Lecture 19: Image-Based Rendering of Immersive Content

Panorama, (Stereo) Virtual Reality/360° Video, and Light Field

Yuhao Zhu

http://yuhaozhu.com
yzhu@rochester.edu

CSC 259/459, Fall 2025
Computer Imaging & Graphics

The Roadmap

Theoretical Preliminaries

Human Visual Systems

Display and Camera

Modeling and Rendering

Sources of Color

Display and Lighting

Photographic Optics

Image Sensor

Image Signal Processing

Image/Video Compression

Immersive Content

Panoramas

A **panorama** (formed from Greek πᾶν "all" + ὅραμα "sight") is a wide-angle view or representation of a physical space

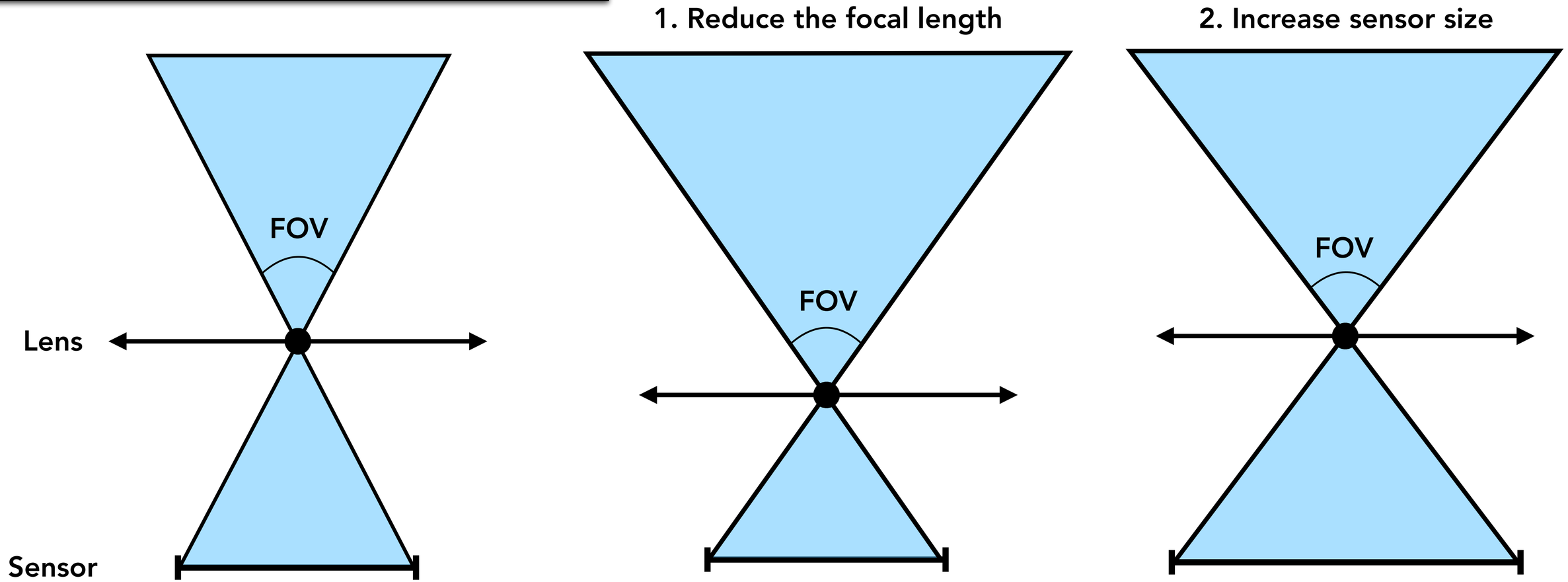
https://unsplash.com/photos/Ym7NVRKgmJc

https://unsplash.com/photos/Ym7NVRKgmJc

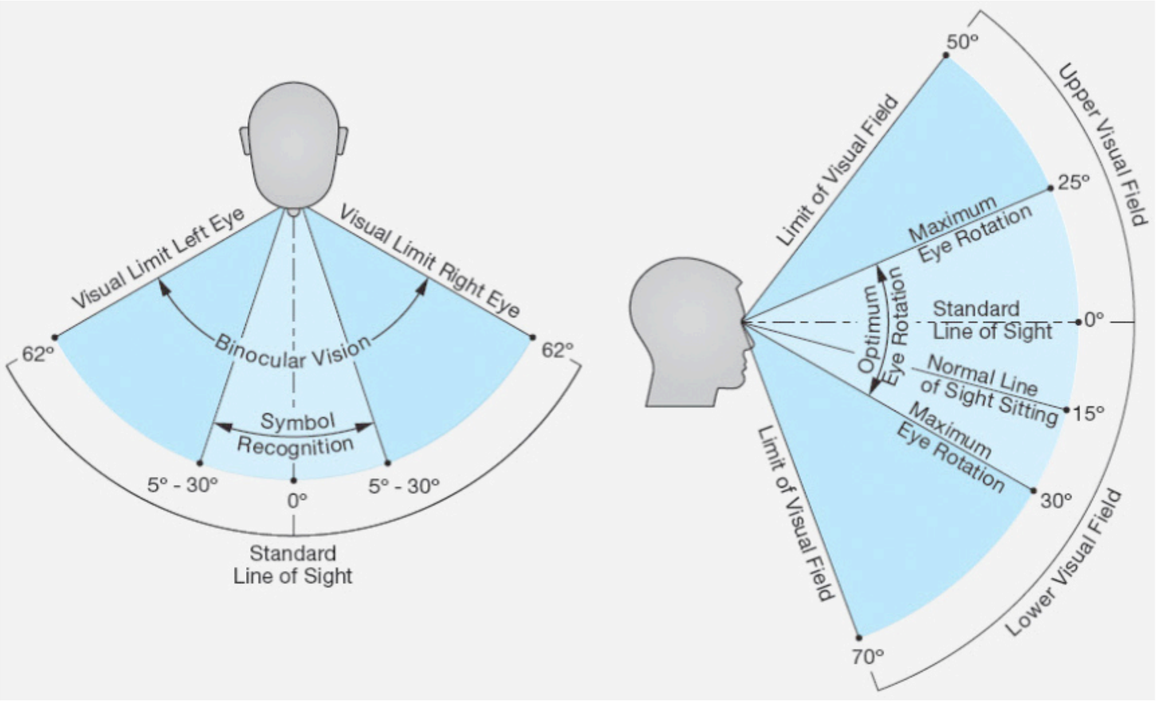
https://www.tapsmart.com/tips-and-tricks/panoramic-photos-capture-panorama-ios-12-guide/

Two Ways to Achieve Wide Field of View

Neither is desirable. The name of the game is to achieve wide FOV with narrow-FOV cameras.



How Much FOV Do We Need?

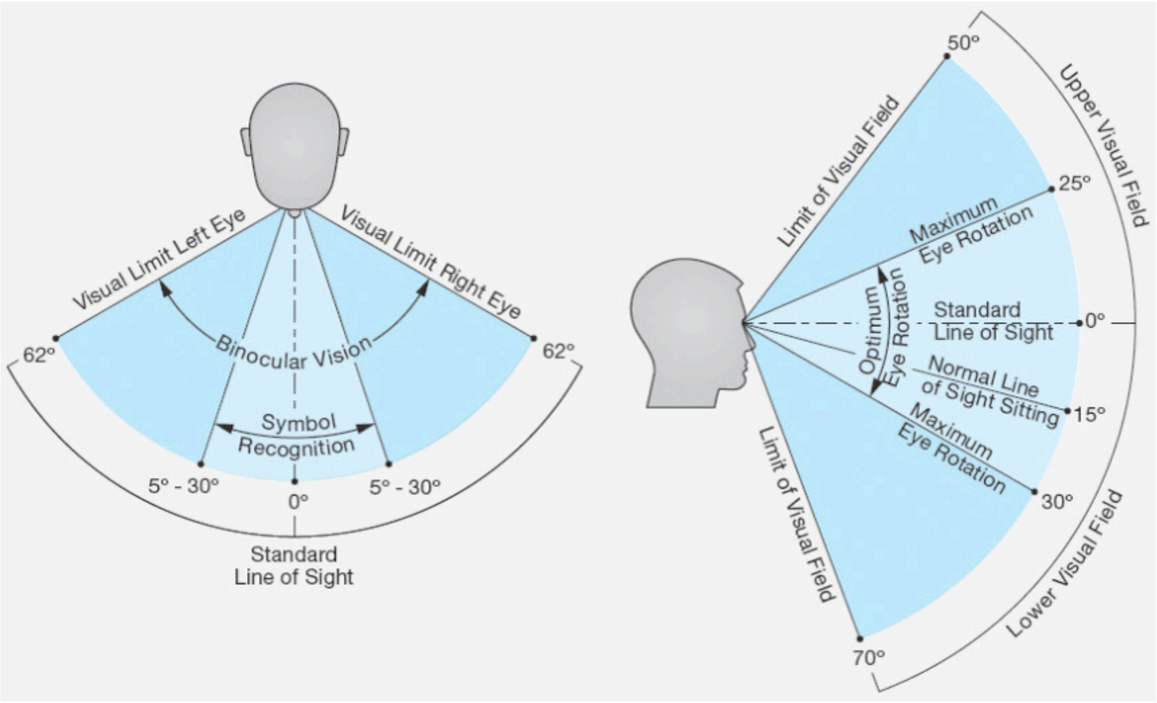


https://blog.vuze.camera/the-ultimate-immersive-experience/

Requirement for content that provides an "**immersive**" experience: human eye FOV.

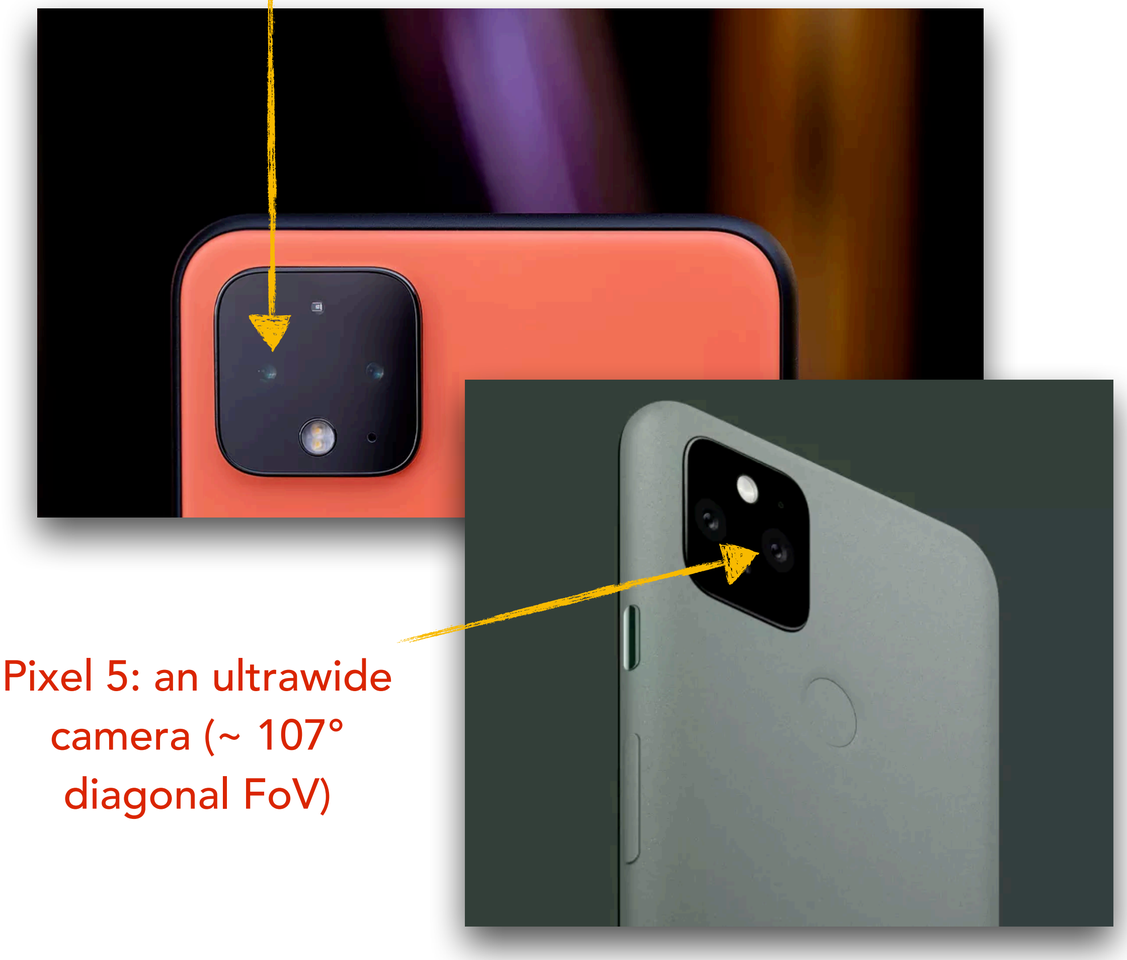
Full eye FOV is 360° horizontal and 180° vertical if we rotate head.

Camera FOV vs. Human Eyes



https://blog.vuze.camera/the-ultimate-immersive-experience/

Pixel 4: f = 28mm, ~75.4° diagonal FoV



Pixel 5: an ultrawide camera (~ 107° diagonal FoV)

https://www.theverge.com/2019/10/21/20924941/google-pixel-upgrade-camera-feature-updates-dual-exposure-live-hdr

https://www.tomsguide.com/news/google-pixel-5-release-date-price-specs-and-big-camera-upgrades

Related: Telephoto (Long-Focus) Lens

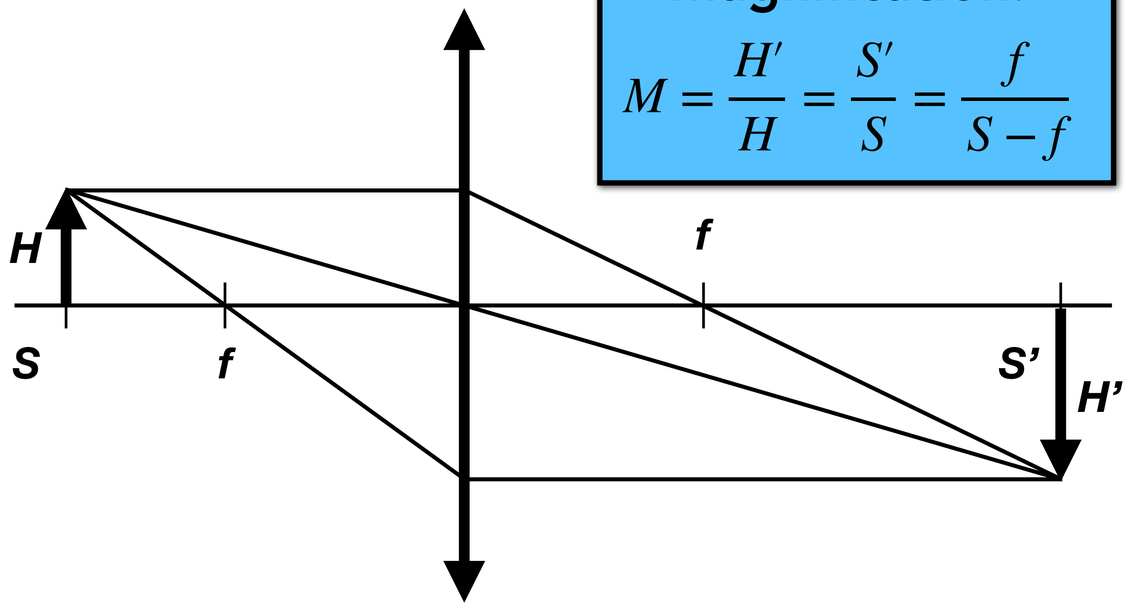
Related: Telephoto (Long-Focus) Lens

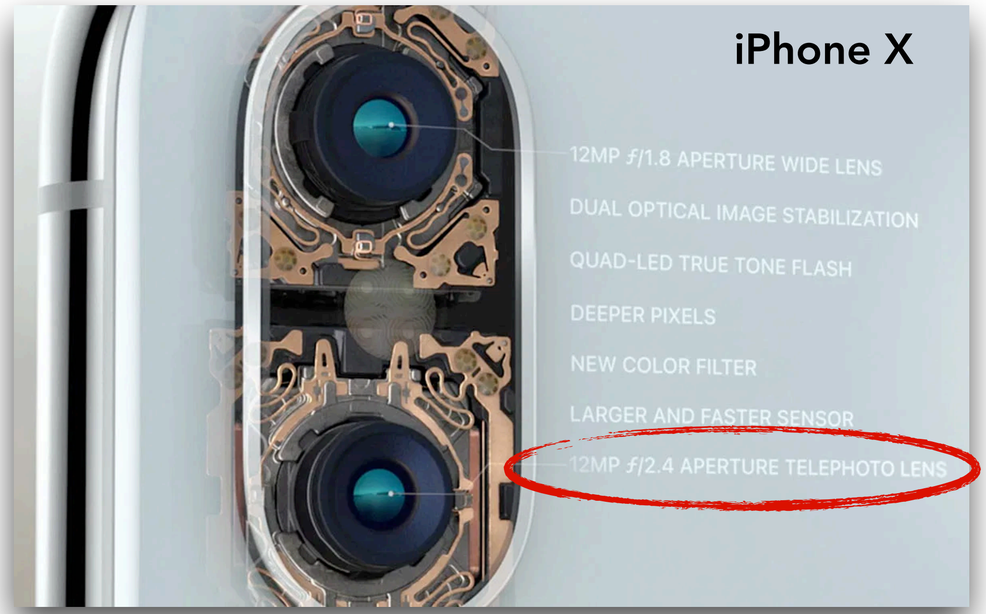
Long-focus lens brings distance object closer by having a large focal length, which leads to a large magnification factor.

- **Trade-off**: also reduces the field of view.

Magnification:

$$M = \frac{H'}{H} = \frac{S'}{S} = \frac{f}{S - f}$$

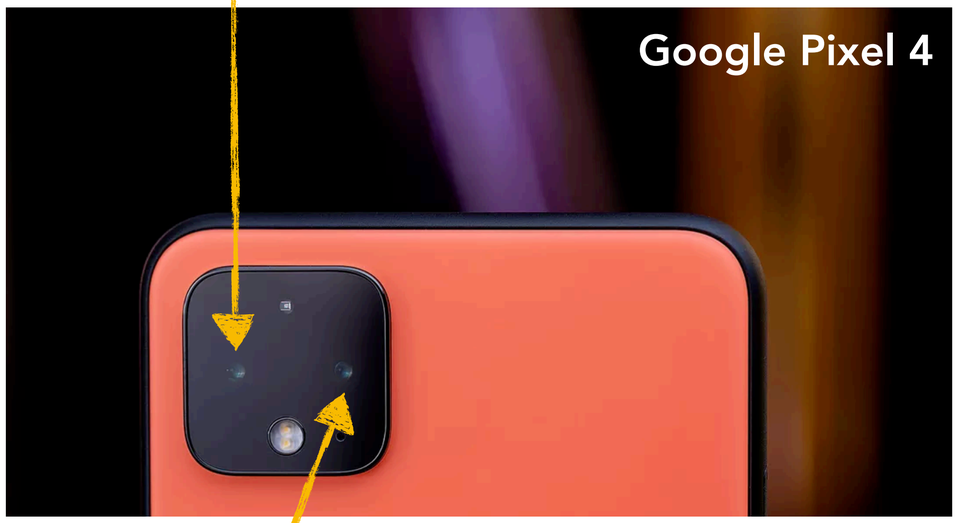




https://9to5mac.com/2018/05/01/coretronics-patent-claim-iphone-x/

Related: Telephoto (Long-Focus) Lens

f = 28mm, ~ 75.4° diagonal FoV



Telephoto lens (f = 48 mm) + computational photography algorithm for super resolution

Front camera is wider: f = 22 mm (FOV ~ 90°). Why?

https://9to5mac.com/2018/05/01/coretronics-patent-claim-iphone-x/ https://www.slashgear.com/google-pixel-4-camera-review-brilliant-and-frustrating-21596272/

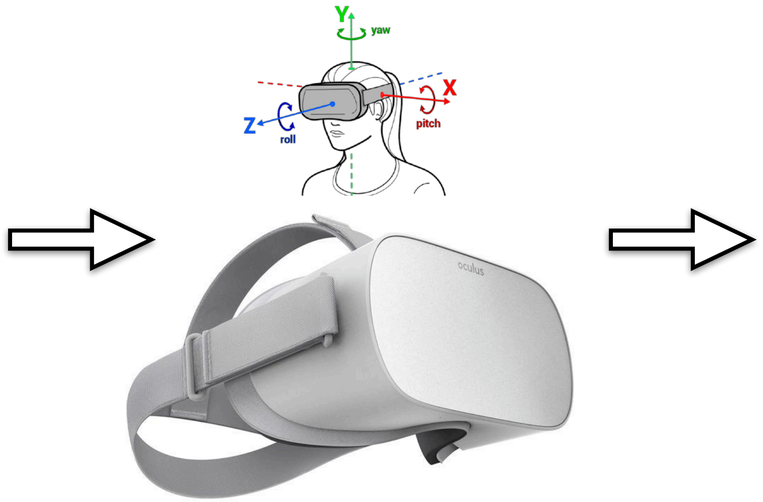
Aside: LiDAR in Modern Smartphones

https://nofilmschool.com/best-apps-scanning-objects-your-phone https://www.phonearena.com/news/iPhone-12-to-have-larger-12MP-sensors_id126865

Two Strategies Presenting Wide-FOV Content

1. Panorama presented to viewer “at once” and consumed “as is”.
   - This is a good option if the FOV is not too wide.

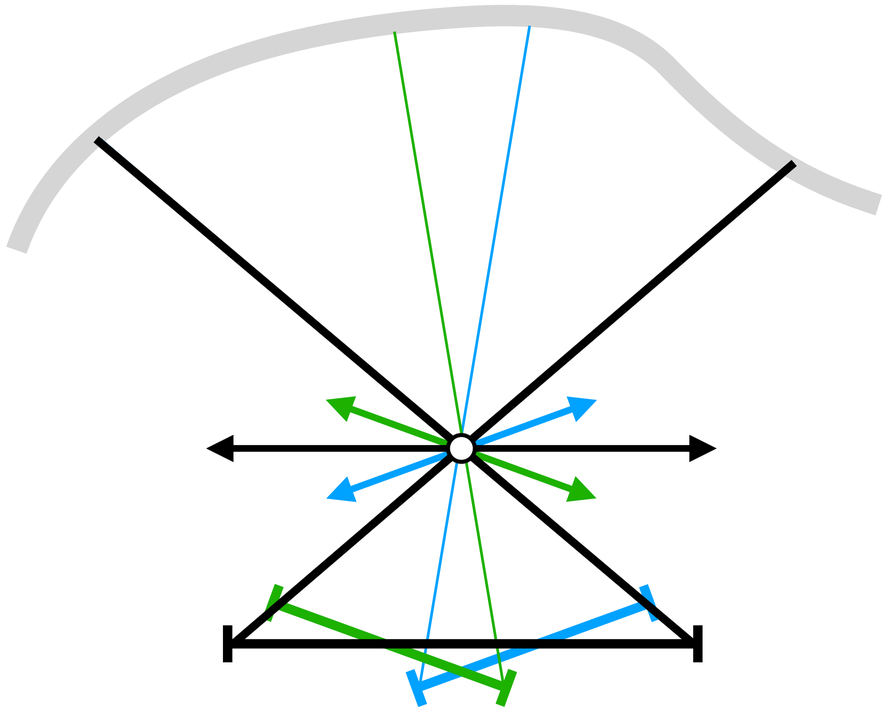
Two Strategies Presenting Wide-FOV Content

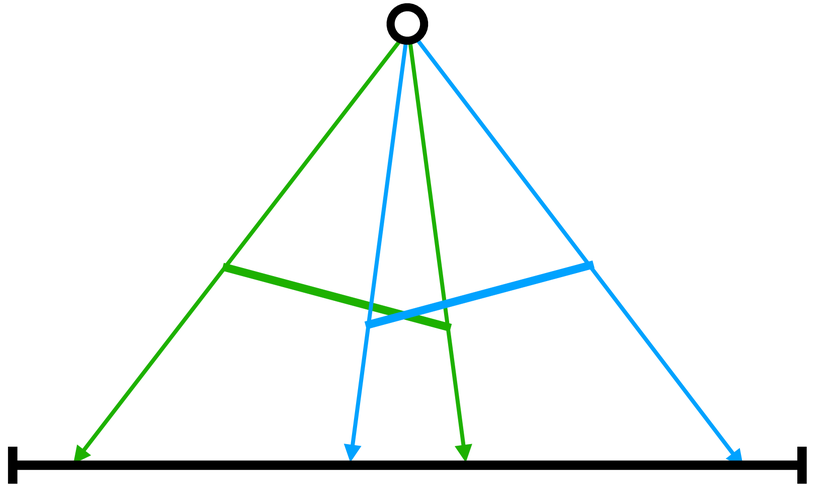


1. Panorama presented to viewer “at once” and consumed “as is”.
   - This is a good option if the FOV is not too wide.

2. Present a small FOV based on head orientation (or gaze direction)
   - This is perhaps the only option for presenting 360 content.

Obtaining Wide FOV Using Narrow-FOV Cameras

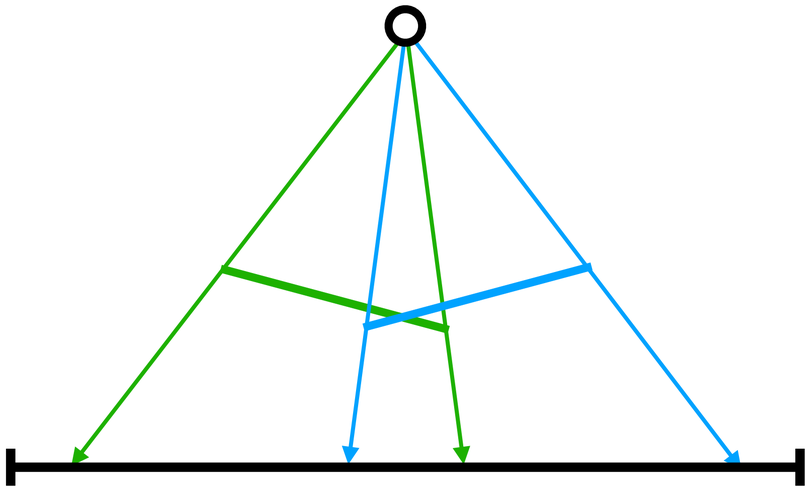




Obtaining Wide FOV Using Narrow-FOV Cameras

Stitching two images to form a wider FOV image is done by re-projecting the two images to a common sensor plane. This is a **perspective projection**.

Maintaining the same “perspective”. The new image looks just like taken with an actual wider FOV camera.

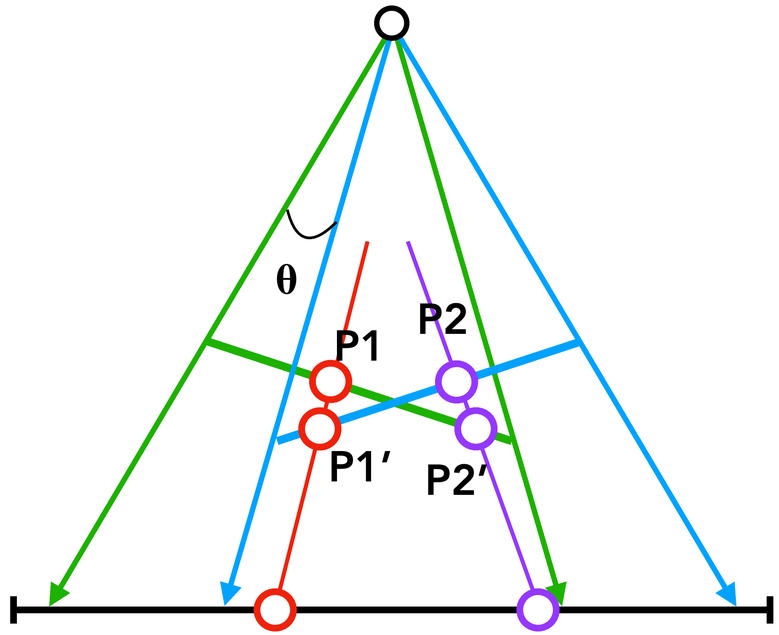


Two Requirements

1. The camera must be rotated around the **"center of perspective"**
   - which is the pinhole in a pinhole camera or the middle of the lens in an ideal thin lens (can't have translational movement).

2. The rotation angle must be known
   - which in practice is hard to know. We usually have to calculate an estimation that best fits the observations.
   - The observations are from **matching** pixels in the overlapping area between two captures.

Some form of stereo matching algorithm (e.g., optical flow) is used to match pixels.

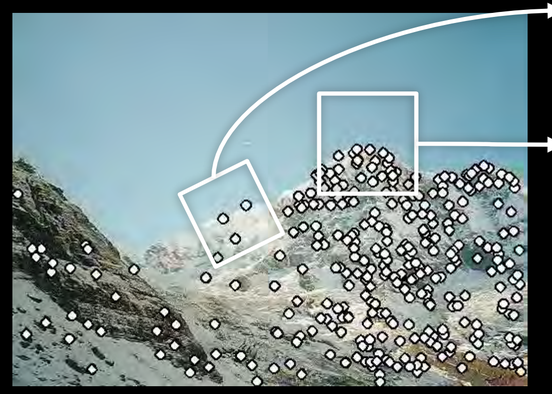


Feature matching

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

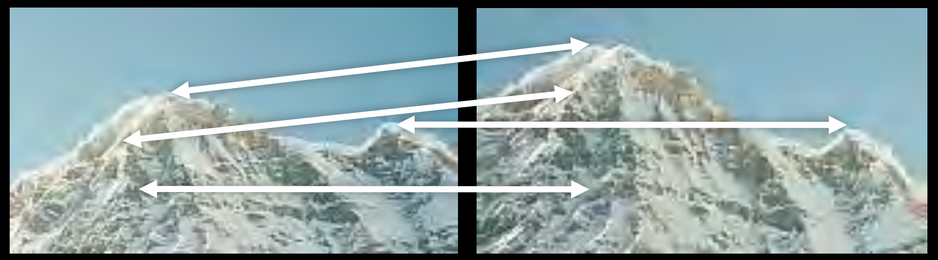
2. **Description:**
   Extract vector feature descriptor surrounding each interest point.

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



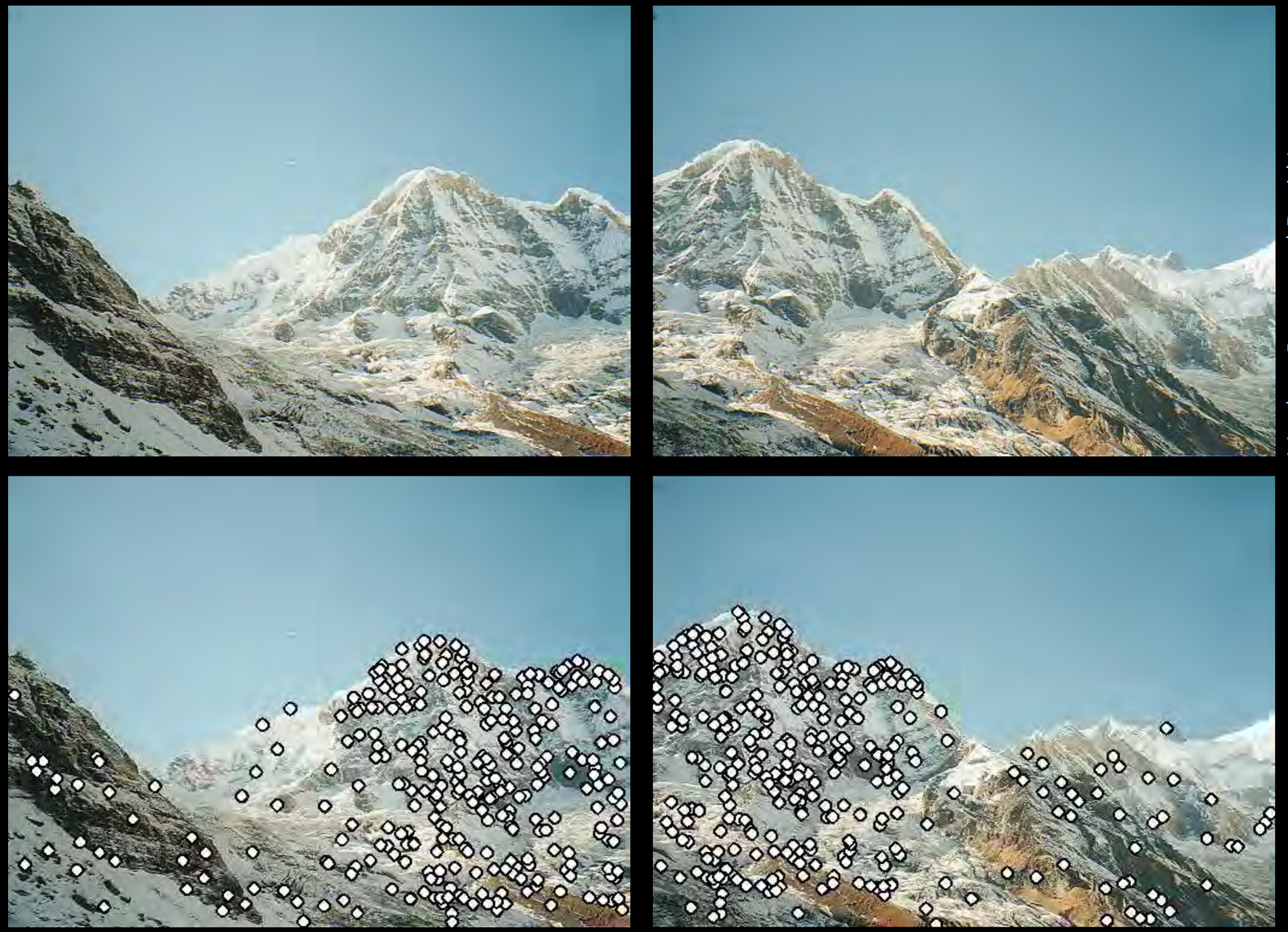
$$\mathbf{x}_1 = \left[x_1^{(1)}, \ldots, x_d^{(1)}\right]$$

$$\mathbf{x}_2 = \left[x_1^{(2)}, \ldots, x_d^{(2)}\right]$$



Slide by Kristen Grauman

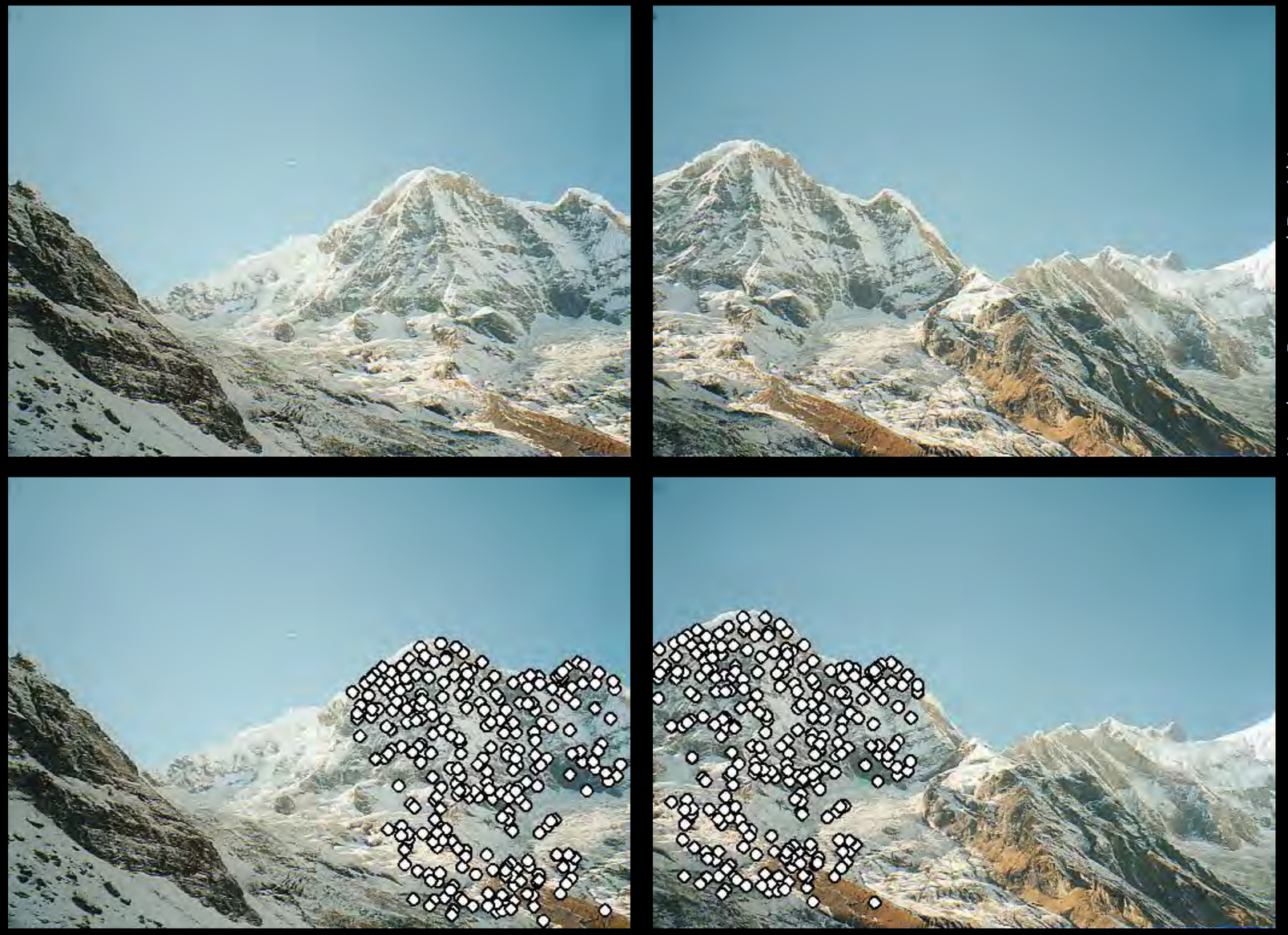
SIFT features



Matthew Brown and David Lowe

Slide by Matthew Brown

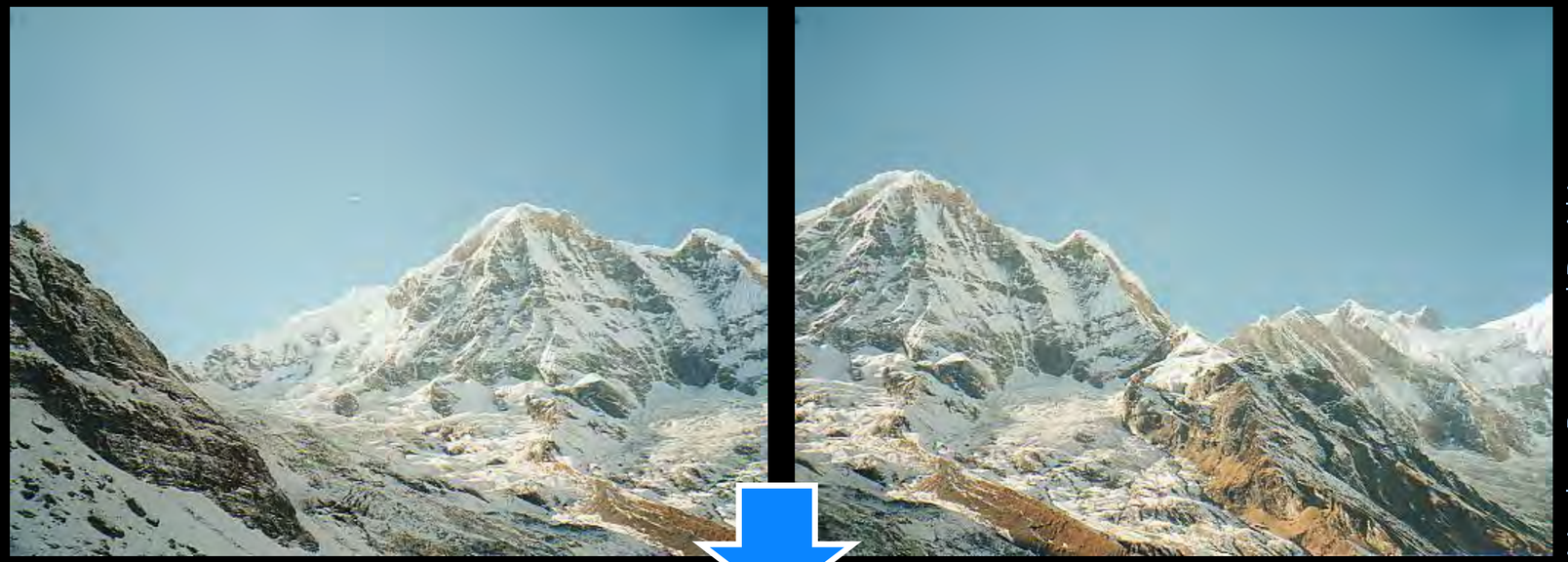
Matched SIFT features



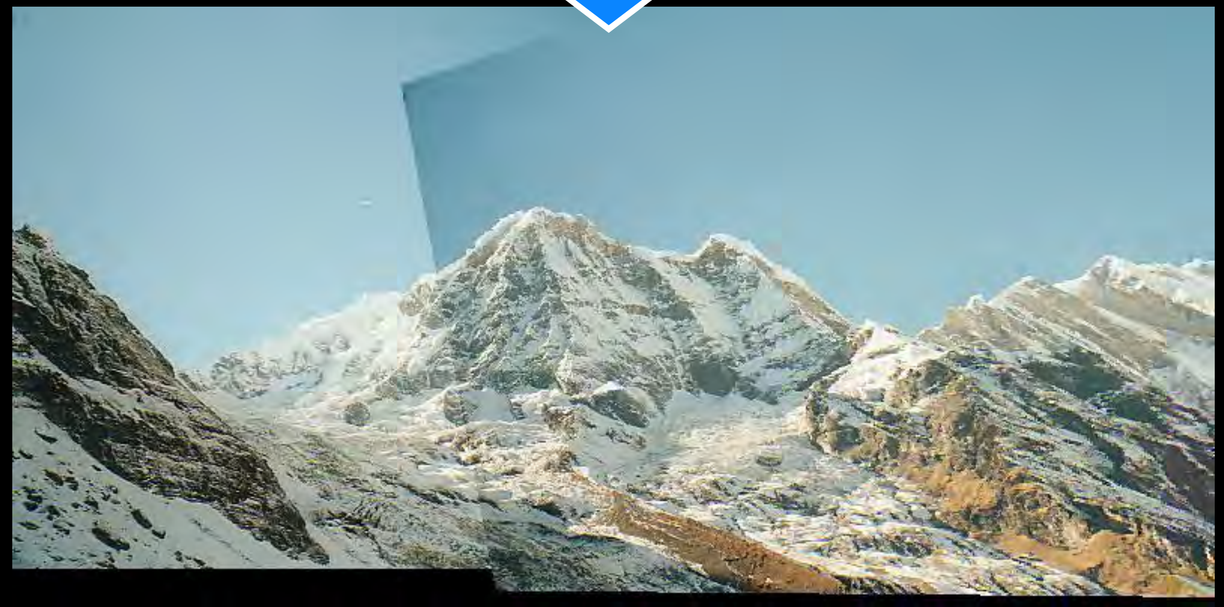
Matthew Brown and David Lowe

Slide by Matthew Brown

Aligned images



Matthew Brown and David Lowe



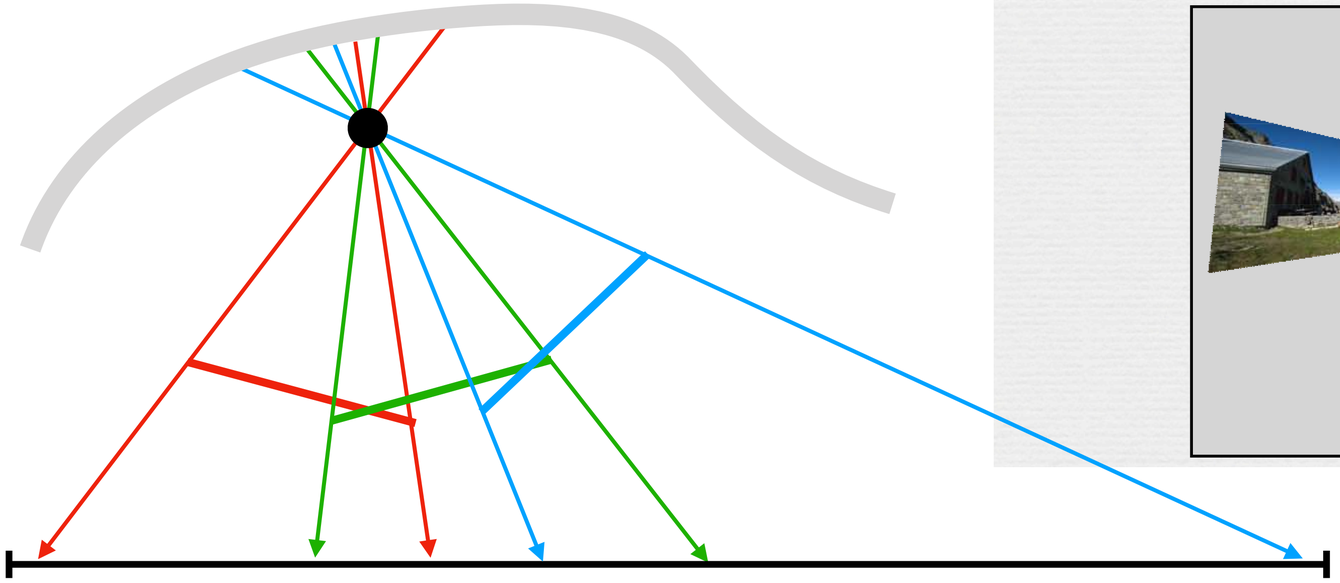
The alignment process is usually formulated as an optimization problem that minimizes the error of matching pairs.

