Chasing a Chimera: from VIN to Real-time High-level Understanding

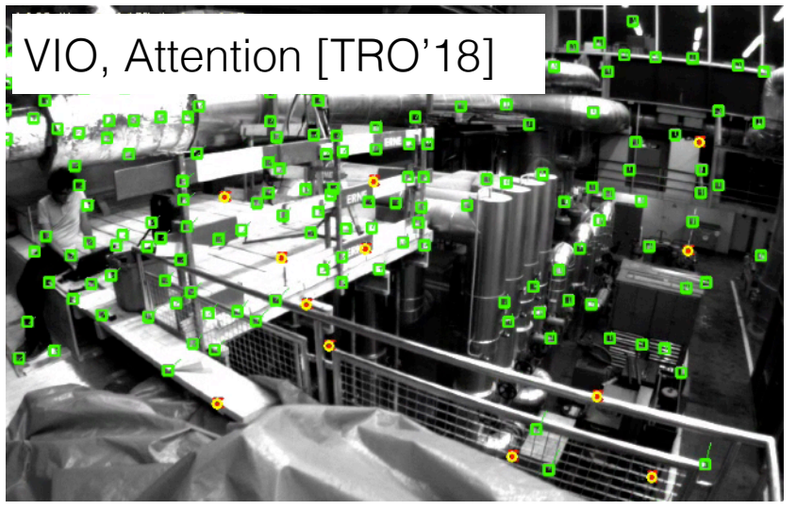
Luca Carlone
Charles Stark Draper Assistant Professor
Massachusetts Institute of Technology

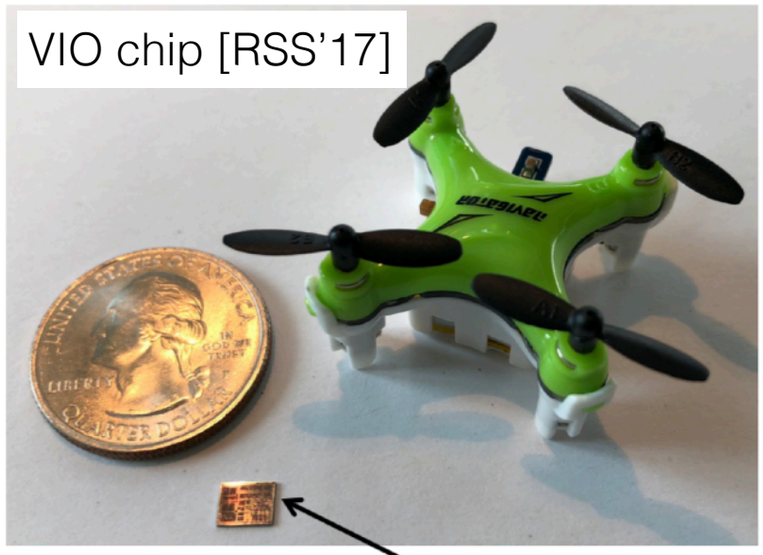
SPARK Lab

Sensing **P**erception **A**utonomy and **R**obot **K**inetics

SPARK Lab

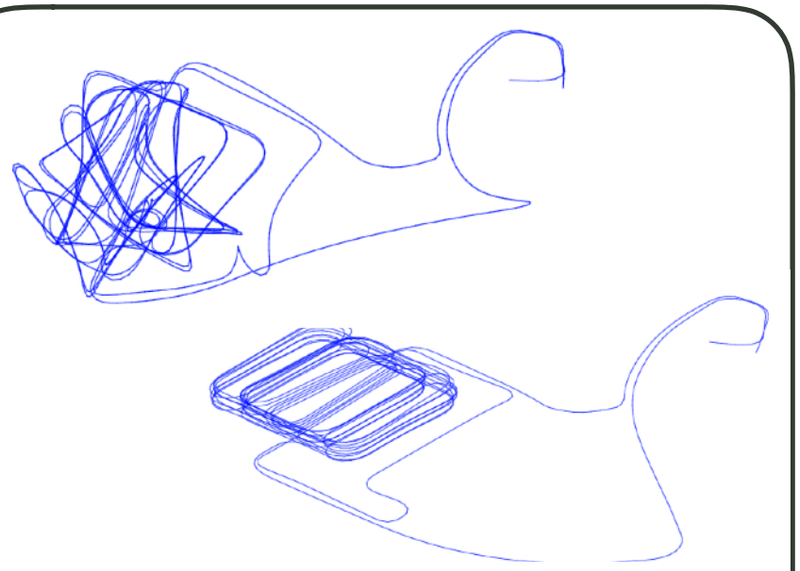
Efficiency

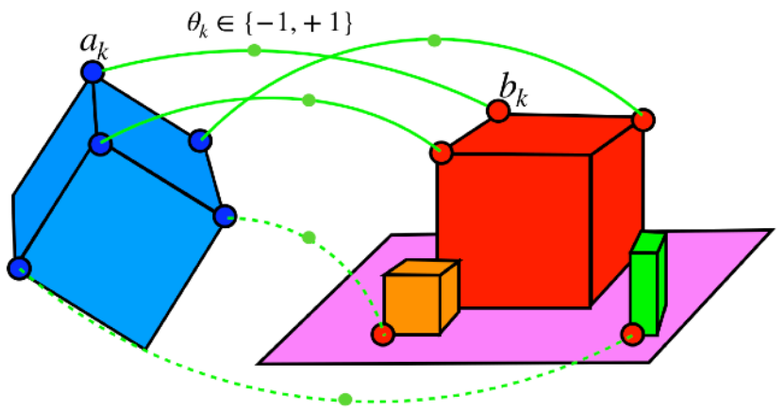

VIO, Attention [TRO'18]




Distributed algorithms [IJRR'17]

Robustness

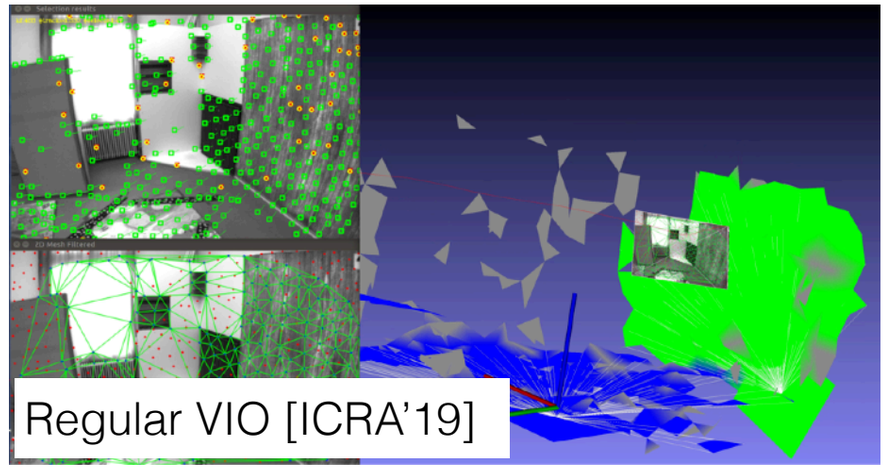

Pose graph optimization: duality [T-RO'15], SE-sync [IJRR'18]

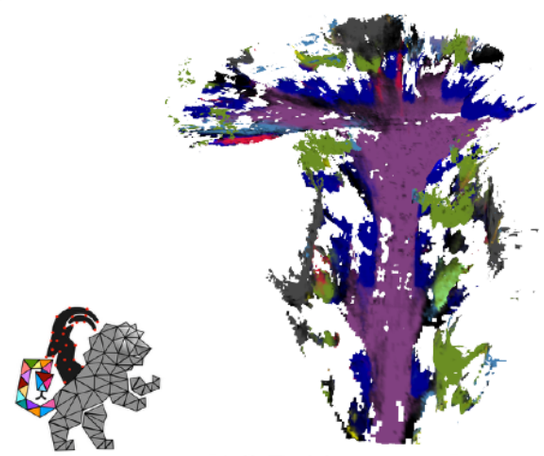


Outlier-robust object detection [RSS'19, RAL'19]

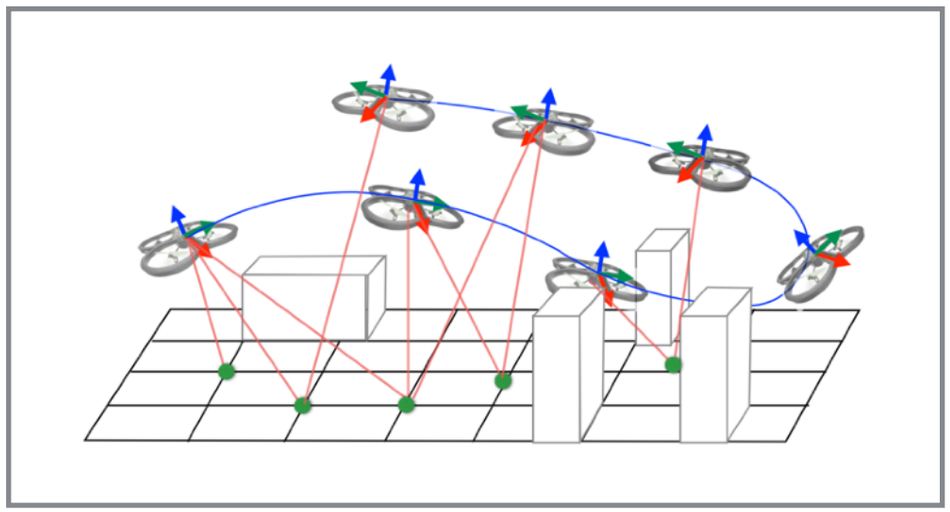
High-level understanding


FUSES [RAL'19]

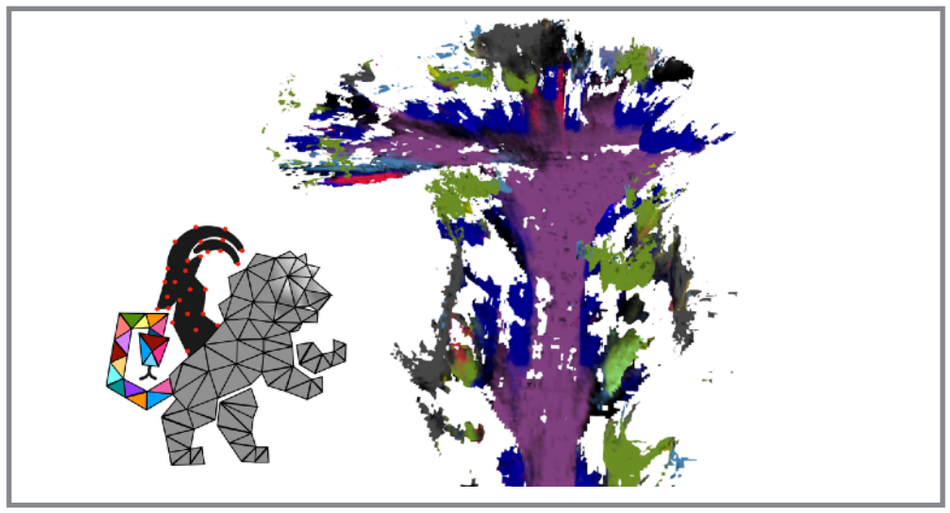

Regular VIO [ICRA'19]


