Lecture 7:

The Light Field, and Capture for VR

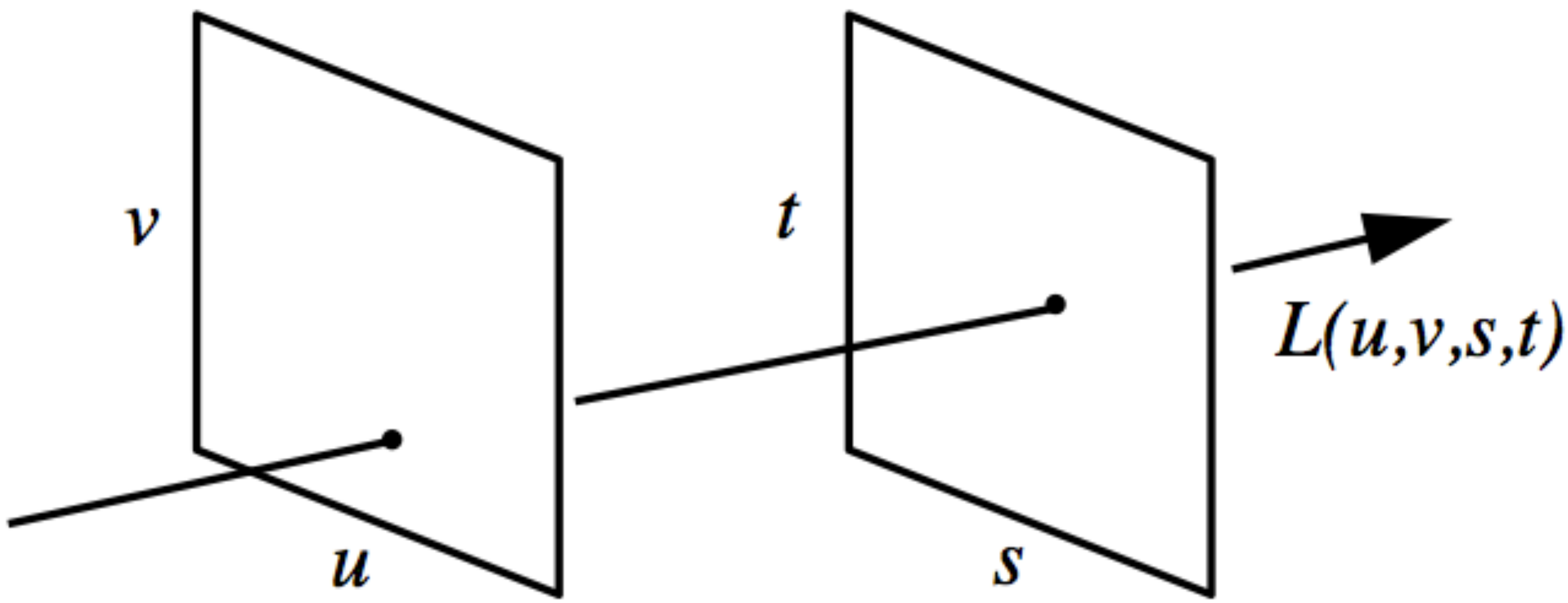
Visual Computing Systems
Stanford CS348K, Fall 2018

Let’s think about all the “rays of light” in this room

Light-field parameterization

[Levoy and Hanrahan 96]
[Gortler et al., 96]

Light field is a 4D function (represents light in free space: no occlusion)



[Image credit: Levoy and Hanrahan 96]

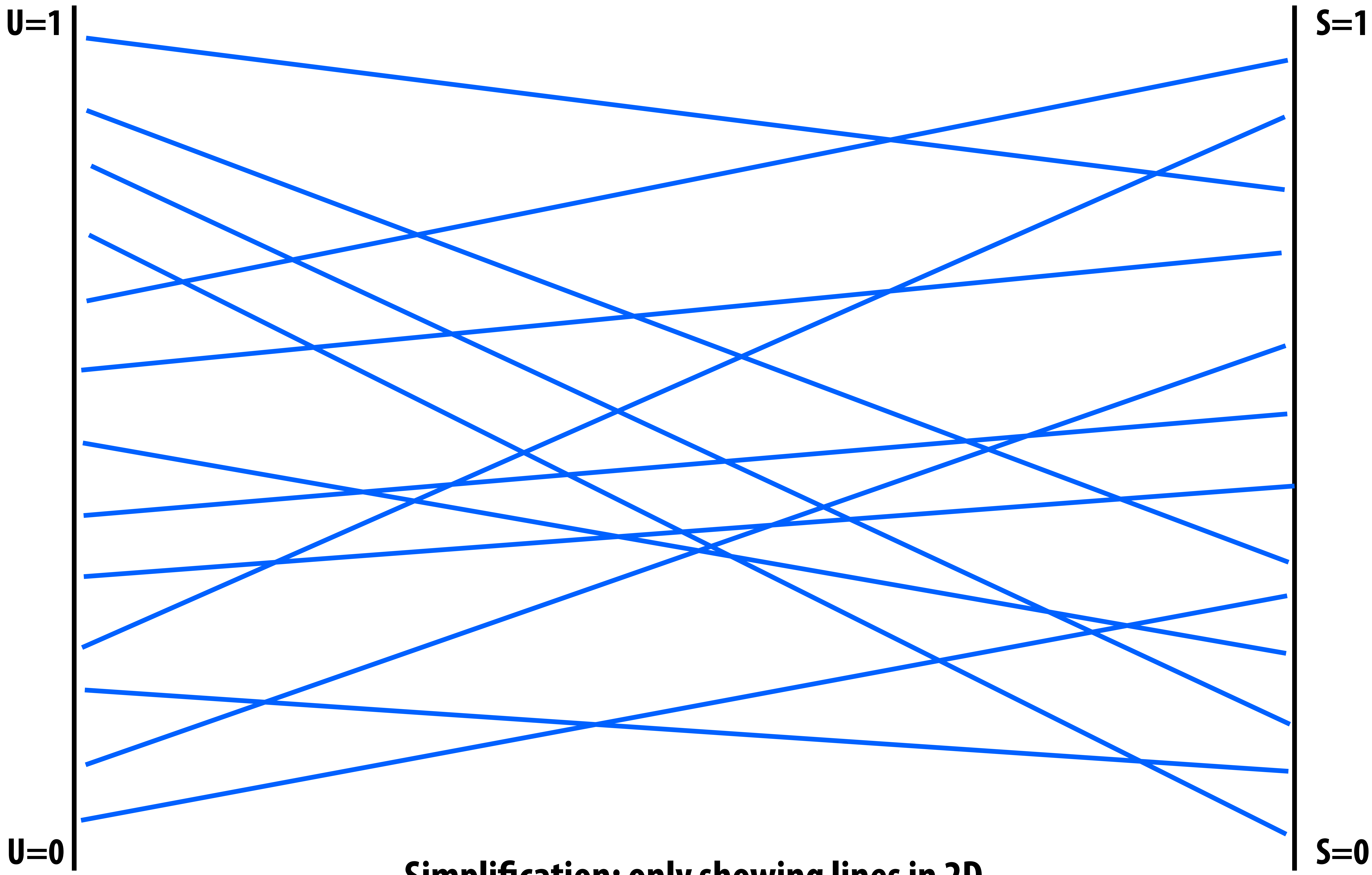
Efficient two-plane parameterization

Line described by connecting point on (u,v) plane with point on (s,t) plane

If one of the planes placed at infinity: point + direction representation

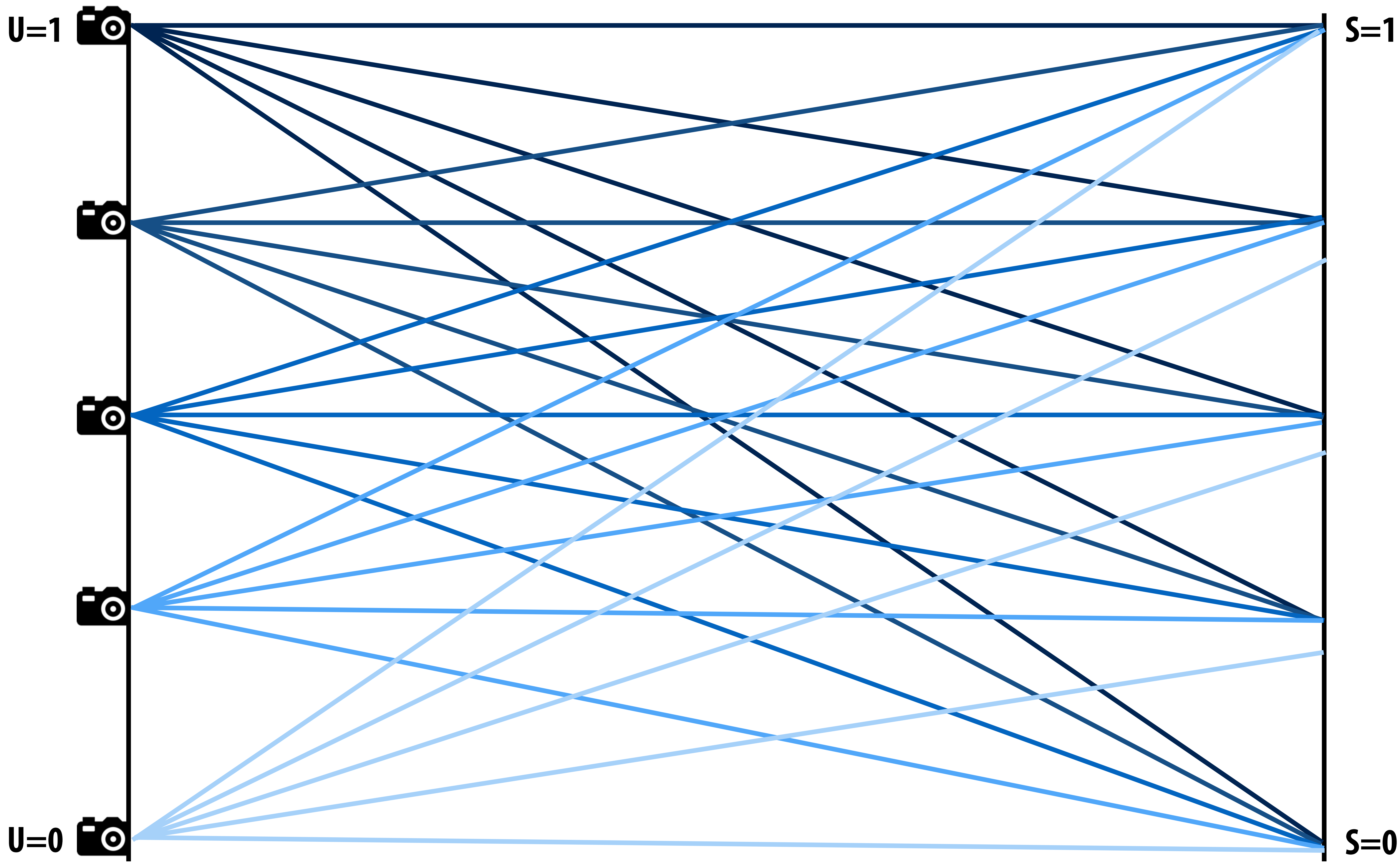
Levoy/Hanrahan refer to representation as a “light slab”: beam of light entering one quadrilateral and exiting another

Sampling the light field



Simplification: only showing lines in 2D
(full light field is 4D function)

Sampling the light field by taking pictures



[Credit: Camera icon by Venkatesh Aiyulu from The Noun Project]

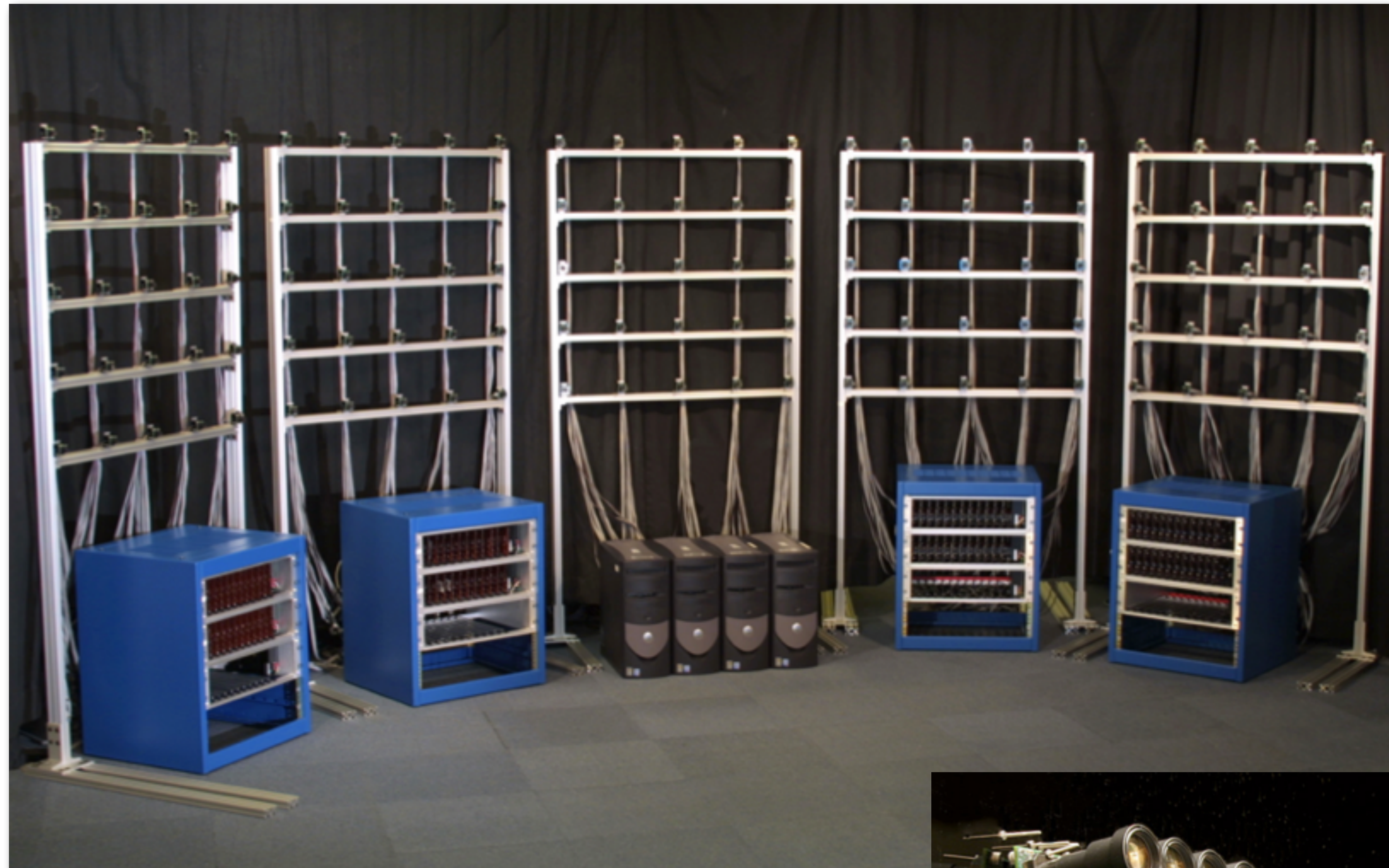
Stanford Camera Array

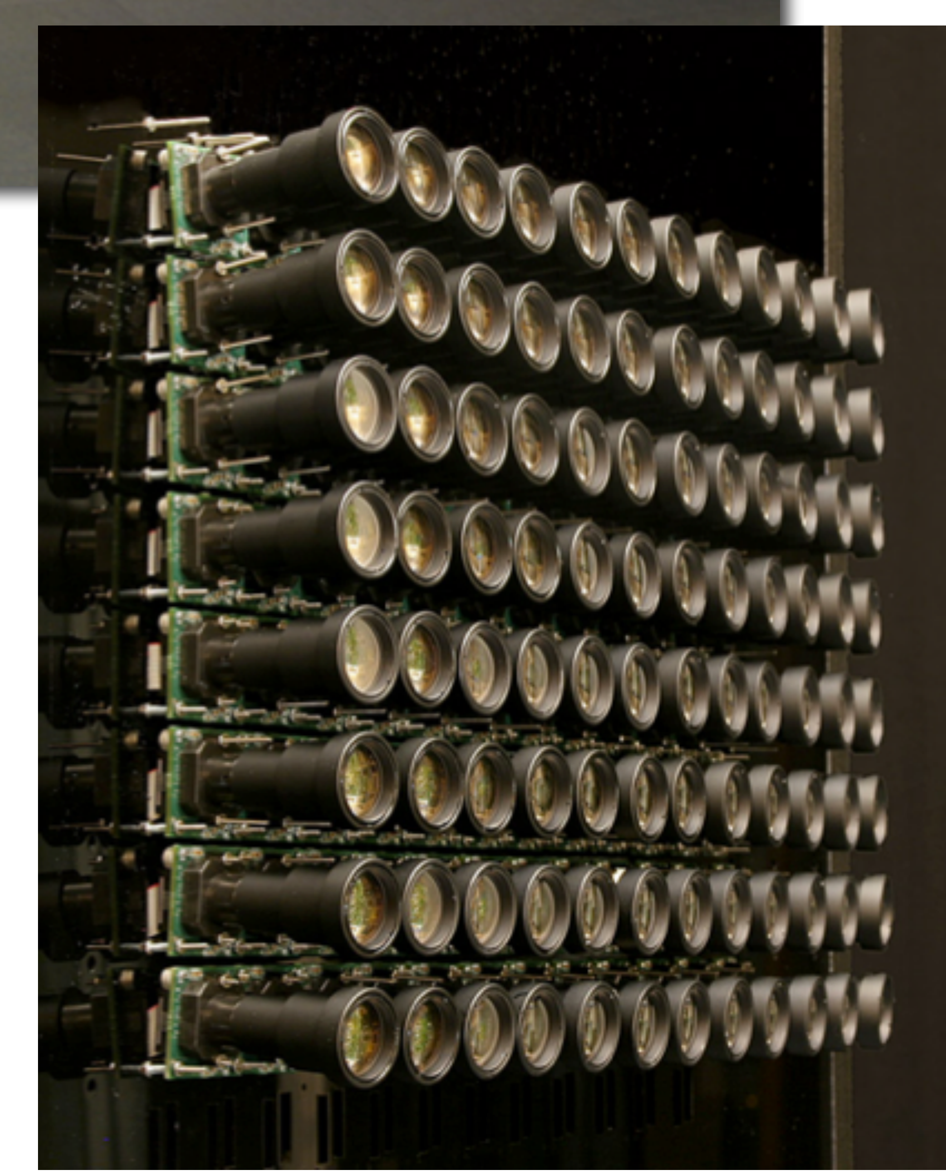
[Wilburn et al. 2005]

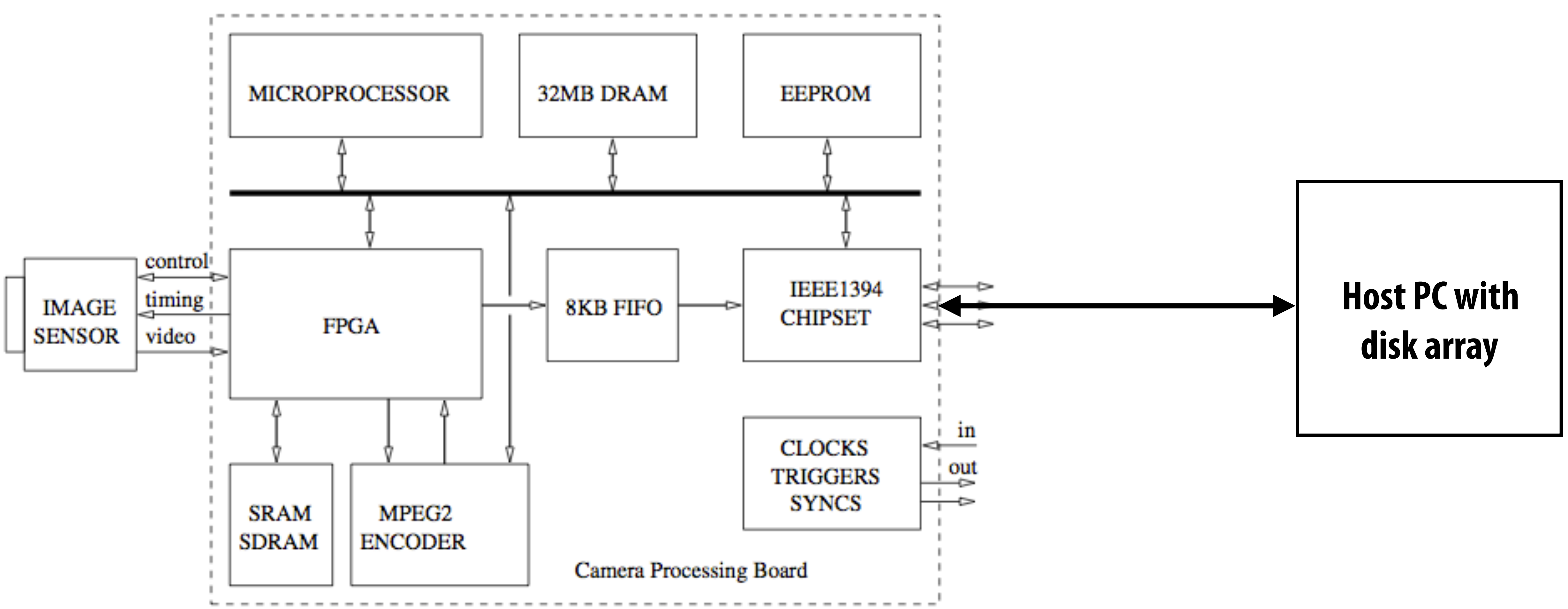
640 x 480 tightly synchronized, repositionable cameras

Custom processing board per camera

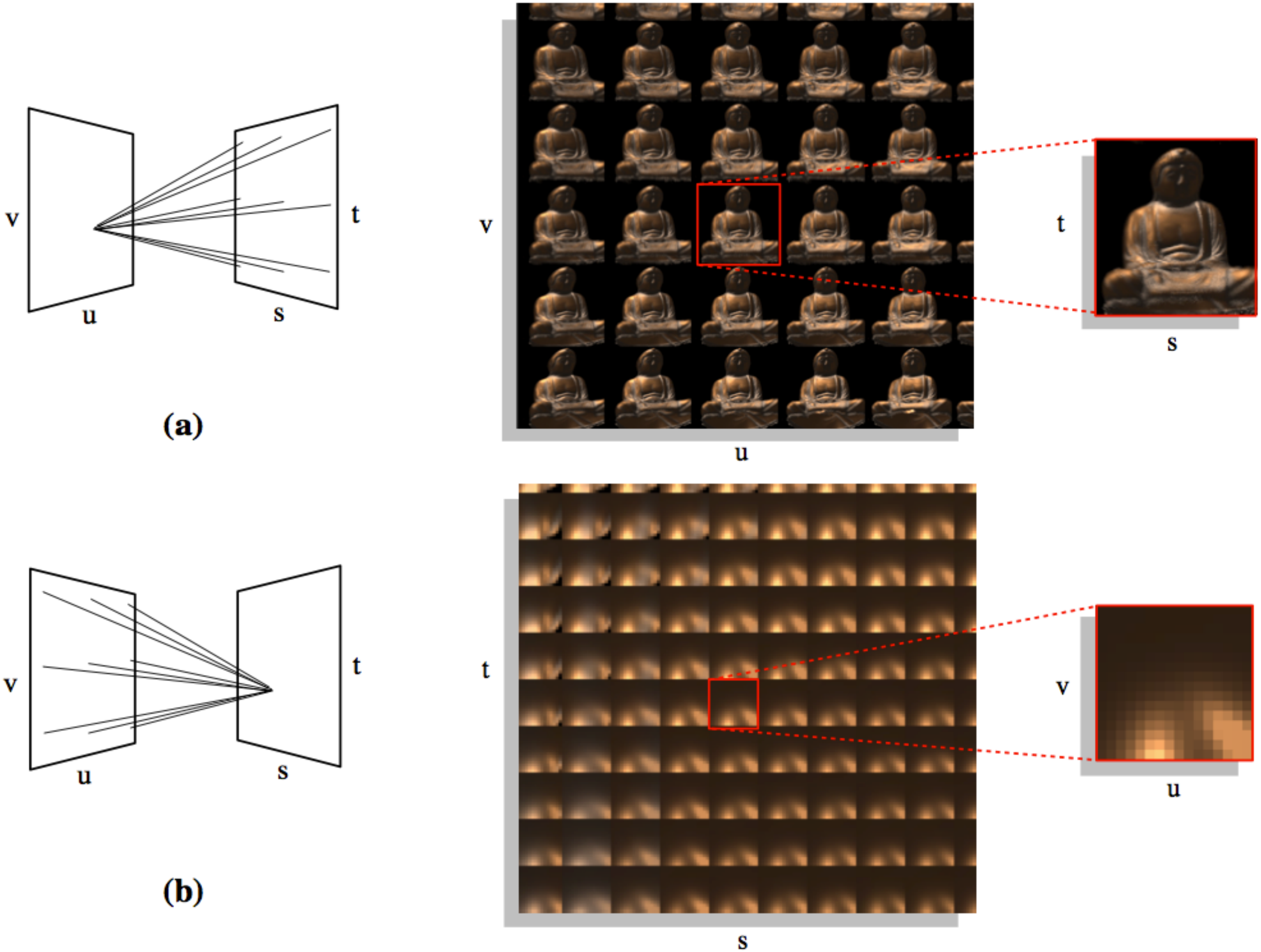
Tethered to PCs for additional processing/storage







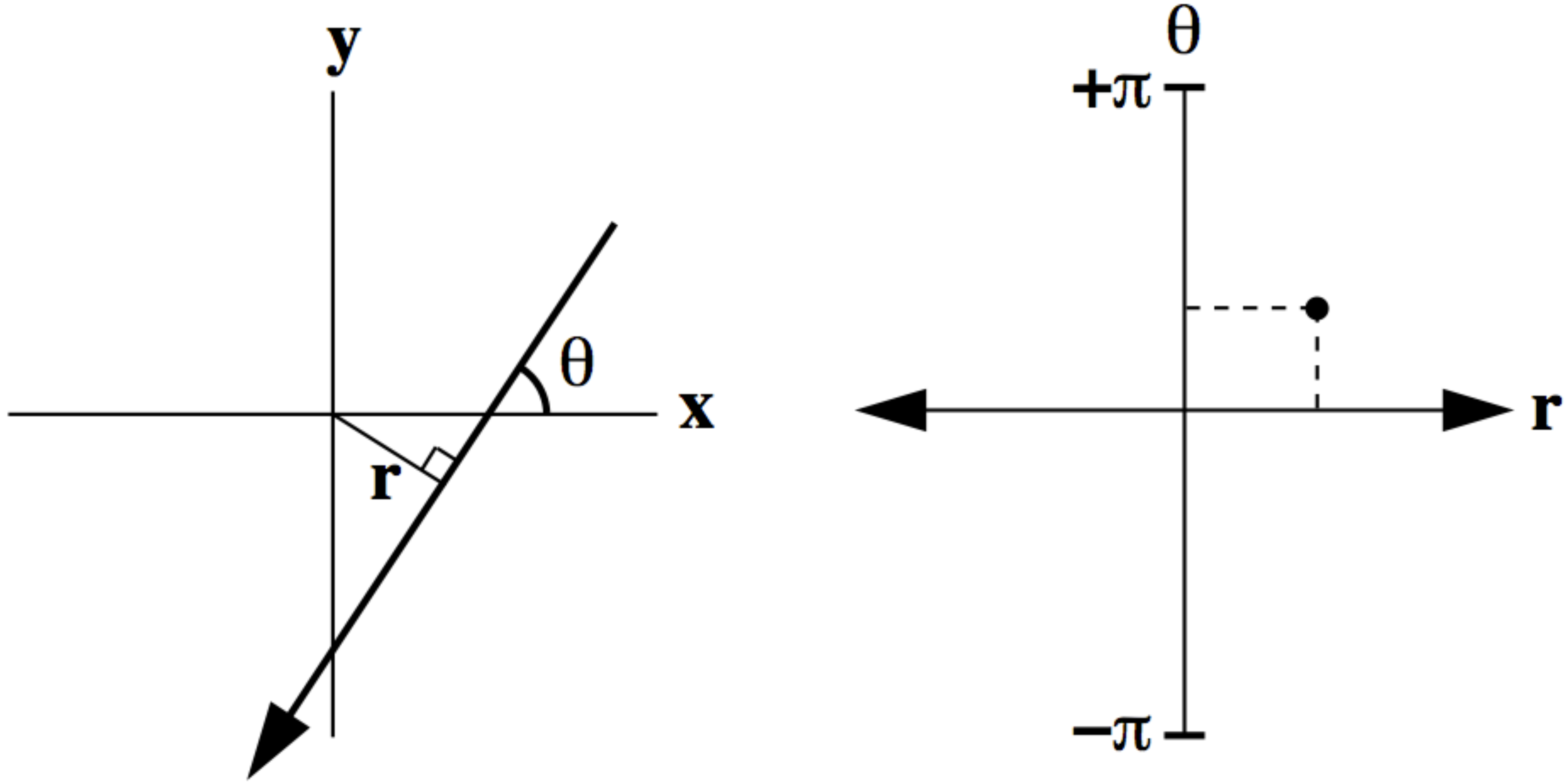
Light field storage layouts



[Image credit: Levoy and Hanrahan 96]

