# ECE276A Project 3 Report

Yuchen Zhang[1]

*Abstract*– **In this project a Simultaneous Localization and Mapping problem based on velocity information from IMU and detected feature points from stereo camera is solved by using the Extended Kalman Filter. Comparison between SLAM and mapping only result revels robustness provided by visual features in localization.**

## I INTRODUCTION

Autonomous driving is desirable to prevent causalities due to driver error. Robust and accurate localization is needed.

In this project, a simultaneous localization and mapping method based on IMU velocity measurement and stereo camera is developed. Resembling how humans sense the world using eyes, the stereo camera information adds robustness to IMU reckoning.

## II PROBLEM FORMULATION

In this project, two approaches for mapping is constructed: visual mapping and visual inertial SLAM. Below describes common assumption(information) available and their difference.

### I Common Assumption / Information

1. Static frame transform ${}_OT_I$ in which $O$ is camera optical frame and $I$ is IMU frame is given.
2. Intrinsic matrix $K = \begin{bmatrix} fs_u & 0 & c_u \\ 0 & fs_v & c_v \\ 0 & 0 & 1 \end{bmatrix}$ and baseline of the stereo camera are known.
3. Known data association between landmark and observed feature. Precisely, the data association function $\Delta_t : \{1, ..., M\} \rightarrow \{1, ..., N_t\}$ that maps landmark number $i$ into observed feature number $j$ is available $\forall t$.
4. Landmarks are static. Or, $m_{t+1} = m_t, \quad \forall t$.
5. Linear and angular velocity measurements from IMU and feature observation $z_t$ synchronized and available $\forall t$

### II Visual Mapping

With additional assumption that the IMU pose ${}_WT_I \in SE(3)$ is known $\forall t$, estimate the landmark positions $\{m_1, ..., m_M\}$ in which $m_i \in \mathbb{R}^3$ and maintain a probability distribution $p(m|z_{1:t}, u_{1:t})$.

### III Visual-Inertial SLAM

Without any additional assumption, estimate the IMU pose ${}_WT_I \in SE(3)$ jointly with the landmark positions $\{m_1, ..., m_M\}$ and maintain a joint probability distribution $p({}_WT_I, m|z_{1:t}, u_{1:t})$

## III TECHNICAL APPROACH

### I Overview

To tackle the visual inertial SLAM problem, I made assumptions and implemented a Extended Kalman Filter to estimate the probability distribution of IMU pose ${}_WT_I$ and landmark locations $\{m_1, \cdots, m_M\}$. The algorithm loops four steps: predict, update, and, state initialization, state removal.

After visual inertial SLAM is completed, slight modification to its implementation yields the dead-reckoning and landmark mapping approach for comparison.

### II State Representation and Assumptions

IMU pose and landmark positions jointly constitutes the EKF state $x = ({}_WT_I, \{m_1, ..., m_M\})$, in which ${}_WT_I \in \text{SE}(3)$ and $m_i \in \mathbb{R}^3$. To utilize the EKF, we made assumption that all states follow a joint gaussian distribution:

$$x = \left[\frac{{}_WT_I}{m}\right] \sim N\left(\underbrace{\left[\frac{\mu_T}{\mu_m}\right]}_{\mu \in (\text{SE3}, \mathbb{R}^{3\times M})}, \underbrace{\begin{bmatrix} \Sigma_{TT} & \Sigma_{Tm} \\ \Sigma_{mT} & \Sigma_{mm} \end{bmatrix}}_{\Sigma \in \mathbb{R}^{(6+3M)\times(6+3M)}}\right)$$

In which $m = \{m_1, ..., m_M\}$ represents the stacked landmark positions, the combined mean $\mu$ consists of a SE(3) portion of mean IMU pose $\mu_T$ and $\mathbb{R}^{3M}$ portion of mean landmark position $\mu_m$. The covariance

$\Sigma$ contains the covariance between the 6 degrees of freedom of IMU pose and $3M$ degrees of freedom of landmark positions.

