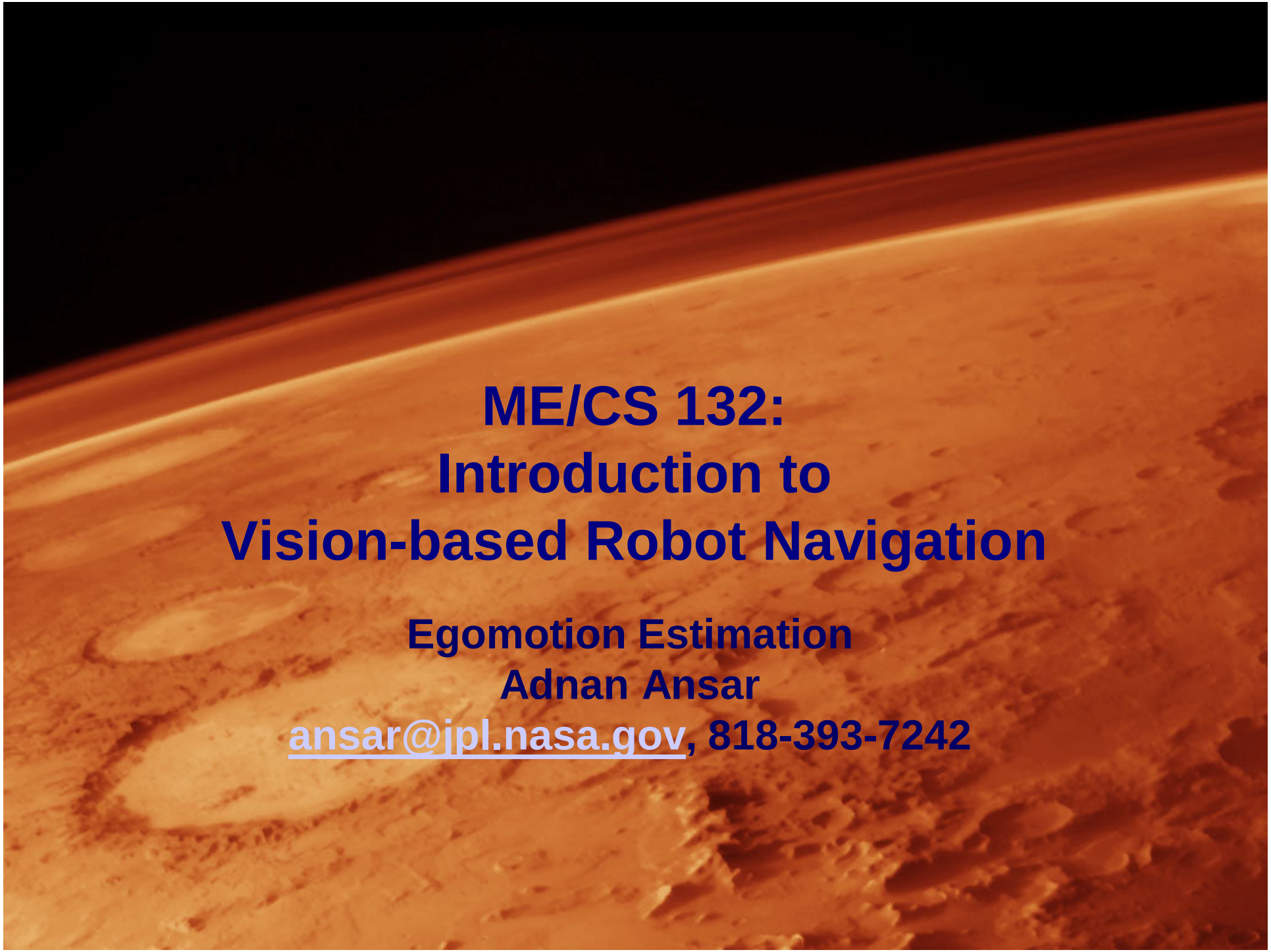


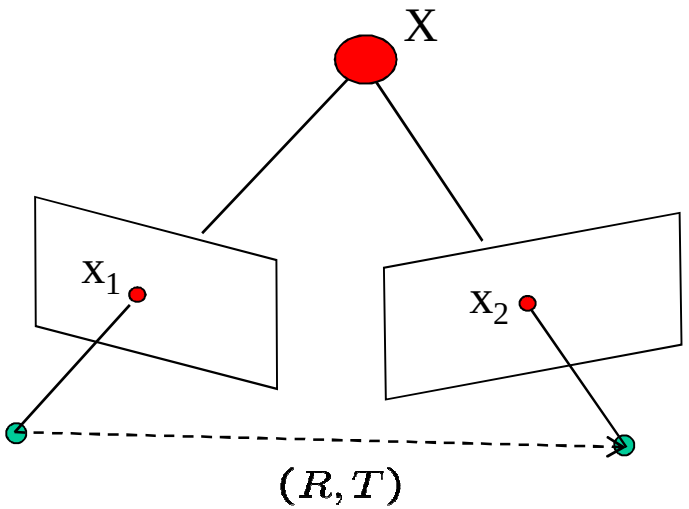
ME/CS 132:
Introduction to
Vision-based Robot Navigation

Egomotion Estimation

Adnan Ansar
ansar@jpl.nasa.gov, 818-393-7242

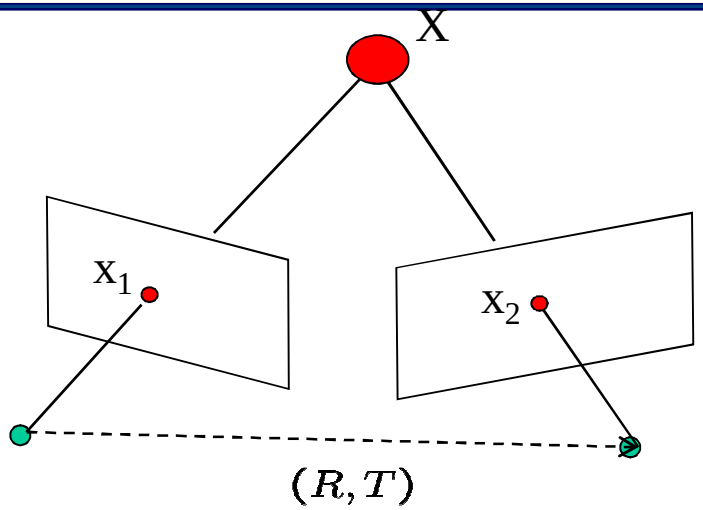
Overview

- Suppose a camera undergoes motion



- Some questions to consider:
  - What sort of constraints are imposed on the imagery?
  - Can you recover camera motion from imagery alone?
  - Can you say anything about the world based on a imagery from a moving camera?
  - What happens if the camera is not calibrated?

