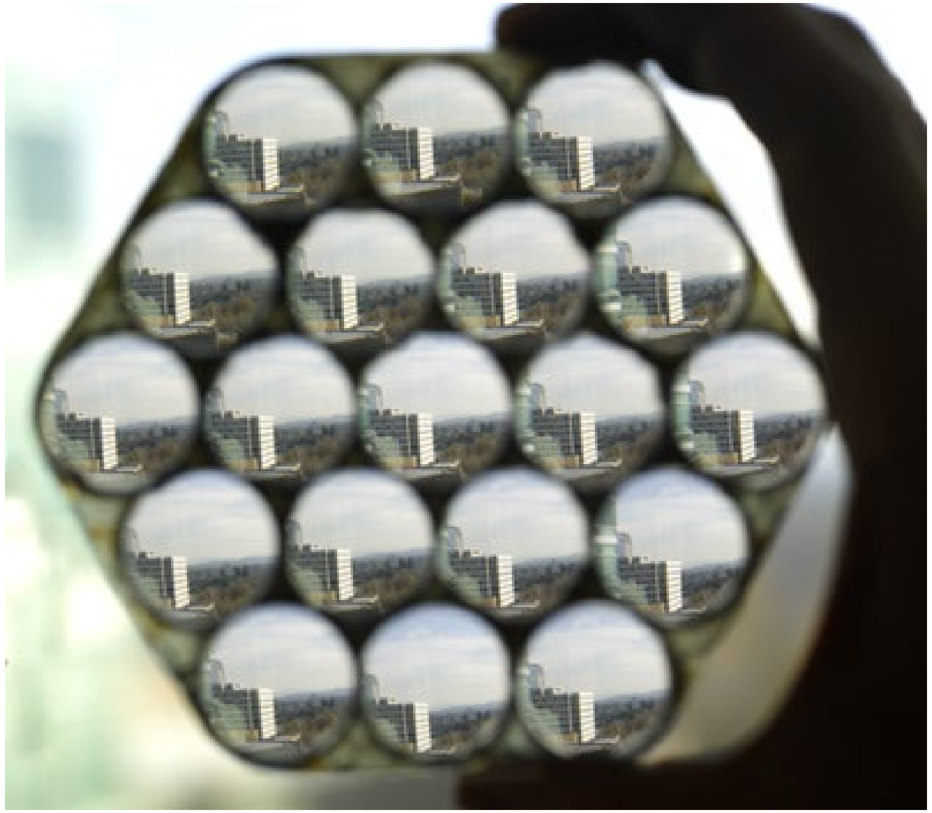
Focal stacks and lightfields



15-463, 15-663, 15-862
Computational Photography
Fall 2023, Lecture 10

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

Course announcements

- Homework assignment 4 is out.
  - Due November 3rd
  - Any questions?
- Mid-semester grades will be posted by 4 pm today.
- Competition results for homework assignment 2 posted.

Overview of today's lecture

- Quick reminder: pinhole vs lens cameras.
- Focal stack.
- Confocal stereo.
- Lightfield.
- Measuring lightfields.
- Plenoptic camera.
- Images from lightfields.
- Some notes on (auto-)focusing.

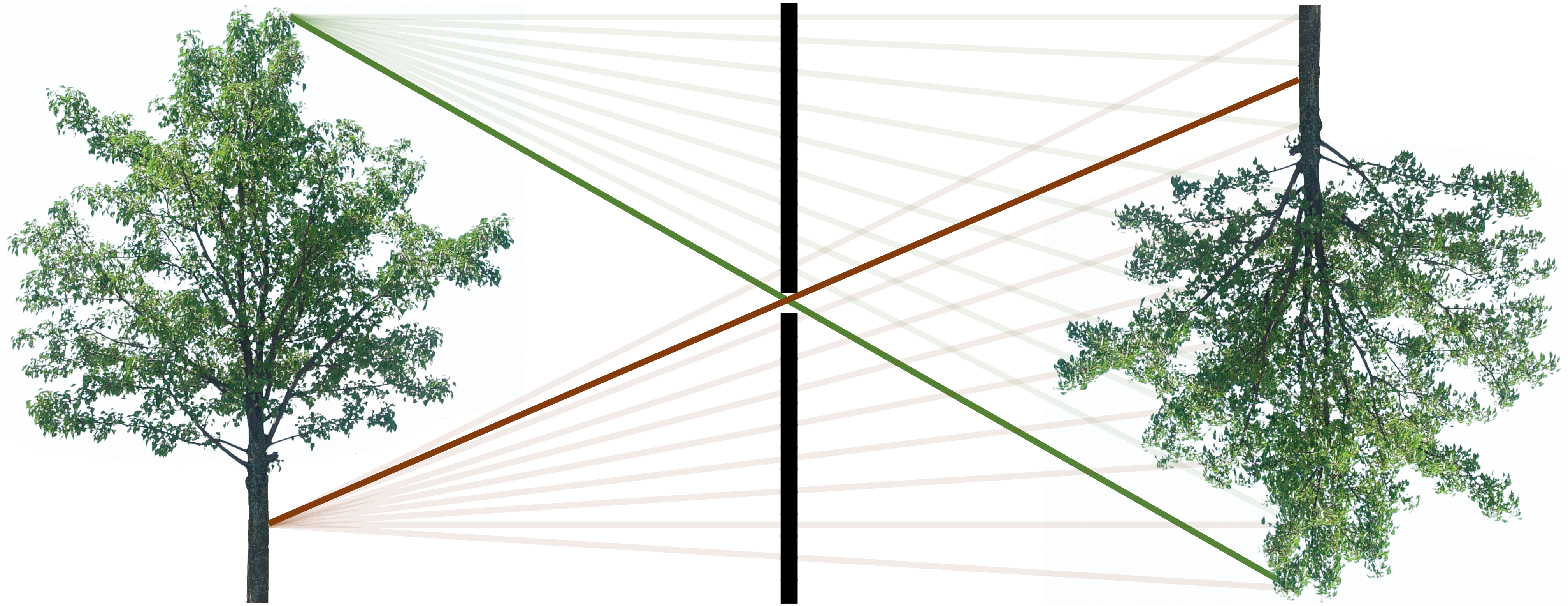
Slide credits

Many of these slides were adapted from:

- Fredo Durand (MIT).
- Gordon Wetzstein (Stanford).
- Marc Levoy (Stanford).

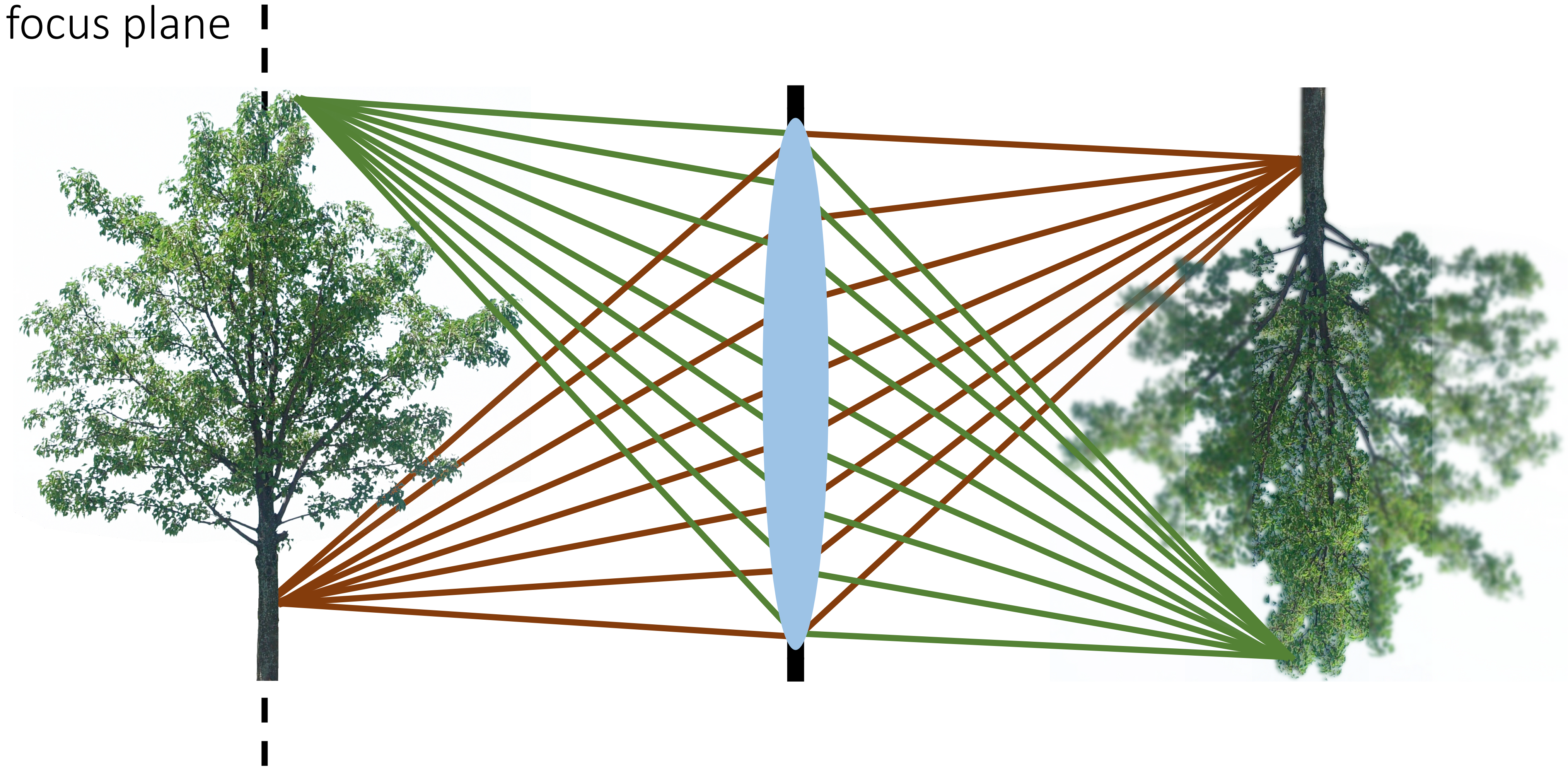
Pinhole vs lens cameras

Pinhole camera



- Everything is in focus.
- Very light inefficient.

Lens camera

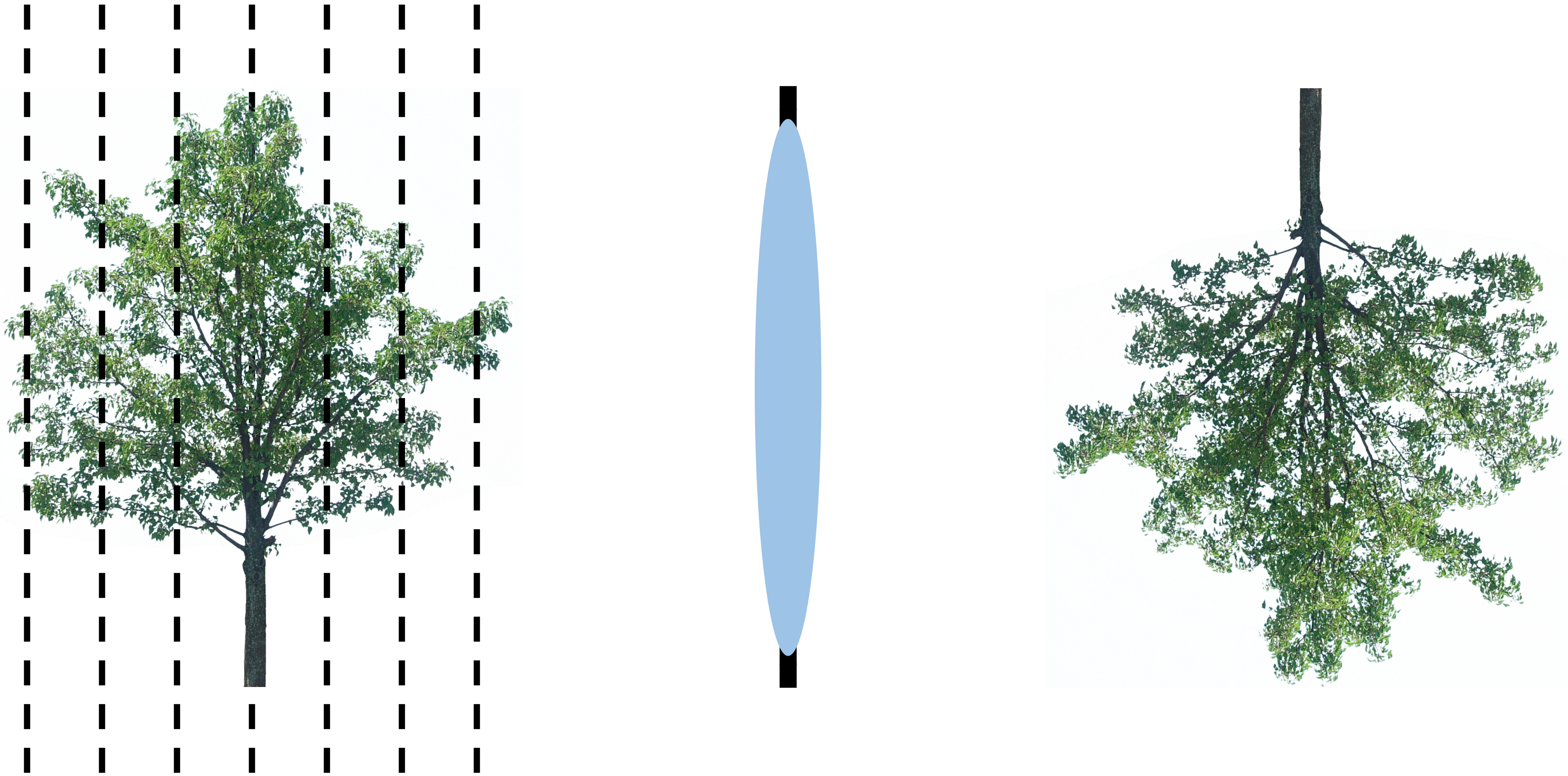


- Only one plane is in focus.
- Very light efficient.

How can we get an all in-focus image?

Focal stack

Focal stack



- Capture images focused at multiple planes.
- Merge them into a single all in-focus image.

Analogous to what we did in HDR

- Focal stack instead of exposure stack.

Homework 4: focal stack imaging

Focal stack imaging

1. Capture a focal stack

2. Merge into an all in-focus image

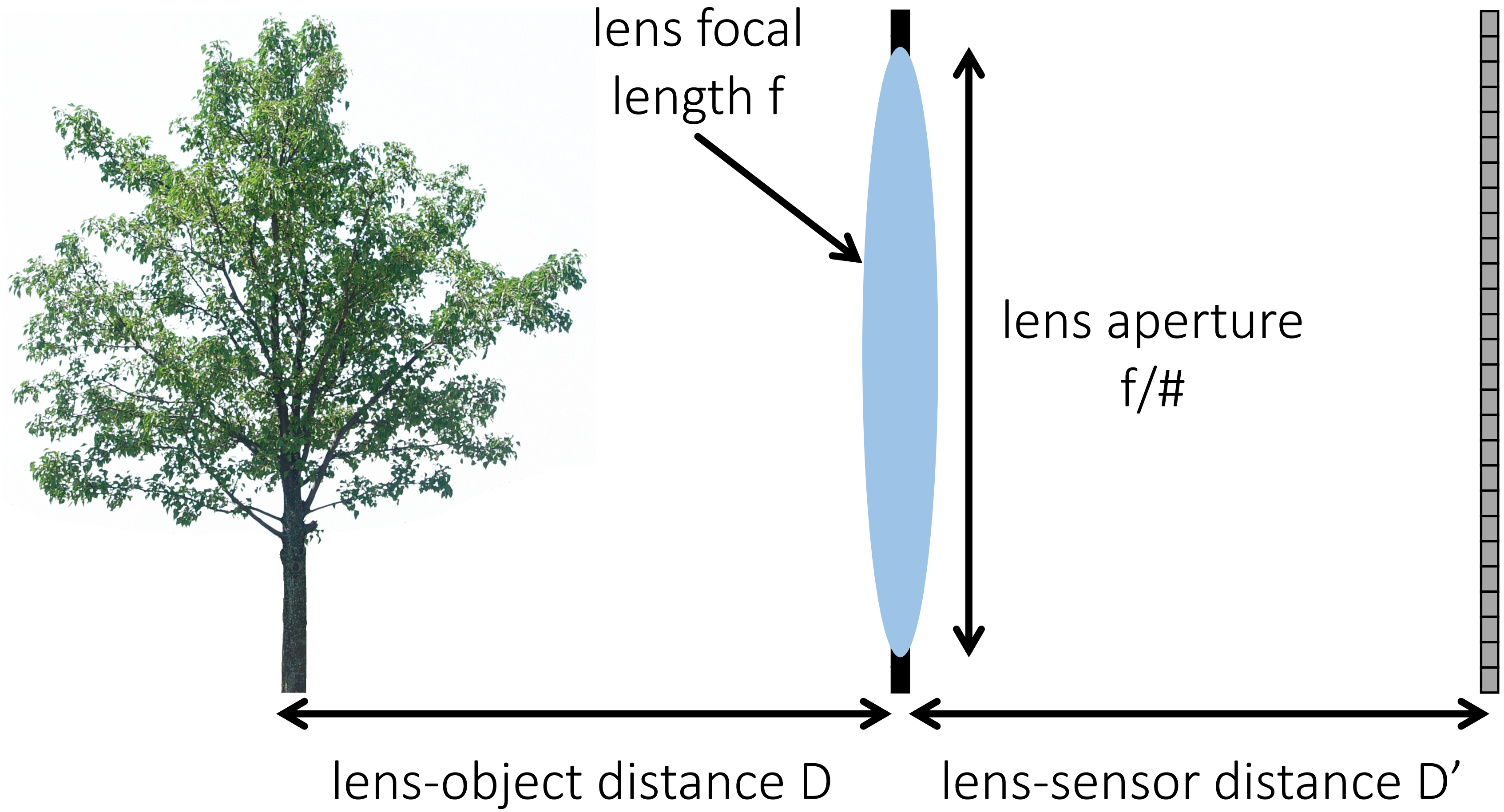
Focal stack imaging

1. Capture a focal stack



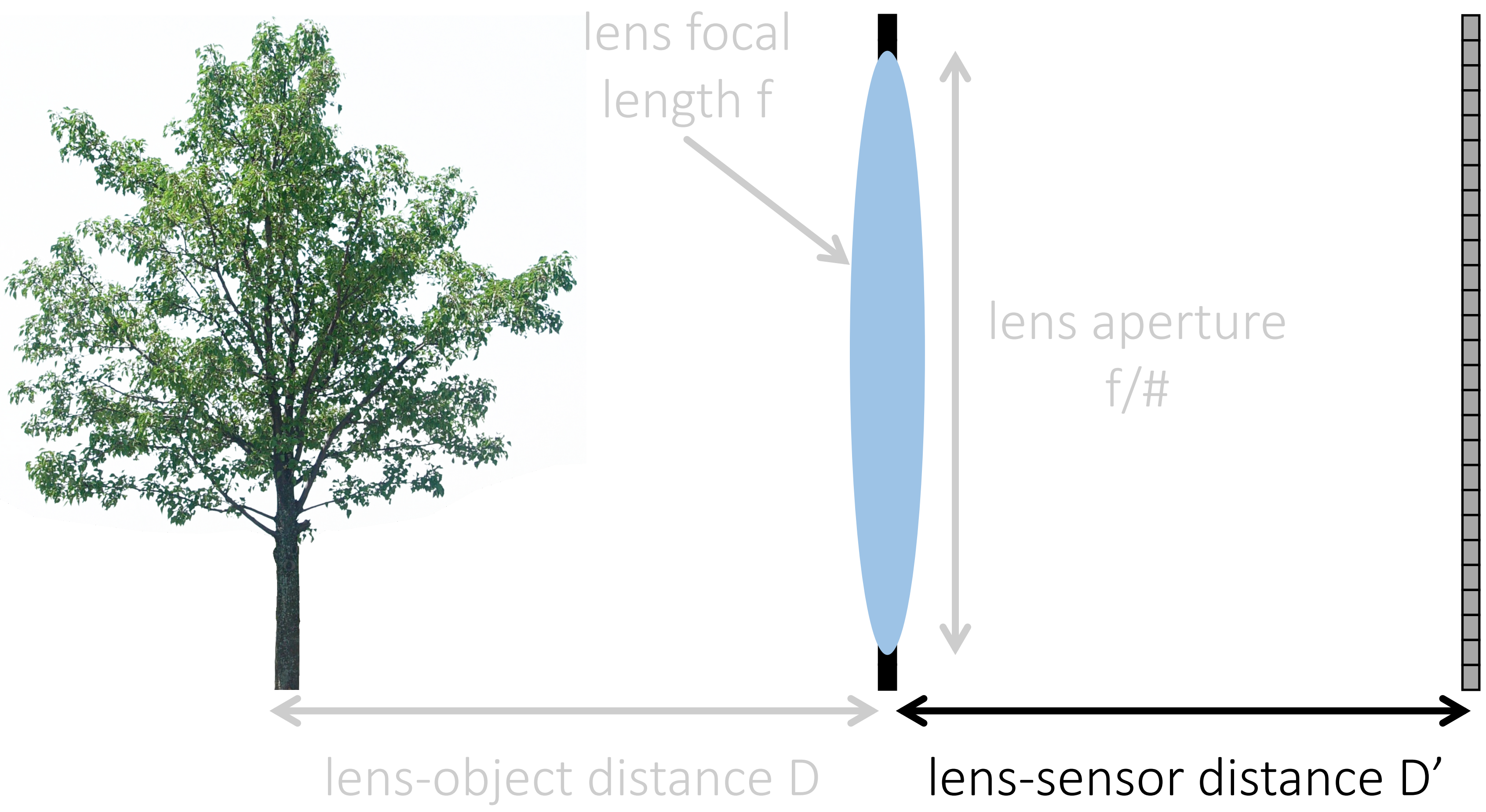
2. Merge into an all in-focus image

How do you capture a focal stack?



Which of these parameters would you change (and how)?

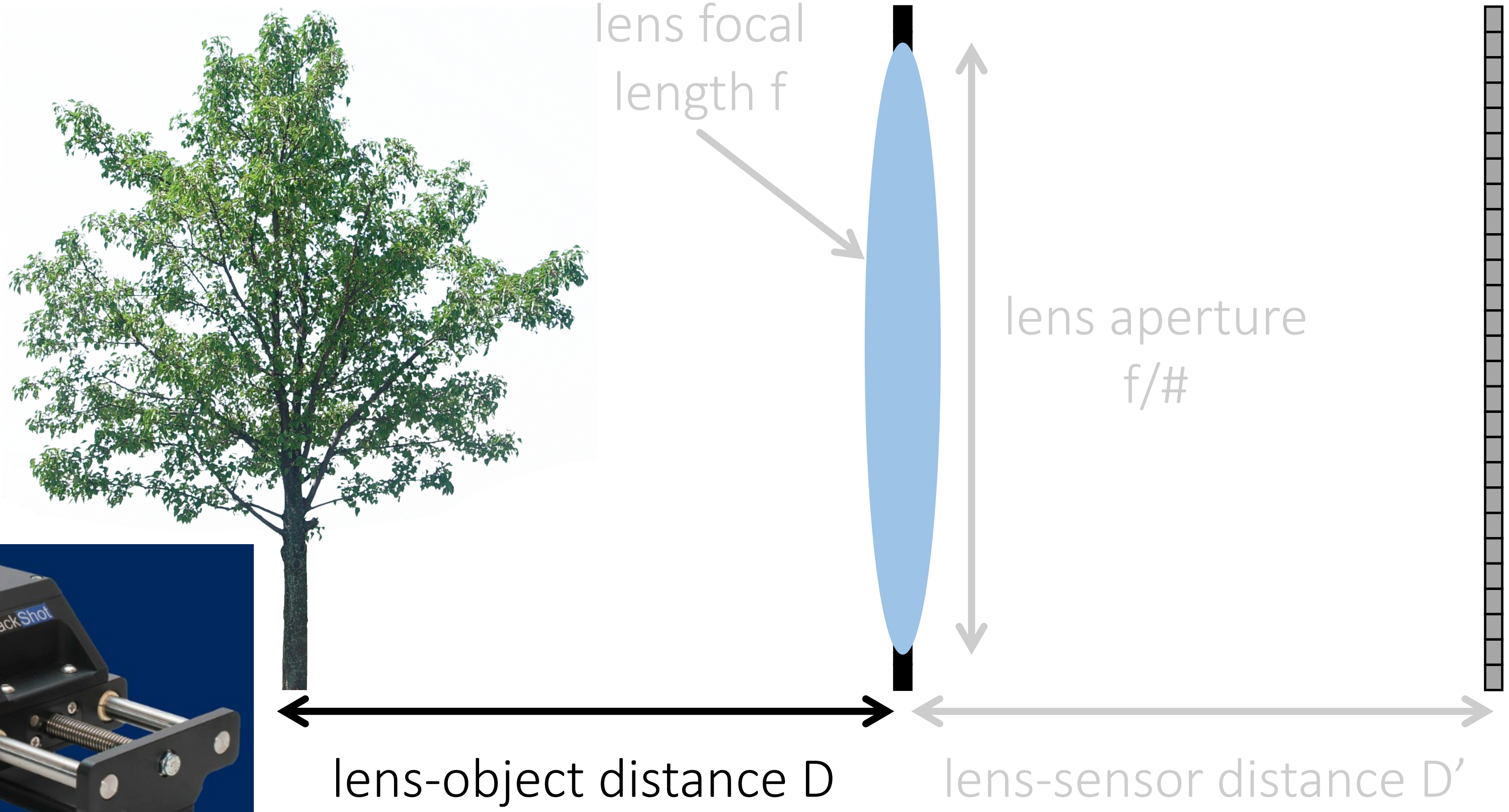
How do you capture a focal stack?



rotate lens focus ring (not zoom!)

Which of these parameters would you change (and how would you achieve that)?

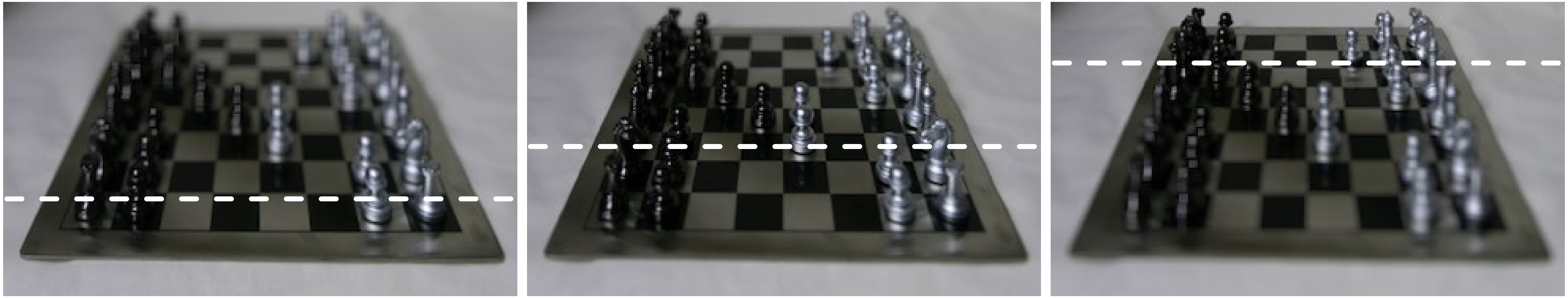
How do you capture a focal stack?



Which of these parameters would you change (and how would you achieve that)?

Capturing a focal stack

In-focus plane in each stack image

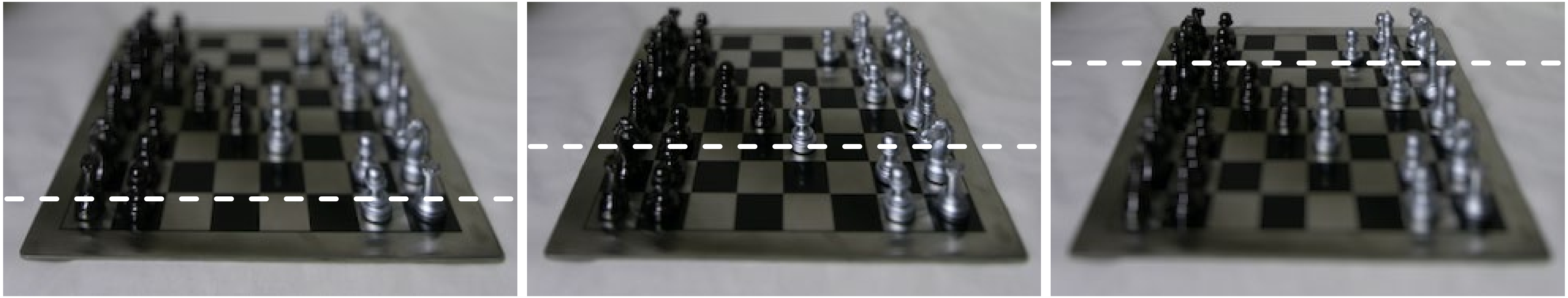


Focal stack imaging

1. Capture a focal stack

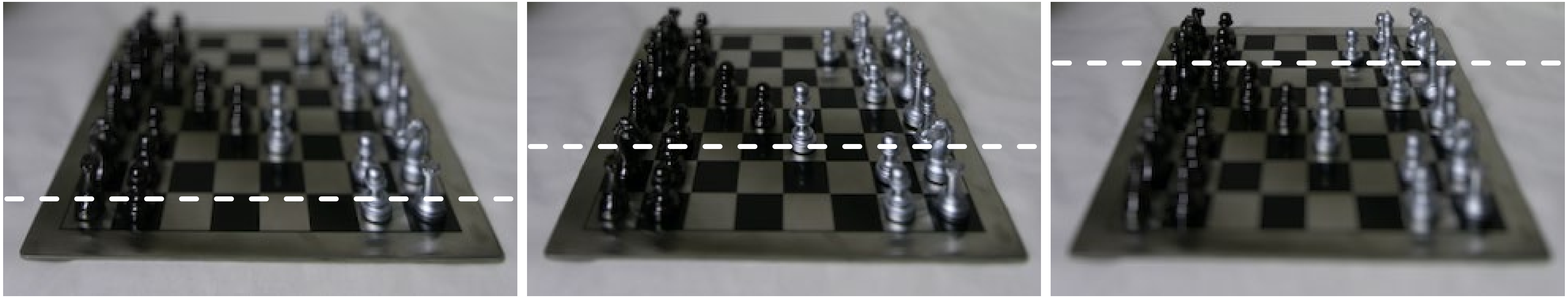
2. Merge into an all in-focus image

How do you merge a focal stack?



How do you merge a focal stack?

1. Align images
2. Assign per-pixel weights representing “in-focus”-ness
3. Compute image average



How do you merge a focal stack?

1. Align images
2. Assign per-pixel weights representing “in-focus”-ness
3. Compute image average

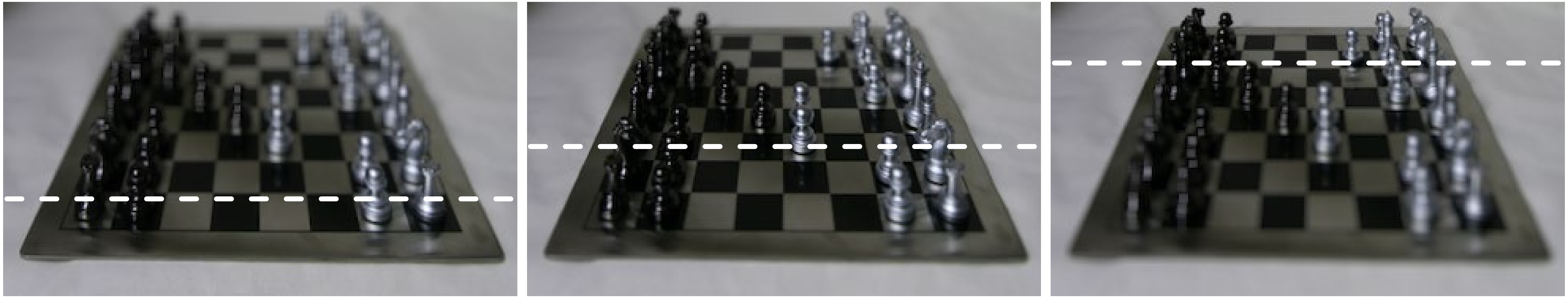


Image alignment

Why do we need to align the images?

Image alignment

Why do we need to align the images?

- When we change focus distance, we also change field of view (magnification).
- Also, scene may not be static (but we will be ignoring this for now).

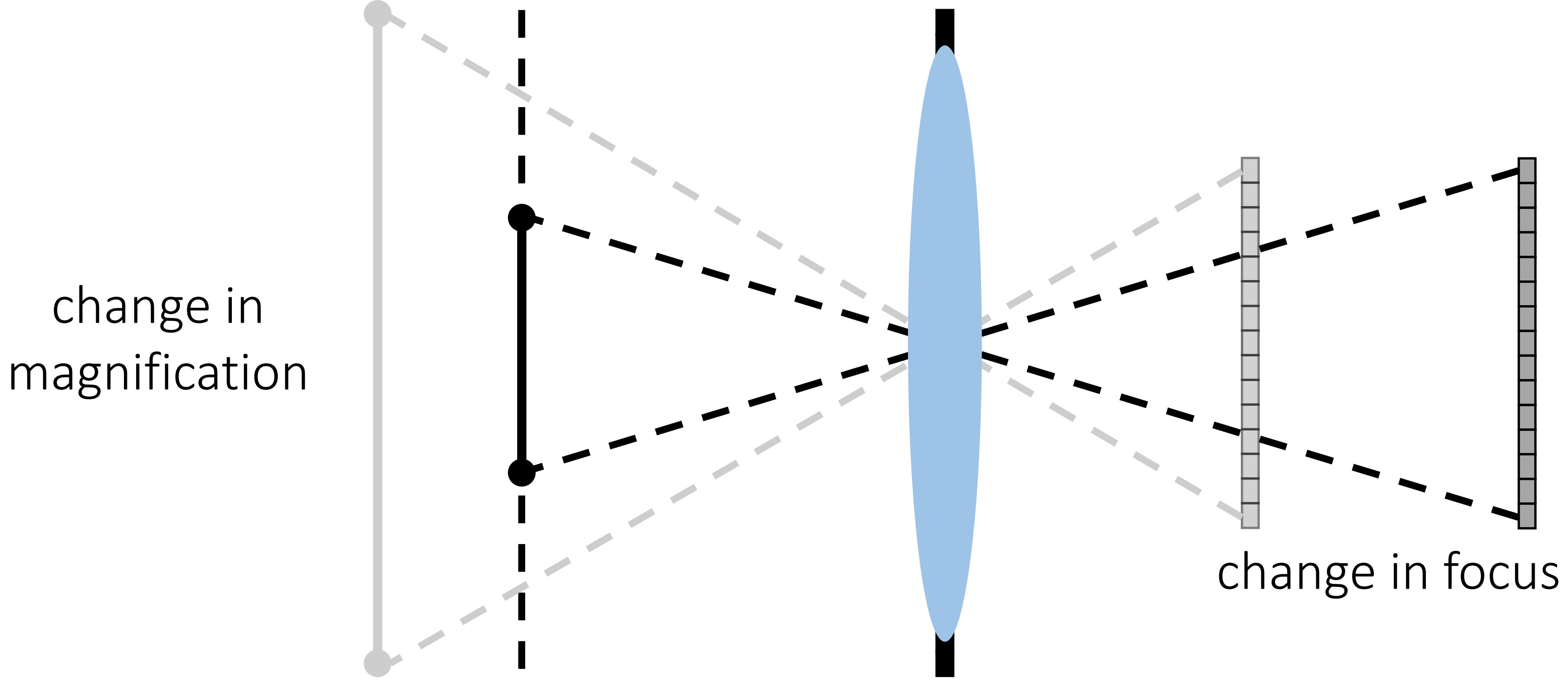


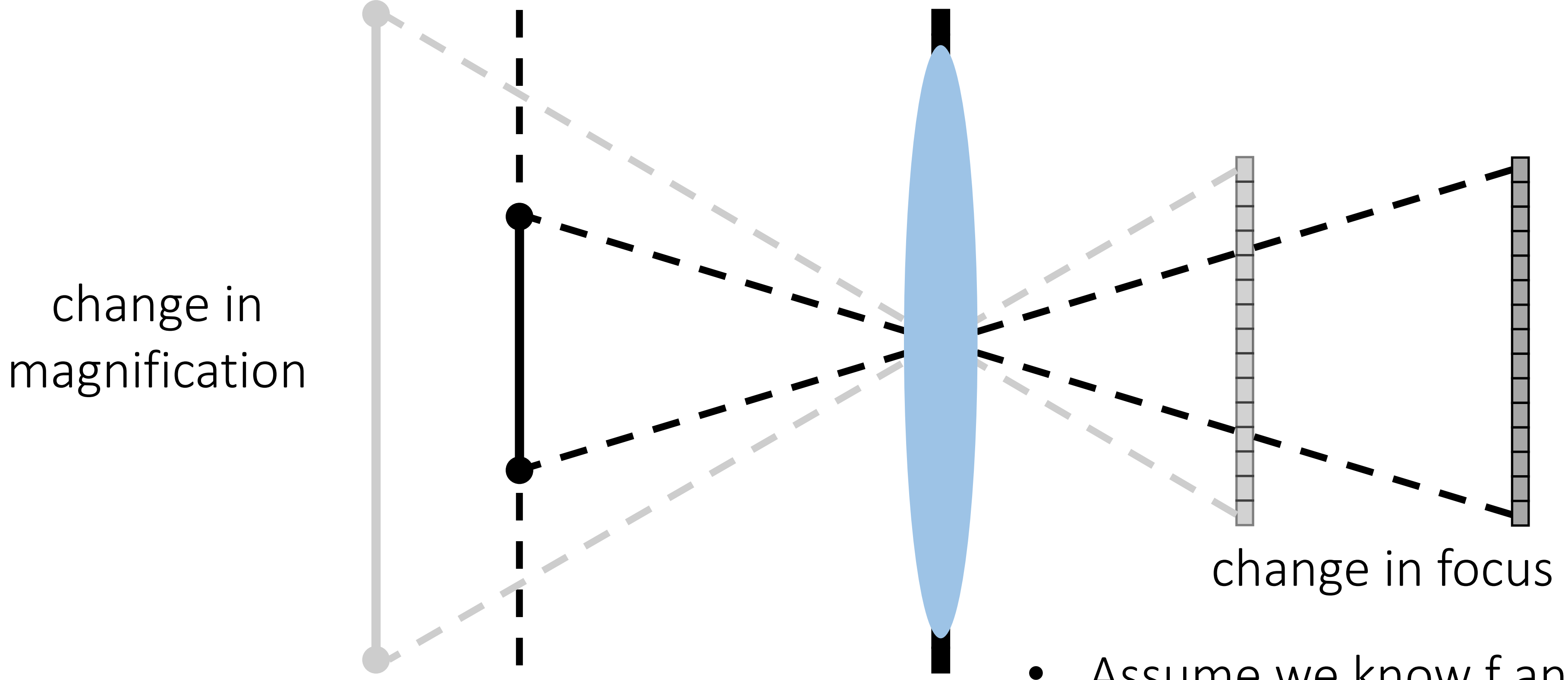
Image alignment

Image alignment

Image alignment

Why do we need to align the images?

- When we change focus distance, we also change field of view (magnification).
- Also, scene may not be static (but we will be ignoring this for now).



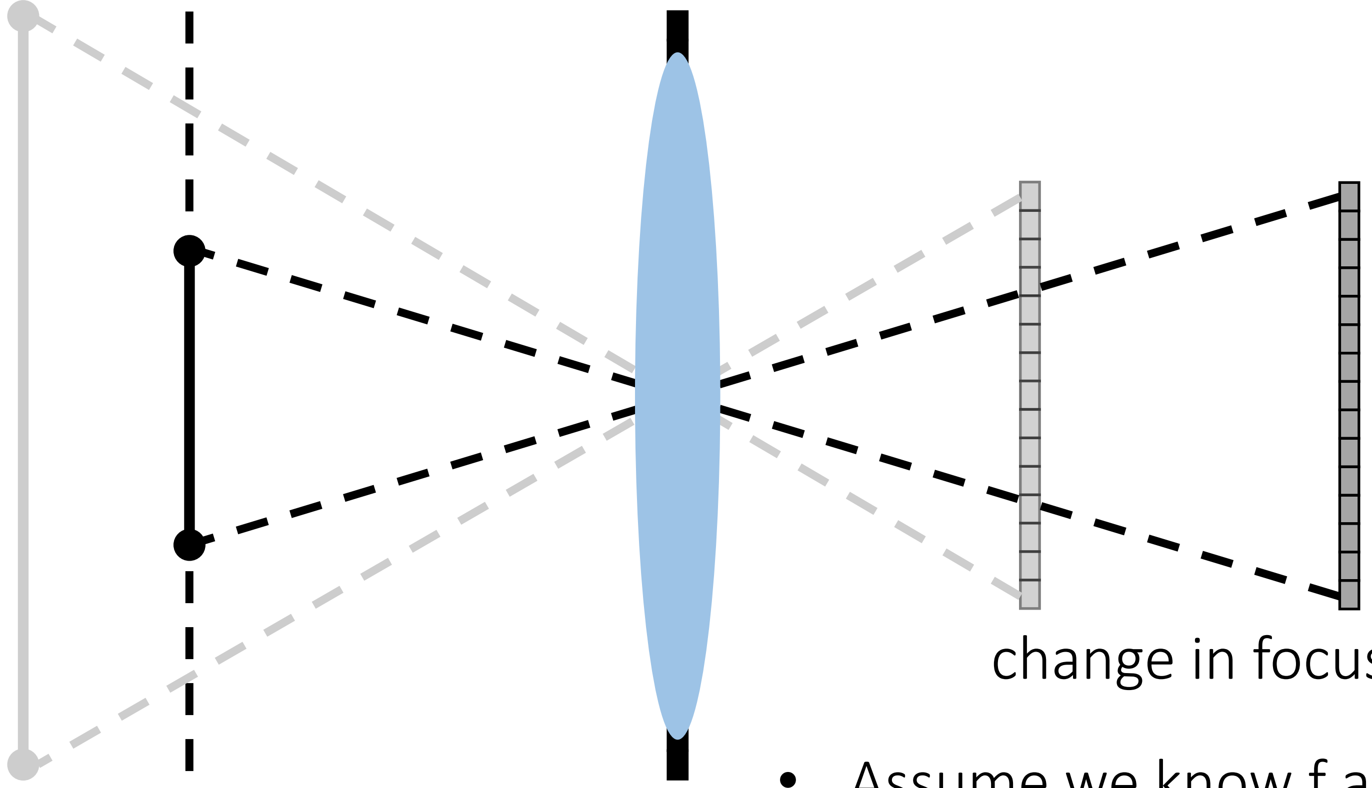
- Assume we know f and all D' values.
- How do can we align the images?

Image alignment

Why do we need to align the images?

- When we change focus distance, we also change field of view (magnification).
- Also, scene may not be static (but we will be ignoring this for now).

change in magnification



change in focus

$$m = \frac{f}{D' - f}$$

$$\frac{1}{D'} + \frac{1}{D} = \frac{1}{f}$$

resize using these equations

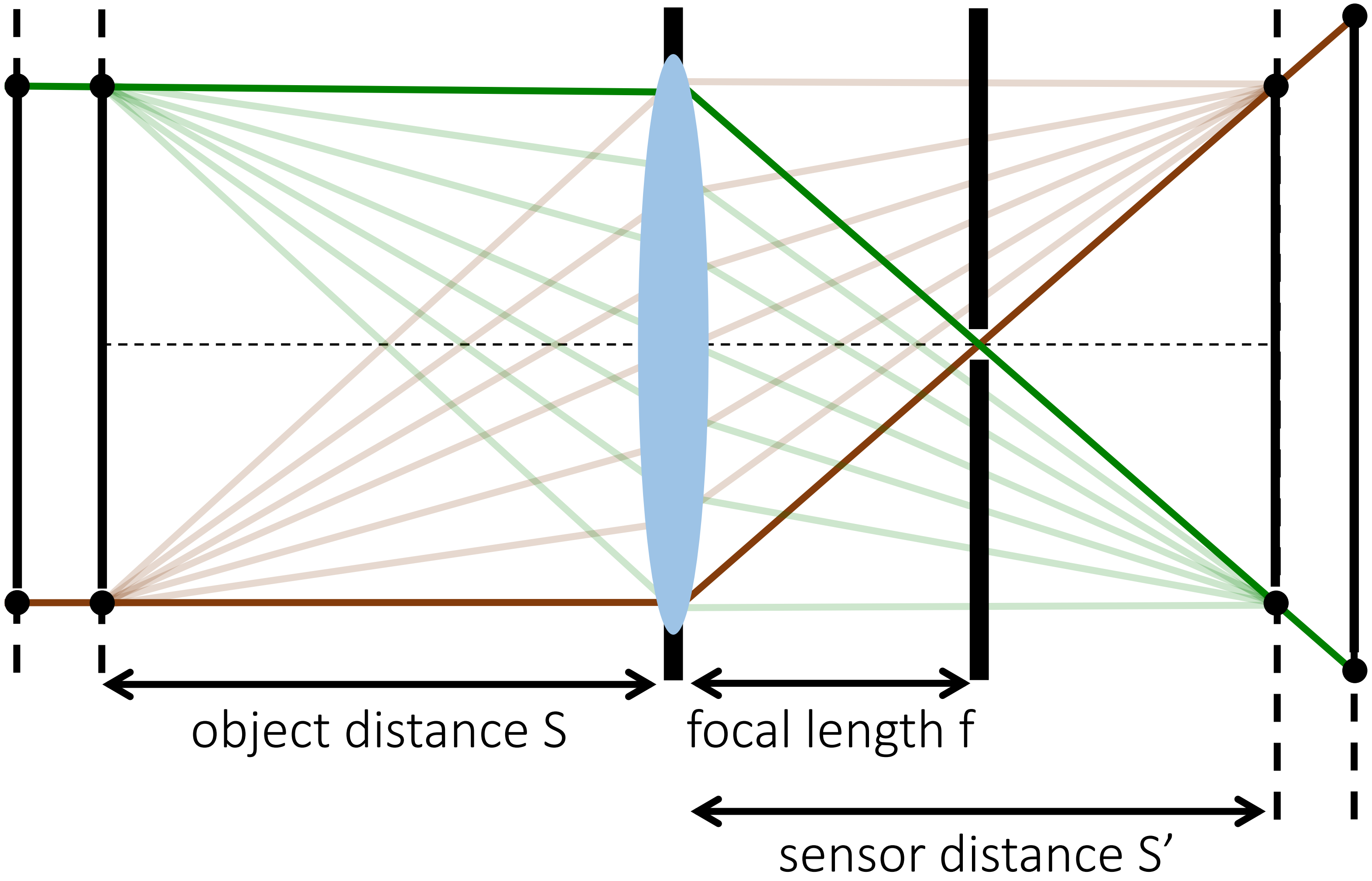
- Assume we know f and all D’ values.
- How do can we align the images?

How can we avoid having to do alignment?

Use a telecentric lens

Place a pinhole at focal length, so that only rays parallel to primary ray pass through.

Magnification independent of object depth.

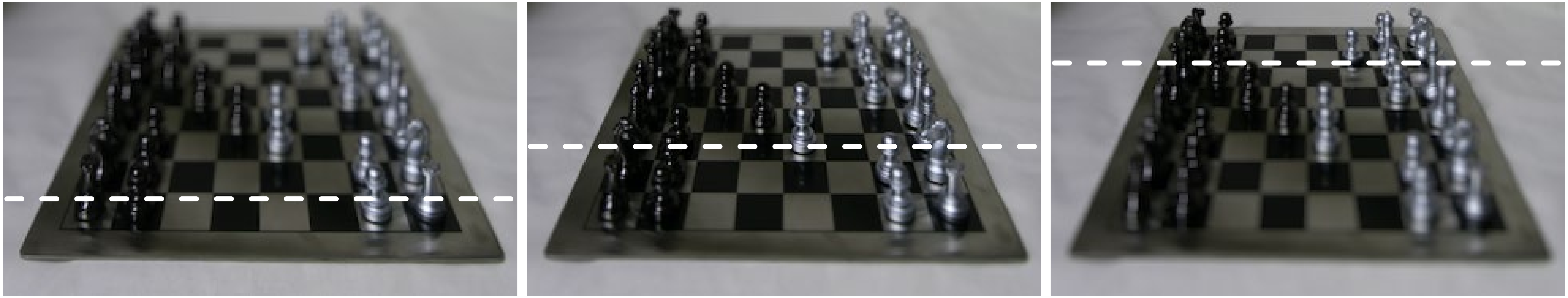


Magnification depends only on sensor-lens distance S’.

[Watanabe and Nayar, PAMI 1997]

How do you merge a focal stack?

1. Align images
2. Assign per-pixel weights representing “in-focus”-ness
3. Compute image average



Weight assignment

How do we measure how much “in-focus” each pixel is?

Weight assignment

How do we measure how much “in-focus” each pixel is?

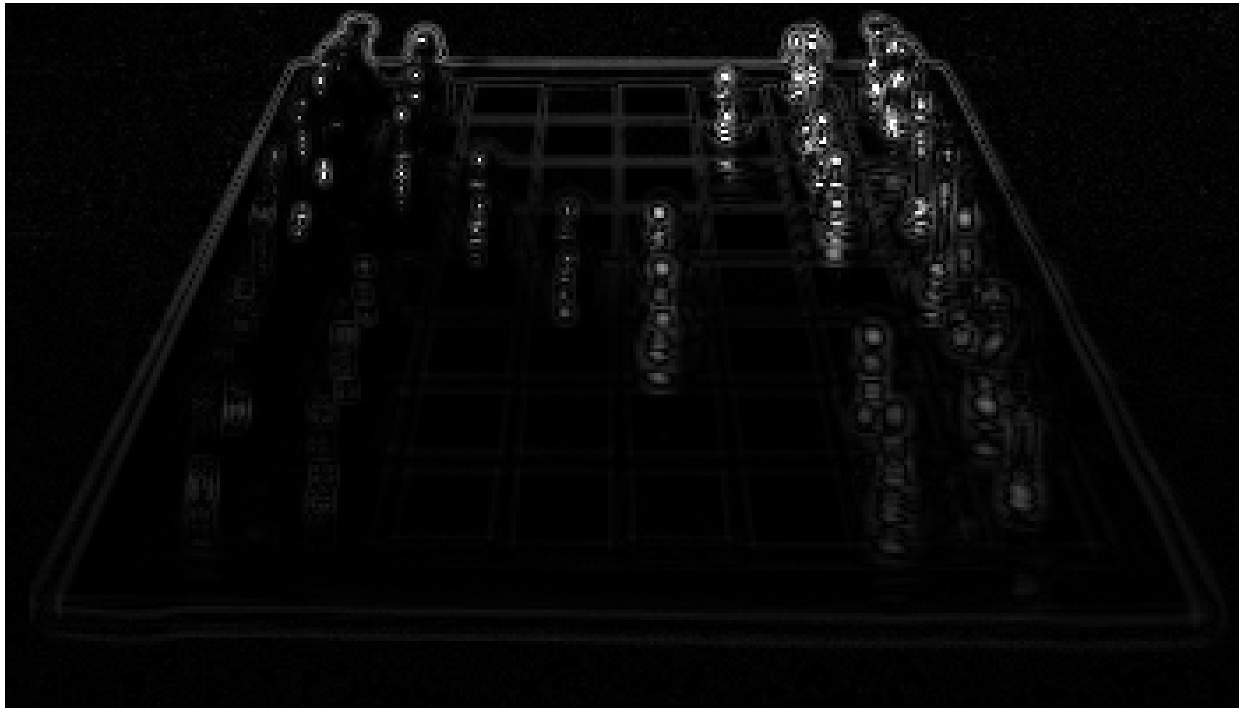
- Measure local sharpness.