Line-space representation

Each line in Cartesian space* is represented by a point in line space



Cartesian space **Line space**

* Shown here in 2D, generalizes to 3D Cartesian lines

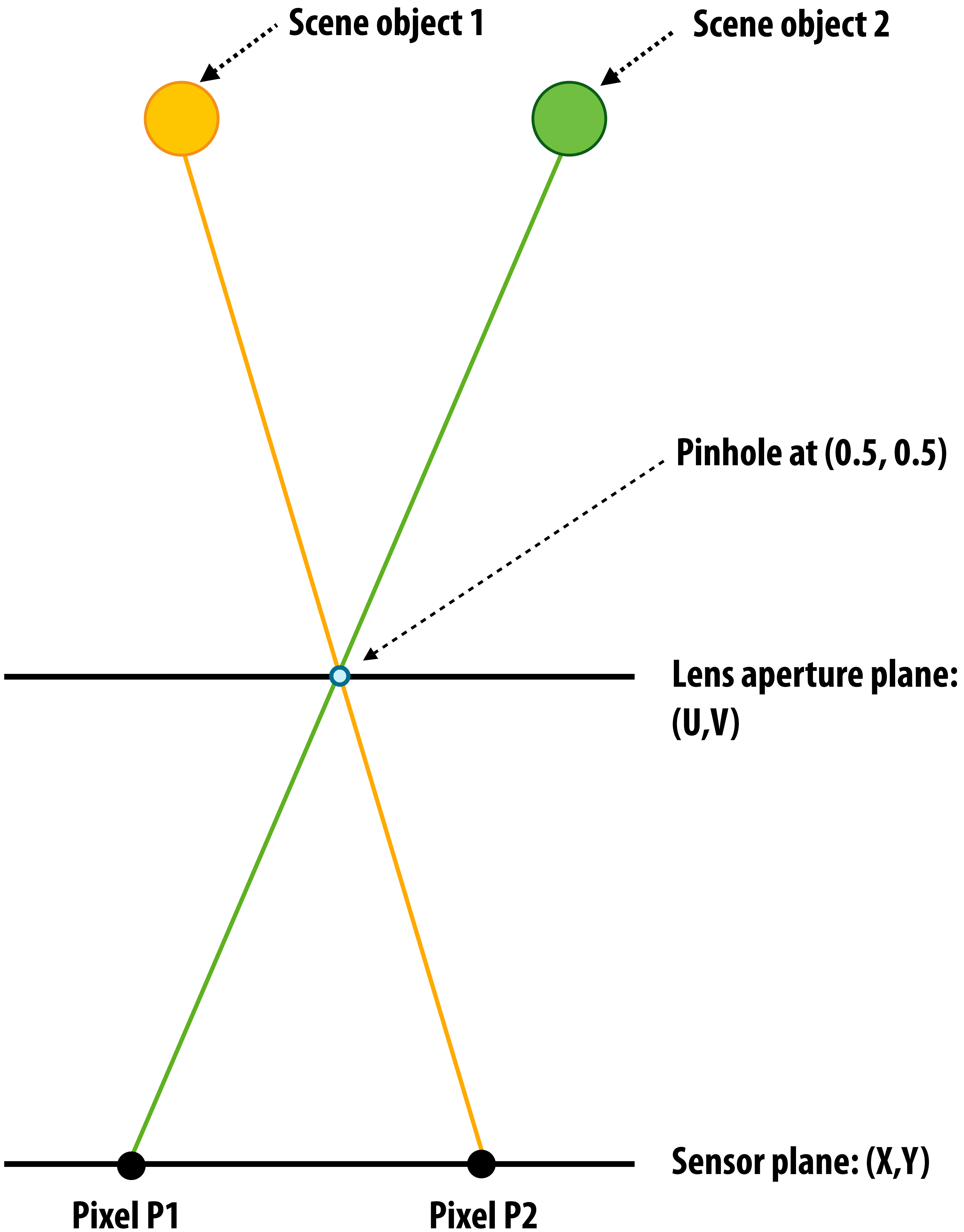
[Image credit: Levoy and Hanrahan 96]

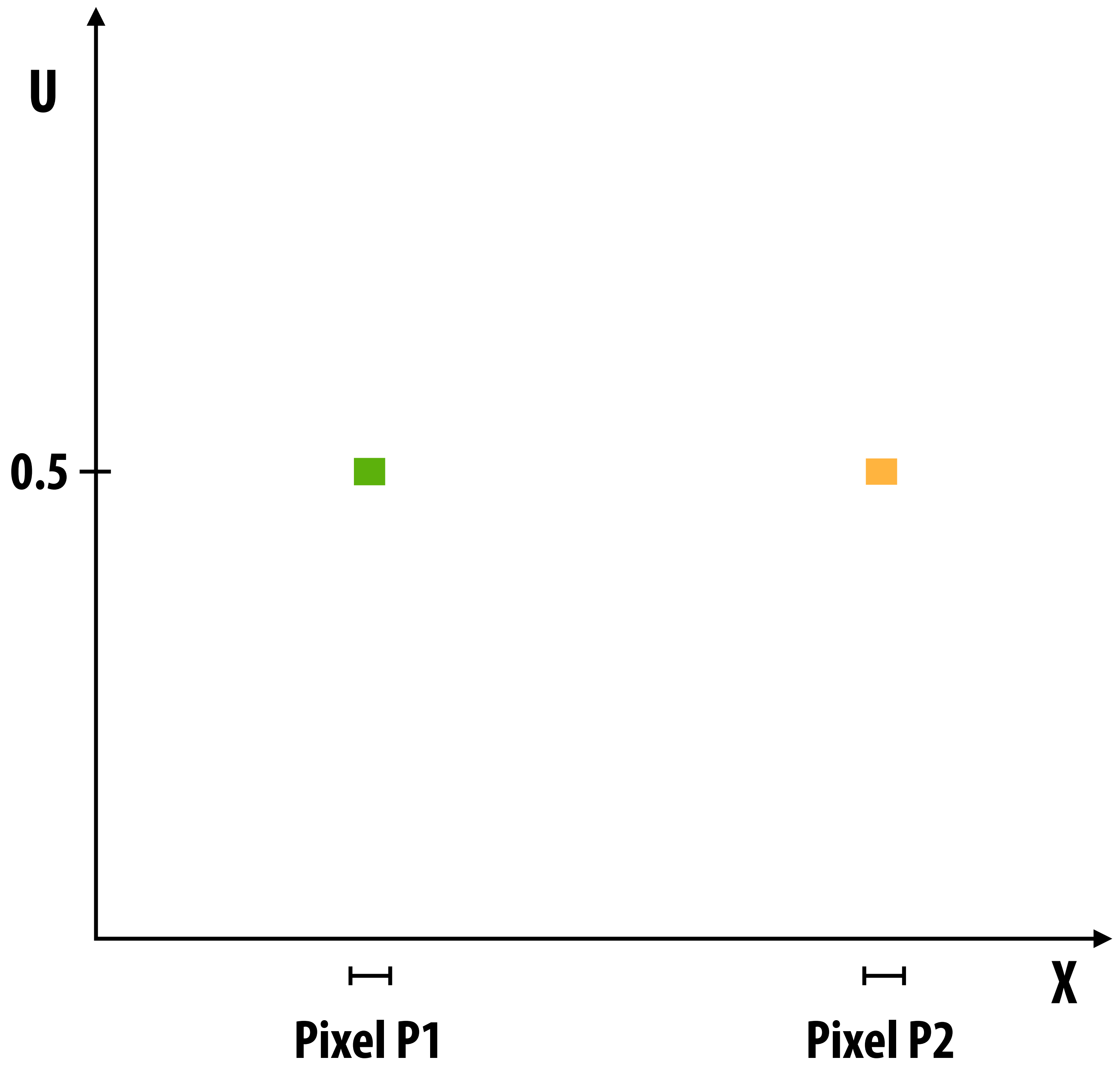
Pinhole camera

https://civilwar150pinholeproject.com/2013/04/13/pinhole-shutter/

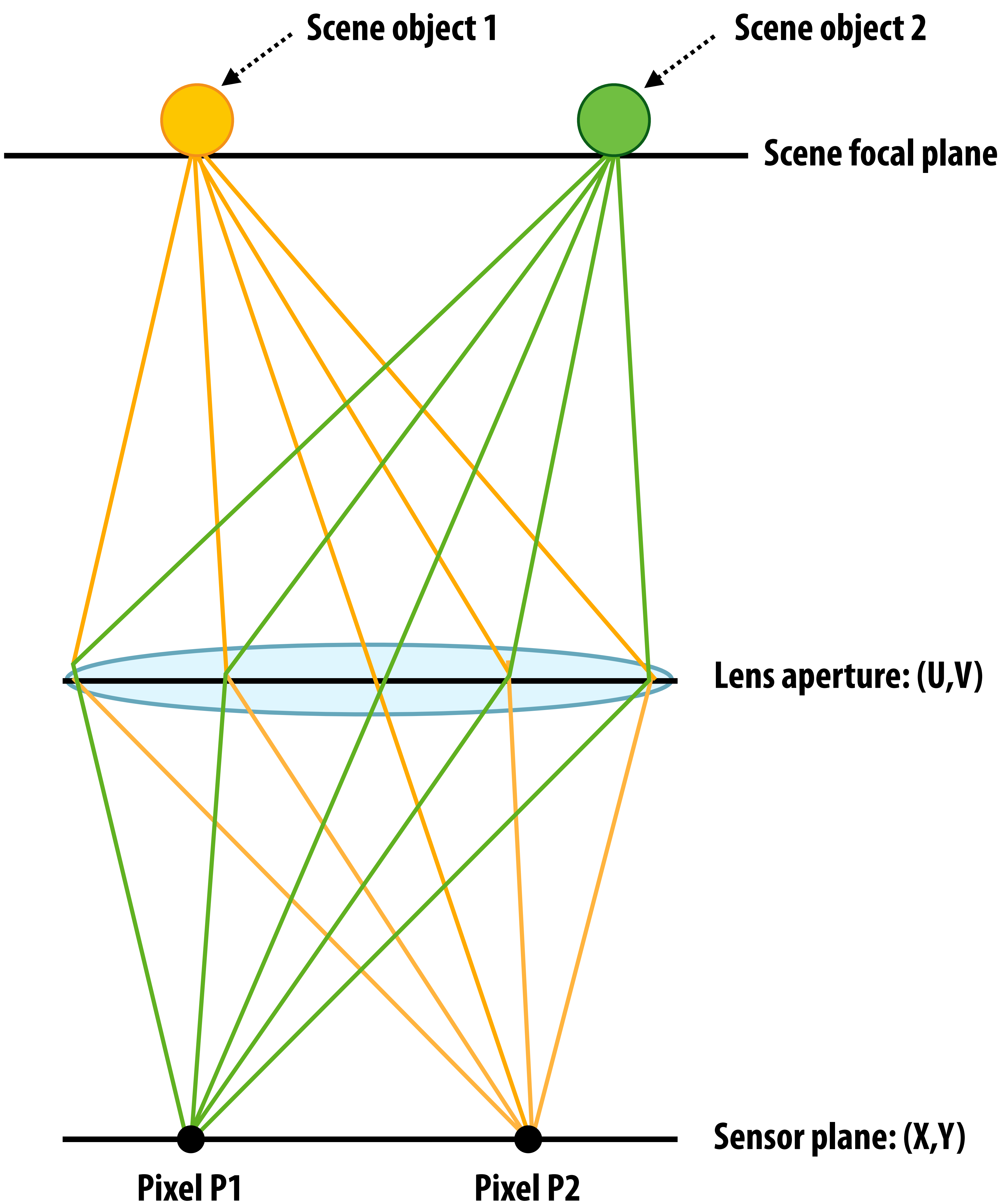
http://brianvds.blogspot.com/2012/08/a-simple-pinhole-camera.html

Light field inside a pinhole camera

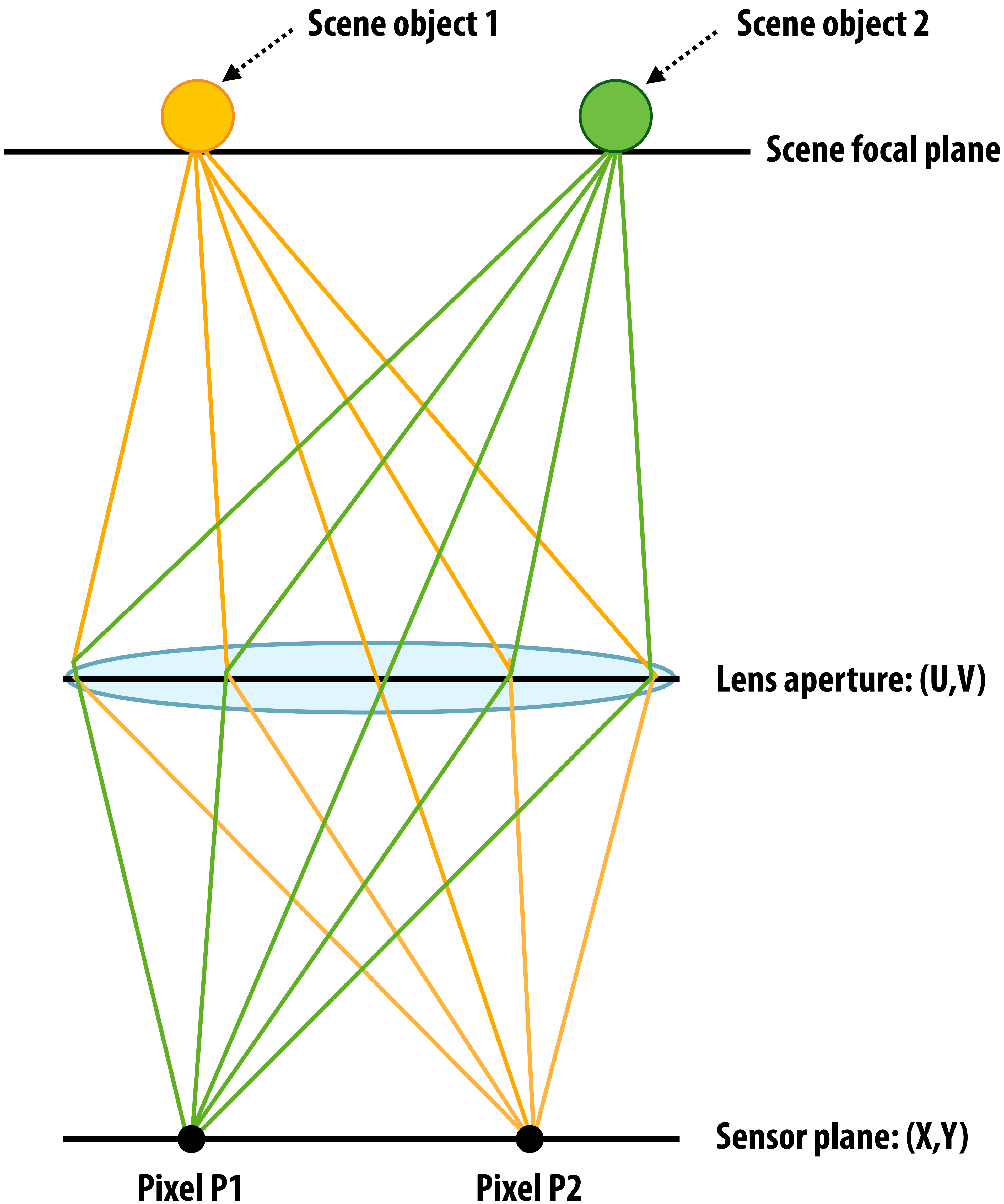




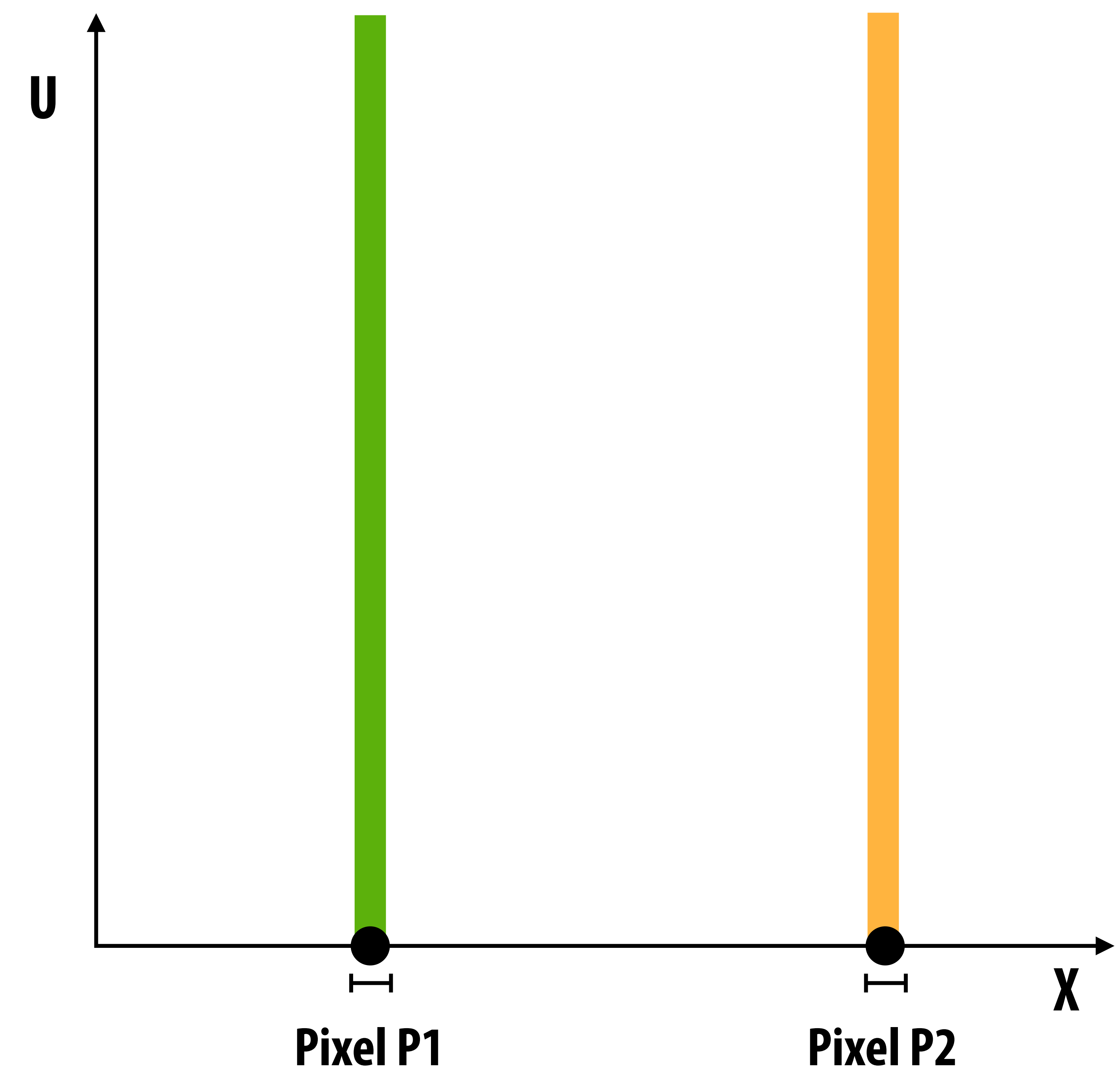
Camera with finite aperture



Light field inside a camera

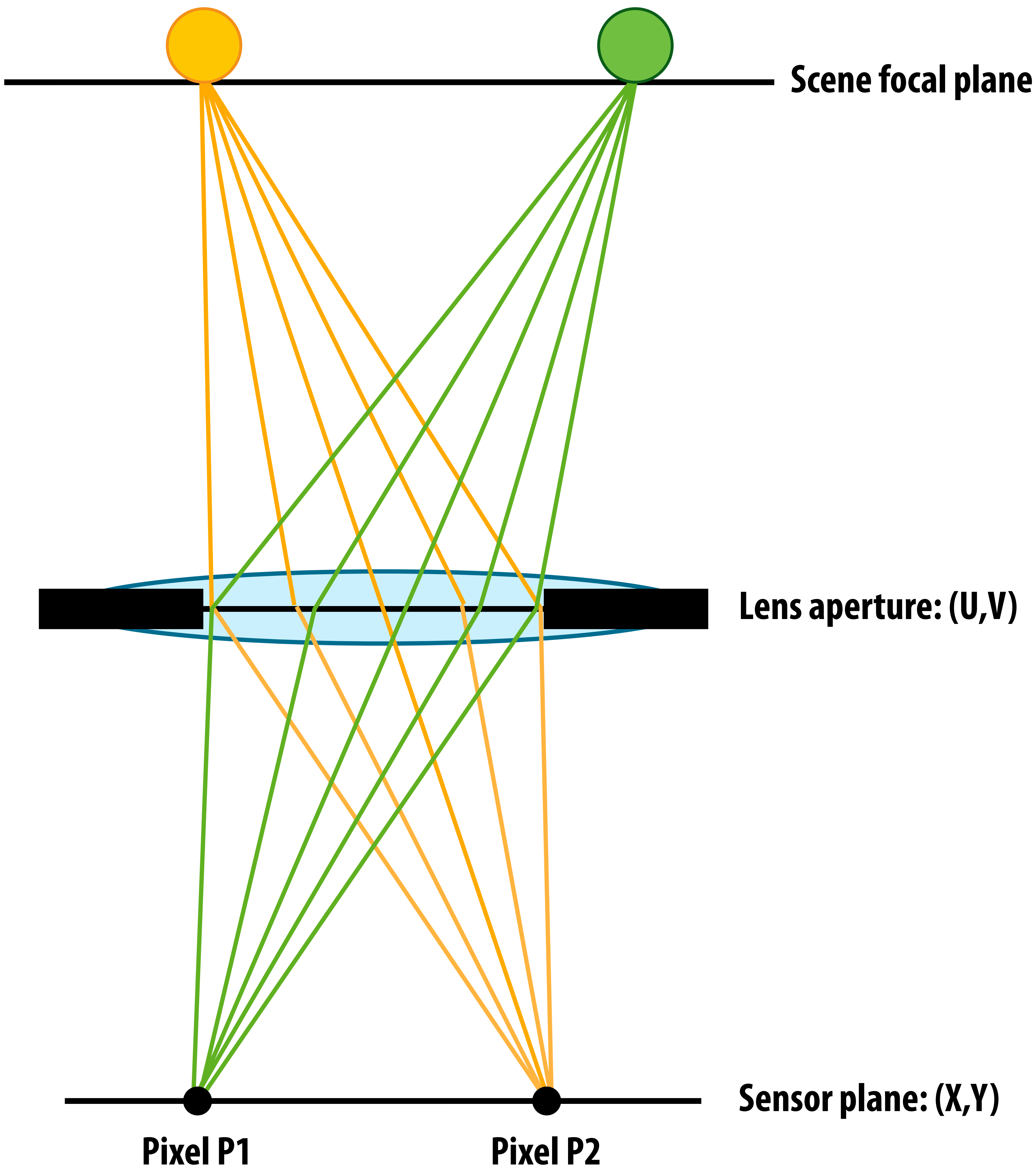


Ray space plot
(only showing 2D X-U projection)

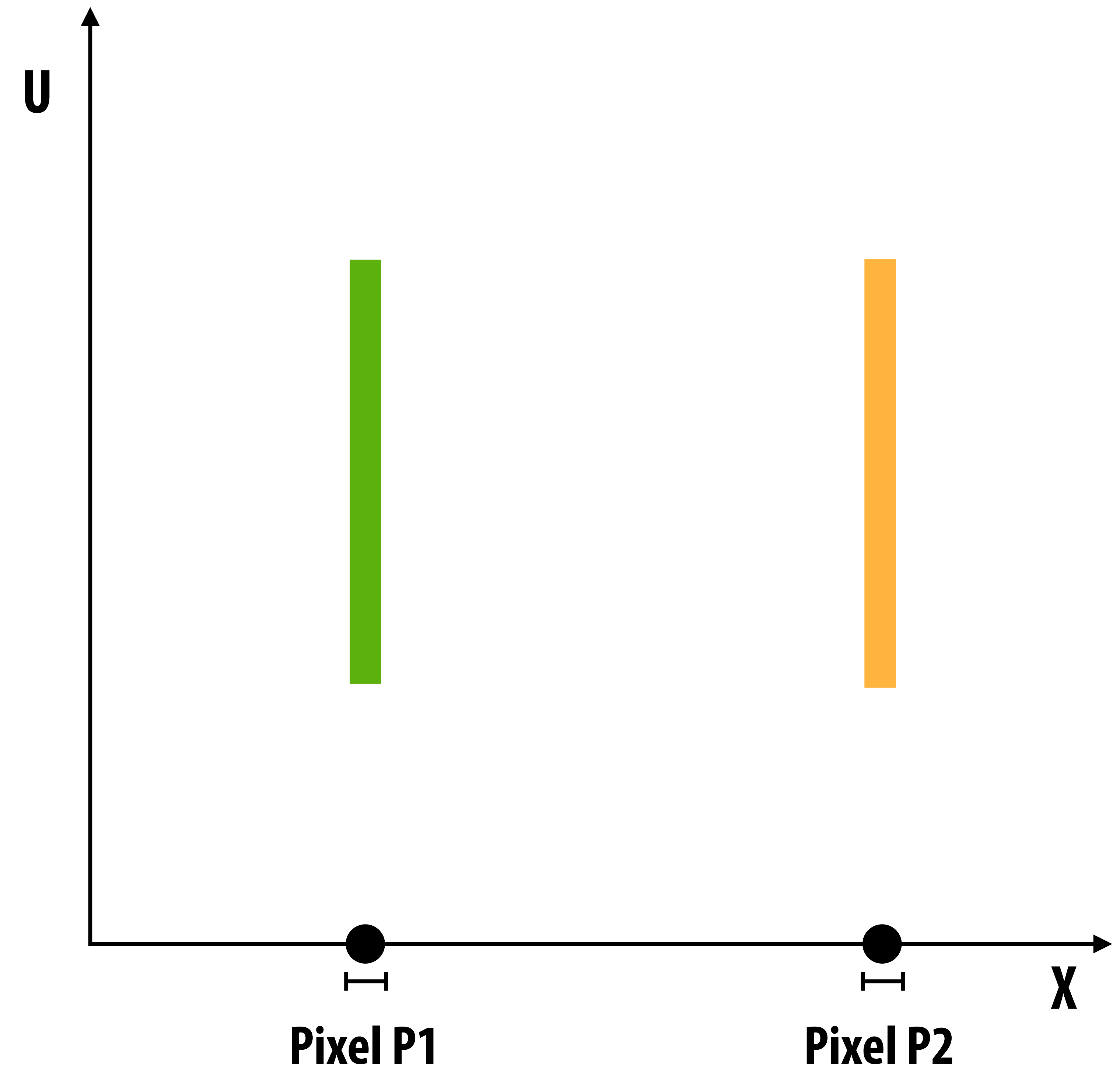


Sensor pixels measure integral of energy from all rays of light passing through points on the aperture and a pixel-sized area of the sensor.

