



Light fields

Advanced Graphics & Image Processing

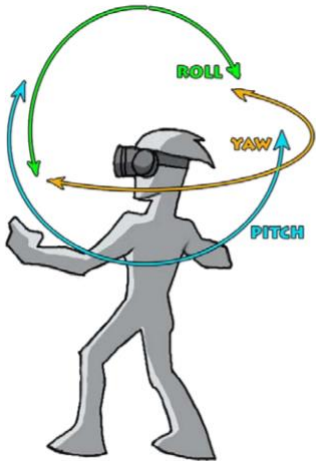
Rafał Mantiuk

Computer Laboratory, University of Cambridge

Motivation: 3DoF vs 6DoF in VR

3DoF

- ▶ Tracking with inexpensive Inertial Measurements Units
- ▶ Content:
 - ▶ Geometry-based graphics
 - ▶ Omnidirectional stereo video
 - ▶ May induce cyber-sickness due to the lack of motion depth cues



6DoF

- ▶ Requires internal (inside-out) or external tracking
- ▶ Content:
 - ▶ Geometry-based graphics
 - ▶ Point-cloud rendering
 - ▶ Image-based rendering
 - ▶ View interpolation
 - ▶ **Light fields**
 - ▶ ...

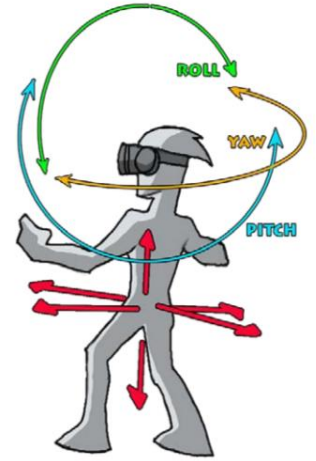
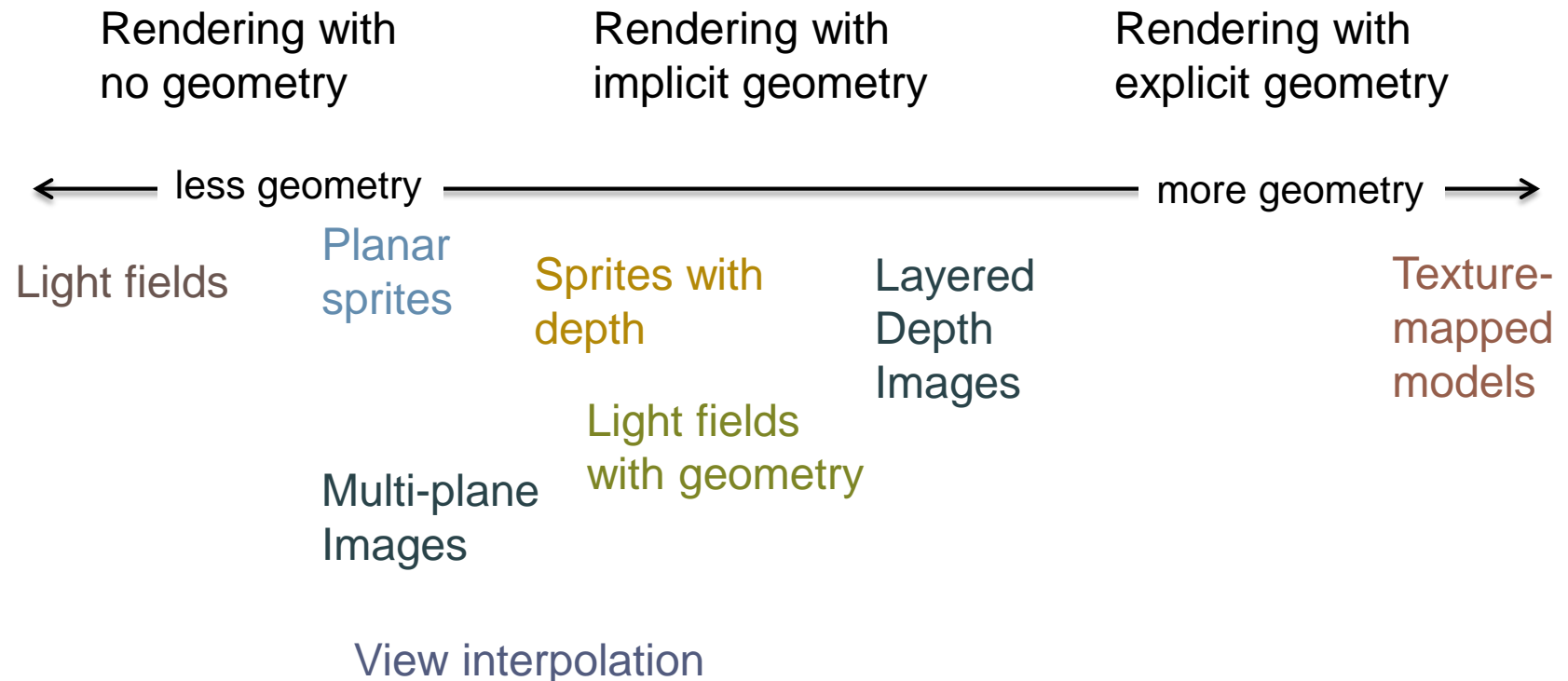


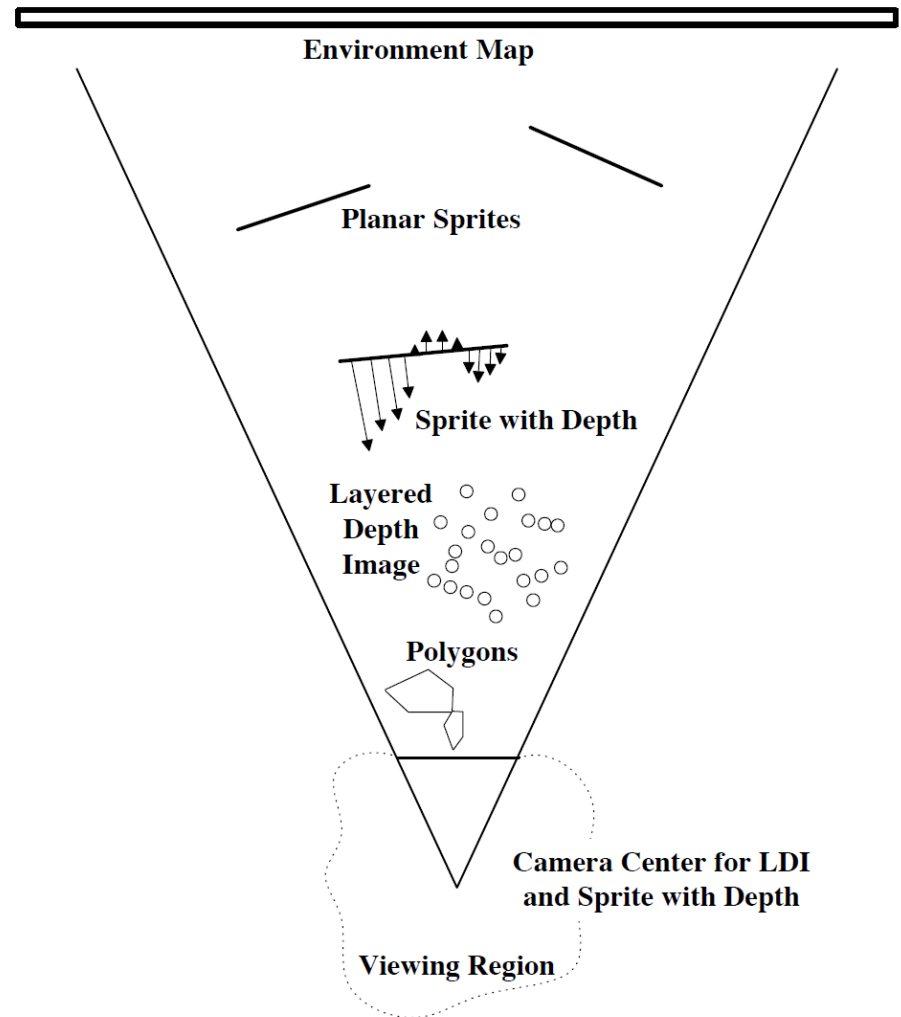
Image Based Rendering

- ▶ Render new views from images/textures with reduced or no geometry



Mixing geometry and IBR primitives

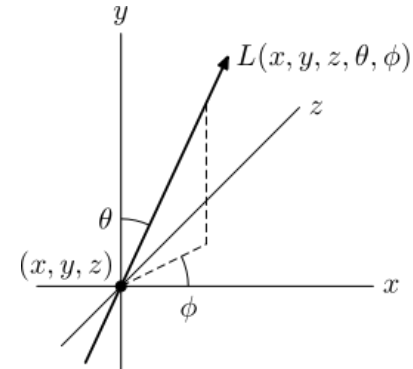
- ▶ Type of rendering method/primitive to use depends on the viewing distance
- ▶ Accurate geometry is less relevant for distant objects
 - ▶ As they are less affected by parallax
- ▶ As we move closer to objects, geometry becomes more important



From a plenoptic function to a light field

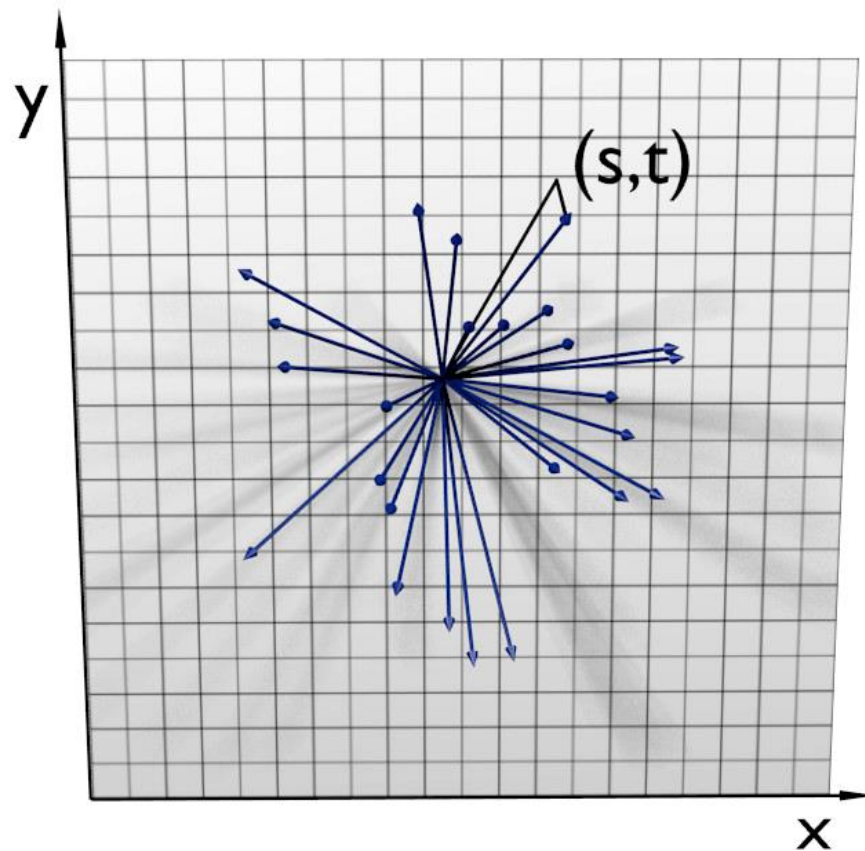
- ▶ Plenoptic function – describes all possible rays in a 3D space

