Visual-Inertial SLAM

Yahsiu Hsieh
Department of Electrical & Computer Engineering
University of California, San Diego
y4hsieh@eng.ucsd.edu

***Abstract*—This paper presented an implementation of simultaneous localization and mapping (SLAM) using Extended Kalman Filter.**

I. Introduction

Simultaneous Localization and Mapping (SLAM) is a topic of study that has been around for many years. It is the fundamental component in AI problem. Various applications can be found in motion planning, unmanned air vehicle, and self driving vehicle.

This project aims to simultaneously localize and map the robot in an unknown indoor environment using IMU odometry data and feature points detected by stereo camera. A extended kalman filter based approach is implemented to achieve this objective. The kalman filter is the optimal linear estimate for linear system models with additive independent white noise in both the prediction and the observation systems. The extended kalman filter adapted techniques from calculus, namely multivariate Taylor Series expansions, to linearize a nonlinear model.

The rest of paper is as follows. First we give the detailed formulations of SLAM problem in Section II. Technical approaches are introduced in Section III. And at last we setup the experiment, results are presented in Section IV.

II. Problem Formulations

A. IMU Pose Estimation

The IMU pose estimation problem is that: given a IMU data $\mathbf{u_t} := \begin{bmatrix}\mathbf{v_t} & \omega_\mathbf{t}\end{bmatrix}^T \in \mathbb{R}^6$ ($\mathbf{v_t}$ is linear velocity and $\omega_\mathbf{t}$ is angular velocity), try to predict the pose of IMU $T_t \in SE(3)$ over time t.

B. Landmark Position Estimation

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

C. SLAM

The SLAM problem is that: given the IMU data containing linear velocity $v_t \in \mathbb{R}^3$ and angular velocity $\omega_t \in \mathbb{R}^3$, and the stereo visual features $\mathbf{z}_t \in \mathbb{R}^{4\times N_t}$ (left and right image pixels), simultaneously localizing the IMU pose and feature mapping in the world frame.

III. Technical Approaches

In this section we discuss about algorithms and models used in this project.

A. IMU Localization via EKF Prediction

Our goal here is having extended kalman filter prediction for IMU localization.

To achieve this task, we will need to know the mean $\mu \in SE(3)$ and covariance $\Sigma \in \mathbb{R}^{6\times 6}$ of our predicted result. Hence, we will need to solve equation (1) and equation (2) with $w_t \sim N(0, W)$.

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

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

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 $\hat{\hat{u}}_t \in \mathbb{R}^{6\times 6}$ is the adjoint of $\hat{u}_t$

To be more specific, we have

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

$$\hat{\hat{u}}_t = \begin{bmatrix}\hat{\omega}_t & \hat{v}_t \\ 0 & \hat{\omega}_t\end{bmatrix} \tag{4}$$

where $\hat{\omega}_t$ is the corresponding skew-symmetric matrix of $\omega_t \in \mathbb{R}^3$ and v_t is the linear velocity $\in \mathbb{R}^3$.

B. Landmark Mapping via EKF Update

Out goal here is using extended kalman filter to obtain the landmark position given IMU and camera data.

To achieve this task, we will need to know the mean $\mu \in \mathbb{R}^{3M}$ and covariance $\Sigma \in \mathbb{R}^{3M\times 3M}$ of our landmark position. Hence, we will need to solve equation (14), (15), and (16).

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

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

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

where $\mathbf{z}_t$ is the current observation, M is the stereo camera calibration matrix, pi is our projection function, and $H_t \in \mathbb{R}^{4N_t\times 3N}$ is the observation model Jacobian evaluated at μ_t, with N_t and N represents the number of current observed landmarks and total landmarks, respectively.

To be more specific, we have

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

$$H_{t,i,j} = \begin{cases} M \frac{d\pi}{dq}({}_oT_i U t \mu_{t,j}) {}_oT_i U_t P^T \\ 0, \qquad \text{otherwise} \end{cases} \tag{9}$$

$H_{t,i,j}$ has value if observation i corresponds to landmark j at time t.

$$\pi(\mathbf{q}) := \frac{1}{q_3}\mathbf{q}, \quad \frac{d\pi}{d\mathbf{q}}(\mathbf{q}) = \begin{bmatrix} 1 & 0 & -\frac{q1}{q3} & 0 \\ 0 & 1 & -\frac{q2}{q3} & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -\frac{q4}{q3} & 1 \end{bmatrix} \tag{10}$$

To obtain observation $\mathbf{z}_t$ in world coordinate, we will need to utilize the following three equations.

$$d = u_L - u_R = \frac{1}{z} f_{su} b \tag{11}$$

$$\begin{bmatrix} u_L \\ v_L \\ d \end{bmatrix} = \begin{bmatrix} f_{su} & 0 & c_u & 0 \\ 0 & f_{sv} & c_v & 0 \\ 0 & 0 & 0 & f_{su}b \end{bmatrix} \frac{1}{z} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \tag{12}$$

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = {}_oR_i R^T (\mathbf{m} - \mathbf{p}) \tag{13}$$

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

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

C. Visual-Inertial SLAM

Our goal here is to combine both the EKF prediction and update step to achieve final IMU localization.

