Project 3: Visual-Inertial SLAM

Shou-Yu Wang
Department of Electrical and Computer Engineering
UC San Diego
PID: A69030868

***Abstract*—This project addresses the challenge of visual-inertial simultaneous localization and mapping (SLAM) for mobile robots. By fusing inertial measurements with stereo camera data using an extended Kalman filter (EKF), the approach simultaneously estimates the robot's trajectory and the positions of environmental landmarks. The resulting framework is crucial for enabling robust navigation and mapping in complex, real-world environments.**

I. Introduction

Simultaneous localization and mapping (SLAM) is a core problem in robotics, where a robot must build a map of an unknown environment while simultaneously determining its own pose within that map. In many applications—from autonomous vehicles to mobile robotics in cluttered or GPS-denied settings—SLAM is fundamental for safe and efficient operation.

In this project, we focus on visual-inertial SLAM, which integrates high-frequency inertial measurement unit (IMU) data with stereo camera observations. The IMU provides robust motion information through linear and angular velocity measurements, while the stereo camera system offers complementary spatial measurements through visual features. Our approach leverages the extended Kalman filter (EKF) framework to fuse these heterogeneous data sources, yielding a reliable estimate of both the robot's pose (represented as an element of SE(3)) and the 3D positions of landmarks in the environment. By addressing challenges such as sensor noise and calibration uncertainties, our method is designed to enhance localization accuracy and map consistency.

II. Problem Formulation

We aim to estimate the state of a mobile robot and its surrounding landmarks using sensor measurements from an inertial measurement unit (IMU) and stereo cameras. This problem involves two interrelated tasks: predicting the robot's pose over time using IMU data and simultaneously mapping the positions of static landmarks via stereo camera observations.

Inputs

- **IMU Measurements:**
  - Linear velocity: $\mathbf{v}_t \in \mathbb{R}^3$
  - Angular velocity: $\boldsymbol{\omega}_t \in \mathbb{R}^3$
- **Time Stamps:**
  - τ_t (in seconds, UNIX time)
- **Stereo Camera Observations:**
  - Visual feature measurements: $\mathbf{z}_t \in \mathbb{R}^{4\times M}$ where each column contains the pixel coordinates

    $$\begin{bmatrix} u_l \\ v_l \\ u_r \\ v_r \end{bmatrix}$$

    for features detected in the left and right camera images. A measurement of $[-1,-1,-1,-1]^\top$ indicates a missing observation.
- **Calibration Parameters:**
  - Intrinsic calibration matrices: $K_l, K_r \in \mathbb{R}^{3\times 3}$ for the left and right cameras.
  - Extrinsic transformations: $T_{\text{extL}}, T_{\text{extR}} \in \mathrm{SE}(3)$ representing the transformation from the IMU frame to the left and right camera frames.

Outputs

- **IMU Pose Trajectory:**
  - A sequence of poses $\{T_t\}_{t=0}^{T}$ where each $T_t \in \mathrm{SE}(3)$ represents the robot's orientation and position.
- **Landmark Positions:**
  - A set of 3D landmark positions $\{\mathbf{m}_i\}_{i=1}^{M}$ with $\mathbf{m}_i \in \mathbb{R}^3$, obtained initially via triangulation and refined through EKF updates.

Mathematical Formulation

- **State Representation:**
  - **IMU Pose:** The robot's state is represented as a homogeneous transformation $T_t \in \mathrm{SE}(3)$.
  - **Landmark State:** Each landmark is modeled by a 3D point $\mathbf{m}_i \in \mathbb{R}^3$.
- **State Propagation (IMU Prediction):**

  $$T_{t+1} = T_t \cdot \exp\left(\Delta t \begin{bmatrix} \mathbf{v}_t \\ \boldsymbol{\omega}_t \end{bmatrix}^\wedge\right),$$

  where Δt is the time interval between measurements and $(\cdot)^\wedge$ denotes the operator mapping a 6D twist vector to a 4×4 matrix in $\mathfrak{se}(3)$.
- **Error Covariance Propagation:**

  $$P_{t+1} = F_t\, P_t\, F_t^\top + G_t\, Q\, G_t^\top,$$

  where $F_t \approx I - \text{axangle2adtwist}(\mathbf{u}_t)$ is the state transition matrix (obtained via a first-order approximation), $G_t =$

$\Delta t\, I$ is the noise Jacobian, and Q is the process noise covariance matrix.

- **Landmark Observation Model:** The predicted stereo measurement for a landmark $\mathbf{m}_i$ is given by:

$$\mathbf{z}_{\text{pred}} = h\left(T_t, \mathbf{m}_i, K_l, K_r, T_{\text{extL}}, T_{\text{extR}}\right),$$

  where the function $h(\cdot)$ projects the 3D landmark into the pixel coordinates of both cameras.

