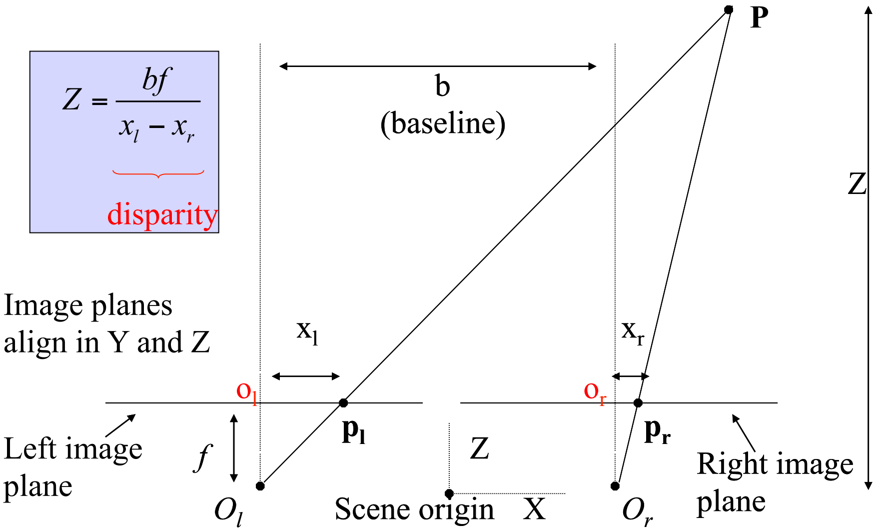
COMP37212:
Stereo basics and Epipolar Geometry (part 2)

Stereo - recap

- Stereo reconstruction is straightforward
- Require to calibrate camera parameters
- Require to find conjugate pairs
    - o Establish correspondences between points in right and left images.
- Edge-based correspondences
    - o Edge features
        - Polarity
        - orientation
    - o Canny
        - Multi-scale search
        - Positional precision
- Correlation matching
    - o Interest points (e.g. corners)
    - o Template matching

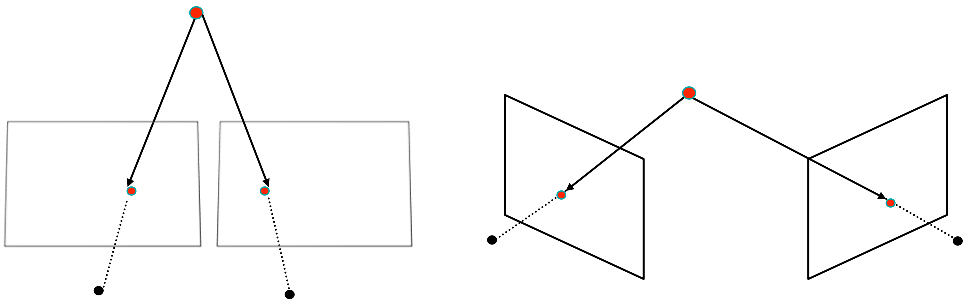
A simple Stereo System

$$Z = \underbrace{\frac{bf}{x_l - x_r}}_{\text{disparity}}$$



General Case With Calibrated Cameras

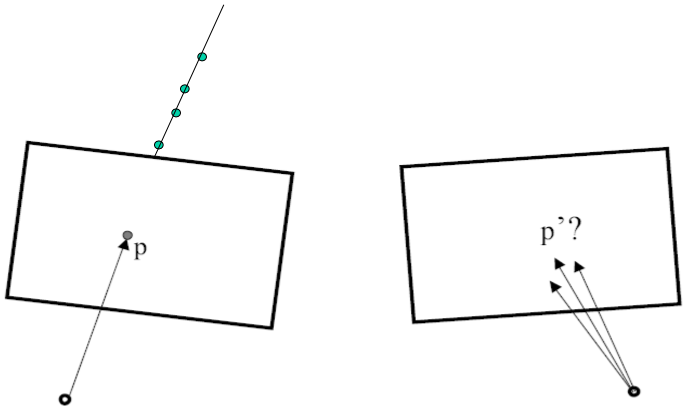
- The two cameras need not have parallel optical axes.



vs.

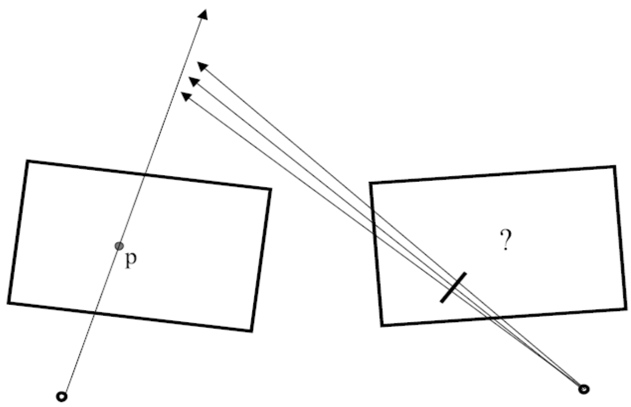
4

Stereo correspondence constraints



- Given p in left image, where can corresponding point p' ' in the right image be?

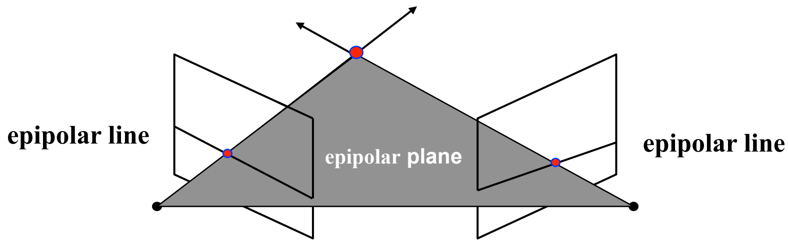
Stereo Correspondence Constraints



6

Stereo Correspondence Constraints

- Geometry of two views allows us to constrain where the corresponding pixel for some image point in the first view must occur in the second view.

