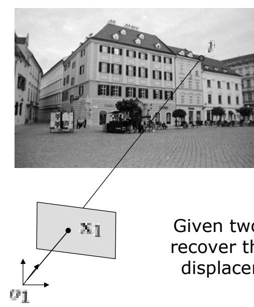
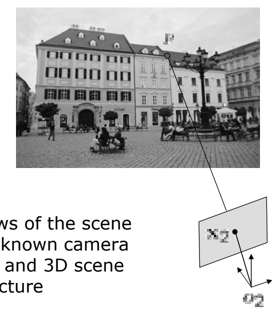
Two-View Geometry - Stereo

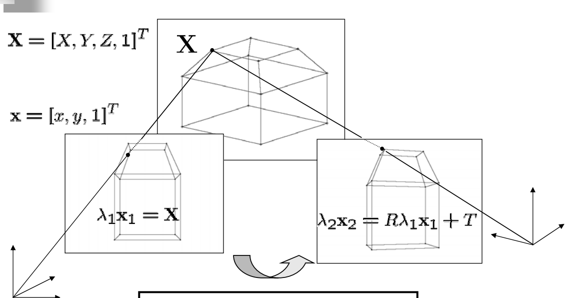
General Formulation



Given two views of the scene recover the unknown camera displacement and 3D scene structure



Rigid Body Motion – Two Views



$$\lambda_2 \mathbf{x}_2 = R\lambda_1 \mathbf{x}_1 + T$$

$$\lambda \mathbf{x} = \Pi \mathbf{X} = [R, T]\mathbf{X} \qquad \Pi = [R, T] \in \Re^{3\times 4}$$

3D Structure and Motion Recovery

Euclidean transformation

$$\lambda_2 \mathbf{x}_2 = R\lambda_1 \mathbf{x}_1 + T$$

measurements unknowns

$$\sum_{j=1}^{n} \|\mathbf{x}_1^j - \pi(R_1, T_1, \mathbf{X})\|^2 + \|\mathbf{x}_2^j - \pi(R_2, T_2, \mathbf{X})\|^2$$

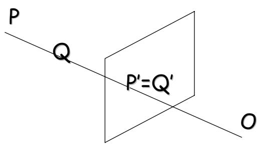
Find such Rotation and Translation and Depth that the reprojection error is minimized

Two views ~ 200 points
6 unknowns – Motion 3 Rotation, 3 Translation
- Structure 200x3 coordinates
- (-) universal scale

Difficult optimization problem

Why Stereo Vision?

- 2D images project 3D points into 2D:

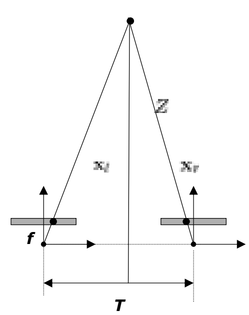


- 3D Points on the same viewing line have the same 2D image:
  - 2D imaging results in depth information loss

Stereo

- Assumes (two) cameras.
- Known positions.
- Recover depth.

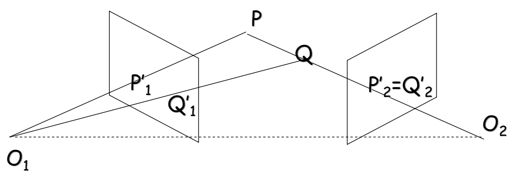
Stereo – Special Configuration



$$\frac{Z}{T} = \frac{Z-f}{T-\mathrm{x}_l-\mathrm{x}_r}$$

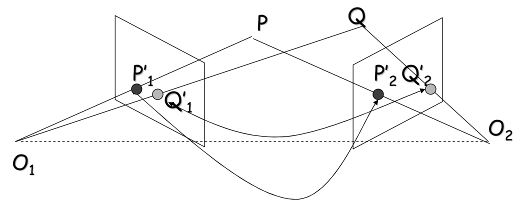
$$Z = \frac{fT}{\text{disparity}}$$

Recovering Depth Information:



Depth can be recovered with two images and triangulation.