- **Cost Function and Goal:** The objective is to minimize the residual error between the observed and predicted measurements:

$$\mathbf{r} = \mathbf{z}_t - \mathbf{z}_{\text{pred}},$$

  by applying the Extended Kalman Filter (EKF) framework. In the EKF update, the landmark states and the IMU pose are iteratively refined by linearizing the observation model and minimizing the measurement residual.

III. Technical Approach

Our approach to visual-inertial SLAM is divided into four main parts:

A. IMU Localization via EKF Prediction

The first step in our visual-inertial SLAM approach is to predict the robot's pose using the IMU measurements. This is achieved by propagating the state through an Extended Kalman Filter (EKF) prediction step based on SE(3) kinematics. Note that we incorporate custom process noise parameters tailored to the IMU sensor characteristics, which are critical for accurately modeling uncertainties in the prediction.

a) 1. IMU Measurement Integration: At each timestep, the IMU provides:

$$\mathbf{v} \in \mathbb{R}^3, \quad \mathbf{w} \in \mathbb{R}^3,$$

which are the linear and angular velocities in the IMU frame, respectively. Given a small time interval Δt, we define the discretized twist vector as:

$$\mathbf{u} = \begin{bmatrix} \mathbf{v}\,\Delta t \\ \mathbf{w}\,\Delta t \end{bmatrix} \in \mathbb{R}^6.$$

b) 2. Exponential Map and Incremental Pose: The twist $\mathbf{u}$ is mapped to an incremental transformation in SE(3) using the exponential map:

$$\Delta T = \exp\left(\mathbf{u}^\wedge\right),$$

where $\mathbf{u}^\wedge \in \mathfrak{se}(3)$ is the 4×4 matrix representation (the "hat" operator) of $\mathbf{u}$. The new pose is then predicted by right-multiplying the previous pose:

$$T_{\text{pred}} = T_{\text{prev}} \cdot \Delta T.$$

c) 3. Covariance Propagation and Custom Noise: Let $P_{\text{prev}} \in \mathbb{R}^{6\times 6}$ be the error covariance at the previous timestep and $Q \in \mathbb{R}^{6\times 6}$ be the process noise covariance matrix. In our implementation, Q is custom-designed to capture the uncertainties specific to the IMU sensor characteristics (e.g., sensor noise in linear and angular velocity measurements). We first approximate the state transition matrix using a first-order Taylor expansion:

$$F \approx I_6 - A,$$

where

$$A = \text{axangle2adtwist}(\mathbf{u}) \in \mathbb{R}^{6\times 6}.$$

Assuming that the process noise enters linearly with Δt, we define the noise Jacobian as:

$$G = \Delta t \cdot I_6.$$

The predicted error covariance is then updated by:

$$P_{\text{pred}} = F\, P_{\text{prev}}\, F^\top + G\, Q\, G^\top.$$

d) Summary of Equations:

- **Discretized twist:**

$$\mathbf{u} = \begin{bmatrix} \mathbf{v}\,\Delta t \\ \mathbf{w}\,\Delta t \end{bmatrix}.$$

- **Incremental pose:**

$$\Delta T = \exp\left(\mathbf{u}^\wedge\right).$$

- **Pose update:**

$$T_{\text{pred}} = T_{\text{prev}} \cdot \Delta T.$$

- **State transition:**

$$F \approx I_6 - \text{axangle2adtwist}(\mathbf{u}).$$

- **Covariance update:**

$$P_{\text{pred}} = F\, P_{\text{prev}}\, F^\top + (\Delta t\, I_6)\, Q\, (\Delta t\, I_6)^\top.$$

These equations form the mathematical backbone of the EKF prediction step, allowing us to propagate both the pose and its associated uncertainty using the IMU measurements, with custom noise parameters ensuring that the model accurately reflects sensor-specific uncertainties.