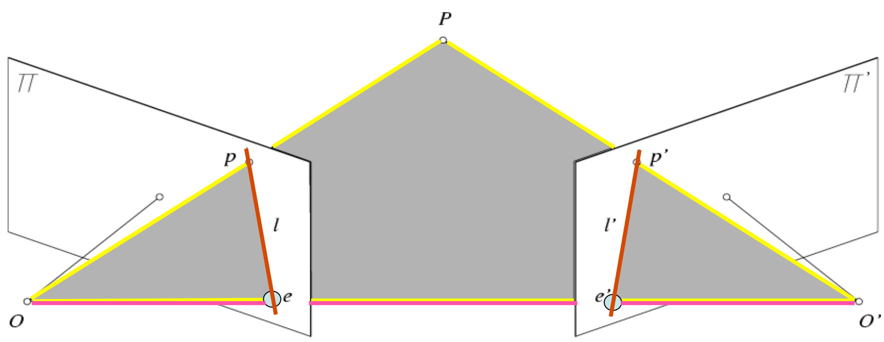


- Epipolar constraint: Why is this useful?
  - o Reduces correspondence problem to 1D search along conjugate epipolar lines.

Slide adapted from Steve Seitz

7

Epipolar Geometry



• Epipolar Plane • Baseline

• Epipoles • Epipolar Lines

Slide adapted from Marc Pollefeys

8

Epipolar Geometry: Terms

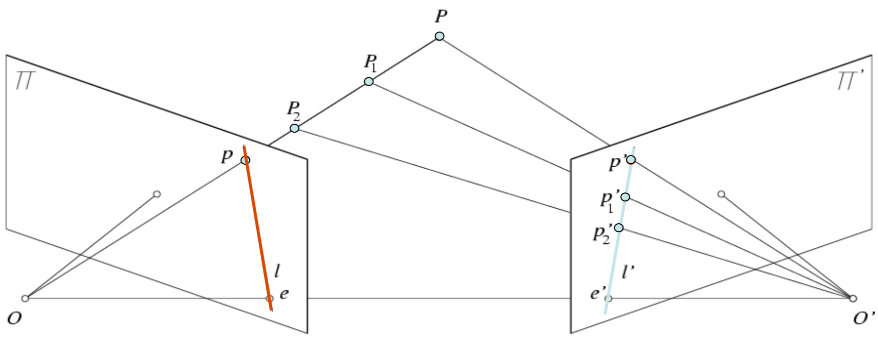
- *Baseline*: line joining the camera centers
- *Epipole*: point of intersection of baseline with the image plane
- *Epipolar plane*: plane containing baseline and world point
- *Epipolar line*: intersection of epipolar plane with the image plane

- All epipolar lines intersect at the epipole.
- An epipolar plane intersects the left and right image planes in epipolar lines.

Adapted from Marc Pollefeys

9

Epipolar Constraint

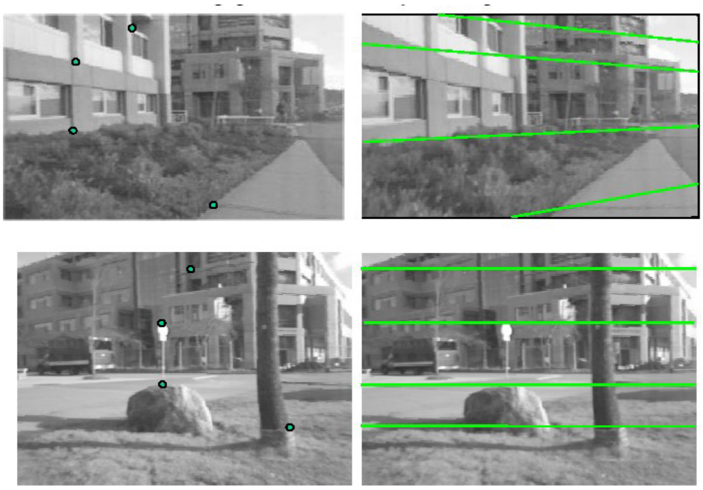


- Potential matches for p have to lie on the corresponding epipolar line l'.
- Potential matches for p' have to lie on the corresponding epipolar line l.

