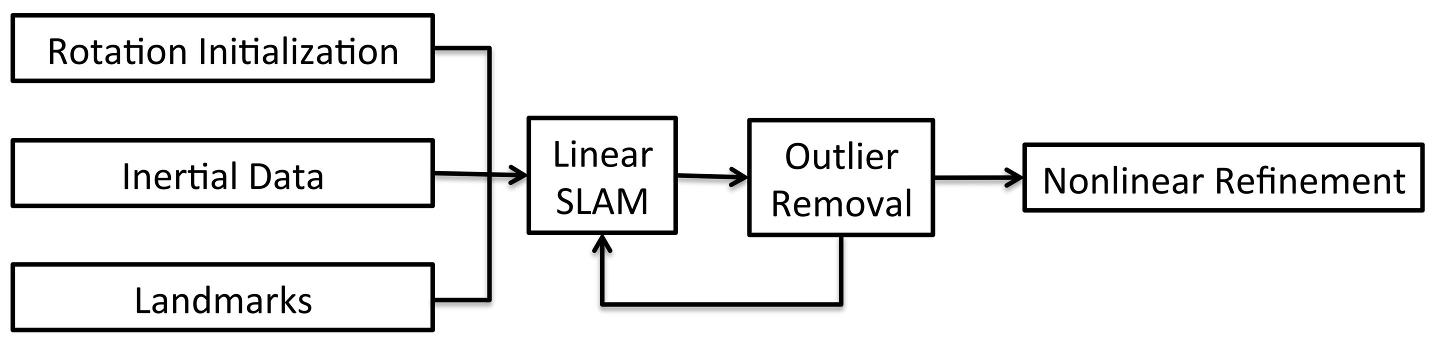
Linear Methods for Visual/Inertial SLAM

Zoran Sjanic*†, Martin Skoglund*, Thomas B. Schön* and Fredrik Gustafsson*
*Linköping University †Saab Aeronautics

Main idea

Structure from Motion and Bundle Adjustment are well established imaging techniques for reconstruction of 3D structures and camera motion simultaneously. In the case of monocular vision and inertial measurements many of these parameters can be estimated fairly well using linear methods. Also, the linear methods can efficiently be used for iterative outlier removal. Combined, these constitute a good initial guess for the total SLAM problem.



Introduction

In this work we present a solution to the simultaneous localisation and mapping (SLAM) problem using a camera and inertial sensors. Our approach is based on a clustering of feature tracks which are fused with inertial data in a linear estimator. The linear solution is combined with an outlier rejection step and requires only a few iterations to converge while removing most outliers. The final step takes the linear solution as a starting point in a standard nonlinear least-squares solver which efficiently solve the total SLAM, including additive sensor biases and global scale. This approach is evaluated using both real and simulated data.

Problem Formulation

Contrary to standard SLAM formulations we consider a parameter estimation problem using camera and inertial measurements on the form

$$\arg\min_{\Theta} \sum_t ||y_t - h(x_t, \Theta)||^2_{R^{-1}},$$
$$\text{subject to } x_t = f(x_{t-1}, \Theta),$$

where the parameter vector $\Theta = [\mathsf{M},\ \mathsf{B},\ v_0,\ \mathsf{A},\ \Omega]$, consist of landmark coordinates M, IMU biases B, initial velocity v_0, and navigation frame accelerations A and angular rates Ω, respectively. The pose and velocity, $x = [p, v, q]^T$, are considered nuisance variables. This estimation problem is nonlinear and a good initial value for Θ is needed. We will show how this can be done efficiently exploiting the inherent linear structures.

