Visual-Inertial SLAM

Baoqian Wang
Department of Electrical and Computer Engineering
University of California, San Diego
San Diego, CA, 92093
Email:bawang@ucsd.edu

***Abstract*—This project investigates using Extended Kalman Filter (EKF) in simultaneous localization and mapping (SLAM) problem. In particular, the robot is localized using the EKF with IMU data which consists of prediction step and update step. Moreover, the environment where the robot is operated is then estimated using EKF with visual information . The experimental studies are conducted using the real IMU data,images, etc., collected by a ground vehicle. The proposed algorithm shows promising performance in SLAM problem.**

I. Introduction

Simultaneous localization and mapping is a hot topic in robotics area. It concerns with two questions of the robot, 'where am I?' and 'what does the environment looks like?'. It plays an important role in many robotic applications such as navigation, environment reconstruction and so on.

In the literature, the SLAM has been widely studied. Early SLAM approaches were mostly based on maximum likelihood estimation (MLE) [1], maximum a posterior estimation (MAP) and bayesian inference (BI)[2]. As the representations of the robot states, map, observations and control inputs affect the SLAM, different SLAM approaches are adopted for different representations. For instance, the map can be represented by landmark-based map, occupancy grid, surfels and polygonal mesh. Popular occupancy grid SLAM algorithms include Fast SLAM[3], which uses a particle filter to maintain the robot trajectory pdf and log-odds mapping to maintain a probabilistic map for every particle. Moreover, another algorithm namely kinect fusion [4] matches consecutive RGBD point clouds using the iterative closest point algorithm and updates a grid discretization of the truncated signed distance function (TSDF) representing the scene surface via weighted averaging. Moreover, examples for popular landmark-based SLAM algorithms include Rao-Blackwellized Particle Filter[5], Kalman Filter[6], Factor Graphs SLAM [7].

In this project, we focus on landmark-based SLAM algorithms. In particular, the features to represent landmarks in the environment are localized and mapped through the EKF, while the ground robot is localized simultaneously using EKF. The experimental studies based on the real data including IMU data and images show the promising performance of the algorithm.

The rest of this paper is organized as follows. Section II formulates the Visual Inertial SLAM problem. Section III presents the EKF SLAM approaches. The performance of the approaches are evaluated comprehensively in Section IV. Section V concludes the paper with a brief summary.

II. Problem Formulation

In this section, the Visual-Inertial SLAM problem is formulated.

A. Environment Representation

The environment is represented with visual features of landmarks in the environment (see Figure 1 as an illustration). In particular, suppose there are N features. A vector $\mathbf{m_i} \in \mathbb{R}^3, \forall i \in \{1,2,...,N\}$ is used to represent the coordinates of the feature in the environment. The image pixels values of the corresponding features i, is represented by $\mathbf{z_i} \in \mathbb{R}^4 \forall i \in \{1,2,...,N\}$.

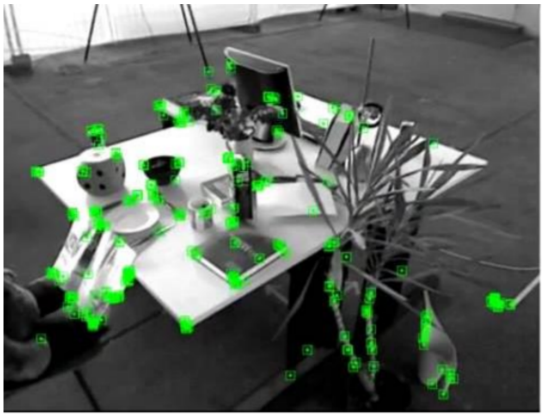


Figure 1. Illustration of image features

B. Observation Model

The observation model relates the image pixels values $\mathbf{z_i}$ with the features coordinates in the world frame $\mathbf{m_i}$ subject to measurement noise $\boldsymbol{v}$, which is captured by

