Visual-Inertial SLAM

Hisu-Wen Yen A59010599 hsyen@ucsd.edu

I. Introduction

SLAM technique is used to build the map around robot and to localize robot's position, and map is very important element in robotics. If we have map, then we can require robot to perform some specific tasks. However, most of the time the map is unknown, then we need to use SLAM to discover the map. In this report contains how to use extended Kalman filter to perform SLAM.

II. Problem Formulation

In this problem, we use notation z as observation, u as control input and x as robot state.

A. IMU Localization via EKF Prediction

Given IMU measurements $u_{0:T}$ and feature observations $z_{0:T}$, we want to estimate the pose of robot T_t over time.

$$u_t = \begin{bmatrix} v_t \\ \omega_t \end{bmatrix} \in R^6 \tag{1}$$

$$T_t =_{world} T_{IMU},_t \in SE(3) \tag{2}$$

We can obtain general velocity from IMU directly, so we don't have motion model to estimate v_{t+1} and ω_{t+1}. Instead, we use kinematic equation to estimate pose T_{t+1}.

We also need to know how to apply gaussian distribution over T_t. This is because in EKF, we force everything into gaussian distribution and apply linear model, so everything will keep in gaussian family. Then, we can use Kalman filter prediction step to compute $\mu_{t+1|t}$ and $\Sigma_{t+1|t}$.

The problem becomes compute

$$T_{t+1|z_{0:t+1},u_{0:t}} \sim N\left(\mu_{t+1|t}, \Sigma_{t+1|t}\right) \tag{3}$$

$$\mu_{t+1|t} \in SE(3) \tag{4}$$

$$\Sigma_{t+1|t} \in R^6 \tag{5}$$

With prior

$$T_{t|z_{0:t},u_{0:t-1}} \sim N\left(\mu_{t|t}, \Sigma_{t|t}\right) \tag{6}$$

B. Landmark Mapping via EKF Update

We also want to build map of all landmarks. Assume our trajectory on part A is correct, then we want to estimate position of landmarks via observation model.

Given T_t, we can use observation model to compute position of landmark m_i

$$m_i = \begin{bmatrix} x_i \\ y_i \\ z_i \end{bmatrix} \in R^3 \tag{7}$$

$$\begin{aligned} z_{t,i} &= h(T_t, m_i) + v_{t,i} \\ &= K_s \pi({}_oT_I T_t^{-1} \underline{m_i}) + v_{t,i} \end{aligned} \tag{8}$$

$v_{t,i}$ is noise of measurement

$$\underline{m_i} = \begin{bmatrix} x_i \\ y_i \\ z_i \\ 1 \end{bmatrix} \in R^4 \tag{9}$$

We are not 100% sure where the landmarks, so we use gaussian distribution to represent its probability.

The problem becomes compute

$$p(m_i | T_t, z_{t,i}) \text{ over time} \tag{10}$$

C. Visual-Inertial SLAM

We are not 100% sure about robot pose T_t and landmark position m_i in reality. Therefore, we want a gaussian distribution for all term. The mean and covariance for this distribution are

$$mean = \begin{pmatrix} \mu_{s,t} \\ \mu_{m,t} \end{pmatrix} \tag{11}$$

$$covariance \in R^{(3M+6)*(3M+6)} \tag{12}$$

M is number of all landmarks

$\mu_{s,t}$ is mean for robot pose

$$\mu_{s,t} \in SE(3) \tag{13}$$

$\mu_{m,t}$ is mean for stacked landmark

$$\mu_{m,t} \in R^{3M} \tag{14}$$

$covariance$ is (3M+6) by (3M+6), first 3M by 3M blocks are covariance of landmarks. Last 6 term are covariance w.r.t robot pose. First 3 are for position, last 3 are for orientation.

The problem becomes compute the mixed $mean$ and $covariance$ over time.

III. Thectival Approches

A. IMU Localization via EKF Prediction

First, we need to solve what is motion model for pose T_t.

$$T_t \in SE(3) = \begin{bmatrix} R_t & P_t \\ 0^T & 1 \end{bmatrix} \in R^{4*4} \mid R_t \in SO(3), P_t \in R^3 \quad (15)$$

We can start from R_t, then extend it to T_t.