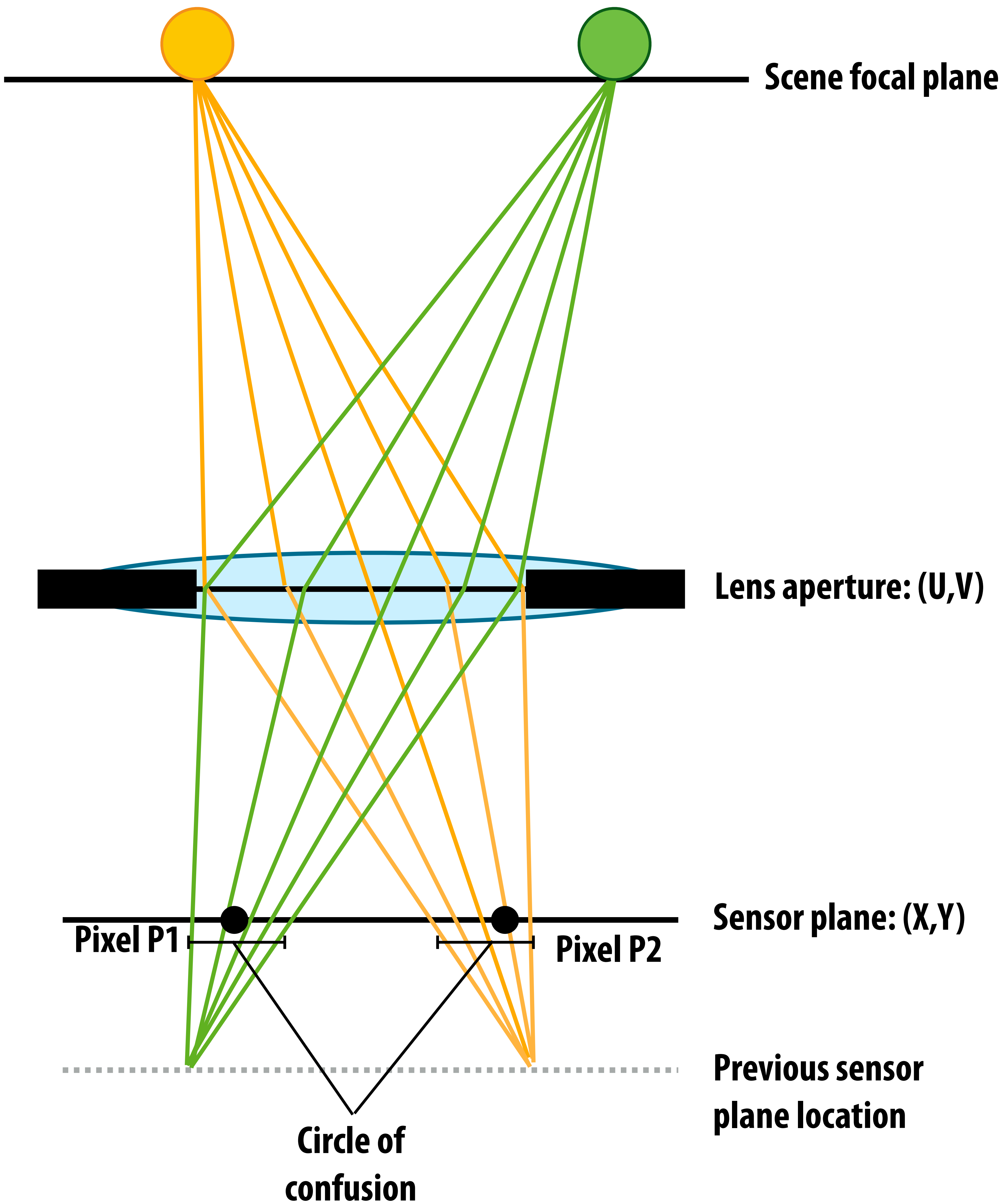
Decrease aperture size



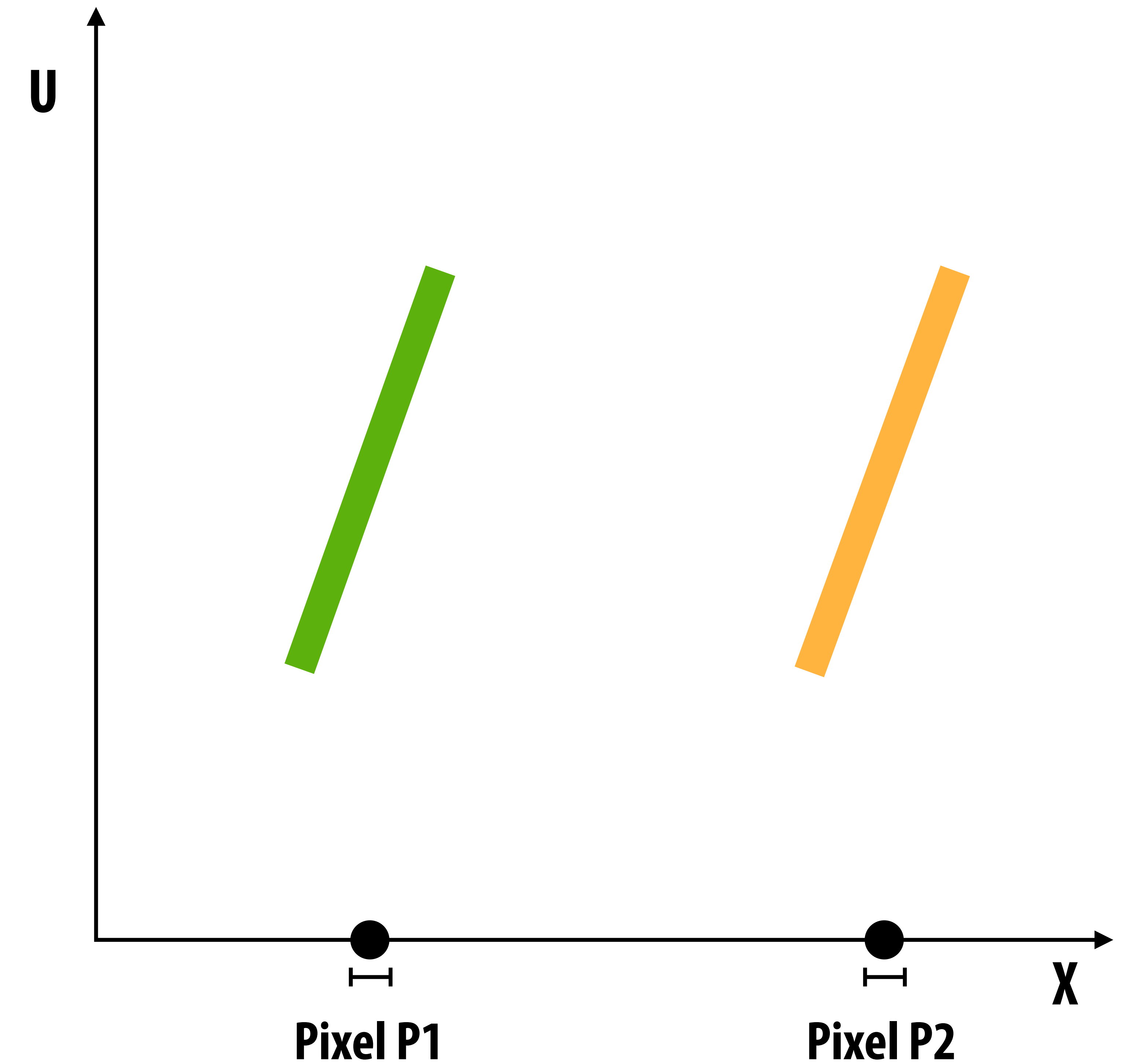
Ray space plot



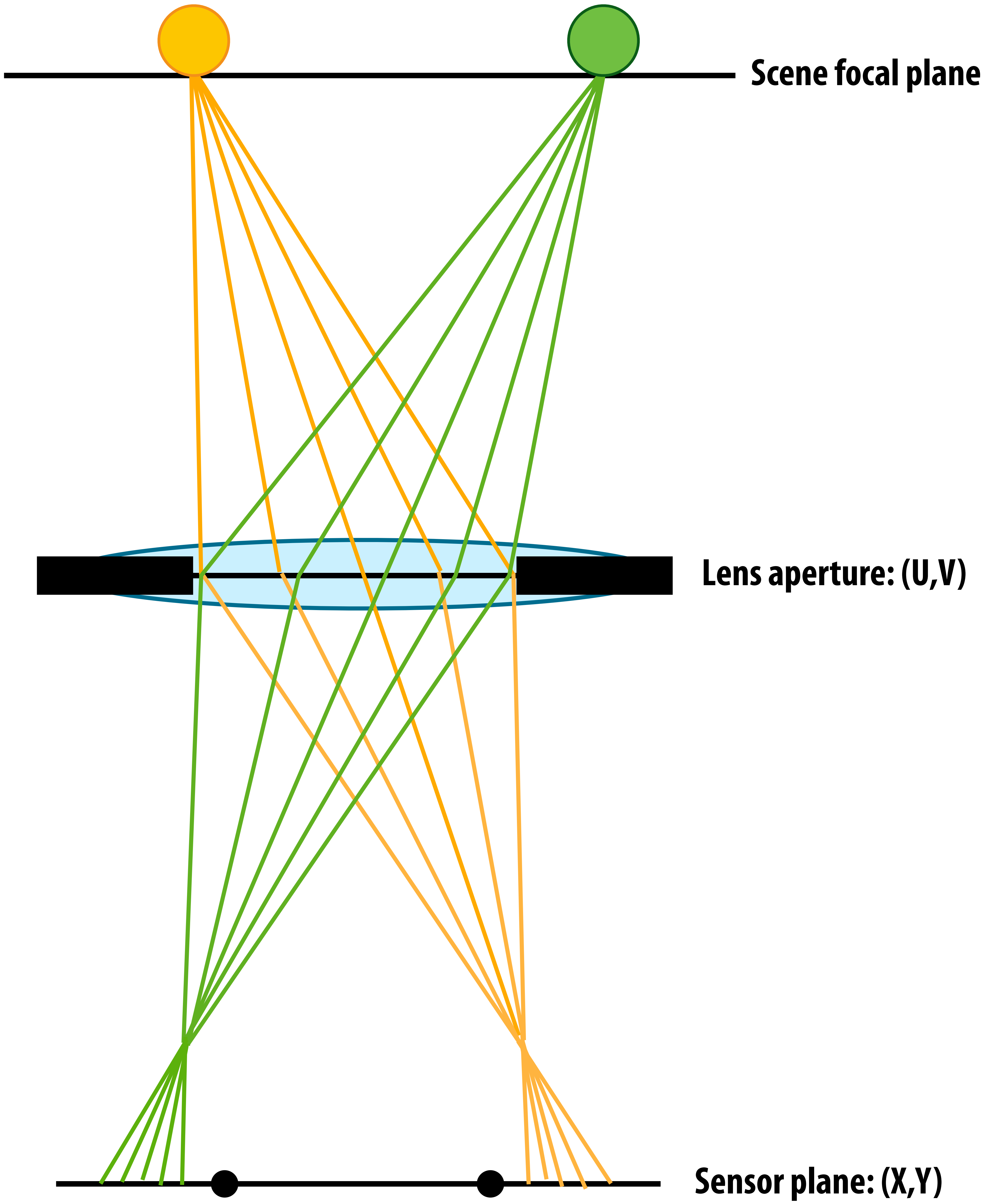
Defocus



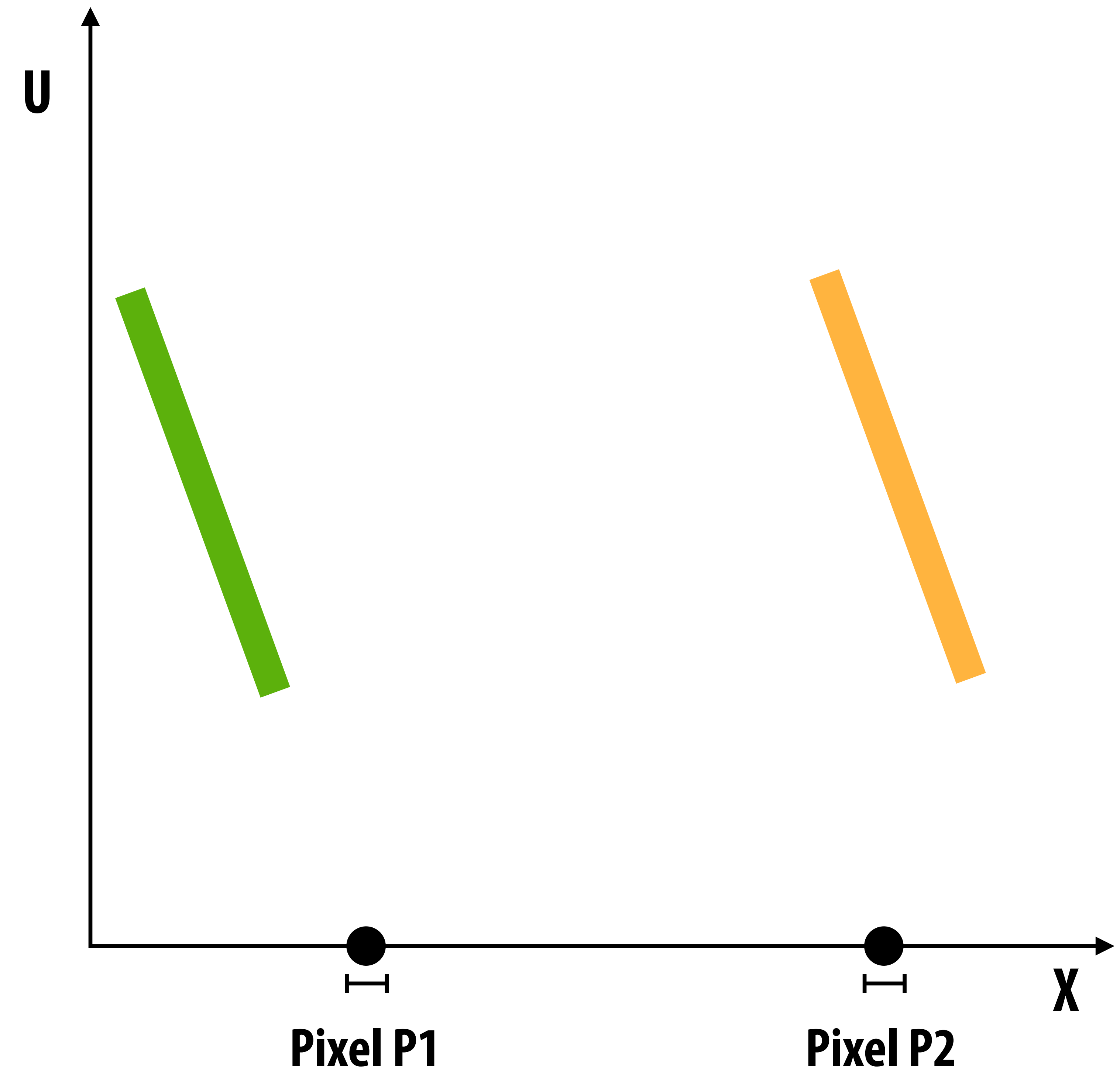
Ray space plot



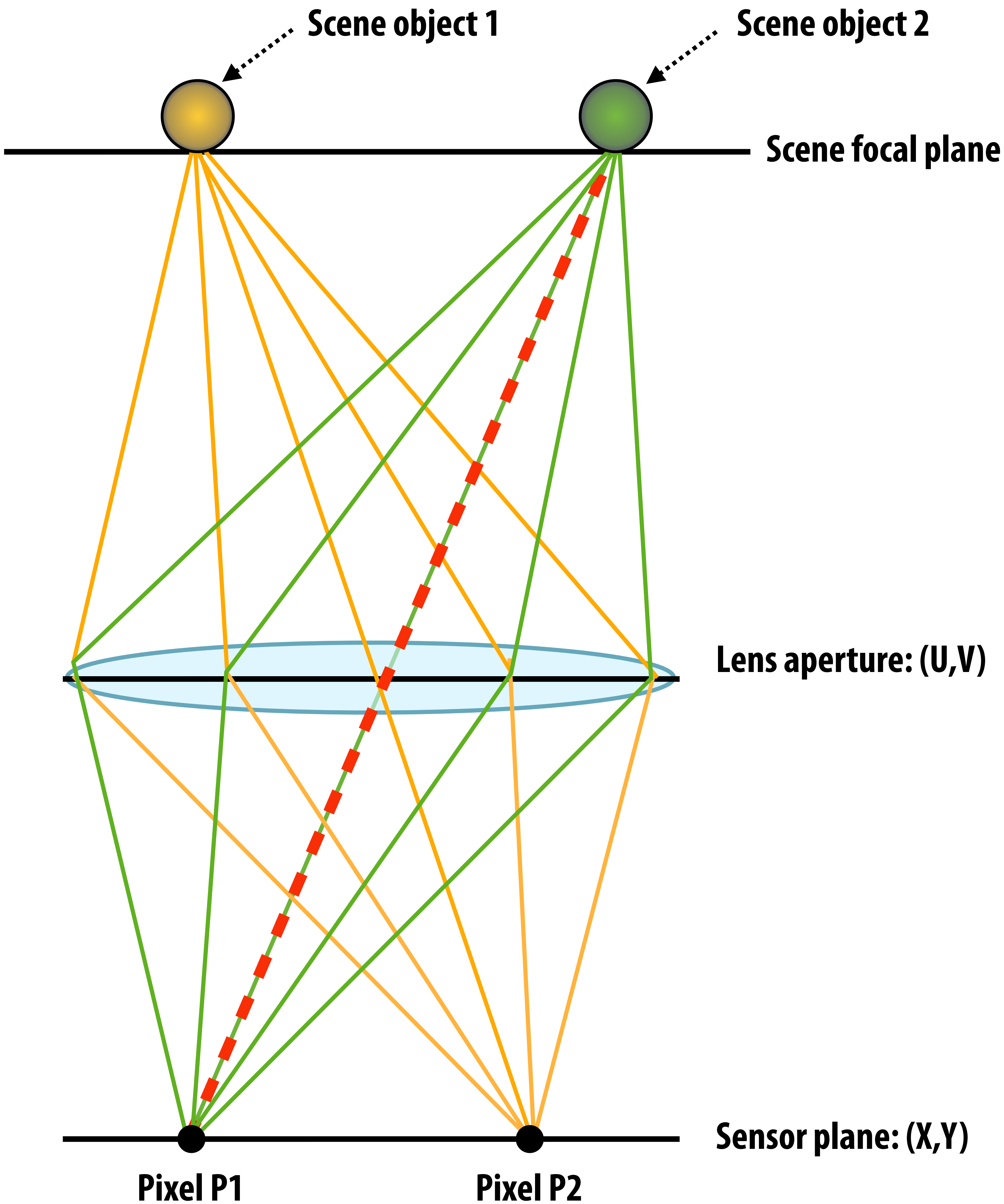
Defocus

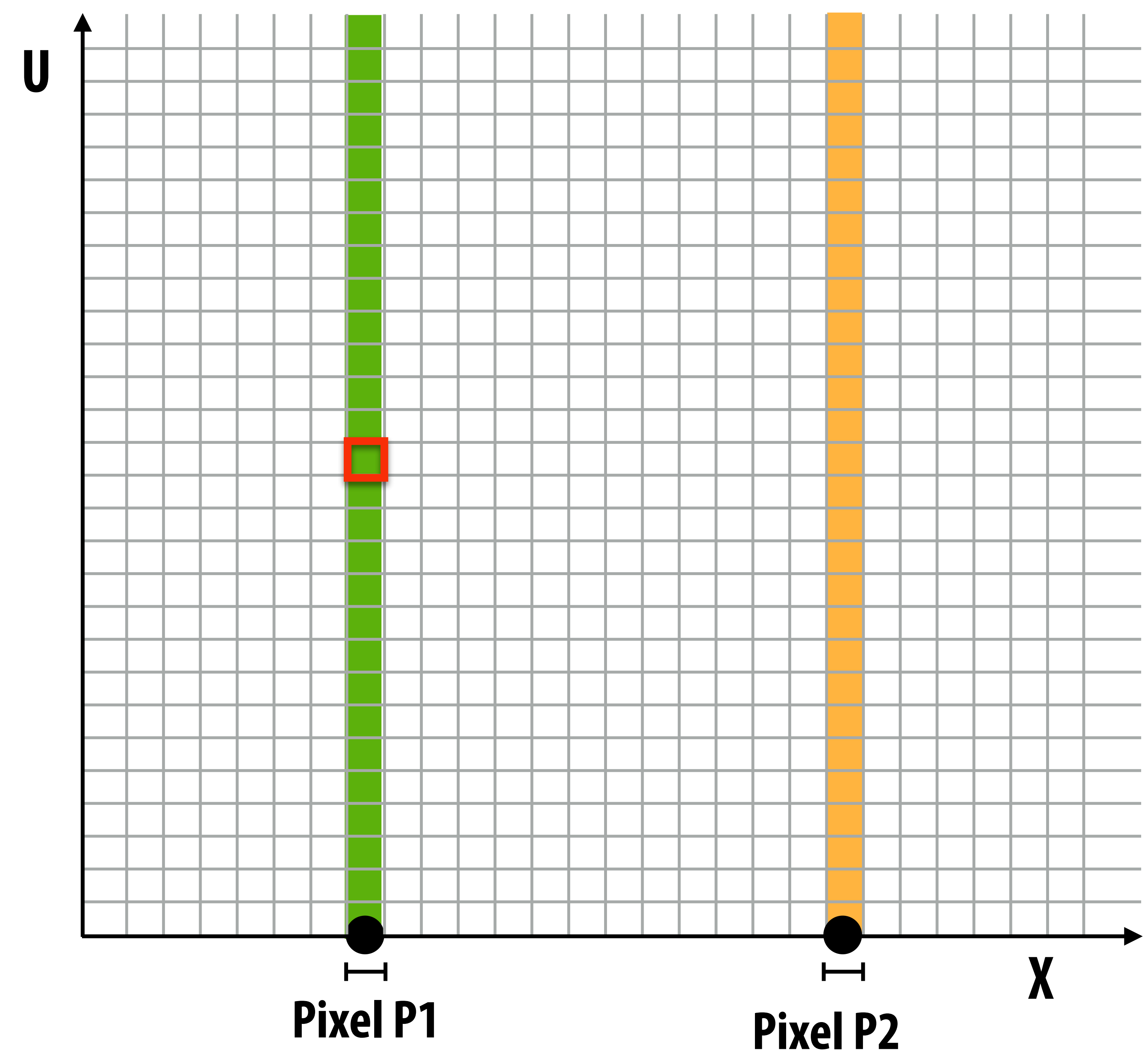


Ray space plot

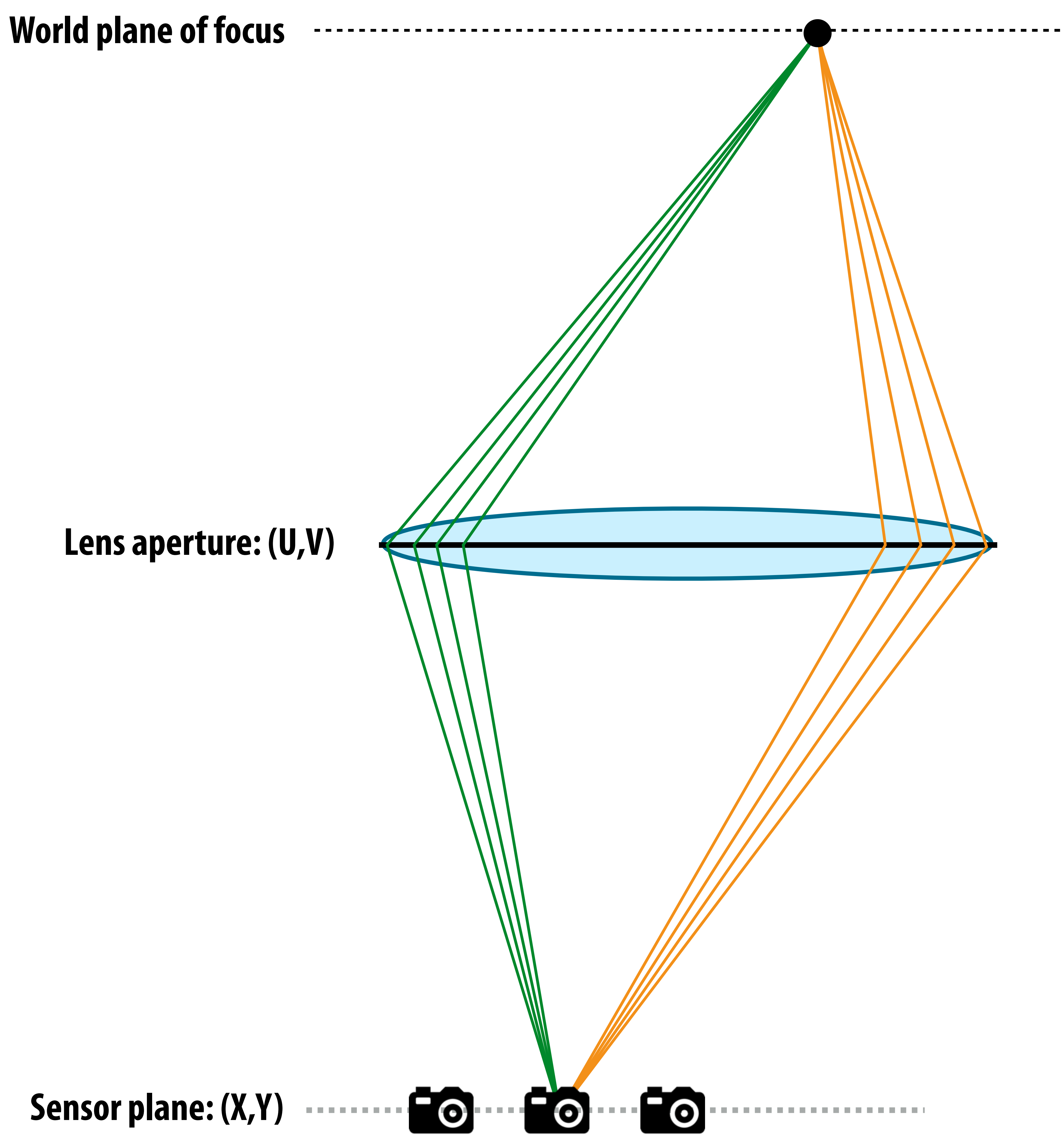


How might we measure the light field inside a camera?