Motion Constraint



- Observe that X, x_1 and x_2 are coplanar, where x_1 and x_2 are in normalized coordinates. This leads to the key observation that:

$$x_2^T T_\times R x_1 = 0 \qquad T_\times = \begin{bmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{bmatrix}$$

- Why?

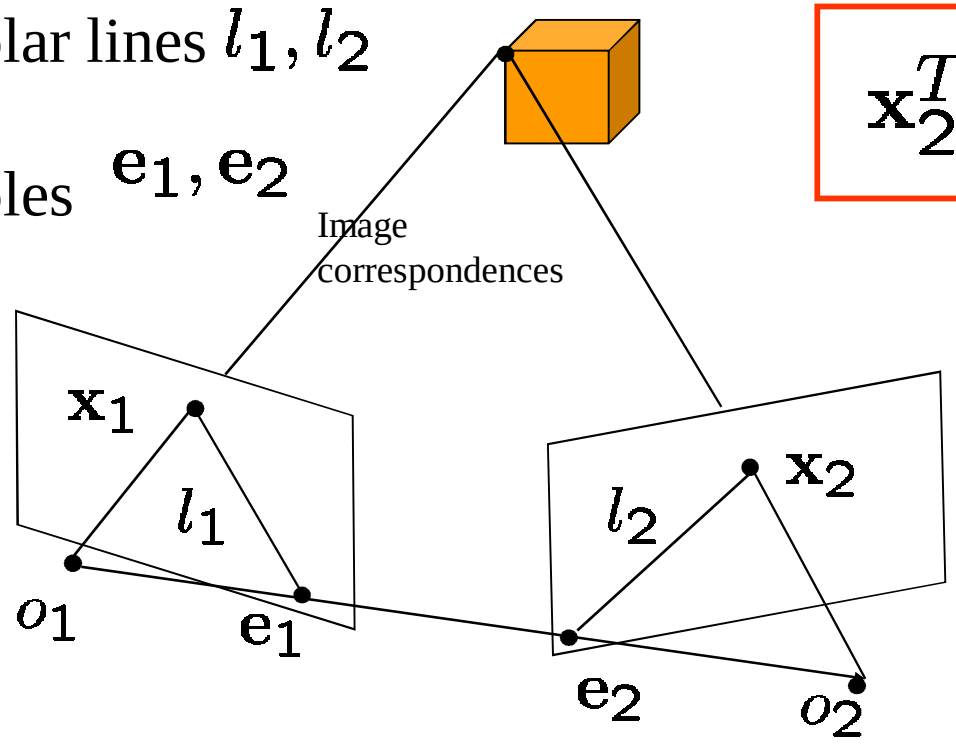
Essential Matrix

- Consider: $E = T_{\times} R$

  E is called the essential matrix. It encodes the epipolar geometry of calibrated pair of cameras.
- Properties:
  - Rank 2
  - Non-zero singular values are equal
  - Encodes only extrinsics
- Note that in the projective plane, we can think of $l_2 = Ex_1$ or $l_1 = x_2^T E$ as projective lines. This gives us a mapping from any point in image 1 to a line (the epipolar line) in image 2 and vice versa.
- Each epipolar line contains the epipole, which is the image of the other camera projection center.

Epipolar Geometry

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



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

$$E = T_\times R$$

From Ma, Soatto, Kosecka and Sastry: Invitation to 3D vision

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

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

Epipolar Geometry (Uncalibrated Case)

- If the camera is uncalibrated, we do not have normalized coordinates. Suppose K is the (unknown) calibration matrix. It still follows that

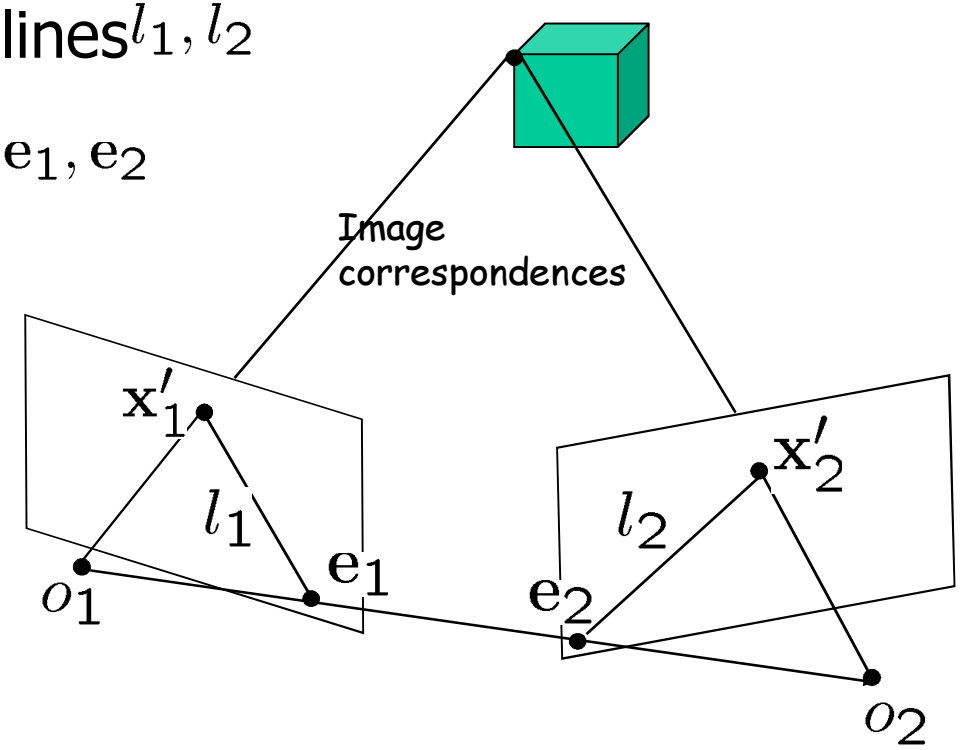
$$(K^{-1}p_2)^T E(K^{-1}p_1) = 0$$
$$p_2^T K^{-T} E K^{-1} p_1 = 0$$
$$p_2^T F p_1 = 0$$

- Still have epipolar constraint
- Properties of F matrix
  - Rank 2
  - Encodes intrinsics and extrinsics

Properties of the Fundamental Matrix

$$\mathbf{x}'^T_2 F \mathbf{x}'_1 = 0$$

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



From Ma, Soatto, Kosecka and Sastry: Invitation to 3D vision

$$l_1 \sim F^T \mathbf{x}'_2 \qquad l_i^T \mathbf{x}'_i = 0 \qquad l_2 \sim F \mathbf{x}'_1$$

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

Solving for Essential / Fundamental Matrix

- 8 Point Algorithm (Longuett-Higgins)
  - Requires at least 8 points (recall that E and F have 9 elements up to scale) in non-degenerate configuration.
  - Let

$$F = \begin{pmatrix} f_1 & f_2 & f_3 \\ f_4 & f_5 & f_6 \\ f_7 & f_8 & f_9 \end{pmatrix} \qquad \underline{f} = \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ \vdots \\ f_9 \end{pmatrix}$$