Slide credit: Marc Pollefeys

10

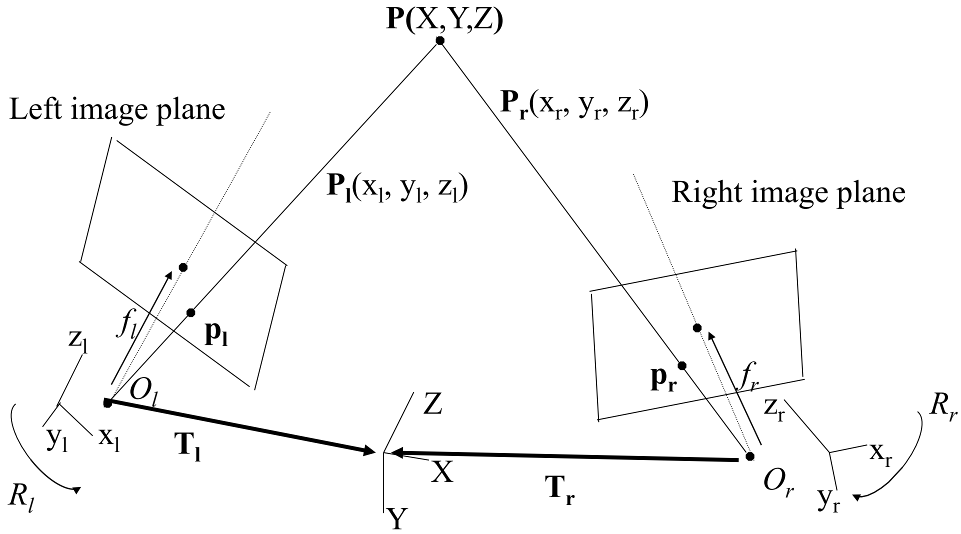
Example



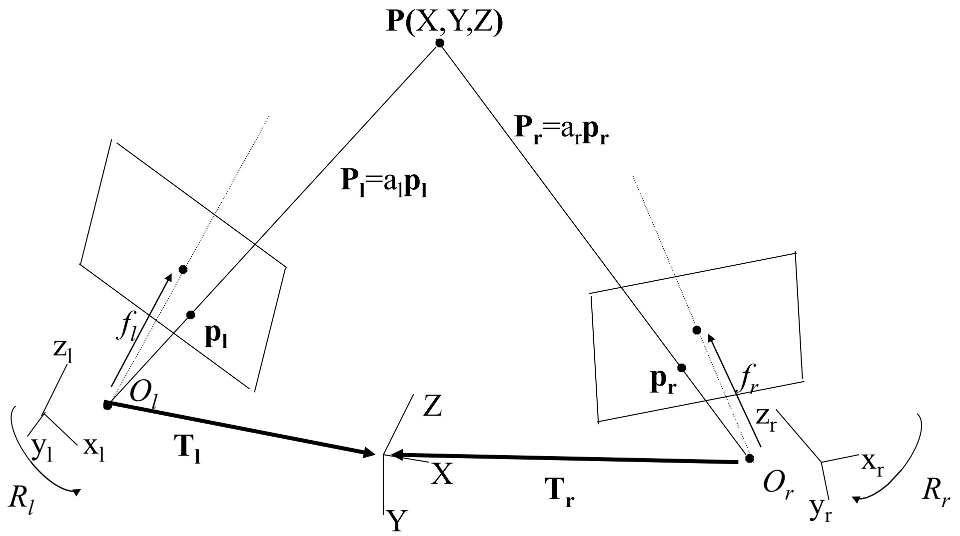
Generalising the Stereo Geometry

- In general, cameras are at arbitrary orientations
  - o Image planes are not parallel
- Separation between optical centres is not parallel to image planes
- Camera coordinate systems are different from each other and from the *scene coordinate system*
- Coordinate systems are related by
  - o Rotation matrices R_l and R_r giving the orientations of each of the camera coordinate systems relative to the scene coordinate system
  - o Translation vectors $\mathbf{T_l}$ and $\mathbf{T_r}$ between the camera origins and the scene origin.

Stereo Geometry



Stereo Reconstruction



Stereo Reconstruction

$\mathbf{P}_l' = \mathbf{T}_l + R_l \mathbf{P}_l \qquad \mathbf{P}_r' = \mathbf{T}_r + R_r \mathbf{P}_r$ In Scene Coordinates

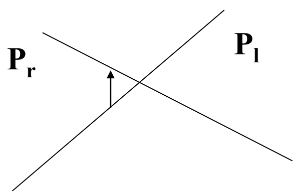
$\mathbf{P}_l' = \mathbf{T}_l + a_l R_l \mathbf{p}_l \qquad \mathbf{P}_r' = \mathbf{T}_r + a_r R_r \mathbf{p}_r$

$\mathbf{p}_l = (x_l, y_l, f_l) \qquad \mathbf{p}_r = (x_r, y_r, f_r)$

$\mathbf{P}_l'$ and $\mathbf{P}_r'$ intersect where $\mathbf{T}_l + a_l R_l \mathbf{p}_l = \mathbf{T}_r + a_r R_r \mathbf{p}_r$

Stereo Reconstruction (continued)

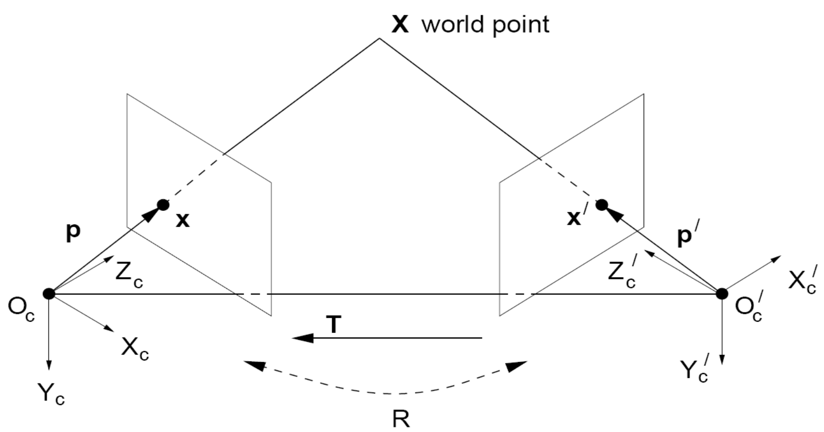
- Not quite that simple
- Because of measurement inaccuracies projected vectors will not coincide.
- We need to find the mid-point of the vector between the closest points on each
- **This can also be found by solving a set of linear equations**



