A Nonlinear Least-Squares Approach to the SLAM Problem

Zoran Sjanic [*,**] **Martin A. Skoglund** [*] **Thomas B. Schön** [*]
Fredrik Gustafsson [*]

[*] *Division of Automatic Control, Linköping University, SE-581 83
Linköping, Sweden (e-mail: {zoran, ms, schon, fredrik}@isy.liu.se).*
[**] *Department of Flight Data and Navigation, Saab Aeronautics,
SE-581 88 Linköping, Sweden*

Abstract:
In this paper we present a solution to the simultaneous localisation and mapping (SLAM) problem using a camera and inertial sensors. Our approach is based on the maximum a posteriori (MAP) estimate of the complete SLAM problem. The resulting problem is posed in a nonlinear least-squares framework which we solve with the Gauss-Newton method. The proposed algorithm is evaluated on experimental data using a sensor platform mounted on an industrial robot. In this way, accurate ground truth is available, and the results are encouraging.

Keywords: Inertial measurement units, Cameras, Smoothing, Dynamic systems, State estimation

1. INTRODUCTION

In this paper we present an optimisation based solution to the simultaneous localisation and mapping (SLAM) problem formulated as nonlinear least-squares, and solved with the Gauss-Newton method. The method aims at providing high quality SLAM estimates which could e.g., be used as priors for computing detailed terrain maps.

SLAM is the problem of estimating a map of the surrounding environment from a moving platform, while simultaneously localising the platform. These estimation problems usually involve nonlinear dynamics and nonlinear measurements of a high dimensional state space. In Dellaert and Kaess [2006] a nonlinear least-squares approach to SLAM, called square root Smoothing and Mapping ($\sqrt{\text{SAM}}$) is presented. We extend this approach by considering a full 6 DOF platform, 3 DOF landmarks, inputs using inertial sensors and camera measurements. The resulting algorithm is evaluated on experimental data from a structured indoor environment and compared with ground truth data.

For more than twenty years SLAM has been a popular field of research and is considered an important enabler for autonomous robotics. An excellent introduction to SLAM is given in the two part tutorial by Durrant-Whyte and Bailey [2006], Bailey and Durrant-Whyte [2006] and for a thorough overview of visual SLAM Chli [2009] is highly recommended. In the seminal work of Smith et al. [1990] the idea of a stochastic map was presented and was first used in Moutarlier and Chatila [1989], where the estimate is computed with an Extended Kalman Filter (EKF). There are by now quite a few examples of successful EKF SLAM implementations, see e.g., Guviant and Nebot [2001], Leonard et al. [2000]. Another popular approach is the FastSLAM method [Montemerlo et al., 2002, 2003] which uses particle filters. These are known to handle nonlinearities very well. Both EKF SLAM and FastSLAM suffer from inconsistencies due to poor data association, linearisation errors [Bailey et al., 2006a] and particle depletion [Bailey et al., 2006b].

Some impressive work where the SLAM problem is solved solely with cameras can be found in Davison et al. [2007], Davison [2003], Eade [2008], Klein and Murray [2007]. The camera only SLAM methods have many similarities with bundle adjustment techniques, [Hartley and Zisserman, 2004, Triggs et al., 2000], and the stochastic map estimation problem can be seen as performing structure from motion estimation [Fitzgibbon and Zisserman, 1998, Taylor et al., 1991]. Without any other sensors measuring the platform dynamics, the image frame rate and the visual information contents in the environment are limiting factors for the ego motion estimation, and hence the map quality.

Recent years' increase in computational power has made smoothing an attractive option to filtering. One of the first SLAM related publications, where the trajectory is not filtered out to a single estimate is Eustice et al. [2006], where the whole time history is estimated with a so called *delayed state information filter*. Other, more optimisation like approaches are Dellaert and Kaess [2006], Kaess et al. [2008], Bibby and Reid [2007], Bryson et al. [2009], which all optimise over the whole trajectory and a feature based map.