$$\mathbf{z}_{t,i} = h\left(U_t, \mathbf{m}_j\right) + \mathbf{v}_{t,i} := M\pi\left(oT_l U_t \mathbf{m}_j\right) + \mathbf{v}_{t,i} \qquad (1)$$

where M is the calibration matrix, πq is the projection function. $v_{t,i}$ is the measurement noise, oT_l is the extrinsic parameter matrix which are represented by

$$M := \begin{bmatrix} fs_u & 0 & c_u & 0 \\ 0 & fs_v & c_v & 0 \\ fs_u & 0 & c_u & -fs_u b \\ 0 & fs_v & c_v & 0 \end{bmatrix} \qquad (2)$$

where f is the focal length, s_u, s_v are the pixel scaling and c_u, c_v are principle point, b is the stereo baseline.

$$\pi(\mathbf{q}) := \frac{1}{q_3}\mathbf{q} \in \mathbb{R}^4 \tag{3}$$

C. Motion Model

The motion model in this project is based on odometry data , i.e., IMU. The model is represented by

$$U_{t+1} = \exp\left(-\tau\left(\left(\mathbf{u}_t + \mathbf{w}_t\right)\right)^\wedge\right) U_t \tag{4}$$

where U_t is the inverse pose of the IMU, $\mathbf{u}_t$ is the IMU data that includes the linear velocity and angular velocity, w_t is the Gaussian noise which follows $\mathcal{N}(0, W)$., τ is the time step. The symbol indicates the Hat-map which is represented by

$$\phi^\wedge = \begin{bmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \end{bmatrix}^\wedge = \begin{bmatrix} 0 & -\phi_3 & \phi_2 \\ \phi_3 & 0 & -\phi_1 \\ -\phi_2 & \phi_1 & 0 \end{bmatrix} \in \mathbb{R}^{3\times 3}, \quad \phi \in \mathbb{R}^3 \tag{5}$$

D. SLAM Problem

Then the SLAM problem consists of two parts, mapping and localization, respectively. The localization problem is formulated as

The mapping is formulated as, given the visual features observations $\mathbf{z}0 : T$, estimate the homogeneous coordinates $\mathbf{m} \in \mathbb{R}^{4\times M}$ in the world frame of the landmarks that generated the visual observations.

The localization problem is formulated as given the IMU measurements $\mathbf{u_{0:T}}$ with $\mathbf{u}_t := \left[\mathbf{v}_t^\top, \omega_t^\top\right]^\top$ and visual feature observations $\mathbf{z}_{0:T}$, estimate the inverse IMU pose $U_t := wT_{l,t}^{-1} \in SE(3)$ over time.

Overall, the SLAM problem is formulated as: based on the above environment representations, motion model and observation model, the SLAM problem is described as follows: Given the IMU data: linear velocity v_t, rotational velocity ω_t, and camera data: visual features $z_t \in \mathbb{R}^{4\times N}$, the IMU pose ${}_wT_I \in SE(3)$ in the world frame over time and the world frame coordinates of the point landmarks $m \in \mathbb{R}^{\nVdash\times\mathbb{N}}$ corresponding to the visual features z_t are generated.

Based on the Markov assumption, the mathematical formulation of the SLAM problem is captured by:

$$\begin{aligned} p\left(\mathbf{U}_{0:t}, \mathbf{m} | \mathbf{z}_{0:t}, \mathbf{u}_{0:t-1}\right) = p_f\left(\mathbf{U}_{0:t} | \mathbf{z}_{0:t}, \mathbf{u}_{0:t-1}\right) \\ \prod_i p_h(\mathbf{m}_i | \mathbf{z}_{0:t}, \mathbf{x}_{0:t}) \end{aligned} \tag{6}$$

where p, p_f, p_h denotes the probability distribution of the trajectory and environment, motion model and observation model respectively.

