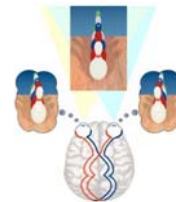


Introduction to Computer Vision

CS / ECE 181B

Tuesday, May 11, 2004



Today

- Multiple view geometry and stereo
- Handout #6 available (check with Isabelle)

Ack: M. Turk and M. Pollefeys

Midterm

Seeing in 3D

- Humans can perceive depth, shape, etc. – 3D properties of the world
 - How do we do it?
- We use many cues
 - Oculomotor convergence/divergence →
 - Accomodation (changing focus)
 - Motion parallax (changing viewpoint)
 - Monocular depth cues
 - ♦ Occlusion, perspective, texture gradients, shading, size
 - Binocular disparity (stereo)
- How can computers perceive depth?

May 2004

Stereo

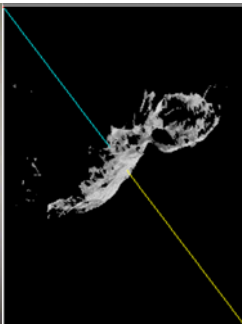
3



May 2004

Stereo

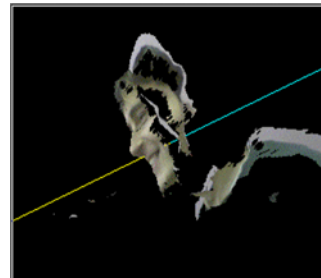
4



May 2004



Stereo



Multiple views and depth



May 2004

Stereo

Why multiple views?

- A camera projects the 3D world into 2D images
- This is not always a problem – humans can figure out a lot from a 2D view!

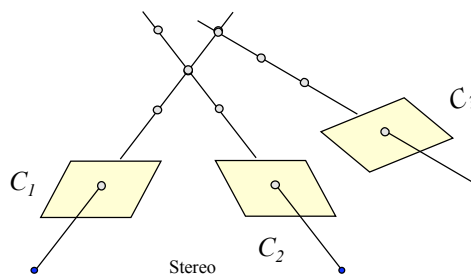


May 2004

7

Why multiple views?

- But precise 3D information (distance, depth, shape, curvature, etc.) is difficult or impossible to obtain from a single view
- In order to measure distances, sizes, angles, etc. we need multiple views (and calibrated cameras!)
 - Monocular $\rightarrow$ binocular $\rightarrow$ trinocular...

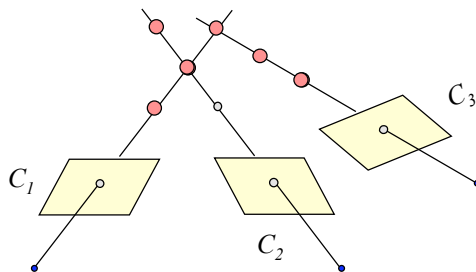


May 2004

8

Multiple view geometry

- There are many possible configurations of scene points that could have created corresponding points in multiple views



- Two big questions for multiple view geometry problems:
 - Which are possible?
 - Which are most likely?

May 2004

Stereo

9

M. Pollefeys

Questions

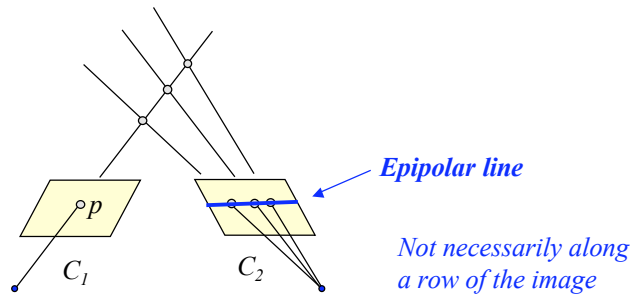
- **Correspondence geometry:** Given an image point x in the first view, how does this constrain the position of the corresponding point x' in the second image?
- **Camera geometry (motion):** Given a set of corresponding image points $\{x_i, x'_i\}$, $i=1, \dots, n$, what are the cameras P and P' for the two views?
- **Scene geometry (structure):** Given corresponding image points x_i, x'_i and cameras P, P' , what is the position of (their pre-image) X in space?

May 2004

Stereo

10

Two-view geometry



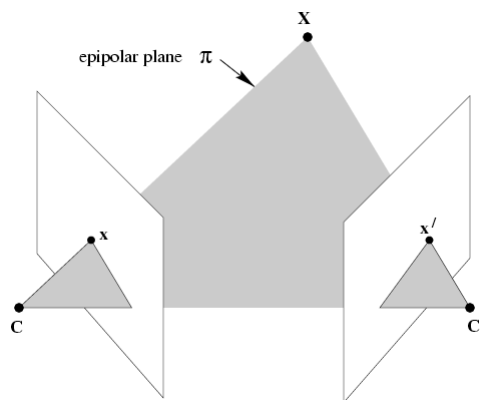
- The **epipolar geometry** is defined by the origins of the camera coordinate frames, the scene point P , and the locations of the image planes

May 2004

Stereo

11

Epipolar geometry

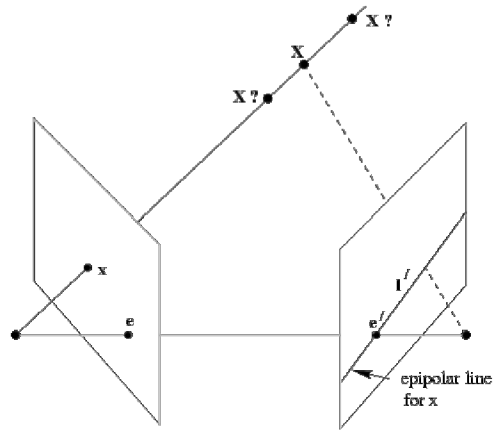


May 2004

Stereo

12

Epipolar Geometry



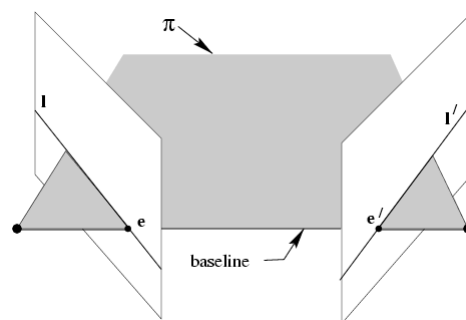
What if only C, C', x are known?