2. PROBLEM FORMULATION

We assume that the dynamic model and the measurements are on the following form

$$x_t = f(x_{t-1}, u_t) + \underbrace{B_w w_t}_{\tilde{w}_t}, \tag{1a}$$

$$l_t = l_{t-1}, \tag{1b}$$

$$y_{t_k} = h(x_{t_k}, l_{t_k}) + e_{t_k}, \tag{1c}$$

where x_t and l_t are vehicle and landmark states, respectively, and the inertial measurements can be modelled as inputs u_t. The meaning of y_{t_k} is a measurement relative to landmark l_{t_k} at time t_k, and this is because the measurements and the dynamic model deliver data in different rates. If we assume that all the measurements and the inputs for $t = \{0 : N\}$ and $k = \{1 : K\}$ $(K \ll N)$ are available and the noise is independent and identically distributed (i.i.d.), then the joint probability density of (1) is

$$p(x_{0:N},\ l_N | y_{1:K},\ u_{1:N}) =$$
$$p(x_0) \prod_{t=1}^{N} p_{\tilde{w}_t}(x_t | x_{t-1}, u_t) \prod_{k=1}^{K} p_{e_{t_k}}(y_{t_k} | x_{t_k}, l_{t_k}). \tag{2}$$

Note that the map, l_N, is static and the estimate is given for the last time step only. Furthermore, the initial platform state x_0 is fixed to the origin without uncertainty. This is a standard SLAM approach and x_0 is therefore treated as a constant. The smoothed maximum a posteriori (MAP) estimate of $x_{0:N}$ and l_N is then

$$[x_{0:N}^*,\ l_N^*] = \underset{x_{0:N},\ l_N}{\arg\max}\ p(x_{0:N},\ l_N | y_{t_1:t_K},\ u_{1:N}) =$$
$$\underset{x_{0:N},\ l_N}{\arg\min}\ -\log p(x_{0:N},\ l_N | y_{t_1:t_K},\ u_{1:N}). \tag{3}$$

If the noise terms $\tilde{w}_t$ and e_{t_k} are assumed to be Gaussian and white, i.e., $e_{t_k} \sim \mathcal{N}(0, R_{t_k})$ and $\tilde{w}_t \sim \mathcal{N}(0, \tilde{Q}_t)$, (3) then becomes

$$[x_{0:N}^*,\ l_N^*] = \underset{x_{0:N},\ l_N}{\arg\min} \sum_{t=1}^{N} ||x_t - f(x_{t-1}, u_t)||^2_{\tilde{Q}_t^{-1}} +$$
$$\sum_{k=1}^{K} ||y_{t_k} - h(x_{t_k}, l_{t_k})||^2_{R_{t_k}^{-1}}, \tag{4}$$

which is a nonlinear least-squares formulation.

3. MODELS

Before we introduce the details of the dynamic model some coordinate frame definitions are necessary:

- Body coordinate frame (b), moving with the sensor and with origin fixed in the IMU's inertial centre.
- Camera coordinate frame (c), moving with the sensor and with origin fixed in the camera's optical centre.
- Earth coordinate frame (e), fixed in the world with its origin arbitrary positioned.

When the coordinate frame is omitted from the states it is assumed that they are expressed in the earth frame e.

3.1 Dynamics

The dynamic model used in this application has 10 states consisting of the position and velocity of the b frame expressed in the e frame, $p^e = [p_x\ p_y\ p_z]^T$ and $v^e = [v_x\ v_y\ v_z]^T$, respectively. The orientation is described using a unit quaternion $q^{be} = [q_0\ q_1\ q_2\ q_3]^T$ defining the orientation of the b frame expressed in the e frame.

The IMU measurements are treated as inputs, reducing the state dimension needed, and we denote the specific force $u_a^b = [a_x^b\ a_y^b\ a_z^b]^T$ and denote the angular rate $u_\omega^b = [\omega_x^b\ \omega_y^b\ \omega_z^b]^T$. The dynamics of the sensor in (1a) is then

$$\underbrace{\begin{bmatrix} p_t^e \\ v_t^e \\ q_t^{be} \end{bmatrix}}_{x_t} =$$

$$\underbrace{\begin{bmatrix} I_3 & TI_3 & 0 \\ 0 & I_3 & 0 \\ 0 & 0 & I_4 \end{bmatrix} \begin{bmatrix} p_{t-1}^e \\ v_{t-1}^e \\ q_{t-1}^{be} \end{bmatrix} + \begin{bmatrix} \frac{T^2}{2}I_3 & 0 \\ TI_3 & 0 \\ 0 & \frac{T}{2} \end{bmatrix} \begin{bmatrix} \mathcal{R}(q_{t-1}^{be})^T u_{a,t}^b + g^e \\ S(u_{\omega,t}^b)q_{t-1}^{be} \end{bmatrix}}_{f(x_{t-1}, u_t)}$$

$$+ \underbrace{\begin{bmatrix} \frac{T^2}{2}I_3 & 0 \\ TI_3 & 0 \\ 0 & \frac{T}{2}\tilde{S}(q_{t-1}^{be}) \end{bmatrix}}_{B_w(x_{t-1})} \underbrace{\begin{bmatrix} w_{a,t}^b \\ w_{\omega,t}^e \end{bmatrix}}_{w_t}, \tag{5}$$