Finding Correspondences:



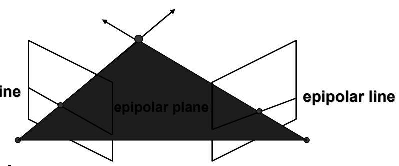
Finding Correspondences:

Simplest Case

- Image planes of cameras are parallel.
- Focal points are at same height.
- Focal lengths same.
- Then, epipolar lines are horizontal scan lines.

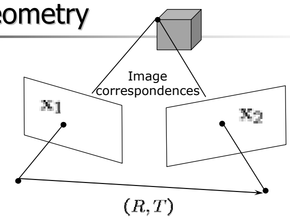
Stereo correspondence

- Determine Pixel Correspondence
  - Pairs of points that correspond to same scene point



- Epipolar Constraint
  - Reduces correspondence problem to 1D search along *corresponding epipolar lines*

Epipolar Geometry



$$\lambda_2 \mathbf{x}_2 = R\lambda_1 \mathbf{x}_1 + T \quad / \widehat{\mathbf{x}_2} T$$

- Algebraic Elimination of Depth [Longuet-Higgins '81]:

$$\mathbf{x}_2^T \underbrace{\widehat{T} R}_{E} \mathbf{x}_1 = 0$$

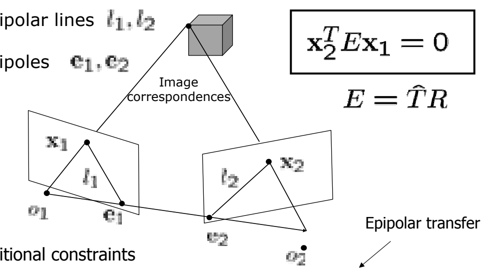
- Essential matrix $E = \widehat{T} R$

Epipolar Geometry

- Epipolar lines l_1, l_2
- Epipoles $\mathbf{e}_1, \mathbf{e}_2$

$$\mathbf{x}_2^T E \mathbf{x}_1 = 0$$

$$E = \widehat{T} R$$



- Additional constraints

$l_1 \sim E^T \mathbf{x}_2 \qquad l_i^T \mathbf{x}_i = 0 \qquad l_2 \sim E\mathbf{x}_1$

$E\mathbf{e}_1 = 0 \qquad l_i^T \mathbf{e}_i = 0 \qquad \mathbf{e}_2 E^T = 0$

Characterization of Essential Matrix

$$\mathbf{x}_2^T \widehat{T} R \mathbf{x}_1 = 0$$

Essential matrix $E = \widehat{T} R$ special 3x3 matrix

$$\mathbf{x}_2^T \begin{bmatrix} e_1 & e_2 & e_2 \\ e_4 & e_5 & e_6 \\ e_7 & e_8 & e_9 \end{bmatrix} \mathbf{x}_1 = 0$$

(Essential Matrix Characterization)
A non-zero matrix E is an essential matrix iff its SVD: $E = U\Sigma V^T$ satisfies: $\Sigma = diag([\sigma_1, \sigma_2, \sigma_3])$ with $\sigma_1 = \sigma_2 \neq 0$ and $\sigma_3 = 0$ and $U, V \in SO(3)$

Estimating Essential Matrix

- Find such Rotation and Translation that the epipolar error is minimized

$$min_E \sum_{j=1}^{n} \mathbf{x}_2^{jT} E \mathbf{x}_1^j$$

- Space of all Essential Matrices is 5 dimensional
- 3 DOF Rotation, 2 DOF – Translation (up to scale !)
- Denote $\mathbf{a} = \mathbf{x}_1 \otimes \mathbf{x}_2$

$$\mathbf{a} = [x_1x_2, x_1y_2, x_1z_2, y_1x_2, y_1y_2, y_1z_2, z_1x_2, z_1y_2, z_1z_2]^T$$

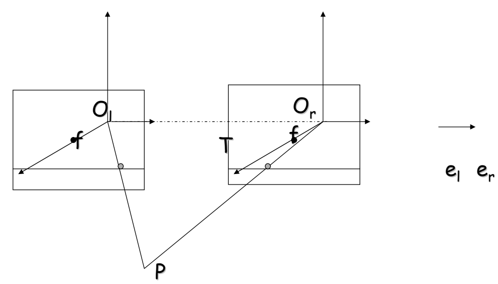
$$E^s = [e_1, e_4, e_7, e_2, e_5, e_8, e_3, e_6, e_9]^T$$