B. Landmark Mapping via EKF Update

In this step, each landmark's 3D position is refined using stereo measurements and an EKF update. As in the IMU prediction step, we also introduce custom noise parameters to model the uncertainties in the visual measurements. The process consists of landmark initialization, measurement prediction, Jacobian calculation, and the EKF update equations, followed by a threshold-based outlier rejection to ensure robustness.

a) 1. Landmark Initialization via Triangulation: For a stereo measurement vector

$$z = \begin{bmatrix} u_l \\ v_l \\ u_r \\ v_r \end{bmatrix},$$

the 3D position of a landmark $m \in \mathbb{R}^3$ is first estimated by triangulation. Using the projection matrices of the left and right cameras, P_l and P_r, we construct a linear system:

$$\begin{aligned} A_0 &= u_l \, P_l(2,:) - P_l(0,:), \\ A_1 &= v_l \, P_l(2,:) - P_l(1,:), \\ A_2 &= u_r \, P_r(2,:) - P_r(0,:), \\ A_3 &= v_r \, P_r(2,:) - P_r(1,:), \end{aligned}$$

and stack them into a 4×4 matrix A:

$$A = \begin{bmatrix} A_0 \\ A_1 \\ A_2 \\ A_3 \end{bmatrix}.$$

The homogeneous solution m_h is found by solving:

$$A \, m_h = 0,$$

and the Cartesian coordinate is obtained by dehomogenization:

$$m = \frac{[m_h]_{1:3}}{[m_h]_4}.$$

b) 2. Measurement Prediction: Given the landmark position m and the current IMU pose T_{imu}, we predict the stereo measurement by projecting m into both cameras. The projection for the left and right cameras is:

$$p_l = P_l \begin{bmatrix} m \\ 1 \end{bmatrix}, \quad h_l(m) = \begin{bmatrix} \frac{p_{l,x}}{p_{l,z}} \\ \frac{p_{l,y}}{p_{l,z}} \end{bmatrix},$$

$$p_r = P_r \begin{bmatrix} m \\ 1 \end{bmatrix}, \quad h_r(m) = \begin{bmatrix} \frac{p_{r,x}}{p_{r,z}} \\ \frac{p_{r,y}}{p_{r,z}} \end{bmatrix}.$$

Thus, the predicted measurement is:

$$h(m) = \begin{bmatrix} h_l(m) \\ h_r(m) \end{bmatrix} \in \mathbb{R}^4.$$

c) 3. Computation of the Residual: The measurement residual is given by:

$$r = z - h(m).$$

d) 4. Jacobian Calculation: The Jacobian $H \in \mathbb{R}^{4\times3}$ of the measurement function $h(m)$ with respect to the landmark position m is computed by differentiating the perspective projection for each camera. For the left camera:

$$\frac{\partial h_l}{\partial m} = \begin{bmatrix} \frac{1}{p_{l,z}} & 0 & -\frac{p_{l,x}}{p_{l,z}^2} \\ 0 & \frac{1}{p_{l,z}} & -\frac{p_{l,y}}{p_{l,z}^2} \end{bmatrix},$$

and similarly for the right camera:

$$\frac{\partial h_r}{\partial m} = \begin{bmatrix} \frac{1}{p_{r,z}} & 0 & -\frac{p_{r,x}}{p_{r,z}^2} \\ 0 & \frac{1}{p_{r,z}} & -\frac{p_{r,y}}{p_{r,z}^2} \end{bmatrix}.$$

Stacking these gives:

$$H = \begin{bmatrix} \frac{\partial h_l}{\partial m} \\ \frac{\partial h_r}{\partial m} \end{bmatrix}.$$

e) 5. EKF Update Equations: Let $P \in \mathbb{R}^{3\times3}$ be the current covariance of the landmark. With the measurement noise covariance $R \in \mathbb{R}^{4\times4}$, where R is custom-tuned based on the stereo camera's noise characteristics, the Kalman gain is computed as:

$$K = P \, H^\top \left(H \, P \, H^\top + R \right)^{-1}.$$

The updated landmark position and covariance are then:

$$m_{\text{new}} = m + K \, r,$$

$$P_{\text{new}} = (I_3 - K \, H) \, P.$$

f) 6. Outlier Rejection via Thresholding: The dataset contains a large number of visual feature measurements, but not all contribute positively to the SLAM solution. To improve robustness, we incorporate a threshold-based outlier rejection mechanism in the EKF update for landmark mapping.

g) 7. Threshold-Based Rejection: After updating a landmark's position m via the EKF update, we compare the Euclidean distance between m_{new} and the current IMU position (extracted from T_{imu}):

$$d = \|m_{\text{new}} - T_{\text{imu}}(1:3, 4)\|.$$

If this distance exceeds a predefined threshold (e.g., $d > d_{\text{th}}$), the update is rejected:

$$\text{if } d > d_{\text{th}}, \quad \text{reject update (or remove the landmark).}$$

This strategy helps in discarding landmarks that become inconsistent with the current IMU estimate.

h) 8. Rationale and Causes of Outliers: Outliers may arise due to several reasons:

- **Noisy Measurements:** Visual feature measurements can be corrupted by sensor noise, leading to erroneous pixel coordinates.
- **Poor Feature Matching:** Incorrect correspondences between stereo images or across time frames can result in mismatches.
- **Calibration Errors:** Inaccuracies in the intrinsic or extrinsic calibration parameters may produce inconsistent projections.
- **Occlusions and Dynamic Elements:** Features on moving objects or those partially occluded may yield inconsistent measurements.

