Visual-Inertial Tracking using Pre-integrated Factors

Tarlan Bakirli, Yihao Wang

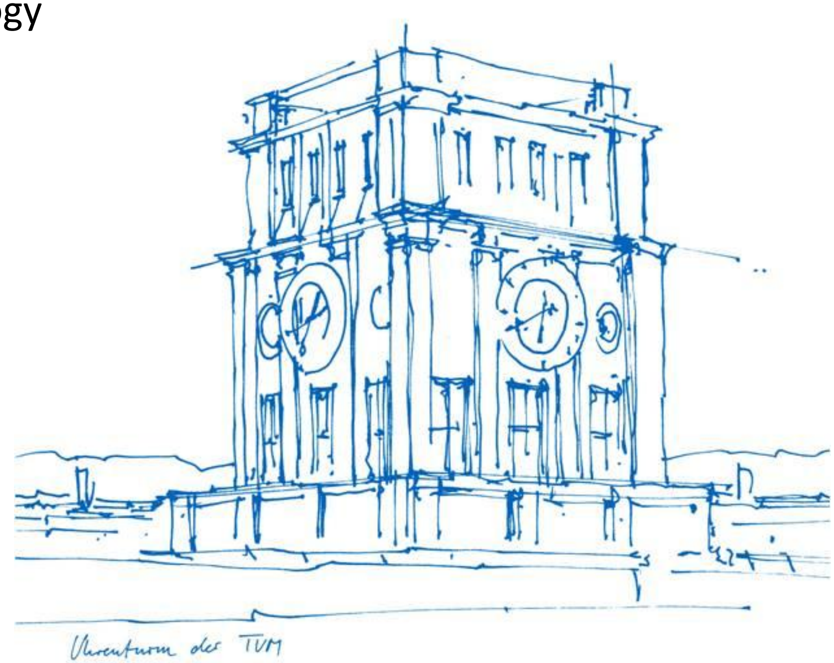
Advisor: Jason Chui

Computer Vision Group

TUM School of Computation, Information and Technology

Technical University of Munich

02.09.23 MI, 30. Januar 2023



Contents

- Background
- Pipeline
  - Data Loading
  - IMU Initialization
  - Preintegration
  - Optimization
  - Evaluation
- Experiments
- Limitations
- Future Work

Background

Visual Odometry:

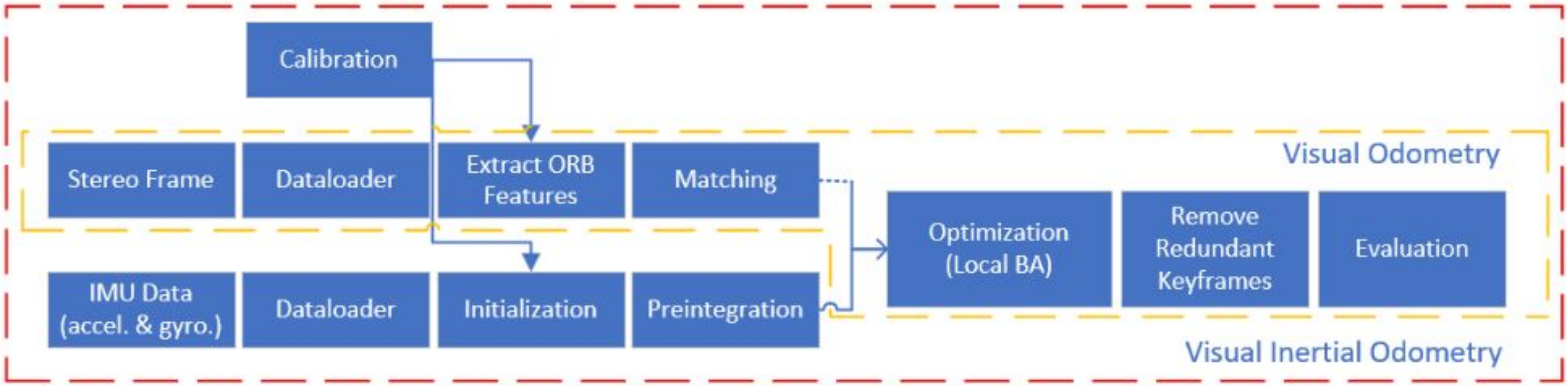
- Camera frames only
- Unknown scale for monocular setup
- Limited output rate
- Cameras precise in slow motion
- Not accurate in high speed scenarios (e.g. highways)

Visual-Inertial Odometry:

- Camera frames + IMU measurements (Complement each other)
- Robust motion estimates
- High output rate
- IMU large uncertainty in slow motion
- Precise in high speed scenarios

Pipeline

- ➔ Dataloading
- ➔ Initialization
- ➔ Preintegration
- ➔ Optimization
- ➔ Evaluation



Data Loading

- Implementation of IMU data loading
- Implementation of ground truth data loading

➢ Calibration:
  - Double Sphere camera model
  - basalt_calibrate & basalt_calibrate_imu
  - generated values adapted into existing calibration file

Basalt Calibration: https://gitlab.com/VladyslavUsenko/basalt/-/blob/master/doc/Calibration.md

IMU Initialization

- Fixed Bias
  - For computation, we set bias_gyro and bias_accel as {0, 0, 0}
- Zero Scale
  - Computed by stereo frames. No need for initialization.
- Fixed Gravity
  - Assign the gravity direction and world frame: $\hat{\mathbf{g}}_{\mathrm{I}} = \{0, 0, -1\}$
  - Set g = {0, 0, -9.81}
- Pose estimation
  - Initial rotation from accel&unit vec.
  - Zero initial translation
  - Estimate T_w_i
- Velocity Estimation
  - The initial frames are still.
  - Set vel_w_i_init = {0, 0, 0}
  - Estimate vel_w_i

