CSCM77
Stereo

Prof. Xianghua Xie

x.xie@swansea.ac.uk

http://csvision.swan.ac.uk

Stereo vision

- Taking two images simultaneously from two different locations
- Objects appear in different position in each image depending on its depth in the scene (distance from the cameras)
- The position difference in the two images is known as **disparity**
- So, there is correlation between depth and disparity

- Stereo vision: inferring 3D structure from a pair of images taken from different viewpoints

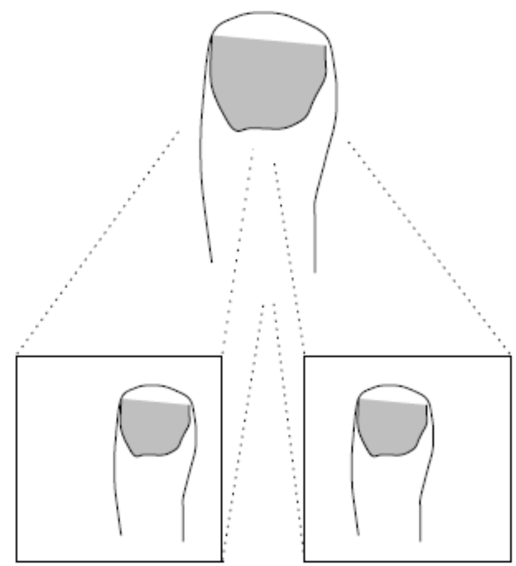


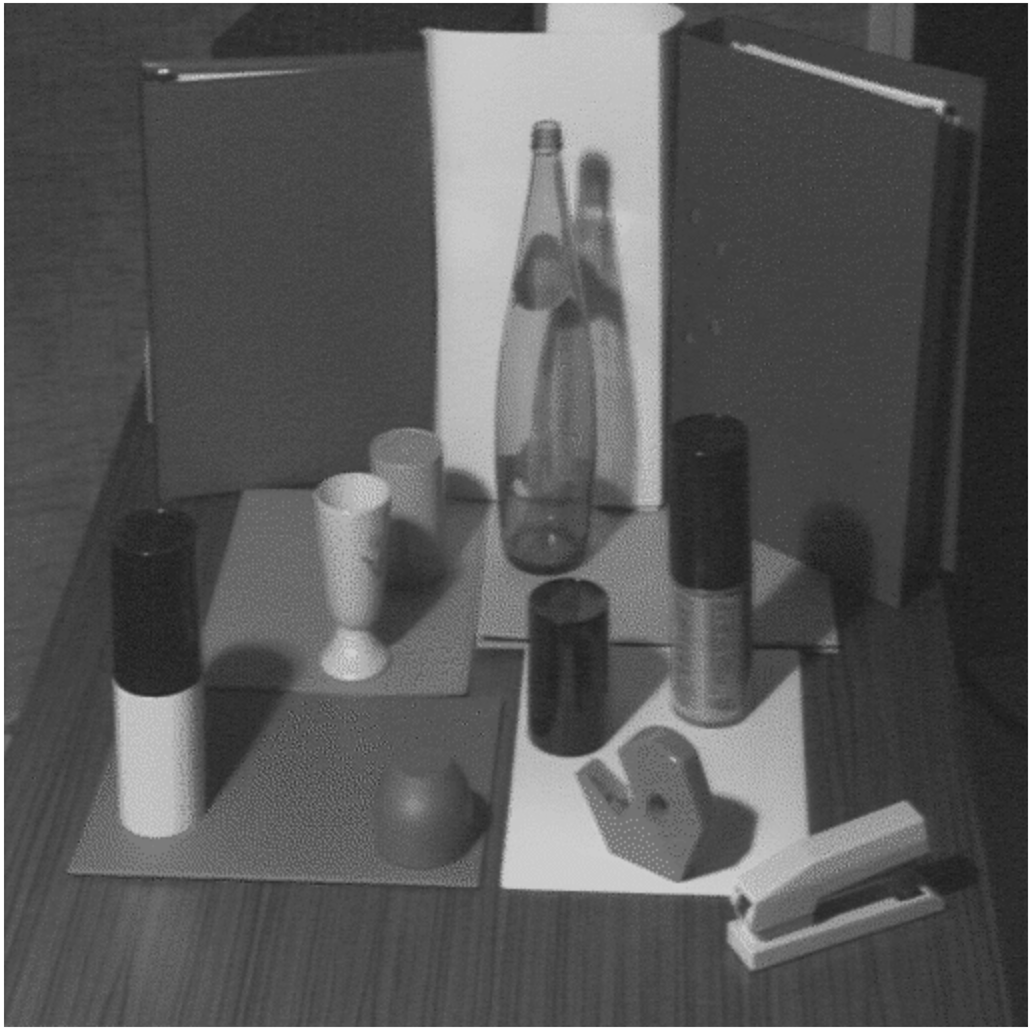
Image credit: Calway

Stereo vision

- A pair of stereo images

Stereo vision

- A pair of stereo images

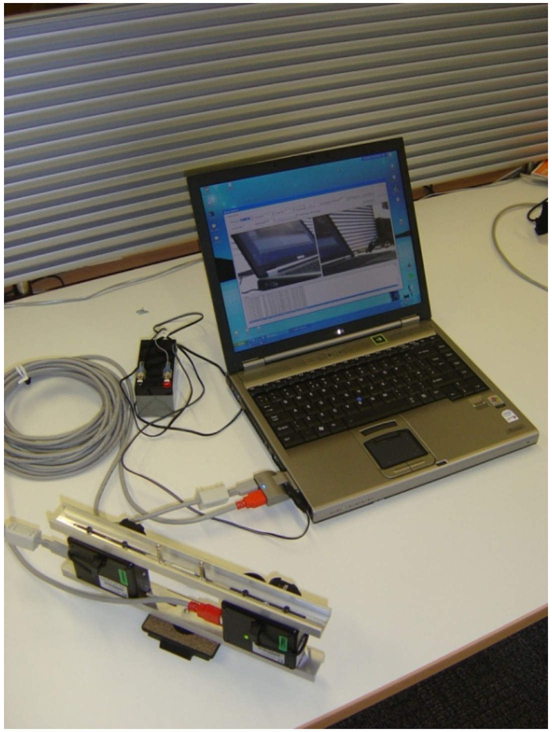


Stereo vision

- An example setup

- The distance between two cameras (central positions) are known as base line
- The example given is narrow base line set up

