Stereo Vision

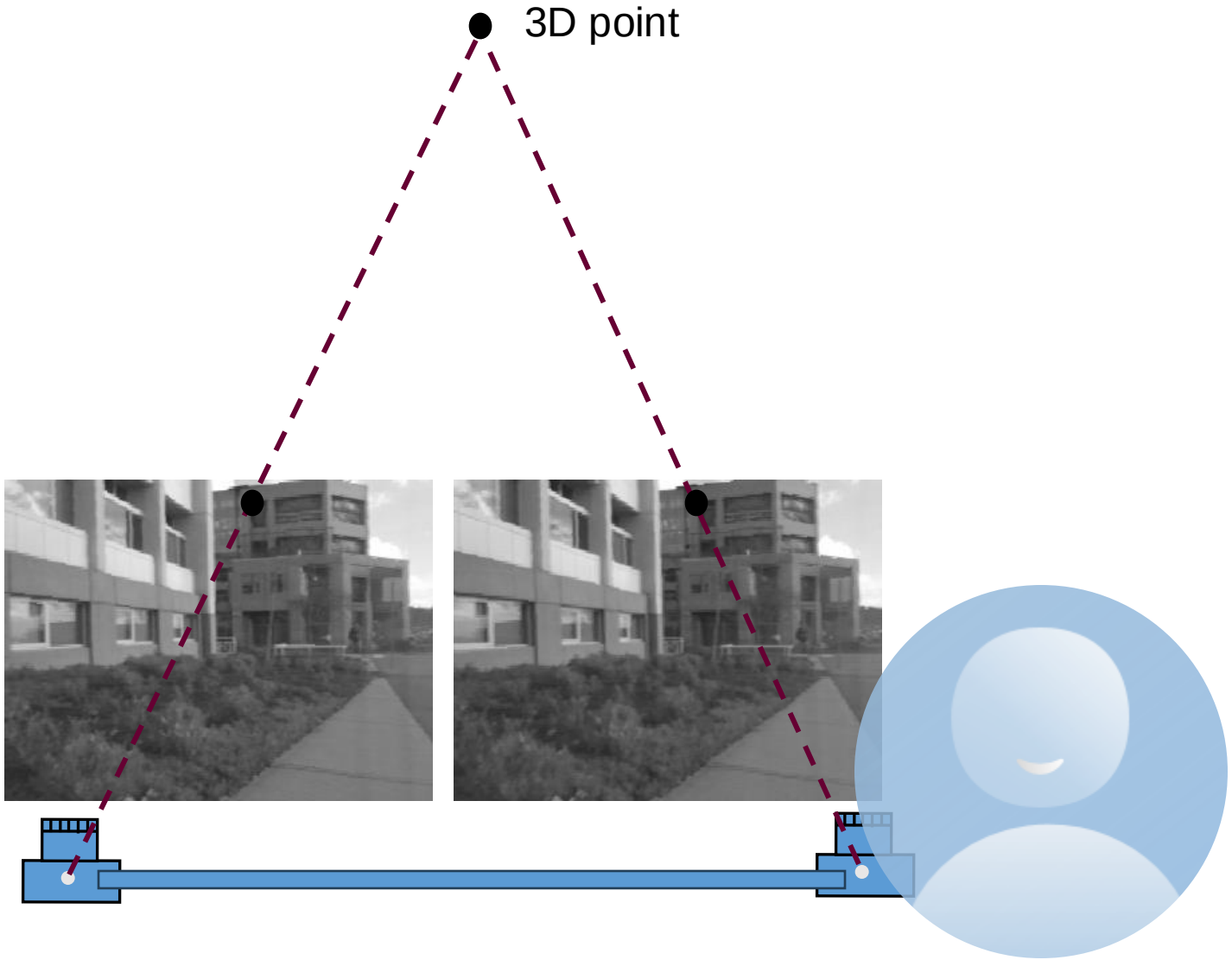
CS 3630

Humans use stereo vision

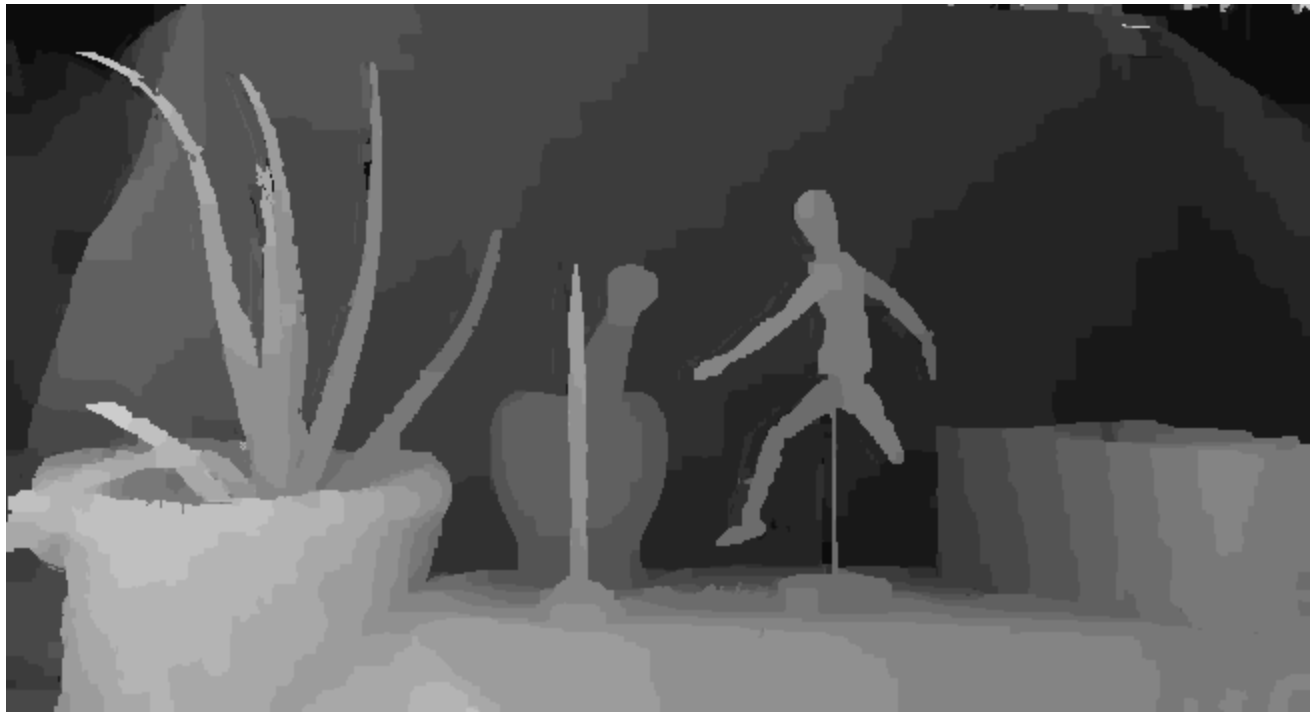
Very useful in computer vision as well as it eliminates scale ambiguity

Effect of Moving Camera

- As camera is shifted (viewpoint changed):
  - 3D points are projected to different 2D locations
  - Amount of shift in projected 2D location depends on depth
- 2D shifts= stereo disparity



Example

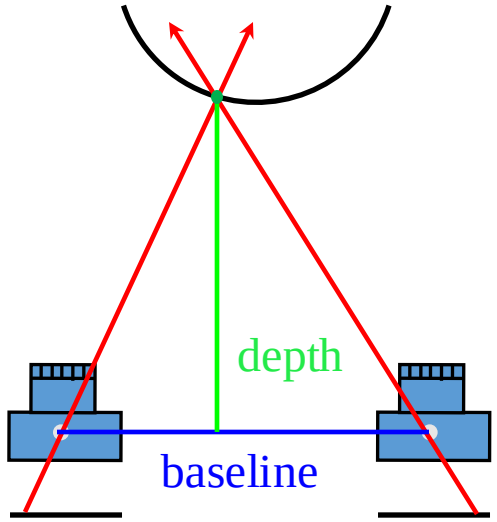


Right image

Left image

View Interpolation

Basic Idea of Stereo



Triangulate the same point on two images to recover depth.