- ▶ Function of position (x, y, z) and ray direction (θ, ϕ)
- ▶ But also wavelength λ and time t
- ▶ Between 5 and 7 dimensions

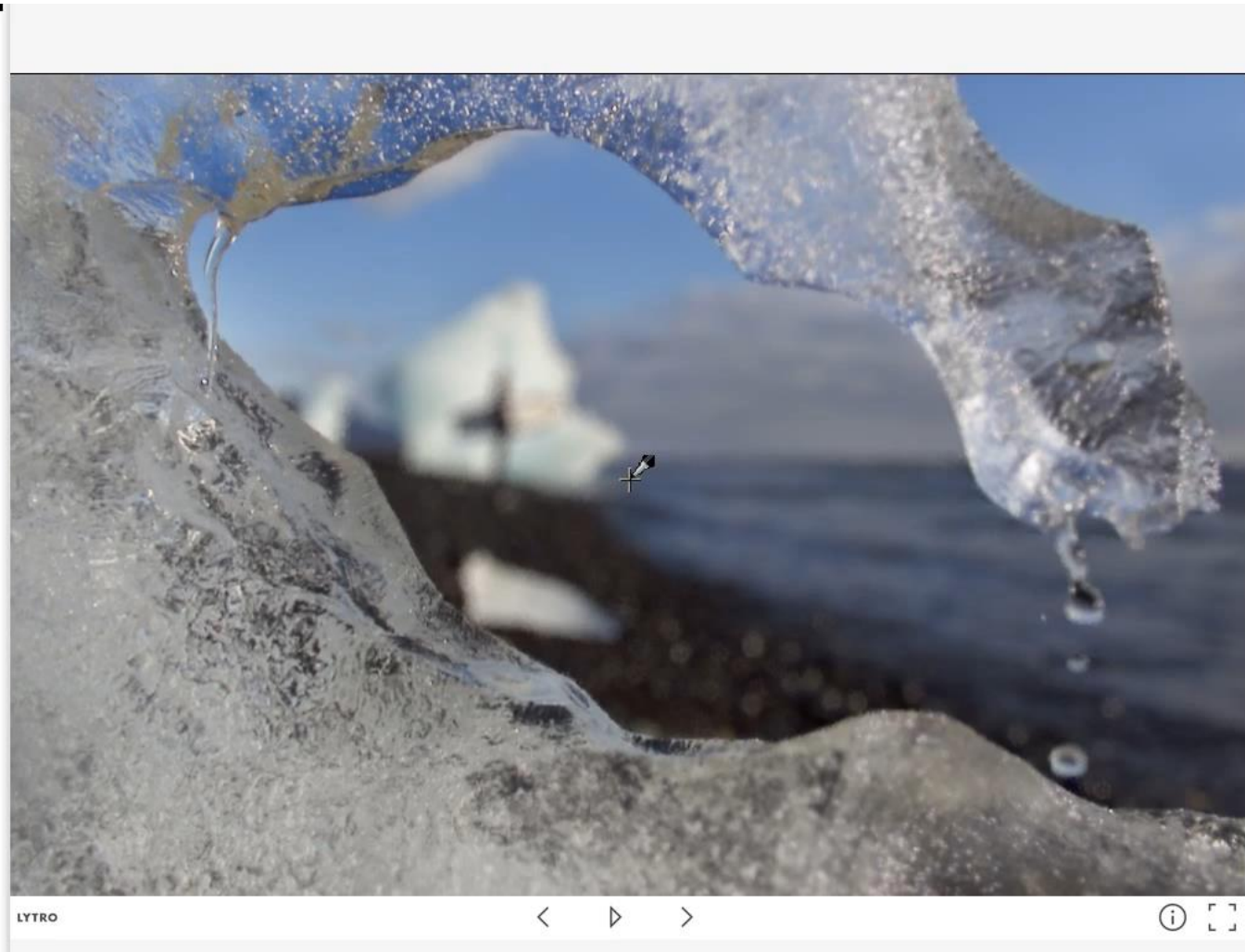


- ▶ But the number of dimensions can be reduced if
 - ▶ The camera stays outside the convex hull of the object
 - ▶ The light travels in uniform medium
 - ▶ Then, radiance L remains the same along the ray (until the ray hits an object)
 - ▶ This way we obtain a **4D light field** or **lumigraph**

4D Light field

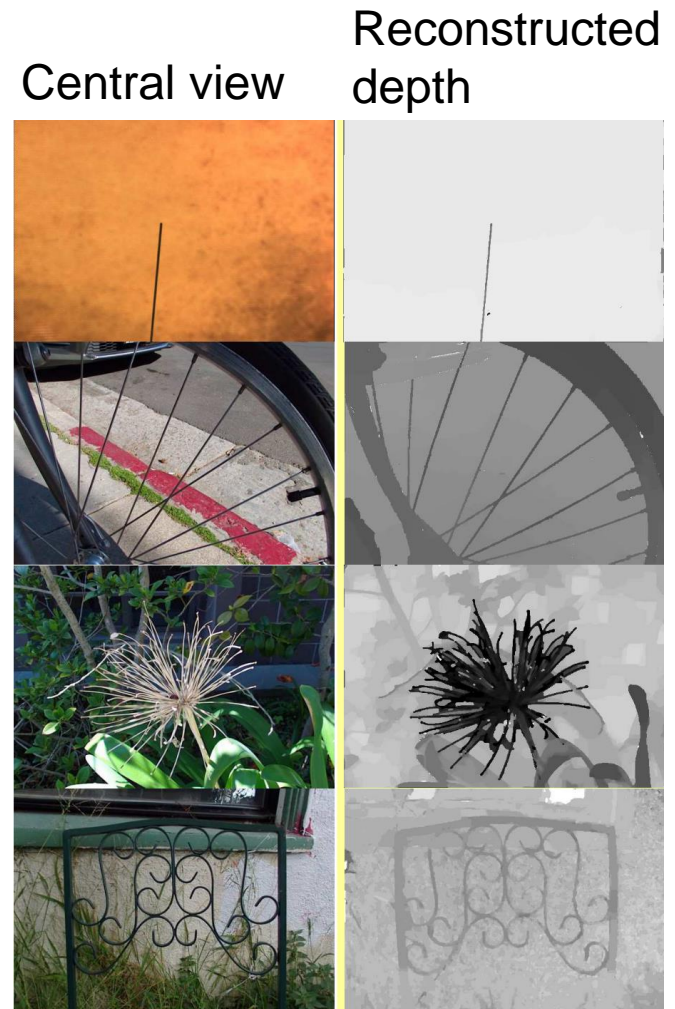


Refocusing and view point adjustment



Depth estimation from light field

- ▶ Passive sensing of depth
- ▶ Light field captures multiple depth cues
 - ▶ Correspondance (disparity) between the views
 - ▶ Defocus
 - ▶ Occlusions



From: *Ting-Chun Wang, Alexei A. Efros, Ravi Ramamoorthi*; The IEEE International Conference on Computer Vision (ICCV), 2015, pp. 3487-3495

Two methods to capture light fields

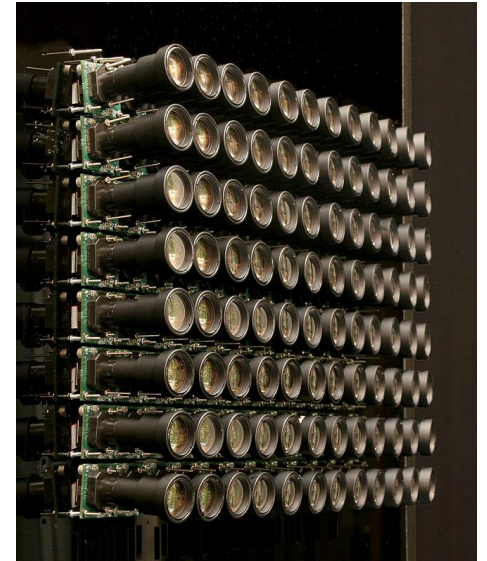
Micro-lens array

- ▶ Small baseline
- ▶ Good for digital refocusing
- ▶ Limited resolution

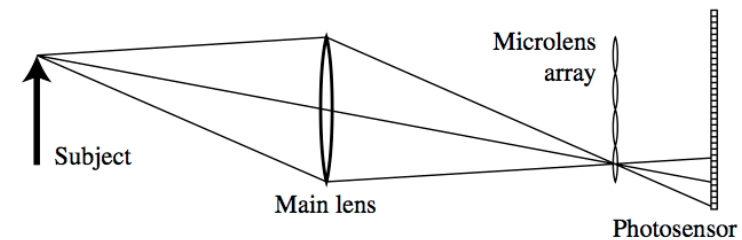
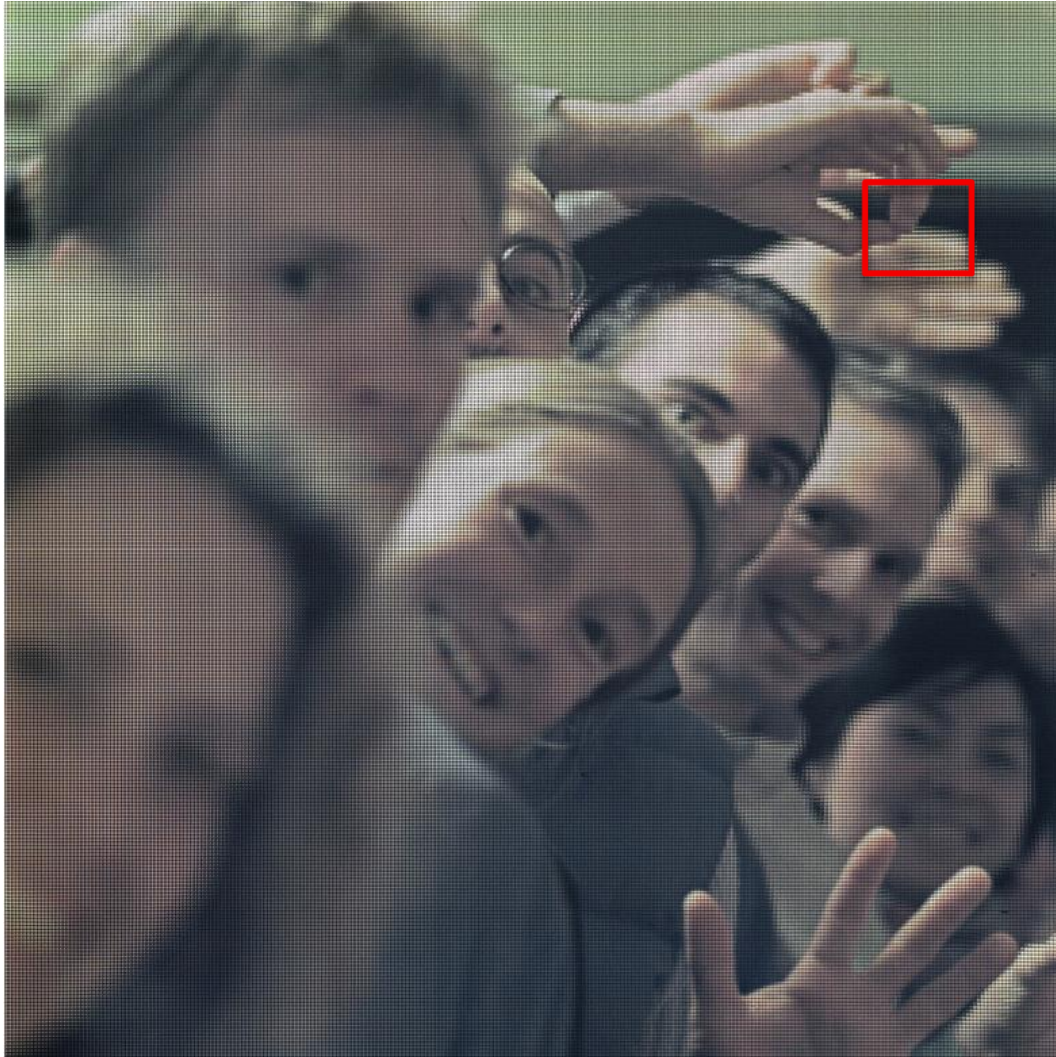


