Solving the SLAM Problem for Unmanned Aerial Vehicles Using Smoothed Estimates

Zoran Sjanic, Martin Skoglund, Thomas B. Schön and Fredrik Gustafsson
Division of Automatic Control
Linköping University
SE-581 83 Linköping, Sweden
Email: {zoran, ms, schon, fredrik}@isy.liu.se

***Abstract*—In this paper we present a solution to the simultaneous localization and mapping (SLAM) problem for unmanned aerial vehicles (UAV) using a camera and inertial sensors. A good SLAM solution is an important enabler for autonomous robots. Our approach is based on an optimization based formulation of the problem, which results in a smoother, rather than a filter. The proposed algorithm is evaluated on experimental data and the results are compared with accurate ground truth data. The results from this comparisons are encouraging.**

I. Introduction

Simultaneous localization and mapping is the problem of estimating a map of the surrounding environment from a moving platform while simultaneously localizing the platform. Usually these estimation problems involve nonlinear dynamics and nonlinear measurements of the environment. When computing SLAM estimates there are many things to consider which will affect quality. To mention a few; How should the dynamics be modelled? Which sensors are best suited for the environment at hand? Are there demands on the algorithm to run in real-time or can the estimates be computed in a post-processing step? What level of detail is needed in the map?

In applications of UAVs a detailed map is necessary if, for example, a landing in an unknown environment is to be performed. The idea is to obtain a good estimate of the trajectory and the map using all the measurements collected during the flight (or an appropriate part of it). In [1] this optimisation approach to SLAM, called square root Smoothing and Mapping (SAM), is described for a two-dimensional problem, where measurements and dynamics run at the same rate.

In this paper we present an iterative solution to the SLAM problem based on square root smoothing of multirate measurements. The method aims at providing high quality SLAM estimates which could e.g. be used as a prior for computing detailed terrain maps. In practical applications different sensors deliver measurements at different rates. We extend the SAM solution by concerning the three-dimensional case, multirate measurements using inertial sensors and camera measurements. The algorithm is evaluated on experimental data from a structured indoor environment and compared with ground truth data.

II. Related Work

SLAM has been a popular field of research for more than twenty years and is considered an important enabler for autonomous robotics. An excellent introduction to SLAM is given in the two part tutorial [2, 3] and for a thorough cover of visual based methods [4] is highly recommended. In the seminal work [5] the idea of a stochastic map is presented. The first implementation using this idea can be found in [6] where the the estimates are computed with and extended Kalman Filter (EKF). There are by now quite a few examples of sucessful EKF SLAM implementations, see e.g., [7, 8]. Another popular approach is the FastSLAM method [9, 10] 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, linearization errors [11] and particle depletion [12]. It is likely that similar problems occur with other methods as well.

Digital cameras are today ubiquitous, and since they deliver information rich measurements at low cost they have become the most commonly used sensors in SLAM. This has spawned a field where the SLAM problem is solved with cameras only, see e.g. [13–16]. The camera only SLAM methods have many similarities with bundle adjustment techniques, [17], and the stochastic map estimation problem can be seen as performing structure from motion estimation [18, 19]. These methods are common in the computer vision domain. 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.

Aircraft applications are very challenging for various reasons; The platform state dimension is usually large and quite involved compared to ground based platforms, such as mobile robots. Aerial vehicles cover large areas fast, resulting in huge maps. Some examples of SLAM the applied to aerial applications can be found in [20–24].

The recent years increase in computational power has made smoothing an attractive option to filtering. One of the first publications where the trajectory is not filtered out to a single estimate is [25] where rather the whole time history is estimated with a delayed state information filter. Other, more optimisation like approaches are [1, 26–28] which all use

maximum likelihood estimates of the whole trajectory and a feature based map.

III. Problem Formulation

The dynamic model and the camera measurements are on the following form,

$$x_t = f(x_{t-1}, u_t) + B_w w_t, \tag{1a}$$
$$l_t = l_{t-1}, \tag{1b}$$
$$y_{t_k} = h_{t_k}(x_{t_k}, l_{t_k}) + e_{t_k}, \tag{1c}$$

where x_t and l_t are vehicle and landmark states, respectively. The inertial measurements are modelled as inputs u_t. The meaning of y_{t_k} is a measurement relative to landmark l_{t_k} at time t_k (this is because the measurements and the dynamic model deliver data in different rates). In our case the measurements are pixel coordinates in an image given by the SIFT feature extractor [29]. The noise terms w_t and e_t are assumed to be Gaussian and white. This is essentially a SLAM formulation that can be solved within a standard EKF framework, see e.g. [2, 5]. The resulting trajectory $x^0_{0:N}$ and landmark estimate l^0_N can be used to linearise the dynamics around x^0_t at time t as

