Schmidt-EKF-based Visual-Inertial Moving Object Tracking

Kevin Eckenhoff, Patrick Geneva, Nathaniel Merrill, and Guoquan Huang
University of Delaware

ICRA 2020

Introduction

- Tracking the motion of a sensor platform moving in an unknown environment is a well-studied problem (Simultaneous Localization and Mapping)

- Being able to additionally perceive **moving objects** is critical in many applications such as obstacle avoidance and autonomous surveillance

- Goal: using cameras and an IMU, simultaneously track the **ego-motion** of the platform and the **3D pose** of a moving rigid body

VIO/Target Tracking

- Standard visual-inertial odometry: estimate the orientation, position, velocity, gyro bias, and accelerometer bias of the IMU

$$\mathbf{x}_I = \begin{bmatrix} {}^I_G\mathbf{q}^\top & {}^G\mathbf{p}_I^\top & {}^G\mathbf{v}_I^\top & \mathbf{b}_g^\top & \mathbf{b}_a^\top \end{bmatrix}^\top$$

- Target states: orientation, position, and a set of **motion parameters**

$$\mathbf{T} = \begin{bmatrix} {}^T_G\bar{q}^\top & {}^G\mathbf{p}_T^\top & \eta^\top \end{bmatrix}^\top$$

- In total, estimate both states in a single Extended Kalman Filter

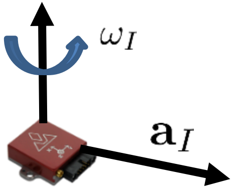
$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_I^\top & \mathbf{T}^\top \end{bmatrix}^\top$$

- Need to define how states evolve (**propagation**) and measurement function (**update**)

Propagation

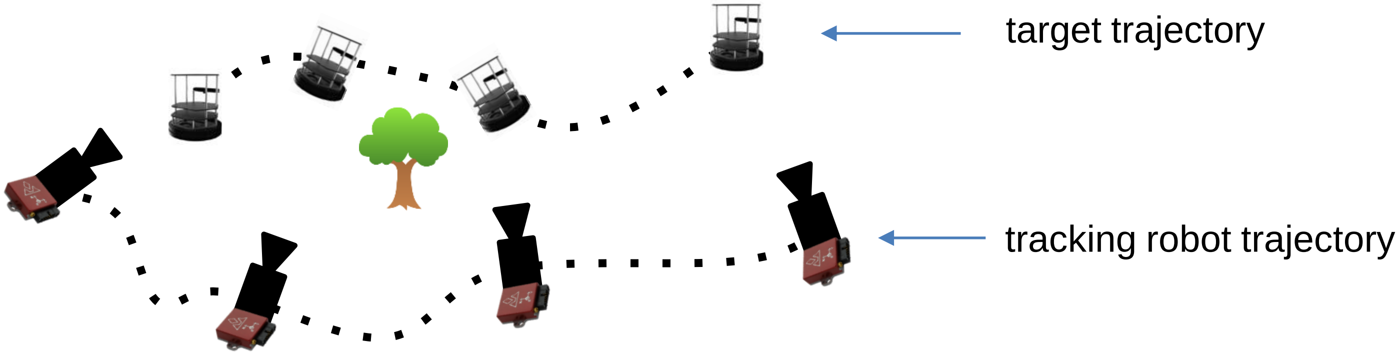
- Propagate IMU state using noisy gyroscope and accelerometer readings:

$$\mathbf{x}_{I_{k+1}} = \mathbf{f}\left(\mathbf{x}_{I_k}, \omega_{I_{k:k+1}}, \mathbf{a}_{I_{k:k+1}}\right)$$



- Target propagates as a function of its **current state** and **model noises**:

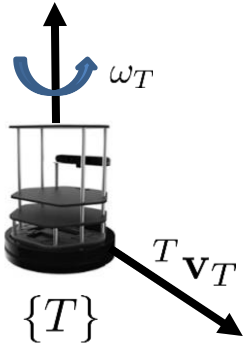
$$\mathbf{T}_{k+1} = \mathbf{g}\left(\mathbf{T}_k\right) + \mathbf{n}_T$$



Propagation

- Example motion model: constant local velocity

$$\eta = \begin{bmatrix} \omega_T^\top & {}^T\mathbf{v}_T^\top \end{bmatrix}^\top$$



- To handle errors in motion assumption, model these parameters as **random walks**