How do we measure local sharpness?

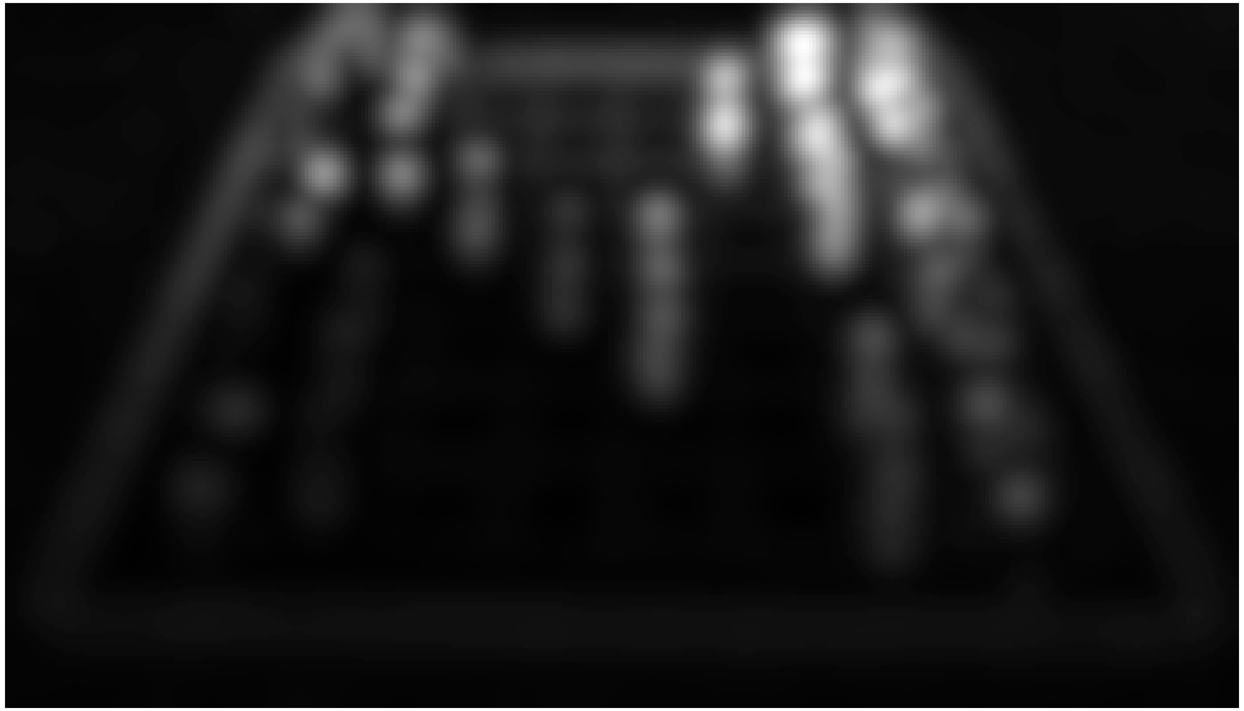
Weight assignment

How do we measure how much “in-focus” each pixel is?

- Measure local sharpness.



run Laplacian operator

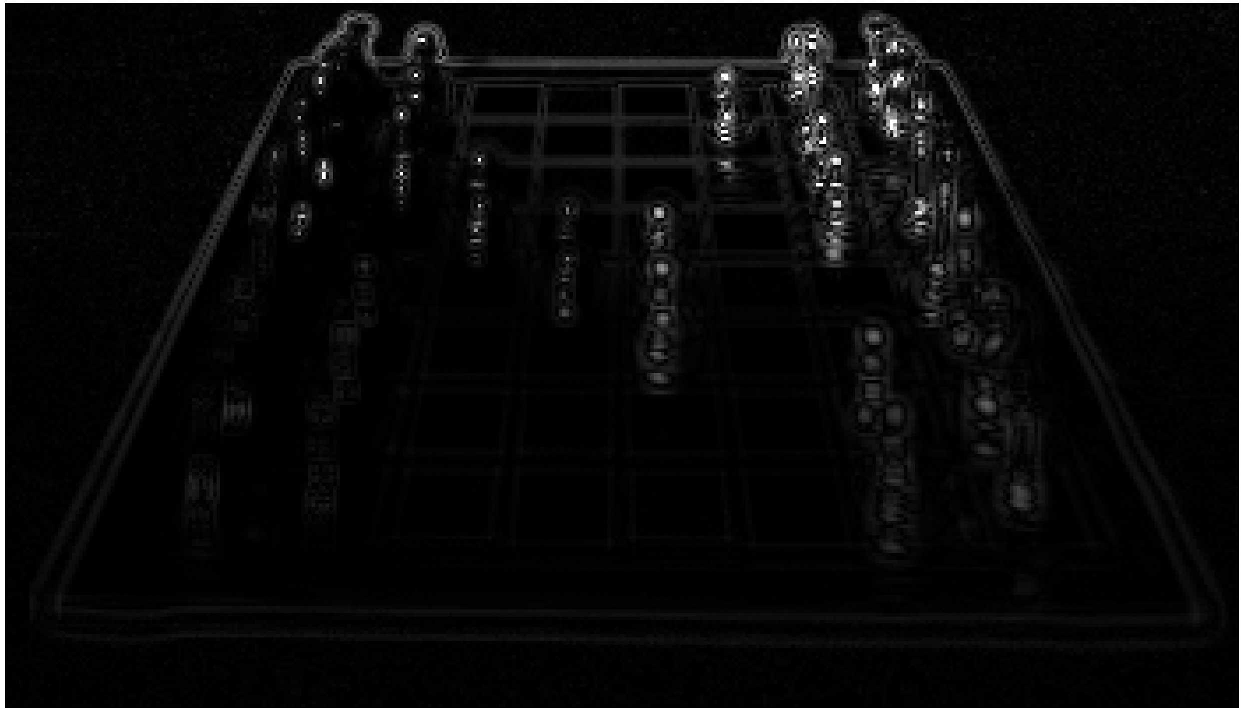


do some Gaussian blurring (why?)

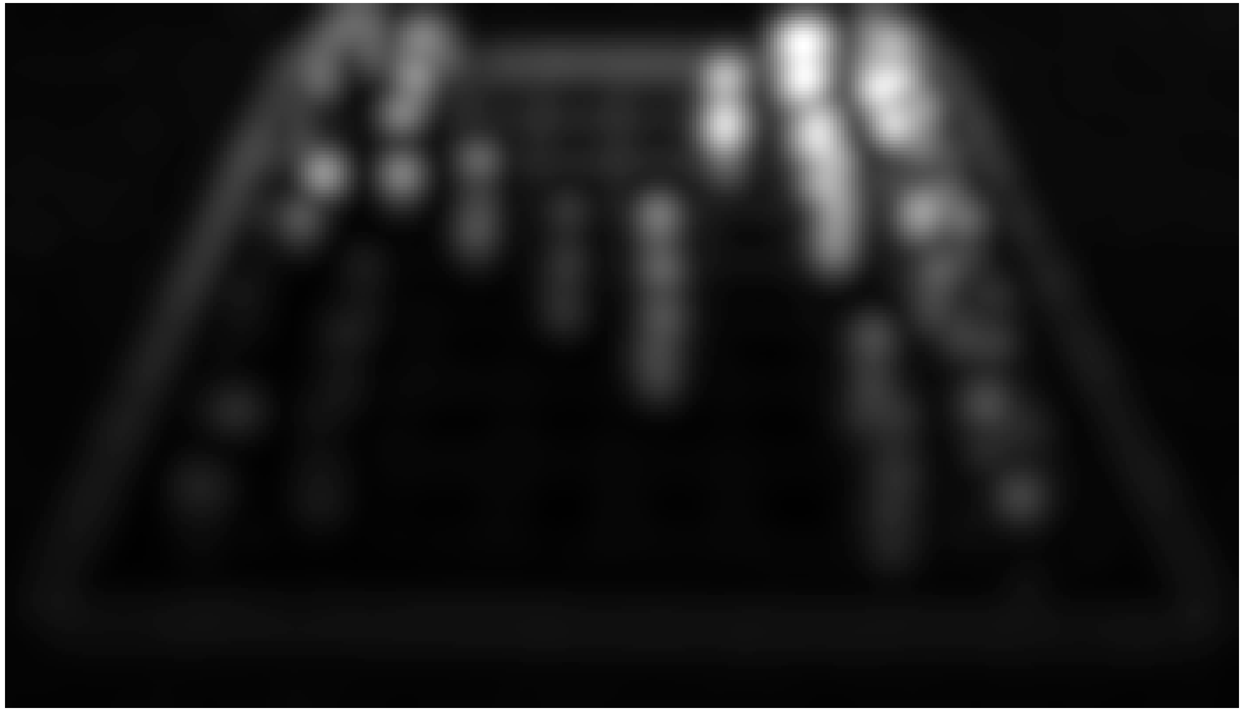
Weight assignment

How do we measure how much "in-focus" each pixel is?

- Measure local sharpness.



run Laplacian operator

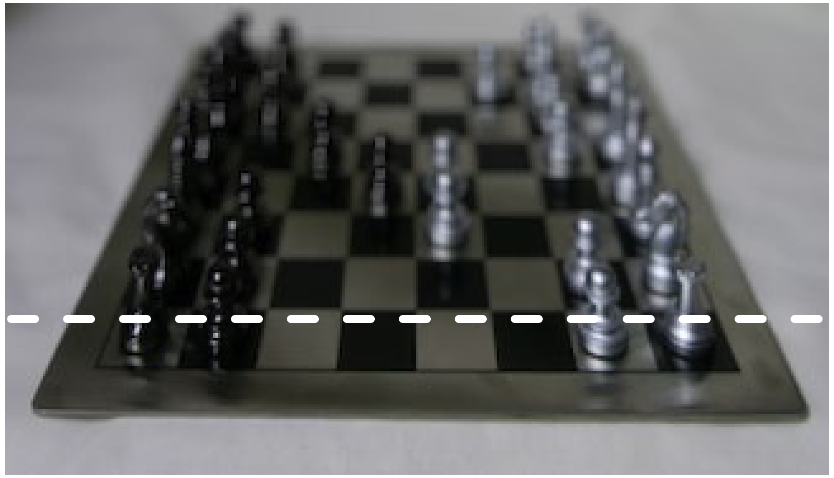


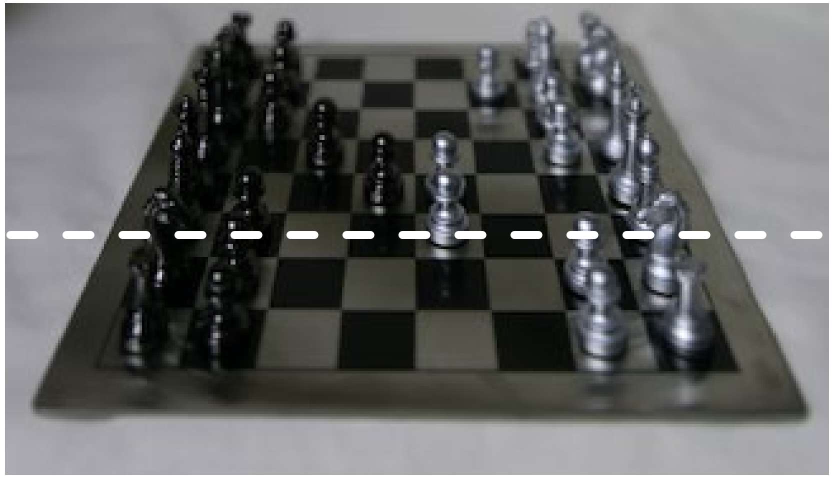
do some Gaussian blurring (so that nearby pixels have similar weights)

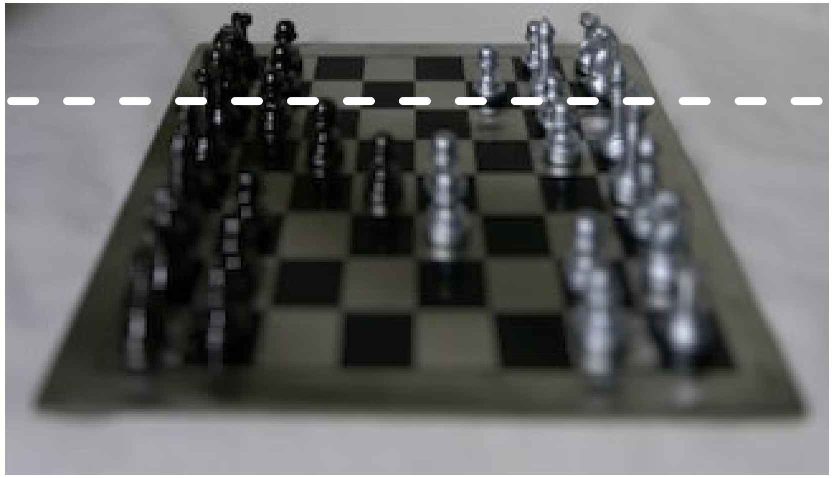
Just one example, many alternatives possible.

How do you merge a focal stack?

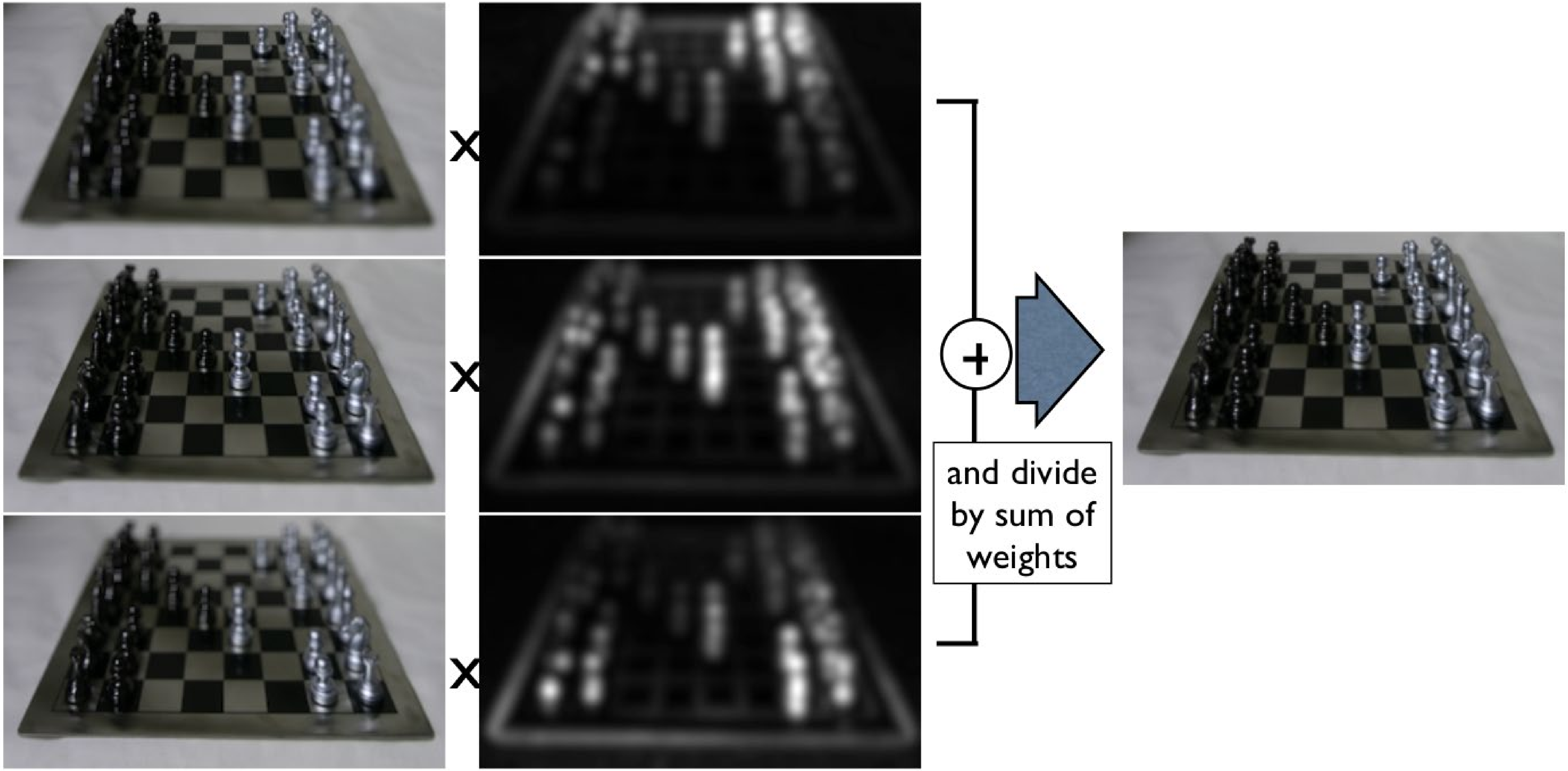
1. Align images
2. Assign per-pixel weights representing "in-focus"-ness
3. Compute image average







Focal stack merging



Some results

example image from stack

all in-focus image

Another example

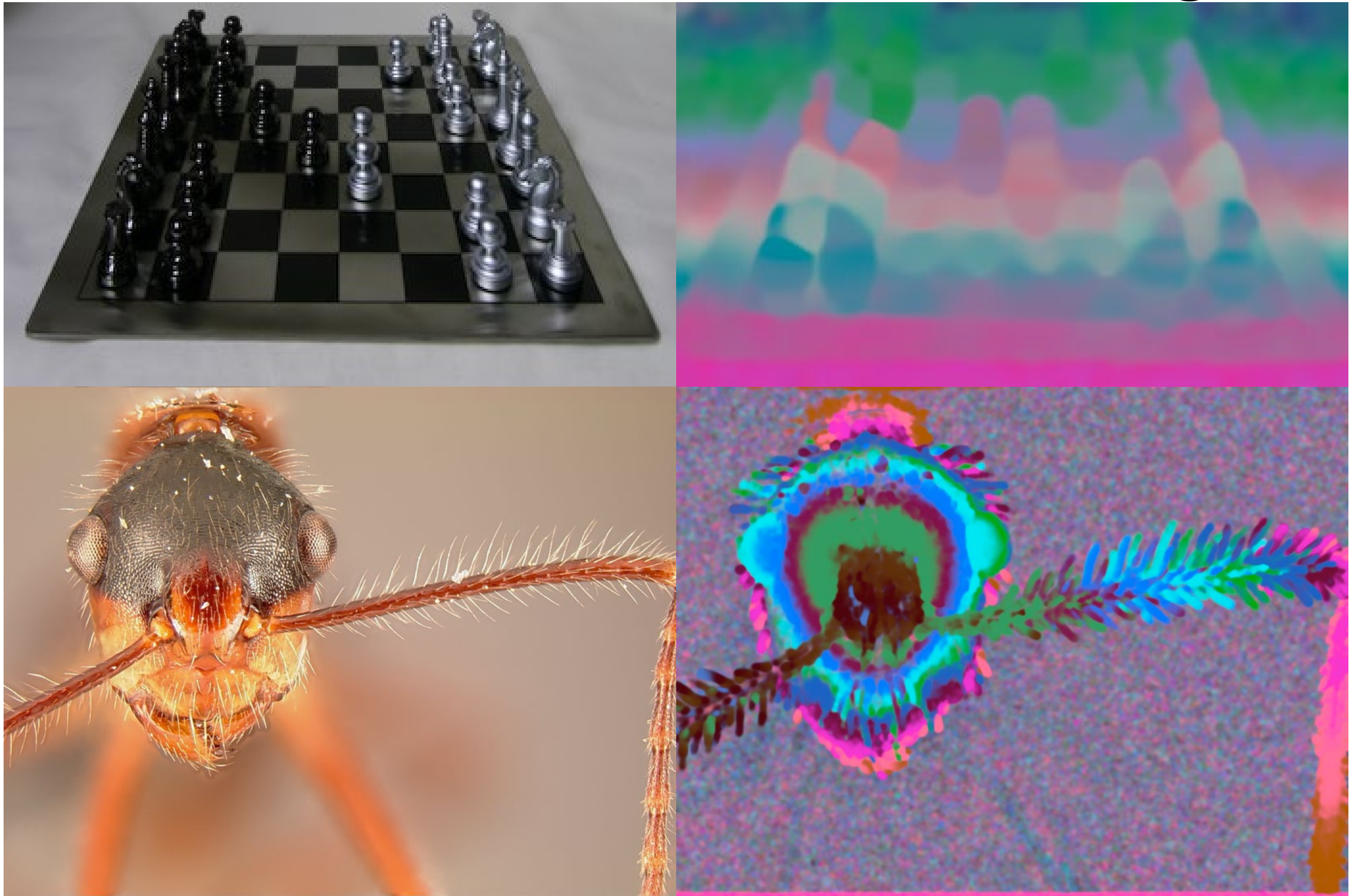
Focal stacking is very useful in macrophotography, where depths of field are very shallow

Another example

middle image from stack

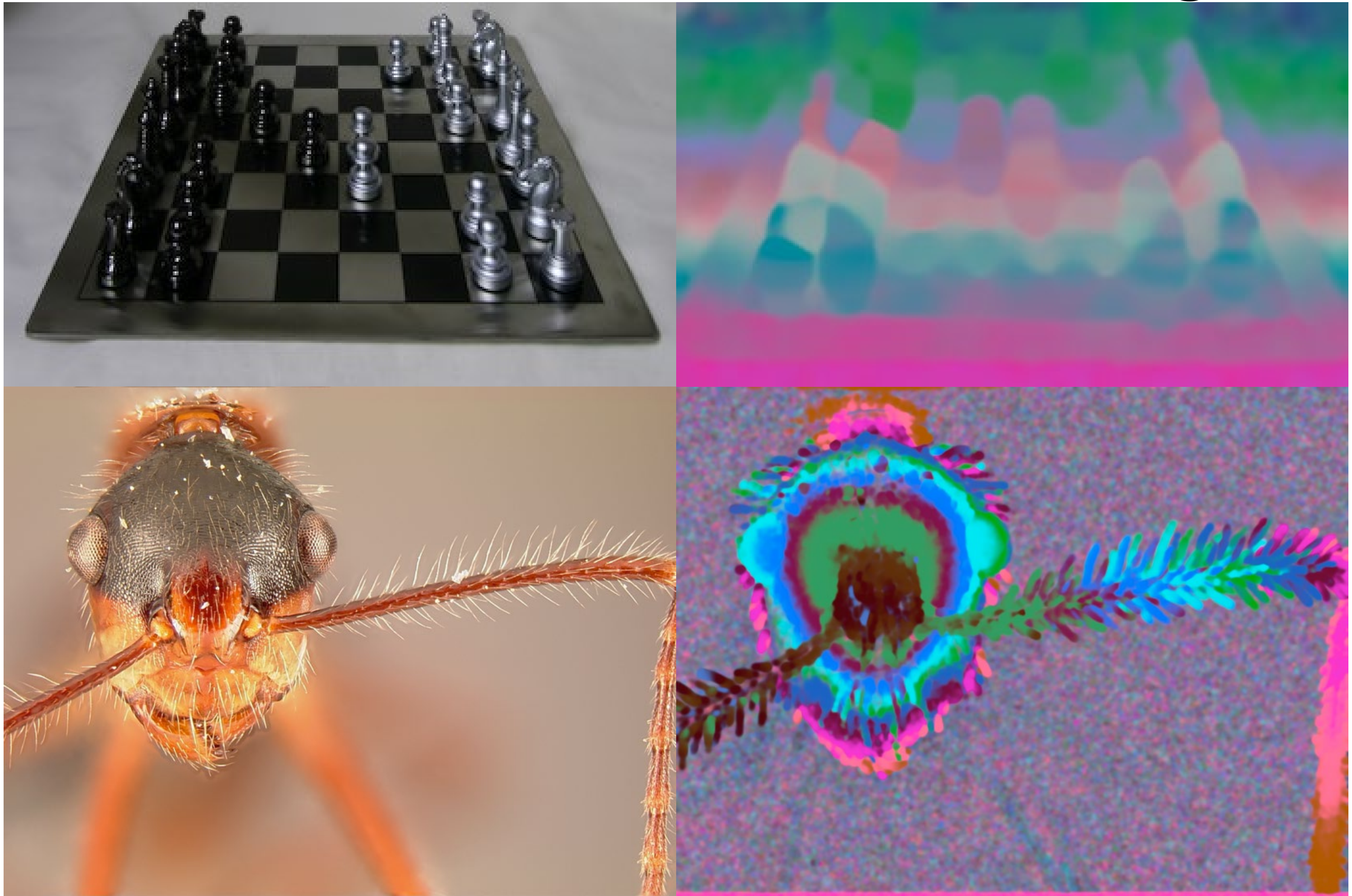
all in-focus image

Another look at the mixing weights



What do the mixing weights look like?

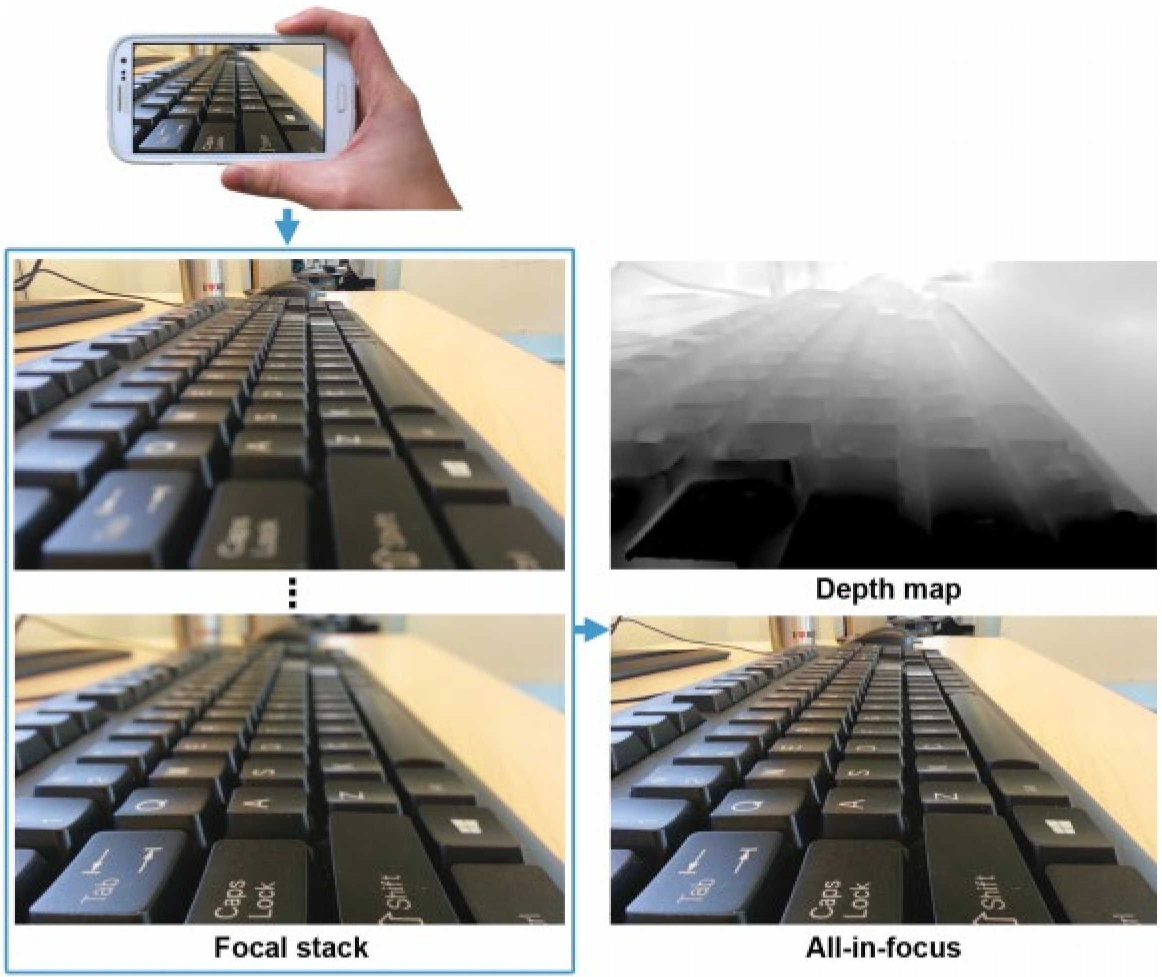
Another look at the mixing weights



Depth from focus = determining sharpest pixel in focal stack

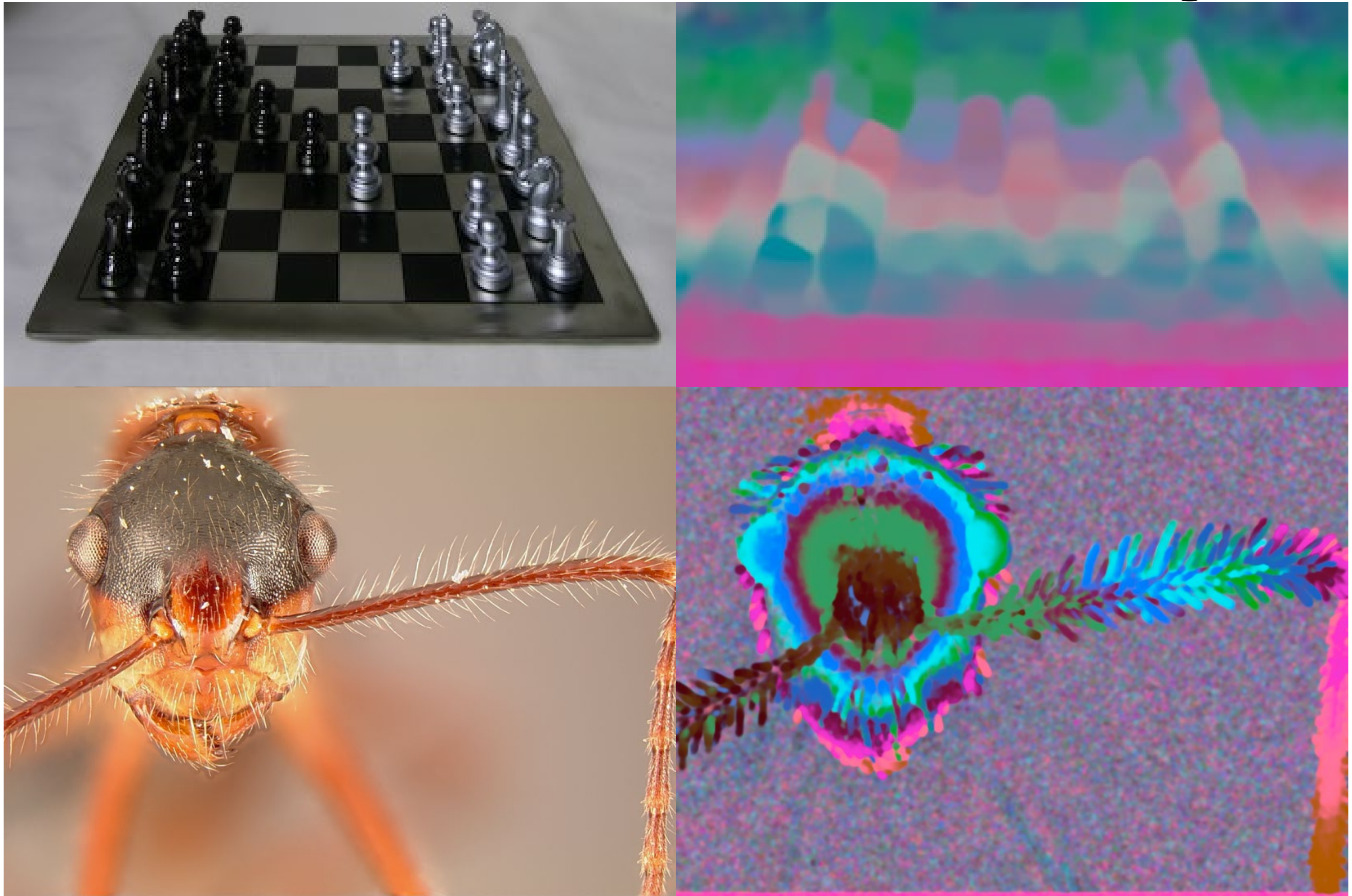
Depth from focus on a mobile phone

Use focal stack from autofocus



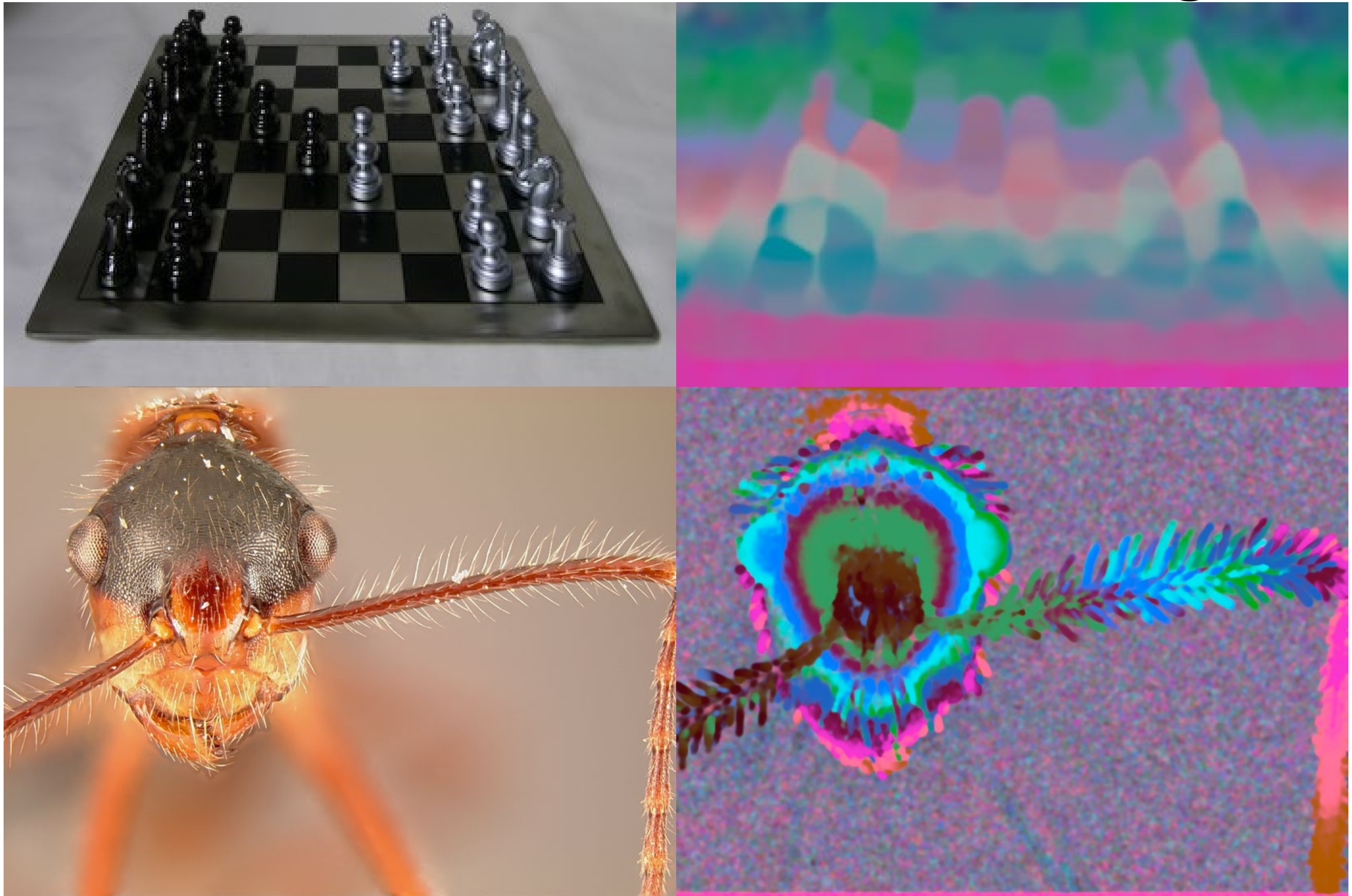
[Suwajanakorn et al., CVPR 2015]

Another look at the mixing weights



What is a problem of these depth maps?

Another look at the mixing weights



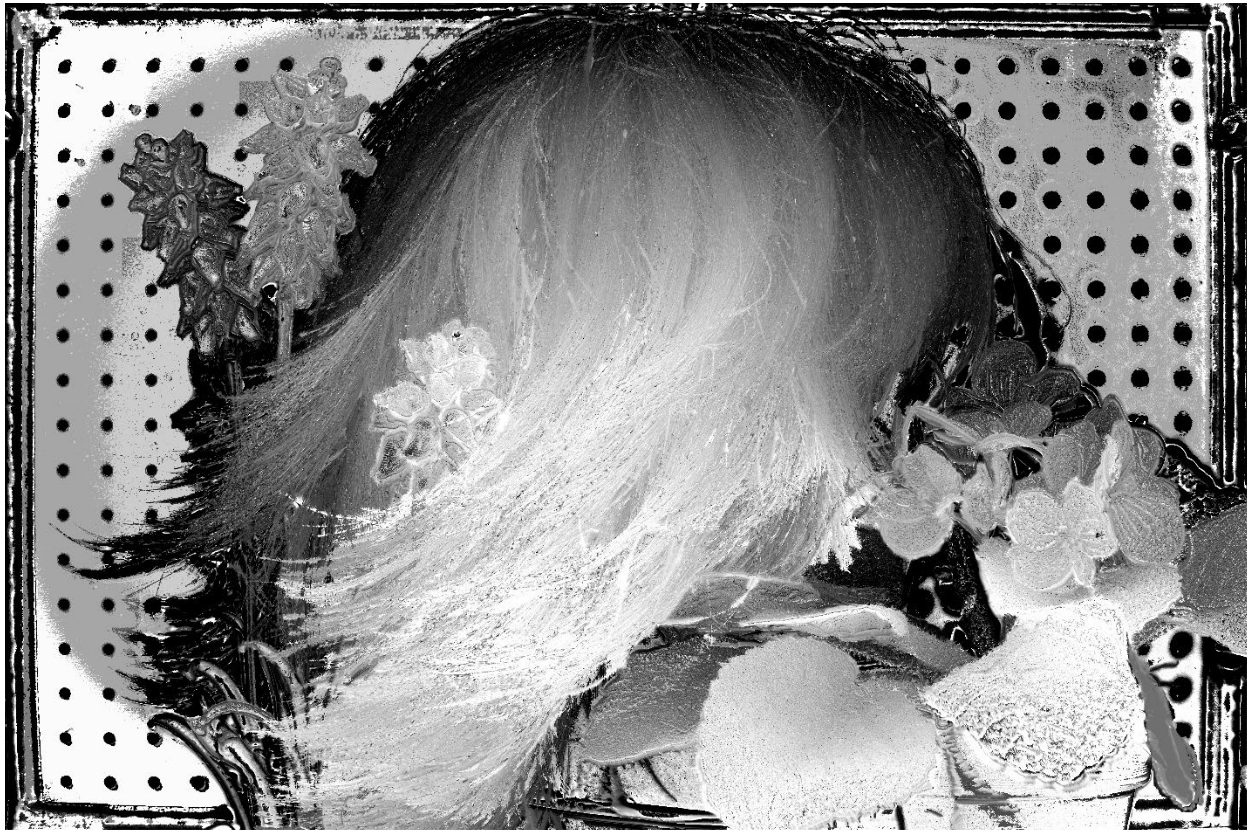
What is a problem of these depth maps?

- Blurry because we to process entire neighborhoods to compute sharpness.
- Can we use any extra information to get *per-pixel* depth?

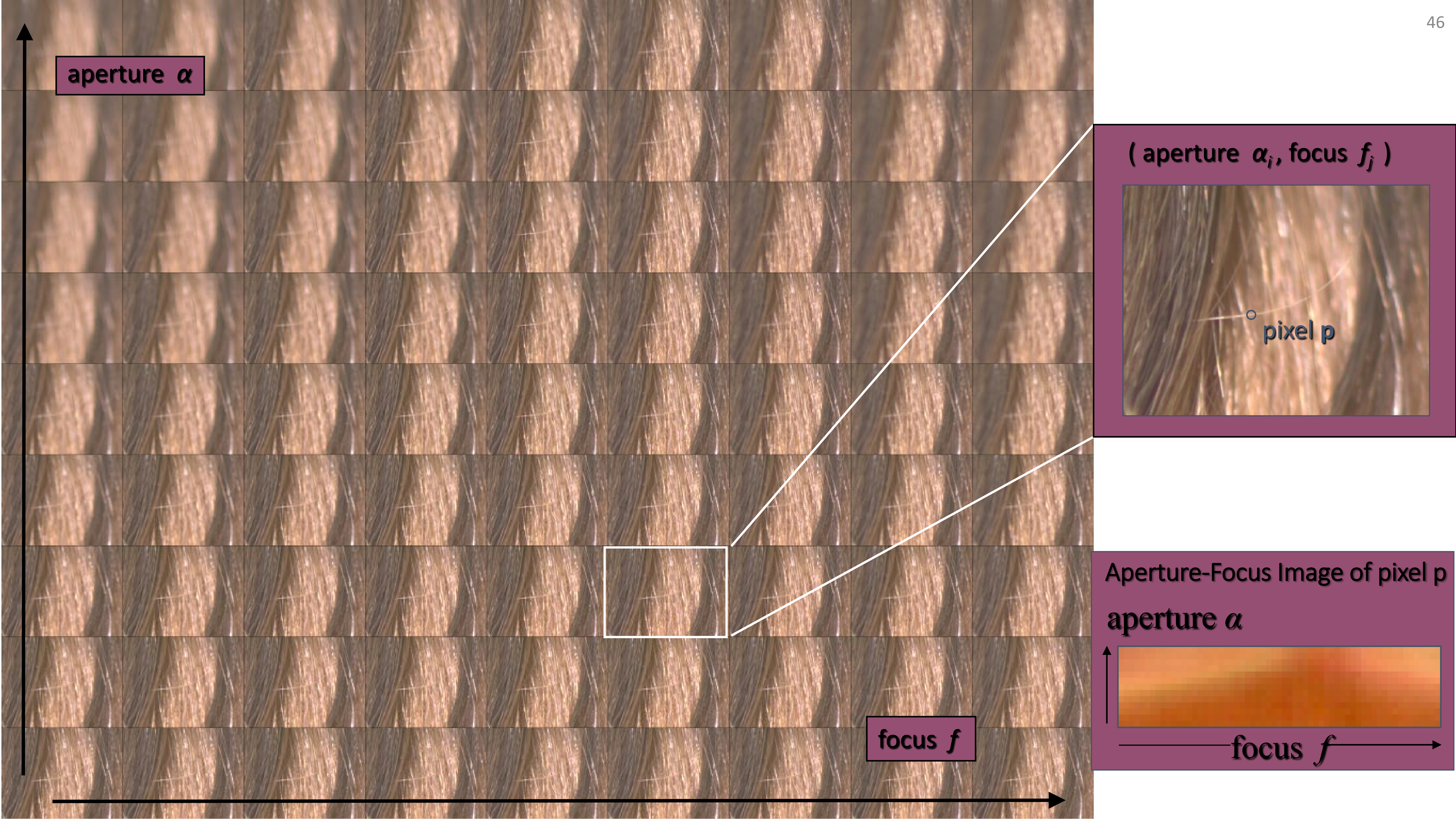
Confocal stereo

Confocal stereo

- Capture a 2D stack by varying *both* focus and aperture
- Analyze each pixel's focus-aperture image to find true depth



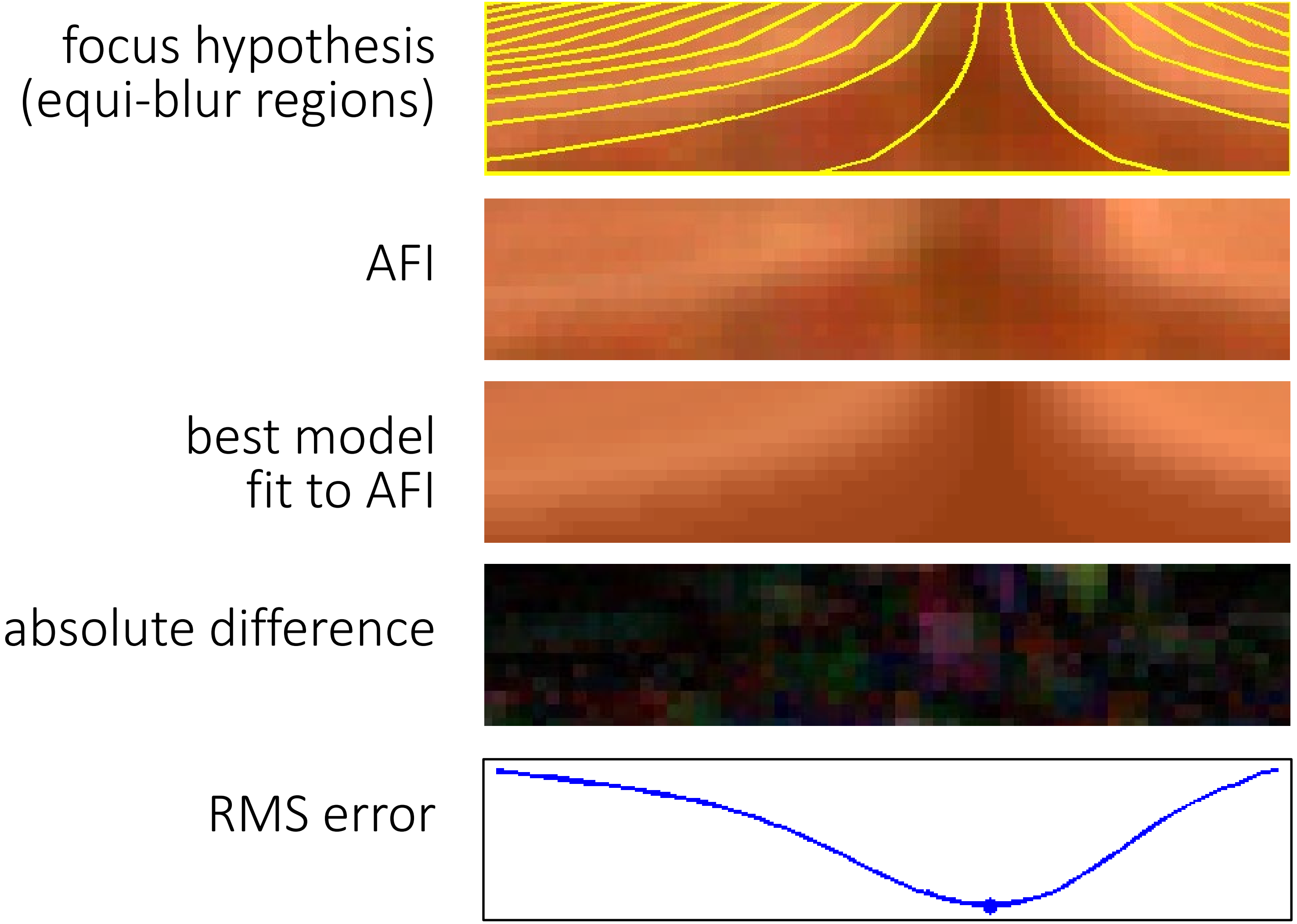
[Hassinof and Kutulakos, ECCV 2006]


aperture α
(aperture α_i, focus f_j)
pixel p
Aperture-Focus Image of pixel p
aperture α
focus f
focus f

AFI model fitting

Key idea: When a point is in focus, its color and intensity remain the same for all apertures.

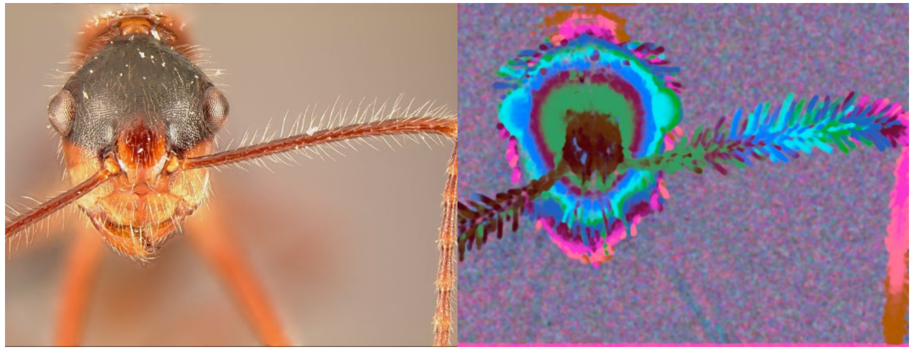
- This property is called *confocal constancy*.
- We can find correct depth by checking intensity variation across apertures for each focus setting.



Depth inference techniques

Confocal stereo:
- Requires focus-aperture stack.
- Gives per-pixel depth estimates.