$$x^0_t + \delta x_t = f(x^0_{t-1}, u_t) + F_t \delta x_{t-1} + B_w w_t, \tag{2}$$

where δx_t is a small deviation from x^0_t and F_t is the derivative of $f(x_{t-1}, u_t)$ with respect to x_{t-1} evaluated in x^0_{t-1}. The measurement equation around l^0_N and $x^0_{t_k}$ for the measurement at time t_k is

$$y_{t_k} = h_{t_k}(x^0_{t_k}, l^0_N) + H_{t_k}\delta x_{t_k} + J_{t_k}\delta l_N + e_{t_k}, \tag{3}$$

where δl_N is a small deviation from l^0_N, H_{t_k} is the derivative of $h_{t_k}(x_{t_k}, l_N)$ with respect to x_{t_k} evaluated at $x^0_{t_k}$ and l^0_N and J_{t_k} is the derivative of $h_{t_k}(x_{t_k}, l_N)$ with respect to l_N evaluated at $x^0_{t_k}$ and l^0_N.

By rearranging the terms in (2) and (3) we can formulate the least-squares problem of finding the trajectory and map deviations that minimises the noise variance as

$$\underset{\delta x_t,\ \delta l_N}{\arg\min} \sum_{t=1}^{N} ||w_t||^2_{\widetilde{Q}_t^{-1}} + \sum_{k=1}^{K} ||e_{t_k}||^2_{R_{t_k}^{-1}}, \tag{4}$$

where $\widetilde{Q}_t = B_w Q_t B_w^T$.

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

- Body coordinate frame (b), moving with the vehicle and with origin fixed in the IMU's inertial center
- Camera coordinate frame (c), moving with the vehicle and with origin fixed in the camera's optical center
- Earth coordinate frame (e), fixed in the world with origin arbitrary positioned. When coordinate frame is omitted from the states it is assumed that they are expressed in the earth frame.

A. Dynamics

The dynamic model used in this application has 10 states consisting of $p = [p_x\ p_y\ p_z]^T$ which is the position of the IMU, $v = [v_x\ v_y\ v_z]^T$ which is the velocity of the IMU and $q^{eb} = [q^{eb}_0\ q^{eb}_1\ q^{eb}_2\ q^{eb}_3]^T$ which is the quaternion defining the rotation of the IMU from the body to the earth frame. Accelerations are modelled as input $u_a = [a_x\ a_y\ a_z]^T$ and angular rates are modelled as time varying parameters $\omega = [\omega_x\ \omega_y\ \omega_z]^T$ rather then treat them as states. Landmark states are according to inverse depth parametrisation, [30], of dimension 6. First three states, x, y and z, represent the 3D position of the vehicle when the landmark was first observed. Last three states describe a vector to the landmark in spherical coordinates parametrised with azimuthal angle φ, elevation angle θ and inverse distance ρ, giving $l = [x\ y\ z\ \theta\ \phi\ \rho]^T$. Note that φ and θ are expressed in the earth coordinate frame, e, with z-axis pointed upwards. This means that a landmark with earth fixed coordinates $[X\ Y\ Z]^T$ is parametrised as

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} + \frac{1}{\rho} m(\varphi, \theta), \tag{5a}$$
$$m(\varphi, \theta) = \begin{bmatrix} \cos\varphi \sin\theta \\ \sin\varphi \sin\theta \\ \cos\theta \end{bmatrix} \tag{5b}$$

and landmark states are created from the normalised pixel coordinates $[u\ v]^T$ in the following way

$$p^c = \begin{bmatrix} x \\ y \\ z \end{bmatrix}, \tag{6a}$$
$$g = \begin{bmatrix} g_x \\ g_y \\ g_z \end{bmatrix} = \mathcal{R}(q^{eb}_t)\mathcal{R}(q^{bc}) \begin{bmatrix} u \\ v \\ 1 \end{bmatrix}, \tag{6b}$$
$$\varphi = \text{atan2}(g_y, g_x), \tag{6c}$$
$$\theta = \text{atan2}\left(\sqrt{g_x^2 + g_y^2}, g_z\right), \tag{6d}$$
$$\rho = \frac{1}{d_0}. \tag{6e}$$

