Shape from Stereo

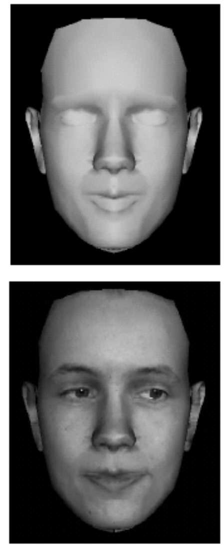
Shape from Stereo

- Problem
  - Infer 3D structure of a scene from two images taken from different viewpoints

- Two Primary Sub-problems in Stereo Vision
  - Correspondence problem
  - Recover Stereo geometry and 3D reconstruction

- Lectures on Stereo Vision
  - Simple Stereo Configuration
  - Stereo Geometry – Epipolar Geometry
  - Correspondence Problem – Two classes of approaches
  - 8-point Algorithm to Solve Epipolar Geometry
  - 3D Reconstruction Problems

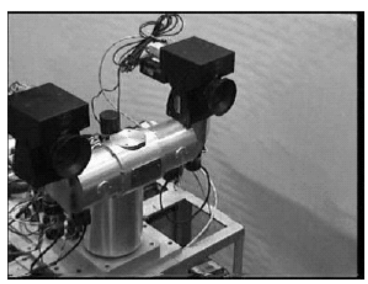
images



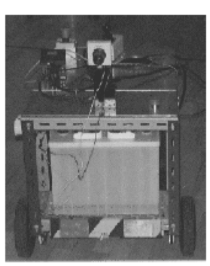
shape

surface
reflectance

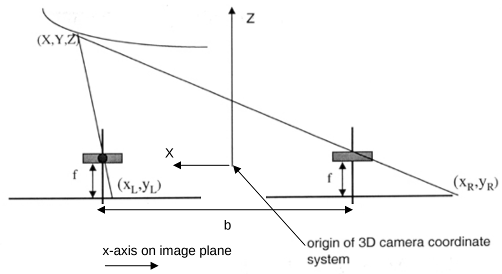
Stereo head



Camera on a mobile vehicle



Simple Stereo Configuration



- Optical axes are parallel, separated by baseline, b.
- Line connecting lens centers is perpendicular to the optical axis, and x-axis is parallel to that line.
- 3D coordinate system is a Cyclopean system centered between the cameras.

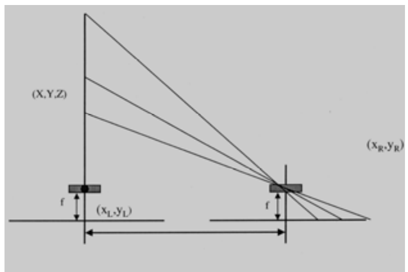
Simple Stereo Configuration

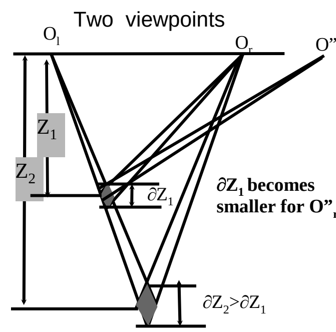
- (X,Y,Z) are the coordinates of P in the Cyclopean coordinate system. *Note: textbook has a slightly different formulation.*
- The coordinates of P in the left camera coordinate system are $(X_L,Y_L,Z_L)=(X-b/2,Y,Z)$.
- The coordinates of P in the right camera coordinate system are $(X_R,Y_R,Z_R)=(X+b/2,Y,Z)$.
- So, the x image coordinates of the projection of P are $x_L= (X-b/2)f/Z$, $x_R = (X+b/2)f/Z$.
- Subtracting x_L from x_R and solving for Z:

  $Z = bf/(x_R - x_L)$
- This lateral displacement $(x_R - x_L)$ is called the horizontal binocular disparity, d.

Simple Stereo Configuration

- Distance $Z = bf/d$ is inversely proportional to disparity



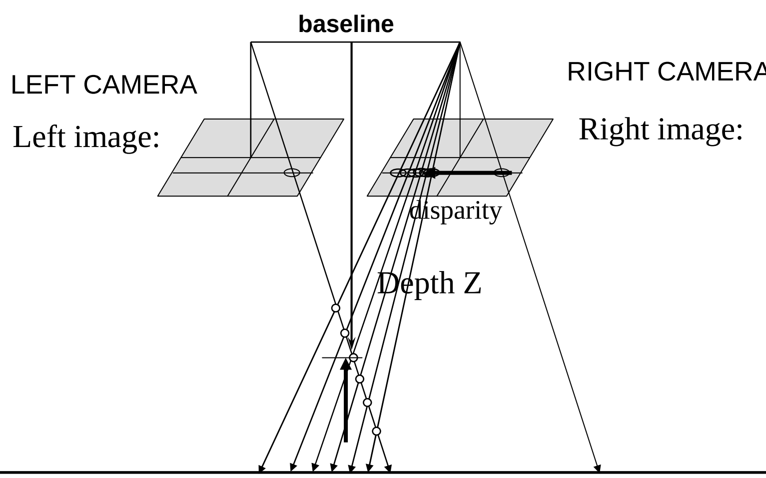


- Disparity is proportional to b
  - The larger b, the further we can accurately range
  - but as b increases, the images decrease in common field of view.

Relative error in Z

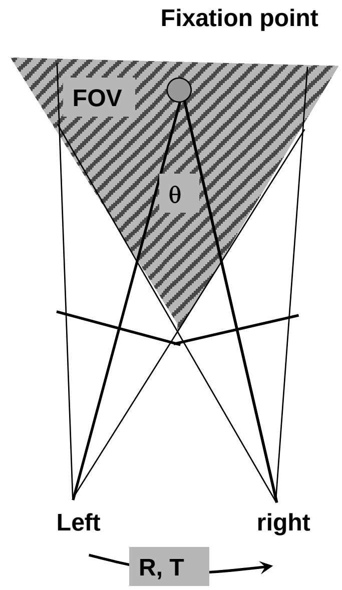
$$\frac{\partial Z}{Z} = \frac{Z}{bf}\partial(d)$$

- Definition: A scene point, P, visible in both cameras gives rise to a pair of image points called a corresponding pair.
  - The corresponding point of a point in the left (right) image must lie on the same image row (line) in the right (left) image (for this simple case).
  - This line is called an epipolar line (for that point). For our simple geometry, all epipolar lines are parallel to the x-axis.

