VISUAL-INERTIAL SLAM:A PROJECT REPORT

Renwei Bai

Department of Mechanical and Aerospace Engineering
University of California, San Diego
rbai@.ucsd.edu

1. INTRODUCTION

Simultaneous Localization and Mapping (SLAM) is a parameter estimation problem,which contains two parts, localization and mapping .And the two parts are performed simultaneously. SLAM has been applied in many areas, such as indoor robots, AR, unmanned aerial vehicle and automatic vehicles. In different areas, SLAM algorithm can implement different functions.For example, the sweeping robot uses the SLAM algorithm to identify and model different scenes and calculate its own location in these scenes. In UAV field, Active obstacle avoidance is a very typical weak application of Slam. The drone only needs to know where the obstacle is to plan the trajectory and bypass the obstacle.

In this project, we will use data from a inertial measurement unit (IMU) and a stereo camera for localization and mapping by implement extended Kalman filter method. IMU is used to measures linear and angler accelerations. The main parts of a IMU are accelerometer and gyroscope,which are the main factors affecting the performance of the IMU. The IMU used in autonomous vehicles can provide real-time location information by updating in high frequency (usually 1kHz).But it has a fatal drawback—its error will increase over time, so it can only rely on IMU for localization in a very short time. To solve this problem, IMU in autonomous vehicles is usually used together with the GNSS(Global Navigation Satellite System).

The rest of paper is arranged as follows. First I give the detailed formulations of the three parts of Visual-Inertial SLAM problems in Section 2.And then I will describe my technical approach to each problem in section 3. At least, I present my results, and discuss them.

2. PROBLEM FORMULATIONS

In this Visual-Inertial Localization and Mapping problem, we need to input linear acceleration $\mathbf{a}_t \in R^3$, rotational velocity$\omega_t \in R^3$ from IMU and features $\mathbf{z}_{t,i} \in R^4$ (left and right image pixels) for $i = 1, \ldots, N_t$ with the assumption that The transformation ${}_CT_I \in SE(3)$ from the IMU to the camera optical frame (extrinsic parameters) and the stereo camera calibration matrix **Ks** (intrinsic parameters) are known:

$$K_s := \begin{bmatrix} f_{s_u} & 0 & c_u & 0 \\ 0 & f_s & c_v & 0 \\ f_{s_u} & 0 & c_u & -f_u b \\ 0 & f_s & c_v & 0 \end{bmatrix} \quad \begin{aligned} f &= \text{ focal length } [m] \\ s_u, s_v &= \text{ pixel scaling[pixels } /m] \\ c_u, c_v &= \text{ principal point [pixels]} \\ b &= \text{ stereo baseline } [m] \end{aligned}$$

to obtain the outputs World-frame IMU pose ${}_WT_I \in SE(3)$ over time and World-frame coordinates $\mathbf{m}_j \in R^3$ for $j = 1, \ldots, M$ of the point landmarks that generated the visual features $\mathbf{z}_{t,i} \in R^4$.

The whole project problem will be completed in three steps by solving three problems: IMU Localization via EKF Prediction, Landmark Mapping via EKF Update, Visual-Inertial SLAM.

2.1. IMU Localization via EKF Prediction

In this part, we are given IMU measurements as control input $\mathbf{u}_{0:T}$ with $\mathbf{u}_t := \left[\mathbf{v}_t^\top, \boldsymbol{\omega}_t^\top\right]^\top \in R^6$, where $\mathbf{v}_t$ is linear velocity and ω_t is angular velocity at time t, to estimate the pose $T_t \mid \mathbf{z}_{0:t}, \mathbf{u}_{0:t-1} \sim \mathcal{N}\left(\boldsymbol{\mu}_{t|t}, \Sigma_{t|t}\right)$ with $\boldsymbol{\mu}_{t|t} \in SE(3)$ and $\Sigma_{t|t} \in R^{6\times 6}$ of the IMU over time t.

