# Visual-Inertial SLAM using Extended Kalman Filter

**Eskil Berg Ould-Saada**
University of California San Diego
La Jolla, CA92093
eskilbo@stud.ntnu.no

## 1 Introduction

In the latter years, Simultaneous Localisation and Mapping (SLAM) has been an important part of computer vision and AI when it comes to autonomous systems with motion planning and estimation in general. Especially for self-driving drones discovering an undiscovered area, a good system for localisation and mapping is immense. In short, SLAM is the problem of creating and updating a map of an unknown area whilst keeping track of an agent's localisation at the same time. To solve this problem, several sensors and different techniques are used.

In this project, Visual-Inertial Simultaneous Localisation and Mapping (VI-SLAM) was implemented using and extended Kalman filter (EKF). Using synchronized measurements from an IMU and a stereo camera as well as camera calibration and the calibration between the two sensors specifying the transformation from the IMU to the left camera frame, the car's localisation was predicted, the visual features were mapped and a visual-inertial SLAM was implemented. Figure 1 shows visual features such as those that are mapped in this project across the left-right cameras and across time. Throughout this report, a problem formulation will be presented in mathematical terms, as well as a description of the technical approach, before the results will be presented.

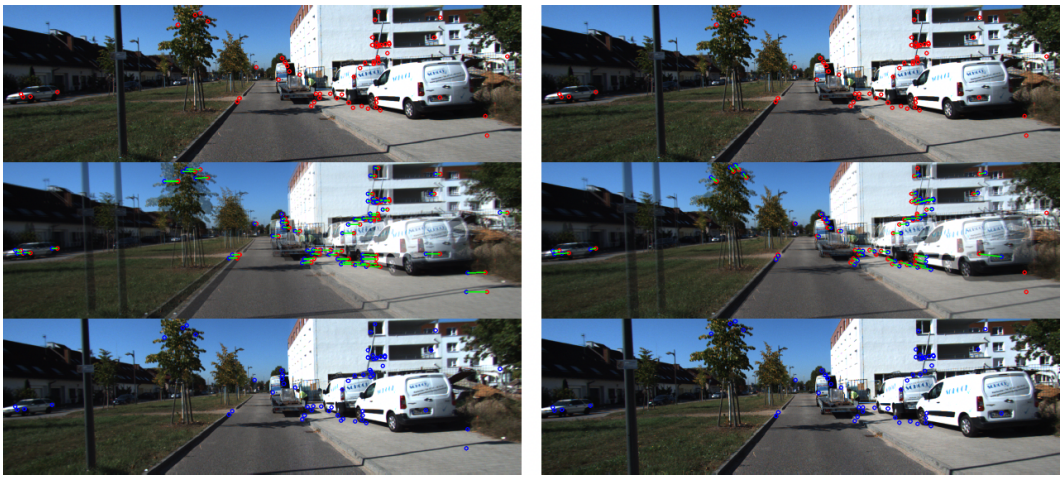

Figure 1: Visual features across left-right cameras and across time.

## 2 Problem Formulation

The problem wanting solving in this project is localization of the driving car as well as visual mapping of landmarks, and completing the visual-inertial SLAM algorithm by combining the localization and mapping.

### 2.1 Dataset

The dataset given in this project consists of four parts. First are the IMU measurements that include the linear velocity $\mathbf{v_t} \in \mathbb{R}^3$ and angular velocity $\omega_\mathbf{t} \in \mathbb{R}^3$ measured in the frame of the IMU. Second are the time stamps $\tau$. Third is the intrinsic calibration that includes the baseline $b$ and calibration matrix

$$K = \begin{bmatrix} fs_u & 0 & c_u \\ 0 & fs_v & c_v \\ 0 & 0 & 1 \end{bmatrix} \tag{1}$$

Lastly, the data includes the extrinsic calibration in form of the transformation ${}_IT_C \in SE(3)$.