Kimera [ICRA'19-20]

Outline

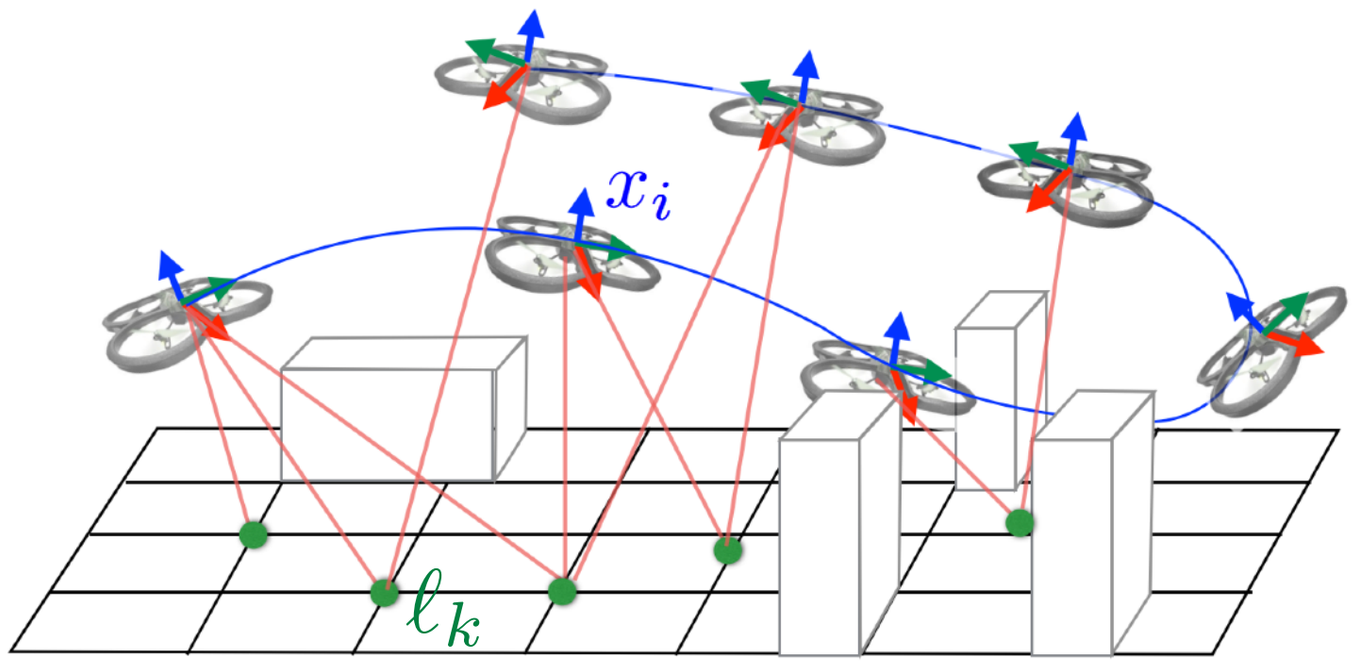


Visual-Inertial Navigation:
an optimization lens



Kimera: real-time
high-level understanding

Maximum A-Posteriori (MAP) Estimation

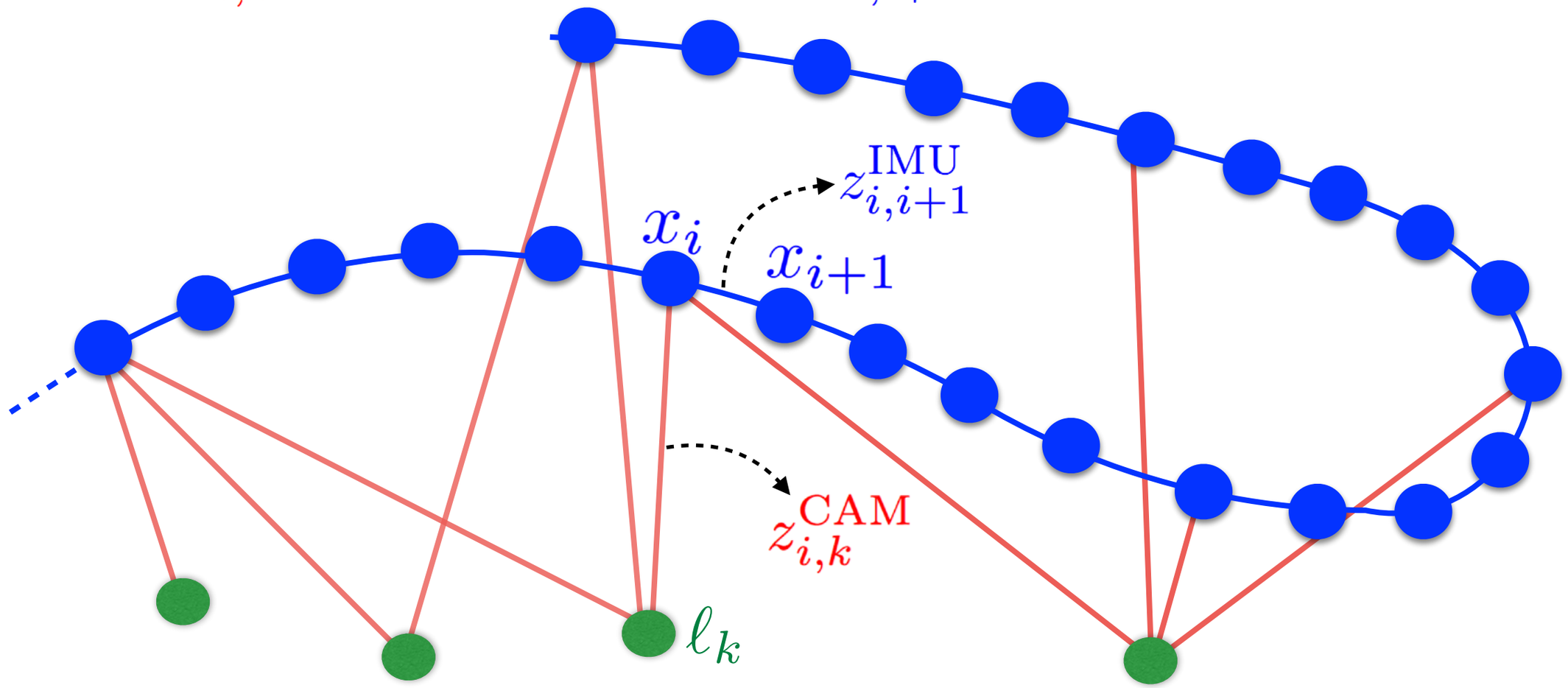


$$X^\star = \underset{x_i, \ell_k \forall i,k}{\arg\min} \sum_{i,k} \| z_{i,k}^{\text{CAM}} - \pi(x_i, \ell_k) \|^2 + \sum_{i,i+i} \| z_{i,i+i}^{\text{IMU}} - f(x_i, x_{i+1}) \|^2$$

Maximum A-Posteriori (MAP) Estimation

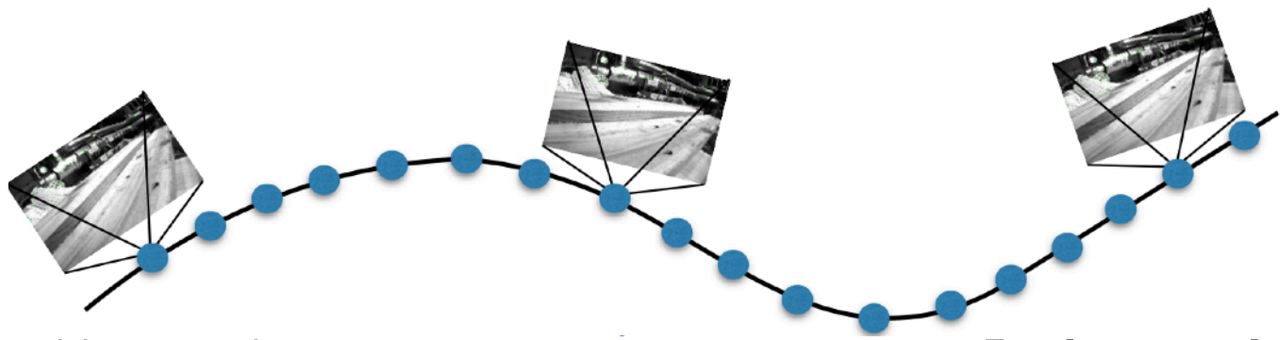
$$X^\star = \underset{x_i, \ell_k \forall i,k}{\arg\min} \sum_{i,k} \| z_{i,k}^{\text{CAM}} - \pi(x_i, \ell_k) \|^2 + \sum_{i,i+i} \| z_{i,i+i}^{\text{IMU}} - f(x_i, x_{i+1}) \|^2$$

Factor graph



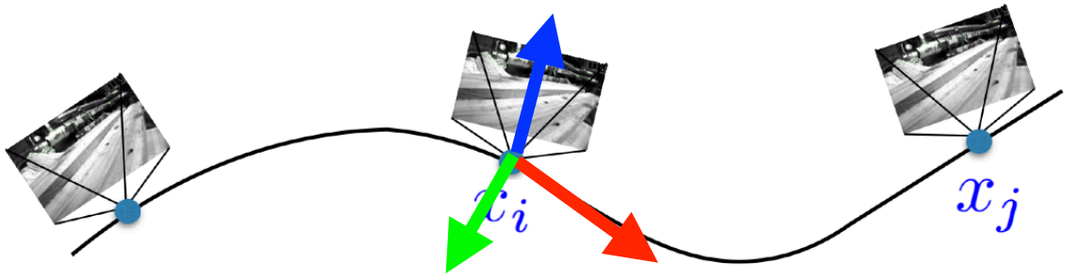
IMU Preintegration

How to save computation?



