UNIVERSITÉ Concordia UNIVERSITY
FACULTY OF ENGINEERING AND COMPUTER SCIENCE
Department of Computer Science and Software Engineering

COMP498G/691G COMPUTER VISION

LECTURE 11
STEREO

Administrative

- Course credits resolution
- Tonight's tutorial topics
  - Camera calibration
- Quiz #1
  - deadline: 26th February 2017
  - LATE SUBMISSION POLICY DOES NOT APPLY

Today's Lecture

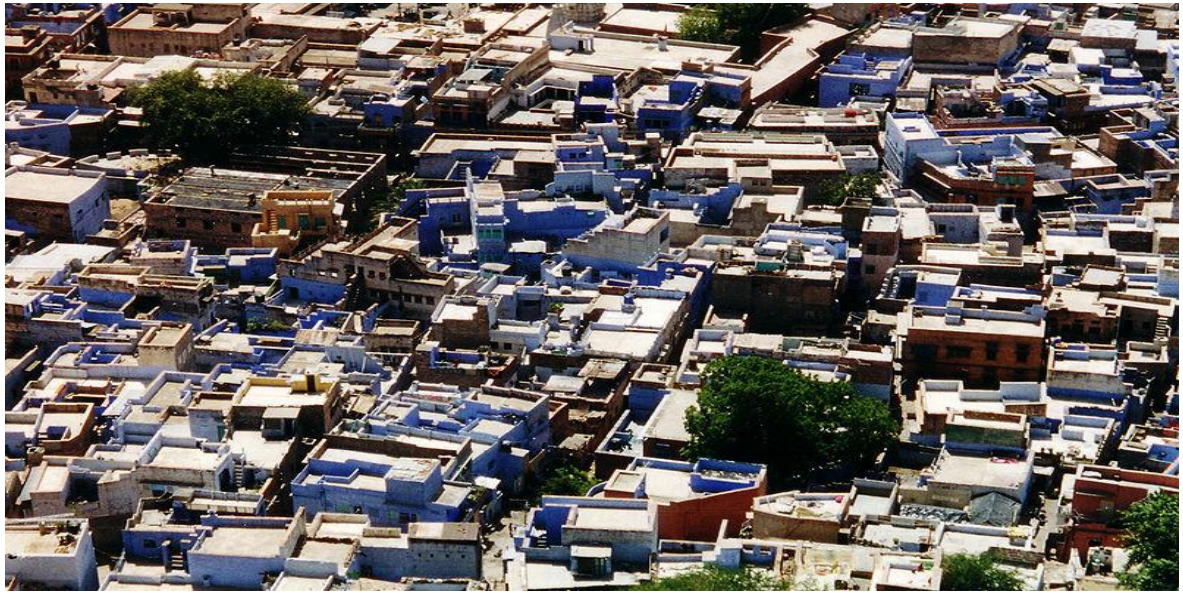
- Introduction to multiple views: Stereo
  - Epipolar Geometry
  - Slides acknowledgment: K. Grauman, J. Hays
- Questions

Miniature faking

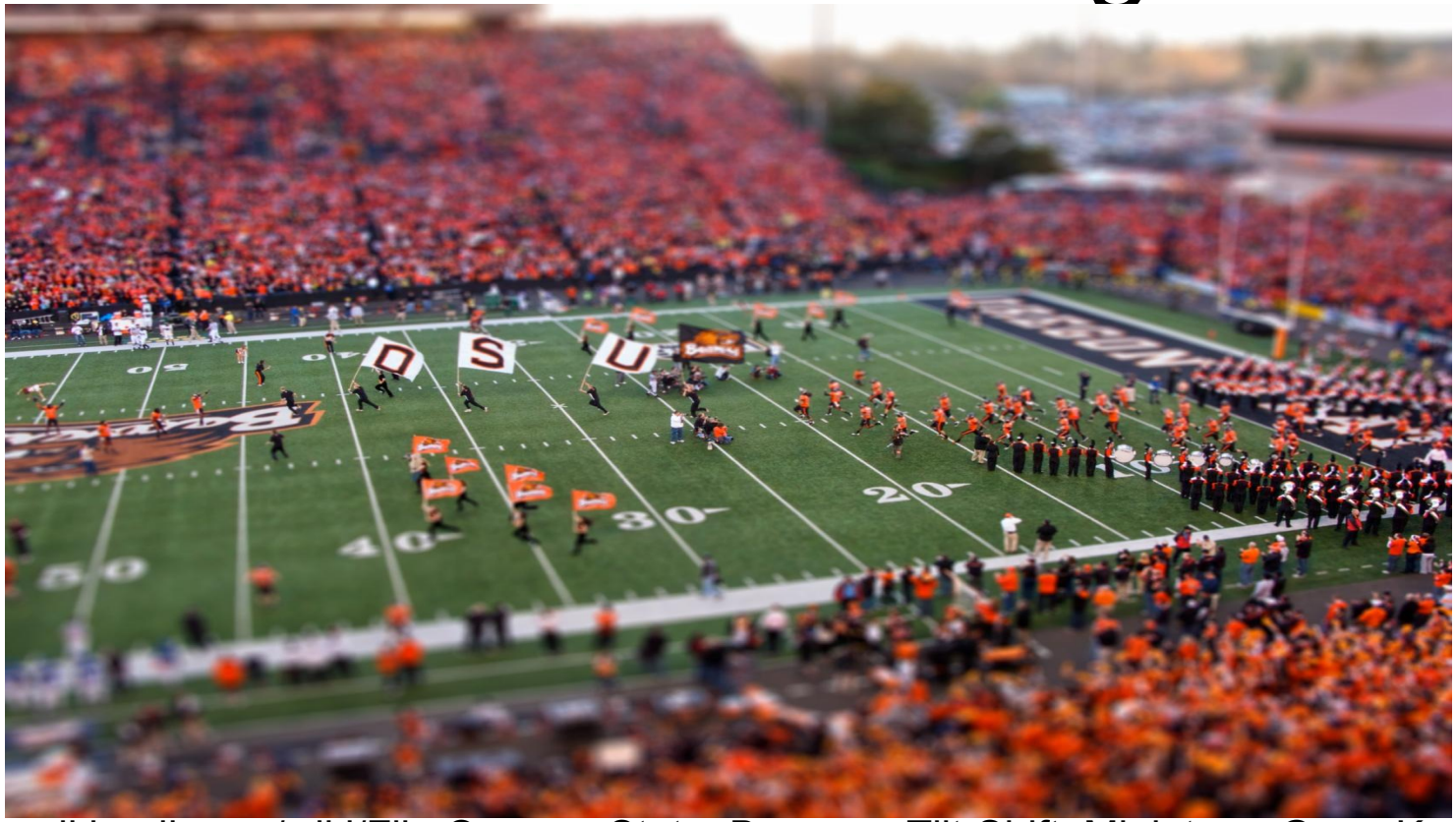
In close-up photo, the depth of field is limited.

http://en.wikipedia.org/wiki/File:Jodhpur_tilt_shift.jpg

Miniature faking

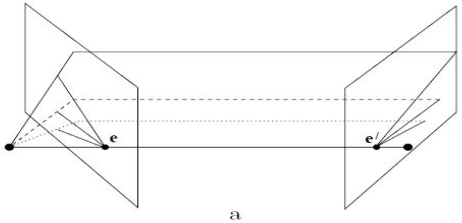


Miniature faking



http://en.wikipedia.org/wiki/File:Oregon_State_Beavers_Tilt-Shift_Miniature_Greg_Keene.jpg

Multiple views



a

Hartley and Zisserman

stereo vision
structure from motion
optical flow

Why multiple views?

- Structure and depth are inherently ambiguous from single views.

Images from Lana Lazebnik

Why multiple views?

- Structure and depth are inherently ambiguous from single views.

Images from Lana Lazebnik

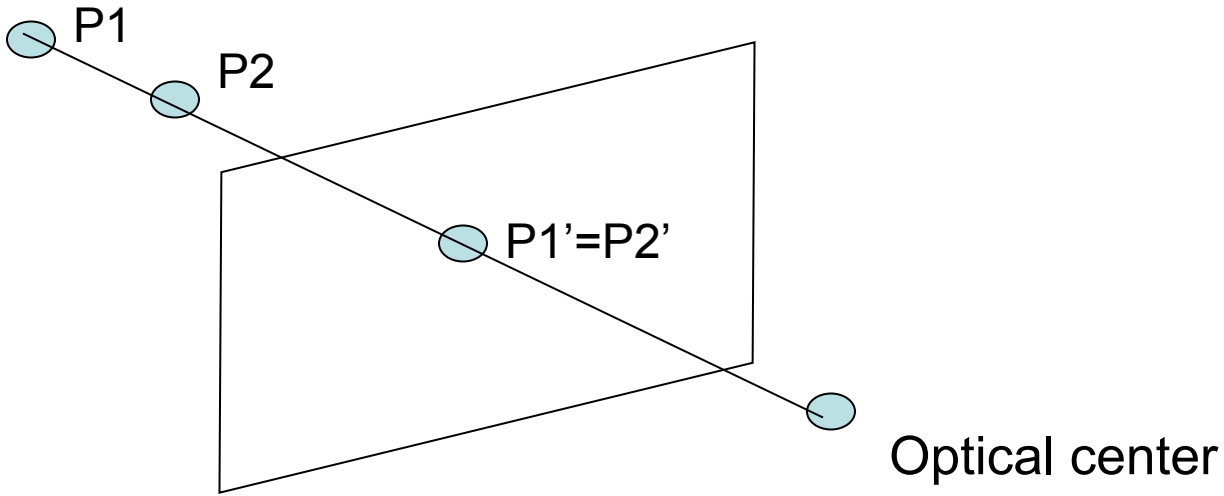
Why multiple views?

- Structure and depth are inherently ambiguous from single views.

Images from Lana Lazebnik

Why multiple views?

- Structure and depth are inherently ambiguous from single views.

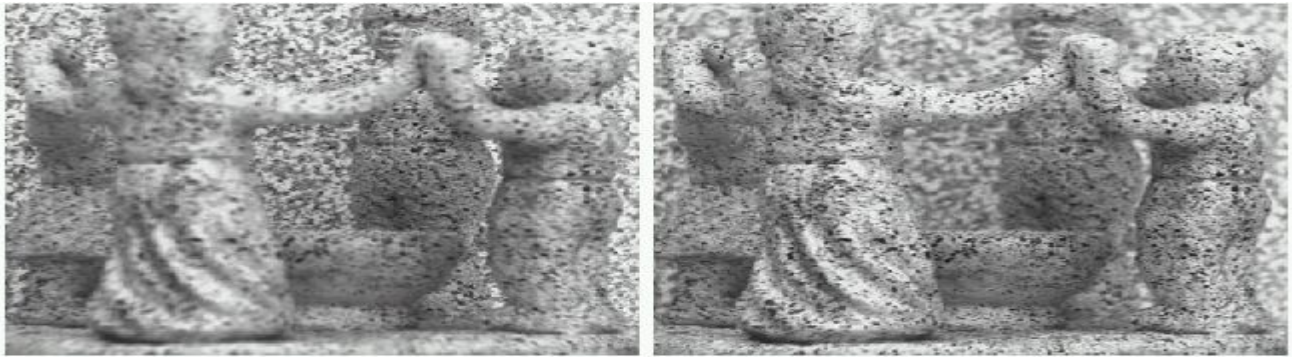


- What cues help us to perceive 3d shape and depth?

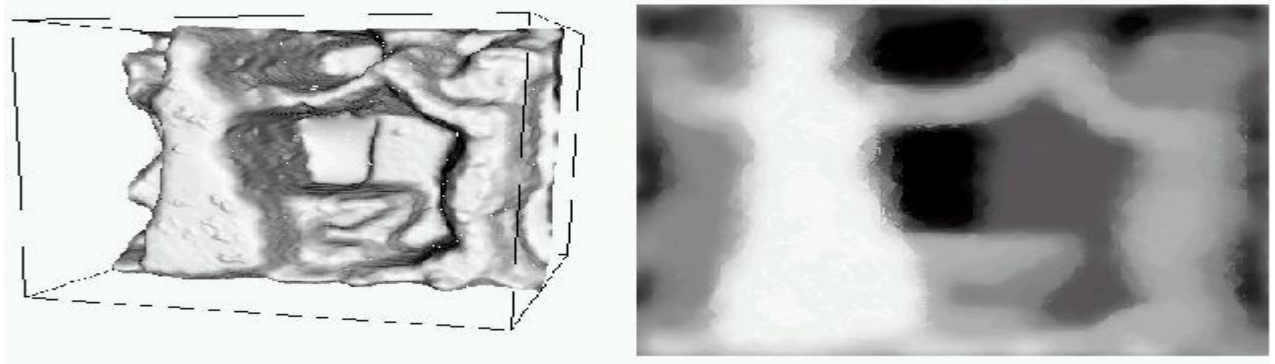
Shading

[Figure from Prados & Faugeras 2006]

Focus/defocus



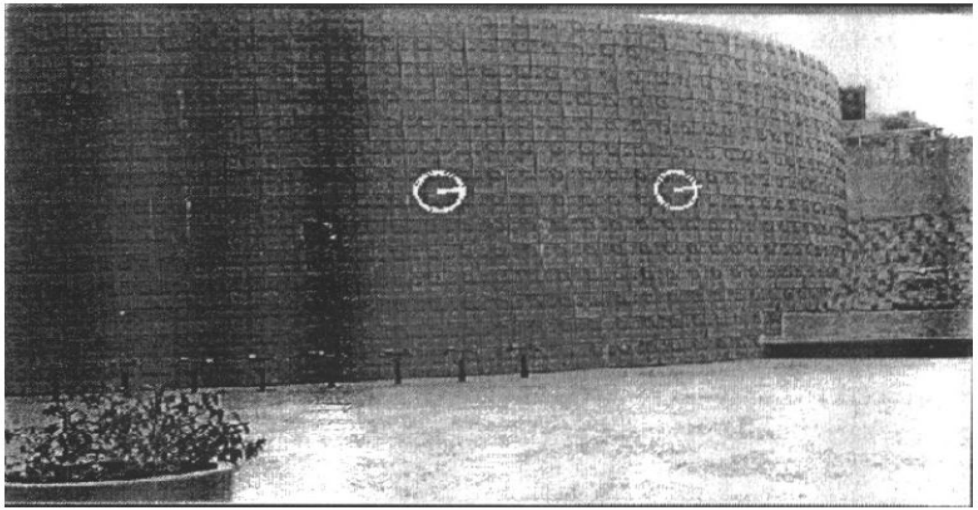
Images from same point of view, different camera parameters

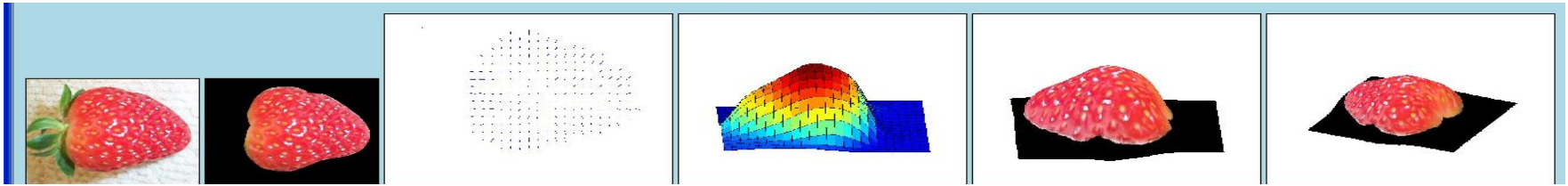


3d shape / depth estimates

[figs from H. Jin and P. Favaro, 2002]

Texture





[From A.M. Loh. The recovery of 3-D structure using visual texture patterns. PhD thesis]

Perspective effects

Image credit: S. Seitz

Motion

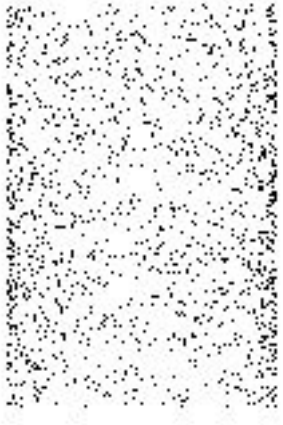


Figures from L. Zhang

http://www.brainconnection.com/teasers/?main=illusion/motion-shape

Motion

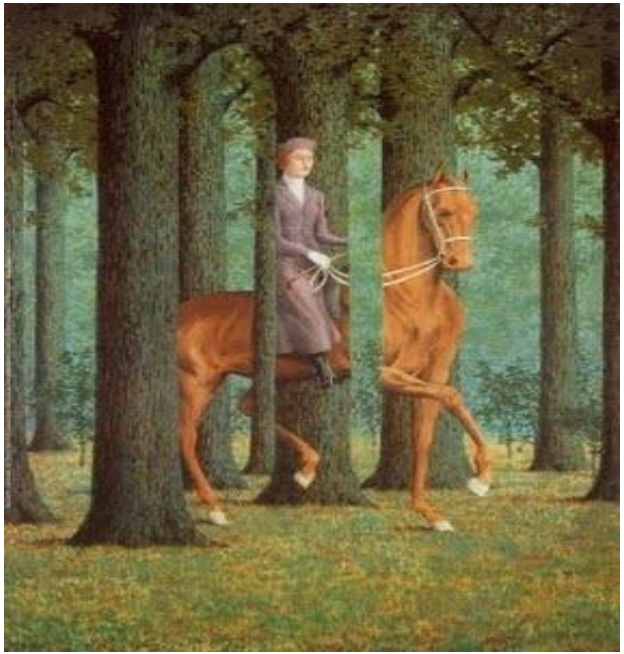


Figures from L. Zhang

http://www.brainconnection.com/teasers/?main=illusion/motion-shape

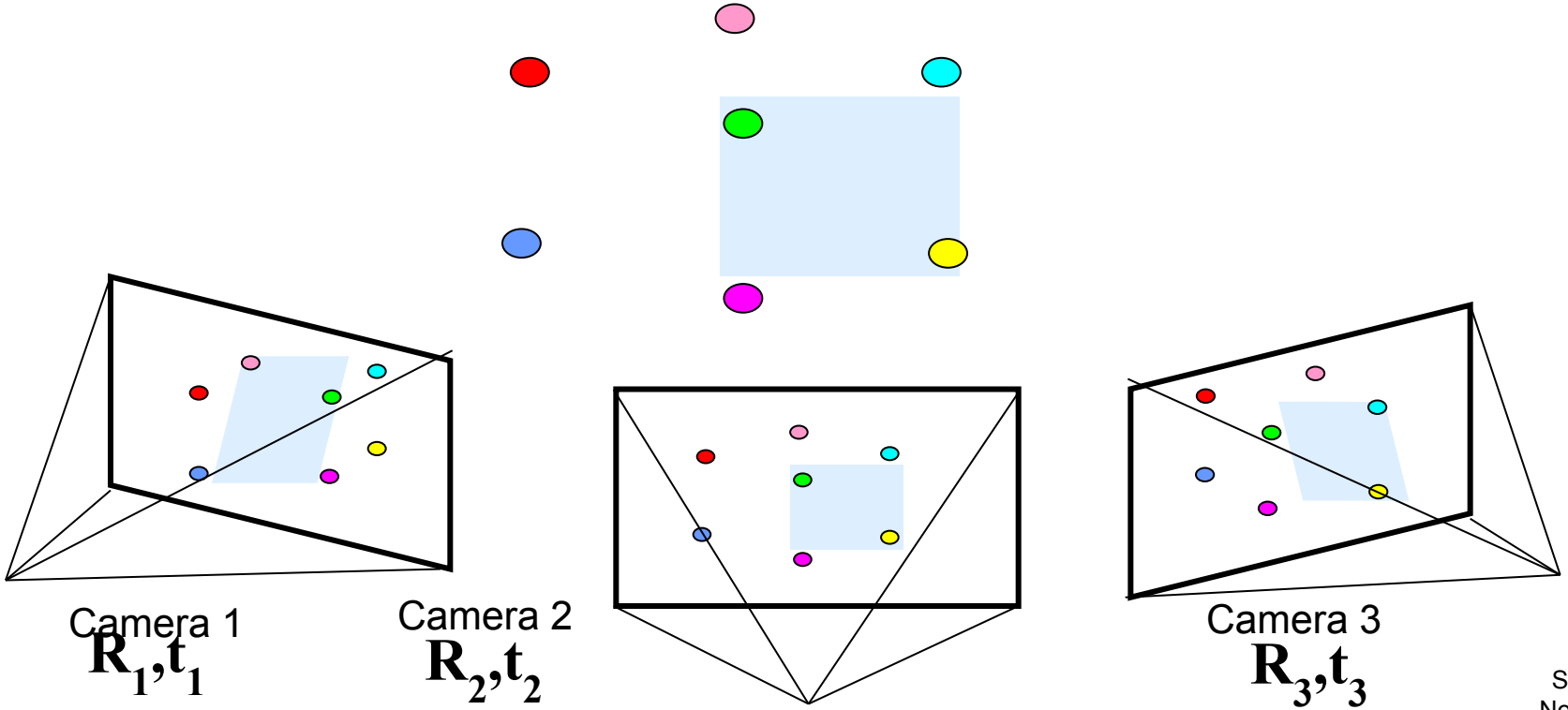
Occlusion



Rene Magritte famous painting *Le Blanc-Seing* (literal translation: "The Blank Signature") roughly translates as "free hand" or "free rein".

