Visual Inertial Odometry using Focal Plane Binary Features (BIT-VIO)

Matthew Lisondra [1,*], Junseo Kim [1,*], Riku Murai [2], Kourosh Zareinia [1] and Sajad Saeedi [1]

[1] Toronto Metropolitan University (TMU).

[2] Imperial College London, Department of Computing.

[*] Both authors contributed equally to this research at TMU.

Table of Contents

I. Background

[1] Rebecq, H., Horstschaefer, T., & Scaramuzza, D. (2017). Real-time visual-inertial odometry for event cameras using keyframe-based nonlinear optimization. Procedings of the British Machine Vision Conference 2017. https://doi.org/10.5244/c.31.16

I. Background – Introduction

Camera technology has advanced and developed since its inception since the 1800s to 1900s and is still continuing to do so today. Camera technology, as we know it, is the basis of so many different fields, responsible for so many different applications.

[1]

There are exciting types that are still being properly and fully developed, one of which is dealing with the event camera system.

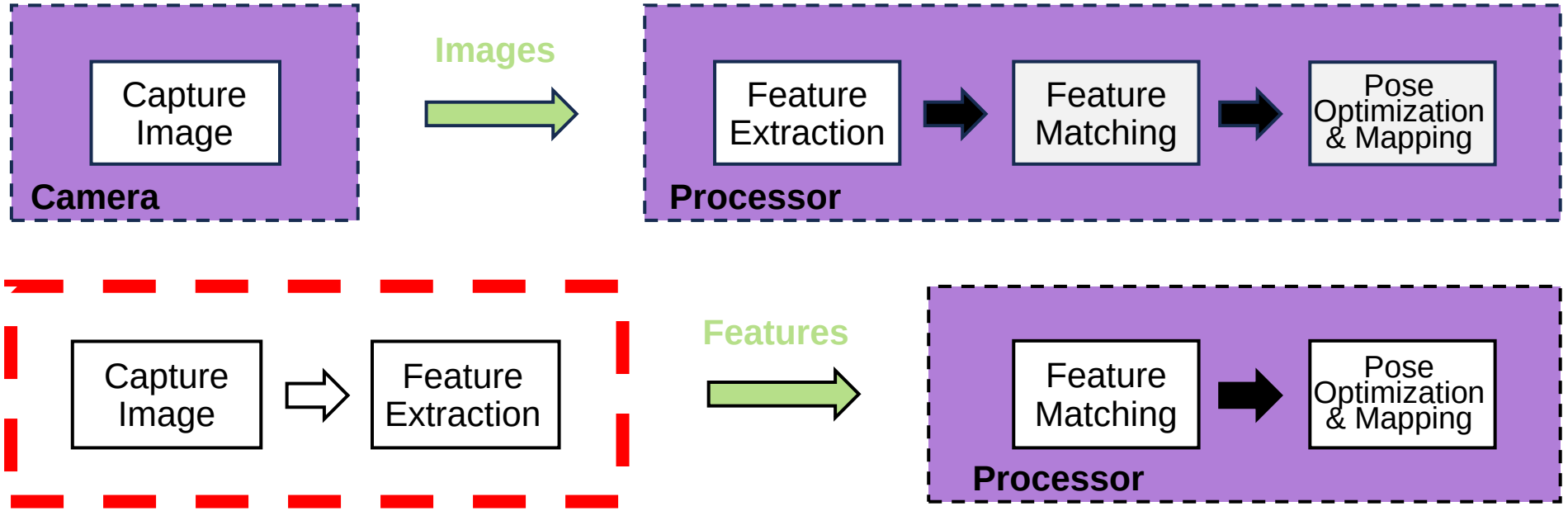
Of particular interest to us today is not the event camera, but the FPSP camera technology, its advantages, and its potential use in robotic localization and mapping.

[1] Murai, R., Saeedi, S. amp; Kelly, P. H. Bit-VO: Visual odometry at 300 fps using binary features from the focal plane. 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (2020).

FPSP vs. Conventional Camera

Moving the Computation Towards the Focal-Plane

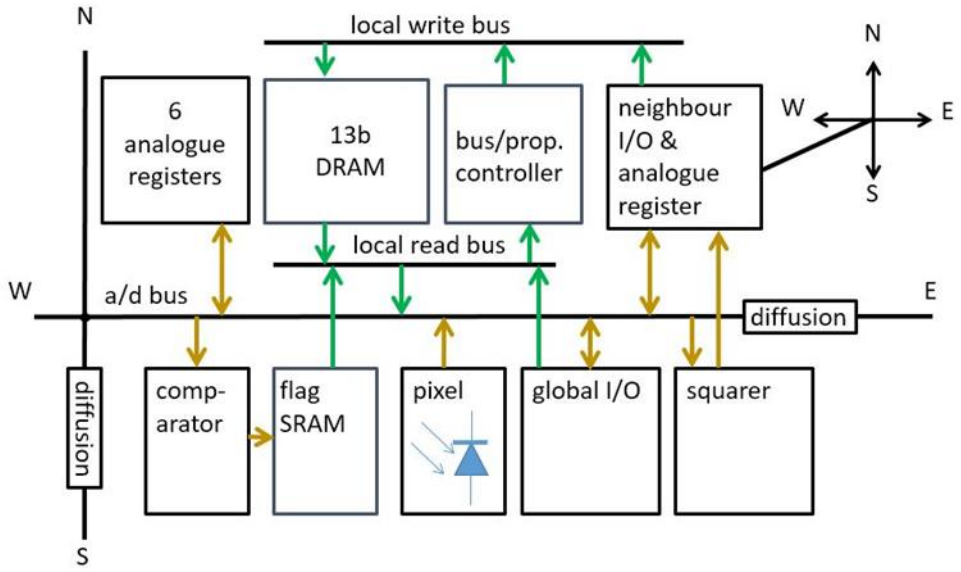
[1]



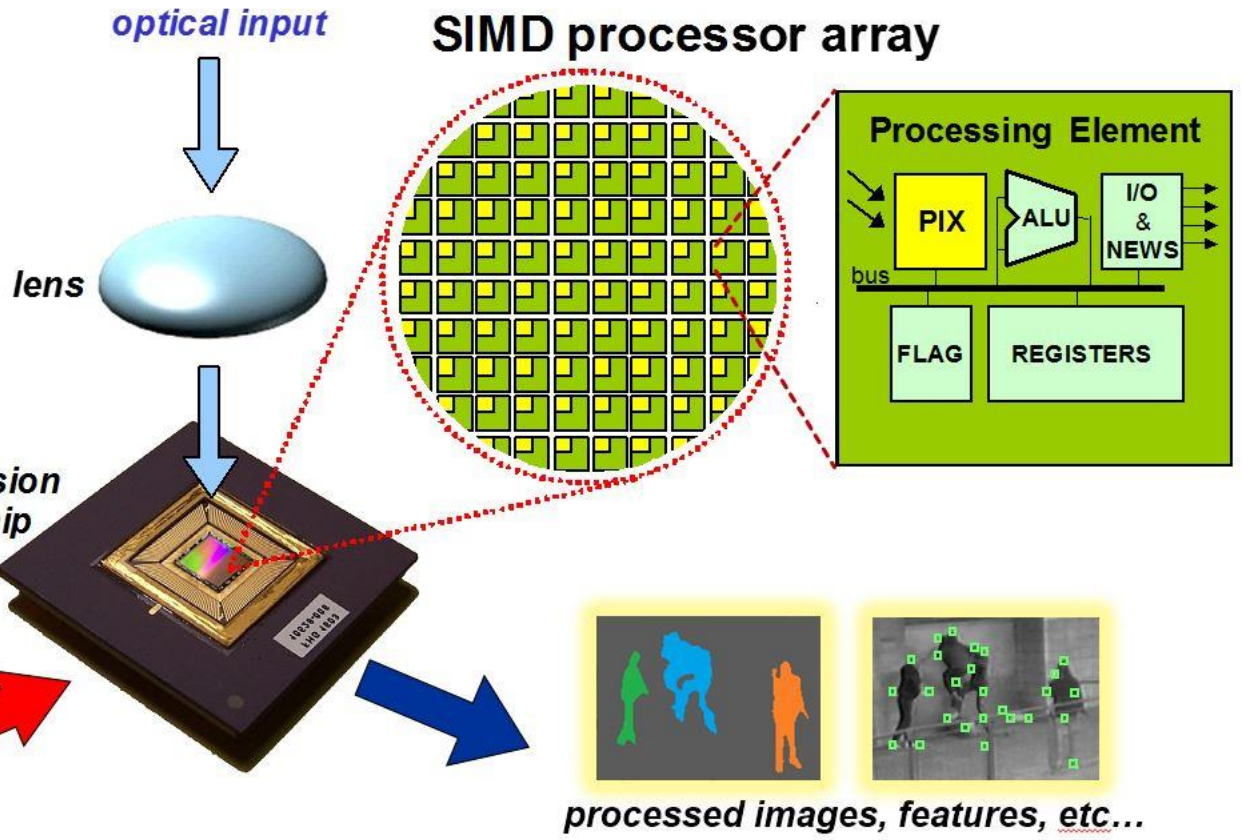
- Data transfer bandwidth is reduced – No intensity information is transferred
- Reduces computation costs on the processor side
- Only meaningful data is transmitted

[3] S.J.Carey, A.Lopich, D.R.W.Barr, B.Wang and P.Dudek, "A 100,000 fps Vision Sensor with Embedded 535 GOPS/W 256x256 SIMD Processor Array", VLSI Circuits Symposium 2013, pp.C182-C183, June 2013

I. Background – FPSP SCAMP-5 Architecture



[3]



The FPSP SCAMP-5 camera technology is designed to provide lower latency with lower power consumption.

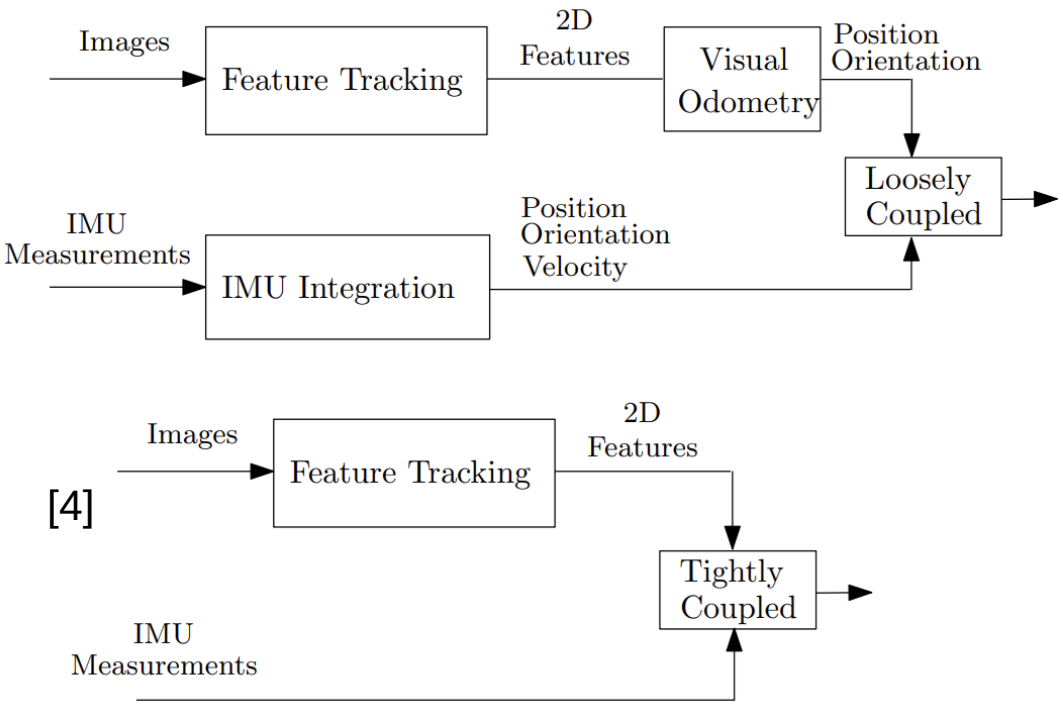
[4] Scaramuzza, D.; Zhang, Z. Aerial robots, visual-inertial odometry of. Encyclopedia of Robotics 1–9 (2020).

I. Background – Visual Inertial Odometry

"Visual-Inertial odometry (VIO) is the process of estimating the state (pose and velocity) of an agent (e.g., an aerial robot) by using only the input of one or more cameras plus one or more Inertial Measurement Units (IMUs) attached to it."

VIO is the only viable alternative to GPS and lidar-based odometry to achieve accurate state estimation.

[4]



[4]

Filtering	Fixed-lag Smoothing	Full smoothing
Only updates the most recent states • (e.g., extended Kalman filter)	Optimizes window of states • Marginalization • Nonlinear least squares optimization	Optimize all states • Nonlinear Least squares optimization
×1 Linearization ×Accumulation of linearization errors ×Gaussian approximation of marginalized states ✓Fastest	✓Re-Linearize ×Accumulation of linearization errors ×Gaussian approximation of marginalized states ✓Fast	✓Re-Linearize ✓Sparse Matrices ✓Highest Accuracy ×Slow (but fast with GTSAM)

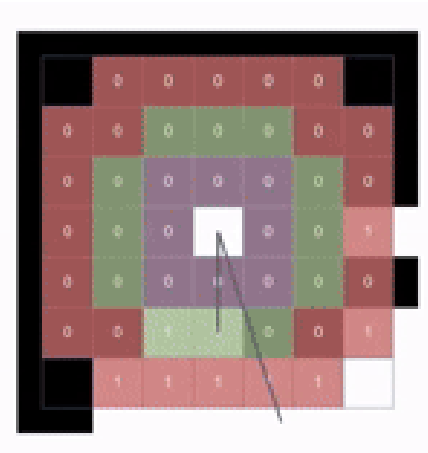
Both VO and VIO would benefit greatly from higher framerates and lower power consumption vision systems. We need FPSPs!

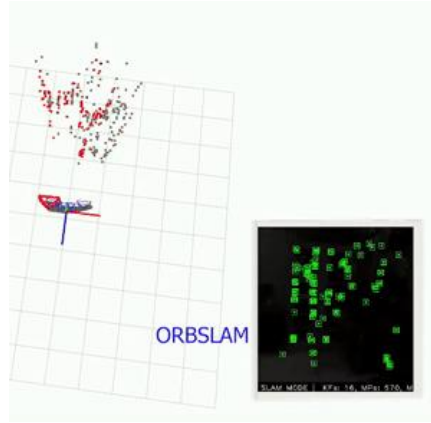
[2] Murai, R., Saeedi, S. amp; Kelly, P. H. Bit-VO: Visual odometry at 300 fps using binary features from the focal plane. 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (2020).

I. Background – Prior Work 6-DOF VO (BIT-VO) on an FPSP

A prior work in 2020 titled BIT-VO: Visual Odometry at 300 FPS by Riku Murai, Sajad Saeedi and Paul H. J. Kelly makes full use of the Focal-plane Sensor-processor camera technology.

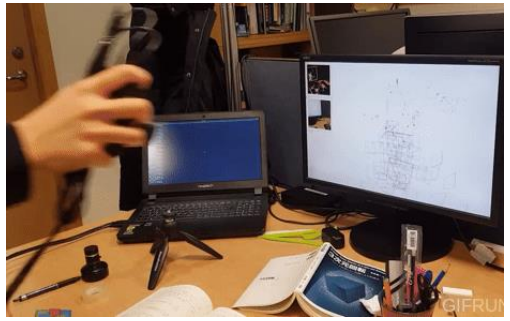
[2]





The contributions of this previous work consisted of:

1. An efficient Visual Odometry which operates at over 300 FPS. Using no intensity information, the proposed method can accurately track the pose, even under difficult situations where the state-of-the-art monocular SLAM fails.
2. A robust feature matching scheme, which uses the novel binary-edge based descriptor. Using a small, 44-bit descriptor, the system can track the noisy feature data computed on the focal plane in the SCAMP-5 FPSP image sensor itself.
3. Extensive evaluation of the system against measurements from a motion capture system, including difficult scenarios such as violently shaking the device 4-5 times in a second.



So, work has been done on an all 6-DOF algorithm with the FPSP camera technology,
but what about more complex estimation, those of which are the more accurate 6-DOF algorithms SLAM, or VIO?

II. Problem Statement

Does there exist a 6-DOF SLAM or VIO algorithm using the advantages of the FPSP camera technology?

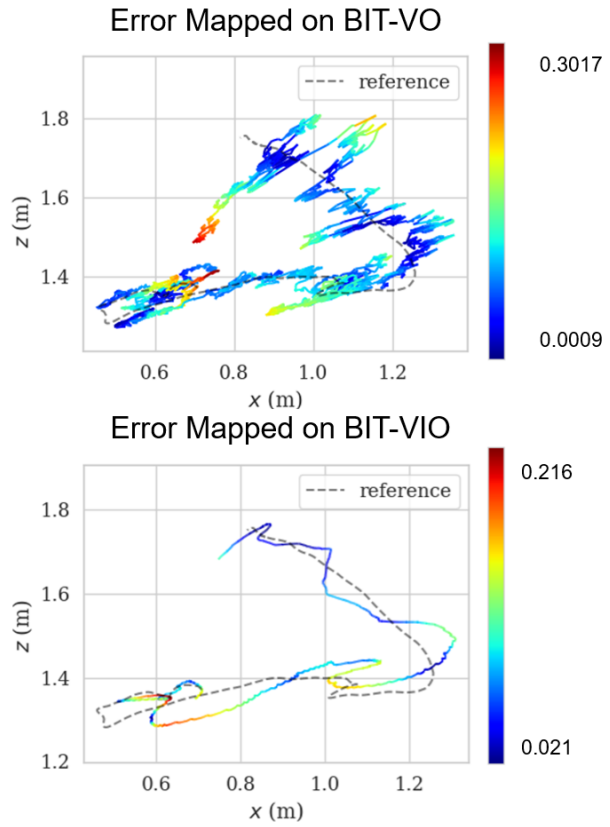
Extending on the previous work BIT-VO, we present BIT-VIO,
the first 6-DOF VIO algorithm to utilize the advantages of the FPSP for vision-IMU-fused state estimation.

The contributions of this work are:

(I) Efficient Visual Inertial Odometry operating and correcting by loosely-coupled sensor-fusion iterated Extended Kalman Filter (iEKF) at 300 FPS with an IMU at 400 Hz.

(II) Uncertainty propagation for BIT-VO's pose as it is based on binary-edge-based descriptor extraction, 2D to 3D re-projection, and transform from pose to pose.

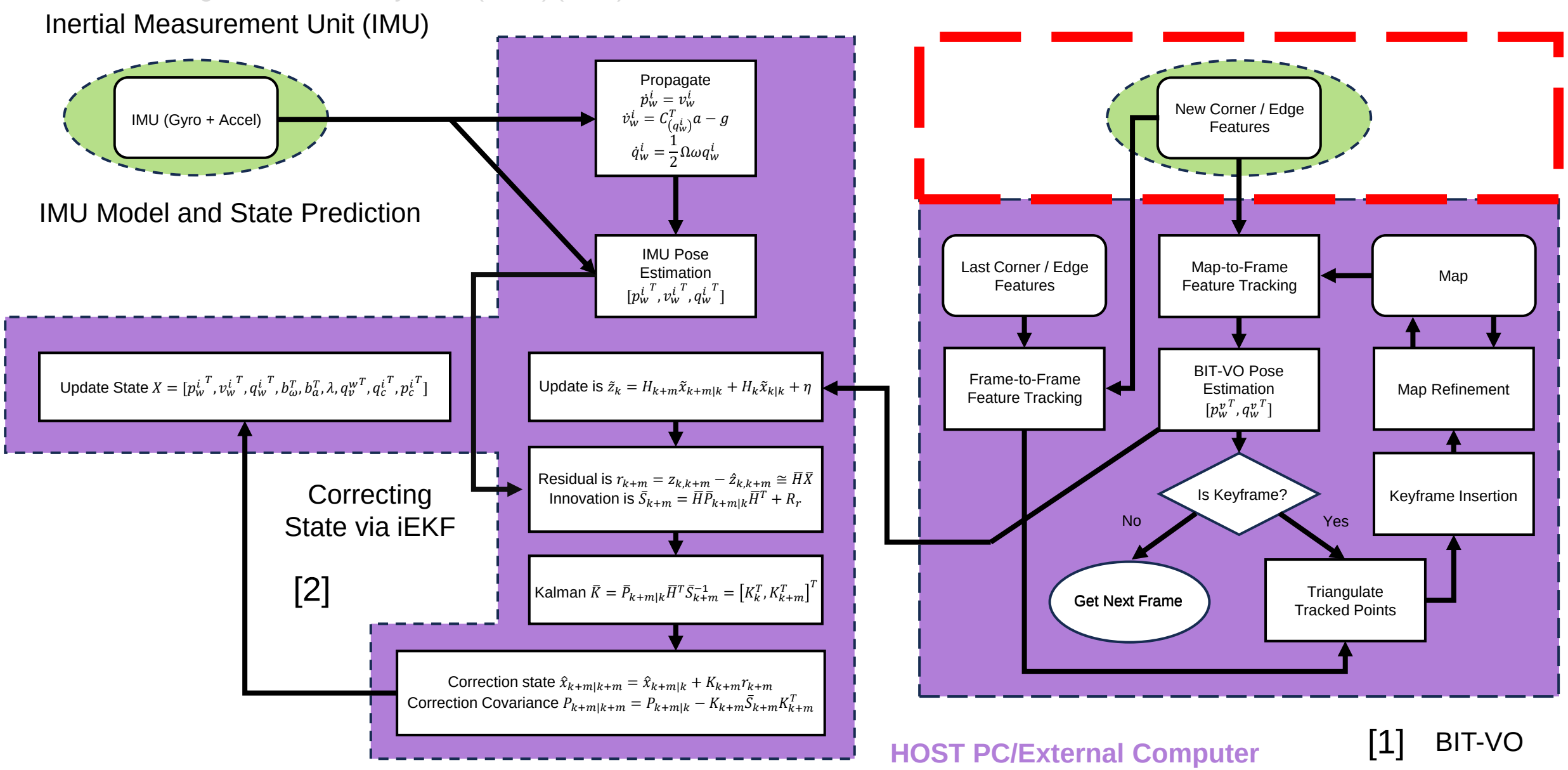
(III) Comparison with BIT-VO and ground-truth, tested on a moving robotic system against a motion capture system.