May 2004

Stereo

13

Epipolar Geometry



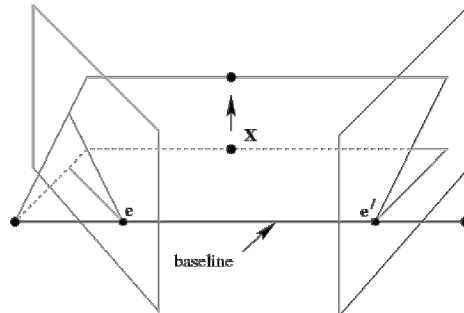
All points on π project on l and l'

May 2004

Stereo

14

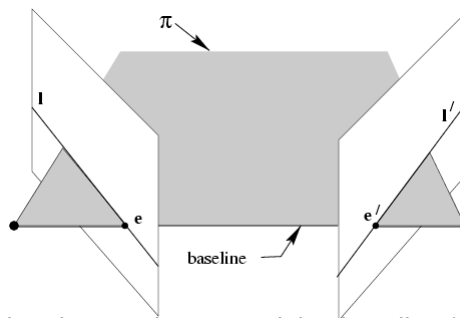
Epipolar Geometry



Family of planes π and lines l and l'
Intersection in e and e'

Epipolar geometry

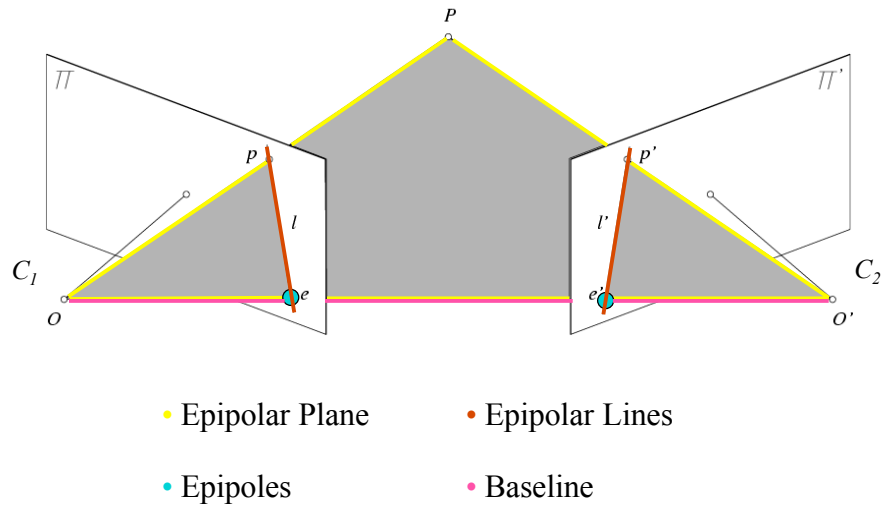
epipoles e, e'
= intersection of baseline with image plane
= projection of projection center in other image
= vanishing point of camera motion direction



an epipolar plane = plane containing baseline (1-D family)

an epipolar line = intersection of epipolar plane with image
(always come in corresponding pairs)

Epipolar geometry

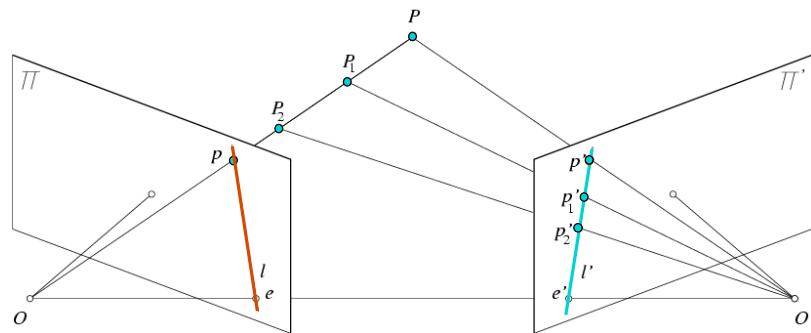


May 2004

Stereo

17

Epipolar constraint



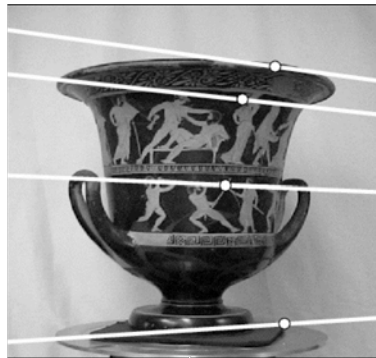
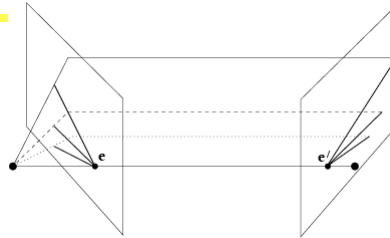
- Potential matches for p have to lie on the corresponding epipolar line l'
- Potential matches for p' have to lie on the corresponding epipolar line l

