Reflections on Image-Based Rendering

Richard Szeliski

Google Research and The University of Washington

Indian Conference on Vision, Graphics, and Image Processing

December 16th, 2023

[30 years of] Reflections on Image-Based Rendering

Richard Szeliski
Google Research and The University of Washington

Indian Conference on Vision, Graphics, and Image Processing

December 16th, 2023

New edition of my book

Computer Vision: Algorithms and Applications, 2nd ed.

© 2022 Richard Szeliski, The University of Washington

https://szeliski.org/Book

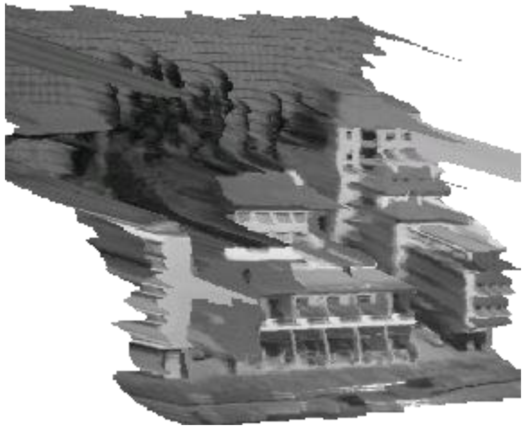
New edition of my book

View Interpolation

- Given two images with correspondences, *morph* (warp and cross-dissolve) between them [Chen & Williams, SIGGRAPH'93]

input

depth image

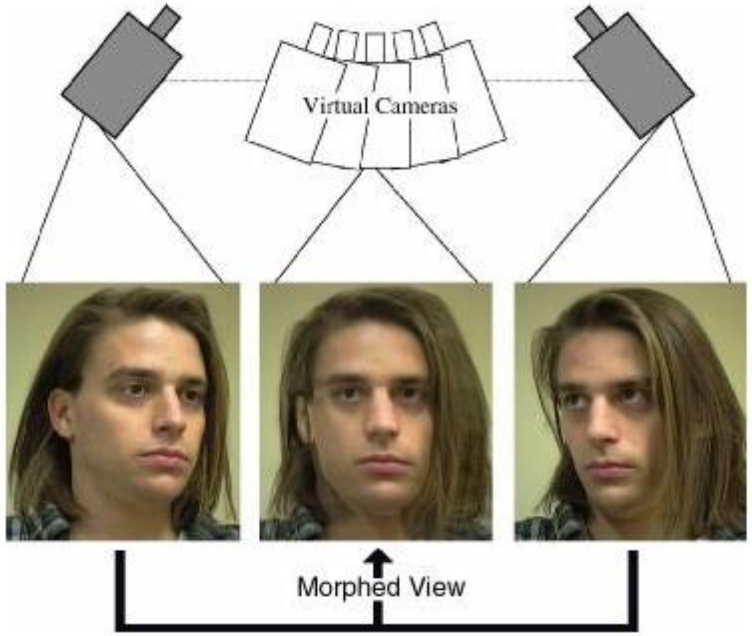


novel view

[Matthies,Szeliski,Kanade'88]

View Morphing

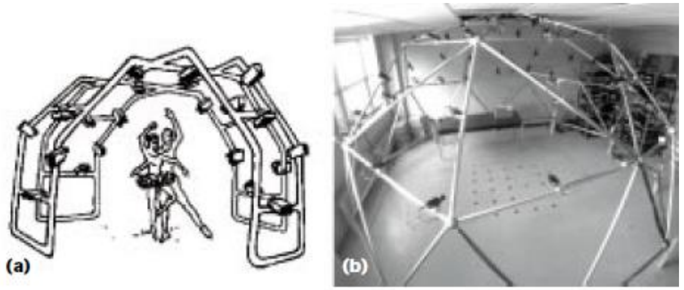
- Morph between pair of images using epipolar geometry [Seitz & Dyer, SIGGRAPH'96]



Video view interpolation (later)

Interactive 3D video scenarios

- Sports events, e.g., CMU's 30-camera "EyeVision" system at SuperBowl XXXV) and 2016
- Concert performances, plays, circus acts
- Games
- Instructional video, e.g., golf, skating, martial arts
- Interactive (Internet) video



Takeo Kanade and Peter Rander
Carnegie Mellon University

P.J. Narayanan
Centre for Artificial Intelligence and Robotics

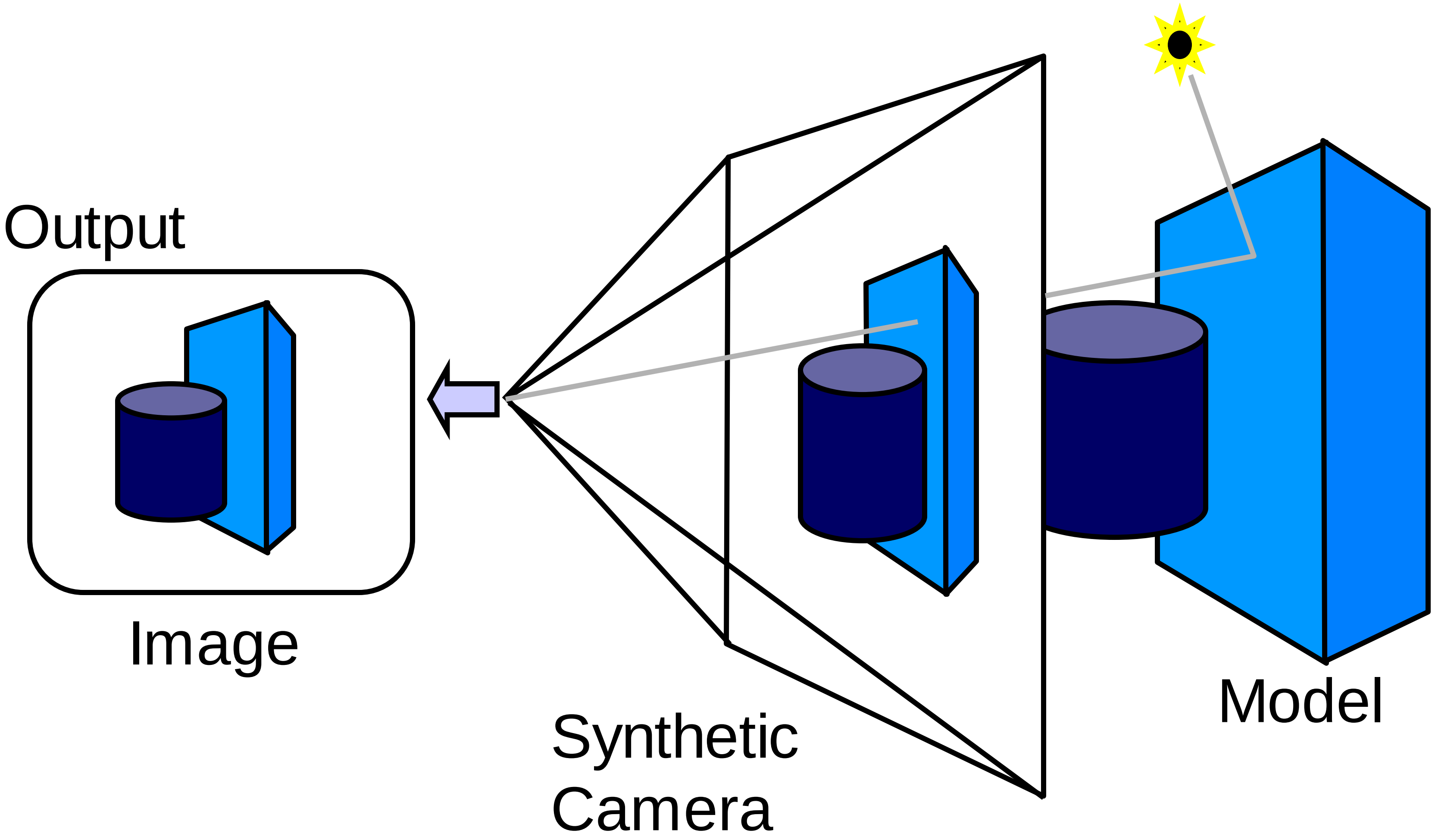
Figure 1. The Virtualized Reality studio: (a) conceptual; (b) 3D Dome.

Outline

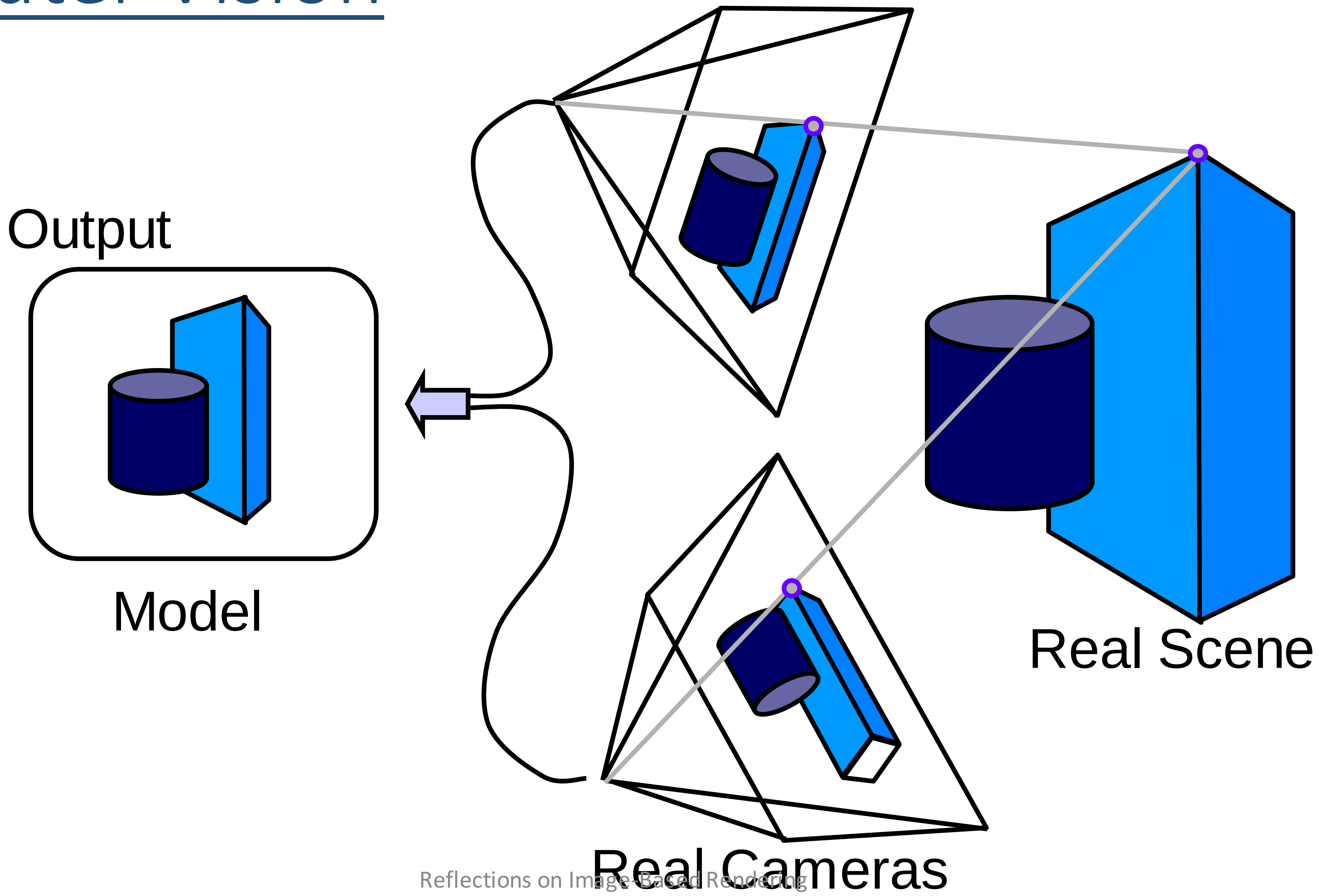
- Image-Based Rendering representations
  - Lumigraphs, Light Fields, Sprites with Depth, and Layers
- Virtual Viewpoint Video
- 360° and 3D Video
- 3D Photos
- Reflections and transparency
- Neural rendering

Image-Based Rendering

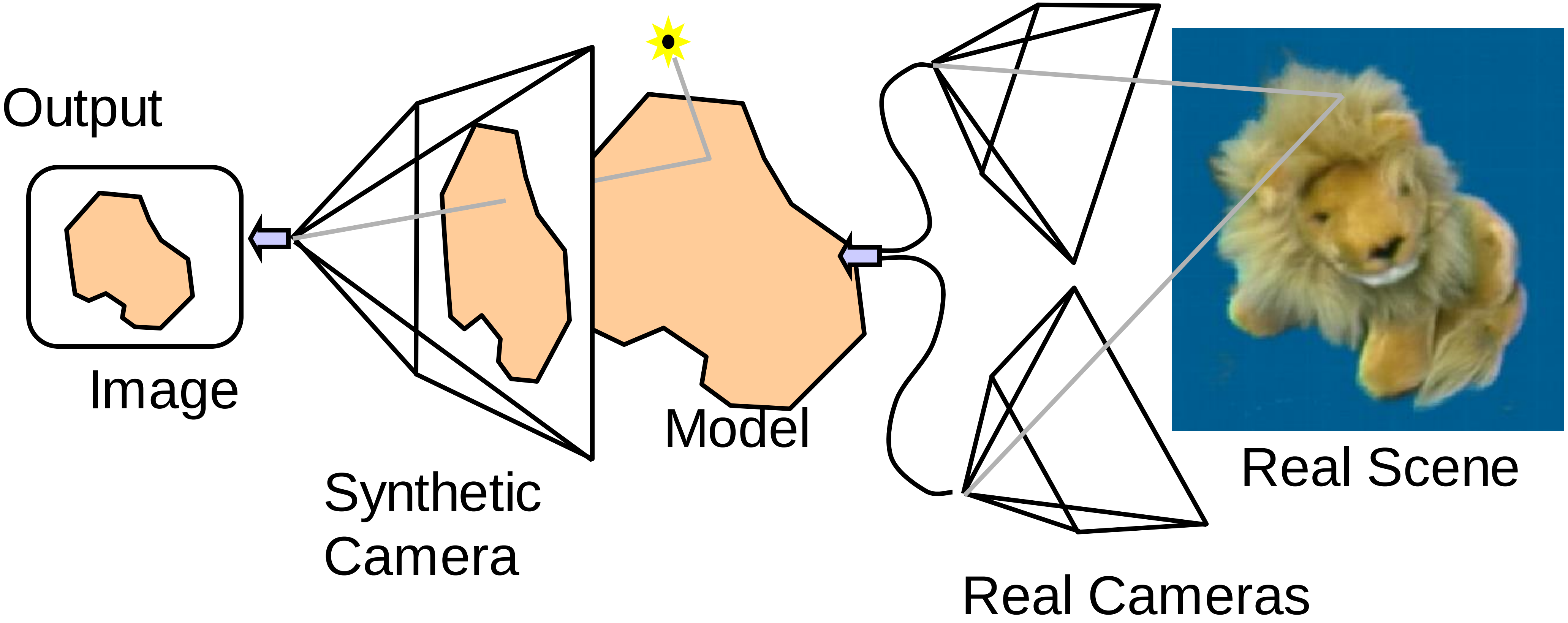
Computer Graphics



Computer Vision



But, vision technology fails



…and so does graphics

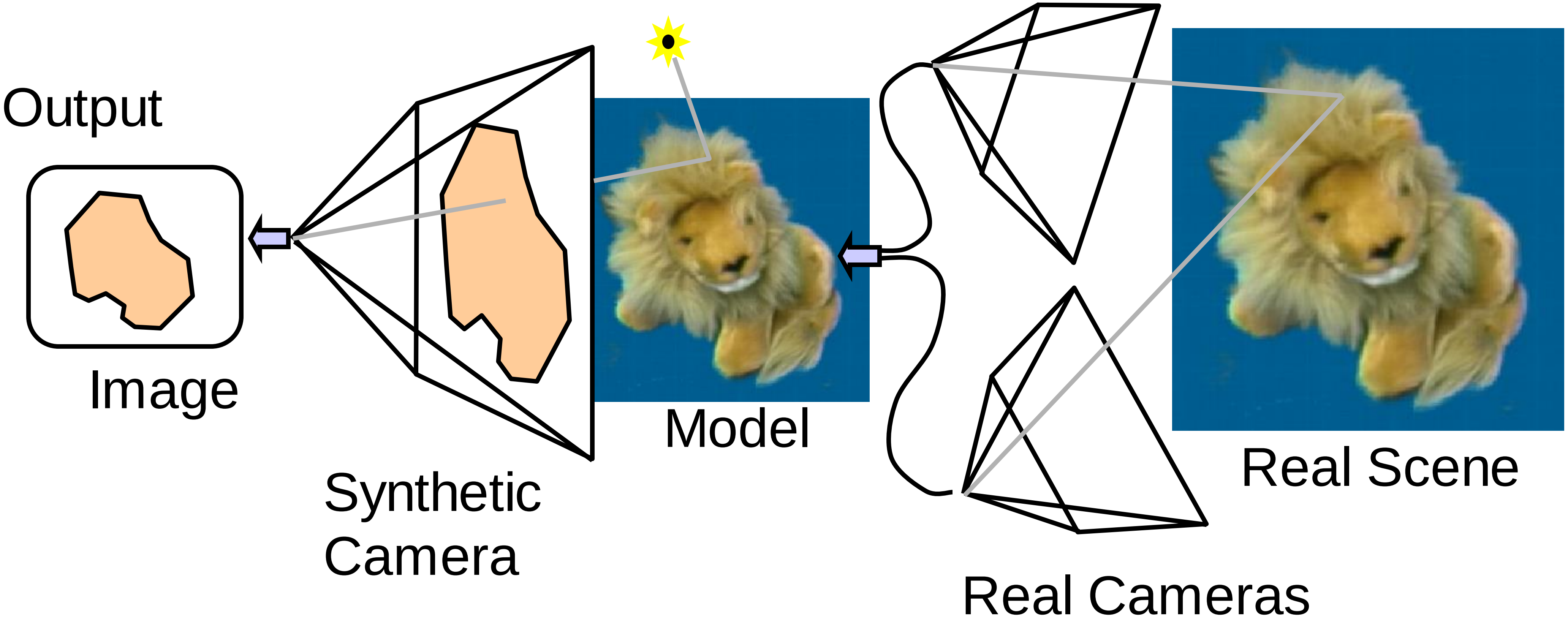
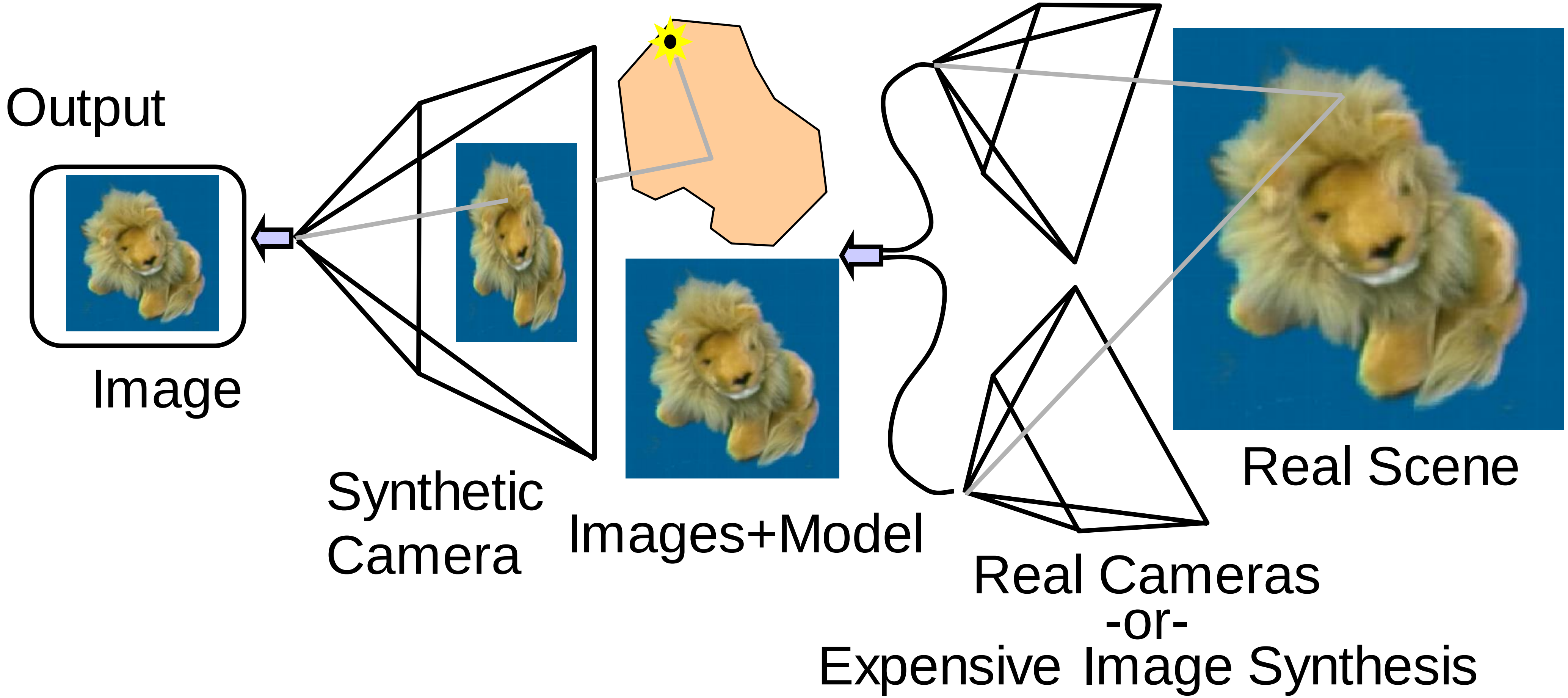
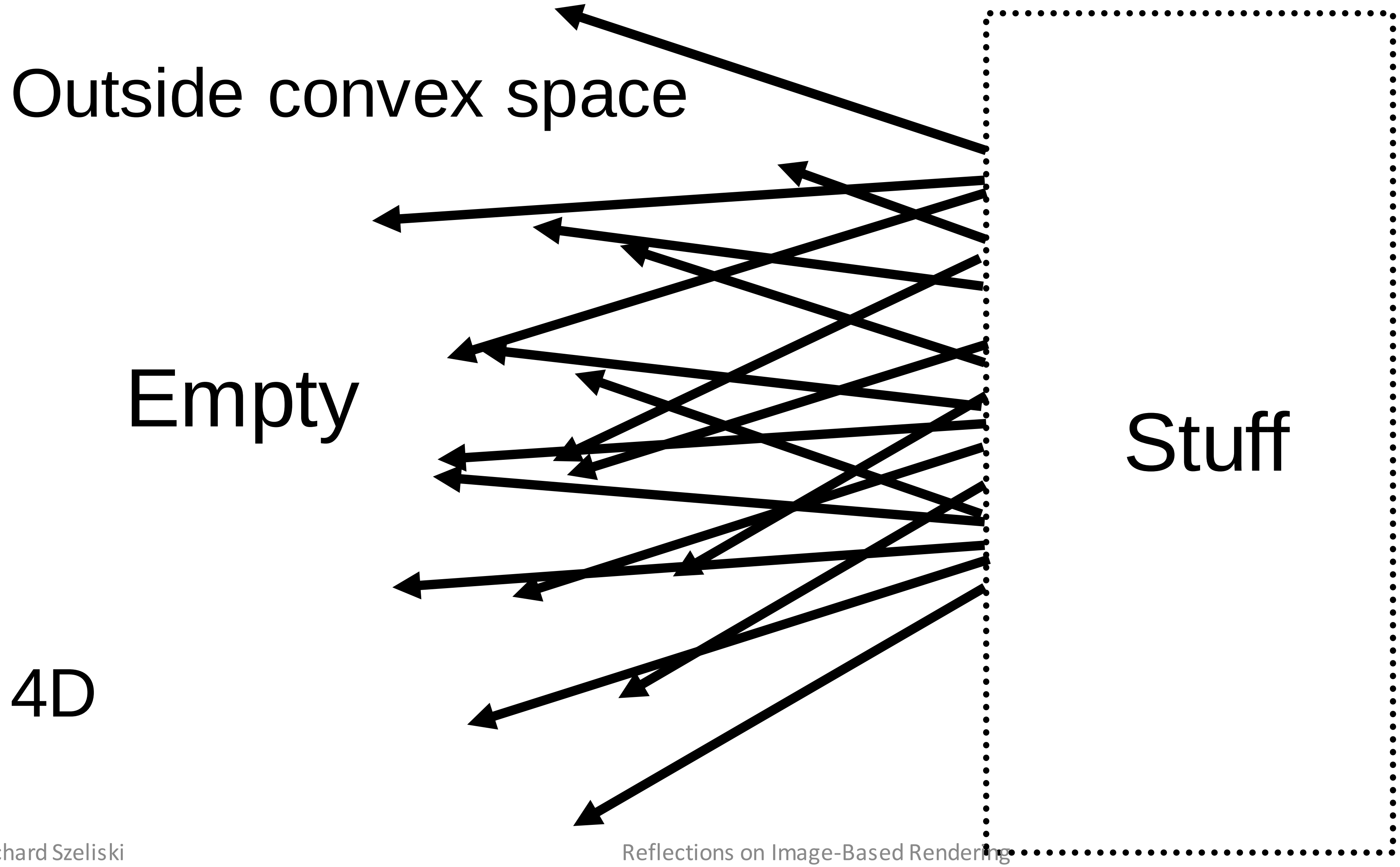