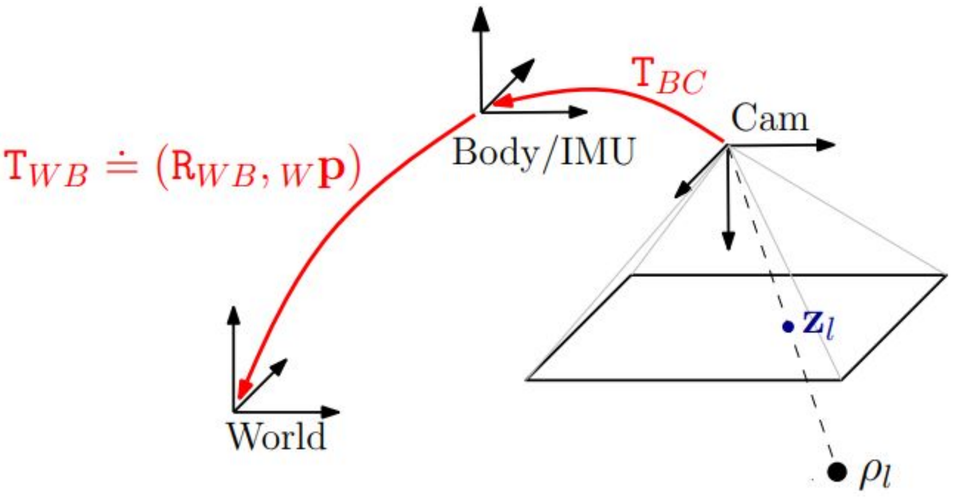
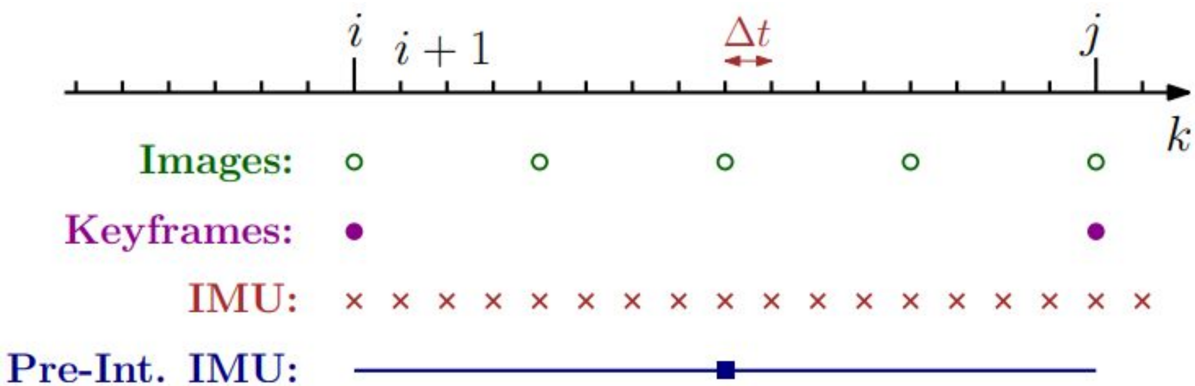


Image: Christian Forster et al. (2016)

Preintegration

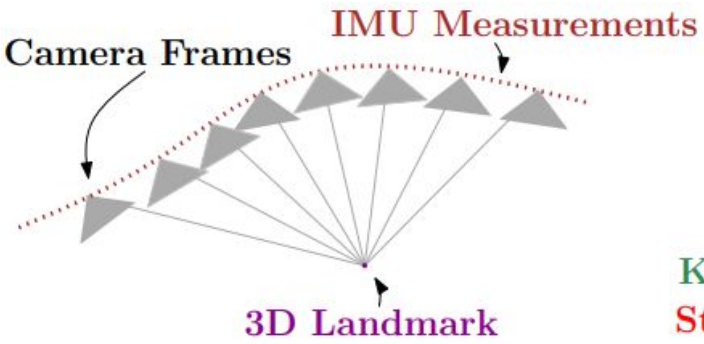
$$
\begin{aligned}
\mathbf{R}_{\mathtt{WB}}^{i+1} &= \mathbf{R}_{\mathtt{WB}}^{i} \Delta\mathbf{R}_{i,i+1} \operatorname{Exp}\left(\left(\mathbf{J}_{\Delta R}^{g} \mathbf{b}_{g}^{i}\right)\right) \\
{}_{\mathtt{W}}\mathbf{v}_{\mathtt{B}}^{i+1} &= {}_{\mathtt{W}}\mathbf{v}_{\mathtt{B}}^{i} + \mathbf{g}_{\mathtt{W}} \Delta t_{i,i+1} \\
&\quad + \mathbf{R}_{\mathtt{WB}}^{i}\left(\Delta\mathbf{v}_{i,i+1} + \mathbf{J}_{\Delta v}^{g} \mathbf{b}_{g}^{i} + \mathbf{J}_{\Delta v}^{a} \mathbf{b}_{a}^{i}\right) \\
{}_{\mathtt{W}}\mathbf{p}_{\mathtt{B}}^{i+1} &= {}_{\mathtt{W}}\mathbf{p}_{\mathtt{B}}^{i} + {}_{\mathtt{W}}\mathbf{v}_{\mathtt{B}}^{i} \Delta t_{i,i+1} + \frac{1}{2} \mathbf{g}_{\mathtt{W}} \Delta t_{i,i+1}^{2} \\
&\quad + \mathbf{R}_{\mathtt{WB}}^{i}\left(\Delta\mathbf{p}_{i,i+1} + \mathbf{J}_{\Delta p}^{g} \mathbf{b}_{g}^{i} + \mathbf{J}_{\Delta p}^{a} \mathbf{b}_{a}^{i}\right)
\end{aligned}
$$

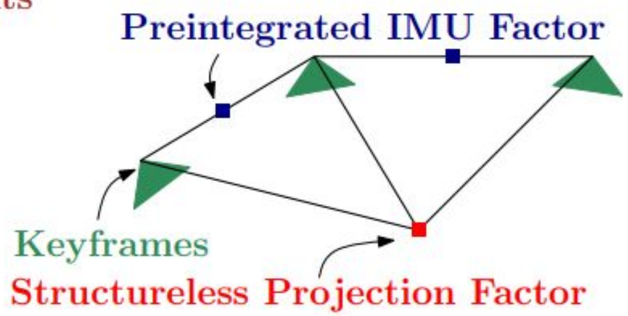


Images: Christian Forster et al. (2016)