Intuition: handheld light field camera

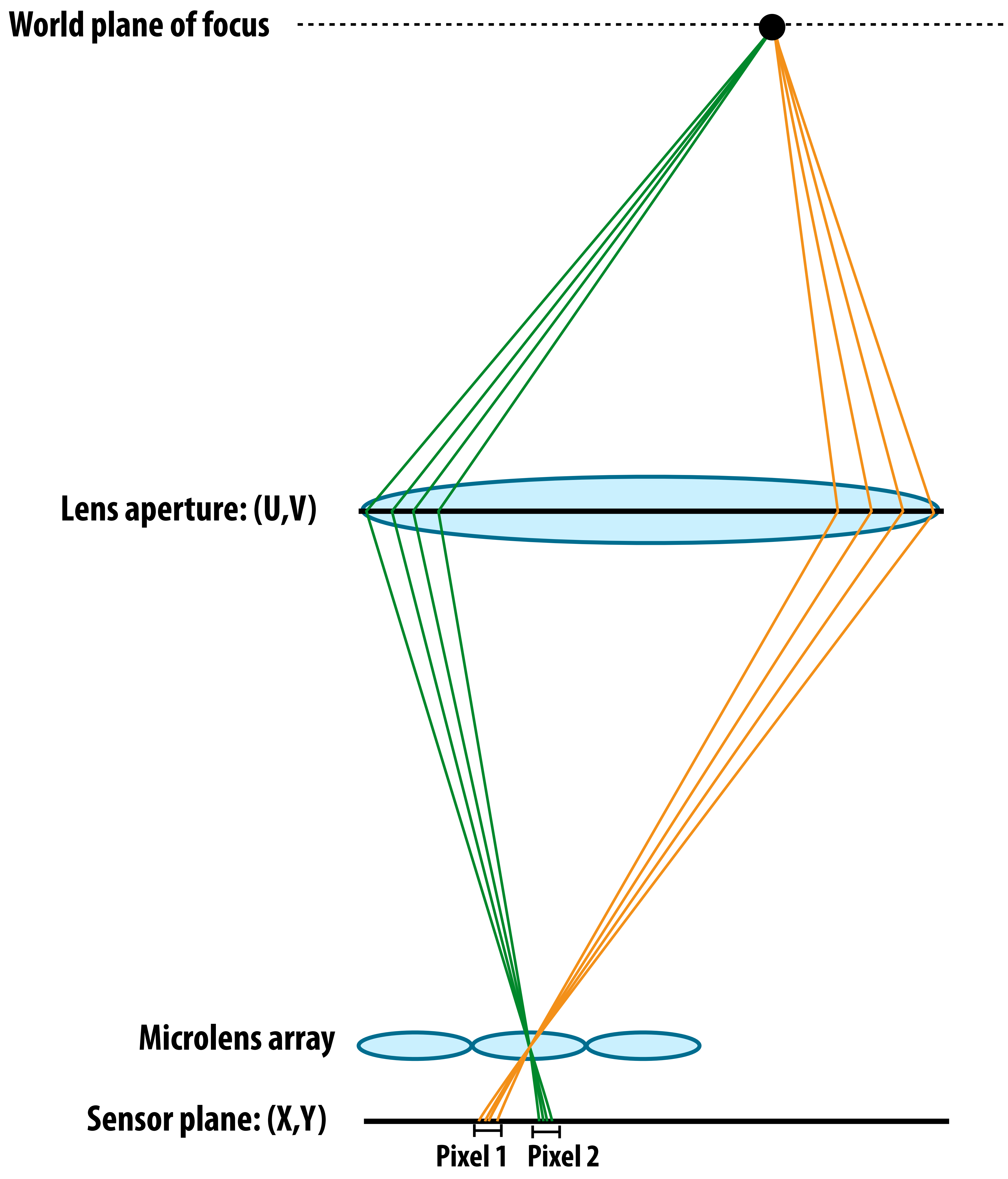


[Ng et al. 2005]

[Adelson and Wang, 1992]

Intuition: build an optical system where each region of the sensor "takes" a picture of the aperture of the main lens

Handheld light field camera

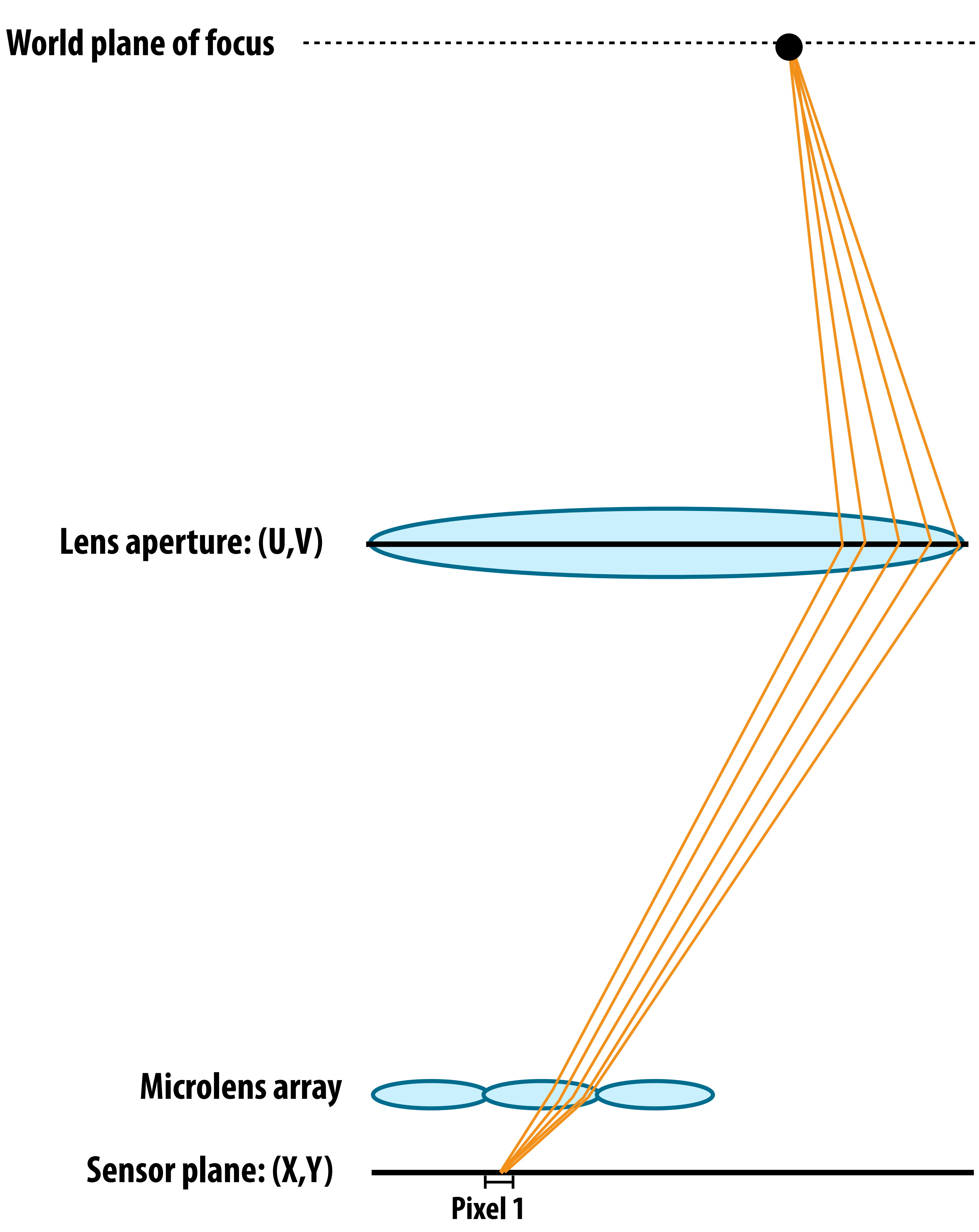


[Ng et al. 2005]

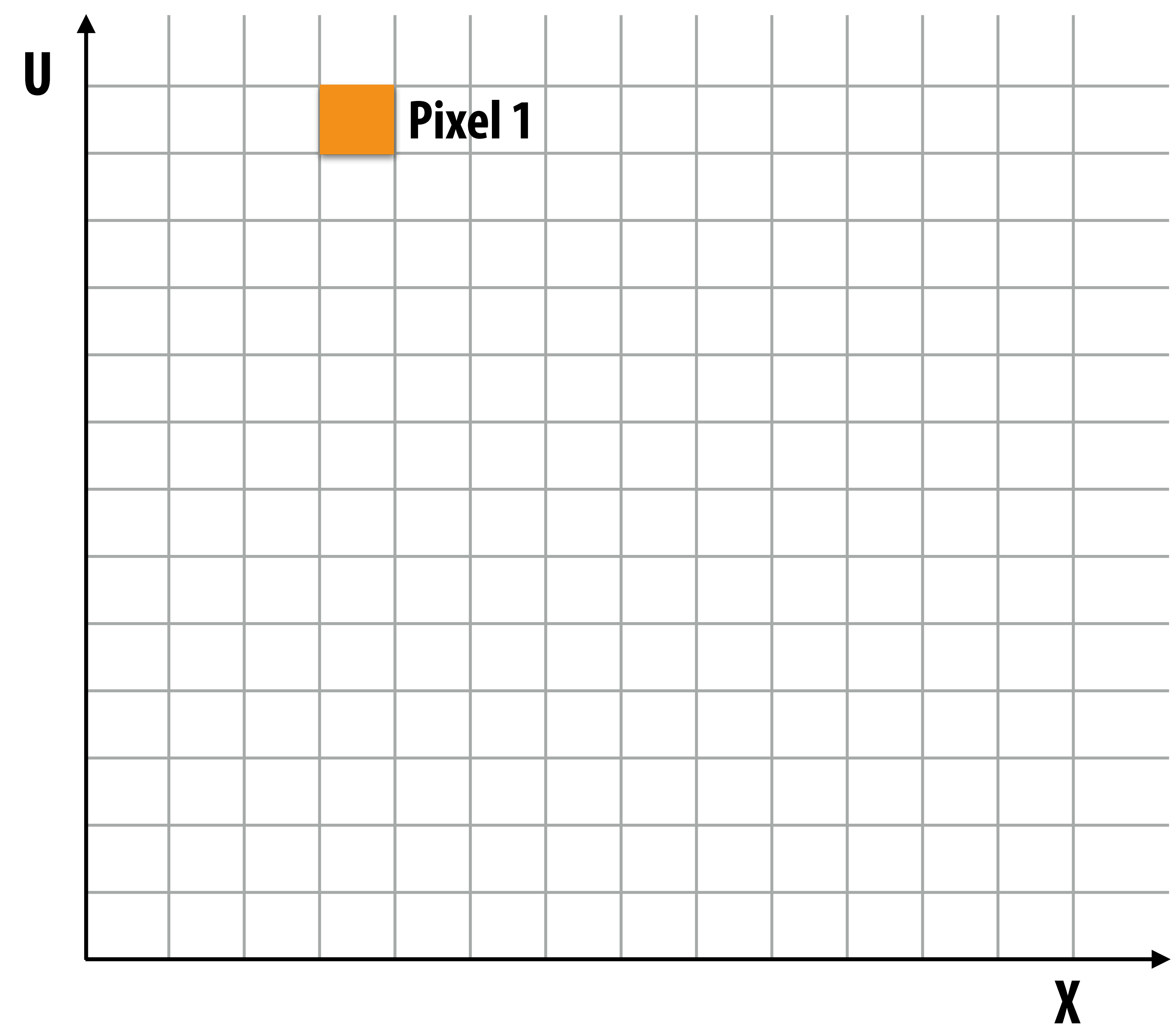
[Adelson and Wang, 1992]

Implementation: microlens array placed just on top of the sensor.

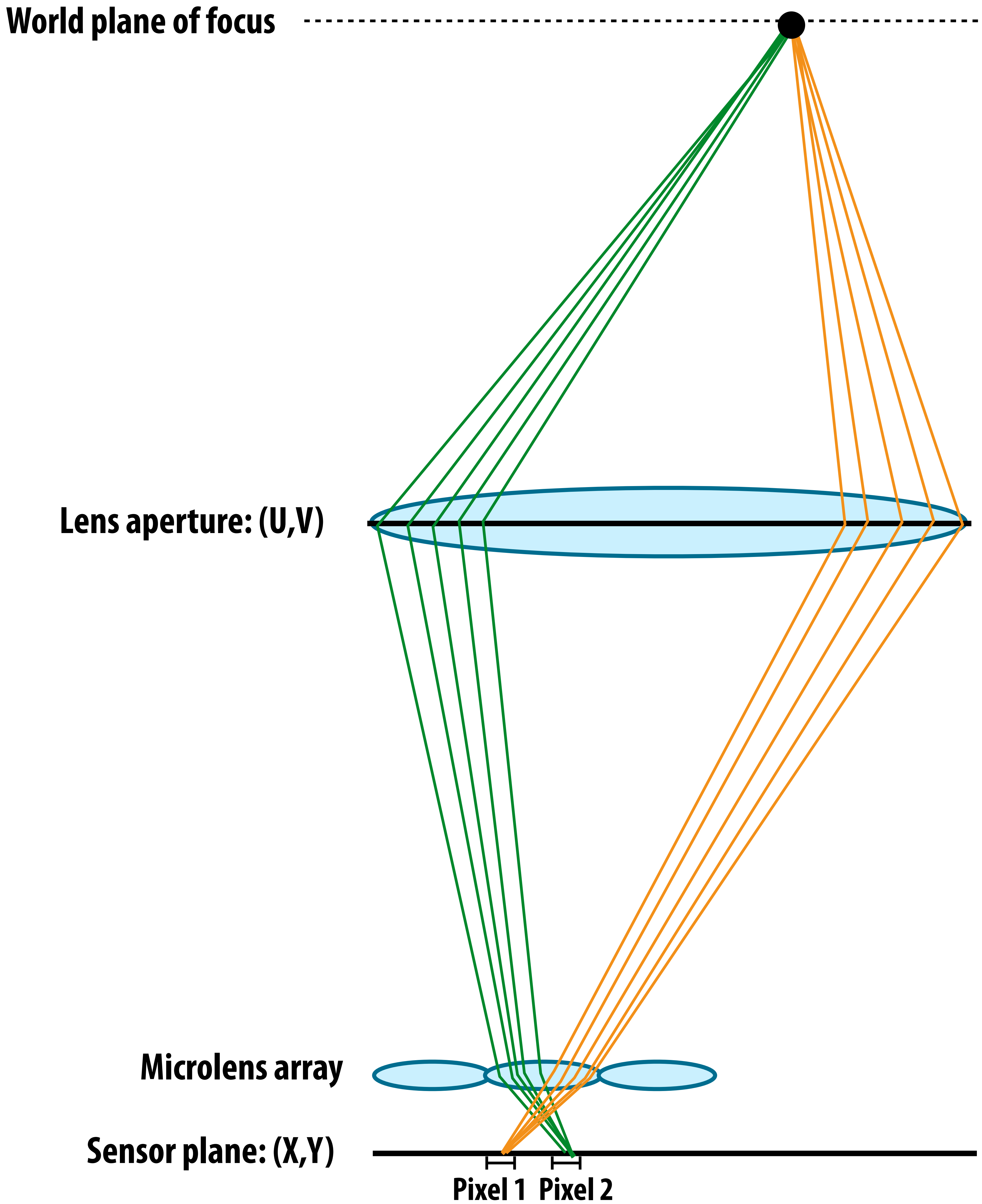
Each sensor pixel records a small beam of light inside the camera

