iLoco: Real-Time Visual SLAM Using an iPhone

EEECS 568: Mobile Robotics - Group 1

Adeep Das, Velu Manohar, Muhammad Khan, Nikhil Sridhar

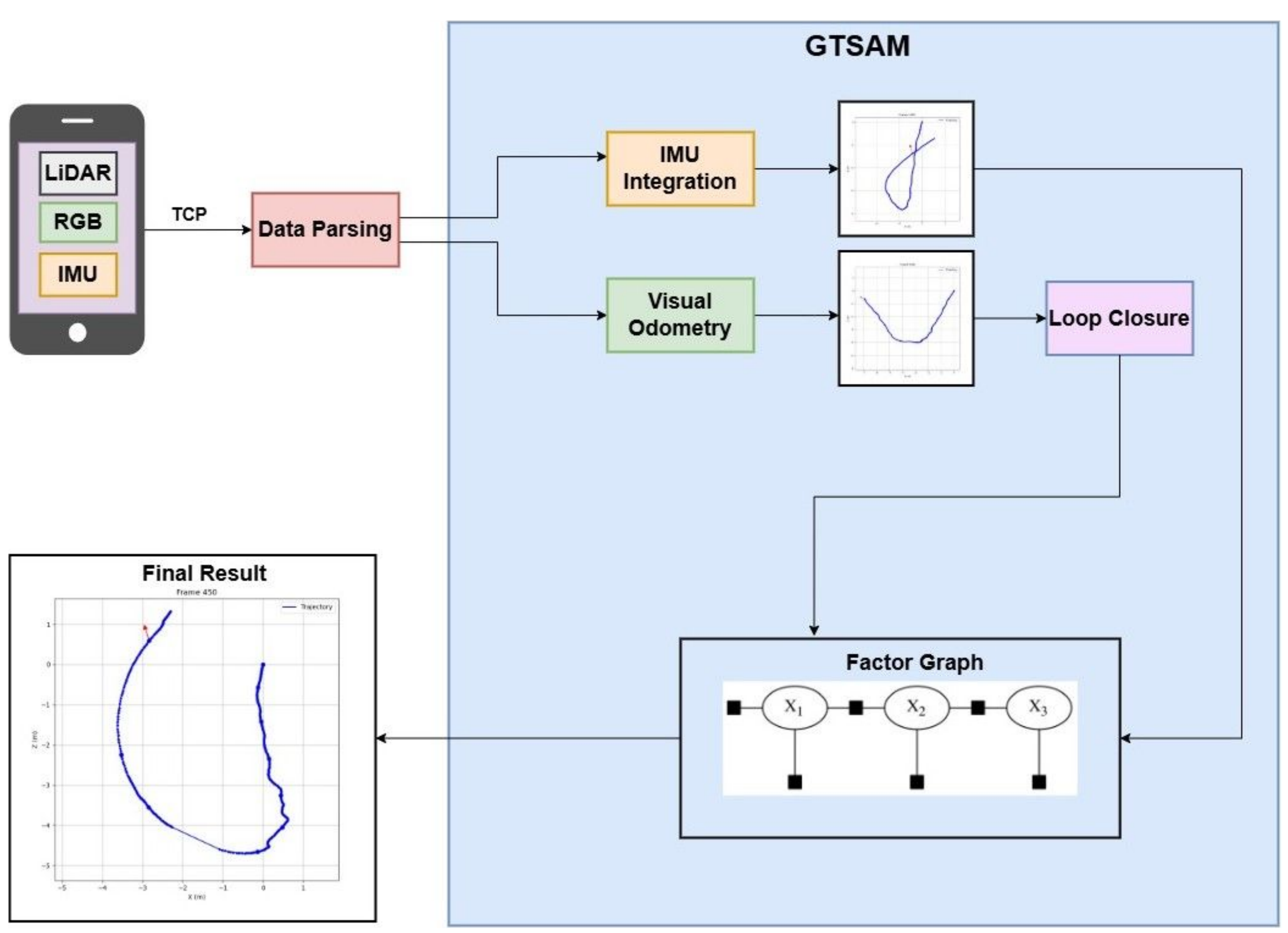
MICHIGAN ENGINEERING
UNIVERSITY OF MICHIGAN

Motivation

While modern smartphones like the iPhone offer advanced sensors—including LiDAR, RGB cameras, and IMUs—developers are often limited to high-level APIs such as ARKit, which conceal the underlying SLAM pipeline. Our motivation with iLoco is to break open this black box.