Preintegration

- Integrating IMU measurements per frame
- Synchronized with camera, measurements at discrete times
- 10 imu measurements per frame
  - Frame rate: 2x20 Hz (Stereo Setup)
  - IMU rate: 200 Hz
- Preintegration of Basalt library
  - propagates state at each imu timestamp
  - predicts next state based on propagated delta state





Images: Christian Forster et al. (2016)
Basalt: https://gitlab.com/VladyslavUsenko/basalt/-/tree/master

Optimization

- **Bundle Adjustment**: joint optimization of feature reprojection error and the IMU error

$$\theta = \left\{ \mathbf{R}^j_{\mathtt{WB}}, {}_{\mathtt{W}}\mathbf{p}^j_{\mathtt{B}}, {}_{\mathtt{W}}\mathbf{v}^j_{\mathtt{B}}, \mathbf{b}^j_g, \mathbf{b}^j_a \right\}$$

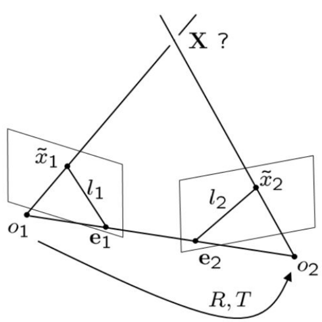
$$\theta^* = \underset{\theta}{\operatorname{argmin}} \left(\sum_k \mathbf{E}_{\text{proj}}(k, j) + \mathbf{E}_{\text{IMU}}(i, j) \right)$$

Feature reprojection error for each keyframe:

ρ: the Huber robust cost function

$$\mathbf{E}_{\text{proj}}(k, j) = \rho \left(\left(\mathbf{x}^k - \pi(\mathbf{X}^k_{\mathtt{C}}) \right)^T \mathbf{\Sigma}_k \left(\mathbf{x}^k - \pi(\mathbf{X}^k_{\mathtt{C}}) \right) \right)$$

$$\mathbf{X}^k_{\mathtt{C}} = \mathbf{R}_{\mathtt{CB}} \mathbf{R}^j_{\mathtt{BW}} \left(\mathbf{X}^k_{\mathtt{W}} - {}_{\mathtt{W}}\mathbf{p}^j_{\mathtt{B}} \right) + {}_{\mathtt{C}}\mathbf{p}_{\mathtt{B}}$$



Transform cameras to IMU (body) frame

IMU error between two frames i and j:

$$\mathbf{E}_{\text{IMU}}(i, j) = \rho \left(\left[\mathbf{e}^T_R \, \mathbf{e}^T_v \, \mathbf{e}^T_p \right] \mathbf{\Sigma}_I \left[\mathbf{e}^T_R \, \mathbf{e}^T_v \, \mathbf{e}^T_p \right]^T \right) + \rho \left(\mathbf{e}^T_b \mathbf{\Sigma}_R \mathbf{e}_b \right)$$

$$\mathbf{e}_R = \operatorname{Log} \left(\left(\Delta \mathbf{R}_{ij} \operatorname{Exp} \left(\mathbf{J}^g_{\Delta R} \mathbf{b}^j_g \right) \right)^T \mathbf{R}^i_{\mathtt{BW}} \mathbf{R}^j_{\mathtt{WB}} \right)$$

$$\mathbf{e}_v = \mathbf{R}^i_{\mathtt{BW}} \left({}_{\mathtt{W}}\mathbf{v}^j_{\mathtt{B}} - {}_{\mathtt{W}}\mathbf{v}^i_{\mathtt{B}} - \mathbf{g}_{\mathtt{W}} \Delta t_{ij} \right) - \left(\Delta \mathbf{v}_{ij} + \mathbf{J}^g_{\Delta v} \mathbf{b}^j_g + \mathbf{J}^a_{\Delta v} \mathbf{b}^j_a \right)$$

$$\mathbf{e}_p = \mathbf{R}^i_{\mathtt{BW}} \left({}_{\mathtt{W}}\mathbf{p}^j_{\mathtt{B}} - {}_{\mathtt{W}}\mathbf{p}^i_{\mathtt{B}} - {}_{\mathtt{W}}\mathbf{v}^i_{\mathtt{B}} \Delta t_{ij} - \frac{1}{2} \mathbf{g}_{\mathtt{W}} \Delta t^2_{ij} \right) - \left(\Delta \mathbf{p}_{ij} + \mathbf{J}^g_{\Delta p} \mathbf{b}^j_g + \mathbf{J}^a_{\Delta p} \mathbf{b}^j_a \right)$$

$$\mathbf{e}_b = \mathbf{b}^j - \mathbf{b}^i$$

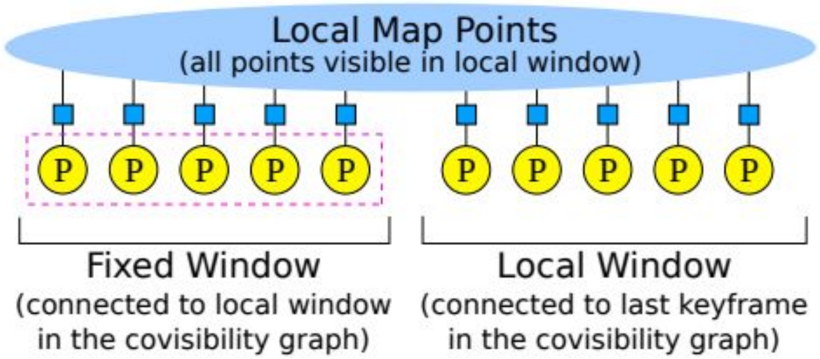
We don’t optimize bias.

