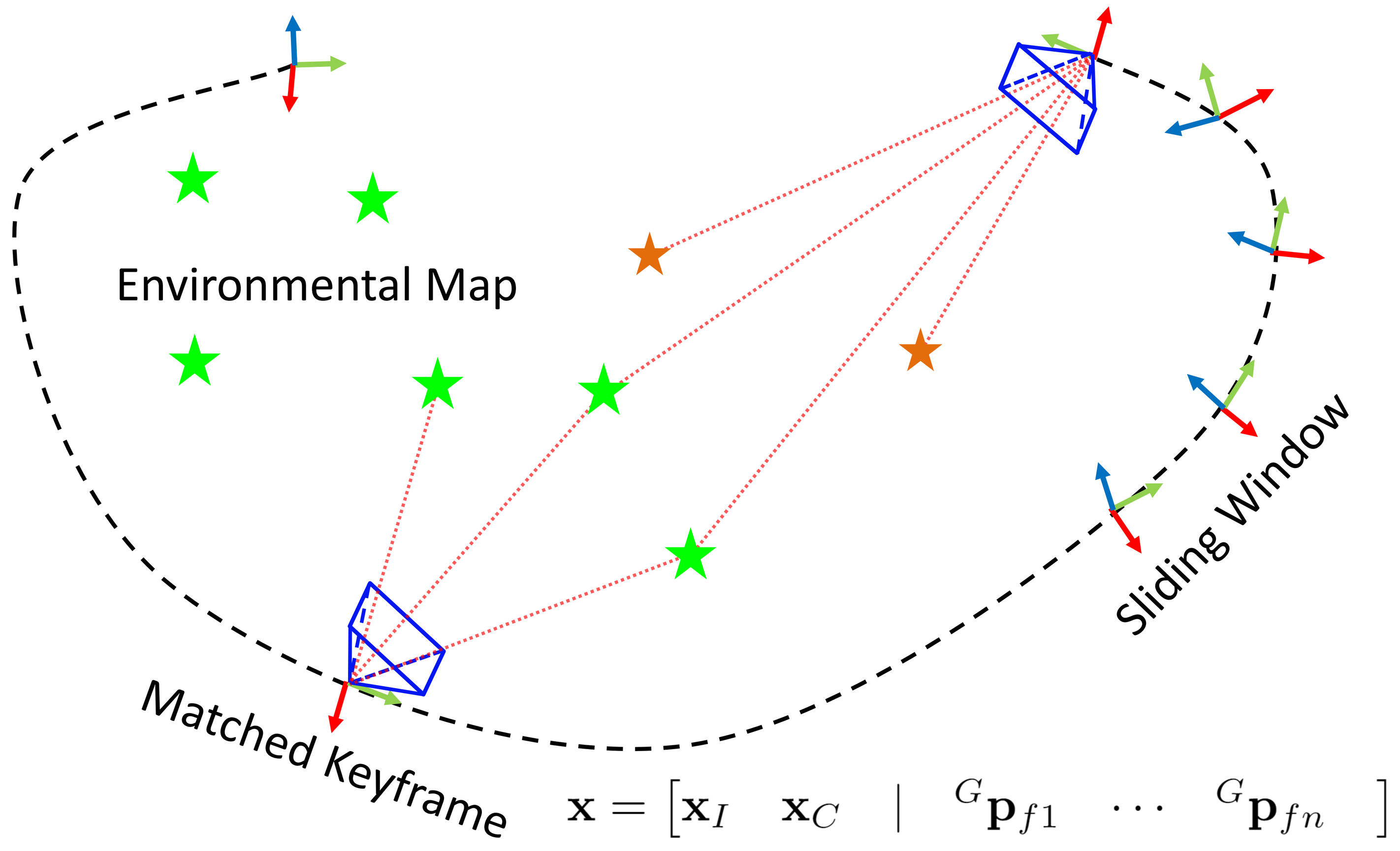
An Efficient Schmidt-EKF for 3D Visual-Inertial SLAM