Here, $\mathcal{R}(q^{bc})$ is the rotation matrix describing the rotation from the camera frame to the body frame, $\mathcal{R}(q^{eb}_t)$ is the rotation from the body frame to the earth frame, p^c is the camera position when the landmark is observed and d_0 is the initial depth for the landmark. Finally, atan2 is a function that gives angles $\theta \in [-\pi, \pi]$ and is a variation of the $\arctan$ function. 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. Because we have no knowledge of the motion of the vehicle and the map is static, the dynamics (1a)-(1b) in our

case becomes

$$\underbrace{\begin{bmatrix} p_{t+1} \\ v_{t+1} \\ q^{eb}_{t+1} \end{bmatrix}}_{x_{t+1}} = \begin{bmatrix} I_3 & TI_3 & 0 \\ 0 & I_3 & 0 \\ 0 & 0 & I_4 + \frac{T}{2}S(\omega_t) \end{bmatrix} \underbrace{\begin{bmatrix} p_t \\ v_t \\ q^{eb}_t \end{bmatrix}}_{x_t} + \begin{bmatrix} \frac{T^2}{2}I_3 \\ TI_3 \\ 0 \end{bmatrix} \underbrace{\left[\mathcal{R}(q^{eb}_t)u_{a,t} + g^e\right]}_{u_t} + \underbrace{\begin{bmatrix} \frac{T^2}{2}I_3 & 0 \\ TI_3 & 0 \\ 0 & \frac{T}{2}\tilde{S}(q^{eb}_t) \end{bmatrix}}_{B_w(x_t)} \underbrace{\begin{bmatrix} w_{a,t} \\ w_{\omega,t} \end{bmatrix}}_{w_t}, \tag{7a}$$

$$l_{t+1} = l_t, \tag{7b}$$

where

$$w_{a,t} \sim \mathcal{N}(0, Q_a),\ Q_a = \sigma_a I_3, \tag{8a}$$

$$w_{w,t} \sim \mathcal{N}(0, Q_w),\ Q_\omega = \sigma_\omega I_3, \tag{8b}$$

$$S(\omega_t) = \begin{bmatrix} 0 & -\omega_{x,t} & -\omega_{y,t} & -\omega_{z,t} \\ \omega_{x,t} & 0 & \omega_{z,t} & -\omega_{y,t} \\ \omega_{y,t} & -\omega_{z,t} & 0 & \omega_{x,t} \\ \omega_{z,t} & \omega_{y,t} & -\omega_{x,t} & 0 \end{bmatrix}, \tag{8c}$$

$$\tilde{S}(q^{eb}_t) = \begin{bmatrix} -q_{1,t} & -q_{2,t} & -q_{3,t} \\ q_{0,t} & -q_{3,t} & q_{2,t} \\ q_{3,t} & q_{0,t} & -q_{1,t} \\ -q_{2,t} & q_{1,t} & q_{0,t} \end{bmatrix} \tag{8d}$$

and $\mathcal{R}(q^{eb}_t)u_{a,t}+g^e$ is the accelerometer measurements rotated from the body coordinate frame, to the earth coordinate frame, where $g^e = [0\ 0\ -9.81]^T$ compensates for the earth gravitational field.

B. Measurements

The measurements in our setup are the features in the images which are of dimension 2. Feature extraction is performed with SIFT, [29], (C-code is downloaded from *http://web.engr.oregonstate.edu/∼hess/*). The extracted features are then matched and associated with landmarks from the state vector. The association during the EKF run is performed in the following way; all landmarks are first projected in the image and the most probable landmarks are chosen by means of spatial constraints. Then the SIFT feature descriptors for the most probable landmarks and features are matched. In this way a data association sequence is created for each image relating measurements and landmarks in the state vector. It is assumed that these associations are good enough, hence they can used in all square root SAM iterations. The whole measurement vector y_{t_k} will have dimension $2 \times n_{\text{associated features}}$ and is expressed in normalised pixel coordinates. A measurement equation relating states and measurements is also needed. It has the general form

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

where

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

For a single landmark j, the measurement (9) (omitting the time index for readability) looks like

$$l^c_j = \begin{bmatrix} l^c_{x,j} \\ l^c_{y,j} \\ l^c_{z,j} \end{bmatrix} = \\ = \mathcal{R}(q^{cb})\mathcal{R}(q^{be})\left(\rho_j\left(p - p^c_j\right) + m(\varphi_j, \theta_j)\right), \tag{11a}$$

$$y^c_j = \begin{bmatrix} \frac{l^c_{x,j}}{l^c_{z,j}} \\ \frac{l^c_{y,j}}{l^c_{z,j}} \end{bmatrix}. \tag{11b}$$

In order not to use non-stable features (i.e. those that are initialised and only measured once or twice), the landmarks are proclaimed usable first if they were measured and associated at least three times. This is also a requirement for observability, since the dimension of the landmark state is 6 while the dimension of the measurement is 2.

