Monocular Visual Odometry

Christian Forster

Robotics and Perception Group
Institute for Informatics
University of Zürich

December 9, 2013

Monocular Visual Odometry

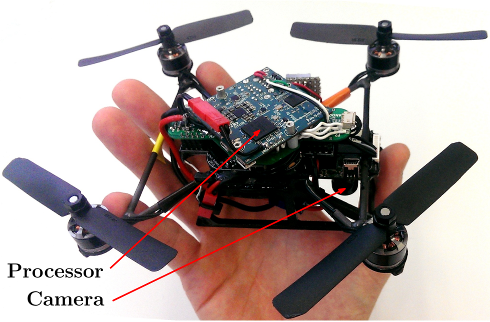
Goal

Estimate the egomotion of the Micro Aerial Vehicle (MAV) using only a single camera.

- Lightweight
- Low power consumption
- Moving platform: Stereo vision not necessary

"Visual Odometry"

Incrementally estimate the pose of the vehicle by examining the changes that motion induces on the images.



Visual Odometry

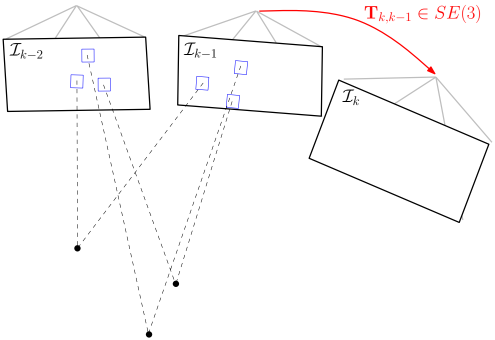
Problem Formulation

Two camera poses at adjacent time instants $k-1$ and k are related by the rigid body transformation

$$\mathbf{T}_{k,k-1} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ 0 & 0 \end{bmatrix} \in SE(3).$$

Concatenation of the relative transformations allows to recover the path:

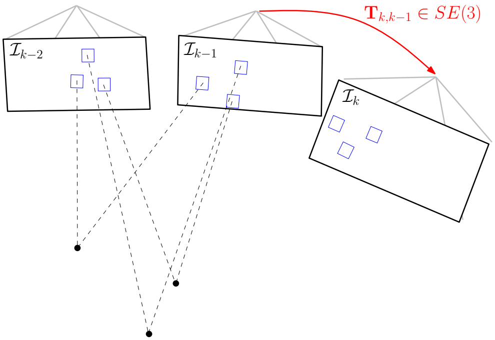
$$\mathbf{T}_{k,0} = \mathbf{T}_{k,k-1} \cdot \mathbf{T}_{k-1,k-2} \cdots \mathbf{T}_{1,0}$$



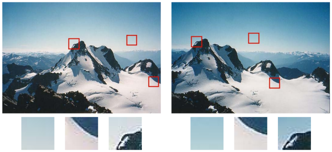
Feature-based Visual Odometry

Pipeline

1. Feature selection
2. Feature matching
3. Pose estimation
4. Pose refinement
5. Triangulation

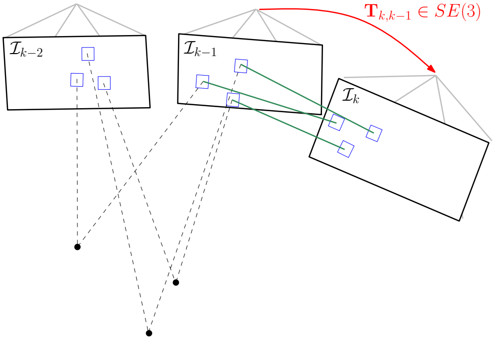


Which features?



Source: Szeliski, "Computer Vision: Algorithms and Applications", Springer 2010.

Feature-based Visual Odometry



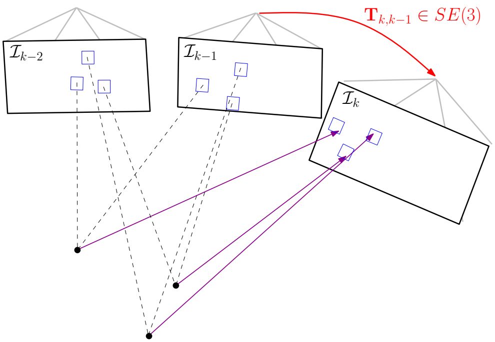
Pipeline

1. Feature selection
2. Feature matching
3. Pose estimation
4. Pose refinement
5. Triangulation

Matching Strategies

- Fast: SSD/NCC over small patch
- Robust: Match invariant feature descriptors, e.g., SIFT [Lowe, 2003]