2.2. Landmark Mapping via EKF Update

Assuming that the predicted IMU trajectory from part one above is correct, we can use the T_t from part 1 and the given observations z_t , where z_t is the pixel coordinates $\mathbf{z}_t := \begin{bmatrix} \mathbf{z}_{t,1}^\top & \cdots & \mathbf{z}_{t,N_t}^\top \end{bmatrix}^\top \in R^{4N_t}$ for $t = 0, \ldots, T$ of detected visual features with correspondences between the left and the right camera frames are provided in the stereo camera data, to estimate the coordinate $\mathbf{m} := \begin{bmatrix} \mathbf{m}_1^\top & \cdots & \mathbf{m}_M^\top \end{bmatrix}^\top \in R^{3M}$ of the landmarks that generated them.

2.3. Visual-Inertial SLAM

In part (3), we need to combine the IMU prediction step,the prior pose $T_{t+1} \mid z_{0:t}, u_{0:t} \sim \mathcal{N}\left(\mu_{t+1|t}, \Sigma_{t+1|t}\right)$ with $\boldsymbol{\mu}_{t+1|t} \in SE(3)$ and $\Sigma_{t+1|t} \in R^{6\times 6}$ from part (1) with the landmark $\mathbf{m} \in R^{3M}$ update step from part (2) and implement

an IMU update step based on the stereo-camera observation model$\mathbf{z}_{t+1,i} = h\left(T_{t+1}, \mathbf{m}_j\right) + \mathbf{v}_{t+1,i}$ with measurement noise $\mathbf{v}_t \sim \mathcal{N}(0, V)$ to update the pose $T_t := wT_{l,t} \in SE(3)$ of the IMU over time.

3. TECHNICAL APPROACH

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

3.1. Extended Kalman Filter

The EKF uses a first-order Taylor series approximation to the motion and observation models around the state and noise means:

$$f\left(\mathbf{x}_t, \mathbf{u}_t, \mathbf{w}_t\right) \approx f\left(\boldsymbol{\mu}_{t|t}, \mathbf{u}_t, \mathbf{0}\right) + \left[\frac{df}{d\mathbf{x}}\left(\boldsymbol{\mu}_{t|t}, \mathbf{u}_t, \mathbf{0}\right)\right] \left(\mathbf{x}_t - \boldsymbol{\mu}_{t|t}\right) + \left[\frac{df}{d\mathbf{w}}\left(\boldsymbol{\mu}_{t|t}, \mathbf{u}_t, \mathbf{0}\right)\right] \left(\mathbf{w}_t - \mathbf{0}\right) \quad (1)$$

$$h\left(\mathbf{x}_{t+1}, \mathbf{v}_{t+1}\right) \approx h\left(\boldsymbol{\mu}_{t+1|t}, \mathbf{0}\right) + \left[\frac{dh}{d\mathbf{x}}\left(\boldsymbol{\mu}_{t+1|t}, \mathbf{0}\right)\right] \left(\mathbf{x}_{t+1} - \boldsymbol{\mu}_{t+1|t}\right) + \left[\frac{dh}{d\mathbf{v}}\left(\boldsymbol{\mu}_{t+1|t}, \mathbf{0}\right)\right] \left(\mathbf{v}_{t+1} - \mathbf{0}\right) \quad (2)$$

Let $F_t := \frac{df}{dx}\left(\mu_{t|t}, \mathbf{u}_t, \mathbf{0}\right)$ and $Q_t := \frac{df}{d\mathbf{w}}\left(\mu_{t|t}, \mathbf{u}_t, \mathbf{0}\right)$so that :

$$f\left(\mathbf{x}_t, \mathbf{u}_t, \mathbf{w}_t\right) \approx f\left(\boldsymbol{\mu}_{t|t}, \mathbf{u}_t, \mathbf{0}\right) + F_t\left(\mathbf{x}_t - \boldsymbol{\mu}_{t|t}\right) + Q_t\mathbf{w}_t \quad (3)$$