For the ease of the following sections, extend addition and subtraction operators for the joint state $({}_W T_I, \{m_1, ..., m_M\})$ as follows:

1. $\oplus : (\mathrm{SE}(3), \mathbb{R}^{3M}) \times (\mathbb{R}^6, \mathbb{R}^{3M}) \to (\mathrm{SE}(3), \mathbb{R}^{3M})$

$$\begin{bmatrix}\mu_T \\ m\end{bmatrix} \oplus \begin{bmatrix}\delta\xi \\ \delta m\end{bmatrix} \mapsto \begin{bmatrix}\mu_T \exp([\delta\xi]_\times) \\ m + \delta m\end{bmatrix}$$

2. $\ominus : (\mathrm{SE}(3), \mathbb{R}^{3M}) \times (\mathrm{SE}(3), \mathbb{R}^{3M}) \to (\mathbb{R}^6, \mathbb{R}^{3M})$

$$\begin{bmatrix}\mu_{T2} \\ m_2\end{bmatrix} \ominus \begin{bmatrix}\mu_{T1} \\ m_1\end{bmatrix} \mapsto \begin{bmatrix}\log(\mu_{T1}^{-1}\mu_{T2})^\vee \\ m_2 - m_1\end{bmatrix}$$

With the pair of operators, we can define differentiation as usual:

$$\frac{\partial f(x_t)}{\partial x_t} = \lim_{\delta x \to 0} \frac{f(x_t \oplus \delta x) \ominus f(x_t)}{\|\delta x\|}$$

## III Motion and Observation model

**Motion** $x_{t+1} = f(x_t, u_t, w_t; \tau)$

Given last state $x_t$, control input $u_t = [v_t, \omega_t]^T$, noise $w_t$ and time difference $\tau$, the non-linear motion model relates current and next state $x_{t+1}$ . It is expressed as:

$$\underbrace{\begin{bmatrix}{}_W T_I^{(t+1)} \\ m^{(t+1)}\end{bmatrix}}_{x_{t+1}} = f(x_t, u_t, w_t; \tau)$$

$$= \begin{bmatrix}{}_W T_I^{(t)} \exp([\tau u_t]_\times) \exp(w_t) \\ m^{(t)}\end{bmatrix}$$

In which $w \sim N(0, W)$ is the motion noise with covariance $W$.

Its jacobian w.r.t state and noise is computed as:

$$F = \frac{\partial f}{\partial \mu_t} = \begin{bmatrix}\tilde{F} & 0 \\ 0 & I_{3M}\end{bmatrix} = \begin{bmatrix}\exp(-\tau \overset{\curlywedge}{u}_t) & 0 \\ 0 & I_{3M}\end{bmatrix}$$

**Observation** $\tilde{z}_t = h(x_t, v_t)$

Given current state $x_t$ which contains the IMU pose ${}_W T_I$ and landmark positions $\{m_1, ..., m_M\}$, theoretical observation $\tilde{z}_{t,i} \in \mathbb{R}^4$ is calculated for each landmark.

$$\tilde{z}_{t,i} = M\pi\left({}_O T_I \cdot {}_W T_I^{-1}\ \underline{m}_{t,i}\right) + v_{t,i}$$

In which $M$ is the combined intrinsic matrix

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

$\pi(\cdot) : \mathbb{R}^4 \to \mathbb{R}^4, x \mapsto \frac{x}{x_3}$ is the projection function, ${}_O T_I$ is the static transformation from IMU frame into left camera optical frame.

At any time instance, a total of $N_t$ landmarks are observed and produced the observation $z_t \in \mathbb{R}^{4N_t}$. By assumption, the landmark association function $\Delta_t : \{1, ..., M\} \to \{1, ..., N_t\}$ is given. Its jacobian w.r.t state $x_t = ({}_W T_I, m)$ can be computed as:

$$H({}_W T_I, m) = \frac{\partial z_t}{\partial x_t} = \left[\frac{\partial z_t}{\partial {}_W T_I}, \frac{\partial z_t}{\partial m}\right] \in \mathbb{R}^{4N_t \times (6+3M)}$$

