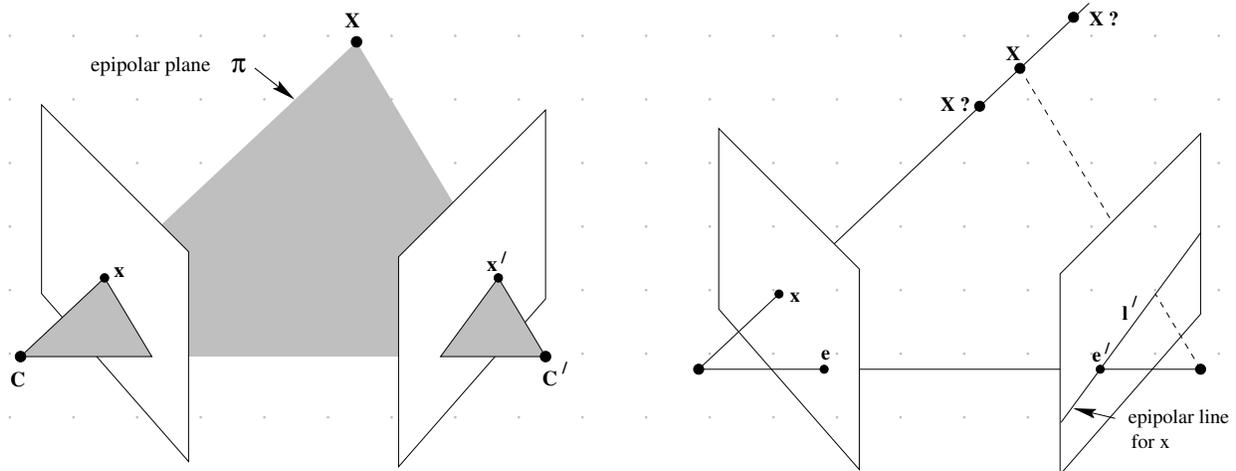


The Epipolar Constraint

A point \mathbf{X} in 3D projects to pixel \mathbf{p} in the left image and pixel \mathbf{p}' in the right image. Given \mathbf{p} , where might \mathbf{p}' be?



Claim: The correspondence search is _____-dimensional, not _____-dimensional.

The epipolar constraint. Figure from Hartley & Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed., Cambridge University Press, 2004.

Epipolar Lines in Real Images

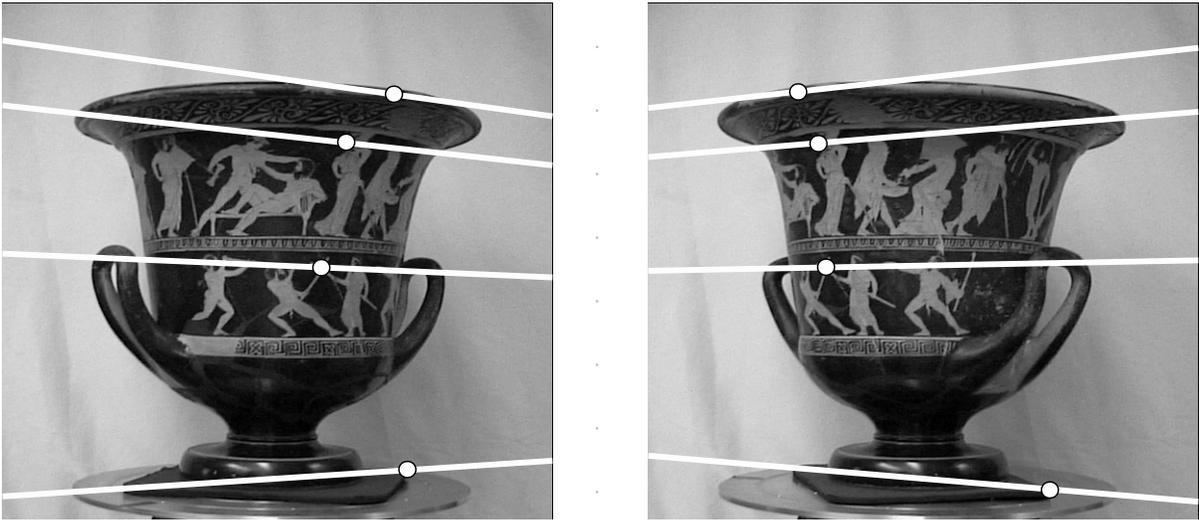


Figure 1: Epipolar lines on a stereo pair. Each point in one image corresponds to a line in the other. Figures from Hartley & Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed., Cambridge University Press, 2004.

