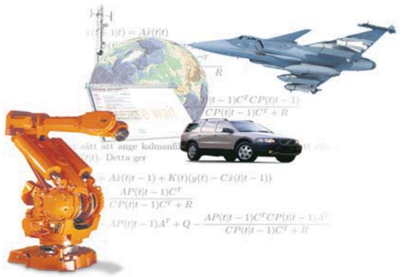
Lecture 4 – Epipolar Geometry and Reconstruction



Thomas Schön,
Division of Automatic Control,
Department of Electrical Engineering,
Linköping University.

Email: schon@isy.liu.se

Accurate noise modeling is just as important as accurate geometric (physical) modeling

Content – Lecture 4

1. Summary of Lecture 3
2. Epipolar geometry (= the geometry of 2 views)
3. Basic reconstruction
4. Reconstruction as an optimization problem (bundle adjustment, SLAM, SfM)
5. Planar homography

Summary – Lecture 3 (Correspondence problem)

Given a point in the 3-D world the *correspondence problem* amounts to finding its 2-D projection in two different images.

That is, find pairs of pixels (one in each image) that corresponds to the same point in the 3-D world.

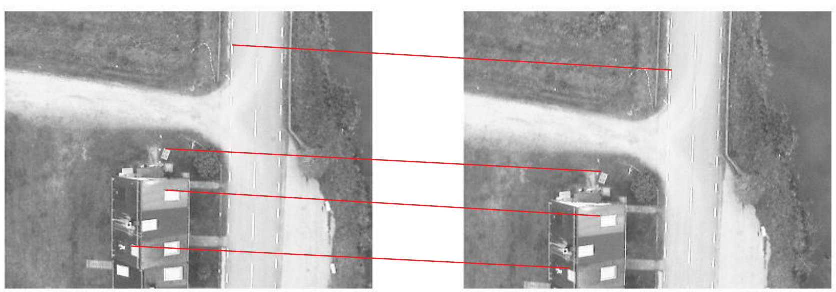
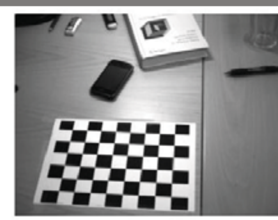


Image 1 Image 2

Summary – Lecture 3 (Harris corners)

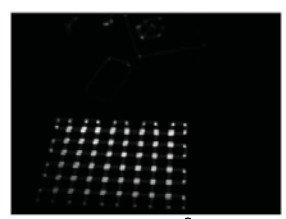


Original image

$$s(x_p, y_p) = \det(H(x_p, y_p)) + k\text{tr}^2(H(x_p, y_p)),$$

$$H(x_p, y_p) = \sum_{x=-N/2}^{N/2} \sum_{y=-N/2}^{N/2} \begin{pmatrix} I_1 & I_{12} \\ I_{12} & I_2 \end{pmatrix} \in \mathbb{R}^{2\times 2}$$

$$\begin{aligned} I_1 &= I_x^2(x_p + x, y_p + y), \\ I_{12} &= I_x(x_p + x, y_p + y) I_y(x_p + x, y_p + y) \\ I_2 &= I_y^2(x_p + x, y_p + y). \end{aligned}$$



$s = \det(H) + k\text{tr}^2(H)$

Summary – Lecture 3 (A first estimation problem)

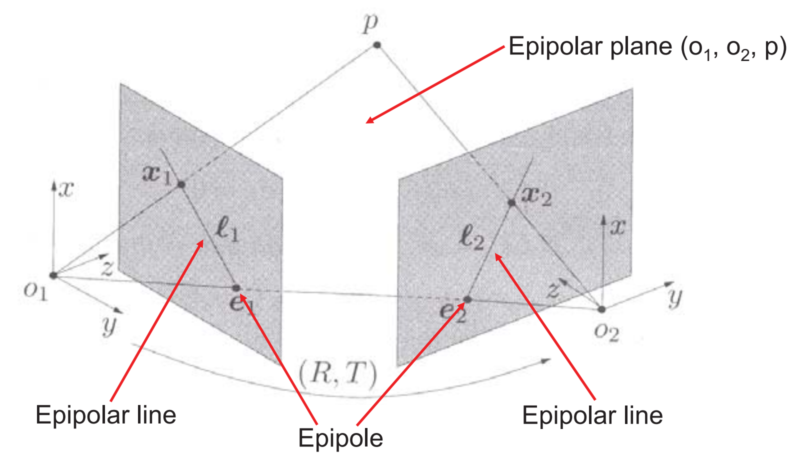
Visual odometry is the process of estimating the pose of a camera using information from the image.

