ECE276A Project 3: Visual-Inertial SLAM

Wilson Liao
Department of Electrical and Computer Engineering
University of California, San Diego
w4liao@ucsd.edu

Ethen Tsao (in discuss)
Department of Electrical and Computer Engineering
University of California, San Diego
ltsao@ucsd.edu

I. Introduction

SLAM (Simultaneous Localization and Mapping) is a technique used in robotics and computer vision to create a map of an unknown environment while at the same time keeping track of the robot's position within that environment. SLAM is a challenging problem because it involves estimating both the position of the robot and the location of landmarks in the environment simultaneously, and dealing with the uncertainty that arises from sensor noise and imperfect knowledge of the environment. It's applications in robotics include autonomous navigation, mapping of unknown environments, and augmented reality.

In this project, we are solving the slam problem for a car in an unknown environment with IMU and landmark feature data obtained from the sensors on the car. We propose a solution for visual-inertial SLAM based on Extended Kalman Filter (EKF) and visual landmarks. The EKF is a version of non-linear Kalman Filter using first-order Talyor series to approximate the motion and observation models around the state and noise means.

II. Problem Formulation

A. Simultaneous localization and mapping

The SLAM problem is a combination of 2 problems, which are mapping and localization. Given robot's trajectory $x_{0:T}$ and sensor measurements $z_{0:T}$ and observation model h, the mapping problem can be written in this way:

$$\min_{m} \sum_{t=0}^{T} ||z_t - h(x_t, m)||_2^2 \tag{1}$$

where m is the map.

Given the map m, sensor measurements $z_{0:T}$, observation model h, control input $u_{0:T-1}$ and motion model f, we can localize the robot, i.e. get the robot's trajectory, by defining the localization problem as follows:

$$\min_{x_{0:T}} \sum_{t=0}^{T} ||z_t - h(x_t, m)||_2^2 + \sum_{t=0}^{T-1} ||x_{t+1} - f(x_t, u_t)||_2^2 \tag{2}$$

where $x_{0:T}$ is the trajectory.

Combining (1) and (2), we can formulate an optimization problem to estimate the orientation trajectory $x_{1:T}$ and map m. The following is the objective function of the slam problem:

$$\min_{x_{1:T}, m} \sum_{t=1}^{T} ||z_t - h(x_t, m)||_2^2 + \sum_{t=0}^{T-1} ||x_{t+1} - f(x_t, u_t)||_2^2 \tag{3}$$

B. Bayes Filter

Bayes Filter is a probabilistic technique for estimating the state x_t of a dynamical system by combining evidence from control inputs u_t and observations z_t using Markov assumptions and Bayes rule. The Bayes filter keeps track of predicted pdf and updated pdf.

- Prediction Step: given a prior pdf $p_{t|t}$ of x_t and control input u_t, we use the motion model p_f to compute the predicted pdf $p_{t+1|t}$ of x_t:

$$p_{t+1|t}(x) = \int p_f(x|s, u_t) p_{t|t}(s) ds \tag{4}$$

- Update Step: given a predicted pdf $p_{t+1|t}$ of $x_t + 1$ and measurement $z_t + 1$, we use the observation model p_h to obtain the updated pdf $p_{t+1|t+1}$ of x_t:

$$p_{t+1|t+1}(x) = \frac{p_h(z_{t+1}|x) p_{t+1|t}(x)}{\int p_h(z_{t+1}|s) p_{t+1|t}(s) ds} \tag{5}$$

C. IMU Pose Estimation

The first problem of the project is the IMU pose estimation problem. Given IMU data $u_t = [v_t, \omega_t]^T \in \mathbb{R}^6$, where v_t and ω are linear velocity and angular velocity respectively, we need to estimate the pose of IMU $T_t \in SE(3)$ over time t.

D. Landmark Mapping

The second problem of the project is the landmark mapping problem. Given the robot poses x_t (obtained from IMU pose estimation) and landmark observations $z_{0:T}$, we need to estimate the position of the landmarks $m \in \mathbb{R}^{3 \times M}$, where M is the number of landmarks.

E. Visual Inertial SLAM

The third problem of the project is the visual inertial slam problem. Given IMU data $u_t = [v_t, \omega_t]^T \in \mathbb{R}^6$, and landmark observations $z_{0:T}$, where $z_T \in \mathbb{R}^{4 \times M}$ (left and right image pixels), we need to simultaneously localizing the robot pose and landmark mapping in the world frame.