We can map $SO(3)$ to $so(3)$ as

$$so(3) = \{\hat{\theta} \in R^{3*3} | \theta \in R^3\} \quad (16)$$

$$R = \exp(\hat{\theta}) \quad (17)$$

We can derive kinematics from its property.

$$R^T(t)R(t) = I \quad (18)$$

$$\dot{R}^T(t)R(t) + R^T(t)R(\dot{t}) = 0 \quad (19)$$

$R^T(t)R(\dot{t})$ is skew-symmetric.

$$R^T(t)R(\dot{t}) = \hat{\omega}(t) \quad (20)$$

$$R(\dot{t}) = R(t)\hat{\omega}(t) \quad (21)$$

Now, we can find first order approximation to rotation matrix.

$$R(t + dt) \approx R(t) + R(t)\hat{\omega}(t)dt \quad (22)$$

Assuming ω is constant over a short period τ :

$$R(t + dt) = R(t)\exp(\tau\hat{\omega}) \quad (23)$$

With (23) we can get discrete rotation kinematics:

$$R_{k+1} = R_k \exp(\tau\hat{\omega}) \quad (24)$$

Second problem here is how is gaussian distribution work for rotation matrix. We can use the idea of perturbation to achieve it.

There are two ways to add perturbation on rotation matrix, one is in $so(3)$, another is in $SO(3)$.

Perturbation in $so(3)$: $\exp((\theta + \delta\theta)^\wedge)$ (25)

Perturbation in $SO(3)$: $\mathrm{R}\exp(\widehat{\delta\varphi})$ or $\exp(\widehat{\delta\varphi})\mathrm{R}$ (26)

We can define a gaussian distribution over a rotation matrix R:

$$R = \exp(\hat{\epsilon})\mu, \ \epsilon \sim N(0, \Sigma) \quad (27)$$

$$\mu \in SO(3) \quad (28)$$

$\epsilon \in R^3$ is a zero-mean gaussian random vector (29)

Let

$$Y = QR = Q\exp(\hat{\epsilon})\mu = \exp((Q\epsilon)^\wedge)Q\mu \quad (30)$$

$$E[Y] = Q\mu \quad (31)$$

$$Var[Y] = Var[Q\epsilon] = Q\Sigma Q^T \quad (32)$$

Now, we can apply similar idea into pose $SE(3)$

We can map pose matrix $T \in SE(3)$ by a position-rotation vector ξ.

$$\xi = \begin{bmatrix} \rho \\ \theta \end{bmatrix} \in R^6 \quad (33)$$

$$T = \exp(\hat{\xi}) \quad (34)$$

$$\hat{\xi} = \begin{bmatrix} \hat{\theta} & \rho \\ 0 & 0 \end{bmatrix} \in R^{4*4} \quad (35)$$

Similar to R, we can derive discrete-time pose kinematics:

$$\varsigma(t) = \begin{bmatrix} v(t) \\ \omega(t) \end{bmatrix} \in R^6 \quad (36)$$

$$T_{k+1} = T_k \exp(\tau\hat{\varsigma}) \quad (37)$$

To consider a gaussian distribution over T, we utilize small perturbation $\widehat{\delta\mu} \in se(3)$ over $\mu \in SE(3)$:

$$T = \mu\exp(\widehat{\delta\mu}) \approx \mu(I + \widehat{\delta\mu}) \quad (38)$$

After detailed derivation we can get discrete-time perturbation:

Nominal : $\mu_{t+1} = \mu_t \exp(\tau\widehat{u_t})$ (39)

Perturbation: $\delta\mu_{t+1} = \exp(-\tau\, \overset{\curlywedge}{u_t})\, \delta\mu_t + w_t$ (40)

$$\overset{\curlywedge}{u_t} = \begin{bmatrix} \hat{\omega} & \hat{v} \\ 0 & \hat{\omega} \end{bmatrix} \in R^{6*6} \quad (41)$$

$w_t \sim N(0, W)$ is noise for motion.