Feature-based Visual Odometry



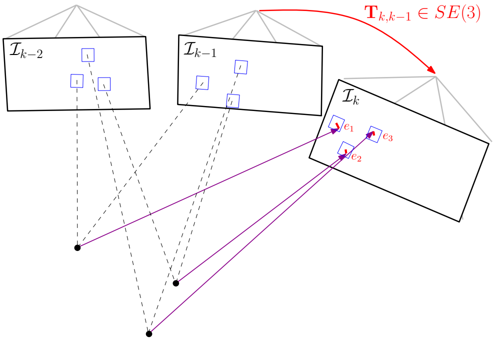
Pipeline

1. Feature selection
2. Feature matching
3. Pose estimation
4. Pose refinement
5. Triangulation

Epipolar Geometry

Three 3D point to 2D feature correspondences are necessary to estimate the 3D camera pose. [Kneip et al., 2011]

Feature-based Visual Odometry



Pipeline

1. Feature selection
2. Feature matching
3. Pose estimation
4. Pose refinement
5. Triangulation

Minimize reprojection errors

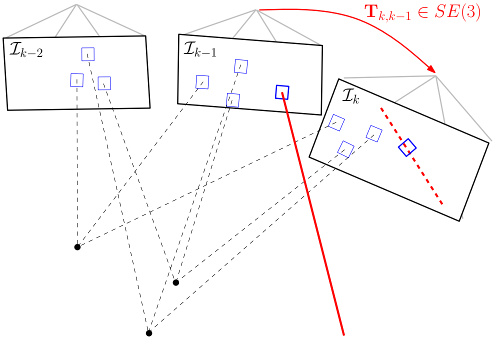
$$\mathbf{T}_{k,k-1} = \arg\min_{\mathbf{T}} \sum_i \rho\left[e_i(\mathbf{T})\right], \qquad \text{typically} \quad \rho[.] \hat{=} \frac{1}{2\sigma^2} \parallel . \parallel^2$$

Can be solved with e.g. Gauss Newton.

Feature-based Visual Odometry

Pipeline

1. Feature selection
2. Feature matching
3. Pose estimation
4. Pose refinement
5. Triangulation



Triangulation

Search along Epipolar line for matching feature.

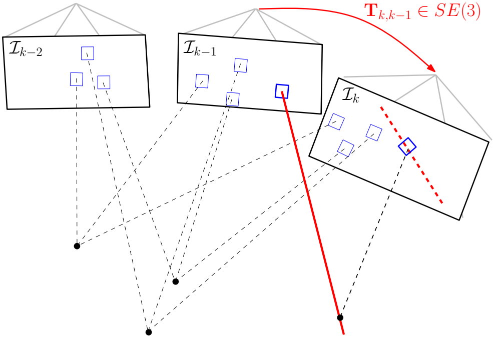
Feature-based Visual Odometry

Pipeline

1. Feature selection
2. Feature matching
3. Pose estimation
4. Pose refinement
5. Triangulation

Triangulation

Search along Epipolar line for matching feature.



Implementation Details

- Make more robust by using many (hundreds) of features.
- Use motion model to speed-up feature matching.
- Use robust estimation techniques to handle wrong matches (e.g., RANSAC [Fischler and Bolles, 1981]).
- Minimize drift through incremental Bundle Adjustment: Joint optimization of frames and 3D points [Mouragnon et al., 2006] .
- Parallelize tracking and mapping [Klein and Murray, 2009] .

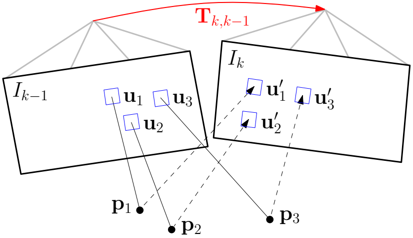
The direct approach

Minimize photometric error

$$\mathbf{T}_{k,k-1} = \arg\min_{\mathbf{T}} \iint_{\bar{\mathcal{R}}} \rho\left[\delta\mathcal{I}(\mathbf{T},\mathbf{u})\right] d\mathbf{u}.$$

where

$$\begin{aligned}\delta\mathcal{I}(\mathbf{T},\mathbf{u}) &= \mathcal{I}_k(\mathbf{u}') - \mathcal{I}_{k-1}(\mathbf{u}) \\ \mathbf{u}' &= \pi\left(\mathbf{T}\cdot\pi^{-1}(\mathbf{u}, z_{\mathbf{u}})\right)\end{aligned}$$



Advantages

- No costly feature extraction
- No costly robust feature matching

Bibliography

Fischler, M. A. and Bolles, R. C. (1981). Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Communications of the ACM*, 24(6):381–395.

Klein, G. and Murray, D. (2009). Parallel Tracking and Mapping on a Camera Phone. *IEEE Int. Symposium on Mixed and Augmented Reality*, pages 83–86.

Kneip, L., Scaramuzza, D., and Siegwart, R. (2011). A Novel Parametrization of the Perspective-Three-Point Problem for a Direct Computation of Absolute Camera Position and Orientation. *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*.

Lowe, D. G. (2003). Distinctive image features from scale-invariant keypoints.

Mouragnon, E., Lhuillier, M., Dhome, M., Dekeyser, F., and Sayd, P. (2006). 3D Reconstruction of Complex Structures with Bundle Adjustment: an Incremental Approach. *Proceedings of the 2006 IEEE International Conference on Robotics and Automation*, pages 3055–3061.