Image-Based Rendering

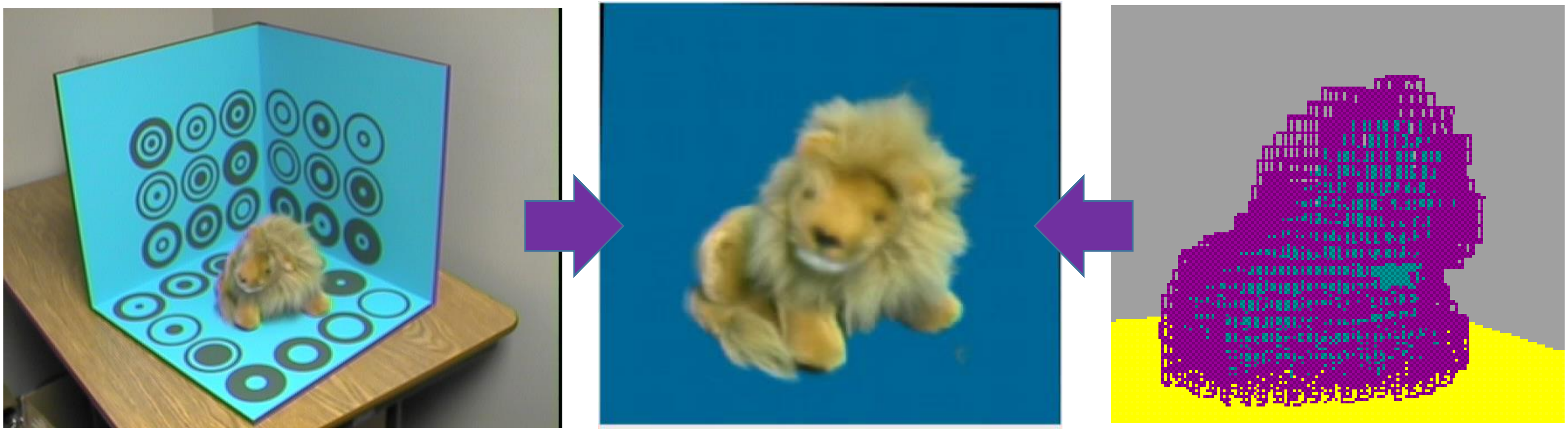


Lumigraph / Light Field [1996]



Lumigraph – Capture

- Convert images into a solid 3D model

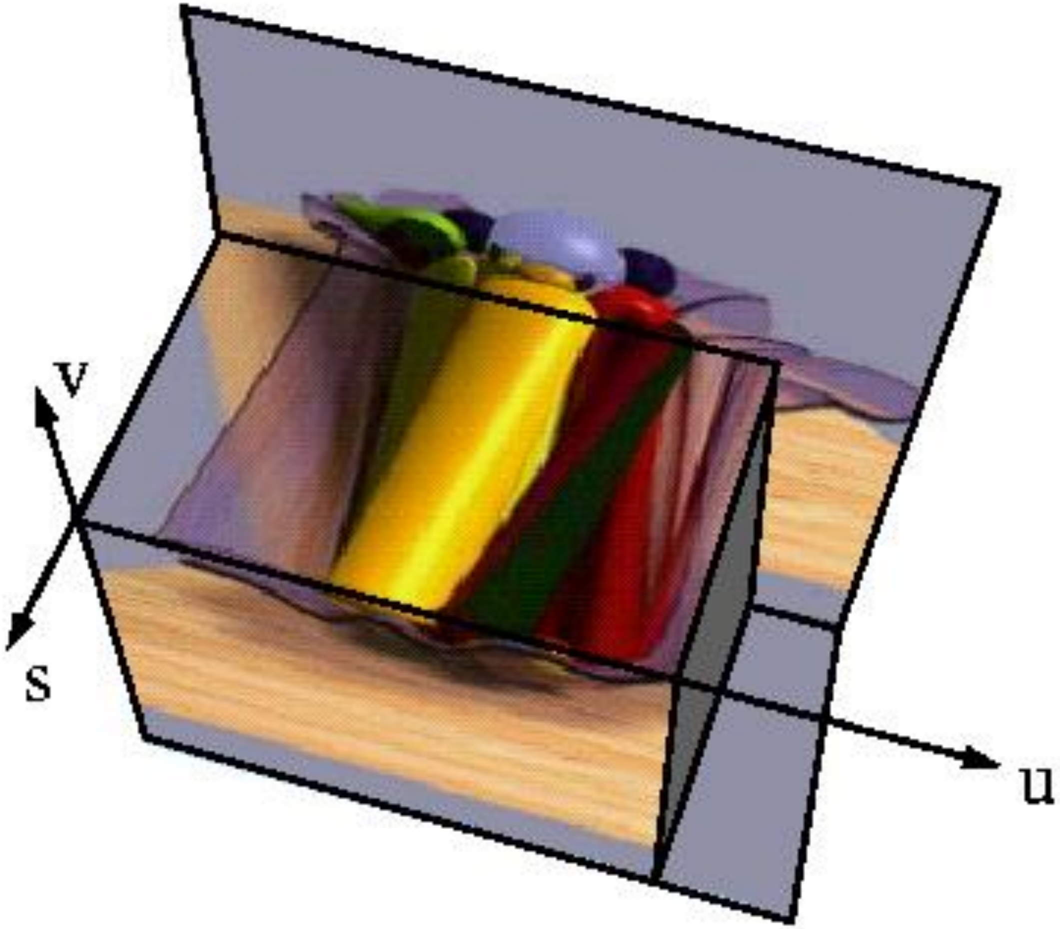


- Render from images and model

Lumigraph – Image Effects

Can model effects such as:

- parallax
- occlusion
- translucency
- refraction
- highlights
- reflections



Unstructured Lumigraph

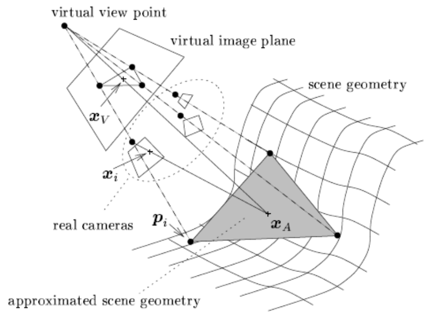
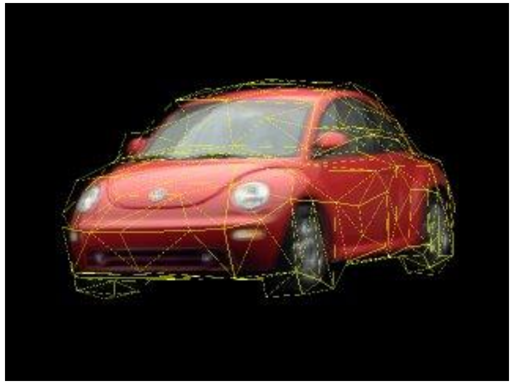
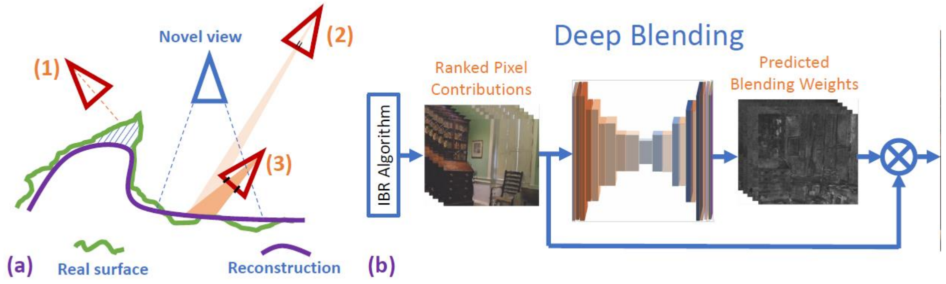


Figure3. Drawing triangles of neighboring projected camera centers and approximating scene geometry by one plane for the whole scene, for one camera triple or by several planes for one camera triple.

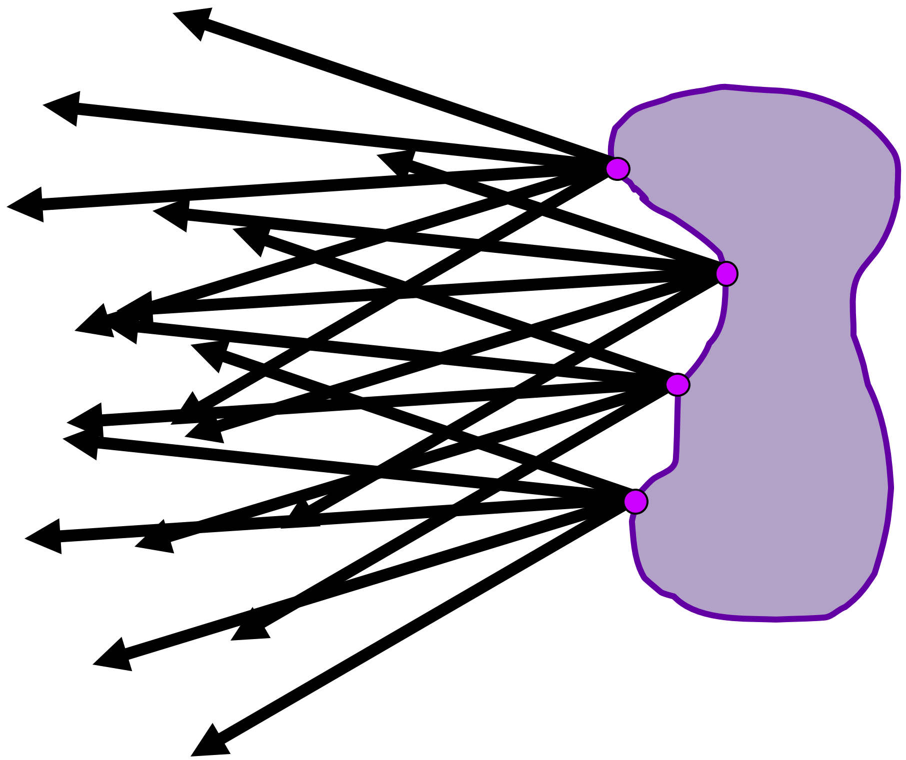
- What if the images aren't sampled on a regular 2D grid?
- Can still re-sample rays
- Ray weighting becomes more complex [Heigl *et al.*,DAGM'99]
- Unstructured Lumigraph [Buehler *et al.*, SIGGRAPH'2000]
- Deep blending [Hedman *et al.*, SG Asia 2018]
- FVS [Riegler & Koltun, ECCV'2020]





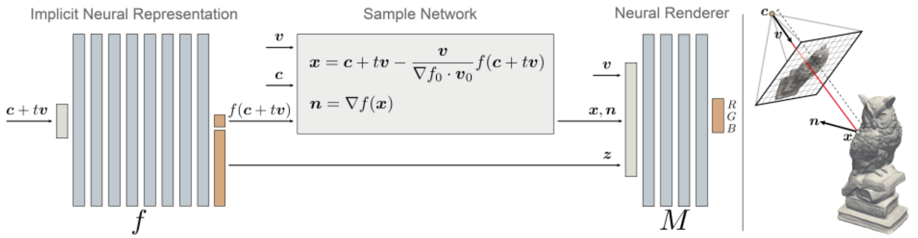
Surface Light Fields

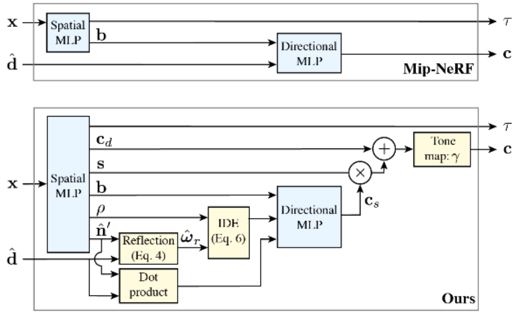
- [Wood et al, SIGGRAPH 2000]
- Turn 4D parameterization around:
  - image @ every surface pt.
- Leverage coherence:
  - compress radiance fn (BRDF * illumination) after rotation by ***n***



Surface Light Fields

- [Wood et al, SIGGRAPH 2000]
- …
- Implicit Differentiable Renderer [Yariv *et al.*, NeurIPS 2020]
- Stable View Synthesis [Riegler and Koltun, 2021]
- Ref-NeRF [Verbin *et al.*, CVPR 2022]





Environment Matting [2000]

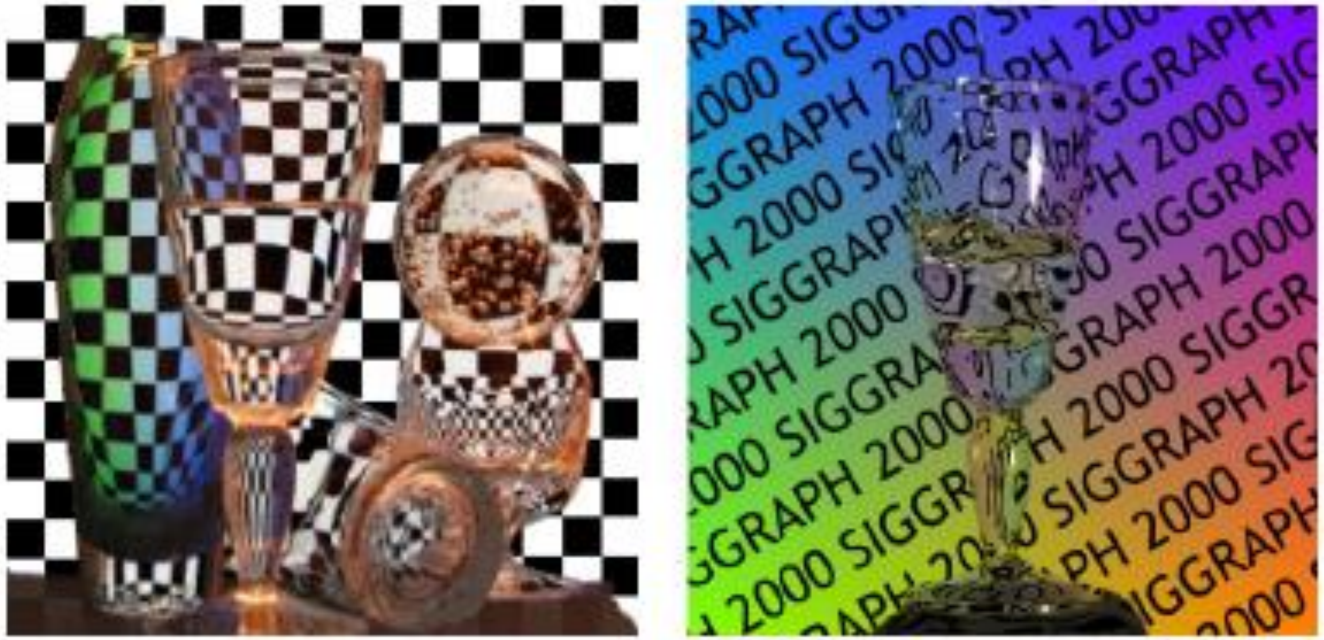
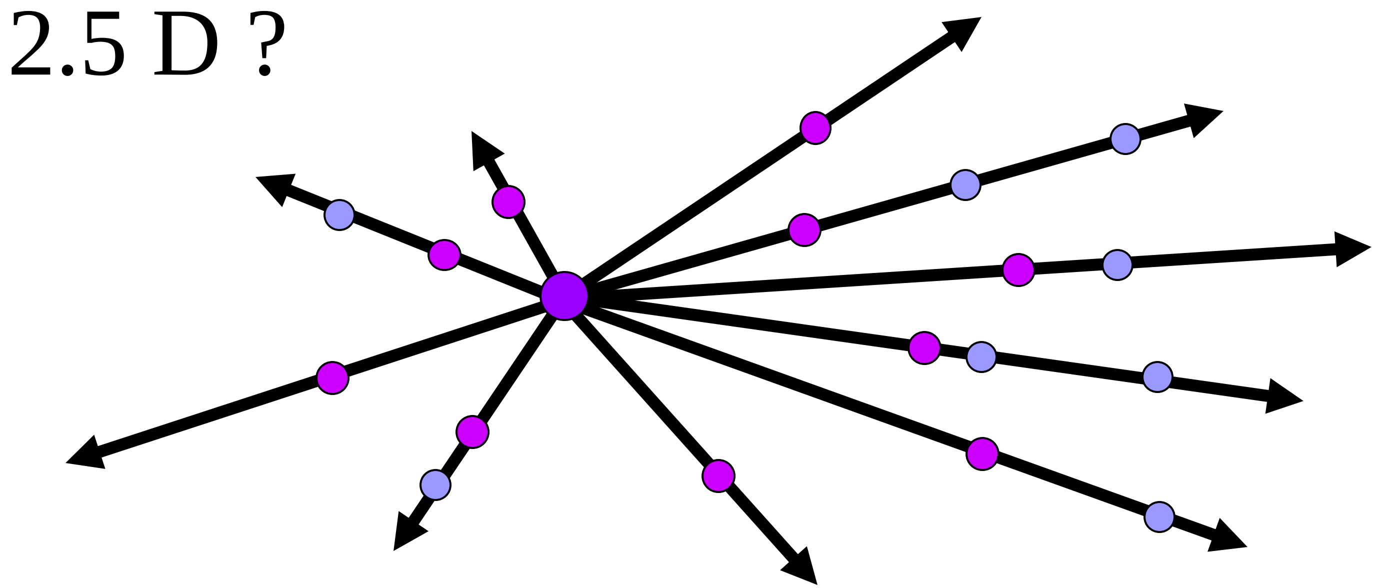


Figure 1 Sample composite images constructed with the techniques of this paper: slow but accurate on the left, and a more restricted example acquired at video rates on the right.

… NeRV: Neural Reflectance and Visibility Fields for Relighting … [Srinivasan *et al.*, CVPR’21]

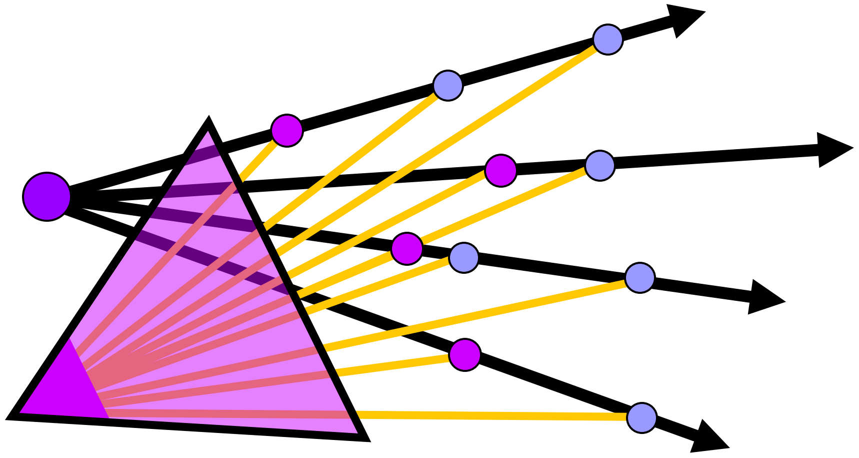
Layered Depth Image



Layered Depth Image

Layered Depth Image

- Rendering from LDI
  [Shade et al., SIGGRAPH'98]



- Incremental in LDI X and Y
- Guaranteed to be in back-to-front order

Sprites with Depth

- Represent scene as collection of cutouts with depth (planes + parallax)
- Render back to front with fwd/inverse warping [Shade *et al.*, SIGGRAPH'98]
- Basis of Virtual Viewpoint Video [Zitnick *et al.* 2004]
- …
- Immersive LFV .. Layered Mesh [Broxton .. SG'20]
- GeLaTO [Martin-Brualla *et al.*, ECCV'20]

Multiplane images

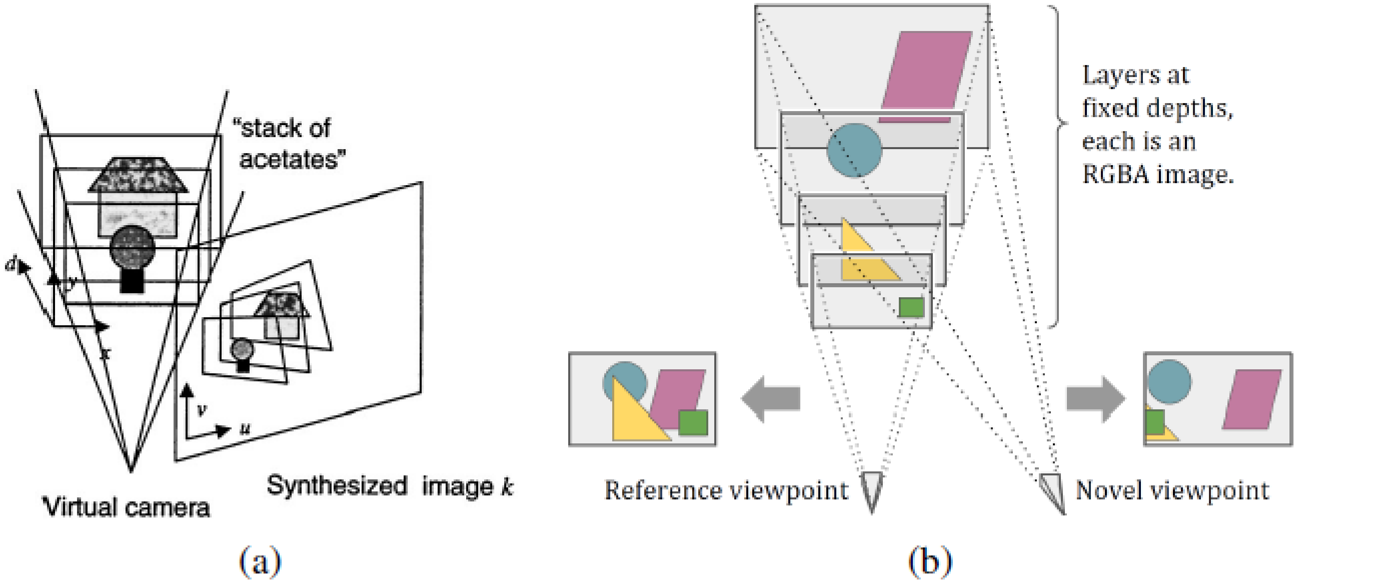
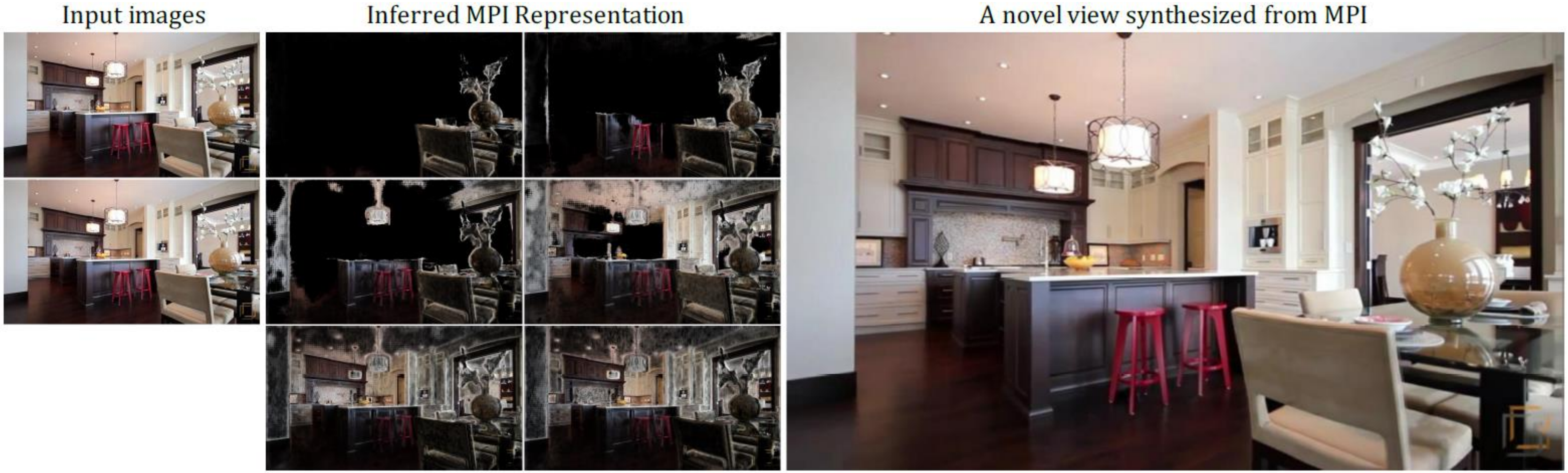


Figure 14.7 *Finely sliced fronto-parallel layers: (a) stack of acetates (Szeliski and Golland 1999) © 1999 Springer and (b) multiplane images (Zhou, Tucker, Flynn* et al. *2018) © 2018 ACM.*

Multiplane images



Stereo Magnification… [Zhou *et al.*, SIGGRAPH 2018]