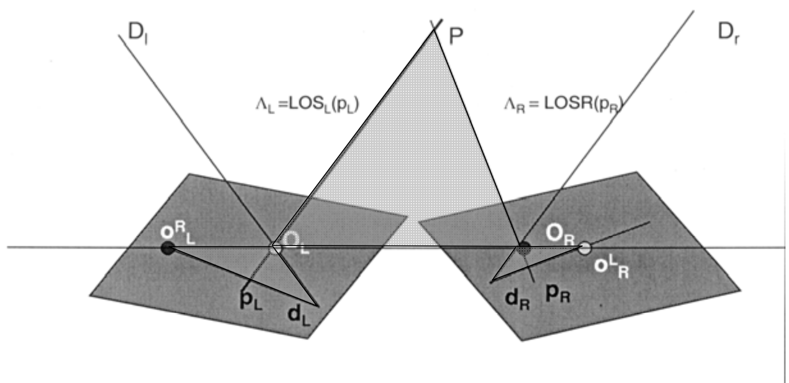


A more practical stereo imaging model

- Difficult, in practice, to
  - have the optical axes parallel
  - have the baseline perpendicular to the optical axes.
- We might want to tilt the camera towards one another to have more overlap in the images
- Required to solve calibration problem – finding the transformation between the two cameras.
  - It is a rigid body motion and can be decomposed into a rotation **R** and a translation **T**.



General stereo imaging (Epipolar Geometry)



NB: O_L and O_R (the red spots) are above the image planes.

General stereo imaging

FACTS

- Point P lies somewhere on the ray (line) Λ_L from p_L through O_L
  - but from the left image alone, we do not know where on this ray P lies
- Since the perspective projection of a line is a line, the perspective projection of Λ_L in the right image is a line
  - the ``first'' point on Λ_L that might correspond to P is O_L
    - any point closer to the left image than O_L would be between the lens and the image plane, and could not be seen
    - the perspective projection of O_L in the right camera is the point o^L_R .
  - the ``last'' point on $\Lambda_L$ that might correspond to P is the point ``infinitely'' far away along the ray Λ_L
    - but its image is the vanishing point of the ray Λ_L in the right camera, d_R
- any other possible location for P will project to a point in R on the line joining o^L_R to d_R.

General stereo imaging

First general conclusion

- Given any point, p_L , in the left image of a stereo pair, its corresponding point must appear on a **line** in the right image
- Furthermore, all of the epipolar lines for all of the points in the left image must pass through a common point in the right image
  - » this point is image of the left lens center in the right image
  - » this point lies on the line of sight for **every** point in the left image
  - » therefore, the epipolar lines must all contain (i.e., pass through) the image of this point
  - » This point is called an epipole.
- Finally, the epipolar line for p_L must also pass through the vanishing point in the right image for the line of sight through p_L
- If we know, based on p_L 's coordinates, the coordinates in the right image of two points on its epipolar line, we know its epipolar line.

Next FACT

- Remember that any three non-collinear points define a plane, and
- The points O_L , p_L , and $o^R{}_L$ (epipole in the left image) are three non-collinear points, so they form a plane, Π, known as the **epipolar plane**.
  - -- the line Λ_L lies on this plane, since two points on the line lie on the plane
- The intersection of this plane with the right image plane is the epipolar line of p_L
  - -- and this would be the image of **any** line on this plane
- Let p'_L be some other point on the line joining p_L and $o^R{}_L$.
  - -- the line of sight through p'_L to P' lies on Π since p'_L lies on the line p_L -$o^R{}_L$ which lies on the plane, and O_L also lies on the plane.
- *Thus*, the epipolar line for p'_L must be the same line as the epipolar line for p_L, or for **any other point** on the line containing p_L and $o^R{}_L$.
- These conjugate epipolar lines p_L - $o^R{}_L$ and p_R - $o^L{}_R$ are the intersection of the epipolar plane Π with the left and right image planes respectively.

General stereo imaging

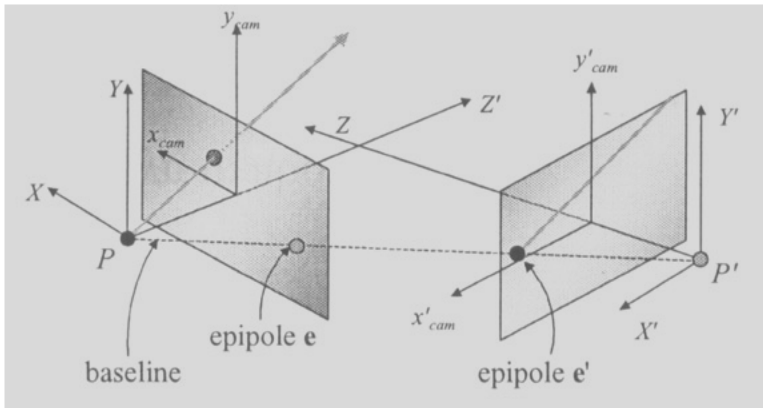
Second general conclusion

- Given any point, p_L , in the left image. Consider
  - -- the line joining p_L and the image of the right camera center in the left image (the epipole in the left image), and
  - -- the epipolar line of p_L in the right camera.
- Given **any point** on either of these two lines, its corresponding pair must lie on the other line
- These lines are the intersection of the epipolar plane Π with the left and right image planes respectively.
- Question: is the epipolar plane different for each point?

demo

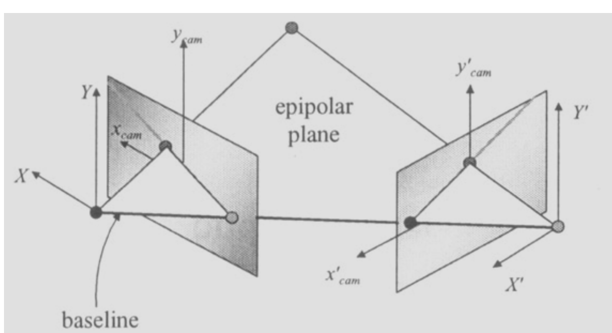
Summary: General stereo imaging

- The **epipole**: is the *point* of intersection of the line joining the optical centres---the *baseline*---with the image plane. The epipole is the image in one camera of the optical centre of the other camera.