May 2004

Stereo

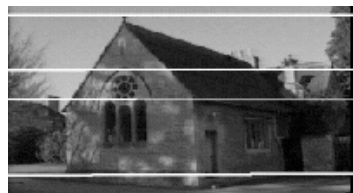
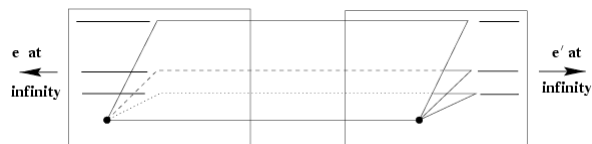
18

Example: converging cameras



May 2004

Example: motion parallel with image plane

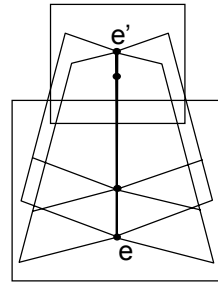
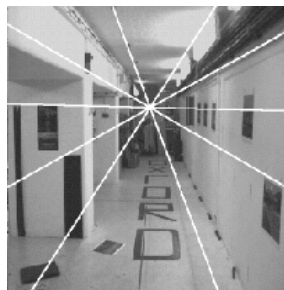
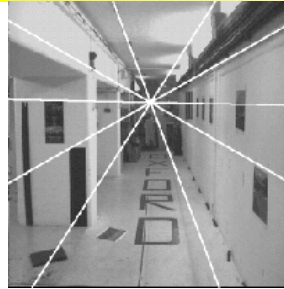


May 2004

Stereo

20

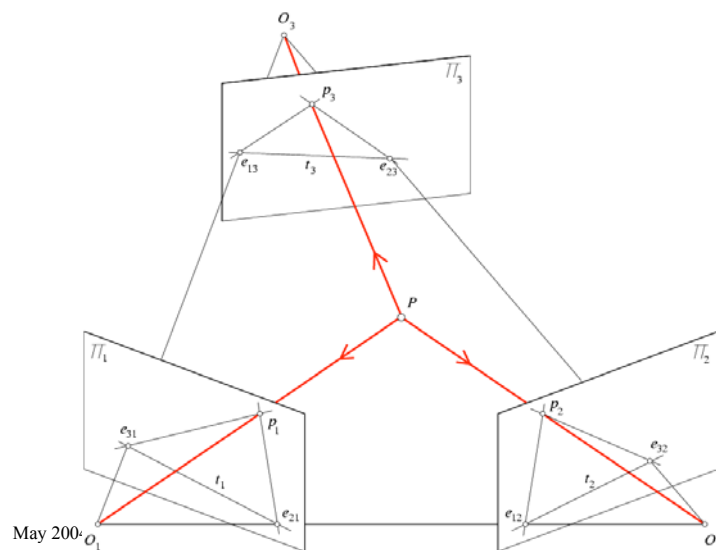
Example: forward motion



May 2004

21

Trinocular epipolar constraint



May 2004

22

Basic approach to stereo vision

- Find features of interest in N image views
 - The “correspondence problem”
- Triangulate
 - A method to measure distance and direction by forming a triangle and using trigonometry
- Reconstruct object/scene depth
 - From dense points
 - From sparse points



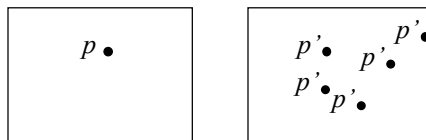
May 2004

Stereo

23

Step 1: The correspondence problem

- Given a “point” in one image, find the location of that same point in a second image (and maybe third, and fourth, ...)



A search problem: Given point p in the left image, where in the right image should we search for a corresponding point?

Sounds easy, huh?

May 2004

Stereo

24

Correspondence problem



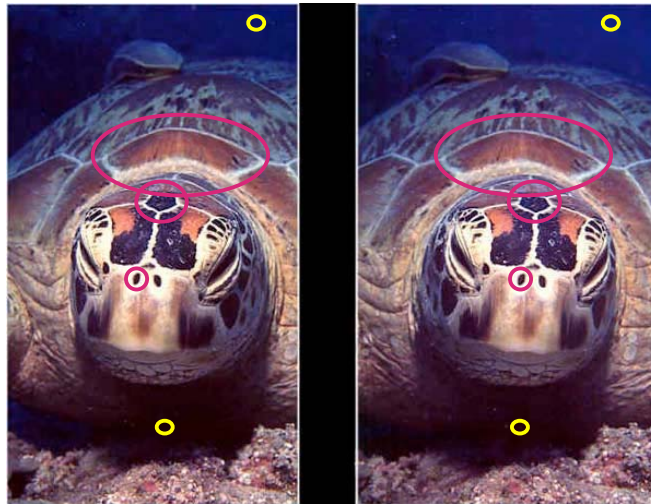
Left image



Right image

- *What is a point?*
- *How do we compare points in different images? (Similarity measure)*

Correspondence problem



Right image

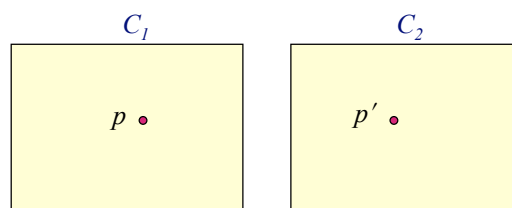
Left image

The correspondence problem

- A classically difficult problem in computer vision
 - Is every point visible in both images?
 - Do we match *points* or *regions* or ...?
 - Are corresponding (L-R) image regions similar?
- Correspondence is easiest when the depth is large compared with the camera baseline distance
 - Because the cameras then have about the same viewpoint
 - But...
- Two classes of stereo correspondence algorithms:
 - Feature based (sparse) – corners, edges, lines, ...
 - Correlation based (dense)
 - ♦ How large a window of support to use?