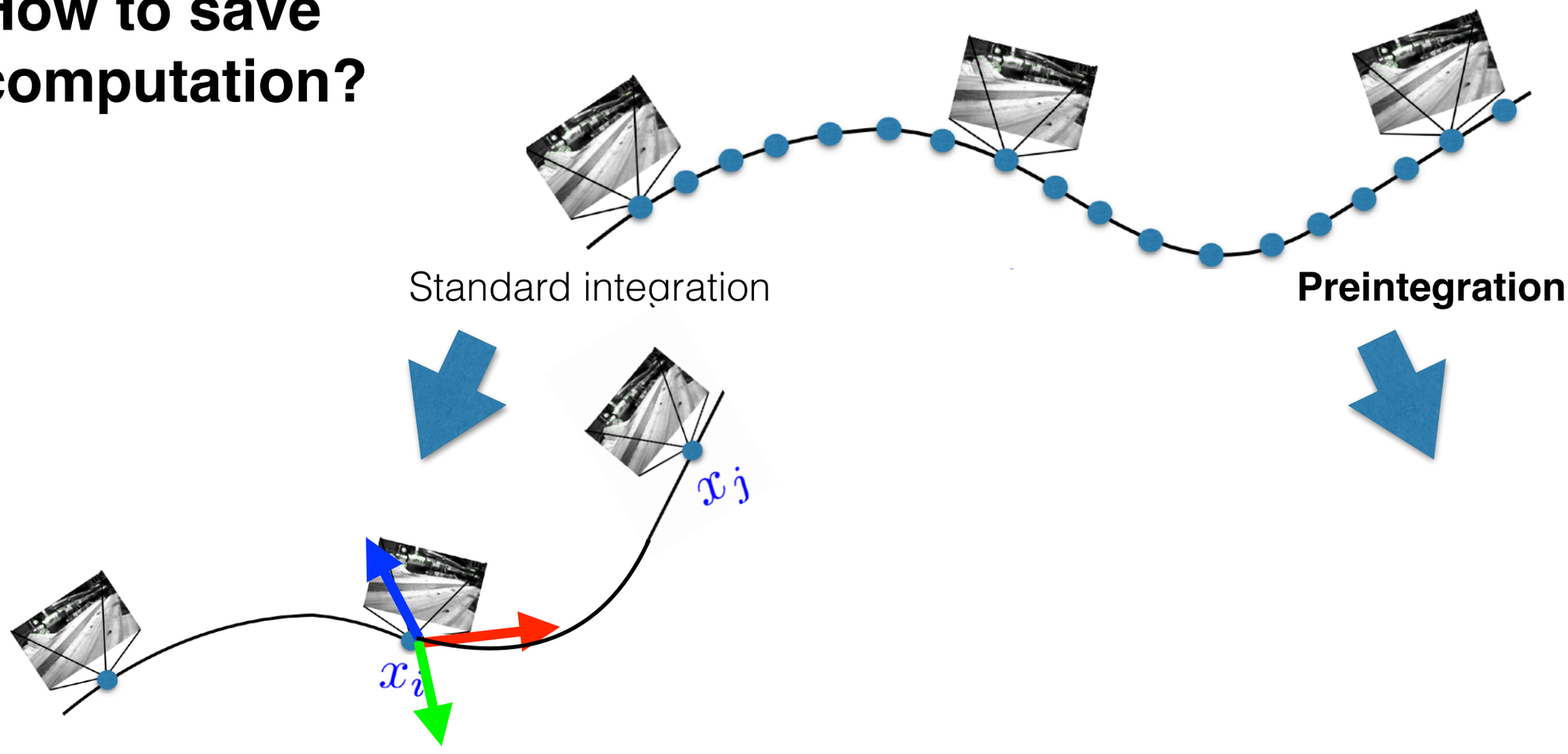
Standard integration

Preintegration



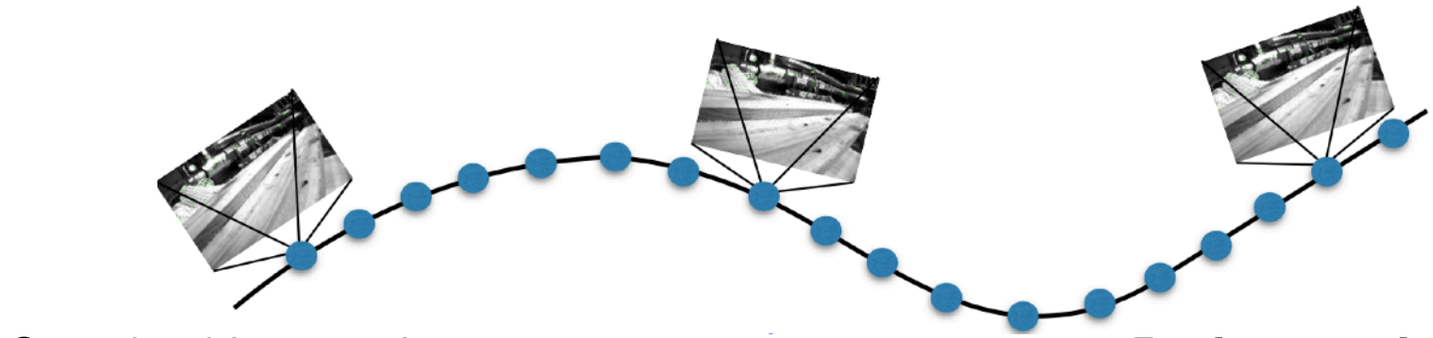
IMU Preintegration

How to save computation?



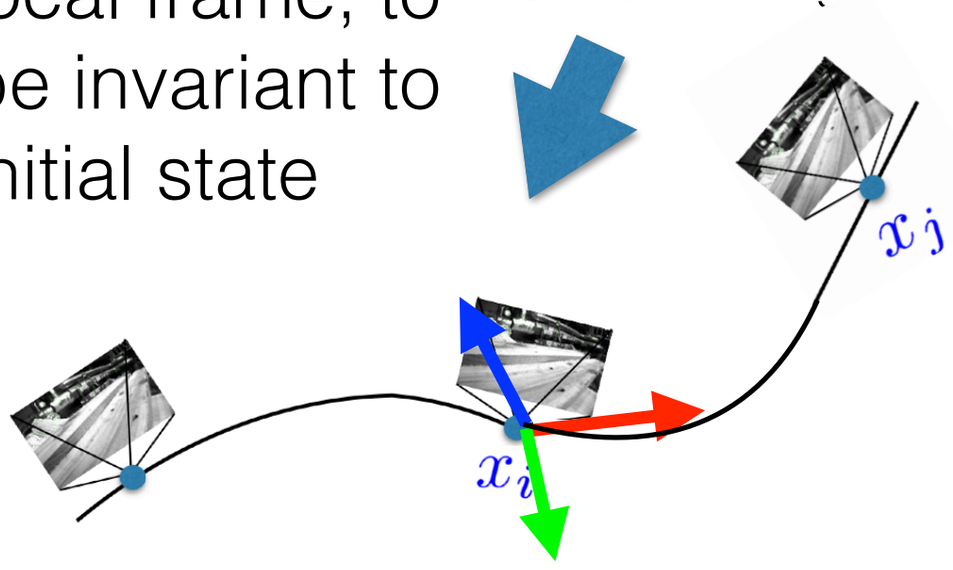
IMU Preintegration

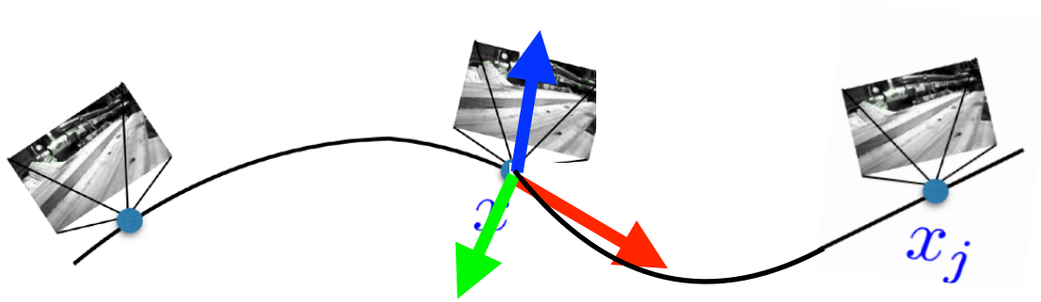
How to save computation?
integration is performed in local frame, to be invariant to initial state



