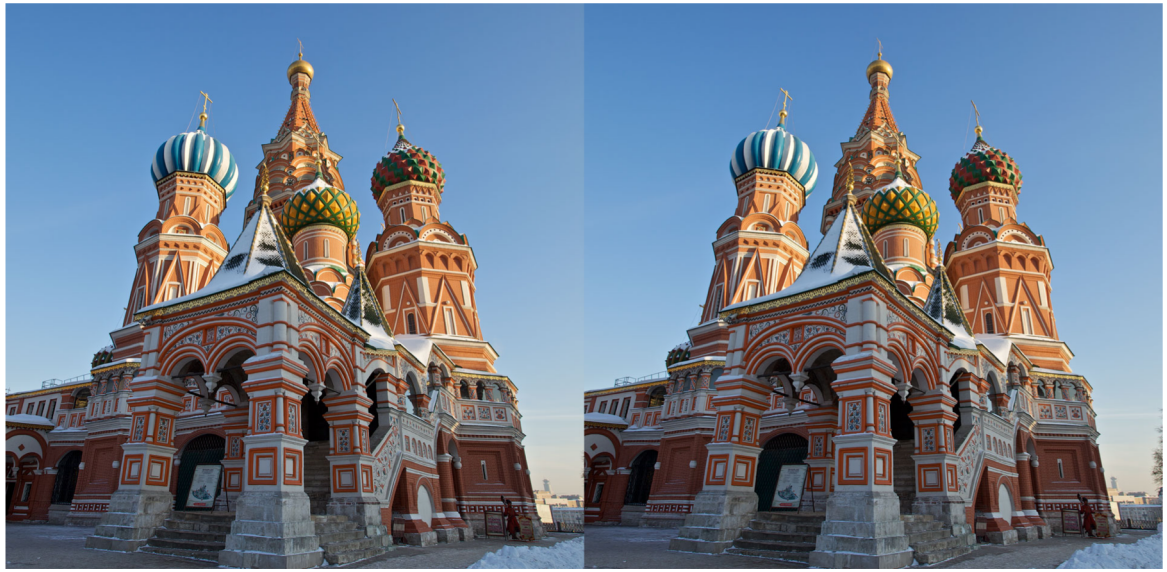
Robotic Perception
Depth Estimation

Yezhou Yang, Ph.D.
Assistant Professor
Zhiyuan Fang, Teaching Assistant
Arizona State University

Ira A. Fulton Schools of Engineering
Arizona State University

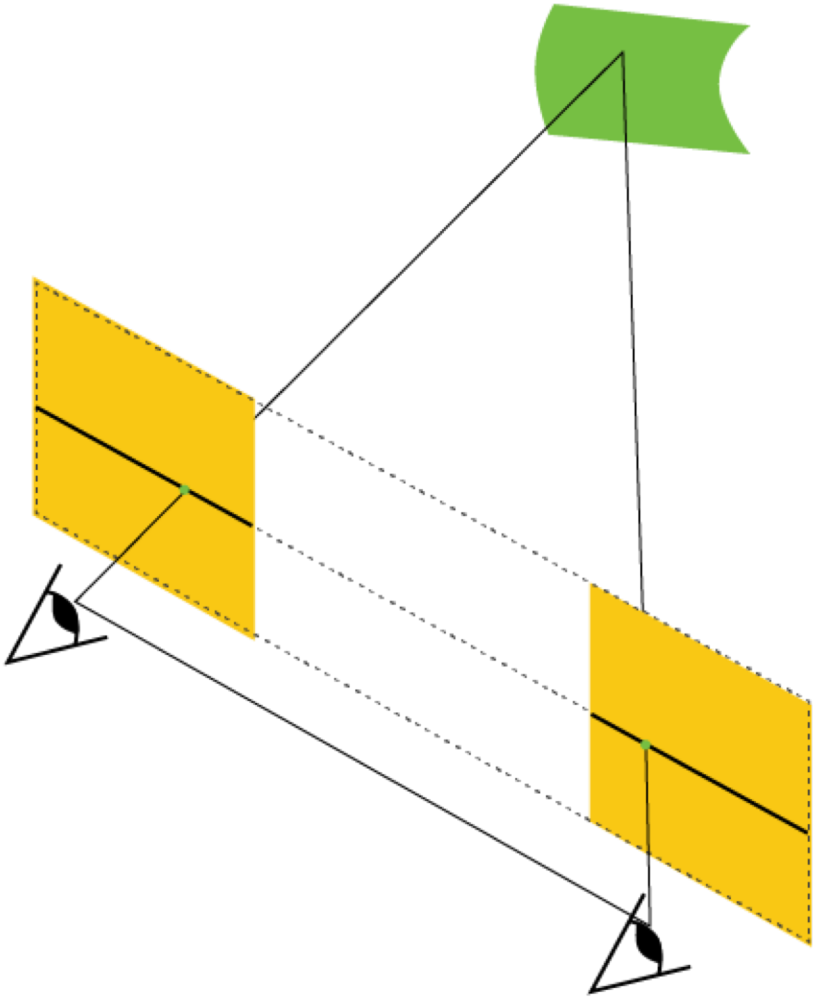
Problem Definition: Binocular Stereo

- **Given a calibrated binocular stereo pair, fuse it to produce a depth image.**
  - Assumes (two) cameras.
  - Known positions.
  - Recover depth.