in which,

$$\frac{\partial z_t}{\partial {}_W T_I}_i = -M \frac{\partial \pi}{\partial s}\left({}_O T_I \cdot {}_W T_I^{-1} \underline{m_i}\right) {}_O T_I ({}_W T_I^{-1} \underline{m_i})^{\odot}$$

$$\frac{\partial z_t}{\partial m}_{i,j} = \begin{cases} M \frac{\partial \pi}{\partial s}\left({}_O T_I \cdot {}_W T_I^{-1} \underline{m_i}\right) {}_O T_I \cdot {}_W T_I^{-1} P^T \\ , \Delta_t(j) = i \\ 0 \\ , \Delta_t(j) \neq i \end{cases}$$

$$\frac{\partial \pi}{\partial s} = \frac{1}{s_3} \begin{bmatrix} 1 & 0 & -\frac{s_1}{s_3} & 0 \\ 0 & 1 & -\frac{s_2}{s_3} & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & -\frac{s_4}{s_3} & 1 \end{bmatrix}$$

$$P = [I_3, 0]$$

**Inverse Observation** $m_i = h^{-1}(z_t; {}_W T_I)$

In current setting, it is possible to recover landmark position from observation $z_t$ given IMU pose ${}_W T_I$. This is used in the state initialization step to provide a appropriate landmark position mean $\mu_{m_i}$.

Let $z_{t,i} = [u_L, v_L, u_R, v_R]^T$, the inverse observation model is expressed as:

$$h^{-1}(z_t; {}_W T_I) =$$

$$P\ {}_W T_I \cdot {}_I T_O \left( \begin{bmatrix}0\\0\\0\\1\end{bmatrix} + P^T \frac{f s_u b}{u_L - u_R} K^{-1} \begin{bmatrix} u_L \\ (v_L + v_R)/2 \\ 1 \end{bmatrix} \right)$$

Its jacobian, which is used to initialize the initial landmark covariance, is:

$$
\begin{aligned}
H' &= \frac{\partial}{\partial z_{t,i}} h^{-1}(z_t; {}_W T_I) \\
&= P_W T_I \cdot {}_I T_O P^T K^{-1} \left(\frac{f s_u b}{u_L - u_R}\right) \\
&\times \begin{bmatrix} -\frac{u_R}{u_L - u_R} & 0 & \frac{u_L}{u_L - u_R} & 0 \\ -\frac{v_L + v_R}{2(u_L - u_R)} & \frac{1}{2} & \frac{v_L + v_R}{2(u_L - u_R)} & \frac{1}{2} \\ -\frac{1}{u_L - u_R} & 0 & \frac{1}{u_L - u_R} & 0 \end{bmatrix}
\end{aligned}
$$

## IV State Initialization

The IMU pose is initialized as:

$$
\mu_T = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}
$$

which is not identity since the IMU is installed up-side-down according to the figure.

When any landmark $z_{new}$ is newly observed, its mean and covariance are initialized through the inverse observation model:

$$
\begin{aligned}
\mu_m &\leftarrow [\mu_m, h^{-1}(z_{new}, \mu_T)]^T \\
\Sigma &\leftarrow \begin{bmatrix} \Sigma_{TT} & \Sigma_{Tm} & 0 \\ \Sigma_{mT} & \Sigma_{mm} & 0 \\ 0 & 0 & H'(z_{new}, \mu_T) V H'(z_{new}, \mu_T)^T \end{bmatrix}
\end{aligned}
$$

And the newly initialized state will NOT be updated again in this iteration. I also experimented initializing the covariance to a constant, which produced better result:

$$
\begin{aligned}
\mu_m &\leftarrow [\mu_m, h^{-1}(z_{new}, \mu_T)]^T \\
\Sigma &\leftarrow \begin{bmatrix} \Sigma_{TT} & \Sigma_{Tm} & 0 \\ \Sigma_{mT} & \Sigma_{mm} & 0 \\ 0 & 0 & 0.5 I_3 \end{bmatrix}
\end{aligned}
$$

More details discussed in section results.