Standard integration

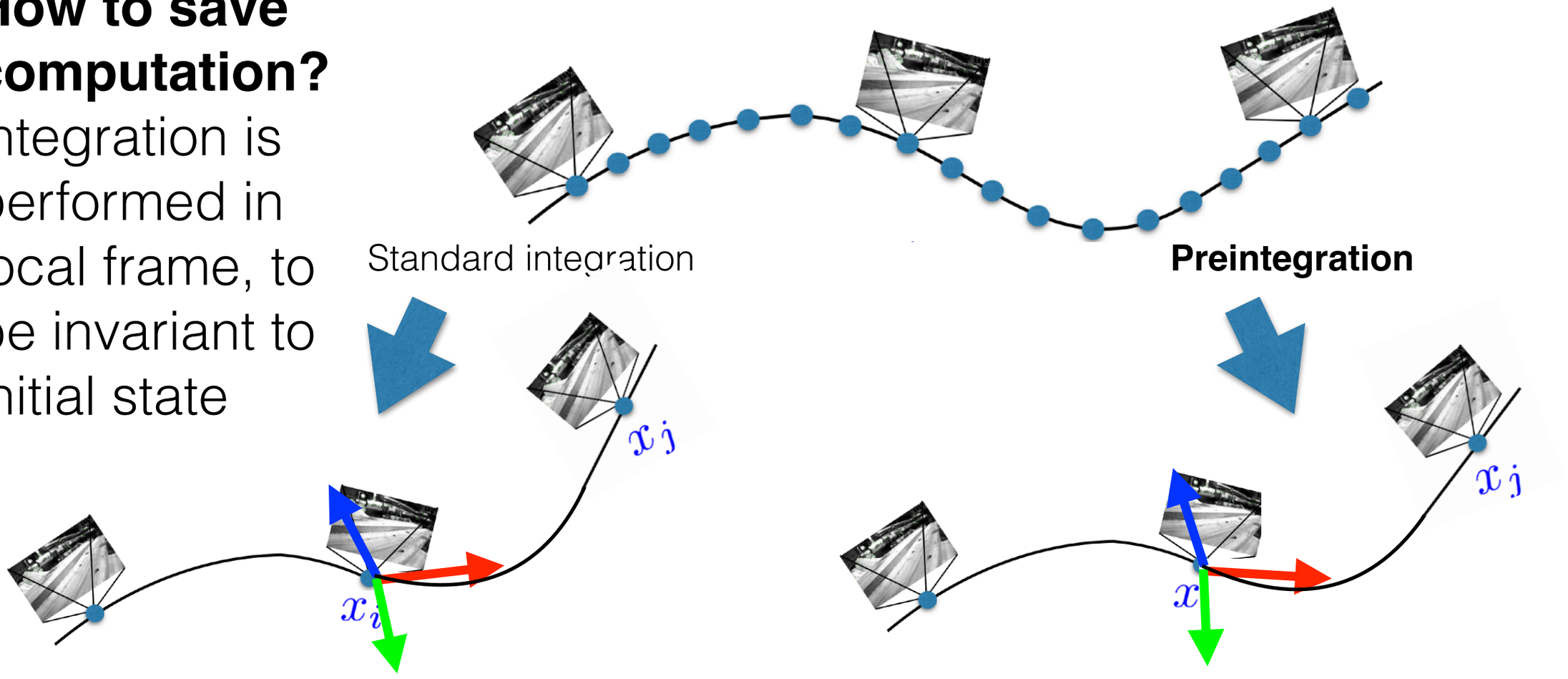
Preintegration



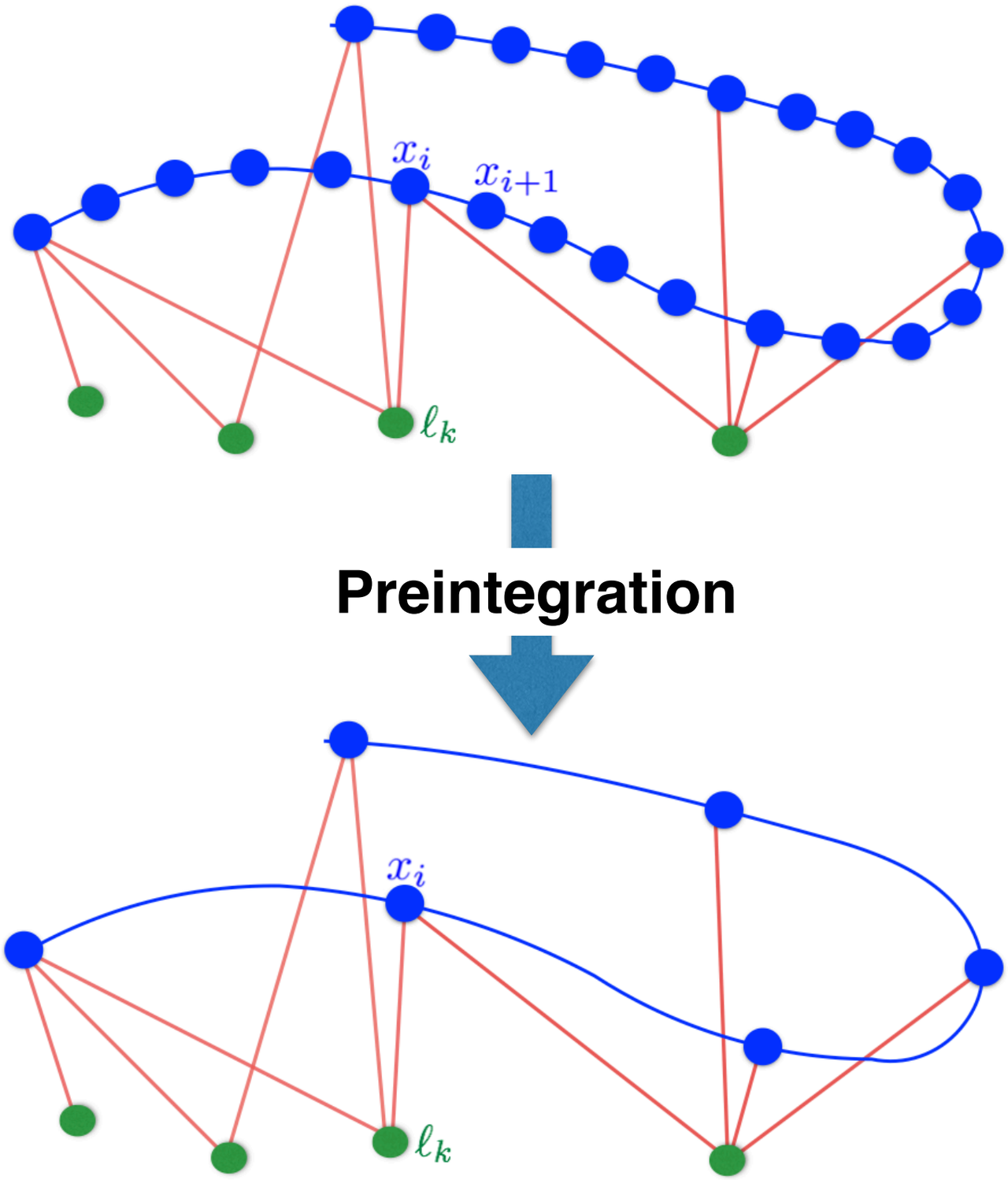


IMU Preintegration

How to save computation?
integration is performed in local frame, to be invariant to initial state



IMU Preintegration



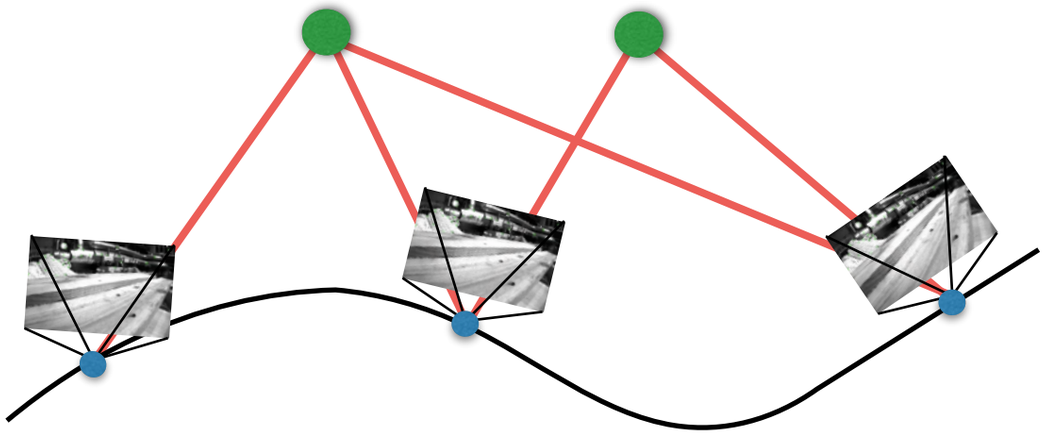
After 10 seconds, original problem has ~10^4 states

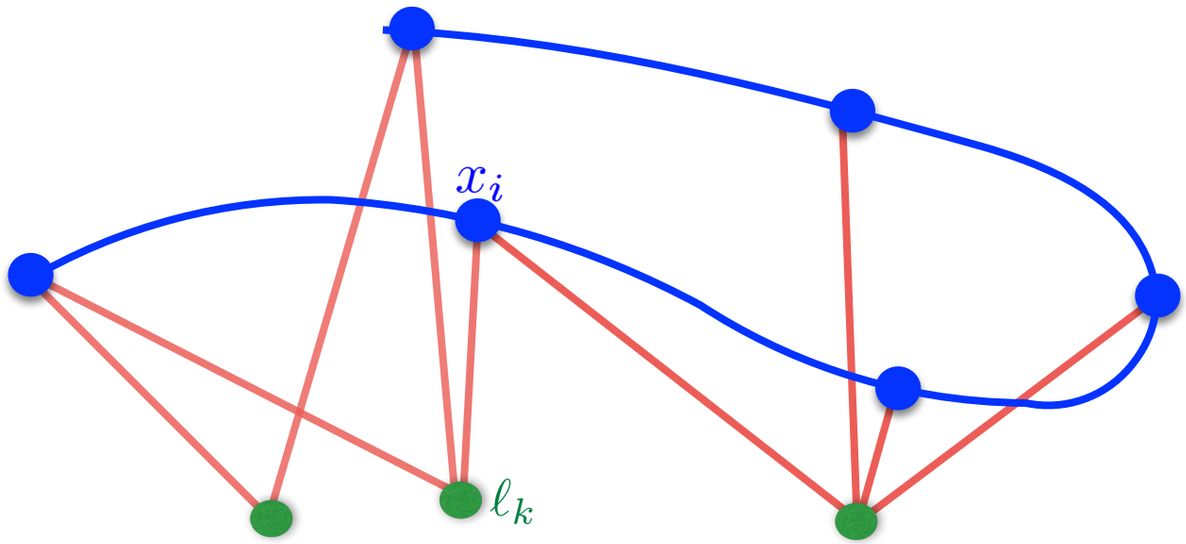
After 10 seconds, preintegrated problem has ~10^2 states

Lupton and Sukkarieh, Visual-inertial-aided navigation for high-dynamic motion in built environments without initial conditions, TRO'12
Forster, Carlone, Dellaert, Scaramuzza, On-Manifold Preintegration for Real-Time Visual-Inertial Odometry, TRO'17 (best paper award)

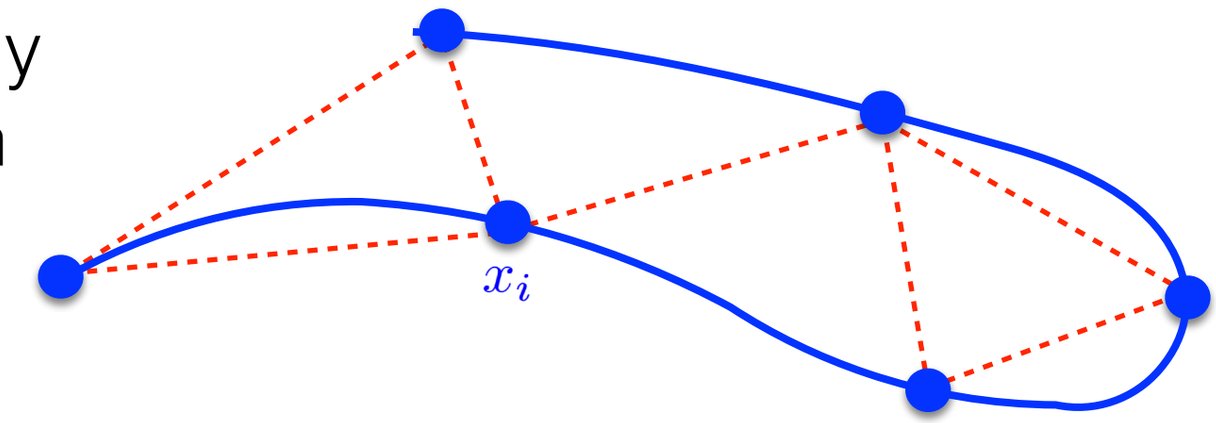
Structureless Vision Model

Conditional independence: each 3D landmark can be computed independently once robot state is known