Multi-sphere and layered meshes

Immersive Light Field Video with a Layered Mesh Representation

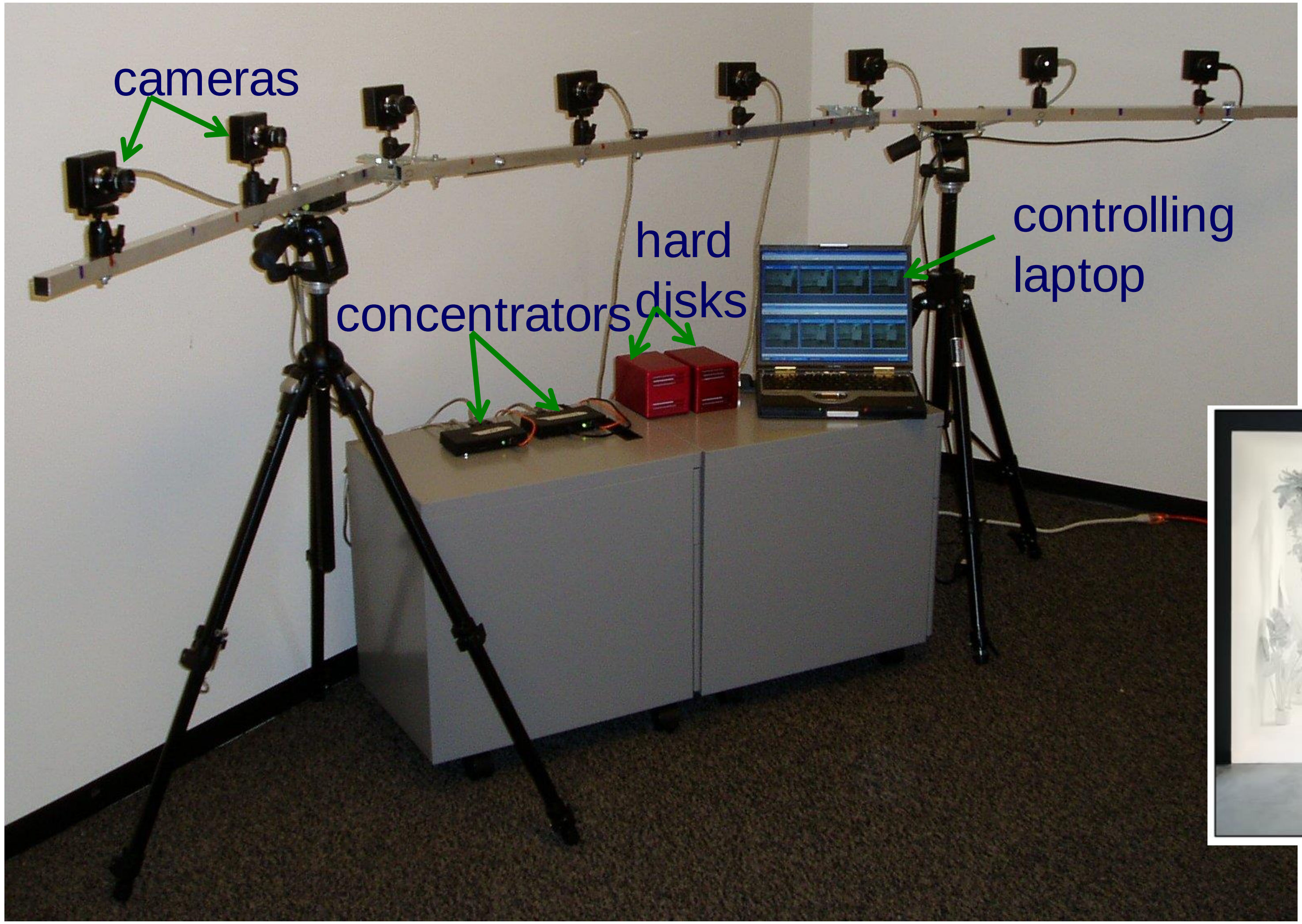
MICHAEL BROXTON*, JOHN FLYNN*, RYAN OVERBECK*, DANIEL ERICKSON*, PETER HEDMAN, MATTHEW DUVALL, JASON DOURGARIAN, JAY BUSCH, MATT WHALEN, and PAUL DEBEVEC, Google

(a) Capture Rig (b) Multi-Sphere Image (c) Layered Mesh Representation

[SIGGRAPH'2020]

Virtual Viewpoint Video

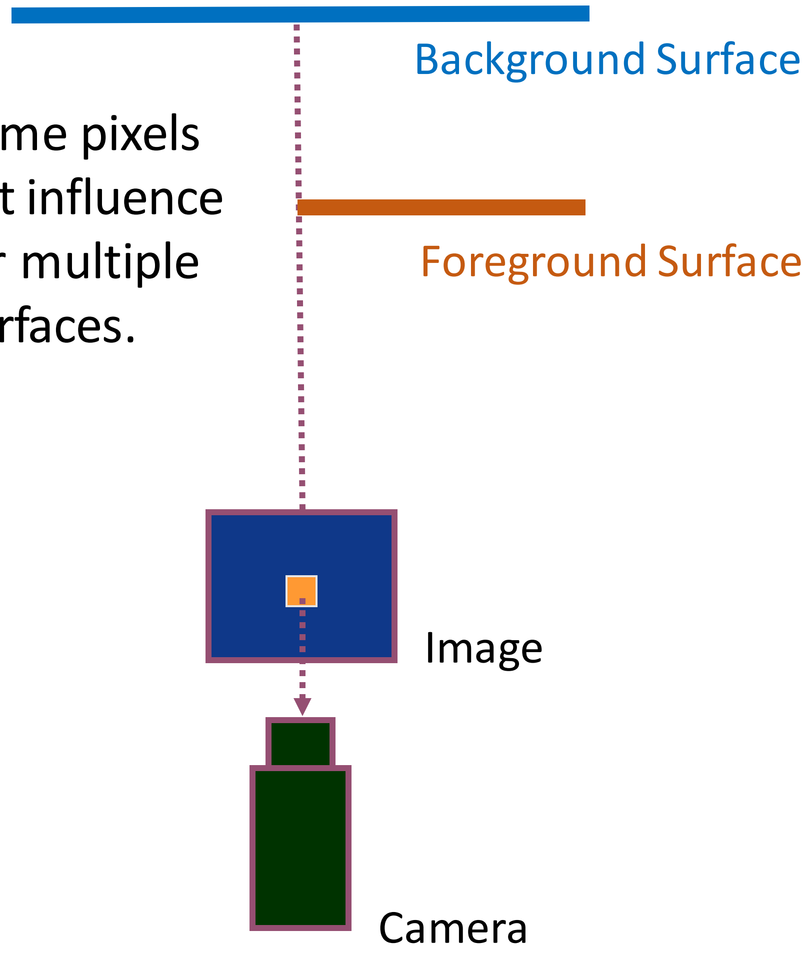
Virtual Viewpoint Video [SIGGRAPH 2004]



[Broxton SG'20]

Matting

Some pixels get influence for multiple surfaces.



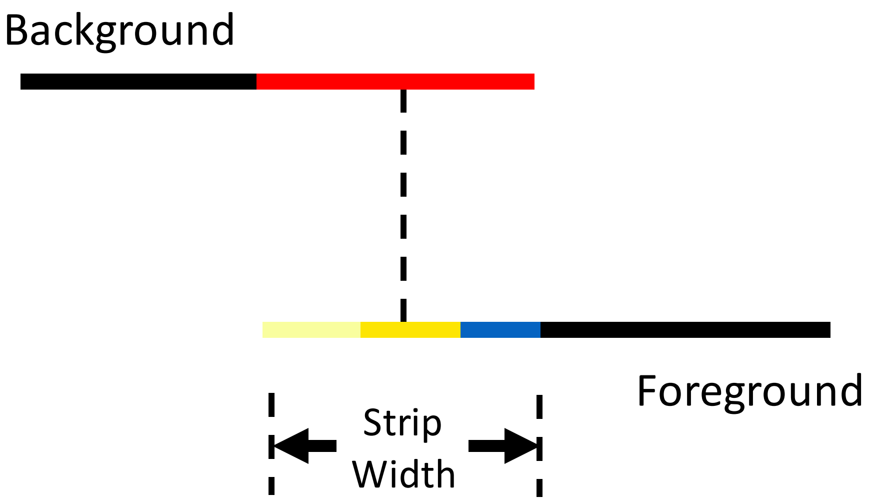
Close up of real image:

Multiple colors and depths at boundary pixels…

Find matting information:

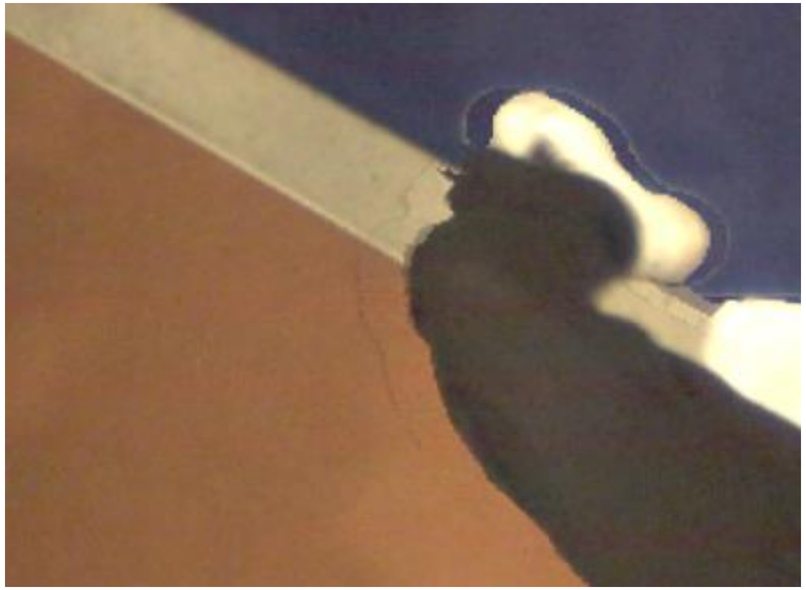
1. Find boundary strips using depth.

2. Within boundary strips compute the colors and depths of the foreground and background object.



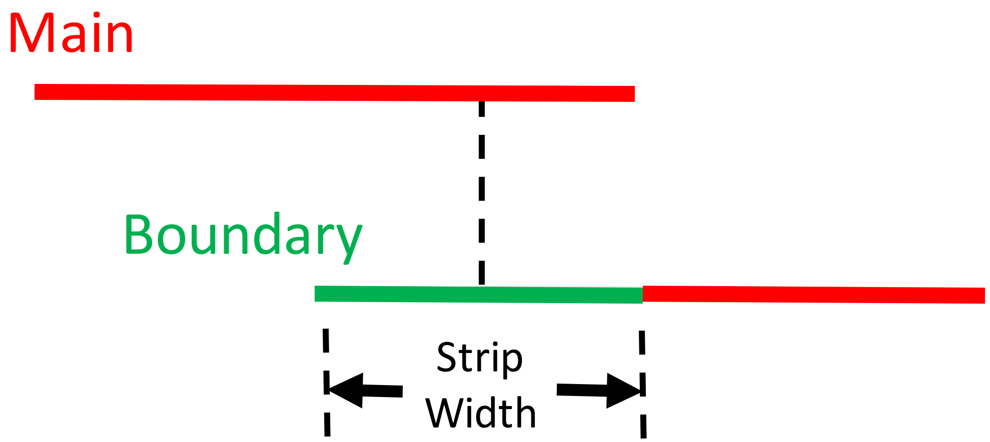
Why matting is important

No Matting



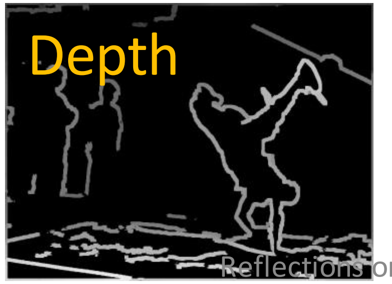
Matting

Virtual Viewpoint Video



Two-layer model with
thin boundary strips
[Zitnick *et al.*, SIGGRAPH'04]

Main Layer: Boundary Layer:

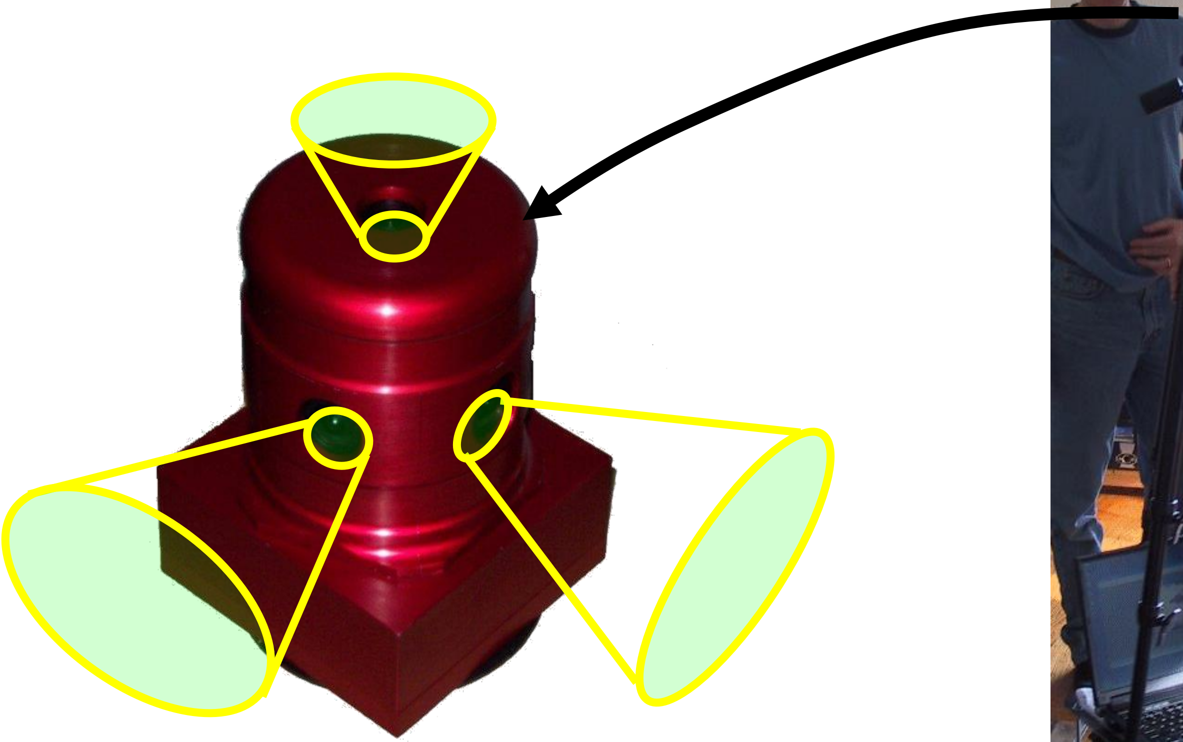



Massive Arabesque

360° Video

360 Video

[Uyttendaele et al. 2004]



Ladybug (six-camera head)

Acquisition platforms (today)

360 Video


ULTRACAM
mobile
K 980 CW

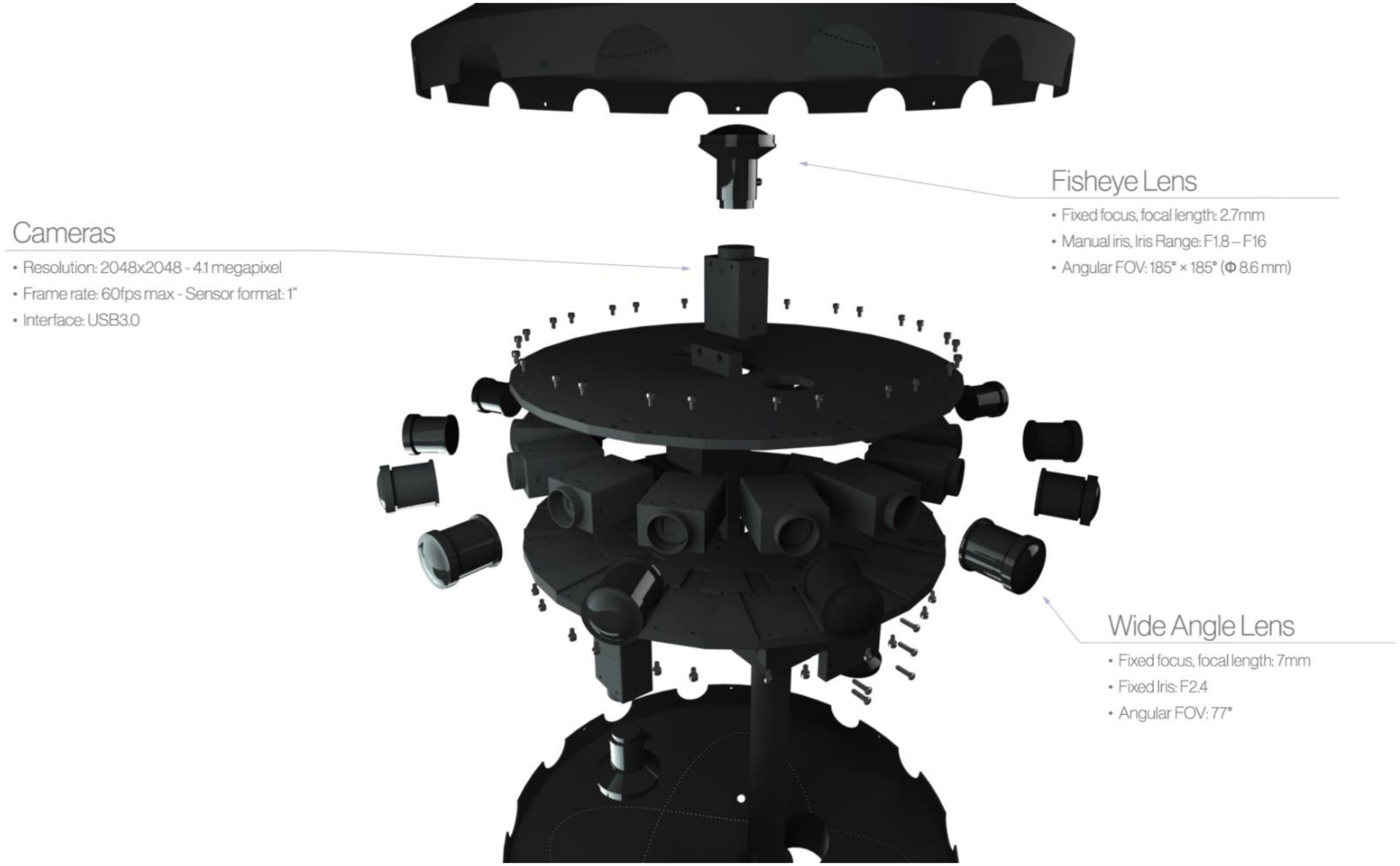
ULTRACAM
mobile

K 980 CW
www.europcar.at

360 Video

Google Jump [2015]

Facebook Surround 360 [2016]



Facebook Surround 360 [2017]

Facebook's new Surround 360 video cameras let you move around inside live-action scenes

The freedom of VR with the fidelity of real life

By Nick Statt | @nickstatt | Apr 19, 2017, 1:15pm EDT

Facebook today announced the second generation of its Surround 360 video camera design, and this time the company is serious about helping potential customers purchase it as an actual product. The Surround 360, which Facebook unveiled last year as an open-source spec guide for others to build off of, has been upgraded as both a larger, more capable unit and a smaller, more portable version.

An Integrated 6DoF Video Camera and System Design

ALBERT PARRA POZO, MICHAEL TOKSVIG, TERRY FILIBA SCHRAGER, and JOYCE HSU, Facebook Inc.
UDAY MATHUR, RED Digital Cinema
ALEXANDER SORKINE-HORNUNG, RICK SZELISKI, and BRIAN CABRAL, Facebook Inc.

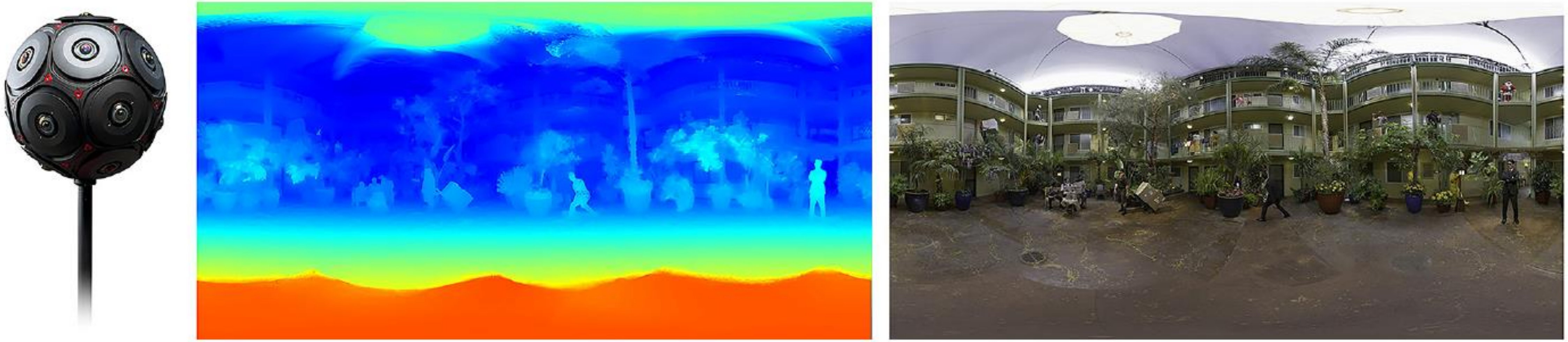
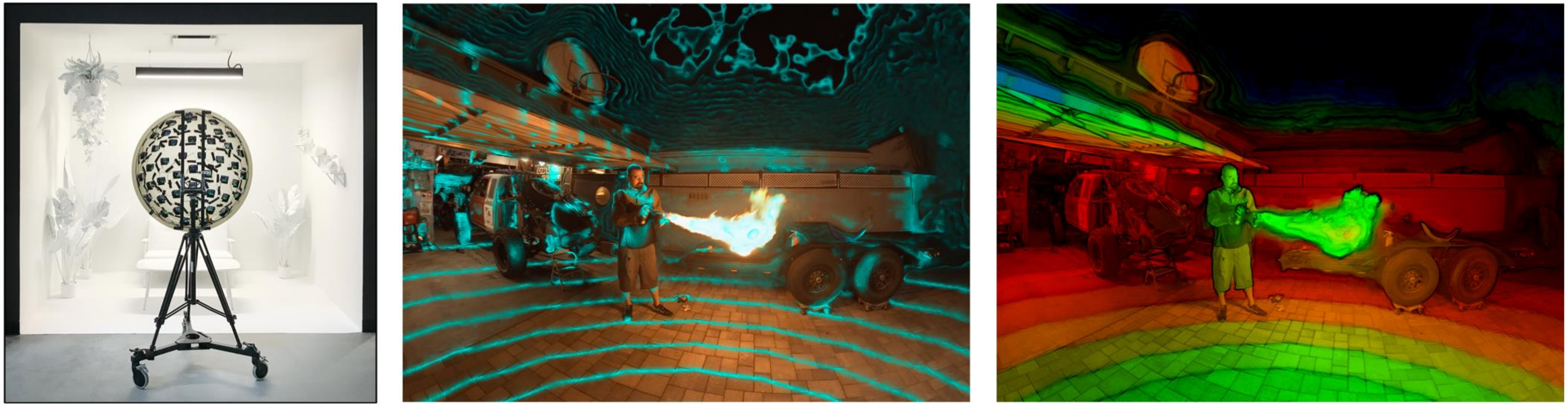


Fig. 1. The commercial 16 camera system, an equirectangular depth map, and final color rendering produced from our system.