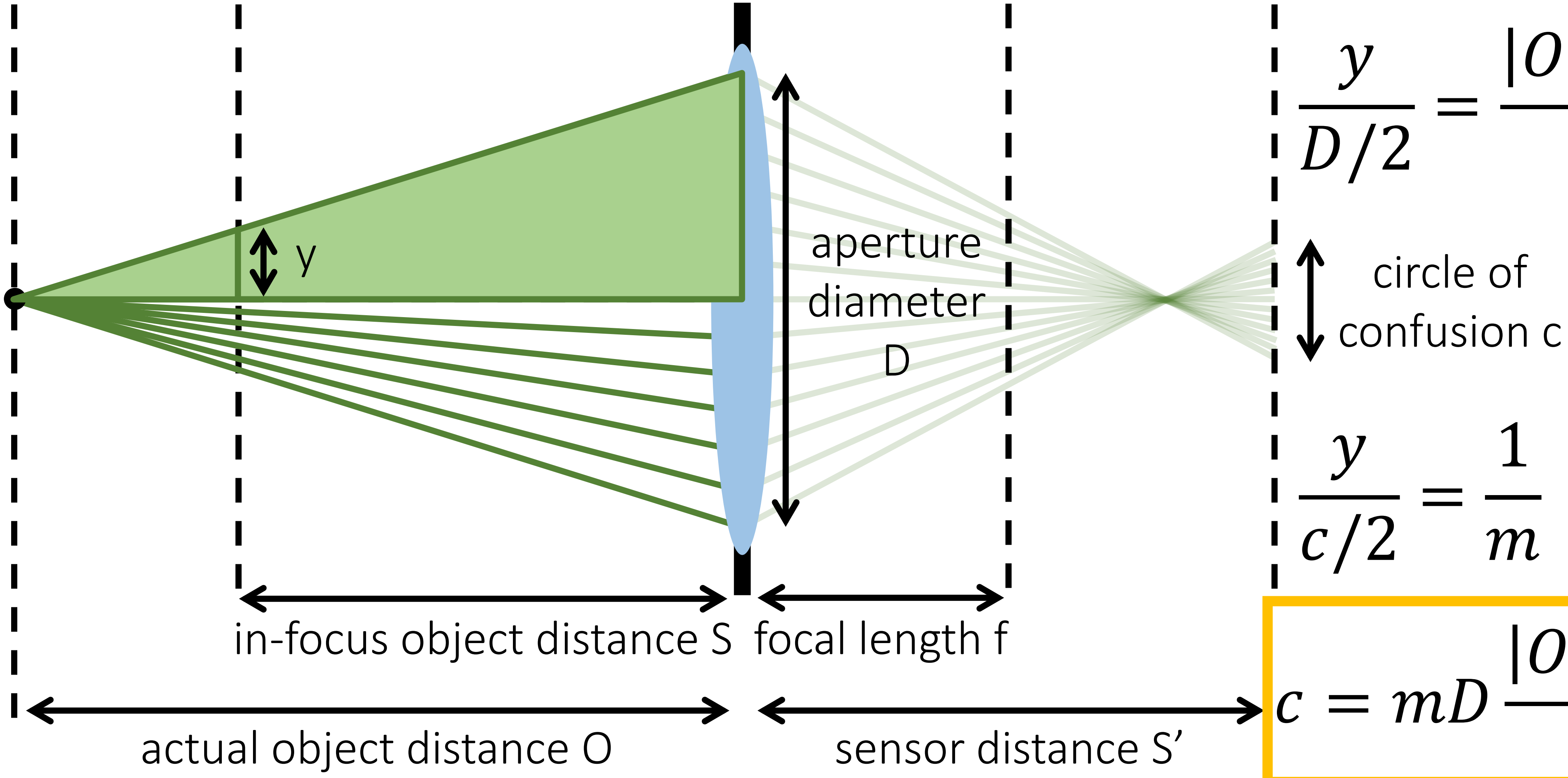
Depth from focus:
- Requires focus stack.
- Gives per-patch depth estimates.



What is a downside of these two approaches?

Reminder: Circle of confusion

Size of circle of confusion depends on distance from in-focus plane.



$$\frac{y}{D/2} = \frac{|O - S|}{O}$$

$$\frac{y}{c/2} = \frac{1}{m}$$

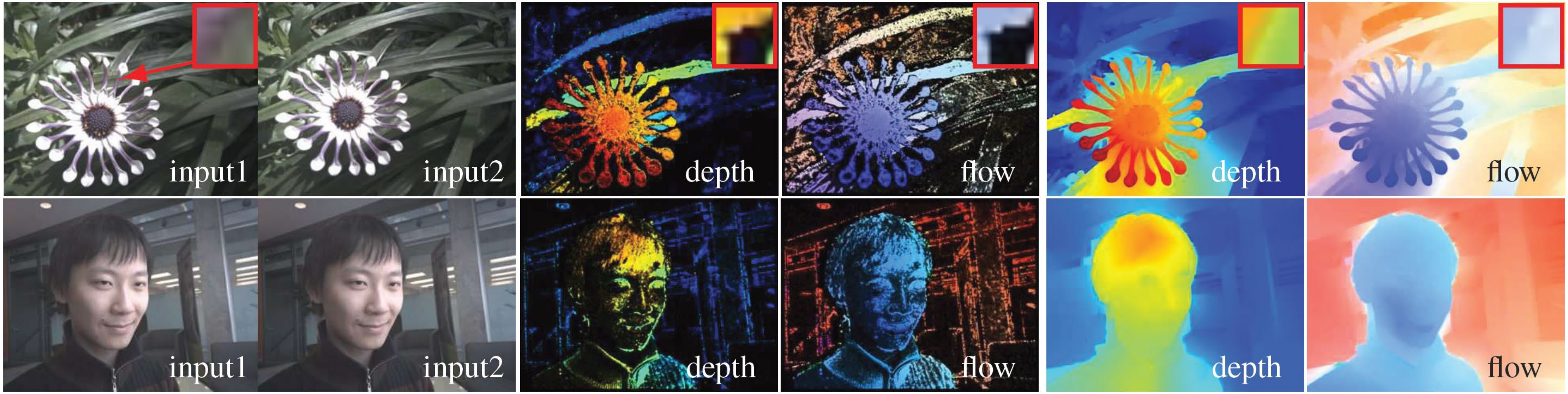
$$\boxed{c = mD\frac{|O - S|}{O}}$$

Depth from defocus

Use as few as two images (*two-frame depth from defocus*):

- Assume circle of confusion can be modeled as a (typically Gaussian) blur kernel with varying σ.
- Use pair of images to estimate how blur size at each pixel changes from one depth to another.
- Relate this blur size to depth.

Requires very elaborate modeling and priors to be robust.

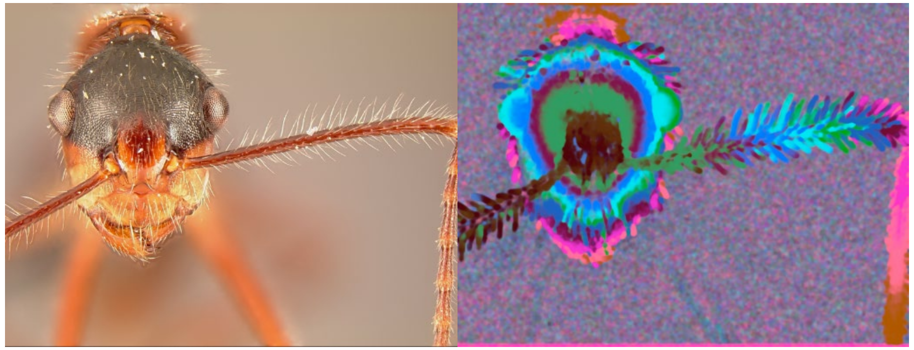


[Tang et al., CVPR 2017]

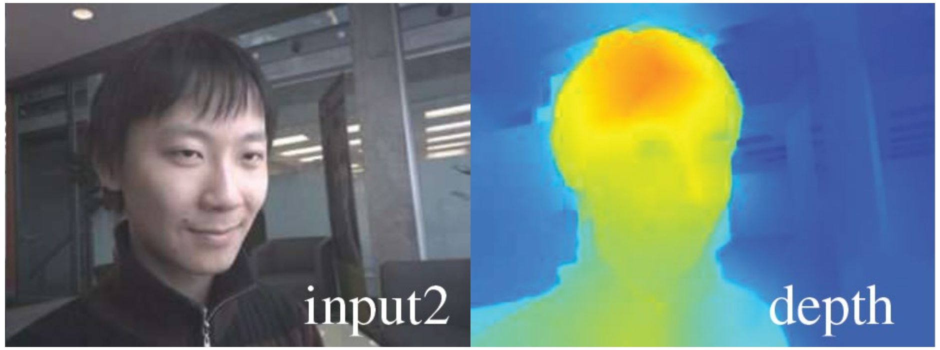
Depth inference techniques

Confocal stereo:
- Requires focus-aperture stack.
- Gives per-pixel depth estimates.

Depth from focus:
- Requires focus stack.
- Gives per-patch depth estimates.



Depth from defocus:
- Requires only two images.
- Gives per-patch depth estimates.



What is a downside of these three approaches?

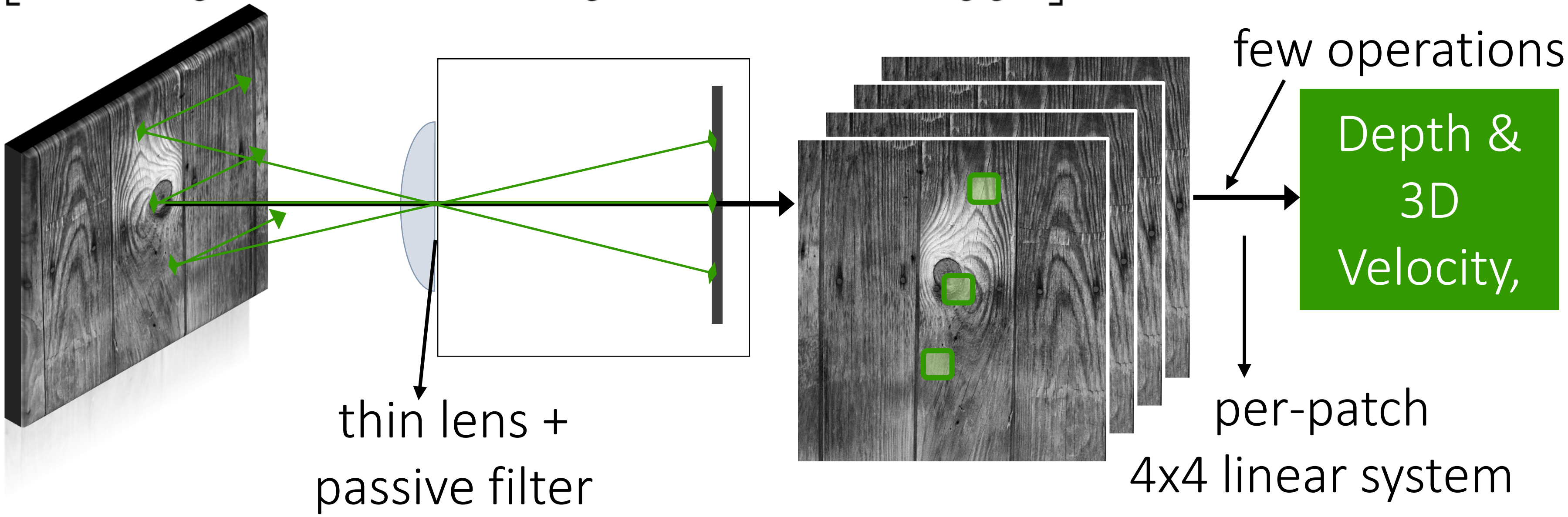
Focal flow

Use a dense focal stack (e.g., from autofocus):

- Assume circle of confusion can be modeled as a (typically Gaussian) blur kernel with varying σ.
- Estimate *optical flow* between successive frames in focal stack.
- Relate optical flow change to depth.

Requires very little computation (no convolutions) but only works on textured patches (why?).

$$\begin{bmatrix} I_x & I_y & (xI_x + yI_y) & (I_{xx} + I_{yy}) \end{bmatrix} \vec{u}_4 + I_t = 0$$



https://www.youtube.com/watch?v=vHdjMgo47BQ

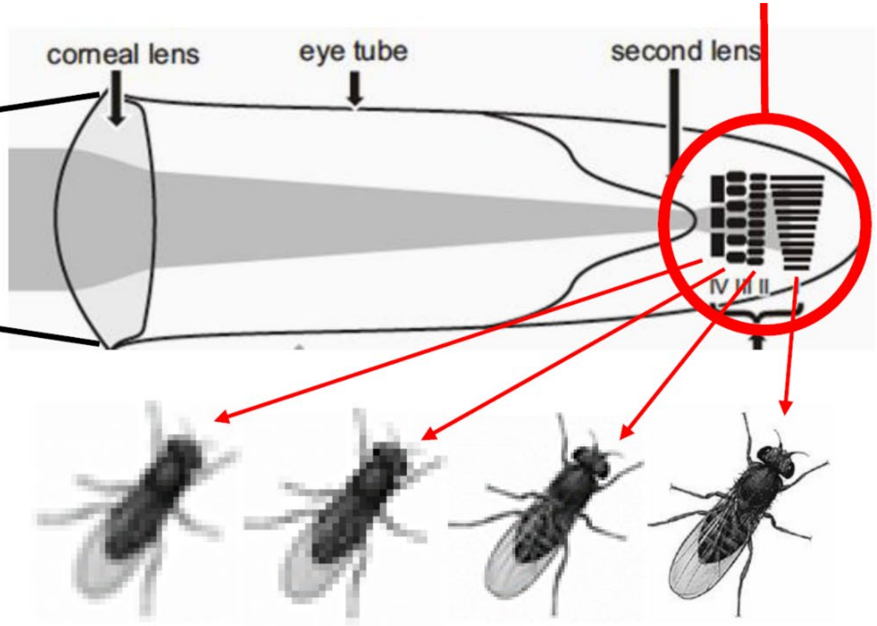
[Alexander et al., ECCV 2016; Guo et al., ICCV 2017]

Focal flow

Use a dense focal stack (e.g., from autofocus):

- Assume circle of confusion can be modeled as a (typically Gaussian) blur kernel with varying σ.
- Estimate *optical flow* between successive frames in focal stack.
- Relate optical flow change to depth.

Requires very little computation (no convolutions) but only works on textured patches (why?).



Biologically-inspired from jumping spiders!

https://www.youtube.com/watch?v=vHdjMgo47BQ

[Alexander et al., ECCV 2016; Guo et al., ICCV 2017]

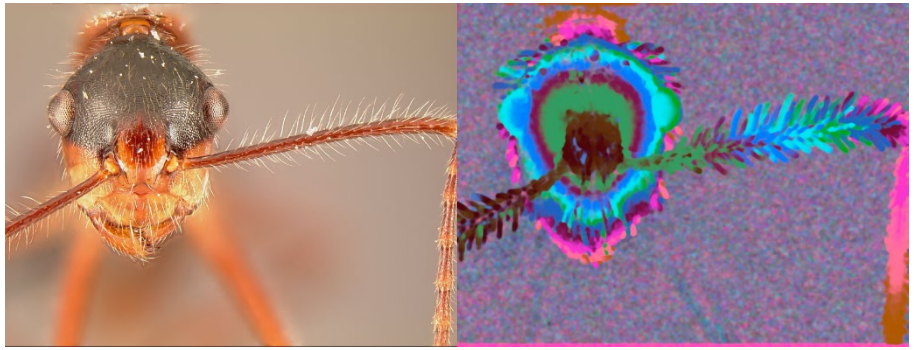
Depth inference techniques

Confocal stereo:

- Requires focus-aperture stack.
- Gives per-pixel depth estimates.

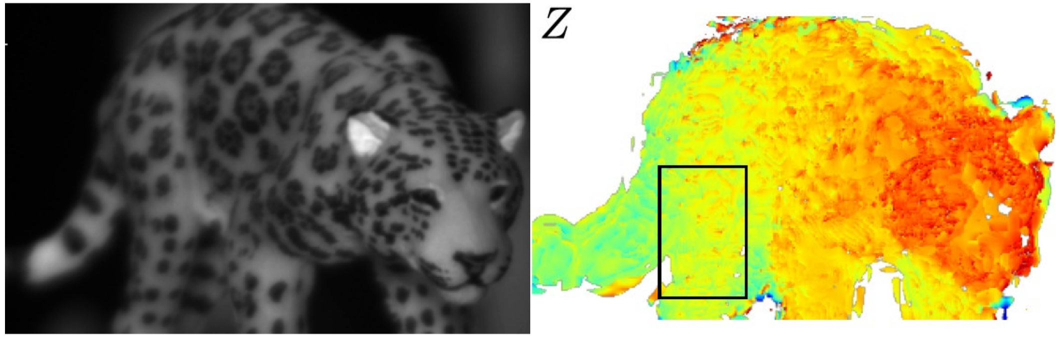
Depth from focus:

- Requires focus stack.
- Gives per-patch depth estimates.



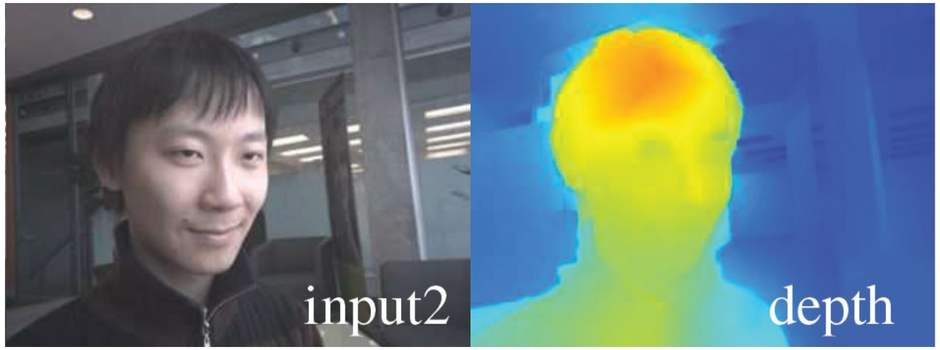
Focal flow:

- Requires dense focus stack.
- Gives *efficient* per-patch depth estimates.



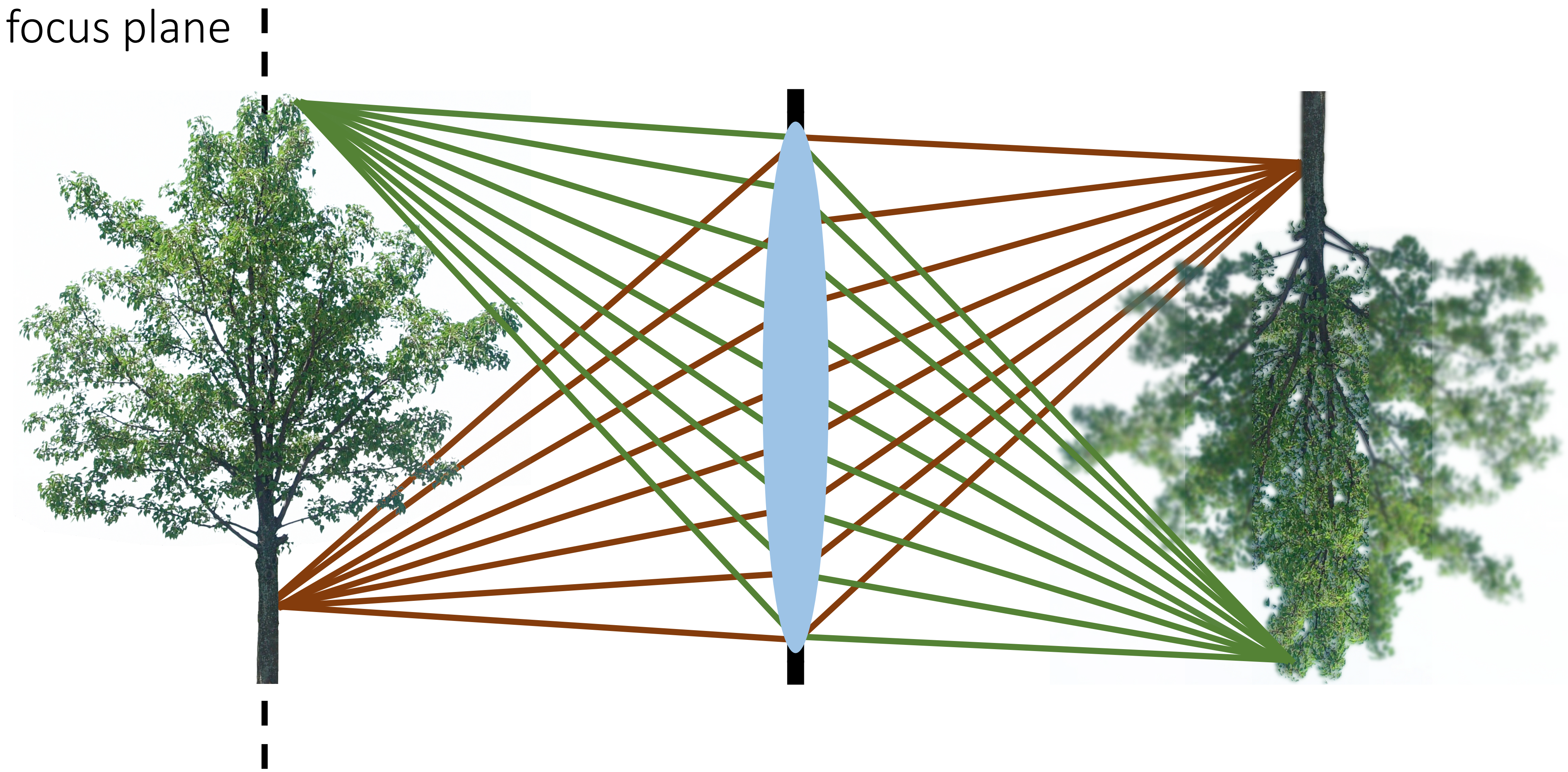
Depth from defocus:

- Requires only two images.
- Gives per-patch depth estimates.



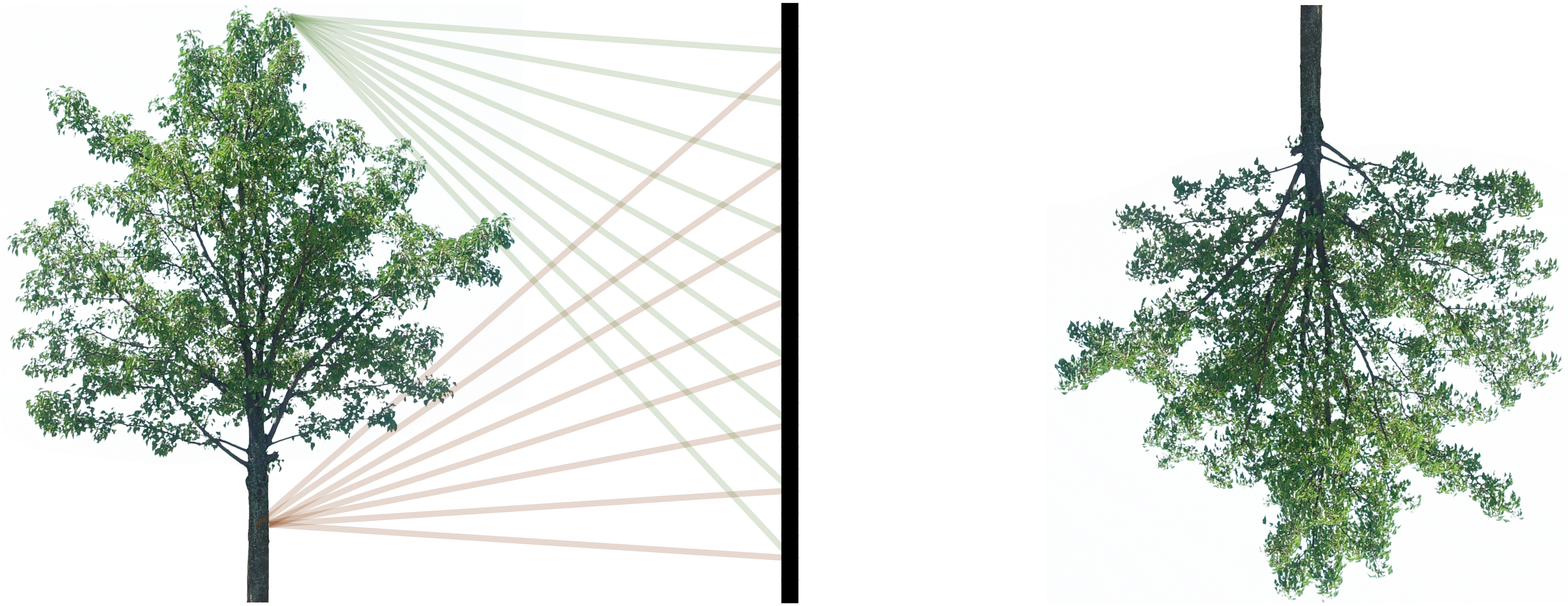
Lightfield

Measuring rays



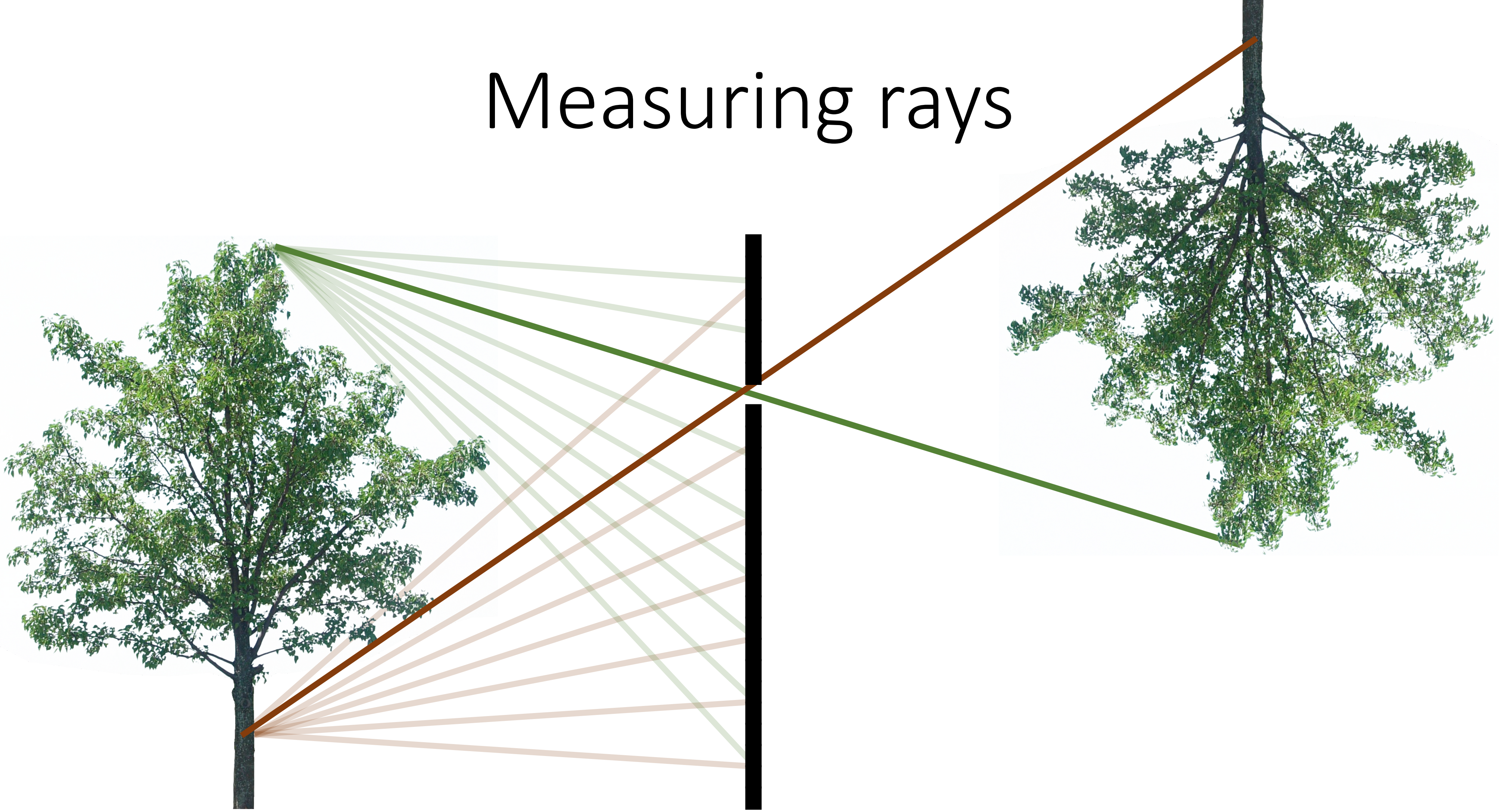
A lens measures all rays radiated from the object (up to aperture size).

Measuring rays



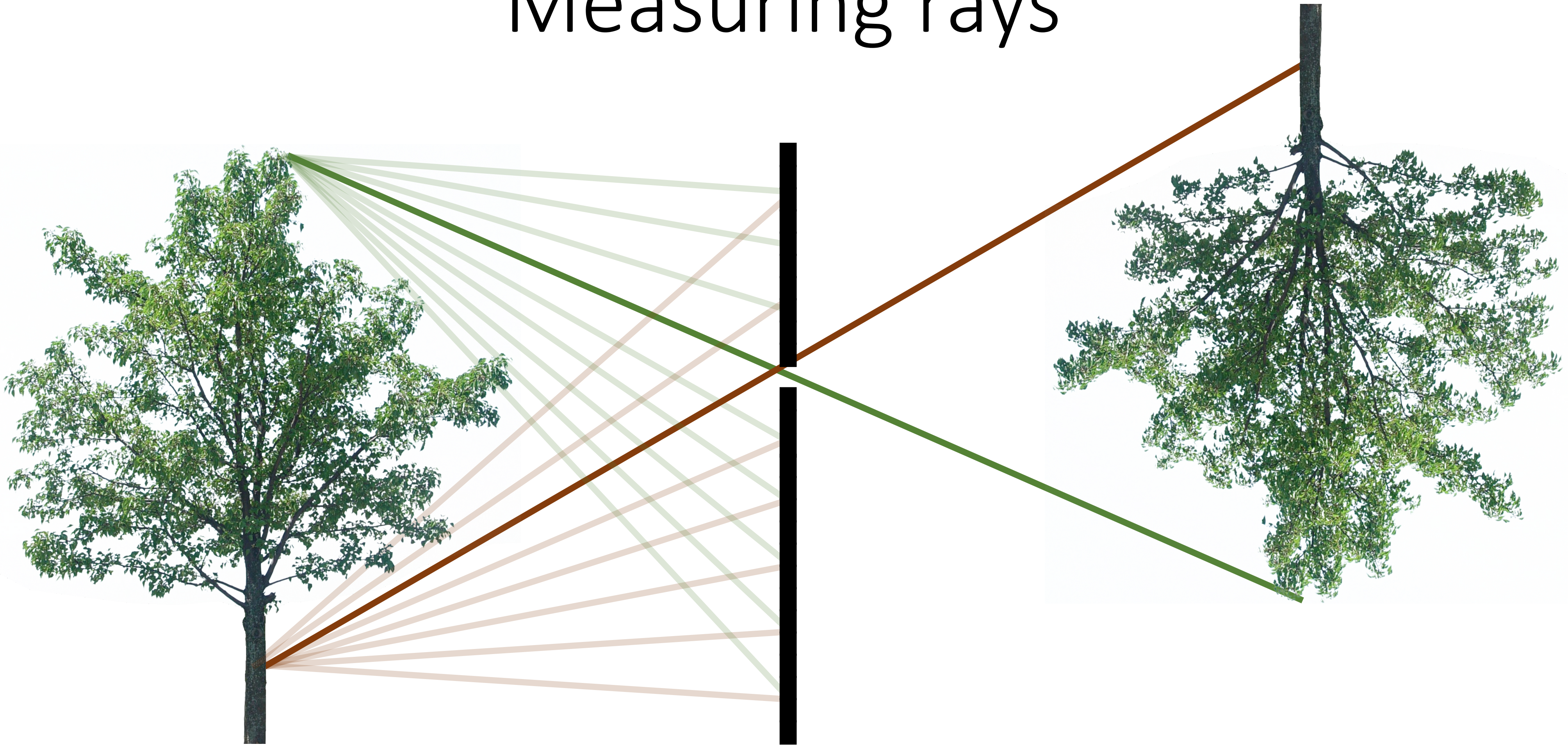
We can capture the same set of rays by using a pinhole camera from multiple viewpoints

Measuring rays



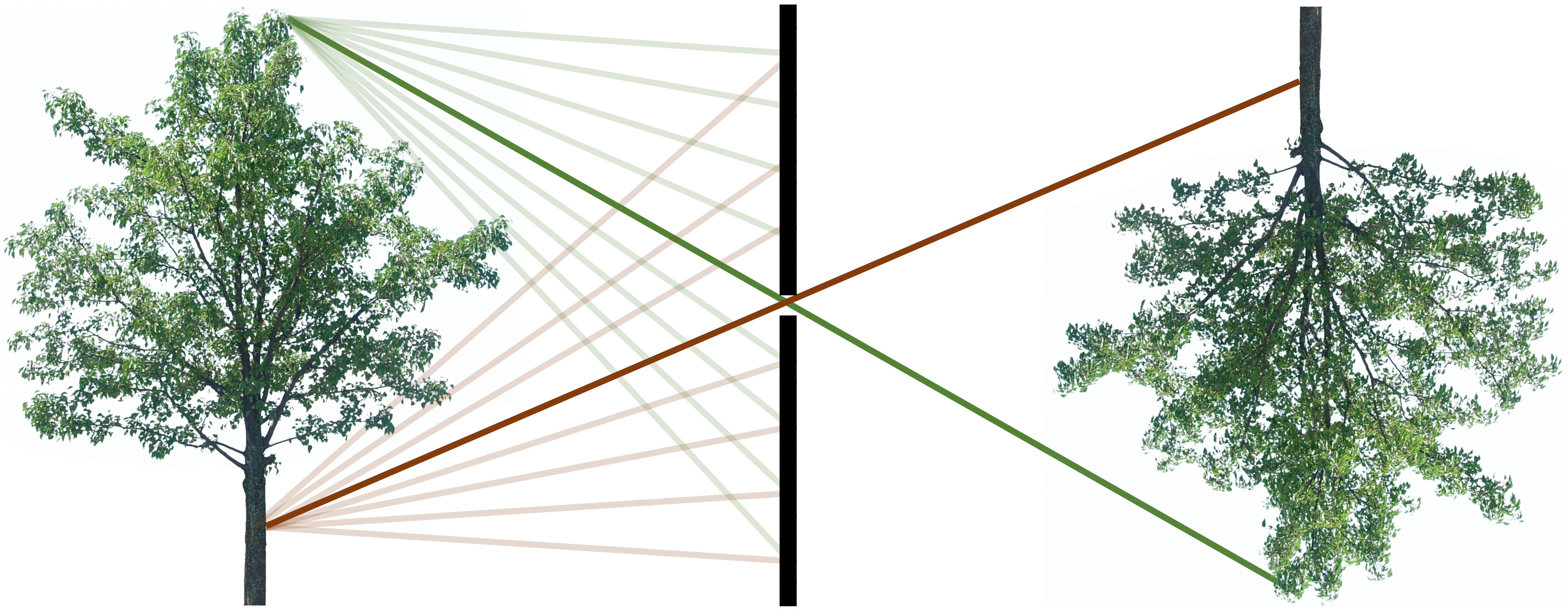
We can capture the same set of rays by using a pinhole camera from multiple viewpoints

Measuring rays



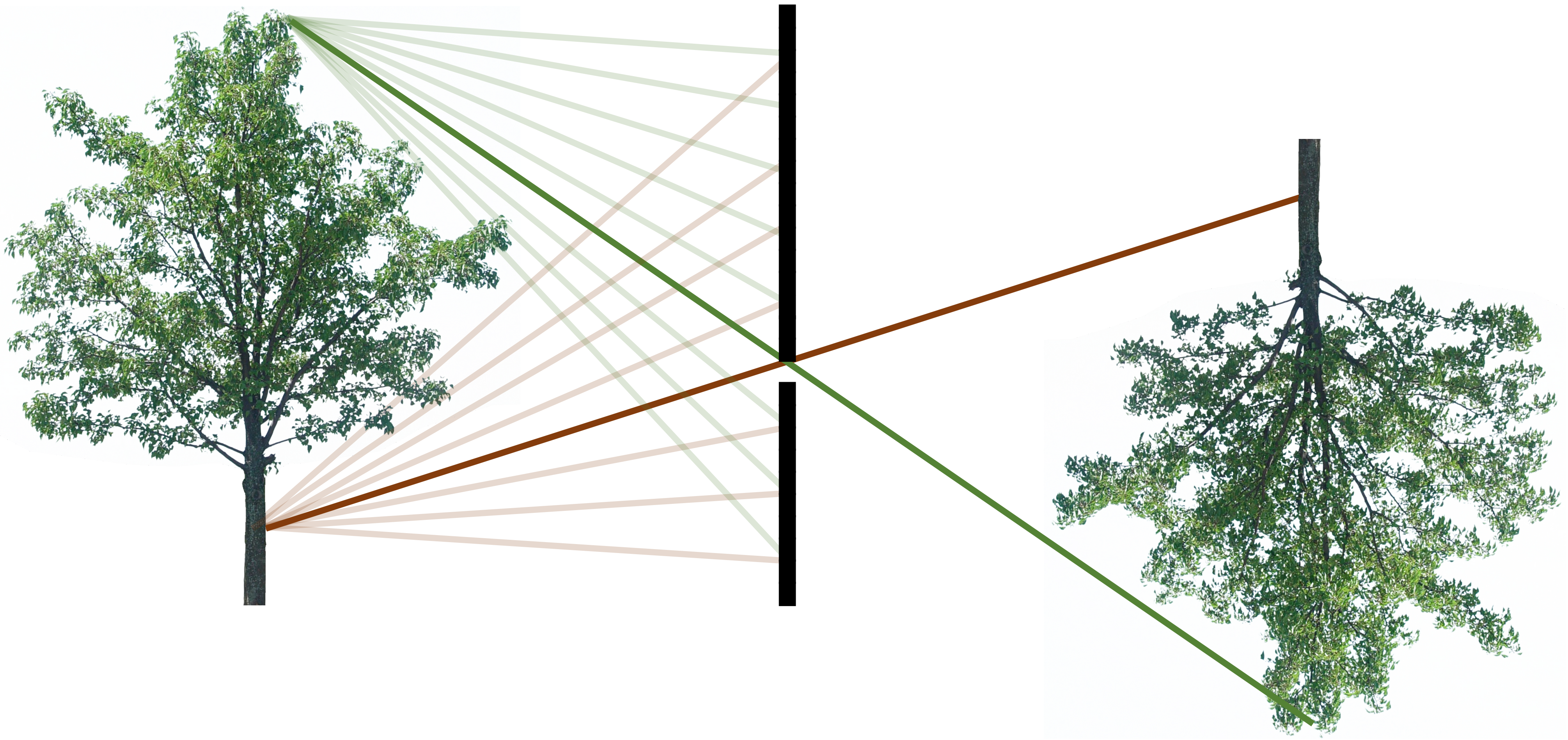
We can capture the same set of rays by using a pinhole camera from multiple viewpoints

Measuring rays



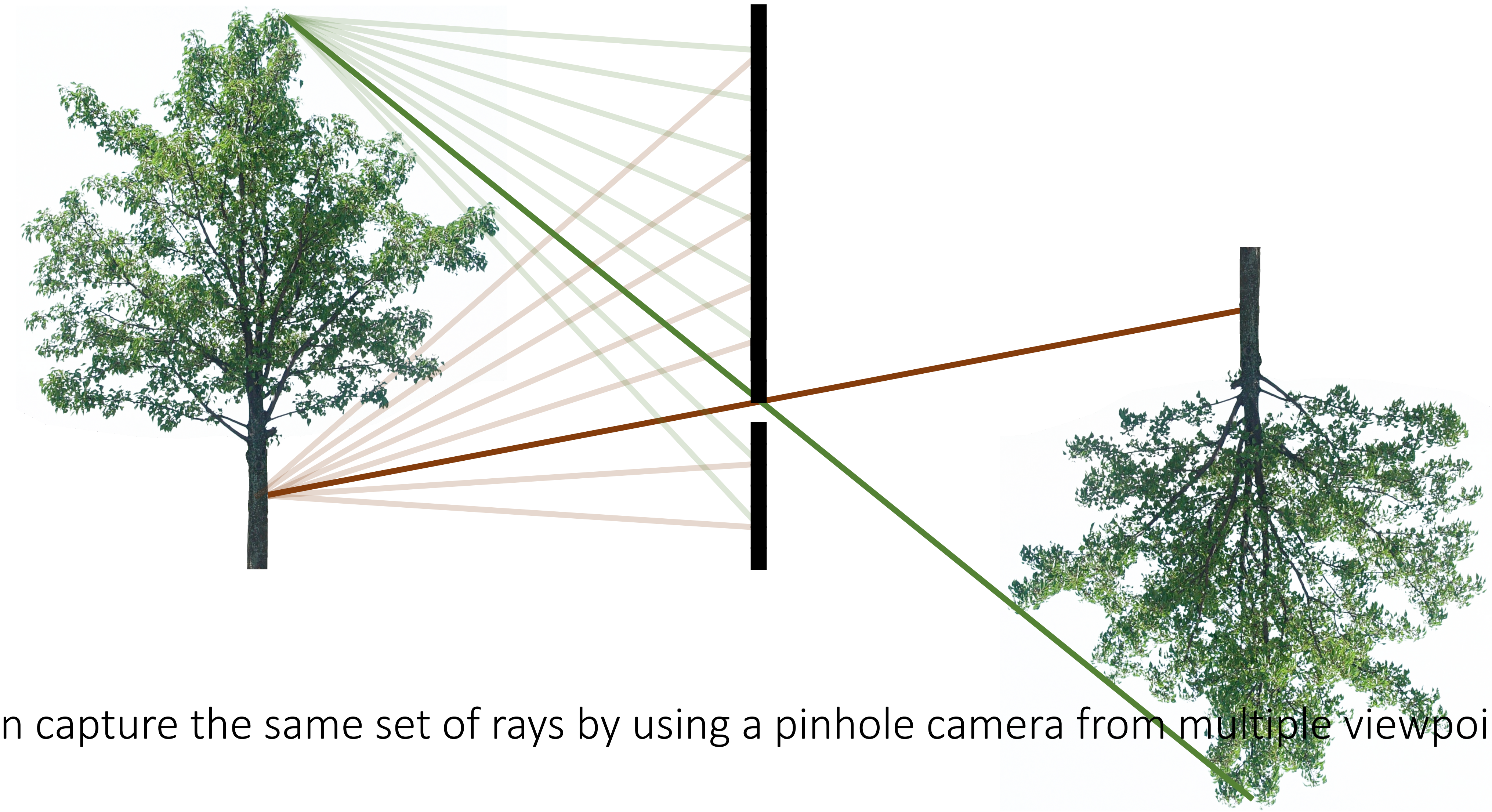
We can capture the same set of rays by using a pinhole camera from multiple viewpoints

Measuring rays



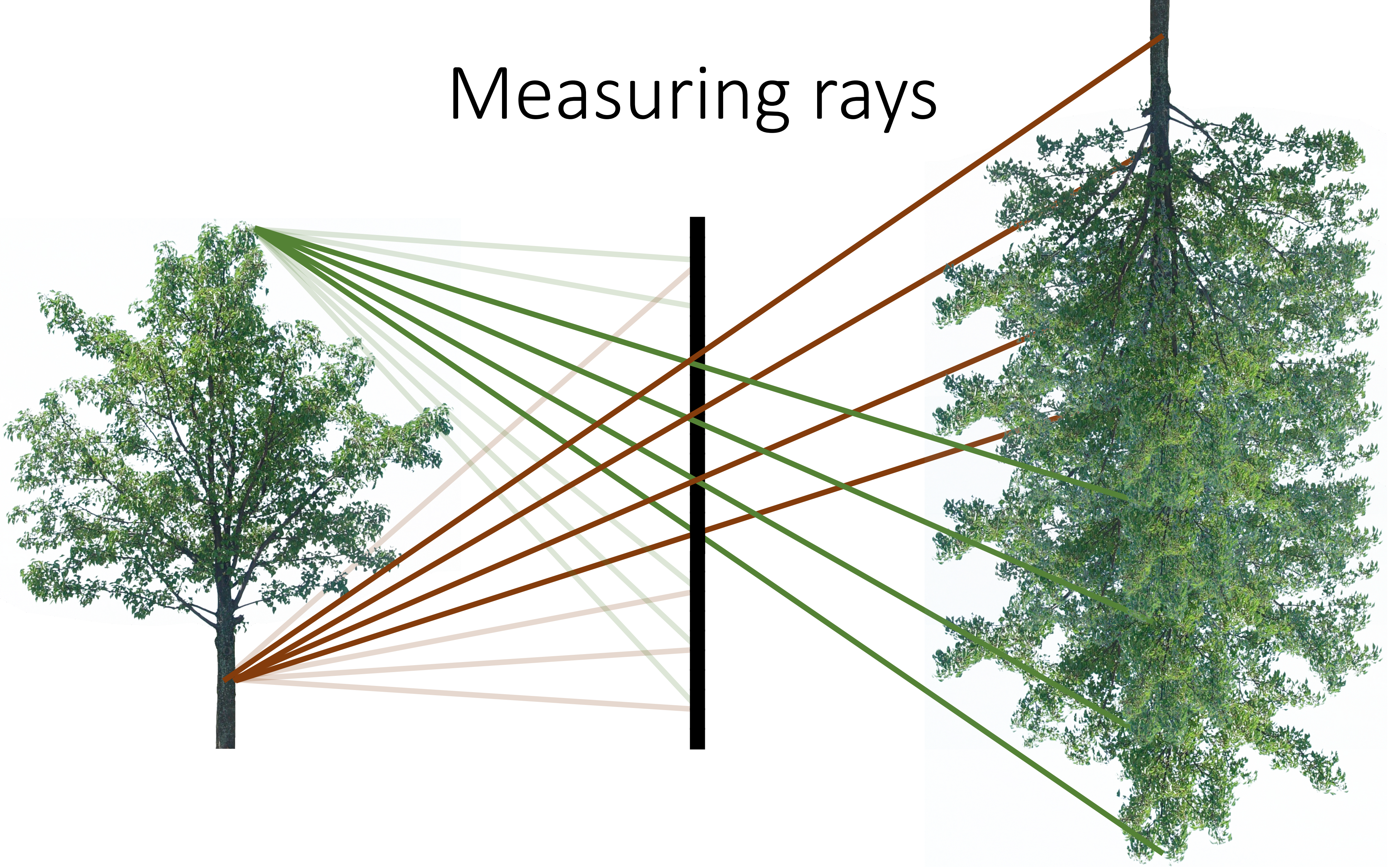
We can capture the same set of rays by using a pinhole camera from multiple viewpoints

Measuring rays



We can capture the same set of rays by using a pinhole camera from multiple viewpoints

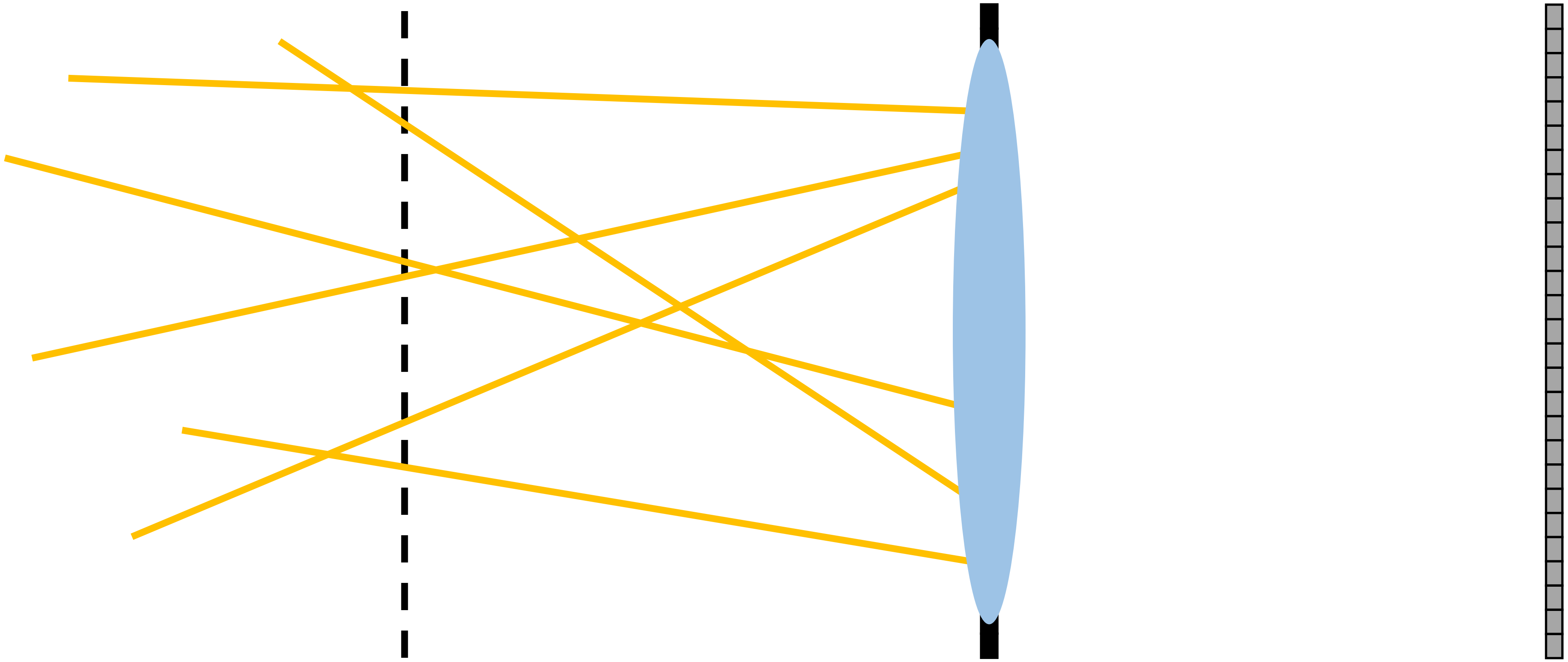
Measuring rays



We can capture the same set of rays by using a pinhole camera from multiple viewpoints

- How would you merge these images into a lens-based, defocused image?

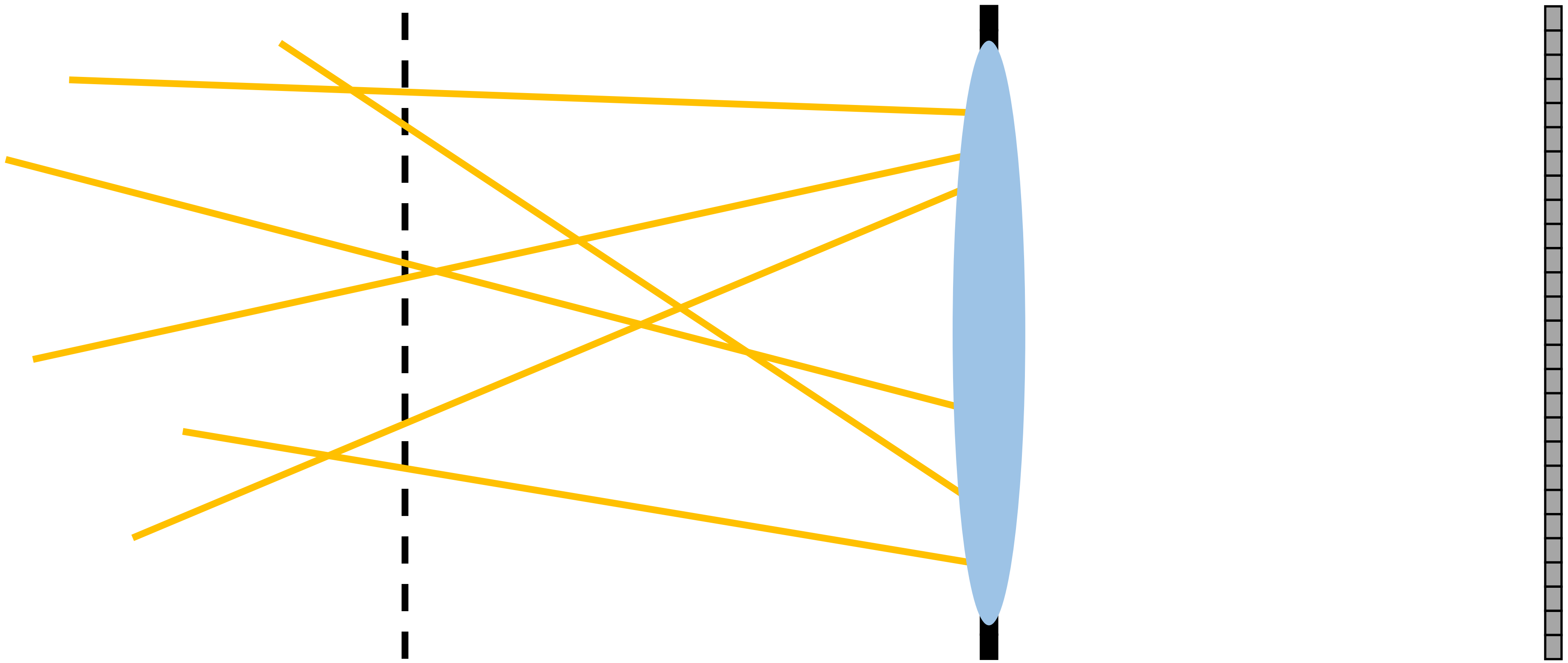
Lightfield: all rays in a scene



What is the dimension of the lightfield?