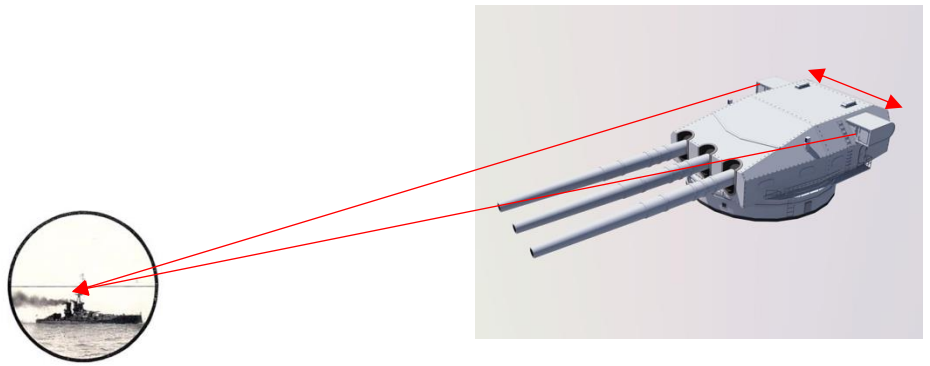
Requires:

- Feature matching across views
- Calibrated cameras

Matching correlation windows across scan lines

Why is Stereo Useful?

- Passive and non-invasive
- Robot navigation (path planning, obstacle detection)
- 3D modeling (shape analysis, reverse engineering, visualization)
- Photorealistic rendering

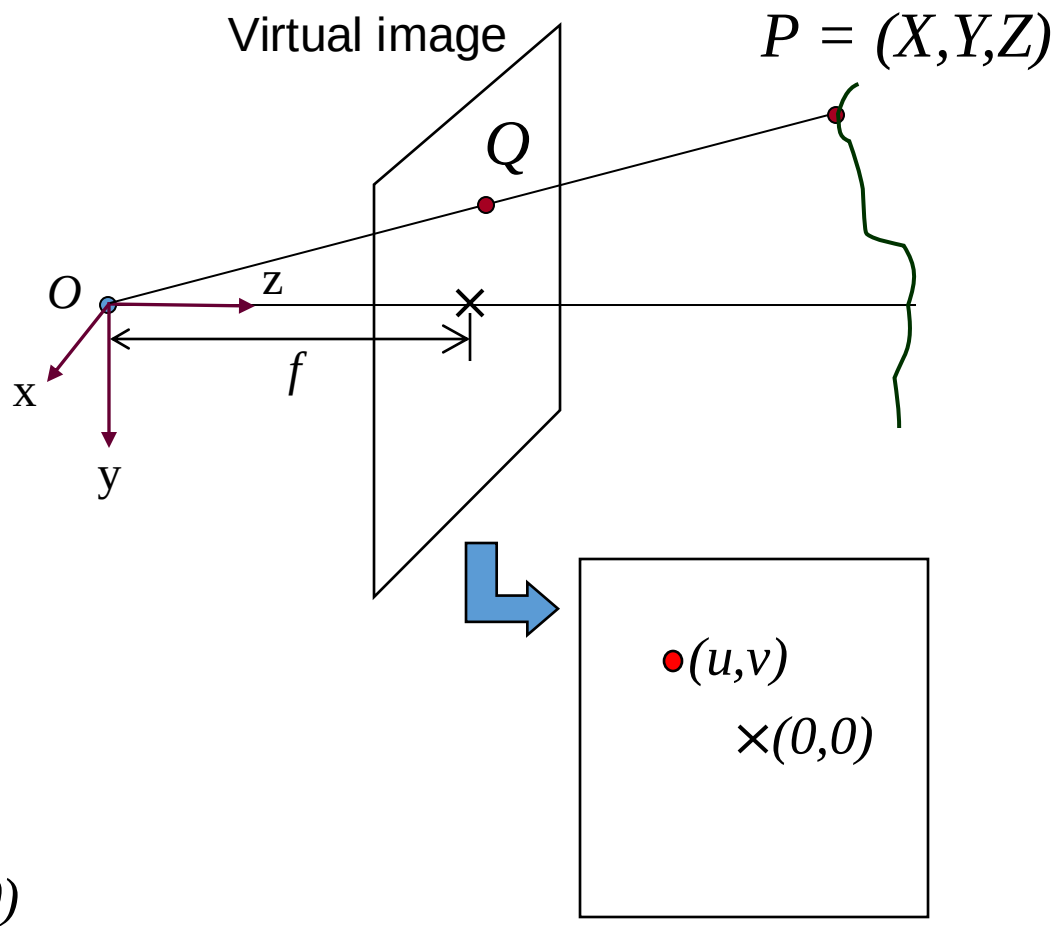


Stereo Geometry

- Recall: Pinhole model
- Now we have two！
- How to recover depth from two measurements?