To achieve this task, we will need to preform prediction and update step. For the prediction step, we do the exact same thing as mentioned in part A. For the update step, equations have changed a bit, the following three equations is what we are solving now.

$$K_{t+1|t} = \Sigma_{t+1|t} H^T_{t+1|t} (H_{t+1|t} \Sigma_{t+1|t} H^T_{t+1|t} + I \otimes V) \tag{14}$$

$$\mu_{t+1|t} = exp(K_{t+1}(\mathbf{z}_{t+1} - \tilde{\mathbf{z}}_{t+1})^\wedge) \tag{15}$$

$$\Sigma_{t+1|t} = (I - K_{t+1|t} H_{t+1|t}) \Sigma_{t+1|t} \tag{16}$$

where $\tilde{\mathbf{z}}_{t+1}$ is our predicted observation, and $H_{t+1|t} \in \mathbb{R}^{4N_t \times 6}$.

To be more specific, we have our predicted observation based on $\mu_{t+1|t}$ and known correspondence π_t

$$\tilde{\mathbf{z}}_{t+1,i} := M\pi({}_oT_i \mu_{t+1|t} \mathbf{m}_j), \; i = 1 \dots N_t \tag{17}$$

$$H_{i,t+1|t} = M \frac{d\pi}{d\mathbf{q}}({}_oT_i \mu_{t+1|t} \mathbf{m}_j) {}_oT_i (\mu_{t+1|t} \mathbf{m}_j)^\odot \tag{18}$$

IV. Results

In this section I will present the results of visual-inertial SLAM for 3 different datasets, including their corresponding paths and landmarks position before filtering.

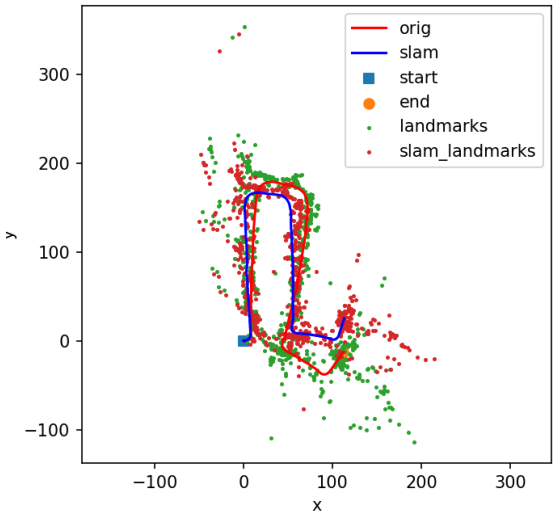

Fig. 1: Dataset 22

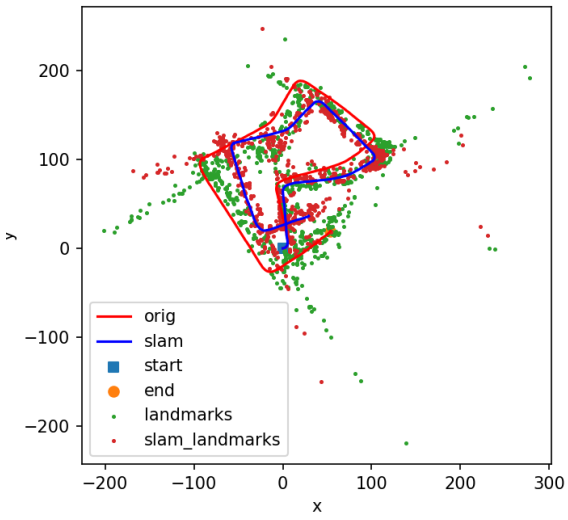

Fig. 2: Dataset 27

A. Conclusion

From the results above, we can see that visual inertial SLAM performs quite well compare to the original IMU data. However, for the dataset 27, we can see that there is a small difference between the start and end position, meaning that there is still

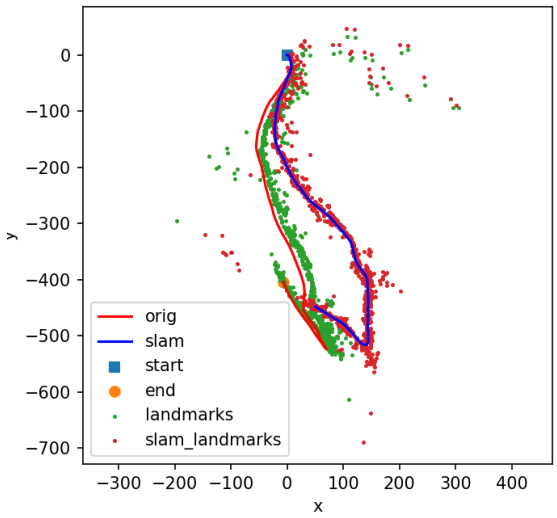

Fig. 3: Dataset 34

room for improvement (the resulting trajectory should be a close loop).

Below are some possible ways to improve:

- Increase the number of landmark to obtain higher accuracy, right now we are only using a quarter of the data (to avoid time penalty).
- Try different noise variance of our motion and observation model.

In conclusion, the result of visual-inertial SLAM is quite well.