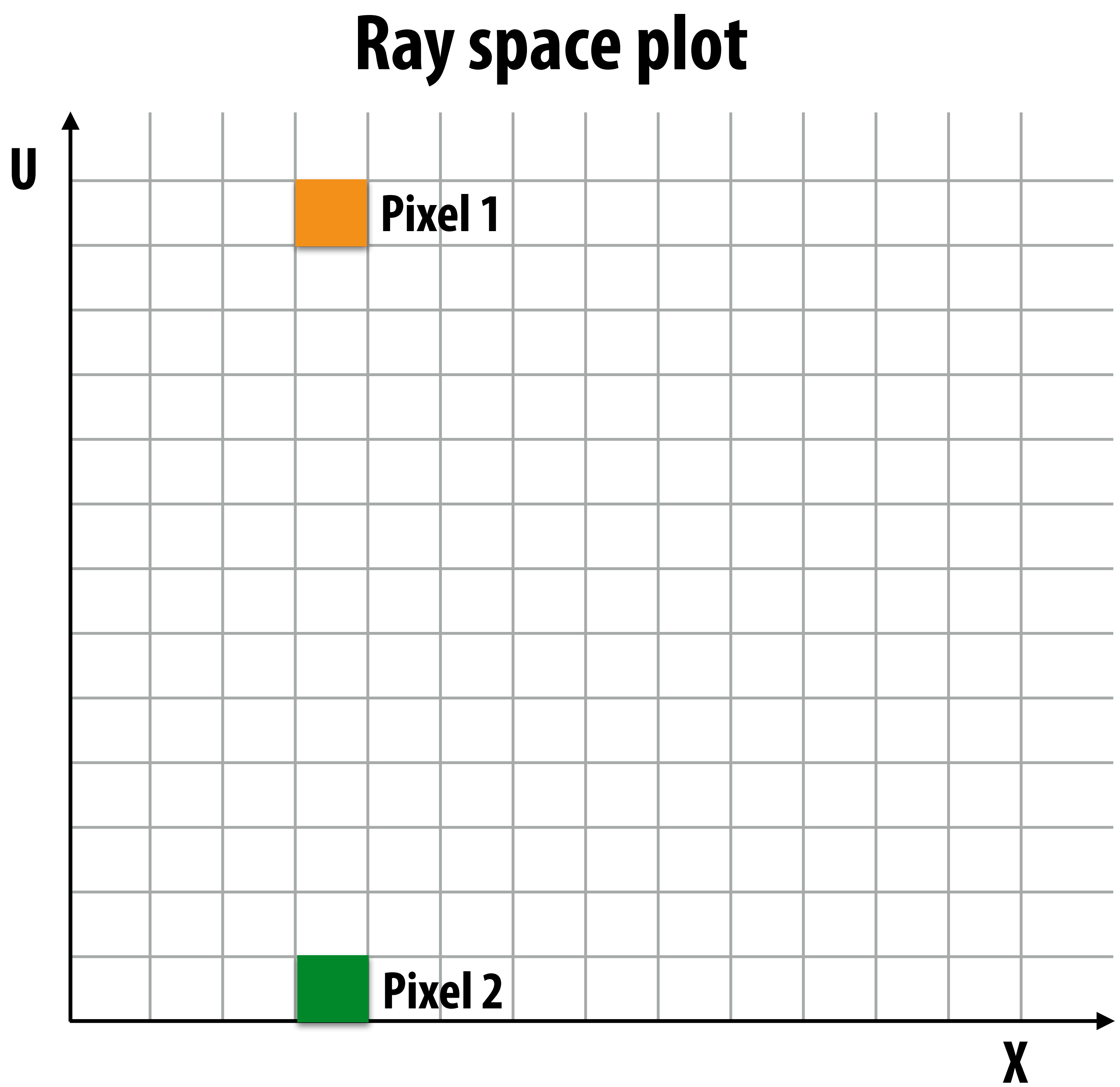


Ray space plot

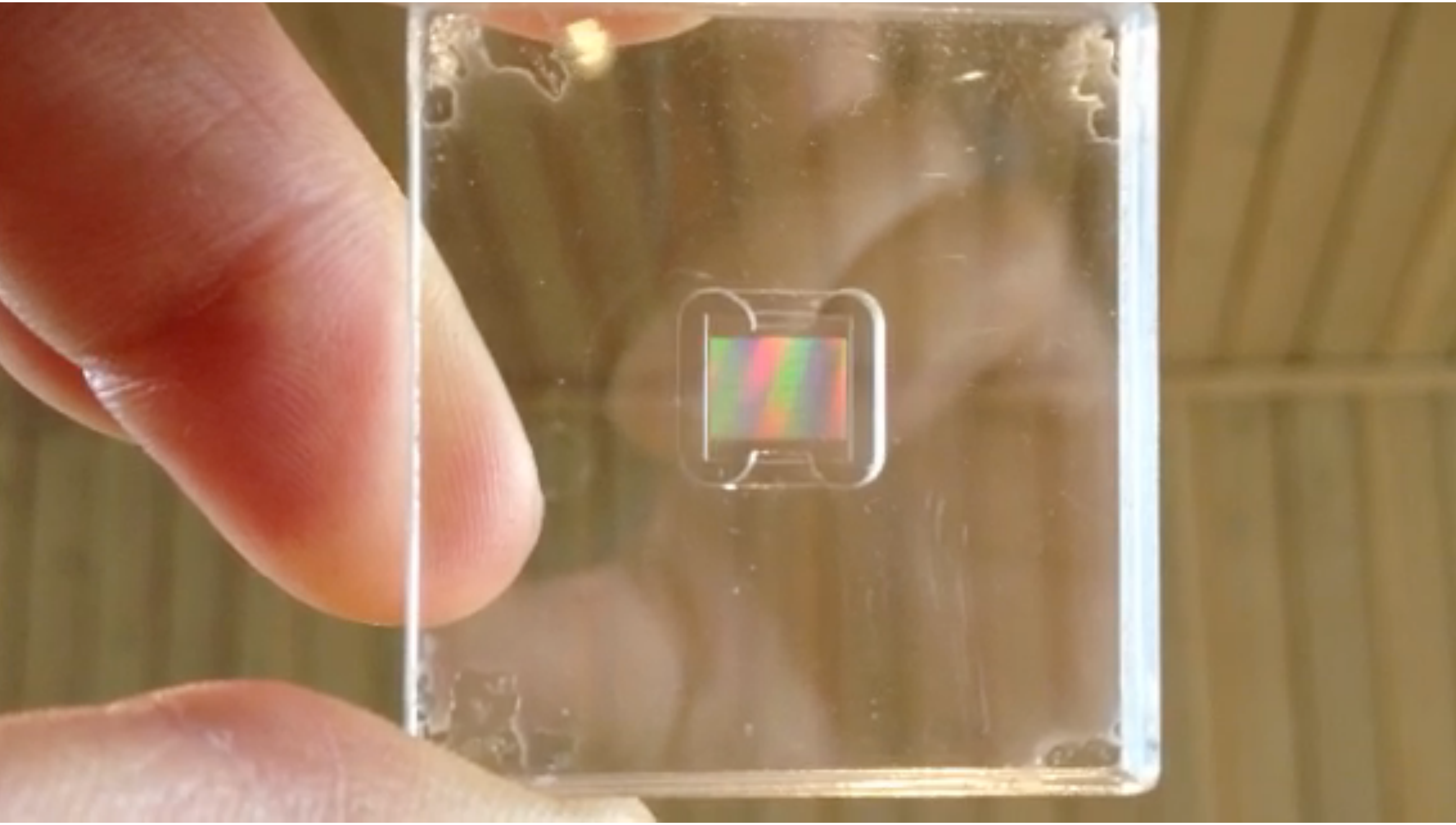


Each sensor pixel records a small beam of light inside the camera





Microlens array

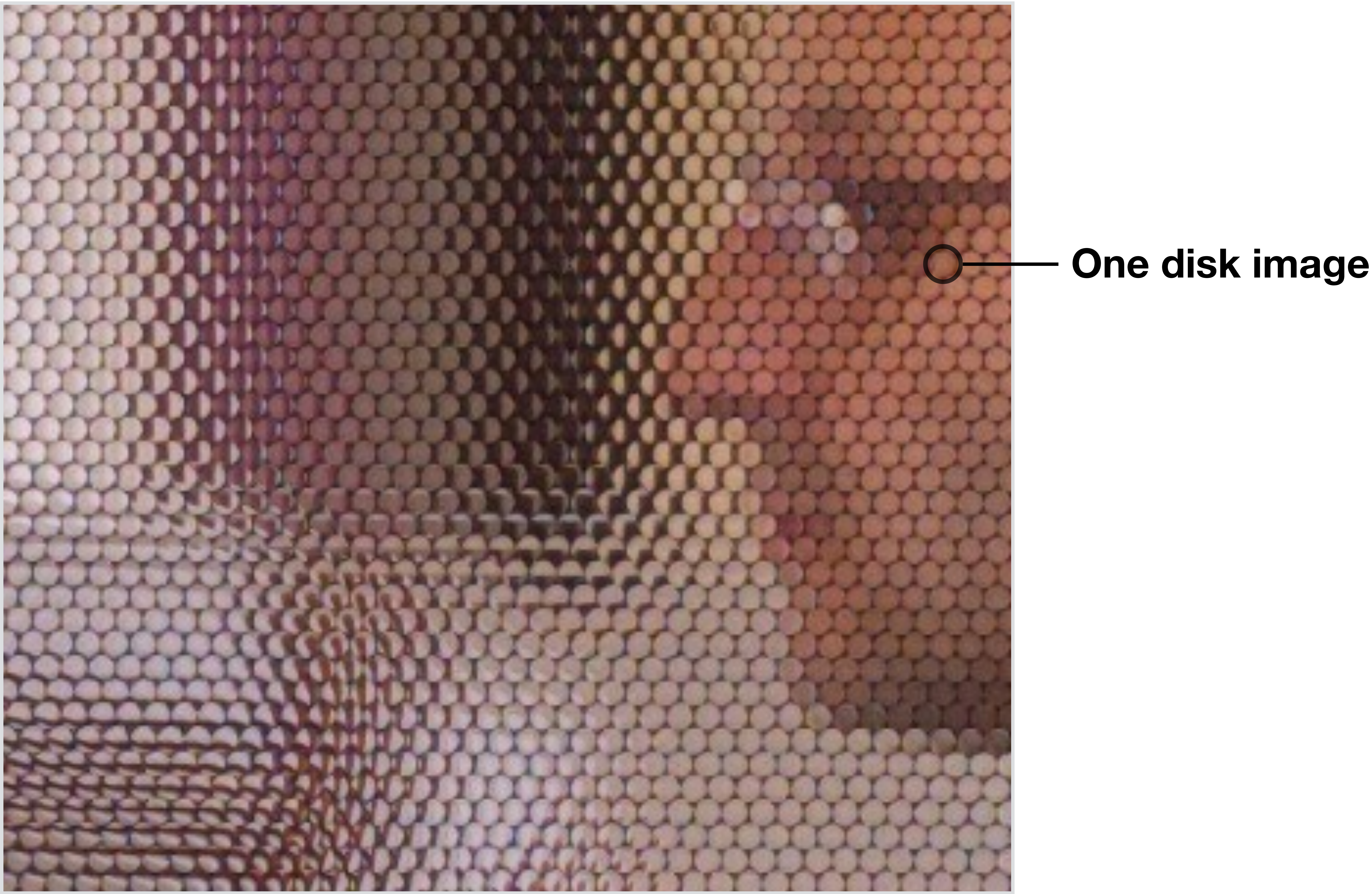


Slide credit: Ren Ng

Raw data from light field sensor

Slide credit: Ren Ng

Raw data from light field sensor



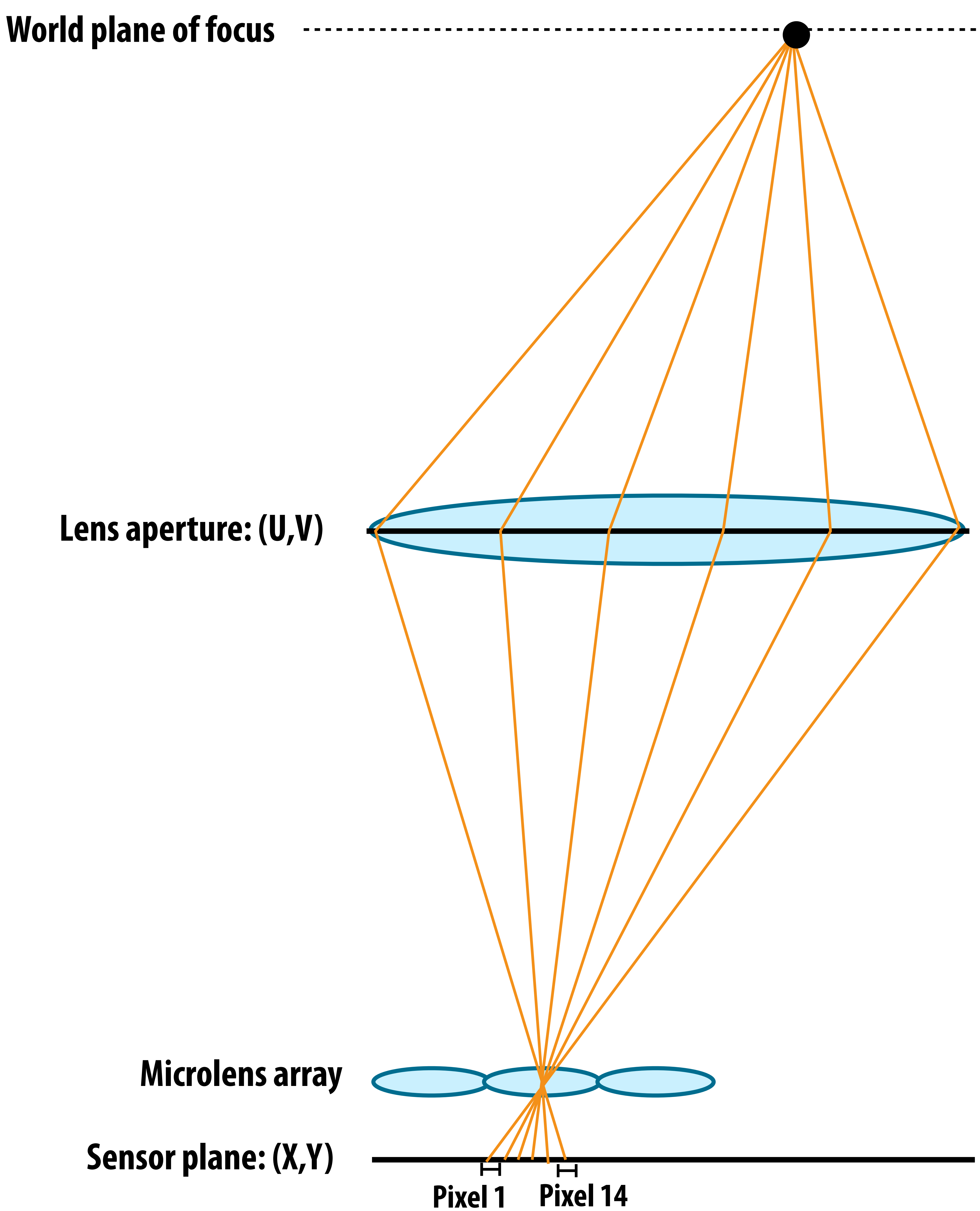
Sub-aperture images

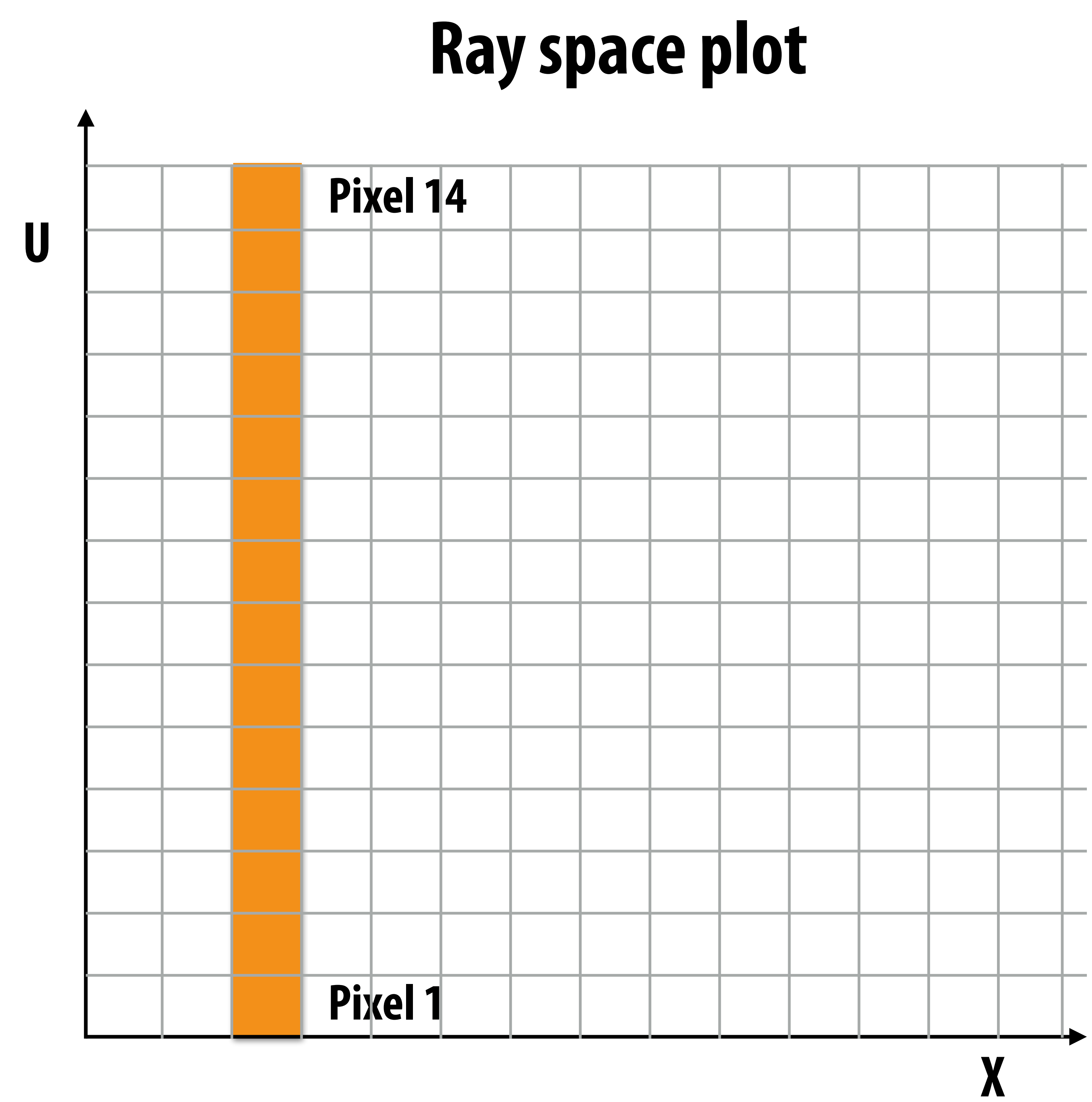
Image from selecting the same pixel under every microlens

Sub-aperture images

Image from selecting the same pixel under every microlens

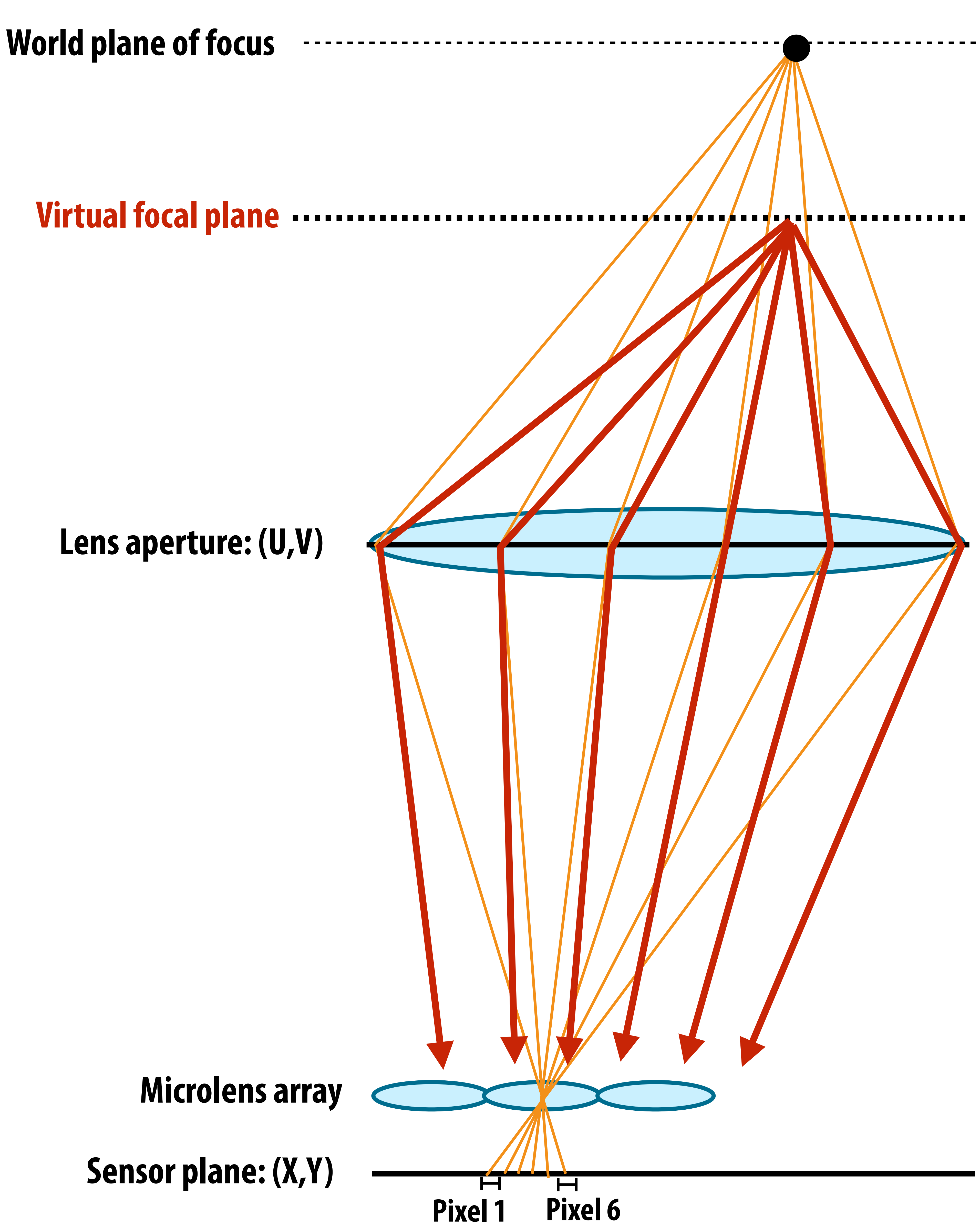
Computing a photograph from a light field

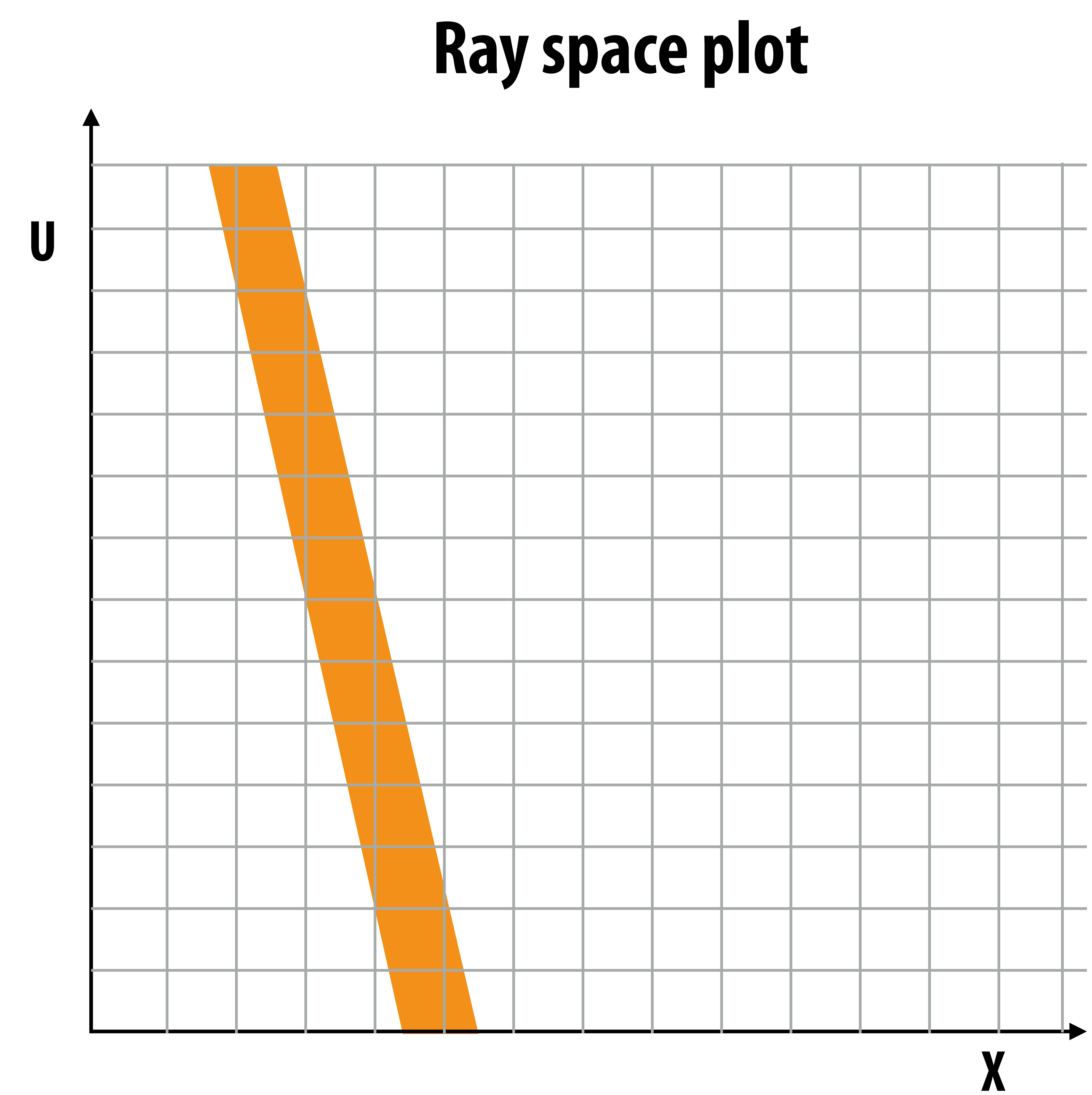




Computing photograph is integral projection (Output image pixel is sum of highlighted light-field sensor pixels)

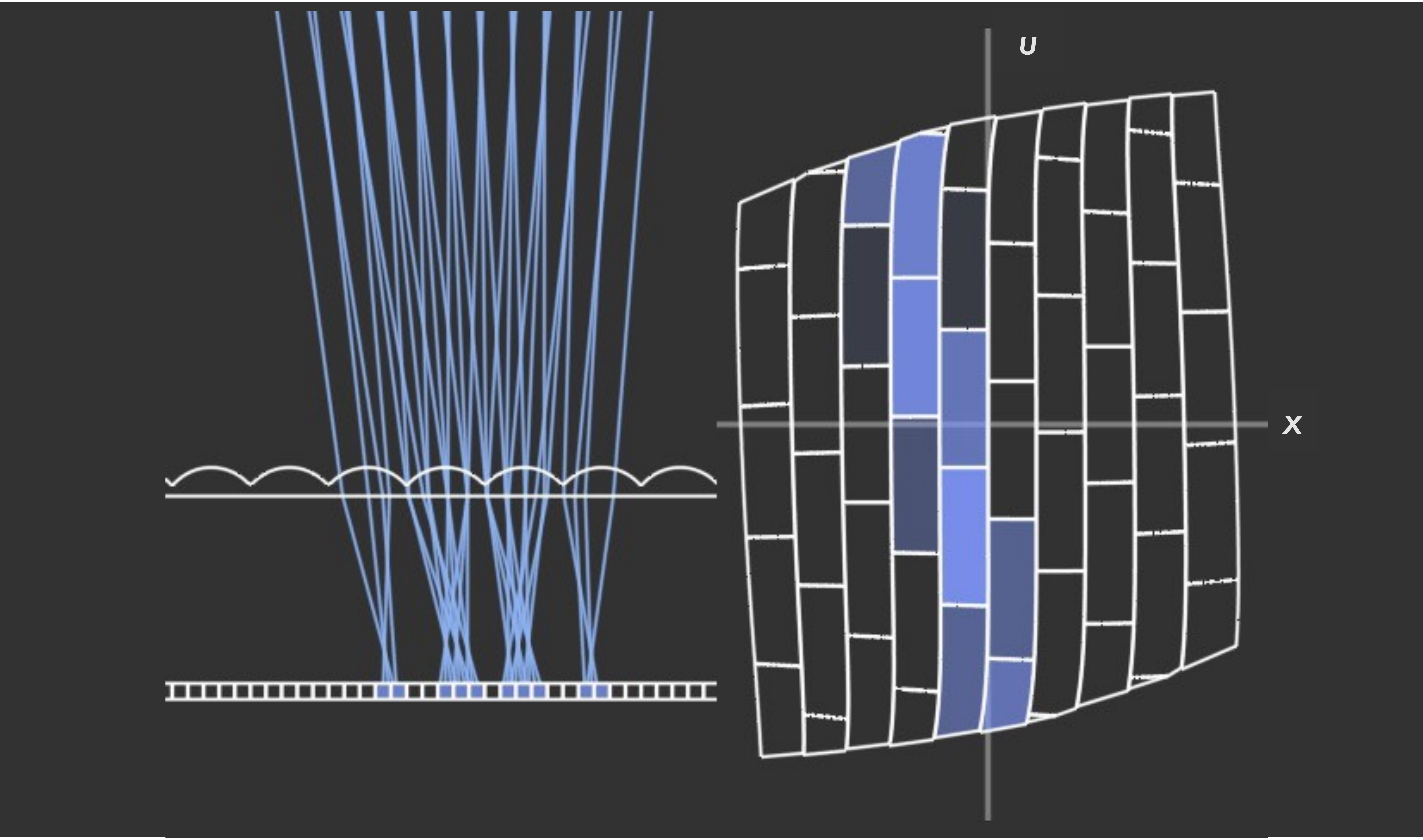
Computing a photograph at a new focal plane





Computing photograph is integral projection (Output image pixel is integral over highlighted region: resample)

Output image pixel is sum of many sensor pixels



Slide credit: Ren Ng

Slide credit: Ren Ng

Recall: split-pixel sensor

Used in cell-phone cameras today to assist with autofocus (also called dual-pixel sensor)

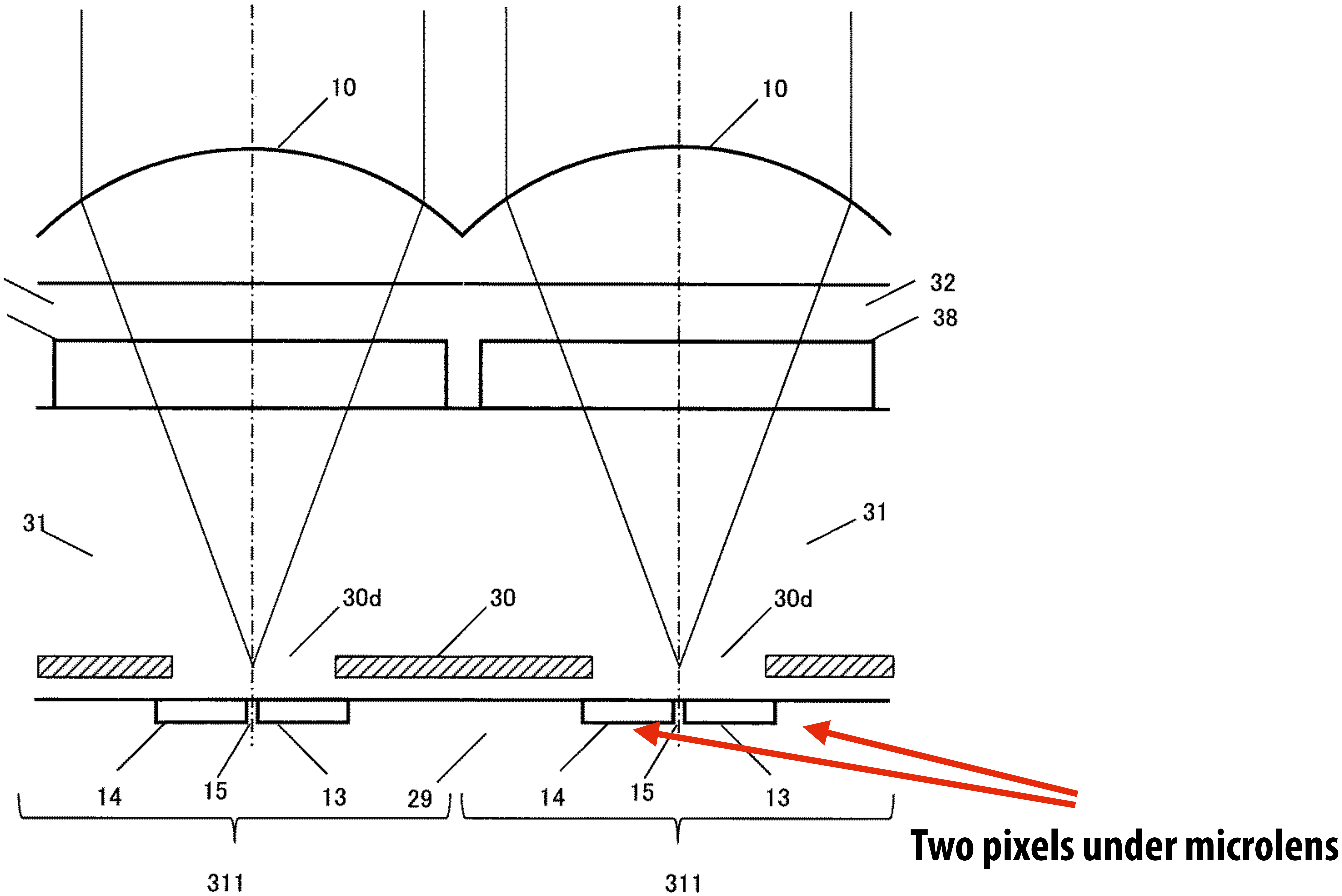


Image credit: Nikon

Virtual reality displays

Virtual reality (VR) vs augmented reality (AR)

VR = virtual reality

User is completely immersed in virtual world (sees only light emitted by display

AR = augmented reality

Display is an overlay that augments user's normal view of the real world (e.g., terminator)

Image credit: Terminator 2 (naturally)

VR headsets

Oculus Rift

HTC Vive

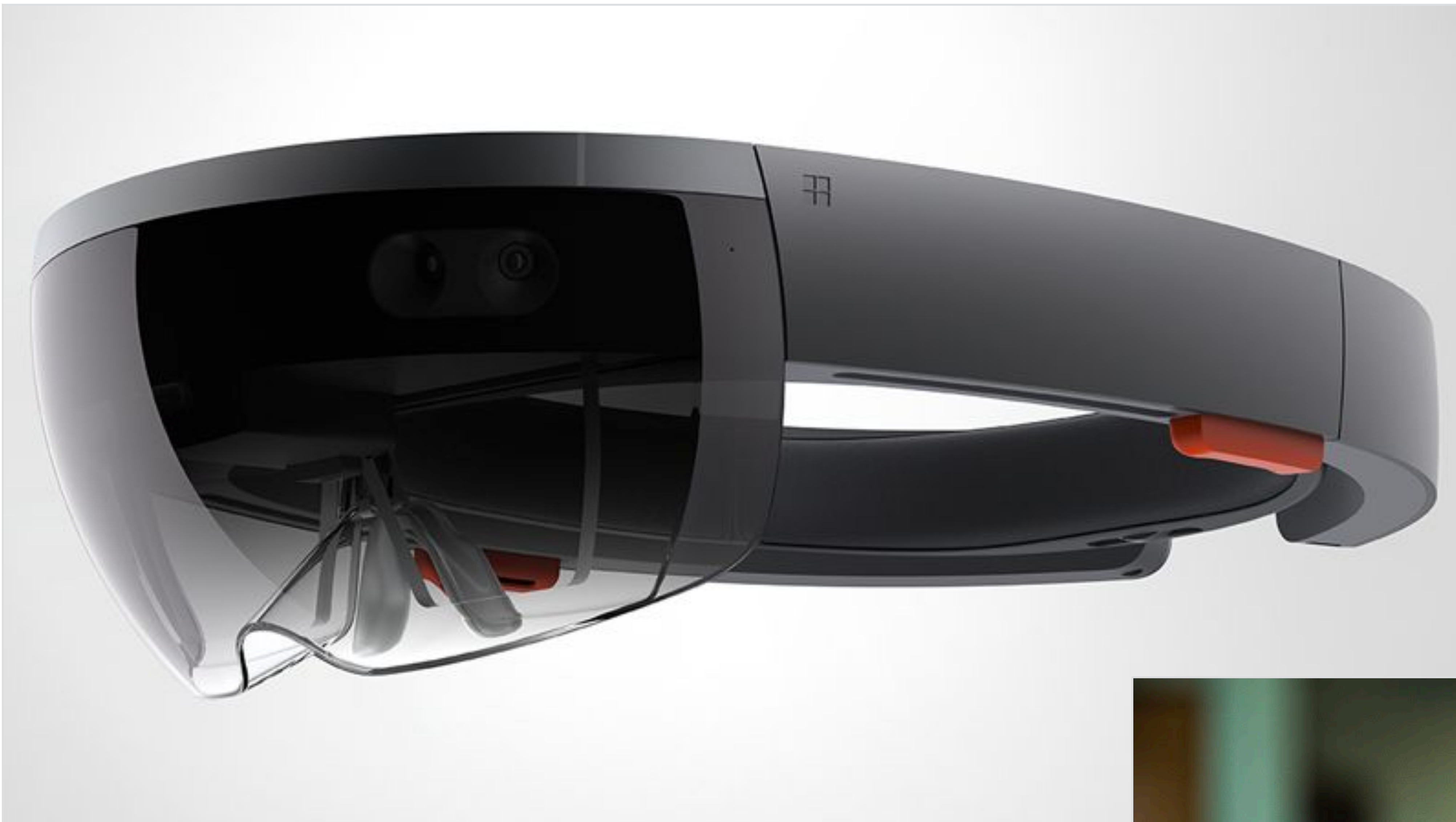
Sony Morpheus

Oculus Go

Google Daydream

Google Cardboard

AR headsets



Microsoft Hololens

Magic Leap One

Oculus Rift CV1

Oculus Rift CV1 headset

Image credit: ifixit.com

Oculus Rift CV1 headset

Image credit: ifixit.com

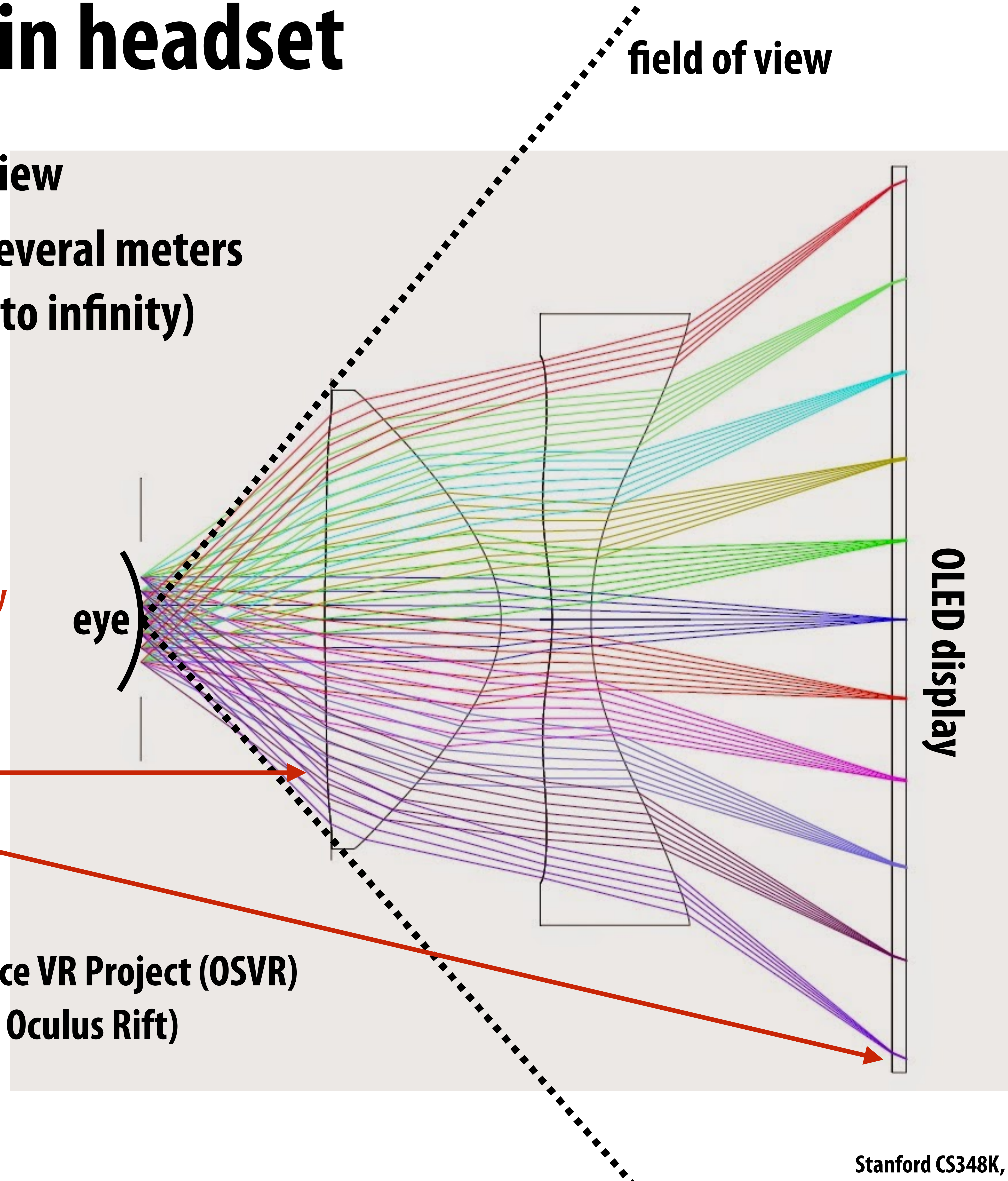
Oculus Rift CV1 headset

1080x1200 OLED display per eye
(2160 x 1200 total pixels)
90 Hz refresh rate
110º field of view

