ECE 276A Project 3: Visual Inertial SLAM

1st Yifan Wu
UCSD ECE
yiw084@ucsd.edu

I. Introduction

Simultaneous localization and mapping (SLAM) is the computational problem of constructing or updating a map of an unknown environment while simultaneously keeping track of an robot's location within the map. It is initially a chicken-and-egg problem whether mapping comes first or localization comes first. Current State of Art solutions includes the particle filter, extended Kalman filter and Graph SLAM. [1]

In this project, we implement the solutions for SLAM based on extended Kalman filter and landmark-based visual mapping. Extend Kalman filter is a nonlinear Bayes filter which forced the predicted and updated pdfs to be Gaussian by evaluating their first and second moments and approximating them with Gaussian with the same moments. IMU and stereo camera measurement are collected to estimate pose trajectory and landmarks position.

II. Problem Formulation

A. Visual Mapping (Observation Model)

Consider the mapping-only problem, we assume the IMU post $T_t \in SE(3)$ from the IMU to the camera optical frame and the stereo camera calibration matrix M are known. we would like yo estimate the corresponding 3D coordinate from the pixel coordinate. Since the sensor does not move sufficiently along the z-axis, the estimation of the z coordinate of the landmarks will not be very good. Therefore, we assume that the z coordinates for all landmarks are 0 and focus only on estimating their xy coordinates.

Given the visual feature extracted by the stereo camera:

$$z_t := \begin{bmatrix} z_{t,1}^T & \dots & z_{t,N_t}^T \end{bmatrix}^T \in \mathbb{R}^{4N_t} \tag{1}$$

where each $z_{t,i}$ contain the pixel coordinate $[u_L,\, v_l,\, u_R,\, v_R]^T$ from both left and right cameras. For t = 0, ..., T, estimate the coordinates of landmarks in world frame:

$$m := \begin{bmatrix} m_1^T & \dots & m_M^T \end{bmatrix}^T \in \mathbb{R}^{3M} \tag{2}$$

In order to simplify the problem, we assume the data association:

$$\Delta_t : \{1, ..., M\} \longrightarrow \{1, ..., N_t\} \tag{3}$$

stipulating that landmark j corresponds to observation $z_{t,i} \in \mathbb{R}^4$ with $i = \Delta_t(j)$ at time t is known or provided by an external algorithm. If the $i_t h$ landmark is not observed at time t, the $i_t h$ column of the observations z_t will be $[-1, -1, -1, -1]^T$

Since the landmarks m_i are static, it is not necessary to consider a motion model or prediction step for landmarks.

1) Landmark Initialization:

Given the robot pose T_t and a series of observations z_t, if the landmark is first-time observed, we need to initiate landmarks position by converting the observation in pixel to to world frame coordinate as follow:

$$\begin{bmatrix} x_o \\ y_o \\ z_o \end{bmatrix} = {}_{cam}T_{imu} \cdot {}_{imu}T_{world} m \tag{4}$$

where x_o, y_o, z_o is the position of landmarks in optical frame.

Observation Model with measurement noise $v_{t,i} \sim \mathcal{N}(0,\, V)$

$$\begin{aligned} z_{t,i} &= h(T_t, m_j) + v_{t,i} \\ &= M\pi(T_I T_t^{-1} \underline{m}_j) + v_{t,i} \end{aligned} \tag{5}$$

2) EKF Update:

If the landmark has already been seen, update the landmark position via EKF

Prior: $m|z_{0:t} \sim \mathcal{N}(\mu_t,\, \Sigma_t)$ with $\mu_t \in \mathbb{R}^{3M}$ and $\Sigma_t \in \mathbb{R}^{3M \times 3M}$

Given a new observation $z_{t+1} \in \mathbb{R}^{4N_{t+1}}$:

$$K_{t+1} = \Sigma_t H_{t+1}^T (H_{t+1} \Sigma_t H_{t+1}^T + I \otimes V)^{-1} \tag{6}$$

$$\mu_{t+1} = \mu_t + K_{t+1}(z_{t+1} - M\pi(T_I T_{t+1}^{-1} \underline{\mu}_t)) \tag{7}$$

$$\Sigma_{t+1} = (I - K_{t+1} H_{t+1}) \Sigma_t \tag{8}$$

Where K is Kalman Gain and H is the observation model Jacobian with respect to m_j evaluated at $\mu_{t,j}$

B. IMU-based Localization (Motion Model)

Considering the localization-only problem, we will simplify the prediction step by using kinematic rather than dynamic equations.

EKF prediction step based on the $SE(3)$ kinematics and control input, $u_t = [v_t\ w_t]^T \in \mathbb{R}^6$ with the linear velocity, $v_t \in \mathbb{R}^3$ and angular velocity $\omega_t \in \mathbb{R}^3$ are performed to estimate the pose $T_t \in SE(3)$ of the IMU over time t from IMU frame to world frame.