  - Each point correspondence gives us one equation of the form

$$p_{1,x}p_{2,x}f_1 + p_{1,x}p_{2,y}f_2 + p_{1,x}p_{2,z}f_3 + p_{1,y}p_{2,x}f_4 + p_{1,y}p_{2,y}f_5 + p_{1,y}p_{2,z}f_6 + p_{1,z}p_{2,x}f_7 + p_{1,z}p_{2,y}f_8 + p_{1,z}p_{2,z}f_9 = 0$$

  - Stack these to get a linear system of the form

$$A\underline{f} = 0$$

  - Right singular vector of A corresponding to smallest singular value is best solution in linear least squares sense

Recall How this was done for Calibration

- Rewrite the last set of equations as linear system in m_{ij}

$$A_{2n\times 12}\,\underline{m}_{12\times 1} = \begin{pmatrix} X_1 & Y_1 & Z_1 & 1 & 0 & 0 & 0 & 0 & -x_1X_1 & -x_1Y_1 & -x_1Z_1 & -x_1 \\ 0 & 0 & 0 & 0 & X_1 & Y_1 & Z_1 & 1 & -y_1X_1 & -y_1X_1 & -y_1X_1 & -y_1 \\ X_2 & Y_2 & Z_2 & 1 & 0 & 0 & 0 & 0 & -x_2X_2 & -x_2Y_2 & -x_2Z_2 & -x_2 \\ 0 & 0 & 0 & 0 & X_2 & X_2 & X_2 & 1 & -y_2X_2 & -y_2Y_2 & -y_2Z_2 & -y_2 \\ . & . & . & . & . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . & . & . & . & . \\ X_n & Y_n & Z_n & 1 & 0 & 0 & 0 & 0 & -x_nX_n & -x_nY_n & -x_nZ_n & -x_n \\ 0 & 0 & 0 & 0 & X_n & Y_n & Z_n & 1 & -y_nX_n & -y_nY_n & -y_nZ_n & -y_n \end{pmatrix} \begin{pmatrix} m_{11} \\ m_{12} \\ m_{13} \\ . \\ . \\ . \\ m_{33} \\ m_{34} \end{pmatrix} = 0_{6\times 1}$$

- For 6 non-coplanar points, matrix A has rank 11, hence 1-dimensional kernel. Let $A = UDV^T$ be SVD. It follows that $\underline{m}$ is column of V corresponding to smallest singular value in D.

Solving for Essential / Fundamental Matrix

- Recall that F has rank 2, while E has rank 2 and equal non-zero singular values
- Let $F = USV^T$
- For fundamental matrix, set smallest singular value to 0. For essential matrix, also set non-zero singular values equal. Let S' be the updated matrix of singular values.
- $F' = US'V^T$ is a good linear estimate of the fundamental (respectively essential) matrix

- How do we recover motion parameters?
  - From F, we can find the epipole as the smallest right singular vector
  - From E, we can find the translation vector as the smallest right singular vector
  - Why?

Decomposing the Essential Matrix

- Algebraically straightforward but messy. (BKP Horn)
- Bottom line:

$$TT^T = \frac{1}{2}\text{Trace}(EE^T)I - EE^T$$

$$T^T TR = \text{Cof}(E)^T - T_\times E$$

- Where Cof(E) is the matrix of cofactors given by

$$E = (e_1 \quad e_2 \quad e_3)$$

$$\text{Cof}(E) = (e_2 \times e_3 \quad e_3 \times e_1 \quad e_1 \times e_2)^T$$

- How many solutions do you get?
  - Positive depth constraint

Decomposing the Essential Matrix

- Linear solution is a good starting point, but in practice, non-linear least squares is always used.
- Consider epipolar constraint directly

$$(R,T) = \arg\min_{(R,T)\in SO(3)\times \mathrm{R}^3} \arg\min \sum_i \| x_{i,2}^T T_\times R x_{i,1} \|^2$$

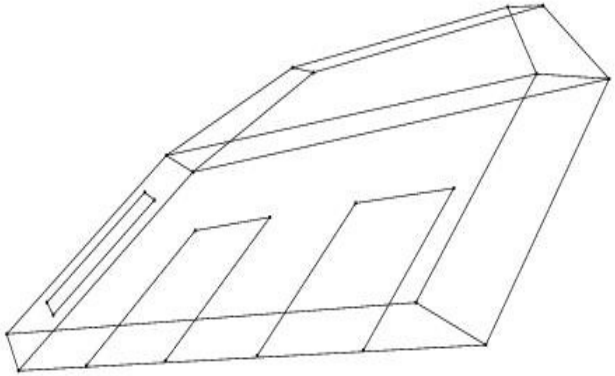
- How do you parameterize R?
  - Euler angles
  - Unit quaternion
  - Skew symmetric matrix. Consider vector Ω with angle of rotation given by θ=||Ω|| and axis of rotation given by ω= Ω\||Ω||.

$$R = \exp(\Omega_\times) = I + \sin(\theta)\omega_\times + (1-\cos(\theta))\omega_\times^2$$

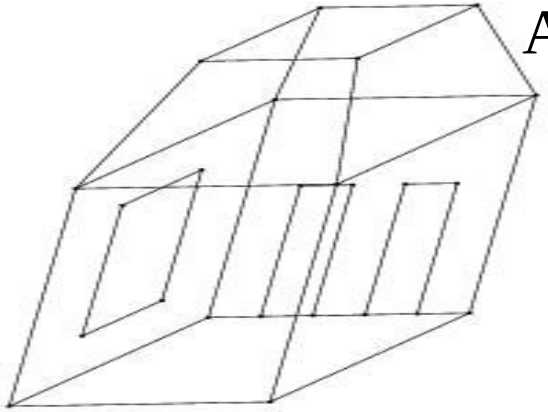
Recovering Scene Structure

- Given a calibrated camera undergoing motion (or two calibrated cameras) we have shown:
  - Epipolar constraint
  - Decomposing into relative motion
- Now triangulate point correspondences to recover 3D up to scale
  - Scale can be constrained with one piece of metric information (e.g. baseline, some object in the world)
- What about uncalibrated?
  - Cannot recover full metric reconstruction, but perspective or affine reconstruction is possible.
  - Self-calibration from motion is a possibility