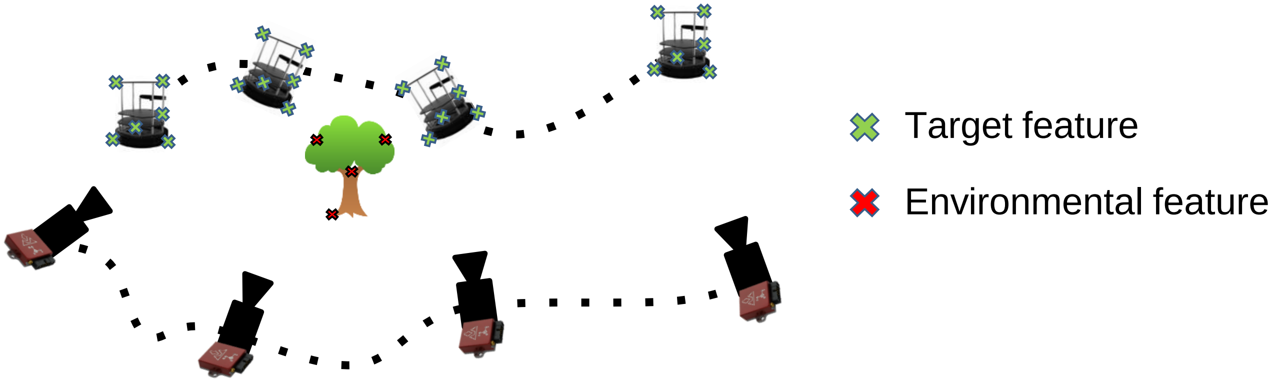
$${}^T\dot{\mathbf{v}}_T = \mathbf{n}_v, \quad \dot{\omega}_T = \mathbf{n}_w$$

$$\mathbf{n}_v \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_v) \quad \mathbf{n}_w \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_w)$$

Update

- Tracking platform’s cameras capture images of both the environment and the target of interest

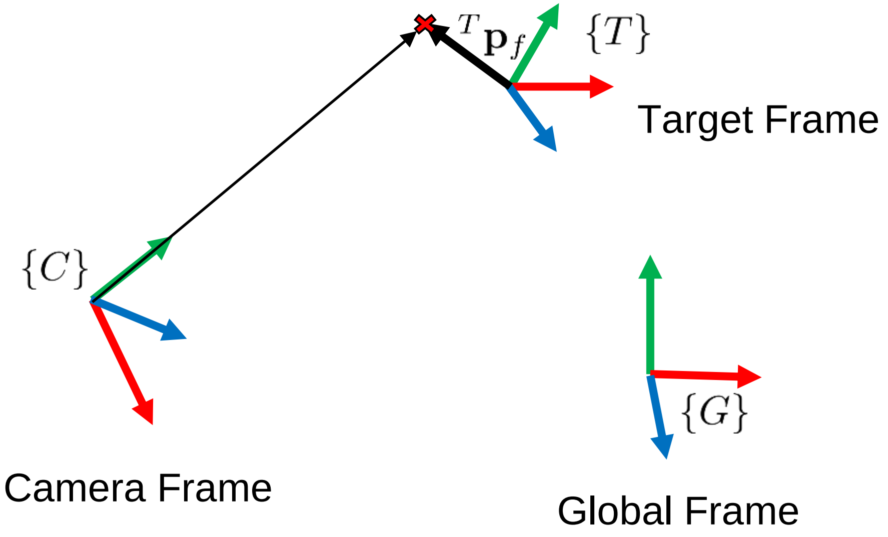
- Camera bearing measurements to environmental features are used in the standard manner



- Target measurements consist of 3D bearing measurements to points contained on the rigid body surface

Update

- Due to rigid body assumption, feature position in the target frame remains **fixed**
- Target measurements are a function of the **camera pose, target pose, and the target feature position**
- Positions for the target's pointcloud **added to the state** to be estimated



Noise Choice

- Assumption of current system: **accurate noise values** for motion model are available

$$ {}^{T}\dot{\mathbf{v}}_T = \mathbf{n}_v, \quad \dot{\omega}_T = \mathbf{n}_w $$

- Performed simulation where a stereo visual-inertial rig followed a planar moving target. Performed a parameter sweep over different **assumed motion noise levels**

- "Good" choice of noise parameters leads to **improved VIO performance** in EKF, poor choice leads to **decreased accuracy**