Mur-Artal, Raúl, and Juan D. Tardós. "Visual-inertial monocular SLAM with map reuse." *IEEE Robotics and Automation Letters* 2.2 (2017): 796-803

Optimization

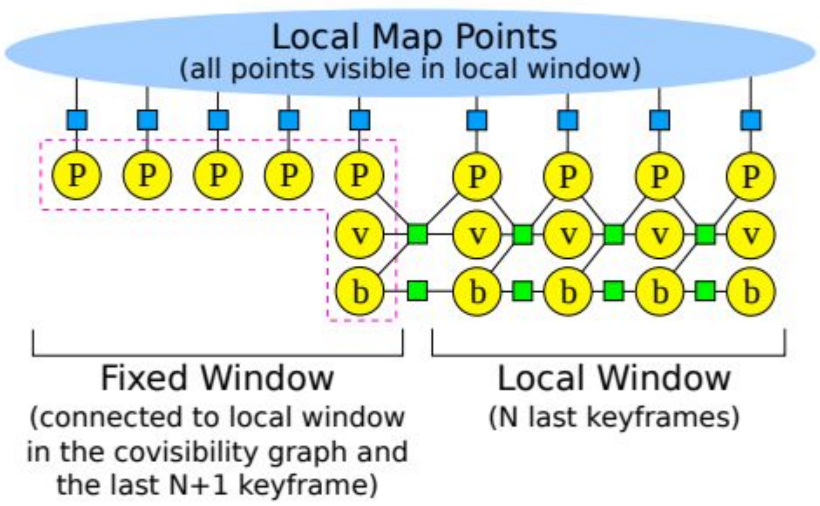
- The increasing data amount for optimization
- The old frames provide weaker information

=> **Sliding windows algorithm**



Visual Odometry’s Local BA

Only consider current states, optimize the frames inside the local window and discard redundant keyframes



Visual Inertial Odometry’s Local BA

For local BA, we firstly choose

- 3 latest frames for IMU error and
- 7 latest keyframes for reprojection error.

Mur-Artal, Raúl, and Juan D. Tardós. "Visual-inertial monocular SLAM with map reuse." *IEEE Robotics and Automation Letters* 2.2 (2017): 796-803

Evaluation

- **Absolute Trajectory Error** (ATE): measures the **difference** between points of the true and the estimated trajectory. - performance of SLAM
  - **Timestamps Association**: associate the estimated poses with GT
  - **Trajectory Alignment**: Align the GT and the estimated trajectory using SVD
  - **Error Calculation**: compute the difference between each pair of poses and GT
  - RMSE (Root Mean Square Error):

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{n}(Y_i - f(x_i))^2}$$

- **Relative Pose Error** (RPE): measure the **drift** of a **visual odometry** system, for example the drift per second - performance of VO
  - Compute the error of relative motion between timestamp pairs
  - Consider **rotation and translation** (pose)

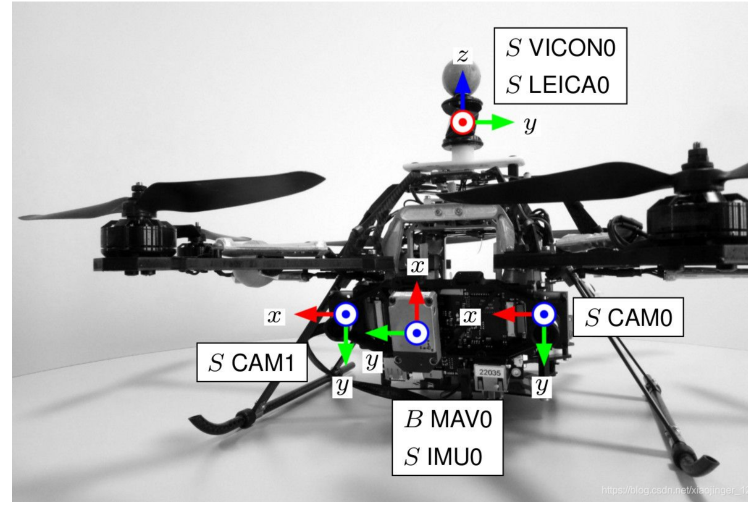
Experiments

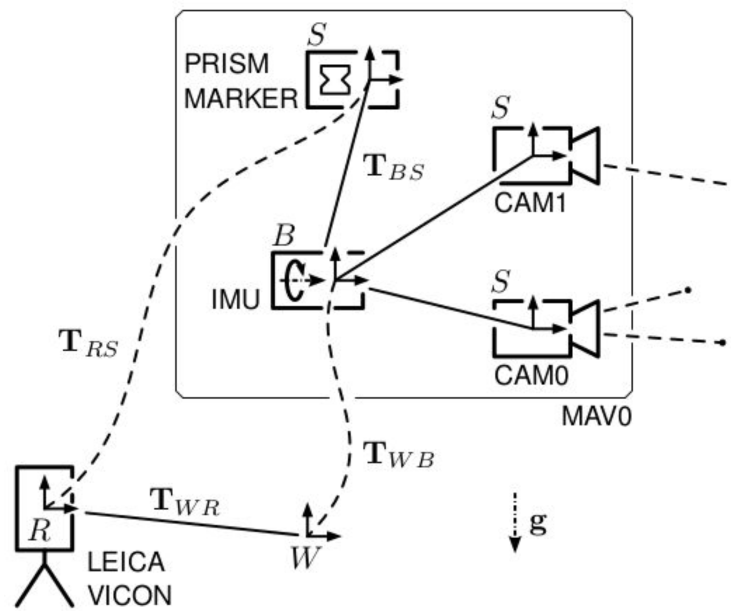
Datasets: EuRoC

- 11 sequences recorded from a micro aerial vehicle
- Classified to different level according to illumination, velocity and so on
- V1_01 - V1_02: Vicon Room
- MH_01 - MH_03: Machine Hall

The data contain:

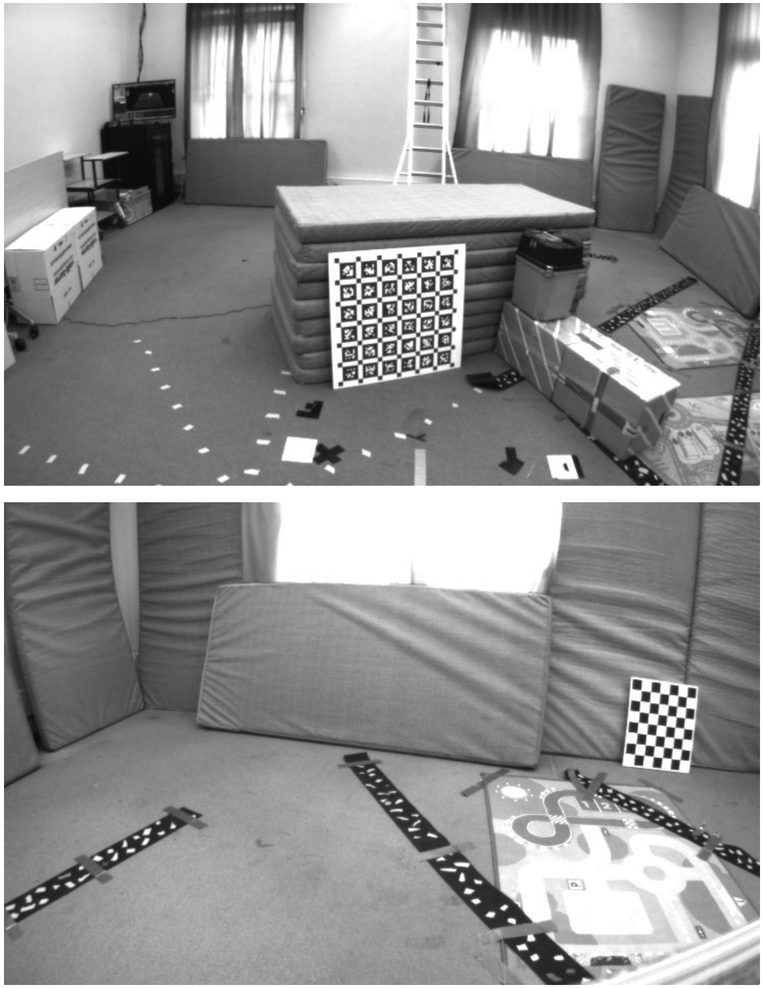
- cam0
- cam1
- imu0
  - gyroscope data
  - acceleration
- leica0
- ground_truth
  - translation
  - rotation matrix
  - velocity
  - bias of gyroscope
  - bias of accelerator



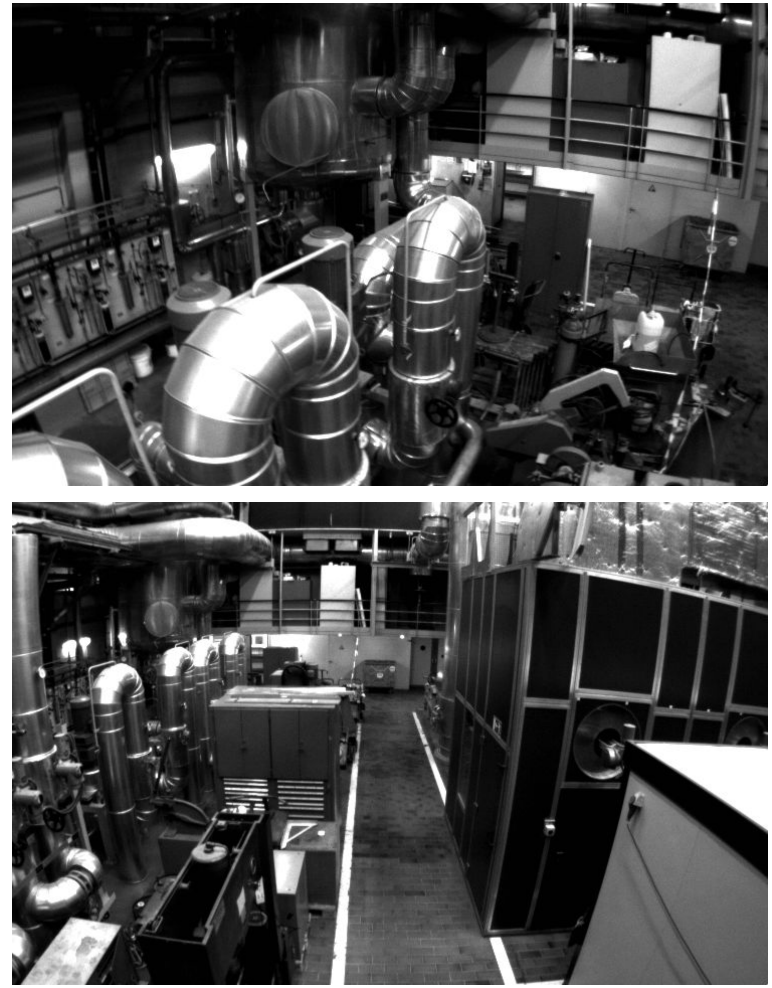


EuRoC dataset: https://projects.asl.ethz.ch/datasets/doku.php?id=kmavvisualinertialdatasets