where

$$w_{a,t}^b \sim \mathcal{N}(0, Q_a), \qquad Q_a = \sigma_a I_3, \tag{6a}$$

$$w_{w,t}^e \sim \mathcal{N}(0, Q_w), \qquad Q_w = \sigma_w I_3, \tag{6b}$$

$$S(u_{\omega,t}^b) = \begin{bmatrix} 0 & -\omega_x & -\omega_y & -\omega_z \\ \omega_x & 0 & \omega_z & -\omega_y \\ \omega_y & -\omega_z & 0 & \omega_x \\ \omega_z & \omega_y & -\omega_x & 0 \end{bmatrix}, \quad \tilde{S}(q_t^{be}) = \begin{bmatrix} -q_1 & -q_2 & -q_3 \\ q_0 & -q_3 & q_2 \\ q_3 & q_0 & -q_1 \\ -q_2 & q_1 & q_0 \end{bmatrix}, \tag{6c}$$

where $\mathcal{R}(q_t^{be}) \in SO(3)$ is the rotation matrix parametrised using the unit quaternion and $\mathcal{R}(q_t^{be})^T u_{a,t}^b + g^e$ is the specific force input expressed in the e frame, where $g^e = [0\ 0\ -9.81]^T$ compensates for the earth gravitational field.

3.2 Landmark State Parametrisation

Landmark states are encoded in the Inverse Depth Parametrisation (IDP) [Civera et al., 2008]. The first three states, x^e, y^e and z^e, represent the 3D position of the camera when the landmark was first observed. The last three states describe a vector to the landmark in spherical coordinates parametrised with azimuthal angle φ^e, elevation angle θ^e and inverse distance ρ^e, giving $l^e = [x\ y\ z\ \theta\ \phi\ \rho]^T$. The angles φ^e, θ^e and the inverse distance ρ^e are expressed in the right handed earth coordinate frame e with z^e-axis pointing upwards. This means that a landmark l, with earth fixed coordinates $[x_l^e\ y_l^e\ z_l^e]^T$ is parametrised as

$$\begin{bmatrix} x_l^e \\ y_l^e \\ z_l^e \end{bmatrix} = \begin{bmatrix} x^e \\ y^e \\ z^e \end{bmatrix} + \frac{1}{\rho^e} m(\varphi^e, \theta^e), \tag{7a}$$

$$m(\varphi^e, \theta^e) = \begin{bmatrix} \cos\varphi^e \sin\theta^e \\ \sin\varphi^e \sin\theta^e \\ \cos\theta^e \end{bmatrix}. \tag{7b}$$

Since the camera is calibrated, as in Zhang [2000] using the toolbox [Bouguet, 2010], the landmark states can be introduced using normalised pixel coordinates $[u\ v]^T$ according to

$$p^e = \begin{bmatrix} x^e \\ y^e \\ z^e \end{bmatrix}, \tag{8a}$$

$$g^e = \begin{bmatrix} g_x^e \\ g_y^e \\ g_z^e \end{bmatrix} = \mathcal{R}(q_t^{be})^T \mathcal{R}(q^{bc}) \begin{bmatrix} u \\ v \\ 1 \end{bmatrix}, \tag{8b}$$

$$\varphi^e = \arctan 2(g_y^e, g_x^e), \tag{8c}$$

$$\theta^e = \arctan 2\left(||[g_x^e \; g_y^e]^T||_2, g_z^e \right), \tag{8d}$$

$$\rho^e = \frac{1}{d_0^e}. \tag{8e}$$

Here, q^{bc} is the unit quaternion describing the fixed rotation from the camera frame to the body frame. Furthermore, p^e is the camera position when the landmark is observed and d_0 is the initial depth for the landmark. Finally, $\theta = \arctan 2(\cdot)$ is the four-quadrant arc-tangent, $\theta \in [-\pi, \pi]$. The complete landmark vector is of the dimension $6 \times n_{\text{landmarks}}$ and $n_{\text{landmarks}}$ will vary depending on when new landmarks are initiated.