Parameterize every ray based on its intersections with two planes.

Lightfield: all rays in a scene

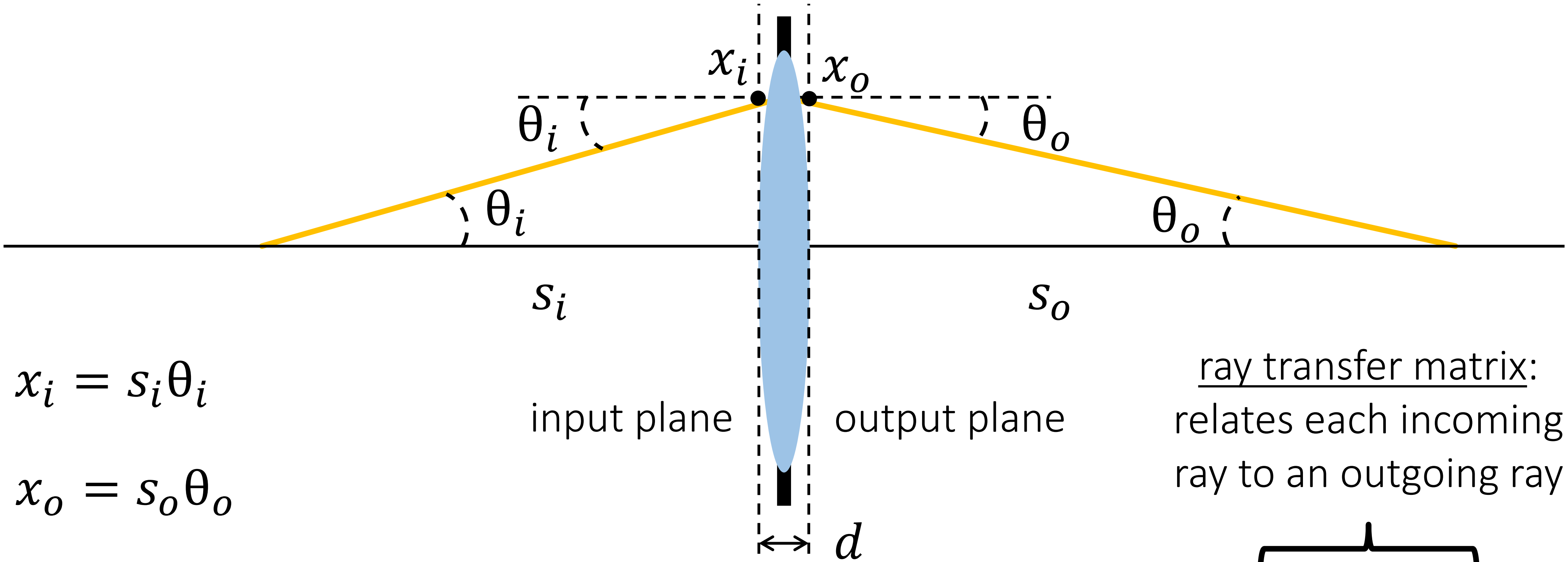


4-dimensional function L(u, v, s, t)

Parameterize every ray based on its intersections with two planes.

- How else can we parameterize the lightfield?
- Where else have we seen it?

Reminder from lecture 4: paraxial optics



$x_i = s_i\theta_i$

$x_o = s_o\theta_o$

$$\frac{1}{s_o} + \frac{1}{s_i} = \frac{1}{f}$$

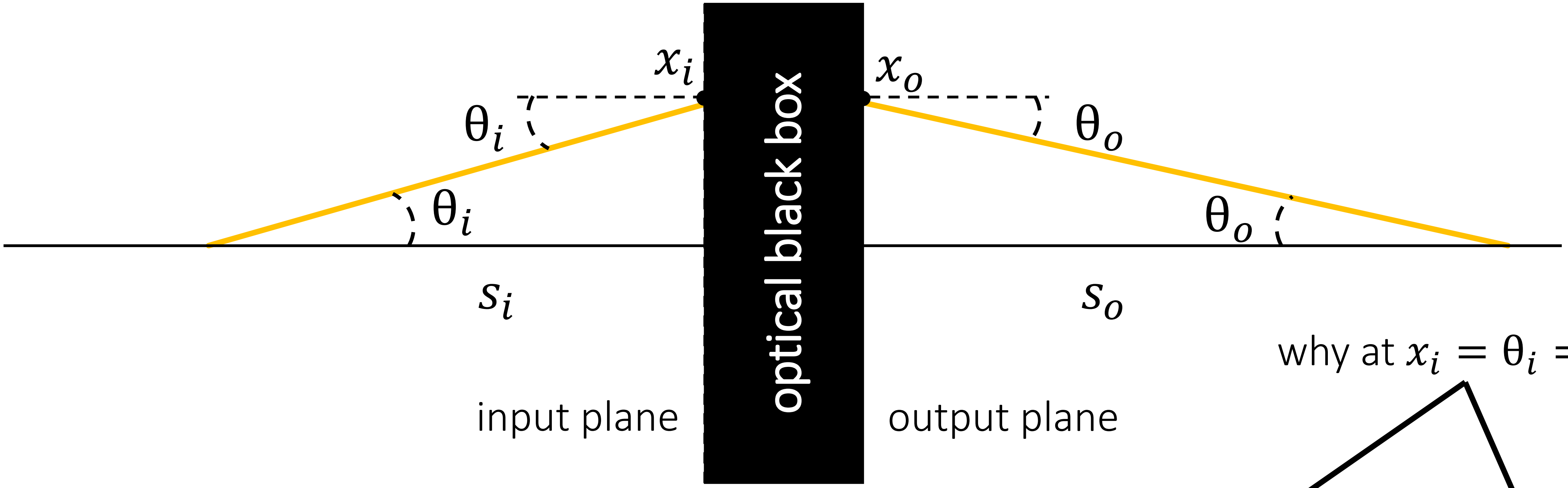
$d \simeq 0 \Rightarrow x_i = x_o$

Putting it all together, we can write:

<u>ray transfer matrix</u>: relates each incoming ray to an outgoing ray

$$\begin{bmatrix} x_o \\ \theta_o \end{bmatrix} = \overbrace{\begin{bmatrix} 1 & 0 \\ -\frac{1}{f} & 1 \end{bmatrix}}^{\text{ray transfer matrix}} \begin{bmatrix} x_i \\ \theta_i \end{bmatrix}$$

Reminder from lecture 4: ray transfer matrix analysis



why at $x_i = \theta_i = 0$?

Under paraxial approximation:

$$\begin{bmatrix} x_o \\ \theta_o \end{bmatrix} = \begin{bmatrix} f(x_i, \theta_i) \\ g(x_i, \theta_i) \end{bmatrix} \simeq \underbrace{\begin{bmatrix} A & B \\ C & D \end{bmatrix}} \begin{bmatrix} x_i \\ \theta_i \end{bmatrix}$$

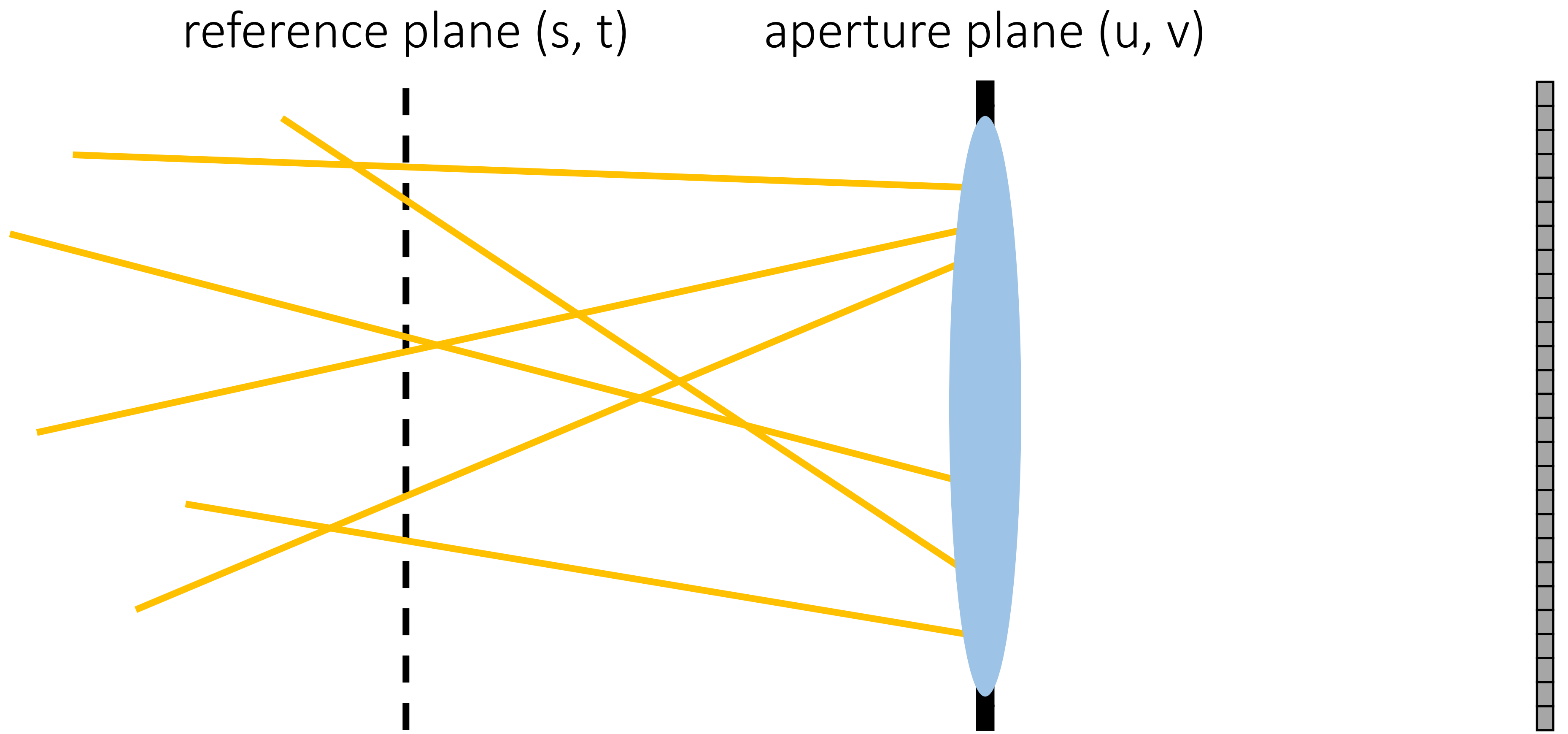
where

$$A = \left. \frac{\partial f}{\partial x_i} \right|_{x_i = \theta_i = 0} \qquad B = \left. \frac{\partial f}{\partial \theta_i} \right|_{x_i = \theta_i = 0}$$

$$C = \left. \frac{\partial g}{\partial x_i} \right|_{x_i = \theta_i = 0} \qquad D = \left. \frac{\partial g}{\partial \theta_i} \right|_{x_i = \theta_i = 0}$$

definition of ray transfer matrix, a.k.a. ABCD matrix

Lightfield: all rays in a scene

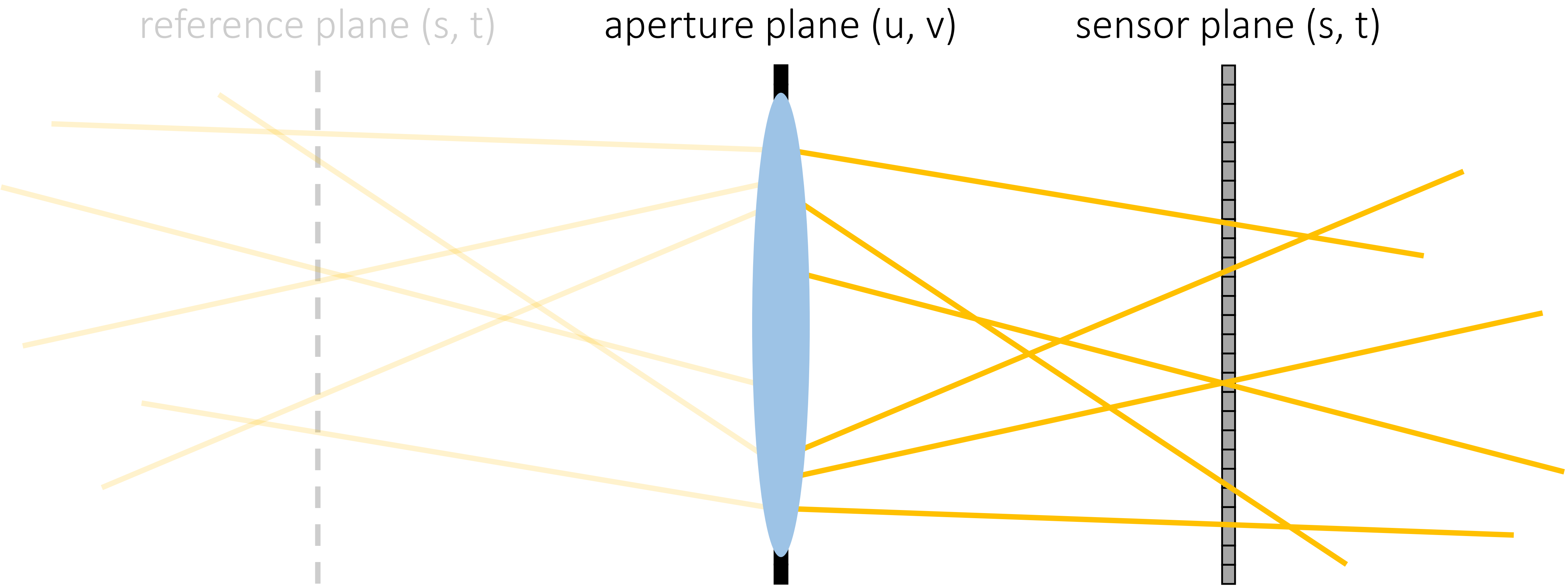


4-dimensional function L(u, v, s, t)

Parameterize every ray based on its intersections with two planes.

- How else can we parameterize the lightfield?
- Where else have we seen it?

Lightfield: all rays in a scene

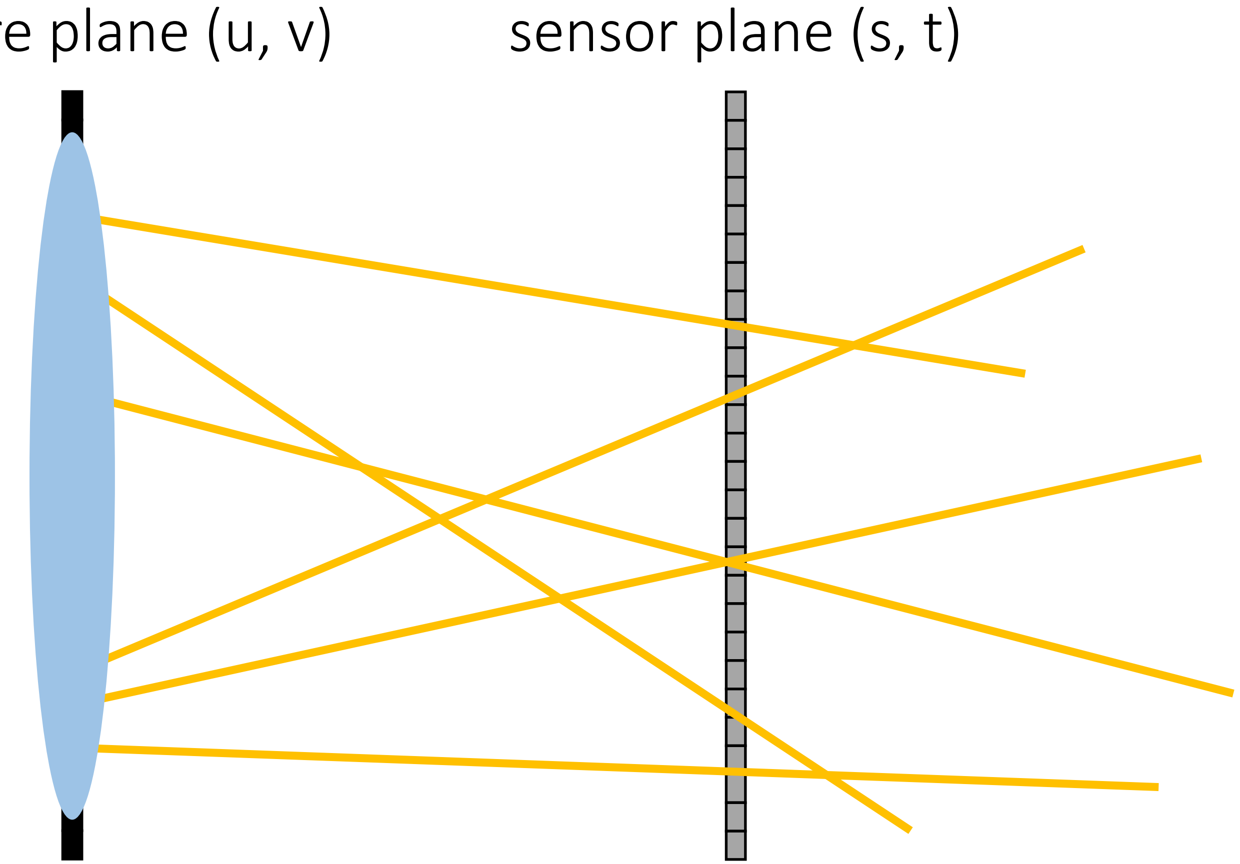


4-dimensional function L(u, v, s, t)
(conjugate of scene-based function)

Parameterize every ray based on its intersections with two planes.

Lightfield slices

What does $L(u = u_o, v = v_o, s, t)$ look like?

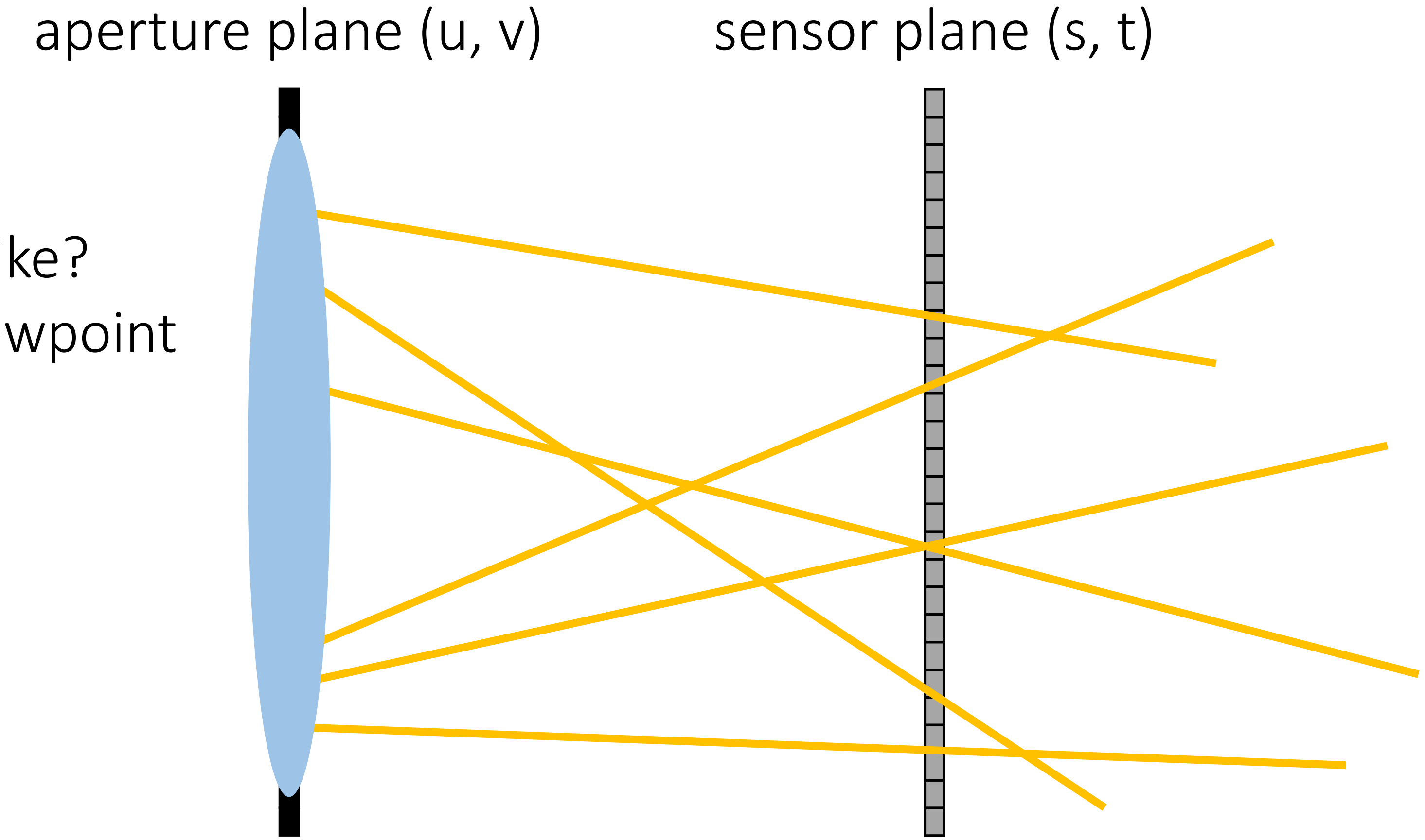


4-dimensional function $L(u, v, s, t)$
(conjugate of scene-based function)

Lightfield slices

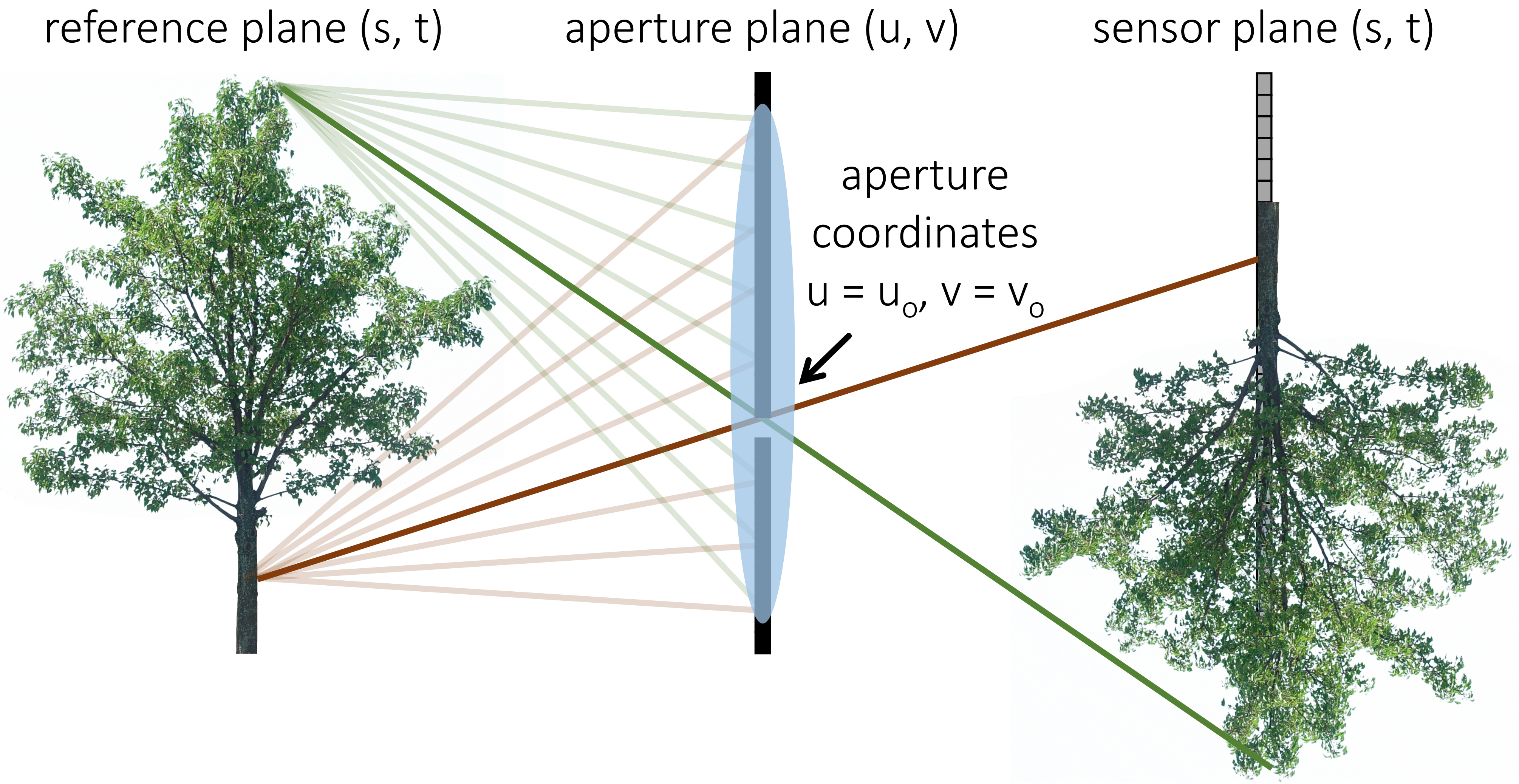
What does $L(u = u_o, v = v_o, s, t)$ look like?

- a pinhole image from a certain viewpoint



4-dimensional function $L(u, v, s, t)$
(conjugate of scene-based function)

Lightfield slices



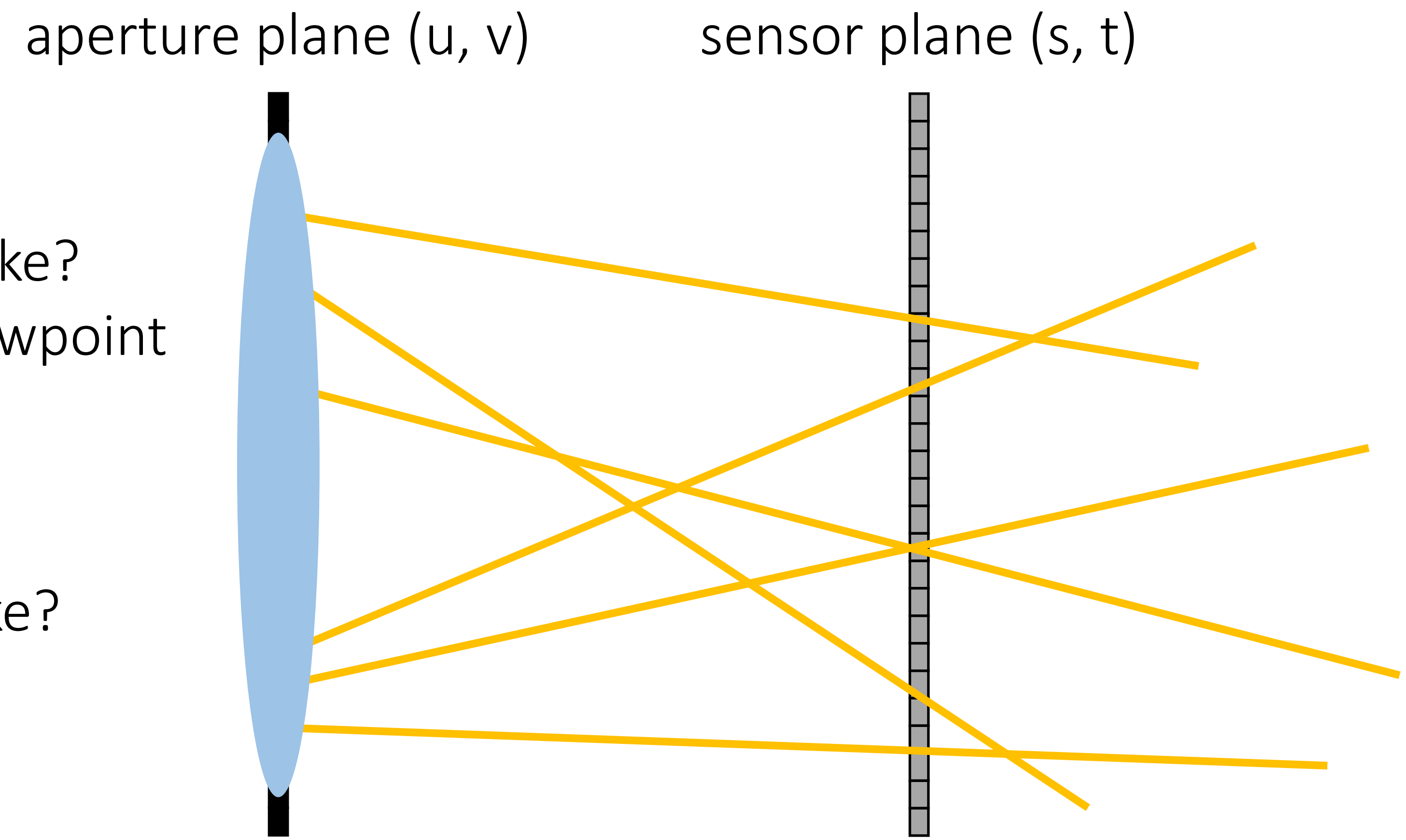
Lightfield slice $L(u = u_o, v = v_o, s, t)$

Lightfield slices

What does $L(u = u_o, v = v_o, s, t)$ look like?

- a pinhole image from a certain viewpoint

What does $L(u, v, s = s_o, t = t_o)$ look like?

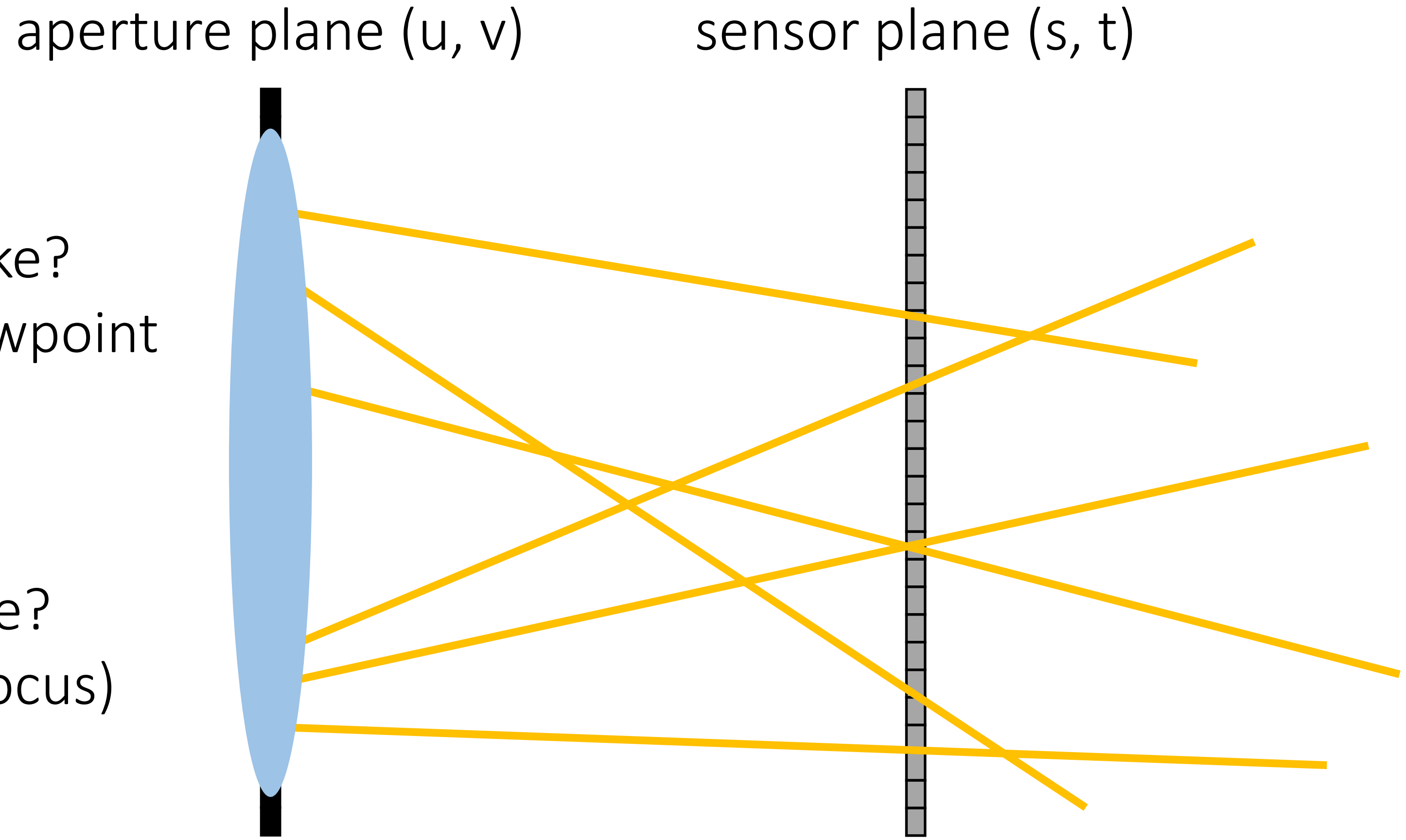


4-dimensional function $L(u, v, s, t)$
(conjugate of scene-based function)

Lightfield slices

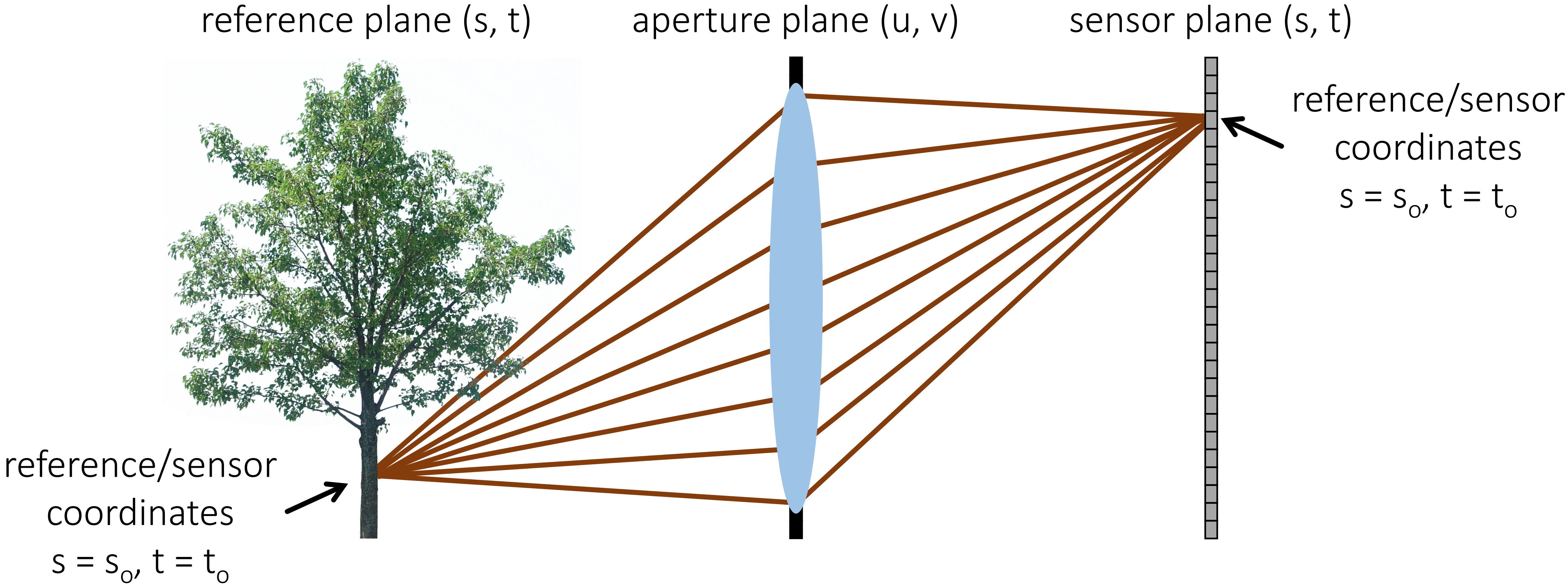
What does $L(u = u_o, v = v_o, s, t)$ look like?
- a pinhole image from a certain viewpoint

What does $L(u, v, s = s_o, t = t_o)$ look like?
- radiance emitted by a certain (in-focus) point at various directions



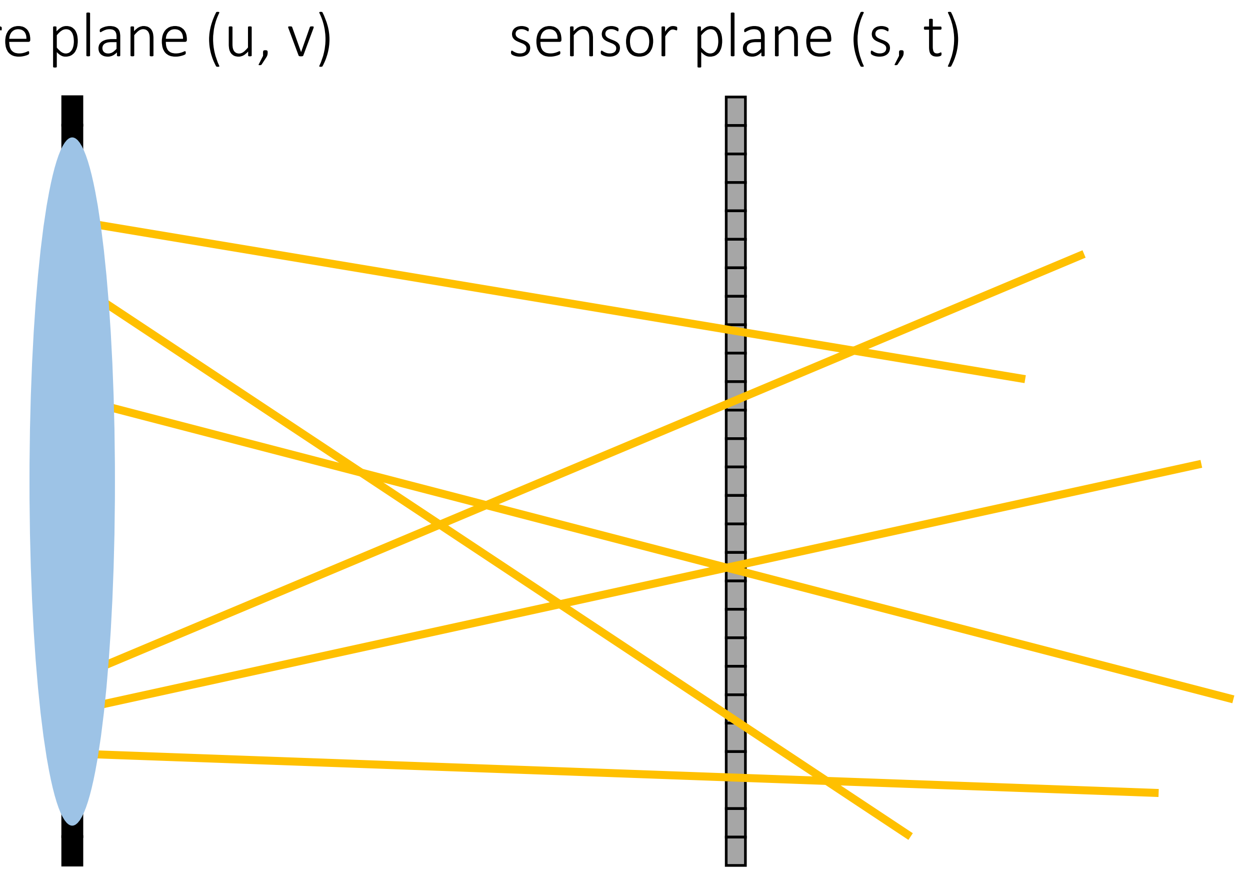
4-dimensional function $L(u, v, s, t)$
(conjugate of scene-based function)

Lightfield slices



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

Lightfield slices

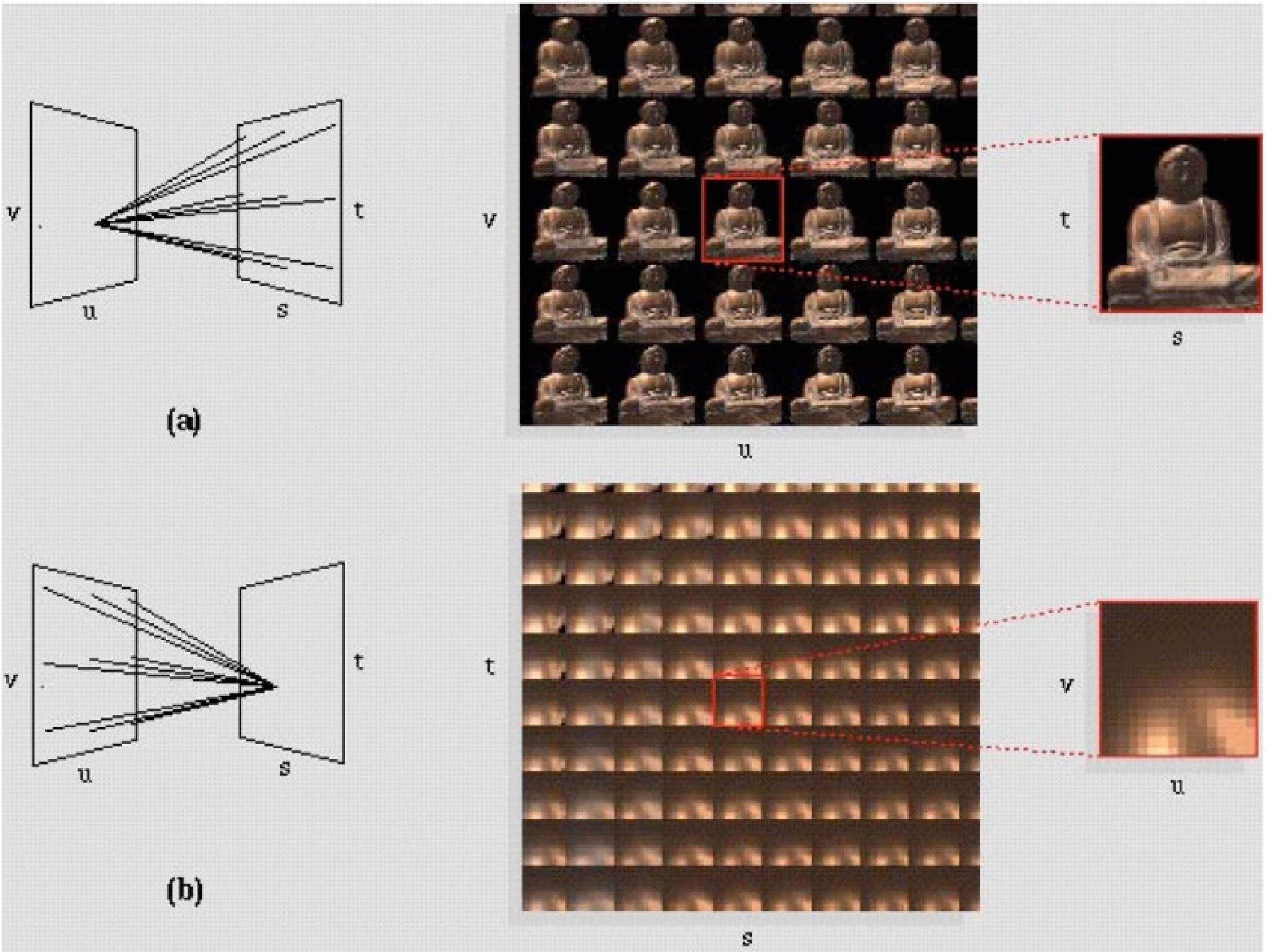


What does $L(u = u_o, v = v_o, s, t)$ look like?
- a pinhole image from a certain viewpoint

What does $L(u, v, s = s_o, t = t_o)$ look like?
- radiance emitted by a certain (in-focus) point in various directions

4-dimensional function $L(u, v, s, t)$
(conjugate of scene-based function)

Lightfield visualization

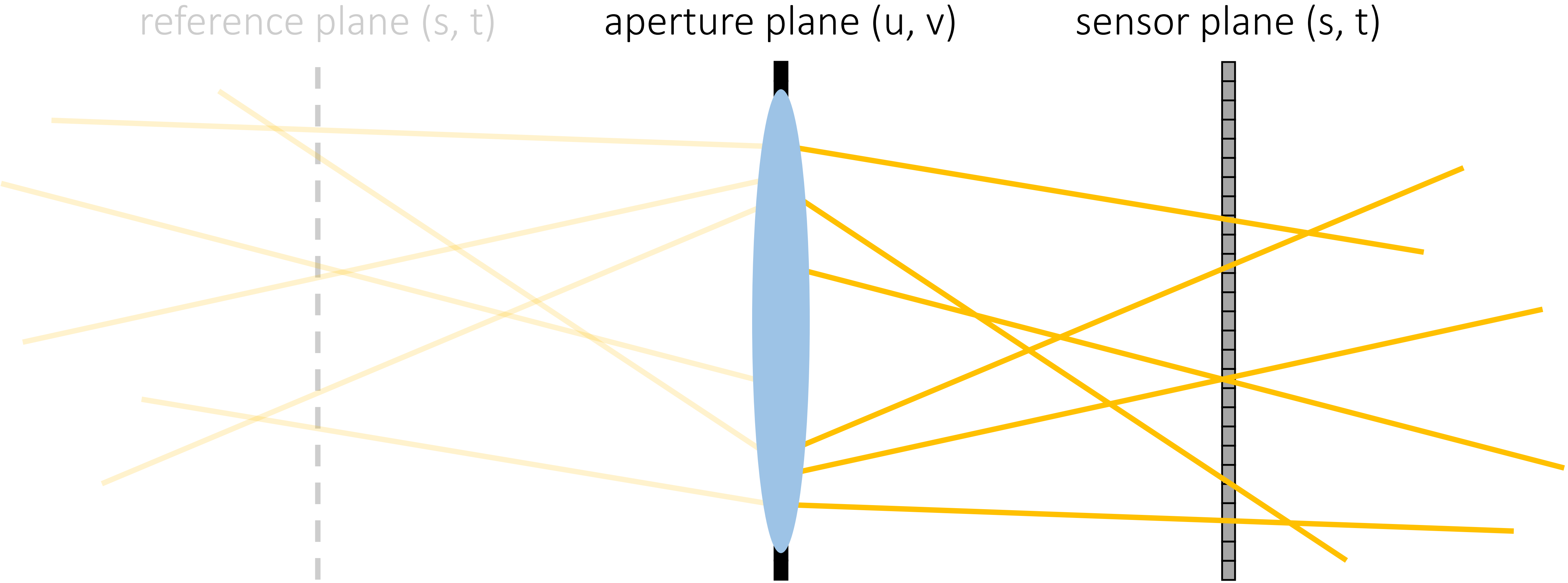


$L(u = u_o, v = v_o, s, t)$ is a pinhole image from a certain viewpoint

$L(u, v, s = s_o, t = t_o)$ is the radiance emitted by a certain (in-focus) point at various directions

Demo:
http://lightfield.stanford.edu/lfs.html

Lightfield: all rays in a scene

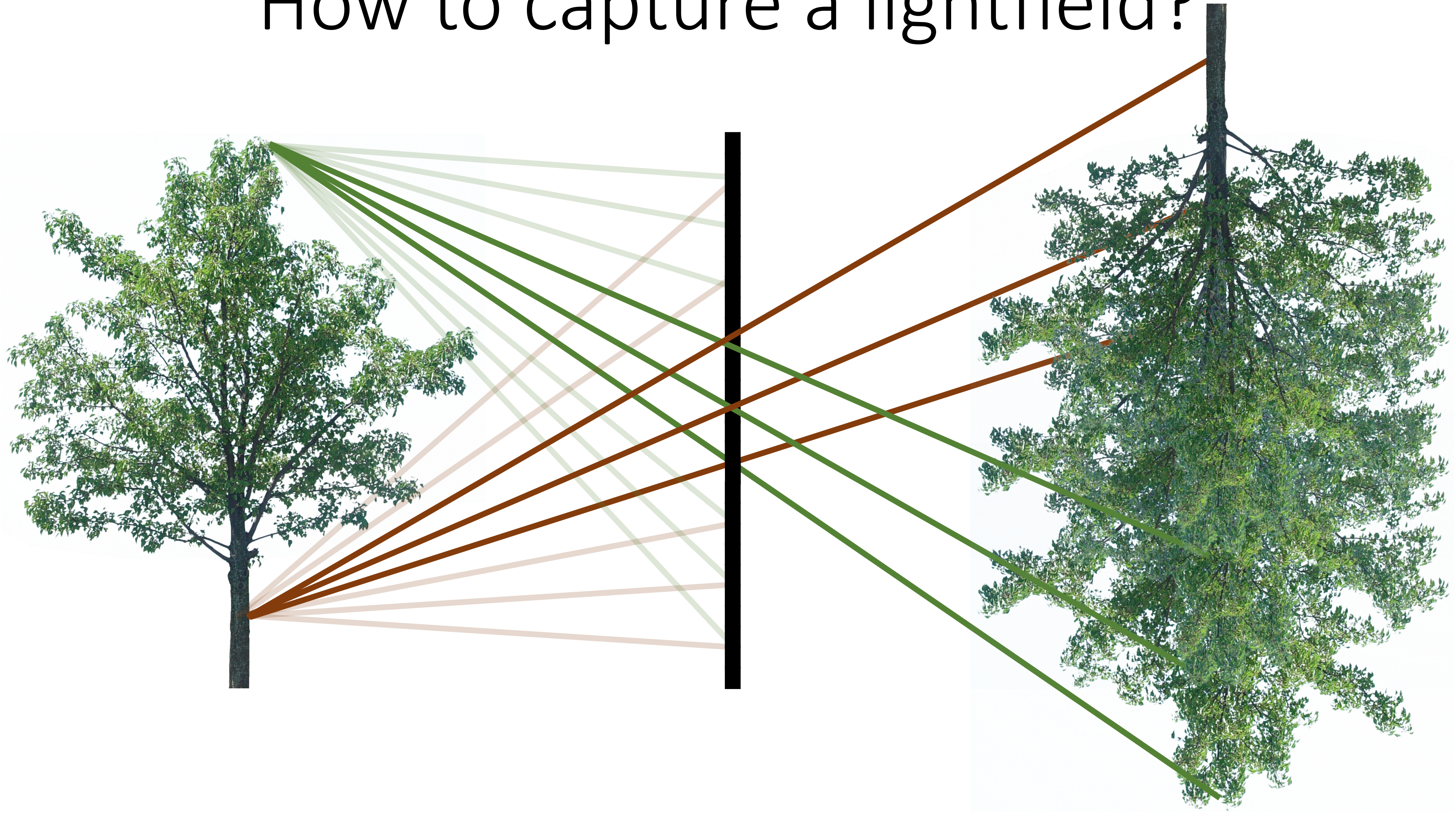


4-dimensional function L(u, v, s, t)
(conjugate of scene-based function)

How can you capture the lightfield of a scene?