Then, the predicted mean and co-variance can be computed in closed form:

$$\begin{aligned}\mu_{t+1|t} &\approx \iint \left(f\left(\mu_{t|t}, \mathbf{u}_t, \mathbf{0}\right) + F_t\left(\mathbf{x} - \mu_{t|t}\right) + Q_t\mathbf{w}\right) \phi\left(\mathbf{x}; \boldsymbol{\mu}_{t|t}, \Sigma_{t|t}\right) \phi(\mathbf{w}; 0, W) d\mathbf{x} d\mathbf{w} \\ &= f\left(\boldsymbol{\mu}_{t|t}, \mathbf{u}_t, \mathbf{0}\right) + F_t\left(\int \mathbf{x}\phi\left(\mathbf{x}; \boldsymbol{\mu}_{t|t}, \Sigma_{t|t}\right) d\mathbf{x} - \boldsymbol{\mu}_{t|t}\right) \\ &\quad + Q_t \int \mathbf{w}\phi(\mathbf{w}; 0, W) d\mathbf{w} = f\left(\boldsymbol{\mu}_{t|t}, \mathbf{u}_t, \mathbf{0}\right) \quad (4)\end{aligned}$$

$$\begin{aligned}\Sigma_{t+1|t} &\approx \iint \left(f\left(\mu_{t|t}, \mathbf{u}_t, \mathbf{0}\right) + F_t\left(\mathbf{x} - \boldsymbol{\mu}_{t|t}\right) + Q_t\mathbf{w}\right) \left(f\left(\boldsymbol{\mu}_{t|t}, \mathbf{u}_t, \mathbf{0}\right) + F_t\left(\mathbf{x} - \boldsymbol{\mu}_{t|t}\right) + Q_t\mathbf{w}\right)^\top \\ &\quad \phi\left(\mathbf{x}; \boldsymbol{\mu}_{t|t}, \Sigma_{t|t}\right) \phi(\mathbf{w}; 0, W) d\mathbf{x} d\mathbf{w} - \mu_{t+1|t}\mu_{t+1|t}^\top \\ &= f\left(\boldsymbol{\mu}_{t|t}, \mathbf{u}_t, \mathbf{0}\right) \left(\int \left(\mathbf{x} - \boldsymbol{\mu}_{t|t}\right)^\top \phi\left(\mathbf{x}; \boldsymbol{\mu}_{t|t}, \Sigma_{t|t}\right) d\mathbf{x}\right) F_t^\top \\ &\quad + F_t\left(\int \left(\mathbf{x} - \boldsymbol{\mu}_{t|t}\right) \phi\left(\mathbf{x}; \mu_{t|t}, \Sigma_{t|t}\right) d\mathbf{x}\right) f\left(\boldsymbol{\mu}_{t|t}, \mathbf{u}_t, \mathbf{0}\right)^\top \\ &\quad + F_t\left(\int \left(\mathbf{x} - \mu_{t|t}\right)\left(\mathbf{x} - \mu_{t|t}\right)^\top \phi\left(\mathbf{x}; \boldsymbol{\mu}_{t|t}, \Sigma_{t|t}\right) d\mathbf{x}\right) F_t^\top \\ &\quad + F_t\left(\int \left(\mathbf{x} - \mu_{t|t}\right)\left(\mathbf{x} - \mu_{t|t}\right)^\top \phi\left(\mathbf{x}; \boldsymbol{\mu}_{t|t}, \Sigma_{t|t}\right) d\mathbf{x}\right) F_t^\top \\ &\quad + Q_t\left(\int \mathbf{w}\mathbf{w}^\top \phi(\mathbf{w}; 0, W) d\mathbf{w}\right) Q_t^\top \\ &= F_t\Sigma_{t|t}F_t^\top + Q_t W Q_t^\top \quad (5)\end{aligned}$$