This work is the first steps toward a 6-DOF SLAM or VIO algorithm using the advantages of the FPSP camera technology.

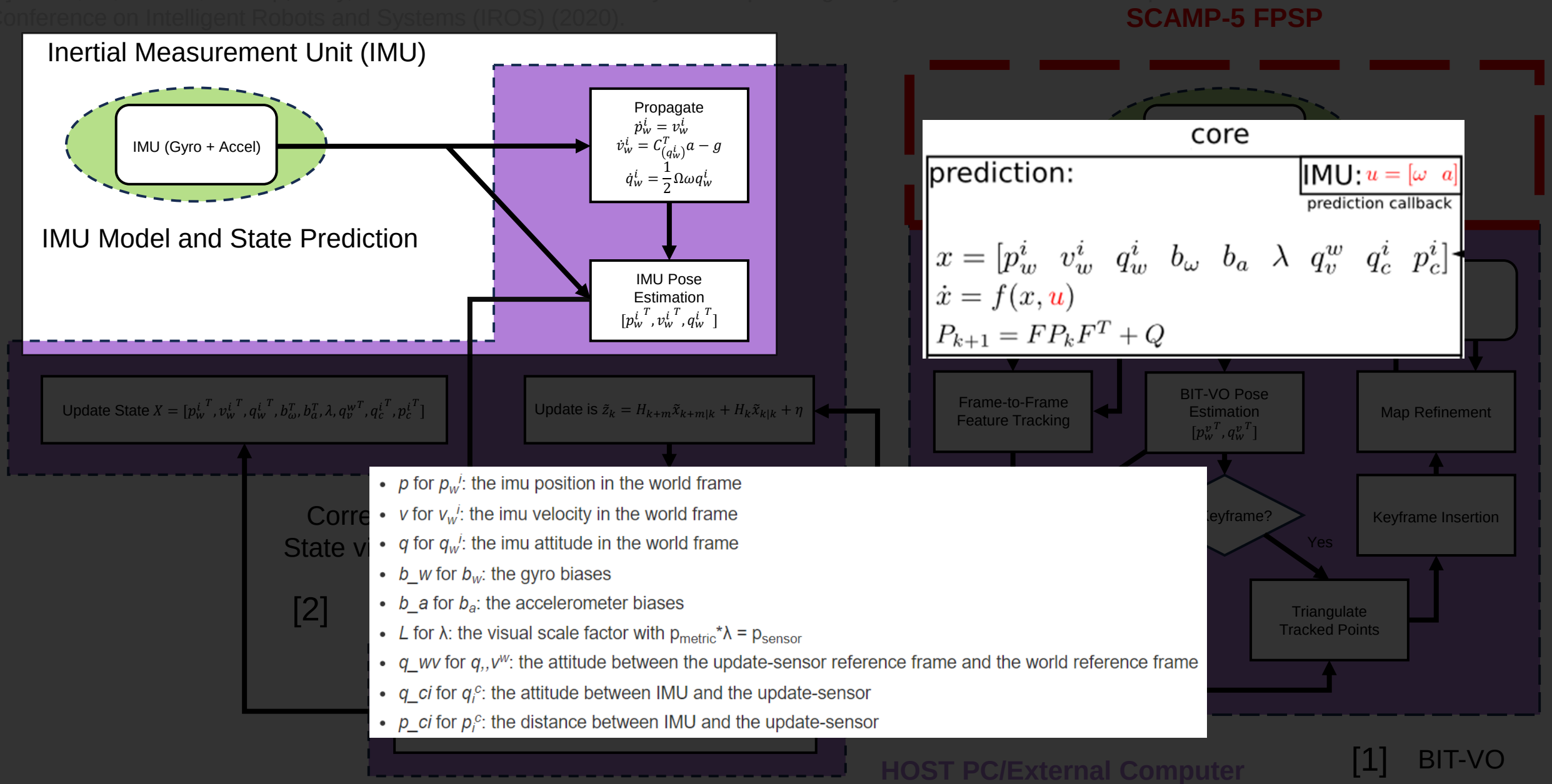
III. Critical Design

[1] Murai, R., Saeedi, S. amp; Kelly, P. H. Bit-VO: Visual odometry at 300 fps using binary features from the focal plane. 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (2020).



[2] Simon Lynen, Markus W Achtelik, Stephan Weiss, Margarita Chli, and Roland Siegwart. A robust and modular multi-sensor fusion approach applied to MAV navigation. In 2013 IEEE/RSJ international conference on intelligent robots and systems, pages 3923–3929. IEEE, 2013.

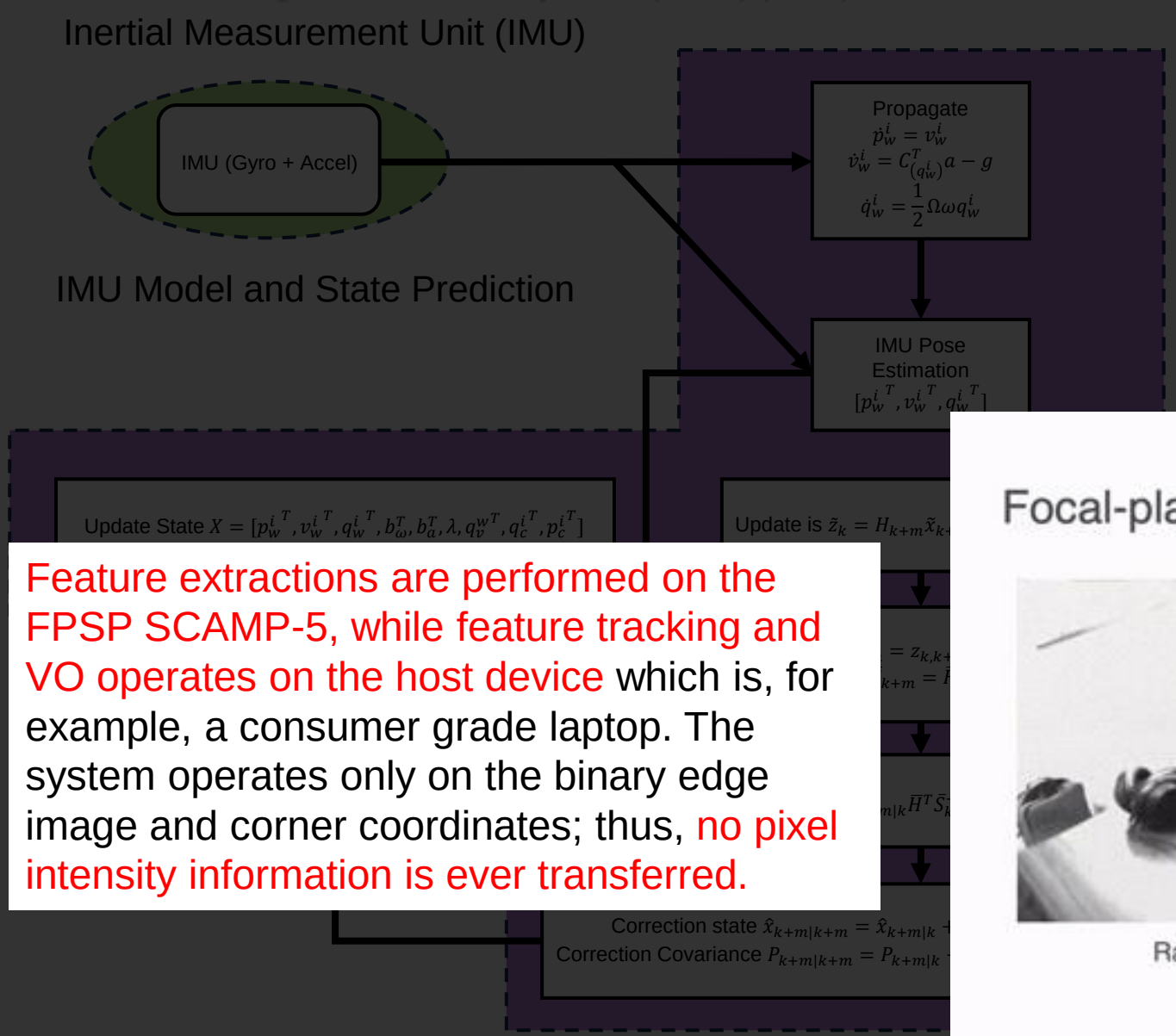
[1] Murai, R., Saeedi, S. amp; Kelly, P. H. Bit-VO: Visual odometry at 300 fps using binary features from the focal plane. 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (2020).

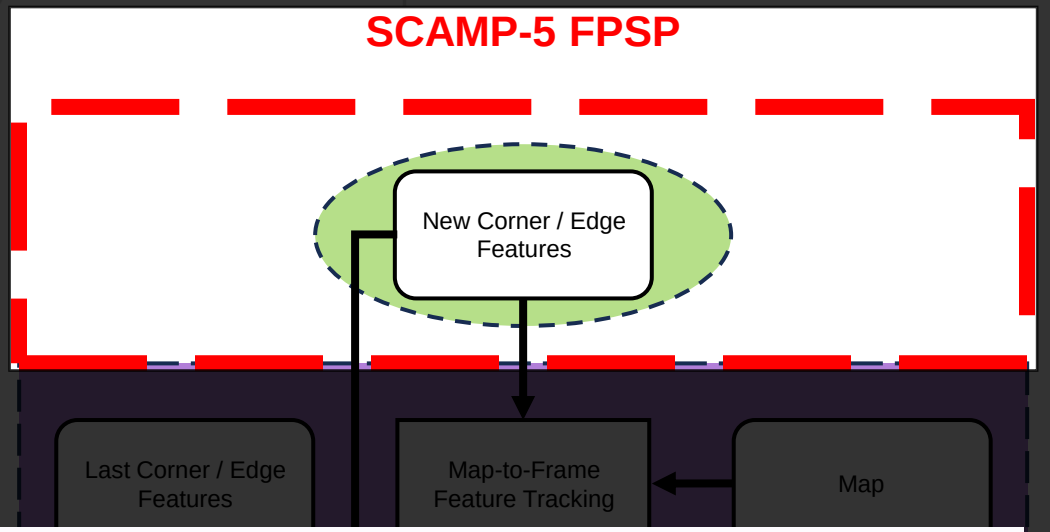


- *p* for p_w^i: the imu position in the world frame
- *v* for v_w^i: the imu velocity in the world frame
- *q* for q_w^i: the imu attitude in the world frame
- *b_w* for b_w: the gyro biases
- *b_a* for b_a: the accelerometer biases
- *L* for λ: the visual scale factor with $p_{metric} * \lambda = p_{sensor}$
- *q_wv* for $q_{,,v}^w$: the attitude between the update-sensor reference frame and the world reference frame
- *q_ci* for q_i^c: the attitude between IMU and the update-sensor
- *p_ci* for p_i^c: the distance between IMU and the update-sensor

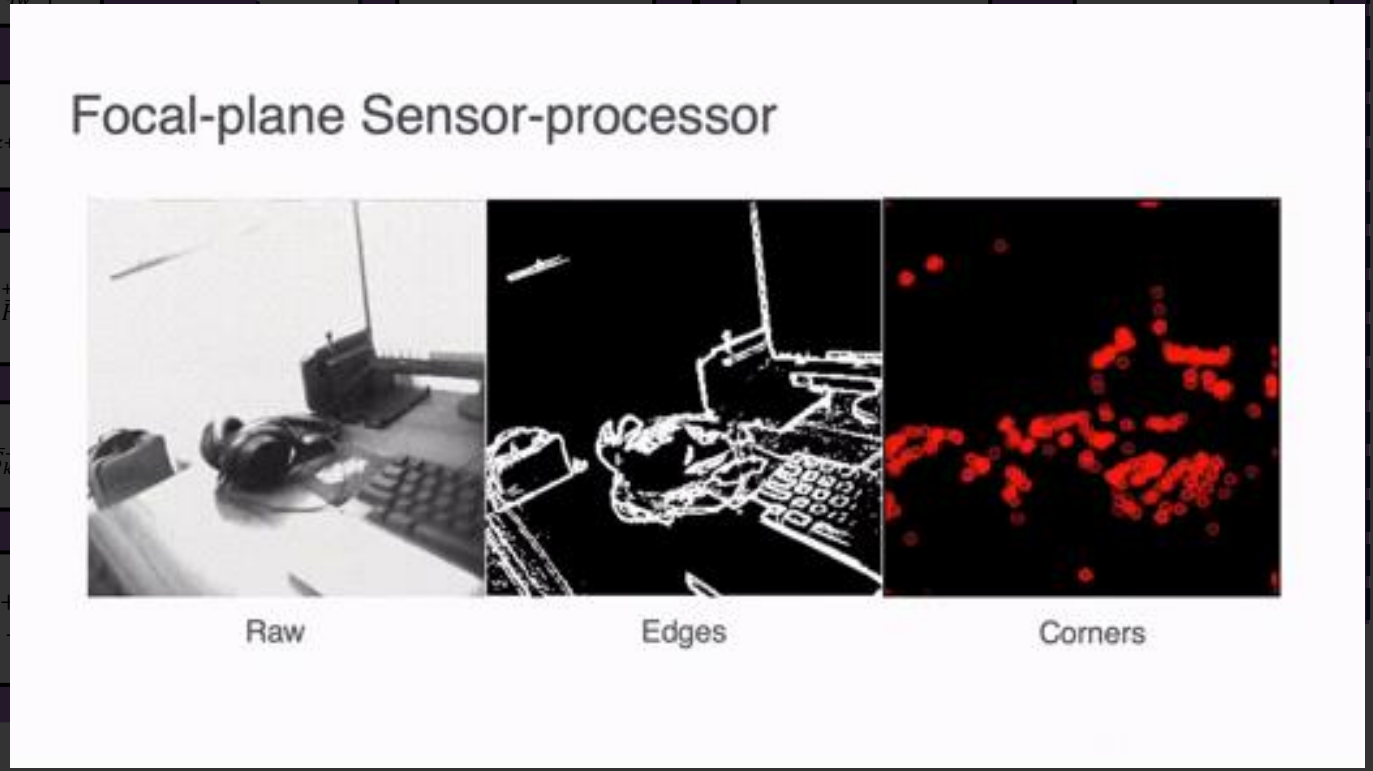
[2] Simon Lynen, Markus W Achtelik, Stephan Weiss, Margarita Chli, and Roland Siegwart. A robust and modular multi-sensor fusion approach applied to MAV navigation. In 2013 IEEE/RSJ international conference on intelligent robots and systems, pages 3923–3929. IEEE, 2013.

[1] Murai, R., Saeedi, S. amp; Kelly, P. H. Bit-VO: Visual odometry at 300 fps using binary features from the focal plane. 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (2020).



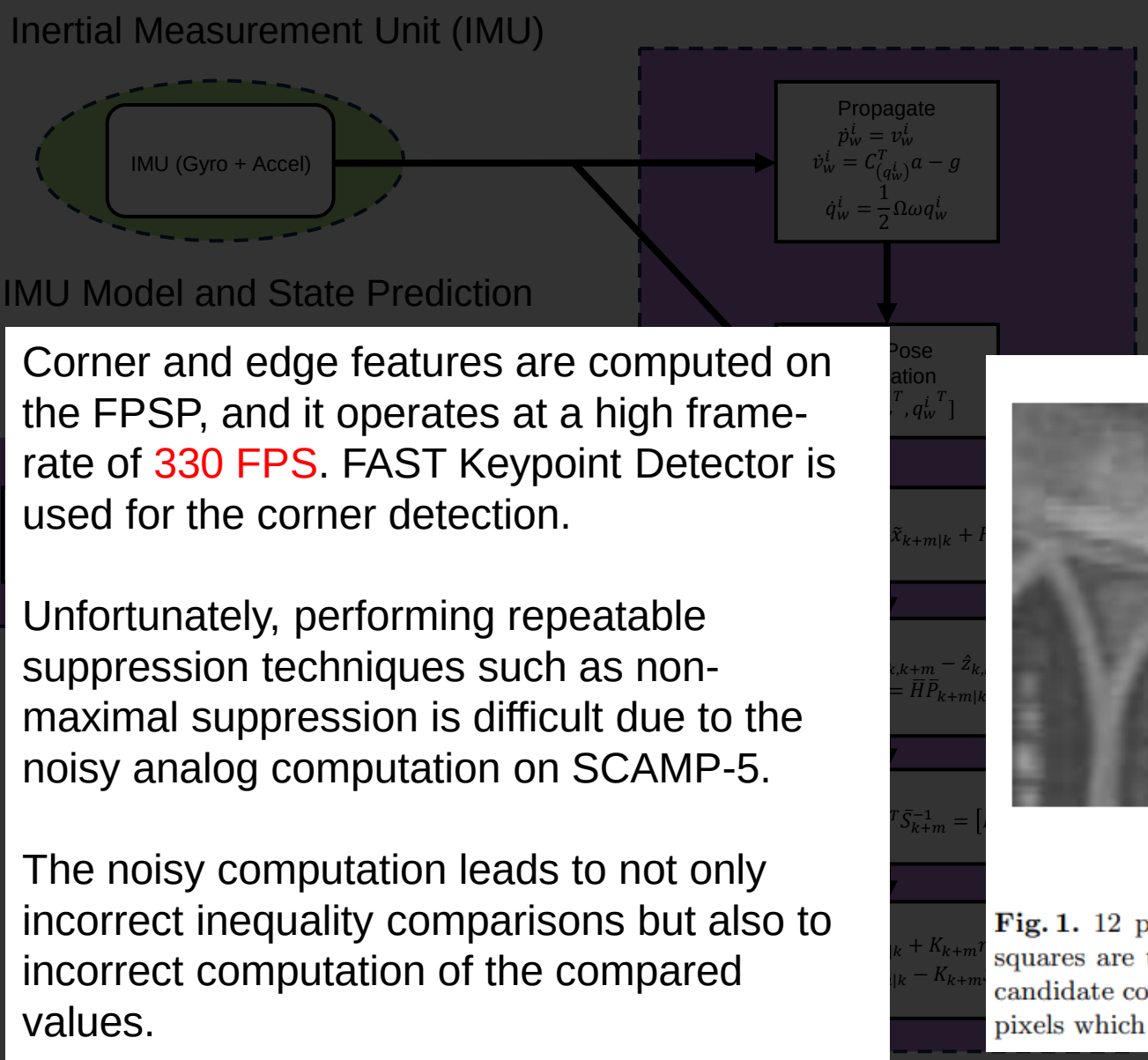


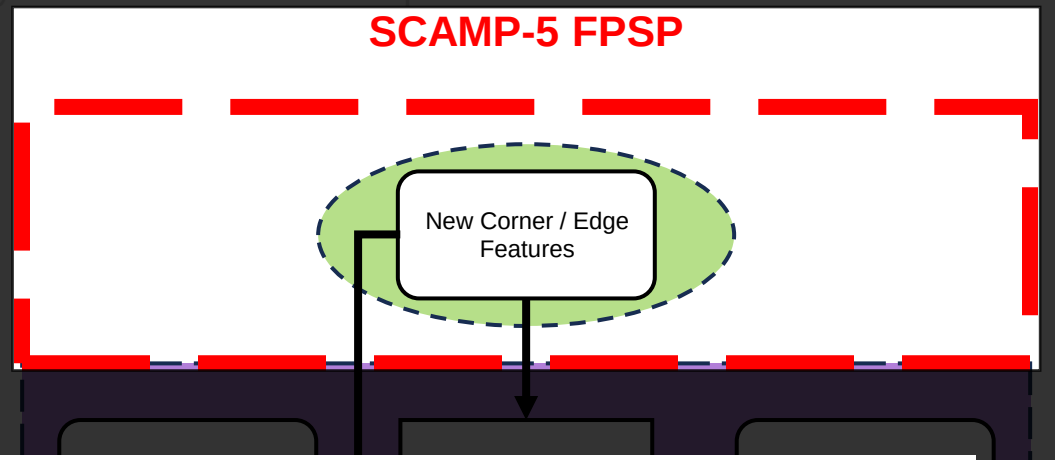
Feature extractions are performed on the FPSP SCAMP-5, while feature tracking and VO operates on the host device which is, for example, a consumer grade laptop. The system operates only on the binary edge image and corner coordinates; thus, no pixel intensity information is ever transferred.



[2] Simon Lynen, Markus W Achtelik, Stephan Weiss, Margarita Chli, and Roland Siegwart. A robust and modular multi-sensor fusion approach applied to MAV navigation. In 2013 IEEE/RSJ international conference on intelligent robots and systems, pages 3923–3929. IEEE, 2013.

[1] Murai, R., Saeedi, S. amp; Kelly, P. H. Bit-VO: Visual odometry at 300 fps using binary features from the focal plane. 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (2020).





Corner and edge features are computed on the FPSP, and it operates at a high frame-rate of 330 FPS. FAST Keypoint Detector is used for the corner detection.

Unfortunately, performing repeatable suppression techniques such as non-maximal suppression is difficult due to the noisy analog computation on SCAMP-5.

The noisy computation leads to not only incorrect inequality comparisons but also to incorrect computation of the compared values.