Review: Pinhole Camera Model

3D scene point P is projected to a 2D point Q in the virtual image plane

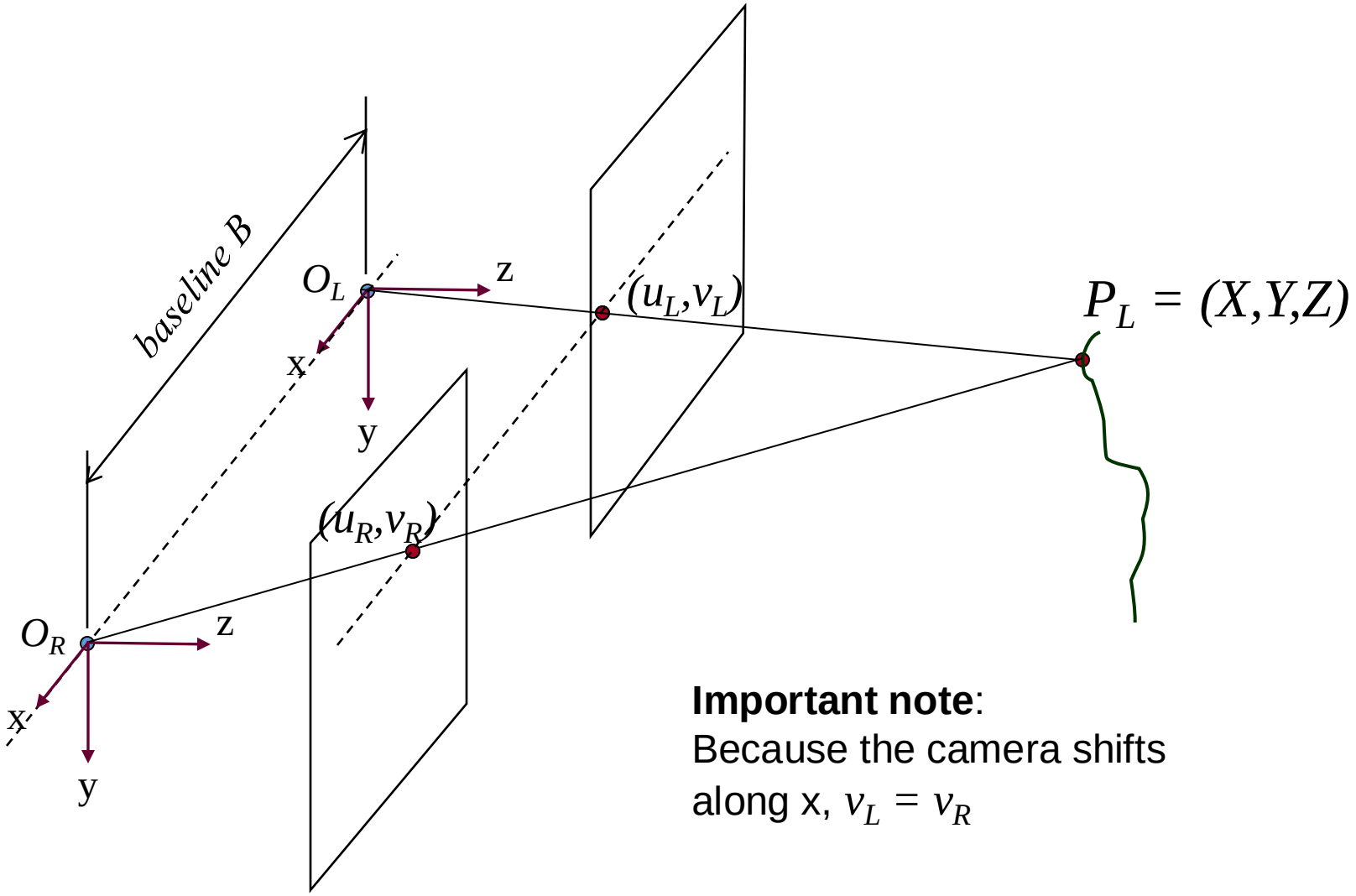


The 2D coordinates in the image are given by

$$(u, v) = \left(F\frac{X}{Z}, F\frac{Y}{Z}\right)$$

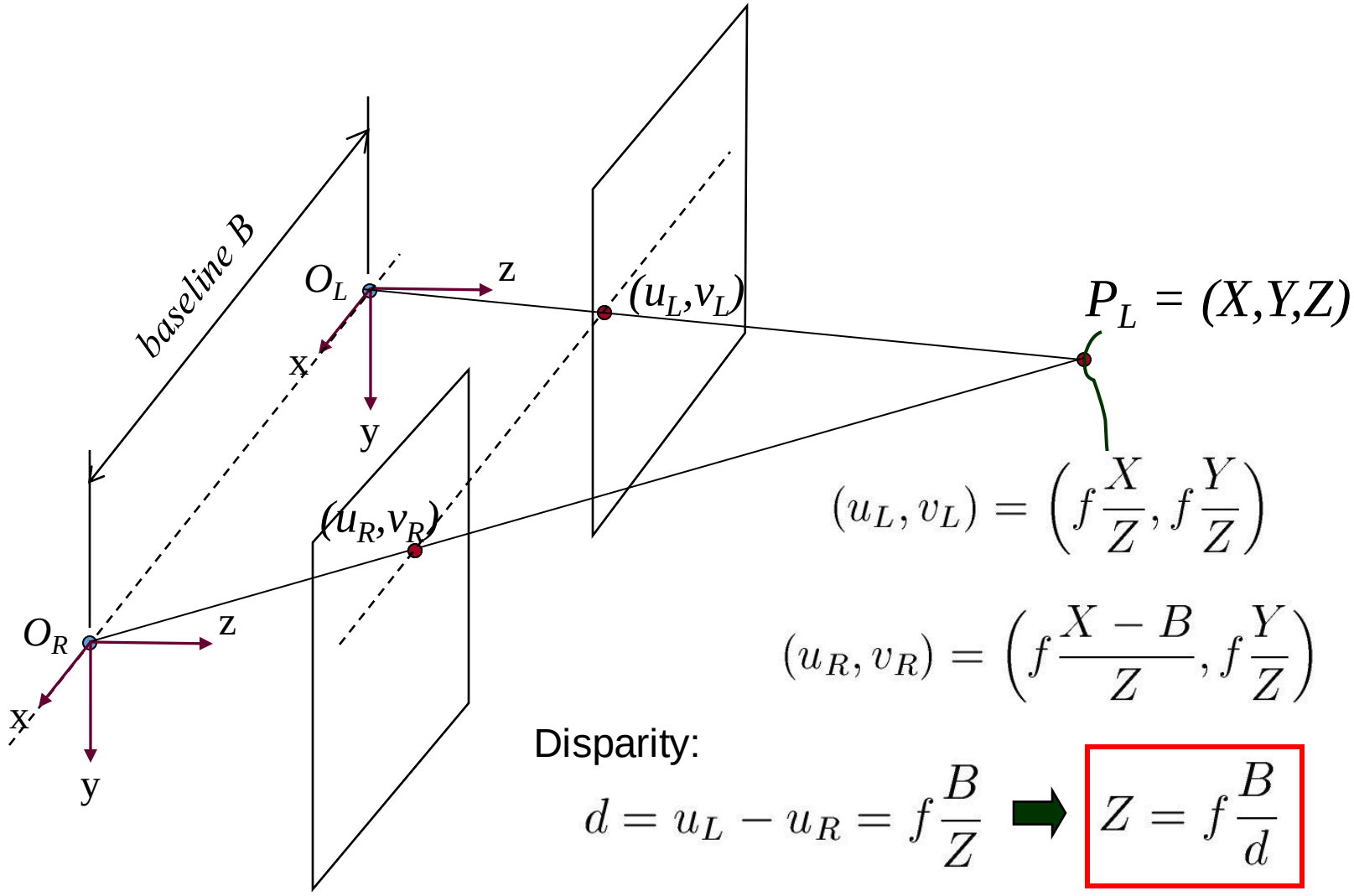
