

Video for Virtual Reality

SIGGRAPH 2017 Course

Christian Richardt

University of Bath

James Tompkin

Brown University

Jordan Halsey

VR Playhouse

Aaron Hertzmann

Adobe Research

Jonathan Starck

Foundry

Oliver Wang

Adobe Research

Course website with slides:

<http://richardt.name/pub/Video4VR/>

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author. Copyright is held by the owner/author(s).

SIGGRAPH 2017 Courses, July 30 – August 3, 2017, Los Angeles, CA, USA

ACM 978-1-4503-5014-3/17/07

<https://doi.org/10.1145/3084873.3084894>

ABSTRACT

Video can capture the dynamic appearance of the real world in a way no other technology does; virtual reality technology, on the other hand, enables the display of dynamic visual content with unparalleled realism and immersion. The fusion of these two technologies—*video for virtual reality (VR)*—promises to enable many exciting photo-realistic experiences.

Over half a day, this course will provide an overview of three aspects of this exciting medium: the technical foundations, current systems in practice, and the potential for future systems of VR video.

In the first section, we will explore the geometric and optical problems underpinning VR video. Then, we will introduce both 360 degree video and stereoscopic video, including how 360 video is captured, analyzed, and stitched, including the mathematics behind how stereo 360 video can be captured.

This background material provides the prerequisites for understanding current systems in use. In the middle hour of the course, we explain how state-of-the-art stereo 360 video is produced from camera systems and computational processing. Then, we will consider the art of storytelling in VR, and how new tools for editing VR video can aid in the craft of this art production. Finally, this section provides an industry perspective covering current production and post-production choices and practice, including CG integration.

The final part of our course focuses on the next generation of video for VR, where we move to 6 degrees-of-freedom (6DoF) experiences. We introduce the basics and challenges behind light field cameras, processing and displays, and see how they can enable 6DoF experiences. This will be followed by another industry perspective on how light field camera arrays have been used to create cutting-edge experiences integrating volumetric live-action elements. To conclude the course, we will see how far we still must go toward the ideal system, in hopes of inspiring the attendees to push the boundary farther to reach it.

We hope this course is useful to a broad audience—at SIGGRAPH and beyond—as we cover the academic, artistic, and production sides of VR video.

COURSE HISTORY AND RELEVANT EXISTING COURSES

This is a new course on a topic that has so far not been covered at SIGGRAPH. While significant advances in video and virtual reality have been made in the past decade, less has been specifically written about this common use of VR technology. We hope to address this with our course.

Some aspects related to our course topics have been covered in previous SIGGRAPH (Asia) courses:

- [Human-centered design for VR interactions](#)¹ by Jason Jerald and Richard Marks (SIGGRAPH 2016) — Considered the perception of VR, including motion sickness and how it can be reduced in interactive VR experiences. While we will mention these issues, they will not be the focus of our course. Instead, we focus on the broader technologies, including (post-)production issues.
- [Augmented reality: principles and practice](#)² by Dieter Schmalstieg and Tobias Höllerer (SIGGRAPH 2016) — We will consider mostly VR technology. While the last ‘future’ part of our course will talk about volumetric live-action elements, which could be used in AR, we will focus more on techniques for production and post-production rather than viewing device fundamentals.
- [User-centric computational videography](#)³ by Christian Richardt, James Tompkin, Jiamin Bai, and Christian Theobalt (SIGGRAPH 2015) — This course covers mostly 2D video, and does not cover VR.
- [Video-based rendering](#)⁴ by Marcus Magnor and Marc Pollefeys (SIGGRAPH 2005) — Video-based rendering is used in video exploration systems, and has greatly improved in rendering quality in the 12 years since this course.

There is also a second set of courses which, while having some overlap, are less relevant to the topics covered in our course:

- [Put on your 3D glasses now: the past, present, and future of virtual and augmented reality](#)⁵ by Douglas Lanman, Henry Fuchs, Mark Mine, Ian McDowall, and Michael Abrash (SIGGRAPH 2014) — A comprehensive survey of VR display technology, with a strong focus on head-mounted displays. We will focus on content creation and processing, and only briefly discuss displays.
- [Computational plenoptic imaging](#)⁶ by Gordon Wetzstein, Ivo Ihrke, Douglas Lanman, Wolfgang Heidrich, Ramesh Raskar, and Kurt Akeley (SIGGRAPH 2012) — The course contains segments on light field capture, and “inside the Lytro camera”, but did not discuss light field video, or take a post-production viewpoint.
- [Lightfield photography: theory and methods](#)⁷ by Todor Georgiev and Andrew Lumsdaine (SIGGRAPH Asia 2009–2010) — The course presents the mathematical foundations of light field cameras. While this may be a useful reference for attendees of our course, and while we will cover some of the basics, we will focus more on the production aspects of video for VR.

¹<https://doi.org/10.1145/2897826.2927320>

²<https://doi.org/10.1145/2897826.2927365>

³<https://doi.org/10.1145/2776880.2792705>

⁴<https://doi.org/10.1145/1198555.1198678>

⁵<https://doi.org/10.1145/2614028.2628332>

⁶<https://doi.org/10.1145/2343483.2343494>

⁷<https://doi.org/10.1145/1900520.1900527>

COURSE SCOPE

In our course, we focus on the video aspect of virtual reality experiences, i.e. the visual component, and will tend not to discuss the audio component. As spatial audio is an important part of immersive VR experience, we feel it would be best discussed in its own right.

INTENDED AUDIENCE

In short: visual computing researchers, video producers, enthusiastic video users.

As VR and video cover a wide range of technologies, our course is targeted at a broad audience and contains information for many parties: enthusiastic video users and video professionals will discover cutting-edge VR video production techniques in the industry perspectives; academics will discover new research work (CHI 2017, EG 2017) and an overview of future 6DoF challenges; video producers will gain insight into the same techniques which may, in the near future, make it into production tools; and students will benefit from the background to current state-of-the-art systems and the problems currently being tackled to bring future VR video systems to use.

PREREQUISITES

A basic knowledge of video processing, computer graphics, and computer vision is useful. However, we expect people without this experience to still find the course engaging—the section on storytelling and tools, for example, will be accessible to all audiences.

LEVEL OF DIFFICULTY

Intermediate.

COURSE NOTES

We will make annotated slides available on our course website <http://richardt.name/pub/Video4VR/>.

COURSE PRESENTER INFORMATION

Christian Richardt — *University of Bath*
christian@richardt.name • <http://richardt.name>

Christian Richardt is a Lecturer (=assistant professor) at the University of Bath. He received a BA and PhD in Computer Science from the University of Cambridge in 2007 and 2012, respectively. He was previously a postdoctoral researcher at Inria Sophia Antipolis, Max Planck Institute for Informatics and the Intel Visual Computing Institute. His research combines insights from vision, graphics and perception to extract and reconstruct visual information from images and videos, to create high-quality visual results and experiences, with a focus on video processing for 360 degree videos, light fields, and for user-centric applications. He has co-organized the 2015 SIGGRAPH course on User-Centric Computational Videography.

James Tompkin — *Brown University*
james_tompkin@brown.edu • <http://www.jamestompkin.com>

James Tompkin is an assistant professor in Computer Science at Brown University. He received an MSci degree in Computer Science from King's College London in 2006, and an EngD degree in Virtual Environments, Imaging, and Visualization from University College London in 2012. He was a postdoctoral researcher at the Max-Planck-Institute for Informatics (2012–2014), and at the Harvard Paulson School of Engineering and Applied Science (2014–2016). His research applies vision, graphics, machine learning, and interaction to create new visual computing tools and experiences. He has previously co-organized the 2015 SIGGRAPH course on User-Centric Computational Videography.

Jordan Halsey — *VR Playhouse*
jordan@vrplayhouse.com • <http://vrplayhouse.com/>

Jordan Reece Halsey is a Los Angeles based video artist and the technical director for VR Playhouse. At VR Playhouse, he has designed and created virtual reality experiences for Time Life, Honda, Boost Mobile, Four Seasons, and many others. Jordan is an expert in projection mapping and has worked at the forefront of emerging technologies for many years, collaborating with companies like Google, Disney, and Ford; and with artists like Madonna, Miley Cyrus and A\$AP Rocky. Other notable projects include Amon Tobin's seminal Isam 2.0 tour, "Mothership 2.0" by Skrillex, and Michael Jackson's final tour and film, "This Is It".

Aaron Hertzmann — *Adobe Systems Incorporated*

hertzman@adobe.com • <http://www.dgp.toronto.edu/~hertzman/>

Aaron Hertzmann is a Principal Scientist at Adobe. He received a BA in Computer Science and Art/Art History from Rice University in 1996, and a PhD in Computer Science from New York University in 2001, respectively. He was a Professor at University of Toronto for ten years, and has also worked at Pixar Animation Studios, University of Washington, Microsoft Research, Mitsubishi Electric Research Lab, Interval Research Corporation and NEC Research. He is an associate editor for ACM Transactions on Graphics. His awards include the MIT TR100 (2004), a Sloan Foundation Fellowship (2006), a Microsoft New Faculty Fellowship (2006), the CACS/AIC Outstanding Young CS Researcher Award (2010), and the Steacie Prize for Natural Sciences (2010). He has spoken at four SIGGRAPH courses (most recently, the 2009 course on Human Body Movement), including one he organized (2004 course on Machine Learning).

Jonathan Starck — *Foundry*

jonathan.starck@foundry.com • <https://www.foundry.com/>

Jonathan Starck is Head of Research at Foundry, a global developer of computer graphics, visual effects and 3D modelling software for the design, visualization, and entertainment industries. He received a MEng in Engineering from the University of Cambridge in 1996 and a PhD in Computer Vision from University of Surrey in 2003. His research focuses on the development of image and video processing tools to accelerate artist workflows in high-end content creation. These technologies have created industry standard tools for movie post-production (NukeX and Ocula). Foundry has recently released a new toolset (CaraVR) to assist artists in the creation of high quality live-action VR content in post production, and are also exploring the video processing pipelines for immersive 6DoF VR experiences.

Oliver Wang — *Adobe Systems Incorporated*

owang@adobe.com • <http://www.oliverwang.info>

Oliver Wang is a Senior Research Scientist at Adobe. He received his PhD in Computer Science in 2010 from the University of California, Santa Cruz. Since then, he has worked in a number of research positions at HP Labs, Industrial Light and Magic and the Max Planck Institut Informatik, and a six year stint at Disney Research in Zurich. During this time his work led to a number of patents, and has been integrated into the production process of several mainstream movies. His research broadly spans the areas of image and video processing, computer vision, and machine learning.

SCHEDULE

—— INTRODUCTION AND BACKGROUND (ABOUT 15 MINUTES) ——

1. Welcome and Introduction

James Tompkin, Brown University, 15 minutes

- Motivation: why care about video in VR?
- Overview: outline of course topics.
- Brief primer on human audio/visual perception.
- Brief history of VR video.

—— TECHNICAL FOUNDATIONS (ABOUT 55 MINUTES) ——

2. 360 Spherical Video

Oliver Wang, Adobe, 20 minutes

- 360 video formats / spherical projection choices.
- Content creation: Fisheye/multicam setups.
- Stitching: Structured/unstructured (panoramic video, handheld), mesh warping (Rich360).
- Analysis: Stabilization, recinematography, pano2vid, saliency (where do people look?), highlight detection.

3. Stereoscopic 3D Videos and Panoramas

Christian Richardt, University of Bath, 20 minutes

- Capturing and displaying stereoscopic 3D videos.
- Viewing comfort: accommodation–vergence conflict & comfort zone.
- Editing stereo 3D videos: StereoBrush, de-anaglyph etc.
- Megastereo: creating high-quality stereoscopic panoramas.

4. Q&A followed by a short break

15 minutes

—— CURRENT PRACTICE (ABOUT 75 MINUTES) ——

5. Industry Perspective: 360 Video Cameras in Production

Jonathan Starck, Foundry, 20 minutes

- Introduction to 360 video production: stories and experience.
- Camera rigs: mono, stereo, Nokia OZO, Google Jump and Facebook Surround 360.
- Challenges in production and the impact on stitching in post-production.

6. Art, Storytelling, and Tools

Aaron Hertzmann, Adobe, 20 minutes

- History of visual storytelling media.
- VR video as an art form: special issues, needs, and artistic considerations.
- Editing VR footage paper highlight:
[Vremiere: In-headset Virtual Reality Video Editing](#)⁸ (CHI 2017)

7. Industry Perspective: Post-production Pipelines for 360 Video

Jonathan Starck, Foundry, 20 minutes

- Introduction to post-production: node-based compositing.
- 360 stitching pipelines: setting up rigs and stitching.
- Challenges in post: clean up and CG integration.

8. Q&A followed by a short break

15 minutes

—— THE FUTURE (ABOUT 40 MINUTES) ——

9. Light Field Video Basics and Challenges

James Tompkin, Brown University, 20 minutes

- Introduction to light fields: What, why, and how?
- Capturing light fields: state-of-the-art cameras and future potential.
- Processing light fields: filtering⁹ (EG 2017), editing, challenges ahead.
- Viewing light fields with virtual and augmented reality displays.

10. Industry Perspective: Light Field Production and Post Production

Jordan Halsey, VR Playhouse, 20 minutes

- Developing immersive experiences using light field camera arrays.
- Connecting Google Jump to post-production workflows.
- Real-time experiences using volumetric live-action elements.

—— CONCLUSION (10 MINUTES) ——

11. Closing and Q&A

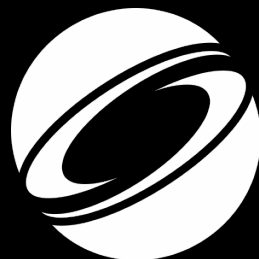
All, 10 minutes

- Summary & outlook
- How far to go to the ideal system? (*James Tompkin*)
- Questions & answers

= TOTAL TIME OF 3 HOURS AND 15 MINUTES

⁸<http://web.cecs.pdx.edu/~fliu/project/vremiere/>

⁹<http://liris.cnrs.fr/~nbonneel/cameraarrays>

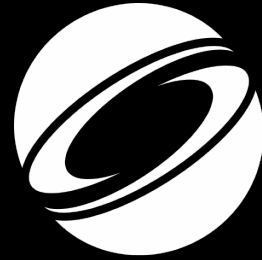


30 JULY – 3 AUGUST *Los Angeles*
SIGGRAPH2017

VIDEO FOR VIRTUAL REALITY

SIGGRAPH 2017 COURSE

<http://richardt.name/pub/Video4VR/>



30 JULY – 3 AUGUST *Los Angeles*
SIGGRAPH2017

This is a sample of the notes that we will present at the course. Unfortunately, some content could not be approved for release on time for the ACM Digital Library deadline, so please visit the course website for the complete version.

VIDEO FOR VR?



-
- Video captures real world dynamic appearance like no other technology.
 - Virtual reality enables the display of *dynamic* visual content with unparalleled immersion.
 - Video + VR is *telepresence*
 - Perceptual teleportation

WHY CARE ABOUT VIDEO IN VR?



- Depictions of the dynamic real world are compelling.
- Put yourself in someone else's shoes
- Storytelling
- Art
- Tourism
- Teleoperation



Imagine we want to look at objects
on a display as if we were looking
at them in real life.

What would we need?

PERCEPTUAL NEEDS



An eye



Variable focus

Two eyes



Binocular stereo
'3D' (*Not really 3D!*)

A neck and a body

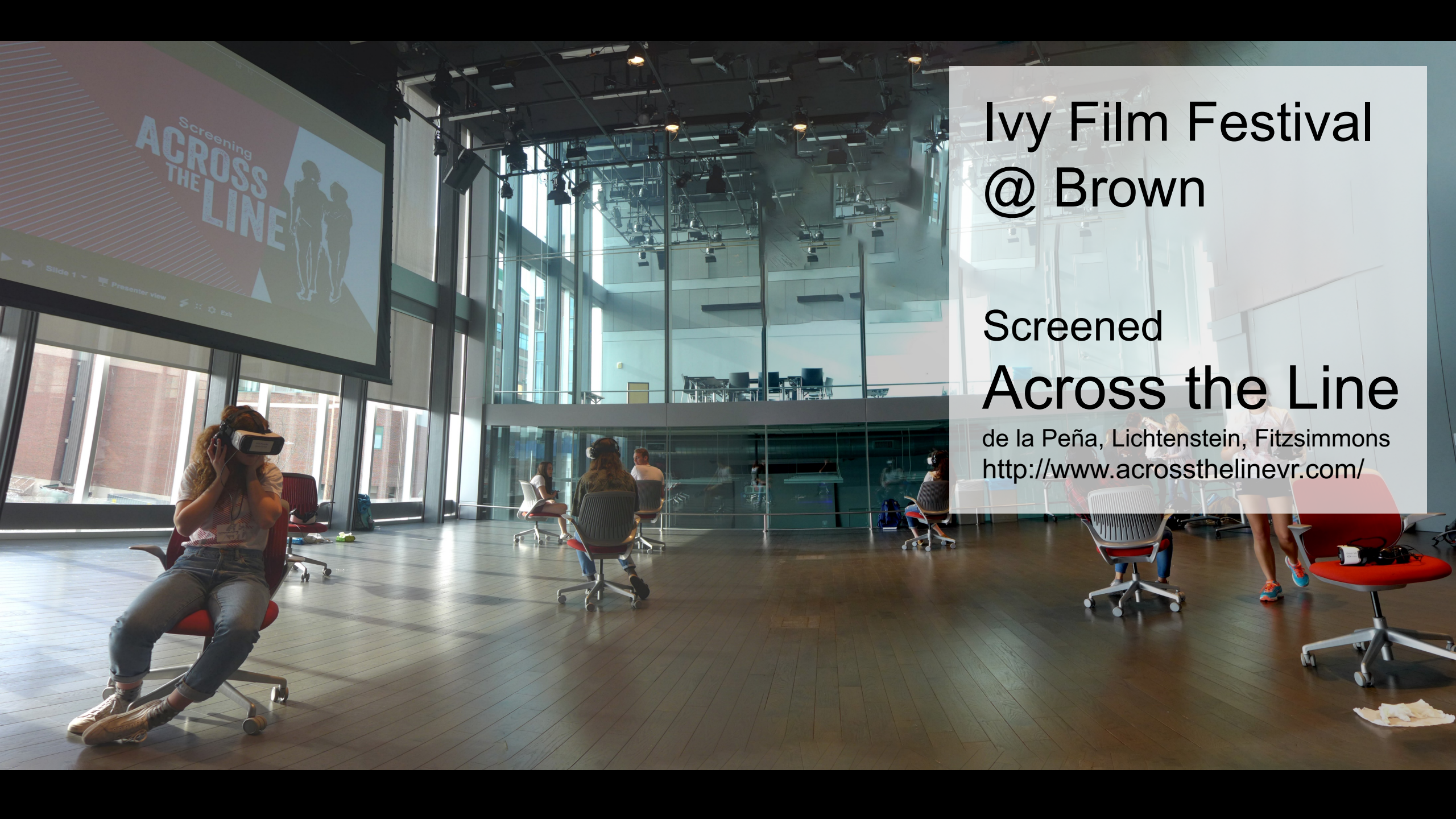


Motion parallax
'6DoF'

WHAT ARE WE COVERING?



- No audio (even though it is critical)
- No haptics / proprioception
- No gustation
- No olfaction - no *smell-o-vision*
- Just the visual sense – the literal video part
 - Plenty of exciting things here



Ivy Film Festival @ Brown

Screened **Across the Line**

de la Peña, Lichtenstein, Fitzsimmons
<http://www.acrossthelinevr.com/>



ADAM BLUMENTHAL



- *Virtual Reality Artist-in-Residence*
@ Brown's Granoff Center for the Arts
- Burning of the Gaspee
- Brown Alpert Medical School
 - VR mannequins, actors
 - Liberian hospital training
- 14 students in VR production.

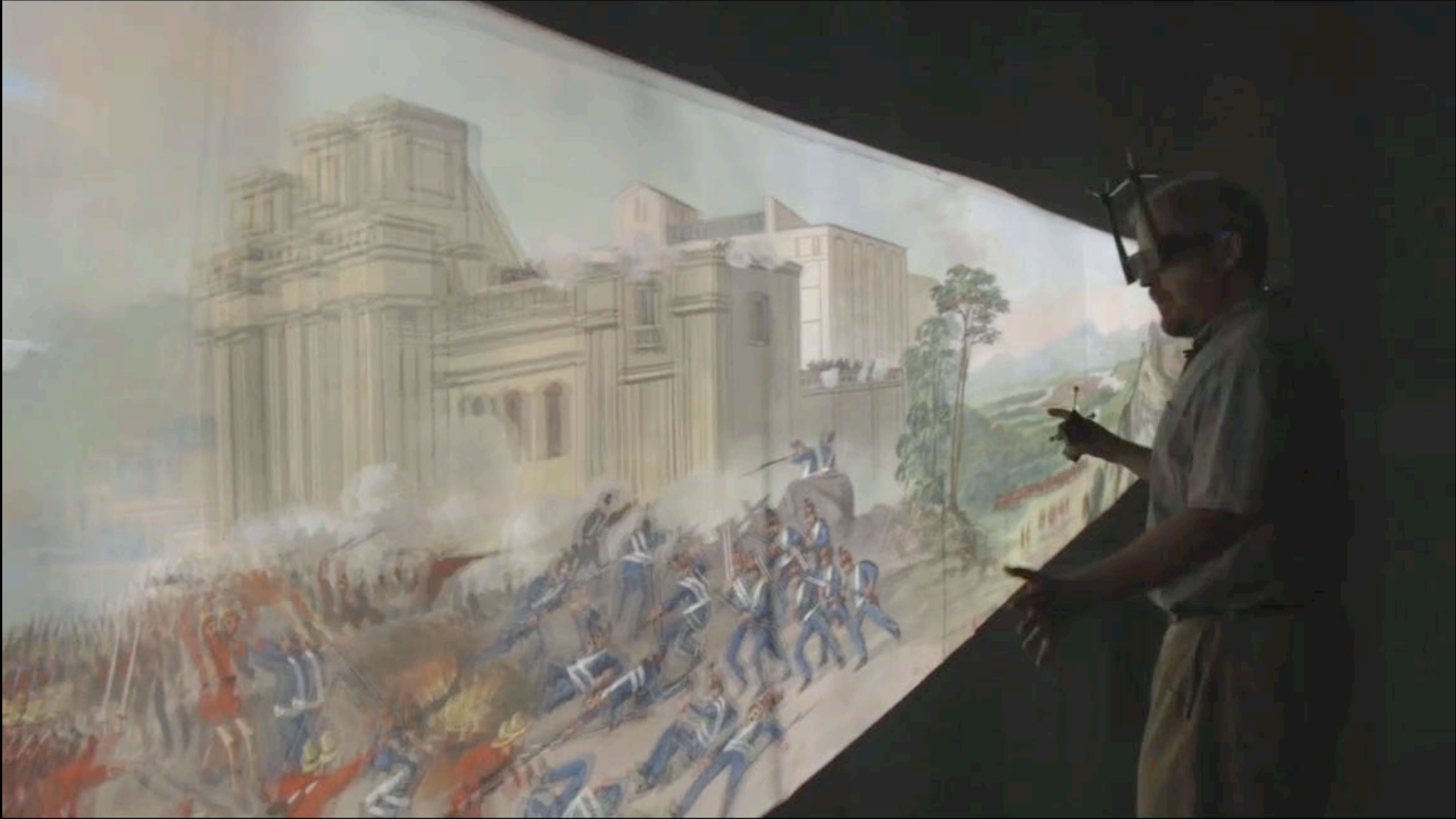




David Laidlaw

VR &
Scientific
Visualization





WHO ARE WE?



Aaron
Hertzmann



Oliver
Wang



Jon
Starck



Jordan
Halsey



Christian
Richardt



James
Tompkin



THE COURSE



- Introduction
- Technical Foundations
 - 2. 360° video background
 - 3. Stereo 3D videos and panoramas

2. 360° VIDEO



- 360° projections
- 360° cameras
 - Stitching
 - Processing
 - Analysis
- Future challenges



Oliver Wang



3. STEREO 3D VIDEOS & PANORAMAS



- Capturing and displaying stereo 3D videos
- Viewing comfort considerations
- Editing stereo 3D videos
- Creating stereo 3D panoramas



Christian
Richardt



THE COURSE



- Introduction
- Technical Foundations
 - 2. 360° video background
 - 3. Stereo 3D videos and panoramas
- Current Practice
 - 4. State-of-the-art cameras
 - 5. Industry post-production pipelines
 - 6. Art, storytelling, and tools

4. CURRENT CAMERAS



- Camera rigs for stereo 360 video capture.
- Stereoscopic 360 video stitching and video formats in production.
- The practical impact on stitching in post.
- Limitations of existing camera rigs.



Jon Starck
FOUNDRY.

5. POST PRODUCTION



- Introduction to live-action VR production.
- Overview of different camera systems in production.
- Post-production tools: node-based compositing.
- 360 stitching pipelines: setting up rigs and stitching.
- Challenges in post: clean up and CG integration.



Jon Starck
FOUNDRY.

6. ART, STORYTELLING, AND TOOLS



- History of visual storytelling media.
- VR video as an art form:
 - Special issues, needs, and artistic considerations.
- Editing VR footage paper highlight:
 - Vremiere: In-headset Virtual Reality Video Editing



Aaron
Hertzmann



THE COURSE



- Introduction
- Technical Foundations
 - 2. 360° video background
 - 3. Stereo 3D videos and panoramas
- Current Practice
 - 4. State-of-the-art cameras
 - 5. Industry post-production pipelines
 - 6. Art, storytelling, and tools
- Future 6DoF Systems
 - 7. Light fields background
 - 8. Light field production and post-production

7. LIGHT FIELD BASICS



- What, why, and how?
- Capturing light fields:
 - State-of-the-art cameras and future potential.
- Processing light fields:
 - filtering, editing, challenges ahead.



James
Tompkin



BROWN
Computer Science

8. LIGHT FIELD PRODUCTION



- Developing immersive experiences using light field camera arrays.
- Connecting Google JUMP to post-production workflows.
- Real-time experiences using volumetric live-action elements.



Jordan
Halsey

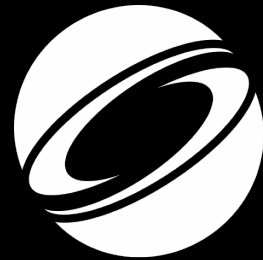
 **Playhouse**



THE COURSE



- Introduction
- Technical Foundations
- Current Practice
- Future 6DoF Systems
- Summary and outlook
 - How far do we have to go?
 - Q & A



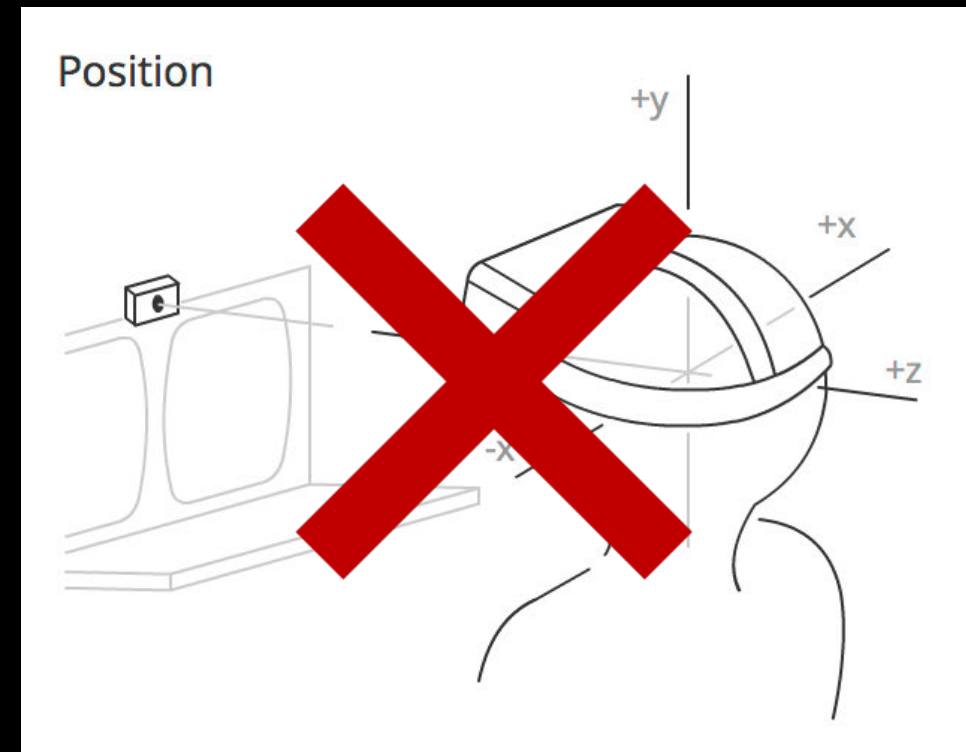
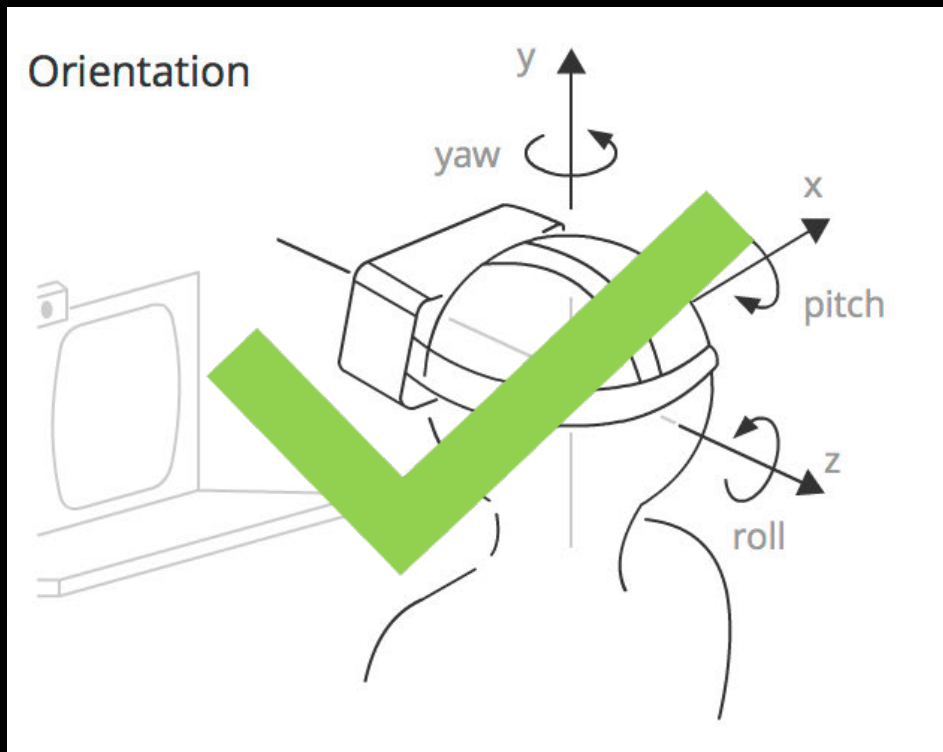
30 JULY – 3 AUGUST *Los Angeles*
SIGGRAPH2017