Summary: General stereo imaging

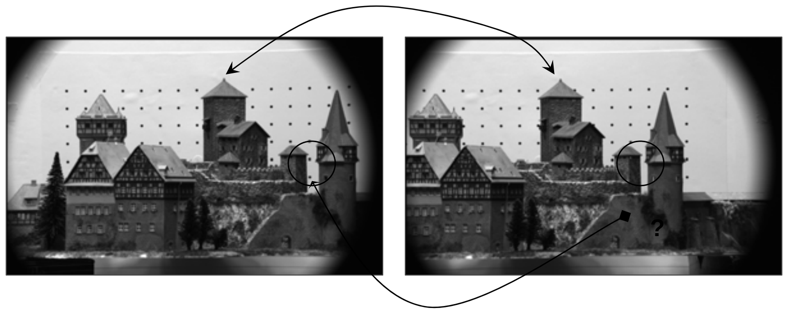
- The **epipolar plane**: is the *plane* defined by a 3D point and the optical centres. Or, equivalently, by an image point and the optical centres.
- The **epipolar line**: is the *straight line* of intersection of the epipolar plane with the image plane. It is the image in one camera of a ray through the optical centre and image point in the other camera. All epipolar lines intersect at the epipole.



Stereo correspondence problem

- Given a point,p, in the left image, find its corresponding point in the right image

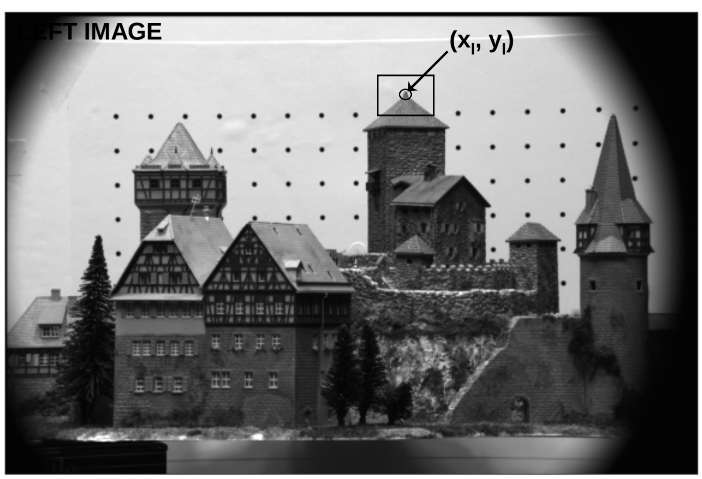
  -- called the stereo correspondence or stereo matching problem



- Broadly classified into two types of approach

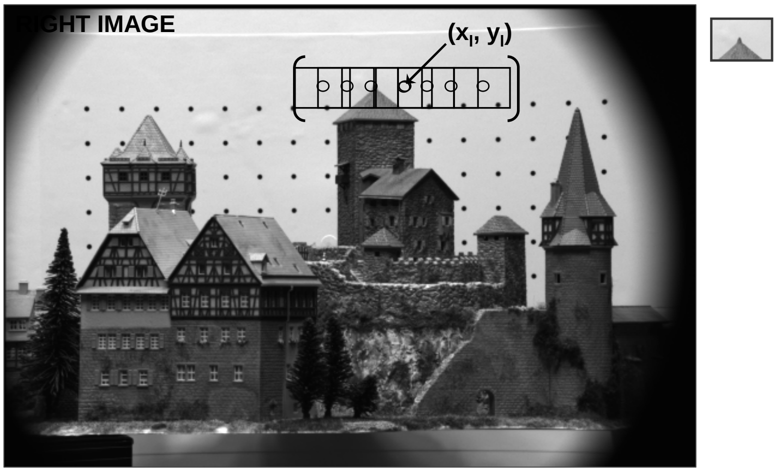
  1. **Intensity based approaches**: Image intensity used directly for correlation.
  2. **Feature based approaches**: Edge, corners (or any other salient points), regions, etc used (typically 200-300 per image) .

Correlation Approach



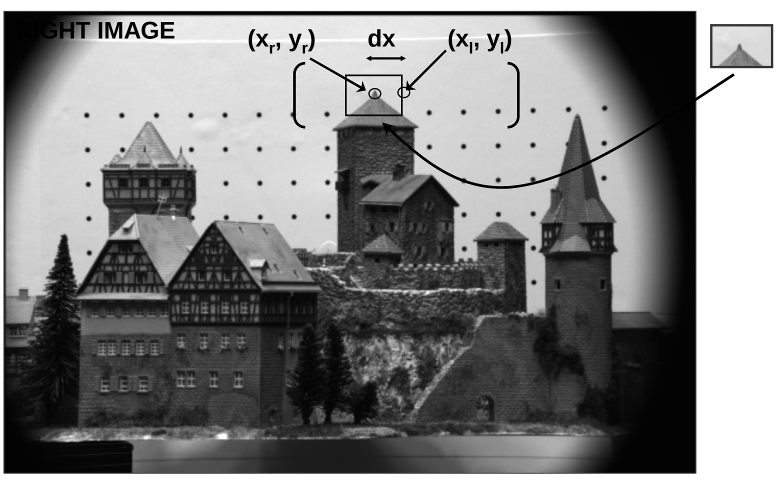
- For Each point (x_l, y_l) in the left image, define a window centered at the point

Correlation Approach



- … search its corresponding point within a search reghon in the right image

Correlation Approach



- … the disparity (dx, dy) is the displacement when the correlation is maximum

Correlation Approach

- Elements to be matched
  - Image window of fixed size centered at each pixel in the left image
- Similarity criterion
  - A measure of similarity between windows in the two images
  - The corresponding element is given by window that maximizes the similarity criterion within a search region
- Search regions
  - Theoretically, search region can be reduced to a 1-D segment, along the epipolar line, and within the disparity range.
  - In practice, search a slightly larger region due to errors in calibration

Correlation Approach

- Equations

$$c(dx, dy) = \sum_{k=-W}^{W} \sum_{l=-W}^{W} \psi(I_l(x_l + k, y_l + l), I_r(x_l + dx + k, y_l + dy + l))$$

- disparity $\bar{\mathbf{d}} = (\bar{d}x, \bar{d}y) = \arg\max_{\mathbf{d} \in R} \{c(dx, dy)\}$

- Similarity criterion
  - Cross-Correlation $\Psi(u, v) = uv$
    - Usually use Normalize Cross-correlation (NCC)
  - Sum of Square Difference (SSD) $\Psi(u, v) = -(u - v)^2$
  - Sum of Absolute Difference(SAD) $\Psi(u, v) = -|u - v|$