III. Technical Approaches

In this section, the approaches to estimate the robot states $\mathbf{U}_{0:T}$ and the environment $\mathbf{m}$ are presented. In particular, the EKF algorithm is used to localize the robot and map the landmark features in the environment.

A. EKF Mapping

In the EKF mapping, the landmark features are assumed to have the prior Gaussian distribution which follows $\mathcal{N}(\mu_t, \sigma_t)$ By using EKF filter, the observation model described in Section II-C is first approximated using the first-order Taylor series approximation using the perturbation $\sigma\mu_{t,j}$, which is represented by

$$\begin{aligned} \mathbf{z}_{t,i} &= M\pi\left(oT_lU_t\left(\boldsymbol{\mu}_{t,j} + \delta\boldsymbol{\mu}_{t,j}\right)\right) + \mathbf{v}_{t,i} \\ &= M\pi\left(oT_lU_t\left(\underline{\boldsymbol{\mu}}_{t,j} + P^\top\delta\boldsymbol{\mu}_{t,j}\right)\right) + \mathbf{v}_{t,i} \\ &\approx M\pi\left(oT_lU_t\underline{\boldsymbol{\mu}}_{t,j}\right) + M\frac{d\pi}{d\mathbf{q}}\left(oT_lU_t\underline{\boldsymbol{\mu}}_{t,j}\right) \\ &\quad oT_lU_tP^\top\delta\boldsymbol{\mu}_{t,j} + \mathbf{v}_{t,i} \end{aligned} \tag{7}$$

Then the mean and covariance are updated by

$$\begin{aligned} K_t &= \Sigma_t H_t^\top\left(H_t\Sigma_t H_t^\top + I\otimes V\right)^{-1} \\ \mu_{t+1} &= \mu_t + K_t\left(z_t - \tilde{z}_t\right) \\ \Sigma_{t+1} &= \left(I - K_tH_t\right)\Sigma_t \end{aligned} \tag{8}$$

where H_t is represented by $M\frac{d\pi}{d\Phi}\left(oT_lU_t\underline{\mu}_{t,j}\right)oT_lU_tP^\top$ if observation i corresponds landmark j at time t, otherwise $H_t = 0$. and $\tilde{\mathbf{z}}_{t,i}$ is represented by

$$\tilde{\mathbf{z}}_{t,i} := M\pi\left(oT_lU_t\underline{\boldsymbol{\mu}}_{t,j}\right) \in \mathbb{R}^4 \tag{9}$$

$$I\otimes V := \begin{bmatrix} V & & \\ & \ddots & \\ & & v \end{bmatrix} \tag{10}$$

B. EKF Localization

In this subsection, the localization of the robot using EKF is introduced. The EKF consists of two steps, the prediction step and update step, which are described as follows

1) Prediction: In the prediction step, the inverse pose of the IMU has the prior distribution $U_t | \mathbf{z}_{0:t}, \mathbf{u}_{0:t-1} \sim \mathcal{N}\left(\boldsymbol{\mu}_{t|t}, \Sigma_{t|t}\right)$ with $\boldsymbol{\mu}_{t|t} \in SE(3)$ and $\Sigma_{t|t} \in \mathbb{R}^{6\times 6}$. By using the EKF, the predicted inverse pose of the robot is represented by

$$\begin{aligned} \boldsymbol{\mu}_{t+1|t} &= \exp\left(-\tau\hat{\mathbf{u}}_t\right)\boldsymbol{\mu}_{t|t} \\ \Sigma_{t+1|t} &= \mathbb{E}\left[\delta\boldsymbol{\mu}_{t+1|t}\delta\boldsymbol{\mu}_{t+1|t}^\top\right] \\ &= \exp\left(-\tau\hat{\mathbf{u}}_t\right)\Sigma_{t|t}\exp\left(-\tau\hat{\mathbf{u}}_t\right)^\top + W \end{aligned} \tag{11}$$