Prior: $T_t|z_{0:t}, u_{0:t-1} \sim \mathcal{N}(\mu_{t|t}, \Sigma_{t|t})$ with $\mu_{t|t} \in SE(3)$ and $\Sigma_{t|t} \in \mathbb{R}^{6\times 6}$. In discrete-time with discretization τ, motion model becomes:

$$\mu_{t+1|t} = \mu_{t|t} \cdot exp(\tau \hat{u}_t) \tag{9}$$

$$\delta\mu_{t+1|t} = exp(-\tau u_t + w_t) \tag{10}$$

where motion noise $w_t \sim \mathcal{N}(0,\, W)$

1) EKF Prediction Step:

$$\mu_{t+1|t} = \mu_{t|t} \cdot exp(\tau \hat{u_t}) \tag{11}$$

$$\Sigma_{t+1|t} = exp(-\tau u_t) \cdot \Sigma_{t+1|t} \cdot exp(-\tau u_t)^T + W \tag{12}$$

where motion noise $w_t \sim \mathcal{N}(0, W)$ which is the Gaussian distribution around the predicted value. τ is the absolute time difference between the current timestamps and the previous timestamps.

$$\hat{u_t} := \begin{bmatrix} \hat{\omega_t} & v_t \\ 0^T & 0 \end{bmatrix} \mathbb{R}^{4\times 4} \tag{13}$$

$$u_t^\curlywedge := \begin{bmatrix} \hat{\omega_t} & \hat{v_t} \\ 0 & \hat{\omega_t} \end{bmatrix} \mathbb{R}^{6\times 6} \tag{14}$$

2) EKF Update Step:
Assuming the position of landmarks are known, update pose mean and covariance. The observation model is the same as in the visual mapping problem, but the variable of interest is the IMU pose $T_{t+1} \in SE(3)$ instead of the landmark positions $m \in \mathbb{R}^3$. Therefore, we need 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 $\mu_{t+1|t}$.

$$K_{t+1} = \Sigma_t H_{t+1}^T (H_{t+1} \Sigma_t H_{t+1}^T + I \otimes V)^{-1} \tag{15}$$

$$\mu_{t+1|t+1} = \mu_{t+1|t}\, exp((K_{t+1}(z_{t+1} - \tilde{z}_{t+1}))^\wedge) \tag{16}$$

$$\Sigma_{t+1|t+1} = (I - K_{t+1} H_{t+1}) \Sigma_t \tag{17}$$

where $\tilde{z}_{t+1}$ is the predicted observation.

C. SLAM

Consider localization and mapping simultaneously, neither the pose T_t or landmark positions m is known. To achieve the goal of estimating position of landmarks and pose of the robot at the same time, the idea is to merge the mean and covariance of IMU Pose and landmarks position:

$$\mu = \begin{bmatrix} \mu_L \\ \mu_{IMU} \end{bmatrix} \in \mathbb{R}^{3M+16} \tag{18}$$

$$\Sigma = \begin{bmatrix} \Sigma_L & C \\ C^T & \Sigma_{IMU} \end{bmatrix} \in \mathbb{R}^{3M+6\times 3M+6} \tag{19}$$

where C is the cross covariance μ_L is the mean of landmarks, μ_{IMU} is the mean of IMU Pose, Σ_L is the covariance of landmarks, Σ_{IMU} is the covariance of IMU Pose. Since the covariance of IMU pose and covariance of landmarks positions are merged into on joint covariance, IMU pose and landmarks positions will become correlated as we perform update.

The prediction step of motion model is same as before, but in update step. There is a single Jacobian Matrix calculated to update both IMU pose and Landmark positions.

$$H_{t+1|t} = [H_{L,t+1|t}\, H_{imu,t+1|t}] \in \mathbb{R}^{4N_t \times (3M+6)} \tag{20}$$

EKF Update:

$$K = \begin{bmatrix} \Sigma_L & C \\ C^T & \Sigma_{IMU} \end{bmatrix} \begin{bmatrix} H_L^T \\ H_{IMU}^T \end{bmatrix} S^{-1} \tag{21}$$

$$K_{t+1} = \Sigma_t H_{t+1}^T (H_{t+1} \Sigma_t H_{t+1}^T + I \otimes V)^{-1} \tag{22}$$

$$\mu_{t+1|t+1} = \begin{bmatrix} \mu_{L,t+1|t} + K_{t+1}(z_{t+1} - \tilde{z}_{t+1}) \\ \mu_{IMU,t+1|t}\, exp((K_{t+1}(z_{t+1} - \tilde{z}_{t+1})^\wedge) \end{bmatrix} \tag{23}$$

$$\Sigma_{t+1|t+1} = (I - K_{t+1} H_{t+1}) \Sigma_{t+1|t} \tag{24}$$

III. Technical Approach

In order to perform robust pose and landmarks position estimation, we will use Extend Kalman Filter to effectively use observation to update or correct our previous estimation.

A. Extended Kalman Filter

Extended Kalman Filter is a Bayes filter with the following assumptions:

- The prior pdf $p_{t|t}$ is Gaussian
- The motion model has Gaussian noise w_t

$$x_{t+1} = f(x_t, u_t, w_t),\ w_t \sim \mathcal{N}(0,\ W) \tag{25}$$

- The observation model has Gaussian noise v_t

$$z_{t+1} = h(x_t, w_t),\ v_t \sim \mathcal{N}(0,\ V) \tag{26}$$

- The motion noise w_t and observation noise v_t are independent of each other, of the state x_t, and across time

The challenge of non-linear Kalman Filter is that the predicted and updated pdfs are not Gaussian and can no longer be evaluated in closed form. It using moment matching which force the predicted and updated pdfs to be Gaussian by evaluating their first and second moments and approximating them with Gaussians with the same moments.