3.3 Camera Measurements

The measurements are sub-pixel coordinates in the images given by the SIFT feature extractor [Lowe, 1999]. The dimension of the measurement vector y_{t_k} is $2 \times n_{\text{af}}$, where n_{af} denotes the number of associated features. The measurements are expressed in normalised pixel coordinates. The camera measurement equation relating states and measurements has the form

$$y_{t_k} = \underbrace{h(x_{t_k}, l_{t_k})}_{y_{t_k}^c} + e_{t_k}, \tag{9}$$

where

$$e_{t_k} \sim \mathcal{N}(0, R_{t_k}), \; R_{t_k} = \sigma_{\text{features}} I_{2 \times n_{\text{af}}}. \tag{10}$$

Using the IDP, (7) and (8), for a single landmark j and omitting time dependency, the measurement (9) is calculated as

$$l_j^c = \begin{bmatrix} l_{x,j}^c \\ l_{y,j}^c \\ l_{z,j}^c \end{bmatrix} =$$
$$\mathcal{R}(q^{bc})^T \mathcal{R}(q^{be}) \left(\rho_j^e \left(p^e - p_j^e - \mathcal{R}(q^{be})^T r^{bc} \right) + m(\varphi_j, \theta_j) \right), \tag{11a}$$

$$y_j^c = \frac{1}{l_{z,j}^c} \begin{bmatrix} l_{x,j}^c \\ l_{y,j}^c \end{bmatrix}, \tag{11b}$$

where p_j^e and ρ_j^e are defined in (8a) and (8e), respectively. The translation r^{bc} and orientation $\mathcal{R}(q^{bc})$ defines the constant relative pose between the camera and the IMU. The parameters in r^{bc} and $\mathcal{R}(q^{bc})$ were estimated in the previous work by Hol et al. [2010].

4. SOLUTION

The proposed solution starts with an initialisation of the states using EKF SLAM and the initial states are then smoothed using nonlinear least squares.

4.1 Initialisation

The nonlinear least-squares algorithm needs an initial estimate $x_{0:N}^0$, l_N^0, which is obtained using EKF SLAM.

The time update is performed with the model (5) in a standard EKF, for details, see e.g., Kailath et al. [2000]. The the landmark states (1b) are stationary and will therefore only be corrected in the measurement update.

The measurement update needs some further explanation. Each time an image is available (which in our experiments is 8 times slower than the specific force and the angular rate inputs) a measurement update is made. The measurement update needs an association between the features extracted from the current image and the landmarks present in the state vector. The associations computed during EKF SLAM are found in the following way; first, all landmarks are projected into the image according to (9) and the most probable landmarks are chosen as the nearest neighbours inside a predefined region. Second, the SIFT feature descriptors for the landmarks and features inside the region are matched. In this way a data association sequence is created for each image, relating the measurements and the landmarks in the state vector. To enhance the feature tracking we discard unstable features (i.e., those that are only measured once or twice) and features are proclaimed usable only if they are found at least three times.

4.2 Nonlinear Least-Squares Smoothing

The nonlinear problem (4) is in our approach solved using the Gauss-Newton method, i.e., at each iteration we solve the linearised version of the problem.

In order to formulate the linearised least-squares smoothing problem for our specific setup we first need some definitions:

$$F_t \triangleq \left. \frac{\partial f(x, u)}{\partial x} \right|_{(x,u) = (x_{t-1}^0, u_t)}, \tag{12}$$

is the Jacobian of the motion model and

$$H_{t_k}^j \triangleq \left. \frac{\partial h(x, l)}{\partial x} \right|_{(x,l) = (x_{t_k}^0, l_j^0)}, \tag{13}$$

is the Jacobian of the measurement k at time t_k with respect to the vehicle states. The IDP gives a special structure to the equations since the measurements of the features are related to the pose where the features where initialised. Therefore, the landmark Jacobian is split into two parts. The first part is

$$J_{x_{t_k}}^j \triangleq \left. \frac{\partial h(x, l)}{\partial x} \right|_{(x,l) = (x_{t_k}^0, l_j^0)}, \tag{14}$$

which is the Jacobian of measurement k at time t_k, with respect to the position where landmark j was initialised. The second part is the Jacobian of measurement k at time t_k of the states ϕ_j, θ_j and ρ_j of landmark j

$$J_{t_k}^j \triangleq \left. \frac{\partial h(x, l)}{\partial l} \right|_{(x,l) = (x_{t_k}^0, l_j^0)}. \tag{15}$$

From the initialisation, Section 4.1, a trajectory $x_{0:N}^0$ and a landmark l_N^0 estimate is given and is therefor treated as a constant. The linearised process model at time t is then

$$x_t^0 + \delta x_t = F_t(x_{t-1}^0 + \delta x_{t-1}) + Bu_t + B_w(x_{t-1}^0)w_t. \tag{16}$$