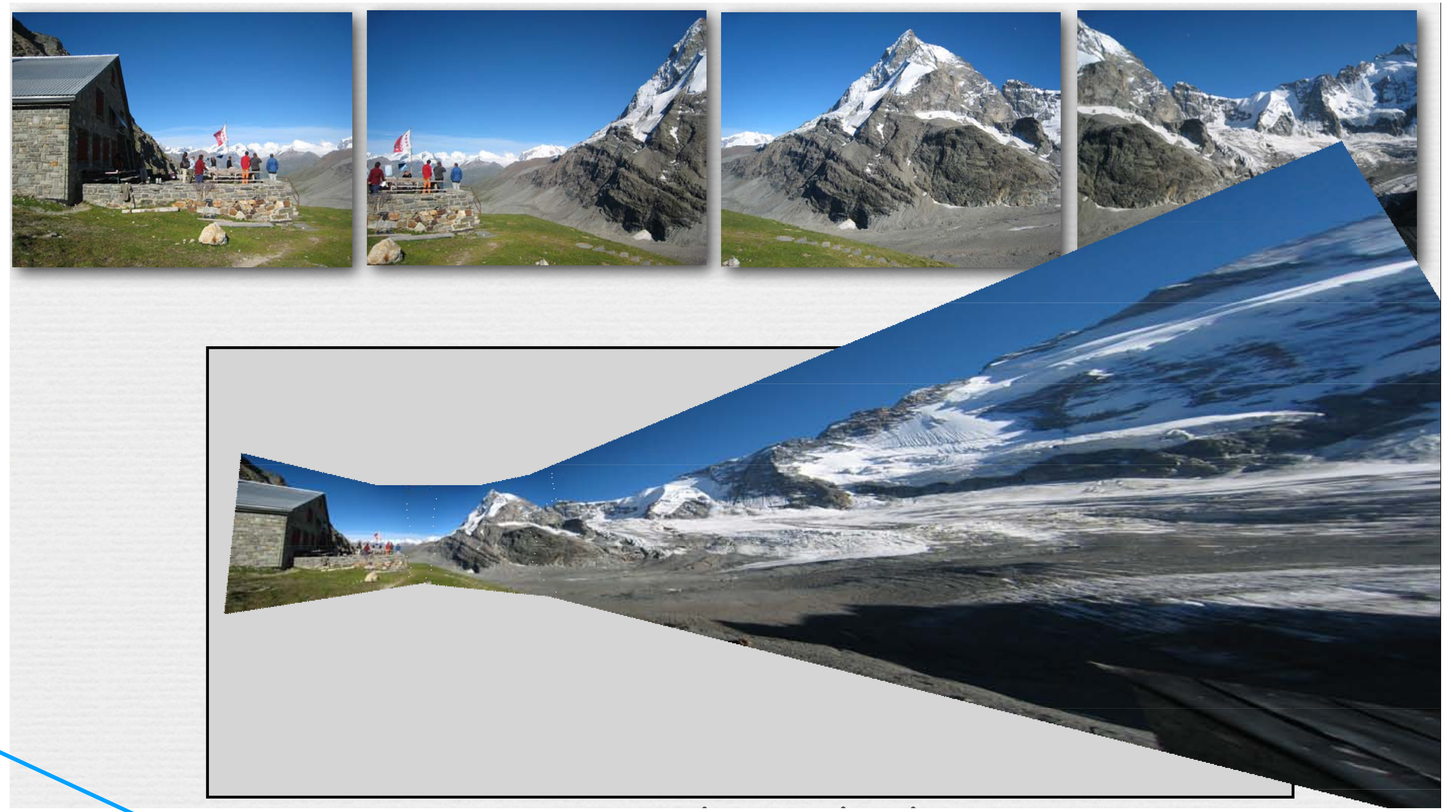
Slide by Matthew Brown

Ghosting: If We Don't Rotate Around the COP

Planar Mapping Distortion

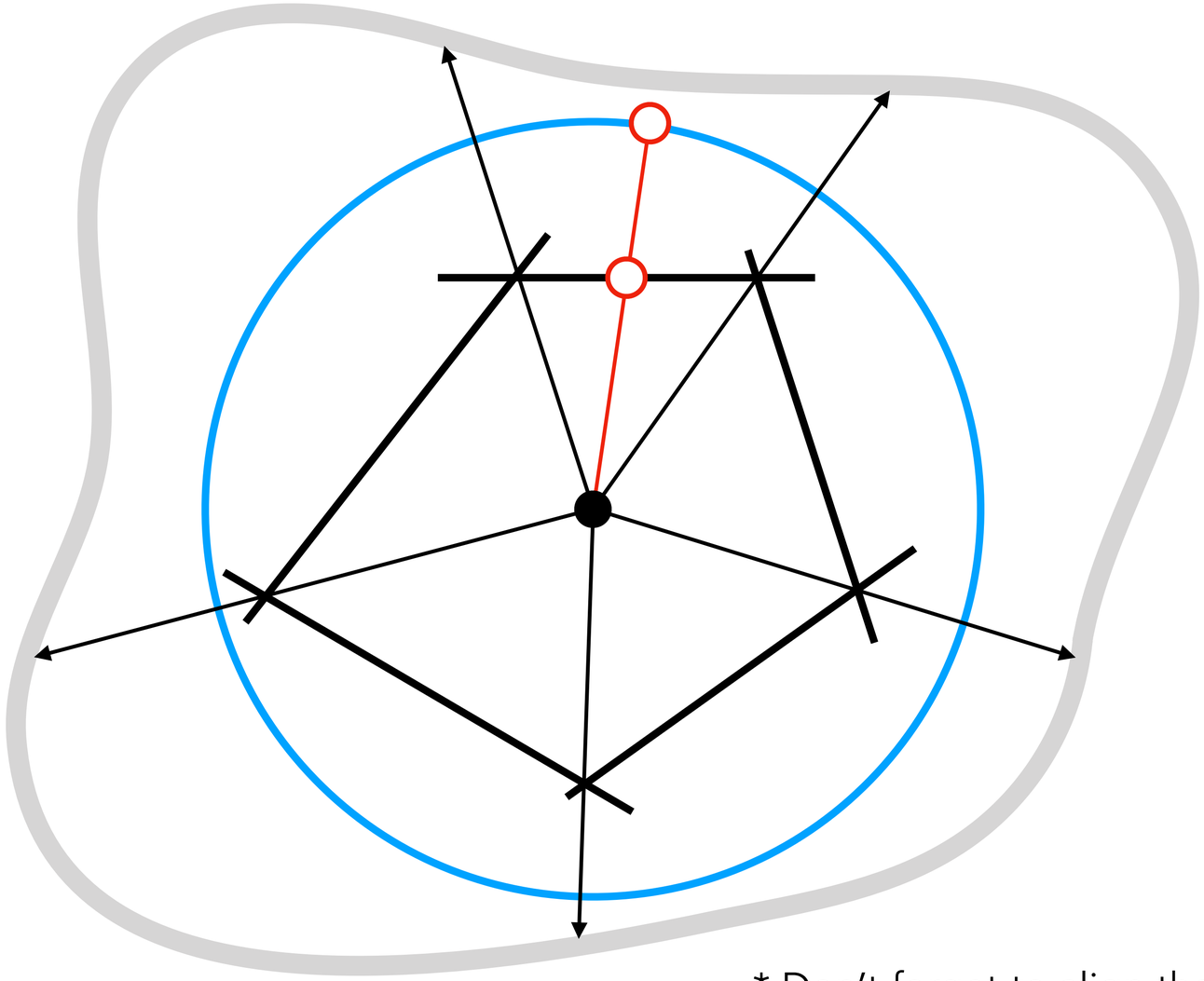
When stitching spans a large FOV, the projected image will look **distorted**. When the FOV is **over 180°**, it's **impossible** to map everything into one single plane!

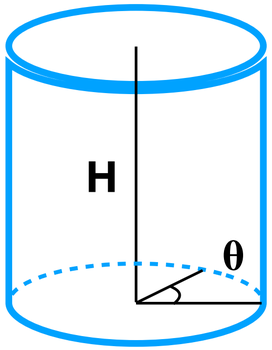




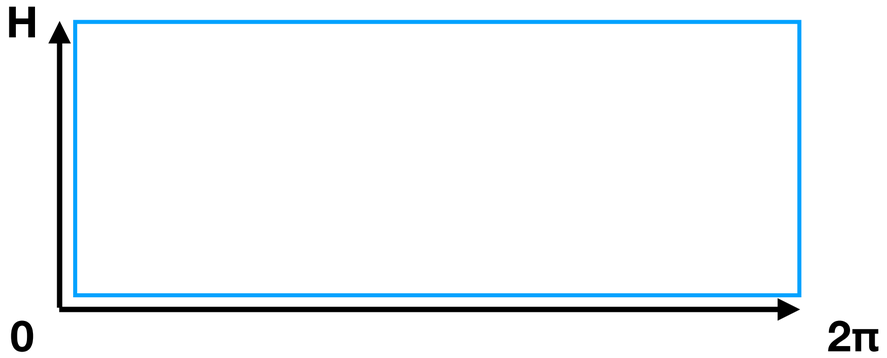
Example credit: Marc Levoy

Cylindrical Mapping





Every point is represented by a [h, **θ**] tuple. So a cylinder can be represented/stored as a conventional 2D image.



* Don’t forget to align the images first!

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

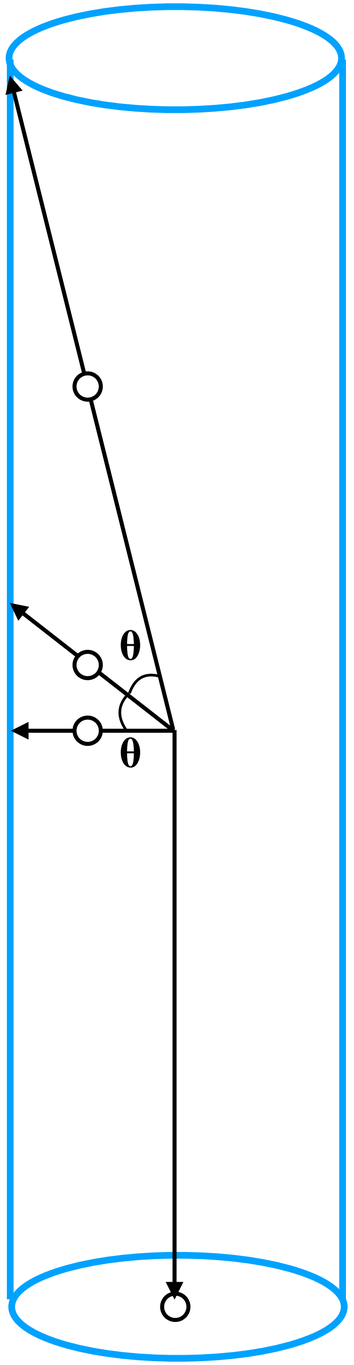
https://commons.wikimedia.org/wiki/File:Jonas_Daniel_Meijerplein_5_-_Amsterdam_-_Rijksmonument_2034_-_panorama_-_cylindrical.jpg

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

Issue 1: straight lines are not straight, because cylindrical projection is not a perspective projection.

Vertical and horizontal straight lines will still be mapped to straight lines, but other straight lines won’t be.

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



Issue 2: vertical distortion is significant near the poles and requires a very large cylinder.

- Extreme case: capturing the north and south poles need an infinitely large cylinder!
- Acceptable for a small vertical FOV.

When to Use Cylindrical Mapping

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

Usually cylindrical panorama is used for landscapes, where

- 1) there aren't many straight lines, and
- 2) we don't usually care about a full 180° vertical FOV (when you shoot panorama, how often do you care about the ground and the sky?)

Applications

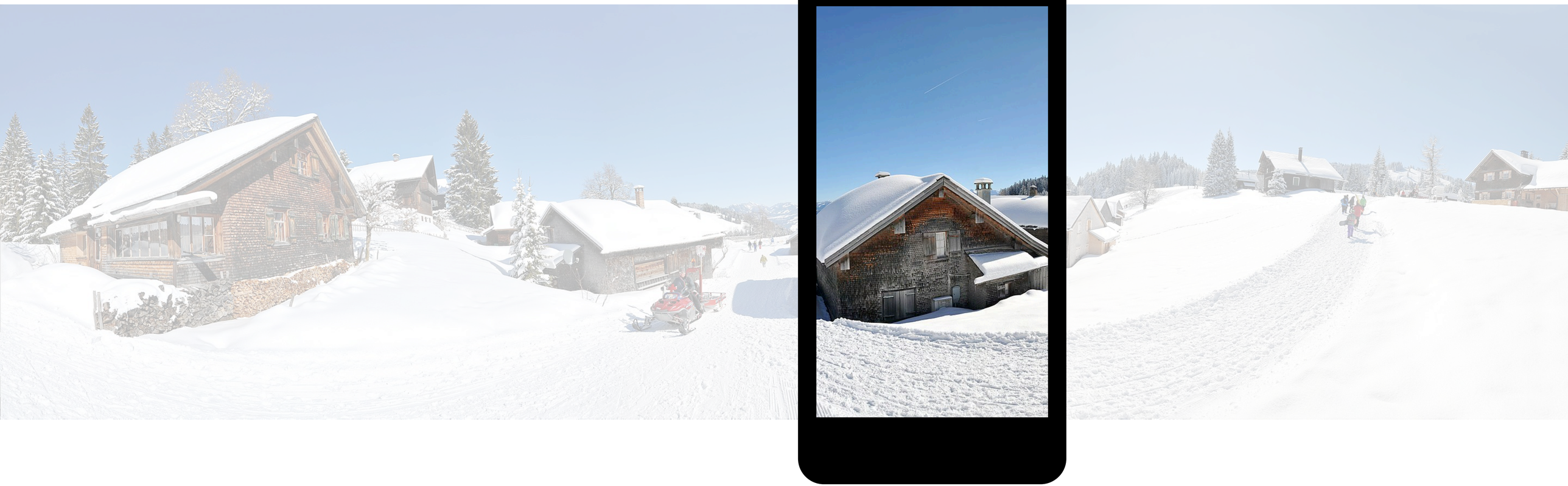
- now built into all mobile phones
- one simple camera sweep
- panorama computed on the fly

Facebook

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

Ben Claremont

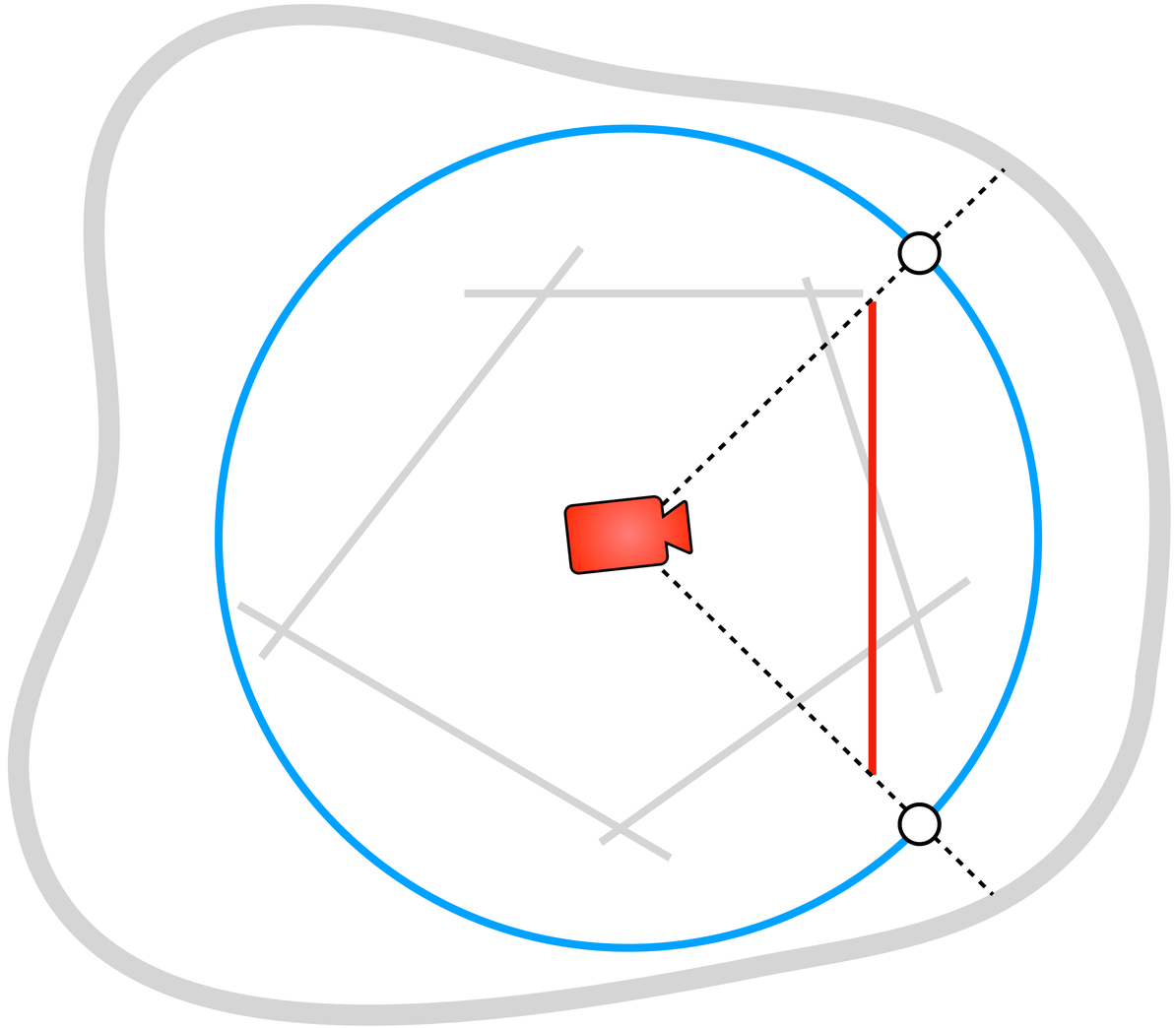
Obtain Small FOV From Cylindrical Panorama



Good idea?

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

Re-projecting Cylindrical Panorama to Small FOV

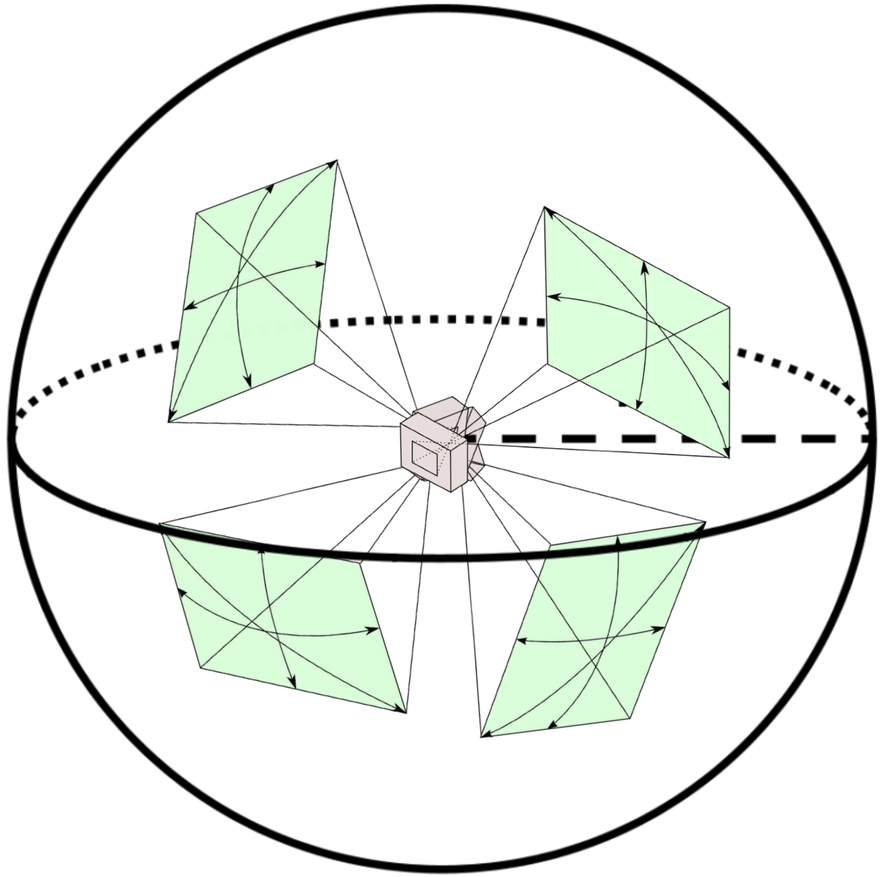


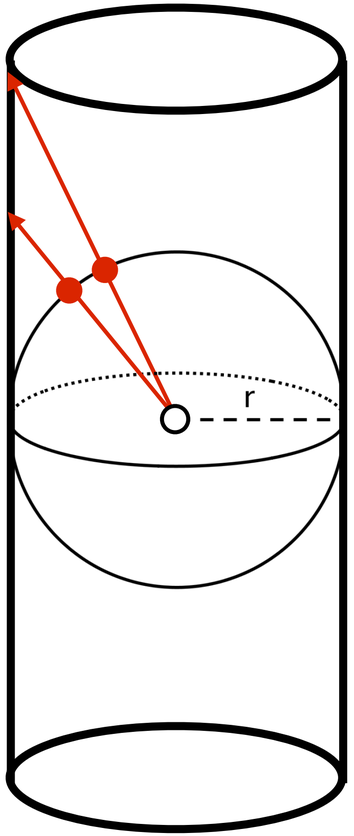
Distortion happens when 3D-2D projection is not perspective.

We could remove distortion by performing a perspective projection from the cylindrical plane to the sensor plane. This will generate an image that looks just like is taken by a camera with a FOV of $\boldsymbol{\theta}$.

- Panning across an image or VR viewing.

Spherical Mapping (for 360° Content)





Use spherical mapping to capture poles. Vertical distortion is smaller.

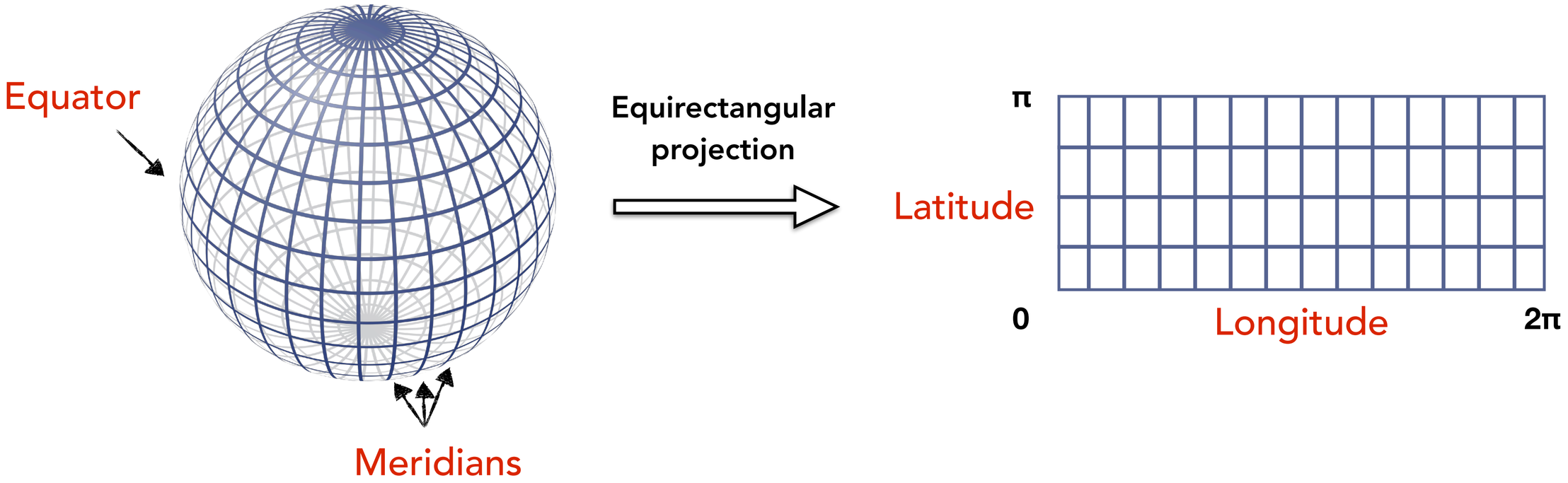
Spherical mapping is typically used for full 360° content, e.g., VR videos.

* Don’t forget to align the images first!

Storing Spherical Data

How to store a spherically mapped scene? That is, how to map a globe?

There are many **map projections**. Equirectangular is the most common.



https://www.artstation.com/marketplace/p/z16q/star-desert-360-degrees-panorama-backgrounds-10000-x-5000-pixels

Distortion in Equirectangular Projection

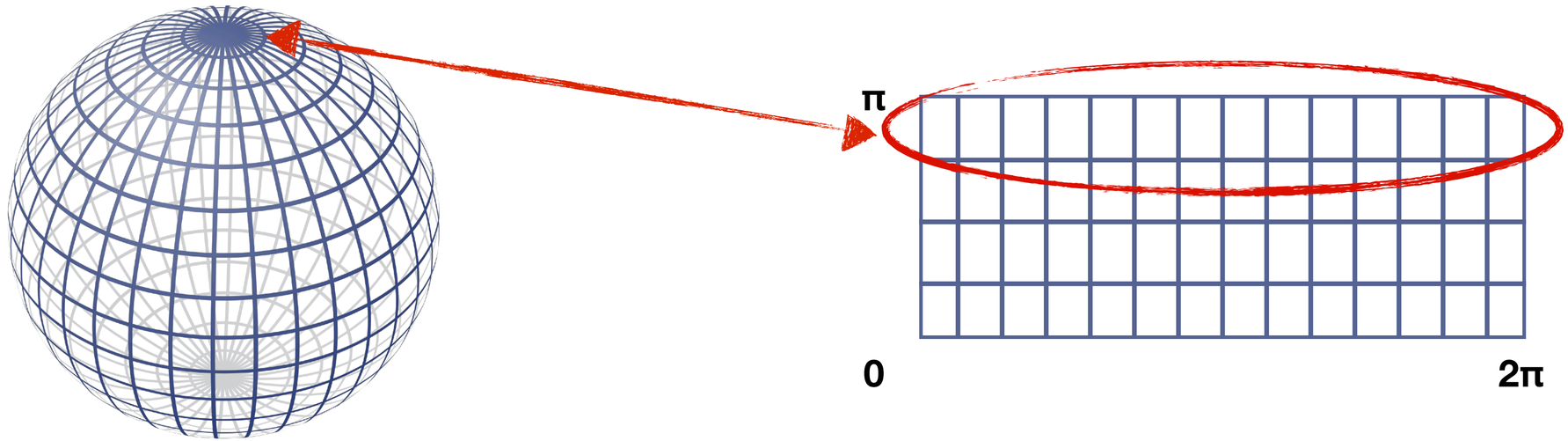
Issues With Equirectangular Projection

Lots of pixels are dedicated to the pole regions.

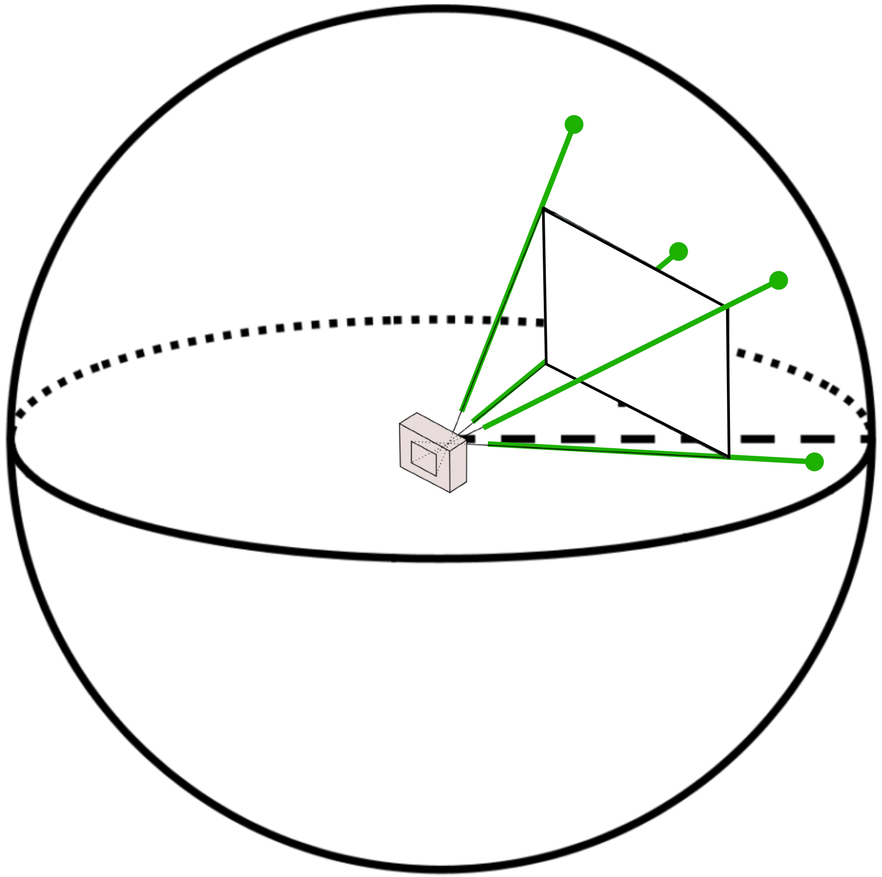
- But perhaps that's where we are least interested in typical scenes.

Pole areas are distorted.

- Makes video compression harder. Requires higher network bandwidth to stream.



Re-projecting Spherical Panorama

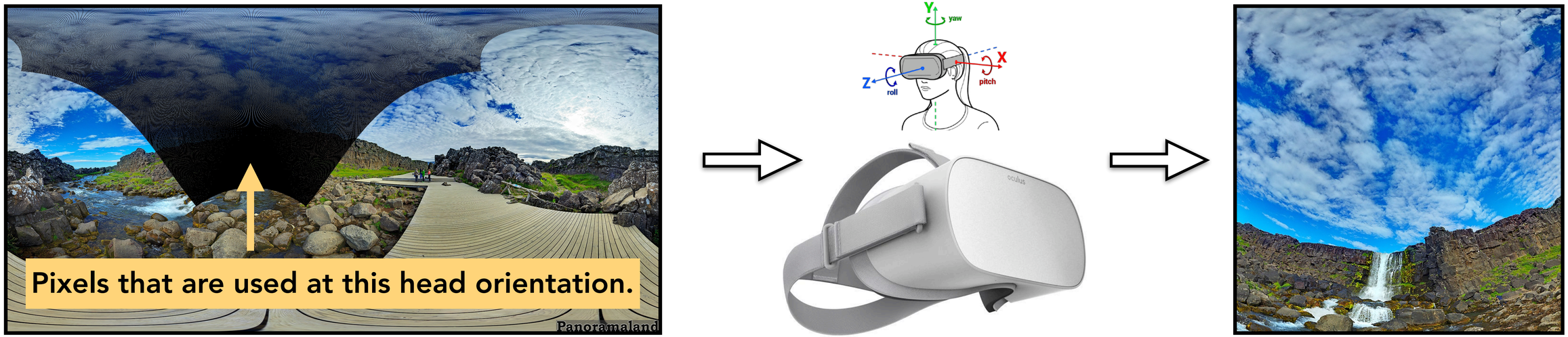


Spherically-mapped 360° content can’t be consumed as is — too much distortion.

When viewing 360° content in a VR headset or panning across the panorama, re-project the part of the sphere that’s under the current FOV.

- Similar to re-projection done for cylindrical panorama, except there re-projection is optional.

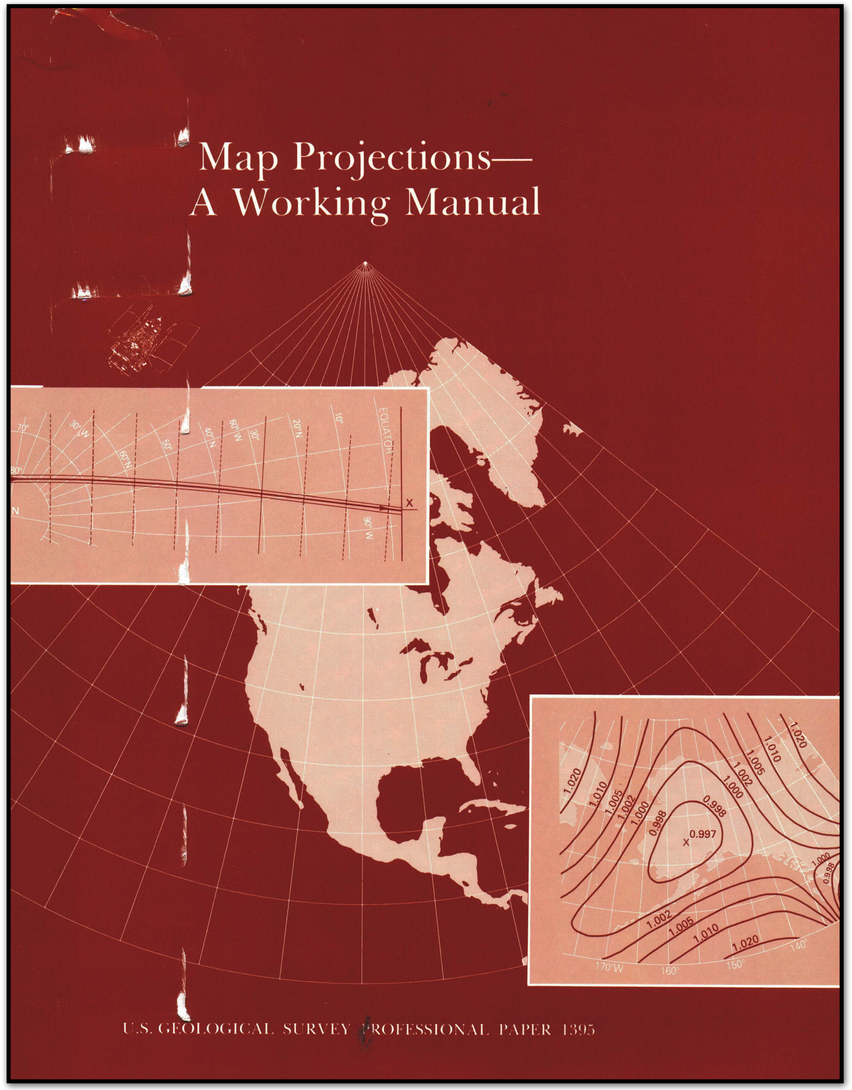
Re-projecting Spherical Panorama

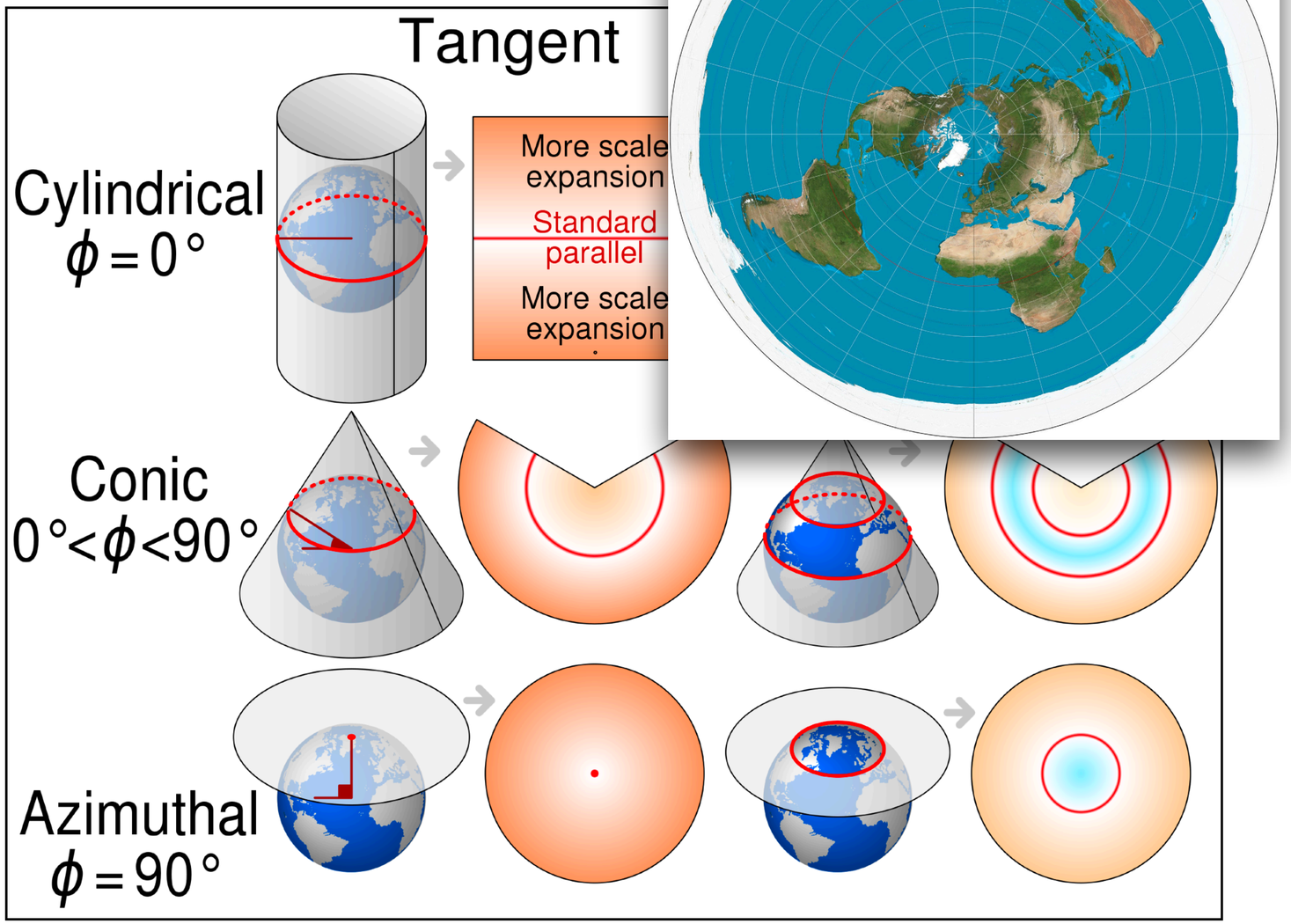


This re-projection is a main performance bottleneck.

- In conventional video playback, once the video frames are decoded/uncompressed they can be directly displayed.
- VR video playback requires an extra step. In practice re-projection is implemented as texture mapping in GPU (more later).