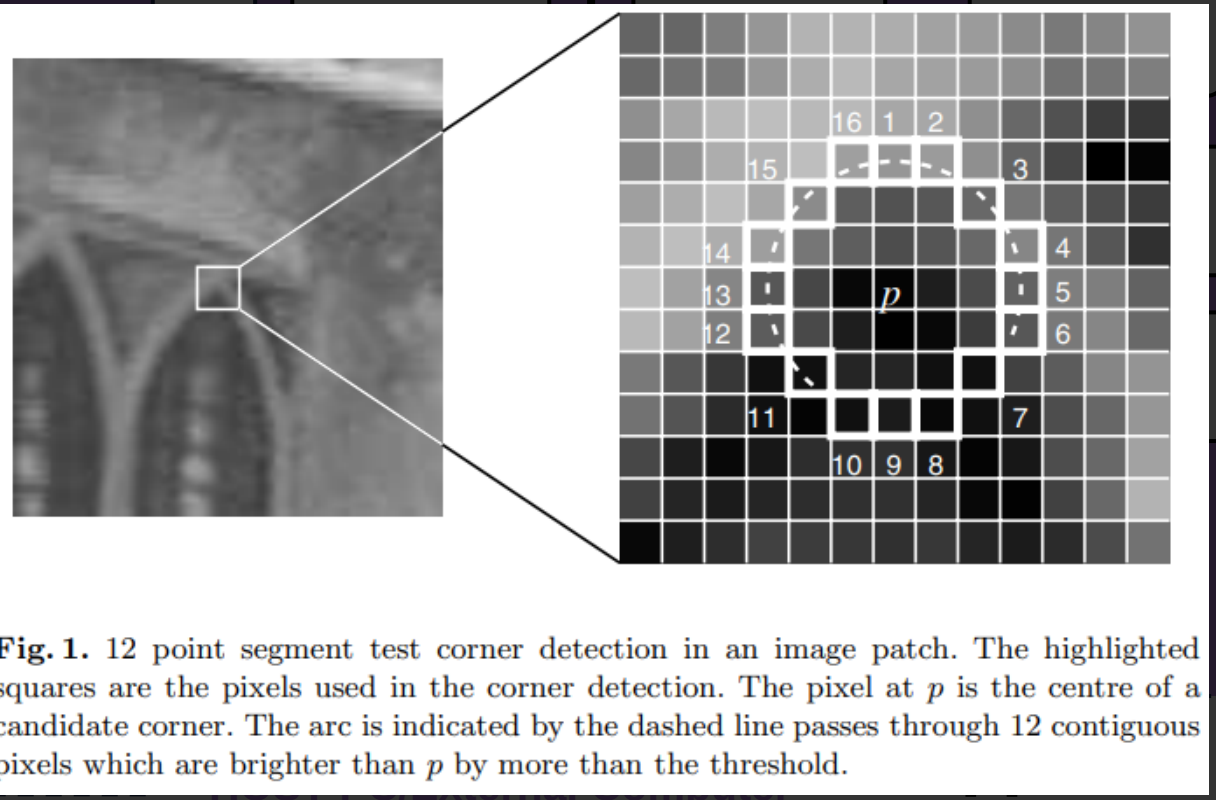


Fig. 1. 12 point segment test corner detection in an image patch. The highlighted squares are the pixels used in the corner detection. The pixel at p is the centre of a candidate corner. The arc is indicated by the dashed line passes through 12 contiguous pixels which are brighter than p by more than the threshold.

[2] Simon Lynen, Markus W Achtelik, Stephan Weiss, Margarita Chli, and Roland Siegwart. A robust and modular multi-sensor fusion approach applied to MAV navigation. In 2013 IEEE/RSJ international conference on intelligent robots and systems, pages 3923–3929. IEEE, 2013.

[1] Murai, R., Saeedi, S. amp; Kelly, P. H. Bit-VO: Visual odometry at 300 fps using binary features from the focal plane. 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (2020).

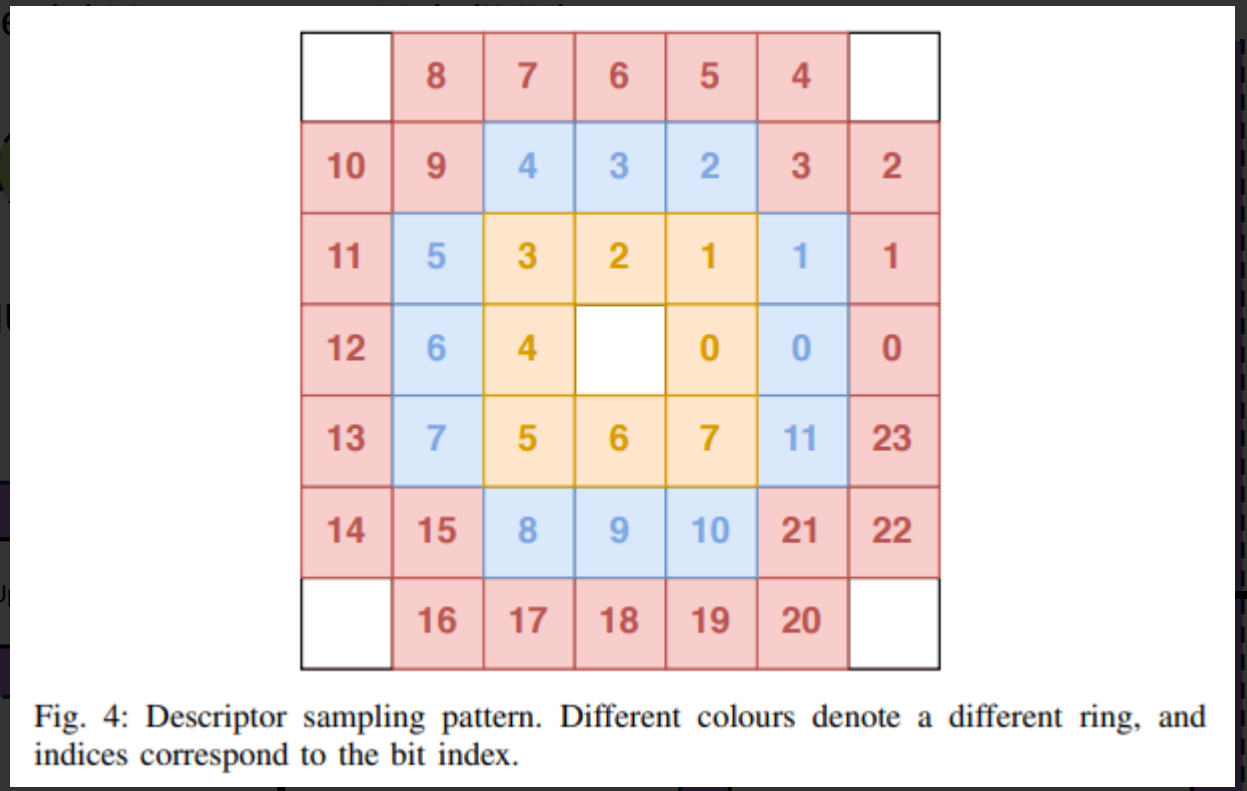
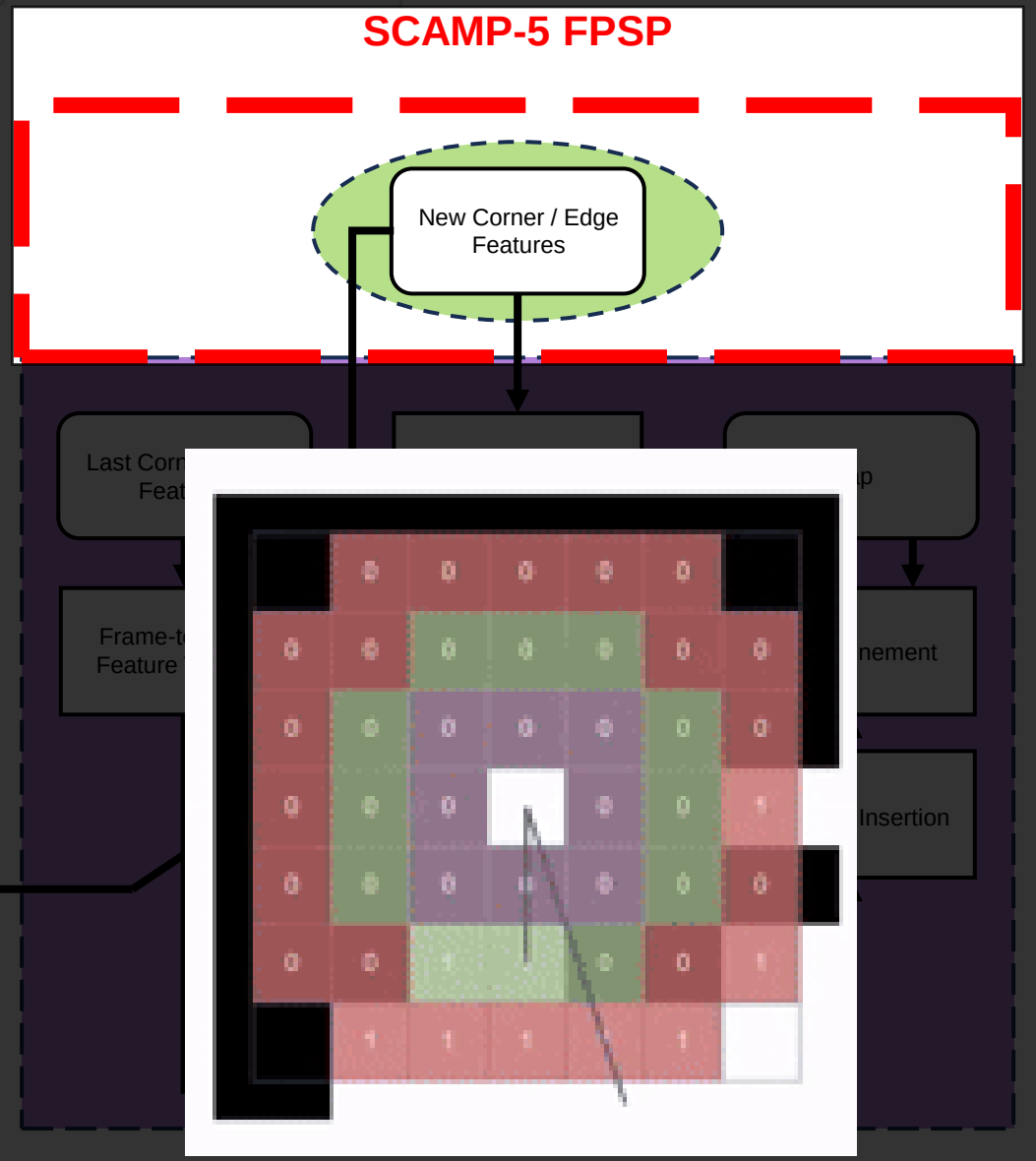


Fig. 4: Descriptor sampling pattern. Different colours denote a different ring, and indices correspond to the bit index.



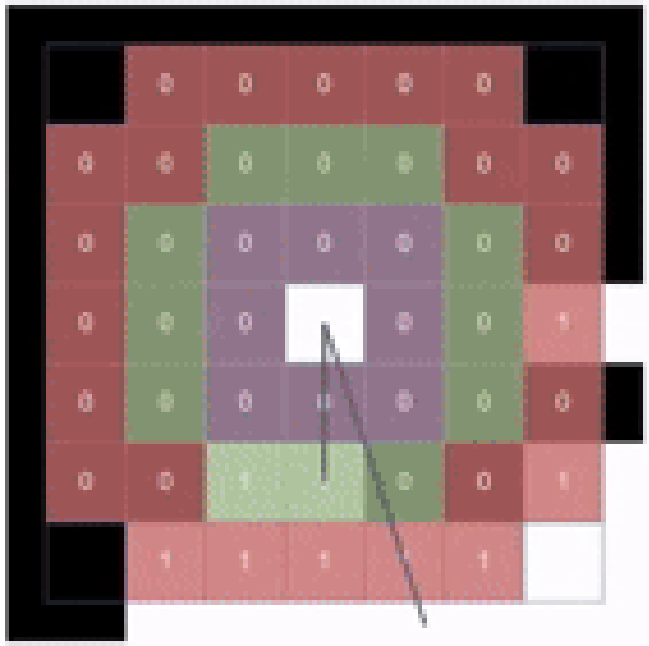
$$\theta = \tan^{-1}\frac{\sum_{x,y} yG(x,y)}{\sum_{x,y} xG(x,y)} \quad (1)$$

$$\theta = \tan^{-1}\frac{\sum_{x,y} yB(x,y)}{\sum_{x,y} xB(x,y)} \quad (2)$$

The rotation invariance is achieved by bit-rotations of the rings independently, based on the orientation θ.

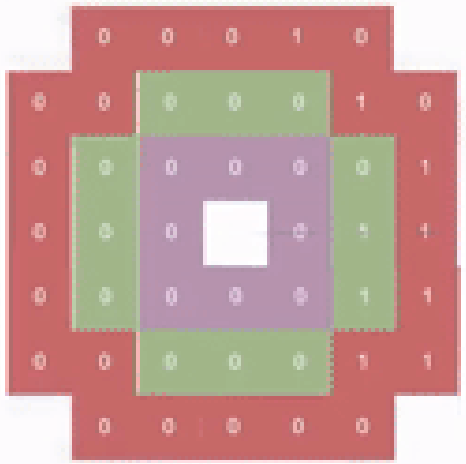
[1] BIT-VO

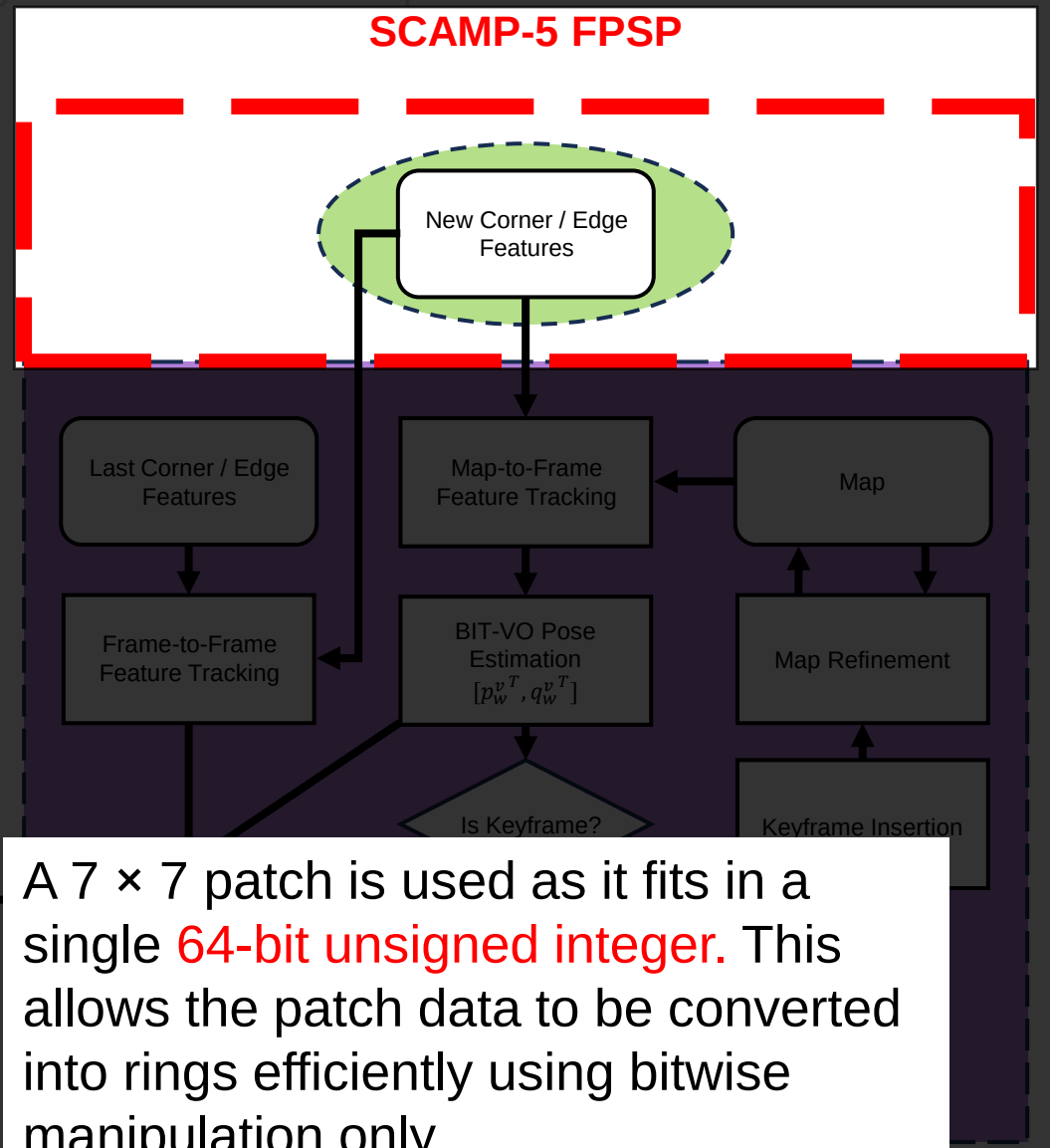
[2] Simon Lynen, Markus W. Achtelik, Stephan Weiss, Margarita Chli, ... to MAV navigation. In 2013 IEEE/RSJ international conference on intell...



$$\theta = \tan^{-1}\frac{\sum_{x,y} yG(x,y)}{\sum_{x,y} xG(x,y)} \quad (1)$$

$$\theta = \tan^{-1}\frac{\sum_{x,y} yB(x,y)}{\sum_{x,y} xB(x,y)} \quad (2)$$

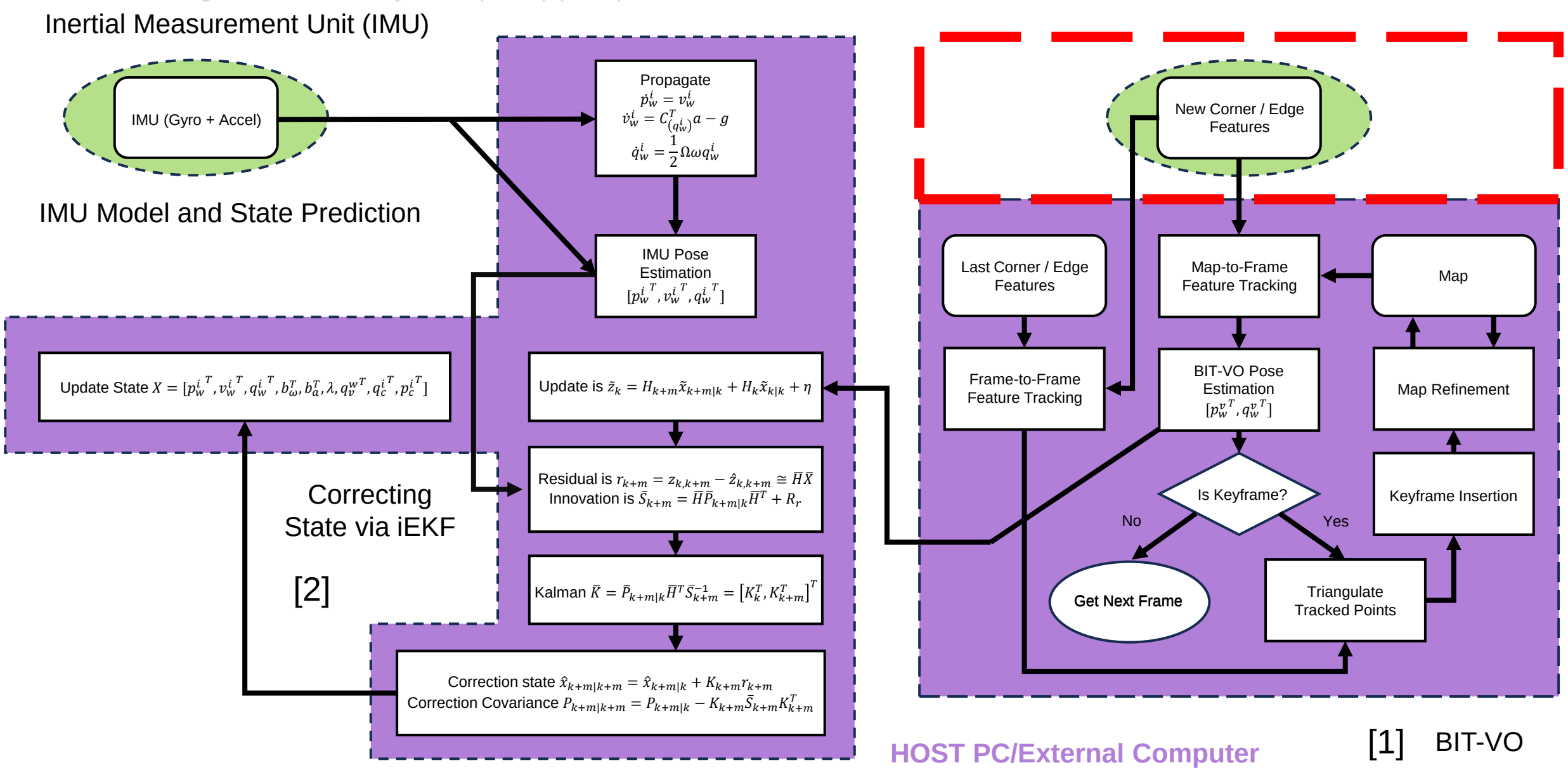




A 7 × 7 patch is used as it fits in a single 64-bit unsigned integer. This allows the patch data to be converted into rings efficiently using bitwise manipulation only.

[1] BIT-VO

[1] Murai, R., Saeedi, S. amp; Kelly, P. H. Bit-VO: Visual odometry at 300 fps using binary features from the focal plane. 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (2020).