Let $H_{t+1} := \frac{dh}{dx}\left(\mu_{t+1|t}, \mathbf{0}\right)$ and $R_{t+1} := \frac{dh}{dv}\left(\boldsymbol{\mu}_{t+1|t}, \mathbf{0}\right)$ so that:

$$h\left(\mathbf{x}_{t+1}, \mathbf{v}_{t+1}\right) \approx h\left(\boldsymbol{\mu}_{t+1|t}, \mathbf{0}\right) + H_{t+1}\left(\mathbf{x}_{t+1} - \boldsymbol{\mu}_{t+1|t}\right) + R_{t+1}\mathbf{v}_{t+1} \quad (6)$$

The joint distribution of $\mathbf{x}_{t+1}$ and $\mathbf{z}_{t+1}$ can be computed in closed form:

$$\begin{aligned}\mathbf{m}_{t+1|t} &:= \iint h(\mathbf{x}, \mathbf{v}) \phi\left(\mathbf{x}; \boldsymbol{\mu}_{t+1|t}, \Sigma_{t+1|t}\right) \phi(\mathbf{v}; 0, V) d\mathbf{x} d\mathbf{v} \\ &\approx h\left(\boldsymbol{\mu}_{t+1|t}, \mathbf{0}\right) \quad (7) \\ S_{t+1|t} &:= \iint \left(h(\mathbf{x}, \mathbf{v}) - \mathbf{m}_{t+1|t}\right)\left(h(\mathbf{x}, \mathbf{v}) - \mathbf{m}_{t+1|t}\right)^\top \\ &\quad \phi\left(\mathbf{x}; \boldsymbol{\mu}_{t+1|t}, \Sigma_{t+1|t}\right) \phi(\mathbf{v}; 0, V) d\mathbf{x} d\mathbf{v} \\ &\approx H_{t+1}\Sigma_{t+1|t}H_{t+1}^\top + R_{t+1}VR_{t+1}^\top \quad (8) \\ C_{t+1|t} &:= \iint \left(\mathbf{x} - \boldsymbol{\mu}_{t+1|t}\right)\left(h(\mathbf{x}, \mathbf{v}) - \mathbf{m}_{t+1|t}\right)^\top \\ &\quad \phi\left(\mathbf{x}; \boldsymbol{\mu}_{t+1|t}, \Sigma_{t+1|t}\right) \phi(\mathbf{v}; 0, V) d\mathbf{x} d\mathbf{v} \\ &\approx \Sigma_{t+1|t}H_{t+1}^\top \quad (9)\end{aligned}$$

The conditional Gaussian distribution of $\mathbf{X}_{t+1} \mid \mathbf{Z}_{t+1}$ is then:

$$\boldsymbol{\mu}_{t+1|t+1} := \boldsymbol{\mu}_{t+1|t} + K_{t+1|t}\left(\mathbf{z}_{t+1} - \mathbf{m}_{t+1|t}\right) \quad (10)$$

$$\Sigma_{t+1|t+1} := \Sigma_{t+1|t} - K_{t+1|t} \cdot S_{t+1|t} \cdot K_{t+1|t}^\top \quad (11)$$

$$K_{t+1|t} := C_{t+1|t} \cdot S_{t+1|t}^{-1} \quad (12)$$

By inserting equations (7)(8)(9) into equations (10)(11)(12), we get:

Update:

$$\mu_{t+1|t+1} = \mu_{t+1|t} + K_{t+1|t}\left(z_{t+1} - h\left(\mu_{t+1|t}, 0\right)\right) \quad (13)$$

$$\Sigma_{t+1|t+1} = \left(I - K_{t+1|t}H_{t+1}\right)\Sigma_{t+1|t} \quad (14)$$