[Video](#)

[SIGGRAPH Asia 2019]

Hemispherical light field capture & playback



(a) Capture Rig (b) Multi-Sphere Image (c) Layered Mesh Representation

IMMERSIVE LIGHT FIELD VIDEO
WITH A LAYERED MESH REPRESENTATION

SIGGRAPH 2020 Technical Paper

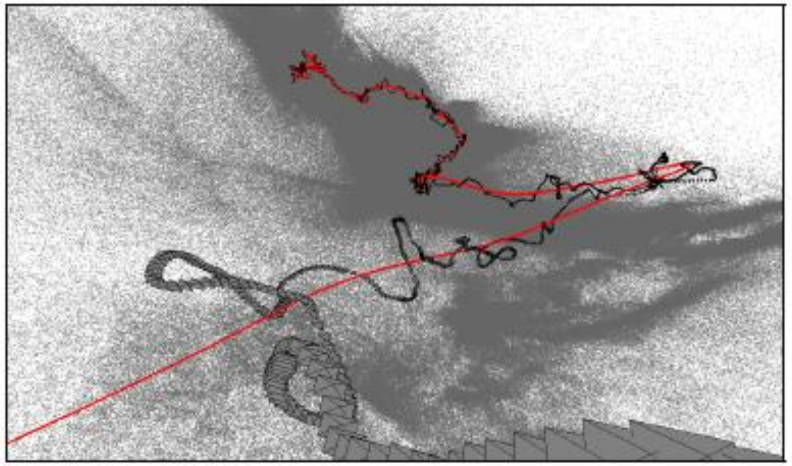
Download PDF

Michael Broxton*, John Flynn*, Ryan Overbeck*, Daniel Erickson*,
Peter Hedman, Matthew DuVall, Jason Dourgarian, Jay Busch, Matt Whalen,
Paul Debevec

Immersive Video Stabilization

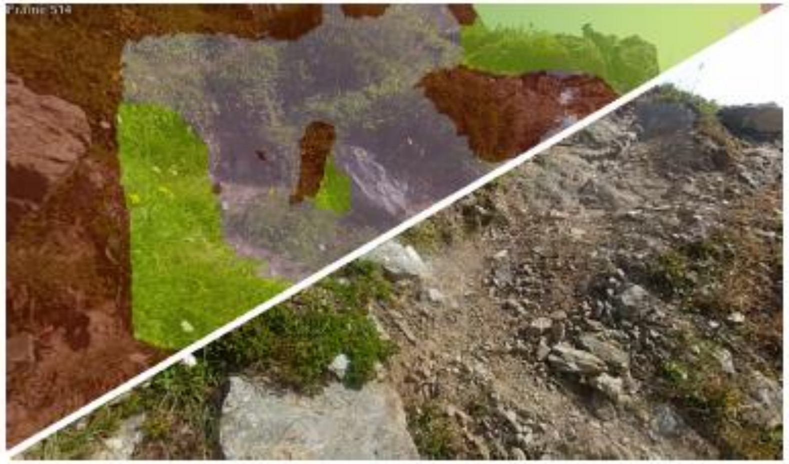
First-person Hyperlapse

Create buttery-smooth "fast forwards" from action videos



(a) Scene reconstruction

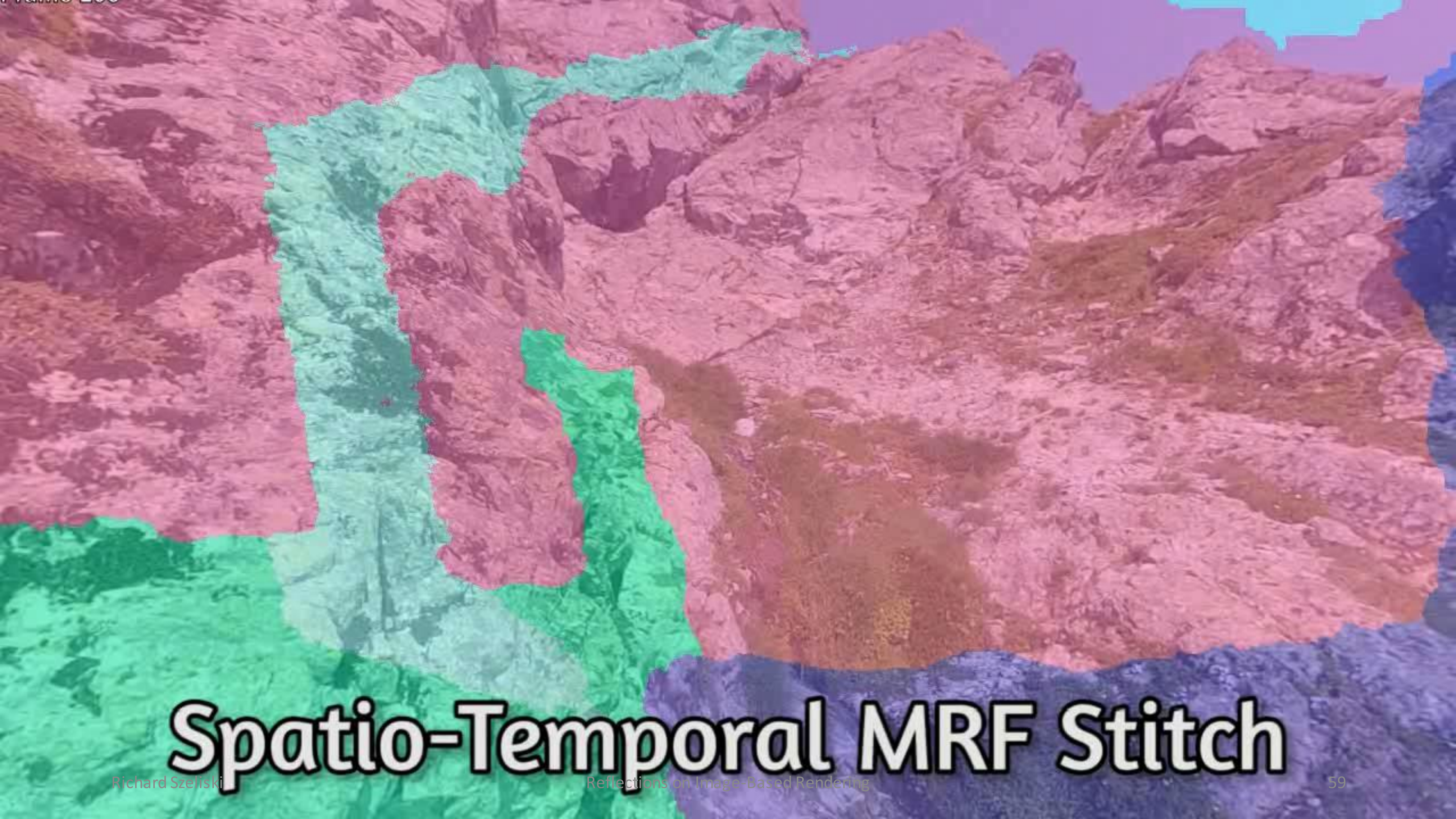
(b) Proxy geometry



(c) Stitched & blended

[Kopf, Cohen, Szeliski, SIGGRAPH 2014]

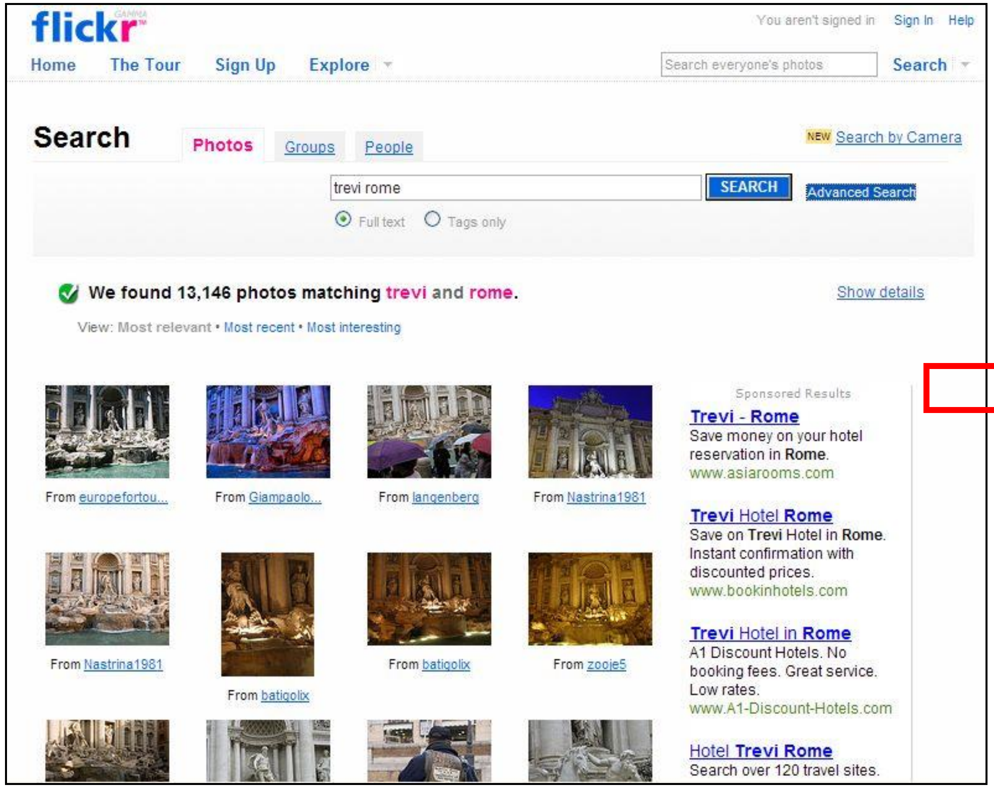

Proxy Geometry
(for a single video frame)

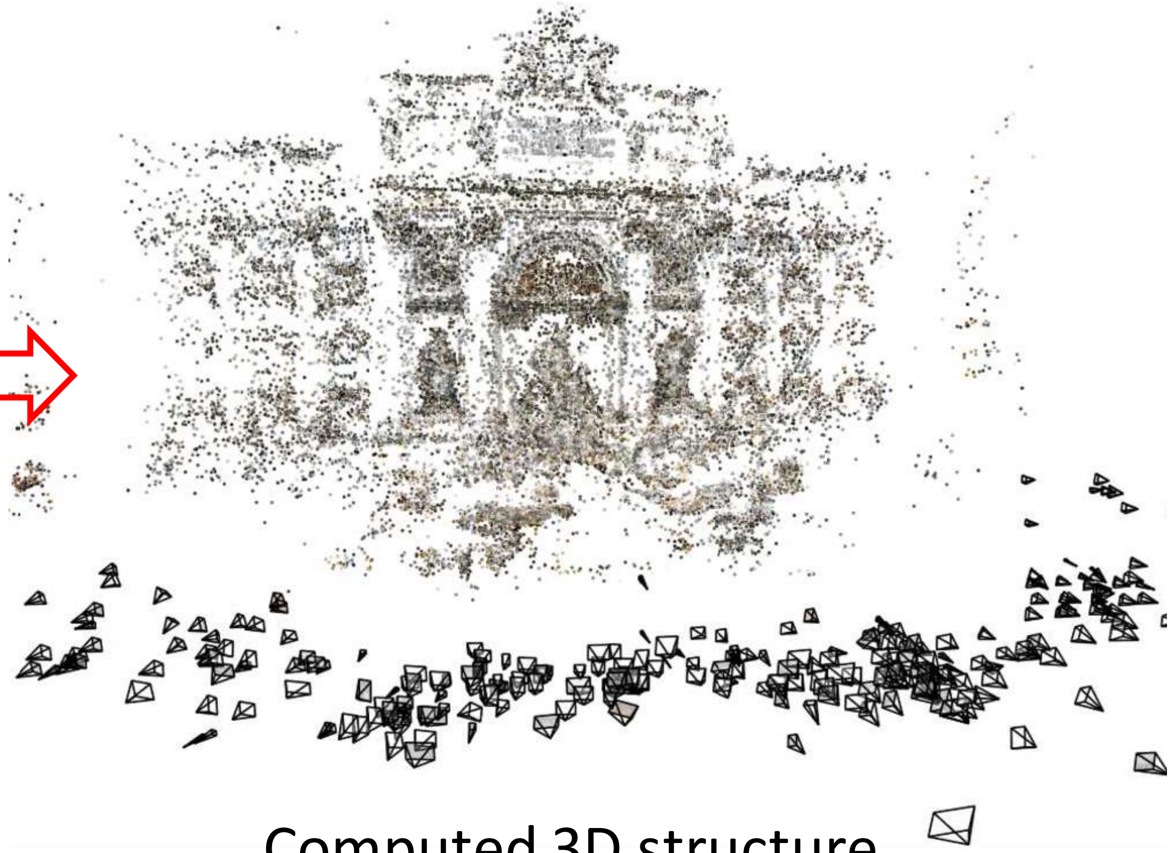

Spatio-Temporal MRF Stitch


Input Video

Large-Scale Reconstruction

Photo Tourism



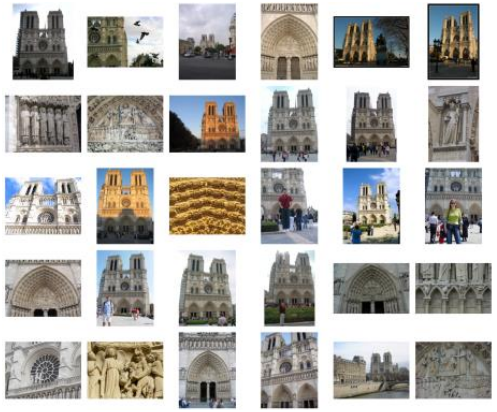


Internet images

Computed 3D structure

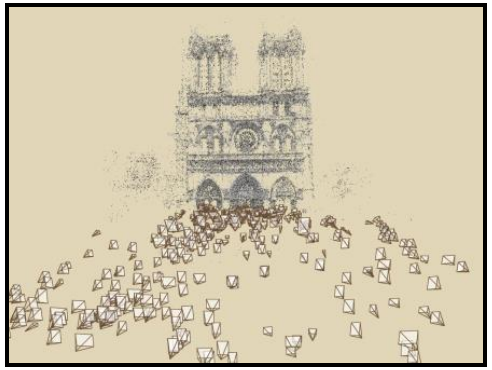
[Snavely, Seitz, Szeliski, SIGGRAPH 2006]

System overview



Input photographs

Scene reconstruction



Relative camera positions and orientations

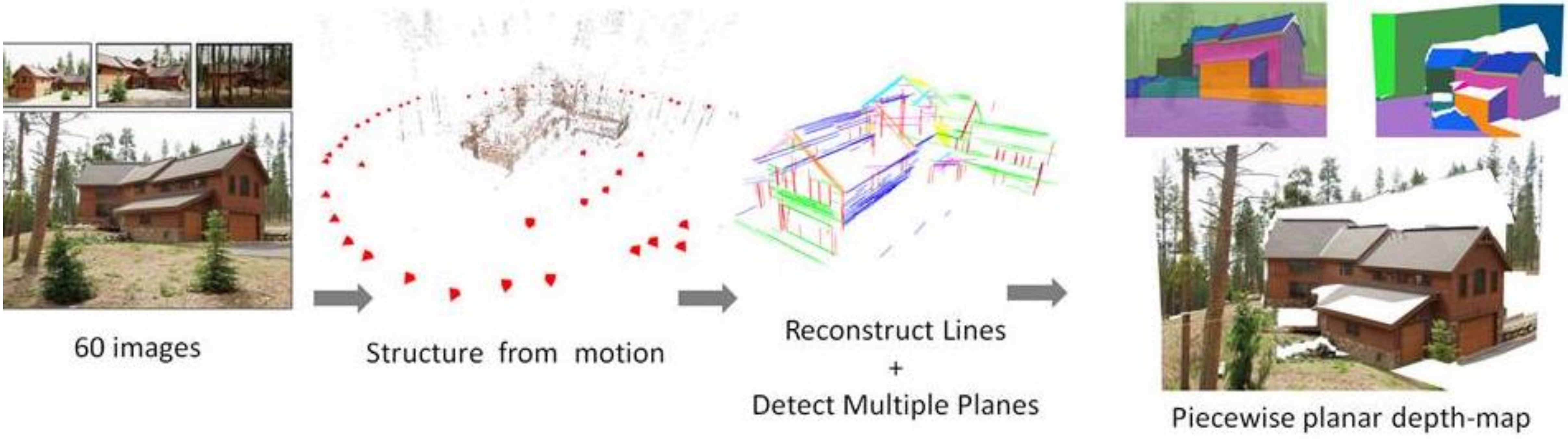
Point cloud

Sparse correspondence

Photo Explorer

Navigation: Prague Old Town Square

Piecewise planar proxies



[Sinha, Steedly, Szeliski ICCV'09]

Photo Tours - 2012

[Kushal *et al.*, 3DIMPVT 2012]

The Visual Turing Test - 2013

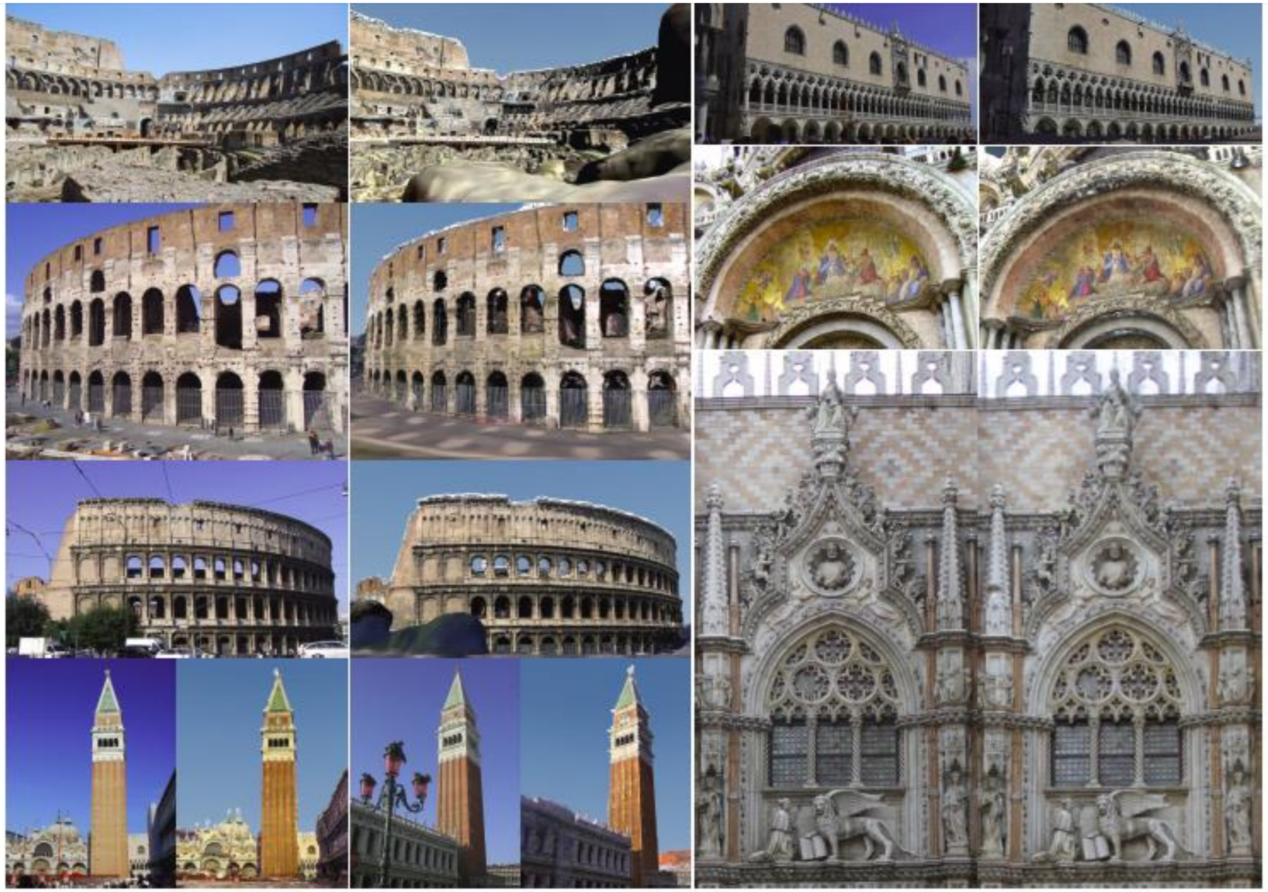
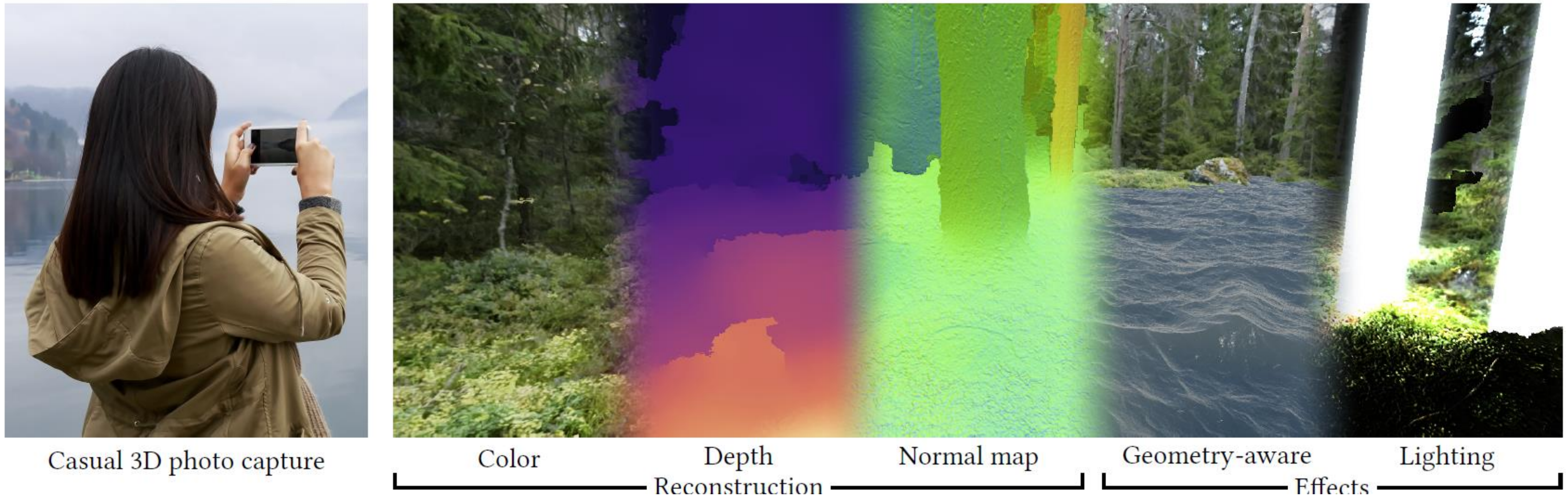


Figure 5. Visual Turing test. In each image pair, the ground truth image is on the left and our result is on the right.

[Shan *et al.*, 3DV 2013]



[Casual 3D Photography]

Peter Hedman, Suhib Alsisan, Richard Szeliski, Johannes Kopf

SIGGRAPH Asia 2017

Casual 3D Photography

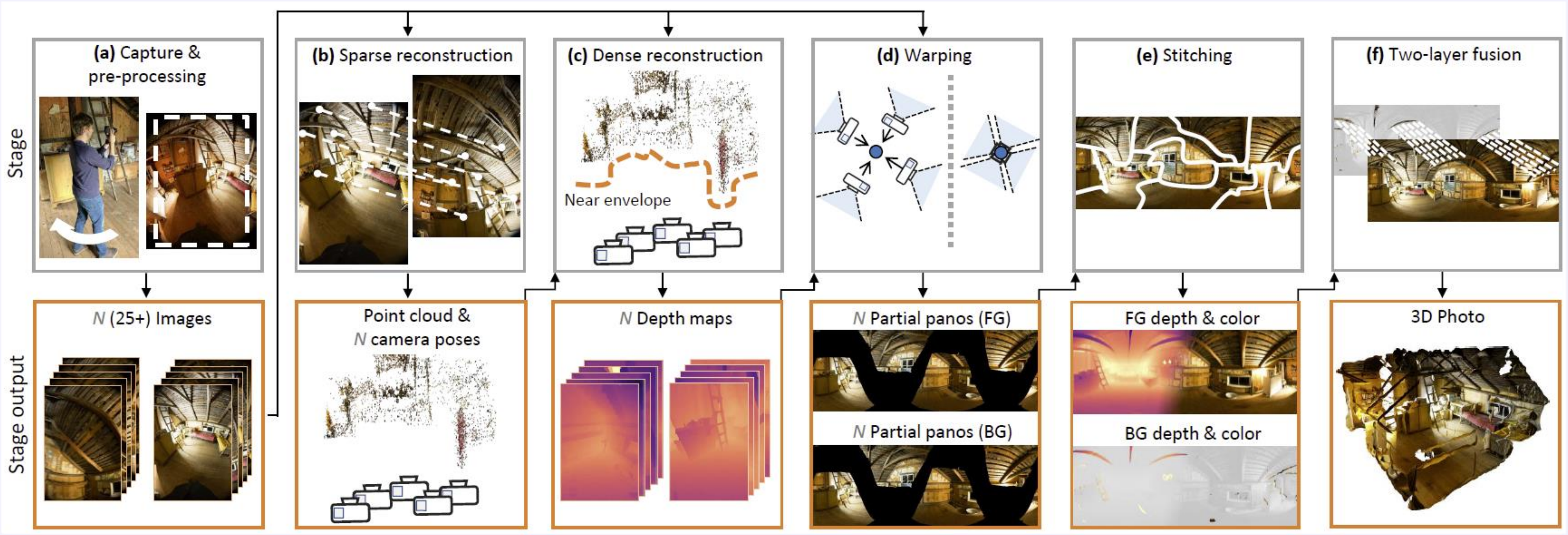
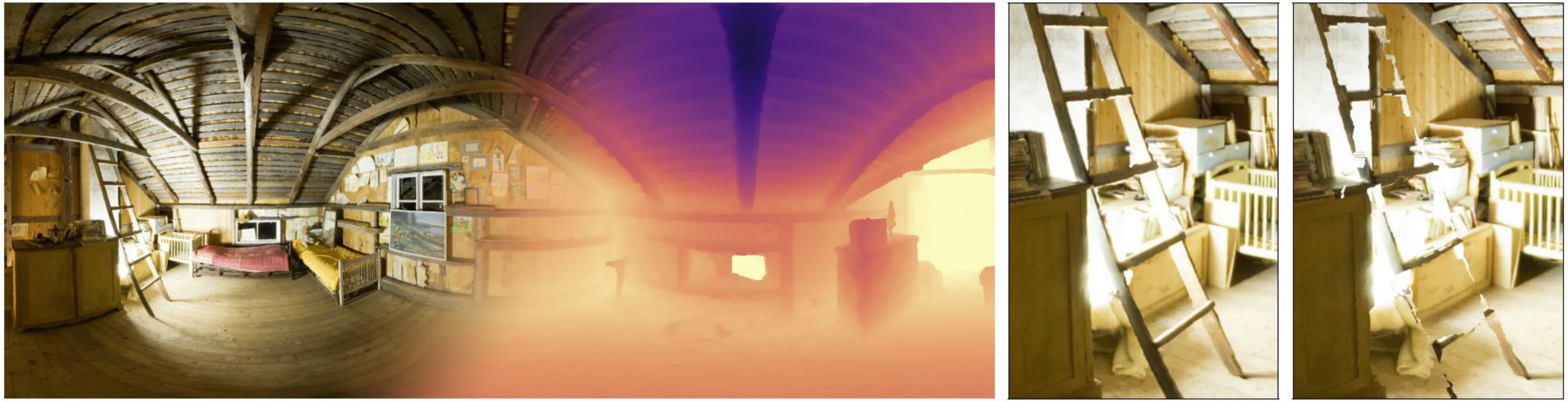


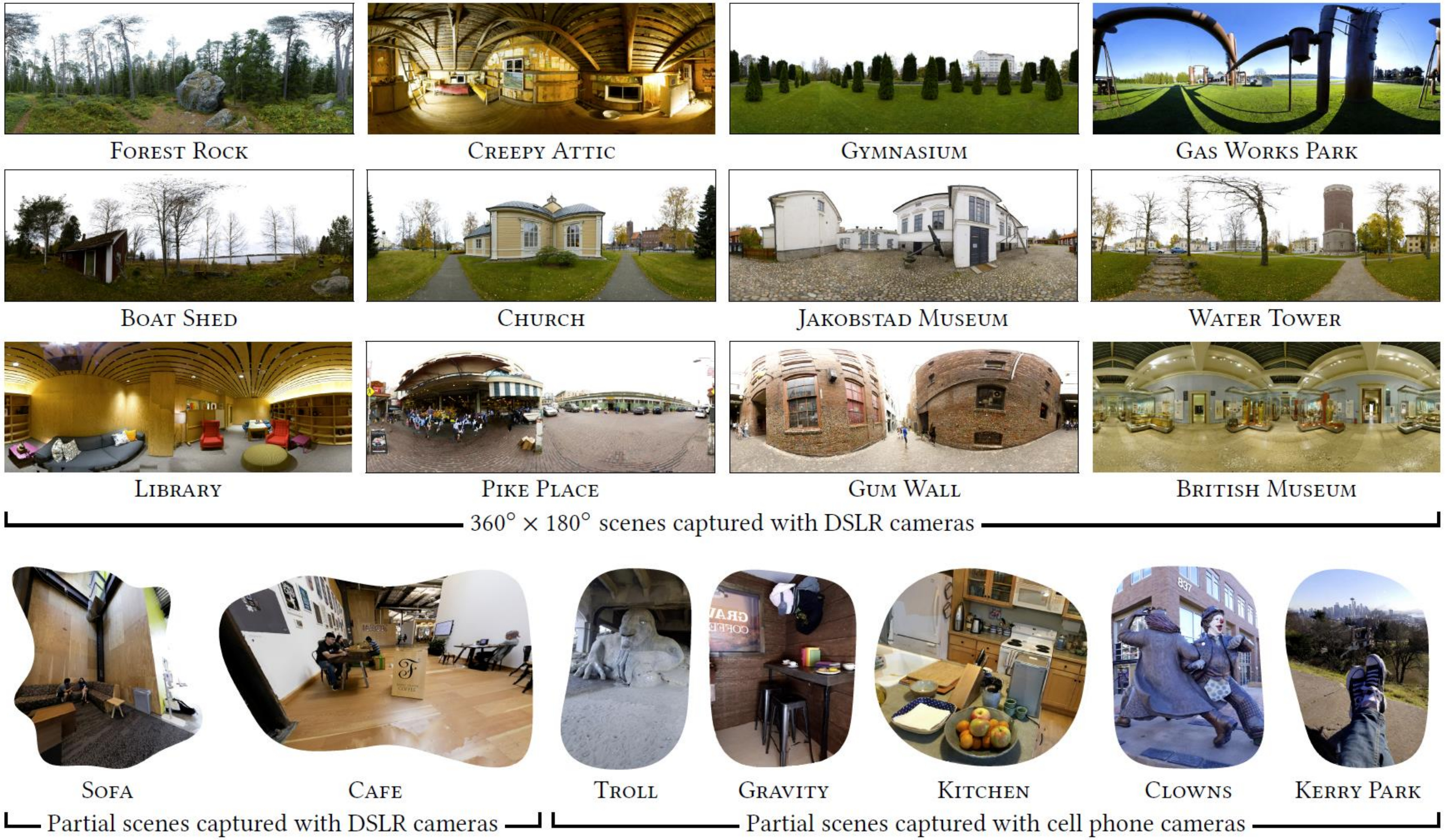
Figure 2: A breakdown of the 3D photo reconstruction algorithm into its six stages, with corresponding inputs and outputs: (a) Capture and pre-processing, Sec. 4.1; (b) Sparse reconstruction, Sec. 4.2; (c) Dense reconstruction, Sec. 4.3; (d) Warping into a central panorama, Sec. 4.4.1; (e) Parallax-tolerant Stitching, Sec. 4.4.2; (f) Two-layer fusion, Sec. 4.4.3.