### 2.2 SLAM Problem Formulation

Given the series of controls $u_t$, observations $z_t$ over time steps $t$, the goal of this project is to estimate the agent's location $x_t$ and a map of the landmarks around the agent $m_t$. Thus, the problem at hand in this project is to compute

$$P\left(m_{t+1}, x_{t+1} \mid z_{1:t+1}, u_{1:t}\right) \tag{2}$$

Using Bayes' rule, one can solve this by computing for the location posteriors

$$P\left(x_t \mid z_{1:t}, u_{1:t}, m_t\right) = \sum_{m_{t-1}} P\left(z_t \mid x_t, m_t, u_{1:t}\right) \sum_{x_{t-1}} P\left(x_t \mid x_{t-1}\right) P\left(x_{t-1} \mid m_t, z_{1:t-1}, u_{1:t}\right) \tag{3}$$

and for the map update one solves

$$P\left(m_t \mid x_t, z_{1:t}, u_{1:t}\right) = \sum_{x_t} \sum_{m_t} P\left(m_t \mid x_t, m_{t-1}, z_t, u_{1:t}\right) P\left(m_{t-1}, x_t \mid z_{1:t-1}, m_{t-1}, u_{1:t}\right) \tag{4}$$

Overall for the project, the big parts is to solve the two latter equations. This will be done using the Extended Kalman Filter as presented below. The problem will be to localizate the IMU using EKF prediction, do visual mapping using EKF update then update the IMU localization on the final part to complete the V-I SLAM algorithm.

## 3 Technical Approach

This section will contain all the technical approach done in this project.

### 3.1 Extended Kalman Filter

To solve the problems stated above, one must use the Extended Kalman Filter prediction and update formulas. Given the prior

$$x_t \mid z_{0:t}, u_{0:t-1} \sim \mathcal{N}\left(\mu_{t|t}, \Sigma_{t|t}\right) \tag{5}$$

the motion model

$$x_{t+1} = f\left(x_t, u_t, w_t\right), \quad w_t \sim \mathcal{N}(0, W) \tag{6}$$

and the observation model

$$z_t = h\left(x_t, v_t\right), \quad v_t \sim \mathcal{N}(0, V) \tag{7}$$

the prediction of the mean and covariances are computed as follows:

$$\begin{aligned} \mu_{t+1|t} &= f\left(\mu_{t|t}, u_t, 0\right) \\ \Sigma_{t+1|t} &= F_t \Sigma_{t|t} F_t^T + Q_t W Q_t^T \end{aligned} \tag{8}$$

where $F_t := \frac{df}{dx}\left(\boldsymbol{\mu}_{t|t}, \mathbf{u}_t, \mathbf{0}\right)$ and $Q_t := \frac{df}{dw}\left(\boldsymbol{\mu}_{t|t}, \mathbf{u}_t, \mathbf{0}\right)$ are the Jacobians.

The update step of the EKF are as follows:

$$\begin{aligned} \mu_{t+1|t+1} &= \mu_{t+1|t} + K_{t+1|t}\left(z_{t+1} - h\left(\mu_{t+1|t}, 0\right)\right) \\ \Sigma_{t+1|t+1} &= \left(I - K_{t+1|t} H_{t+1}\right) \Sigma_{t+1|t} \end{aligned} \tag{9}$$

where $H_{t+1} := \frac{dh}{dx}\left(\mu_{t+1|t}, 0\right)$ and $R_{t+1} := \frac{dh}{dv}\left(\mu_{t+1|t}, 0\right)$ are the Jacobians and

$$K_{t+1|t} = \Sigma_{t+1|t} H_{t+1}^T \left(H_{t+1} \Sigma_{t+1|t} H_{t+1}^T + R_{t+1} V R_{t+1}^T\right)^{-1} \tag{10}$$

is the Kalman gain.

### 3.2 IMU Localization using EKF Prediction

Given the IMU measurements $u_t$ and the landmark observations $z_t$, the prior IMU pose is

$$T_t \mid \mathbf{z}_{0:t}, \mathbf{u}_{0:t-1} \sim \mathcal{N}\left(\boldsymbol{\mu}_{t|t}, \Sigma_{t|t}\right) \text{ with } \boldsymbol{\mu}_{t|t} \in SE(3) \text{ and } \Sigma_{t|t} \in \mathbb{R}^{6\times 6} \tag{11}$$

and the EKF prediction step becomes

$$\begin{aligned} \boldsymbol{\mu}_{t+1|t} &= \boldsymbol{\mu}_{t|t} \exp\left(\tau_t \mathbf{u}_{\mathbf{skew}\, t}\right) \\ \Sigma_{t+1|t} &= \mathbb{E}\left[\delta\boldsymbol{\mu}_{t+1|t} \delta\boldsymbol{\mu}_{t+1|t}^\top\right] = \exp\left(-\tau\hat{\mathbf{u}}_t\right) \Sigma_{t|t} \exp\left(-\tau\hat{\mathbf{u}}_t\right)^\top + W \end{aligned} \tag{12}$$