VIDEO FOR VIRTUAL REALITY
360° VIDEO

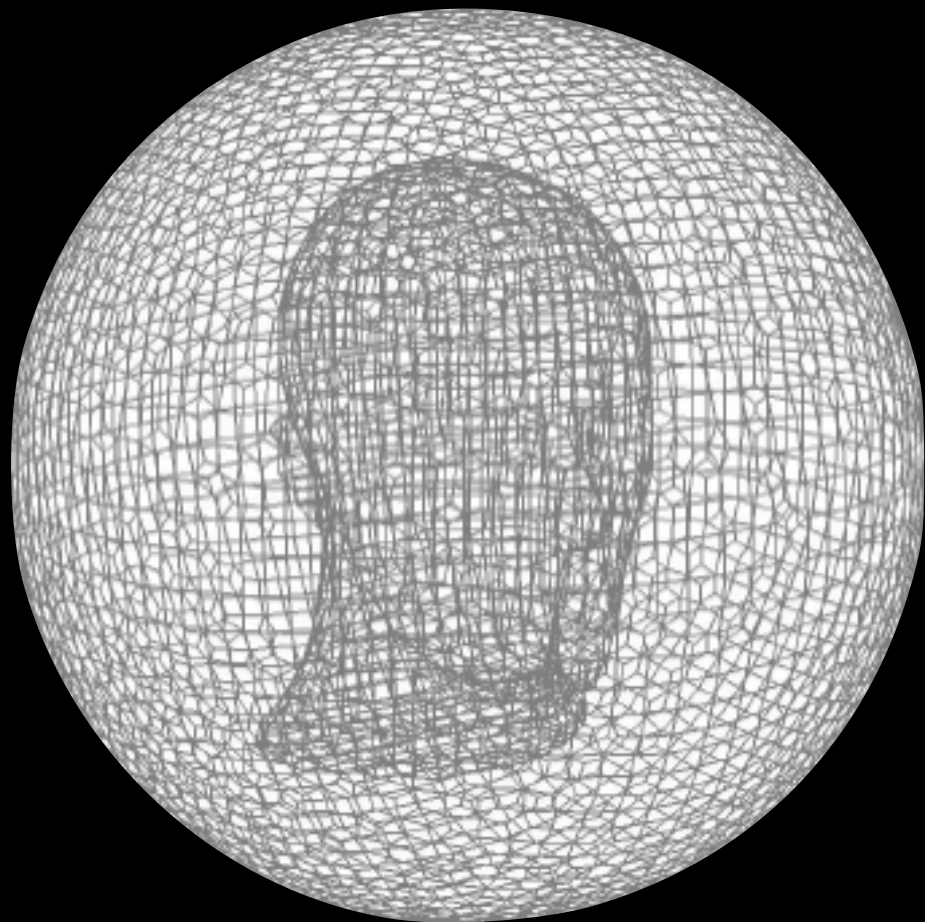
OLIVER WANG



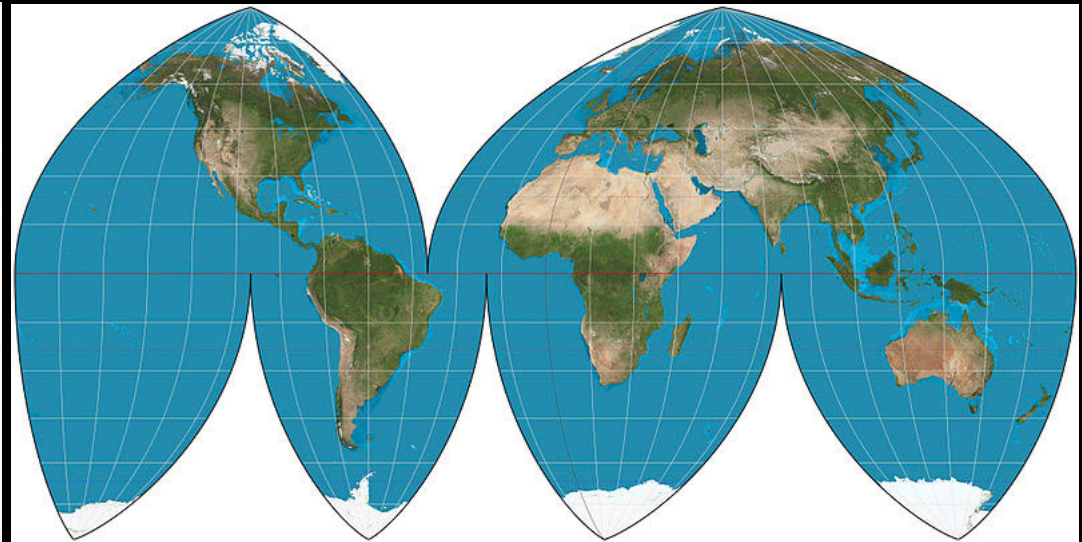
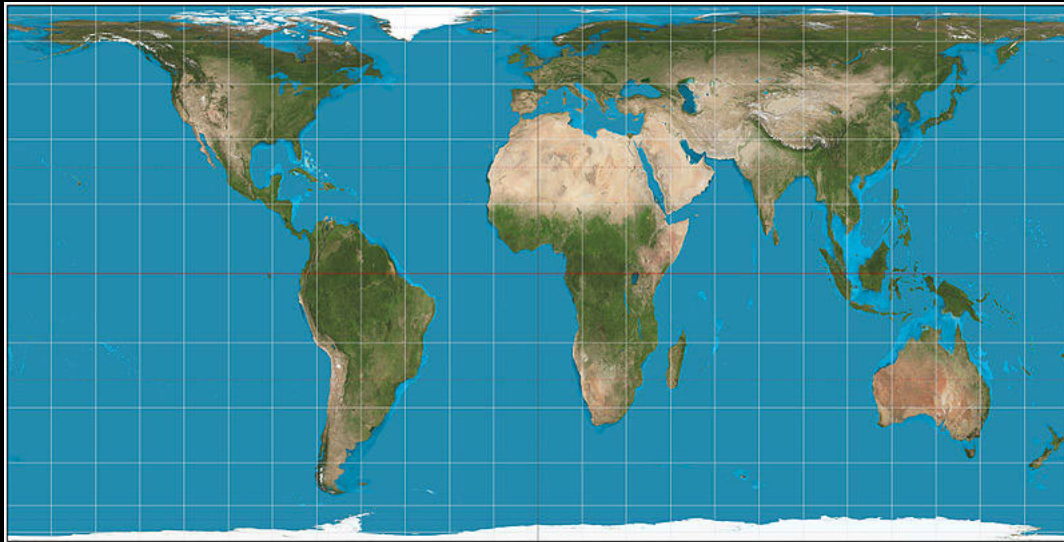
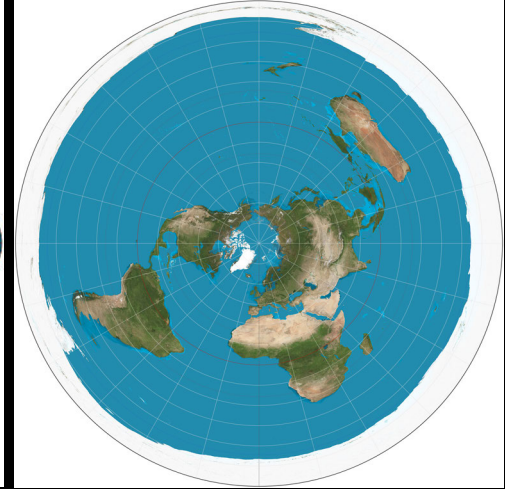
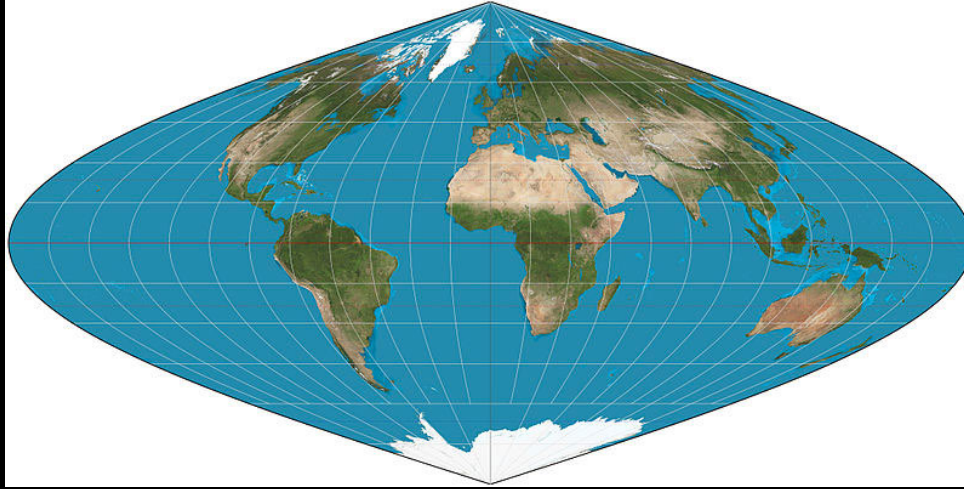
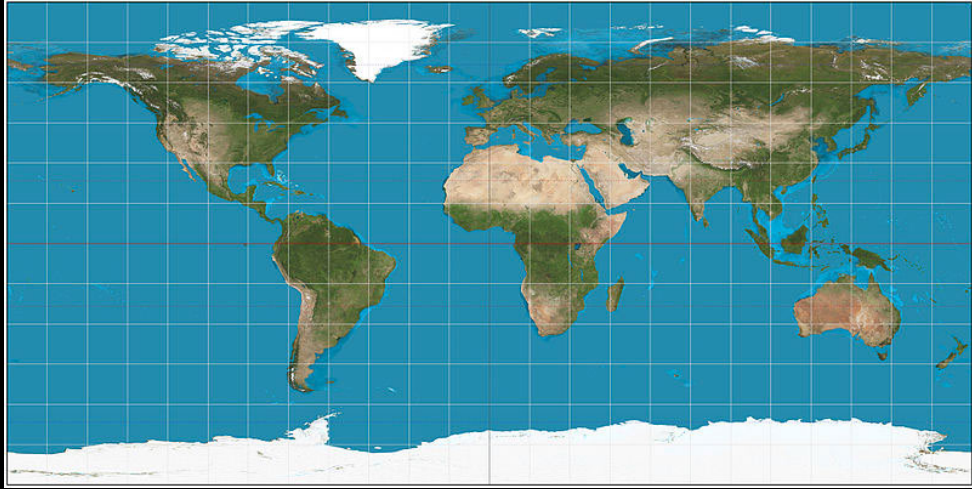
360° VIDEO



360° VIDEO FORMATS



SPHERICAL PROJECTION

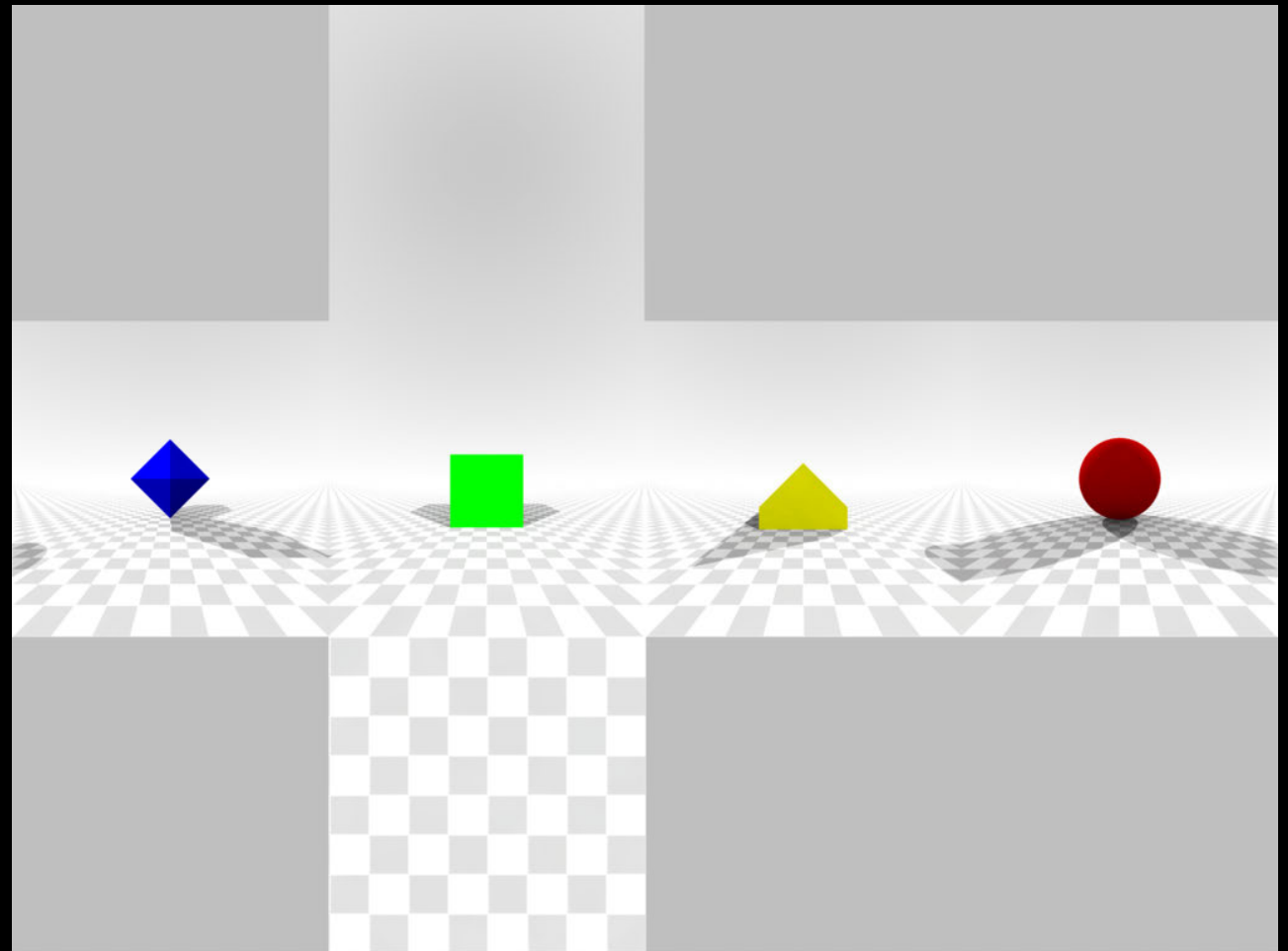
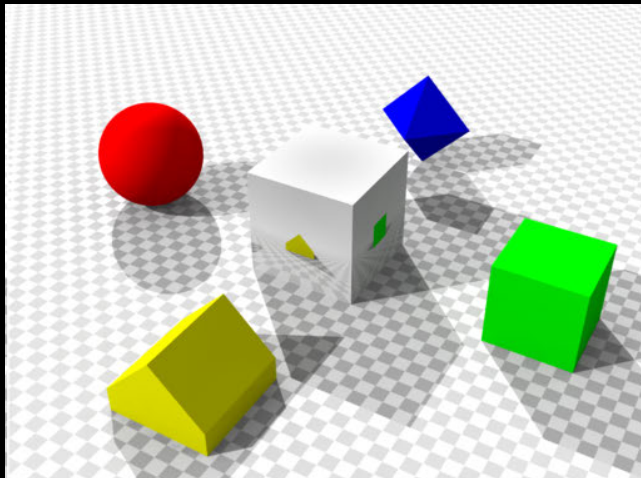
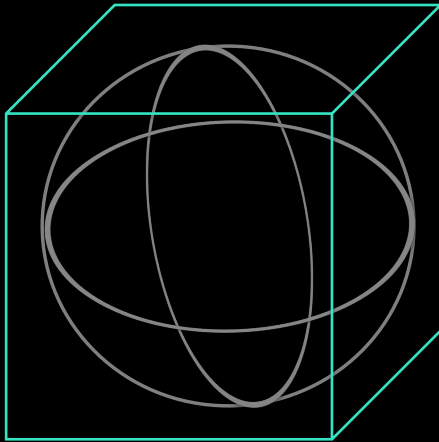


EFFICIENT STORAGE + STREAMING



- Could just store everything equirectangular
 - Inefficient storage at poles
 - Inefficient streaming as user can't look everywhere at once
- Alternatives:
 - Cube maps – 25% storage reduction over equirectangular
 - Google: equi-angular cube map
 - Facebook: multi-pyramid encoding

CUBE MAP

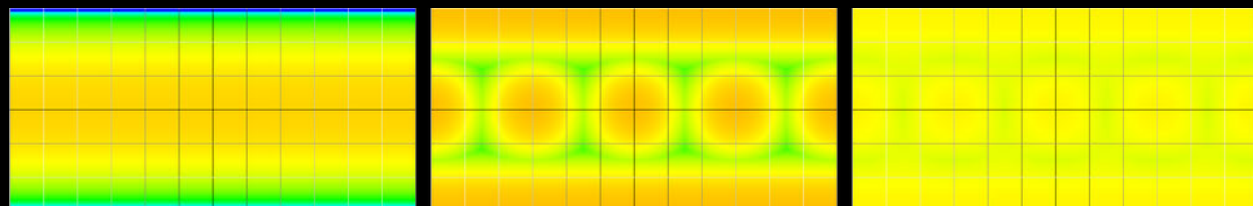


GOOGLE'S APPROACH



- Cube map is not equal angle per pixel – fix it.
- Angle density:

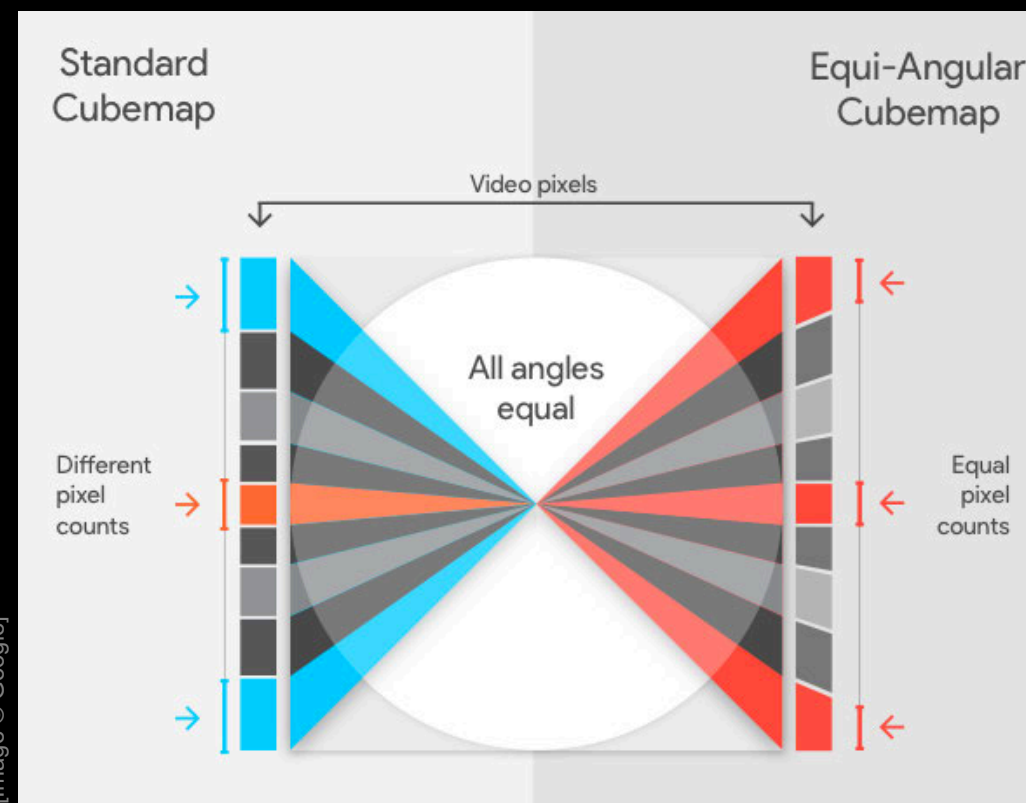
[Image © Google]



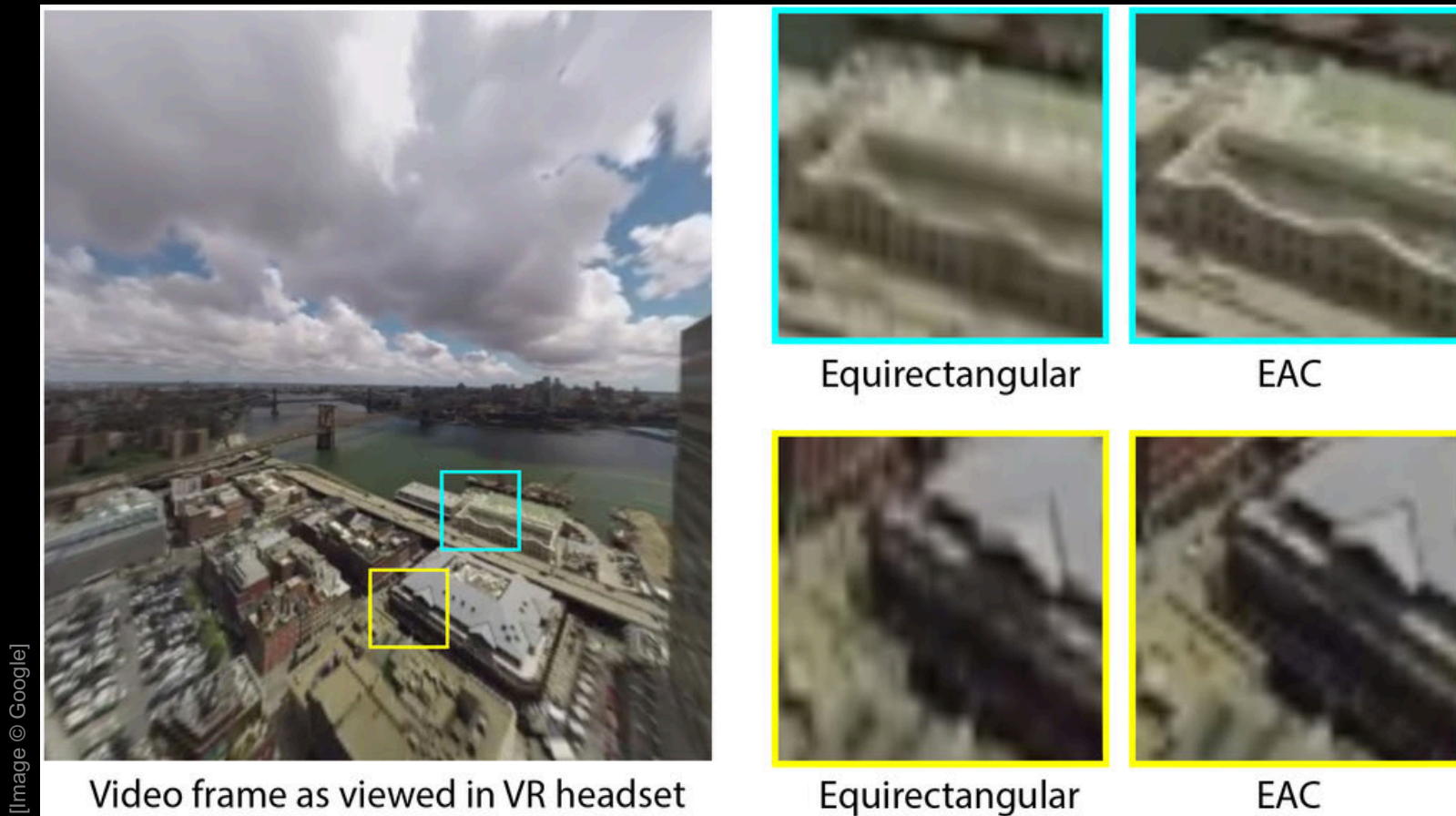
Equirectangular

Standard Cubemap

Equi-Angular Cubemap



GOOGLE'S APPROACH



FACEBOOK'S APPROACH

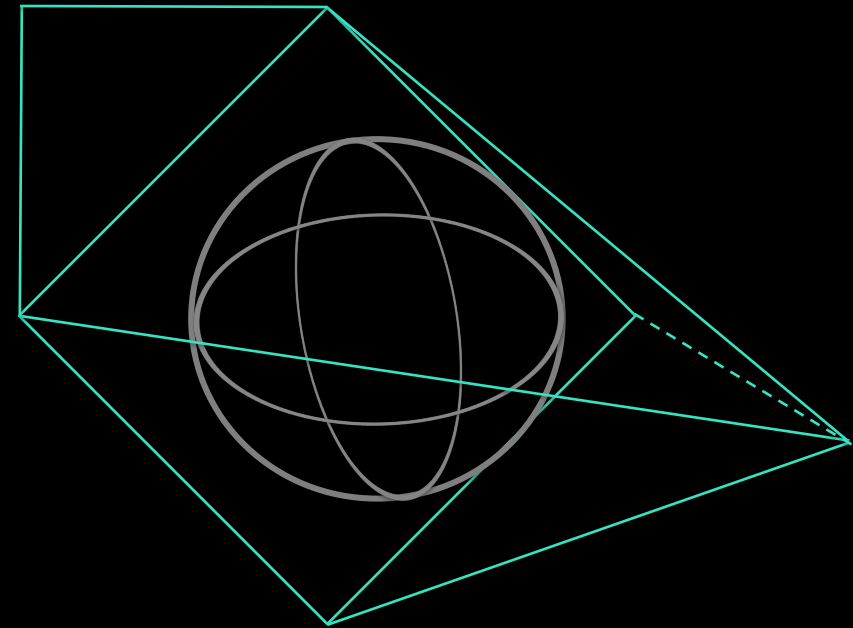
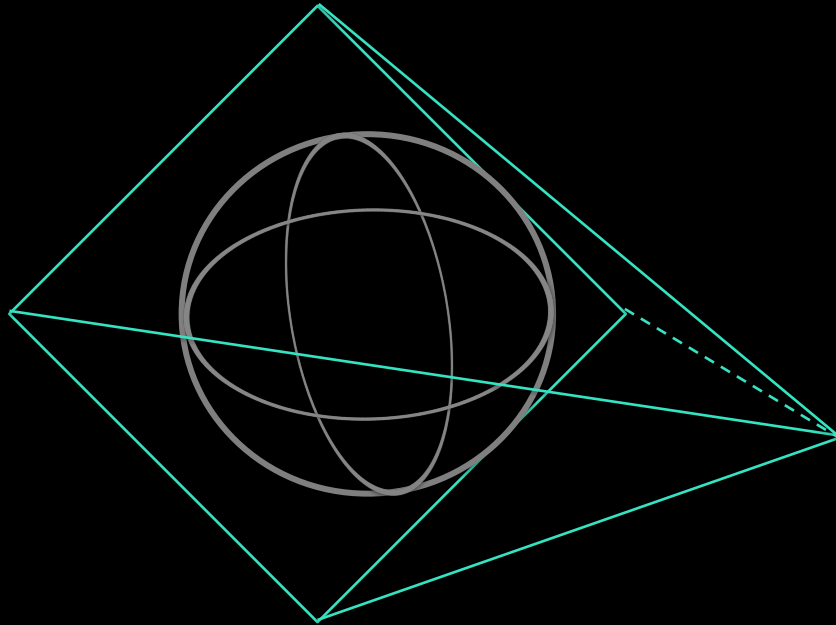


- Split up 360° into 30 directional streams
- Switch between directional streams at view time
 - Reduces bandwidth
- Format for stream
 - One full-res cube-map face + four quarter-res pyramid faces

FACEBOOK'S APPROACH



- Each of the 30 streams is one of these:

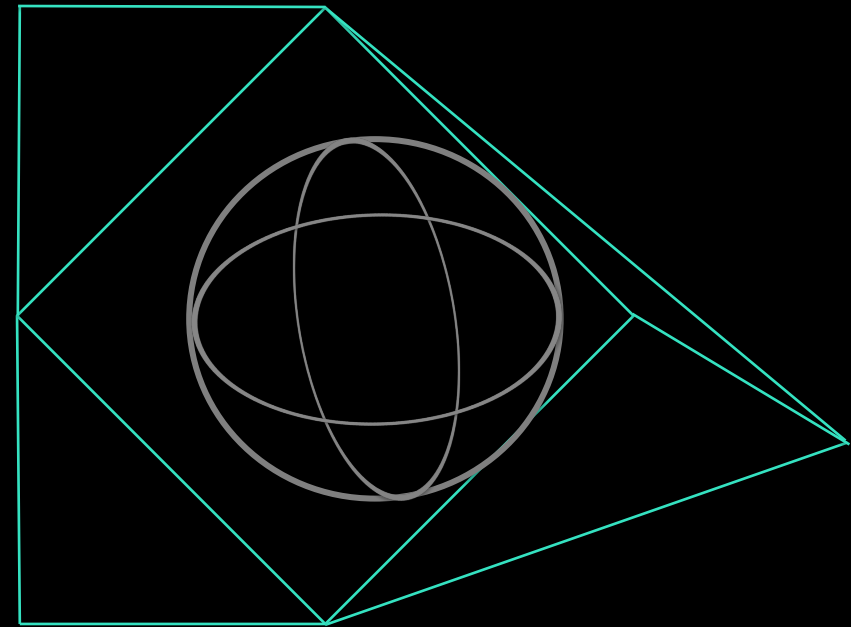
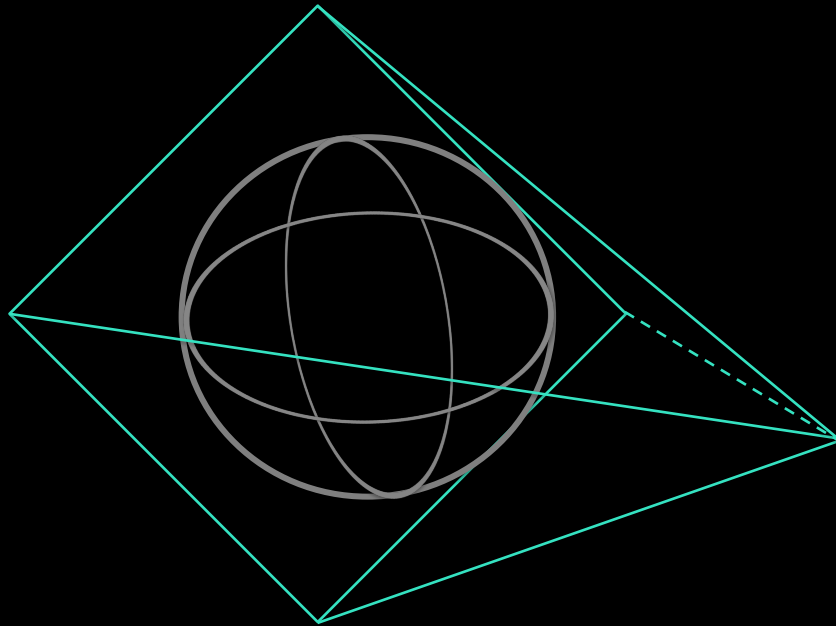


80% reduction in streaming bandwidth

FACEBOOK'S APPROACH



- Each of the 30 streams is one of these:

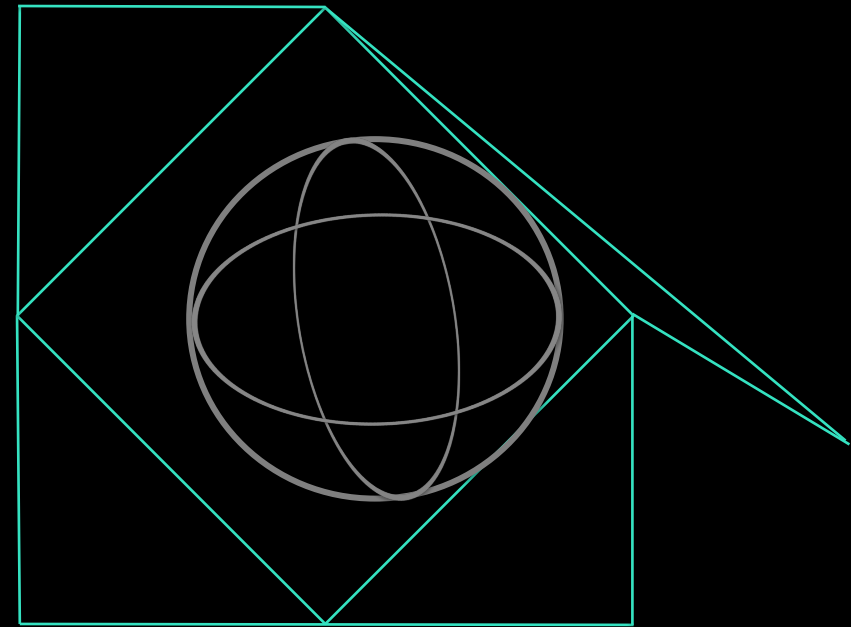
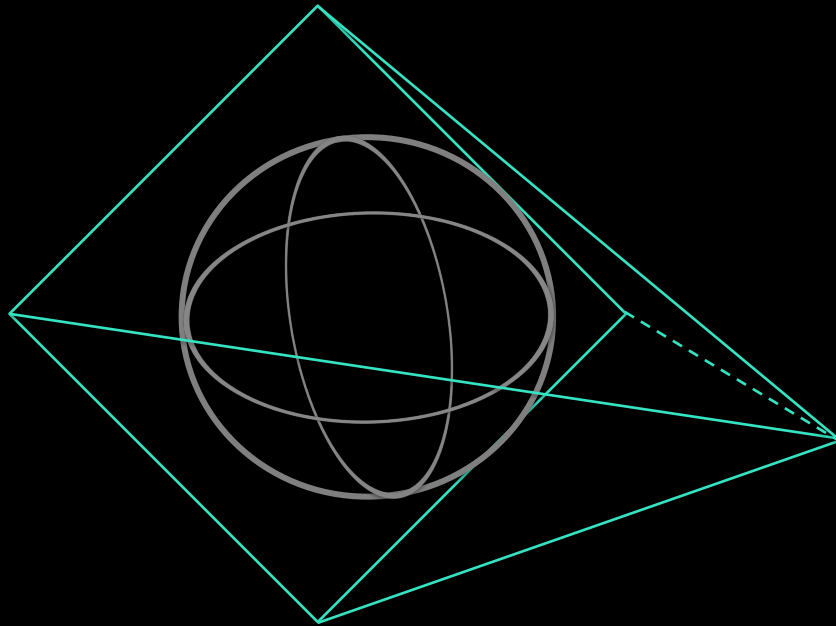