Scene Structure in Uncalibrated Case

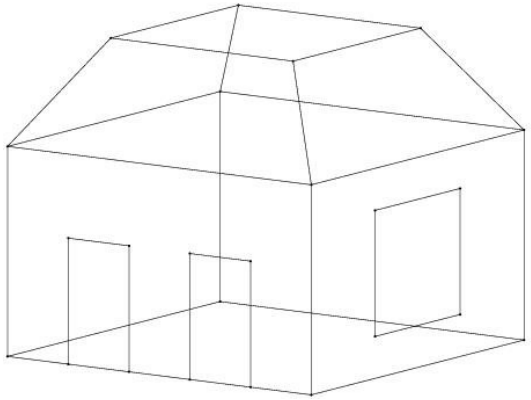


Projective



Affine

Metric

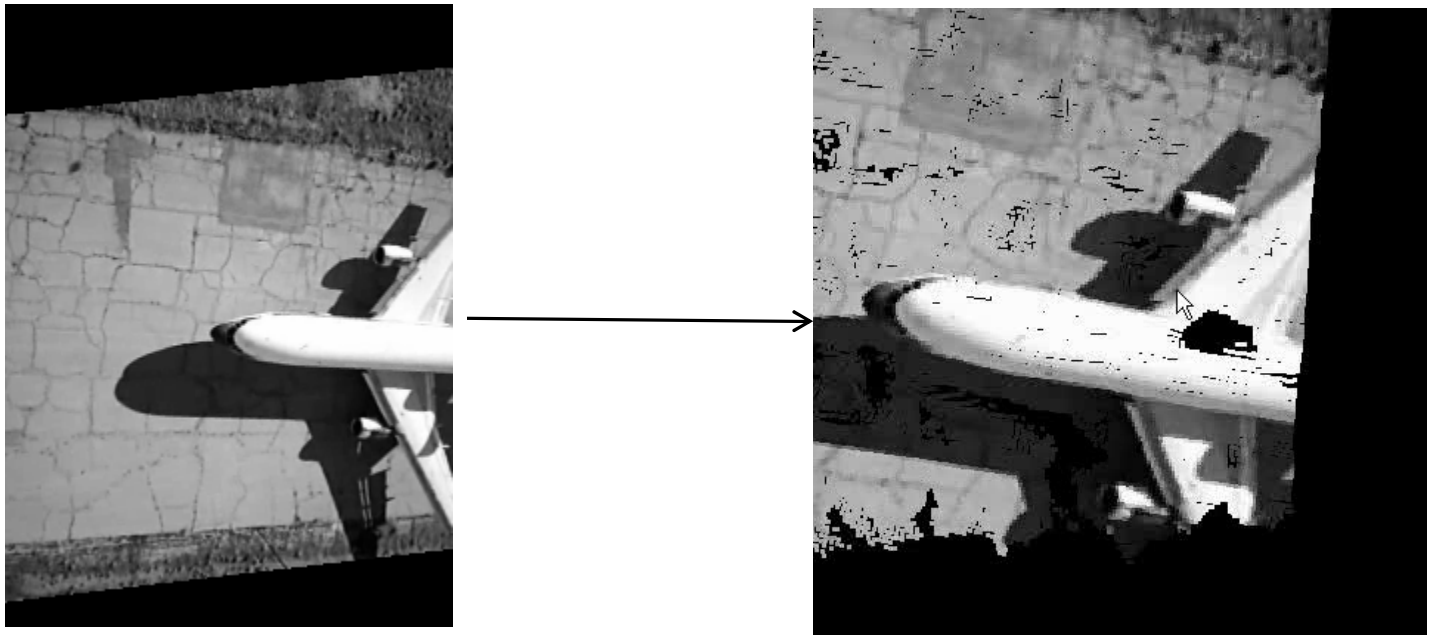


Upgrade requires additional knowledge: (motion, scene structure, calibration)

Example of Motion Recovery from Epipolar

- Motion recovered directly from epipolar geometry

$$(R,T) = \underset{(R,T)\in SO(3)\times \mathrm{R}^3}{\arg\min} \sum_i \| x_{i,2}^T T_\times R x_{i,1} \|^2$$



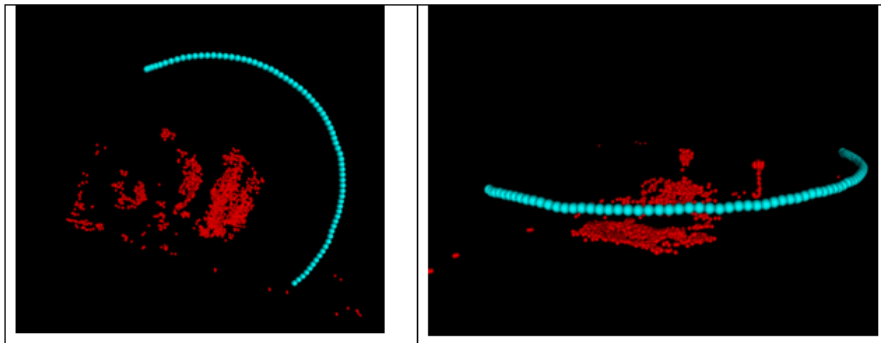
Other Considerations for Motion Estimation

- An alternative to decomposing the essential matrix is to bypass it entirely.
- Directly parameterize point triangulation based on motion estimate
- Drive parameter search by reprojection error.
  - For current motion estimate, find 3D points by triangulation
  - Reproject points into the image and compute error
  - Use some optimization technique to adjust motion estimate to reduce error
  - Typical: Levenberg-Marquardt algorithm
- In general case (many frames), called bundle adjustment
  - Drawbacks: potentially slow, potentially poorly conditioned, potentially lots of outliers.

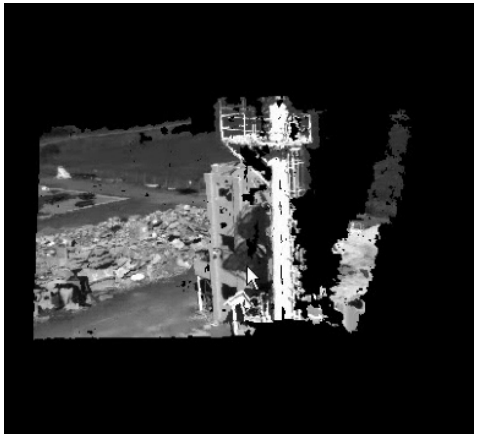
Reconstruction from Bundle Adjustment

- Imagery from a single camera is used to estimate the motion of the aircraft
- The estimated motion allows triangulation of scene points using dense stereo matching techniques

Scene



Reconstructed trajectory (blue) and tracked scene points (red)



Dense Scene Reconstruction using Vision Based Motion Estimation

Outlier Rejection Using Fundamental Matrix

- Use RANSAC with epipolar constraint.
  - Requires only point correspondences and the assumption of perspective projection
  - Does not require calibration or prior motion knowledge.
- Error term:

$$\varepsilon = p_2^T F p_1$$

- Note: Scale of error is dependent on scale of F and image scale
  - Normalize F -> straightforward
  - Normalize points -> Haralick