By applying the threshold, we ensure that only reliable landmarks, which remain close to the IMU position, are used for further SLAM updates. This selective strategy, along with the use of custom measurement noise parameters, helps to maintain a robust and computationally efficient system.

i) Summary of Equations:

- **Triangulation:**

$$A\,m_h = 0, \quad m = \frac{[m_h]_{1:3}}{[m_h]_4}.$$

- **Measurement Prediction:**

$$h(m) = \begin{bmatrix} \frac{p_{l,x}}{p_{l,z}} \\ \frac{p_{l,y}}{p_{l,z}} \\ \frac{p_{r,x}}{p_{r,z}} \\ \frac{p_{r,y}}{p_{r,z}} \end{bmatrix}.$$

- **Residual:**

$$r = z - h(m).$$

- **Jacobian:**

$$H = \begin{bmatrix} \frac{\partial h_l}{\partial m} \\ \frac{\partial h_r}{\partial m} \end{bmatrix}.$$

- **Kalman Gain:**

$$K = P\,H^\top \left(H\,P\,H^\top + R\right)^{-1}.$$

- **Update:**

$$m_{\text{new}} = m + K\,r, \quad P_{\text{new}} = (I_3 - K\,H)\,P.$$

This formulation, derived from our implementation in `ekf_update.py`, forms the mathematical basis for updating the landmark states in the visual-inertial SLAM system while robustly rejecting outliers.

C. Visual-Inertial SLAM

The final SLAM system cyclically combines the IMU prediction with landmark updates and uses the refined landmarks to correct the IMU pose. Custom noise parameters are applied at each stage to account for sensor-specific uncertainties, ensuring robust and accurate estimation. At each timestep t, the following steps are executed:

a) 1. IMU Pose Prediction (EKF Prediction): Using the previous pose T_{t-1} and IMU measurements, the predicted pose is computed as:

$$T_t^{\text{pred}} = T_{t-1} \cdot \exp\left(\Delta t \begin{bmatrix} \mathbf{v}_{t-1} \\ \boldsymbol{\omega}_{t-1} \end{bmatrix}^\wedge\right),$$

with the corresponding error covariance:

$$P_t^{\text{pred}} = F\,P_{t-1}\,F^\top + G\,Q\,G^\top,$$

where

$$F \approx I - \text{axangle2adtwist}(\mathbf{u}_{t-1}), \quad G = \Delta t\,I.$$

Here, Q is the custom-tuned process noise covariance matrix that reflects the uncertainties in the IMU measurements.

b) 2. Landmark Update (EKF Update for Landmarks): For each landmark i observed in the stereo images, with measurement

$$\mathbf{z}_{t,i} = \begin{bmatrix} u_l \\ v_l \\ u_r \\ v_r \end{bmatrix},$$

the current estimate $\mathbf{m}_{i,t}$ is updated as follows:

$$\mathbf{z}_{\text{pred},i} = h\left(T_t^{\text{pred}}, \mathbf{m}_{i,t}, K_l, K_r, T_{\text{extL}}, T_{\text{extR}}\right),$$

$$\mathbf{r}_i = \mathbf{z}_{t,i} - \mathbf{z}_{\text{pred},i},$$

and the corresponding EKF update is:

$$K_i = P_{i,t}\,H_i^\top \left(H_i\,P_{i,t}\,H_i^\top + R\right)^{-1},$$

$$\mathbf{m}_{i,t}^{\text{new}} = \mathbf{m}_{i,t} + K_i\,\mathbf{r}_i, \quad P_{i,t}^{\text{new}} = (I - K_i\,H_i)\,P_{i,t},$$

where H_i is the Jacobian of the measurement function with respect to $\mathbf{m}_i$ and R is the custom-tuned measurement noise covariance matrix for the stereo camera. In addition, a threshold-based rejection is applied: if the Euclidean distance

$$d = \|\mathbf{m}_{i,t}^{\text{new}} - T_{\text{imu}}(1:3,4)\|$$

exceeds a predefined threshold d_{th}, the update for that landmark is rejected.

c) 3. IMU Pose Correction via Visual-Inertial Fusion: To reduce drift and refine the predicted IMU pose using the updated landmarks, we incorporate a visual measurement update. For all valid landmark observations at timestep t, we:

1) Compute the overall measurement residual:

$$\mathbf{r} = \mathbf{z}_t - h\left(T_t^{\text{pred}}, \{\mathbf{m}_{i,t}^{\text{new}}\}\right),$$

where $\mathbf{z}_t$ stacks all stereo measurements and $h(\cdot)$ projects the corresponding landmarks.