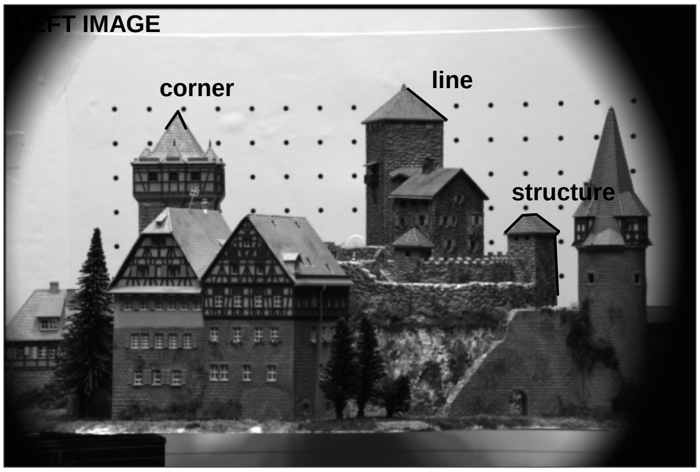
Correlation Approach

- PROS
  - Easy to implement
  - Produces dense disparity map
  - Maybe slow
- CONS
  - Needs textured images to work well
  - Inadequate for matching image pairs from very different viewpoints due to illumination changes
  - Window may cover points with quite different disparities
  - Inaccurate disparities on the occluding boundaries

Feature-based Approach

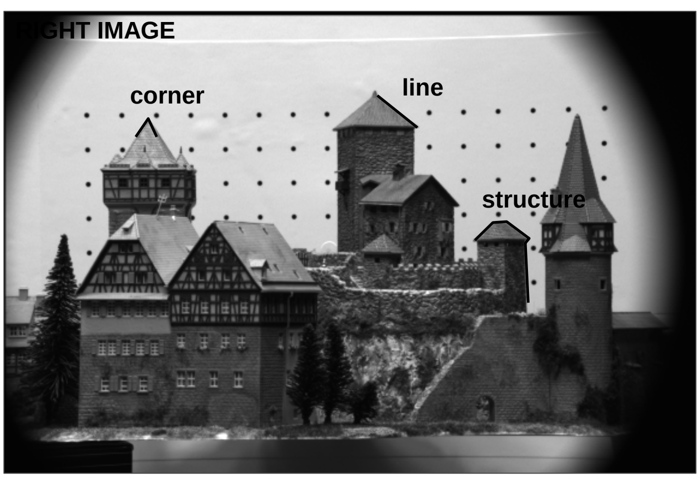
- Features
  - Edge points
  - Lines (length, orientation, average contrast)
  - Corners (Y-junction, L-junction, A-junction etc)
- Matching algorithm
  - Extract features in the stereo pair
  - Define similarity measure
  - Search correspondences using similarity measure and the epipolar geometry

Feature-based Approach



- For each feature in the left image…

Feature-based Approach

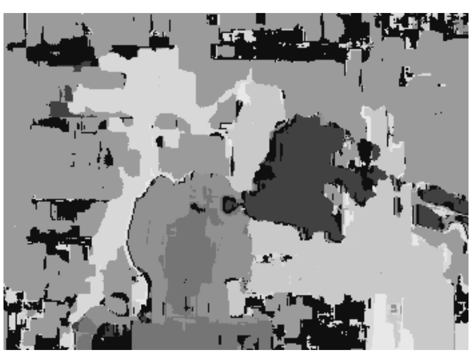


- Search in the right image… the disparity (dx, dy) is the displacement when the similarity measure is maximum

Feature-based Approach

- PROS
  - Relatively insensitive to illumination changes
  - Good for man-made scenes with strong lines but weak texture or textureless surfaces
  - Work well on the occluding boundaries (edges)
  - Could be faster than the correlation approach
- CONS
  - Only sparse depth map
  - Feature extraction may be tricky
    - Lines (Edges) might be partially extracted in one image
    - How to measure the similarity between two lines?

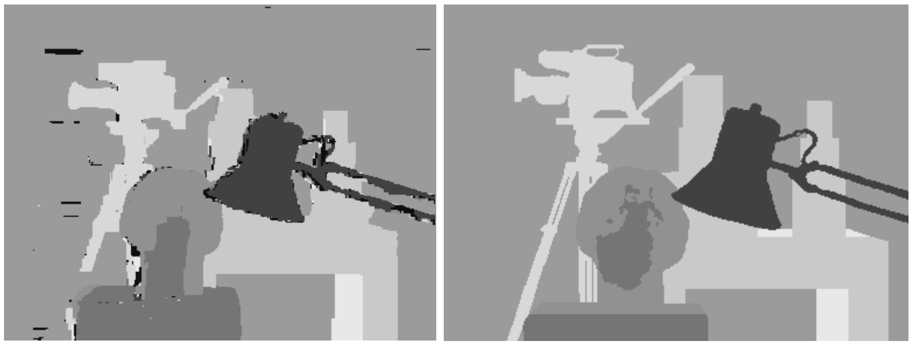
Results with window correlation



Window-based matching (best window size)

Ground truth

Results with better modern method



Graph cut method

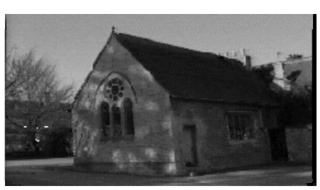
Boykov et al., "Fast Approximate Energy Minimization via Graph Cuts", PAMI 23(11) p1222-1239, 2001.