C. Filtering

An initial EKF run is performed with this model in the standard way with the state and covariance matrix time update, see e.g. [31],

$$\hat{x}_{t+1|t} = F_t\hat{x}_{t|t} + Bu_{a,t}, \tag{12a}$$

$$\hat{l}_{t+1|t} = \hat{l}_{t|t}, \tag{12b}$$

$$P^{xx}_{t+1|t} = F_t P^{xx}_{t|t} F^T_t + B_w(\hat{x}_{t|t})QB^T_w(\hat{x}_{t|t}), \tag{12c}$$

where

$$Q = \begin{bmatrix} Q_a & 0 \\ 0 & Q_\omega \end{bmatrix}, \tag{13a}$$

$$P_t = \begin{bmatrix} P^{xx}_t & P^{xl}_t \\ P^{lx}_t & P^{ll}_t \end{bmatrix}. \tag{13b}$$

Note that in the time update only the vehicle state covariance matrix is updated. Also note that the noise covariance matrix $B_wQB^T_w$ is singular. This is due to two reasons; first, the four parameter quaternion is an over parametrisation of the three corresponding angles, and second, the velocity is neither measured nor an input, hence the covariance $B_wQB^T_w$ will at most be of rank 6. This does not influence the EKF calculations, but needs to be addressed in the case of the square root smoother. The measurement update is performed each time an image is available (which here is 4 times slower than accelerations and angular rates) in the usual way

$$\begin{bmatrix} \hat{x}_{t|t} \\ \hat{l}_{t|t} \end{bmatrix} = \begin{bmatrix} \hat{x}_{t|t-1} \\ \hat{l}_{t|t-1} \end{bmatrix} + K_t(y_t - h(\hat{x}_{t|t-1}, \hat{l}_{t|t-1})), \tag{14a}$$

$$K_t = P_{t|t-1}C^T_t(C_tP_{t|t-1}C^T_t + R_t)^{-1}, \tag{14b}$$

$$P_{t|t-1} = P_{t|t-1} - K_tC_tP_{t|t-1} \tag{14c}$$

where

$$C_t = \left[\tfrac{\partial}{\partial x_t}h(x_t, l_t) \quad \tfrac{\partial}{\partial l_t}h(x_t, l_t)\right]\Big|_{(x_t,l_t)=(\hat{x}_{t|t-1},\hat{l}_{t|t-1})}. \tag{15}$$

D. Square Root SAM

Rather than merging measurements and predictions in a filter the basic concept of SAM is to minimize all the measurement and trajectory errors in a least-squares sense. It would be possible to pose the nonlinear least squares problem given a dynamic model and measurement model which could then be solved iteratively with for example Gauss-Newton. In our setup we choose to consider the linearised dynamic and measurement models resulting in a linear least-squares problem.

In order to formulate the square root SAM problem we first need some definitions:

$$F_t \triangleq \left.\frac{\partial f(x_{t-1}, u_t)}{\partial x_{t-1}}\right|_{x_{t-1}=x^0_{t-1}} \tag{16}$$

is the Jacobian of the motion model.

$$H^j_{t_k} \triangleq \left.\frac{\partial h_{t_k}(x_{t_k}, l_j)}{\partial x_{t_k}}\right|_{(x_{t_k},l_j)=(x^0_{t_k},l^0_j)} \tag{17}$$

is the measurement Jacobian of measurement k at time t_k.

$$J^j_{x_{t_k}} \triangleq \left.\frac{\partial h_{t_k}(x_{t_k}, l_j)}{\partial x_{t_k}}\right|_{(x_{t_k},l_j)=(x^0_{t_k},l^0_j)} \tag{18}$$

is the Jacobian of measurement k at time t_k with respect to the position where landmark j was initialised.

$$J^j_{t_k} \triangleq \left.\frac{\partial h_{t_k}(x_{t_k}, l_j)}{\partial l_j}\right|_{(x_{t_k},l_j)=(x^0_{t_k},l^0_j)} \tag{19}$$

is the Jacobian of measurement k at time t_k of the states ϕ_j, θ_j and ρ_j of landmark j.

