Stereo Vision

Fiseha B. Tesema, PhD

Recap

- Pinhole cameras
  - Perspective projection
  - Orthographic Projection
- Camera Modeling
  - What is the effect of varying aperture size?
  - Cameras and lenses
- Properties of Perspective Projection
- Propertied of Orthographic Projection
- Going to digital image space
  - Intrinsic and Extrinsic parameter

$$P' = K \begin{bmatrix} R & T \end{bmatrix} P_w = M P_w \tag{10}$$

- Camera Calibration

Outline

- Stereo Vision
- Geometry for a simple stereo system
- Epipolar geometry of the stereo system
- Multiview Stereo

Stereo Vision

- Close one eye and hold your finger up. Switch eyes. your finger appears to shift against the background.
- Each eye sees your finger from a slightly different angle (like how astronomers observe stars from different points in Earth's orbit).
- The closer an object (your finger), the **larger** the apparent shift (parallax effect).
- Distant objects (background) appear to move **less** because their parallax is smaller.
- **Parallax** is the apparent shift in the position of an object when viewed from two different points.

parallax

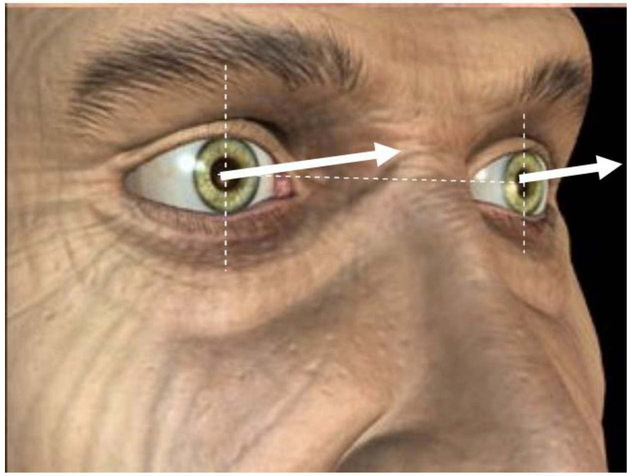
Stereo Vision

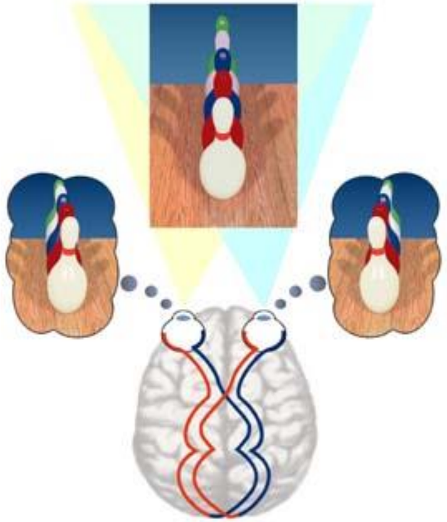
- Stereopsis is the component of depth perception retrieved by means of binocular disparity through binocular vision.
- The term "stereopsis" comes from Ancient Greek στερεός (stereós) 'solid', and ὄψις (ópsis) 'appearance, sight'.

[https://en.wikipedia.org/wiki/Stereopsis]

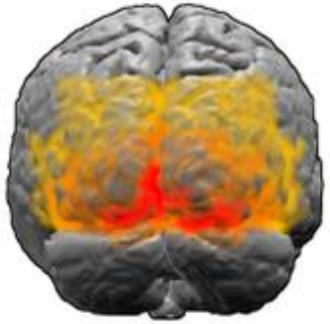
Human Stereoscopic Vision

Binocular vision occurs because each eye (left and right) receives a different image due to their slightly different positions in one's head. These positional differences are referred to as "horizontal disparities" or, more generally, "binocular disparities". Disparities are processed in the visual cortex of the brain to yield depth perception.





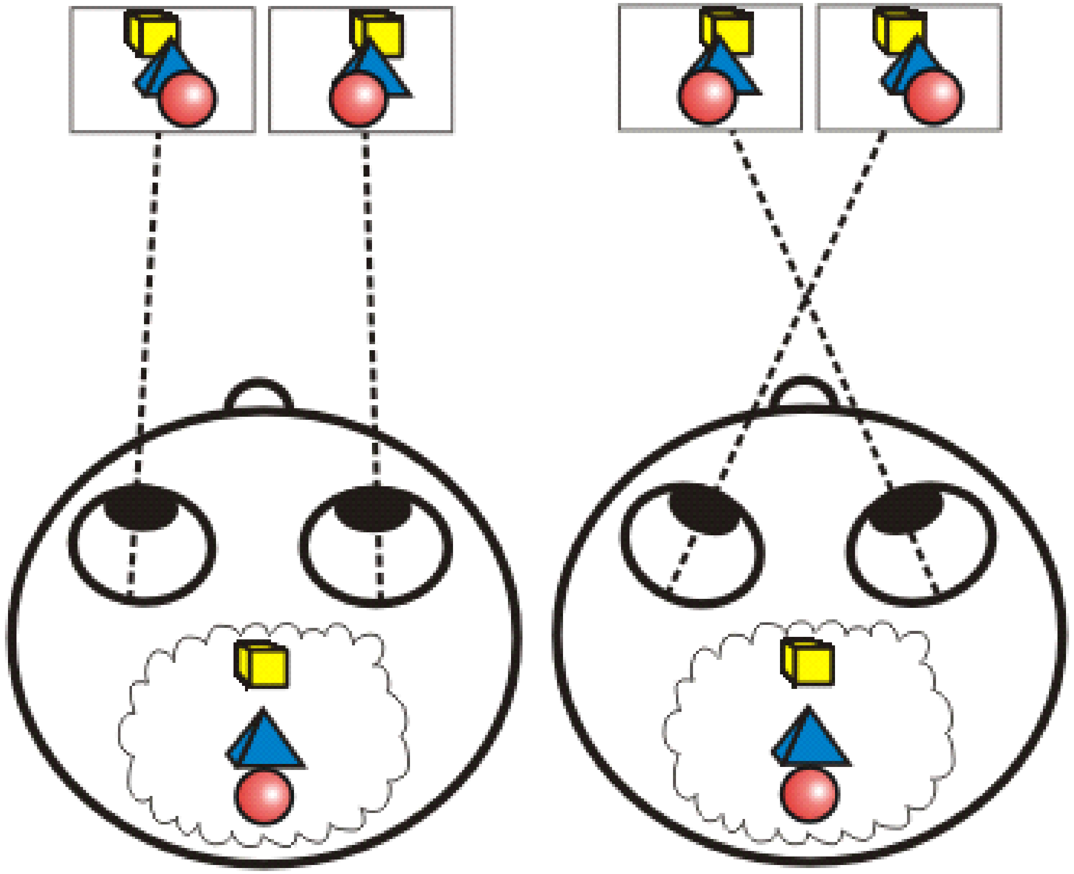
[http://scecinfo.usc.edu]



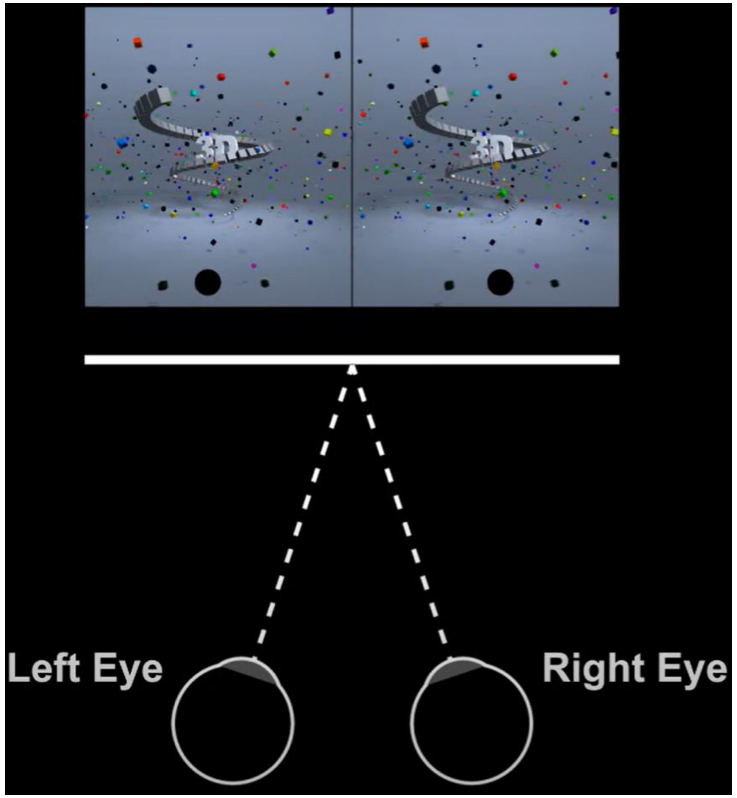
Visual Cortext

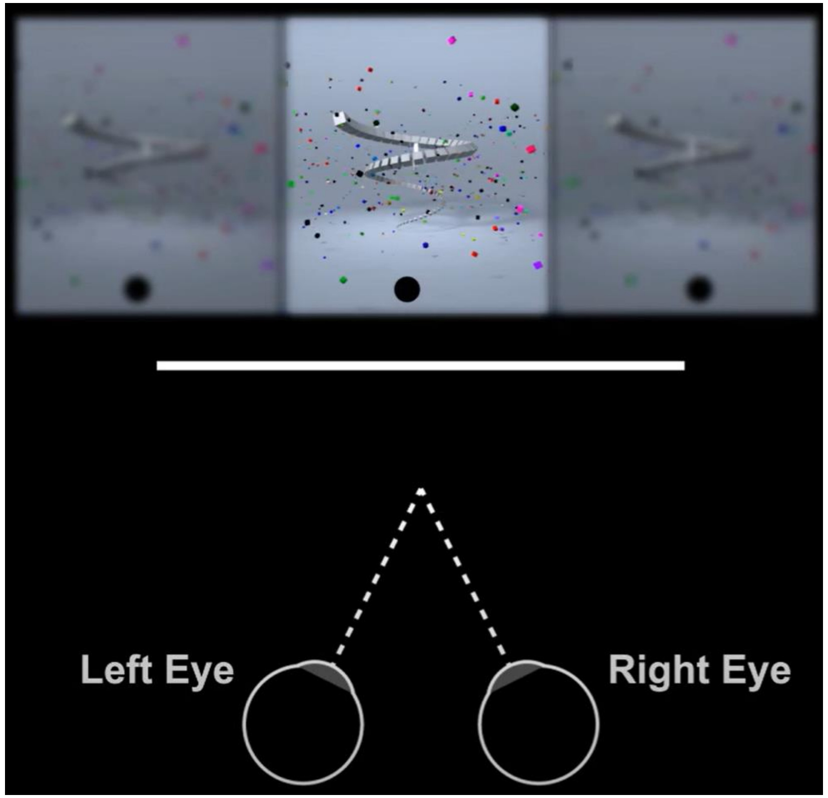
https://en.wikipedia.org/wiki/Stereopsis

Parallel/Cross-eyed depth perception



Parallel/Cross-eyed depth perception





While watching the following movie, try to focus your eyes and this bringing the two black dots at the bottom together.
https://www.youtube.com/watch?v=zBa-bCxsZDk
https://www.youtube.com/watch?v=ppL8SrHq9VM

Stereoscope

- In 1840, Sir Charles Wheatstone developed the stereoscope.
- Stereograms were popular in the early 1900's
- Simulating 3D by artificially presenting two different images separately to each eye using a method called stereoscopy.

Brewster-type stereoscope, 1870

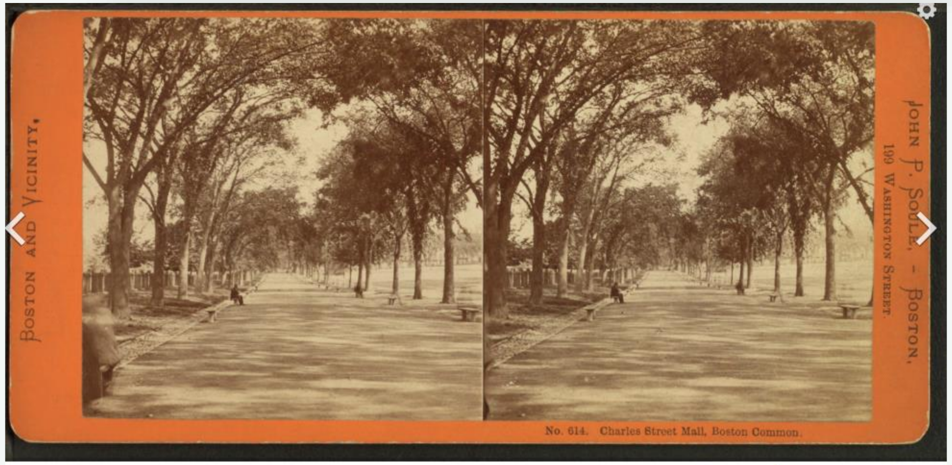
More details

Alessandro Nassiri - Museo della Scienza e della Tecnologia "Leonardo da Vinci"

Visore stereoscopico portatile di tipo Brewster, J. Fleury - Hermagis, 1870, con messa a fuoco manuale. Per la visione di lastre e stampe stereoscopiche 8,5x17cm. Museo nazionale della scienza e della tecnologia Leonardo da Vinci, Milano.

File: IGB 006055 Visore stereoscopico portatile Museo scienza e tecnologia Milano.jpg

Created: 1 July 2014



View of Boston, c. 1860; an early stereoscopic card for viewing a scene from nature

More details

Soule, John P., 1827-1904 -- Photographer - This image is available from the New York Public Library's Digital Library under the digital ID G90F336_113F: digitalgallery.nypl.org → digitalcollections.nypl.org

File: Charles Street Mall, Boston Common, by Soule, John P., 1827-1904 3.jpg

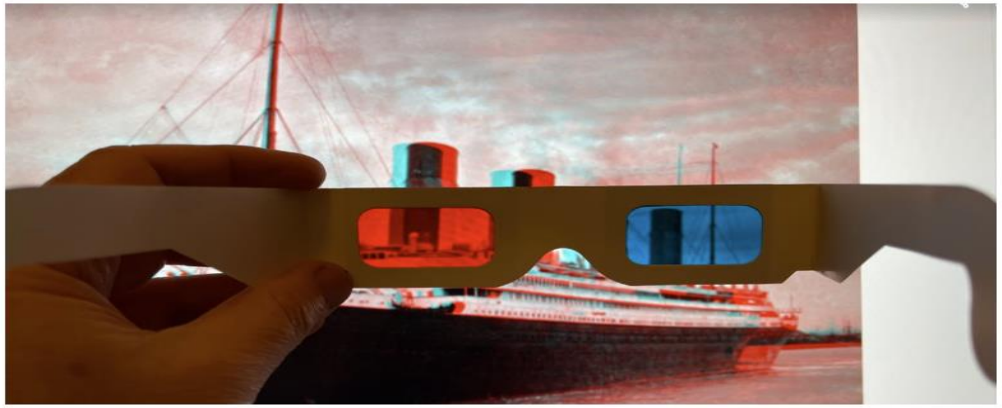
Created: Coverage: 1860?-1890?. Source Imprint: 1860?-1890?. Digital item published 7-28-2005; updated 4-23-2009.

Anaglyphs

- Anaglyphs are a way of encoding parallax in a single picture. Two slightly different perspectives of the same subject are superimposed on each other in contrasting colors, producing a three-dimensional effect when viewed through two correspondingly colored filters.
- 3D movies were popular in the 1950's
- The left and right images were displayed as red and blue.