Kalman Gain:

$$K_{t+1|t} := \Sigma_{t+1|t}H_{t+1}^\top \left(H_{t+1}\Sigma_{t+1|t}H_{t+1}^\top + R_{t+1}VR_{t+1}^\top\right)^{-1} \quad (15)$$

3.2. IMU Localization via EKF Prediction

In this part, we use prediction step of Extended Kalman Filter to estimate the pose $T_t \mid \mathbf{z}_{0:t}, \mathbf{u}_{0:t-1} \sim \mathcal{N}\left(\boldsymbol{\mu}_{t|t}, \Sigma_{t|t}\right)$ with $\boldsymbol{\mu}_{t|t} \in SE(3)$ and $\Sigma_{t|t} \in R^{6\times6}$ of the IMU over time t. This means that

$$T_t = \boldsymbol{\mu}_{t|t} \exp\left(\delta\hat{\boldsymbol{\mu}}_{t|t}\right) \quad (16)$$

with perturbation $\delta\boldsymbol{\mu}_{t|t} \sim \mathcal{N}\left(0, \Sigma_{t|t}\right)$

The motion model is nominal kinematics of $\mu_{t|t}$ and perturbation kinematics of $\delta\boldsymbol{\mu}_{t|t}$ with time discretization τ_t:

$$\boldsymbol{\mu}_{t+1|t} = \boldsymbol{\mu}_{t|t} \exp\left(\tau_t \hat{\mathbf{u}}_t\right) \quad (17)$$

$$\delta\boldsymbol{\mu}_{t+1|t} = \exp\left(-\tau_t \hat{\mathbf{u}}_t\right) \delta\boldsymbol{\mu}_{t|t} + \mathbf{w}_t \quad (18)$$

Applying equations (4) and (5) with equation (17) and (18), the **EKF Prediction Step** with $\mathbf{w}_t \sim \mathcal{N}(0, W)$ is :

$$\boldsymbol{\mu}_{t+1|t} = \boldsymbol{\mu}_{t|t} \exp\left(\tau_t \hat{\mathbf{u}}_t\right) \quad (19)$$

$$\Sigma_{t+1|t} = E\left[\delta\boldsymbol{\mu}_{t+1|t} \cdot \delta\boldsymbol{\mu}_{t+1|t}^{\top}\right]$$

$$= \exp\left(-\tau\hat{\mathbf{u}}_t\right) \Sigma_{t|t} \exp\left(-\tau\hat{\mathbf{u}}_t\right)^{\top} + W \quad (20)$$

where, $\mathbf{u}_t := \begin{bmatrix} \mathbf{v}_t \\ \boldsymbol{\omega}_t \end{bmatrix} \in R^6$ $\hat{\mathbf{u}}_t := \begin{bmatrix} \hat{\boldsymbol{\omega}}_t & \mathbf{v}_t \\ \mathbf{0}^{\top} & 0 \end{bmatrix} \in R^{4\times4}$ $\hat{\mathbf{u}}_t := \begin{bmatrix} \hat{\boldsymbol{\omega}}_t & \hat{\mathbf{v}}_t \\ 0 & \hat{\boldsymbol{\omega}}_t \end{bmatrix} \in R^{6\times6}$

3.3. Landmark Mapping via EKF Update

In this part,the observation model with measurement noise $\mathbf{v}_{t,i} \sim \mathcal{N}(0, V)$ is :

$$\mathbf{z}_t = K_s \pi\left(oT_I T_t^{-1} \underline{\mathbf{m}}\right) + \mathbf{v}_t$$

Where π is projection function:

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

Observation model is generated from Stereo Camera , which is the features in the form of $z_{t,i} = [u_L \quad v_L \quad u_R \quad v_R]^T$ provided in the **.npz** files. Together with the Stereo Camera Model we can obtain the optical-frame coordinates $[x \quad y \quad z]^T$ of landmarks $\mathbf{m}_i$:

$$\begin{bmatrix} u_L \\ v_L \\ u_R \\ v_R \end{bmatrix} = \underbrace{\begin{bmatrix} fs_u & 0 & c_u & 0 \\ 0 & fs_v & c_v & 0 \\ fs_u & 0 & c_u & -f_{s_u}b \\ 0 & f_{s_v} & c_v & 0 \end{bmatrix}}_{Ks} \frac{1}{z} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