Linear Models

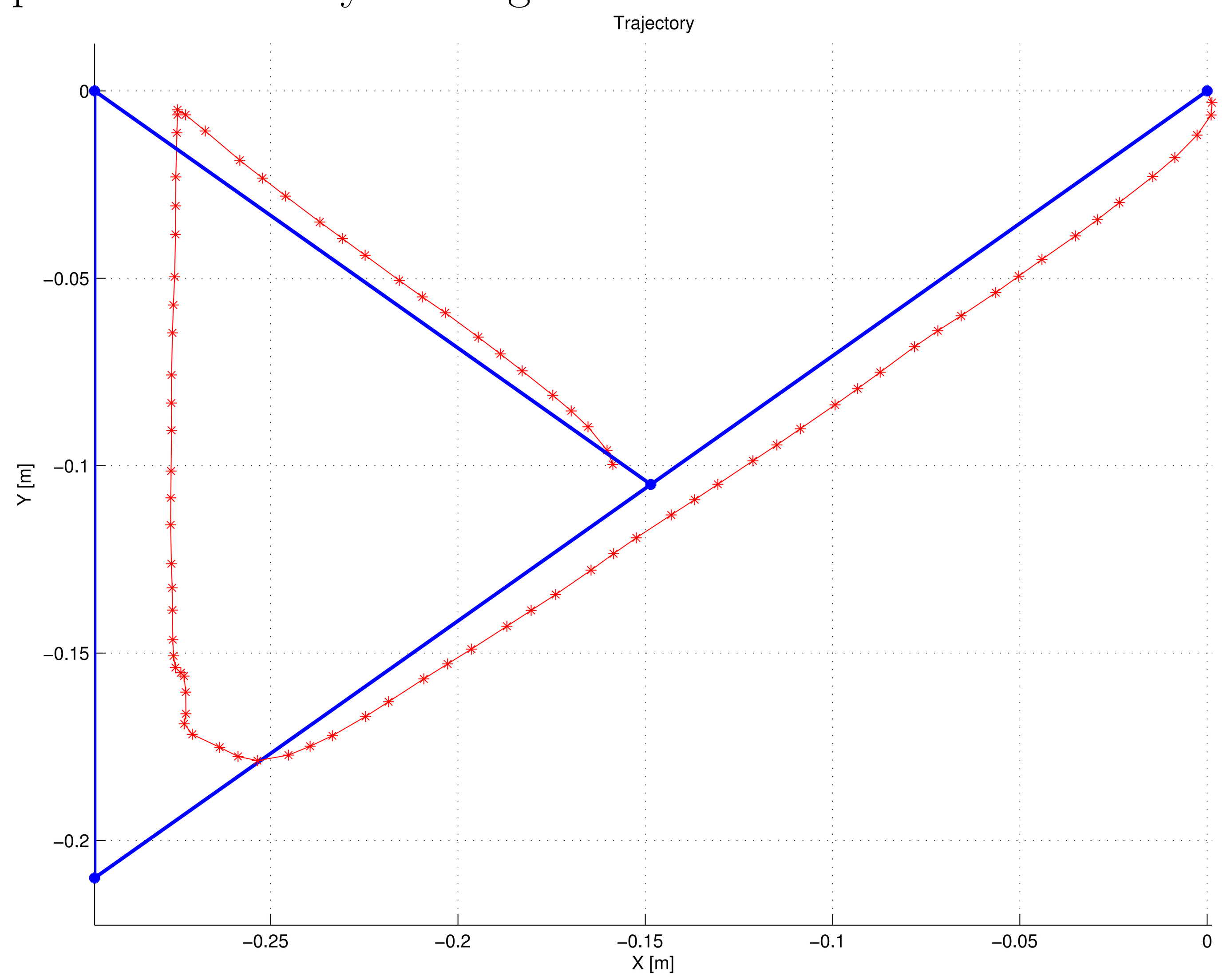
A calibrated perspective camera act as mapping $P([X, Y, Z]) = [X/Z, Y/Z]$ and normalized camera measurement $y_t^i = [u_t^i, v_t^i]^T$ of landmark i at time t is then

$$y_t^i = P(\underbrace{R_t^{cw}(m_t^i - p_t)}_{[X^c, Y^c, Z^c]^T}) + e_t \tag{1}$$

which relates the absolute pose of the camera to the 3D location of the point. Note that the pose is a function of Θ. From (1), a linear system with parameter dependent noise is given by

$$\begin{bmatrix} -1 & 0 & u \\ 0 & -1 & v \end{bmatrix} \begin{bmatrix} X^c \\ Y^c \\ Z^c \end{bmatrix} = Z^c \begin{bmatrix} e^u \\ e^v \end{bmatrix}. \tag{2}$$

If if two or more measurements are available (2) can be iteratively solved by initializing a $\bar{Z}^c$ for the right hand side and solve the resulting linear least squares problem and then re-weight the problem with the new Z^{c*}. This procedure usually converges after a few iterations.