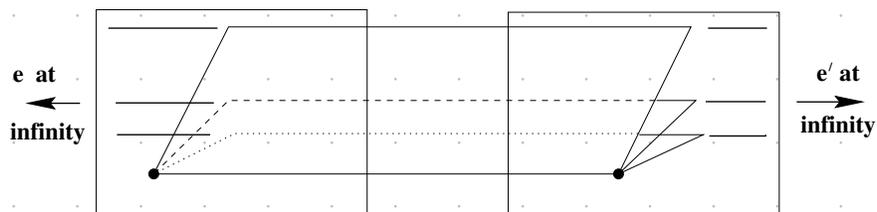
Rectified Stereo

A **rectified stereo pair** is a special configuration:

- Two cameras with parallel optical axes
- Separated by a horizontal baseline B
- Same focal length f and image size

Draw the setup (top-down view):

Describe the epipolar lines in rectified stereo:



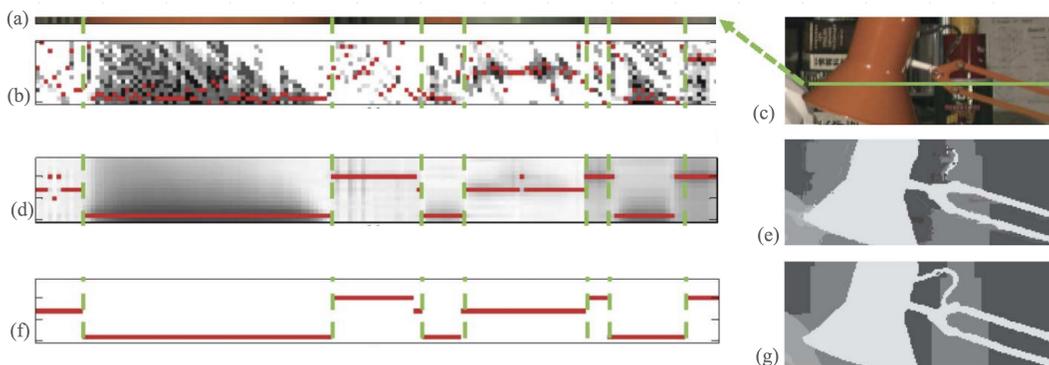
Disparity

Definition: For a pixel (x, y) in the left image, the **disparity** d tells us where the corresponding pixel is in the right image:

Left pixel (x, y) \longleftrightarrow Right pixel $(\text{_____}, \text{_____})$

Draw a diagram showing disparity for a single 3D point:

Aside: if we wrap this idea with enough hacks, we can compute disparity without any machine learning! (OpenCV implements this as `StereoBM` and `StereoSGBM` [link]).



Issues with intensity-based stereo matching. (a) One line of one image, (c), of stereo pair. (b) Intensity matching error for different disparities, plotted vertically, as a function of horizontal position. Red shows disparities of best intensity match. (d) Smoothed matching errors, with best matches shown in red, and resulting depth image, (e). (f) Ground truth matches and depth (g). Dataset from Scharstein & Szeliski (2002). Figure from Torralba, Isola, & Freeman, *Foundations of Computer Vision*, MIT Press, 2024. (CC-BY-NC-ND).

Photometric Consistency

For the depth2 project, we train a network to predict disparity from a *single* image, but we use *stereo pairs* for supervision.

Core idea: If the predicted disparity is correct, we can reconstruct the left image from the right image.

The warping operation:

Given disparity map D (defined on the left image), synthesize the left image from the right:

$$\hat{I}_L(x, y) = I_R(\text{-----}, \text{-----})$$

Draw a system block diagram for training depth prediction from stereo pairs using photometric consistency:

The Photometric Loss

Training signal: Compare the synthesized left image to the actual left image.

L1 loss:

$$\mathcal{L} = \underline{\hspace{10em}}$$

Limitations

What can go wrong with photometric consistency?

Extensions (smoothness loss, SSIM, left-right consistency) are covered in the project documentation.