## V Prediction

When a IMU data arrives, a prediction step is triggered. Let $u_t = [v_t, \omega_t]^T$ be the generalized velocity:

$$
\begin{aligned}
\mu_{t+1} &= \mu_t \oplus \begin{bmatrix} \tau u_t \\ 0 \end{bmatrix} = \begin{bmatrix} \mu_T^{(t)} \exp([\tau u_t]_\times) \\ \mu_m^{(t)} \end{bmatrix} \\
\Sigma_{t+1} &= \underbrace{\begin{bmatrix} \tilde{F} & 0 \\ 0 & I \end{bmatrix}}_{F} \underbrace{\begin{bmatrix} \Sigma_{TT} & \Sigma_{Tm} \\ \Sigma_{mT} & \Sigma_{mm} \end{bmatrix}}_{\Sigma_t} \underbrace{\begin{bmatrix} \tilde{F} & 0 \\ 0 & I \end{bmatrix}^T}_{F^T} + \begin{bmatrix} W & 0 \\ 0 & 0 \end{bmatrix} \\
&= \begin{bmatrix} \tilde{F} \Sigma_{TT} \tilde{F}^T + W & \tilde{F} \Sigma_{Tm} \\ \Sigma_{mT} \tilde{F}^T & \Sigma_{mm} \end{bmatrix}
\end{aligned}
$$

In which $\tilde{F} = \exp(-\tau \overset{\curlywedge}{u}_t)$

## VI Update

When a stereo observation arrives, an update step is triggered. Let $z_t \in \mathbb{R}^{4N_t}$. With known data assumption, let $\mu'_m$ be obtained from the the EKF state mean $\mu_m$ of these observed features.

$$
\begin{aligned}
\tilde{z}_t &= h(z_t, \mu'_m) \\
K &= \Sigma_t H(\mu_T, \mu_m)^T \left[ H(\mu_T, \mu_m) \Sigma_t H(\mu_T, \mu_m)^T + V \right]^{-1} \\
\mu_t &= \mu_t \oplus [K(z_t - \tilde{z}_t)] \\
\Sigma_t &= (I - K H(\mu_T, \mu_m)) \Sigma_t
\end{aligned}
$$

## VII State Removal

As times goes, EKF will have an increasing number of states that ultimately reach a suboptimal runtime requirement. In my algorithm, landmarks are removed from the state representation if both conditions are true:

1. Landmark $X, Y$ position covariance (read from $\Sigma_t$) $\sigma_{xx}^2 + \sigma_{yy}^2 \leq 0.01$

2. Consecutively unobserved for 300 time steps

## VIII Static Variables

| Variable | Value |
|---|---|
| W | $\begin{bmatrix} 0.001 I_3 & 0 \\ 0 & 0.0005 I_3 \end{bmatrix}$ |
| V | $0.25 I_4$ |

## IX Reduce to Dead-reckoning and Landmark Mapping

### IX.1 Predict

The prediction step is the same as SLAM.

$$
\mu_{t+1} = \mu_t \oplus \begin{bmatrix} \tau u_t \\ 0 \end{bmatrix} = \begin{bmatrix} \mu_T^{(t)} \exp([\tau u_t]_\times) \\ \mu_m^{(t)} \end{bmatrix}
$$

### IX.2 Update

To reduce SLAM back to to dead-reckoning and landmark mapping, I modified the $H$ matrix to

$$H({}_WT_I, m) = \frac{\partial z_t}{\partial x_t} = \left[0 * \frac{\partial z_t}{\partial {}_WT_I}, \frac{\partial z_t}{\partial m}\right]$$

With the assumption that covariance term $\Sigma_{mT}$, $\Sigma_{Tm}$ are all zero at the first time step, the Kalman gain $K$ in update step is computed as:

$$\begin{aligned} K &= \Sigma_t \left[0, \frac{\partial z_t}{\partial m}\right]^T \left[\left[0, \frac{\partial z_t}{\partial m}\right] \begin{bmatrix} \Sigma_{TT} & 0 \\ 0 & \Sigma_{mm} \end{bmatrix} \left[0, \frac{\partial z_t}{\partial m}\right]^T + V\right]^{-1} \\ &= \Sigma_t \left[0, \frac{\partial z_t}{\partial m}\right]^T \left[\tilde{H}\Sigma_{mm}\tilde{H}^T + V\right]^{-1} \\ &= \begin{bmatrix} 0 \\ \Sigma_{mm}\tilde{H}^T \left(\tilde{H}\Sigma_{mm}\tilde{H}^T + V\right)^{-1} \end{bmatrix} \end{aligned}$$

