ECE 276A: Sensing and Estimation in Robotics Project 3

Lin-Hsi Tsao (PID: A59015911)
Department of Electrical and Computer Engineering
University of California, San Diego
ltsao@ucsd.edu

Acknowledge: Wilson Liao (w4liao@ucsd.edu)

I. INTRODUCTION

In this project, we are required to implement Visual-Inertial SLAM.

A. Visual-Inertial SLAM

SLAM (Simultaneous Localization and Mapping) is a computational problem in robotics that involves building a map of an unknown environment and simultaneously determining the location of the robot within that environment.

In this project, we used the data obtained from the IMU odometry and feature points detected by the stereo camera. Given these data, we implemented the Extended Kalman Filter (EKF), which utilizes multivariate Taylor Series expansions to linearize a nonlinear model, to perform the prediction step and the update step for the SLAM problem.

The rest of the paper is as follows. The detailed formulations of the SLAM problem are in Section II. Technical approaches are introduced in Section III. The experiment, results are presented in Section IV.

II. PROBLEM FORMULATION

A. IMU localization via EKF prediction

The IMU pose estimation problem could be expressed as: given an IMU data $u_t = [v_t \, \omega_t]^T \in R^6$, we will predict the pose of IMU $T_t \in$ SE(3) over time t, where v_t is linear velocity and ω_t is the angular velocity.

B. Landmark mapping via EKF update

The landmark position estimation problem could be expressed as: given the visual feature observations $z_{0:T}$ and the inverse IMU pose $U_t \in SE(3)$, estimate the homogeneous coordinates $m \in R^{4\times M}$ in the world frame of the landmarks that generated the visual observation, where M is the total number of landmarks.

C. Visual-Inertial SLAM

The Visual-Inertial SLAM problem could be expressed as: given the IMU data and the stereo visual features $z_t \in R^{4\times N_t}$, including the left and right image pixels, we simultaneously localize the IMU pose and feature mapping in the world frame.

III. TECHNICAL APPROACH

A. IMU localization via EKF prediction

The task is to compute the mean $\mu \in SE(3)$ and covariance $\sigma \in R^{6\times 6}$ of our predicted result. The following equation (1) and equation (2) with $w_t \sim N(0, W)$ are needed to be solved.

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

$$\Sigma_{t+1|t} = exp(-\tau\hat{\hat{u}})\sigma_{t|t}exp(-\tau\hat{\hat{u}})^T + W \tag{2}$$

where τ denotes the time time discretization for t_t - t_{t+1}, $\hat{u} \in R^{4\times 4}$ denotes the hat map of the control input u_t, and $\hat{\hat{u}}$ denotes the adjoint of $\hat{u}$. The $\hat{u}$ and $\hat{\hat{u}}$ could be expressed as:

$$\hat{u} = \begin{bmatrix} \hat{\omega_t} & v_t \\ 0 & 0 \end{bmatrix} \tag{3}$$

$$\hat{\hat{u}} = \begin{bmatrix} \hat{\omega_t} & \hat{v_t} \\ 0 & \hat{\omega_t} \end{bmatrix} \tag{4}$$

where $\hat{\omega_t}$ denotes corresponding skew-symmetric matrix of $w_t \in R^3$ and $v_t \in R^3$ denotes the linear velocity.

B. Landmark mapping via EKF update

The task is to compute the mean $\mu \in R^{3M}$ and covariance $\Sigma \in R^{3M\times 3M}$ of our predicted result. Hence, equations (5) (6) (7) are needed to be solved.

$$K_{t+1|t} = \Sigma_{t|t}H_t^T(H_t\Sigma_{t|t}H_t^T + I \otimes V) \tag{5}$$

$$\mu_{t+1|t} = \mu_{t|t} + K_t(z_t - M\pi({}_oT_iU_t\mu_{t|t}) \tag{6}$$

$$\Sigma_{t+1|t} = (I - K_tH_t)\Sigma_{t|t} \tag{7}$$

where z_t is the current observation, M is the stereo camera calibration matrix, π is the projection function, and $H_t \in R^{4N_t\times 3N}$ is the observation model Jacobian evaluated at μ_t, with N_t and N represents the number of currently observed landmarks and total landmarks.s

In more details, M could expressed as:

$$M = \begin{bmatrix} f_{su} & 0 & c_u & 0 \\ 0 & f_{sv} & c_v & 0 \\ f_{su} & 0 & c_u & -f_{su}b \\ 0 & f_{sv} & c_v & 0 \end{bmatrix} \tag{8}$$

$$H_{t,i,j} = Md\frac{d\pi}{dq}({}_oT_iUt\mu_{t,j}){}_oT_iU_tP^T \tag{9}$$

$H_{t,i,j} = 0$ if observation i does not have corresponding landmark j at time t. The projection function could be expressed as:

$$\pi(q) = \frac{1}{q_3} q \tag{10}$$

$$\frac{d\pi}{dq}(q) = \begin{bmatrix} 1 & 0 & \frac{q_1}{q_3} & 0 \\ 0 & 1 & \frac{q_2}{q_3} & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{q_4}{q_3} & 1 \end{bmatrix} \tag{11}$$

To obtain observation z_t in the world coordinate, the following equations needed to be utilized:

$$d = u_L - u_R = \frac{1}{z} f_{su} b \tag{12}$$

$$\begin{bmatrix} u_L \\ v_L \\ d \end{bmatrix} = \begin{bmatrix} f_{su} & 0 & c_u & 0 \\ 0 & f_{sv} & c_v & 0 \\ 0 & 0 & 0 & f_{su} b \end{bmatrix} \frac{1}{z} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \tag{13}$$

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = {}_oR_i R^T (m - p) \tag{14}$$

where ${}_oR_i$ is the rotation matrix transform the IMU frame to the camera frame, R^T is the rotation matrix transform the world frame to the IMU frame, p is the current IMU position in the world frame, the remaining parameters can be obtained from camera data, and m is the current observation z_t in the world frame.

