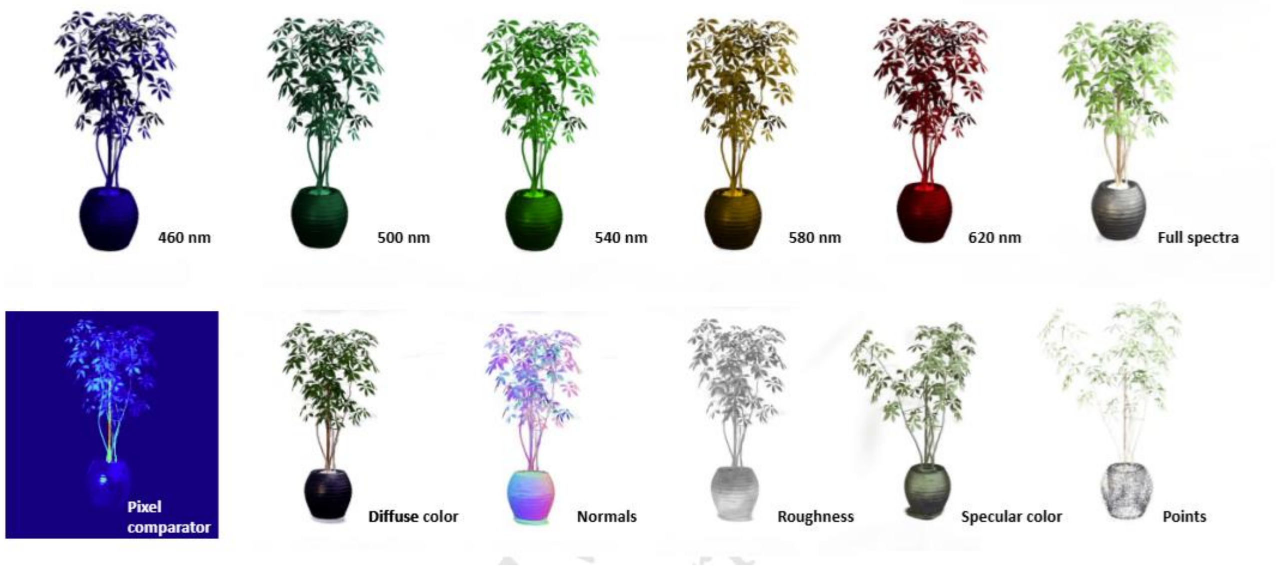


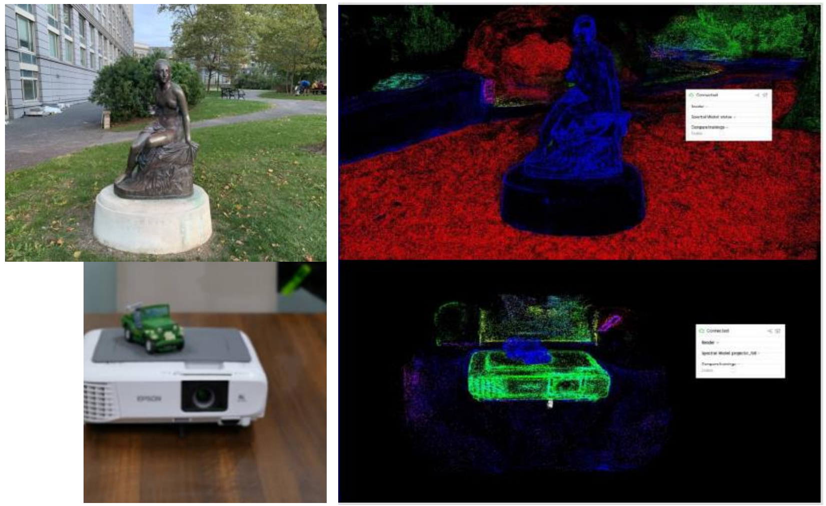
S. N. Sinha, H. Graf, M. Weinmann. SpectralGaussians: A relightable spectral Gaussian splatting framework to generate photorealistic semantic Gaussian splats in different spectra, ongoing work

Extension to Multi-Spectral Scene Representation

Results:

Method	Spectral NeRF Synthetic Dataset[27]						Average
	kitchen	Living room	Digger	Spaceship	Vintage car	Cartoon knight	
	PSNR ↑						
NeRF[7]	34.583	33.172	30.658	30.126	33.478	34.485	32.400
Mip-NeRF[8]	-	-	33.301	31.495	33.883	35.102	33.945
Aug-NeRF[37]	34.480	32.540	31.538	30.929	33.639	33.908	32.677
SpectralNeRF[27]	35.115	33.665	33.378	31.951	34.480	34.915	33.610
Ours	37.035	37.989	40.218	41.233	42.636	36.723	38.456
	SSIM ↑						
NeRF[7]	0.8943	0.9929	0.9187	0.9358	0.7958	0.9273	0.9123
Mip-NeRF[8]	-	-	0.9290	0.9475	0.8166	0.9572	0.9126
Aug-NeRF[37]	0.9026	0.9649	0.9248	0.9402	0.8002	0.9287	0.9163
SpectralNeRF[27]	0.9117	0.9931	0.9357	0.9482	0.8169	0.9573	0.9349
Ours	0.9747	0.9733	0.9923	0.9951	0.9893	0.9572	0.9801
	LPIPS ↓						
NeRF[7]	0.1650	0.0578	0.0413	0.0275	0.1319	0.1545	0.0722
Mip-NeRF[8]	-	-	0.0435	0.0535	0.1747	0.1526	0.1061
Aug-NeRF[37]	0.1603	0.0706	0.0341	0.0389	0.1536	0.1705	0.0973
SpectralNeRF[27]	0.1637	0.0479	0.0259	0.0250	0.1499	0.1510	0.0733
Ours	0.0739	0.0525	0.0109	0.0084	0.0527	0.0741	0.0438





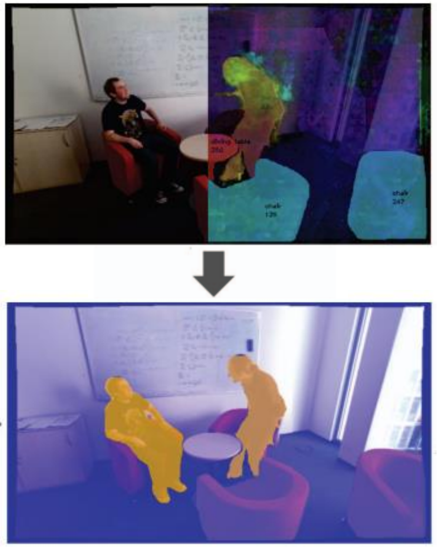
S. N. Sinha, J.Kühn, H. Graf, M. Weinmann. SpectralSplatsViewer: An Interactive Web-Based Tool for Visualizing Cross-Spectral Gaussian Splats, Web3D, 2024

S. N. Sinha, H. Graf, M. Weinmann. SpectralGaussians: A relightable spectral Gaussian splatting framework to generate photorealistic semantic Gaussian splats in different spectra, ongoing work

Further Trends

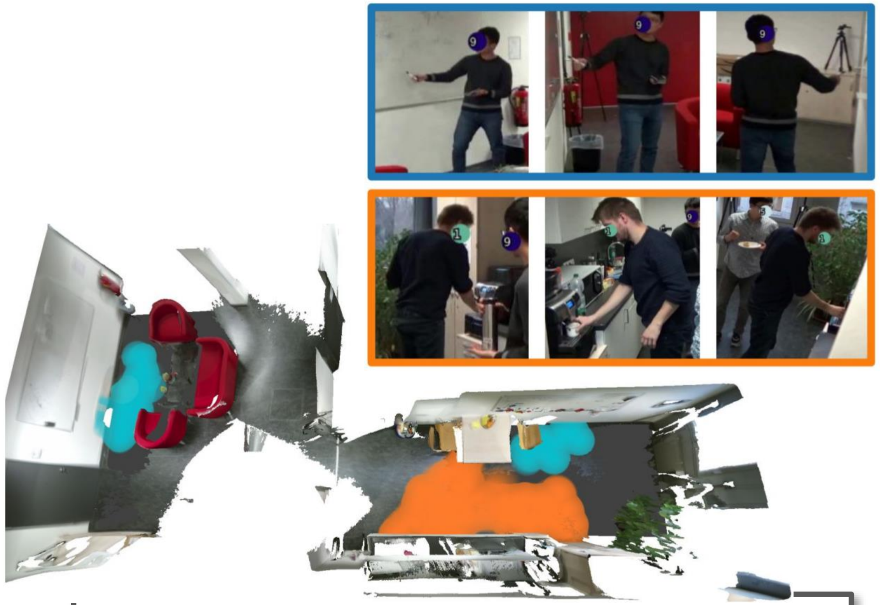
Tighter integration of context into inference/capture