Measuring lightfields

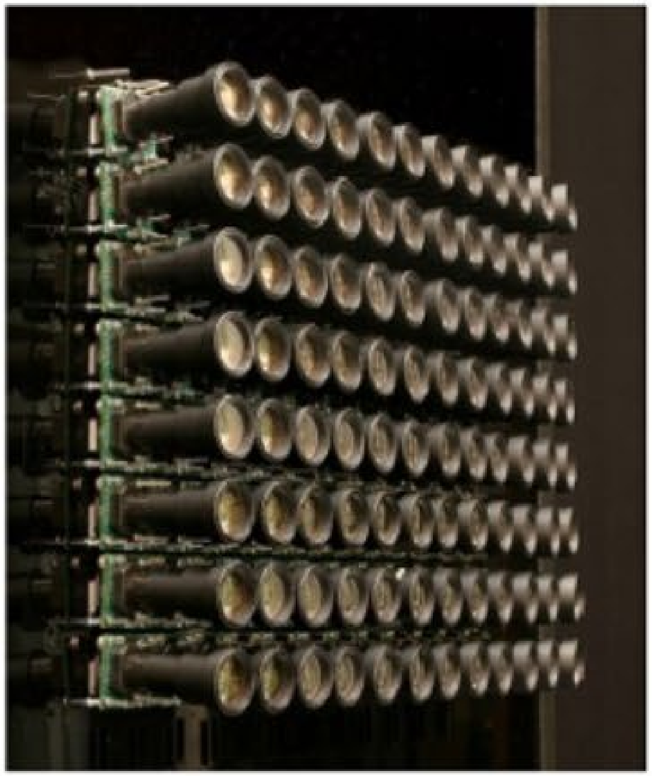
How to capture a lightfield?

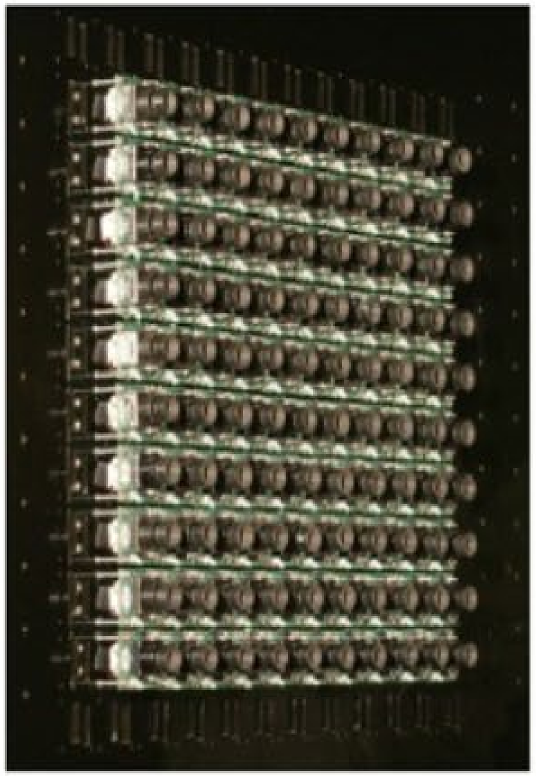


How can you do this?

Option 1: use multiple cameras

Stanford camera array



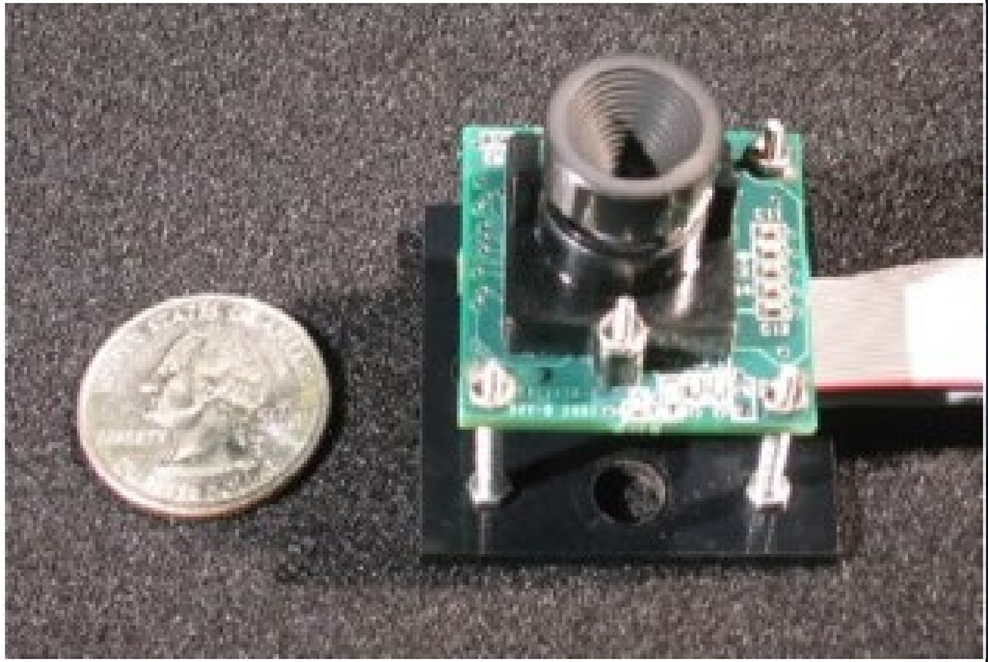


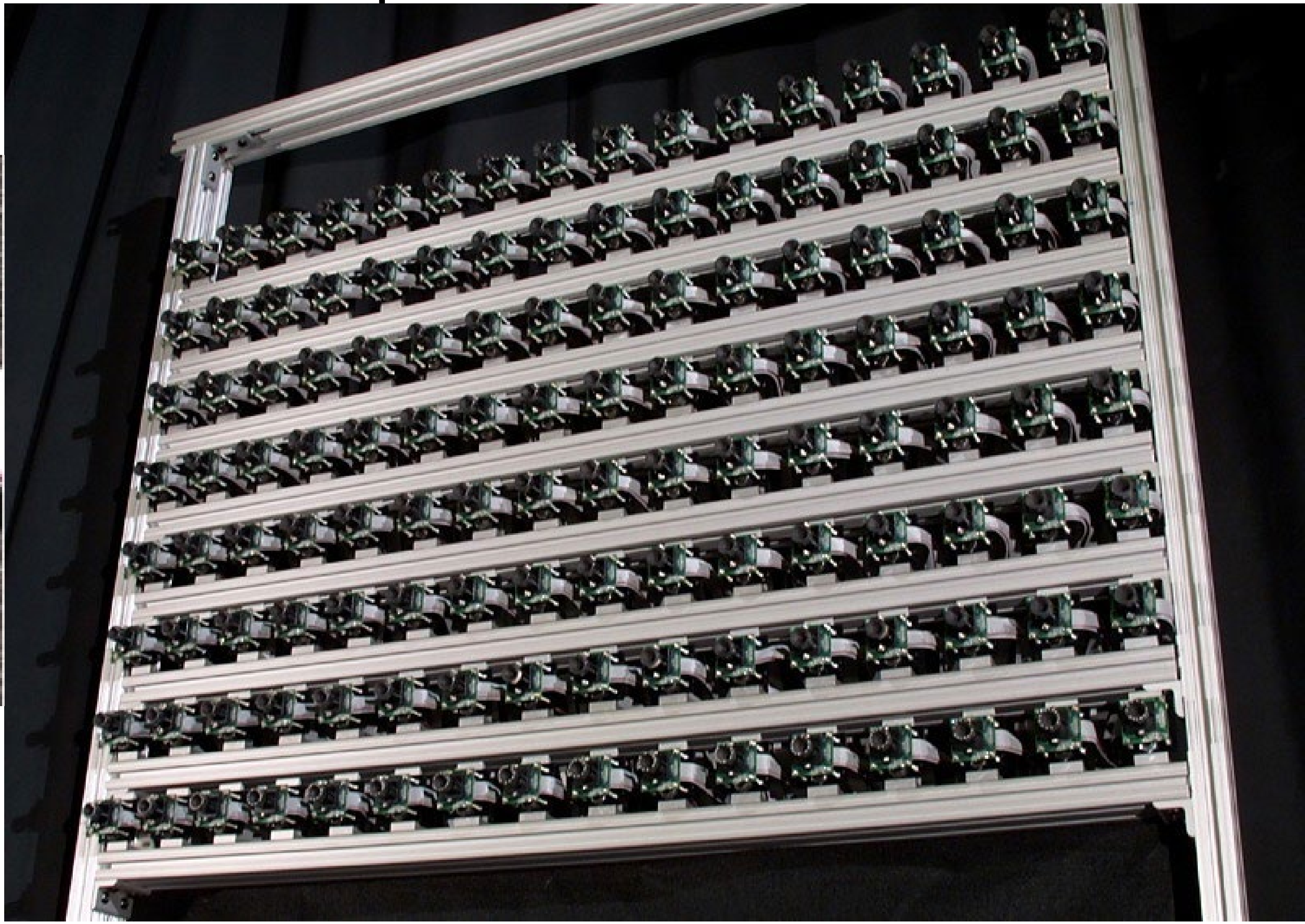
(“synthetic aperture”)

[Willburn et al., SIGGRAPH 2005]

Option 1: use multiple cameras

Stanford camera array





What kind of lens would you use for this?

(“synthetic aperture”)

[Willburn et al., SIGGRAPH 2005]

Compound Eye vs Light Field

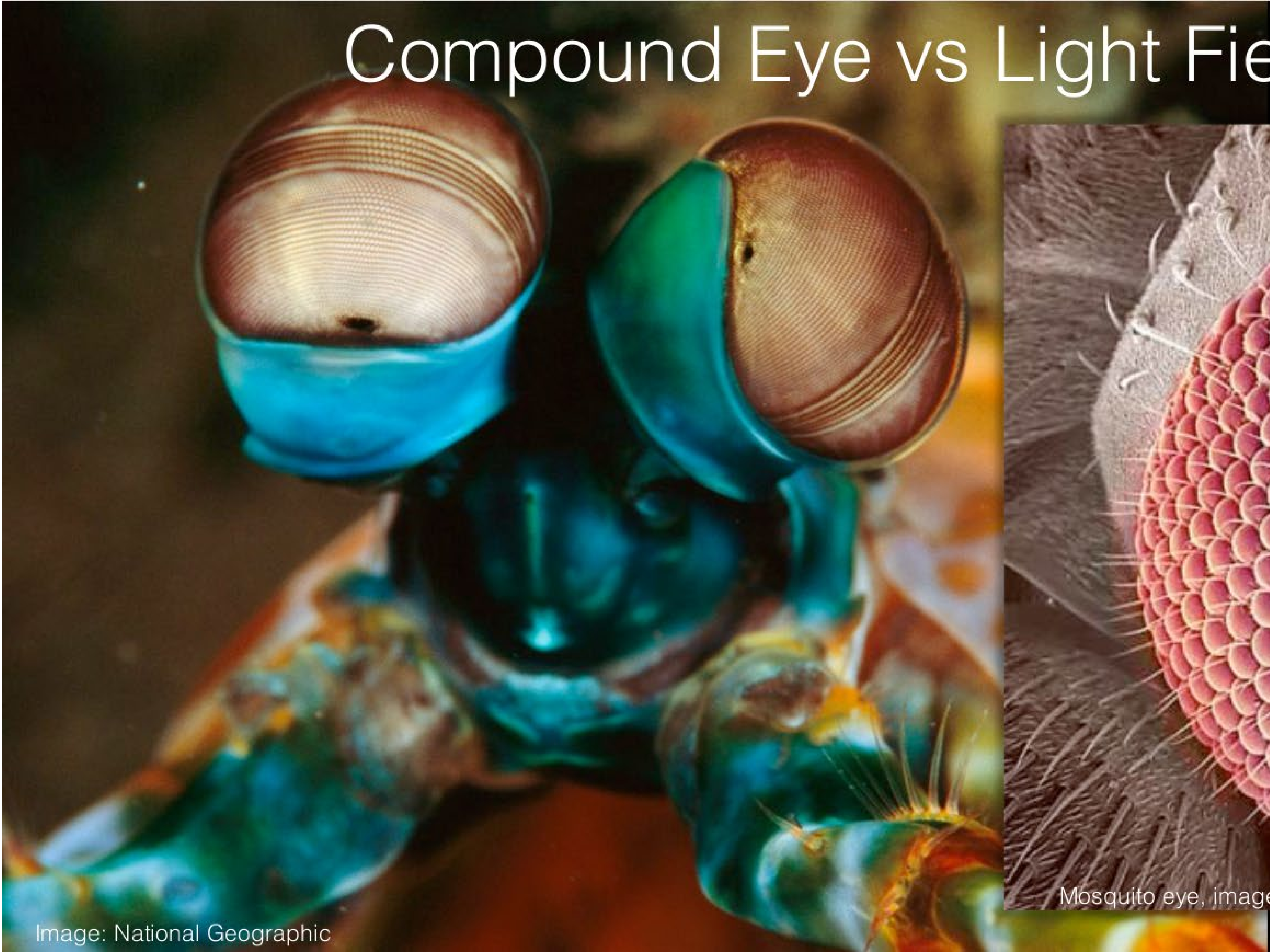
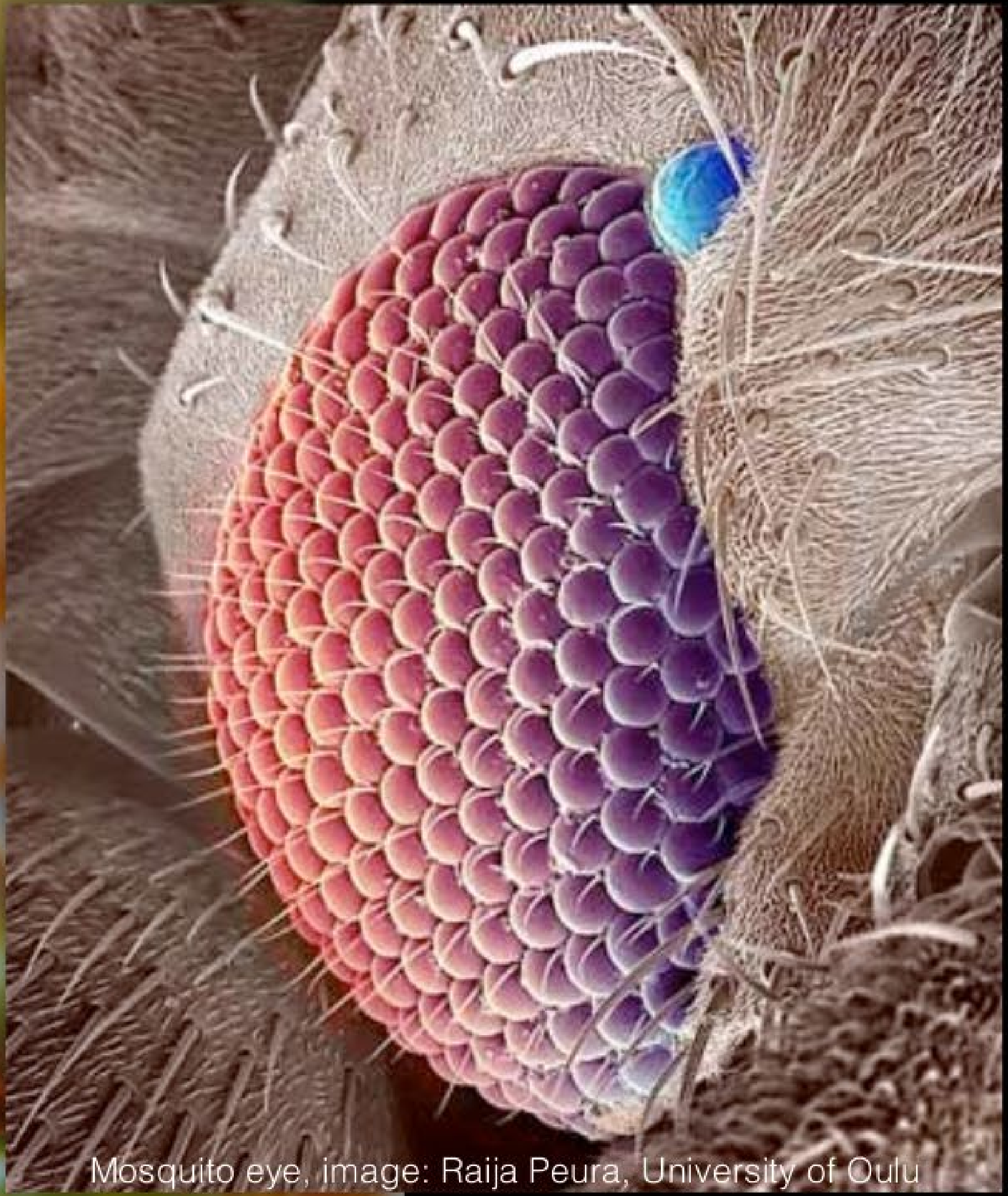


Image: National Geographic

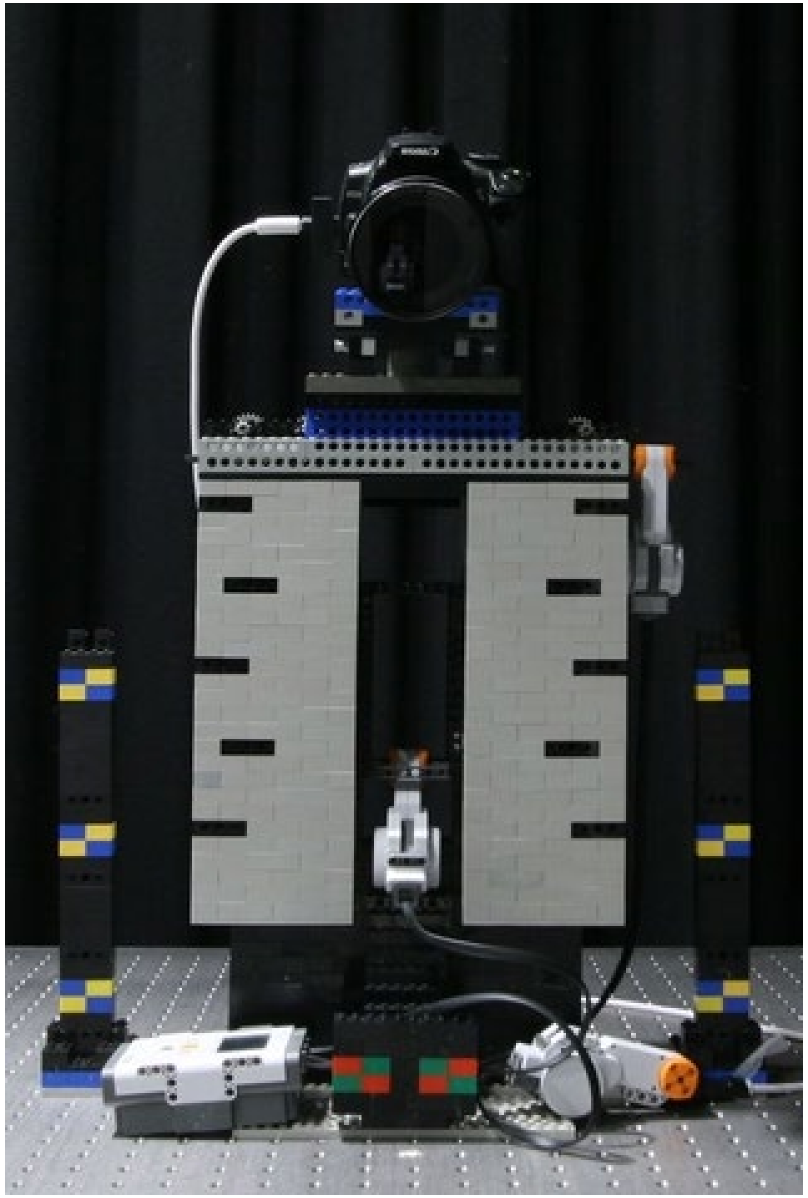


Mosquito eye, image: Raija Peura, University of Oulu

Option 2: take multiple images with one camera

Single camera mounted on LEGO motor. Demo: http://lightfield.stanford.edu/aperture.swf?lightfield=data/self_portrait_lf/preview.zip&zoom=1

How would you move the camera?

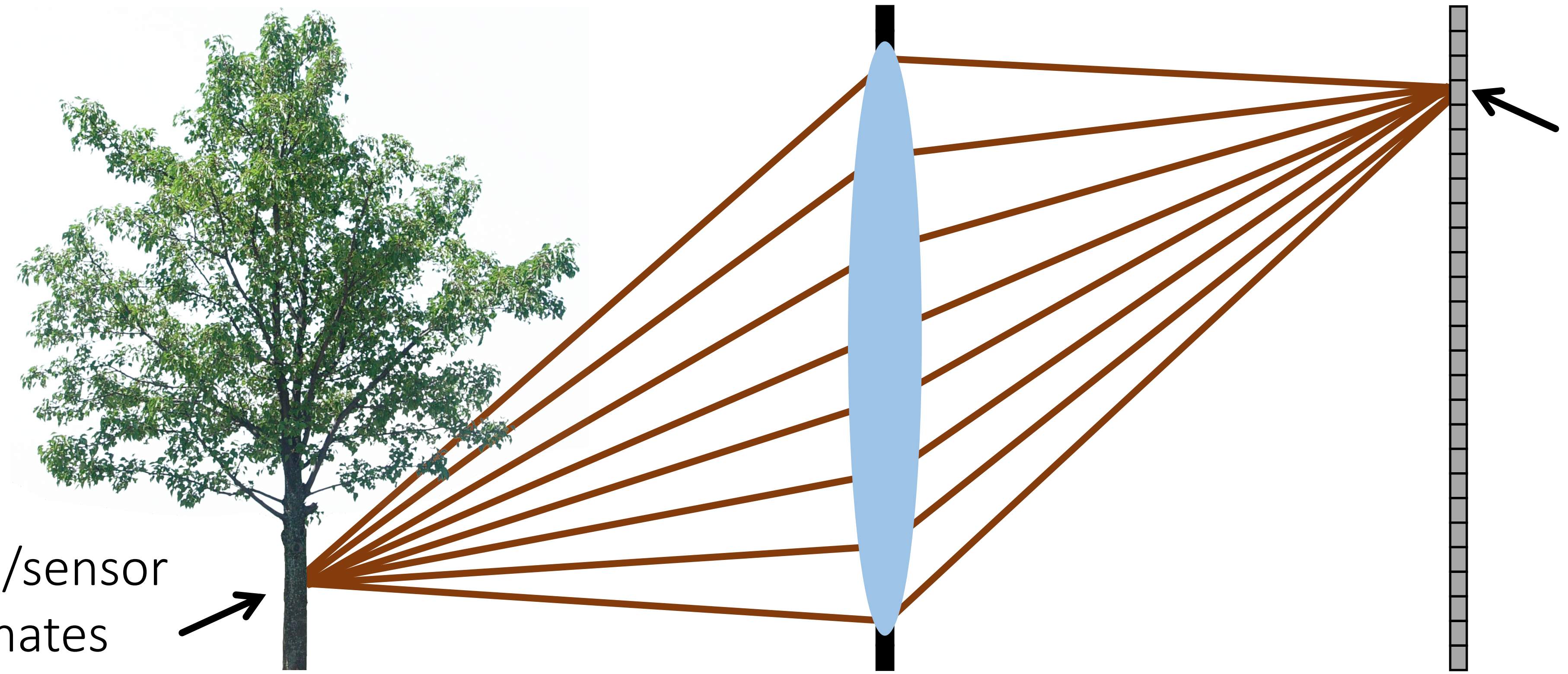


Plenoptic camera

Option 3: use a plenoptic camera

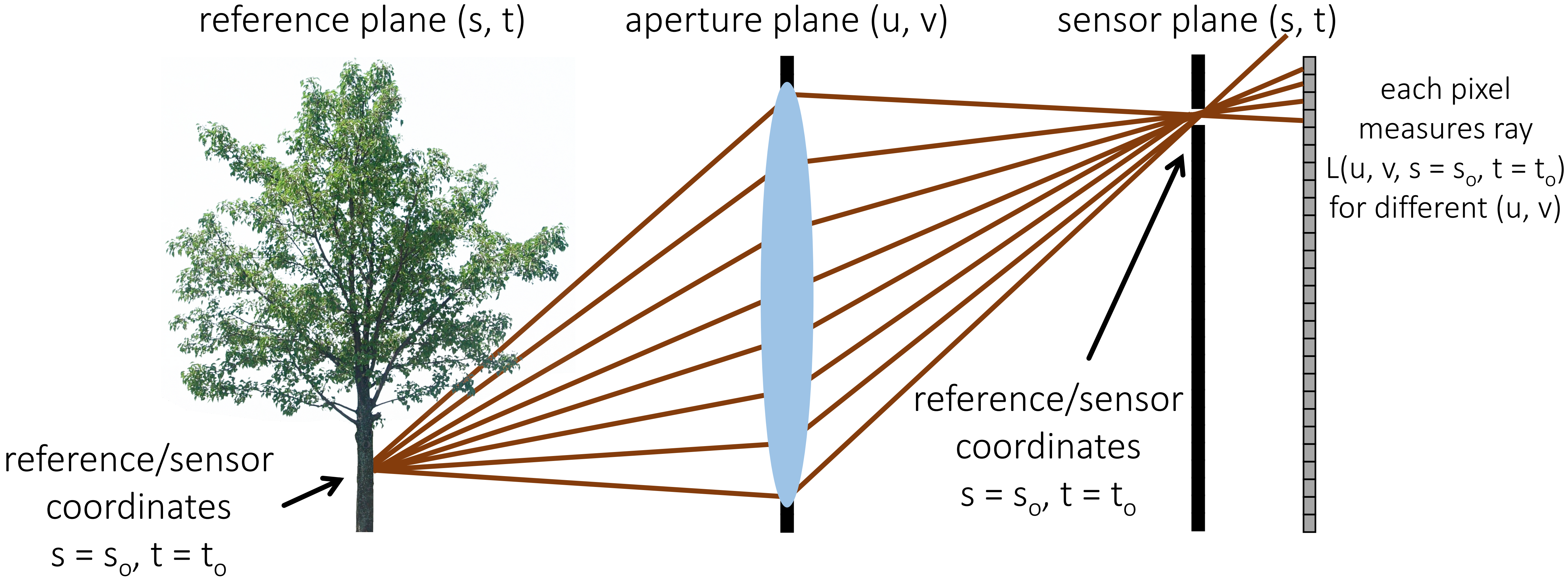
plenoptic = plenus (Latin for “full”) + optic (Greek for “seeing”, in this case)

Making a plenoptic camera



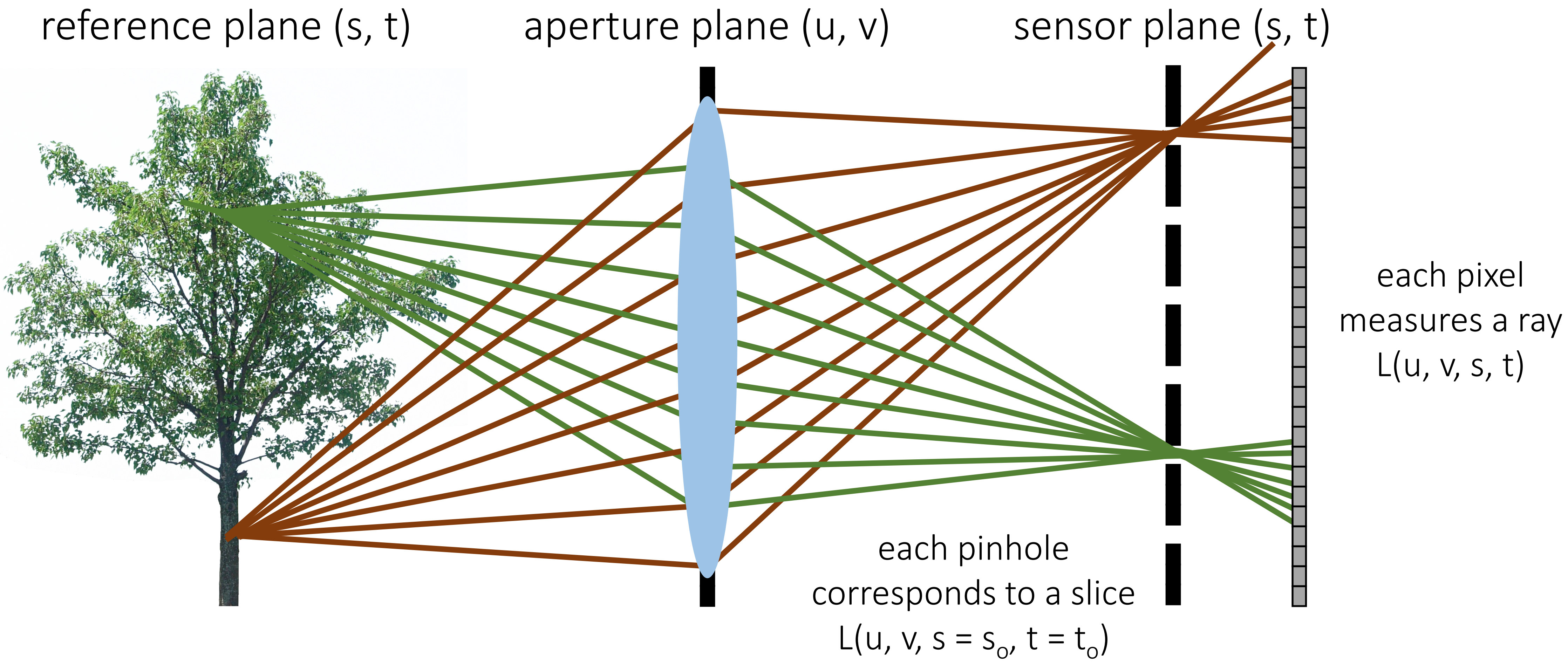
Lightfield slice $L(u, v, s = s_o, t = t_o)$

Making a plenoptic camera



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

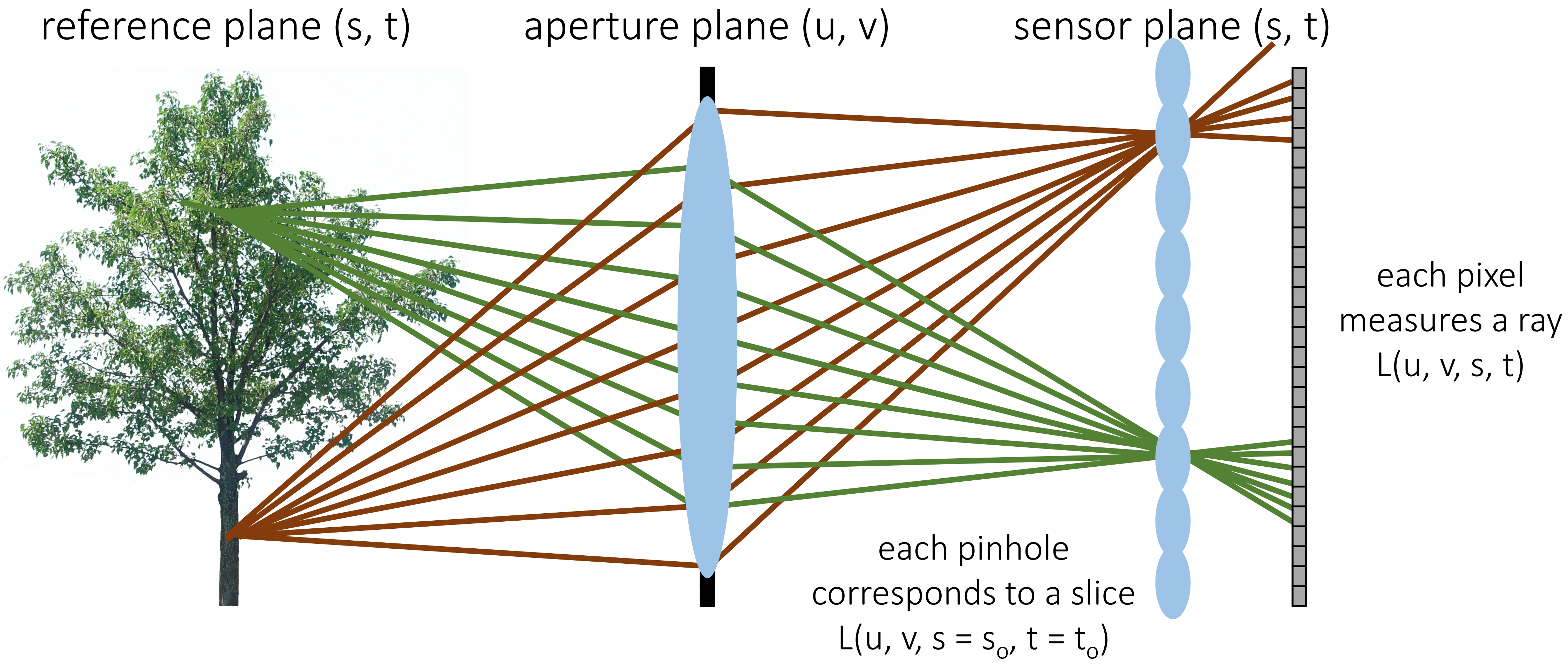
Making a plenoptic camera



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

History of the plenoptic camera

164 SCIENTIFIC AMERICAN August 19, 1911

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'*, as indicated in the diagram. Consider any globule. Rays from the points *M M'* 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'*. 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'*, 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

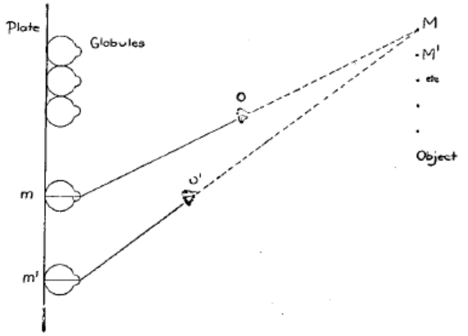


Diagram to illustrate the principle of Lippmann's integral photography.

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.

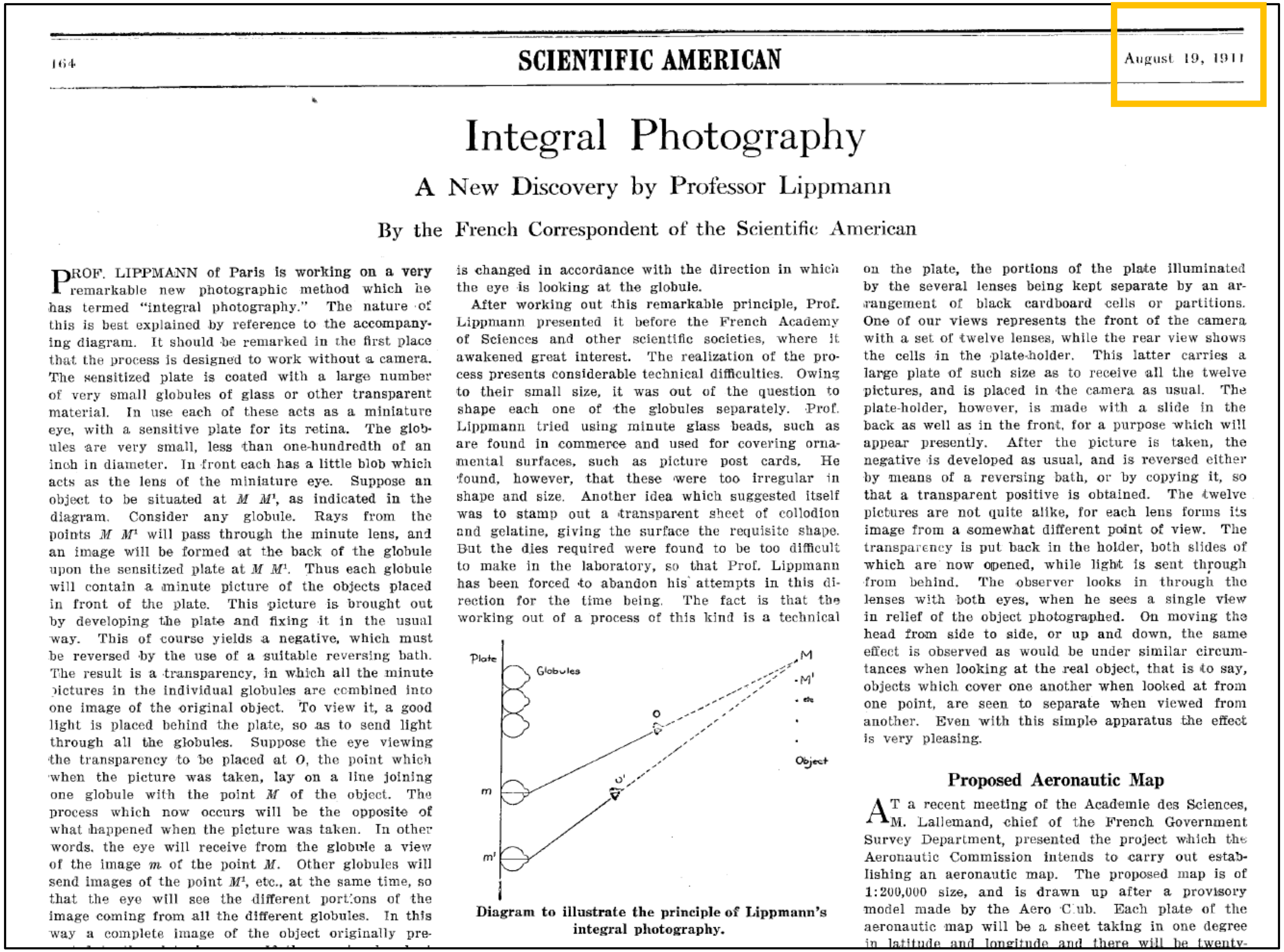
Proposed Aeronautic Map

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-

First conceptualized by Gabriel Lippmann, who called it *integral photography*.

- Original article appeared in French journal three years earlier, what is shown here is an American re-print.

History of the plenoptic camera



164 SCIENTIFIC AMERICAN August 19, 1911

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*', as indicated in the diagram. Consider any globule. Rays from the points *M M*' 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*'. 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*', 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.

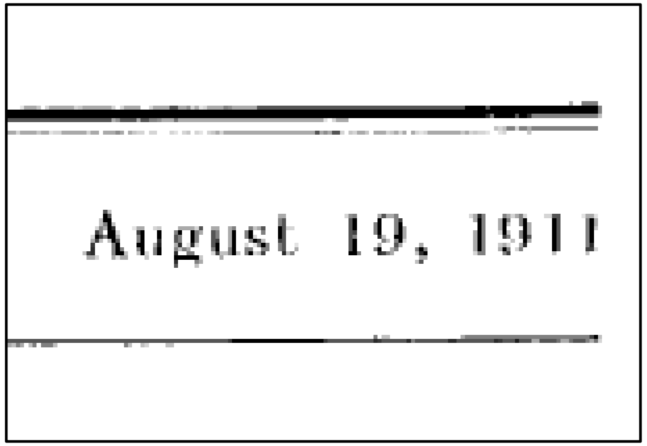
on the plate, the portions of the plate illuminated by the several lenses being kept separate by an arrangement of black cardboard cells or partitions. One of our views represents the front of the camera with a set of twelve lenses, while the rear view shows the cells in the plate-holder. This latter 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-

First conceptualized by Gabriel Lippmann, who called it *integral photography*.

- Original article appeared in French journal three years earlier, what is shown here is an American re-print.



Notice the date: more than a century ago.

History of the plenoptic camera

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.

History of the plenoptic camera

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.

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

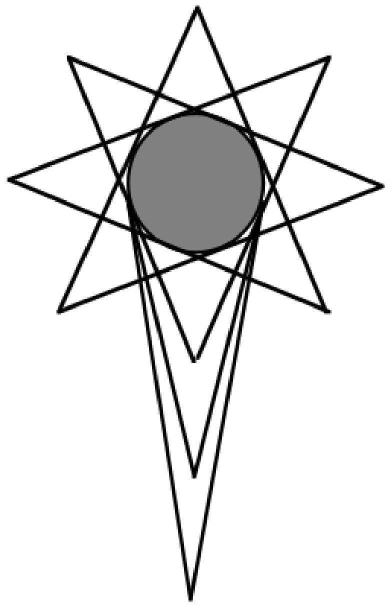


Fig. 1. Diagram from Leonardo's notebooks illustrating the fact that the light rays leaving an object's surface may be considered to form a collection of cones (which Leonardo calls "pyramids"), each cone constituting an image that would be seen by a pinhole camera at a given location.

Re-introduced to computer vision and graphics in 1992, which was also when the term *plenoptic camera* was coined.

History of the plenoptic camera

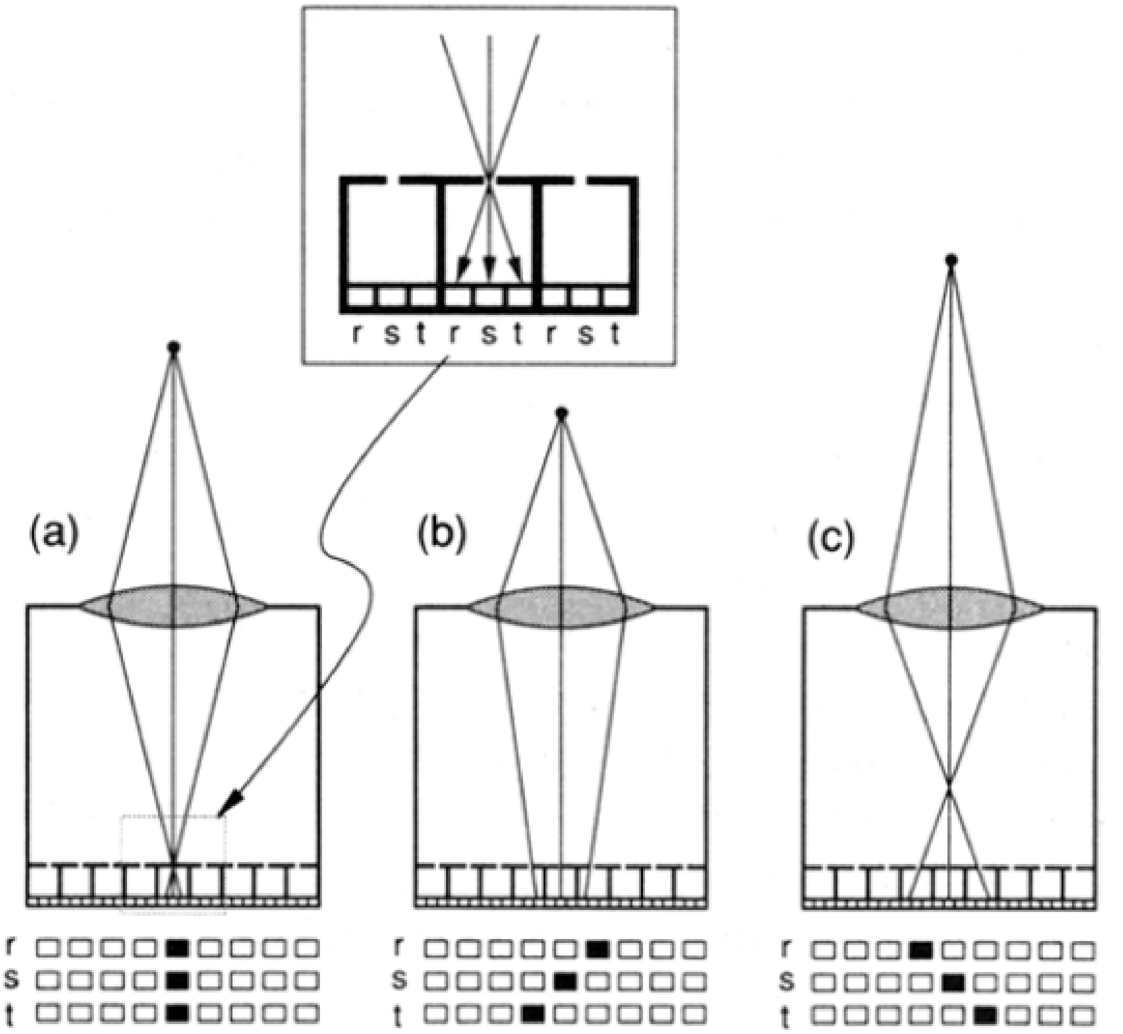


Fig. 5. Array of miniature pinhole cameras placed at the image plane can be used to analyze the structure of the light striking each macropixel.

Figure from the 1992 paper, which shows the version of the plenoptic camera with pinholes instead of lenslets on the image plane.

History of the plenoptic camera

Stanford Tech Report CTSR 2005-02

Light Field Photography with a Hand-held Plenoptic Camera

Ren Ng* Marc Levoy* Mathieu Brédif* Gene Duval† Mark Horowitz* Pat Hanrahan*
*Stanford University †Duval Design

Abstract

This paper presents a camera that samples the 4D light field on its sensor in a single photographic exposure. This is achieved by inserting a microlens array between the sensor and main lens, creating a plenoptic camera. Each microlens measures not just the total amount of light deposited at that location, but how much light arrives along each ray. By re-sorting the measured rays of light to where they would have terminated in slightly different, synthetic cameras, we can compute sharp photographs focused at different depths. We show that a linear increase in the resolution of images under each microlens results in a linear increase in the sharpness of the refocused photographs. This property allows us to extend the depth of field of the camera without reducing the aperture, enabling shorter exposures and lower image noise. Especially in the macrophotography regime, we demonstrate that we can also compute synthetic photographs from a range of different viewpoints. These capabilities argue for a different strategy in designing photographic imaging systems.

To the photographer, the plenoptic camera operates exactly like an ordinary hand-held camera. We have used our prototype to take hundreds of light field photographs, and we present examples of portraits, high-speed action and macro close-ups.

Keywords: Digital photography, light field, microlens array, synthetic photography, refocusing

Externally, our hand-held light field camera looks and operates exactly like a conventional camera: the viewfinder, focusing mechanism, length of exposure, etc. are identical. Internally, we augment the 2D photosensor by placing a microlens array in front of it, as proposed by Adelson and Wang [1992] in their work on the "plenoptic camera" (They did not build this device, but prototyped a non-portable version containing a relay lens.) Each microlens forms a tiny sharp image of the lens aperture, measuring the directional distribution of light at that microlens.

This paper explains the optical recipe of this camera in detail, and develops its theory of operation. We describe an implementation using a medium format digital camera and microlens array. Using this prototype, we have performed resolution experiments that corroborate the limits of refocusing predicted by the theory. Finally, we demonstrate examples of refocusing and view-point manipulation involving close-up macro subjects, human portraits, and high-speed action.

2 Related Work

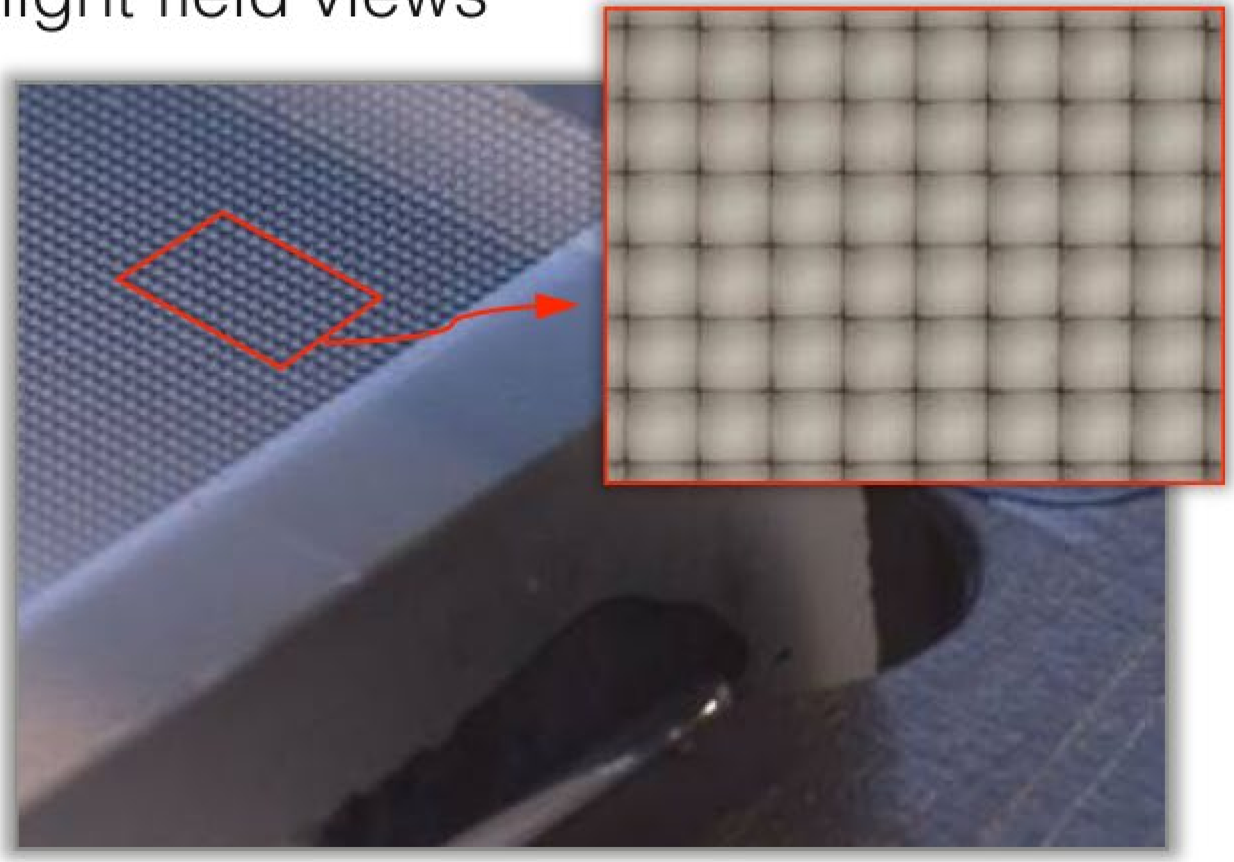
The optical design of our camera is very similar to that of Adelson and Wang's plenoptic camera [1992]. Compared to Adelson and Wang, our prototype contains two fewer lenses, which significantly shortens the optical path, resulting in a portable camera. These differences are explained in more detail Section 3.1 once sufficient technical background has been introduced. The other main differ-

Re-re-introduced to computer vision and graphics in 2005, when the term *light field photography* was also coined.

