
thrive
SIGGRAPH2019
LOS ANGELES • 28 JULY – 1 AUGUST
Capture4VR
From VR Photography to VR Video
richardt.name/Capture4VR

Copyright notice

SIGGRAPH 2019 Courses, July 28 – August 1, 2019, Los Angeles, CA, USA

ACM ISBN 978-1-4503-6307-5/19/07.

https://doi.org/10.1145/3305366.3328028

Who are we?

Christian Richardt

UNIVERSITY OF BATH

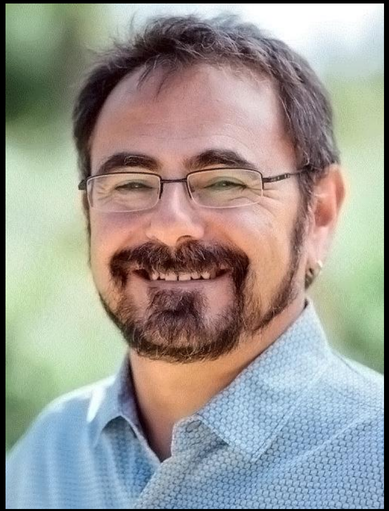
Peter Hedman

UCL

Ryan S. Overbeck

Google

Brian Cabral



facebook

Robert Konrad

Stanford University

Steve Sullivan

Microsoft

Virtual reality

- Offers unparalleled immersion
- Still a young medium
- We need new ways:
  - to author content
  - to capture from the real world
- In this course, we:
  - provide an overview of progress in VR photography and VR video
  - discuss state-of-the-art systems by Facebook, Google + Microsoft

Oculus

Course schedule

Start		Topic	Speaker
14:00	1.	Introduction	Christian Richardt, Bath
14:20	2.	360° (Stereo) Panoramas	Christian Richardt, Bath
14:40	3.	3D Photography	Peter Hedman, UCL
15:00	4.	Light Field Photography	Ryan S. Overbeck
15:20		Q&A + *Break*	
15:35	5.	360 and ODS Video	Brian Cabral, Facebook
15:55	6.	Live ODS Video	Robert Konrad, Stanford
16:15	7.	6-DoF Video	Brian Cabral, Facebook
16:35	8.	MR Capture Studios	Steve Sullivan, Microsoft
16:55	9.	Conclusion + Q&A	All presenters

1. Introduction

Christian Richardt

- Welcome + introduction of co-presenters
- Overview and structure of this course
- A brief history of panoramas and stereoscopy

UNIVERSITY OF BATH

2. 360° (Stereo) Panoramas

- Traditional panorama stitching
  - applications in consumer devices
- Omnidirectional stereo (ODS) panoramas
  - Omnistereo
  - Megastereo
- Panoramas with motion parallax
  - Parallax360
  - MegaParallax

Christian Richardt

UNIVERSITY OF BATH

3. 3D Photography

Peter Hedman

- VR photography with textured 3D reconstructions from hand-held photos:
  - Casual 3D photography
  - Instant 3D Photography
- More immersive exploration in VR:
  - enables full 6 degrees-of-freedom (6-DoF) head motion
  - allows users to look around freely in VR

4. Light field photography

- Introduction to light fields
- Google's panoramic light fields:
  - capturing ~1000 photos of a scene
  - process using structure-from-motion and multi-view stereo
  - real-time rendering for VR HMDs
- Extensions to light field video
- DeepView:
  - view synthesis with multi-plane images

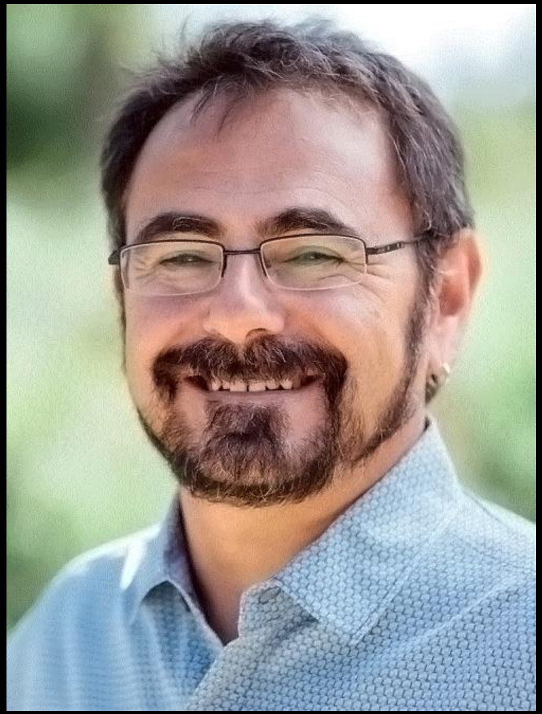
Ryan S. Overbeck

Google

5. 360 and ODS video

- Moving from still images to moving pictures
- 360° video
  - affordable consumer 360° cameras
  - but lacks depth
- Omnidirectional stereo (ODS)
  - produces stereoscopic 360° video
  - but requires multi-camera rigs
  - e.g. Facebook Surround 360 or Google Jump

Brian Cabral



facebook

6. Live ODS video

- ODS video approaches usually require expensive off-line processing
  - prevents live streaming
- Overview of live streaming ODS:
  - live streaming ODS camera arrays
  - single-shot ODS systems
  - rotating systems for ODS capture

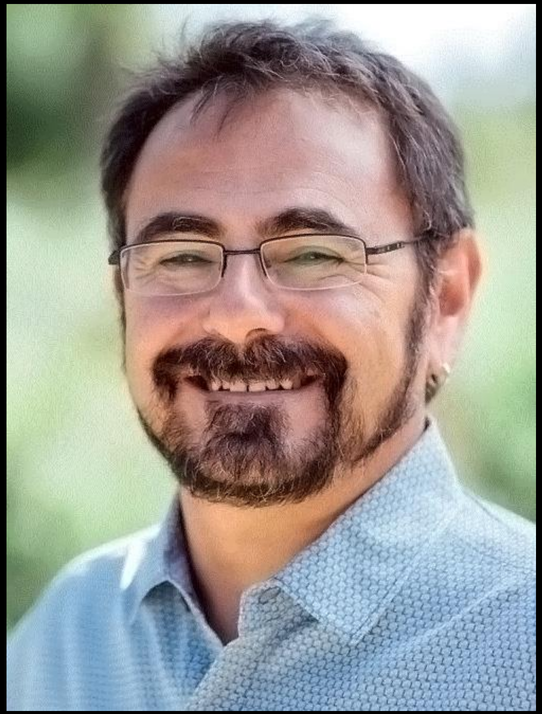
Robert Konrad

Stanford University

7. 6-DoF video

- Need support for 6-DoF motion within VR video
  - so that viewers can move their head around freely to explore a scene
- Facebook’s latest camera rigs and techniques:
  - x6, x24 and Manifold cameras
  - 6-DoF video techniques and results

Brian Cabral



facebook

8. Microsoft Mixed Reality Capture Studios

- We also want to capture objects realistically
  – e.g. people and animals
- Volumetric video capture using outside-in camera arrangement
- Microsoft Mixed Reality Capture Studios:
  – state-of-the-art commercial facilities
  – overview of underlying technology
  – ‘holograms’ can be inserted into VR/AR experiences

Steve Sullivan

Microsoft

9. Conclusion + Q&A

- Short summary
- Discussion of remaining challenges towards ubiquitous 6-DoF VR photography and video
- Questions & answers

Christian Richardt

UNIVERSITY OF BATH

Course schedule

Start		Topic	Speaker
14:00	1.	Introduction	Christian Richardt, Bath
14:20	2.	360° (Stereo) Panoramas	Christian Richardt, Bath
14:40	3.	3D Photography	Peter Hedman, UCL
15:00	4.	Light Field Photography	Ryan S. Overbeck
15:20		Q&A + *Break*	
15:35	5.	360 and ODS Video	Brian Cabral, Facebook
15:55	6.	Live ODS Video	Robert Konrad, Stanford
16:15	7.	6-DoF Video	Brian Cabral, Facebook
16:35	8.	MR Capture Studios	Steve Sullivan, Microsoft
16:55	9.	Conclusion + Q&A	All presenters

thrive
SIGGRAPH2019
LOS ANGELES • 28 JULY – 1 AUGUST

Christian Richardt

A Brief History of VR Photography + Video

CAMERA
Centre for the Analysis of Motion, Entertainment Research and Applications

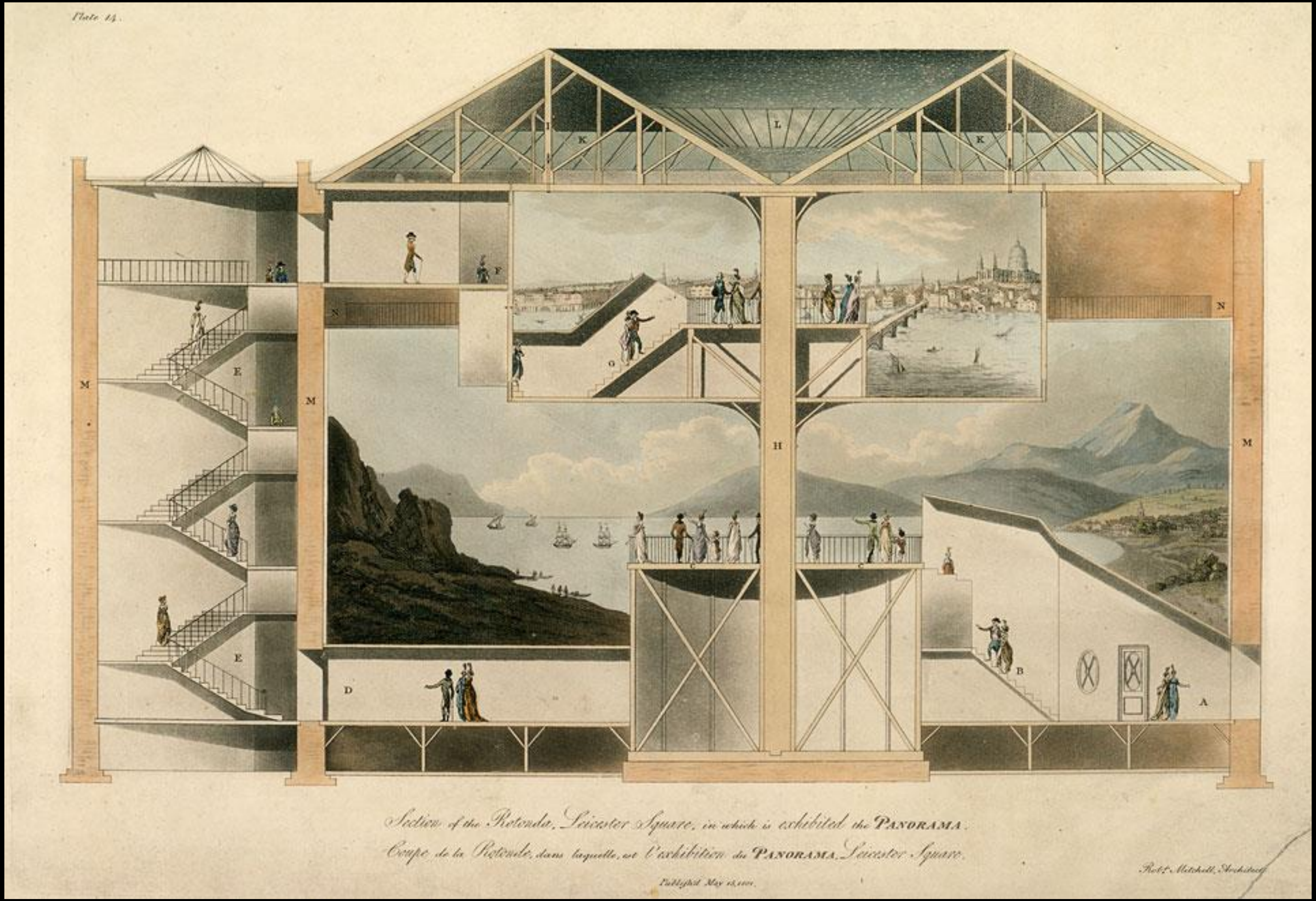
UNIVERSITY OF BATH

Panorama

- formed from Greek πᾶν "all" + ὅραμα "sight"
- term coined by painter Robert Barker in 1792
- definition: any wide-angle view or representation of a physical space

Panoramic view of London, from the top of Albion Mills, by Robert Barker, 1792

The Rotunda in Leicester Square (1793–1863)



Robert Mitchell, 1901

Panorama Mesdag (1881)

120 × 14 metres — painted in 1880–1881 by Hendrik Willem Mesdag

Gettysburg Cyclorama (1883)

115 × 13 metres — painted in 1882–1883 by Paul Philippoteaux

Stereoscopy

- Technique for creating or enhancing the illusion of depth in an image by means of stereopsis for binocular vision

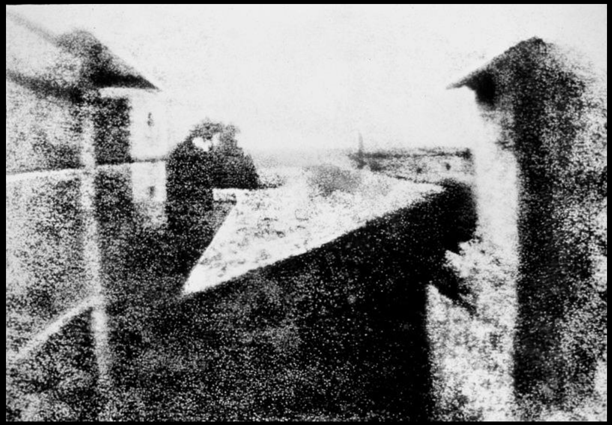
- from Greek:
  - στερεός (stereos) — 'firm, solid'
  - σκοπέω (skopeō) — 'to look, to see'

Charles Wheatstone
(1802–1875)

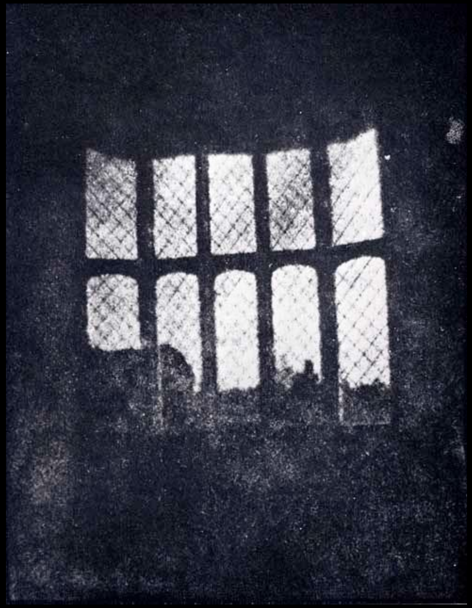
Wheatstone stereoscope (1838)

© David Tett/King's College London

A brief history of photography



1826
First photograph
(Nicéphore Niépce)



1835
First negative
(Henry Fox Talbot)

1839
Daguerrotype
(Louis Daguerre)

Wheatstone stereoscope (1838)

© David Tett/King's College London

Brewster stereoscope (1849)

© The Bill Douglas Cinema Museum, University of Exeter

David Brewster
(1781–1868)

Holmes stereoscope (1861)

Oliver Wendell Holmes
(1809–1894)

Holmes stereoscope (1861)

Stereo photography = Stereography

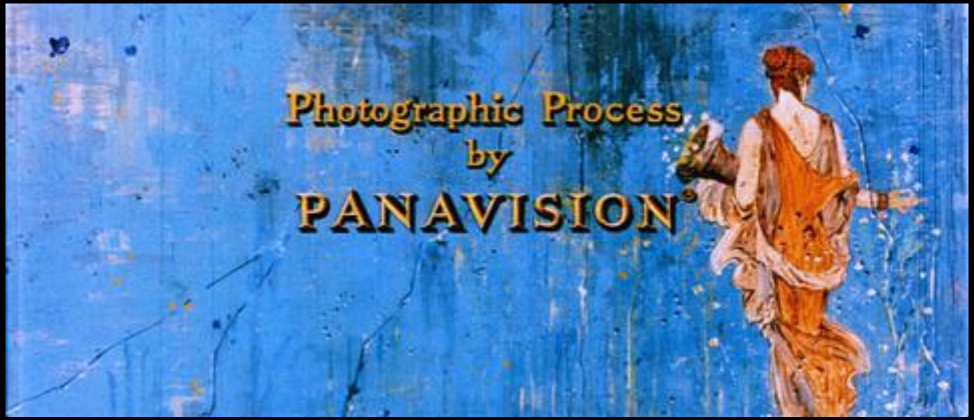
Underwood & Underwood, Publishers.
New York, London, Toronto-Canada, Ottawa-Kansas.
Works and Studios
Arlington, N.J. Littleton, N.H. Washington, D.C.
Victoria, Queen and Empress from 1837 to 1901 - England's most beloved and longest reigning Sovereign.
Copyright 1901 by Underwood & Underwood.
(40) Avenue of Palms, Los Angeles, California.
Copyright 1899 by Strohmeyer & Wyman.
East River Bridge, N. Y.
Trocadero Entrance to the Exposition, Colonial Section in Foreground, Paris, 1900.
(88) Milan's Cathedral, Italy.
Copyright 1897 by Underwood & Underwood.
A Mighty Monument to Pagan Brutality—the Colosseum (E.), Rome, Italy.
1302—The Taj Mahal, Agra, India.
Copyright 1906, by W. S. Smith.

Motion Pictures (1890s)

The U.S. National Archives (1926)

Widescreen motion picture film formats

- First success in late 1920s and before 1932 Great Depression:
  - e.g. NaturalVision, Fox Grandeur, Magnifilm (2:1 aspect ratio)
  - Warner Vitascope, MGM Realife
- Second wave in 1950s and 60s:
  - Ultra Panavision 70 (2.76:1)
    - Ben-Hur (1959)
  - CinemaScope (2.35:1 – 2.55:1)
    - Lady and the Tramp (1955)
  - Super Panavision 70 (2.20:1)
    - 2001: A Space Odyssey (1968)

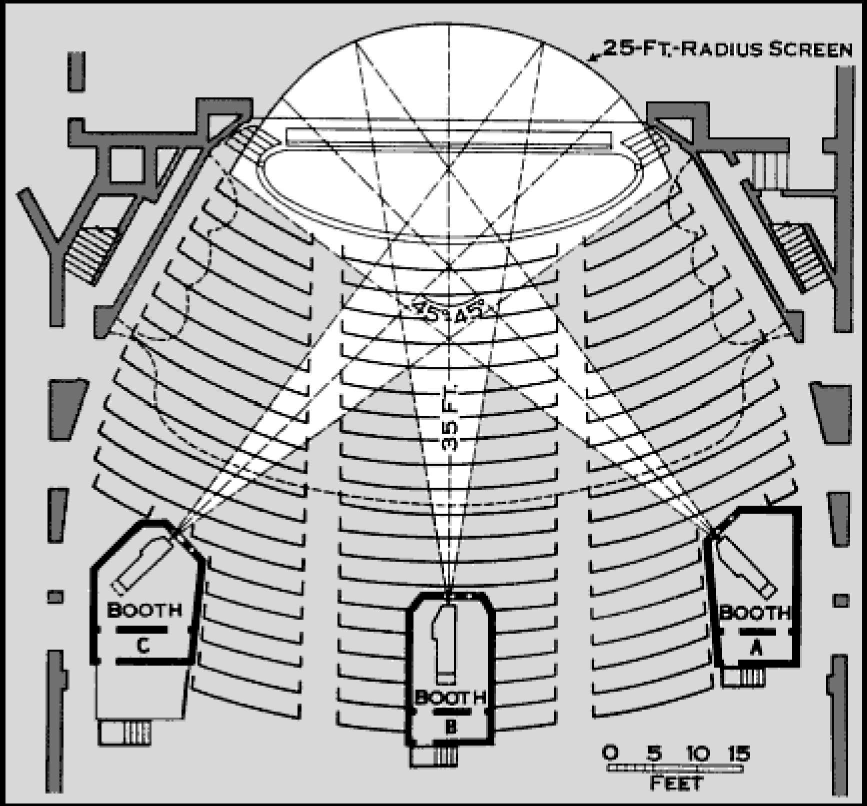


Screenshot of "The Big Fisherman" (1959), the first film released using the Super Panavision 70 process.

Multi-projector projection

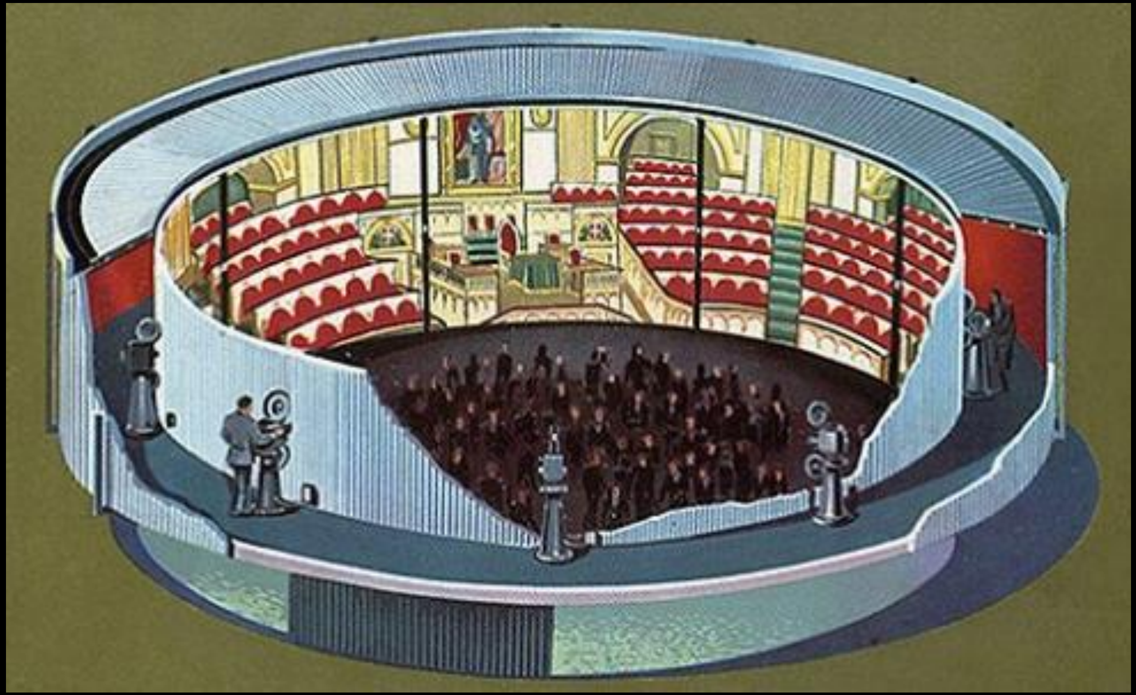
Cinerama (1952)

3× 35mm projectors



Circle-Vision 360° (1955)

9× 35mm projectors



https://www.themeparktourist.com/

thrive
SIGGRAPH2019
LOS ANGELES • 28 JULY – 1 AUGUST

Christian Richardt

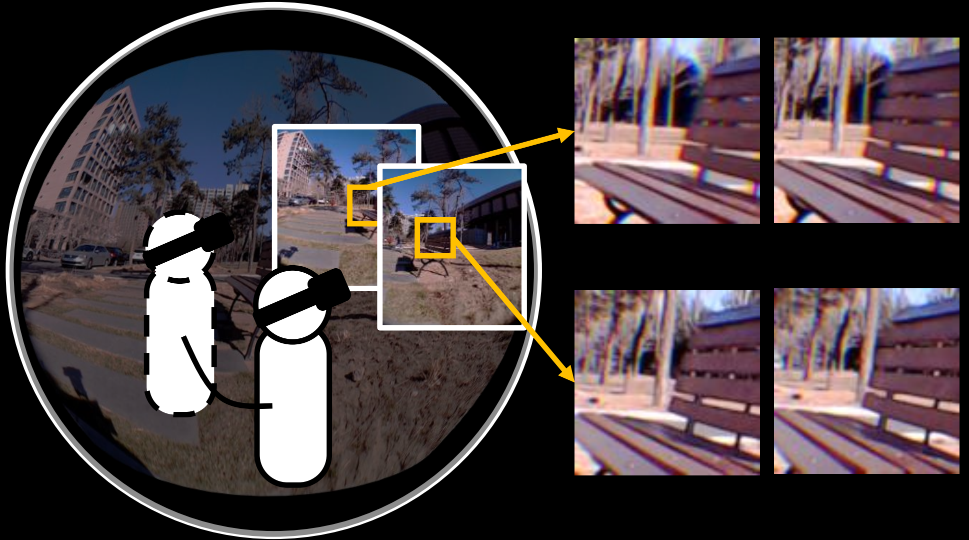
360° (Stereo) Panoramas

CAMERA
Centre for the Analysis of Motion, Entertainment Research and Applications

UNIVERSITY OF BATH

360° (Stereo) Panoramas

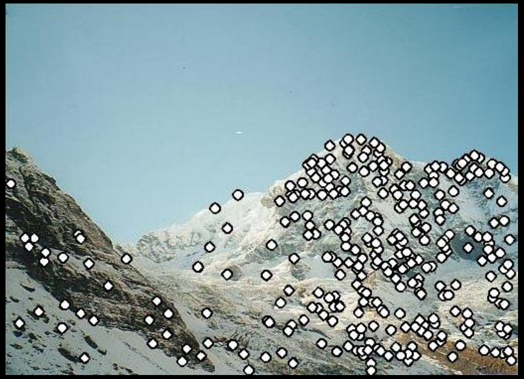
1. 360° panoramas
   - alignment + stitching [Brown & Lowe 2007]
   - parallax-aware stitching [Zhang & Liu, 2014]
2. Stereo panoramas
   - Omnistereo [Peleg et al. 2001]
   - MegaStereo [Richardt et al. 2013]
3. Towards 6-DoF with motion parallax
   - Parallax360 [Liu et al. 2018]
   - MegaParallax [Bertel et al. 2019]



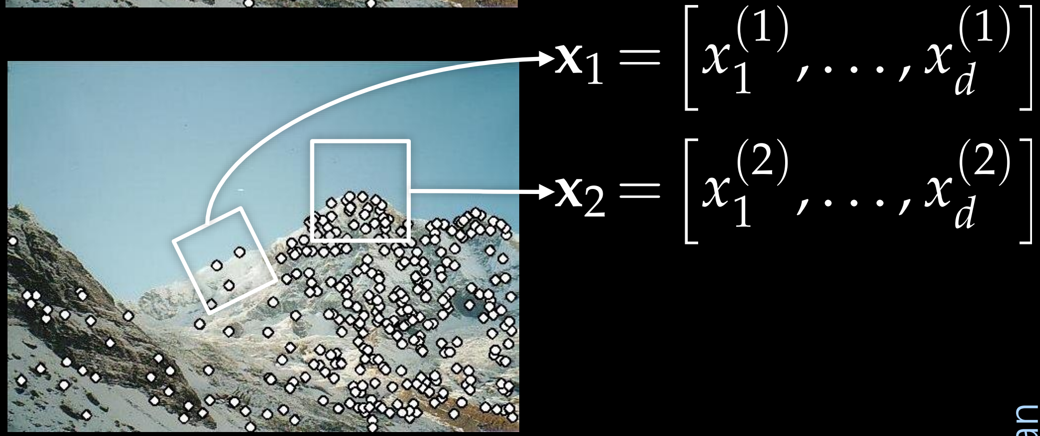
Luo et al., TVCG 2018

Feature matching

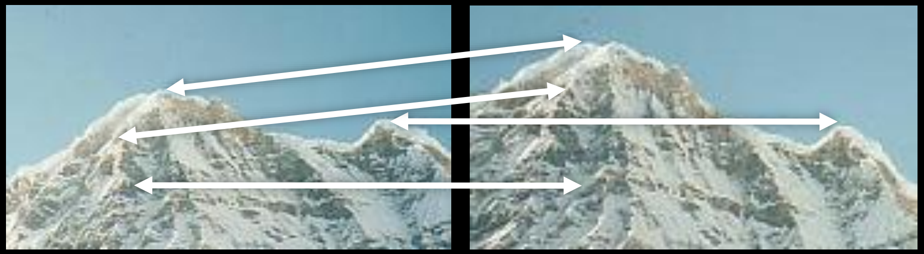
1. **Detection:**
   Identify the interest points



2. **Description:**
   Extract vector feature descriptor surrounding each interest point.

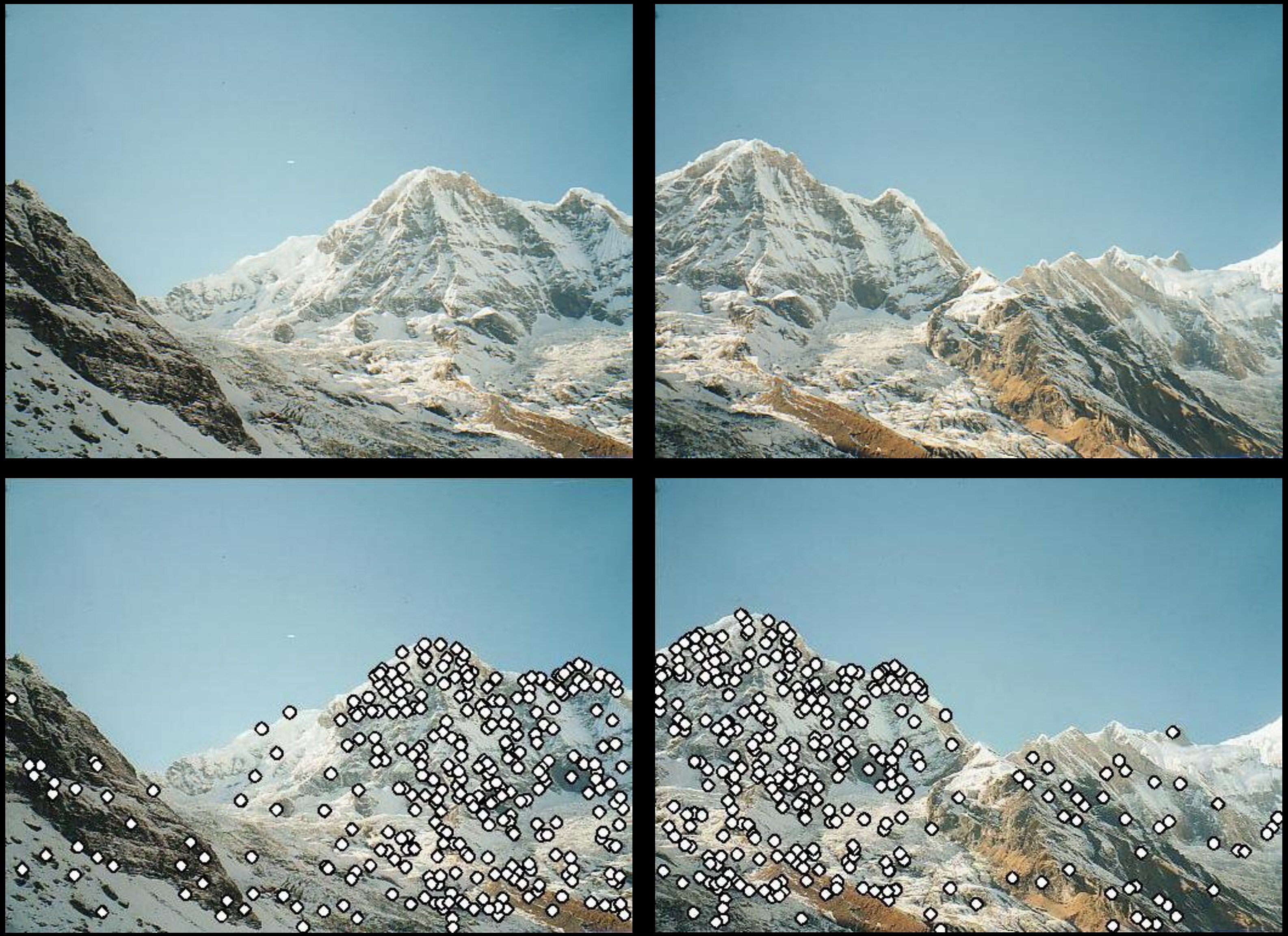


3. **Matching:**
   Determine correspondence between descriptors in 2 views



Slide by Kristen Grauman

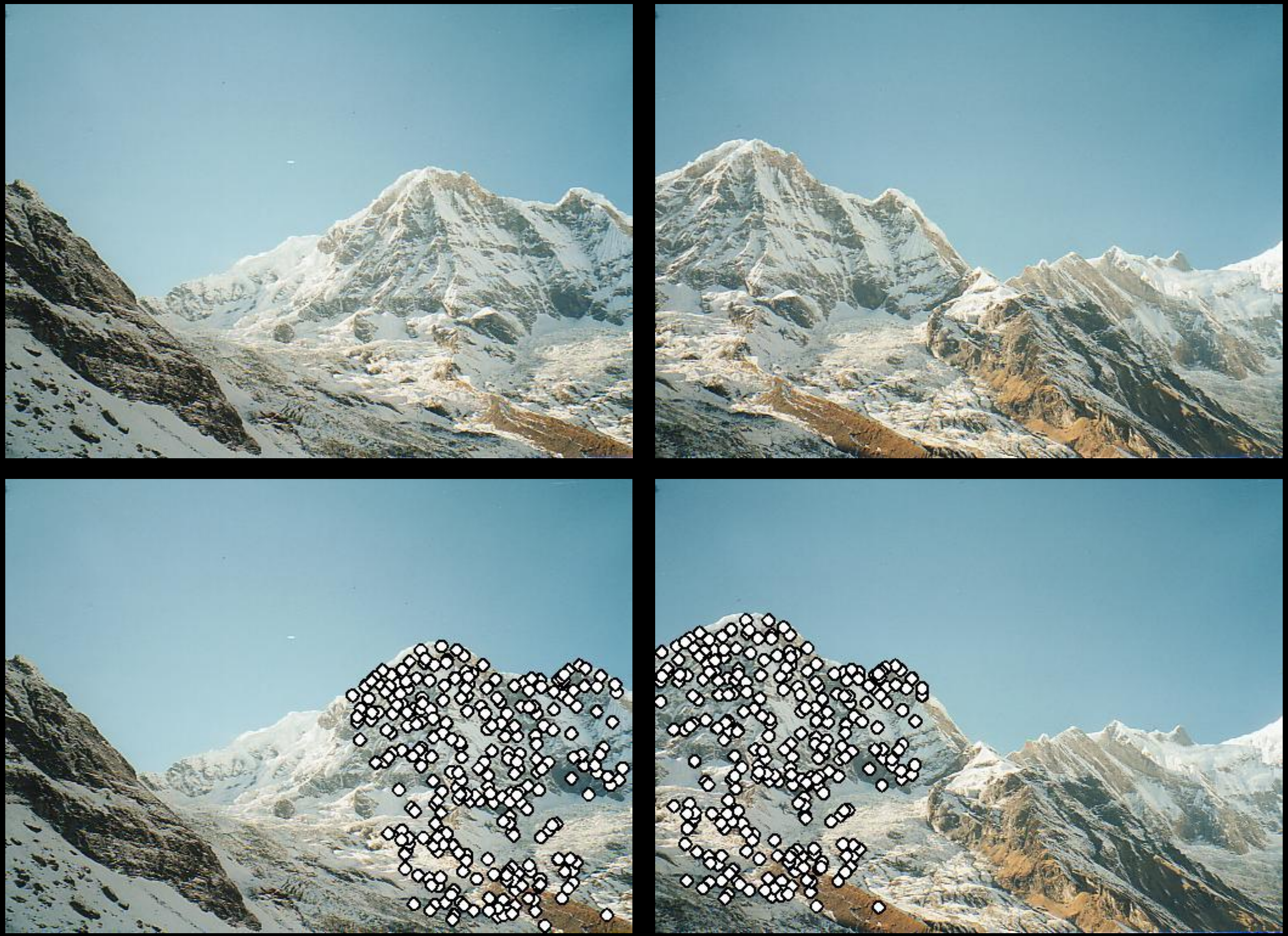
SIFT features



Matthew Brown and David Lowe

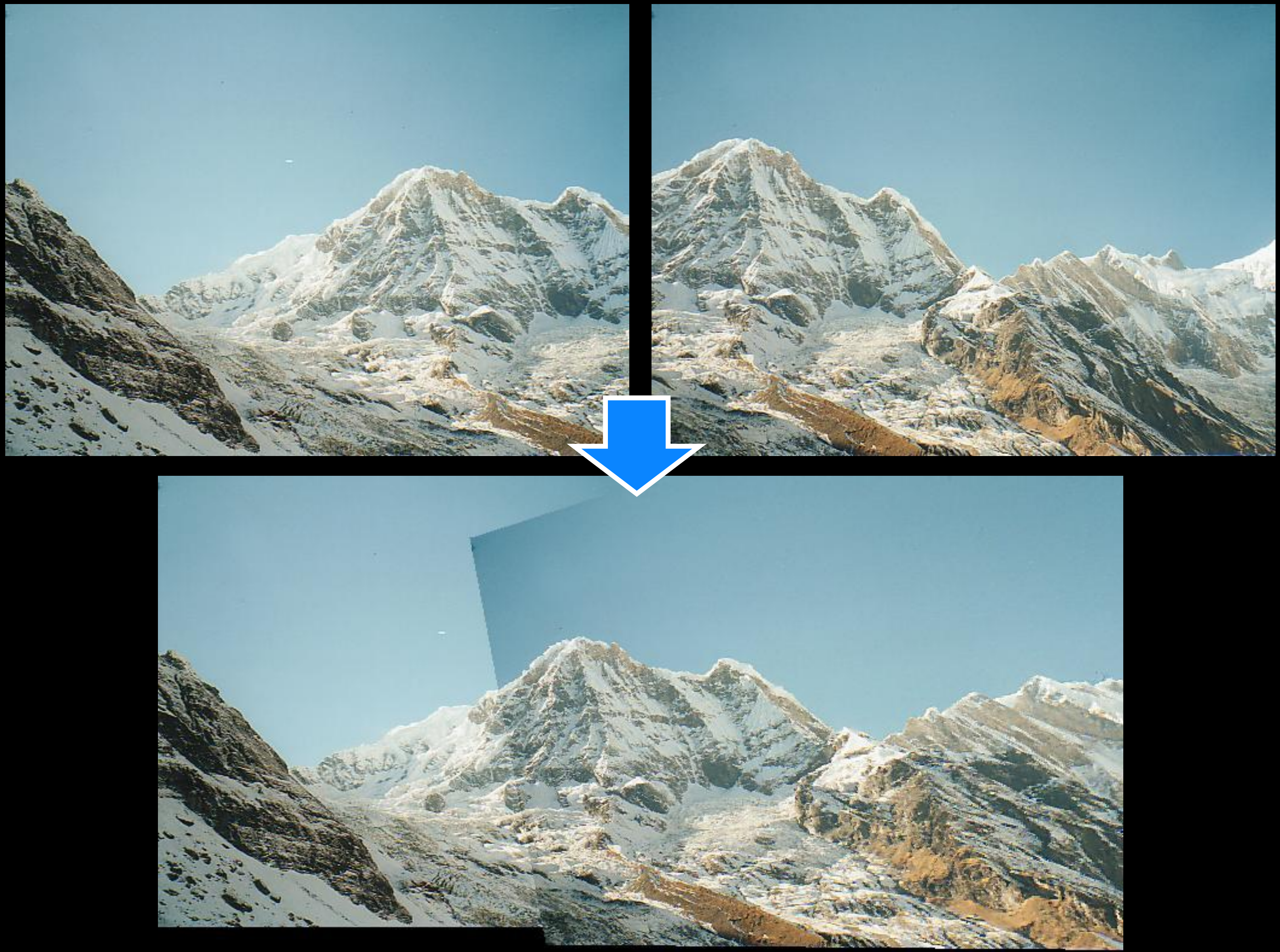
Slide by Matthew Brown

Matched SIFT features



Slide by Matthew Brown

Aligned images



Matthew Brown and David Lowe

Slide by Matthew Brown

Image alignment

Slide by Matthew Brown

Image blending

- Multi-band blending [Burt & Adelson, TOG 1983]

Automatic Panoramic Image Stitching using Invariant Features
Matthew Brown & David G. Lowe
International Journal of Computer Vision, 2007

Image alignment and stitching: a tutorial
Richard Szeliski
Foundations and Trends in Computer Graphics and Vision, 2006

Slide by Matthew Brown

Parallax-aware stitching

- image alignment generally relies on homography estimates
  - perfect for camera rotation or planar scene content
  - but problematic for photos that are captured handheld
- need to explicitly handle parallax between images
  - e.g. Parallax-tolerant Image Stitching [Zhang & Liu, CVPR 2014]

Brown & Lowe 2007

Zhang & Liu 2014

Applications

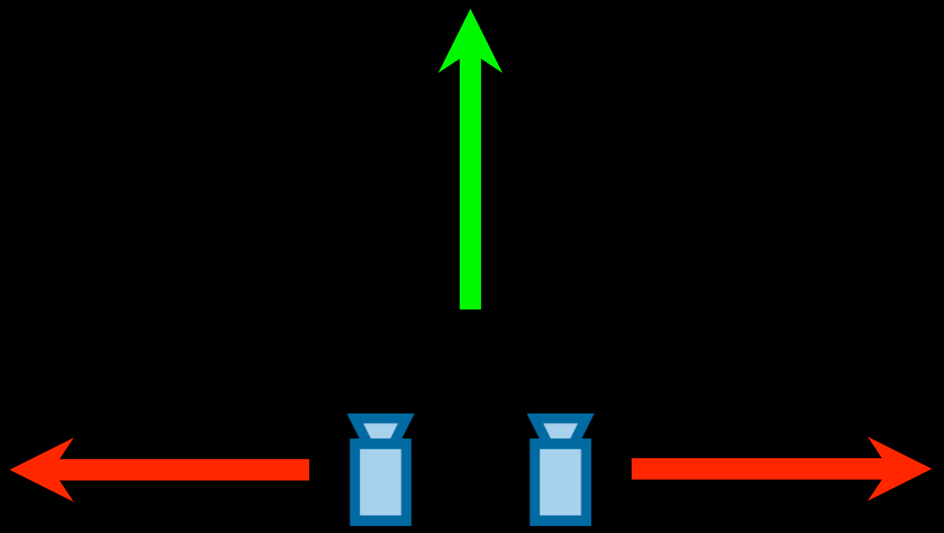
- now built into all mobile phones
- one simple camera sweep
- panorama computed on the fly

Facebook

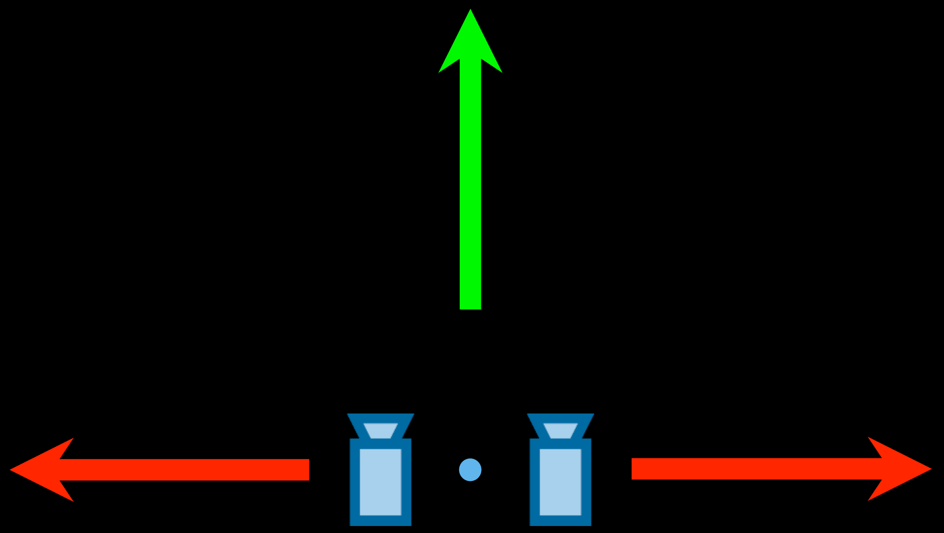
- consumer 360° cameras
- stitch views of two 180°+ fisheye cameras
- capturing photos and videos

Ben Claremont

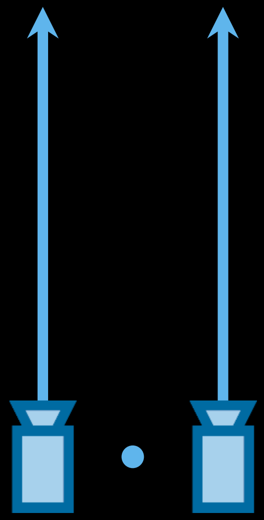
Capturing stereo panoramas



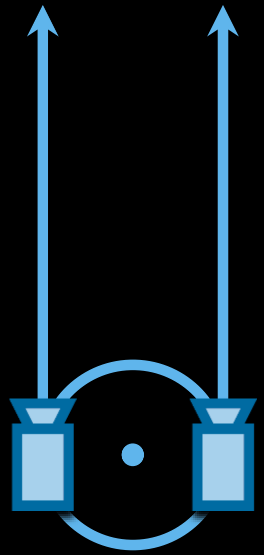
Capturing stereo panoramas



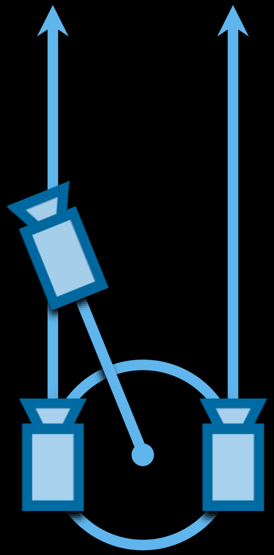
Capturing stereo panoramas



Capturing stereo panoramas

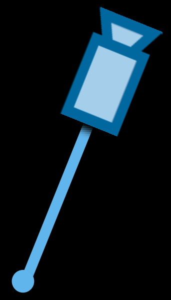


Capturing stereo panoramas



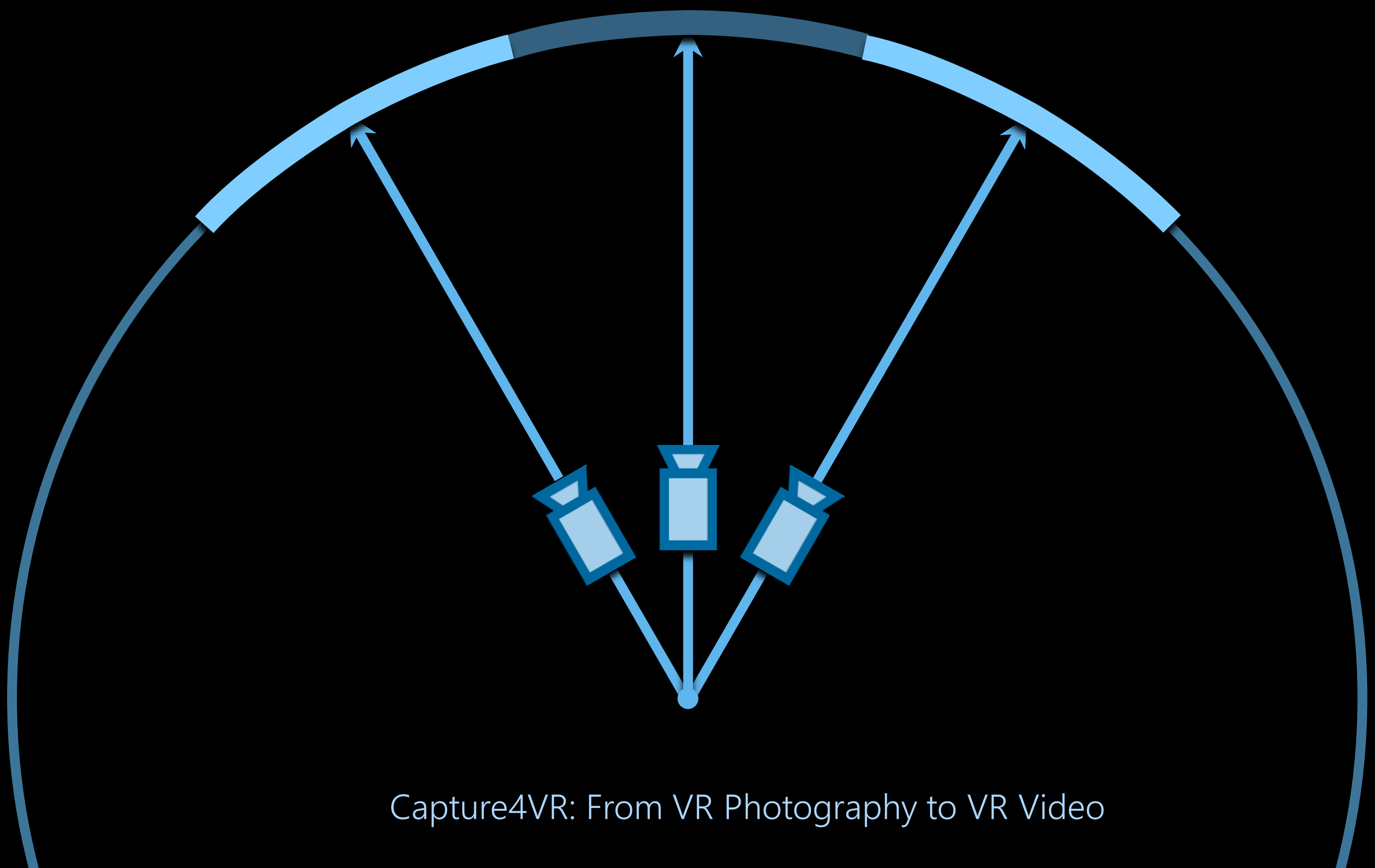
Omnistereo: Panoramic Stereo Imaging
Peleg et al., *IEEE TPAMI 2001*

Capturing stereo panoramas

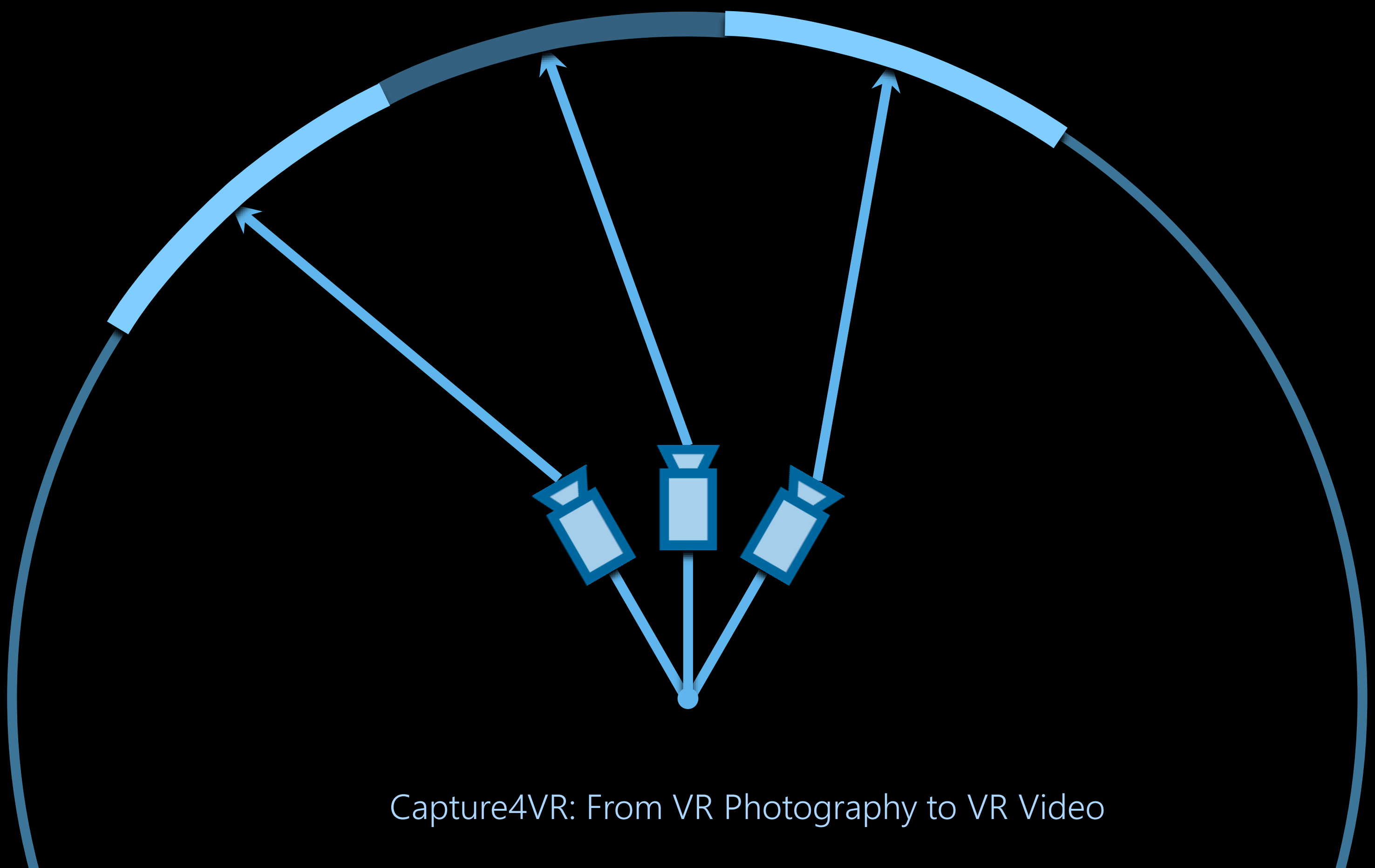


Omnistereo: Panoramic Stereo Imaging
Peleg et al., *IEEE TPAMI 2001*

Capturing stereo panoramas



Capturing stereo panoramas



Capturing stereo panoramas

Input video:

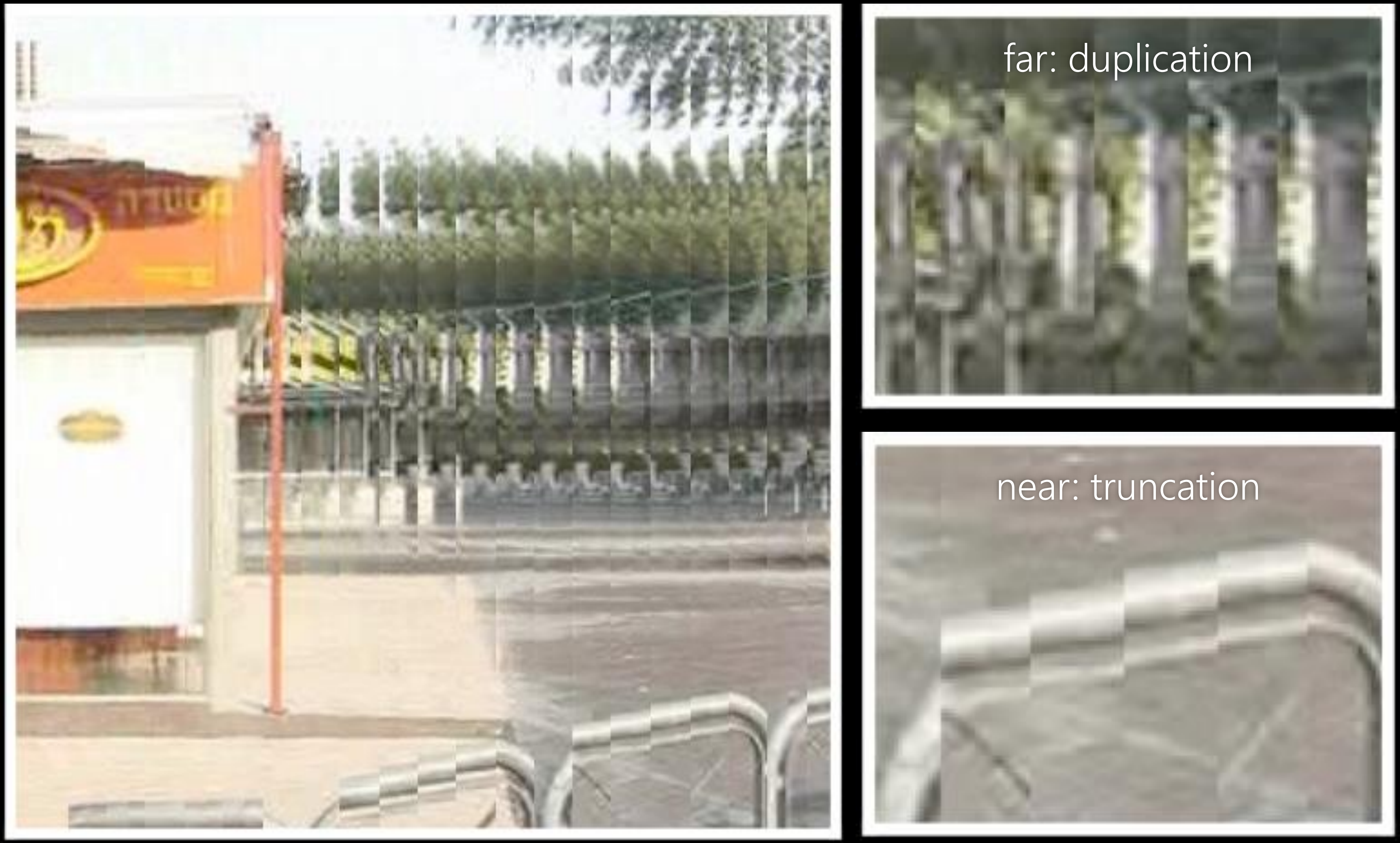
Capturing stereo panoramas

Image alignment

image-based alignment

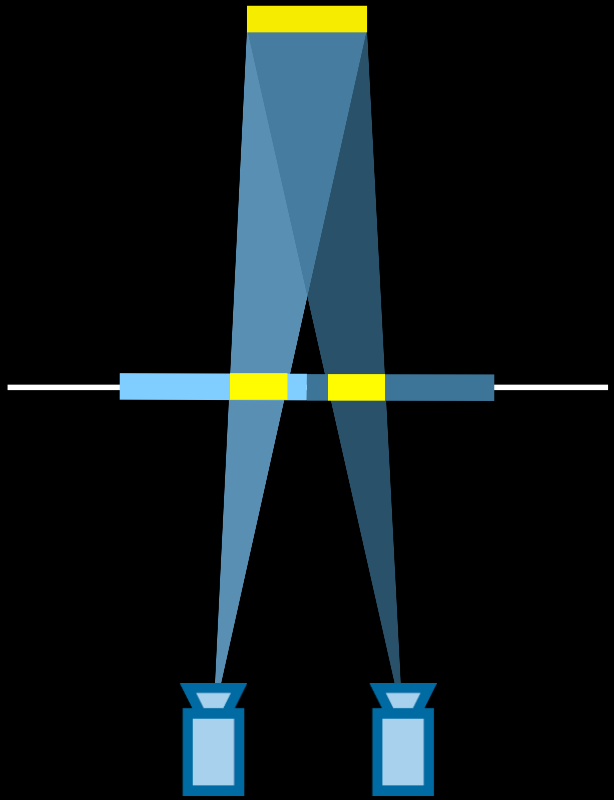
SfM-based alignment

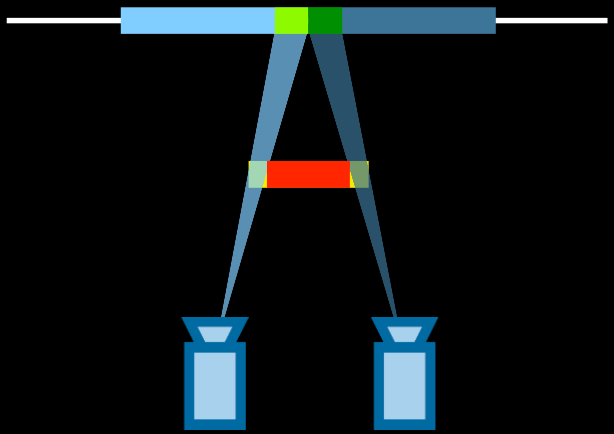
Strip blending artefacts



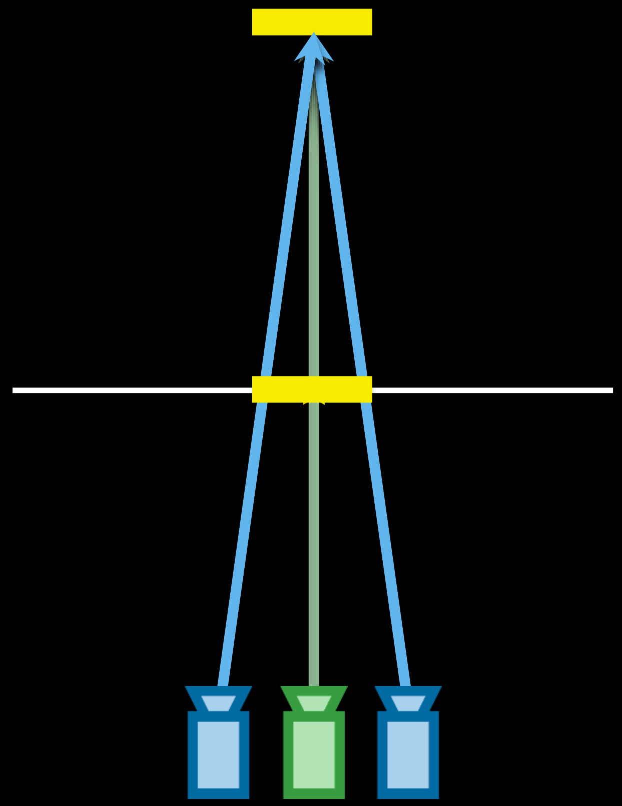
© dataset 'refaim' by Rav-Acha et al., IJCV 2008

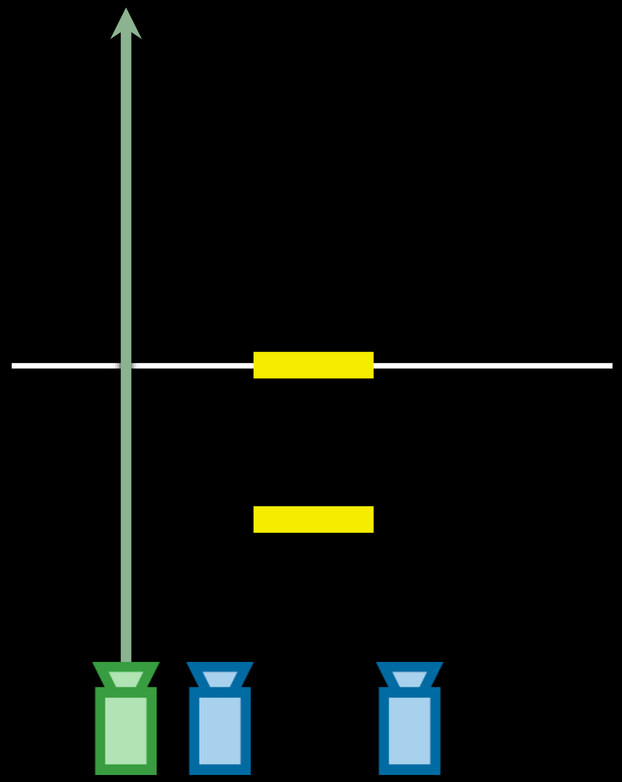
Duplication + truncation



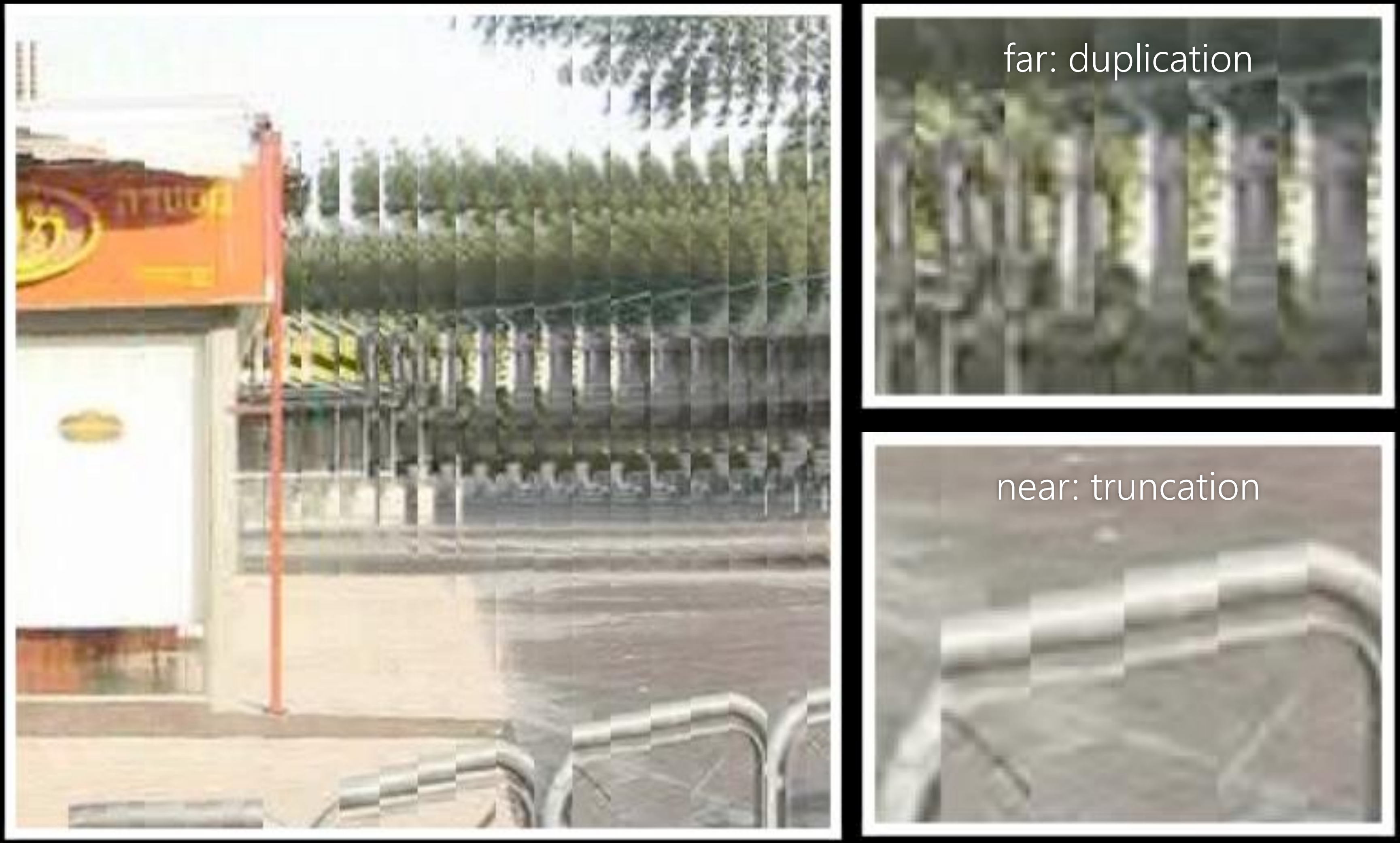


Flow-based ray interpolation





Strip blending artefacts



Flow-based blending

Blending comparison

No blending

Flow-based blending

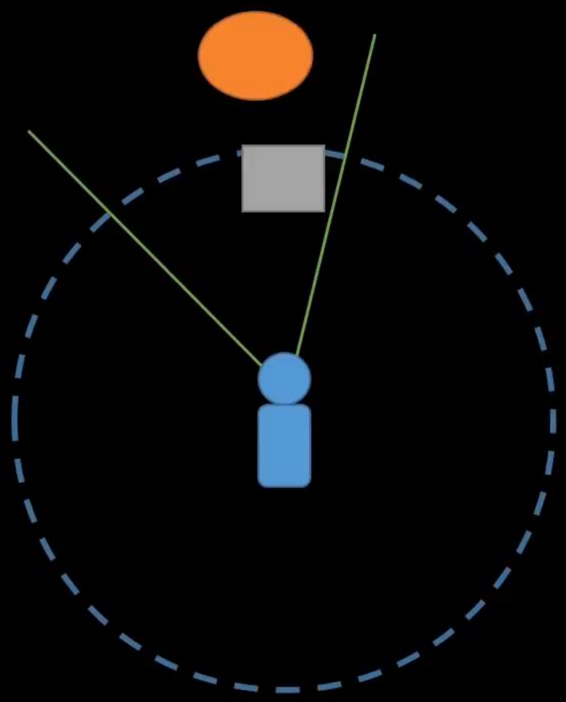
Stereo 3D panorama

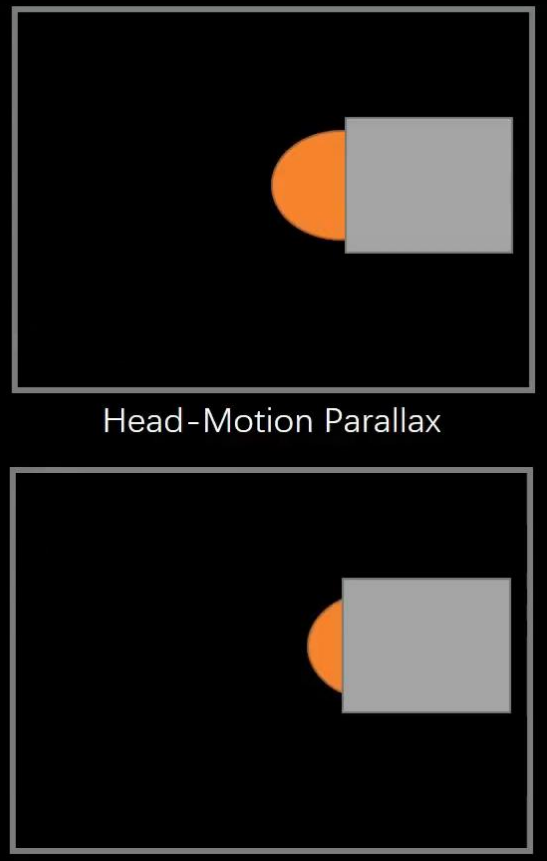
Megastereo: Constructing High-Resolution Stereo Panoramas
Richardt et al., *CVPR 2013*

Megastereo: Constructing High-Resolution Stereo Panoramas

Richardt et al., *CVPR 2013*

Motion parallax

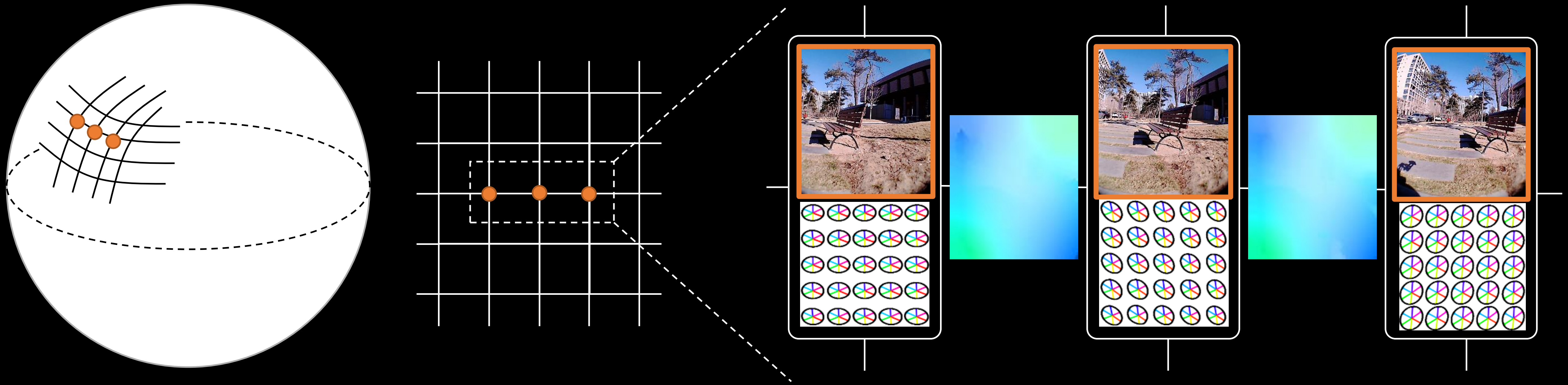




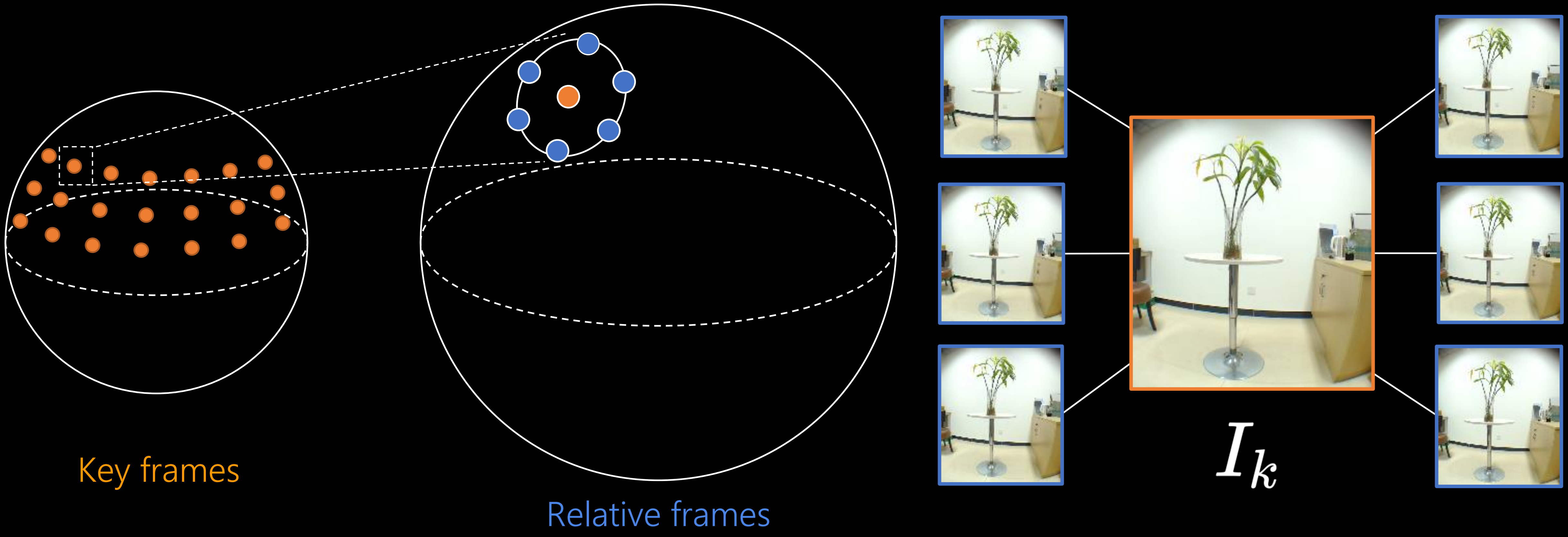
Luo et al., TVCG 2018

Parallax360: Scene representation

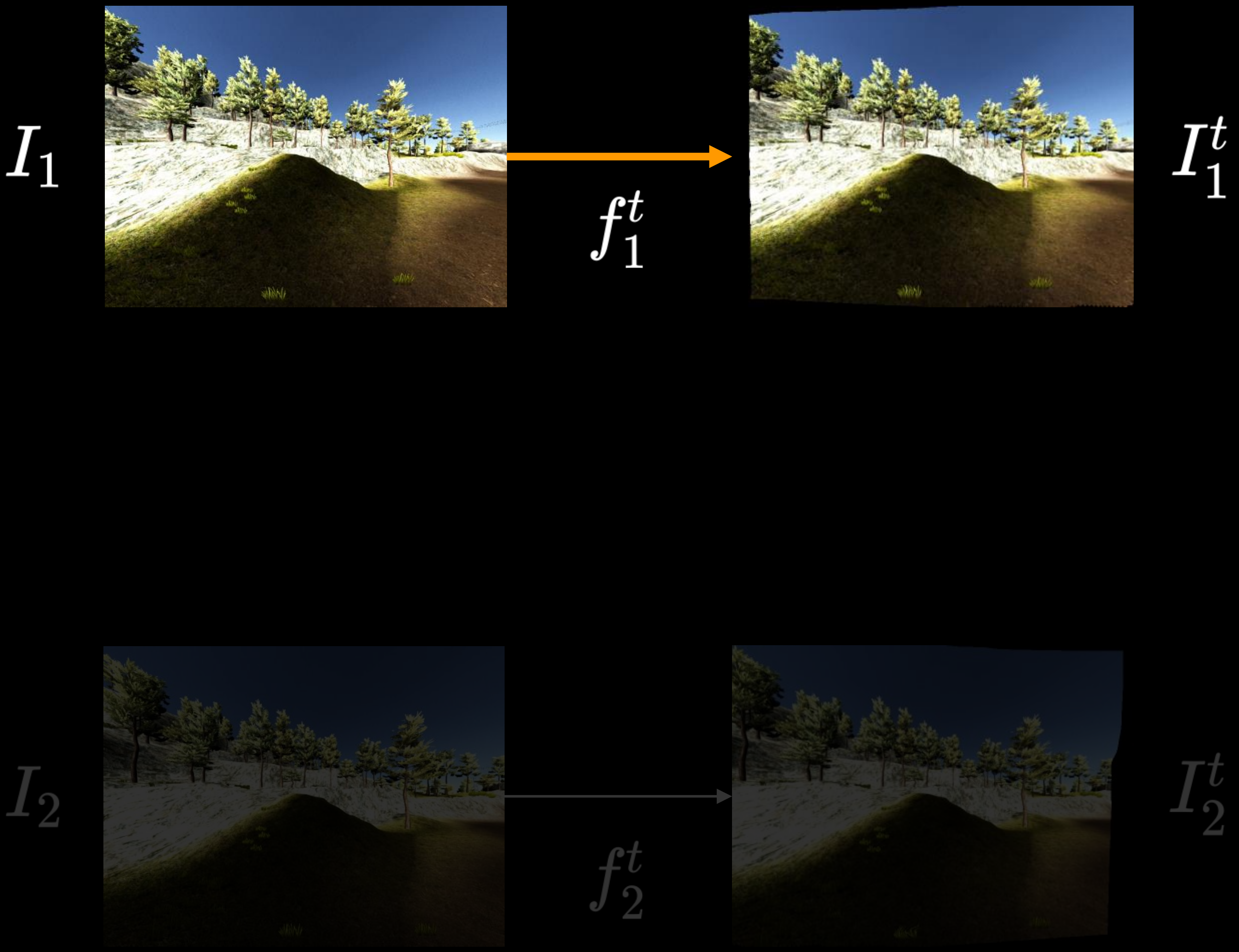
- **Key frames:** colour information of the scene
- **Disparity motion fields:** implicit 3D information at each key frame
- **Pairwise motion fields:** efficient and smooth viewpoint transitions in novel-view synthesis