Multiple views

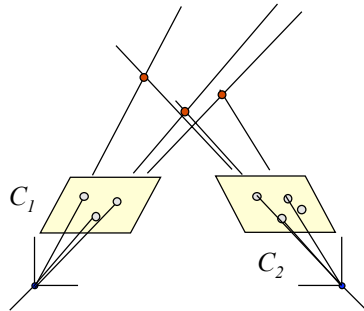
- What do you need to know in order to calculate the depth (or location) of the point that causes p and p' ?



- Values of $p = (u, v)$ and $p' = (u', v')$
- Locations of C_1 and C_2 (full extrinsic parameters)
 - Rigid transformation between C_1 and C_2
- Intrinsic parameters of C_1 and C_2

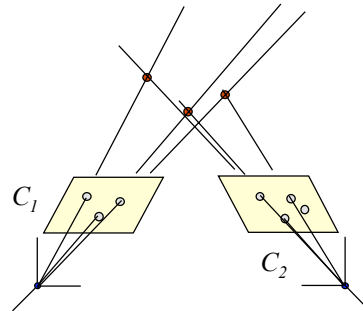
Duality: Calibration and stereo

- Given calibrated cameras, we can find depth of points
- Given corresponding points, we can calibrate the cameras



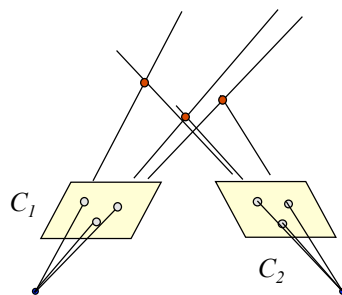
May 2004

Stereo



29

Example: Extrinsic parameters from 3 points



1 known point

2 known points

3 known points

In this case, we know the point correspondences and the point distances.

If we only know the correspondences, we'll need at least five points

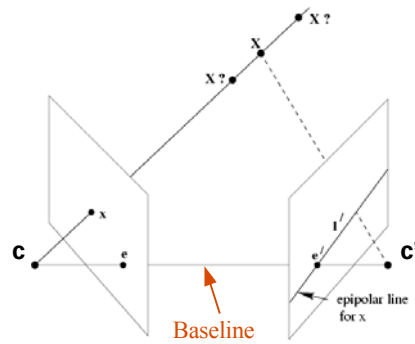
May 2004

Stereo

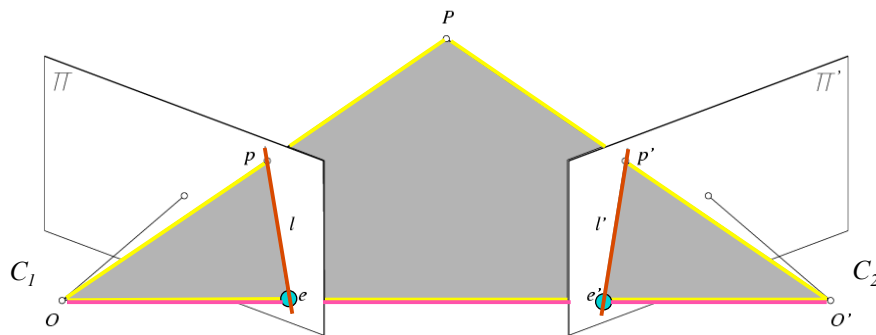
30

The geometry of multiple views

- Epipolar Geometry
 - The Essential Matrix
 - The Fundamental Matrix
- The Trifocal Tensor
- The Quadrifocal Tensor

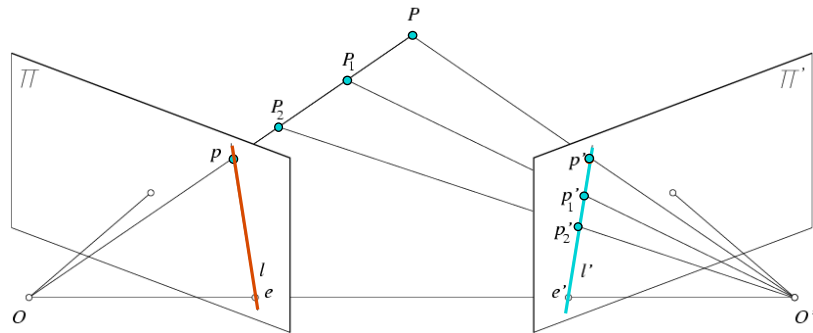


Epipolar geometry



- Epipolar Plane
- Epipolar Lines
- Epipoles
- Baseline

Epipolar constraint



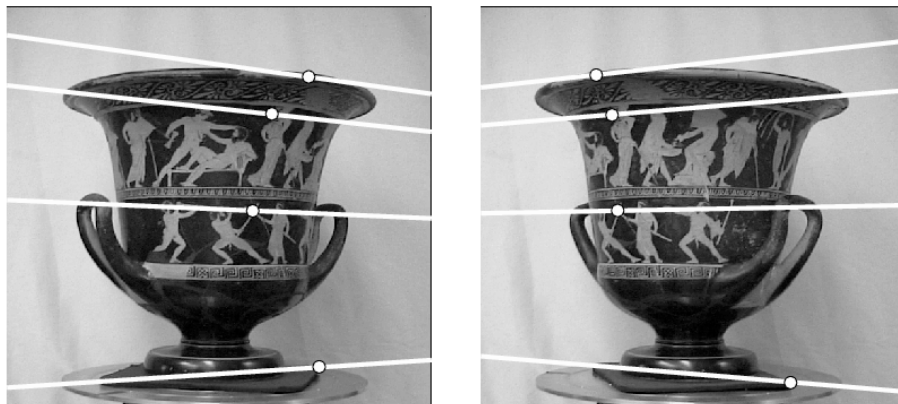
- Potential matches for p have to lie on the corresponding epipolar line l'
- Potential matches for p' have to lie on the corresponding epipolar line l