More general case

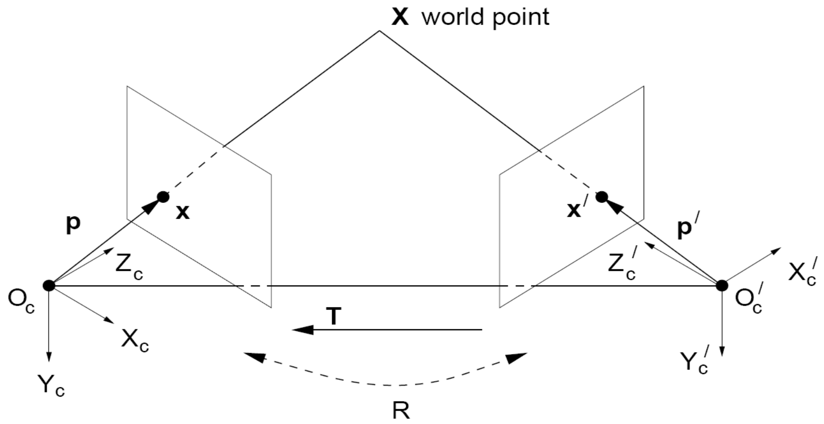
- Defining the **Essential matrix**

Stereo geometry, with calibrated cameras



Main idea

Stereo geometry, with calibrated cameras

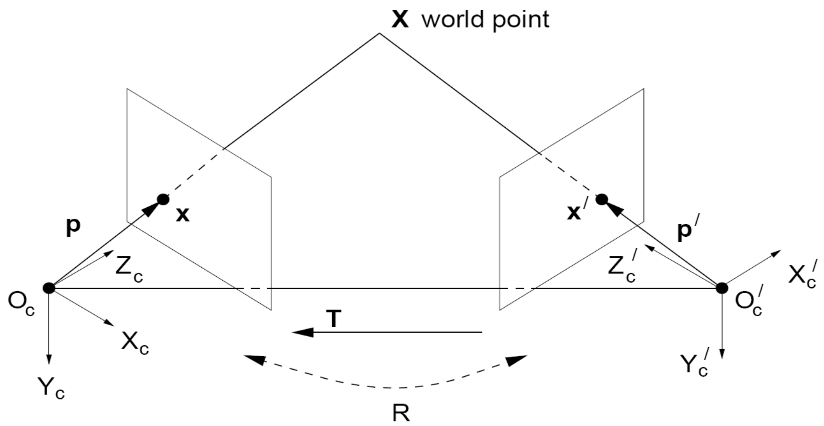


If the stereo rig is calibrated, we know :

how to **rotate** and **translate** camera reference frame 1 to get to camera reference frame 2.

Rotation: 3 x 3 matrix **R**; translation: 3 vector **T**.

Stereo geometry, with calibrated cameras



If the stereo rig is calibrated, we know :

how to **rotate** and **translate** camera reference frame 1 to get to camera reference frame 2.

$$\mathbf{X'}_c = \mathbf{R}\mathbf{X}_c + \mathbf{T}$$

Recall from Algebra: cross product

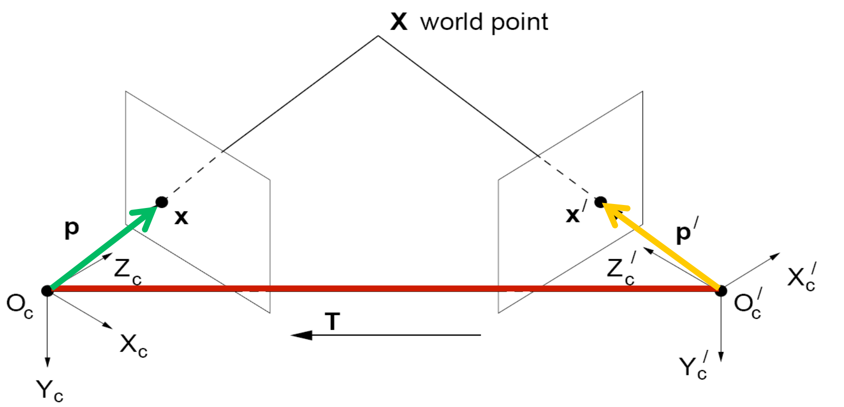
$$\vec{a} \times \vec{b} = \vec{c}$$

$$\vec{a} \cdot \vec{c} = 0$$
$$\vec{b} \cdot \vec{c} = 0$$

Vector cross product takes two vectors and returns a third vector that's perpendicular to both inputs.

So here, c is perpendicular to both a and b, which means the dot product = 0.

From geometry to algebra



$$\boxed{\mathbf{X}'} = \mathbf{R}\boxed{\mathbf{X}} + \boxed{\mathbf{T}}$$

$$\underbrace{\mathbf{T} \times \mathbf{X}'}_{\text{Normal to the plane}} =$$

$$= \mathbf{T} \times \mathbf{R}\mathbf{X}$$

$$\mathbf{X}' \cdot (\mathbf{T} \times \mathbf{X}') = \mathbf{X}' \cdot (\mathbf{T} \times \mathbf{R}\mathbf{X})$$
$$= 0$$

Again from algebra: Matrix form of cross product

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

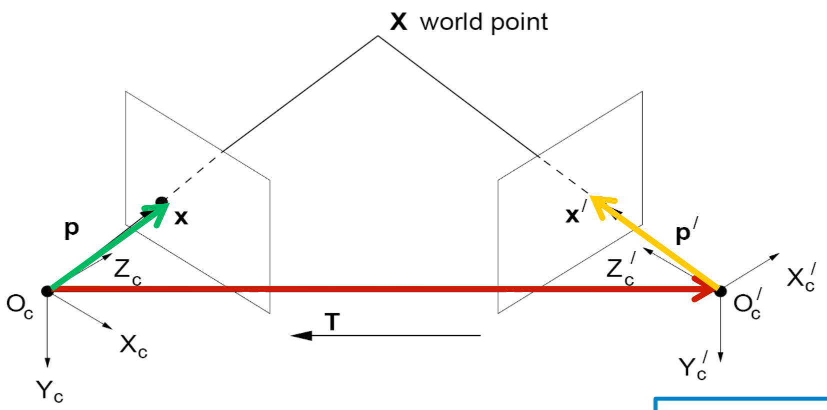
$$\vec{a} \cdot \vec{c} = 0$$
$$\vec{b} \cdot \vec{c} = 0$$

Can be expressed as a matrix multiplication.

$$[a_x] = \begin{bmatrix} 0 & -a_3 & a_2 \\ a_3 & 0 & -a_1 \\ -a_2 & a_1 & 0 \end{bmatrix}$$

$$\boxed{\vec{a} \times \vec{b} = [a_x]\vec{b}}$$

From geometry to algebra



$$\mathbf{X'} = \mathbf{RX} + \mathbf{T}$$

$$\underbrace{\mathbf{T} \times \mathbf{X}'}_{\text{Normal to the plane}} = \mathbf{T} \times \mathbf{RX} + \mathbf{T} \times \mathbf{T}$$