Where K_s is from stereo baseline **b** (in meters) and camera calibration matrix:

$$K = \begin{bmatrix} fs_u & 0 & c_u \\ 0 & fs_v & c_v \\ 0 & 0 & 1 \end{bmatrix}$$

So, the prior $\mathbf{m} \mid \mathbf{z}_{0:t} \sim \mathcal{N}(\boldsymbol{\mu}_t, \Sigma_t)$ with $\boldsymbol{\mu}_t \in R^{3M}$ and $\Sigma_t \in R^{3M\times3M}$ can be calculated as:

$$\begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = {}_cT_i T_t^{-1} \underline{\mu_t}$$

Where ${}_cT_i$ is the transformation from the left camera to the IMU frame and T_t is the pose of IMU in world frame.
And now, with the assumption that the landmarks $\mathbf{m}_i$ are static. In other words, it is not necessary to consider a motion model or a prediction step for $\mathbf{m}_i$. Considering the landmark mapping problem, the EKF update step here is:

Kalman Gain:

$$K_{t+1} = \Sigma_t H_{t+1}^{\top}\left(H_{t+1}\Sigma_t H_{t+1}^{\top} + I \otimes V\right)^{-1} \quad (21)$$

Update:

$$\boldsymbol{\mu}_{t+1} = \boldsymbol{\mu}_t + K_{t+1}(\mathbf{z}_{t+1} - \underbrace{K_s \pi\left({}_oT_I T_{t+1}^{-1}\underline{\mu}_t\right)}_{\tilde{z}_{t+1}}) \quad (22)$$

$$\Sigma_{t+1} = \left(I - K_{t+1}H_{t+1}\right)\Sigma_t \quad (23)$$

$\tilde{\mathbf{z}}_{t+1} \in R^{4N_{t+1}}$ is the predicted observation based on the landmark position estimates μ_t at time **t**:

$$\tilde{\mathbf{z}}_{t+1,i} := K_s \pi\left({}_oT_l T_{t+1}^{-1}\underline{\boldsymbol{\mu}}_{t,j}\right) \in R^4 \quad \text{for } i = 1, \ldots, N_{t+1}$$

We need the observation model Jacobian $H_{t+1} \in R^{4N_t\times3M}$ evaluated at μ_t with block elements $H_{t+1,i,j} \in R^{4N_t\times3M}$: Applying the chain rule:

$$H_{t+1,i,j} = \begin{cases} K_s \frac{d\pi}{dq}\left({}_oT_l T_{t+1}^{-1}\underline{\mu}_{t,j}\right) {}_oT_I T_{t+1}^{-1} P^{\top} & \text{if } \Delta_t(j) = i \\ \mathbf{0}, & \text{otherwise} \end{cases} \quad (24)$$

Where $\mathbf{q}_{t+1,j} = {}_oT_I T_{t+1}^{-1}\underline{\mathbf{m}}_j$ and $P = \begin{bmatrix} I & 0 \end{bmatrix} \in R^{3\times4}$.