Parallax360: Image capture scheme

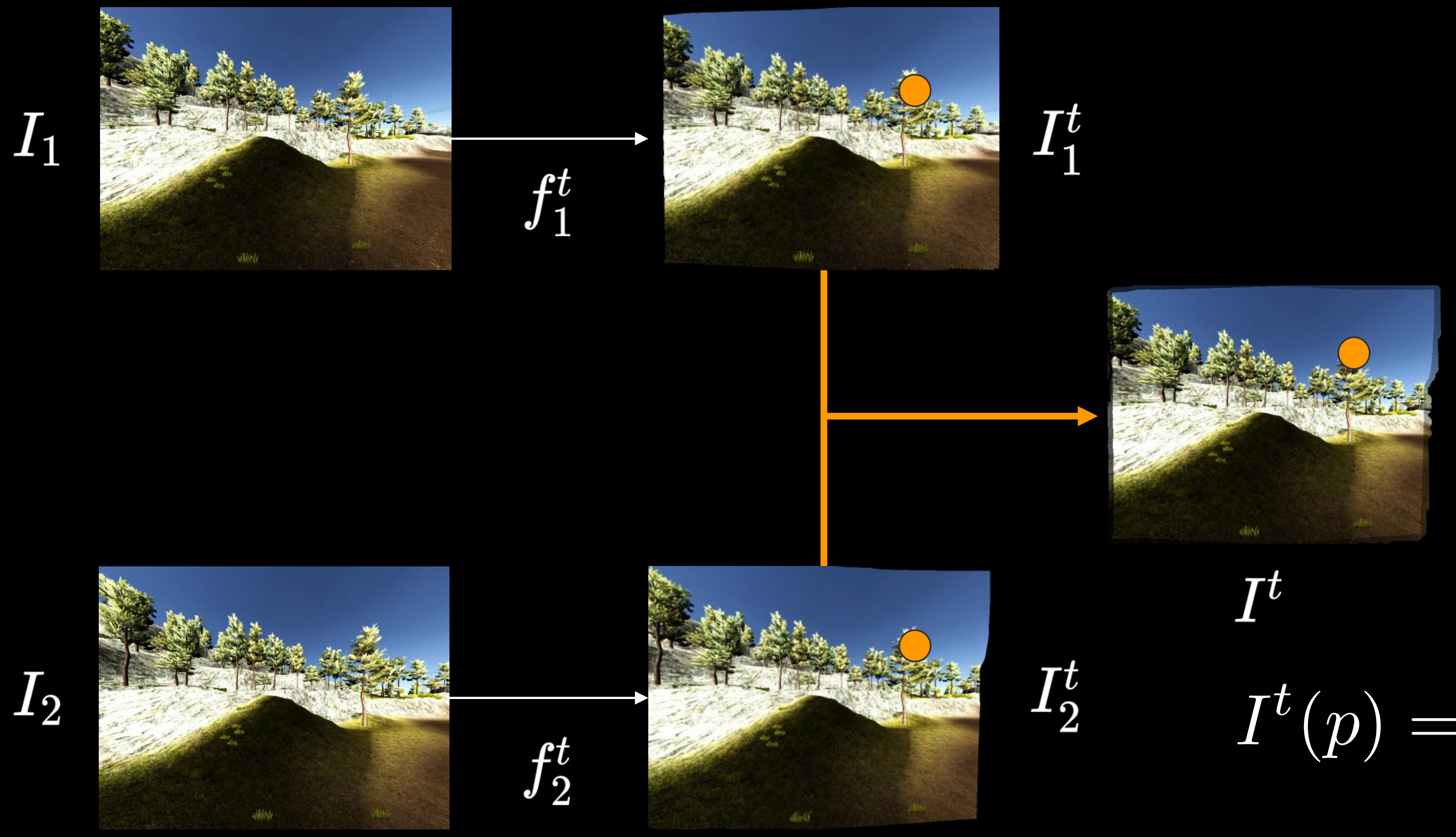


Parallax360: Novel-view synthesis



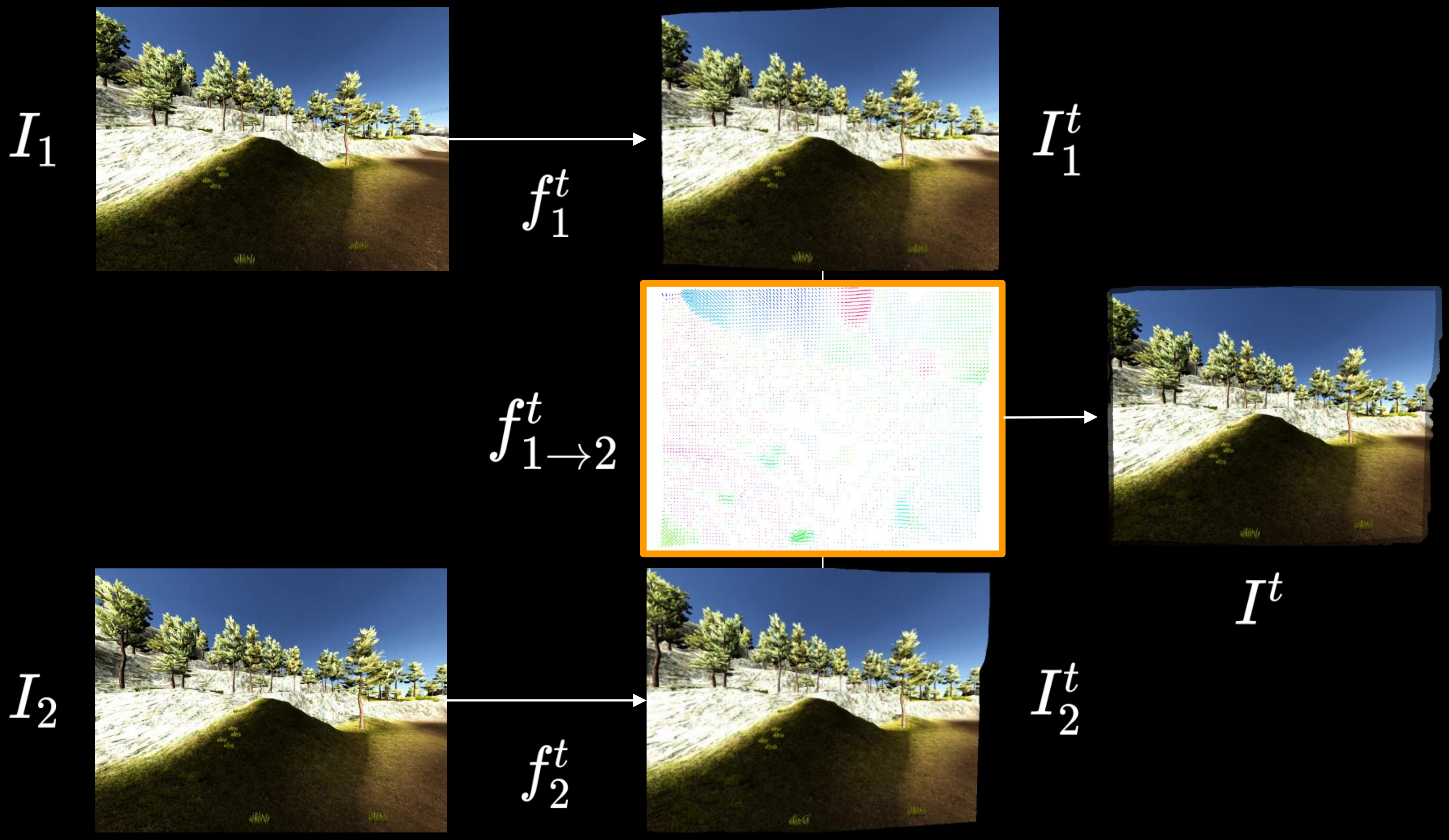
$$I_1^t(p) = I_1\big((f_1^t)^{-1}(p)\big)$$

Parallax360: Novel-view synthesis



$$I^t(p) = \sum_{k=1}^{K} \overline{w}_k I_k^t(p_k)$$

Parallax360: Novel-view synthesis



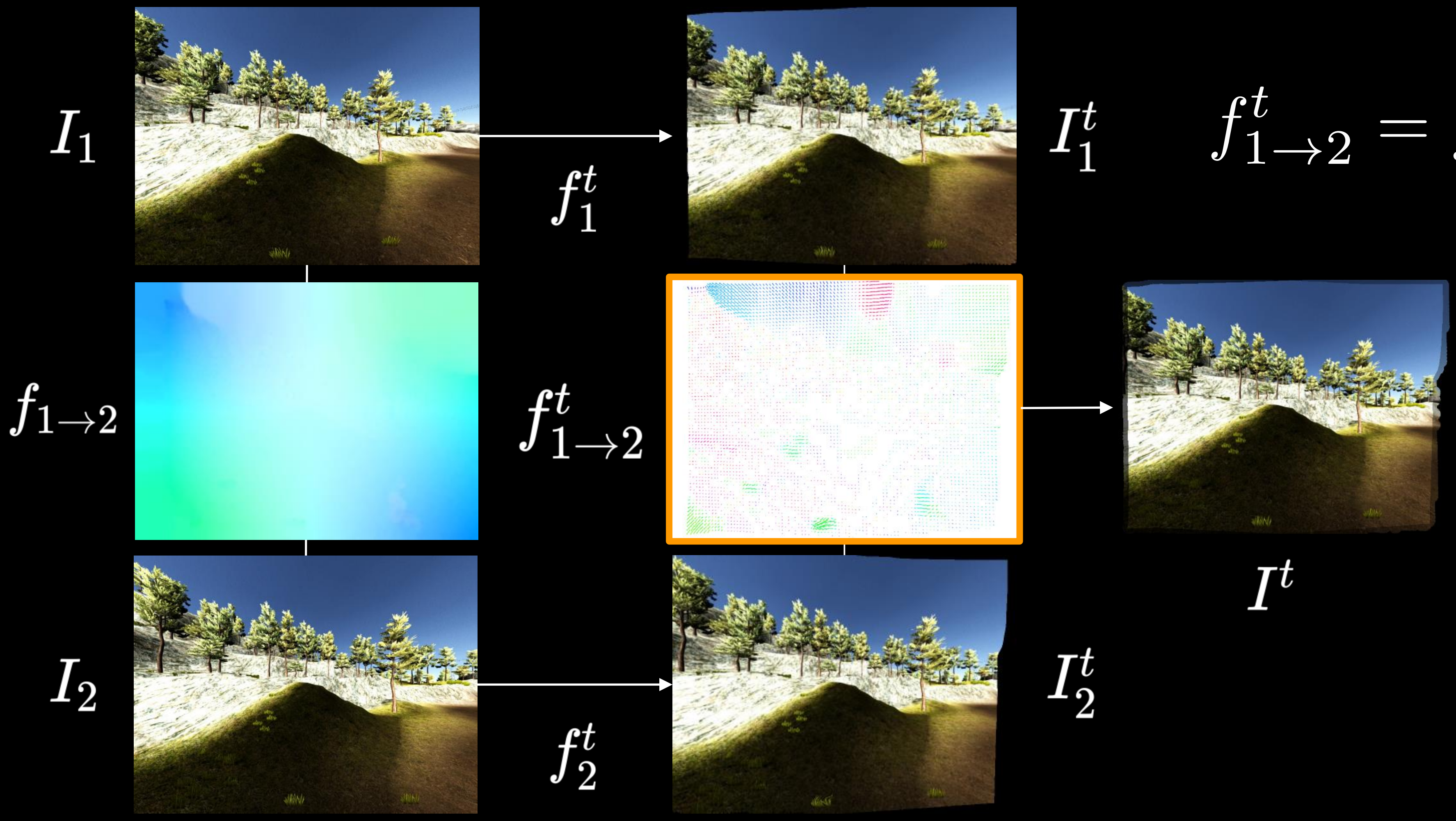
Experiments and Results

Evaluation of view synthesis quality:

Flow-Based Blending

Alpha Blending

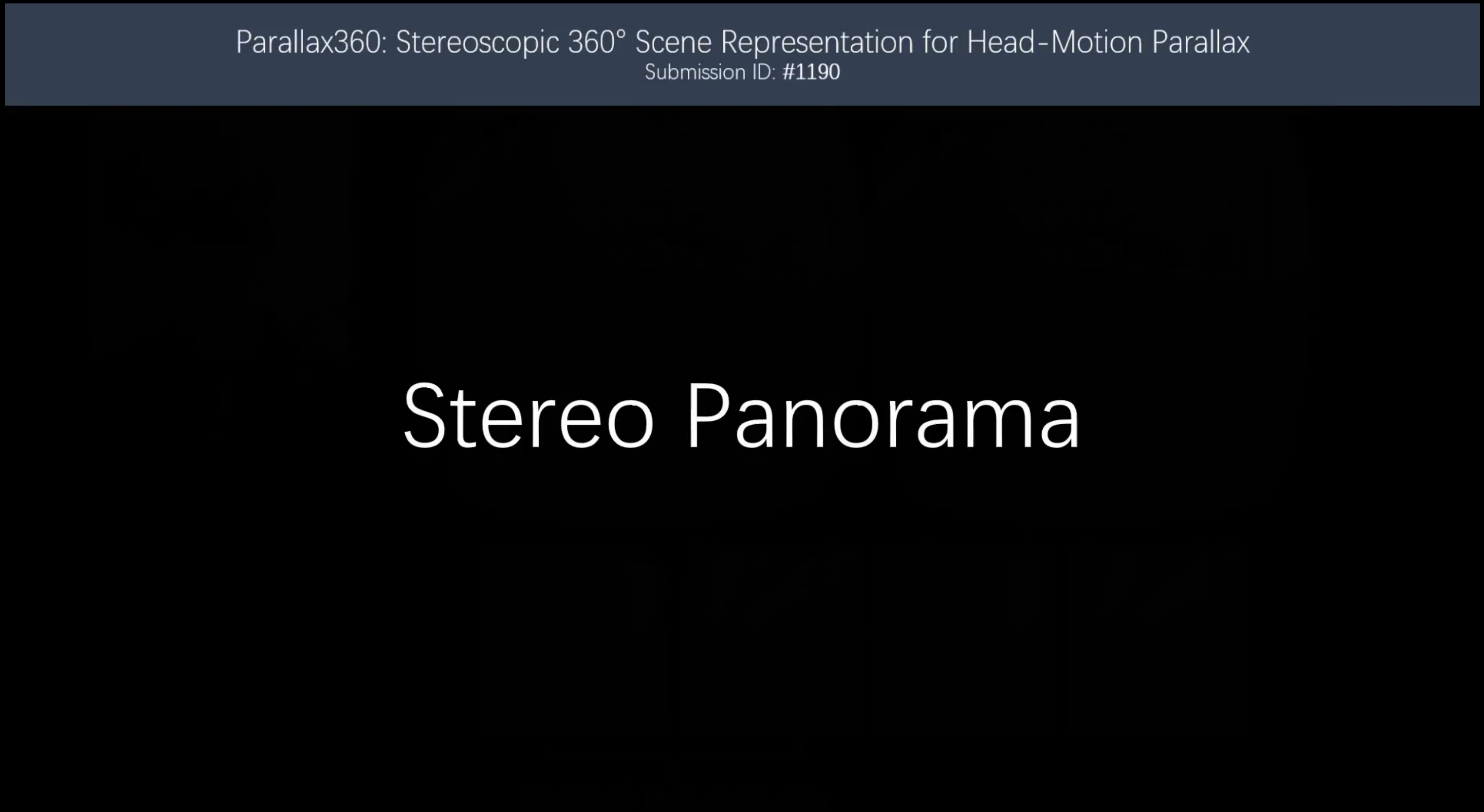
Parallax360: Novel-view synthesis



$$f_{1\to2}^t = f_2^t \circ f_{1\to2} \circ \left(f_1^t\right)^{-1}$$

Parallax360: Results

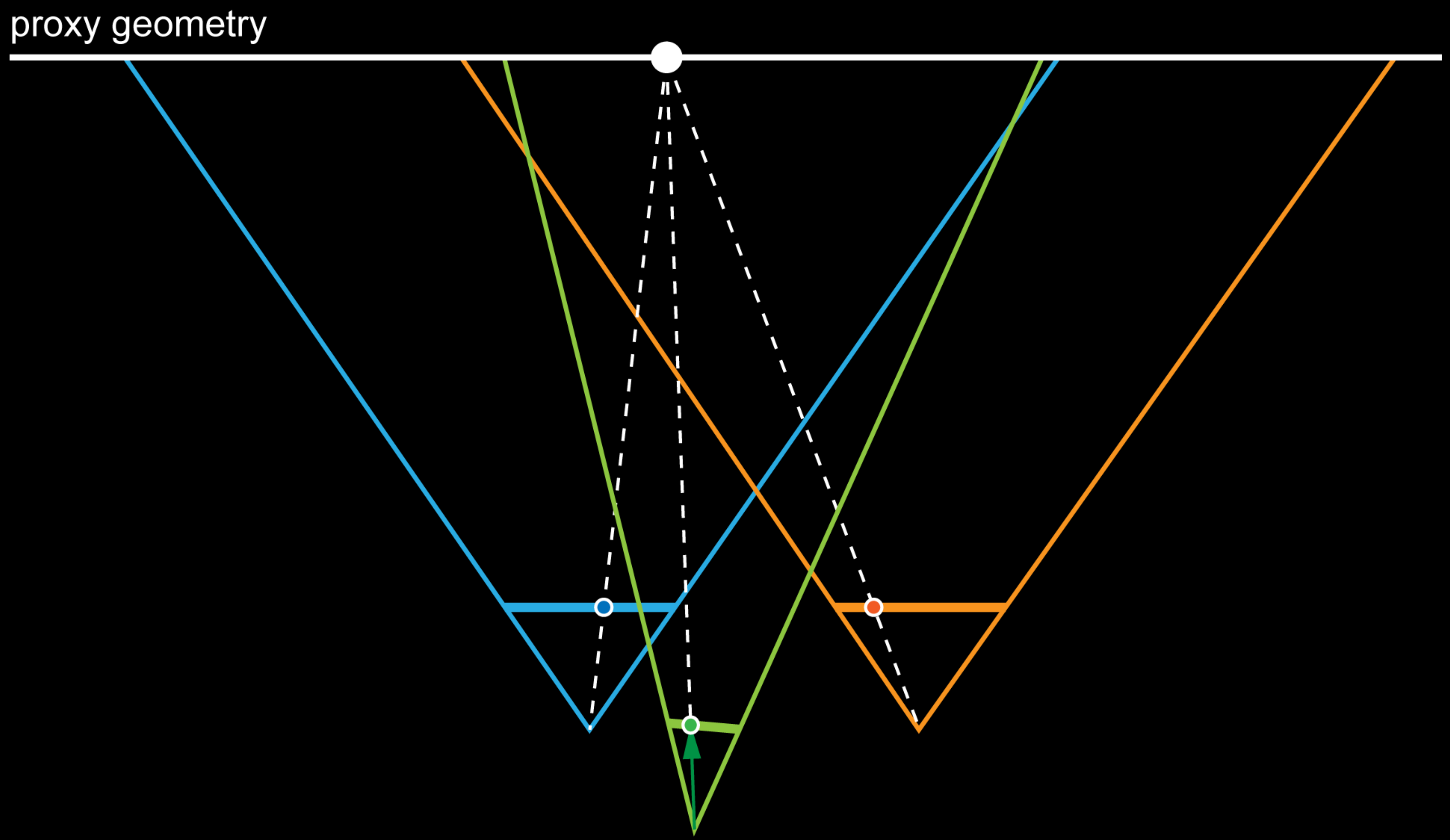
Comparison on real-world scenes:



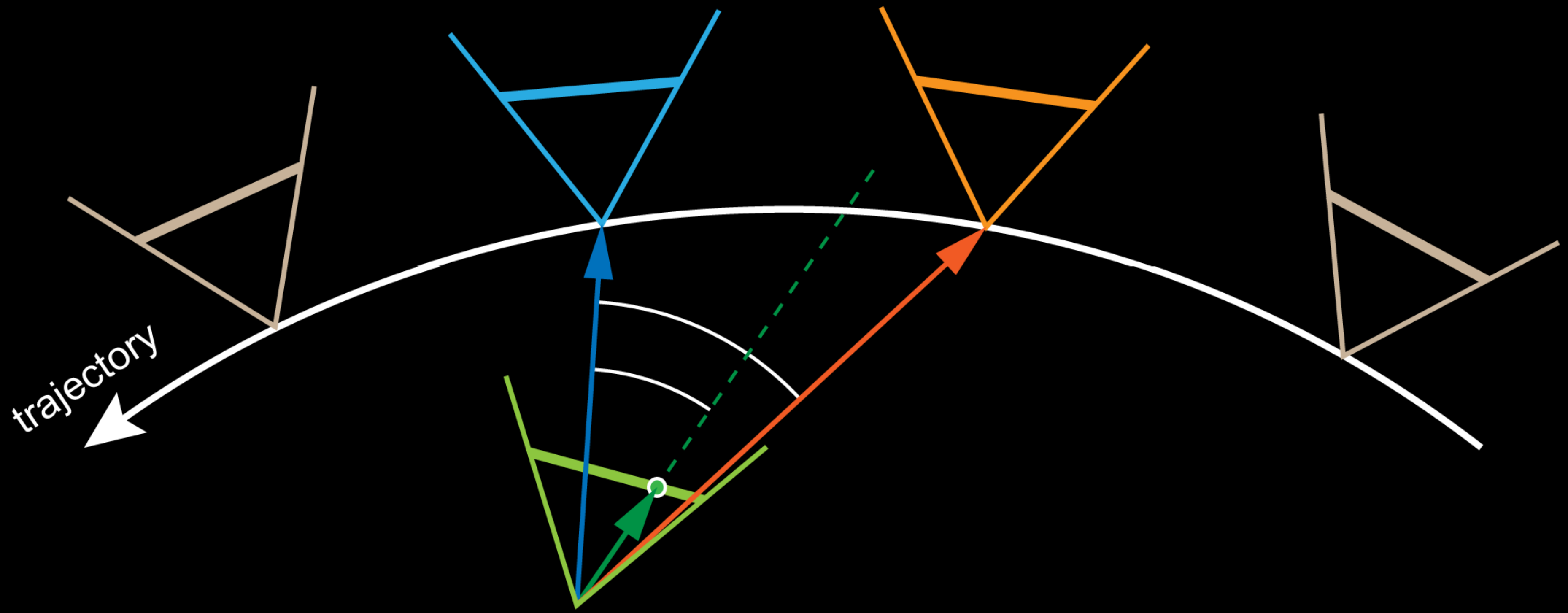
Input video

Dataset:	Rooftop
Capture:	rig
Resolution:	960×1280
Field of view:	88°×104°
Images:	360
Radius:	1.22 m

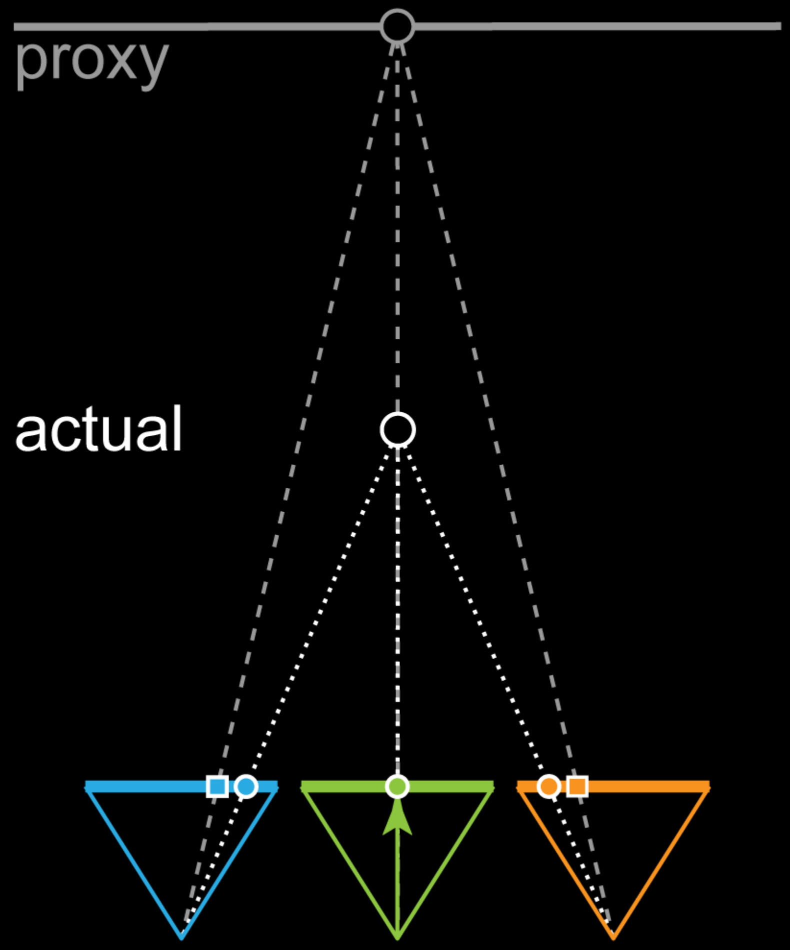
MegaParallax: Proxy-based novel-view synthesis



MegaParallax: Per-ray novel-view synthesis



MegaParallax: Flow-based blending



MegaParallax: Forward–backward motion

Parallax360 view synthesis

[Luo et al., 2018] (constant perspective)

MegaParallax

[Bertel et al., 2019] (with changing perspective)

Bertel et al., MegaParallax, TVCG 2019

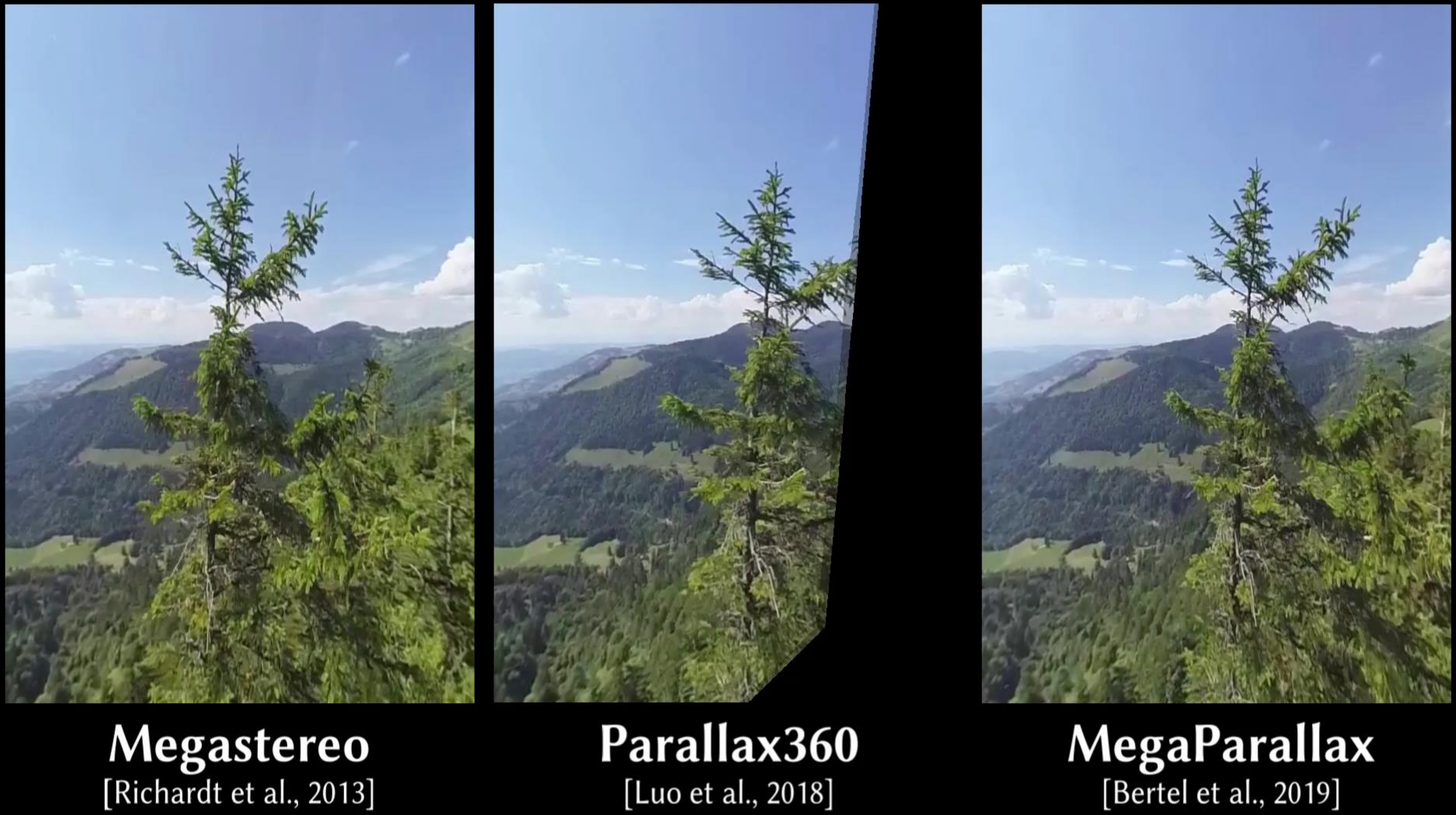
MegaParallax: Input video

Bertel et al., MegaParallax, TVCG 2019

MegaParallax: result

Bertel et al., MegaParallax, TVCG 2019

MegaParallax: Lateral translation



Bertel et al., MegaParallax, TVCG 2019

Panoramas summary

- Panoramas:
  - widespread adoption in smartphones + 360 cameras
  - but flat appearance due to lack of depth
- Stereo panoramas:
  - appearance of depth in all directions
  - extended to stereo 360 video [Anderson et al. 2016, Schroers et al. 2018]
  - but no support for head translation (or depth at poles)
- Motion parallax:
  - additional degrees of freedom allow more immersive exploration

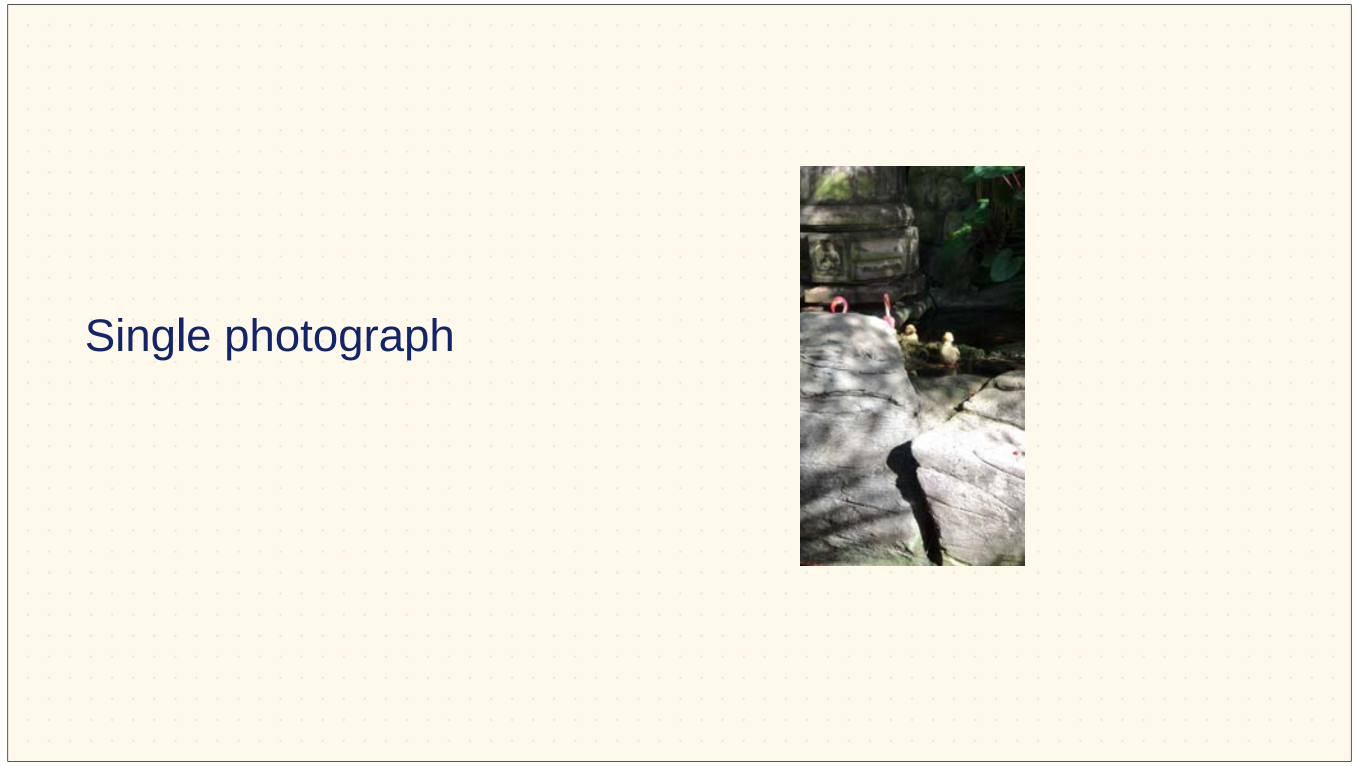
Next up

Start		Topic	Speaker
14:00	1.	Introduction	Christian Richardt, Bath
14:20	2.	360° (Stereo) Panoramas	Christian Richardt, Bath
14:40	3.	3D Photography	Peter Hedman, UCL
15:00	4.	Light Field Photography	Ryan S. Overbeck
15:20		Q&A + *Break*	
15:35	5.	360 and ODS Video	Brian Cabral, Facebook
15:55	6.	Live ODS Video	Robert Konrad, Stanford
16:15	7.	6-DoF Video	Brian Cabral, Facebook
16:35	8.	MR Capture Studios	Steve Sullivan, Microsoft
16:55	9.	Conclusion + Q&A	All presenters

I’m here to talk about 3D photography…

In other words, technologies that enable anyone to capture a place using a camera that they already own. Converting the place into a digital representation that allows anyone to later revisit it in virtual reality.

- As shown here.
- In this talk, we focus on armchair VR experiences, where you can sit comfortably and lean form side to side to peek behind objects.

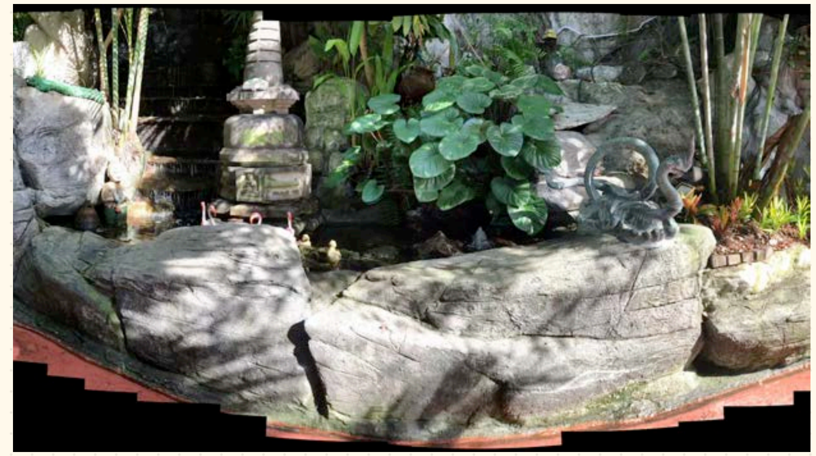


But before we delve into the details, let’s briefly talk about alternative representations for this type of content.

A normal photograph is not enough, as it has a limited field of view.

Panorama stitching

APAP	[Zaragoza2013]
Parallax-tolerant	[Zhang2014]
SEAGULL	[Lin2016]
...	



You see more of the scene by stitching many images into a seamless panorama.

However, the result lacks depth cues and looks flat in VR. We'd like to provide a more immersive experience

Omnistereo

Binocular parallax

Stereo panoramas	[Peleg1999]
Megastereo	[Richardt2013]
Google Jump	[Anderson2016]
SpinVR	[Konrad2016]

...

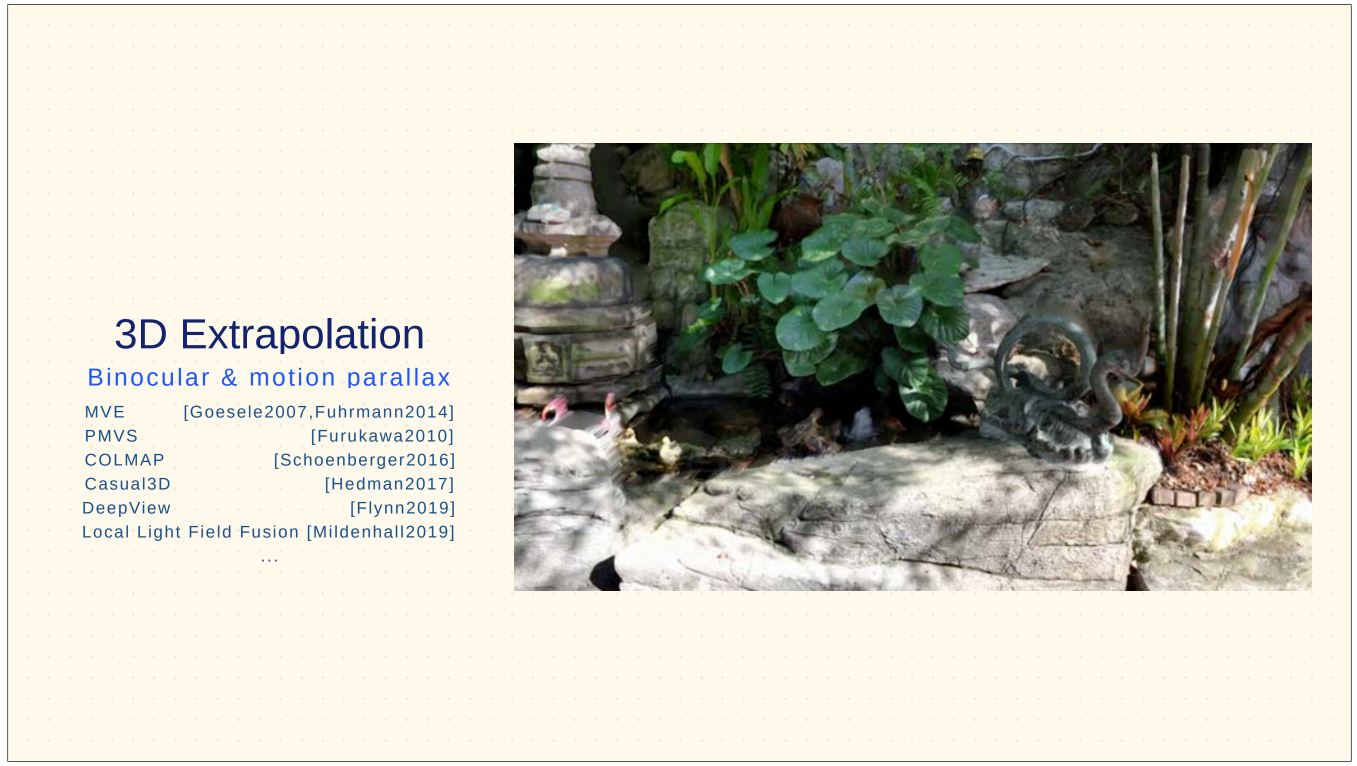
The omnistereo representation stitches two separate panoramas —one for the left eye and one for the right eye. This provides depth perception through binocular parallax, but the perspective is still fixed in the scene.

3D Interpolation

Binocular & motion parallax

Concentric mosaics [Shum1999]
Parallax360 [Luo2018]
MegaParallax [Bertel2019]
...

In the previous talk, we saw representations that allowed for 3D interpolation between camera positions. This is more immersive — and provides both binocular and motion parallax. Unfortunately, these representations impose constraints on the range of motion when viewing the scene.



In this talk, we'll be discussing representations that allow for slight viewpoint extrapolation —which enable viewers to freely move their heads around and peek behind objects.

... Specifically, we’ll be talking about the “3D photo” representation. It is similar to a panorama, ...

<depth appears>

... but with extra channels for depths, ...

<normal map appears>

... and sometimes normals.

<Color re-appears, camera starts swinging with parallax>

It also has multiple layers, so we can move the camera in 3D, experience motion parallax and peek behind objects.

<camera flies up>

The illusion breaks if you move really far away, ...

<...and back down>

…, but 3D photos look great when staying close to the original views.

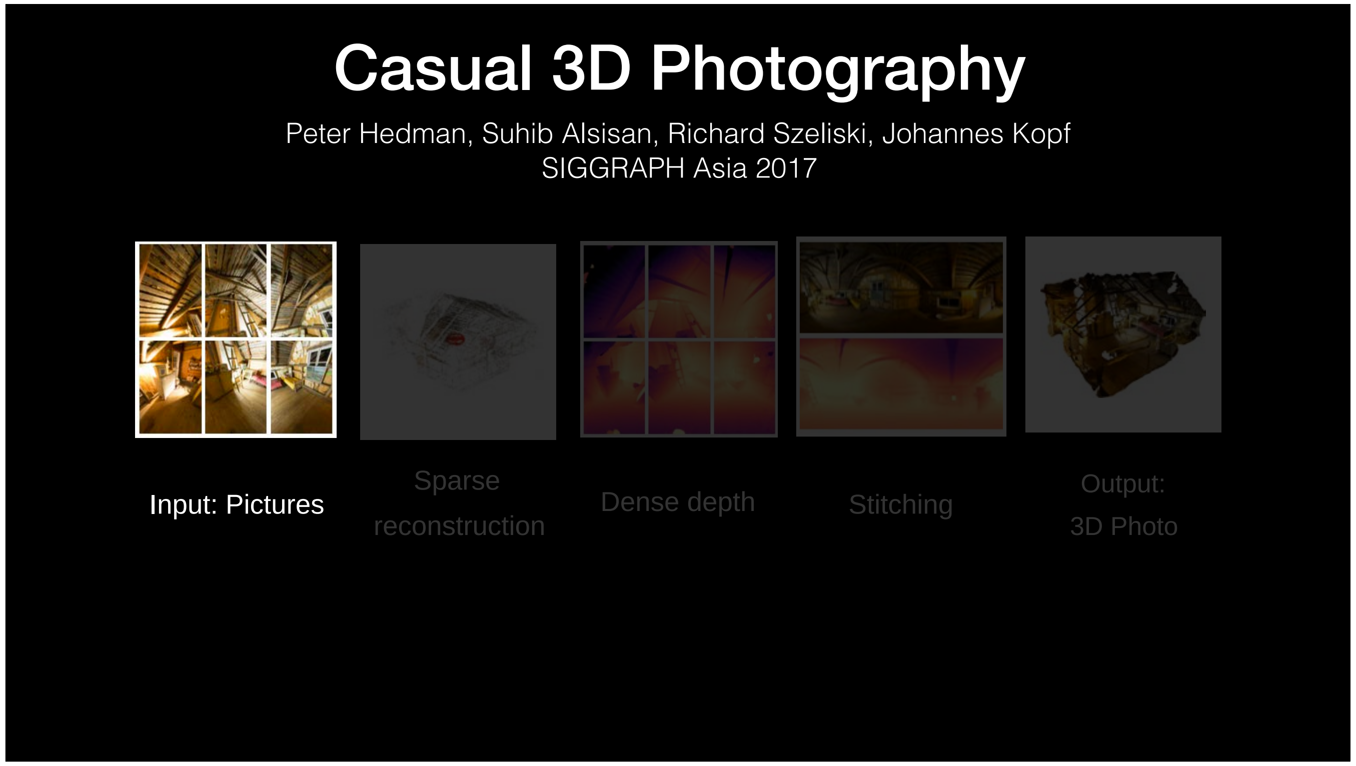
Because we have explicit 3D geometry we can interact with the 3D photo in ways that are not possible with normal photos.

<water effect appears>

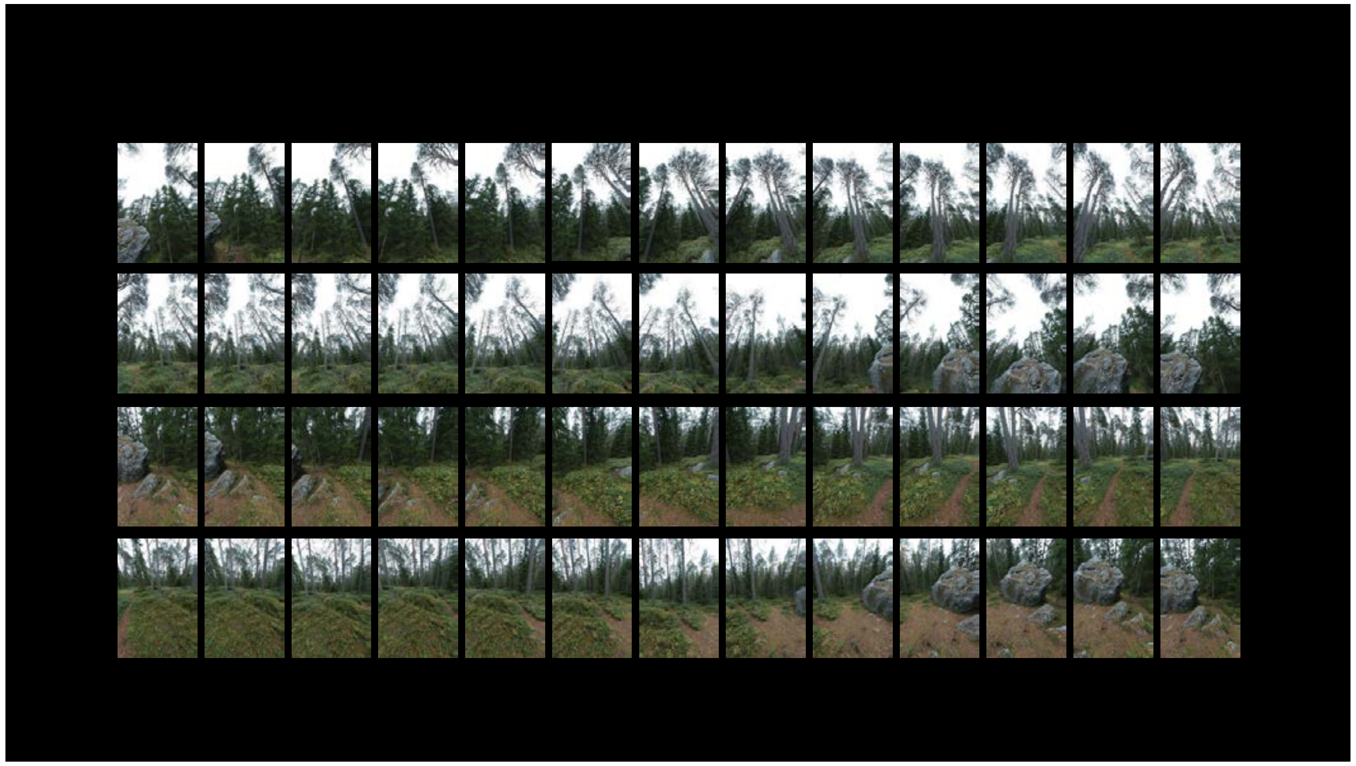
For example, we can do things like flood the scene half-way with water.

<light effect appears>

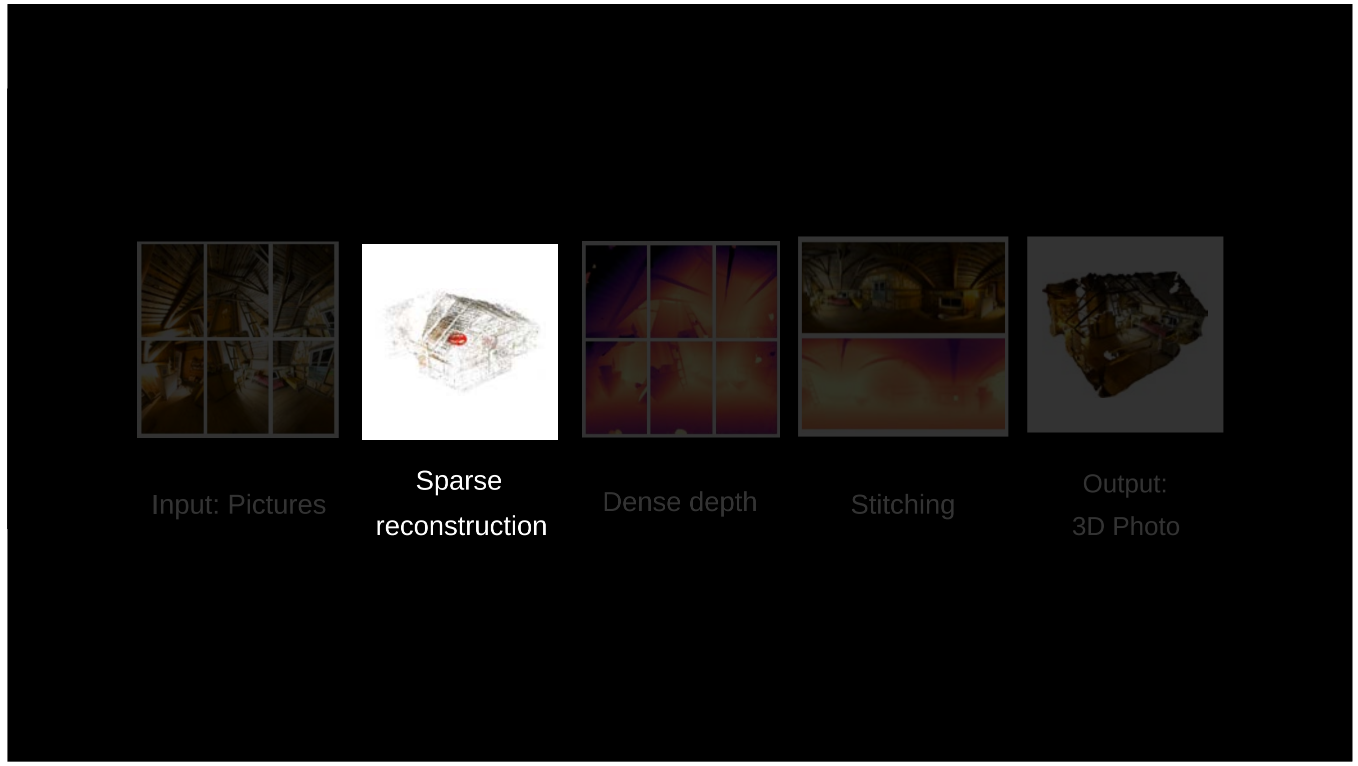
And our if we have normals, we can also play around with the lighting. We can turn day into night and shoot laser beams out of the ground!

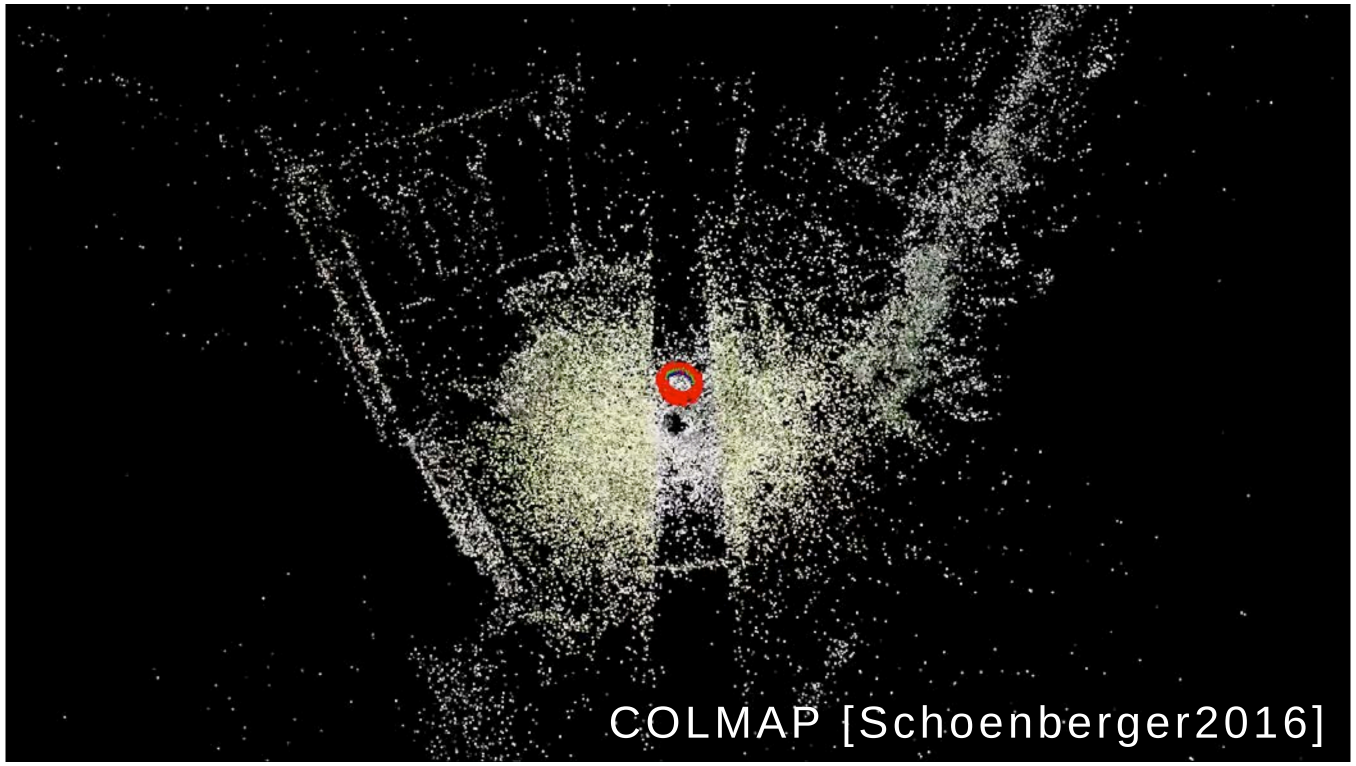


Let's first talk about a pipeline which builds 3D photos from images captured with any off-the-shelf camera.

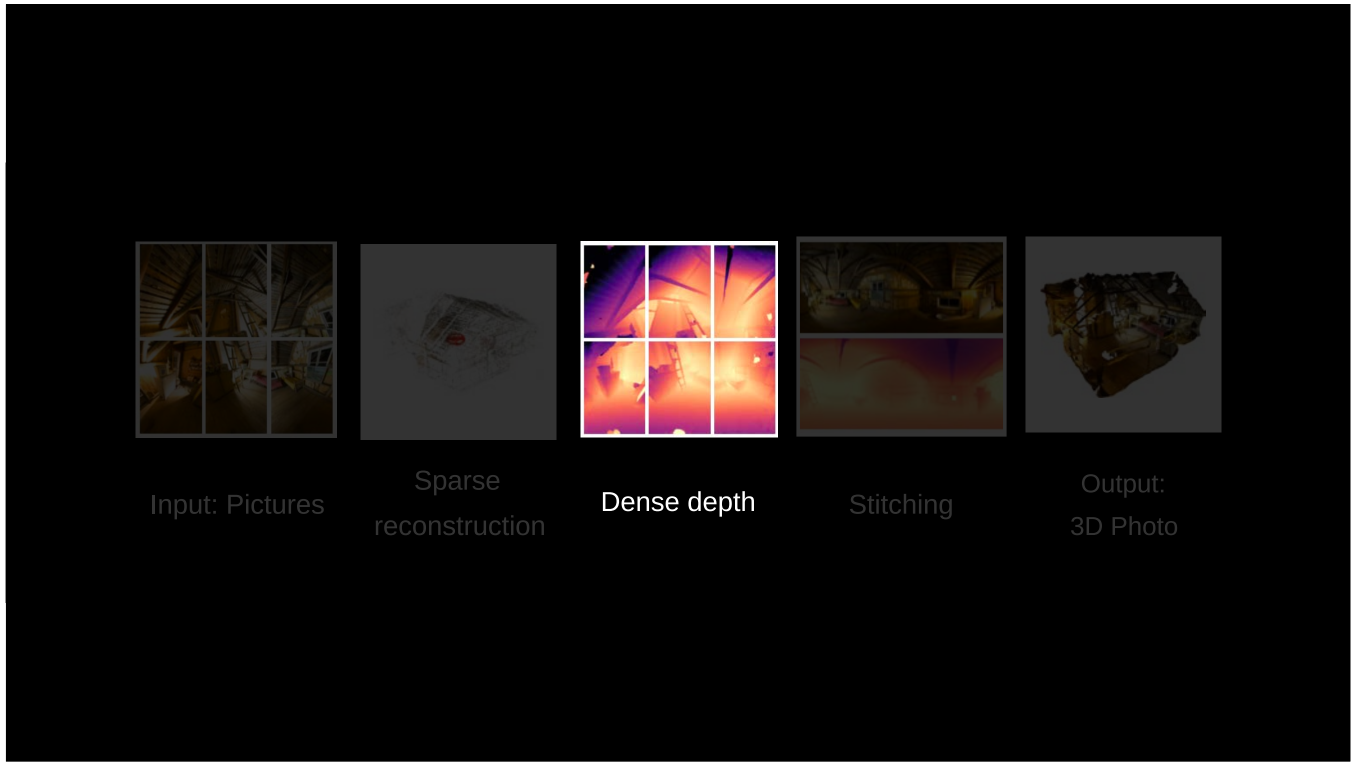


- In order to make the reconstruction work, we need some overlap in the images.
- Every point in the scene should be seen in 3-5 images so we can reconstruct its depth reliably.
- So, all in all, with a DSLR and fisheye you need about two rings of images with 20 - 25 images each to capture a complete 360 x 180 panorama that covers all directions.

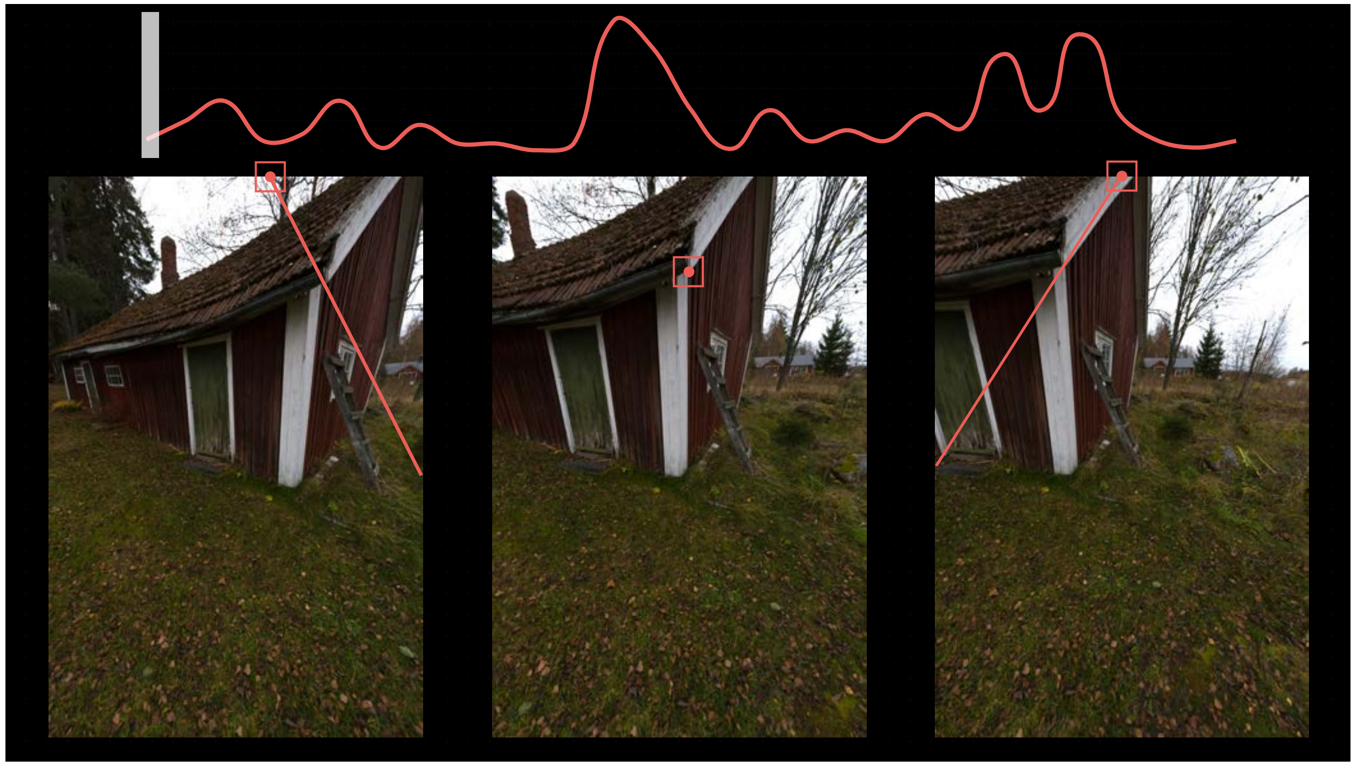

Input: Pictures
Sparse reconstruction
Dense depth
Stitching
Output: 3D Photo



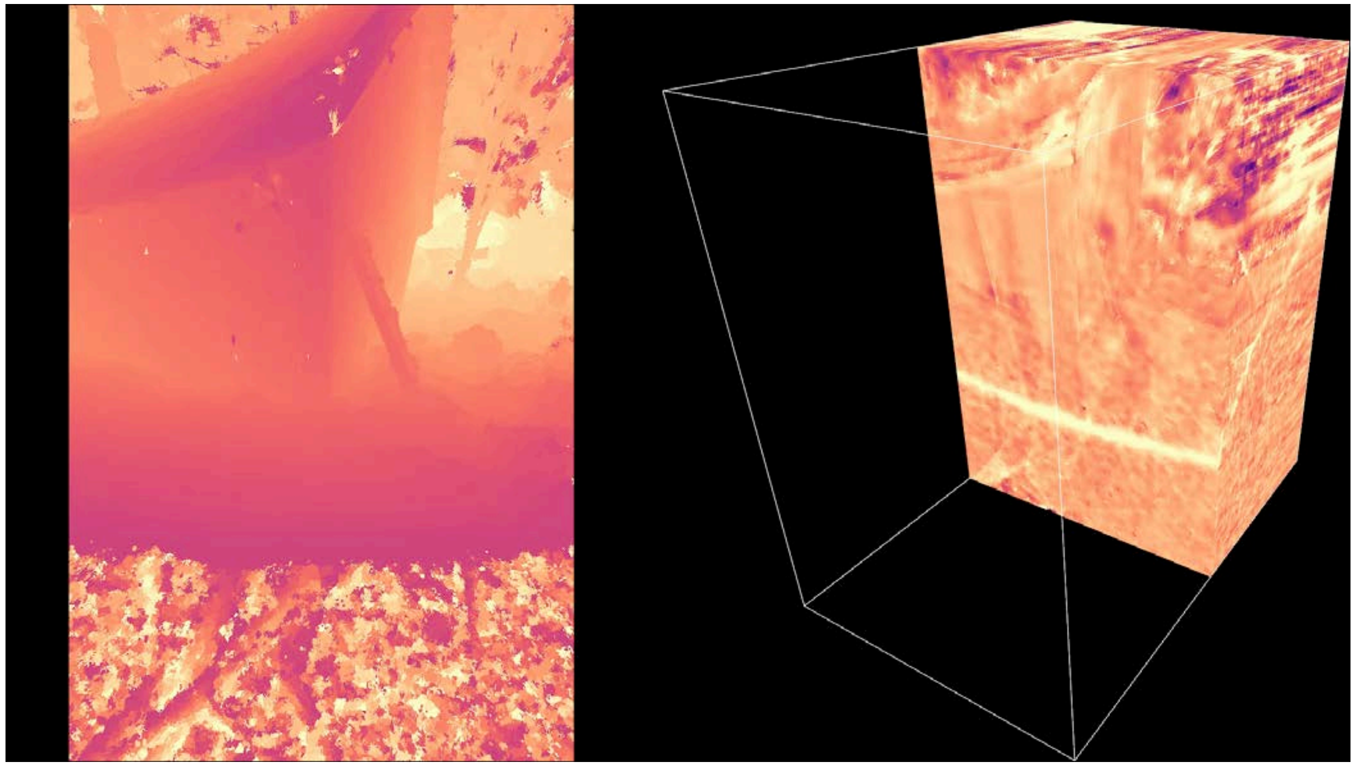
- After we've captured the images, we need to answer the questions "where were the images captured from"?
- This is a fairly standard problem in computer vision, and we use a state-of-the-art structure-from-motion algorithm to do this.
- The algorithm triangulates the camera locations: things, just from the image data:
  - This is visualized here with red crosses. You can nicely see how the algorithm recovered the structure of the rings on which I captured the images.
  - A side-effect of this triangulation is a sparse point cloud of the scene.


Input: Pictures
Sparse
reconstruction
Dense depth
Stitching
Output:
3D Photo

- The next step estimate dense depth maps for every single image in the scene. In other words, we want to estimate the 3D location of every single pixel.
- Here we visualize this information with depth maps, where bright pixels are far away, and dark pixels are closer to the camera.
- This is also a well-known problem in computer vision called “multi-view stereo”.



- Say we want to reconstruct the depth of the corner of the roof seen in the central image.
- The sparse reconstruction tells us how the cameras are position with respect to each other. Based on this, we know that the roof corner will land somewhere along these lines in the neighboring images.
- So we can search along these lines, and look for patches which look similar to the corner in the central image.
- Based on how similar these patches are, we can compute a matching confidence.
- We start our search far away, and move closer to the camera. And hopefully, the confidence reaches a maximum at the correct location, where both patches lie on the corner.
- However, as we try to match patches closer to the camera, it is often likely that we get false positive regions of high confidence.



- Let's visualize this confidence score for every single pixel in the image. I will show you an example of "winner-takes-all" optimization, where we keep track of the best depth we've seen so far.
- We start off by assuming that the depth is infinitely far away, as you can see with the very bright depth map on the left.
- On the right you can see a confidence image, which tells us how confident we are that this far-away depth is correct for each pixel.
- As we move closer to the camera, you can see how the depth map gradually forms.
- If we pause the process half-way, you can see how the depth map is not yet complete, as the ground in the foreground is missing.
- But look at the confidence volume on the right! You can see a bright white line, which tells us that we're very confident that these pixels should have this exact depth.
- As we continue the sweep, we fill in the ground and end up with a complete depth map.

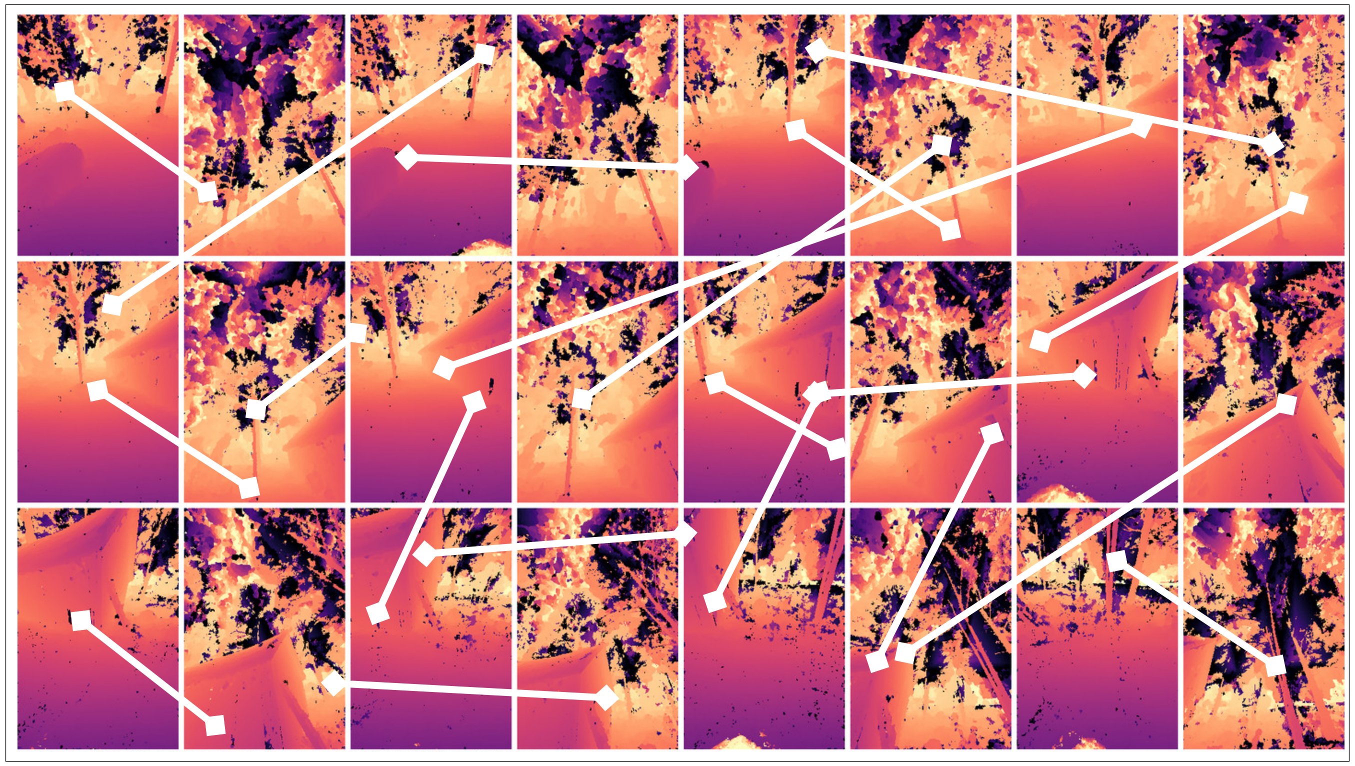
However, this depth map contains artifacts. For example, the dark regions in the top-right corner.

If we take a look at the confidence volume, you can see how these regions correspond to noisy false-positive regions.

These are caused by our capture conditions, since we captured our images one at a time, and the scene is not completely static (for example, trees can be swaying in the wind) it's possible that certain regions will never match completely between images.

Furthermore, as you get closer to the central camera, most of the other images will not be able to see this location. Since we have to compute the confidence score using fewer images, this means that the confidence volume is inherently less reliable nearby.

If we visualize this depth map in 3D, you can see how the dark regions create the so-called "flying pixels" effect, which is very distracting.



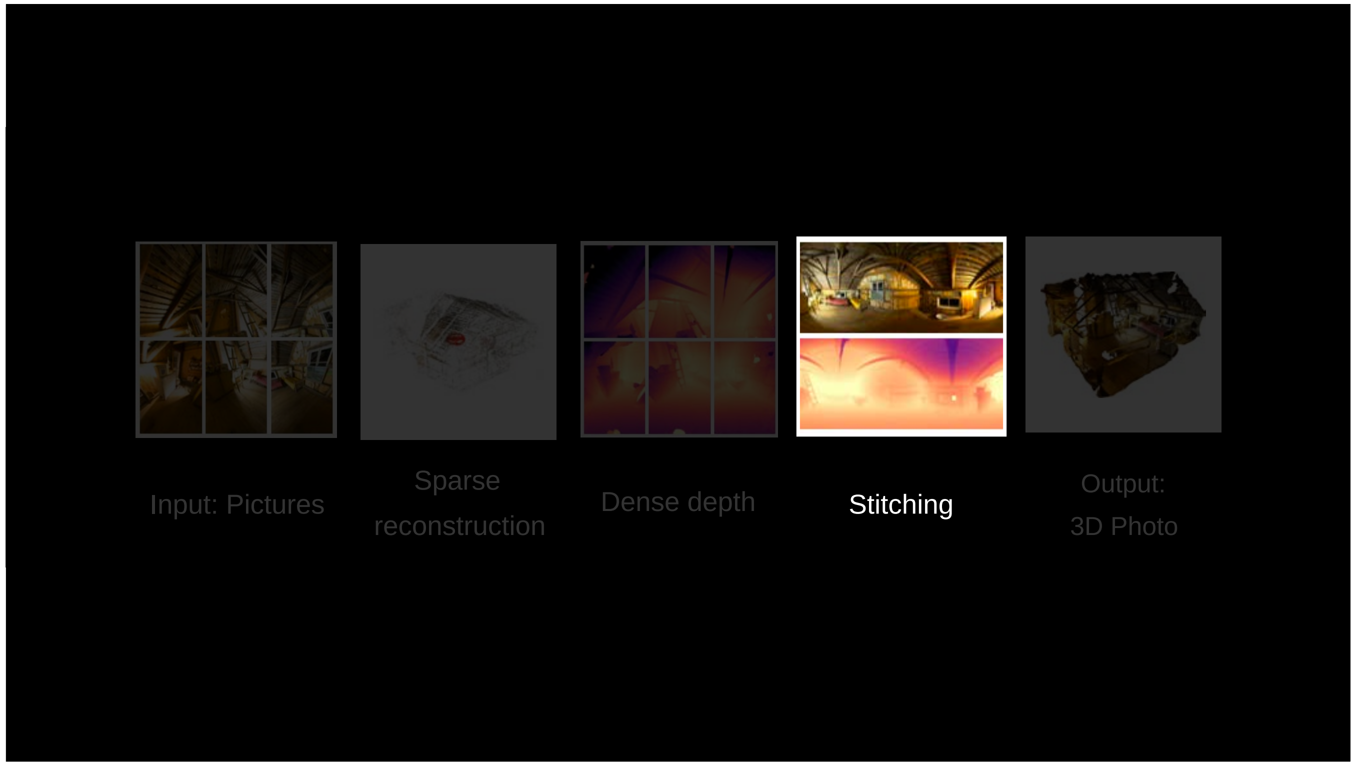
There are many different ways of alleviating these artifacts.