If stereo were critical for depth perception, navigation, recognition, etc., then this would be a problem

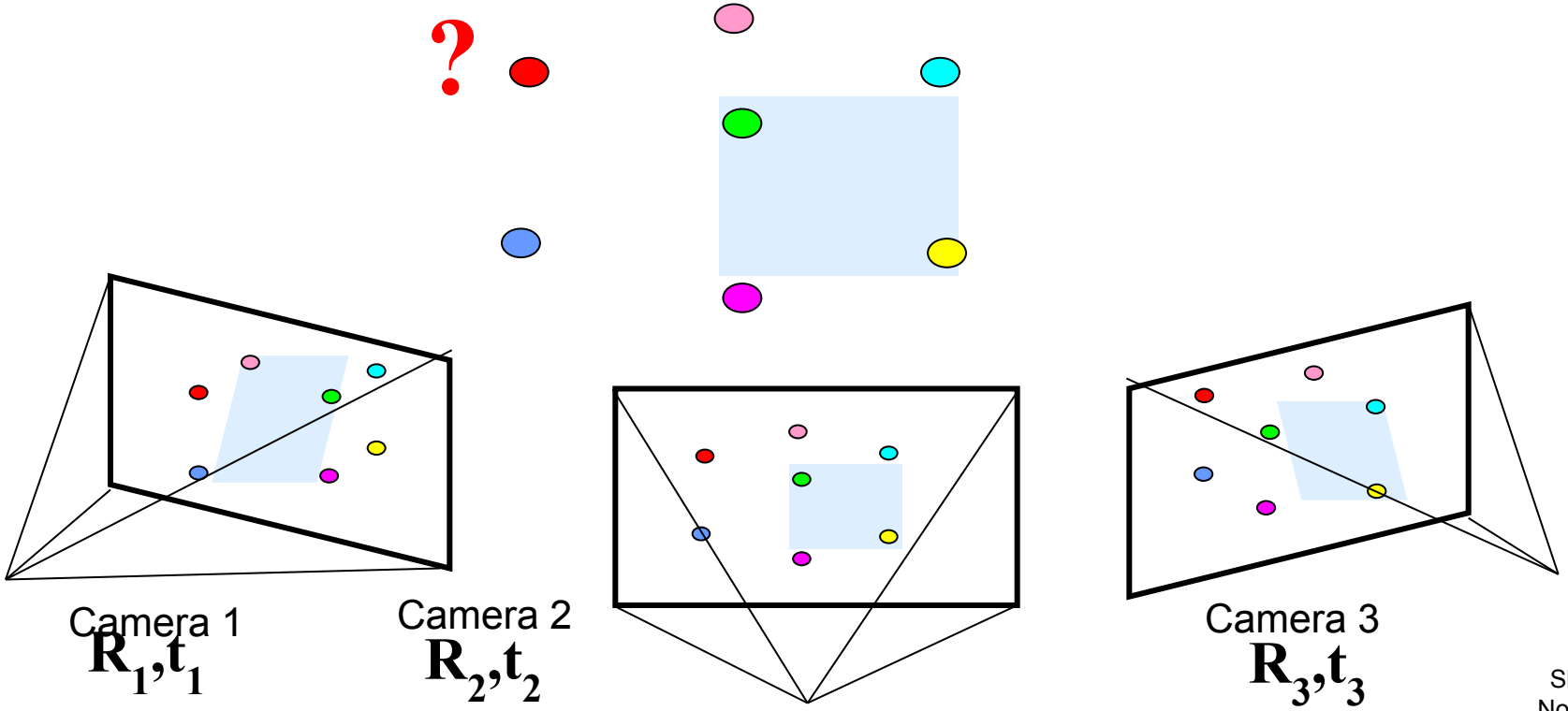
Multi-view geometry problems



Slide credit: Noah Snavely

Multi-view geometry problems

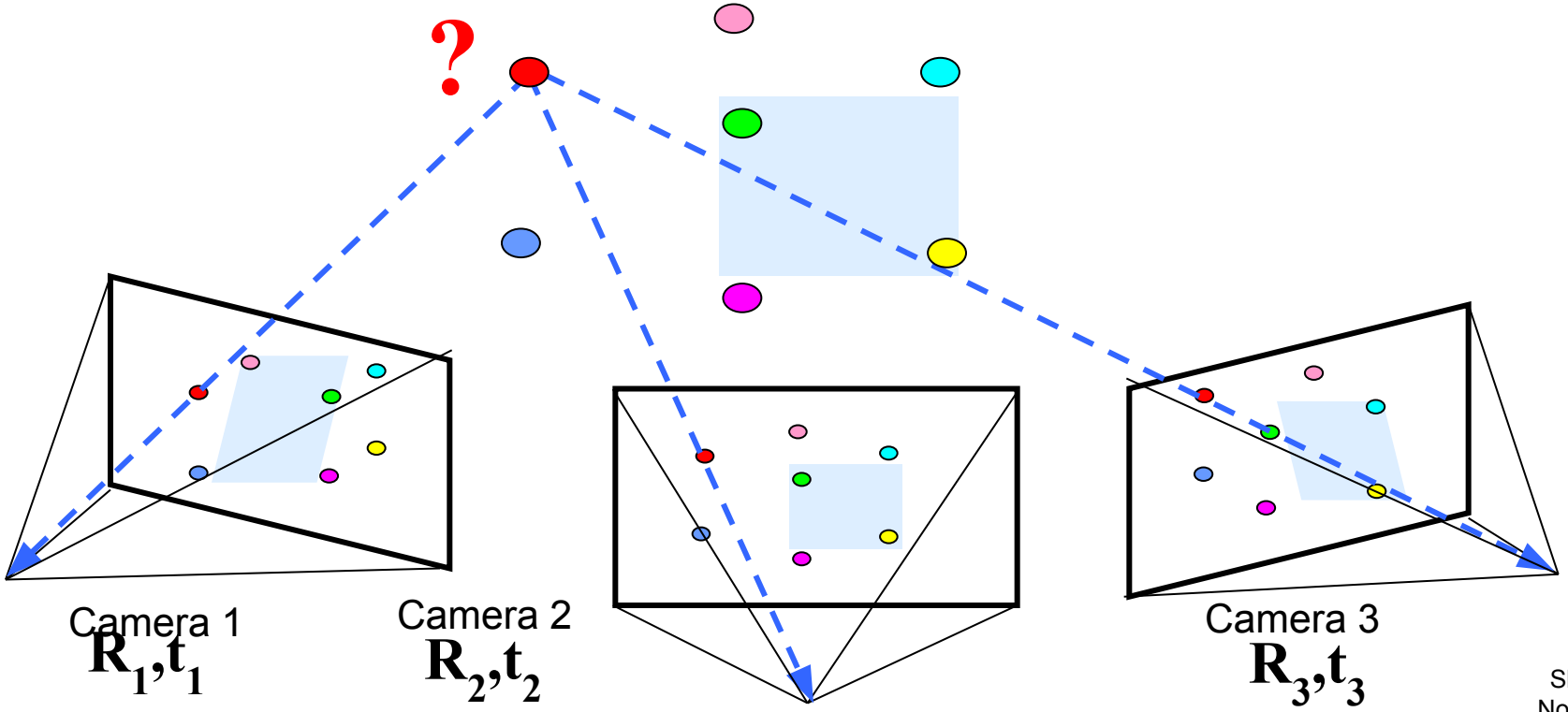
- **Structure:** Given projections of the same 3D point in two or more images, compute the 3D coordinates of that point



Slide credit: Noah Snavely

Multi-view geometry problems

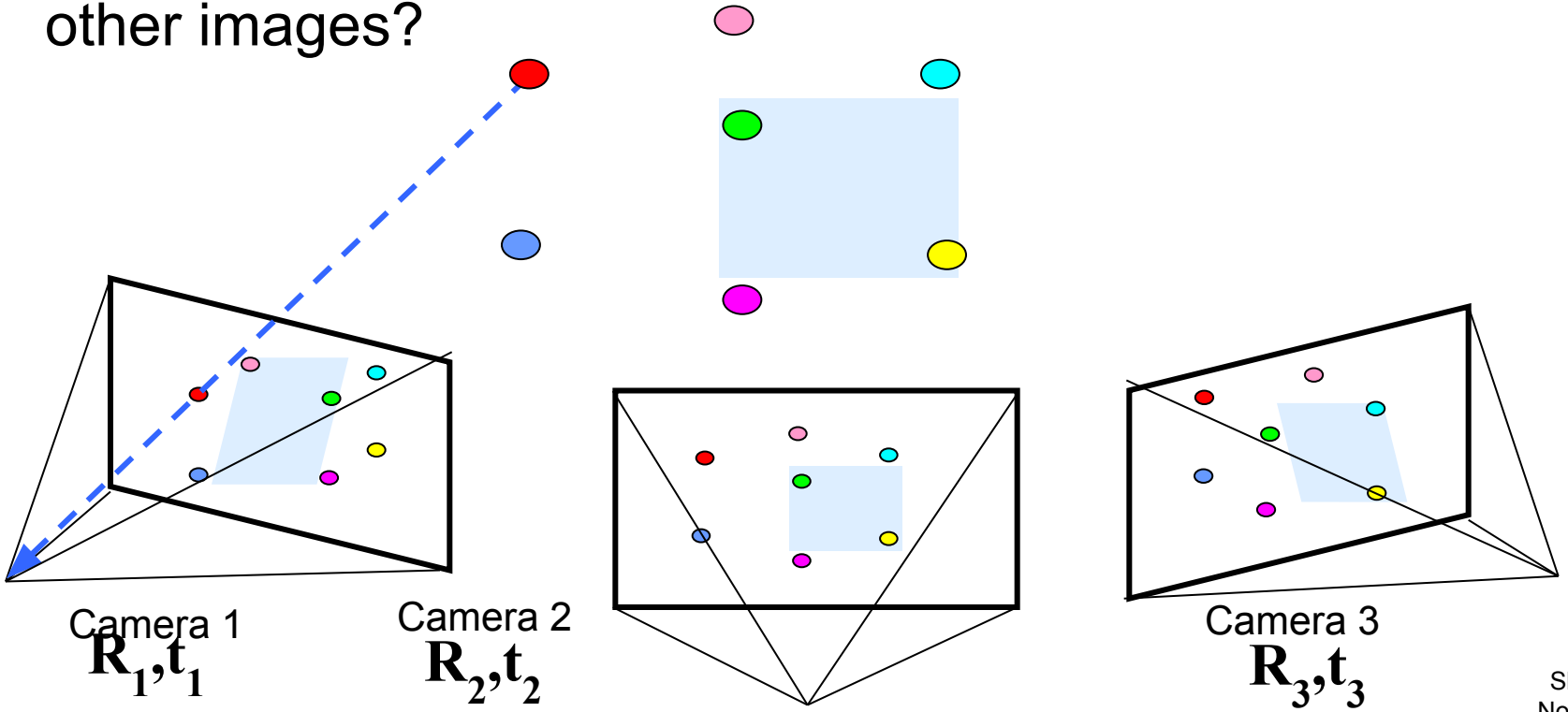
- **Structure:** Given projections of the same 3D point in two or more images, compute the 3D coordinates of that point



Slide credit: Noah Snavely

Multi-view geometry problems

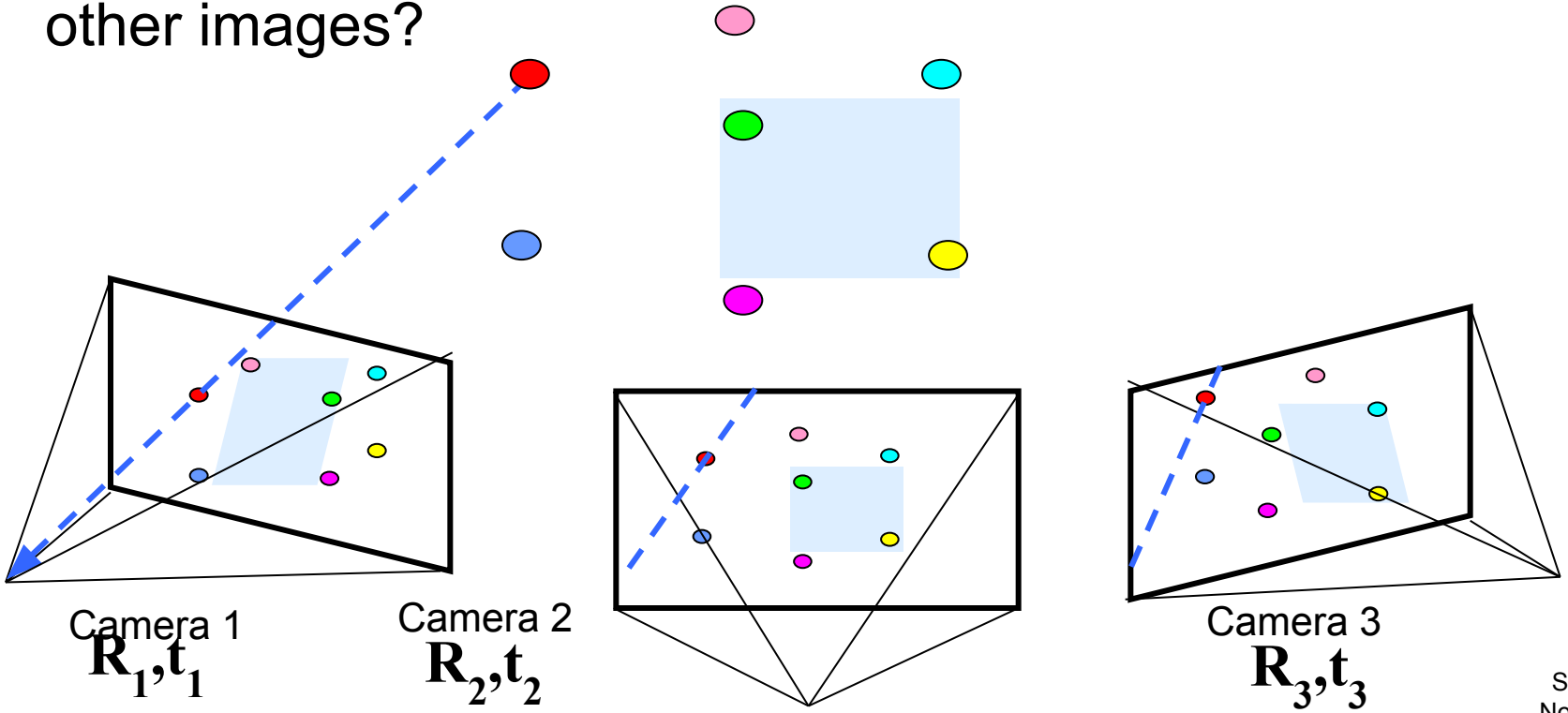
- **Stereo correspondence:** Given a point in one of the images, where could its corresponding points be in the other images?



Slide credit: Noah Snavely

Multi-view geometry problems

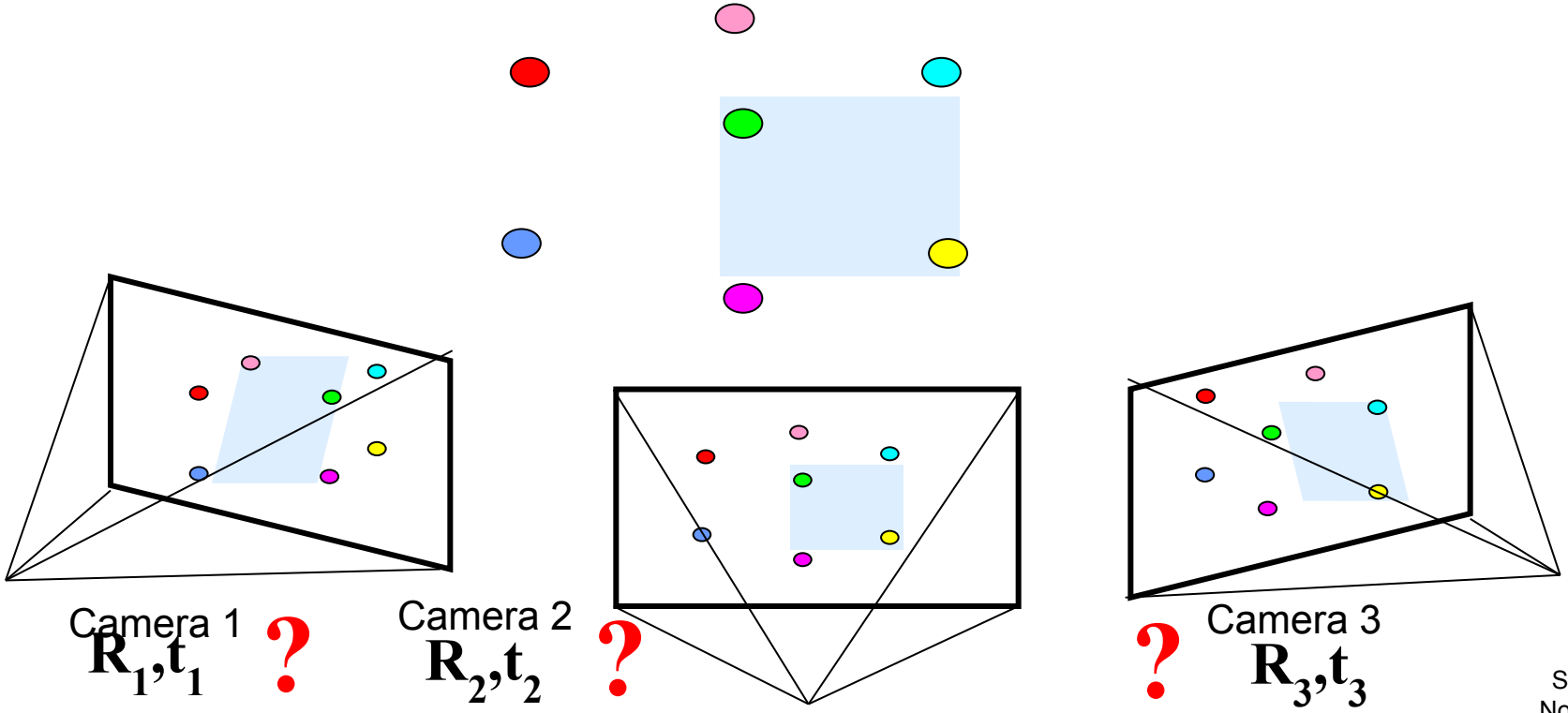
- **Stereo correspondence:** Given a point in one of the images, where could its corresponding points be in the other images?