Camera array

- ▶ Large baseline
- ▶ High resolution
- ▶ Rendering often requires approximate depth



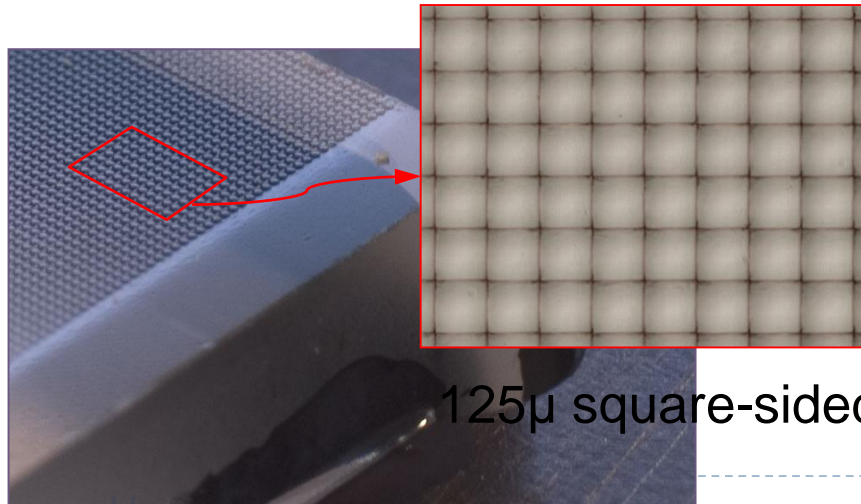
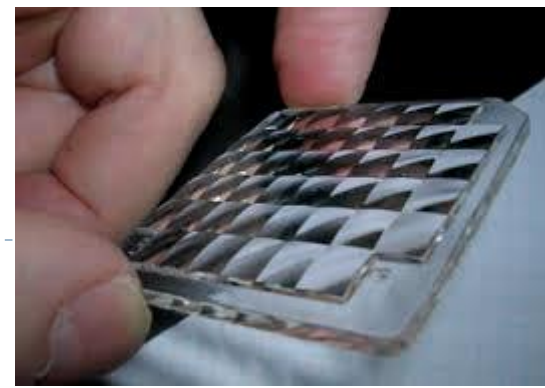
Light field image – with microlens array



Digital Refocusing using Light Field Camera



Lenslet
array



125 μ square-sided microlenses

[Ng et al 2005]

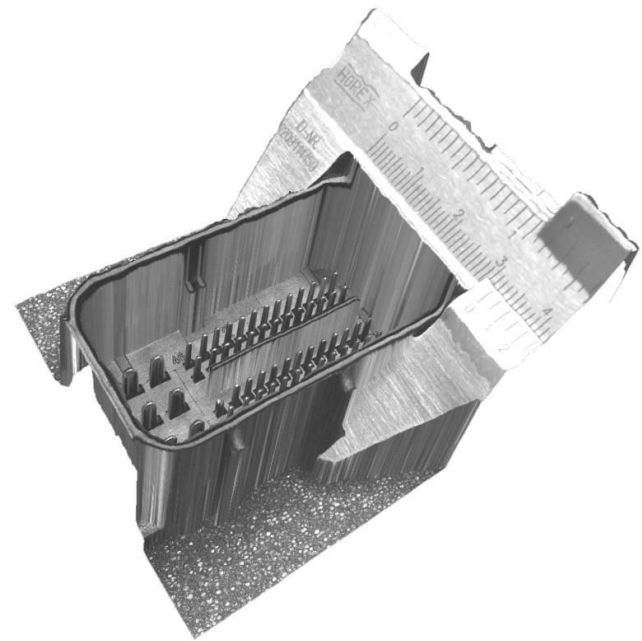
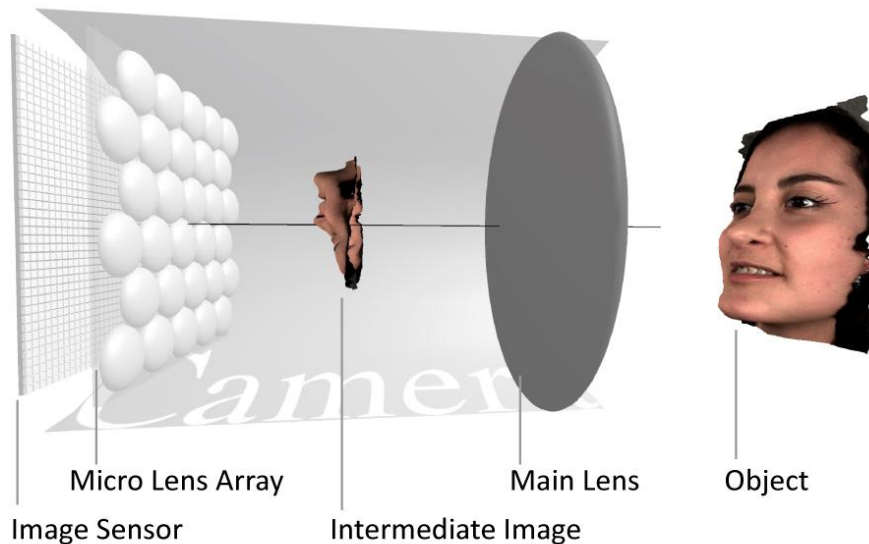
Lytro-cameras

- ▶ First commercial light-field cameras
- ▶ Lytro illum camera
 - ▶ 40 Mega-rays
 - ▶ 2D resolution: 2450 x 1634 (4 MPixels)



Raytrix camera

- ▶ Similar technology to Lytro
- ▶ But profiled for computer vision applications

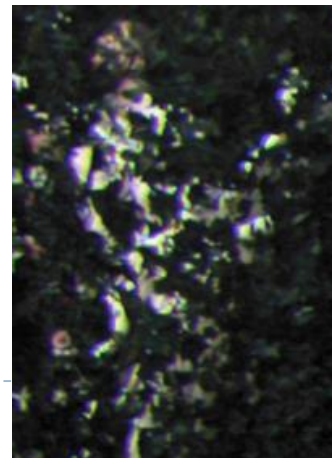


Stanford camera array



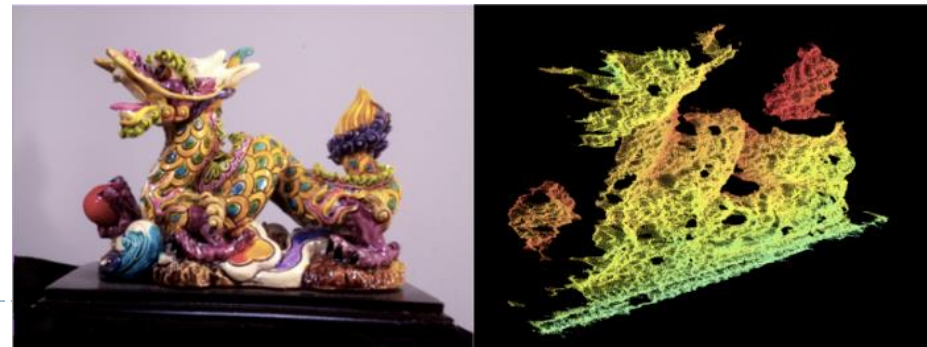
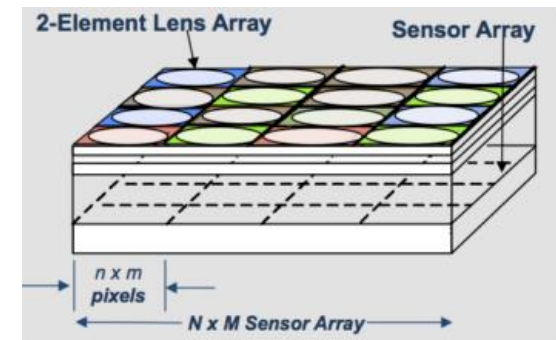
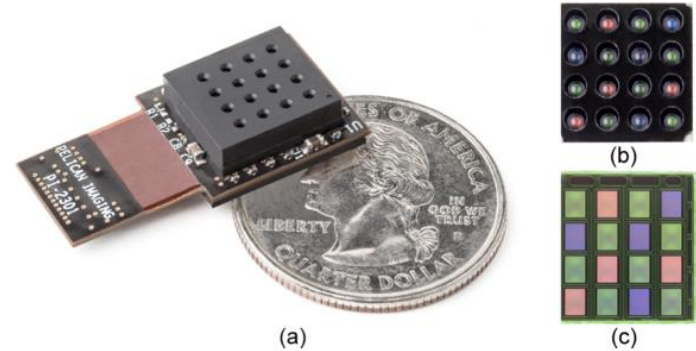
96 cameras