where

$$\begin{aligned} \mathbf{u}_t &:= \begin{bmatrix} \mathbf{v}_t \\ \omega_t \end{bmatrix} \in \mathbb{R}^6 \\ \hat{\mathbf{u}}_t &:= \begin{bmatrix} \hat{\omega}_t & \mathbf{v}_t \\ \mathbf{0}^\top & 0 \end{bmatrix} \in \mathbb{R}^{4\times 4} \\ \hat{\mathbf{u}}_t &:= \begin{bmatrix} \hat{\omega}_t & \hat{\mathbf{v}}_t \\ 0 & \hat{\omega}_t \end{bmatrix} \in \mathbb{R}^{6\times 6} \end{aligned} \tag{12}$$

2) Update: In the update step, given the predicted inverse pose $U_{t+1}|z_{0:t}, u_{0:t} \sim \mathcal{N}\left(\boldsymbol{\mu}_{t+1|t}, \Sigma_{t+1|t}\right)$ with $\boldsymbol{\mu}_{t+1|t} \in SE(3)$ and $\Sigma_{t+1|t} \in \mathbb{R}^{6\times 6}$. The observation is approximated using the first-taylor series by

$$\begin{aligned}
\boldsymbol{z}_{t+1,i} &= M\pi\left(oT_1 \exp\left(\hat{\boldsymbol{p}}\boldsymbol{\mu}_{t+1|t+1}\right)\boldsymbol{\mu}_{t+1|t}\mathbf{m}_t\right) + \mathbf{v}_{t+1,i} \\
&\approx M\pi\left(oT_t\left(I + \hat{\boldsymbol{\mu}}\boldsymbol{\mu}_{t+1|t+1}\right)\boldsymbol{\mu}_{t+1|t}\mathbf{m}_t\right) + \mathbf{v}_{t+1,i} \\
&= M\pi\left(oT_t\boldsymbol{\mu}_{t+1|t}\mathbf{m}_j + oT_t\left(\boldsymbol{\mu}_{t+1|t}\mathbf{m}_j\right)^{\odot}\delta\boldsymbol{\mu}_{t+1|t+1}\right) \\
&\quad + \mathbf{v}_{t+1,i} \\
&\approx M\pi\left(oT_1\boldsymbol{\mu}_{t+1|t}\mathbf{m}_j\right) + M\frac{d\pi}{d\mathbf{q}}\left(oT_1\boldsymbol{\mu}_{t+1|t}\mathbf{m}_j\right) \\
&\quad oT_t\left(\boldsymbol{\mu}_{t+1|t}\mathbf{m}_j\right)^{\odot}\delta\boldsymbol{\mu}_{t+1|t+1} + \mathbf{v}_{t+1,i}
\end{aligned} \tag{13}$$

The inverse pose of the robot is the updated by

$$\begin{aligned}
K_{t+1|t} &= \Sigma_{t+1|t}H_{t+1|t}^{\top}\left(H_{t+1|t}\Sigma_{t+1|t}H_{t+1|t}^{\top} + I \otimes V\right)^{-1} \\
\boldsymbol{\mu}_{t+1|t+1} &= \exp\left(\left(K_{t+1|t}\left(\mathbf{z}_{t+1} - \tilde{\mathbf{z}}_{t+1}\right)\right)^{\wedge}\right)\boldsymbol{\mu}_{t+1|t} \\
\Sigma_{t+1|t+1} &= \left(I - K_{t+1|t}H_{t+1|t}\right)\Sigma_{t+1|t}
\end{aligned} \tag{14}$$

where $H_{t+1|t} = \begin{bmatrix} H_{1,t+1|t} \\ \vdots \\ H_{N_{t+1},t+1|t} \end{bmatrix}$, $H_{i,t+1|t}$ is represented by

$$H_{i,t+1|t} = M\frac{d\pi}{d\mathbf{q}}\left(oT_1\boldsymbol{\mu}_{t+1|t}\mathbf{m}_j\right)oT_1\left(\boldsymbol{\mu}_{t+1|t}\mathbf{m}_j\right)^{\circ} \tag{15}$$

and $\tilde{\mathbf{z}}_{t+1,i}$ is captured by

$$\tilde{\mathbf{z}}_{t+1,i} := M\pi\left(oT_1\boldsymbol{\mu}_{t+1|t}\mathbf{m}_j\right) \tag{16}$$