Slide credit: Noah Snavely

Multi-view geometry problems

- **Motion:** Given a set of corresponding points in two or more images, compute the camera parameters



Slide credit: Noah Snavely

Human eye

Rough analogy with human visual system:

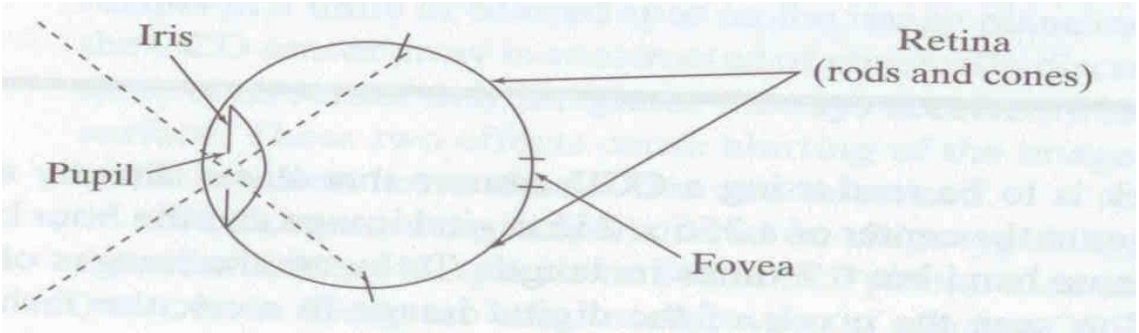
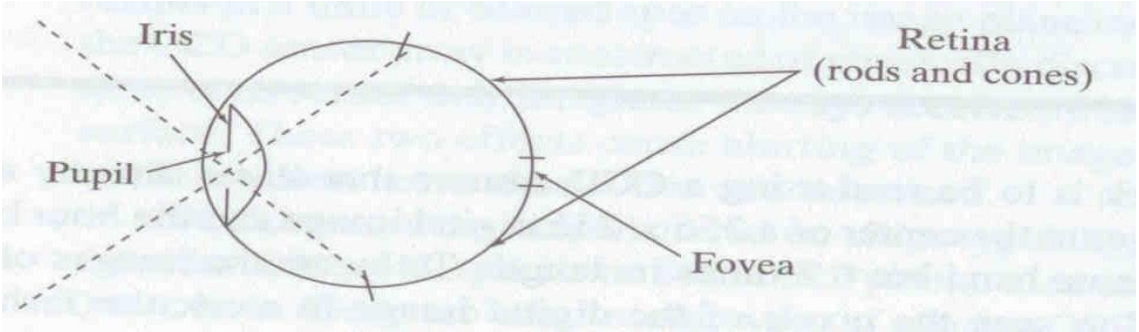


Fig from Shapiro and Stockman

Human eye

Rough analogy with human visual system:



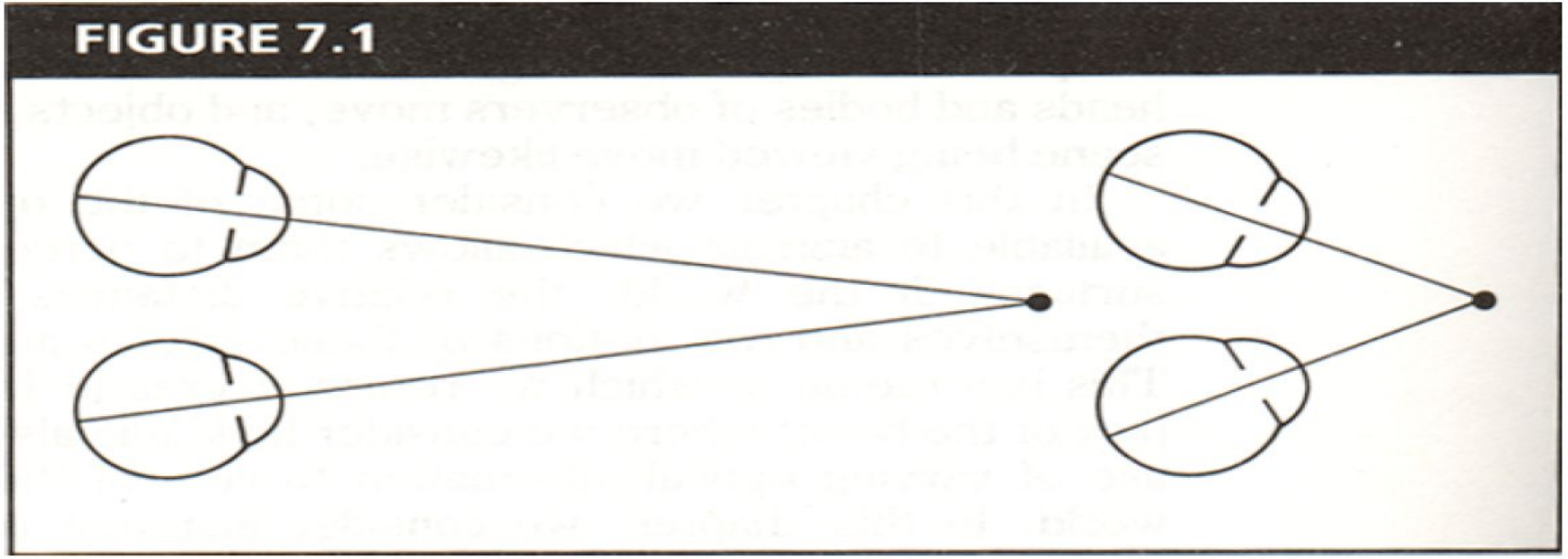
Pupil/Iris – control amount of light passing through lens

Retina - contains sensor cells, where image is formed

Fovea – highest concentration of cones

Fig from Shapiro and Stockman

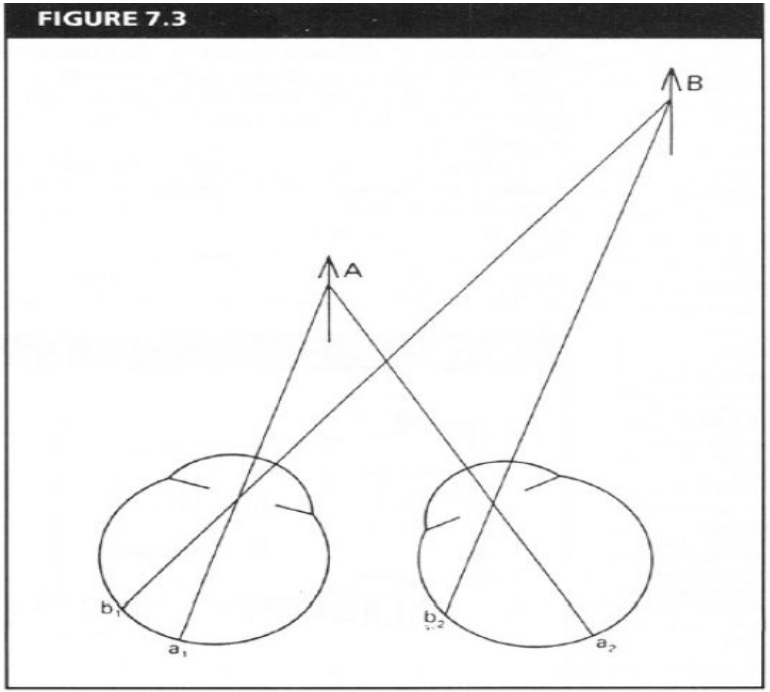
Human stereopsis: disparity



From Bruce and Green, Visual Perception, Physiology, Psychology and Ecology

Human eyes **fixate** on point in space – rotate so that corresponding images form in centers of fovea.

Human stereopsis: disparity



From Bruce and Green, Visual Perception, Physiology, Psychology and Ecology

Disparity occurs when eyes fixate on one object; others appear at different visual angles

Adapted from David Forsyth, UC Berkeley

Random dot stereograms

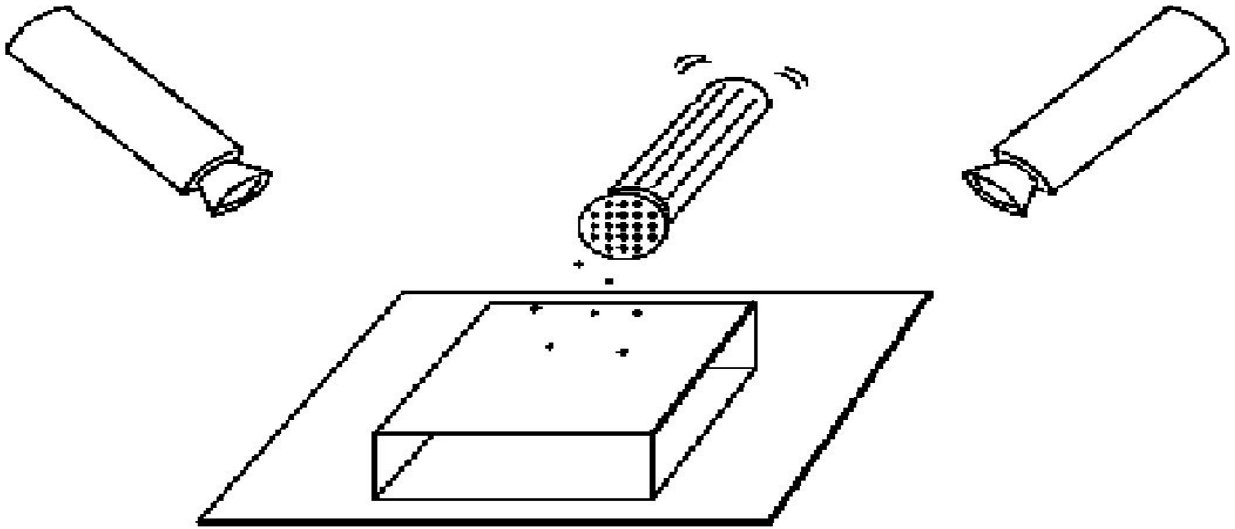
- Julesz 1960: Do we identify local brightness patterns before fusion (monocular process) or after (binocular)?

Random dot stereograms

- Julesz 1960: Do we identify local brightness patterns before fusion (monocular process) or after (binocular)?

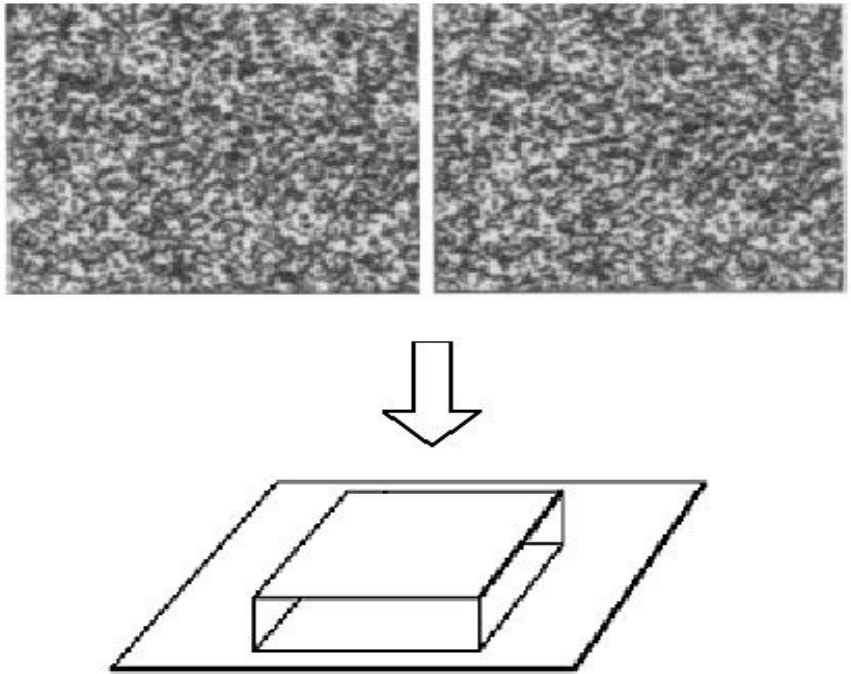
- To test: pair of synthetic images obtained by randomly spraying black dots on white objects

Random dot stereograms



Forsyth & Ponce

Random dot stereograms



Random dot stereograms

- When viewed monocularly, they appear random; when viewed stereoscopically, see 3d structure.

Random dot stereograms

- When viewed monocularly, they appear random; when viewed stereoscopically, see 3d structure.
- Human binocular fusion not directly associated with the physical retinas; must involve the central nervous system (V2, for instance)

Random dot stereograms

- When viewed monocularly, they appear random; when viewed stereoscopically, see 3d structure.
- Human binocular fusion not directly associated with the physical retinas; must involve the central nervous system (V2, for instance)
- Imaginary "*cyclopean retina*" that combines the left and right image stimuli as a single unit

Random dot stereograms

- When viewed monocularly, they appear random; when viewed stereoscopically, see 3d structure.
- Human binocular fusion not directly associated with the physical retinas; must involve the central nervous system (V2, for instance)
- Imaginary "*cyclopean retina*" that combines the left and right image stimuli as a single unit
- High level scene understanding not required for Stereo

Random dot stereograms

- When viewed monocularly, they appear random; when viewed stereoscopically, see 3d structure.
- Human binocular fusion not directly associated with the physical retinas; must involve the central nervous system (V2, for instance)
- Imaginary "*cyclopean retina*" that combines the left and right image stimuli as a single unit
- High level scene understanding not required for Stereo
- But, high level scene understanding is arguably *better* than stereo

Stereo photography and stereo viewers

Take two pictures of the same subject from two slightly different viewpoints and display so that each eye sees only one of the images.

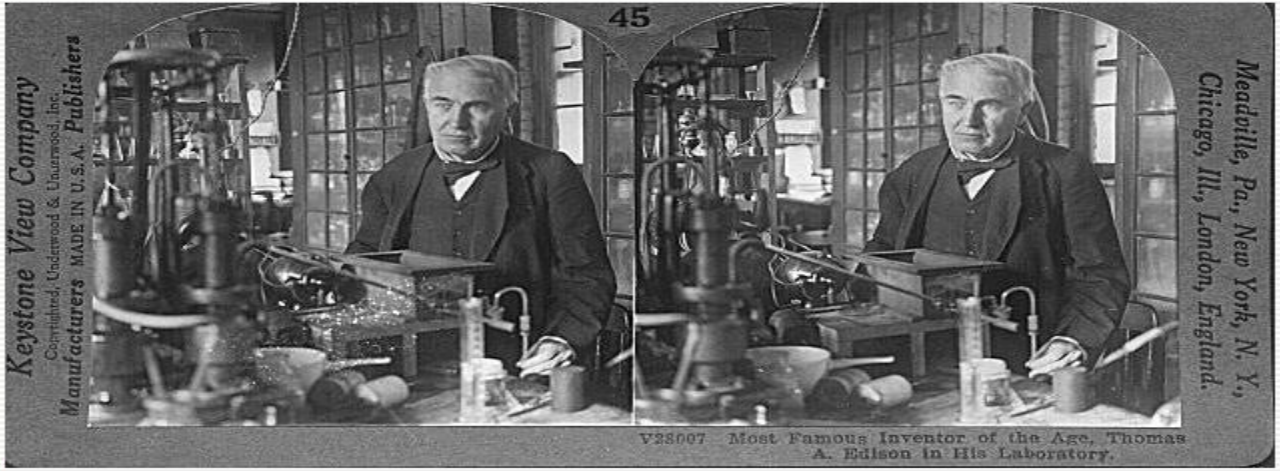
Invented by Sir Charles Wheatstone, 1838

Stereo photography and stereo viewers

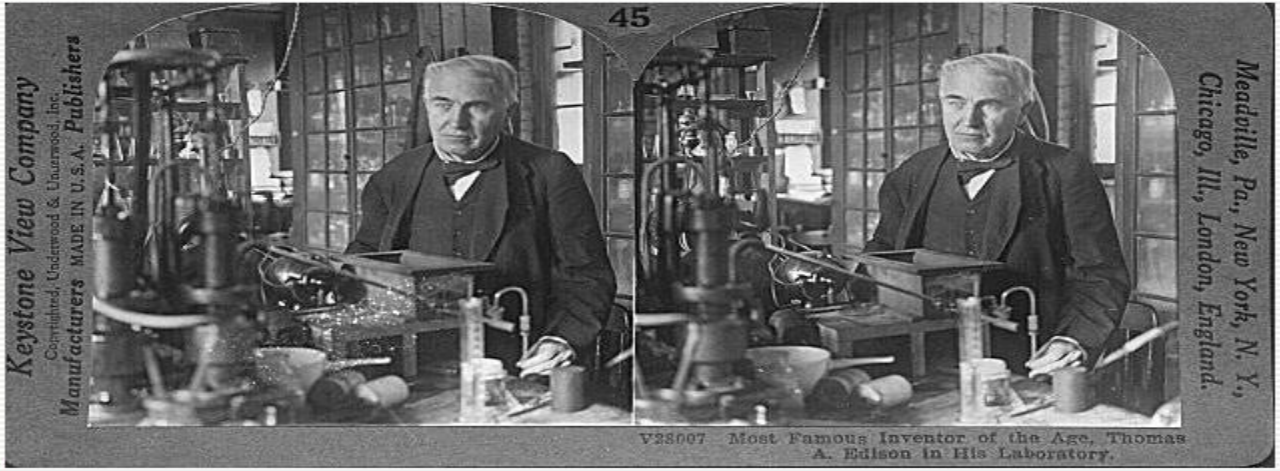
Take two pictures of the same subject from two slightly different viewpoints and display so that each eye sees only one of the images.

Invented by Sir Charles Wheatstone, 1838

Image from fisher-price.com



http://www.johnsonshawmuseum.org

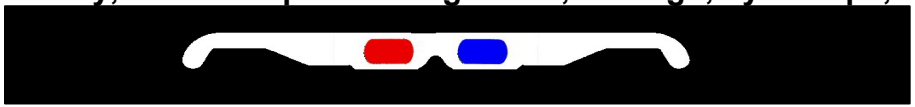


http://www.johnsonshawmuseum.org

http://www.johnsonshawmuseum.org

http://www.johnsonshawmuseum.org

Public Library, Stereoscopic Looking Room, Chicago, by Phillips, 1923



http://www.well.com/~jimg/stereo/stereo_list.html

Autostereograms

Exploit disparity as depth cue using single image.