C. Visual-Inertial SLAM

In Visual-Inertial SLAM, we combine both the EKF prediction and update step to achieve final IMU localization. To achieve this task, we will need to perform the prediction and update steps. As for the prediction step, the equations could be expressed as:

$$\mu_{t+1|t} = exp(-\tau \hat{u}) \mu_{t|t} \tag{15}$$

$$\Sigma_{t+1|t} = exp(-\tau \hat{\hat{u}}) \sigma_{t|t} exp(-\tau \hat{\hat{u}})^T + W \tag{16}$$

As for the update step, the equations could be expressed as:

$$K_{t+1} = \Sigma_{t+1|t} H_{t+1|t}^T (H_{t+1} \Sigma_{t+1|t} H_{t+1}^T + I \otimes V)^{-1} \tag{17}$$

$$\mu_{t+1|t+1} = exp(K_{t+1}(z_{t+1} - (\widetilde{z}_{t+1})^{\wedge}) \tag{18}$$

$$\Sigma_{t+1|t+1} = (I - K_{t+1} H_{t+1}) \Sigma_{t+1|t} \tag{19}$$

where $\widetilde{z}_{t+1}$ is our predicted observation, and $H_{t+1|t} \in R^{4N_t \times 6}$. To be more specific, we have our predicted observation based on $\mu_{t+1|t}$ and known correspondence π_t

$$\widetilde{z}_{t+1,i} = M\pi({}_oT_i \mu_{t+1|t} m_j), i = 1, ..., N_t \tag{20}$$

$$H_{i,t+1|t} = M \frac{d\pi}{dq}({}_oT_i \mu_{t+1|t} m_j) {}_oT_i (\mu_{t+1|t} m_j)^{\odot} \tag{21}$$

IV. Results

Below are the results of the landmarks. From the results, we can see that the slam pattern is different from the original pattern, which is obtained from the IMU data. To improve the result, we could sample more visual features and fine-tune the noise of the prediction step and the update step.

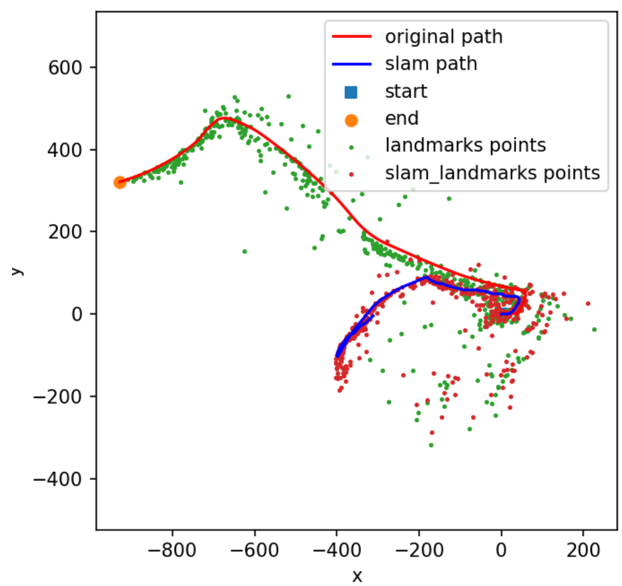

Fig. 1. Compared landmarks on dataset 3

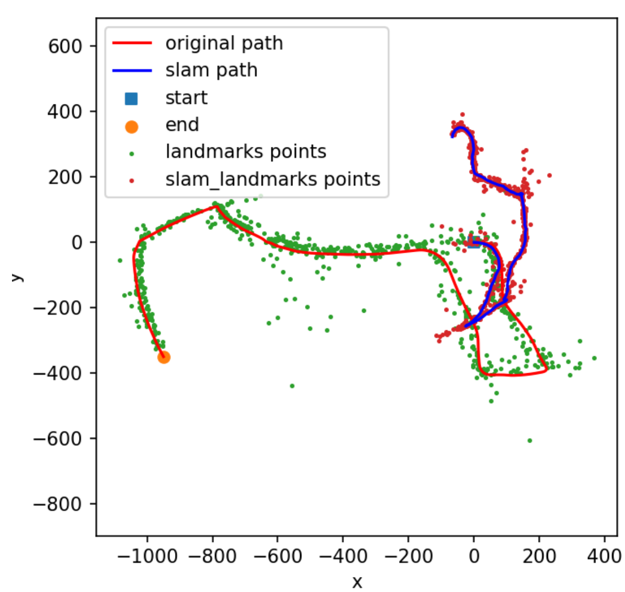

Fig. 2. Compared landmarks on dataset 10