“stereo-pairs for cross eyed or parallel viewing methods,” urixblog.com. [Online]. Available: http://urixblog.com/en/html5-stereo-viewer-en/. [Accessed: 21-Oct-2014].

Simplest Case: Parallel Images

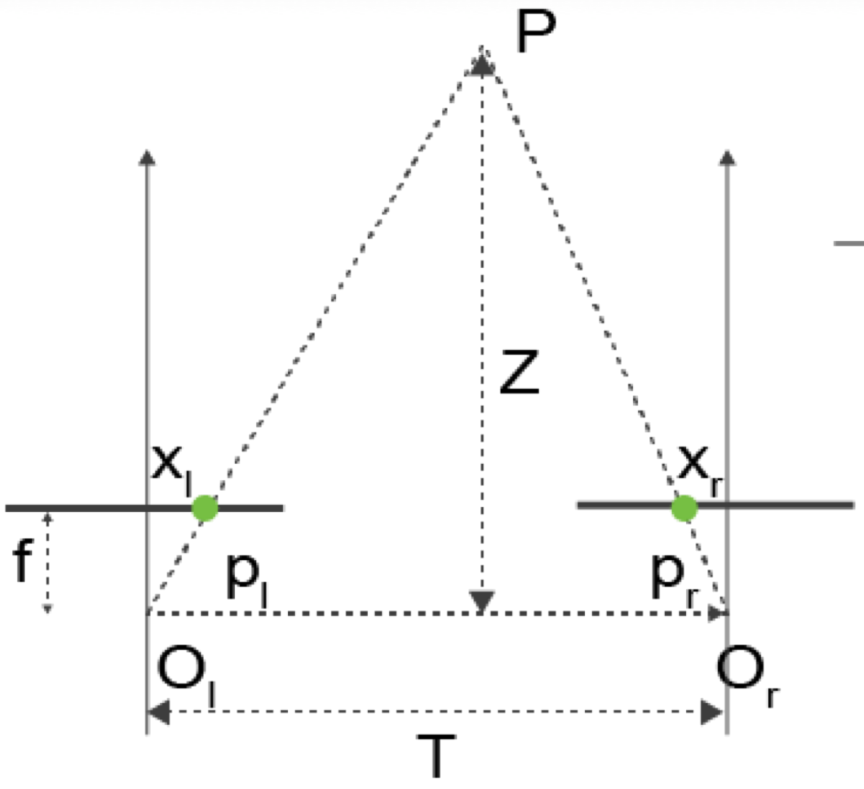


Parallel Images

- Image planes of cameras are to each other and to the baseline.
- Camera centers are at same height.
- Focal lengths are the same.
- Then epipolar lines fall along the horizontal scan lines of the images.

Image modified from: “Epipolar Geometry and Stereo Vision - ppt download,” SlidePlayer. [Online]. Available: https://slideplayer.com/slide/12649237/. [Accessed: 10-May-2019].