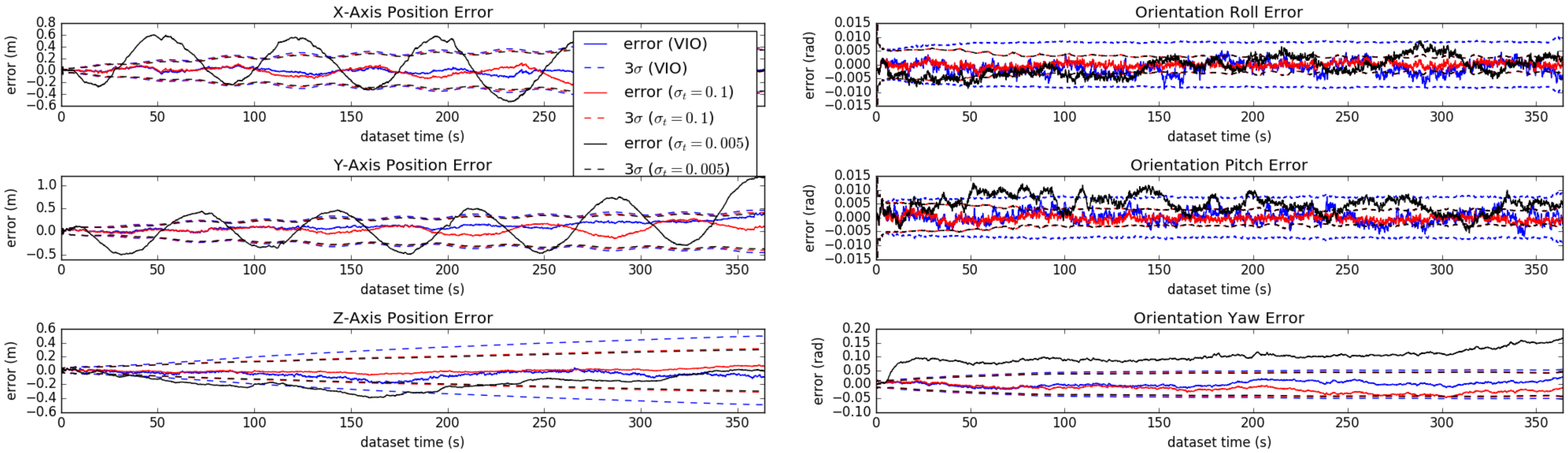
Absolute Trajectory Error (m/deg)

σ_t	Target	IMU	
0.001	1.883 / 8.944	1.839 / 8.687	Severely Degraded
0.005	0.259 / 1.547	0.254 / 1.566	
0.010	0.220 / 1.322	0.217 / 1.388	
0.050	0.203 / 1.313	0.199 / 1.341	Improved
0.100	0.197 / 1.291	0.194 / 1.293	
0.500	0.235 / 1.370	0.232 / 1.378	
VIO	× / ×	0.231 / 1.397	Standard VIO

Coupled Estimation

- Tightly-coupled VIO and target tracking (red) can lead to improved performance given correct target motion models over standard VIO (blue)

- **Overconfident** motion models cause estimator to become **inconsistent** (black)



Coupled Estimation

University of Delaware

- **Schmidt Kalman Filter** (SKF): Does not update part of state while **consistently tracking correlations**

- Using SKF to not update standard navigation states during target measurements leads to robust VIO

R: robot states

T: target states

EKF

$$\begin{bmatrix} \hat{\mathbf{x}}_T^+ \\ \hat{\mathbf{x}}_R^+ \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{x}}_T^- \\ \hat{\mathbf{x}}_R^- \end{bmatrix} + \begin{bmatrix} \mathbf{L}_T \\ \boxed{\mathbf{L}_R} \end{bmatrix} \mathbf{S}^{-1}\mathbf{r}$$

$$\begin{bmatrix} \mathbf{P}_{TT}^+ & \mathbf{P}_{TR}^+ \\ \mathbf{P}_{RT}^+ & \mathbf{P}_{RR}^+ \end{bmatrix} = \begin{bmatrix} \mathbf{P}_{TT}^- & \mathbf{P}_{TR}^- \\ \mathbf{P}_{RT}^- & \mathbf{P}_{RR}^- \end{bmatrix} - \begin{bmatrix} \mathbf{L}_T\mathbf{S}^{-1}\mathbf{L}_T^\top & \mathbf{L}_T\mathbf{S}^{-1}\mathbf{L}_R^\top \\ \mathbf{L}_R\mathbf{S}^{-1}\mathbf{L}_T^\top & \boxed{\mathbf{L}_R\mathbf{S}^{-1}\mathbf{L}_R^\top} \end{bmatrix}$$