Extended Kalman Filter uses a first-order Taylor series approximation to the motion models around the state and noise means:

$$\begin{aligned} & f(x_t, u_t, w_t) \simeq f(\mu_{t|t}, u_t, 0) \\ & + \left[\tfrac{df}{dx}(\mu_{t|t}, u_t, 0)\right](x_t - \mu_{t|t}) \\ & + \left[\tfrac{df}{dw}(\mu_{t|t}, u_t, 0)\right](w_t - 0) \\ & = f(x_t, u_t, 0) + F_t(x_t - \mu_{t|t}) + Q_t w_t \end{aligned} \tag{27}$$

where

$$\begin{aligned} F_t &= \tfrac{df}{dx}(\mu_{t|t}, u_t, 0) \\ Q_t &= \tfrac{df}{dw}(\mu_{t|t}, u_t, 0) \end{aligned} \tag{28}$$

Extended Kalman Filter uses a first-order Taylor series approximation to the motion observation models around the state and noise means:

$$\begin{aligned} &h(x_{t+1}, v_{t+z}) \simeq h(\mu_{t+1|t}, 0) \\ &+\left[\frac{dh}{dx}(\mu_{t+1|t}, 0)\right](x_{t+1} - \mu_{t+1|t}) \\ &+\left[\frac{dh}{dw}(\mu_{+1t|t}, 0)\right](V_{t+1} - 0) \\ &= h(\mu_{t+1|t}, 0) + H_{t+1}(x_{t+1} - \mu_{t+1|t}) + R_{t+1}V_{t+1} \end{aligned} \tag{29}$$

where

$$\begin{aligned} H_{t+1} &= \tfrac{dh}{dx}(\mu_{t+1|t}, 0) \\ R_{t+1} &= \tfrac{dh}{dw}(\mu_{t+1|t}, 0) \end{aligned} \tag{30}$$

Base on the equation above, the Extended Kalmen Filter will have following equations:

Prior:

$$x_t|z_{0:t}, u_{0:t} \sim \mathcal{N}(\mu_{t|t}, \Sigma_{t|t}) \tag{31}$$

Motion Model:

$$\begin{aligned} x_{t+1} &= f(x_t, u_t, w_t) \\ w_t &\sim \mathcal{N}(0, W) \\ F_t &= \tfrac{df}{dx}(\mu_{t|t}, u_t, 0) \\ Q_t &= \tfrac{df}{dw}(\mu_{t|t}, u_t, 0) \end{aligned} \tag{32}$$

Observation Model:

$$\begin{aligned} z_t &= h(x_t, v_t) \\ v_t &\sim \mathcal{N}(0, V) \\ H_{t+1} &= \tfrac{dh}{dx}(\mu_{t+1|t}, 0) \\ R_{t+1} &= \tfrac{dh}{dw}(\mu_{t+1|t}, 0) \end{aligned} \tag{33}$$

Prediction:

$$\begin{aligned} \mu_{t+1|t} &= f(\mu_{t|t}, u_t, 0) \\ \Sigma_{t+1|t} &= F_t\Sigma_{t|t}F_t^T + Q_tWQ_t^T \end{aligned} \tag{34}$$

Update:

$$\begin{aligned} \mu_{t+1|t+1} &= \mu_{t+1|t} + K_{t+1|t}(z_{t+1} - h(\mu_{t+1|t}, 0)) \\ \Sigma_{t+1|t+1} &= (I - K_{t+1|t}H_{t+1})\Sigma_{t+1|t} \end{aligned} \tag{35}$$

Kalman Gain:

$$K_{t+1|t} = \Sigma_{t+1|t}H_{t+1}(H_{t+1}\Sigma_{t+1|t}H_{t+1}^T + R_{t+1}VR_{t+1}^T)^{-1} \tag{36}$$

B. Landmark Mapping via EKF Update

For the mapping-only, we assume IMU pose $T_t \in SE(3)$ is known and the data association:

$$\Delta_t : \{1, ..., M\} \longrightarrow \{1, ..., N_t\} \tag{37}$$

stipulating that landmark j corresponds to observation $z_{t,i} \in \mathbb{R}^4$ with $i = \Delta_t(j)$ at time t is known or provided by an external algorithm.

Prior: $m|z_{0:t} \sim \mathcal{N}(\mu_t, \Sigma_t)$ with $\mu_t \in \mathbb{R}^{3M}$ and $\Sigma_t \in \mathbb{R}^{3M\times 3M}$

Observation Model with measurement noise $v_{t,i} \sim \mathcal{N}(0, V)$

$$\begin{aligned} z_{t,i} &= h(T_t, m_j) + v_{t,i} \\ &= M\pi(T_IT_t^{-1}\underline{m}_j) + v_{t,i} \end{aligned} \tag{38}$$

To initialize the 3D landmarks position in world frame, we use the projection equation to project back the observation point from stereo camera frame to world frame. The steps are state as following:

- Calculate disparity $d = u_L - u_R$
- Calculate z_o coordinate in optical frame $z_o = \frac{fs_u \cdot b}{d}$
- Calculate x_o and y_o with respect to equation (4)
- Transform the observation in optical frame to to world frame coordinate with respect to equation (5)
- Initialize the landmark prior $\mu_j = \begin{bmatrix} x_w \\ y_w \\ z_w \end{bmatrix}$ and landmark covariance $\Sigma_j = I_{3\times 3}$