Image credit: ifixit.com

Role of optics in headset

1. Create wide field of view
2. Place focal plane at several meters away from eye (close to infinity)



Note: parallel lines reaching eye converge to a single point on display (eye accommodates to plane near infinity)

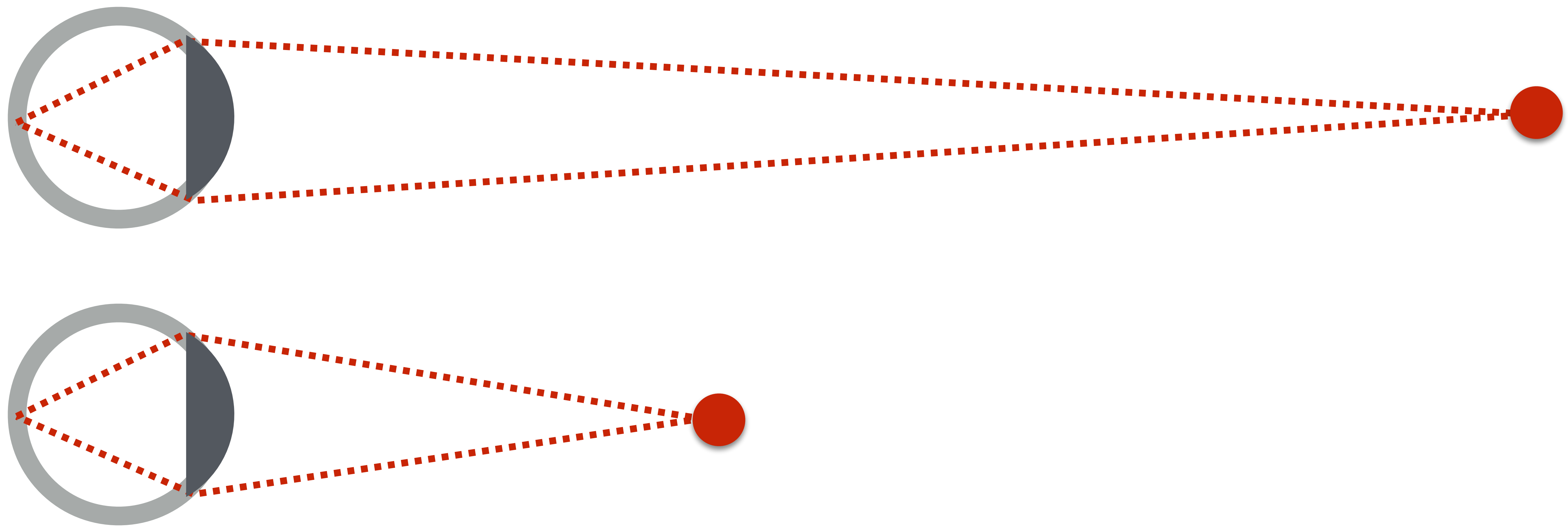
Lens diagram from Open Source VR Project (OSVR)
(Not the lens system from the Oculus Rift)
http://www.osvr.org/

Accommodation and vergence

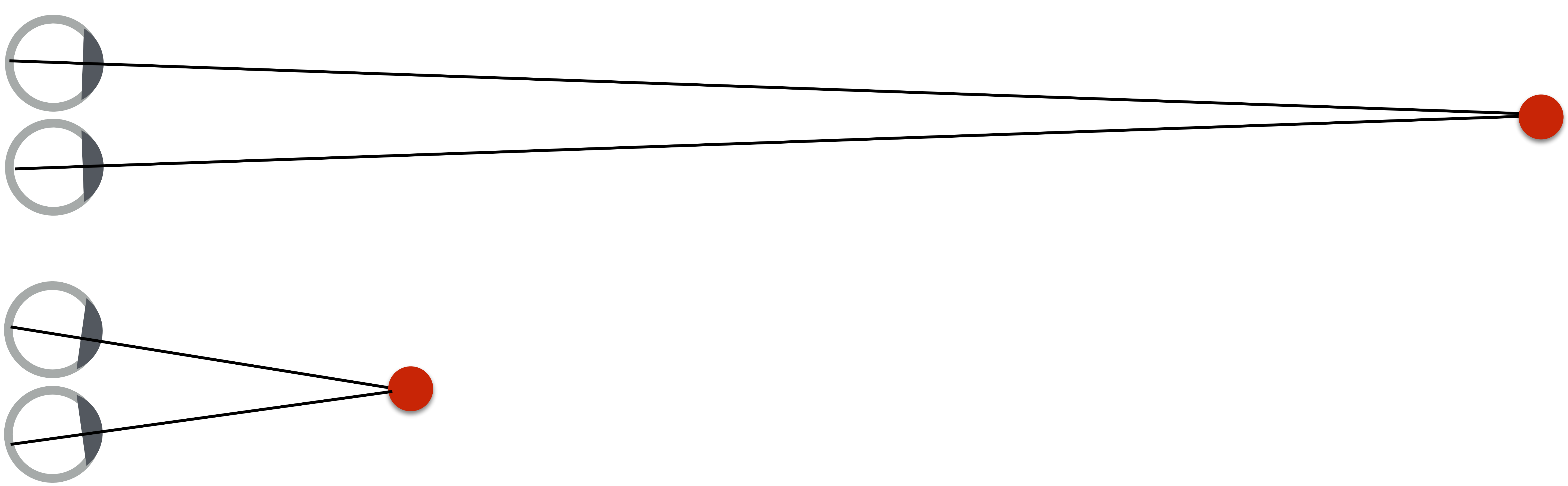
Accommodation: changing the optical power of the eye to focus at different distances

Eye accommodated at far distance

Eye accommodated at near distance

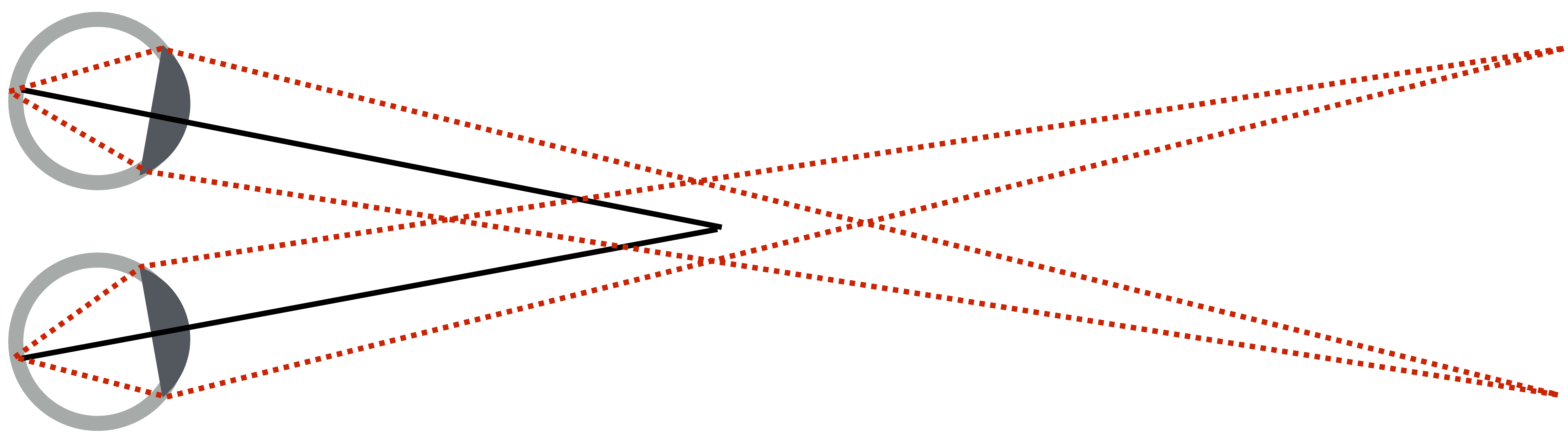


Vergence: rotation of eye to ensure projection of object falls in center of retina



Accommodation - vergence conflict

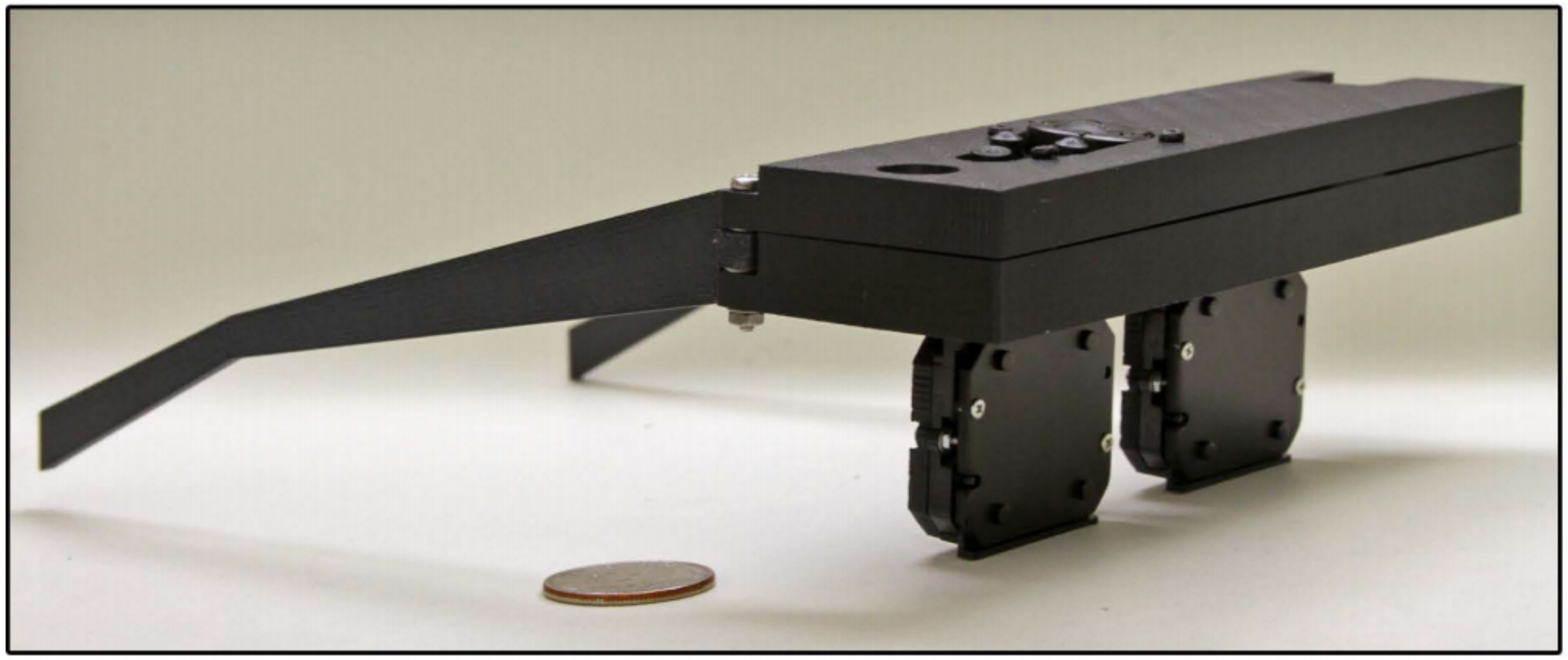
- **Given design of current VR displays, consider what happens when objects are up-close to eye in virtual scene**
  - Eyes must remain accommodated to near infinity (otherwise image on screen won't be in focus)
  - But eyes must converge in attempt to fuse stereoscopic images of object up close
  - Brain receives conflicting depth clues… (discomfort, fatigue, nausea)

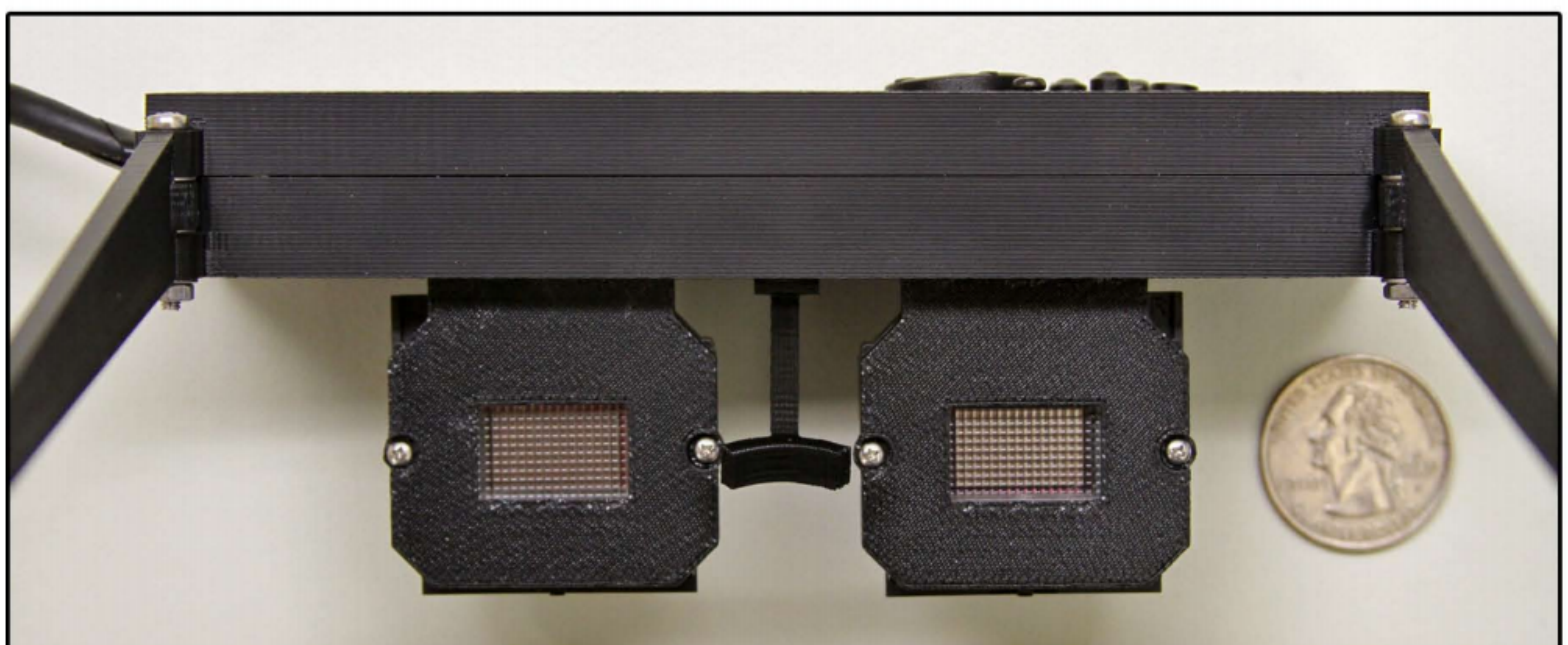


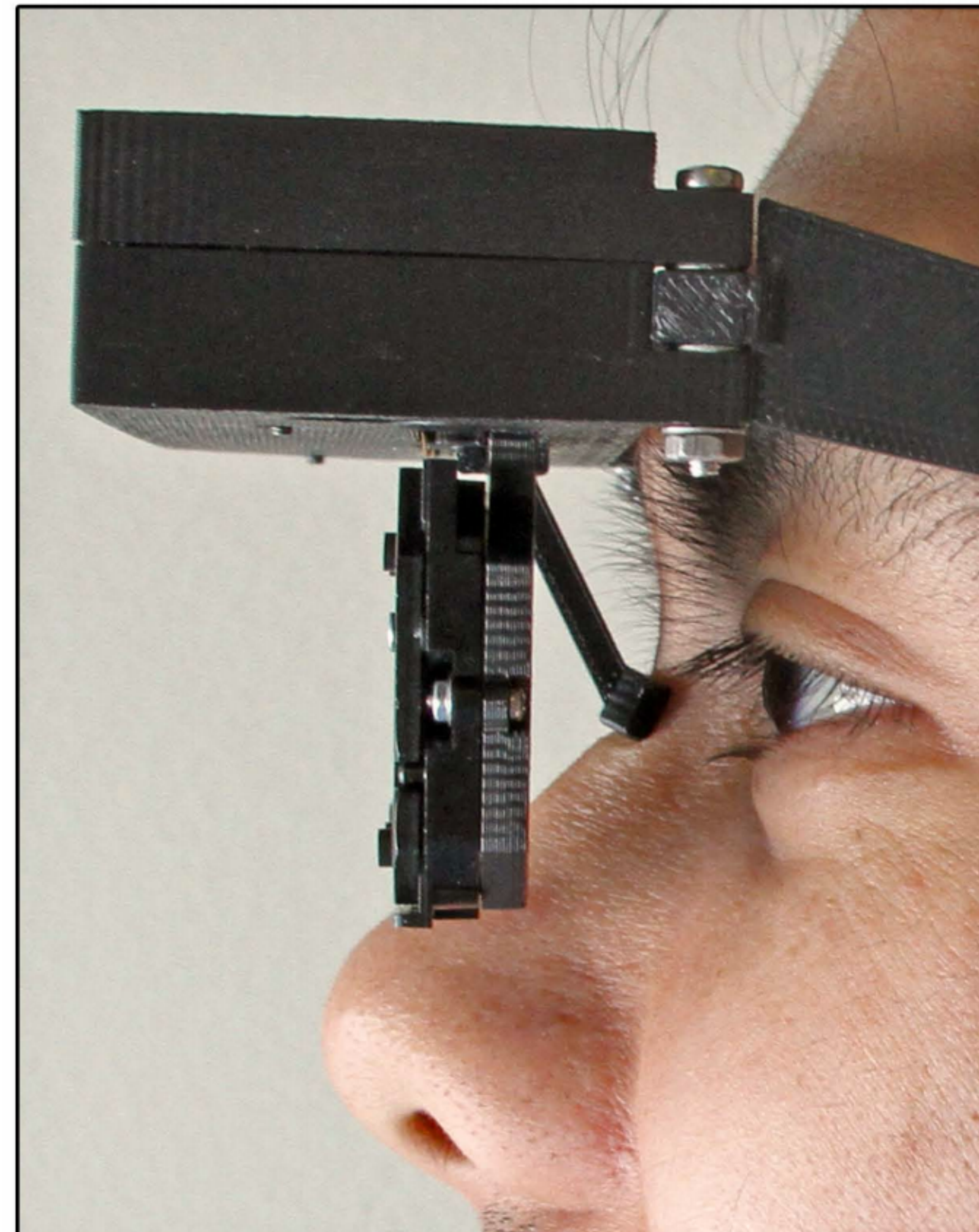
This problem stems from nature of display design. If you could just make a display that emits the light field that would be produced by a virtual scene, then you could avoid the accommodation - vergence conflict…

Aside: near-eye light field displays

Goal: recreate light field in front of eye







Oculus CV1 IR camera and IR LEDs

60Hz IR Camera
(measures absolute position of headset 60 times a second)

Headset contains:
IR LEDs (tracked by camera)
Gyro + accelerometer (1000Hz) (for rapid relative positioning)

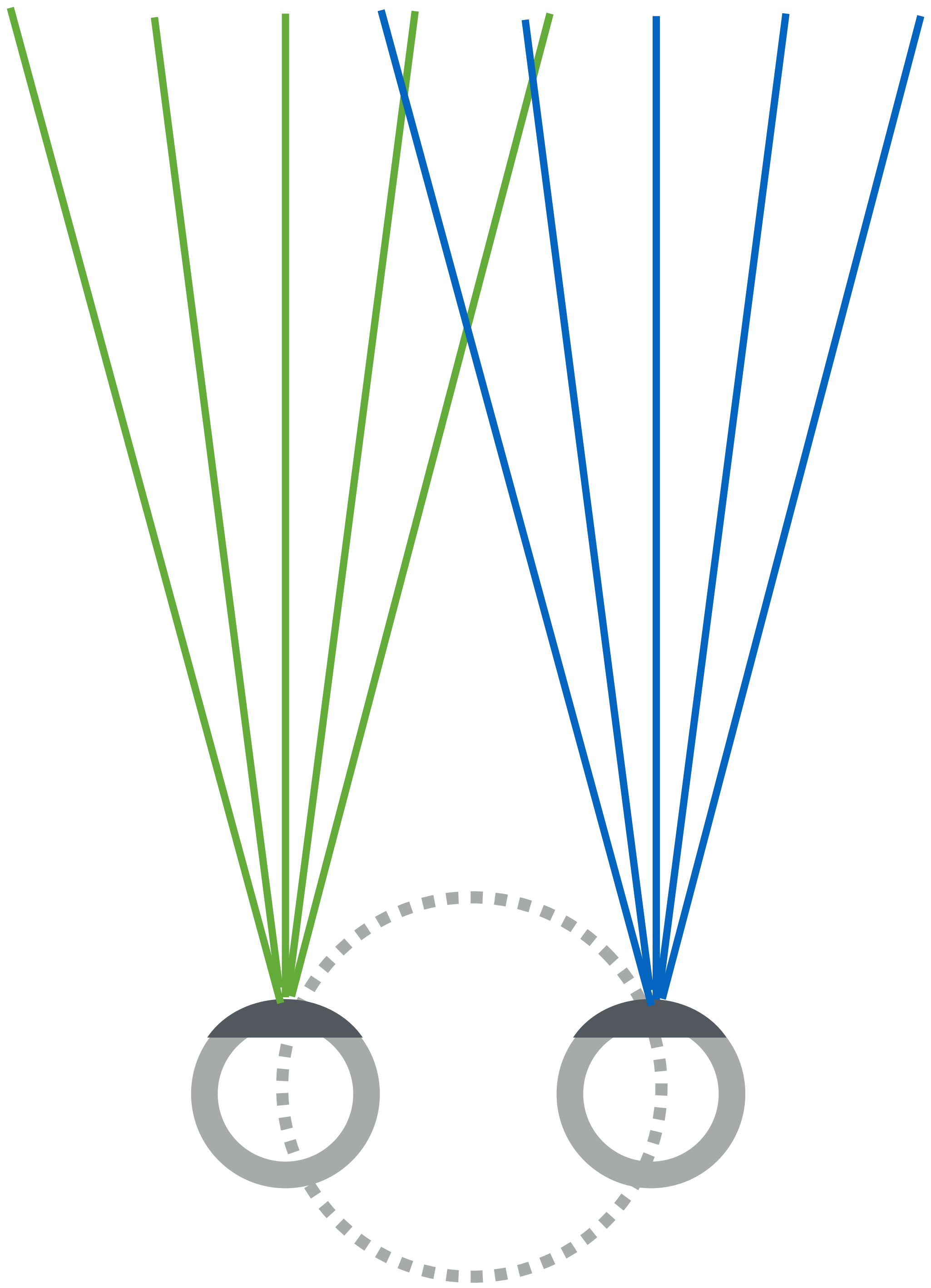
Image credit: ifixit.com

Acquiring VR content

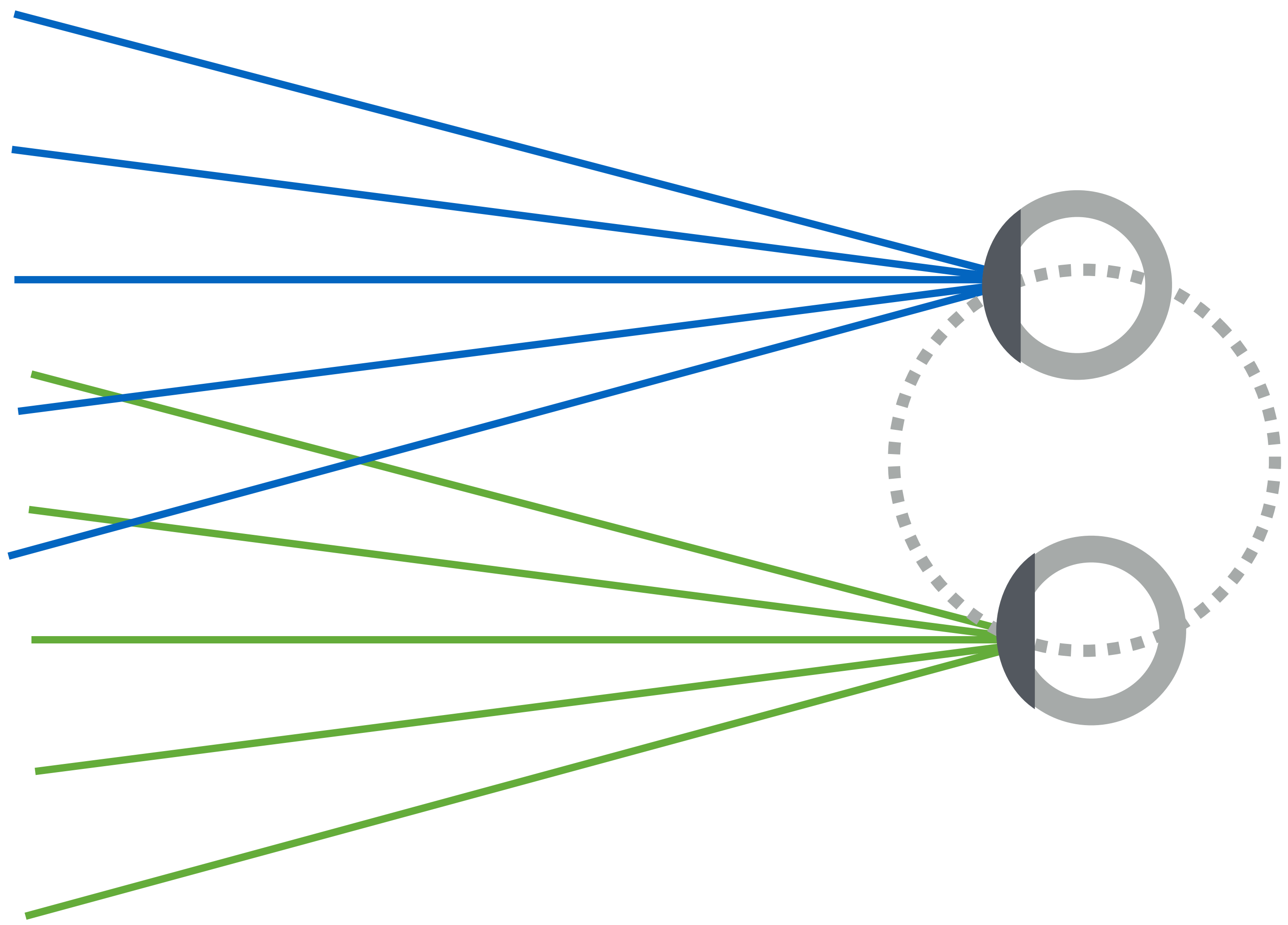
Google's Jump VR video:
Yi Halo Camera (17 cameras)