Casual 3D Photography



(a) Front color-and-depth panorama (b) Front detail (c) Back detail

Casual 3D Photography



Instant 3D Photography

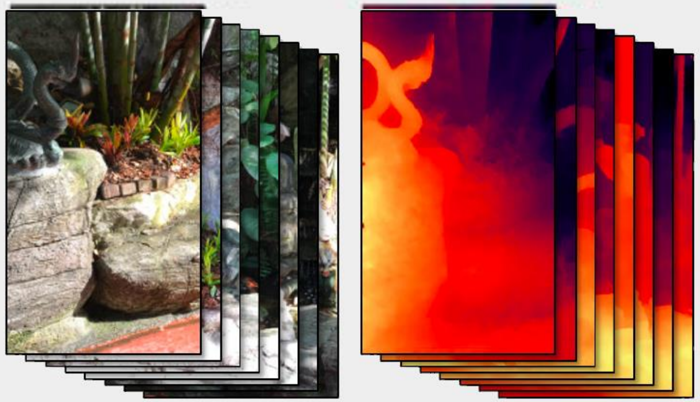
Peter Hedman
University College London *

Johannes Kopf
Facebook

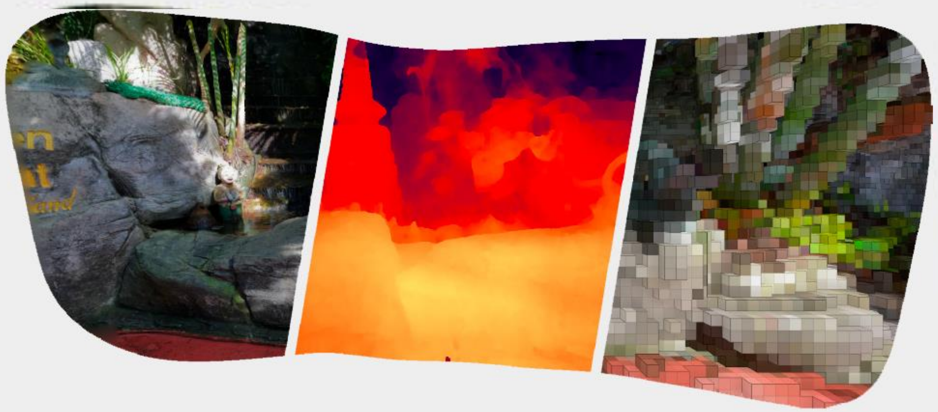
GENERATIONS / SIGGRAPH2018

* This work was done while Peter was working as a contractor for Facebook.

Dual camera phone



Input burst of 34 color-and-depth photos, captured in 34.0 seconds



Our 3D panorama (showing color, depth, and a 3D effect), generated in 34.7 seconds.

Our work enables practical and casual 3D capture with regular dual camera cell phones. Left: A burst of input color-and-depth image pairs that we captured with a dual camera cell phone at a rate of one image per second. Right: 3D panorama generated with our algorithm in about the same time it took to capture. The geometry is highly detailed and enables viewing with binocular and motion parallax in VR, as well as applying 3D effects that interact with the scene, e.g., through occlusions (right).

Practical 3D Photography

Johannes Kopf Suhib Alsisan Francis Ge Yangming Chong Kevin Matzen
Ocean Quigley Josh Patterson Jossie Tirado Shu Wu Michael F. Cohen
Facebook

Practical 3D Photography

Johannes Kopf, Suhib Alsisan, Francis Ge, Yangming Chong, Kevin Matzen, Ocean Quigley, Josh Patterson, Jossie Tirado, Shu Wu, Michael F. Cohen

CVPR Workshop on Computer Vision for Augmented and Virtual Reality, Long Beach, CA, 2019.

PDF

#spotlight, #demo

(a) Input (setup) (100 ms)

(b) LDI (inpainted color / depth) (1100 ms)

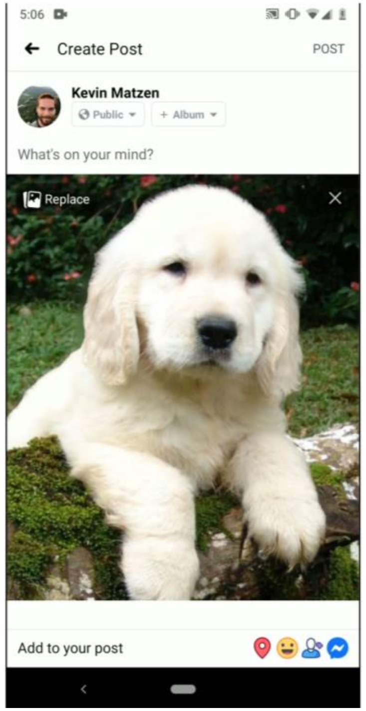
(d) Triangle Mesh (100 ms)

(e) Novel view (30fps)

Figure 1. 3D Photo Creation. Runtime measured on iPhone X.

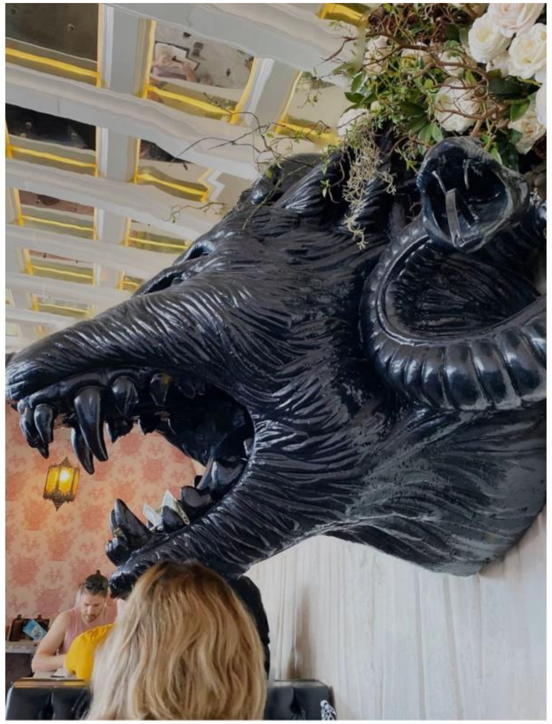
3D Photos on Facebook

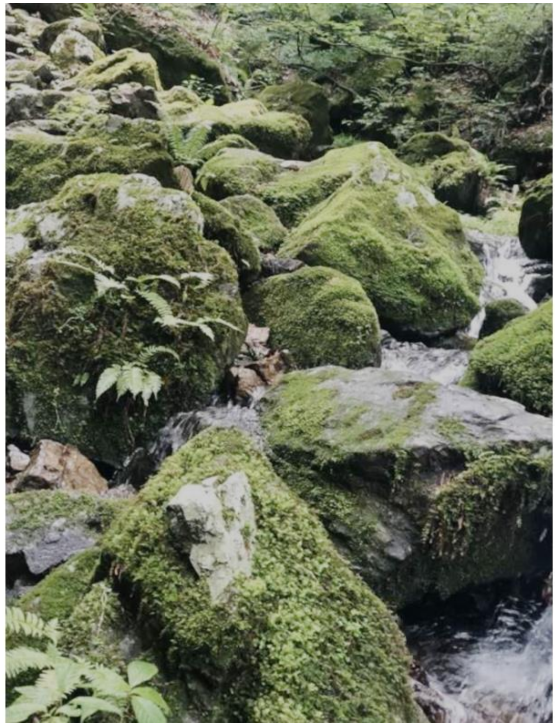
Estimate depth map from photo to create an interactive animation



3D Photos on Facebook

Estimate depth map from photo to create an interactive animation





Google Photos cinematic effect

Jamie Aspinall

Product Manager, Google Photos

Published Dec 15, 2020

Relive the moment with Cinematic photos

Cinematic photos help you relive your memories in a way that feels more vivid and realistic—so you feel like you're transported back to that moment. To do this, we use machine learning to predict an image's depth and produce a 3D representation of the scene—even if the original image doesn't include depth information from the camera. Then we animate a virtual camera for a smooth panning effect—just like out of the movies.

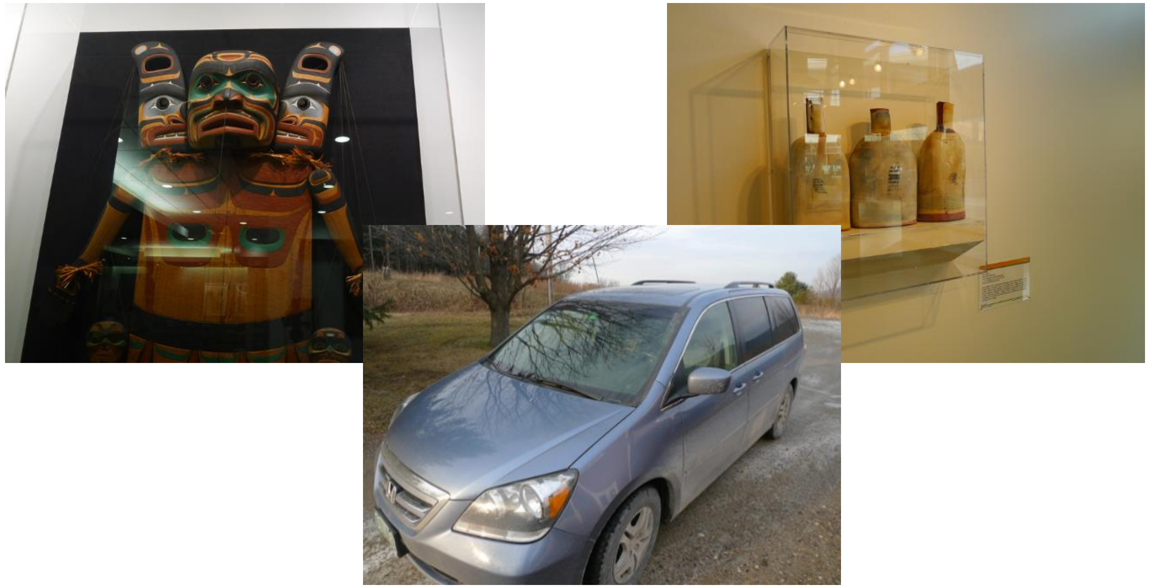
https://blog.google/products/photos/new-cinematic-photos-and-more-ways-relive-your-memories/

What’s missing?

Reflections and Transparency

Image-Based Rendering with Reflections

- Reflections, gloss, and highlights are everywhere




- How do these affect image-based modeling / rendering?

  [Sinha *et al.*, SIGGRAPH 2012]

Standard IBR with Reflections

Our New Rendering System


Two Layers
One Layer


Our Result
Two Layers
Standard IBR
One Layer

Input

Front Depth

Rear Depth

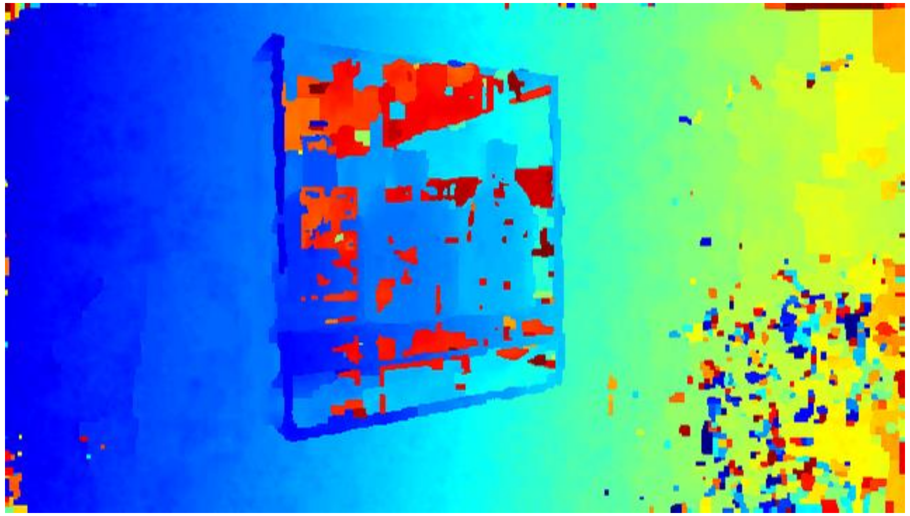
Input

Front Layer

Rear Layer

Image-Based Rendering in the Gradient Domain

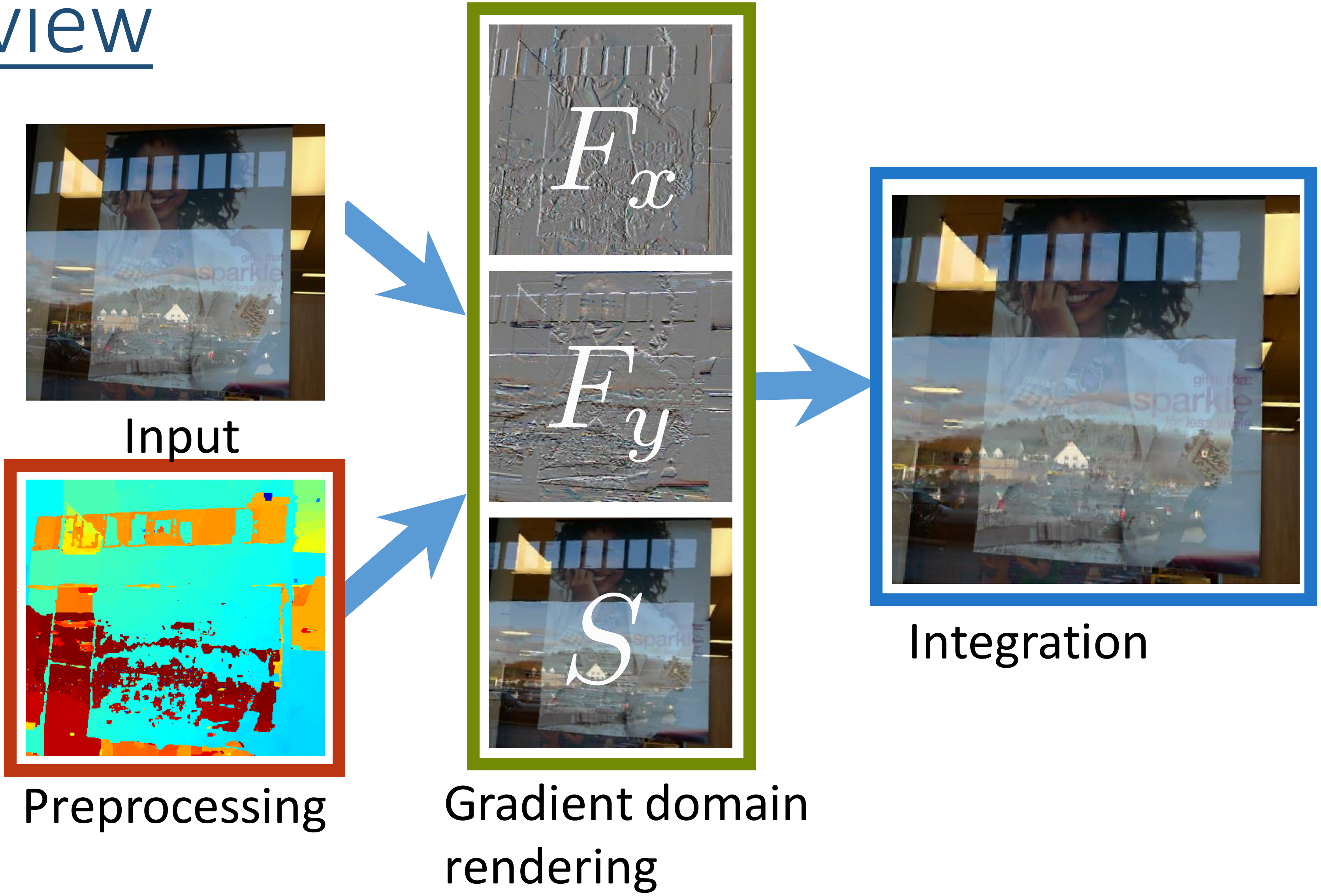
- Wrong depth for textureless or transparent areas



- Solve by reconstructing depth at gradients and re-integrating

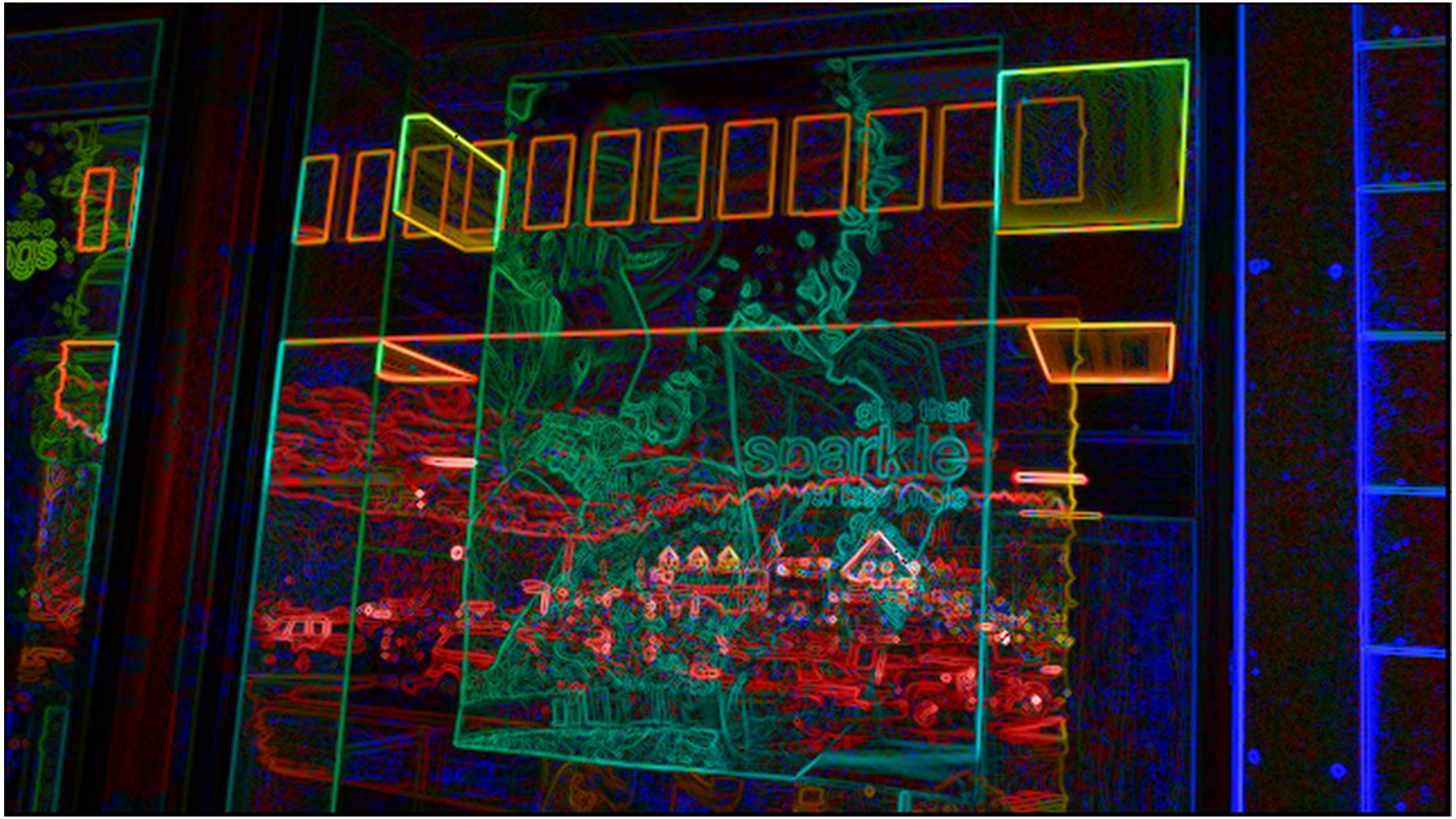
  [Kopf *et al.* SIGGRAPH Asia 2013]

Overview

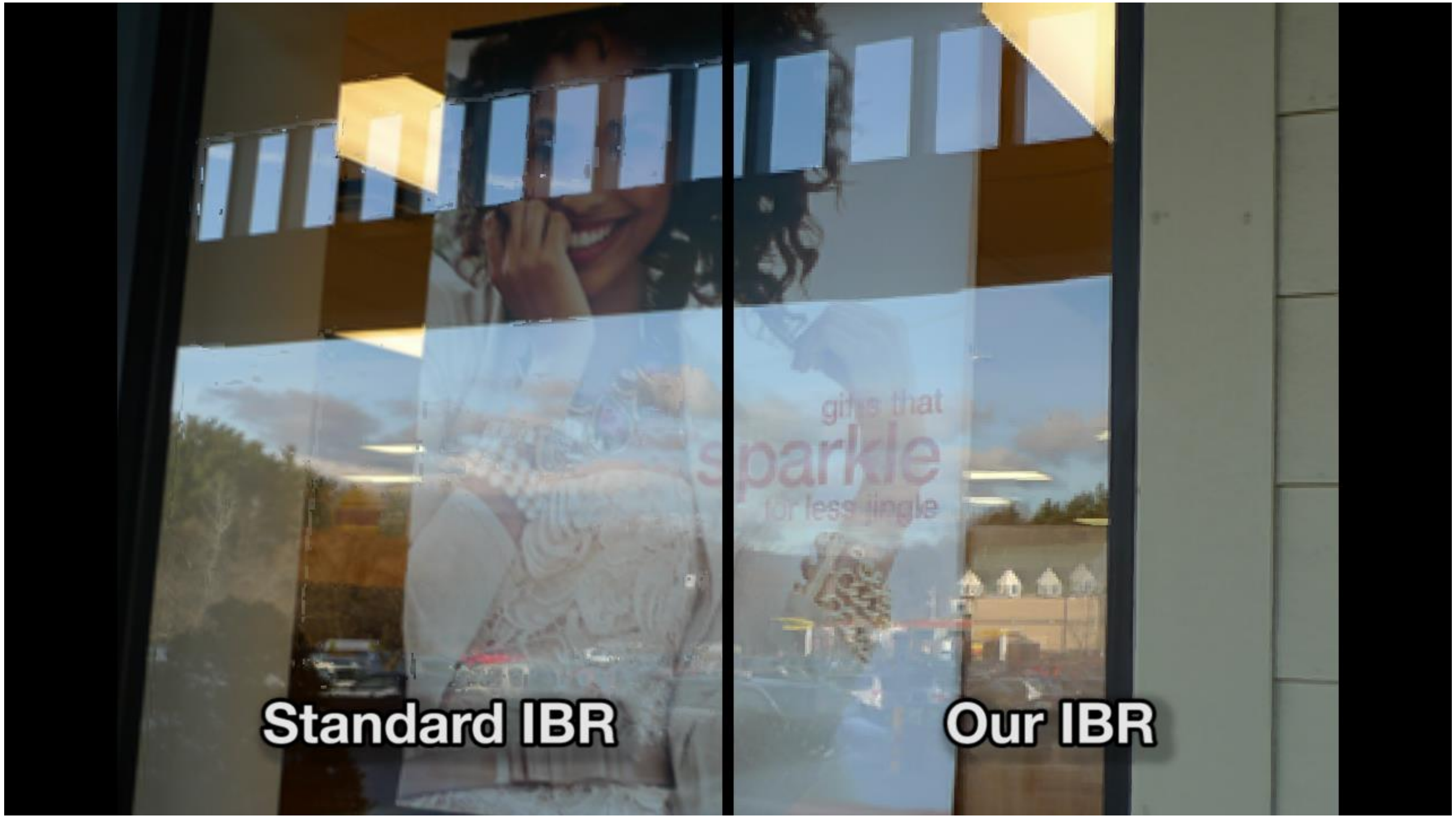


Input
F_x
F_y
S
Preprocessing
Gradient domain rendering
Integration

Gradient Domain



Our Method



Open issues

- Improve stereo matching
  - Plane + parallax representation
- Reflectivity (β) estimation
  - Iterative Refinement
- Handle distorted reflections
  - [See next slide]
- Model real-valued reflectivity
  - Fresnel reflection

Real-World Normal Map Capture for Nearly Flat Reflective Surfaces

Bastien Jacquet[1], Christian Häne[1], Kevin Köser[12*], Marc Pollefeys[1]

ETH Zürich[1]
Zürich, Switzerland

GEOMAR Helmholtz Centre for Ocean Research[2]
Kiel, Germany

Abstract

Although specular objects have gained interest in recent years, virtually no approaches exist for markerless reconstruction of reflective scenes in the wild. In this work, we present a practical approach to capturing normal maps in real-world scenes using video only. We focus on nearly planar surfaces such as windows, facades from glass or metal, or frames, screens and other indoor objects and show how normal maps of these can be obtained without the use of an artificial calibration object. Rather, we track the reflections of real world straight lines, while moving with a hand held

Figure 1. Real-world glass reflection. Notice that reflection in different windows on the same facade can appear very different due to minor deformations and normal variations. Our goal is to capture normal maps of real windows to faithfully reproduce this effect.

