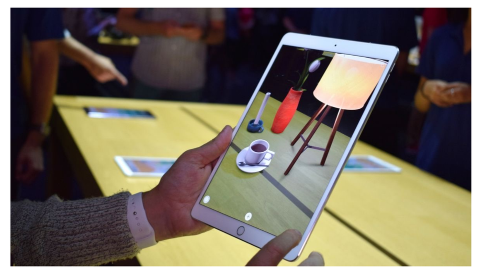
NRI: FND: Consistent Distributed Visual-Inertial Estimation and Perception for Cooperative Unmanned Aerial Vehicles

Guoquan (Paul) Huang / University of Delaware

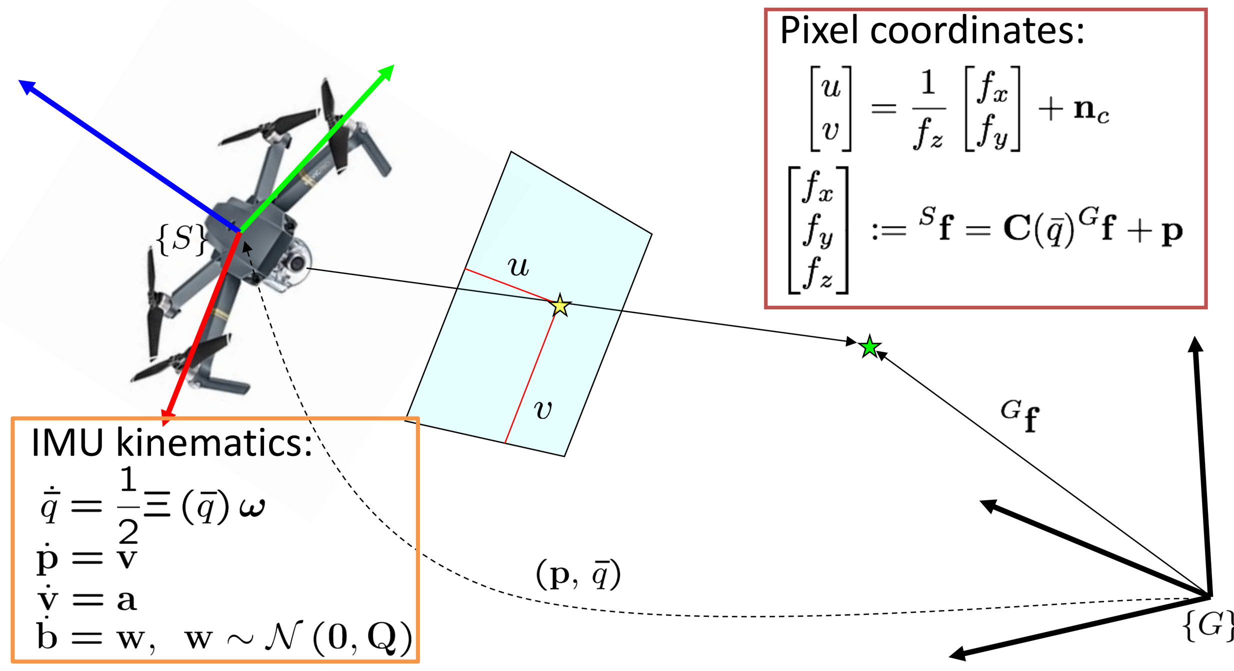
https://sites.udel.edu/robot/

1. Visual-inertial estimation

- Visual-inertial navigation system (VINS) or SLAM holds great potentials in practical applications:



- Goal: To estimate 3D motion & scene understanding using IMUs and cameras onboard unmanned aerial vehicles (UAVs)



Pixel coordinates:

$$\begin{bmatrix} u \\ v \end{bmatrix} = \frac{1}{f_z}\begin{bmatrix} f_x \\ f_y \end{bmatrix} + \mathbf{n}_c$$

$$\begin{bmatrix} f_x \\ f_y \\ f_z \end{bmatrix} := {}^{S}\mathbf{f} = \mathbf{C}(\bar{q})\,{}^{G}\mathbf{f} + \mathbf{p}$$

IMU kinematics:

$$\dot{\bar{q}} = \frac{1}{2}\Xi(\bar{q})\,\omega$$
$$\dot{\mathbf{p}} = \mathbf{v}$$
$$\dot{\mathbf{v}} = \mathbf{a}$$
$$\dot{\mathbf{b}} = \mathbf{w}, \quad \mathbf{w} \sim \mathcal{N}(0, \mathbf{Q})$$

