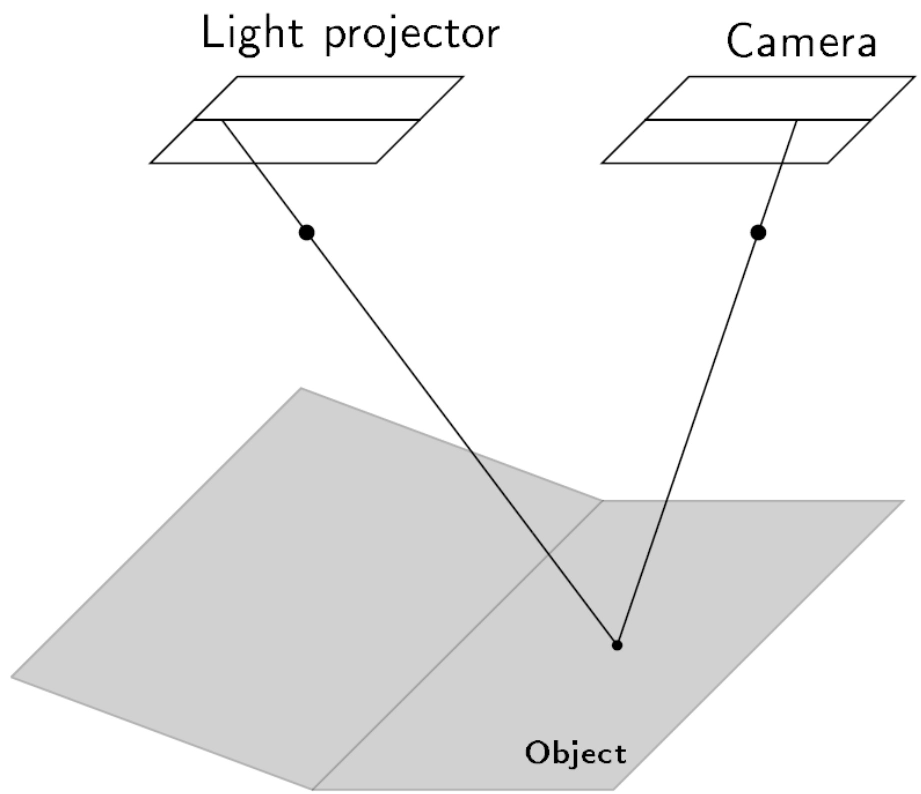
It’s all about depth/range

- Lidar (laser scanners)
- RGB-D
- Stereo Vision

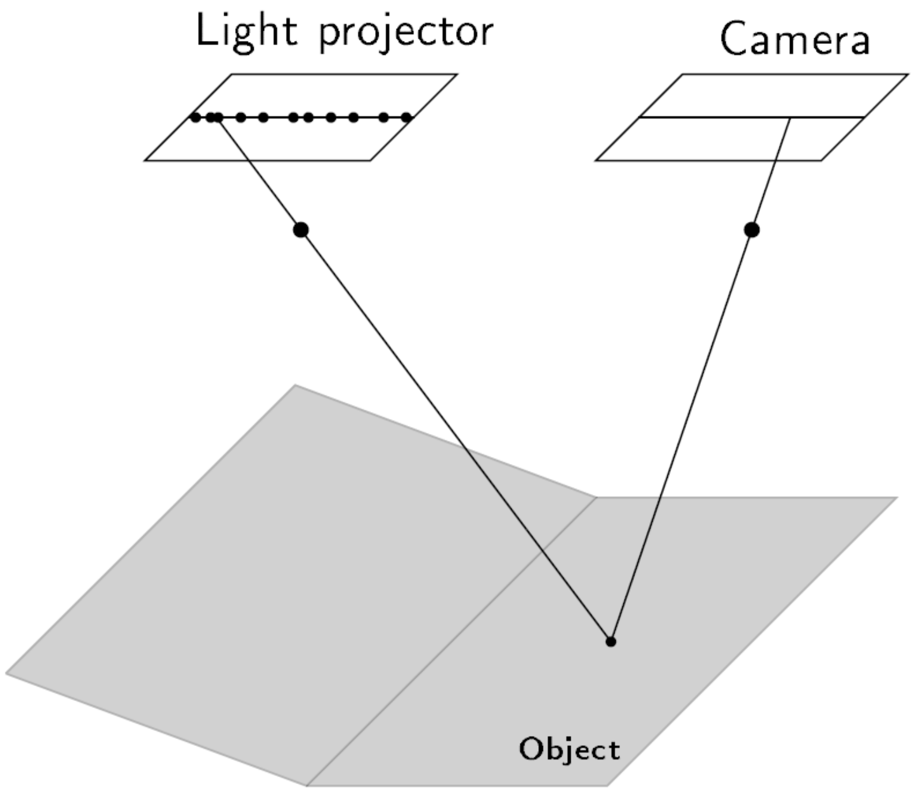
Structured light

- Light patterns are projected onto the scene



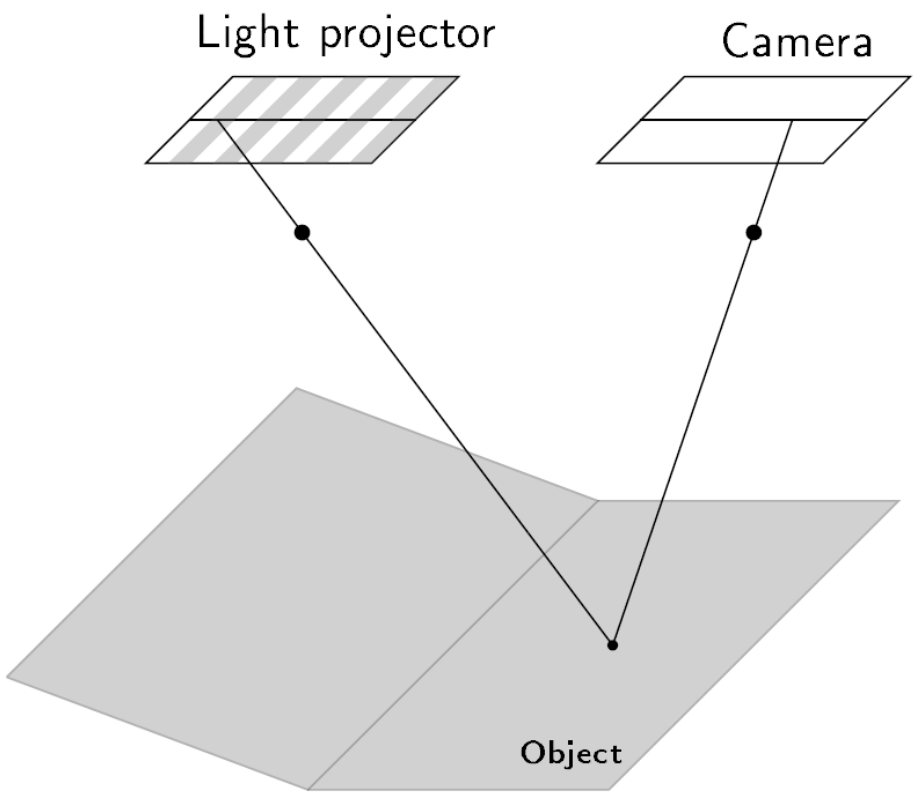
Structured light

- The patterns are made so that each token of light is distinguishable from the others



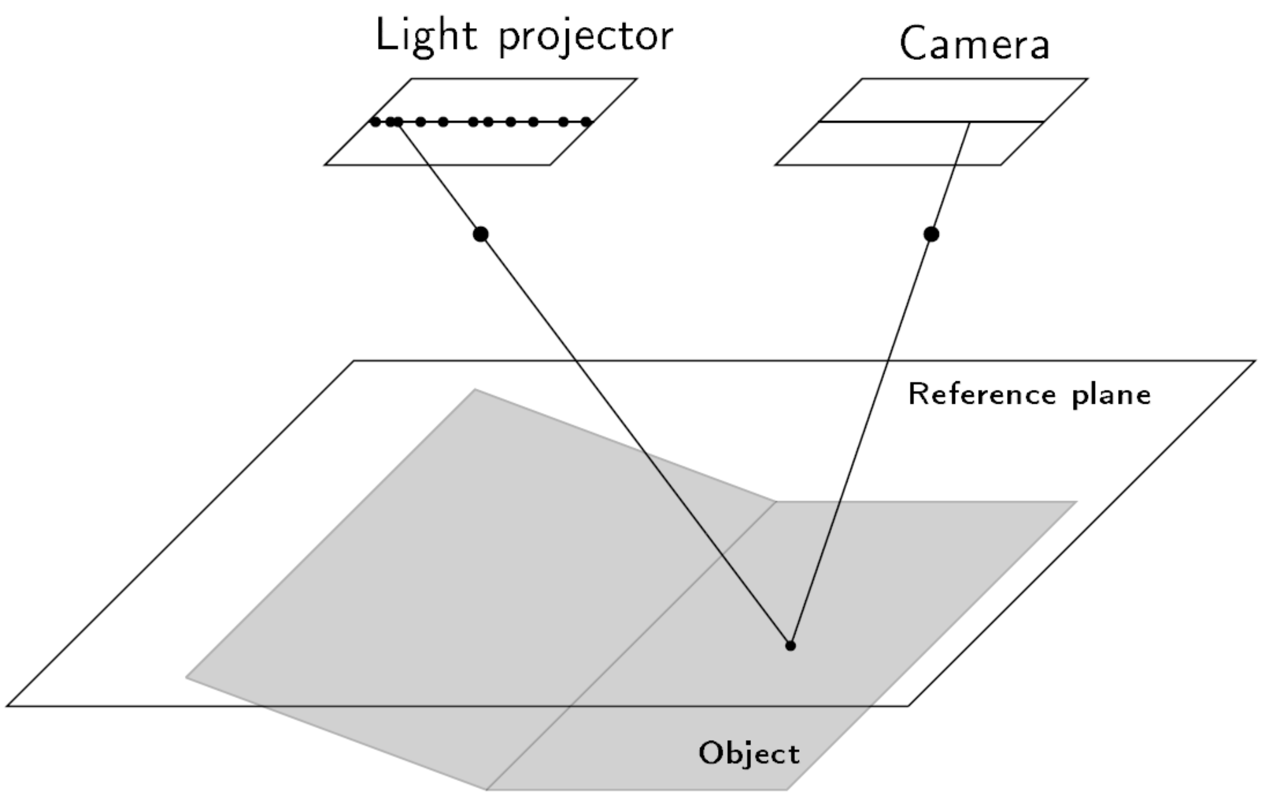
Structured light

- The patterns are made so that each token of light is distinguishable from the others