Point Normalization

- Suppose you have a set of point $\{p_i\}$ that are distributed in the image plane with some bias. It would be preferable to have points $\{q_i\}$ that are centered at the origin and are normalized.
- IF there was a linear transformation L with the property that

$$Lp_i = q_i$$

we could write

$$\varepsilon = q_2^T L^{-T} F L^{-1} q_1 = q_2^T F' q_1$$

Point Normalization

- Define

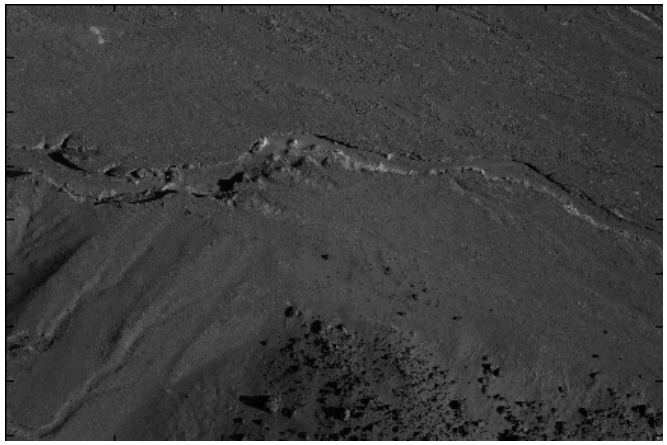
$$\overline{p} = \frac{p - \text{mean}(\{p_i\})}{\text{std.dev.}(\{p_i\})}$$

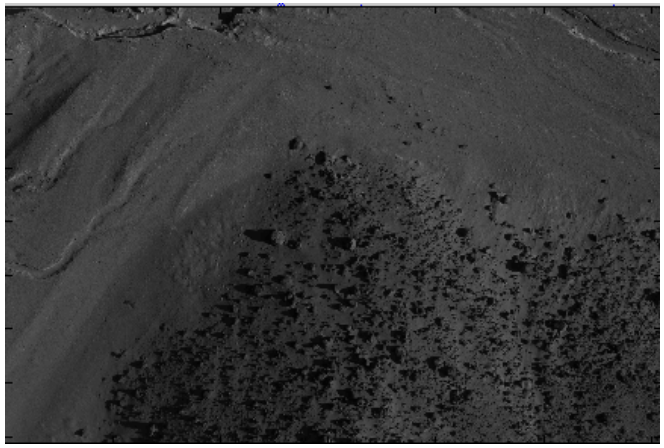
$$L = \overline{P} P^{+}$$

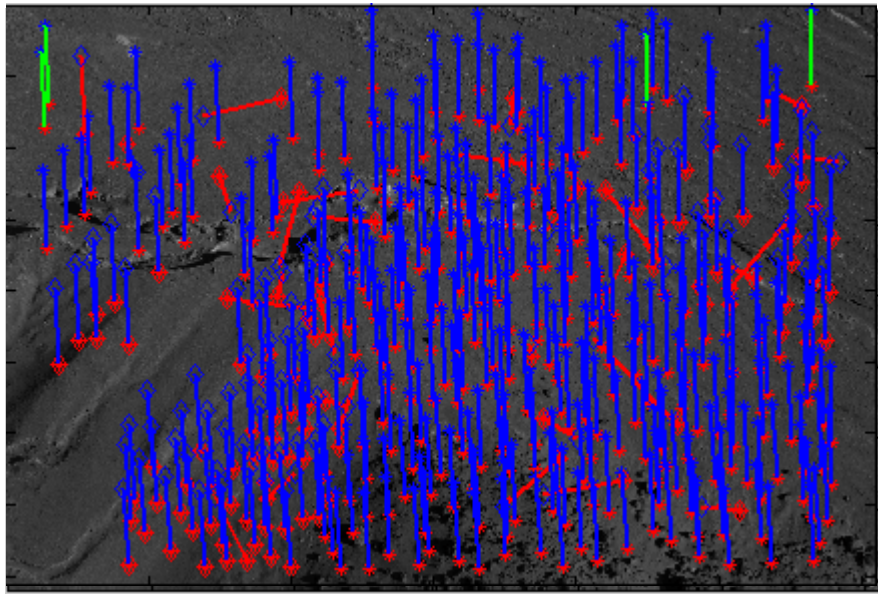
- Where P is the 3 x n whose columns consist of $\{p_i\}$ and P^{+} is the Moore-Penrose pseudoinverse.

- Now just work with $\{q_i\} = \{Lp_i\}$

Homography Example Revisited





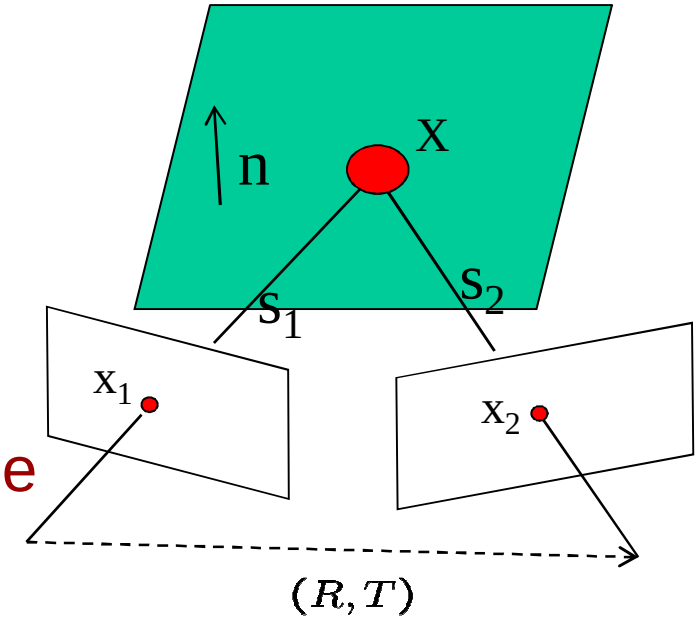


Homography based outlier rejection

Homography Revisited

- Recall that a homography relates a plane in space to the two images.
- Equation of a plane:

$$n^T X = D$$



- $X = s_1 x_1$ in left camera frame
- $X = s_2 x_2$ in right camera frame
- A little algebra shows:

$$s_2 x_2 = (R + \frac{1}{D} T n^T) s_1 x_1 \longrightarrow H = (R + \frac{1}{D} T n^T)$$

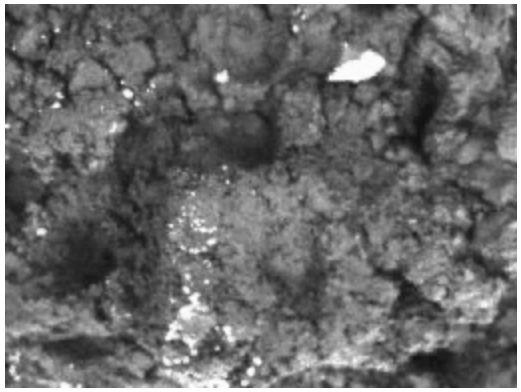
- Decomposition of H into (R,T) is more messy algebra. (Longuett-Higgins)