The linearised measurement equations are given by

$$y_{t_k}^j = h(x_{t_k}^0, l_j^0) + H_{t_k}^j \delta x_{t_k} + J_{x_{t_k}}^j \delta x_{t_k} + J_{t_k}^j \delta l_j + e_{t_k}^j. \tag{17}$$

The linearised least-squares problem for the prediction and measurement errors is then

$$[\delta x_t^*, \ \delta l_j^*] = \arg\min_{\delta x_t, \delta l_j} \sum_{t=1}^{N} ||F_t \delta x_{t-1} - I \delta x_t - a_t||_{\tilde{Q}_t^{-1}}^2$$

$$+ \sum_{k=1}^{K} ||H_{t_k}^j \delta x_{t_k} + J_{x_{t_k}}^j \delta x_{t_k} + J_{t_k}^j \delta l_j - c_{t_k}^j||_{R_{t_k}^{-1}}^2 \qquad (18)$$

where $a_t = x_t^0 - F_t x_{t-1}^0 - Bu_t$ and $c_{t_k}^j = y_{t_k}^j - h(x_{t_k}^0, l_j^0)$. Here a_t and $c_{t_k}^j$ are the prediction errors of the linearised dynamics around x_t^0 and the innovations, respectively. The stacked version of the problem (18) can be solved iteratively according to

$$\eta^{i+1} = \arg\min_{\eta} ||A(\eta^i)\eta - b(\eta^i)||_2^2, \quad \eta^0 = 0, \qquad (19)$$

where we define $\eta = [\delta x_t, \ \delta l_j]$, and $A(\eta)$ and $b(\eta)$ is the matrix part and the vector part of (18), respectively.

The structure of the A matrix is perhaps best explained using an example:

$$A(\eta) = \begin{bmatrix} A_{11} & 0 \\ A_{21} & A_{22} \end{bmatrix} =$$

$$\begin{bmatrix} -I & & & & & & \\ F_2 & -I & & & & & \\ & F_3 & \ddots & & & & \\ & & \ddots & -I & & & \\ & & & F_6 & \ddots & & \\ & & & & \ddots & -I & \\ & & & & & F_{10} & \ddots \\ & & & & & & \ddots & -I \\ \hline J_{x_5}^1 & H_5^1 & & & & & J_5^1 & \\ J_{x_5}^2 & H_5^2 & & & & & J_5^2 \\ J_{x_9}^2 & & & H_9^2 & & & & J_9^2 \\ J_{x_{13}}^1 & & & & & H_{13}^1 & J_{13}^1 \end{bmatrix}, \qquad (20)$$

$t_k = 1$: Two landmarks are seen for the first time giving the landmarks' initialisation positions, i.e., the columns where the Jacobians (14) are placed.

$t_k = 5$: The second camera measurement arrives, landmarks 1 and 2 are observed and the first two block rows of A_{21} and A_{22} are added.

$t_k = 9$: Camera measurement 3 arrives, landmark 2 is observed and the third block row of A_{21} and A_{22} is added.

$t_k = 13$: Camera measurement 4 arrives, landmark 1 is observed and the fourth block row of A_{21} and A_{22} is added.

A single iteration of the nonlinear least-squares smoothing algorithm can be summarised in pseudo code as seen in Algorithm 1.

The least-squares problem is weighted, so it is assumed that all of the terms in (18) are multiplied with the corresponding matrix square root of the inverse of the covariance matrices for the process and the measurement noise, respectively. Note that the covariance matrix of the process noise, $\tilde{Q}_t = B_w(x_t^0) Q_t B_w(x_t^0)^T$, is singular rendering the use of normal inversion impossible. In order

---

Algorithm 1 Nonlinear Least-Squares Smoothing for SLAM

Input: x^0, l^0 (trajectory and map from previous iteration), u (inputs), data association
Output: x^s, l^s (smoothed estimate of the trajectory and the map)
N = # IMU measurements
$A = [\ ], \ a = [\ ], \ c = [\ ]$
for $i = 1$ to N **do**
  predict states, $x_i = f(x_{i-1}^0, u_i)$
  if image available **then**
    use the data association and calculate $h(x_i^0, l_i^0)$
    calculate $A_{11} = [A_{11} \ A_{11}^i]^T$,
    $A_{21} = [A_{21} \ A_{21}^i]$ and
    $A_{22} = [A_{22} \ A_{22}^i]$ according to (12) - (20)
    calculate $a_i = x_i^0 - x_i$
    calculate $c_i = y_i - h(x_i^0, l_i^0)$
    set $a = [a^T \ a_i^T]^T$
    set $c = [c^T \ c_i^T]^T$
  else
    calculate $A_{11} = [A_{11} \ A_{11}^i]^T$
    calculate $a_i = x_i^0 - x_i$
    set $a = [a^T \ a_i^T]^T$
  end if
end for
Assemble up A according to (20) and $b = [a^T \ c^T]^T$
solve the least squares problem (19)
calculate $[x^{sT}, \ l^{sT}]^T = [x^{0T}, \ l^{0T}]^T + \eta$