- We will only talk about narrow base line stereo



Stereo vision

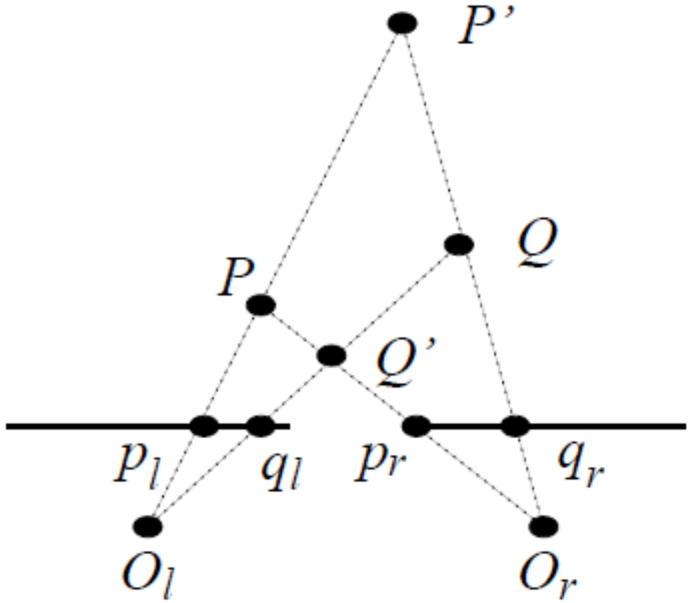
- The goal is to estimate depth from a pair of stereo images

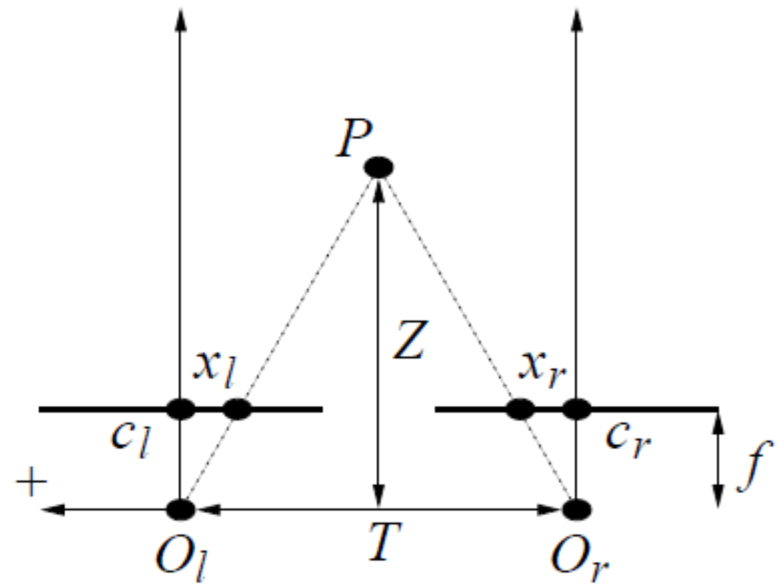
We need to solve two main problems

- Correspondence problem
  - For all elements in the left image, find their corresponding elements in the right image (or vice versa)
  - "elements": pixels, features (such as edge & corner), regions, objects and so on
- Reconstruction problem
  - Use the estimated disparities between elements to reconstruct 3D structure of the scene
  - Additional information about the camera and scene is necessary

A simple stereo system

- Parallel optical axes
  - Two cameras are perfectly aligned
- Use triangulation to work out the depth