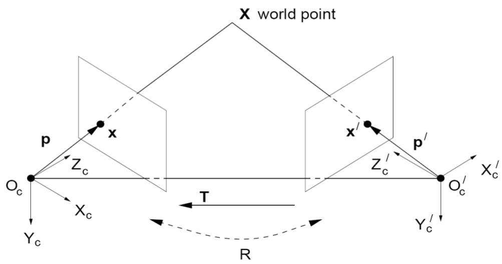
$$= \mathbf{T} \times \mathbf{RX}$$

$$\mathbf{X}' \cdot (\mathbf{T} \times \mathbf{X}') = \mathbf{X}' \cdot (\mathbf{T} \times \mathbf{RX}) = 0$$

Essential matrix

$$\boxed{\mathbf{X}' \cdot (\mathbf{T} \times \mathbf{R}\mathbf{X}) = 0}$$

$$\mathbf{X}' \cdot ([\mathrm{T}_x]\mathbf{R}\mathbf{X}) = 0$$



Let $\mathbf{E} = [\mathrm{T}_x]\mathbf{R}$

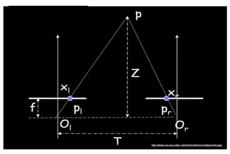
$$\mathbf{X}'^T \mathbf{E}\mathbf{X} = 0$$

E is called the **essential matrix**, and it relates corresponding image points between both cameras, given the rotation and translation.

If we observe a point in one image, its position in other image is constrained to lie on line defined by above.

Note: these points are in **camera coordinate systems**.

Essential matrix example: parallel cameras



$\mathbf{R} =$

$\mathbf{T} =$

$\mathbf{E} = [\mathrm{T}_x]\mathbf{R} =$

$\mathbf{p} = [x, y, f]$

$\mathbf{p}' = [x', y', f]$

$$\mathbf{p}'^{\mathrm{T}} \mathbf{E}\mathbf{p} = 0$$

For the parallel cameras, image of any point must lie on same horizontal line in each image plane.

image I(x,y)

Disparity map D(x,y)

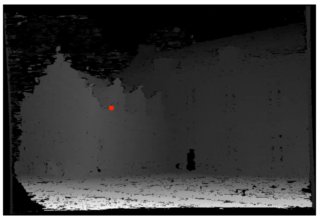
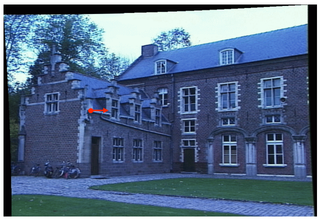


image I′(x′,y′)



$$(x', \boxed{y'}) = (x + D(x,y), \boxed{y})$$

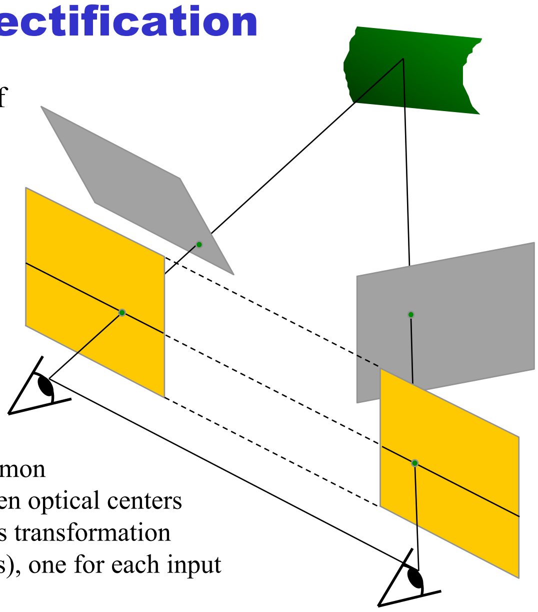
What about when cameras' optical axes are not parallel?

Rectification

- Search for correspondences is a 1-D search along epipolar lines.
  - o Lines at arbitrary angles through the epipoles.
- Search is easier if the images are *rectified*.
- Knowing R_l and R_r we can transform (warp) the images so that the image planes are parallel
  - o Epipoles are at infinity
  - o Epipolar lines are parallel to the horizontal image axis

Stereo image rectification

In practice, it is convenient if image scanlines (rows) are the epipolar lines.



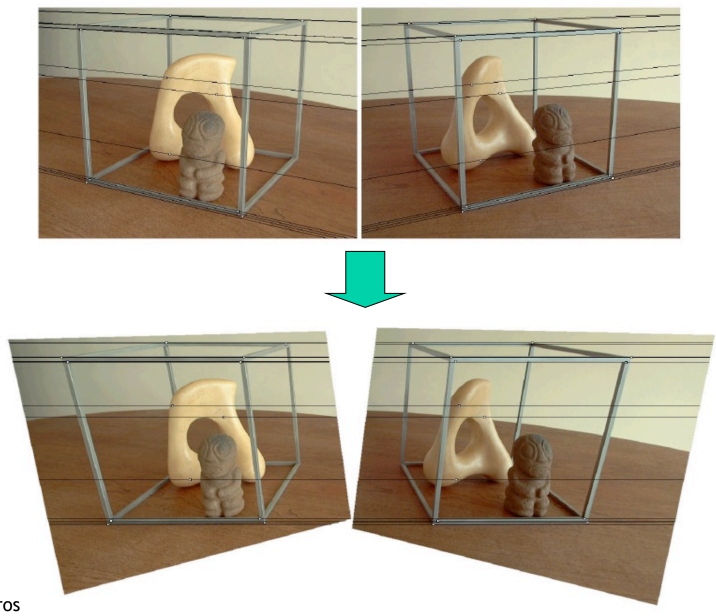
reproject image planes onto a common plane parallel to the line between optical centers

pixel motion is horizontal after this transformation

two homographies (3x3 transforms), one for each input image reprojection

Slide credit: Li Zhang

Stereo Image Rectification: Example



Source: Alyosha Efros

30

Summary so far

- Depth from stereo: main idea is to triangulate from corresponding image points.
- Epipolar geometry defined by two cameras
  - We've assumed known extrinsic parameters relating their poses
- Epipolar constraint limits where points from one view will be imaged in the other
  - Makes search for correspondences quicker