Other Cartographic Map Projections



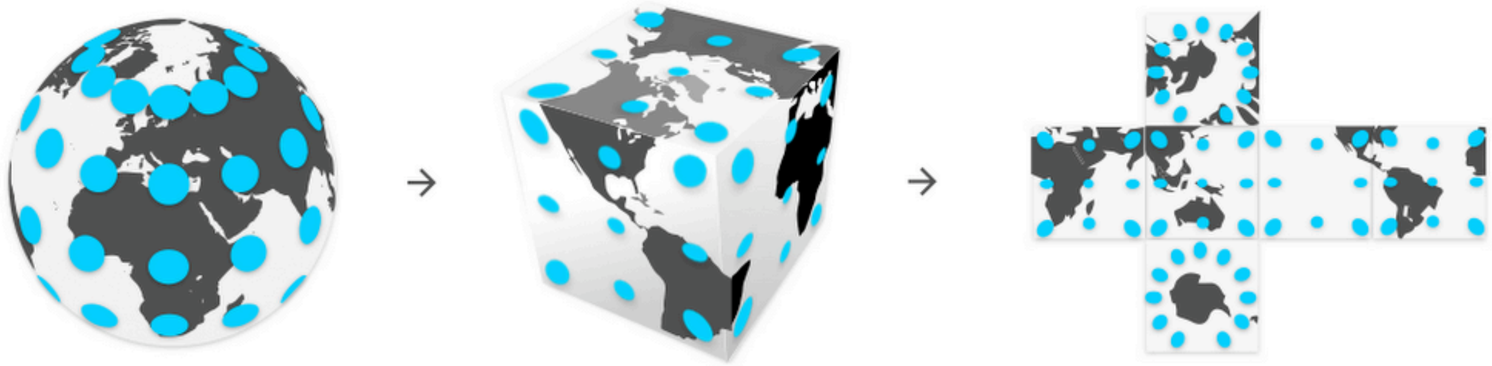


https://en.wikipedia.org/wiki/Map_projection

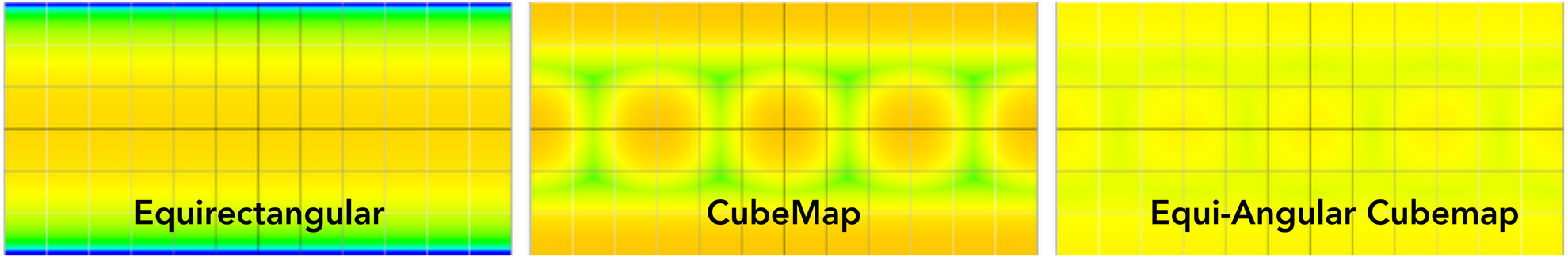
Other Map Projections for Storing VR Content

Goal is to allocate pixels evenly across the sphere and minimize distortion.

CubeMap projection

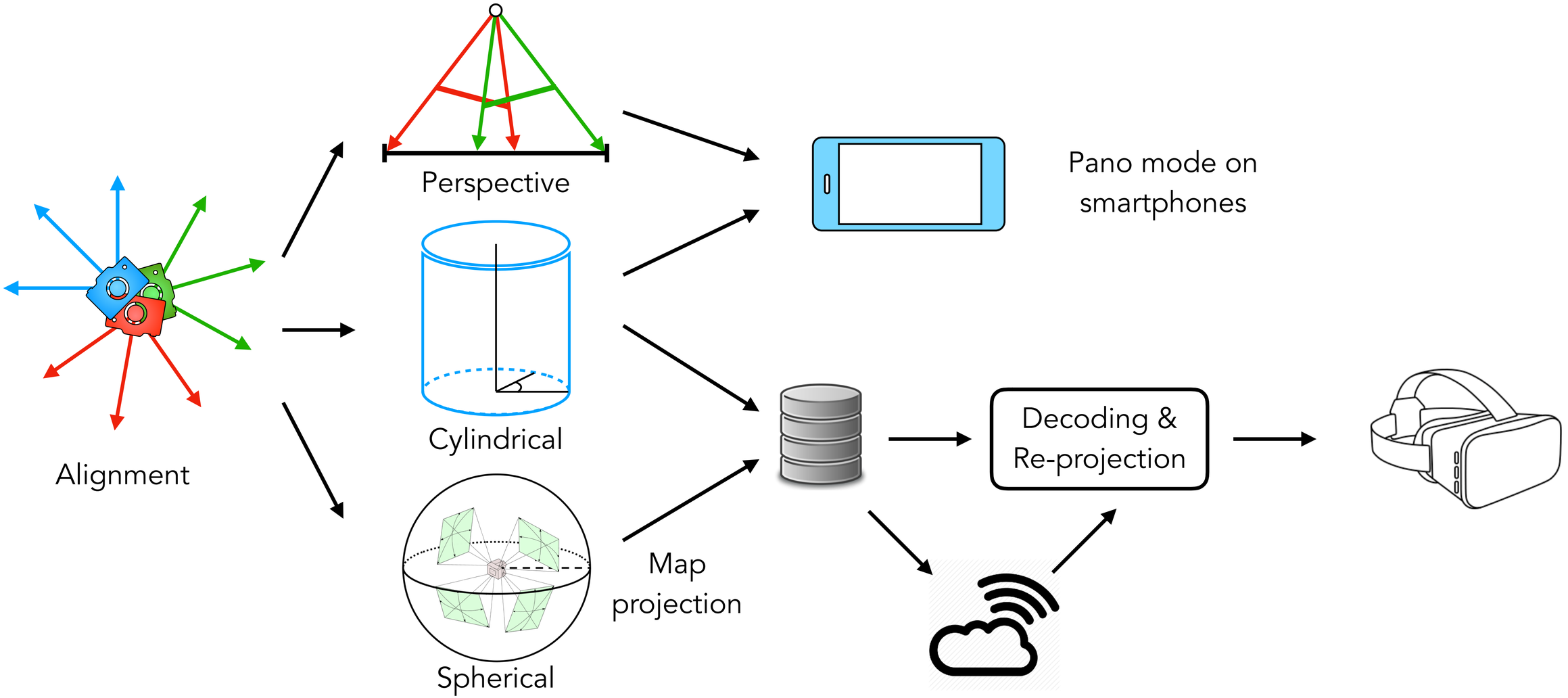


Color indicates density

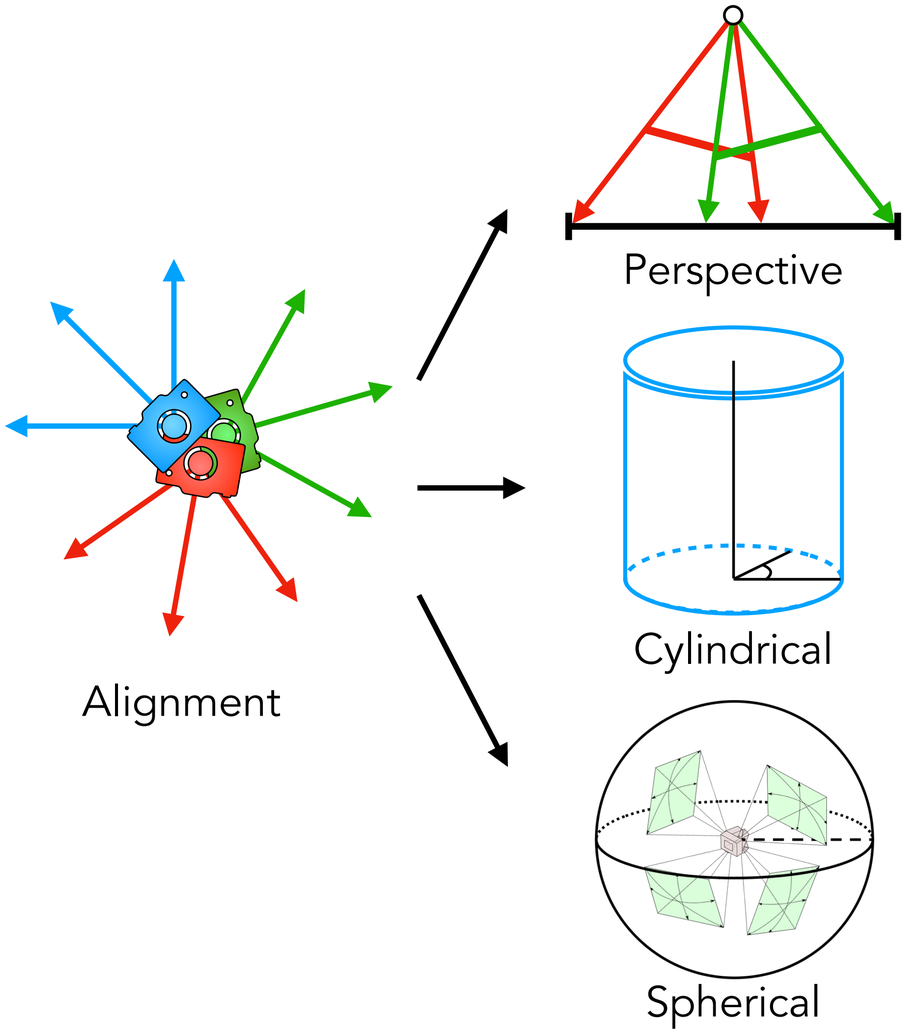


https://blog.google/products/google-vr/bringing-pixels-front-and-center-vr-video/

Recap (So Far)


Perspective
Cylindrical
Spherical
Alignment
Map
projection
Pano mode on
smartphones
Decoding &
Re-projection

What Do They Have in Common?

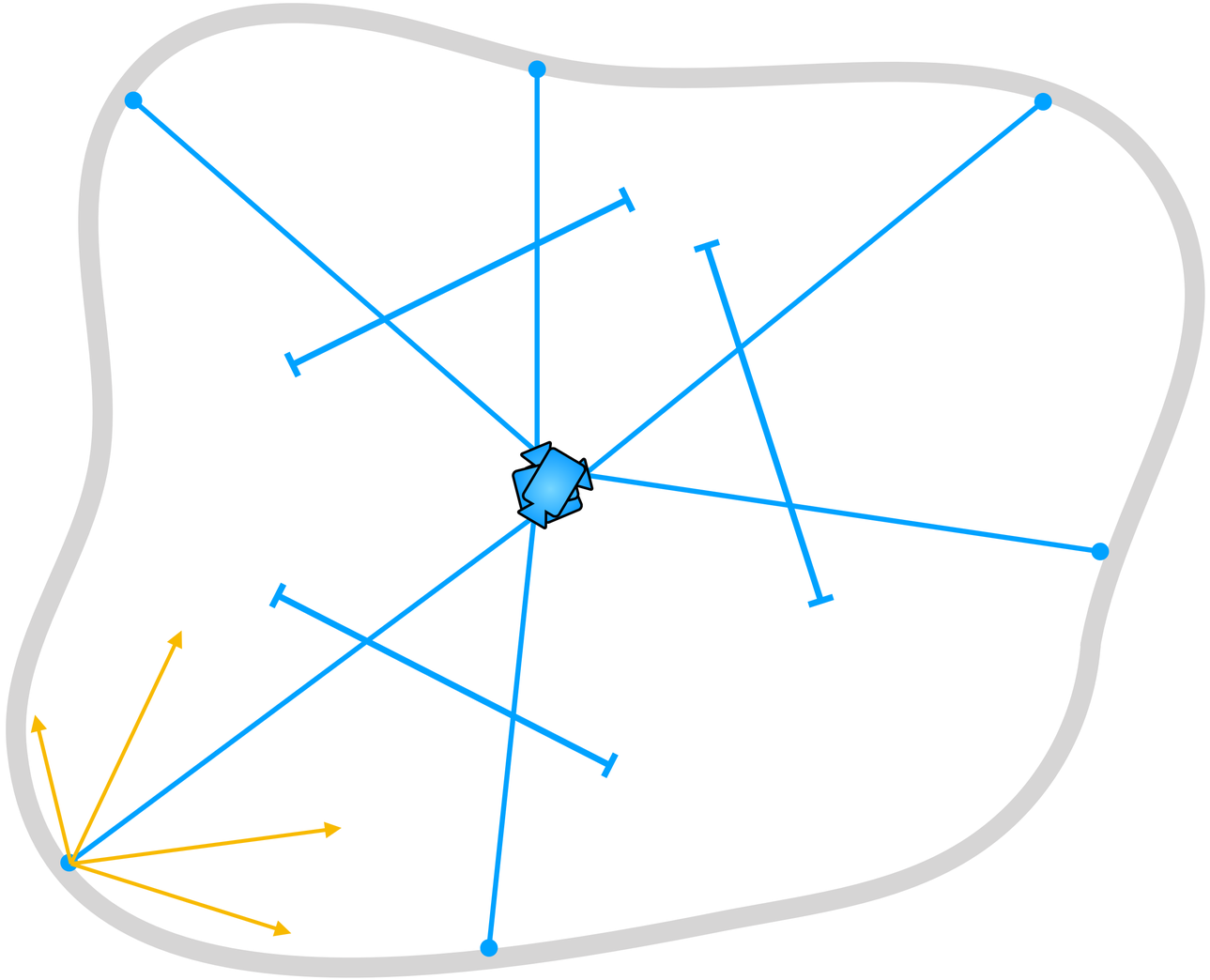


They have the same goal: **simulating a virtual camera** based on information from real camera captures.

To do that, it's all about capturing **rays** and calculating radiance of individual rays so that we can simulate a virtual camera.

What rays have we been tracing?

One Ray for Each Scene Point

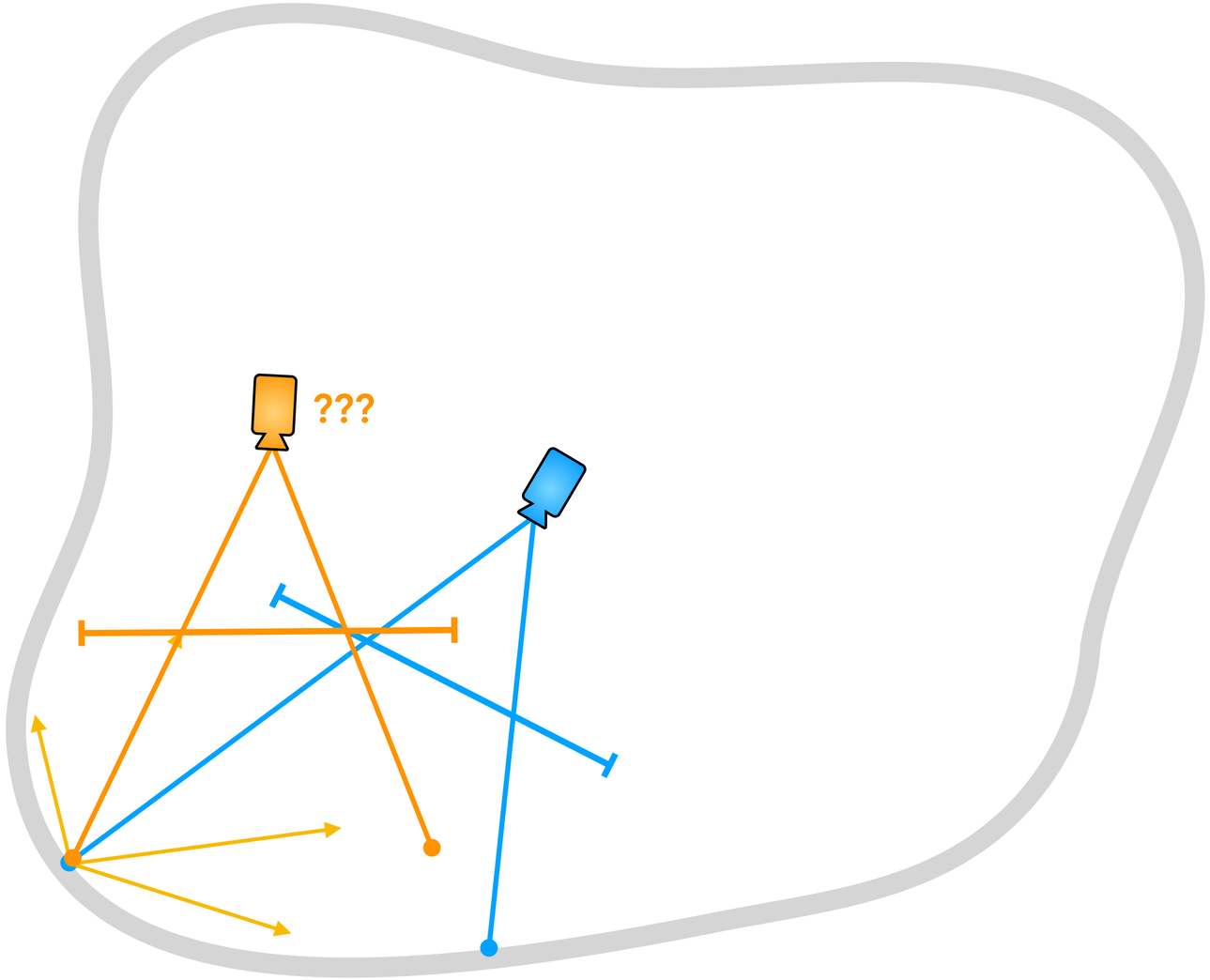


For each scene point, we “calculate”/need the radiance of just one ray from that point.

- Assuming pinhole camera
- The radiance is calculated from real camera captures
- Interpolate between points

Why one ray per point? The camera position doesn’t change; it just rotates, no translation.

3 DOF So Far During Playback (Rendering)



Because we trace one ray from each point, we can't simulate virtual camera at other positions.

- During playback, the headset tracks head **rotation**, but **not translation**.

Rotation and translation is parameterized by three Degrees of Freedom (DoF) each.

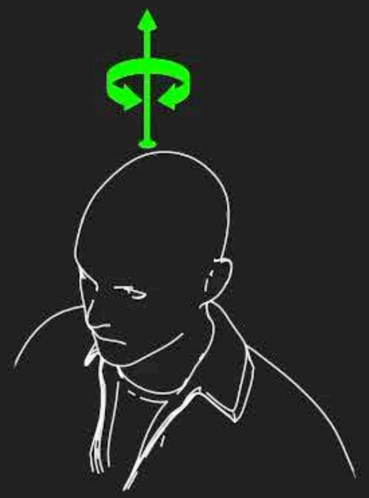
- So this is 3 DOF.

Recall: Motion = Translation + Rotation

Responsible for rotation

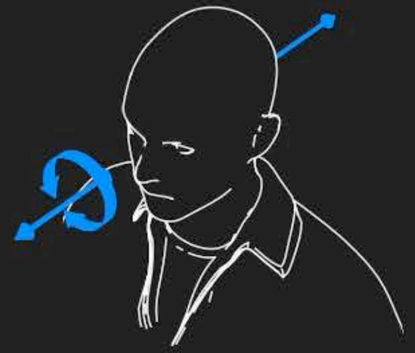
$$[x, y, z, 1] \times \begin{bmatrix} T_{00} & T_{01} & T_{02} & 0 \\ T_{10} & T_{11} & T_{12} & 0 \\ T_{20} & T_{21} & T_{22} & 0 \\ T_{30} & T_{31} & T_{32} & 1 \end{bmatrix} = [x', y', z', 1]$$

Responsible for translation

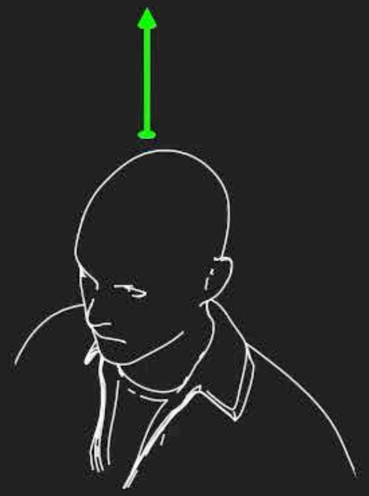


Y AXIS YAW

X AXIS PITCH

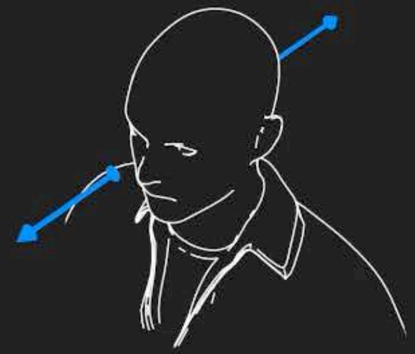


Z AXIS ROLL



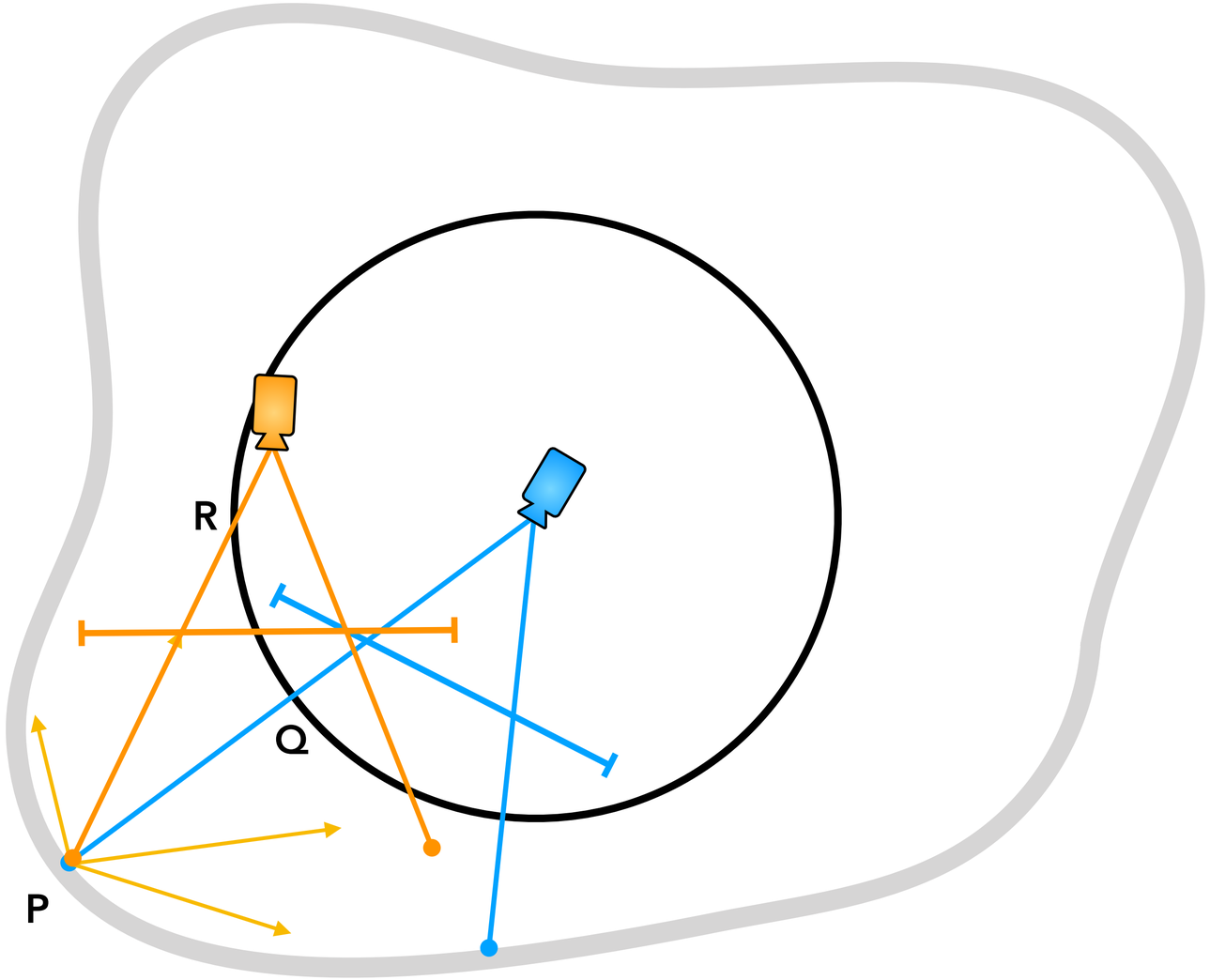
Y AXIS UP/DOWN

X AXIS LEFT/RIGHT



Z AXIS FRONT/BACK

A Non-Solution

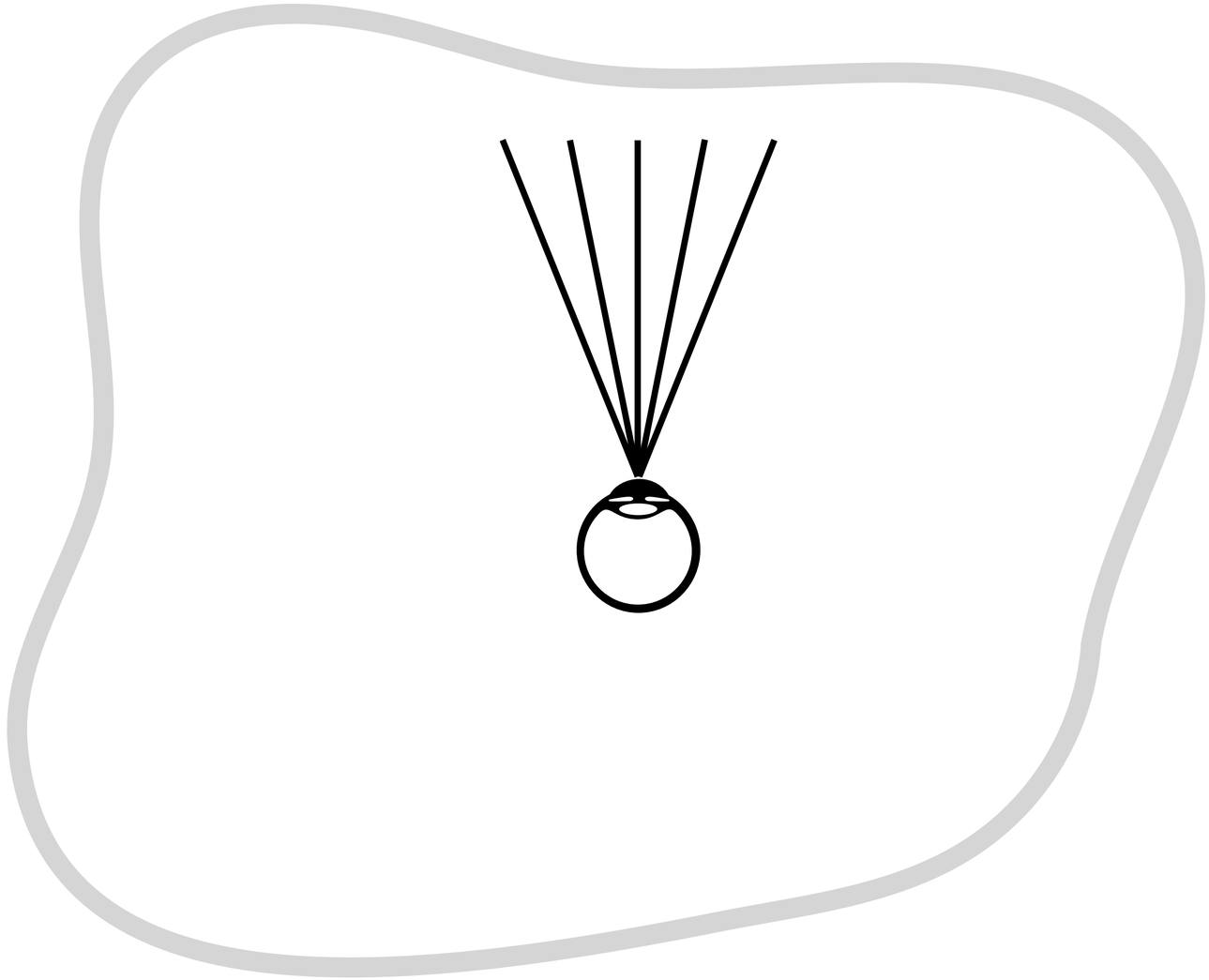


Can’t we use Q to approximate R — assuming P is diffuse?

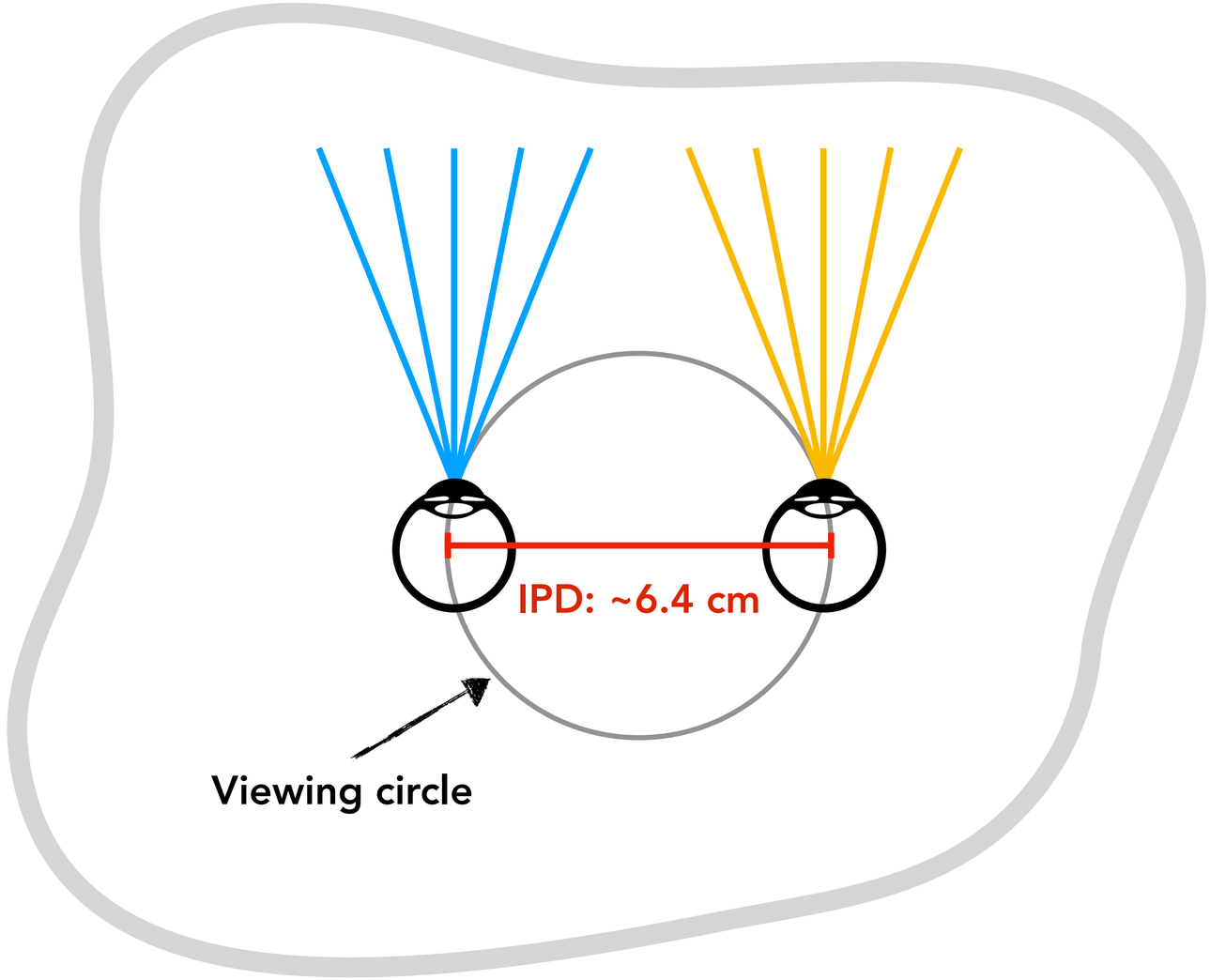
The exact pixel position of Q (in the panorama) depends on the depth of P, which at rendering time we don’t know (not stored in the panorama).

- Unless we perform a full-blown 3D reconstruction of the scene (photogrammetry; later)

Stereoscopic 360° (Limited/Fixed Translation)



Stereoscopic 360° (Limited/Fixed Translation)

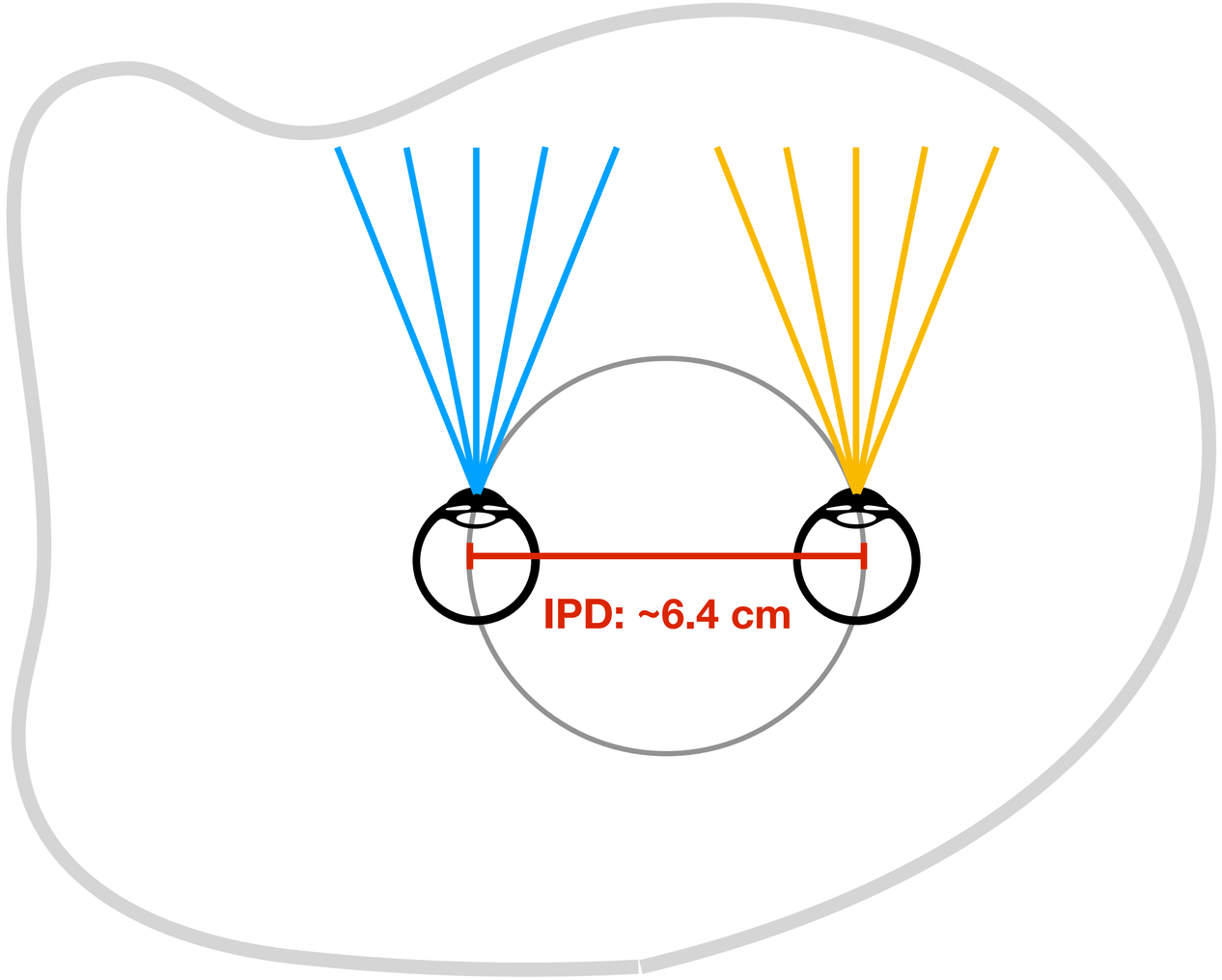


Wheatstone Stereoscope (1838)

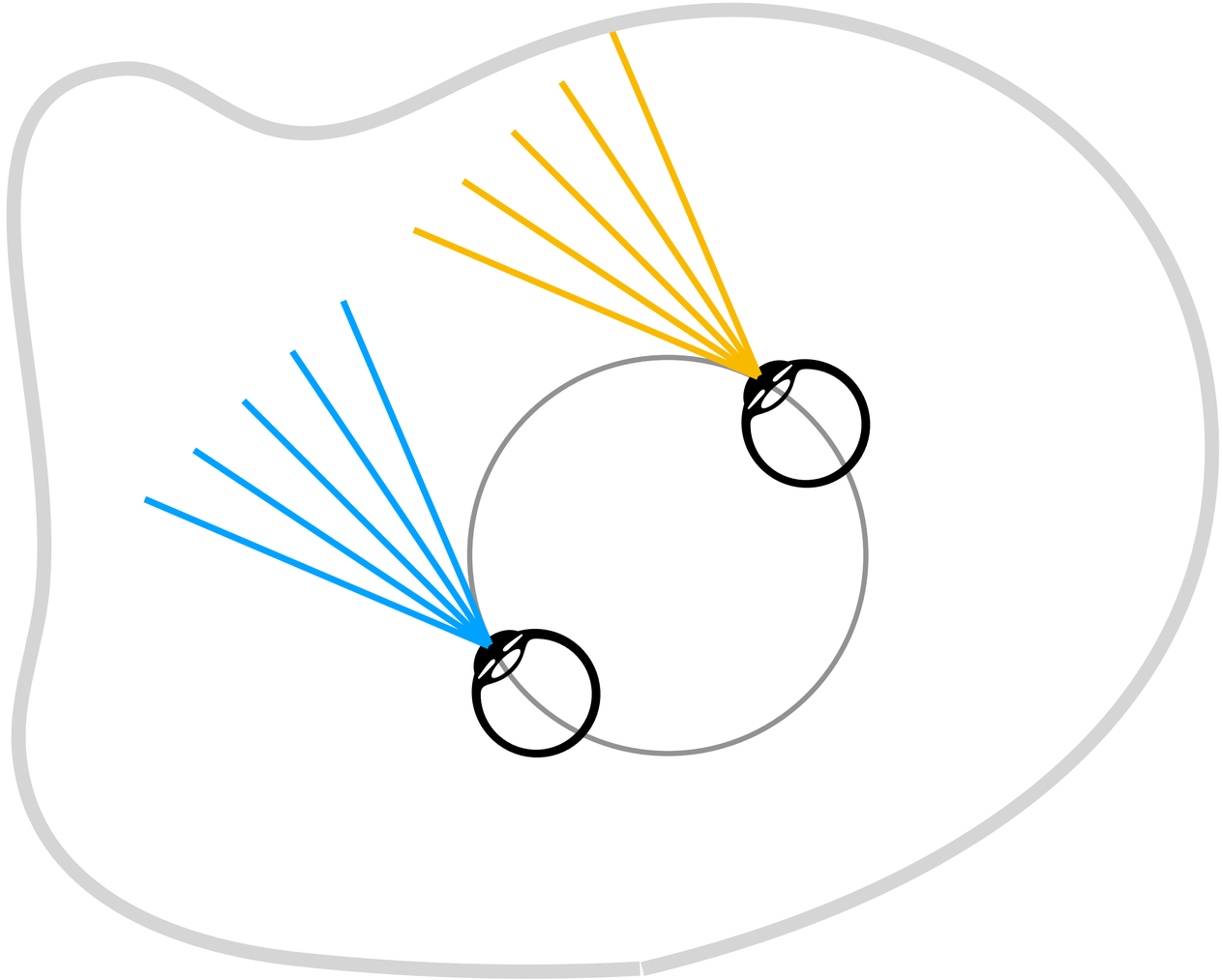
Brewster Stereoscope (1849)

David Brewster
(1781–1868)