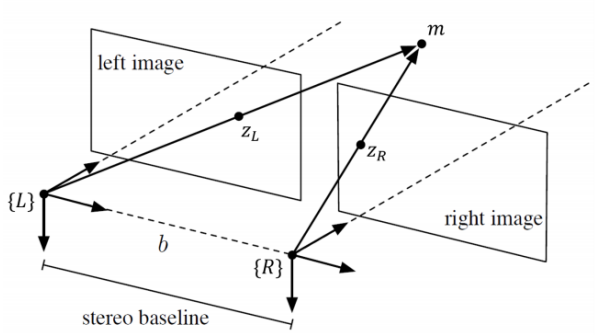

Fig. 1. *Stereo Camera Model*

.

$$\begin{bmatrix} u_L \\ v_L \\ u_L - u_R \end{bmatrix} = \begin{bmatrix} fs_u & 0 & c_u & 0 \\ 0 & fs_v & c_v & 0 \\ 0 & 0 & 0 & fs_ub \end{bmatrix} \frac{1}{z_o} \begin{bmatrix} x_o \\ y_o \\ z_o \\ 1 \end{bmatrix} \tag{39}$$

$$\begin{bmatrix} x_o \\ y_o \\ z_o \end{bmatrix} = {}_{cam}T_{imu} \cdot {}_{imu}T_{world} \begin{bmatrix} x_w \\ y_w \\ z_w \end{bmatrix} \tag{40}$$

where f is the focal length, s_u, s_v are pixel scaling, c_u and c_v are the principle points, and b is the stereo baseline. x_o, y_o, z_o is the position of landmarks in optical frame.

The calibration matrix M:

$$M = \begin{bmatrix} fs_u & 0 & c_u & 0 \\ 0 & fs_v & c_v & 0 \\ 0 & 0 & 0 & fs_u b \end{bmatrix} \tag{41}$$

where f is the focal length, s_u, s_v are pixel scaling, c_u and c_v are the principle points, and b is the stereo baseline.

Homogeneous coordinates:

$$\underline{m}_j = \begin{bmatrix} m_j \\ 1 \end{bmatrix} \tag{42}$$

Projection Function and its derivative:

$$\begin{aligned} \pi(q) &:= \tfrac{1}{q3} q \in \mathbb{R}^4 \\ \tfrac{d\pi}{dq}(q) &= \tfrac{1}{q3} \begin{bmatrix} 1 & 0 & -\frac{q1}{q3} & 0 \\ 0 & 1 & -\frac{q2}{q3} & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -\frac{q4}{q3} & 1 \end{bmatrix} \in \mathbb{R}^{4\times 4} \end{aligned} \tag{43}$$

All observations, stacked as a $4Nt$ vector, at time t with notation abuse:

$$\begin{aligned} z_t &= M\pi(T_I T_t^{-1} \underline{m}) + v_t \\ v_t &\sim \mathcal{N}(0, I \otimes V) \\ I &= \begin{bmatrix} V & & \\ & \dots & \\ & & V \end{bmatrix} \end{aligned} \tag{44}$$

Given a new observation $z_{t+1} \in \mathbb{R}^{4N_{t+1}}$:

$$K_{t+1} = \Sigma_t H_{t+1}^T (H_{t+1} \Sigma_t H_{t+1}^T + I \otimes V)^{-1} \tag{45}$$

$$\mu_{t+1} = \mu_t + K_{t+1}(z_{t+1} - \tilde{z}_{t+1}) \tag{46}$$

$$\Sigma_{t+1} = (I - K_{t+1} H_{t+1}) \Sigma_t \tag{47}$$

where $\tilde{z}_{t+1} \in \mathbb{R}^{4N_{t+1}}$ is the predicted observation based on the landmark position estimates μ_t at time t

$$\tilde{z}_{t+1} = M\pi(T_I T_{t+1}^{-1} \underline{\mu}_t) \tag{48}$$

We need the observation model Jacobian $H_{t+1} \in \mathbb{R}^{4N_t \times 3M}$ evaluate at μ_t with block elements $H_{t+1,i,j} \in \mathbb{R}^{4\times 3}$:

$$H_{t+1,i,j} = \begin{cases} \frac{d}{dm_j} h(T_{t+1}, m_j) | m_j = \mu_{t,j} & if \delta_t(j) = i \\ 0 & otherwise \end{cases} \tag{49}$$

Here $H_{t+1,i,j}$ represents the ith observation in time t corresponding to jth landmarks. The entire H_{t+1} is a block matrix which fills with $H_{t+1,i,j} \in \mathbb{R}^{4\times 3}$. Since we assume all the landmarks are static, there is no need to perform EKF prediction step for mapping.