- Rewrite $\mathbf{a}^T E^s = 0$
- Collect constraints from all points

$$\chi E^s = 0$$

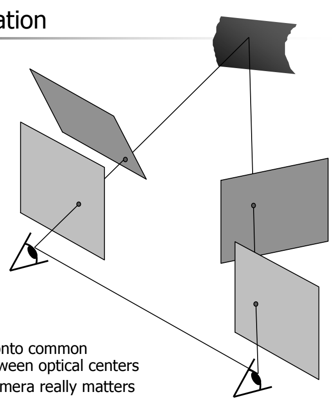
$$min_E \sum_{j=1}^{n} \mathbf{x}_2^{jT} E \mathbf{x}_1^j \Rightarrow min_{E^s} \|\chi E^s\|^2$$

Epipolar Geometry for Parallel Cameras



Epipoles are at infinite
Epipolar lines are parallel to the baseline

Image rectification



- Image Reprojection
  - reproject image planes onto common plane parallel to line between optical centers
- Notice, only focal point of camera really matters

(Seitz)

Epipolar rectification

- Rectified Image Pair
- Corresponding epipolar lines are aligned with the scan-lines
- Search for dense correspondence is a 1D search

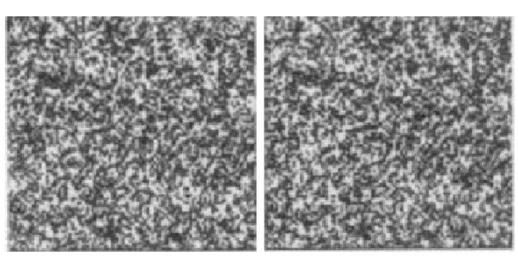
Epipolar rectification

Rectified Image Pair

Correspondence: What should we match?

- Objects?
- Edges?
- Pixels?
- Collections of pixels?

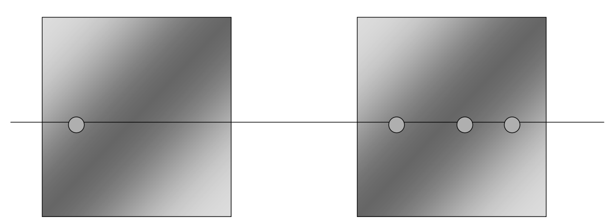
Random dot stereograms



B. Julesz: Showed that correspondence is not needed for stereo.

Correspondence: Epipolar constraint.

Corresponding point has to lie on the epipolar line



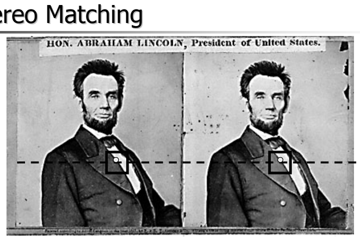
Correspondence Problem

- Two classes of algorithms:
  - Correlation-based algorithms
    - Produce a DENSE set of correspondences
  - Feature-based algorithms
    - Produce a SPARSE set of correspondences

Correspondence: Photometric constraint

- Same world point has same intensity in both images.
  - Lambertian fronto-parallel
  - Issues:
    - Noise
    - Specularity
    - Foreshortening

Stereo Matching



For each epipolar line
  For each pixel in the left image
  - compare with every pixel on same epipolar line in right image
  - pick pixel with minimum match cost
  - This will never work, so:

Improvement: match ***windows***

(slides O. Camps)

Comparing Windows:

f ?= g

$$SSD = \sum_{[i,j] \in R} (f(i,j) - g(i,j))^2$$

$$C_{fg} = \sum_{[i,j] \in R} f(i,j)g(i,j)$$

} Most popular

For each window, match to closest window on epipolar line in other image.

