Initialisation and Estimation Methods for Batch Optimisation of Inertial/Visual SLAM

Martin Skoglund*, Zoran Sjanic* † and Fredrik Gustafsson*
*Linköping University †Saab Aeronautics

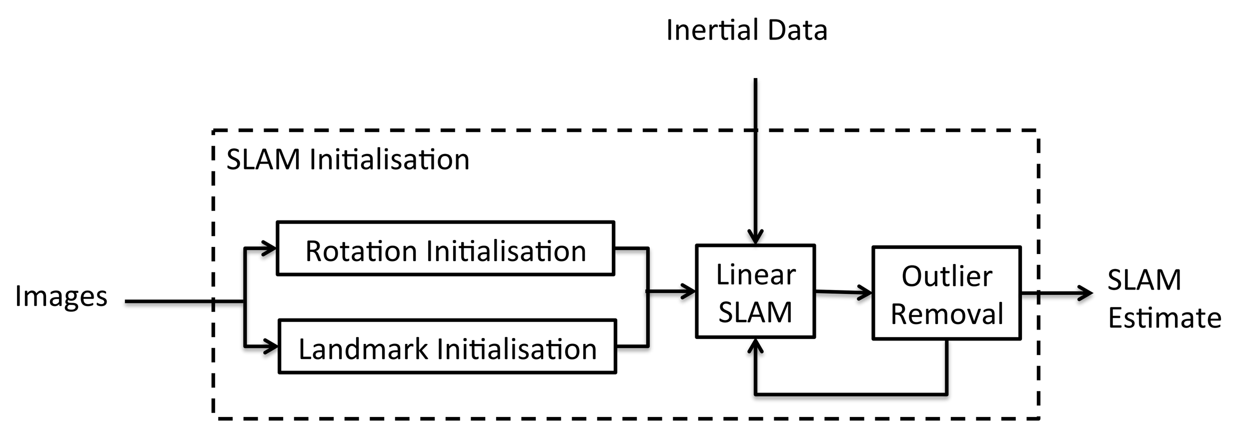
Problem description

Batch SLAM using a monocular camera and inertial sensors (IMUs) is a highly non-convex problem and requires a good initial estimate. Finding such an estimate is a non-trivial task due to the inherently different nature of the sensors.

For instance; Dead-reckoning of low grade IMU data is only good for a very short time. A potential solution is to initialise the system with an EKF. However, this is not a viable option for large problems. There is also a higher risk of getting stuck in local optima since the map and trajectory are too coupled by the inertial data.

Contribution

We propose a multi-stage iterative procedure that exploits the different characteristics of the sensors alone in such a way that outliers will be rather distinct when the estimates are combined. An important fact is that the SLAM problem is linear if the platform's rotations are known. Also, if feature correspondences are known these rotations can be estimated from images alone. Furthermore, by utilising feature tracking and clustering of feature tracks, loop closures are automatically detected.



Landmarks and Rotations

The landmark map and the rotation sequence of the platform are initialised using only the camera measurements. This is done as follows:

1. Feature tracks are found by solving a sequence small linear assignment problems using SIFT descriptors only.
2. Loop closures are computed by single-linkage clustering of the feature tracks.
3. From sets of correspondences defined by the feature tracks (1) rotations are computed using the well known Eight-point algorithm.
4. Initial landmarks are the ones that remain after feature track initialisation (1) and track clustering (2).

Linear SLAM

The projection is defined as $P([X,Y,Z]^T) = [X/Z, Y/Z]^T$ and a normalised camera measurement $y_t^m = [u_t, v_t]^T$ of a landmark, m, at time t is then

$$y_t^m = P(\mathsf{R}^{ce}(m - p_t)) + e_t \tag{1}$$

which relates the absolute pose of the camera w.r.t. the 3D location of the point. Denoting the argument of the projection by $\delta(p, m) = [\delta_x,\ \delta_y,\ \delta_z]^T$ which is the difference between landmark and camera position expressed in the camera coordinate system i.e., $P(\mathsf{R}^{ce}(m - p)) = P(\delta)$. We then have

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \frac{\delta_x}{\delta_z} \\ \frac{\delta_y}{\delta_z} \end{bmatrix} + \begin{bmatrix} e_u \\ e_v \end{bmatrix} \Rightarrow \qquad \begin{bmatrix} u\delta_z \\ v\delta_z \end{bmatrix} = \begin{bmatrix} \delta_x \\ \delta_y \end{bmatrix} + \begin{bmatrix} e_u\delta_z \\ e_v\delta_z \end{bmatrix} \tag{2}$$

which is linear in the unknown parameters m and p since R is assumed known, but has noise that is dependent on the parameters. With δ explicit (2) becomes

$$\mathsf{R}_{3,:}(m - p) \begin{bmatrix} u - e_u \\ v - e_v \end{bmatrix} = \begin{bmatrix} \mathsf{R}_{1,:}(m - p) \\ \mathsf{R}_{2,:}(m - p) \end{bmatrix}, \tag{3}$$

where $\mathsf{R}_{i,:}$ denotes the i:th row of the rotation matrix R^{ce}.