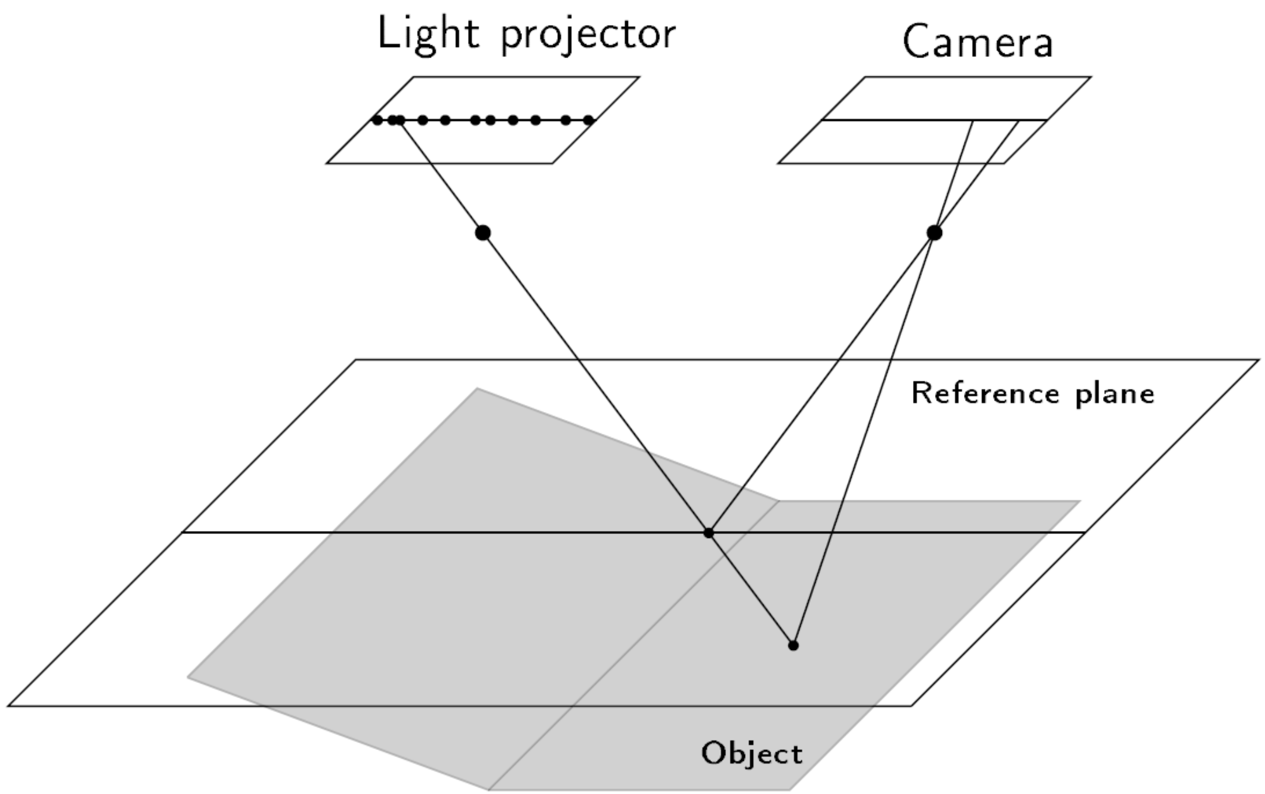
Structured light

- A reference image at known depth is captured using the camera



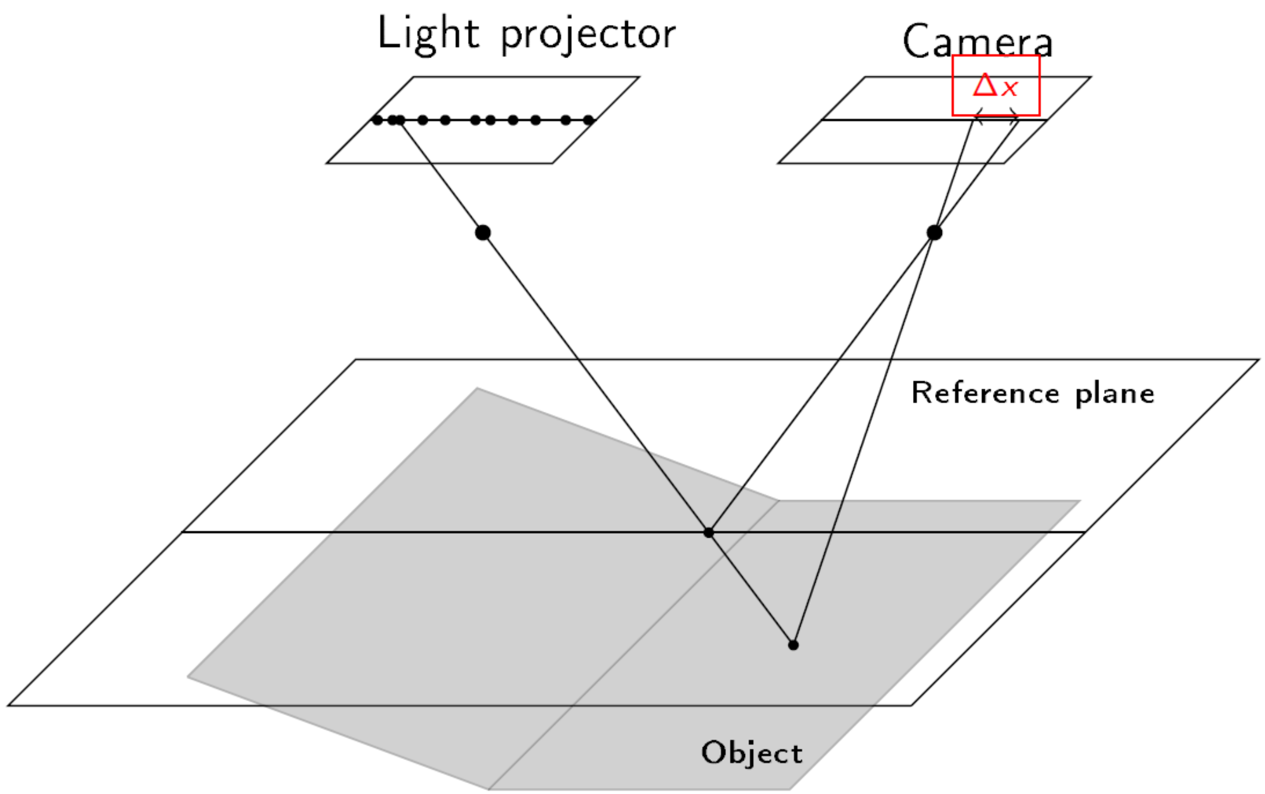
Structured light

- A reference image at known depth is captured using the camera

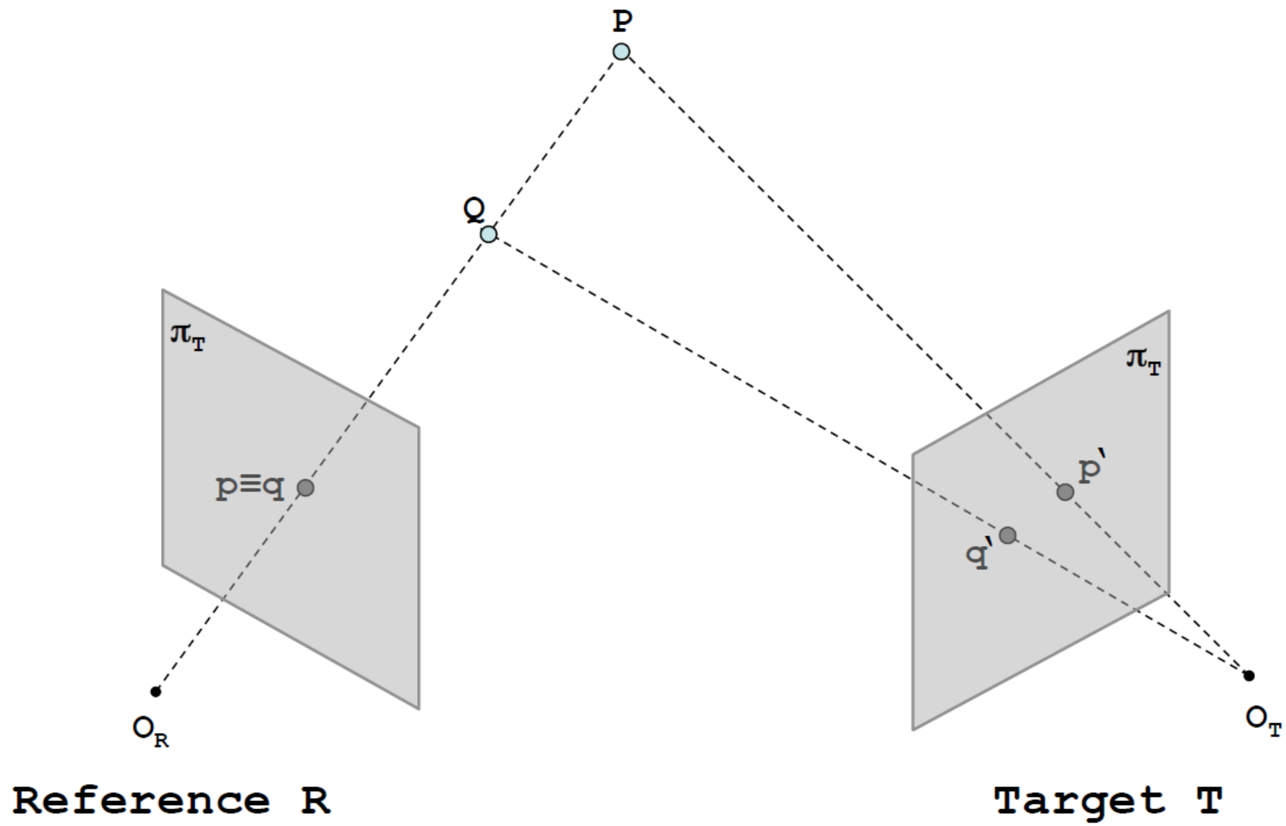


Structured light

- The shift Δx is proportional to the depth of the object



Stereo camera



With two (or more) cameras we can infer depth, by means of triangulation, if we are able to find corresponding (homologous) points in the two images

The objective

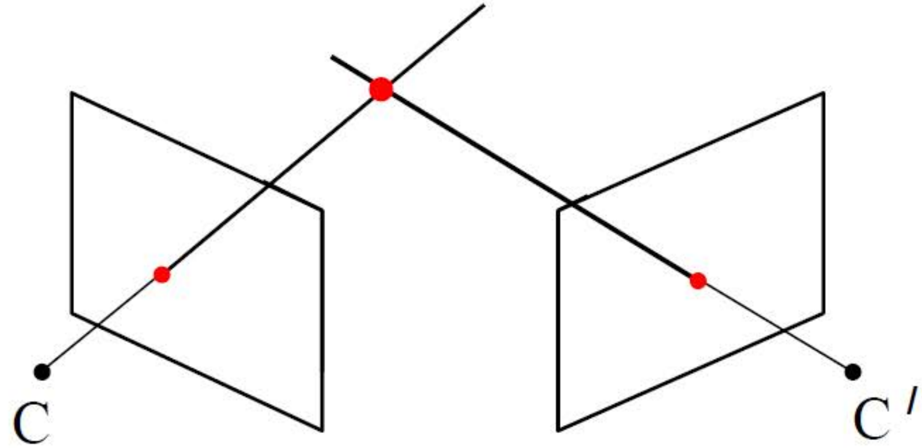
Given two images of a scene acquired by known cameras compute the 3D position of the scene (structure recovery)

Basic principle: triangulate from corresponding image points

- Determine 3D point at intersection of two back-projected rays

Corresponding points are images of the same scene point

Triangulation



The back-projected points generate rays which intersect at the 3D scene point

An algorithm for stereo reconstruction

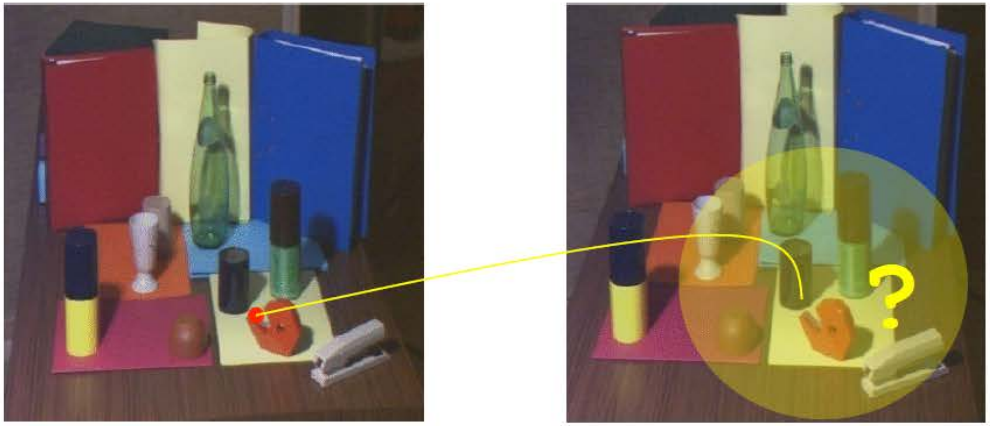
1. For each point in the first image determine the corresponding point in the second image

   (this is a search problem)

2. For each pair of matched points determine the 3D point by triangulation

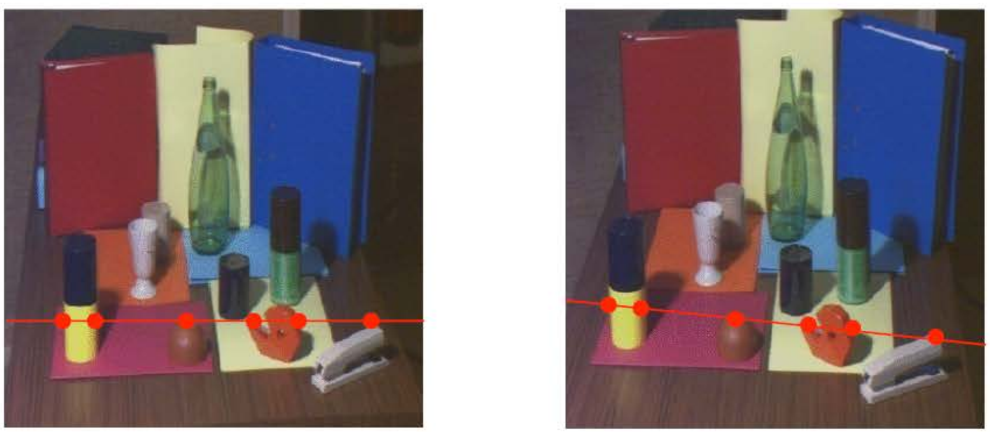
   (this is an estimation problem)

How to solve the correspondence problem ?