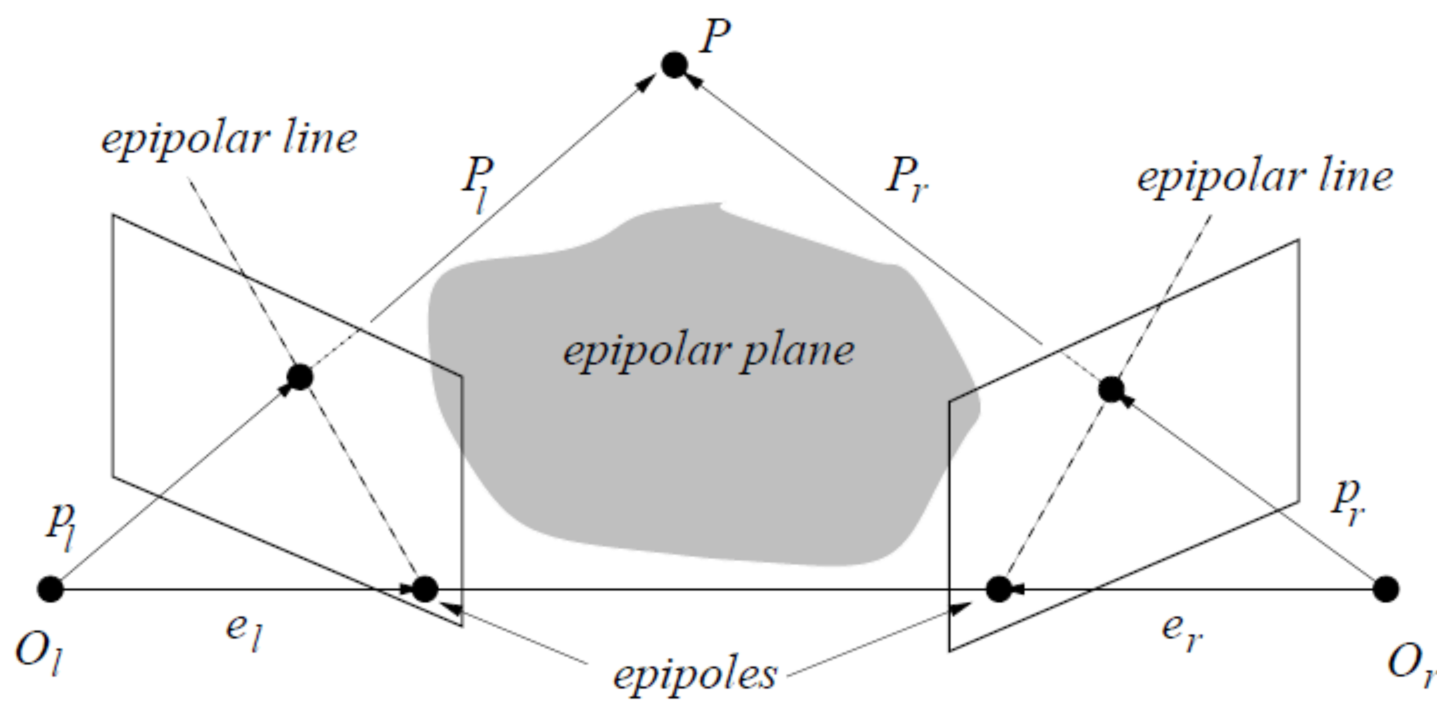




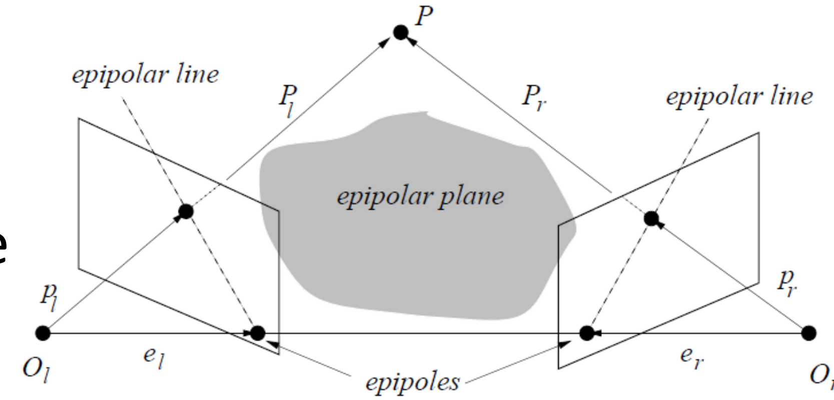
  - Similar triangles: $\frac{T+x_l-x_r}{Z-f} = \frac{T}{Z} \quad \rightarrow \quad Z = \frac{fT}{x_r-x_l}$
  - Unknowns: $d = x_r - x_l$, f, T, c_l and c_r

General stereo problem

- Epipolar geometry
  - Use pin-hole camera model
  - It is the intrinsic projective geometry between two views.
  - It is independent of scene structure, and only depends on the cameras' intrinsic parameters and relative pose (stereo extrinsic).



Stereo extrinsic parameters



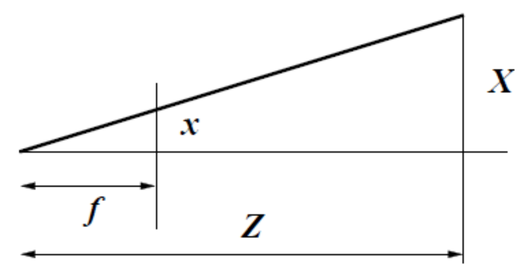
- Vectors P_l and P_r refer to the same 3D point P w.r.t. left and right camera frames, respectively
- The relationship between these two vectors are given by rotation matrix $\boldsymbol{R}$ and translation vector T:

$$\mathrm{P}_r = R(\mathrm{P}_l - \mathrm{T}) \quad (1)$$

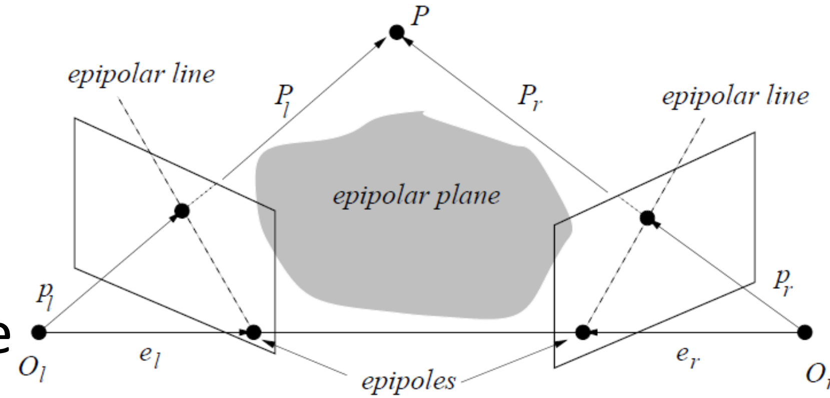
- The rotation matrix and translation vector define the extrinsic parameters of the stereo system
  - It basically tells us how cameras are positioned to each other

- Using perspective equation from previous lecture, we have:

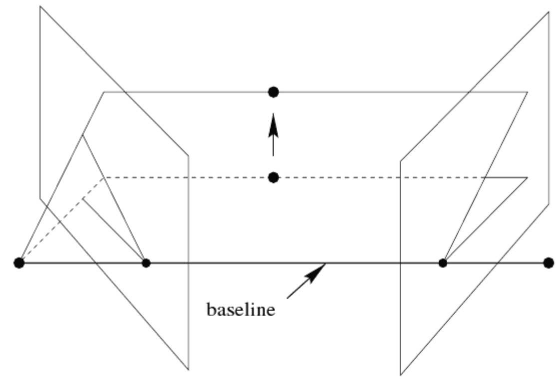
$$\mathrm{p}_l = f_l \mathrm{P}_l / Z_l \qquad \mathrm{p}_r = f_r \mathrm{P}_r / Z_r \quad (2)$$



Epipolar constraint



- Each 3D point defines an epipolar plane intersecting with each image plane along the epipolar line
- Given a point in the left image, the corresponding point in the right image must lie along its epipolar line
  - This is known as epipolar constraint