In which $\tilde{H} = \frac{\partial z_t}{\partial m}$. This modification will lead the EKF match all update equations that should be implemented in dead-reckoning and Landmark Mapping approach:

$$\begin{aligned} K &= \Sigma_{mm}\tilde{H}^T \left(\tilde{H}\Sigma_{mm}\tilde{H}^T + V\right)^{-1} \\ \mu_m &= \mu_m + K(z_t - h({}_WT_I, \mu_m)) \\ \Sigma_{mm} &= (I - K\tilde{H})\Sigma_{mm} \end{aligned}$$

## IV RESULTS

### I Dataset processing

Features are scaled down by 3 for runtime.

Notice that my initial IMU pose is

$$\mu_T = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Which corresponds to $z$ axis of IMU pointing down. After plotting the angular velocity measurement and compared with raw camera image, I concluded that the angular velocity is either flipped or not given in IMU frame. *I rotated the linear and angular velocity data to the IMU frame by assuming they are given in the Vehicle frame.*

### II SLAM result

Videos for dead-reckoning and VI-SLAM process are available in video storage folder.

**Dead-reckoning and Landmark Mapping** The final path and mapped landmarks:

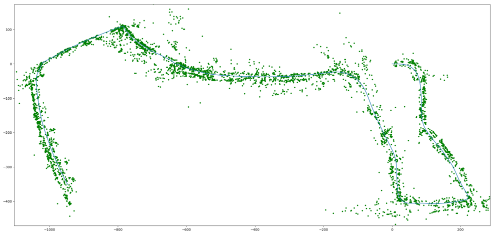

Figure 1: Dead-reckoning and Landmark Mapping

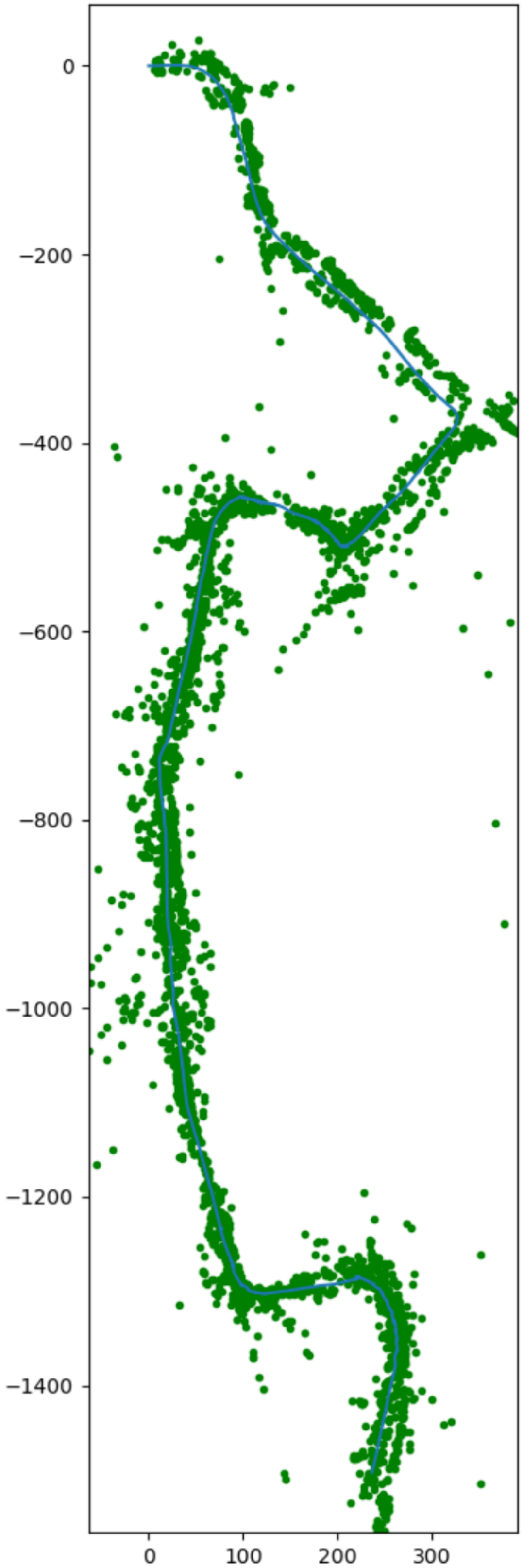


Figure 2: Inverse observation jacobian