Now, we can apply EKF Prediction Step to estimate $\mu_{t+1|t}$ and $\Sigma_{t+1|t}$:

$$\mu_{t+1|t} = \mu_t \exp(\tau\widehat{u_t}) \quad (42)$$

$$\Sigma_{t+1|t} = \exp(-\tau\, \overset{\curlywedge}{u_t})\Sigma_{t|t} \exp(-\tau\, \overset{\curlywedge}{u_t})^T + W \quad (43)$$

B. Landmark Mapping via EKF Update

To compute $p(m_i|T_t, z_{t,i})$ over time, we utilize observation model. In this part, we assume that landmarks are independent; therefore, each landmark m_i has its own mean $\mu_{t,i}$ and covariance $\Sigma_{t,i}$.

However, we didn't get any prior information about landmark. We use first observation of each landmark to create a prior.

From camera model we know that

$$\begin{bmatrix} u_L \\ v_L \\ d \end{bmatrix} = \begin{bmatrix} fs_u & 0 & c_u & 0 \\ 0 & fs_v & c_v & 0 \\ 0 & 0 & 0 & fs_u * b \end{bmatrix} \frac{1}{z_o} \begin{bmatrix} x_o \\ y_o \\ z_o \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} x_o \\ y_o \\ z_o \\ 1 \end{bmatrix} = \begin{bmatrix} oR_r R^T & -oR_r R^T P \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix}$$

Solving the equation, we can get x_w, y_w, z_w. Set estimated landmark position in world frame as $\mu_{t,i}$, and $\Sigma_{t,i}$ to identify matrix for any unseen landmark.

Because of projection function π in observation model isn't linear, we need to use EKF to estimate $\mu_{t+1,i}$ and $\Sigma_{t+1,i}$.

To get $\frac{\partial}{\partial m_i} h(T_{t+1}, m_i)$, we use chain rule:

$$\pi(q) = \frac{1}{q_3} q \in R^4 \quad (44)$$

$$\frac{d\pi}{dq}(q) = \frac{1}{q_3}\begin{bmatrix} 1 & 0 & -\frac{q_1}{q_3} & 0 \\ 0 & 1 & -\frac{q_2}{q_3} & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -\frac{q_4}{q_3} & 1 \end{bmatrix} \in R^{4*4} \quad (45)$$

$$\begin{aligned} \frac{\partial}{\partial m_i} h(T_{t+1}, m_i) &= K_s \frac{\partial \pi}{\partial q}(q_i) \frac{\partial q_i}{\partial m_i} \\ &= K_s \frac{\partial \pi}{\partial q}\left(oT_I T_t^{-1} \underline{m_i} \right) oT_I T_{t+1}^{-1} \frac{\partial \underline{\underline{m_i}}}{\partial m_i} \\ &= K_s \frac{\partial \pi}{\partial q}\left(oT_I T_t^{-1} \underline{m_i} \right) oT_I T_{t+1}^{-1} P^T = \mathrm{H} \end{aligned} \quad (46)$$

$$P = [I\ 0] \in R^{3*4} \quad (47)$$

We can plug this into EKF update step:

$$K_{t+1} = \Sigma_t H_{t+1}^T (H_{t+1} \Sigma_t H_{t+1}^T + V) \quad (48)$$

$$\mu_{t+1} = \mu_t + K_{t+1}(z_{t+1} - \tilde{z}_{t+1}) \quad (49)$$

$$\Sigma_{t+1} = (I - K_{t+1} H_{t+1})\, \Sigma_t \quad (50)$$

$\tilde{z}_{t+1}$ is predicted observation:

$$\tilde{z}_{t+1} = K_s \pi({}_o T_I T_{t+1}^{-1} \underline{\mu_t}) \in R^4 \quad (51)$$

C. Visual-Inertial SLAM

To compute EKF update step for IMU pose, we need to find derivative of observation model with respect to pose. Here, we directly show the result:

$$H_{t+1,i} = -K_s \frac{d\pi}{dq}(oT_I \mu_{t+1|t}^{-1} \underline{m_i}) oT_I (\mu_{t+1|t}^{-1} \underline{m_i})^{\odot} \in R^{4*6} \quad (52)$$

$$\begin{bmatrix} s \\ 1 \end{bmatrix}^{\odot} = \begin{bmatrix} I & -\hat{s} \\ 0 & 0 \end{bmatrix} \in R^{4*6} \quad (53)$$

$\mu_{t+1|t} \in SE(3)$ obtained from part A.