---

to overcome this, we simply regularise the problem by adding a diagonal matrix ΔI to the covariance matrix, with Δ being a small number, rendering the covariance matrix invertible. Furthermore, it is assumed that the associations from the initialisation is good enough and that we do not have to compute new associations after each iterate.
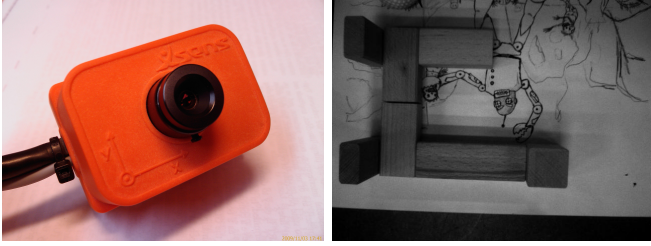
5. EXPERIMENTS

The implementation is done in MATLAB, except for the SIFT binaries, where we use a C code library from Hess [2010].

5.1 Experimental Setup

For the purpose of obtaining high quality ground truth motion data we used an IRB 1400 industrial robot from ABB. In an industrial robot the rotation and translation of the end tool can be logged with high accuracy. This gives an excellent performance evaluation possibility, which is otherwise difficult. The actual robot trajectory was not possible to acquire during the experiment. However, since the industrial robot is very accurate the actual output of the robot will be very close to the programmed trajectory.

We constructed a small synthetic environment with known topography to obtain realistic ground truth map data, see Fig. 1b. We use a combined IMU/camera sensor unit, shown in Fig. 1a. The sensor unit is mounted at the end tool position of the industrial robot. The IMU measurements are sampled at 100 Hz and images of size 640×480 pixels are sampled at 12.5 Hz.

(a) The combined strap down IMU and camera system.

(b) An image from the camera during the experiment.

Fig. 1. The IMU/camera sensor unit used in the experiments and an image from the camera over-viewing the synthetic environment.

5.2 Results

The resulting trajectories and map obtained with the data from an experiment are presented in Fig. 2. The Ground truth trajectory is a reference trajectory for the robot. From these plots it is clearly visible that the smoothed estimate is closer to the true trajectory than the initial estimate. The improvement is also visible if the initial estimate and the final smoothed landmark estimate are compared as in Fig. 3. Note that some landmark positions are already quite accurately estimated since the change is small after the smoothing. The smoothed estimate also has a more accurate universal scale.
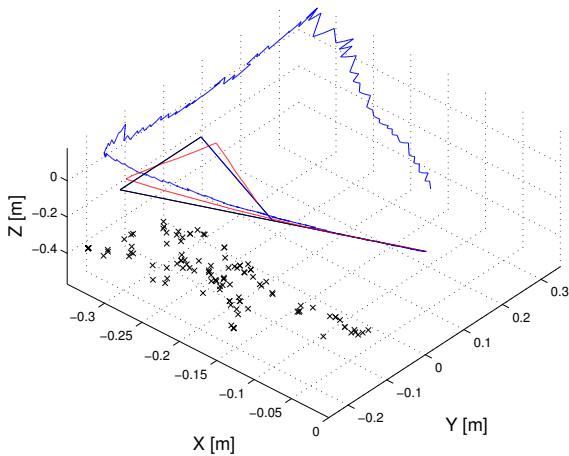

Fig. 2. The smoothed trajectory in red, the initial EKF trajectory in blue and the ground truth trajectory in black. The black crosses are the smoothed landmark estimates.

Both the smoothed horizontal speed of the platform, defined as $||[v_t^x \ v_t^y]^T||_2$, and resulting estimate from the initialisation are plotted in Fig. 4. We see that the smoothed speed is much closer to 0.1 m/s, which is the true speed.