Prototype plenoptic camera

- predecessor of Lytro
- resolution: 292x292px, 14x14 light field views

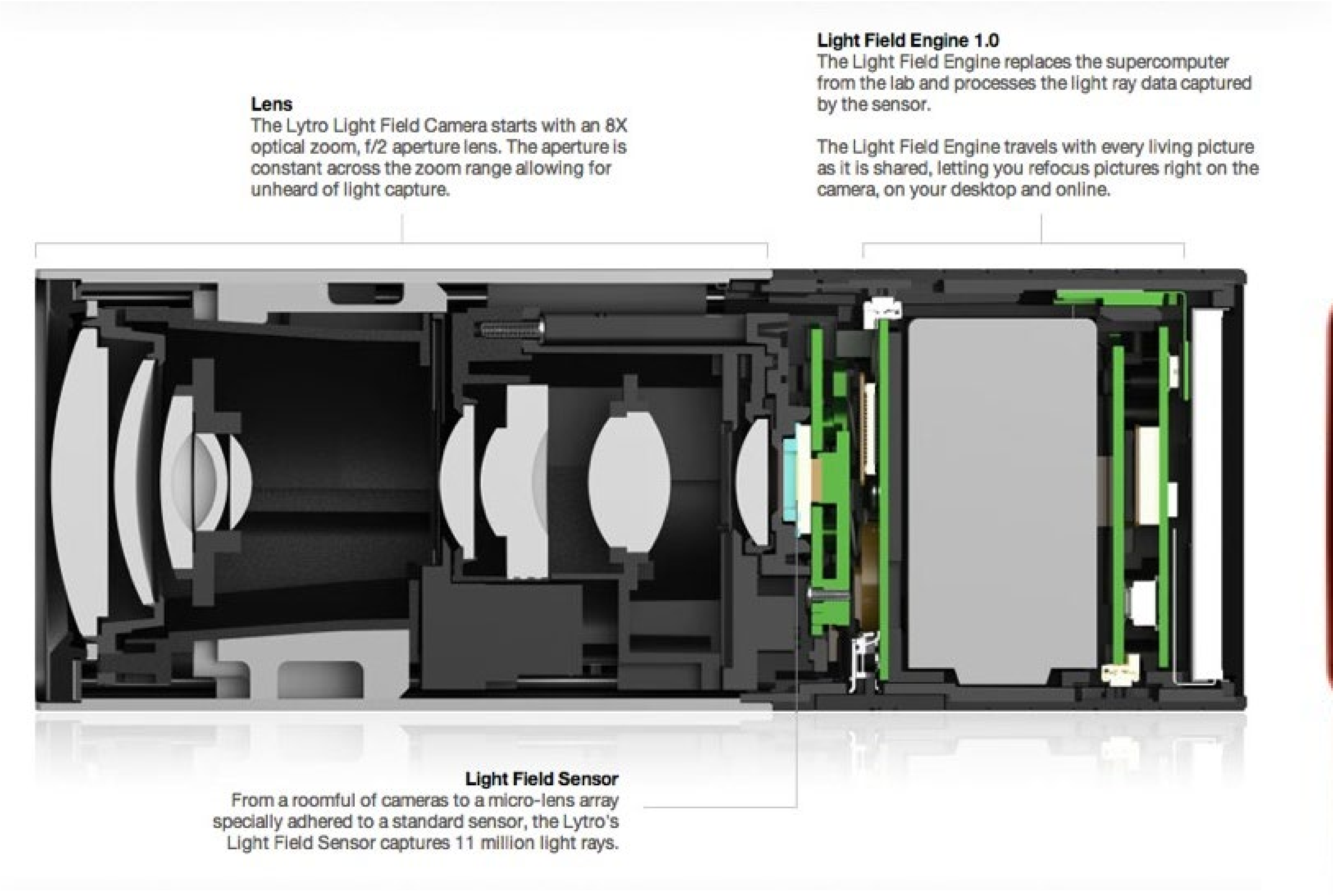
Kodak 16-megapixel sensor



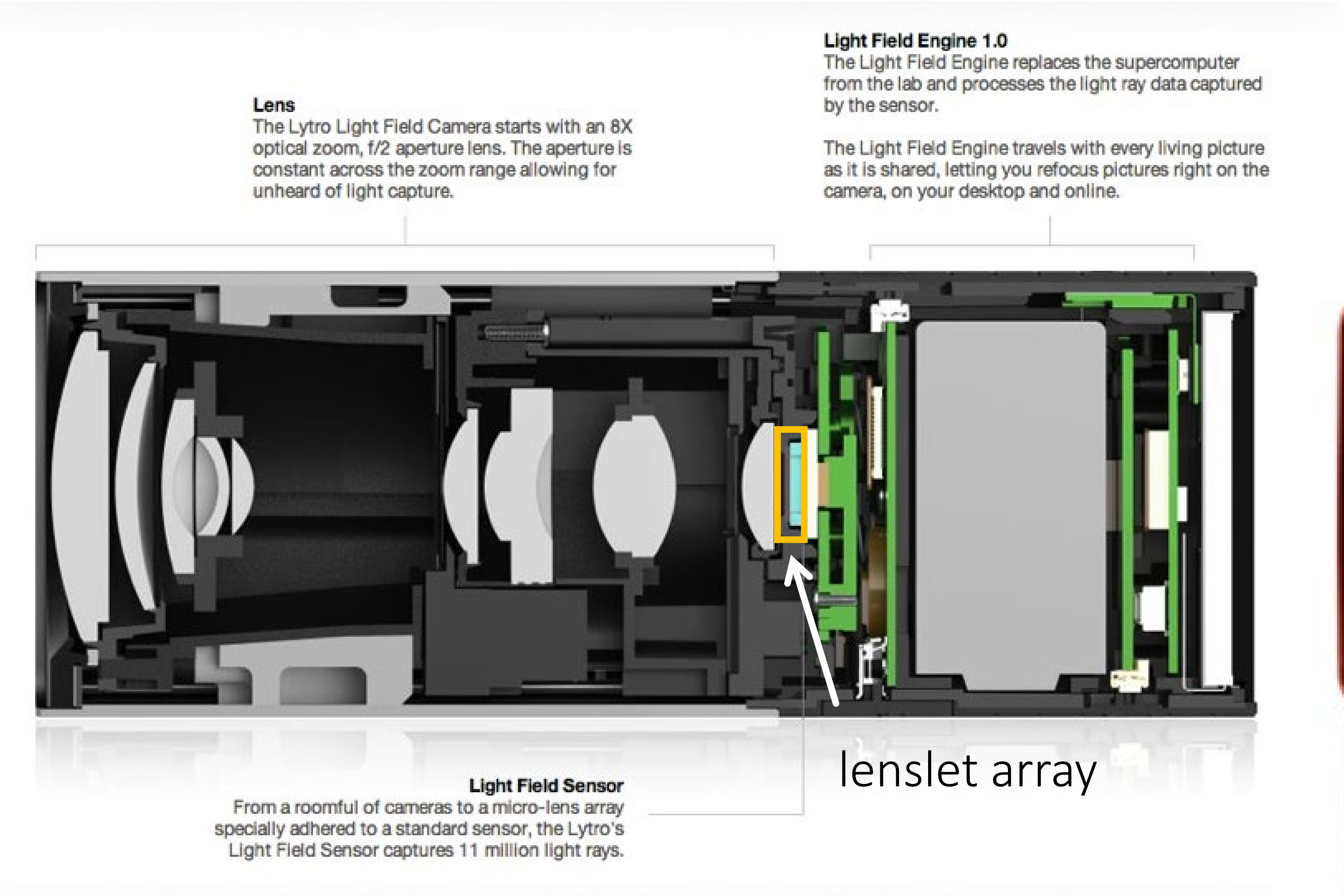
125μ square-sided microlenses

[Ng et al., Stanford Technical Report 2005]

Commercial plenoptic camera



Commercial plenoptic camera



Commercial plenoptic camera

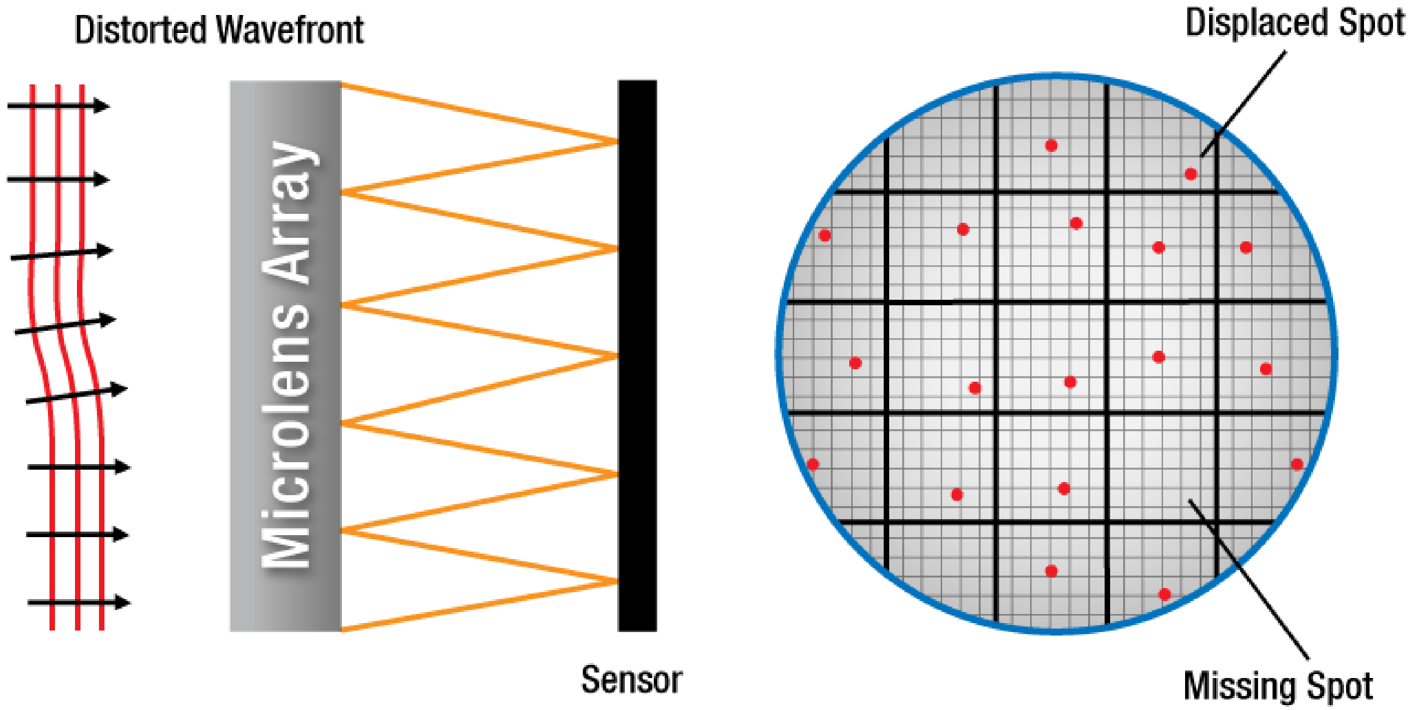
newer version with
higher resolution

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.
- <u>Very</u> expensive.

Shack-Hartmann wavefront sensors



- Completely different use: measuring how close a wavefront is to being planar
- Exactly the same optics as a plenoptic camera.
- Common instrument in optics labs and medical imaging.

Question

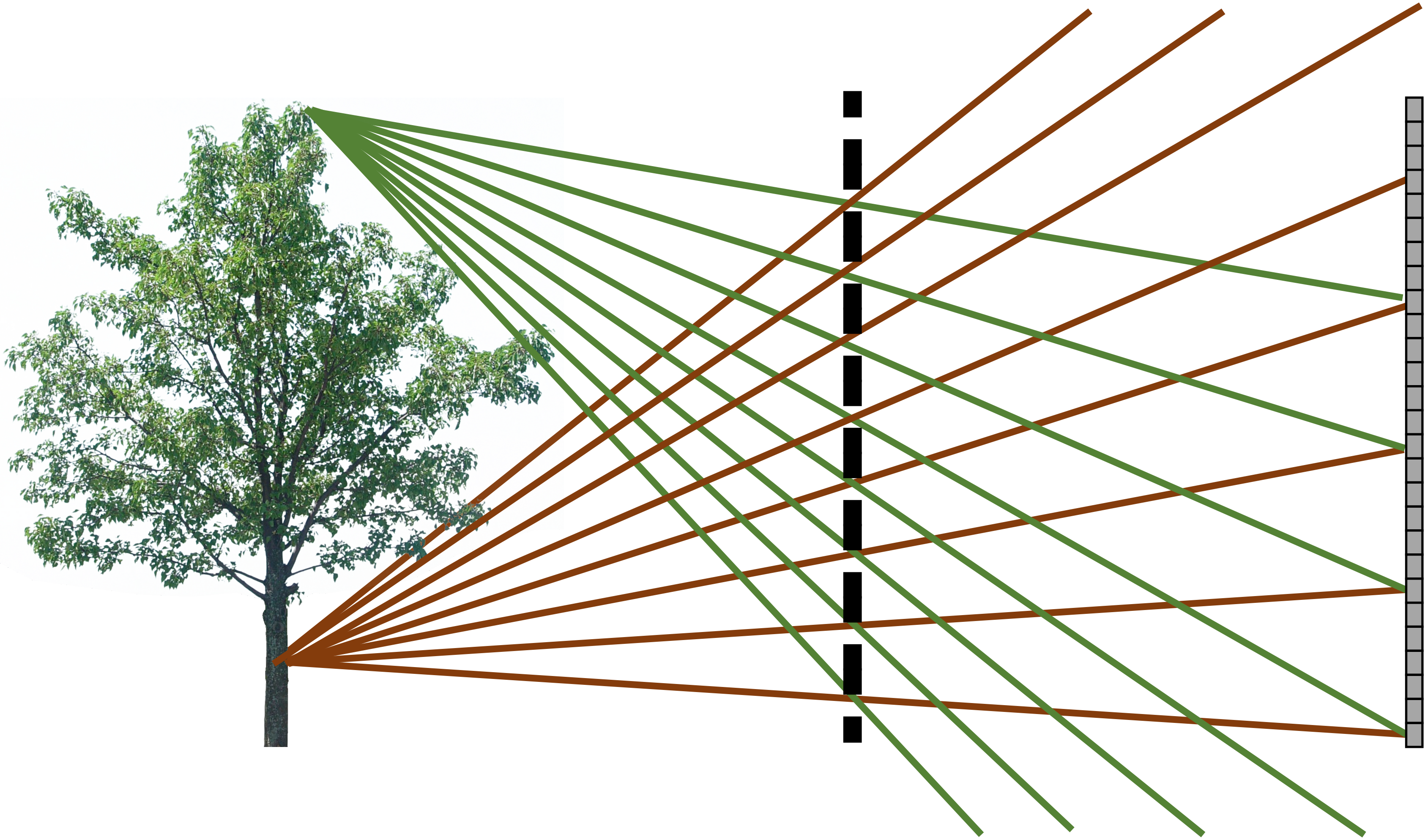
Given that the plenoptic camera design was known since 1908, why did it take a century for such a camera to be made commercially available?

Question

Given that the plenoptic camera design was known since 1908, why did it take a century for such a camera to be made commercially available?

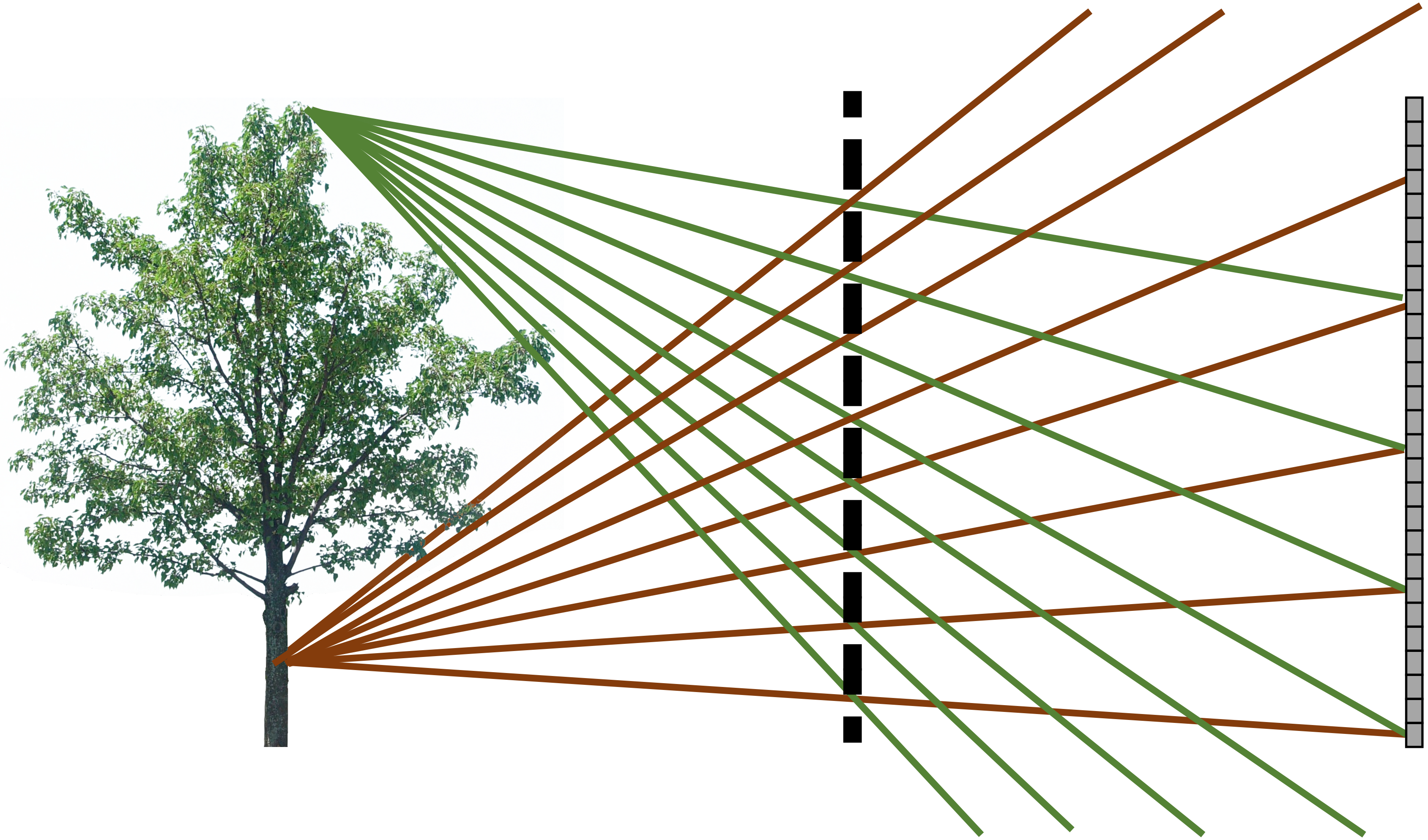
- Difficult to manufacture good lenslet arrays at high resolution (unless attached to sensors).
- Digital sensors did not have sufficiently high resolution.
- No known good use for them besides depth sensing.
  - The introduction of the handheld plenoptic camera was strongly influenced by the 1996 lightfield papers, which demonstrated all of the useful photographic operations one could replicate by having access to the entire lightfield.
- And of course, nobody thought of commercializing them.

Making a plenoptic camera



Is it possible to make a plenoptic camera using an array of pinholes?

Making a plenoptic camera



Need to be able to separate different rays falling on same pixel.

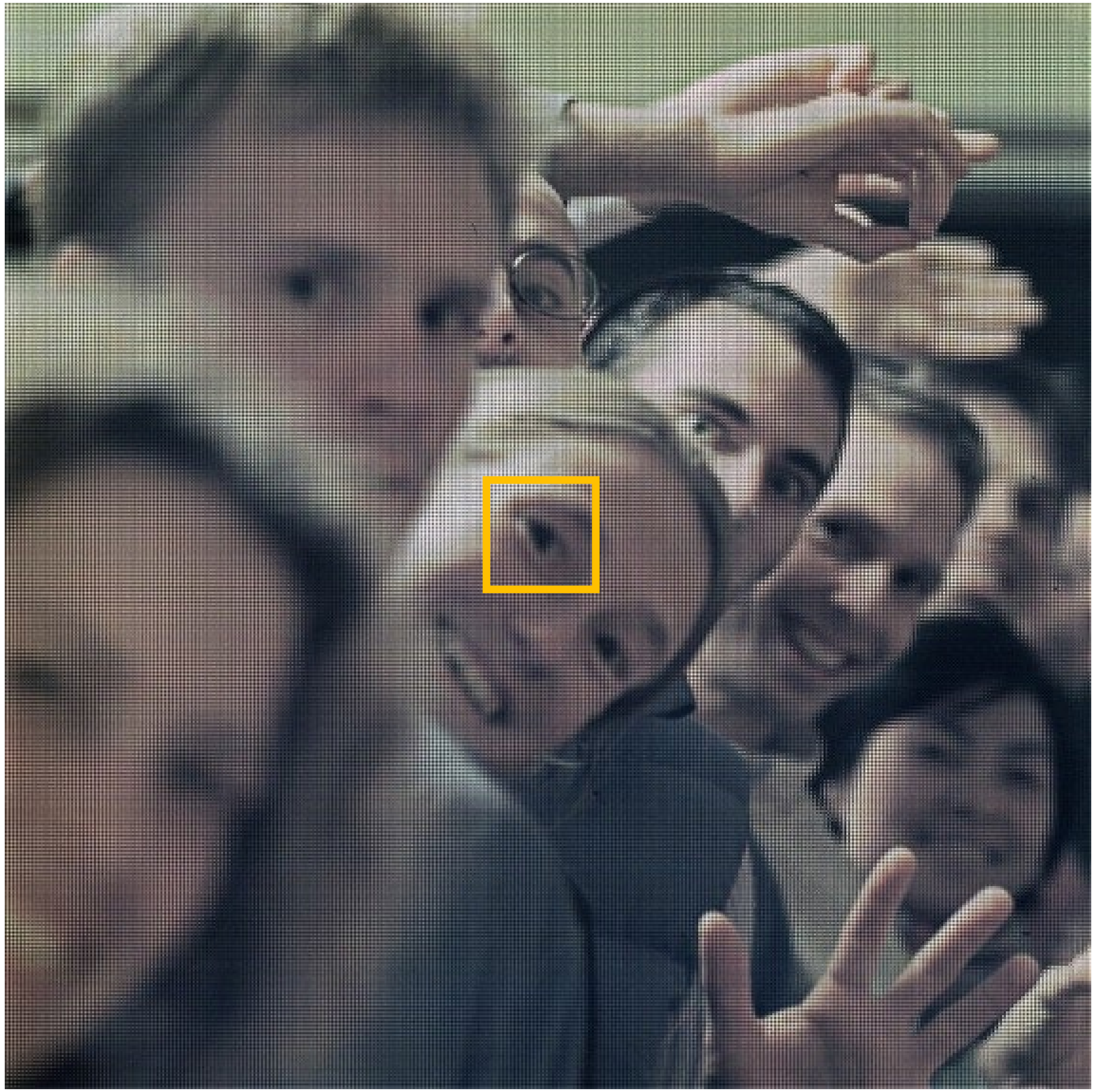
Is it possible to make a plenoptic camera using an array of pinholes?

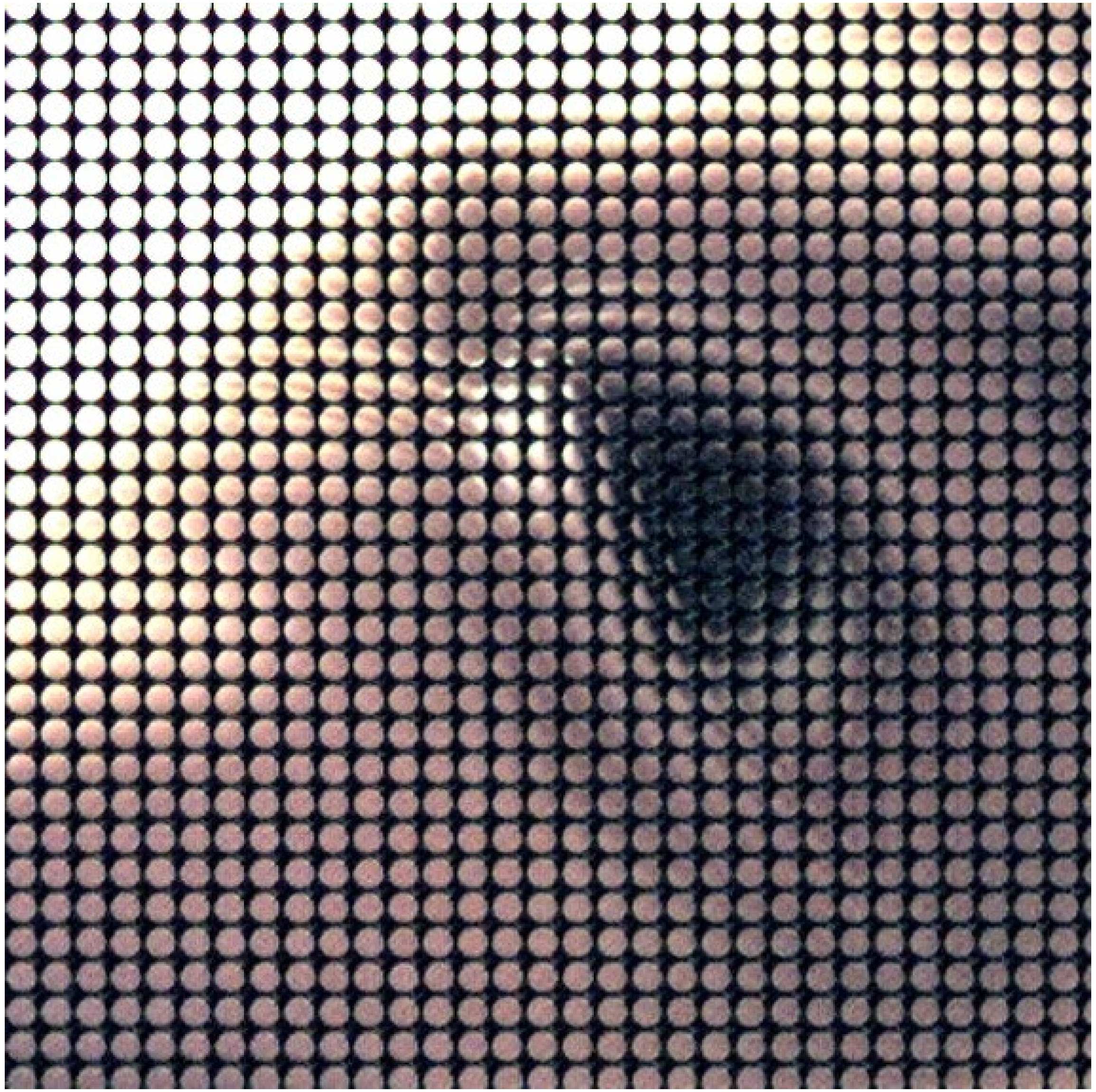
- Yes, under certain conditions on the images being captured.

[Georgiev, ECCV 2008]

Images from lightfields

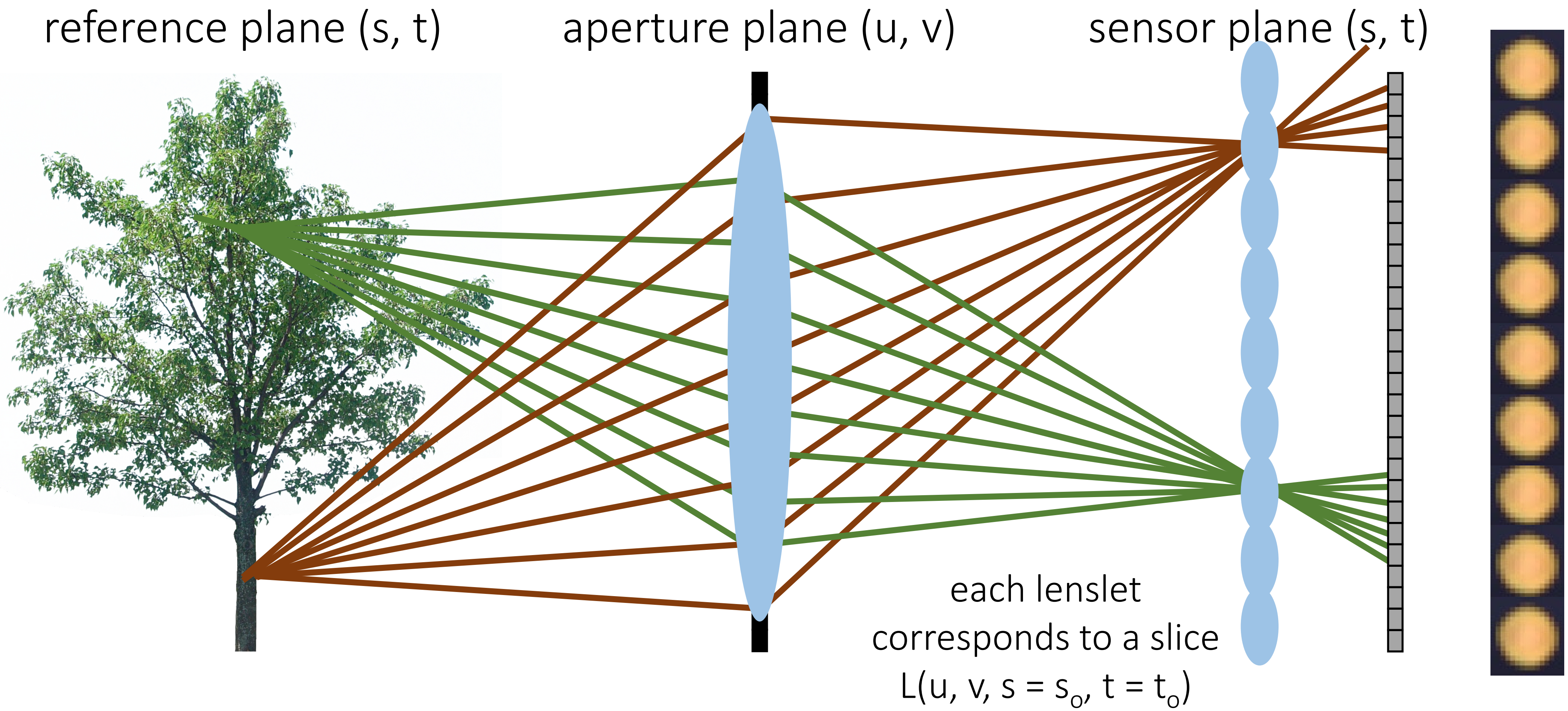
A plenoptic "image"





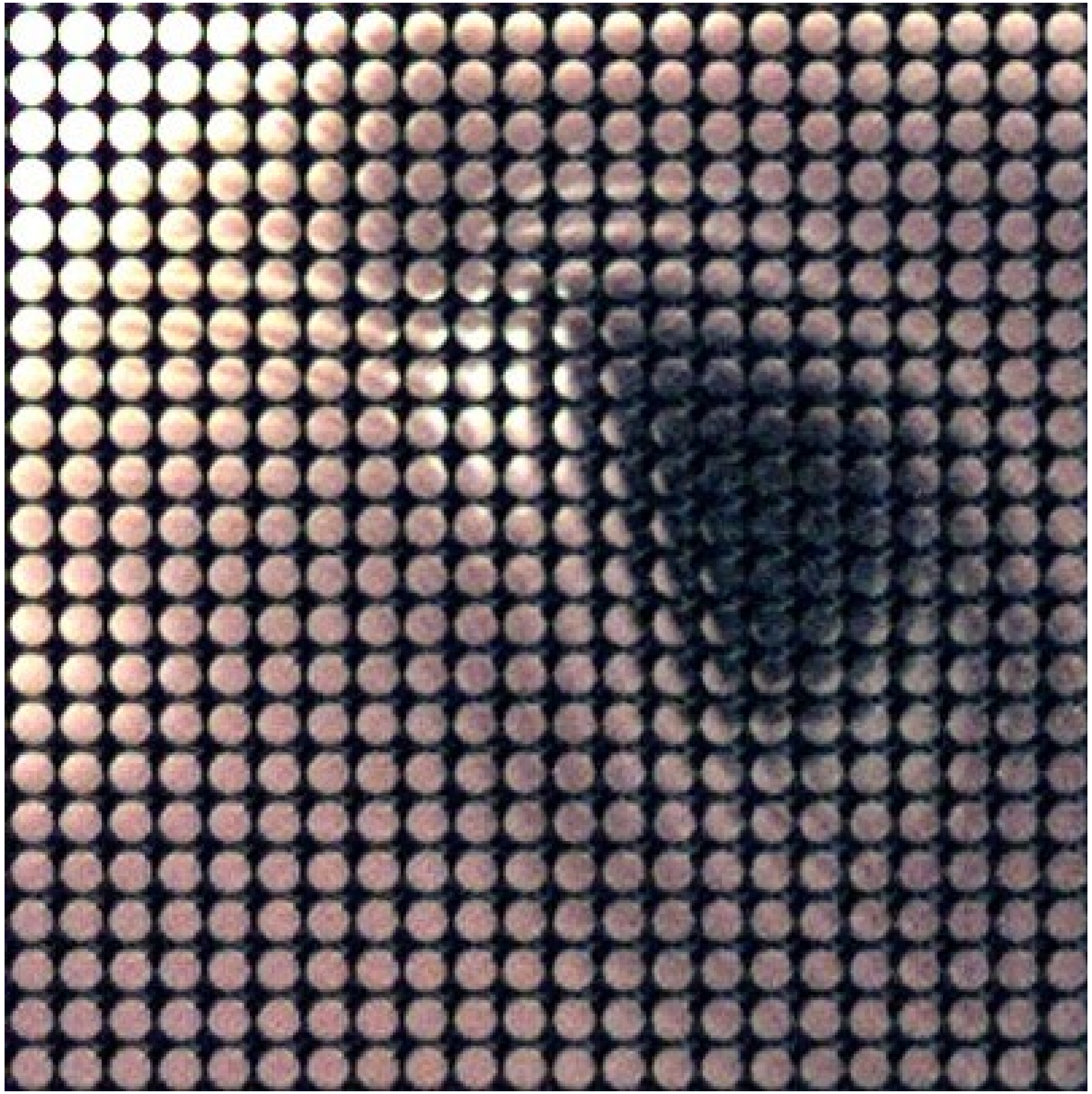
What are these circles?

A plenoptic camera



Lightfield L(u, v, s, t)

The plenoptic image



Which coordinates do I change when I move from one circle to another?

Which coordinates do I change when I move within each circle?

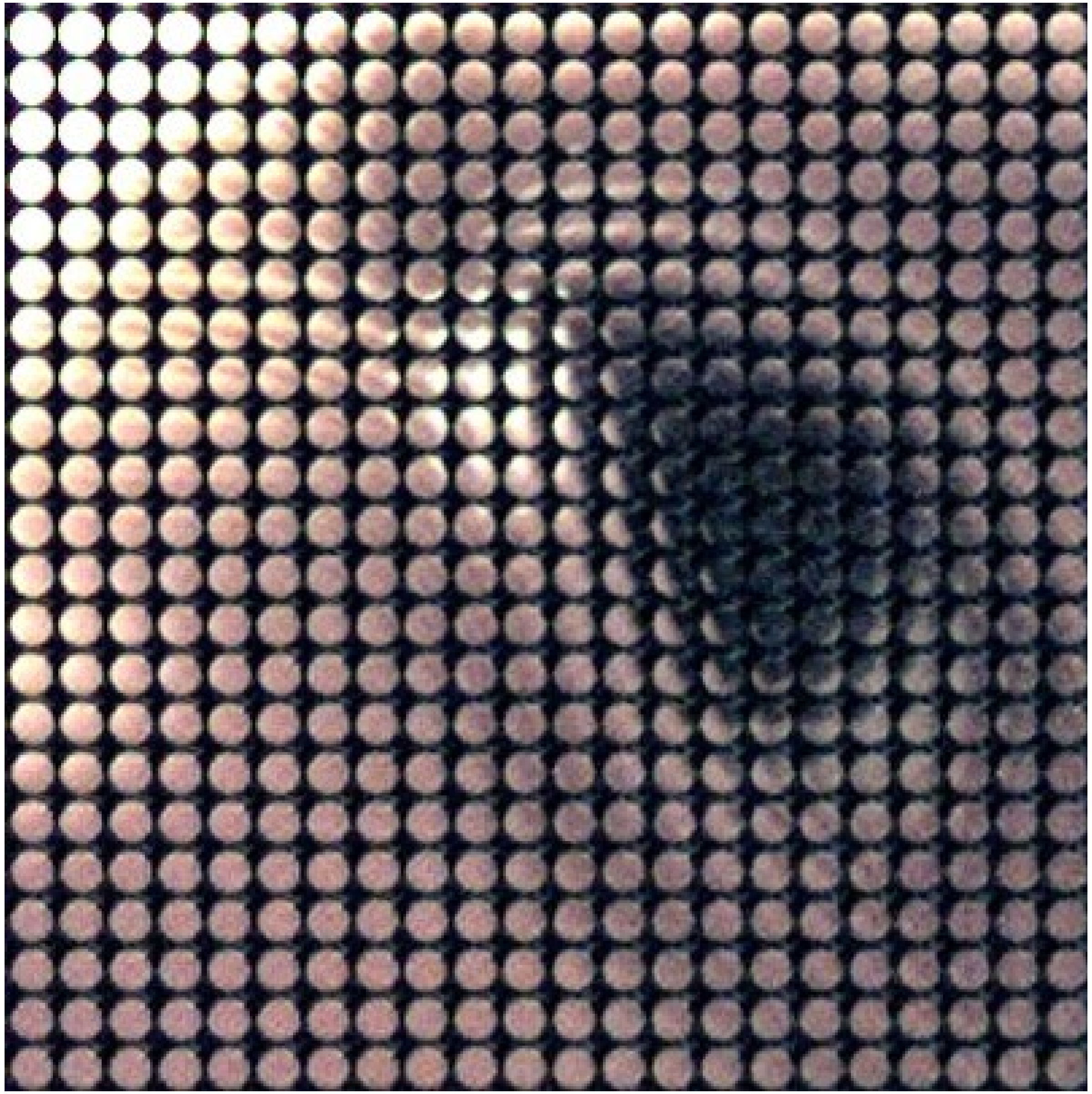
The plenoptic image

Which coordinates do I change when I move from one circle to another?

- I change s, t (sensor plane) coordinates.

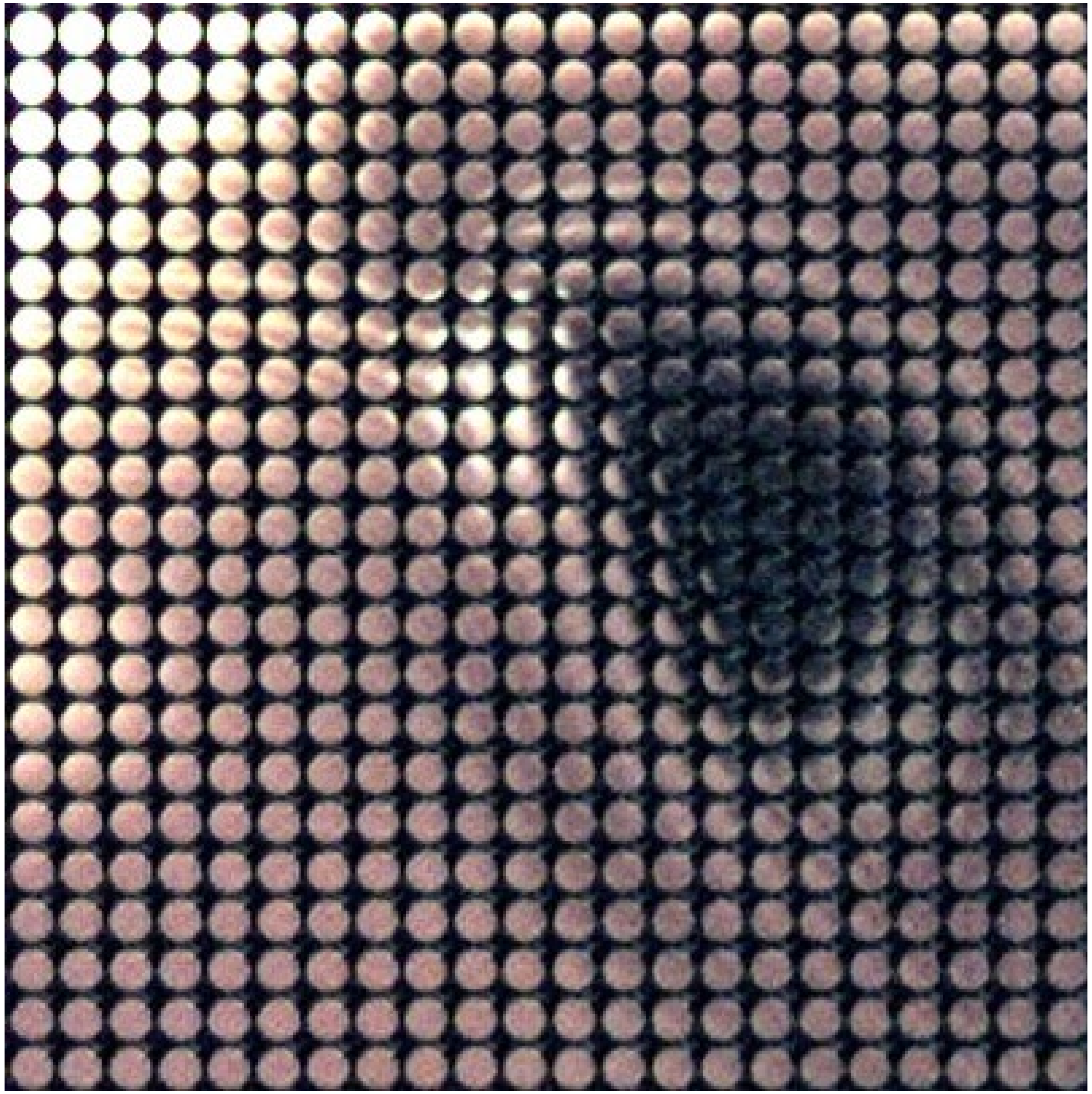
Which coordinates do I change when I move within each circle?

- I change u, v (aperture plane) coordinates.



How do I…

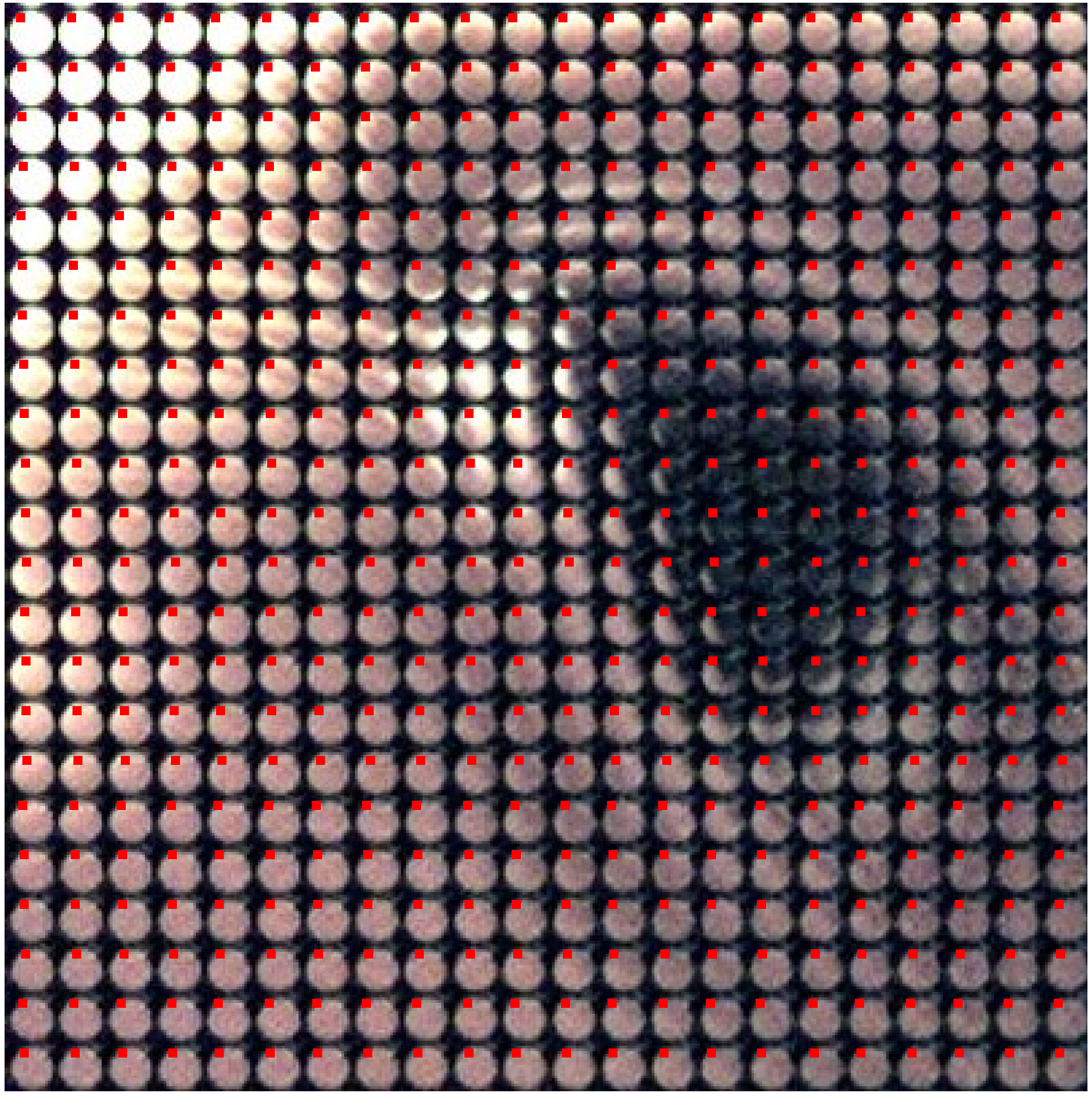
Simulate different viewpoints?



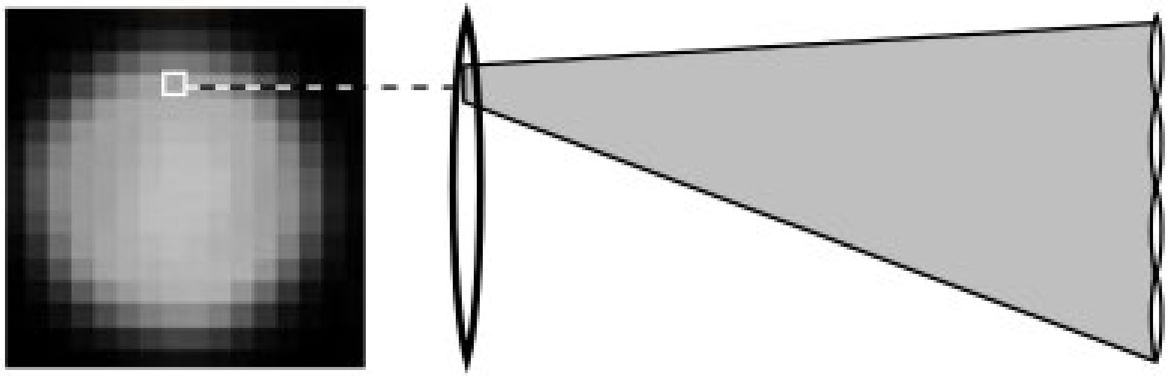
How do I…

Simulate different viewpoints?

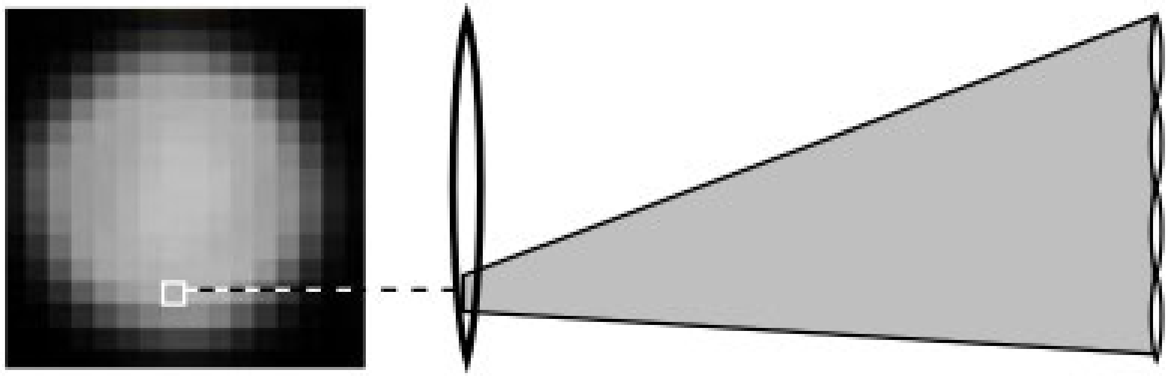
- Pick same pixel within each lenslet view



Changing the viewpoint



Viewpoint change is limited by the aperture of each of the lenslets.