Application: Reconstruction of occluded surfaces

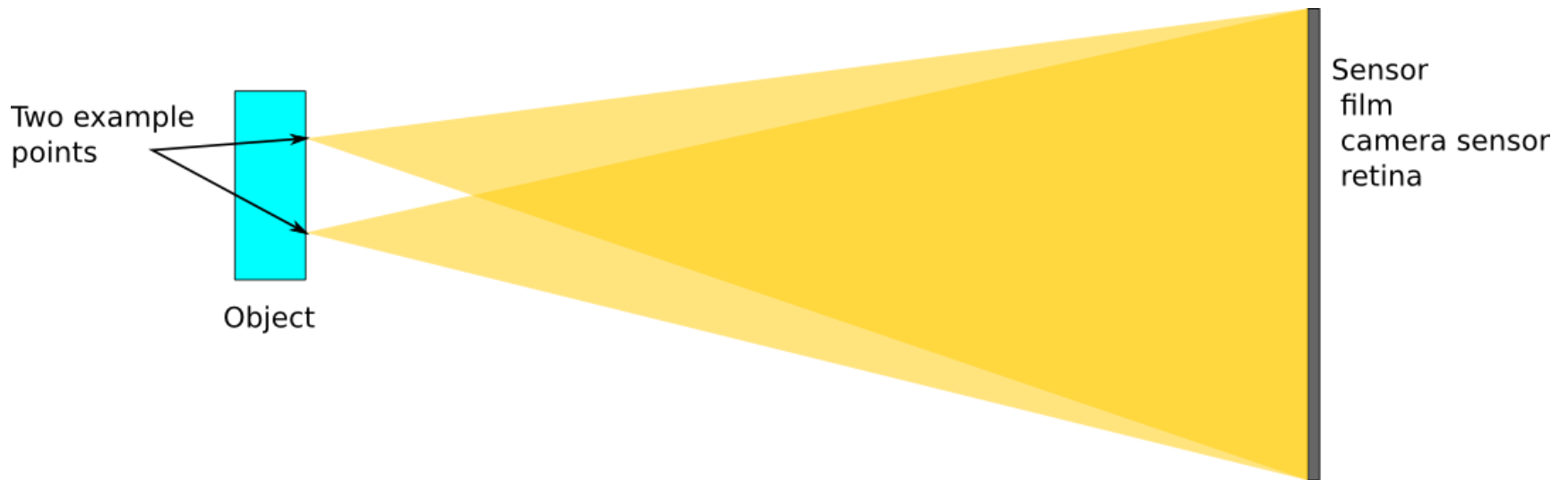


PiCam camera array module

- ▶ Array of 4×4 cameras on a single chip
- ▶ Each camera has its own lens and senses only one spectral colour band
 - ▶ Optics can be optimized for that band
- ▶ The algorithm needs to reconstruct depth

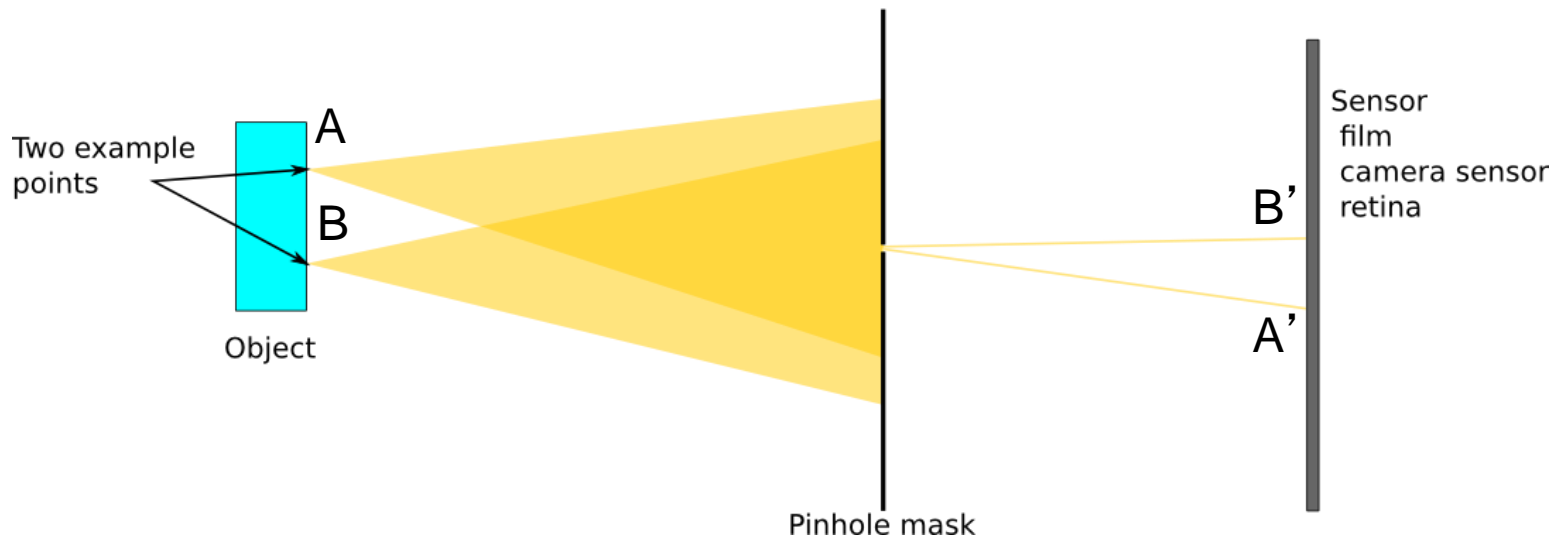


Imaging – without lens



Every point in the scene illuminates every point (pixel) on a sensor. Everything overlaps - no useful image.

Imaging – pinhole camera

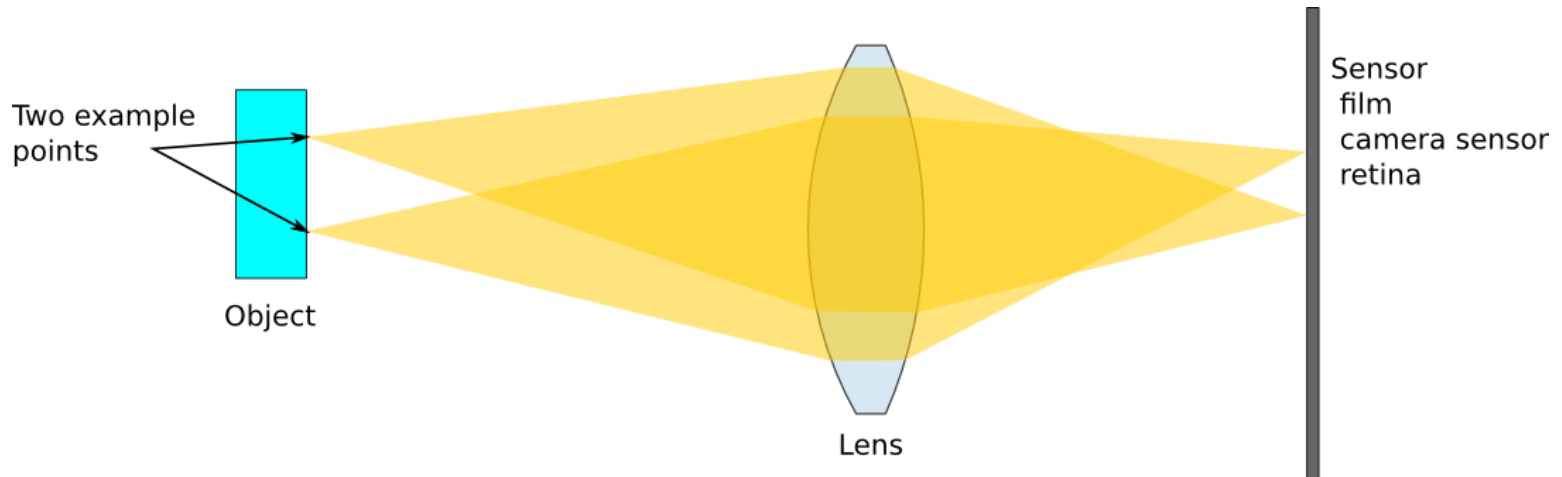


Pinhole masks all but only tiny beams of light. The light from different points is separated and the image is formed.

But very little light reaches the sensor.



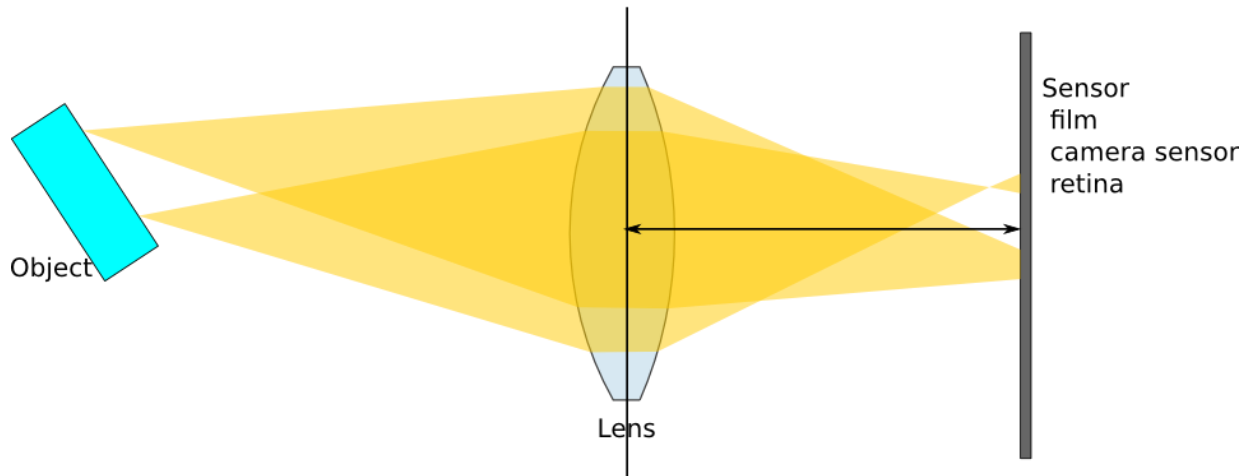
Imaging – lens



Lens can focus a beam of light on a sensor (focal plane).

Much more light-efficient than the pinhole.

Imaging – lens



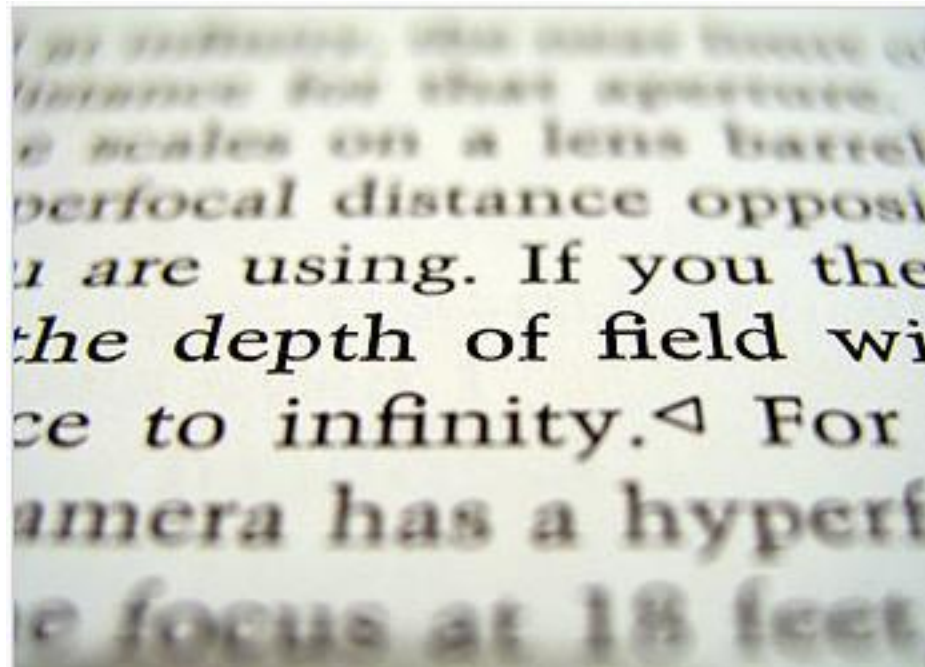
But if the light beams coming from different distances are not focused on the same plane.