6. CONCLUSIONS AND FUTURE WORK

In this work we have presented the SLAM problem formulated as nonlinear least-squares. For evaluation we have used a combined camera and IMU sensor unit mounted at the manipulator of an industrial robot which gives accurate ground truth.
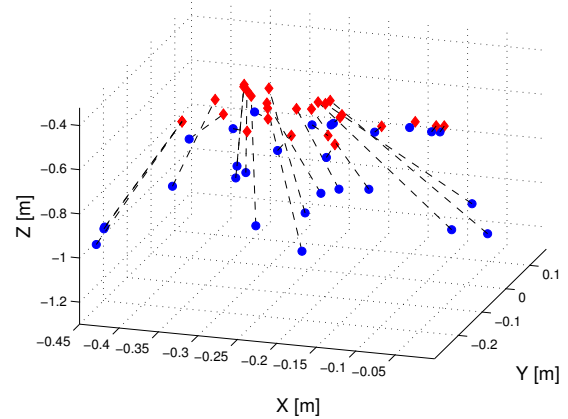

Fig. 3. The initial landmark estimates given by the EKF in blue bullets and the final smoothed estimate in red diamonds, where the black dashed line illustrate the relative displacement.
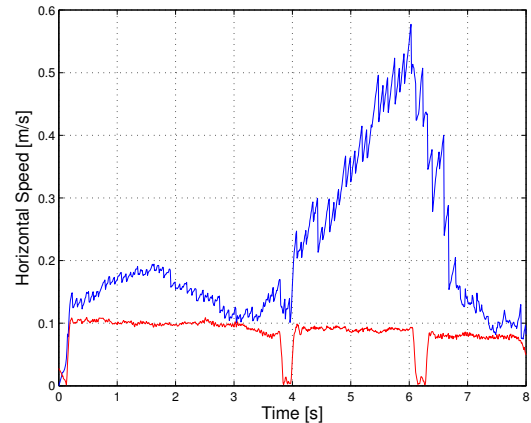

Fig. 4. The smoothed horizontal speed of the camera in red and EKF in blue. The true speed is 0.1 m/s except for when the robot stops and changes direction, this happens at about 4 seconds and 6 seconds.

The experimental results in Section 5.2 show that the nonlinear least-squares trajectory, Fig. 2, and the speed estimate, Fig. 4, show a significant improvement of the initial estimate. The sparse point cloud in Fig. 3, illustrating the initial landmark estimate and final smoothed estimate, shows also an improvement. The universal scale of the environment is improved since the landmarks have moved towards more probable positions.

For a long-term solution another initialisation procedure is necessary, since EKF SLAM is intractable for large maps. A possible alternative is to use IMU supported visual odometry to get a crude initial estimate. This approach needs a supporting global data association scheme.

laboratory, Linköping University, for expert advice on computer vision.

REFERENCES

T. Bailey and H. Durrant-Whyte. Simultaneous Localization and Mapping (SLAM): Part II. *IEEE Robotics & Automation Magazine*, 13(3):108–117, September 2006.

T Bailey, J Nieto, J. E. Guivant, M. Stevens, and E. M. Nebot. Consistency of the EKF-SLAM algorithm. In *Proceedings of the International Conference on Intelligent Robots and Systems (IROS)*, pages 3562–3568, Beijing, China, 2006a.

T. Bailey, J. Nieto, and E. M. Nebot. Consistency of the FastSLAM algorithm. In *International Conference on Robotics and Automation (ICRA)*, pages 424–429, Orlando, Florida, USA, 2006b.

C. Bibby and I. Reid. Simultaneous Localisation and Mapping in Dynamic Environments (SLAMIDE) with reversible data association. In *Proceedings of Robotics: Science and Systems (RSS)*, Atlanta, GA, USA, 2007.

J.-Y. Bouguet. http://www.vision.caltech.edu/bouguetj/calib_doc/, 2010.

M. Bryson, M. Johnson-Roberson, and S. Sukkarieh. Airborne smoothing and mapping using vision and inertial sensors. In *Proceedings of the International Conference on Robotics and Automation (ICRA)*, pages 3143–3148, Kobe, Japan, 2009. IEEE Press. ISBN 978-1-4244-2788-8.

M. Chli. *Applying Information Theory to Efficient SLAM*. PhD thesis, Imperial College London, 2009.

J. Civera, A.J. Davison, and J. Montiel. Inverse Depth Parametrization for Monocular SLAM. *IEEE Transactions on Robotics*, 24(5):932–945, Oct. 2008. ISSN 1552-3098. doi: 10.1109/TRO.2008.2003276.

A J Davison. Real-time simultaneous localisation and mapping with a single camera. In *Proceedings of the Ninth IEEE International Conference on computer vision*, pages 1403–1410, 2003.

A. J. Davison, I. D. Reid, N. D. Molton, and O. Stasse. MonoSLAM: Real-time single camera SLAM. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(6):1052–1067, 2007.