C. Visual-Inertial SLAM

The Visual SLAM is to perform the EKF localization and EKF mapping simultaneously. Based on the EKF mapping and localization procedures described in the above sections. The Visual-Inertial SLAM algorithm is described as follows

IV. Results

In this section, the proposed Visual Inertial SLAM algorithm is implemented on the real data collected by ground robot.

A. Data descriptions

The data includes the IMU measurements which includes linear velocity $v_t \in \mathcal{R}^3$ and angular velocity $\omega_t \in \mathcal{R}^3$ measured in the body frame of the IMU. Stereo camera images and time stamps, intrinsic calibration and extrinsic calibration.

Algorithm 1: Visual-Inertial SLAM

```
Input: Stereo camera observations
       $\mathbf{z}_t,\ \forall t \in \{0, 1, ...., T\}$, IMU data
       $\mathbf{u}_t,\ \forall t \in \{0, 1, ...., T\}$, Calibration matrix
       $M$
Output: Trajectories of inverse pose of robot $\mathbf{U_t}$,
        Landmark features coordinate, $\mathbf{m}$
// Step 1: Initialization
Initialize prior inverse pose
  $U_t|\mathbf{z}_{0:t}, \mathbf{u}_{0:t-1} \sim \mathcal{N}\left(\boldsymbol{\mu}_{t|t}, \Sigma_{t|t}\right)$
Initialize landmark features $m_t \sim \mathcal{N}(\mu_t, \Sigma_t)$
for $t$=0:$T$ do
    // Step 2: Localization
    $U_{t+1|t} \leftarrow$
      $EKF\_Localization\_Prediction(u_t, M, U_t)$
      (Using Equation 11)
    $U_{t+1|t+1} \leftarrow$
      $EKF\_Localization\_Update(u_t, M, U_{t+1|t}, z_t)$
      (Using Equation 14)
    // Step 3: Mapping
    $\mathbf{m}_{t+1} \leftarrow EKF\_Mapping(\mathbf{m}_t, U_{t+1|t}, M, \mathbf{z_t})$
      (Using Equation 8)
Return $\mathbf{U}_{0:T}, \mathbf{m_T}$
```

B. Localization Results

The localization results of three different data set that only use the EKF prediction are shown in Figure 2. As we can see from the Figure, trajectories of the IMU are reasonable.

C. Mapping Result

Based on the predicted pose trajectory of the IMU, the EKF mapping is then performed. The landmark features of the different data set are shown in Figure 3. As we can see, the features matches with the pose trajectory well.

D. Visual Inertial SLAM Result

To further evaluate the performance of the proposed SLAM algorithm, the visual inertial slam is also implemented which consists of three steps as described in Section III including EKF prediction, EKF update, and EKF mapping. The trajectories of the IMU pose are shown in Figure 4 and while the Landmark features mapping results are shown in Figure 5. Compared with the pose only obtained from prediction step, the pose obtained from SLAM is more accurate. Similarly, the mapping results of SLAM are also more accurate than that obtained only from EKF mapping.

V. Conclusion

This project investigates using Extended Kalman Filter (EKF) in simultaneous localization and mapping (SLAM) problem. In particular, the robot is localized using the EKF with IMU data which consists of prediction step and update step. Moreover, the environment where the robot is operated is then estimated using EKF with visual information

. The experimental studies are conducted using the real IMU data,images, etc., collected by a ground vehicle. The proposed algorithm shows promising performance in SLAM problem. Moreover, the complete visual inertial SLAM has better performance than pure prediction and mapping

References

[1] Y. Tian, H. Suwoyo, W. Wang, D. Mbemba, and L. Li, “An aekf-slam algorithm with recursive noise statistic based on mle and em,” *Journal of Intelligent & Robotic Systems*, vol. 97, no. 2, pp. 339–355, 2020.