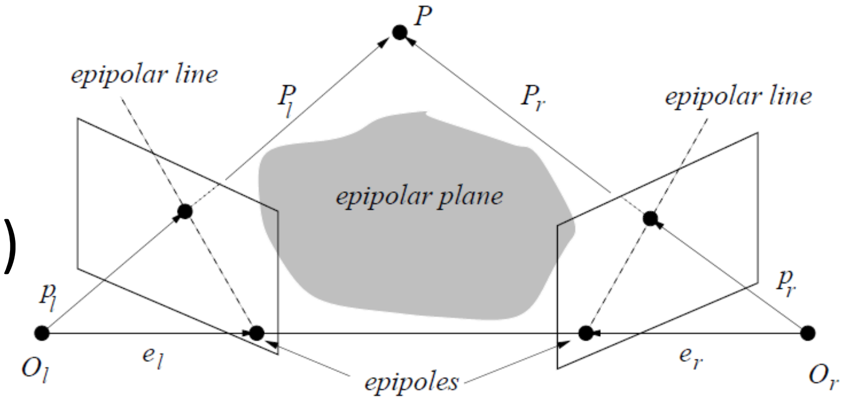


- It constrains locations of possible matches for points in either image
- It reduces a 2D correspondence search problem to 1D

Epipolar geometry

- So, epipolar constraint can help us to solve correspondence problem
- However, we need to know the intrinsic and extrinsic parameters
  - This is solved based on known correspondence: **calibration**
- Two important matrices
  - The Essential Matrix: defines relationship between an image point w.r.t. **camera coordinates** and the epipolar line
    - Given a point in one image plane, determine the epipolar line in the other image plane
  - The Fundamental Matrix: defines relationship between an image point w.r.t. **pixel coordinates** and the epipolar line
    - Given a point in pixel coordinates (actual digital image), determine the epipolar line in the other

The essential matrix



- The three vectors P_l, T and (P_l - T) all lie in the epipolar plane
- Thus:
  - Cross product (T X P_l) gives a vector perpendicular to the epipolar plane
  - Hence dot product between (P_l - T) and the cross product is 0

$$(\mathrm{P}_l - \mathrm{T})^T(\mathrm{T} \times \mathrm{P}_l) = 0 \quad \rightarrow \quad (R^T\mathrm{P}_r)^T(\mathrm{T} \times \mathrm{P}_l) = 0 \qquad (3)$$

- Cross product can be written as (to simplify notation):

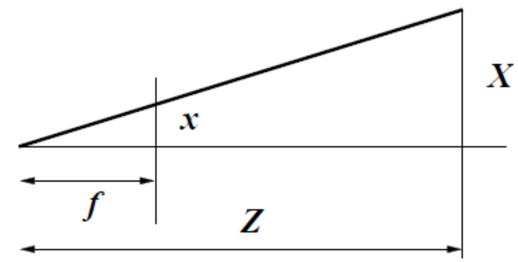
$$\mathrm{T} \times \mathrm{P}_l = S\mathrm{P}_l \;\rightarrow\; S = \begin{bmatrix} 0 & -T_z & T_y \\ T_z & 0 & -T_x \\ -T_y & T_x & 0 \end{bmatrix} \quad (4)$$

- Hence (3): $\mathrm{P}_r^T R(\mathrm{T} \times \mathrm{P}_l) = 0 \;\rightarrow\; \mathrm{P}_r^T E\mathrm{P}_l = 0 \quad E = RS$
  - E is the Essential Matrix

The essential matrix

- The essential matrix defines epipolar constraint in terms of extrinsic parameters of the stereo system

$$\mathrm{P}_r^T E \mathrm{P}_l = 0$$



- Recall perspective equation (2):

$$\mathrm{p}_l = f_l \mathrm{P}_l / Z_l \qquad \mathrm{p}_r = f_r \mathrm{P}_r / Z_r \qquad (2)$$

- Divide the top equation by $Z_r Z_l / f_r f_l$ and using (2):

$$\mathrm{p}_r^T E \mathrm{p}_l = 0$$

  – For a point p_l, the corresponding point must satisfy the above equation

The fundamental matrix

- Recall the mapping from image plane to pixel coordinates, given by (3) in the previous lecture:

$$x = (x_{im} - o_x)s_x \qquad y = (y_{im} - o_y)s_y \quad (3)$$

- Thus, we have:

$$\mathrm{p}_l = M_l^{-1}\bar{\mathrm{p}}_l \qquad M_l = \begin{bmatrix} 1/s_x & 0 & o_x/f \\ 0 & 1/s_y & o_y/f \\ 0 & 0 & 1 \end{bmatrix}$$