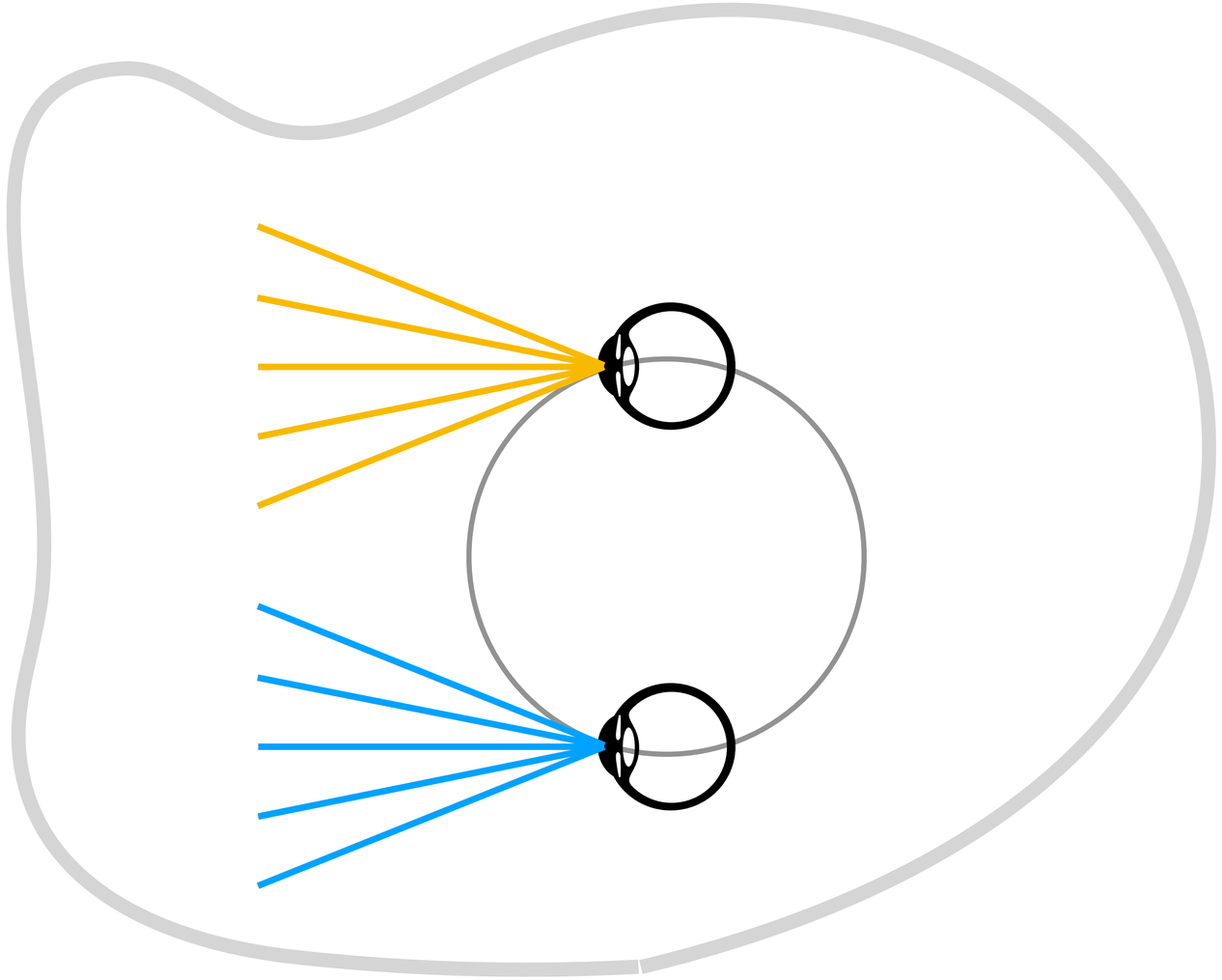
Stereo VR Video Acquisition



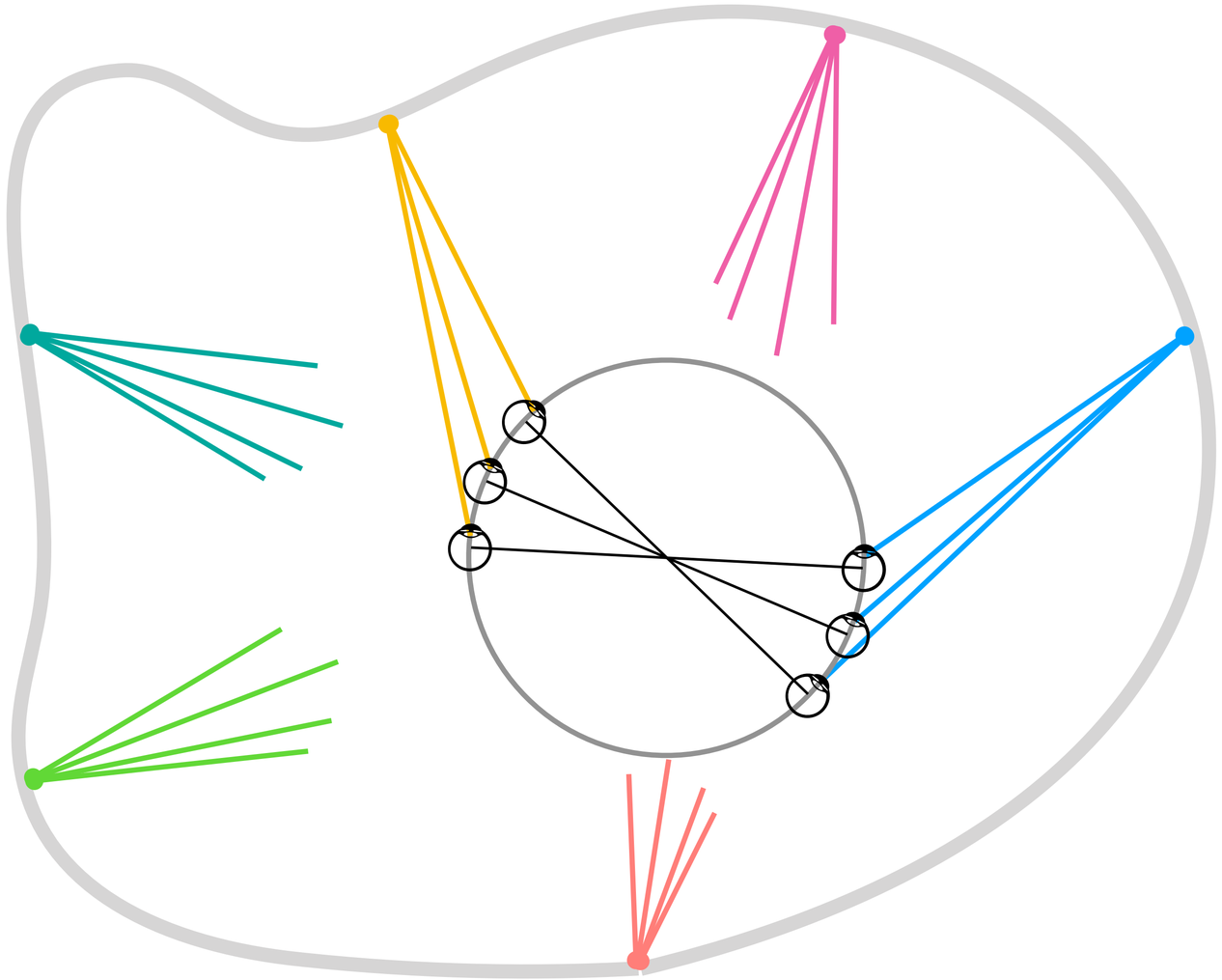
Stereo VR Video Acquisition



Stereo VR Video Acquisition



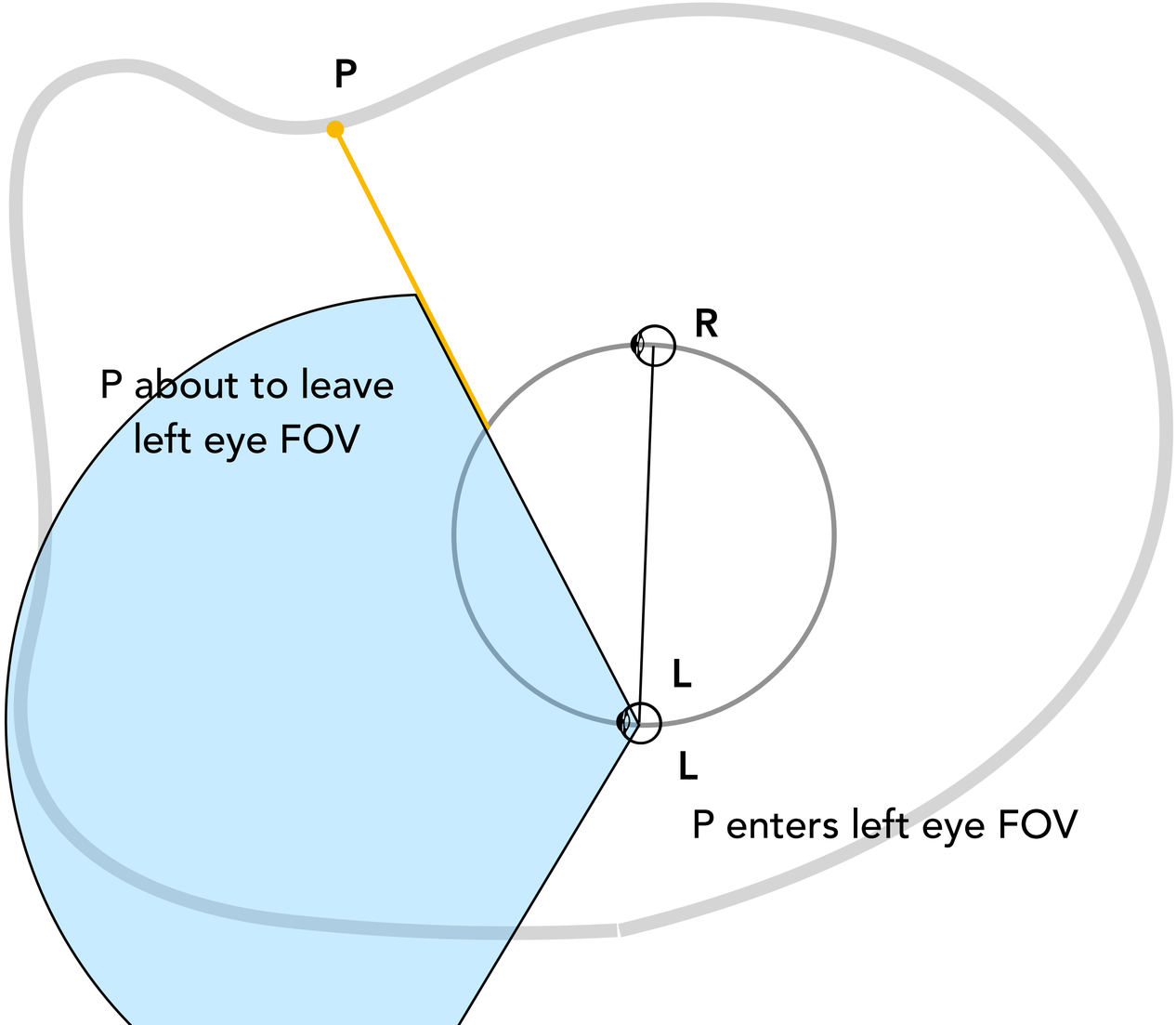
Need to Capturing Multiple Rays from a Point



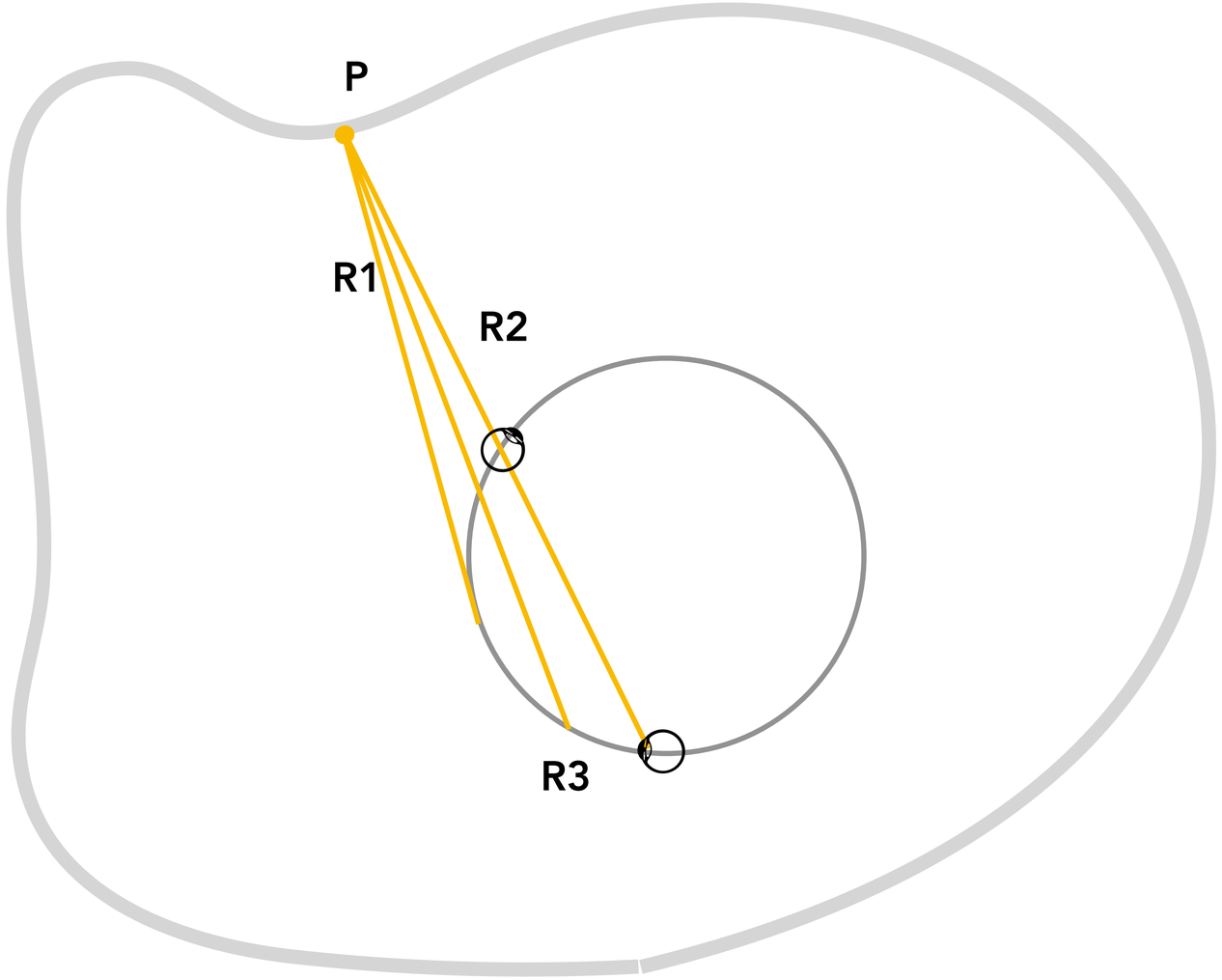
Many rays to trace (calculate radiance) and to store.

- Would increase content creation time (even if it's done offline).
- Perhaps more important, it would also increase **storage** overhead. Instead of storing one single color/radiance for each point, we now have to store many colors/radiances.

Which Rays Do We Need?



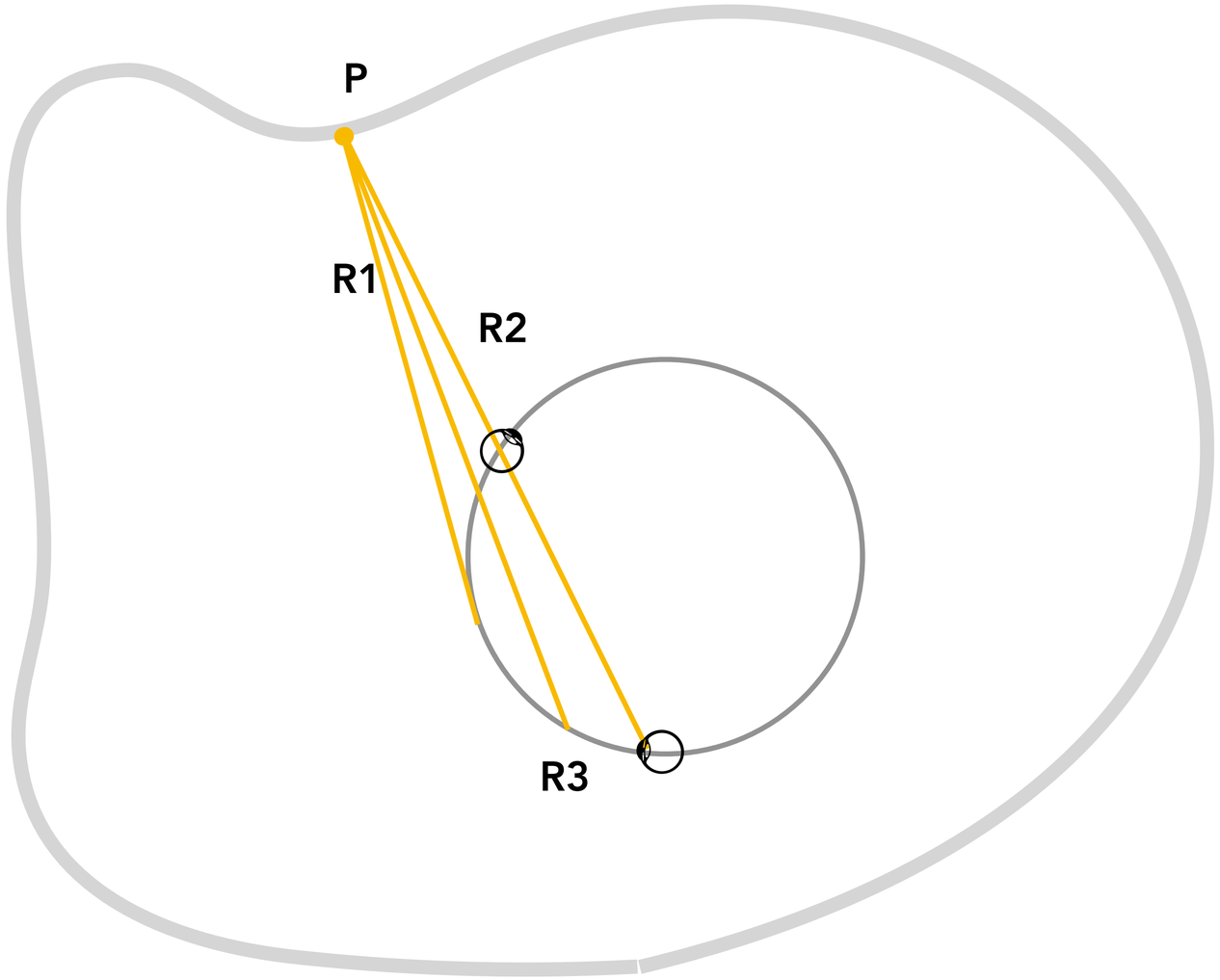
Which Rays Do We Need?



A small bundle of rays between R1 and R2 are ideally what we need to capture.

- Other rays from P won't be seen by the left eye.
- R1 is tangential to the viewing circle.

Ray Approximation

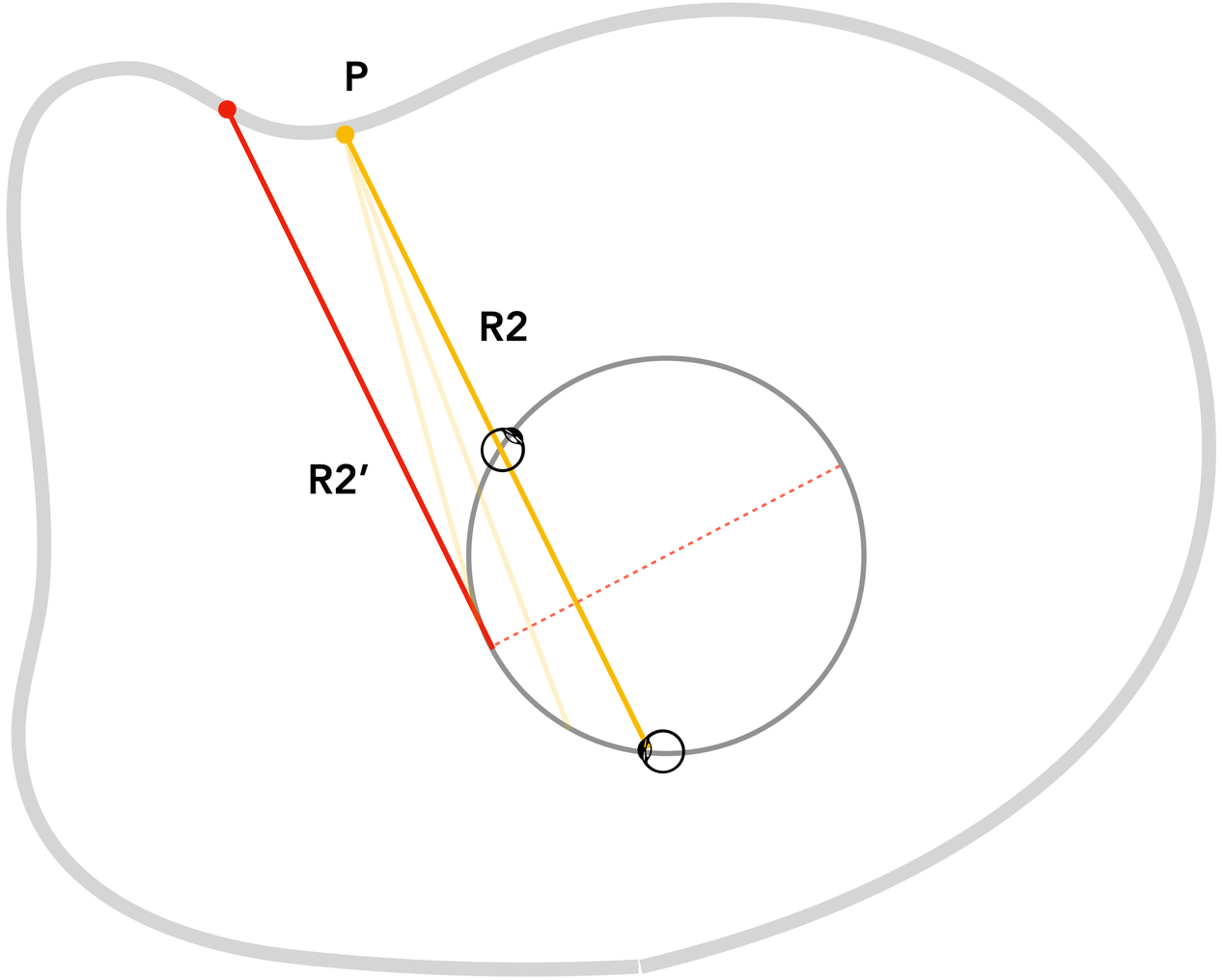


We approximate each ray by a parallel ray that:

- originates from the viewing cycle
- is tangential to the viewing cycle

R1 is not approximated since it’s already a tangential ray.

Ray Approximation



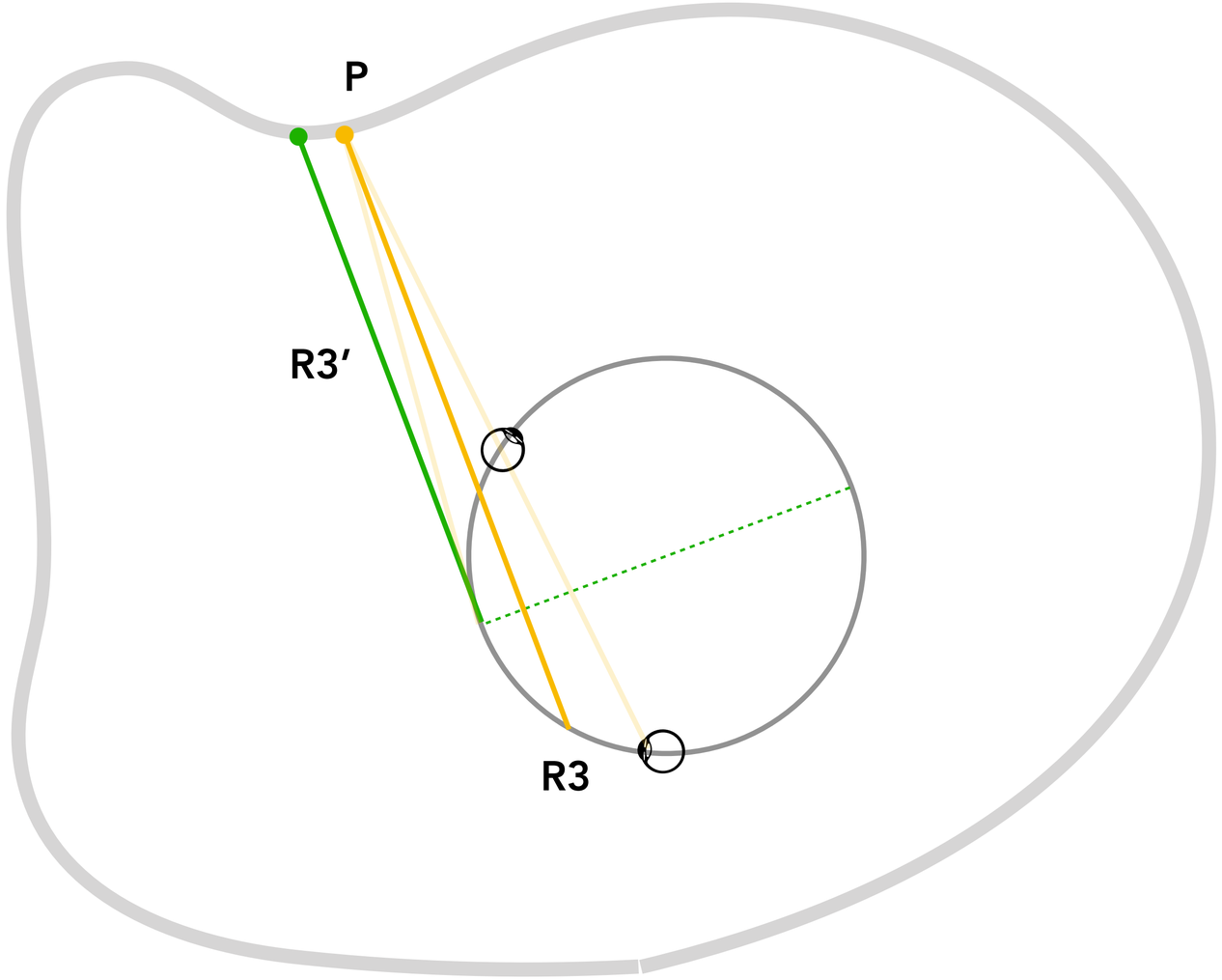
We approximate each ray by a parallel ray that:

- originates from the viewing cycle
- is tangential to the viewing cycle

For instance:

- R2 is approximated by R2’

Ray Approximation



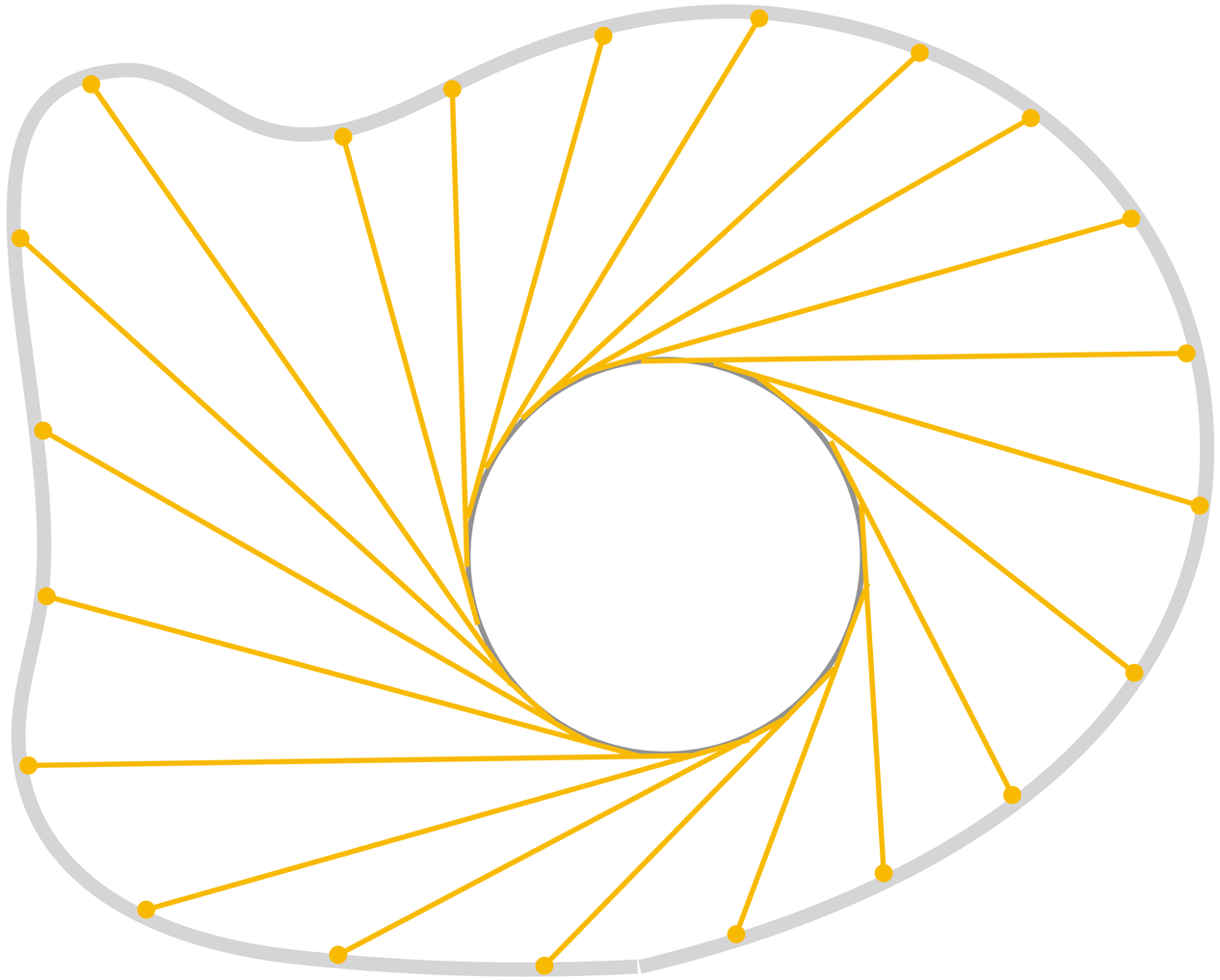
We approximate each ray by a parallel ray that:

- originates from the viewing cycle
- is tangential to the viewing cycle

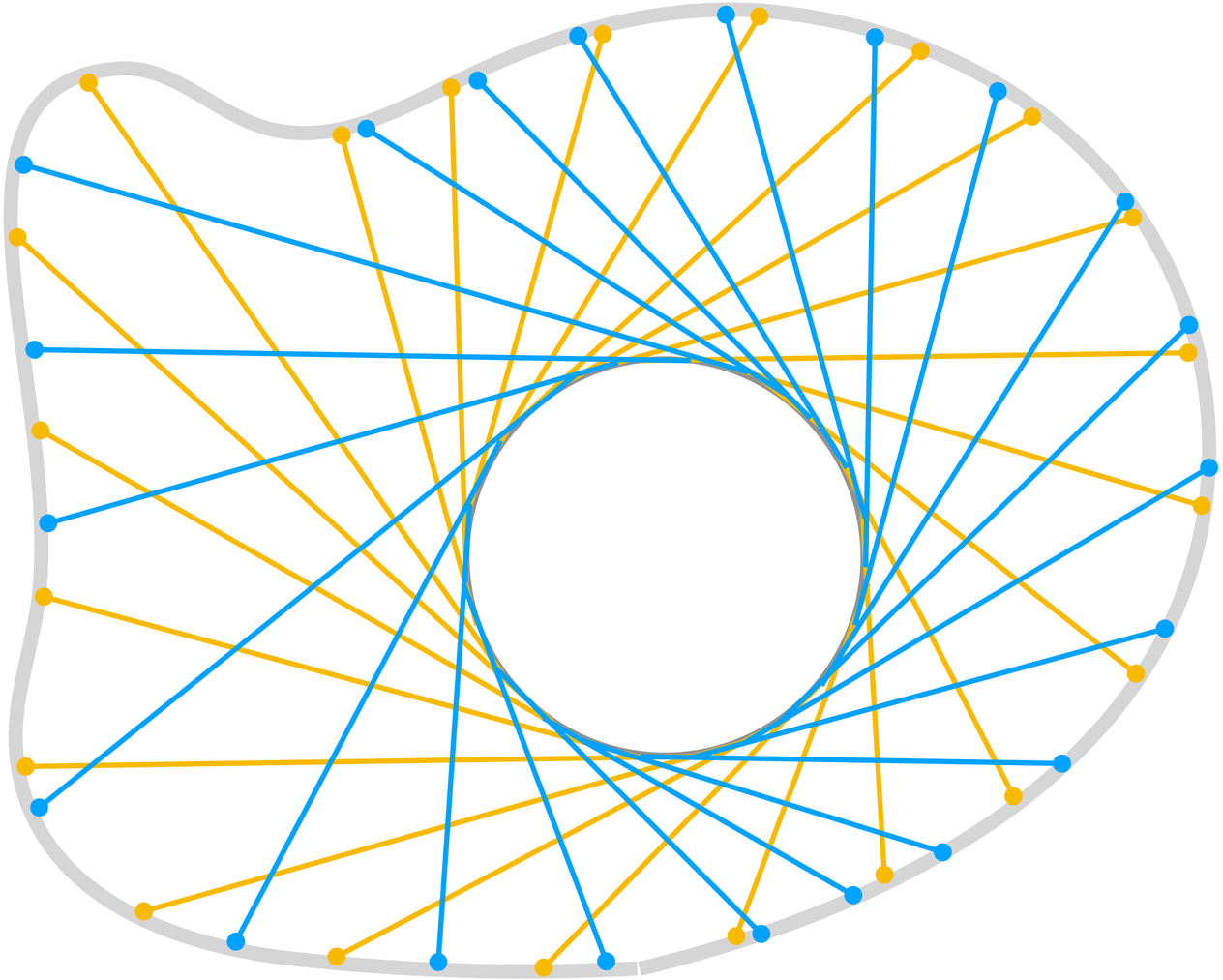
For instance:

- R2 is approximated by R2’
- R3 is approximated by R3’

Omni-Directional Stereo (ODS) Capture



Omni-Directional Stereo (ODS) Capture



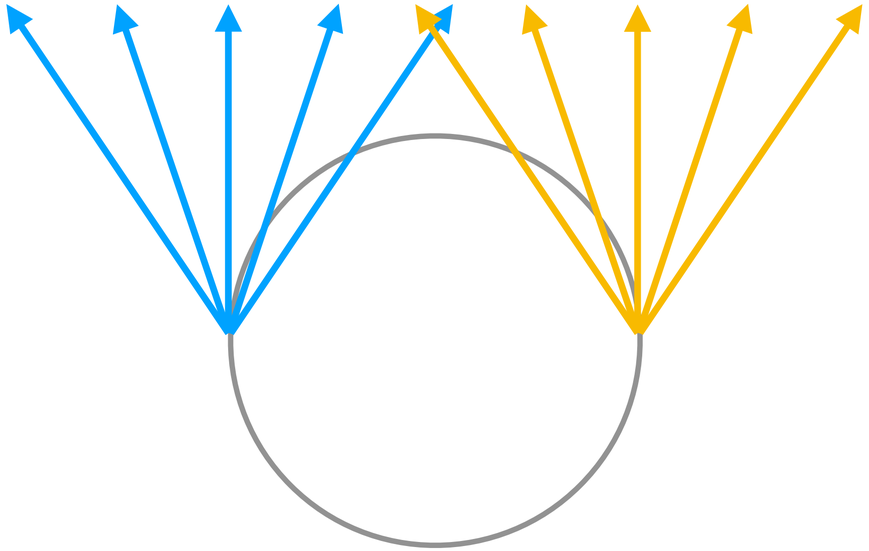
Sample rays tangential to the viewing circle for both eyes.

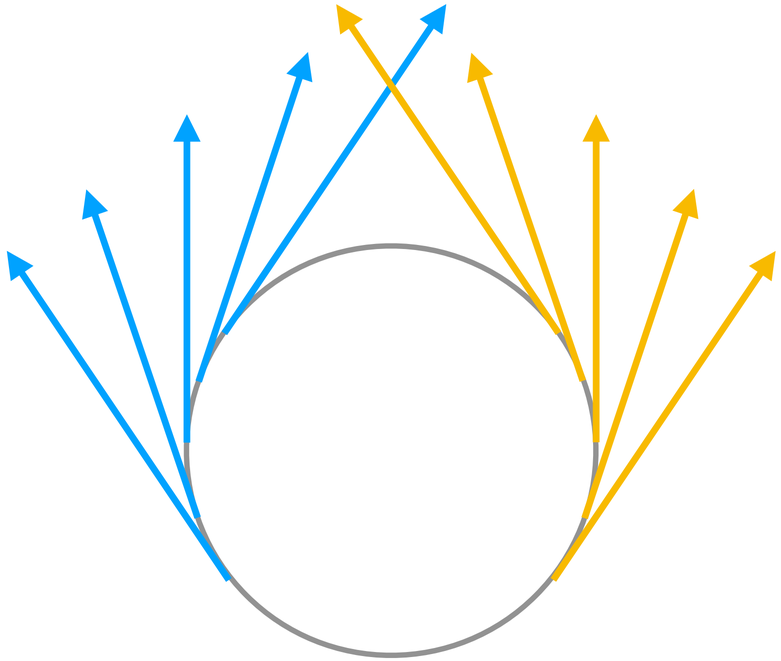
If sampled with infinite resolution, any ray in the scene can be captured/ approximated.

ODS == Ignoring Ray Origin

Technically ray = original + direction. In ODS rendering, all rays that have the same direction are approximated as the same ray regardless of directions.

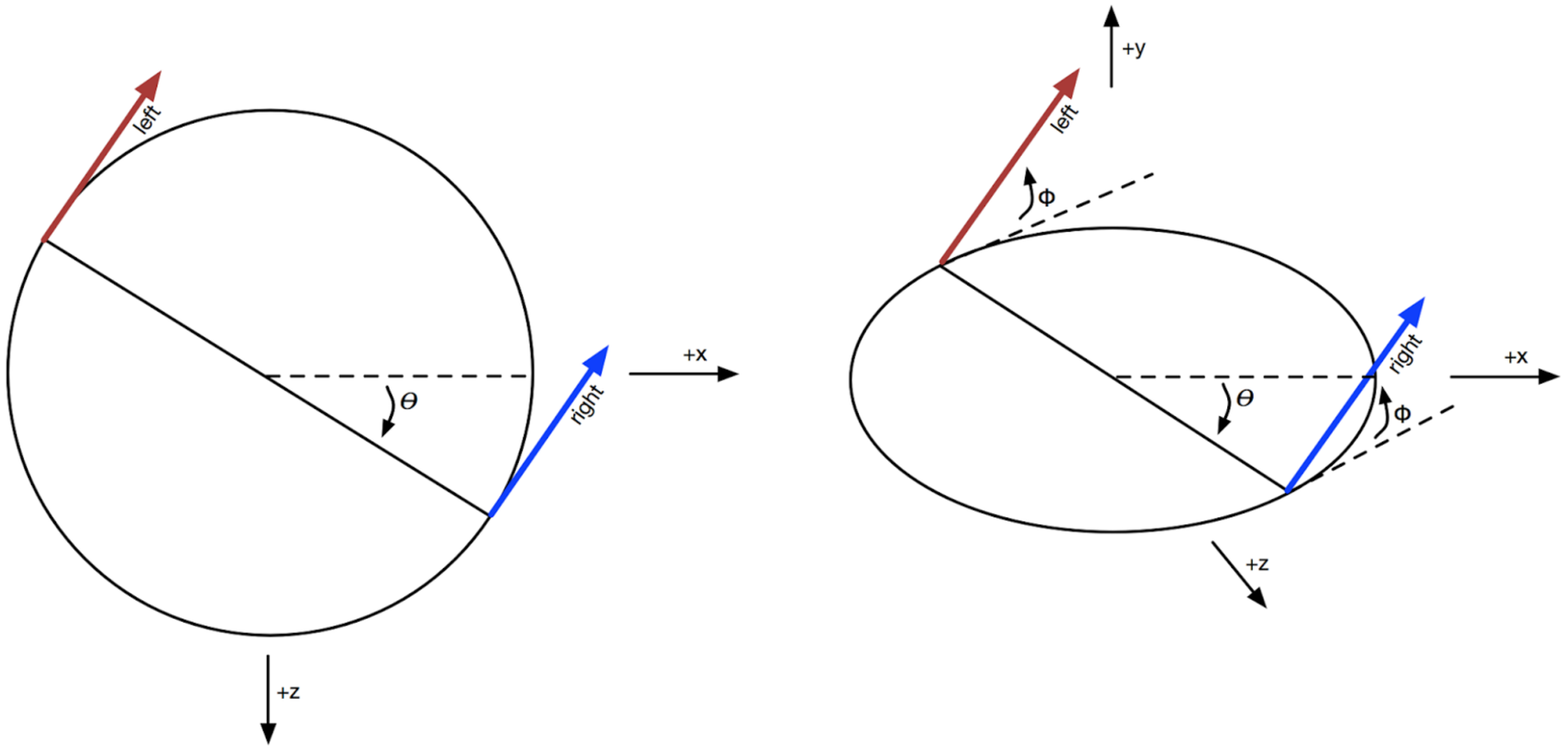
- Ignore the ray origin and consider only the direction





At Capturing Time

The captured (tangential) rays always originate from the circle where y = 0. At each point on the circle, we sweep ϕ from $-\pi/2$ to $-\pi/2$



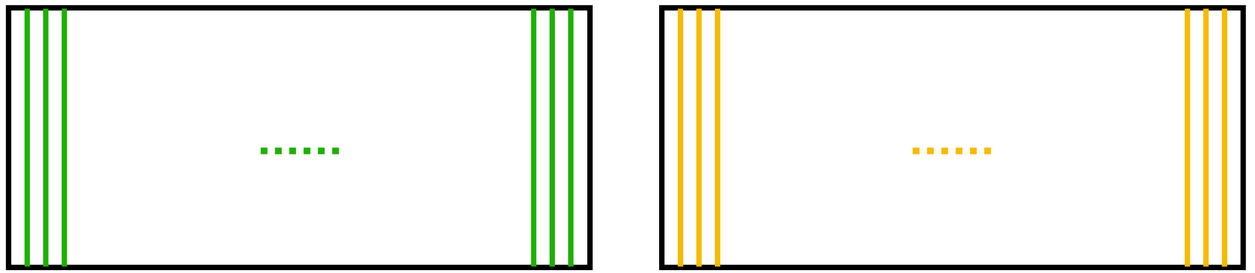
https://developers.google.com/vr/jump/rendering-ods-content.pdf?authuser=0

Storing ODS VR Content

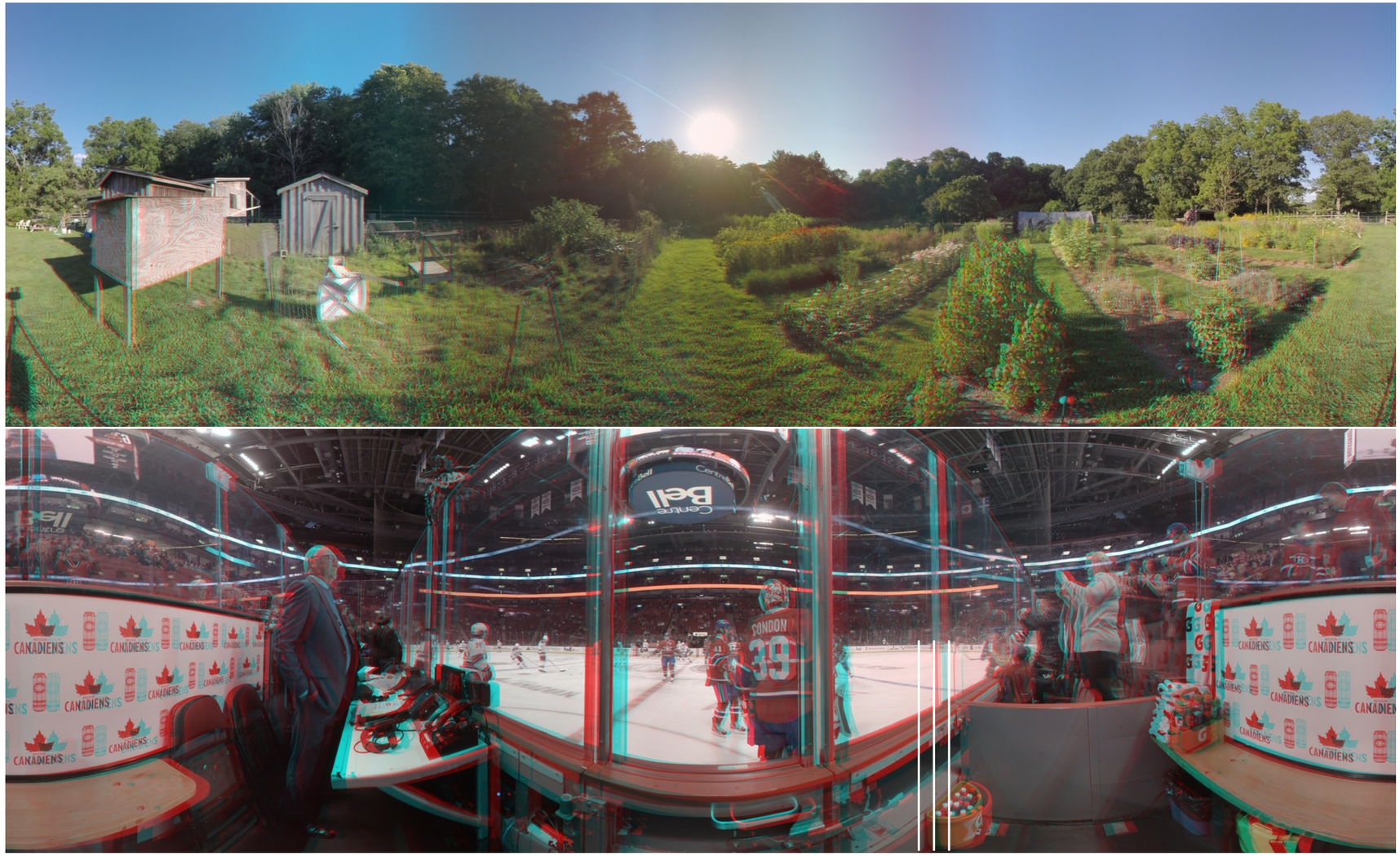
One panorama for each eye (e.g., equirectangular projection).