3.4. Visual-Inertial SLAM

In this part, we will combine the IMU prediction step from part (2) with the landmark update step from part (3) and implement an IMU update step based on the stereo-camera observation model to obtain a complete visual-inertial SLAM algorithm. The prior $\boldsymbol{\mu}_{t+1|t} \in SE(3)$ and $\Sigma_{t+1|t} \in R^{6\times6}$ are

calculated in equation (19)(20) in part (2). And the predicted observation based on $\mu_{t+1|t}$ and known correspondences t is:

$$\tilde{\mathbf{z}}_{t+1,i} := K_s \pi \left({}_O T_I \mu_{t+1|t}^{-1} \underline{\mathbf{m}}_j \right) \quad \text{for } i = 1, \ldots, N_{t+1} \quad (25)$$

Where $\underline{\mathbf{m}}_j$ is the position of landmarks in world frame, which are computed in part (3).And the Jacobian of $\tilde{\mathbf{z}}_{t+1,i}$,i with respect to T_{t+1} evaluated at $\Sigma_{t+1|t}$ is :

$$H_{t+1,i} = -K_s \frac{d\pi}{d\mathbf{q}} \left({}_O T_I \mu_{t+1|t}^{-1} \underline{\mathbf{m}}_j \right) {}_O T_I \left(\boldsymbol{\mu}_{t+1|t}^{-1} \underline{\mathbf{m}}_j \right)^{\odot} \in R^{4\times 6} \quad (26)$$

Perform the EKF update:

$$K_{t+1} = \Sigma_{t+1|t} H_{t+1}^{\top} \left(H_{t+1} \Sigma_{t+1|t} H_{t+1}^{\top} + I \otimes V \right)^{-1} \quad (27)$$

$$\mu_{t+1|t+1} = \mu_{t+1|t} \exp\left(\left(K_{t+1} \left(\mathbf{z}_{t+1} - \tilde{\mathbf{z}}_{t+1} \right) \right)^{\wedge} \right) \quad (28)$$

$$\Sigma_{t+1|t+1} = \left(I - K_{t+1} H_{t+1} \right) \Sigma_{t+1|t} \quad (29)$$

Where $H_{t+1} = \begin{bmatrix} H_{t+1,1} \\ \vdots \\ H_{t+1,N_{t+1}} \end{bmatrix}$ (30)

4. RESULTS

The pictures shown below are the images of the predicted trajectory of 03 **Fig.1**, the estimated 2-D positions Of the visual features of 03 **Fig.2** , the predicted trajectory of 10 **Fig.3** and the estimated 2-D positions Of the visual features of 10 **Fig.4** over time for the Visual-inertial SLAM problem. At first, in visual mapping problem, I calculated all the parameters in a three dimension matrix, for example $\mu_t \in R^{M\times 3\times 1}$ and $H_{t+1} \in R^{N\times 4\times 3}$ and update all the landmarks together in the form of three dimension matrices. But the matrix $H_{t+1}\Sigma_t H_{t+1}^{\top} + I \otimes V$ always became singular after some timestamps, which is not allowed to compute the inverse in equation (21). So I had to adjust my initial Σ to a very large value and reduced the numbers of the landmarks used in mapping, which made my previous results a mess. Under these circumstances, I changed the update algorithm and rewrote my code. The final edition of my program is a for-loop,which only updated one landmark for a time. I choose one landmark for every four of them. Finally, the mapping result looks good.

Unfortunately, there is some error in my code of Visual-Inertial SLAM part. Something wrong happens in computing $\boldsymbol{\mu}_{t+1|t+1}$ part. The value of my estimated trajectory is too large (see **Fig.5**). But, I do not have enough time to modify my code.