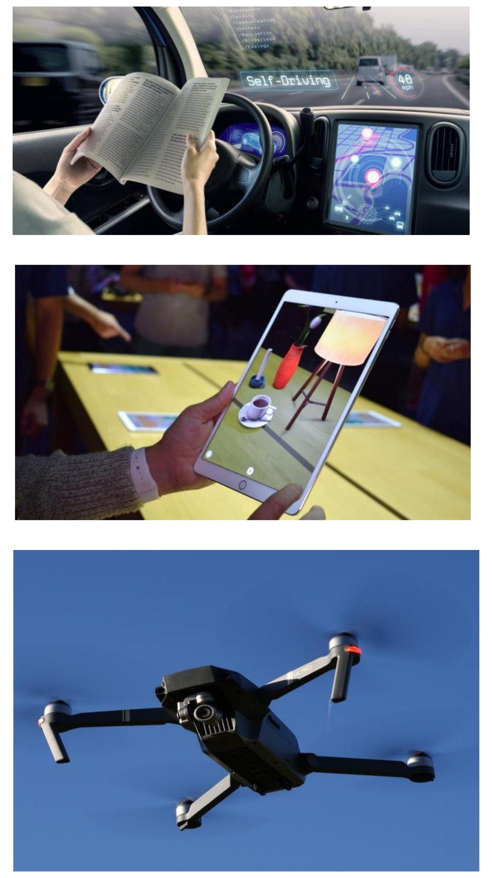
Patrick Geneva, James Maley, and Guoquan (Paul) Huang
Robot Perception and Navigation Group (RPNG), University of Delaware (UD)

CVPR LONG BEACH CALIFORNIA June 16-20, 2019

Introduction

- **Visual-inertial navigation** aims to estimate 3D motion of the sensor platform using an IMU and camera, which is critical for many technologies.





- **Contributions**
  - Adapt Schmidt-KF to achieve **linear complexity** of Visual-Inertial SLAM
  - Leverage **2D-to-2D matching** of historical keyframes to find map matches
  - Perform extensive validations to show the proposed Schmidt-EKF VI-SLAM outperforms the state-of-the-art visual-inertial systems

Schmidt-EKF

- Key idea: Don't refine mature SLAM features. Treat as nuisance parameters and update *only* the active states and their cross-correlations
- Method: Set the Kalman gain of "nuisance" parameters to zero

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

$$\begin{bmatrix}\mathbf{P}_{AA}^+ & \mathbf{P}_{AS}^+\\ \mathbf{P}_{SA}^+ & \mathbf{P}_{SS}^+\end{bmatrix} = \begin{bmatrix}\mathbf{P}_{AA}^- & \mathbf{P}_{AS}^-\\ \mathbf{P}_{SA}^- & \mathbf{P}_{SS}^-\end{bmatrix} - \begin{bmatrix}\mathbf{L}_A\mathbf{S}^{-1}\mathbf{L}_A^\top & \mathbf{L}_A\mathbf{S}^{-1}\mathbf{L}_S^\top\\ \mathbf{L}_S\mathbf{S}^{-1}\mathbf{L}_A^\top & \boxed{\mathbf{L}_S\mathbf{S}^{-1}\mathbf{L}_S^\top}\end{bmatrix}$$

Standard EKF

vs

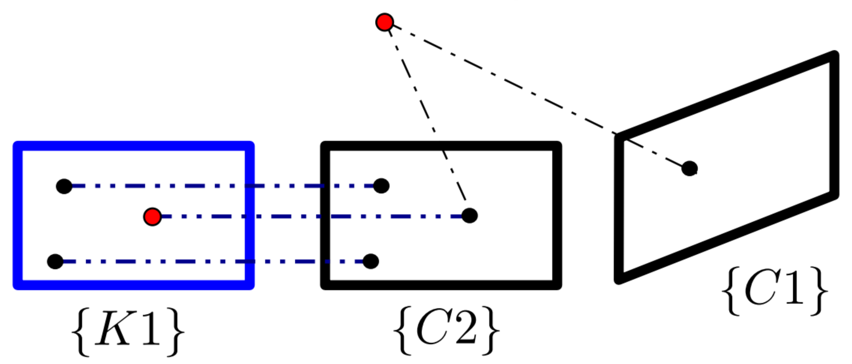
$$\begin{bmatrix}\hat{\mathbf{x}}_A^+\\ \hat{\mathbf{x}}_S^+\end{bmatrix} = \begin{bmatrix}\hat{\mathbf{x}}_A^-\\ \hat{\mathbf{x}}_S^-\end{bmatrix} + \begin{bmatrix}\mathbf{L}_A\\ \boxed{\mathbf{0}}\end{bmatrix}\mathbf{S}^{-1}\mathbf{r}$$

$$\begin{bmatrix}\mathbf{P}_{AA}^+ & \mathbf{P}_{AS}^+\\ \mathbf{P}_{SA}^+ & \mathbf{P}_{SS}^+\end{bmatrix} = \begin{bmatrix}\mathbf{P}_{AA}^- & \mathbf{P}_{AS}^-\\ \mathbf{P}_{SA}^- & \mathbf{P}_{SS}^-\end{bmatrix} - \begin{bmatrix}\mathbf{L}_A\mathbf{S}^{-1}\mathbf{L}_A^\top & \mathbf{L}_A\mathbf{S}^{-1}\mathbf{L}_S^\top\\ \mathbf{L}_S\mathbf{S}^{-1}\mathbf{L}_A^\top & \boxed{\mathbf{0}}\end{bmatrix}$$