Reference (R)   Target (T)

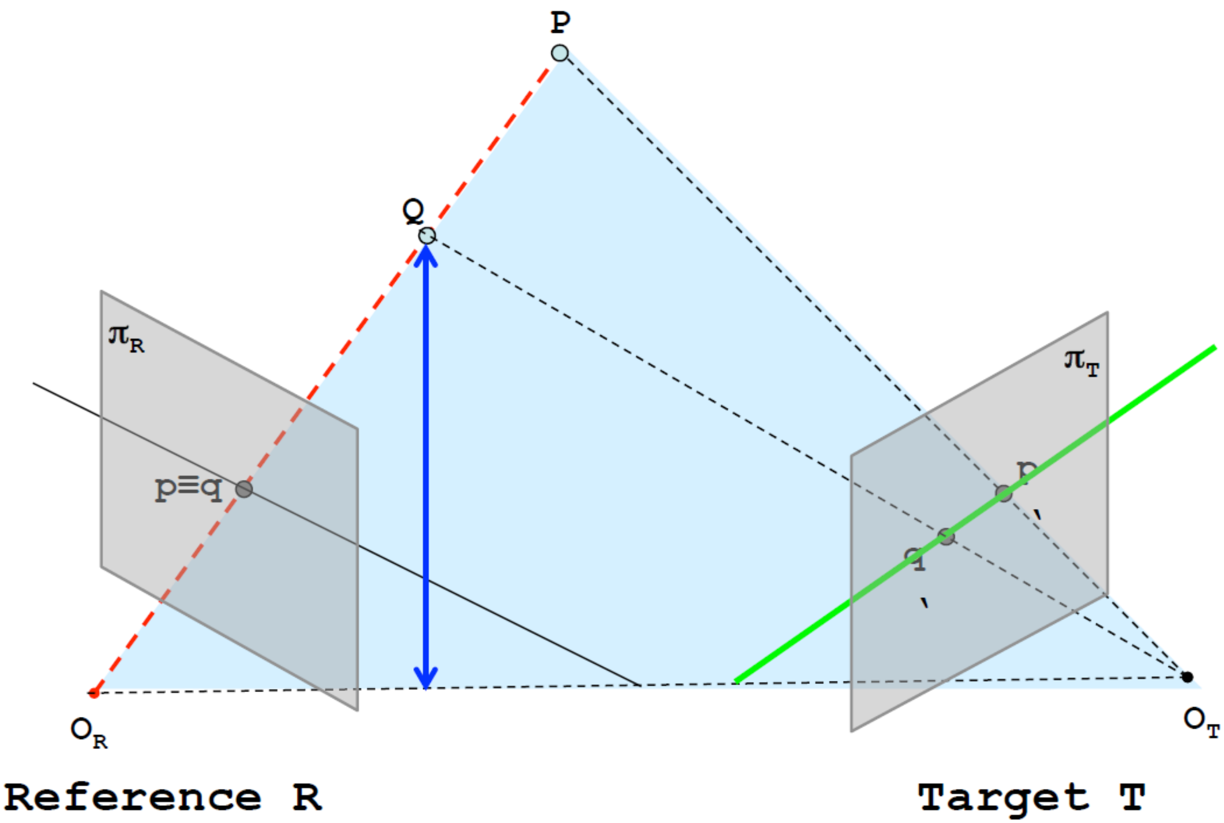
2D search domain ?



Reference (R)   Target (T)

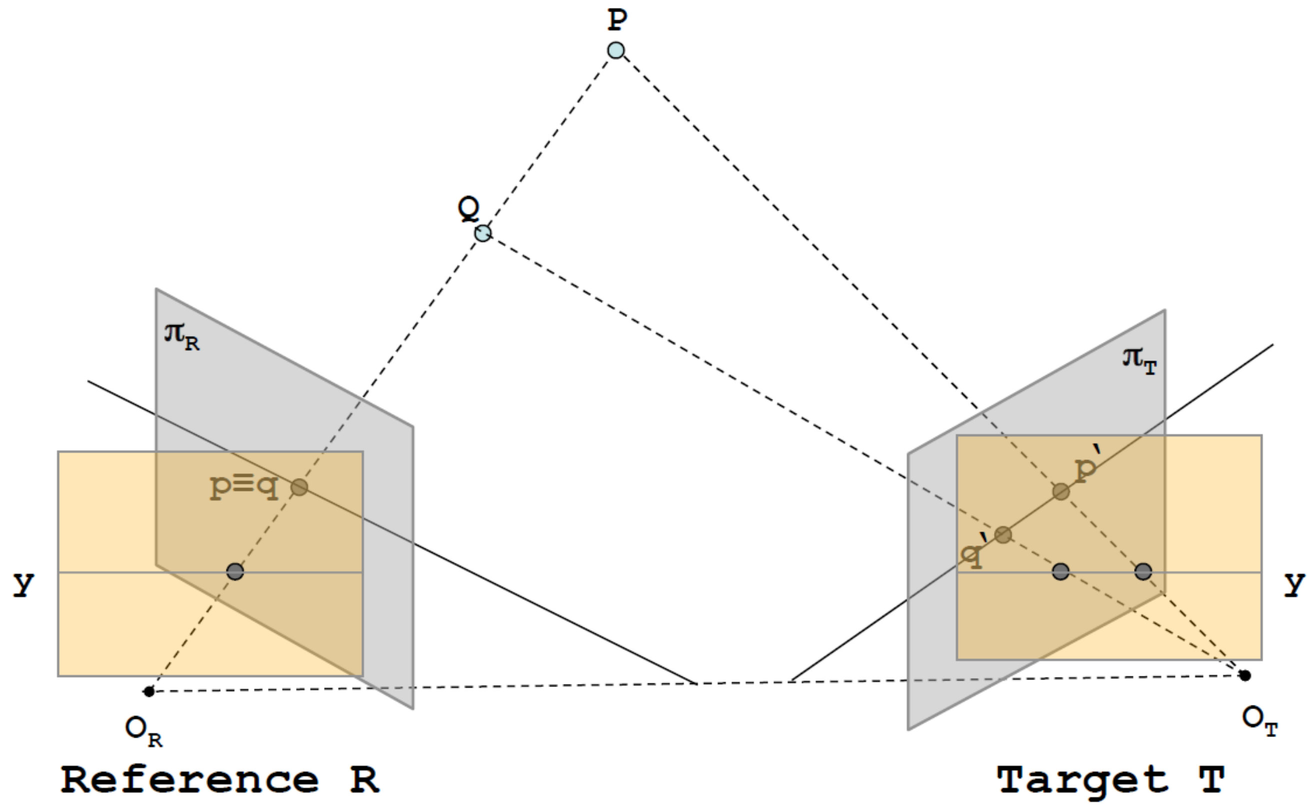
No!! Thanks to the epipolar constraint

Epipolar constraint

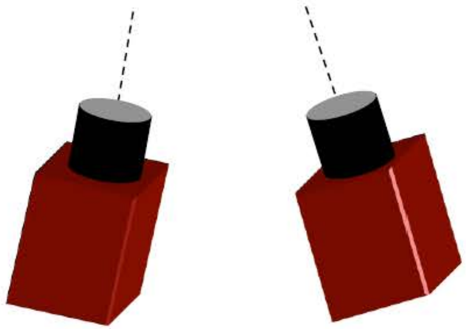


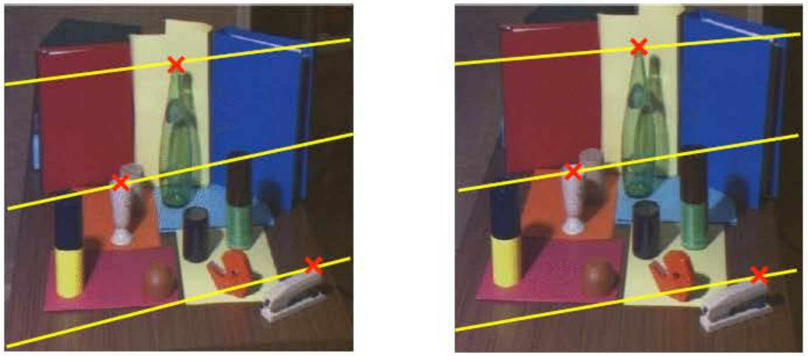
- Consider two points P and Q on the same line of sight of the reference image R (both points project into the same image point p≡q on image plane π_R of the reference image)
- The epipolar constraint states that the correspondence for a point belonging to the (red) line of sight lies on the green line on image plane π_T of target image

Stereo camera in standard form

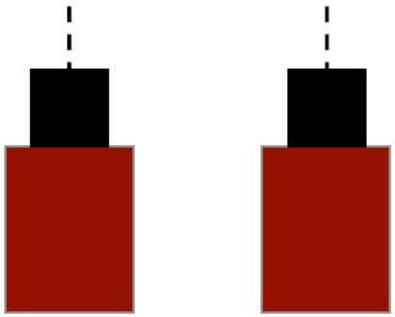


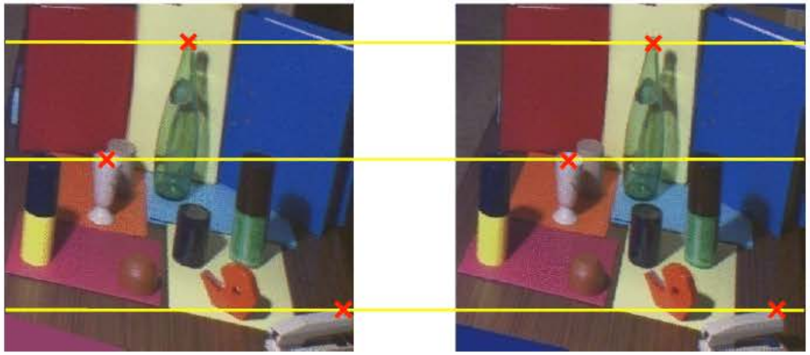
Once we know that the search space for corresponding points can be narrowed from 2D to 1D, we can put (virtually) the stereo rig in a more convenient configuration (standard form) - corresponding points are constrained on the same image scanline





Original stereo pair

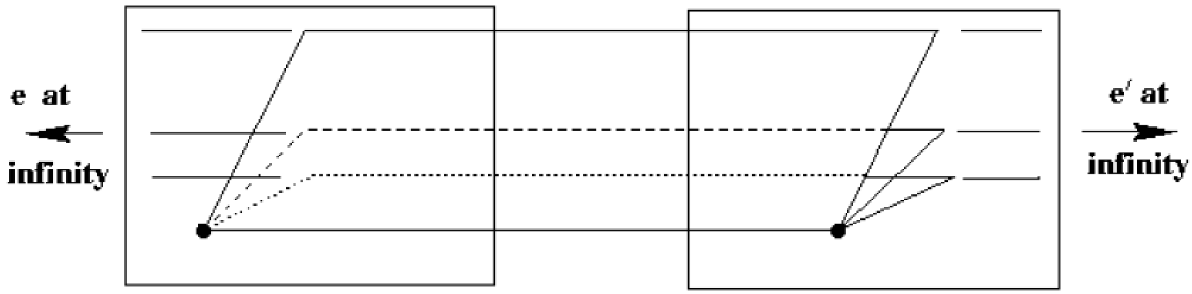


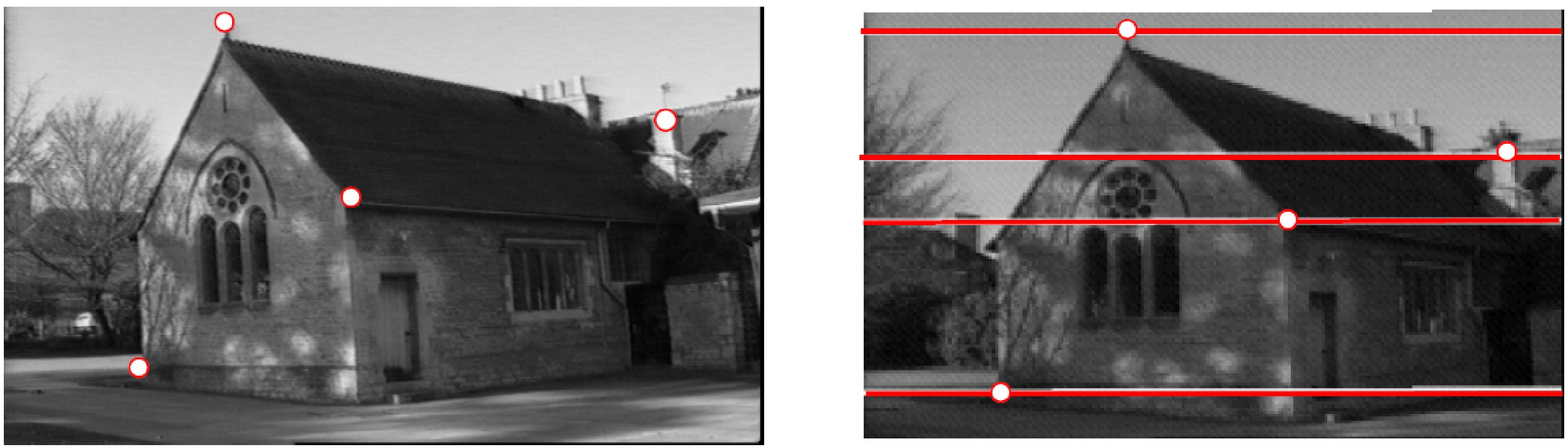


Stereo pair in standard form

Cameras are "*perfectly*" aligned and with the same *focal length*

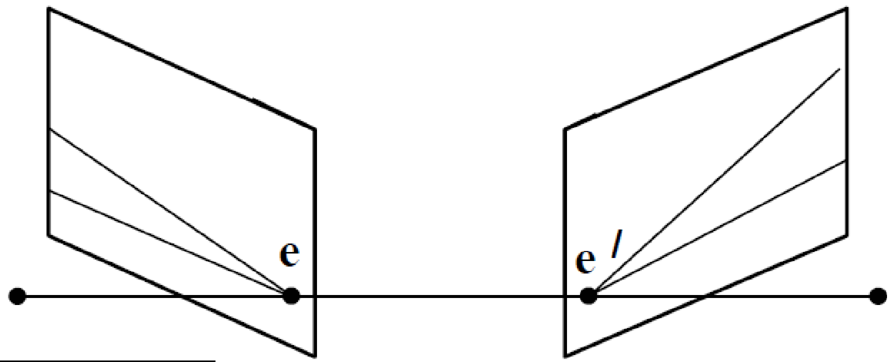
Epipolar geometry example I: parallel cameras