These points will appear blurry in the resulting image.

Camera needs to move lens to focus an image on the sensor.

Depth of field

- ▶ Depth of field – range of depths that provides sufficient focus



Defocus blur is often desirable

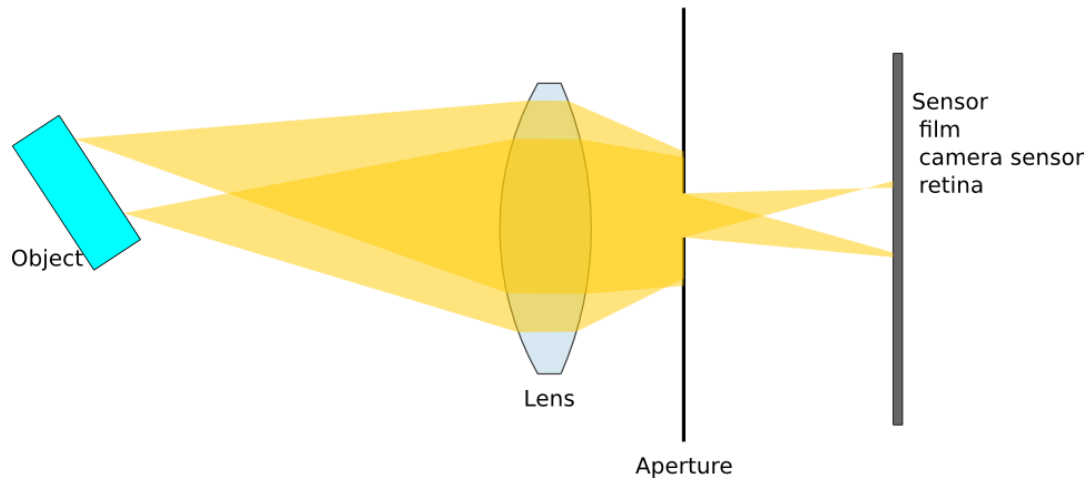


To separate the object of interest from background



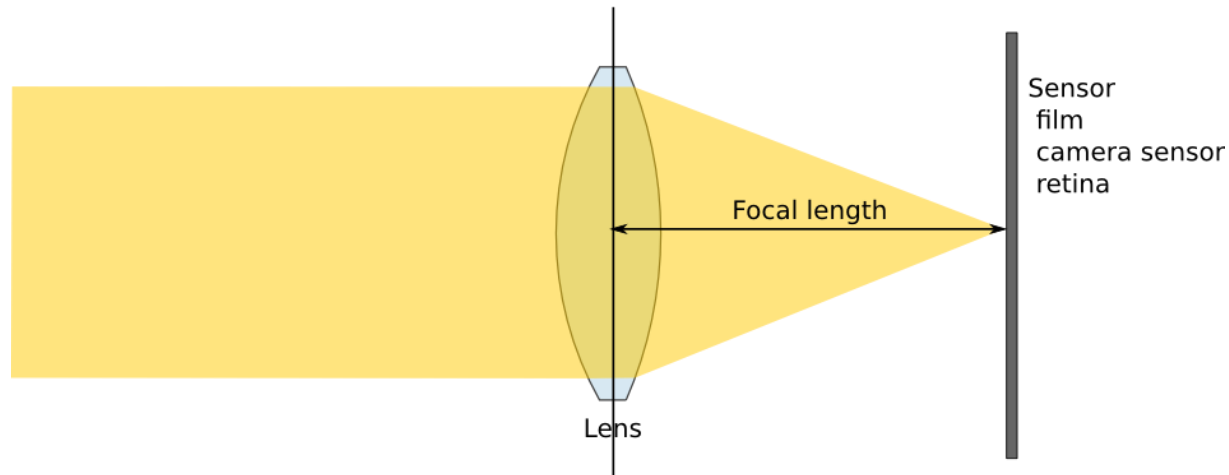
Defocus blur is a strong depth cue

Imaging – aperture



Aperture (introduced behind the lens) reduces the amount of light reaching sensor, but it also reduces blurriness from defocus (increases depth-of-field).

Imaging – lens

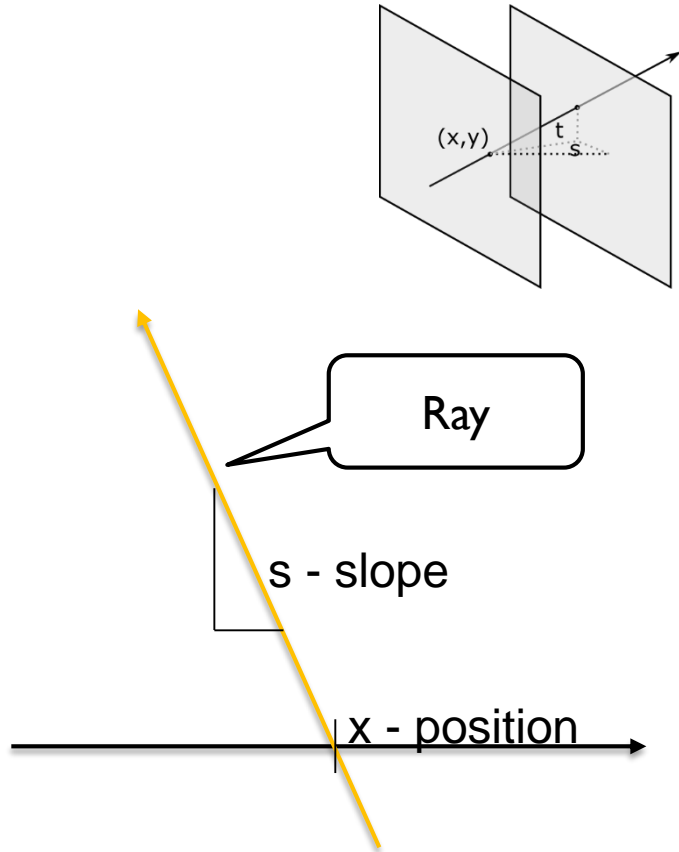


Focal length – length between the sensor and the lens that is needed to focus light coming from an infinite distance.

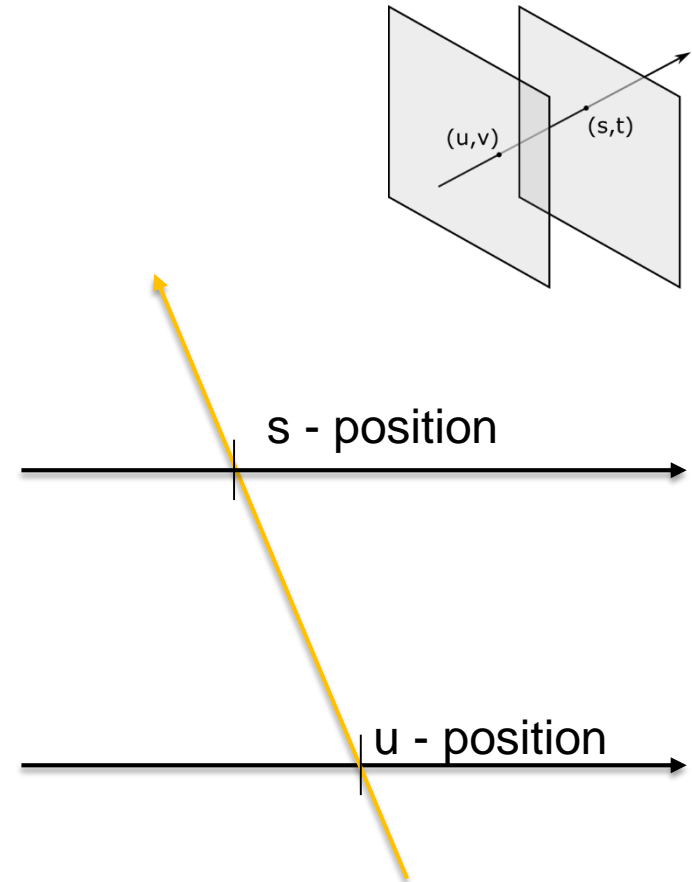
Larger focal length of a lens – more or less magnification?

Light fields: two parametrisations

(shown in 2D)

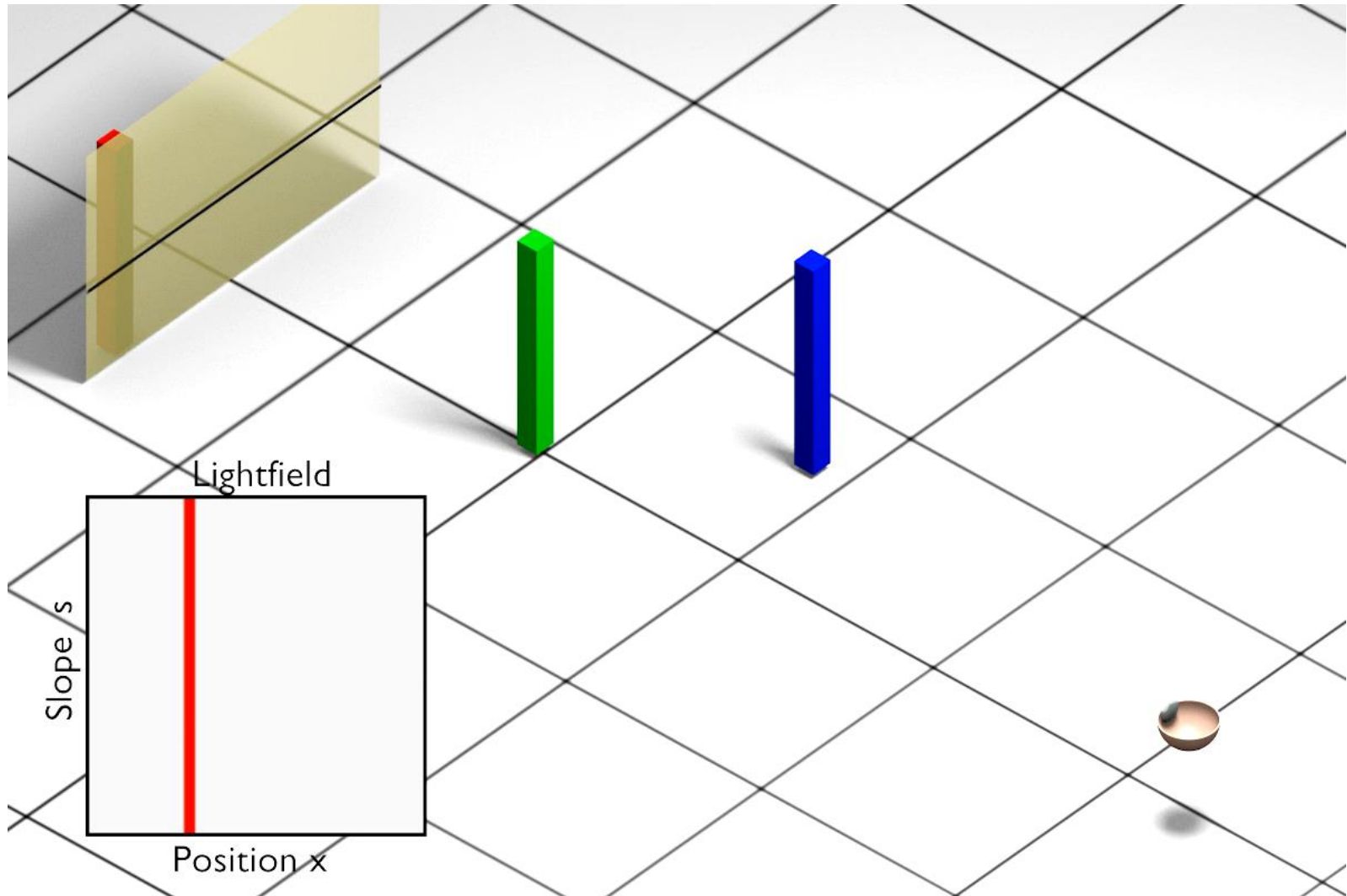


Position and slope
(slope - tangent of the angle)