Pose Estimation

- Suppose you have a calibrated camera with known 3D to 2D point correspondences. How do you recover the camera extrinsics?
- Same technique used for calibration (DLT)
  - This is not very robust -> ignores the scene geometry and uses an algebraic constraint
  - Could try RANSAC type approach, but that only handles outliers, not noise
  - Try to incorporate scene geometry.
- Typical pseudo-linear approaches write projection equations in terms of 3D knowledge (Quan, etc.)
- Leasts squares optimization works (Can be made very efficient, e.g. Lu, Hager and Mjolsness)

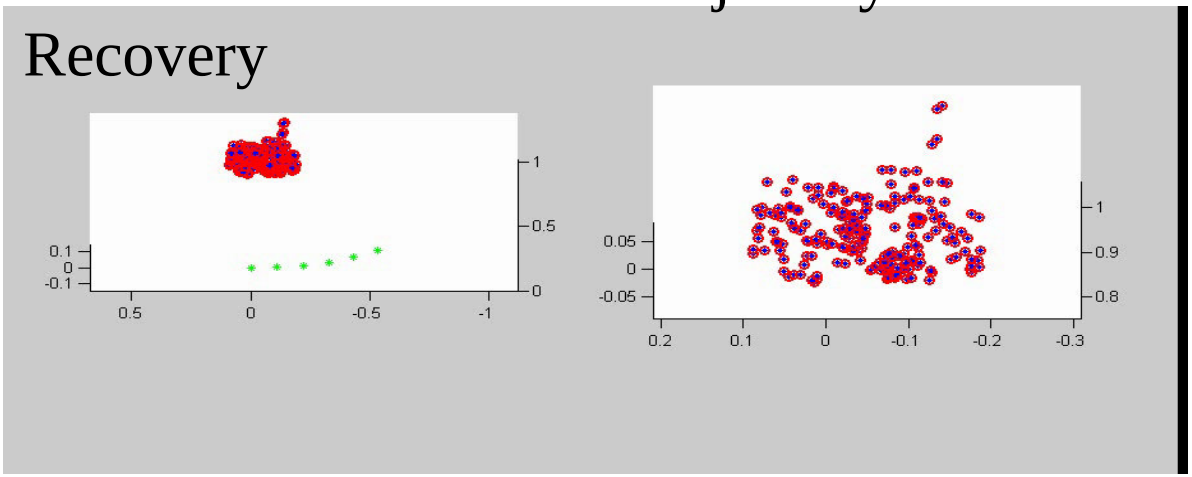
Bundle Adjustment on Small Body

- Catalog generation
  - Initial orbits used to build catalog of landmarks with associated descriptors
  - Bundle adjustment techniques + partial state information (attitude from star tracker, distance from altimeter, etc.) used to find 3D positions of landmarks in frame of small body



Test imagery

3D Reconstruction and Trajectory Recovery



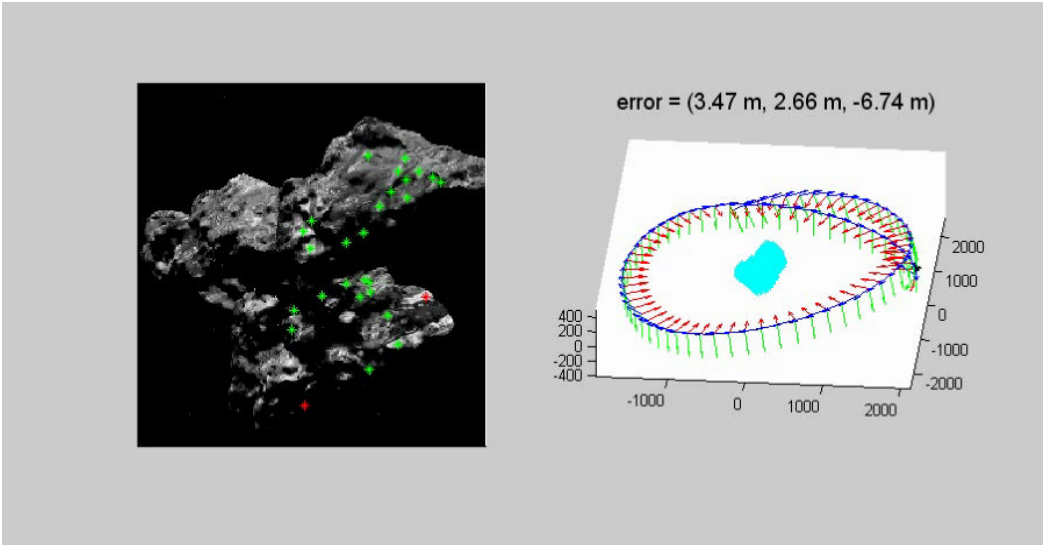
Blue points = all 3D points recovered

Red points = 3D points recovered/refined in current image

Green points = reconstructed camera positions

Pose Estimation on Small Body

- Localization
  - Landmarks are identified during later orbits and matched to catalog
  - Bearing angles of landmarks from imagery and 3D coordinates from catalog are supplied to onboard estimator
  - Robust vision based solution of 6 DoF pose (position and attitude) from matched landmarks used as sanity check on estimator



Landmark Detection and vision based localization

- Synthetic imagery of small body (~500 m radius) from 2 km orbit with known trajectory
- Vision based localization error ~10 m. Estimator will do even better.

Left pane: Imagery with detected landmarks shown. Red = rejected as outlier. Green = accepted

Right pane: Recovered position overlayed on ground truth trajectory. Vision based position errors shown for each frame.

Mathematical Digression

- Suppose you have two sets of 3D points $\{P_i\}$ and $\{Q_i\}$ related by

$$Q_i = RP_i + T$$

- How do you find (R,T)?
- Problem is referred to as exterior orientation (BKP Horn)
  - Subtract mean from $\{P_i\}$ and $\{Q_i\}$
  - Solve orthogonal Procrustes problem to find R
  - Find T in terms of means of $\{P_i\}$ and $\{Q_i\}$

Mathematical Digression

- Let

$$\overline{P} = \text{mean}(\{P_i\}) \quad \overline{Q} = \text{mean}(\{Q_i\})$$

$$P'_i = P_i - \overline{P} \quad Q'_i = Q_i - \overline{Q}$$

- View $\{P'_i\}$ and $\{Q'_i\}$ as column vectors and stack them into 3 x n matrices P' and Q'.