80% reduction in streaming bandwidth

FACEBOOK'S APPROACH



- Each of the 30 streams is one of these:

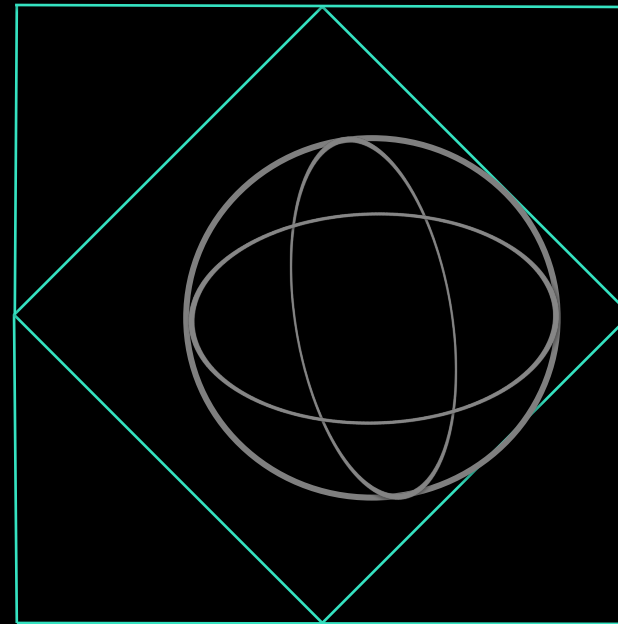
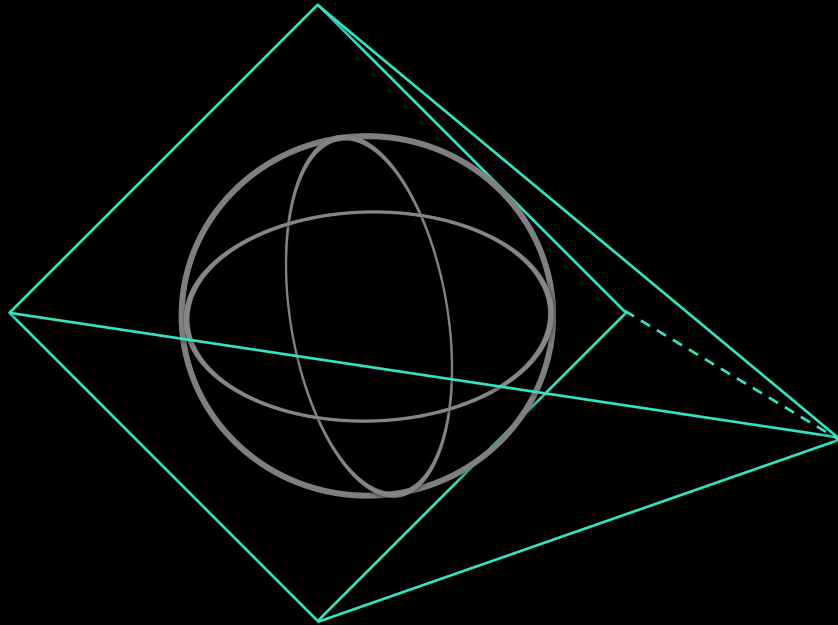


80% reduction in streaming bandwidth

FACEBOOK'S APPROACH



- Each of the 30 streams is one of these:



80% reduction in streaming bandwidth

360° VIDEO CONTENT CREATION



- Fisheye
 - Minimize number of seams
 - Resolution limited by sensor and angular density
- Multi-camera setups
 - Very high resolution possible
 - More complex stitching problem
 - Structured vs. unstructured

CONSUMER FISHEYE CAMERAS



Kodak SP360



Samsung Gear 360



Ricoh Theta S

MULTI-CAMERA SETUPS



- Cover sphere with many individual cameras
- Stitch together in post



Google Jump camera – Yi Halo from Yi Technology
<https://g.co/jump>

MULTI-CAMERA SETUPS



- Cover sphere with many individual cameras
- Stitch together in post
- Creative designs
 - Head mount

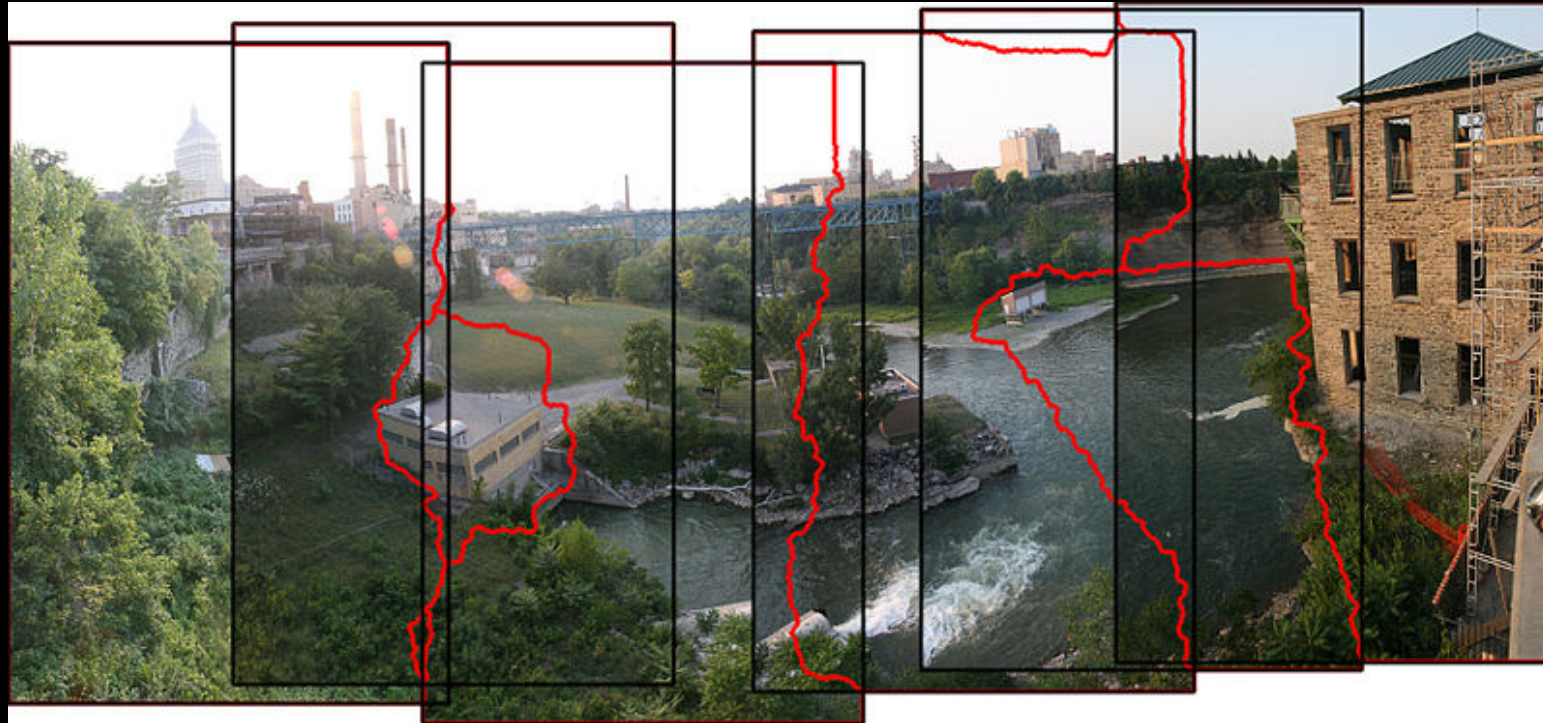


[Alessandro Dimas – Dimas Technologies GmbH
<http://www.360-degree-video.com/pages/360deg-video-helmet.php>]

STITCHING



- Panoramic stitching forefather to 360 degree video



STITCHING



-
- Panoramic stitching forefather to 360 degree video
 - Challenge of parallax

STITCHING



-
- Panoramic stitching forefather to 360 degree video
 - Challenge of parallax
 - Solutions
 - Optical flow based:
 - *Systems and Experiment Paper: Construction of Panoramic Image Mosaics with Global and Local Alignment* (Shum & Szeliski, IJCV 2000)
 - *Panoramic Video from Unstructured Camera Arrays* (Perazzi et al., Eurographics 2015)
 - Mesh warping based:
 - *Rich360: Optimized Spherical Representation from Structured Panoramic Camera Arrays* (Lee et al., SIGGRAPH 2016)

PROCESSING



-
- New challenges (and opportunities) for 360° video
 - *360° Video Stabilization* (Kopf, SIGGRAPH Asia 2016)
 - More background visible means robust motion estimation
 - Rotation stabilization only; no cropping required

PROCESSING

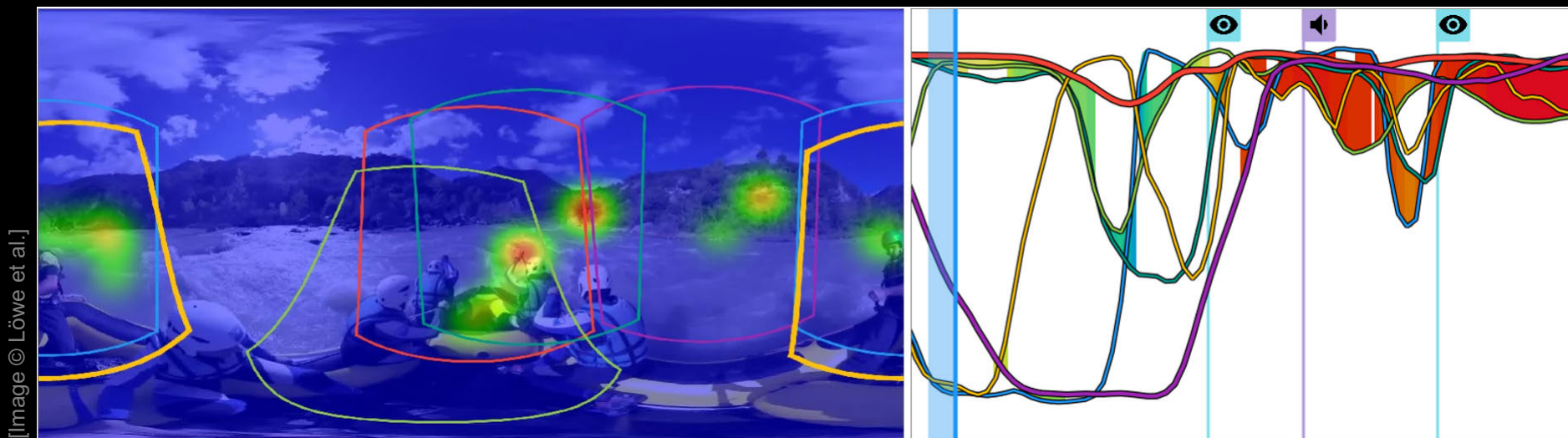


- 360° for traditional video
- Get everything at capture time
 - Highlight detection:
Making 360° Video Watchable in 2D: Learning Videography for Click Free Viewing
(Su & Grauman, CVPR 2017)
 - Hyperlapse:
Semantic-driven Generation of Hyperlapse from 360° Video (Lai et al., arXiv 2017, 1703.10798)

PROCESSING



- Where do people look?
- Cinematographic challenges
 - Have to learn new rules (YouTube Creators)
 - Saliency (Löwe)



[Image © Löwe et al.]

[Löwe et al., Visualization and Analysis of Head Movement and Gaze Data for Immersive Video in Head-mounted Displays, *Workshop on Eye Tracking and Visualization (ETVIS)*, October 2015.]

ANALYSIS



- Can take advantage of, and introduce direction back into 360 video
 - Recinematography opportunities

FUTURE PROBLEMS



- Cinematographic rules:
 - Cuts, Zooms?
- Perceptual studies: dizziness, nausea
- Tools: Compositing
- Spatial Audio



30 JULY – 3 AUGUST *Los Angeles*
SIGGRAPH2017

Christian Richardt

Stereoscopic 3D Videos and Panoramas



CAMERA

Centre for the Analysis of Motion,
Entertainment Research and Applications



UNIVERSITY OF
BATH

Stereoscopic 3D videos and panoramas

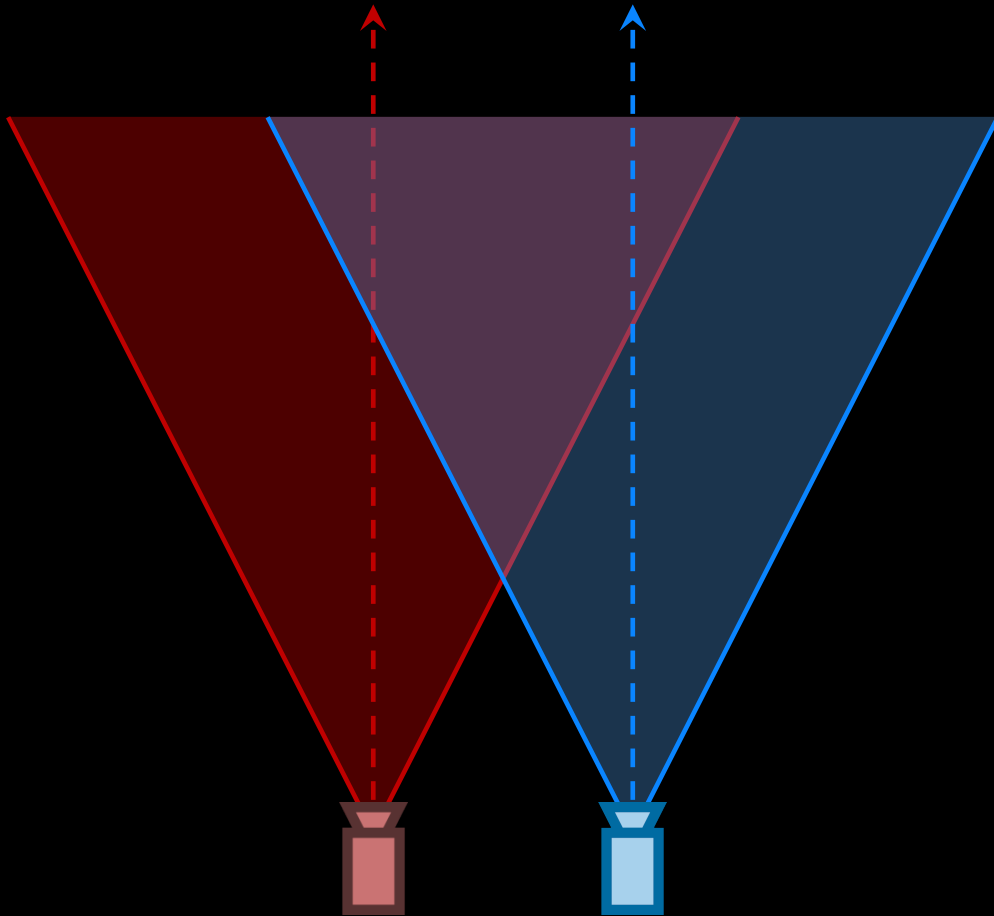
1. Capturing and displaying stereo 3D videos
2. Viewing comfort considerations
3. Editing stereo 3D videos
4. Creating stereo 3D panoramas

Content / Schedule

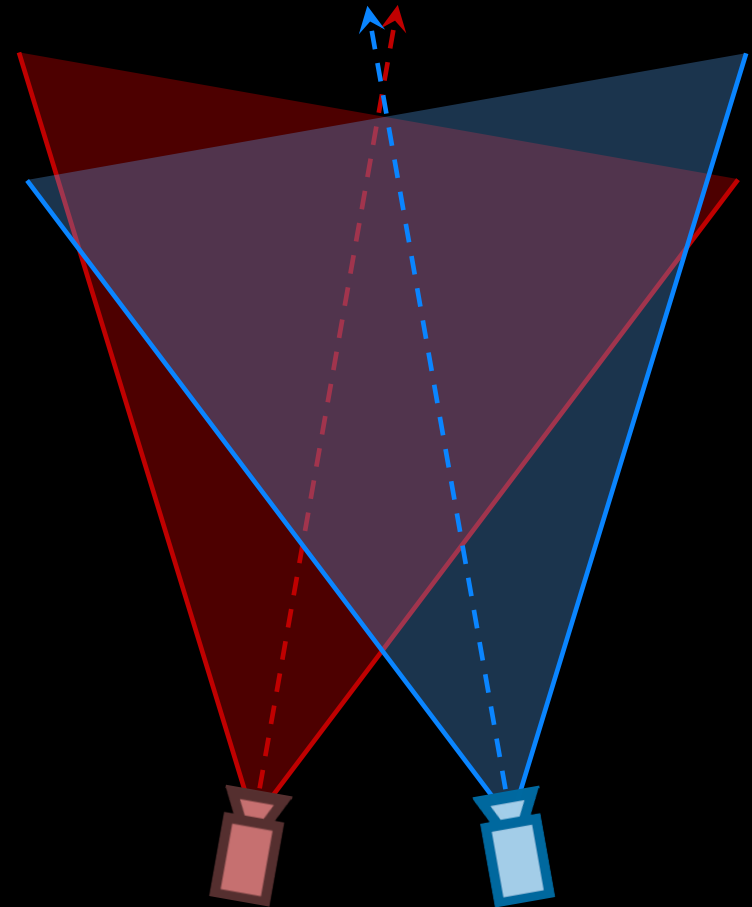
- 5–7 mins – Capturing and displaying stereo 3D videos
- 3–4 mins – Stereo viewing comfort (total: 8–11 mins)
- 4–6 mins – Editing stereo 3D videos (total: 12–17 mins)
- 6–8 mins – Creating stereo panoramas (total: 18–25 mins)

Stereo camera rigs

Parallel

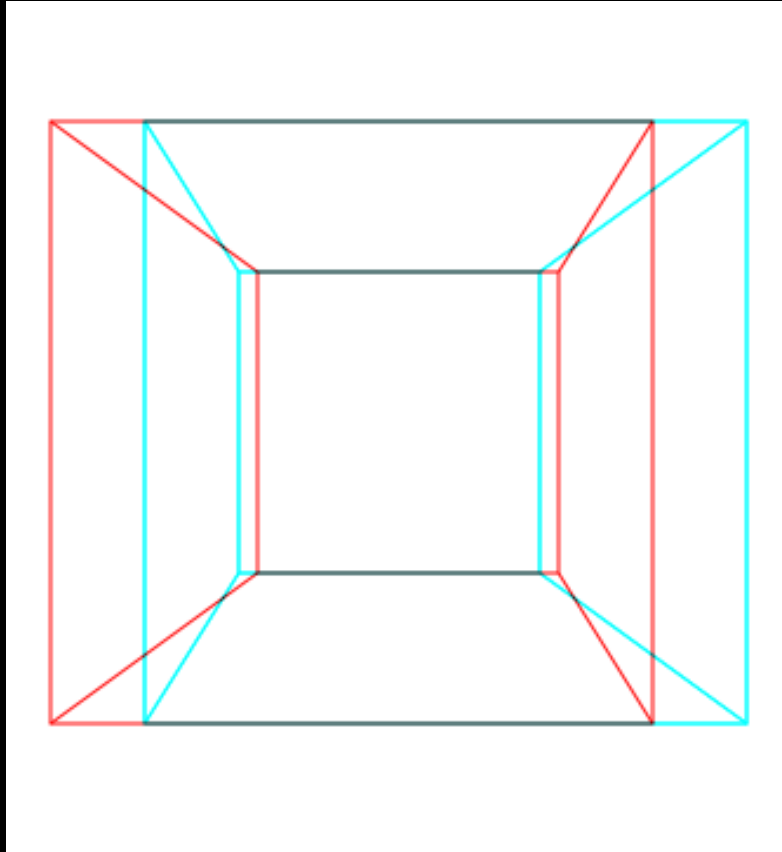


Converged



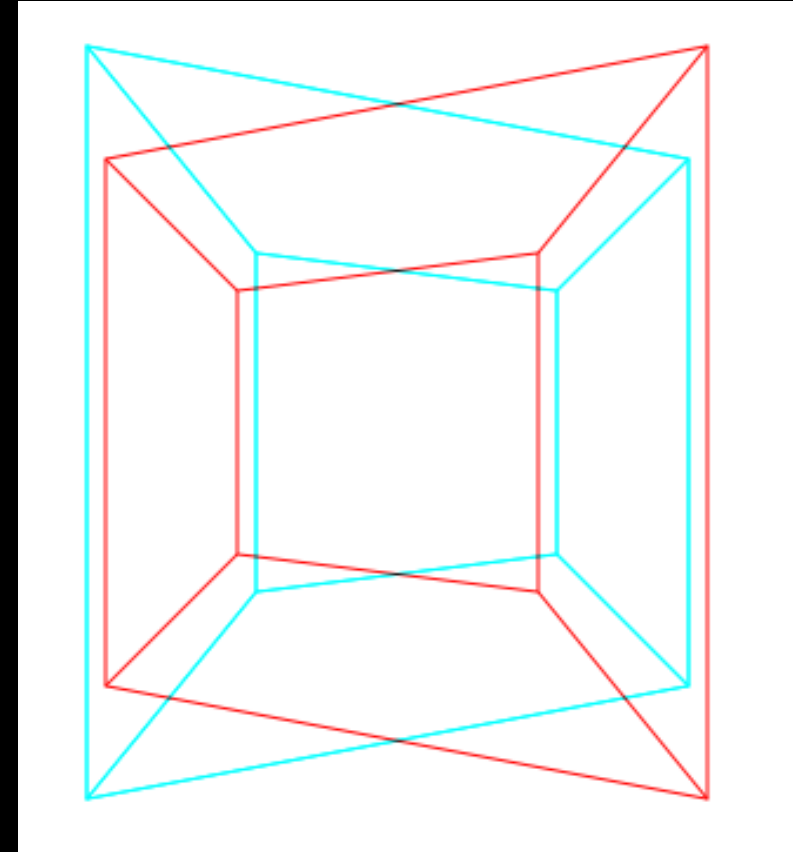
Stereo camera rigs

Parallel



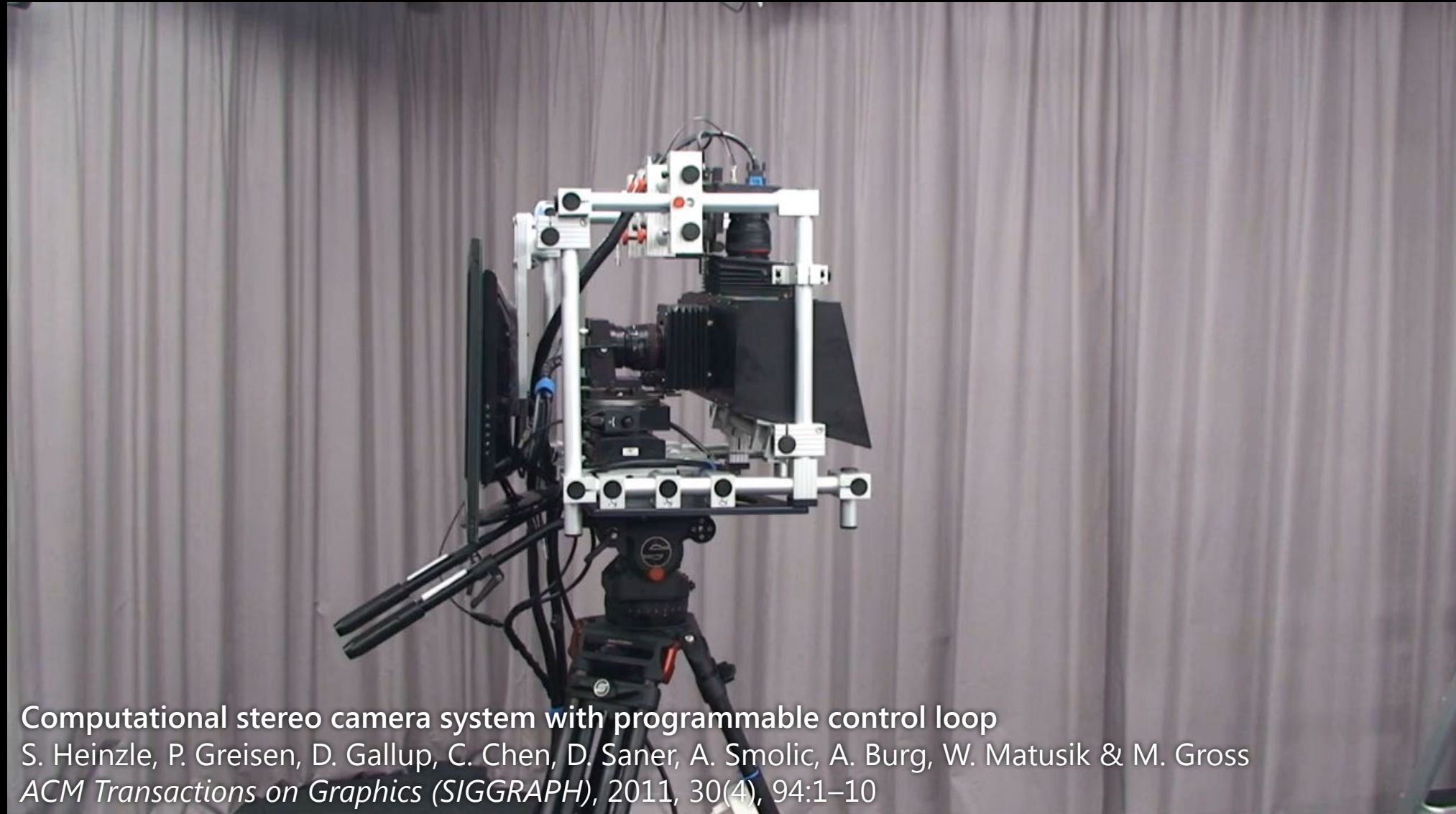
©2012 Oliver Kreylos

Converged



©2012 Oliver Kreylos

Computational stereo 3D camera system



Computational stereo camera system with programmable control loop

S. Heinzle, P. Greisen, D. Gallup, C. Chen, D. Saner, A. Smolic, A. Burg, W. Matusik & M. Gross

ACM Transactions on Graphics (SIGGRAPH), 2011, 30(4), 94:1–10

©2011 Heinzle et al./ACM

Commercial stereo 3D projection

Polarised projection

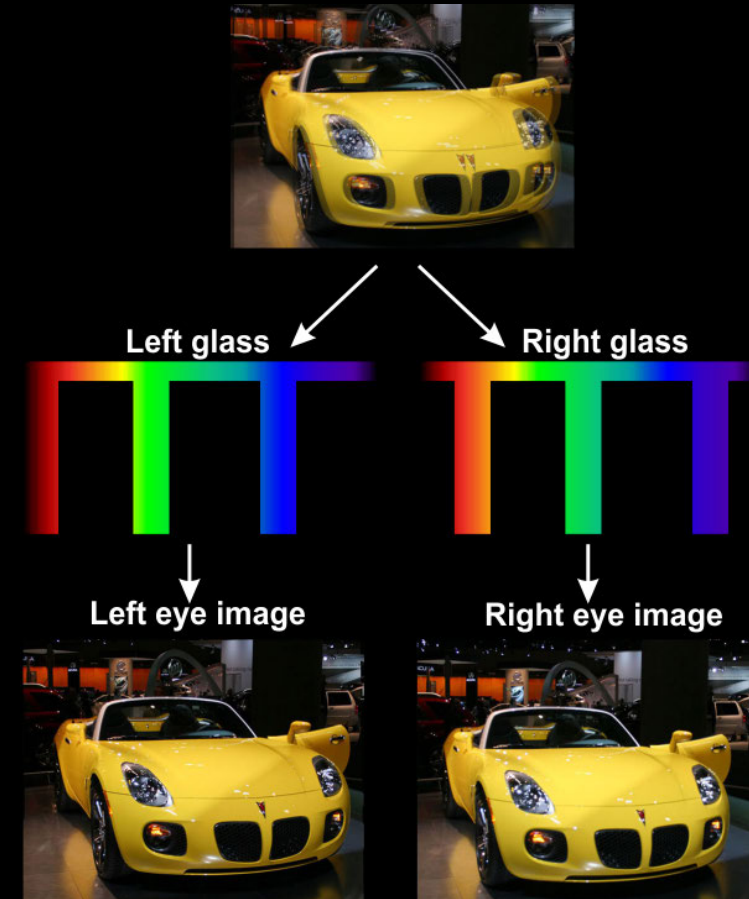
e.g. RealD 3D, MasterImage 3D



©2011 Scott Wilkinson/Sound and Vision

Wavelength multiplexing

e.g. Dolby 3D



©Raoul NK, 3dnatureguy/Wikimedia Commons/CC-BY-SA-3.0

Medium-scale stereo 3D displays

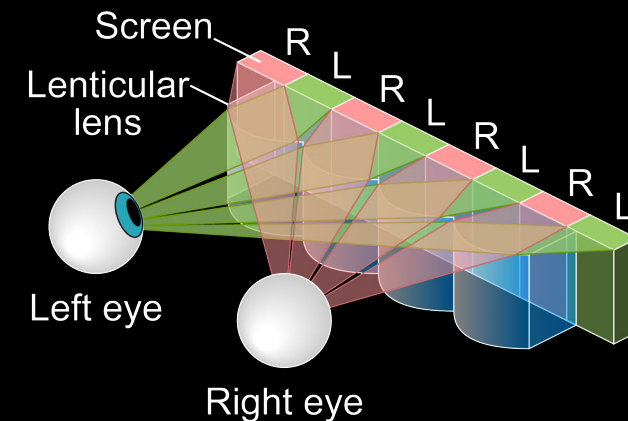
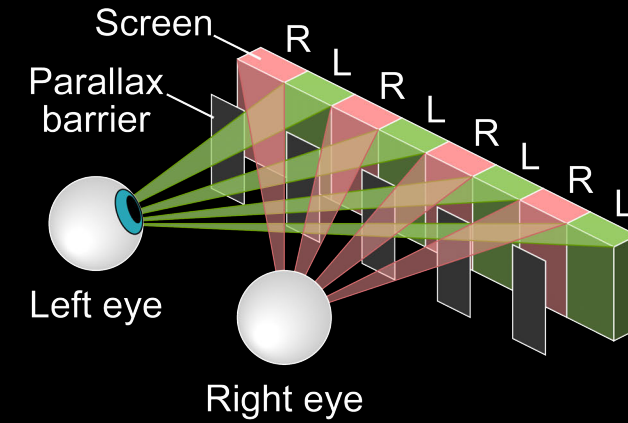
Active shutter glasses

e.g. NVIDIA 3D Vision, 3D TVs



©2011 MTBS3D/NVIDIA

Autostereoscopy



©Cmglee/Wikimedia Commons/CC-BY-SA-3.0

Other stereo 3D displays

Head-mounted displays (HMDs)