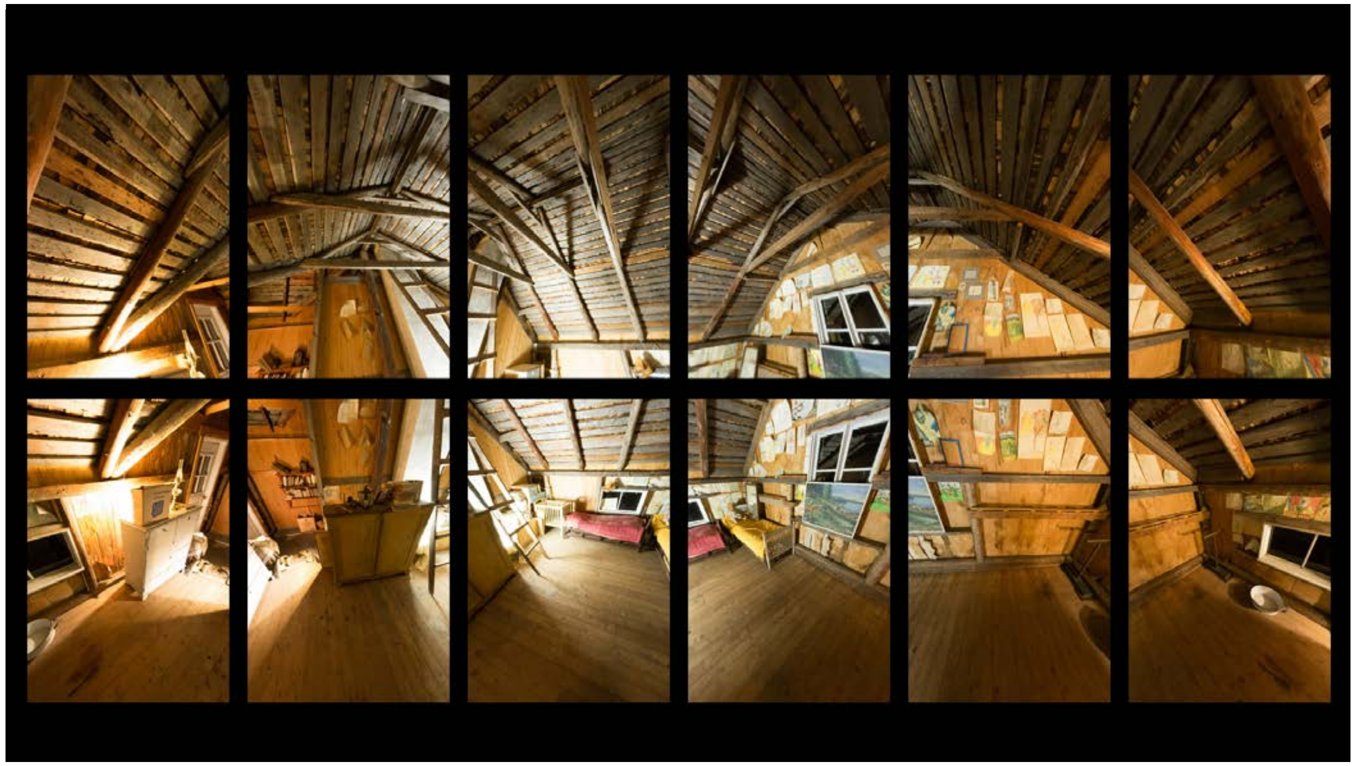
One way I found particularly effective was to first use this cheap winner-takes-all strategy to quickly compute noisy depth maps for each input image.

Then we look for consistencies and inconsistencies between all depth maps. Eventually forming a notion of the free space in the scene.

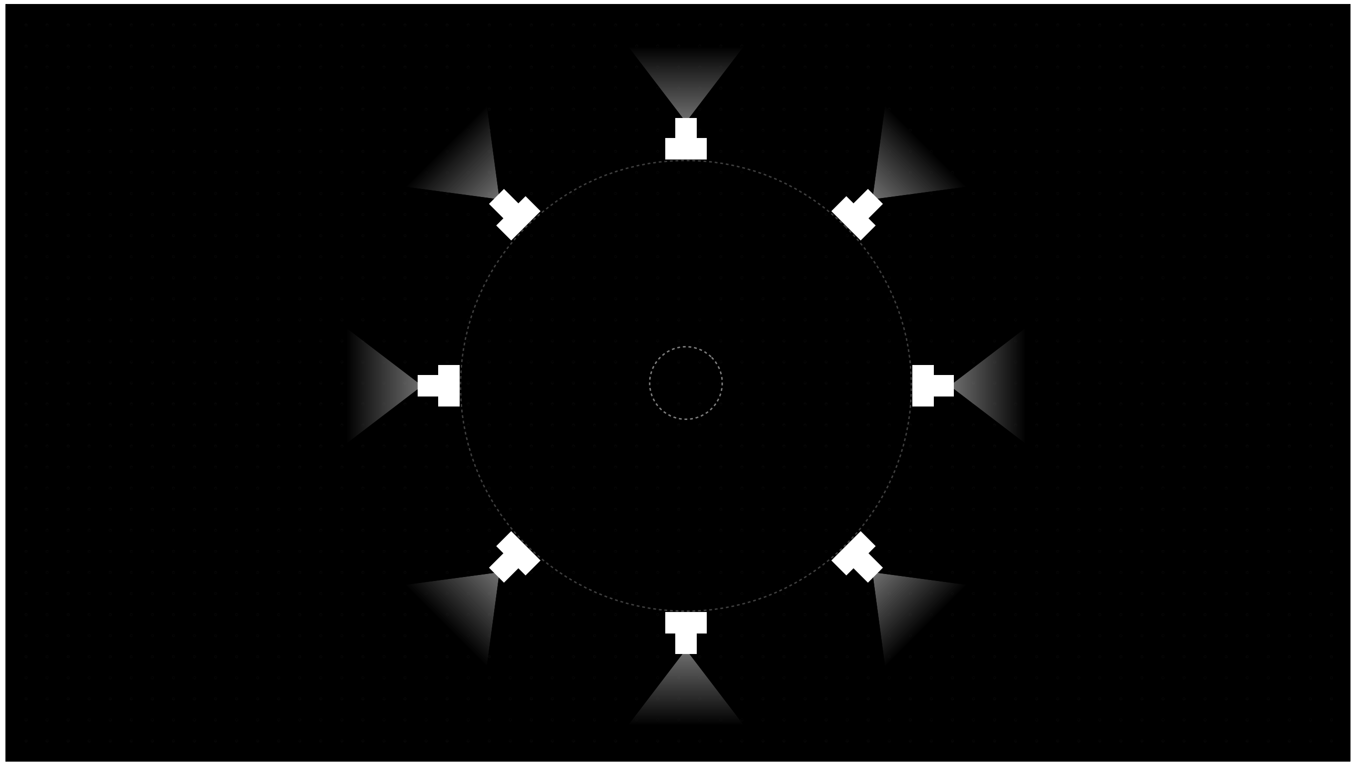
By simultaneously respecting free space constraints in the scene, and also regularizing the depth map to be smooth, it’s possible to extract a better looking depth map without artifacts.

Let's take a look at it in 3D, as you can see it looks much more believable.

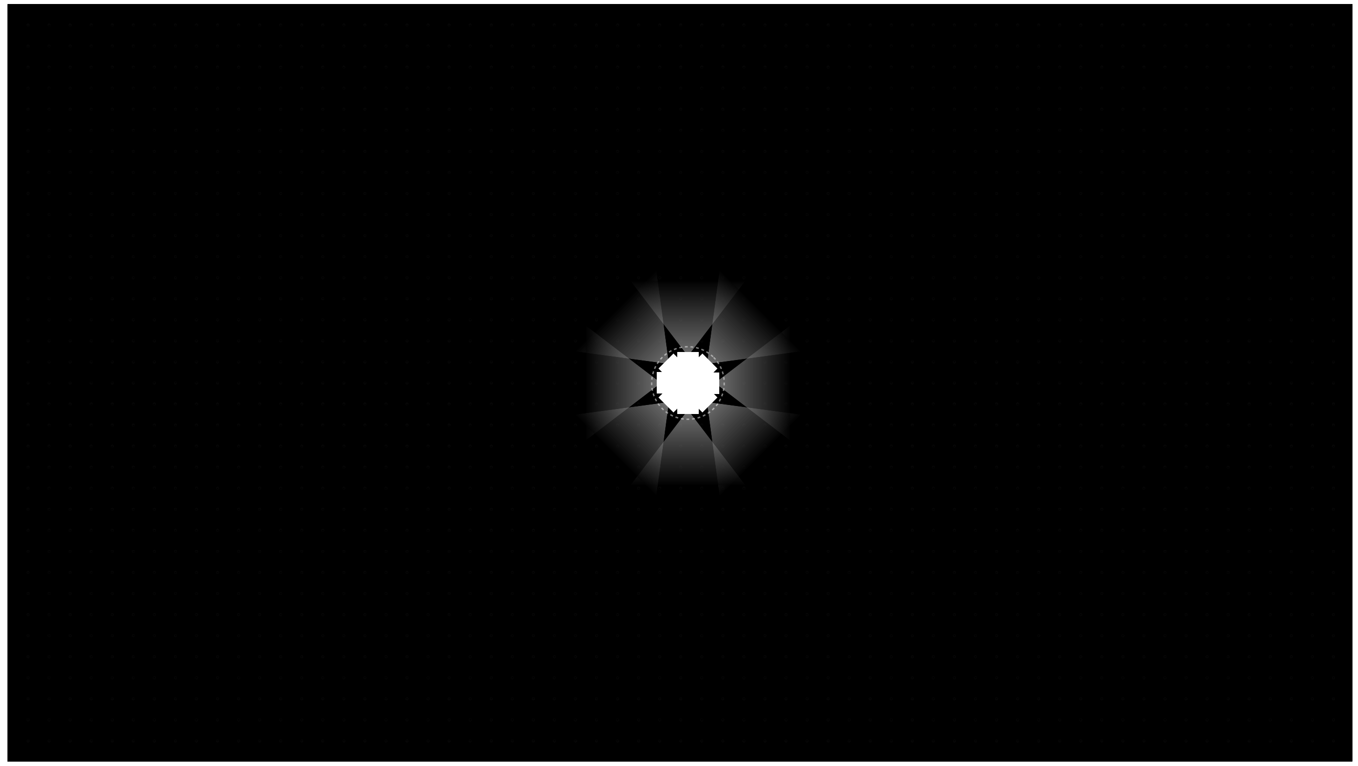

Input: Pictures
Sparse reconstruction
Dense depth
Stitching
Output: 3D Photo



- Now we can compute depth maps that are relatively artifact free, and we're done with the analysis of our images. The depth maps tell us almost everything there is to say about the scene, at least in terms of geometry. So, the question is how we can stitch them together into a panorama. put them together to create a 3D photo.
- However, standard panorama stitchers will not work as the images were captured from different locations. So they will not align, because of parallax.

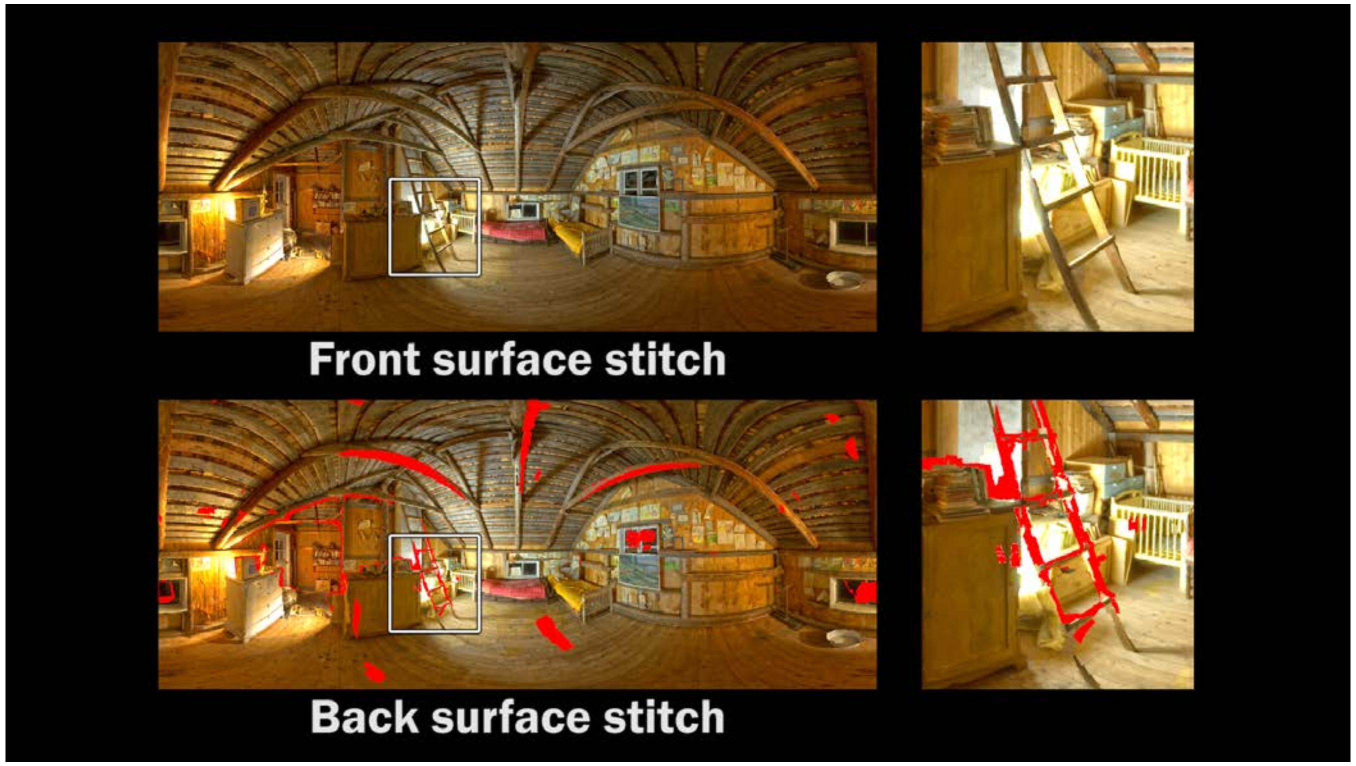


- But thanks to the depth maps, we can warp the images and **re-render** them as if they were all taken from a common central viewpoint.



- Let's take a look at how the images look in the central panorama.
- Take note of how well they are fitting together now.
- For each pixel in the panorama, we need to know from which input image we should fetch depth and color.
- We formulate this as a labeling problem, and use discrete optimization to produce a panorama with both depths and colors.

- Just producing a pano with depth is not enough. As you can see, this only represents the foreground objects in the scene and we cannot peek behind corners.



- We use a cool trick where we invert the depth-test to produce a back-surface stitch. Take a look at the paper if you're interested in the details.
- This second stitch looks very similar to the front-surface stitch, except it shows unique content that is not visible in the foreground layer.
  <Click... details appear>
- Focus your attention on the red highlights here. This is where the back surface stitch contains unique content.
- See how the ladder appears eroded away, and we see some reconstructed background that we can use in the 3D photo.

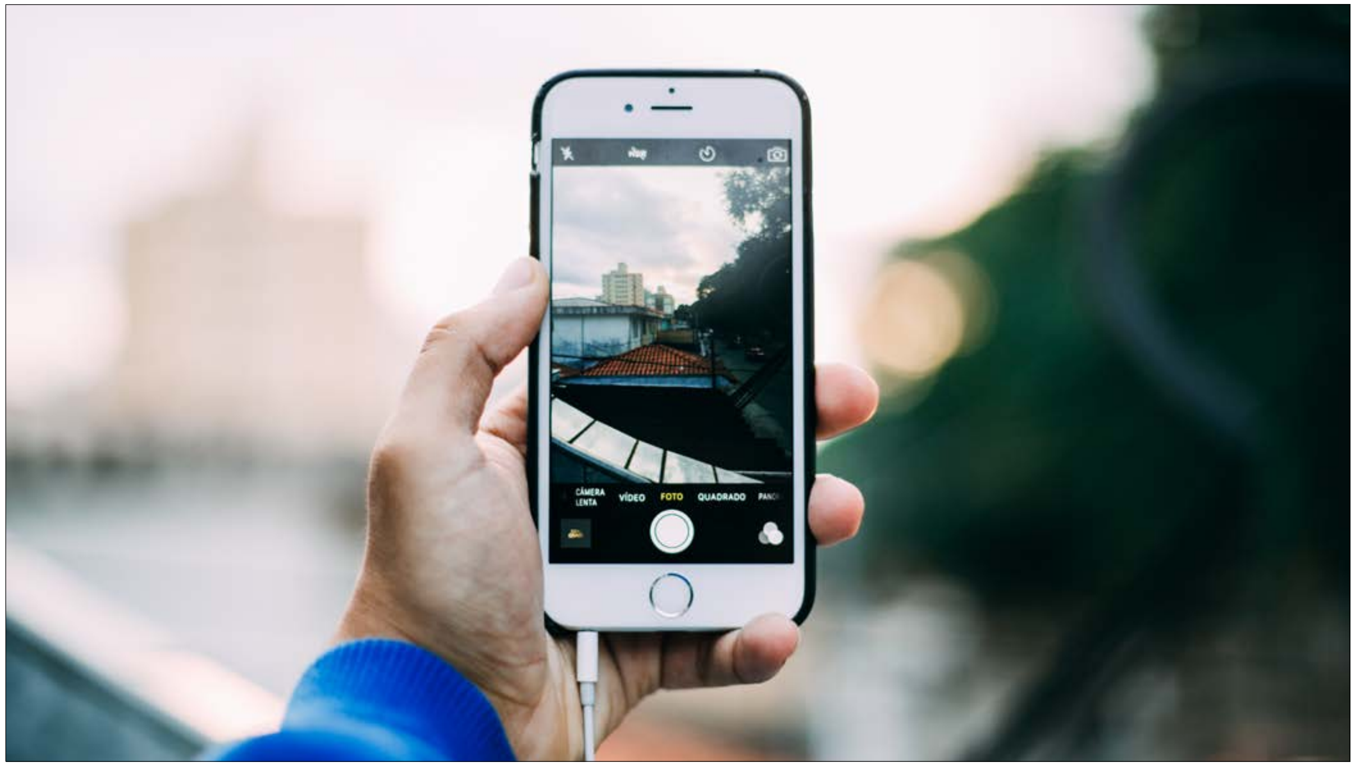
- Let's take a closer look at this. This is the foreground stitch. I will in a moment switch over to showing the back stitch. Watch what happens to the tree...
  <Click... Back stitch appears.>
- Now it is gone. This is the back background layer.
- We can fuse both of these layers together.
  <Click... Connect front/back stitch appears.>
- And now we have a representation where you can peek behind objects.

<Click... Expanded back layer appears.>
- To allow for even more camera motion, we extend the background layer and smoothly fill in the colors.
  <Click... Parallax animation appears.>
- This is all we need to display our scene in VR and enable free viewpoint changes.

- Let's now take a look at our results.
- I should be telling you that I captured a large variety of scenes to show the robustness of our approach and compare it with other approaches: Indoor scenes, outdoor scenes, thin structures, reflective objects and so on.
- However, I really just wanted to go back to Finland and capture my home town in 3D. You'll see the attic in my old summer cottage, the church in the town center and even my old high school.
- So lean back, and enjoy the beautiful views of Jakobstad.
- If you're interested in more details and results, please take a look at our supplemental material.

However, it takes a WHILE to process these 3D reconstructions. It takes up to 4 hours to process a scene!

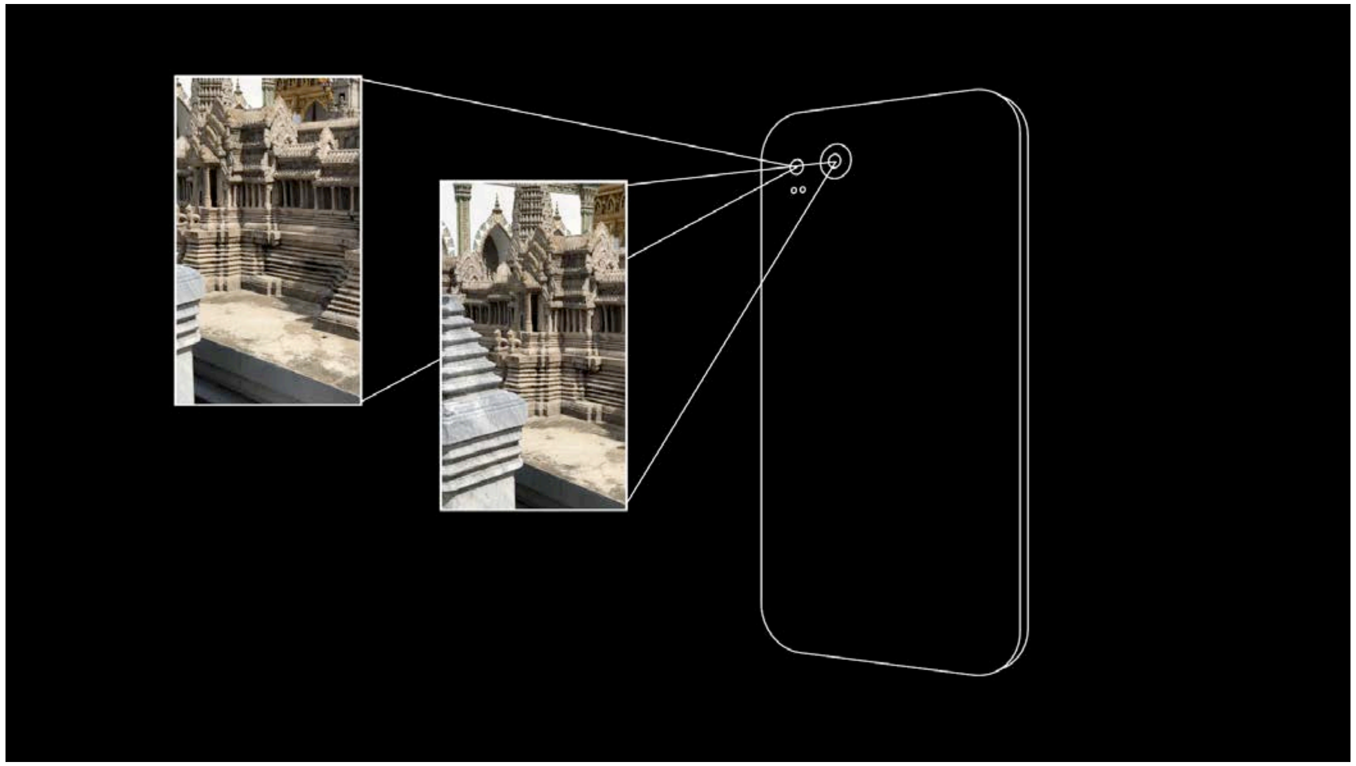
This is like the good old days of film photography: It's easy to take pictures, but you have to wait for the film to be developed before you see the result.



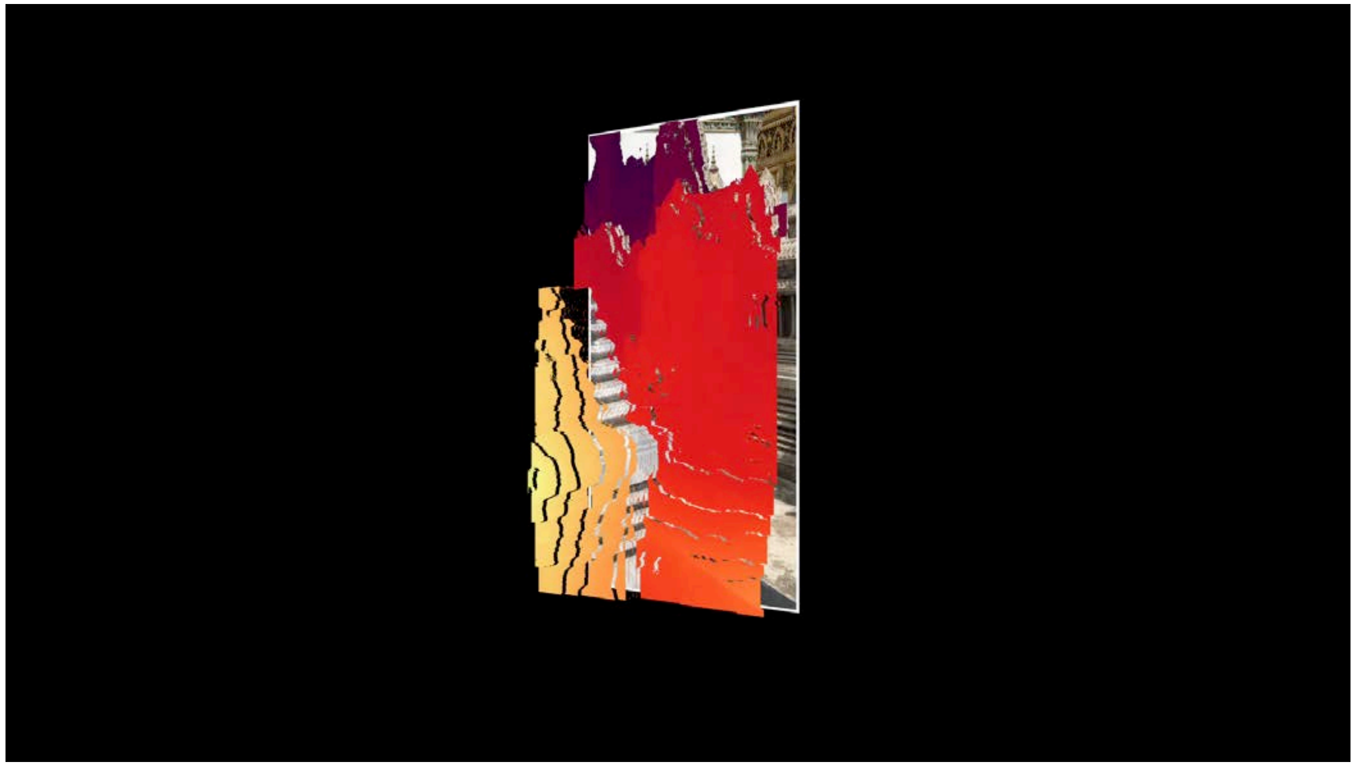
Capture becomes easier with instant feedback, much like taking a picture with a digital camera or a cell phone.

With this in mind, we designed a very fast approach to 3D reconstruction.

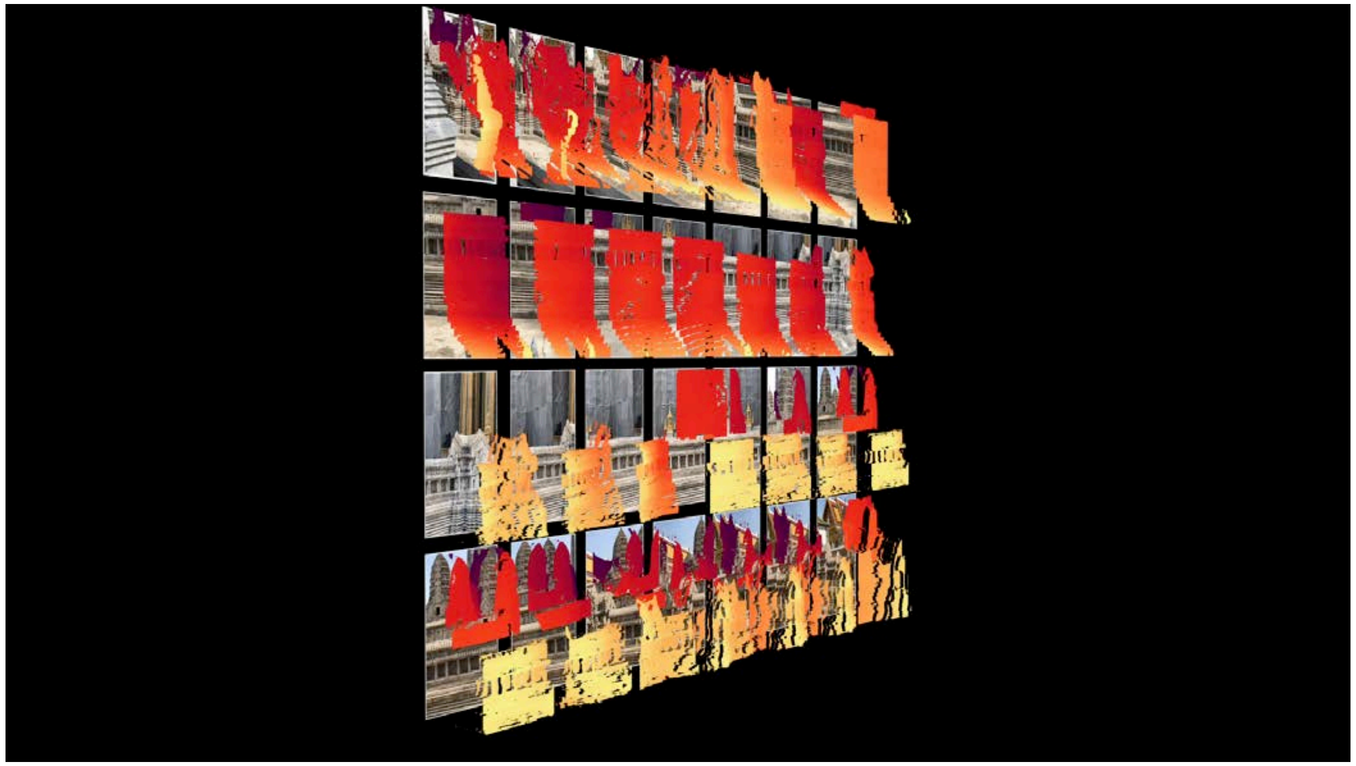
The first design decision was to use a dual camera cell phone which takes two images at every single location.



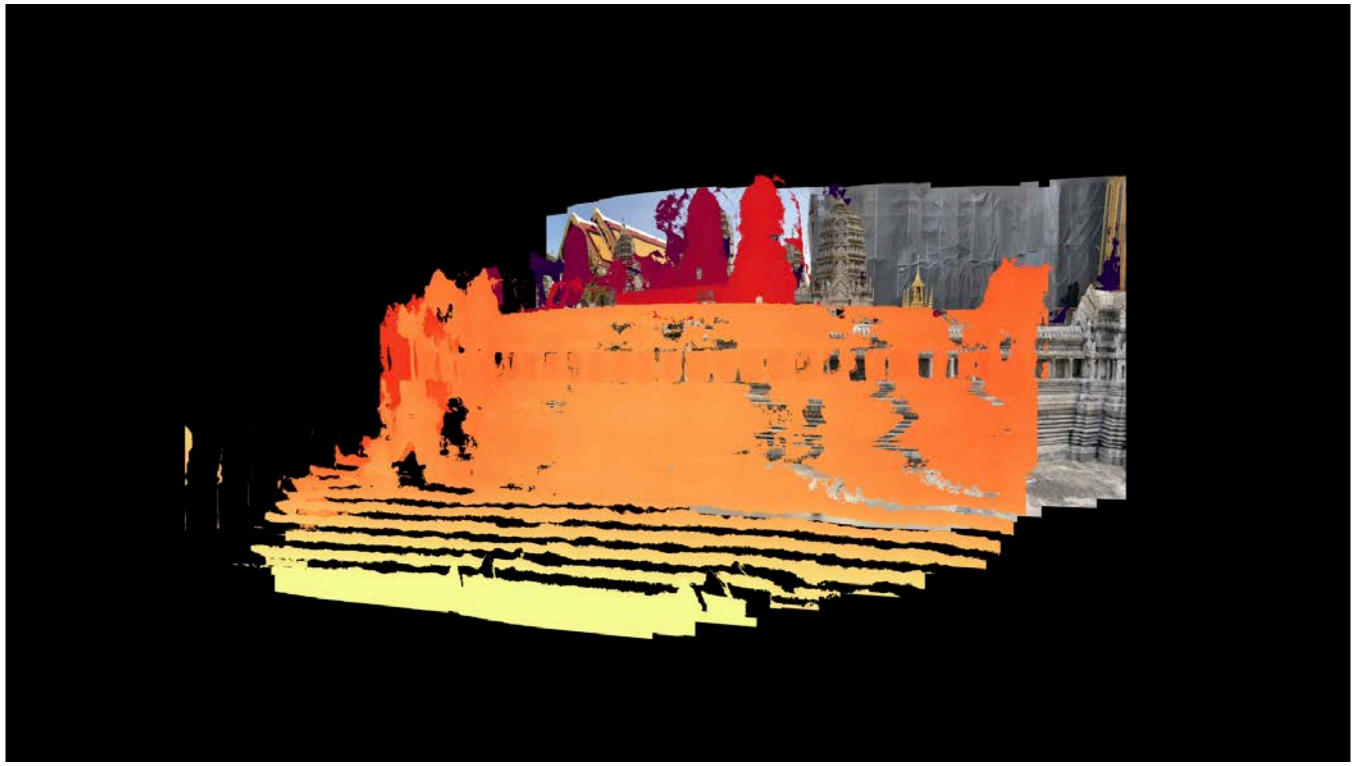
These cell phones also provide fast depth estimation algorithm, which quickly computes a depth map at each location.



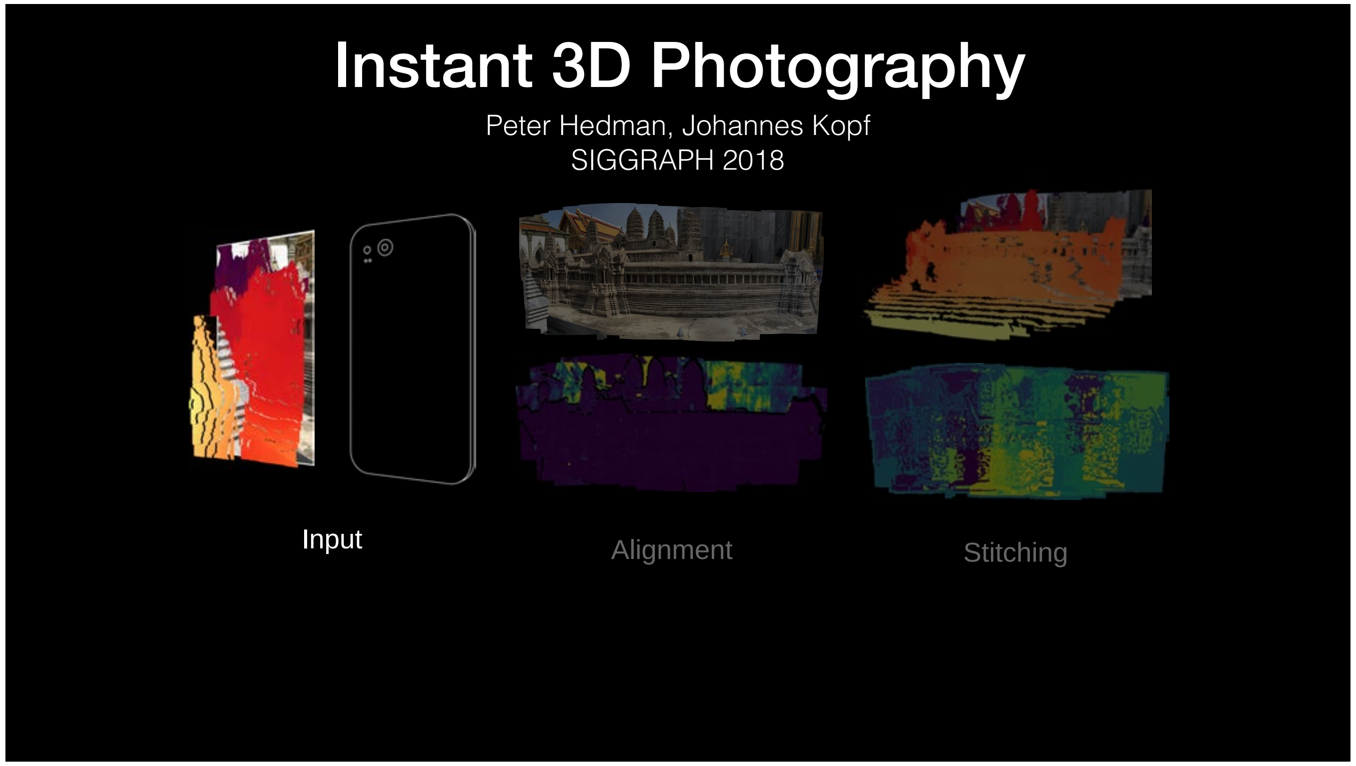
We developed an application that captures a burst of these color-and-depth photos.



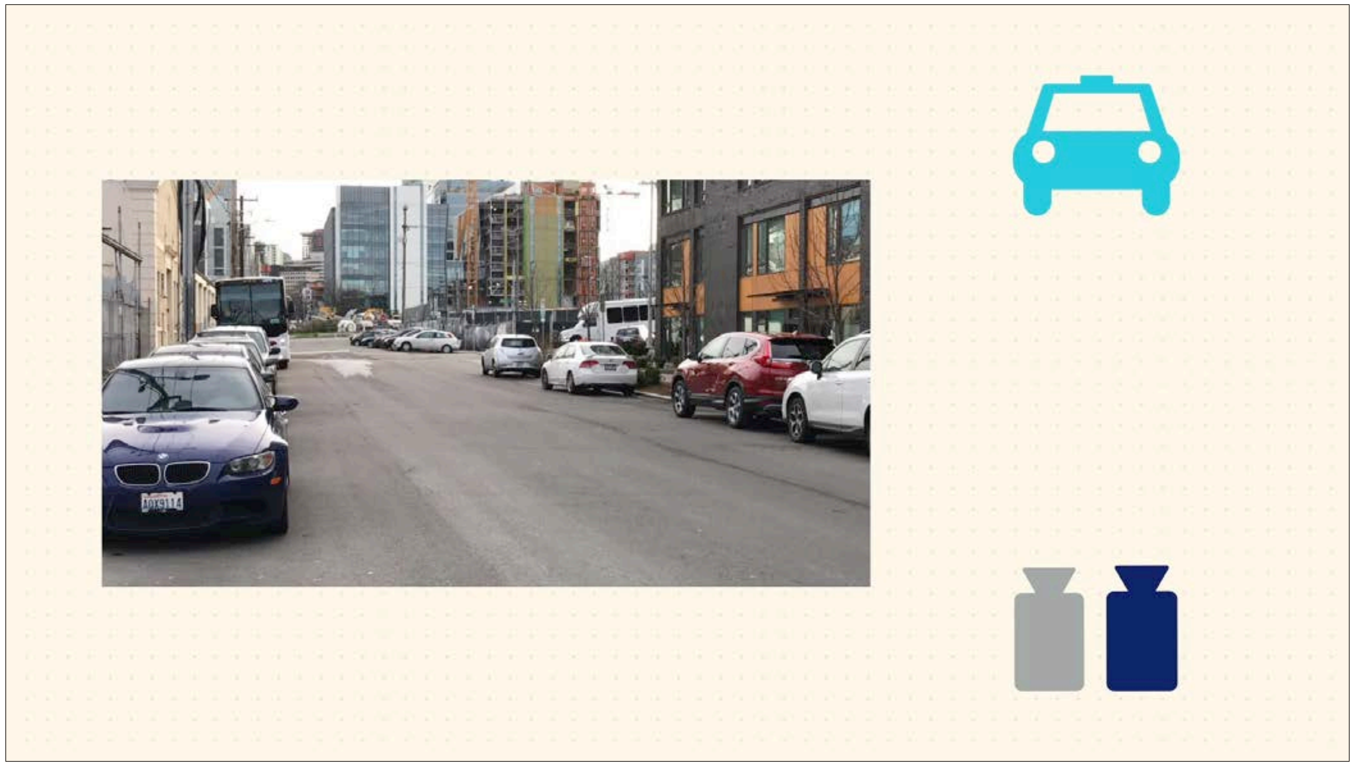
And the main technical contribution is a fast approach to align these photos in 3D, combining them into a color-and-depth panorama.



And then, with a meshing approach we talked about earlier, we turn this panorama into a 3D mesh that enables viewpoint extrapolation in VR.



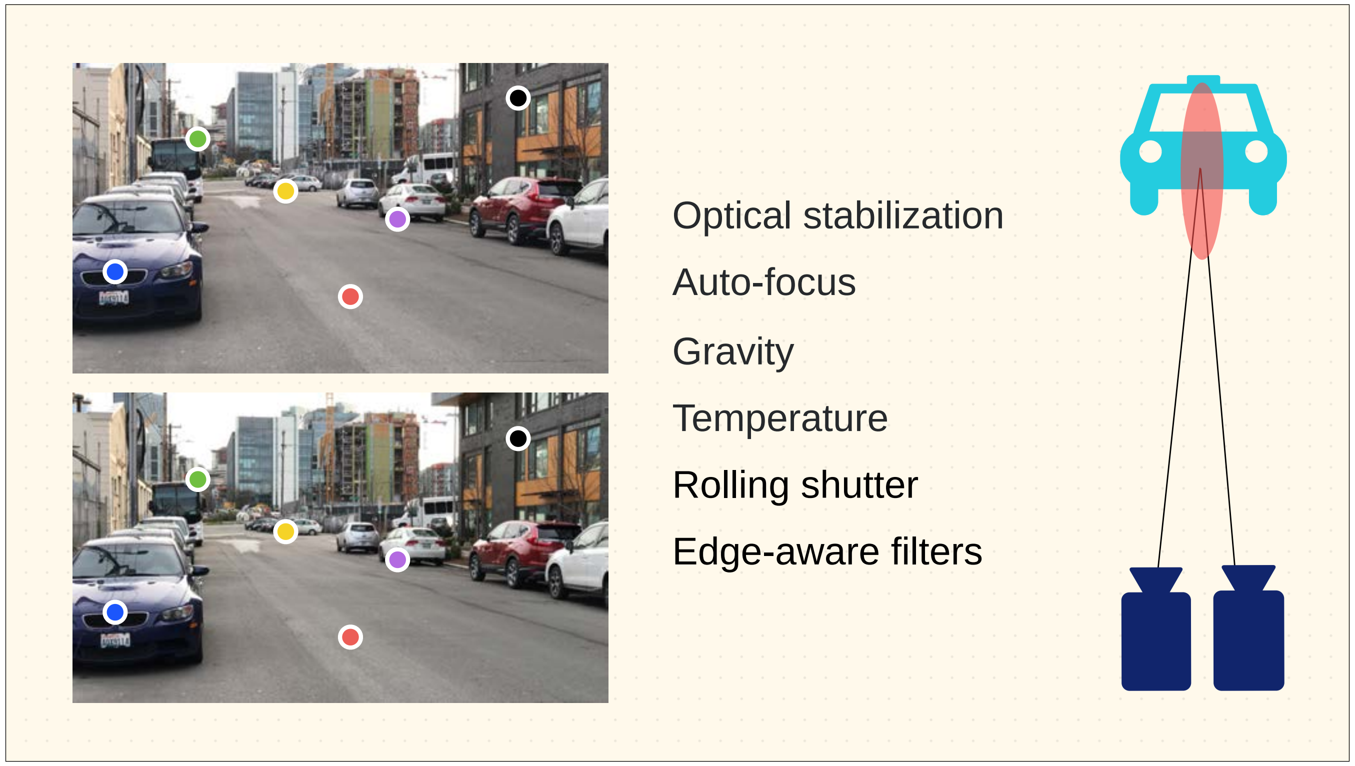
Let’s first discuss the characteristics of our input data.



This is the image pair you get from a dual-camera phone.

Since the distance between the two cameras --- the baseline --- is so small, the images are barely different.

This makes stereo matching fast, as you can keep the search region small.



However, the triangulation angle is so narrow, which often makes the reconstructed depth unreliable.

And it gets worse: In practice the cameras move and rotate independently of each other.

This happens because of otherwise desirable features in cell-phone cameras:
- Auto-focus
- Optical stabilization
- Rolling shutter

In fact, even gravity plays a role here.

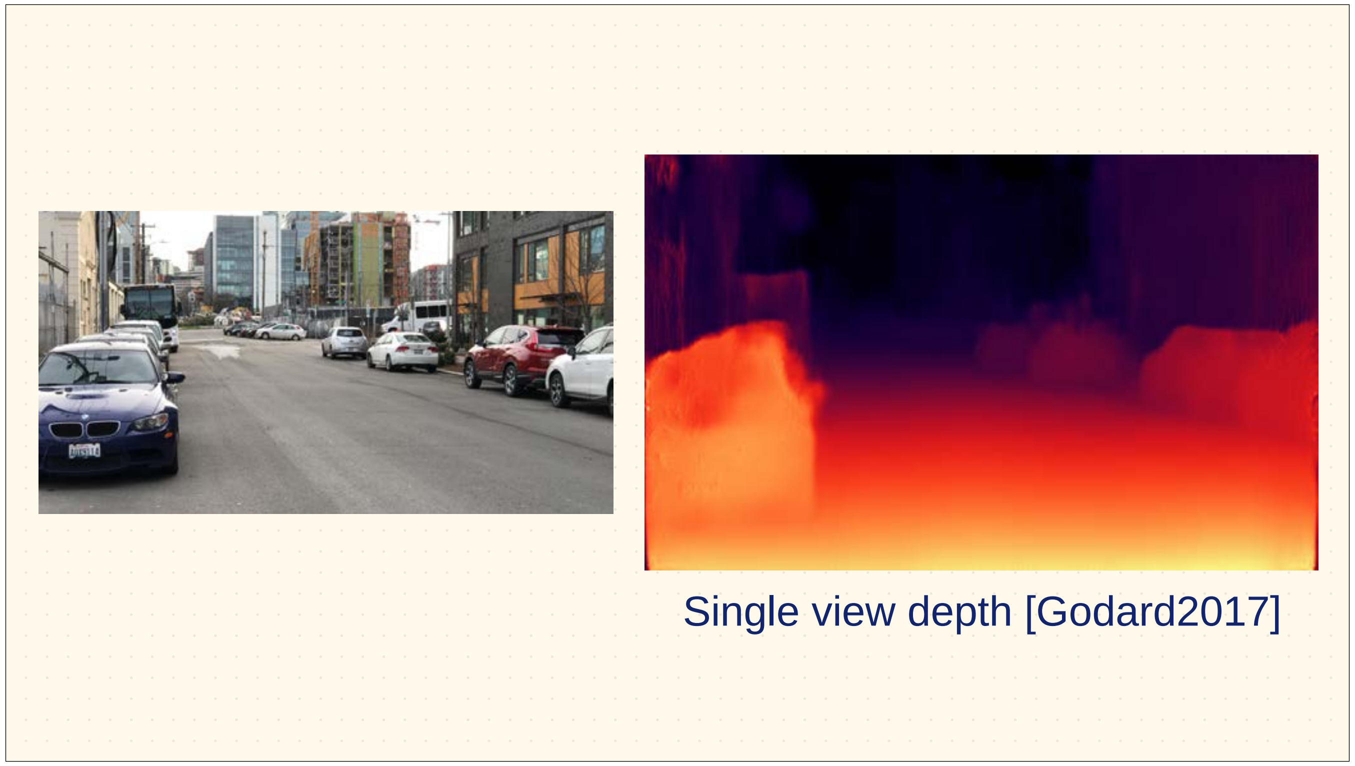
This is why most algorithms use aggressive edge-aware filtering

resulting in smooth depth maps that also respect object boundaries.

Unfortunately, this introduces a large low-frequency error on the estimated depth values.

We found that other small-baseline depth estimation methods also suffer from similar degradations

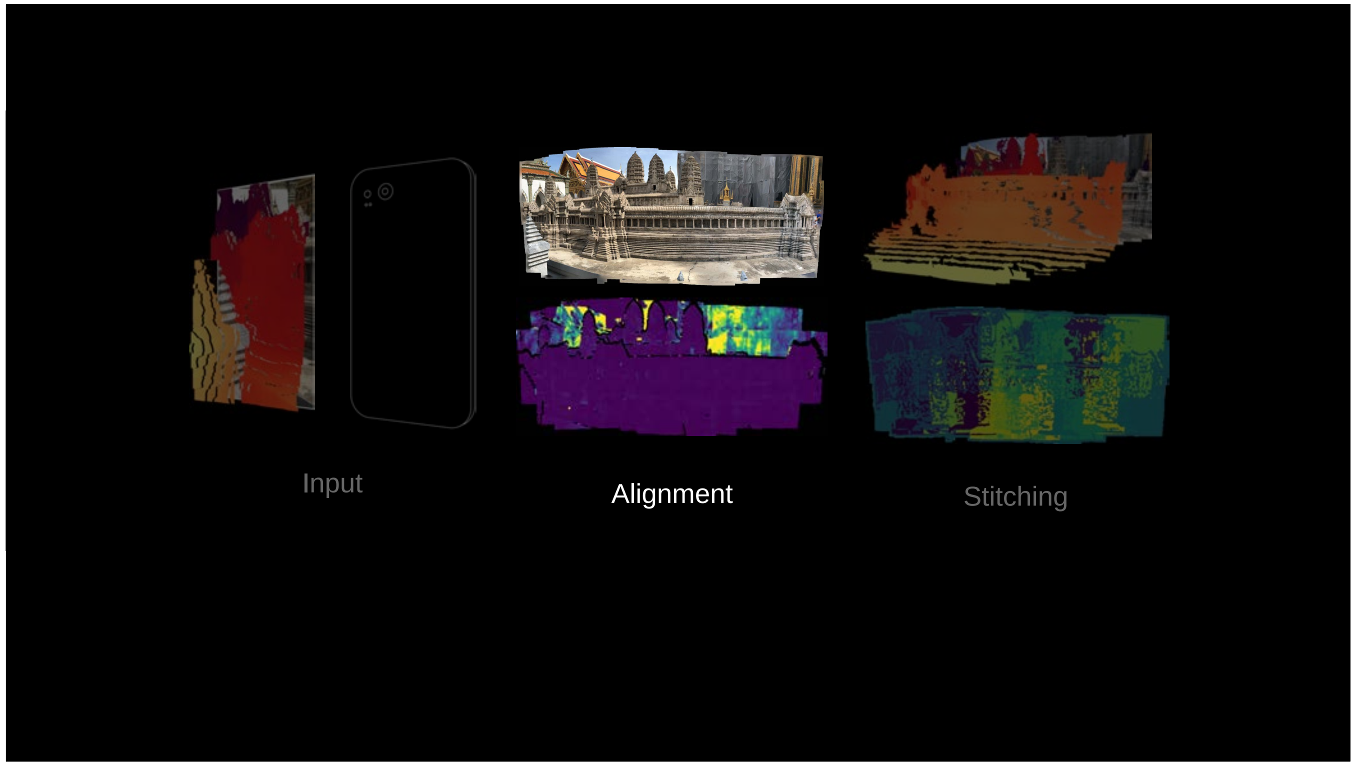
With the help of aggressive edge-aware filtering



For example: single-view depth estimation using convolutional neural networks

Small motion clip [Ha2016]

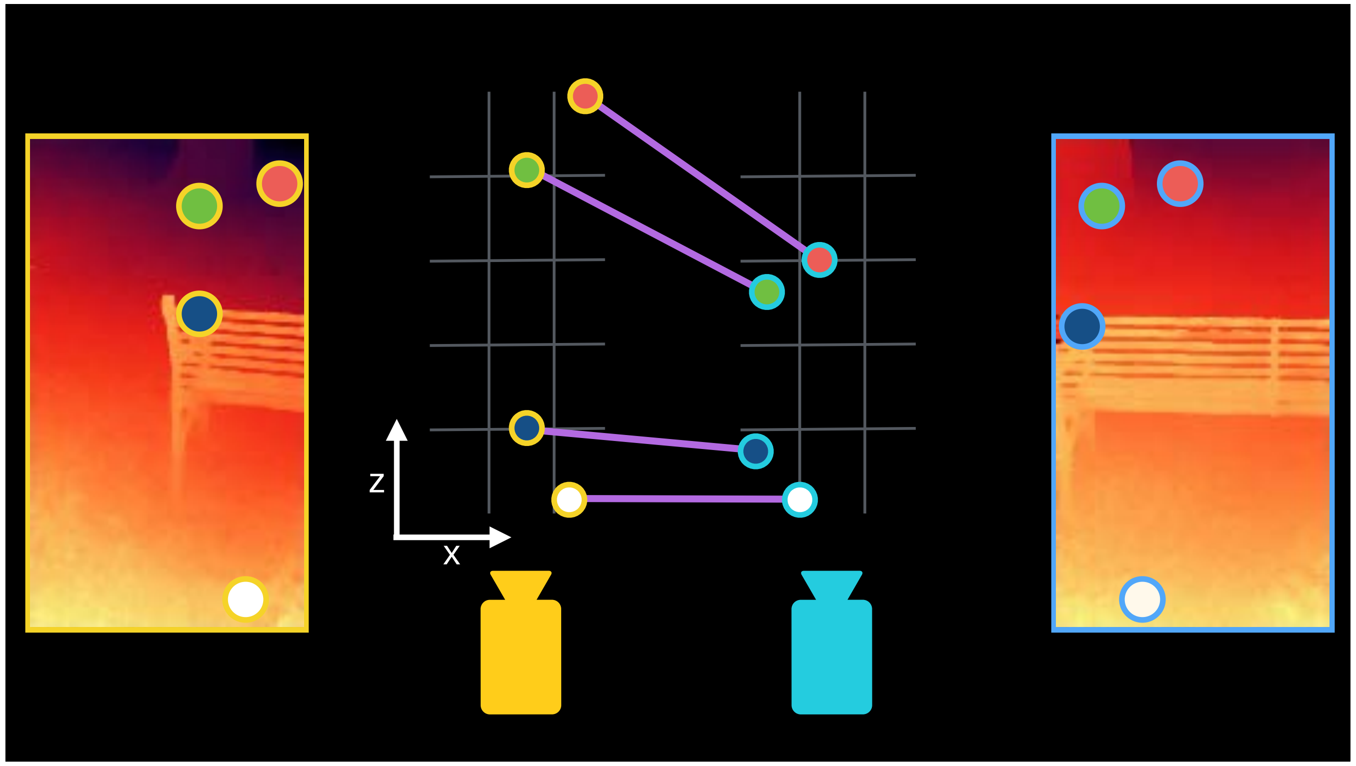
Or: Depth estimation from a short video clip



So how can we get rid of these deformations?

Remember that we have an application which captures a burst of these color-and-depth images.

This provides redundant information, and we can establish feature point correspondences between the images to reason about the 3D structure in the scene.

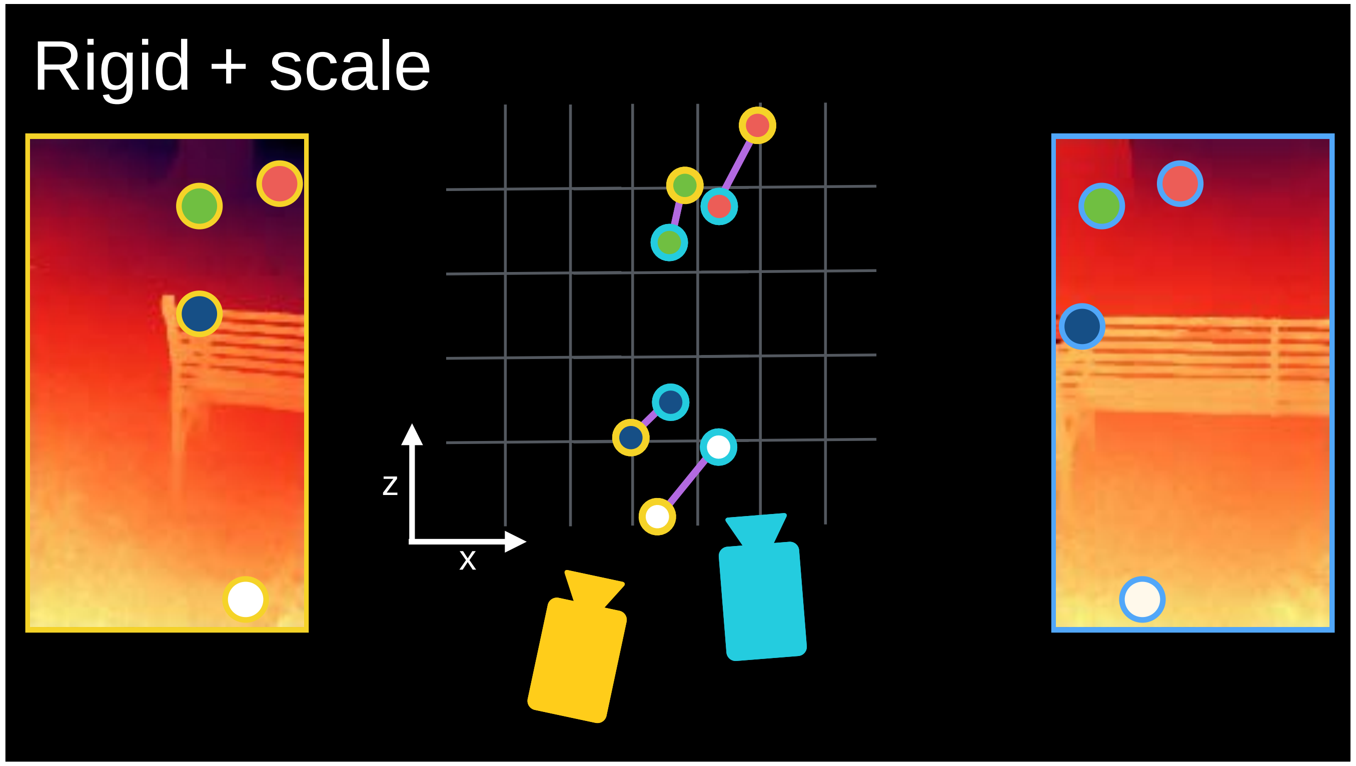


Let’s take a look at two images and a few matching feature points.

Since we know the camera intrinsics and have depth maps, we can project these points out into 3D, forming two point clouds.

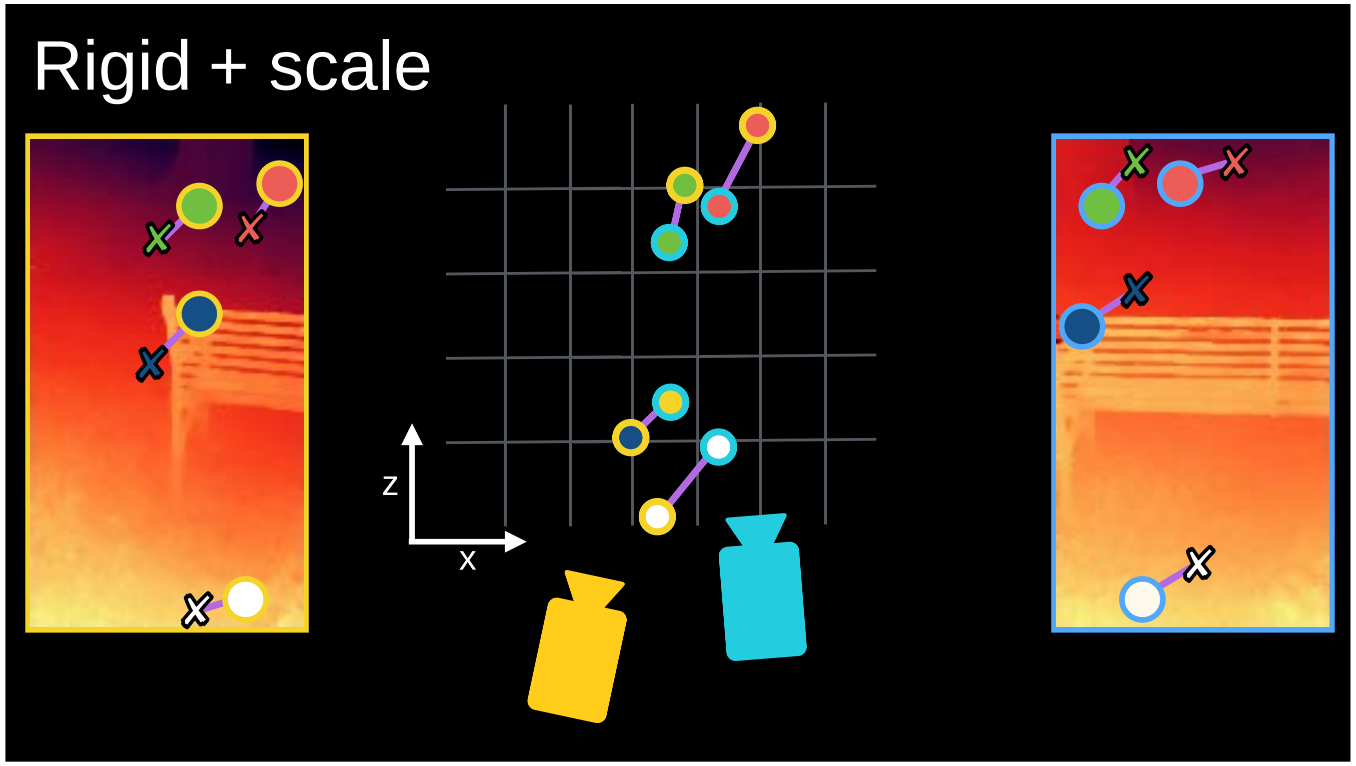
Already here we can see that there’s a scale difference between the two depth maps.

We can now try to better align these two images, optimising for a rigid transformation that minimises the 3D distance between matching feature points.

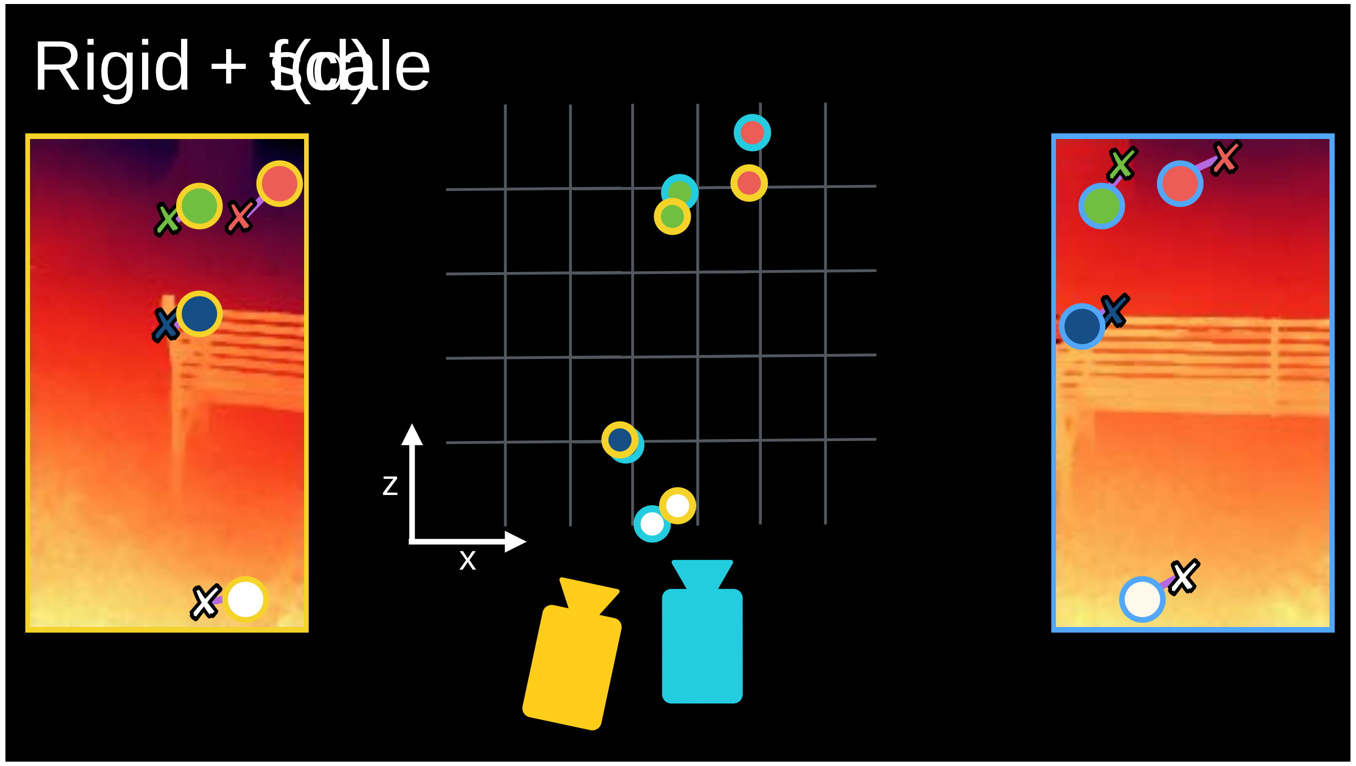


But this is not quite enough, as there is a scale difference between the images.

If we also optimize for scale, we run into trouble —we can easily reduce the error to zero by shrinking the scene!

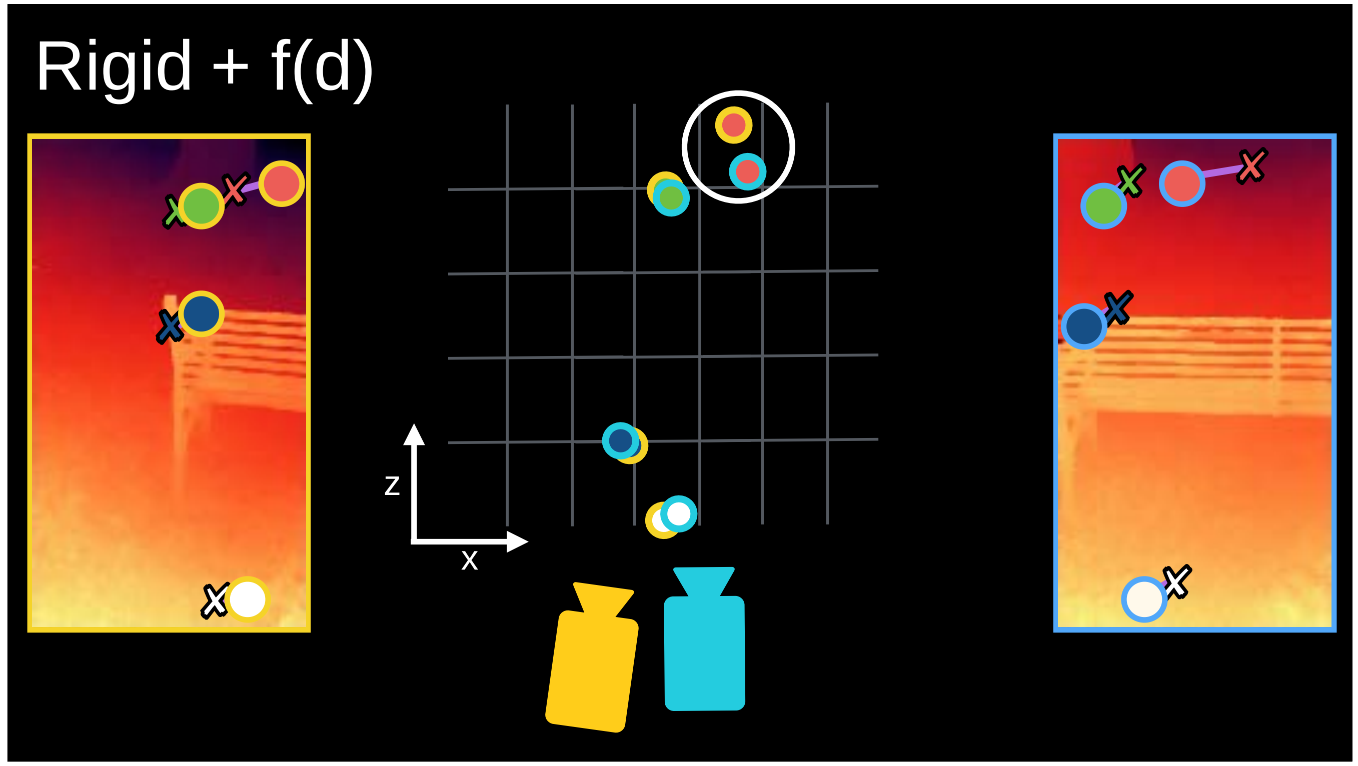


Instead, we found it better to measure the 2D distance between feature points. We use the depths to project each corresponding point into the other image. This allows us to optimize for scale without having the trivial solution of shrinking the scene.



Unfortunately, the deformations on depth cannot be explained by just a per-image scale factor.

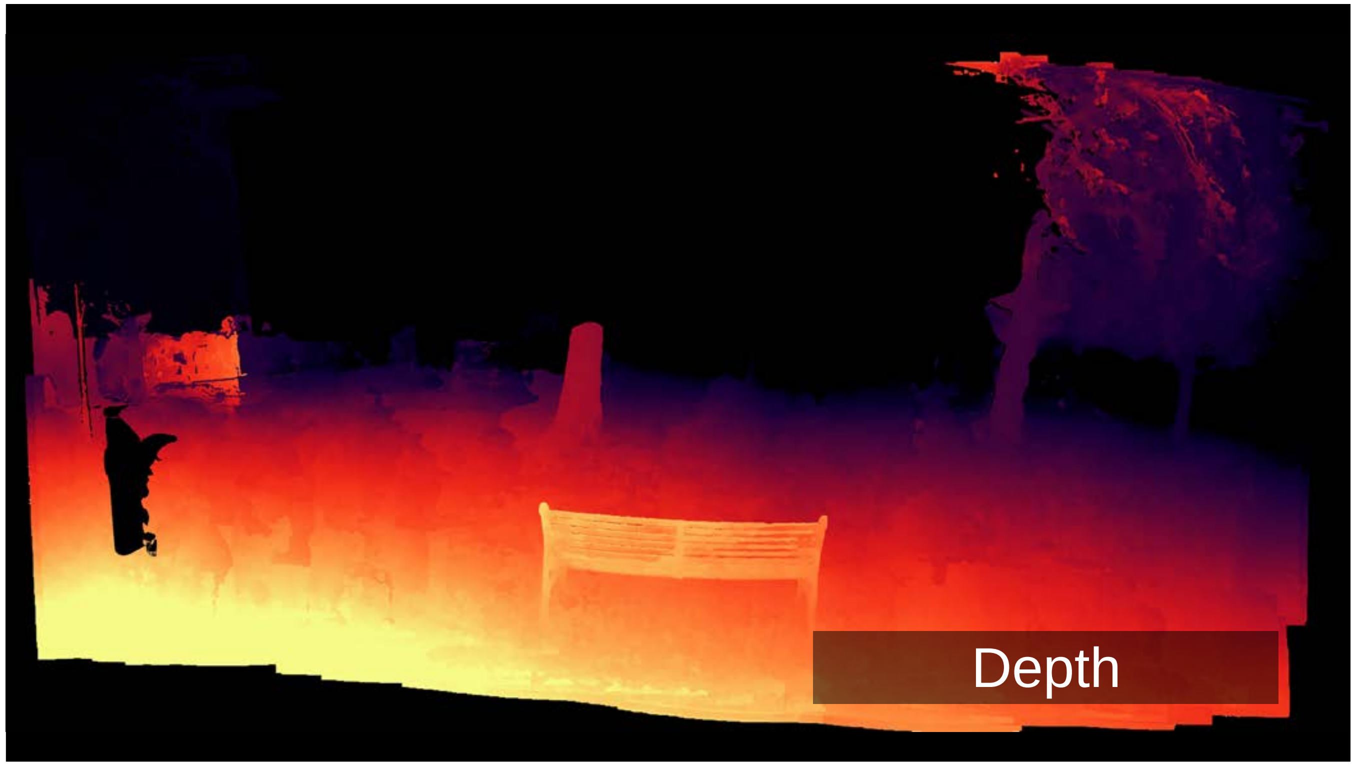
We experimented with more general depth correction functions, and found that an affine transformation on disparities worked best.



But there is still quite a bit of residual error.

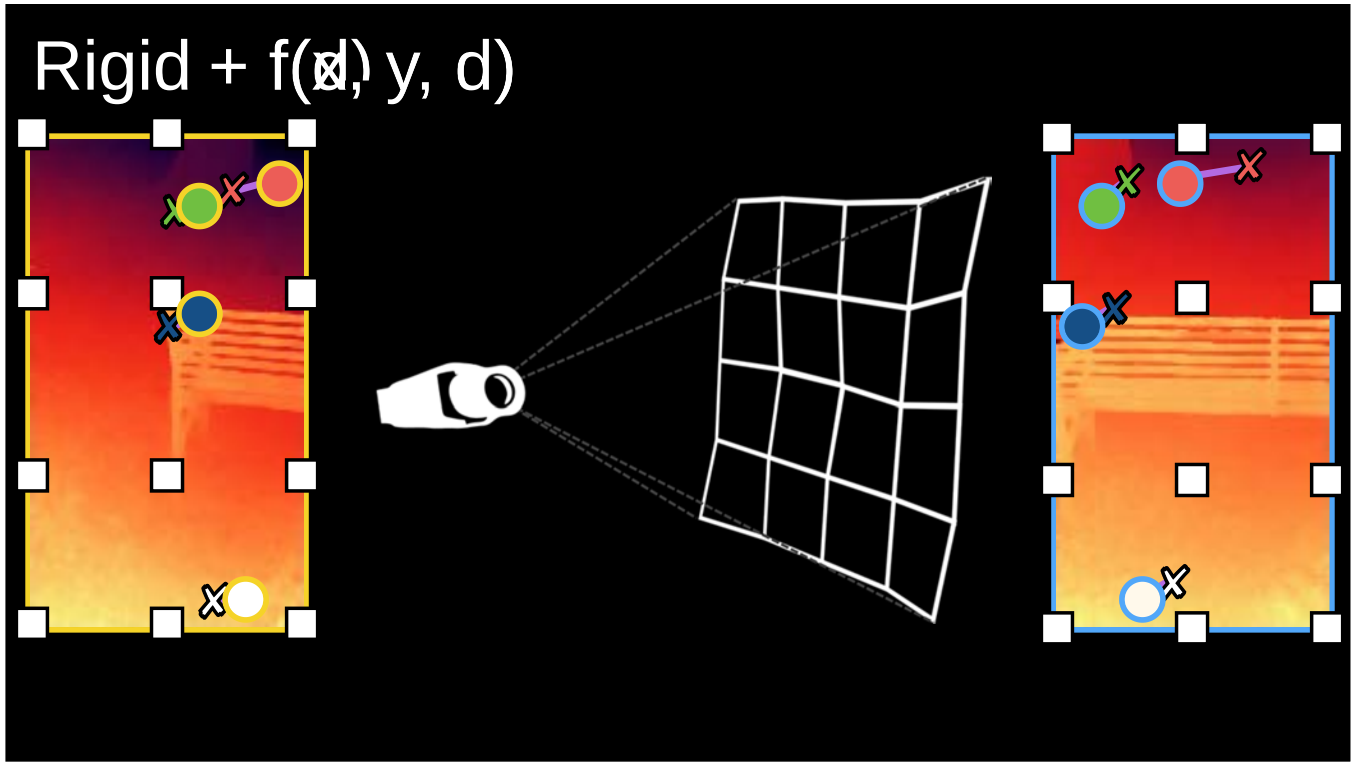
… But maybe this isn't too big of a deal. What if we simply used this alignment to stitch a panorama?

The color panorama looks great.



But unfortunately the depth channel isn't doing quite as well.

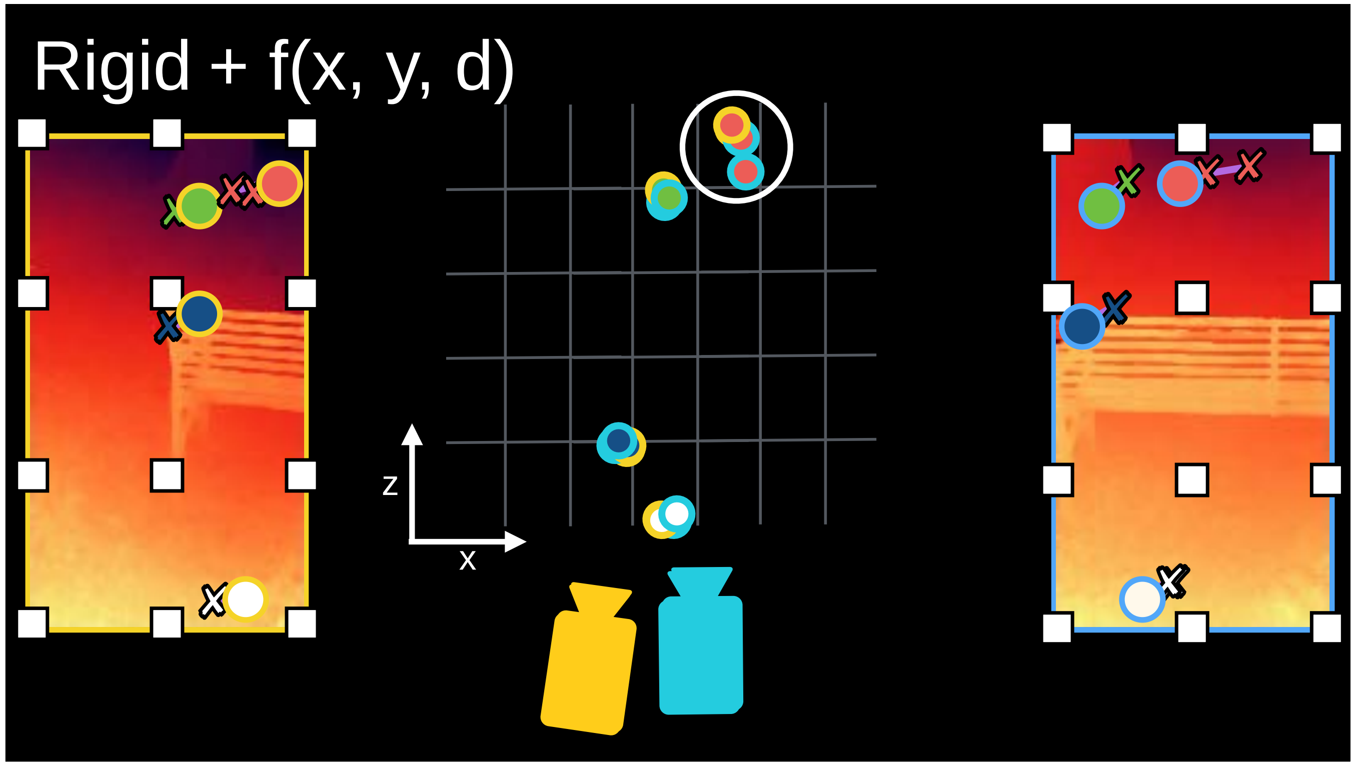
There are a lot of misalignments on the ground, and there are big black regions, where the only feasible solution was to push these parts of the scene out to infinity.