Stereo images of the Titanic (Anaglyph method)

(a)

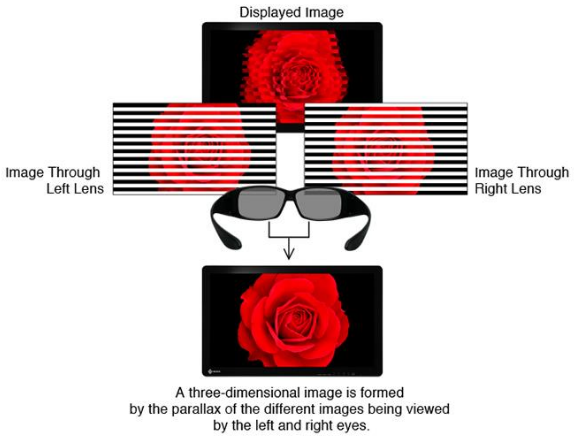


(b)

Figure 1.1: (a) Stereo anaglyph of the ocean liner, the Titanic [McManus2022]. The red image shows the right eye's view, and cyan the left eye's view. When viewed through stereo red/cyan stereo glasses, as in (b), the cyan contrast appears in the left eye image and the red variations appear to the right eye, creating a the perception of 3d.

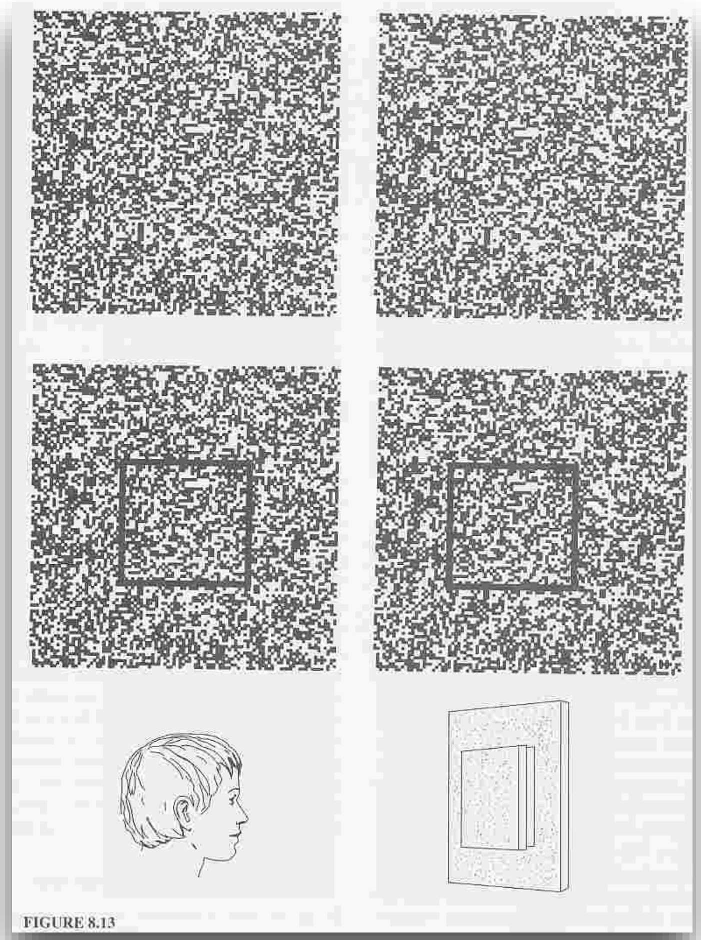
Polarized Glass

- Current technology for 3D movies and computer display is to use polarized glasses
- The viewer wears eyeglasses with a polarizing filter for each eye. The left and right filters have different polarizations, so each eye receives only the image with the matching polarization.



https://www.eizo.com/library/healthcare/the_advantages_of_using_polarized_3d_technology_for_surgeries/

Depth without objects

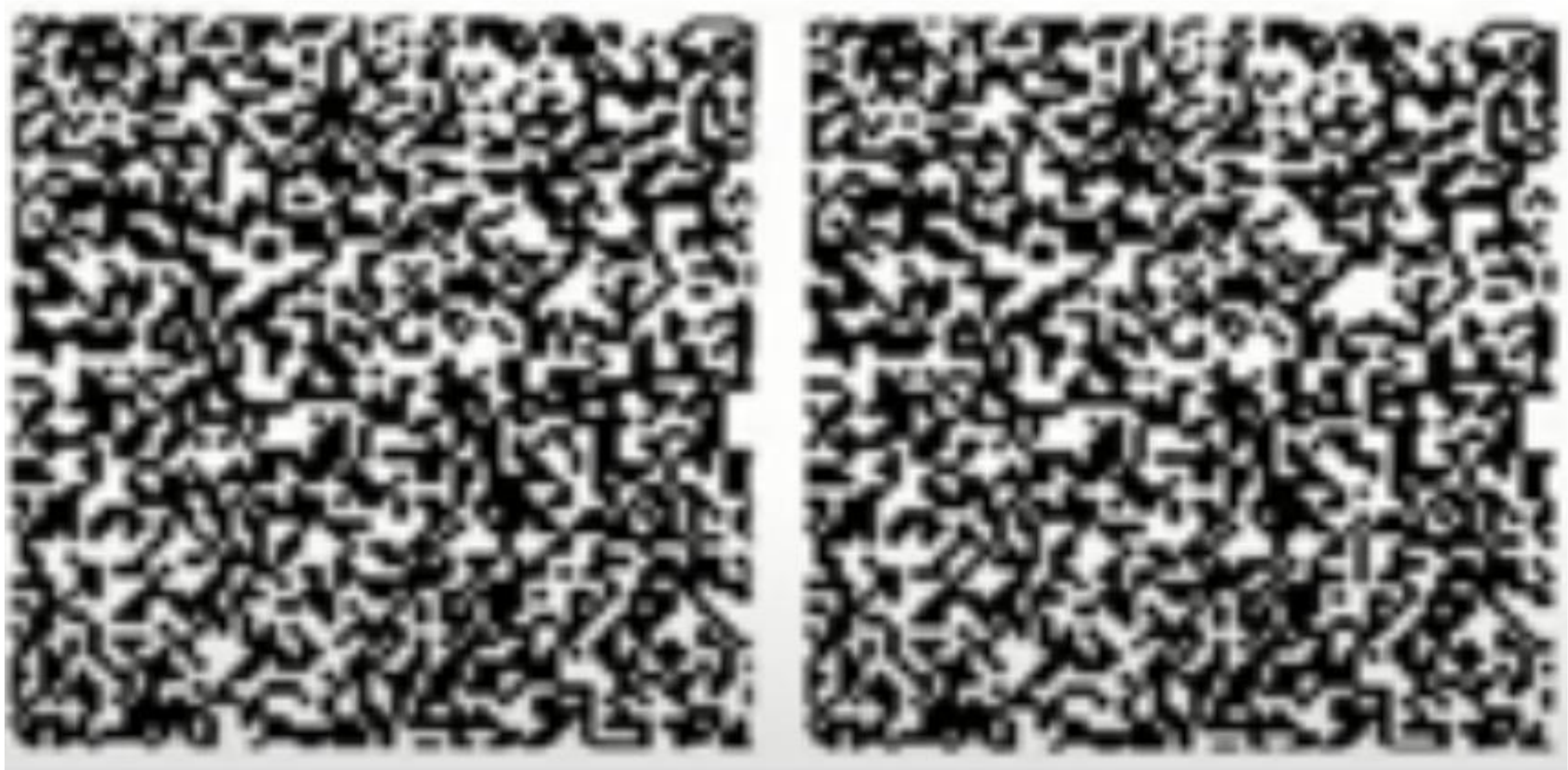


Random dot stereograms (Bela Julesz)

- The first person to put it in digital format.

Julesz, 1971

What Can you see?



Want develop your first random-dot stereogram?

1. Create an image of suitable size. Fill it with random dots. Duplicate the image.

Want develop your first random-dot stereogram?

2. Select a region in one image, in this case, in the right image.

Want develop your first random-dot stereogram?

3. Shift this region horizontally by one or two dot diameters and fill in the empty region with new random dots. The stereogram is complete.

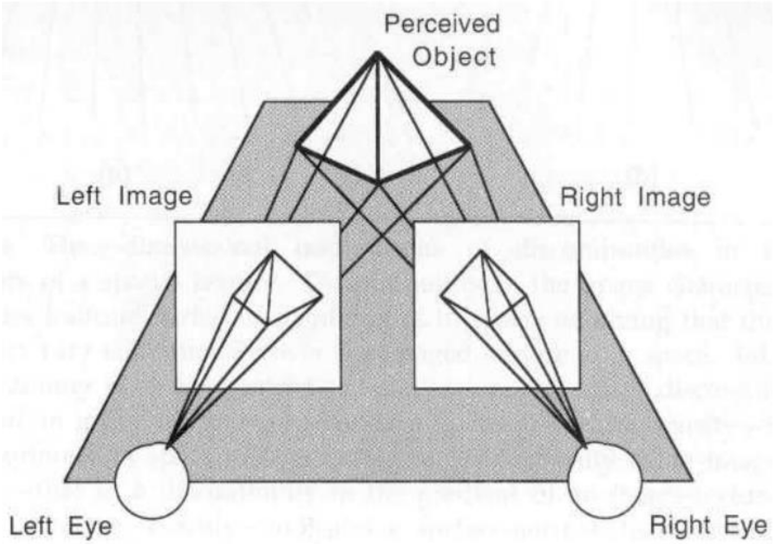
Geometry for a simple stereo system

Simple stereo

- Simple stereo vision (Horizontal stereo)
  - Simple method for recovering the three dimensional structure of a scene from two images.

Simple (Calibrated) Stereo vision

- The recovery of the 3D structure of a scene using two or more images of the 3D scene, each acquired from a different viewpoint in space.
- The images can be obtained using multiple cameras or one moving camera.
- The term binocular vision is used when two cameras are employed.



The two problems of stereo

- The correspondence problem.
  - Finding pairs of matched points such that each point in the pair is the projection of the same 3D point.
  - Triangulation depends crucially on the solution of the correspondence problem.
- The reconstruction problem.
  - Given the corresponding points, we can compute the disparity map.
  - The disparity map can be converted to a 3D map of the scene (i.e., recover the 3D structure) if the stereo geometry is known.

Let's consider one camera

One camera

Two cameras

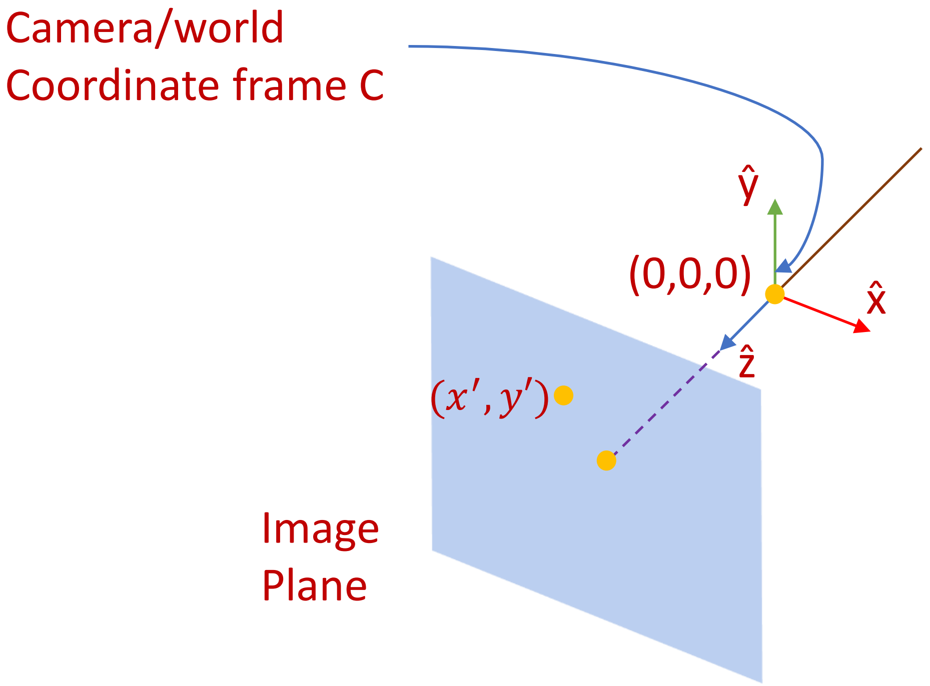
N cameras

Backward Projection: From 2D to 3D

- Given a calibrated camera, can we find the 3D scene point from a single 2D image?

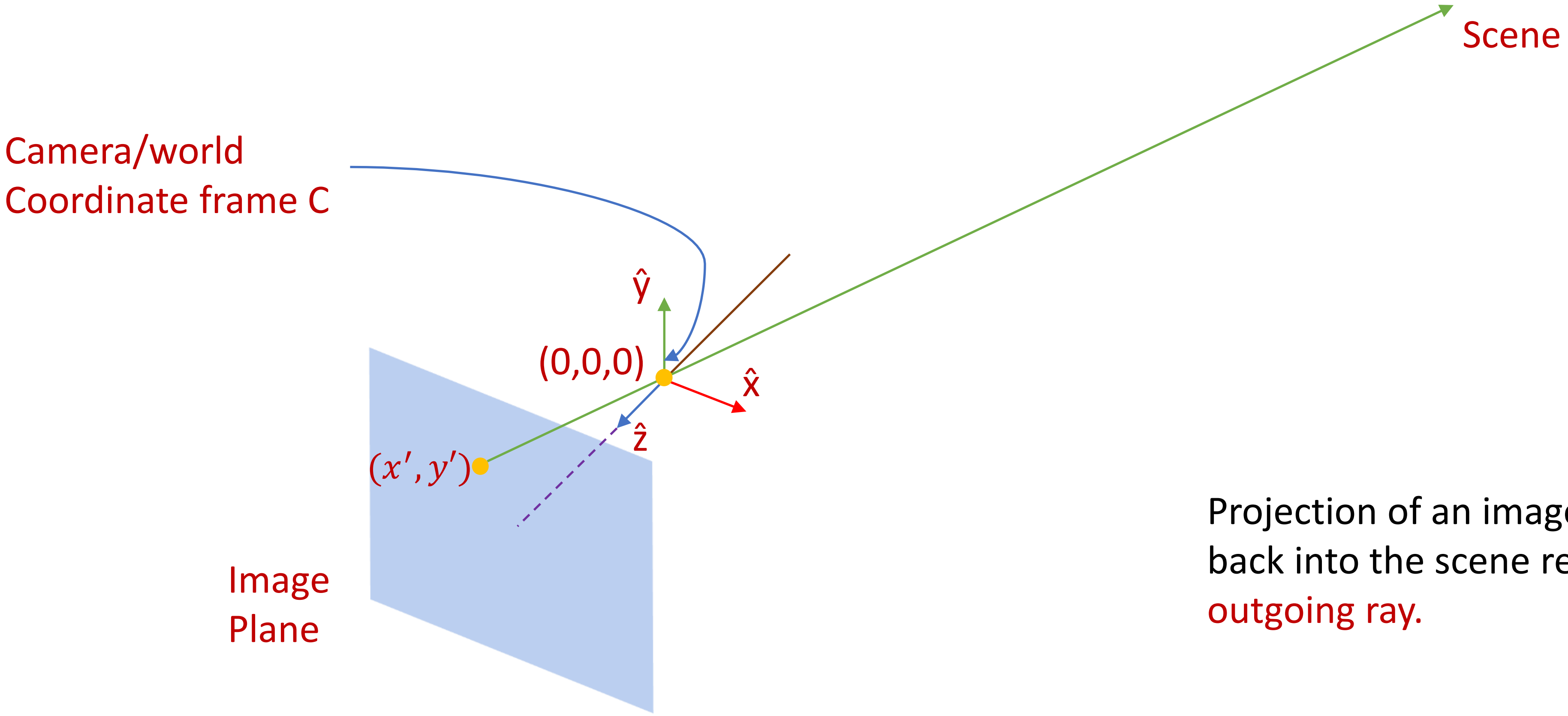
Backward Projection: From 2D to 3D

- Given a calibrated camera, can we find the 3D scene point from a single image? NO!!!!