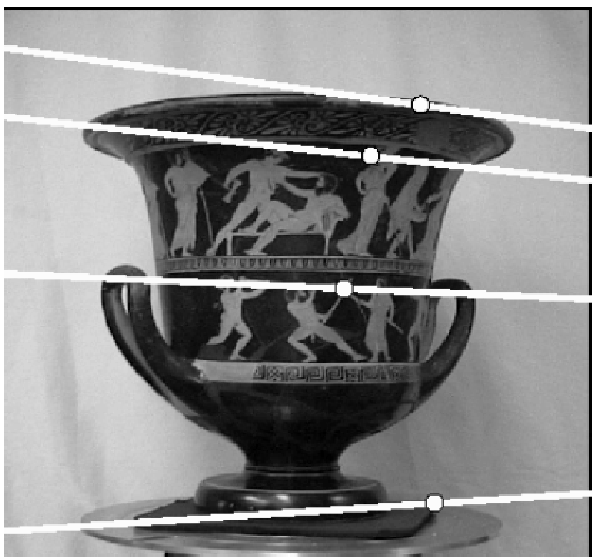


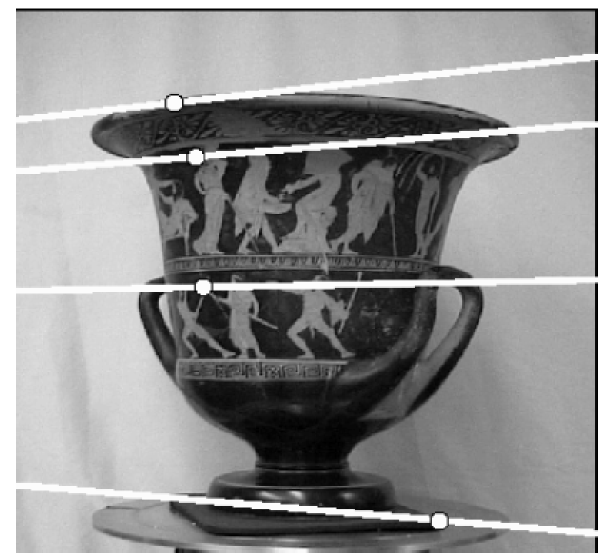


Epipolar geometry depends only on the relative pose (position and orientation) and internal parameters of the two cameras, i.e. the position of the camera centres and image planes. It does not depend on the scene structure (3D points external to the camera).

Epipolar geometry example II: converging cameras

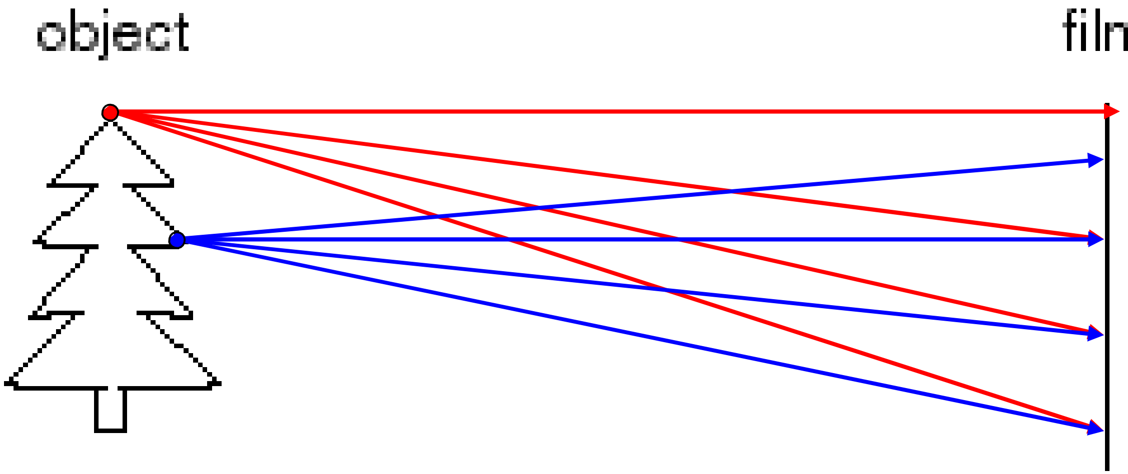






Note, epipolar lines are in general not parallel

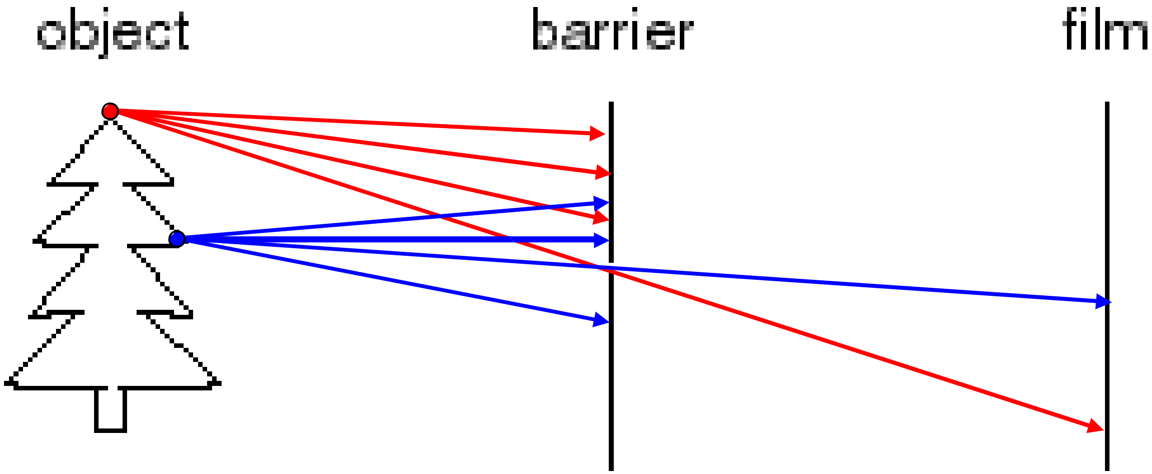
How do we see the world?



Design a camera

- Idea 1: put a piece of film in front of an object
- Do we get a reasonable image?

Pinhole camera

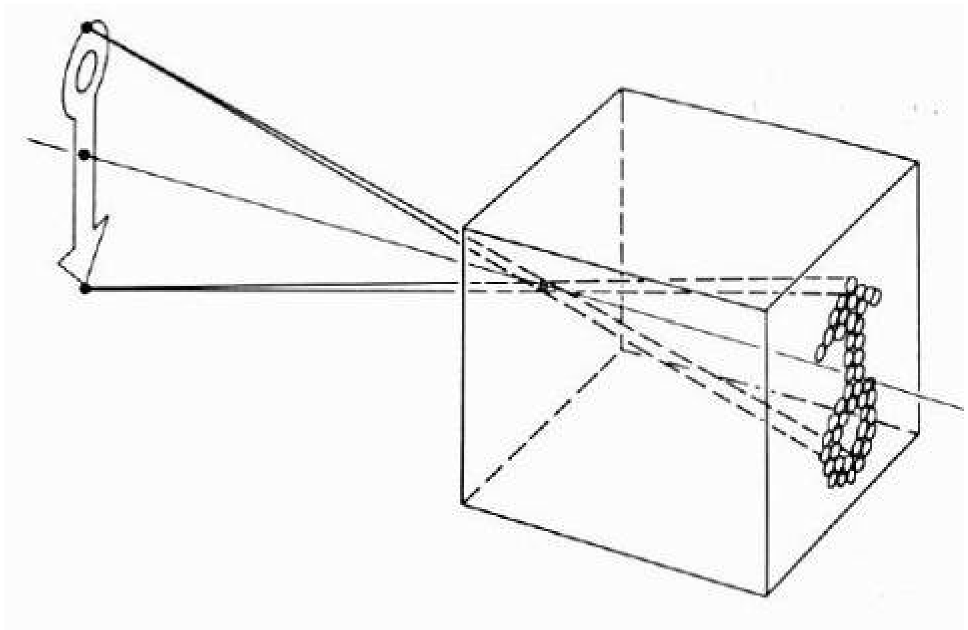


Add a barrier to block off most of the rays

- This reduces blurring
- The opening known as the **aperture**
- How does this transform the image?

Slide by Steve Seitz

Pinhole camera model

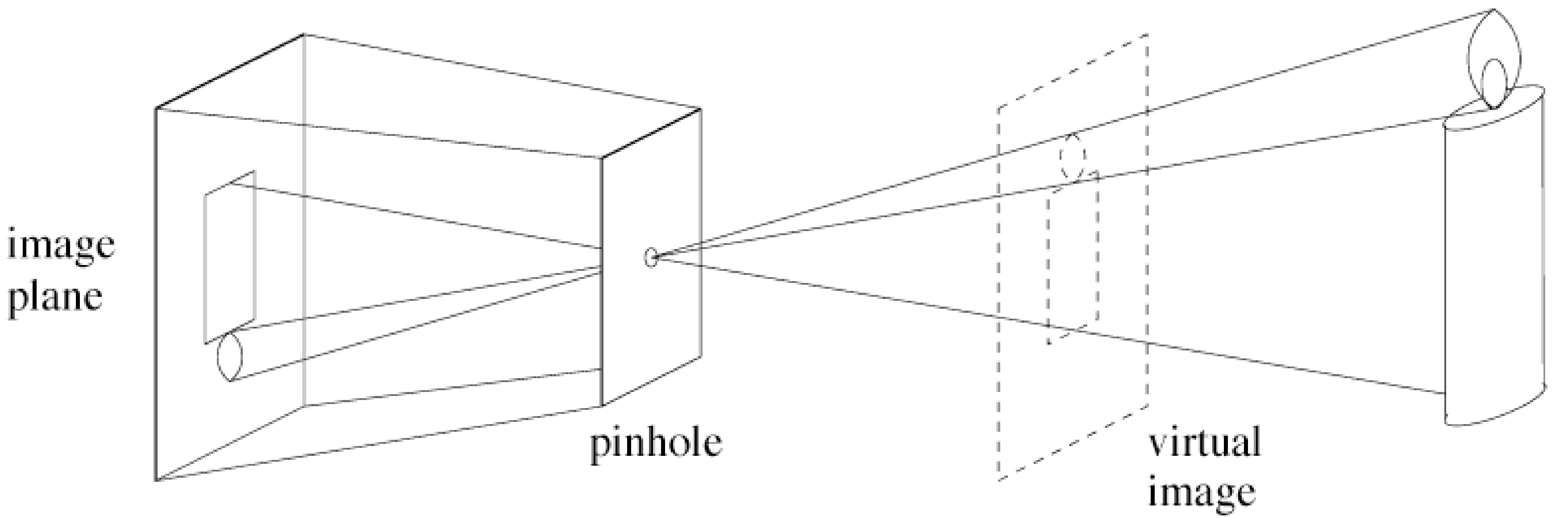


Pinhole model:

- Captures **pencil of rays** – all rays through a single point
- The point is called **Center of Projection (COP)**
- The image is formed on the **Image Plane**
- **Effective focal length *f*** is distance from COP to Image Plane

Slide by Steve Seitz

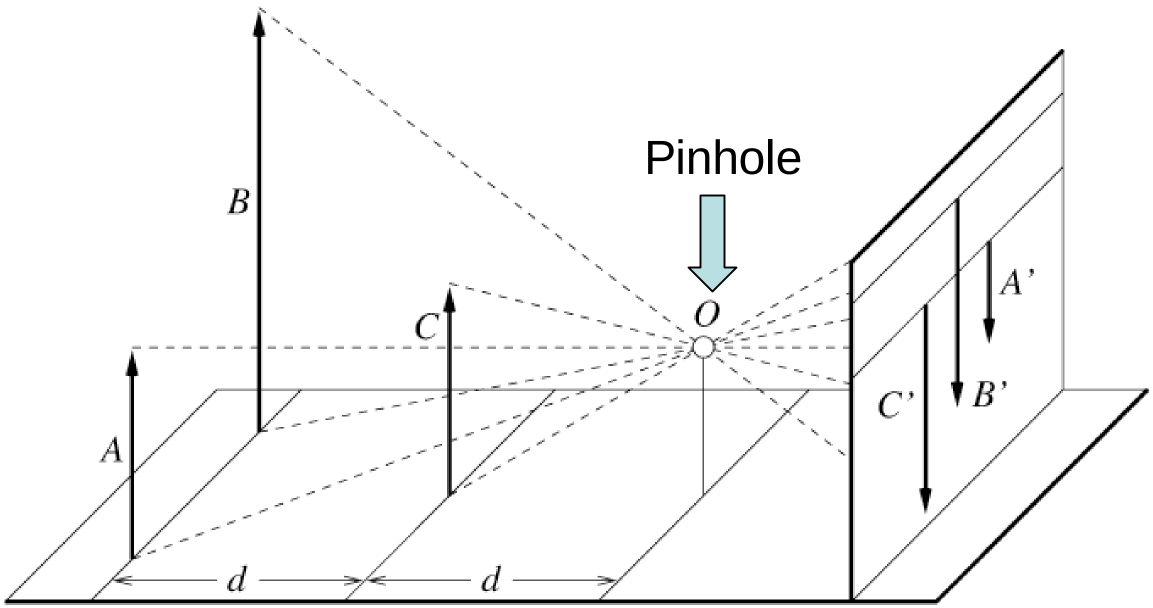
We'll use the pinhole camera model to describe image formation