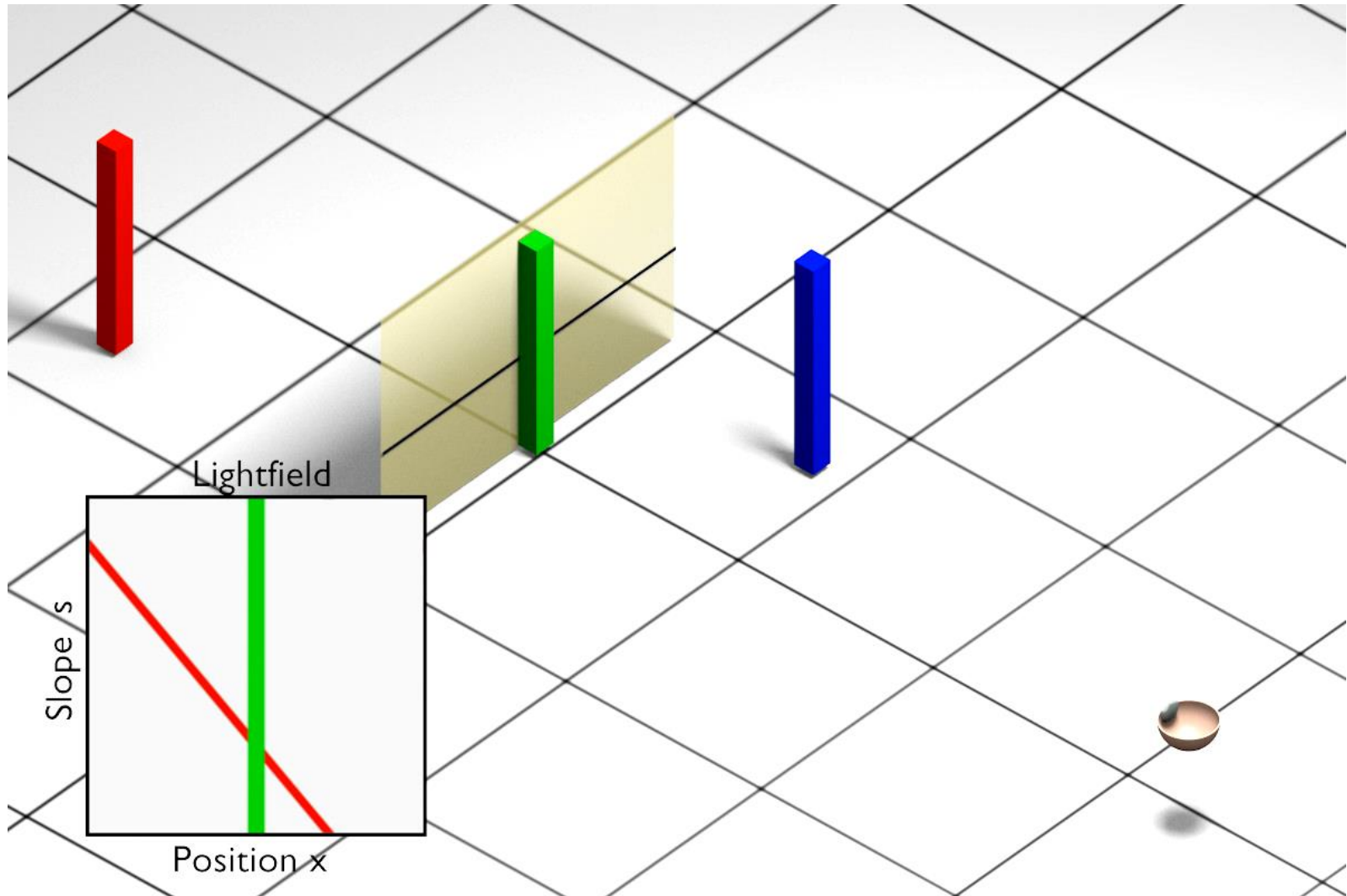


Two planes

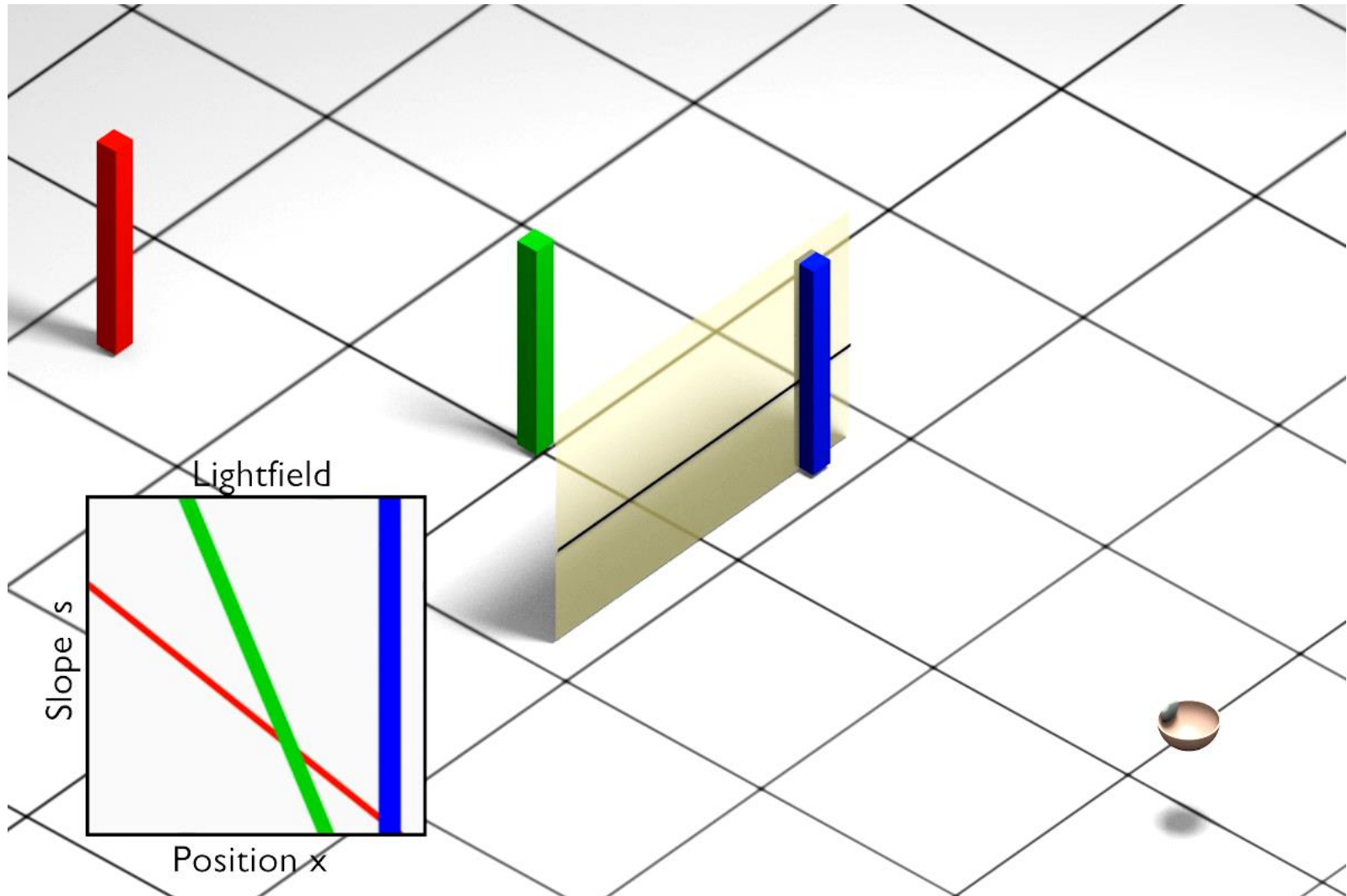
Lightfield - example



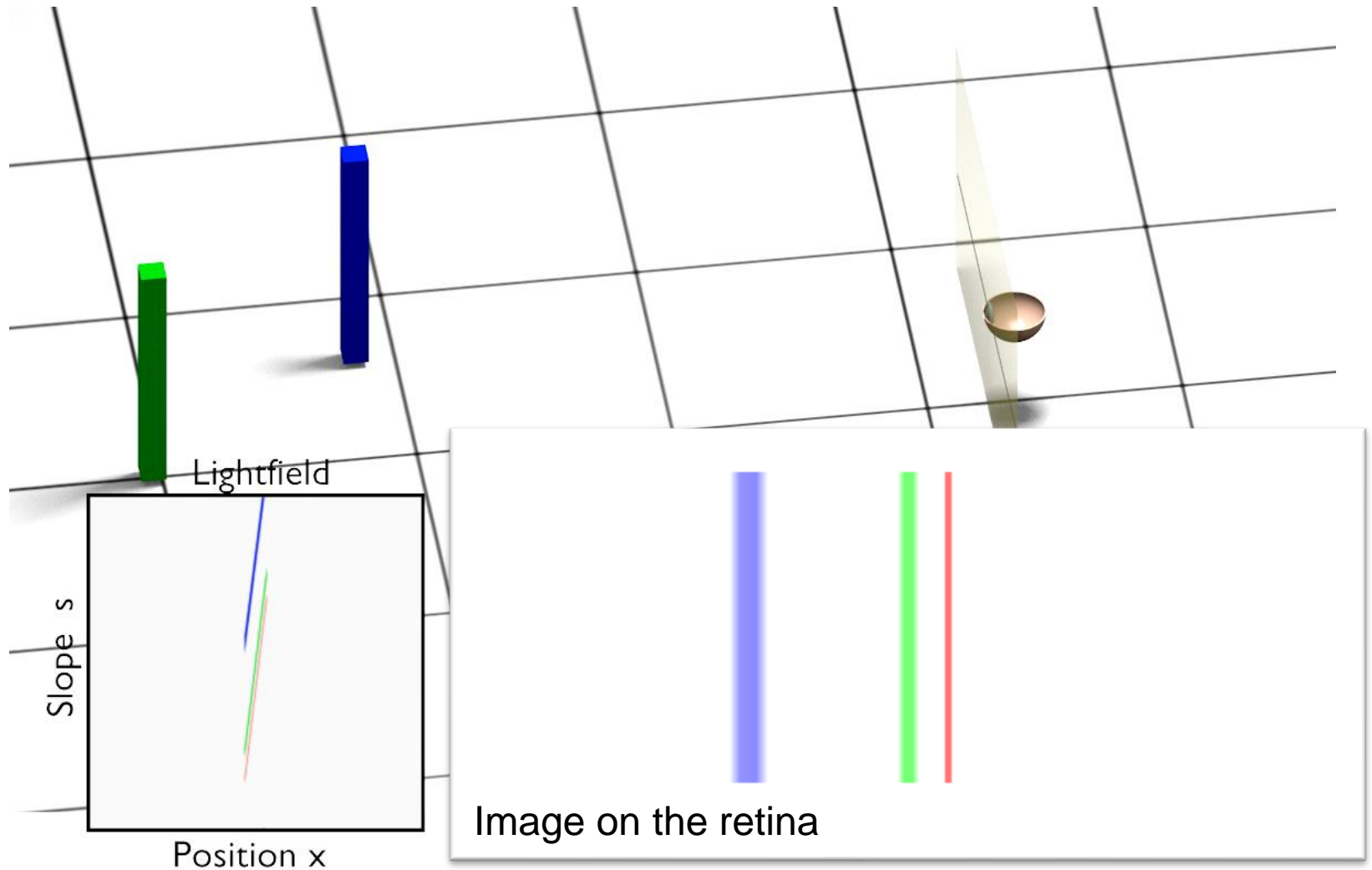
Lightfield - example



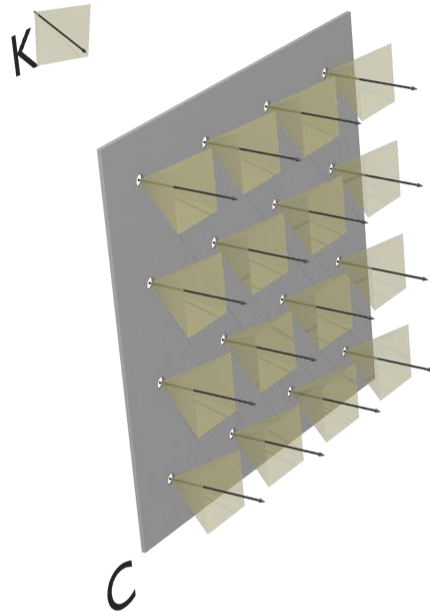
Lightfield - example



Lightfield - example



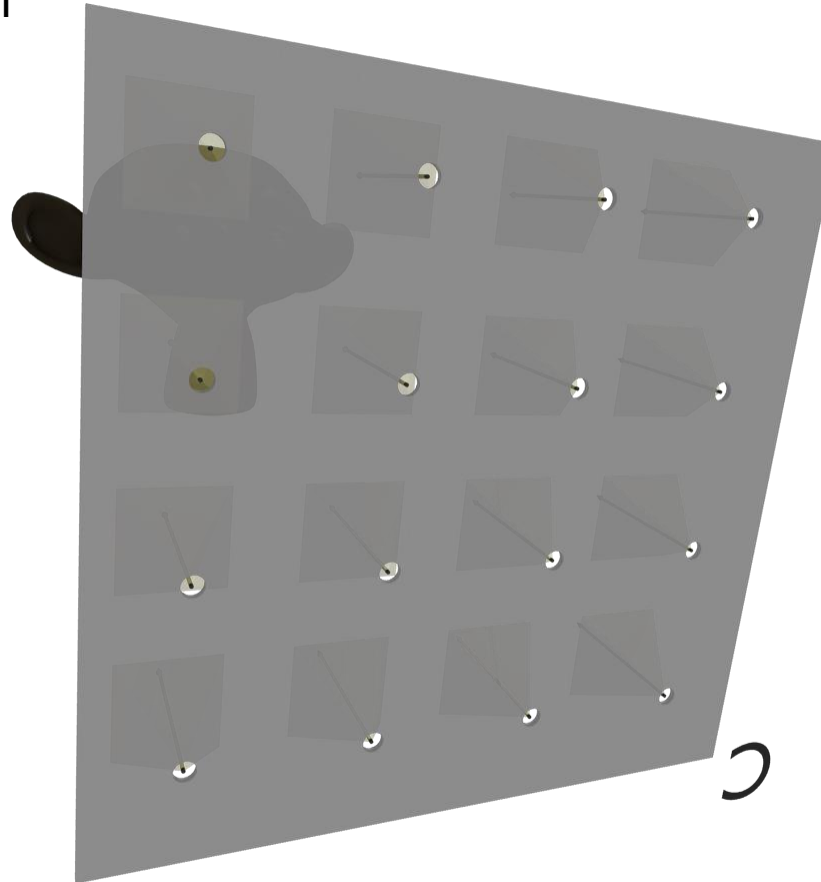
Light field rendering (1 / 3)



We want to render a scene (Blender monkey) as seen by camera K. We have a light field captured by a camera array. Each camera in the array has its aperture on plane C.

Light field rendering (2/3)