C. IMU-based Localization via EKF Prediction

The localization problem aims to estimate the pose $T_t \in SE(3)$ of the robot over time. Motion Model for the continuous-time IMU pose T(t) with noise w(t):

$$\dot{T} = T(\hat{u} + \hat{w}) \tag{50}$$

where $u_t = [v_t \ w_t]^T \in \mathbb{R}^6$ with the linear velocity, $v_t \in \mathbb{R}^3$ and angular velocity $\omega_t \in \mathbb{R}^3$

Using $T = \mu exp(\delta\hat{\mu}) \approx \mu(I + \delta\hat{\mu})$, the above equation can be split into nominal and perturbation kinematics:

- Nominal: $\dot{\mu} = \mu\hat{u}$
- Perturbation: $\delta\dot{\hat{\mu}} = -\hat{u}\delta\mu + w$

In discrete-time with discretization τ, motion model becomes:

$$\mu_{t+1|t} = \mu_{t|t} \cdot exp(\tau\hat{u}_t) \tag{51}$$

$$\delta\mu_{t+1|t} = exp(-\tau u_t + w_t) \tag{52}$$

where motion noise $w_t \sim \mathcal{N}(0, W)$.

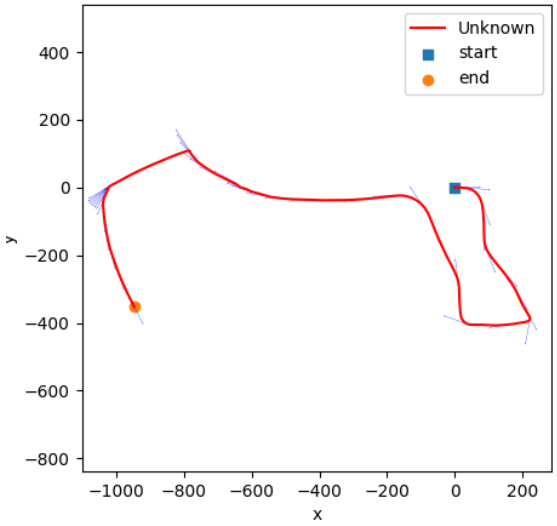

Fig. 2. *Dead Reckoning*

1) EKF Prediction Step(Dead Reckoning):

Prior: $T_t | z_{0:t}, u_{0:t-1} \sim \mathcal{N}(\mu_{t|t}, \Sigma_{t|t})$ with $\mu_{t|t} \in SE(3)$ and $\Sigma_{t|t} \in \mathbb{R}^{6\times 6}$.

$$\mu_{t+1|t} = \mu_{t|t} \cdot exp(\tau\hat{u}_t) \tag{53}$$

$$\Sigma_{t+1|t} = exp(-\tau u_t^{\curlywedge}) \cdot \Sigma_{t+1|t} \cdot exp(-\tau u_t^{\curlywedge})^T + W \tag{54}$$

where motion noise $w_t \sim \mathcal{N}(0, W)$

$$\hat{u_t} := \begin{bmatrix} \hat{\omega_t} & v_t \\ 0^T & 0 \end{bmatrix} \mathbb{R}^{4\times 4} \tag{55}$$

$$u_t^{\curlywedge} := \begin{bmatrix} \hat{\omega_t} & \hat{v_t} \\ 0 & \hat{\omega_t} \end{bmatrix} \mathbb{R}^{6\times 6} \tag{56}$$

2) EKF Update Step:

Observation Model with measurement noise $v_{t,i} \sim \mathcal{N}(0, V)$:

$$z_{t,i} = h(T_t, m_j) + v_{t,i} = M\pi(T_I T_t^{-1} \underline{m}_j) + v_{t,i} \tag{57}$$

The observation model is the same as in the visual mapping problem, but the variable of interest is the IMU pose $T_{t+1} \in SE(3)$ instead of the landmark positions $m \in \mathbb{R}^3$. Therefore, we need 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 $\mu_{t+1|t}$.

The first-order Taylor series approximation of observation i at time $t+1$ using an IMU pose perturbation $\delta\mu$ is:

$$\begin{aligned} z_{t+1,i} &\approx M\pi(T_I \mu_{t+1|t}^{-1} \underline{m}_j) \\ &- M \tfrac{d\pi}{dq}(T_I \mu_{t+1|t}^{-1} \underline{m}_j) T_I (\mu_{t+1|t}^{-1} \underline{m}_j)^{\odot} \delta\mu \\ &+ v_{t+1,i} \end{aligned} \tag{58}$$

where for homogeneous coordinate $\underline{s} \in \mathbb{R}^5$ and $\hat{\xi} \in se(3)$:

$$\begin{aligned} \hat{\xi}\underline{s} &= \underline{s}^{\odot}\xi \\ \underline{s} &:= \begin{bmatrix} I & -\hat{s} \\ 0 & 0 \end{bmatrix} \end{aligned} \tag{59}$$

Jacobian of $z_{t+1,i}$ with respect to T_{t+1} evaluated at $\mu_{t+1|t}$:

$$H_{t+1,i} = -M \tfrac{d\pi}{dq}(T_I \mu_{t+1|t}^{-1} \underline{m}_j) T_I (\mu_{t+1|t}^{-1} \underline{m}_j)^{\odot} \in \mathbb{R}^{4\times 6}$$

$$H_{t+1} = \begin{bmatrix} H_{t+1,1} \\ . \\ . \\ H_{t+1,N_{t+1}} \end{bmatrix} \tag{60}$$

Prior: $T_{t+1}|z_{0:t}, u_{0:t} \sim \mathcal{N}(\mu_{t+1|t}, \Sigma_{t+1|t})$ with $\mu_{t+1|t} \in SE(3)$ and $\Sigma_{t+1|t} \in \mathbb{R}^{6\times 6}$.