SKF

$$\begin{bmatrix} \hat{\mathbf{x}}_T^+ \\ \hat{\mathbf{x}}_R^+ \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{x}}_T^- \\ \hat{\mathbf{x}}_R^- \end{bmatrix} + \begin{bmatrix} \mathbf{L}_T \\ \boxed{\mathbf{0}} \end{bmatrix} \mathbf{S}^{-1}\mathbf{r}$$

$$\begin{bmatrix} \mathbf{P}_{TT}^+ & \mathbf{P}_{TR}^+ \\ \mathbf{P}_{RT}^+ & \mathbf{P}_{RR}^+ \end{bmatrix} = \begin{bmatrix} \mathbf{P}_{TT}^- & \mathbf{P}_{TR}^- \\ \mathbf{P}_{RT}^- & \mathbf{P}_{RR}^- \end{bmatrix} - \begin{bmatrix} \mathbf{L}_T\mathbf{S}^{-1}\mathbf{L}_T^\top & \mathbf{L}_T\mathbf{S}^{-1}\mathbf{L}_R^\top \\ \mathbf{L}_R\mathbf{S}^{-1}\mathbf{L}_T^\top & \boxed{\mathbf{0}} \end{bmatrix}$$

Coupled Estimation

- Schmidt Kalman Filter (SKF): Does not update part of state while **consistently tracking correlations**

- Using SKF to not update standard navigation states during target measurements leads to robust VIO

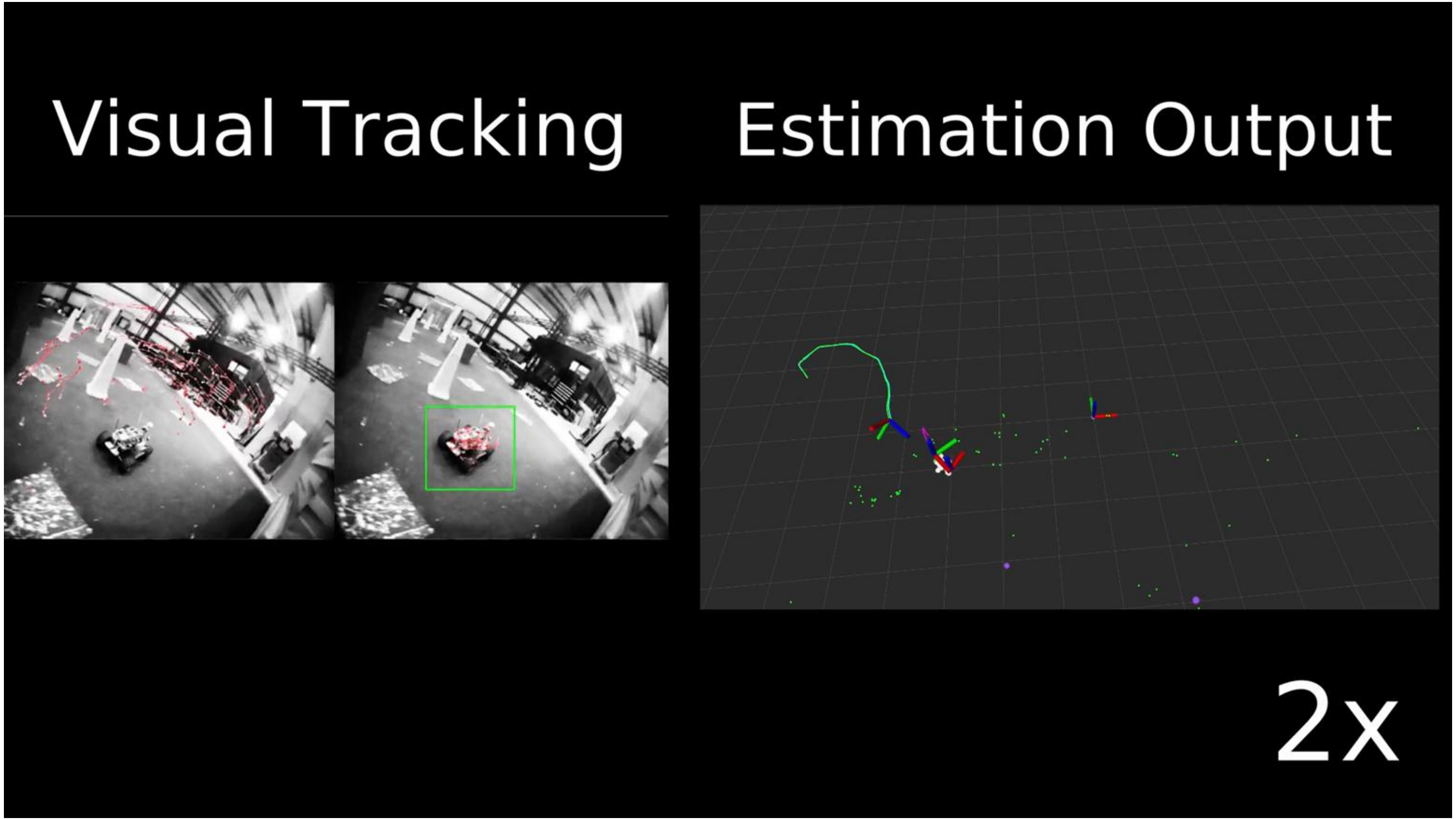
Absolute Trajectory Error (m/deg)