From the EKF run an initial trajectory x^0 and landmark l^0 estimate is given and is therefor treated as a constant. The linearised process model at time t is then

$$x^0_t + \delta x_t = F_t \cdot (x^0_{t-1} + \delta x_{t-1}) + Bu_t + B_w(x^0_{t-1})w_t. \tag{20}$$

The linearised measurement equations are given by

$$y^j_{t_k} = h_{t_k}(x^0_{t_k}, l^0_j) + H^j_{t_k}\delta x_{t_k} + J^j_{x_{t_k}}\delta x_{t_k} + J^j_{t_k}\delta l_j + e^j_{t_k}. \tag{21}$$

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

$$[\delta x^*_t,\ \delta l^*_j] = \arg\min_{\delta x_t, \delta l_j} \sum_{t=1}^{N} ||F_t\delta x_{t-1} - I\delta x_t - a_t||^2_{\widetilde{Q}^{-1}_t} + \sum_{k=1}^{K} ||H^j_{t_k}\delta x_{t_k} + J^j_{x_{t_k}}\delta x_{t_k} + J^j_{t_k}\delta l_j - c^j_{t_k}||^2_{R^{-1}_{t_k}} \tag{22}$$

where $a_t = x^0_t - F_t x^0_{t-1} - Bu_t$ and $c^j_{t_k} = y^j_{t_k} - h_{t_k}(x^0_{t_k}, l^0_j)$ and $\widetilde{Q}_t = B_w(x^0_{t-1})QB^T_w(x^0_{t-1})$. a_t and $c^j_{t_k}$ are prediction errors of the linearised dynamics around x^0_t and innovations respectively. Note that the second term in (22) could include several Jacobians corresponding to different sensors at possibly different sampling rates. Now the problem can be solved according to

$$\theta^{i+1} = \arg\min_{\theta} ||A(\theta^i)\theta - b(\theta^i)||^2_2, \quad \theta^0 = 0. \tag{23}$$

Following most of the notation from [1] an example of how the structure of the A matrix looks like is illustrated in (24) and it is computed in the following way:

1) At time $t = 1$ the two landmarks are seen the first time, but landmarks are not initialised in the map until they are measured a second time.
2) At time $t = 5$ the second camera measurement arrives and landmark 1 and 2 are observed again and hence initialised in the map.
3) At time $t = 9$ camera measurement three arrives and landmark 2 is observed.
4) At time $t = 13$ camera measurement four arrives and landmark 1 is observed.

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

$$\left[\begin{array}{ccccccccc|cc}
-I & & & & & & & & & & \\
F_2 & -I & & & & & & & & & \\
 & F_3 & \ddots & & & & & & & & \\
 & & \ddots & -I & & & & & & & \\
 & & & F_6 & \ddots & & & & & & \\
 & & & & \ddots & -I & & & & & \\
 & & & & & F_{10} & \ddots & & & & \\
 & & & & & & \ddots & -I & & & \\
\hline
J^1_{x_1} & & & H^1_5 & & & & & & J^1_5 & \\
J^2_{x_1} & & & H^2_5 & & & & & & & J^2_5 \\
J^2_{x_9} & & & & & H^2_9 & & & & & J^2_9 \\
J^1_{x_{13}} & & & & & & & H^1_{13} & & J^1_{13} &
\end{array}\right] \tag{24}$$

The square root SAM algorithm can be summarised in pseudo code as seen in Algorithm 1.

As in [1], the least squares problem is weighted, so it is assumed that all of the terms in (22) are multiplied with the corresponding matrix square root of the inverse of the covariance matrices for the system and the measurement noise. As already noted the covariance matrix of the system noise is singular rendering the use of normal inversion impossible. In order to overcome this, we simply regularise the problem adding a matrix of the form ΔI to the covariance matrix, with Δ being a small number, rendering the covariance matrix invertible.

Square root SAM is implemented in general according to [1] but with some adjustments to cope with multirate signals like IMU measurements that are sampled in 100 Hz and images of size 240×320 pixels that are sampled in 25 Hz. This algorithm requires an initial trajectory, map and association between map landmarks and features in the images. An initial trajectory is obtained with EKF SLAM, which also gives a map and data association. It is assumed that these will be sufficient for the initiation of the SAM iterations.

The 6 parameters of the inverse depth parametrisation also needs to be handled, since the 2 dimensional measurements

will cause rank deficiency if not enough measurements of the landmarks are available. The inverse depth parametrisation will also make the structure of the matrix in the square root problem different to [1] since measurements of the features are related to the to the pose where the feature was initialised.

IV. Experiments

A. Experimental Setup

For the purpose of obtaining realistic data with good ground truth a synthetic environment was build up, see Fig. 1. An IMU/Camera system mounted on the arm of the industrial robot is then "flown" above it. Fig. 2 illustrates the IMU/camera and an image from the camera during the experiment.

In a industrial robot the rotation and translation of the end tool is usually known with high accuracy. This allows for excellent performance evaluation, which will be difficult if a real flight data is used where GPS or DGPS must be used as a ground truth and orientation is of less good quality as well.