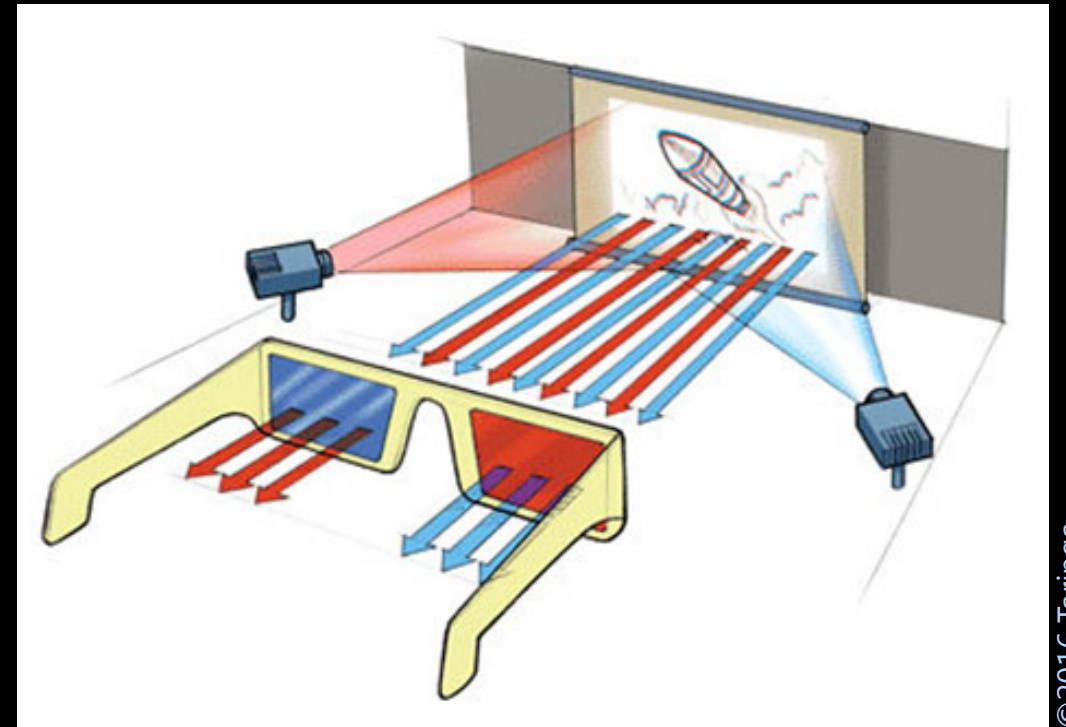
e.g. HTC Vive, Oculus Rift, Google Cardboard



©2016 HTC Corporation

Anaglyph stereo

e.g. red cyan glasses, ColorCode 3-D, Inficolor 3D

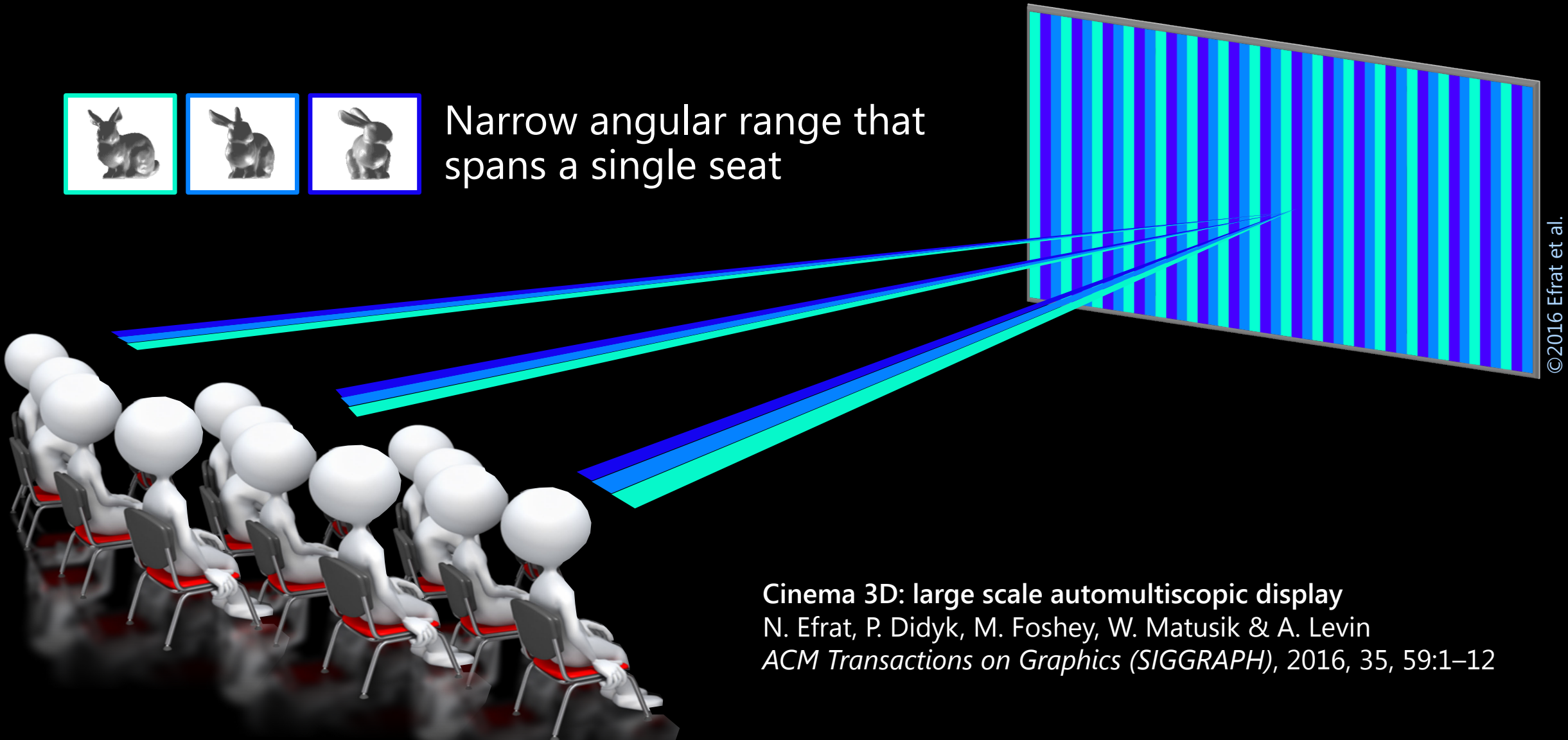


©2016 Taringa

Cinema 3D



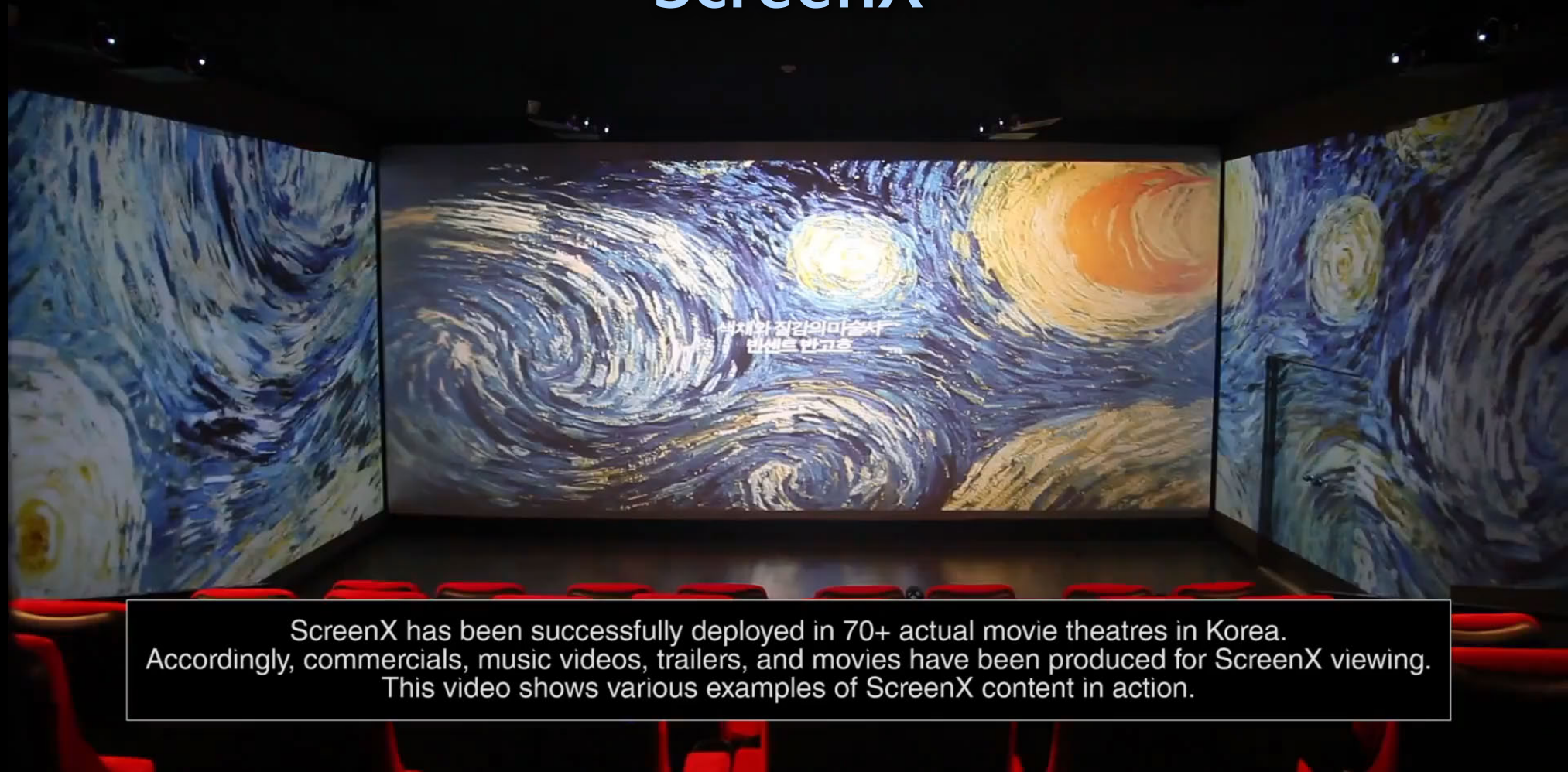
Narrow angular range that spans a single seat



©2016 Efrat et al.

Cinema 3D: large scale automultiscopic display
N. Efrat, P. Didyk, M. Foshey, W. Matusik & A. Levin
ACM Transactions on Graphics (SIGGRAPH), 2016, 35, 59:1–12

ScreenX



ScreenX has been successfully deployed in 70+ actual movie theatres in Korea. Accordingly, commercials, music videos, trailers, and movies have been produced for ScreenX viewing. This video shows various examples of ScreenX content in action.

©2017 Lee et al./KAIST Visual Media Lab

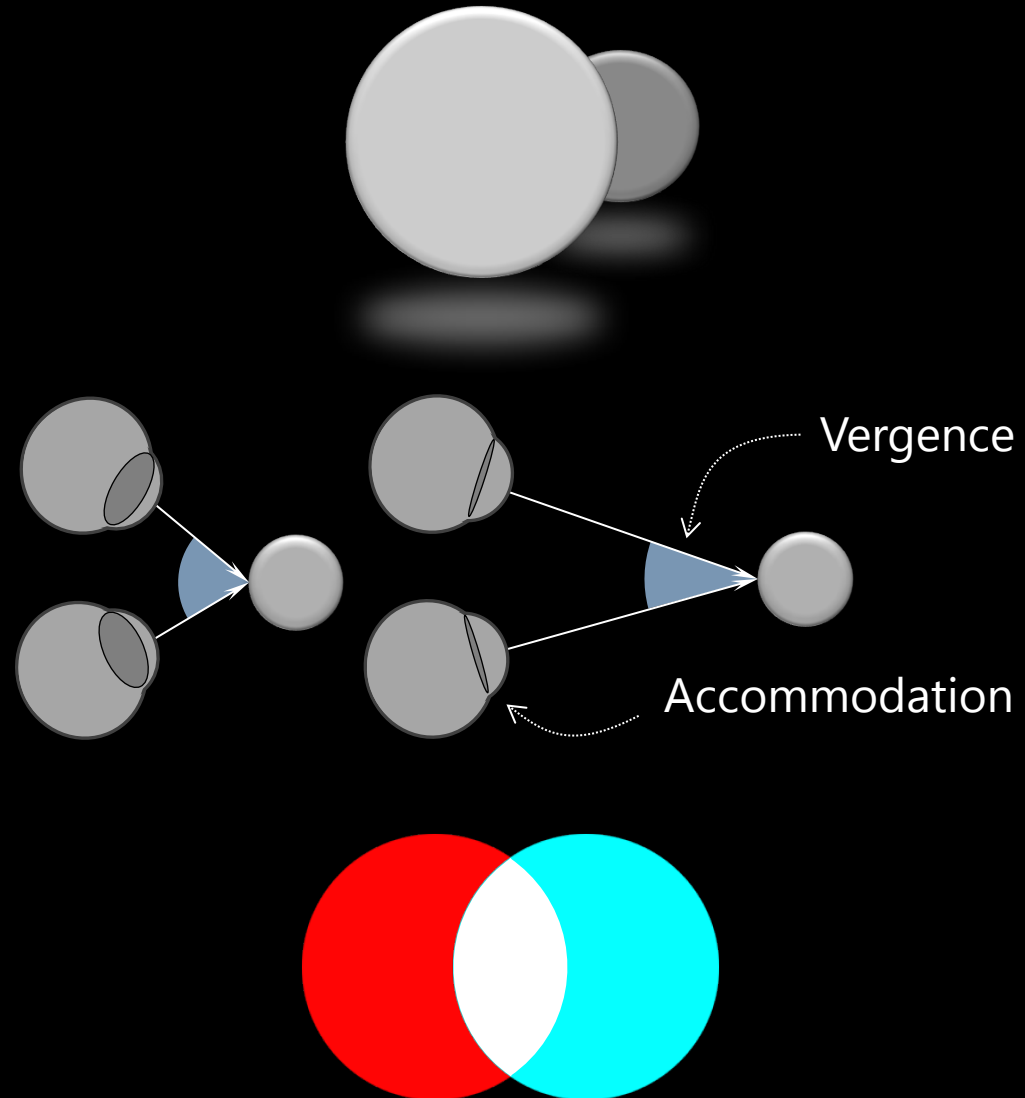
ScreenX: public immersive theatres with uniform movie viewing experiences

J. Lee, S. Lee, Y. Kim & J. Noh

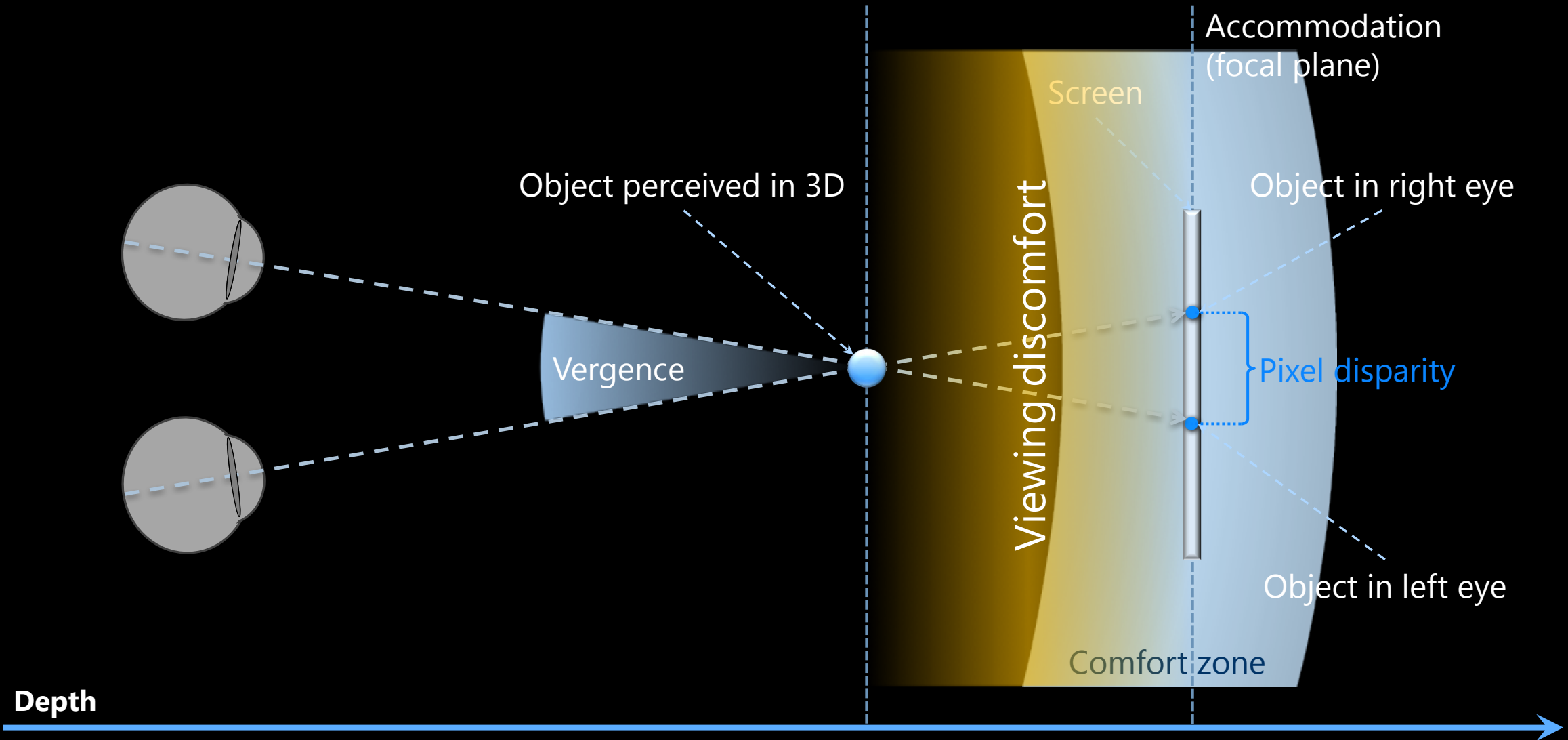
IEEE Transactions on Visualization and Computer Graphics, 2017, 23(2), 1124–1138

Depth cues

- Pictorial cues:
 - size, occlusion, perspective, aerial perspective, texture gradient, motion parallax, depth of field, ...
- Ocular cues:
 - Accommodation
 - Vergence
- Binocular disparity

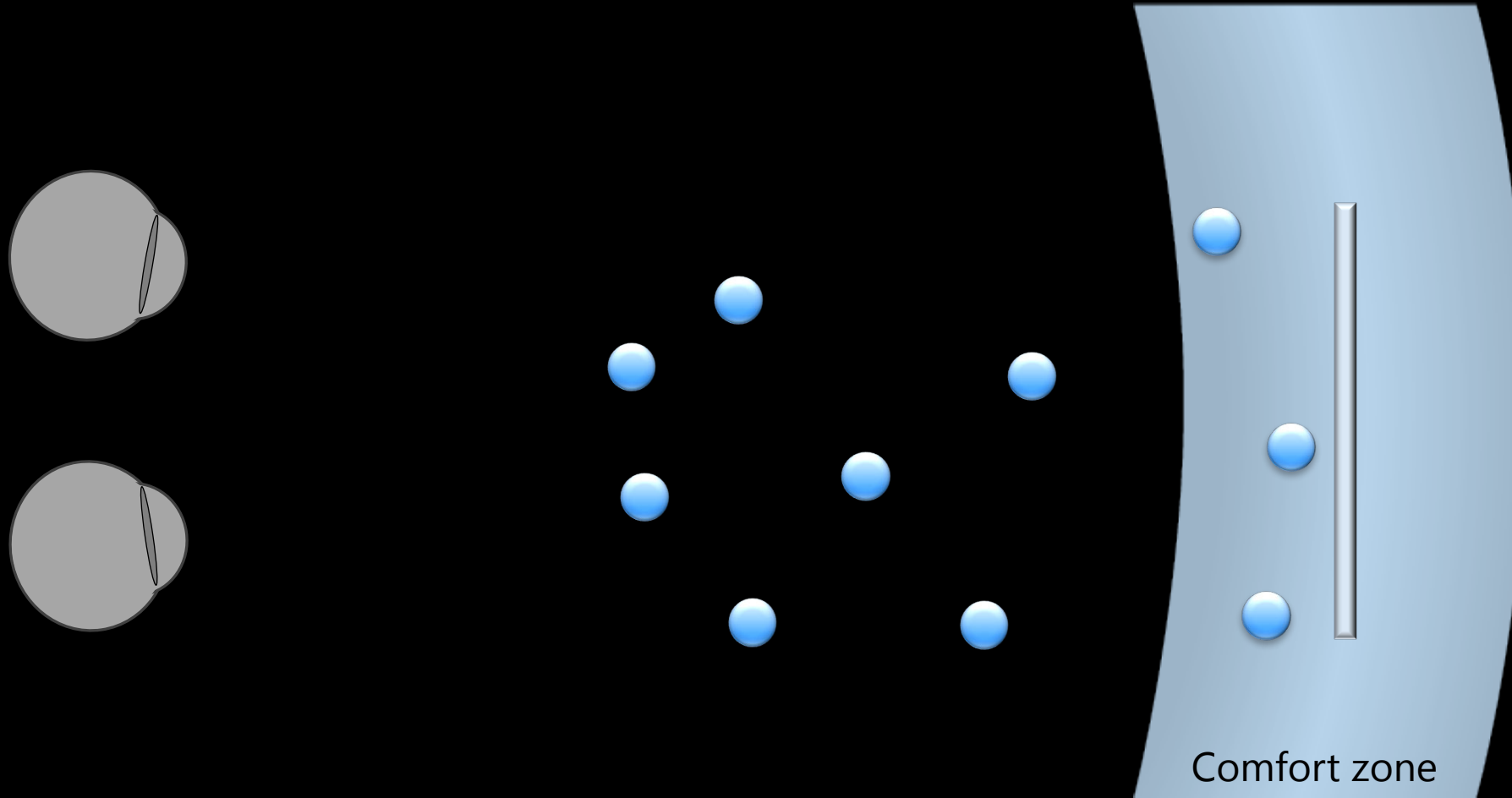


How does disparity work?

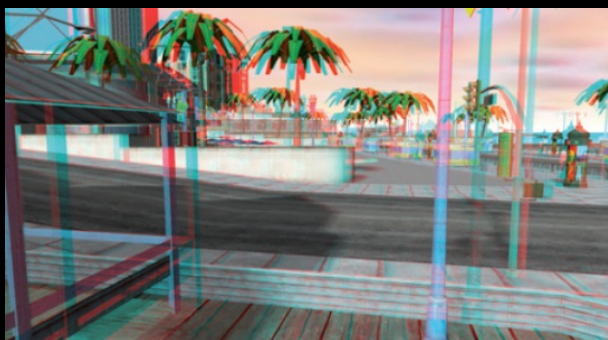
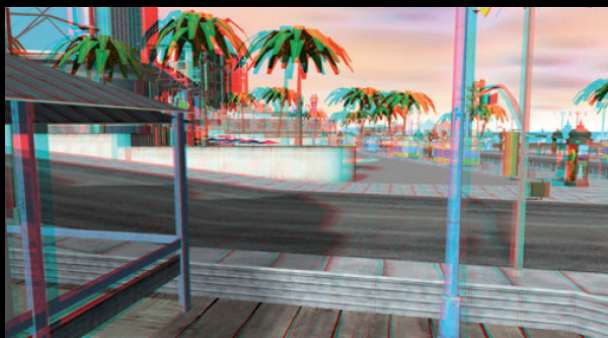


Depth

Preventing viewing discomfort

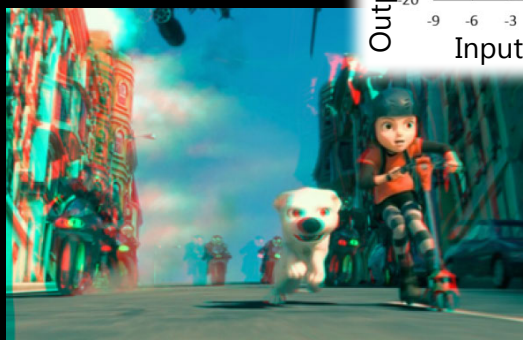
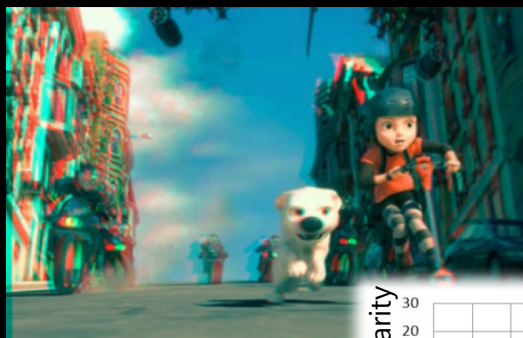


Disparity manipulation



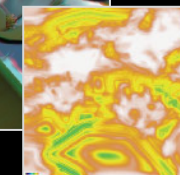
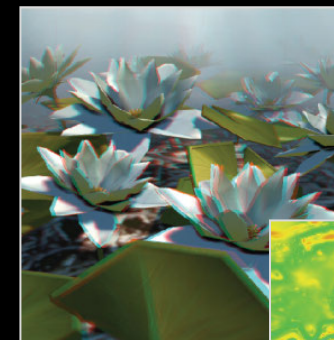
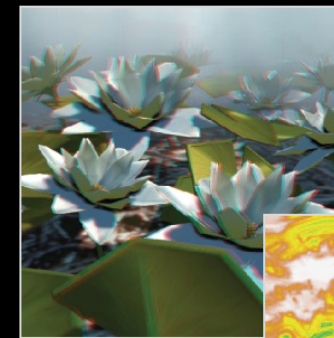
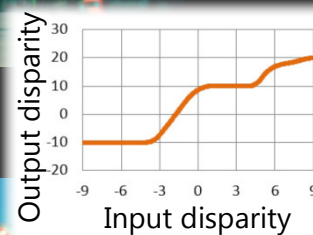
OSCAM – Optimized stereoscopic camera control for interactive 3D

Oskam et al., *SIGGRAPH Asia 2011*



Nonlinear disparity mapping for stereoscopic 3D

Lang et al., *SIGGRAPH 2010*



Perceived distortions



Perceived distortions

A perceptual model for disparity

Didyk et al., *SIGGRAPH 2011*

Additional reading

- **Production rules for stereo acquisition**
Zilly et al., *Proc. IEEE* 2011
- **Predicting stereoscopic viewing comfort using a coherence-based computational model**
Richardt et al., *CAe* 2011
- **A luminance-contrast-aware disparity model and applications**
Didyk et al., *SIGGRAPH Asia* 2012
- **A metric of visual comfort for stereoscopic motion**
Du et al., *SIGGRAPH Asia* 2013
- **Modeling and optimizing eye vergence response to stereoscopic cuts**
Templin et al., *SIGGRAPH* 2014
- **What makes 2D-to-3D stereo conversion perceptually plausible?**
Kellnhofer et al., *SAP* 2015
- **GazeStereo3D: seamless disparity manipulations**
Kellnhofer et al., *SIGGRAPH* 2016
- **Causes of discomfort in stereoscopic content: a review**
Terzic & Hansard, *arXiv:1703.04574*

2D-to-3D conversion



©2016 Leimkühler et al.

Perceptual real-time 2D-to-3D conversion using cue fusion
Leimkühler et al., *IEEE TVCG* 2017



©2011 Wang et al./ACM

StereoBrush: interactive 2D to 3D conversion
using discontinuous warps
Wang et al., *SBIM* 2011

2D-to-3D conversion – Additional reading

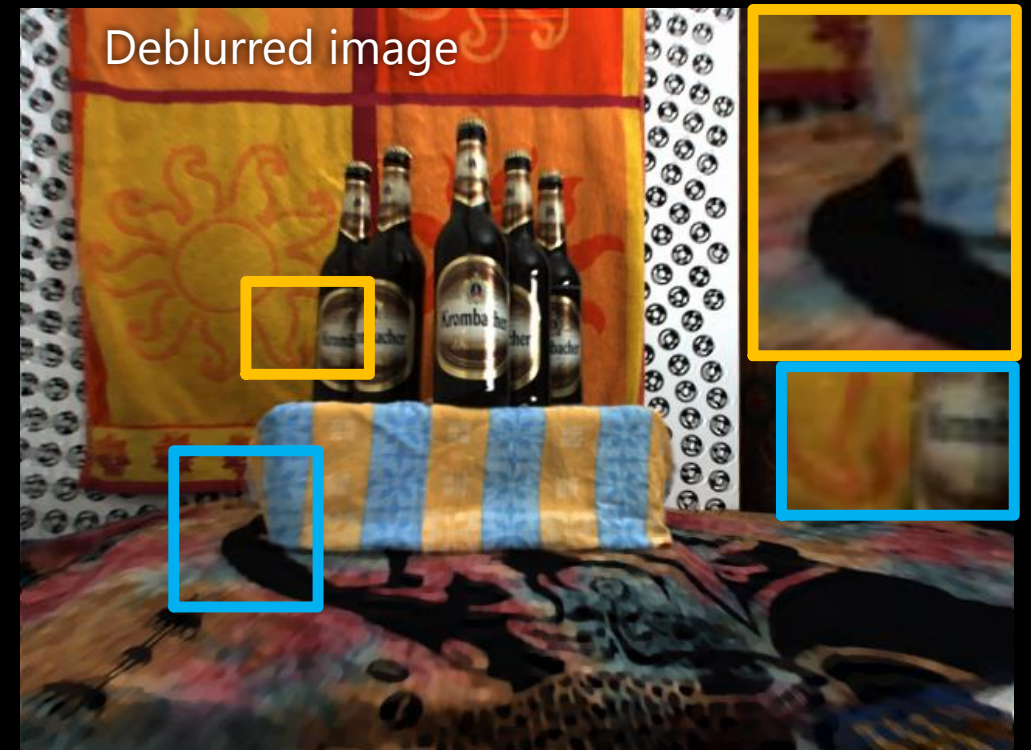
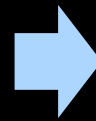
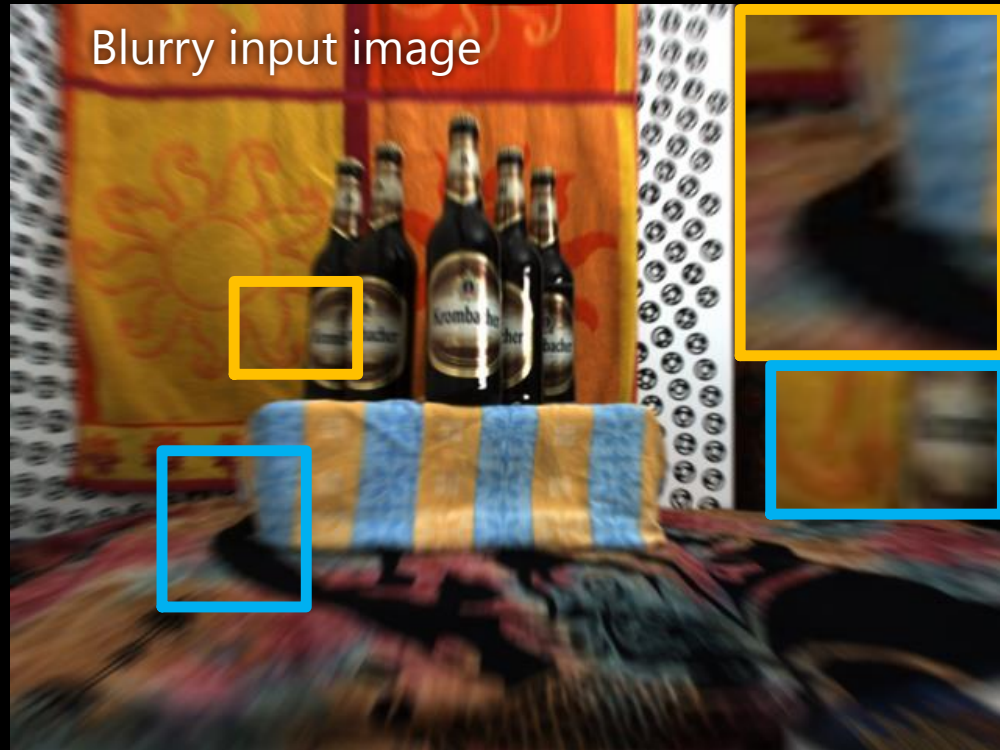
- **Deep3D: fully automatic 2D-to-3D video conversion with deep convolutional neural networks**
Xie et al., *ECCV 2016*
- **Hallucinating stereoscopy from a single image**
Zeng et al., *CGF (Eurographics) 2015*
- **Video stereolization: combining motion analysis with user interaction**
Liao et al., *IEEE TVCG 2012*
- **Depth Director: a system for adding depth to movies**
Ward et al., *IEEE CG&A 2011*
- **Stereoscopic video synthesis from a monocular video**
Zhang et al., *IEEE TVCG 2007*