[2] J. Mullane, B.-N. Vo, M. D. Adams, and B.-T. Vo, “A random-finite-set approach to bayesian slam,” *IEEE transactions on robotics*, vol. 27, no. 2, pp. 268–282, 2011.

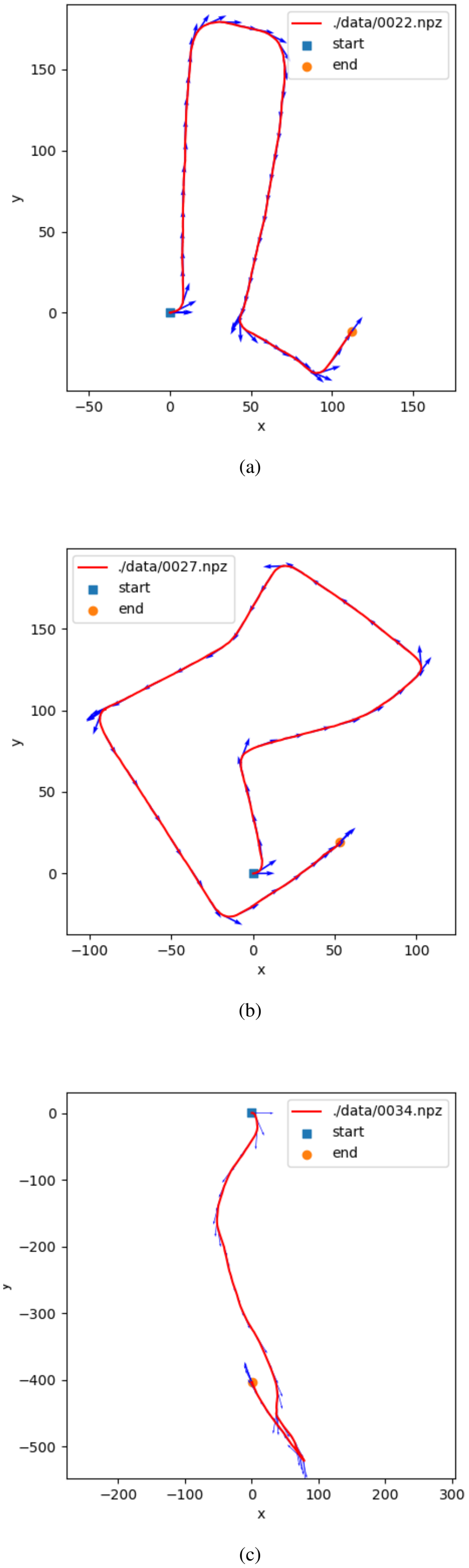
[3] M. Montemerlo, S. Thrun, D. Koller, B. Wegbreit *et al.*, “Fastslam: A factored solution to the simultaneous localization and mapping problem,” *Aaai/iaai*, vol. 593598, 2002.

[4] H. Roth and M. Vona, “Moving volume kinectfusion.” in *BMVC*, vol. 20, no. 2, 2012, pp. 1–11.

[5] G. Grisetti, C. Stachniss, and W. Burgard, “Improved techniques for grid mapping with rao-blackwellized particle filters,” *IEEE transactions on Robotics*, vol. 23, no. 1, pp. 34–46, 2007.

[6] S. Huang and G. Dissanayake, “Convergence and consistency analysis for extended kalman filter based slam,” *IEEE Transactions on robotics*, vol. 23, no. 5, pp. 1036–1049, 2007.

[7] N. Carlevaris-Bianco, M. Kaess, and R. M. Eustice, “Generic node removal for factor-graph slam,” *IEEE Transactions on Robotics*, vol. 30, no. 6, pp. 1371–1385, 2014.