(slides O. Camps)

Comparing Windows:

Minimize $\sum_{[i,j]\in R} (f(i,j) - g(i,j))^2$ Sum of Squared Differences

Maximize $C_{fg} = \sum_{[i,j]\in R} f(i,j)g(i,j)$ Cross correlation

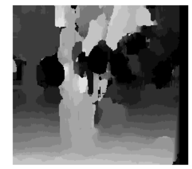
It is closely related to the SSD:

$$SSD = \sum_{[i,j]\in R} (f-g)^2 =$$

$$= \sum_{[i,j]\in R} f^2 + \sum_{[i,j]\in R} g^2 - 2 \sum_{[i,j]\in R} fg$$

Window size

W = 3



W = 20

- Effect of window size

  Better results with *adaptive window*

  - T. Kanade and M. Okutomi, *A Stereo Matching Algorithm with an Adaptive Window: Theory and Experiment*,, Proc. International Conference on Robotics and Automation, 1991.
  - D. Scharstein and R. Szeliski. Stereo matching with nonlinear diffusion. International Journal of Computer Vision, 28(2):155-174, July 1998

(S. Seitz)

Stereo results

- Data from University of Tsukuba

Scene

Ground truth

(Seitz)

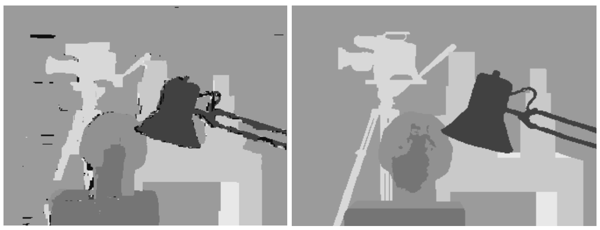
Results with window correlation

Window-based matching (best window size)

Ground truth

(Seitz)

Results with better method



State of the art   Ground truth

Boykov et al., Fast Approximate Energy Minimization via Graph Cuts, International Conference on Computer Vision, September 1999.

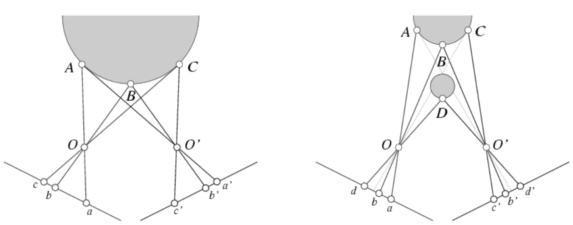
(Seitz)

Ordering constraint

- Usually, order of points in two images is same.
- Is this always true?
- If we match pixel i in image 1 to pixel j in image 2, no matches that follow will affect which are the best preceding matches.
- *Example with pixels (a la Cox et al.).*

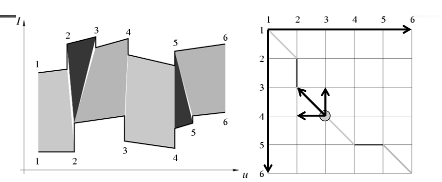
The Ordering Constraint

Points on the epipolar lines appear in the same order



But it is not always the case ...
This enables dynamic programming

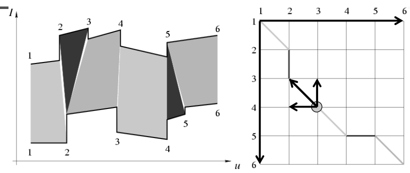
Dynamic Programming (Baker and Binford, 1981)



Find the minimum-cost path going monotonically down and right from the top-left corner of the graph to its bottom-right corner.

- Nodes = matched feature points (e.g., edge points).
- Arcs = matched intervals along the epipolar lines.
- Arc cost = discrepancy between intervals.