Since the rotations are assumed known, the gyro measurements are not included here. Thus, the position, p_t, of the platform can be expressed as a linear function of the initial velocity v_0 and the accelerations $a_{1:t}$. The following unconstrained problem is proposed

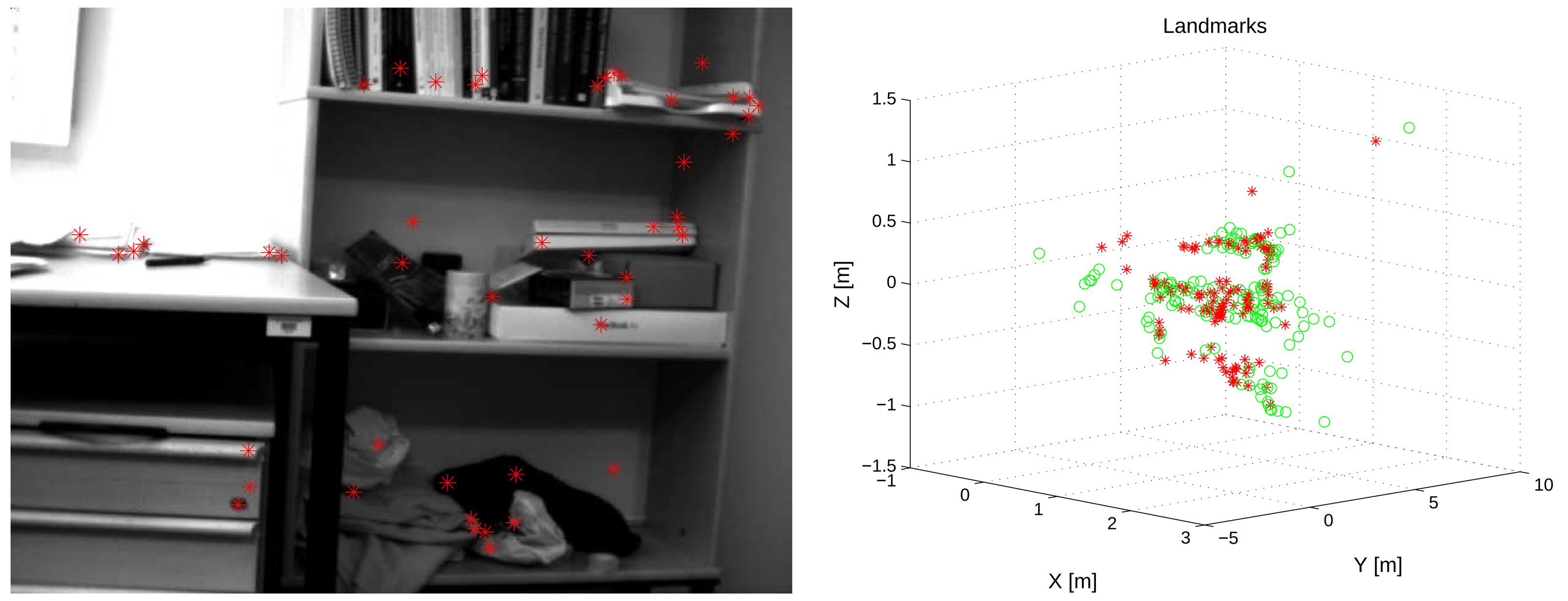
$$\theta_{\text{init}} = \arg\min_{\theta} \sum_{t=1}^{N} \|y_t^a - \mathsf{R}_t^{ce}(a_t - g^e) - b_a\|^2_{R_a^{-1}} + \|y_t^m \delta_z(v_0, a_{1:t}, m) - \delta_{x,y}(v_0, a_{1:t}, m)\|^2_{\tilde{R}_m^{-1}}, \tag{4}$$

where, $\theta = [a_{1:N}, v_0, b_a, m]^T$ and $\tilde{R}_m = \delta_z(v_0, a_{1:t}, m)^2 R_m$. Since the landmark measurements noise depends on the parameters we have an Iterative Reweighted Least Squares (IRLS) formulation. This can be treated in an iterative fashion where δ_z from the previous iterate is used for weighting the noise covariance.