2) Numerically estimate the Jacobian J_T of the measurement function with respect to the IMU pose:

$$J_T = \left.\frac{\partial h(T, \{\mathbf{m}_{i,t}^{\text{new}}\})}{\partial T}\right|_{T=T_t^{\text{pred}}}.$$

3) Compute the gain:

$$K_T = P_t^{\text{pred}}\,J_T^\top \left(J_T\,P_t^{\text{pred}}\,J_T^\top + R_{\text{meas}}\right)^{-1},$$

where R_{meas} is the aggregated measurement noise covariance, tuned for the visual sensor.

4) Update the IMU pose:

$$\delta\mathbf{x} = K_T\,\mathbf{r},$$

$$T_t^{\text{updated}} = T_t^{\text{pred}} \cdot \exp\left((\delta\mathbf{x})^\wedge\right),$$

and update the covariance:

$$P_t^{\text{updated}} = (I - K_T\,J_T)\,P_t^{\text{pred}}.$$

d) 4. Summary of the Iterative Process: At each timestep, the visual-inertial SLAM system performs:

1) **Prediction:** Compute T_t^{pred} and P_t^{pred} using IMU data with custom process noise.
2) **Landmark Update:** Refine each landmark $\mathbf{m}_{i,t}$ using EKF update equations with custom measurement noise and threshold-based outlier rejection.
3) **Pose Correction:** Adjust the predicted pose T_t^{pred} via a visual update using aggregated residuals and numerical Jacobian estimation.

This integration of IMU and visual measurements, with careful tuning of noise parameters at each stage, forms a robust framework that continuously corrects for drift and enhances the accuracy of both the pose and the landmark estimates.

IV. Results

In this section, we present and analyze the outcomes of each major step in our visual-inertial SLAM pipeline. We begin with the IMU-only trajectory estimation (EKF prediction), followed by landmark mapping via EKF update, and finally demonstrate the fully integrated visual-inertial SLAM results. In our implementation, we set the process noise covariance Q based on typical IMU sensor specifications: for instance, a linear velocity standard deviation of $\sigma_v = 0.1$ m/s and an angular velocity standard deviation of $\sigma_\omega = 0.01745$ rad/s (approximately 1°). For each dataset (`00`, `01`, and `02`), we provide plots and a brief discussion of the performance.

A. IMU Localization via EKF Prediction

1) Dataset `00`: Figure 1 shows the estimated IMU trajectory for `dataset00` using only the EKF prediction step. The red curve represents the predicted path, while the blue segments indicate orientation samples (drawn at fixed intervals). The starting and ending positions are marked with a square and a circle, respectively. We observe that the overall trajectory remains coherent, with minimal drift for this dataset.

2) Dataset `01`: In Figure 2, we show the trajectory estimation for `dataset01`. There is a more pronounced drift in certain segments, likely due to faster and more complex robot maneuvers. However, the overall shape of the trajectory is still consistent with the path that the robot traversed.

3) Dataset `02`: Finally, Figure 3 illustrates the predicted IMU trajectory for `dataset02`. This dataset contains more significant turns and potential occlusions in the environment. While the EKF prediction alone does exhibit drift toward the end of the run, the overall path still follows the general ground-truth route.

a) Discussion: Across all three datasets, the IMU-based prediction offers a reasonably good initial pose estimate. The custom noise parameters help capture the uncertainties in linear and angular velocity measurements, which is crucial for maintaining stability in the filter. Nevertheless, we observe that longer trajectories tend to accumulate drift, highlighting the importance of incorporating visual measurements to correct the pose estimate and reduce cumulative errors.

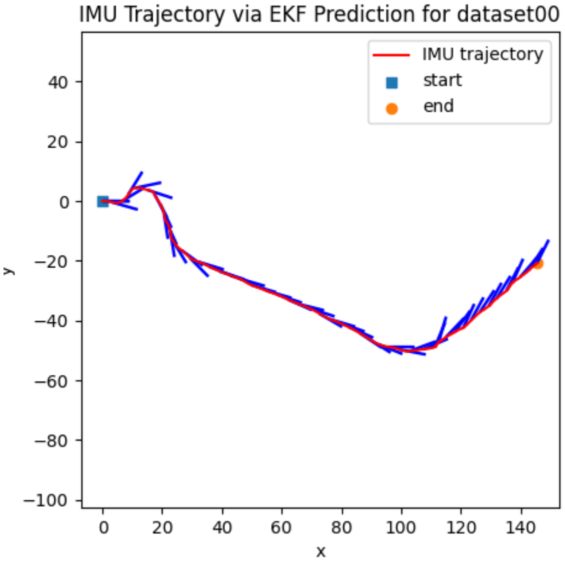

Fig. 1: IMU trajectory via EKF prediction for `dataset00`.