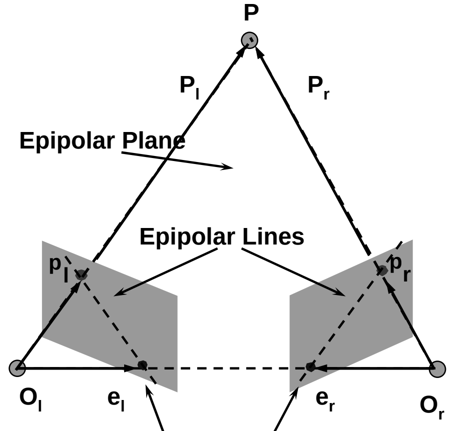
Ground truth

- See Middlebury stereo site: www.middlebury.edu/stereo/ for various techniques
- Read “A Comparative Study of Energy Minimization Methods for Markov Random Fields with Smoothness-Based Priors”. Szeliski et al. PAMI2008 for a more recent survey

Constraints for Stereo correspondence

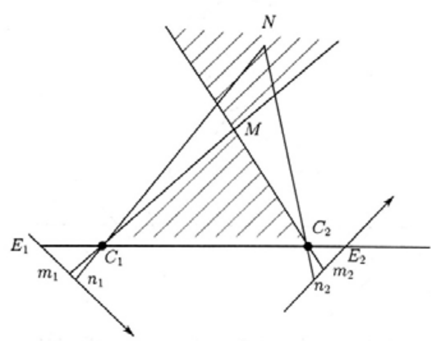
- What constraints simplify this problem?
  - -- Epipolar constraint - need only search for the corresponding point on the epipolar line





Constraints for Stereo correspondence

- Continuity or ordering constraint - if we are looking at a continuous surface, images of points along a given epipolar line will be ordered the same way

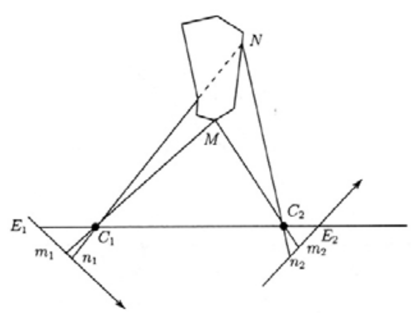


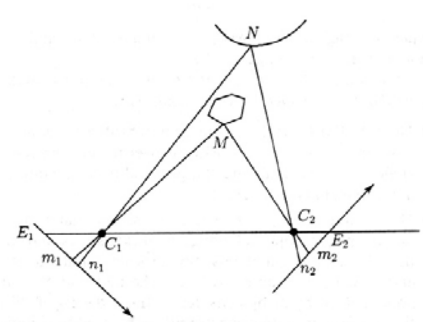
Choose any point N in the cross-hatched zone.

Order along epipolar lines: (m_1,n_1) is L to R, & (m_2,n_2) R to L: reverse ordering.

Cross-hatched zone is the *forbidden zone* attached to M (if M and N are on same opaque object of nonzero thickness)

Constraints for Stereo correspondence





M and N cannot be seen simultaneously by retinas 1 and 2.

In practice, difficult to eliminate whole cross-hatched zone.

M and N belong to different objects. Ordering constraint does not apply.

more reasonable to force only neighbors of M within a small neighborhood to belong to non-forbidden zone.

Constraints for Stereo correspondence

- Disparity gradient constraint - disparity changes slowly over most of the image.
  - -- Exceptions occur at and near occluding boundaries where we have either discontinuities in disparity or large disparity gradients as the surface recedes away from sight.
- Psychophysics experiments showed that human perception imposes similar constraint.

Why is the correspondence problem hard?

- Foreshortening (perspective) effects
  - A square match window in one image will be distorted in the other if disparity is not constant - complicates correlation
- How to handle this problem?

  -- Match images at low resolutions

  -- use these estimates as initial conditions for matching edges at next finest scale

Why is the correspondence problem hard?

- Occlusion
  - Even for a smooth surface, there might be points visible in one image and not the other
    - depends on rate of change of depth
    - separation of optical centres

  -- Consider aerial photo pair of urban area - vertical walls of buildings might be visible in one image and not the other
    - scene with depth discontinuities (lurking objects) violate continuity constraint and introduces occlusion

•Variations in intensity between images due to

-- uncorrelated noise effects

-- specularities

2^{nd} stage: Stereo Algorithm to recover Z

- Need to recover R & T first.
- Let $P_L = (X_L, Y_L, Z_L)^T$ be the position of P in the left system.
- Let $P_R = (X_R, Y_R, Z_R)^T$ be the position of P in the right system.
- Then $P_R = R\ (P_L - T)$ or $P_L = (R^T P_R + T)$ where R is a 3x3 orthogonal matrix ($RR^T=I$) representing the rotation and T is the translation vector $O_R - O_L$.
- This represents a set of 3 linear equations in 12 unknowns, so a set of 4 non-coplanar pairs of *known* 3-D points is sufficient, in general to solve for the transformation.
  - In fact, orthogonality of R means only 3 points are required (if we use the fact that the 3^{rd} row of R can be obtained by $r_3 = r_1 \times r_2$).

2^{nd} stage: Stereo Algorithm to recover Z

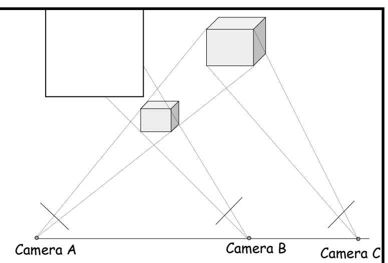
- Generally, we only have the corresponding pairs, not the 3-D coordinates of the scene points.
- For a given corresponding pair, we have:

  $r_{11}\, x_R + r_{21}\, y_R + r_{31}\, f + t_1\, f / Z_R = x_L\, Z_L / Z_R$

  $r_{12}\, x_R + r_{22}\, y_R + r_{32}\, f + t_2\, f / Z_R = y_L\, Z_L / Z_R$

  $r_{13}\, x_R + r_{23}\, y_R + r_{33}\, f + t_3\, f / Z_R = f\, Z_L / Z_R$
- This is 3 equations in 14 unknowns
- Each additional corresponding pair adds 3 equations, but also adds two unknowns (the Z coordinates)
- For n points, we have 3n equations and 12+2n unknowns. So we need at least 12 points to solve the equations.
- Later, we use a different formulation to obtain a 8-point algo.