(Single image random dot stereogram, Single image stereogram)

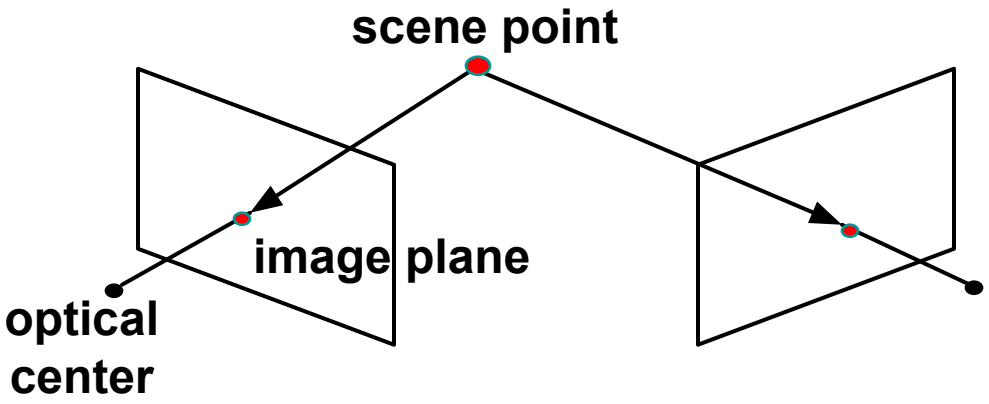
Images from magiceye.com

Autostereograms

Images from magiceye.com

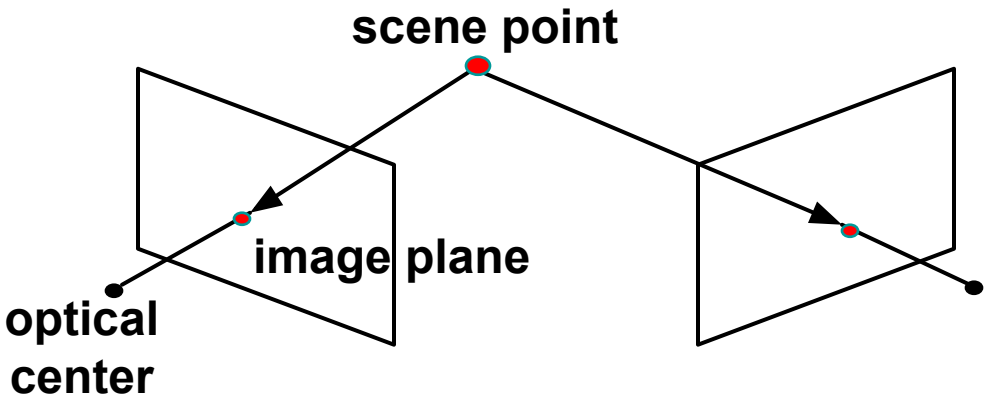
Estimating depth with stereo

- **Stereo**: shape from “motion” between two views



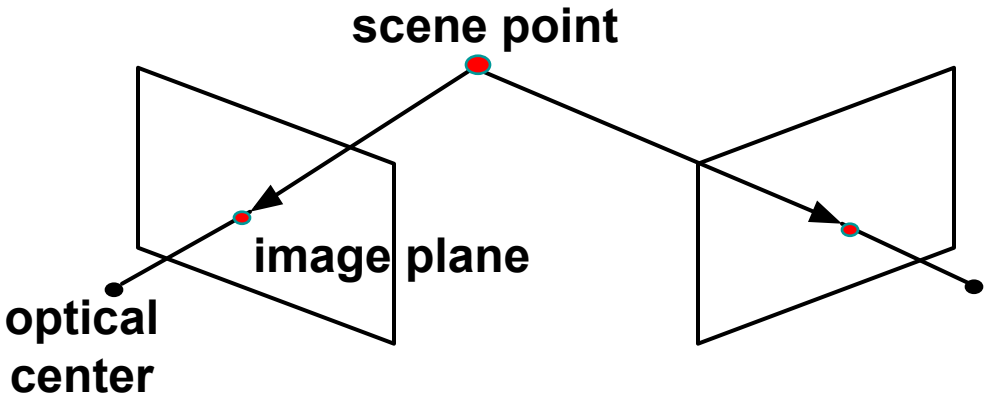
Estimating depth with stereo

- **Stereo**: shape from “motion” between two views
- We’ll need to consider:
- Info on camera pose (“calibration”)



Estimating depth with stereo

- **Stereo**: shape from “motion” between two views
- We’ll need to consider:
- Info on camera pose (“calibration”)
- Image point correspondences

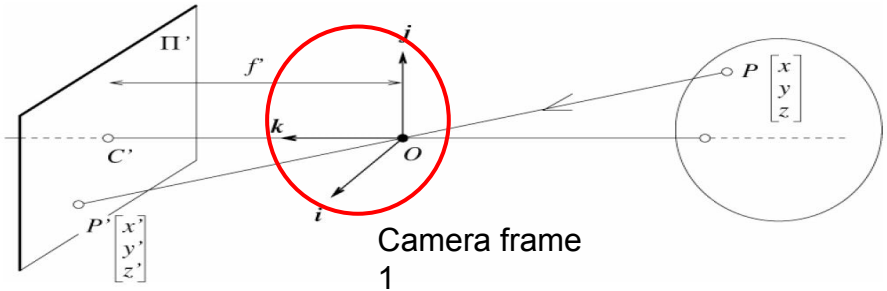


Stereo vision

Two cameras, simultaneous views

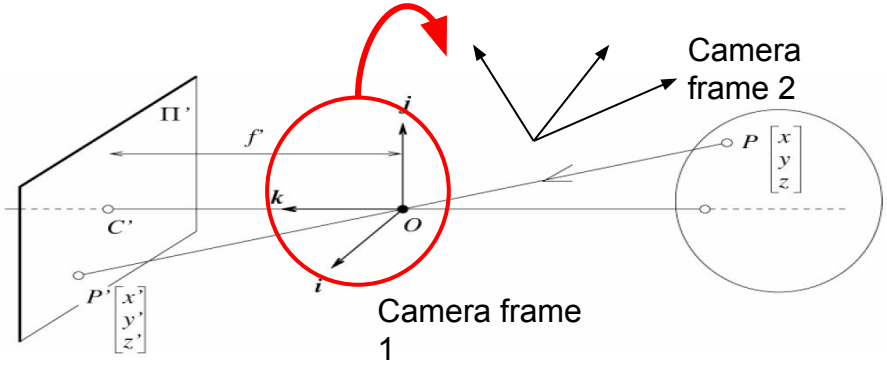
Single moving camera and static scene

Camera parameters



- *Extrinsic* params: rotation matrix and translation vector
- *Intrinsic* params: focal length, pixel sizes (mm), image center point, radial distortion parameters

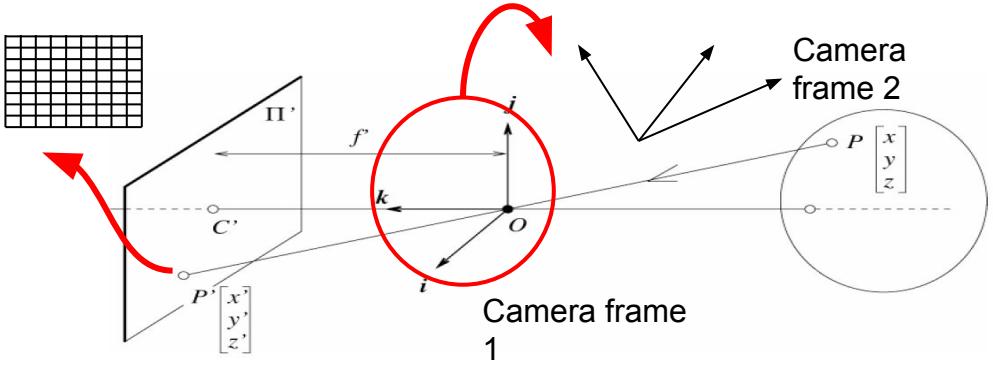
Camera parameters



Extrinsic parameters:
Camera frame 1 ⟷ Camera frame 2

- *Extrinsic* params: rotation matrix and translation vector
- *Intrinsic* params: focal length, pixel sizes (mm), image center point, radial distortion parameters

Camera parameters

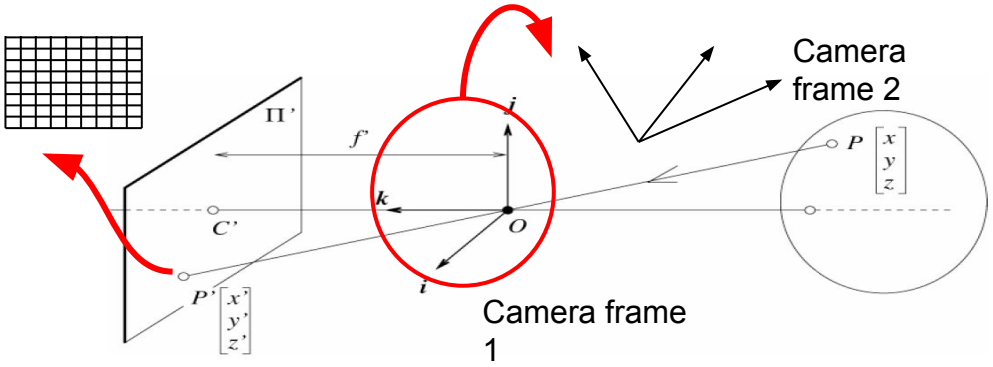


Extrinsic parameters:
Camera frame 1 ⟷ Camera frame 2

Intrinsic parameters:
Image coordinates relative to camera ⟷ Pixel coordinates

- *Extrinsic* params: rotation matrix and translation vector
- *Intrinsic* params: focal length, pixel sizes (mm), image center point, radial distortion parameters

Camera parameters



Extrinsic parameters:
Camera frame 1 ⟷ Camera frame 2

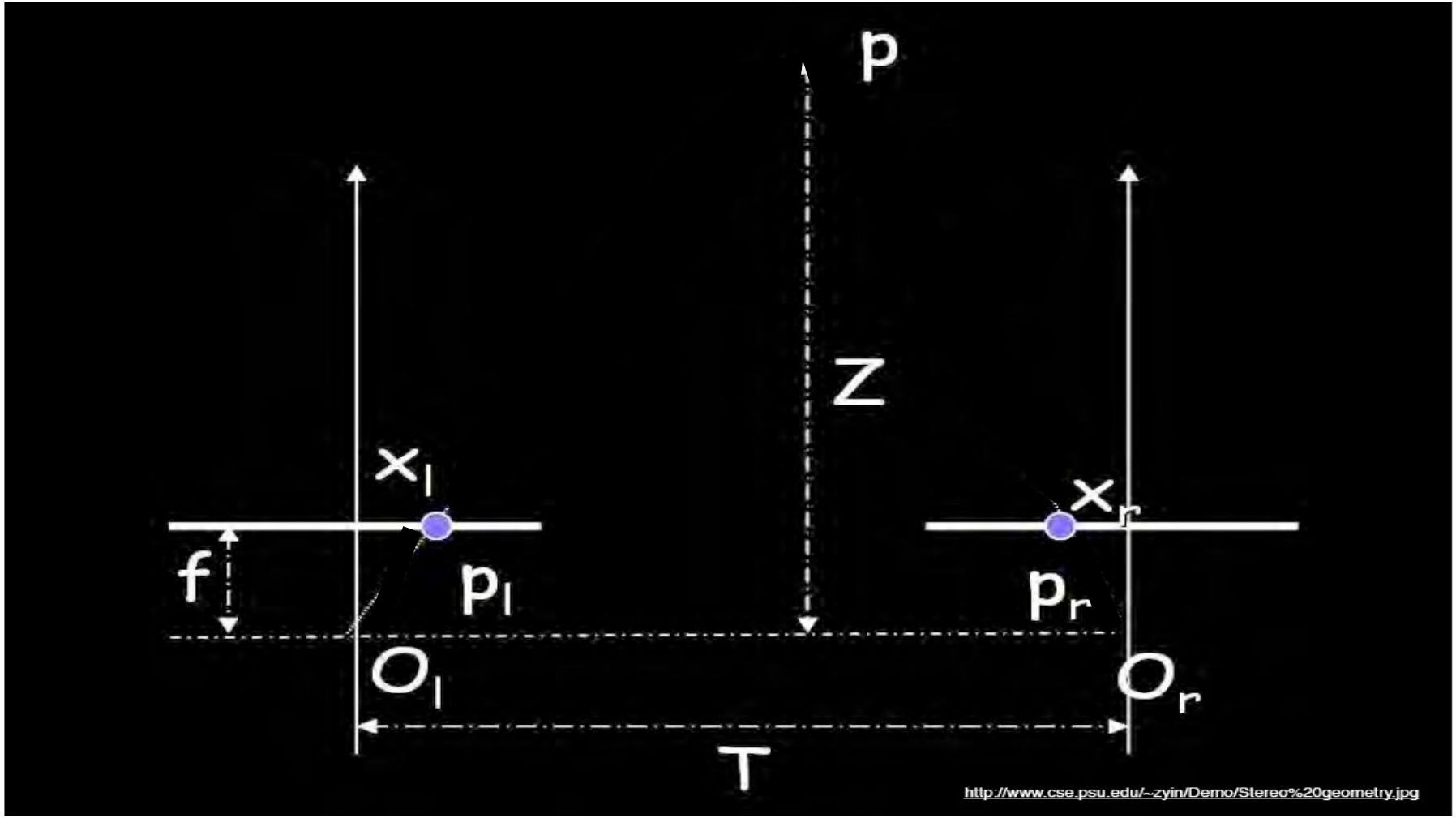
Intrinsic parameters:
Image coordinates relative to camera ⟷ Pixel coordinates

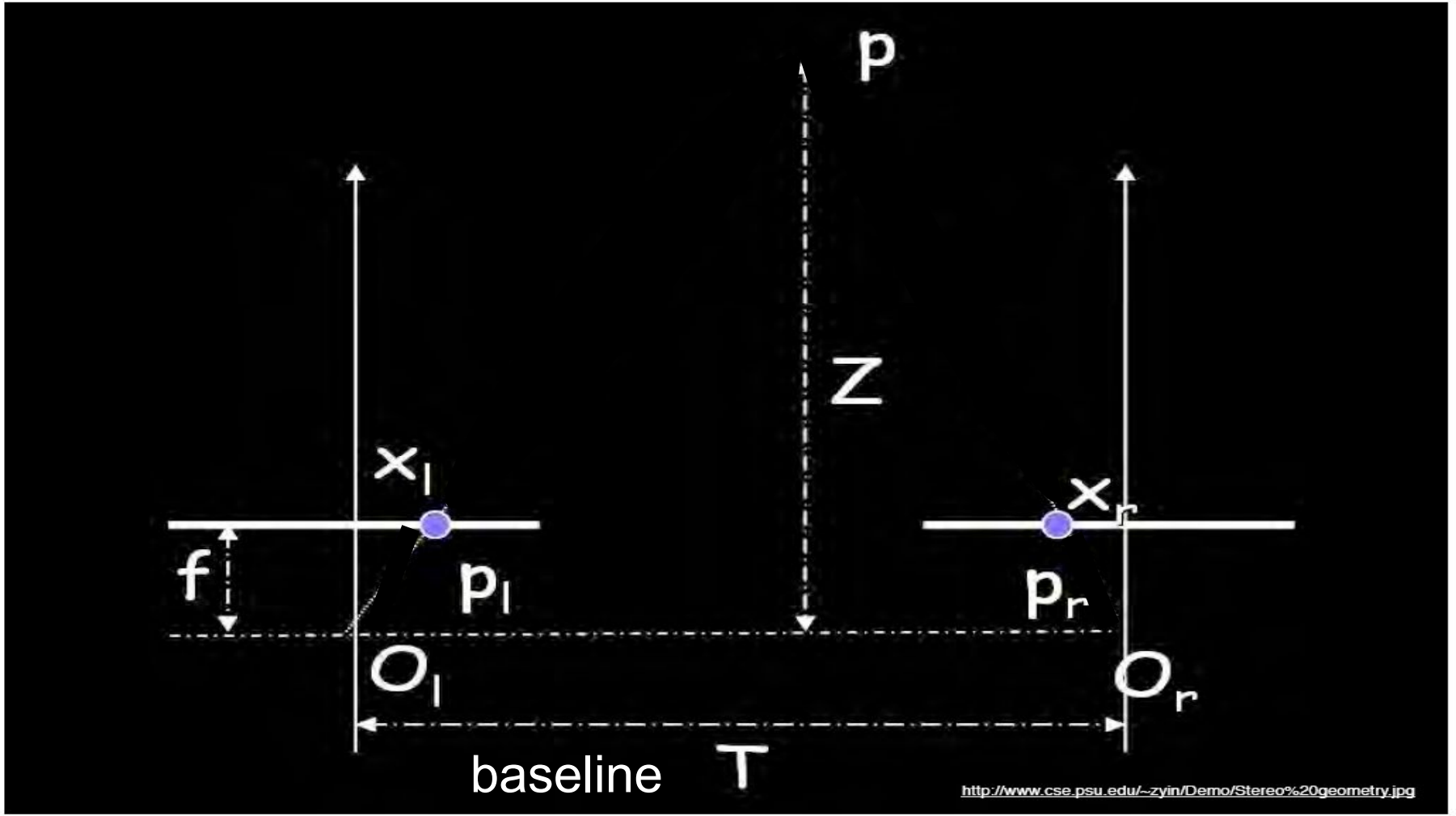
- *Extrinsic* params: rotation matrix and translation vector
- *Intrinsic* params: focal length, pixel sizes (mm), image center point, radial distortion parameters

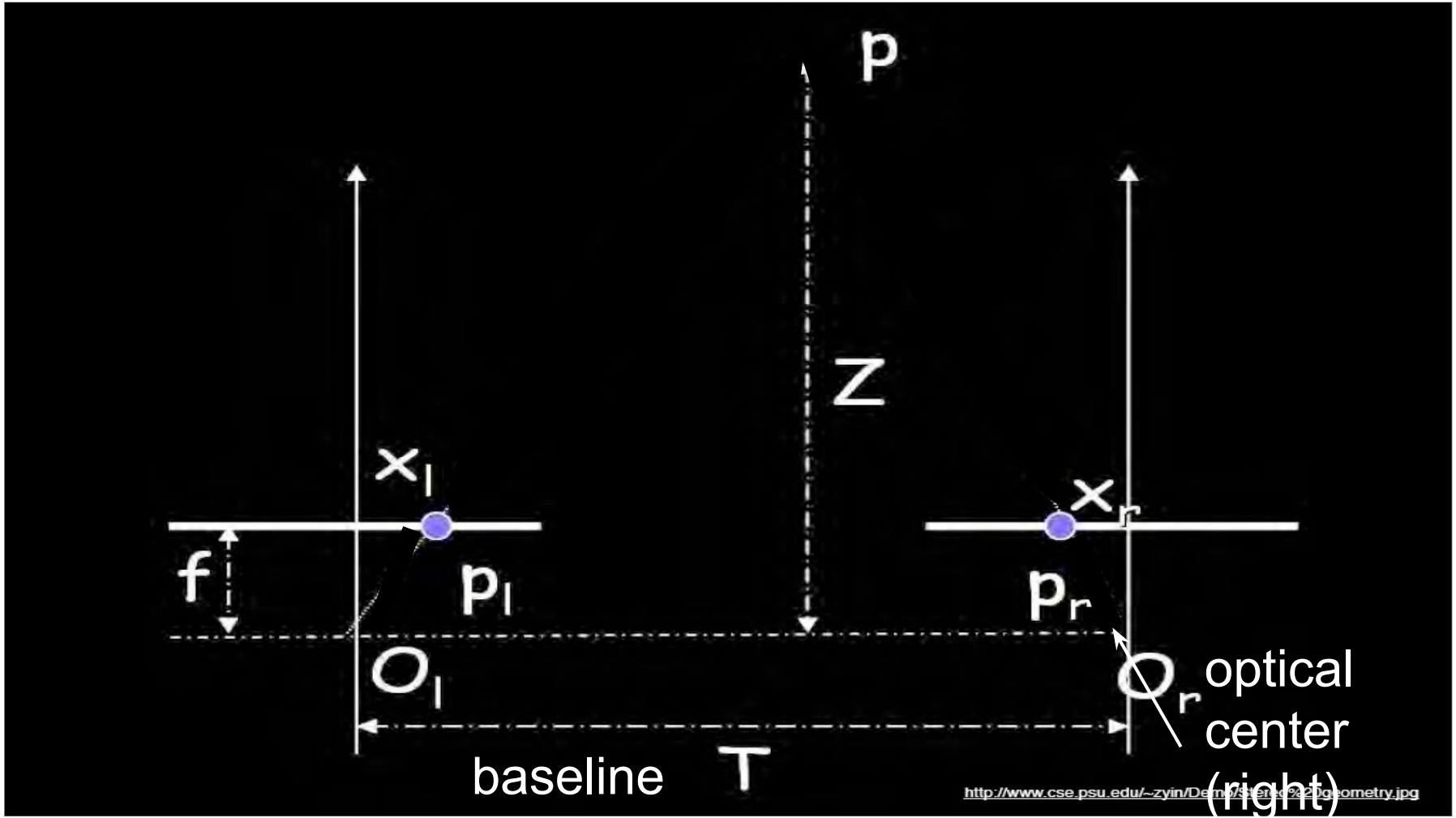
We'll assume for now that these parameters are given and fixed.