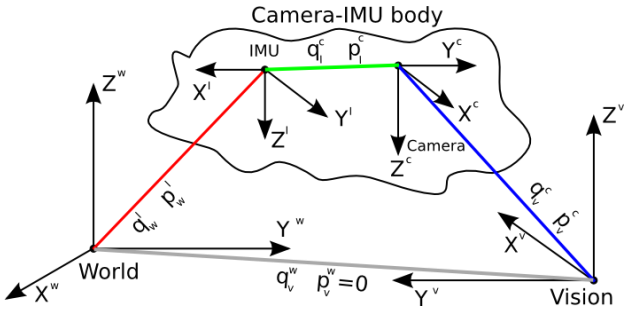


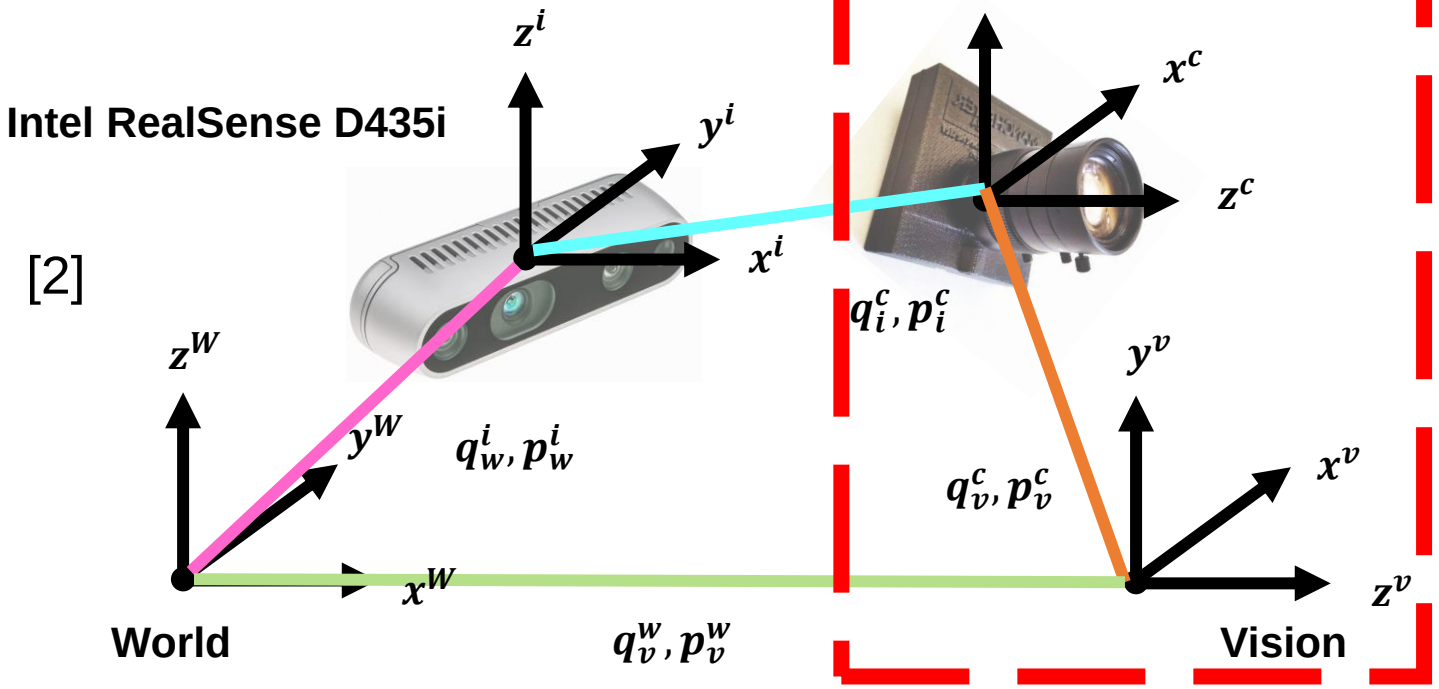
[2] Simon Lynen, Markus W Achtelik, Stephan Weiss, Margarita Chli, and Roland Siegwart. A robust and modular multi-sensor fusion approach applied to MAV navigation. In 2013 IEEE/RSJ international conference on intelligent robots and systems, pages 3923–3929. IEEE, 2013.

[2] Simon Lynen, Markus W Achtelik, Stephan Weiss, Margarita Chli, and Roland Siegwart. A robust and modular multi-sensor fusion approach applied to MAV navigation. In 2013 IEEE/RSJ international conference on intelligent robots and systems, pages 3923–3929. IEEE, 2013.

IMU and SCAMP-5 FPSP Frames

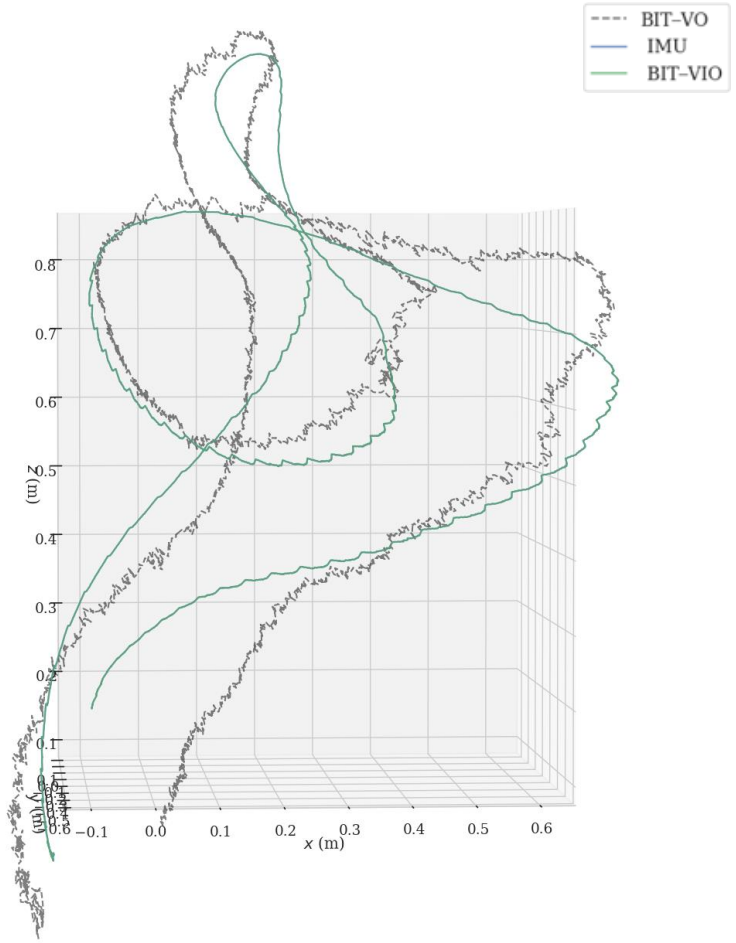
IMU and SCAMP-5 FPSP frames. Intel RealSense D435i IMU is used in this work, and SCAMP-5 is the FPSP used as a camera sensor. Four coordinate frames, with two being a part of SCAMP-5 FPSP (camera and vision frames).



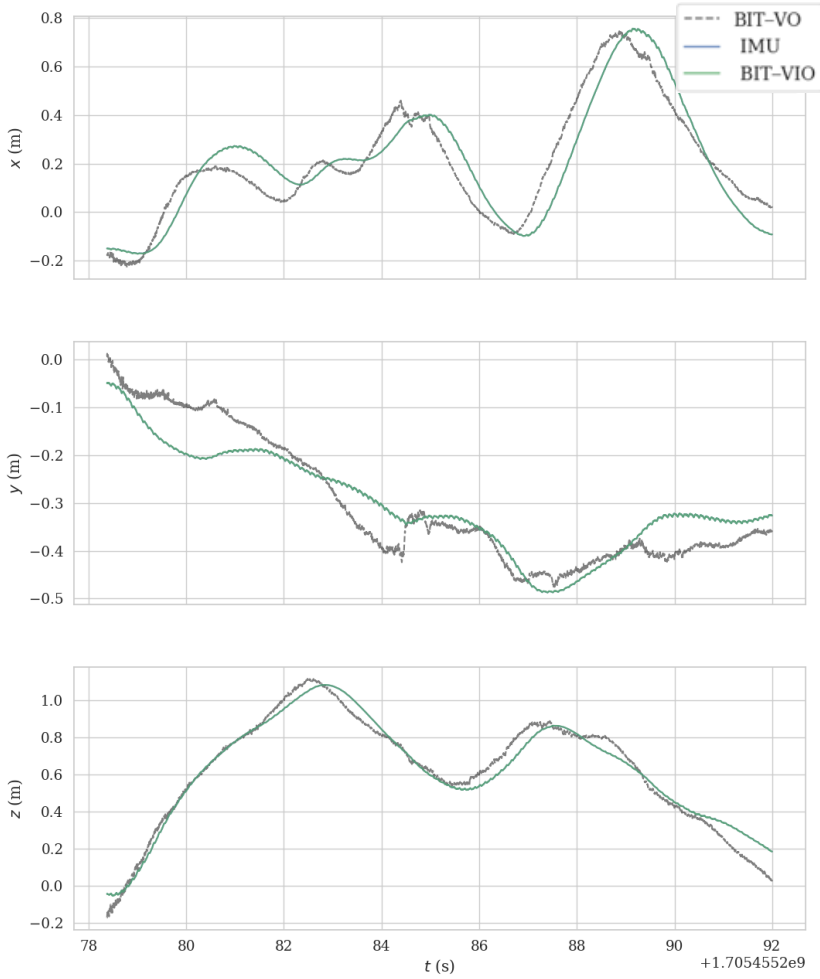


IV. Results and Progress

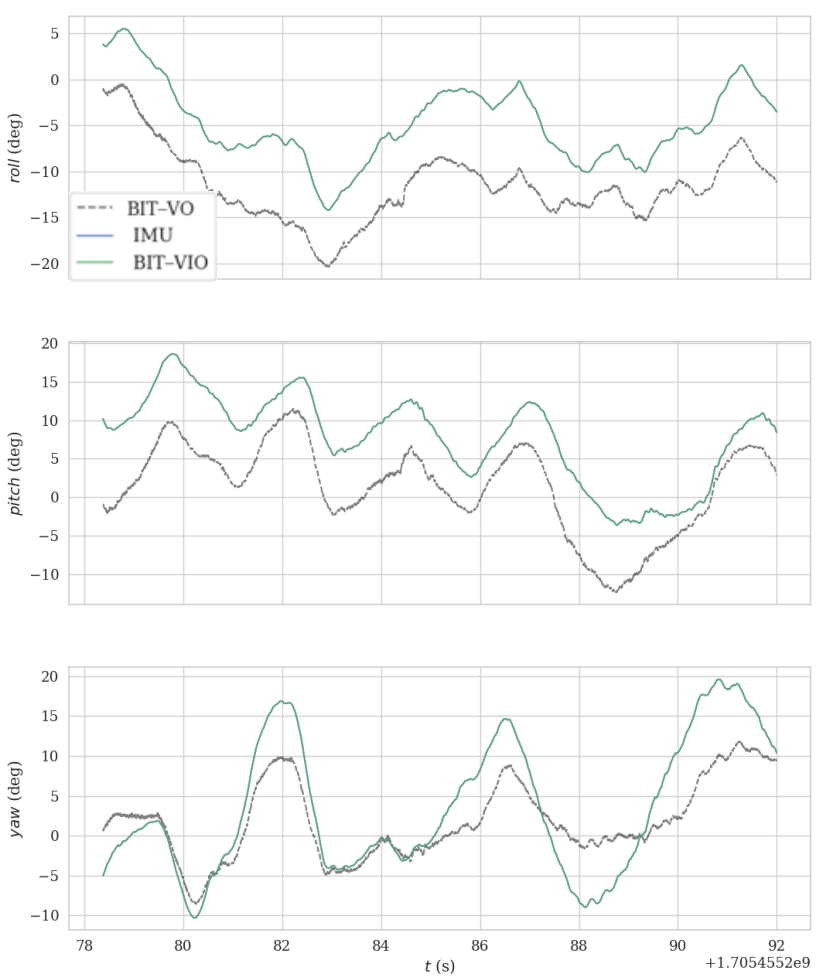
Compared just BIT-VO with BIT-VIO and IMU



Traj 9/20 ~ 3.4 m



XYZ strong

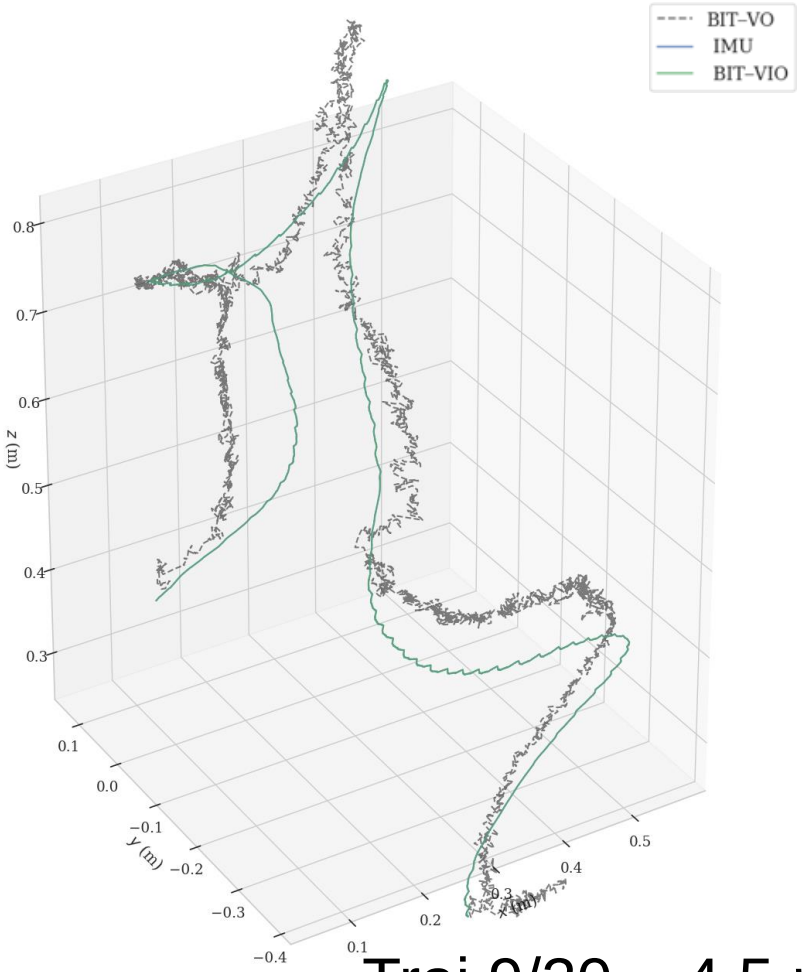


RPY trends (~5deg)

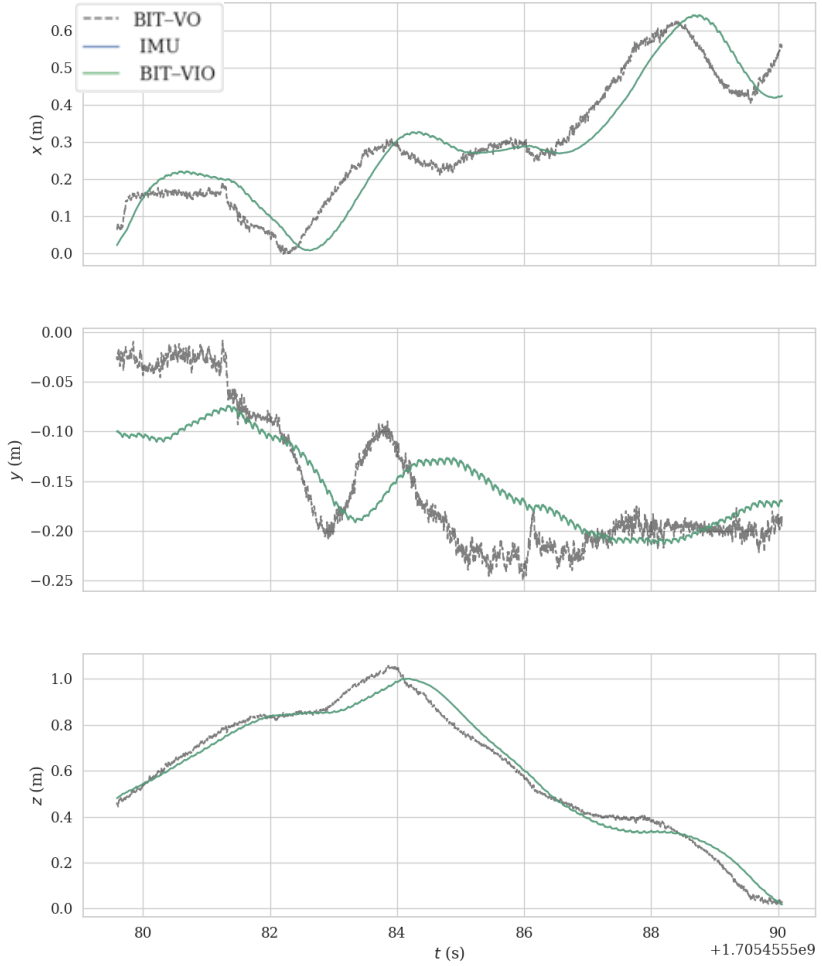
Full 6-DOF long trajs., fast motions
Is fully achievable as we solved for both
Now time to run experiments with GT.

Our BIT-VIO silences the high frequency noise of prior BIT-VO.

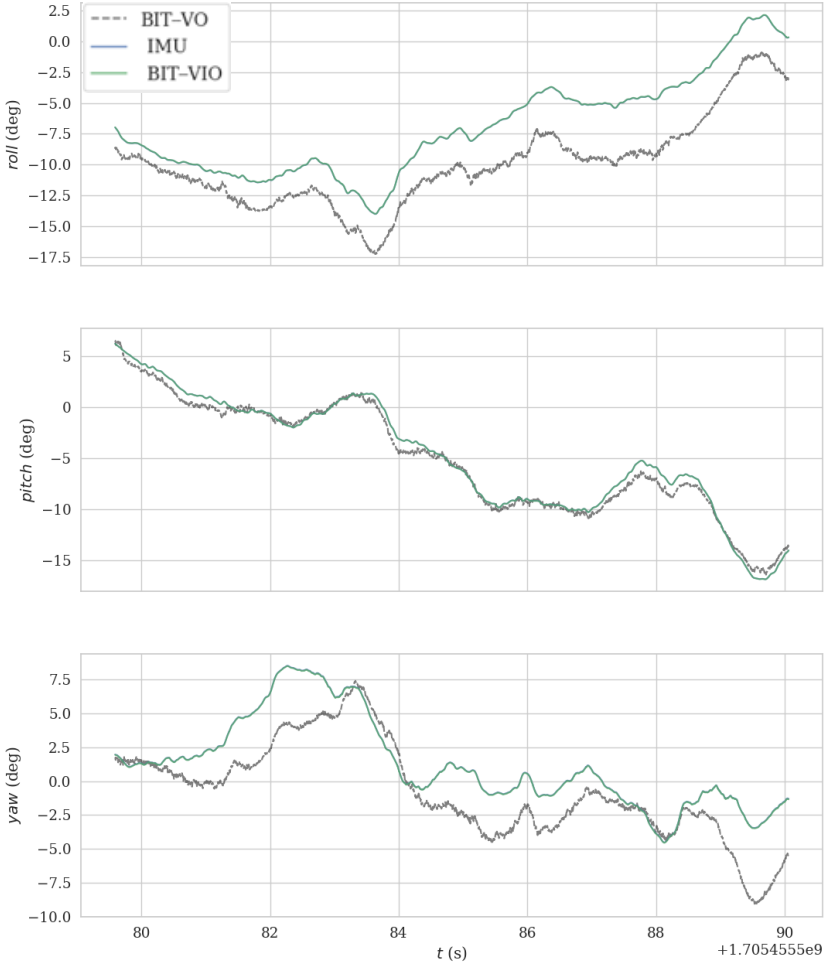
Compared just BIT-VO with BIT-VIO and IMU



Traj 9/20 ~ 4.5 m



XYZ strong

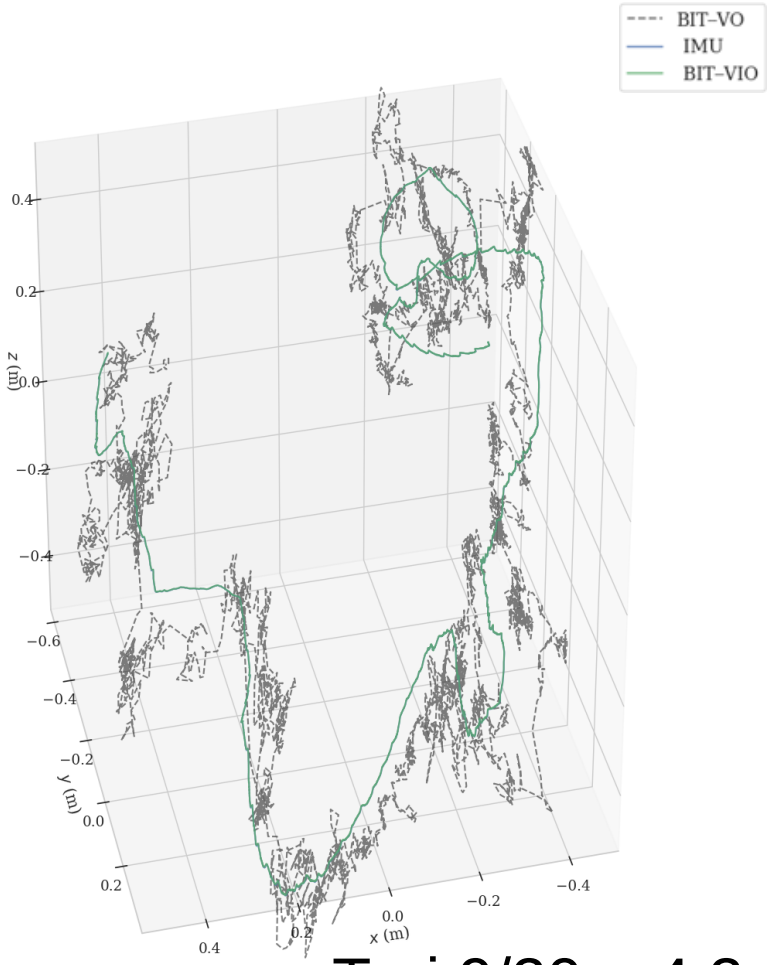


RPY trends (~5deg)

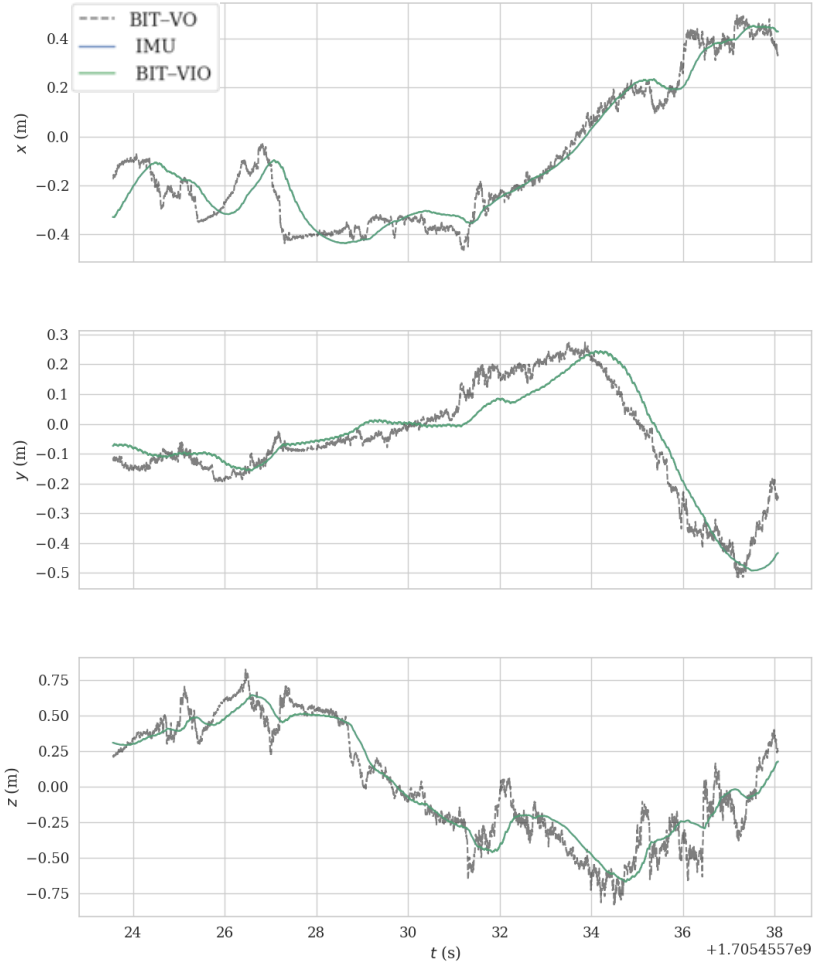
Full 6-DOF long trajs., fast motions
Is fully achievable as we solved for both
Now time to run experiments with GT.