To solve this, we use a depth correction function that varies across each image. We do not want to introduce new structures into the depth maps, so we want it to be smooth.

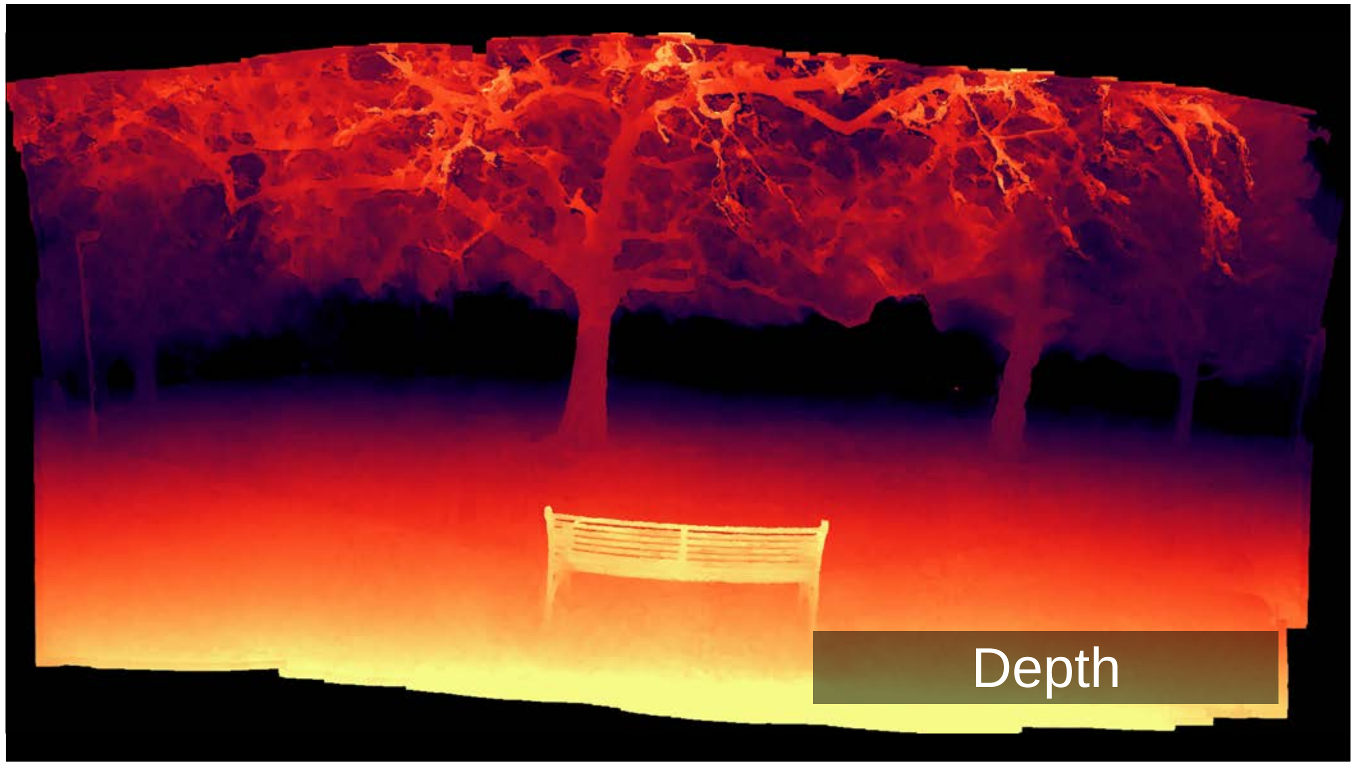
We optimize for coefficients at a regular grid of locations across the image, and use bilinear interpolation to form this smoothly varying depth correction function.

With this approach, we can isolate the feature points in the top right corner of each image, and use a slightly different depth correction there.



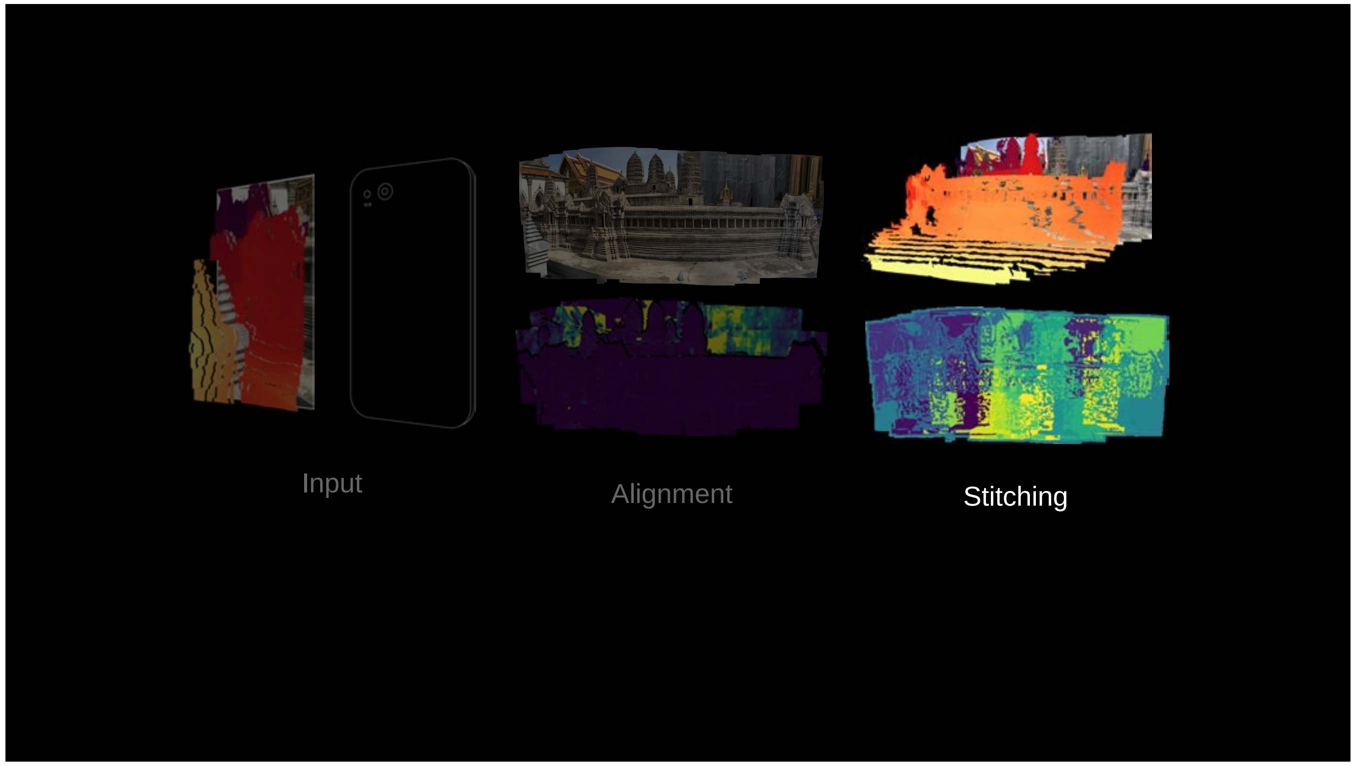
This brings everything into much better alignment.

And if we take a look at the resulting panoramas, the colors still look good.



But there’s a huge difference in the depth channel.

The ground is much smoother, and there are no big black regions anymore.



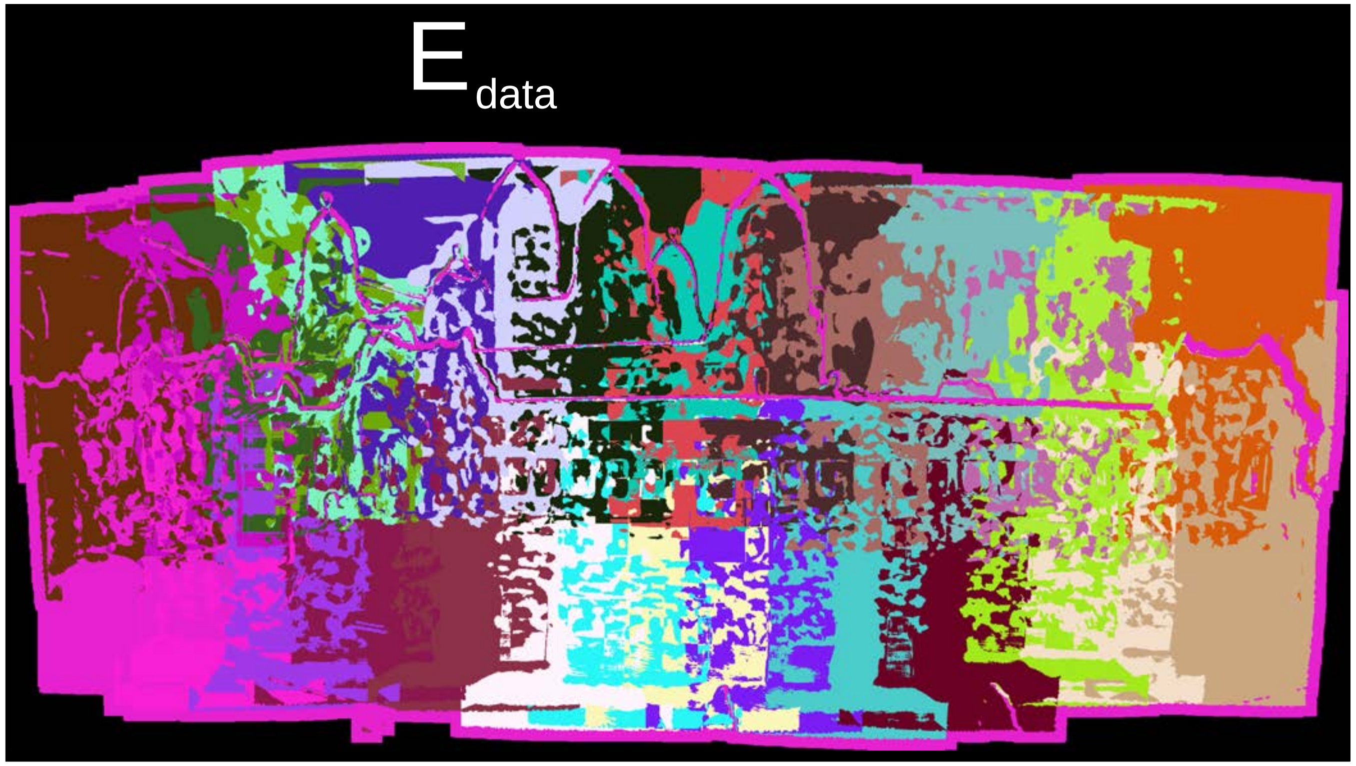
This allows us to quickly align the images in 3D. But how can we speed up stitching?

Earlier in this talk we used a seam-hiding stitching approach, which optimizes for a per-pixel data cost and a seam-hiding smoothness cost.

The data cost tells us how well each image works at each pixel.

The smoothness cost reduces the number of image transitions in the panorama, which reduces the opportunities to make mistake. It also tries to hide these transitions in regions where they are hard to spot.

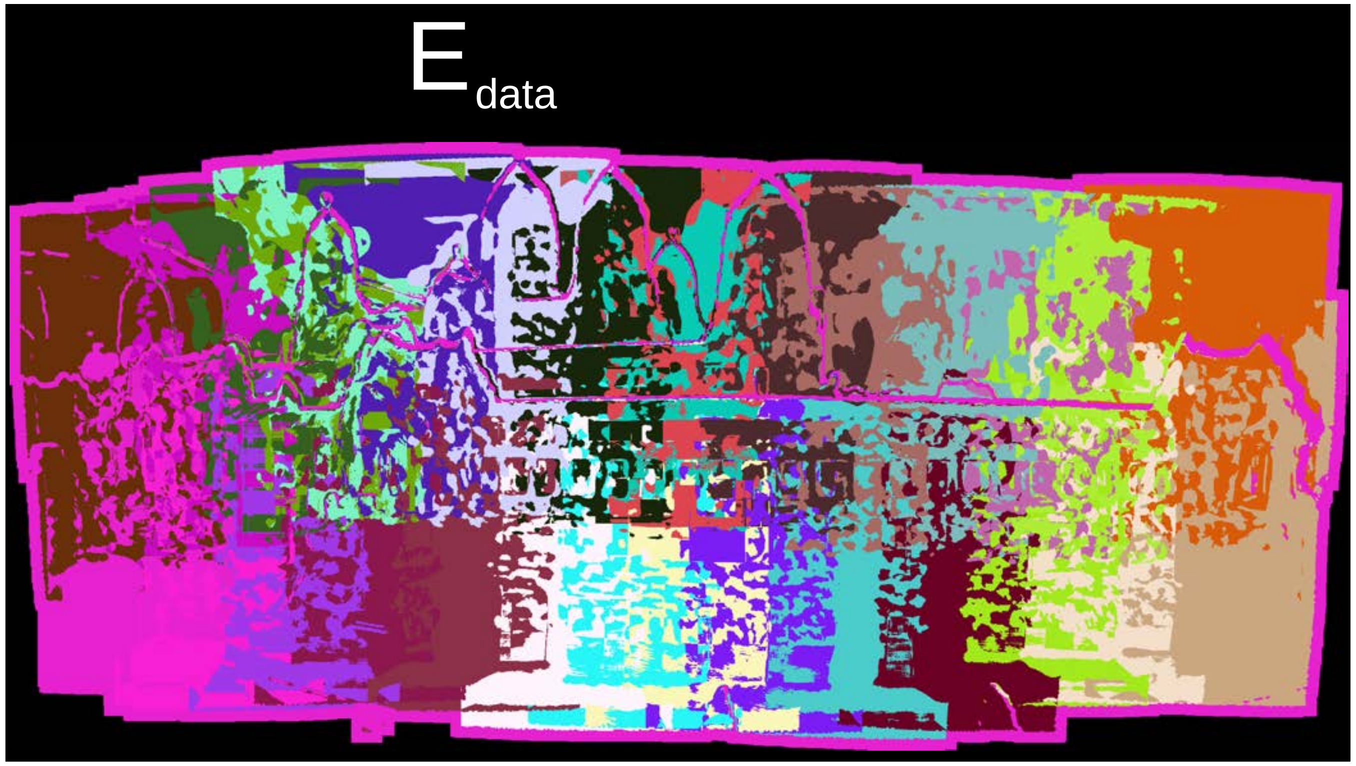
Unfortunately, it is very expensive to optimize for this type of smoothness —using algorithms such as alpha expansion.



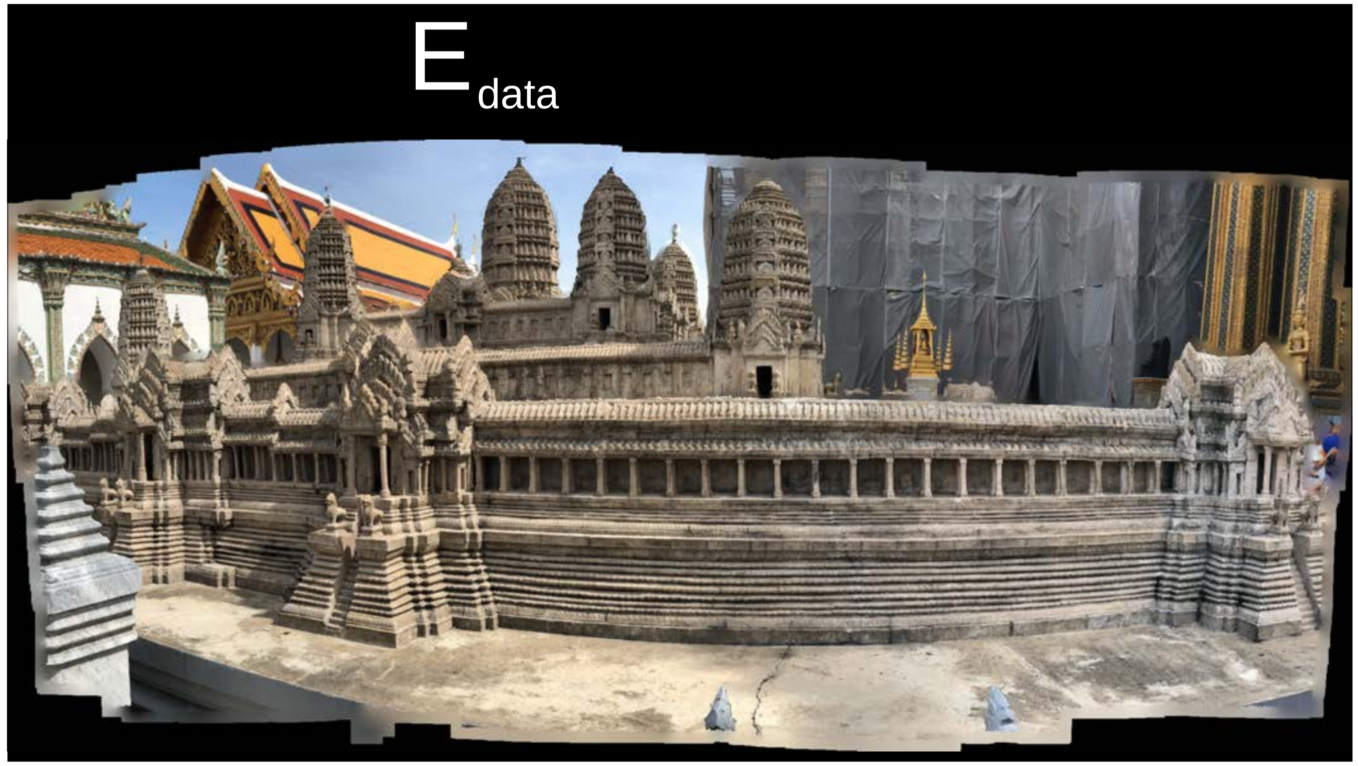
It would be much faster if we could only consider the per-pixel data cost. Which can be quickly obtained for each pixel independently with parallel computation.

However, as you can see, this introduces a lot more image transitions —providing ample opportunity to make mistakes and introduce visible seams.

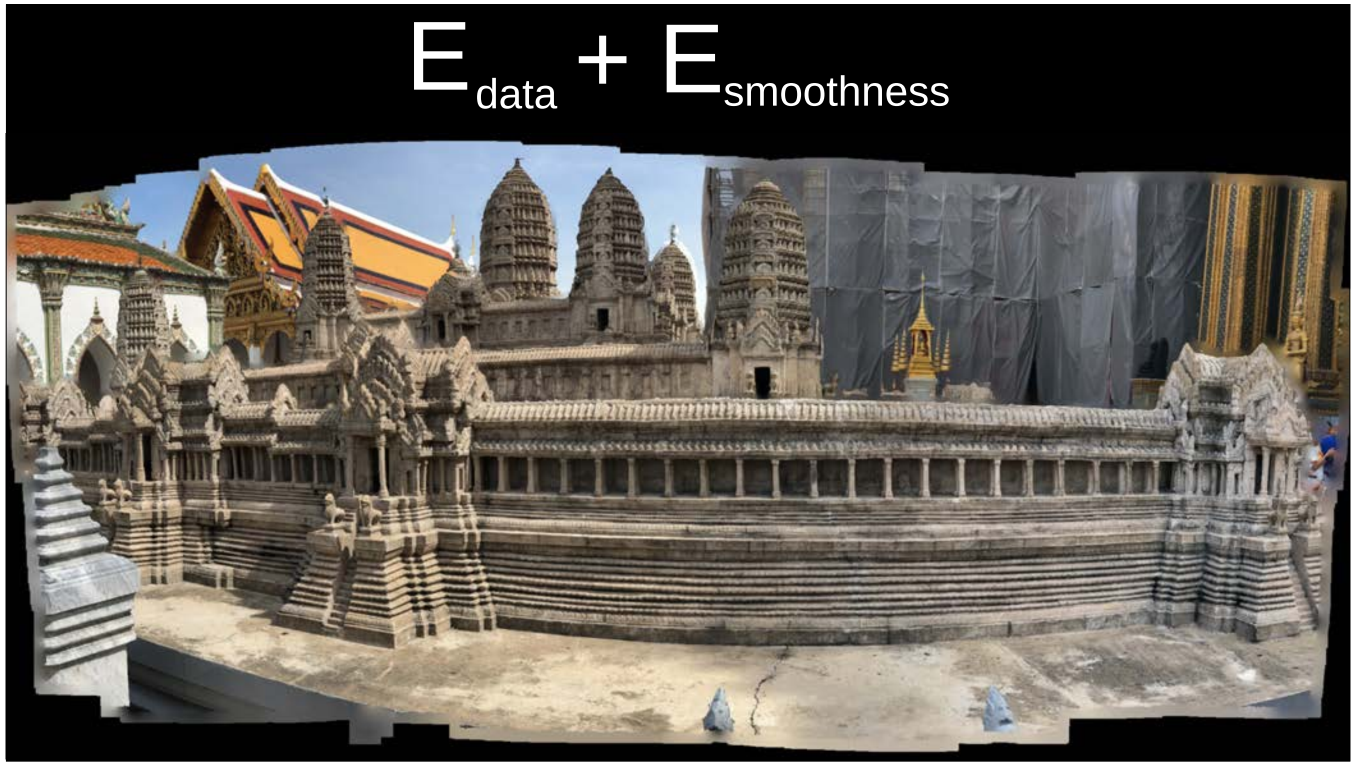
Thankfully our alignment is very precise!



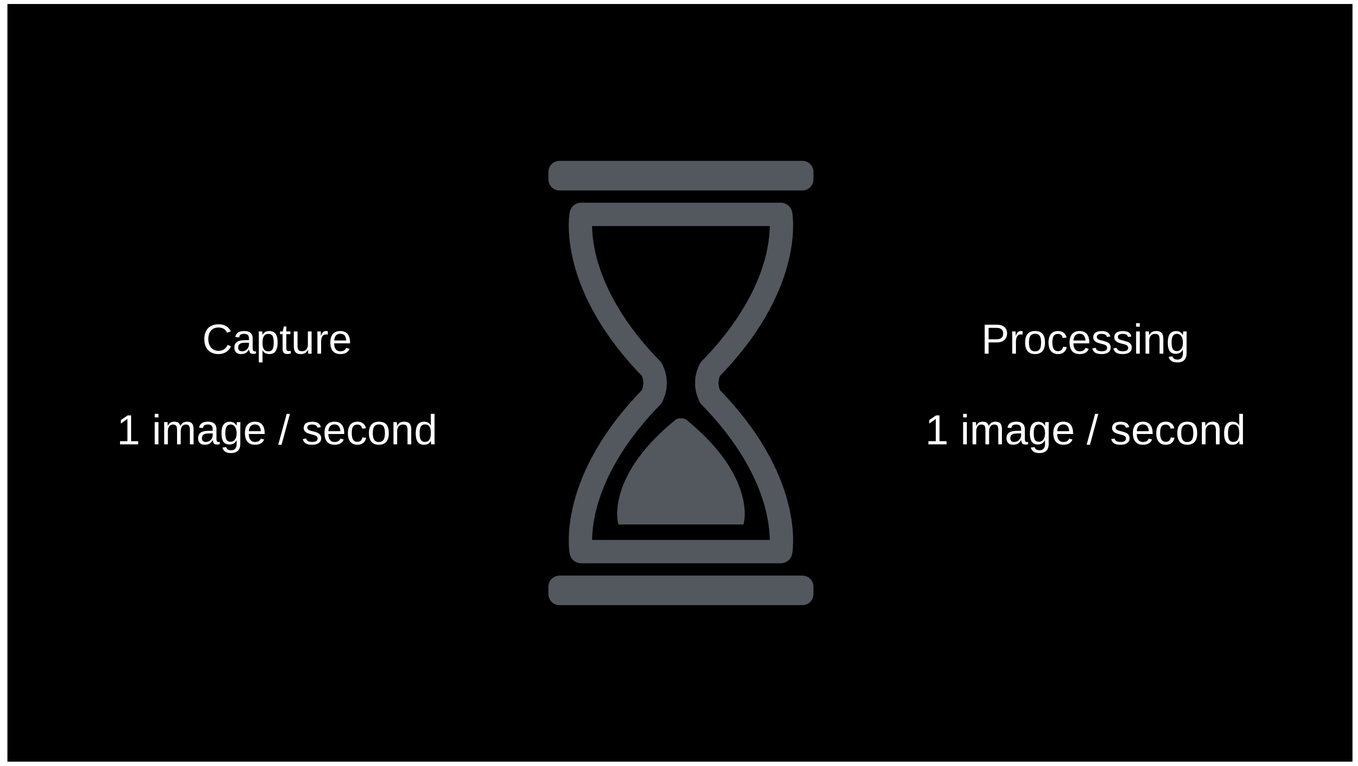
So if we optimize only the data cost



We still get an output panorama that looks great



And is very hard to distinguish from the result which uses slow seam-hiding stitching.



Putting everything together we obtain a very fast pipeline.

The capture application captures roughly one image per second, and the processing takes about one second per image!

Full disclosure, for now, we’ve only run processing on a desktop PC. But this was using unoptimized CPU code, and most bottlenecks can trivially be ported to the GPU.

With these optimizations in mind, and also by interleaving processing with capture, I am confident that you could produce a result almost immediately after capture.

Close, but no cigar

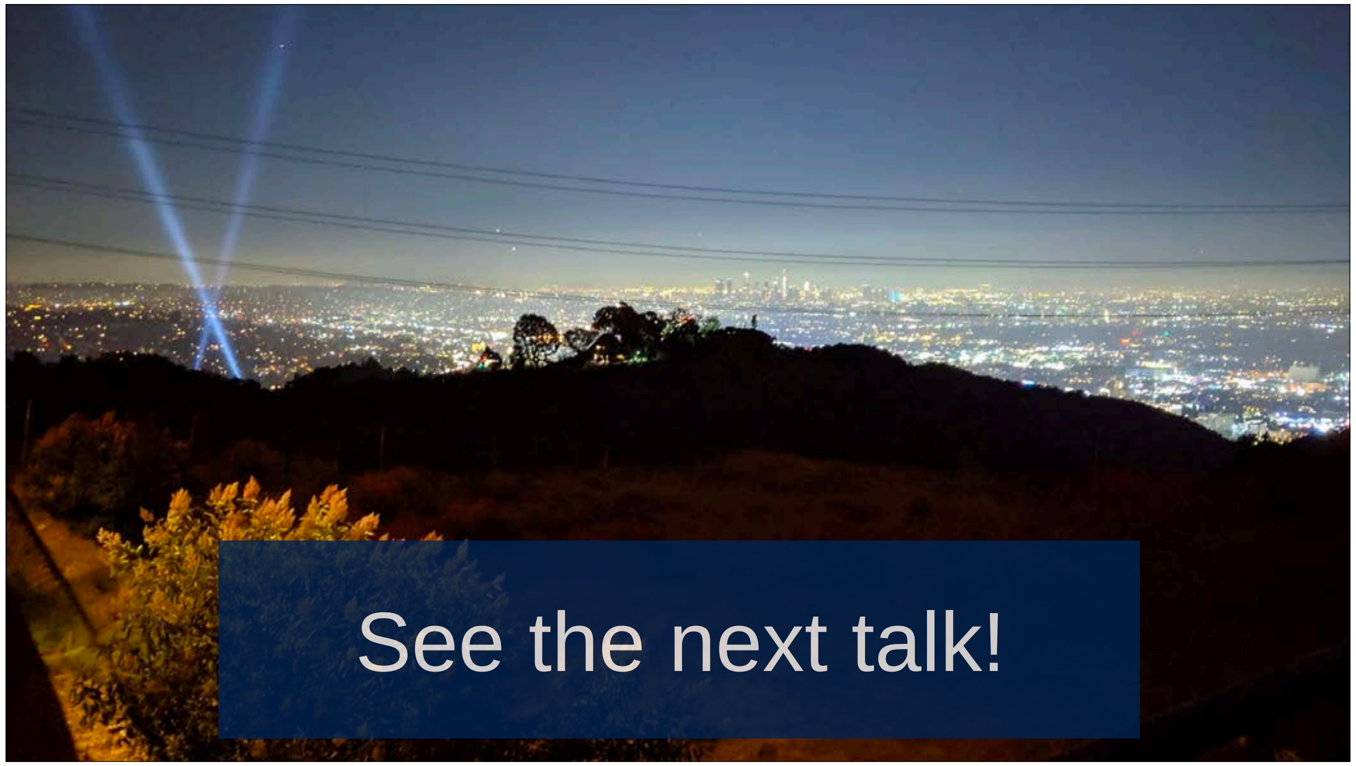
- Overall, the easy capture process, and the fast processing pipeline changed the way I captured 3D photos. I became more opportunistic, and started experimenting more — capturing scenes whenever I felt like it.
- Most of the scenes you'll see here were captured during my vacations. You'll see Bangkok, and some snowy scenes that I captured in Finland over Christmas.

But let's take a step back. In this talk I've talked about technologies that focus on casual capture — enabling anyone to easily capture 3D photos for VR. However, when it comes to quality I can only say: “Close, but no cigar”

- The 3D photo representation discussed in this talk does not support moving highlights. Instead it bakes them into the texture map, resulting in a “painted on” appearance.

These systems also share the limitations of most multi-view stereo approaches. We're unable to reconstruct partially transparent objects such as glass.

Our scene representation also contains only two-layers, so we wouldn't to be able to represent the multiple layers of glass you can see here.



- Finally, we're unable to represent volumetric effects, such as the beautiful Hollywood lights shining through the fog.

For to reach a higher level of realism and quality, we need to bring out the big guns and use a custom built capture rig to obtain full light fields.

Stick around for the next talk where Ryan Overbeck describes exactly this!


Light Field Photography (and Video)
Ryan Overbeck
Google
SIGGRAPH 2019

SIGGRAPH ASIA 2018 TOKYO

A System for Acquiring, Processing, and Rendering Panoramic Light Field Stills for Virtual Reality

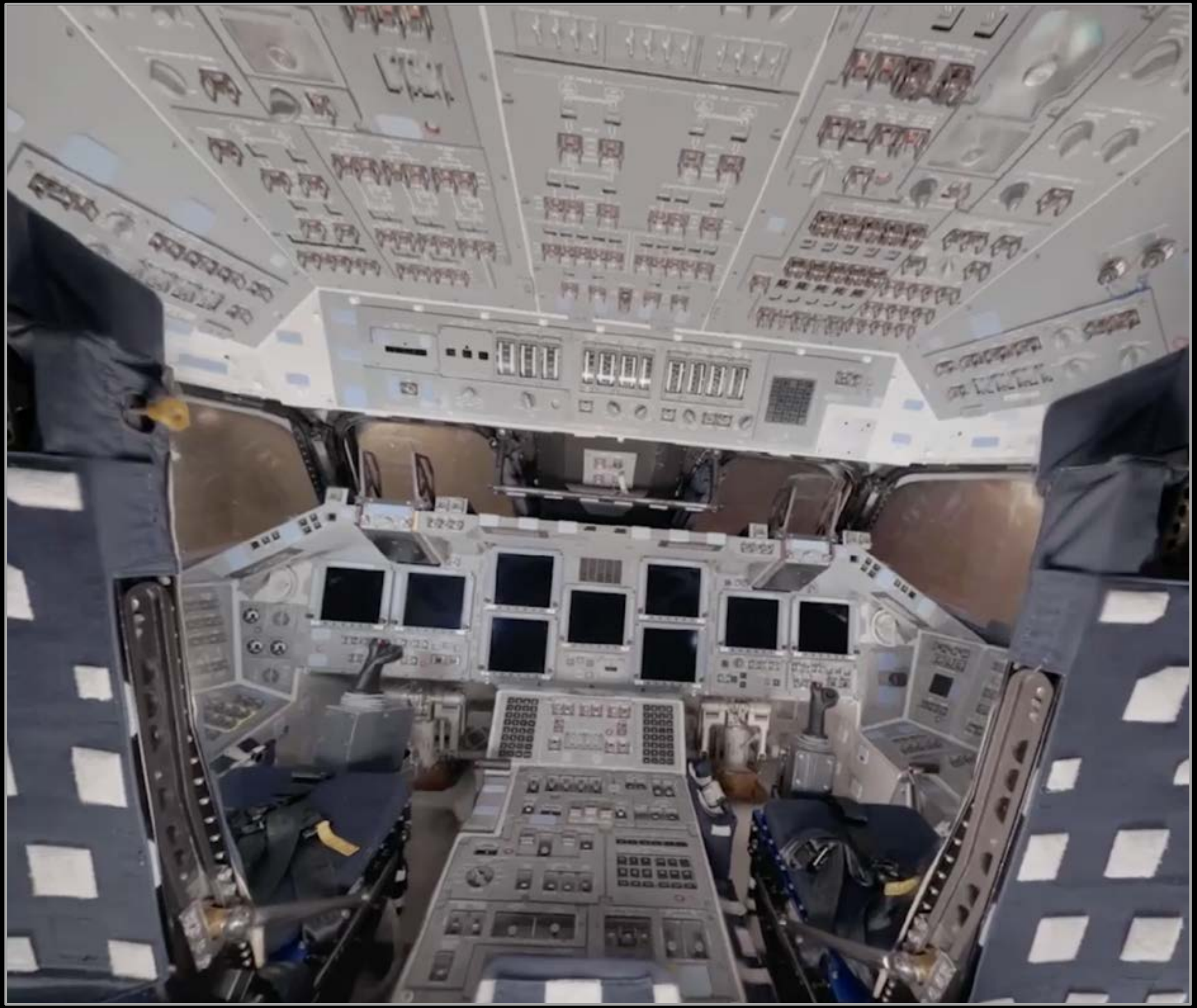
Ryan Overbeck, Daniel Erickson, Daniel Evangelakos, Matt Pharr, and Paul Debevec

Immersive Photography

- **Photogrammetry**
  - 6 DOF

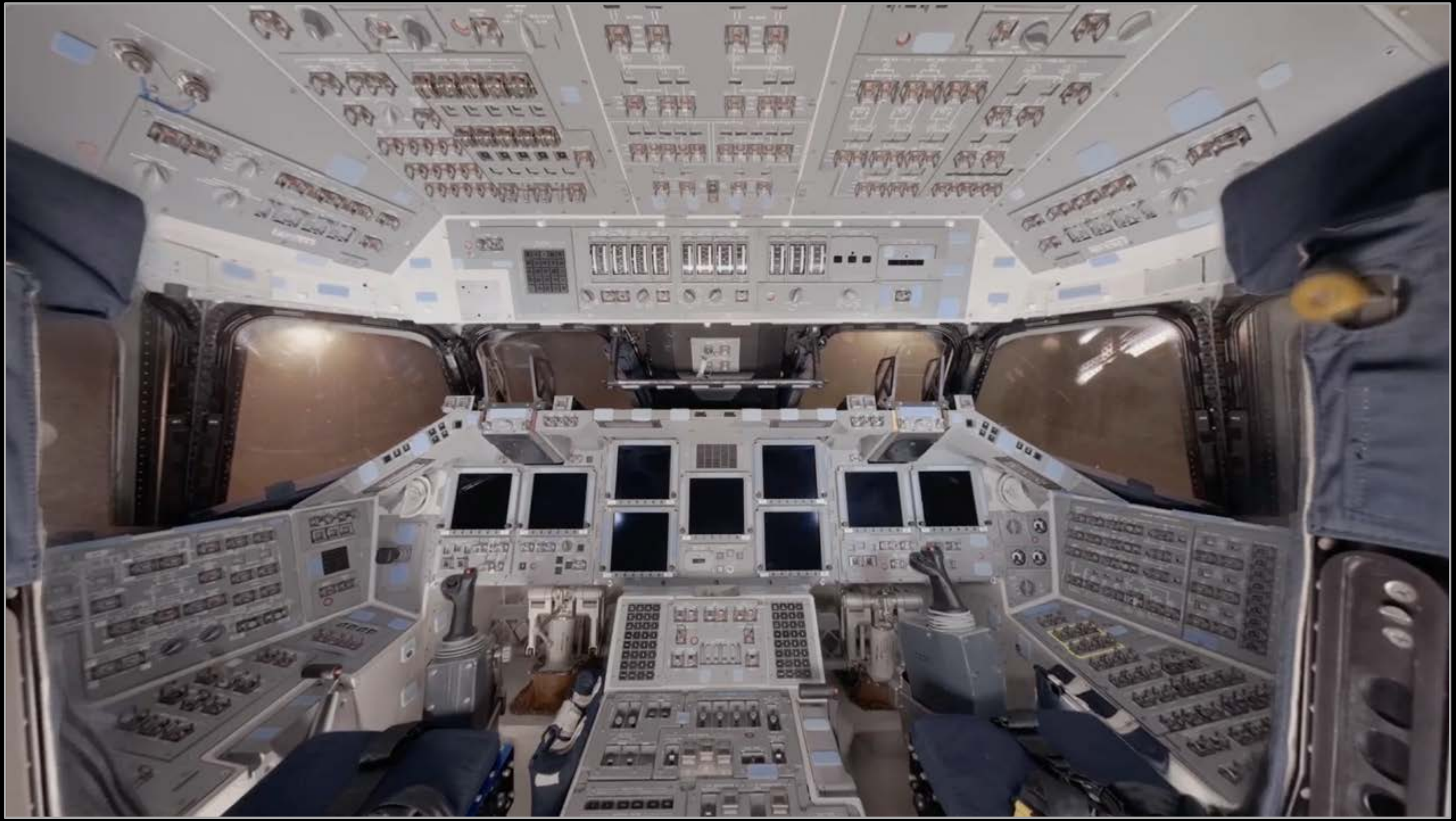
Immersive Photography

- Photogrammetry
  - 6 DOF
- **Stereo Panoramas**
  - Photo-Realism

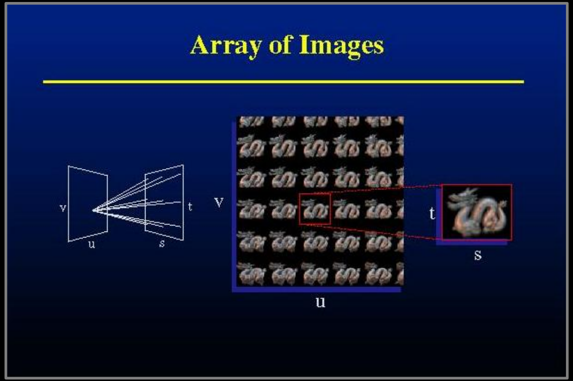


Immersive Photography

- Photogrammetry
  - 6 DOF
- Stereo Panoramas
  - Photo-Realism
- **Light Fields**
  - 6DOF
  - Photo-Realism

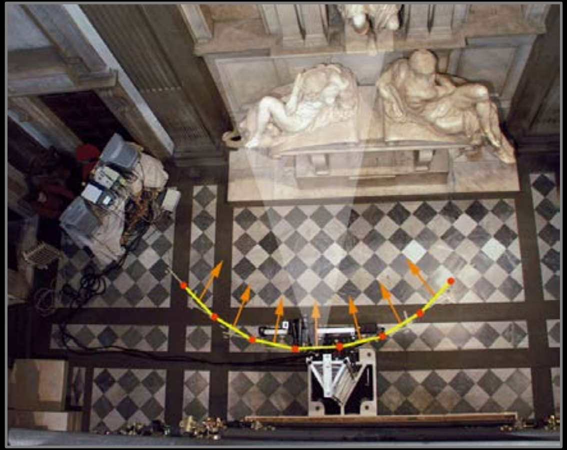


Marc Levoy and Pat Hanrahan. Light Field Rendering. SIGGRAPH 1996



Steven J. Gortler et al. *The Lumigraph.* SIGGRAPH 1996

Levoy et al. The digital Michelangelo project: 3D scanning of large statues. SIGGRAPH 2000. (Light Field of "Night")



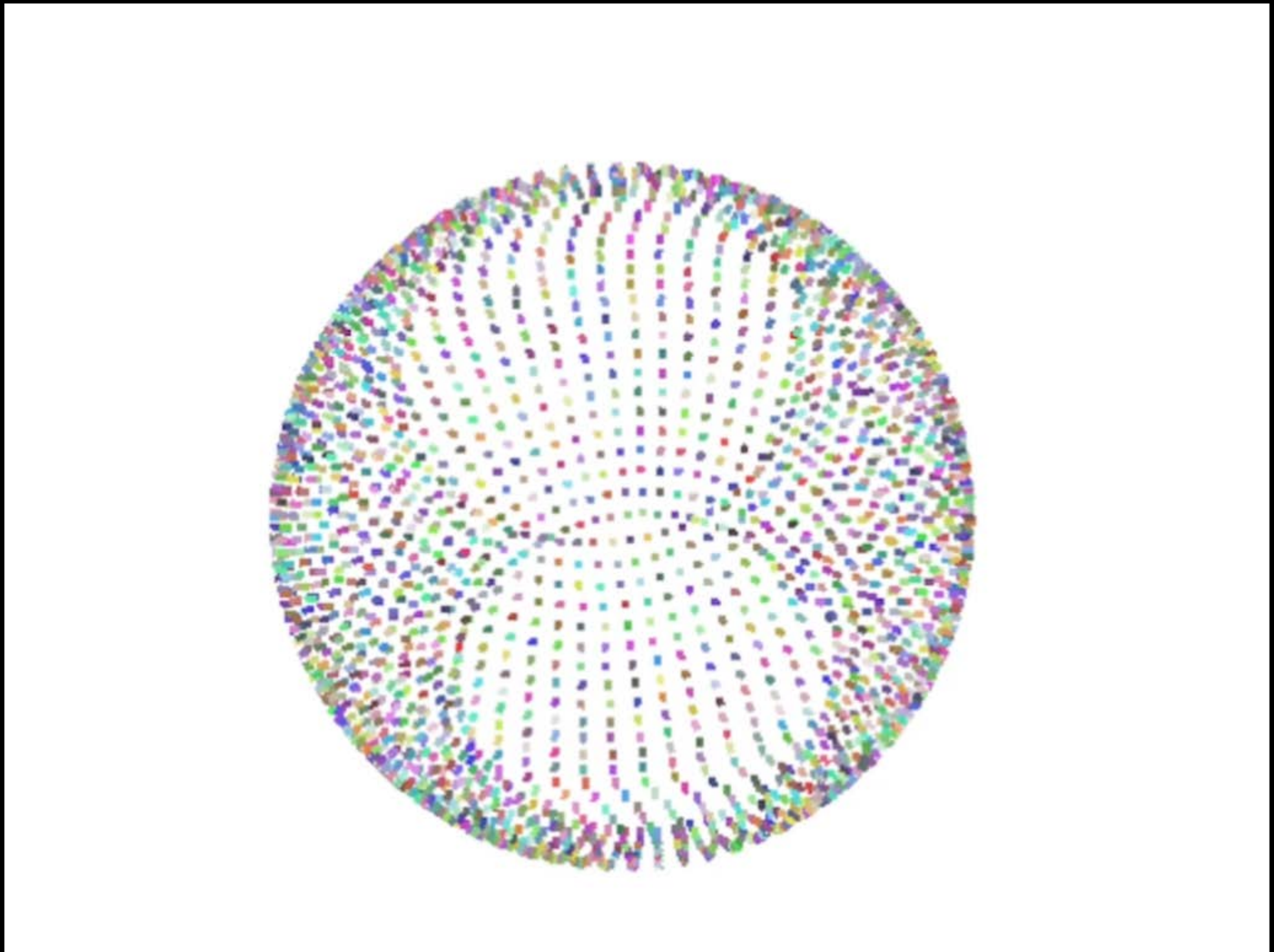
Debevec et al. *Spherical light field environment capture for virtual reality using a motorized pan/tilt head and offset camera.* SIGGRAPH 2015 Posters

T Milliron et al. Hallelu*jah: The World's First Lytro VR Experience.* SIGGRAPH 2017 VR Village

The System

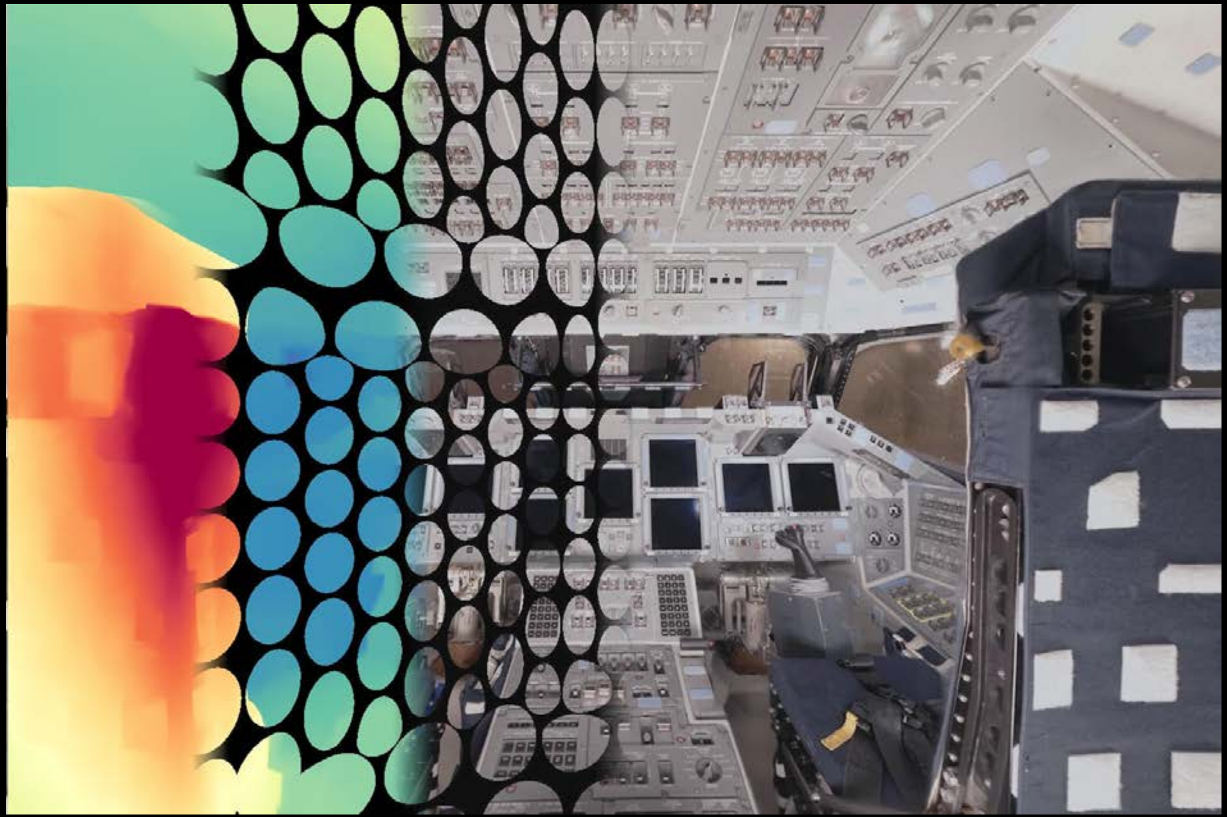
- **Acquire**

- Acquire
- Process
  - Calibrate

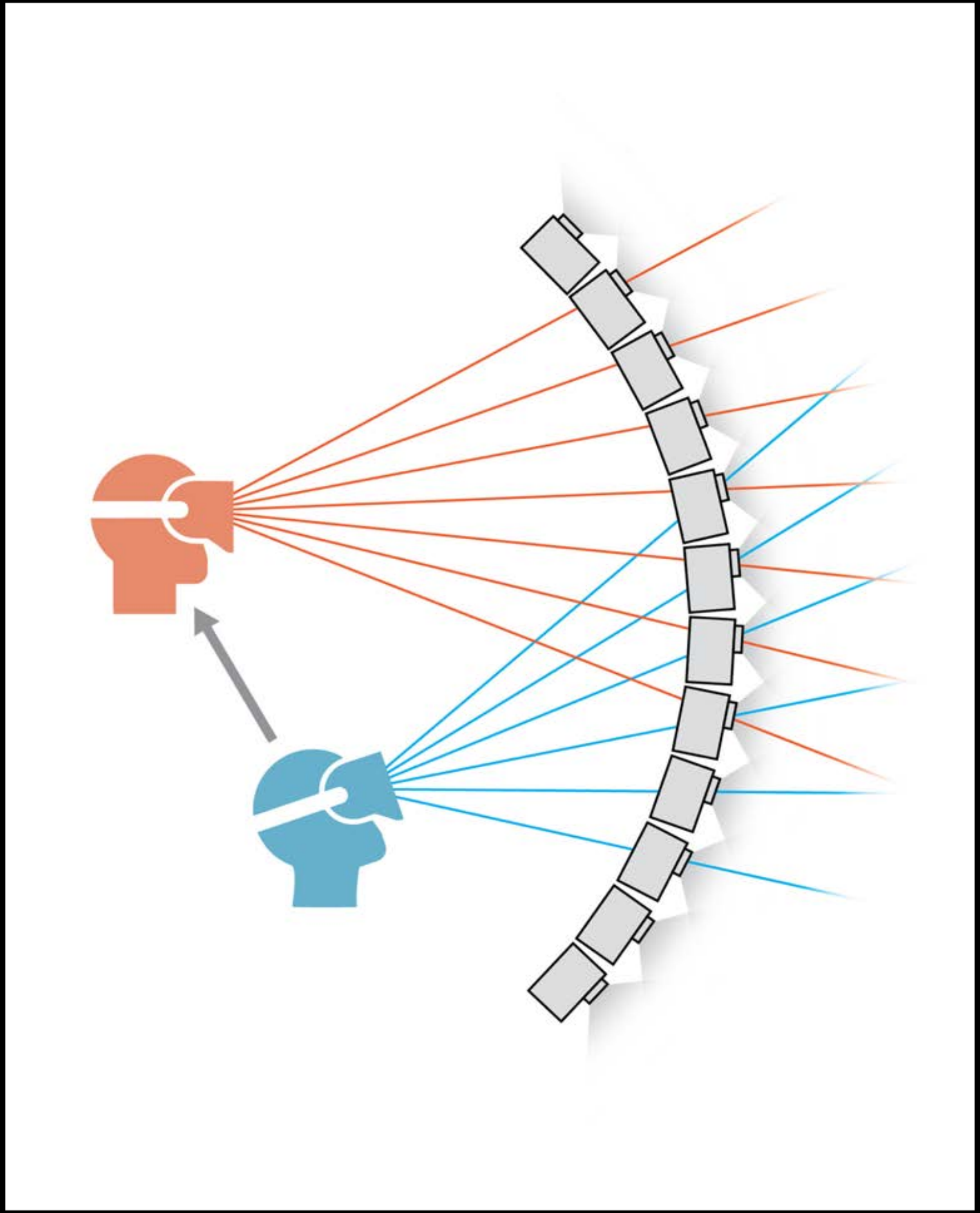


- Acquire
- Process
  - Calibrate
  - Geometry
  - Prefilter
  - Compress

- Acquire
- Process
  - Calibrate
  - Geometry
  - Prefilter
  - Compress
- Render

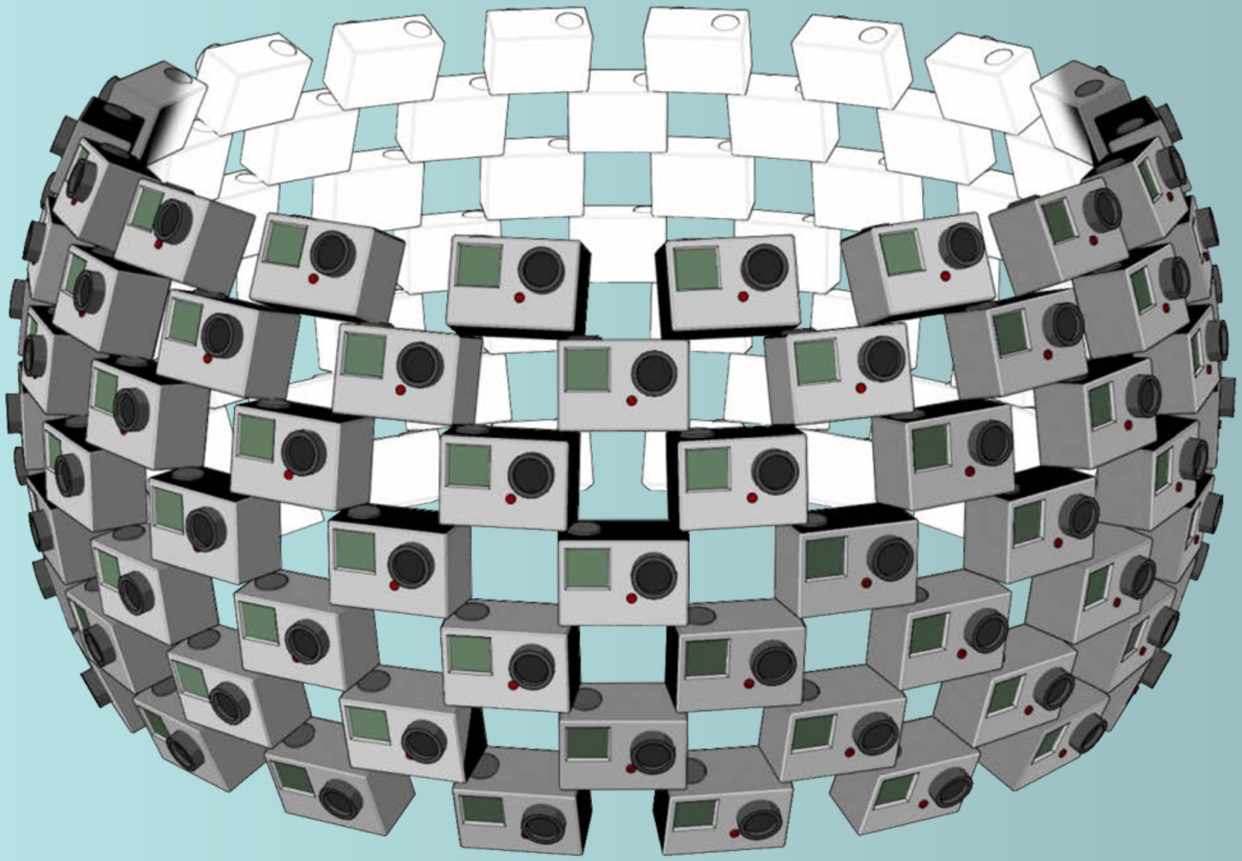


- **Acquire**
- Render
- Prefilter
- Compress



Light Field

- Acquire sparse images on surface.
- Render views inside surface.

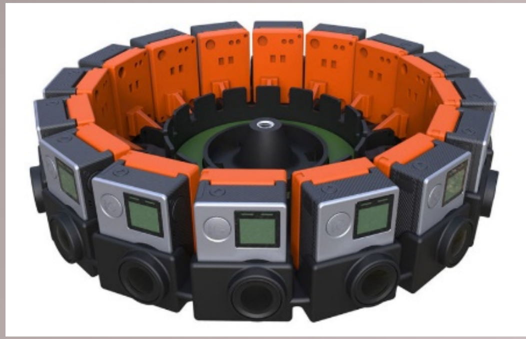


18 x 7 fisheye cylindrical/spherical camera array

(>1Gpixel/frame at 4K each camera)

light field video playback with panoramic stereo and full parallax

(from Stanford SCIEN Workshop on Light Field Imaging, 2/12/2015)



JUMP Odyssey 360 Stereo Camera

Jump: Virtual Reality Video
Anderson et al, SIGGRAPH Asia 2016

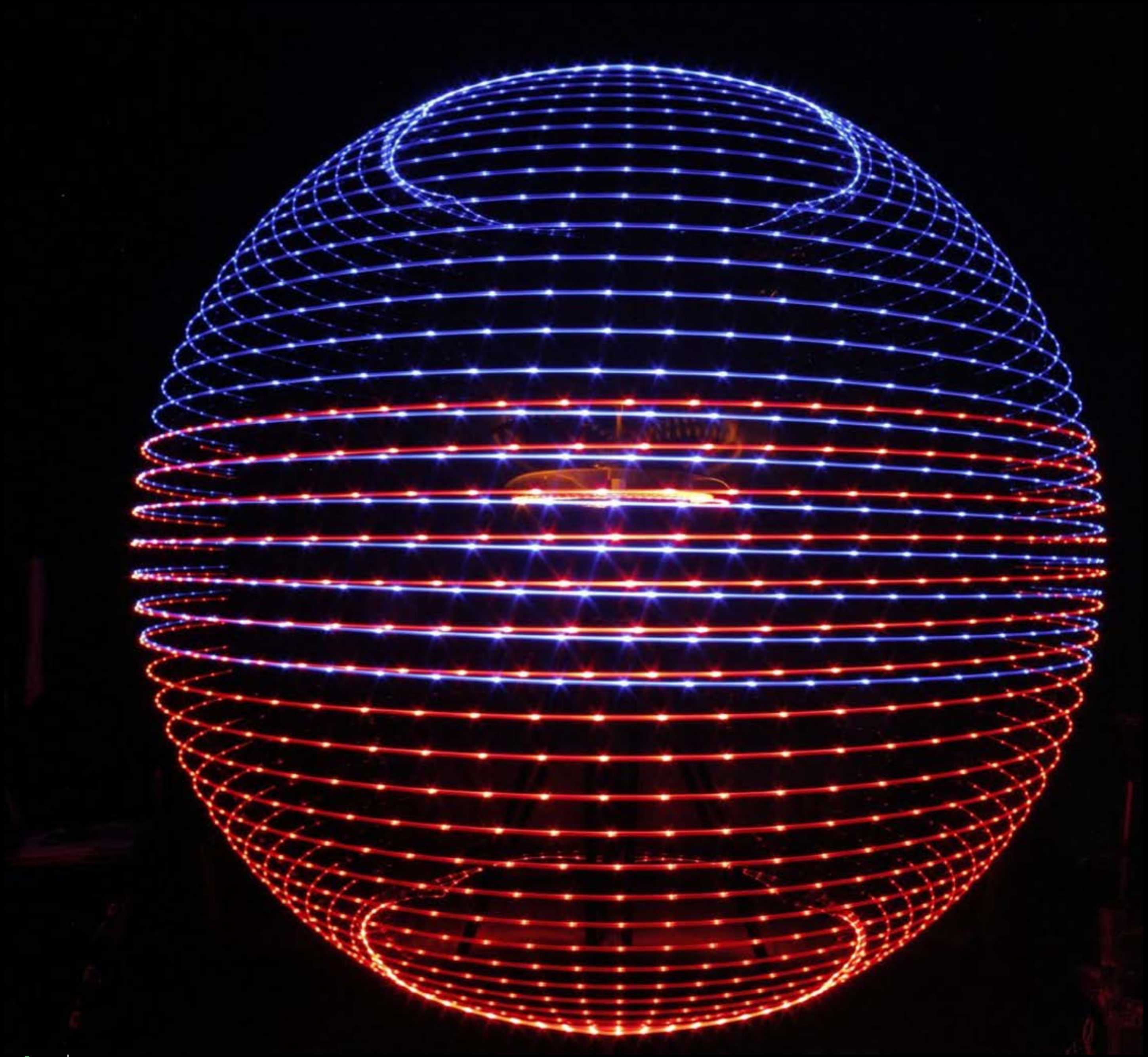

16xGoPro rig (5x speed)


HERO4

Light Field Portraiture

- GoProx16 rig


2X DSLR




2xDSLR rig (5x speed)


cam 1 HDR images


cam 2 HDR images

30-90 seconds
People + Outdoors

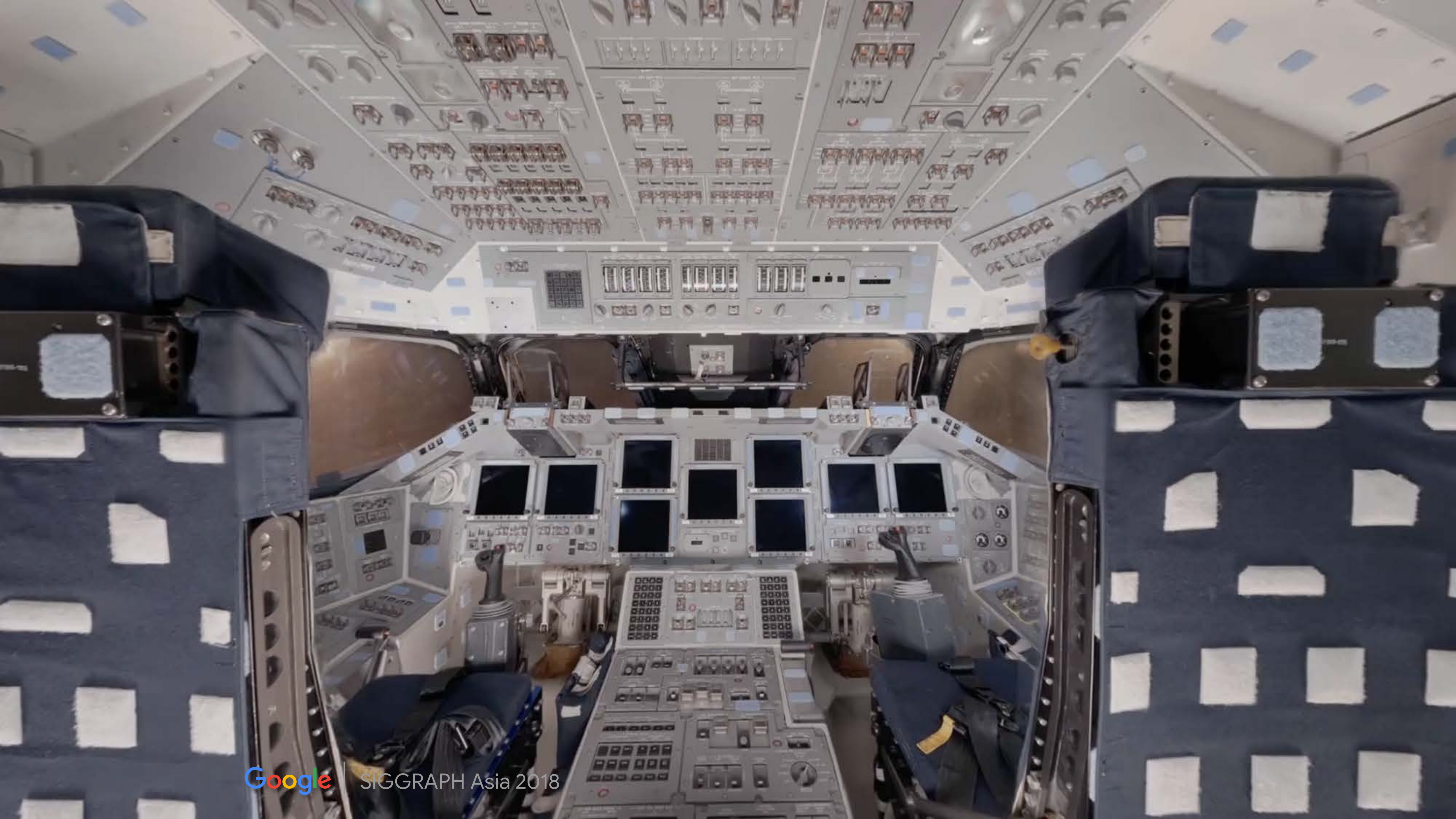
10-40 minutes
Highest Quality Pixels

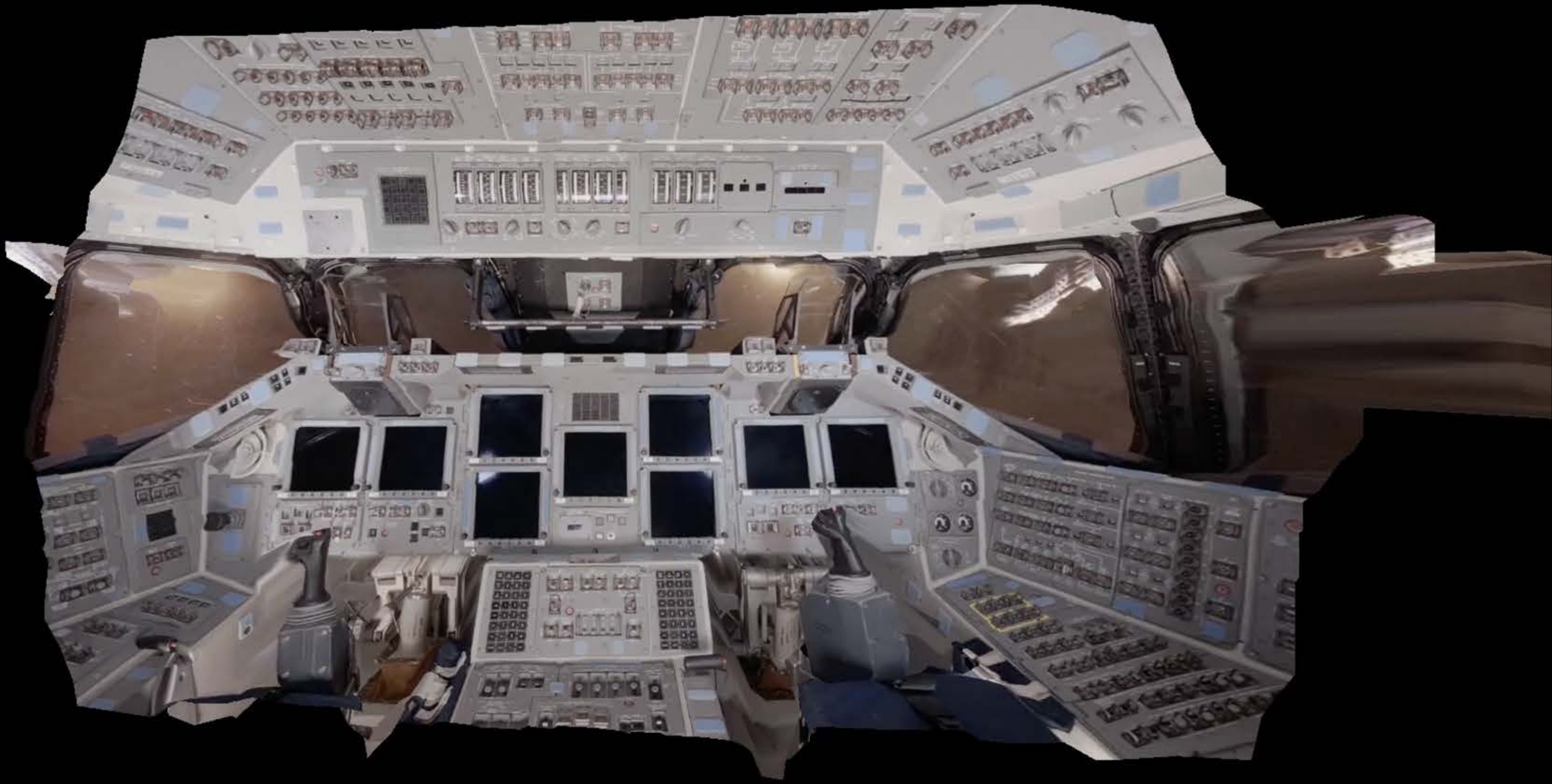
- Acquire
- **Render**
- Prefilter
- Compress