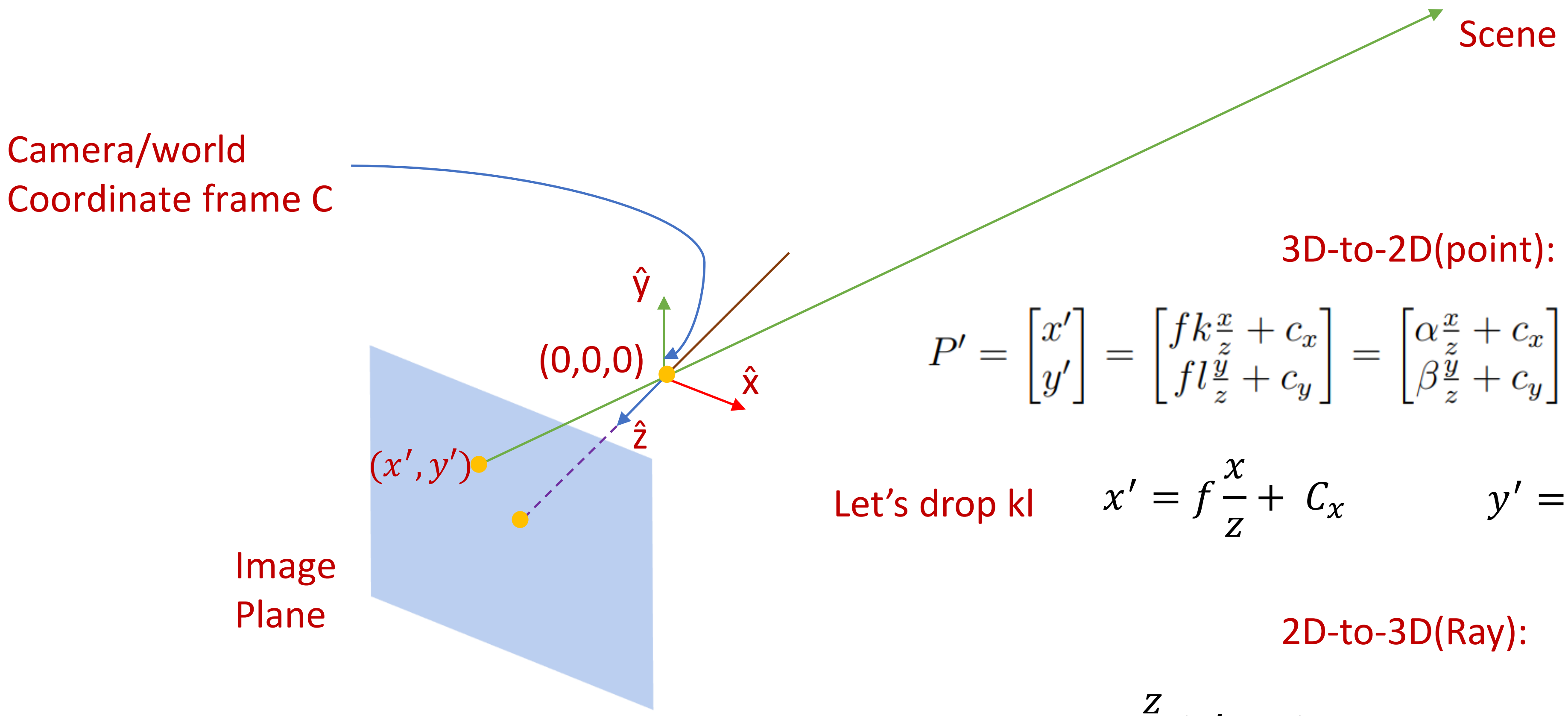
Backward Projection: From 2D to 3D

- Given a calibrated camera, can we find the 3D scene point from a single image? NO!!!!



Projection of an image point back into the scene results in an outgoing ray.

Backward Projection: From 2D to 3D



3D-to-2D(point):

$$P' = \begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} fk\frac{x}{z} + c_x \\ fl\frac{y}{z} + c_y \end{bmatrix} = \begin{bmatrix} \alpha\frac{x}{z} + c_x \\ \beta\frac{y}{z} + c_y \end{bmatrix} \tag{4}$$

Let's drop kl

$$x' = f\frac{x}{z} + C_x \qquad y' = f\frac{y}{z} + C_y$$

2D-to-3D(Ray):

$$x = \frac{z}{f}(x' - C_x) \qquad y = \frac{z}{f}(y' - C_y)$$

$$z > 0$$

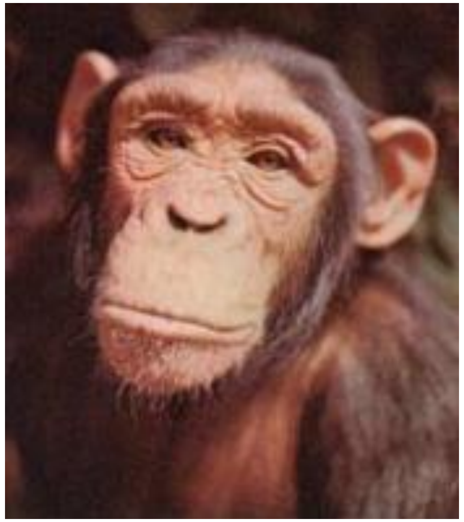
We can’t compute 3D from single 2D image

We need more information!

Let’s consider two eyes (Stereo Vision

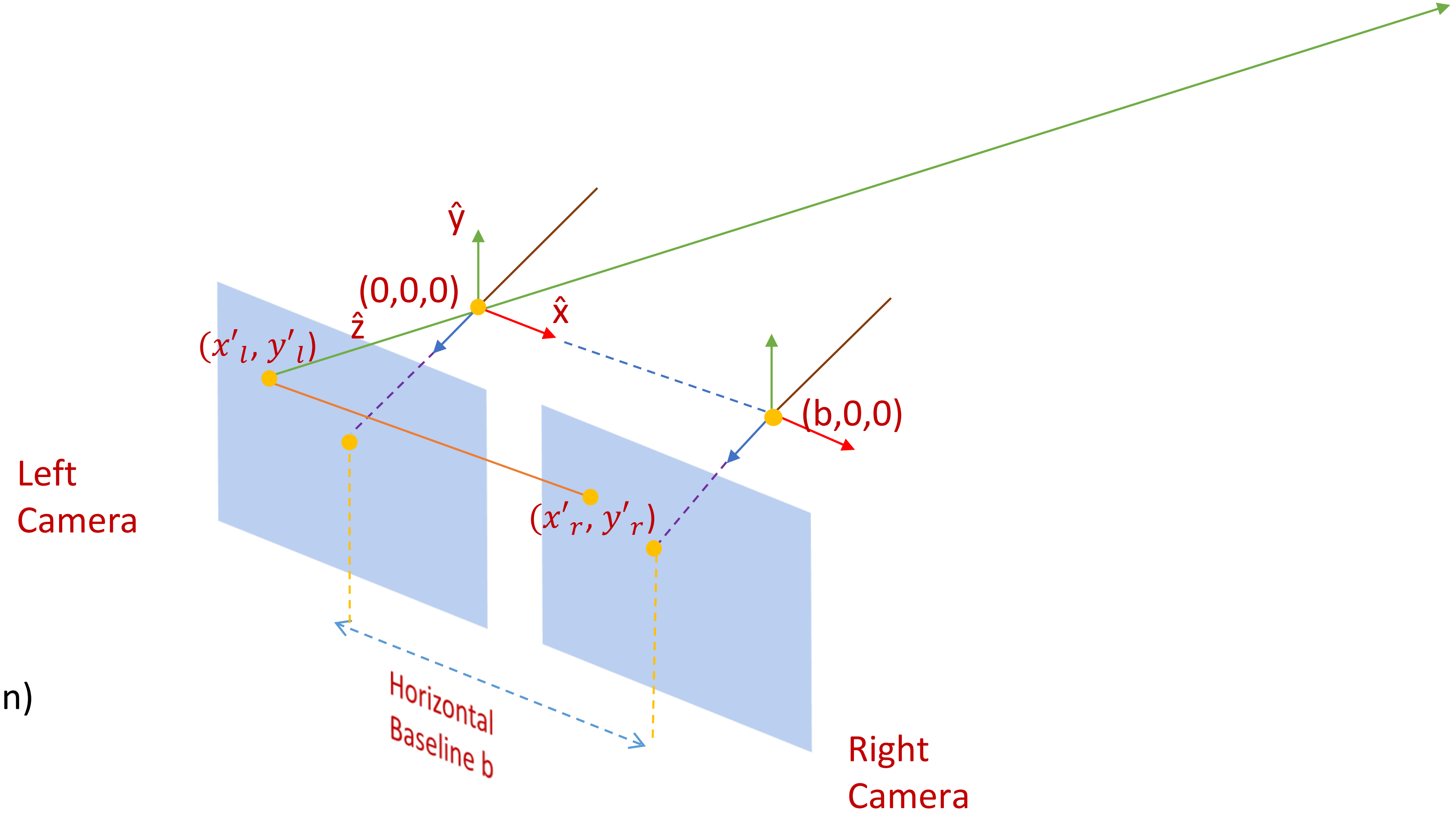
One camera

Two cameras



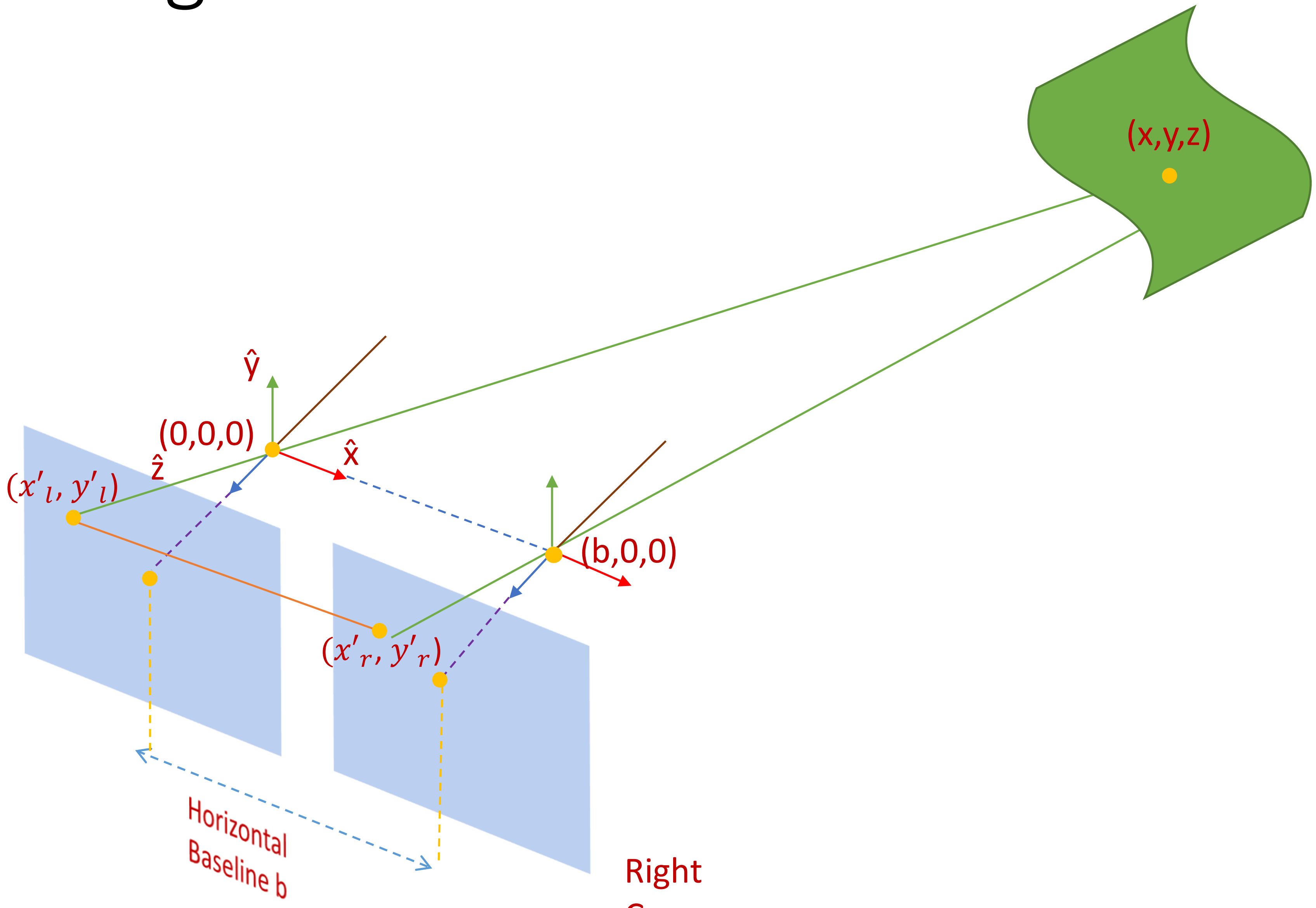
N cameras

Triangulation using Two Cameras



Stereo System
(Binocular Vision)

Triangulation using Two Cameras



Left
Camera

Right
Camera

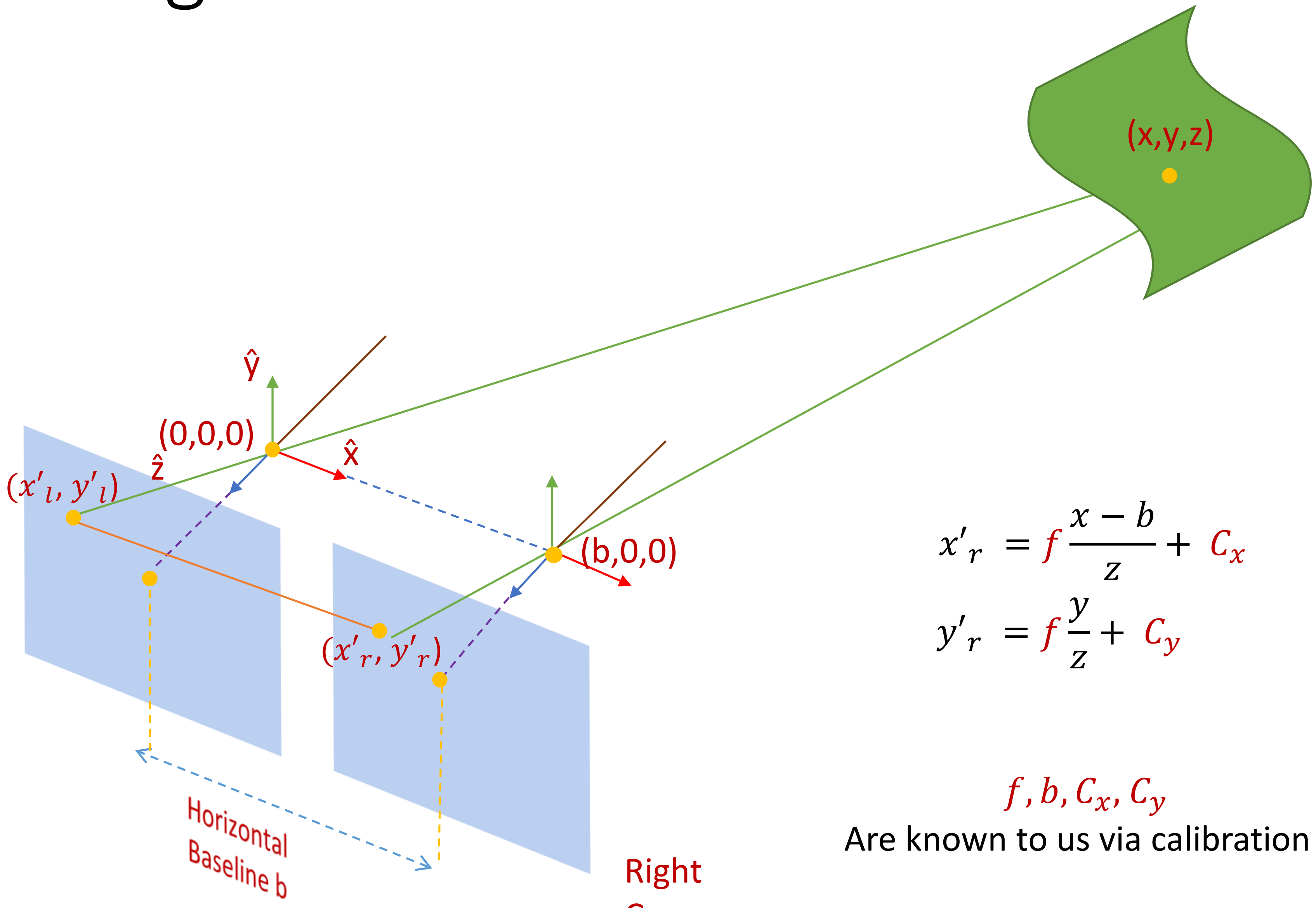
Stereo System
(Binocular Vision)