Scale Ambiguity in Z



- Note that in the equations

  $r_{11}\, x_R + r_{21}\, y_R + r_{31}\, f + t_1\, f / Z_R = x_L\, Z_L / Z_R$

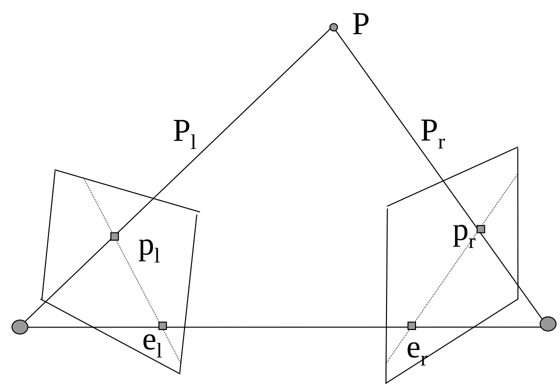
  $r_{12}\, x_R + r_{22}\, y_R + r_{32}\, f + t_2\, f / Z_R = y_L\, Z_L / Z_R$

  $r_{13}\, x_R + r_{23}\, y_R + r_{33}\, f + t_3\, f / Z_R = f\, Z_L / Z_R$

  the unknowns t_i and Z_{Lj}, Z_{Rj} appear only in ratios.
- So if T, Z_{Lj}, Z_{Rj} is a solution, then so is kT, kZ_{Lj}, kZ_{Rj} for any k not equal to 0
- This means that absolute range cannot be determined – you will see this again in motion.
- Must have one more constraint to get a unique solution (e.g. T.T=1)

Stereo Algorithm

- Toward a better algorithm by factoring out Z
- Consider equation of the epipolar plane
  - Co-planarity condition of vectors P_l, T and P_l-T



$$\mathbf{P_r} = R\,(\mathbf{P_l} - \mathbf{T})$$

Stereo Algorithm

This line basically says that the triple product of 3 coplanar vectors are zero.

$$(P_l - T)^T\, T \times P_l = (P_l - T)^T J(T) P_l = 0$$

$$\Rightarrow (R^T P_r)^T\, J(T) P_l = 0$$

$\mathbf{P_r} = R\,(\mathbf{P_l} - \mathbf{T})$

$$P_r^T R J(T) P_l = 0$$

$$P_r^T E P_l = 0$$

$$J(T) = \begin{bmatrix} 0 & -T_z & T_y \\ T_z & 0 & -T_x \\ -T_y & T_x & 0 \end{bmatrix}$$

In the textbook, J(T) is called S. Any cross product between 2 vectors can be written in this form.

E=RJ(T) known as **Essential Matrix.**

Multiply by $f_r f_l$ / ($Z_r Z_l$), we can now relate pts in image plane:

$$\mathbf{p}_l = \frac{f_l}{Z_l}\mathbf{P_l}$$

$$\mathbf{p}_r = \frac{f_r}{Z_r}\mathbf{P}_r$$

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

$\mathbf{p_l}$ = [x_l, y_l, f_l]^T is the projection of $\mathbf{P_l}$ in the left image plane, expressed in the left camera reference frame, i.e., $\mathbf{p_l}$ = f_l $\mathbf{P_l}$ /Zl.

Stereo Algorithm

- Bringing in pixel coordinates:

$$\mathrm{p}_l = M_l^{-1}\overline{\mathrm{p}}_l \quad ; \quad \mathbf{p}_r = M_r^{-1}\overline{\mathbf{p}}_r .$$

$$M = \begin{pmatrix} \frac{-f}{s_x} & 0 & o_x \\ 0 & \frac{-f}{s_y} & o_y \\ 0 & 0 & 1 \end{pmatrix}$$

- Substituting into $\mathbf{p}_r^T E \mathbf{p}_l = 0$

$$\Rightarrow \overline{\mathrm{p}}_r^T F \overline{\mathrm{p}}_l = 0,$$

where $F = M_r^{-T} E M_l^{-1}$.

- The matrix F is called the **Fundamental matrix.**
- Both E and F have rank 2 because
  - E = R J(T) and J(T) has rank 2, and $F = M_r^{-T} E M_l^{-1}$.
  - E has only 5 d.o.f.; this translates into further constraints compared to F: E=U diag(σ, σ, 0) V^T.
- The matrix F maps points in the left image onto the corresponding epipolar line in the right image.

Duality in $\mathcal{P}^2$: Points and lines

- In $\mathcal{P}^2$, there are objects like points and lines.
- Recall a point is defined by 3 numbers, $\mathbf{x}=(x_1, x_2, x_3)^T$.
- A line is also defined by 3 numbers $\mathbf{u}=(u_1, u_2, u_3)^T$ such that $\mathbf{u}^T \mathbf{x} = 0$. Intuitively corresponds to the following:

  Inhomogeneous coord: ax+by+c =0

  Homogeneous coord: ax+by+cz =0
- Formally there is no difference between points & lines in $\mathcal{P}^2$.
  - Known as the principle of duality.
- A point represented by **x** can be thought as the set of lines through it. These lines are represented by **u**, satisfying $\mathbf{u}^T \mathbf{x} = 0$.
- Inversely, a line represented by **u** can be thought as the set of points represented by **x** and satisfying the same equation.
- Now interpret $\overline{\mathrm{p}}_r^T F \overline{\mathrm{p}}_l = 0$

Recover F from point correspondences

- If we know F, then we know the epipolar geometry. What does that mean?
  - Given a point in the left image, we know how to map to the corresponding epipolar line in the right image.
  - We know where the epipoles are (compute L & R epipoles using null(F) and null(F^T) respectively). Self-reading (sect 7.3.6).
- How to recover F from a few point matches?
  - 7-point algo, 8-point algo, normalized 8-point algo, nonlinear algo minimizing distance between point and epipolar lines, etc.
  - We focus on the classical 8-point algorithm.
- If the intrinsic parameters are known, easy to recover E from F :

$$F = M_r^{-T} E M_l^{-1}.$$

Recover F from point correspondences

The Eight-Point Algorithm:

- Write $\bar{p}_r^T F \bar{p}_l = 0$ as

$$\begin{aligned} &x_l x_r F_{11} + x_l y_r F_{21} + x_l F_{31} + \\ &y_l x_r F_{12} + y_l y_r F_{22} + y_l F_{32} + \\ &x_r F_{13} + y_r F_{23} + F_{33} = 0 \end{aligned}$$

- Collecting these equations for all corresponding points (minimum 8 pairs): $A f' = 0$;

  where f' denotes entries of F collected in a column vector.

- Use SVD ($A=UDV^T$) to solve for this homogeneous linear system. f' is the column of V corresponding to the only null singular value (in practice, the least singular value) of A.