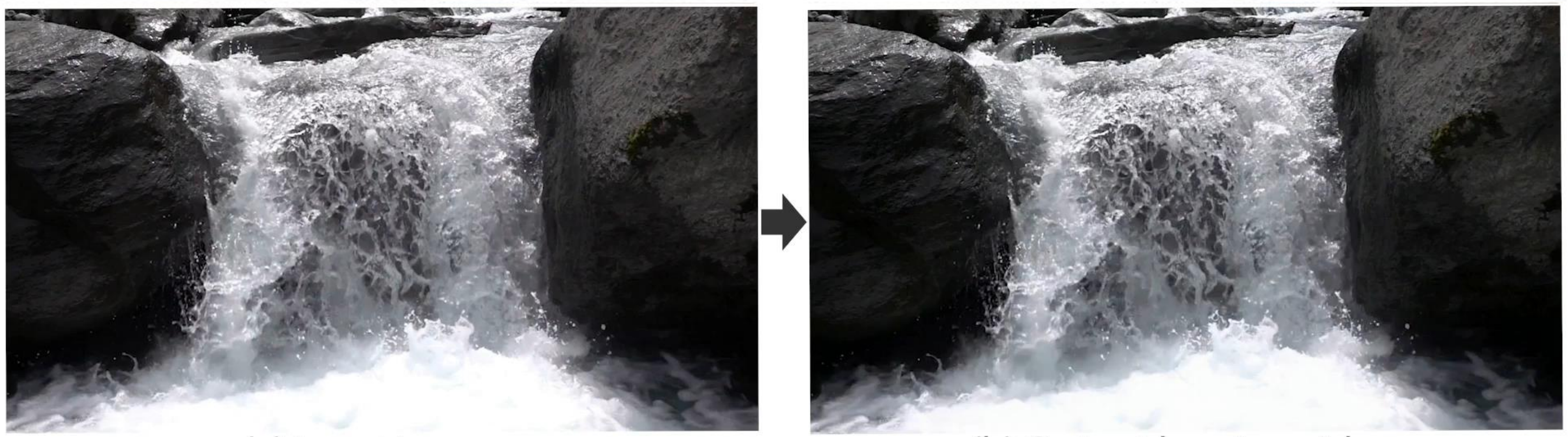
Neural Rendering

Animating Pictures

Animating Pictures with Eulerian Motion Fields

Aleksander Holynski[1], Brian Curless[1], Steven M. Seitz[1], Richard Szeliski[2]
[1]University of Washington, [2]Facebook

Paper | arXiv | Video | Code (coming soon)



(a) Input image (b) Output looping video

https://eulerian.cs.washington.edu/

Animating Pictures

Figure 14.17 *Video textures (Schödl, Szeliski et al. 2000) © 2000 ACM: (a) a clock pen-*

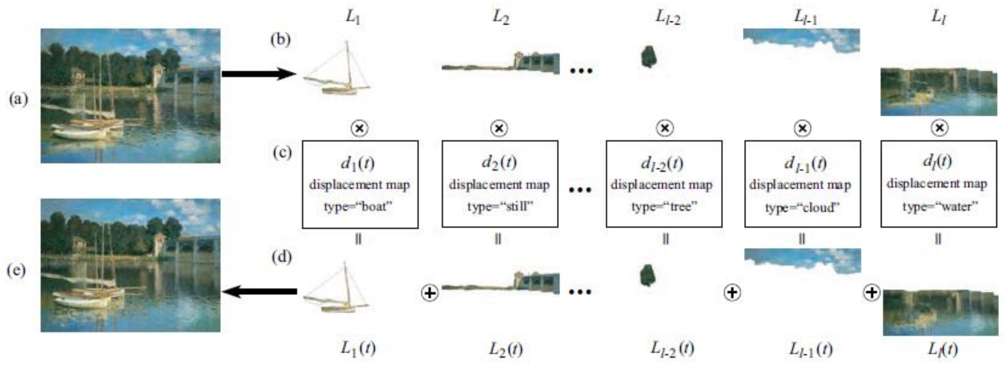


Figure 14.18 *Animating still pictures (Chuang, Goldman et al. 2005) © 2005 ACM. (a) The input still image is manually segmented into (b) several layers. (c) Each layer is then animated with a different stochastic motion texture (d) The animated layers are then composited to produce (e) the final animation*

Animating Pictures

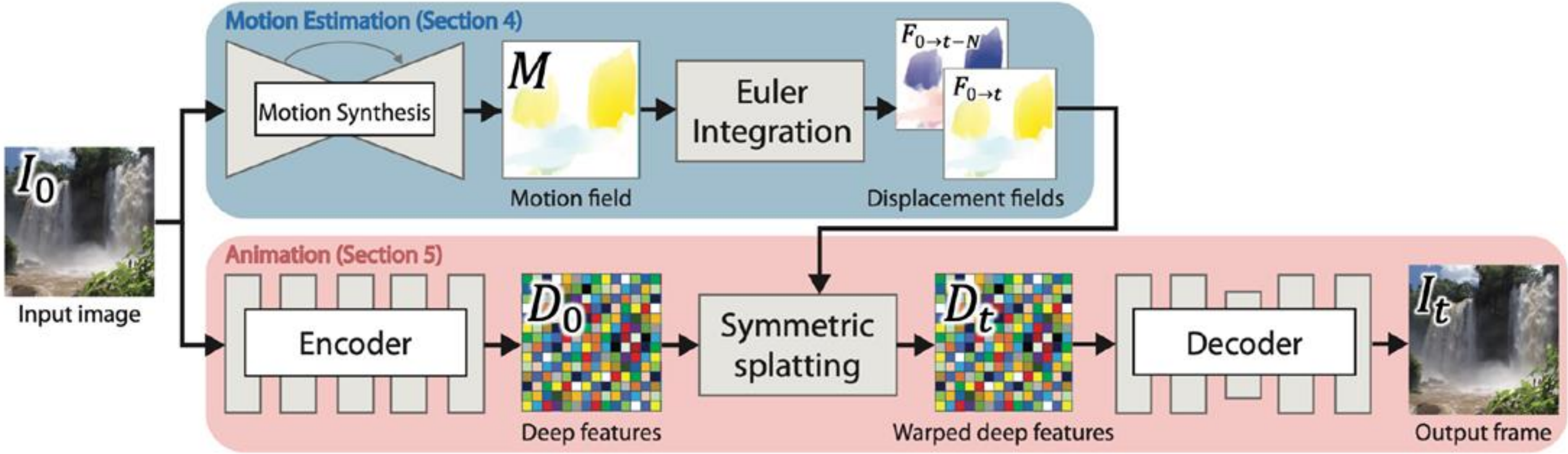


Figure 2: **Overview:** Given an input image I_0, our motion estimation network predicts a motion field M. Through Euler integration, M is used to generate future and past displacement fields $F_{0\to t}$ and $F_{0\to t-N}$, which define the source pixel locations in all other frames t. To animate the input image using our estimated motion, we first use a feature encoder network to encode the image as a feature map D_0. This feature map is warped by the displacement fields (using a novel symmetric splatting technique) to produce the corresponding warped feature map D_t. The warped features are provided to the decoder network to create the output video frame I_t.

Animating Pictures

Animating Pictures

Figure 3: **Deep warping:** Above: Naïve splatting of RGB pixels results in increasingly large unknown regions over time, shown in magenta. Below: For the same frames, our deep warping approach synthesizes realistic texture in these unknown regions.

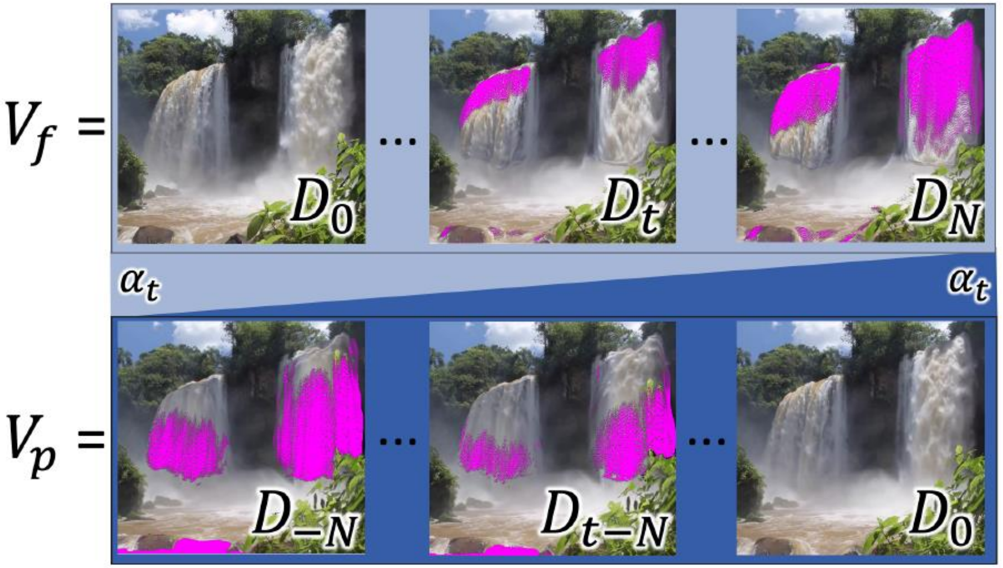


Figure 4: **Seamless looping**: An illustrated example of how seamless loops are created. Two feature videos are created by warping D_0. The first, V_f, contains the result of integrating the motion field M, resulting in a video starting with the input image and animating into the future. The second, V_p, instead uses $-M$, resulting in a video starting in the past and ending with the input frame. These two videos typically contain complementary unknown regions (shown in magenta). Before decoding, we combine the two

Animating Pictures

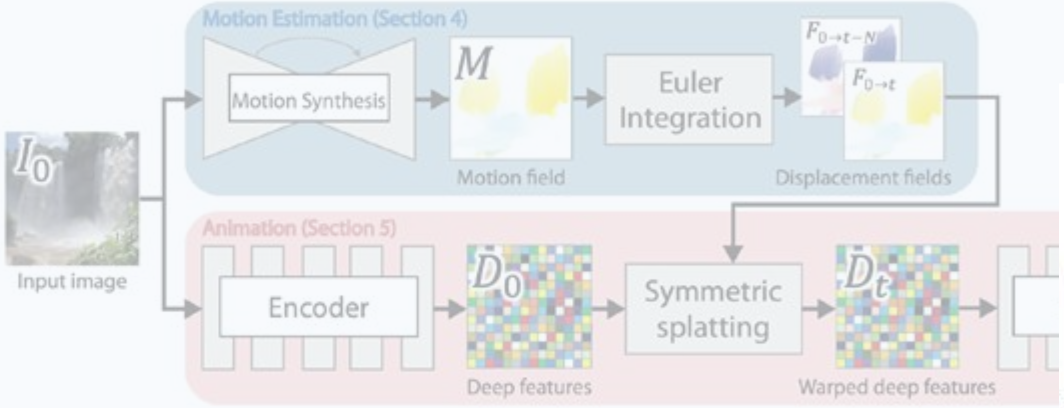


Figure 2: **Overview:** Given an input image I_0, our motion estimation network predicts a motion field
is used to generate future and past displacement fields $F_{0\to t}$ and $F_{0\to t-N}$, which define the source pi
To animate the input image using our estimated motion, we first use a feature encoder network to encod
This feature map is warped by the displacement fields (using a novel symmetric splatting technique) to p
feature map D_t. The warped features are provided to the decoder network to create the output video fra

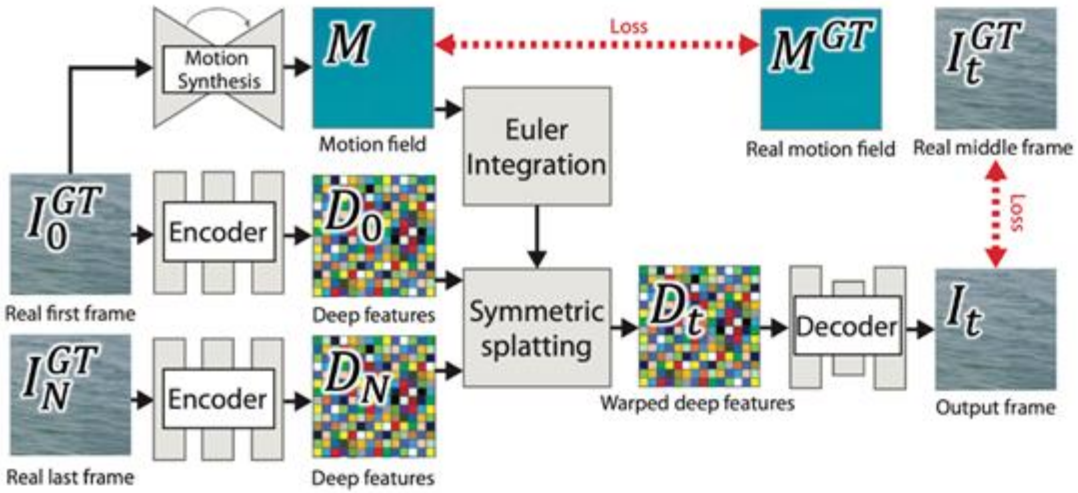


Figure 5: **Training:** As described in Section 5.1, each frame in our generated looping video is composed of textures from two warped frames. To supervise this process during training, i.e., to have a real frame to compare against, we perform our symmetric splatting using the features from two different frames, I_0 and I_N (instead of I_0 twice, as in inference). We enforce the motion field M to match

Animating Pictures

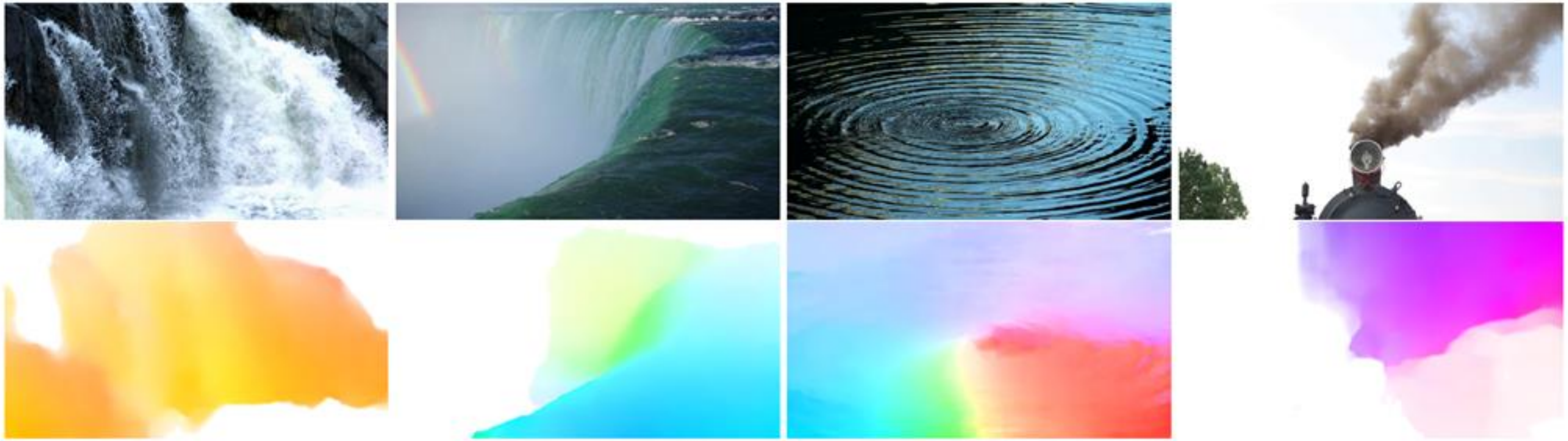


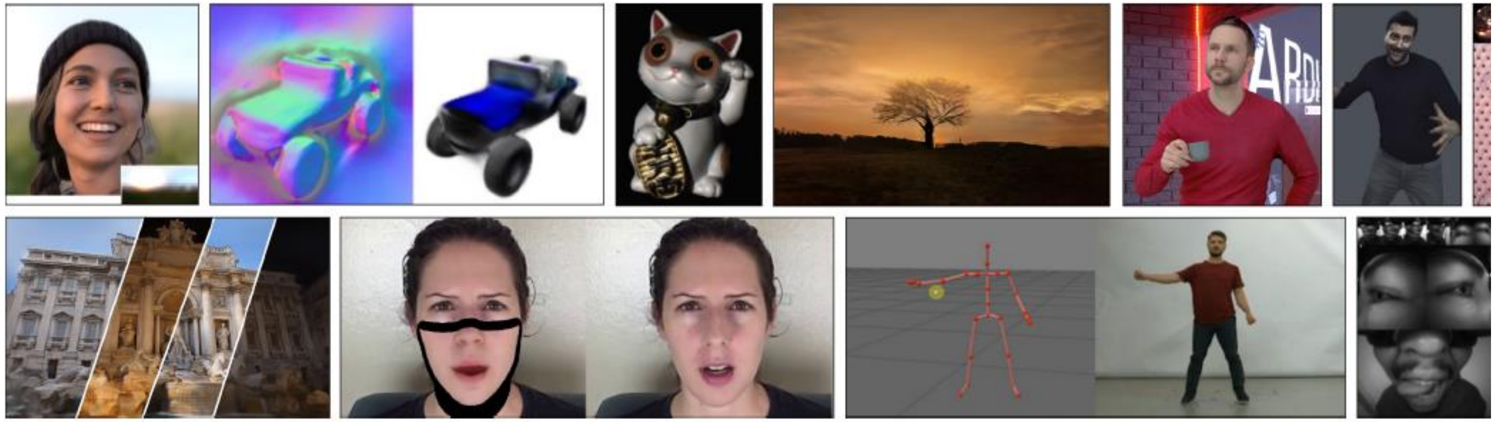
Figure 6: Examples of the input images (top), alongside their corresponding synthesized motion fields (bottom). Full resolution images, along with their corresponding animated videos, can be found in the supplementary video.

https://eulerian.cs.washington.edu/

State of the Art on Neural Rendering (2020)

A. Tewari[1*] O. Fried[2*] J. Thies[3*] V. Sitzmann[2*] S. Lombardi[4] K. Sunkavalli[5] R. Martin-Brualla[6] T. Simon[4] J. Saragih[4] M. Nießner[3]
R. Pandey[6] S. Fanello[6] G. Wetzstein[2] J.-Y. Zhu[5] C. Theobalt[1] M. Agrawala[2] E. Shechtman[5] D. B Goldman[6] M. Zollhöfer[4]

[1]MPI Informatics [2]Stanford University [3]Technical University of Munich [4]Facebook Reality Labs [5]Adobe Research [6]Google Inc *Equal



... us methods. Images from [S

... ut combines genera-
... om computer graph-
... twork training. With
... neural rendering is

Neural Rendering

CVPR 2020 tutorial.

09:00–09:15	Welcome and Introduction	Michael Zollhöfer
09:15–09:30	Fundamentals, Taxonomy, Neural Rendering	Ayush Tewari
Semantic Photo Synthesis and Manipulation		
09:30–09:40	Overview	Jun-Yan Zhu
09:40–10:00	Semantic Image Synthesis with Spatially-Adaptive Normalization	Taesung Park
10:00–10:30	Coffee Break	
Facial Reenactment & Body Reenactment		
10:25–10:35	Overview	Justus Thies
10:35–11:00	Neural Rendering for High-Quality Synthesis of Human Portrait Video and Images	Christian Theobalt
11:00–11:20	Neural Rendering for Virtual Avatars	Aliaksandra Shysheya

Novel View Synthesis		
11:20–11:35	Overview	Vincent Sitzmann
11:30–11:50	Neural Rerendering in the Wild	Moustafa Meshry
11:50–12:10	NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis	Ben Mildenhall
12:10–13:20	Lunch Break	
Learning to Relight		
13:20–13:30	Overview	Zexiang Xu
13:30–13:50	Multi-view Relighting Using a Geometry-Aware Network	Julien Philip
13:50–14:10	Neural Inverse Rendering	Abhimitra Meka
Free Viewpoint Videos		
14:10–14:20	Overview	Sean Fanello
14:20–14:40	Neural Rendering for Performance Capture	Rohit K. Pandey
14:40–15:00	Neural Volumes: Learning Dynamic Renderable Volumes from Images	Stephen Lombardi
15:00–15:30	Coffee Break	
15:30–15:45	Social Implications, Open Challenges, Conclusion	Ohad Fried
15:45–16:15	Followup Discussion	

A. Tewari & O. Fried & J. Thies et al. / State of the Art on Neural Rendering

Method	Required Data	Network Inputs	Network Outputs	Contents	Controllable Parameters	Explicit Control	CG Module	Generality	Multi-modal Synthesis	Temporal Coherence	
Bau et al. [BSP*19a]	IS	IS	I	RE	S	✗	✗	✓	✗	✗	Semantic Photo Synthesis (Section 6.1)
Brock et al. [BLRW17]	I	N	I	S	R	✓	✗	✓	✗	✗	
Chen and Koltun [CK17]	IS	S	I	RE	S	✗	✗	✓	✓	✗	
Isola et al. [IZZE17]	IS	S	I	ES	S	✗	✗	✓	✗	✗	
Karacan et al. [KAEE16]	IS	S	I	E	S	✗	✗	✓	✓	✗	
Park et al. [PLWZ19b]	IS	S	I	RE	S	✗	✗	✓	✓	✗	
Wang et al. [WLZ*18b]	IS	S	I	RES	S	✗	✗	✓	✓	✗	
Zhu et al. [ZKSE16]	I	N	I	ES	RT	✓	✗	✓	✓	✗	
Aliev et al. [AUL19]	ID	R	I	RS	C	✓	N	✗	✗	✗	Novel View Synthesis (Section 6.2)
Eslami et al. [ERB*18]	IC	IC	I	RS	C	✓	✗	✓	✗	✗	
Hedman et al. [HPP*18]	V	I	I	RES	C	✓	N	✓	✗	✗	
Meshry et al. [MGK*19]	I	IL	I	RE	CL	✓	N	✗	✓	✗	
Nguyen-Phuoc et al. [NPLBY18]	ICL	E	I	S	CL	✓	N	✓	✗	✗	
Nguyen-Phuoc et al. [NLT*19]	I	NC	I	S	C	✓	✗	✓	✓	✗	
Sitzmann et al. [STH*19]	V	IC	I	S	C	✓	D	✗	✗	✗	
Sitzmann et al. [SZW19]	IC	IC	I	S	C	✓	D	✓	✗	✗	
Thies et al. [TZT*20]	V	IRC	I	S	C	✓	N	✗	✗	✗	
Xu et al. [XBS*19]	IC	IC	I	S	C	✓	D	✓	✗	✗	
Lombardi et al. [LSS*19]	VC	IC	I	HPS	C	✓	D	✗	✗	✗	Free Viewpoint Video (Section 6.3)
Martin-Brualla et al. [MBPY*18]	VDC	R	V	P	C	✓	N	✓	✗	✓	
Pandey et al. [PTY*19]	VDI	IDC	I	P	C	✓	✗	✓	✗	✗	
Shysheya et al. [SZA*19]	V	R	I	P	CP	✓	✗	✓	✗	✗	
Meka et al. [MHP*19]	IL	IL	I	H	L	✓	✗	✓	✗	✗	Relighting (Section 6.4)
Philip et al. [PGZ*19]	I	IL	I	E	L	✓	N	✓	✗	✗	
Sun et al. [SBT*19]	IL	IL	IL	H	L	✓	✗	✓	✗	✗	
Xu et al. [XSHR18]	IL	IL	I	S	L	✓	✗	✓	✗	✗	
Zhou et al. [ZHSJ19]	IL	IL	IL	H	L	✓	✗	✓	✗	✗	
Fried et al. [FTZ*19]	VT	VR	V	H	H	✓	N	✗	✗	✓	Facial Reenactment (Section 6.5)
Kim et al. [KGT*18]	V	R	V	H	PE	✓	N	✗	✗	✓	
Lombardi et al. [LSSS18]	VC	IMC	MX	H	CP	✓	N	✗	✗	✗	
Thies et al. [TZN19]	V	IRC	I	HS	CE	✓	D	✗	✗	✗	
Wei et al. [WSS*19]	VC	I	MX	H	CP	✓	D	✗	✗	✗	
Zakharov et al. [ZSBL19]	I	IK	I	H	PE	✗	✗	✓	✗	✗	
Aberman et al. [ASL*19]	V	J	V	P	P	✗	✗	✗	✗	✓	Body Reenactment (Section 6.5)
Chan et al. [CGZE18]	V	J	V	P	P	✗	✗	✗	✗	✓	
Liu et al. [LXZ*19]	VM	R	V	P	P	✓	N	✗	✗	✓	
	Inputs and Outputs				Control		Misc				

Table 1: *Selected methods presented in this survey. See Section 6 for explanation of attributes in the table and their possible value.*

Advances in Neural Rendering (2022)

A. Tewari[1,6*] J. Thies[2*] B. Mildenhall[3*] P. Srinivasan[3*] E. Tretschk[1] W. Yifan[4,8] C. Lassner[5] V. Sitzmann[6] R. Martin-Brualla[3]
S. Lombardi[5] T. Simon[5] C. Theobalt[1] M. Nießner[7] J. T. Barron[3] G. Wetzstein[8] M. Zollhöfer[5] V. Golyanik[1]

[1]MPI for Informatics [2]MPI for Intelligent Systems [3]Google Research [4]ETH Zürich [5]Reality Labs Research
[6]MIT [7]Technical University of Munich [8]Stanford University *Equal contribution.