Notice how the image is inverted

(Image from Slides by Forsyth)

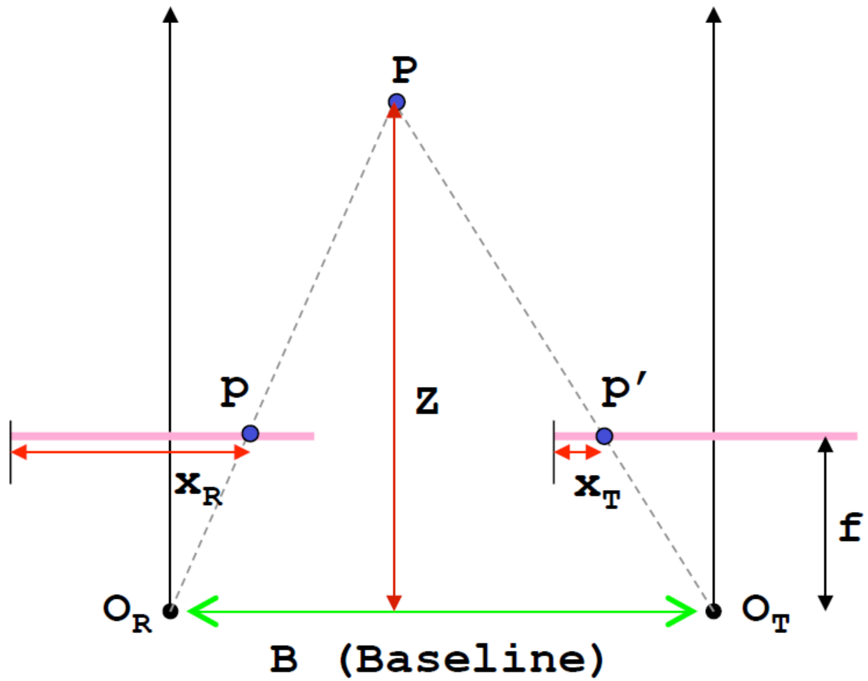
Projection Effects



- Height of objects depends on the distance from the pinhole (O)

(Image from Slides by Forsyth)

Disparity and depth

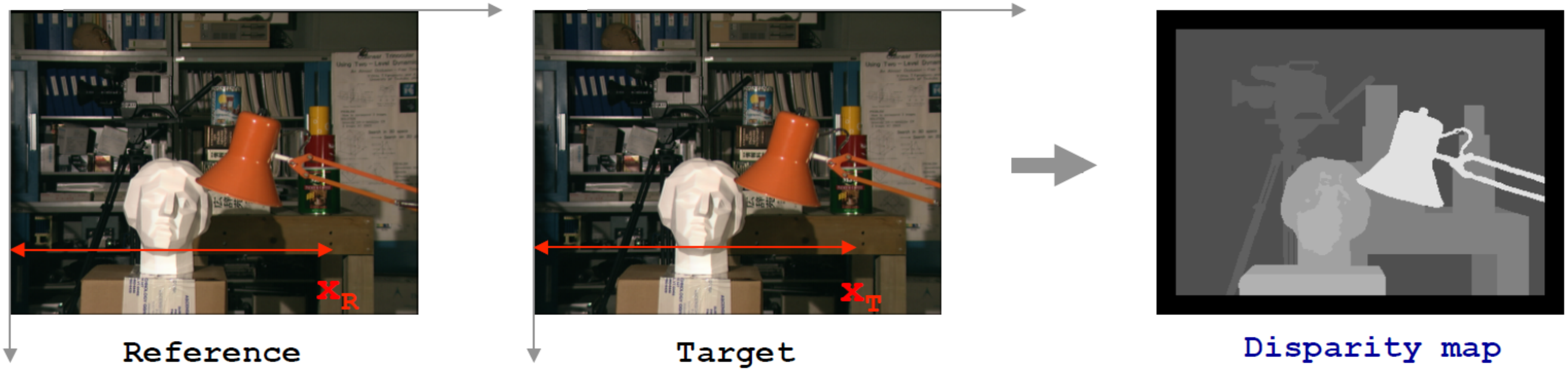


With the stereo rig in standard form and by considering similar triangles (PO_RO_T and Ppp'):

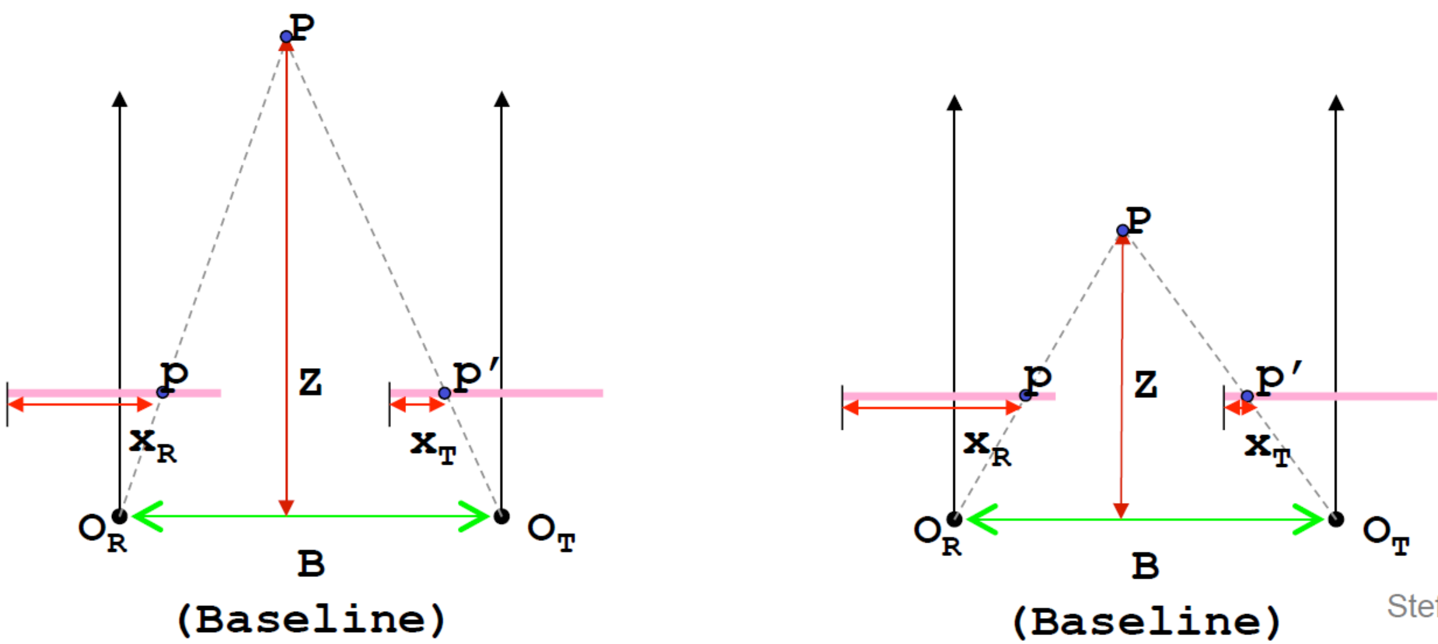
$$\frac{b}{Z} = \frac{(b + x_T) - x_R}{Z - f} \quad \Rightarrow \quad Z = \frac{b \cdot f}{x_R - x_T} = \frac{b \cdot f}{d}$$

$x_R - x_T$ is the **disparity**

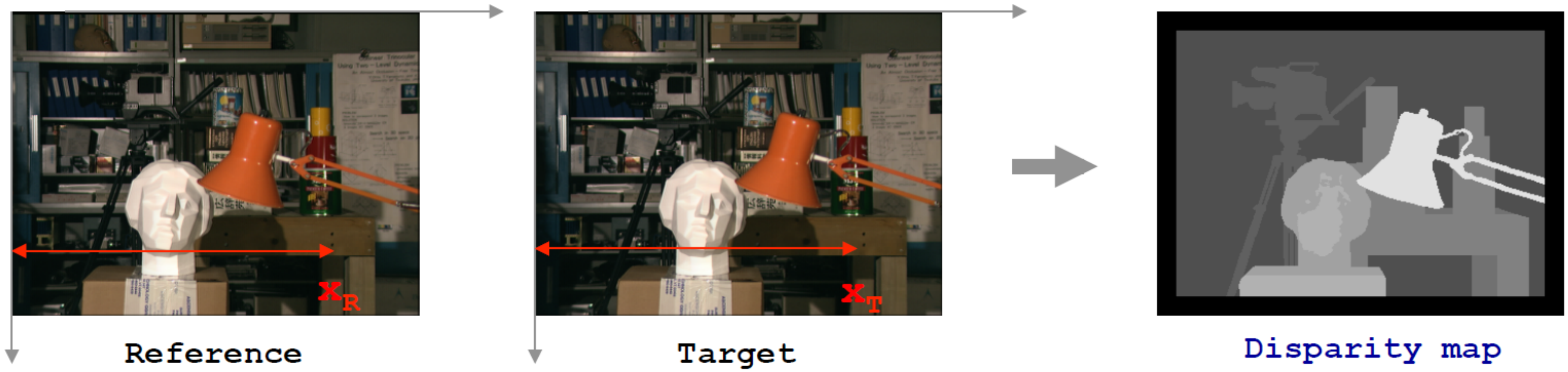
The disparity is the difference between the x coordinate of two corresponding points; it is typically encoded with greyscale image (closer points are brighter).



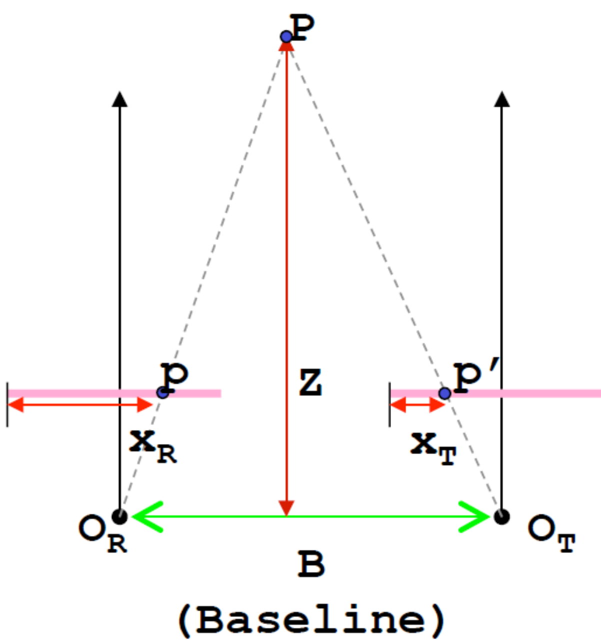
Disparity is higher for points closer to the camera

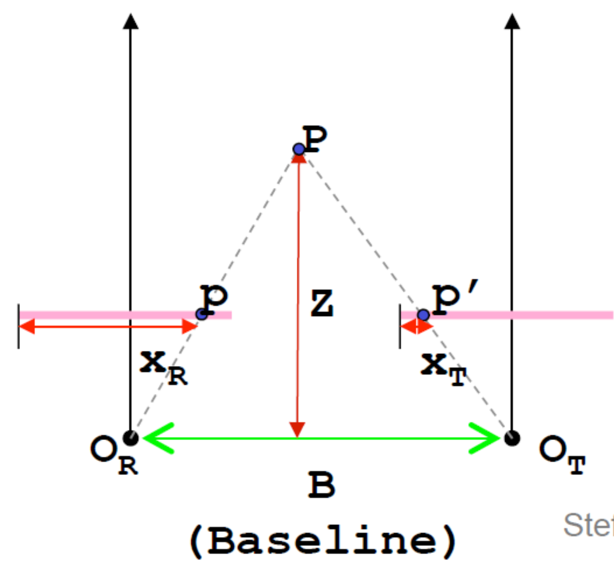


The disparity is the difference between the x coordinate of two corresponding points; it is typically encoded with greyscale image (closer points are brighter).