- Ideal matrix F has rank 2. Noise corrupts, so rank of the computed F (call it $\tilde{F}$) not 2. How to enforce singularity?
- By SVD, decompose $\tilde{F}$

$$\tilde{F} = UDV^T$$
$$= \begin{bmatrix} & & \\ & & \\ & & \end{bmatrix} \begin{bmatrix} \sigma_1 & & \\ & \sigma_2 & \\ & & \sigma_3 \end{bmatrix} \begin{bmatrix} & & \\ & & \\ & & \end{bmatrix}$$

- Partitioning, we get $UDV^T = U'D'V'^T + U'' D'' V''^T$

where
$$U = \begin{array}{cc} [U' & U''] \\ 2 & 1 \end{array} \; 3 \qquad D = \begin{array}{c} \begin{bmatrix} D' & 0 \\ 0 & D'' \end{bmatrix} \\ \begin{array}{cc} 2 & 1 \end{array} \end{array} \begin{array}{c} 2 \\ 1 \\ \end{array} \qquad V^T = \begin{array}{c} \begin{bmatrix} V'^T \\ V''^T \end{bmatrix} \\ 3 \end{array} \begin{array}{c} 2 \\ 1 \\ \end{array}$$

- By rank theorem, ideal matrix F has rank 2. i.e. $D'' = \sigma_3 = 0$.
- $U'' D'' V''^T$ due to noise only, and $U'D'V'^T$ represents best approximation of the ideal matrix F with rank 2 as required.

RANSAC based on 8-point Algorithm

- How to choose the 8 corresponding pairs to determine F? Recall correspondence is likely to contain errors.
- The following algorithm (RANSAC) has been proven successful. Given a number of extracted points in two images:
  1. Choose at random 8 points from each image and calculate the essential / fundamental matrix.
  2. For each point in the first image find a corresponding point (if possible) on the corresponding epipolar line.
  3. Repeat from 1 until a sufficient number of correspondences fulfilling the epipolar constraint has been obtained.
- It is advisable to attach a score to each match, calculated by correlation and use these scores to get a measure of the quality of each match.

The Epipolar Geometry: Example

- The fundamental matrix of a stereo pair is

$$\mathbf{F} = \begin{bmatrix} 0 & -625 & 31450 \\ 625 & 0 & -24925 \\ -31450 & 24925 & 0 \end{bmatrix}$$

- Compute the epipoles, and for a point (20,30) in the left image, what is the equation of the epipolar line in the right image?

Solution:

- For the epipoles, we compute the null(**F**) and null($\mathbf{F}^T$). We get [-0.6210 -0.7836 -0.0156]^T, for both. This is [39.88 50.32] in the image coordinate system (dividing by -0.0156). Note that the epipole position may not be the same in general, it just happened for this example. Let the equation of the epipolar line $au' + bv' + c = 0$, then

$$[a\ b\ c]^T = \mathbf{F}[20\ 30\ 1]^T$$

Solving, we get $1.000u' - 0.9783v' + 9.3504 = 0$.

The Epipolar Geometry: Example

- Write down the fundamental matrix for a simple parallel camera stereo rig. You are given baseline = t_x and the two intrinsic matrices are

$$M_L = M_R = \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

What does the equation $\overline{\mathbf{p}}_{\mathbf{r}}{}^{\mathbf{T}}\mathbf{F}\overline{\mathbf{p}}_{\mathbf{l}} = 0$ reduce to?

- **Solution:**

$$F = M_r^{-T}(RJ(T))M_l^{-1} = \begin{bmatrix} 1/f & 0 & 0 \\ 0 & 1/f & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -t_x \\ 0 & t_x & 0 \end{bmatrix} \begin{bmatrix} 1/f & 0 & 0 \\ 0 & 1/f & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix}$$

$\overline{\mathbf{p}}_{\mathbf{r}}{}^{\mathbf{T}}\mathbf{F}\overline{\mathbf{p}}_{\mathbf{l}} = 0$ reduces to $y_r = y_l$, i.e. (horizontal disparity)

The Epipolar Geometry: Example

- The epipole e is the null space vector of F (exercise) such that Fe=0. In this case:

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \mathbf{0}$$

Interpret the geometric meaning of this $e = (1,0,0)^T$.

3-D Reconstruction

- Given a set of corresponding points, need to compute 3-D coordinates. Complexity of problem depends on whether stereo system is calibrated.
  - We study the easier case of a calibrated system (intrinsic parameters known) : Calibrated reconstruction.
- Also - reconstruction will only result in sparse set of 3-D points
  - need to fill in the gaps between feature points, eg by surface fitting.
  - Filling in requires model of 3-D structure, eg piecewise planar, B-splines, etc.

L 

 R

VRML model of reconstructed head

Calibrated Reconstruction

Briefly, the following steps are involved:

- Identify a number (at least 8) of point correspondences
- Estimate the fundamental matrix F using the 8-point algorithm (for robustness, use RANSAC)
- Calculate the essential matrix E from F and the camera calibration matrices M_l, M_r.
- Extract the rotation and translation components from E.
- Determine the 3D point locations.

Calibrated Reconstruction

- Essential matrix E can be determined from the F and the camera calibration matrices M_l, M_r :

$$E = M_r^{\ T} F M_l$$

- Determine the SVD of E:

$$E = U S V^T$$

- Rotation and translation are given by

$$R = U W V^T \text{ or } U W^T V^T$$

$$T = v_3 \text{ or } - v_3$$

where v_3 is the last column of V and $W = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

Calibrated Reconstruction

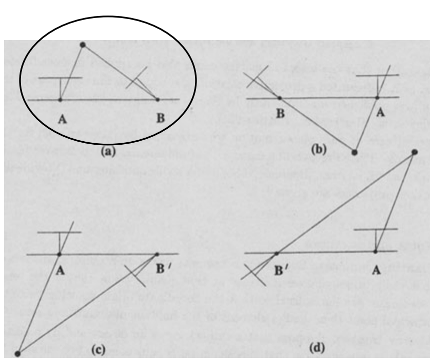
- The result for T is merely from the observation that the epipole is given by the null space of E *(Textbook: sect 7.3.6, p156, self-reading)* and the null space of E is given by v_3 *(property 4 of SVD: textbook p323)*.

- *For detailed proof of the rest: see Hartley, p239 (not required). Note that in Hartley, the translation* $\mathbf{t}$ *is our* $-RT$ *(Hartley, p143), and thus its solution for "translation"* $\mathbf{t}$ *is different, given by* u_3 or - u_3 *instead. You should be able to derive one from the other, however.*

- *Textbook p165 also gives an alternative solution for R. Not required.*

Calibrated Reconstruction

- So we have 4 possible combinations of (R , T):
  ($U\,W\,V^T$, v_3), ($U\,W\,V^T$, -v_3), ($U\,W^T\,V^T$, v_3), ($U\,W^T\,V^T$, -v_3)
- We can disambiguate between the solutions by choosing the one which corresponds to having all of the recovered points in front of both cameras (positive depth constraint).