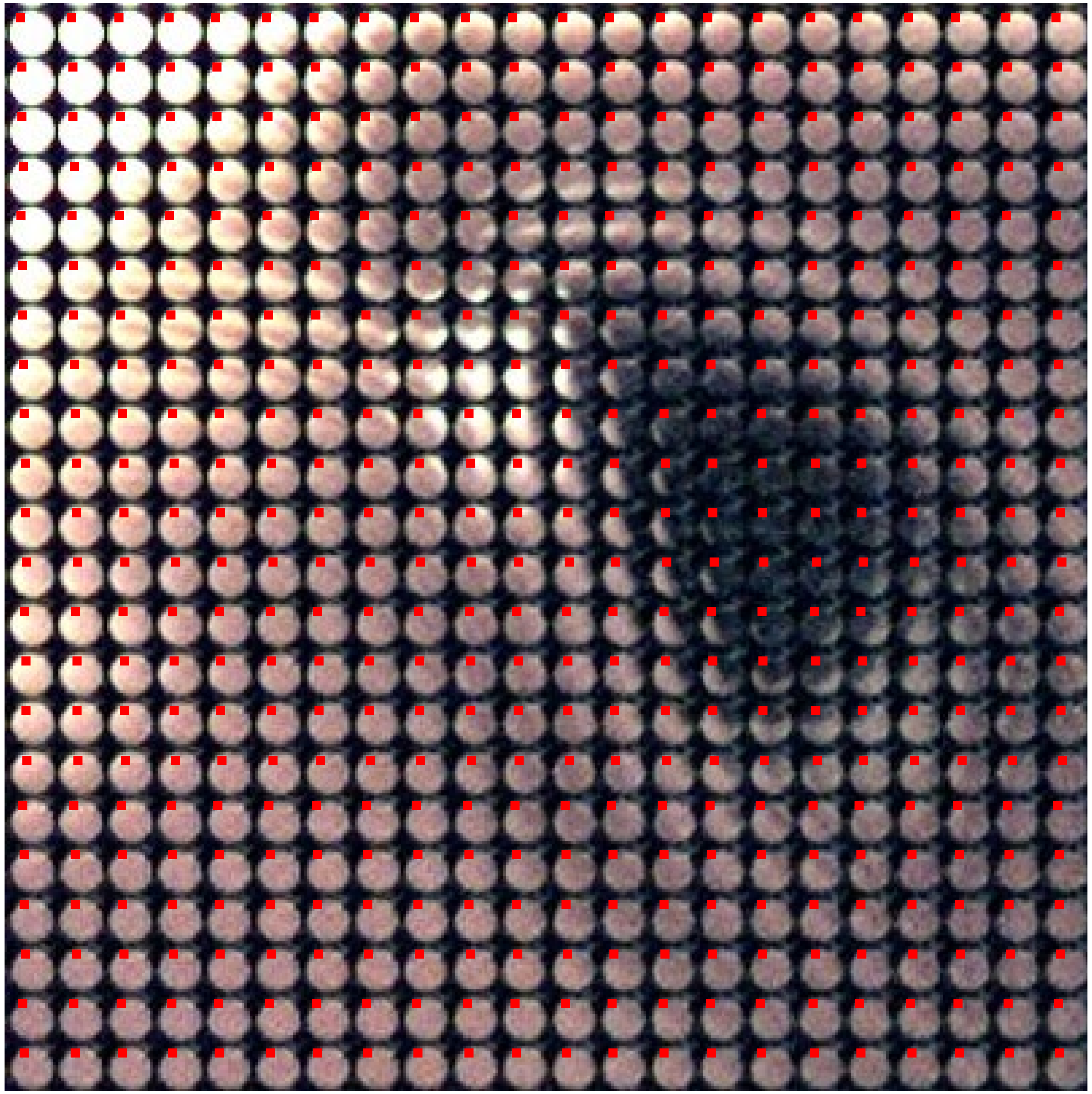


How do I…

Simulate different viewpoints?

- Pick same pixel within each lenslet view

Simulate different aperture sizes?



How do I…

Simulate different viewpoints?
- Pick same pixel within each lenslet view

Simulate different aperture sizes?
- Sum more than one pixels within each lenslet view

How do I…

Simulate different viewpoints?

- Pick same pixel within each lenslet view

Simulate different aperture sizes?

- Sum more than one pixels within each lenslet view

Simulate lens at current focus setting?

How do I…

Simulate different viewpoints?

- Pick same pixel within each lenslet view

Simulate different aperture sizes?

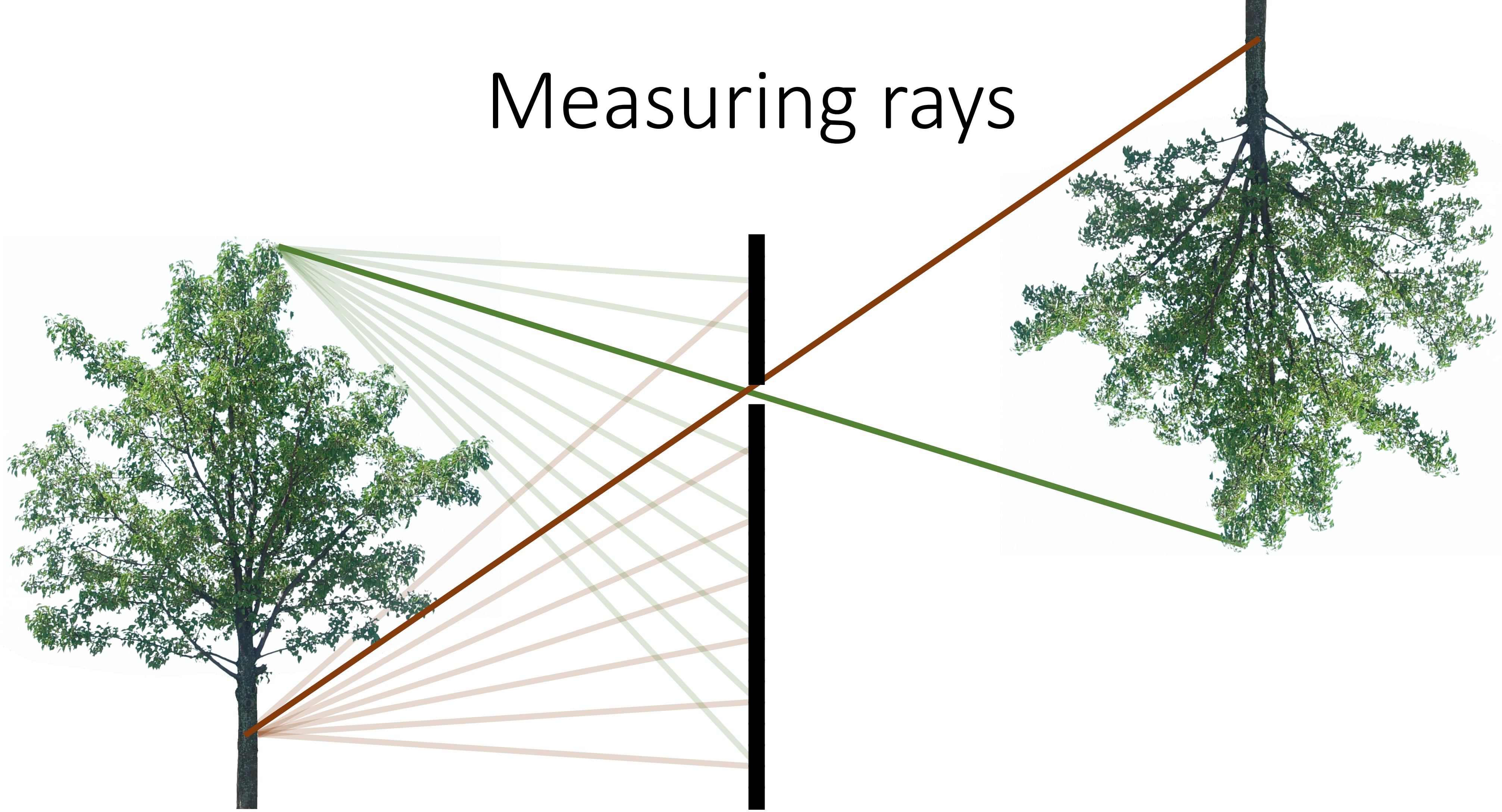
- Sum more than one pixels within each lenslet view

Simulate lens at current focus setting?

- Same as above. Sum all pixels for max aperture setting.

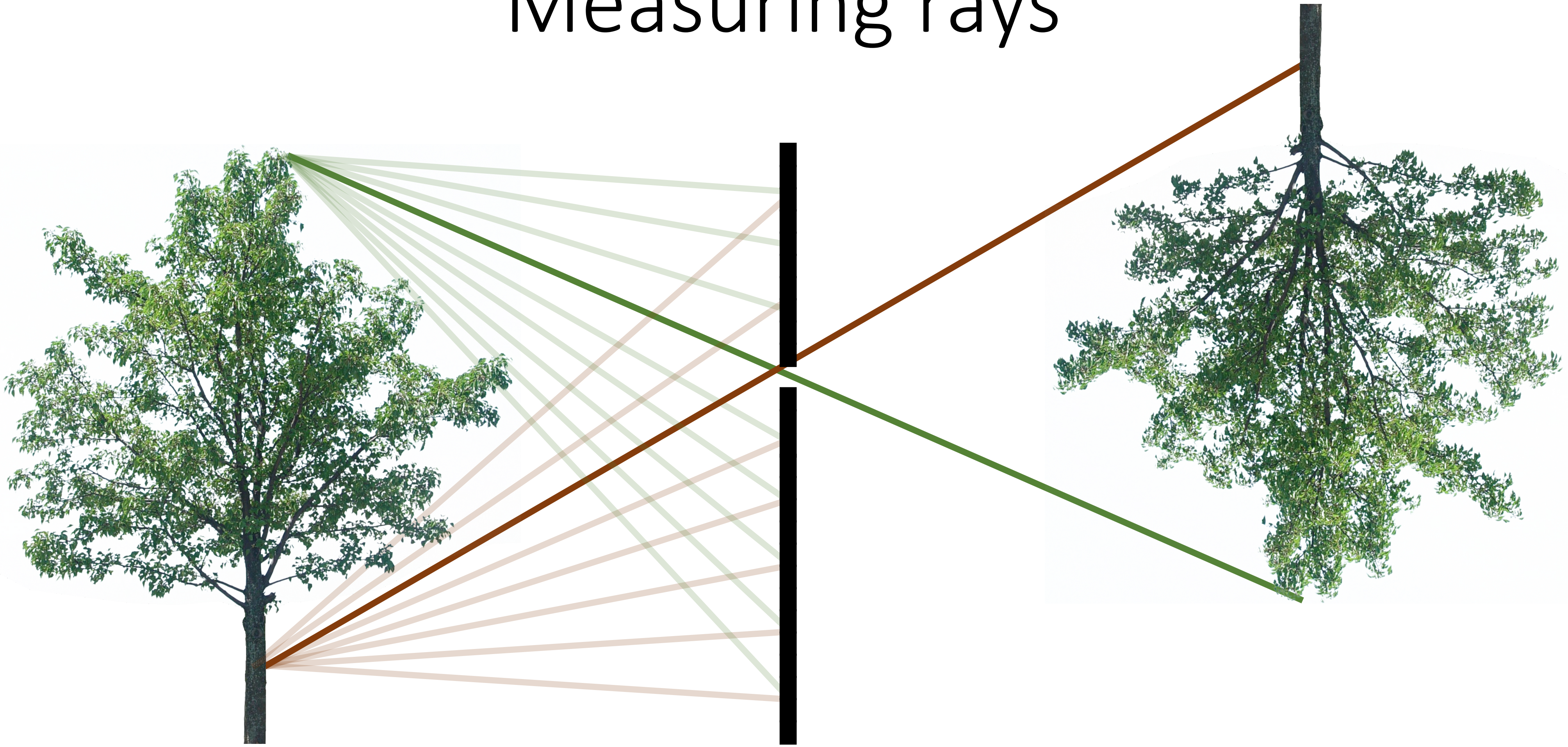
How do I change focus setting?

Measuring rays



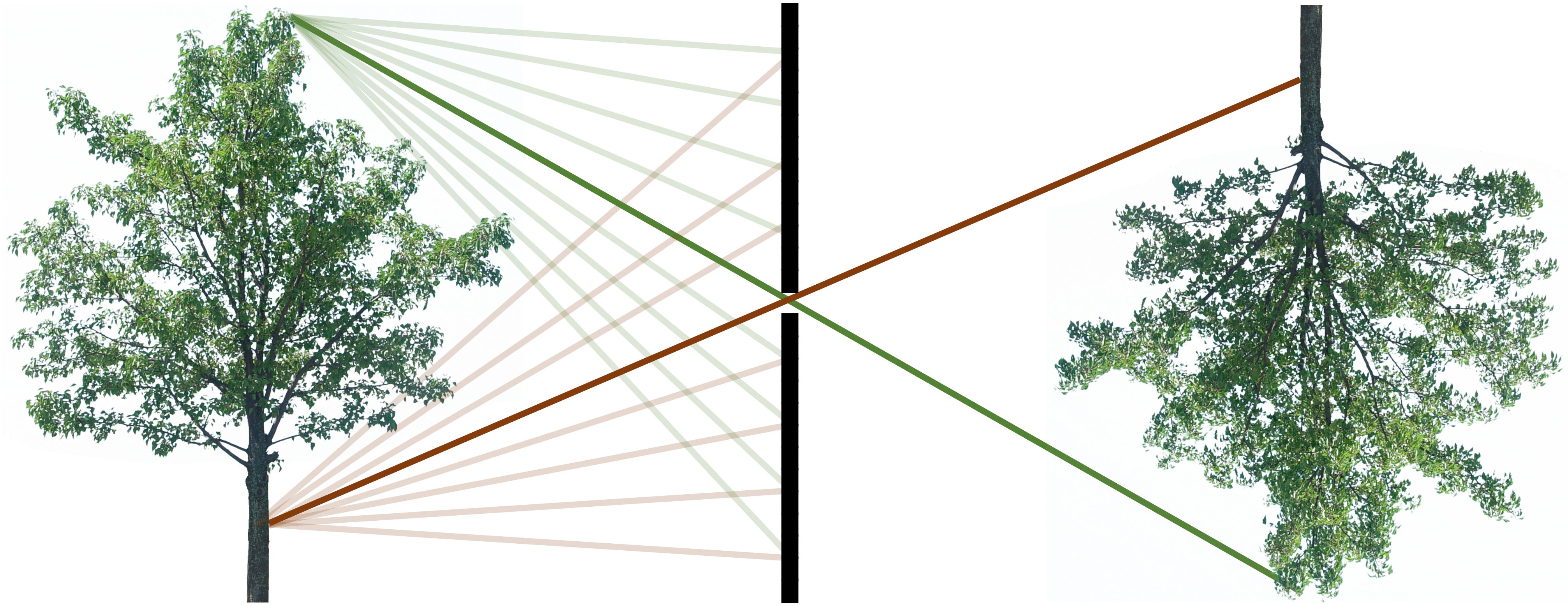
We can capture the same set of rays by using a pinhole camera from multiple viewpoints

Measuring rays



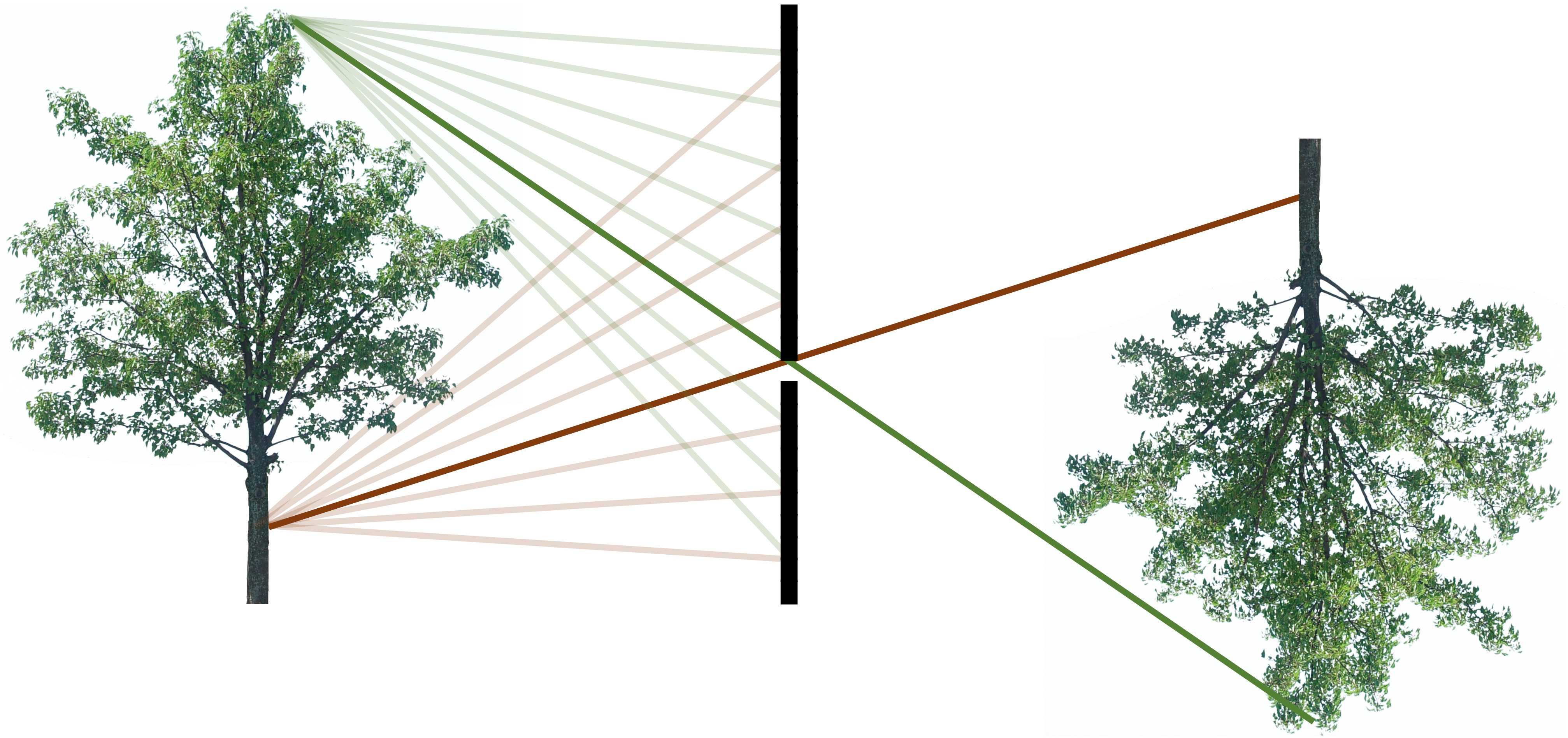
We can capture the same set of rays by using a pinhole camera from multiple viewpoints

Measuring rays



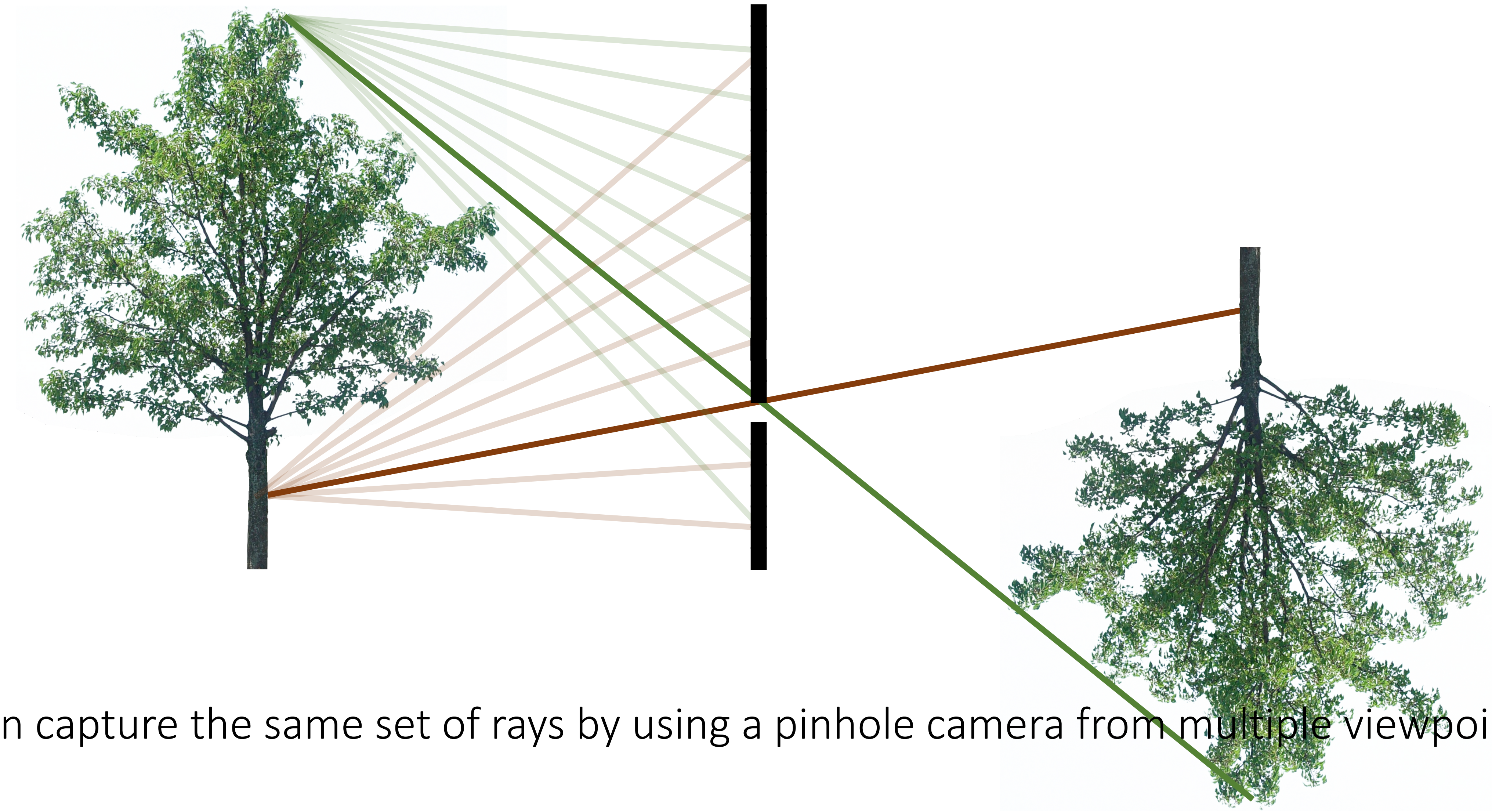
We can capture the same set of rays by using a pinhole camera from multiple viewpoints

Measuring rays



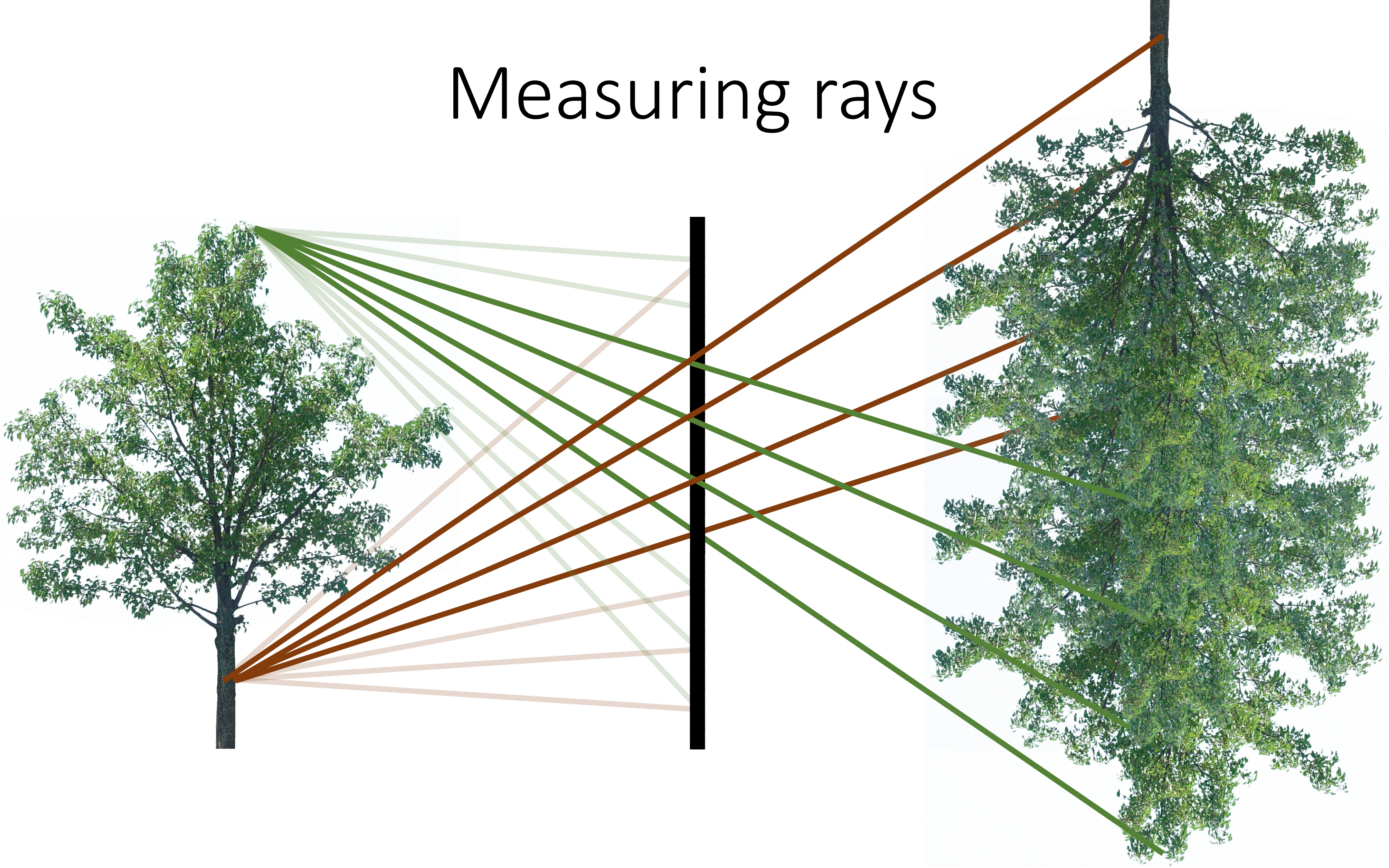
We can capture the same set of rays by using a pinhole camera from multiple viewpoints

Measuring rays



We can capture the same set of rays by using a pinhole camera from multiple viewpoints

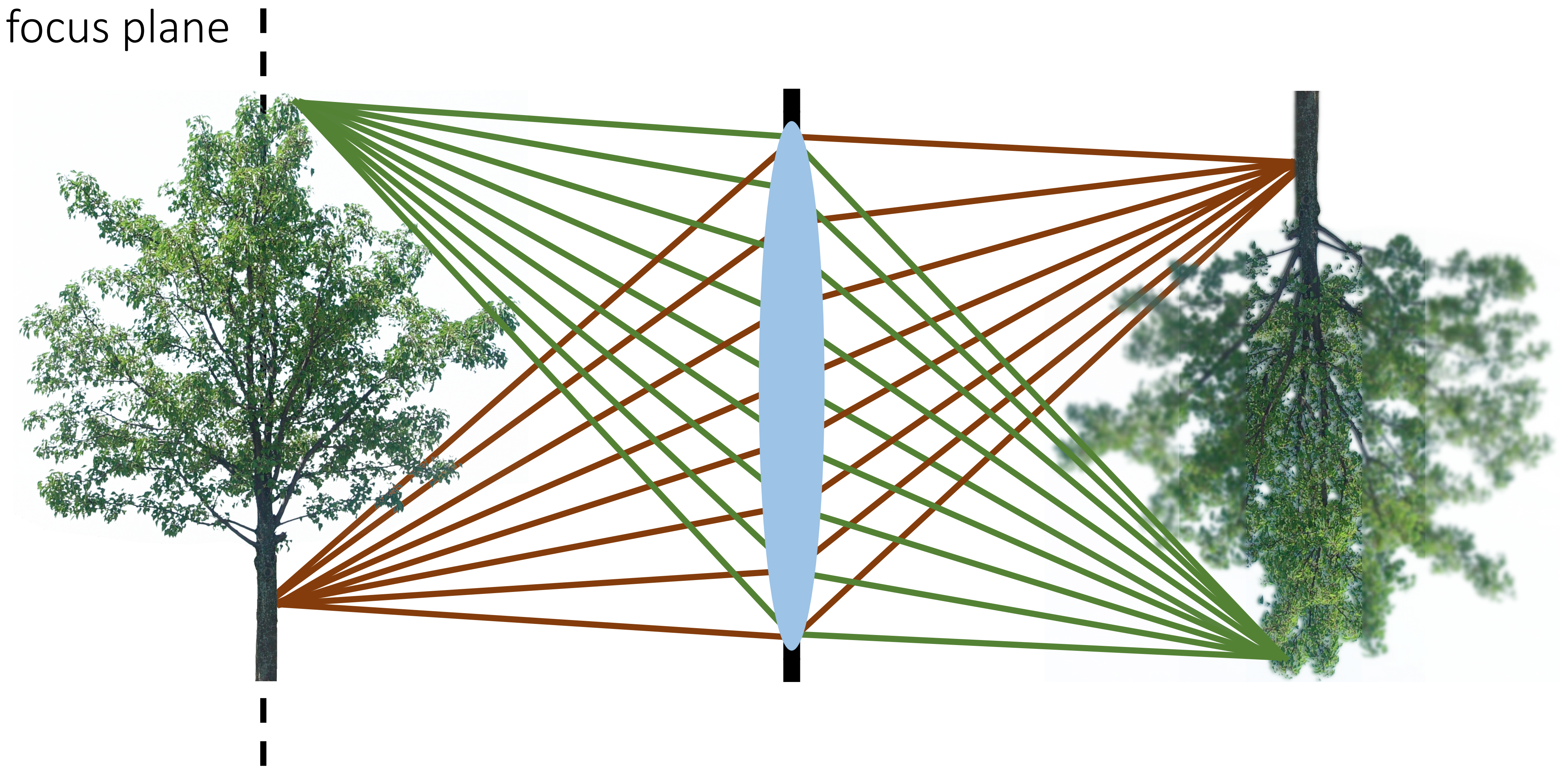
Measuring rays



We can capture the same set of rays by using a pinhole camera from multiple viewpoints

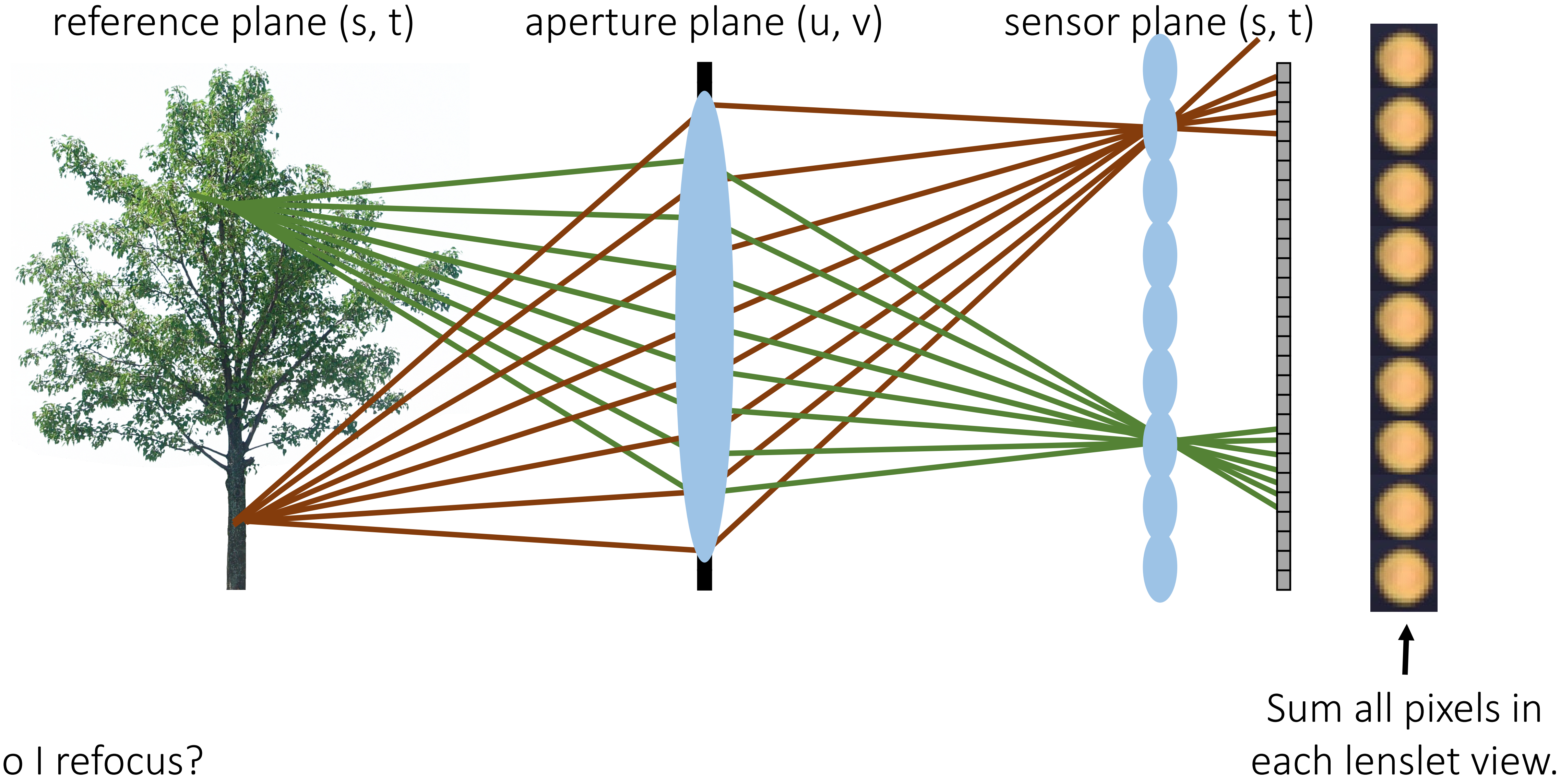
- How would you merge these images into a lens, defocused image?

Measuring rays



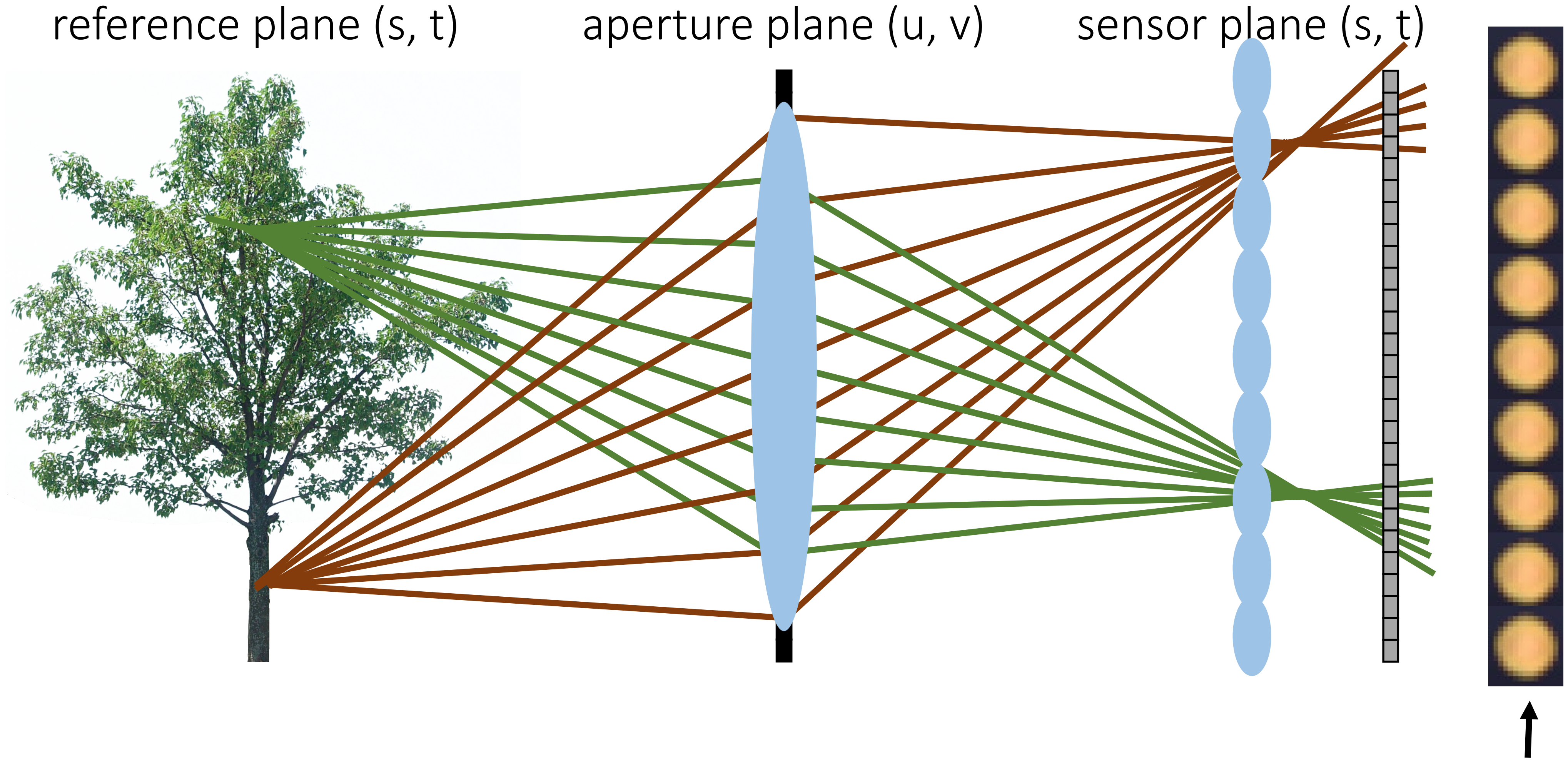
Sum all pixels in each lenslet view.

Form lens image



How do I refocus?

Form lens image



Sum all pixels in each lenslet view.

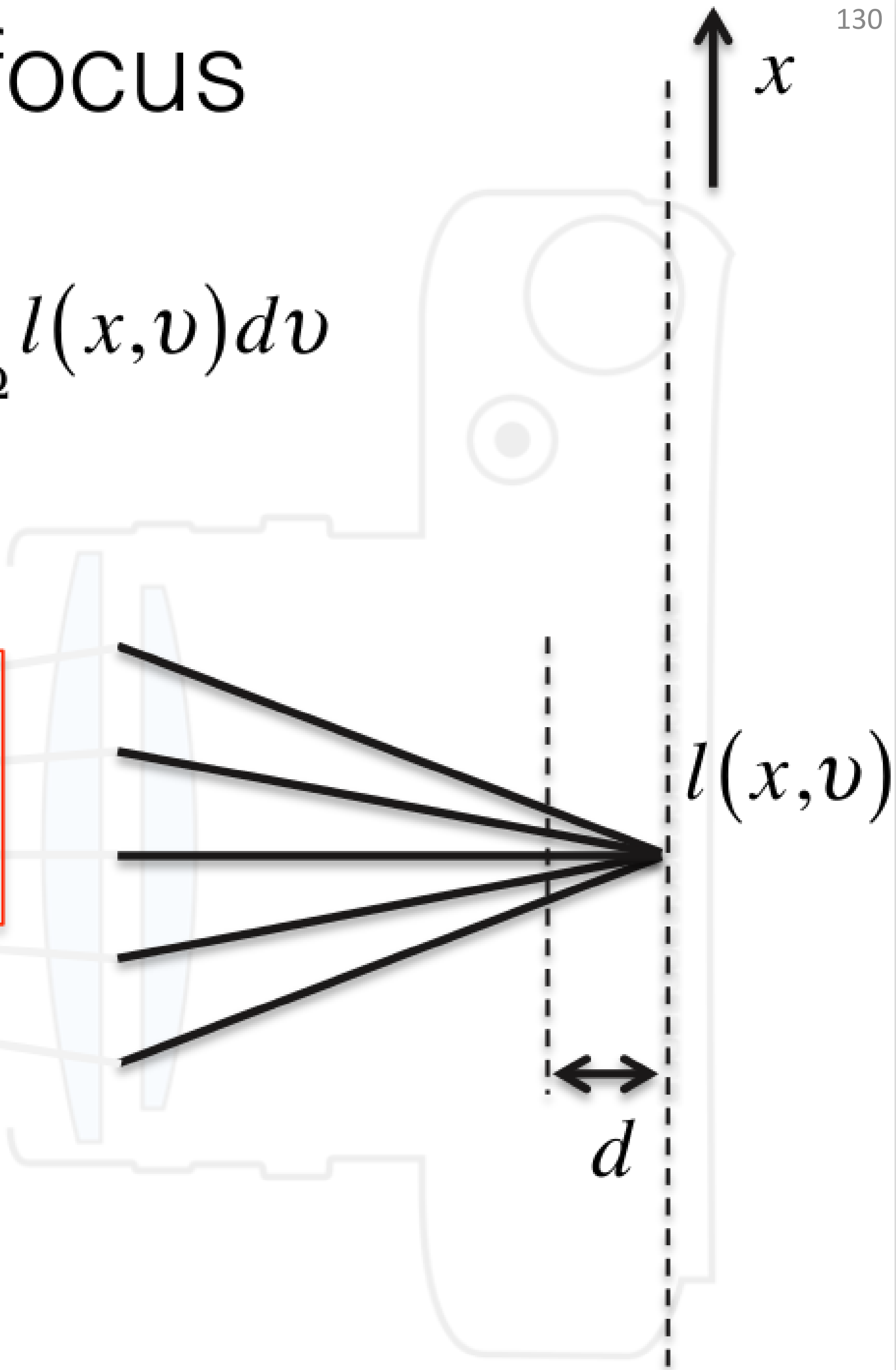
How do I refocus?

- Need to move sensor plane to a different location.

Understanding Refocus

- consider light field inside camera
- synthesize image on sensor $i_{d=0}(x) = \int_{\Omega} l(x,\upsilon)\,d\upsilon$

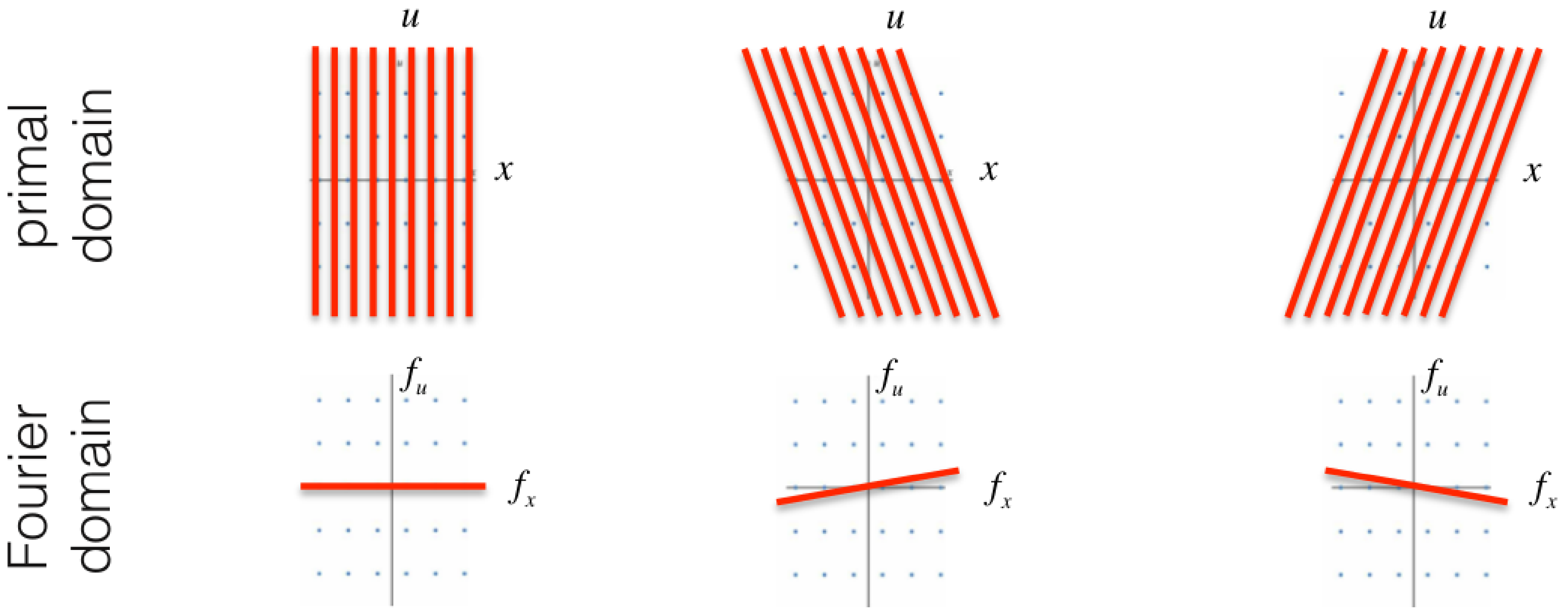
$$\boxed{i_d(x) = \int_{\Omega} l(x + d\upsilon, \upsilon)\,d\upsilon}$$



Understanding Refocus – Fourier Slicing

- Fourier slice theorem: projection in primal is slicing in Fourier space

$$i_d(x) = \int_\Omega l(x + dv, v)\,dv \longleftrightarrow \hat{i}_d(f_x) = \hat{l}(f_x, f_v - df_x)$$



[Ng, SIGGRAPH 2006]

Refocusing example

Refocusing example

Refocusing example

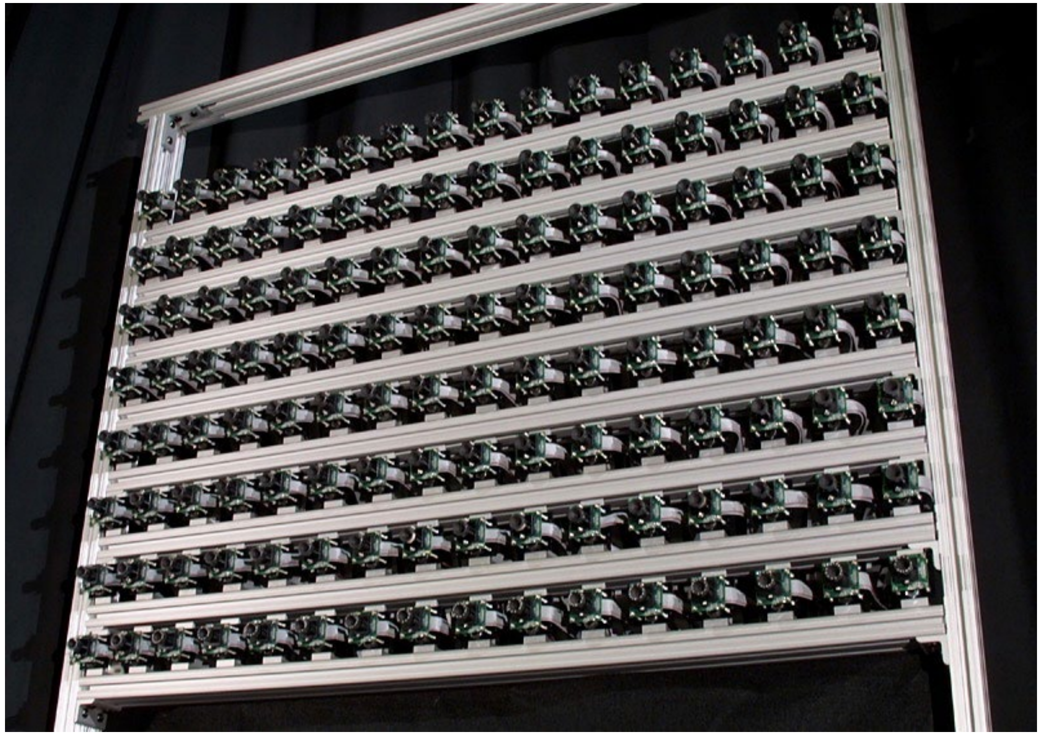
Synthesizing images from a lightfield

Many more examples with demo: http://lightfields.stanford.edu/

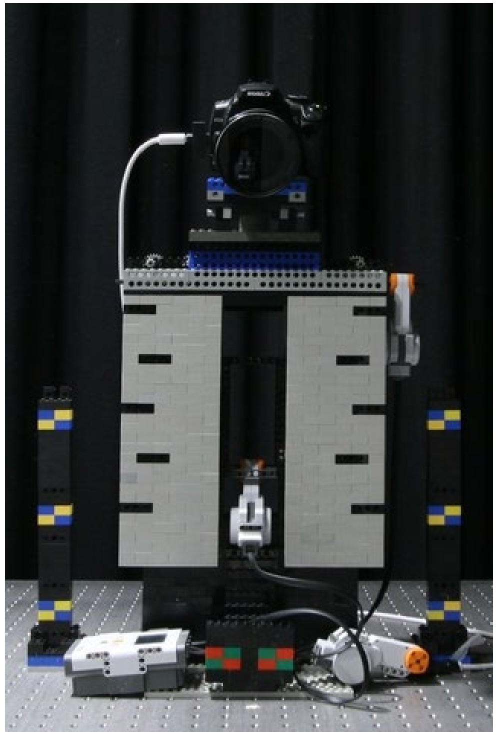
Three ways to measure a lightfield

1) Use a plenoptic camera

2) Use a camera array



3) Use one camera multiple times



What are the pros and cons of each?

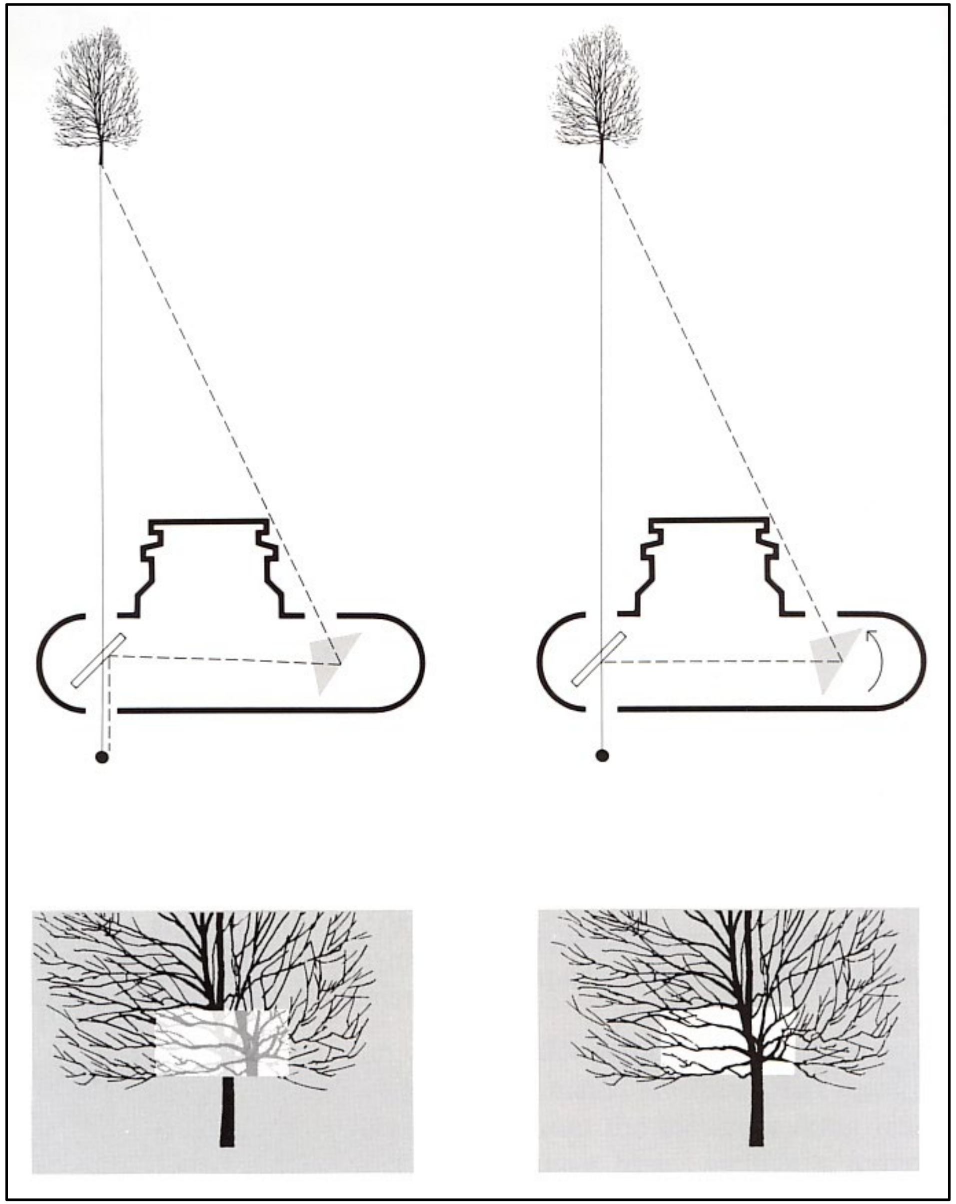
Some notes on (auto-)focusing

Different cameras have different focusing processes

Manual focus in rangefinder cameras

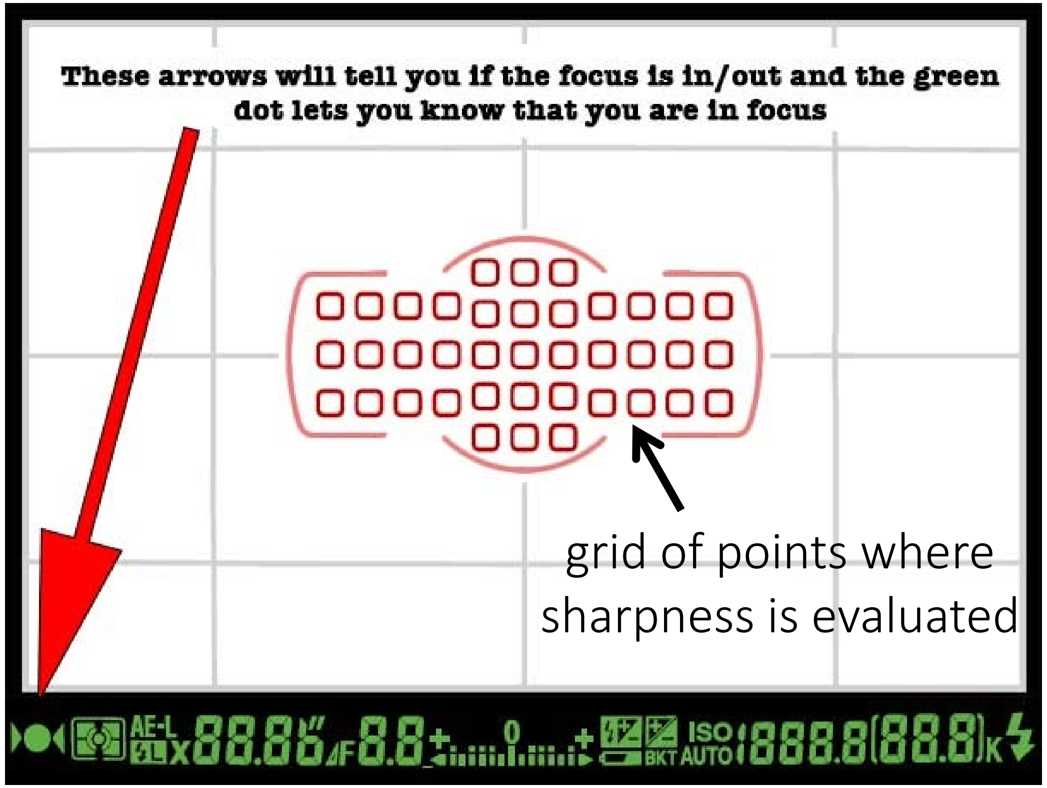
- Focusing based on triangulation: when the image is in focus, you will see the two copies aligned.
- Very accurate but very painstaking.
- Different perspective than that of the main lens.