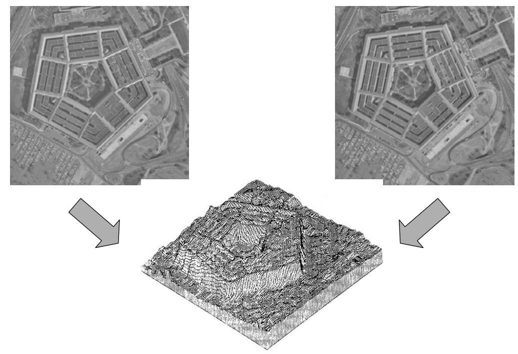
Dynamic Programming (Baker and Binford, 1981)



```
% Loop over all nodes (k,l) in ascending order.
for k = 1 to m do
  for l = 1 to n do
    % Initialize optimal cost C(k,l) and backward pointer B(k,l).
    C(k,l) ← +∞; B(k,l) ← nil;
    % Loop over all inferior neighbors (i,j) of (k,l).
    for (i,j) ∈ Inferior − Neighbors(k,l) do
      % Compute new path cost and update backward pointer if necessary.
      d ← C(i,j) + Arc − Cost(i,j,k,l);
      if d < C(k,l) then C(k,l) ← d; B(k,l) ← (i,j) endif;
      endfor;
    endfor;
  endfor;
% Construct optimal path by following backward pointers from (m,n).
P ← {(m,n)}; (i,j) ← (m,n);
while B(i,j) ≠ nil do (i,j) ← B(i,j); P ← {(i,j)} ∪ P endwhile.
```

onically … the … ge points). … lar lines. … s.

Dynamic Programming (Ohta and Kanade, 1985)



Reprinted from "Stereo by Intra- and Intet-Scanline Search," by Y. Ohta and T. Kanade, IEEE Trans. on Pattern Analysis and Machine Intelligence, 7(2):139-154 (1985). © 1985 IEEE.

Other constraints

- Smoothness: disparity usually doesn't change too quickly.
  - Unfortunately, this makes the problem 2D again.
  - Solved with a host of graph algorithms, Markov Random Fields, Belief Propagation, ….
- Uniqueness constraint (each feature can at most have one match)
- Occlusion and disparity are connected.

Feature-based Methods

- Conceptually very similar to Correlation-based methods, but:
  - They only search for correspondences of a sparse set of image features.
  - Correspondences are given by the most similar feature pairs.
  - Similarity measure must be adapted to the type of feature used.

Feature-based Methods:

- Features most commonly used:
  - Corners
    - Similarity measured in terms of:
      - surrounding gray values (SSD, Cross-correlation)
      - location
  - Edges, Lines
    - Similarity measured in terms of:
      - orientation
      - contrast
      - coordinates of edge or line's midpoint
      - length of line

Example: Comparing lines

- l_l and l_r: line lengths
- θ_l and θ_r: line orientations
- (x_l,y_l) and (x_r,y_r): midpoints
- c_l and c_r: average contrast along lines
- $\omega_l\ \omega_\theta\ \omega_m\ \omega_c$: weights controlling influence

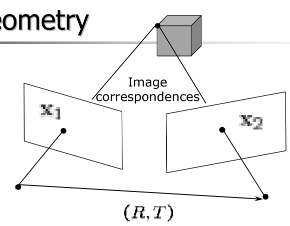
$$S = \frac{1}{\omega_l(l_l - l_r)^2 + \omega_\theta(\theta_l - \theta_r)^2 + \omega_m[(x_l - x_r)^2 + (y_l - y_r)^2] + \omega_c(c_l - c_r)^2}$$

The more similar the lines, the larger S is!

Summary

- First, we understand constraints that make the problem solvable.
  - Some are hard, like epipolar constraint.
    - Ordering isn't a hard constraint, but most useful when treated like one.
  - Some are soft, like pixel intensities are similar, disparities usually change slowly.
- Then we find optimization method.
  - Which ones we can use depend on which constraints we pick.