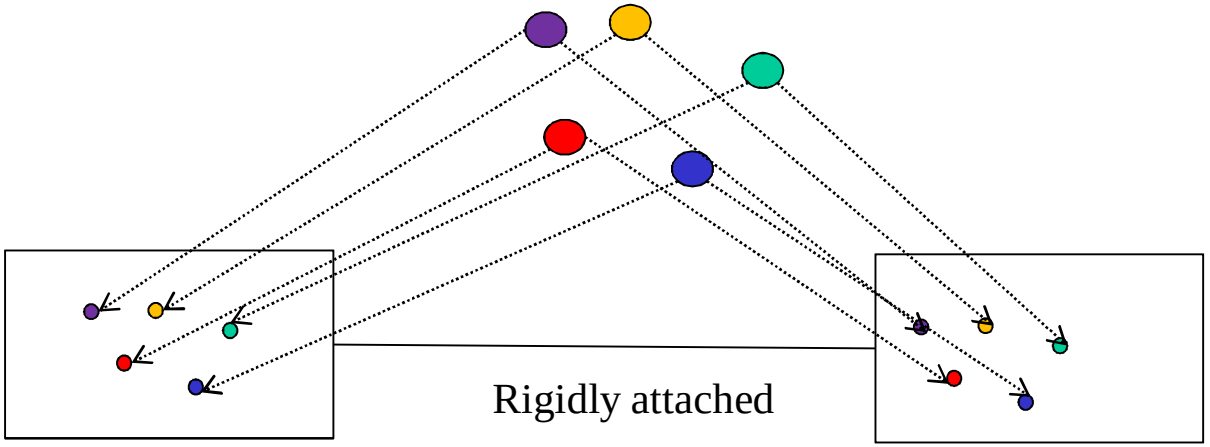
$$\widehat{R} = Q' P'^T = USV^T$$

$$R = UV^T$$

$$T = \overline{Q} - R\overline{P}$$

Visual Odometry using Stereo

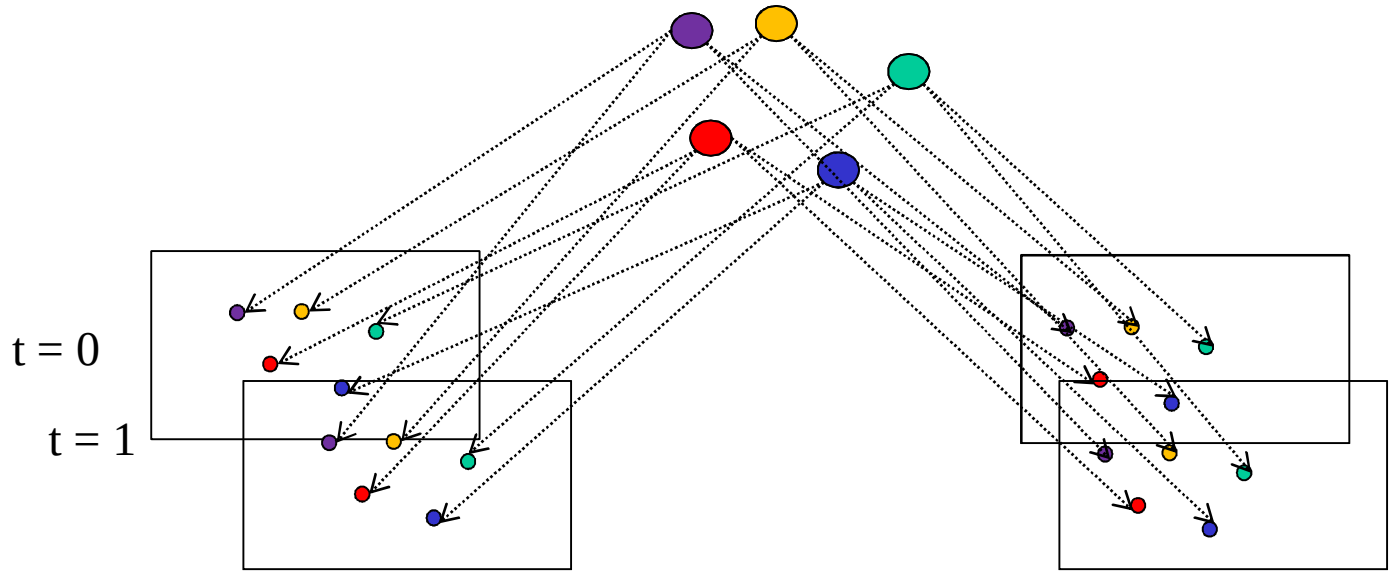
- Given a calibrated stereo pair, you can measure vehicle ego-motion



- Triangulation from Stereo gives coordinates of world points in camera reference frame.

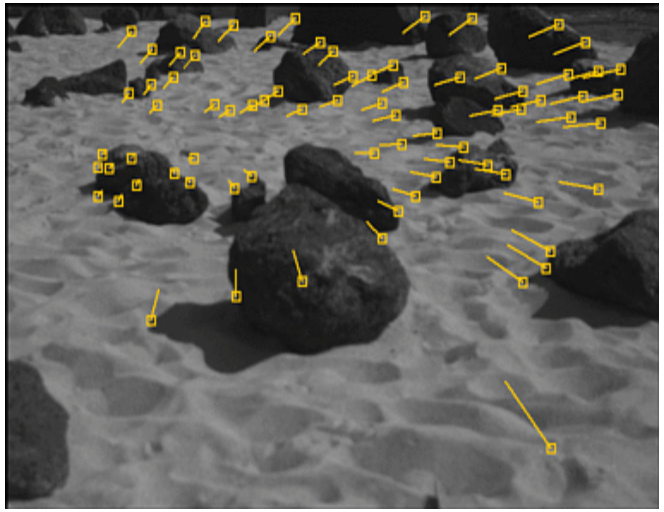
Visual Odometry using Stereo

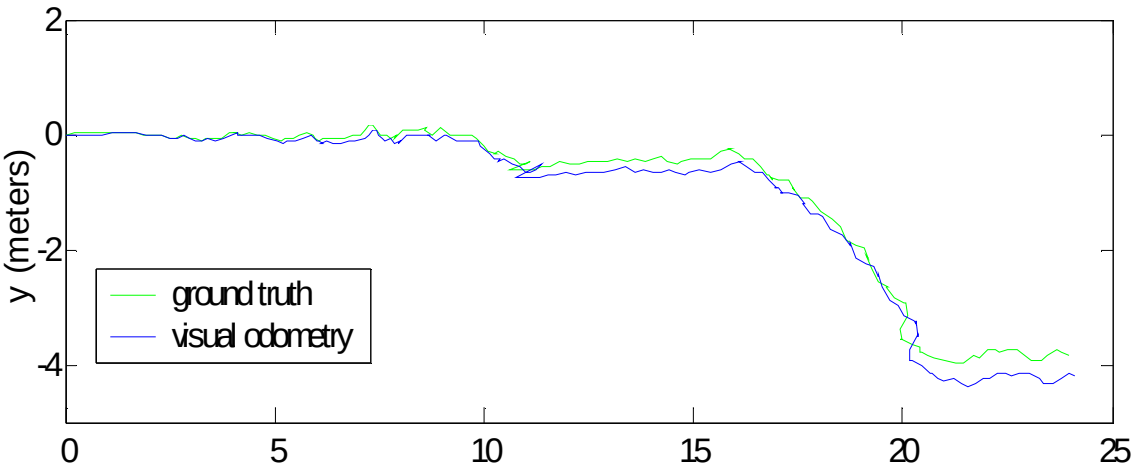
- Given a calibrated stereo pair, you can measure vehicle ego-motion



- Get solution for 3D points at timestep 0 and timestep 1.
- Solving for Euclidean motion between points equivalent to solving for (e.g. left) camera motion.