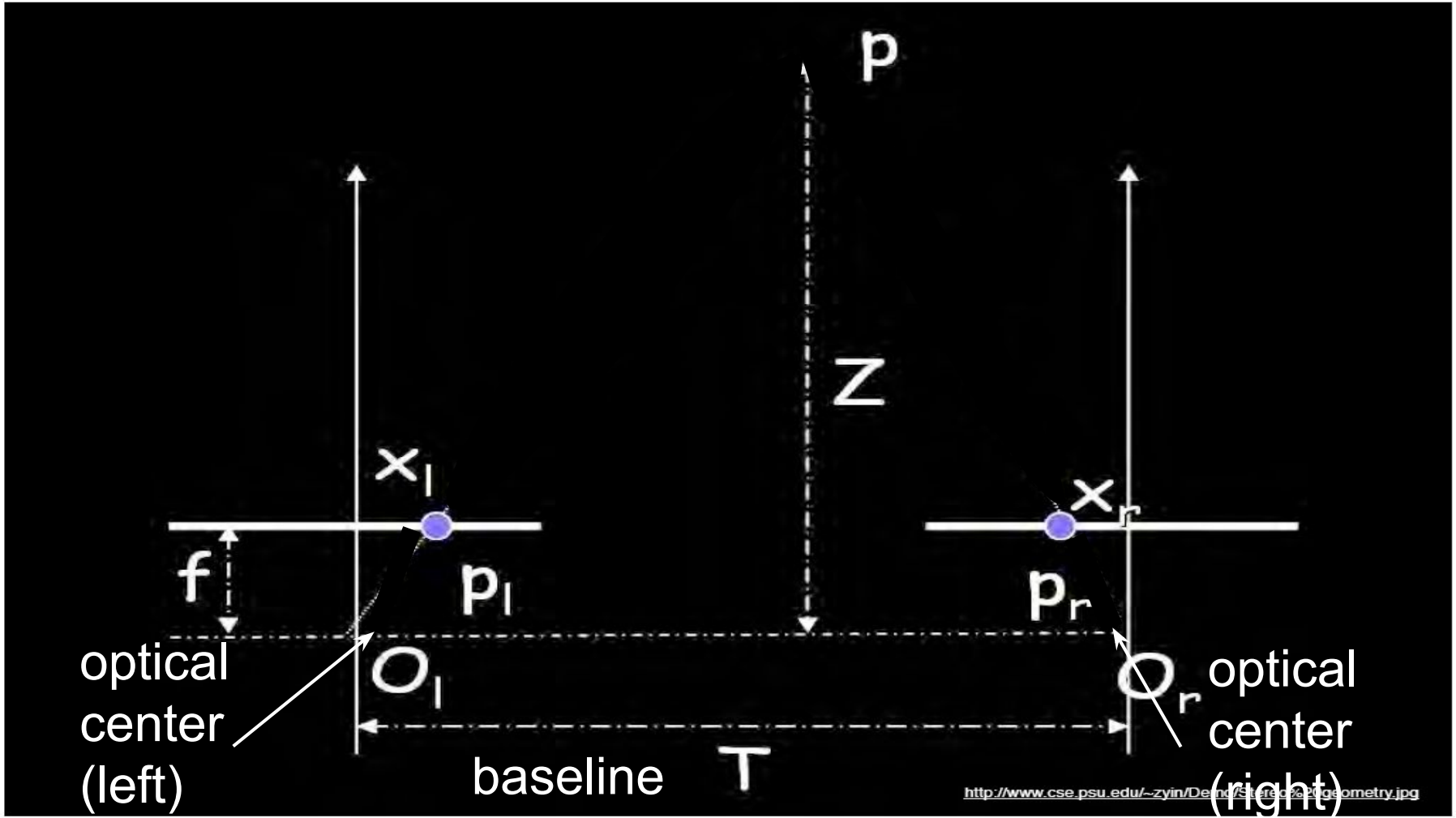
Geometry for a simple stereo system

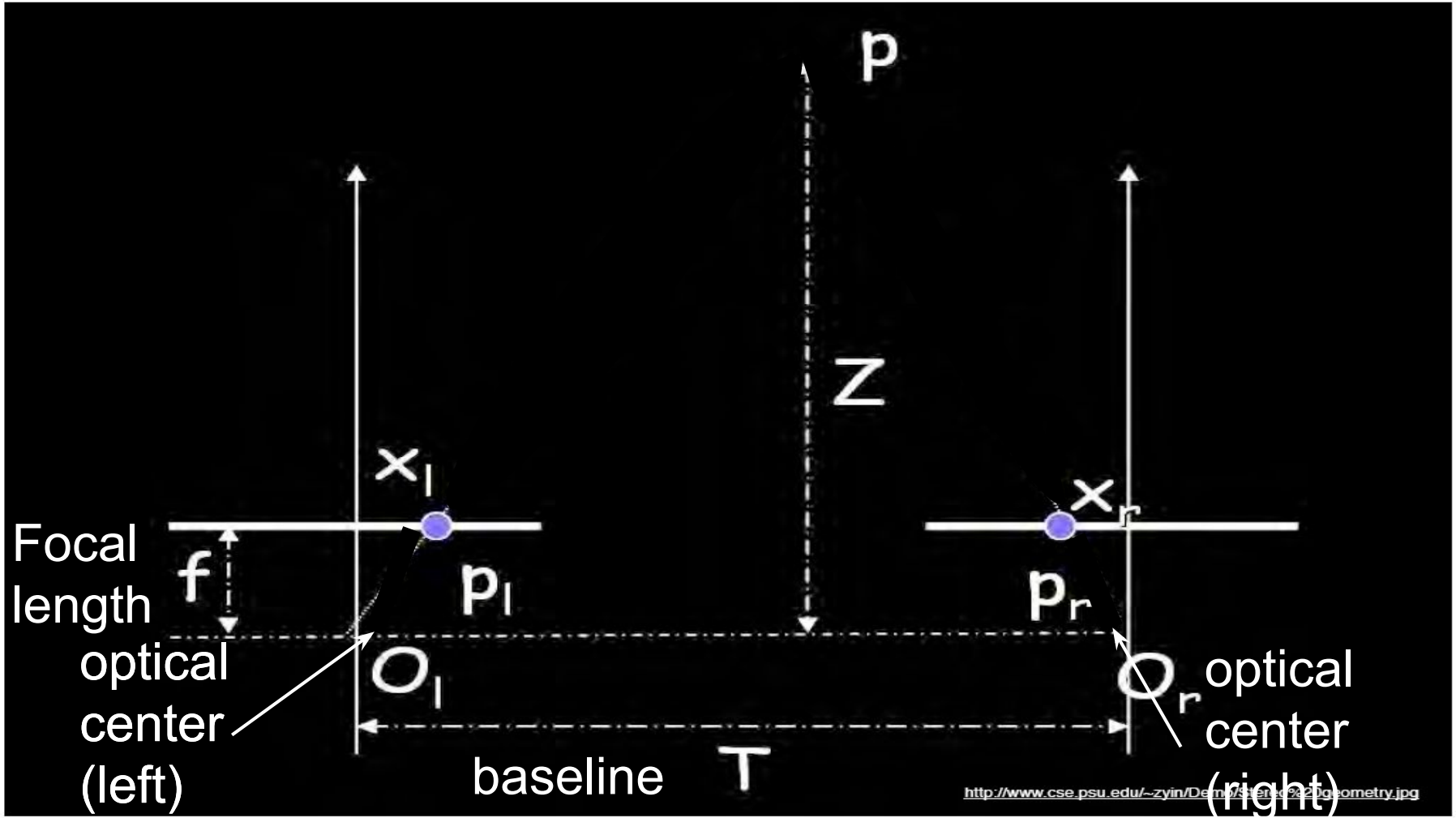
- First, assuming parallel optical axes, known camera parameters (i.e., calibrated cameras):

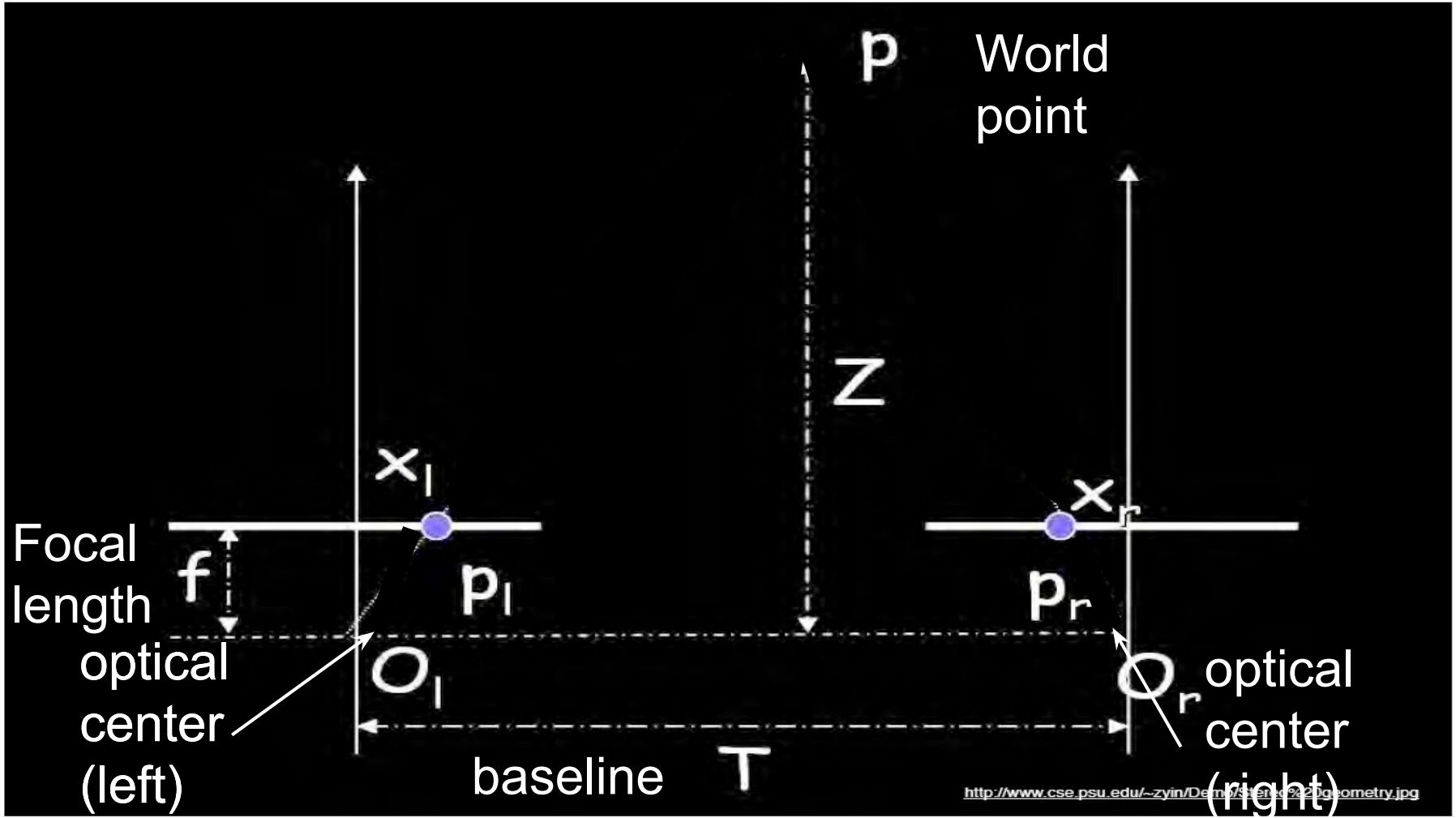

p
Z
x_l
x_r
f
p_l
p_r
O_l
O_r
T
http://www.cse.psu.edu/~zyin/Demo/Stereo%20geometry.jpg

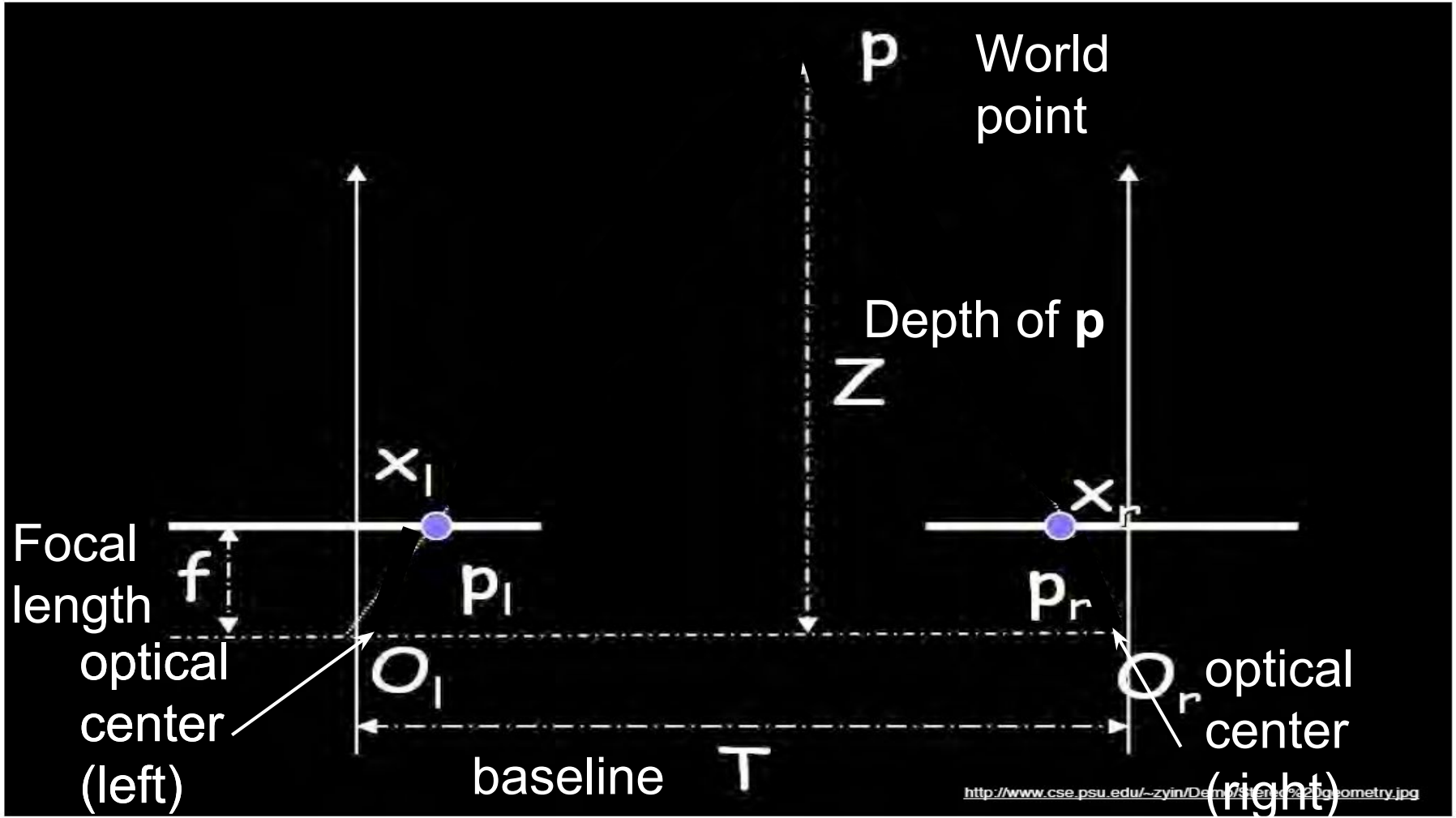

P
Z
x_l
x_r
f
p_l
p_r
O_l
O_r
baseline
T
http://www.cse.psu.edu/~zyin/Demo/Stereo%20geometry.jpg