### Visual-Inertial SLAM

**Landmark Covariance Initialization** I experimented with two approaches to initialize landmark covariance - either through jacobian of the inverse observation model or a given constant. Below is the result for initializing covariance to a fixed constant:

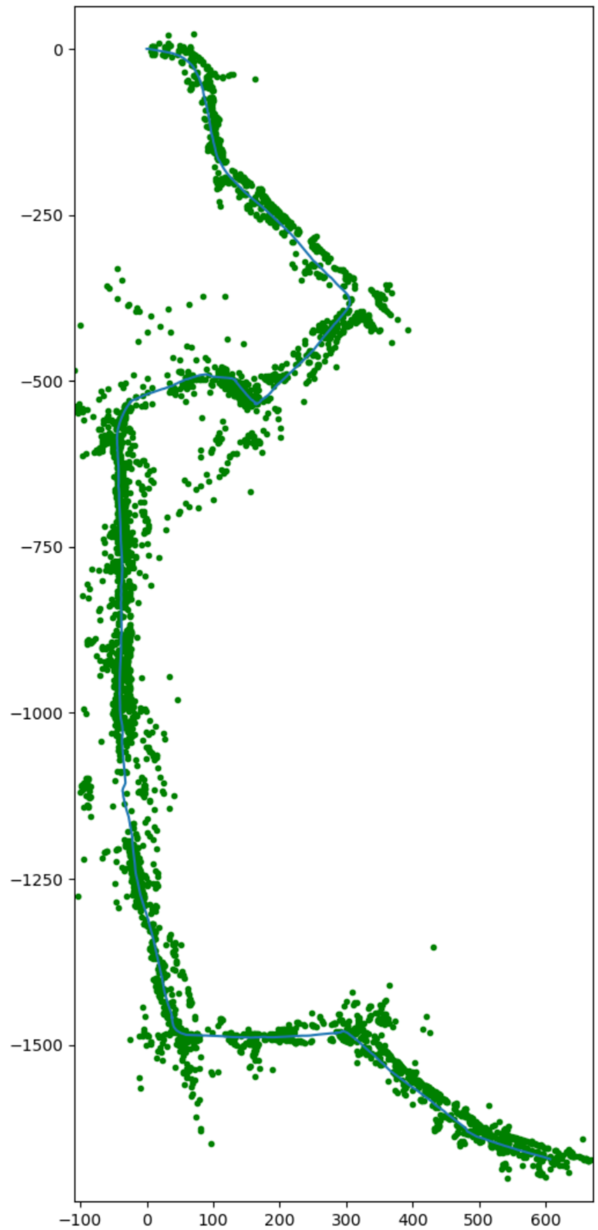


Figure 3: Fixed initial landmark covariance

## III Discussion

Firstly, the result from VI-SLAM differs greatly from that of dead-reckoning and landmark mapping. With human inspection of the camera image, the vehicle made several right-angle turns. Compared to dead-reckoning, VI-SLAM successfully corrected the IMU prediction based on stereo camera feature inputs and caught those right-angle turns. VI-SLAM achieved better result.

Secondly, method of landmark position covariance initialization indeed affect localization result. When initializing with the jacobian of inverse observation and the observation covariance $V$, there exists cases during the SLAM where a landmark is updated to a position behind the camera, which is impossible. This phenomenon occurs mostly during turning. I performed unit test on the jacobian and found no error. Potential cause may be:

1. Biased observation covariance $V$

2. Observation covariance $V$ is not constant - $V$ may be correlated with rotation movement, when scenes in front the camera changes drastically.