where

$$\mathbf{u}_t := \begin{bmatrix} \mathbf{v}_t \\ \boldsymbol{\omega}_t \end{bmatrix} \in \mathbb{R}^6 \quad \mathbf{u}_{\mathbf{skew}\, t} := \begin{bmatrix} \hat{\boldsymbol{\omega}}_t & \mathbf{v}_t \\ \mathbf{0}^\top & 0 \end{bmatrix} \in \mathbb{R}^{4\times 4} \quad \hat{\mathbf{u}}_t := \begin{bmatrix} \hat{\boldsymbol{\omega}}_t & \hat{\mathbf{v}}_t \\ 0 & \hat{\boldsymbol{\omega}}_t \end{bmatrix} \in \mathbb{R}^{6\times 6} \tag{13}$$

### 3.3 Visual Mapping using EKF Update

For the visual mapping, the following observation model is used:

$$\mathbf{z}_{t,i} = h\left(T_t, \mathbf{m}_j\right) + \mathbf{v}_{t,i} := K_s \pi\left(o T_l T_t^{-1} \underline{\mathbf{m}}_j\right) + \mathbf{v}_{t,i} \tag{14}$$

where $\underline{\mathbf{m}}_j$ is the homogeneous coordinate of the observed landmarks.

Here,

$$K_s := \begin{bmatrix} fs_u & 0 & c_u & 0 \\ 0 & fs_v & c_v & 0 \\ fs_u & 0 & c_u & -fs_u b \\ 0 & fs_v & c_v & 0 \end{bmatrix} \tag{15}$$

and

$$\pi(\mathbf{q}) := \frac{1}{q_3}\mathbf{q} \in \mathbb{R}^4 \tag{16}$$

All the observations end up as

$$\mathbf{z}_t = K_s \pi\left(o_I T_t^{-1} \underline{\mathbf{m}}\right) + \mathbf{v}_t \quad \mathbf{v}_t \sim \mathcal{N}(\mathbf{0}, I \otimes V) \quad I \otimes V := \begin{bmatrix} V & & \\ & \ddots & \\ & & V \end{bmatrix} \tag{17}$$

The prior for the mapping is:

$$\mathbf{m} \mid \mathbf{z}_{0:t} \sim \mathcal{N}\left(\boldsymbol{\mu}_t, \Sigma_t\right) \text{ with } \boldsymbol{\mu}_t \in \mathbb{R}^{3M} \text{ and } \Sigma_t \in \mathbb{R}^{3M \times 3M} \tag{18}$$

which leads to the updates for the map given a new observation $\mathbf{z}_{t+1} \in \mathbb{R}^{4N_{t+1}}$:

$$\begin{aligned} K_{t+1} &= \Sigma_t H_{t+1}^\top \left(H_{t+1} \Sigma_t H_{t+1}^\top + I \otimes V\right)^{-1} \\ \boldsymbol{\mu}_{t+1} &= \boldsymbol{\mu}_t + K_{t+1}(\mathbf{z}_{t+1} - \underbrace{K_s \pi\left(o T_I T_{t+1}^{-1} \boldsymbol{\mu}_t\right)}_{\tilde{z}_{t+1}}) \\ \Sigma_{t+1} &= \left(I - K_{t+1} H_{t+1}\right) \Sigma_t \end{aligned} \tag{19}$$

where $H_{t+1}$ is the Jacobian of the predicted observation with respect to $m_j$:

$$H_{t+1,i,j} = \begin{cases} K_s \frac{d\pi}{d\mathbf{q}}\left(o T_l T_{t+1}^{-1} \underline{\mu}_{t,j}\right) o T_l T_{t+1}^{-1} P^\top & \text{if observation i corresponds to landmark j at time t} \\ 0, & \text{otherwise} \end{cases} \tag{20}$$