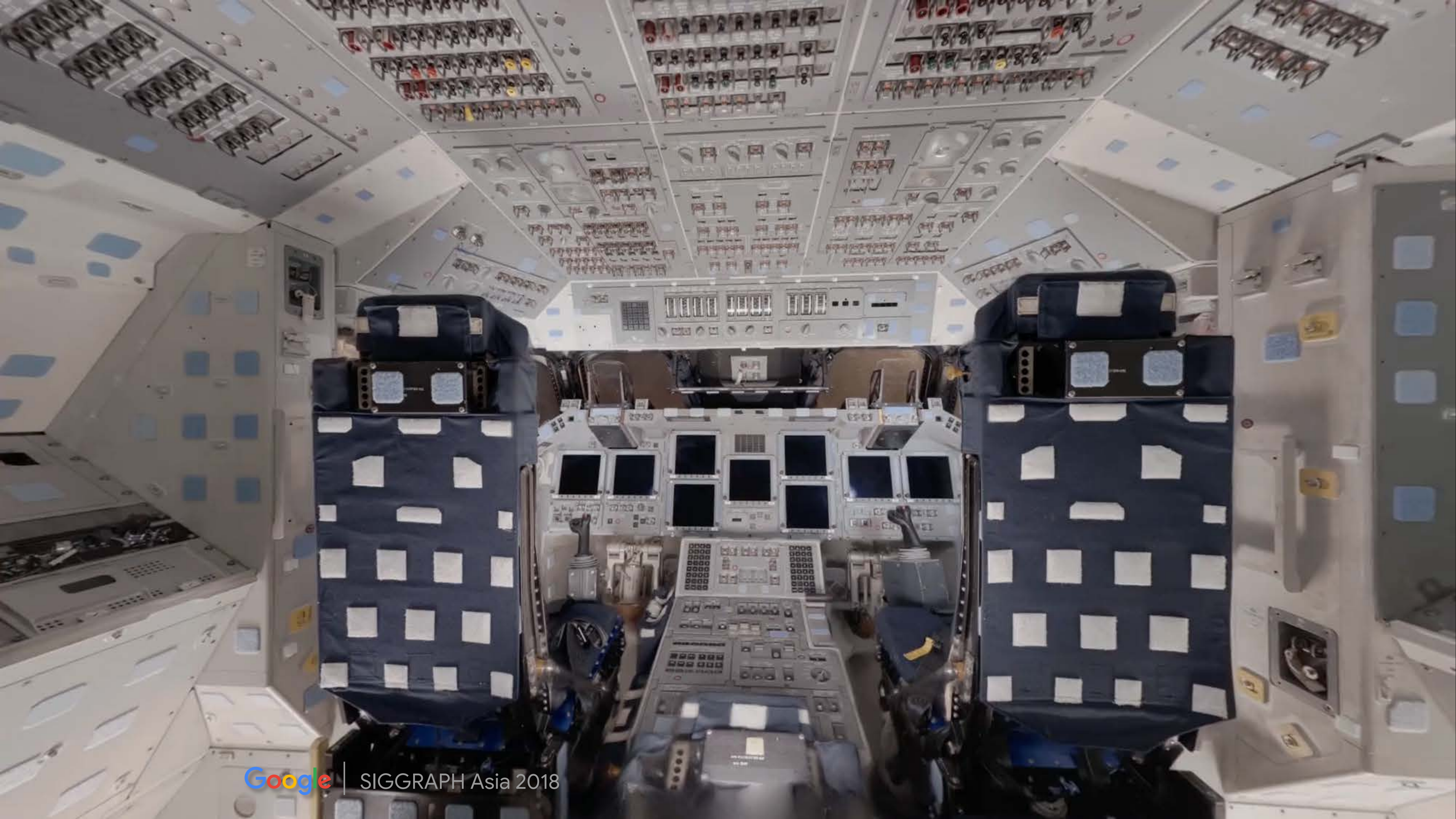
Quality

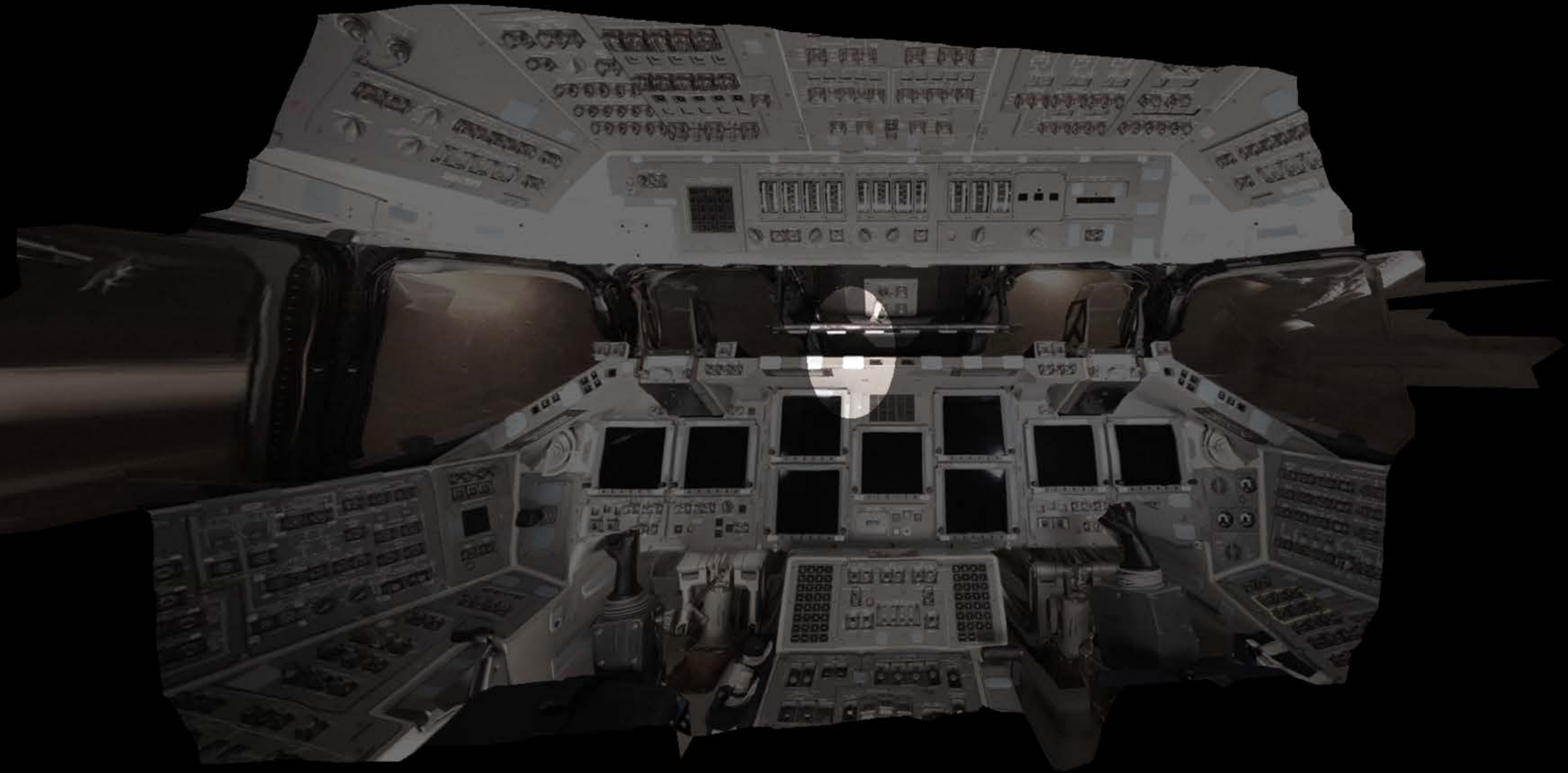
Speed

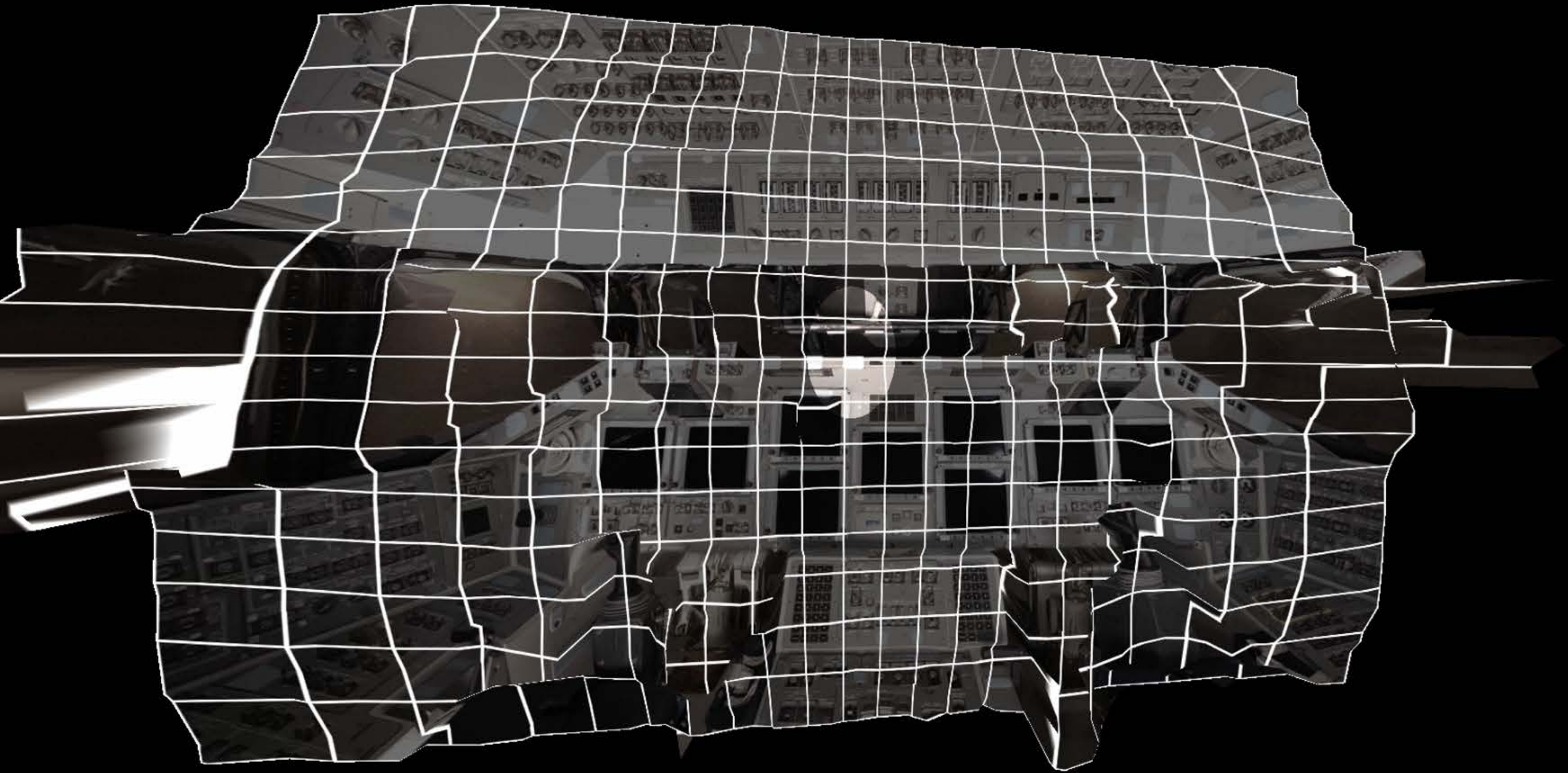
Stereo views @90 Hz

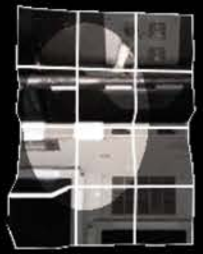




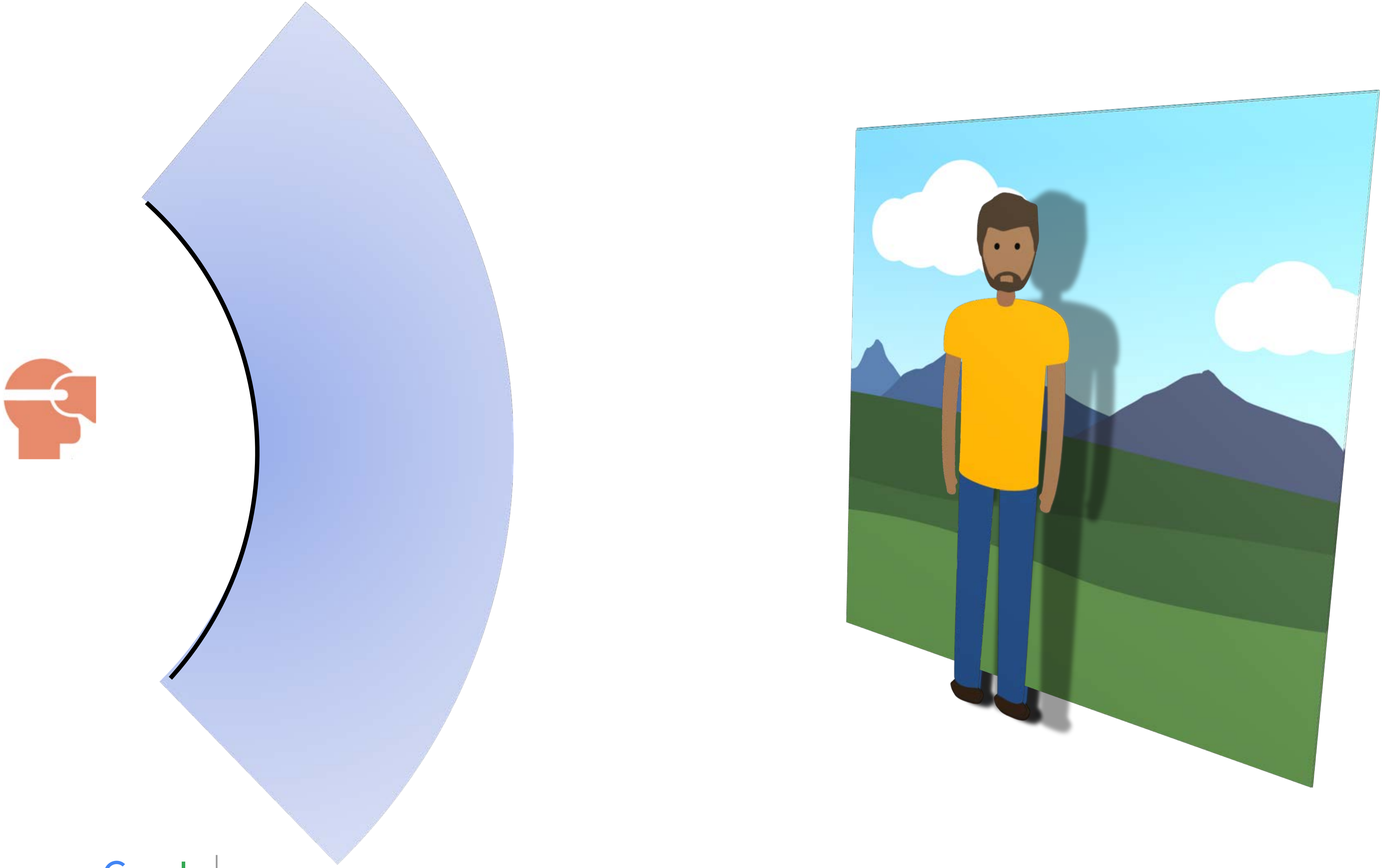


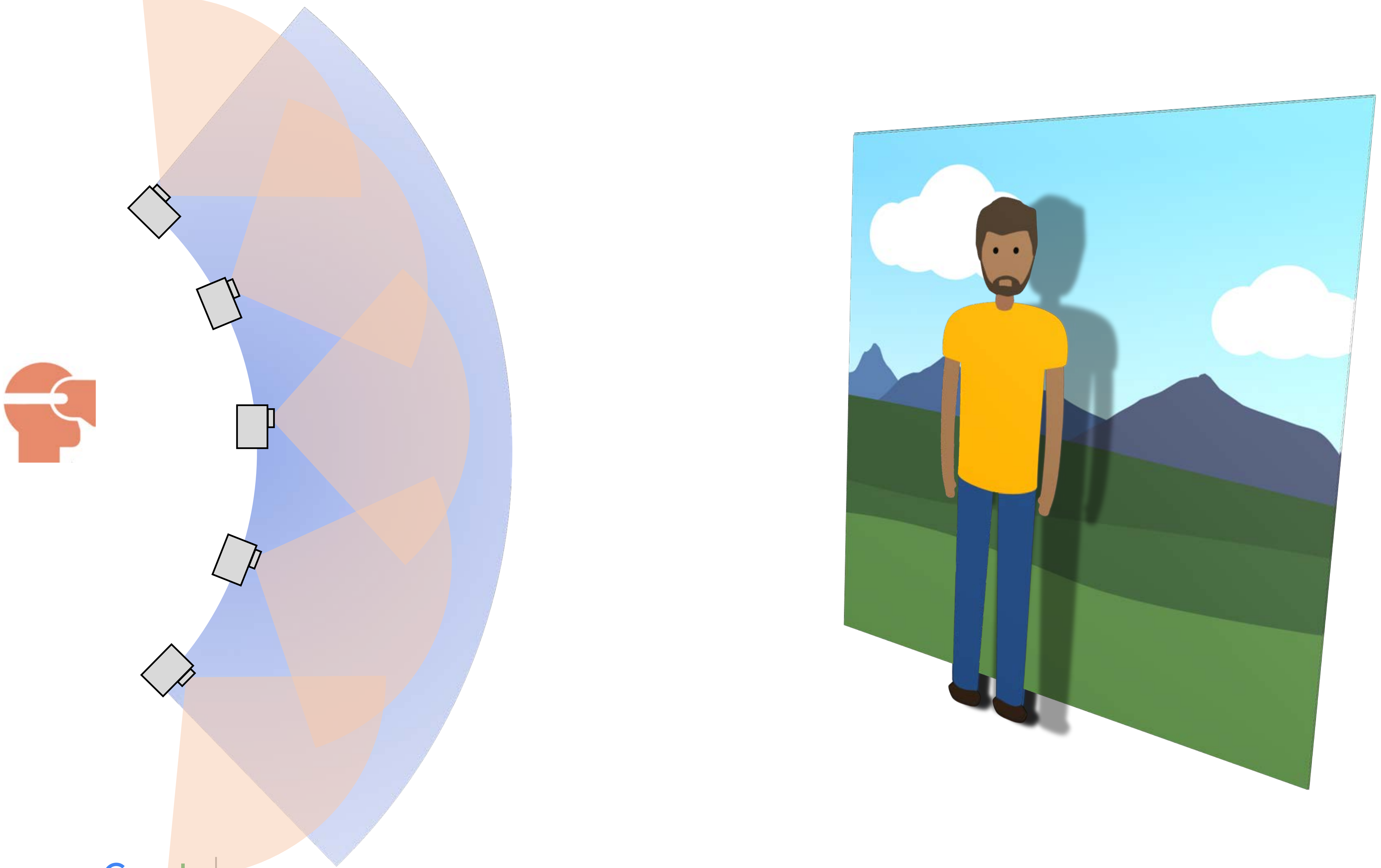


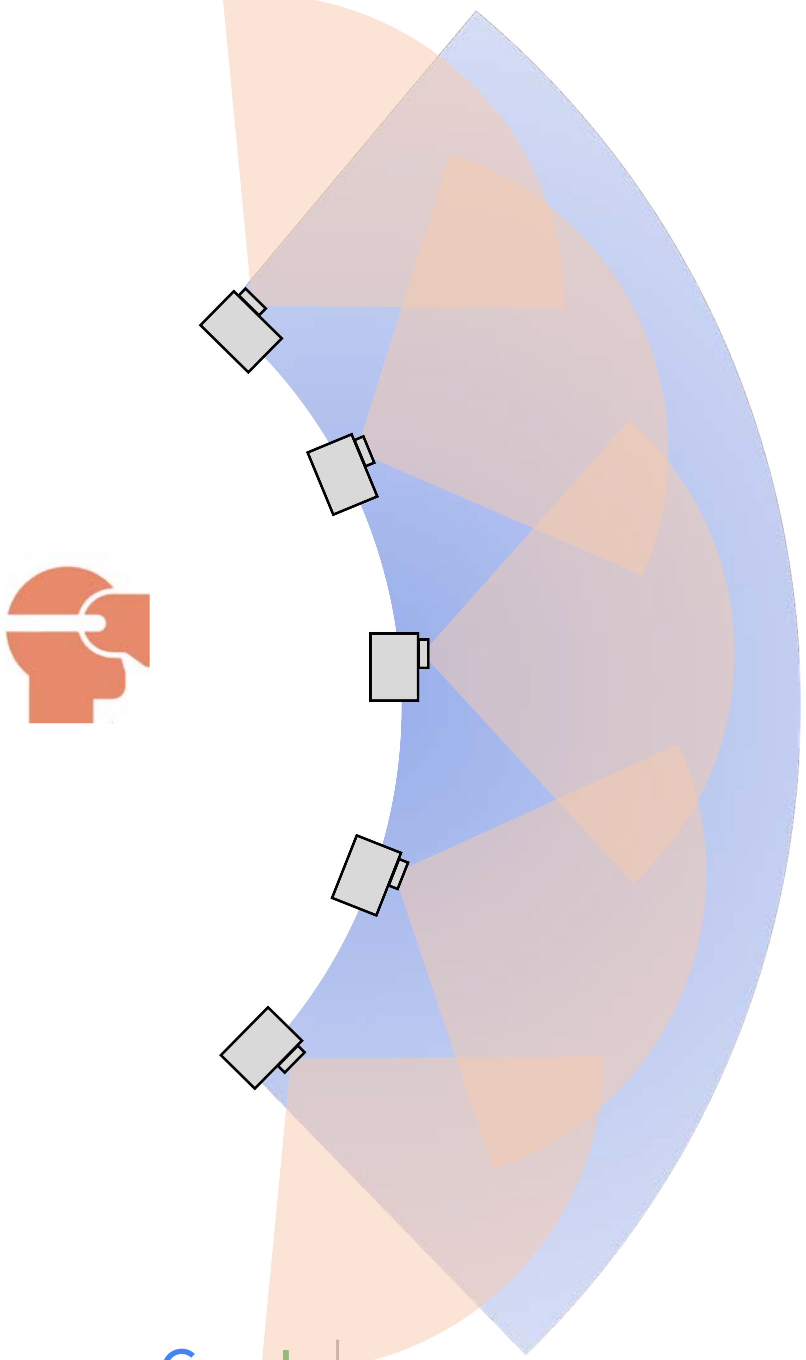


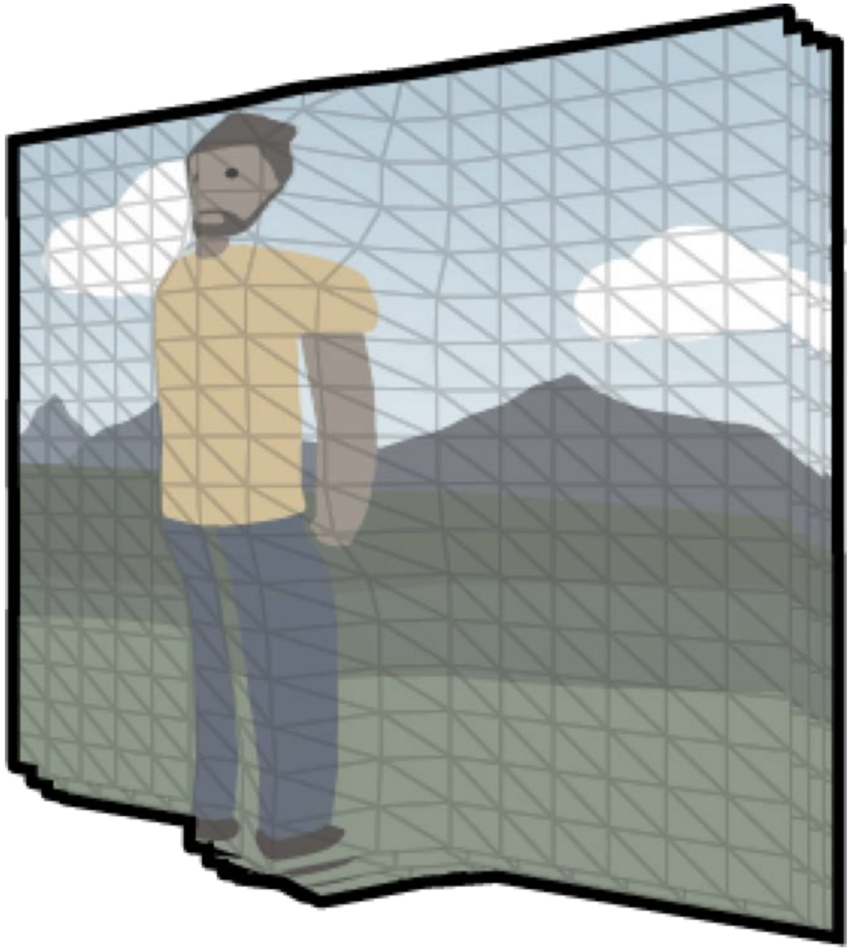


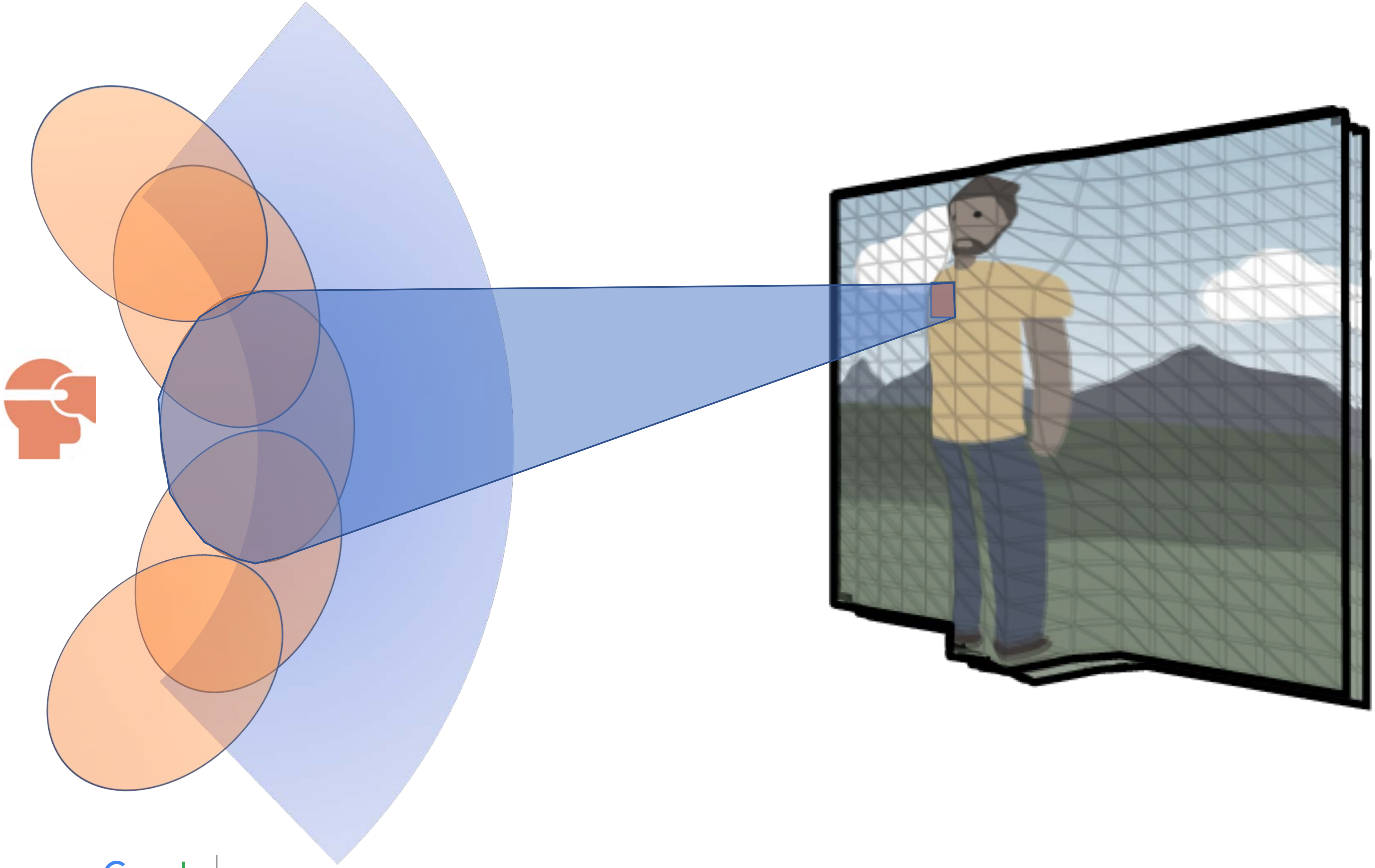
- Acquire
- Render
- **Prefilter**
- Compress

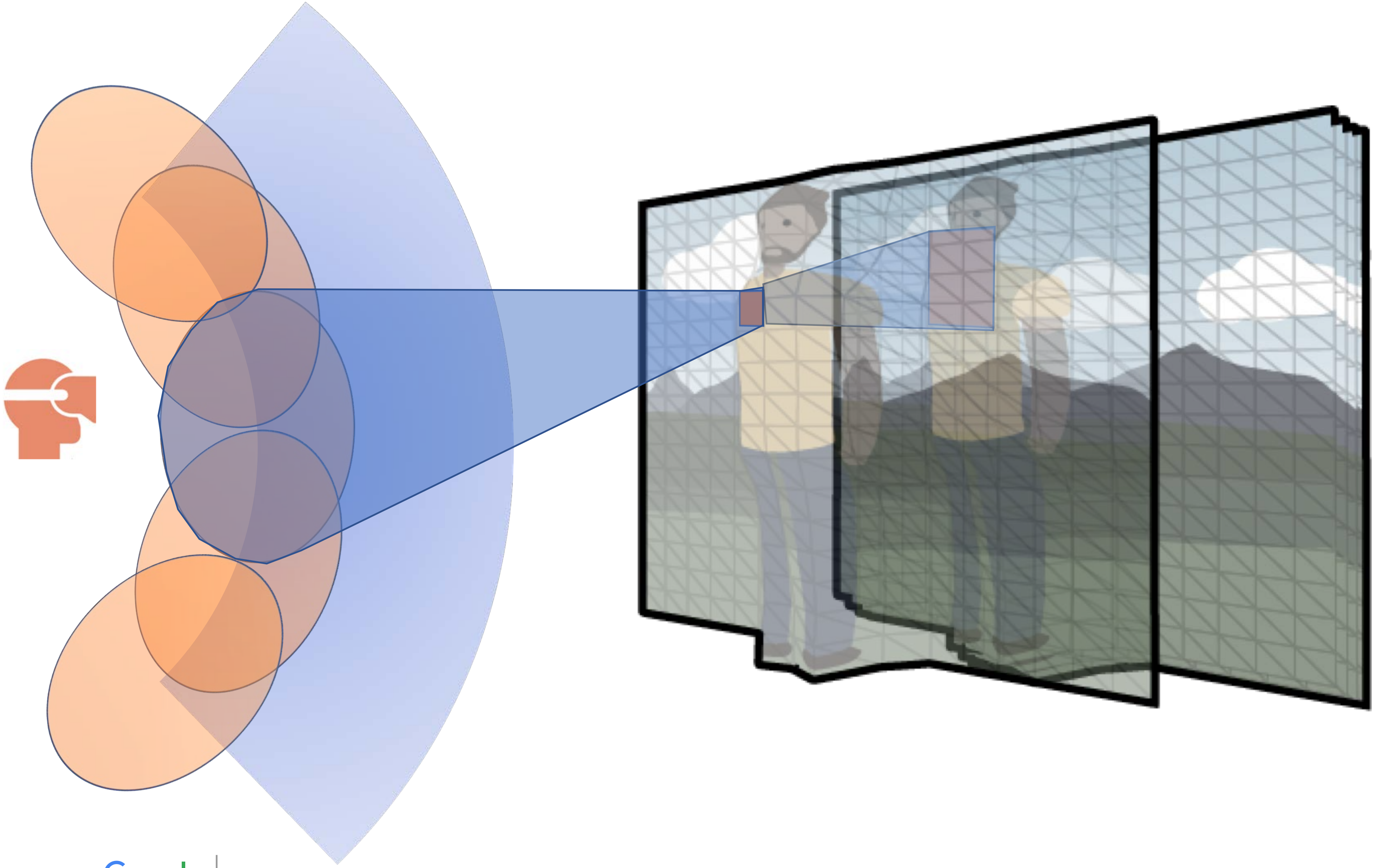













Without Prefilter
Google
SIGGRAPH 2019


With Prefilter

Low Res Geometry Without Prefilter


Low Res Geometry
With Prefilter
Google | SIGGRAPH 2019


With Prefilter

- Acquire
- Render
- Prefilter
- **Compress**

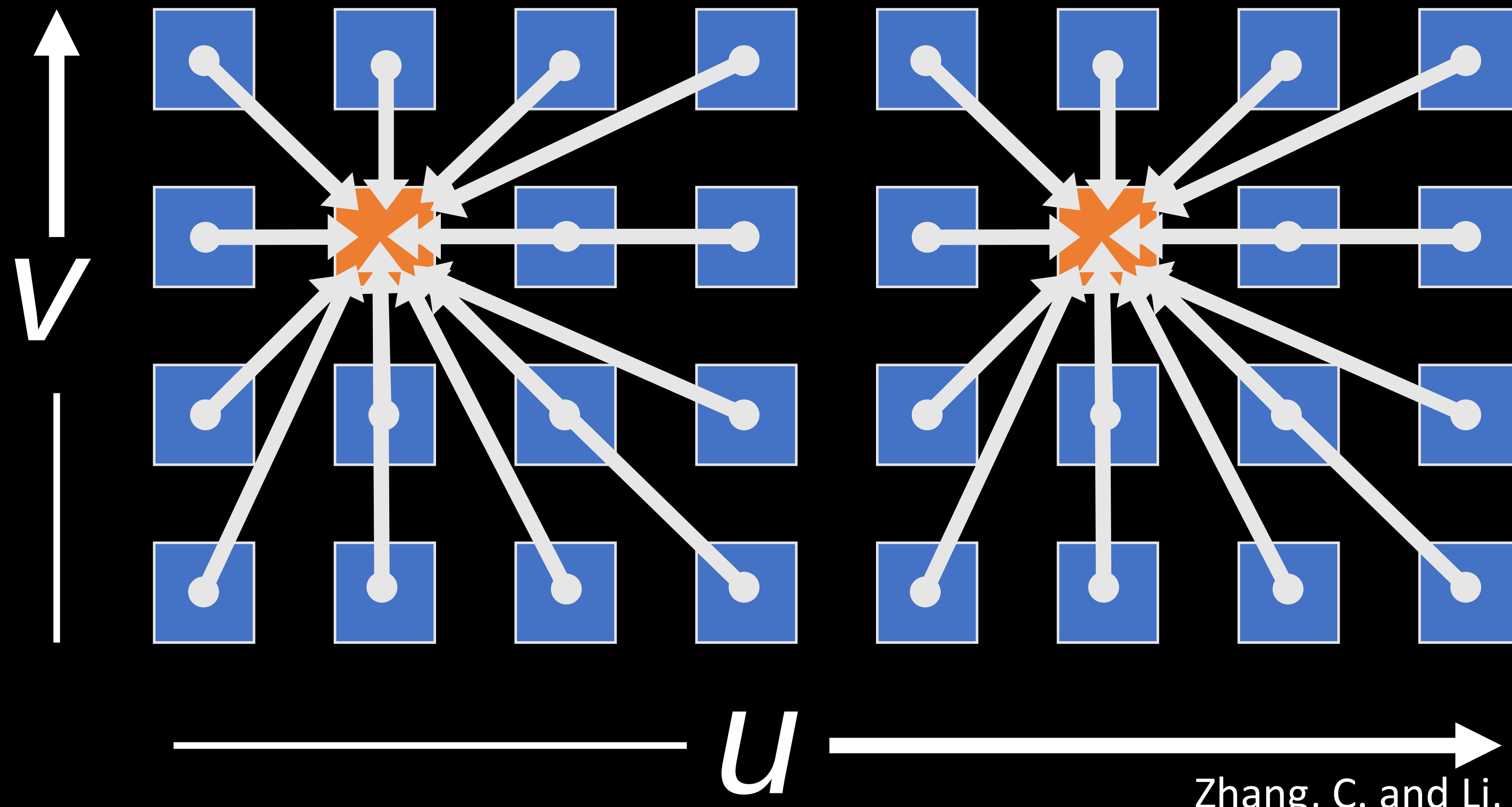
~1000-2000 Images X 1280x1024 X RGB

= ~4-8GB Uncompressed

IDEA: Use Video Compression

We need random access to tiles

Motion Compensated Prediction (MCP) for Light Fields



Zhang, C. and Li, J., 2000. Compression of lumigraph with multiple reference frame (MRF) prediction and just-in-time rendering.

VP9

+

1. Random access to tiles
2. Unlimited reference images

AV1

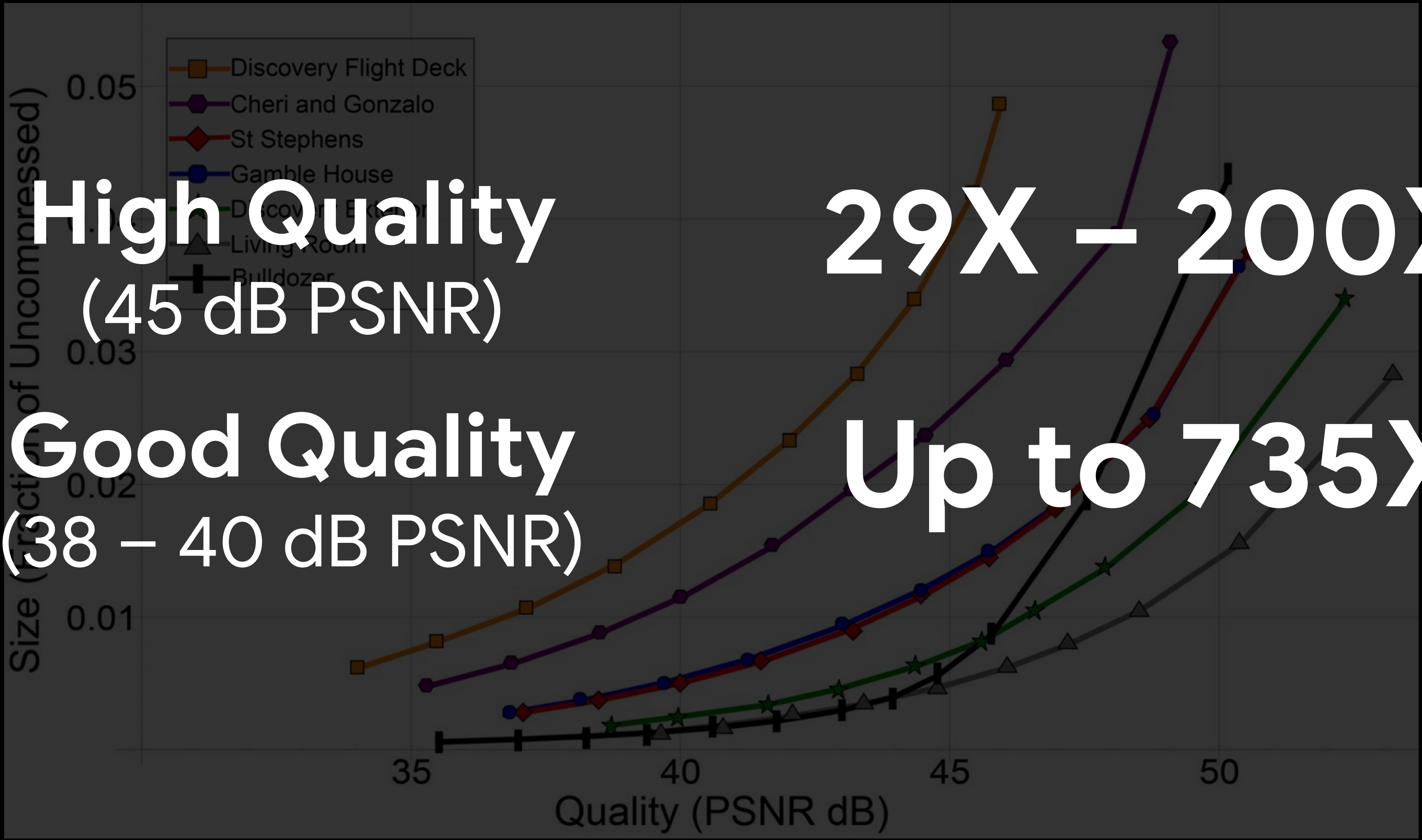
"Large Scale Tile Decoding"

High Quality
(45 dB PSNR)

29X – 200X*

Good Quality
(38 – 40 dB PSNR)

Up to 735X*

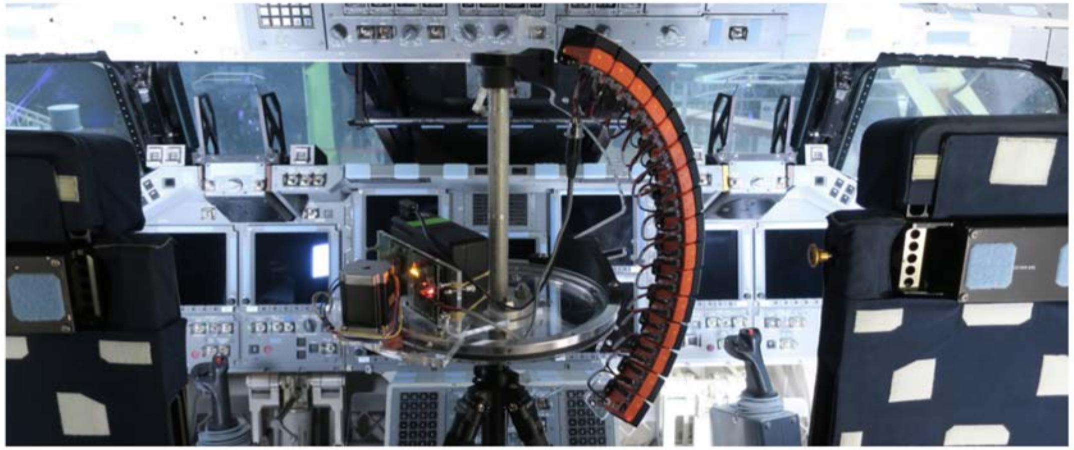


* Results on our datasets which are sparser than most light field datasets


Welcome to Light Fields

GOOGLE AR AND VR

Experimenting with Light Fields



Connectivity

VR is still a novelty, but Google's light-field technology could make it serious art

MIT Technology Review

A new VR app lets you explore worlds with surprising depth and detail.

by Rachel Metz March 14, 2018

STEAM STORE COMMUNITY ABOUT SUPPORT

Your Store Games Software Hardware Videos News

All Games > Casual Games > Welcome to Light Fields

Welcome to Light Fields


Saint Stephen’s Church


Airstream Selfie

Light Field Video

Light Field Video

Sparser Cameras

1000s -> 10s

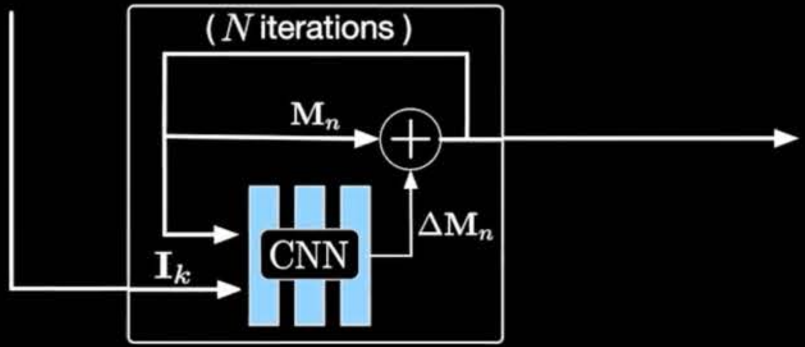
Need better view synthesis

CVPR LONG BEACH CALIFORNIA June 16-20, 2019

DeepView: View Synthesis with Learned Gradient Descent

John Flynn, Michael Broxton, Paul Debevec, Matthew Duvall, Graham Fyffe, Ryan Overbeck, Noah Snavely, Richard Tucker

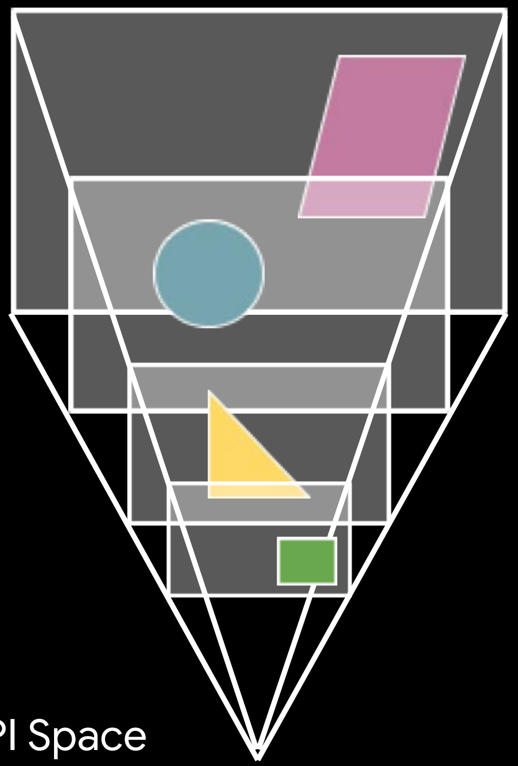
A sparse set of input images from different viewpoints



Learned gradient descent

Network generated multi-plane image

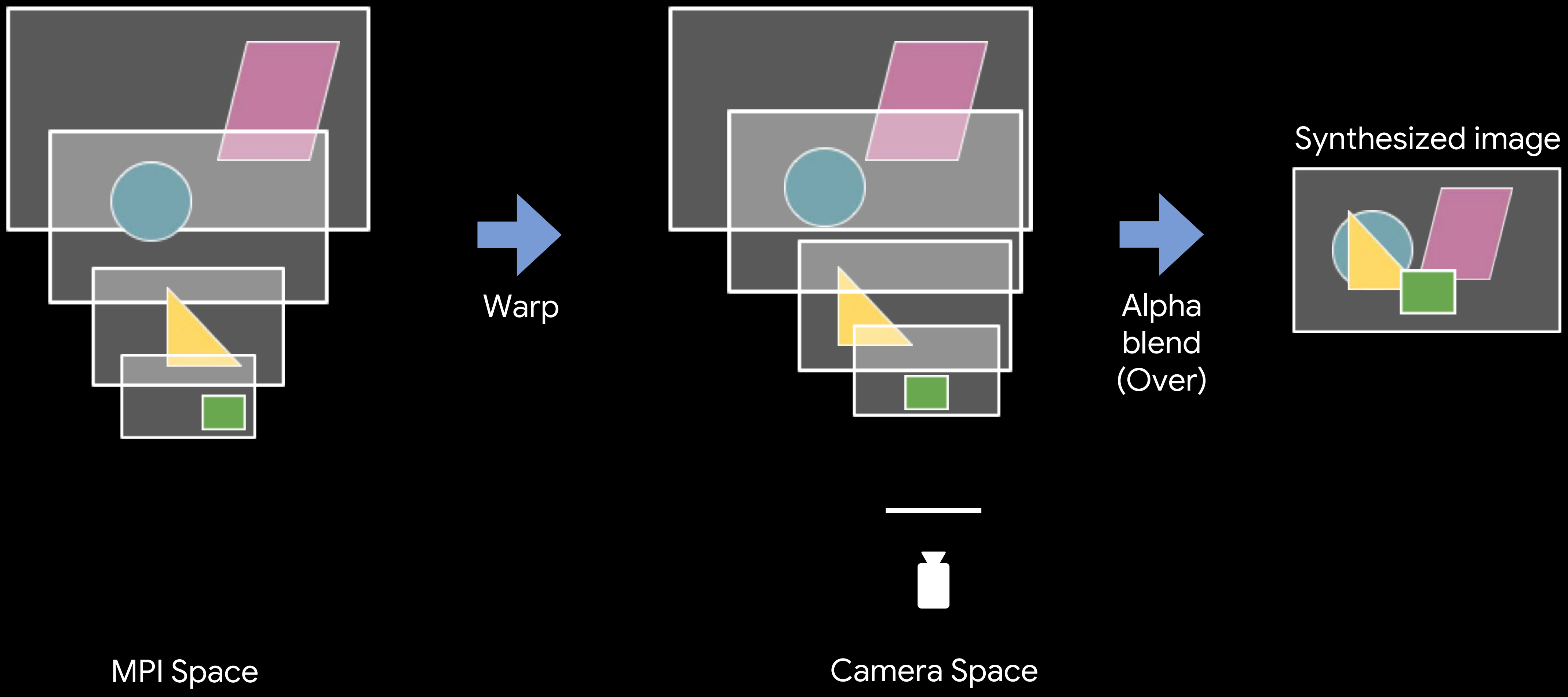
Multiplane Image



Each plane is at a fixed depth and encoded by an RGBA image

Zhou et al. *Stereo magnification: Learning view synthesis using multiplane images*, SIGGRAPH 2018.

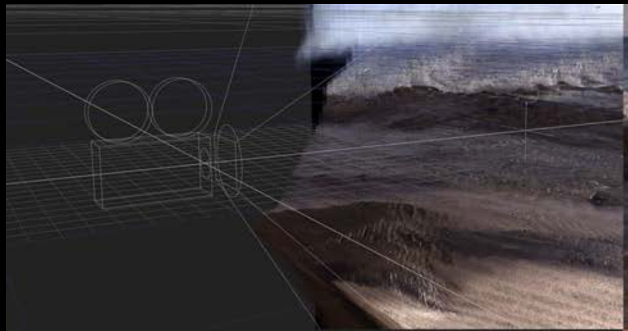
Multiplane Image




EXIT

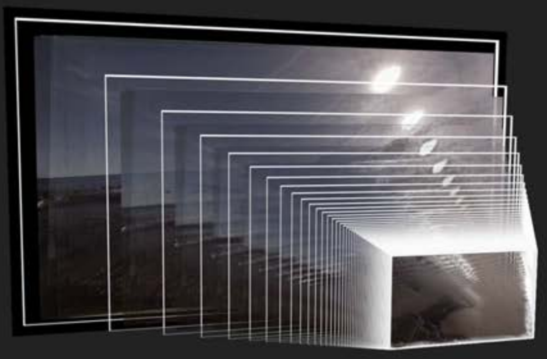
thrive
SIGGRAPH2019
LOS ANGELES • 28 JULY – 1 AUGUST

POSTERS



Compositing Light Field Video Using Multiplane Images

Matthew DuVall, John Flynn, Michael Broxton, Paul Debevec
Google, Inc.



VR@50 Light Fields SIGGRAPH 2018

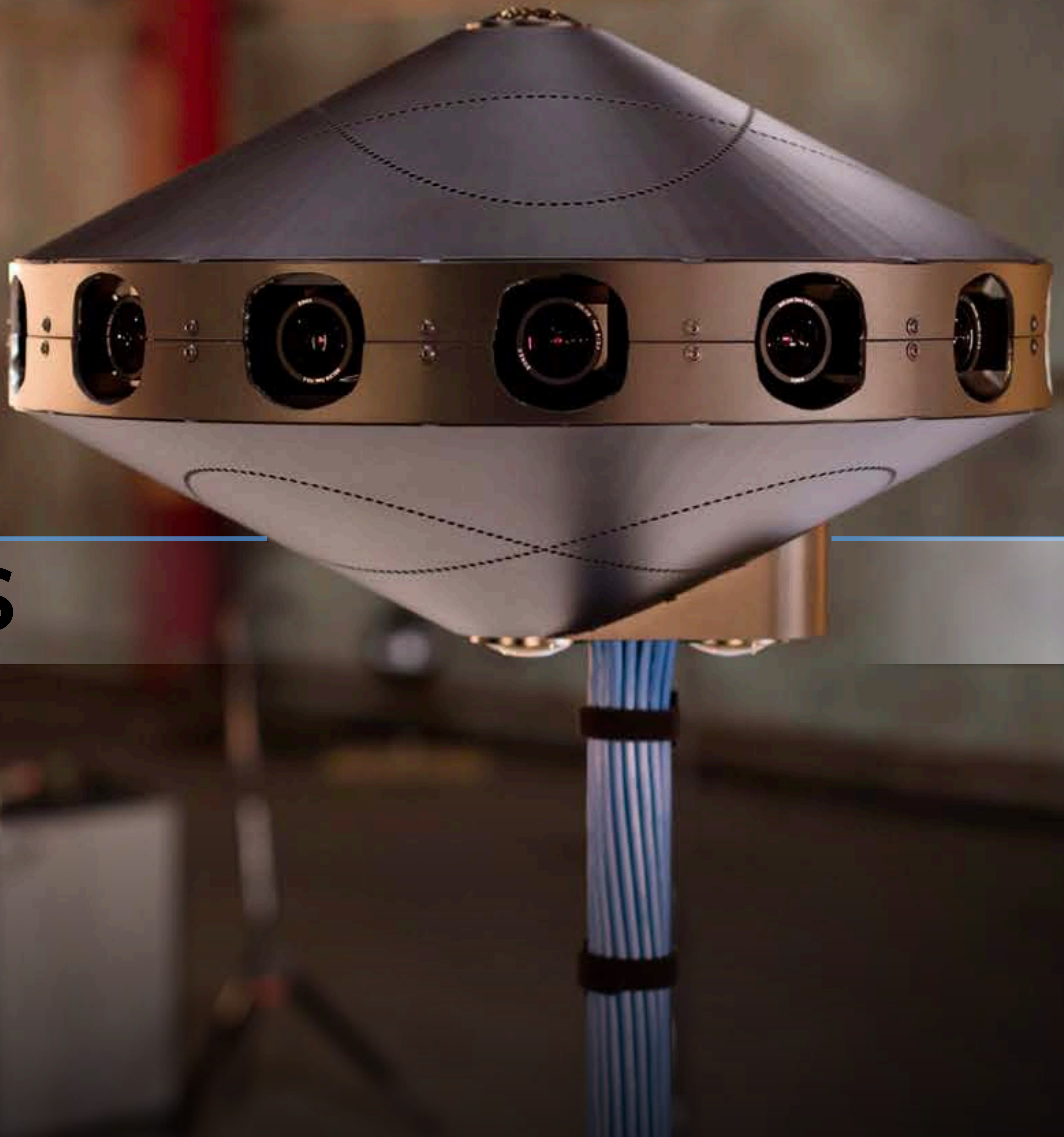
VR@50 Light Fields at
SIGGRAPH 2018
Download at: https://augmentedperception.github.io/deepview/

Thank you!

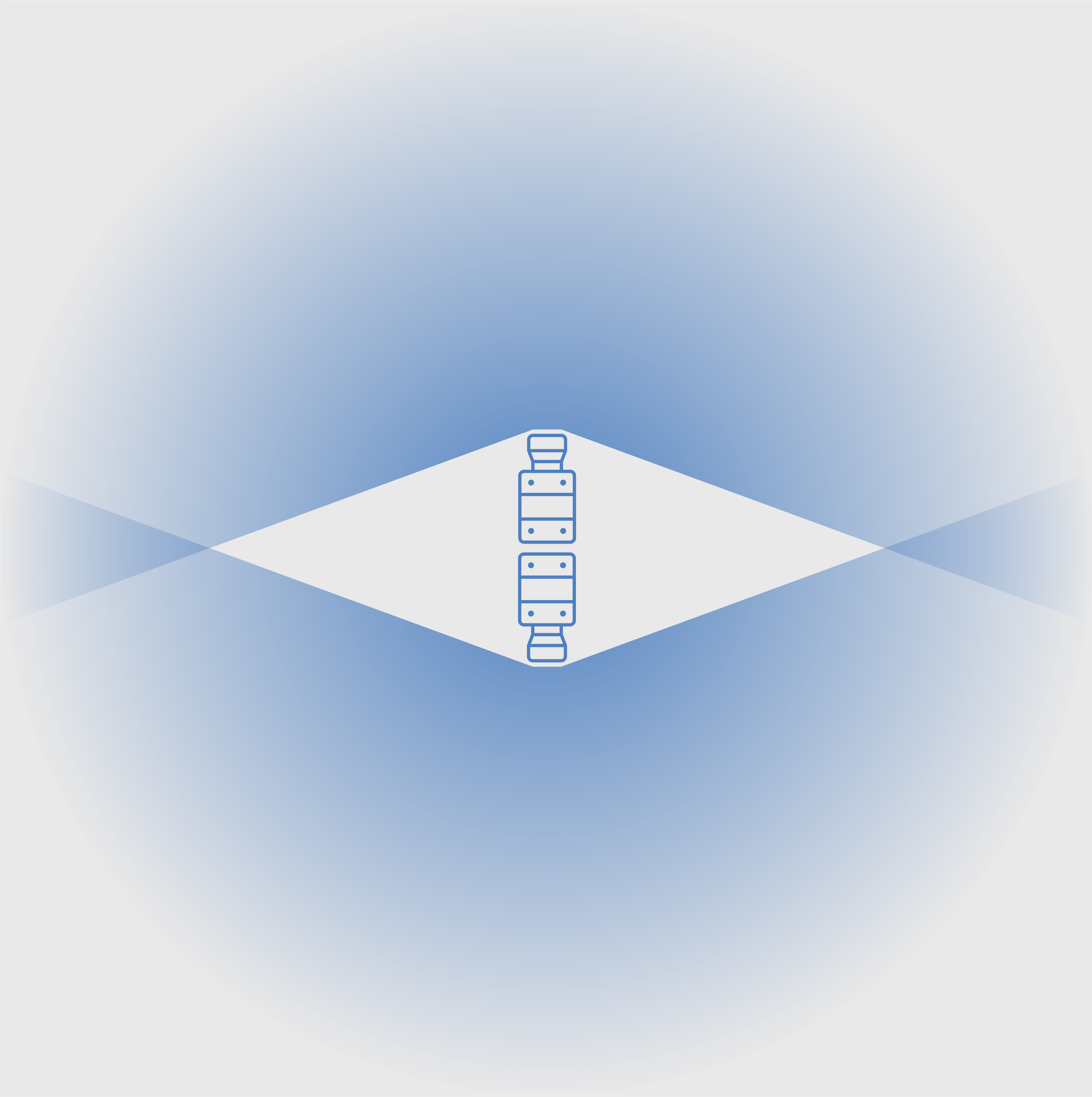
360 and ODS

VIDEO CAPTURE SYSTEMS

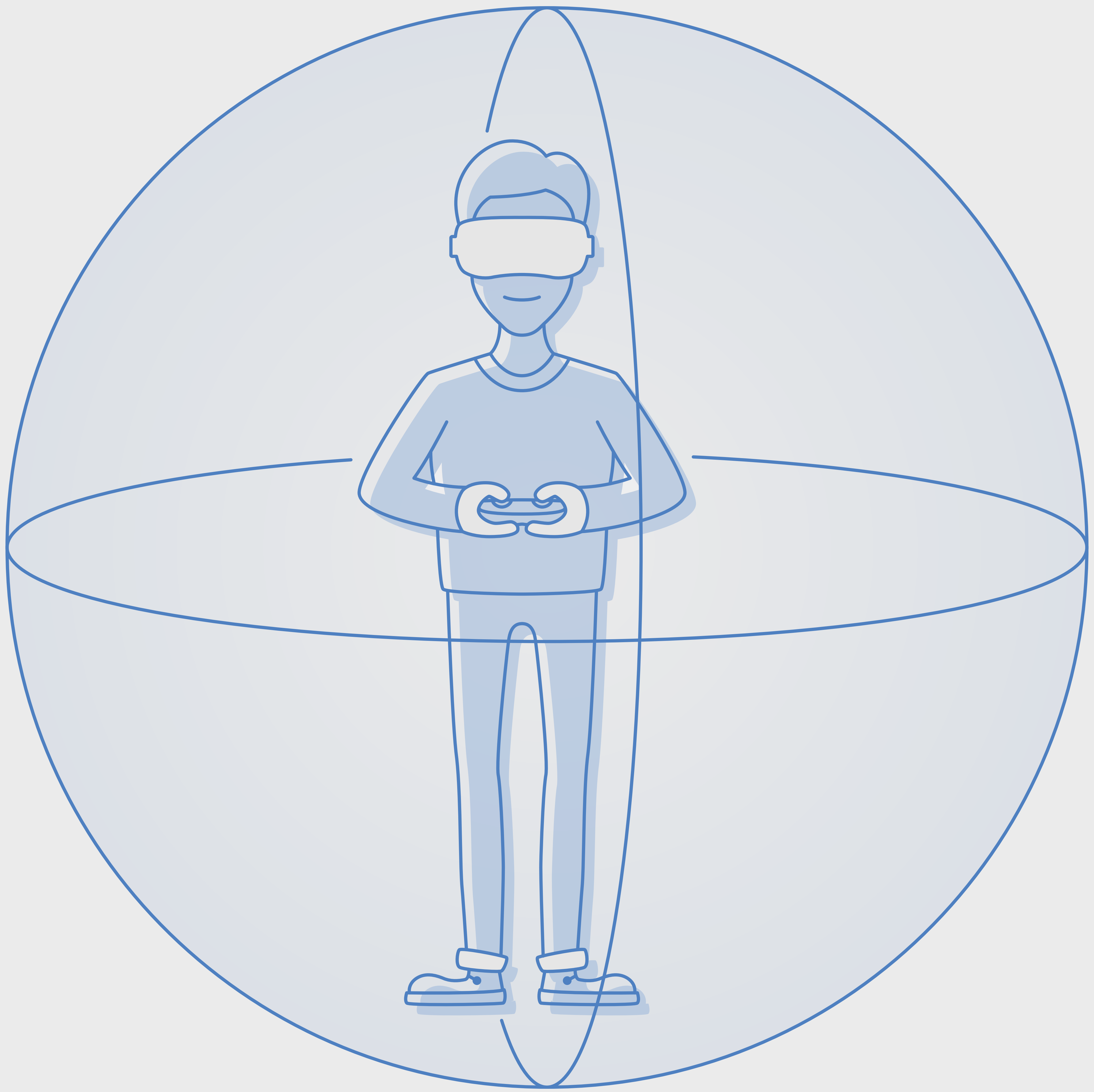
Brian Cabral facebook



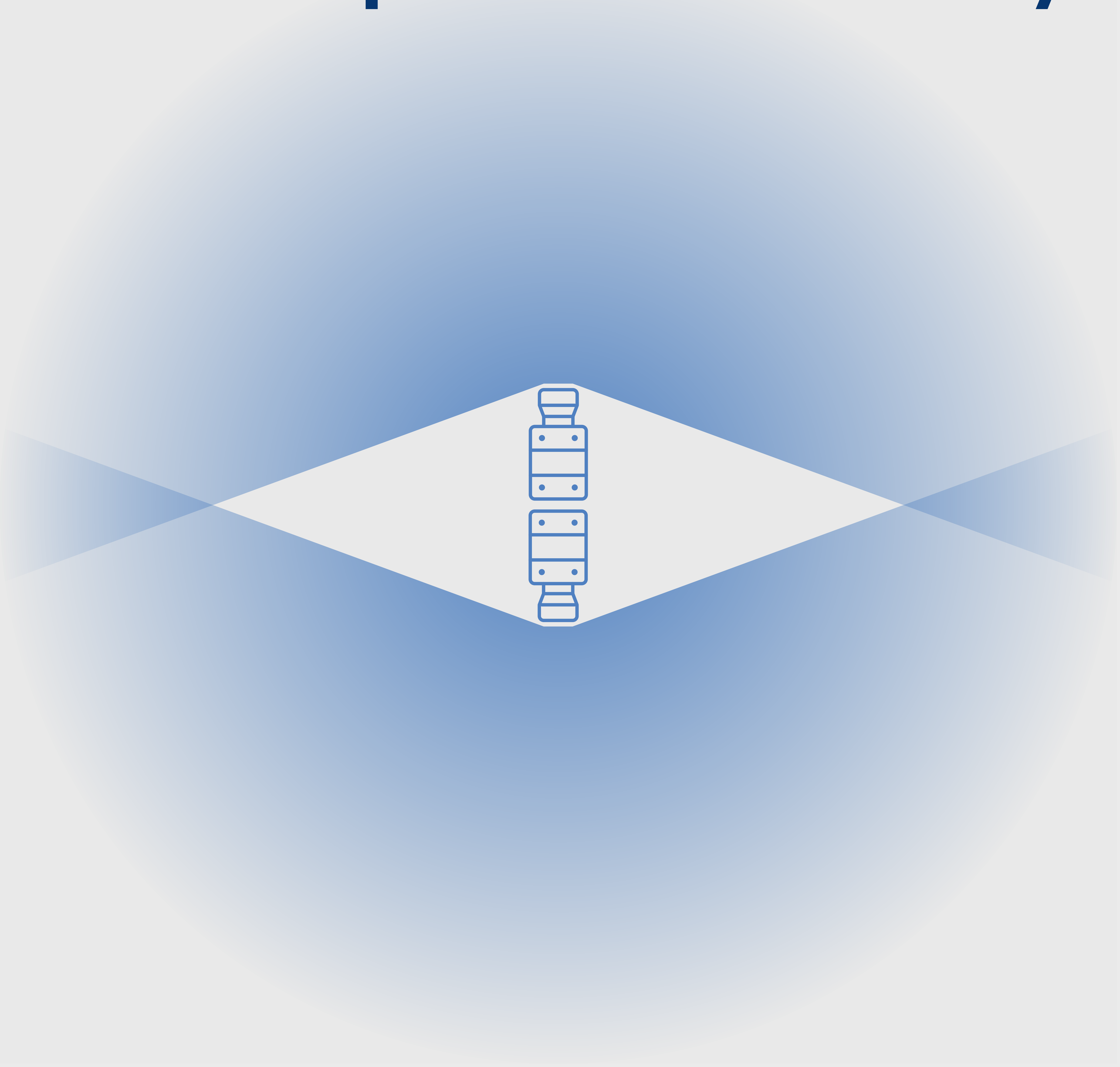
Monoscopic 360° capture



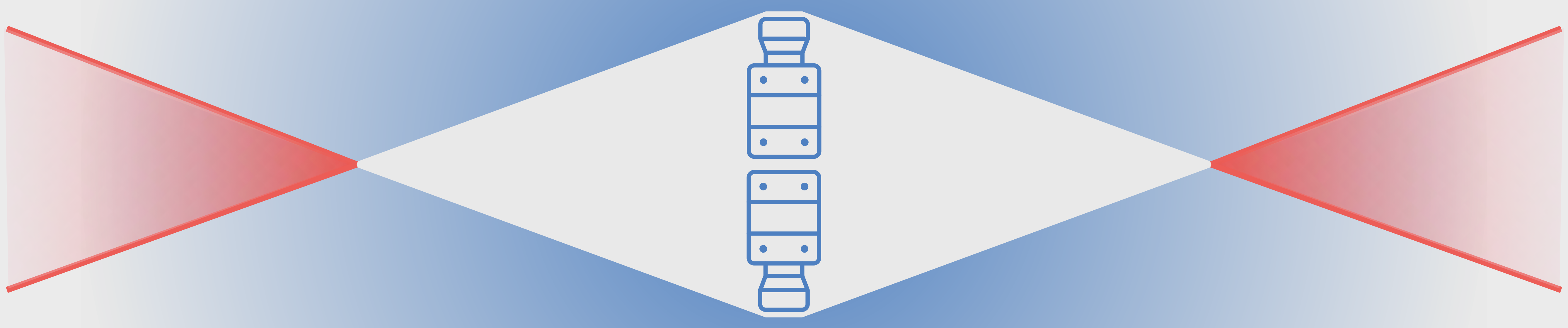
Monoscopic 360° viewing



Assumption of infinity

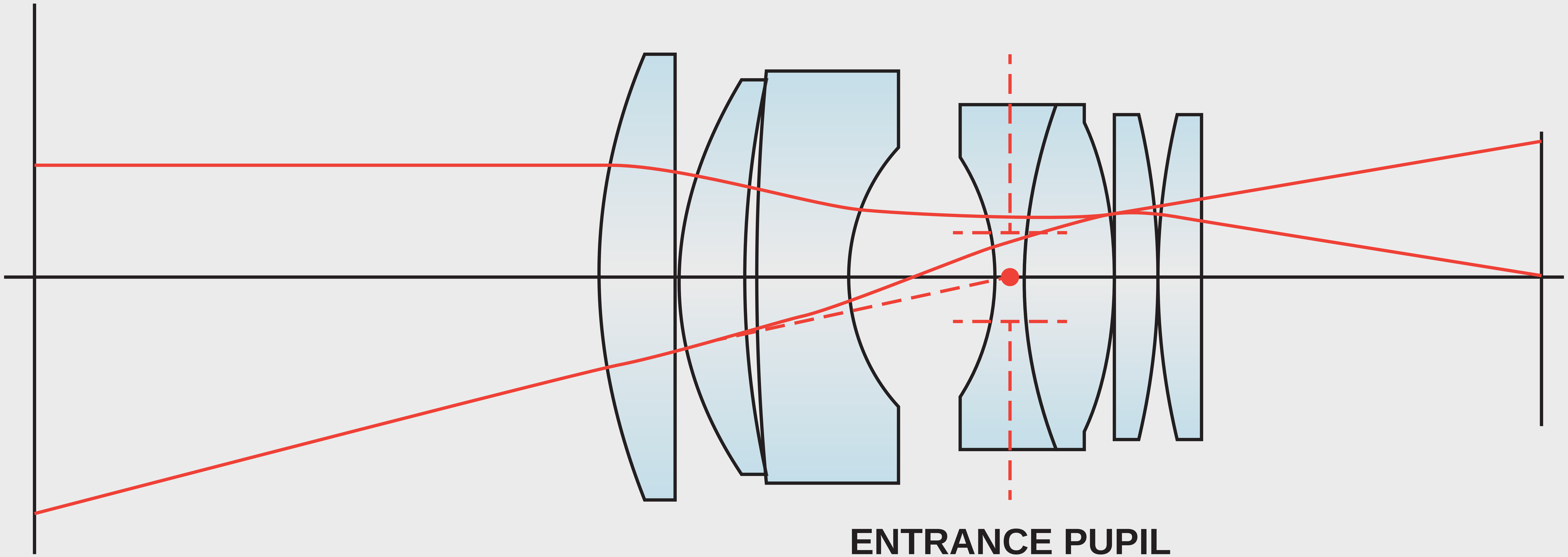


Assumption of infinity



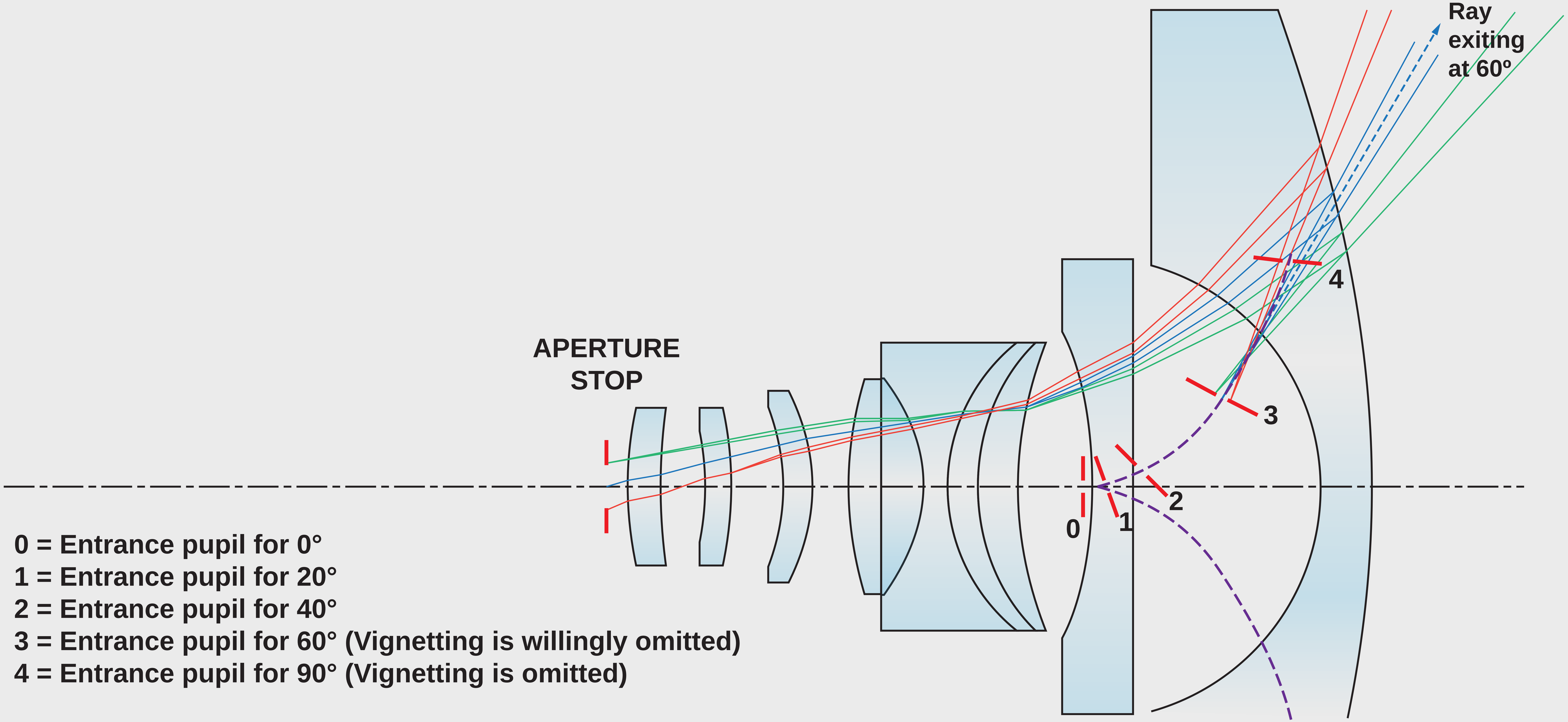
Regions of overlap have parallax
near and within hyperfocal distance

The entrance pupil the point of zero parallax



f tan(θ) lenses have single entrance pupil

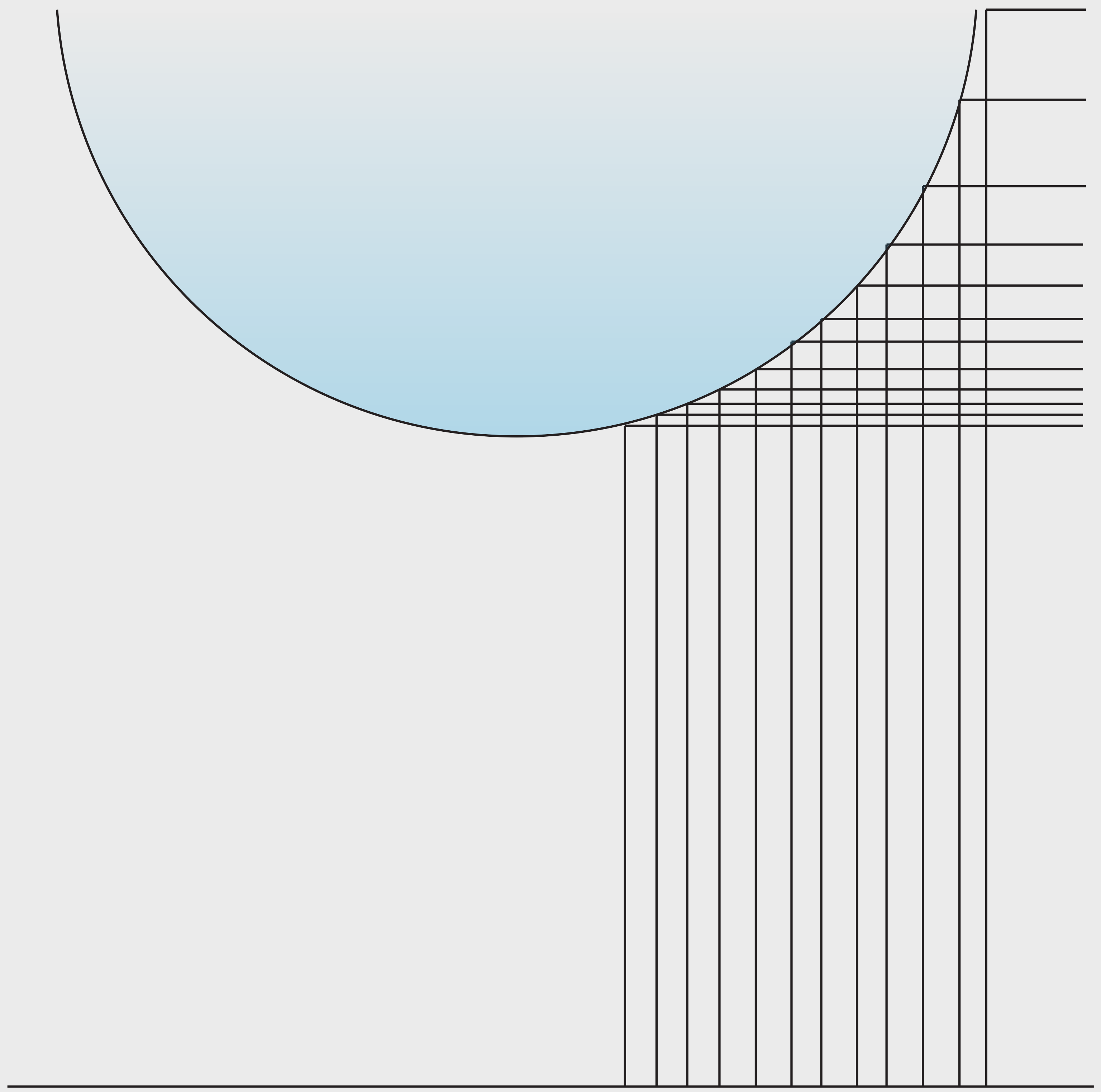
The entrance pupil Fisheye lenses



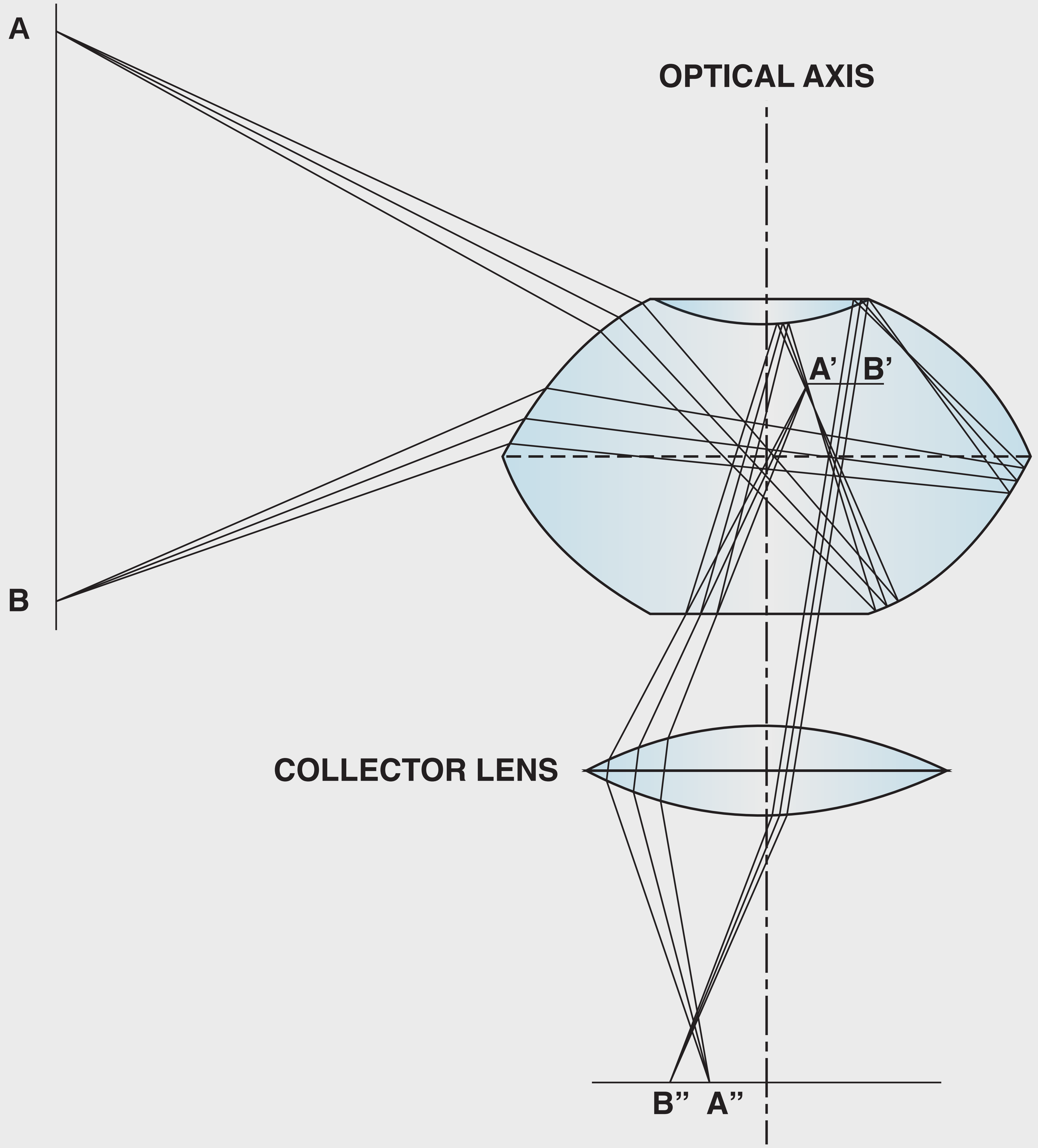
0 = Entrance pupil for 0°
1 = Entrance pupil for 20°
2 = Entrance pupil for 40°
3 = Entrance pupil for 60° (Vignetting is willingly omitted)
4 = Entrance pupil for 90° (Vignetting is omitted)

f θ lenses have no single entrance pupil

Catadioptric



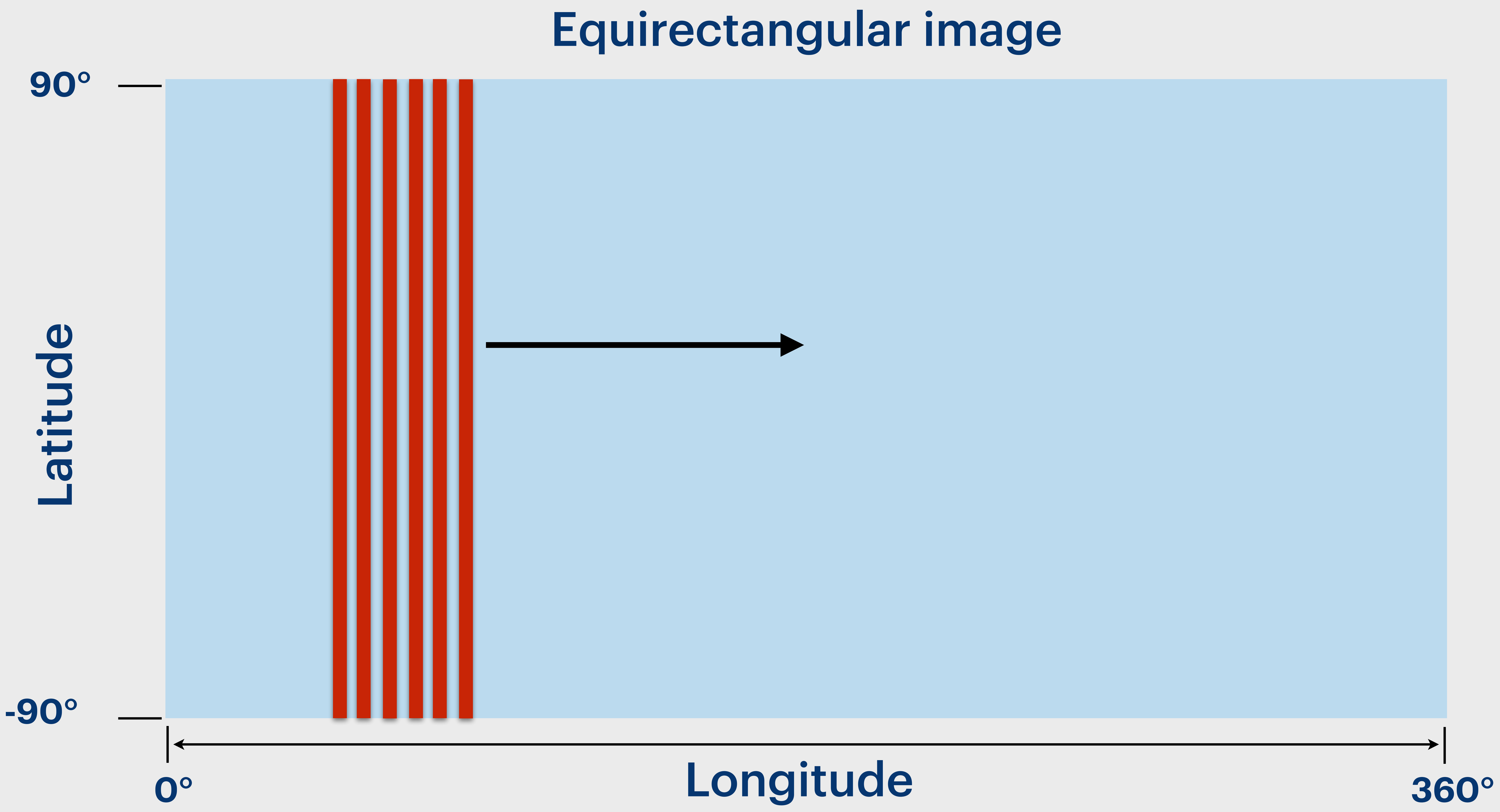
Panoramic Annular Optics



These are really cylindrical capture but no stitch lines!