Video de-anaglyph



Temporally Coherent Video De-Anaglyph
Roo & Richardt, *SIGGRAPH Talks 2014*

Stereo 3D video deblurring

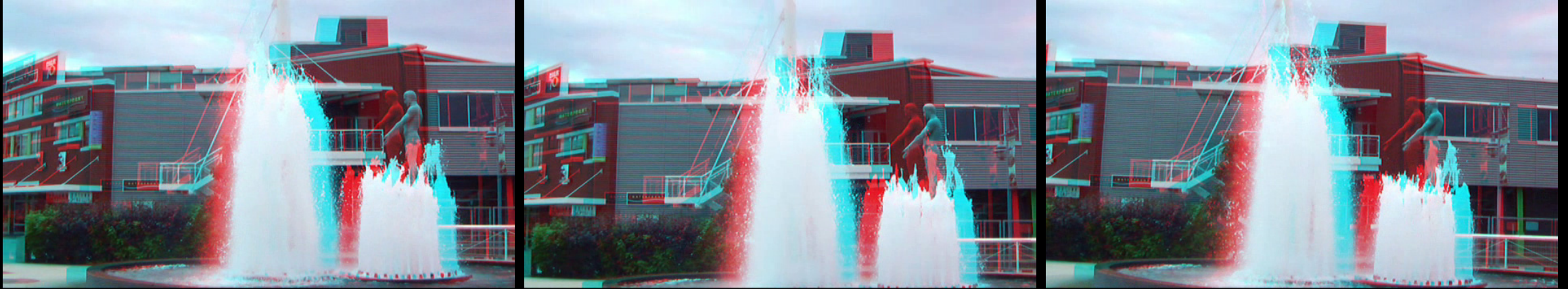


Simultaneous stereo video deblurring and scene flow estimation ↑
Pan et al., *CVPR 2017*

Stereo Video Deblurring
Sellent et al., *ECCV 2016*

Stereo 3D video stabilisation

Input video frames (anaglyph)



Stabilised video frames (anaglyph)

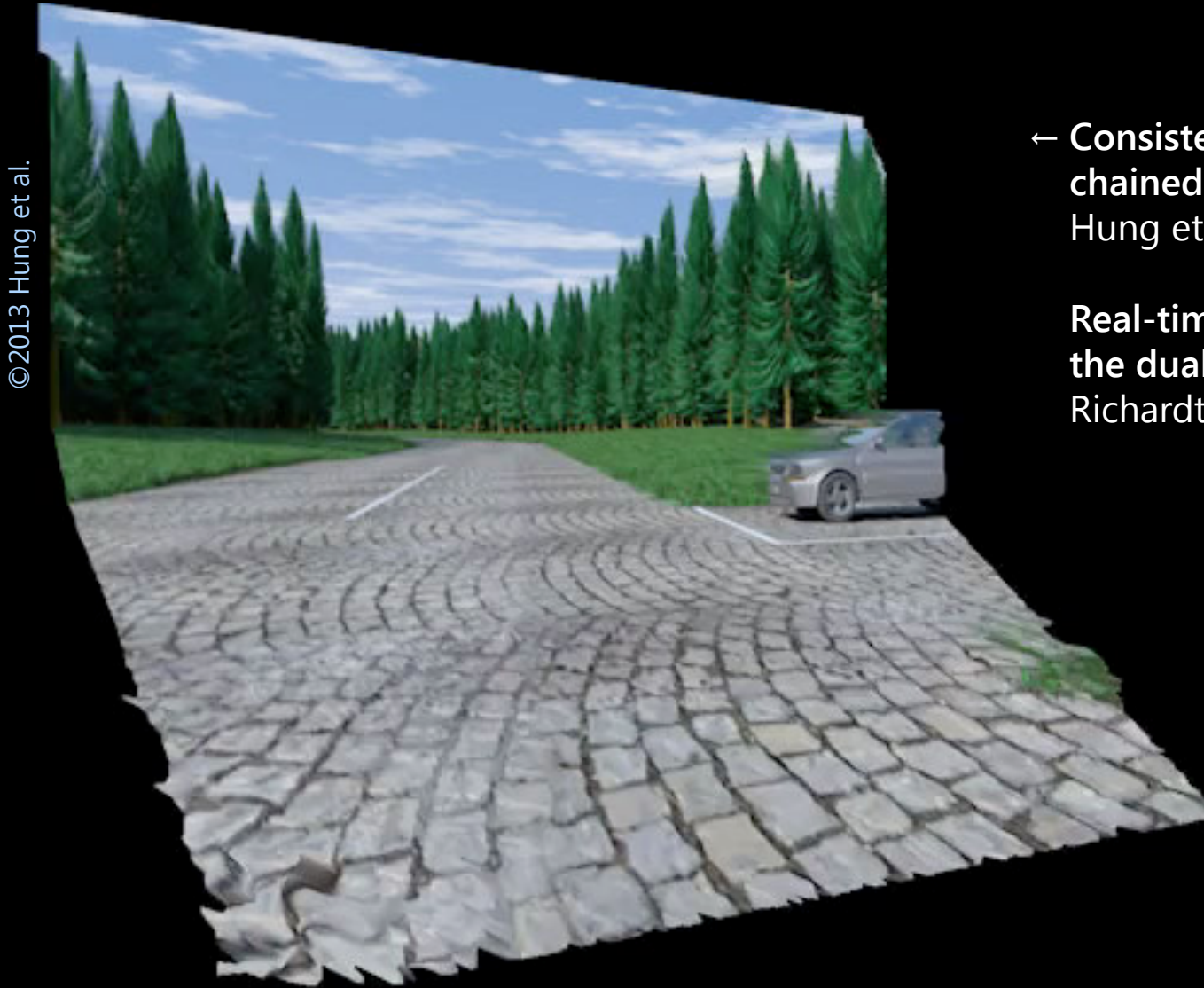


Joint Subspace Stabilization for Stereoscopic Video

Liu et al., *ICCV 2013*

©2013 Liu et al./IEEE

Correspondence finding



©2013 Hung et al.

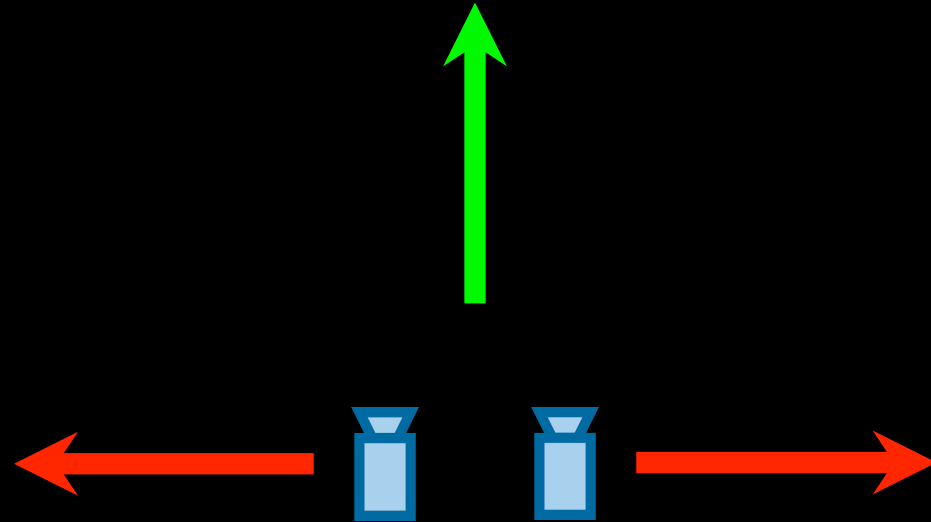
← Consistent binocular depth and scene flow with chained temporal profiles
Hung et al., *IJCV* 2013

Real-time spatiotemporal stereo matching using the dual-cross-bilateral grid
Richardt et al., *ECCV* 2010

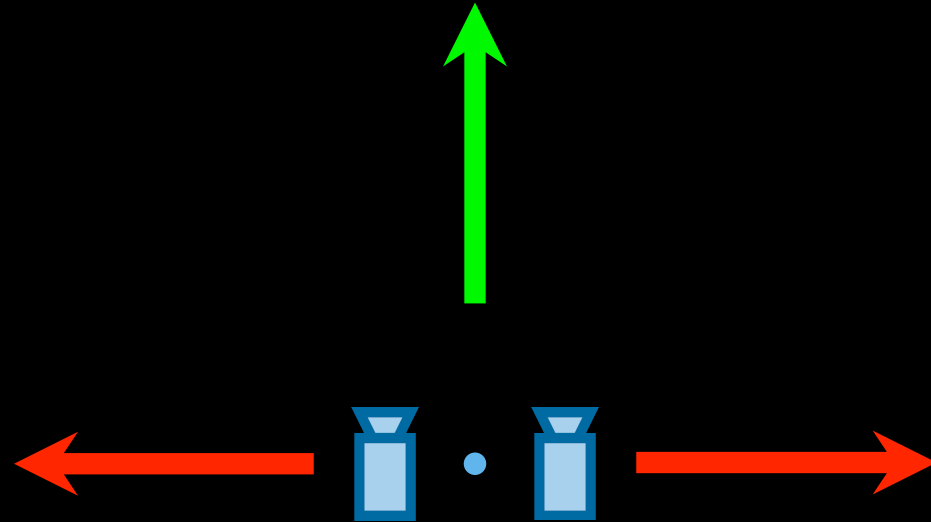
Image-only techniques

- Retargeting:
 - **Object-coherence warping for stereoscopic image retargeting**
Lin et al., *IEEE TCSVT* 2014
 - **Stereo seam carving a geometrically consistent approach**
Basha et al., *IEEE TPAMI* 2013
- Compositing:
 - **StereoPasting: interactive composition in stereoscopic images**
Tong et al., *IEEE TVCG* 2013
 - **Stereoscopic 3D copy & paste**
Lo et al., *SIGGRAPH Asia* 2010
- Warping:
 - **Perspective-aware warping for seamless stereoscopic image cloning**
Luo et al., *SIGGRAPH Asia* 2012
 - **Enabling warping on stereoscopic images**
Niu et al., *SIGGRAPH Asia* 2012

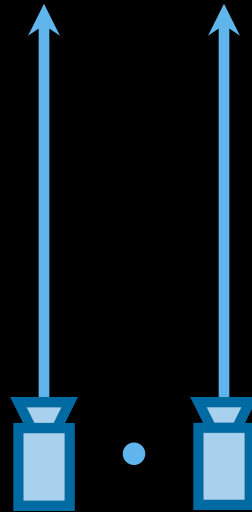
Capturing 3D panoramas



Capturing 3D panoramas



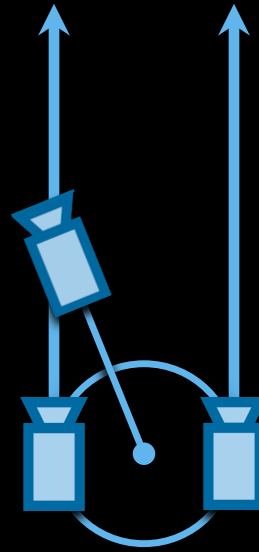
Capturing 3D panoramas



Capturing 3D panoramas

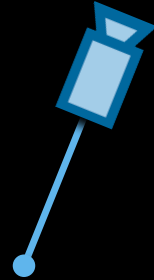


Capturing 3D panoramas



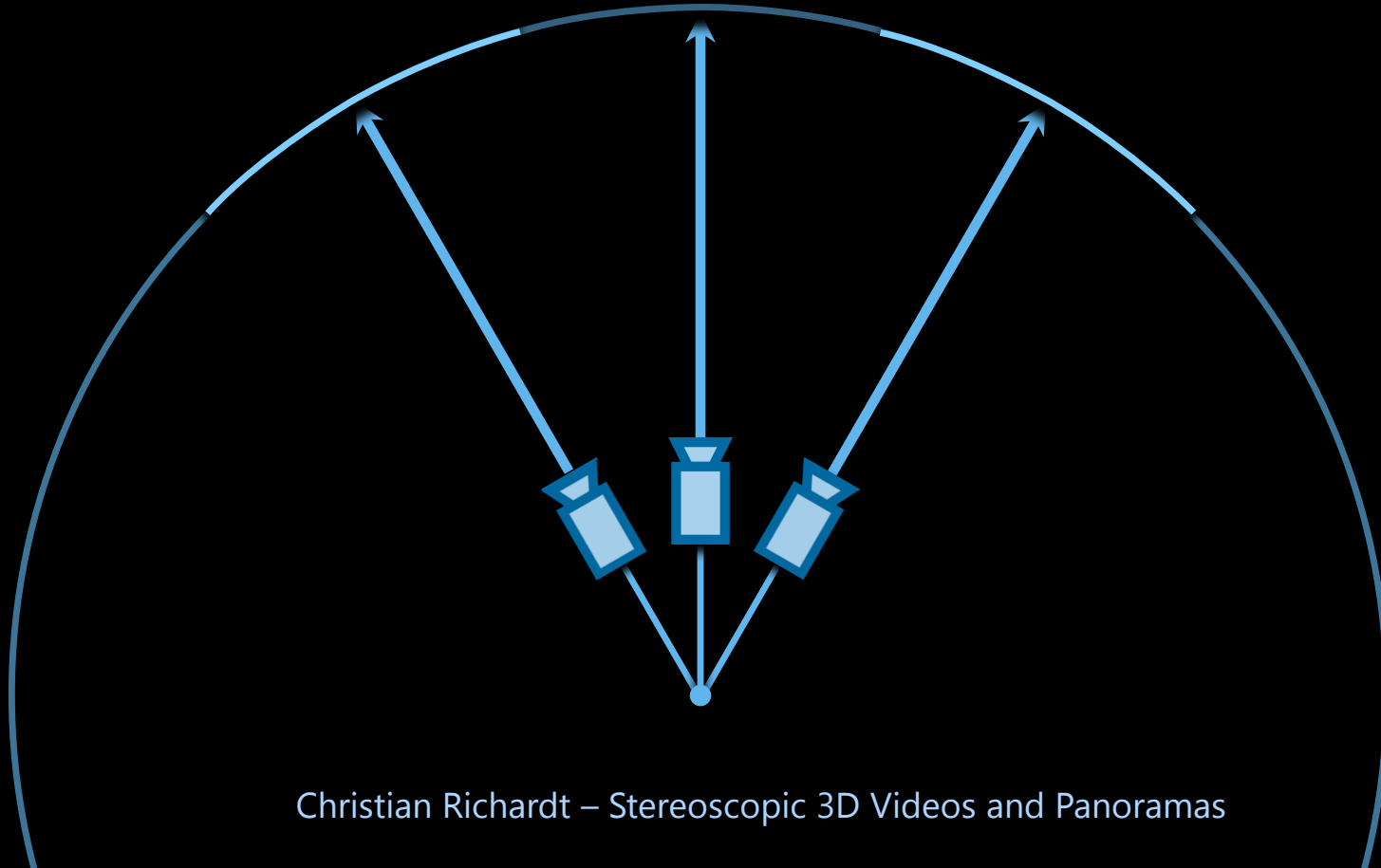
Omnistereo: Panoramic Stereo Imaging
Peleg et al., *IEEE TPAMI* 2001

Capturing 3D panoramas

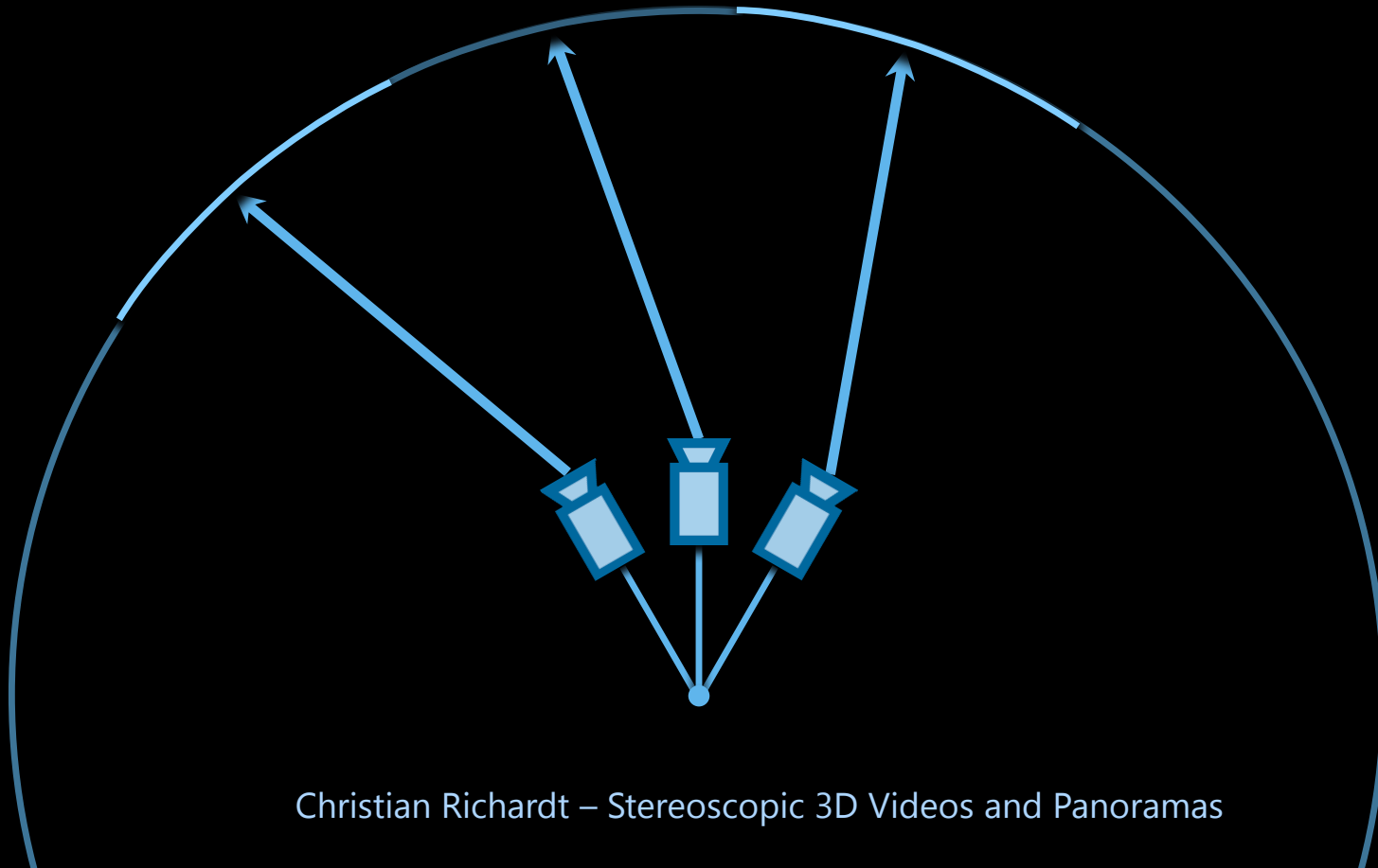


Omnistereo: Panoramic Stereo Imaging
Peleg et al., *IEEE TPAMI* 2001

Capturing 3D panoramas



Capturing 3D panoramas



Capturing 3D panoramas

Input video:



©2013 Richardt et al.

Capturing 3D panoramas



Megastereo: Constructing High-Resolution Stereo Panoramas
Richardt et al., CVPR 2013

©2013 Richardt et al.

Image alignment



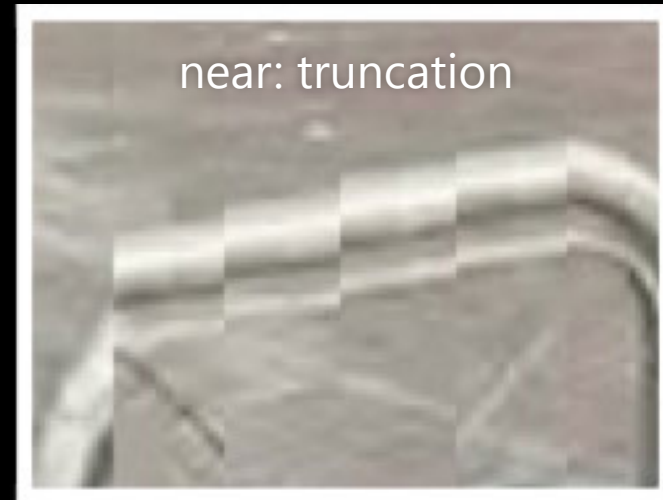
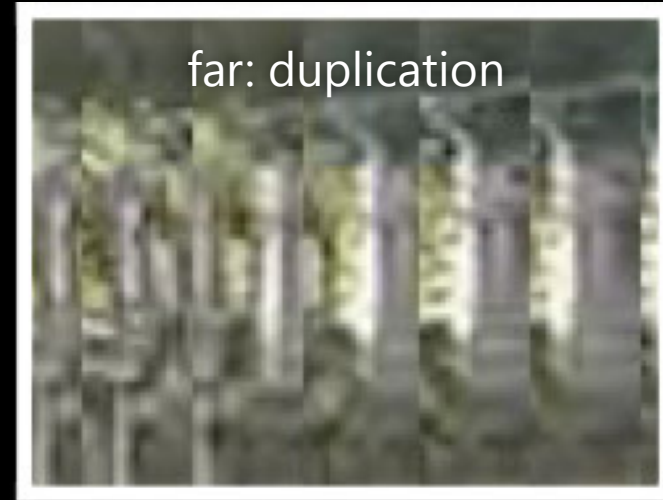
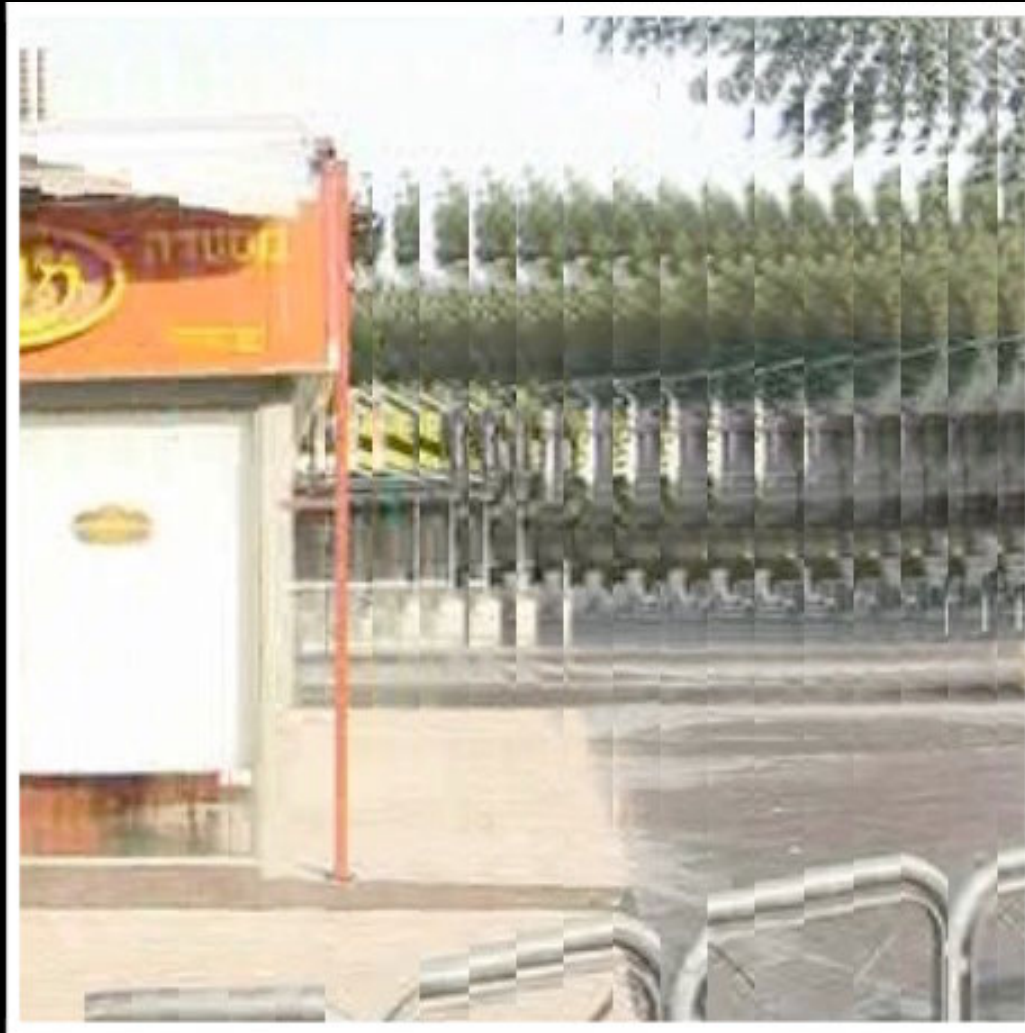
image-based alignment



improved alignment

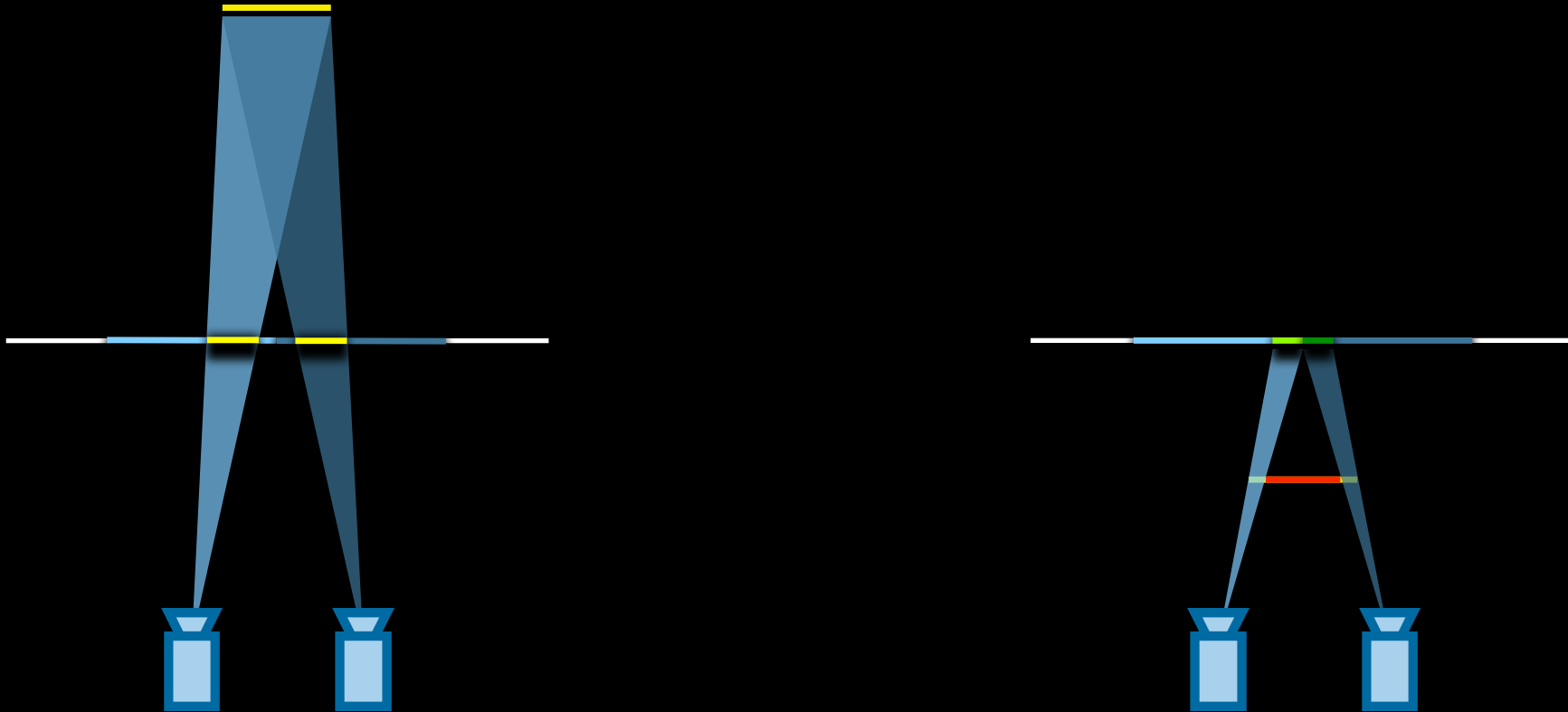
©2013 Richardt et al.