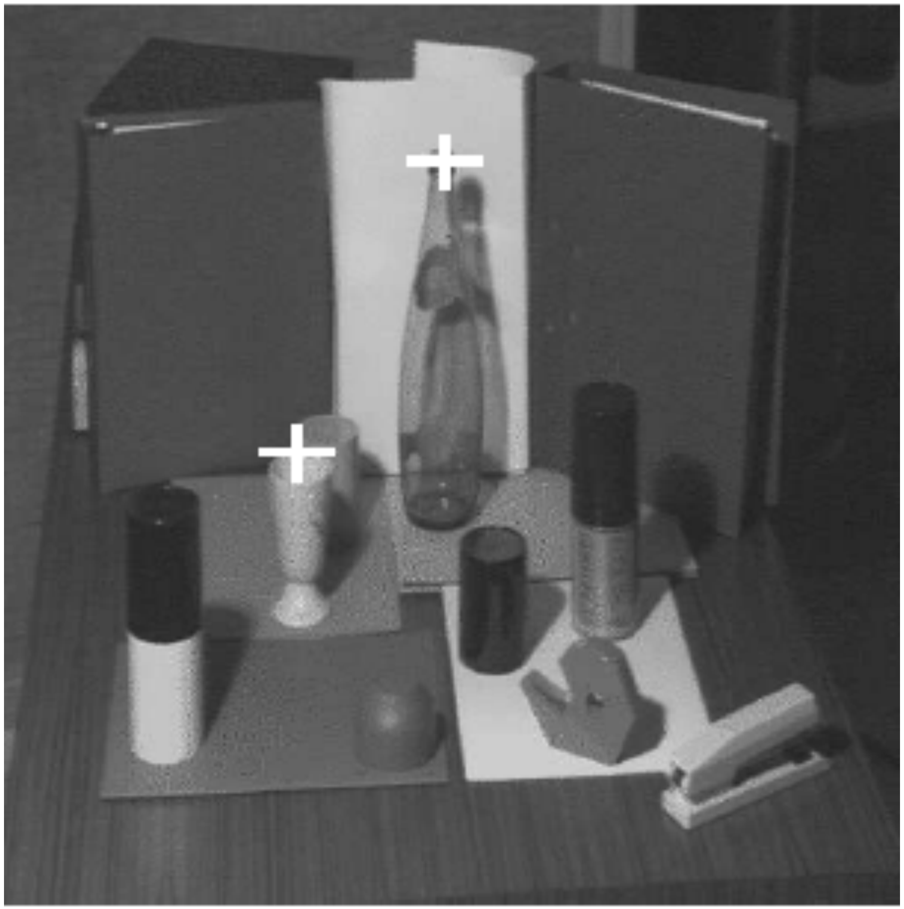
- $\mathrm{p}_r^T E \mathrm{p}_l = 0$ can then be re-written as:

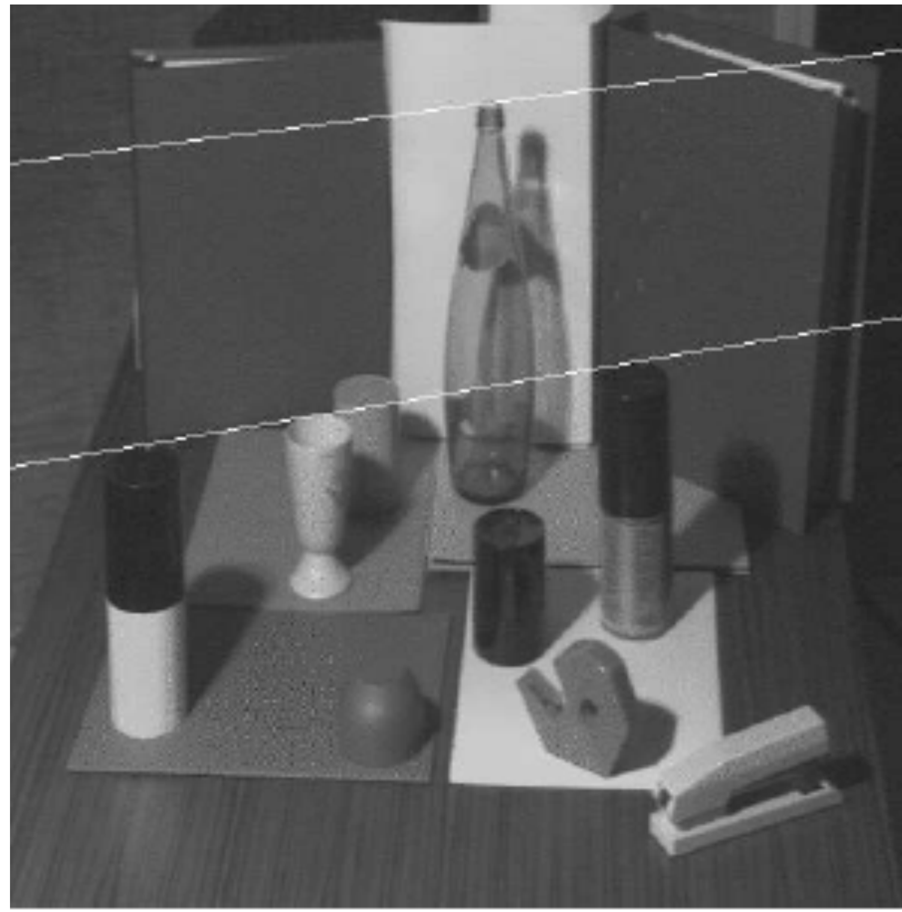
$$\bar{\mathrm{p}}_r^T F \bar{\mathrm{p}}_l = 0, \quad F = (M_r^{-1})^T E M_l^{-1}$$

  – F is known as the fundamental matrix, which defines epipolar line in pixel coordinates
  – Let $\bar{\mathrm{u}}_r = F\bar{\mathrm{p}}_l$, so the epipolar line in pixel coordinates is:

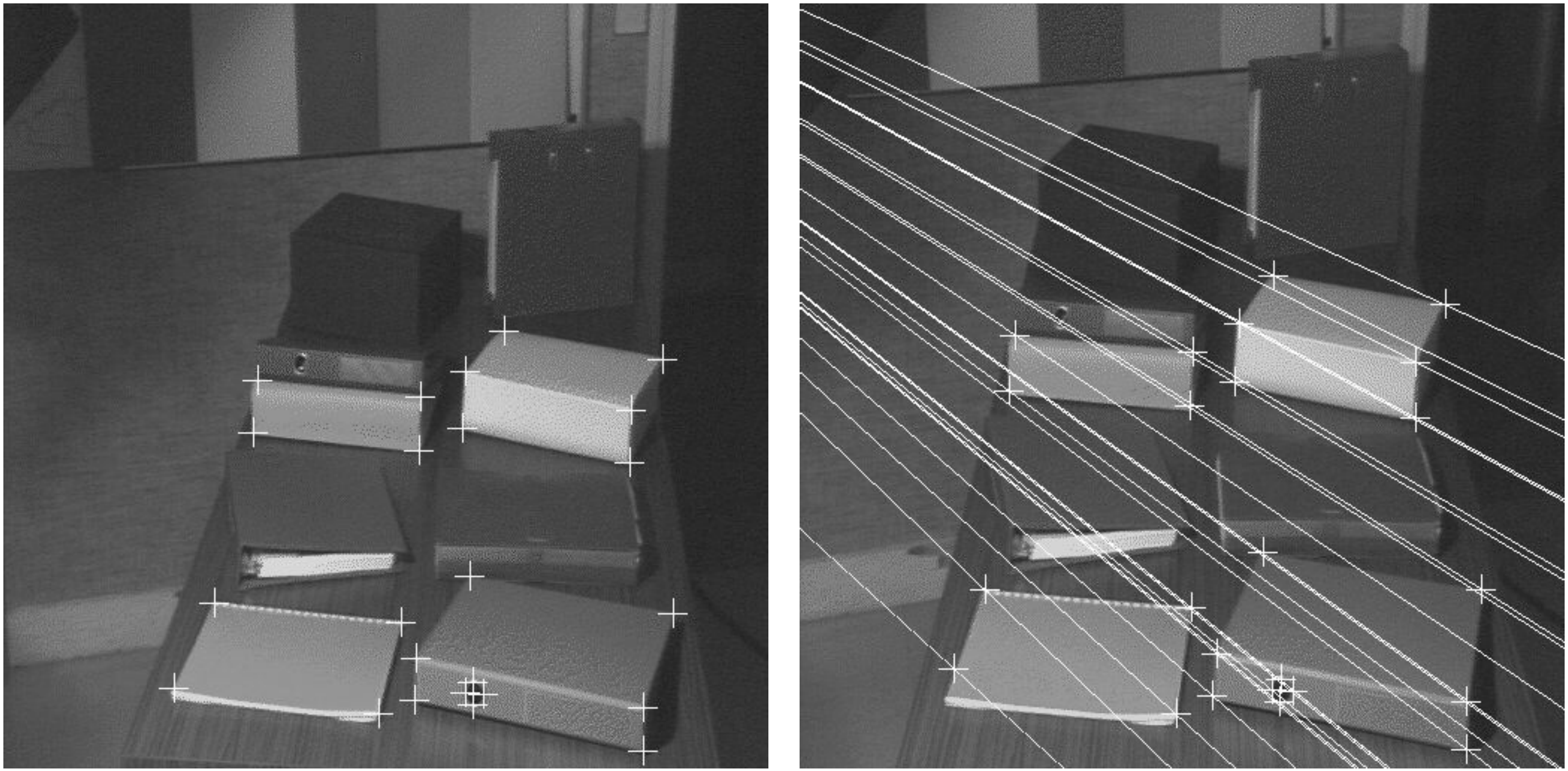
$$\bar{x}_r \bar{u}_{rx} + \bar{y}_r \bar{u}_{ry} + f\bar{u}_{rz} = 0$$