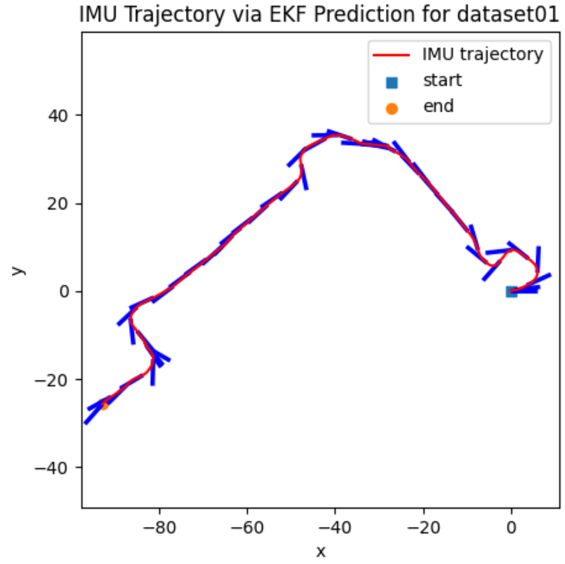

Fig. 2: IMU trajectory via EKF prediction for `dataset01`.

B. Landmark Mapping via EKF Update

In this section, we demonstrate the results of landmark mapping using stereo observations, assuming that the IMU trajectory is fixed and does not receive corrections from the visual measurements. We initialize and update landmarks at each timestep via EKF, applying threshold-based outlier rejection to discard erroneous features. For datasets that include stereo feature measurements, we set:

- $\sigma_{\text{px}} = 4$, leading to a measurement noise covariance $R = \text{diag}(16, 16, 16, 16)$.

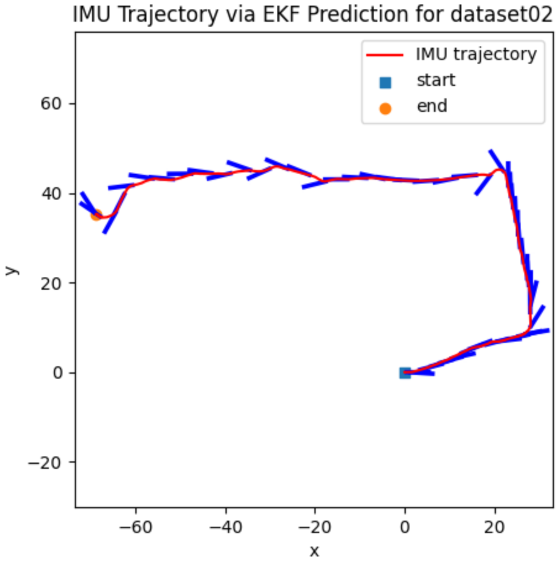

Fig. 3: IMU trajectory via EKF prediction for `dataset02`.

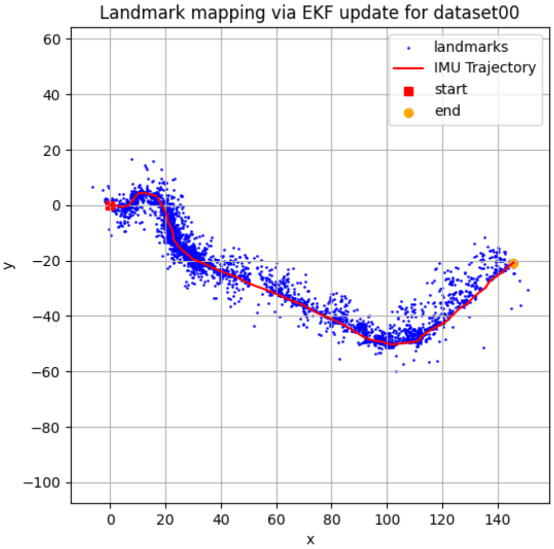

Fig. 4: Landmark mapping via EKF update for `dataset00` with threshold $= 20$.