B. Results

The results obtained with the above mentioned data set are presented in Fig. 3 and Fig. 4 below. Ground truth trajectory is a reference trajectory for the robot. The actual robot trajectory

Algorithm 1 Multirate square root SAM

Input: x^0 (initial trajectory and map), u (inputs), data association
Output: x^s (smoothed estimate of the trajectory and map)
N = # IMU measurements
$A = [\ ]$
$a = [\ ]$
$c = [\ ]$
for i = 1 to N **do**
  if image available **then**
    predict states, $x_i = f(x_{i-1}^0, u_i)$
    use 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 (16) - (24)
    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
    predict states, $x_i = f(x_{i-1}^0, u_i)$
    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
build up A according to (24) and $b = [a^T\ c^T]^T$
solve the least squares problem (23)
calculate $x^s = x^0 + \theta^*$

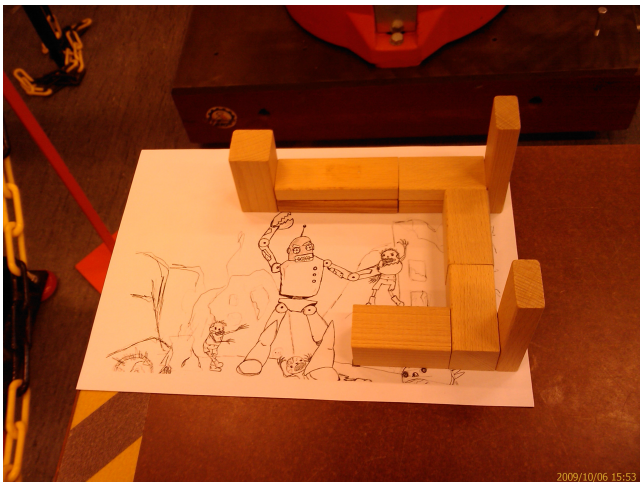
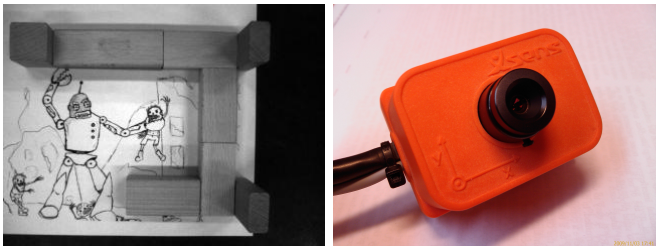


Fig. 1: Synthetic environment used in the experiments.



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

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

Fig. 2: An image of the experimental environment in (a) acquired from the camera/IMU in (b).

was not possible to acquire during the experiment. However, since the industrial robot is very accurate, this should not be a problem.

We see that there are few landmarks in the elevated areas i.e. the wooden blocks. This is because the image resolution was only 240×320 pixels and the SIFT features are not stable in those areas. Both the smoothed speed of the camera ($\mathbf{v}_t = \sqrt{(v_t^x)^2 + (v_t^y)^2}$) and resulting estimate from the EKF are plotted in Fig. 5. We see that the smoothed speed is much closer to 0.1 m/s, which is the true speed.

V. Conclusions and Future Work

The experimental results in Section IV-B indicates that the square root SAM trajectory, Fig. 4, and the speed estimates, Fig. 5 is an improvement of the initial EKF SLAM run. The sparse point cloud in Fig. 4, representing the landmarks estimate, of are good in areas at $-0.5m$ height. This is due to the previous mentioned problem of finding stable SIFT features on the wooden blocks.

As mentioned in Section III-C, the covariance matrix of the system noise is singular and thus not invertible, making the problem impossible to solve unless some workaround is implied. Our solution at the moment is to regularise the covariance matrix and in that way make it invertible. This

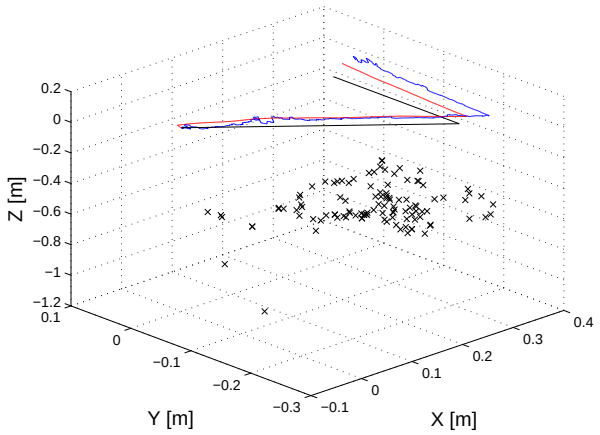


Fig. 3: Smoothed trajectory in red, EKF trajectory in blue, "Ground Truth" trajectory in black. The black crosses are landmarks.