Strip blending artefacts

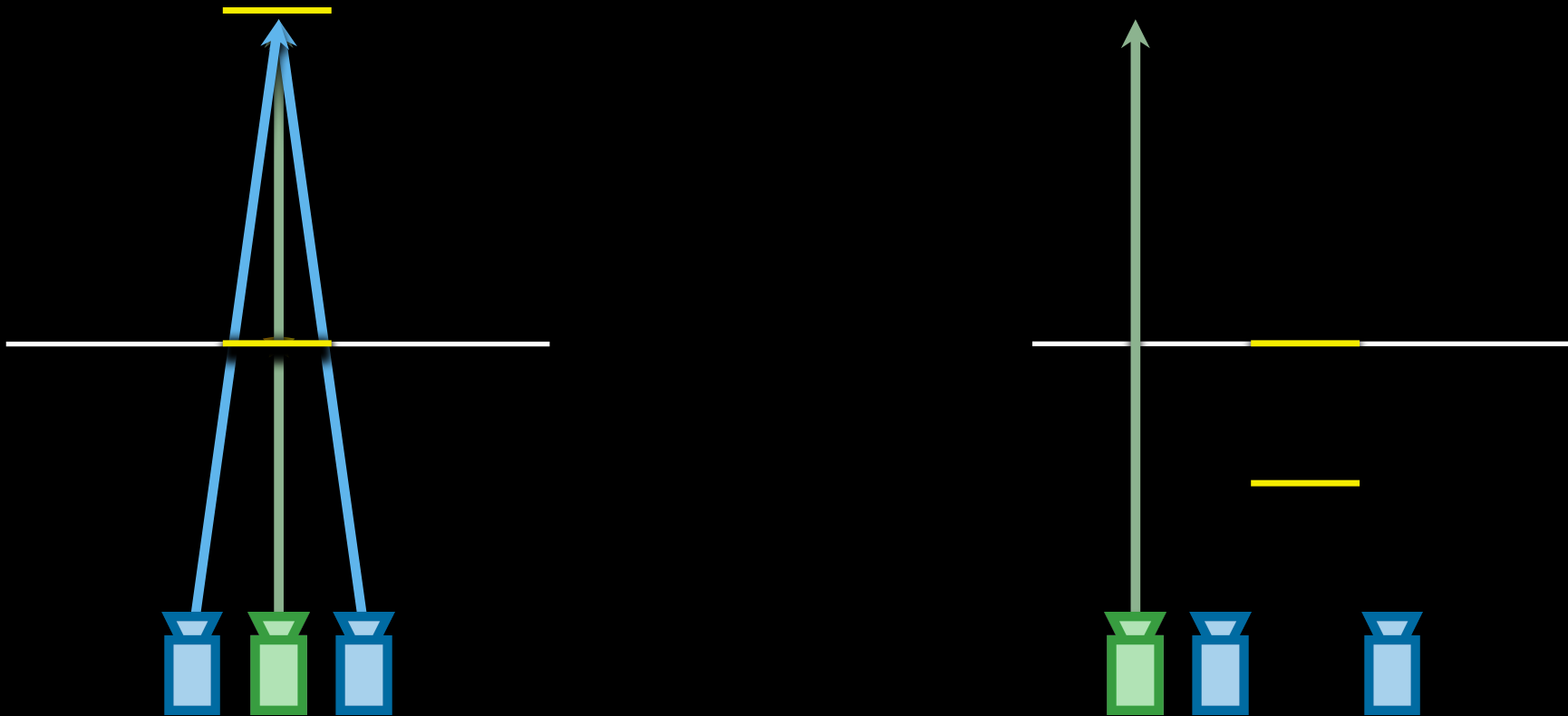


© dataset 'refaim' by Rav-Acha et al., IJCV 2008

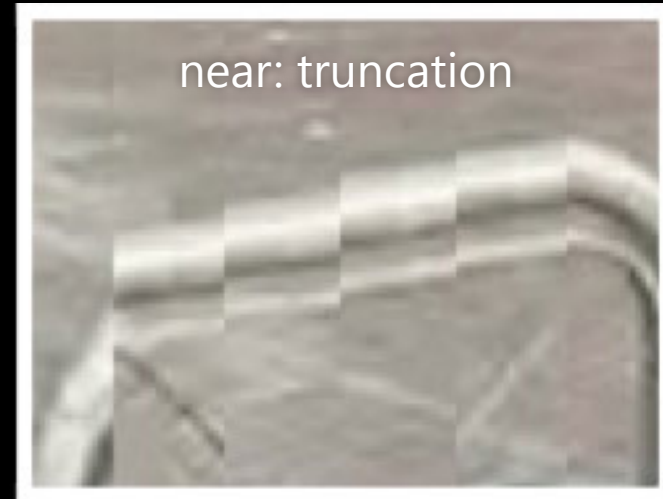
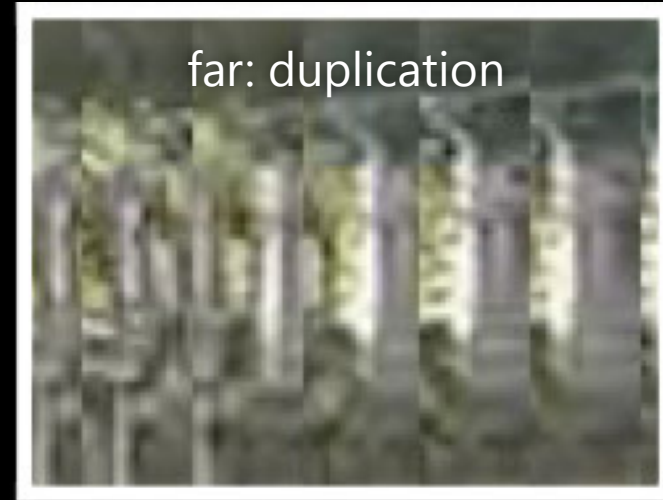
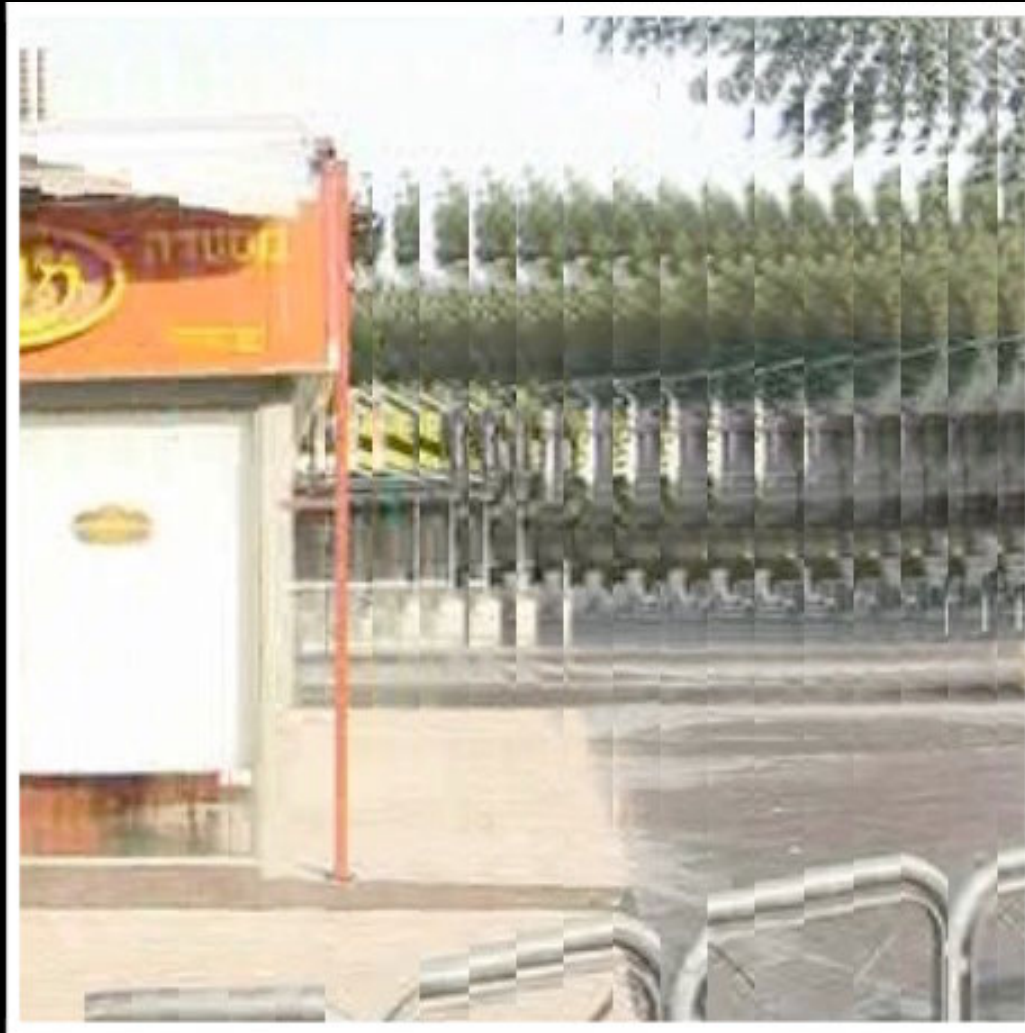
Duplication + truncation



Flow-based ray interpolation

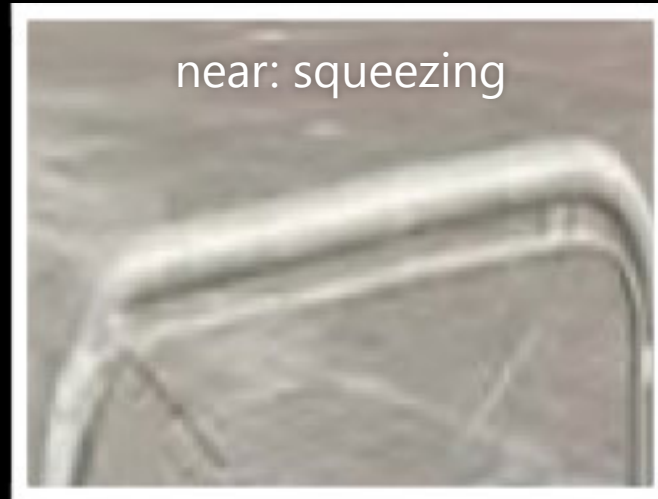


Strip blending artefacts



© dataset 'refaim' by Rav-Acha et al., IJCV 2008

Flow-based blending



©2013 Richardt et al.; dataset 'refaim' by Rav-Acha et al., IJCV 2008

Blending comparison

No blending



Flow-based blending



©2013 Richardt et al.

Stereo 3D panorama



Megastereo: Constructing High-Resolution Stereo Panoramas
Richardt et al., *CVPR 2013*



360°

140 MP stereo panorama



©2013 Richardt et al.

100% zoom





Questions?

Christian Richardt

Stereoscopic 3D Videos and Panoramas

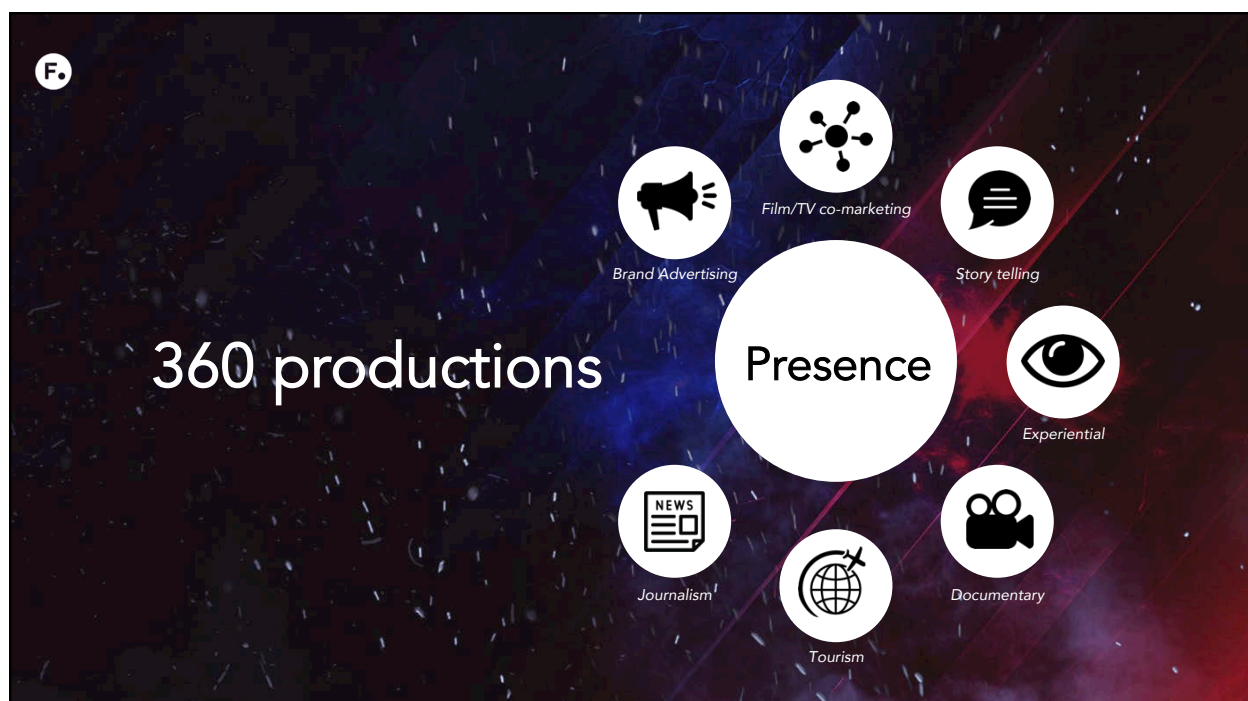




30 JULY – 3 AUGUST *Los Angeles*
SIGGRAPH2017

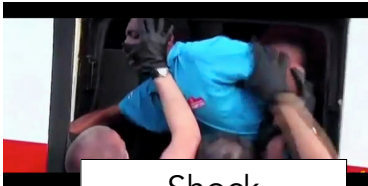
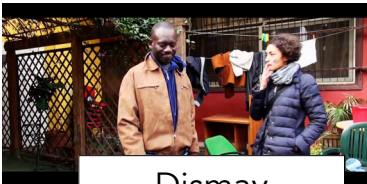
360 Video in Production

Jonathan Starck
Head of Research, Foundry





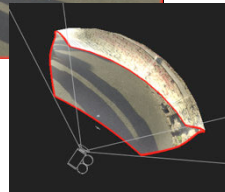
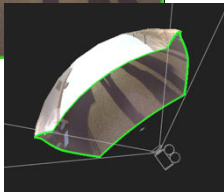
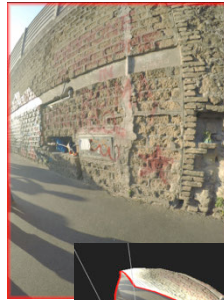
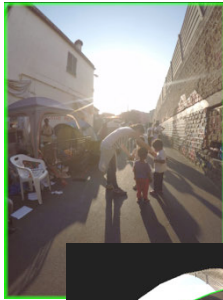
F. Shot planning – staging is critical

 Immersion	 Shock	 Support
 Dismay	 Emptiness	 Hope

... but not always possible

F.

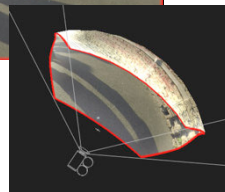
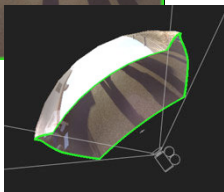
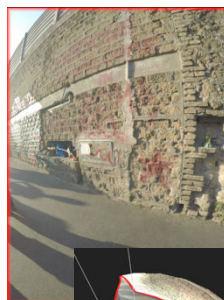
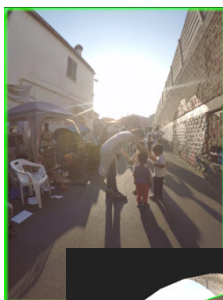
Shot planning – impact on image stitching



Converged at 10m

F.

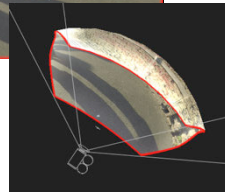
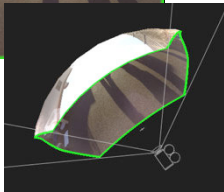
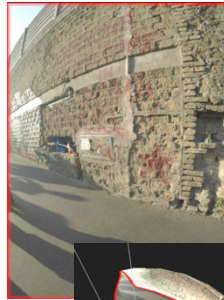
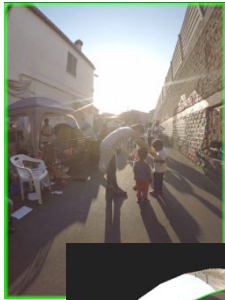
Shot planning – impact on image stitching



Converged at 2m

F.

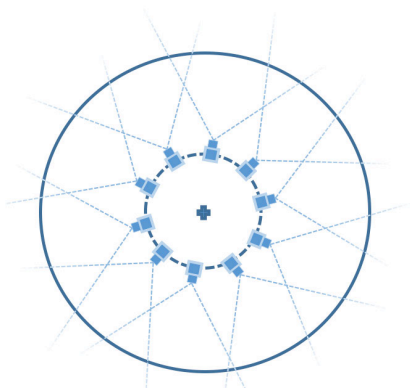
Shot planning – impact on image stitching



Converged at 1m

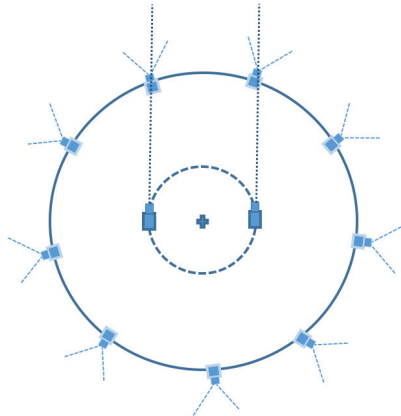
F.

Monoscopic cameras – design considerations

mono-stitch
removing parallax

F.

Stereoscopic cameras – design considerations



stereo-stitch
using parallax



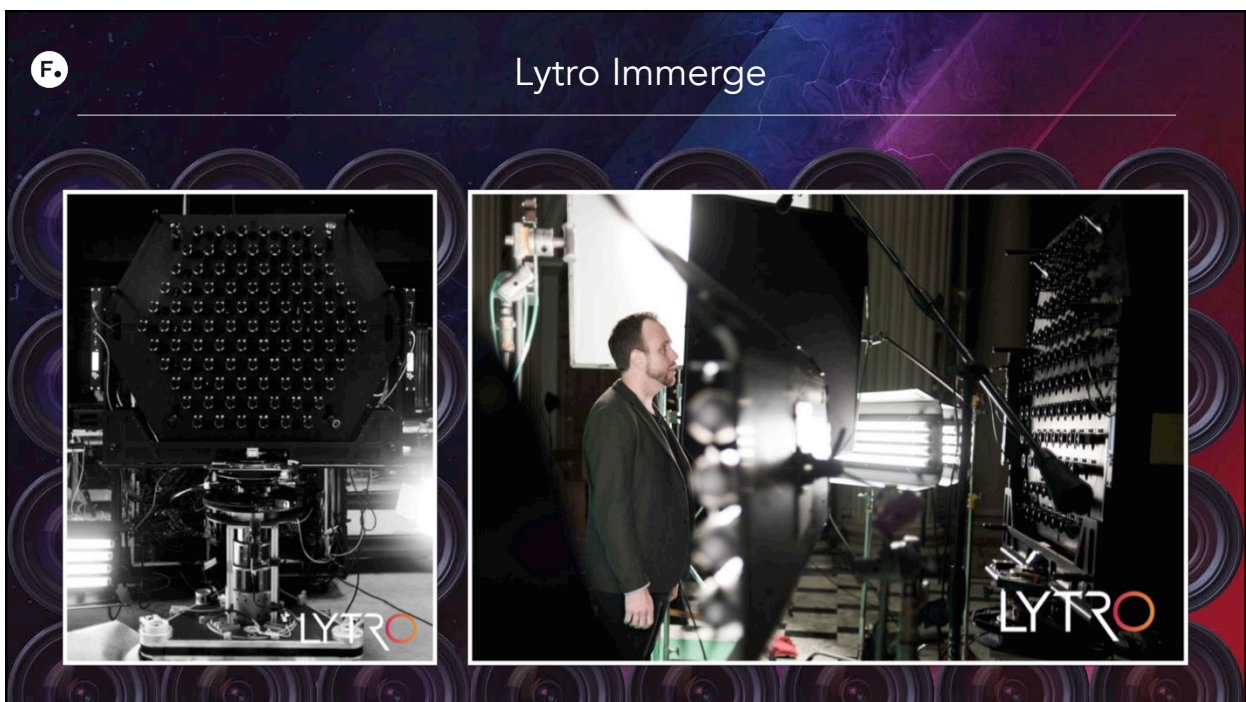
F.

Yi Halo



JUMP
YIHALO





We covered

- 360 productions
- Staging in production
- Shot planning and stitching
- Monoscopic cameras
- Stereoscopic cameras

FOUNDRY.
imagination engineered



30 JULY – 3 AUGUST *Los Angeles*
SIGGRAPH2017

Post-production for 360 Video

Jonathan Starck
Head of Research, Foundry

F.

Challenges for 360 video



Camera rig setup



Colour correction



Stitching



Stitch correction



Stabilisation



Match-move



Headset review



F.

Stitching



F.

Corrections



F.

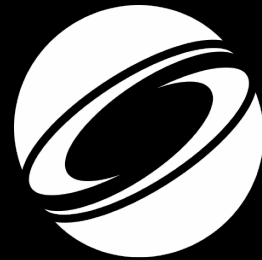
Enhancement



We covered

- Tasks in post-production
- Node-based compositing
- Setting up a "stitch"
- Manual corrections
- CG integration

FOUNDRY.
imagination engineered



30 JULY – 3 AUGUST *Los Angeles*
SIGGRAPH2017

ART, STORYTELLING, AND TOOLS

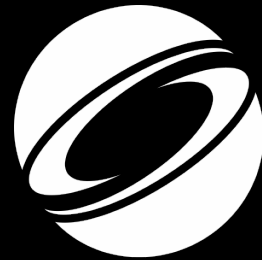
AARON HERTZMAN





Unfortunately, we could not obtain rights release on this content before the deadline to upload course notes to the ACM Digital Library.

Please see our course website for the final material:
<http://richardt.name/pub/Video4VR/>



30 JULY – 3 AUGUST *Los Angeles*
SIGGRAPH2017

VIDEO FOR VIRTUAL REALITY
LIGHT FIELD VIDEO

JAMES TOMPKIN



BROWN
Computer Science

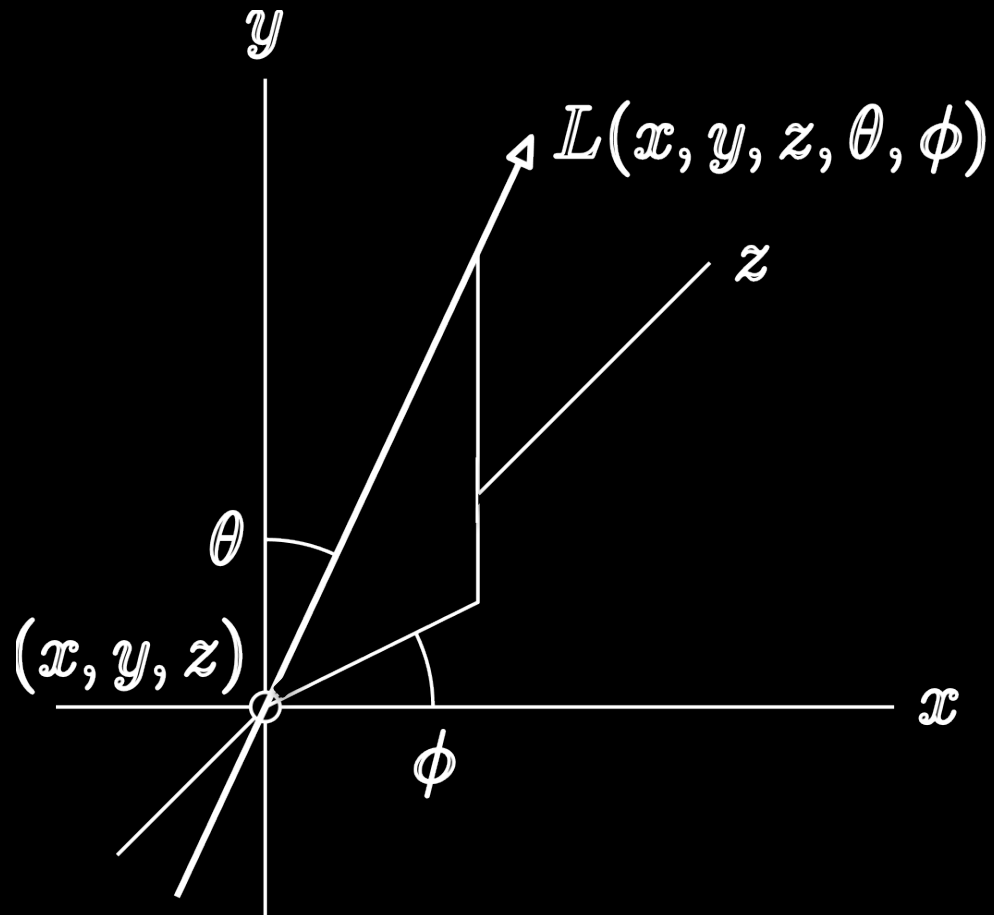


WHAT IS A LIGHT FIELD?



-
- Seems to have turned into a catch-all term for many advanced camera/display technologies.
 - Can mean different things; typically
 - Parallax
 - Focus / refocus
 - How should we think about this?

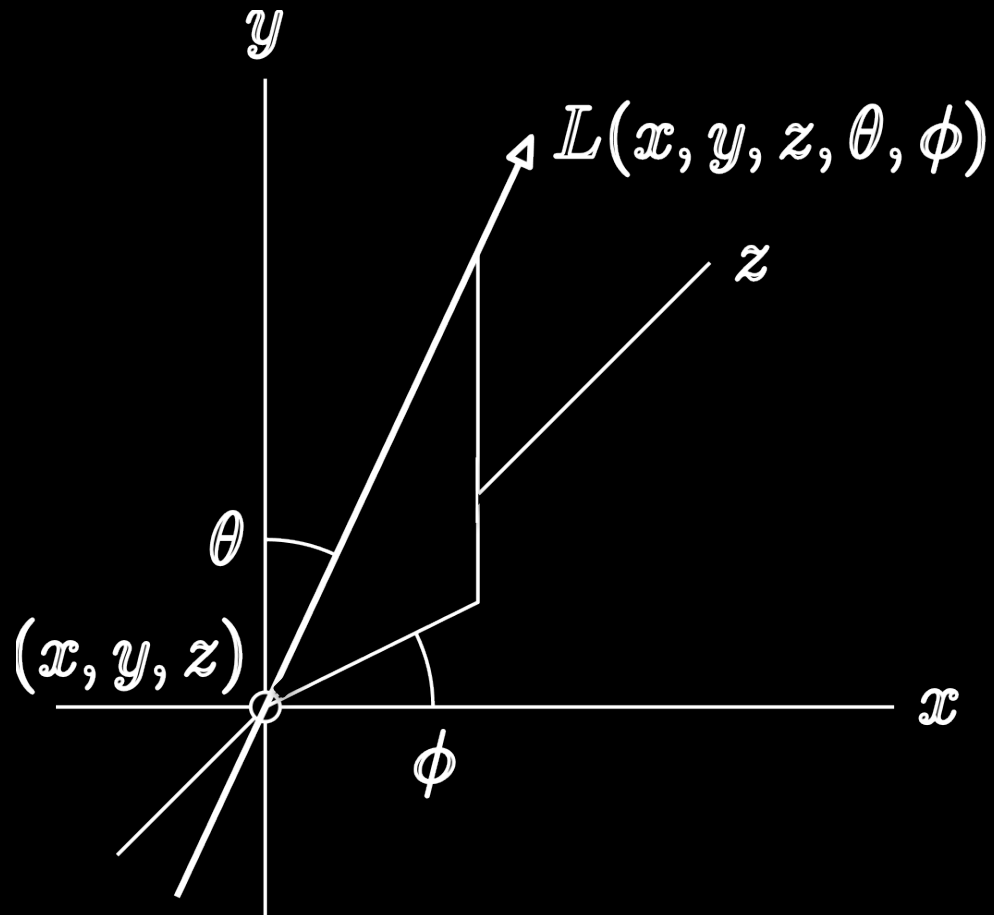
PLENOPTIC FUNCTION



5+ dimensional function
representing all light everywhere

- x, y, z -> position in space
- θ, ϕ -> angle on sphere
- L -> amount of light (radiance)

PLENOPTIC FUNCTION



5+ dimensional function
representing all light everywhere

- x, y, z -> position in space
- θ, ϕ -> angle on sphere
- L -> amount of light (radians)
- Video needs t -> time

ALL TECHNIQUES ARE SAMPLINGS

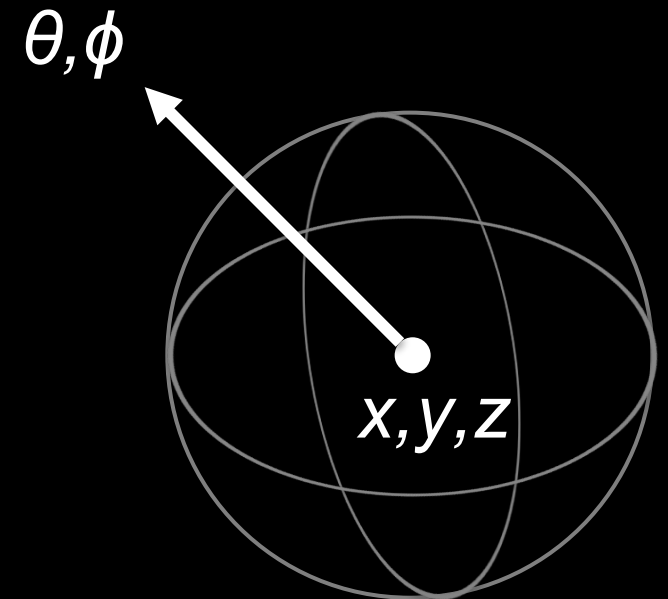


- Impossible to actually capture a light field.
- Different camera configurations provide different *samplings* of the plenoptic function.

ALL TECHNIQUES ARE SAMPLINGS



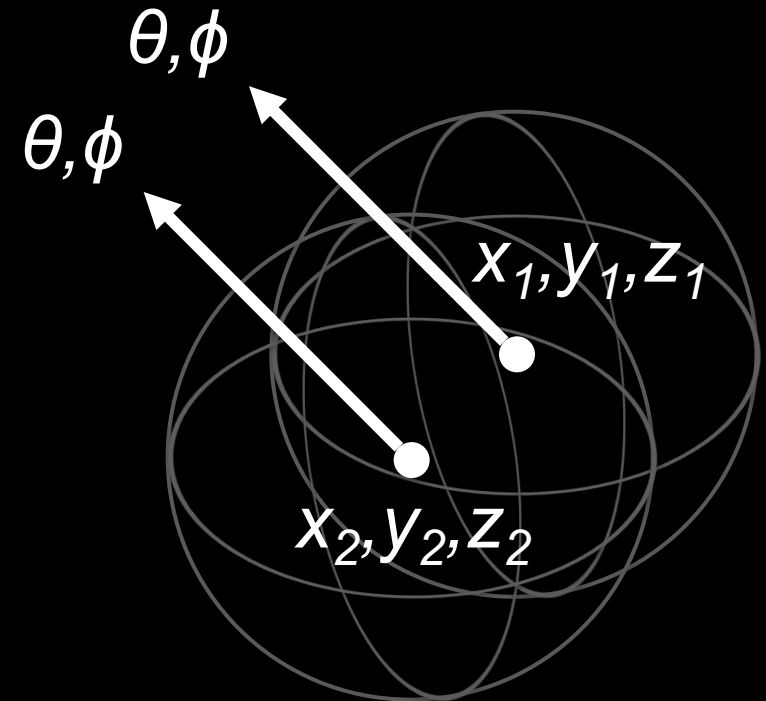
- Impossible to actually capture a light field.
- Different camera configurations provide different *samplings* of the plenoptic function.
- Idealized 360° video
 - Single x, y, z sample
 - ‘Complete’ sampling of θ, ϕ



ALL TECHNIQUES ARE SAMPLINGS



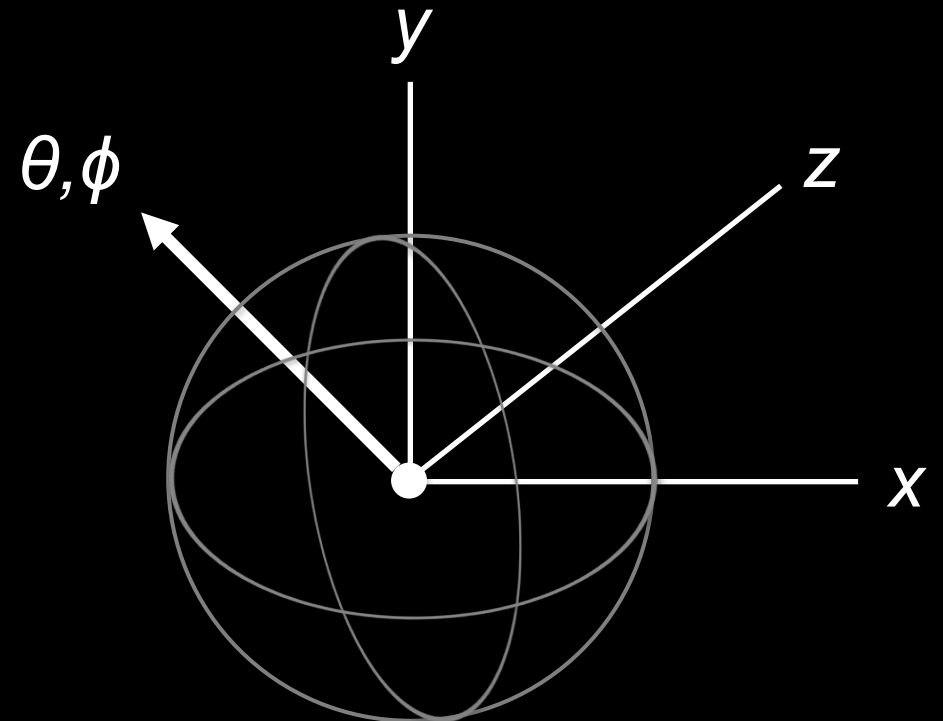
- Impossible to actually capture a light field.
- Different camera configurations provide different *samplings* of the plenoptic function.
- Idealized stereo 3D 360° video
 - Two x, y, z samples
 - ‘Complete’ sampling of θ, ϕ
- *What happens when you tilt your head?*



ALL TECHNIQUES ARE SAMPLINGS



- Impossible to actually capture a light field.
- Different camera configurations provide different *samplings* of the plenoptic function.
- A 'light field' generally implies
 - Many x, y, z samples
 - Some sampling of θ, ϕ
- *Many options for how to sample!*



WHAT IS THE RIGHT WAY TO THINK ABOUT LIGHT FIELDS?



- A space of possible samplings!
 - Leads to a space of possible camera setups
 - Each is some trade-off in the sampling space
- Different configurations for different applications
 - Each sampling misses some information...
 - ...but we have a suite of algorithms to (try to) accommodate.