Facebook Manifold
(16 8K cameras)

Stereo, 360-degree viewing



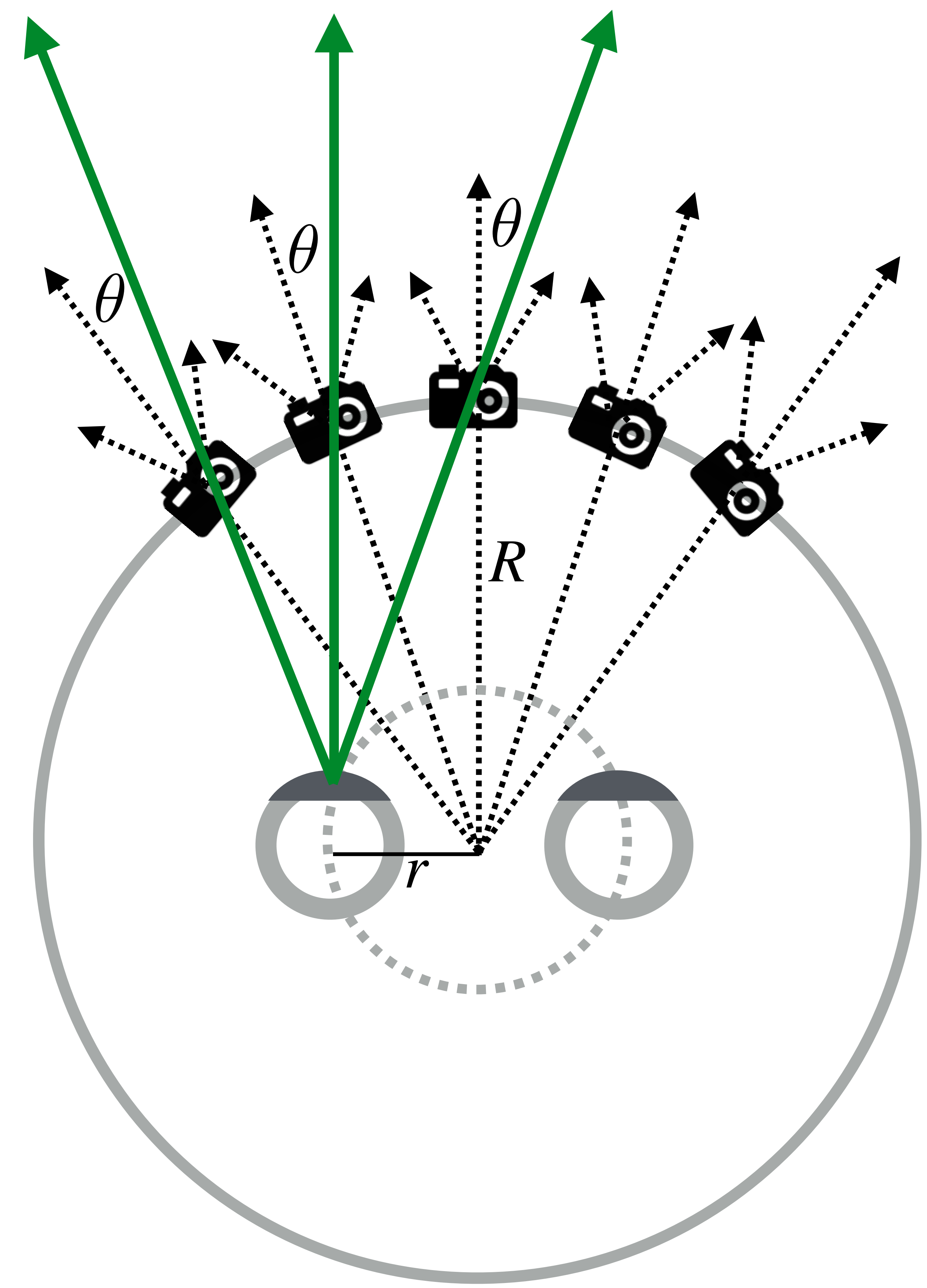
Stereo, 360-degree viewing



Measuring light arriving at left eye

Left eye

$$\sin\theta = r/R$$

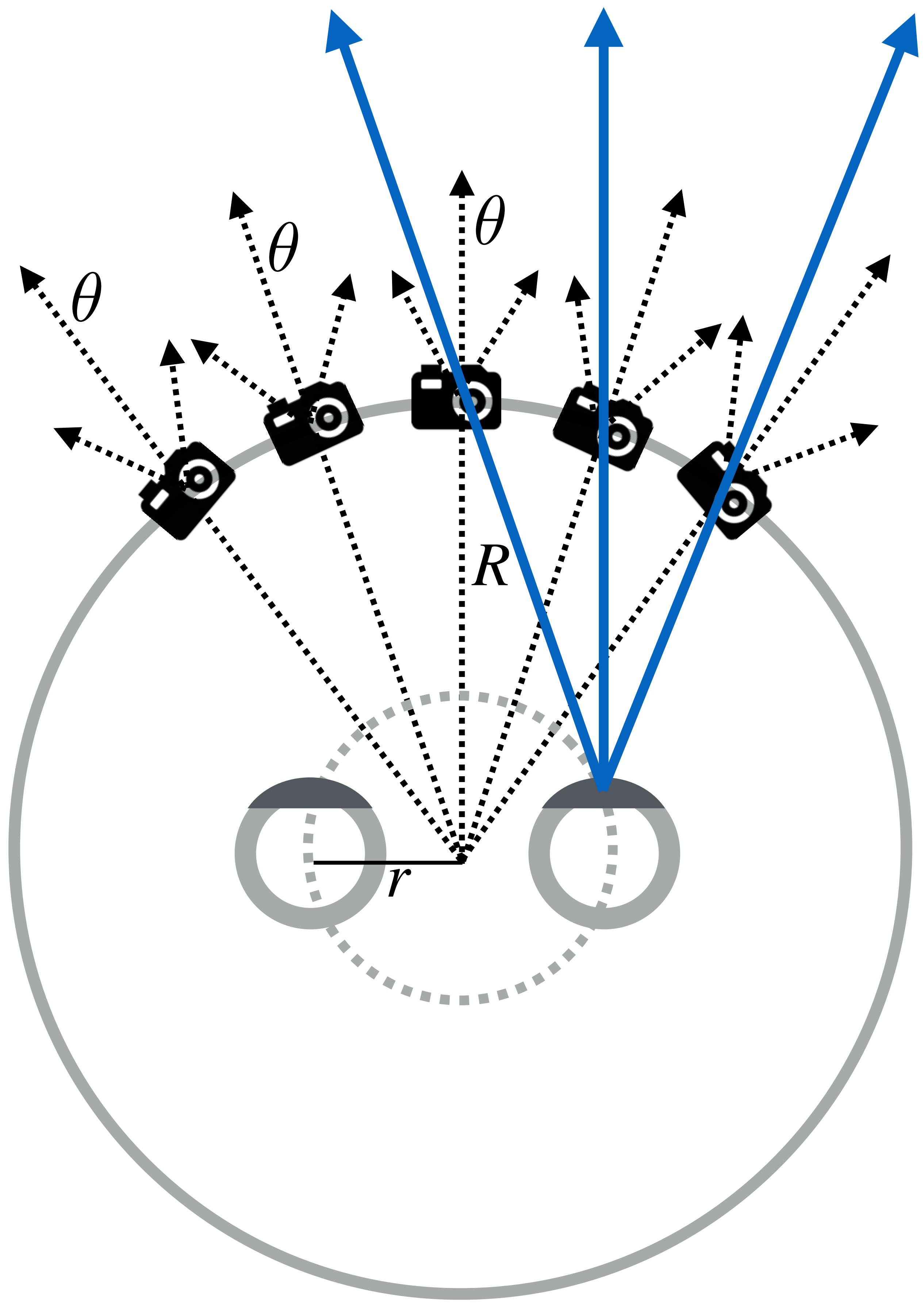


[Credit: Camera icon by Venkatesh Aiyulu from The Noun Project]

Measuring light arriving at right eye

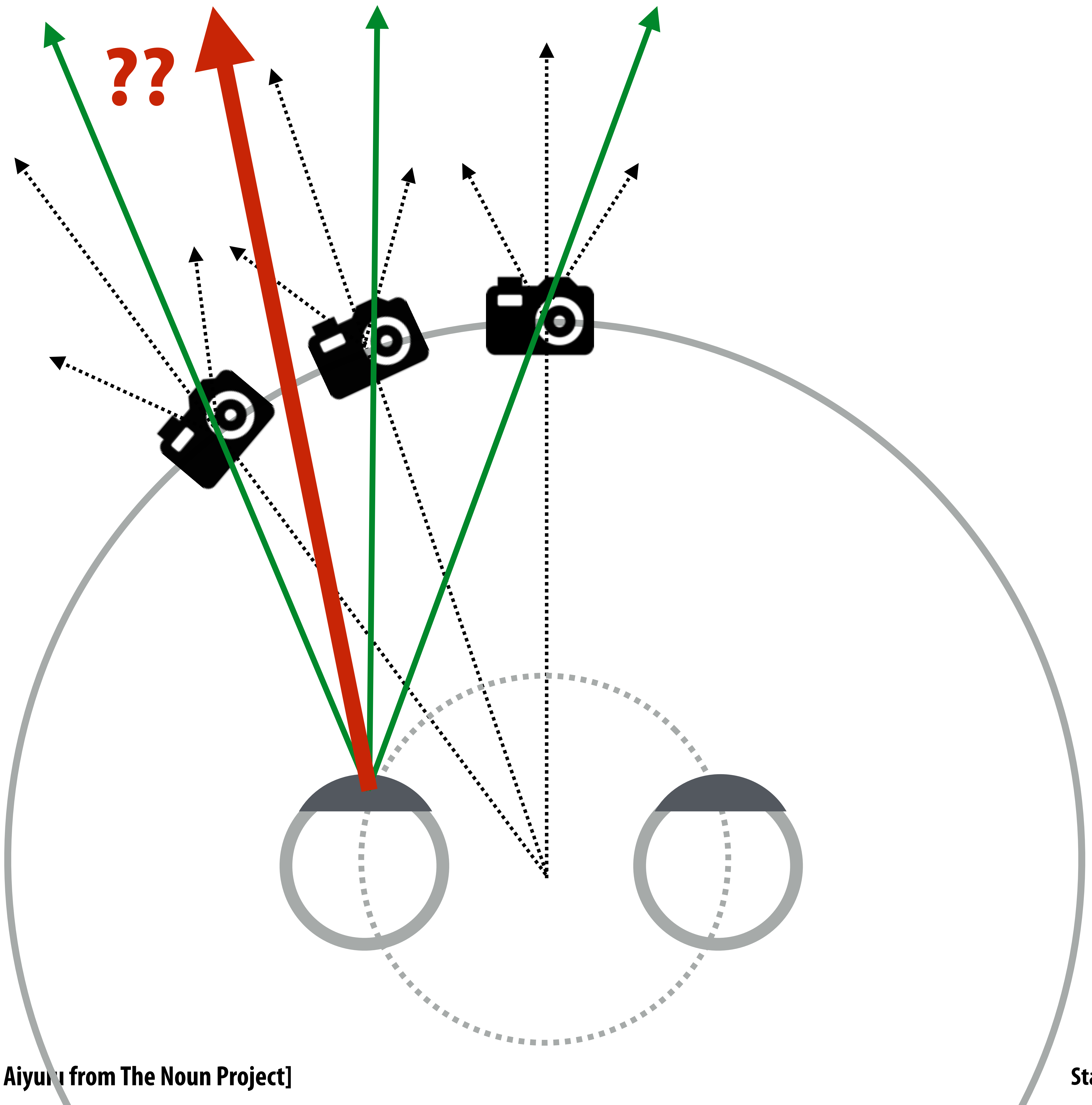
Right eye

$$\sin\theta = -r/R$$



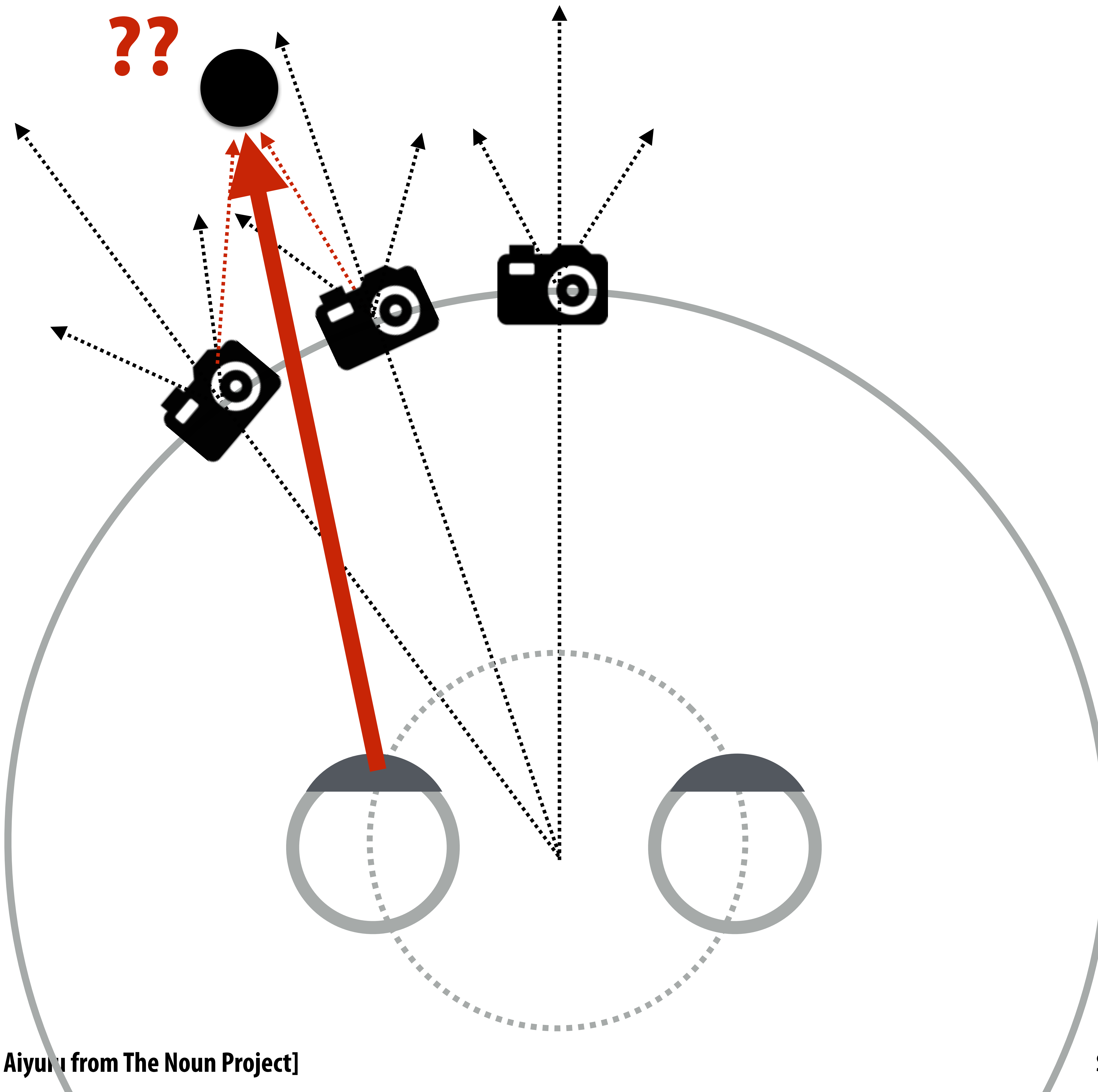
[Credit: Camera icon by Venkatesh Aiyulu from The Noun Project]

How to estimate rays at “missing” views?



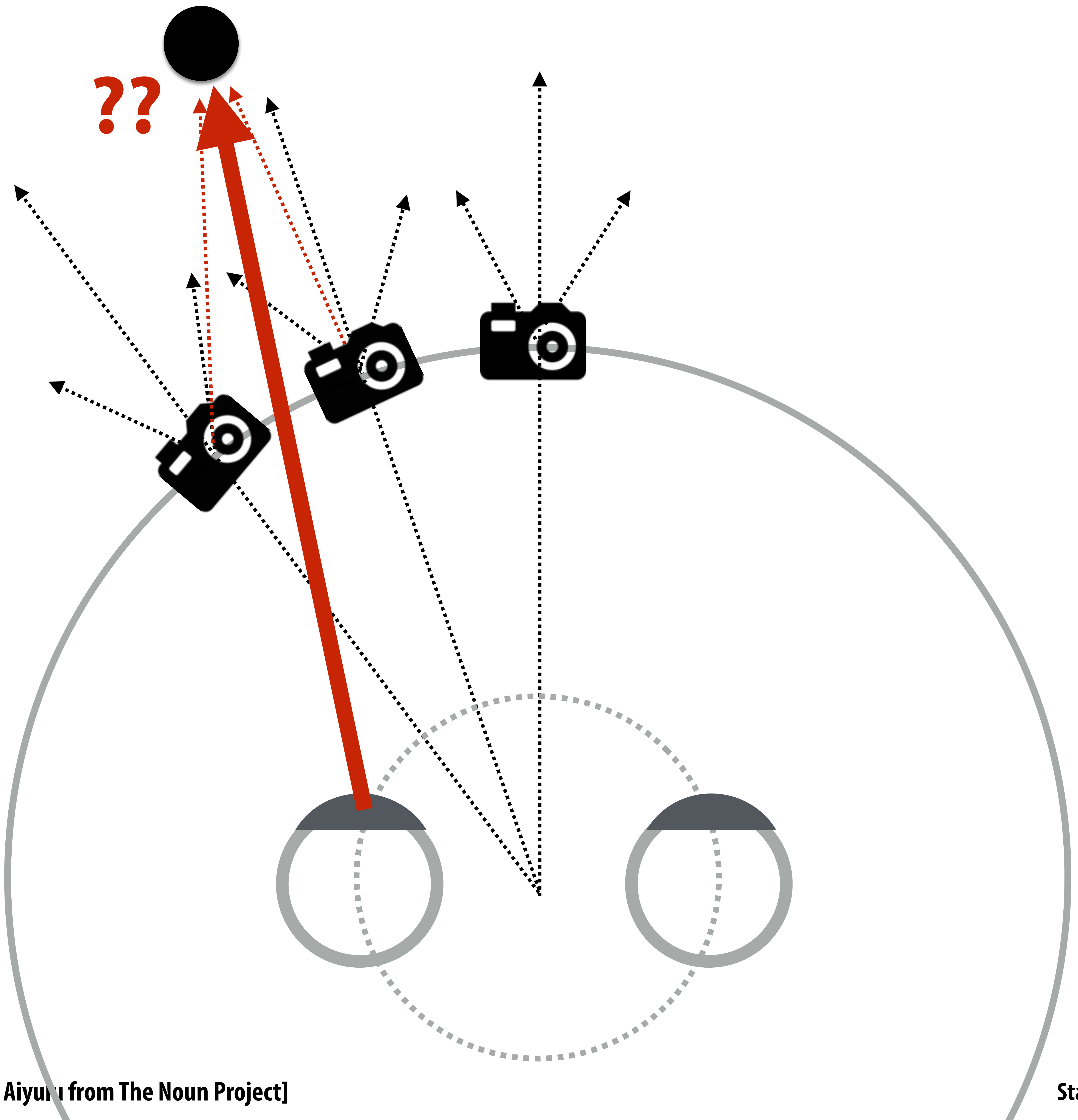
[Credit: Camera icon by Venkatesh Aiyulu from The Noun Project]

Interpolation to novel views depends on scene depth



[Credit: Camera icon by Venkatesh Aiyulu from The Noun Project]

Interpolation to novel views depends on scene depth



[Credit: Camera icon by Venkatesh Aiyulu from The Noun Project]

Computing depth of scene point from two images

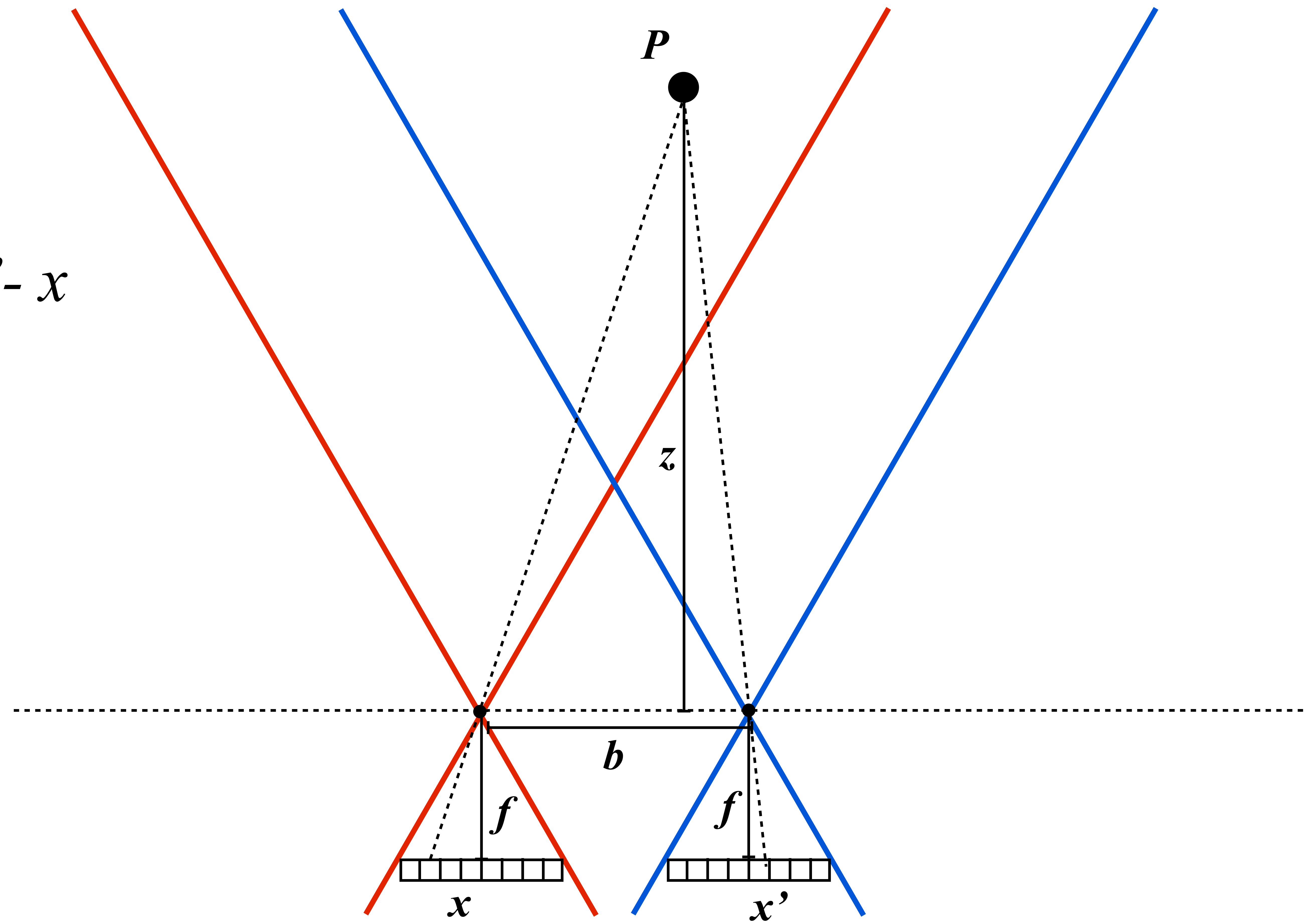
Binocular stereo 3D reconstruction of point *P*: depth from disparity

Focal length: f

Baseline: b

Disparity: $d = x' - x$

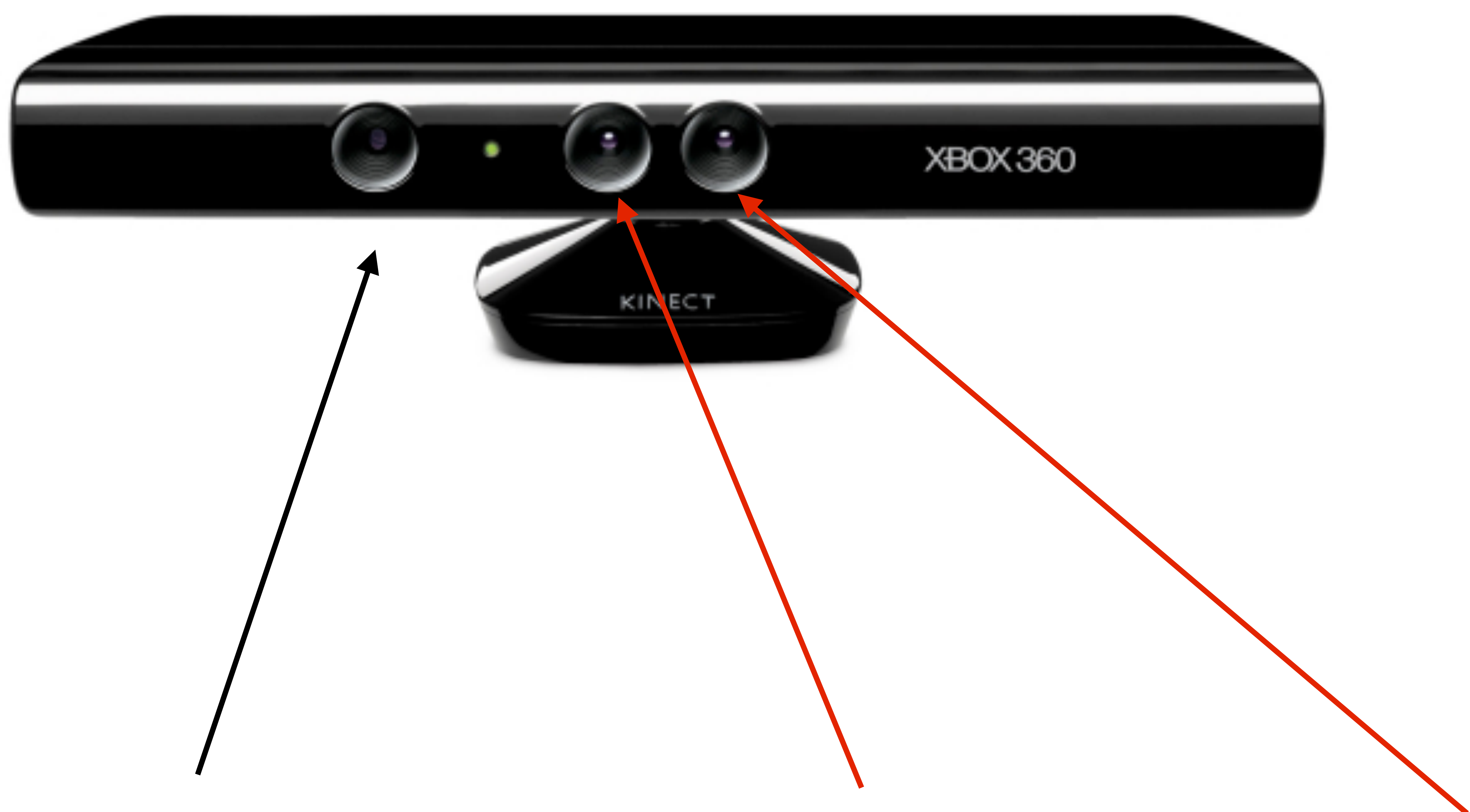
$$z = \frac{bf}{d}$$



Simple reconstruction example: cameras aligned (coplanar sensors), separated by known distance, same focal length