Conceptual solution:

1. Initialization
2. Acquire an image and correct for possible lens distortion
3. Predict the positions of the previously detected map entries
4. Data association, match the predicted interest points to the corresponding descriptors obtained in the new image
5. Remove outliers
6. Search for new interest points in areas where there are no interest points
7. Update the state estimates
8. Repeat from 2

(Home work 2 studied step 4 and 6)

Epipolar Geometry (the geometry of two views)



The eight-point algorithm

Longuet-Higgins, H. C. **A computer algorithm for reconstructing a scene from two projections,** *Nature, 293:133-135, 1981*.

For a given set of correspondences $(x_1^j, x_2^j), j = 1, \dots, n (n \geq 8)$ the eight-point algorithm recovers $(R, T) \in SE(3)$ which satisfy the epipolar constraint,

$$(x_2^j)^T \hat{T} R x_1^j = 0, \qquad j = 1, \dots, n$$

1. Compute a first approximation of the essential matrix

Form $\chi = \begin{pmatrix} a^1 & a^2 & \dots & a^n \end{pmatrix}^T \in \mathbb{R}^{n \times 9}$ from correspondences x_1^j and x_2^j according to

$$a^j = x_1^j \otimes x_2^j \in \mathbb{R}^9$$

Find the vector $E^s \in \mathbb{R}^9$ of unit length such that $\|\chi E^s\|$ is minimized as follows: compute the SVD of $\chi = U_\chi \Sigma_\chi V_\chi^T$ and define E^s to be the ninth column of V_χ. Typically, $E^s \notin \varepsilon$

The eight-point algorithm

2. Project onto the essential space

Compute the singular value decomposition of the matrix E

$$E = U \text{diag}(\sigma_1, \sigma_2, \sigma_3) V^T$$

where $\sigma_1 \geq \sigma_2 \geq \sigma_3 \geq 0$ and $U, V \in SO(3)$ In general, since E may not be an essential matrix, $\sigma_1 \neq \sigma_2$ and $\sigma_3 \neq 0$. However, its projection onto the normalized essential space is $U \Sigma V^T$, where $\Sigma = \text{diag}(1, 1, 0)$.

3. Recover the motion from the essential matrix

We only need U and V in order to extract R and T from the essential matrix E

$$R = U R_Z^T \left(\pm \frac{\pi}{2}\right) V^T, \qquad \hat{T} = U R_Z \left(\pm \frac{\pi}{2}\right) U^T.$$

Comments on the eight-point algorithm

- The epipolar constraint is valid both for E and $-E$. Hence, there are 4 possible solutions for (R, T). However, only one of these solution is physically possible. The others give rise to negative depths of the reconstructed points.
- The eight points have to be in a “general position”, that is we have to excite the problem.
- Recall that the solution is only known up to an unknown scale factor.

Problems with the eight-point algorithm

- There is no guarantee that the estimated pose is as close as possible to the true solution.
- Since the correspondences are noisy, the reconstruction might very well be inconsisten.

A Dangerous Word – Optimal

It is important to understand that an optimal solution is just a solution to an optimization problem. It does **NOT** reveal whether the solution is good or bad!!

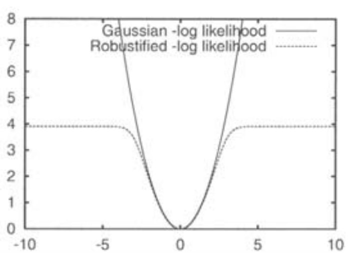
“Unfortunately, there is no “correct”, uncontroversial, universally accepted objective function, and the choice of discrepancy measure is part of the design process, since it depends on what assumptions are made on the residuals w_i^j. Different assumptions result in different choices of discrepancy measures, which eventually result in different “optimal” solutions.”

Choosing the cost function

If outliers exist, non-robust models such as Gaussians should NOT be used. Better to use Cauchy or other robust densities.

The ML framework is naturally robust, there is no need to “robustify” it as long as realistic error distributions are used. Hence, **accurate noise modeling** is just as important as accurate geometric modeling.