F. Dellaert and M. Kaess. Square Root SAM: Simultaneous Localization and Mapping via Square Root Information Smoothing. *International Journal of Robotics Research*, 25(12):1181–1203, 2006. ISSN 0278-3649. doi: http://dx.doi.org/10.1177/0278364906072768.

H. Durrant-Whyte and T. Bailey. Simultaneous Localization and Mapping: Part I. *IEEE Robotics & Automation Magazine*, 13(12):99–110, June 2006.

E. Eade. *Monocular Simultaneous Localisation and Mapping*. PhD thesis, Cambridge University, 2008.

R. M. Eustice, H. Singh, and J. J. Leonard. Exactly sparse delayed-state filters for view-based SLAM. *IEEE Transactions on Robotics*, 22(6):1100–1114, 2006.

A. W. Fitzgibbon and A. Zisserman. Automatic Camera Recovery for Closed or Open Image Sequences. In Hans Burkhardt and Bernd Neumann, editors, *ECCV (1)*, volume 1406 of *Lecture Notes in Computer Science*, pages 311–326. Springer, 1998. ISBN 3-540-64569-1.

J.E. Guviant and E.M. Nebot. Optimization of the simultaneous localization and map-building algorithm for real-time implementation. *IEEE Transactions on Robotics and Automation*, 17(3):242–257, June 2001. ISSN 1042-296X.

R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, second edition, 2004. ISBN 0-521-54051-8.

R. Hess. http://blogs.oregonstate.edu/hess/code/sift/, 2010.

J. Hol, T. B. Schön, and F. Gustafsson. Modeling and Calibration of Inertial and Vision Sensors. *The International Journal of Robotics Research*, 29(2), February 2010.

M. Kaess, A. Ranganathan, and F. Dellaert. iSAM: Incremental Smoothing and Mapping. *IEEE Transactions on Robotics*, 24(6):1365–1378, Dec 2008.

T. Kailath, A. H. Sayed, and B. Hassibi. *Linear Estimation*. Prentice-Hall, Uppser Saddle River, New Jersey, 2000.

G. Klein and D. Murray. Parallel tracking and mapping for small AR workspaces. In *Proceedings of the International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 225–234, Nara, Japan, 2007.

J. J. Leonard, H. Jacob, and S. Feder. A Computationally Efficient Method for Large-Scale Concurrent Mapping and Localization. In *Proceedings of the Ninth International Symposium on Robotics Research*, pages 169–176, Salt Lake City, Utah, 2000. Springer-Verlag.

D. Lowe. Object Recognition from Local Scale-Invariant Features. In *Proceedings of the Seventh International Conference on Computer Vision (ICCV'99)*, pages 1150–1157, Corfu, Greece, 1999.

M. Montemerlo, S. Thrun, D. Koller, and B. Wegbreit. FastSLAM: A Factored Solution to the Simultaneous Localization and Mapping Problem. In *Proceedings of the AAAI National Conference on Artificial Intelligence*, Edmonton, Canada, 2002. AAAI.

M. Montemerlo, S. Thrun, D. Koller, and B. Wegbreit. FastSLAM 2.0: An Improved Particle Filtering Algorithm for Simultaneous Localization and Mapping that Provably Converges. In *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence (IJCAI)*, Acapulco, Mexico, 2003.

P Moutarlier and R Chatila. Stochastic multisensory data fusion for mobile robot location and environment modelling. In *5th International Symposium on Robotics Research*, pages 207–216, Tokyo, Japan, 1989.

R. Smith, M. Self, and P. Cheeseman. Estimating uncertain spatial relationships in robotics. In *Autonomous robot vehicles*, pages 167–193. Springer-Verlag New York, Inc., New York, NY, USA, 1990. ISBN 0-387-97240-4.

C.J. Taylor, David Kriegman, and P. Anandan. Structure and Motion in Two Dimensions from Multiple Images: A Least Squares Approach. In *Proceedings of the IEEE Workshop on Visual Motion*, pages 242–248, Princeton, NJ , USA, October 1991.

B. Triggs, P. Mclauchlan, R. Hartley, and A. Fitzgibbon. Bundle adjustment - a modern synthesis. In B. Triggs, A. Zisserman, and R. Szeliski, editors, *Vision Algorithms: Theory and Practice*, volume 1883 of *Lecture Notes in Computer Science*, pages 298–372. Springer-Verlag, 2000.

Z. Zhang. A flexible new technique for camera calibration. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(11):1330–1334, November 2000. ISSN 0162-8828. doi: 10.1109/34.888718.