Basic Math



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

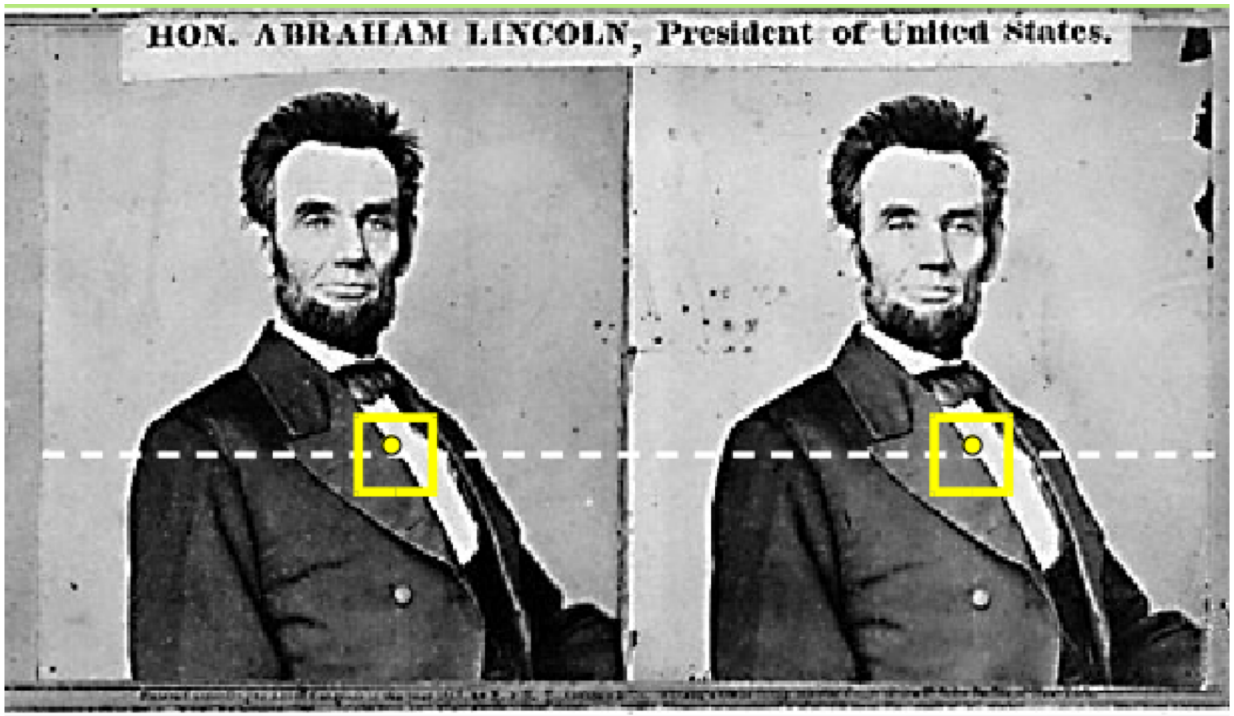
Disparity: $\boldsymbol{d = x_l - x_r}$

$$Z = f\frac{T}{d}$$

Then given T and d, we can compute Z

- **T** is the stereo baseline
- **d** measures the difference in retinal position between corresponding points

Window Matching



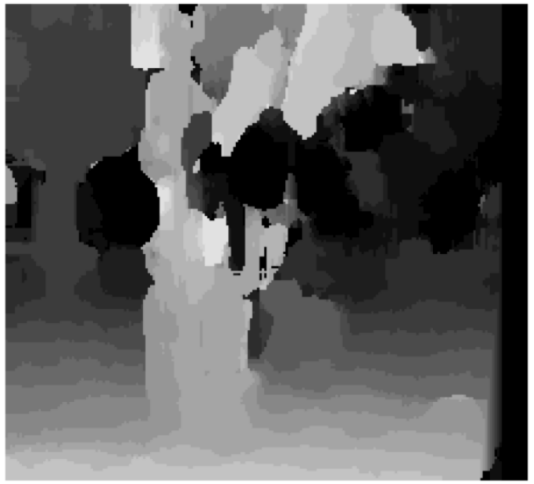
Match Windows (image patches)

Effect of Window Size

W = 3

- **Smaller window**
  - More detail
  - More noise

Effect of Window Size

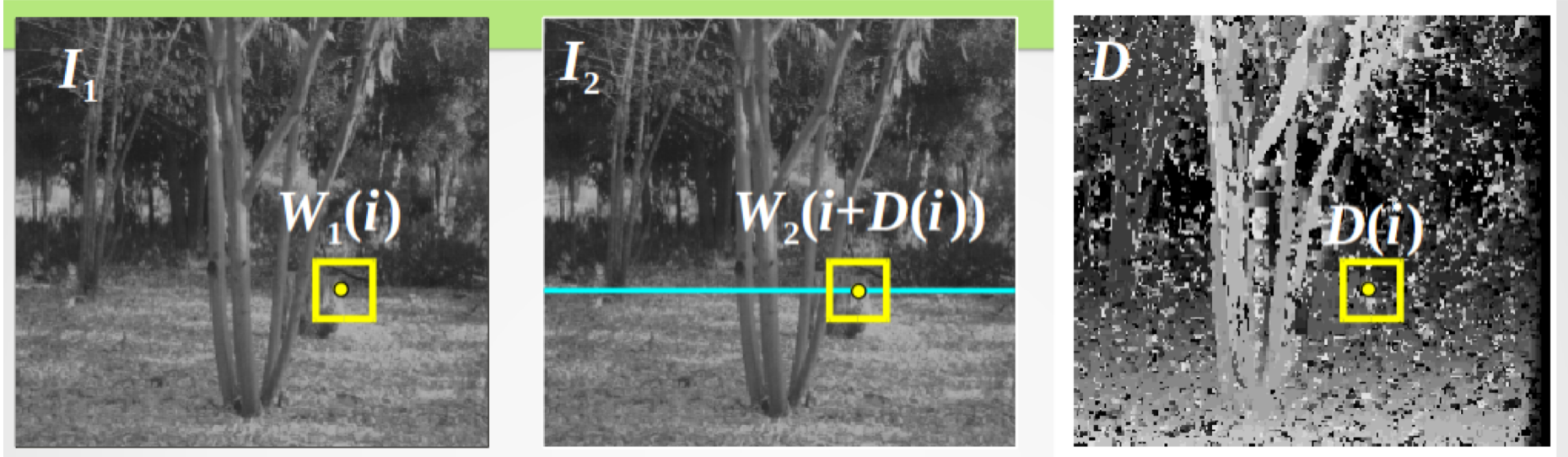


W = 10

Larger window

- Smoother disparity maps
- Less detail

Stereo Matching as Energy Minimization



$$E(D) = \sum_{i} \left(W_1(i) - W_2(i + D(i))\right)^2 + \lambda \sum_{\text{neighbors } i,j} \rho\left(D(i) - D(j)\right)$$

data term — trade-off param — smoothness term

Y. Boykov, O. Veksler, and R. Zabih, Fast Approximate Energy Minimization via Graph Cuts, PAMI 2001