We built a SLAM system that streams raw RGB-D and IMU data from an iPhone and fuses it using GTSAM and Bag-of-Words. This allows for a transparent and customizable SLAM pipeline that runs in real time, visualizing keyframes, feature points, and trajectories live. By making the full SLAM process visible and modifiable, iLoco empowers students, educators, and researchers to understand, evaluate, and improve the core components of SLAM on everyday hardware.

System Architecture

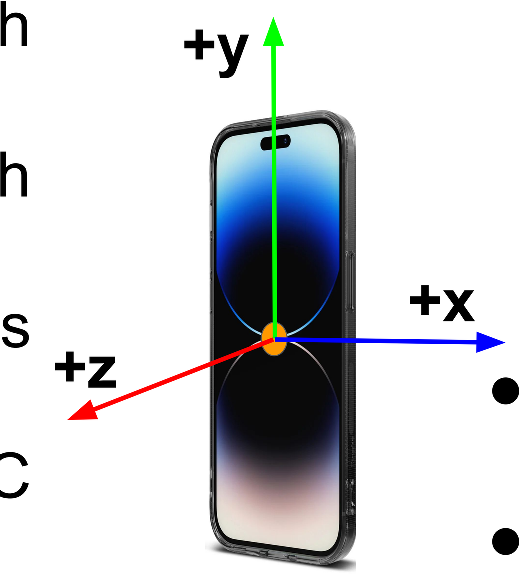


IMU Odometry

- Collected timestamped iPhone IMU data
- Built skew-symmetric matrices for rotation updates
- Used Rodrigues' formula with Taylor expansions for small-angle Gamma functions
- Propagated IMU state via discrete-time integration using zero-order hold
- Computed rotation with exponential map; updated velocity and position
- Integrated bias-corrected IMU data to estimate full trajectory