- ... behavioral patterns



hybrid (static+dynamic) scene representation

L. V. Holland, P. Stotko, S. Krumpen, R. Klein, M. Weinmann. Efficient 3D Reconstruction, Streaming and Visualization of Static and Dynamic Scene Parts for Multi-client Live-telepresence in Large-scale Environments, ICCV Workshops 2023



activity maps and action recognition

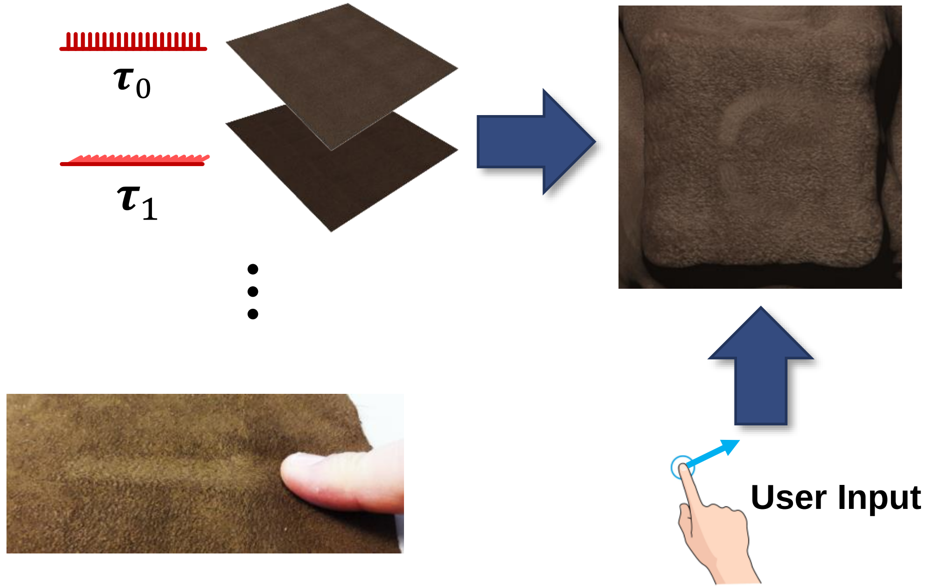
J. Tanke et al., Bonn Activity Maps: Dataset Description, arXiv preprint arXiv:1912.06354, 2019

human-aware placement of service robots

L. Bruckschen, K. Bungert, M. Wolter, S. Krumpen, M. Weinmann, R. Klein, M. Bennewitz. Where can i help? Human-aware placement of service robots, International Conference on Robot and Human Interactive Communication, 2020

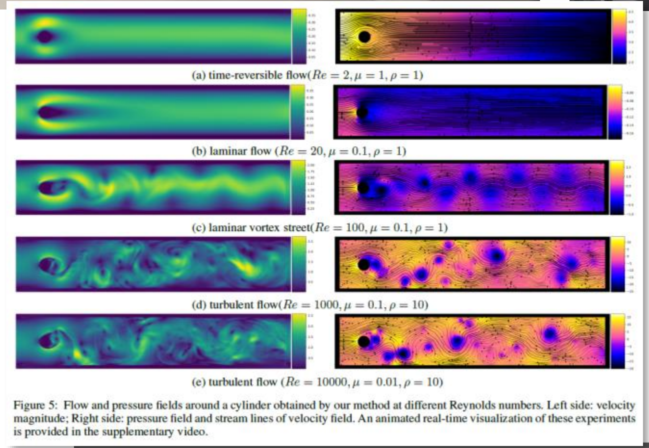
Further Trends

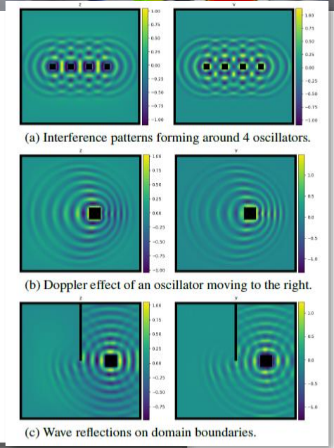
Combination with simulations



interactive simulation of material appearance

S. Krumpen, M. Weinmann, R. Klein. Interactive Appearance Manipulation of Fiber-based Materials, International Conference on Computer Graphics Theory and Applications 2017





Physics-informed machine learning

N. Wandel, M. Weinmann, R. Klein. Learning Incompressible Fluid Dynamics from Scratch-Towards Fast, Differentiable Fluid Models that Generalize, ICLR 2021