Example epipolar lines

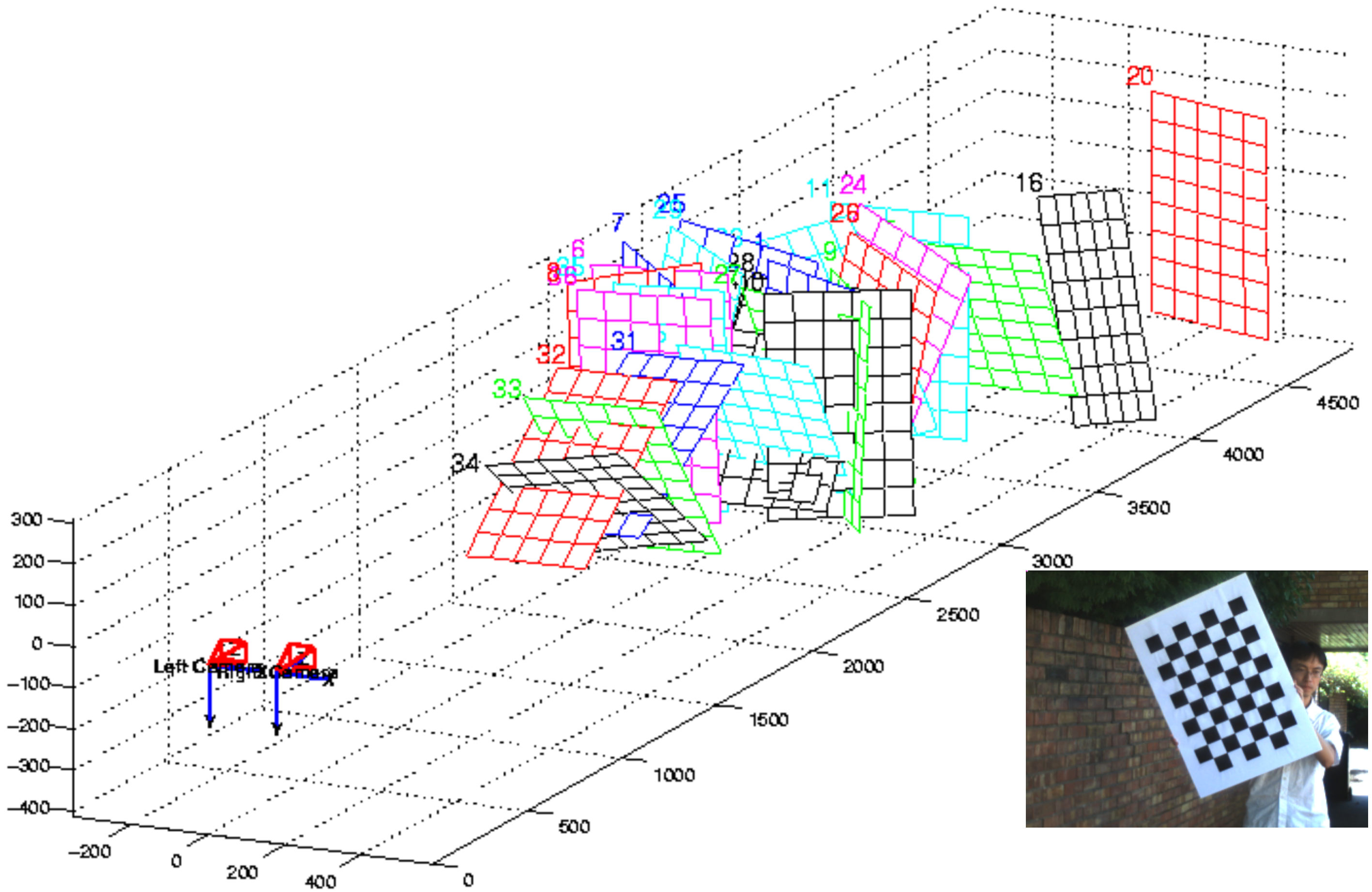


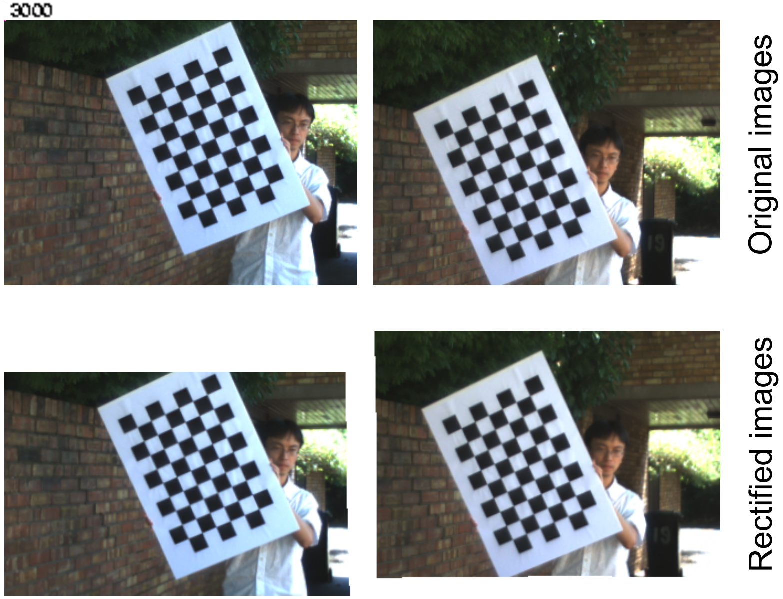


Example epipolar lines



Stereo Calibration





Stereo vision

- The goal is to estimate depth from a pair of stereo images

We need to solve two main problems

- Correspondence problem
  - For all elements in the left image, find their corresponding elements in the right image (or vice versa)
  - "elements": pixels, features (such as edge & corner), regions, objects and so on
- Reconstruction problem
  - Use the estimated disparities between elements to reconstruct 3D structure of the scene
  - Additional information about the camera and scene is necessary

The correspondence problem

- For all elements in the left image, find their corresponding elements in the right image (or vice versa)
- “elements”: pixels, features (such as edge & corner), regions, objects and so on

- Locations of corresponding points can be used to determine the 3D positions of those points

- Three main design decisions
  – Which element to match: pixel, edge, corner, region …
  – How to search for matching elements
  – How to determine a match

- A variety of approaches: depends on application and scene

Challenges

- Occlusions
  - some elements may not have corresponding element due to shape of object or presence of another object
  - depends on rate of change in depth
  - separation of optical centres
- Ambiguities (false matches)
  - similar elements in vicinity of ‘true’ match
- Change in element characteristics between views, due to:
  - change in lighting direction (change in intensity)
  - foreshortening effects (change in size)
  - uncorrelated noise effects

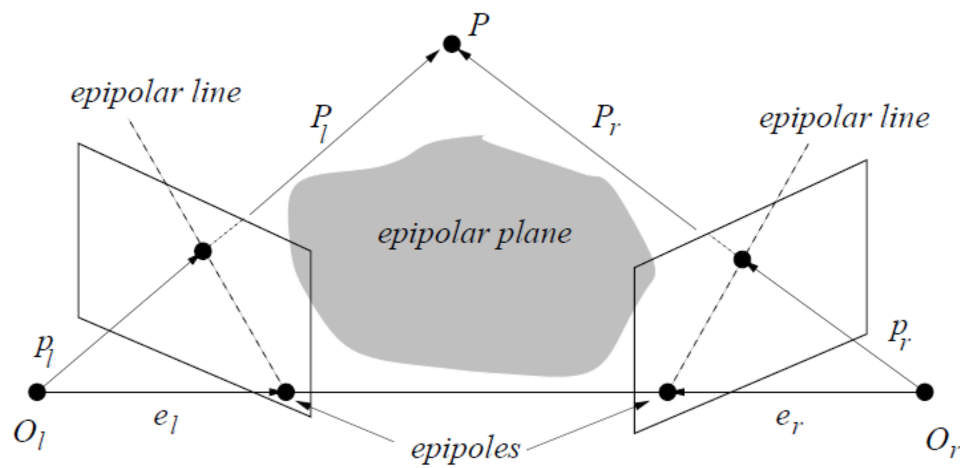
- Requires use of additional constraints and assumptions

Feature based

- Key steps
  - Feature extraction, e.g. edge, corner detection and thresholding to get reliable features
  - For feature in the left image, compute similarity with all features in the right image within certain region
  - Choose correspondent point with the maximum similarity measure
  - Repeat above process for all feature points in the left image
- Performance
  - Choice of feature is very important
  - Compared to correlation based, generally more robust to lighting and size changes
  - Perform poorly on textured regions, particularly random textures
  - Produce sparse disparities