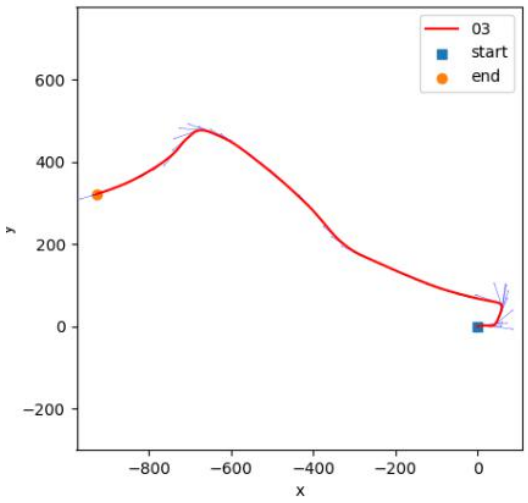

Fig.1 The predicted trajectory of 03

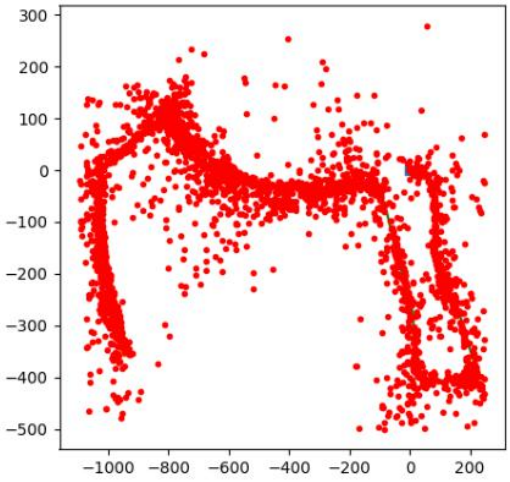

Fig.4 The estimated 2-D positions
Of the visual features of 10

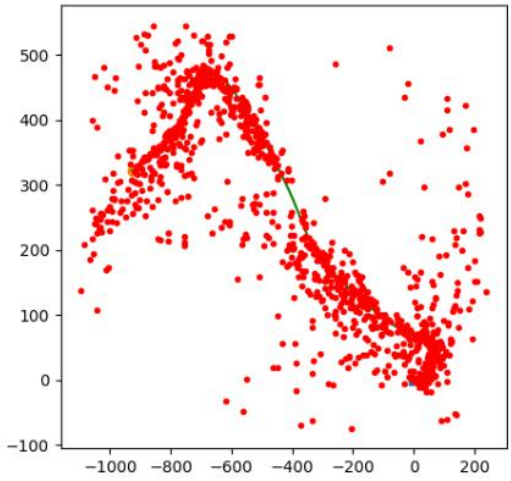

Fig.2 The estimated 2-D positions
Of the visual features of 03

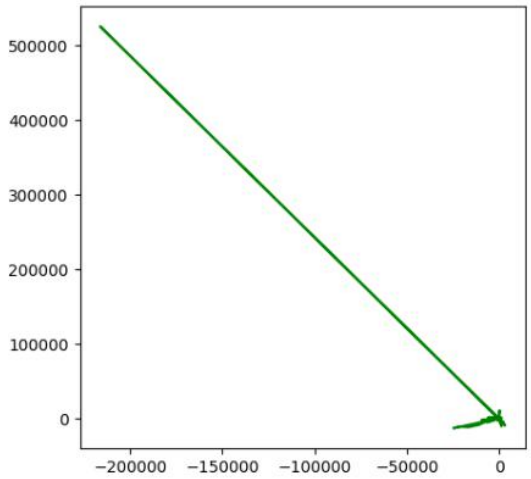

Fig.5 The estimated trajectory of 03

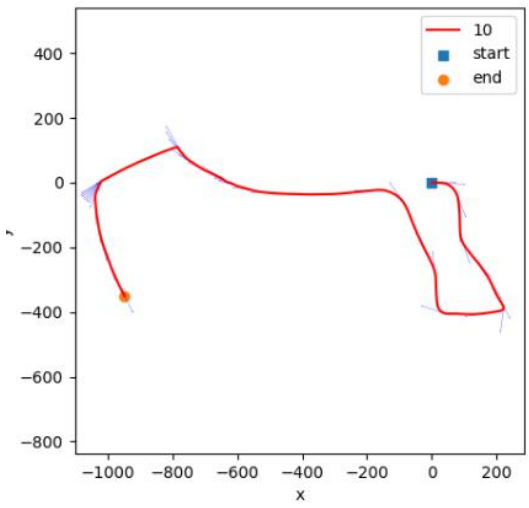

Fig.3 The predicted trajectory of 10