Triangulation using Two Cameras

$$x'_l = f\frac{x}{z} + C_x$$

$$y'_l = f\frac{y}{z} + C_y$$

Left Camera



Right Camera

$$x'_r = f\frac{x-b}{z} + C_x$$

$$y'_r = f\frac{y}{z} + C_y$$

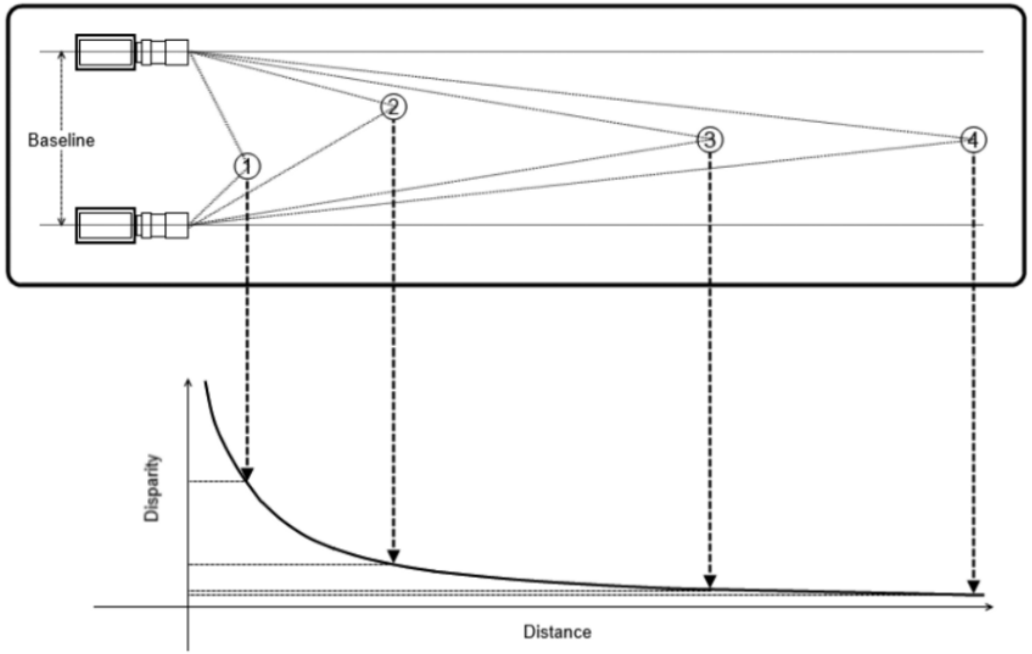
Stereo System
(Binocular Vision)

f, b, C_x, C_y
Are known to us via calibration

Simple Stereo: Depth and Disparity

$$(x'_l, y'_l) = \left(f\frac{x}{z} + C_x,\ f\frac{y}{z} + C_y\right) \qquad (x'_r, y'_r) = \left(f\frac{x-b}{z} + C_x, f\frac{y}{z} + C_y\right)$$

$$x = \frac{b(x'_l - C_x)}{(x'_l - x'_r)} \qquad y = \frac{b(y' - C_y)}{(x'_l - x'_r)} \qquad \boxed{z = \frac{fb}{(x'_l - x'_r)}}$$



Where $(x'_l - x'_r)$ is called Disparity.

Depth z is inversely proportional to Disparity.
Disparity/Parallax is proportional to Baseline.

How we drive X,Y,Z?

$x'_l = f\frac{x}{z} + C_x$	$x'_r = f\frac{x-b}{z} + C_x$
$x = \frac{z}{f}(x'_l - C_x)$	$x = \frac{z}{f}(x'_r - C_x) + b$
$= \frac{z}{f}(x'_l - \cancel{C_x})$	$= \frac{z}{f}(x'_r - \cancel{C_x}) + b$

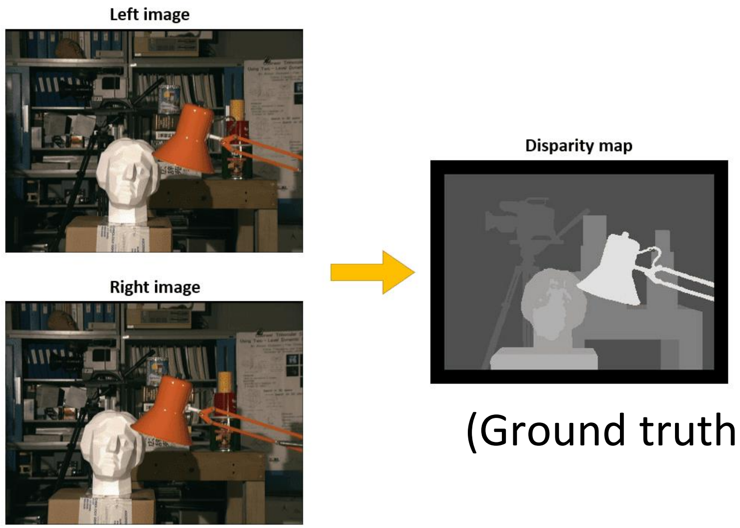
$$z = \frac{fb}{(x'_l - x'_r)}$$

Derivation of X,Y,

$x = \frac{z}{f}(x'_r - C_x) + b$	**Replce z with** $z = \dfrac{fb}{(x'_l - x'_r)}$
$x = \dfrac{b(x'_l - C_x)}{(x'_l - x'_r)}$	
$y = \dfrac{z}{f}(y' - C_y)$	**Replce z with** $z = \dfrac{fb}{(x'_l - x'_r)}$
$y = \dfrac{b(y' - C_y)}{(x'_l - x'_r)}$	

Stereo Matching(Finding correspondence): leads to finding Disparities

- Goal: Find the disparity between left and right stereo pairs.

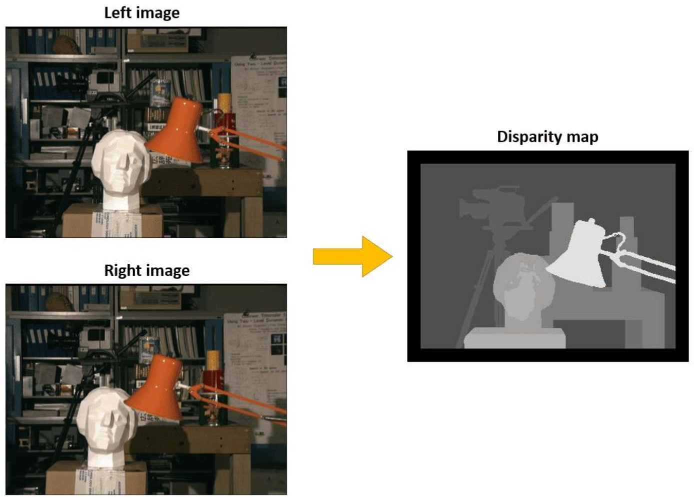


(Ground truth)

- The ground truth it a 3D dimensional scene measured by active illumination method.
- The close the points the greater the disparity and the brighter in the disparity map.

Stereo Matching(Finding correspondence): leads to finding Disparities

- The intersecting thing about the horizontal stereo system is that there is no disparity in vertical direction

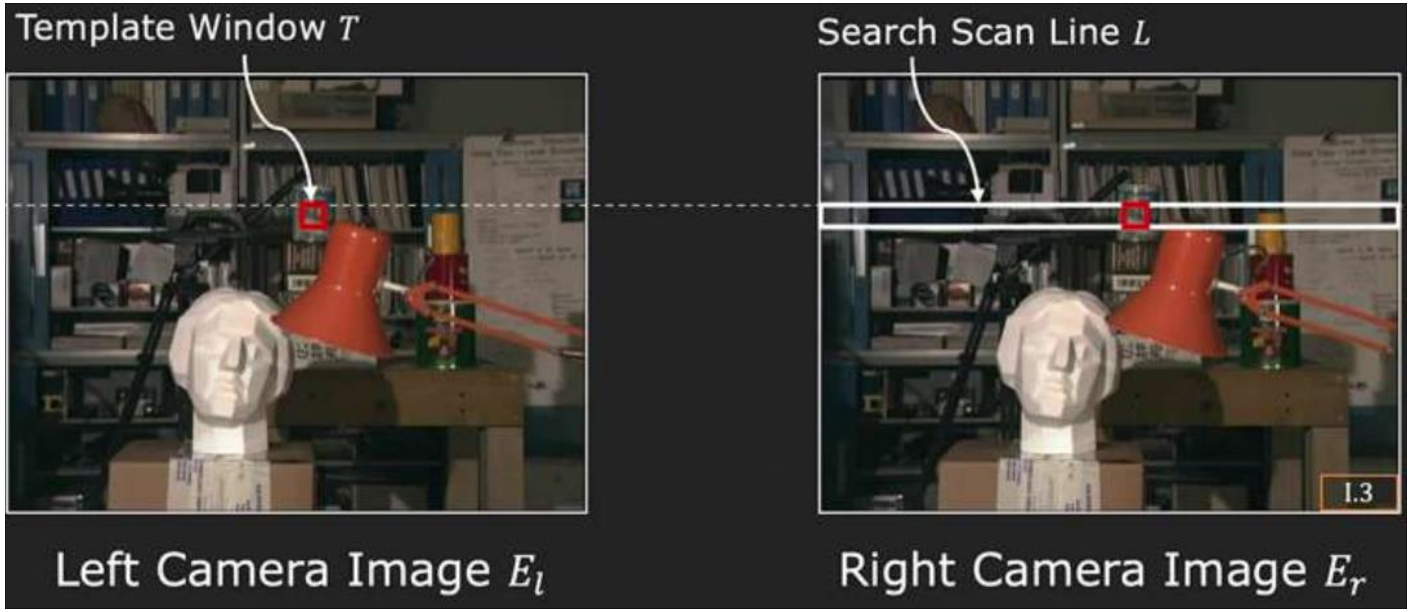


From perspective projection: $y'_l = y'_r = f\frac{x}{z} + C_y$

Corresponding scene point lies on the same horizontal scan line

Stereo Matching: Finding Disparities

- Windows Based Methods
  - Determine disparity using Templet Matching



Disparity : $\mathrm{d} = (x'_l - x'_r)$

Depth: $z = \dfrac{fb}{(x'_l - x'_r)}$

Similarity Metrics for Template Matching:

- Similarity Metrics for Template Matching:
  - Find pixel$(k, l) \in L$ with Minimum Sum of Absolute Differences:

$$SAD(k, l) = \sum |E_l(i, j) - E_r\,(i + k, j + 1)|$$

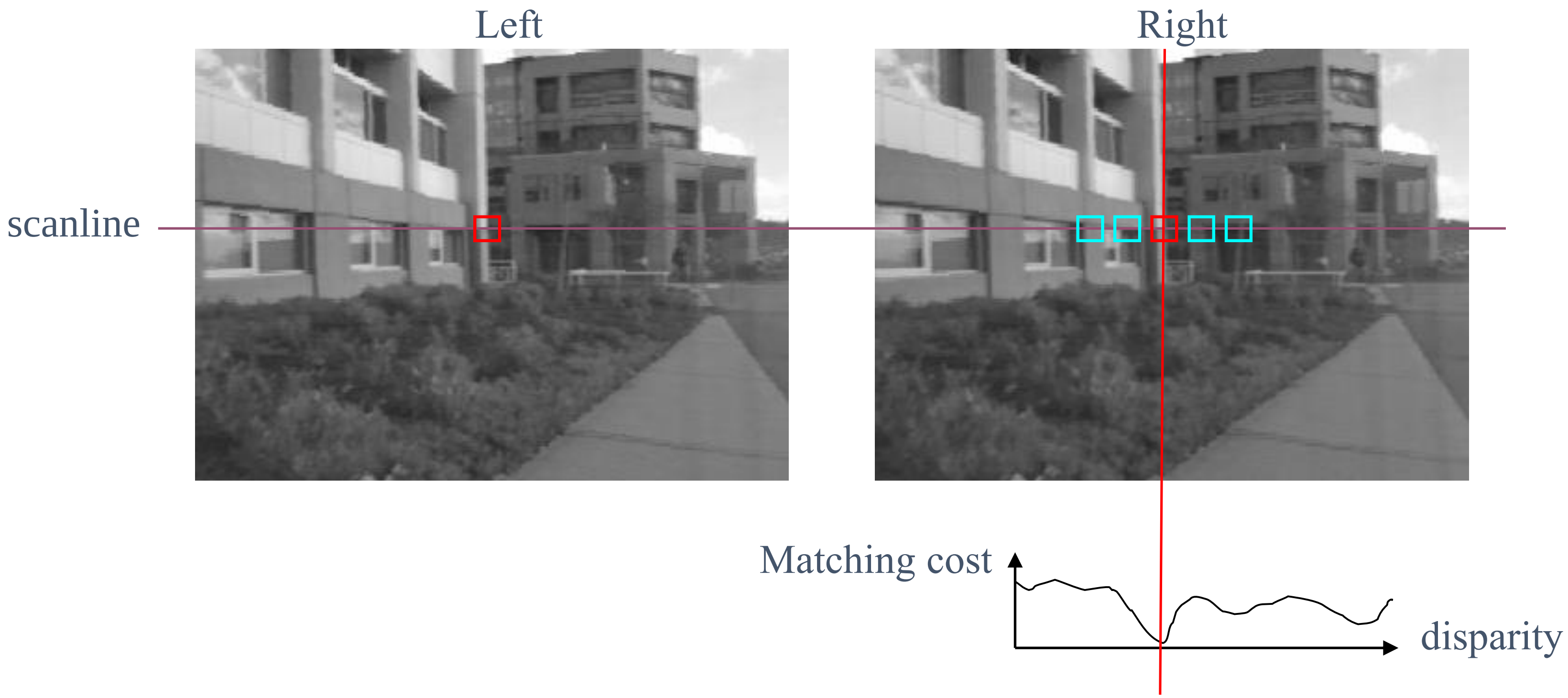
  - Find pixel$(k, l) \in L$ with Minimum Sum of Squared Differences:

$$SSD(k, l) = \sum_{(i,j)\in T} |E_l(i, j) - E_r\,(i + k, j + 1)|^2$$

  - Find pixel$(k, l) \in L$ with Maximum of Normalized Cross-Correlation

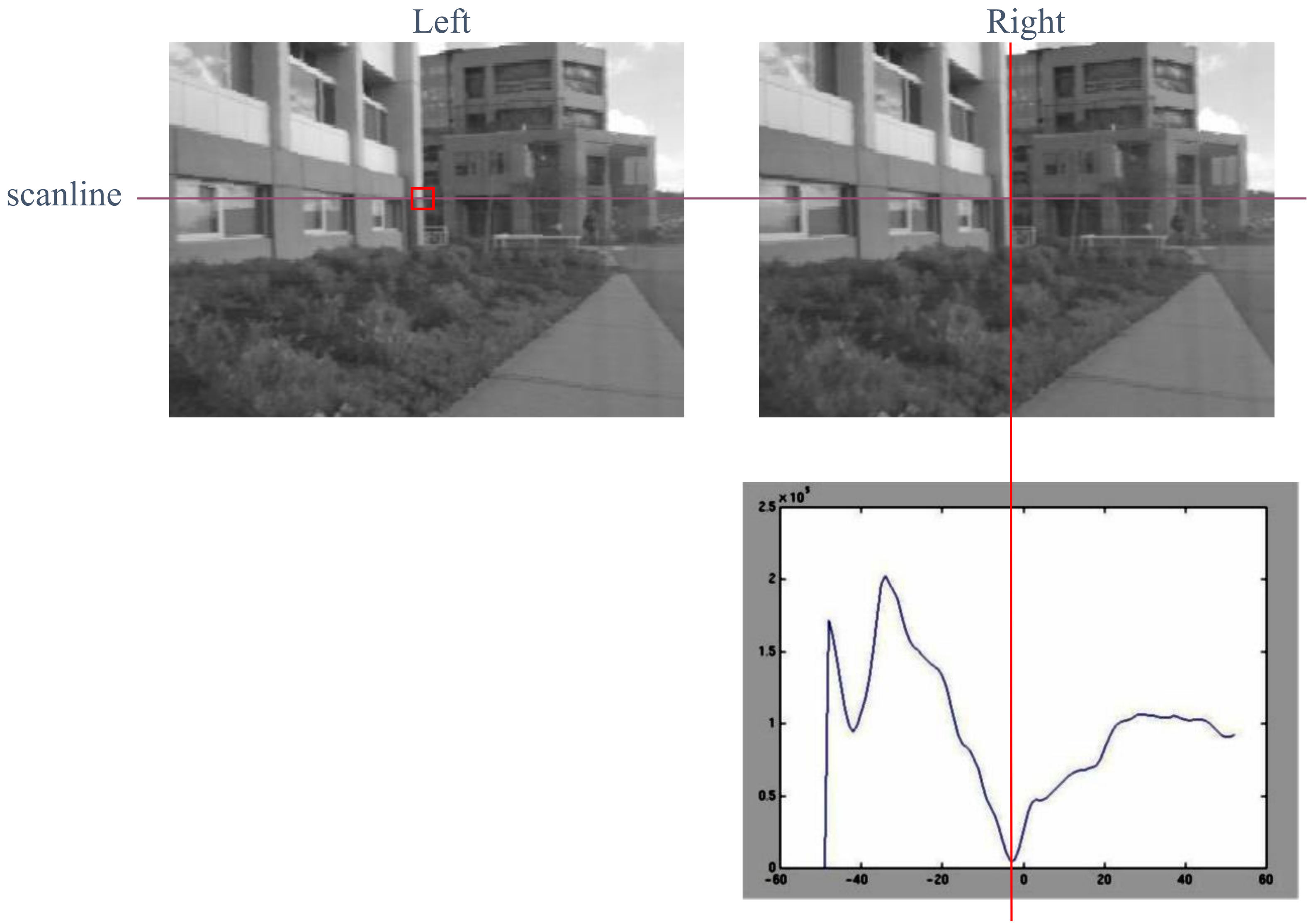
$$\text{NCC}(k, l) = \frac{\sum |E_l(i,j) - E_r\,(i+k, j+1)|}{\sqrt{\sum_{(i,j)\in T} E_l(i,j)^2 \;\; \sum_{(i,j)\in T} E_r\,(i+k, j+1)^2}}$$

Correspondence search



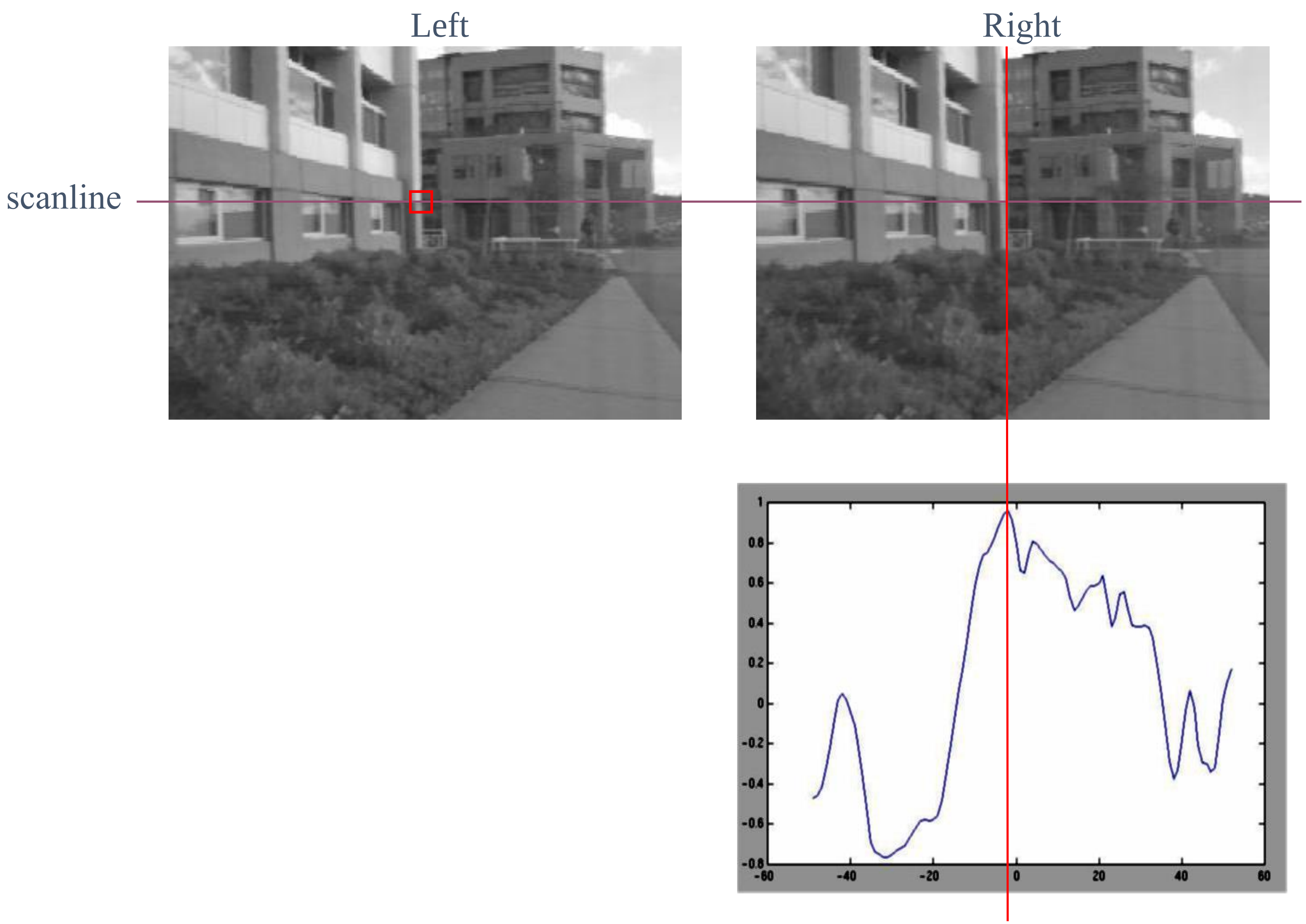
- Slide a window along the right scanline and compare contents of that window with the reference window in the left image
- Matching cost: SSD, SAD, or normalized correlation

Correspondence search



SSD

Correspondence search



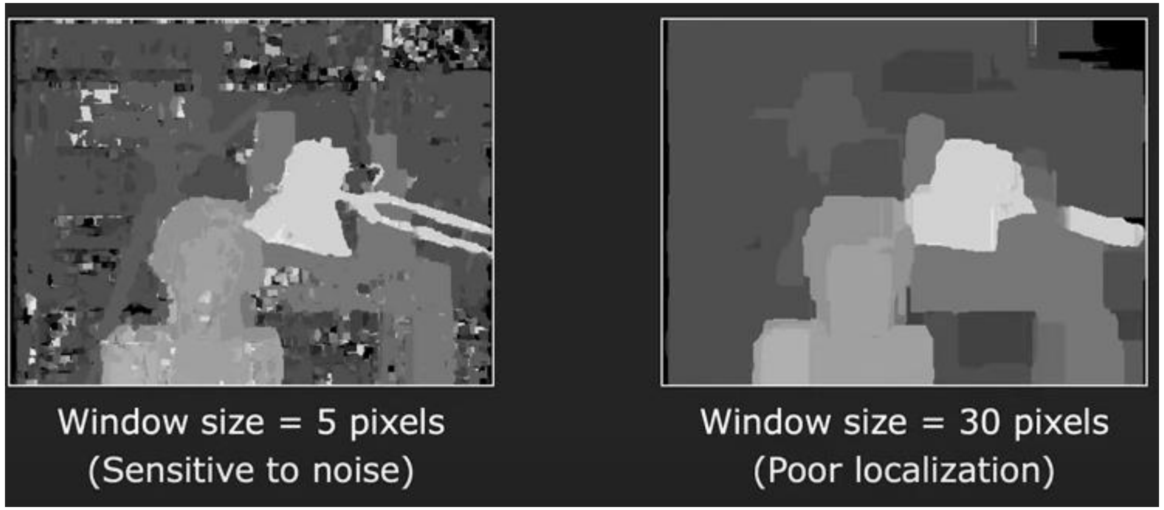
Norm. corr

Basic Stereo Matching Algorithm/Compute depth map

1. **Rectify** the stereo images to align epipolar lines. (not required for basic stereo)
2. For each pixel in the left image:
   - **Find the corresponding pixel** in the right image along the scanline.
   - **Compute disparity** $d = x - x'$.
3. **Triangulate** to compute depth $z = \frac{f.B}{d}$
4. **Create a depth map** by storing depth values for all pixels.

How Large Should Window be?

- The Smaller the window the less descriptive the pattern is.
  - leads to noisy disparity map



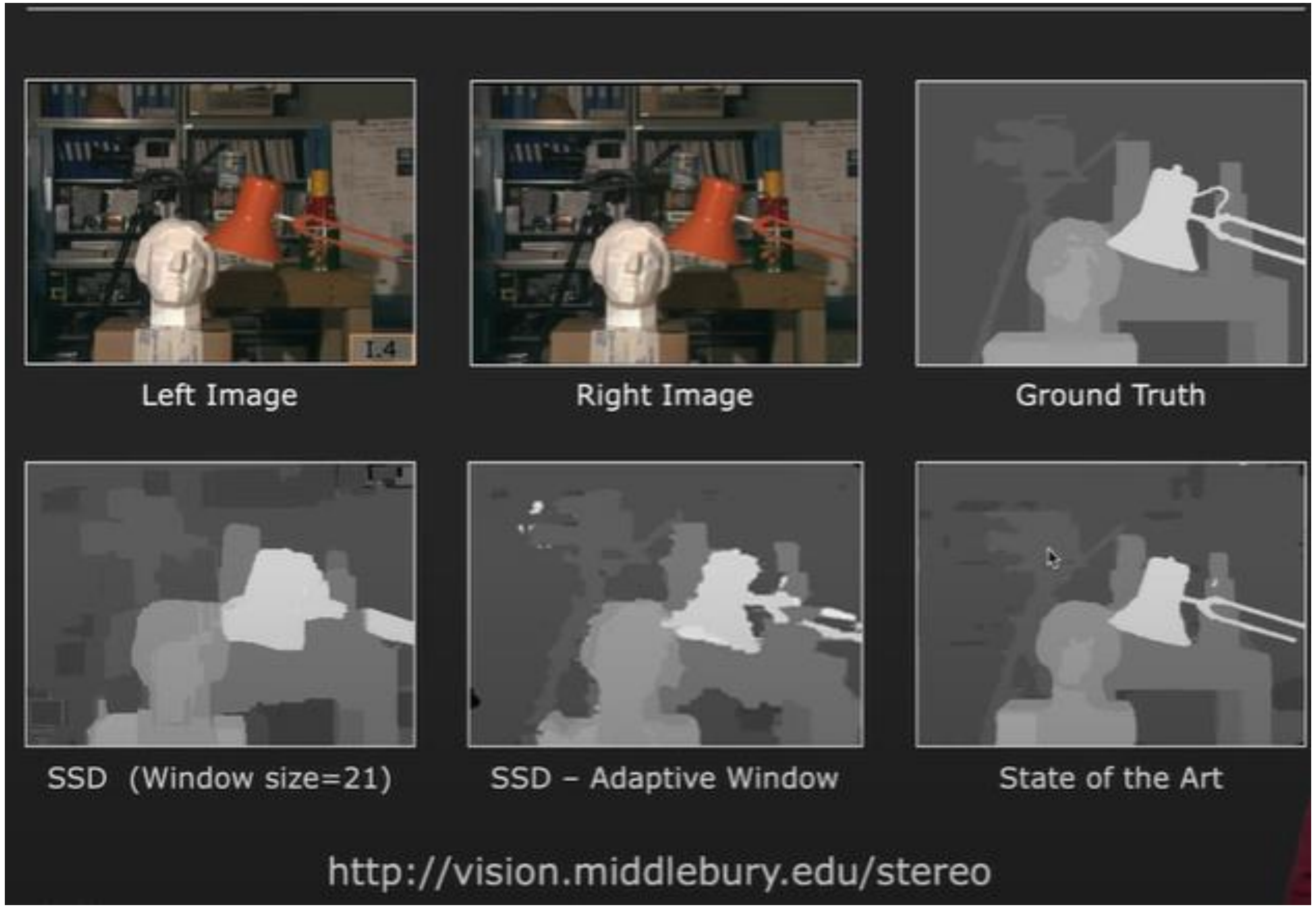
- You will get more robust must in terms of the depth values but the disparity map is more blurred.
  - Poor localizations

Solution is:

Multiple window sizes called the Adaptive window Method solution:

- For each point, match using windows of multiple size and use the disparity that is a result of the best similarity measure (minimize the SSD per pixel)

Window Based Methods: Results

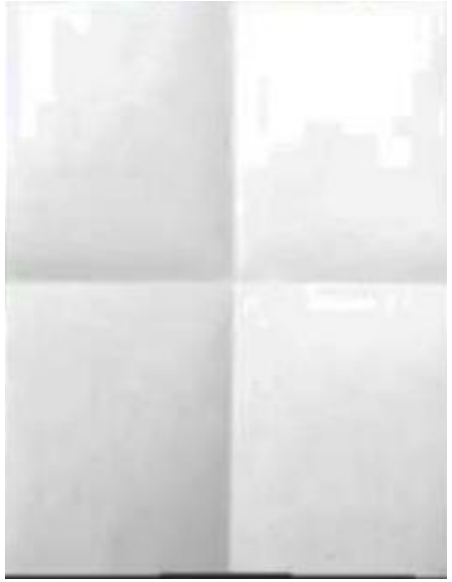


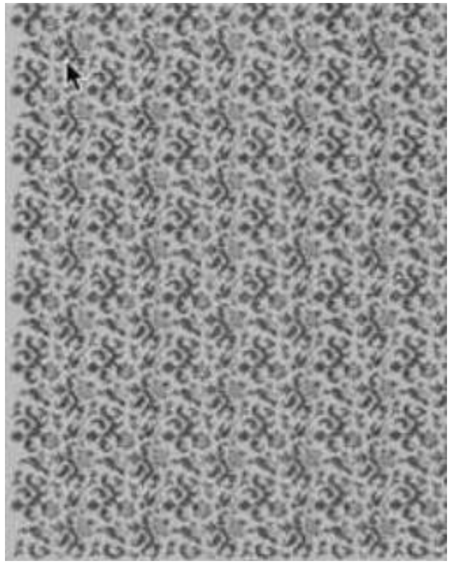
Applications of Stereo Vision

- Autonomous vehicle:
  - https://www.youtube.com/watch?v=XOt2iRUeDag
  - https://www.youtube.com/watch?v=UypJLwgsPxk&t=12s
- Robotics: https://www.youtube.com/watch?v=WSDU8giz6ik
- AR/VR: https://www.youtube.com/watch?v=ZE8FVm_ZIAk

Issue with Stereo Matching/ What will cause errors?