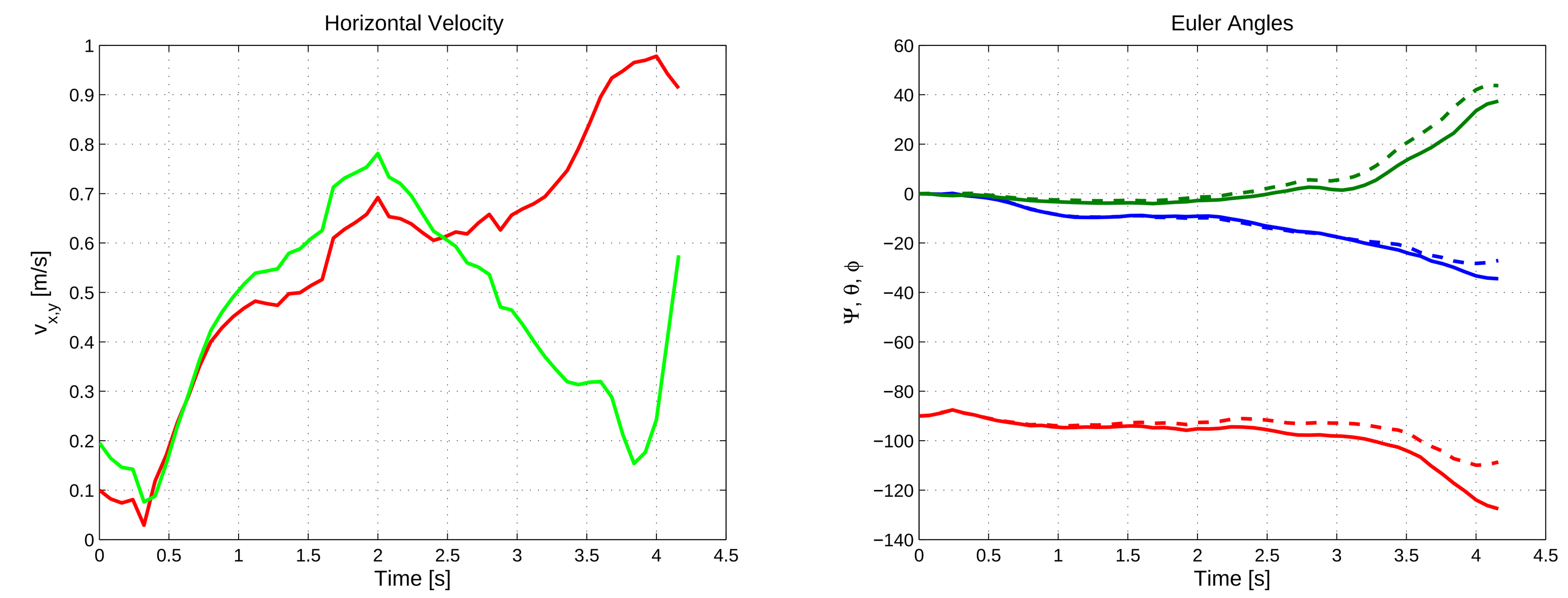
The landmark initialisation will introduce outlier associations which are removed using the IMU data, which describe the motion independently of cluster appearance. We iterate solving (4) and then remove the largest projected landmark residual from each image such that the errors in each image are of similar size. This procedure is terminated when all landmark residuals are within a predefined gate.

Results

The initialisation method and subsequent NLS refinement are demonstrated on some real data.



Left: Example image from the free-hand run with extracted features shown as red stars. Right: Landmark estimates from the linear estimation (green) and from nonlinear refinement (red). Three distinct layers can be recognised corresponding to the levels of the bookshelf.



Left: Estimated horizontal velocity from the linear estimation (green) and from the nonlinear refinement (red). Right: Estimated Euler angles in degrees from camera (dash-dotted) and after nonlinear refinement (solid). Blue is the yaw, green is the pitch and red is the roll angle.

This work has been submitted to IEEE Transactions on Robotics. For more details and more results please refer to the report [1].

References

[1] M. A. Skoglund, Z. Sjanic, and F. Gustafsson. Initialisation and Estimation Methods for Batch Optimisation of Inertial/Visual SLAM. Technical Report LiTH-ISY-R-3065, Department of Electrical Engineering, Linköping University, 2013.