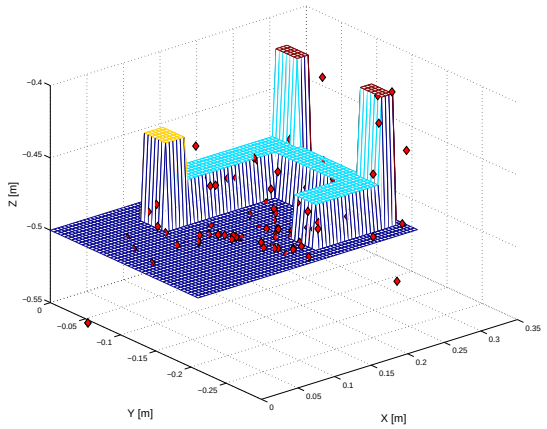


Fig. 4: Estimated smoothed landmarks and ground truth of the environment.

solution works, as seen in the Section IV-B, but it has some shortcomings. First, the regularisation will allow the states where dynamics is exact to deviate from the model, but since the weights on those states are large, the deviation will not be big. In this case, since the dynamics is exact, it should be better to optimise over as an equality constraint. This would result in a constrained least squares problem. Depending on the problem at hand, minimization of the prediction and measurement errors in (23) can be rewritten as standard optimization problems with for example equality constraints in both the measurements and the dynamics. By adding different regularisation terms in the minimization various kinds of noise descriptions could possibly be handled better than in a filter.

The Second thing that needs to be improved is the map. As it is now the landmarks that are created based on the SIFT features are too sparse. This is due to the low camera resolution and that SIFT cannot find stable features in some areas. This

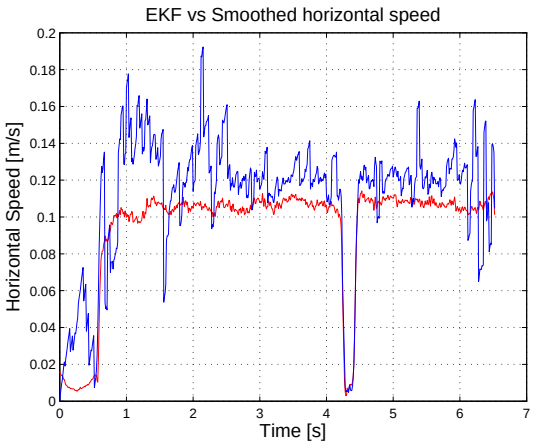


Fig. 5: Smoothed speed of the camera in red and EKF in blue.

prevents us from creating a good terrain profile that could be used for landing applications. For this purpose a more dense map of the environment is needed. In [32, 33] dense terrain maps and a refined trajectory are computed simultaneously. The same method can be used here, except that there is no need to estimate the trajectory during map estimation.

REFERENCES

[1] F. Dellaert and M. Kaess, "Square Root SAM: Simultaneous Localization and Mapping via Square Root Information Smoothing," *Int. J. Rob. Res.*, vol. 25, no. 12, pp. 1181–1203, 2006.

[2] H. Durrant-Whyte and T. Bailey, "Simultaneous Localization and Mapping: Part I," *IEEE Robotics & Automation Magazine*, vol. 13, no. 12, pp. 99–110, June 2006.

[3] T. Bailey and H. Durrant-Whyte, "Simultaneous localization and mapping (SLAM): Part II," *IEEE Robotics & Automation Magazine*, vol. 13, no. 3, pp. 108–117, Sep. 2006.

[4] M. Chli, "Applying information theory to efficient SLAM," Ph.D. dissertation, Imperial College London, 2009.

[5] R. Smith, M. Self, and P. Cheeseman, *Estimating uncertain spatial relationships in robotics*. New York, NY, USA: Springer-Verlag New York, Inc., 1990.

[6] P. Moutarlier and R. Chatila, "Stochastic multisensory data fusion for mobile robot location and environment modelling," in *In 5th International Symposium on Robotics Research*, 1989, pp. 85–94.

[7] J. Guviant and E. Nebot, "Optimization of the simultaneous localization and map-building algorithm for real-time implementation," *IEEE Transactions on Robotics and Automation*, vol. 17, no. 3, pp. 242–257, June 2001.

[8] 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*. Springer-Verlag, 2000, pp. 169–176.