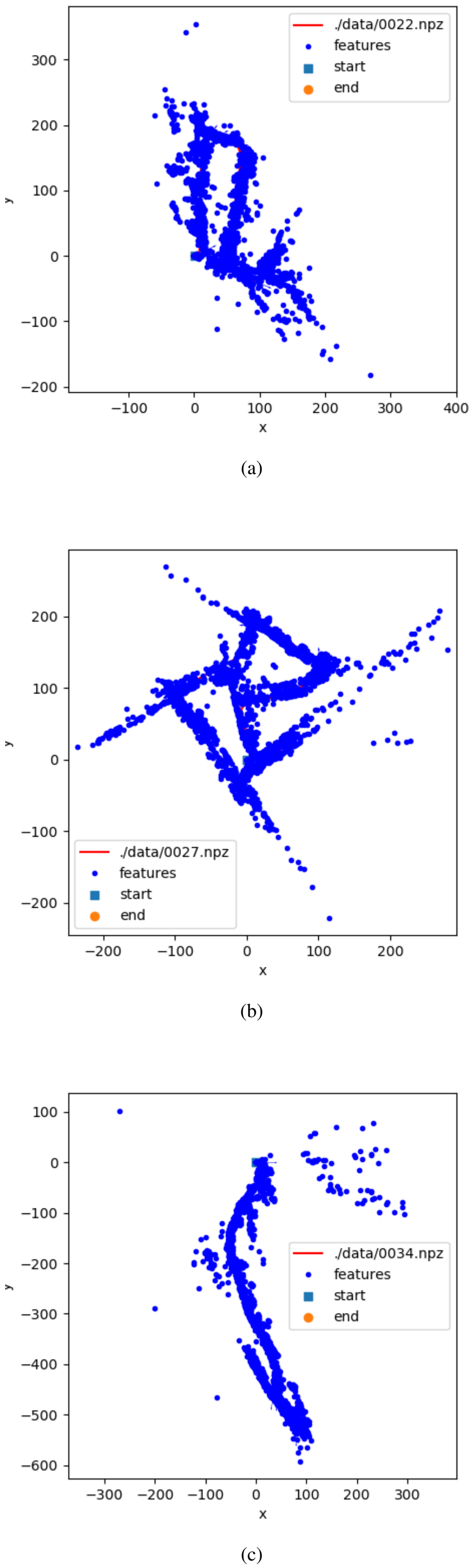


(a)

(b)

(c)

Figure 2. Localization using EKF prediction using data set a) 0022 b) 0027 c) 0034