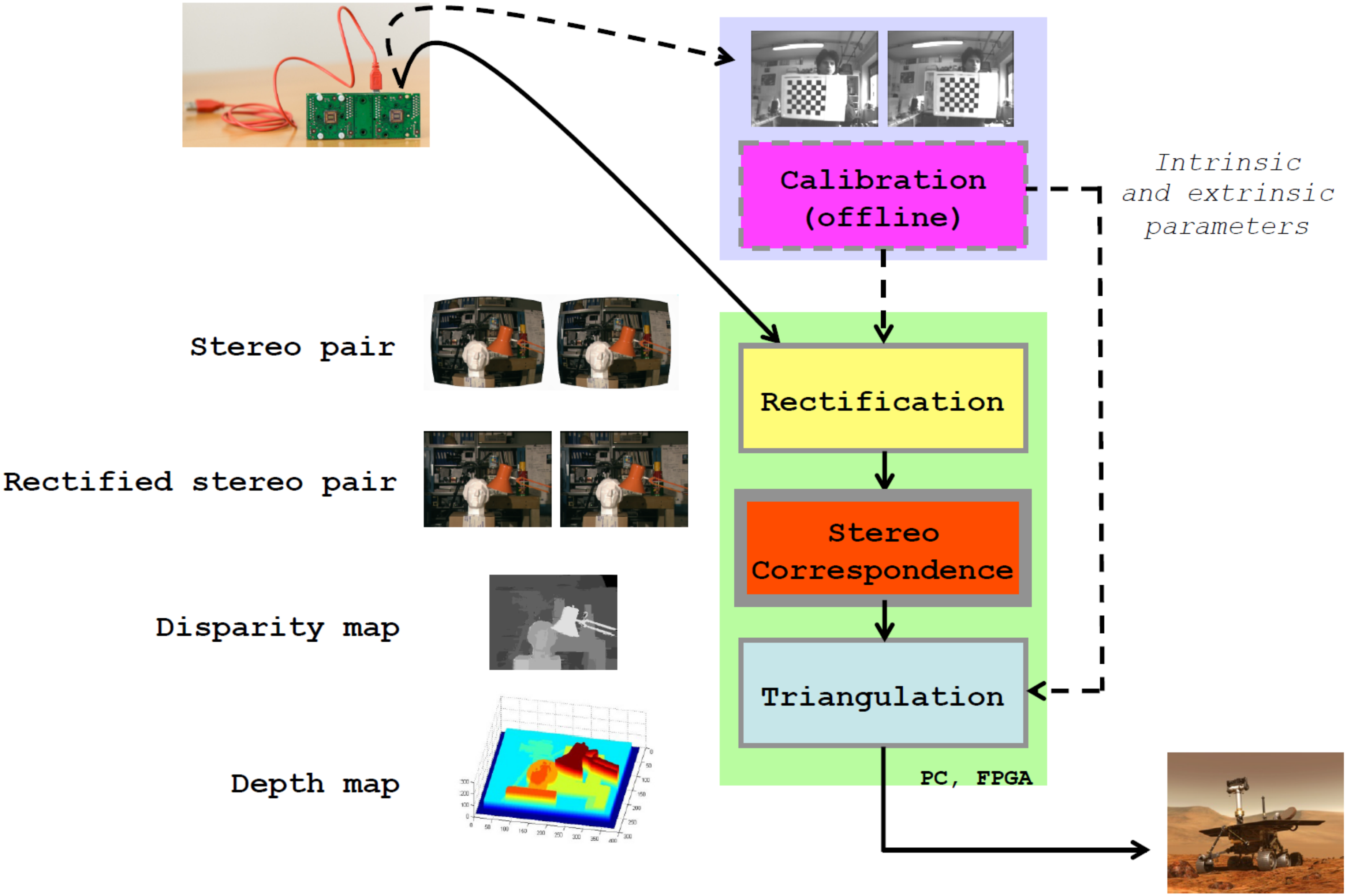


Disparity is higher for points closer to the camera

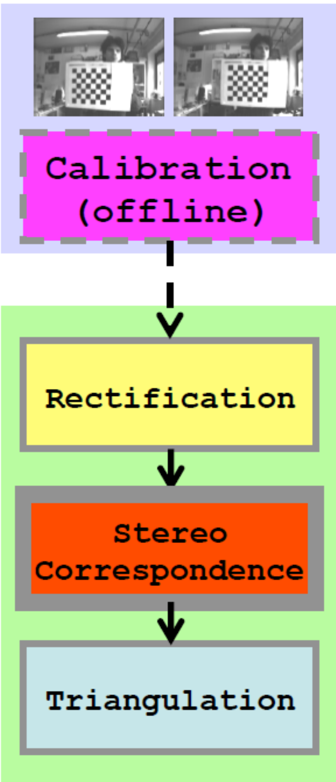




Overview of a stereo vision system

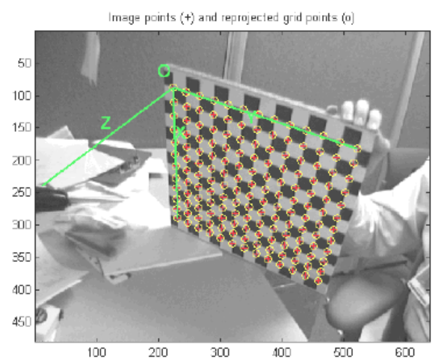


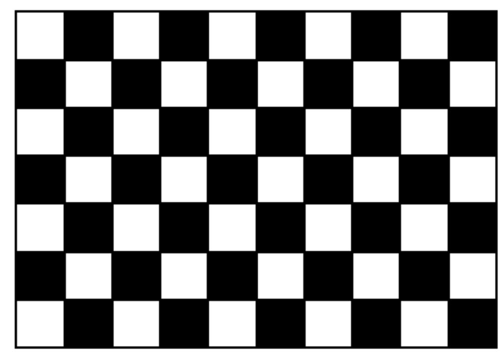
Calibration (offline)



Offline procedure aimed at finding:

- Intrinsic parameters of the two cameras (focal length, image center, parameters of lenses distortion, etc)
- Extrinsic parameters (R and T that aligns the two cameras)

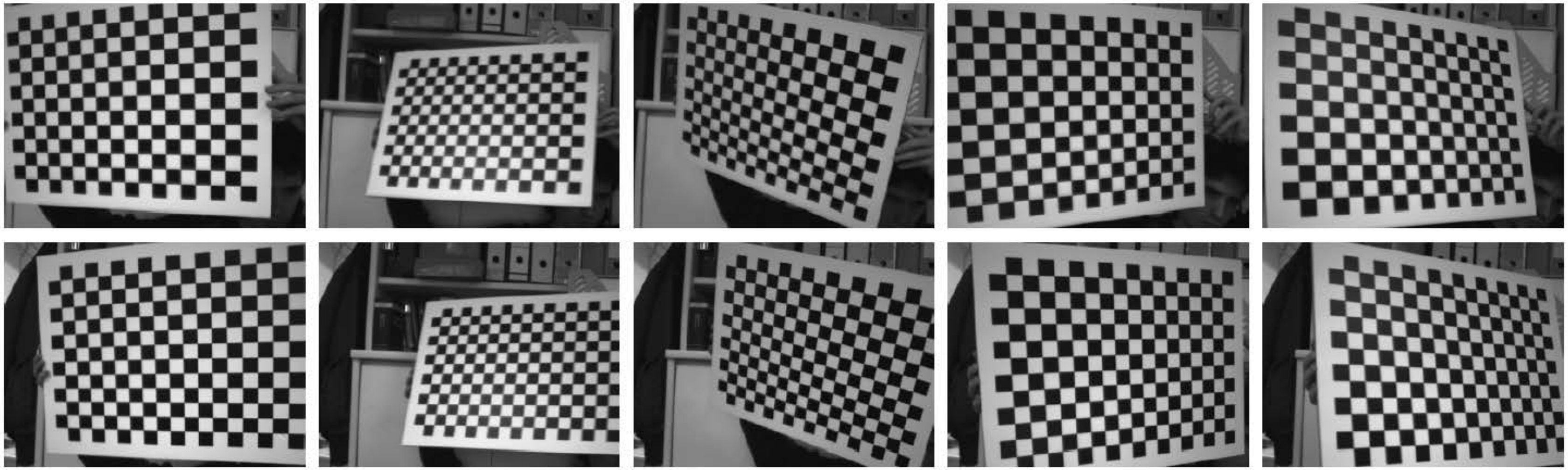




Calibration is carried out acquiring and processing 10+ stereo pairs of a known pattern (typically a checkerboard)

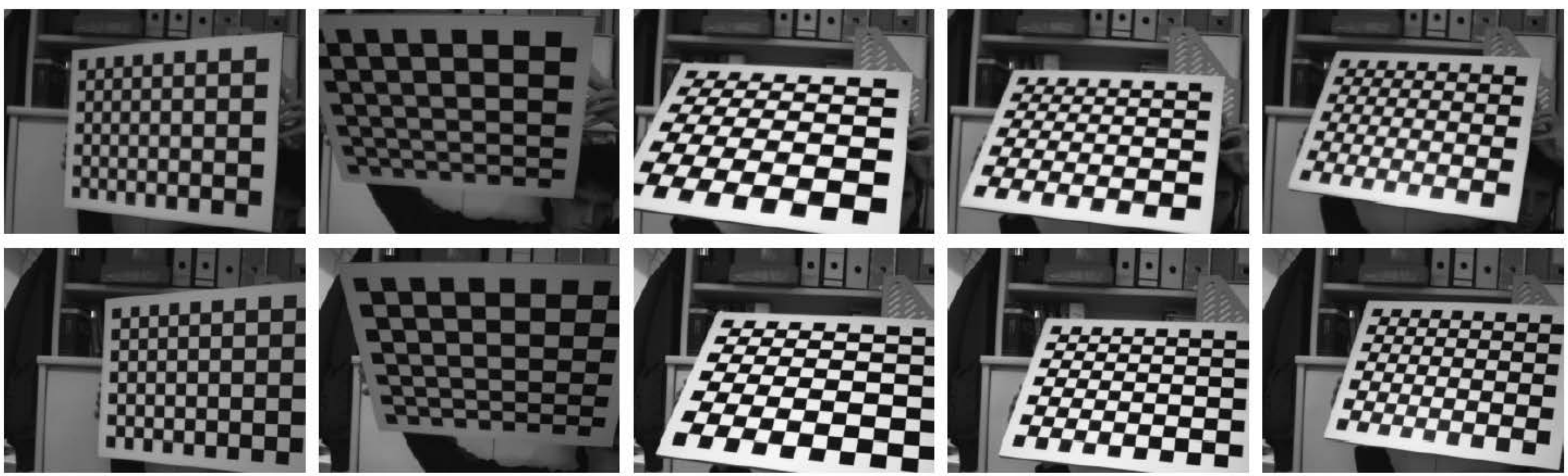
R

T



R

T

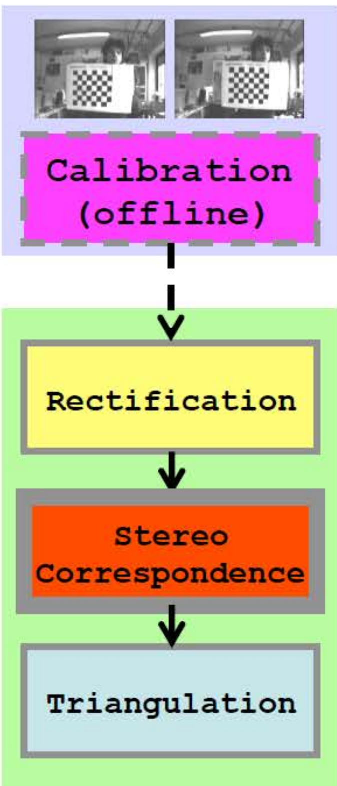


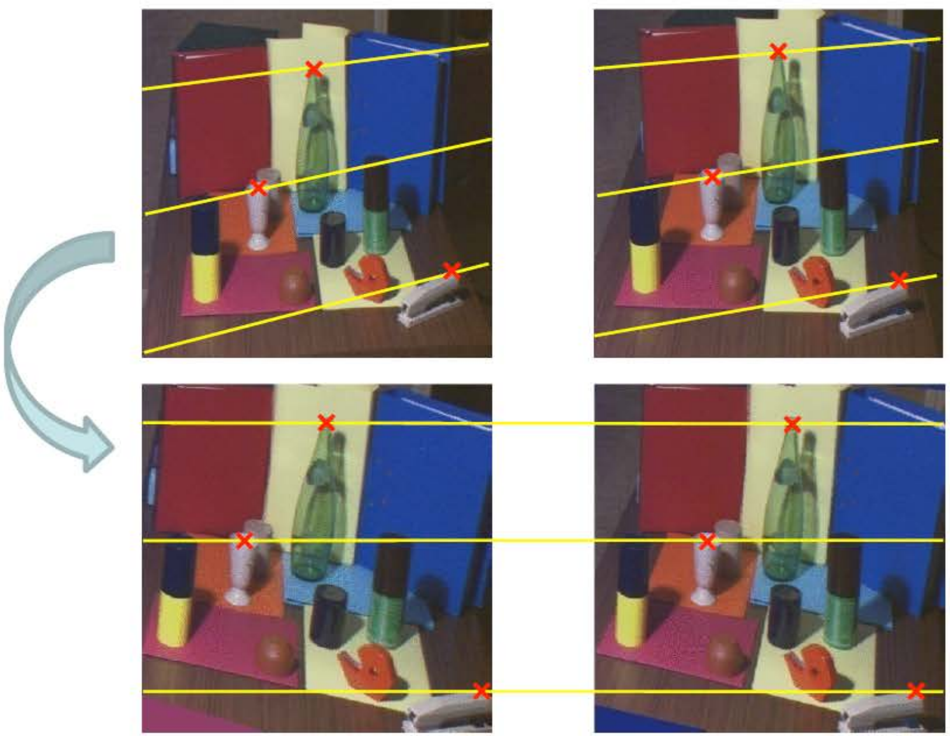
Rectification

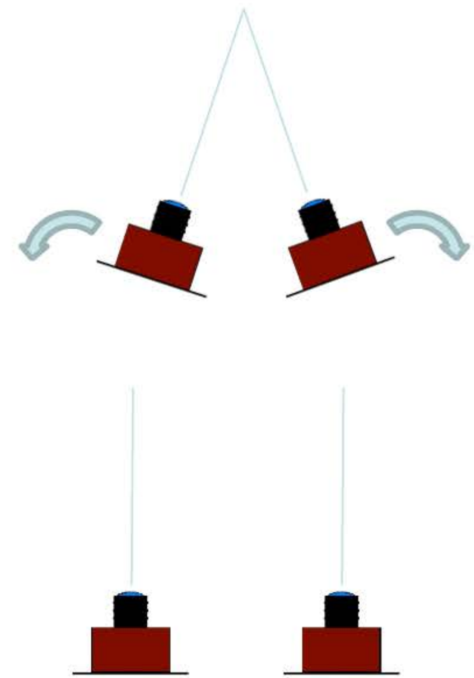
Using the information from the calibration step:

a) removes lens distortions

b) turns the stereo pair in standard form



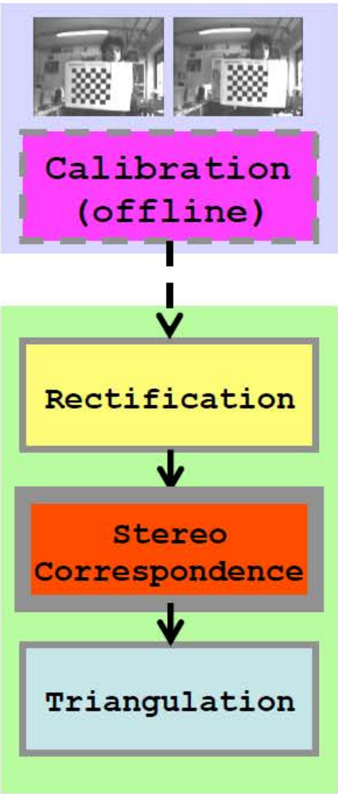


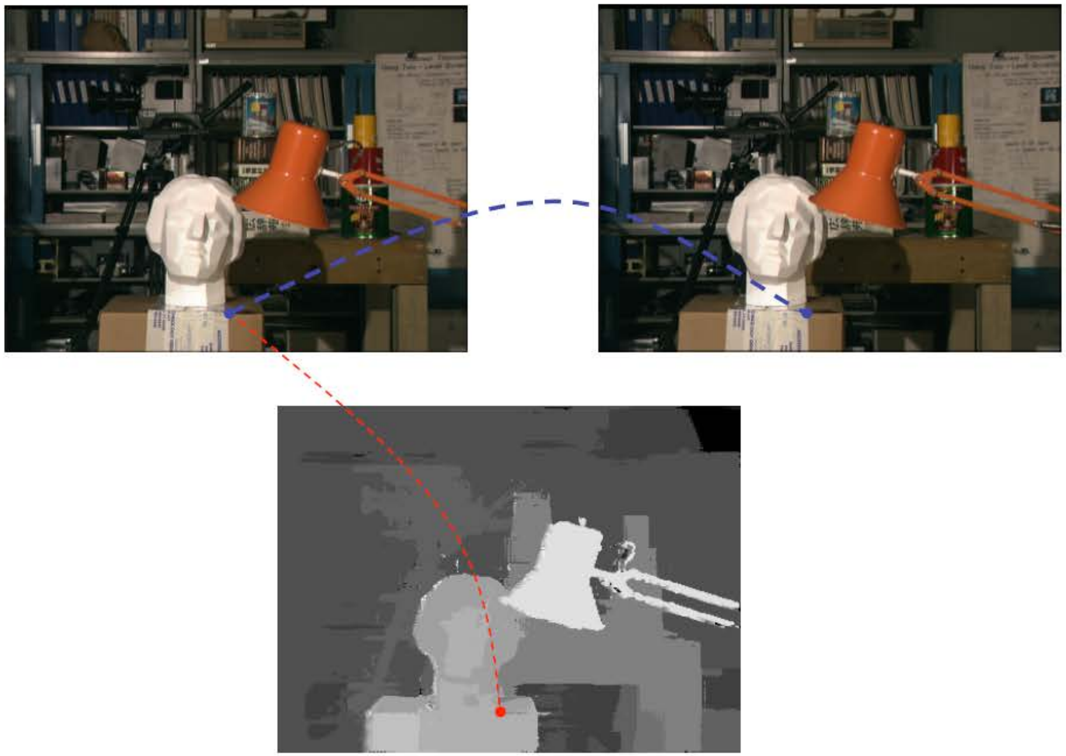


Stereo camera in standard form

Stereo correspondence

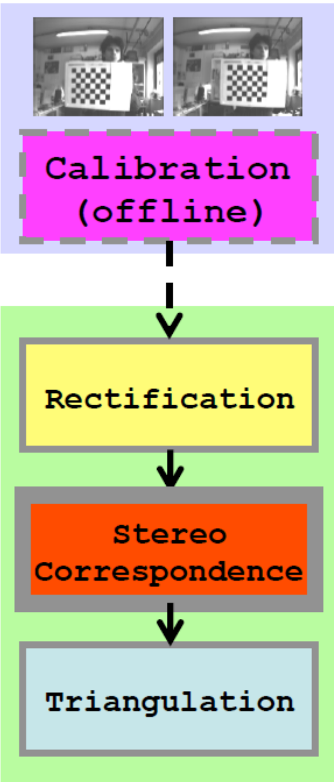
Aims at finding homologous points in the stereo pair.



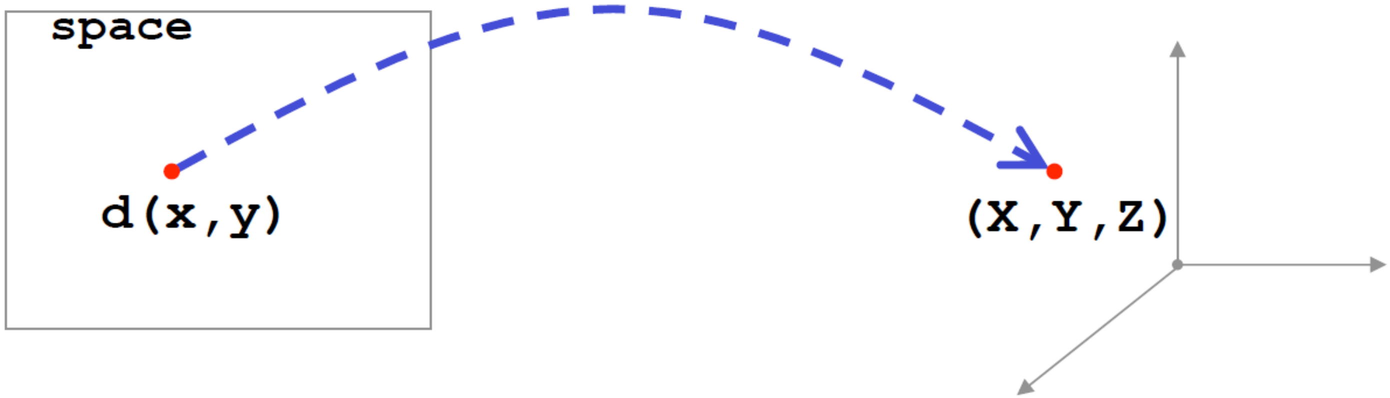


disparity map

Triangulation



Given the disparity map, the baseline and the Focal length (calibration): triangulation computes the position of the correspondence in the 3D space

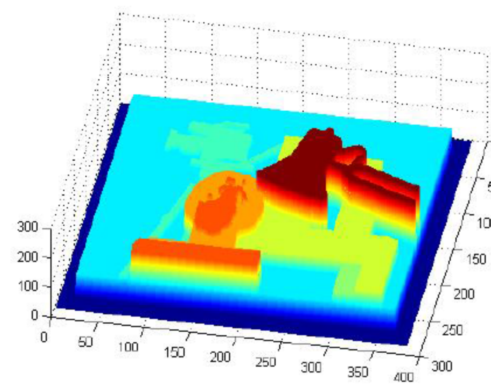


disparity map

$$Z = \frac{b \cdot f}{d}$$

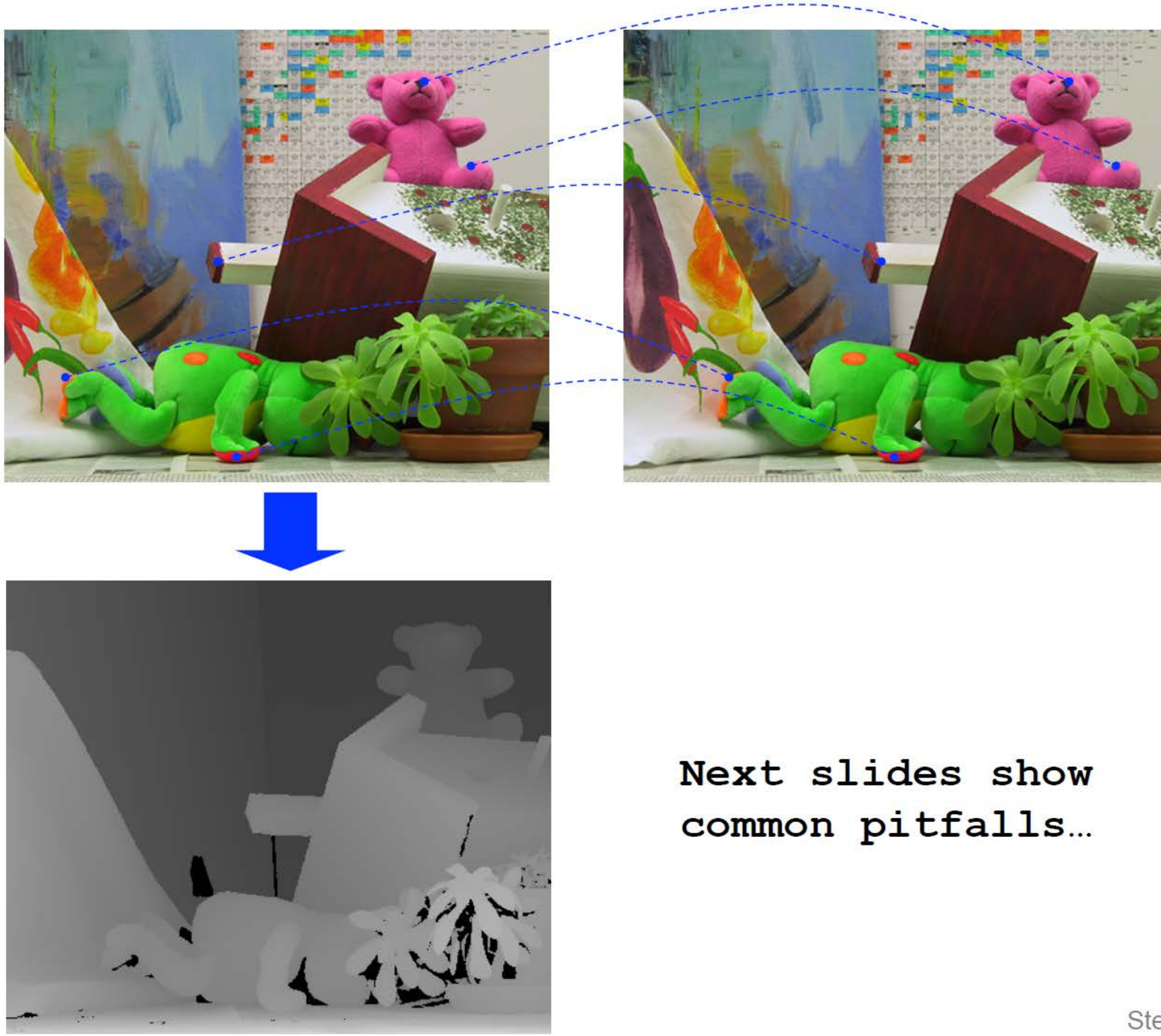
$$X = Z\frac{x_R}{f}$$

$$Y = Z\frac{y_R}{f}$$



depth map

Why is stereo correspondence so challenging ?