May 2004

Stereo

33

Epipolar lines example



May 2004

Stereo

34

Matrix form of cross product

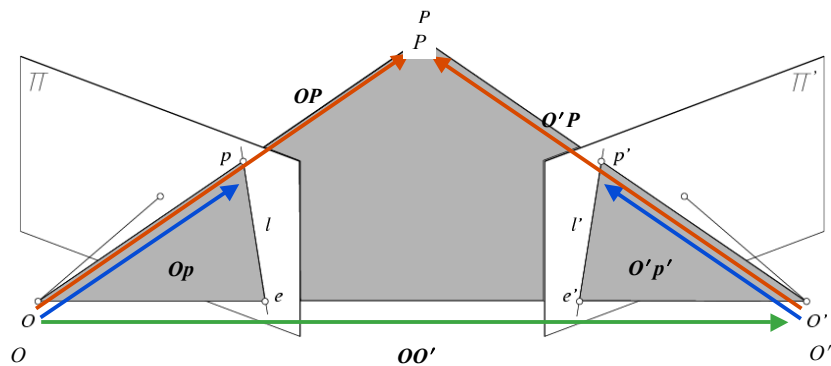
- The cross product of two vectors is a third vector, perpendicular to the others (right hand rule)

$$a \times b = \begin{bmatrix} a_2 b_3 - a_3 b_2 \\ a_3 b_1 - a_1 b_3 \\ a_1 b_2 - a_2 b_1 \end{bmatrix} = \begin{bmatrix} 0 & -a_3 & a_2 \\ a_3 & 0 & -a_1 \\ -a_2 & a_1 & 0 \end{bmatrix} b = [a_{\times}] b$$

$$a \cdot (a \times b) = 0$$

$$b \cdot (a \times b) = 0$$

Case 1: Calibrated camera



$$Op \cdot (OO' \times O'p') = ?$$

$$Op \cdot (OO' \times O'p') = 0$$

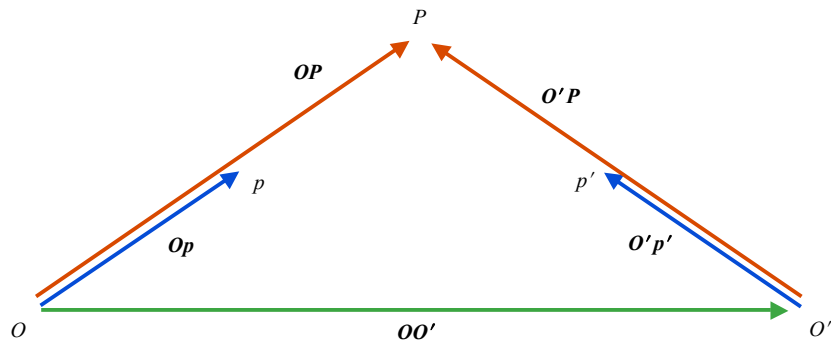
$[R \ t]$ – rigid trans. from O to O'

$$p \cdot (t \times R p') = 0$$

This can be written in matrix form as:

$$p^T E p' = 0$$

Essential Matrix



$$p \cdot (t \times R p') = 0$$

$$p^T E p' = 0$$

$$E = \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix} R = \begin{pmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{pmatrix} R \quad \text{\textcolor{brown}{E} - Essential Matrix}$$

May 2004

Stereo

37

The Essential Matrix

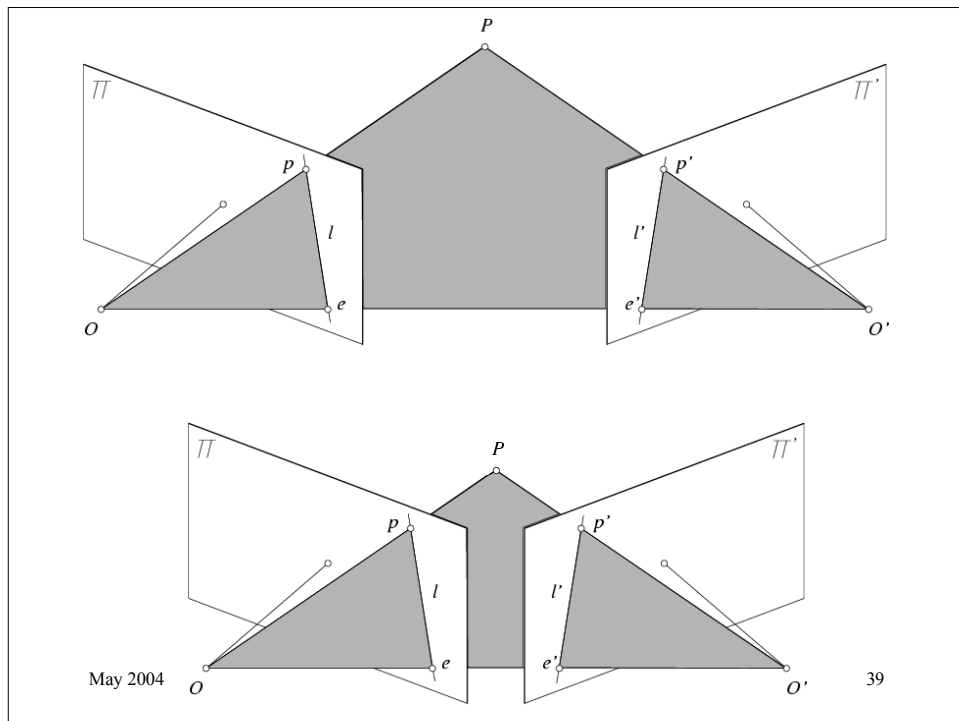
- E describes the transformation between camera coordinate frames
- E has five degrees of freedom
 - Defined up to a scale factor, since

$$p^T E p' = 0$$
- Why only five?
 - A rigid transformation has six degrees of freedom
- 3 rotation parameters, 2 translation direction parameters
 - Why only translation direction?