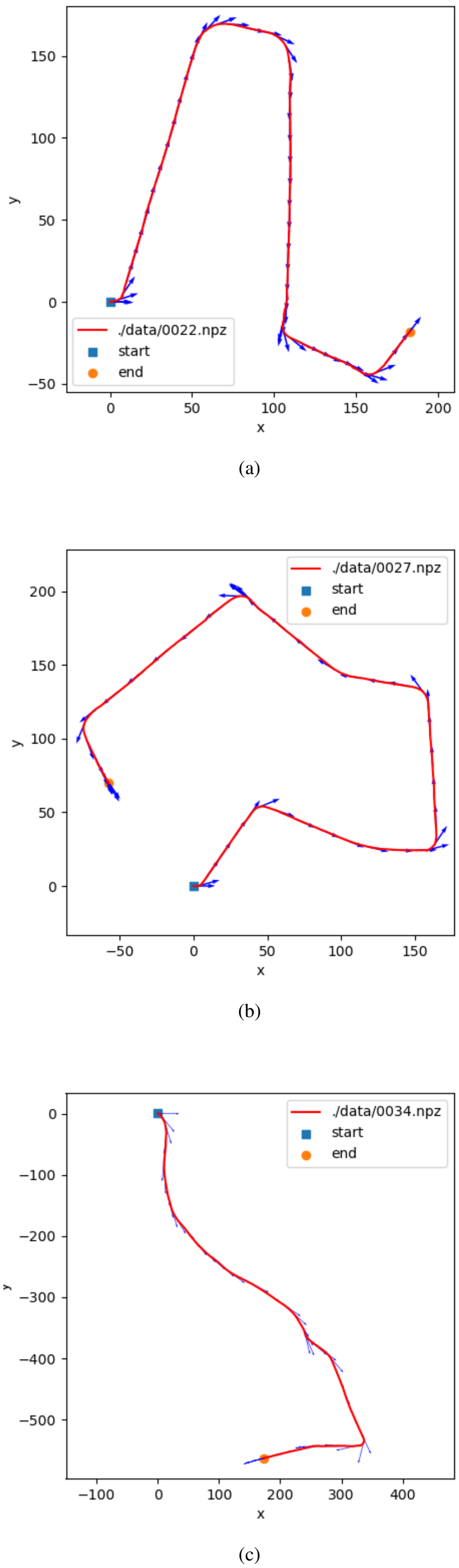


(a)

(b)

(c)

Figure 3. EKF mapping results using data set a) 0022 b) 0027 c) 0034.

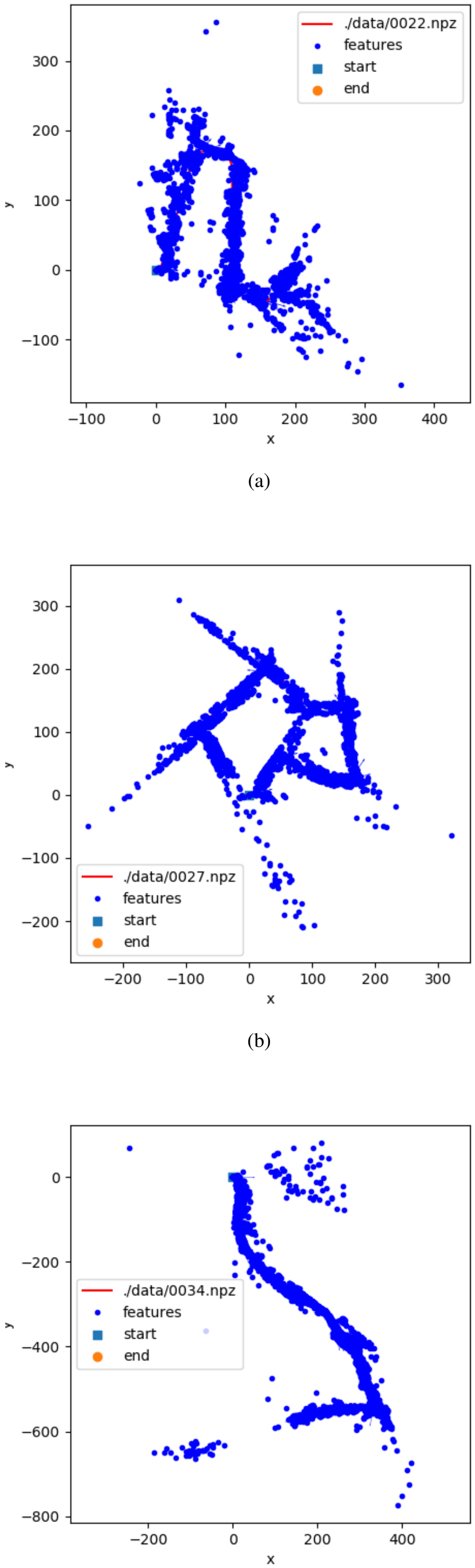


(a)

(b)

(c)

Figure 4. Localization using Visual Inertial SLAM with data set a) 0022 b) 0027 c) 0034



(a)

(b)

(c)

Figure 5. Visual Inertial SLAM results with data set a) 0022 b) 0027 c) 0034.