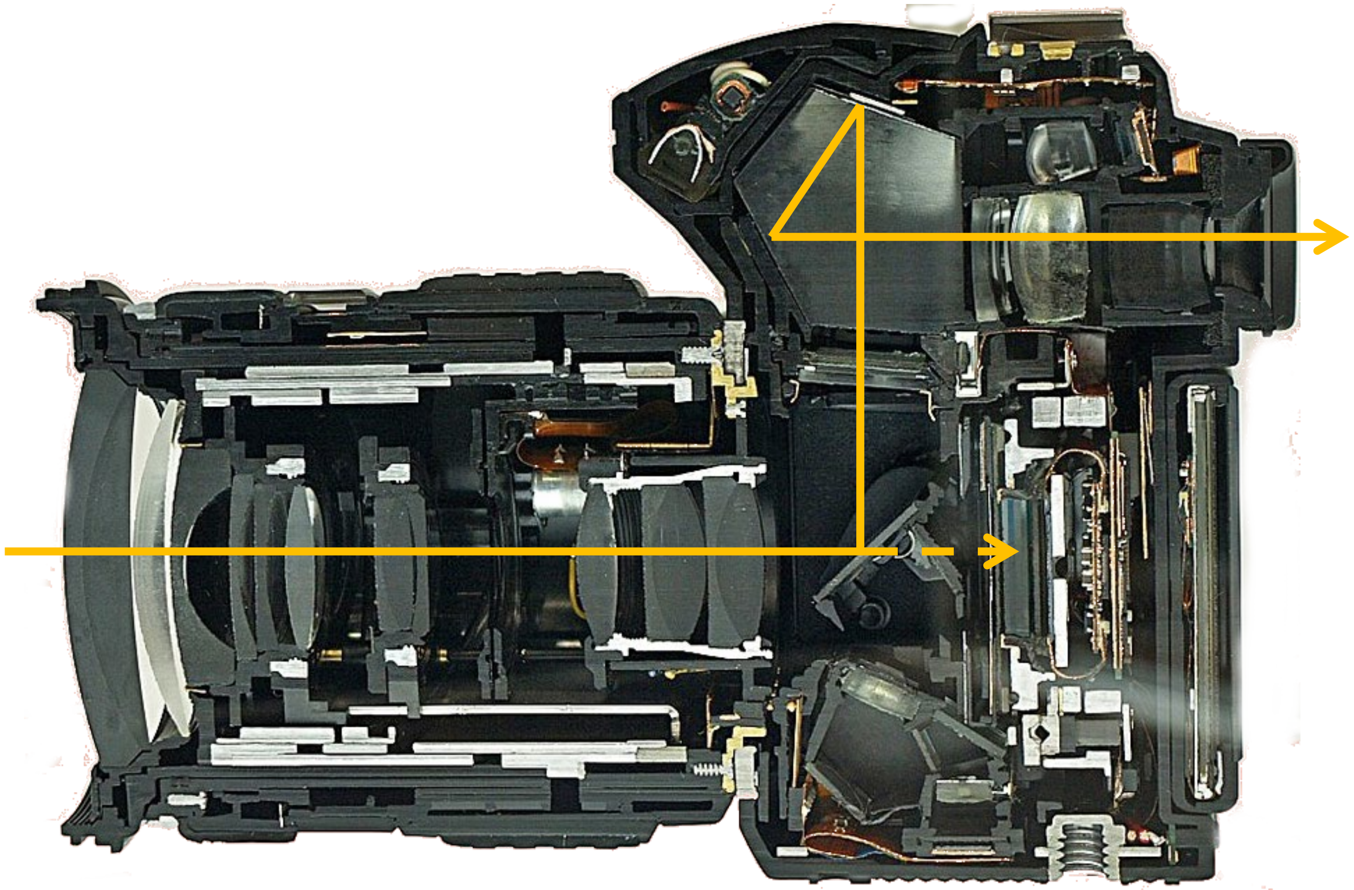
standard in Leica cameras



Manual focus in (D)SLR cameras

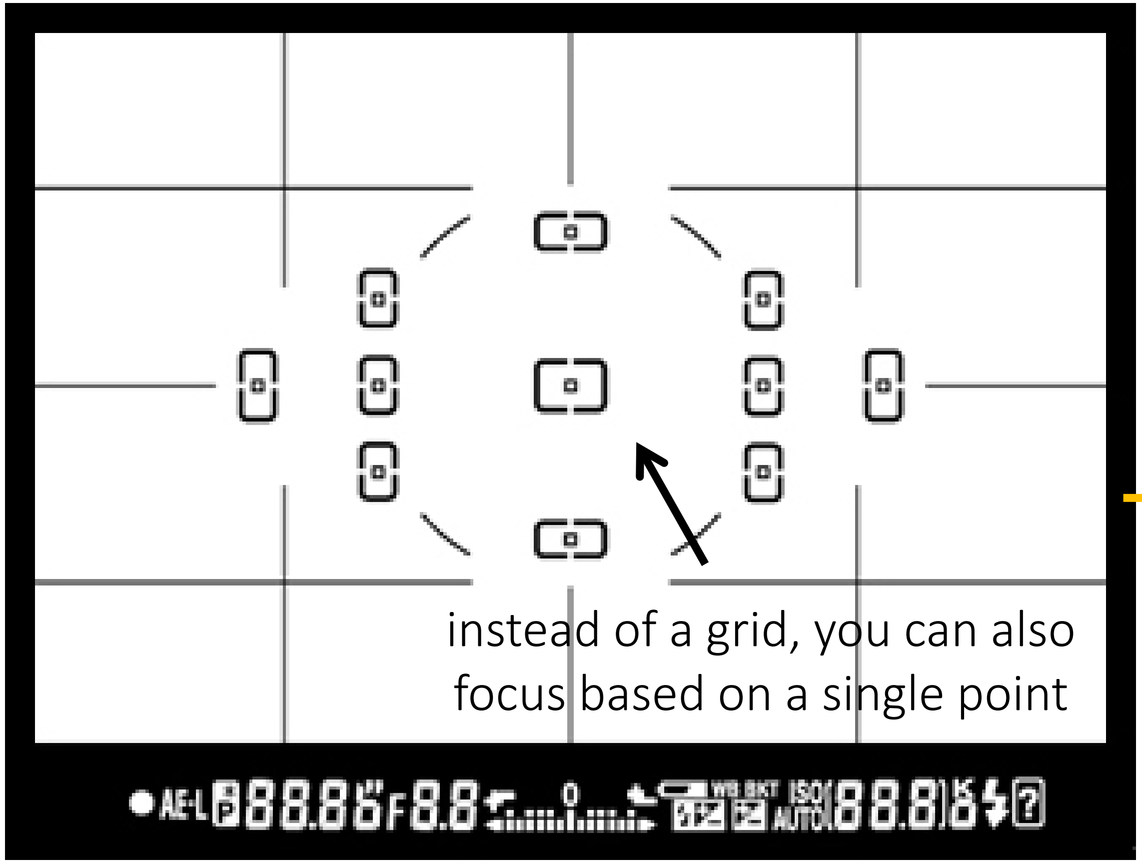
- Same view as main lens.
- Just rotate the focusing ring until you are satisfied by the sharpness.
- Viewfinder indicators can help this process.

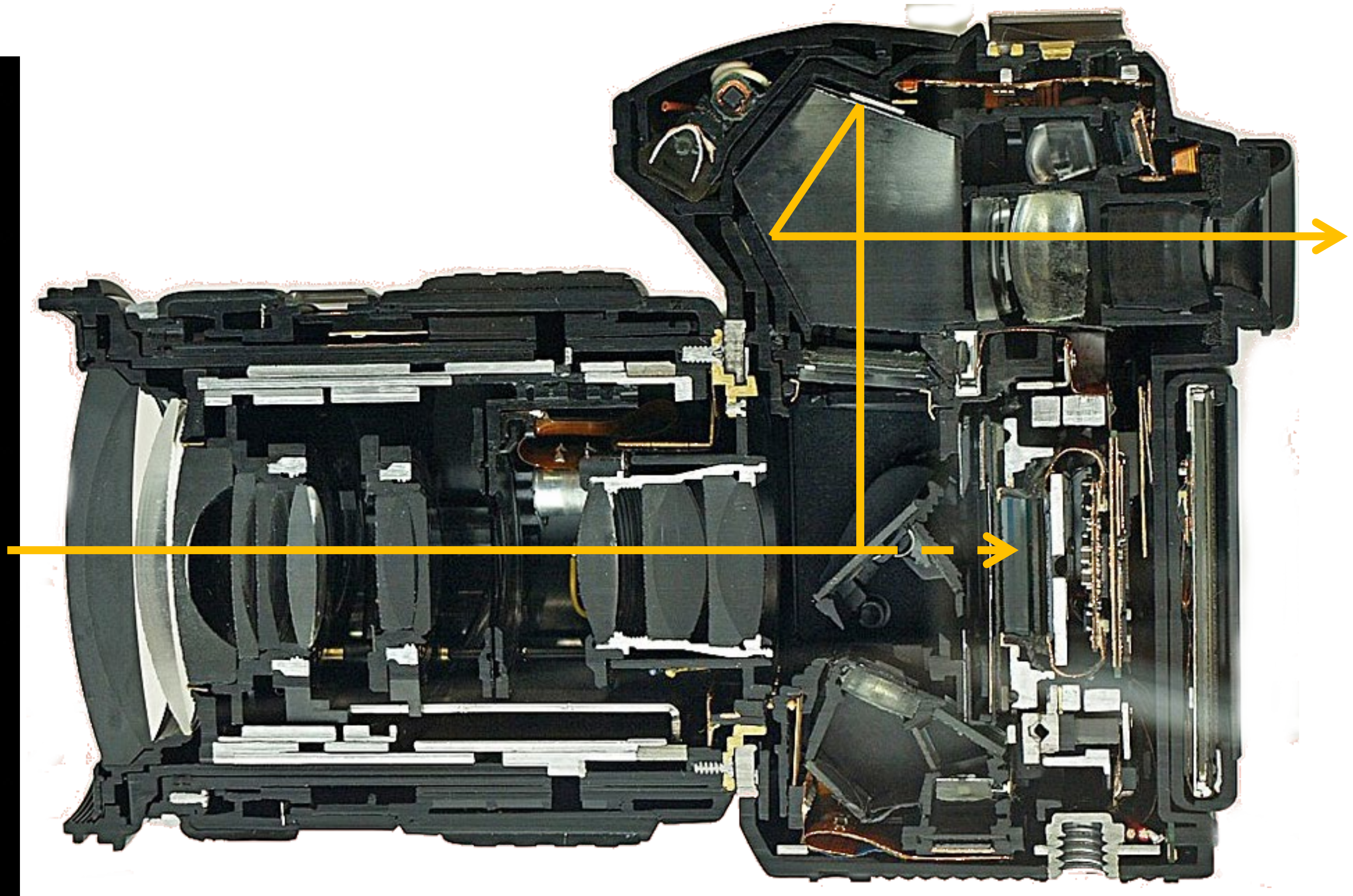




Manual focus in (D)SLR cameras

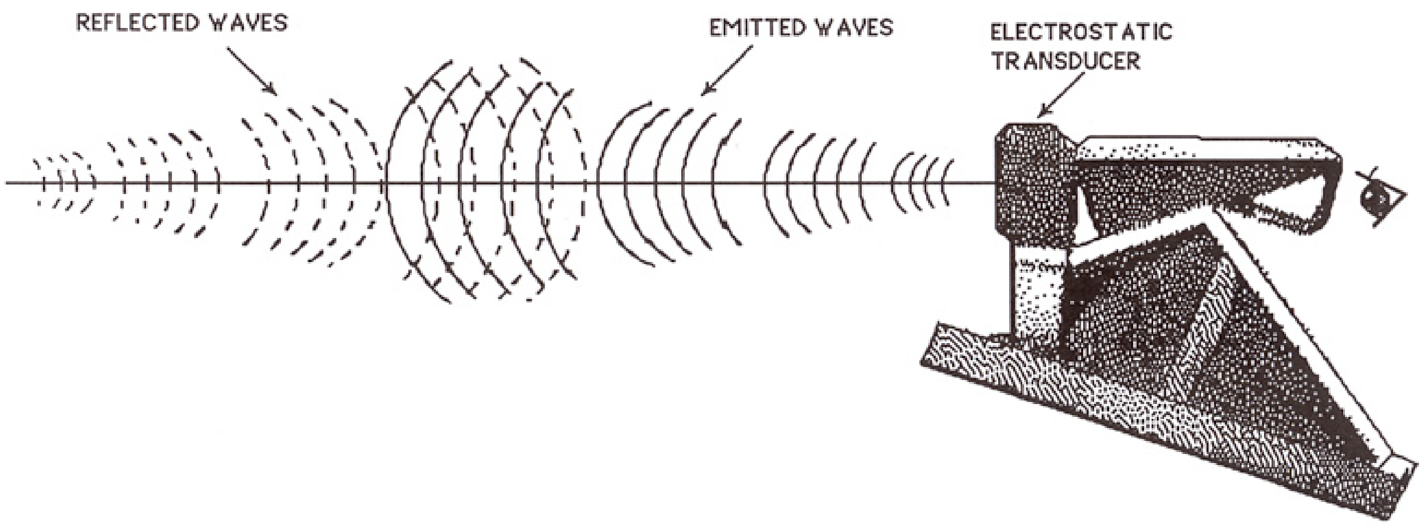
- Same view as main lens.
- Just rotate the focusing ring until you are satisfied by the sharpness.
- Viewfinder indicators can help this process.





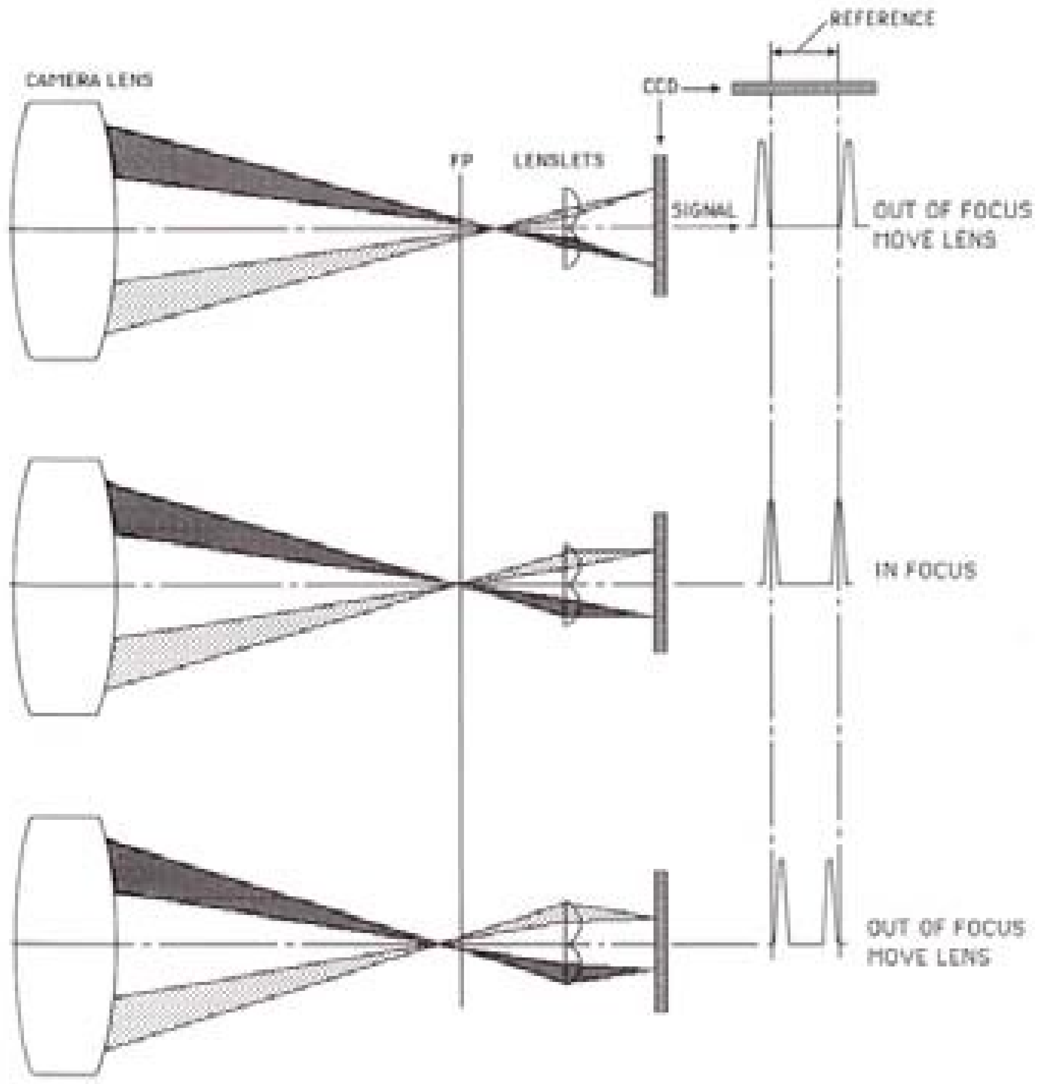
Active auto-focus: time-of-flight sensors

- Basically how SONAR works (we'll also see time-of-flight sensors later in class.
- Method used in Polaroid cameras, which used ultrasound waves.
- Energy inefficient.
- Limited range.
- Multi-path interference (e.g., glass surfaces back-reflected the waves).



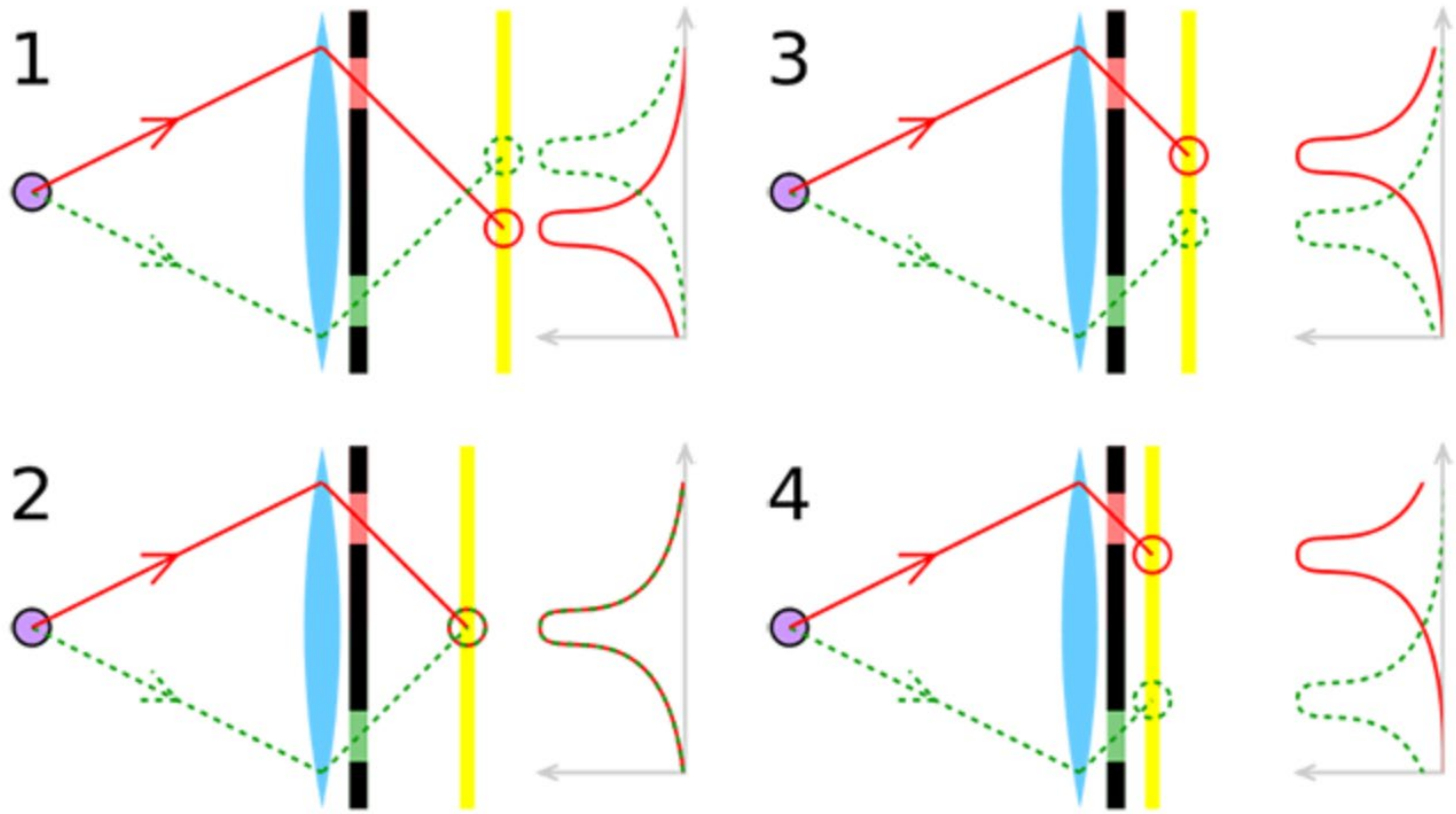
Passive auto-focus: phase detection

- As the lens moves, ray bundles from an object converge to a different point in the camera and change in angle.
- This change in angle causes them to refocus through two lenslets to different positions on a separate AF sensor.
- A certain spacing between these double images indicates that the object is “in focus”.



Demo: http://graphics.stanford.edu/courses/cs178/applets/autofocuspd.html

Passive auto-focus: phase detection



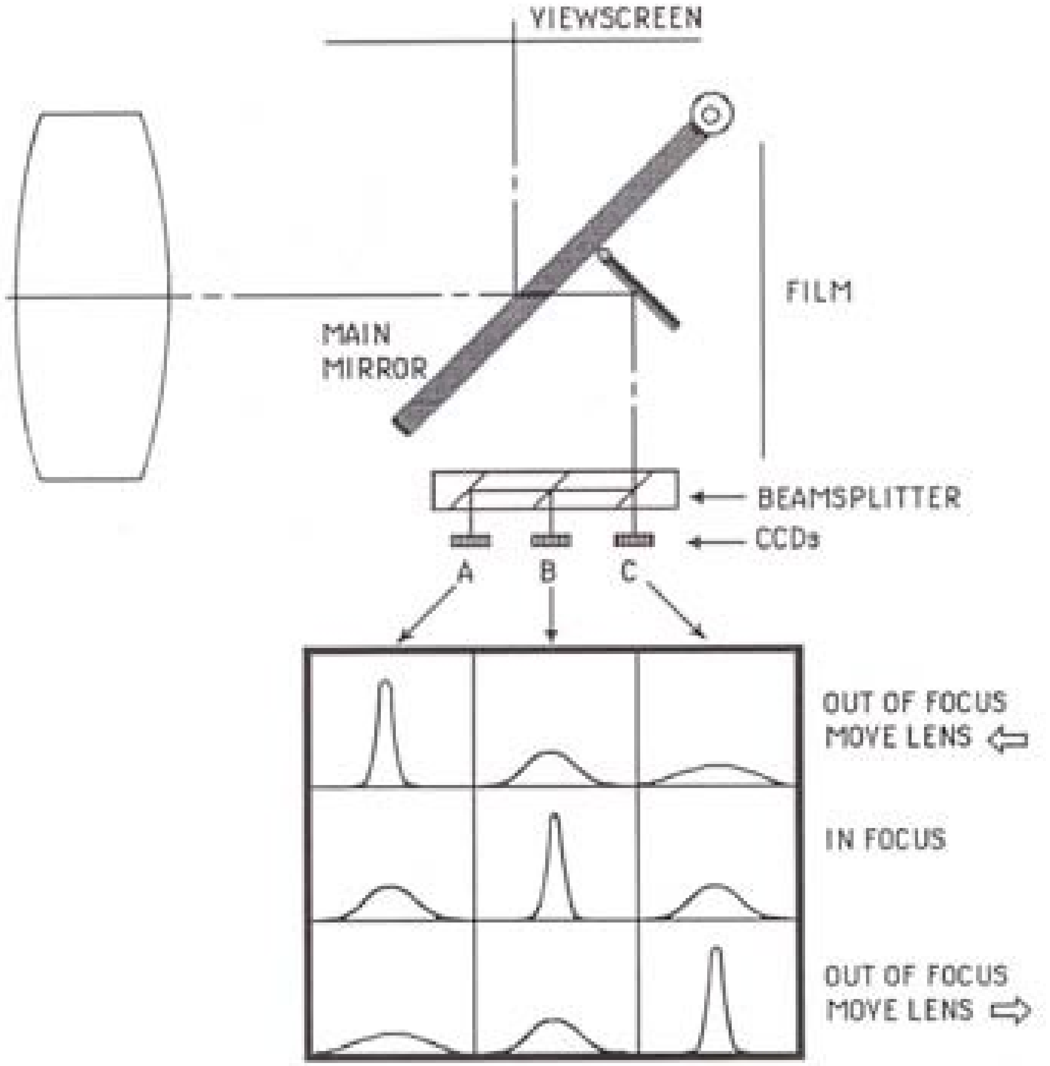
Each yellow box indicates *two* sensors, each measuring light from different parts of the aperture.

- Which one is correct focusing?
- How do you need to move the lens or sensor to get correct focusing?

Passive auto-focus: contrast detection

- Sensors at different image distances will see the same object at high-contrast if it's in focus, or low-contrast if it's not.
- Move the lens until the high-contrast sub-image falls on the middle sensor, which is conjugate to the camera's main sensor.
- Compute contrast using local differences of pixel values.

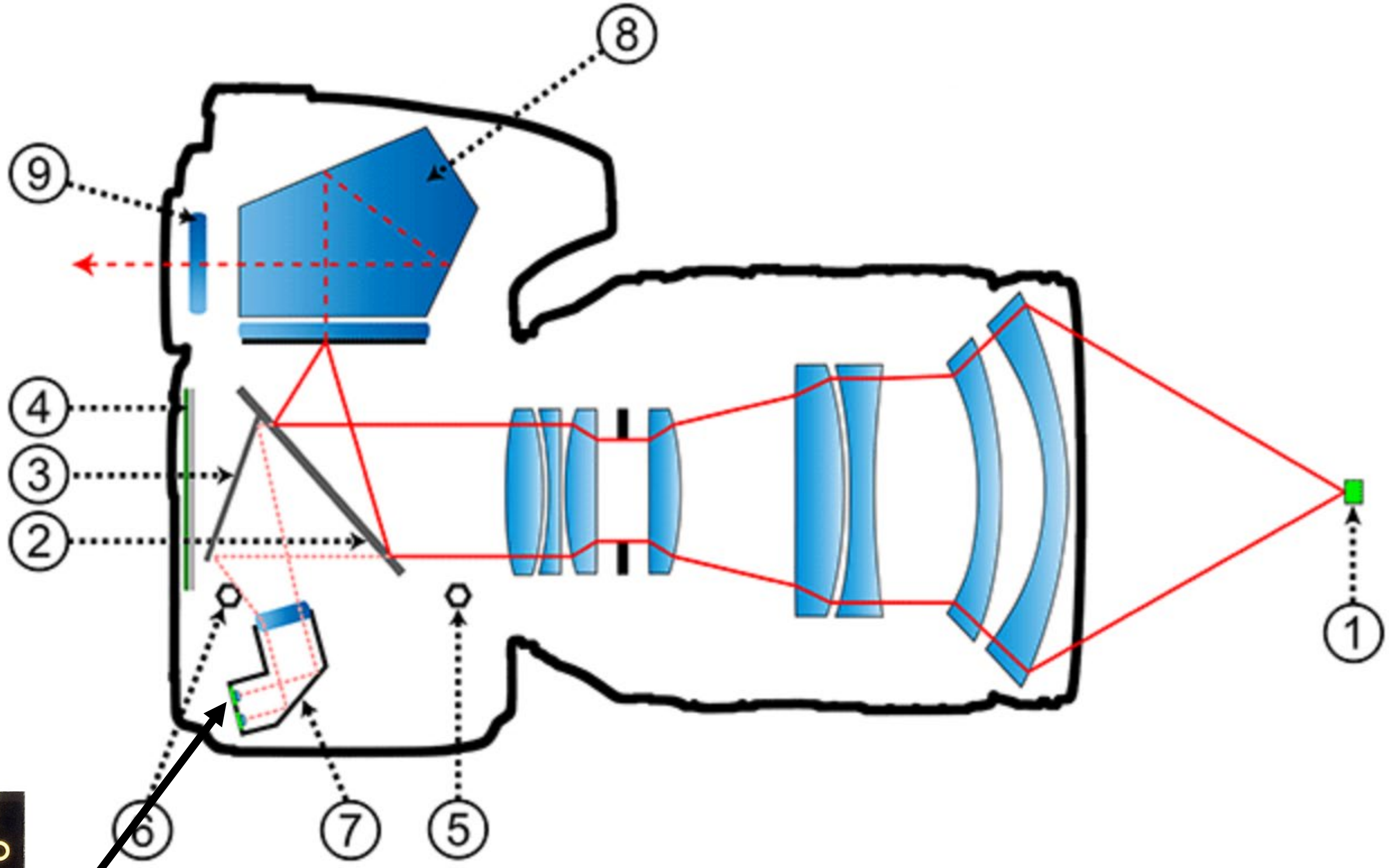
Effectively the same as depth from focus.

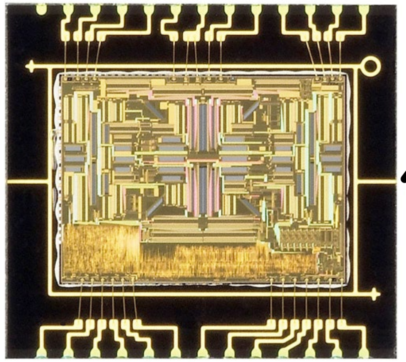


Demo: http://graphics.stanford.edu/courses/cs178/applets/autofocuscd.html

High-end DSLRs use phase detection

- Distance between sub-images allows lens to move directly into focus, without hunting
- Many AF points corresponding to different points on imaging sensor, complicated algorithms for choosing among them: generally use closest point, but also consider position in FOV.





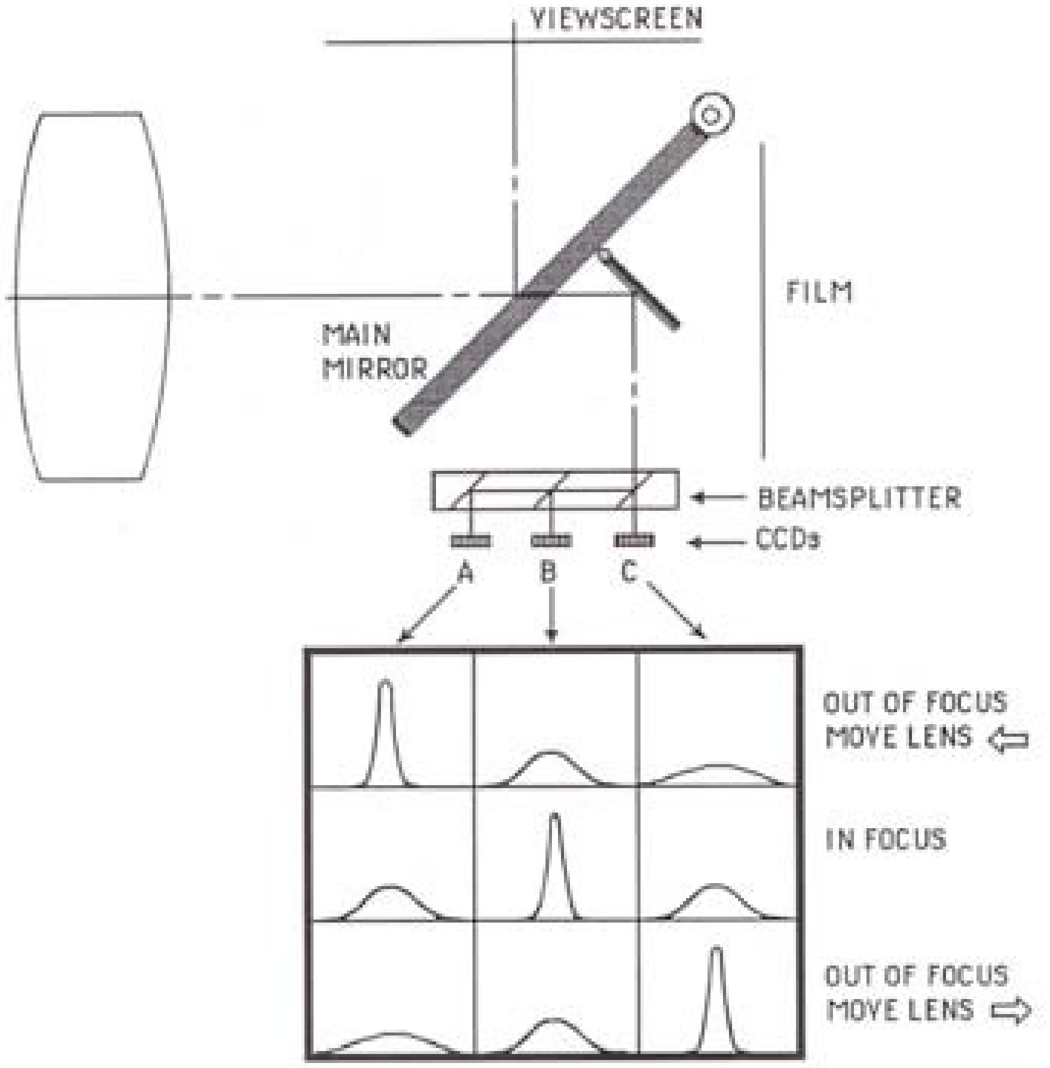
DSLR mirror has a translucent part that allows some light to make it to the AF sensor.

Low-end cameras (and phones) use contrast detection

- Nowadays it is mostly done using main camera sensor instead of dedicated sensors.
- Requires repeated measurements as lens moves, which are captured using the main sensor (an "autofocus stack").
- Slow, requires hunting, suffers from overshooting.

But

- People have come up with creative uses for the autofocus stack (depth-from-focus on a cell phone, HDR+ on Android).



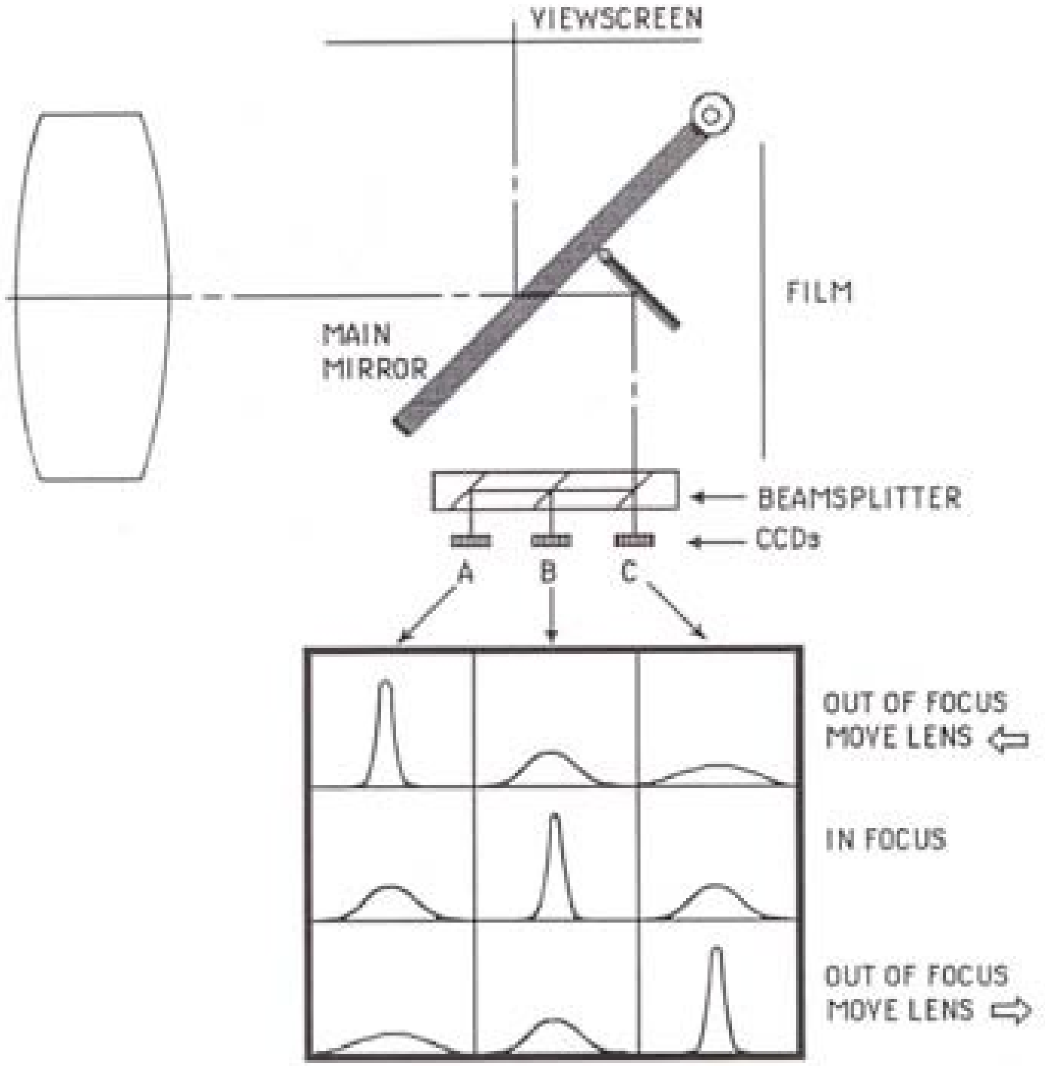
[Suwajanakorn et al., CVPR 2015; Hasinoff et al., SIGGRAPH Asia 2016]

Low-end cameras (and phones) use contrast detection

- Nowadays it is mostly done using main camera sensor instead of dedicated sensors.
- Requires repeated measurements as lens moves, which are captured using the main sensor (an "autofocus stack").
- Slow, requires hunting, suffers from overshooting.

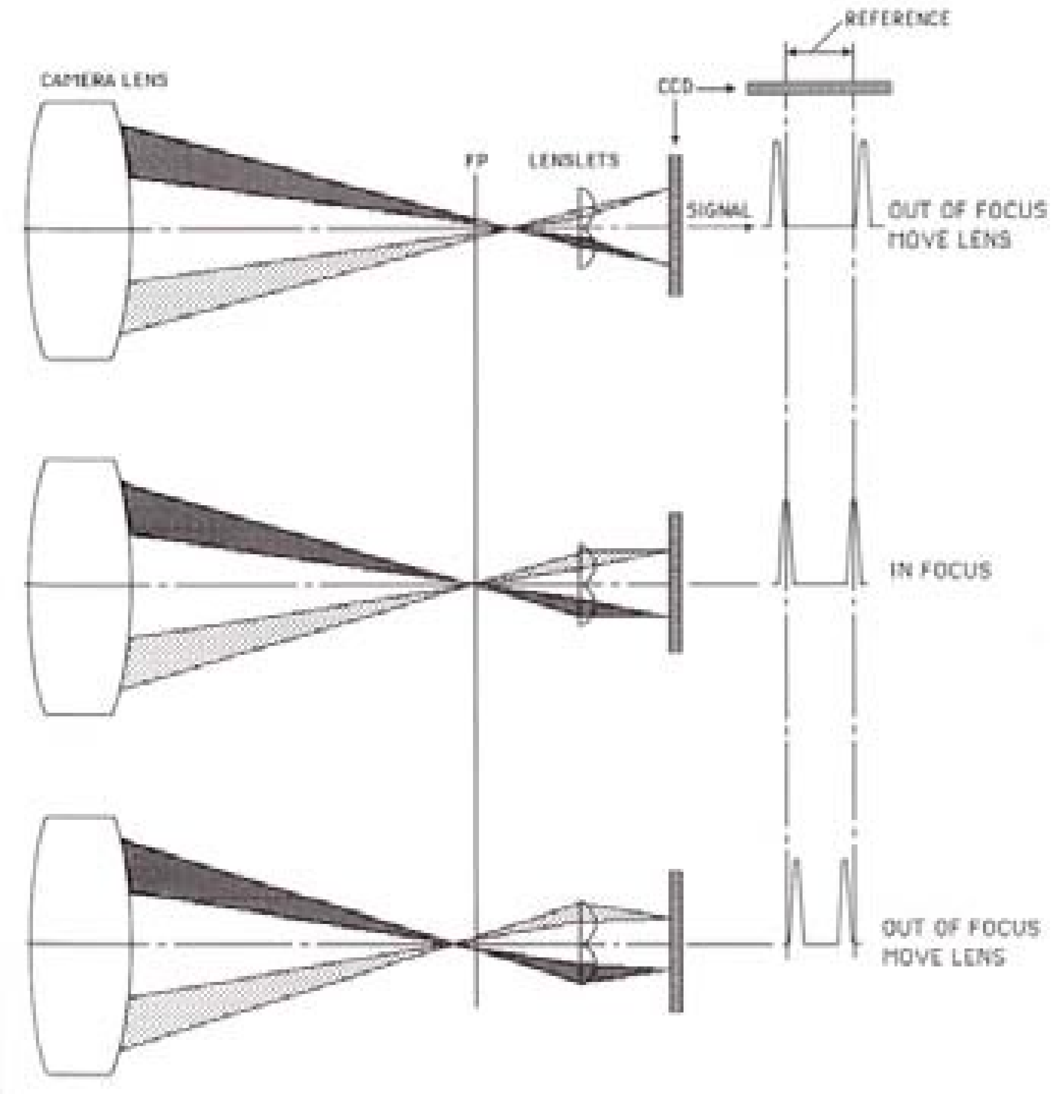
But

- People have come up with creative uses for the autofocus stack (depth-from-focus on a cell phone, HDR+ on Android).



[Suwajanakorn et al., CVPR 2015; Hasinoff et al., SIGGRAPH Asia 2016]

Modern mirror-less cameras use phase detection

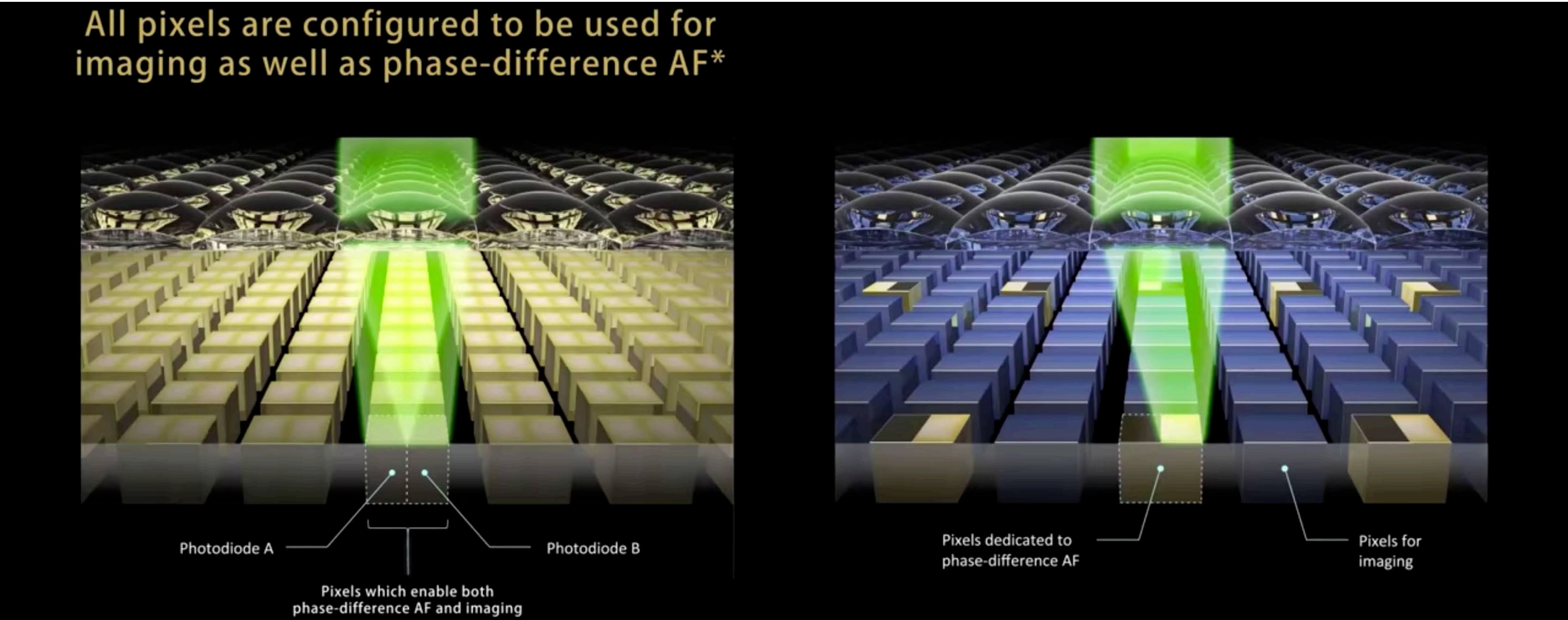


- Dedicate a small number of pixels on the imaging sensor to work for phase detection.
- Do this at different parts of the sensor to be able to autofocus at different parts of the image.

Any downsides?

Dual-pixel phase detection autofocus

- Split each pixel into two independent photodiodes.
- Use different pixels for phase detection.
- Many other interesting opportunities (more on this during the stereo lecture).



Should you use autofocus?

Should you use autofocus?

Quick answer: Yes.

Should you use autofocus?

Quick answer: Yes.

More detailed answer: Yes, except for certain special circumstances.

- You are using a lens that does not have an autofocus motor (e.g., vintage or otherwise old lenses, high-end lenses, industrial and machine vision lenses).
- You are trying to capture an image under conditions where autofocus is prone to fail (e.g., macrophotography, poorly-lit scenes, imaging through glass or occluders).
- You intentionally want some part of the scene to be out of focus (e.g., for artistic effect, or because you want a face or other featured to be obscured).
- You are in an once-in-a-lifetime opportunity to photograph something, and you cannot afford to risk autofocus failing. This additionally assumes that:
  - Your scene is static enough that you can take the time to focus manually.
  - You are experienced enough so that the probability of manual focus failing is smaller than the probability of autofocus failing.

References

Basic reading:
- Szeliski textbook, Section 12.1.3, 13.3.

Additional reading:

- Nayar et al., “Real-Time Focus Range Sensor,” PAMI 1996.
  One of the (relatively) early papers on depth from defocus.
- Suwajanakorn et al., “Depth from Focus with Your Mobile Phone,” CVPR 2015.
  Implementing depth from defocus on a mobile phone using the autofocus focal stack.
- Hasinoff and Kutulakos, “Confocal Stereo,” IJCV 2009.
  The paper on high resolution depth from a focus and aperture stack.
- Kuthirummal et al., “Flexible Depth of Field Photography,” PAMI 2010.
  Continuously change focus within one exposure, without stopping to capture a stack.
- Levoy and Hanrahan, “Light Field Rendering,” SIGGRAPH 1996.
- Gortler et al., “The Lumigraph,” SIGGRAPH 1996.
  The two papers introducing the light field.
- Adelson and Wang, “Single Lens Stereo with a Plenoptic Camera,” PAMI 1992.
  The paper (re)-introducing the plenoptic camera to computer vision and graphics.
- Ng et al., “Light field photography with a hand-held plenoptic camera,” Stanford TR 2005.
  The paper (re)-(re)-introducing the plenoptic camera, and the precursor to Lytro.
- Ng, “Fourier Slice Photography,” SIGGRAPH 2005.
  The paper on frequency-space analysis of refocusing and lightfield measurements.
- Wilburn et al., “High Performance Imaging Using Large Camera Arrays,” SIGGRAPH 2005.
  The camera array paper.
- Georgiev et al., “Unified Frequency Domain Analysis of Lightfield Cameras,” ECCV 2008.
  Make a lightfield camera from a pinhole array (and many other interesting stuff about lightfield cameras).