May 2004

Stereo

38

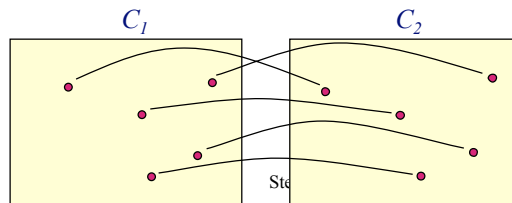


“Up to a scale factor”

- This is always the case with camera calibration and stereo
 - Shrink everything 10x and it all looks the same!
- Typically there is *something* we know that we can use to specify the scale factor
 - E.g., the baseline, the size of an object, the depth of a point/plane

Camera calibration from E

- With five unknowns, theoretically we can recover the essential matrix E by writing $p^T E p' = 0$ for five corresponding pairs of points
 - 5 equations and 5 unknowns
 - We don't need to know anything about the points (e.g., their depth), only that they project to p_i and p'_i
 - There are, however, limitations...
- This is used for camera calibration (extrinsic parameters)



May 2004

41

Case 2: Uncalibrated camera

- Intrinsic parameters not known

$$p = K_1 \hat{p}$$

$$p' = K_2 \hat{p}'$$

Points in the *normalized* image plane

$$K = \begin{bmatrix} \alpha & -\alpha \cot \theta & u_0 \\ 0 & \frac{\beta}{\sin \theta} & v_0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\hat{p}^T E \hat{p}' = 0$$

$$(K_1^{-1} p)^T E (K_2^{-1} p') = 0$$

$$(p^T K_1^{-T}) E (K_2^{-1} p') = 0$$

$$p^T F p' = 0$$

$$F = K_1^{-T} E K_2^{-1} \quad \text{Fundamental Matrix}$$

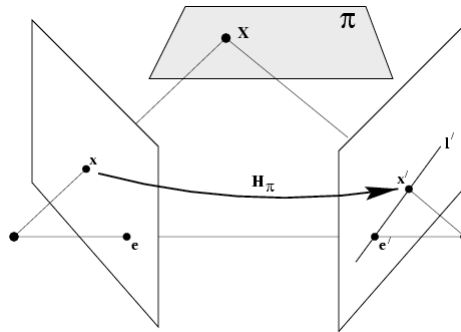
May 2004

Stereo

42

Fundamental Matrix F

geometric derivation



$$x' = H_{\pi} x$$

$$l' = e' \times x' = [e']_{\times} H_{\pi} x = Fx$$

mapping from 2-D to 1-D family (rank 2) ⁴³

May 2004

The Fundamental Matrix

- F has seven independent parameters
- A simple, linear technique to recover F from corresponding point locations is the “eight point algorithm”
- From F , we can recover the epipolar geometry of the cameras
 - Not saying how...
- This is called *weak calibration*

May 2004

Stereo

44

The eight-point algorithm

$$p^T F p' = 0$$

$$(u, v, 1) \begin{pmatrix} F_{11} & F_{12} & F_{13} \\ F_{21} & F_{22} & F_{23} \\ F_{31} & F_{32} & F_{33} \end{pmatrix} \begin{pmatrix} u' \\ v' \\ 1 \end{pmatrix} = 0 \quad \Rightarrow \quad (uu', uv', u, vv', vv', v, u', v', 1) \begin{pmatrix} F_{11} \\ F_{12} \\ F_{13} \\ F_{21} \\ F_{22} \\ F_{23} \\ F_{31} \\ F_{32} \\ F_{33} \end{pmatrix} = 0$$

$$\begin{pmatrix} u_1 u'_1 & u_1 v'_1 & u_1 & v_1 u'_1 & v_1 v'_1 & v_1 & u'_1 & v'_1 \\ u_2 u'_2 & u_2 v'_2 & u_2 & v_2 u'_2 & v_2 v'_2 & v_2 & u'_2 & v'_2 \\ u_3 u'_3 & u_3 v'_3 & u_3 & v_3 u'_3 & v_3 v'_3 & v_3 & u'_3 & v'_3 \\ u_4 u'_4 & u_4 v'_4 & u_4 & v_4 u'_4 & v_4 v'_4 & v_4 & u'_4 & v'_4 \\ u_5 u'_5 & u_5 v'_5 & u_5 & v_5 u'_5 & v_5 v'_5 & v_5 & u'_5 & v'_5 \\ u_6 u'_6 & u_6 v'_6 & u_6 & v_6 u'_6 & v_6 v'_6 & v_6 & u'_6 & v'_6 \\ u_7 u'_7 & u_7 v'_7 & u_7 & v_7 u'_7 & v_7 v'_7 & v_7 & u'_7 & v'_7 \\ u_8 u'_8 & u_8 v'_8 & u_8 & v_8 u'_8 & v_8 v'_8 & v_8 & u'_8 & v'_8 \end{pmatrix} \begin{pmatrix} F_{11} \\ F_{12} \\ F_{13} \\ F_{21} \\ F_{22} \\ F_{23} \\ F_{31} \\ F_{32} \end{pmatrix} = - \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \quad \text{Invert and solve for } F$$

May 2004 Stereo 45

Least squares approach

If $n > 8$

Minimize:

$$\sum_{i=1}^n (\mathbf{p}_i^T \mathcal{F} \mathbf{p}'_i)^2$$

under the constraint $|\mathbf{F}|^2 = 1$

Nonlinear least-squares approach

Minimize

$$\sum_{i=1}^n [d^2(\mathbf{p}_i, \mathcal{F}\mathbf{p}'_i) + d^2(\mathbf{p}'_i, \mathcal{F}^T \mathbf{p}_i)]$$

Point in image 1
Epipolar line in image 1 caused by $\mathbf{p}'$

with respect to the coefficients of $\mathbf{F}$

Nonlinear – initialize it from the results of the eight-point algorithm

Least squares 8-point algorithm



(a)

Hartley's normalized 8-point alg.