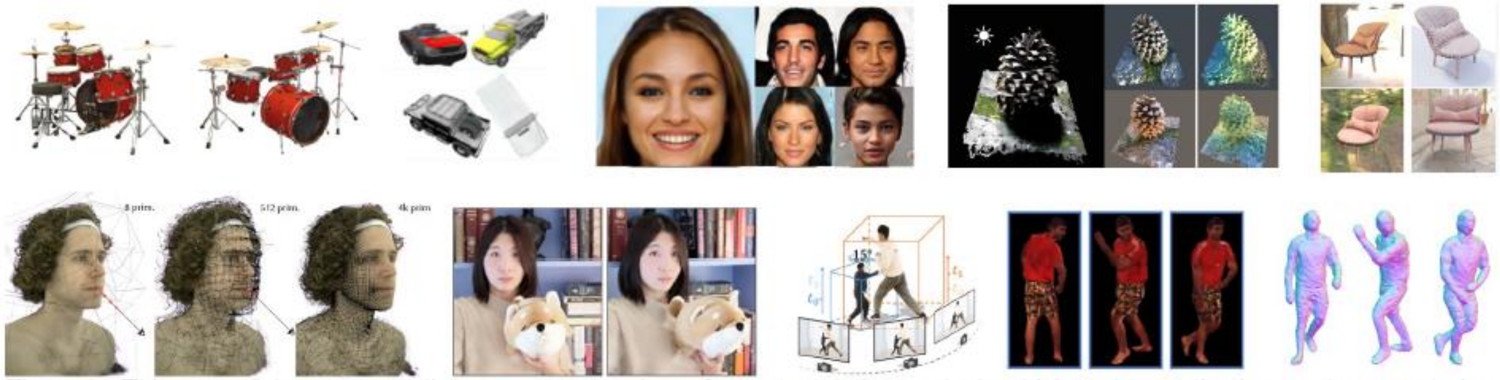


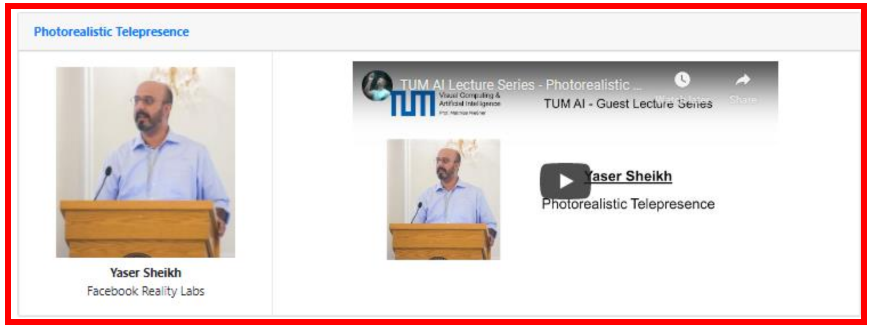
Figure 1: This state-of-the-art report discusses a large variety of neural rendering methods which enable applications such as novel-view synthesis of static and dynamic scenes, generative modeling of objects, and scene relighting. See Section 4 for more details on the various methods. Images adapted from [MST*20, TY20, CMK*21, ZSD*21, BBJ*21, LSS*21, PSB*21, JXX*21, PDW*21] ©2021 IEEE.

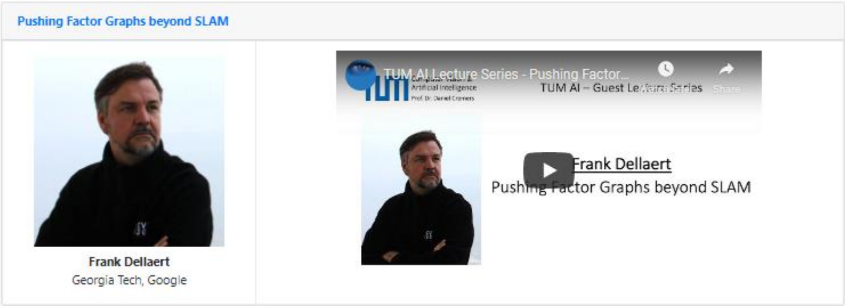
Advances in Neural Rendering

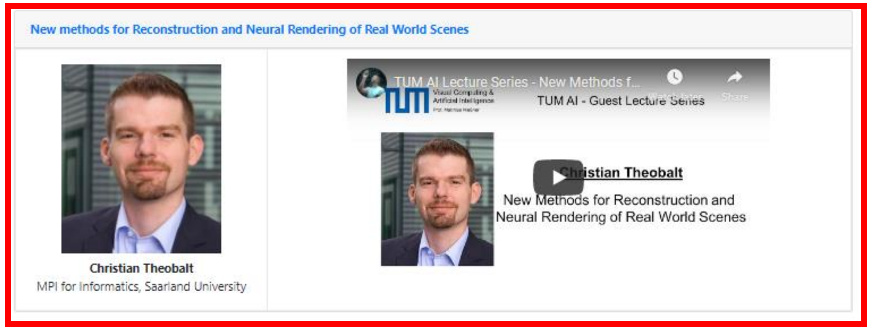
SIGGRAPH 2021 Course

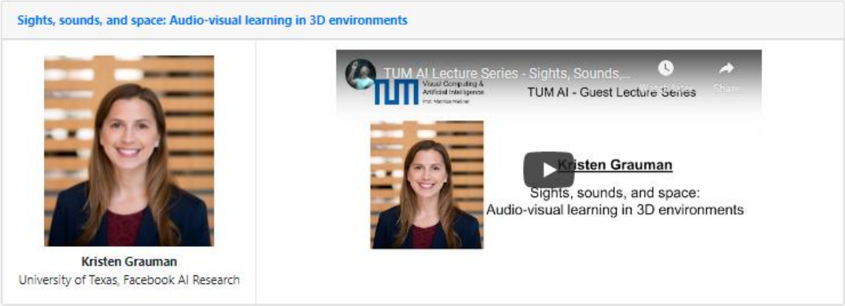
Introduction	
Fundamentals	Michael Zollhöfer
Generative Adversarial Networks	
Loss Functions	Jun-Yan Zhu
GANs with 3D Control	Ayush Tewari
Neural Scene Representations	
Neural Scene Representations	Gordon Wetzstein
Novel View Synthesis for Objects and Scenes	
Introduction	Vincent Sitzmann
Neural Volumetric Rendering	Ben Mindenhall
Fast Rendering of NeRFs	Lingjie Liu
Towards Instant 3D Capture	Dan Goldman
Deformable NeRFs	Keunhong Park
Learning to Relight	
Relightable and Editable Neural Rendering	Zexiang Xu
Total Relighting	Sergio Orts-Escolano
Relightable NeRFs	Pratul Srinivasan
Compositional Scene Representations	
Compositional Scene Representations	Michelle Guo
Free Viewpoint Videos	
Overview of NeRFs for General Dynamic Scenes	Edgar Tretschk
Efficient Neural Rendering of Dynamic Humans and Scenes	Stephen Lombardi
Neural Rendering for Dynamic Performance Capture	Rohit Pandey
Facial and Body Rendering	
Neural Rendering of Faces and Bodies	Justus Thies
Neural Rendering and Video-based Animation of Human Actors	Christian Theobalt
Neural Rendering for Animatable Avatars	Tomas Simon

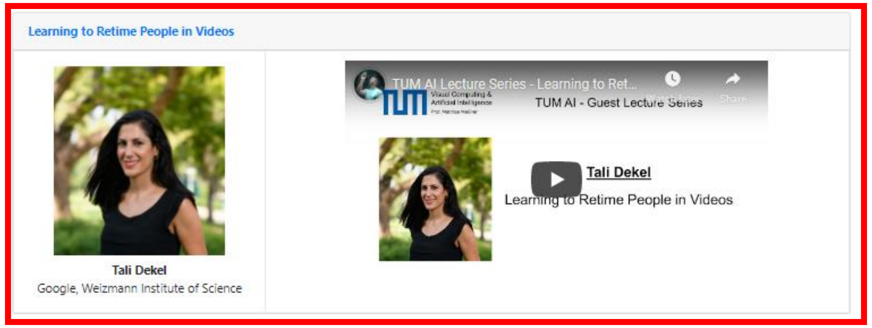
TUM AI Lecture series (2020-2021)



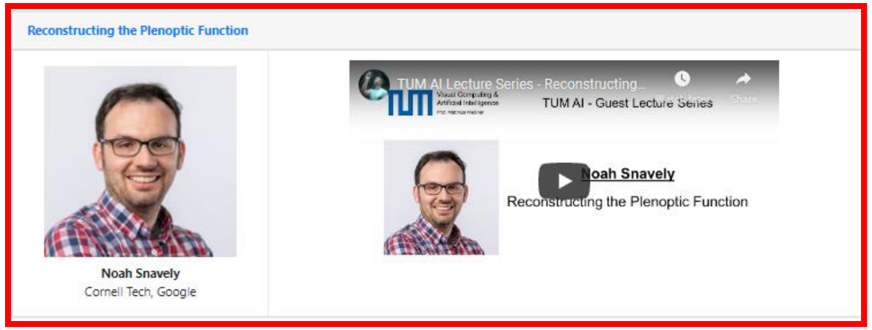


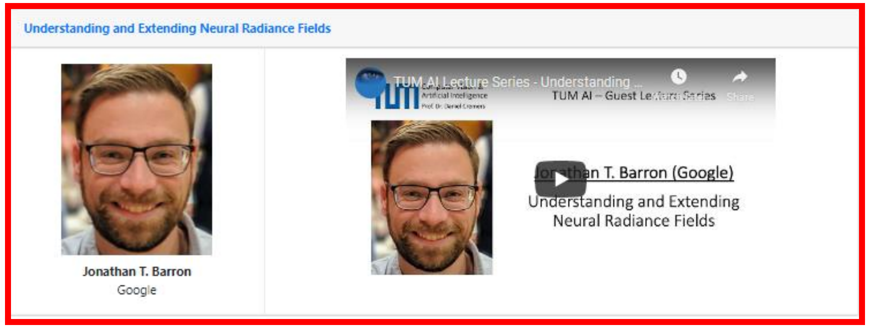


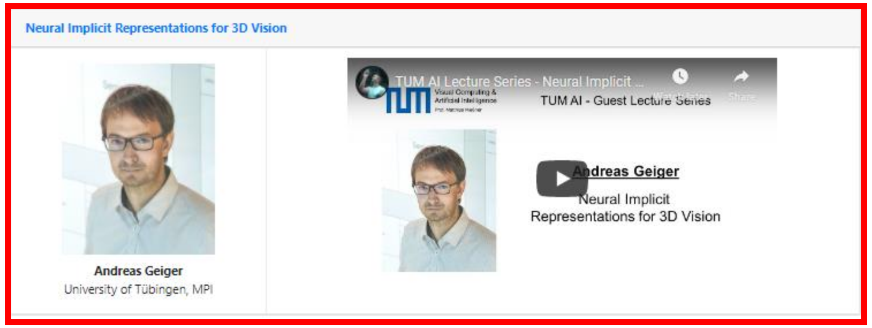




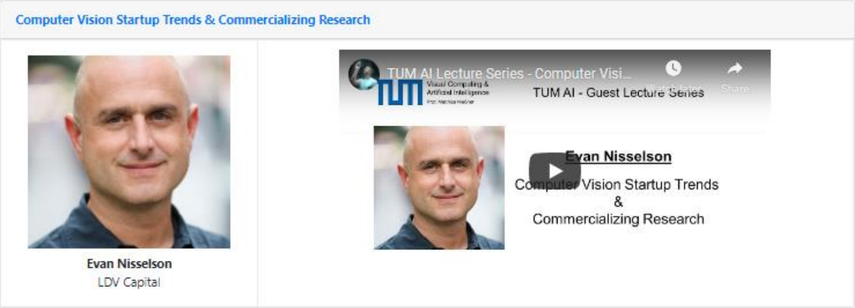
TUM AI Lecture series (2020-2021)





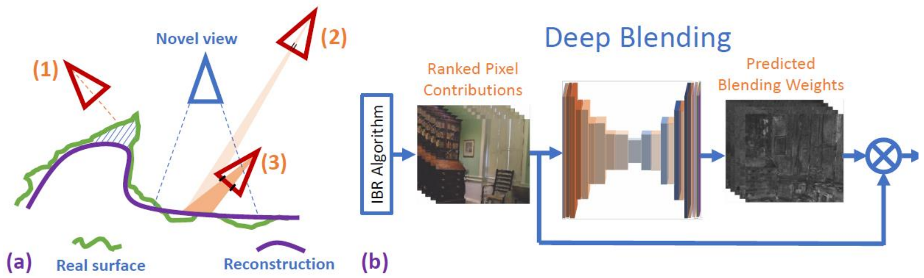


TUM AI Lecture series (2020-2021)

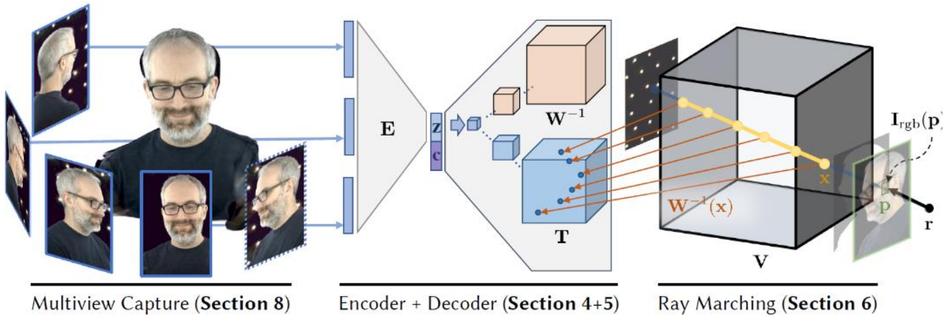


3D representations for neural rendering

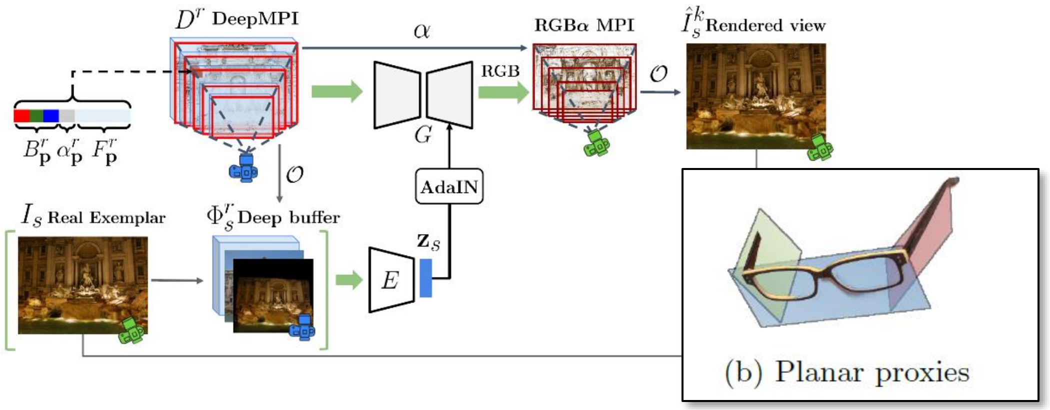
- 3D models & textures



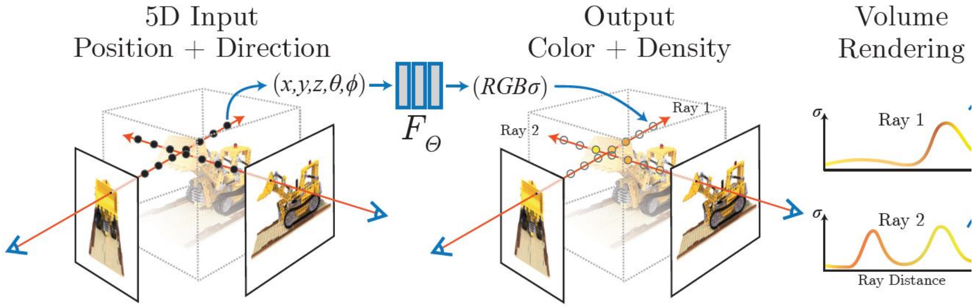
- Voxels



- Depth images and layers



- Implicit functions (MLPs)



Free View Synthesis

2 G. Riegler and V. Koltun

Fig. 1: Novel view synthesis from unstructured input images. The first three images show our synthesized results on the *Truck* scene from Tanks and Temples [21]. The unstructured image sequence was recorded using a handheld cam-

Free View Synthesis 7

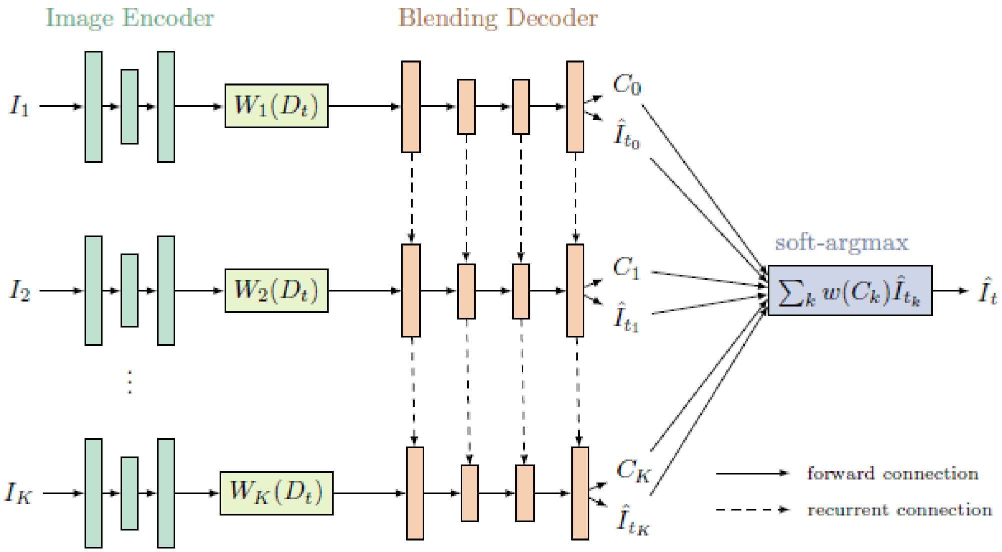
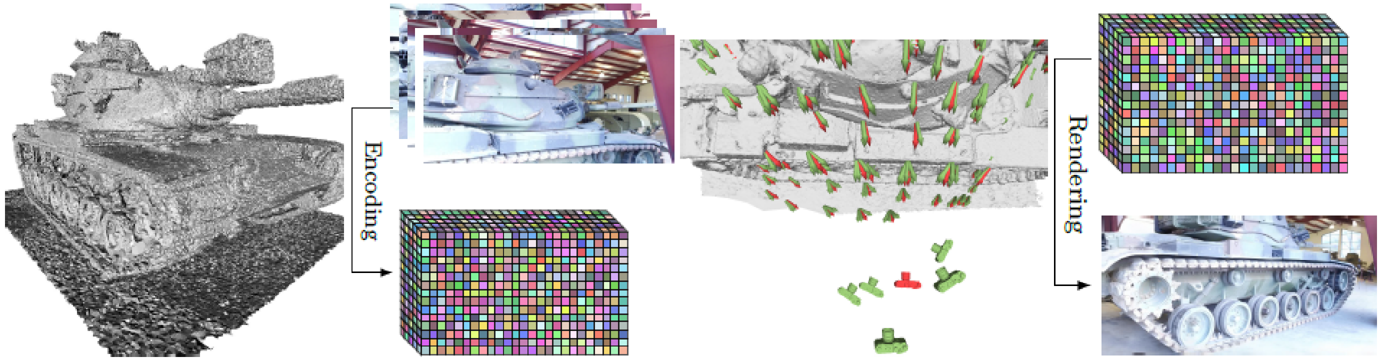


Fig. 3: Overview of the recurrent mapping and blending network. The input is a

Stable View Synthesis [Riegler and Koltun]



(a) Geometric scaffold (b) Encoding source images (c) On-surface aggregation (d) Rendering in the target view

Figure 2: **Overview of Stable View Synthesis.** (a) A geometric scaffold of the scene is constructed using structure-from-motion, multiple-view stereo, and meshing. (b) All source images are encoded into feature tensors via a convolutional network. (c) Given a new target view (red camera), feature vectors from the sou
the geometric scaffold. Red arrows map 3D points to the target view, green arr
(d) The output image in the target view is rendered from a tensor of synthesize

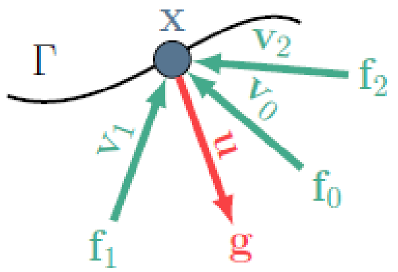


Figure 3: **On-surface aggregation.** A 3D point $\mathbf{x}$ on the geometric scaffold Γ is seen in a set of source images. Each such image contributes a feature vector $\mathbf{f}_k$ along a ray $\mathbf{v}_k$

VK

Vladlen Koltun <vkoltun@gmail.com>

To Richard Szeliski

Cc Gernot Riegler

2/18/2021

Your book is coming along! I read over the monodepth section and the neural rendering section. Both look good to me. They come across as fair and comprehensive. The neural rendering section is a bit like covering Waterloo in the middle of the battle. You can describe the position and movements of the different pieces, but everything is in motion and it's not clear what will survive, what will be remembered, and what shape this will all settle into.

Best,
Vladlen

NeRF++

NERF++: ANALYZING AND IMPROVING NEURAL RADIANCE FIELDS

Kai Zhang Cornell Tech
Gernot Riegler Intel Labs
Noah Snavely Cornell Tech
Vladlen Koltun Intel Labs

(a) bounding volume for the truck only (b) bounding volume for the entire scene

Figure 5: For 360° captures of unbounded scenes, NeRF's parameterization of space either models only a portion of the scene, leading to significant artifacts in background elements (a), or models the full scene and suffers from an overall loss of detail due to finite sampling resolution (b).

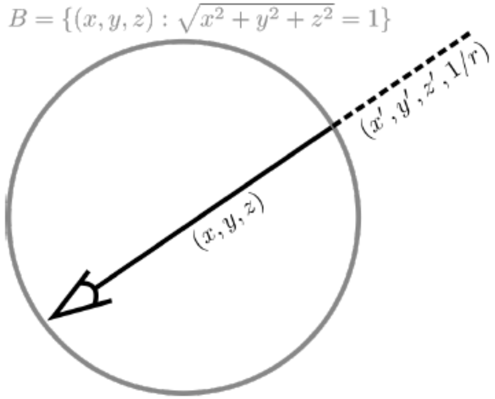


Figure 6: NeRF++ applies different parameterizations for scene contents inside and outside the unit sphere.