Visual Odometry in Practice





Status

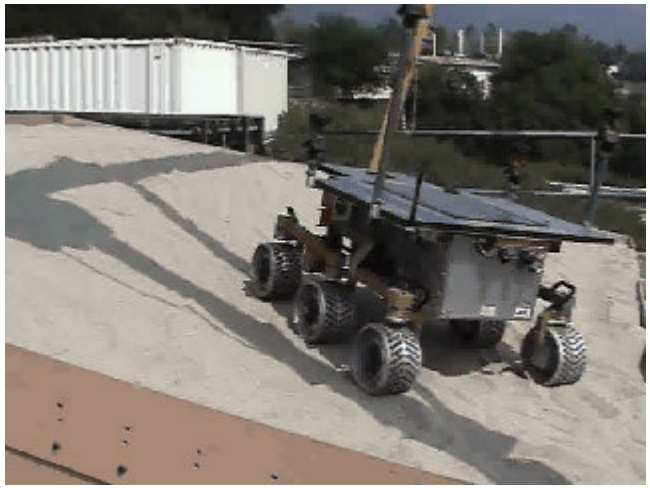
Used extensively on MER

256x256 imagery processed in ~ 3 minutes on RAD6000 MER flight processor

Baselined for MSL; needs to run faster (eg. 2-6x). Approach: either RAD750 or Mobility Avionics Module

Visual Odometry for Slip Estimation and Control

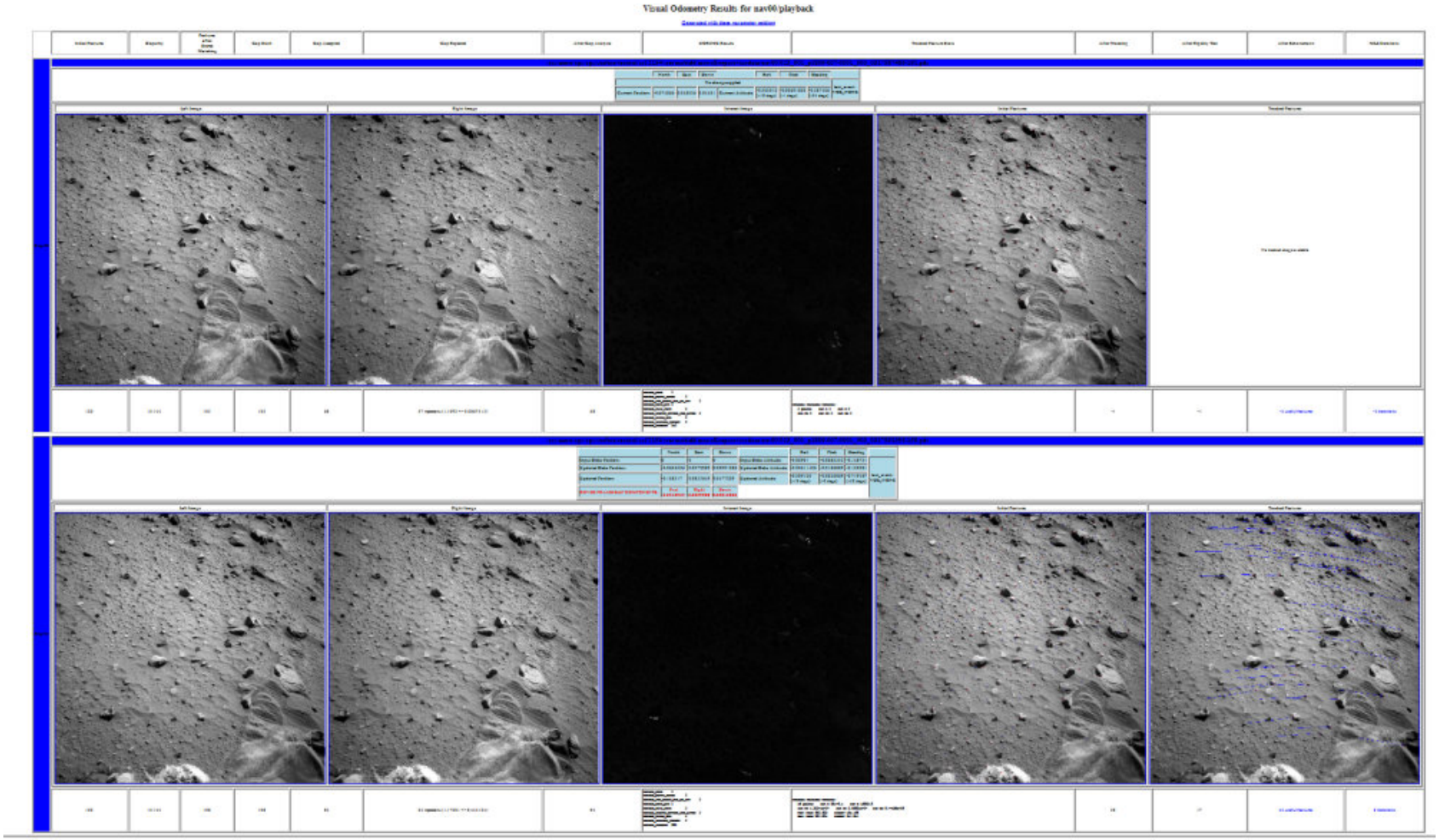
"Open loop" driving using
wheel odometry



Closed loop driving using
visual odometry

Visual Odometry on Mars

- ..\..\..\..\tmp\homog\paolo\index.html



Homework

- Read Szeliski 7.2.
- Problems (due 2/8/11):

1. Implement the 8-point algorithm.
2. For the included dataset from Mars Hill, plot the expression $\varepsilon = p_2^T F p_1$ over all point pairs. Can you easily pick out the outliers?
3. Extra credit: Implement RANSAC using the epipolar constraint instead of homographies. Show that the former has more inliers on the Mars Hill dataset than the latter.

References

- R. Hartley, A Zisserman: Multi View Geometry. Cambridge University Press, 2000.
- O. Faugeras: Three-Dimensional Computer Vision. MIT press, 1993.
- H. Longuet-Higgins: The reconstruction of a plane surface from two perspective projections. In: *Proc. Roy. Soc. London Series B 227 (1249)*, pp. 399–410, 1986.
- H. Longuet-Higgins: A computer algorithm for reconstructing a scene from two projections. *Nature (293)*, pp. 133–135, 1981.
- L. Quan and Z. Lan: Linear N-point camera pose determination. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 21(8):774--780, August 1999.
- BKP. Horn: Recovering Baseline and Orientation from the Essential Matrix.
- B. K. P. Horn: Closed-form solution of absolute orientation using unit orthonormal matrices, *Journal of the Optical Society of America A*, 5(7):1127-1135, 1988.