### 3.4 Visual-Inertial SLAM

The two previous parts are combined to complete the VI SLAM algorithm. The prior pose of the VI SLAM is

$$T_{t+1} \mid z_{0:t}, u_{0:t} \sim \mathcal{N}\left(\boldsymbol{\mu}_{t+1|t}, \Sigma_{t+1|t}\right) \text{ with } \boldsymbol{\mu}_{t+1|t} \in SE(3) \text{ and } \Sigma_{t+1|t} \in \mathbb{R}^{6\times 6} \tag{21}$$

The observation model with measurement noise $\mathbf{v}_t \sim \mathcal{N}(0, V)$ is

$$\mathbf{z}_{t+1,i} = h\left(T_{t+1}, \mathbf{m}_j\right) + \mathbf{v}_{t+1,i} := K_s \pi\left(o T_l T_{t+1}^{-1} \underline{\mathbf{m}}_j\right) + \mathbf{v}_{t+1,i} \tag{22}$$

To update the model, one needs the observation model Jacobian $H_{t+1} \in \mathbb{R}^{4N_{t+1}\times 6}$ with respect to the IMU pose $T_{t+1}$, evaluated at $\mu_{t+1|t}$ where each element of $H_{t+1}$ corresponds to different observations.

The prior mean and covariance for the update are

$$\mu_{t+1|t} \in SE(3) \text{ and } \Sigma_{t+1|t} \in \mathbb{R}^{6\times 6} \tag{23}$$

By computing the predicted observation

$$\tilde{\mathbf{z}}_{t+1,i} := K_s \pi\left(o T_I \mu_{t+1|t}^{-1} \underline{\mathbf{m}}_j\right) \quad \text{for } i = 1, \ldots, N_{t+1} \tag{24}$$

and the Jacobians

$$H_{t+1,i} = -K_s \frac{d\pi}{d\mathbf{q}}\left(o T_I \mu_{t+1|t}^{-1} \underline{\mathbf{m}}_j\right) o T_I \left(\boldsymbol{\mu}_{t+1|t}^{-1} \underline{\mathbf{m}}_j\right)^\odot \in \mathbb{R}^{4\times 6} \tag{25}$$

one gets to the update equations to conclude the VI SLAM algorithm:

$$
\begin{aligned}
K_{t+1} &= \Sigma_{t+1|t} H_{t+1}^{\top} \left( H_{t+1} \Sigma_{t+1|t} H_{t+1}^{\top} + I \otimes V \right)^{-1} \\
\mu_{t+1|t+1} &= \mu_{t+1|t} \exp\left( \left( K_{t+1} \left( \mathbf{z}_{t+1} - \tilde{\mathbf{z}}_{t+1} \right) \right)^{\wedge} \right) \\
\Sigma_{t+1|t+1} &= \left( I - K_{t+1} H_{t+1} \right) \Sigma_{t+1|t}
\end{aligned}
\qquad
H_{t+1} = \begin{bmatrix} H_{t+1,1} \\ \vdots \\ H_{t+1,N_{t+1}} \end{bmatrix}
\tag{26}
$$

.

At this point the IMU Localization has been updated and the algorithm is complete.

# 4 Results

In this part, the results of the SLAM will be presented.

## 4.1 Prediction

Figure 2 shows how the trajectories of the two different datasets were predicted using the approach presented in the technical approach part. When comparing to the video of the trajectories given on Piazza, it seems the predictor did pretty well. The figure also shows the orientation of the IMU along the path it predicts.

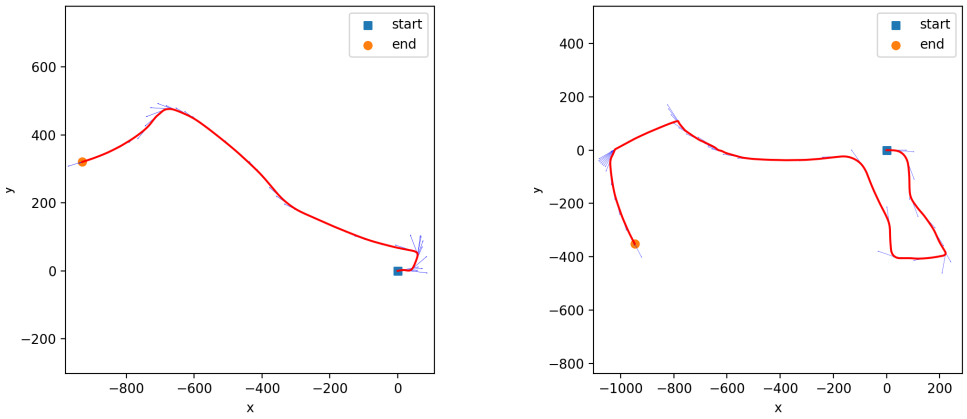


Figure 2: Trajectories from prediction on datasets 03 and 10.