$$K_{t+1} = \Sigma_t H_{t+1}^T (H_{t+1} \Sigma_t H_{t+1}^T + I \otimes V)^{-1} \tag{61}$$

$$\mu_{t+1|t+1} = \mu_{t+1|t}\, exp((K_{t+1}(z_{t+1} - \tilde{z}_{t+1}))^{\wedge}) \tag{62}$$

$$\Sigma_{t+1|t+1} = (I - K_{t+1} H_{t+1}) \Sigma_{t+1|t} \tag{63}$$

D. SLAM

Consider localization and mapping simultaneously, neither the pose T_t or landmark positions m is known. To achieve the goal of estimating position of landmarks and pose of the robot at the same time, the idea is to merge the mean and covariance of IMU Pose and landmarks position:

$$\mu = \begin{bmatrix} \mu_L \\ \mu_{IMU} \end{bmatrix} \in \mathbb{R}^{3M+16} \tag{64}$$

$$\Sigma = \begin{bmatrix} \Sigma_L & C \\ C^T & \Sigma_{IMU} \end{bmatrix} \in \mathbb{R}^{3M+6 \times 3M+6} \tag{65}$$

where μ_L is the mean of landmarks, μ_{IMU} is the mean of IMU Pose, Σ_L is the covariance of landmarks, Σ_{IMU} is the covariance of IMU Pose, C is the cross-covariance. The prediction step of motion model is same as before, but in update step.

$$\mu_{t+1|t} = \begin{bmatrix} \mu_{L,t+1|t} \\ \mu_{IMU,t+1|t} \end{bmatrix} = \begin{bmatrix} \mu_{L,t+1|t} \\ exp(-\tau\hat{u}_t)\mu_{IMU,t|t} \end{bmatrix} \tag{66}$$

$$\begin{aligned} \Sigma_{t+1|t} &= F_t \Sigma_{t|t} F_t^T + W \\ &= \begin{bmatrix} I & 0 \\ 0 & \tilde{F}_t \end{bmatrix} \begin{bmatrix} \Sigma_L & C \\ C^T & \Sigma_{IMU} \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & \tilde{F}_t^T \end{bmatrix} + W \\ &= \begin{bmatrix} \Sigma_L & \tilde{F}_t C \\ C^T \tilde{F}_t^T & \Sigma_{IMU} \end{bmatrix} + W \end{aligned} \tag{67}$$

where $\tilde{F}_t$ is the perturbation, $W_p \in \mathbb{R}^{6\times 6}$ is the process noise.

$$\tilde{F}_t = exp(-\tau u^{\curlywedge}) \tag{68}$$

$$W = \begin{bmatrix} 0 & 0 \\ 0 & W_p \end{bmatrix} \tag{69}$$

There is a single Jacobian Matrix calculated to update both IMU pose and Landmark positions.

$$H_{t+1|t} = [H_{L,t+1|t}\, H_{imu,t+1|t}] \in \mathbb{R}^{4N_t \times (3M+6)} \tag{70}$$

EKF Update:

$$K_{t+1} = \Sigma_t H_{t+1}^T (H_{t+1} \Sigma_t H_{t+1}^T + I \otimes V)^{-1} \tag{71}$$

$$\mu_{t+1|t+1} = \begin{bmatrix} \mu_{L,t+1|t} + K_{t+1}(z_{t+1} - \tilde{z}_{t+1}) \\ \mu_{IMU,t+1|t}\, exp((K_{t+1}(z_{t+1} - \tilde{z}_{t+1})^{\wedge}) \end{bmatrix} \tag{72}$$

$$\Sigma_{t+1|t+1} = (I - K_{t+1} H_{t+1}) \Sigma_{t+1|t} \tag{73}$$

IV. Result

The EKF Visual Inertial SLAM algorithm is performed in dataset 10.npy which is collected by KITTI-360 from real-world driving scenario. Due to the limit of computational power of test machine, a subset of representative feature are selected from the original feature. We use 33 percent and 50 percent of feature to test our algorithm. Since there is only one sample of the real-world driving scenario, the hyper-parameter might not be optimal for other case. The detail of hyper-parameter can be found in Table 1.

TABLE I
HYPER-PARAMETERS FOR EKF SLAM

Parameter	Description	Value
Σ_L	Prior landmark covariance	$0.5 \cdot I_{3\times 3}$
Σ_{IMU}	Prior IMU covariance	$0.01 \cdot I_{6\times 6}$
V	Observation noise covariance	$100 \cdot I_{4N_t \times 4N_t}$
W	Process noise covariance	$1e-10 \cdot I_{6\times 6}$

The result of IMU Prediction step (Dead Reckoning) is shown in Figure 2 in section III (TECHNICAL APPROACH). The red line is the predicted trajectory over time t form IMU