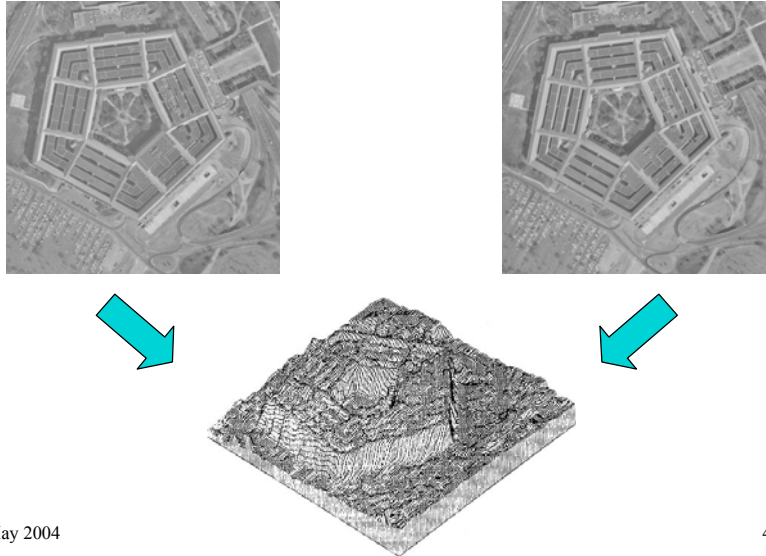
(b)

	Linear Least Squares	(Hartley, 1995)	(Luong et al., 1993)
Av. Dist.	2.33 pixels	0.92 pixels	0.86 pixels

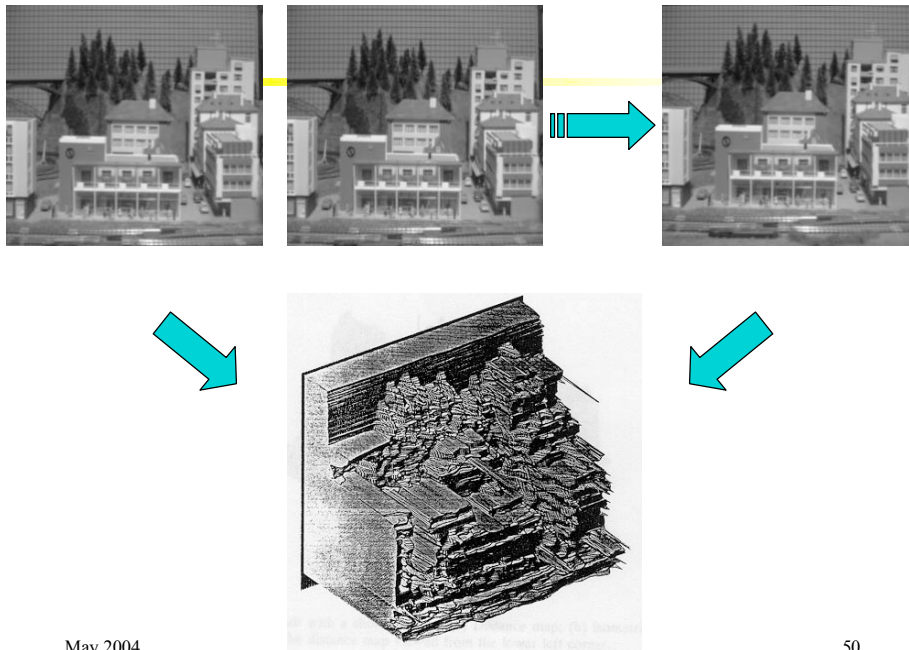
Figure 10-4

Weak-calibration experiment using 37 point correspondences between two images of a toy house. The figure shows the epipolar lines found by (a) the least-squares version of the eight-point algorithm, and (b) the normalized variant of this method proposed by Hartley (1995). Note, for example, the much larger error in (a) for the feature point close to the bottom of the mug. Quantitative comparisons are given in the table, where the average distances between the data points and corresponding epipolar lines are shown for both techniques as well as the nonlinear algorithm of Luong *et al.* (1993). *Data courtesy of Boubakeur Boufama and Roger Mohr.*

Stereo vision (Stereopsis)



49

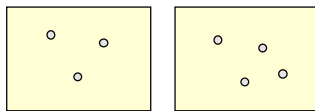


50

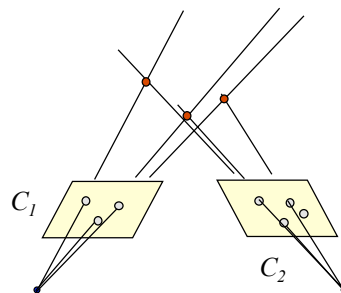
Basic approach to stereo vision

- Find features of interest in N image views
 - The “correspondence problem”
- Triangulate
 - A method to measure distance and direction by forming a triangle and using trigonometry
- Reconstruct object/scene depth
 - From dense points
 - From sparse points