Using the language of optimization this corresponds to add regulatization terms and looking at robust methods.



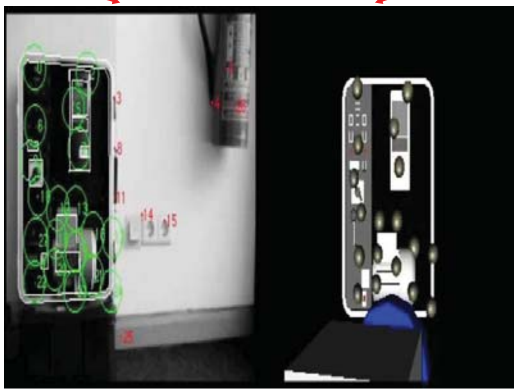
A good introduction to bundle adjustment is provided in,

Triggs, B., McLauchlan, P., Hartley, R. and Fitzgibbon, A. **Bundle adjustment – a modern synthesis,** *Proceedings of the International Workshop on Vision Algorithms: Theory and Practice, 1999,* Corfu, Greece.

Reconstruction (Visual SLAM) example (I/II)

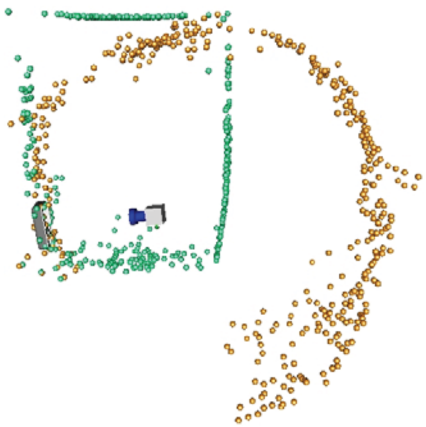
Feature tracking overlayed on the image (green – tracked, red – not tracked)

3D view of the reconstruction, with covariance ellipsoids for the features



Courtesy of Gabriele Bleser

Reconstruction (Visual SLAM) example (II/II)



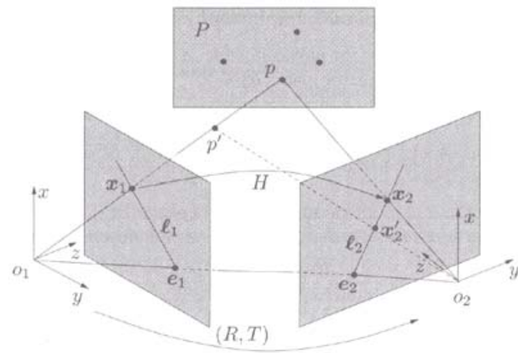
Orange – Reconstruction problem solved using simple triangulation.

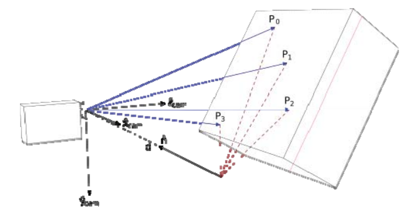
Green – The extended Kalman filter is used to refine the 3D points and some feature quality control is used.

Courtesy of Gabriele Bleser

Homography

A **homography** is an invertible transformation from a projective plane to a projective plane that maps straight lines to straight lines.





Homography

x_1, x_2 - 2 images of a point $p \in P$

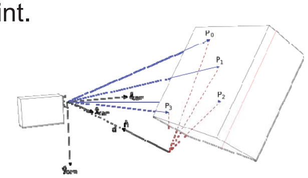
X_1, X_2 - coordinates of p relative to camera 1 and 2, respectively

$$X_2 = RX_1 + T$$

The epipolar constraint implies $x_2^T \hat{T} R x_1 = 0$

Points on the same plane P share an **additional** constraint.

$n = \begin{pmatrix} n_1 \\ n_2 \\ n_3 \end{pmatrix}$ - Unit normal of the plane P w.r.t. camera 1



d - Distance from P to camera 1 (o_1)

$$n^T X_1 = d \qquad \frac{1}{d} n^T X_1 = 1, \qquad \forall X_1 \in P$$

$$X_2 = RX_1 + T\frac{1}{d}n^T X_1 = \underbrace{(R + \frac{1}{d} T n^T)}_{H} X_1 \qquad X_2 = HX_1$$