ECE 276A Project 3 - Visual-Inertial SLAM

Yu-Hao Liu
dept. of Electrical and Computer Engineering
l*University of California, San Diego*
San Diego, CA
yul133@eng.ucsd.edu

***Abstract*—In this project, we implemented visual-inertial simultaneous localization and mapping (SLAM) using the Extended Kalman Filter. The data comes from a car with two sensors, an IMU and a stereo camera.**

Keywords—SLAM, Extended Kalman Filter, Stereo Camera, IMU

I. Introduction

In this paper, we aims to build a map that shows the trajectory where a car traveled and landmarks that were observed by the car. To reach the target, filtering is needed to derive a result that resists the noise coming from the environment and sensors. Therefore, we introduced a classic method of SLAM, which applies the Extended Kalman Filter (EKF) to estimate the true location of the car as well as the right position of the features it observed. Kalman Filtering is one of the Bayes Filtering. However, it has assumptions that the motion model and the observation model is linear in the state and affected by Gaussian noise. The EKF method expanded the assumptions by canceling the linear limitations, but forcing the predicted probability density functions and updated probabilities density functions to be Gaussian. There are two steps for implementing EKF, including a prediction step and a update step. We would estimate the positions of the car by its IMU pose and the position of landmarks by the pixel observed from the stereo camera. Through the EKF algorithm, the positions of the car and the landmarks would be updated to a more likely option.

II. Problem Formulation

A. IMU-based Localization via EKF Prediction

In the first problem, we are trying to localize the car at each time step based on IMU pose only. Therefore, we would apply the EKF prediction step.

- Input: $SE(3)$ kinematics
- Output 1: IMU pose over time t $T_t \in SE(3)$,
- Output 2: the covariance of the IMU pose Σ_{imu}

The process contains defining a Gaussian distribution given the previous state and observations., $U_t|z_{0:t}, u_{0:t-1}$, that represents the inverse IMU pose, predicting its next state U_{t+1} using the motion model.

B. Landmark Mapping via EKF Update

For the second problem, we assume the predicted IMU trajectory from part A is correct and focus on estimating the landmark positions.

- Input 1: Unknown landmarks positions $m \in \mathbb{R}^{3M}$
- Input 2: Visual Observation at time t z_t
- Output 1: the mean of landmarks μ_m
- Output 2: the covariance of landmarks Σ_m

The process contains representing the landmark as a Gaussian distribution $m \sim \mathcal{N}(\mu_t, \Sigma_t)$, predicting observations based on the mean at the last step, then updating the mean and the covariance using the EKF method.

C. Visual-Inertial SLAM

For the last problem, we combined what we've done in the previous two parts: the IMU prediction step from part A with the landmark update step from part B. In addition, we update the IMU pose in this part based on the stereo camera observation model to obtain a complete visual-inertial SLAM algorithm.

- Input 1: IMU pose over time t $T_t \in SE(3)$
- Input 2: Unknown landmarks positions $m \in \mathbb{R}^{3M}$
- Input 3: Visual Observation at time t z_t
- Output 1: the mean of landmarks and IMU μ
- Output 2: the covariance of landmarks and IMU Σ

III. Technical Approach

In this section, I would introduce how Extended Kalman Filter is applied to visual-inertial SLAM. I break down this task into three parts according to the request of the assignment: Localization-only problem, mapping-only problem, and the combination visual-inertial SLAM.

A. Nonlinear Kalman Filter

A nonlinear Kalman filter is a Bayes filter that has several assumptions. The prior probabilities density function (pdf) is Gaussian. The motion model and observation model are affected by Gaussian noise. The process noise w_t and measurement noise v_t are independent of each other, of the state x_t and across time. Lastly, the posterior pdf is forced to be Gaussian via approximation. These assumptions can be described as followed:

Prior:

$$x_t|\, z_{0:t}, u_{0:t-1} \sim \mathcal{N}(\mu_{t|t}, \Sigma_{t|t})$$

Motion model:

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

Observation model:

$$z_t = h(x_t, v_t),\ v_t \sim \mathcal{N}(0, V)$$

B. Visual-Inertial Odometry via the EKF

If we consider the localization-only problem, we will simply the prediction step by using kinematic rather than dynamic equations. Objective is to estimate the inverse IMU pose $U_t = \ {}_{W}T_{I,t}^{-1} \in SE(3)$ over time given the IMU measurements $u_{0:T}$ with $u_t \coloneqq [v_t^T\ \omega_t^T]^T$

First, there would be three assumptions. Linear velocity $v_t \ \in \ \mathbb{R}^3$ instead of linear acceleration $a_t \in \mathbb{R}^3$ measurements are available. The world-frame landmark coordinates $m \in \mathbb{R}^{3xM}$ are known. Also, the data association $\pi_t: \{1, \ldots, M\} \ \rightarrow \{1, \ldots, N_t\}$ stipulating which landmarks were observed at each time t is known or provided by an external algorithm.

The motion model with time discretization τ and noise $w_t \sim \mathcal{N}(0, W)$ is described as following:

$$U_{t+1} = \exp\left(-\tau\big((u_t + w_t)\big)^{\wedge}\right) U_t$$

Note that since that U_t is the inverse IMU pose, $u_t + w_t$ is negative.

Using the perturbation idea and converting to the discrete-time form again, we can re-write the motion model in terms of nominal kinematics of the mean of T_t and zero-mean perturbation kinematics. Therefore, the EKF prediction step is listed as below:

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

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

where

$$\hat{u}_t \coloneqq \begin{bmatrix} \hat{\omega}_t & v_t \\ 0^T & 0 \end{bmatrix} \in \ \mathbb{R}^{4x4}$$

$$\breve{u}_t \coloneqq \begin{bmatrix} \hat{\omega}_t & \hat{v}_t \\ 0 & \hat{\omega}_t \end{bmatrix} \in \ \mathbb{R}^{6x6}$$

Next, we look at the observation model:

$$z_{t+1} \coloneqq h\big(U_{t+1}, m_j\big) + v_{t+1,i}$$

$$\coloneqq M\pi\big({}_{O}T_I U_{t+1} m_j\big) + v_{t+1,i}$$

where projection function π is defined as:

$$\pi(\boldsymbol{q}) \coloneqq \frac{1}{q3}\boldsymbol{q} \ \in \ \mathbb{R}^4$$

and stereo camera model is defined as:

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

To update, we use the Taylor series to expand approximate the observation i at time $t+1$ using an inverse IMU pose perturbation $\delta\mu_{t+1|t+1}$ is:

$$z_{t+1,i} \ = M\pi\big({}_{O}T_I \exp\big(\widehat{\delta\mu}_{t+1|t+1}\big)\, \mu_{t+1} m_j\big) + v_{t+1,i}$$
$$\approx M\pi\big({}_{O}T_I \mu_{t+1} m_j\big) + H\delta\mu_{t+1|t+1} + v_{t+1,i}$$

where Jacobian H is the first-order Taylor series approximation of observation:

$$H = \ M\frac{d\pi}{dq}\,\big({}_{O}T_I \mu_{t+1} m_j\big)\, {}_{O}T_I\big(\mu_{t+1} m_j\big)^{\odot}$$

The derivative of projection matrix is:

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

Note that for homogeneous coordinates $\underline{s} \in \ \mathbb{R}^4$:

$$\begin{bmatrix} \underline{s} \\ 1 \end{bmatrix}^{\odot} \coloneqq \begin{bmatrix} I & -\hat{s} \\ 0 & 0 \end{bmatrix} \ \in \ \mathbb{R}^{4x6}$$

From above equations, we can perform the EKF updates. For each time step, we first predict observations based $\mu_{t+1|t}$ and known correspondences π_t:

$$\tilde{z}_{t+1,i} = \ M\pi\big({}_{O}T_I U_{t+1} m_j\big)\ for\ i = 1\,, \ldots, N_t$$

which, N_t indicates the number of observations observed at current time step.