"Disparity" is the distance between object's projected position in the two images: x - x'

Microsoft XBox 360 Kinect



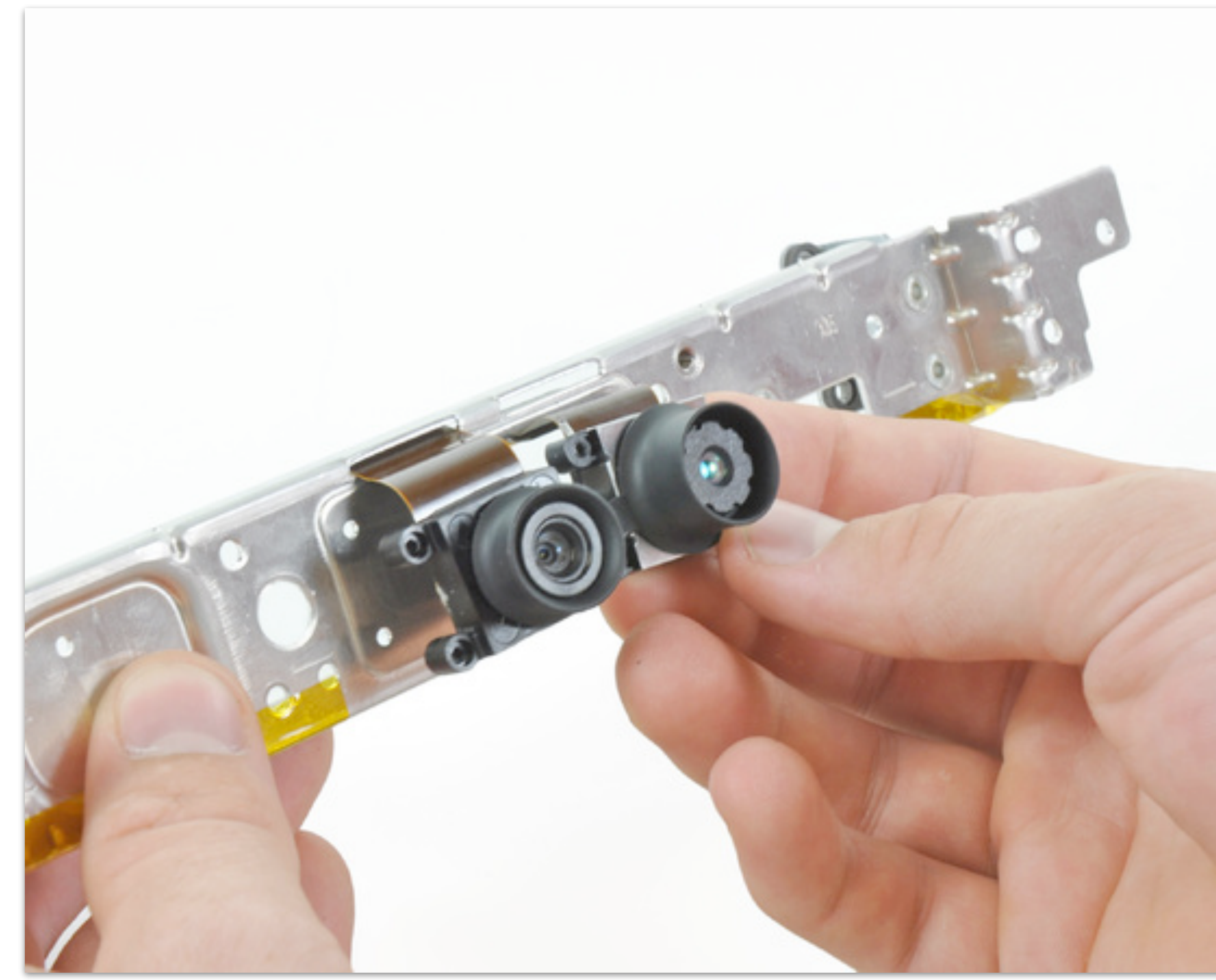


Image credit: iFixIt

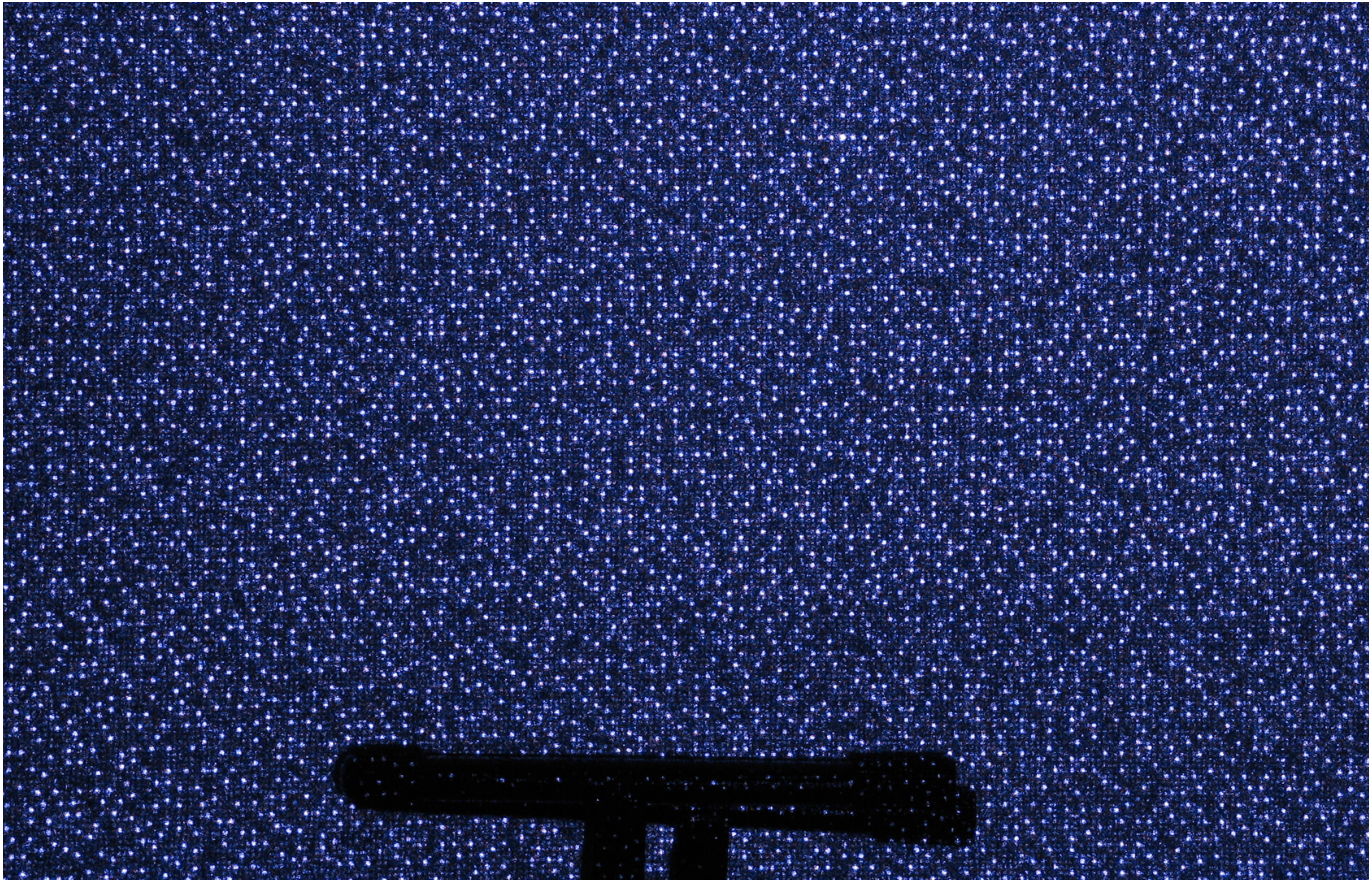
Illuminant
(Infrared Laser + diffuser)

RGB CMOS Sensor
640x480 (w/ Bayer mosaic)

Monochrome Infrared
CMOS Sensor
(Aptina MT9M001)
1280x1024 **

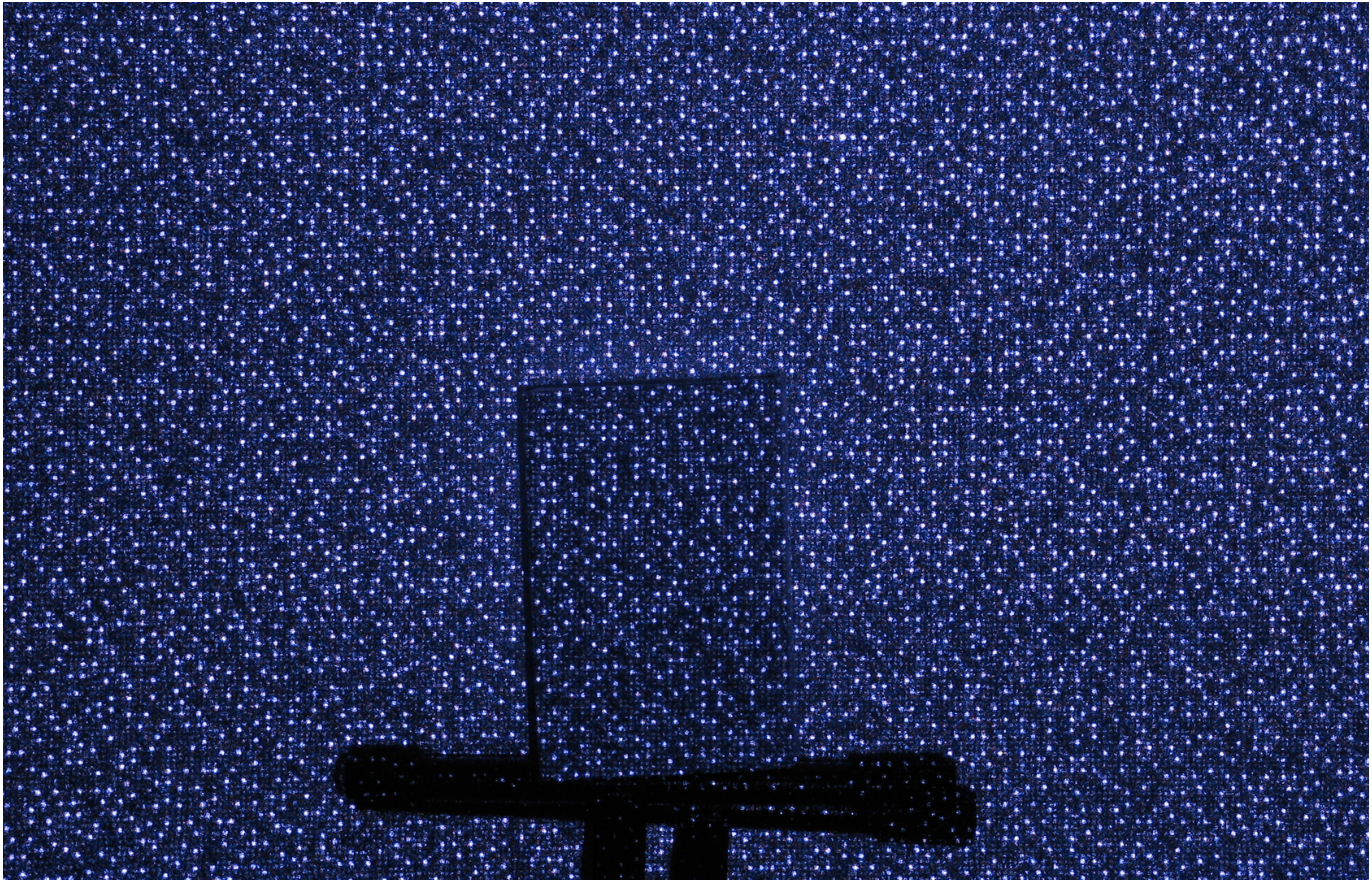
** Kinect returns 640x480 disparity image, suspect sensor is configured for 2x2 pixel binning down to 640x512, then crop

Infrared image of Kinect illuminant output



Credit: www.futurepicture.org

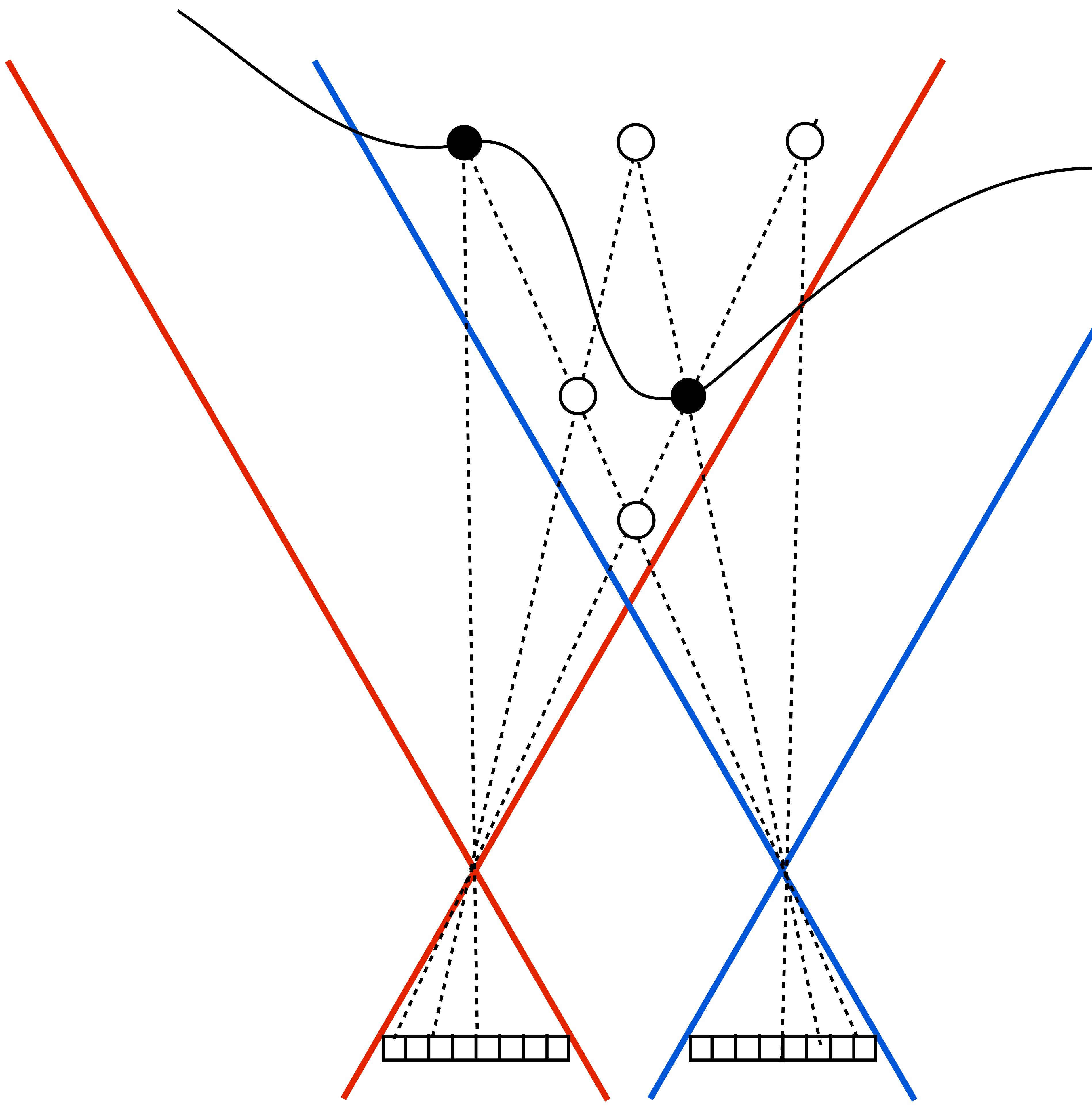
Infrared image of Kinect illuminant output



Credit: www.futurepicture.org

Correspondence problem

How to determine which pairs of pixels in image 1 and image 2 correspond to the same scene point?

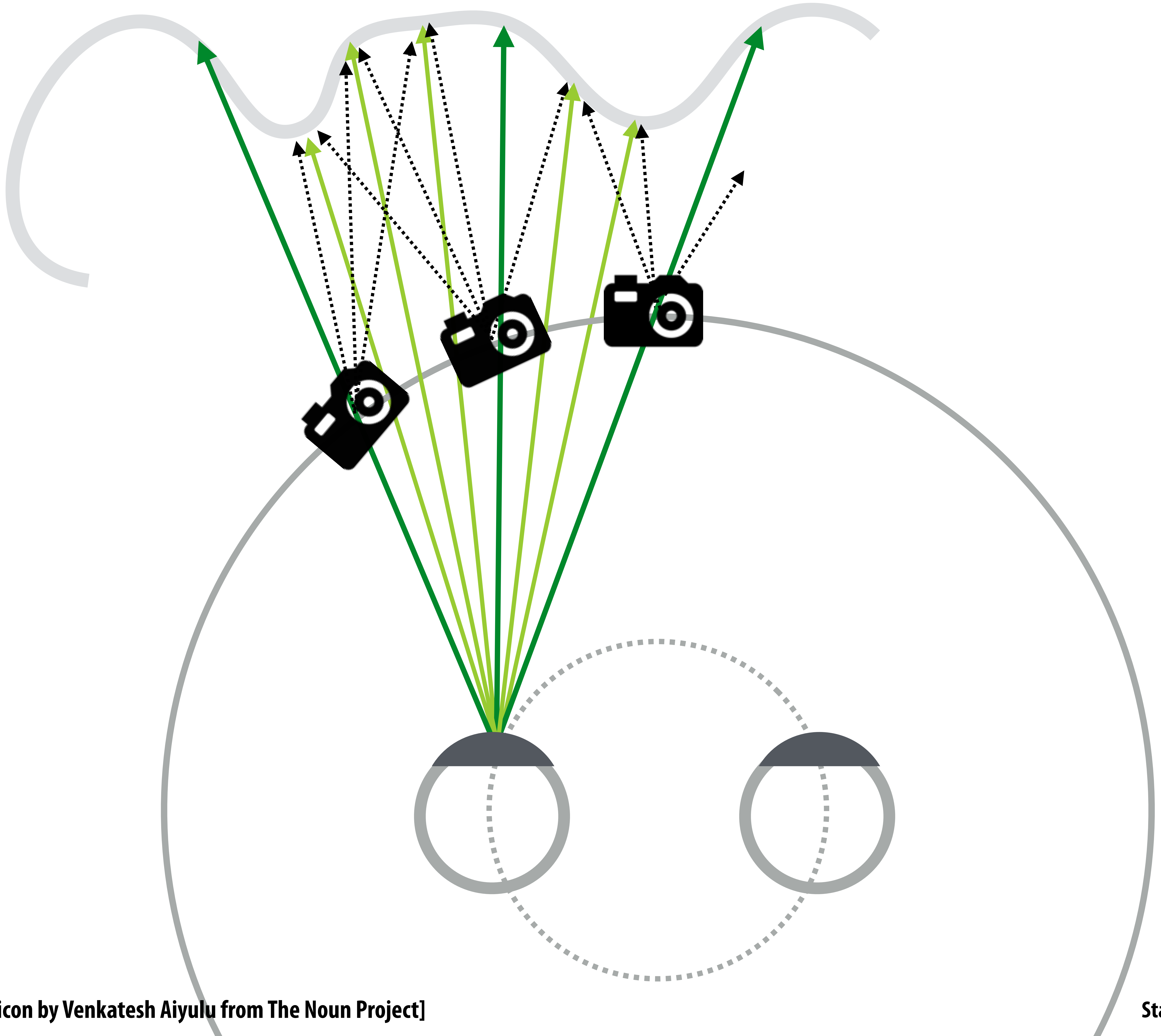


Correspondence problem = compute “flow” between adjacent cameras

- For each pixel in frame from camera *i*, find closest pixel in camera *i+1*
- Google’s Jump pipeline uses a coarse-to-fine algorithm: align 32x32 blocks by searching over local window, then perform per-pixel alignment
  - Recall: H.264 motion estimation, HDR+ burst alignment (same correspondence challenge, but here we are aligning different perspectives at the same time to estimate unknown scene depth, not estimating motion of camera or scene over time)
  - Additional tricks to ensure temporal consistency of flow over time (see papers)

Image credit: Andersen et al. 2016

Left eye: with interpolated rays



[Credit: Camera icon by Venkatesh Aiyulu from The Noun Project]

Omnidirectional stereo (ODS) representation

- Unique panorama of size W x H for left and right eye
- Good: can be saved, compressed, edited as normal video
- Column *j* of pixels corresponds to column from interpolated camera at ring position at angle: $\frac{2\pi j}{W}$

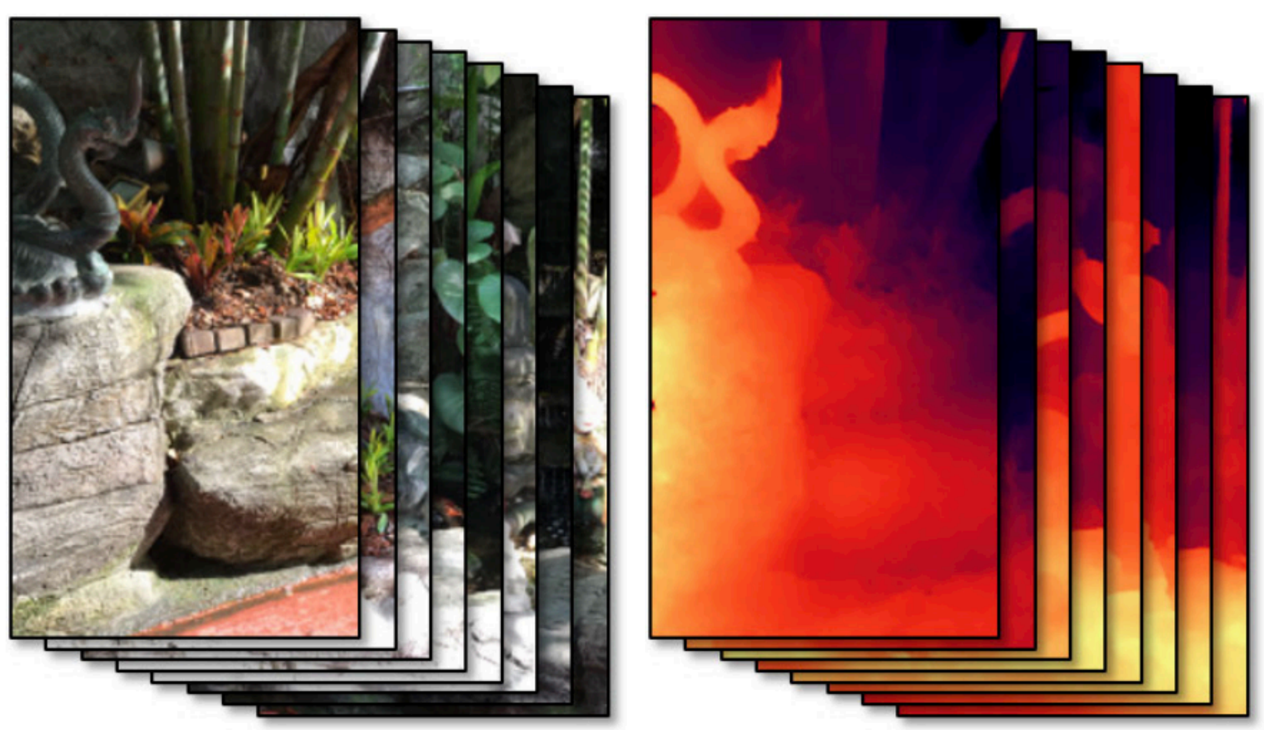
Overlay of Left and Right eye ODS panoramas

Image credit: Andersen et al. 2016

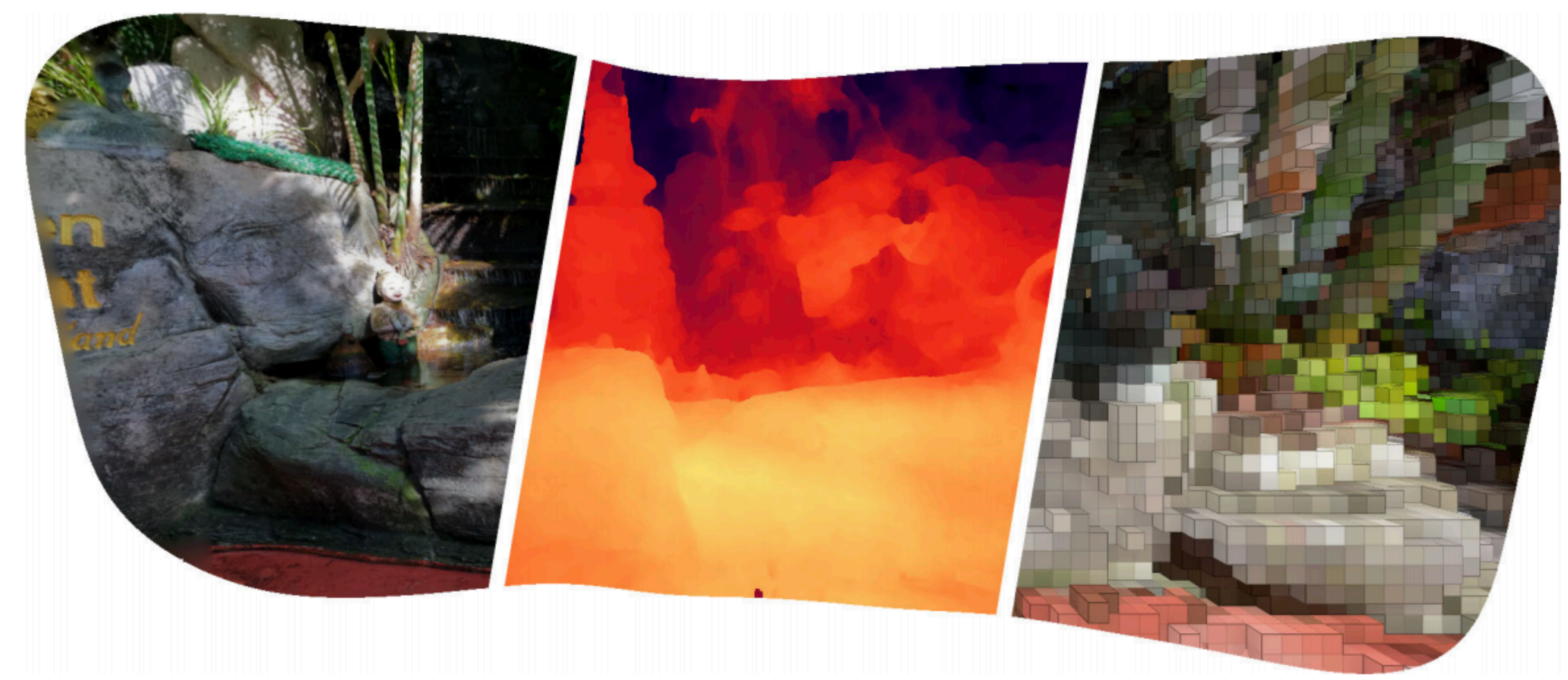
“Casual 3D photography”

- **Acquisition: wave a smartphone camera around to acquire images of scene from multiple viewpoints**
- **Processing: construct 3D representation of scene from photos**
  - Render a textured triangle mesh

Dual-camera Smartphone



Burst of photos + depth maps



Stitch photos into depth panorama, create 3D mesh + textures, render during VR viewing