Schmidt EKF

Schmidt-EKF Visual-Inertial SLAM (SEVIS) Algorithm

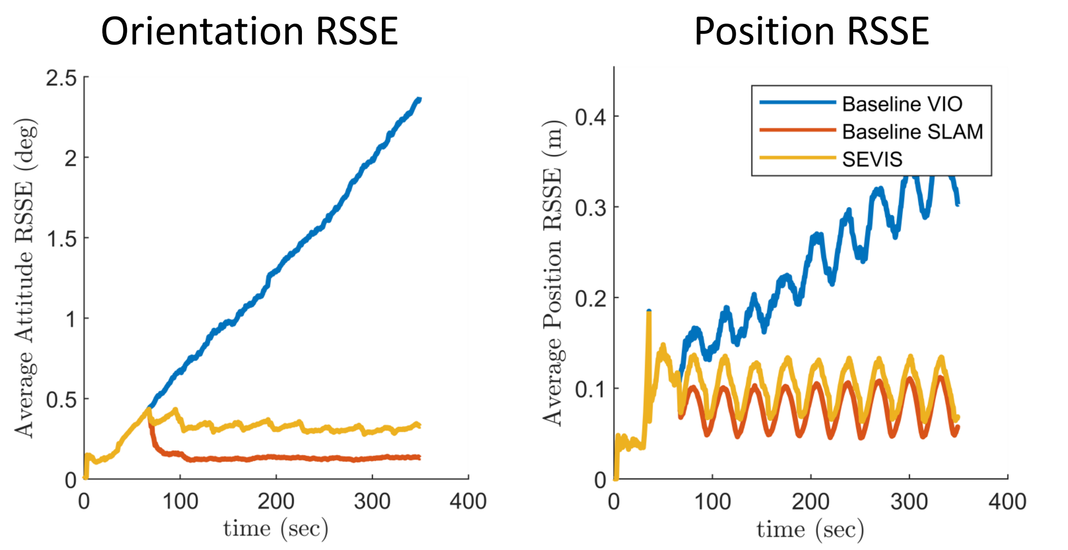
- **State Propagation**
  1. Propagate IMU navigation state
  2. Propagate the active covariance $\mathbf{P}_{AA}$ and cross-correlation $\mathbf{P}_{AS}$
- **State Update**
  1. Stochastically clone current IMU state
  2. Track visual features into newest frame
  3. Perform 2D-to-2D matching to find map feature correspondences
     - Query keyframe database, DBoW2, for matching keyframe
     - Match current tracks to keyframe
     - Associate active features with map features seen from keyframe

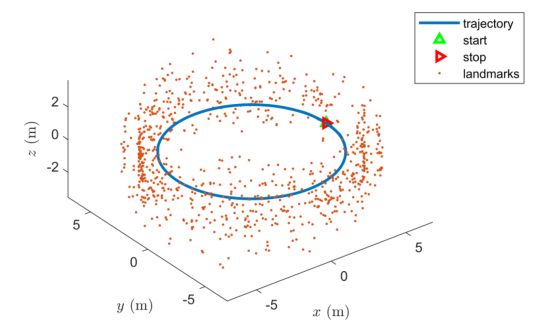


  4. Perform MSCKF feature update with non-map features
  5. Initialize new SLAM features if needed
  6. Schmidt-EKF for measurements that are functions of map features, update sliding window without updating map
- **Feature and Keyframe Management**
  1. Lost SLAM features are either moved to Schmidt or marginalized
  2. Marginalize the oldest clone from MSCKF sliding window
  3. Marginalization of map features bounds map size
  4. Insert new keyframe into database if the current frame sees map features
  5. Remove keyframes that had all of their seen features marginalized out

Monte-Carlo Simulation:

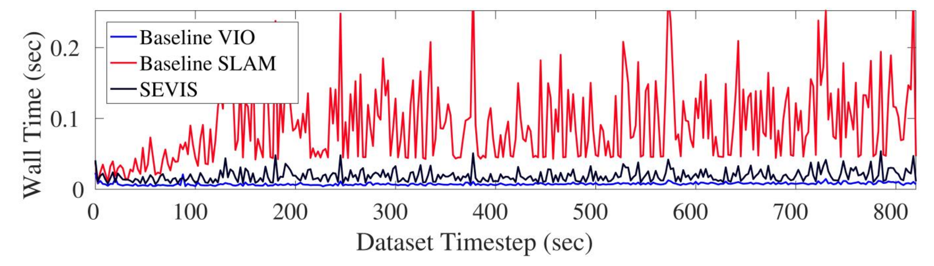
- Simulated realistic sensor values
- SEVIS is able to perform close to full EKF-SLAM filter with bounded accuracy
- *Baseline VIO* – MSCKF with no loop closure information
- *Baseline SLAM* – Full EKF-SLAM in which map features update



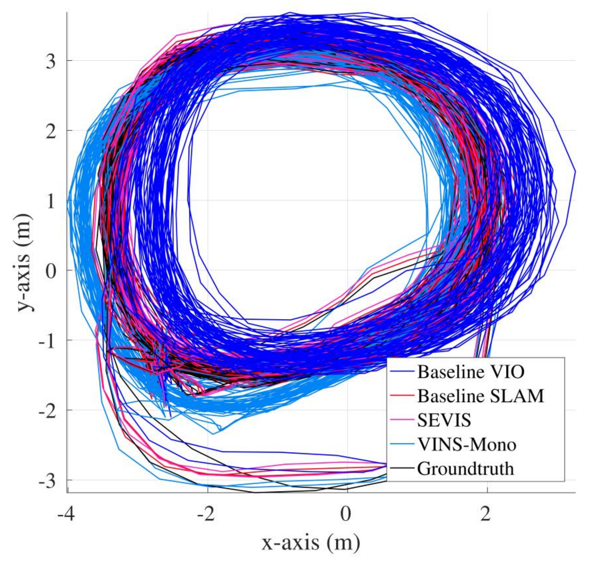


Vicon Loops Dataset (1.2km):

- Timing analysis shows SEVIS remains real-time while full EKF-SLAM filter becomes computationally infeasible, max of 600 features
- SEVIS outperforms VINS-Mono [Qin2018], with accuracy close to full EKF-SLAM

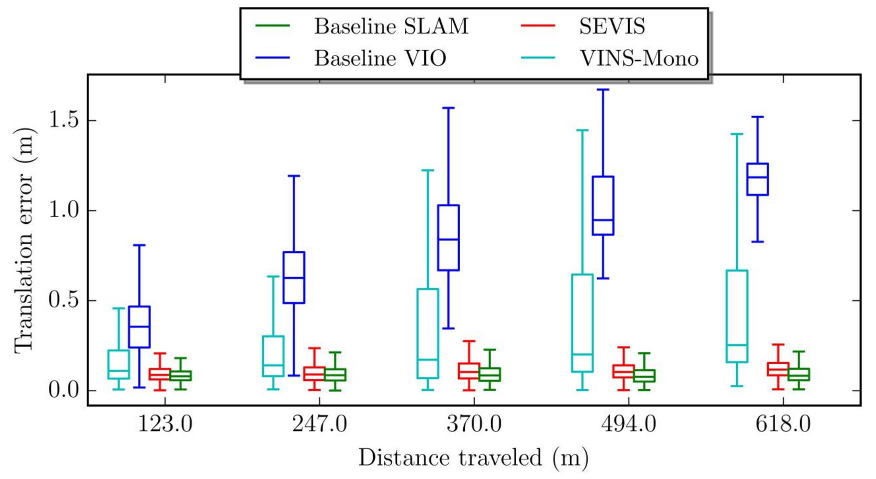


**Single thread on Intel Xeon E3-1505Mv6 @ 3.00GHz



Segment Length	Baseline VIO	Baseline SLAM	SEVIS	VINS-Mono
123m	0.383	0.102	0.111	0.184
247m	0.645	0.099	0.108	0.238
370m	0.874	0.104	0.123	0.325
494m	1.023	0.095	0.121	0.381
618m	1.173	0.107	0.139	0.425
ATE	0.779	0.121	0.128	0.323

Relative Pose Error (RPE) and Absolute Trajectory Error (ATE)



Nighttime Multi-Floor Dataset (1.5km):

- Can handle max of 700 map features at real-time
- Final start-end error was **4.67m (0.31%)** for the baseline VIO, with SEVIS having **0.37m (0.02%)**
- Limited drift over entire dataset with large accuracy gain over standard MSCKF