Epipolar constraint

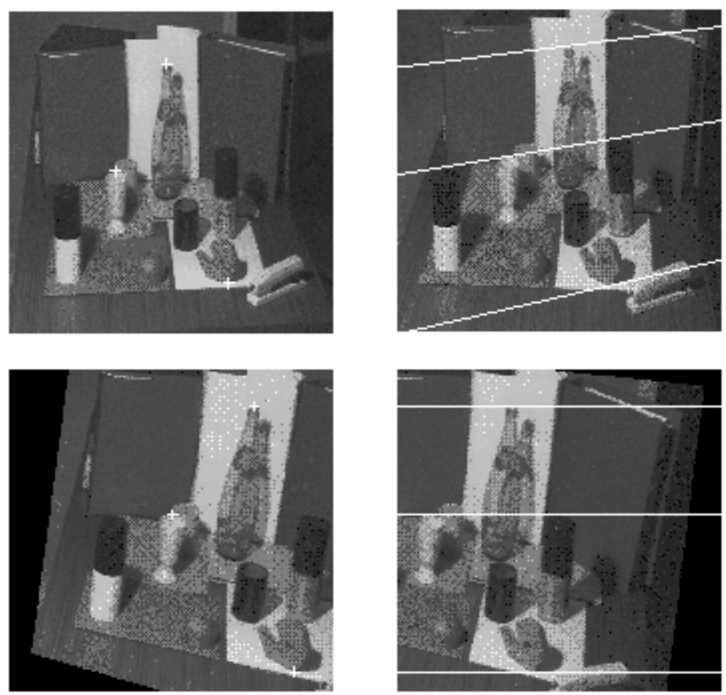
- According to epipolar constraint, corresponding points must lie on the epipolar lines



- Thus limit the search for corresponding region or feature
  - Can then apply correlation or feature based methods
  - 2D search becomes 1D search

Epipolar constraint

- Need to know all the intrinsic and extrinsic parameters.
- Rectification
  - Often both images are warped so that all the epipolar lines are horizontal
  - Only need to search correspondence along the same row in the two images

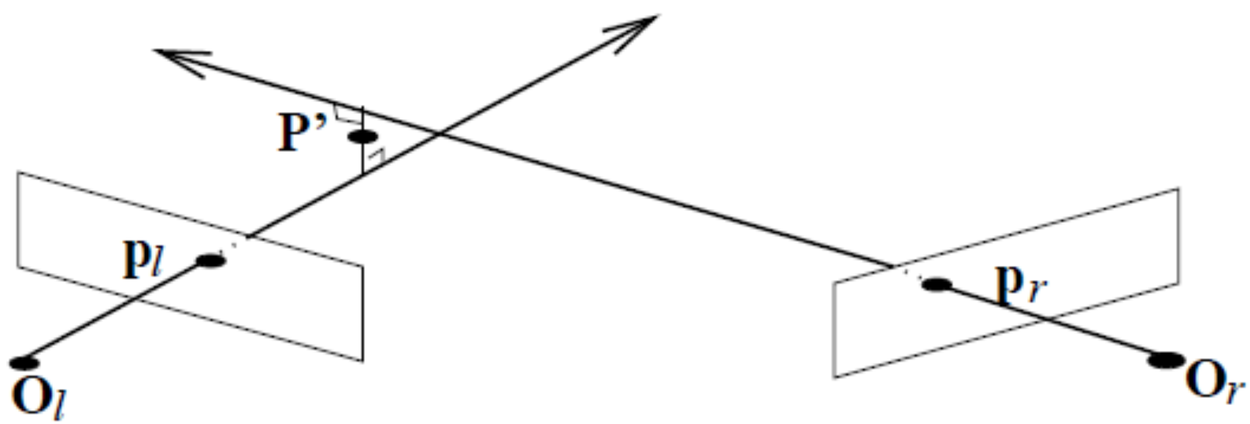


Other constraints

- Spatial continuity
  - Smooth variation of disparity over image plane, i.e. smooth variation of depth
  - Be careful with genuine depth discontinuities
- Uniqueness
  - Pixel in one image can only correspond to only one pixel in the other
  - Exceptions, e.g. occlusions
- Ordering constraint
  - Matching points along corresponding epipolar lines should be in the same order along both lines
  - Exceptions, e.g. occlusions
- Left-right consistency
  - Perform left-to-right match and then right-to-left; only accept matches supported by both

3D reconstruction

- We assume that the stereo is calibrated, i.e. the parameters are known
- Use triangulation to determine the 3D points
- However, in general they do not intersect
  - Need to find the closest point to both rays (mid point of the ray perpendicular to both)

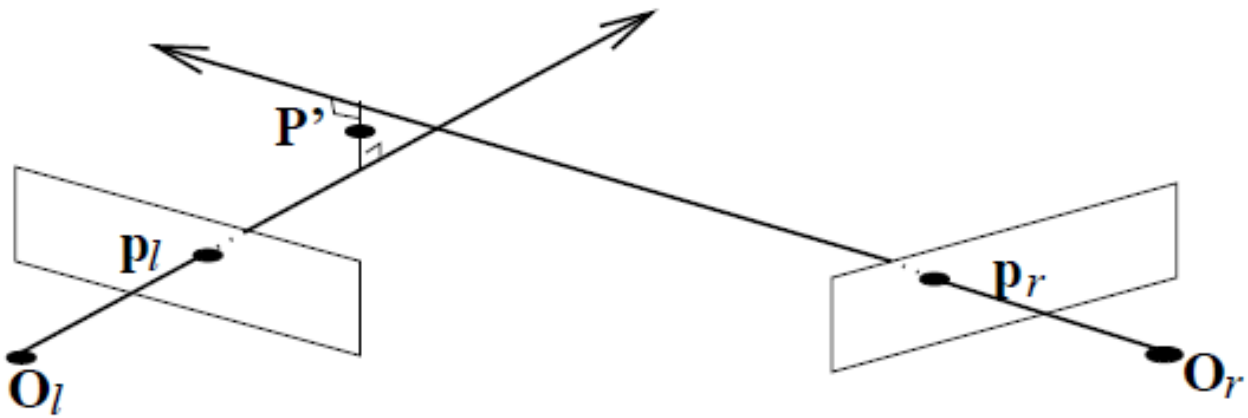


3D reconstruction

Exercise

- Prove that find the mid-point is to solve the following linear equation (find a, b, c scaling factors; 3 linear equations)

$$a\mathrm{p}_l - bR^T\mathrm{p}_r - \mathrm{T} + c(\mathrm{p}_l \times R^T\mathrm{p}_r) = 0$$



Depth estimation using stereo

- Example (video)