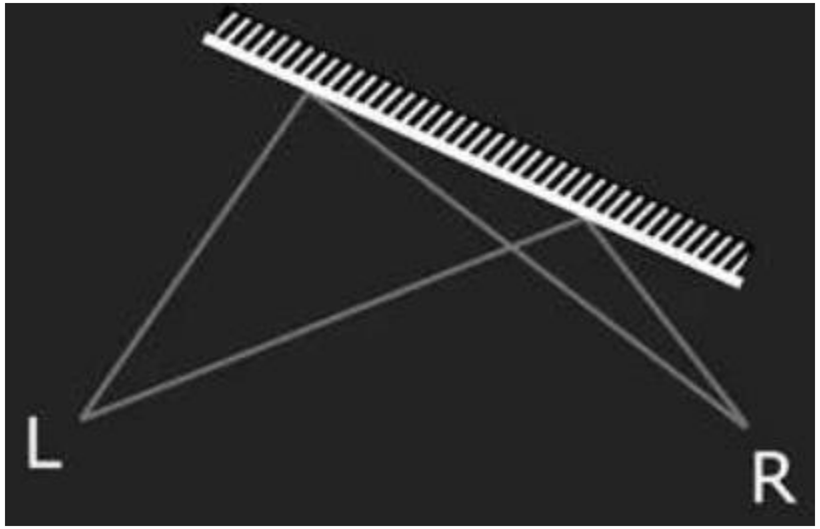
- Chose the image that have texture and non repetitive texture





Issue with Stereo Matching/ What will cause errors?

- Foreshortening effect makes matching challenging



Issue with Stereo Matching/ What will cause errors?

- Violations of brightness constancy (specular reflections)

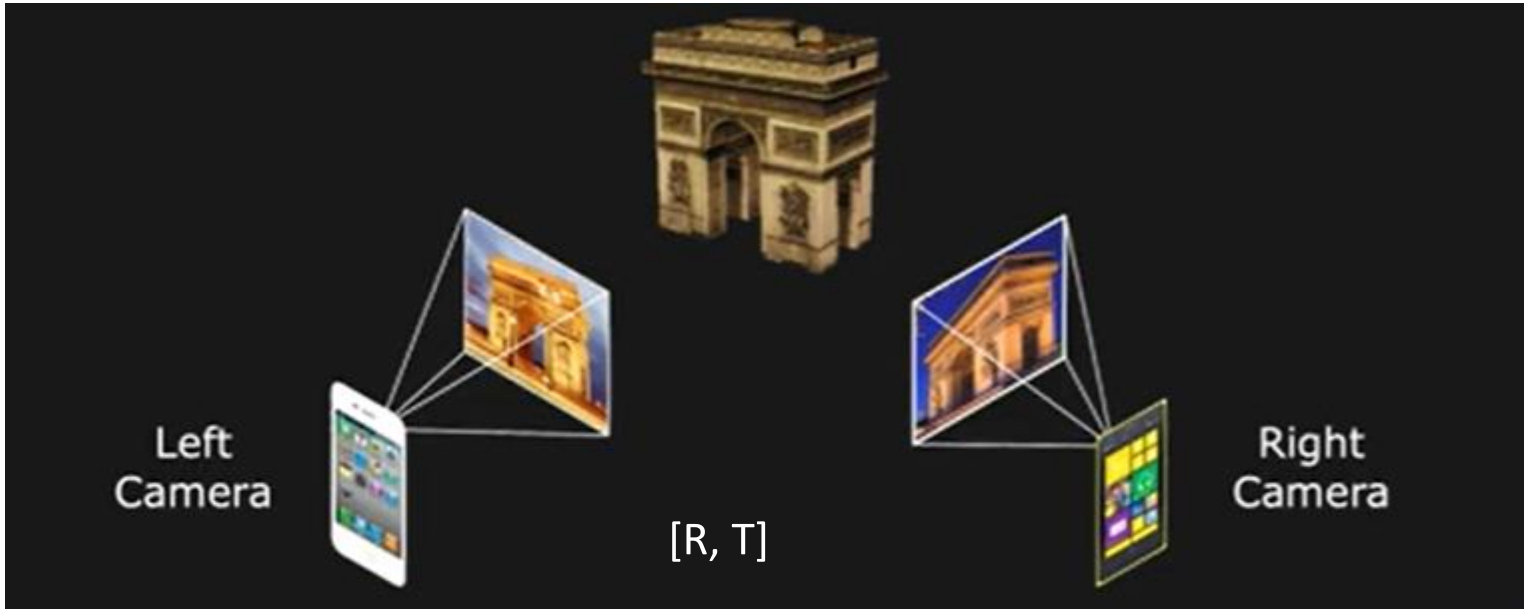
Issue with Stereo Matching/ What will cause errors?

- Camera calibration errors
- Poor image resolution
- Occlusions
- Large motions
- Low-contrast image regions

Next: Epipolar geometry of the stereo system

Compute 3D structure from two arbitrary views

- Compute 3D structure of static scene from two arbitrary views

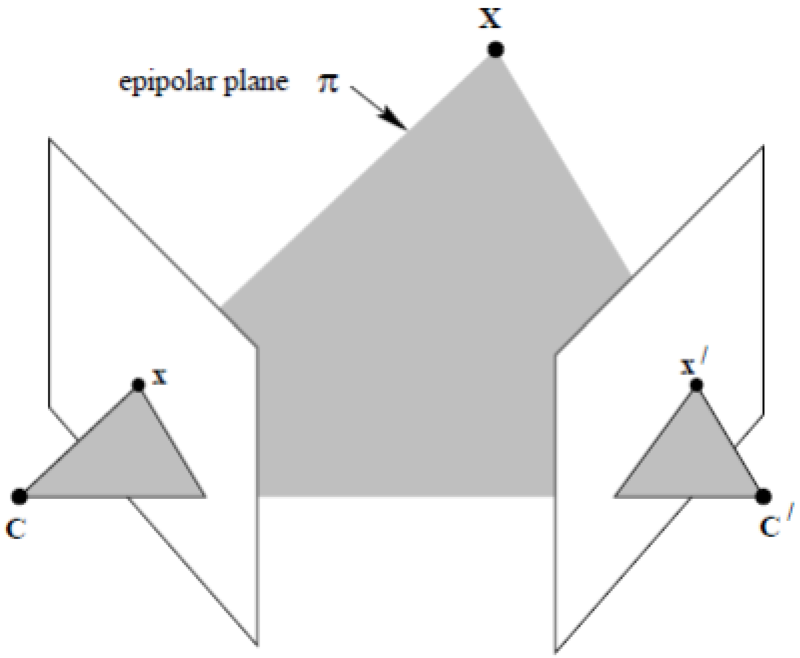


- Calibrated case: Intrinsic (f_x, f_y, c_x, c_y, θ) and Extrinsic (relative position/orientation of cameras) are known for both view/cameras?
- Uncalibrated case: Intrinsic known and Extrinsic (relative position/orientation of camera are unknown?

Correspondence Problem

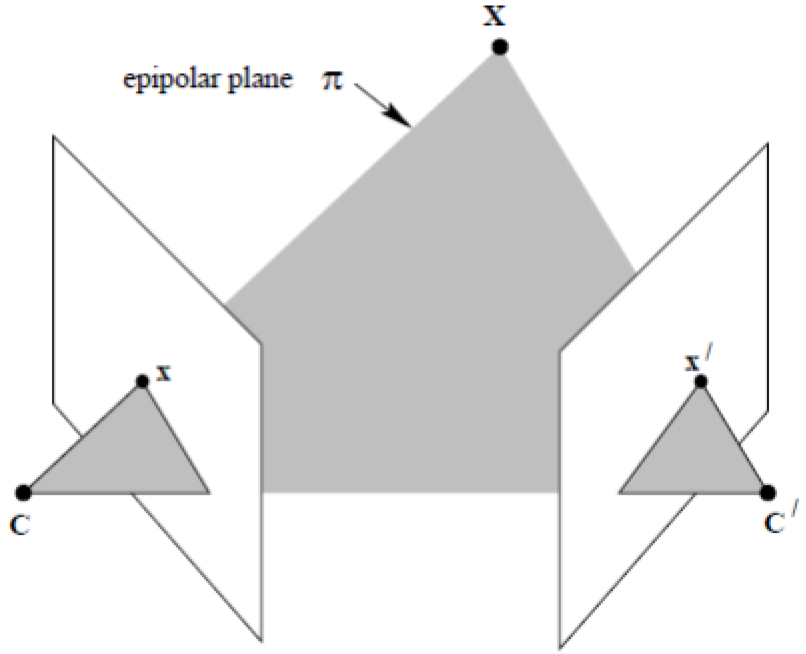
- We have two images taken from cameras at different positions?
- How do we match a point in the first image to a point in the second? How can we constrain our search?

Epipolar geometry



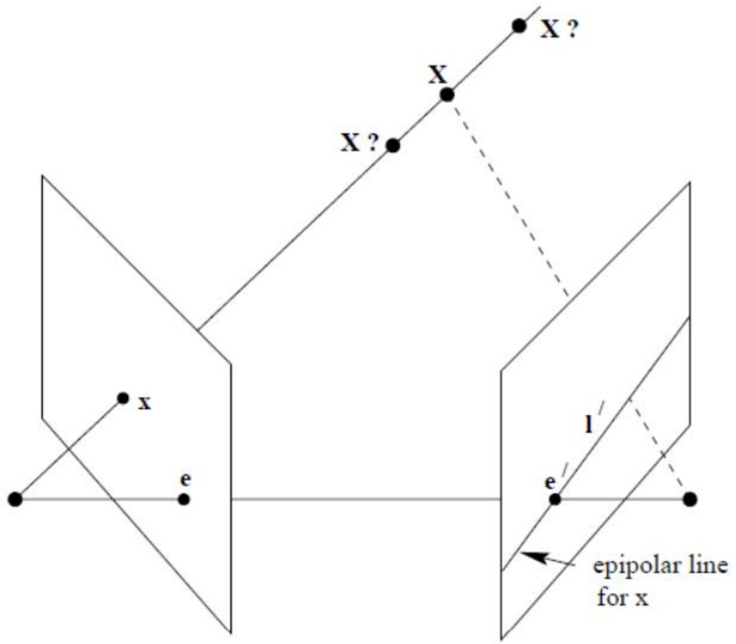
- The point in space X and the two camera centres $C\ and\ C'$ are coplanar.
- We call this plane the epipolar plane π.

Epipolar geometry



- The rays back-projected from $x\ and\ x$' and intersect at X.
- The rays are coplanar, lying in π .
- Suppose that we only know x. Where do we expect to find the corresponding point in x' the second view?

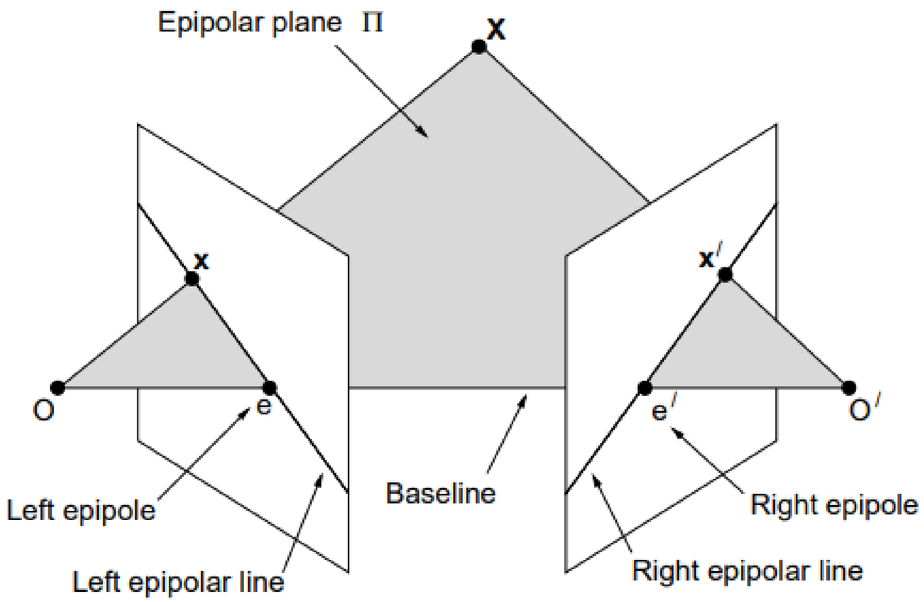
Epipolar geometry



- The ray corresponding to the (unknown) point x' lies in π. Therefore, the point lies on the intersection between the plane and the second image plane. Such intersection is a line called the epipolar line corresponding to x.

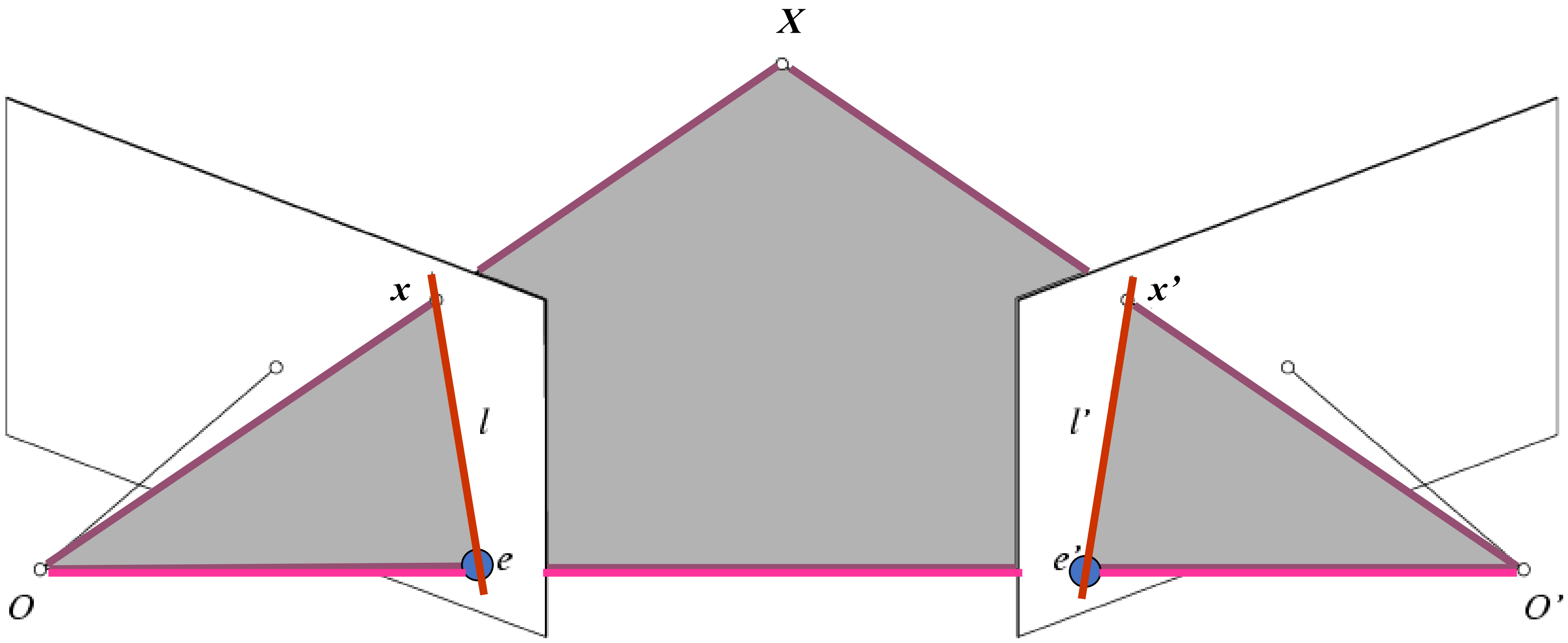
Epipolar Geometry of stereo pair

- The baseline is the line segment joining the two camera centers
- The epipole is the point of intersection of the baseline with the image plane
- An epipolar plane is any plane containing the baseline.
- An epipolar line is the intersection of an epipolar plane with an image plane. All the epipolar lines intersect at the epipole.