The constraints can be removed since the translational part of the dynamics is linear

$$x_t = F x_{t-1} + G\,\mathsf{a}_t$$

then, for an arbitrary time instance

$$x_t = F^t x_0 + \sum_{k=0}^{t-1} F^k G\,\mathsf{a}_{k+1} = \mathsf{D}\,\Theta$$

which is a linear system in Θ. This assumes that the rotations, q, are solved, which can be done using the e.g., the 8-point algorithm.

Iterative Outlier Removal

Feature tracks contain outliers which are defined in a reprojection error sense. A simple strategy, similar to L_∞ outlier removal, is used. It is assumed that the errors can be detected from the linear solution alone. By successively removing landmarks with the largest reprojection errors within an image and then update the linear solution. This metod converges when all errors are of similar size.

Nonlinear Refinement

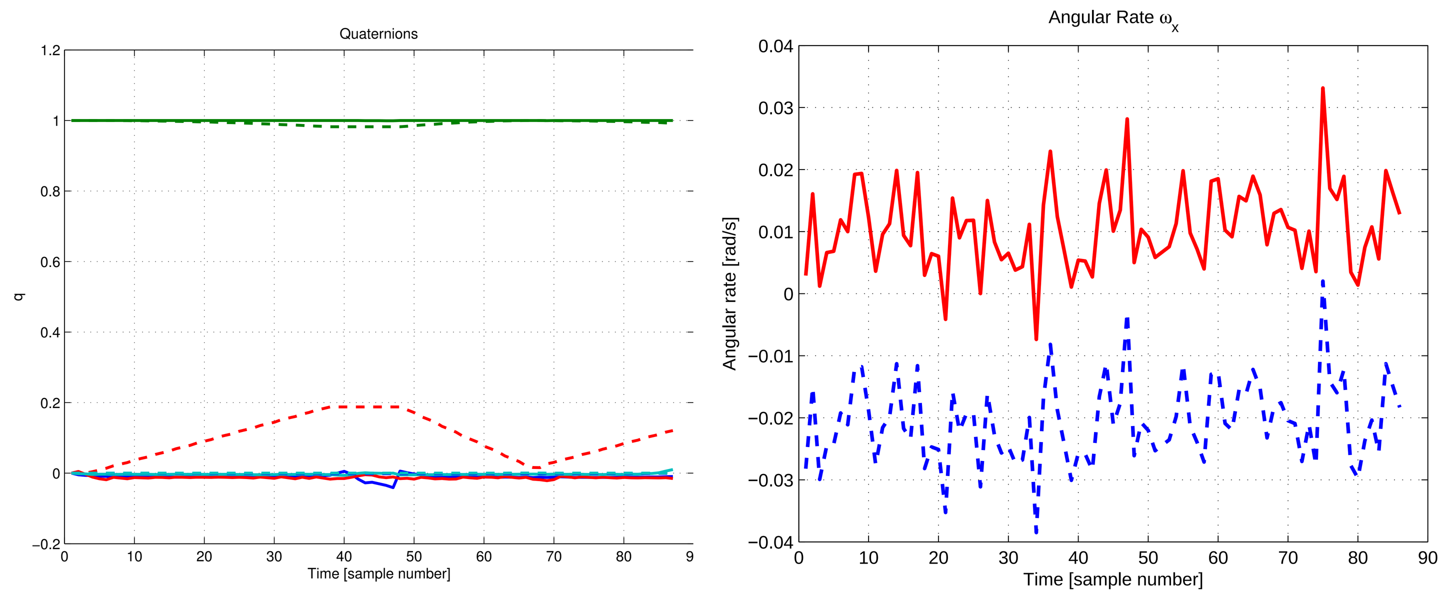
As a final step, a nonlinear least-squares solver is used to optimize the unconstrained problem

$$\arg\min_{\Theta} \sum_t ||y_t - h(\Theta)||^2_{R^{-1}},$$

which also accounts for sensor biases, the initial velocity and the rotation.

Results

The left image shows the camera-only initial rotation (dashed) and the final estimated rotation (solid) which is very close to the truth. The right image shows the angular rate before (dashed) and after bias compensation (solid).



Data set	OL left	IL removed	OL landmarks	SLAM iterations
1	0	5.6%	11%	16
2	0	3.8%	21%	6
3	2	1.1%	8%	7
Real data	1	1.7%	8%	18

- Future work includes solving the problem using Expectation Maximization.
- Elaborate on joint estimation of the full SLAM problem including data association.