- **Terms**: epipole, epipolar plane / lines, disparity, rectification, intrinsic/extrinsic parameters, essential matrix, baseline

Topics of This Lecture

- Geometric vision
  - Visual cues
  - Stereo vision
- Epipolar geometry
  - Depth with stereo
  - Geometry for a simple stereo system
  - Case example with parallel optical axes
  - General case with calibrated cameras
- Stereopsis & 3D Reconstruction
  - Correspondence search
  - Additional correspondence constraints
  - Possible sources of error
  - Applications

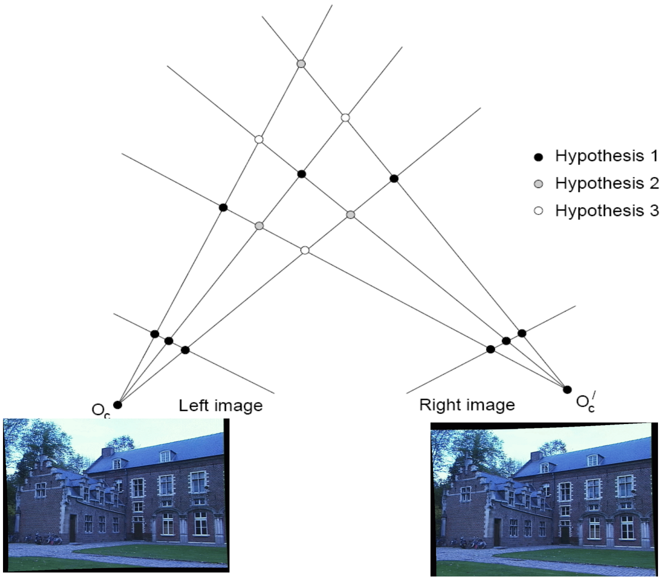
32

Stereo Reconstruction

- Main Steps
  - o Calibrate cameras
  - o Rectify images
  - o Compute disparity
  - o Estimate depth

33

Correspondence Problem



Multiple match hypotheses satisfy epipolar constraint, but which is correct?

34

Figure from Gee & Cipolla 1999

Correspondence problem

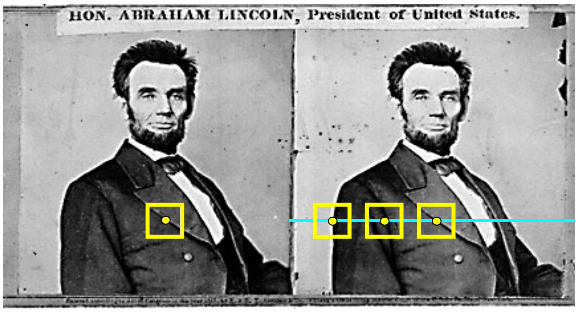
- Beyond the hard constraint of epipolar geometry, there are "soft" constraints to help identify corresponding points
  - o Similarity
  - o Uniqueness
  - o Ordering
- To find matches in the image pair, we will assume
  - o Most scene points visible from both views
  - o Image regions for the matches are similar in appearance

Additional Correspondence Constraints

- Similarity
- Uniqueness
- Ordering

36

Dense Correspondence Search



- For each pixel in the first image
  - o Find corresponding epipolar line in the right image
  - o Examine all pixels on the epipolar line and pick the best match (e.g. SSD, correlation)
  - o Triangulate the matches to get depth information
- This is easiest when epipolar lines are scanlines
  - ⇒ Rectify images first

adapted from Li Zhang

Example: Window Search

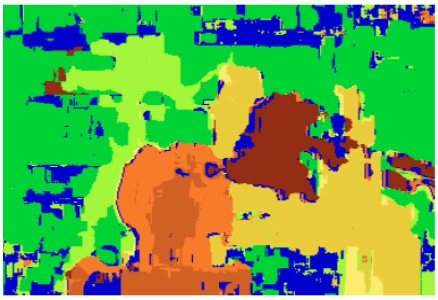
- Data from University of Tsukuba

Scene

Ground truth

Example: Window Search

- Data from University of Tsukuba



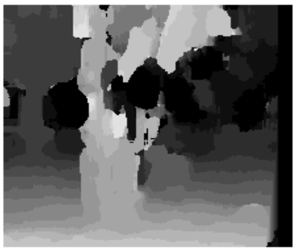
Window-based matching (best window size)

Ground truth

39

Effect of Window Size

W = 3



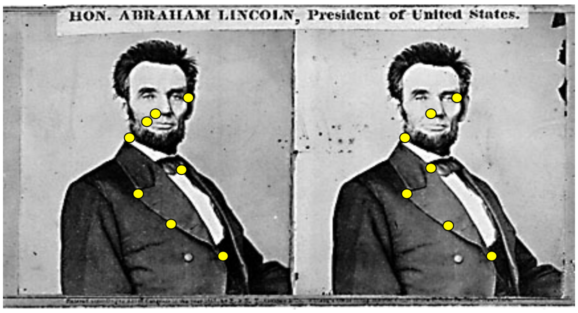
W = 20

Want window large enough to have sufficient intensity variation, yet small enough to contain only pixels with about the same disparity.

40

Figures from Li Zhang

Sparse Correspondence Search



- Restrict search to sparse set of detected features
- Rather than pixel values (or lists of pixel values) use *feature descriptor* and an associated *feature distance*
- Still narrow search further by epipolar geometry

What would make good features?