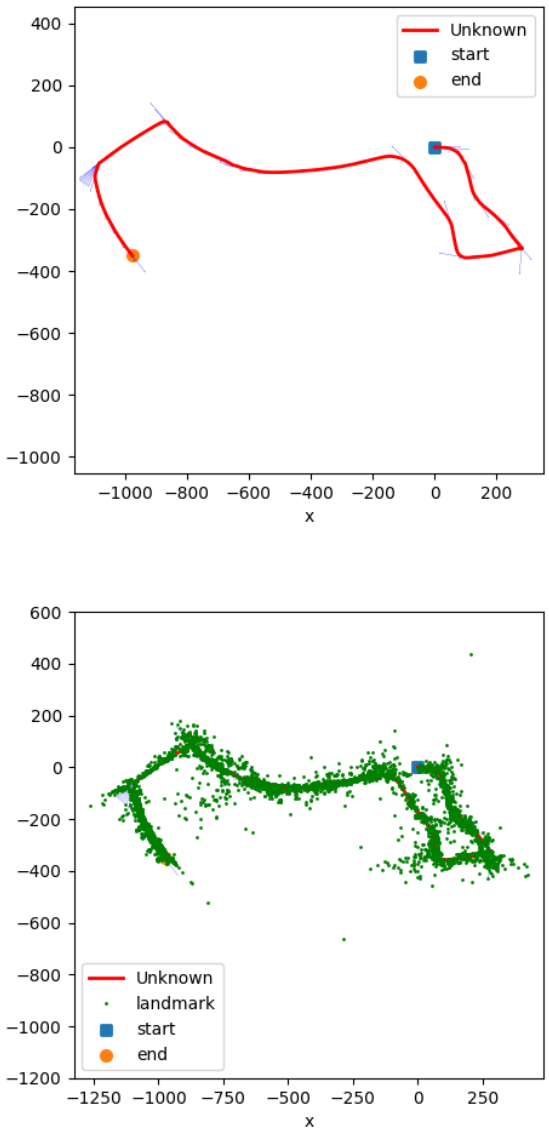

Fig. 3. *Visual-Inertial SLAM with 33% features*

pose $T_t \in SE(3)$, where the blue point is the starting position and orange point is the end position.

If we separate the mapping and localization, the EKF Visual Inertial SLAM algorithm did work well, but it is less accurate since we lost the correlation between IMU pose and landmarks positions, which essentially giving the assumption that the joint covariance can be written as:

$$\Sigma = \begin{bmatrix} \Sigma_L & 0 \\ 0 & \Sigma_{IMU} \end{bmatrix} \in \mathbb{R}^{3M+6 \times 3M+6} \tag{74}$$

Where cross covariance C is zero. It updates Σ_{IMU} and Σ_L independently.

In parctice, C can never be zero. Since the covariance of IMU pose and covariance of landmarks positions are merged into on joint covariance, IMU pose and landmarks positions will become correlated as we perform update. Therefore, it is necessary to keep the covaiance as a full $(3M+6)\times(3M+6)$ large matrix. It is more accurate then perform the block matrix calculation in Kalmen Gain. However, the trade-off is the computational speed and memory usage. The result of SLAM is shown in Figure 3.

$$K = \begin{bmatrix} \Sigma_L & C \\ C^T & \Sigma_{IMU} \end{bmatrix} \begin{bmatrix} H_m^T \\ H_I^T \end{bmatrix} S^{-1} \tag{75}$$

Since in SLAM problem, the robot doesn't have priori knowledge of the map or trajectories, It is very sensitive to noise, choosing the process noise covariance is also important. For the observation noise, we are assuming observation has 10 pixels standard deviation. Figure 4 shows that, as a larger

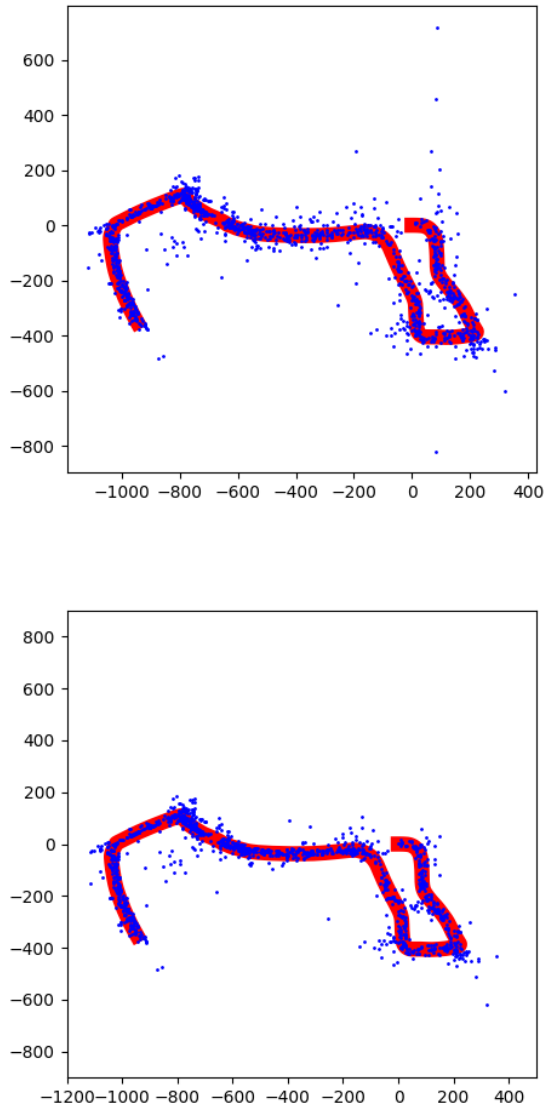

Fig. 4. *Affect of Observation Noise*

observation noise is added, landmarks points in the map is more concentrated. The resulting path will look more like the dead-reckoning, if the motion noise is small. Other other hand increasing W by a little bit will cause the SLAM result to diverge from the dead-reckoning dramatically.