Experiments



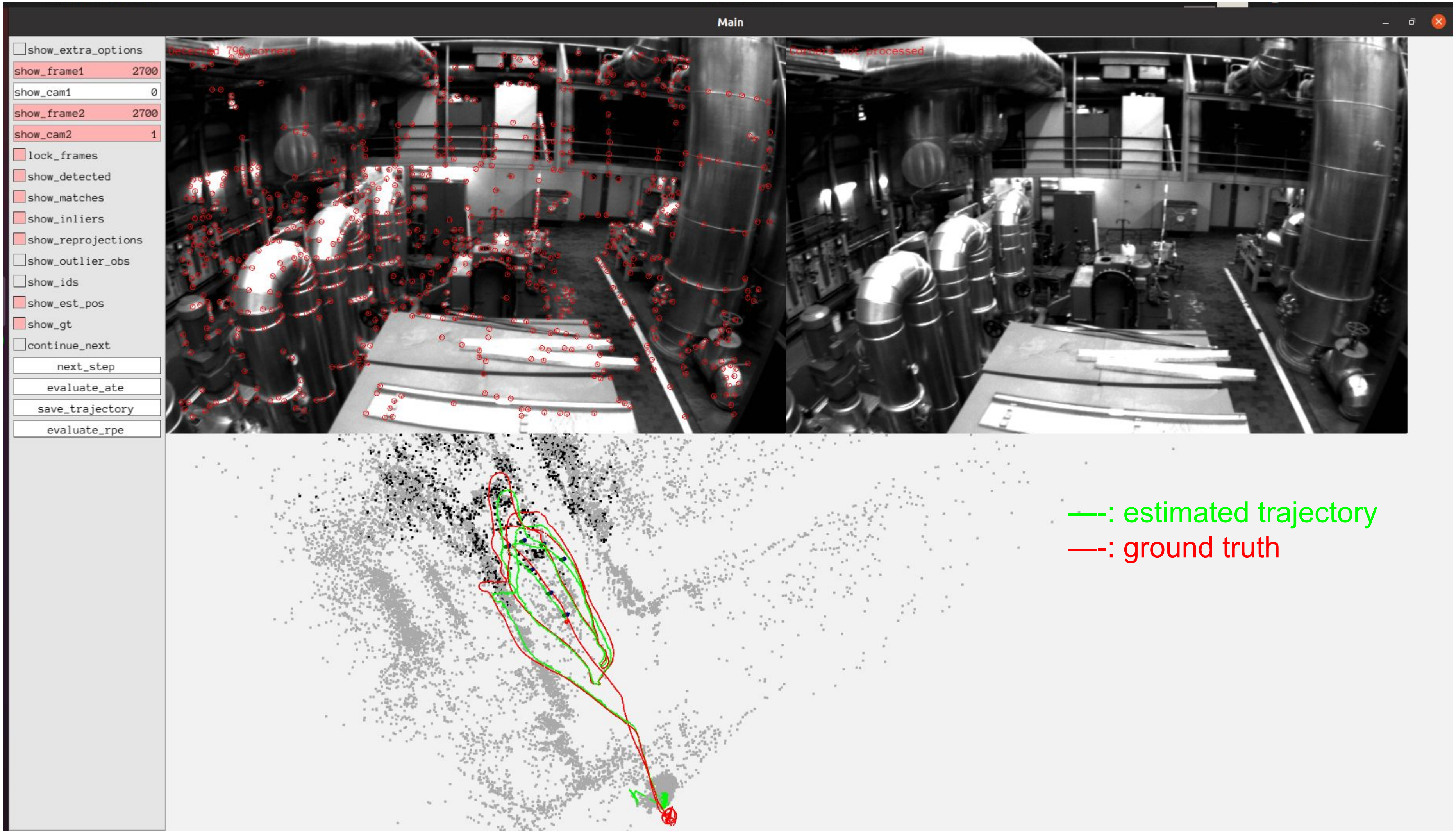
Vicon Room 1



Machine Hall

EuRoC dataset: https://projects.asl.ethz.ch/datasets/doku.php?id=kmavvisualinertialdatasets

Experiments

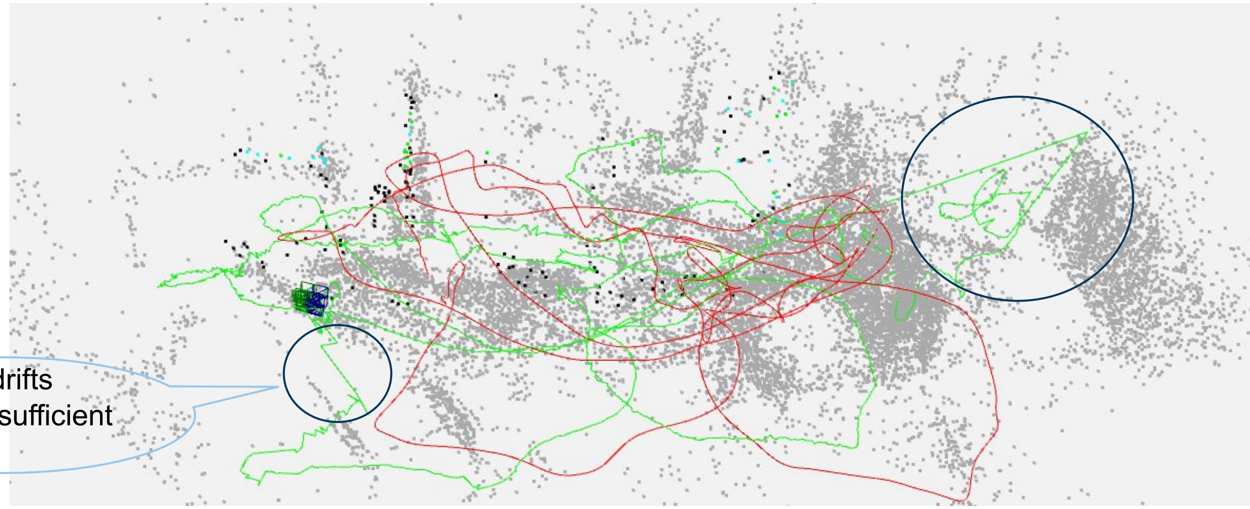


Experiments

V1_01_easy

Without IMU:

ATE: 0.715

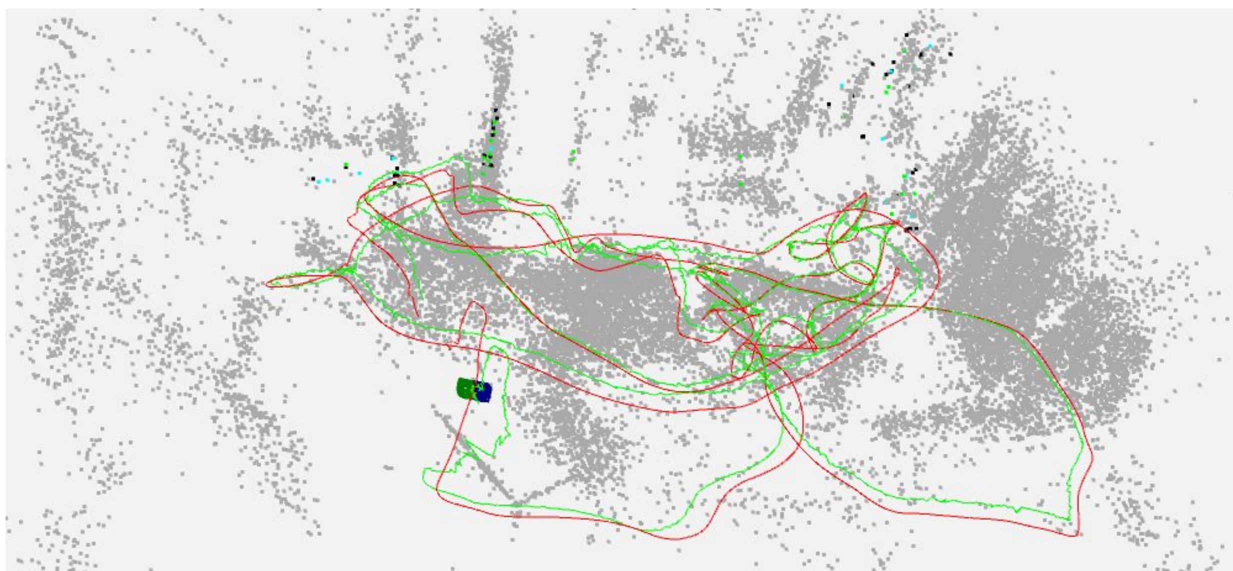


With IMU:

ATE: 0.186

IMU data constrain the camera motion

stable and robust!

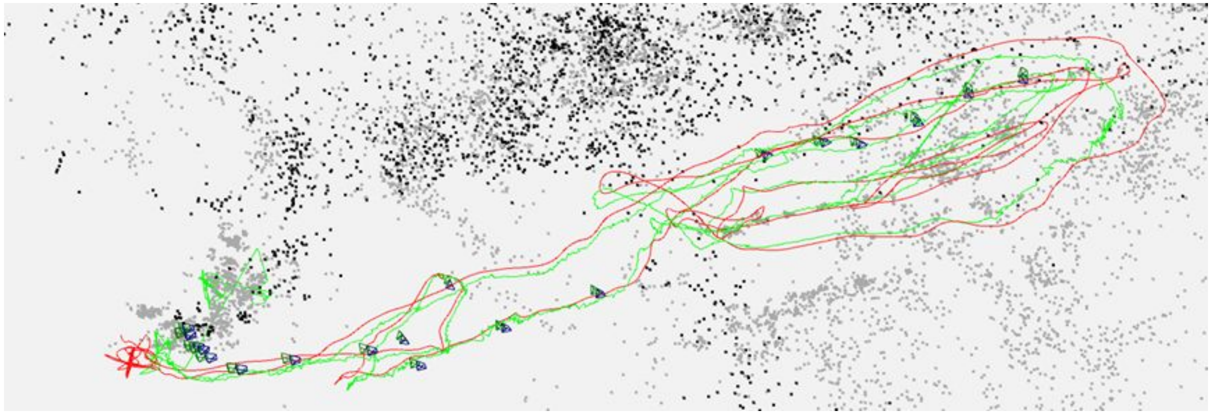


Experiments

MH_01_Easy (with IMU):

RMSE ATE: 0.35

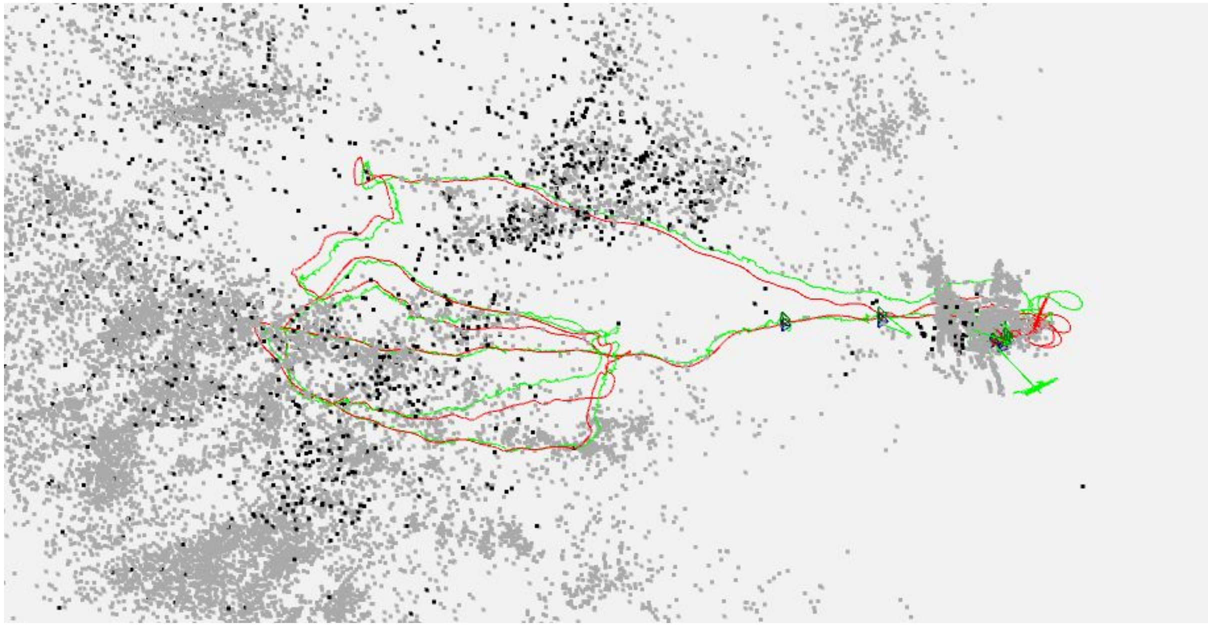
Mean translation RPE: 4.35



MH_02_Easy (with IMU):