Epipolar Geometry



$$\lambda_2 \mathbf{x}_2 = R\lambda_1 \mathbf{x}_1 + T \quad / \widehat{\mathbf{x}_2}T$$

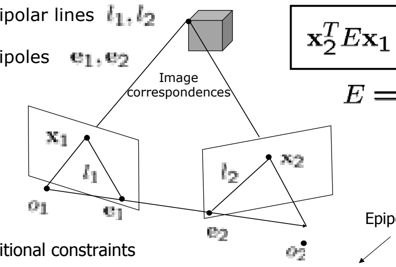
- Algebraic Elimination of Depth [Longuet-Higgins '81]:

$$\mathbf{x}_2^T \underbrace{\hat{T}R}_{E} \mathbf{x}_1 = 0$$

- Essential matrix $E = \hat{T}R$

Epipolar Geometry

- Epipolar lines l_1, l_2
- Epipoles $\mathbf{e}_1, \mathbf{e}_2$



$$\mathbf{x}_2^T E \mathbf{x}_1 = 0$$

$$E = \hat{T}R$$

- Additional constraints

$$l_1 \sim E^T \mathbf{x}_2 \qquad l_i^T \mathbf{x}_i = 0 \qquad l_2 \sim E\mathbf{x}_1$$

$$E\mathbf{e}_1 = 0 \qquad l_i^T \mathbf{e}_i = 0 \qquad \mathbf{e}_2 E^T = 0$$

Characterization of Essential Matrix

$$\mathbf{x}_2^T \hat{T} R \mathbf{x}_1 = 0$$

Essential matrix $E = \hat{T}R$ special 3x3 matrix

$$\mathbf{x}_2^T \begin{bmatrix} e_1 & e_2 & e_2 \\ e_4 & e_5 & e_6 \\ e_7 & e_8 & e_9 \end{bmatrix} \mathbf{x}_1 = 0$$

(Essential Matrix Characterization)
A non-zero matrix E is an essential matrix iff its SVD: $E = U\Sigma V^T$ satisfies: $\Sigma = diag([\sigma_1, \sigma_2, \sigma_3])$ with $\sigma_1 = \sigma_2 \neq 0$ and $\sigma_3 = 0$ and $U, V \in SO(3)$

Estimating Essential Matrix

- Find such Rotation and Translation that the epipolar error is minimized

$$min_E \sum_{j=1}^n \mathbf{x}_2^{jT} E \mathbf{x}_1^j$$

- Space of all Essential Matrices is 5 dimensional
- 3 DOF Rotation, 2 DOF – Translation (up to scale !)
- Denote $\mathbf{a} = \mathbf{x}_1 \otimes \mathbf{x}_2$

$$\mathbf{a} = [x_1x_2, x_1y_2, x_1z_2, y_1x_2, y_1y_2, y_1z_2, z_1x_2, z_1y_2, z_1z_2]^T$$

$$E^s = [e_1, e_4, e_7, e_2, e_5, e_8, e_3, e_6, e_9]^T$$

- Rewrite $\mathbf{a}^T E^s = 0$
- Collect constraints from all points

$$\chi E^s = 0$$

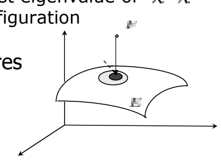
$$min_E \sum_{j=1}^n \mathbf{x}_2^{jT} E \mathbf{x}_1^j \Rightarrow min_{E^s} \|\chi E^s\|^2$$

Estimating Essential Matrix

$$min_E \sum_{j=1}^n \mathbf{x}_2^{jT} E \mathbf{x}_1^j \Rightarrow min_{E^s} \|\chi E^s\|^2$$

Solution E_s
- Eigenvector associated with the smallest eigenvalue of $\chi^T \chi$
- If $rank(\chi^T \chi) < 8$ degenerate configuration

E^s estimated using linear least squares
unstack E^s -> F



Projection on to Essential Space

(Project to Essential Manifold)
If the SVD of a matrix $F \in \mathcal{R}^{3\times3}$ is given by $F = Udiag(\sigma_1, \sigma_2, \sigma_3)V^T$ then the essential matrix E which minimizes the Frobenius distance $\|E - F\|_f^2$ is given by $E = Udiag(\sigma, \sigma, 0)V^T$ with $\sigma = \frac{\sigma_1 + \sigma_2}{2}$

Pose Recovery from Essential Matrix

Essential matrix $E = \hat{T}R$

(Pose Recovery)
There are two relative poses (R,T) with $T \in \mathcal{R}^3$ and $R \in SO(3)$ corresponding to a non-zero matrix essential matrix.