- **Technical challenges when extending to multi-UAVs:**
  - ❑ Consistent, distributed, cooperative visual-inertial estimation under resource constraints

2. Distributed visual-inertial cooperative localization [IROS 2021]

- Multi-robot systems can **collaborate** to accomplish missions more efficiently and robustly
- Visual-Inertial Cooperative Localization (CL) leveraging **additional geometric constraints** in multi-robot systems can improve the localization accuracy of the whole group
- Distributed estimation is more **scalable, robust, and efficient** as compared to centralized estimators
- **Our approach**: An efficient distributed state estimator for multi-robot CL which does not require simultaneous viewing of the common features:
  - ❑ Consistent estimation w/ Covariance Intersection (CI)
  - ❑ Efficient common SLAM and VIO feature updates
  - ❑ Historical loop-closure matching across time and robots to reduced localization drift
- Results: https://www.youtube.com/watch?v=boHBcVoMKk8&t=62s

Table 1: Relative pose error (RPE) on TUM-VI datasets in degrees / meters averaged over all robots for the dataset.

Algorithm	40m	60m	80m	100m	120m
indp-slam	1.818 / 0.093	2.833 / 0.126	2.604 / 0.154	2.774 / 0.185	2.716 / 0.215
ce-cmsckf	**1.358** / 0.071	**1.321** / 0.091	1.357 / 0.108	0.843 / 0.128	0.932 / 0.140
ce-cmsckf-cslam	1.758 / **0.069**	1.350 /**0.079**	**1.027** / **0.100**	**0.718** / 0.119	0.938 / **0.130**
dc-cmsckf	1.662 / 0.075	2.005 / 0.104	1.605 / 0.129	1.142 / 0.141	1.531 / 0.170
dc-cmsckf-cslam	1.800 / 0.080	2.642 / 0.093	2.233 / 0.106	1.544 / **0.114**	0.934 / 0.157
dc-full-window	1.768 / 0.075	2.218 / 0.091	1.788 / 0.109	1.257 / 0.123	**0.854** / 0.159
dc-full-history	**1.213** / **0.067**	**1.232** / **0.061**	**1.029** / **0.065**	**1.004** / **0.068**	**0.784** / **0.072**

Inclusion of common features always improves both centralized and decentralized estimators.

Historical matching able to outperform all other methods (even the centralized)!

3. Versatile and resilient multi-IMU multi-camera (MIMC)-VINS [TRO 2022]

- **Goal:**
  - ❑ To design versatile and resilient MIMC-VINS that seamlessly fuses multi-modal information from an *arbitrary* number of uncalibrated cameras and IMUs, while providing smooth, uninterrupted, and accurate 3D motion tracking even if some sensors fail
- **Key ideas:**
  - ❑ To perform high-order on-manifold state interpolation to efficiently process all available visual measurements without increasing the computational burden due to estimating additional sensors' poses at asynchronous imaging times

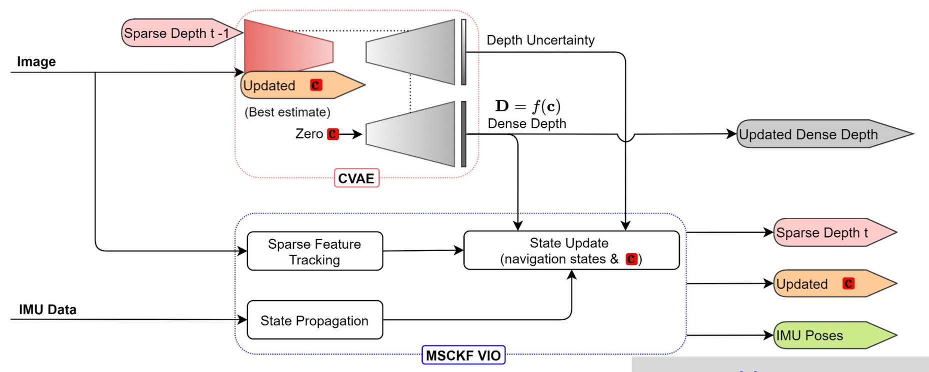
$$^{G}\mathbf{p}_{I(t)} \approx (1-\lambda)\,^{G}\mathbf{p}_{I(t_a)} + \lambda\,^{G}\mathbf{p}_{I(t_b)}$$
$$^{I(t)}_{G}\mathbf{R} \approx \mathrm{Exp}\left(\lambda \mathrm{Log}\left(^{I(t_b)}_{G}\mathbf{R}\,^{G}_{I(t_a)}\mathbf{R}\right)\right)\,^{I(t_a)}_{G}\mathbf{R}$$
$$\lambda = \frac{t - t_a}{t_b - t_a}$$