Schur complement



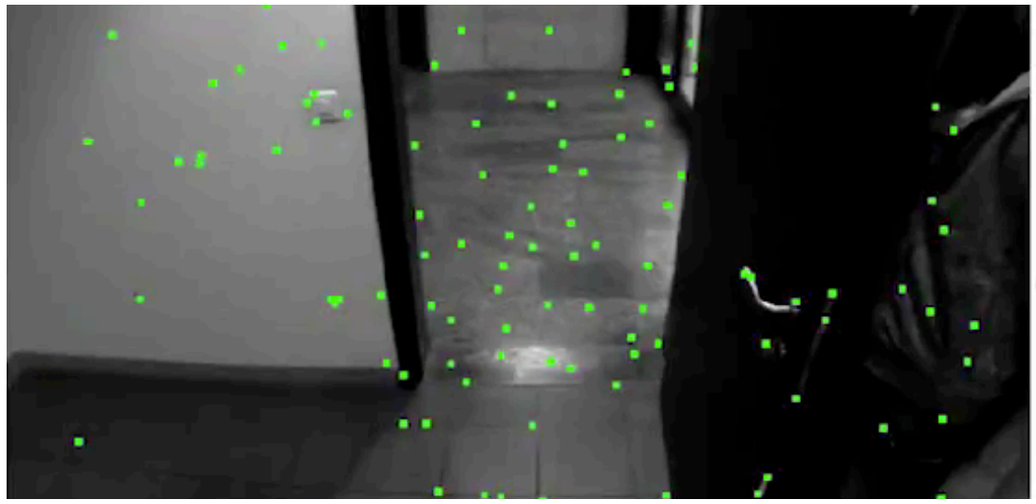
Schur complement trick:

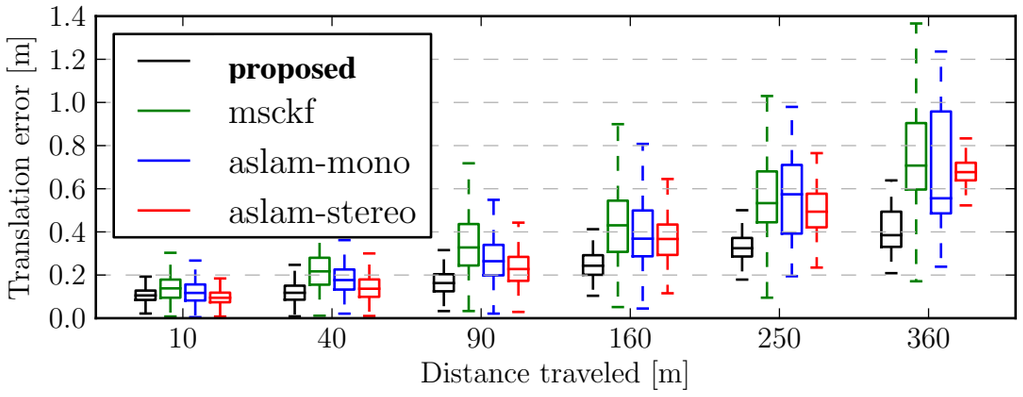
- solve for each landmark separately
- substitute back in the optimization

see also:
Null space trick [Roumeliotis and Mourikis]
Schur complement trick in computer vision

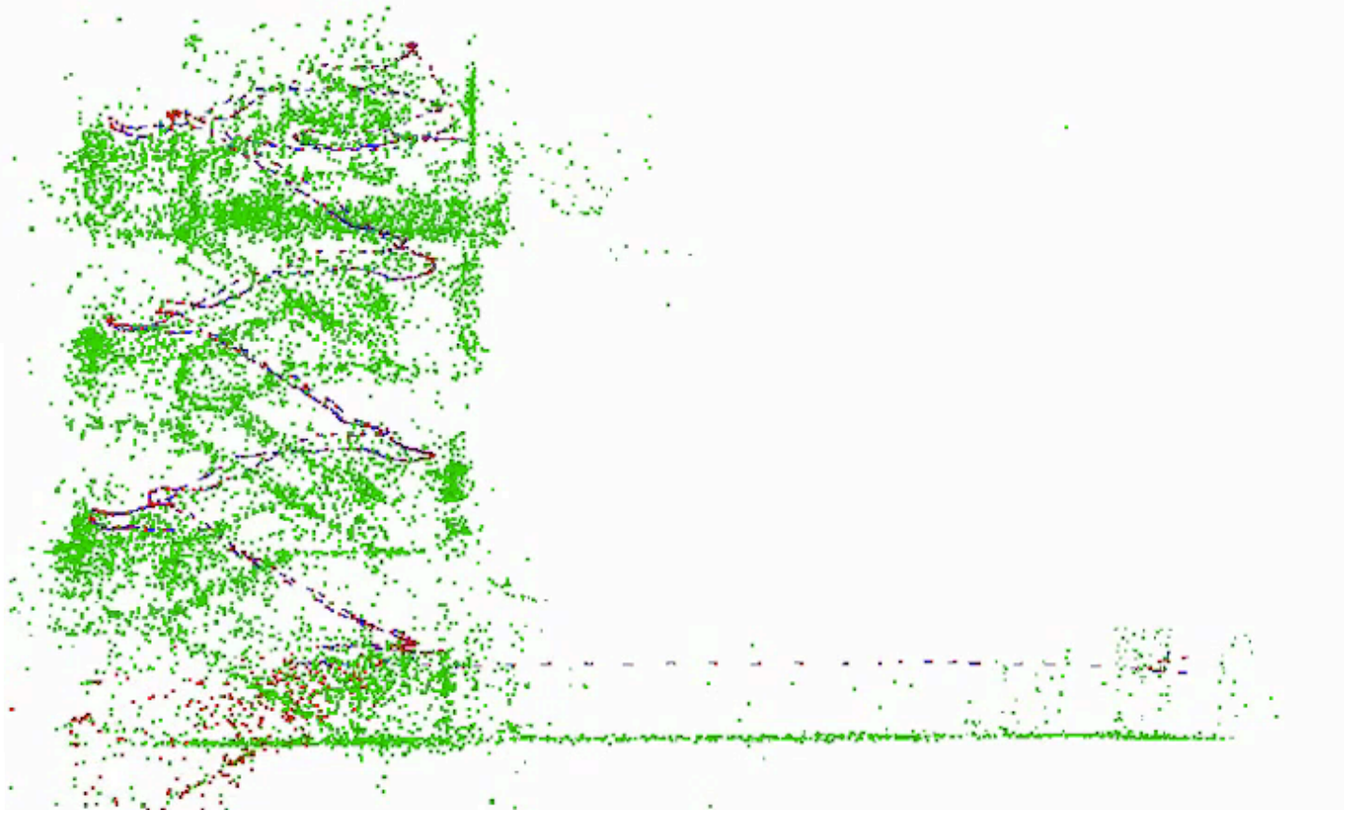
Carlone, Alcantarilla, Chiu, Zsolt, Dellaert, *Mining structure fragments for smart bundle adjustment*, BMVC'14
Mourikis and Roumeliotis. A multi-state constraint Kalman filter for vision-aided inertial navigation. ICRA, 2007.

Results: multiple platforms

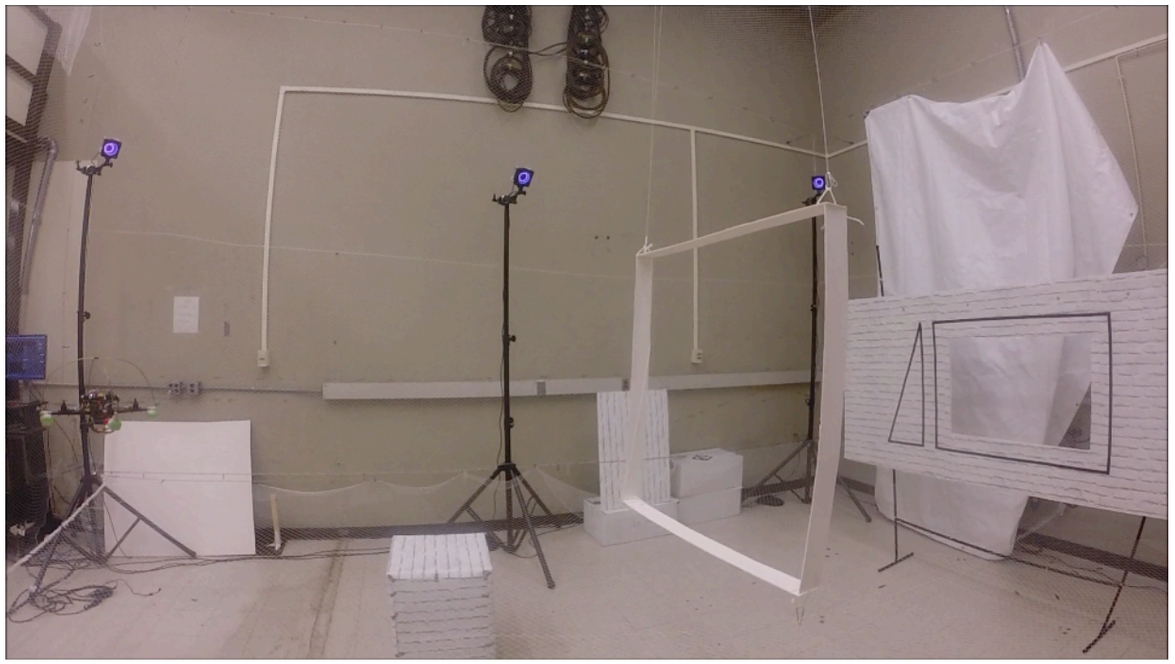




< 0.5% position drift



[2014-2015]



[2016-2017]