$$E = U\Sigma V^T$$

$$(\hat{T}_1, R_1) = (UR_Z(+\tfrac{\pi}{2})\Sigma U^T, UR_Z^T(+\tfrac{\pi}{2})V^T)$$
$$(\hat{T}_2, R_2) = (UR_Z(-\tfrac{\pi}{2})\Sigma U^T, UR_Z^T(-\tfrac{\pi}{2})V^T)$$

$$\Sigma = diag([1,1,0]) \quad R_z(+\tfrac{\pi}{2}) = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

- Twisted pair ambiguity $(R_2, T_2) = (e^{\widehat{u}\pi}R_1, -T_1)$

Two view linear algorithm - summary

$E = \{\hat{T}R | R \in SO(2), T \in S^2\}$

- Solve the LLSE problem:

$min_E \sum_{j=1}^{n} \mathbf{x}_2^{jT} E \mathbf{x}_1^j \Rightarrow \chi E^s = 0$

- Solution eigenvector associated with smallest eigenvalue of $\chi^T\chi$
- Compute SVD of F recovered from data

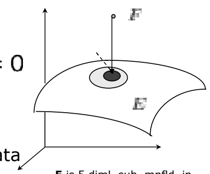
$E^s \rightarrow F \qquad F = U\Sigma V^T$

- Project onto the essential manifold:

$\Sigma' = diag(1,1,0) \;\; E = U\Sigma' V^T$

- Recover the unknown pose:

$$(\hat{T}, R) = (UR_Z(\pm\tfrac{\pi}{2})\Sigma U^T, UR_Z^T(\pm\tfrac{\pi}{2})V^T)$$



E is 5 diml. sub. mnfld. in

- 8-point linear algorithm

Pose Recovery

- There are two pairs (R,T) corresponding to essential matrix E.
- There are two pairs (R,T) corresponding to essential matrix $-E$.
- Positive depth constraint disambiguates the impossible solutions
- Translation has to be non-zero
- Points have to be in general position
  - degenerate configurations – planar points
  - quadratic surface
- Linear 8-point algorithm
- Nonlinear 5-point algorithms yield up to 10 solutions

3D Structure Recovery

$$\underline{\lambda_2}\mathbf{x}_2 = R\underline{\lambda_1}\mathbf{x}_1 + \underline{\gamma}T$$

unknowns

- Eliminate one of the scale's

$$\lambda_1^j \widehat{\mathbf{x}_2^j} R \mathbf{x}_1^j + \gamma \widehat{\mathbf{x}_2^j} T = 0, \quad j = 1, 2, \ldots, n$$

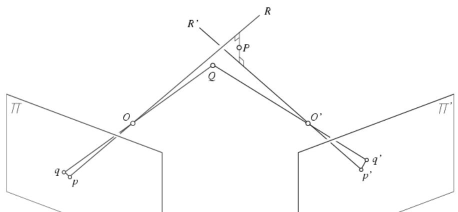
- Solve LLSE problem

$$M^j \bar{\lambda}^j \doteq \left[\widehat{\mathbf{x}_2^j} R \mathbf{x}_1^j, \; \widehat{\mathbf{x}_2^j} T\right] \begin{bmatrix} \lambda_1^j \\ \gamma \end{bmatrix} = 0$$

If the configuration is non-critical, the Euclidean structure of the points and motion of the camera can be reconstructed up to a universal scale.

- Alternatively recover each point depth separately

3D Reconstruction in general case



• Linear Method: find P such that

$$\begin{aligned}\lambda_1 \mathbf{x}_1 &= \Pi_1 \mathbf{X} \\ \lambda_2 \mathbf{x}_2 &= \Pi_2 \mathbf{X}\end{aligned} \Rightarrow \begin{aligned}\mathbf{x}_1 \times \Pi_1 \mathbf{X} &= 0 \\ \mathbf{x}_2 \times \Pi_2 \mathbf{X} &= 0\end{aligned} \Rightarrow A\mathbf{X} = 0$$

• Non-Linear Method: find Q minimizing

Example

Two views


Point Feature Matching


Example Epipolar Geometry


Camera Pose
and
Sparse Structure Recovery