σ_t	Target	IMU
0.001	0.234 / 1.720	0.231 / 1.397
0.005	0.234 / 1.524	0.231 / 1.397
0.010	0.234 / 1.511	0.231 / 1.397
0.050	0.234 / 1.399	0.231 / 1.397
0.100	0.234 / 1.413	0.231 / 1.397
0.500	0.234 / 1.617	0.231 / 1.397
VIO	$\times$ / $\times$	0.231 / 1.397

Target Detection

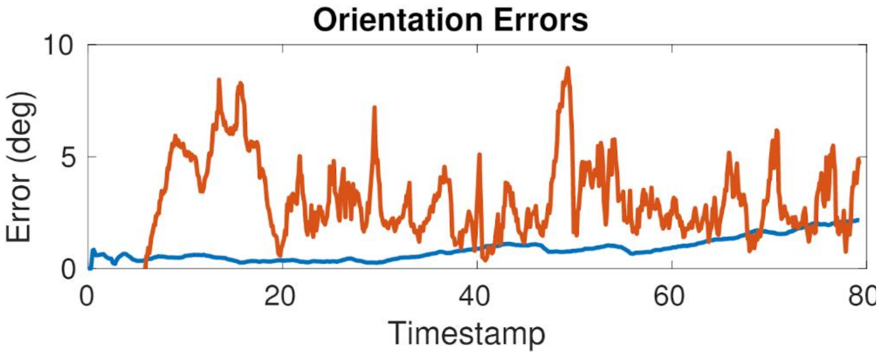
- Based on U-Net architecture, mask used to classify features **on/off the moving target**
- 279 images were hand labeled for training and validation (90/10 split)

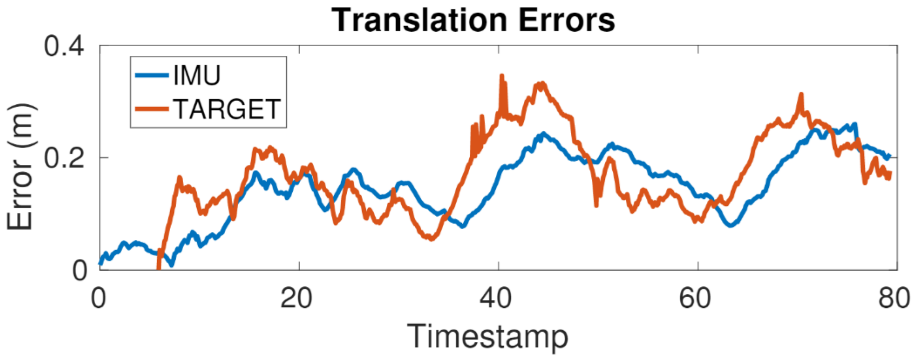
Results



Results

- Proposed Schmidt-EKF-based method **outperforms tightly-coupled method** in presence of overconfident target model:
  - Schmidt-EKF RMSE: **0.153m / 1.091° IMU, 0.183m / 3.443° Target**
  - Tightly-coupled EKF RMSE: **0.409m / 1.640° IMU, 0.492m / 3.037° Target**





Conclusion

- Investigated the effect of tight-coupling of visual-inertial ego-motion and target tracking performance
- Showed that while a proper noise model for the target led to **improved** localization performance, overconfident model selection led to **severely degraded** estimation
- Utilized **Schmidt Kalman Filtering** to prevent target measurements from updating the IMU state, while still conservatively modeling all correlations
- Proposed solution validated in **real-world** visual-inertial moving object experiments