Visual Odometry

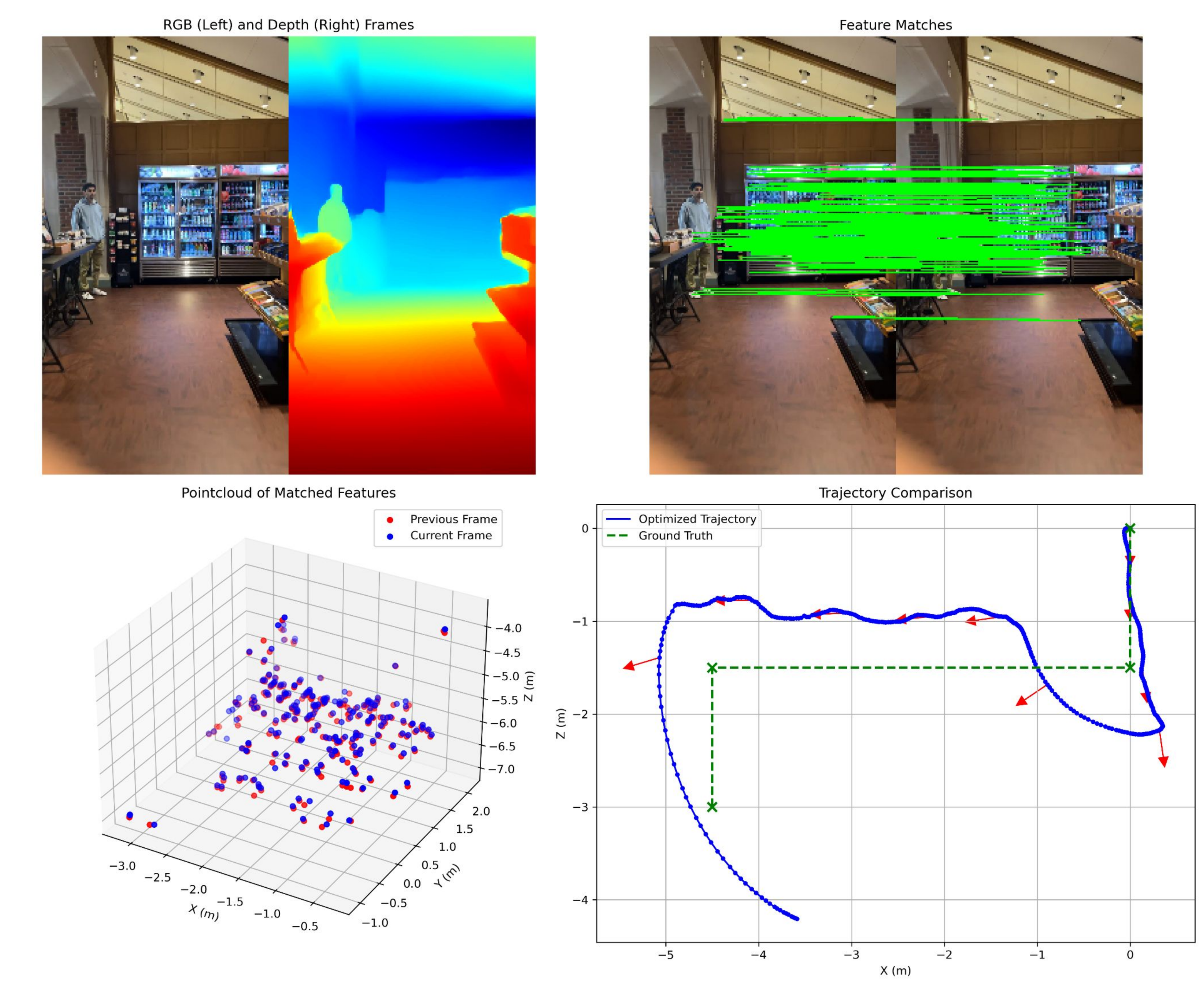
- Collected iPhone RGB and LiDAR depth data
- Matched ORB features using FLANN with Lowe's ratio test
- Converted depth maps to 3D point clouds and extracted valid matches
- Estimated 3D transform with RANSAC and SVD
- Accumulated transforms to recover absolute poses



GTSAM & BoW

- Developed iPhone localization using GTSAM with BoW-based loop closure
- Initialized factor graph with pose, velocity, and bias priors; used ISAM2 for optimization
- Trained ORB vocabulary for loop detection
- Added IMU preintegration and visual odometry constraints per frame
- Detected loop closures via BoW and added loop edges
- Re-optimized graph incrementally and extracted final trajectory

Results



- 2x2 grid showing RGB/depth frames, feature matches, 3D point cloud fitting, and GTSAM optimized trajectory
- iPhone handheld data collection within indoor environments for ~ 30 sec (i.e kitchen, snack room, etc)
- Simplified to yaw-only rotation for simpler visualization

Conclusions

Performance Analysis:

- IMU tracking more reliable for long distances (>5m)
- VO performs poorly on turns due to plain walls
- GTSAM optimization more stable with mini-batches
- iLoco outperforms individual IMU/VO but performs worse than ARKit in rotation accuracy
- Future work
  - Build map to allow for more robust loop closure
  - Integrate image and depth (LiDAR) separately

Acknowledgements: Our team would like to thank Dr. Maani Ghaffari, Dr. Minghan Zu, and the EECS 568 course staff for their guidance and making this project possible.