We then compute the Jacobian $H_{i,t+1}$ of $\tilde{z}_{t+1,i}$ with respect to U_{t+1} evaluated at $\mu_{t+1|t}$. Lastly, we update the Kalman Gain, the mean and the covariance of the IMU pose accordingly as followed:

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

$$\mu_{t+1|t+1} = \exp\left(\big(K_{t+1|t}(z_{t+1} - \tilde{z}_{t+1})\big)^{\wedge}\right)\mu_{t+1|t}$$

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

C. Visual Mappinp

In this part, visual mapping uses the observation model similar with previous section. However, there are still some differences. First, the observation i is approximated by the first-order Taylor series at time t using the perturbation $\delta\mu_{t,j}$ is:

$$z_{t+1,i} \ = M\pi\big({}_{O}T_I U_t(\mu_t + \delta\mu_{t,j})\big) + v_{t+1,i}$$
$$\approx M\pi\big({}_{O}T_I U_t \mu_t\big) + H\delta\mu_{t,j} + v_{t+1,i}$$

where Jacobian H of predicted observations is:

$$H = \ M\frac{d\pi}{dq}\,\big({}_{O}T_I U_t \mu_{t+1}\big)\, {}_{O}T_I U_t P^T$$

Note that P is a projection matrix:

$$P = [I\ 0]$$

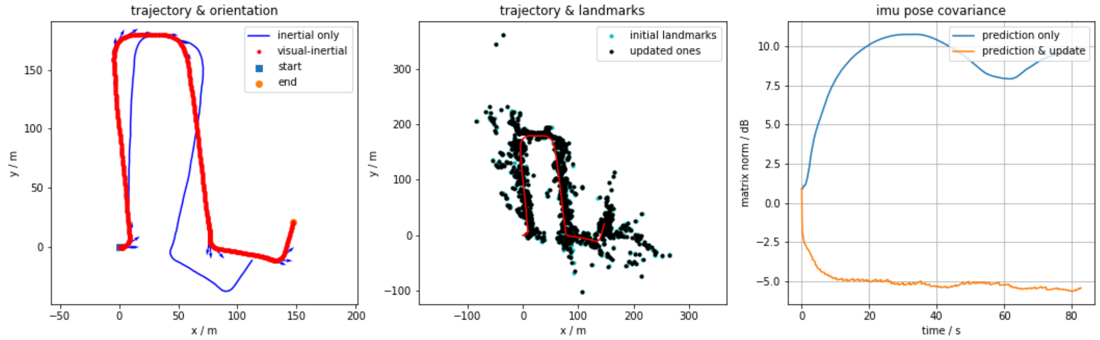

Fig. 1. Data 0022

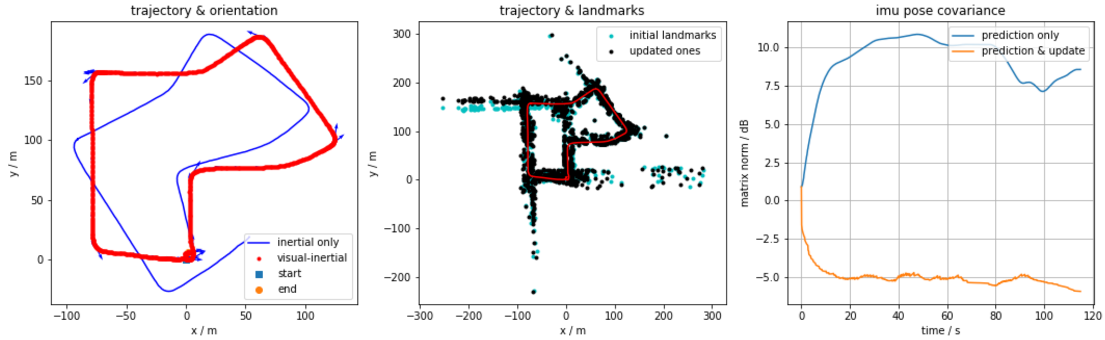

Fig. 2. Data 0027

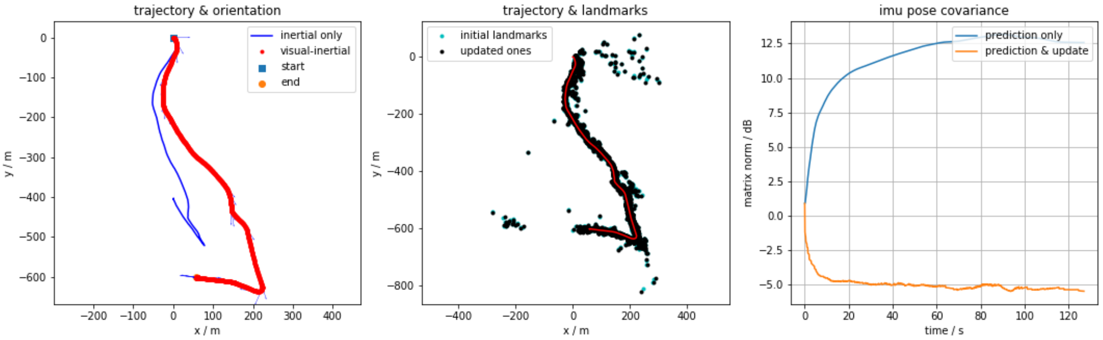

Fig. 3. Data 0034

Therefore, we can first predict the observations based on μ_t and known correspondences π_t:

$$\tilde{z}_{t+1,i} = \ M\pi\left({}_O T_I U_{t+1}\mu_{t,j}\right) for\ i = 1\ , \dots , N_t$$

After calculating the Jacobian using the equations above, we can perform the EKF update:

$$K_t \ = \Sigma_t H_t^T (H_t \Sigma_t H_t^T \ + I \otimes V)^{-1}$$
$$\mu_{t+1} = \mu_t + K_t (z_t - \tilde{z}_t)$$
$$\Sigma_{t+1} = (I - K_t H_t)\Sigma_t$$

where

$$I \otimes V = \begin{bmatrix} V & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & V \end{bmatrix}$$

D. Visual-Inertial SLAM

For the last part, we combine the IMU pose and visual observations together to derive a more accurate estimation on the trajectory and the positions of landmarks. The implementation is actually the combination of previous two sections, which includes a prediction on IMU pose, a prediction on the observed landmarks, and an update on both IMU pose and the landmarks.