Our BIT-VIO silences the high frequency noise of prior BIT-VO.

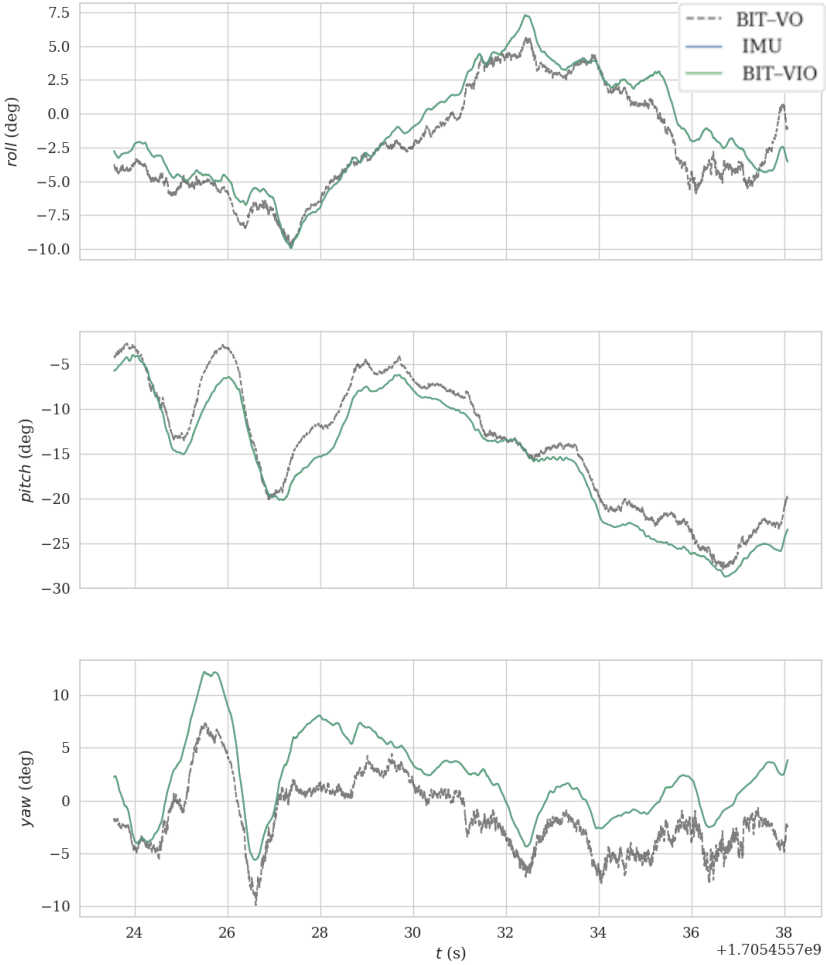
Compared just BIT-VO with BIT-VIO and IMU



Traj 9/20 ~ 4.2 m



XYZ strong

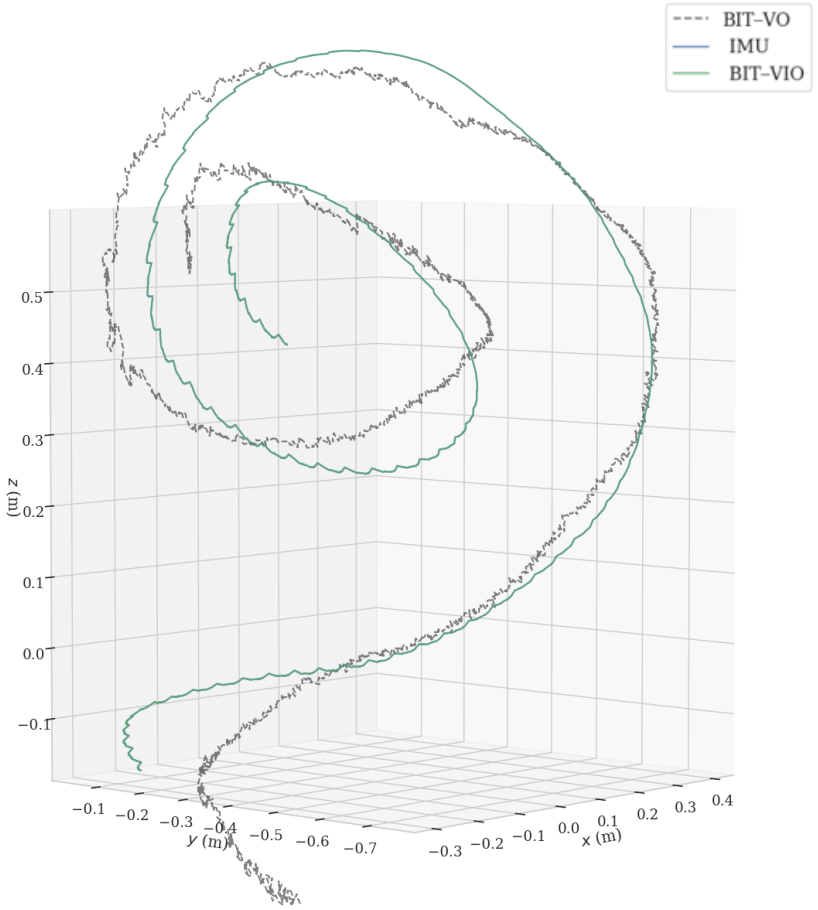


RPY trends (~5deg)

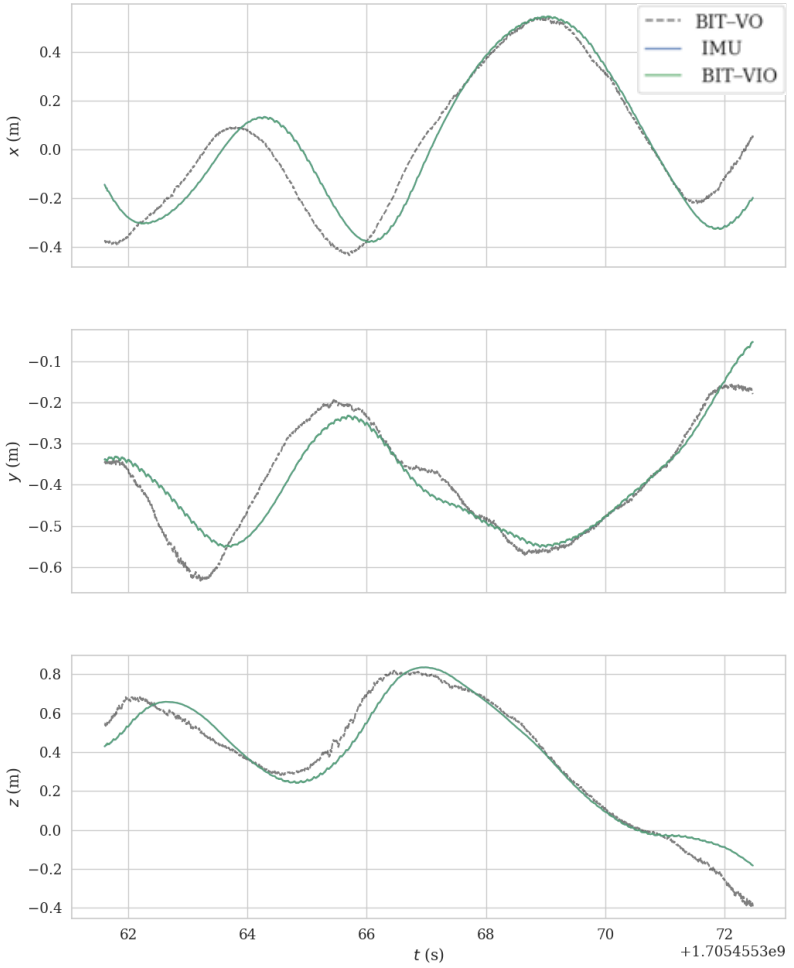
Full 6-DOF long trajs., fast motions
Is fully achievable as we solved for both
Now time to run experiments with GT.

Our BIT-VIO silences the high frequency noise of prior BIT-VO.

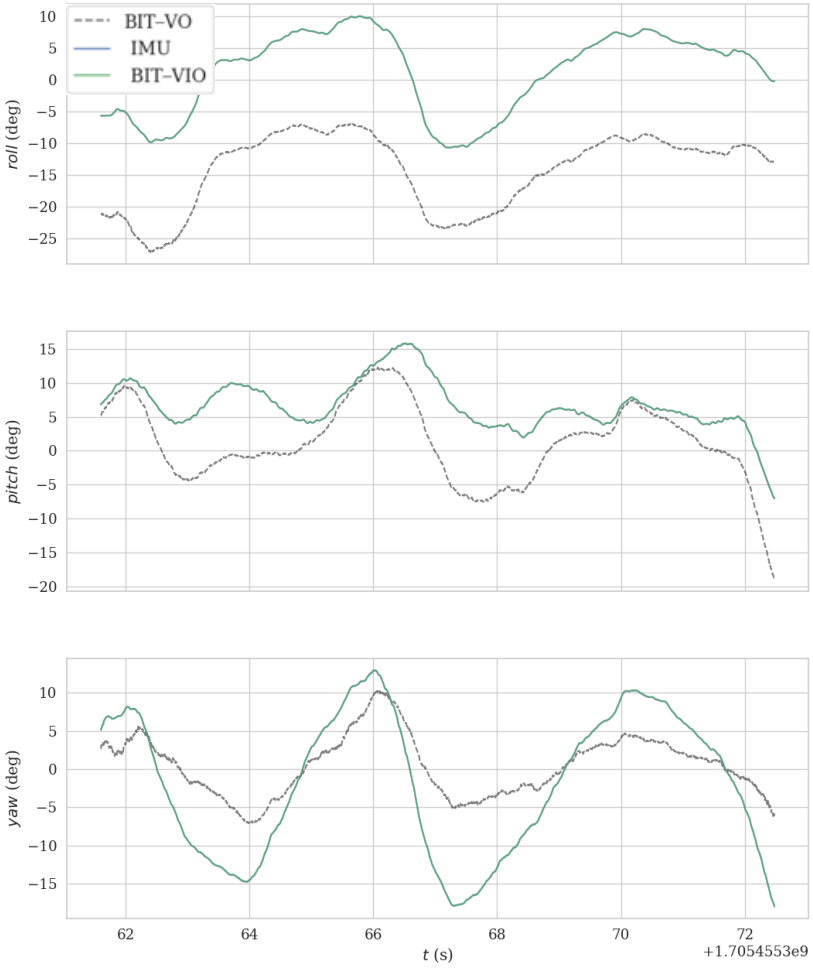
Compared just BIT-VO with BIT-VIO and IMU



Traj 6/20 ~ 5.4 m



XYZ strong



RPY trends (~5deg)

Full 6-DOF long trajs., fast motions
Is fully achievable as we solved for both
Now time to run experiments with GT.

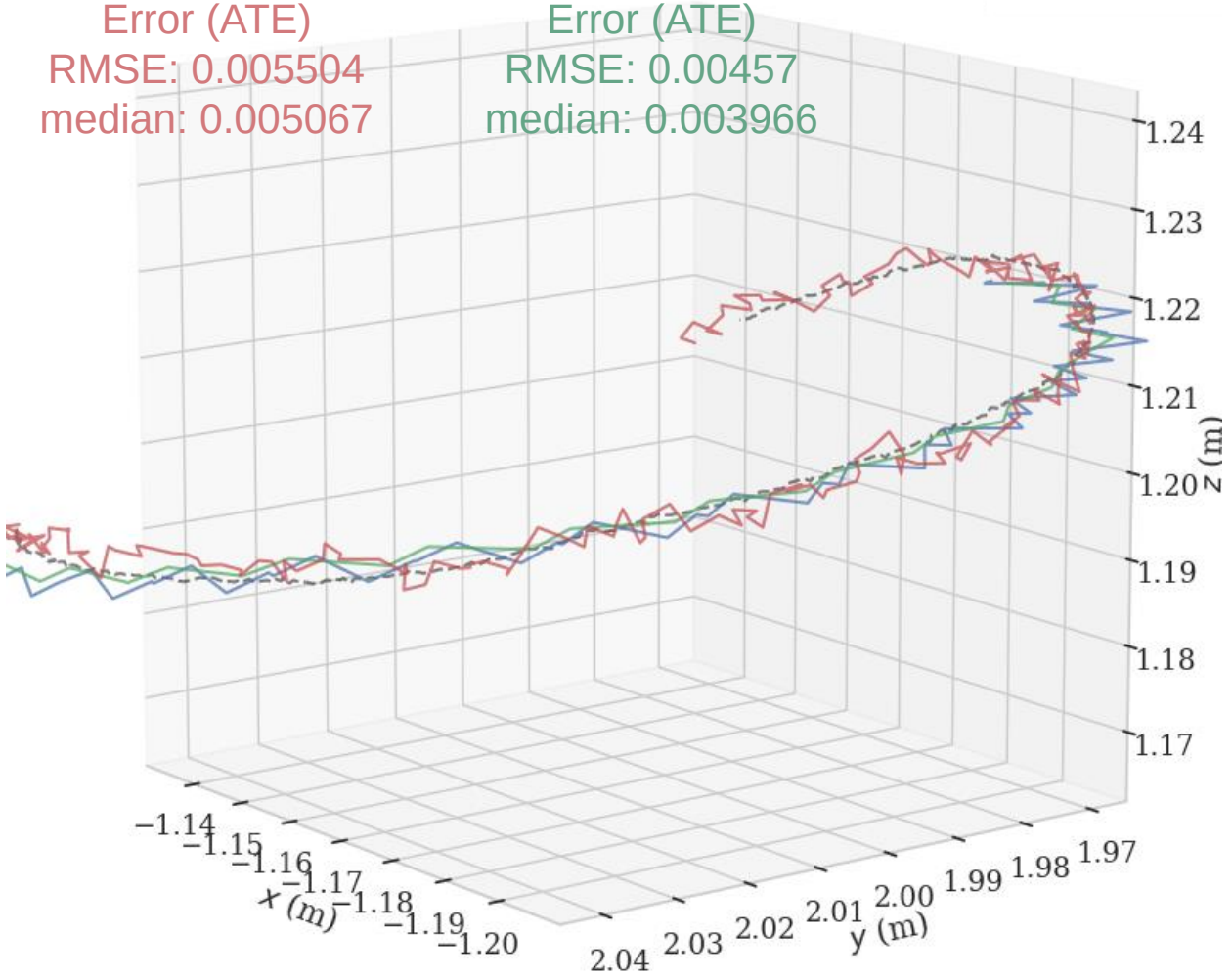
Our BIT-VIO silences the high frequency noise of prior BIT-VO.

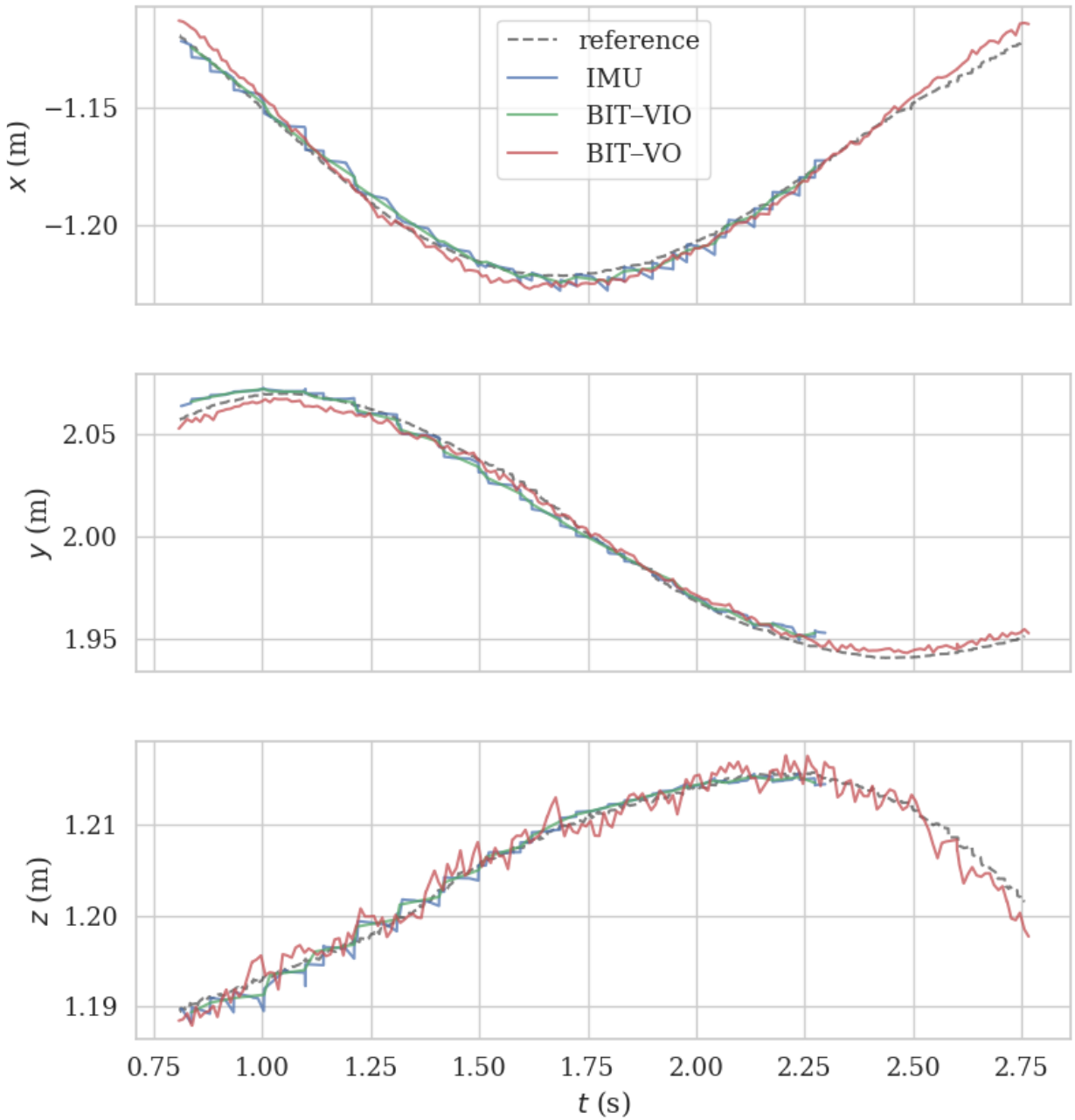
BIT-VO:

Absolute Trajectory Error (ATE)
RMSE: 0.005504
median: 0.005067

BIT-VIO:

Absolute Trajectory Error (ATE)
RMSE: 0.00457
median: 0.003966





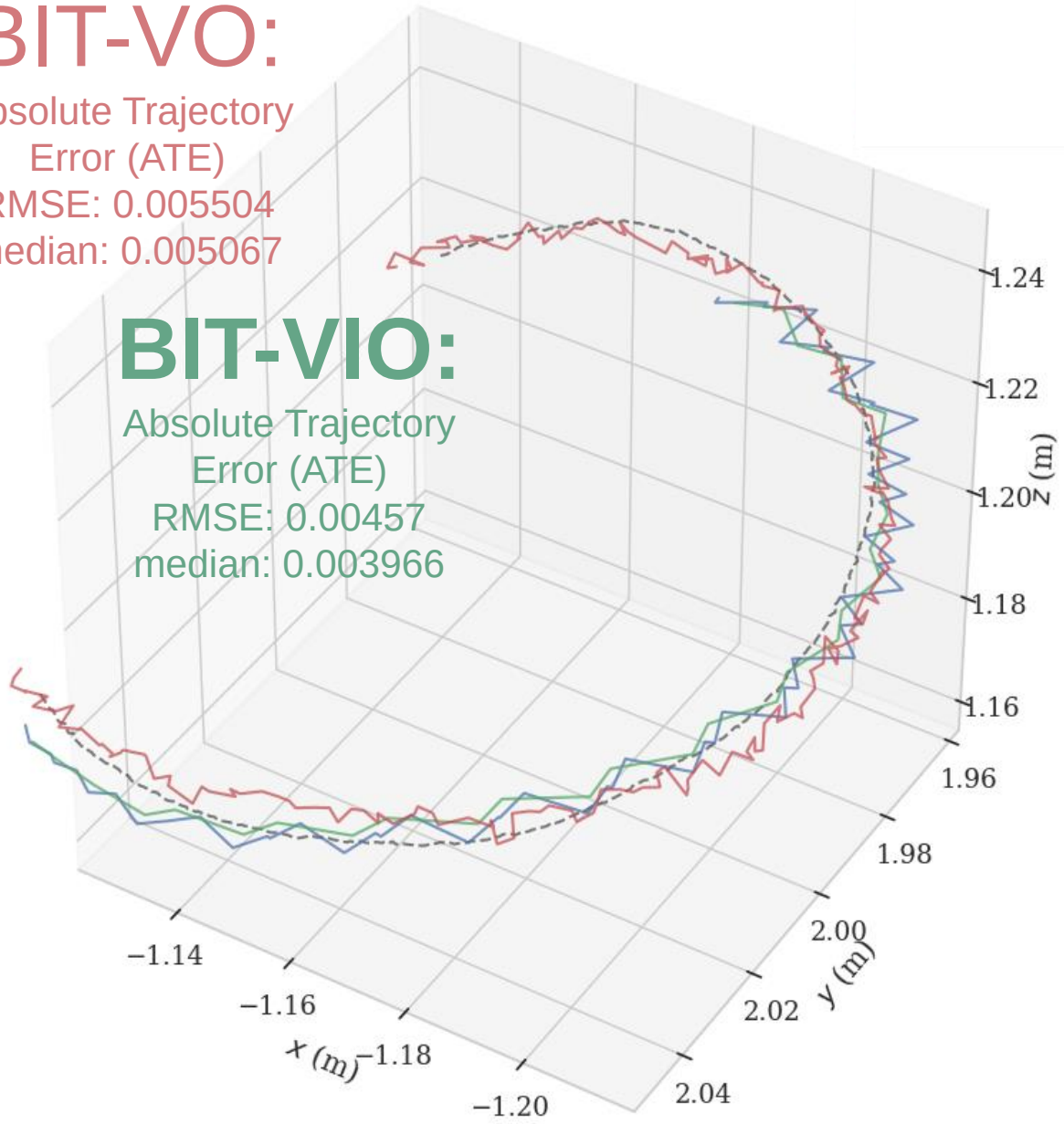
Comparison of the proposed BIT-VIO with inertial (IMU) and visual odometry (BIT-VO) overlaid on the reference ground-truth trajectory. Our BIT-VIO estimates are closer to the ground-truth compared to predictions from IMU or BIT-VO. Notice that our BIT-VIO effectively removes the high frequency noise visible in BIT-VO's trajectory.

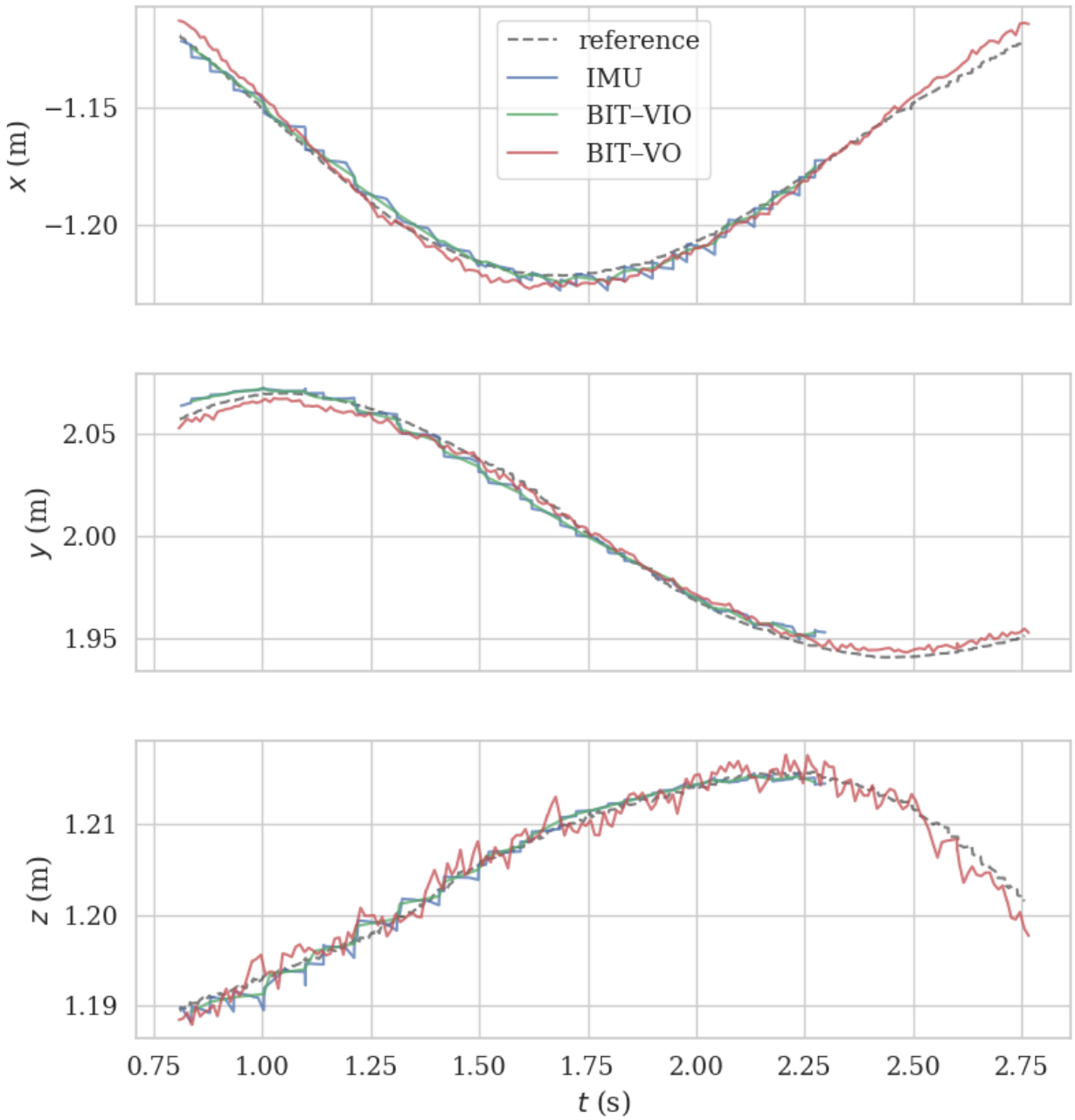
BIT-VO:

Absolute Trajectory Error (ATE)
RMSE: 0.005504
median: 0.005067

BIT-VIO:

Absolute Trajectory Error (ATE)
RMSE: 0.00457
median: 0.003966





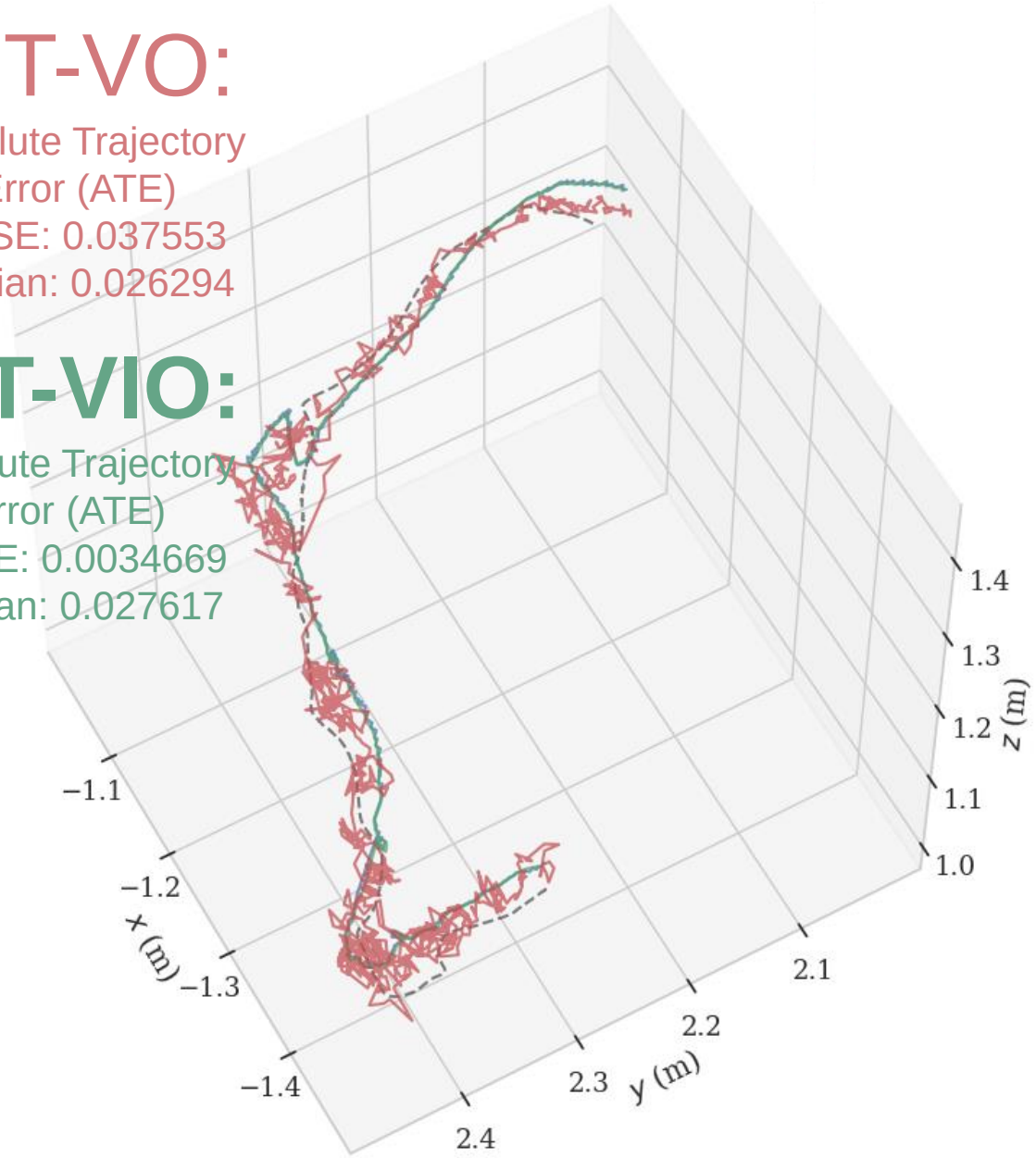
Comparison of the proposed BIT-VIO with inertial (IMU) and visual odometry (BIT-VO) overlaid on the reference ground-truth trajectory. Our BIT-VIO estimates are closer to the ground-truth compared to predictions from IMU or BIT-VO. Notice that our BIT-VIO effectively removes the high frequency noise visible in BIT-VO's trajectory.

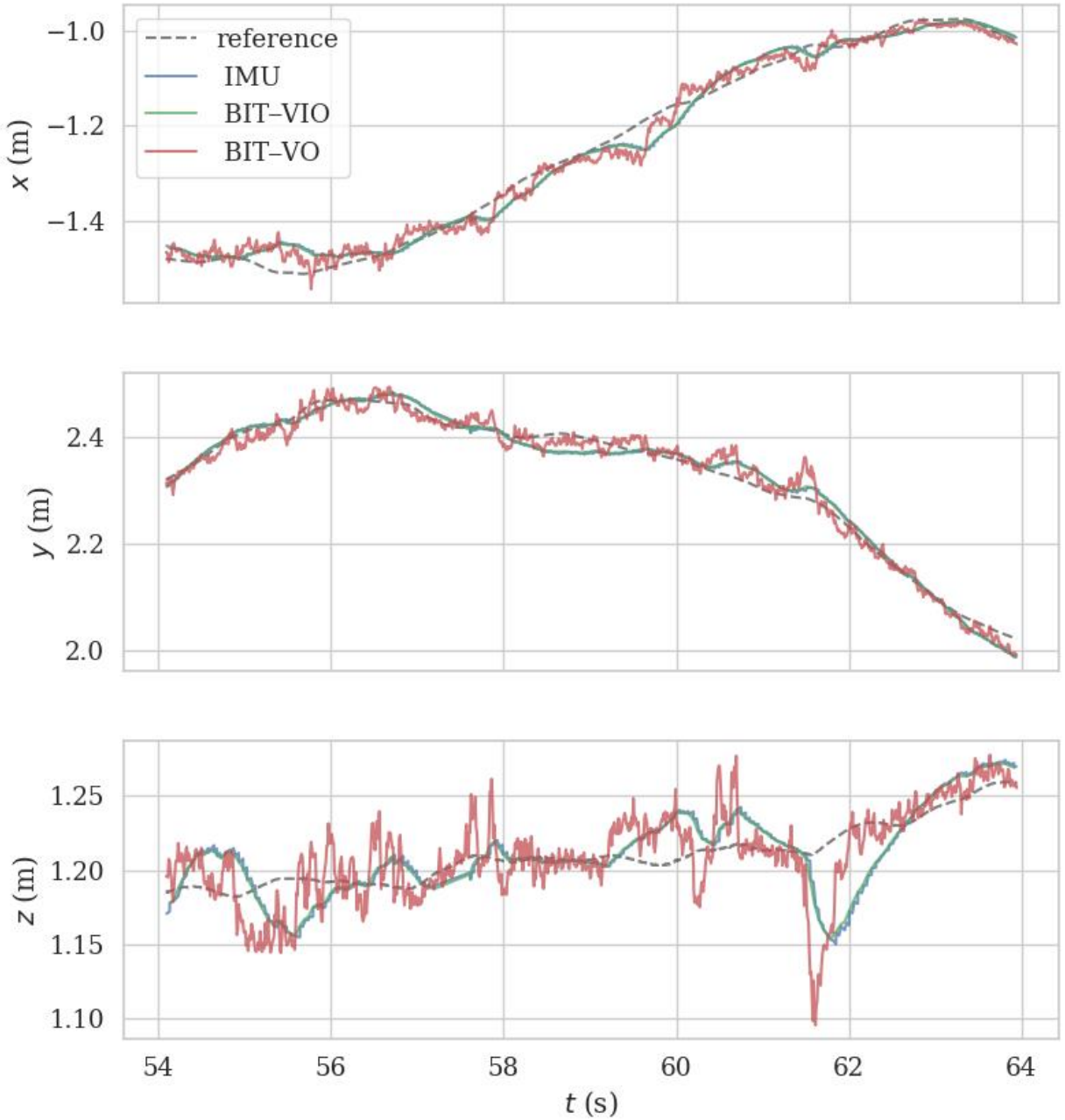
BIT-VO:

Absolute Trajectory Error (ATE)
RMSE: 0.037553
median: 0.026294

BIT-VIO:

Absolute Trajectory Error (ATE)
RMSE: 0.0034669
median: 0.027617





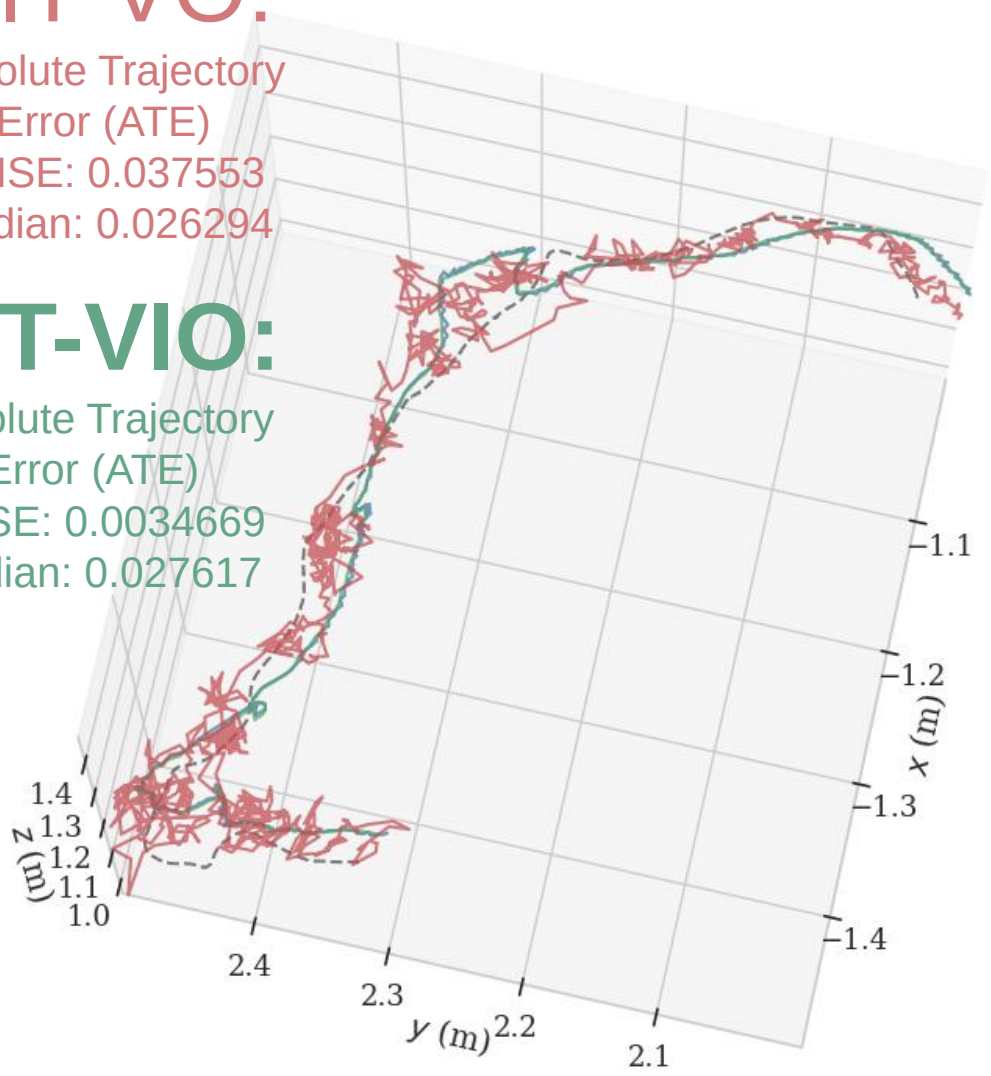
Comparison of the proposed BIT-VIO with inertial (IMU) and visual odometry (BIT-VO) overlaid on the reference ground-truth trajectory. Our BIT-VIO estimates are closer to the ground-truth compared to predictions from IMU or BIT-VO. Notice that our BIT-VIO effectively removes the high frequency noise visible in BIT-VO's trajectory.