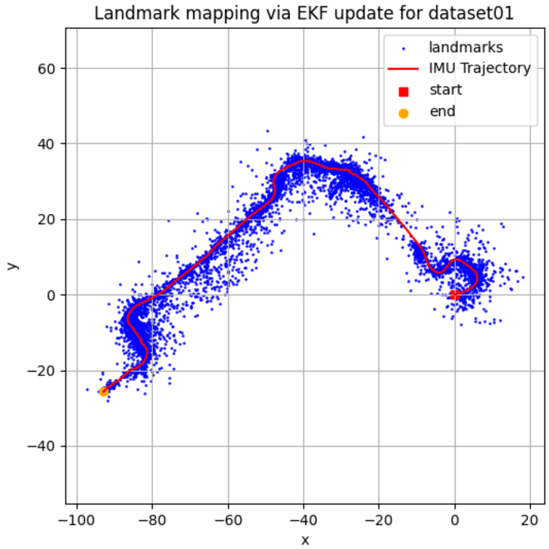

Fig. 5: Landmark mapping via EKF update for `dataset01` with threshold $= 20$.

- A large initial covariance for newly initialized landmarks, init_cov $= 10000 \cdot I_{3\times3}$.
- An outlier rejection threshold threshold $= 20$, where any landmark whose updated position is more than 20 m away from the IMU pose is discarded.

1) Dataset `00`: Figure 4 shows the 2D projection of the updated landmark positions (blue dots) alongside the IMU trajectory (red line) for `dataset00`. The start and end poses are marked with a square and a circle, respectively. We observe that the landmarks are densely populated around the robot's trajectory, indicating that the stereo measurements successfully track multiple features in the environment.

2) Dataset `01`: Figure 5 illustrates a similar result for `dataset01`. Although the robot undergoes more complex maneuvers and the path covers a larger area, the landmark positions remain reasonably consistent around the estimated IMU path. The threshold-based outlier rejection effectively removes spurious landmarks that might arise from poor feature matching or noisy observations.

3) Dataset `02`: For `dataset02`, we did not have pre-computed feature tracks; hence, without completing the optional feature detection and matching step, we were unable to perform the EKF update for landmarks. Consequently, no landmark mapping results are shown for `dataset02`.

a) Discussion: The results indicate that initializing landmarks with a large covariance ($10000\,I$) allows the filter to gradually converge as measurements are accumulated. By setting $\sigma_{\text{px}} = 4$, we assume moderate pixel noise, which helps avoid overly aggressive updates that could destabilize landmark estimates. Meanwhile, the outlier rejection threshold of 20 m provides a balance between retaining legitimate landmarks and discarding clearly inconsistent ones. This strategy is especially useful in challenging conditions where some measurements might be corrupted by dynamic objects, occlusions, or inaccurate stereo matching. Overall, the EKF landmark mapping step produces a dense set of features around the robot's trajectory, paving the way for further refinement in the full visual-inertial SLAM integration.

C. Visual-Inertial SLAM

In the final stage, we integrate the IMU prediction with landmark-based updates to simultaneously refine the IMU pose and map the environment. By allowing the pose to be corrected via stereo observations, we reduce the drift that accumulates in IMU-only estimation.

1) Dataset `00`*:* Figure 6 illustrates the visual-inertial SLAM result for `dataset00`. The green line indicates the predicted IMU trajectory (from the purely inertial EKF prediction), the red line shows the updated trajectory after incorporating stereo measurements, and the blue dots represent the estimated landmark positions. We observe that two trajectories are almost aligned, but the updated one has slightly more subtle maneuvers, showing more detailed movement through updating step.

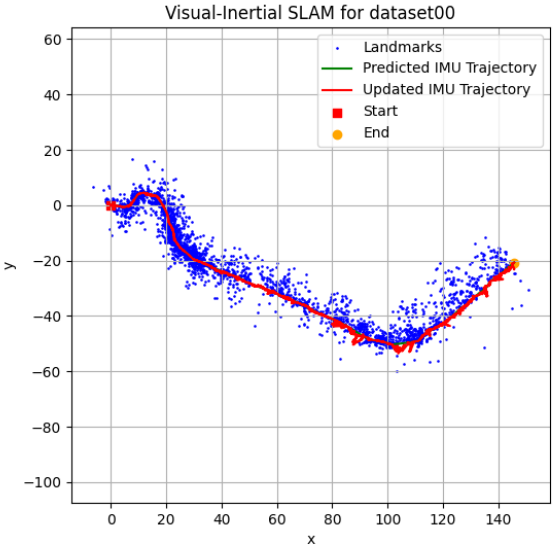

Fig. 6: Visual-inertial SLAM for `dataset00`. Green: predicted IMU trajectory; red: updated IMU trajectory; blue: landmarks.