[9] 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: AAAI, 2002.

[10] ——, "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: IJCAI, 2003.

[11] T. Bailey, J. Nieto, J. E. Guivant, M. Stevens, and E. M. Nebot, "Consistency of the ekf-slam algorithm," in *IROS*, October 9-15, Beijing, China, 2006, pp. 3562–3568.

[12] T. Bailey, J. Nieto, and E. M. Nebot, "Consistency of the fastslam algorithm," in *ICRA*, May 15-19, Orlando, Florida, USA, 2006, pp. 424–429.

[13] 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*, no. 29, 2007.

[14] A. J. Davison, "Real-time simultaneous localisation and mapping with a single camera," in *In Proceedings. Ninth IEEE International Conference on computer vision*, 2003, pp. 1403–1410.

[15] E. Eade, "Monocular simultaneous localisation and mapping," Ph.D. dissertation, Cambridge University, 2008.

[16] G. Klein and D. Murray, "Parallel tracking and mapping for small ar workspaces," in *In ISMAR 2007: Proceedings of the Sixth IEEE and ACM International Symposium on Mixed and Augmented Reality*, 2007, pp. 225–234.

[17] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge University Press, ISBN: 0521540518, 2004.

[18] A. W. Fitzgibbon and A. Zisserman, "Automatic camera recovery for closed or open image sequences," in *ECCV (1)*, ser. Lecture Notes in Computer Science, H. Burkhardt and B. Neumann, Eds., vol. 1406. Springer, 1998, pp. 311–326.

[19] C. Taylor, D. 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*, October 1991.

[20] R. Karlsson, T. Schön, D. Törnqvist, G. Conte, and F. Gustafsson, "Utilizing model structure for efficient simultaneous localization and mapping for a uav application," in *Proceedings 2008 IEEE Aerospace Conference*, Big Sky, MT, USA, March 2008.

[21] F. Caballero, L. Merino, J. Ferruz, and A. Ollero, "Vision-based odometry and slam for medium and high altitude flying uavs," *Journal of Intelligent and Robotics Syststems*, vol. 54, no. 1-3, pp. 137–161, 2009.

[22] M. Bryson and S. Sukkarieh, "Architectures for cooperative airborne simultaneous localisation and mapping," *Journal of Intelligent and Robotic Systems*, vol. 55, no. 4-5, pp. 267–297, 2009.

[23] T. Lupton and S. Sukkarieh, "Efficient integration of inertial observations into visual slam without initialization," in *IROS*. IEEE, 2009, pp. 1547–1552.

[24] ——, "Removing scale biases and ambiguity from 6dof monocular slam using inertial," in *ICRA*. IEEE, 2008, pp. 3698–3703.

[25] R. M. Eustice, H. Singh, and J. J. Leonard, "Exactly sparse delayed-state filters for view-based slam," *IEEE Transactions on Robotics*, vol. 22, no. 6, pp. 1100–1114, 2006.

[26] M. Kaess, A. Ranganathan, and F. Dellaert, "iSAM: Incremental smoothing and mapping," *IEEE Trans. on Robotics, TRO*, vol. 24, no. 6, pp. 1365–1378, Dec 2008.

[27] C. Bibby and I. Reid, "Simultaneous localisation and mapping in dynamic environments (slamide) with reversible data association," in *In Proceedings of Robotics: Science and Systems*, 2007.

[28] M. B, M. Johnson-Roberson, and S. Sukkarieh, "Airborne smoothing and mapping using vision and inertial sensors," in *ICRA'09: Proceedings of the 2009 IEEE international conference on Robotics and Automation*. Piscataway, NJ, USA: IEEE Press, 2009, pp. 3143–3148.

[29] D. Lowe, "Object recognition from local scale-invariant features," in *Proceedings of the Seventh International Conference on Computer Vision (ICCV'99)*, 1999, pp. 1150–1157.

[30] J. Civera, A. Davison, and J. Montiel, "Inverse Depth Parametrization for Monocular SLAM," *Robotics, IEEE Transactions on*, vol. 24, no. 5, pp. 932–945, Oct. 2008.

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

[32] J. F. Montgomery, A. E. Johnson, S. I. Roumeliotis, and L. H. Matthies, "The Jet Propulsion Laboratory Autonomous Helicopter Testbed: A platform for planetary exploration technology research and development," *Journal of Field Robotics*, vol. 23, no. 3-4, pp. 245–267, 2006.

[33] R. A. Newcombe and A. J. Davison, "Live dense reconstruction with a single moving camera," in *To be published on IEEE Conference on Computer Vision and pattern Recognition 2010*, 2010.