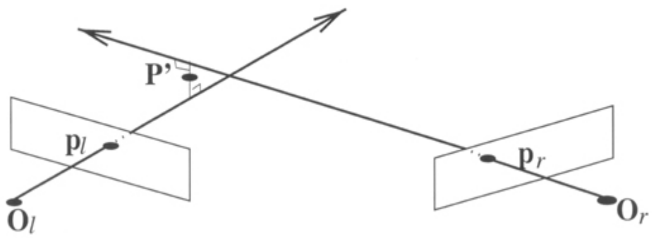


The relationship between the rotation solutions are not so obvious, but it can be shown that (Hartley, p240) they are a twisted pair, ie, one can be obtained from the other by a rotation through 180° about the line joining the 2 camera centers.

Reconstruction by Triangulation

- Given corresponding points p_l and p_r, and camera extrinsic parameters T and R, need to determine where rays ap_l and $b\,p_r$ intersect, ie need to find a and b.
- w.r.t. left camera frame: rays given by ap_l and $T + bR^T p_r$
- In general, rays will not intersect due to errors in calibration and correspondences, and pixelization

  -- find closest point to both rays.



Finding the Best 3-D Point

- Denote rays ap_l and $T + bR^T p_r$ by l and r.
- Required point P' is then the mid-point of segment which is perpendicular to l and r AND joins l and r.
- We can find the endpoints of this segment, say $a_o\, p\, l$ and $T + b_o R^T p_r$, by solving the following system of 3 linear equations:

$$ap_l - (T + bR^T p_r) + c\,(p_l \times R^T p_r) = 0$$

  - $ap_l - (T + bR^T p_r)$ is a segment joining l and r.
  - $c\,(p_l \times R^T p_r)$ is perpendicular to l and r.

- Choose a, b and c to make their difference = 0 therefore gives the required segment and hence the mid-point P'.

Summary

- Equation of the epipolar plane
  - Co-planarity condition of vectors P_l, T and P_l-T

$$(\mathbf{P_l}-\mathbf{T})^T\mathbf{T}\times\mathbf{P_l}=0$$

$$\mathbf{P_r}=\mathbf{R}(\mathbf{P_l}-\mathbf{T})$$

- Essential Matrix E = RS
  - 3x3 matrix constructed from R and T (extrinsic only)
    - Rank (E) = 2, two equal nonzero singular values

$$S=\begin{bmatrix} 0 & -T_z & T_y \\ T_z & 0 & -T_x \\ -T_y & T_x & 0 \end{bmatrix}$$

Rank (S) =2

$$\mathbf{P_r}^{\mathbf{T}}\mathbf{E}\mathbf{P_l}=0$$

$$\mathbf{p}_l=\frac{f_l}{Z_l}\mathbf{P_l} \qquad \mathbf{p}_r=\frac{f_r}{Z_r}\mathbf{P}_r$$

$$\mathbf{p_r}^{\mathbf{T}}\mathbf{E}\mathbf{p_l}=0$$

Summary

- From Camera to Pixels: Matrices of intrinsic parameters

$$\mathbf{M}_{\text{int}}=\begin{bmatrix} -f_x & 0 & o_x \\ 0 & -f_y & o_y \\ 0 & 0 & 1 \end{bmatrix}$$

Rank (M_{int}) =3

$$\mathbf{p_r}^{\mathbf{T}}\mathbf{E}\mathbf{p_l}=0$$

$$\mathbf{p_l}=\mathbf{M_l^{-1}}\bar{\mathbf{p}}_\mathbf{l} \qquad \mathbf{p}_r=\mathbf{M}_r^{-1}\bar{\mathbf{p}}_r$$

$$\bar{\mathbf{p}}_\mathbf{r}^{\mathbf{T}}\mathbf{F}\bar{\mathbf{p}}_\mathbf{l}=0$$

$$\mathbf{F}=\mathbf{M}_r^{-\mathbf{T}}\mathbf{E}\mathbf{M}_l^{-1}$$

- Fundamental Matrix
  - Rank (F) = 2
  - Encodes info on both intrinsic and extrinsic parameters

$$\bar{\mathbf{p}}_\mathbf{r}^{\mathbf{T}}\mathbf{F}\bar{\mathbf{p}}_\mathbf{l}=0 \Rightarrow (x_{im}^{(l)} \quad y_{im}^{(l)} \quad 1)\begin{bmatrix} f11 & f12 & f13 \\ f21 & f22 & f23 \\ f31 & f32 & f33 \end{bmatrix}\begin{pmatrix} x_{im}^{(r)} \\ y_{im}^{(r)} \\ 1 \end{pmatrix}=0$$

Summary

- Input: n point correspondences (n >= 8)
  - Construct homogeneous system Ax= 0 from $\bar{\mathbf{p}}_{\mathbf{r}}^{\mathbf{T}}\mathbf{F}\bar{\mathbf{p}}_{\mathbf{l}} = 0$
    - x = (f_{11},f_{12}, ,f_{13}, f_{21},f_{22},f_{23} f_{31},f_{32}, f_{33}) : entries in F
    - Each correspondence give one equation
    - A is a nx9 matrix
  - Obtain estimate F^ by SVD of A $\mathbf{A} = \mathbf{UDV}^T$
    - x (up to a scale) is column of V corresponding to the least singular value
  - Enforce singularity constraint: since Rank (F) = 2
    - Compute SVD of F^ $\hat{\mathbf{F}} = \mathbf{UDV}^T$
    - Set the smallest singular value to 0: D -> D'
    - Correct estimate of F : $\mathbf{F'} = \mathbf{UD'V}^T$
- Output: the estimate of the fundamental matrix, F'
- Similarly we can compute E given intrinsic parameters
- To find left epipole: $\mathbf{F}\bar{\mathbf{e}}_{\mathbf{l}} = 0$ find the vector that maps F to 0

Reference: Trucco

- Sect 7.1, 7.2 for introduction & correspondence
  - (7.2.3 will be covered in the chapter on motion since it is not really specific to stereo)
- Sect 7.3 for epipolar geometry (up to and including sect 7.3.6).
- Sect 7.3.7 (rectification) is not required, but would be useful if you want to apply a transformation to make all epipolar lines horizontal (after which, say, Marr-Poggio's algorithm can then be applied).
- Sect 7.4 for reconstruction (up to and including sect 7.4.2).