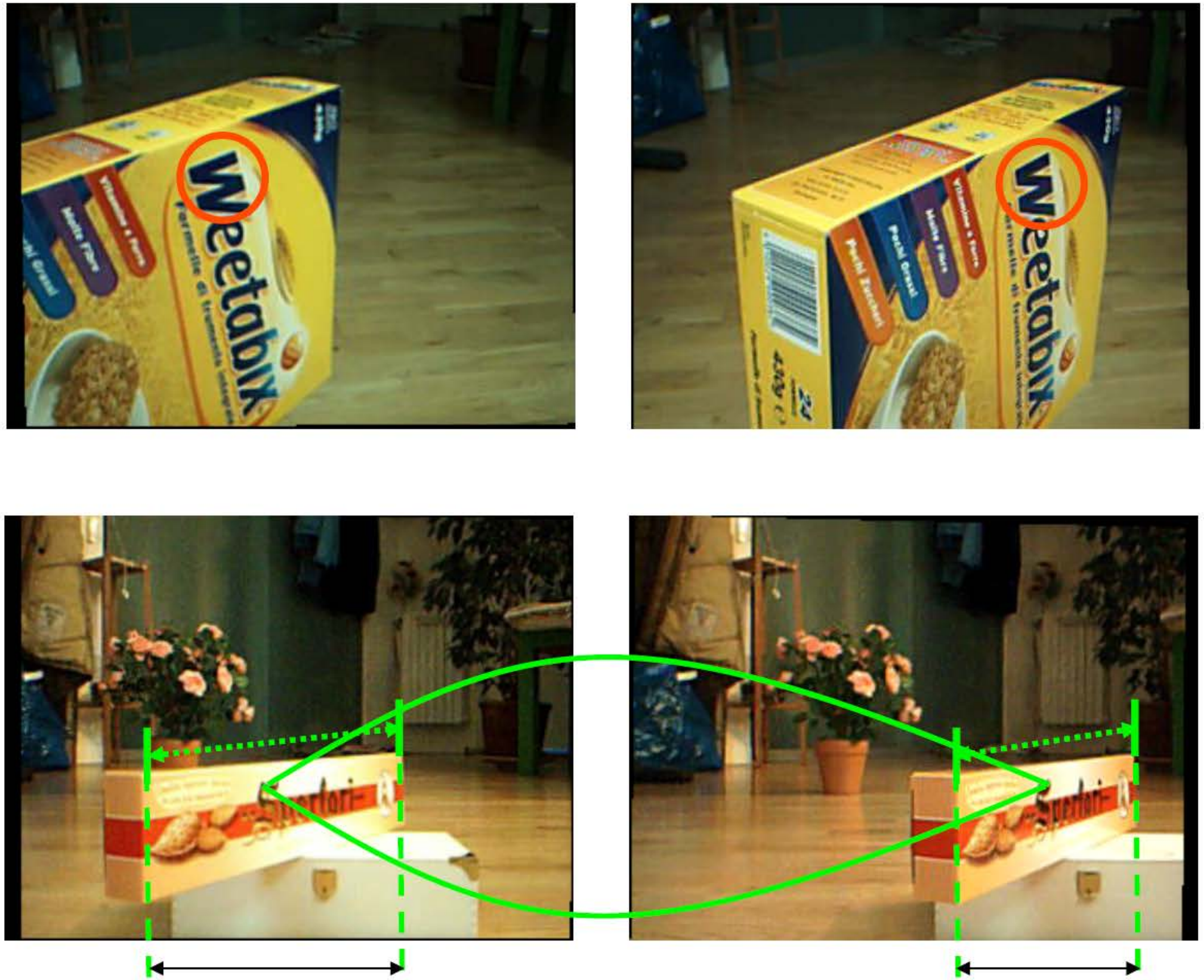
Next slides show common pitfalls…

Photometric distortions and noise

Specular surfaces

Stefano Mattoccia

Foreshortening

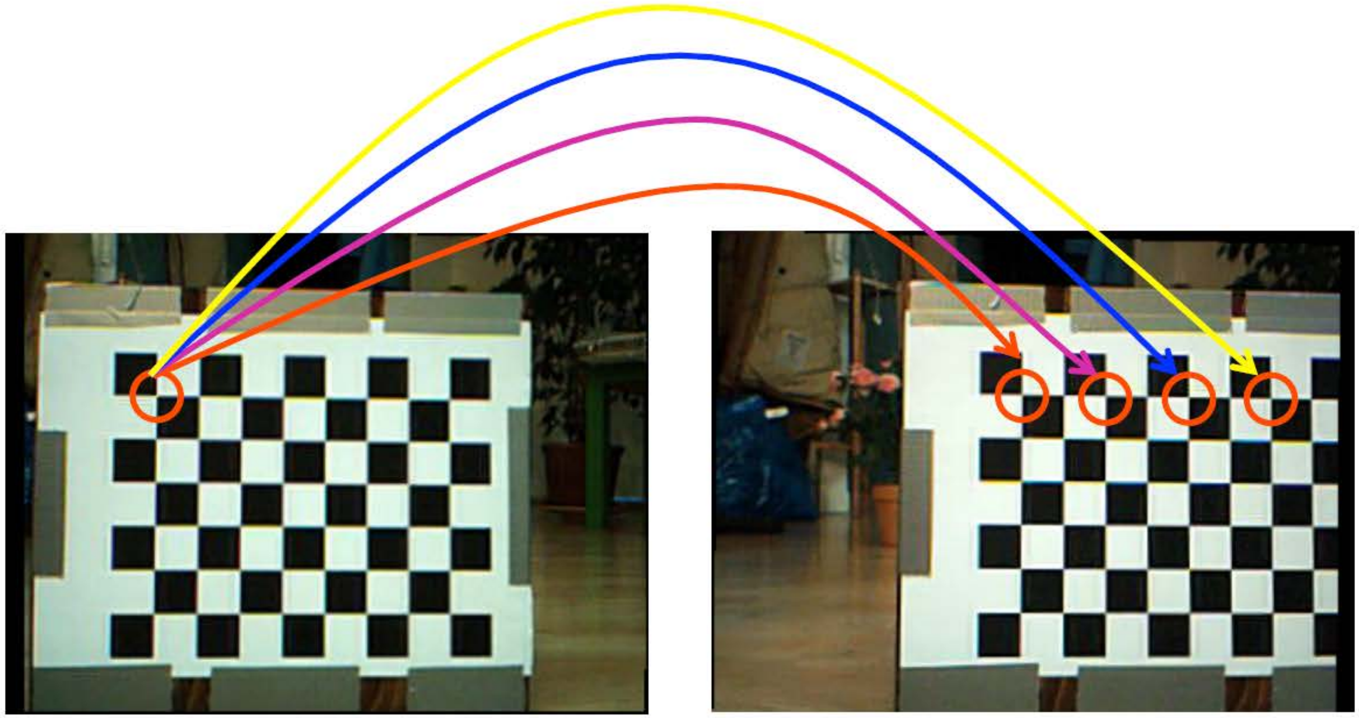


Uniqueness constraint ? :-(

Perspective distortions

Uniform/ambiguous regions

Repetitive/ambiguous patterns

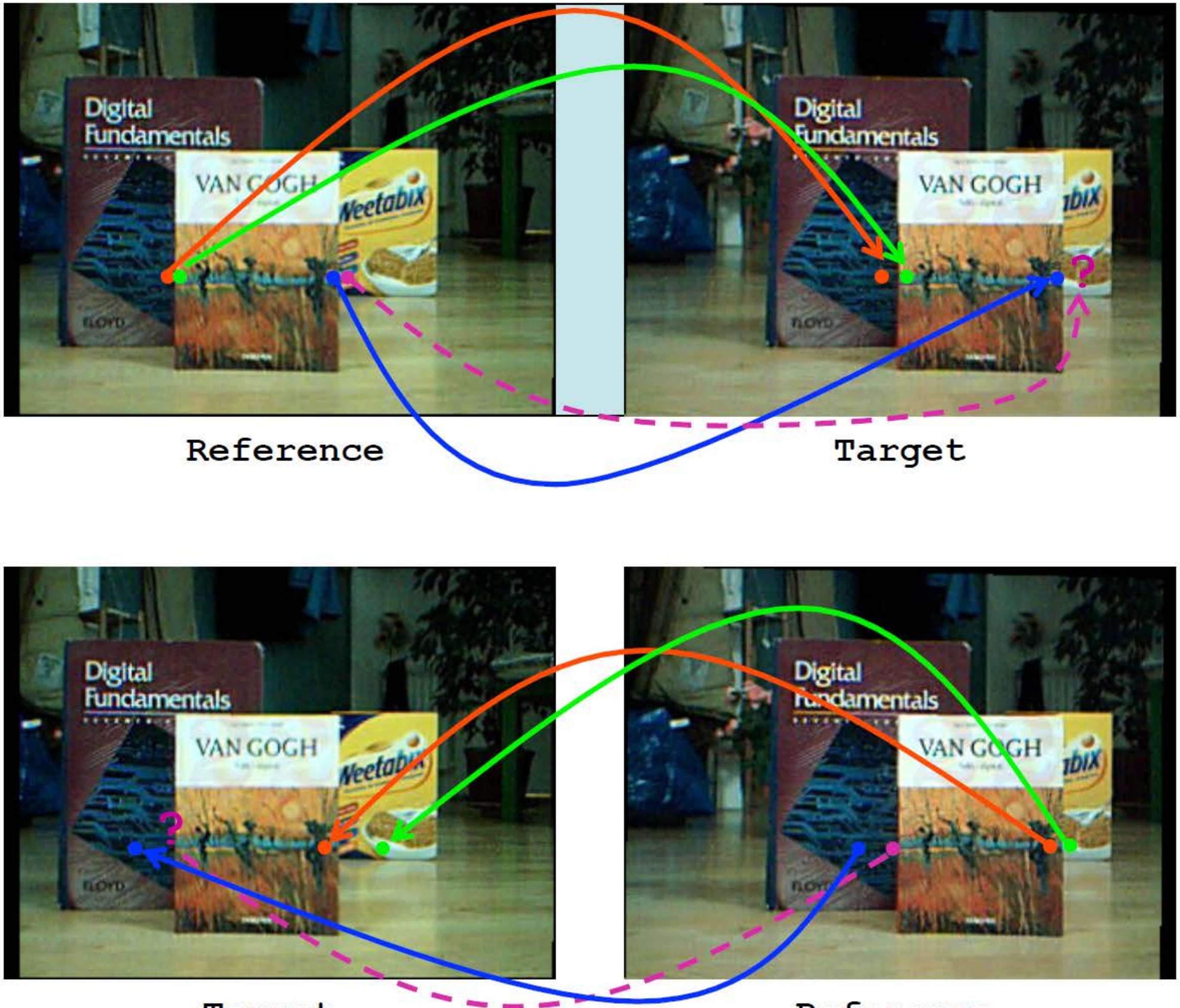


How to reduce ambiguity... ?

Transparent objects

Occlusions and discontinuities 1/2

Occlusions and discontinuities 2/2



Computing depth from images

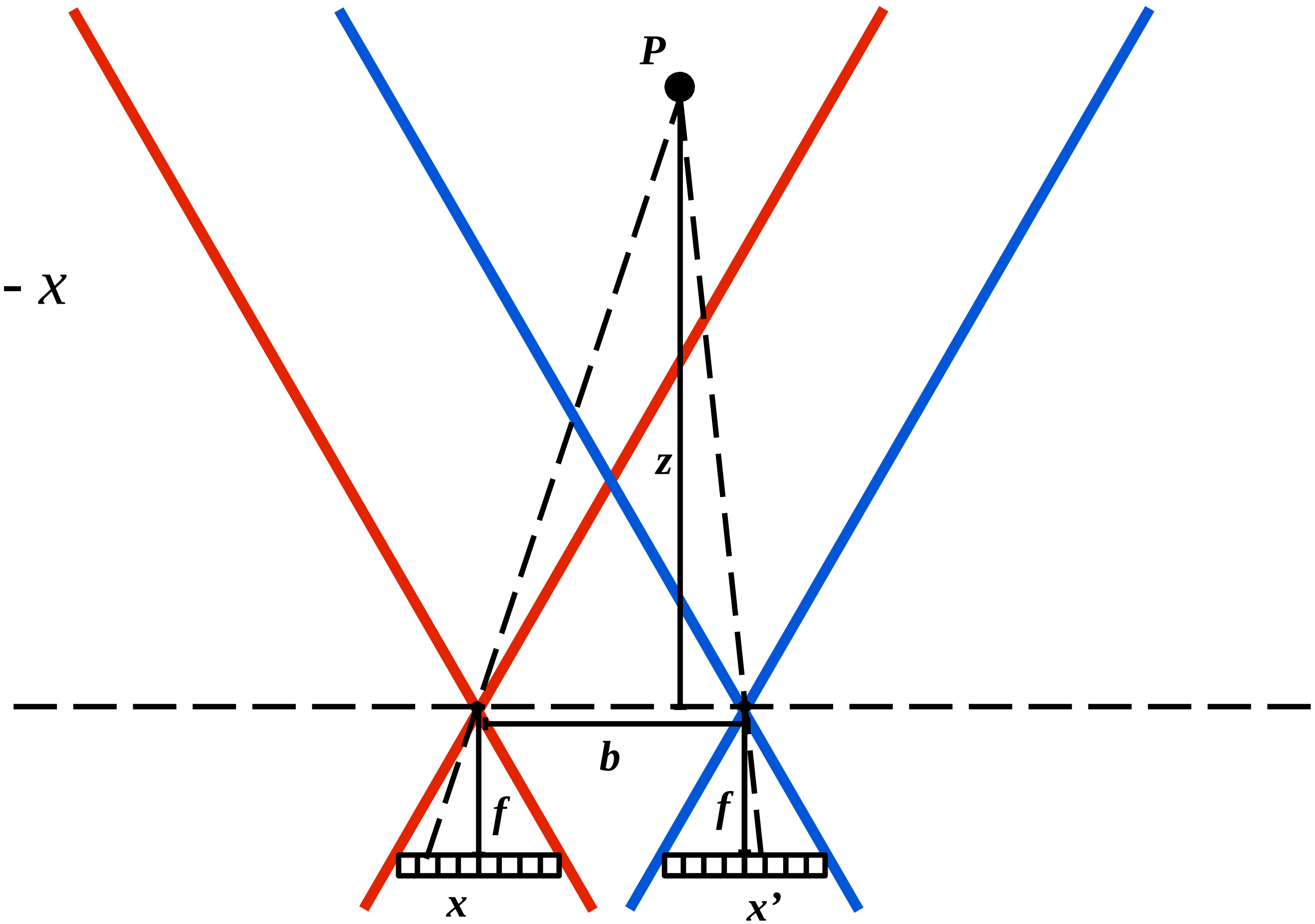
Binocular stereo 3D reconstruction of P: depth from disparity

Focal length: f

Baseline: b

Disparity: $d = x' - x$

$$z = \frac{bf}{d}$$



Simple reconstruction example: cameras aligned (coplanar sensors), separated by known distance, same focal length