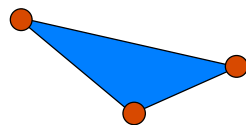
1. Correspondence



2. Triangulation

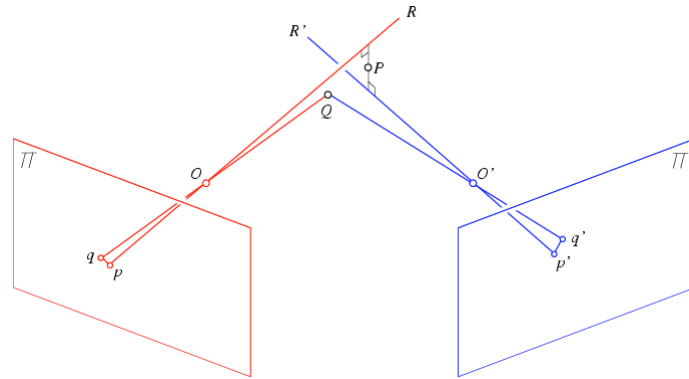


3. Reconstruction



Problem...

- Measurement error causes point Q to be seen at location p rather than the correct location q
 - A least squares method will triangulate to point P



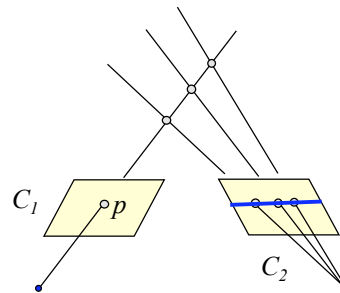
May 2004

Stereo

53

Correspondence

- Knowing the epipolar geometry certainly helps
 - Look on (and near) the epipolar line



- But correspondence is hard!
- Two approaches
 - Try to *improve* correspondence matching
 - Try to *avoid* correspondence matching

May 2004

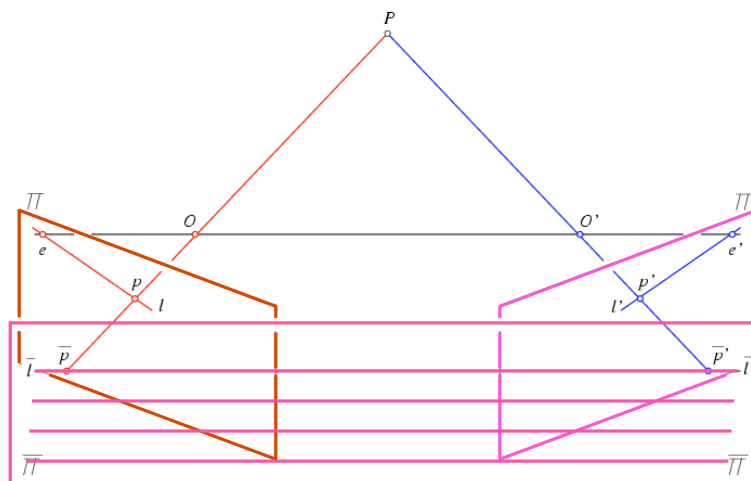
Stereo

54

Image rectification

- Stereo calculations can be much simplified if the two images are *rectified* – replaced by two equivalent images with a common image plane parallel to the baseline
- Single, common image plane
- Epipolar lines are image scan lines

Rectification example



Correlation based stereo matching



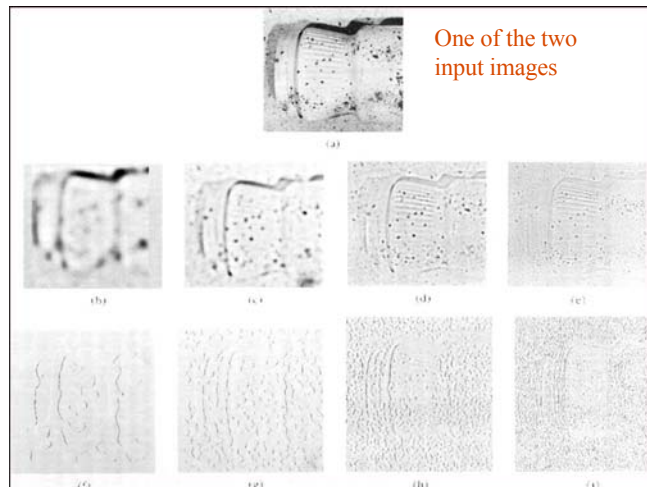
Texture-mapped
reconstructed surface

Multiscale edge-based stereo

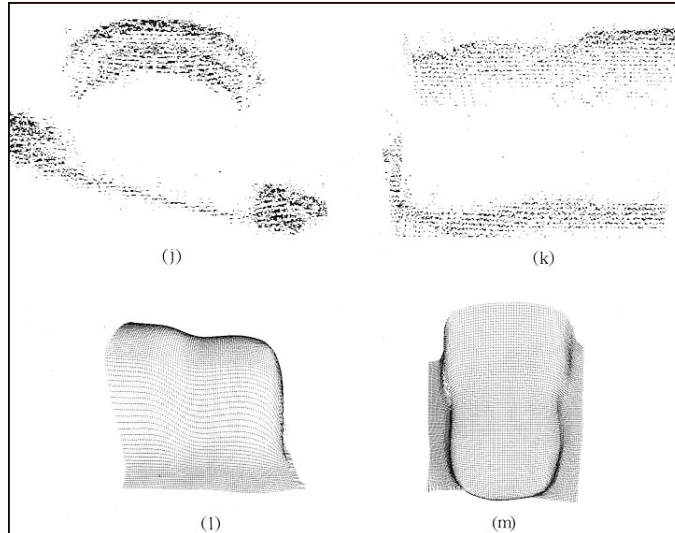
- Discussion session Friday/Monday

Laplacian filtering
at four scales

Zero crossings



Depth map



Reconstructed
surface

May 2004