IV. RESULTS

For the last part, I would present three graphs. Each graph represents a result from a dataset and contains three subplots. First one is the comparison of two trajectories. The middle one indicates the difference between original landmarks and updated ones. The right plot refers to the covariance of predicted-only IMU pose as well as the version of a complete EKF process.

A. *Trajectory comparison*

We have three trajectories to compare the performance of the visual-inertial SLAM. From the video of the first data (Data 0022), we can find that the car drove from a lane to another parallel lane and made two turns in the end. Look back to the result, the updated trajectory is apparently more accurate than the predicted trajectory, which accumulated bias through time and drifted even more when having a turn.

The third data could have a clearer comparison between two trajectories. As you can see, the predicted one has a big mistake when the car was making a turn. It shows that the car made a dramatic turn but it didn't.

Also, to visualize the difference of the covariance of the IMU pose, I plot the norm of the covariance on the right graph. You can find out that the covariance of the updated one has decreased dramatically through time, which means that the filter became more and more confident after iterations while predicted one would be uncertain whenever there is a relatively big change.

Therefore, we can conclude that the EKF-based visual-inertial SLAM has enhanced the accuracy of the estimation.

B. *Landmark update*

As for the landmarks, the first data and the third data has a relatively small change. Nevertheless. we can observe a more apparent difference from the second data (Data 0027). In the demo videos, the car turned for six times. In the fifth turns, the initial observed landmarks changed a lot to the updated ones. This can be explained by looking at the trajectories. In the prediction-only trajectory of the data, it starts to drift after the second turn and accumulated the error, which affects the accuracy of the landmark position. However, by applying the EKF update, all landmarks has return to a reasonable position.