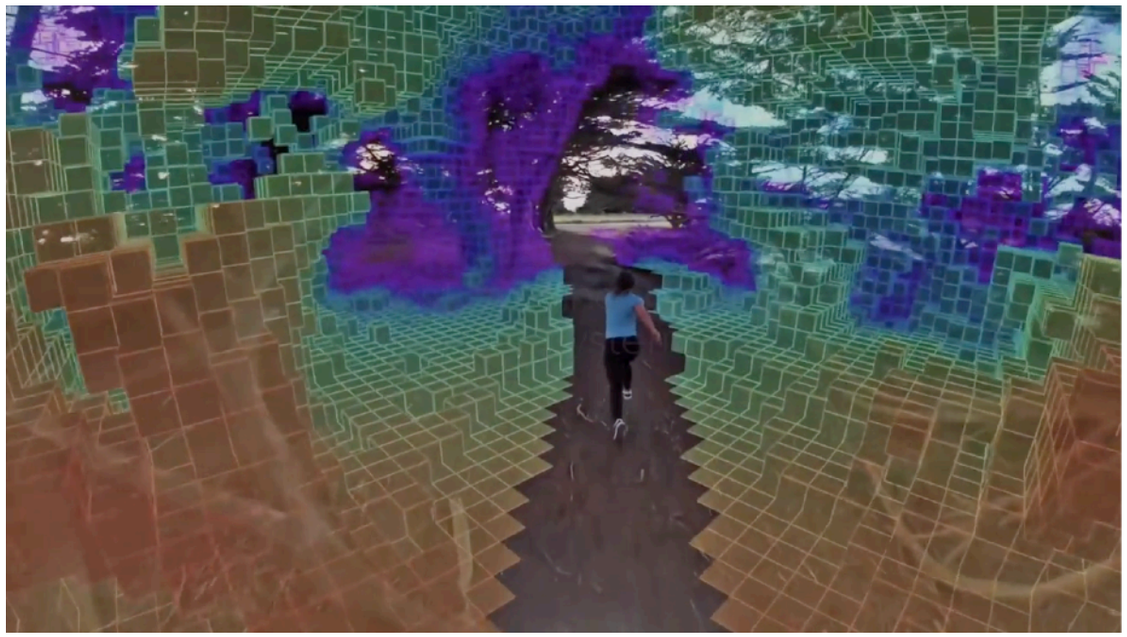
Implemented in standard libraries

- **GTSAM**
- **VI ORB-SLAM**
- **VINS-mono**
- **…**

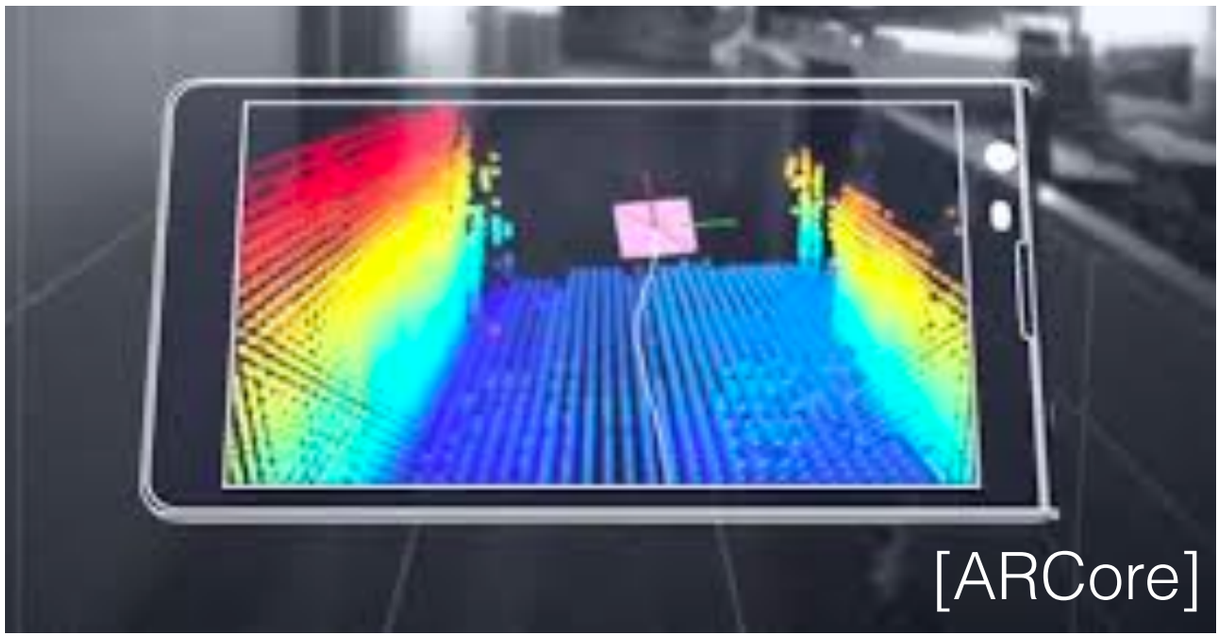
Forster, Carlone, Dellaert, Scaramuzza, *IMU Preintegration on Manifold for Efficient Visual-Inertial Maximum-a-Posteriori Estimation*, RSS'15 (best paper finalist)
Forster, Carlone, Dellaert, Scaramuzza, On-Manifold Preintegration for Real-Time Visual-Inertial Odometry, TRO'17 (best paper award)