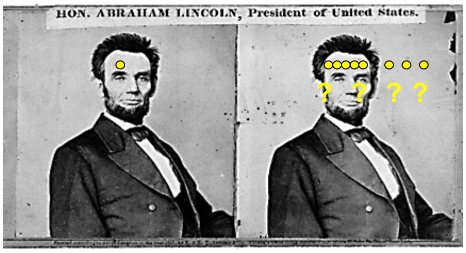
41

Dense vs. Sparse

- Sparse
  - o Efficiency
  - o Can have more reliable feature matches, less sensitive to illumination than raw pixels
  - o But…
    - Have to know enough to pick good features
    - Sparse information
- Dense
  - o Simple process
  - o More depth estimates, can be useful for surface reconstruction
  - o But…
    - Breaks down in textureless regions anyway
    - Raw pixel distances can be brittle
    - Not good with very different viewpoints

42

Difficulties in Similarity Constraint



Untextured surfaces

Occlusions

43

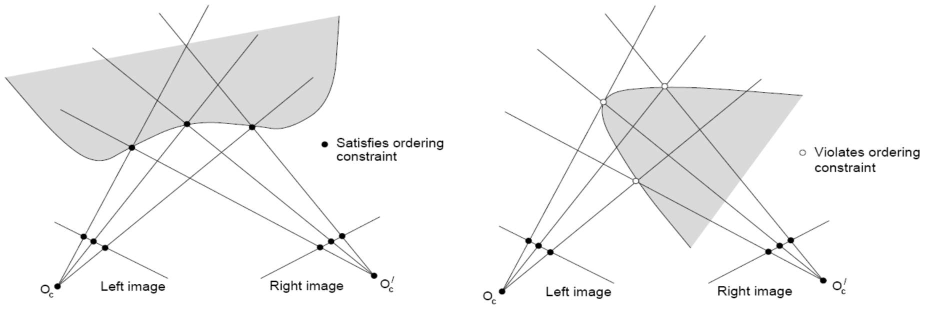
Additional Correspondence Constraints

- Similarity
- Uniqueness
- Ordering

44

Ordering

- Points on *same surface* (opaque object) will be in same order in both views



45

Figure from Gee & Cipolla 1999

Additional Correspondence Constraints

- Similarity
- Uniqueness
- Ordering

- Epipolar lines constrain the search to a line, and these appearance and ordering constraints further reduce the possible matches.

46

Possible Sources of Error?

- Low-contrast / textureless image regions
- Occlusions
- Camera calibration errors
- Violations of *brightness constancy* (e.g., specular reflections)
- Large motions

47

Application: Depth for Segmentation

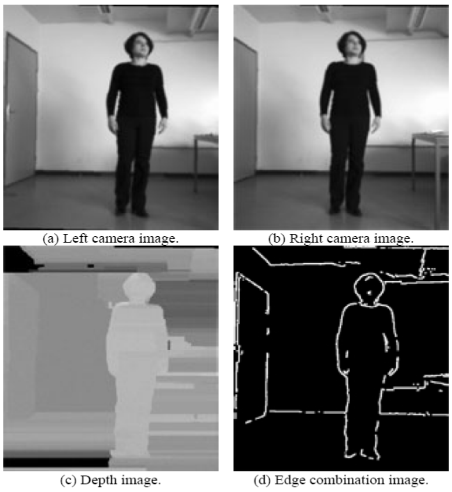


Figure 3 Stereo video frames with computed depth map and edge combination result.

Edges in disparity in conjunction with image edges enhances contours found

Image Source: Danijela Markovic and Margrit Gelautz

Depth for Segmentation


Original image with snake initialization.


(b) Final snake on original image.


(c) Final snake on depth image.


(d) Original image with snake from (c) overlaid.


Final snake on edge combination image.


(f) Original image with snake from (e) overlaid.

Image Source: Danijela Markovic and Margrit Gelautz

Application: View Interpolation

Right Image

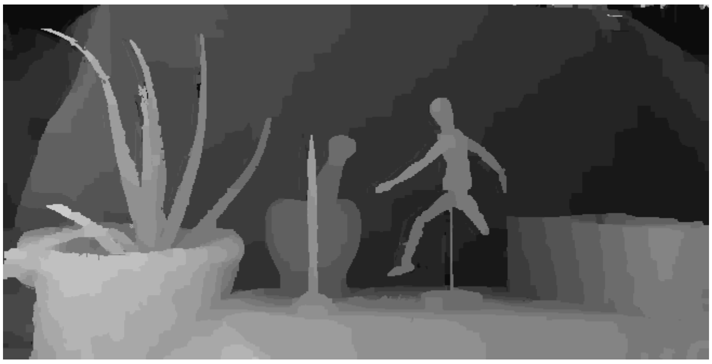
Slide credit: Svetlana Lazebnik

Application: View Interpolation

Left Image

Slide credit: Svetlana Lazebnik

Application: View Interpolation



Disparity

Slide credit: Svetlana Lazebnik

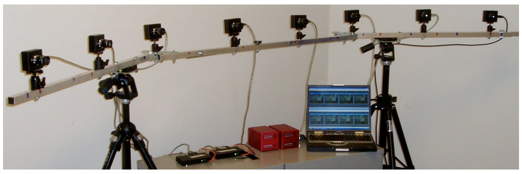
Application: View Interpolation

Slide credit: Svetlana Lazebnik

Applications: Virtual Viewpoint Video

L. Zitnick et al, High-quality video view interpolation using a layered representation, SIGGRAPH 2004.

Applications: Virtual Viewpoint Video



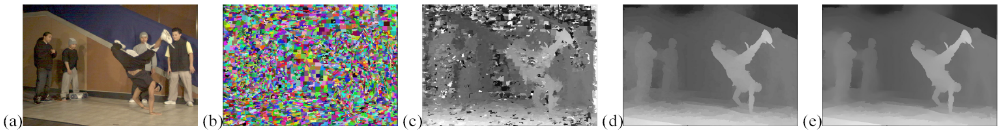


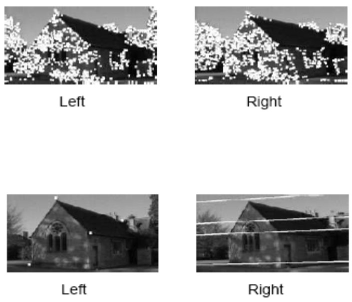
Figure 6: *Sample results from stereo reconstruction stage: (a) input color image; (b) color-based segmentation; (c) initial disparity estimates $\hat{d}_{ij}$; (d) refined disparity estimates; (e) smoothed disparity estimates $d_i(x)$.*