Covariance in part C contains IMU pose and landmarks, so we need to combine H from part B.

$$H = \left[H_l\ H_p \right] \in R^{4M*3M+6} \quad (54)$$

H_l is derivative about landmarks from (46).

H_p is derivative about from (52).

$$H = \begin{bmatrix} H_{l,1} & \cdots & 0 & H_{p,1} \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & H_{l,M} & H_{p,M} \end{bmatrix}$$

Then, we can plug (54) into EKF Update Step:

$$K_{t+1} = \Sigma_{t+1|t} H_{t+1}^T (H_{t+1} \Sigma_{t+1|t} H_{t+1}^T + V) \quad (55)$$

$$\mu_{t+1,i} = \mu_{t,i} + K_{t+1,i}(z_{t+1,i} - \tilde{z}_{t+1,i}) \text{ for landmark} \quad (56)$$

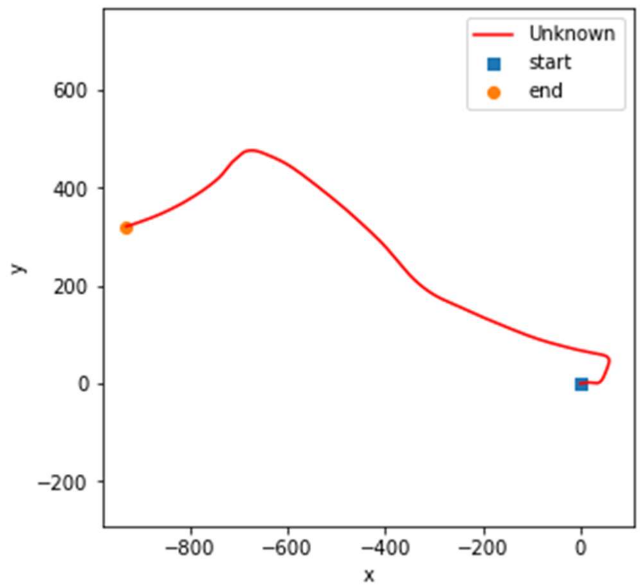
$$\mu_{t+1|t+1} = \mu_{t+1|t} \exp\left((K_{t+1}(z_{t+1,i} - \tilde{z}_{t+1,i}))^{\wedge} \right) \text{ for pose} \quad (57)$$

$$\Sigma_{t+1|t+1} = (I - K_{t+1} H_{t+1})\, \Sigma_{t|t} \quad (58)$$

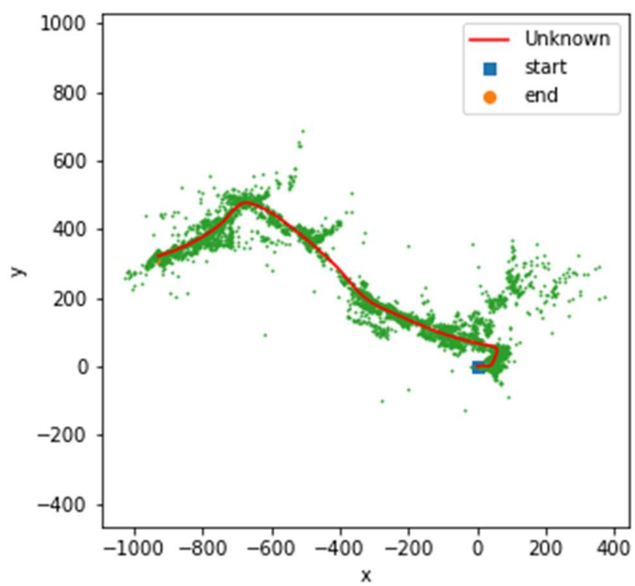
$*$ mean and covariance here are defined in (12) and (13).

IV. Results

A. IMU Localization via EKF Prediction



B. Landmark Mapping via EKF Update



Observation noise V is identity matrix.