From the view point of
camera K

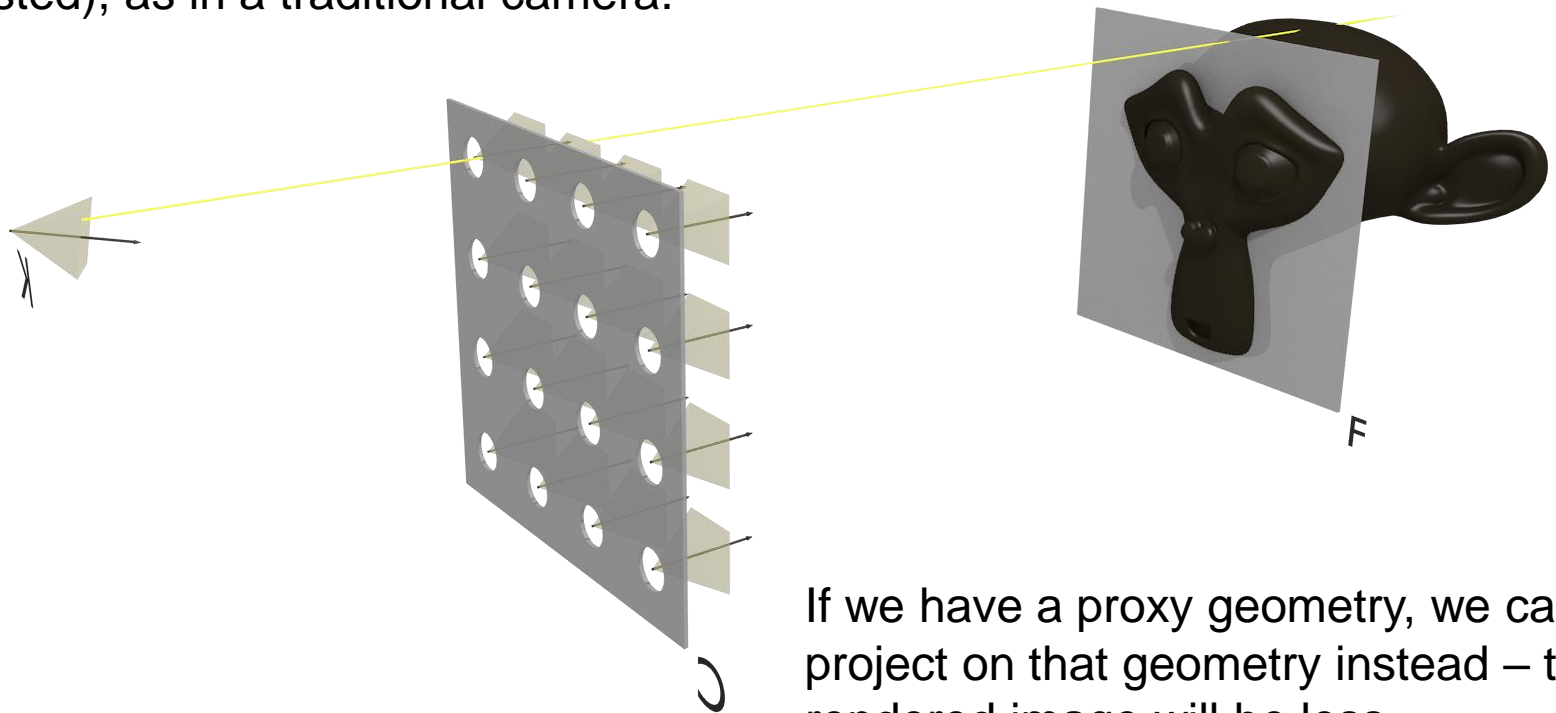


Each camera in the array provides accurate light measurements only for the rays originating from its pinhole aperture.

The missing rays can be either interpolated (reconstructed) or ignored.

Light field rendering (3/3)

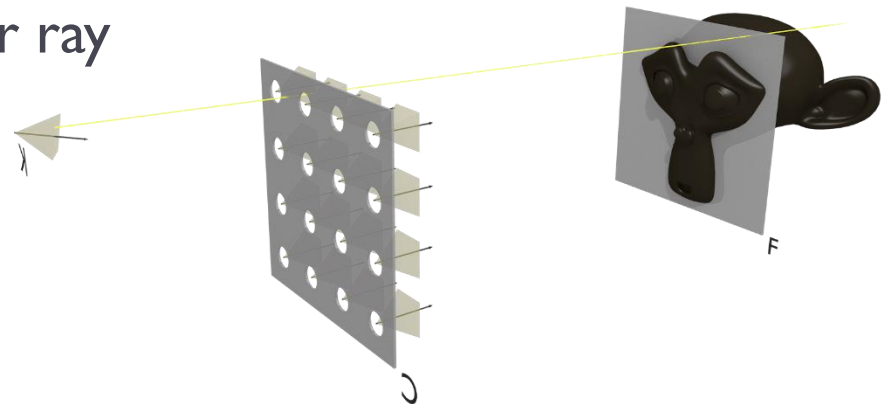
The rays from the camera need to be projected on the focal plane F . The objects on the focal plane will be sharp, and the objects in front or behind that plane will be blurry (ghosted), as in a traditional camera.