III. Technical Approach

A. IMU Localization via EKF prediction

We adopt EKF to solve the SLAM problem. Here we are using the prediction step of EKF for IMU localization. We aim to update the mean $\mu \in SE(3)$ and covariance $\Sigma \in \mathbb{R}^{6\times 6}$ of IMU position in the prediction step. The predicted mean and covariance can be written with motion noise $\omega_t \sim N(0, W)$ as follows:

$$\mu_{t+1|t} = exp(-\tau \hat{u}_t)\mu_{t|t} \tag{6}$$

$$\Sigma_{t+1|t} = exp(-\tau \overset{\curlywedge}{u}_t)\Sigma_{t|t} exp(-\tau \overset{\curlywedge}{u}_t)^T + W \tag{7}$$

where τ is the time discretization, $\hat{u}_t \in \mathbb{R}^{4\times 4}$ is the hat map of the control input u_t, and $\overset{\curlywedge}{u}_t \in \mathbb{R}^{6\times 6}$ is the adjoint of $\hat{u}_t$, which are shown as follows:

$$\hat{u}_t = \begin{bmatrix} \hat{\omega}_t & v_t \\ 0 & 0 \end{bmatrix} \tag{8}$$

$$\overset{\curlywedge}{u}_t = \begin{bmatrix} \hat{\omega}_t & \hat{v}_t \\ 0 & \hat{\omega}_t \end{bmatrix} \tag{9}$$

where $\hat{\omega}_t \in \mathbb{R}^{3\times 3}$ is the hat map of $\omega_t \in \mathbb{R}^3$ and $\hat{v}_t \in \mathbb{R}^{3\times 3}$ is the hat map of $v_t \in \mathbb{R}^3$.

B. Landmark Mapping via EKF update

We are using the update step of EKF for landmark mapping. We aim to update the mean $\mu \in \mathbb{R}^{3M}$ and covariance $\Sigma \in \mathbb{R}^{3M\times 3M}$ of landmark position in the update step. The predicted mean and covariance can be written with observation noise $v_t \sim N(0, V)$ as follows:

$$K_{t+1|t} = \Sigma_{t|t} H_t^T (H_t \Sigma_{t|t} H_t^T + I \otimes V) \tag{10}$$

$$\mu_{t+1|t} = \mu_{t|t} + K_t(z_t - \tilde{z}_t) \tag{11}$$

$$\Sigma_{t+1|t} = (I - K_t H_t)\Sigma_{t|t} \tag{12}$$

where z_t is the current observation, $\tilde{z}_t$ is the predicted observation, K_t is the kalman gain, $H_t \in \mathbb{R}^{4N_t\times 3N}$ is the observation model Jacobian where N_t and N are the number of current observed landmarks and total landmarks respectively. H_t and $\tilde{z}_t$ are shown as follows:

$$\tilde{z}_t = M\pi({}_oT_i U_t \mu_{t|t}) \tag{13}$$

$$H_{t,i,j} = \begin{cases} M\frac{d\pi}{dq}({}_oT_i U_t \mu_{t,j}) {}_oT_i U_t P^T & \\ 0 & \text{otherwise} \end{cases} \tag{14}$$

where the camera calibration matrix M is defined as follows:

$$M = \begin{bmatrix} fs_u & 0 & c_u & 0 \\ 0 & fs_b & c_b & 0 \\ fs_u & 0 & c_u & -fs_b \\ 0 & fs_u & c_v & 0 \end{bmatrix} \tag{15}$$

The derivative of projection function $\pi(\mathbf{q})$ can be written as follows:

$$\frac{d\pi}{d\mathbf{q}}(\mathbf{q}) = \begin{bmatrix} 1 & 0 & -\frac{q1}{q3} & 0 \\ 0 & 1 & -\frac{q2}{q3} & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -\frac{q4}{q3} & 1 \end{bmatrix} \tag{16}$$

The observation z_t in the world frame can be obtained as follows:

$$d = u_L - u_R = \frac{1}{z} f s_u b \tag{17}$$

$$\begin{bmatrix} u_L \\ v_L \\ d \end{bmatrix} = \begin{bmatrix} fs_u & 0 & c_u & 0 \\ 0 & fs_b & c_b & 0 \\ fs_u & 0 & c_u & -fs_b \\ 0 & fs_u & c_v & 0 \end{bmatrix} \frac{1}{z} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \tag{18}$$

$$\begin{bmatrix} u_L \\ v_L \\ d \end{bmatrix} = {}_oR_i R^T(\mathbf{m} - \mathbf{p}) \tag{19}$$

where ${}_oR_i$ is the rotation matrix of IMU to camera, R^T is the rotation matrix of world to IMU, $\mathbf{p}$ is the current IMU position in the world frame, and all other parameters are from camera data.