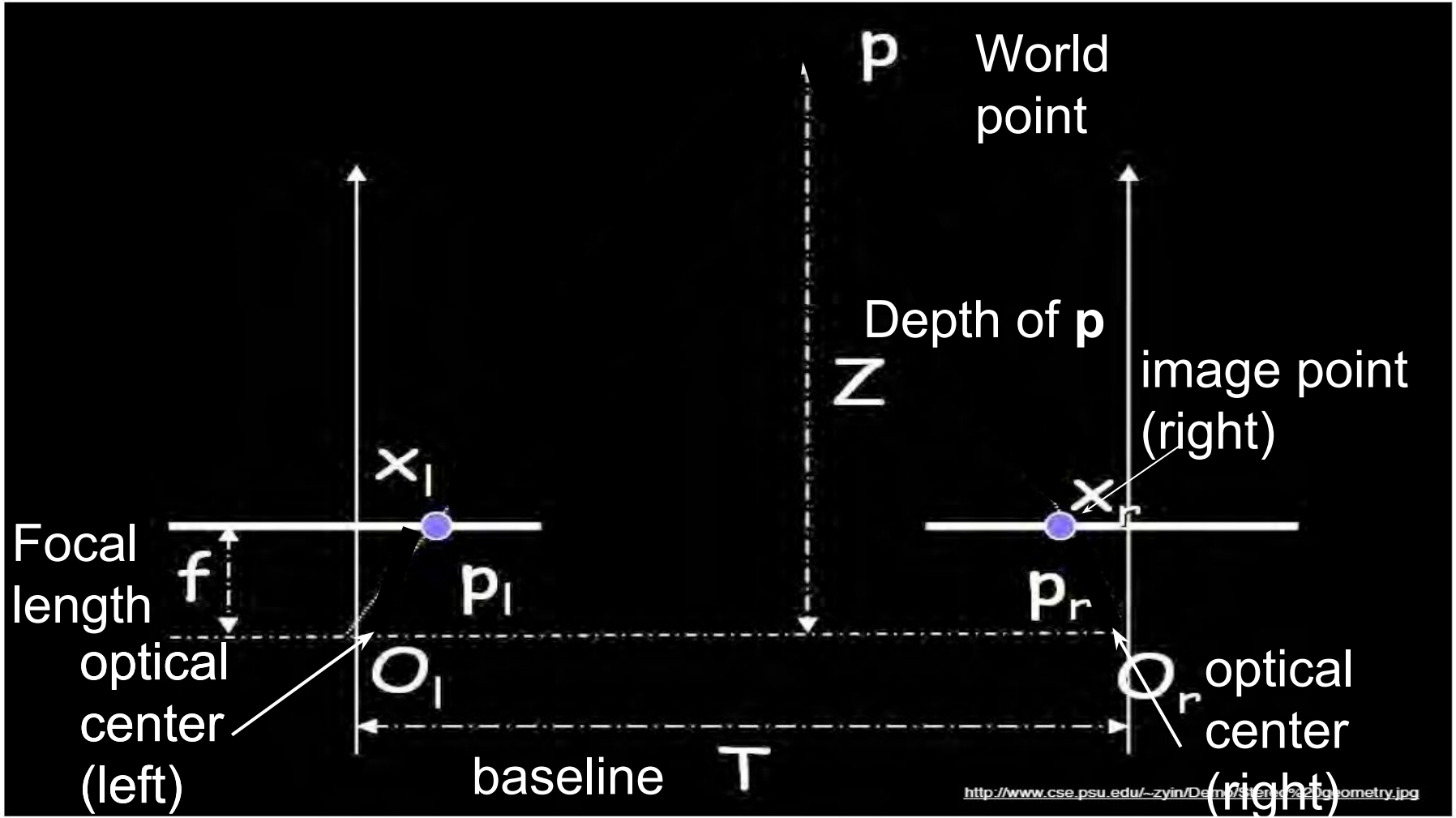

P
Z
x_l
x_r
f
p_l
p_r
O_l
O_r
optical center (right)
baseline
T
http://www.cse.psu.edu/~zyin/Demo/Stereo%20geometry.jpg

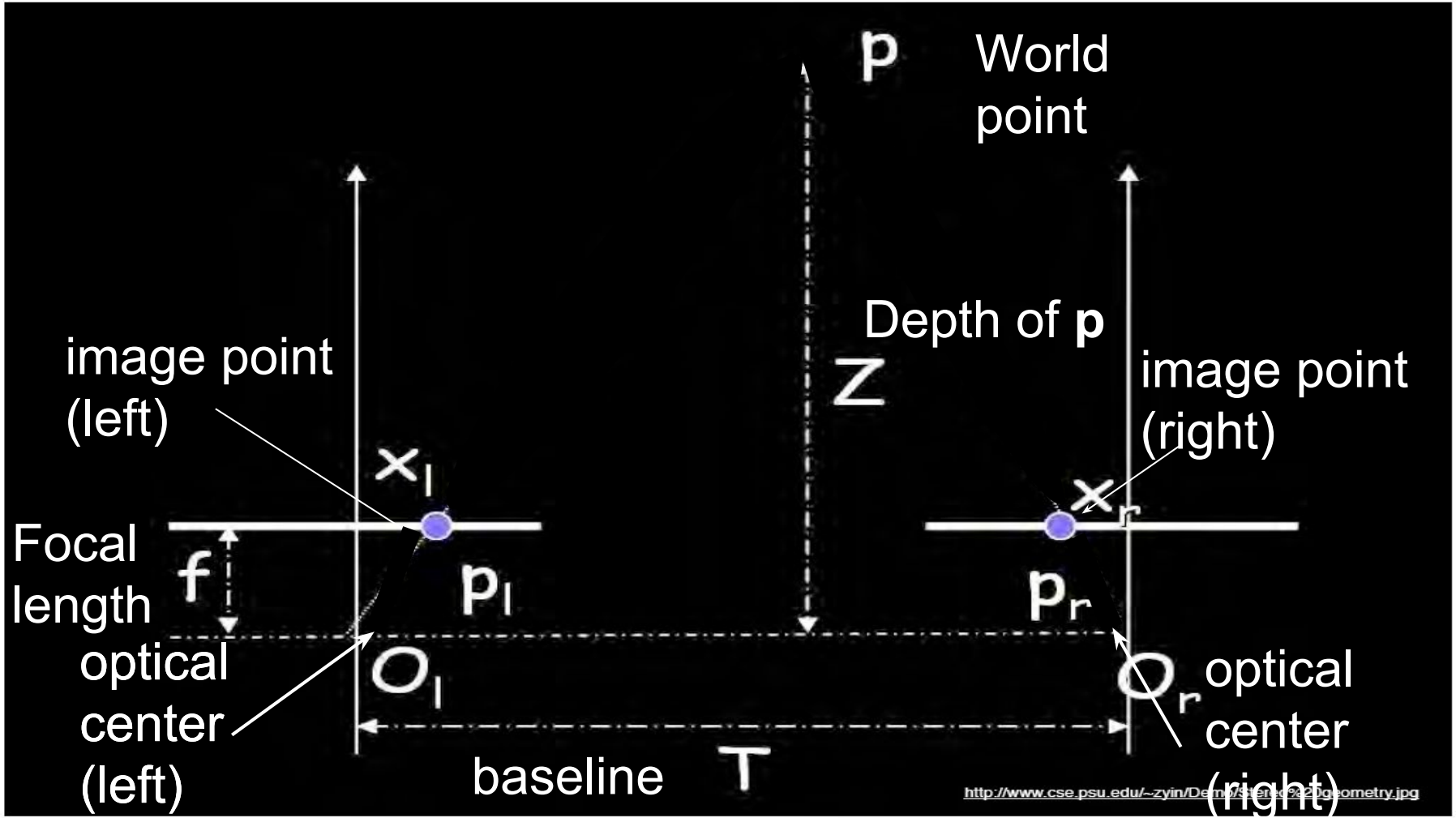

P
Z
x_l
x_r
f
p_l
p_r
O_l
O_r
optical center (left)
optical center (right)
baseline
T
http://www.cse.psu.edu/~zyin/Demo/Stereo%20geometry.jpg


P
Z
x_l
x_r
Focal length
f
p_l
p_r
O_l
O_r
optical center (left)
optical center (right)
baseline
T
http://www.cse.psu.edu/~zyin/Demo/Stereo%20Geometry.jpg


P
World point
Z
x_l
x_r
Focal length
f
p_l
p_r
O_l
O_r
optical center (left)
optical center (right)
baseline
T
http://www.cse.psu.edu/~zyin/Demo/Stereo%20Geometry.jpg

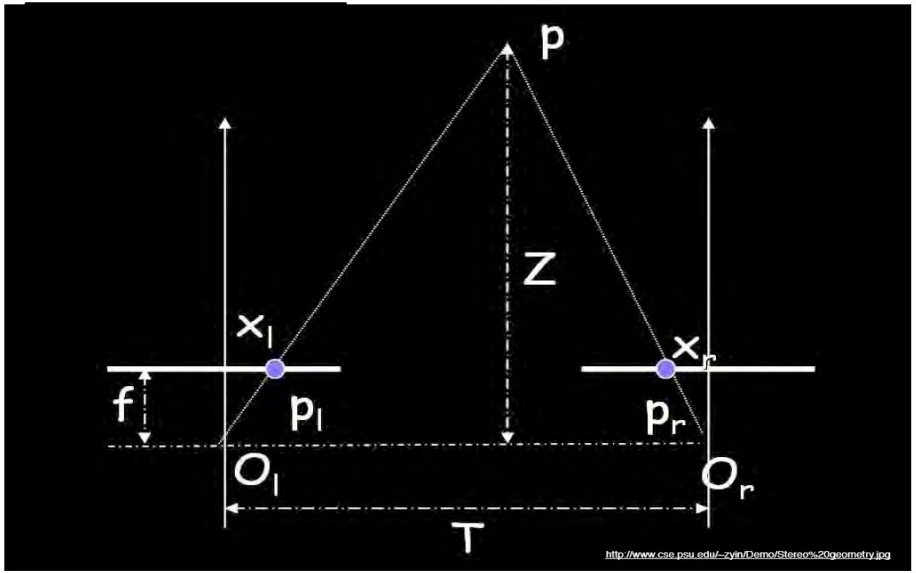

p
World point
Depth of p
Z
x_l
x_r
Focal length
f
p_l
p_r
O_l
O_r
optical center (left)
optical center (right)
baseline
T
http://www.cse.psu.edu/~zyin/Demo/Stereo%20Geometry.jpg


p
World point
Depth of p
Z
image point (right)
x_l
x_r
Focal length
f
p_l
p_r
O_l
O_r
optical center (left)
optical center (right)
baseline
T
http://www.cse.psu.edu/~zyin/Demo/Stereo%20Geometry.jpg


p
World point
Depth of p
Z
image point (left)
image point (right)
x_l
x_r
Focal length
f
p_l
p_r
O_l
O_r
optical center (left)
optical center (right)
baseline
T
http://www.cse.psu.edu/~zyin/Demo/Stereo%20Geometry.jpg

Geometry for a simple stereo system

- Assume parallel optical axes, known camera parameters (i.e., calibrated cameras). **What is expression for Z?**



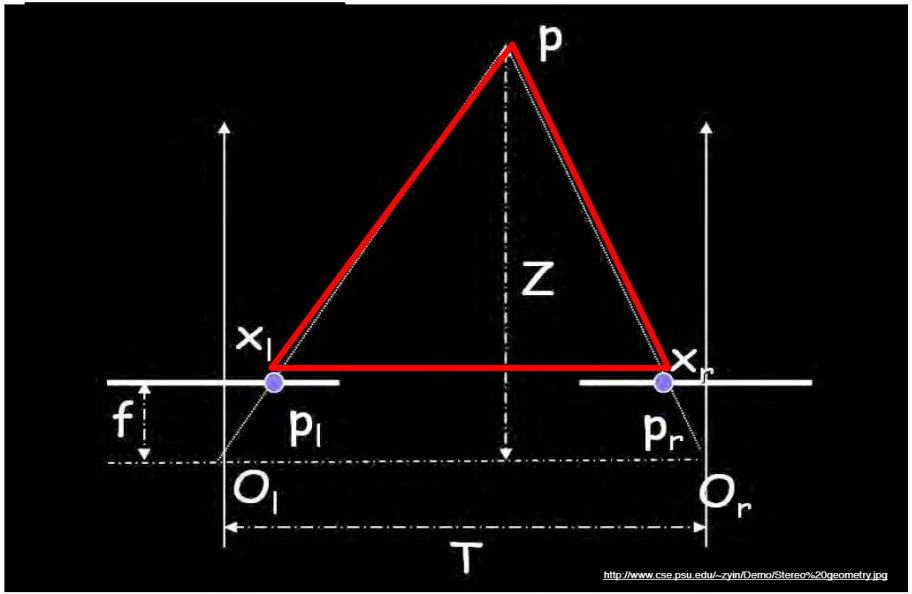
Similar triangles (p_l, P, p_r) and (O_l, P, O_r):

$$\frac{T + x_l - x_r}{Z - f} = \frac{T}{Z}$$

$$Z = f\frac{T}{x_r - x_l}$$

Geometry for a simple stereo system

- Assume parallel optical axes, known camera parameters (i.e., calibrated cameras). **What is expression for Z?**



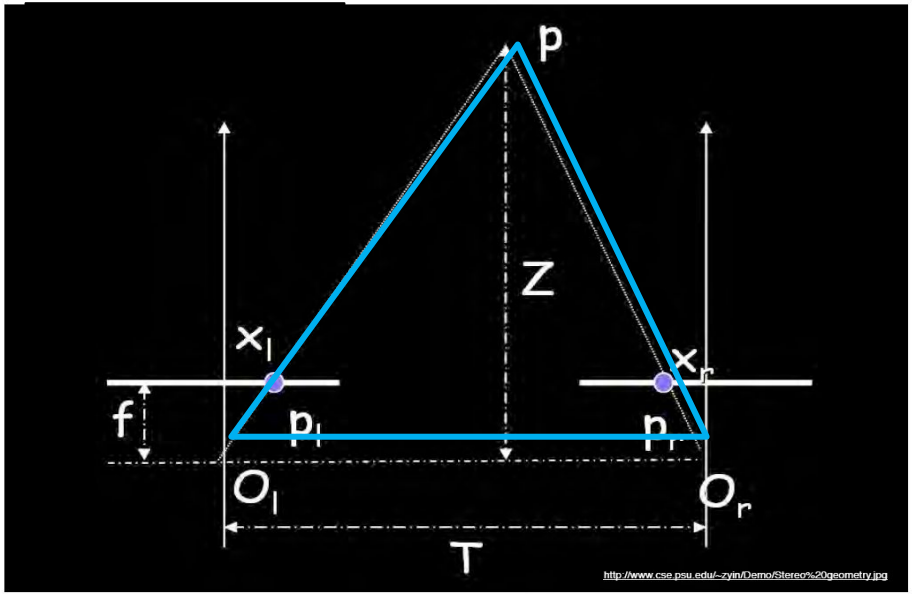
Similar triangles (p_l, P, p_r) and (O_l, P, O_r):

$$\frac{T + x_l - x_r}{Z - f} = \frac{T}{Z}$$

$$Z = f\frac{T}{x_r - x_l}$$

Geometry for a simple stereo system

- Assume parallel optical axes, known camera parameters (i.e., calibrated cameras). **What is expression for Z?**



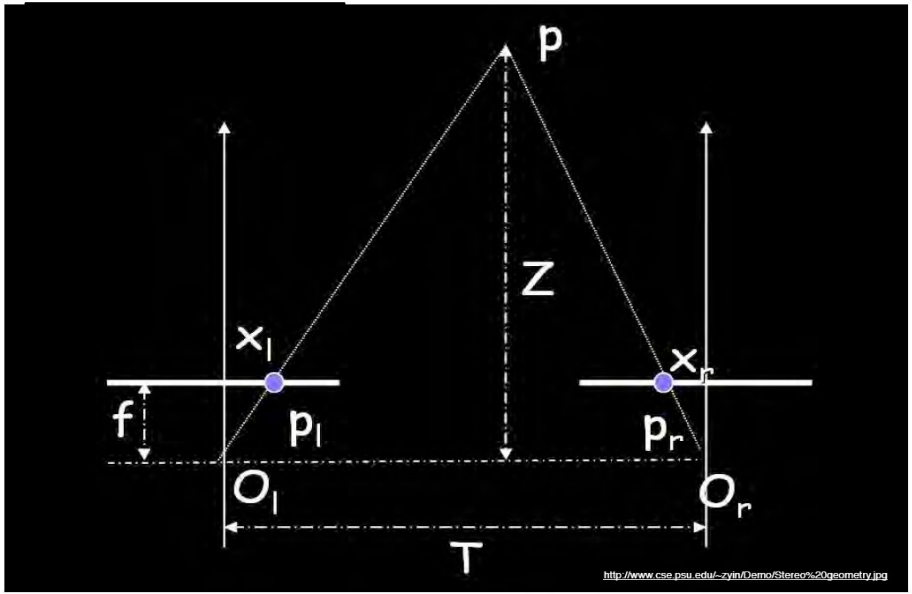
Similar triangles (p_l, P, p_r) and (O_l, P, O_r):

$$\frac{T + x_l - x_r}{Z - f} = \frac{T}{Z}$$

$$Z = f\frac{T}{x_r - x_l}$$

Geometry for a simple stereo system

- Assume parallel optical axes, known camera parameters (i.e., calibrated cameras). **What is expression for Z?**



Similar triangles (p_l, P, p_r) and (O_l, P, O_r):

$$\frac{T + x_l - x_r}{Z - f} = \frac{T}{Z}$$

$$Z = f\frac{T}{x_r - x_l}$$

disparity

Depth from disparity

image I(x,y)

Disparity map D(x,y)

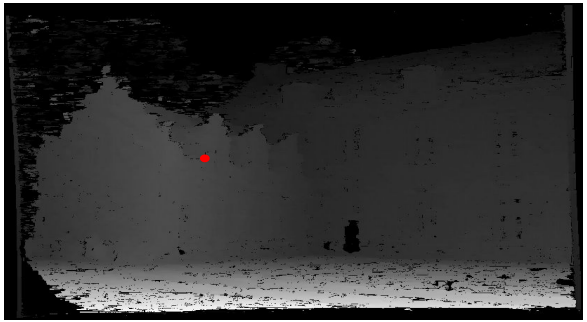


image I´(x´,y´)

$$(x´,y´)=(x+D(x,y),\ y)$$

Depth from disparity

image I(x,y)

Disparity map D(x,y)

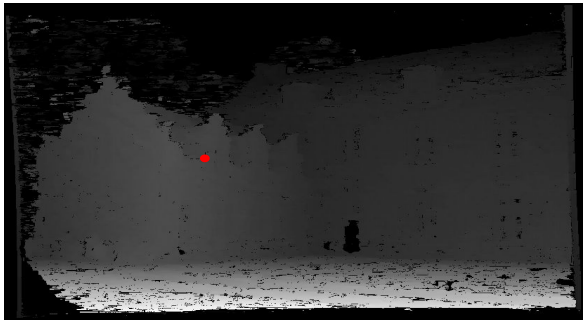
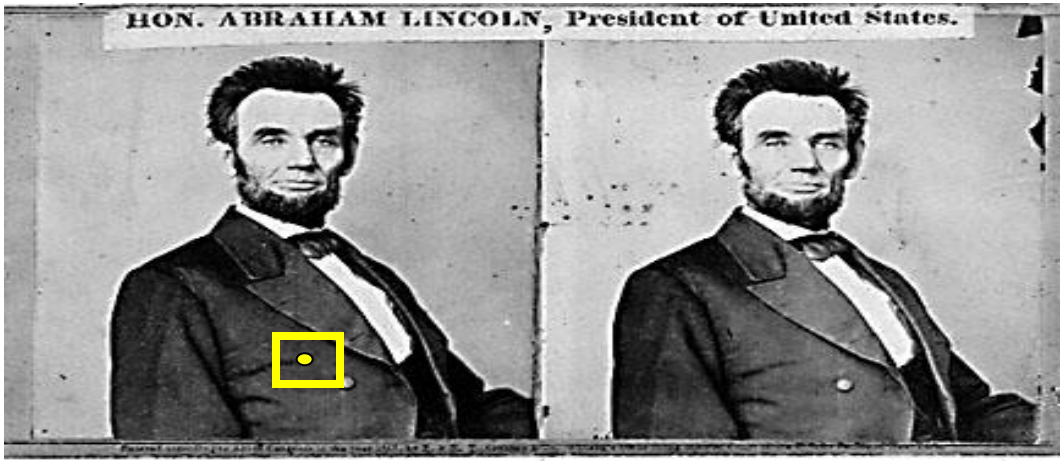


image I´(x´,y´)

$$(x´,y´)=(x+D(x,y), y)$$

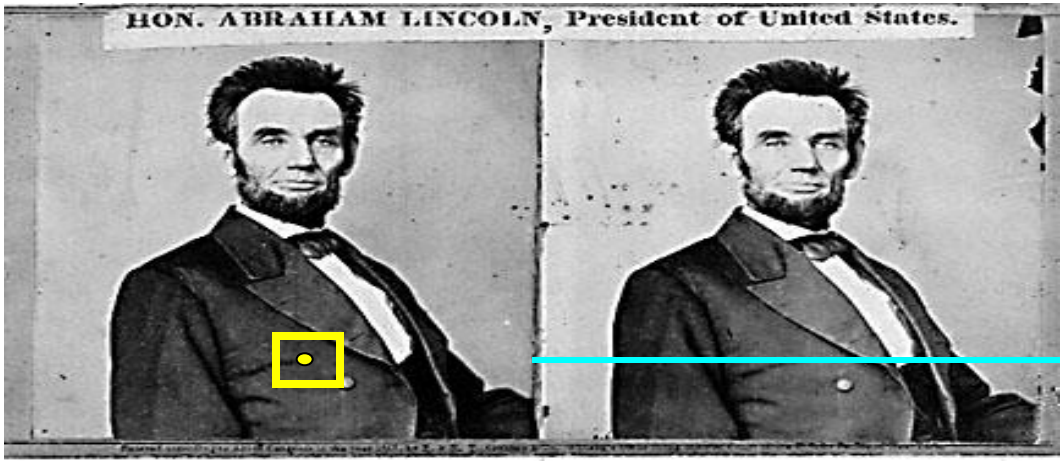
So if we could find the **corresponding points** in two images, we could **estimate relative depth**…