In the joint update step in EKF Visual Inertial SLAM, a better prior assumption or estimation on the corvariance matrix is crucial. Figure 5 and Figure 6 demonstrate the affect of prior joint covariance. Therefore, initialization of Σ_L and Σ_{IMU} in joint covariance is important. Since our imu sensor is accurate, we set Σ_{IMU} to be $0.01 \cdot I_{6\times 6}$ which means we trust more on IMU.

Alternative approach can be using Bayes's smoothing to track all history of state instead of Bayes's filtering.

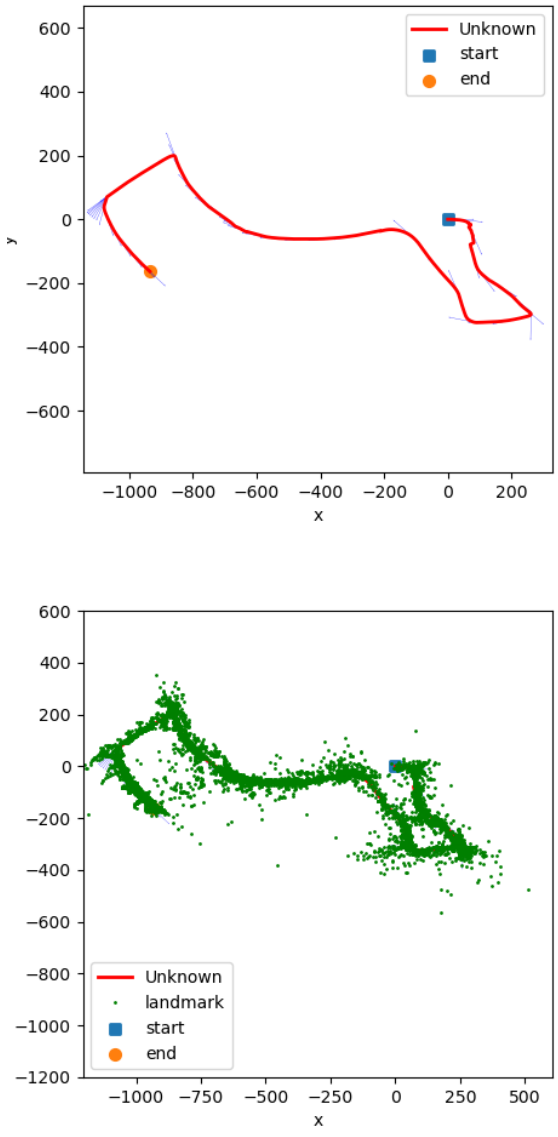

Fig. 5. *Visual-Inertial SLAM with 50% features, Large Prior COV*

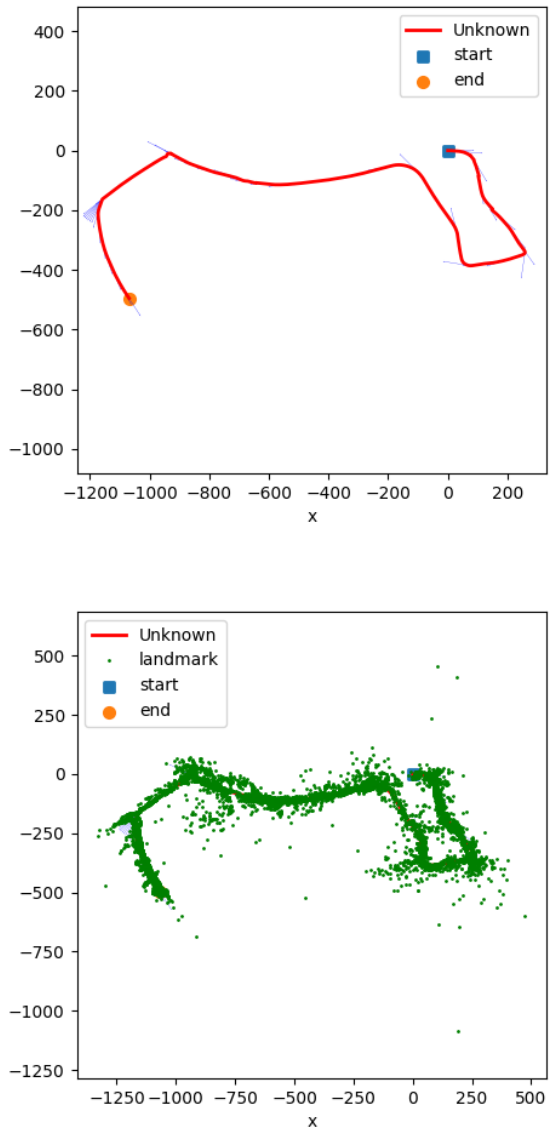

Fig. 6. *Visual-Inertial SLAM with 50% features, Small Prior COV*

REFERENCES

[1] M. Montemerlo, S. Thrun, D. Koller, B. Wegbreit *et al.*, "Fastslam: A factored solution to the simultaneous localization and mapping problem," *Aaai/iaai*, vol. 593598, 2002.