The final videos are stored in conventional format and can be compressed and delivered using existing systems.

- VR video compression and delivery are active areas for research.



Storing ODS VR Content



Overlaying left and right images as an anaglyph

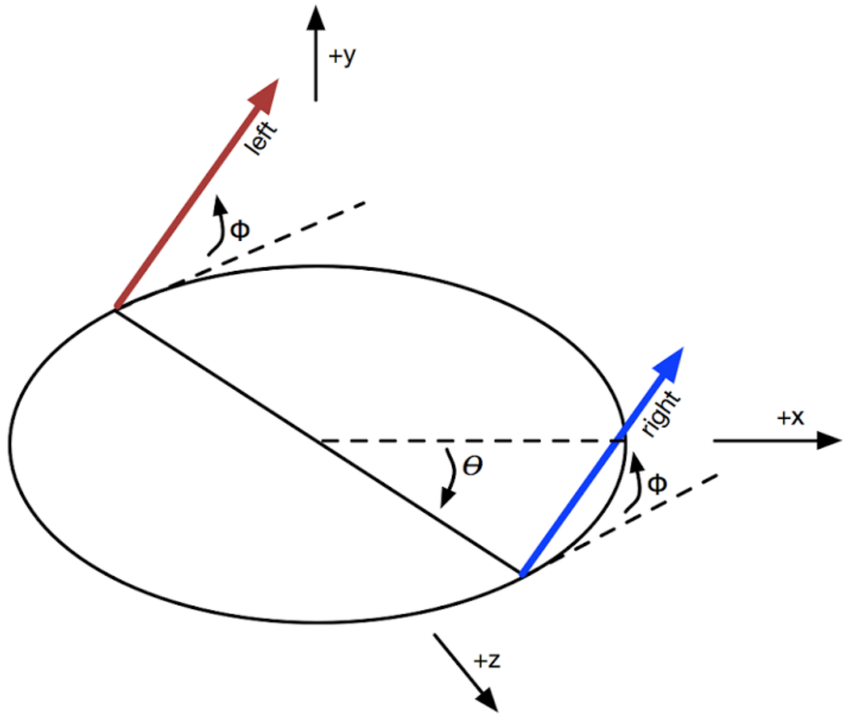
At Rendering Time

For any ray R that we need to render, find its approximate ray R’

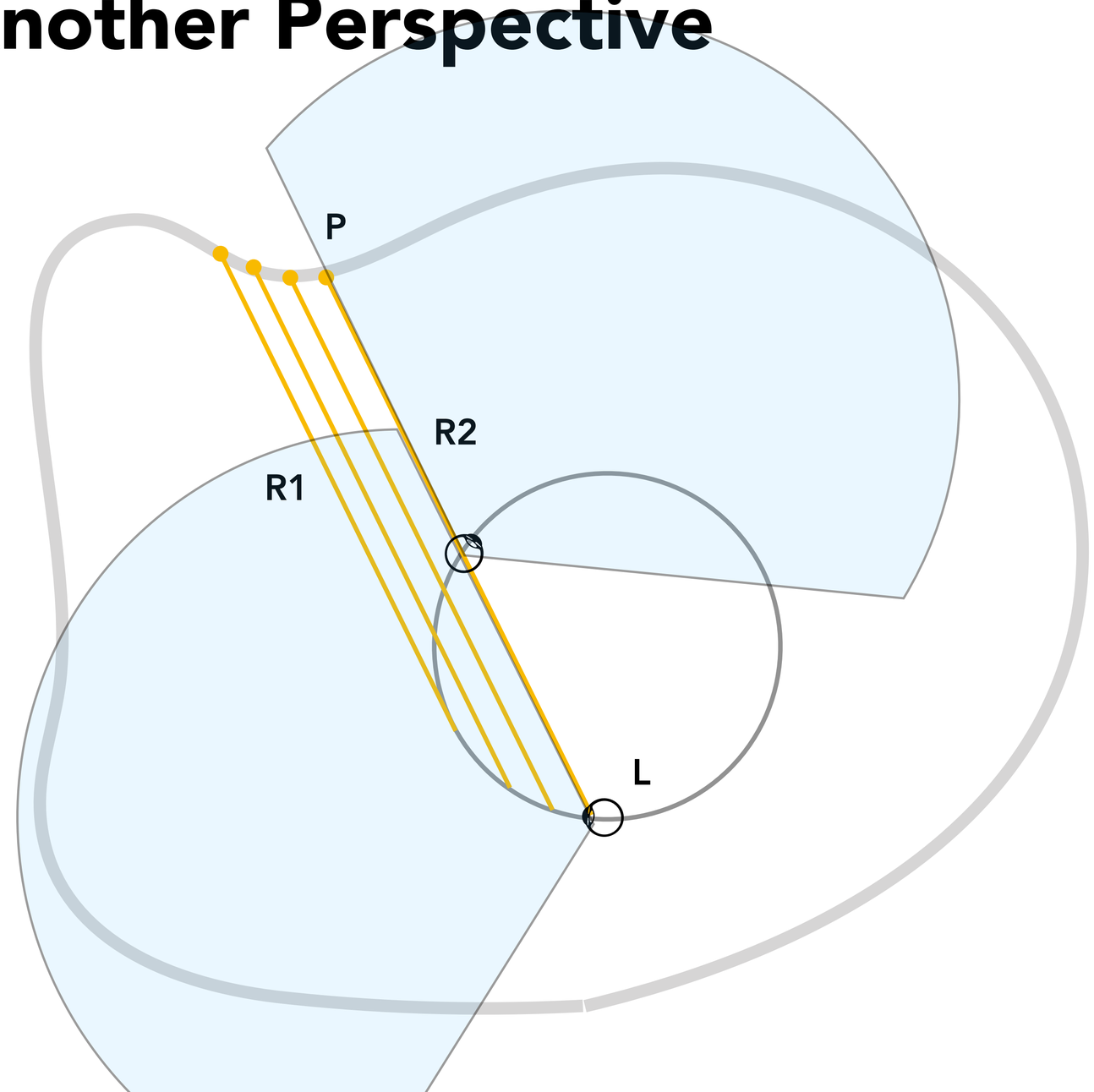
- The **R** rays do not necessarily originate from the viewing circle and might not be tangential.
- **R** is parameterized by azimuth θ and elevation ϕ, using which we find the corresponding **R’** originated from the viewing circle.

Read the color of R’ from the ODS image

- with potential filtering (after all it’s a signal sample + reconstruction problem).



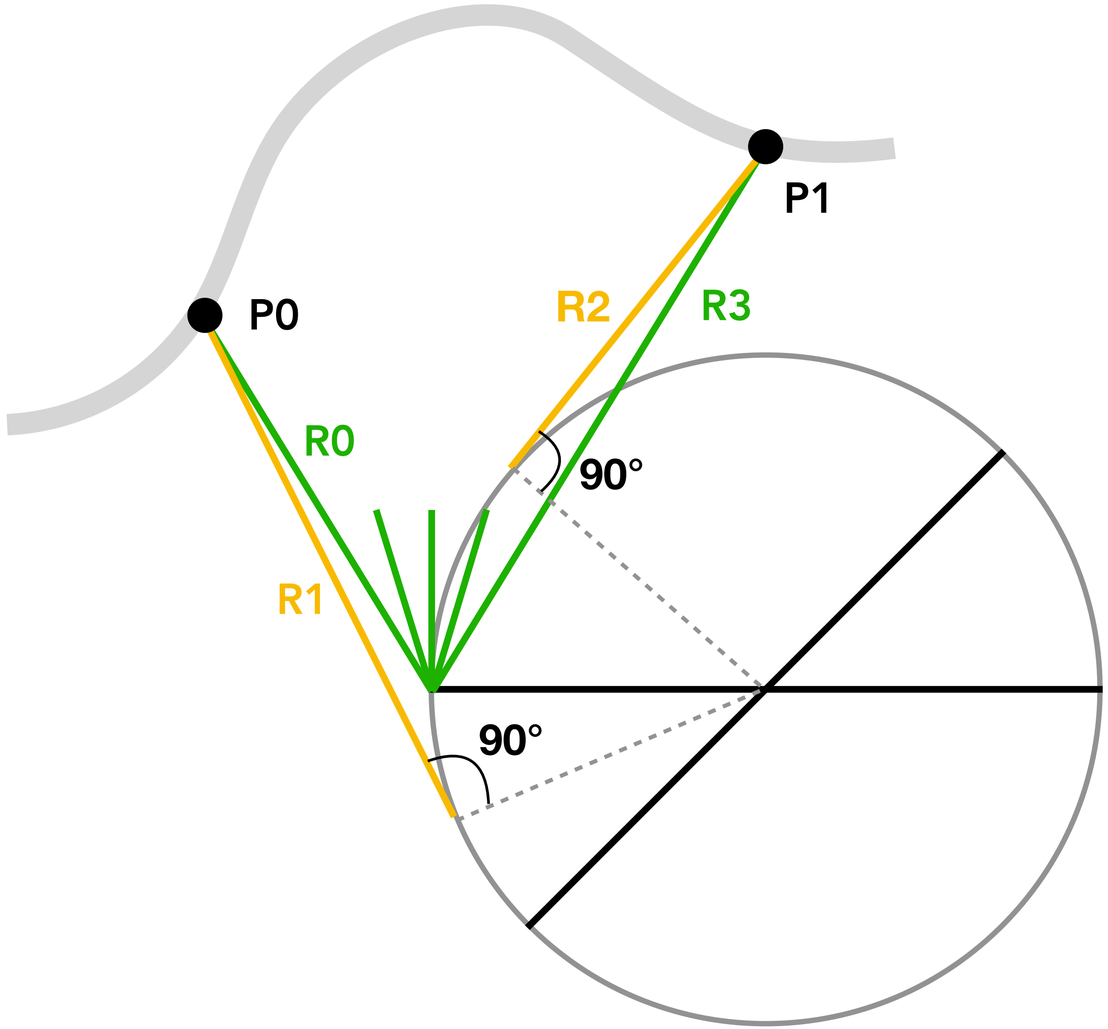
Another Perspective



All the parallel rays between R1 and R2 are approximated by R1.

- Other rays that have the same direction won’t be seen by the left eye.

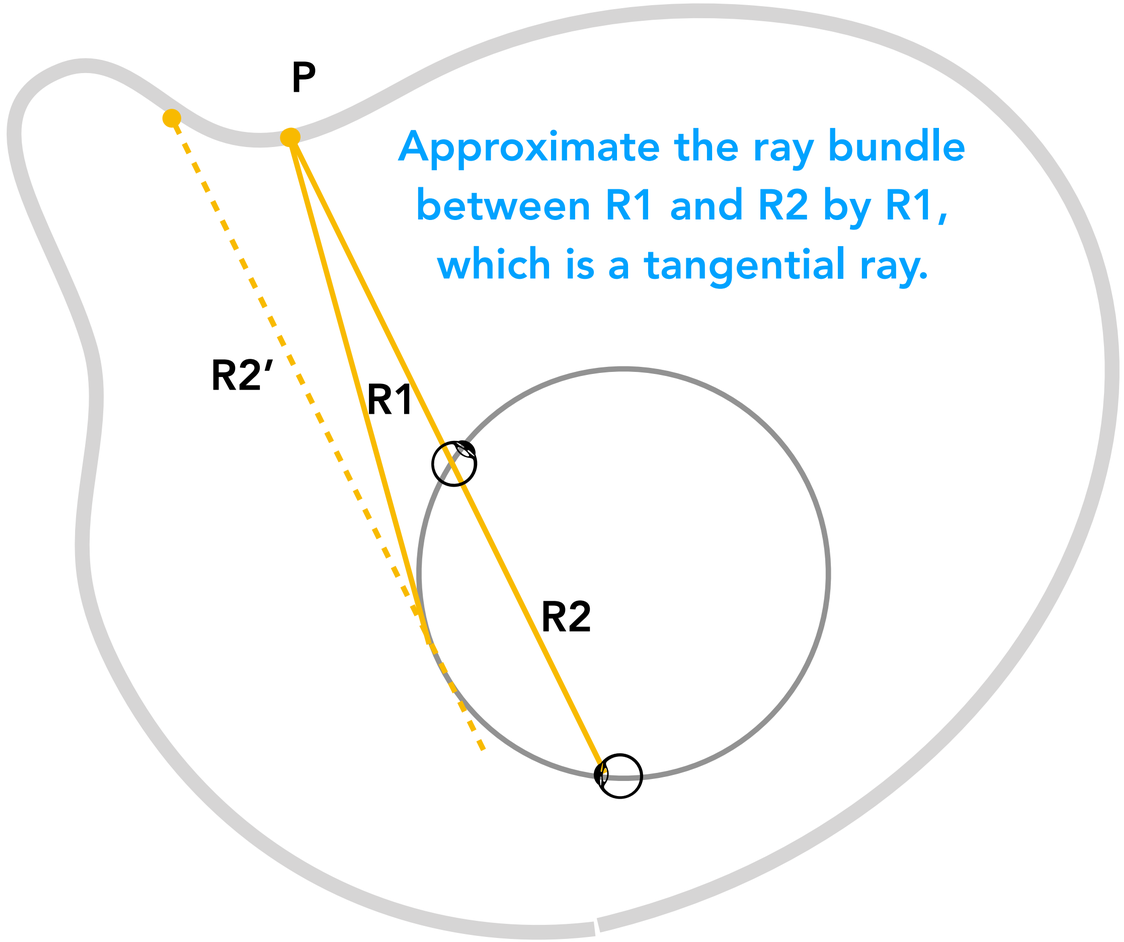
How About This Approximation?



Assumption: radiances of ray R0 and ray R1 are approximately similar.

- Same with R2 and R3.
- This is true if the scene point is perfectly diffuse, and in general it's OK (unless it's a mirror): remember the scene depth is much larger than the IPD. So it's really a small bundle of rays that we are approximating here.

How About This Approximation?

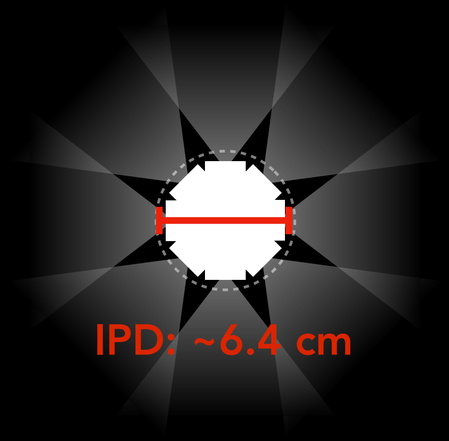


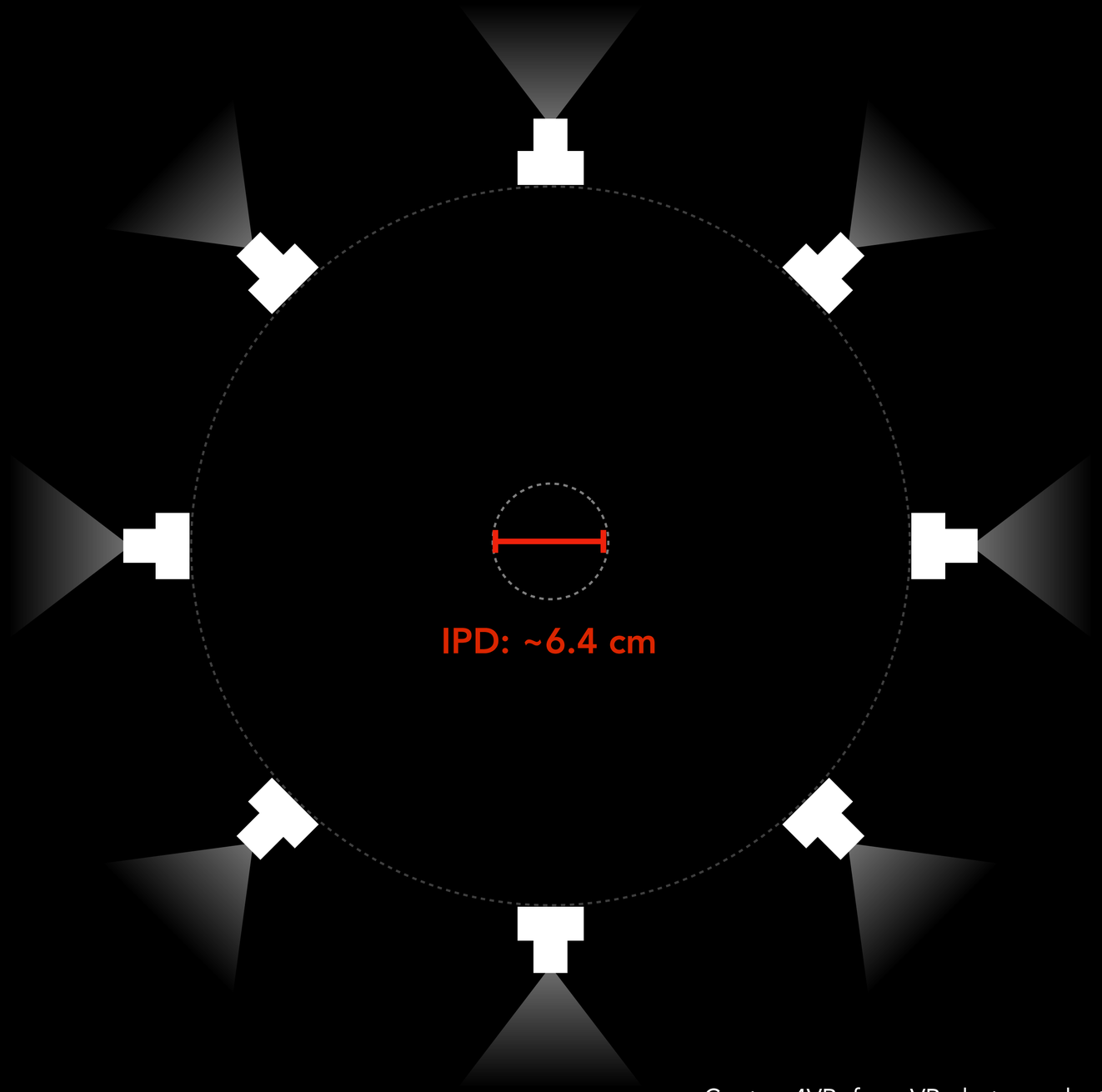
Pros: approx. error is small

- All rays approximated by R1 originate from the same point P
- The bundle is very small

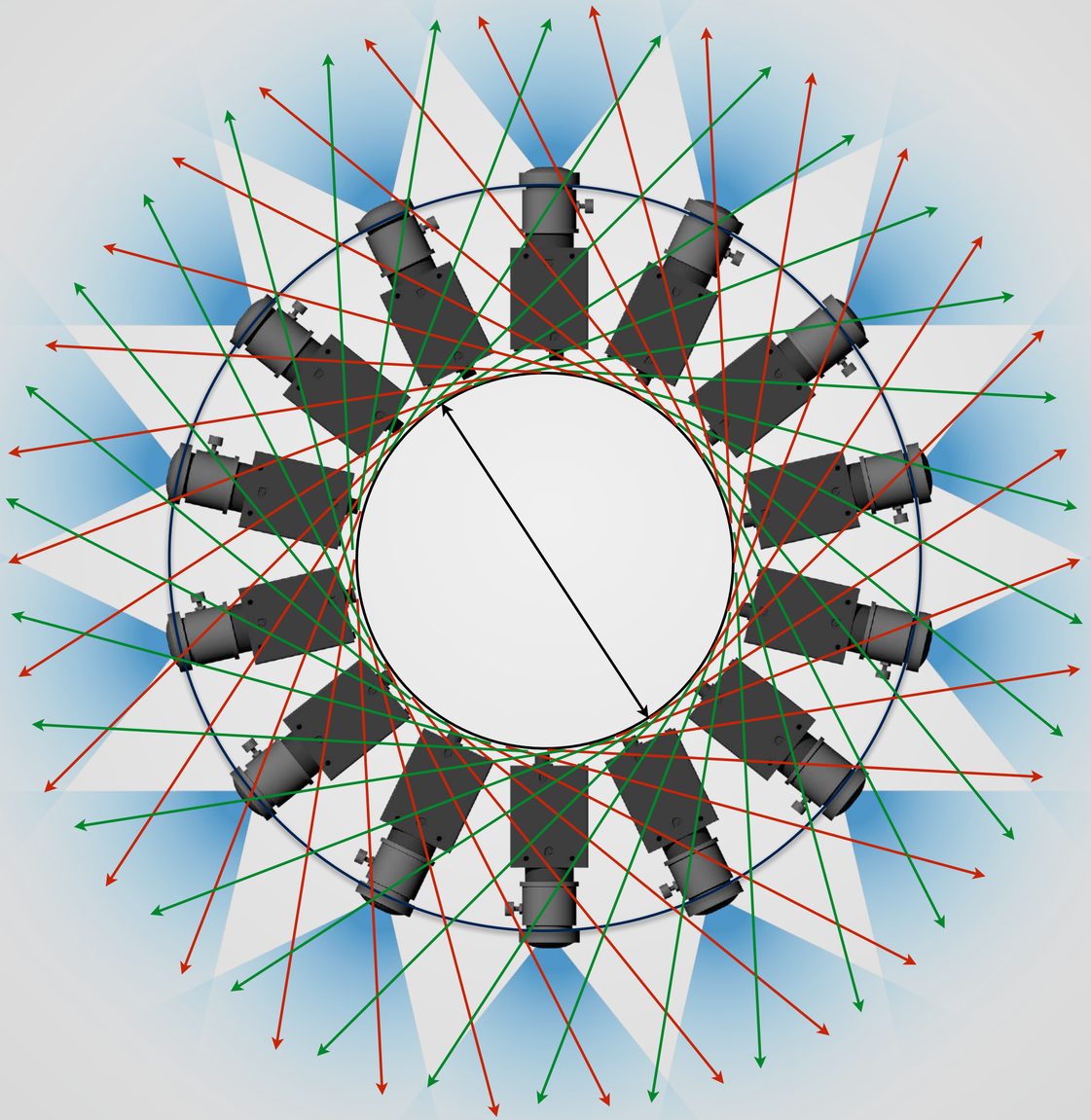
Con: need to know the depth of P for us to find R1 from R2.

- Could be done (e.g., 3D reconstruction and then store the mesh)
- c.f. conventional ODS, where given R2, R2’ is independent of P

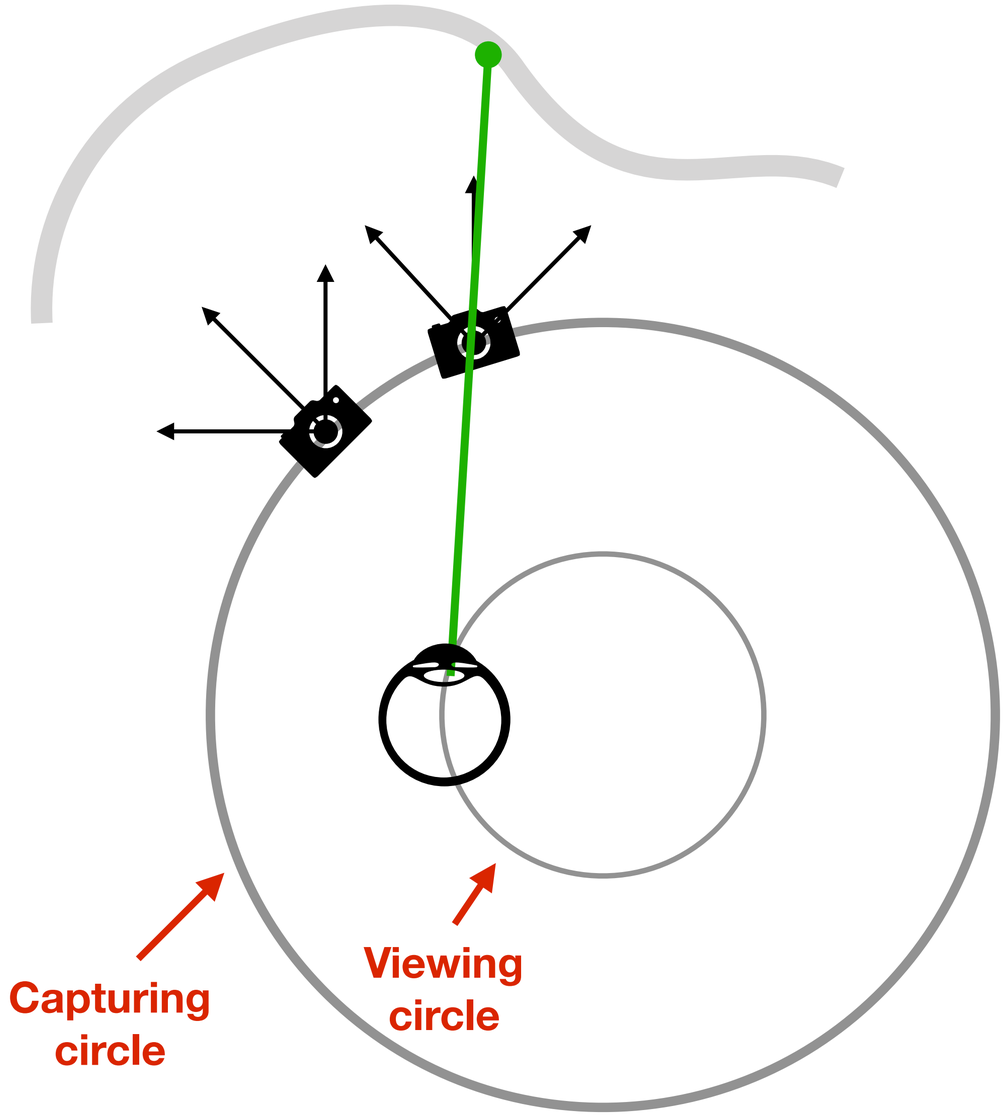

IPD: ~6.4 cm


IPD: ~6.4 cm

Creating ODS with a fixed array of cameras



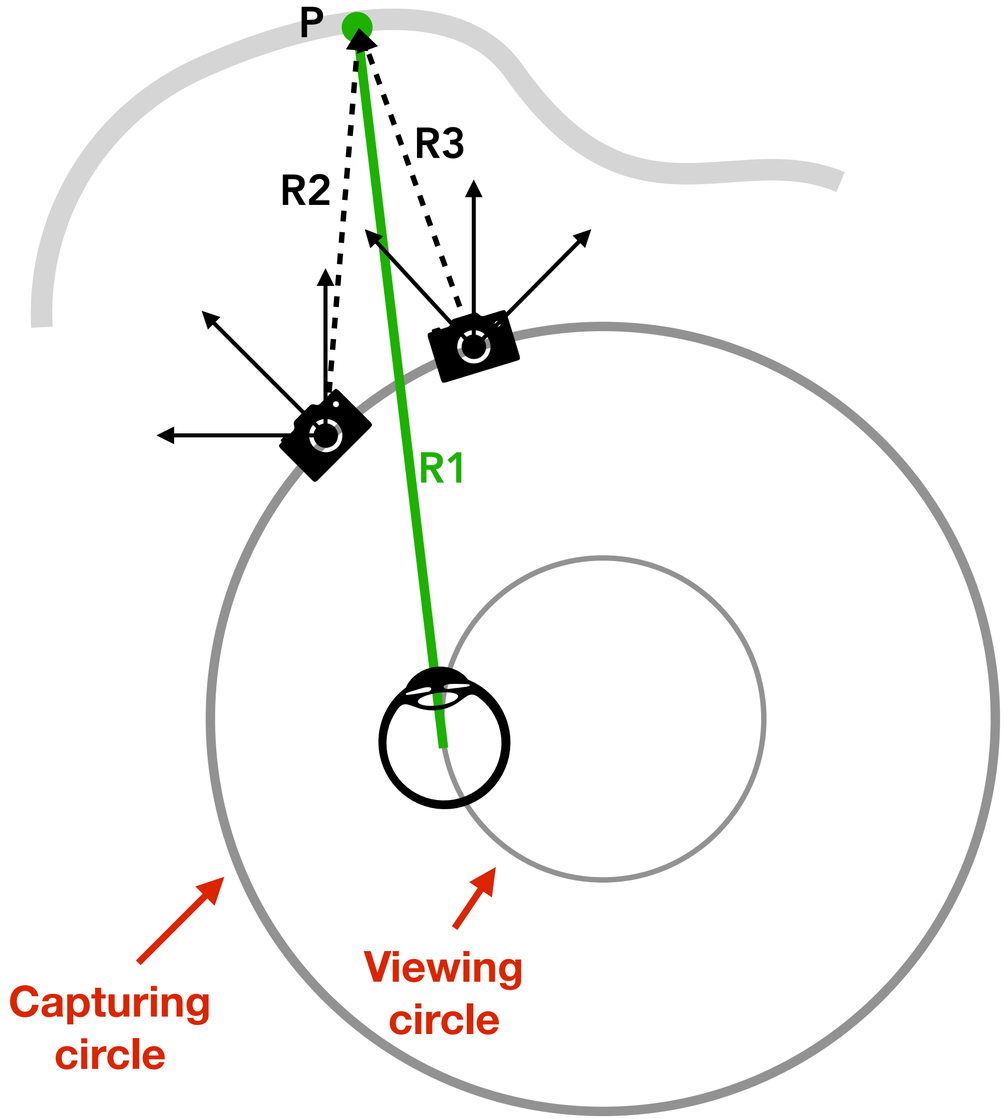
Capturing Infinite Rays with Finite Cameras



If a tangential ray passes through the center of a real camera, we can simply use the corresponding pixel.

If a tangential ray is not captured by any camera, estimate from the two adjacent cameras.

Capturing Infinite Rays with Finite Cameras

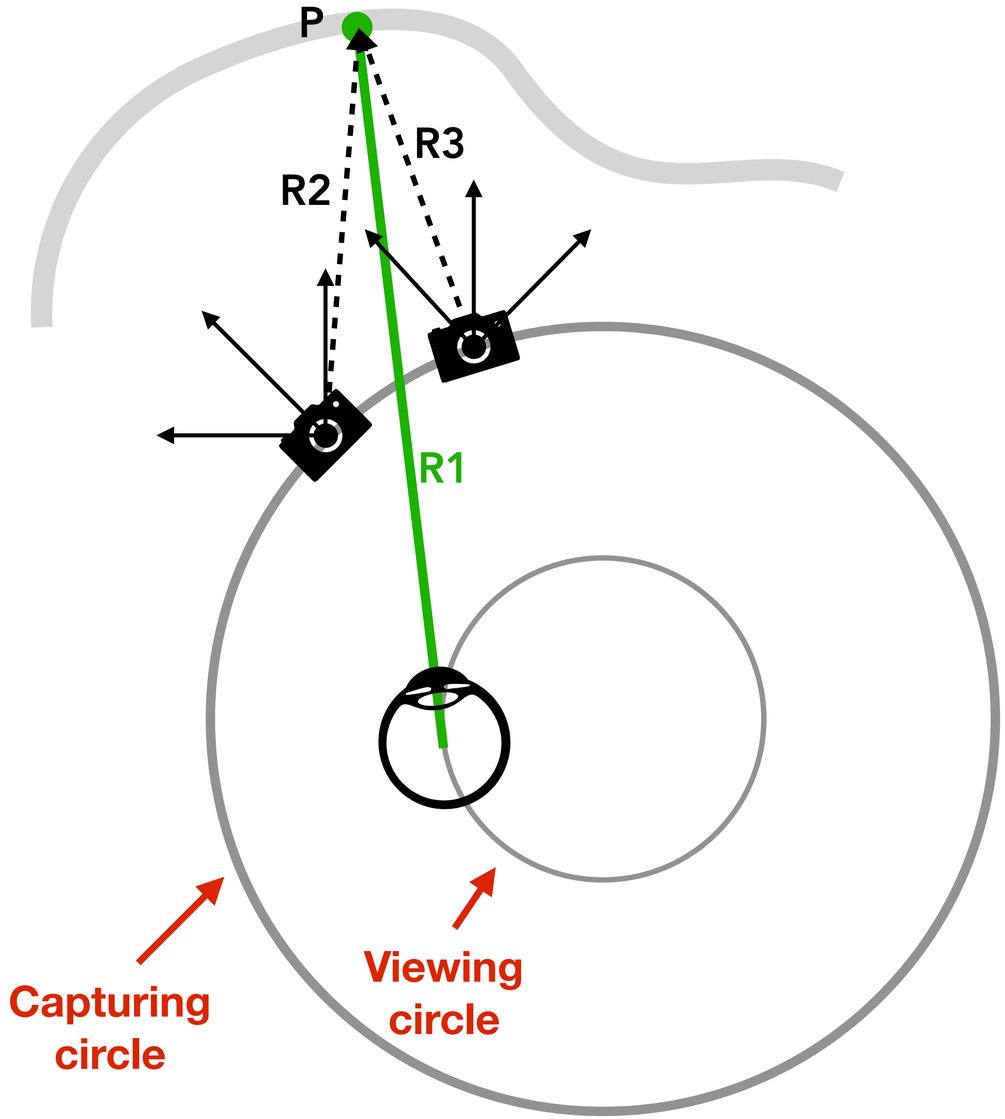


R1 is not directly captured by any camera, but two other rays, R2 and R3 from the same scene point P are, so:

- 3D Reconstruct the scene from the actual camera captures to find P
- Identify R2 and R3
- Interpolate to get R1

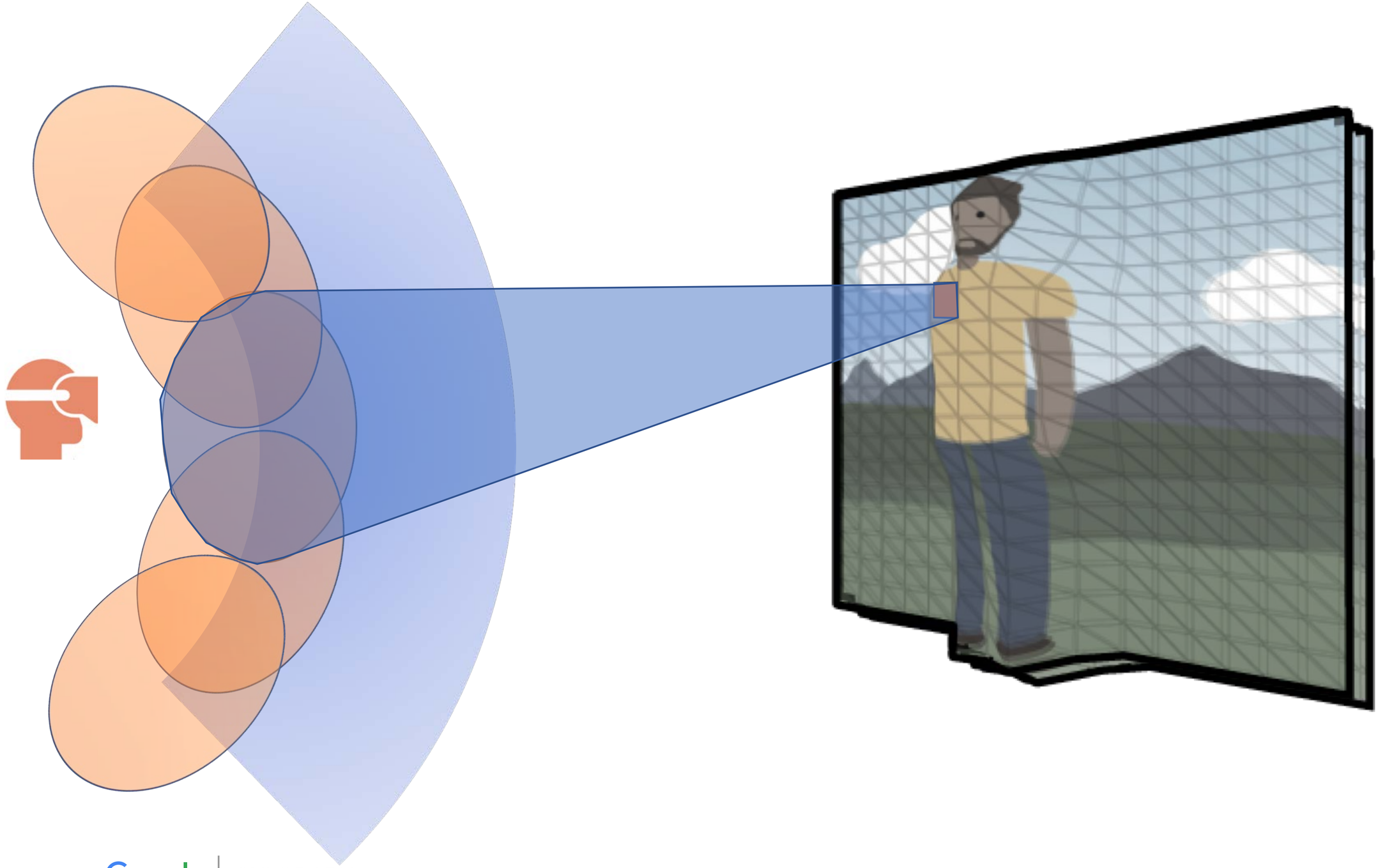
Or simply interpolate between the two camera views to synthesize a virtual camera that does capture R1.

Capturing Infinite Rays with Finite Cameras



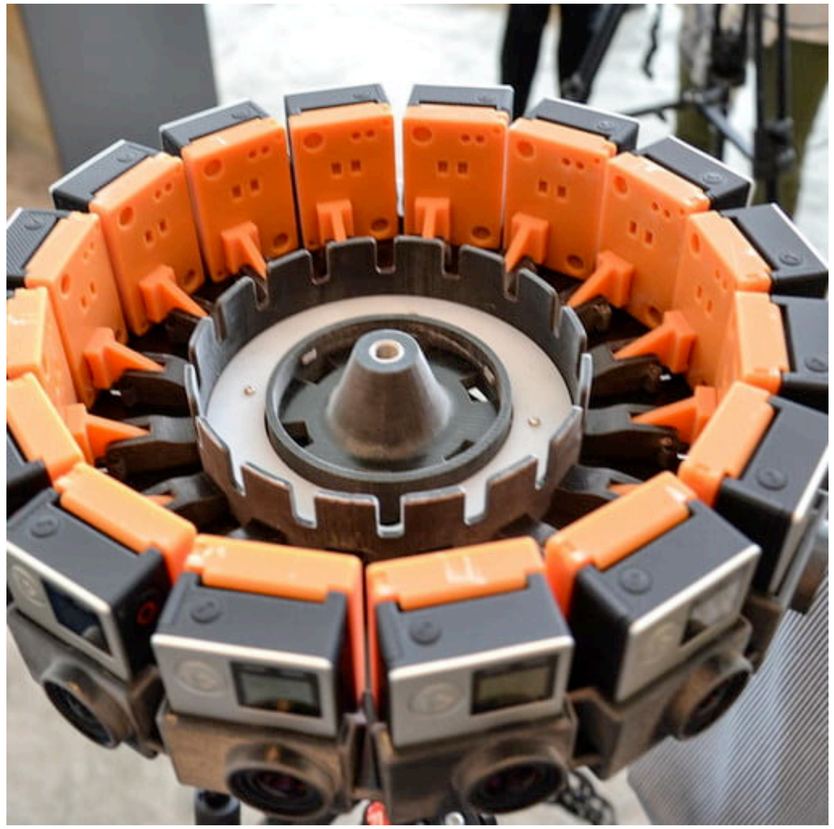
This is fundamentally a signal sampling and reconstruction problem.