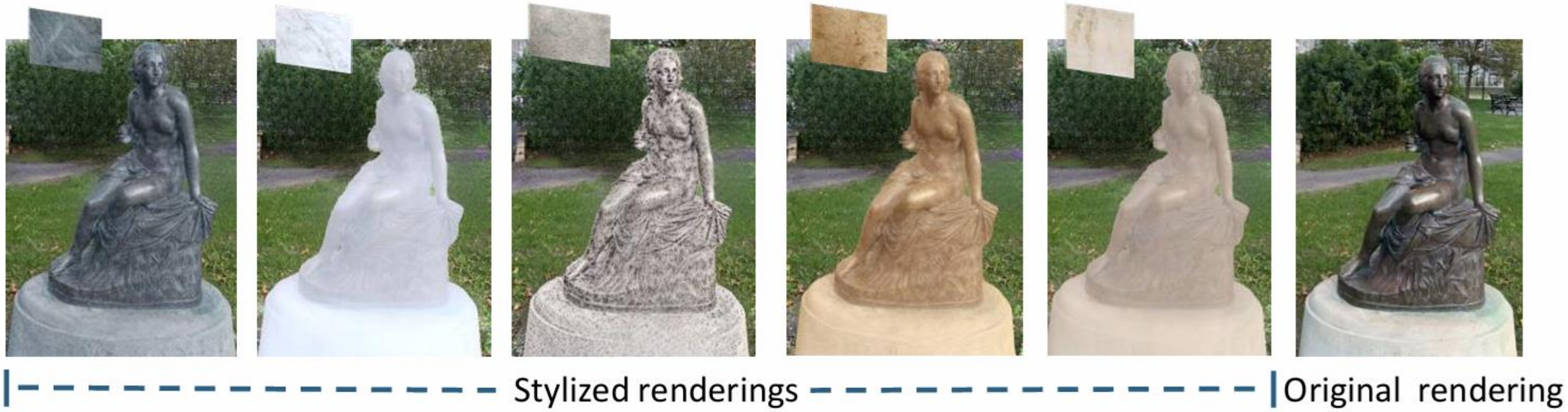
N. Wandel, M. Weinmann, R. Klein. Teaching the incompressible Navier–Stokes equations to fast neural surrogate models in three dimensions, Physics of Fluids, 2021

N. Wandel, M. Weinmann, M. Neidlin, R. Klein. Spline-PINN: Approaching PDEs without data using fast, physics-informed Hermite-spline CNNs, AAAI 2022