Epipolar Geometry: Notation

How to compute: Compute the Epipolar line l?



Calibrated case: Compute the Epipolar line

- Intrinsic ($f_x, f_y, c_x, c_y, \theta$) and Extrinsic (relative position/orientation of cameras) are know for both view/cameras.
- Given a point x on one image, the corresponding epipolar line l’ on the other image can be easily computed using Essential Matrix E.
- Essential Matrix can be obtained through parameters of the two cameras.

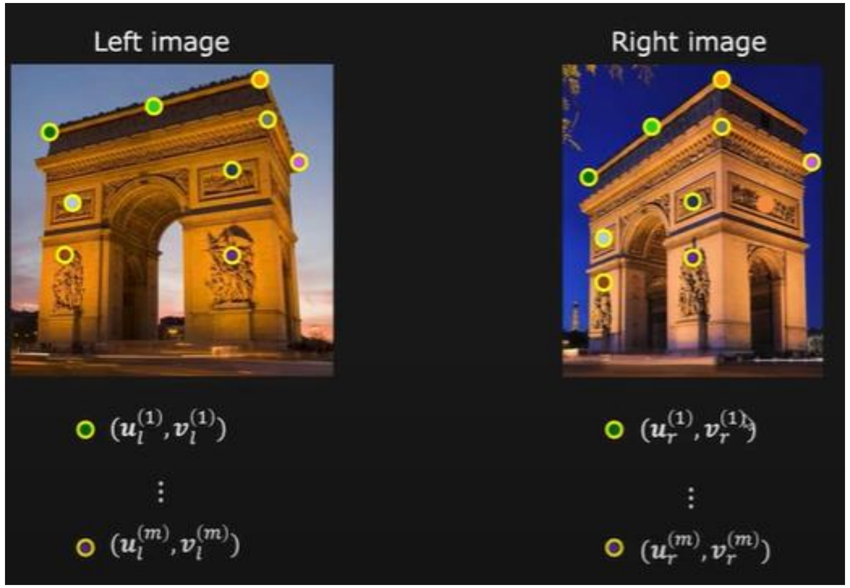
$$E = [T] \times \ R$$

where [T]× is the skew-symmetric

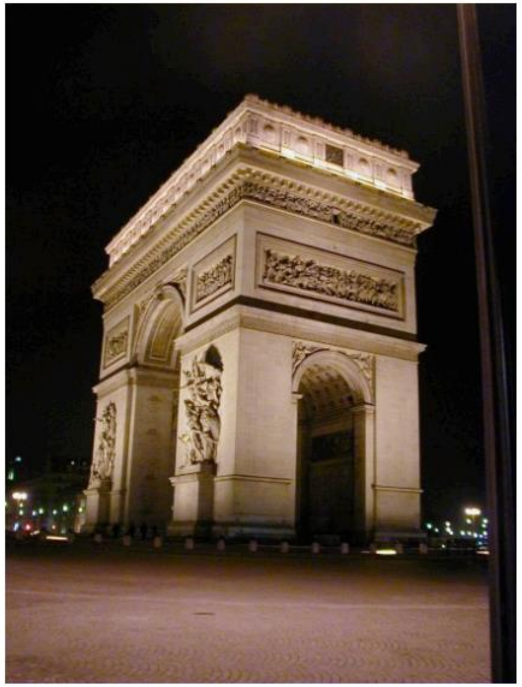
- ***l' = E x***
- Similarly, ***l = E^T x’***

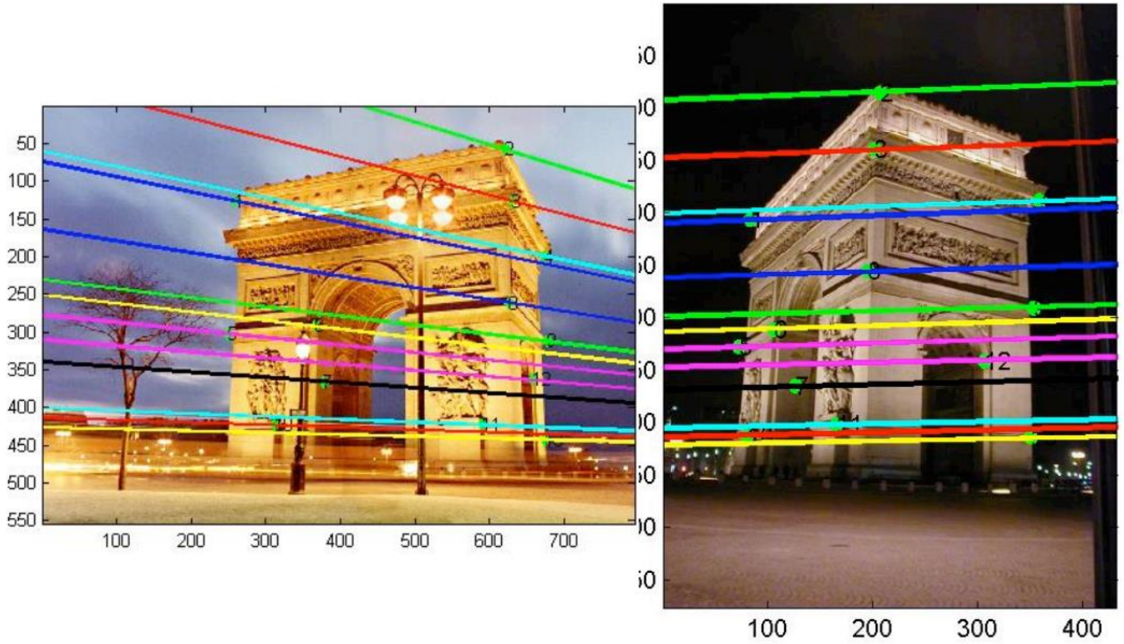
Calibrated case: Compute the Epipolar line

- Uncalibrated case
  - Given a point x on one image, the corresponding epipolar line l’ on the other image can be easily computed using Fundamental Matrix F. => ***l’ = F x***
  - Similarly, $\boldsymbol{l = F^T x'}$
  - Fundamental Matrix can be obtained using 8 point algorithms (need 8 matching points)
    - E.g. using SIFT or hand-picked)

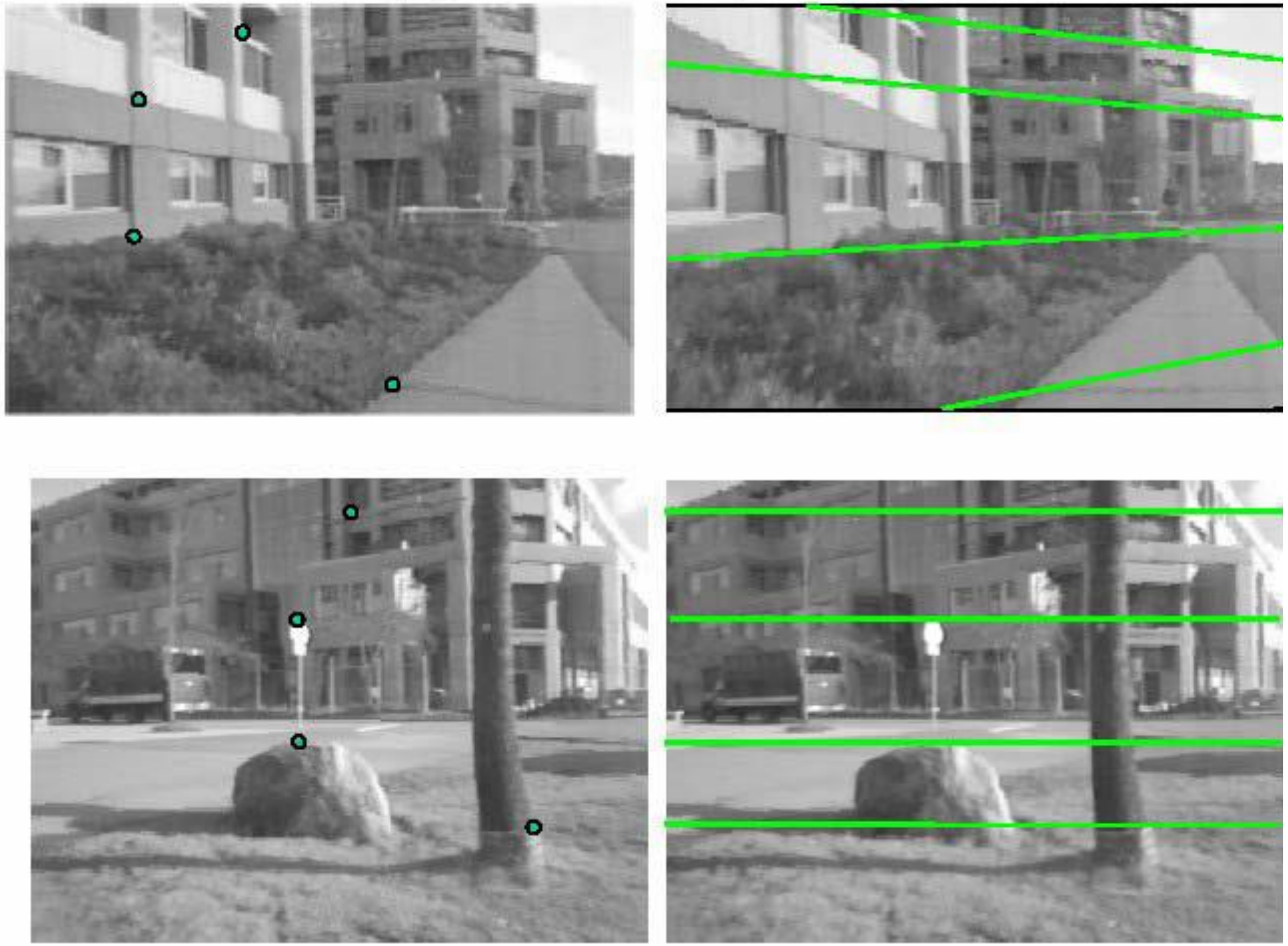


Example

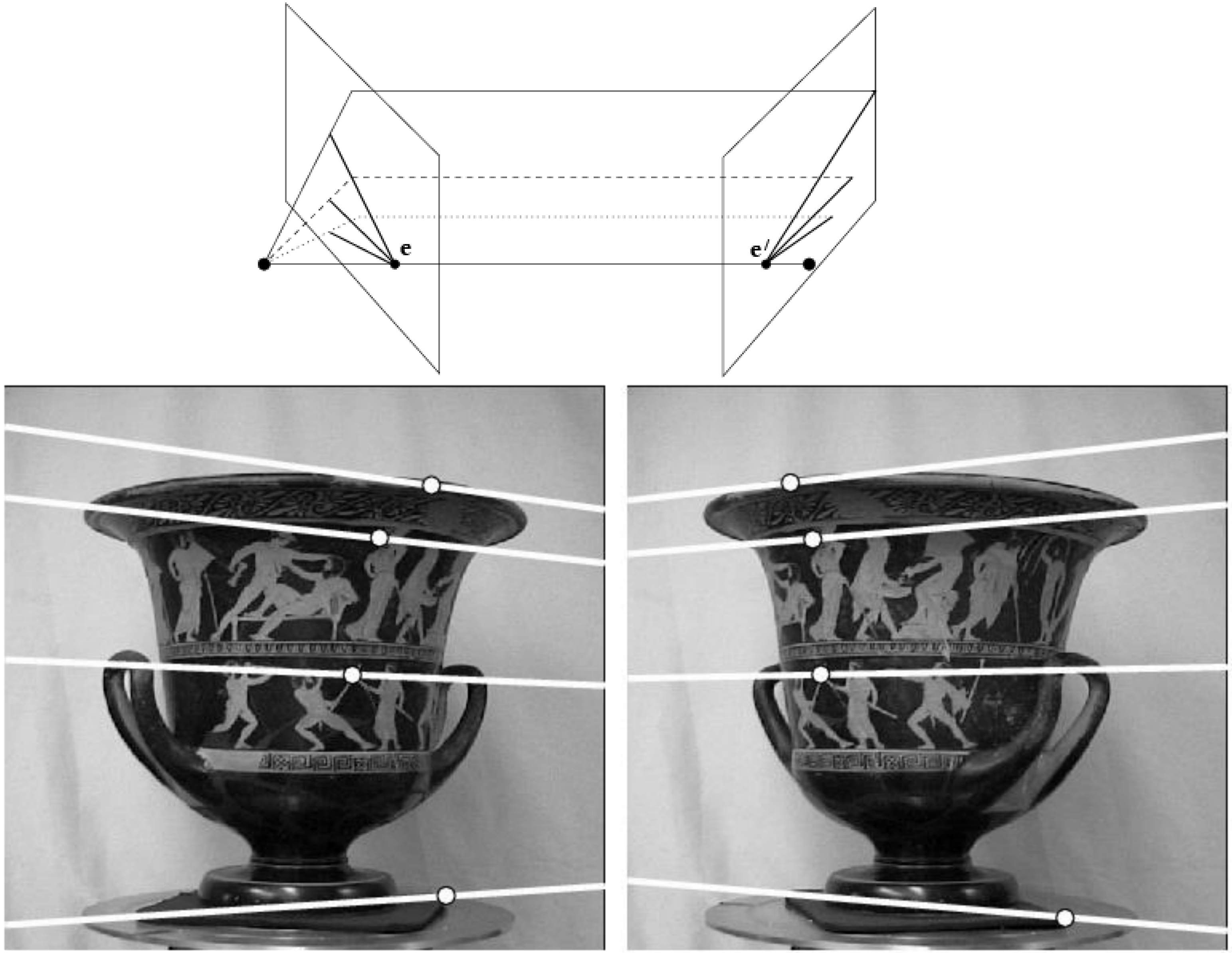



50
100
150
200
250
300
350
400
450
500
550
100
200
300
400
500
600
700
100
200
300
400

Example



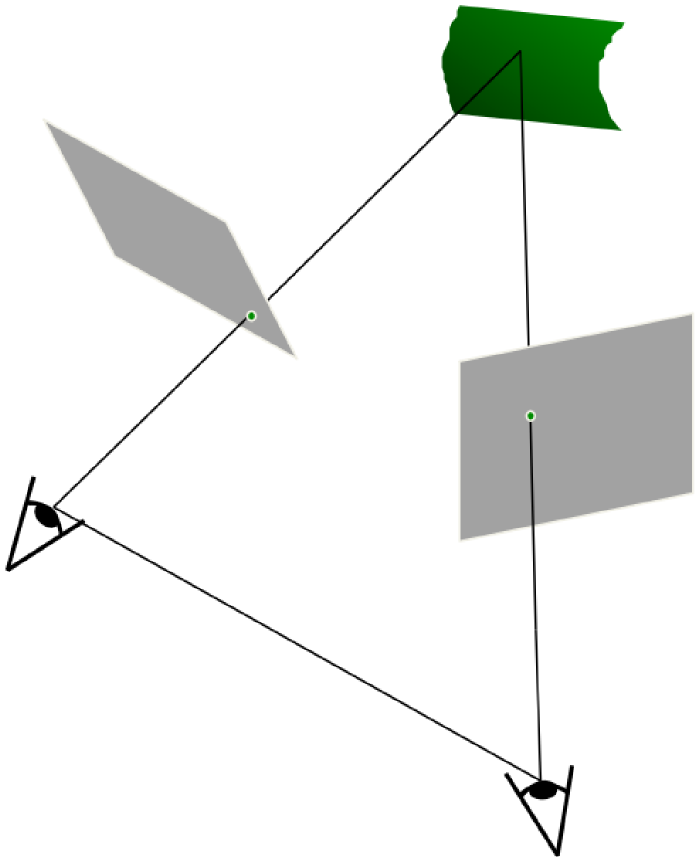
Example: Converging Cameras



Basic Stereo Matching Algorithm

1. **Rectify** the stereo images to align epipolar lines.
2. For each pixel in the left image:
   - **Find the corresponding pixel** in the right image along the scanline.
   - **Compute disparity** $d = x - x'$.
3. **Triangulate** to compute depth $z = \frac{f.B}{d}$
4. **Create a depth map** by storing depth values for all pixels.