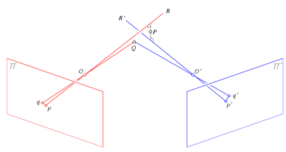
L. Zitnick et al, High-quality video view interpolation using a layered representation, SIGGRAPH 2004.

55

So far: Stereo Reconstruction

- **Main Steps**
  - o Calibrate cameras
  - o Rectify images
  - o Compute disparity
  - o Estimate depth
- **So far, we have only considered calibrated cameras…**
- **Next …**
  - o Camera parameters
  - o Uncalibrated cameras



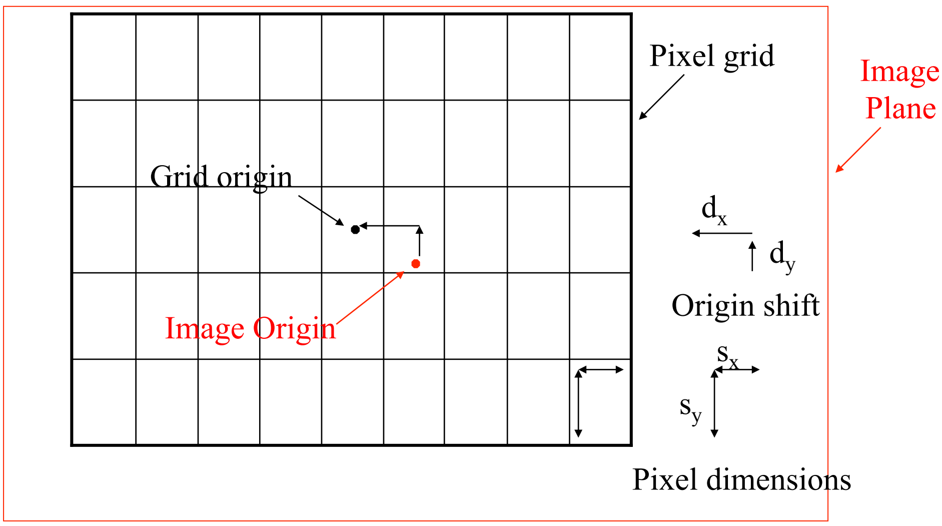


56

Camera Parameters

- Extrinsic parameters
  - o Rotation matrix *R* (3X3) (3 free parameters)
  - o Translation vector (T_x, T_y, T_z)
- Intrinsic parameters
  - o Relate pixel coordinates to image coordinates
  - o Pixel size (s_x, s_y): pixels may not be square
  - o Origin offset (d_x, d_y): pixel origin may not be on optic axis
  - o Focal length, *f*.
  - o Not totally independent. (Need d_x, d_y, *f*, s_x/s_y)

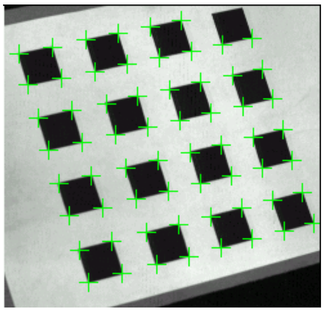
Intrinsic Camera Parameters



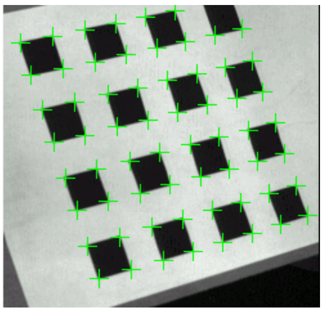
Stereo Calibration

- We need to know the camera parameters
  - o *R*, **T** and *f* to calculate the triangulation
  - o d_x, d_y, f, s_x/s_y, to calculate image coordinates from pixel coordinates
- We can calculate these parameters if we know the scene coordinates of sufficient image points
- Calibration using target image
  - o accurately measured feature positions
  - o Reliable location on images

Camera Calibration Target



Left Image



Right Image

Corners of squares can be located accurately in two directions by edge finding.

Calibration algorithms

- Several algorithms exist
- Compromises between
  - o Accuracy of parameter estimation
  - o Robustness of parameter estimation
  - o Complexity of calculation
    - Least squares…non-linear optimisations
  - o Engineering requirement of target
    - Points on plane
    - Points throughout 3D volume.

Uncalibrated Stereo

- Calibration is necessary to determine absolute 3D positions
- We can determine *relative* 3D positions (up to a scale factor) without calibration
- If at least 8 correspondences in the scene are known (avoiding "degenerate" configurations) sufficient camera parameters can be estimated
- c.f. human stereopsis

Uncalibrated Views

[redacted]

Relative Depth

Summary

- Stereopsis can provide accurate 3D measurements of positions
- Need to establish correspondences
  - o Epipolar lines
  - o Matching criteria
    - Edges, corners, interest points
  - o Constraints
- Calibration
  - o Extrinsic parameters: R, $\mathbf{T}$
  - o Intrinsic parameters: d_x, d_y, f, s_x/s_y
- Reconstruction by triangulation

References and Further Reading

- Background information on epipolar geometry and stereopsis can be found in Chapters 10.1-10.2 and 11.1-11.3 of

D. Forsyth, J. Ponce,
Computer Vision - A Modern Approach.
Prentice Hall, 2003