EXAMPLE: LYTRO CINEMA

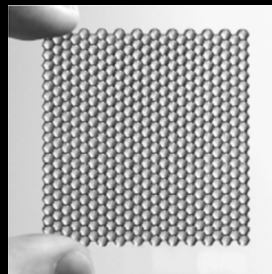


[Image © Lytro]

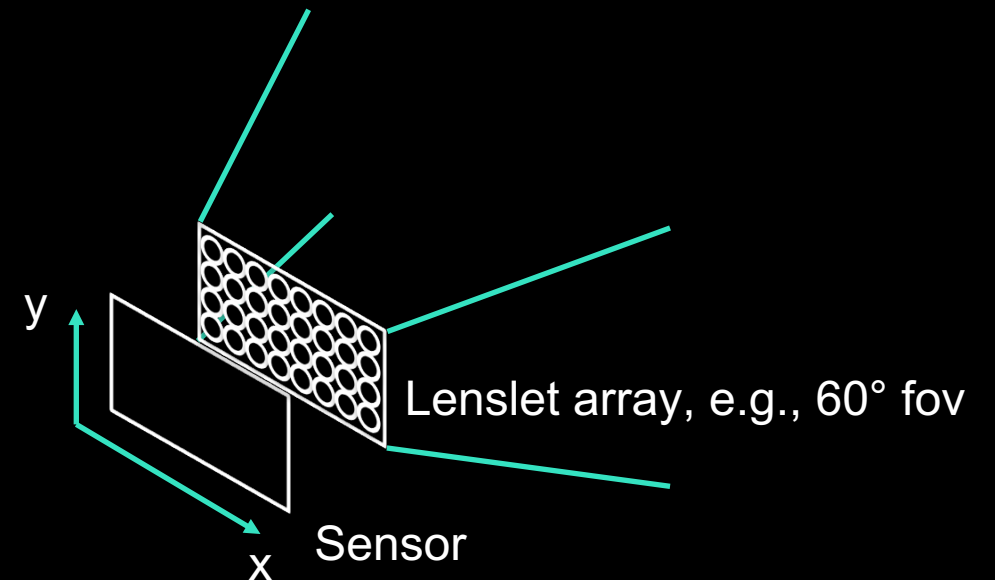


Lenslet-based light field imager
755 MPixel sensor

A lenslet array



- Small baseline – dense sampling
 - Little parallax/head motion
- Small θ, ϕ extent
 - Only sees part of scene



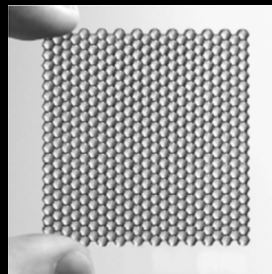
EXAMPLE: LYTRO CINEMA



]Image © Lytro]



Lenslet-based light field imager
755 MPixel sensor



A lenslet array

- Small baseline – dense sampling
 - Little parallax/head motion
- Small θ, ϕ extent
 - Only sees part of scene
- Still a light field camera...
...but not intended for VR.
 - Computational movie camera
 - Helps post-production for synthetic aperture, segmentation

PERCEPTUAL NEEDS



An eye



Variable focus

Two eyes



Binocular stereo
'3D' (*Not really 3D!*)

A neck and a body



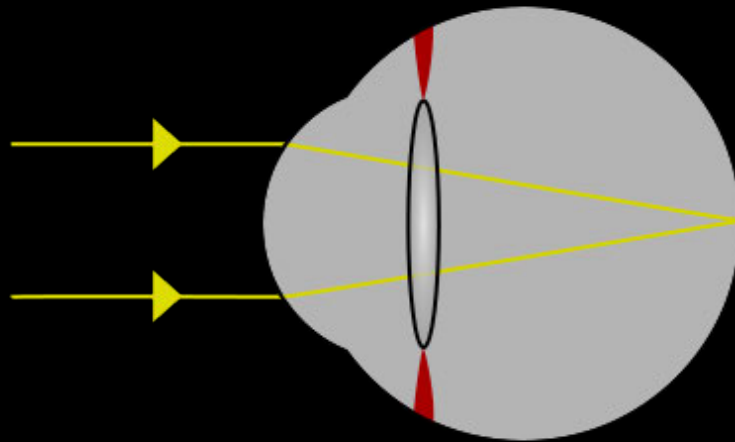
Motion parallax
'6DoF'

SIDE NOTE ABOUT *ACCOMMODATION*

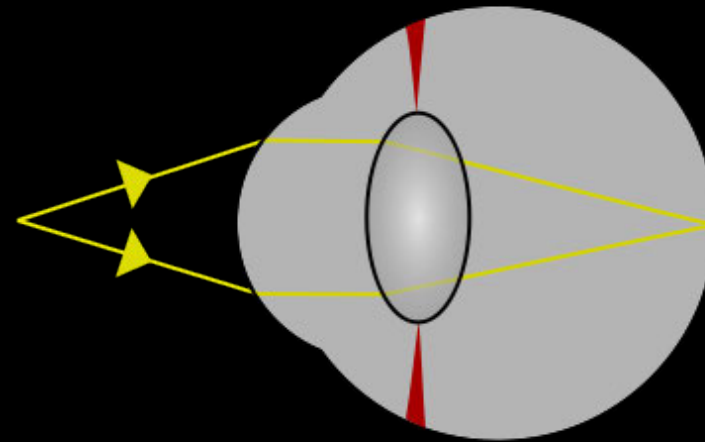


Ability of the eye to focus by changing lens shape.

Human eye:



Object at infinity



Object close up

Requires two *different* rays from same point on object to converge on retina in different positions (be out of focus) for us to accommodate.

SIDE NOTE ABOUT *ACCOMMODATION*



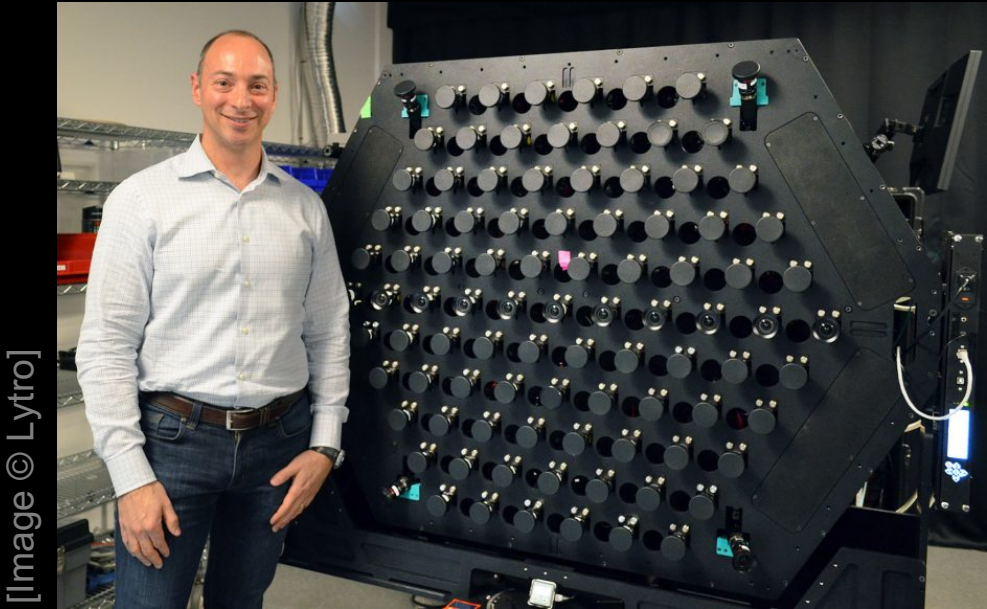
- Accommodation allows a *focusable* image.
- Requires very high angular resolution, as at least two rays from the same point on an object must *both* enter a single eye.
- Image needs a ray every 0.3-0.15 degree.
 - Desktop display at 1-2m distance requires ~15000dpi for 1080p
 - This is at diffraction limit!
 - Lanman et al.'s head mounted 'Near-eye Light Field Display' claim accommodation effects with 120 ppm and 0.14 degree per ray (SIGGRAPH Asia 2013).

LIGHT FIELD VIDEO GOALS



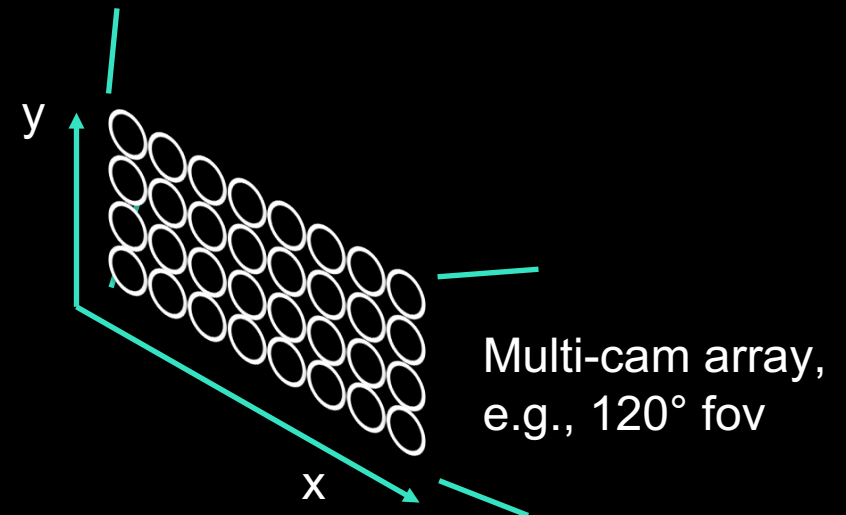
- High immersion
 - 6 degrees of freedom
 - Correct stereo vision everywhere, not just horizontal
- High realism
 - Looks like real life, not like a game
 - Video capture + offline CG

EXAMPLE: LYTRO IMMERGE

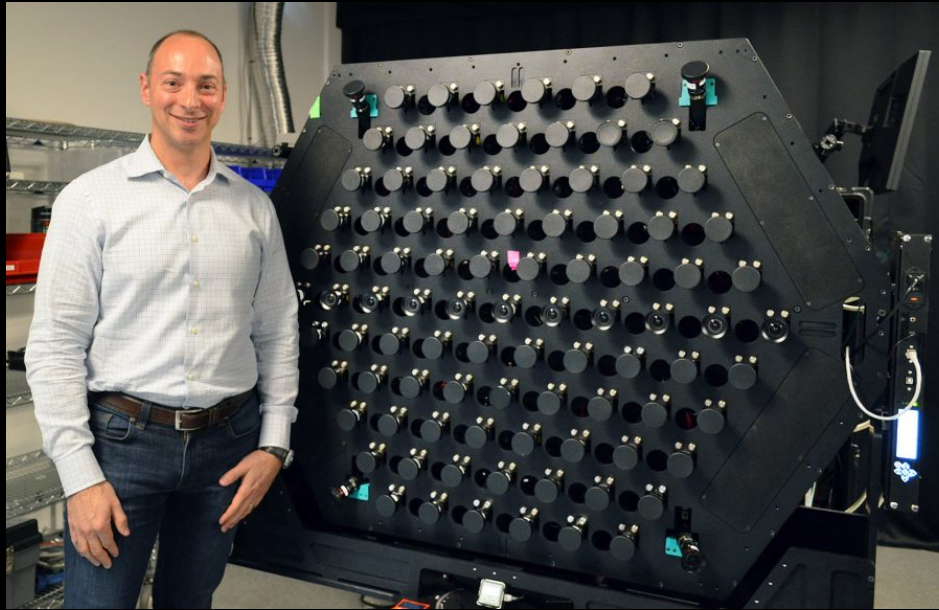


Multi-camera light field imager
95 cameras

- Larger baseline
 - Sufficient parallax/head motion for seated(+) VR
 - Easier to scale resolution
- Still not $360^\circ \theta, \phi$ extent
 - Wider FOV lenses



EXAMPLE: LYTRO IMMERGE

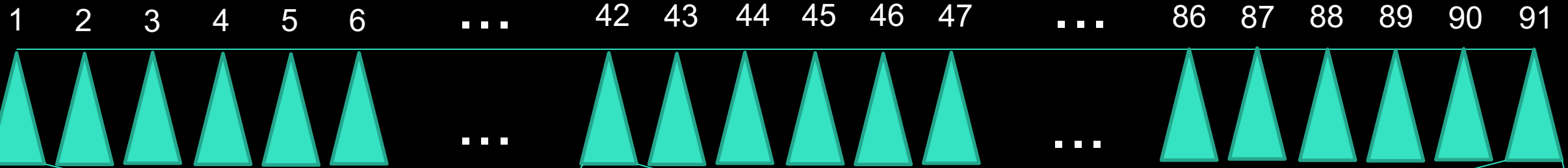


[Image © Lytro]

Multi-camera light field imager
95 cameras

- Larger baseline
 - Sufficient parallax/head motion for seated(+) VR
 - Easier to scale resolution
- Still not $360^\circ \theta, \phi$ extent
 - Wider FOV lenses
- Better for VR
 - Larger 'headbox'
 - But need to interpolate farther between views

EXAMPLE UNPROCESSED LIGHT FIELD VIDEO

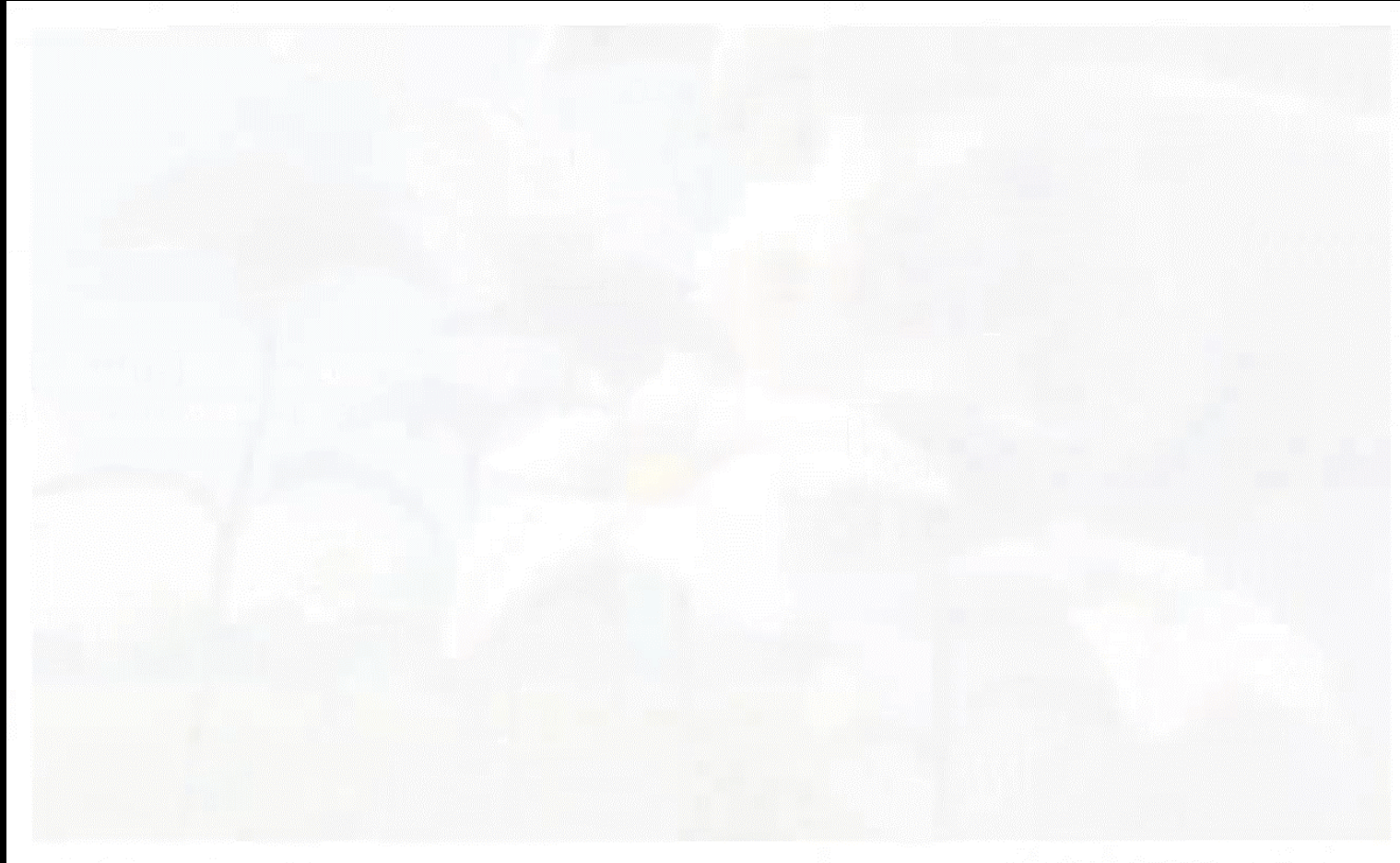


PROCESSING LIGHT FIELDS



- A lot of data!
 - 100x the cameras = 100x the data
 - VR needs high frame rate – 90+ fps
 - 100 camera 1080p 8bit RGB 90fps
= 52 Gbytes per second RAW
- Many operations we take for granted on 1-camera video are not well defined on light fields!
 - Example: filtering

LIGHT FIELD VIDEO – CYCLE THROUGH CAMERA VIEWS



[Bonneel, Tompkin, Wang et al., Eurographics 2017,
Consistent Video Filtering for Camera Arrays]

[Big Buck Bunny / Kovács et al. / Holografika]

LIGHT FIELD VIDEO FILTERING INTRINSIC DECOMPOSITION (SHADING)



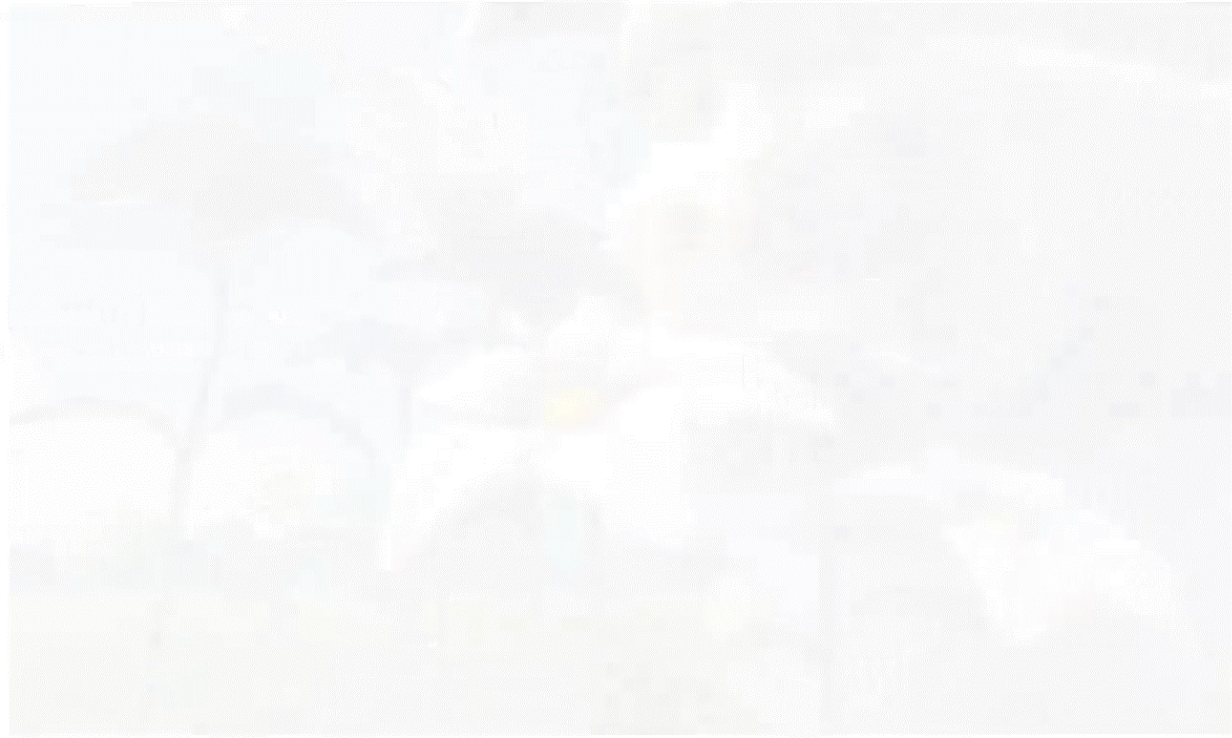
Inconsistency across space and across time!

[Bonneel, Tompkin, Wang et al., Eurographics 2017,
Consistent Video Filtering for Camera Arrays]

[Big Buck Bunny / Kovács et al. / Holografika]

LIGHT FIELD VIDEO FILTERING

INTRINSIC DECOMPOSITION (SHADING)



Input



Processed

[Bonneel, Tompkin, Wang et al., Eurographics 2017,
Consistent Video Filtering for Camera Arrays]

[Big Buck Bunny / Kovács et al. / Holografika]

LIGHT FIELD VIDEO FILTERING

SPATIO-TEMPORAL CONSISTENCY



Our result

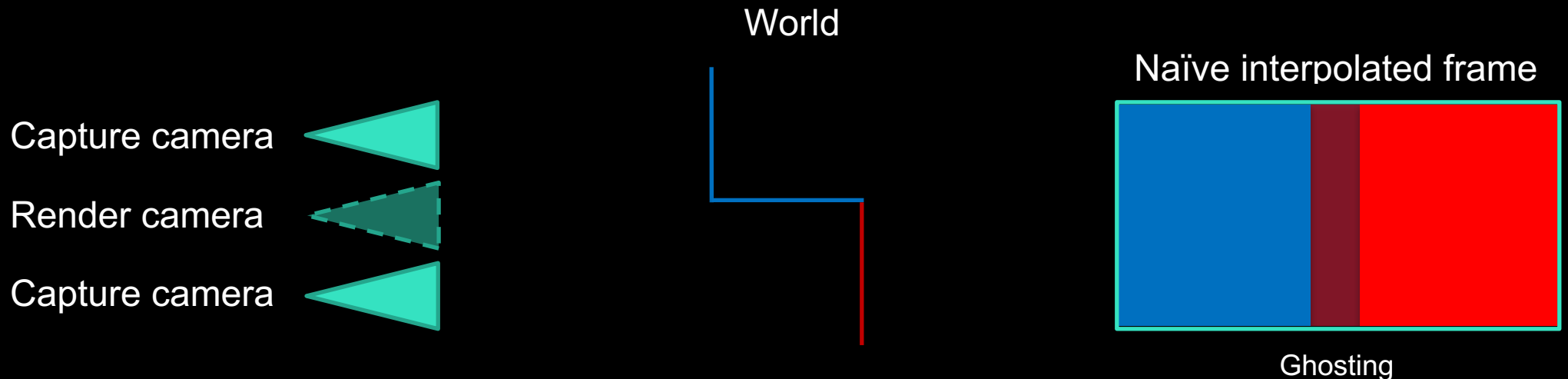
[Bonneel, Tompkin, Wang et al., Eurographics 2017,
Consistent Video Filtering for Camera Arrays]

[Big Buck Bunny / Kovács et al. / Holografika]

PROCESSING LIGHT FIELDS



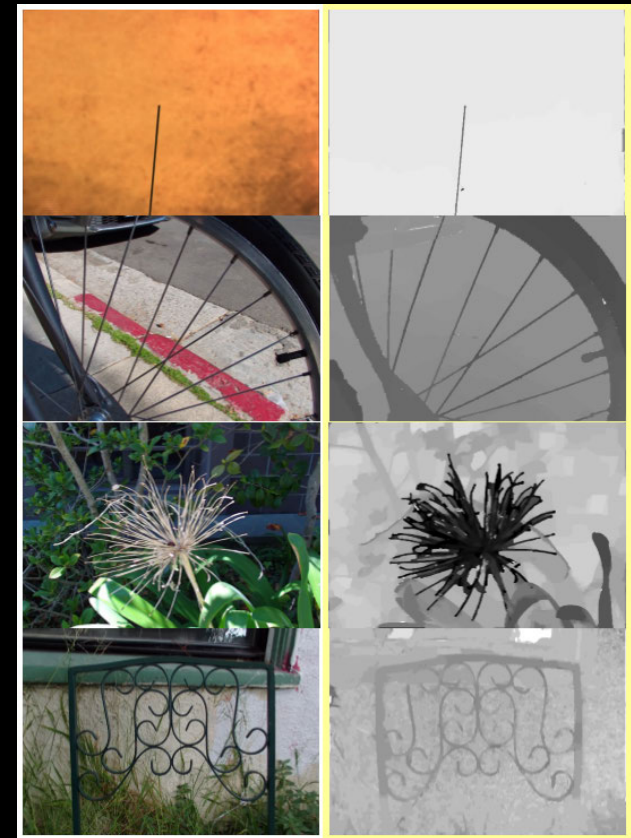
- Let's go back to spatial interpolation.
- What happens when my VR head goes 'in between' the cameras?
 - Interpolation -> ghosting; use flow to reduce
 - Instead, try to recover depth and geometry



DEPTH MAP / GEOMETRY EXTRACTION



- Same as in stereo case, but now with more cameras
 - More data to process but more views lead to less occluded pixels + robustness
- Many techniques:
 - Yucer et al., SIGGRAPH 2016
 - Wang et al., TPAMI 2016
 - Jeon et al., CVPR 2015
 - Kim et al., SIGGRAPH 2013, 3DV 2016



LF input

Wang et al.

PROCESSING LIGHT FIELDS



-
- Let's go back to spatial interpolation.
 - What happens when my VR head goes 'in between' the cameras?
 - What happens when my VR head goes 'beyond' my baseline?
 - Reveal content not captured by any cameras
 - Must 'fake it' or 'hallucinate' – *inpainting only goes so far*

OTHER TECHNOLOGIES: LIDAR, DEPTH CAMERAS, OTHER SCANNERS

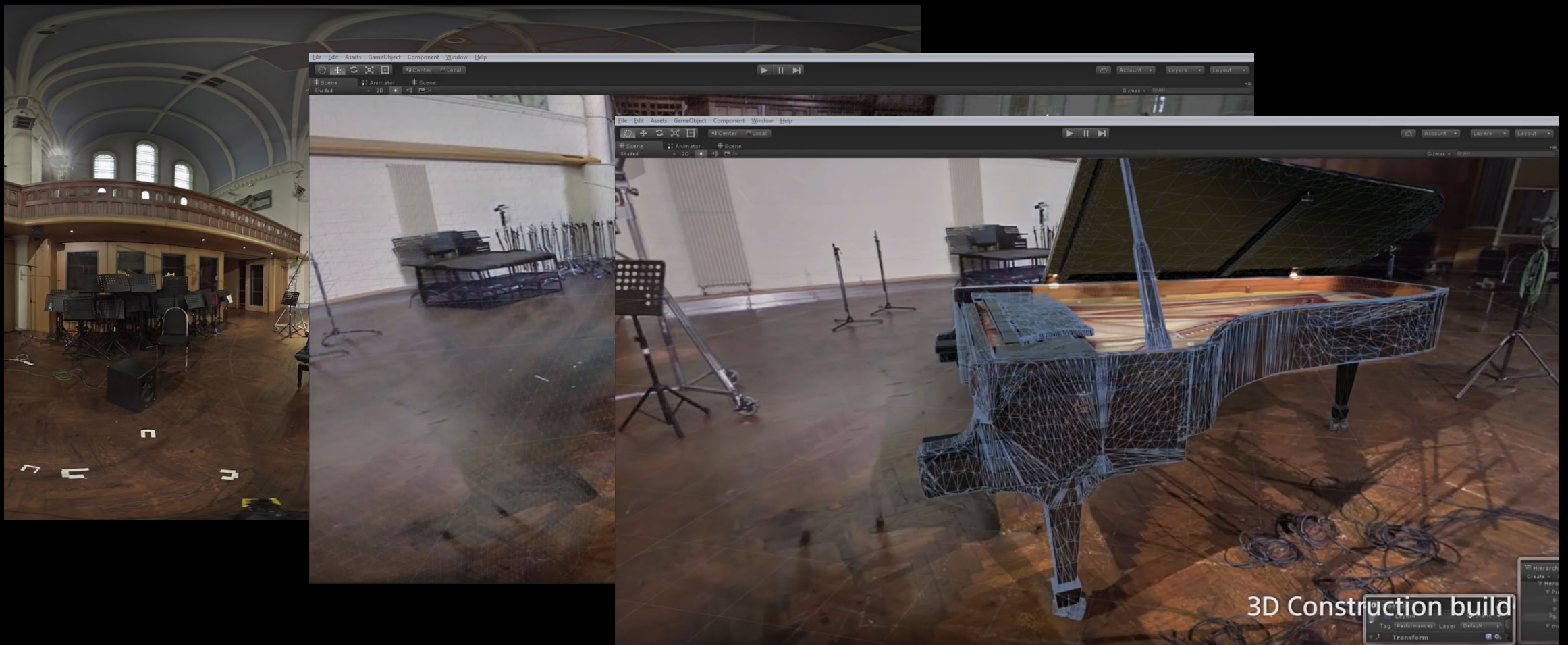


- Use other modes to fill in:
 - Geometry from lidar to generate ‘clean backplate’ for depth.
 - Fit known geometric models to scene.
- Even with dense sampling, sometimes we can’t reconstruct depth or geometry:
 - Other capture modes can help
 - Complex reflection, e.g., anisotropic
 - Complex materials, e.g., translucent



[FARO Focus laser scanner]

SONY 'JOSHUA BELL'