C. Visual-Inertial SLAM

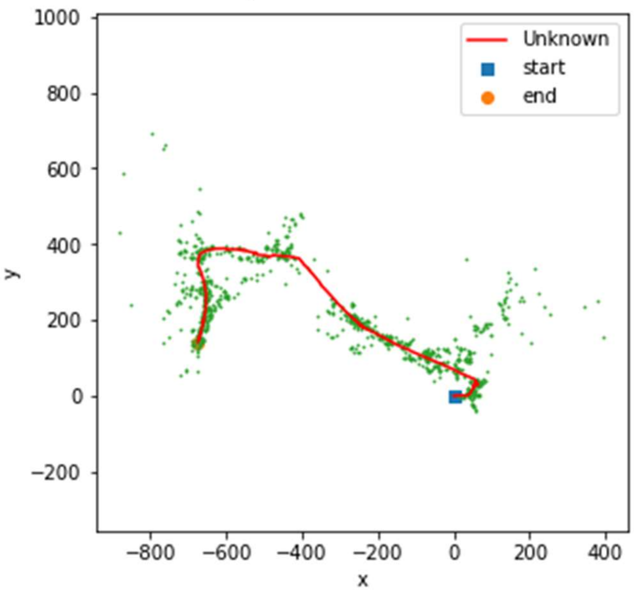
Setup:

Observation noise V is identity matrix.
Landmark prior covariance is identity matrix.

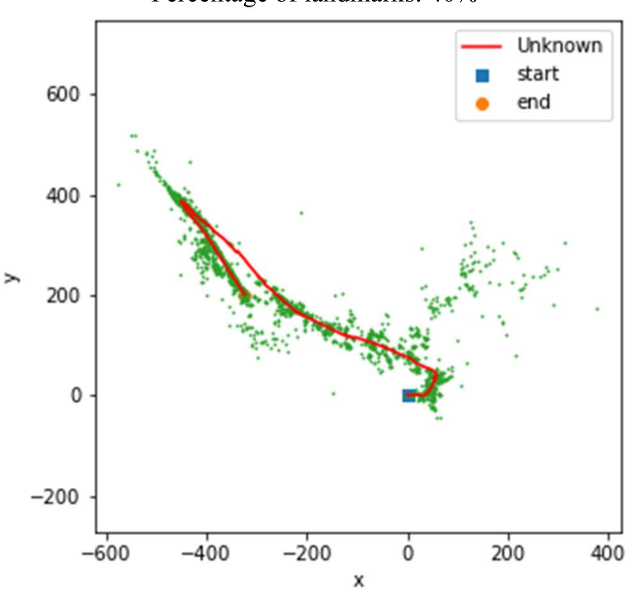
Motion noise W is $\begin{bmatrix} 0.5 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.03 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.03 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.03 \end{bmatrix}$

According to TA, the IMU is flipped, so I tried to flip landmarks back.