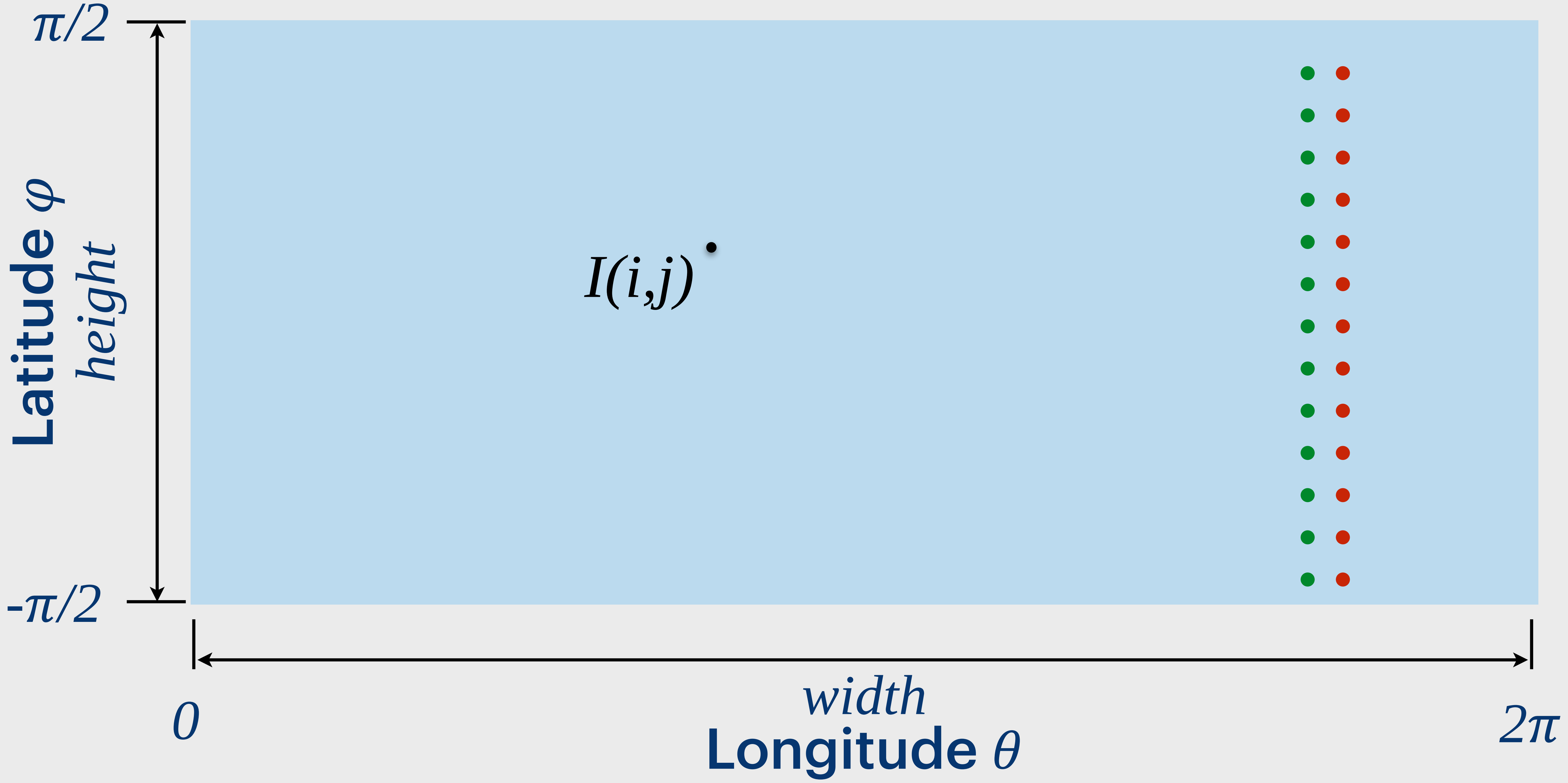
The slit-scan camera model

Another way to create a 360° image

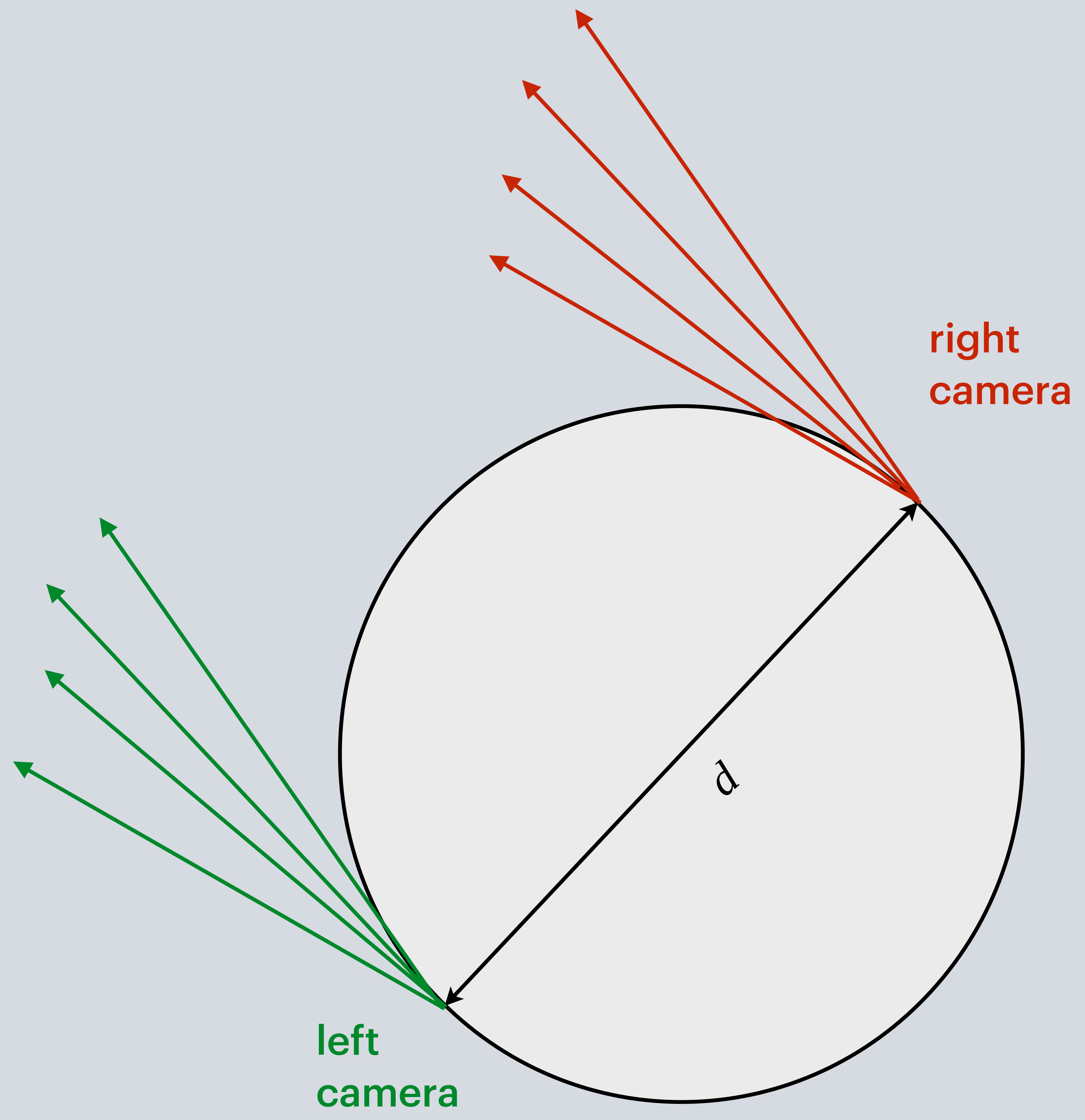


Omni-Directional Stereo (ODS)

Do slit photography for each eye



$$\varphi = \pi \; i \,/\, height - \pi/2$$
$$\theta = 2\pi \; j \,/\, width$$



where p is the inter ocular distance, $r = d/2$

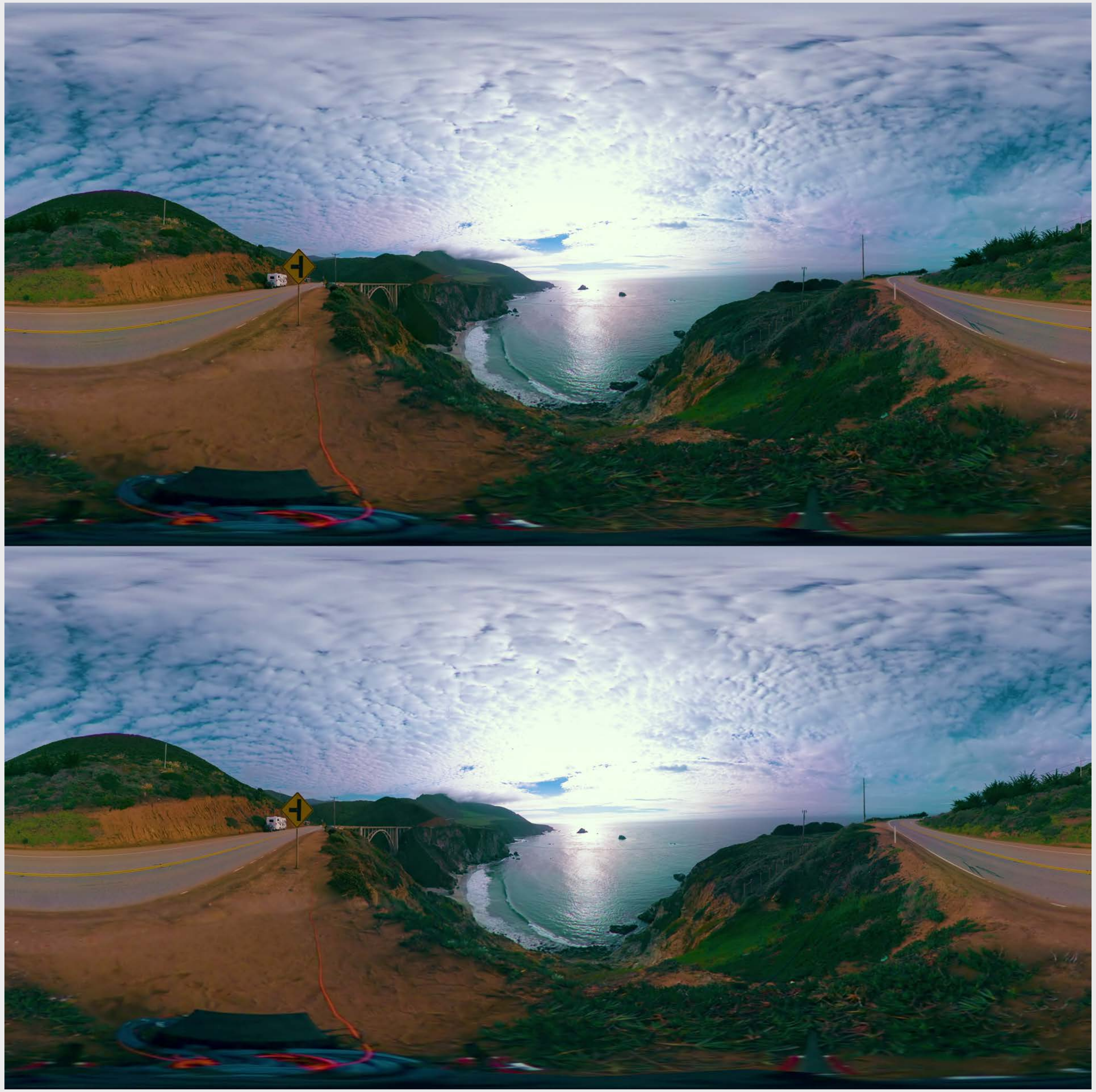
Left Eye Rays

l.x = sin(φ) * cos(θ) - r*sin(θ)
l.y = sin(φ) * sin(θ) - r*cos(θ)
l.z = cos(φ)

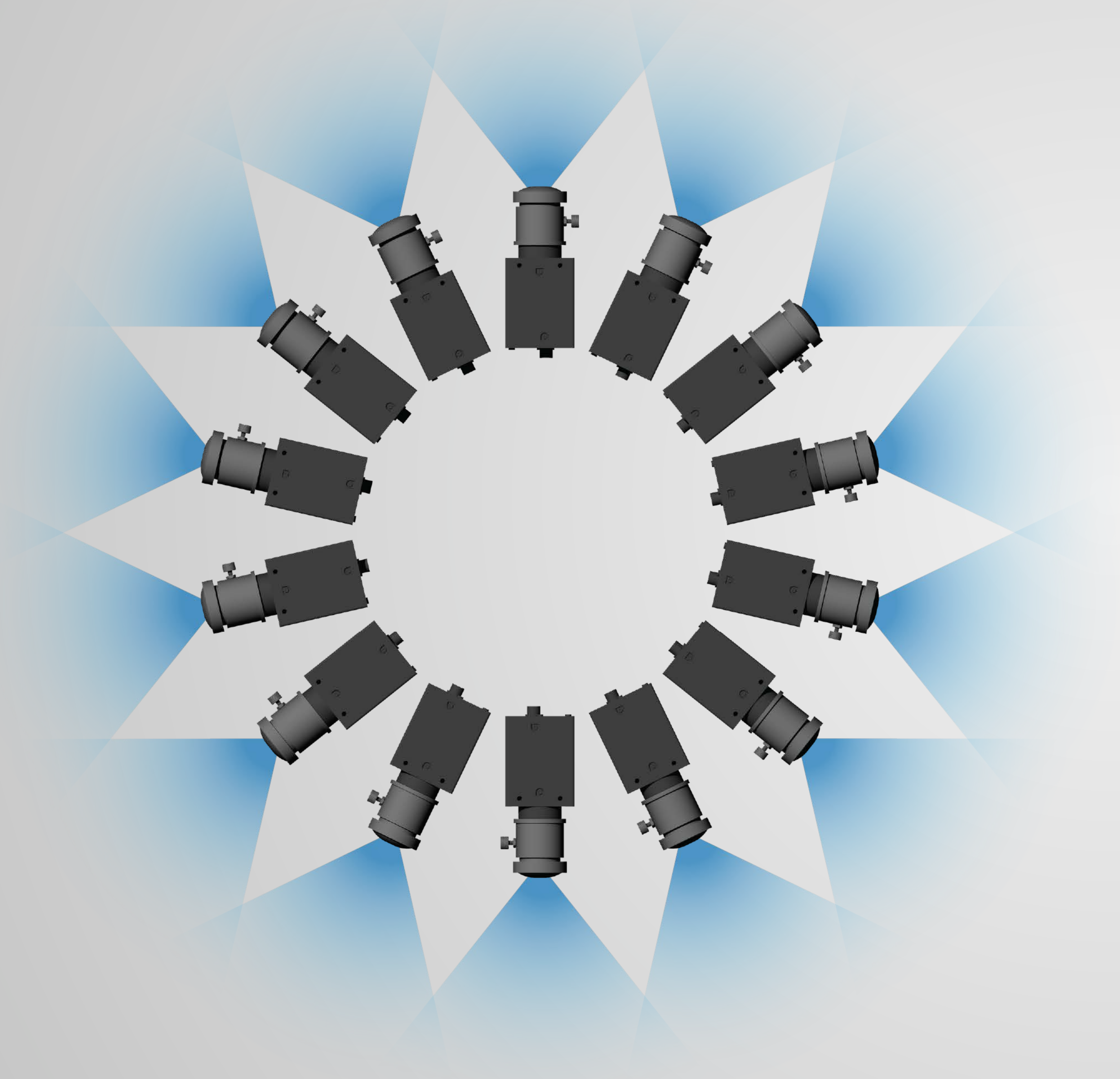
Right Eye Rays

r.x = sin(φ) * cos(θ) + r*sin(θ)
r.y = sin(φ) * sin(θ) + r*cos(θ)
r.z = cos(φ)

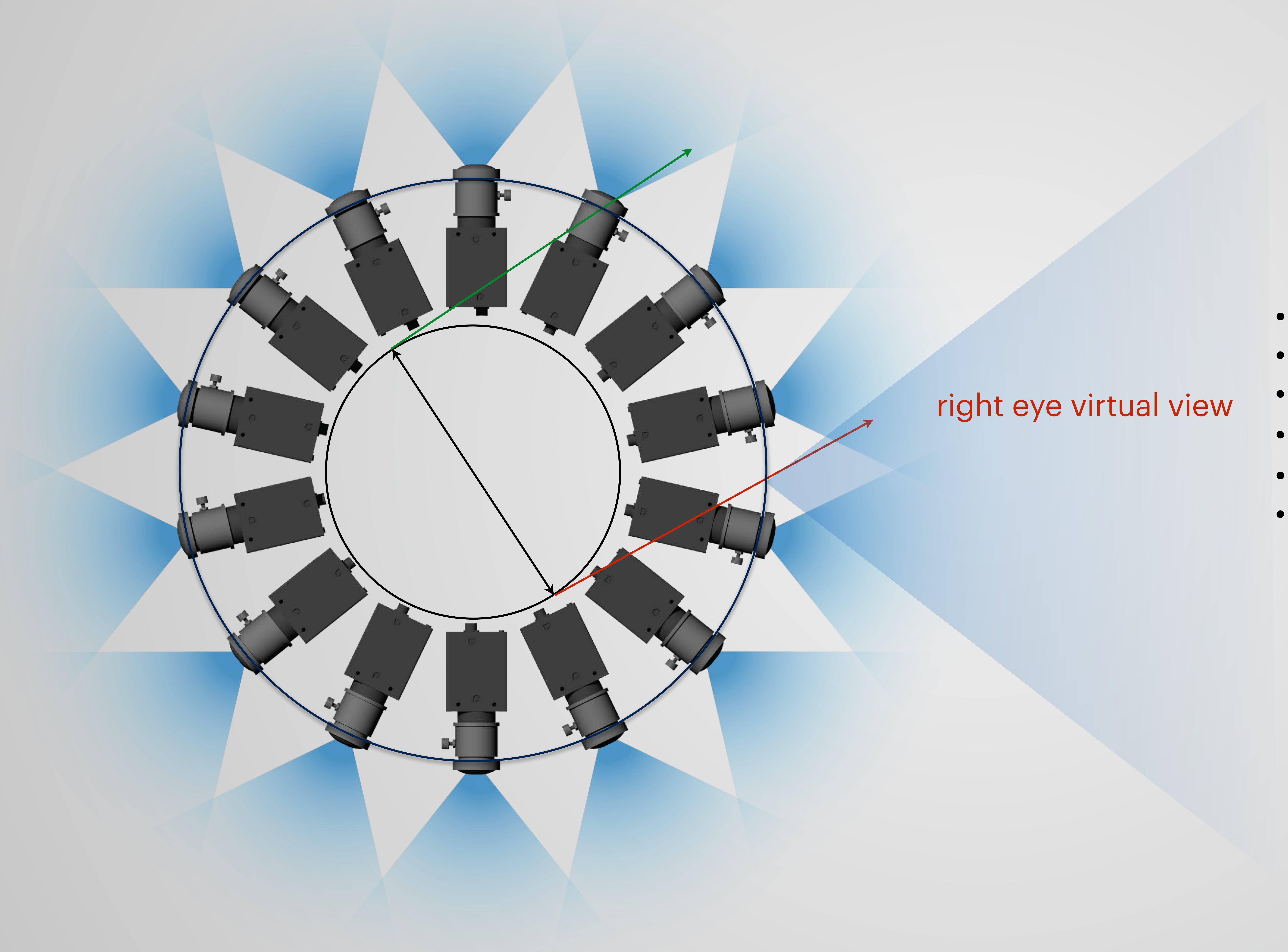
Left-right, top-bottom ODS stereo pair



Creating ODS with a fixed array of cameras

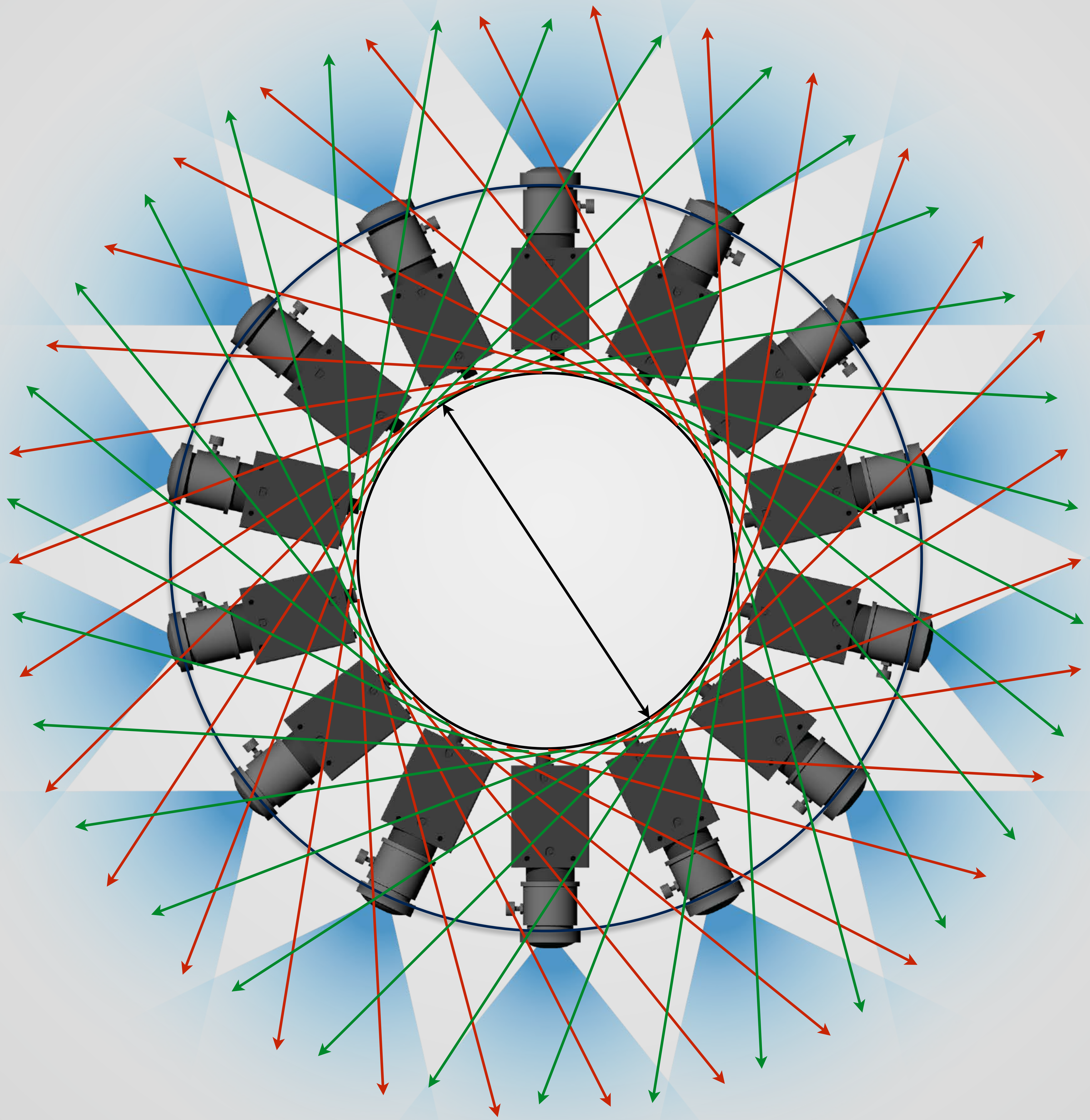


Creating ODS with a fixed array of cameras



- Warp/interpolate nearest 2 images
- Only need to do it for each specific slit
- The virtual camera is modeled as pinhole
- There are $2 * width$ slits
- Blend between cameras
- Handle ghosting via disparity clustering

Creating ODS with a fixed array of cameras



Optical flow between two images

$$I(x, y, t) = I(x + \Delta x, y + \Delta y, t + \Delta t)$$

$$I(x + \Delta x, y + \Delta y, t + \Delta t) = I(x, y, t) + \frac{\partial I}{\partial x}\Delta x + \frac{\partial I}{\partial y}\Delta y + \frac{\partial I}{\partial t}\Delta t + \ldots$$

$$\frac{\partial I}{\partial x}\Delta x + \frac{\partial I}{\partial y}\Delta y + \frac{\partial I}{\partial t}\Delta t = 0$$

$$\frac{\partial I}{\partial x}\frac{\Delta x}{\Delta t} + \frac{\partial I}{\partial y}\frac{\Delta y}{\Delta t} + \frac{\partial I}{\partial t}\frac{\Delta t}{\Delta t} = 0$$

$$I_x V_x + I_y V_y = -I_t$$

Horn and Schunck

$$E = \int\int [(I_x V_x + I_y V_y + I_t)^2 + \alpha^2(||\nabla V_x||^2 + ||\nabla V_y||^2)] dx dy$$

Solving the 3-D Euler-Lagrange equations

$$I_x(I_x^{-k} V_x + I_y^{-k} V_y + I_t) - \alpha^2 \Delta V_x = 0$$
$$I_y(I_x^{-k} V_x + I_y^{-k} V_y + I_t) - \alpha^2 \Delta V_y = 0$$

Using finite difference approximations and rearranging

$$(I_x^2 + 4\alpha^2) V_x + I_x I_y V_y = 4\alpha^2 \overline{V_x} - I_x I_t$$
$$(I_y^2 + 4\alpha^2) V_y + I_x I_y V_x = 4\alpha^2 \overline{V_y} - I_y I_y$$

Solving for the next flow time step

$$V_x^{k+1} = V_x^{-k} - \frac{I_x(I_x^{-k} V_x + I_y^{-k} V_y + I_t)}{4\alpha^4 + I_x^2 + I_y^2}$$

$$V_y^{k+1} = V_y^{-k} - \frac{I_y(I_x^{-k} V_x + I_y^{-k} V_y + I_t)}{4\alpha^4 + I_x^2 + I_y^2}$$

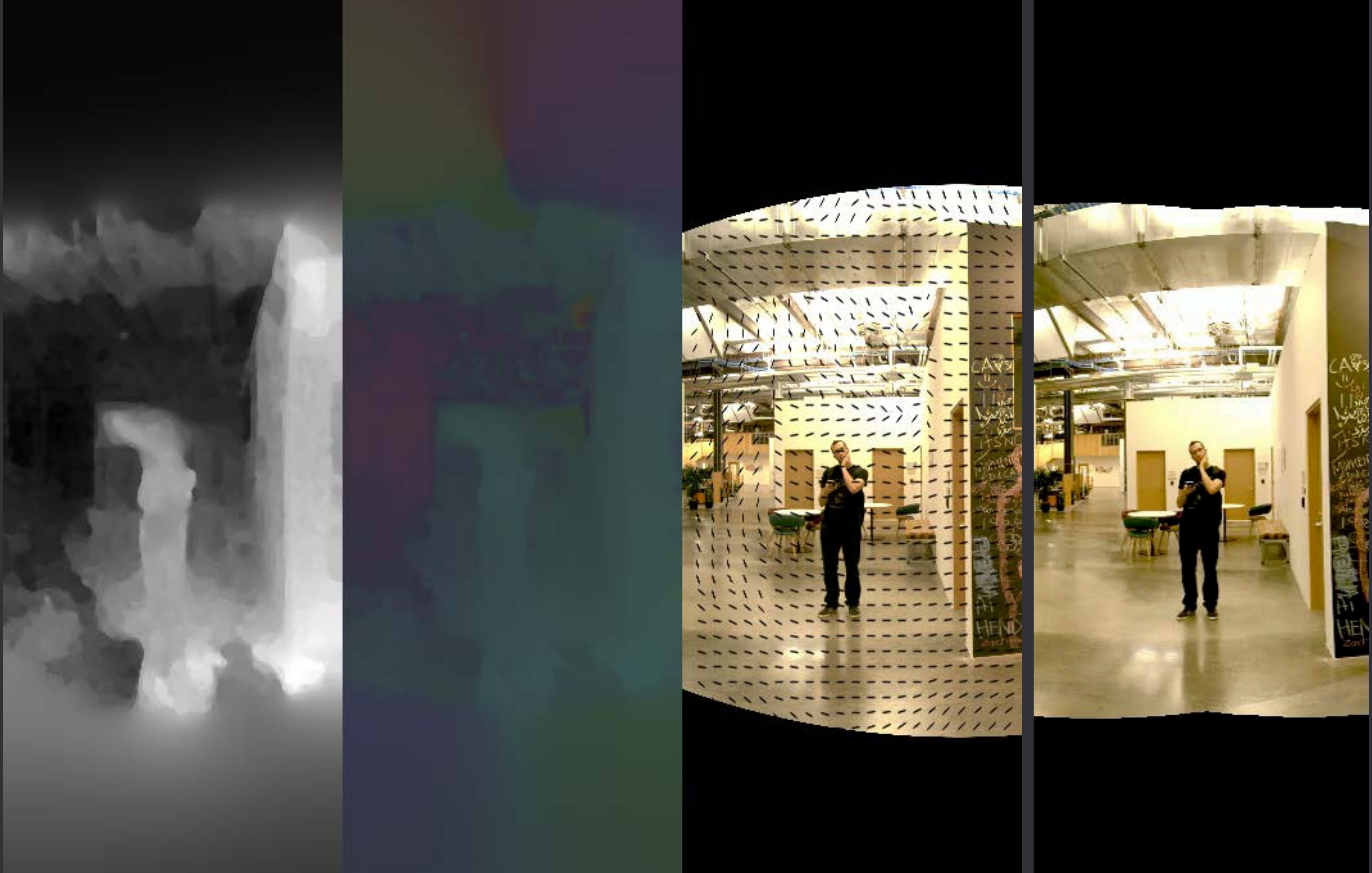

HENDRIK!


Riya
HENDRIK!

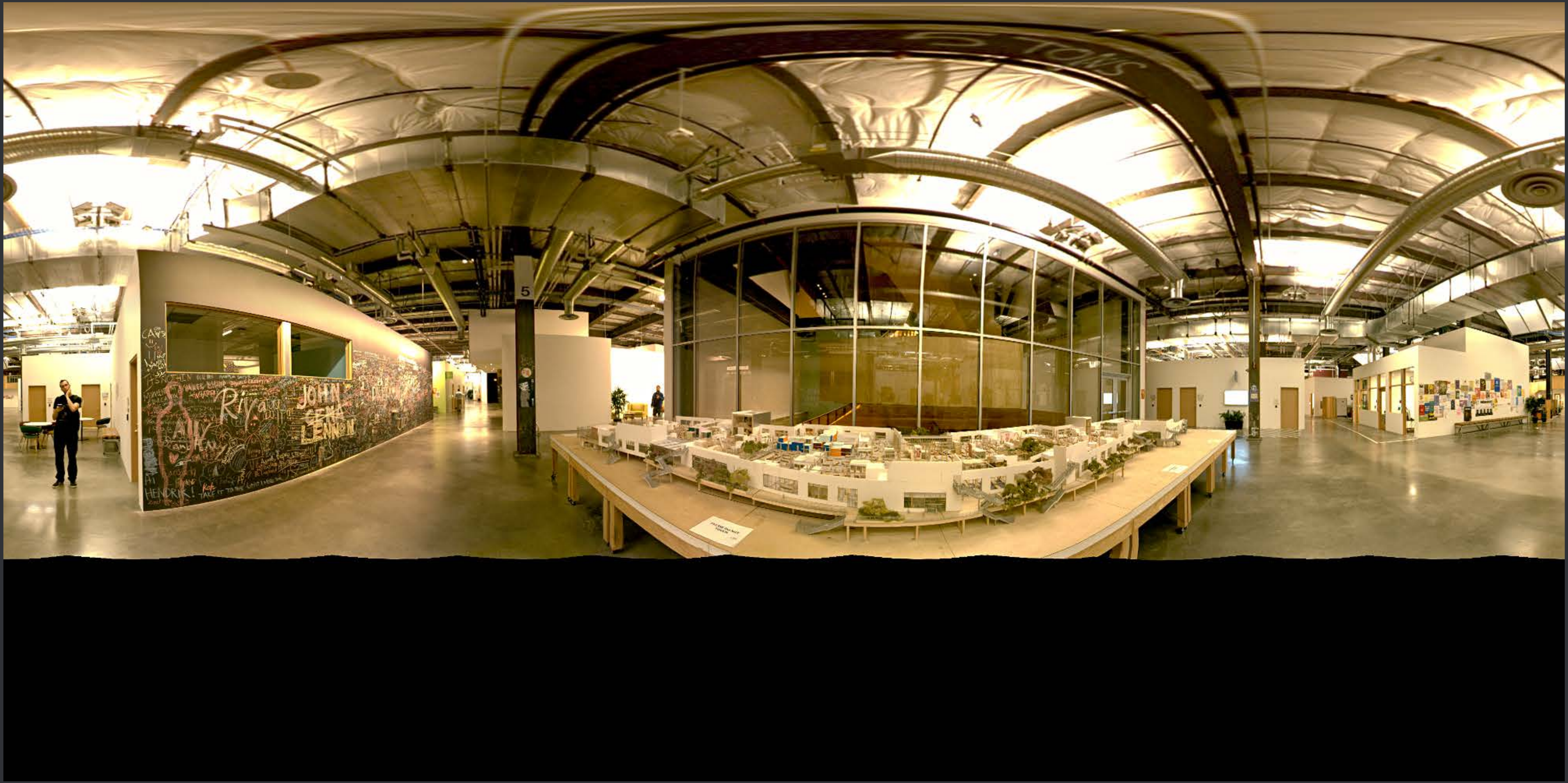

JOHN
LENNON
HENDRIK!


5
JOHN
LENNON
HENDRIK!

Spherical projections align s.t. parallax = 0 @ infinity



Sharpen (Periodic Boundary Aware)




NEW HAVEN LINE DEPARTURES
INFORMATION

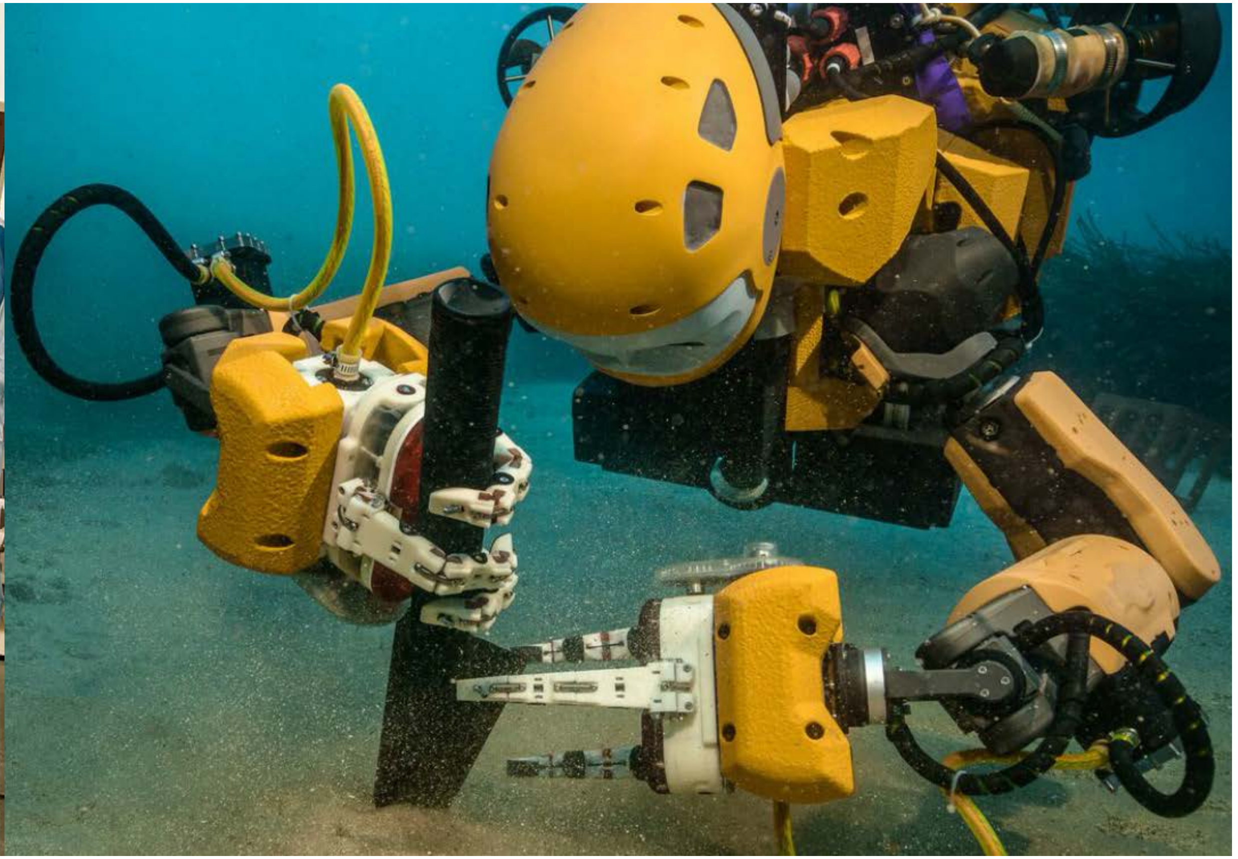

Thank you

Live ODS Video

Robert Konrad


PHILIPS ARENA
Aaron's Own it.
StateFarm
StateFarm
HAWKS 52
RAPTORS 54
RAPTORS 31
ATLANTA
PANIC 20
Aaron's

Live ODS Video



(Some) Stereo VR Cameras

Z-Cam V1 Pro

Kandao Obsidian R

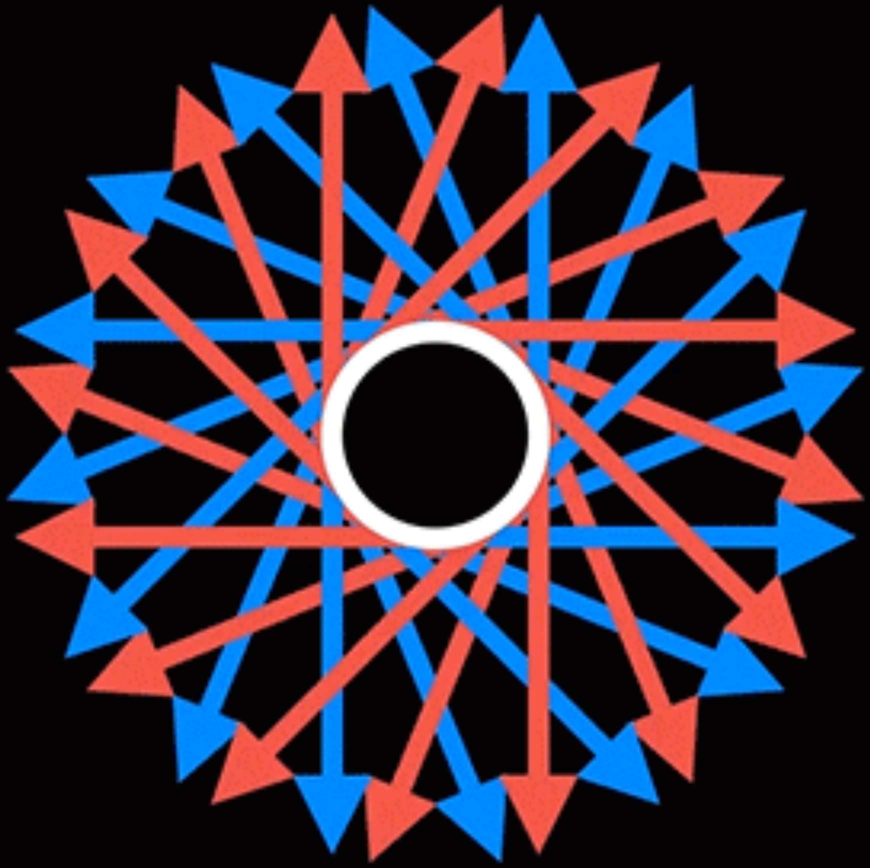
Insta360 Titan

Facebook Surround 360

Facebook Manifold

Yi Halo

Omnidirectional Stereo (ODS)



Ishiguro et al. 1990
Peleg et al. 2001

Getting these tangential rays is non-trivial from camera arrays!

[Anderson et. al, SIGGRAPH Asia 2016]

CAMERA 1

[Anderson et. al, SIGGRAPH Asia 2016]

Challenges for optical flow

Transparency, reflections

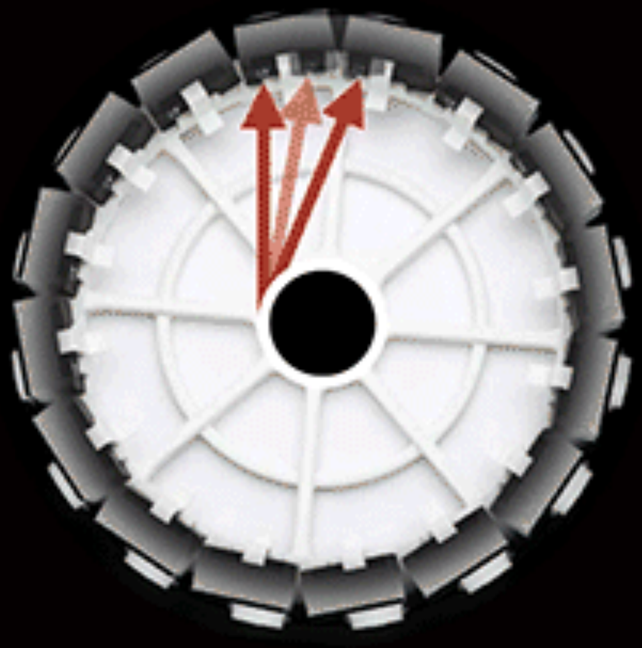
Flow mismatches

Fine Structures

Too Close

[Anderson 2016]

INTERPOLATION



Per-Frame Flow

Temporally Consistent Flow

Anderson et. al, SIGGRAPH Asia 2016

Google Jump – Computation Time Breakdown

60 sec / frame

→

(single machine)

[Anderson et. al, SIGGRAPH Asia 2016]

Google Jump – Computation Time Breakdown

Video Length		Total time to process
1 hour	60 sec / frame → (single machine)	75 days

[Anderson et. al, SIGGRAPH Asia 2016]

Google Jump – Computation Time Breakdown

Video Length		Total time to process
1 hour	60 sec / frame → (single machine)	75 days
1 hour	321 sec / frame → (parallelized on 1000 cores)	10 hours

[Anderson et. al, SIGGRAPH Asia 2016]

Google Jump – Computation Time Breakdown

Video Length		Total time to process
1 hour	60 sec / frame → (single machine)	75 days
1 hour	321 sec / frame → (parallelized on 1000 cores)	10 hours

Operation	Time (sec.)
Flow computation	183
Compositing	54
Frame IO and rectification	40
Flow compression/decompression	38
Post processing/one off setup	6
Total	321

Google Jump – Computation Time Breakdown

Video Length		Total time to process
1 hour	60 sec / frame (single machine)	75 days
1 hour	321 sec / frame (parallelized on 1000 cores)	10 hours

Operation	Time (sec.)
Flow computation	183
Compositing	54
Frame IO and rectification	40
Flow compression/decompression	38
Post processing/one off setup	6
Total	321

Google Jump – Computation Time Breakdown

Video Length		Total time to process
1 hour	60 sec / frame → (single machine)	75 days
1 hour	321 sec / frame → (parallelized on 1000 cores)	10 hours

Operation	Time (sec.)
Flow computation	183
Compositing	54
Frame IO and rectification	40
Flow compression/decompression	38
Post processing/one off setup	6
Total	321

What is real time?

processing time $\leq$ capture rate

capture rate $\neq$ display rate

Live Streaming Camera Arrays

Single Shot ODS capture

Rotating ODS capture

Live Streaming Camera Arrays

Kandao Obsidian R

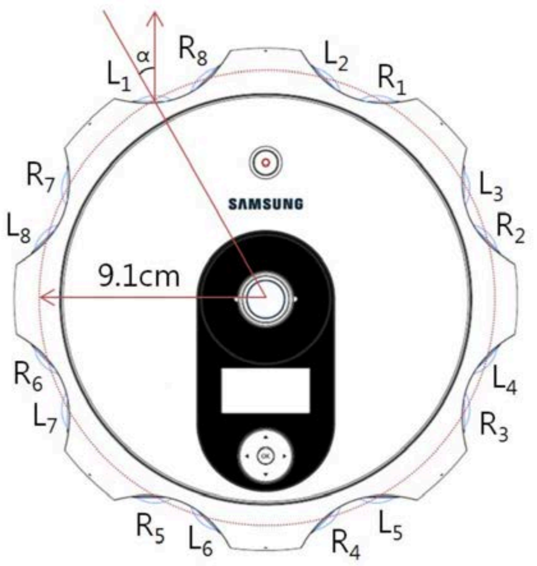
Samsung Round

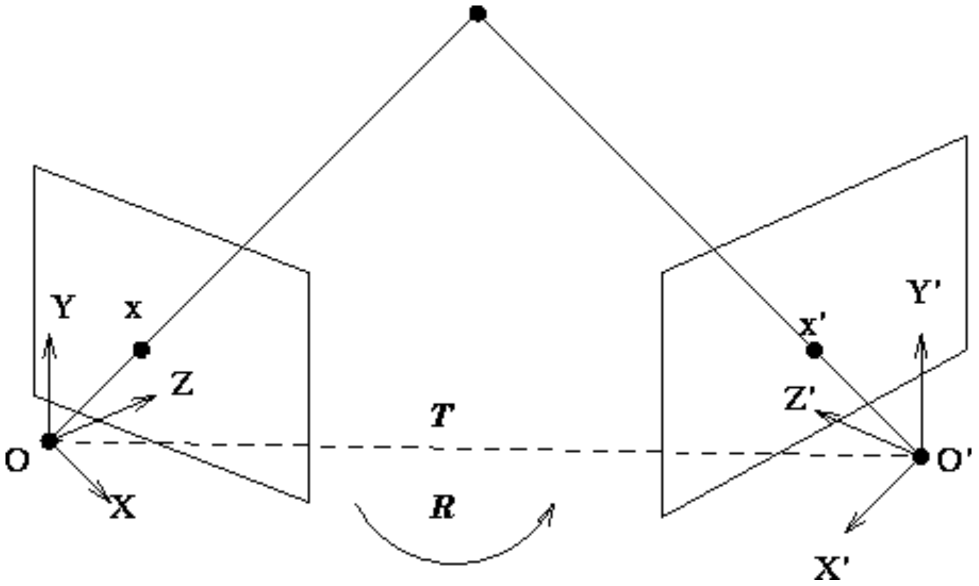
Z-Cam V1 Pro

LivePlanet

Insta360 Titan

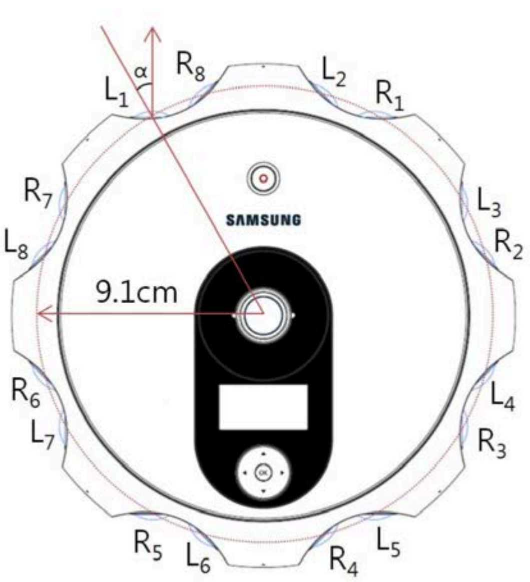
Rely on calibration

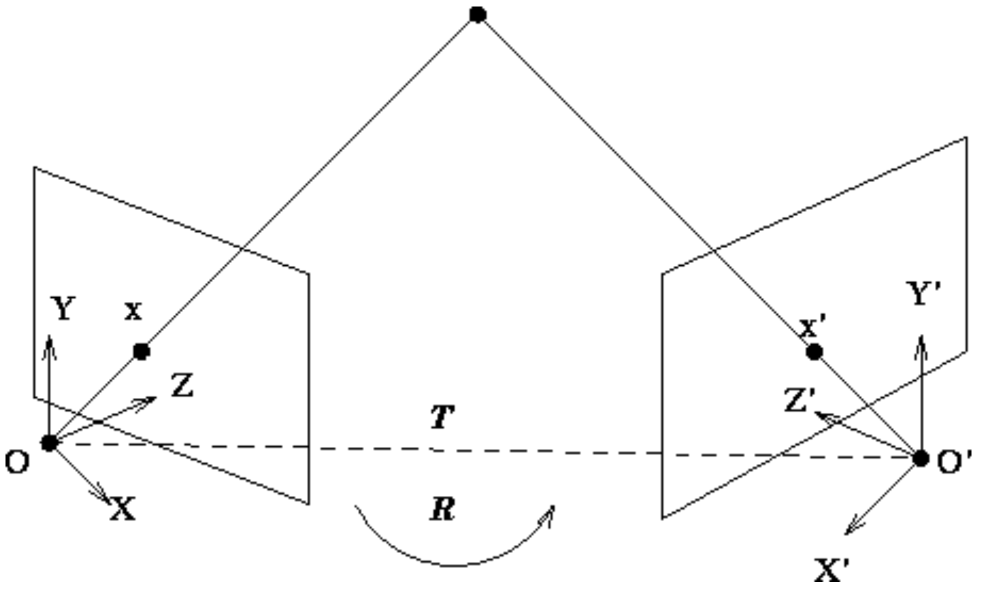


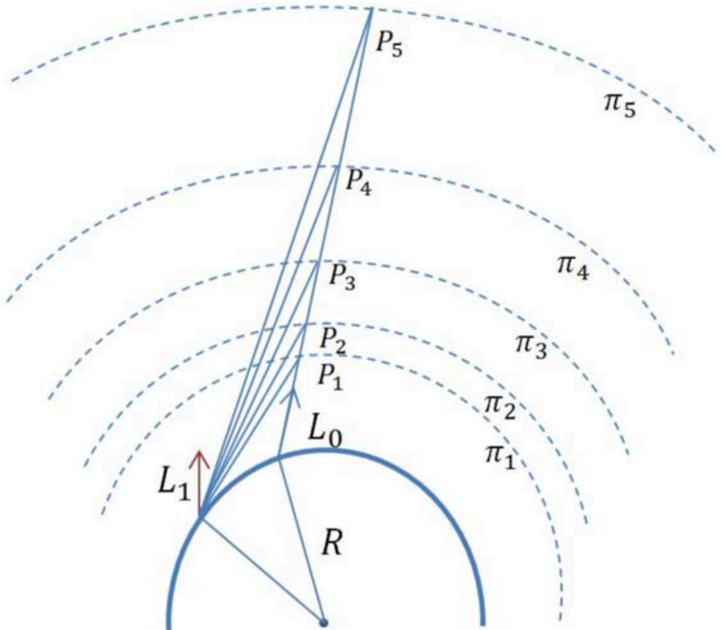


[Limonov, ICCE 2018]

Rely on calibration

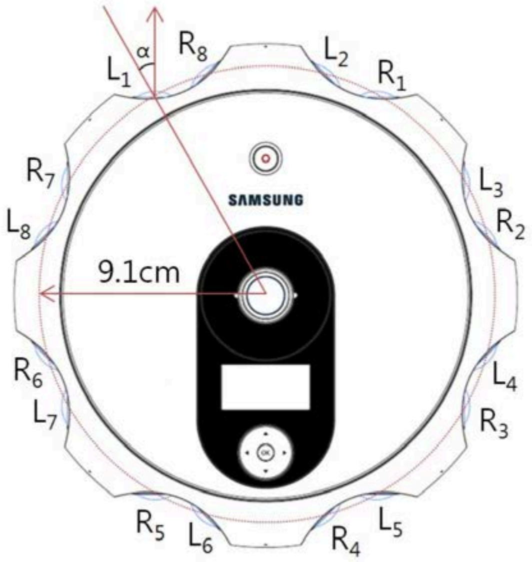


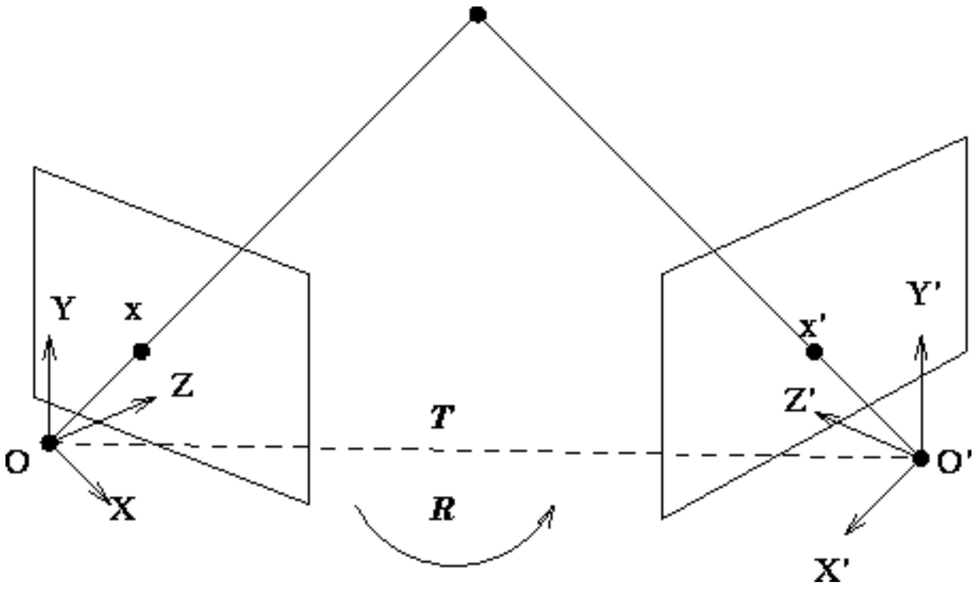


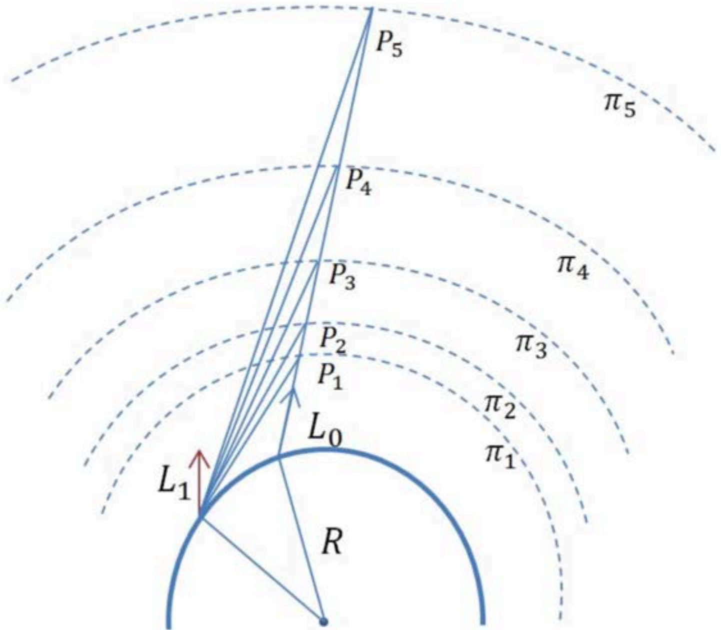


[Limonov, ICCE 2018]

Rely on calibration



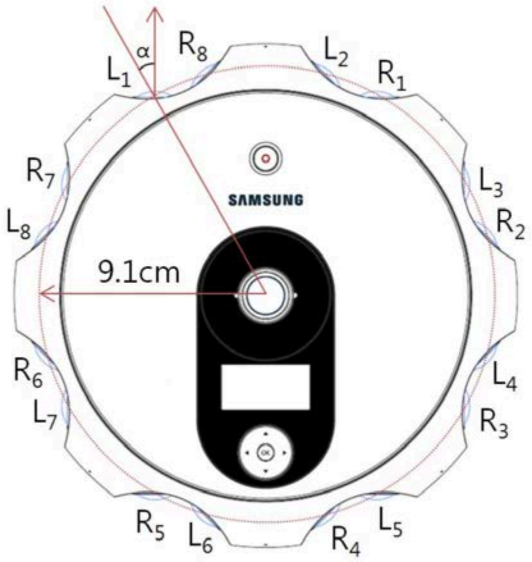


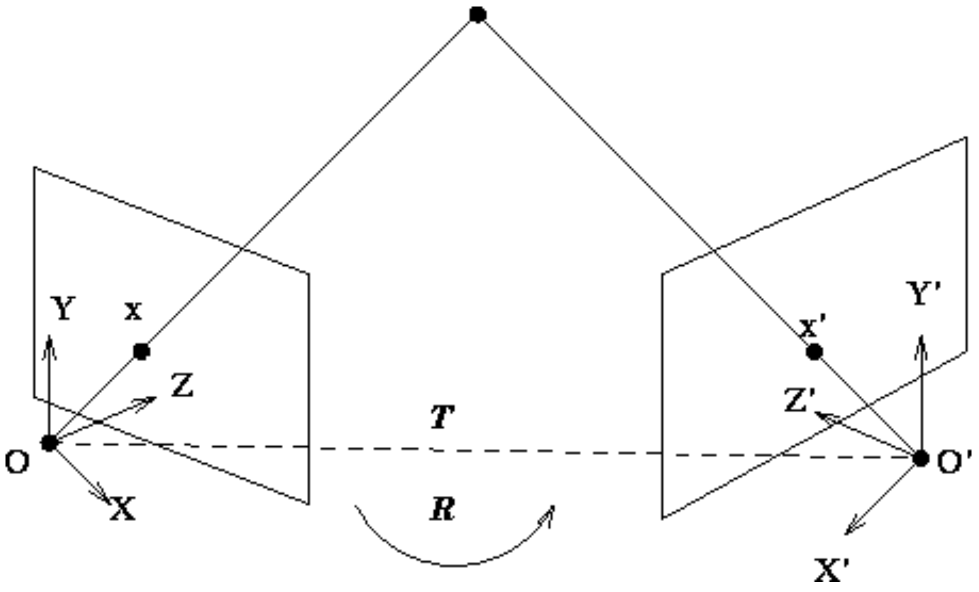


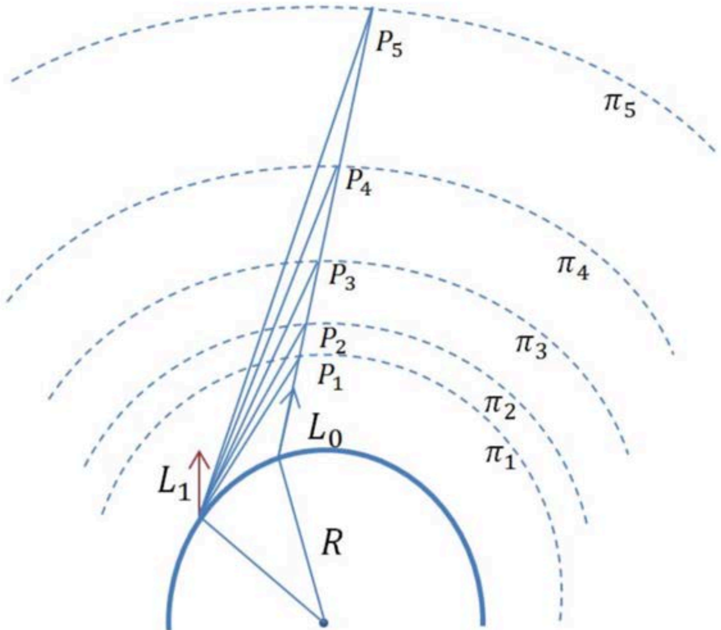
Don't compute 2D optical flow

[Limonov, ICCE 2018]

Rely on calibration





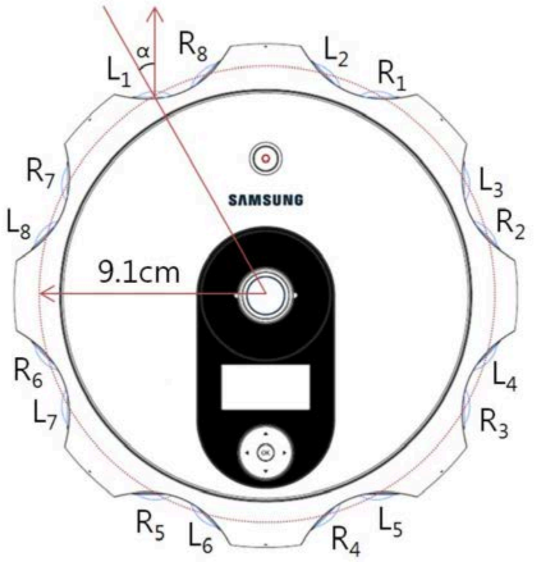


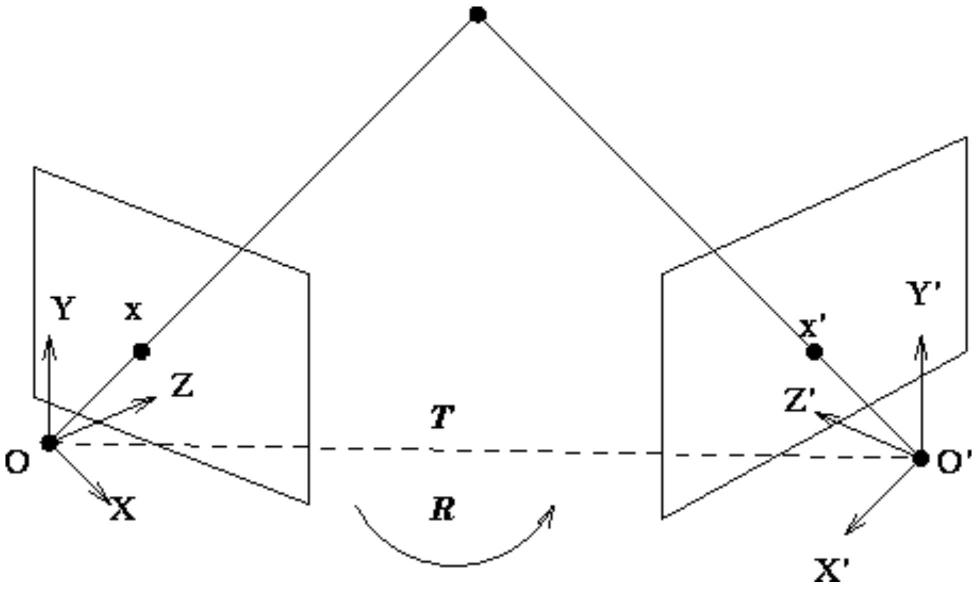
Don't compute 2D optical flow

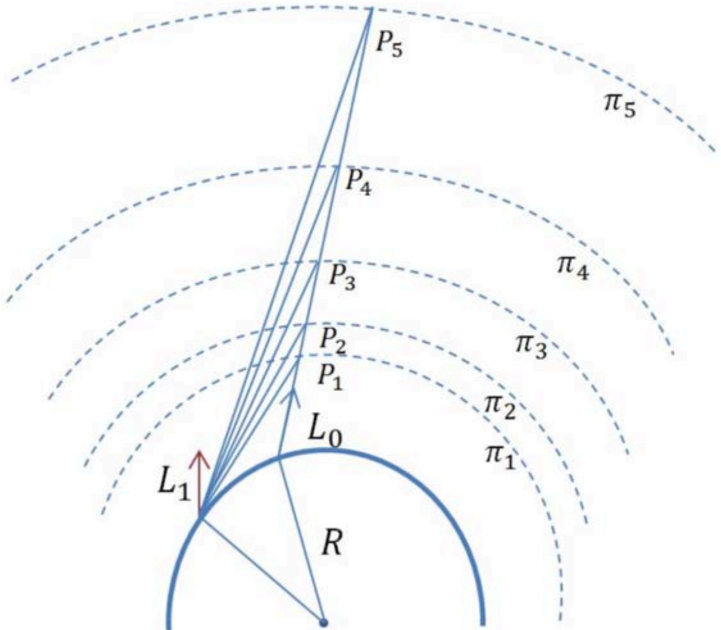
Don't compute 2D optical flow

[Limonov, ICCE 2018]

Rely on calibration







Don't compute 2D optical flow

Don't compute 2D optical flow

No guarantee on temporal consistency

[Limonov, ICCE 2018]

Artifacts

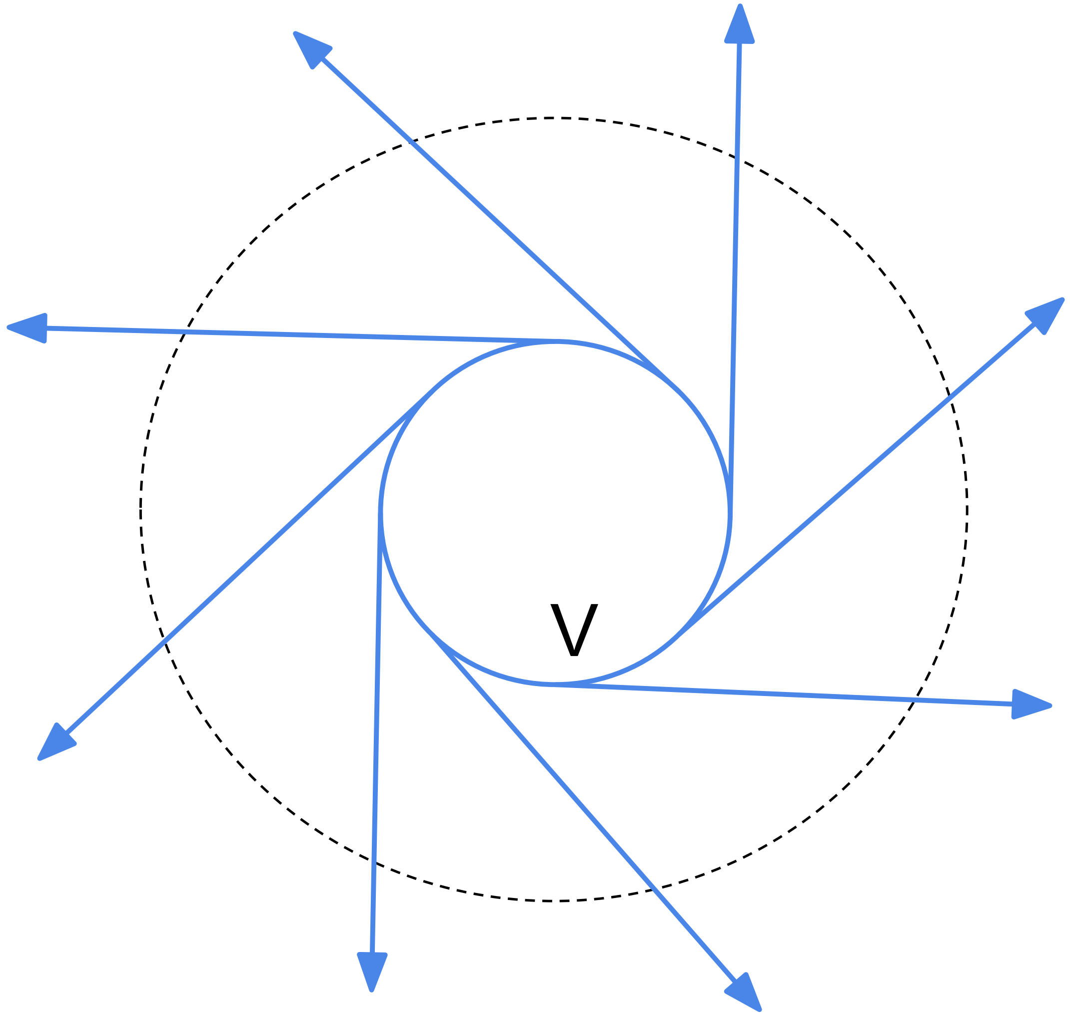
(video courtesy of CreatorUp Youtube Channel)

Live Streaming Camera Arrays

Single Shot ODS capture 1

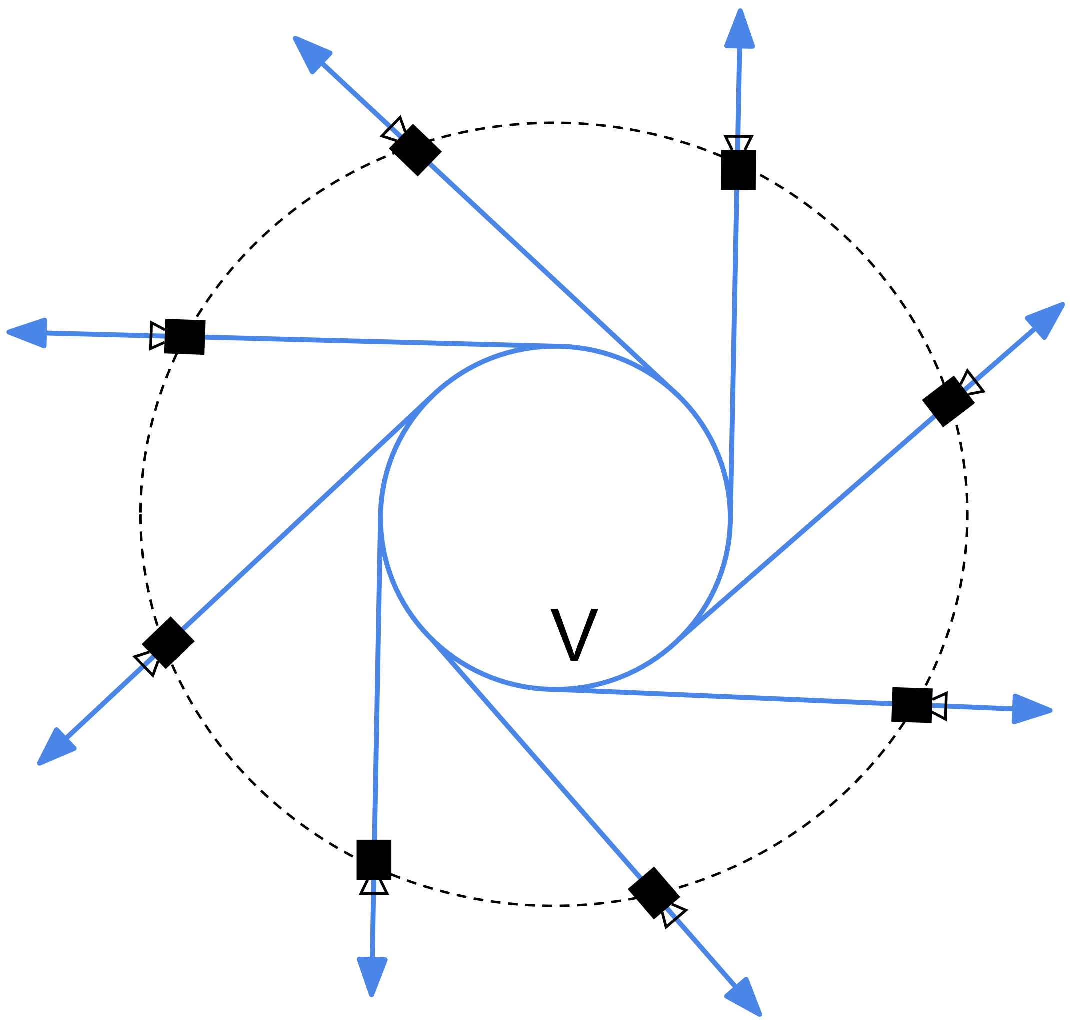
Rotating ODS capture

Single Shot Stereo ODS



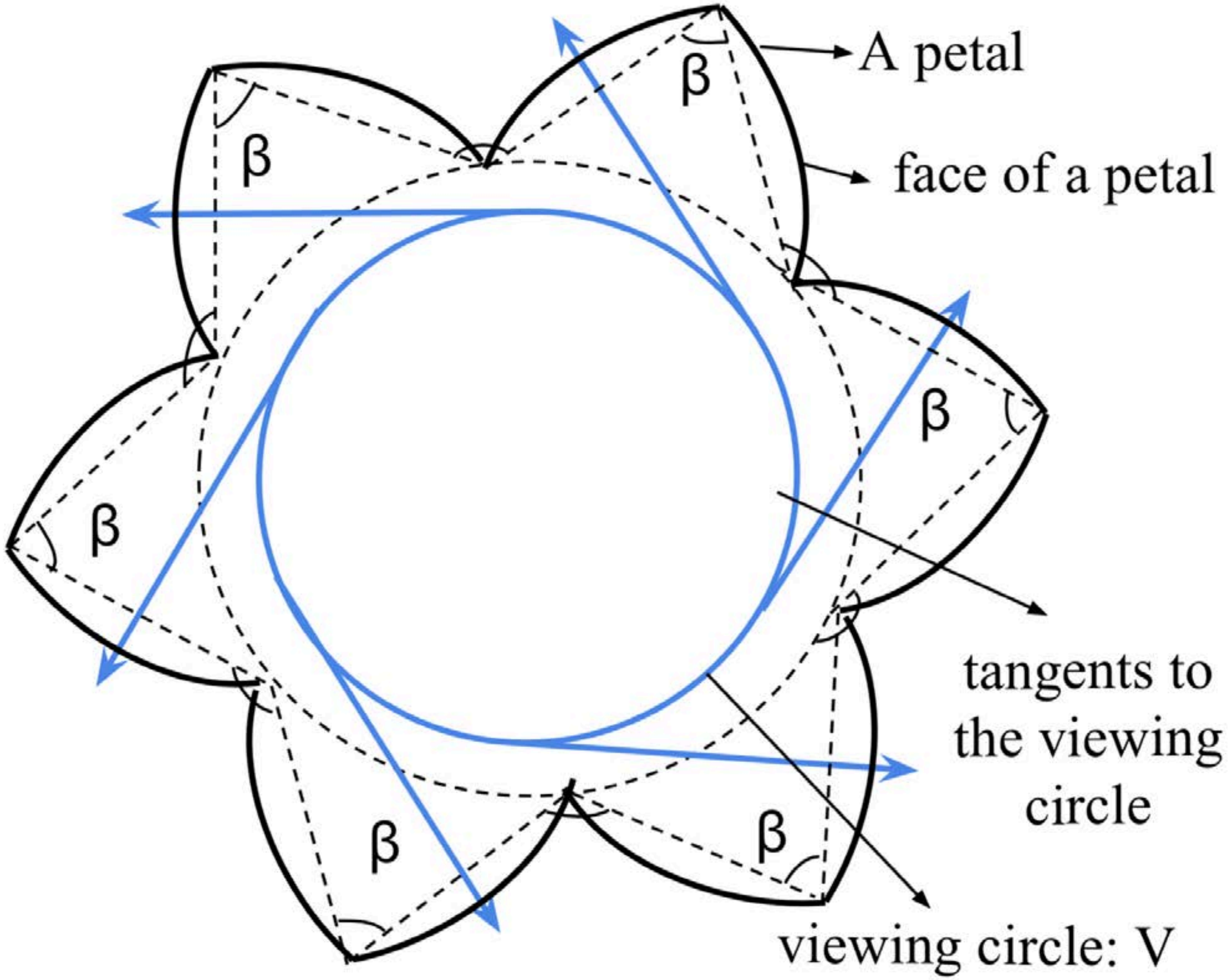
[Aggarwal, CVPR 2016]