Engineered Solutions / Applications

Skydio R1 drone

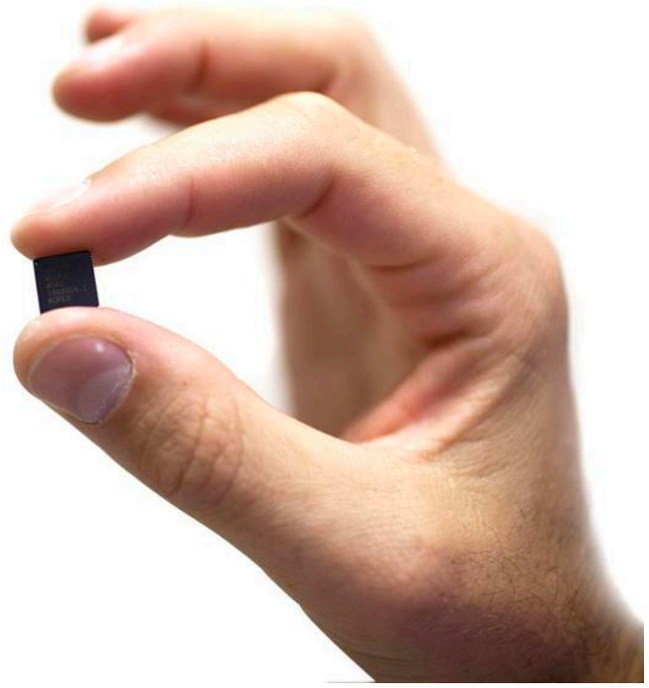


Google Tango



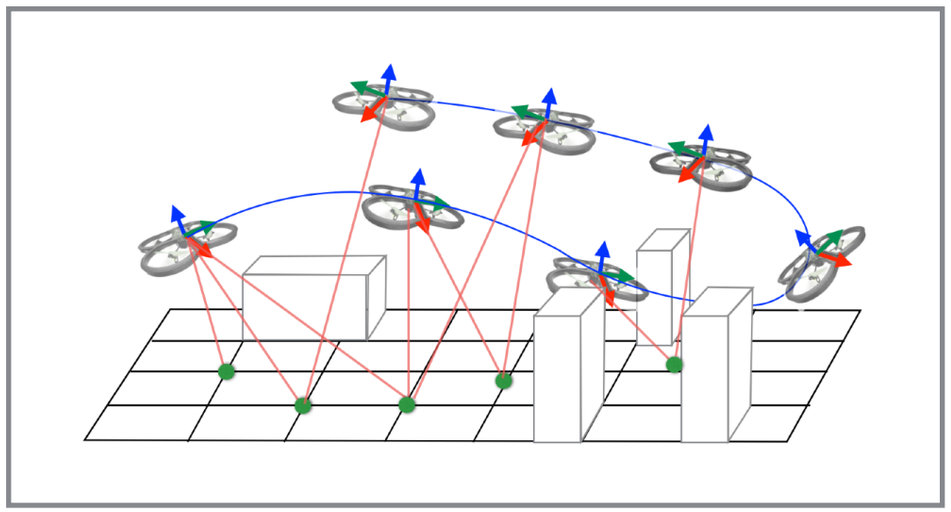
Oculus Rift Goggles

Pokemon Go

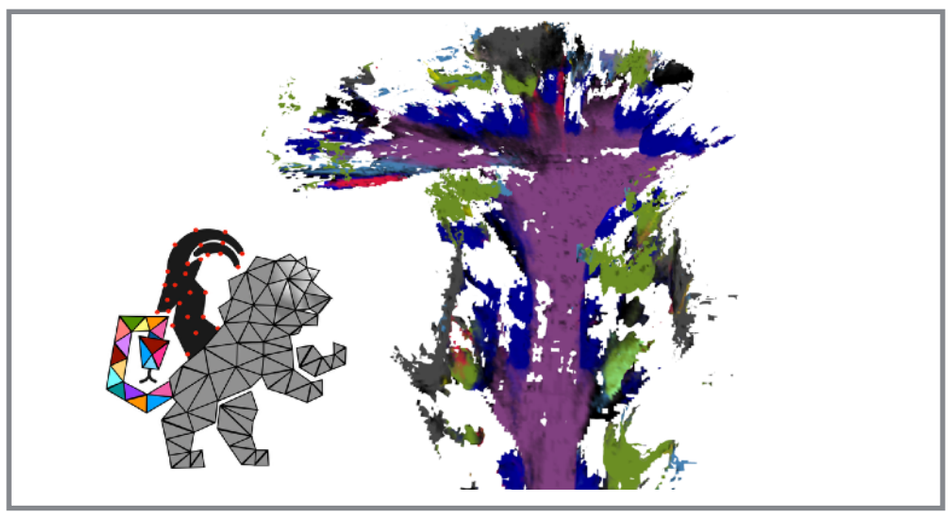


Navion Chip
2017

Outline

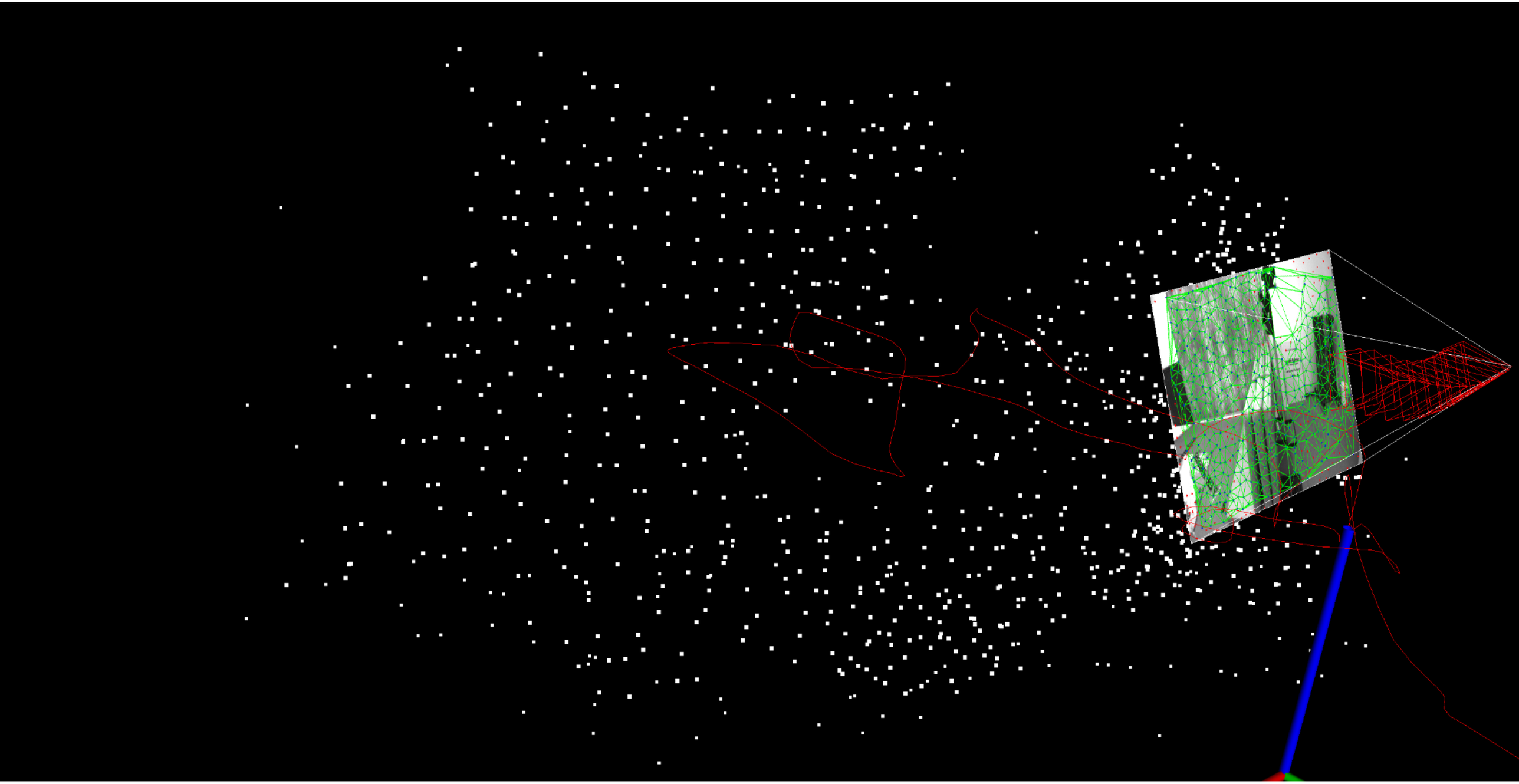


Visual-Inertial Navigation:
an optimization lens



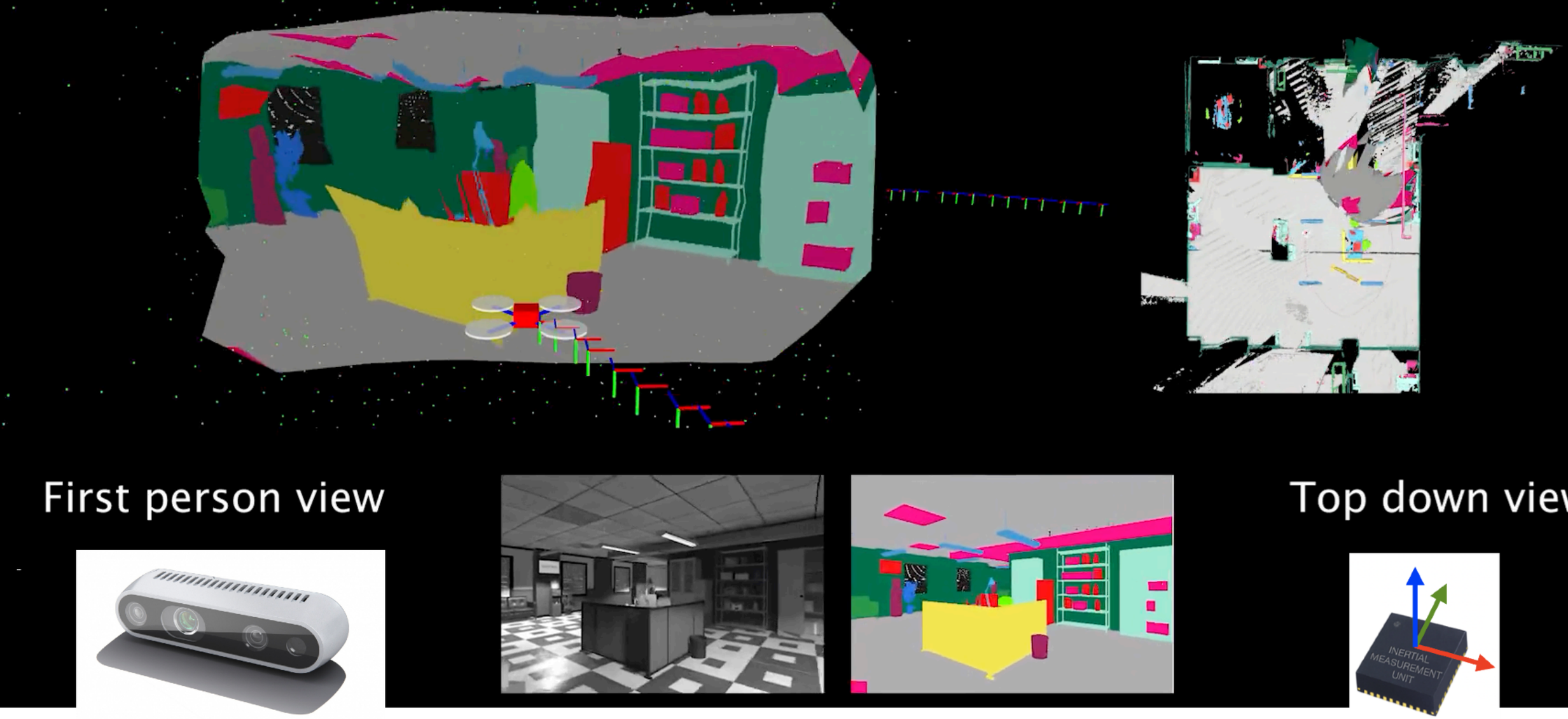
Kimera: real-time
high-level understanding

From VIO to High-level Understanding



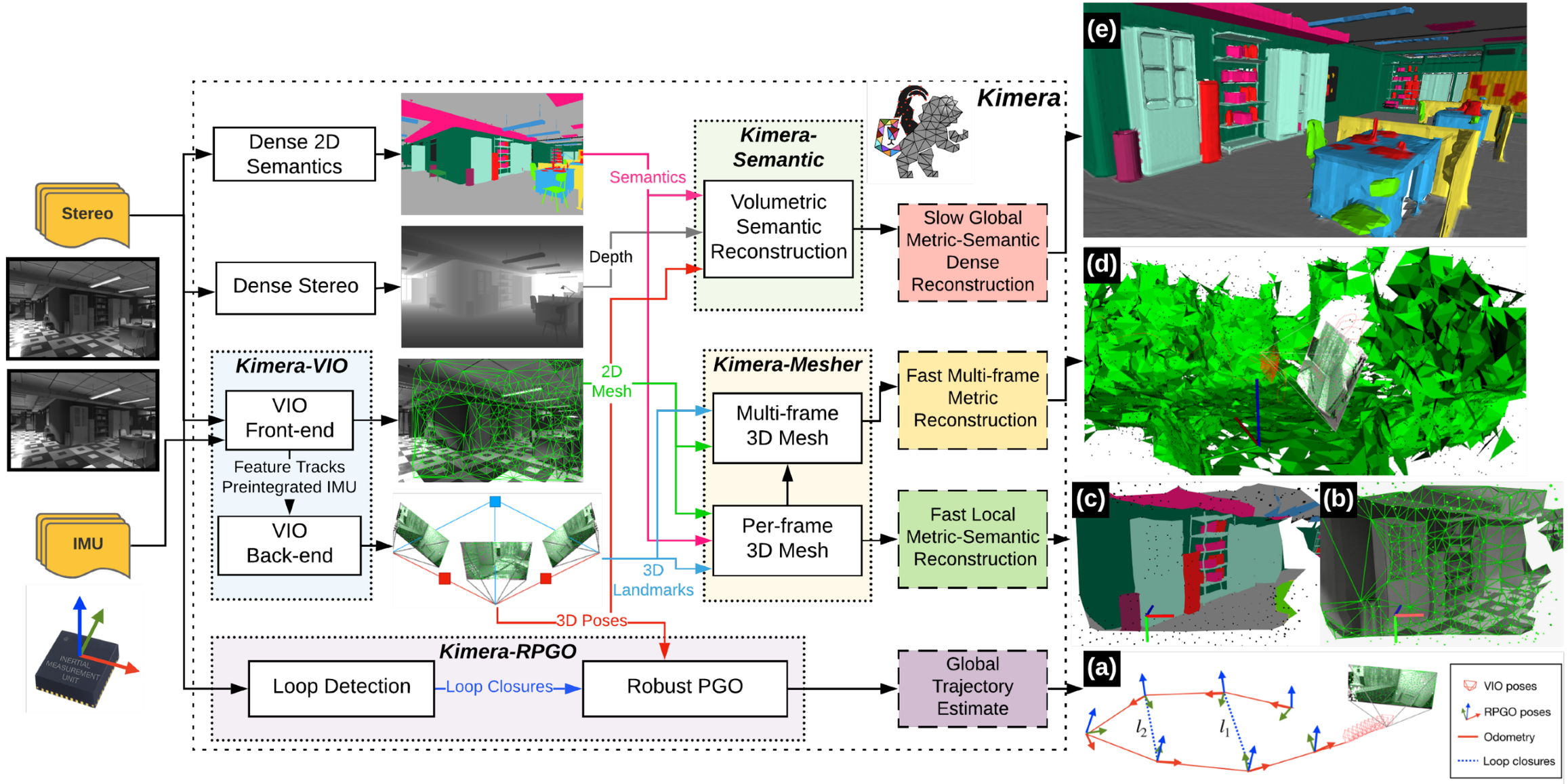
Releasing Kimera

Real-time metric-semantic visual-inertial SLAM



A. Rosinol, M. Abate, Y. Chang, and L. Carlone. Kimera: an open-source library for real-time metric-semantic localization and mapping. Arxiv 1910.02490, 2019.