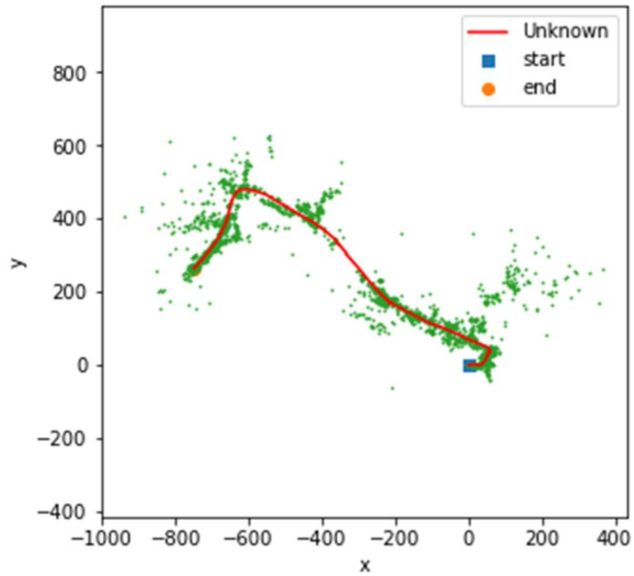
Percentage of landmarks: 20%



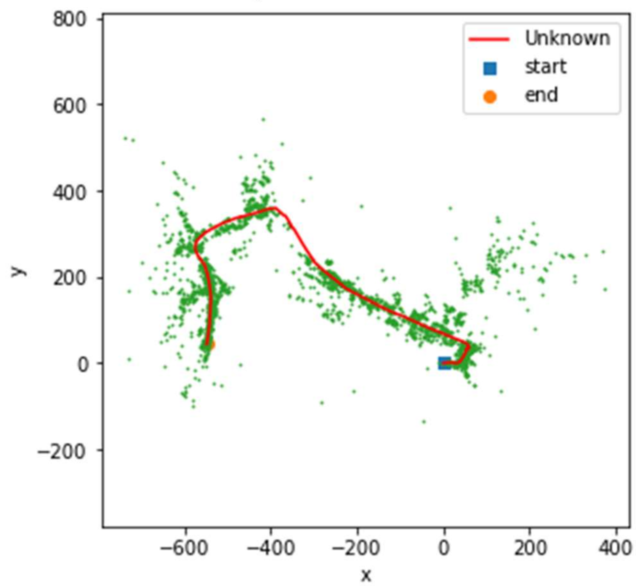
Percentage of landmarks: 40%



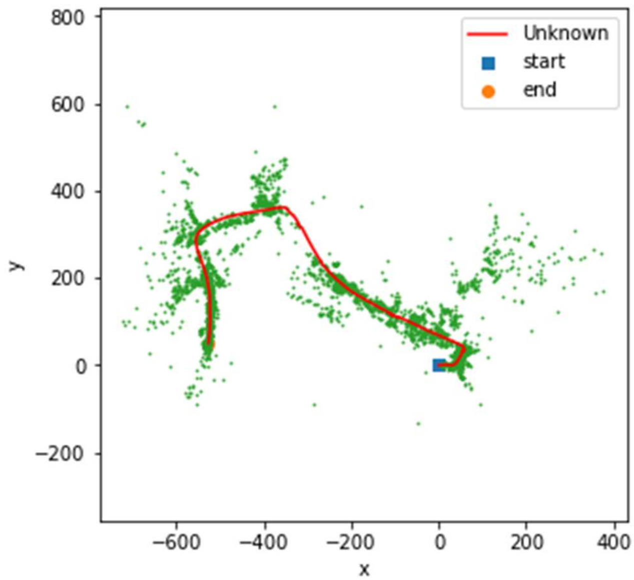
Percentage of landmarks: 50%



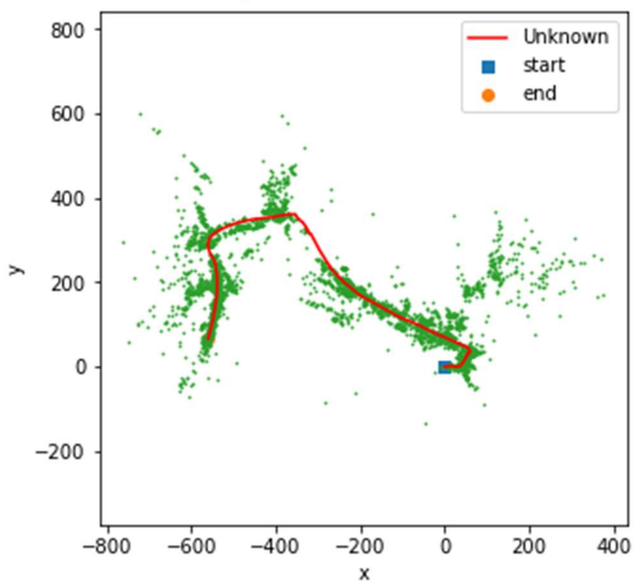
Percentage of landmarks: 60%



Percentage of landmarks: 80%



Percentage of landmarks: 100%



Landmarks are randomly chosen every time. We can observe some interesting conclusions.
Even the number of landmarks is small, the prediction can still perform well with proper choice of landmarks. It’s obvious for 20% and 40%.
When the number of landmarks is sufficient, the prediction tends to be convergent. We can see it from 60%, 80% and 100%.

V. REFERENCE

1. Professor Nikolay Atanasov’s slides (lecture11~14)