Single Shot Stereo ODS



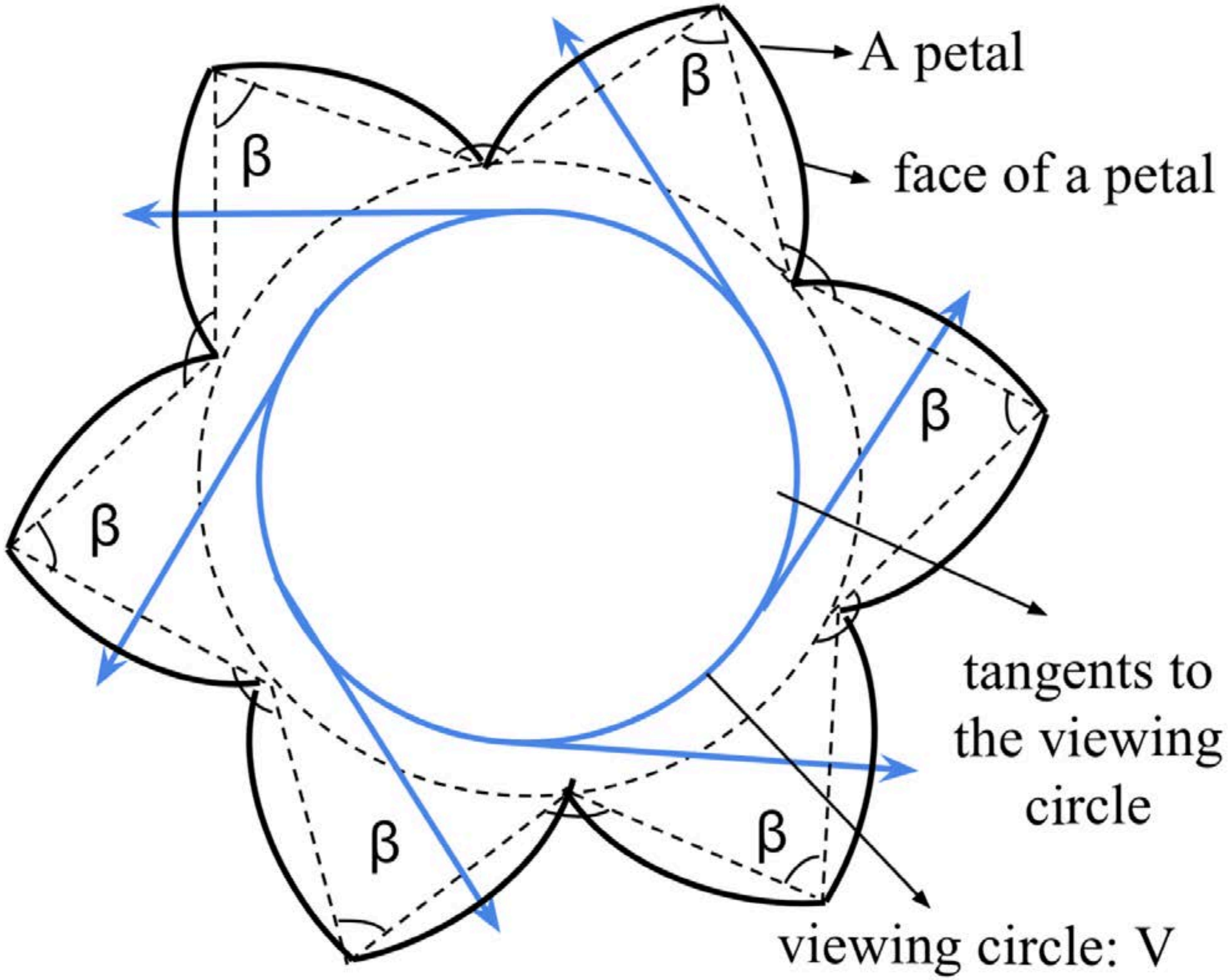
[Aggarwal, CVPR 2016]

Single Shot Stereo ODS



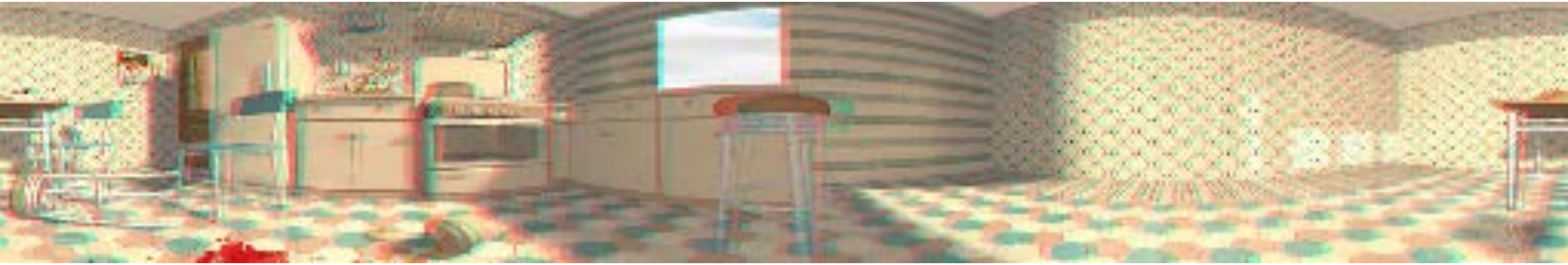
[Aggarwal, CVPR 2016]

Single Shot Stereo ODS

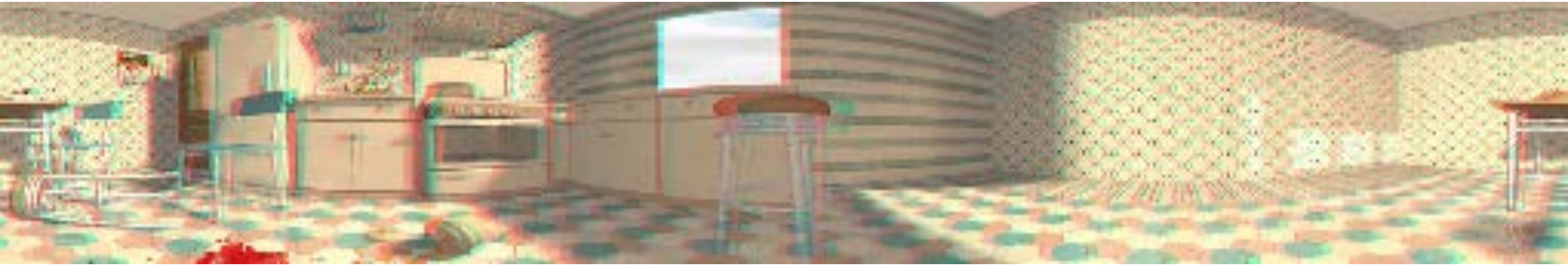


[Aggarwal, CVPR 2016]

Simulated Capture

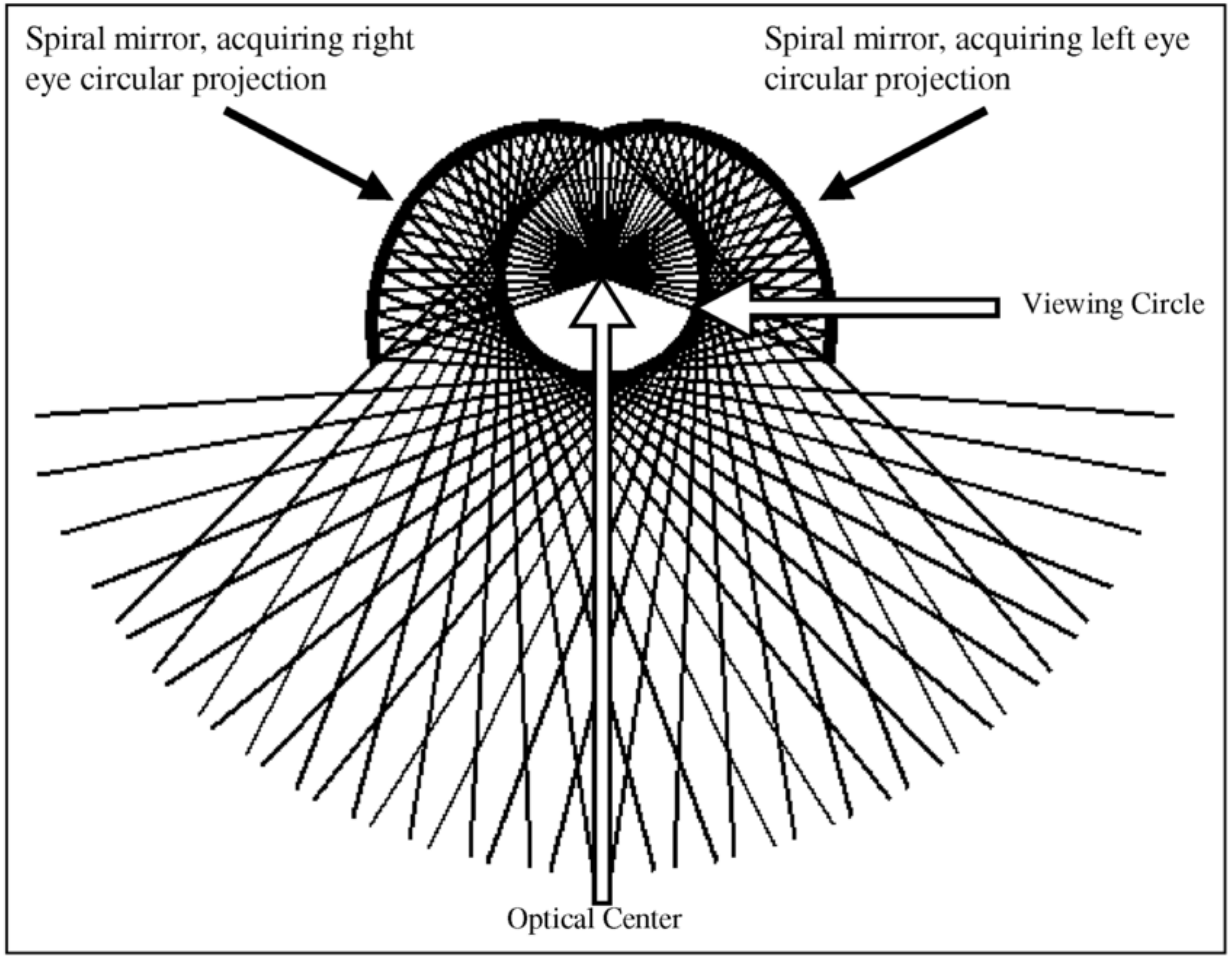


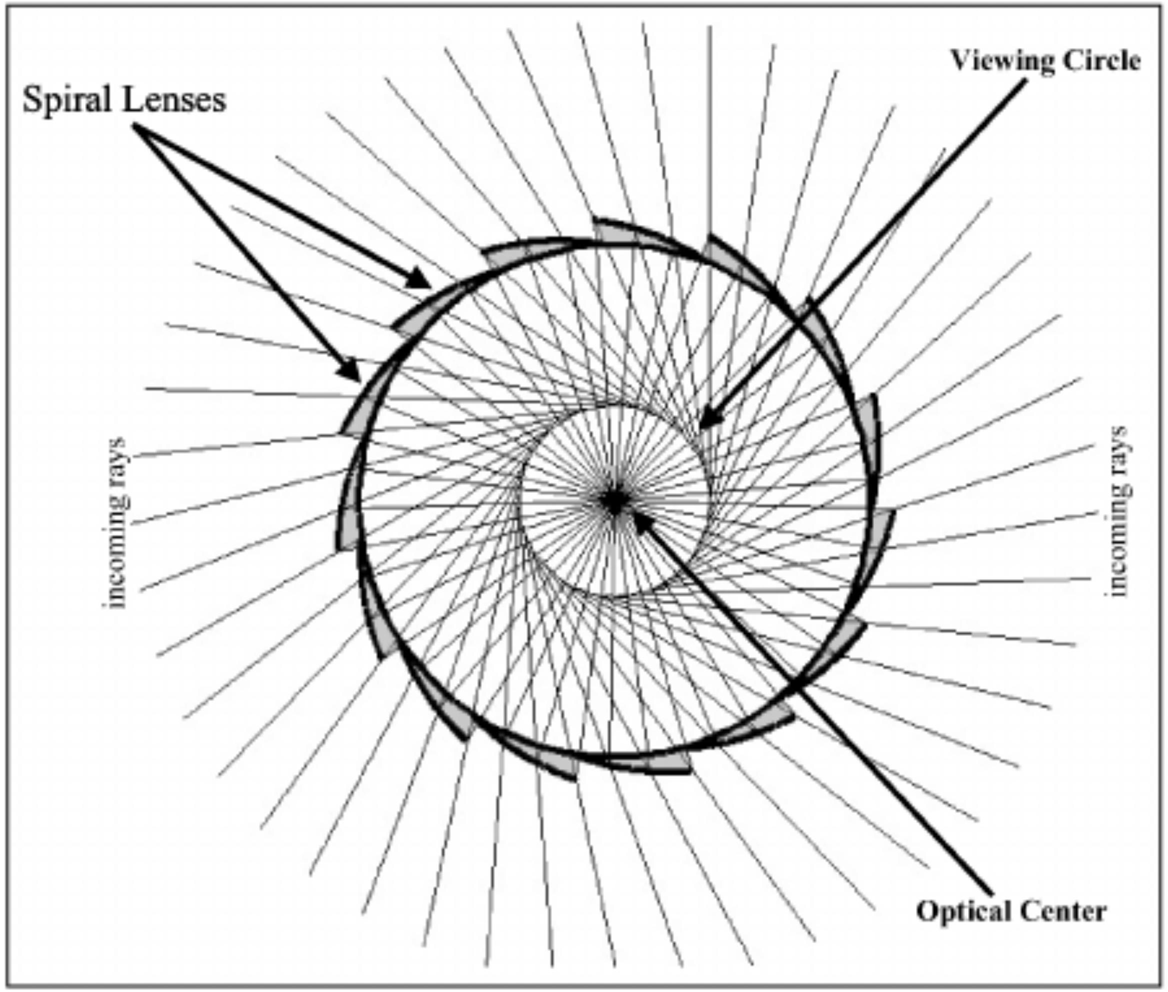
Simulated Capture



Real Capture

Other Optical Designs





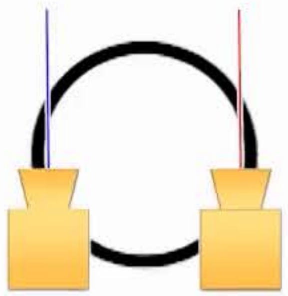
[Peleg, IEEE TPAM 2001]

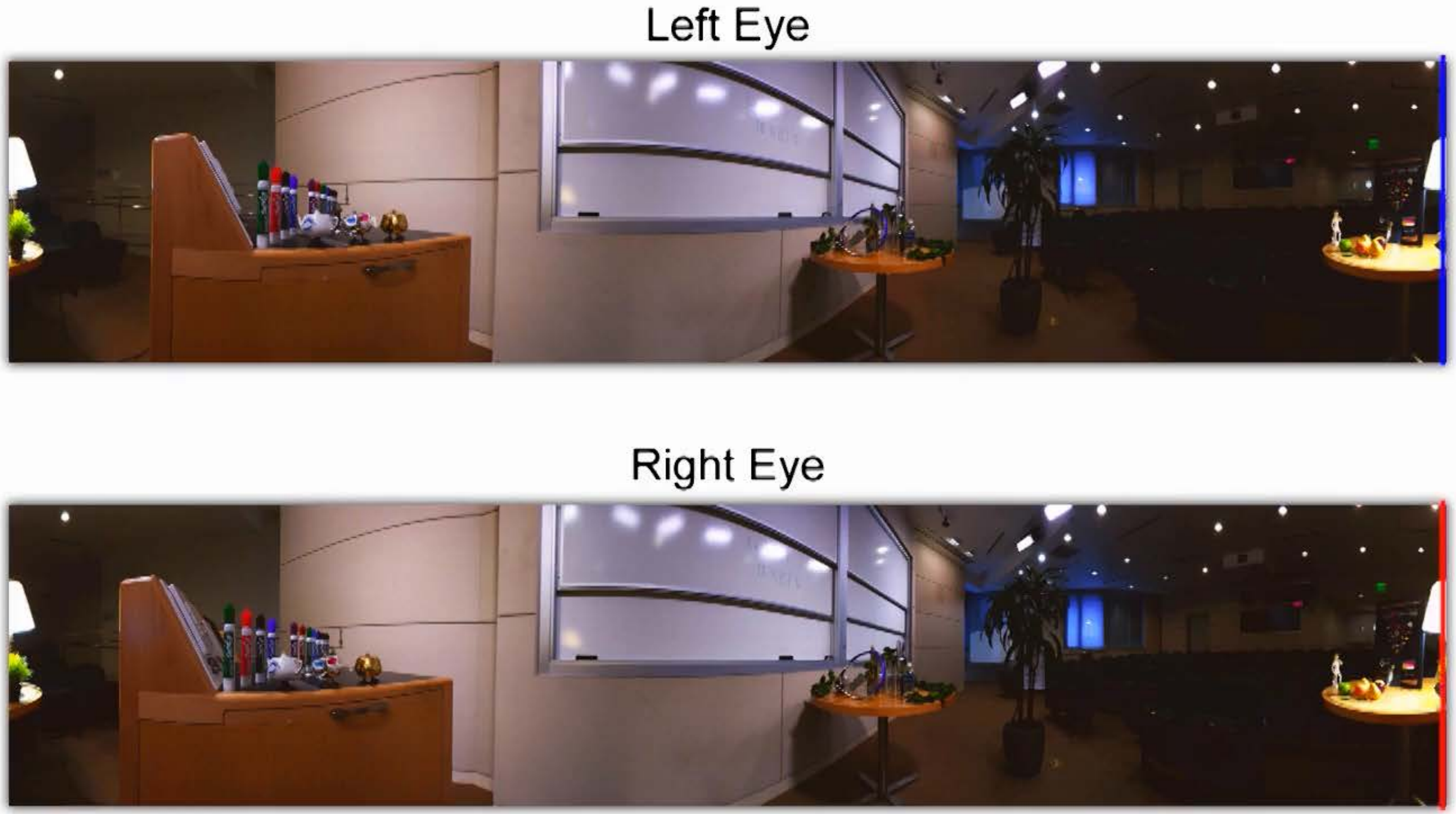
Live Streaming Camera Arrays

Single Shot ODS capture

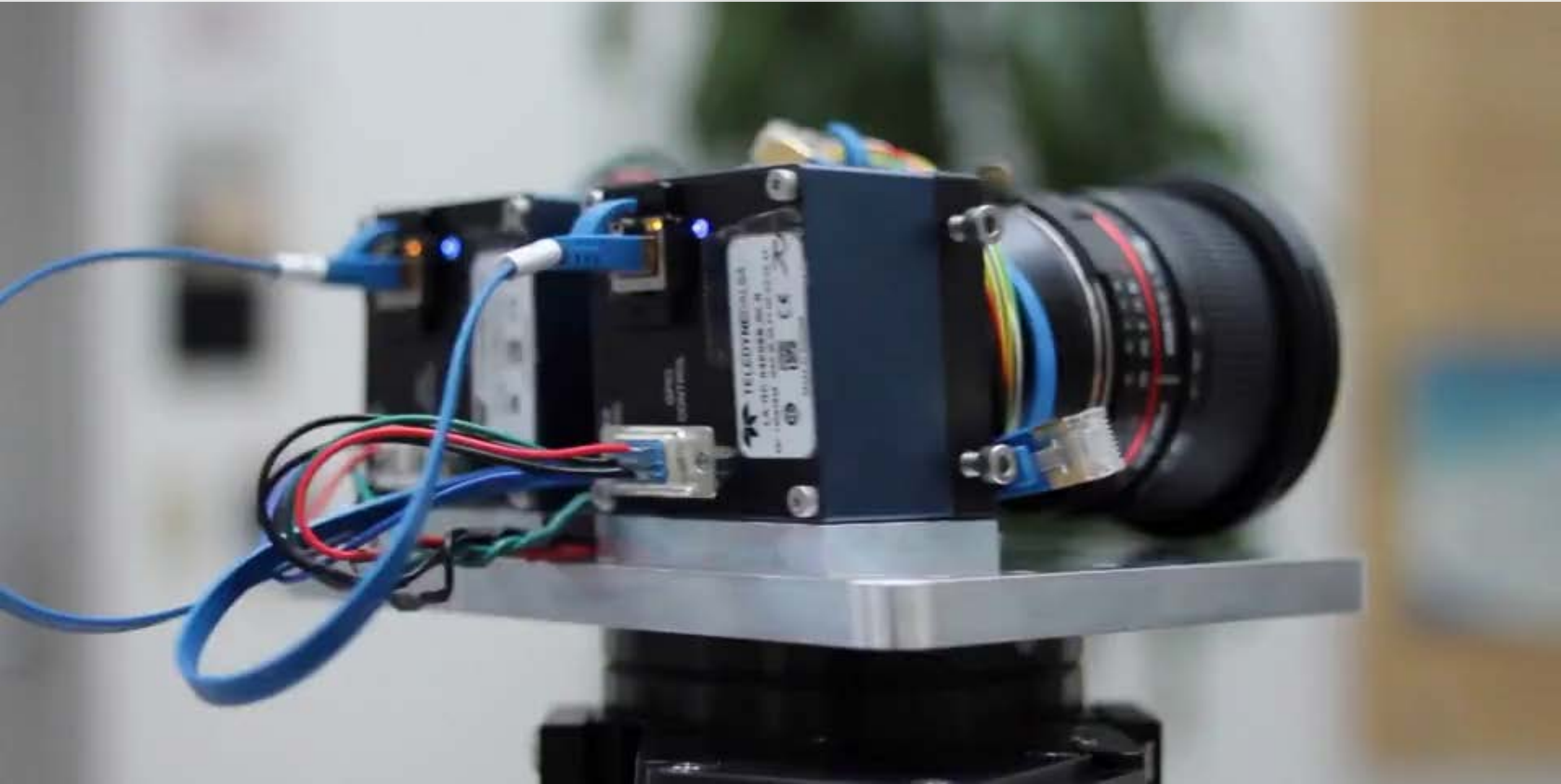
Rotating ODS capture

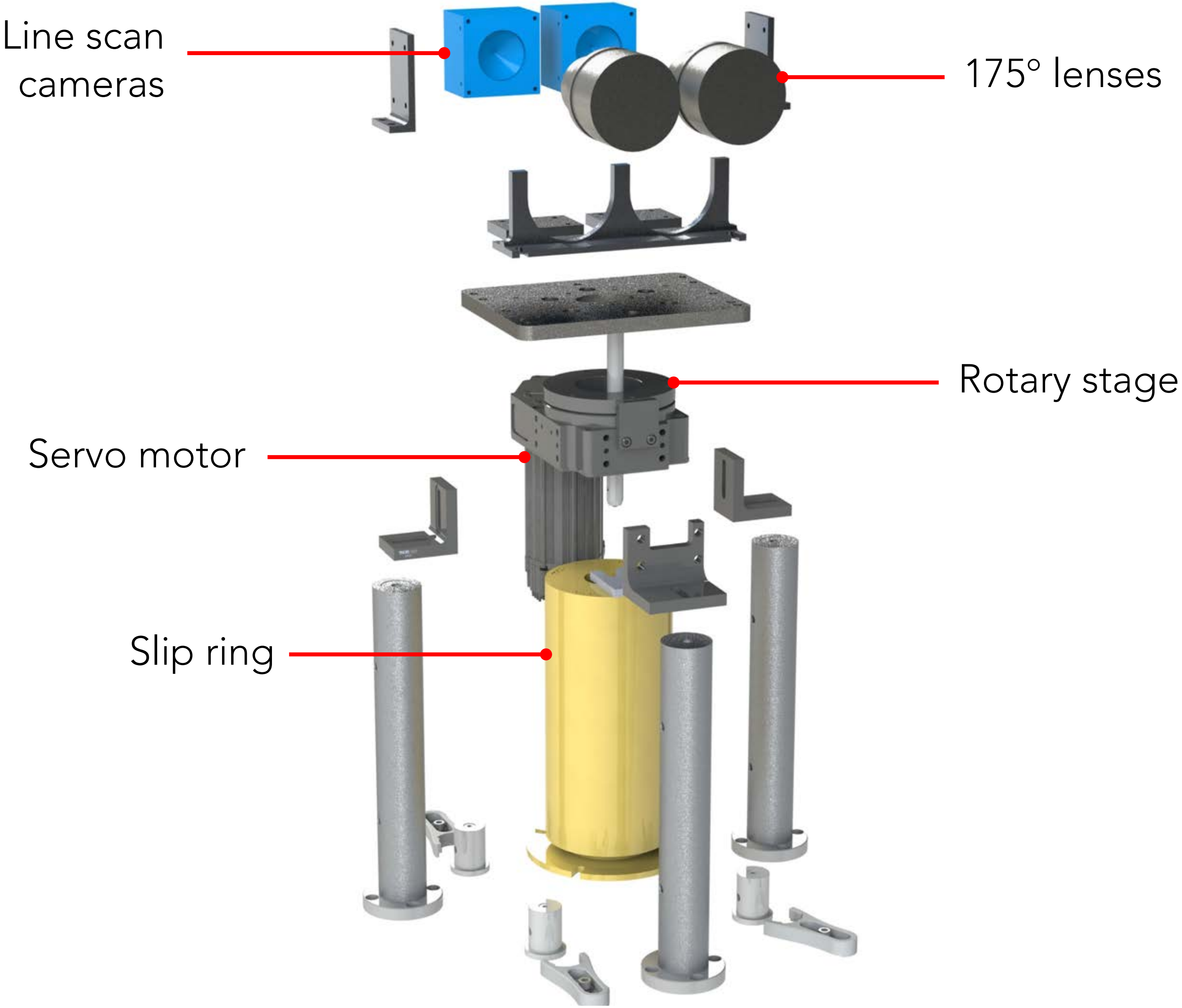
Omnidirectional Stereo (ODS)



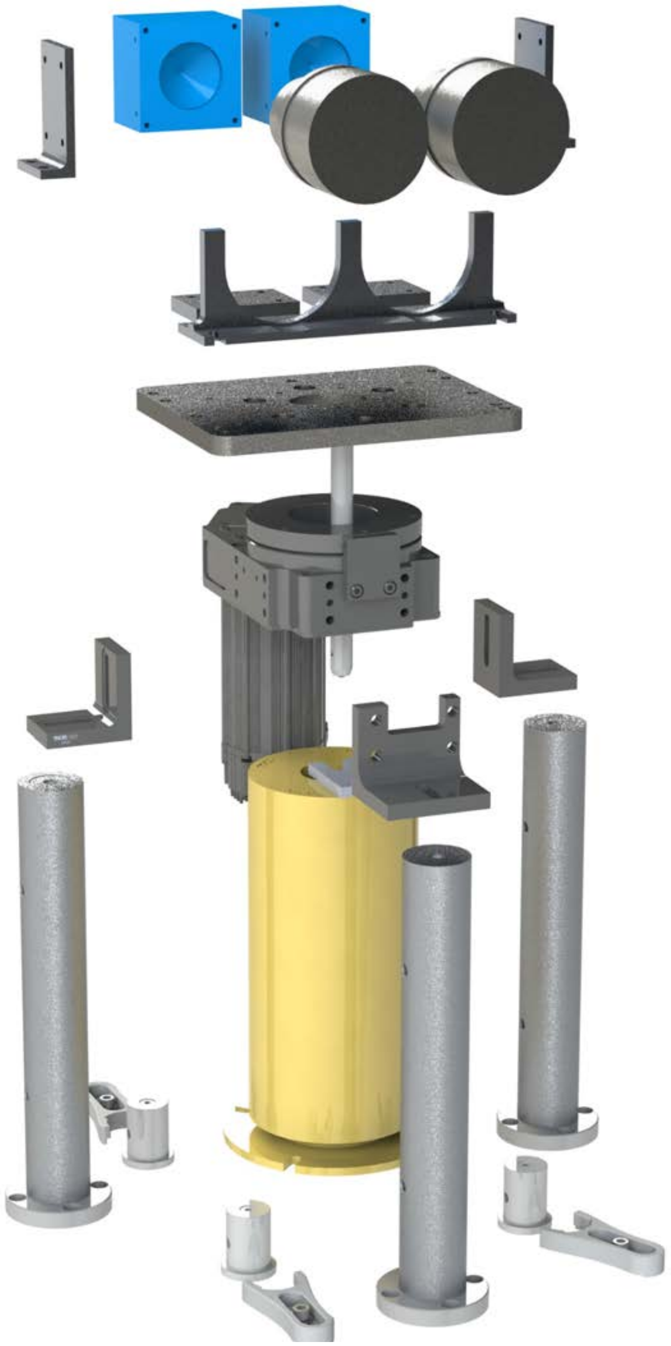


SpinVR




Line scan
cameras
175° lenses
Rotary stage
Servo motor
Slip ring

[Konrad 2017]



Pipeline Comparison

Minimal capture BW

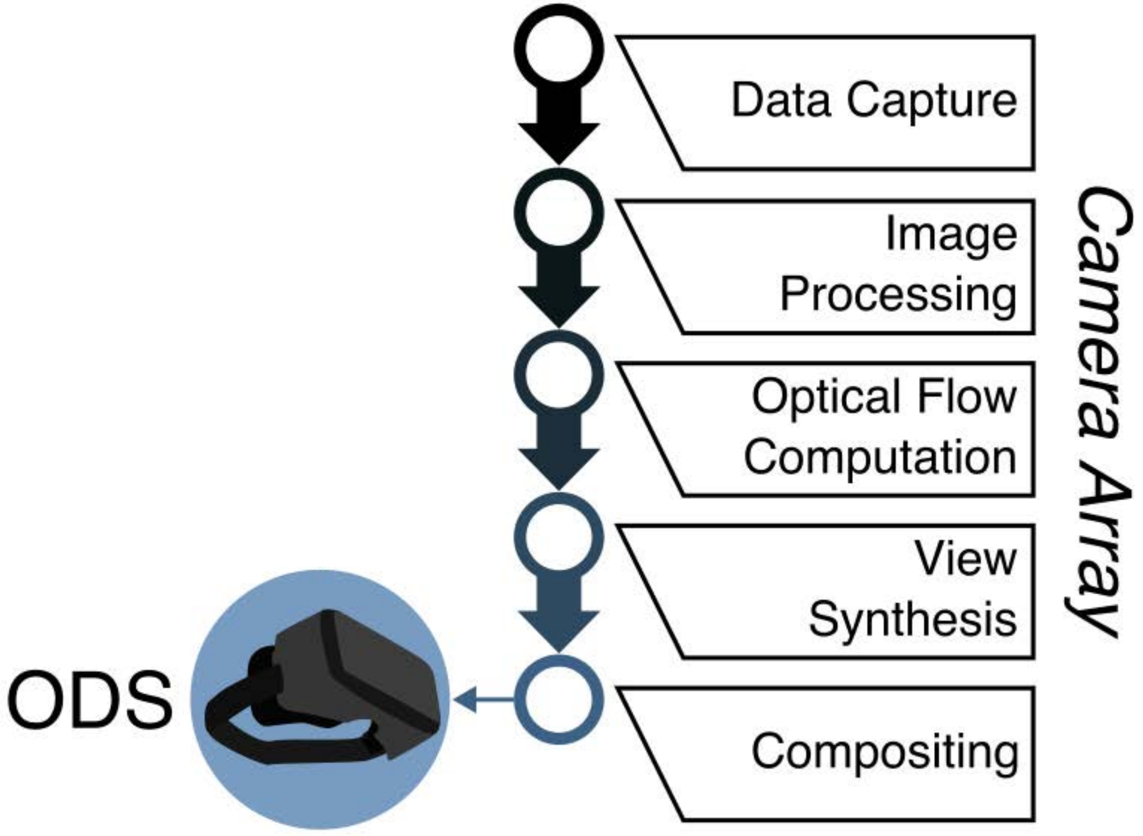
Minimal compute

→ Solves artifact issues*

→ No calibration*

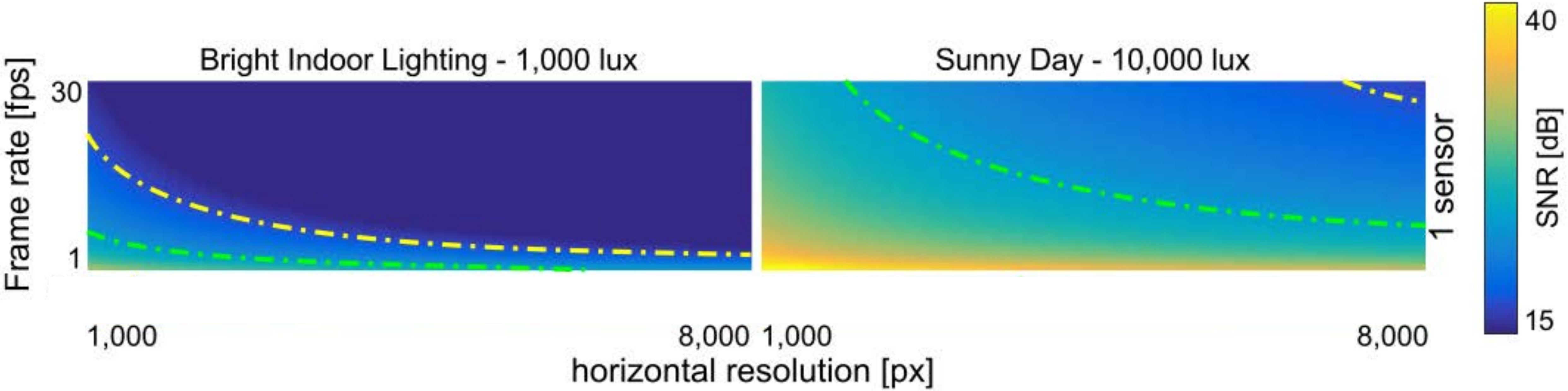
*Some new challenges

*New design space



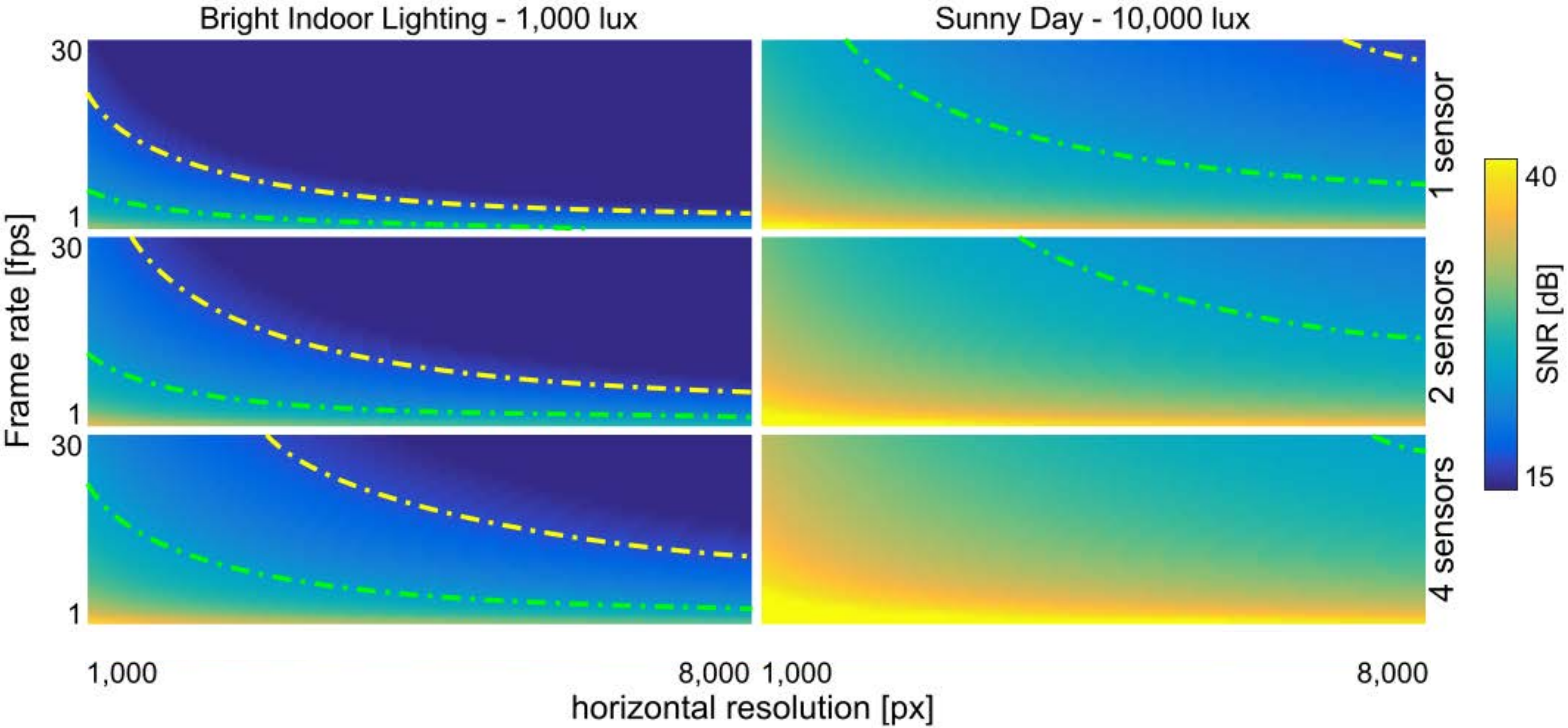
[Konrad 2017]

Design Trade-offs



[Konrad 2017]

Design Trade-offs



[Konrad 2017]

8192 x 4096 x 1/26 fps

4096 x 4096 x 1.11 fps

2048 x 4096 x 16.67 fps

8192 x 4096 x 1/26 fps

4096 x 4096 x 1.11 fps

2048 x 4096 x 16.67 fps


Live Streaming

Challenging Scenes

Reflections

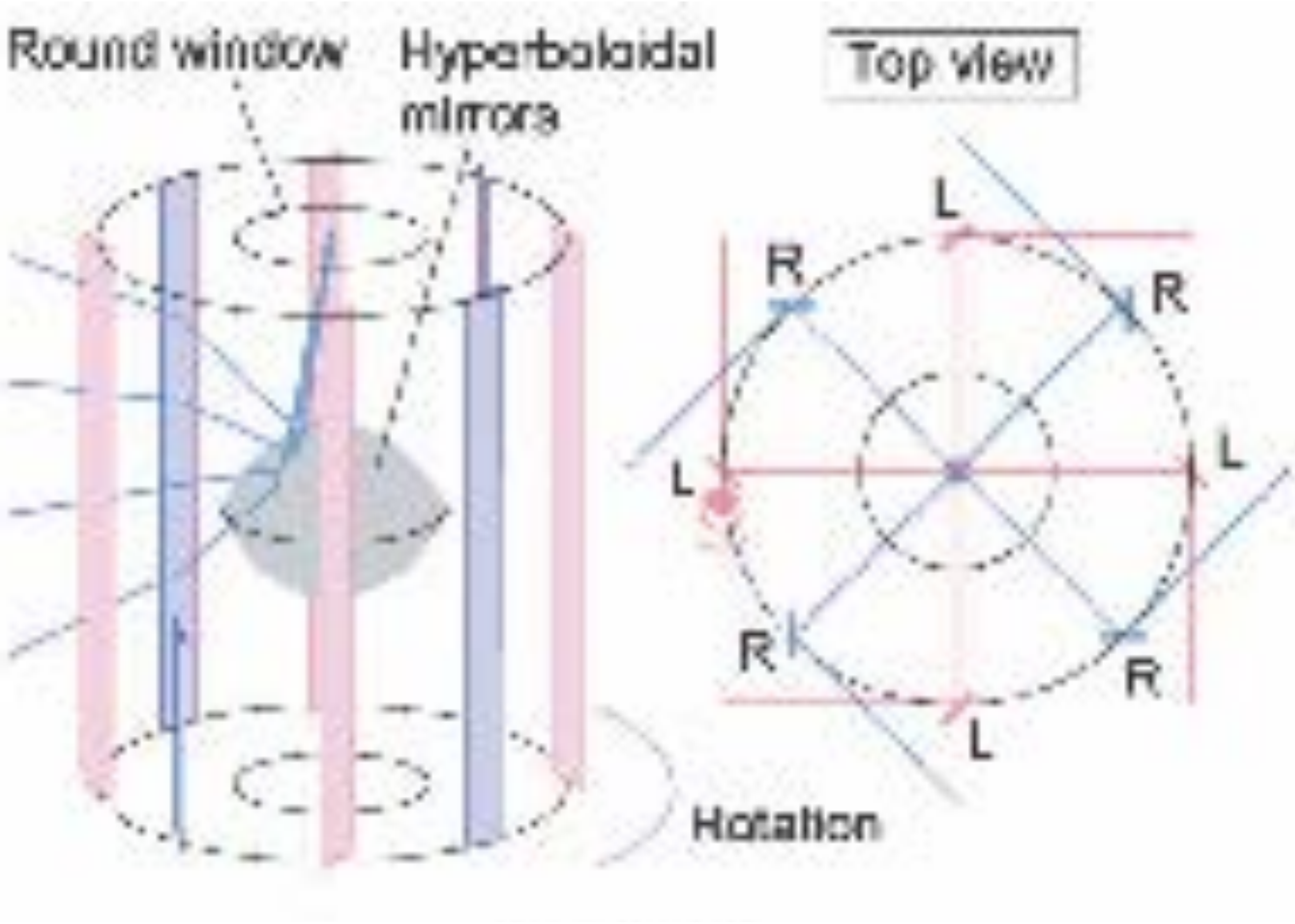
Transparency

Occlusions

Fine structure

[Konrad 2017]

TORNADO

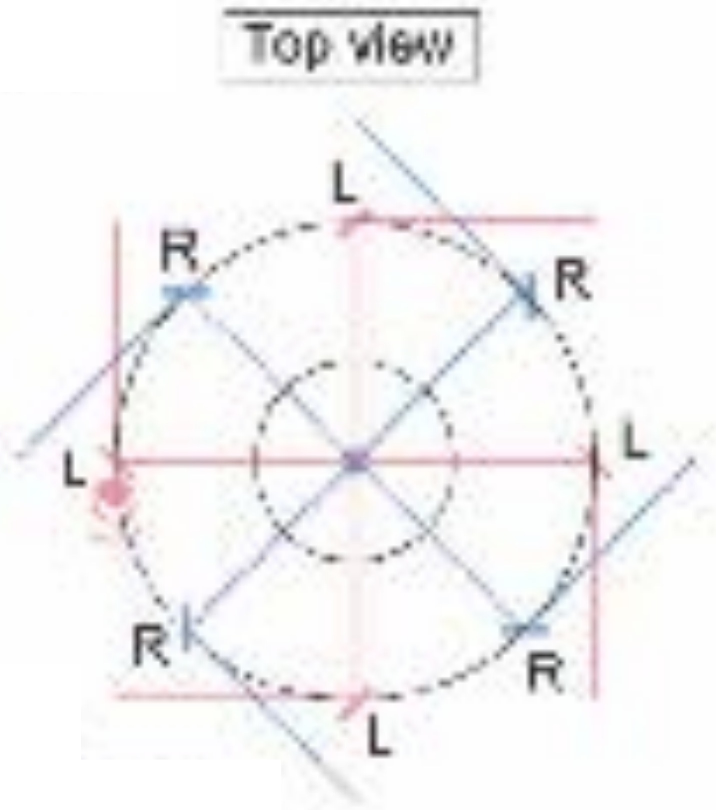


[Tanaka 2005]

Live ODS Video (is hard)

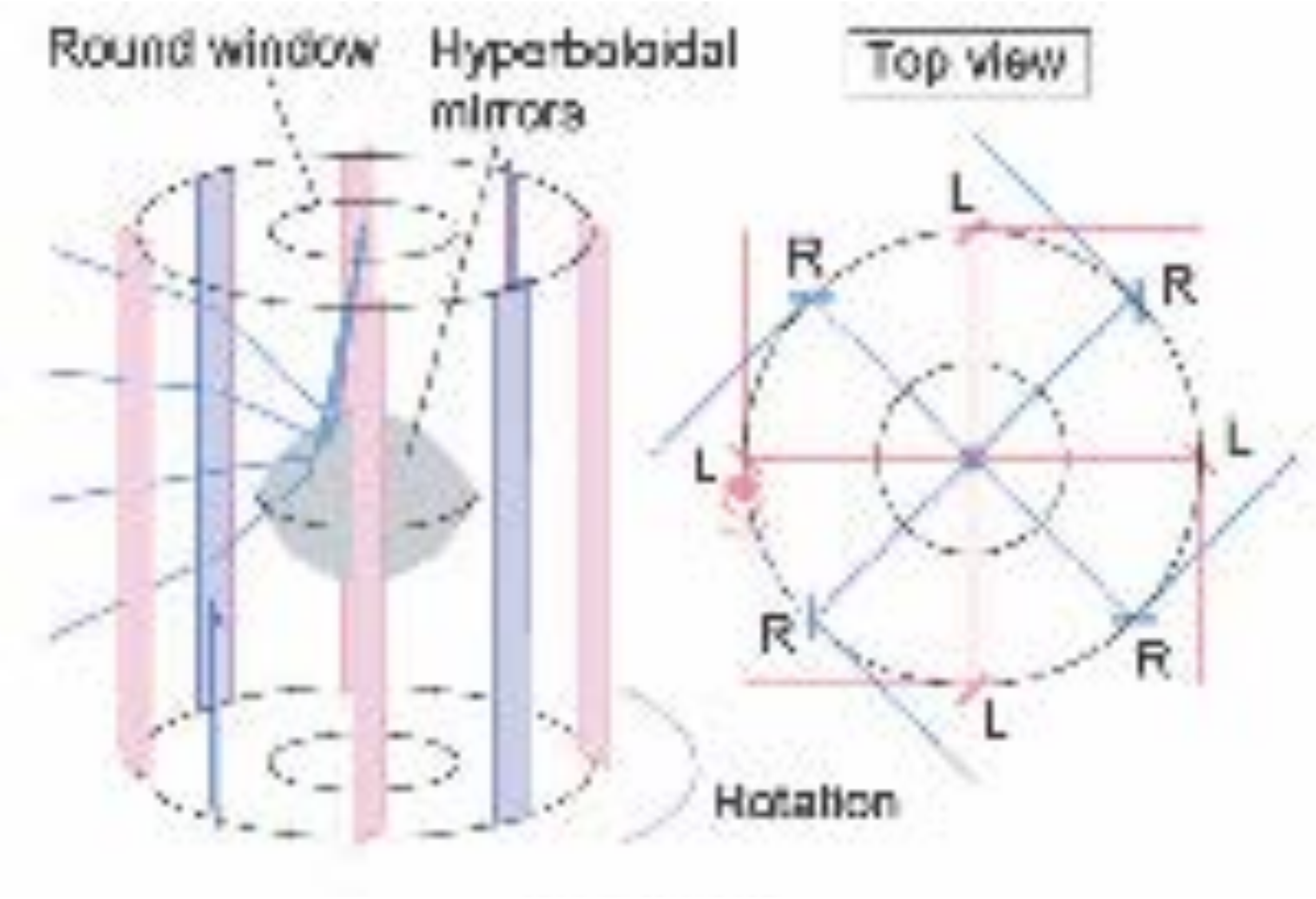
Robert Konrad

TORNADO



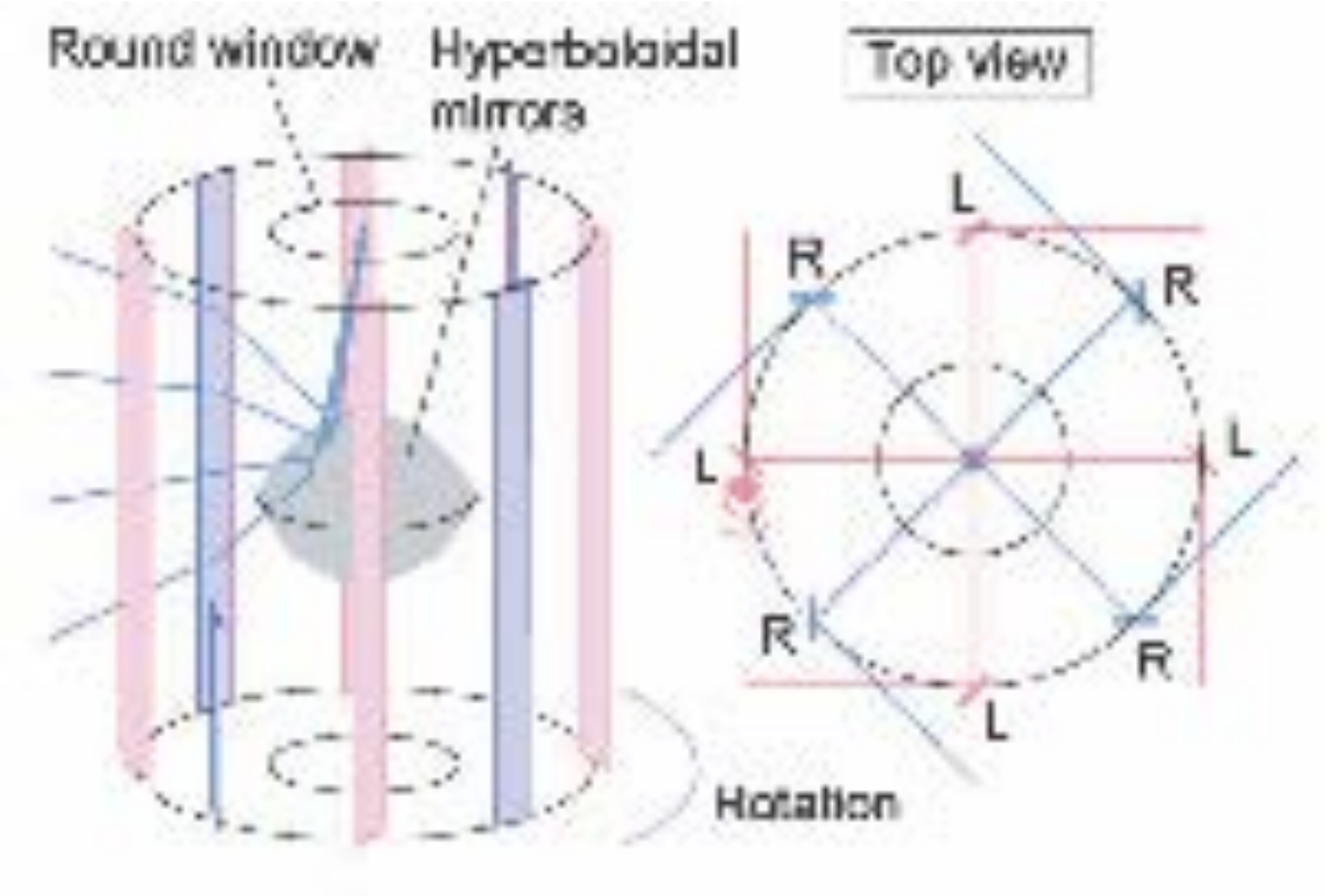
[Tanaka 2005]

TORNADO

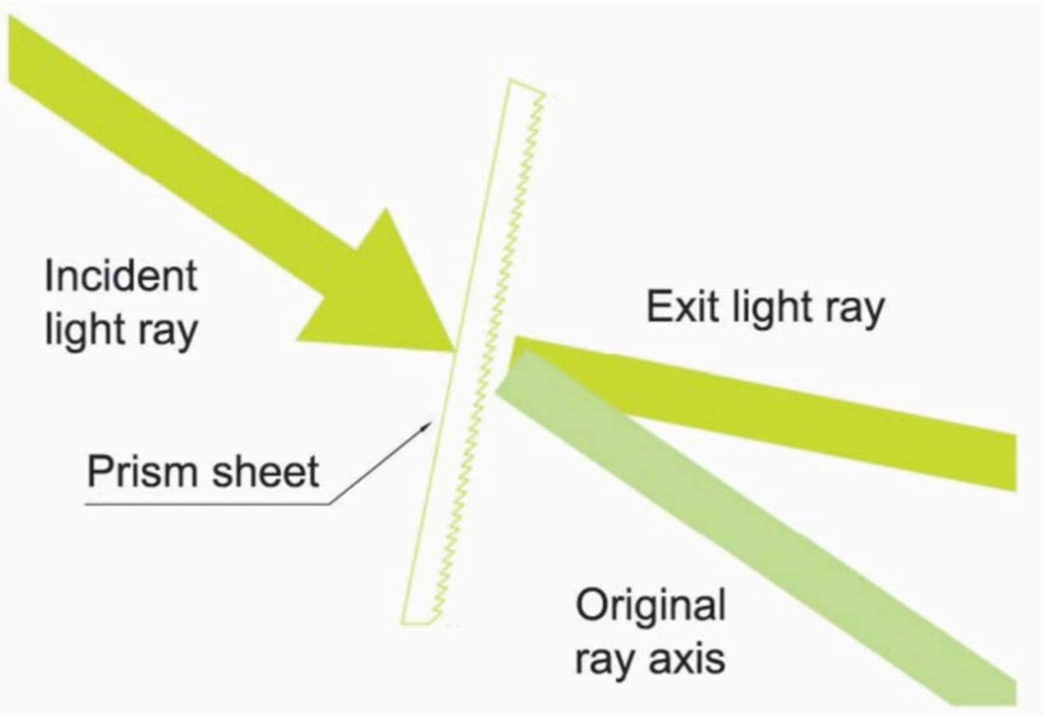


[Tanaka 2005]

TORNADO

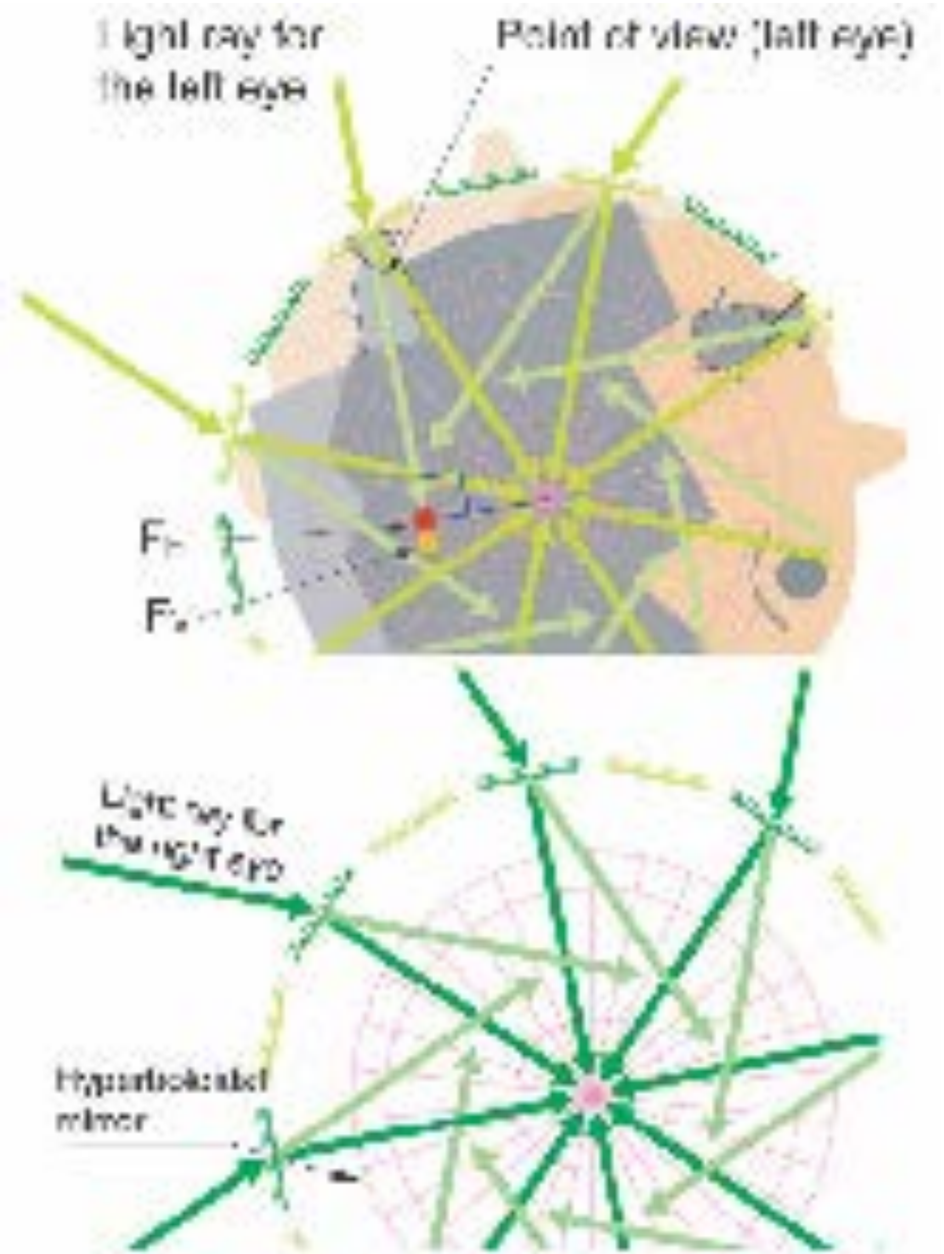


Prism Sheet Working Principle



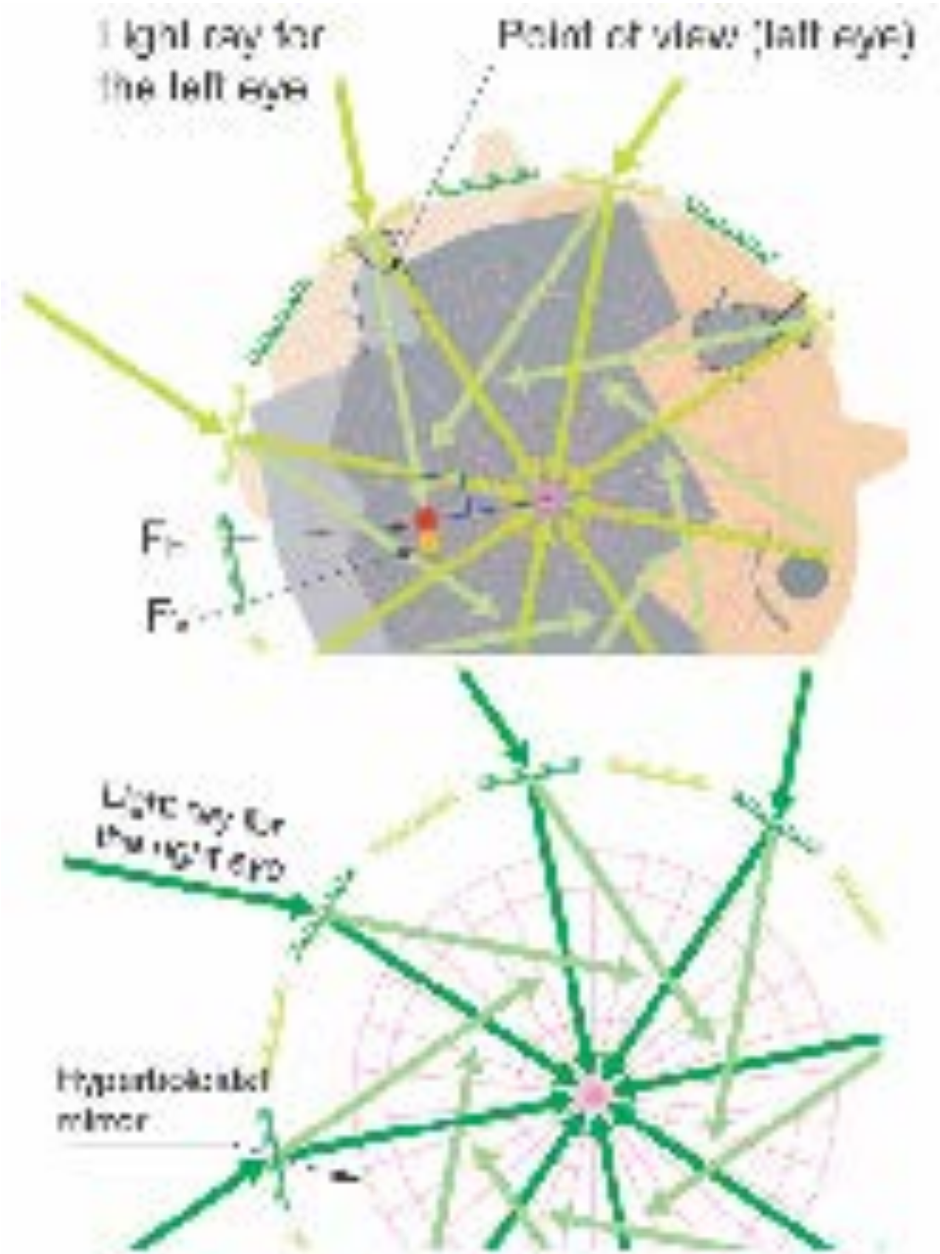
[Tanaka 2005]

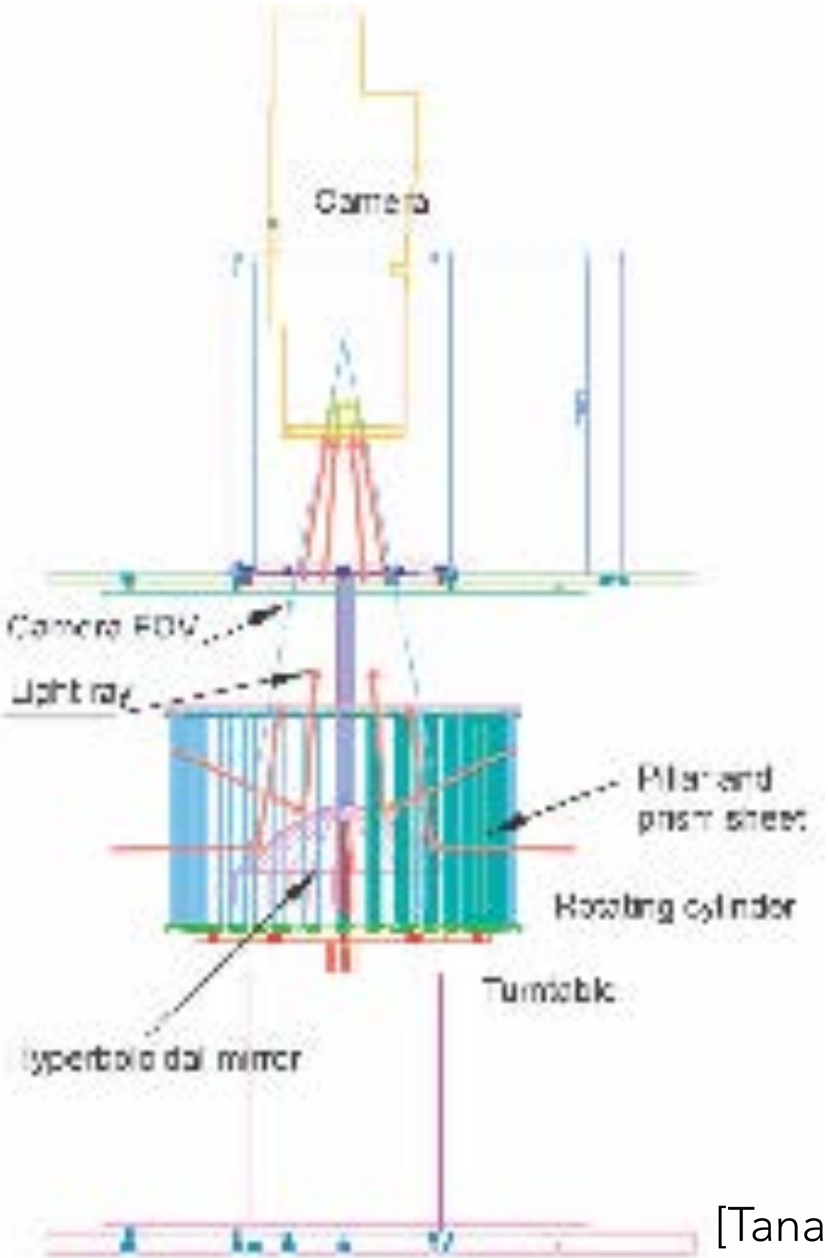
TORNADO



[Tanaka 2005]

TORNADO



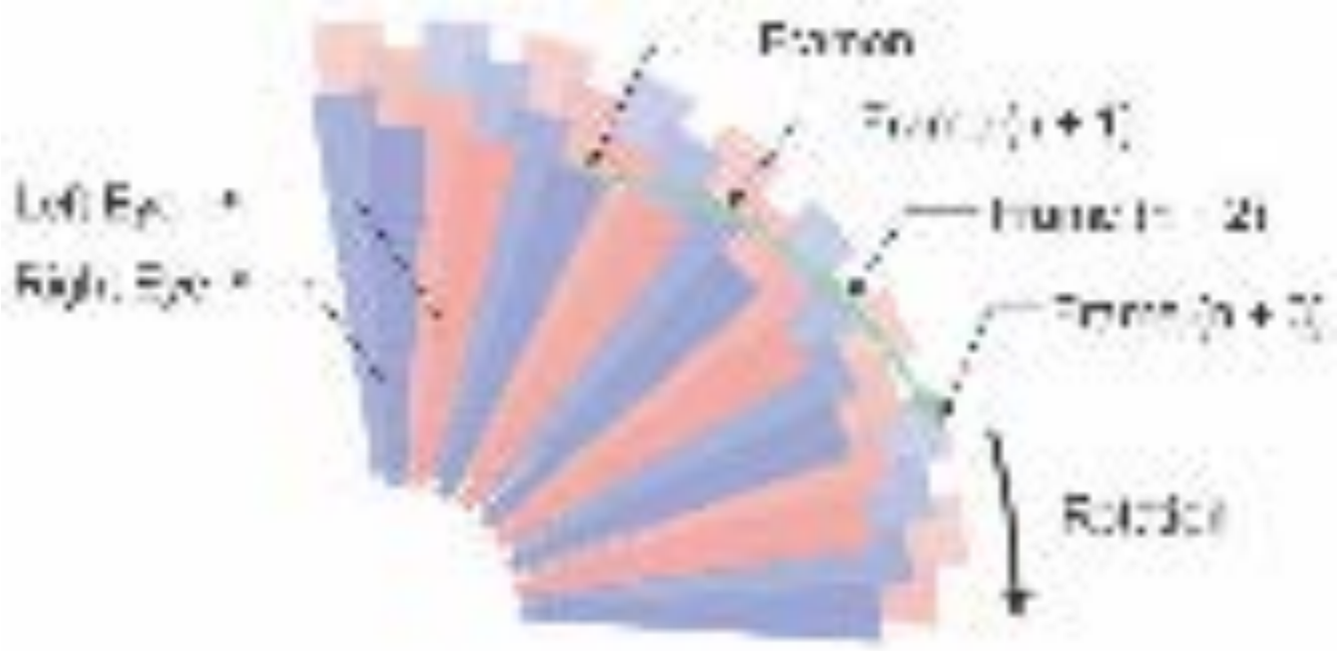


[Tanaka 2005]

TORNADO

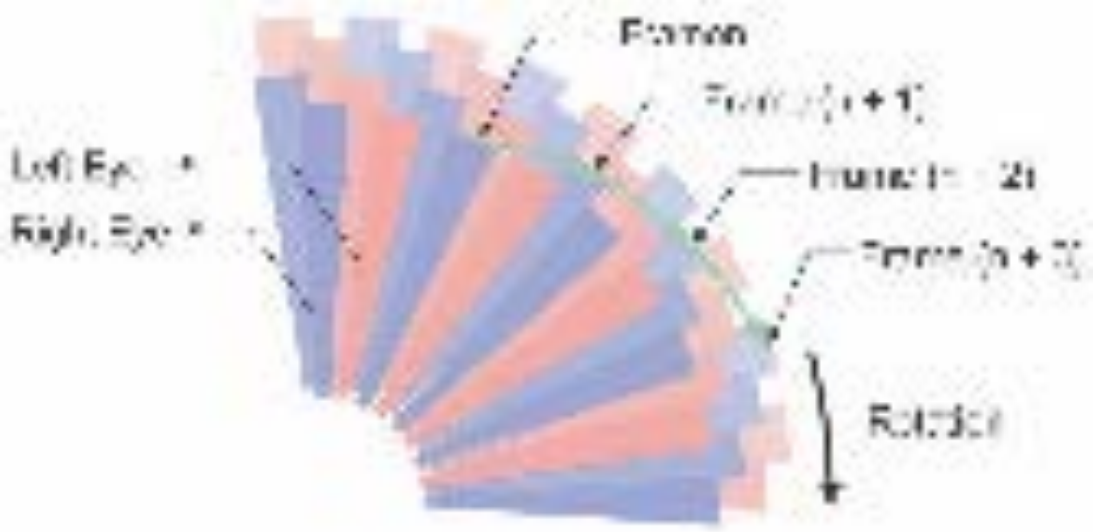
[Tanaka 2005]

TORNADO

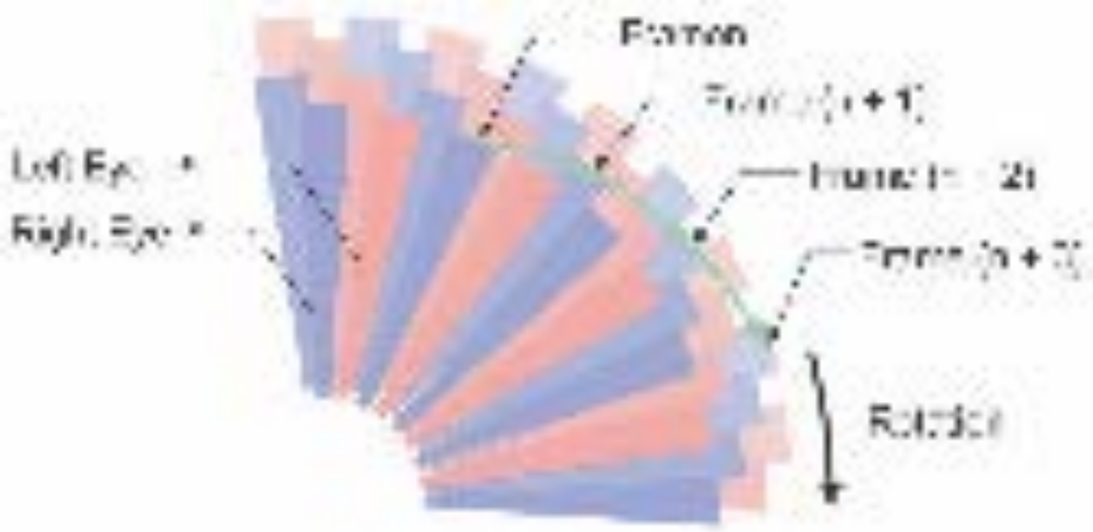


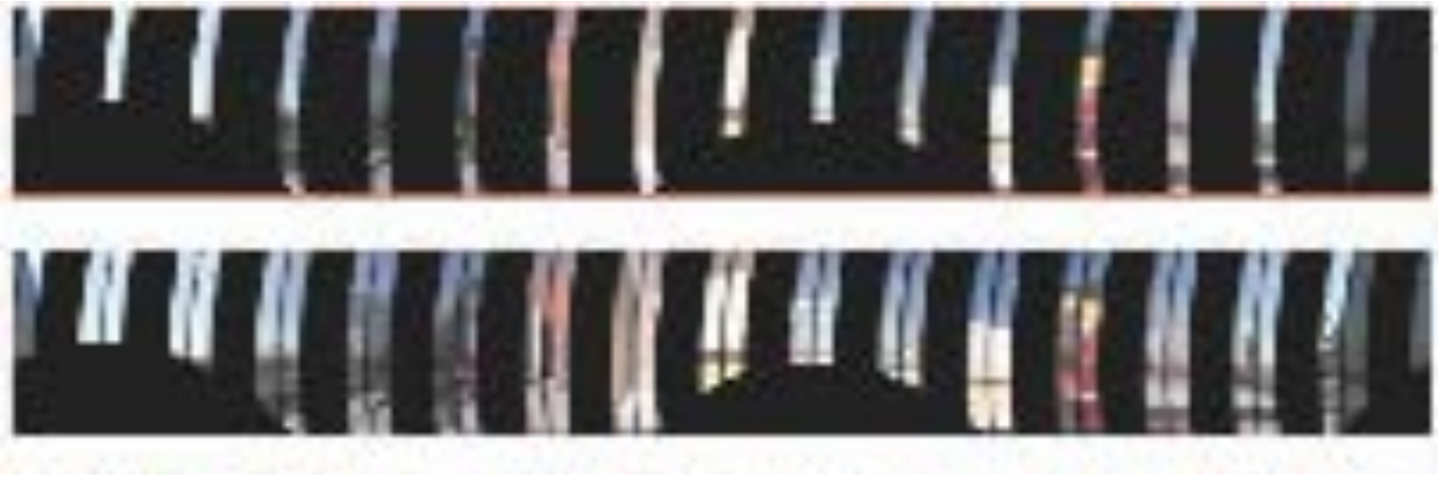
[Tanaka 2005]