- *Adapt volume to camera configuration (MPI ~ volumetric grid)*

NeRV and NeRD (2021)

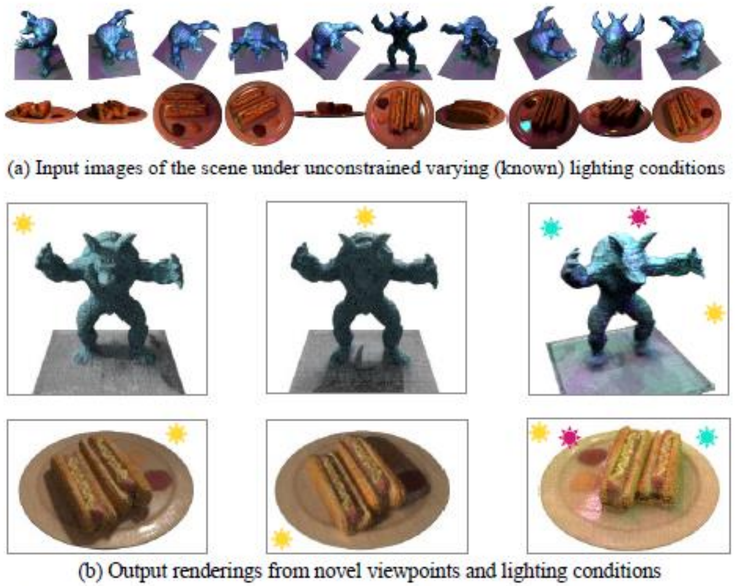


Figure 1: We optimize a Neural Reflectance and Visibility Field (NeRV) 3D representation from a set of input images of a scene illuminated by known but unconstrained lighting. Our NeRV representation can be rendered from novel views under arbitrary lighting conditions not seen during training. Here, we visualize example input data and renderings for two scenes. The first two output rendered images for each scene are from the same viewpoint, each illuminated by a point light at a different location, and the last image is from a different viewpoint under a random colored illumination.

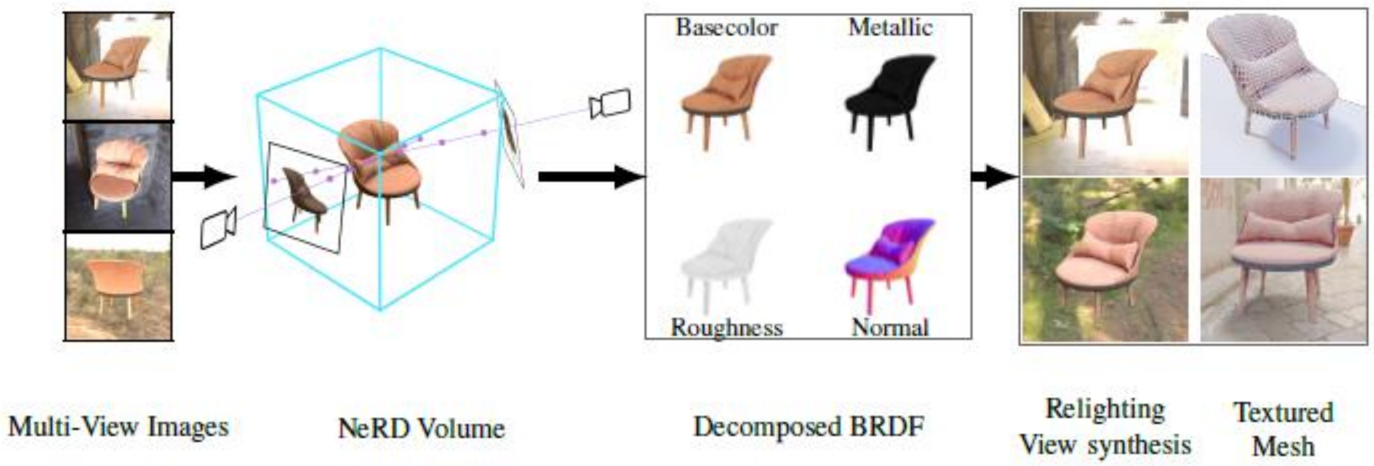
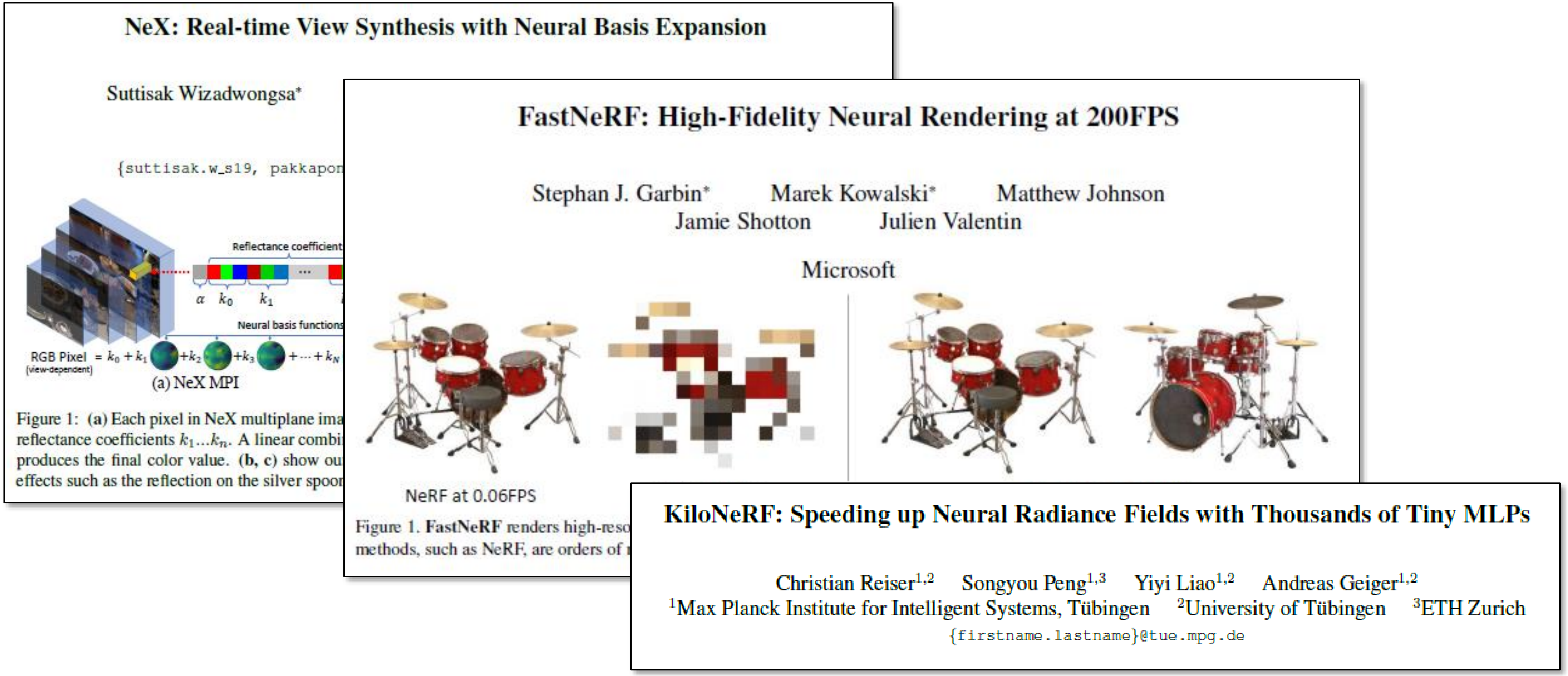


Figure 1: **Neural Reflectance Decomposition.** Multiple views of an object under varying or fixed illumination are encoded into the NeRD volume. During the encoding process, information provided by all samples is decomposed into geometry, spatially-varying BRDF parameters and a rough approximation of the incident illumination in a globally consistent way. This decomposition can be easily extracted and re-rendered under a novel illumination condition in real-time.

Real-time Neural Based Rendering



NeX: Real-time View Synthesis with Neural Basis Expansion

Suttisak Wizadwongsa*

{suttisak.w_s19, pakkapon

(a) NeX MPI

Figure 1: (a) Each pixel in NeX multiplane ima… reflectance coefficients $k_1 \ldots k_n$. A linear combi… produces the final color value. (b, c) show ou… effects such as the reflection on the silver spoon…

FastNeRF: High-Fidelity Neural Rendering at 200FPS

Stephan J. Garbin* Marek Kowalski* Matthew Johnson
Jamie Shotton Julien Valentin

Microsoft

Figure 1. FastNeRF renders high-reso… methods, such as NeRF, are orders of …

KiloNeRF: Speeding up Neural Radiance Fields with Thousands of Tiny MLPs

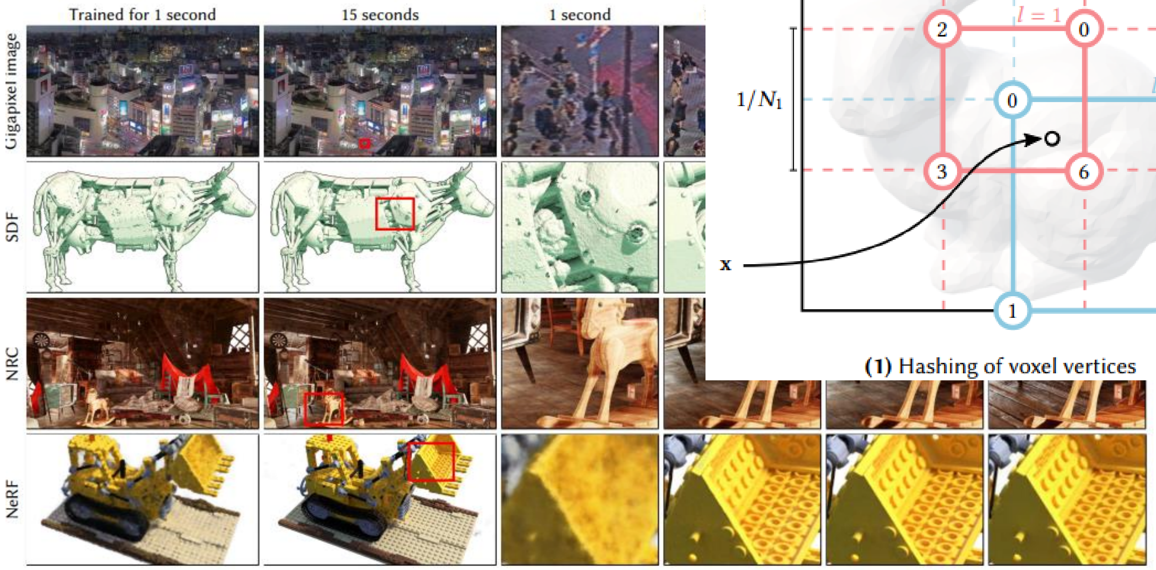
Christian Reiser[1,2] Songyou Peng[1,3] Yiyi Liao[1,2] Andreas Geiger[1,2]
[1]Max Planck Institute for Intelligent Systems, Tübingen [2]University of Tübingen [3]ETH Zurich
{firstname.lastname}@tue.mpg.de

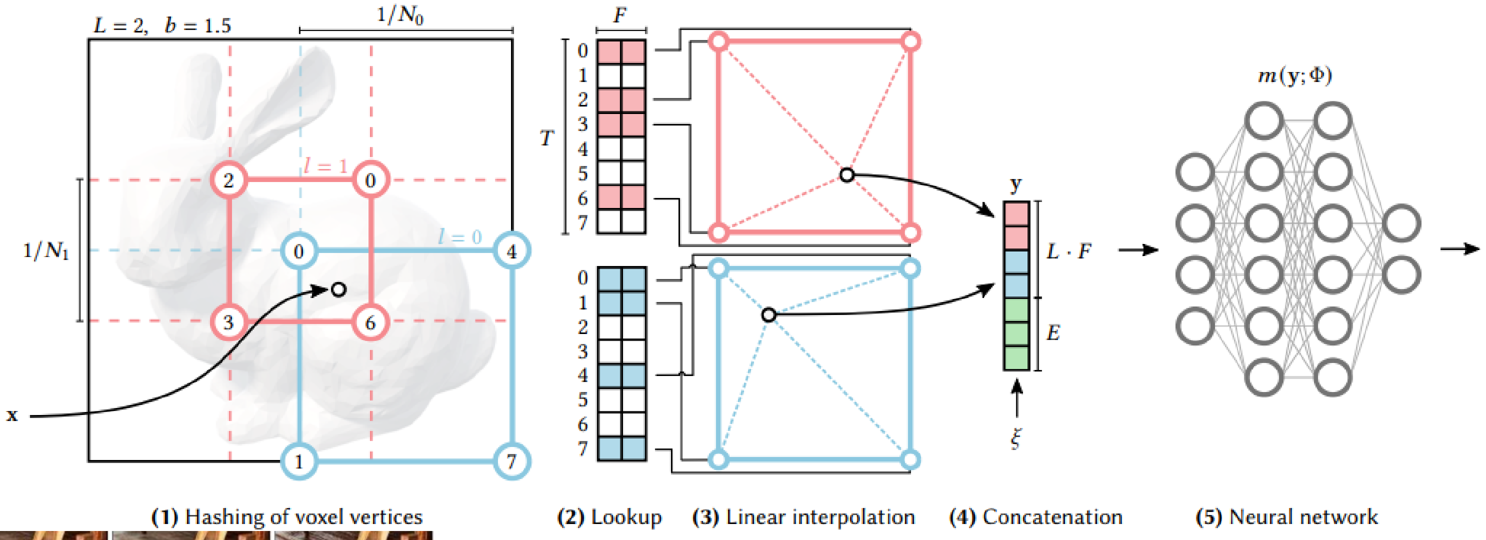
Instant Neural Graphics Primitives (2022)

Instant Neural Graphics Primitives with a Multiresolution Hash Encoding

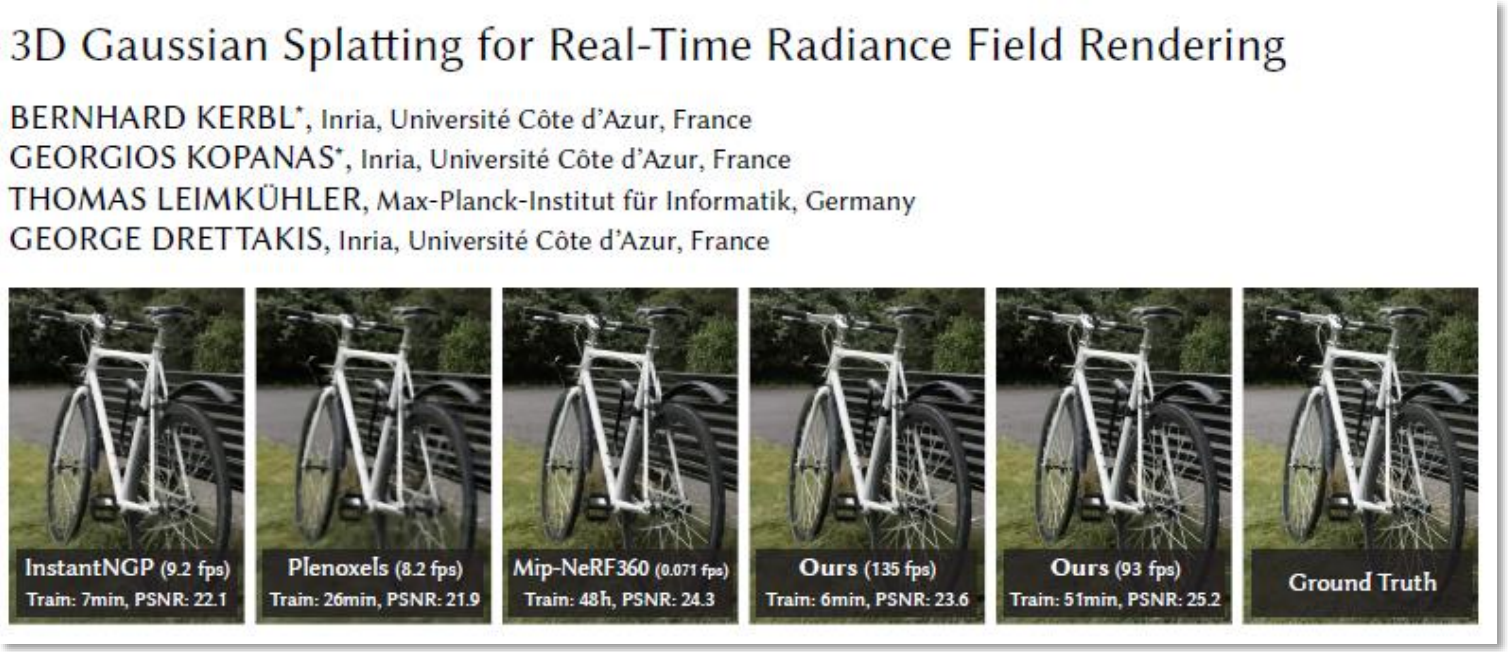
THOMAS MÜLLER, NVIDIA, Switzerland
ALEX EVANS, NVIDIA, United Kingdom
CHRISTOPH SCHIED, NVIDIA, USA
ALEXANDER KELLER, NVIDIA, Germany

https://nvlabs.github.io/instant-ngp

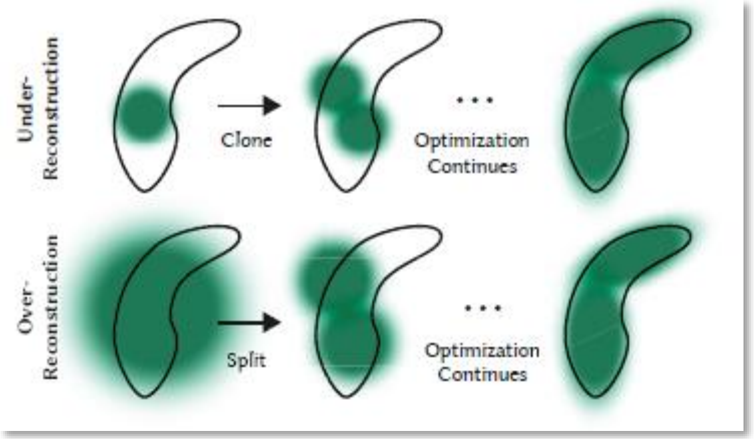


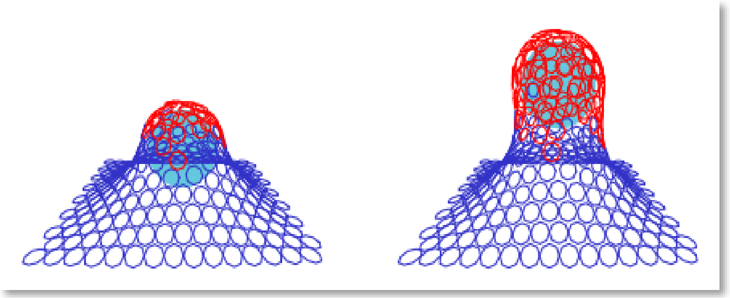


3D Gaussian Splatting (2023)

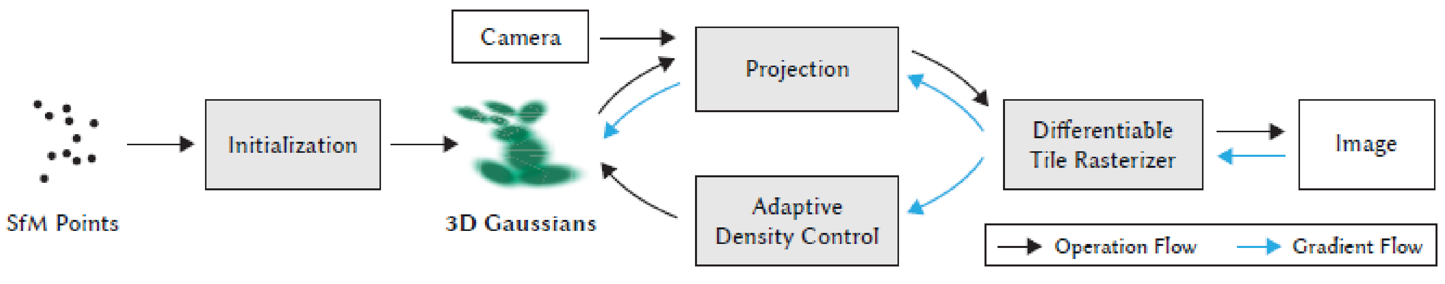

3D Gaussian Splatting for Real-Time Radiance Field Rendering

BERNHARD KERBL*, Inria, Université Côte d'Azur, France
GEORGIOS KOPANAS*, Inria, Université Côte d'Azur, France
THOMAS LEIMKÜHLER, Max-Planck-Institut für Informatik, Germany
GEORGE DRETTAKIS, Inria, Université Côte d'Azur, France





[Szeliski & Tonnesen, SG'92]



SMERF: Streamable … Radiance Fields

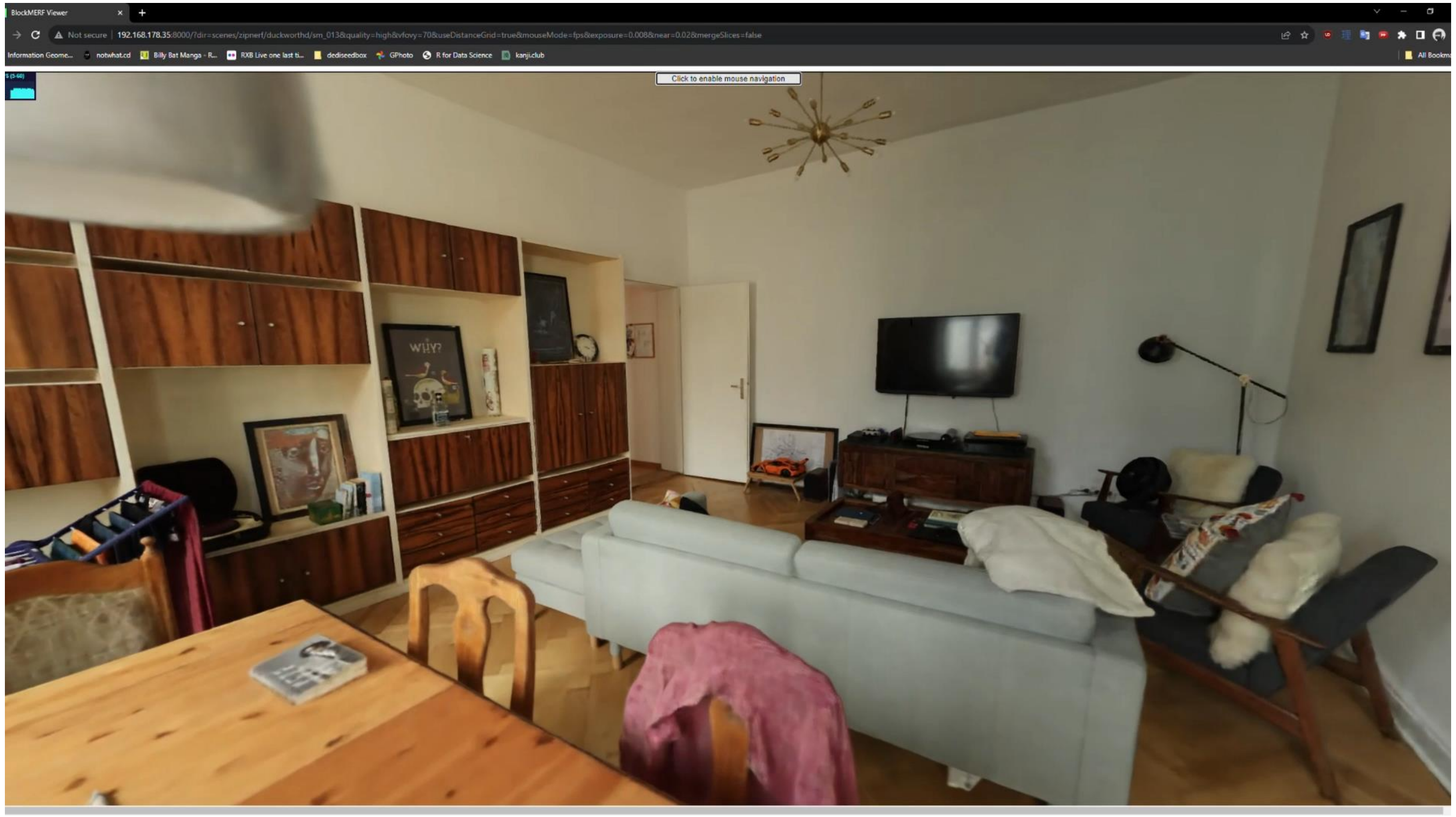
SMERF: Streamable Memory Efficient Radiance Fields for Real-Time Large-Scene Exploration

Daniel Duckworth[1*] Peter Hedman[2*] Christian Reiser[2,4,5] Peter Zhizhin[2]
Jean-François Thibert[3] Mario Lučić[1] Richard Szeliski[2] Jon Barron[2]

[1] Google DeepMind [2] Google Research [3] Google Inc.
[4] Tübingen AI Center [5] University of Tübingen

SMERF: Streamable … Radiance Fields

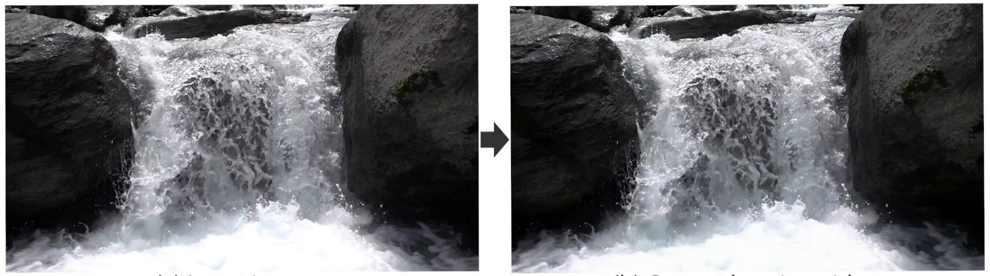
SMERF: Streamable Memory Efficient Radiance Fields for Real-Time Large-Scale Exploration



… wrapping up …

Outline

- Image-Based Rendering
  - Lumigraphs, Light Fields, Sprites with Depth, and Layers
- Virtual Viewpoint Video
- 360° and 3D Video
- 3D Photos
- Reflections and transparency
- Neural rendering



(a) Input image (b) Output looping video

3D for Image-Based Rendering & Novel View Synth.

- Many real-world, highly used applications in:
  - Video Processing
  - Computational Photography
  - Virtual Reality
- Many remaining challenges
  - Specularities and reflections
  - Textureless and thin object

Thank you