- Light emitted from P is the underlying signal
- R2 and R3 are two samples from it, and we want to calculate R1
- Ideally: first reconstruct the underlying signal and then resample it for R1
- In practice, filter R2 and R3 for R1



Commercial VR Video Camera Rig

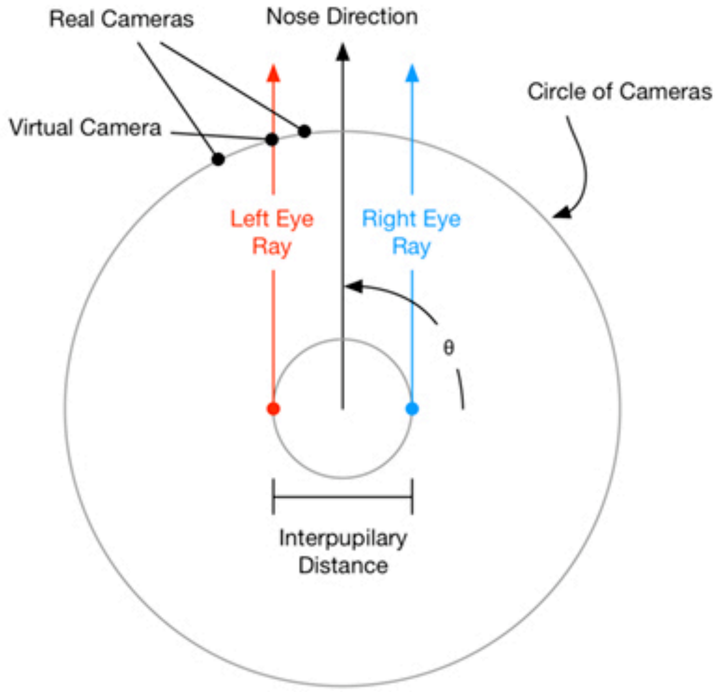
Google Jump VR (discontinued)

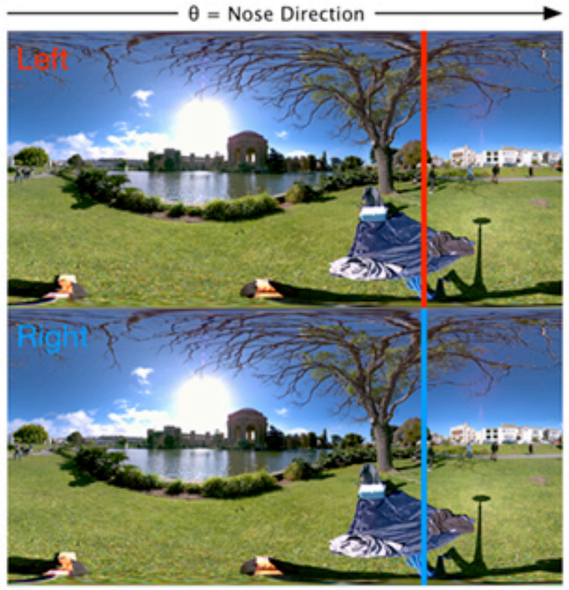


A Jump rig built by Yi Halo

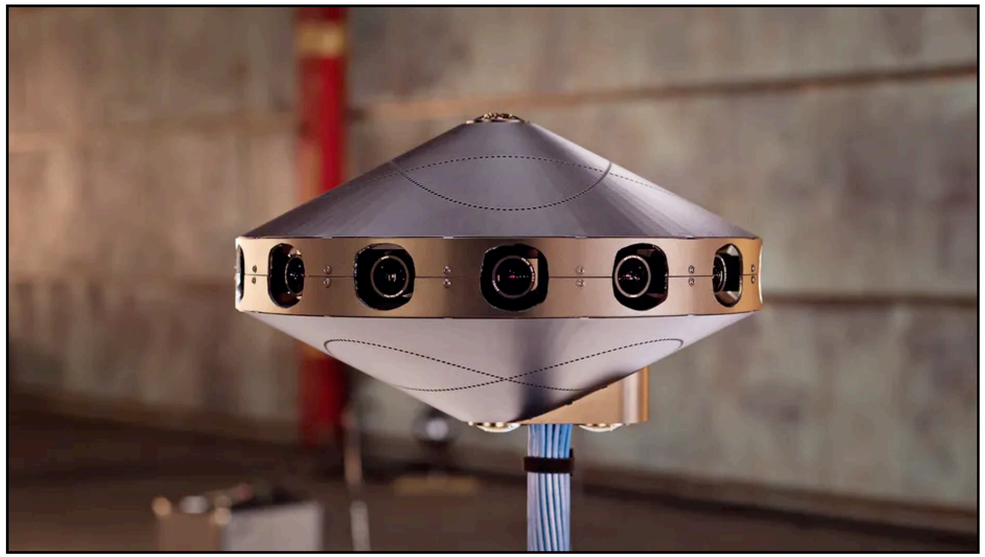
https://static.googleusercontent.com/media/research.google.com/en//pubs/archive/45617.pdf
https://twitter.com/yitechnology/status/918116570559336448

Commercial VR Video Camera Rig





Facebook Surround 360 System (mostly open-source)



https://engineering.fb.com/2016/07/26/video-engineering/surround-360-is-now-open-source/

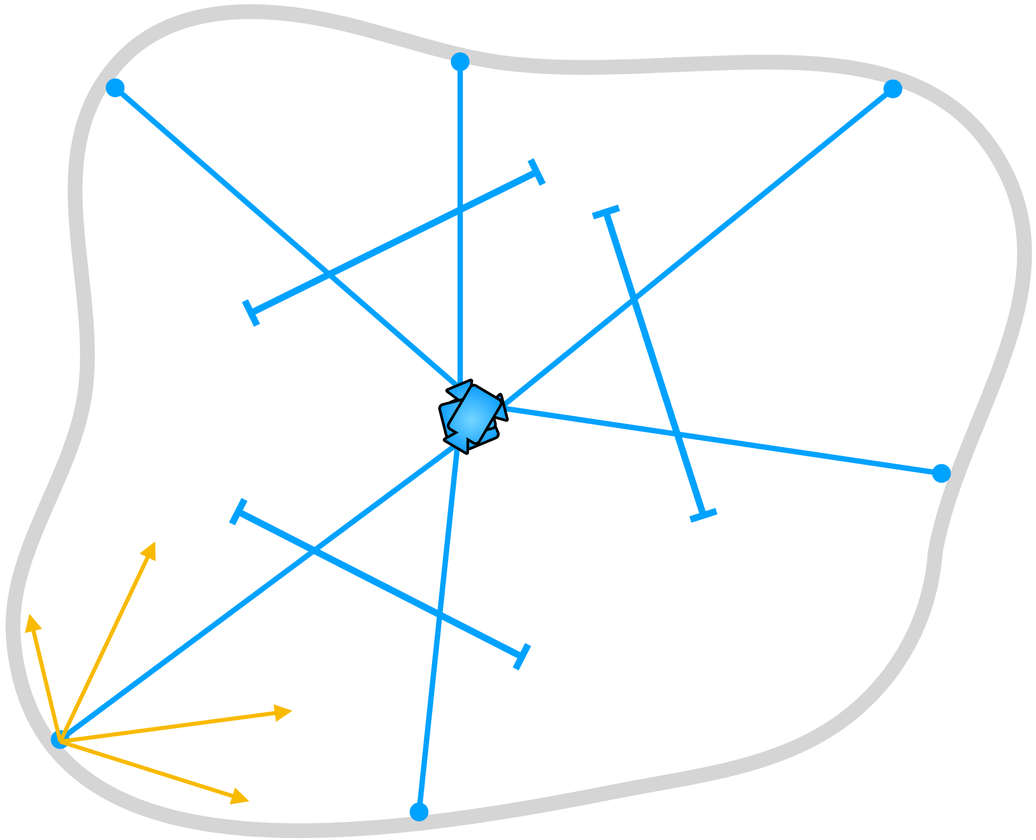
Live-Streaming VR Cameras

Z-Cam V1 Pro

Recap (So Far)

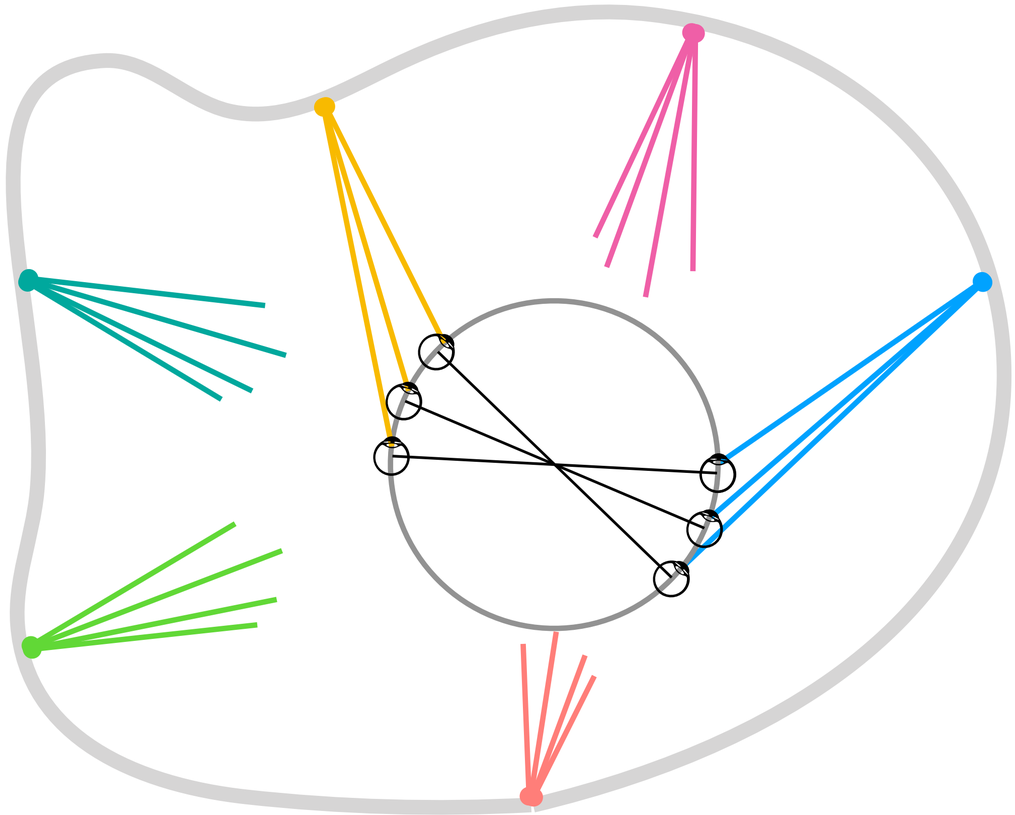
Traditional 360°: 3 DOF

Need & capture one ray per scene point.



ODS: 3 DOF + stereo

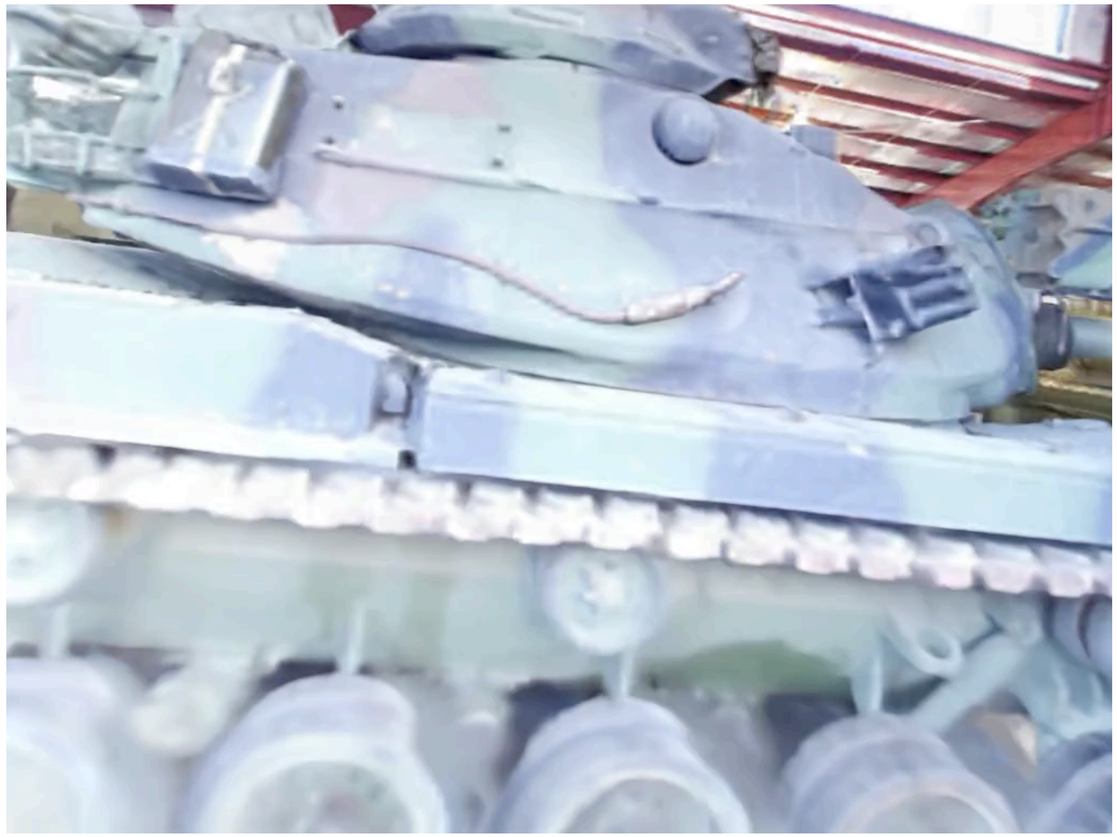
Need a small ray bundle per scene point
(but capture only one ray per point).



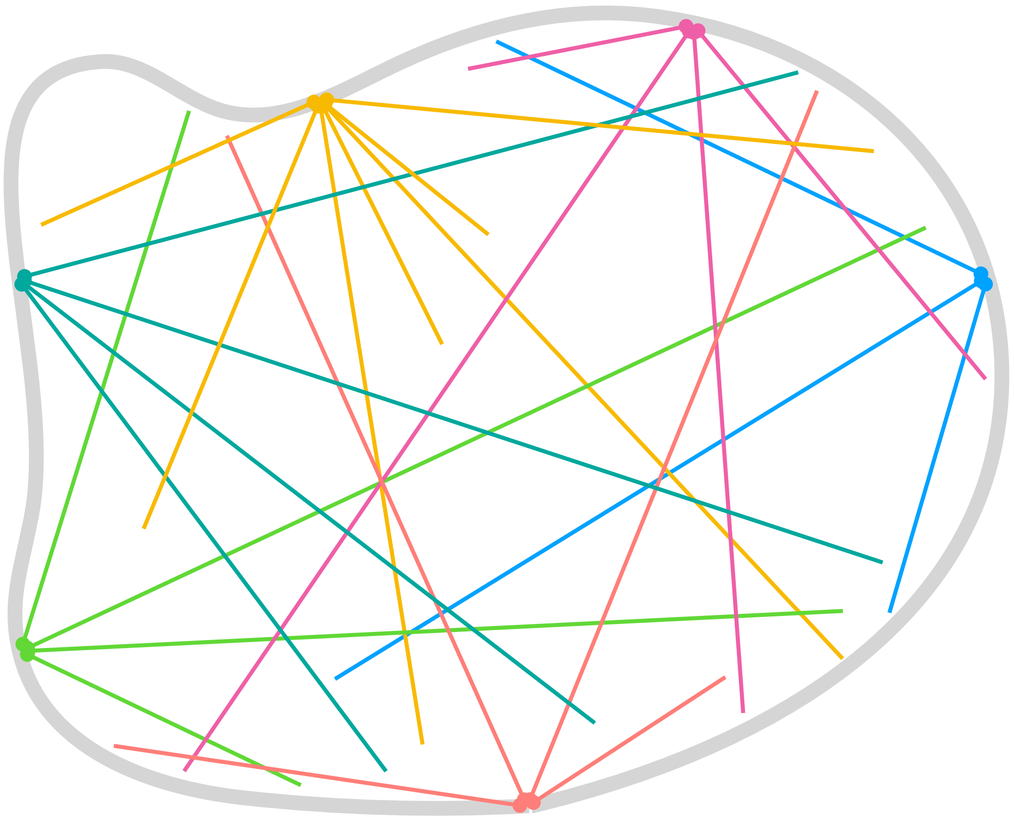
* Actually a small ray bundle subtended by the lens unless using a pinhole camera.

What Does It Take to Achieve This?

What Does It Take to Achieve This?

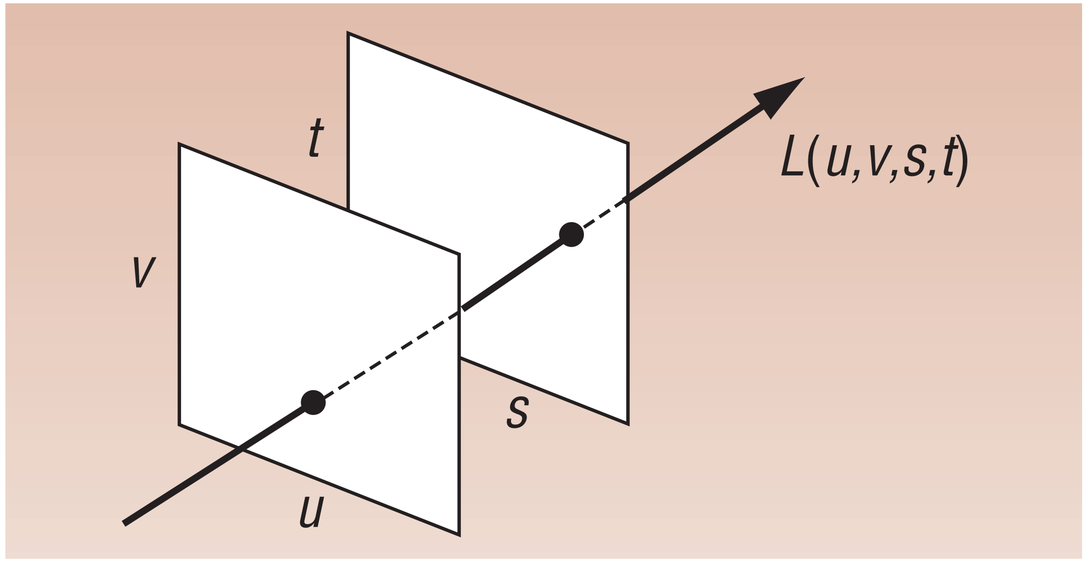


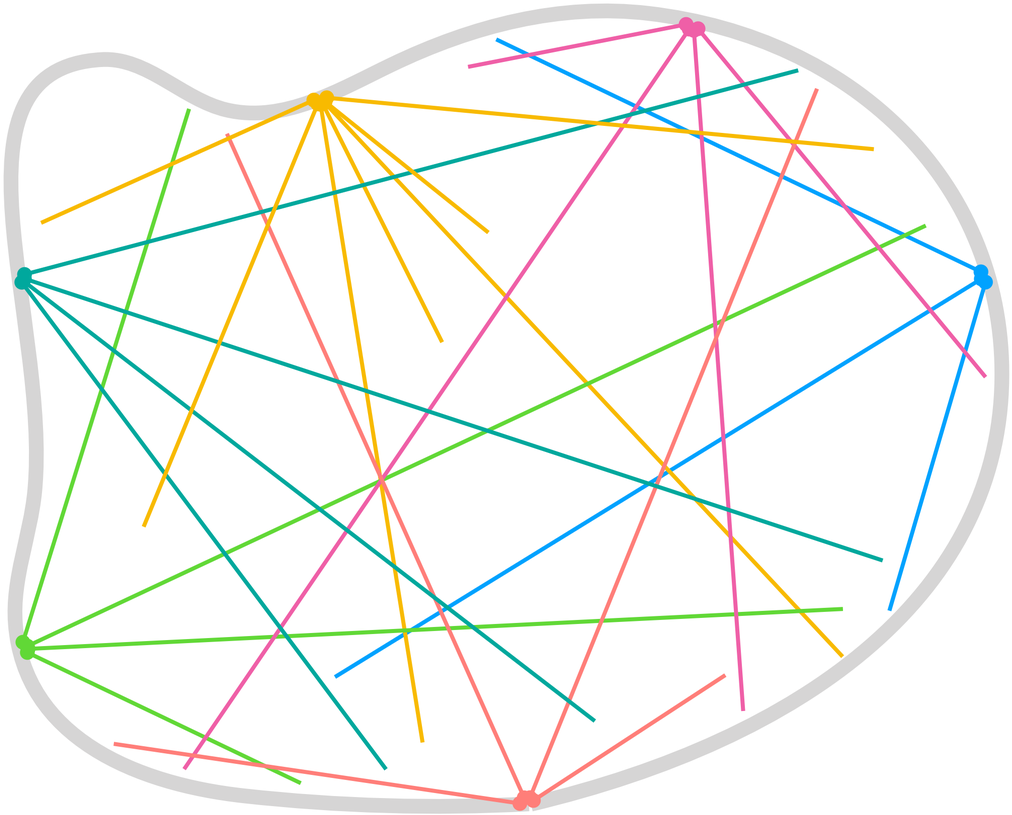
We Need to Trace Every Possible Ray in the Scene



Light Field (a.k.a., Lumigraph)

Plenoptic function: a **4D** function describing a ray in the free space.

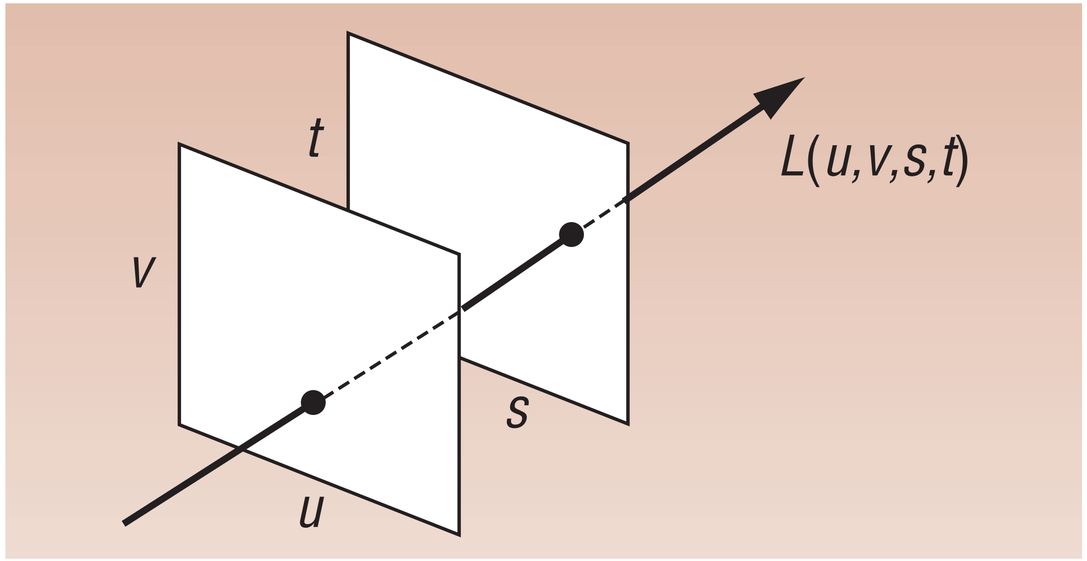




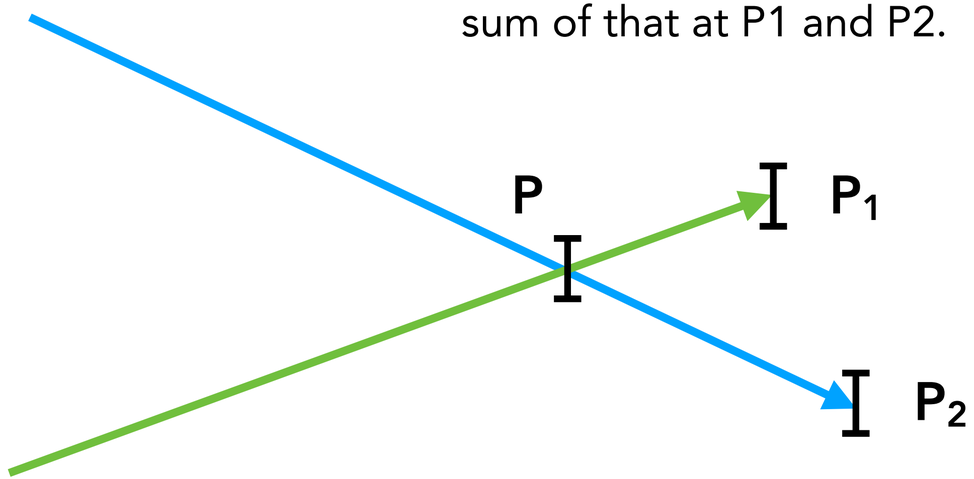
Light Fields and Computational Imaging, IEEE Computer, 2006

A 4D Parameterization (Why?)

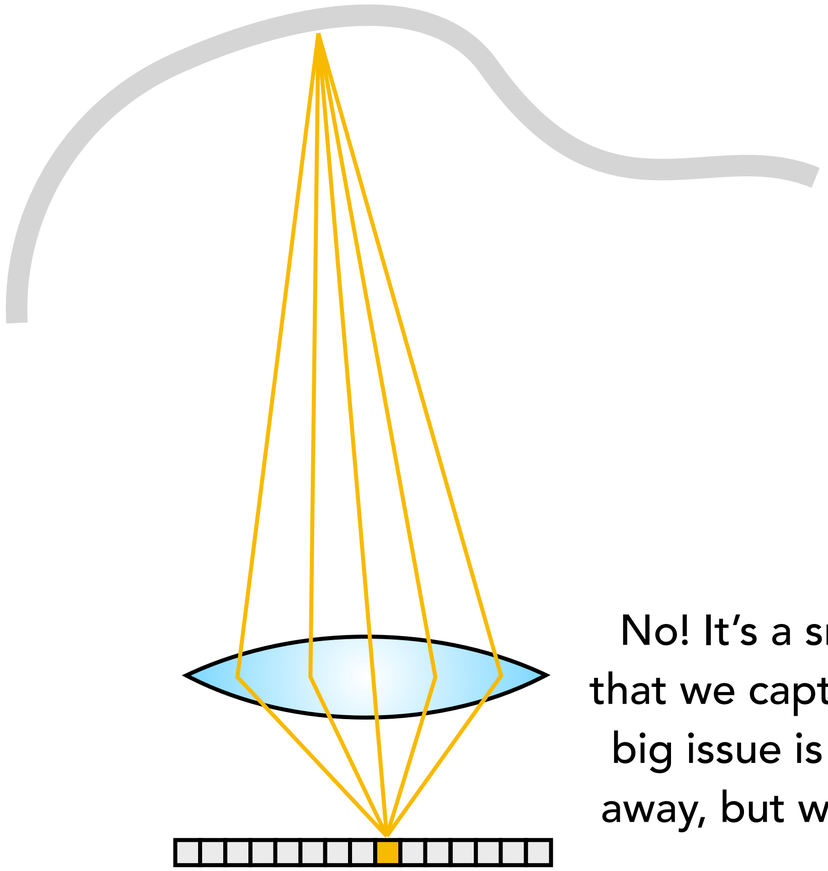
Light field: light leaving every point of the scene and traveling in every direction.
In vacuum ray radiance doesn't change along the ray direction without occlusion.



Additive: energy at P is the sum of that at P1 and P2.

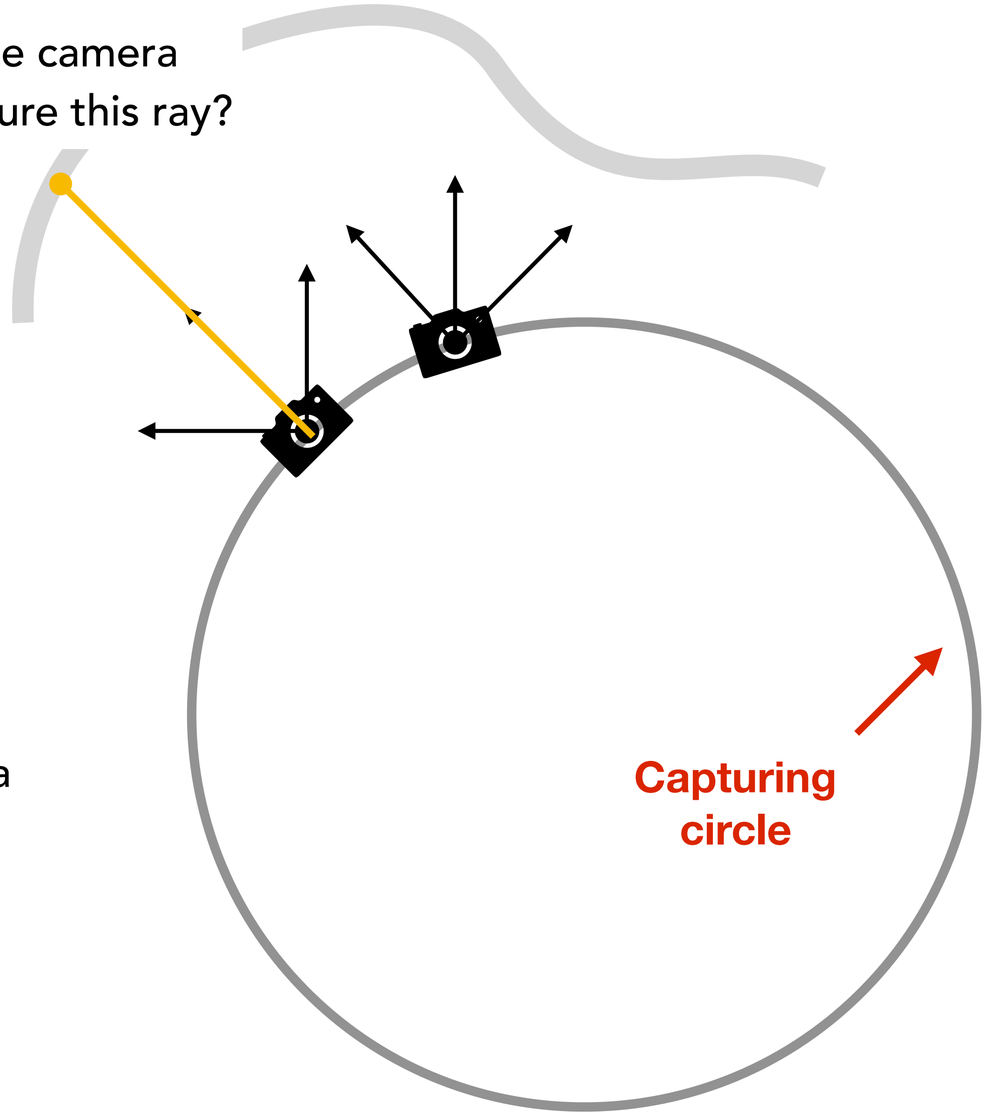


How to Capture Light Field?

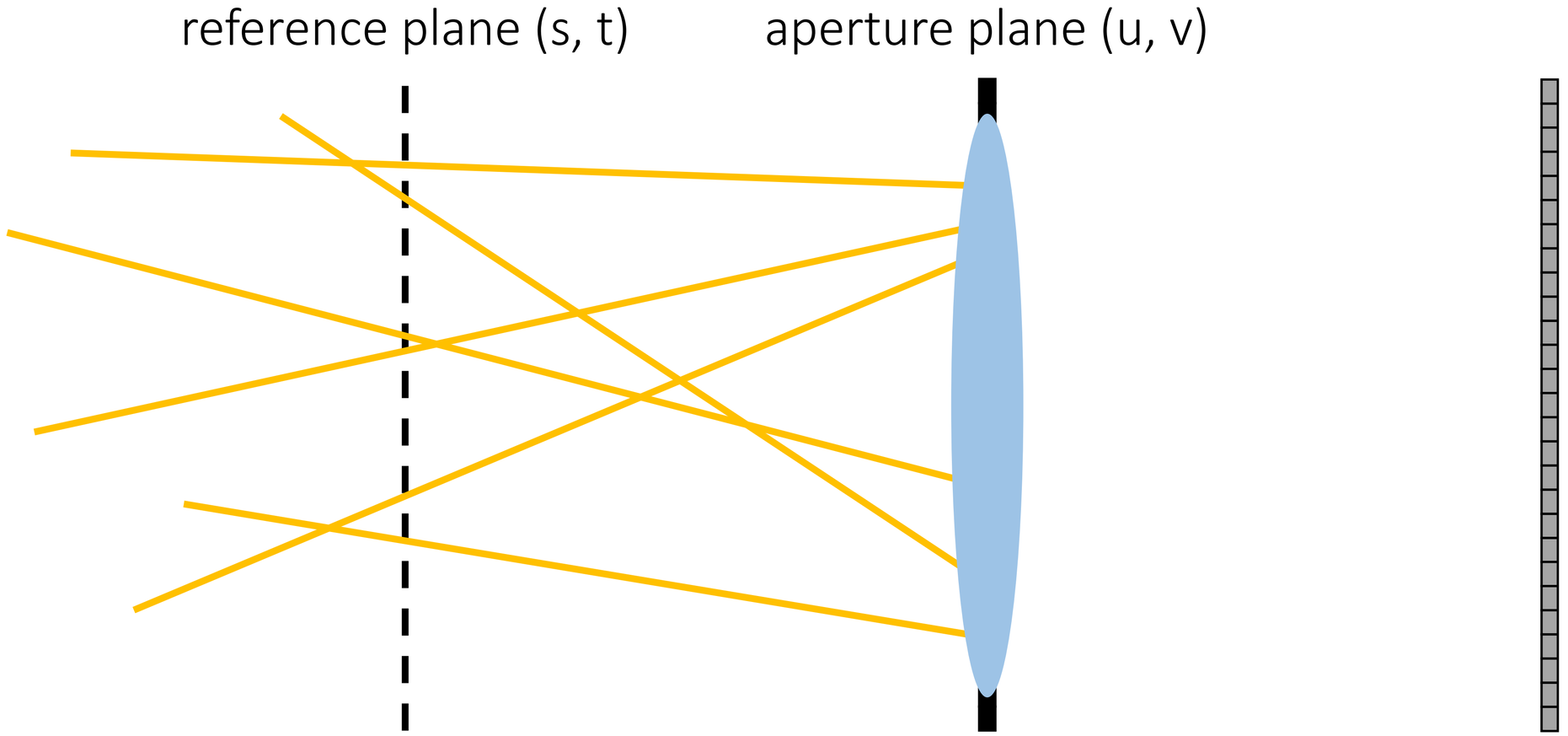


No! It's a small ray bundle that we capture. This is not a big issue is the scene is far away, but will be otherwise.

Does the camera really capture this ray?



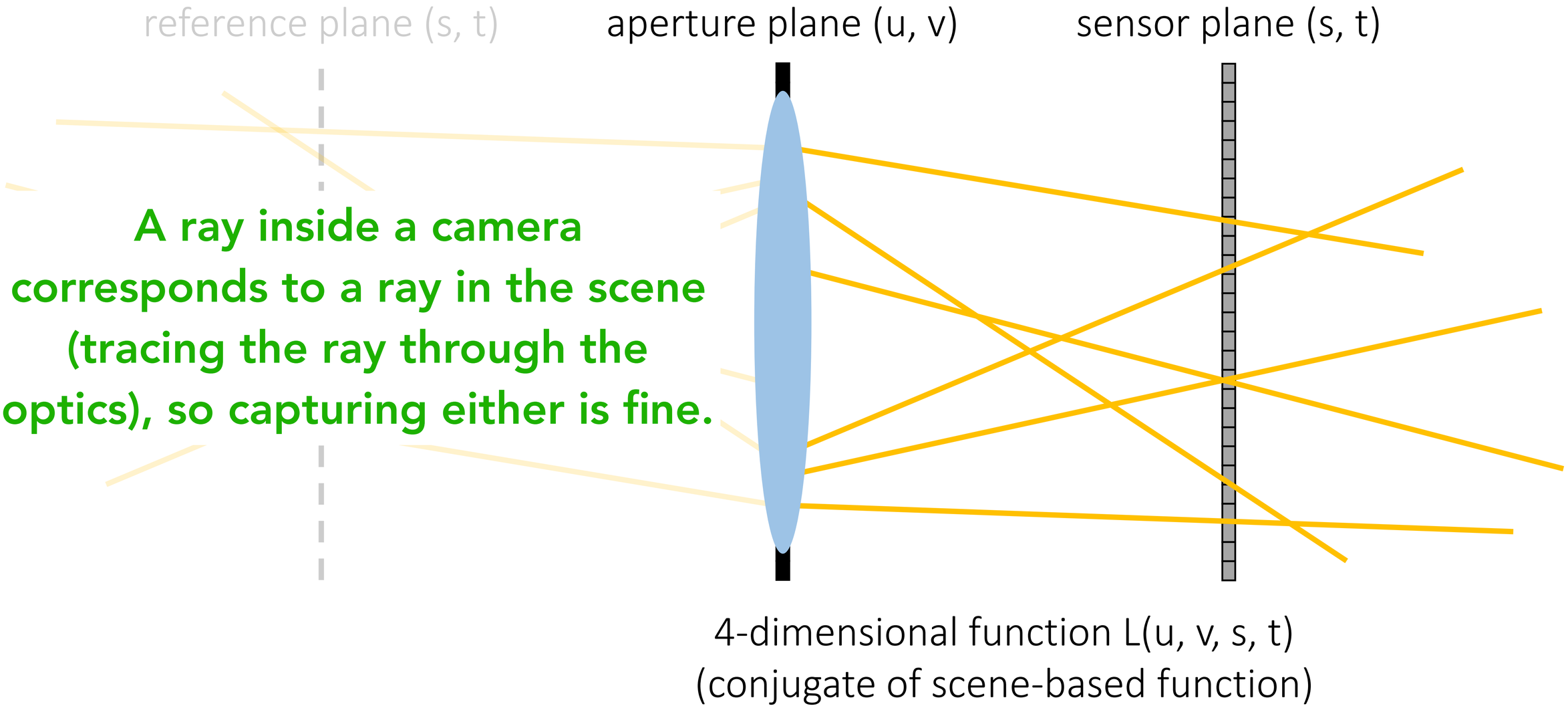
How to Capture the Light Field?



4-dimensional function L(u, v, s, t)

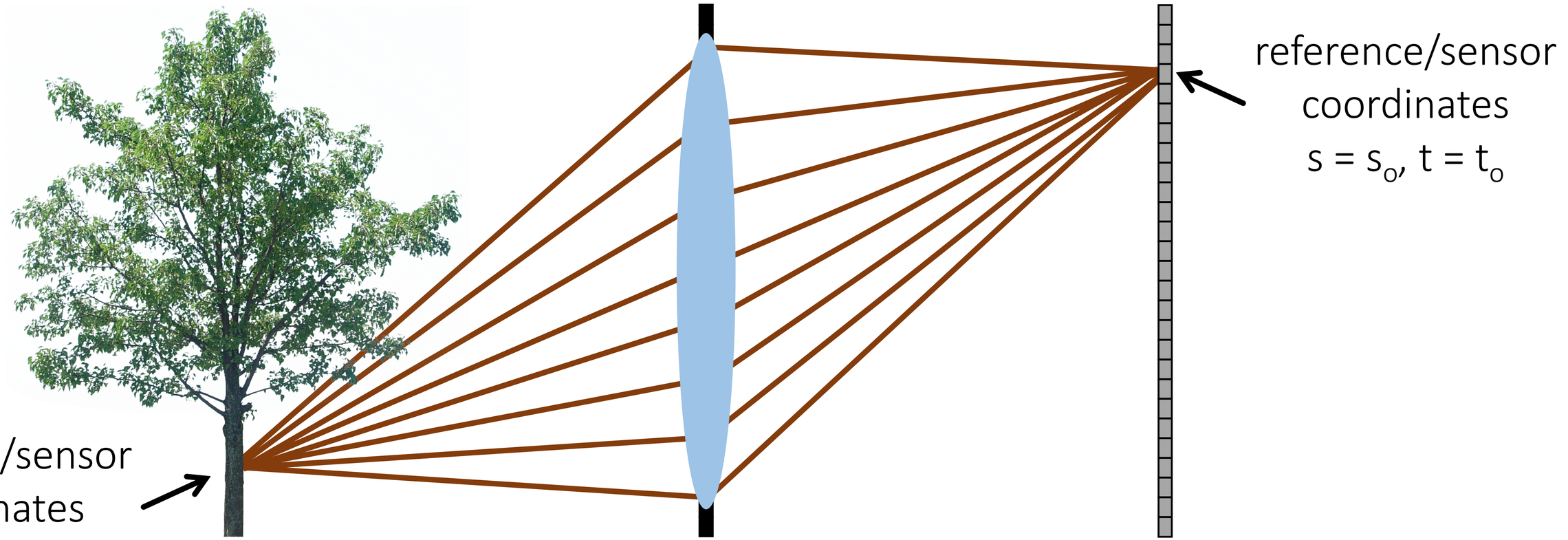
*** Simplifying 4D to 2D**

Capturing the Light Field Inside a Camera



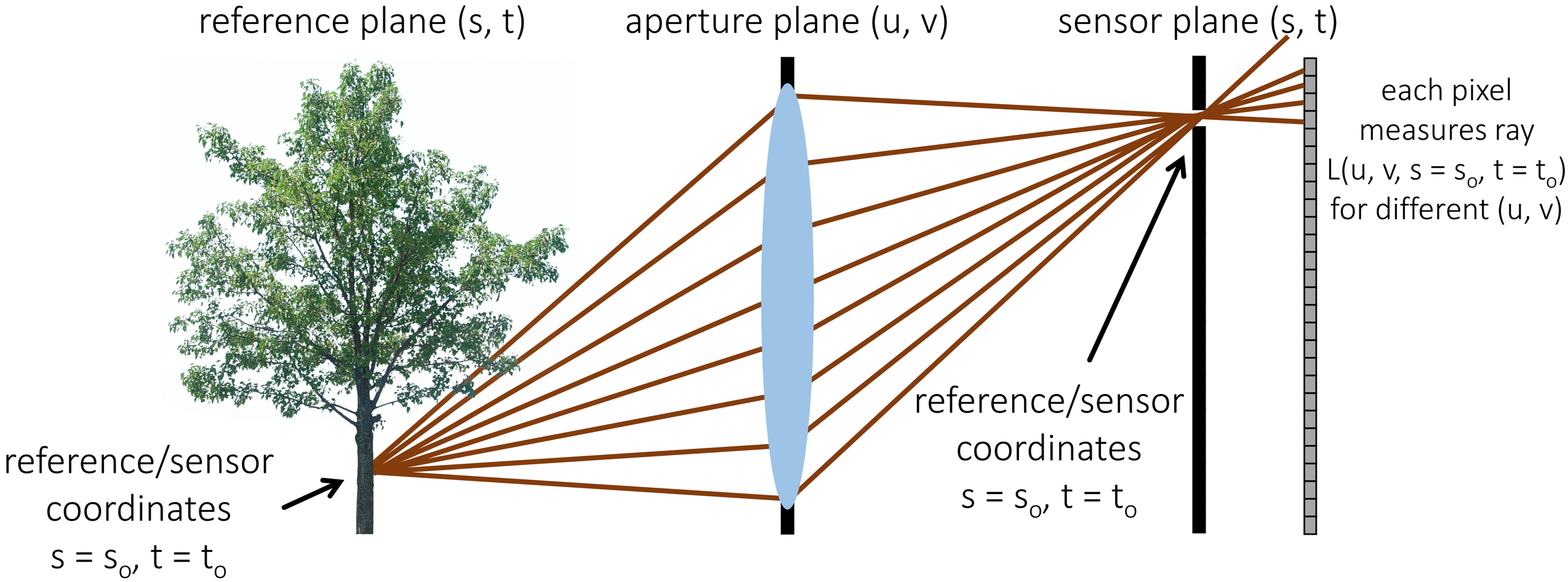
*** Simplifying 4D to 2D**

Making a plenoptic camera



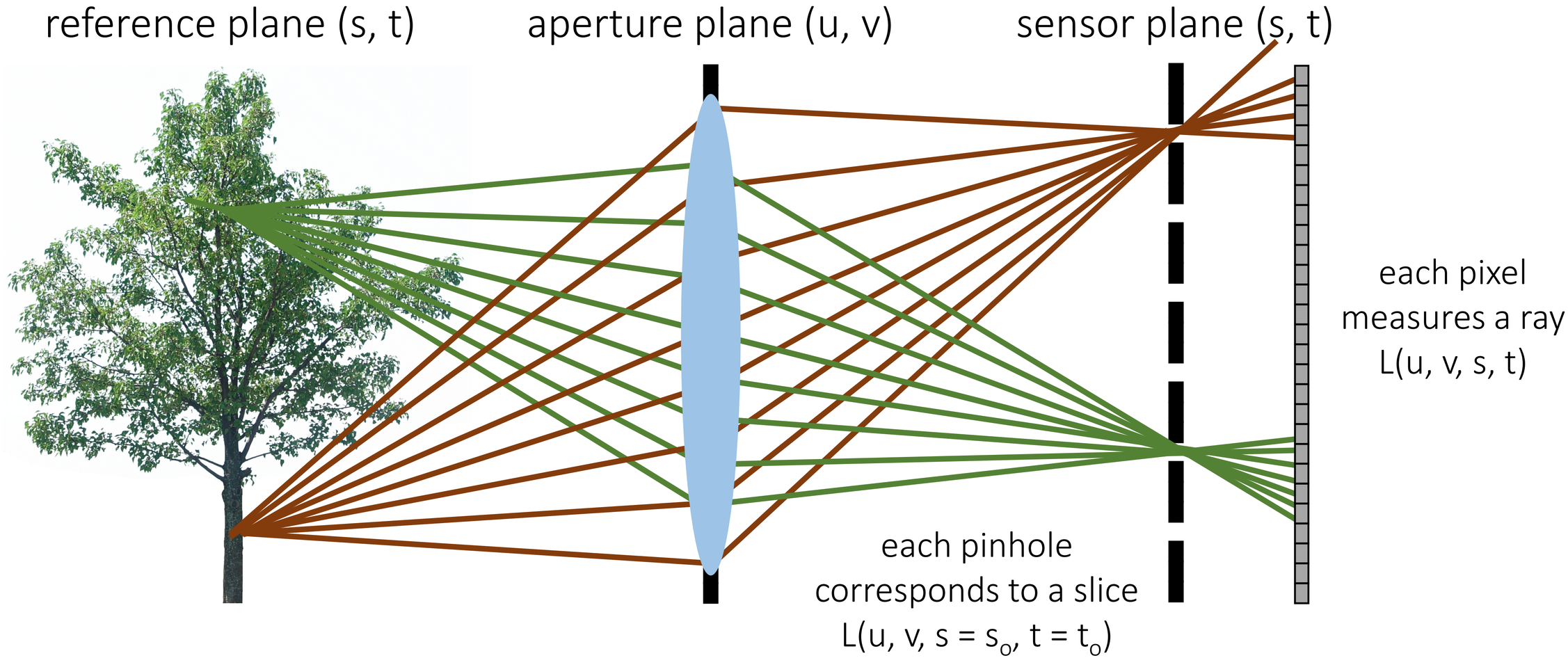
Lightfield slice $L(u, v, s = s_o, t = t_o)$

Making a plenoptic camera



Lightfield slice L(u, v, s = s_o, t = t_o)

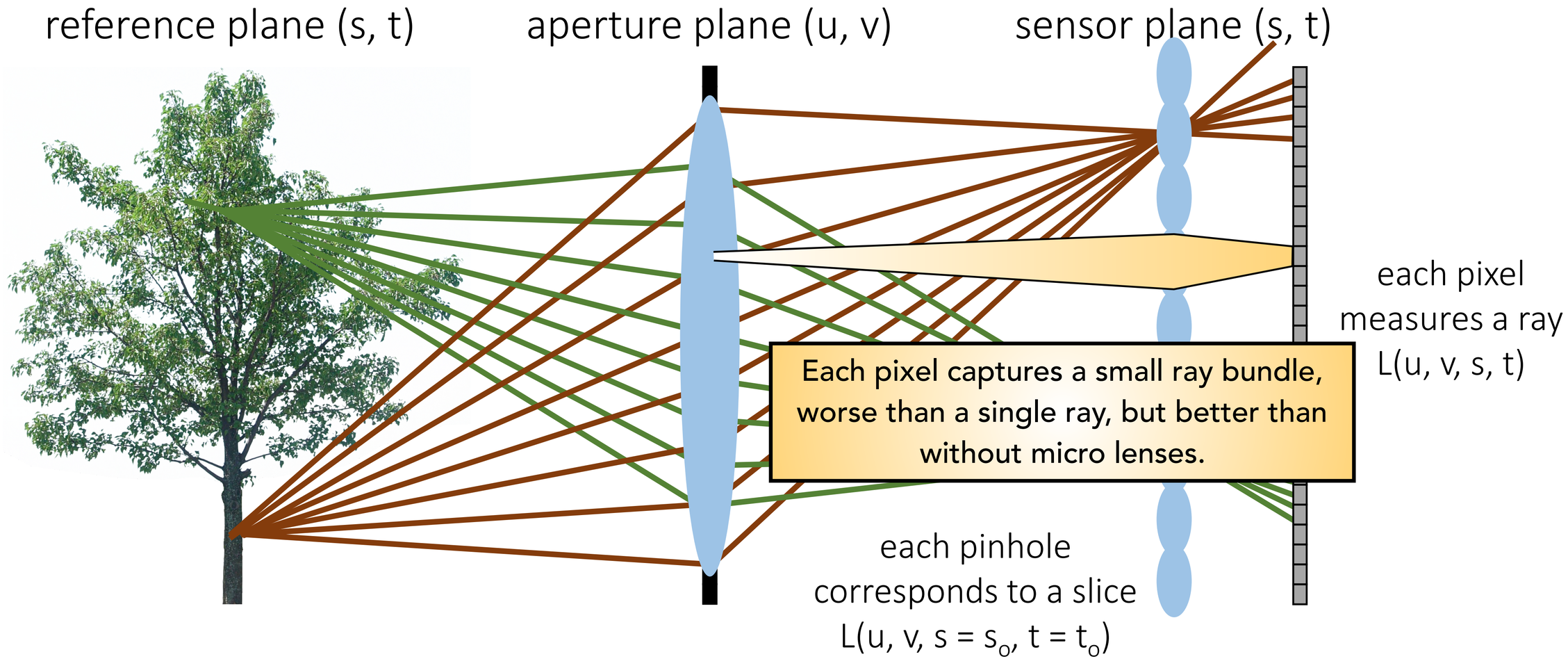
Making a plenoptic camera



Lightfield L(u, v, s, t)

How can we make this more light efficient?