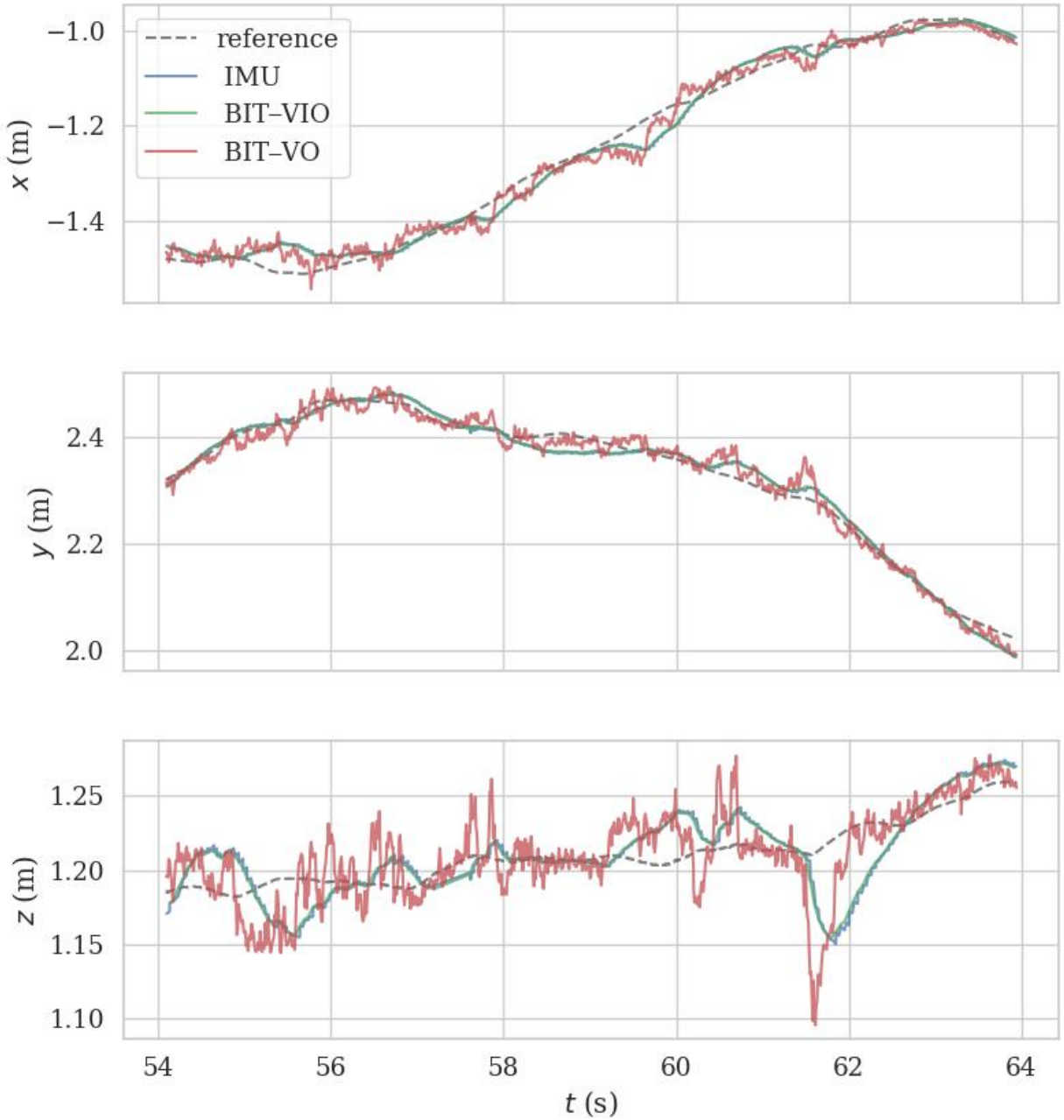
BIT-VO:

Absolute Trajectory Error (ATE)
RMSE: 0.037553
median: 0.026294

BIT-VIO:

Absolute Trajectory Error (ATE)
RMSE: 0.0034669
median: 0.027617





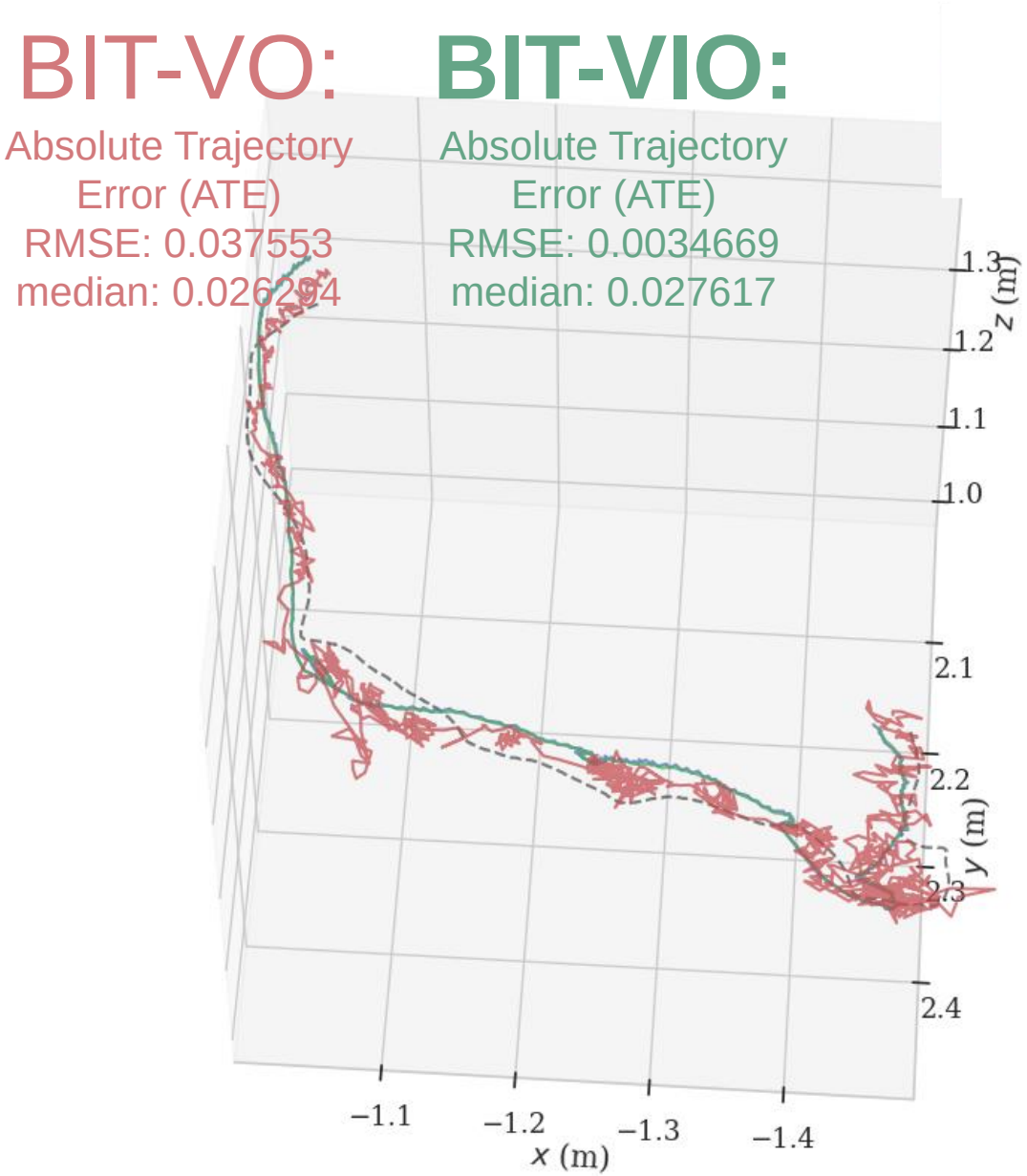
Comparison of the proposed BIT-VIO with inertial (IMU) and visual odometry (BIT-VO) overlaid on the reference ground-truth trajectory. Our BIT-VIO estimates are closer to the ground-truth compared to predictions from IMU or BIT-VO. Notice that our BIT-VIO effectively removes the high frequency noise visible in BIT-VO's trajectory.

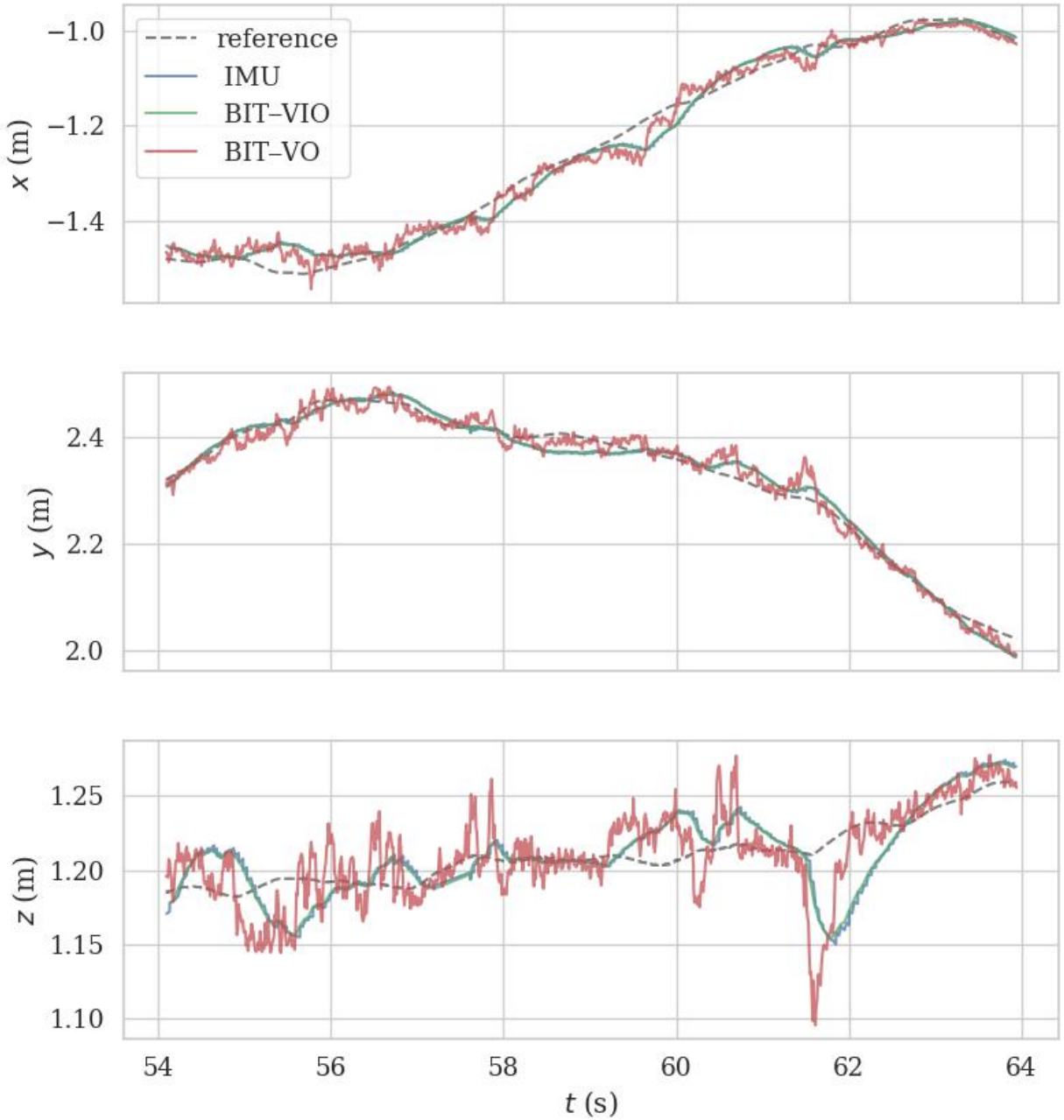
BIT-VO:

Absolute Trajectory Error (ATE)
RMSE: 0.037553
median: 0.026294

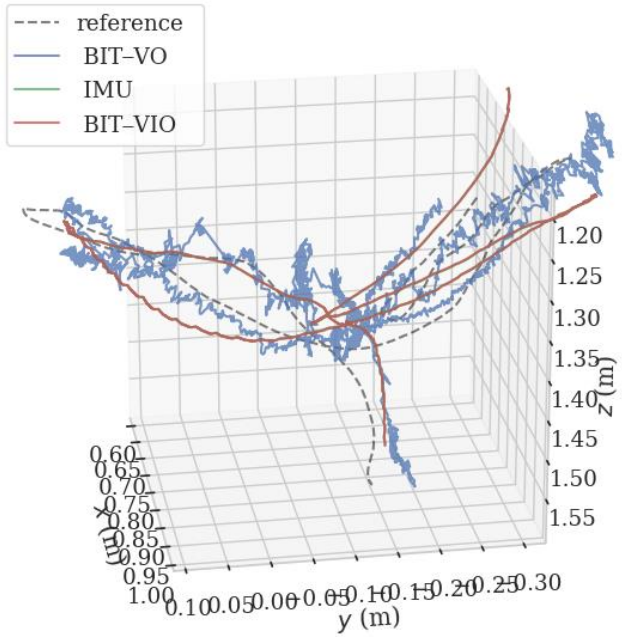
BIT-VIO:

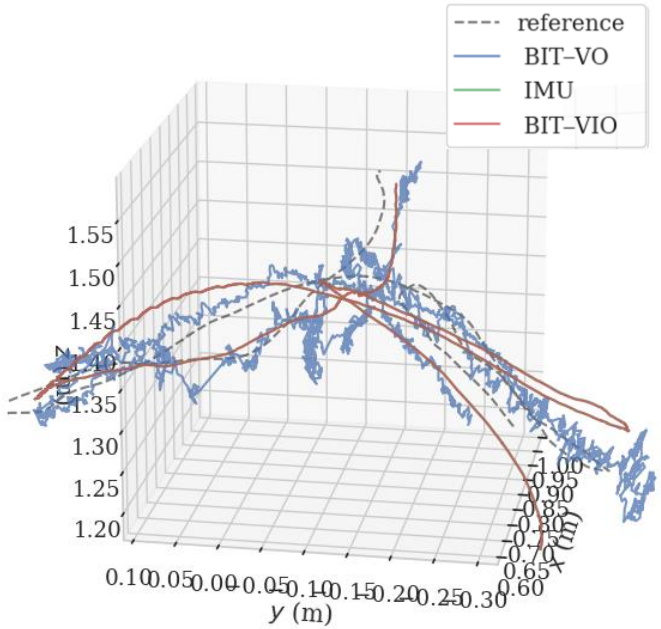
Absolute Trajectory Error (ATE)
RMSE: 0.0034669
median: 0.027617

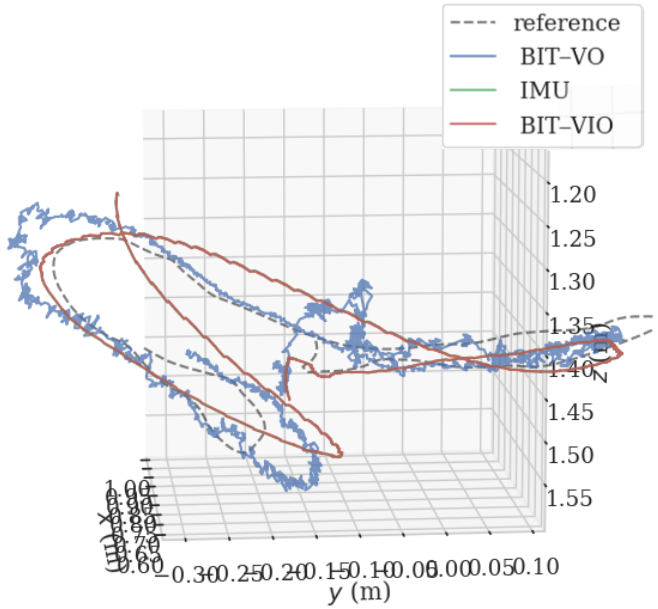




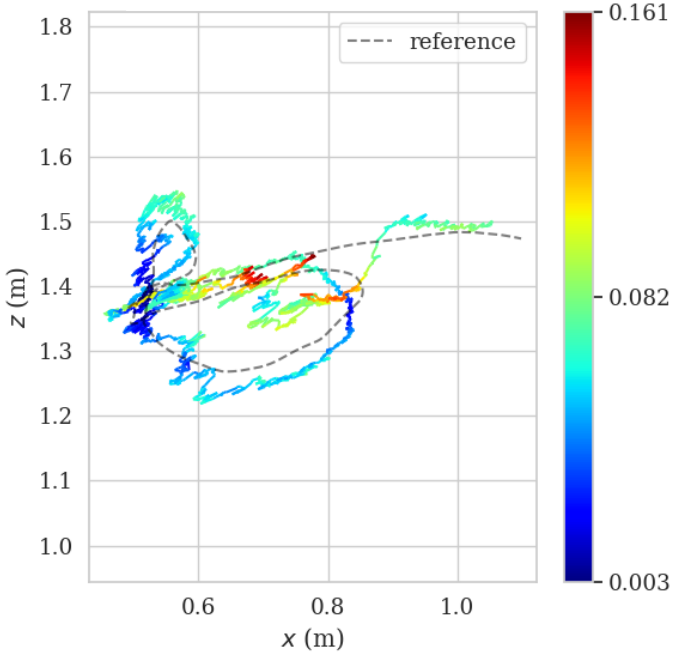
Comparison of the proposed BIT-VIO with inertial (IMU) and visual odometry (BIT-VO) overlaid on the reference ground-truth trajectory. Our BIT-VIO estimates are closer to the ground-truth compared to predictions from IMU or BIT-VO. Notice that our BIT-VIO effectively removes the high frequency noise visible in BIT-VO's trajectory.



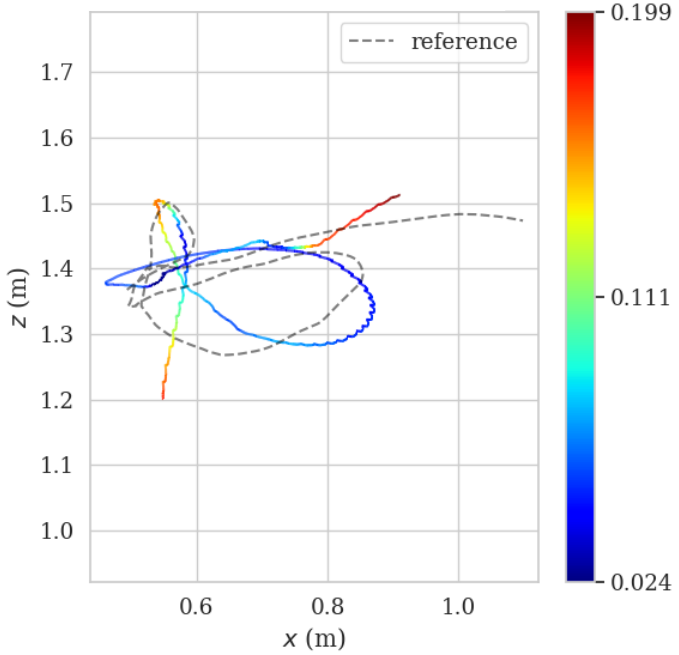




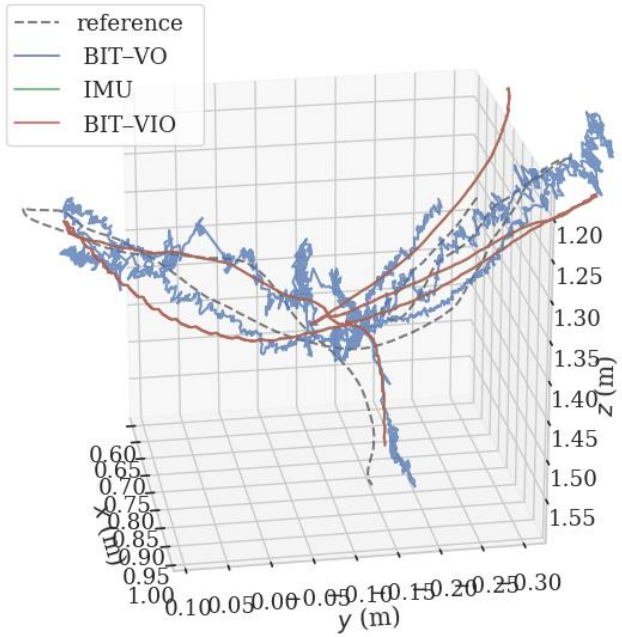
Error Mapped on BIT-VO

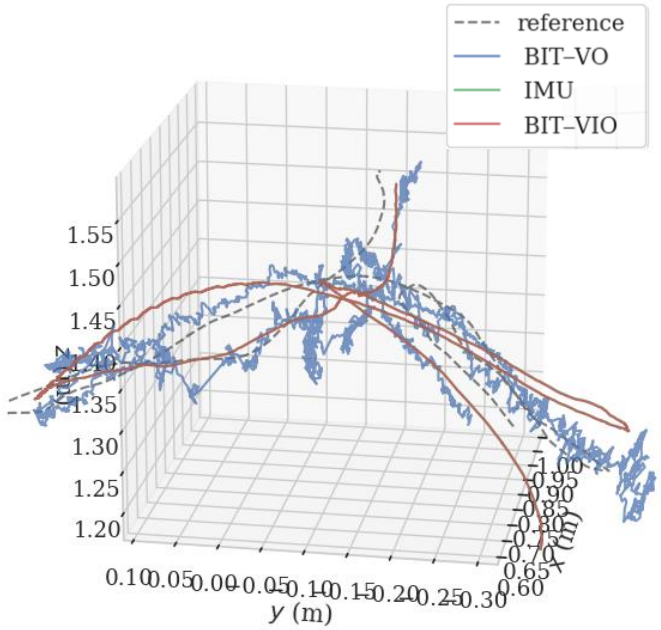


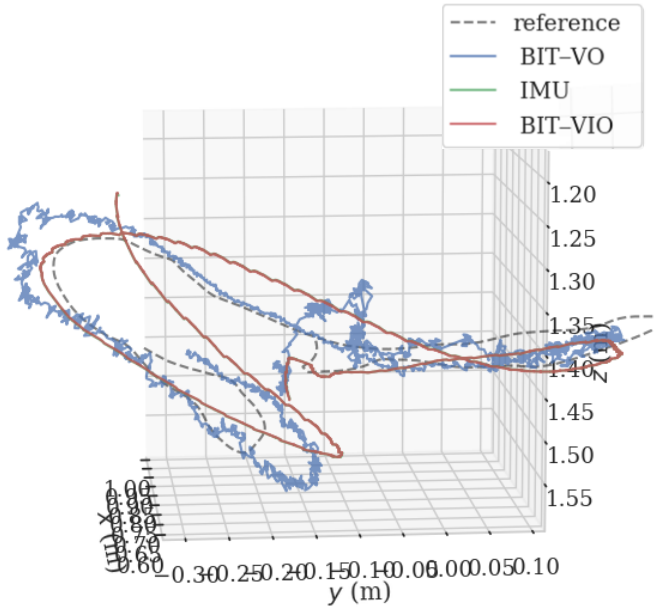
Error Mapped on BIT-VIO



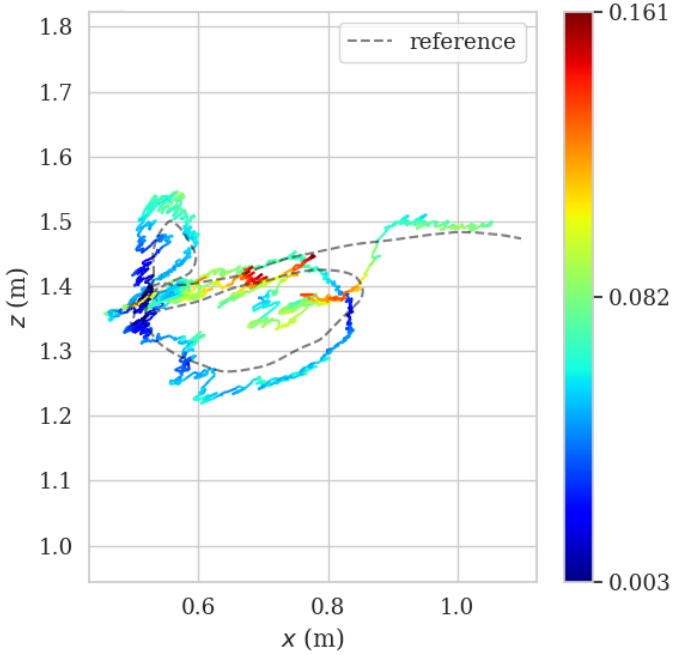
When projecting the error on BIT-VO (left) and BIT-VIO (right), we see that BIT-VO has high frequency noise with red tail-ends on its trajectory when compared to BIT-VIO's stabler closer to ground-truth trajectory with little regions of large ATE error (in red).