Further Trends

Editable representations

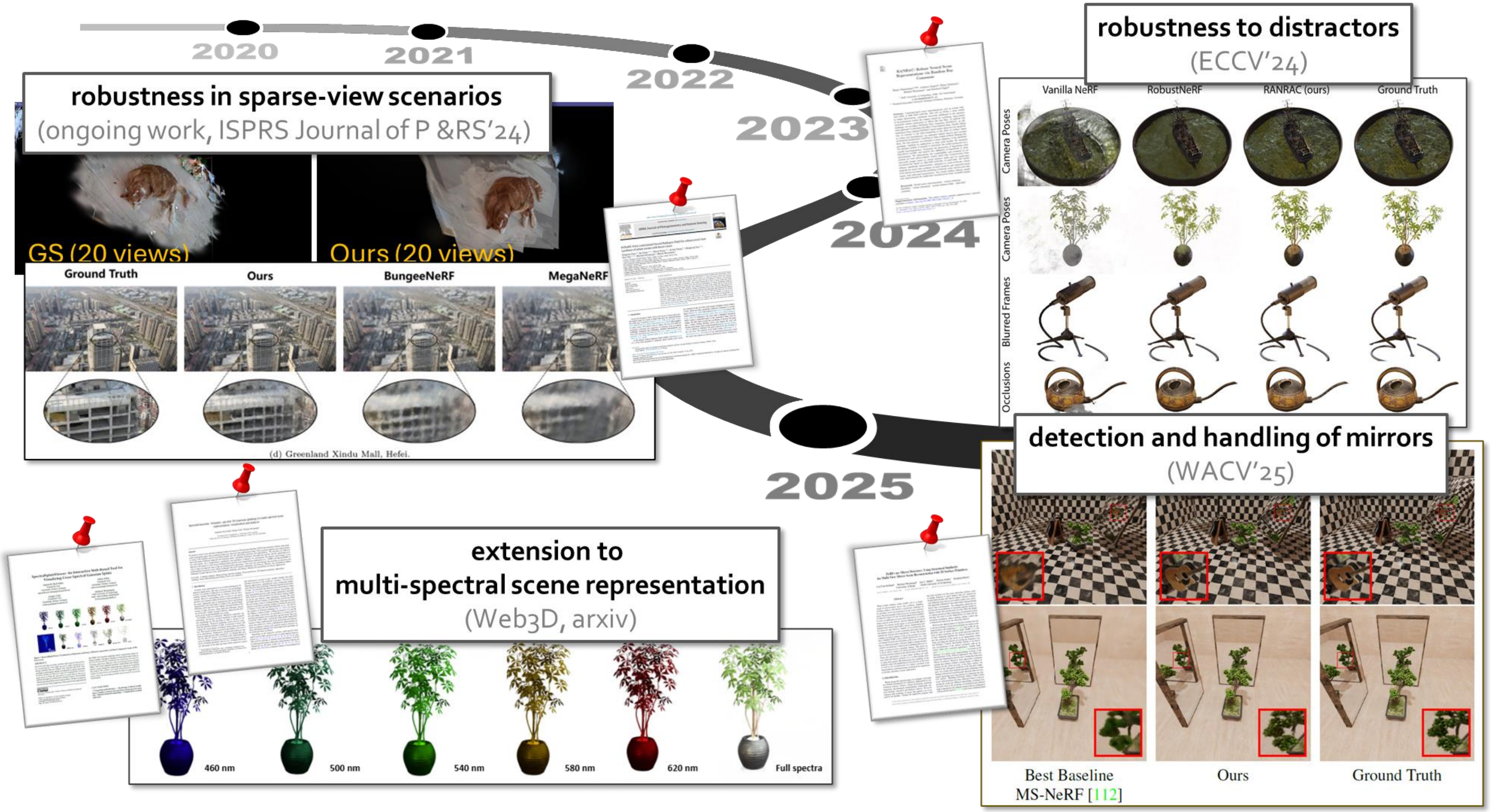
- Semantic scene stylization



S. N. Sinha, H. Graf, M. Weinmann. SemanticSplatStylization: Semantic scene stylization based on 3D Gaussian splatting and class-based style transfer, GCH 2024

Summary

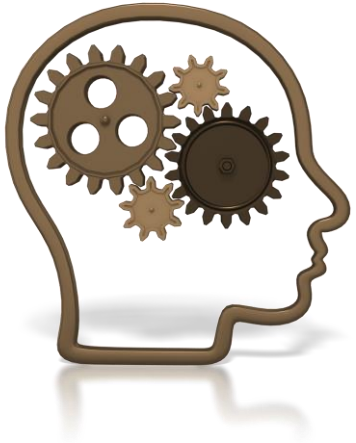
Robust extensions of NeRFs and Gaussian Splatting



Conclusions

AI …

- … offers breakthrough potential for numerous applications
- … leads to paradigm shift in capture technology
- … will make us re-think conventional processes
- … will help to tackle more challenging scenarios



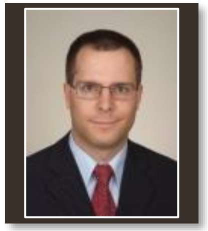
But …

- … cannot solve everything
- … may be outperformed in certain conditions
- … may benefit from coupling to traditional principles
- … we should not forget to use natural intelligence

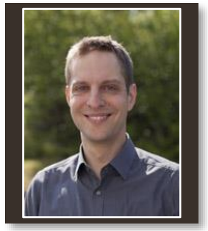
Acknowledgements

+ numerous more collaborators

Reinhard Klein



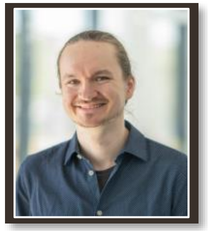
Martin Weinmann



Matthias B. Hullin

Ralf Sarlette

Julian Iseringhausen



Leif Van Holland

Andre Rochow

Roland Ruiters

Christopher Schwarz

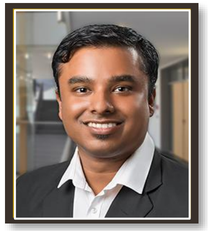
Patrick Stotko

Stefan Krumpen

Benno Buschmann

Jan Uwe Müller

Max Schwarz



Saptarshi Neil Sinha

Holger Graf

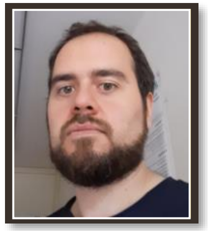
Nils Wandel

Sebastian Merzbach

Elena Trunz

Lukas Bode

Elmar Eisemann



Dennis Haitz

Sven Behnke

Thank you for your attention!

Michael Weinmann
TU Delft, Netherlands

m.weinmann@tudelft.nl