Making a plenoptic camera



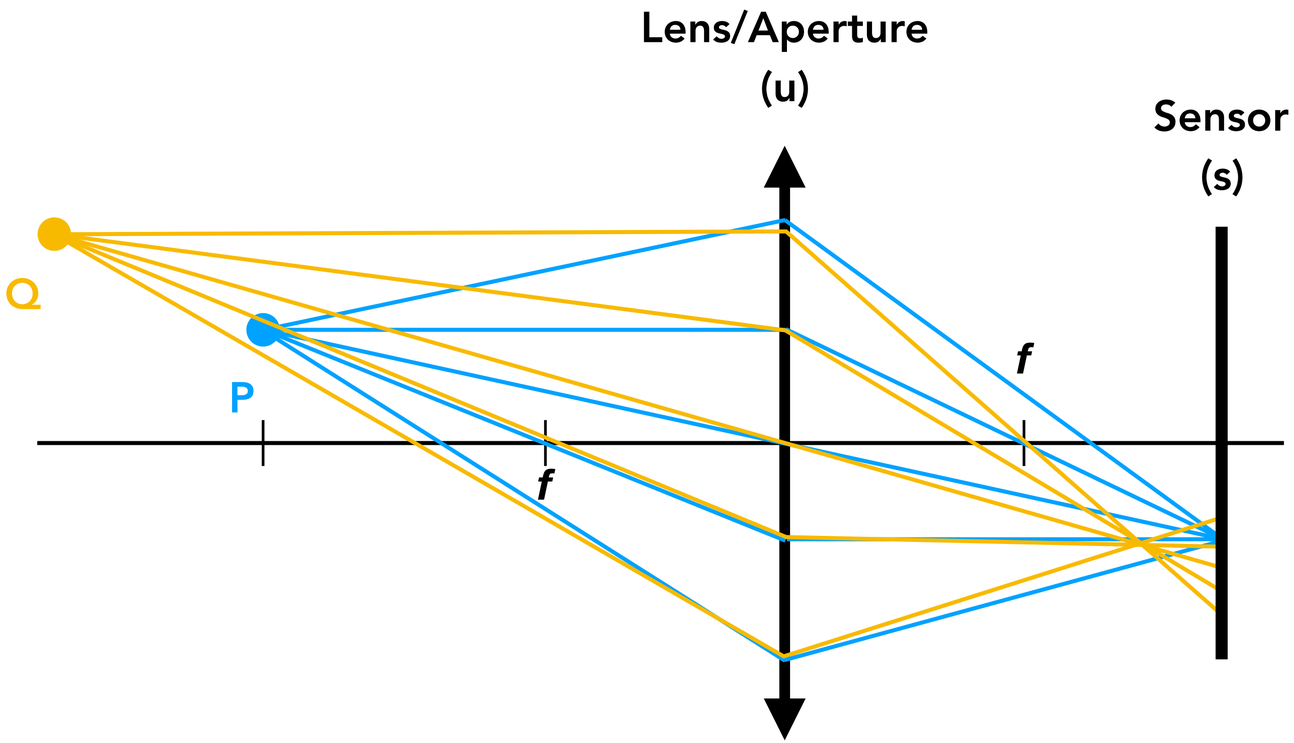
Lightfield L(u, v, s, t)

How can we make this more light efficient?

- replace pinholes with lenslets

http://graphics.cs.cmu.edu/courses/15-463/lectures/lecture_10.pdf

Applications of Light-Field Cameras



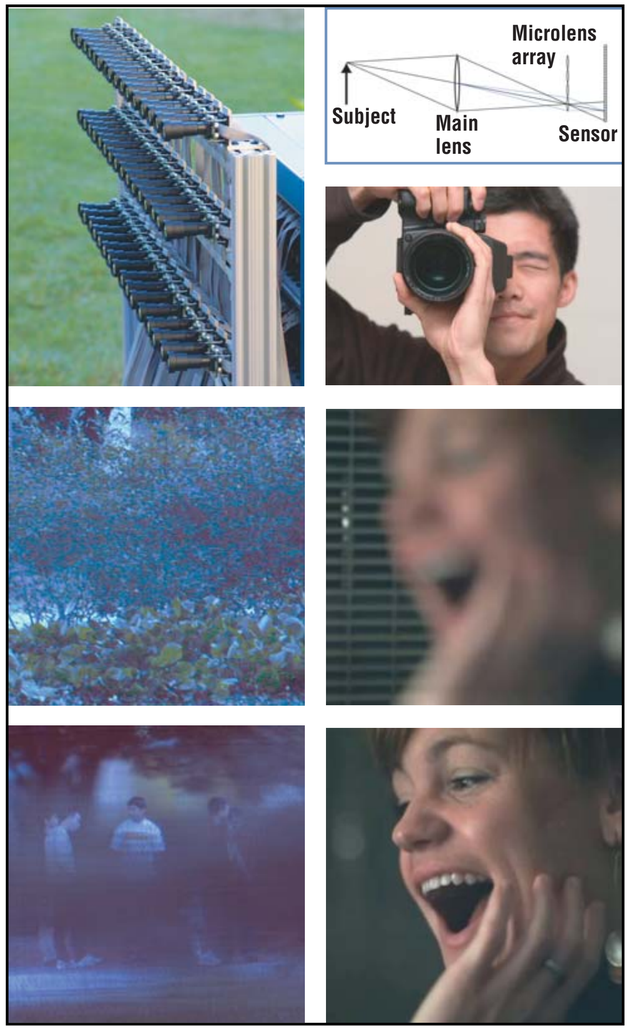
Digital re-focusing

- P is in-focus but Q is out of focus
- If we can estimate the radiance of all the yellow rays inside the camera, we can digitally sum up all those rays to synthesize the color of Q as if the sensor was placed to focus on Q.

Extended depth of field

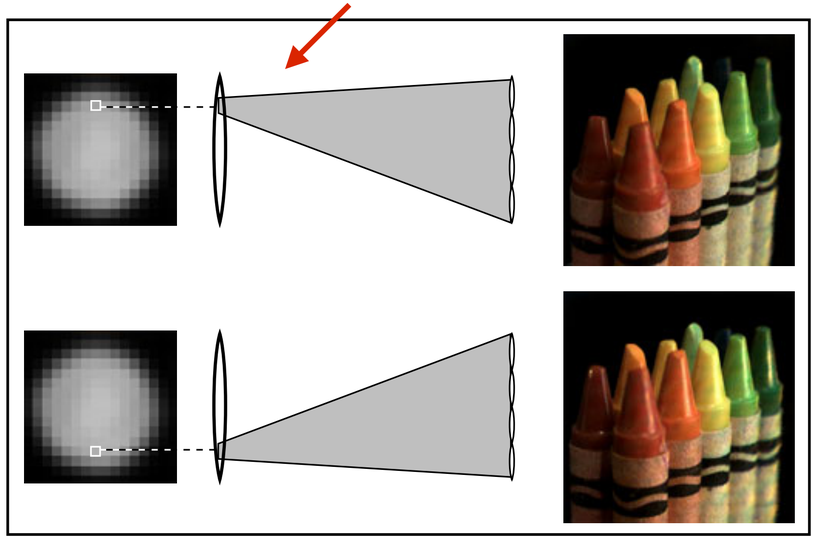
- Focusing on P and Q together.

Applications of Light-Field Camera



Refocusing

Changing perspective (extracting the shown pixel under each microlens)

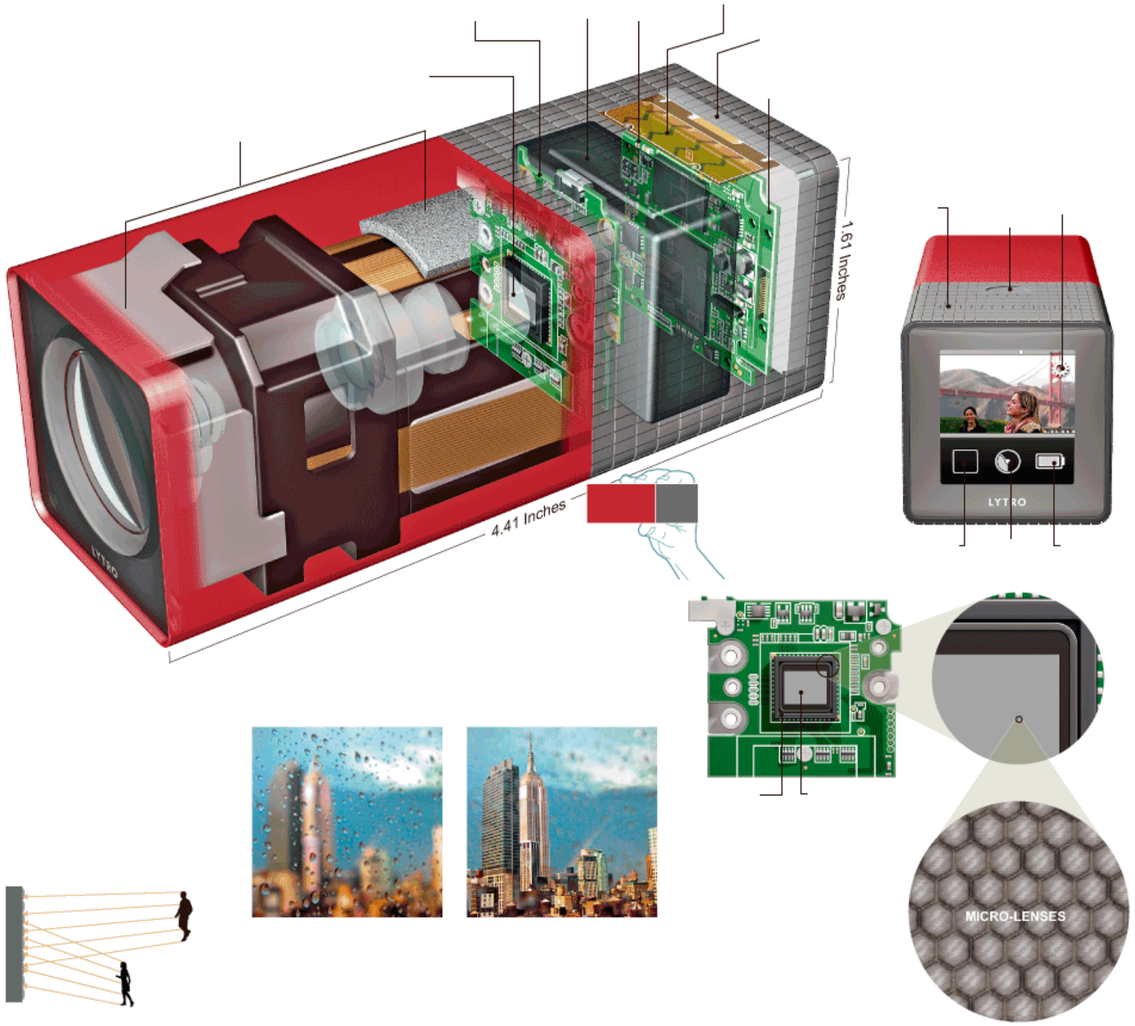


Extended Depth of Field (stacking photos focused at different depths.)

Light Fields and Computational Imaging, IEEE Computer, 2006

Light Field Photography with a Hand-held Plenoptic Camera, Stanford CSTR, 2005

Commercial Plenoptic Camera: Lytro



Acquired by Google; pivoting from consumer photography to immersive (VR) content creation.

https://archive.nytimes.com/www.nytimes.com/interactive/2012/03/01/business/inside-the-lytro.html?_r=0

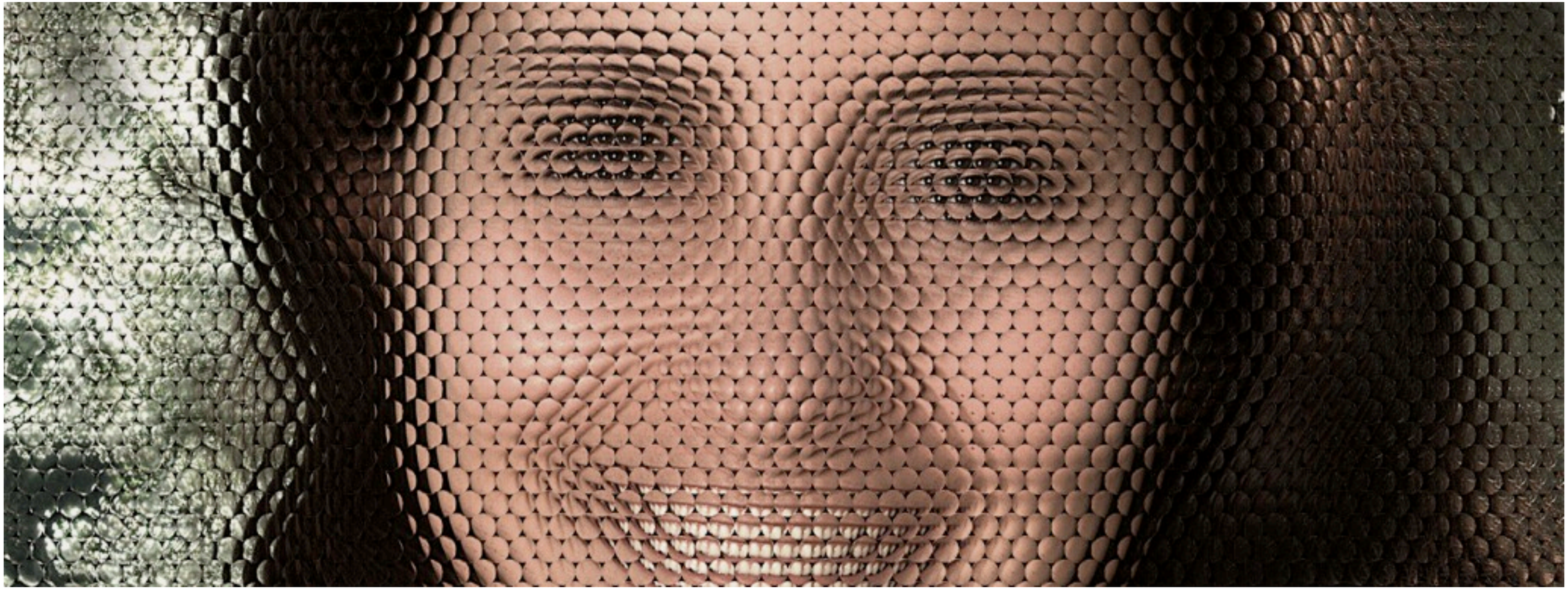
Industrial plenoptic cameras

Plenoptic cameras have become quite popular in lab and industrial settings.

- Much higher resolution, both spatial and angular, than commercial cameras.
- Support interchangeable lenses.
- Can do video.
- Very expensive.

Image Captured by a Plenoptic Camera

Can you explain why we see this pattern?



How Things Evolved

Gabriel Lippmann
(Noble Physics Prize; 1908)

Integral Photography

A New Discovery by Professor Lippmann

By the French Correspondent of the Scientific American

PROF. LIPPMANN of Paris is working on a very remarkable new photographic method which he has termed "integral photography." The nature of this is best explained by reference to the accompanying diagram. It should be remarked in the first place that the process is designed to work without a camera. The sensitized plate is coated with a large number of very small globules of glass or other transparent material. In use each of these acts as a miniature eye, with a sensitive plate for its retina. The globules are very small, less than one-hundredth of an inch in diameter. In front each has a little blob which acts as the lens of the miniature eye. Suppose an object to be situated at M M^1, as indicated in the diagram. Consider any globule. Rays from the points M M^1 will pass through the minute lens, and an image will be formed at the back of the globule upon the sensitized plate at M M^1. Thus each globule will contain a minute picture of the objects placed in front of the plate. This picture is brought out by developing the plate and fixing it in the usual way. This of course yields a negative, which must be reversed by the use of a suitable reversing bath. The result is a transparency, in which all the minute pictures in the individual globules are combined into one image of the original object. To view it, a good light is placed behind the plate, so as to send light through all the globules. Suppose the eye viewing the transparency to be placed at O, the point which when the picture was taken, lay on a line joining one globule with the point M of the object. The process which now occurs will be the opposite of what happened when the picture was taken. In other words, the eye will receive from the globule a view of the image m of the point M. Other globules will send images of the point M^1, etc., at the same time, so that the eye will see the different portions of the image coming from all the different globules. In this way a complete image of the object originally pre-

is changed in accordance with the direction in which the eye is looking at the globule.

After working out this remarkable principle, Prof. Lippmann presented it before the French Academy of Sciences and other scientific societies, where it awakened great interest. The realization of the process presents considerable technical difficulties. Owing to their small size, it was out of the question to shape each one of the globules separately. Prof. Lippmann tried using minute glass beads, such as are found in commerce and used for covering ornamental surfaces, such as picture post cards. He found, however, that these were too irregular in shape and size. Another idea which suggested itself was to stamp out a transparent sheet of collodion and gelatine, giving the surface the requisite shape. But the dies required were found to be too difficult to make in the laboratory, so that Prof. Lippmann has been forced to abandon his attempts in this direction for the time being. The fact is that the working out of a process of this kind is a technical

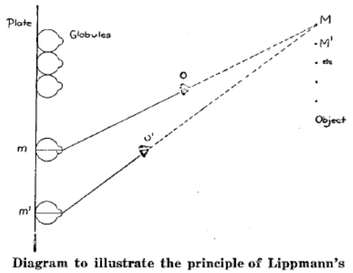


Diagram to illustrate the principle of Lippmann's integral photography.

on the plate, the portions of the plate illuminated by the several lenses being kept separate by an arrangement of black cardboard cells or partitions. One of our views represents the front of the camera with a set of twelve lenses, while the rear view shows the cells in the plate-holder. This latter carries a large plate of such size as to receive all the twelve pictures, and is placed in the camera as usual. The plate-holder, however, is made with a slide in the back as well as in the front, for a purpose which will appear presently. After the picture is taken, the negative is developed as usual, and is reversed either by means of a reversing bath, or by copying it, so that a transparent positive is obtained. The twelve pictures are not quite alike, for each lens forms its image from a somewhat different point of view. The transparency is put back in the holder, both slides of which are now opened, while light is sent through from behind. The observer looks in through the lenses with both eyes, when he sees a single view in relief of the object photographed. On moving the head from side to side, or up and down, the same effect is observed as would be under similar circumstances when looking at the real object, that is to say, objects which cover one another when looked at from one point, are seen to separate when viewed from another. Even with this simple apparatus the effect is very pleasing.

Proposed Aeronautic Map

AT a recent meeting of the Academie des Sciences, M. Lallemand, chief of the French Government Survey Department, presented the project which the Aeronautic Commission intends to carry out establishing an aeronautic map. The proposed map is of 1:200,000 size, and is drawn up after a provisory model made by the Aero Club. Each plate of the aeronautic map will be a sheet taking in one degree in latitude and longitude and there will be twenty-

EPREUVES RÉVERSIBLE

donc un large faisceau qui converge vers A (voir *fig.* 1) : c'est un faisceau large, puisqu'il a pour base toute la plaque sensible, ou du moins toute la partie de cette plaque d'où le point A était visible (1).

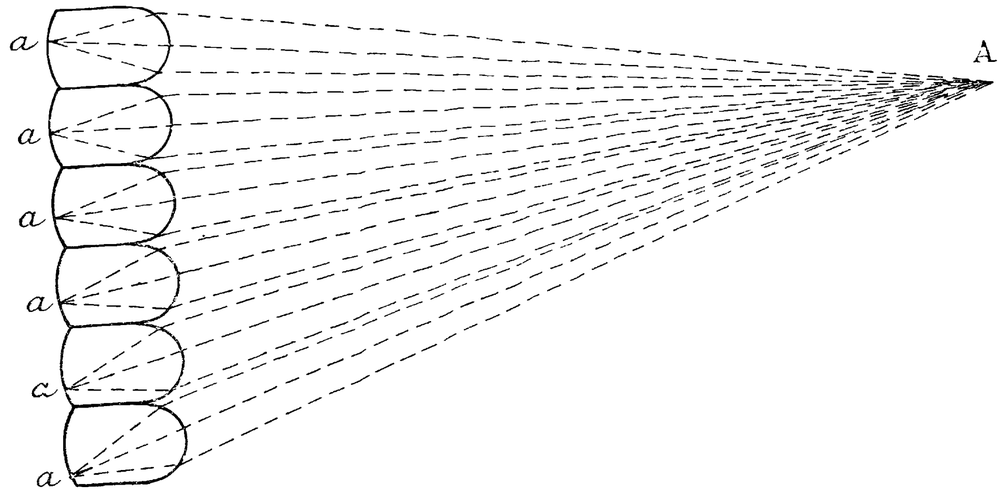


FIG. 1.

How Things Evolved

PARALLAX PANORAMAGRAMS MADE WITH A LARGE DIAMETER LENS

By Herbert E. Ives

Parallax panoramagrams—pictures showing stereoscopic relief through a wide range of distances and angles of observation—have heretofore been made by any one of several kinds of camera, in which the object, the photographic lens, the opaque line grating and the sensitive plate are given a relative motion during the progress of the exposure. Details of these cameras, and of the characteristics of the pictures made by them may be obtained by reference to an earlier publication.[1] In all cases, whether separation of the grating and plate is depended on to cause the development of the panoramic strips by parallax, or a relative motion is given to plate and grating, a common characteristic is the motion of the lens with respect to the object, whereby different points of view are successively projected upon the grating and plate. This note is for the purpose of describing a method in which the moving lens is dispensed with, making possible a parallax panoramagram camera containing no moving parts whatever.

OPTICAL TECHNOLOGY VOLUME 37, NUMBER 7 JULY 1970

Autostereoscopy and Integral Photography

UDC 778.39 + 778.4

Yu.A. Dudnikov

A brief description of autostereoscopic techniques for obtaining three-dimensional images is presented; the advantages and faults of integral photography are analyzed. A comparison is made between holography and integral photography, and the prospects of this latter method are indicated.

The viewing of stereo-photographs without the aid of special optics (hence the term autostereoscopy), is a long-sought goal. As of now, the recently developed holography technique produces stereo images with the largest amount of information presently possible. However, its practical utilization is difficult since it requires coherent illumination.

The concept of holography, i.e., fixation of an image by photographing the pattern of interference of waves incident on the subject and reflected by it [1], derives from the color (non-stereoscopic) photography of G. Lippmann (1894), in which interference nodes and antinodes of the light waves incident on the photographic plate and reflected by its back mirror surface are fixed within the light-sensitive layer.

Autostereoscopic methods which do not require coherent illumination, such as the parallax stereogram of Berthier (1896) [2] and the parallax panoramogram of Kanolt (1915) [3], were successful in their time, and in improved form are being used even today.

The parallax stereogram is a replica from a normal stereoscopic pair, made optically through a line pattern, i.e., through a screen of alternating vertical transparent and opaque lines. The lenses of the objective of the printing device superimpose the images of the two stereoscopic negatives on one another, but the screen subdivides them into narrow alternating strips. One must view the combined image through the same screen, set up in such a way that the right eye sees the strips belonging to the left negative and the left eye those of the right negative (Fig. 1). The image thus produced is half as bright as that obtained with the same illumination in a stereoscope, and contains less visual information, since only half of the area of each frame is viewed (and this without magnification). The stereoscopic effect is satisfactory but in order not to lose it, both the photograph and ones head must be perfectly stationary.

Reappeared under different forms and names throughout the century.

- The left paper is from 1930, the right one from 1970.

http://graphics.cs.cmu.edu/courses/15-463/lectures/lecture_10.pdf

How Things Evolved

IEEE PAMI 1992; built for depth estimation, like phase-detection AF.

IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 14, NO. 2, FEBRUARY 1992

Single Lens Stereo with a Plenoptic Camera

Edward H. Adelson and John Y.A. Wang

Abstract—Ordinary cameras gather light across the area of their lens aperture, and the light striking a given subregion of the aperture is structured somewhat differently than the light striking an adjacent subregion. By analyzing this optical structure, one can infer the depths of objects in the scene, i.e., one can achieve "single lens stereo." We describe a novel camera for performing this analysis. It incorporates a single main lens along with a lenticular array placed at the sensor plane. The resulting "plenoptic camera" provides information about how the scene would look when viewed from a continuum of possible viewpoints bounded by the main lens aperture. Deriving depth information is simpler than in a binocular stereo system because the correspondence problem is minimized. The camera extracts information about both horizontal and vertical parallax, which improves the reliability of the depth estimates.

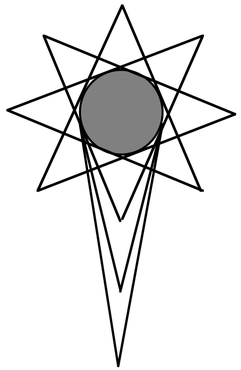
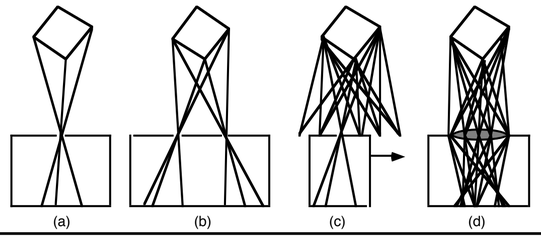


Fig. 1. Diagram from Leonardo's notebooks illustrating the fact that the light rays leaving an object's surface may be considered to form a collection of cones (which Leonardo calls "pyramids"), each cone constituting an image that would be seen by a pinhole camera at a given location.

I. INTRODUCTION

"EVERY BODY in the light and shade fills the surrounding air with infinite images of itself; and these, by infinite pyramids diffused in the air, represent this body throughout space and on every side." Leonardo da Vinci [1] uses these words, together with the drawing in Fig. 1, to describe the relationship between objects' light, and image formation. The object in the drawing sends off rays of light in all directions, and if we choose to place a pinhole camera at any given point in space, we will discover that an image is formed. The image is the projection of a cone of light that Leonardo called a "visual pyramid." The space surrounding an object is densely filled with these pyramids, each representing an image of the object from a slightly different point of view. These infinitely multiplexed



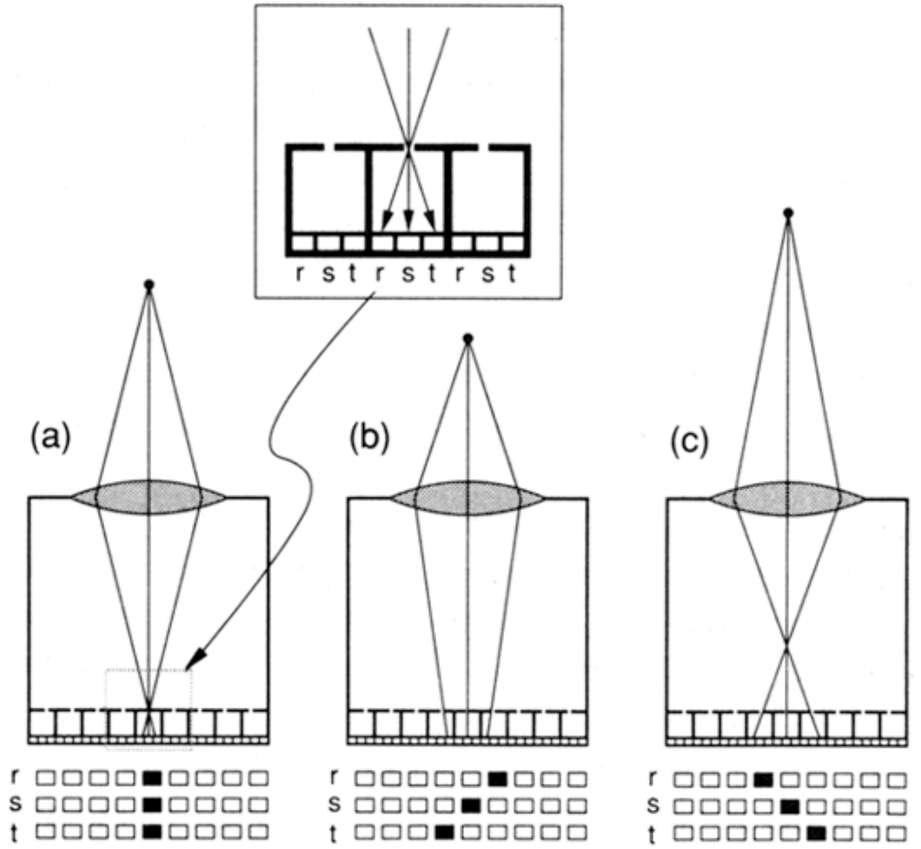
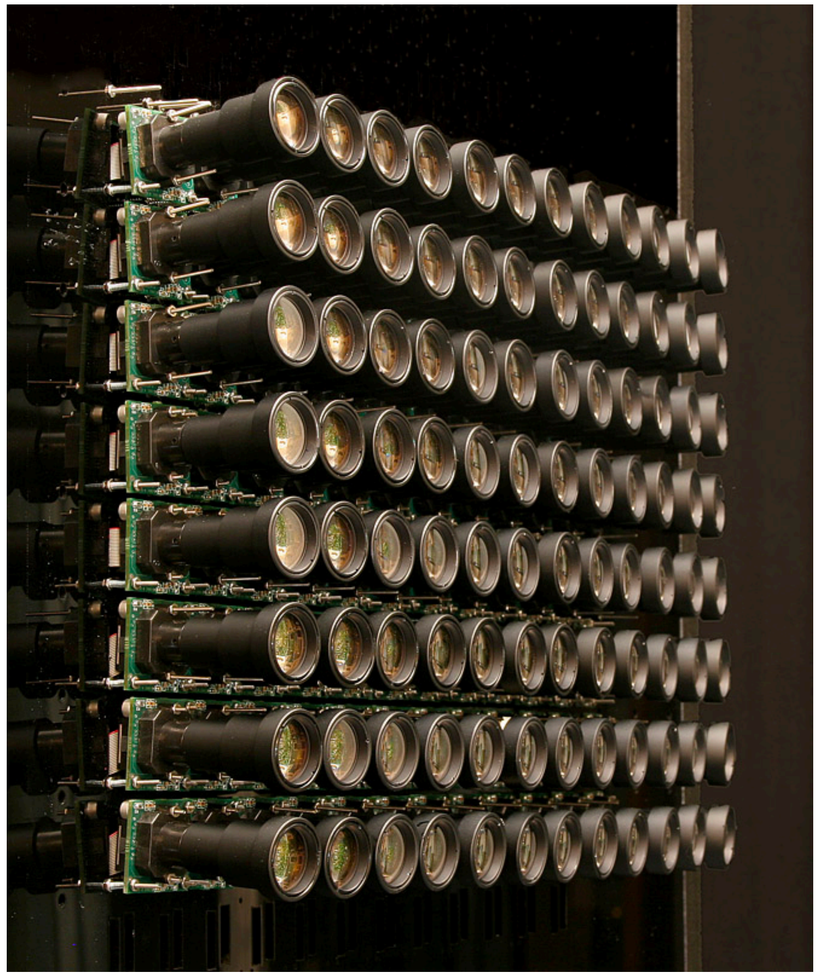


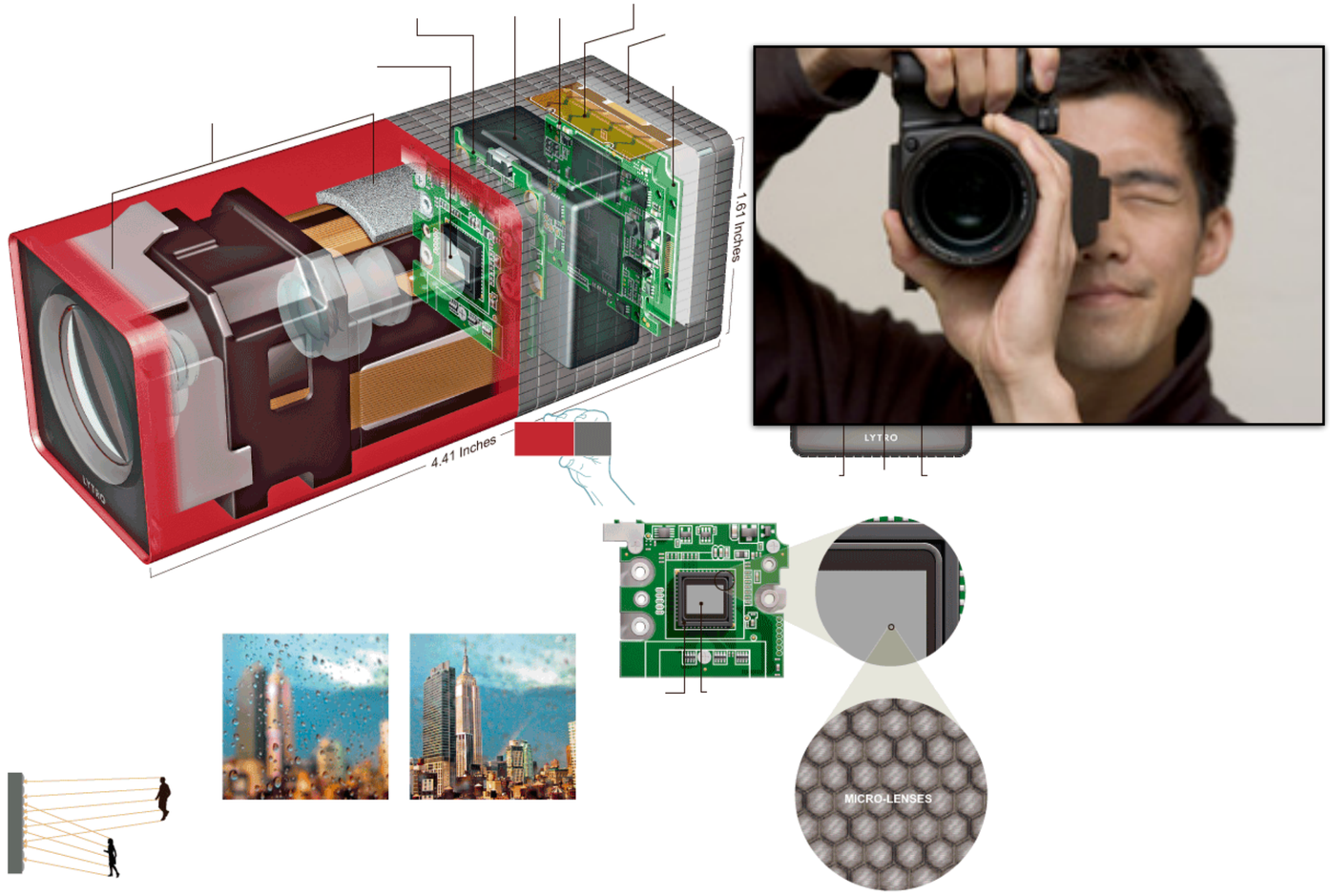
Fig. 5. Array of miniature pinhole cameras placed at the image plane can be used to analyze the structure of the light striking each macropixel.