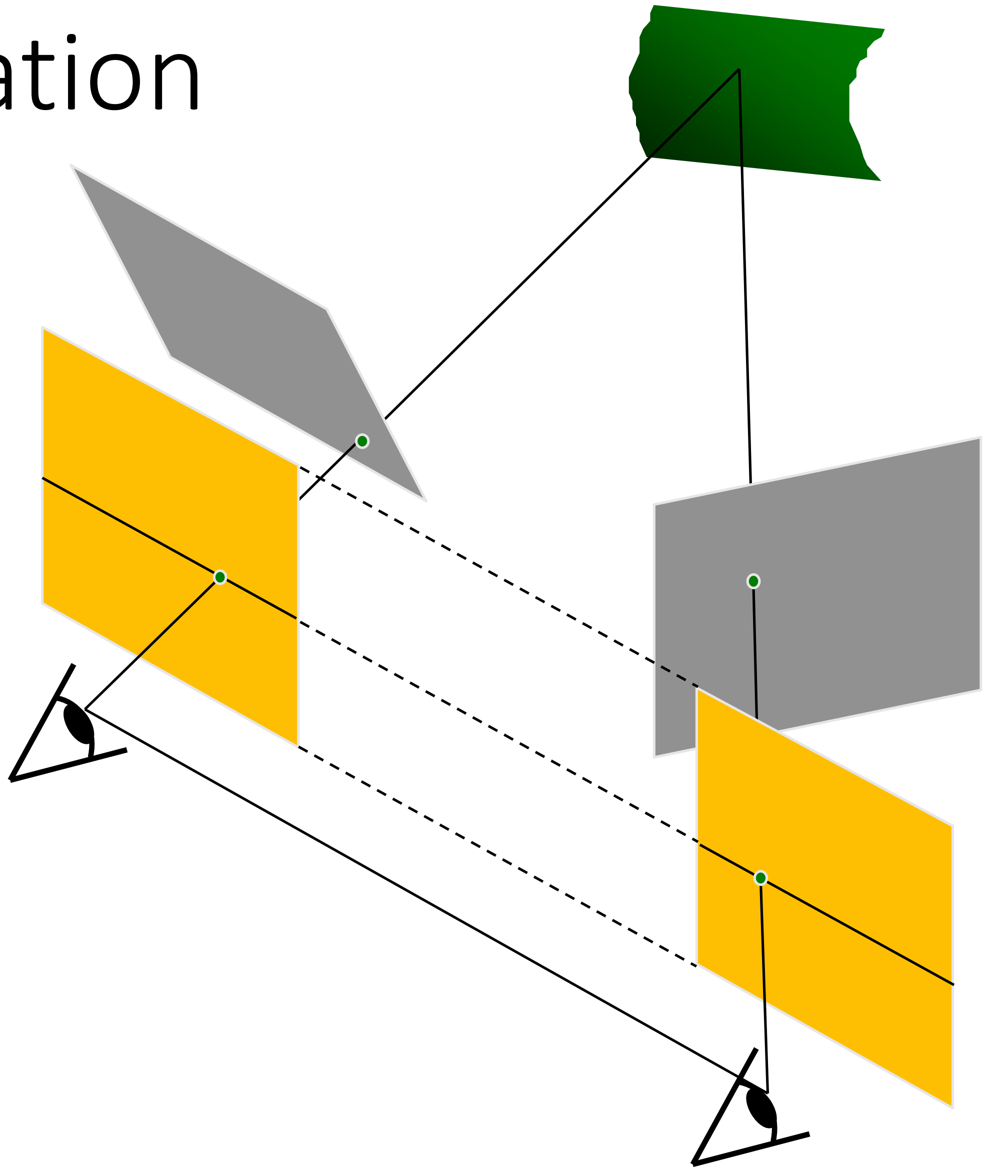
Stereo Image rectification



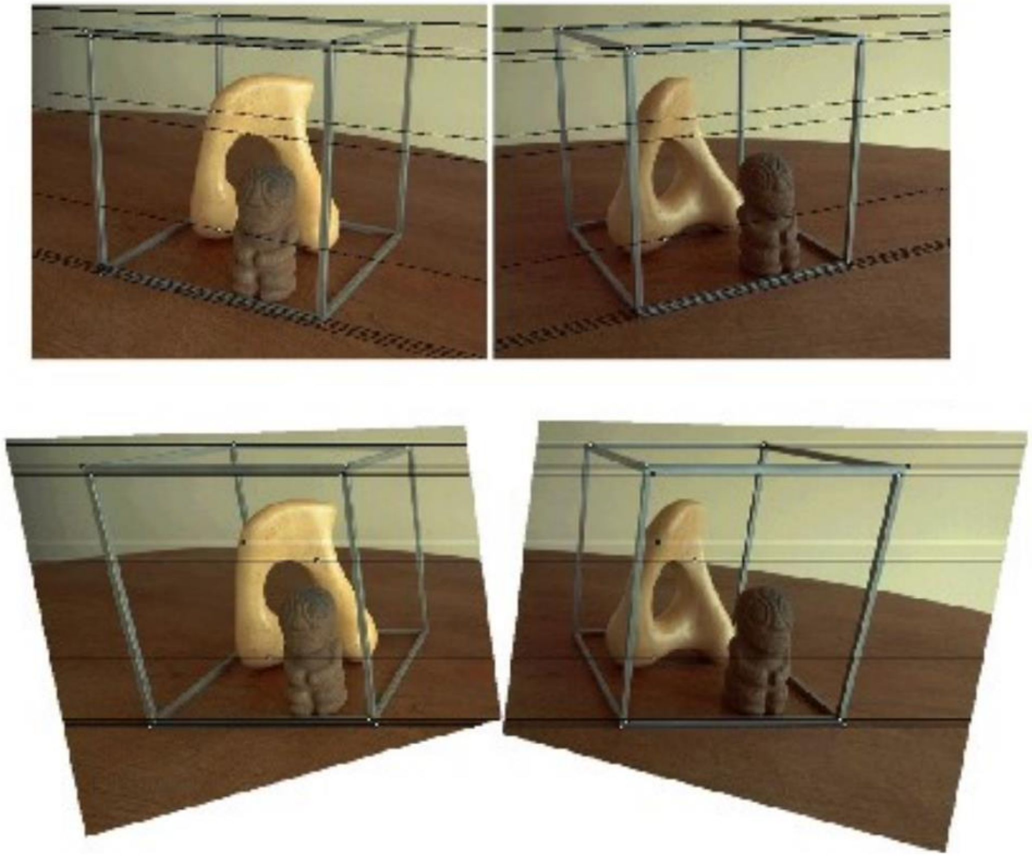
Stereo Image Rectification

- Reproject image planes onto a common plane parallel to the line between camera centers.
- Pixel motion is horizontal after this transformation.

➢ C. Loop and Z. Zhang. [Computing Rectifying Homographies for Stereo Vision](). IEEE Conf. Computer Vision and Pattern Recognition, 1999.



Stereo image rectification



Let’s consider N eyes (Multi View Stereo)

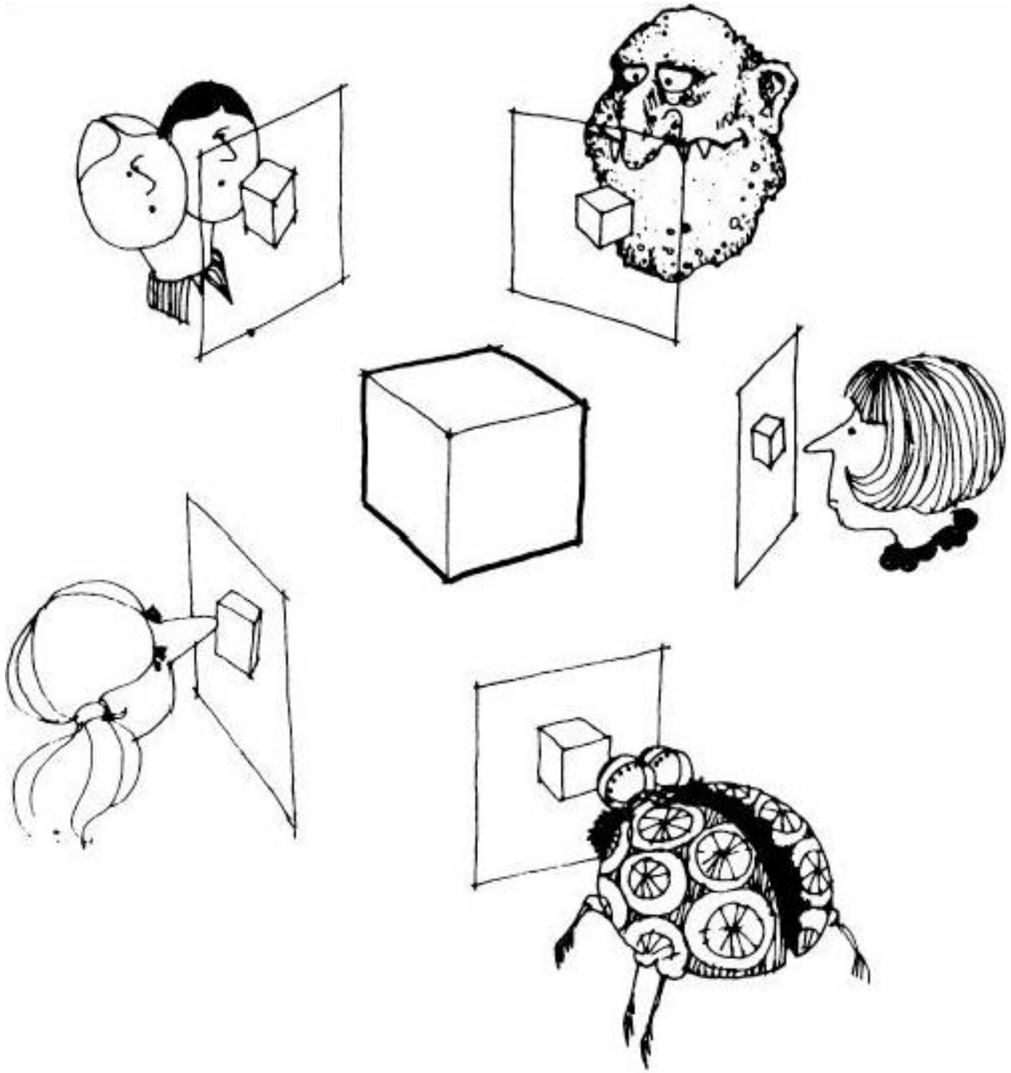
One camera

Two cameras

N cameras

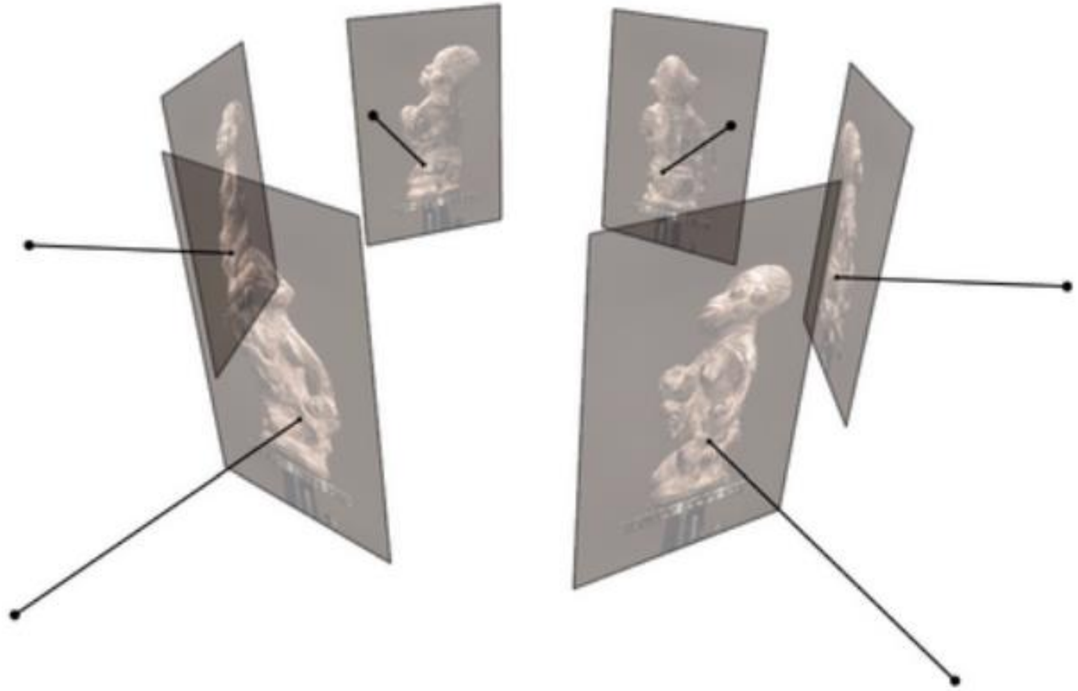
Multi View Stereo

Multi-View Stereo ?

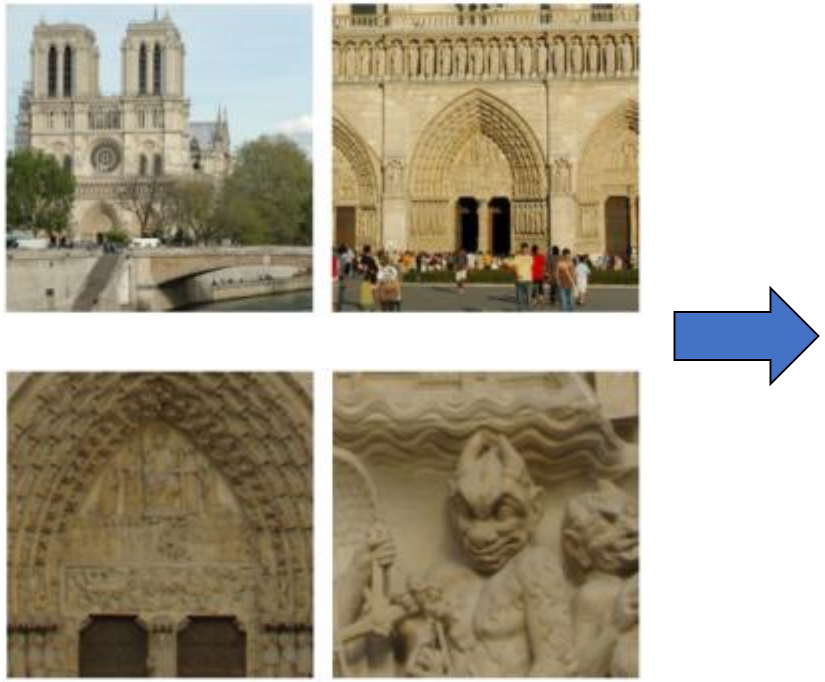


Multi-view stereo

- Goal: given several images of the same object or scene, compute a representation of its 3D shape



Using More Than Two Images



[Multi-View Stereo for Community Photo Collections]
M. Goesele, N. Snavely, B. Curless, H. Hoppe, S. Seitz
Proceedings of [ICCV 2007],

Recap

- Stereo Vision
- Geometry for a simple stereo system
- Epipolar geometry of the stereo system
- Multiview Stereo

Reference

- Szeliski 12.3-12.5

Next: Optical Flow