2) Dataset `01`*:* A similar result for `dataset01` is shown in Figure 7. The robot traverses a more extended path with more pronounced maneuvers, with two trajectories almost overlap. The landmarks positions remained quite stable, which correspond to actual scenario as well.

a) Discussion: By comparing the predicted IMU trajectory (green) with the updated trajectory (red), we observe that the visual measurements serve as an external reference to pull the state estimate back toward a more consistent global frame. The landmarks, initialized and refined through stereo observations, provide spatial constraints that limit the IMU's unbounded drift. Specifically:

- **Reduced Pose Error:** The updated trajectory exhibits notably less drift in both `dataset00` and `dataset01`, particularly over longer time horizons.
- **Consistent Landmark Distribution:** Landmarks cluster around the robot's actual path, reinforcing the validity of the corrected pose estimate.

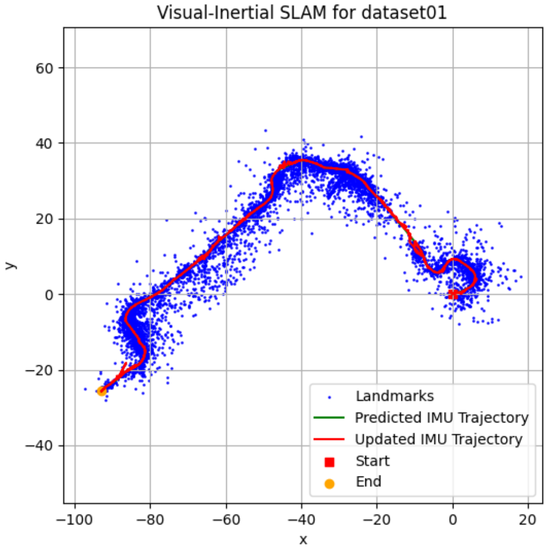

Fig. 7: Visual-inertial SLAM for `dataset01`. Green: predicted IMU trajectory; red: updated IMU trajectory; blue: landmarks.

- **Effect of Noise Tuning:** Proper tuning of both the IMU process noise and the stereo measurement noise is essential for achieving stable corrections without causing filter divergence.

In summary, the visual-inertial fusion effectively balances the high-frequency inertial predictions with the spatially accurate but lower-frequency stereo updates, yielding a more robust and accurate SLAM solution.

V. Conclusion and Future Work

In this project, we presented a visual-inertial SLAM framework that fuses high-frequency IMU data with stereo camera measurements using an Extended Kalman Filter (EKF). Our approach predicts the robot's pose via IMU-only estimation, refines landmark positions using stereo observations with custom noise parameters, and integrates these components to correct drift in the pose estimate. The qualitative results on datasets `00` and `01` demonstrate that the updated trajectory aligns more closely with the actual path, while the landmark distribution provides additional spatial constraints.

A. Discussion and Limitations

Although our system shows promising results, there are several areas where the report and implementation could be improved:

- **Quantitative Evaluation:** The current analysis is mainly qualitative. Future evaluations should include quantitative metrics (e.g., RMSE against ground truth) to rigorously assess performance.
- **Feature Dependence:** The system's performance relies heavily on the quality of stereo feature measurements. As

observed with `dataset02`, incomplete feature detection and matching prevent effective landmark mapping.

- **Computational Scalability:** The EKF update for a large number of landmarks can be computationally intensive. Efficient data management or sparsification techniques are necessary for scalability.
- **Calibration Sensitivity:** Small inaccuracies in intrinsic and extrinsic calibration can significantly affect both triangulation and pose correction.
- **Noise Parameter Tuning:** While custom process and measurement noise parameters are used, their manual tuning could be improved by adaptive strategies.

B. Future Work

Based on the current findings, several avenues for future research are identified:

- **Adaptive Noise Tuning:** Develop methods to dynamically adjust process and measurement noise parameters based on motion dynamics and real-time sensor feedback.
- **Enhanced Feature Detection:** Integrate more robust feature detection and matching algorithms to improve landmark tracking, particularly in challenging scenarios with sparse or occluded features.
- **Optimization Back-End:** Incorporate advanced non-linear optimization frameworks (e.g., factor graphs or bundle adjustment) to refine the state estimates and ensure global consistency.
- **Loop Closure:** Implement loop closure detection to correct long-term drift and maintain a consistent global map over extended operations.
- **Real-Time Processing:** Explore parallel processing or more efficient implementations to achieve real-time performance, especially for high-dimensional landmark states.

Overall, our work demonstrates the potential of visual-inertial SLAM for robust localization and mapping in complex environments. Despite certain limitations, the integration of custom noise parameters and threshold-based outlier rejection provides a strong foundation for future improvements in autonomous navigation.