The resulting $\mathbf{m}$ is our current observation z_t in the world frame.

C. Visual-Inertial SLAM

We combine the EKF prediction step and update step to localize IMU pose and also map the landmarks. The prediction step remains the same in part A. The update step is similar to part B, but is modified as follows:

$$K_{t+1|t} = \Sigma_{t|t} H_{t+1|t}^T (H_{t+1|t} \Sigma_{t|t} H_{t+1|t}^T + I \otimes V) \tag{20}$$

$$\mu_{t+1|t} = exp(K_{t+1}(z_{t+1} - \tilde{z_{t+1}})^{\wedge}) \tag{21}$$

$$\Sigma_{t+1|t} = (I - K_{t+1|t} H_{t+1|t})\Sigma_{t+1|t} \tag{22}$$

where $z_t \tilde{+} 1$ is the predicted observation, $H_{t+1|t} \in \mathbb{R}^{4N_t\times 6}$. $H_{t+1|t}$ and $\tilde{z_{t+1}}$ are shown as follows:

$$\tilde{z_t}, i = M\pi({}_oT_i \mu_{t+1|t} \mathbf{m}_j), \quad i = 1 \ldots N_t \tag{23}$$

$$H_{i,t+1|t} = M\frac{d\pi}{dq}({}_oT_i \mu_{t+1|t} \mathbf{m}_j) {}_oT_i (\mu_{t+1|t} \mathbf{m}_j)^{\odot} \tag{24}$$

IV. Results

The following are the results of visual-inertial SLAM for 2 different datasets, including the original dead-reckoning path with landmark positions and visual SLAM path with landmark positions.

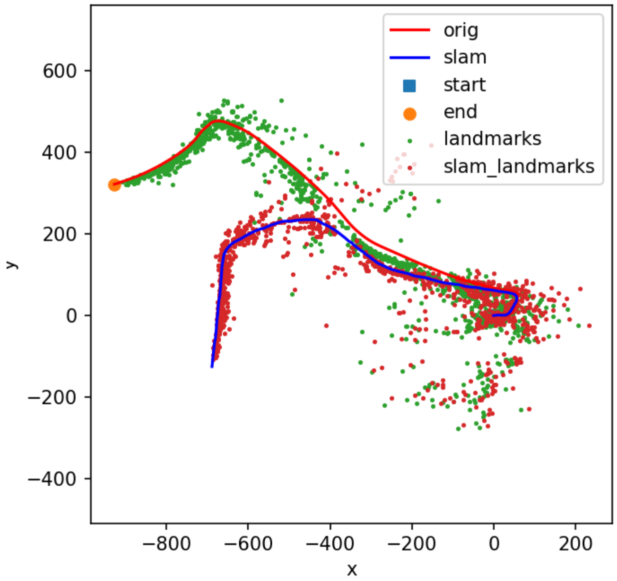

Fig. 1. dataset 3

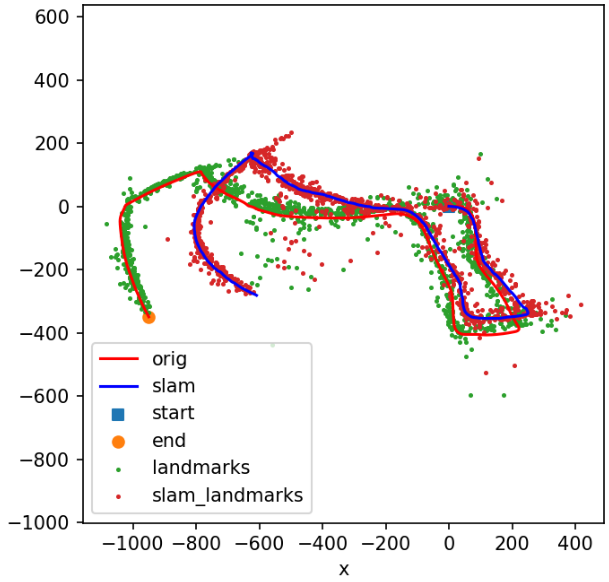

Fig. 2. dataset 10

From the result above, we can see that the trajectory gained from visual SLAM have a shift from the ones gained from dead-reckoning, which we viewed it as an improve from using only dead-reckoning without visual data. Although the result seems to be fine, there are still some points we can improve:

- Use more feature points. Due to the computational issue, we only adopt a small portion ($\frac{1}{10} \sim \frac{1}{4}$) of the data.
- Fine-tune the covariance of motion and observation noise, since we found EKF is quite noise sensitive.