How Things Evolved

The Stanford Multi-Camera Array, 2006

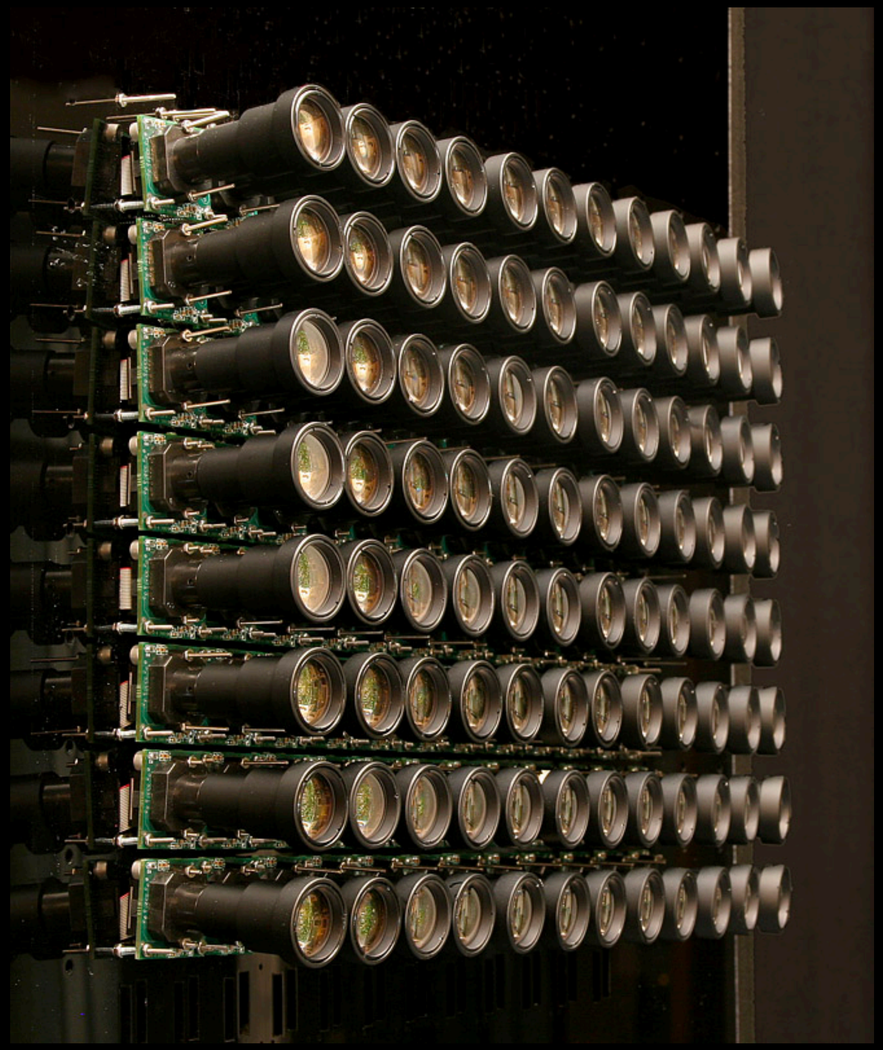


Hand-Held Plenoptic Camera and Lytro, ~2005



Stanford Multi-Camera Array

Early 2000s

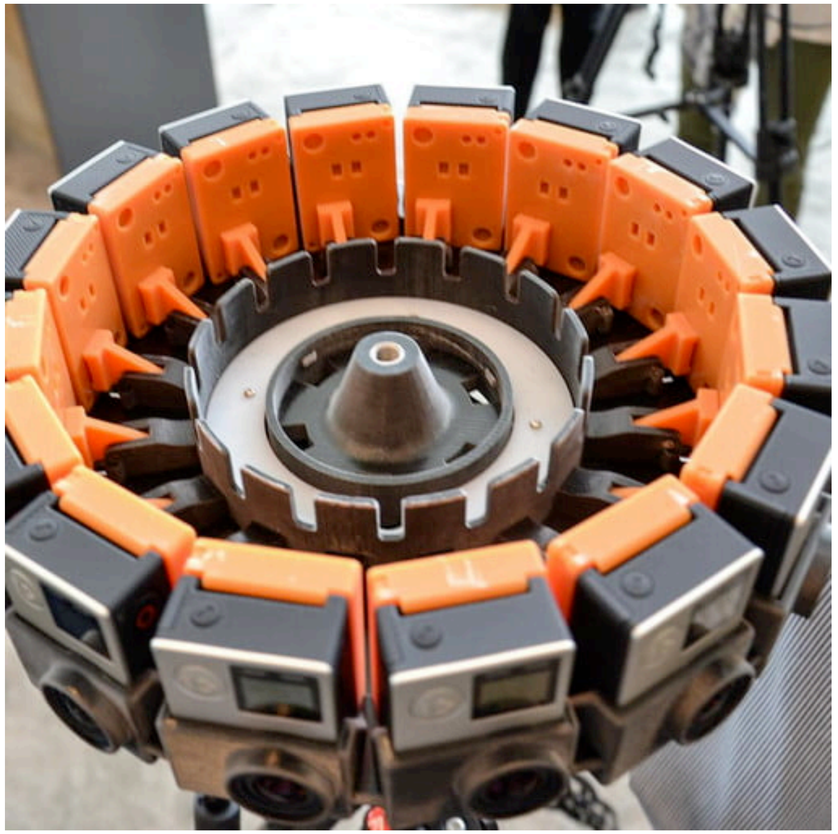


* All the VR camera rigs we saw before fundamentally capture light field, but not at a high angular resolution since each pixel integrates a bundle of rays.

http://graphics.stanford.edu/projects/array/

How Things Evolved

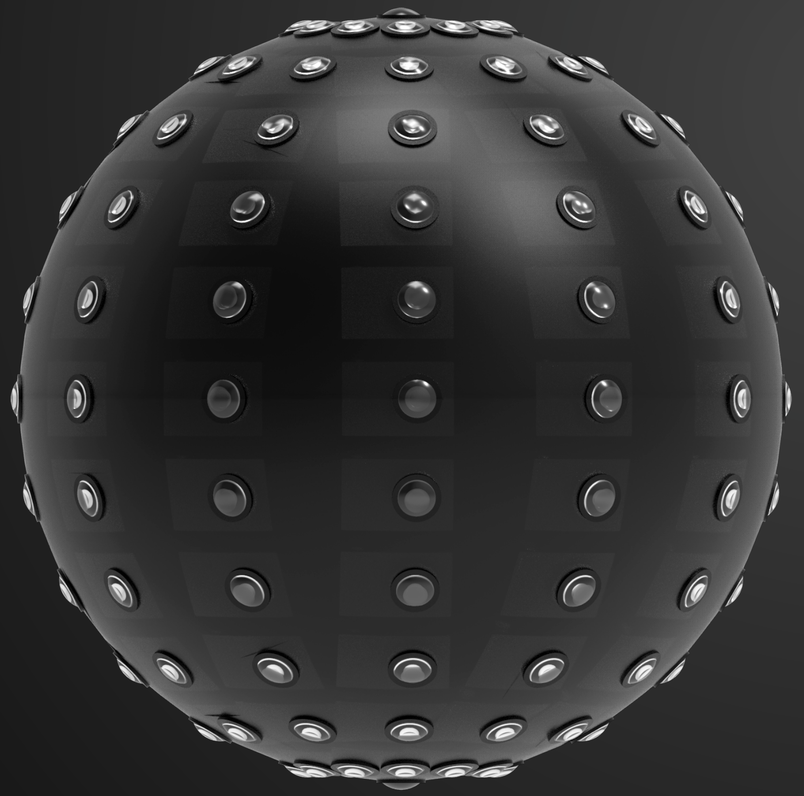
Google Jump VR (2015)

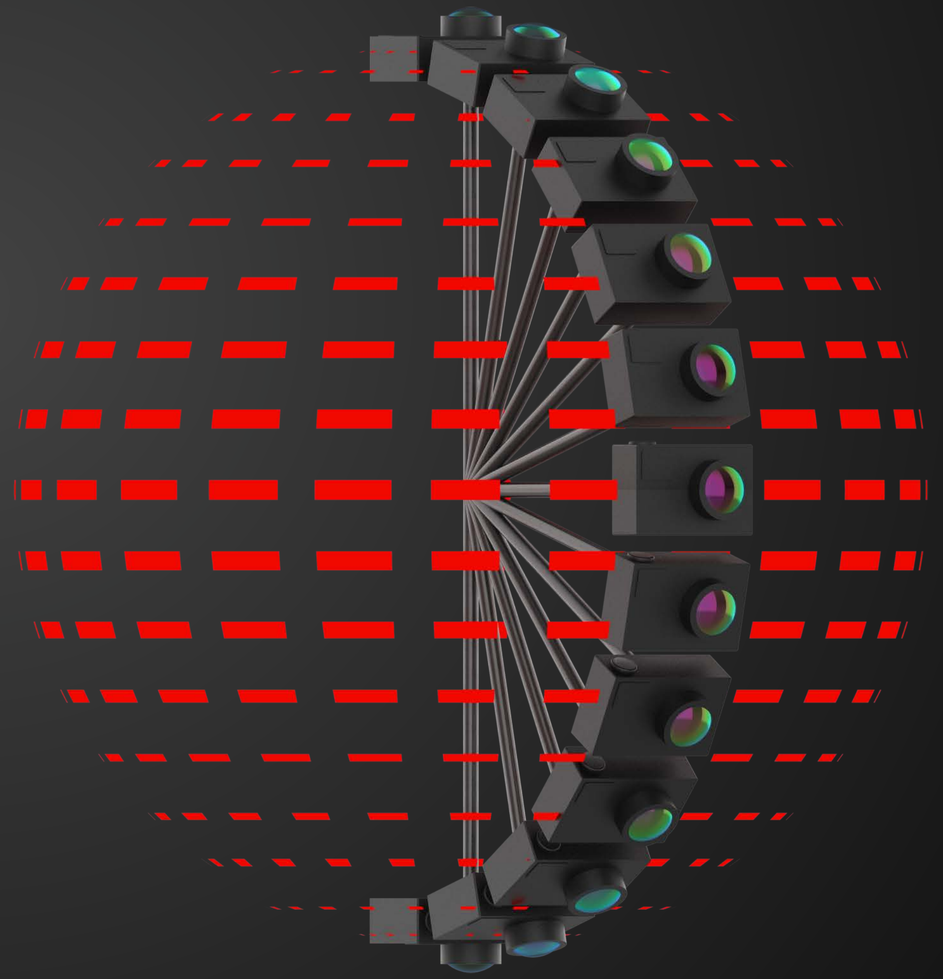


Facebook Surround 360 (2nd gen/2017)

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

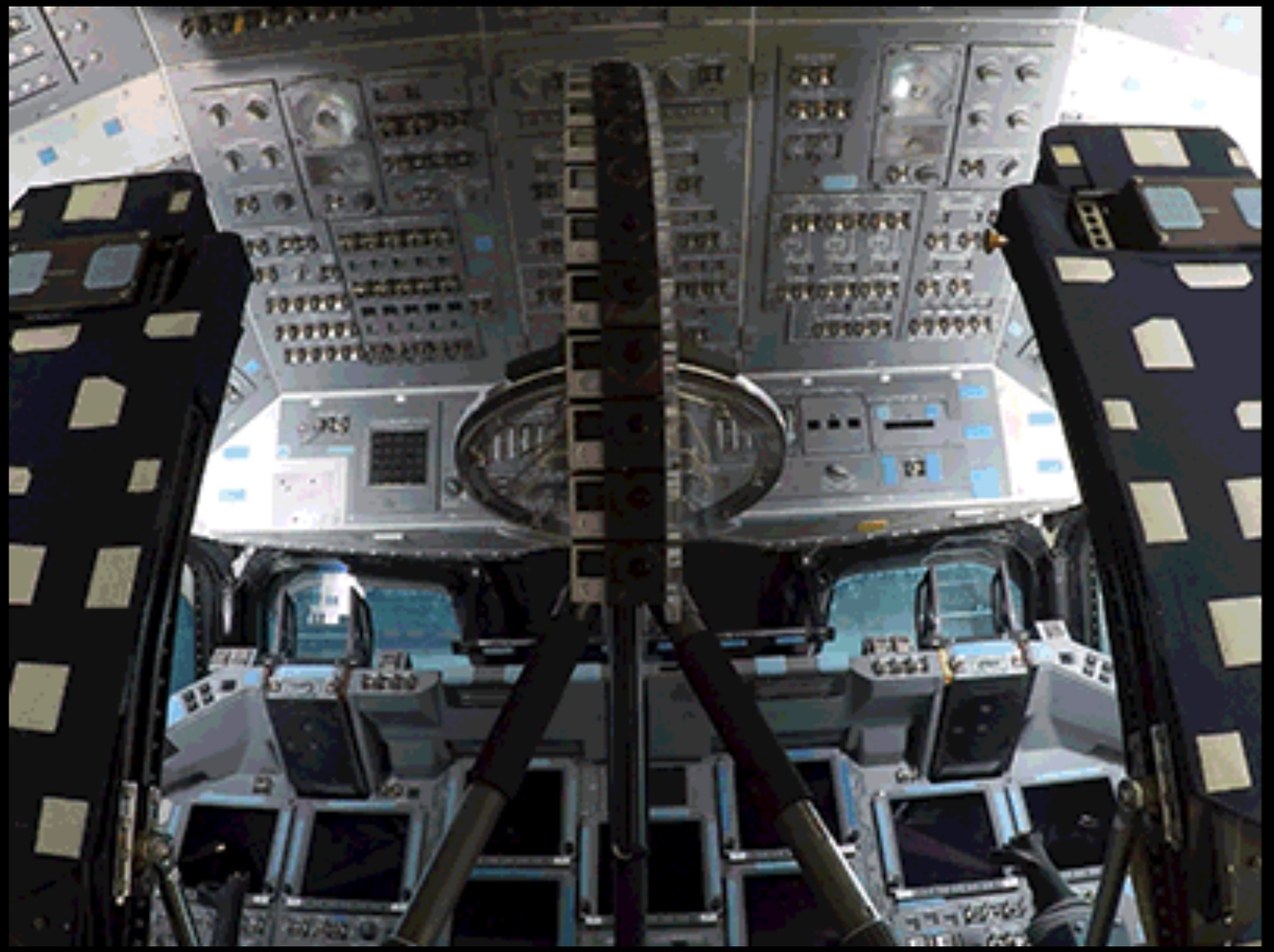


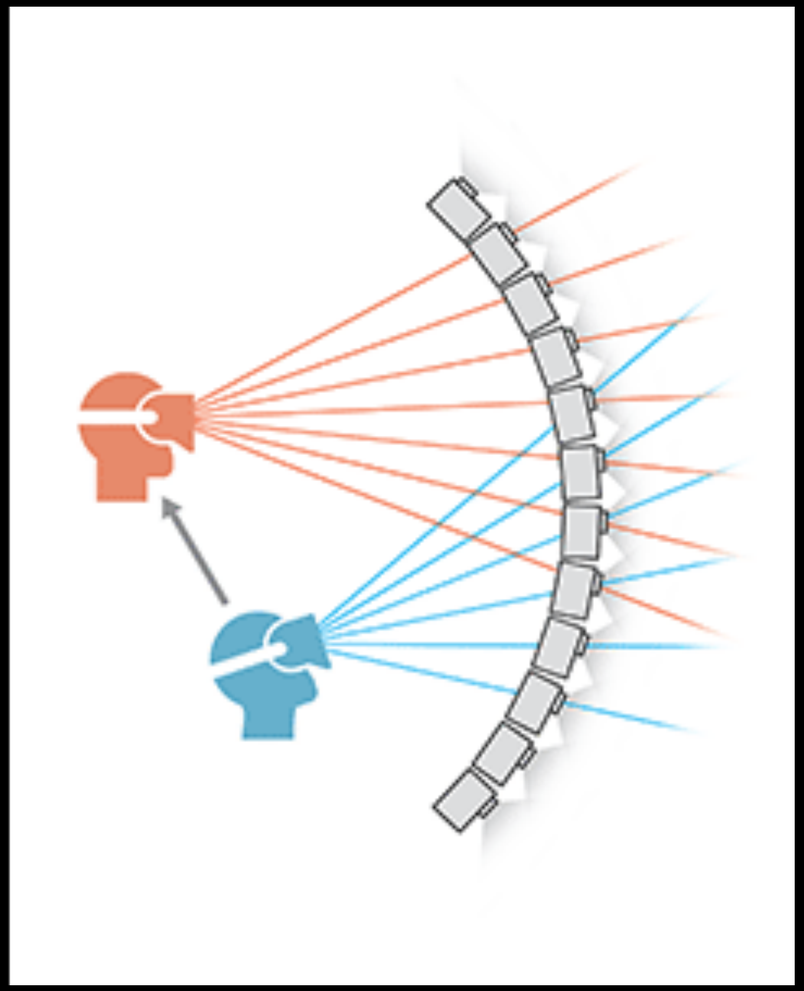


The alternative is to use a sparse, high resolution array

- 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

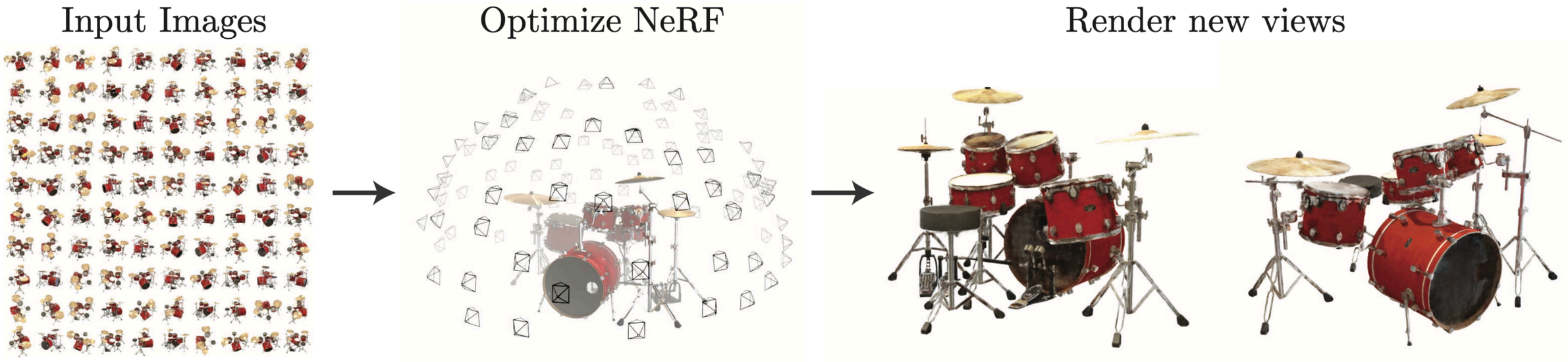
Immersive Light Field Video with a Layered Mesh Representation, SIGGRAPH 2020





https://blog.google/products/google-ar-vr/experimenting-light-fields/

Light Field Today: Neural Radiance Field (NeRF)

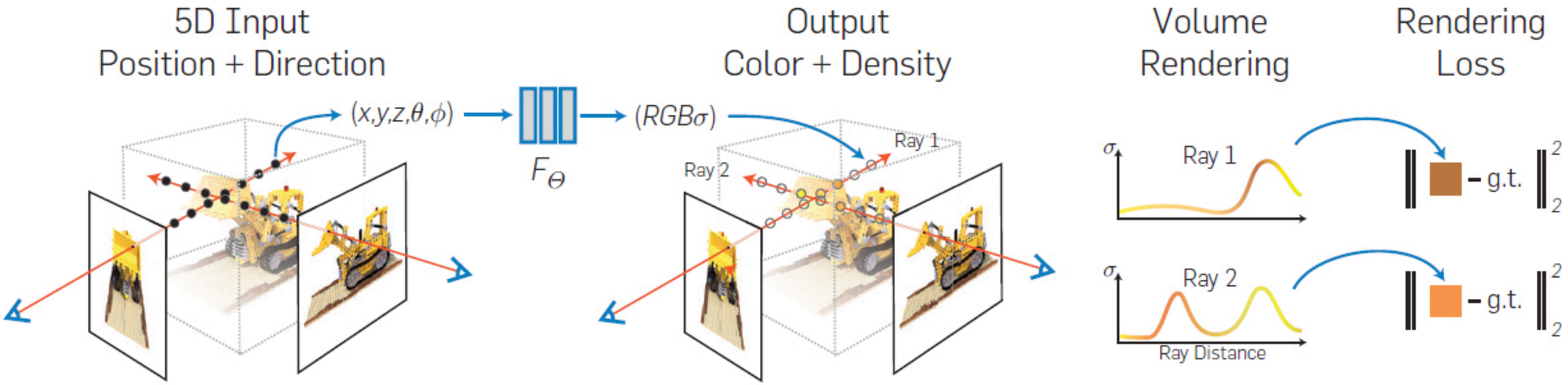


What you need: photos taken from different 6 DoF positions of the scene

- Camera poses are well-calibrated

What you get: photo taken by a virtual camera at any 6 DoF position

Light Field Today: Neural Radiance Field (NeRF)

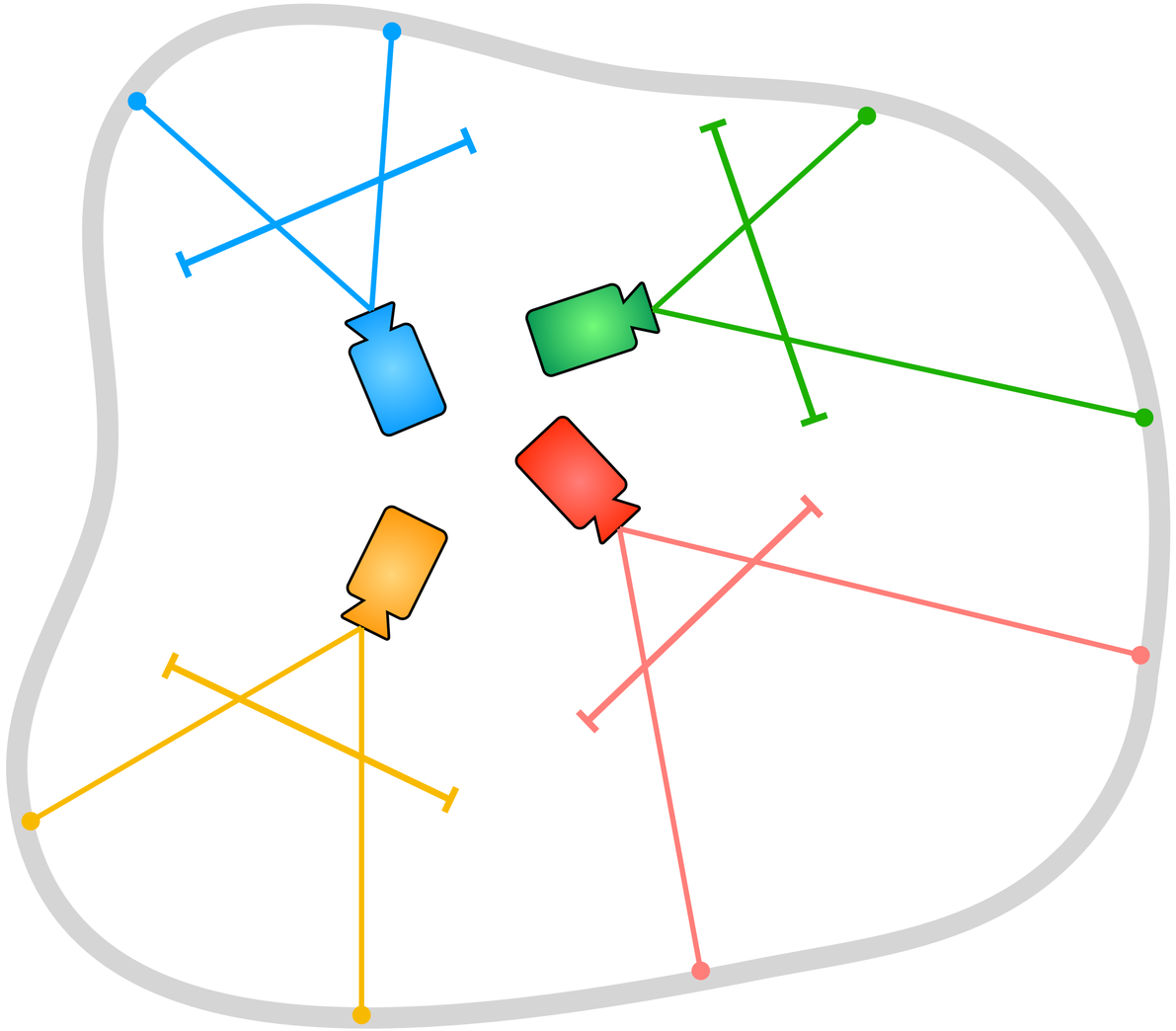


How it works:

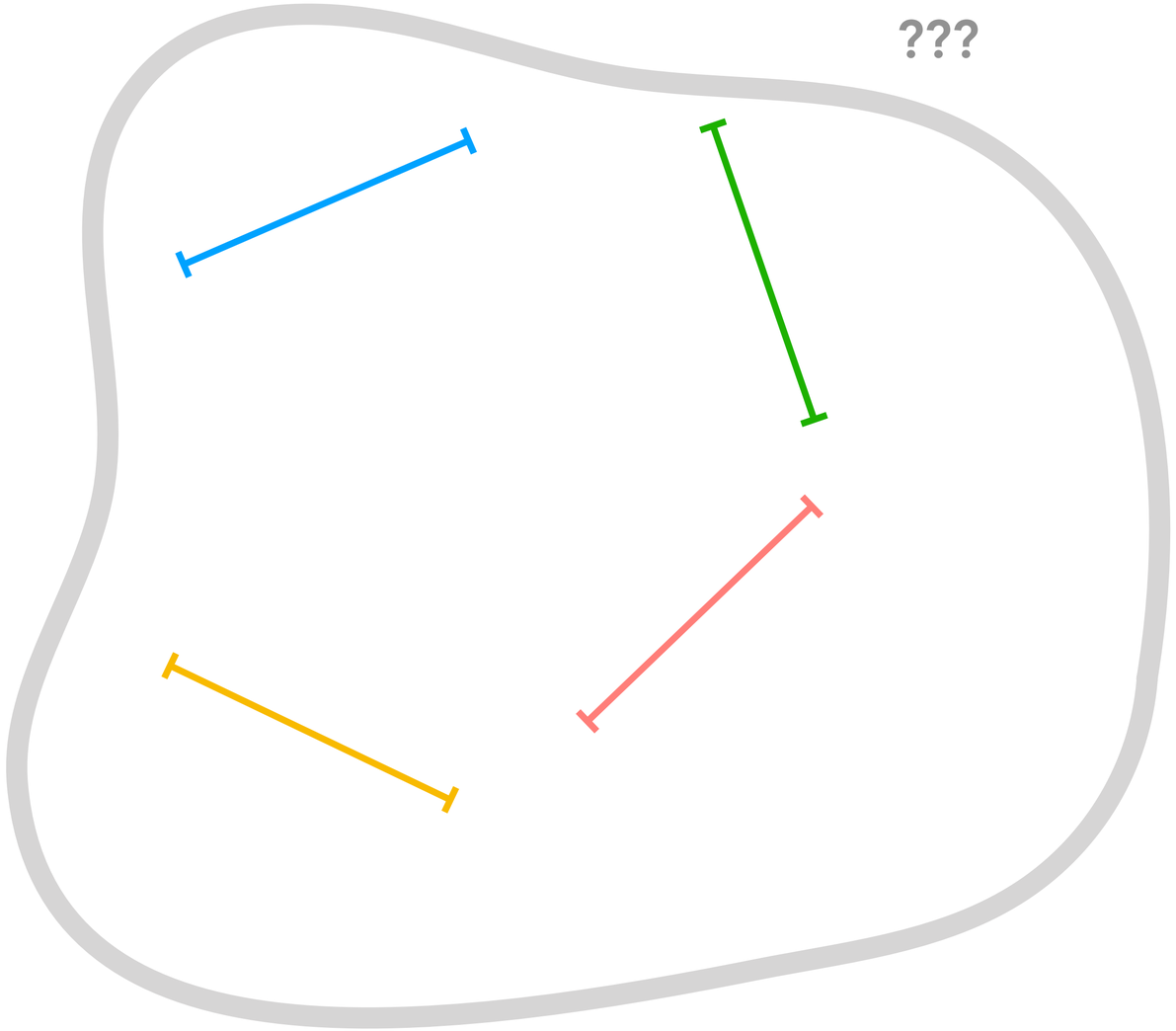
- Train a DNN (fully connected; MLP) for each scene
- DNN takes a spatial position of a ray and its direction and predicts the color
- Use volumetric rendering (later lectures) to render images

https://cacm.acm.org/magazines/2022/1/257450-nerf/fulltext

Aside: Photogrammetry



Aside: Photogrammetry



Constructing 3D scene from uncoordinated photos of the scene.

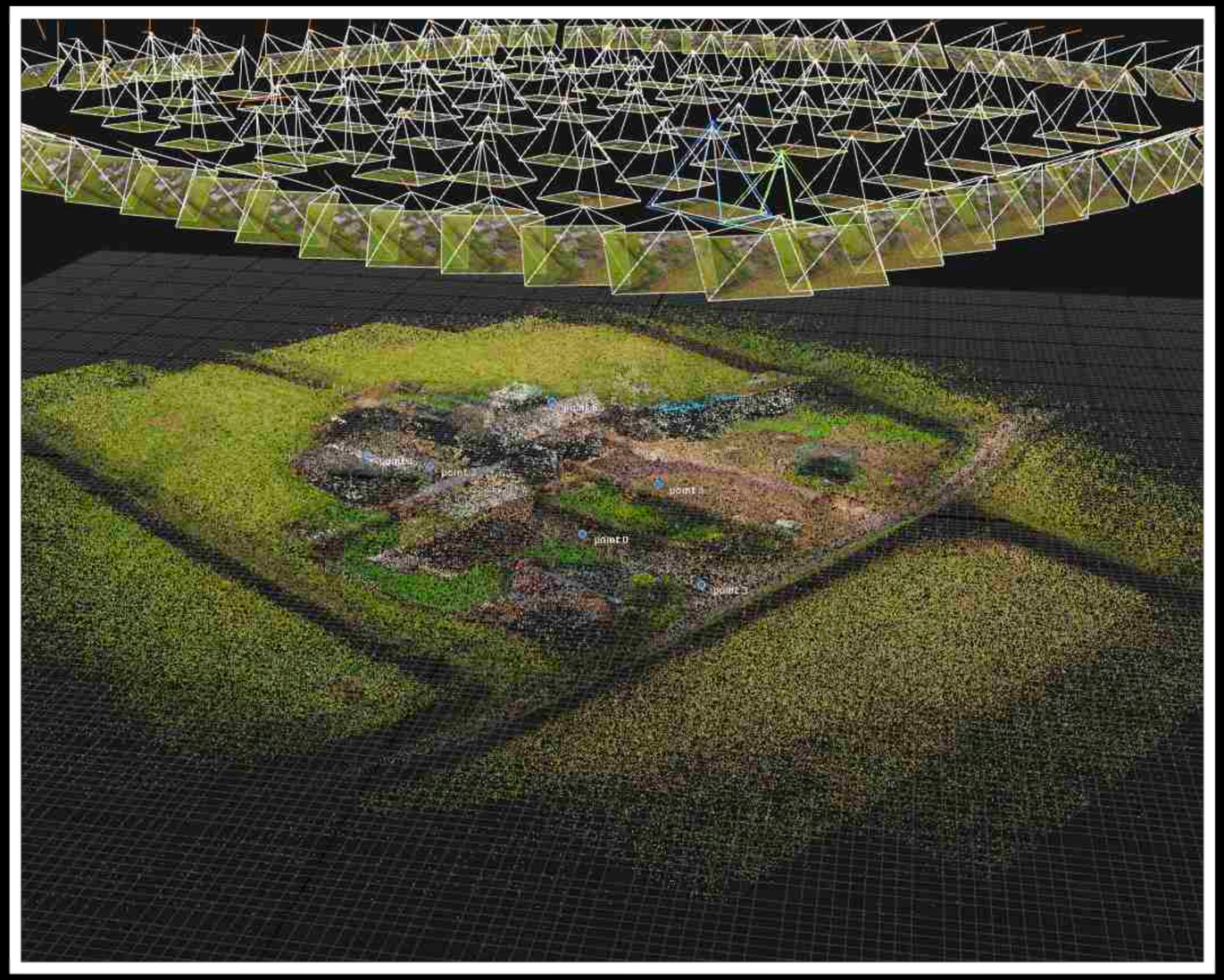
- "Photogrammetry"
- The technical term is "structure from motion" (SfM) or "mapping"
- What we get is colors of scene points, **not light field**.

A large-scale optimization problem

- jointly optimizes scene points and camera poses.
- Also used in robotics localization.

Building Rome in a Day

https://www.cs.cornell.edu/~snavely/bundler/



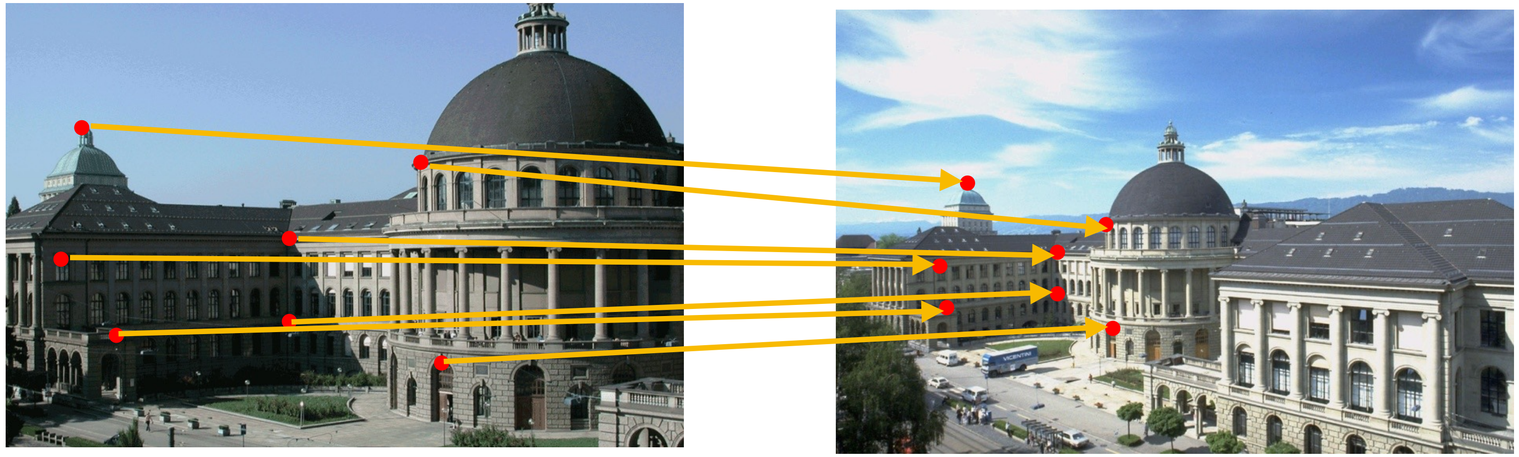
https://crucialservices.net/dronesurvey/drone-mapping/

Working Principle

Take many photos of the scene (in theory from any arbitrary positions)

Align photos (calculating matching pixels across photos)

Bundle Adjustment: find the best-fit camera pose and scene point positions



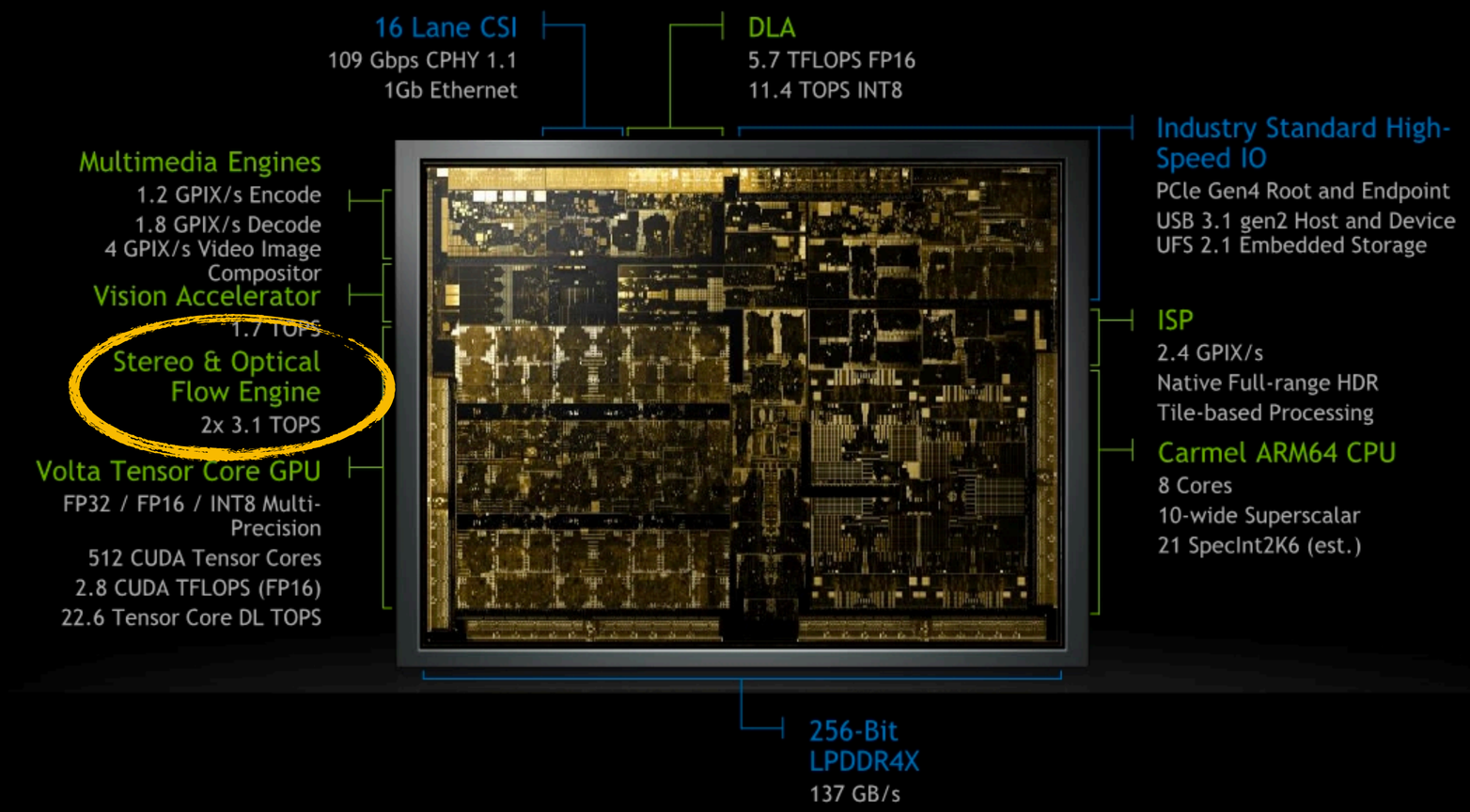
The matching process is in principle the same as in depth estimation from stereo, panorama stitching, and ODS creation. Usually use some form of stereo matching or optical flow algorithm.

XAVIER

World's First Autonomous Machines Processor

Dedicated stereo matching/optical flow hardware!

Throughput:
12.4 TOPS 8-bit
6.2 TOPS 16-bit



Most Complex SOC Ever Made | 9 Billion Transistors, 350mm², 12FFN | ~8,000 Engineering Years

https://en.wikichip.org/wiki/nvidia/tegra/xavier#Stereo_.26_Optical_Flow_Engine

Bundle Adjustment

Take many photos of the scene (in theory from any arbitrary positions)

Align photos (calculating matching pixels across photos)

Bundle Adjustment: find the best-fit camera pose and scene point positions

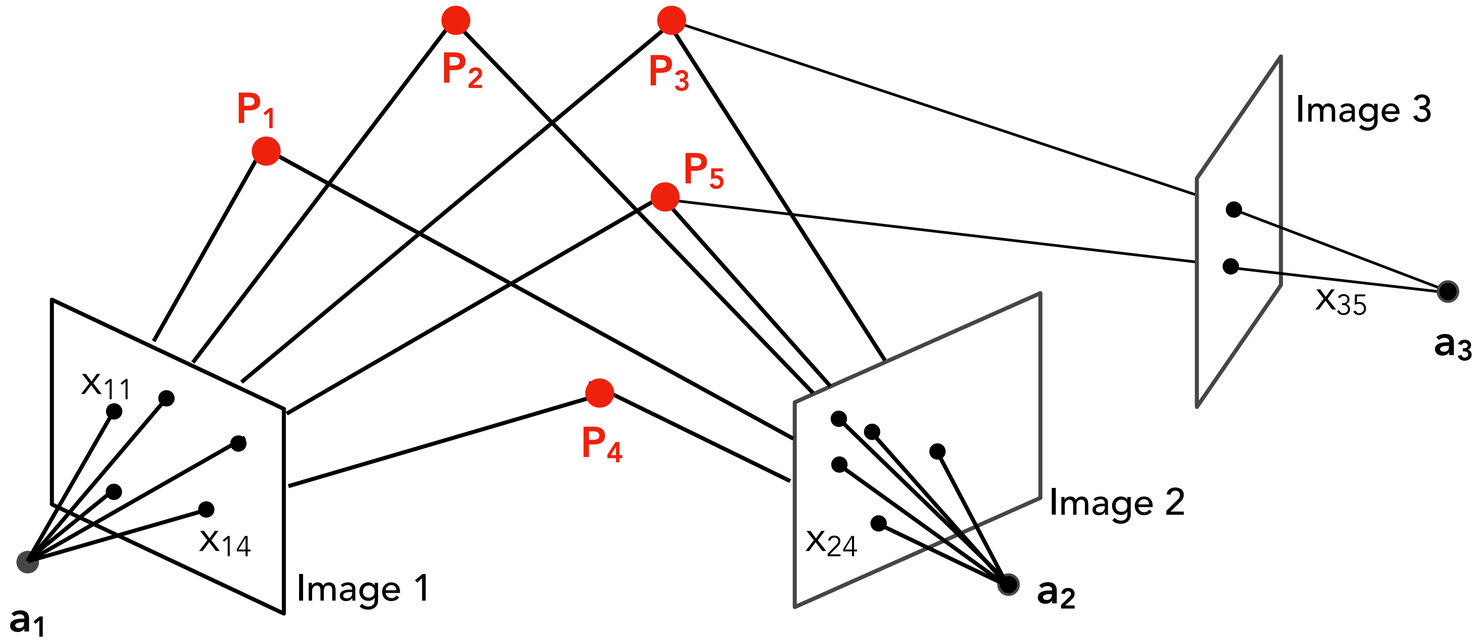
Perspective projection of point P_j to camera a_i

$$\min_{a_i,\, P_j} \sum_{i=1}^{m} \sum_{j=1}^{n} v_{ij} \, \|\mathbf{Q}(a_i, P_j), x_{ij}\|$$

Camera 6 DoF pose

Scene point coordinates

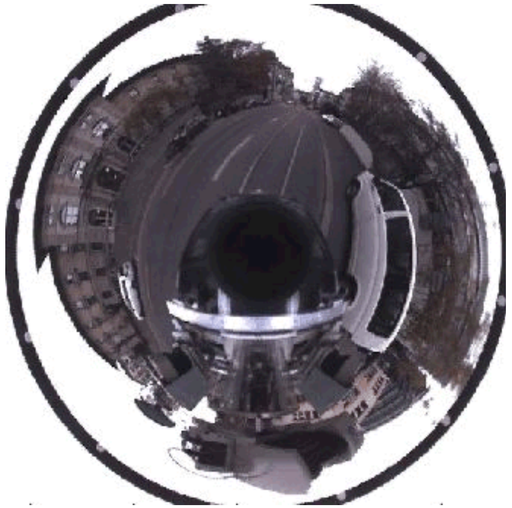
Whether point P_j is scene by camera a_i

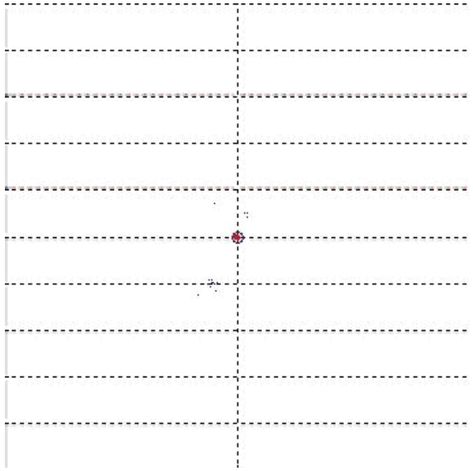


Aside: Robotic Localization

BA can be used in robotic localization.

- Actually it's **Simultaneous Localization and Mapping** (SLAM), because from BA we get both a map of the environment and the robot location.
- The camera poses are essentially agent positions (rotation and translation).
- Also called ego-motion or pose estimation.
- Apply BA in a local, small window to reduce overhead. Global BA is very slow.
- Many other SLAM algorithms exist (take a robotics course).





Summary: All are Light-Field Imaging and Image-Based Rendering in Disguise

Image-based rendering is all about signal sample and reconstruction.

The underlying signal is the light-field, which is hard to directly measure.

Each pixel in an image is a sample of a small portion of the light field (with integration/box filter applied on the signal first).

Rendering a new image at a given camera pose (novel-view synthesis) is ideally best done by first reconstructing the underlying light field and then resample it at the given camera pose.

Summary: All are Light-Field Imaging and Image-Based Rendering in Disguise

Panoramic photography
(3 DoF)

Stereo VR
(3 DoF + Stereo)

NeRF
(6 DoF)