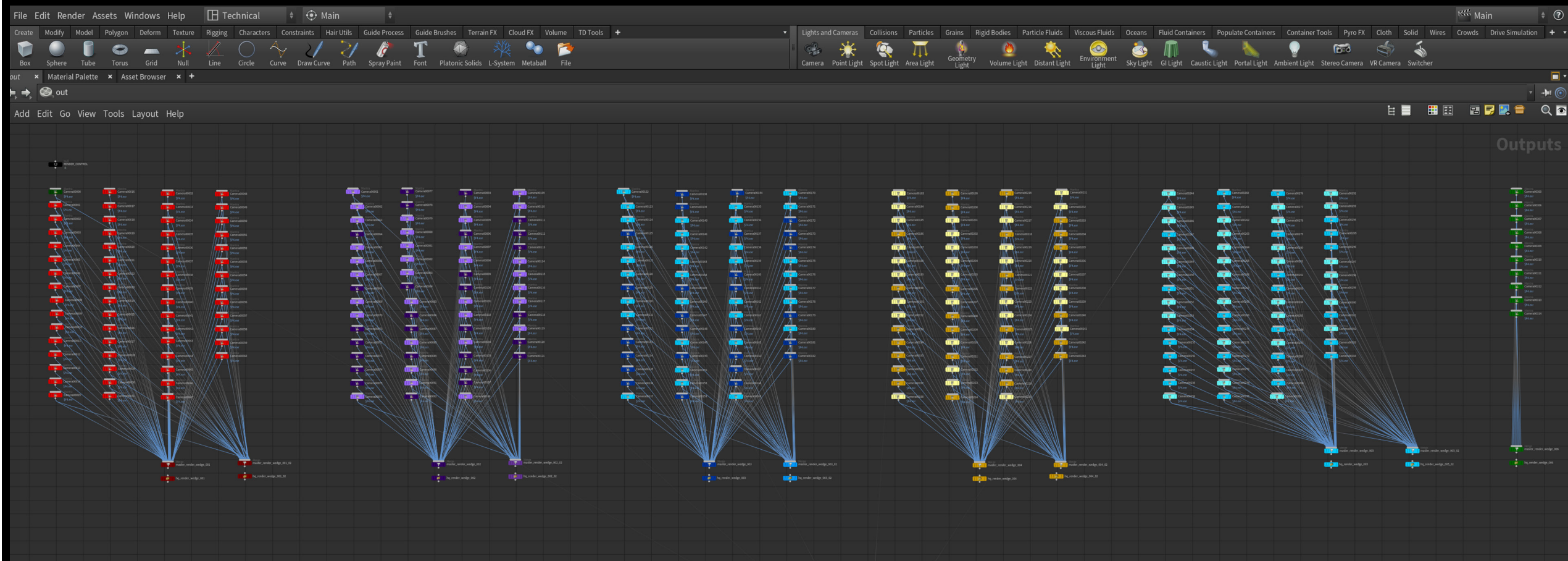




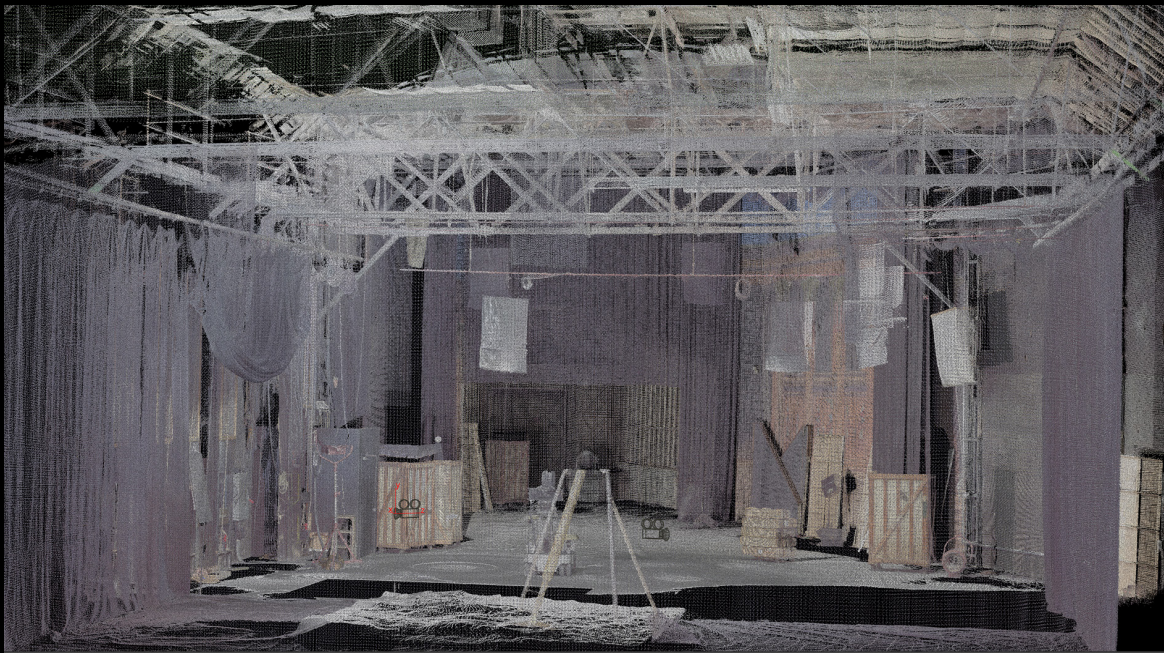
VR Playhouse

NOTICE
NO LOITERING OR
OPEN ALCOHOLIC
CONTAINERS

- Rendering For Multi-Camera Lightfield Production



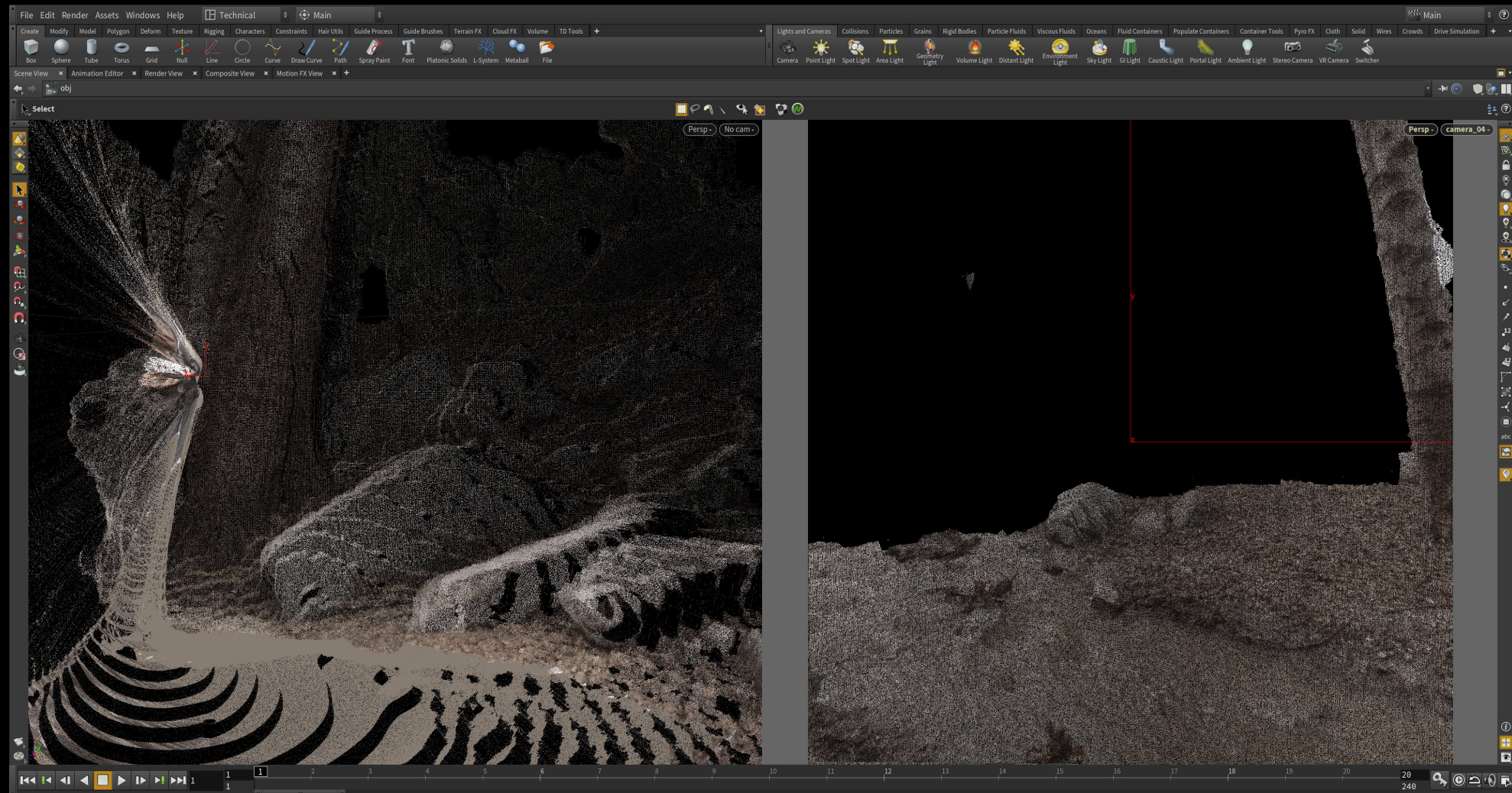
- Geometry Reconstruction Techniques
- Filling In Missing or Incomplete data
- Photogrammetry For Object Re-Construction



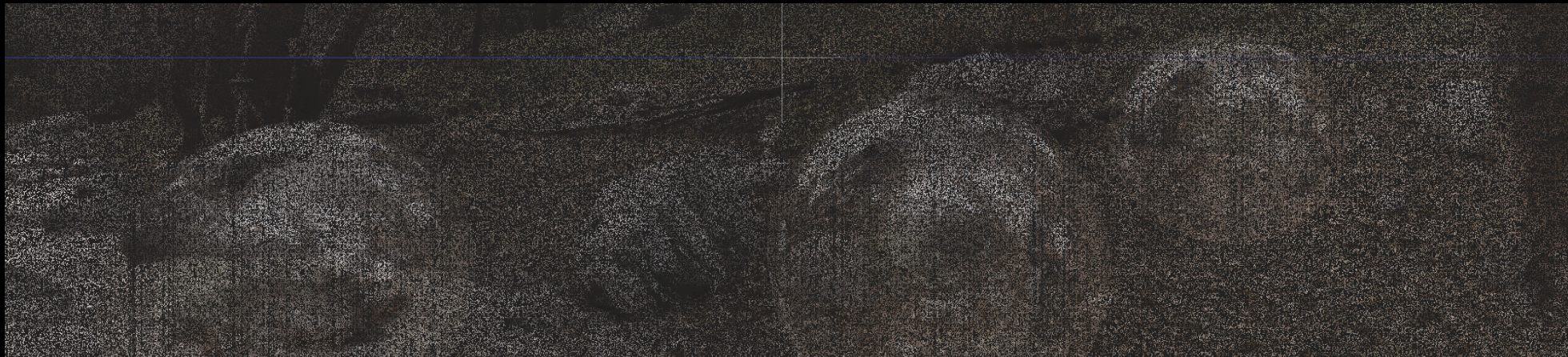
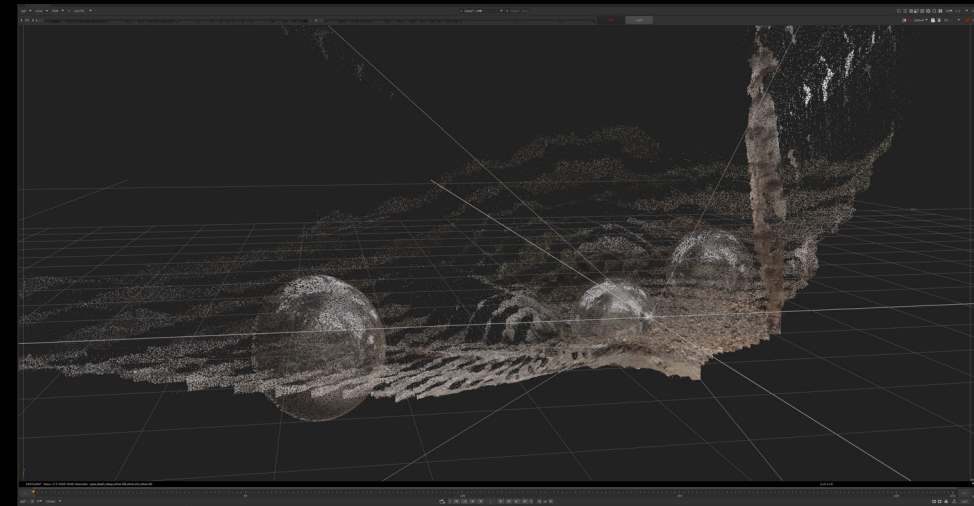
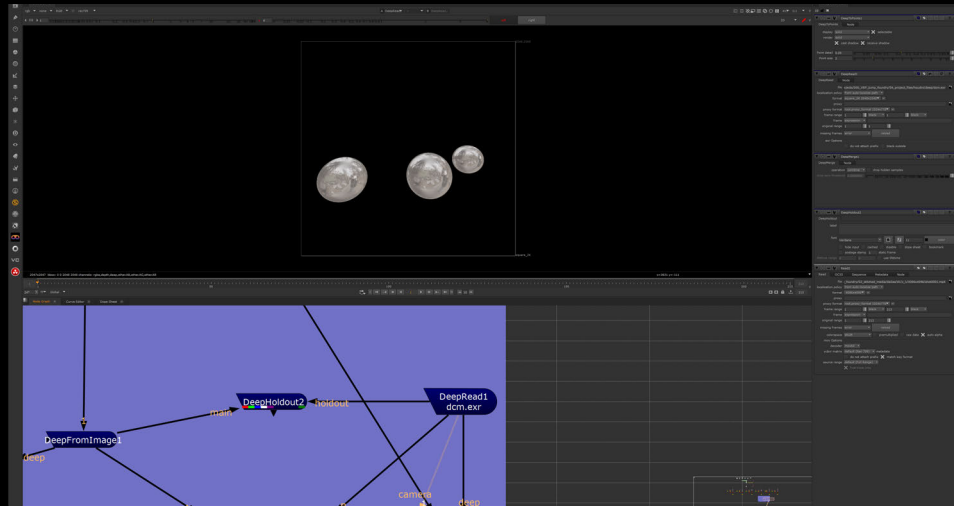
- Top Bottom Stereo
- Outputs Depth as 8 bit Float Data
- Nadir Fixes and Stereo to Mono Blending



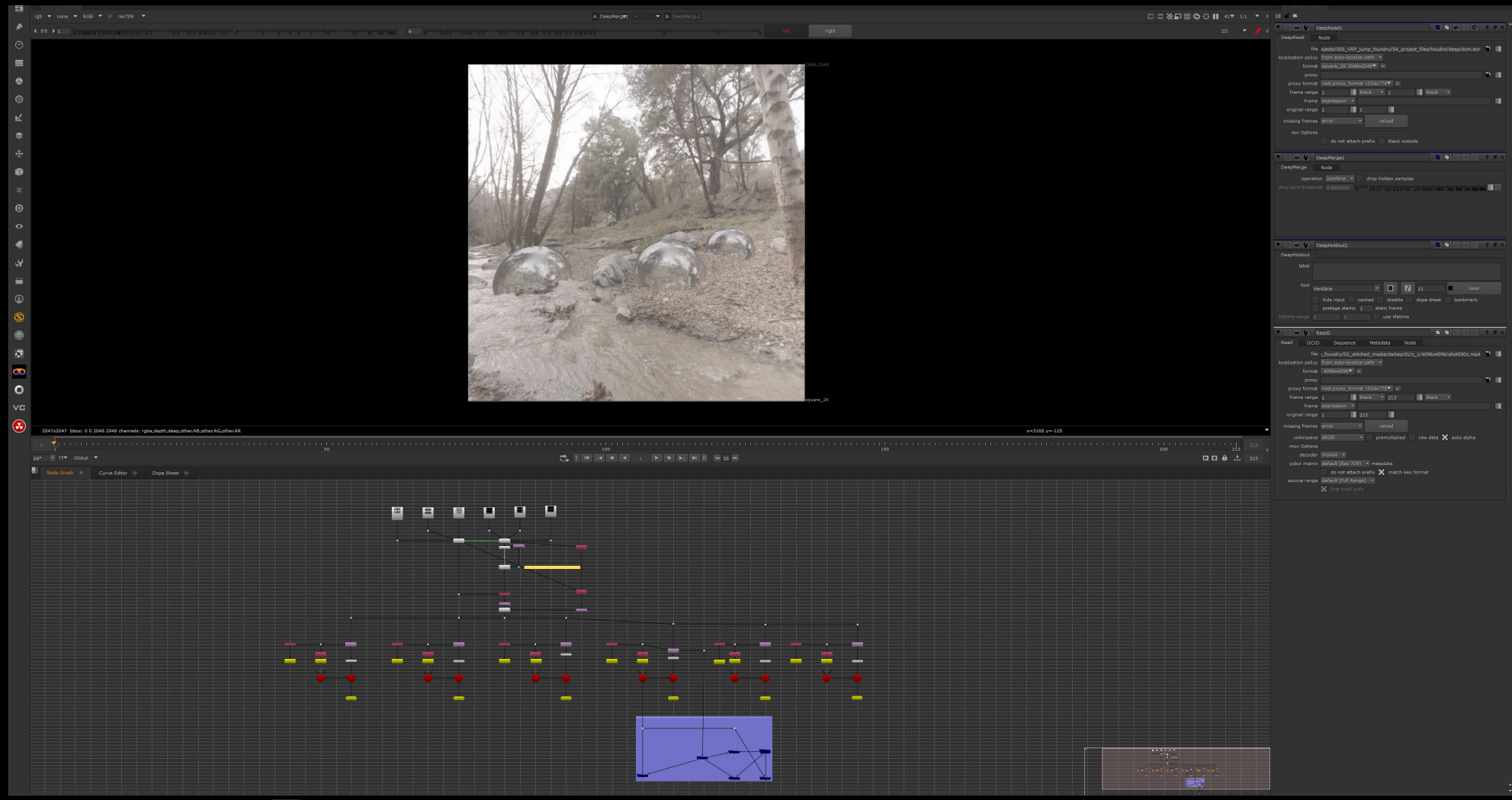
- Point Cloud Re-Construction In Houdini for CGI Integration
- Point Cloud As Seen Through Both The Viewing Camera And In The Scene



- Point Cloud Re-Construction In Houdini for CGI Integration
- Point Cloud As Seen Through Both The Viewing Camera And In The Scene

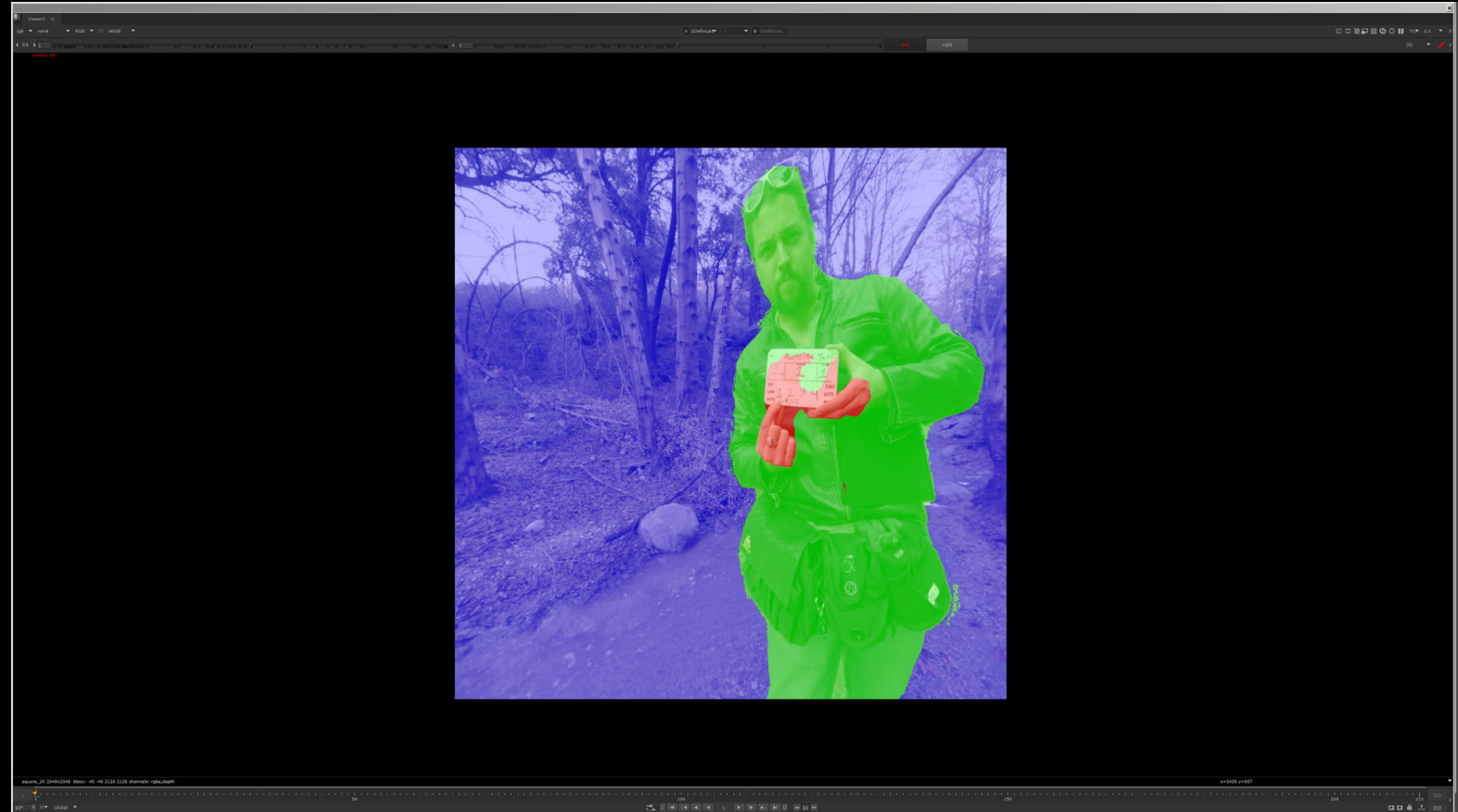
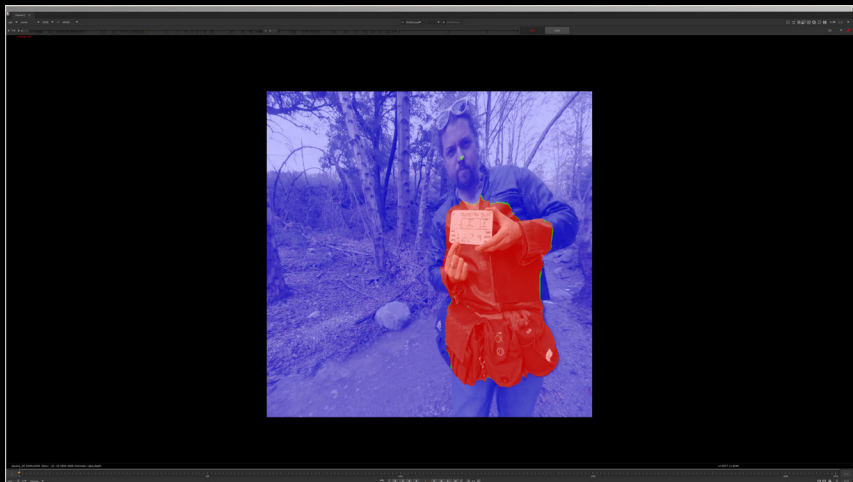


- Deep Based Compositing Techniques for CGI Integration
- Spatial Rectification Between The Lat-Long World and The CGI World
- Matching Z-Depth For Correct Stereo Between CGI and Shot Lat Long



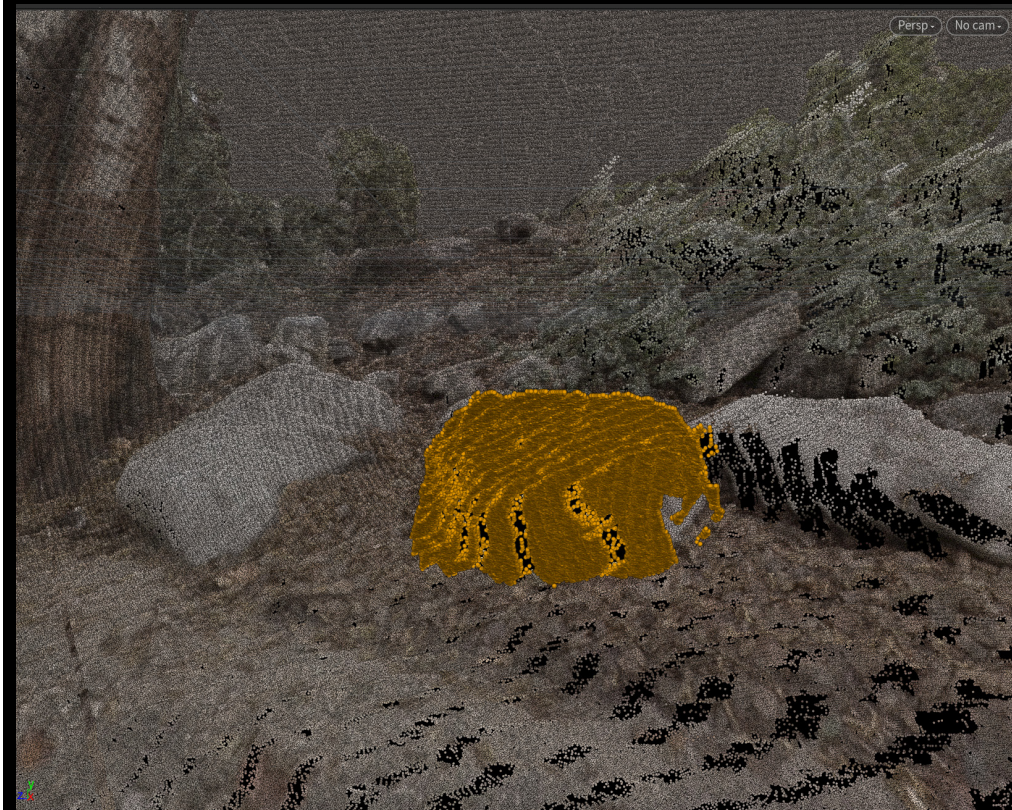
OBJECT EXTRACTION AND ISOLATION THROUGH DEPTH SLICING

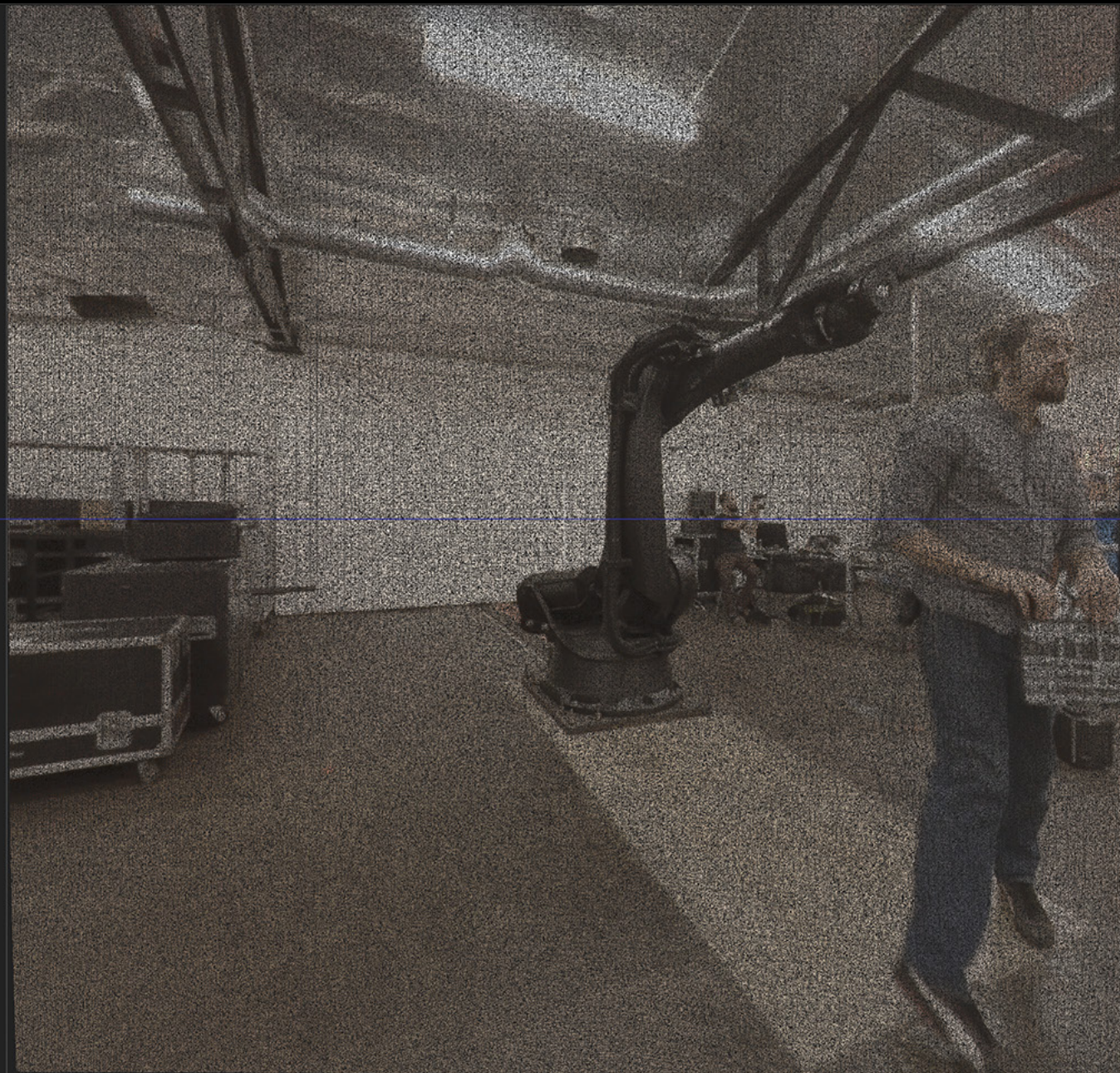
- Automatic Rotoscoping Through Depth Data
- Altering Depth of Field
- Relighting Effects Through Normals

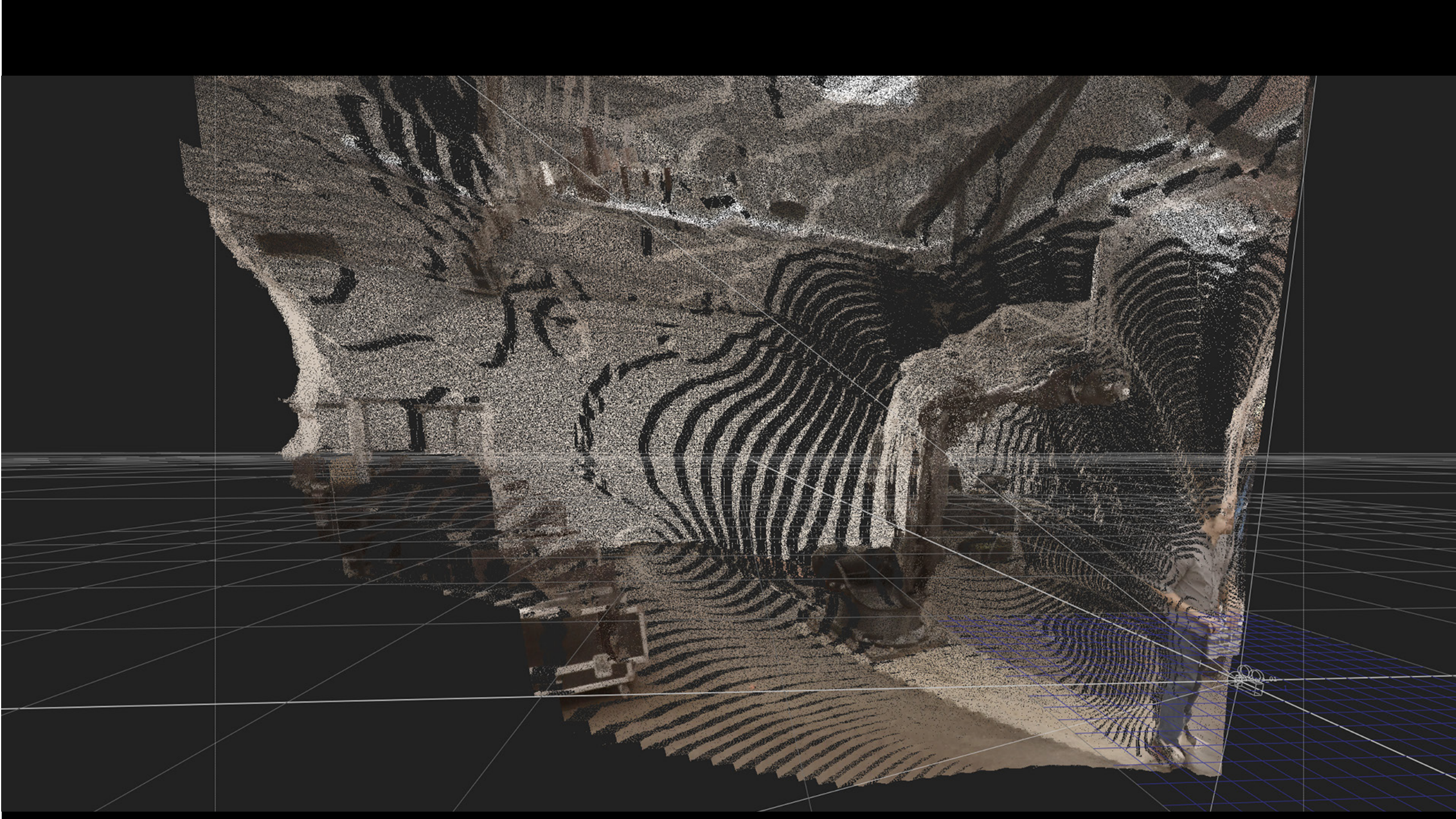


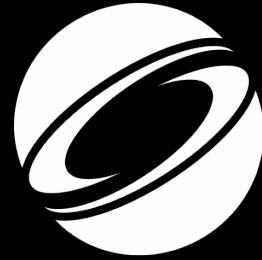
GEOMETRY RE-CONSTRUCTION THROUGH POINT CLOUD MESHING AND TEXTURE RE-PROJECTION

- Using the Google Jump Point Clouds For Geometry Isolation
- Using the Google Jump Point Clouds For Geometry Creation
- Texture Projection









30 JULY – 3 AUGUST *Los Angeles*
SIGGRAPH2017

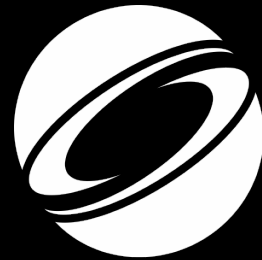
VIDEO FOR VIRTUAL REALITY
SUMMARY AND OUTLOOK

JAMES TOMPKIN

THE COURSE



- Introduction
- Technical Foundations
- Current Practice
- Future Systems
- Summary and outlook
 - How far do we have to go?
 - Q & A



30 JULY – 3 AUGUST *Los Angeles*
SIGGRAPH2017

Final reminder: This is a sample of the notes that we will present at the course. Unfortunately, some content could not be approved for release on time for the ACM Digital Library deadline, so please visit the course website for the complete version.



30 JULY – 3 AUGUST *Los Angeles*
SIGGRAPH2017

VIDEO FOR VIRTUAL REALITY

SIGGRAPH 2017 COURSE

<http://richardt.name/pub/Video4VR/>