TORNADO

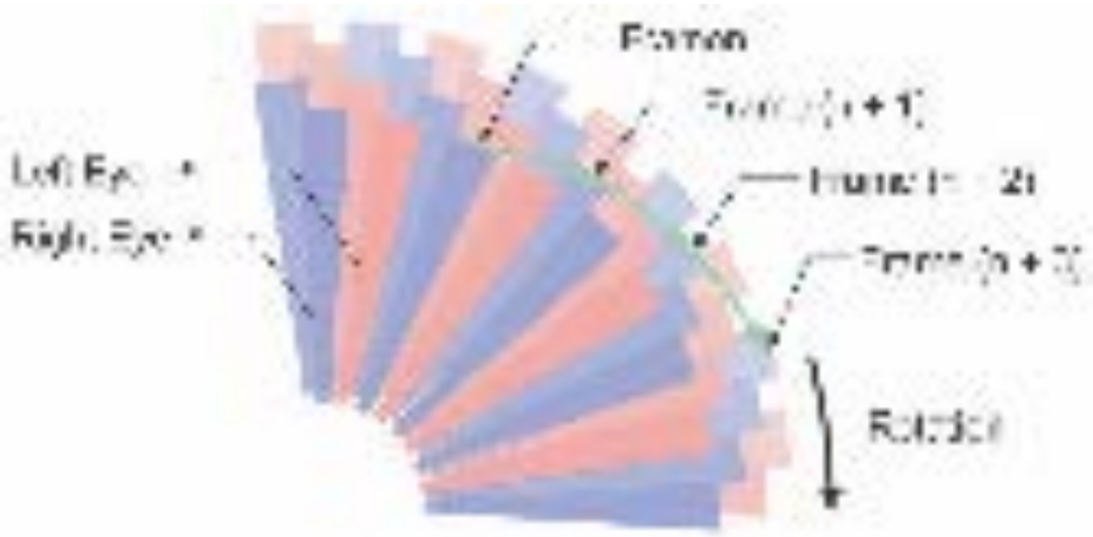


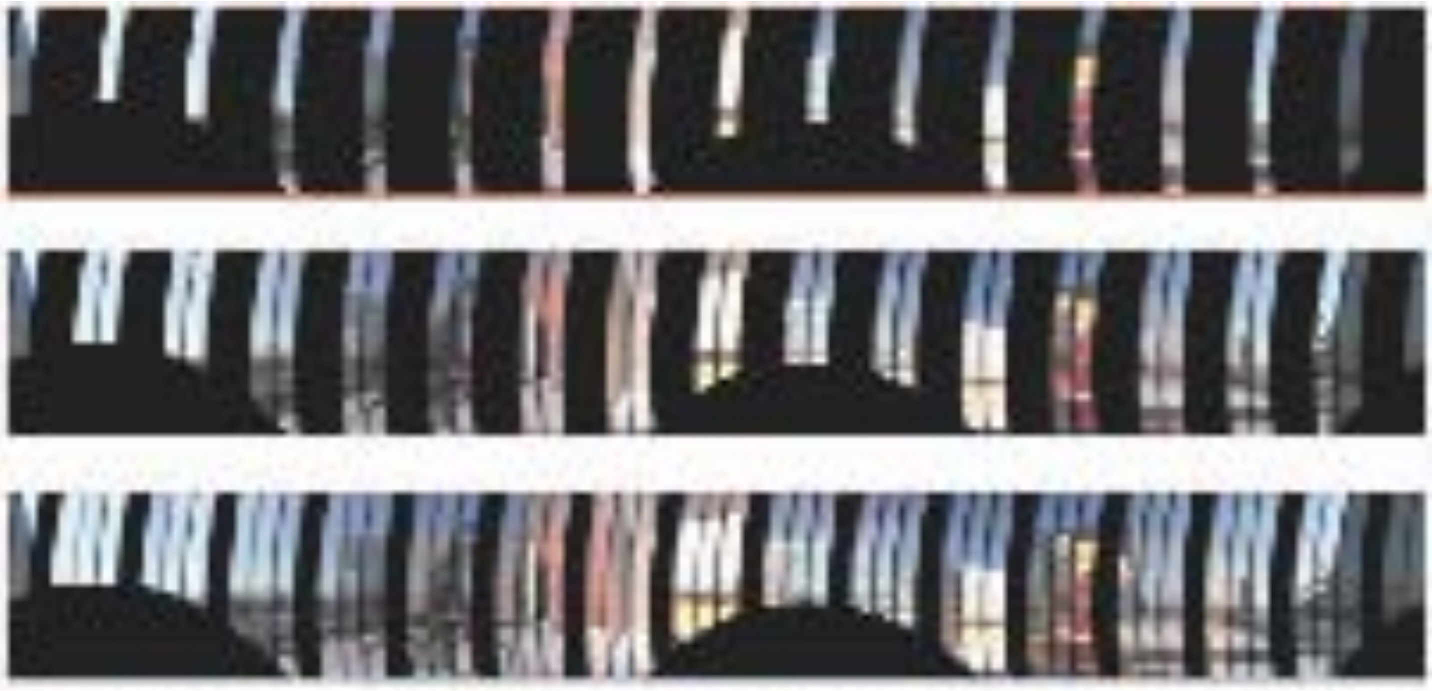
TORNADO



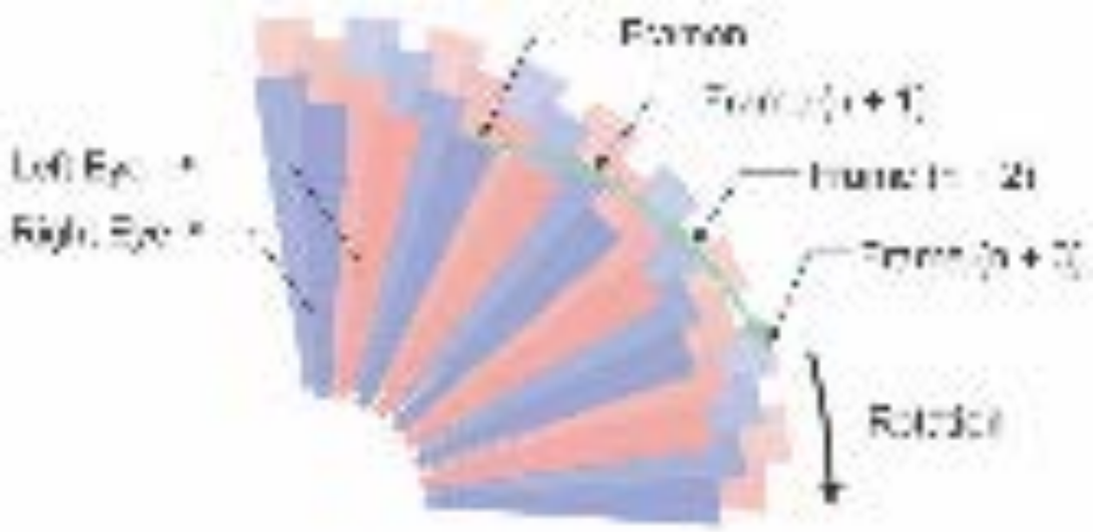


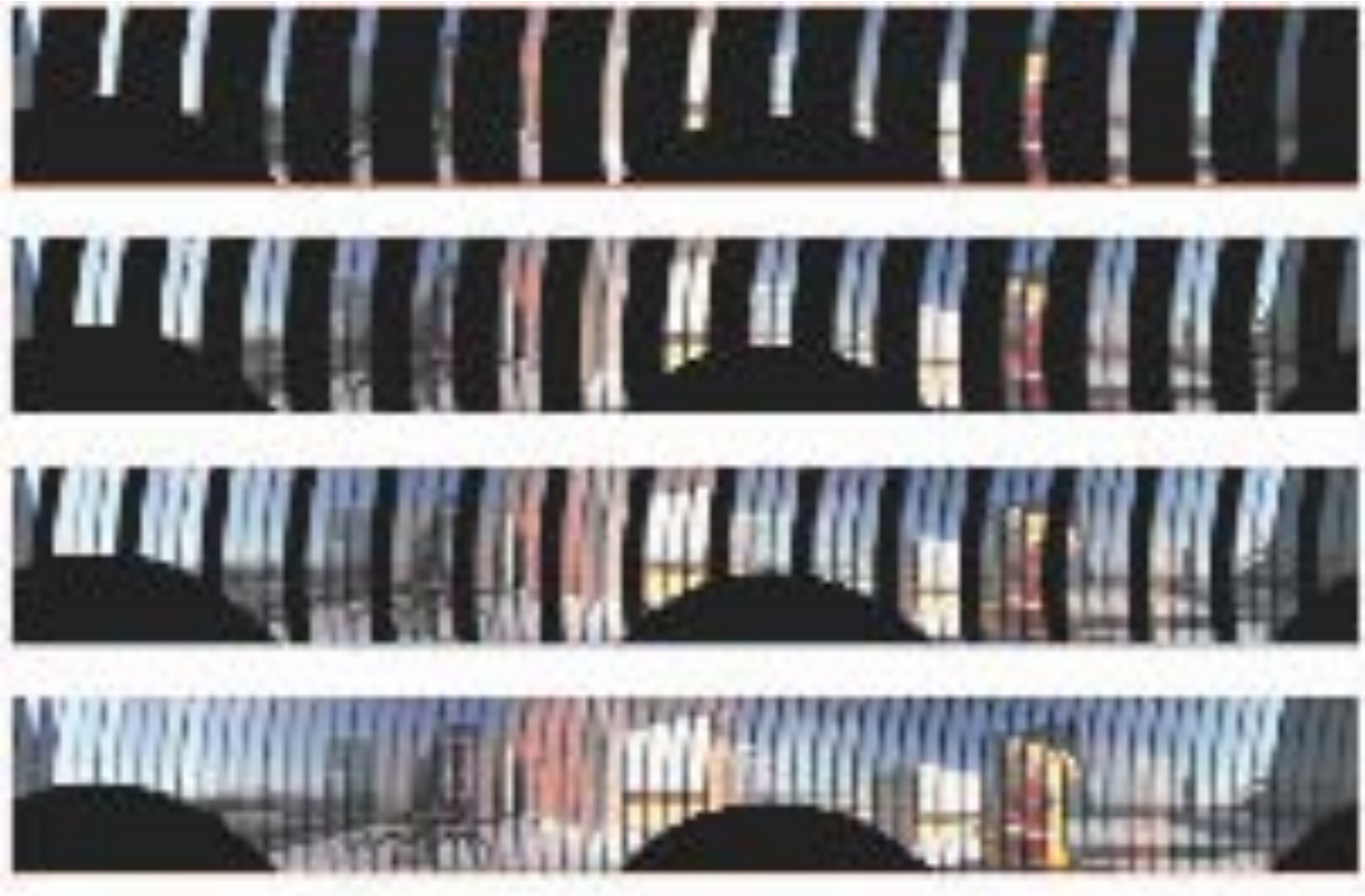
TORNADO



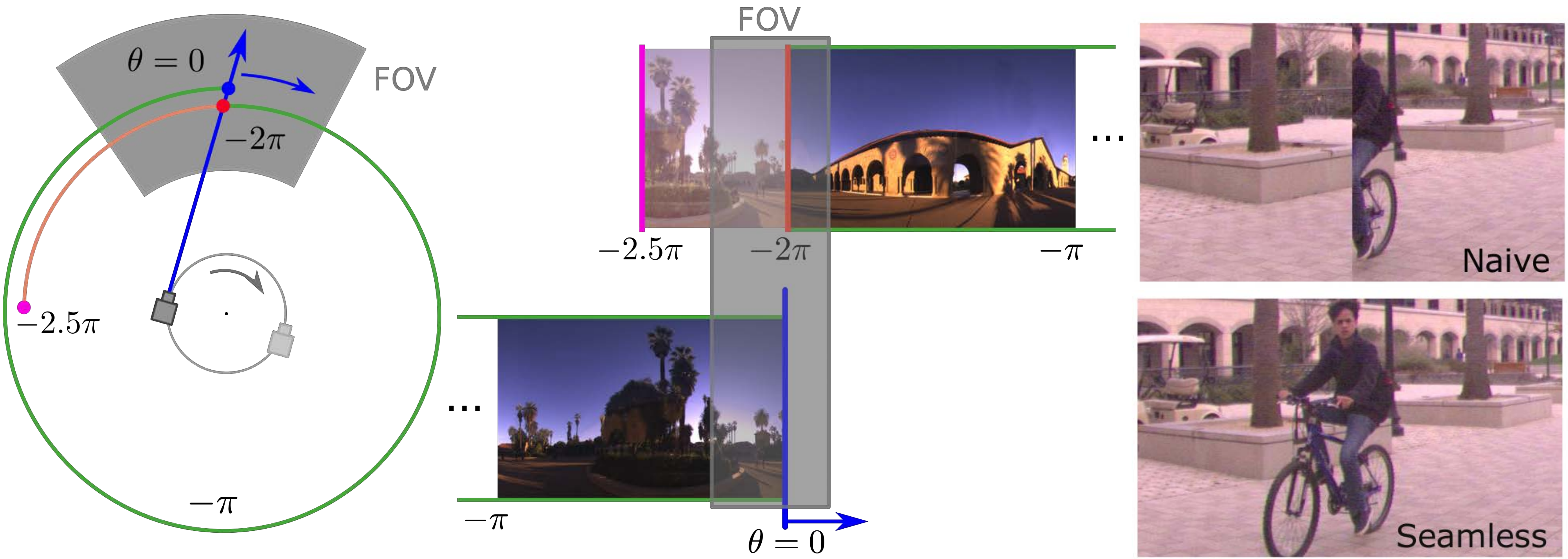


TORNADO

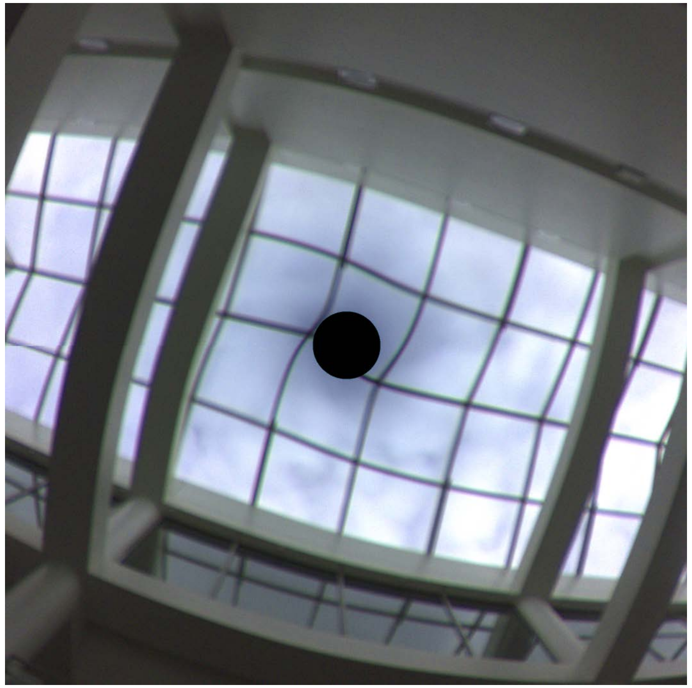


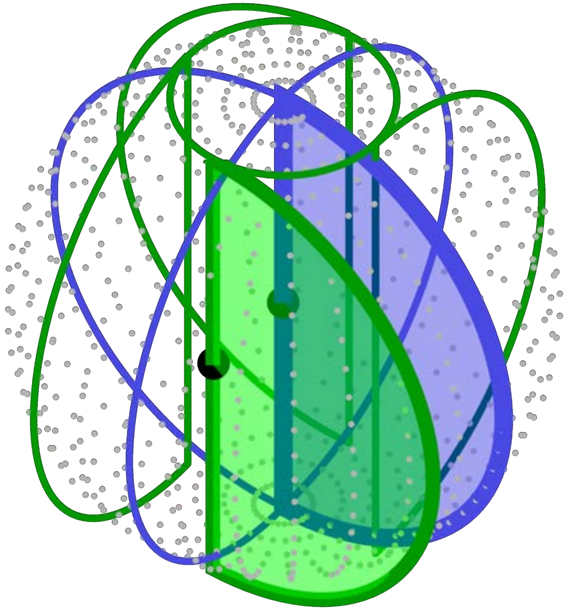


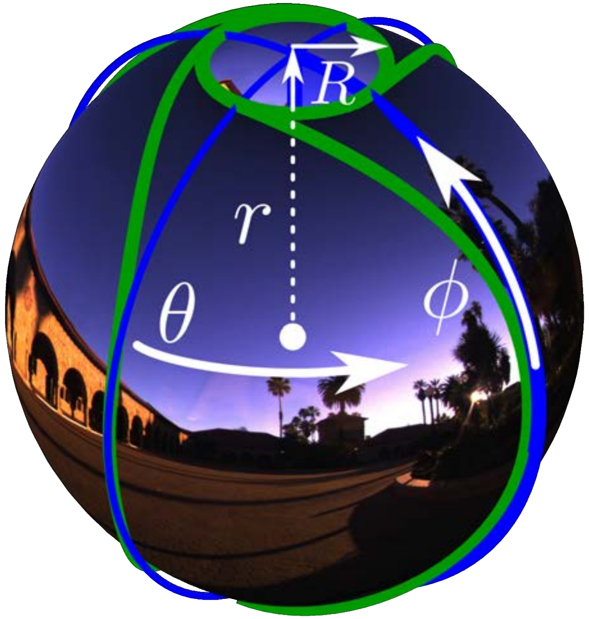
Seamless Rendering



Warping at the Poles







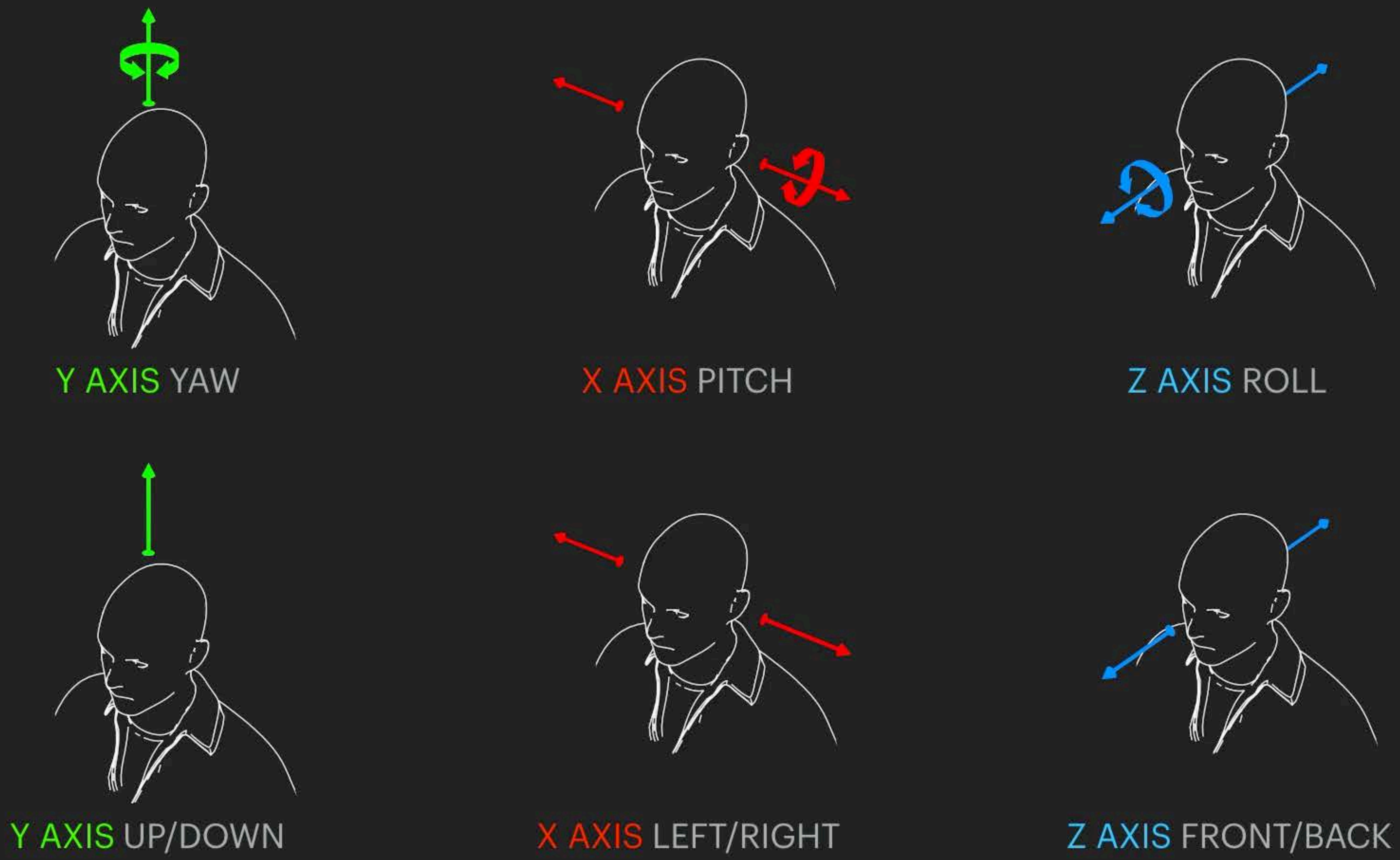
$$\Delta\theta = \tan^{-1}\left(\frac{R/r}{\cos\phi}\right)$$

$$\Delta\phi = \tan^{-1}\left(\frac{\sin\phi}{\sqrt{(R/r)^2 + \cos^2\phi}}\right) - \phi$$

Warping at the Poles

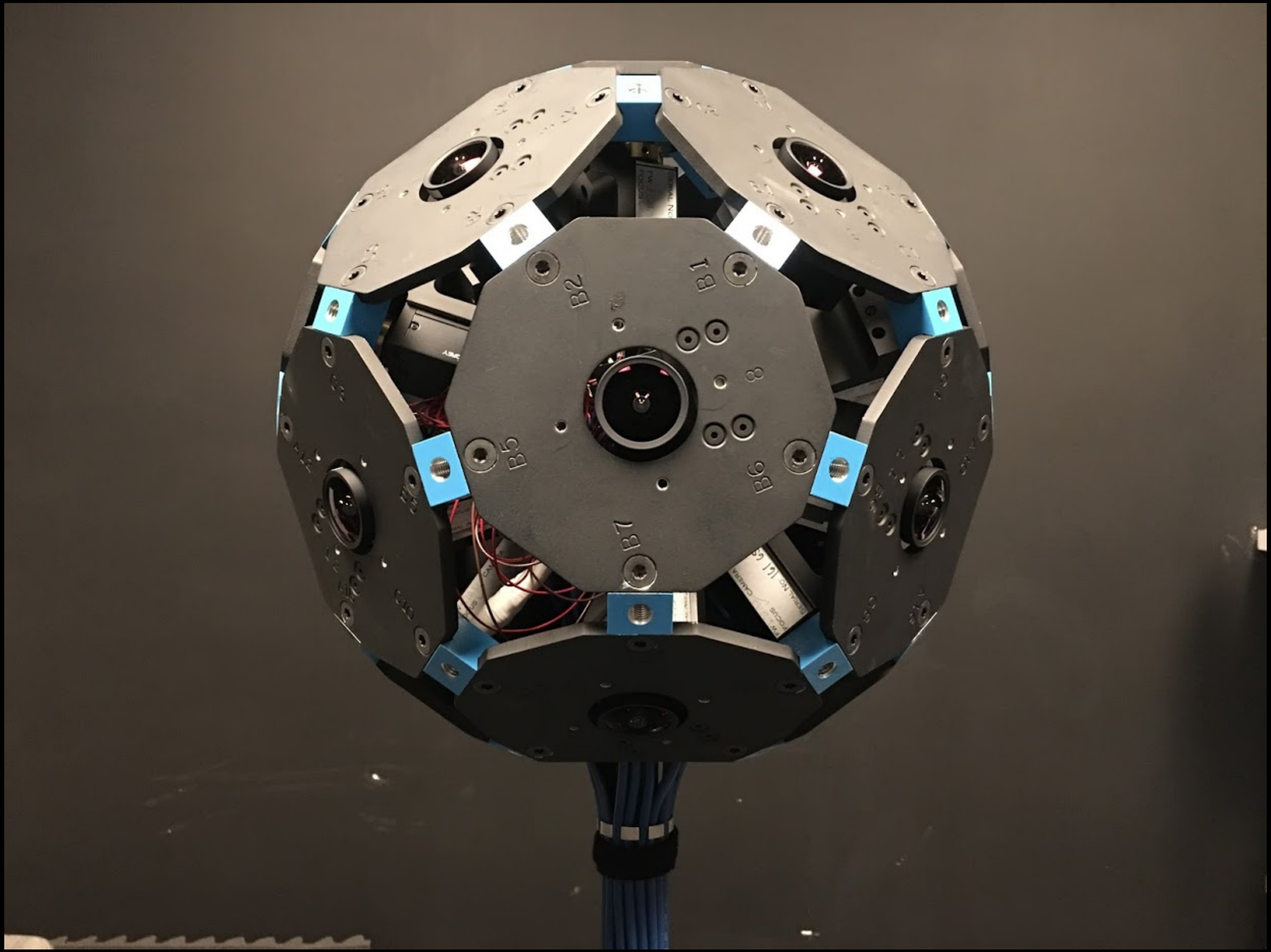
Six Degrees of Freedom Video

Brian Cabral
Facebook


Y AXIS YAW
X AXIS PITCH
Z AXIS ROLL
Y AXIS UP/DOWN
X AXIS LEFT/RIGHT
Z AXIS FRONT/BACK

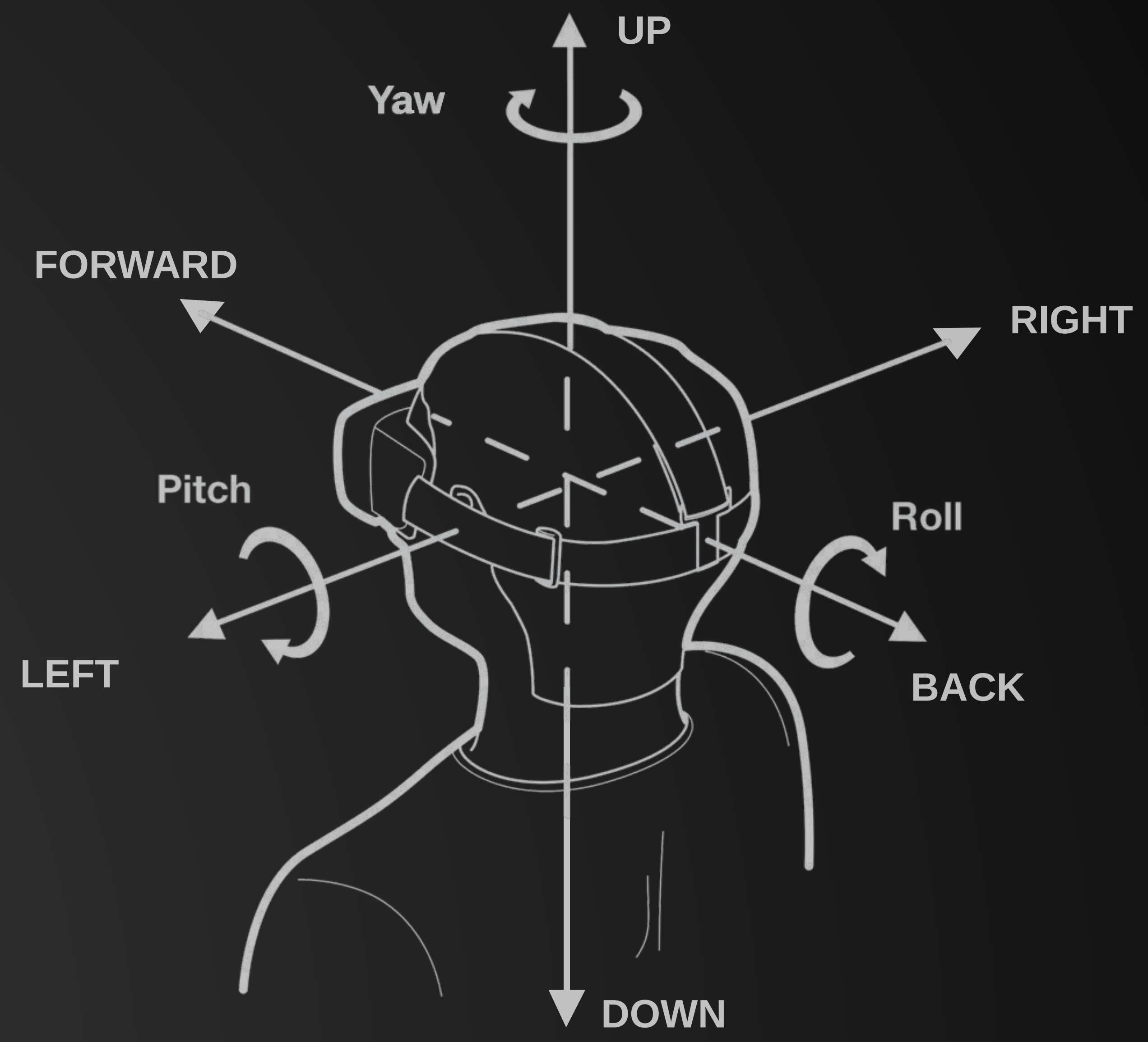
This can't do 6DoF - why not?

This rig can do 6DoF - why?



6DoF vs ODS

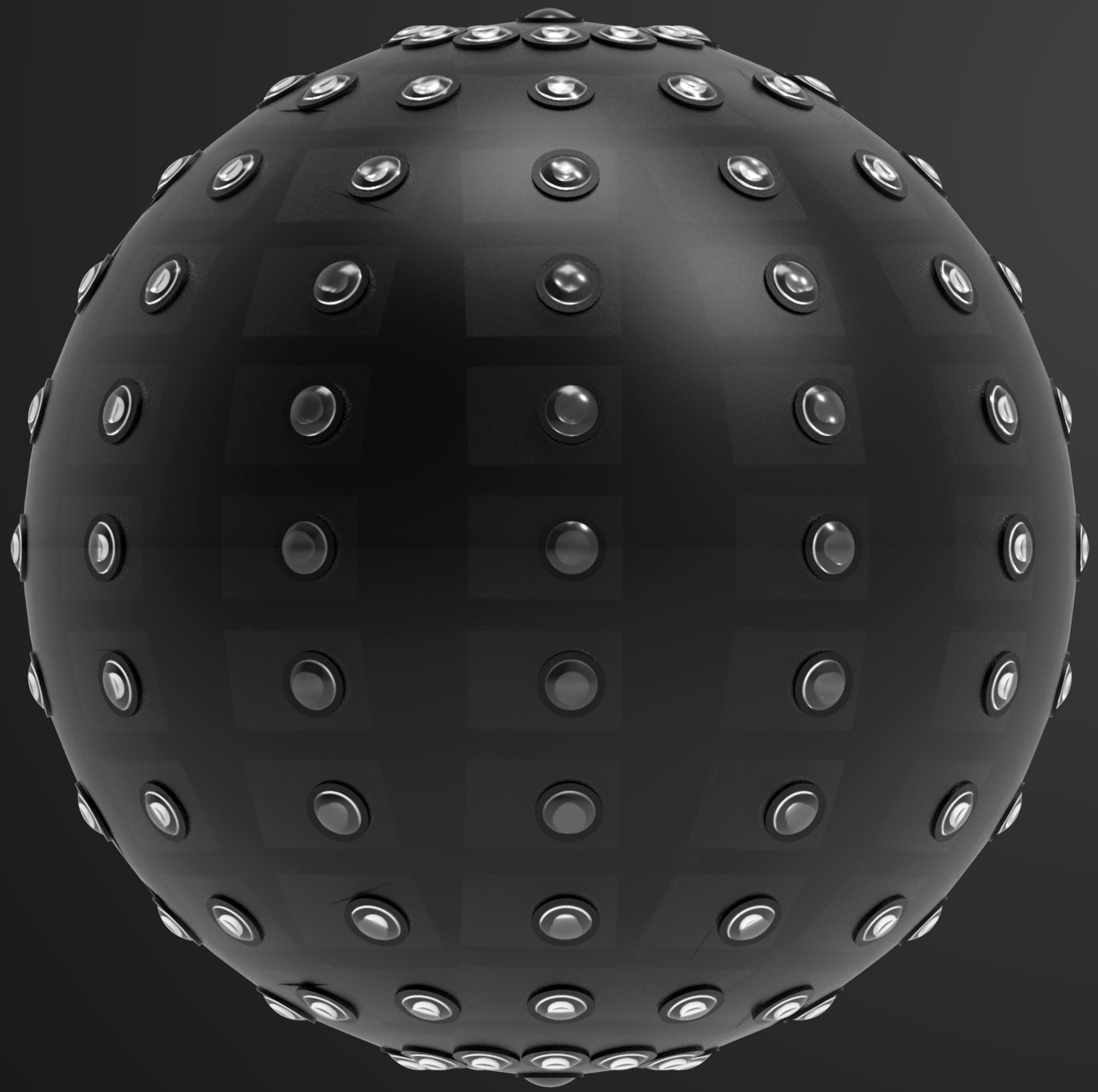
The 6-DOF + time Challenge

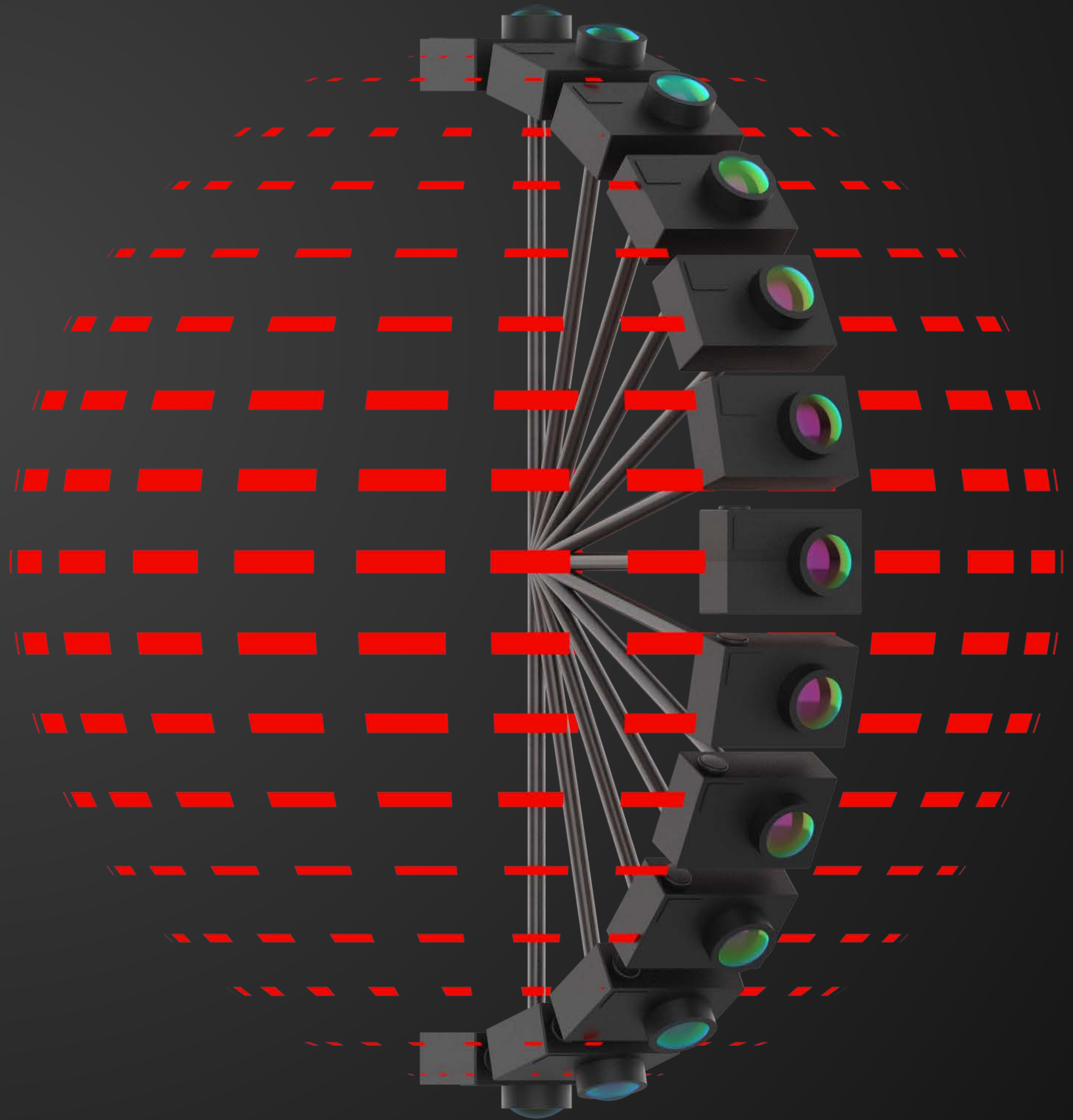


- You need to capture a moving depth map
- Several approaches
    - Trades off between "angular" resolution and spatial
    - Light Field cameras use many sub-apertures to gain more angular resolution
    - We use more spatial resolution by assuming band-limited BRDF's & occlusions

Spherical Lightfields

- A LOT, 100'S OF CAMERAS PLACED IN A SPHERE
- USE A SPINNING GANTRY OF ONE OR MORE CAMERAS
  - GOOD FOR STILL LIFES - STUNNING RESULTS
- HARD TO DO FOR VIDEO



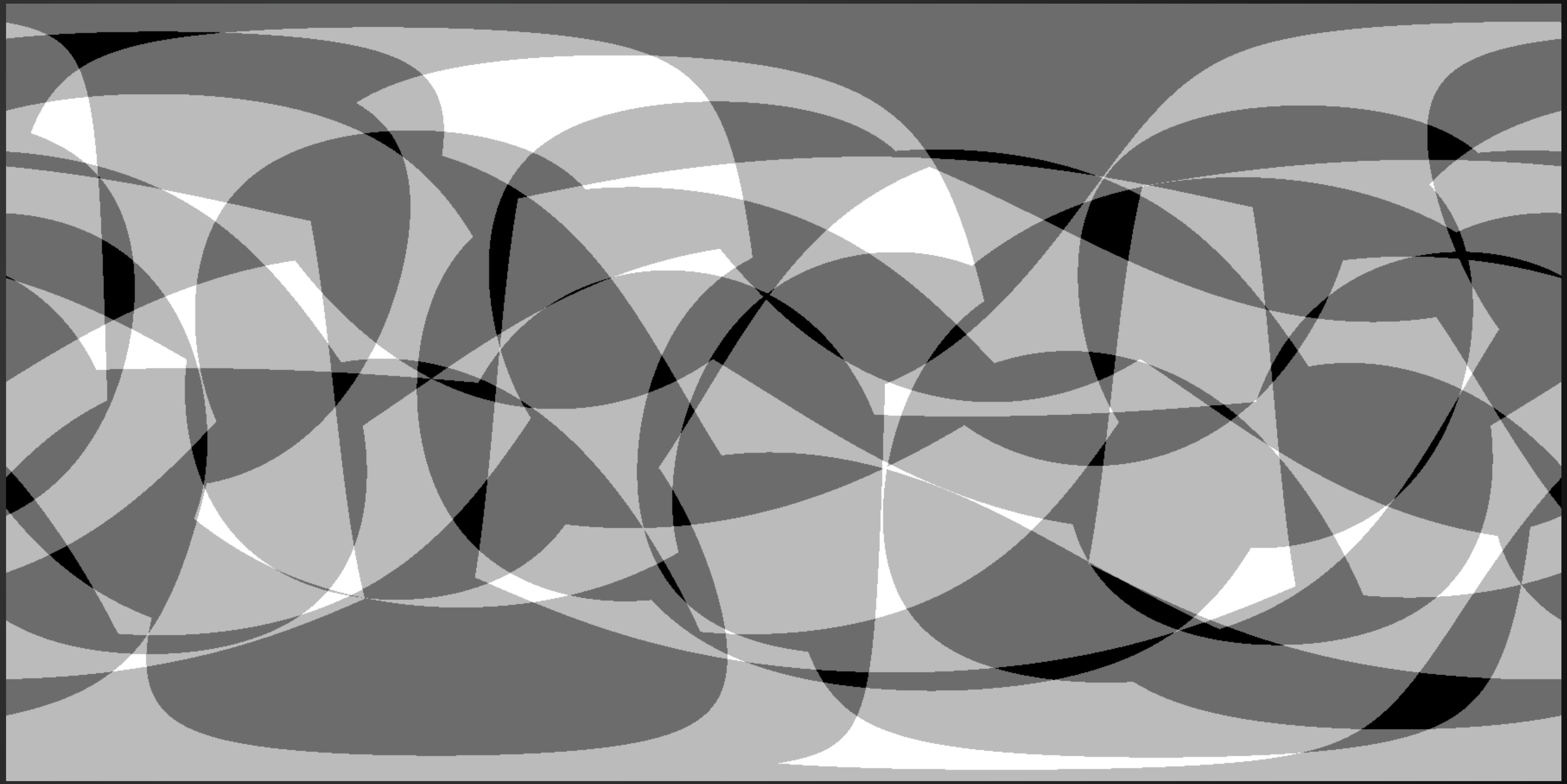


The alternative is to use a sparse, high resolution array

- We need to solve the novel view synthesis problem
- One approach: estimate depth and re-project
  - This not the only approach
- Depth estimation is hard ill-posed problem

How do build such a camera

- Maximize camera overlap
- Minimize camera count
- Maximize head-box
- Minimize weight
- Deal with thermal issues
- Posses high reliability
- Have live preview
- Have great pixels (high SNR)

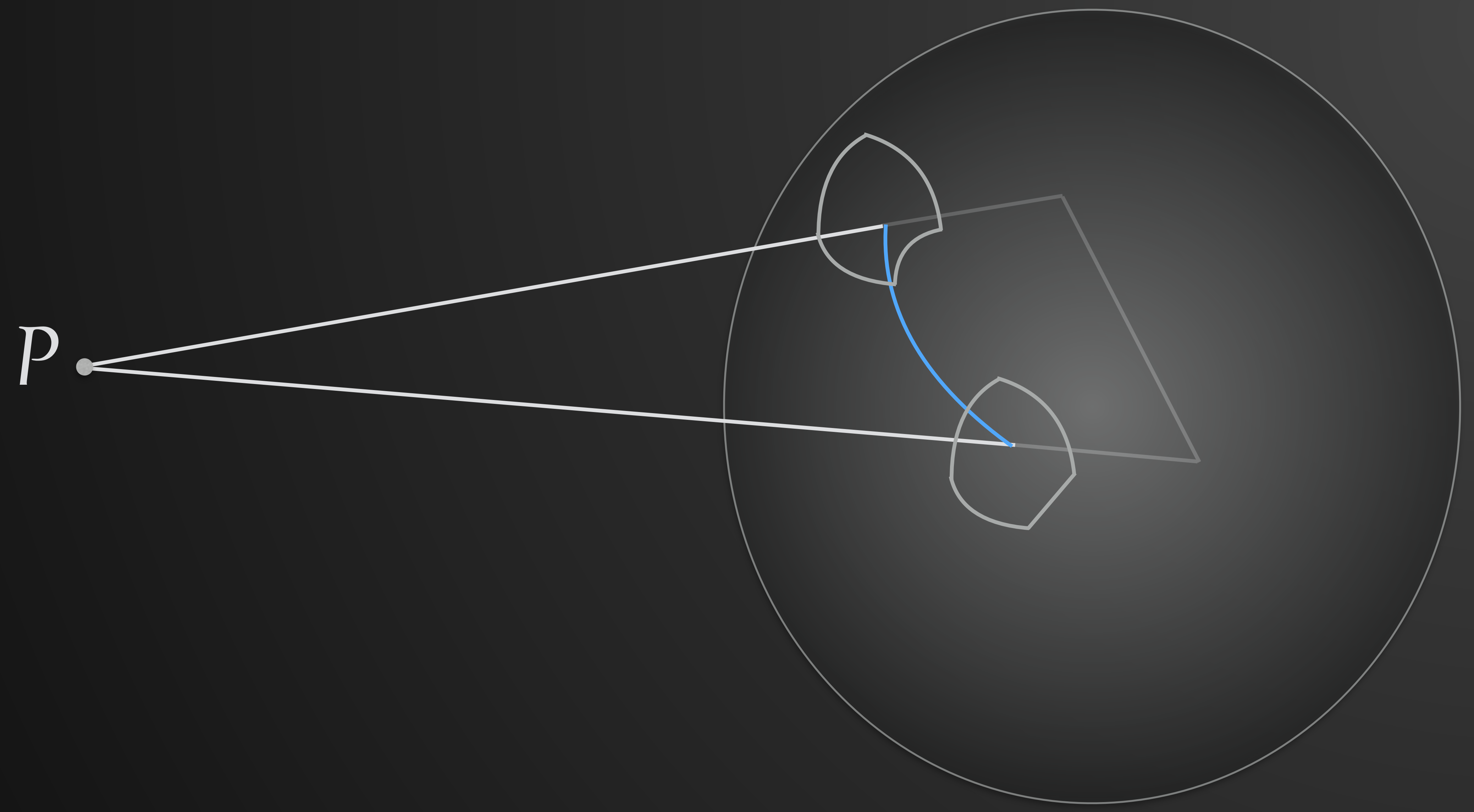


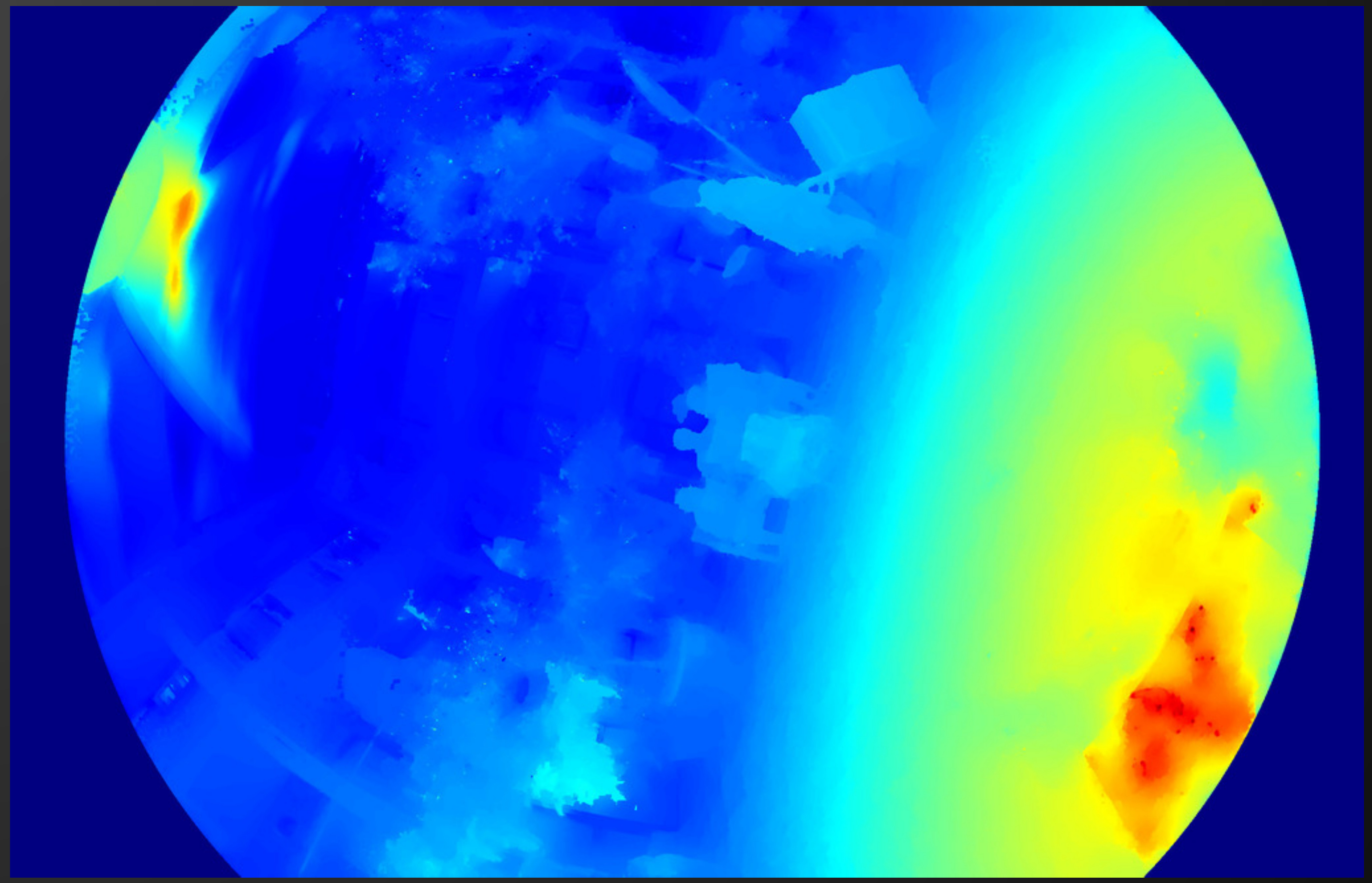
Calibration is critical

- You would like to make it easy to do in the field
- We developed a marker-less system
- Use classic bundle adjustment (good up to “scale”)

Create depth maps per camera

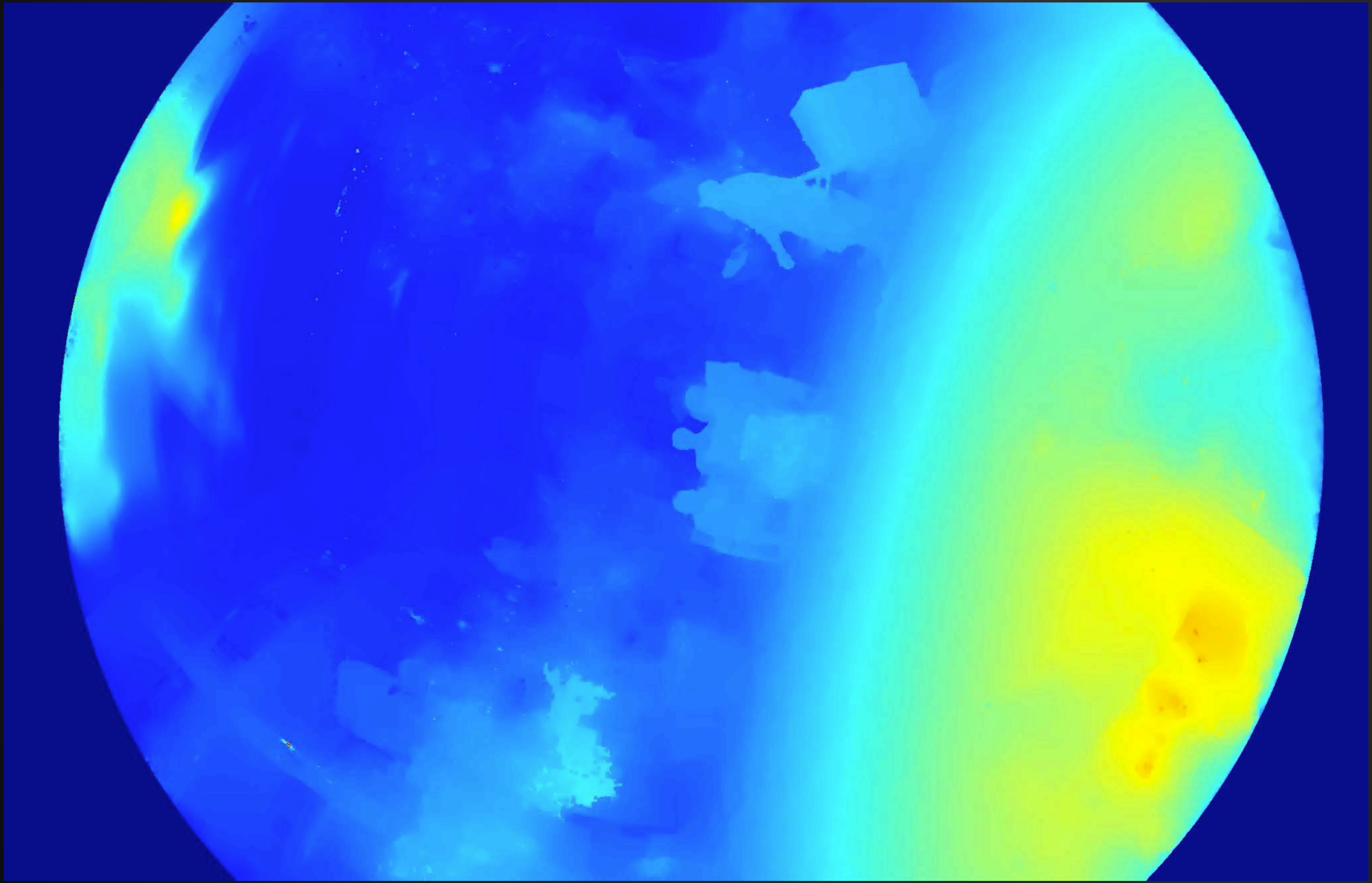
- Even though our mapping is non-linear all the homographic properties hold
- Epipolar "lines" become great arcs and curves
- A depth map per camera





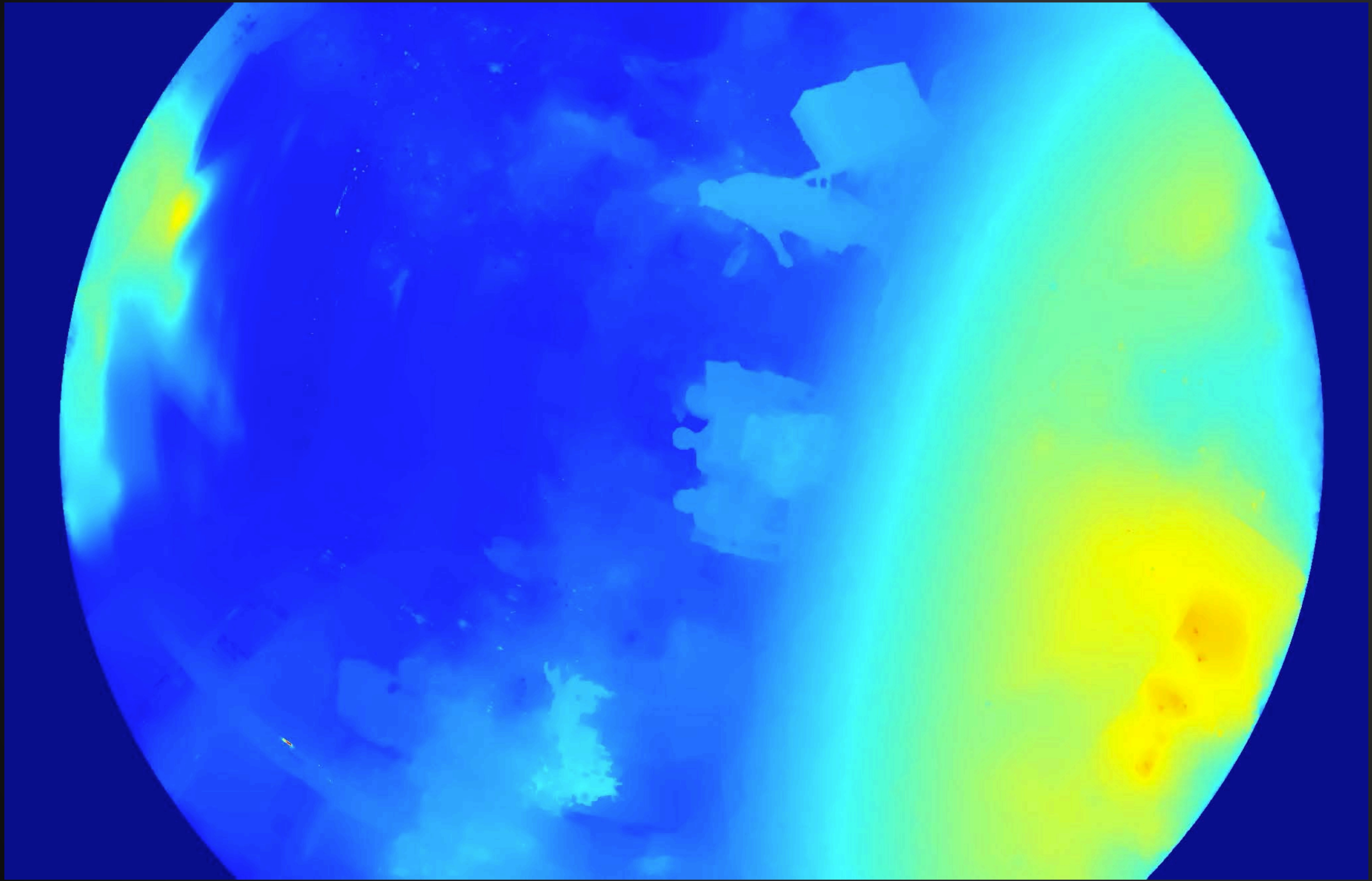
Must filter in time

- Without temporal filtering



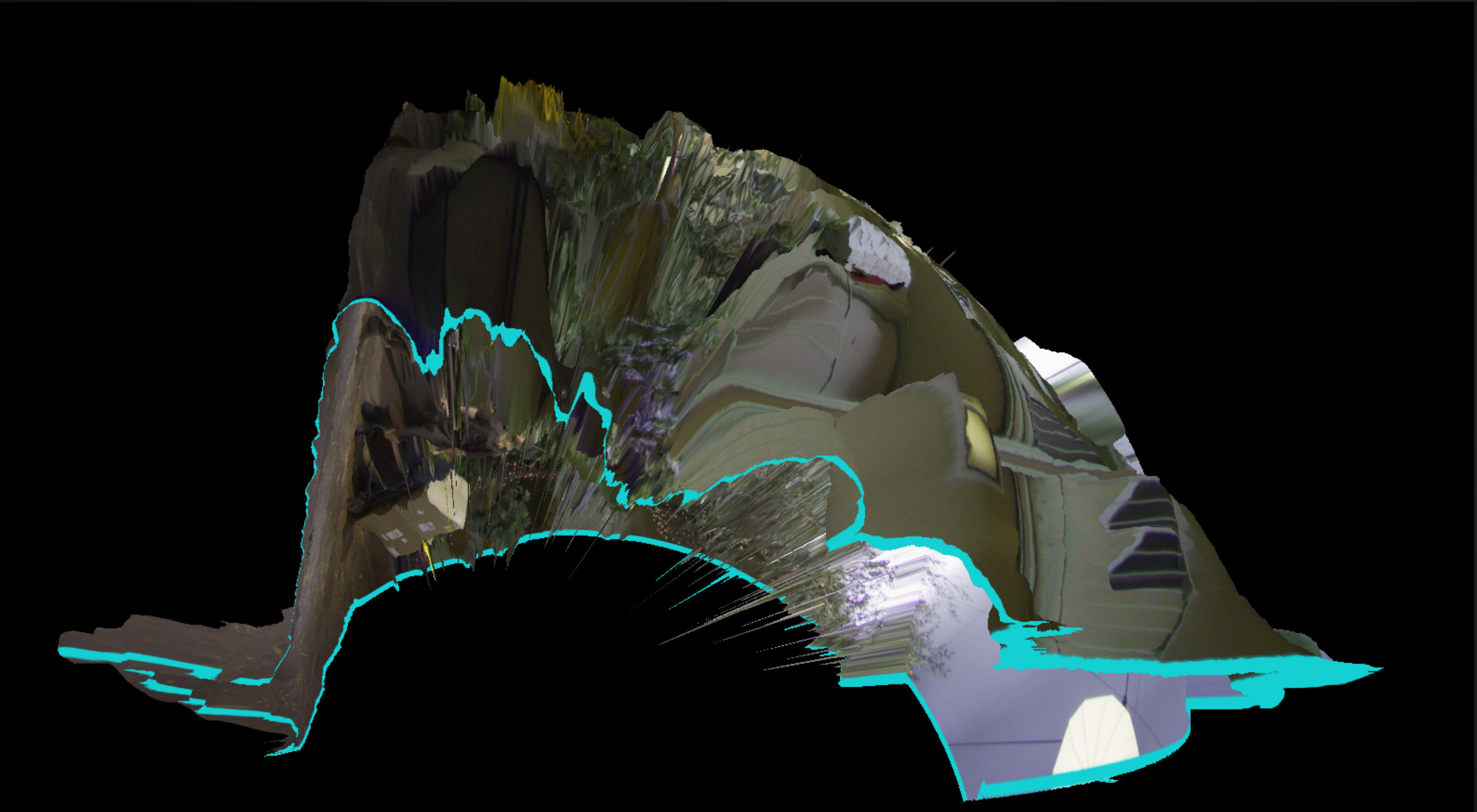
Must filter in time

- Without temporal filtering



We can re-project this into a single space

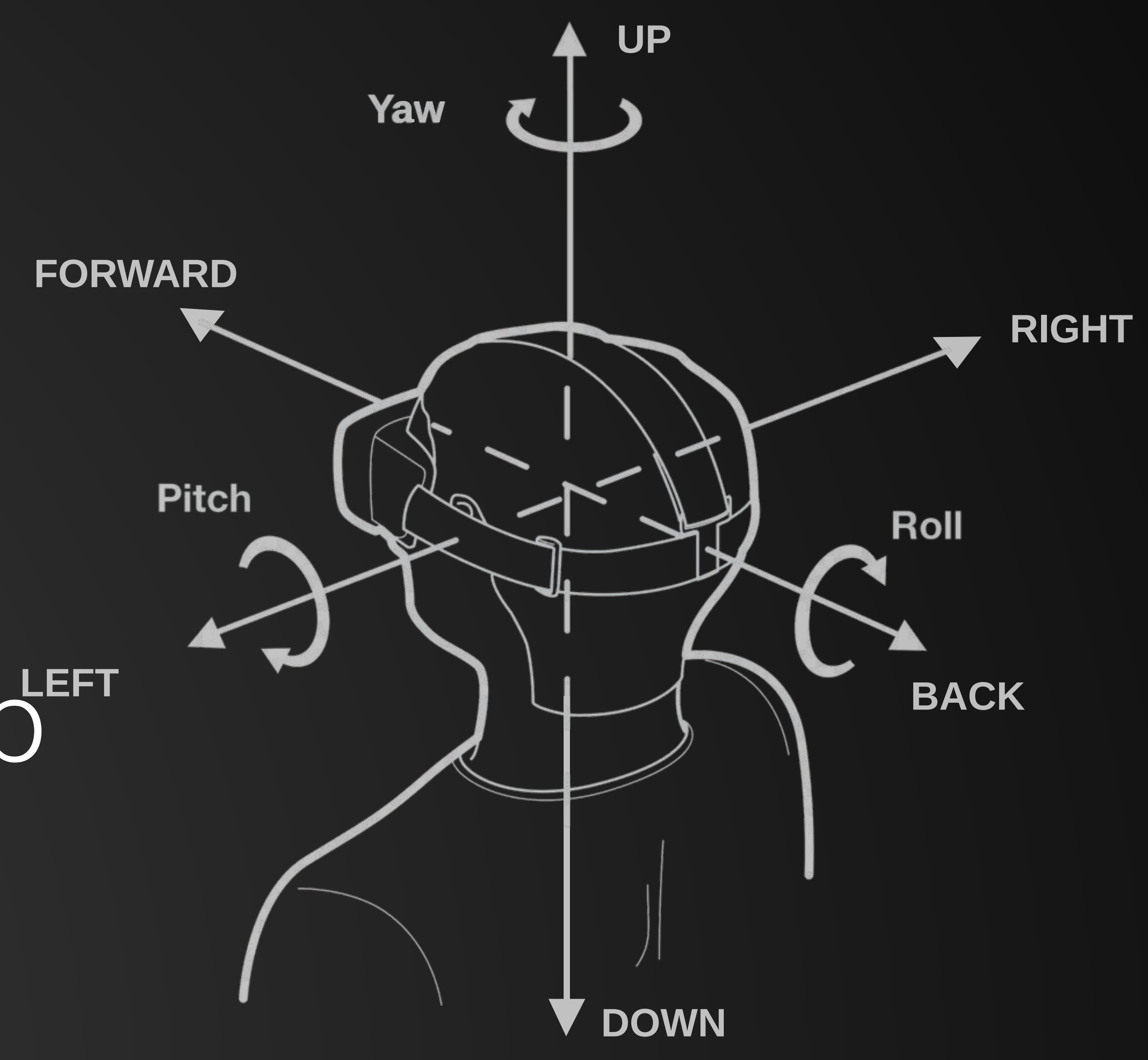
- Each camera's view creates a draped canopy



We can re-project this into a single space

- Streaky triangles induced by depth discontinuities

The 6-DOF + time Challenge



- Closing notes
  - You often still need edge and depth/disparity info to clean up edges for pure light field approaches
  - Sparse cameras don’t preclude image based approaches
  - Good, reliable, calibrated cameras are a necessary for professional capture

Microsoft

Mixed Reality Capture Studios

HOLOGRAMS FROM REAL LIFE

View moments in time from every angle imaginable. From the professional to the personal.

SPRING 2019

ARTS AND ENTERTAINMENT

EDUCATION AND TRAINING

COMMERCE

PERSONAL MEMORIES

Los Angeles/ Metastage

Professional Soundstages

Licensed technology stack

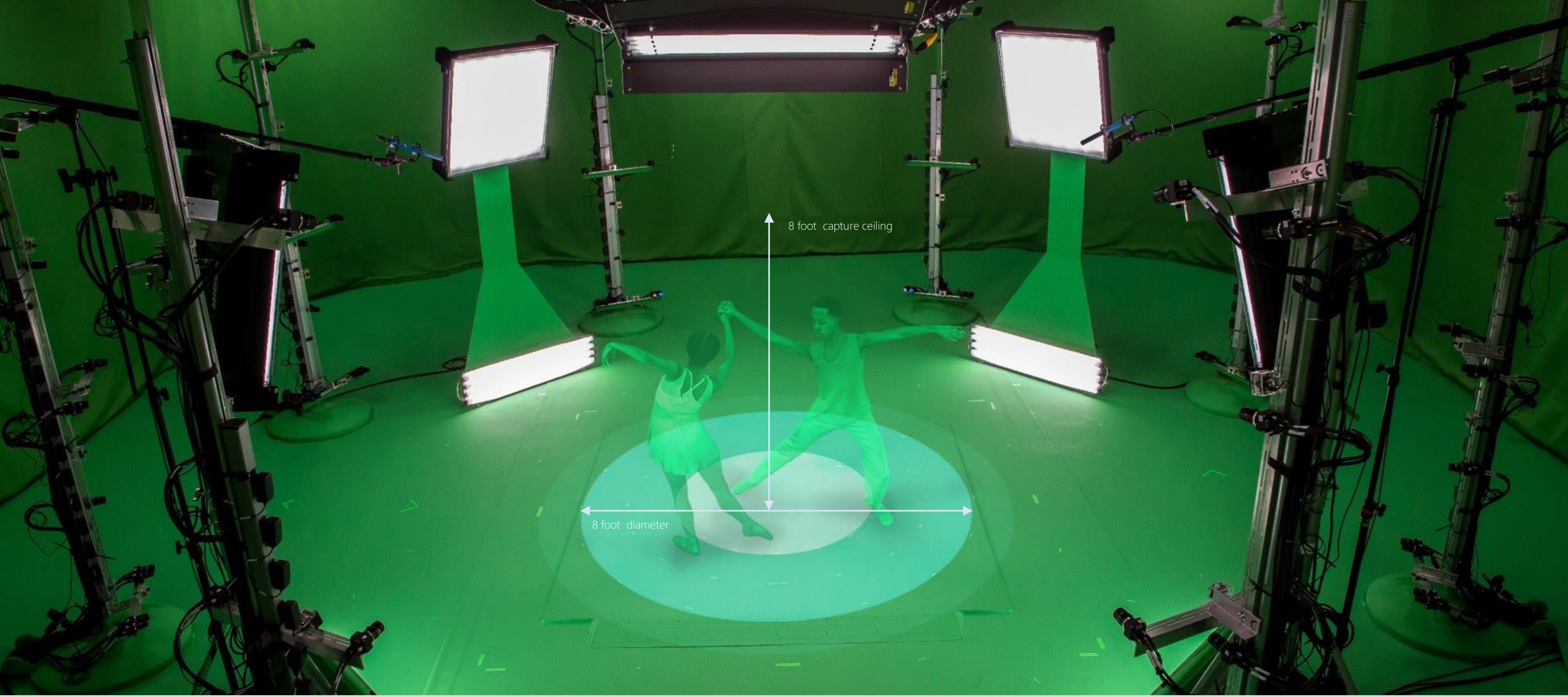
SAN FRANCISCO

LOS ANGELES

LONDON

NEW YORK

2019 TBA



BY THE NUMBERS

Standard configuration is an 8-foot-diameter circle
- Smaller configurations provide higher resolution
- Larger configurations (up to 10 ft) possible for some scenarios

Capture more than 1 person at a time
Customizable lighting
Suspension rigging capable

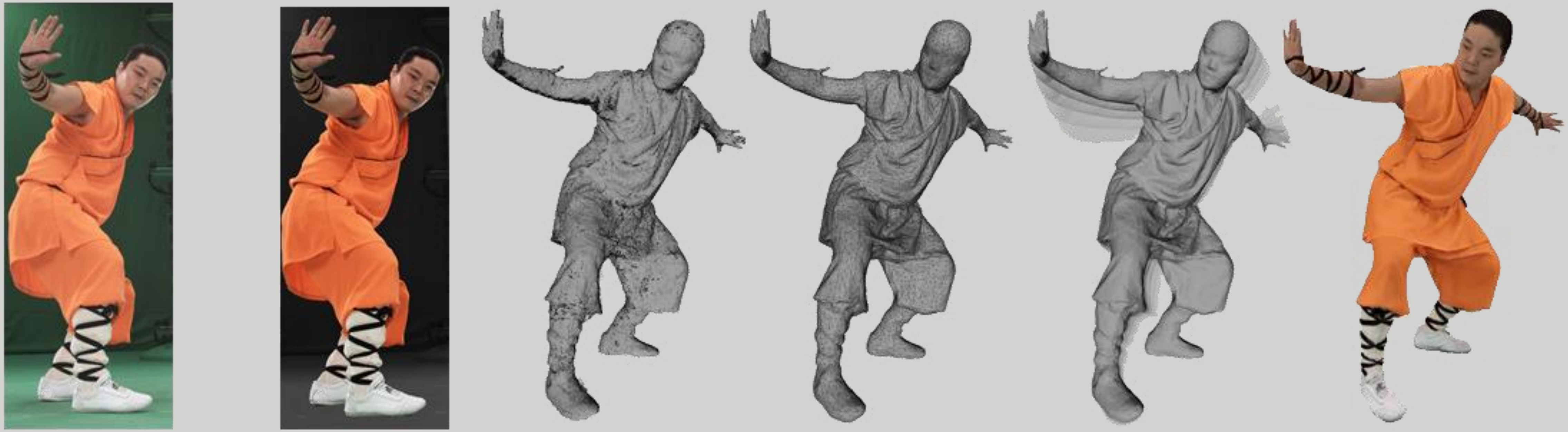
106 synchronized cameras
- 53 IR + 53 RGB cameras
- 4MP sensors
- 30 fps standard
- Timecode

8 High-quality directional mics

WHAT WE DO

Our multi-stage process uses multiple data sets to generate a high quality refined point cloud that produces an exceptional mesh and texture that is compressible to HD video like file sizes.

For more, see our 2015 SIGGRAPH VIDEO



Shoot

CAMERA RESULTS
- 53 RGB
- 53 IR

PREPROCESSED IMAGES
- Background masks

POINT CLOUD
- Millions of points

SMOOTH MESH
- 100s of thousands of triangles
- Decimated to desired result
- Detail preservation (i.e. face)

TEMPORAL MESH
- Mesh tracking creates keyframes
- OBJ mesh
- PNG texture

MP4 FILE
- H264 Compression
- Streamable

PLATFORM SUPPORT

unity	UNREAL ENGINE	Native
Windows	Windows	Windows
iOS, Mac, Android	magic leap, Android	PlayStation.VR, iOS, Android

Post-processing

EDITING

Edit mesh and texture sequences with standard DCC tools like Maya, Photoshop, and Nuke.
Re-encode edited content back to a single material MP4 with our proprietary tool.

AUDIO

We provide basic audio capture and a scratch mix for review. Your audio engineer will be able to sweeten and enhance the source files to create your perfect mix.

RELIGHT

It's possible to use Maya, Arnold, V-Ray, etc. to render complex lighting information like sub-surface scattering post-capture, and then bake back out to our compressed MP4 format. There are limitations, so be sure to chat with us prior to capture.

Post-processing

GAZE RETARGETING

Shader-based mesh deformation can automatically change the angle of a presenter's head to be more in line with a viewer's location without needing to rig and animate.

ROTOSCOPING

We have a Maya workflow and tool for adding simple animated objects post-capture. Props like golf clubs, glasses, swords, and other elements are hard for us to capture well, but they are good candidates for adding in post.


billboard
MUSIC AWARDS
ARTS AND ENTERTAINMENT

The New York Times Magazine

GREAT PERFORMERS — IN — AUGMENTED REALITY

LAKEITH STANFIELD'S BALANCING ACT

SEE A LIFE-SIZE HOLOGRAM OF THE ACTOR TEETERING ON AN IRON BEAM, HIGH ABOVE THE CITY, IN AUGMENTED REALITY.

DEC. 5, 2018

ARTS AND ENTERTAINMENT


CANNES LIONS
18 22 JUNE 2018
DIGITAL CRAFT GRAND PRIX
WINNER
isobar
ARTS AND ENTERTAINMENT

START VR

ARTS AND ENTERTAINMENT


ELEVATOR ACCESS
EXIT
BEGINNINGS
Smithsonian Institution
EDUCATION AND TRAINING


EDUCATION AND TRAINING


Anaphylaxis
Stage 1
Stage 2
Stage 3
100
9
R
142/84
BP
100.4
°F
T
Pearson
EDUCATION AND TRAINING

Alibaba Group
阿里巴巴集团

COMMERCE AND RETAIL

Microsoft

Mixed Reality Capture Studios

thrive SIGGRAPH2019
LOS ANGELES • 28 JULY – 1 AUGUST

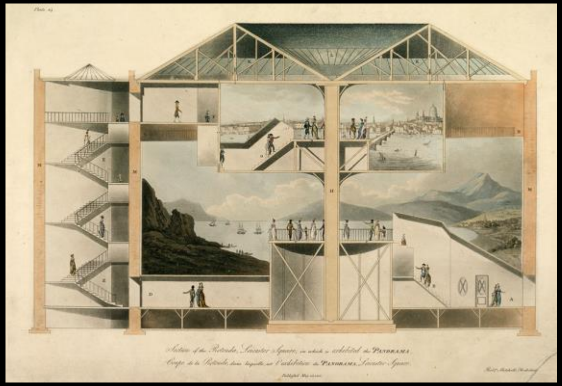
Christian Richardt

Course Conclusion

CAMERA
Centre for the Analysis of Motion, Entertainment Research and Applications

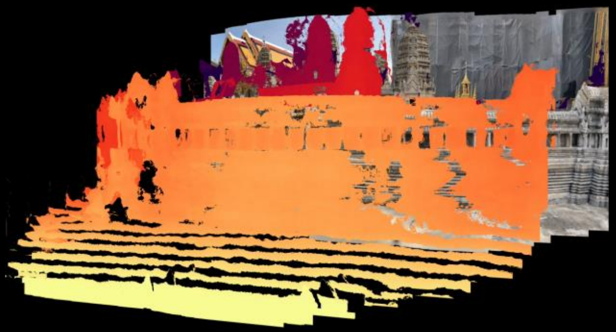
UNIVERSITY OF BATH

Visual summary


Robert Mitchell, 1901


David Tett/KCL


Facebook


Hedman et al., 2018


Overbeck et al., 2018


Facebook


Konrad et al., 2017

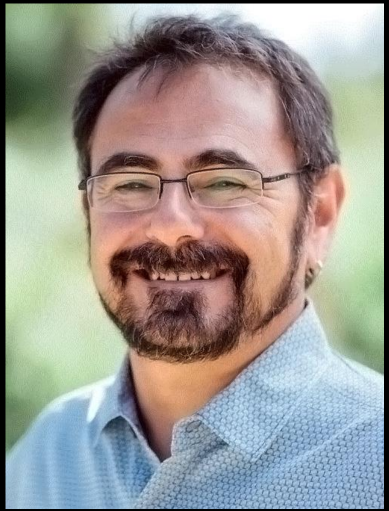

CNET/Facebook


Microsoft

Asking our Team

Christian Richardt | Peter Hedman | Ryan S. Overbeck | Brian Cabral | Robert Konrad | Steve Sullivan



University of Bath | UCL | Google | facebook | Stanford University | Microsoft

What single thing would most improve VR photography or VR video?

Robert Konrad

“I guess the single biggest thing that is missing from cinematic VR currently is supporting 6DOF.

The difference in experience in going from a single viewpoint to allowing for translations is huge.

After that I think real-time streaming of content and supporting focus cues would be the two next most important considerations.”

Steve Sullivan

Microsoft

"From my perspective doing volumetric,
I'd answer in two parts:

- more syncable camera options
  (to speed innovation in capture systems)

- more support from large-scale social platforms
  (to reduce consumption friction and speed adoption)"

Peter Hedman

UCL

1. "A benchmark dataset for VR multi-view stereo (MVS). If we had a 3D reconstruction benchmark [...] for VR capture, it would encourage research into robust MVS algorithms that could be used in any VR capture system.

2. A robust solution to narrow-baseline SfM would be incredibly useful for hand-held capture, facilitating even easier capture with a larger variety of cameras and lenses."

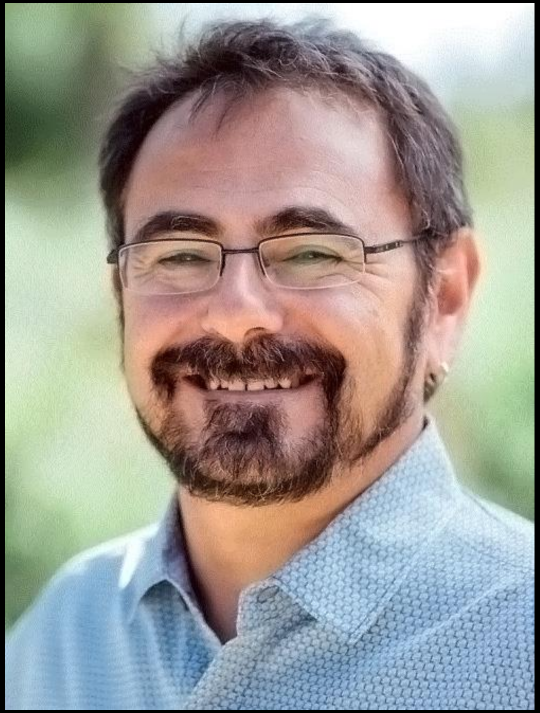
Ryan S. Overbeck

Google

"We need to build a sustainable flywheel for VR photo and video content. This is a bit of a cop-out on the "single thing" part of the question because we need to improve the whole chain.

We need faster and better capture methods and editing workflows, we need more quality content, we need high quality distribution platforms with solid monetization, and, of course, we need more consumer headsets in the wild."

Brian Cabral



facebook

"We talk a lot about the crossing the 'uncanny valley' that limits model based VR capture.

I claim there is a 'artifact barrier' limiting model free capture such as presented in this course. If we could just capture the 3D scene or lightfield without artifacts as we do 2D videos we would have a much wider and quicker adoption.

Crossing that barrier is the THE challenge that is holding back VR video capture."

Christian Richardt

“We need new, improved algorithms that can reconstruct dynamic 360-degree environments from multiple video streams to produce lifelike 6-DoF renderings in real time.

Overall, the entire VR video pipeline, from high-quality 360-degree environment video capture, over reconstruction and processing, to rendering and display in 6DoF with light field displays requires more research and engineering.


thrive
SIGGRAPH2019
LOS ANGELES • 28 JULY – 1 AUGUST
Thank you!
Questions?
Capture4VR
From VR Photography to VR Video
richardt.name/Capture4VR