Basic stereo matching algorithm



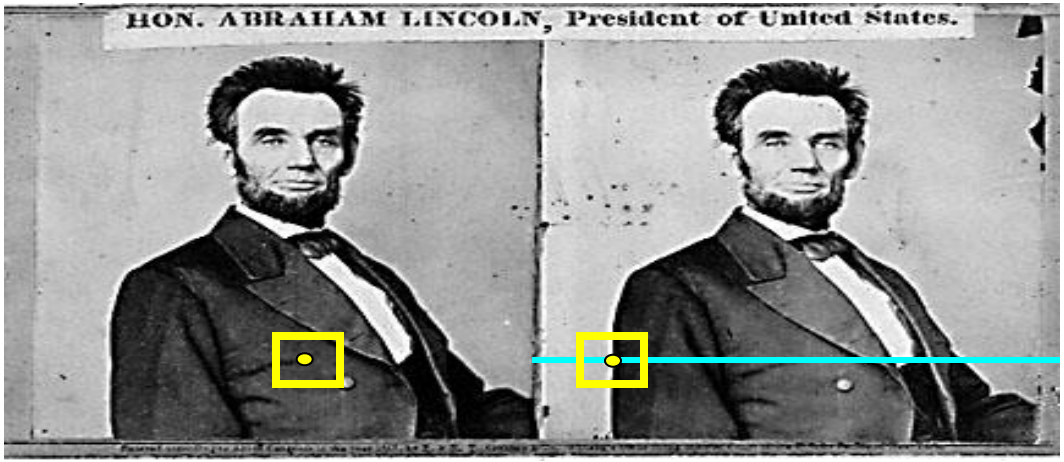
- If necessary, rectify the two stereo images to transform epipolar lines into scanlines
- For each pixel x in the first image
  - Find corresponding epipolar scanline in the right image
  - Examine all pixels on the scanline and pick the best match x'
  - Compute disparity x-x' and set depth(x) = fB/(x-x')

Basic stereo matching algorithm



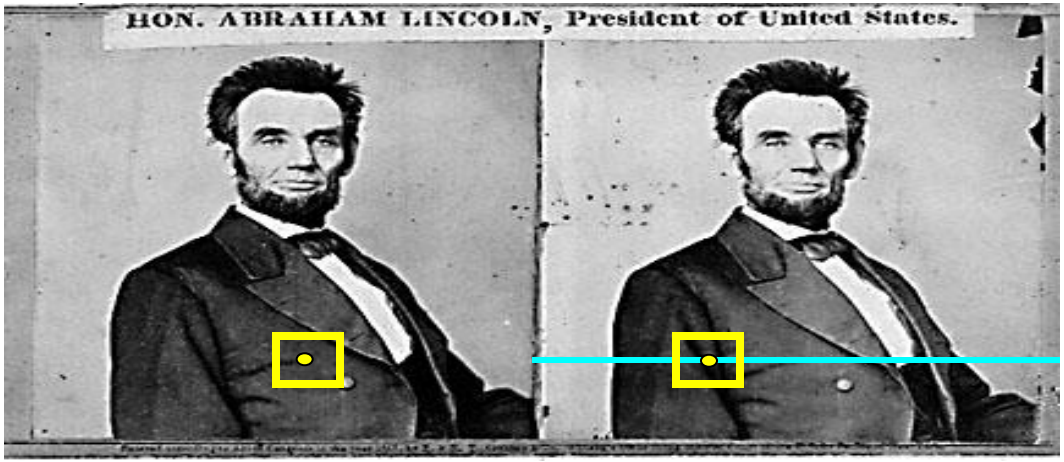
- If necessary, rectify the two stereo images to transform epipolar lines into scanlines
- For each pixel x in the first image
  - Find corresponding epipolar scanline in the right image
  - Examine all pixels on the scanline and pick the best match x’
  - Compute disparity x-x’ and set depth(x) = fB/(x-x’)

Basic stereo matching algorithm



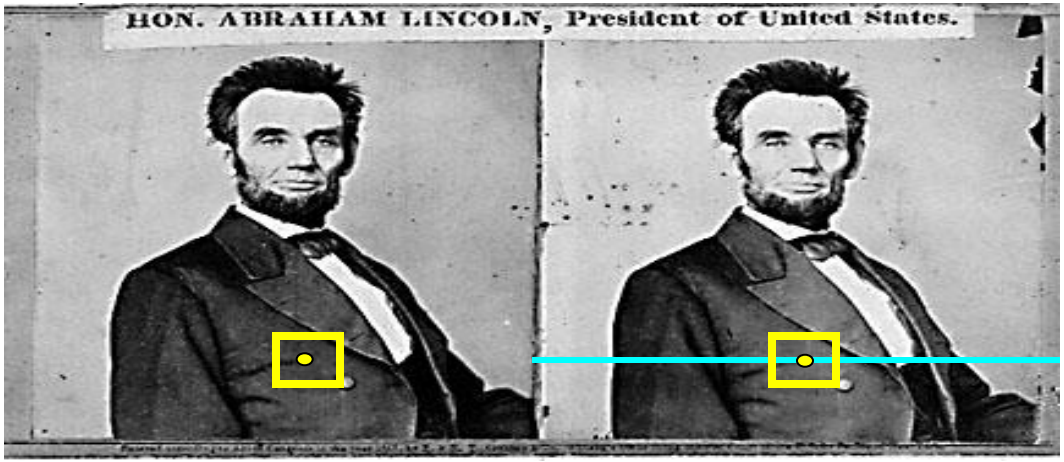
- If necessary, rectify the two stereo images to transform epipolar lines into scanlines
- For each pixel x in the first image
  - Find corresponding epipolar scanline in the right image
  - Examine all pixels on the scanline and pick the best match x'
  - Compute disparity x-x' and set depth(x) = fB/(x-x')

Basic stereo matching algorithm



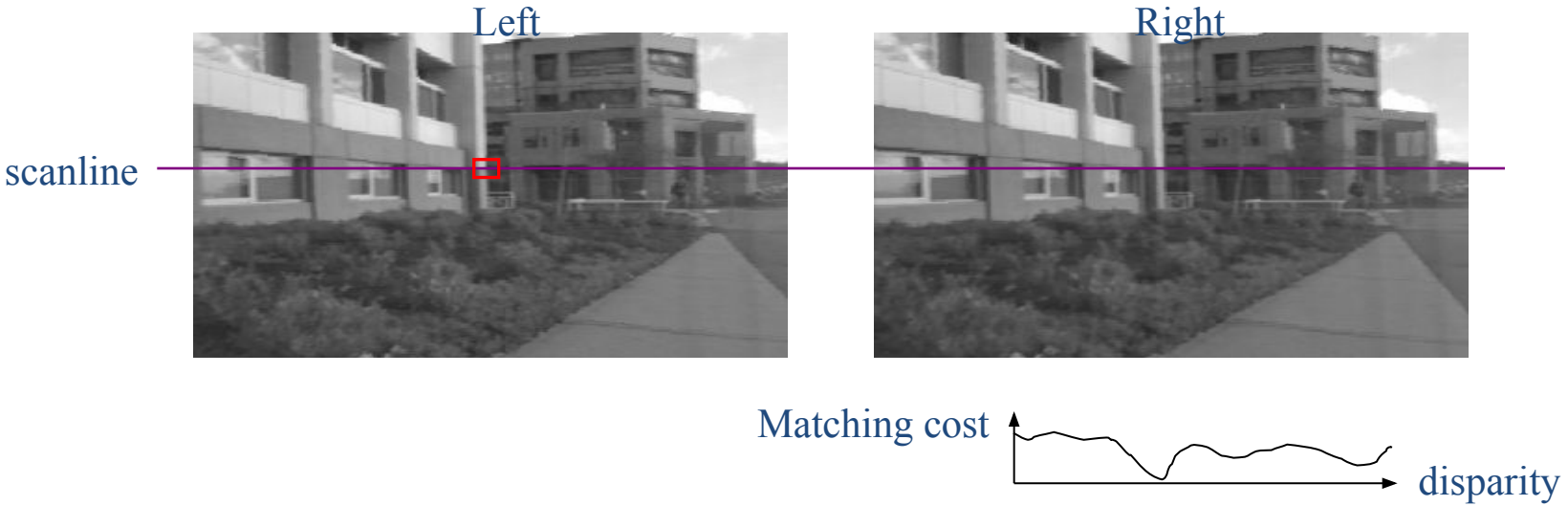
- If necessary, rectify the two stereo images to transform epipolar lines into scanlines
- For each pixel x in the first image
  - Find corresponding epipolar scanline in the right image
  - Examine all pixels on the scanline and pick the best match x'
  - Compute disparity x-x' and set depth(x) = fB/(x-x')

Basic stereo matching algorithm



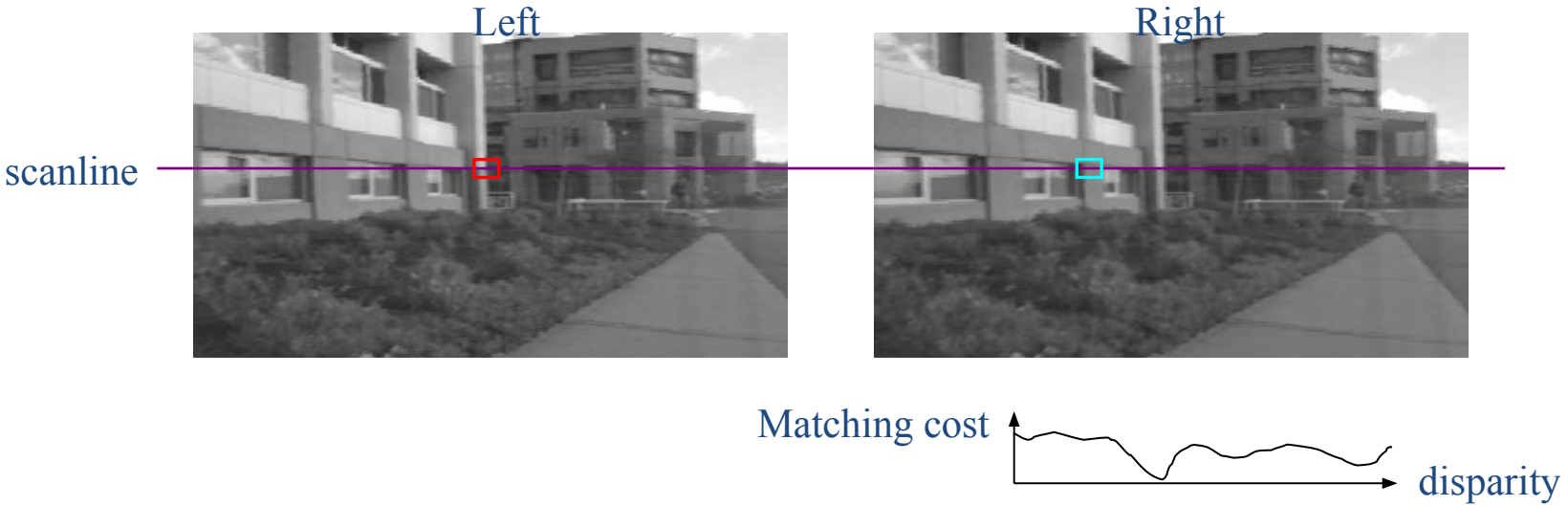
- If necessary, rectify the two stereo images to transform epipolar lines into scanlines
- For each pixel x in the first image
  - Find corresponding epipolar scanline in the right image
  - Examine all pixels on the scanline and pick the best match x’
  - Compute disparity x-x’ and set depth(x) = fB/(x-x’)

Correspondence search



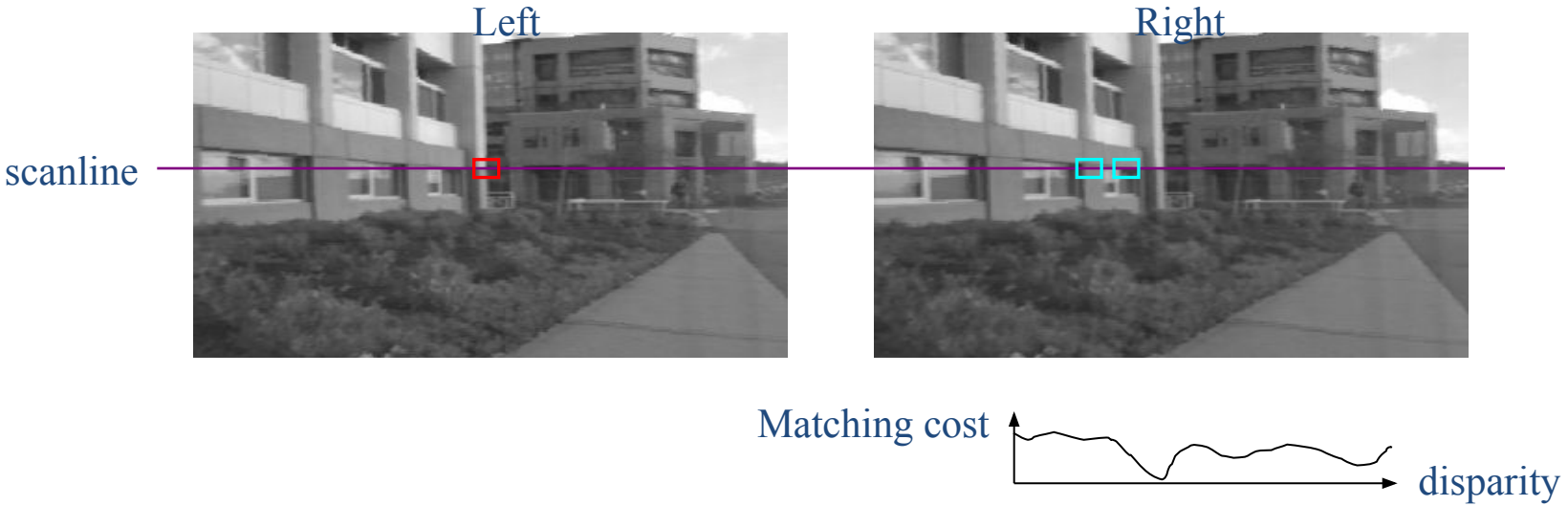
- Slide a window along the right scanline and compare contents of that window with the reference window in the left image
- Matching cost: SSD or normalized correlation

Correspondence search



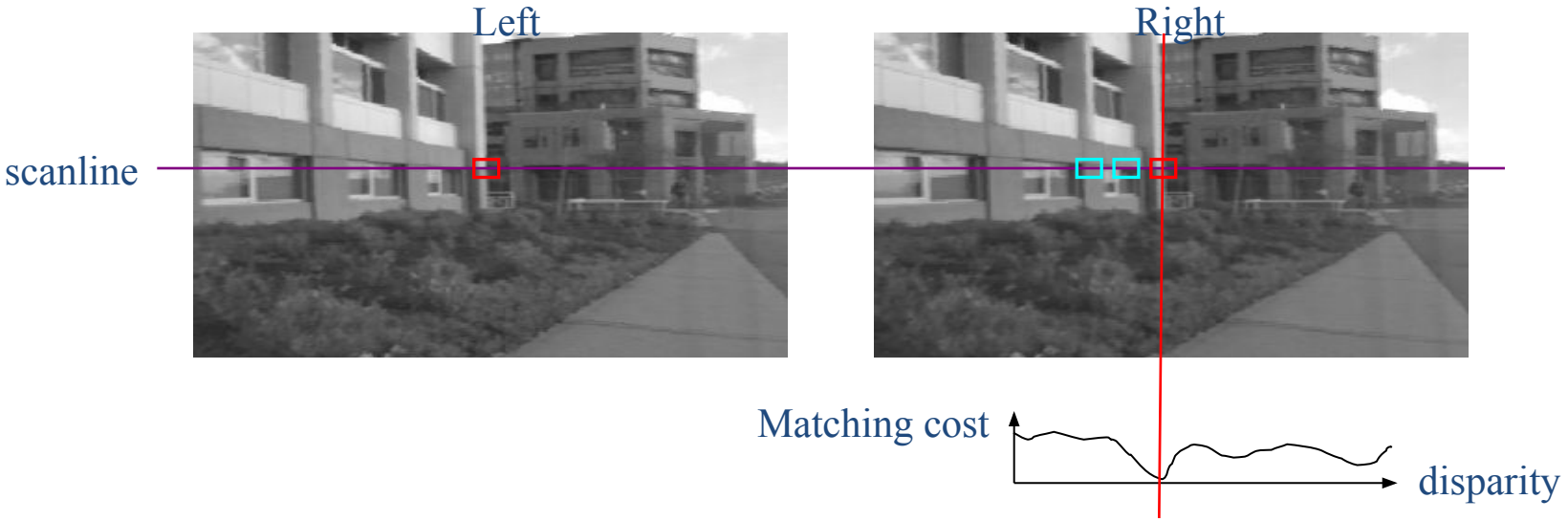
- Slide a window along the right scanline and compare contents of that window with the reference window in the left image
- Matching cost: SSD or normalized correlation

Correspondence search



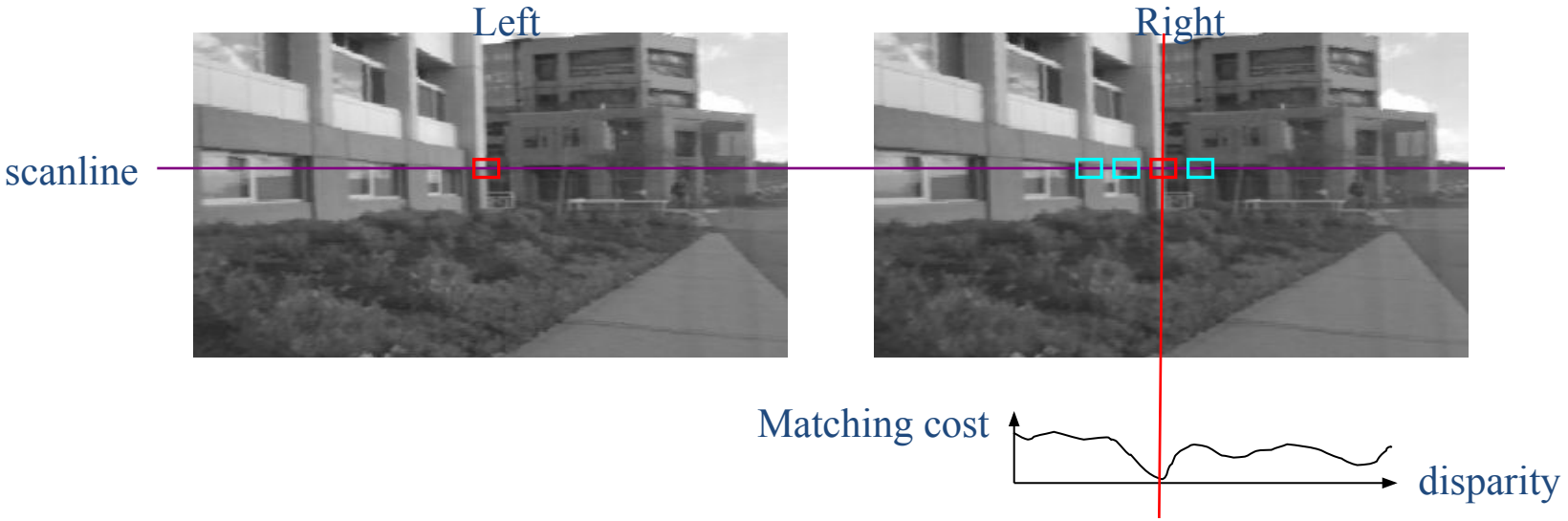
- Slide a window along the right scanline and compare contents of that window with the reference window in the left image
- Matching cost: SSD or normalized correlation