  - ❑ To propagate a joint system consisting of all IMU states while enforcing rigid-body constraints between the IMUs during the filter update stage
  - ❑ To estimate online both spatiotemporal extrinsic and visual intrinsic parameters to be robust to errors in prior sensor calibration

https://www.youtube.com/watch?v=my1ljdJ4irY&t=18s

4. Code-VIO [ICRA 2021] *(Best Paper Finalist in Robot Vision)*

- Real-time perception of dense mapping w/ monocular camera and IMU is essential for autonomous navigation.
- Estimate compact optimizable dense depth (code) in VIO and improve code pipeline for accurate and efficient performance.



https://www.youtube.com/watch?v=tr_vflFjghY

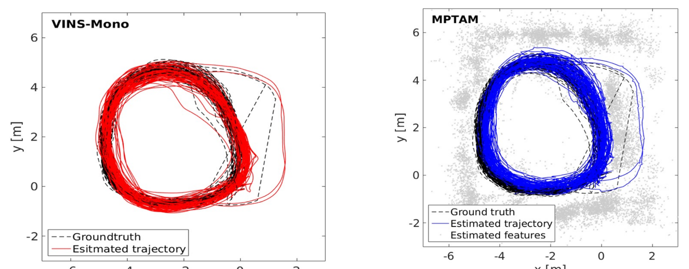
5. Markov parallel tracking and mapping [ICRA 2021]

- The proposed MPTAM estimator:
  - ❑ Retains the advantage of parallel processing.
  - ❑ Follows the Markov assumption of probabilistic SLAM problem.
  - ❑ Different sliding-window estimators can be directly used for realizing the front end.
  - ❑ The efficiency of the back end can be improved with practical approximation (optional).

TABLE I: Real-time pose accuracy (RMSE) and average computation time (of the front end) on EuRoC dataset.

	VINS-Mono					MPTAM				
	w/o loop closure		w/ loop closure			w/o loop closure		w/ loop closure		
	Orientation [deg]	Translation [m]	Orientation [deg]	Translation [m]	Time [ms]	Orientation [deg]	Translation [m]	Orientation [deg]	Translation [m]	Time [ms]
V1_01_easy	6.30	0.08	**6.07**	0.14	69	6.25	0.10	**5.67**	**0.09**	13
V1_02_medium	3.27	0.11	**2.92**	**0.08**	50	3.67	0.09	**2.23**	**0.08**	9
V1_03_difficult	6.07	0.18	7.39	0.31	36	4.76	0.18	**2.29**	**0.15**	5
V2_01_easy	2.06	0.08	**1.72**	0.10	36	4.70	0.18	**1.14**	**0.15**	10
V2_02_medium	4.24	0.16	**3.44**	**0.14**	47	2.69	0.18	**1.71**	**0.10**	7
V2_03_difficult	3.21	0.27	**3.09**	**0.20**	28	3.78	0.40	**2.92**	**0.38**	3
MH_01_easy	1.20	0.15	**1.09**	0.16	68	4.58	0.28	**2.41**	**0.24**	11
MH_02_easy	1.20	0.18	1.32	0.22	64	5.61	0.45	**3.50**	**0.41**	6
MH_03_medium	1.55	0.19	1.78	**0.11**	65	4.93	0.30	**1.14**	**0.16**	11
MH_04_difficult	1.49	0.34	**1.32**	0.41	58	2.35	0.77	**2.29**	**0.67**	4
MH_05_difficult	0.69	0.30	**0.69**	**0.27**	61	5.10	0.69	**1.14**	**0.46**	10

The underlined value means the corresponding accuracy is *not* improved as expected after loop-closure corrections.



6. Broader impact

- Offer great social benefits by enabling UAVs to work in human non-accessible or unspecified environments
- Foster innovative applications in robotics such as aerial transportation during humanitarian aid and disaster relief, thus boosting economic development
- Promote hands-on learning in undergraduate education and enrich graduate curriculum, and create opportunities for students to perform meaningful robotics research