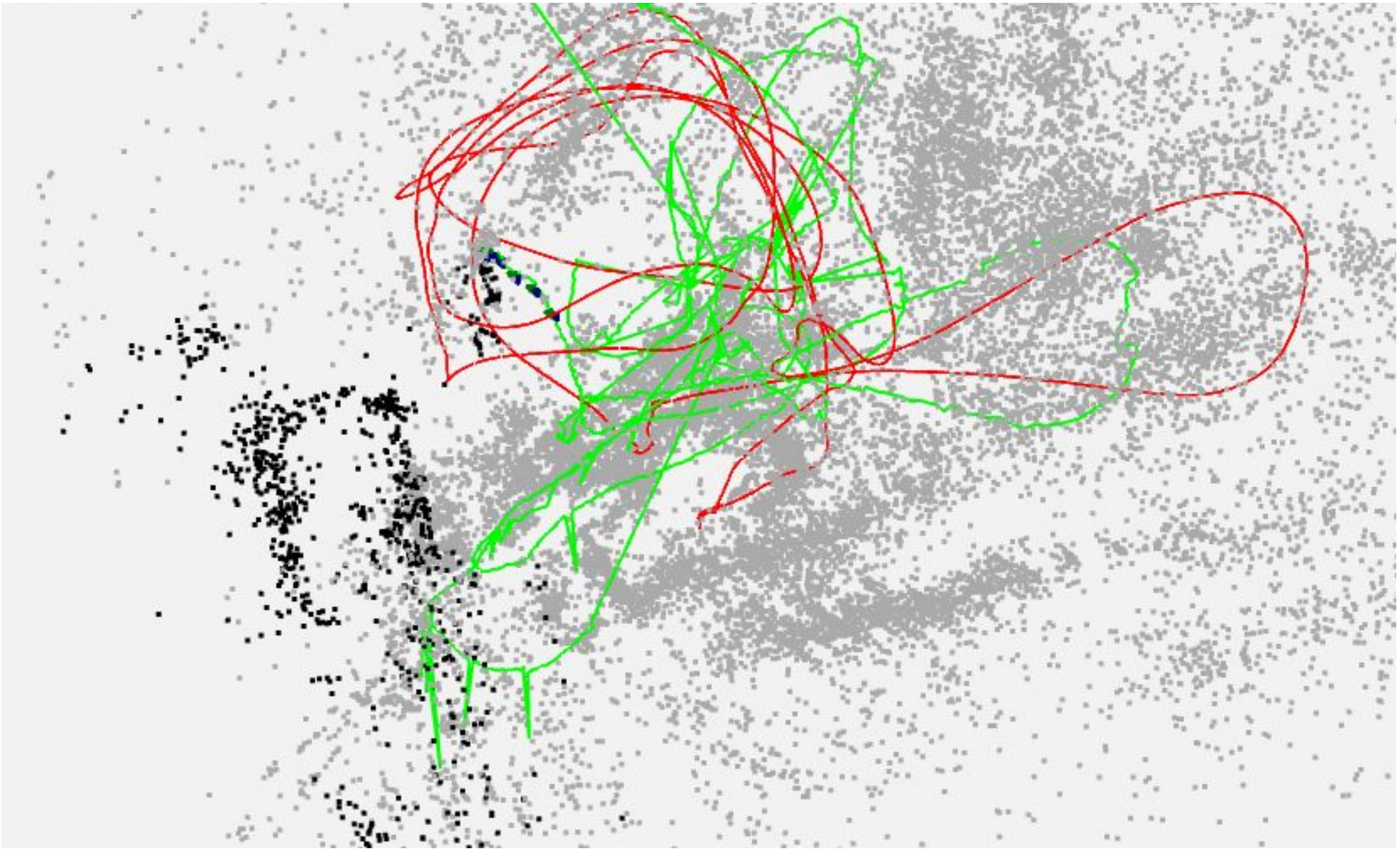
RMSE ATE: 0.33

Mean translation RPE: 6.06



Experiments

Failed case for MH_03_medium:



Experiments

Values of ATE and RPE of the estimated trajectory in meters:

	VO		VIO	
	ATE (RMSE)	RPE (Translation Mean)	ATE (RMSE)	RPE (Translation Mean)
V1_01_easy	0.715	3.770	**0.186**	**3.301**
V1_02_medium	2.037	**3.356**	**1.943**	3.358
MH_01_easy	0.421	6.088	**0.352**	**4.351**
MH_02_easy	**0.254**	**4.064**	0.326	6.055
MH_03_medium	3.251	**5.350**	**3.240**	5.875

Limitations

Compared with visual-only system, our VIO …

- is more computationally-costly, so the window of the BA should be smaller than VO

Besides, for our model itself:

- Initialization is simplified, which is not accurate enough.
  - The gravity, velocity and bias are restricted.

Future Work

- Initialization: consider more complex situation
  - Contains bias of accelerator and gyroscope
  - Contains initial velocity: MAV is not still at the beginning

- Run on loosely-coupled datasets
  - Find the nearest timestamps and do spline interpolation

- Expand to different datasets
  - monocular: consider the initialization of scale
  - RGB-D: consider the fusion of different data

- Expand to different situations
  - Loop Closure

References

1. Basalt: https://gitlab.com/VladyslavUsenko/basalt/-/tree/master
2. Useful tools for the RGB-D benchmark: Computer Vision Group - Useful tools for the RGB-D benchmark (tum.de)
3. Christian Forster et al., On-Manifold Preintegration for Real-Time Visual-Inertial Odometry, 2016
4. EuRoC: https://projects.asl.ethz.ch/datasets/doku.php?id=kmavvisualinertialdatasets
5. Mur-Artal, Raúl, and Juan D. Tardós. "Visual-inertial monocular SLAM with map reuse." *IEEE Robotics and Automation Letters* 2.2 (2017): 796-803

Thanks!
Q&A