Correspondence search



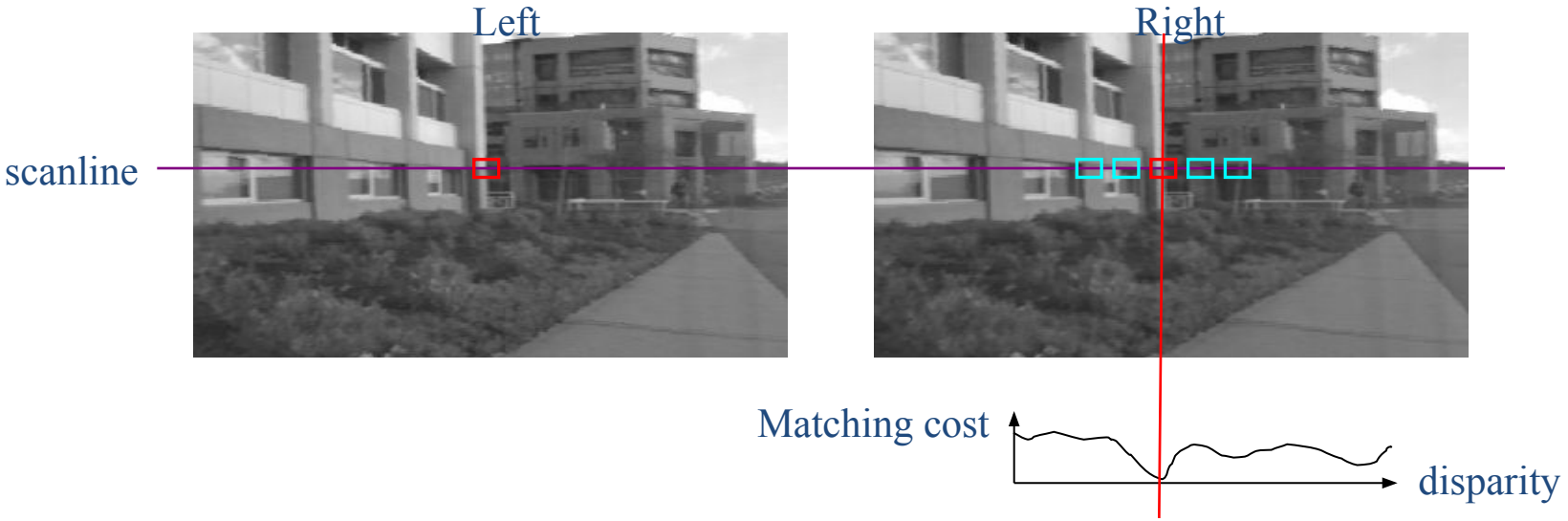
- Slide a window along the right scanline and compare contents of that window with the reference window in the left image
- Matching cost: SSD or normalized correlation

Correspondence search



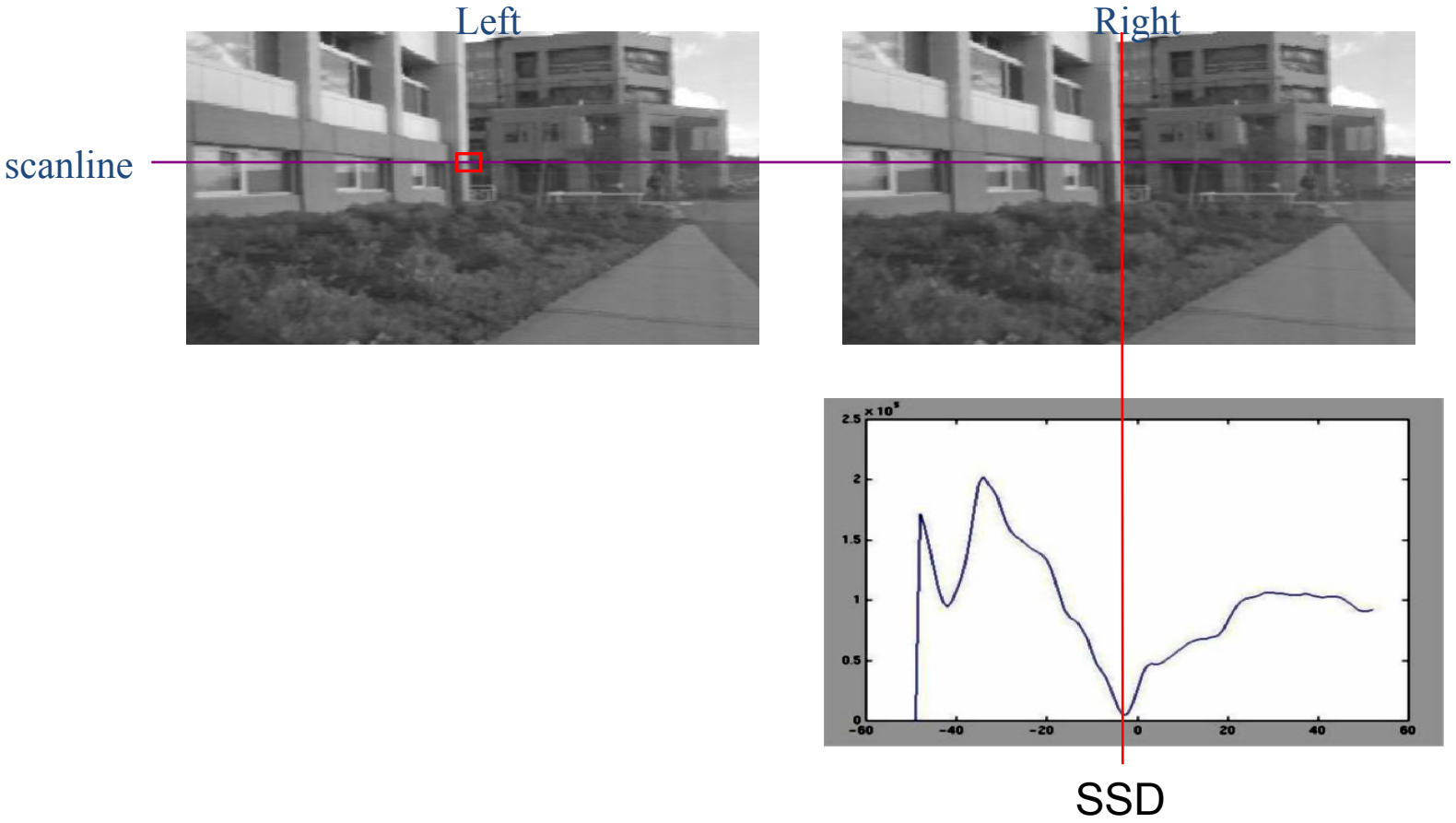
- Slide a window along the right scanline and compare contents of that window with the reference window in the left image
- Matching cost: SSD or normalized correlation

Correspondence search

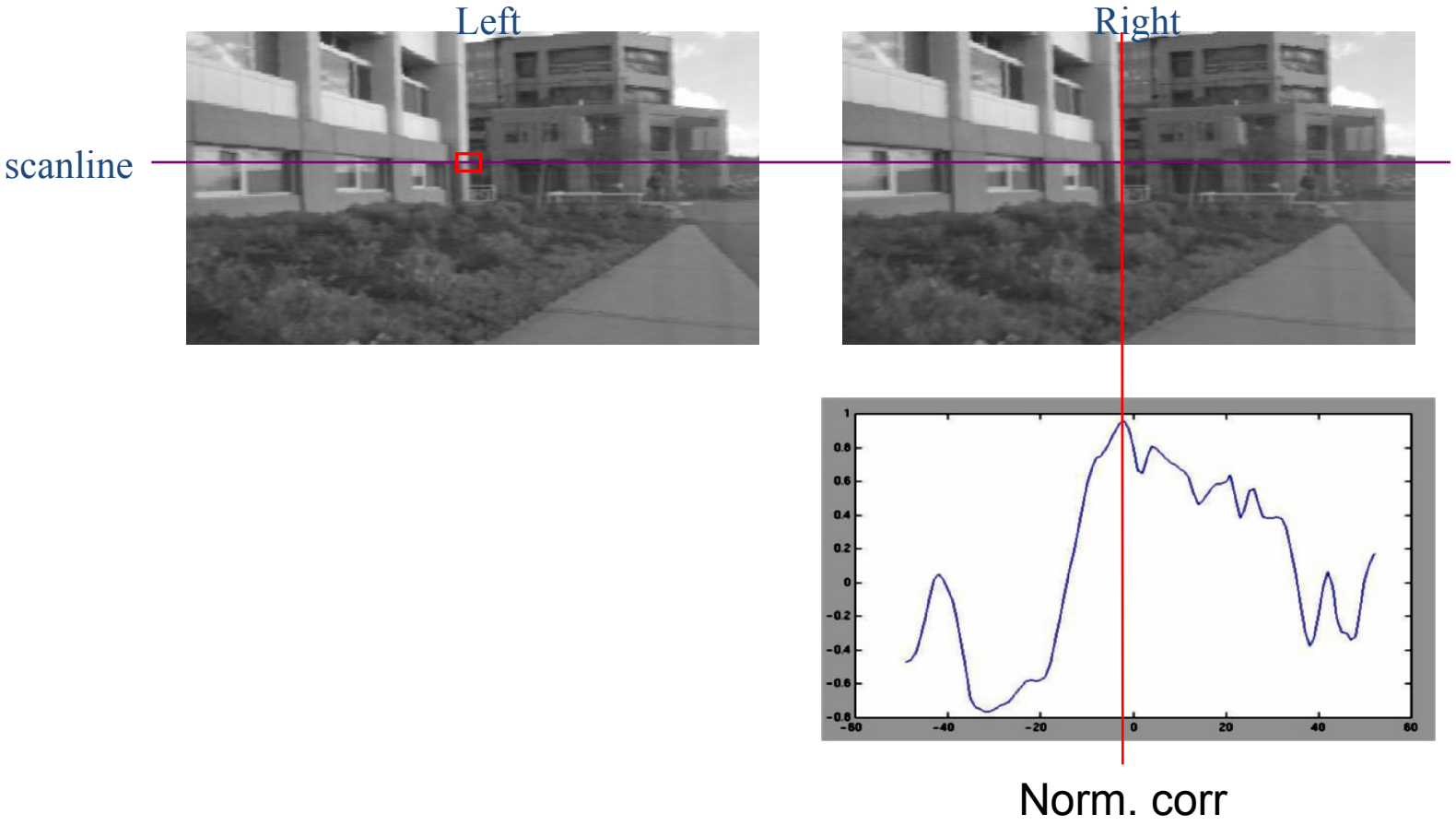


- Slide a window along the right scanline and compare contents of that window with the reference window in the left image
- Matching cost: SSD or normalized correlation

Correspondence search



Correspondence search

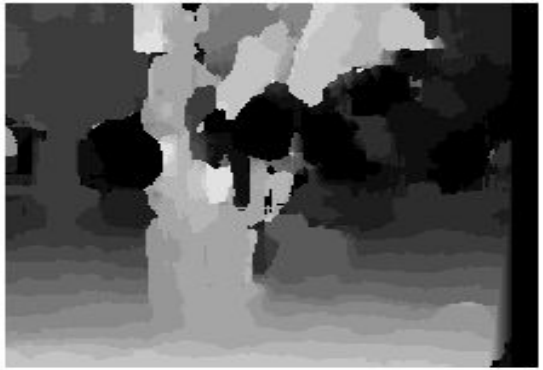


Effect of window size

Effect of window size

Effect of window size

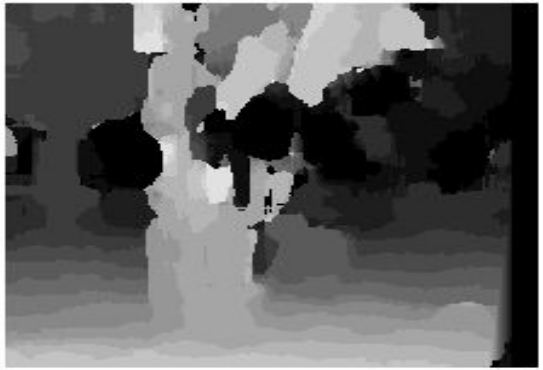
W = 3



W = 20

Effect of window size

W = 3

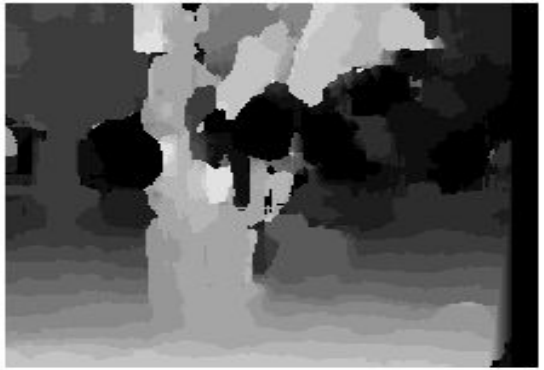


W = 20

- Smaller window
  - \+ More detail
  - – More noise

Effect of window size

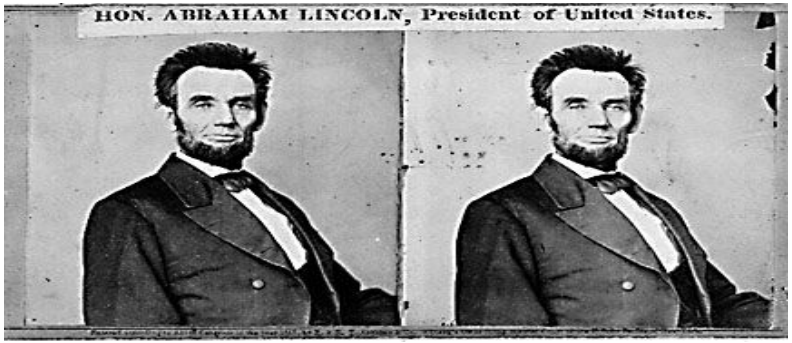
W = 3



W = 20

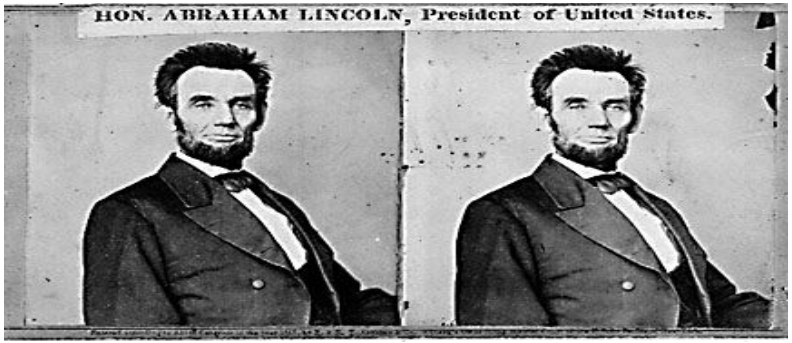
- Smaller window
  - \+ More detail
  - – More noise
- Larger window
  - \+ Smoother disparity maps
  - – Less detail

Failures of correspondence search



Textureless surfaces

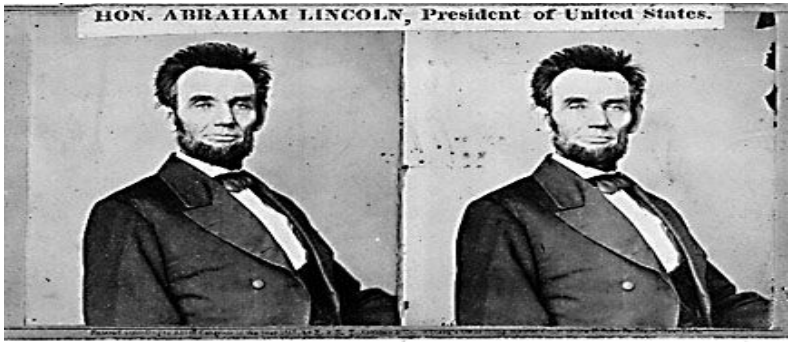
Failures of correspondence search



Textureless surfaces

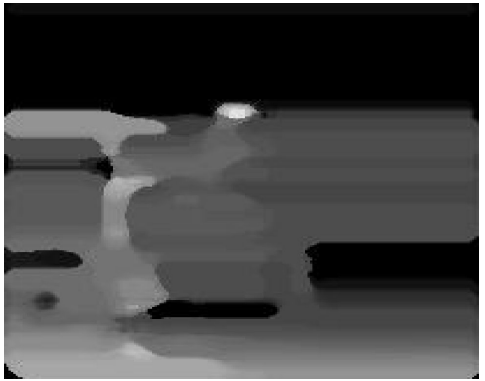
Occlusions, repetition

Failures of correspondence search



Textureless surfaces

Occlusions, repetition

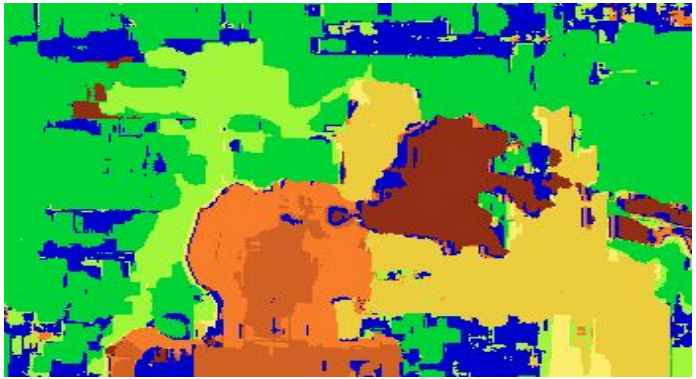


Non-Lambertian surfaces, specularities

Results with window search

Data

Window-based matching



Ground truth

How can we improve window-based matching?

- So far, matches are independent for each point
- What constraints or priors can we add?

Review

- Depth from stereo: main idea is to triangulate from corresponding image points.
- Epipolar geometry defined by two cameras
  - We've assumed known extrinsic parameters relating their poses
- Epipolar constraint limits where points from one view will be imaged in the other
  - Makes search for correspondences quicker
- **Terms**: epipole, epipolar plane / lines, disparity, rectification, intrinsic/extrinsic parameters, essential matrix, baseline

CONCORDIA.CA