### 4.2 Visual Mapping

After implementing as presented in the technical approach part, the mapping of the landmarks were as follows in Figure 3 for the two different datasets. One observes that the landmarks closely follow the trajectory of the car, which indicates that the mapping was done successfully. These results were done after downsampling the features to reduce computing time, and for dataset 3 every tenth feature is used, while for dataset 10 every 30th feature is used.

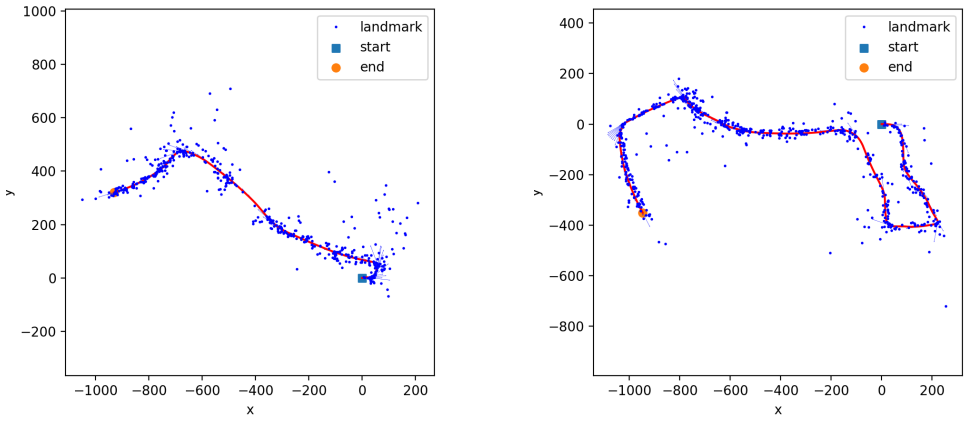


Figure 3: Trajectories and landmark mapping on datasets 03 and 10.

### 4.3 Visual-Inertial SLAM

Figure 4 shows the combined output for the two datasets for the whole visual-inertial SLAM. As one observes, the output is not as it is supposed to be, as it is not nearly as close to the predicted trajectory as it should be. This points to a small error in the implementation of the two previous parts that made the plots not be as intended. The actual plots would probably be a lot closer to the predicted trajectory, and would have corrected what was predicted wrongfully in the prediction step as the update step would take care of the errors. Even though it is clear that something is wrong here, one can observe that the landmarks follow the path reasonably closely, so if the error in the update would have been fixed, the result would be pretty good. Unfortunately, the time was a constraint here and the error was not able to be fixed in time.

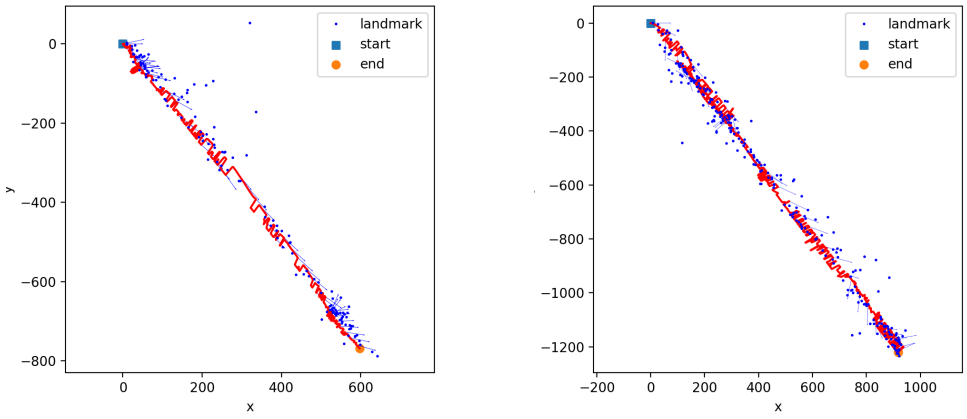


Figure 4: VI SLAM on datasets 03 and 10.