Note: image center is *(0,0)*

Basic Stereo Derivations



Important note:
Because the camera shifts along x, $v_L = v_R$

Basic Stereo Formula



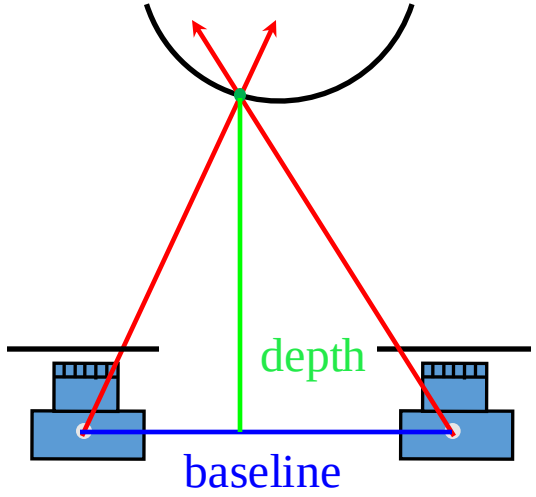
$$(u_L, v_L) = \left(f\frac{X}{Z}, f\frac{Y}{Z}\right)$$

$$(u_R, v_R) = \left(f\frac{X-B}{Z}, f\frac{Y}{Z}\right)$$

Disparity:

$$d = u_L - u_R = f\frac{B}{Z} \Rightarrow \boxed{Z = f\frac{B}{d}}$$

6. Stereo Algorithm



$$Z(x,y) = \frac{f\,B}{d(x,y)}$$

$Z(x, y)$ is depth at pixel (x, y)
$d(x, y)$ is disparity

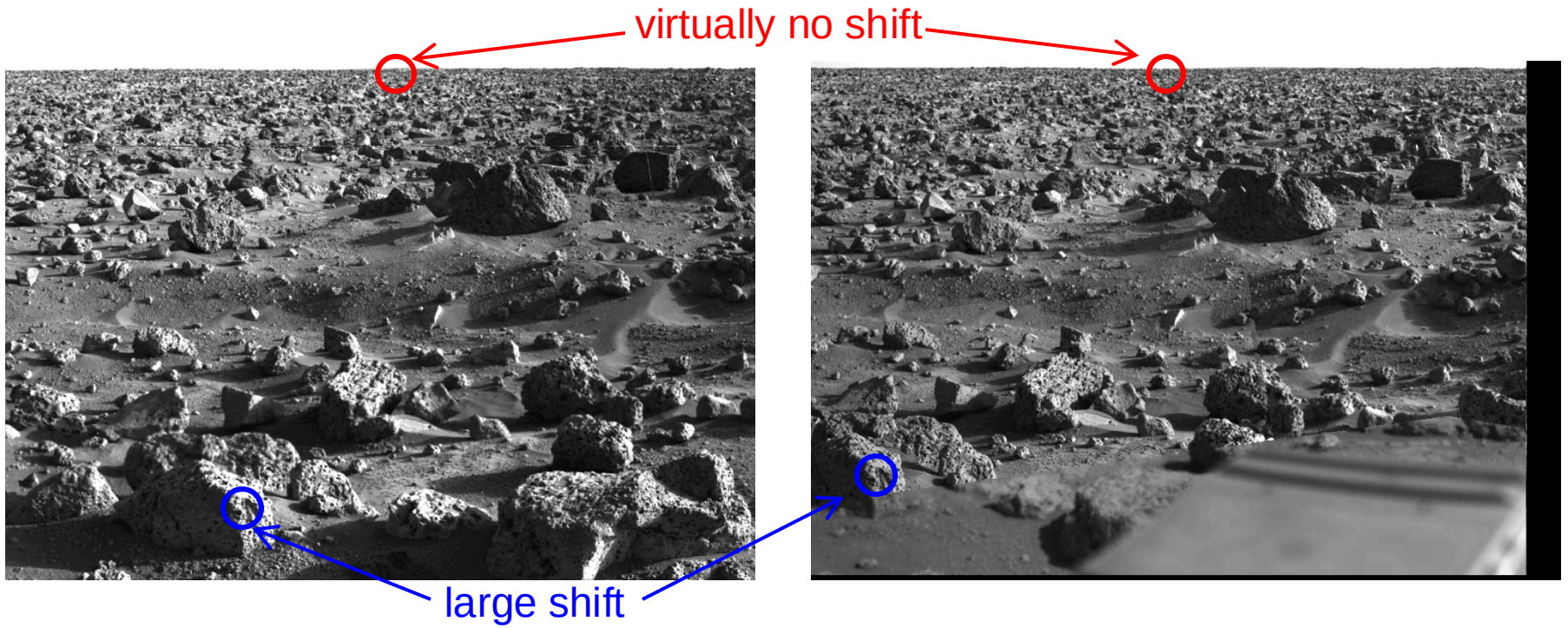
Left

Right

Matching correlation windows across scan lines

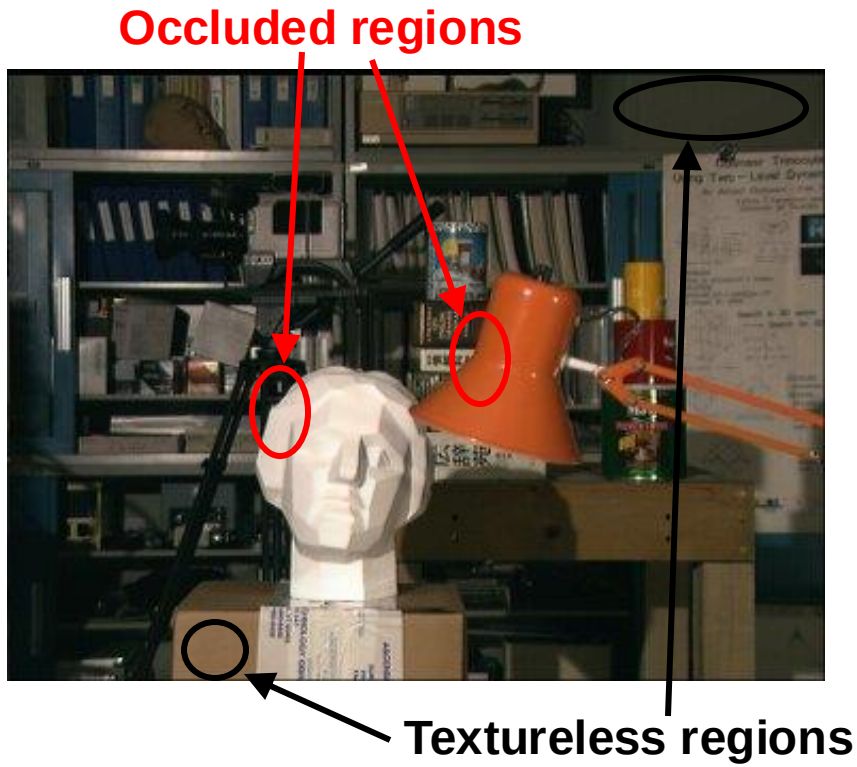
Stereo Correspondence

- Search over disparity to find correspondences
- Range of disparities can be large



Challenges

- Textureless regions create ambiguities
- Occlusions result in missing data
- Repetitive patterns can lead to false matches
- Very sensitive to calibration (mechanical vibrations and temperature affect camera alignment)
- Significant computational cost to processing stereo images



Stereo Cameras

- Mimics human binocular vision
- Implemented by using two cameras mounted a fixed and known distance apart. The two cameras capture images of the same scene from slightly different positions.
- Stereo cameras capture 3D information by comparing the differences between the two images (disparity)
  - When an object is captured by both cameras, it appears at slightly different horizontal positions in each image
  - The closer an object is, the larger this positional difference (disparity) will be
  - Using triangulation and knowing the camera parameters, the system can calculate the actual distance to each point
- The benefit of providing depth information is somewhat outweighed by computational costs for image processing and by sensitivity to calibration.
- LIDAR offers a compelling alternative!