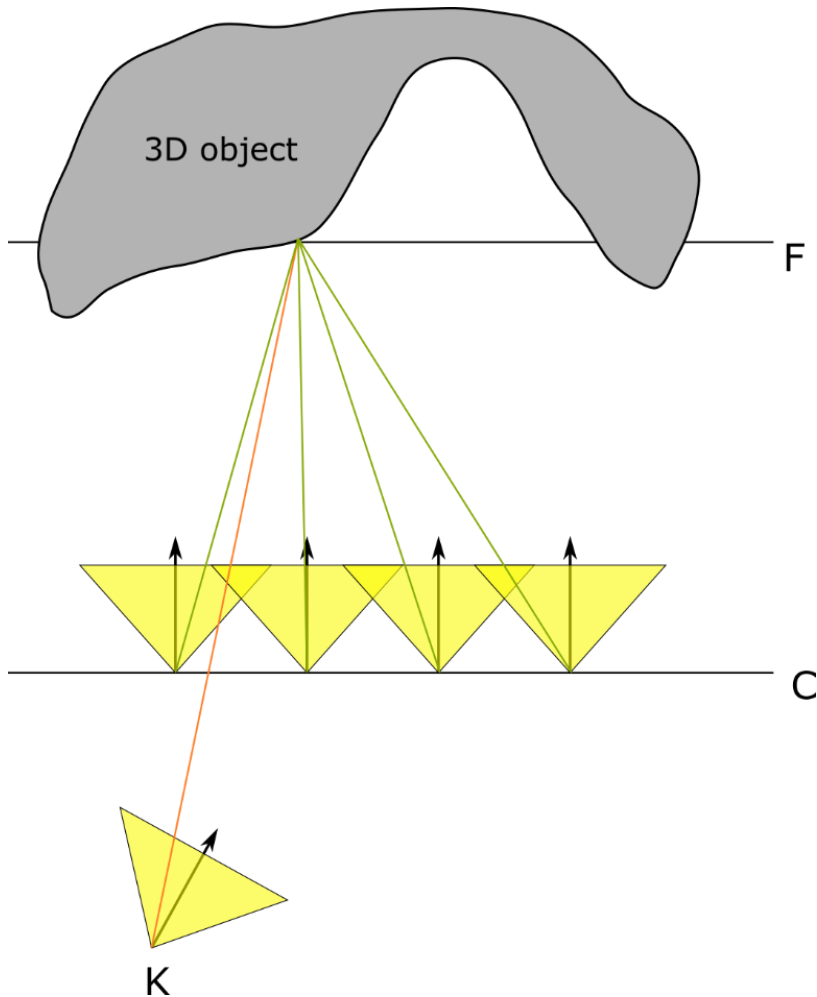
If we have a proxy geometry, we can project on that geometry instead – the rendered image will be less ghosted/blurry

Intuition behind light field rendering

- ▶ For large virtual aperture (use all cameras in the array)
 - ▶ Each camera in the array captures the scene
 - ▶ Then, each camera projects its image on the focal plane F
 - ▶ The virtual camera K captures the projection
- ▶ For small virtual aperture
 - ▶ For each ray from the virtual camera
 - ▶ interpolate rays from 4 nearest camera images
 - ▶ Or use the nearest-neighbour ray

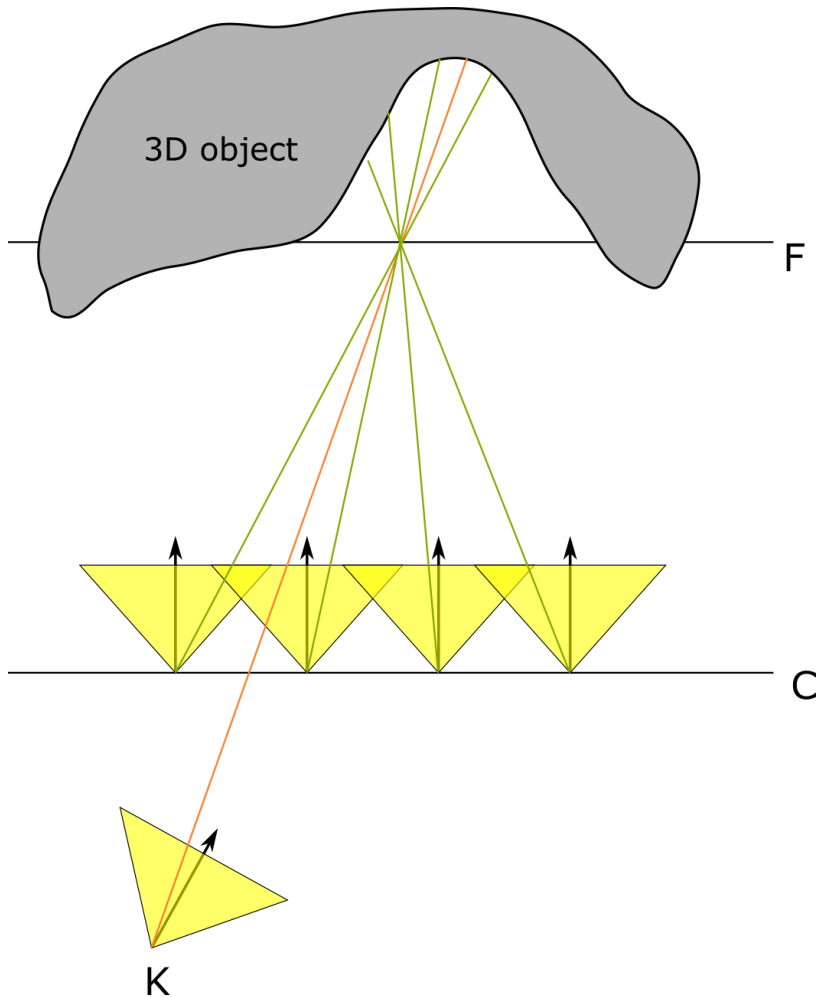


LF rendering – focal plane



- ▶ For a point on the focal plane, all cameras capture the same point on the 3D object
- ▶ They also capture approximately the same colour (for diffuse objects)
- ▶ Averaged colour will be the colour of the point on the surface

LF rendering – focal plane



- ▶ If the 3D object does not lie on the focal plane, all cameras capture different points on the object
- ▶ Averaging colour values will produce a „ghosted” image
- ▶ If we had unlimited number of cameras, this would produce a depth-of-field effect

Finding homographic transformation 1/3

- ▶ For the pixel coordinates p_k of the virtual camera K, we want to find the corresponding coordinates p_i in the camera array image
- ▶ Given the world 3D coordinates of a point w :

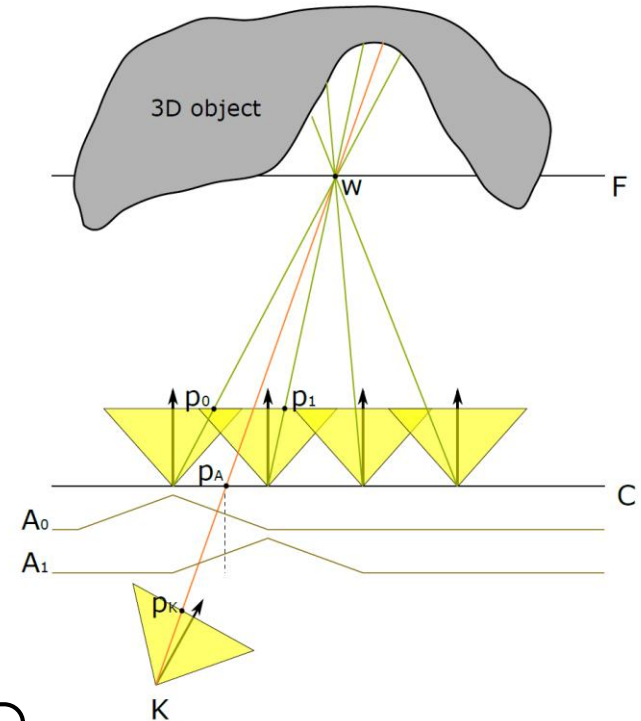
$$p_i = K P V_i w$$

Intrinsic
camera matrix

Projection
matrix

View
matrix

$$\begin{bmatrix} x_i \\ y_i \\ w_i \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} v_{11} & v_{12} & v_{13} & v_{14} \\ v_{21} & v_{22} & v_{23} & v_{24} \\ v_{31} & v_{32} & v_{33} & v_{34} \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$



Finding homographic transformation 2/3

- ▶ A homography between two views is usually found as:

$$\begin{aligned} \mathbf{p}_K &= \mathbf{K}_K \mathbf{P} \mathbf{V}_K \mathbf{w} \\ \mathbf{p}_i &= \mathbf{K}_i \mathbf{P} \mathbf{V}_i \mathbf{w} \end{aligned}$$

hence

$$\mathbf{p}_i = \mathbf{K}_i \mathbf{P} \mathbf{V}_i \mathbf{V}_K^{-1} \mathbf{P}^{-1} \mathbf{K}_K^{-1} \mathbf{p}_K$$

- ▶ But, $\mathbf{K}_K \mathbf{P} \mathbf{V}_K$ is not a square matrix and cannot be inverted
 - ▶ To find the correspondence, we need to constrain 3D coordinates \mathbf{w} to lie on the plane:

$$\mathbf{N} \cdot (\mathbf{w} - \mathbf{w}_F) = 0 \quad \text{or} \quad d = \begin{bmatrix} n_x & n_y & n_z & -\mathbf{N} \cdot \mathbf{w}_F \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

Finding homographic

The plane in
the camera coordinates
(not world coordinates)

n 3/3

- ▶ Then, we add the plane equation to the projection matrix

$$\begin{bmatrix} x_i \\ y_i \\ d_i \\ w_i \end{bmatrix} = \begin{bmatrix} f_x & 0 & 0 & c_x \\ 0 & f_y & 0 & c_y \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ n_x^{(c)} & n_y^{(c)} & n_z^{(c)} & -N^{(c)} \cdot w_F^{(c)} \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} v_{11} & v_{12} & v_{13} & v_{14} \\ v_{21} & v_{22} & v_{23} & v_{24} \\ v_{31} & v_{32} & v_{33} & v_{34} \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

\hat{p}_i
 \hat{K}_i
 \hat{P}
 V_i
 w

- ▶ Where d_i is the distance to the plane (set to 0)
- ▶ Hence

$$\hat{p}_i = \hat{K}_i \hat{P} V_i V_K^{-1} \hat{P}^{-1} \hat{K}_K^{-1} p_K$$

References

▶ Light fields

▶ Micro-lens array

- ▶ Ng, Ren and Levoy, Marc and Bredif, M. and D., & Gene and Horowitz, Mark and Hanrahan, P. (2005). *Light field photography with a hand-held plenoptic camera*.

▶ Camera array

- ▶ OVERBECK, R.S., ERICKSON, D., EVANGELAKOS, D., PHARR, M., AND DEBEVEC, P. 2018. A system for acquiring, processing, and rendering panoramic light field stills for virtual reality. *ACM Transactions on Graphics* 37, 6, 1–15.
- ▶ ISAKSEN, A., MCMILLAN, L., AND GORTLER, S.J. 2000. Dynamically reparameterized light fields. *Proc of SIGGRAPH '00*, ACM Press, 297–306.