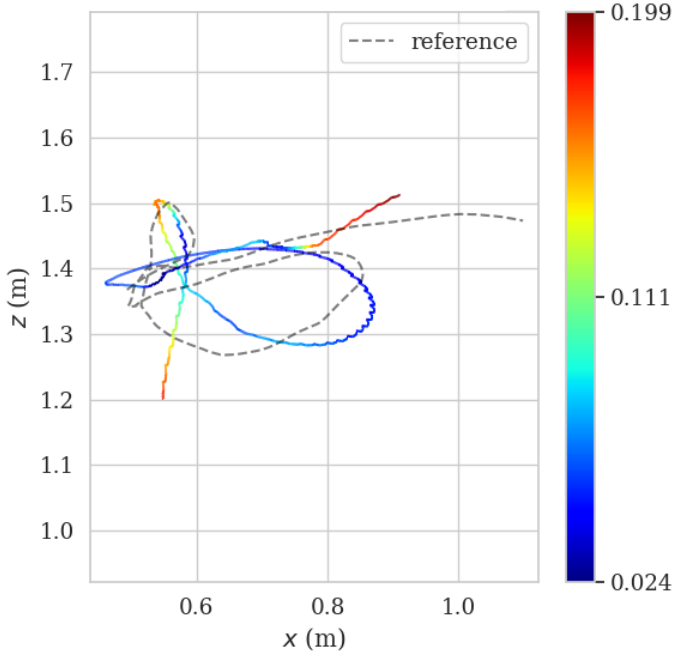




Error Mapped on BIT-VO



Error Mapped on BIT-VIO

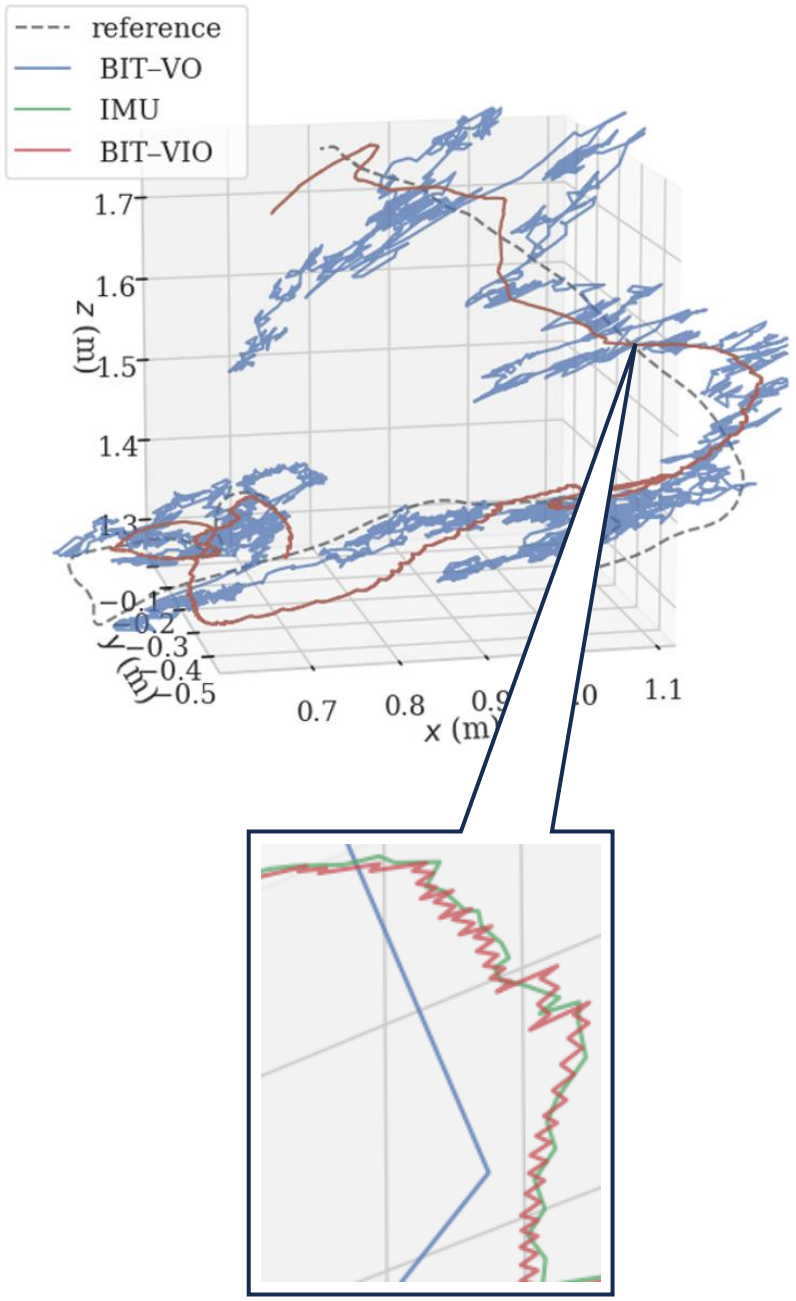


TRAJ. K

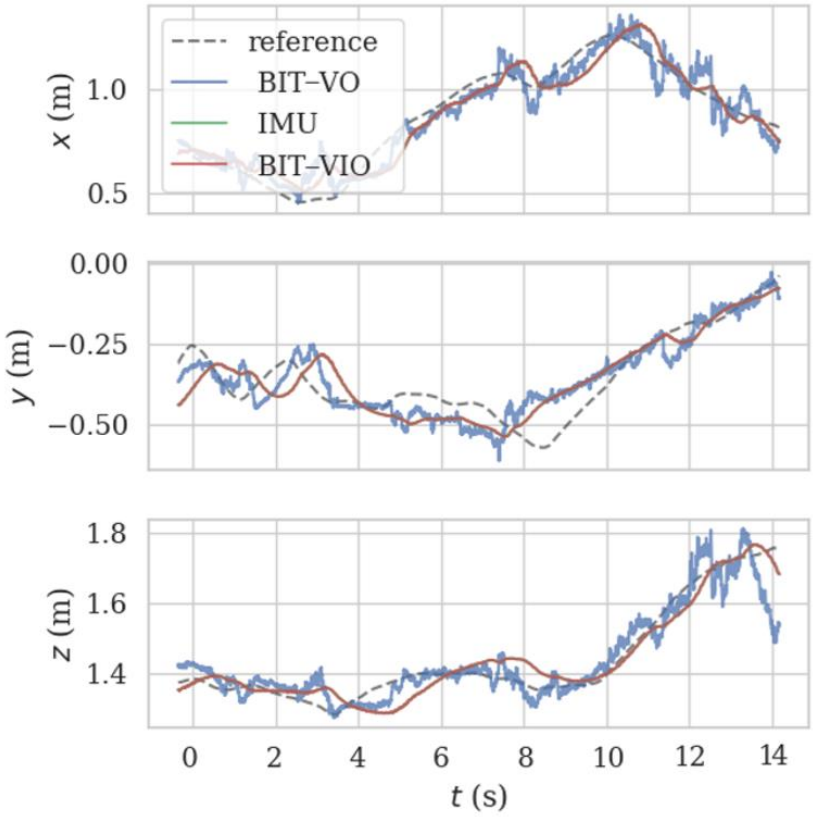
Doing fast, hostile motions, BIT-VO still has such high frequency noise.

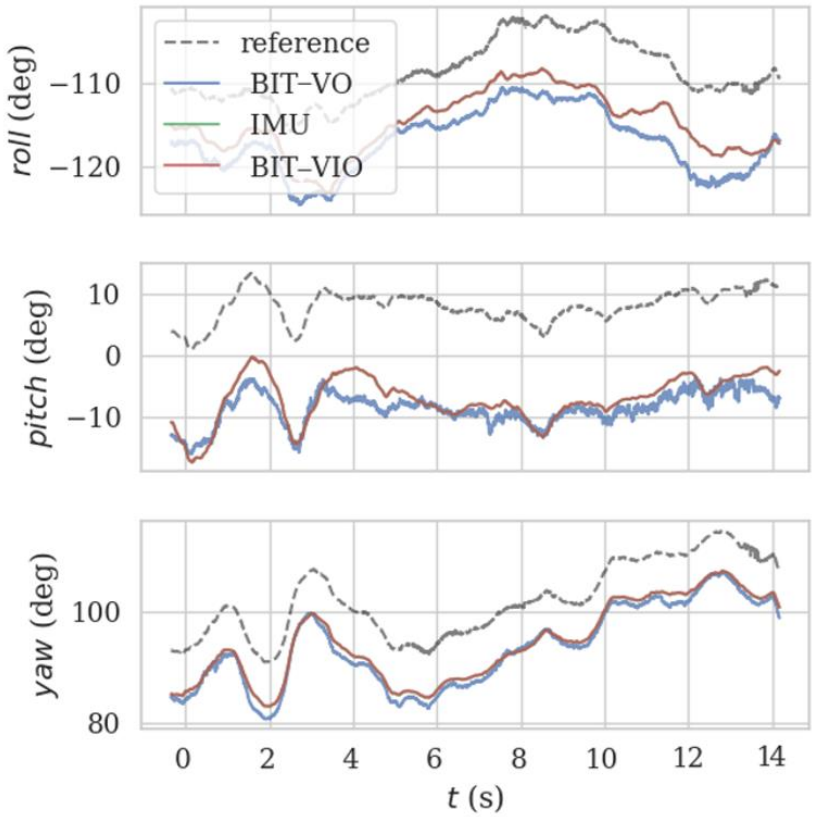
Our BIT-VIO algorithm can effectively remove this while having lower ATE error.

APE with IMU at 400 Hz, BIT-VO at 300 FPS					
Traj.	Type	BIT-VO APE (m)	IMU APE (m)	BIT-VIO APE (m)	Length (m)
K	RMSE:	0.075214	0.092716	**0.092355**	2.48
	median:	0.067571	0.062949	**0.062812**	

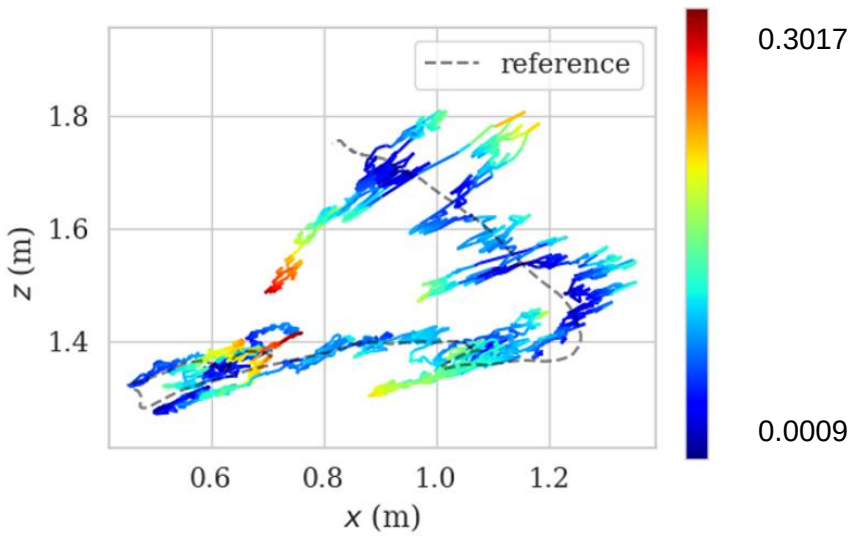


IMU as prediction, BIT-VO as update, BIT-VIO correcting via iEKF

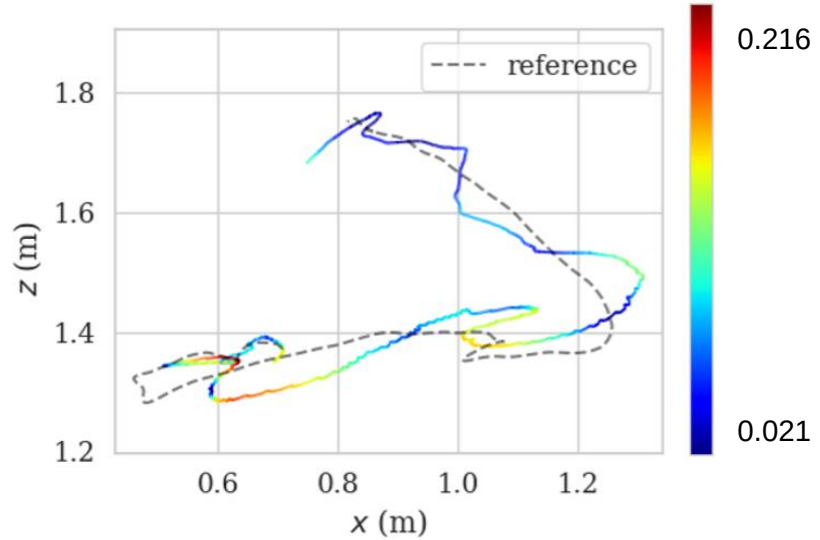




Error Mapped on BIT-VO



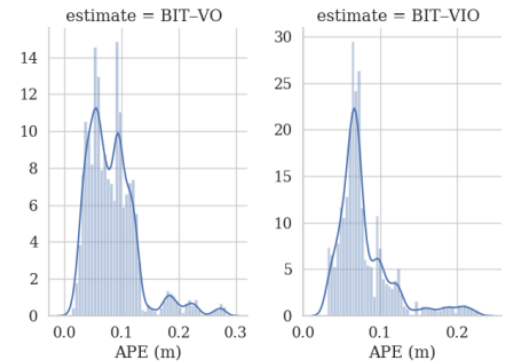
Error Mapped on BIT-VIO

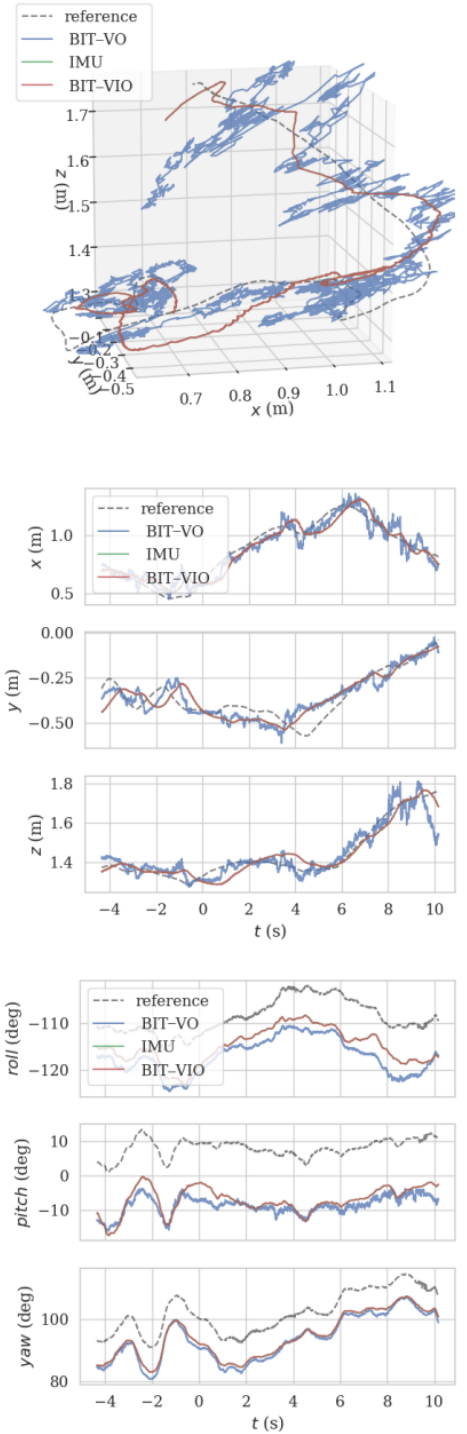


The Absolute Trajectory Error (ATE or APE) of our BITVIO <<< of BIT-VO.

TABLE I
IMU AT 400 HZ AND BIT-VO AT 300 FPS. BIT-VIO HAS LOWER APE.

APE with IMU at 400 Hz, BIT-VO at 300 FPS					
Traj.	Type	BIT-VO APE (m)	IMU APE (m)	BIT-VIO APE (m)	Length (m)
A	RMSE:	0.156247	0.156138	**0.156138**	3.6
	median:	0.138958	0.138315	**0.133202**	
B	RMSE:	**0.144674**	0.151913	0.150545	3.42
	median:	**0.096717**	0.106351	0.09959	
C	RMSE:	0.134617	0.121777	**0.12071**	2.6
	median:	0.119079	0.112582	**0.111856**	
D	RMSE:	0.05454	0.06597	**0.063498**	1.1
	median:	0.050929	0.053554	**0.049605**	
E	RMSE:	0.094479	0.08533	**0.086911**	1.7
	median:	0.07561	0.067952	**0.068756**	
F	RMSE:	0.175323	0.154923	**0.153335**	2.12
	median:	0.174444	0.142438	**0.140952**	
G	RMSE:	0.206866	0.196163	**0.195263**	2.4
	median:	0.15714	0.149788	**0.149103**	
H	RMSE:	0.088185	0.090189	**0.090376**	3.3
	median:	0.070431	0.07014	**0.070212**	
I	RMSE:	**0.134361**	0.13838	0.134328	2.01
	median:	**0.116587**	0.12805	0.124618	
J	RMSE:	0.057402	0.058275	**0.057461**	1.6
	median:	0.044689	0.044147	**0.042597**	
K	RMSE:	0.075214	0.092716	**0.092355**	2.48
	median:	0.067571	0.062949	**0.062812**	
L	RMSE:	0.10864	0.104469	**0.104366**	2.55
	median:	0.089689	0.088151	**0.08788**	
M	RMSE:	**0.094062**	0.108688	0.109156	2.9
	median:	**0.075766**	0.093151	0.09367	
N	RMSE:	0.101385	0.107546	**0.105227**	2.71
	median:	0.09016	0.089743	**0.087536**	





V. Conclusion and Next Steps

We have presented BIT-VIO, the first-ever 6-Degrees of Freedom (6-DOF) Visual Inertial Odometry (VIO) algorithm, which utilizes the advantages of the FPSP for vision-IMU-fused state estimation.

Our BIT-VIO algorithm operates and corrects by loosely-coupled sensor-fusion iterated Extended Kalman Filter (iEKF) at 300 FPS with an IMU at 400 Hz.

We evaluate BIT-VIO qualitatively against BIT-VO and demonstrate improvements in ATE across many trajectories. Moreover, the high-frequency noise evident in BIT-VO is effectively filtered out, resulting in a smoother estimated trajectory.

We plan to take the next steps toward a tightly-coupled VIO approach with the FPSP. Feature extraction would be integrated directly with IMU instead of two separate pose estimates, making the trajectories more robust and loop-closure better achievable.

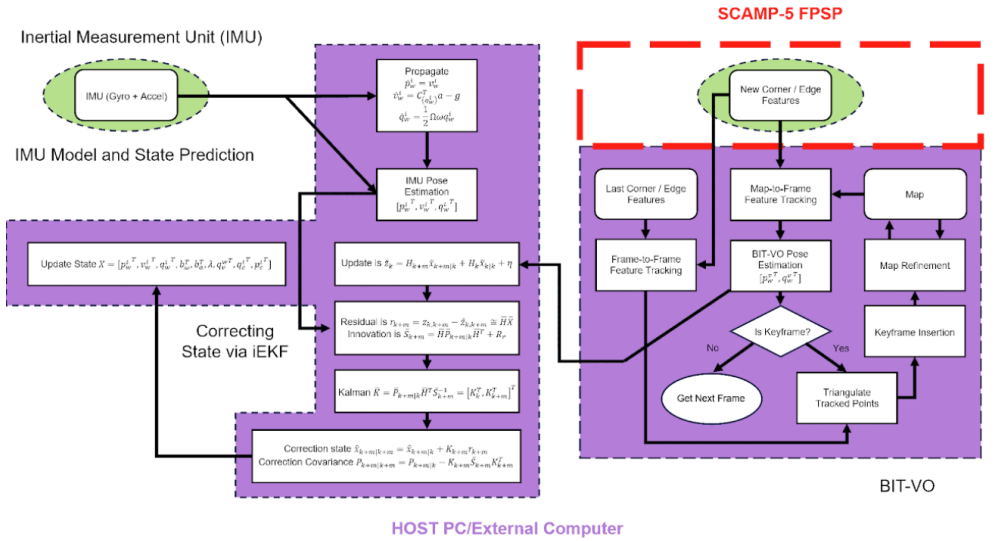


FIGURE 1: Pipeline of BIT-VIO. Multi-Sensor Fusion (iEKF-MSF) is to the left. BIT-VO is to the right. From the BIT-VO algorithm, the vision sensor utilizes the FPSP, highlighted in red. New Corner/Edge Features are detected via the FPSP, off-putting computational load by allowing some image and signal processing to be done on the chip before transferring to a PC host or other external device to be further processed. i.e., some of the framework is alleviated on FPSP.