Architecture

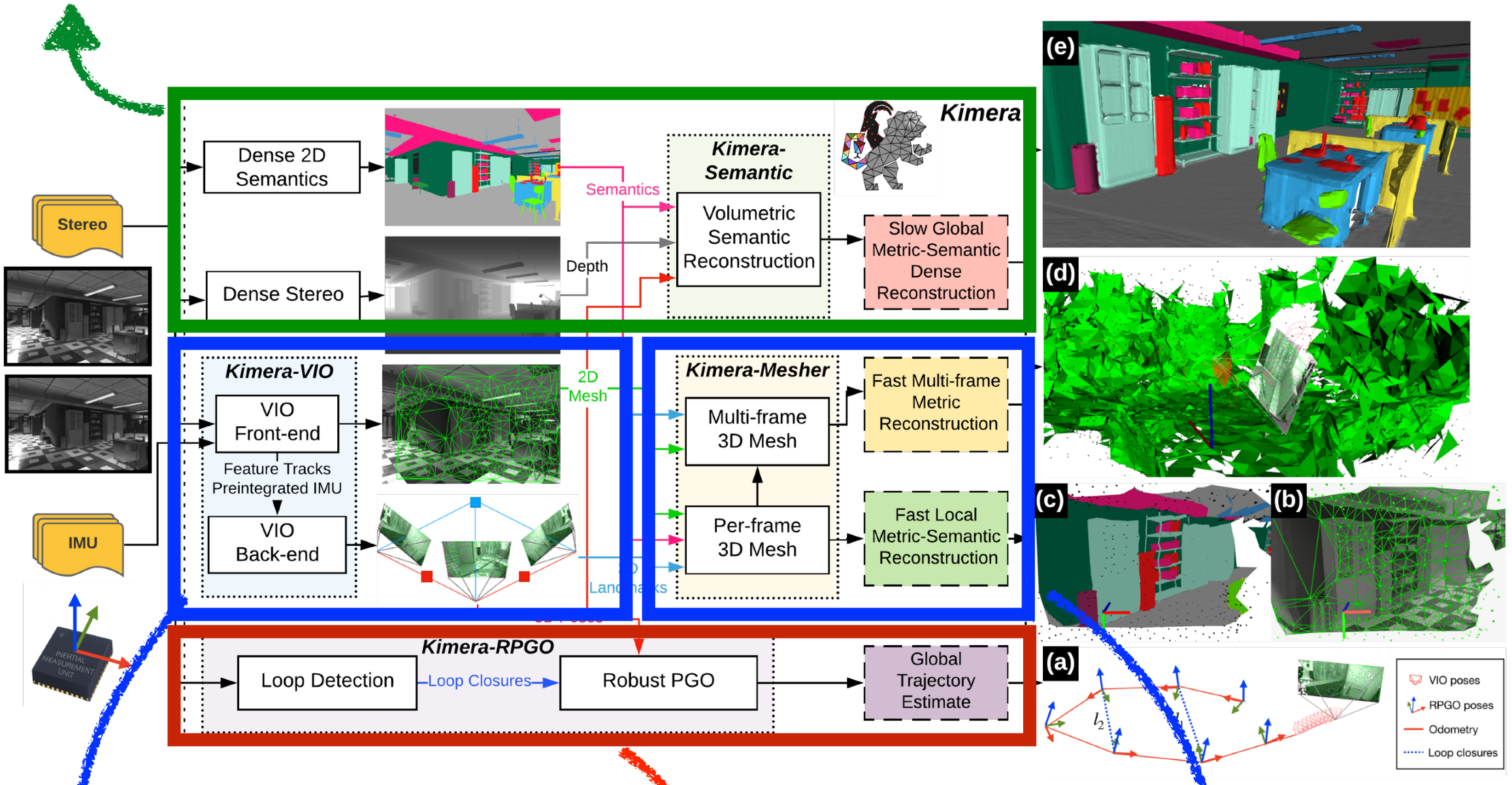


Outputs:

- high-rate state estimates (@IMU rate)
- local mesh (@50Hz)
- global trajectory estimate including loop closures (<10Hz)
- global mesh reconstruction (~1Hz)

Architecture

Kimera-Semantics



Kimera-VIO

use OpenGV!

Kimera-RPGO

Kimera-Mesher

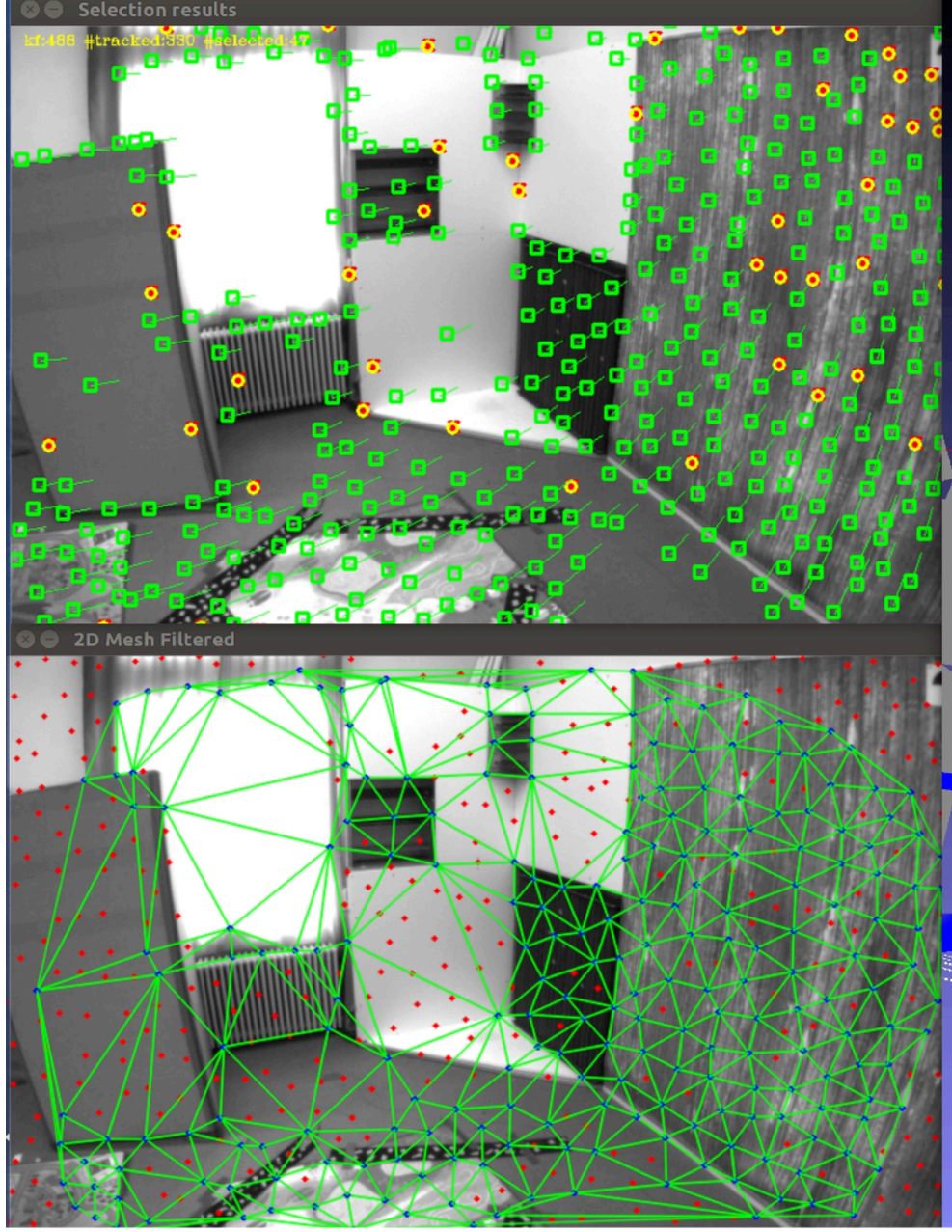
GTSAM-based!

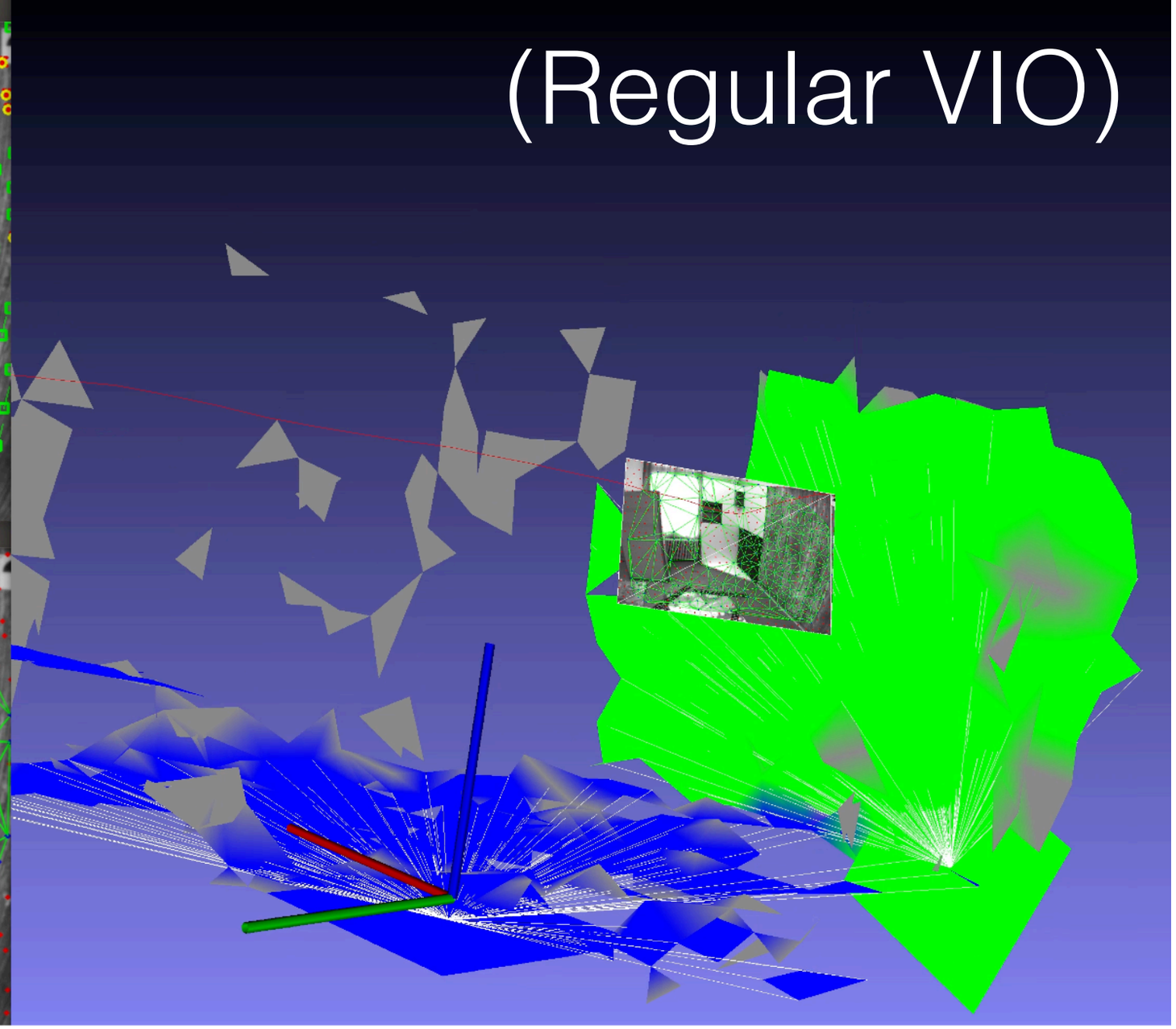
Kimera-VIO & Kimera-Mesher

Kimera-VIO tracks sparse 3D landmarks for fast and accurate state estimation

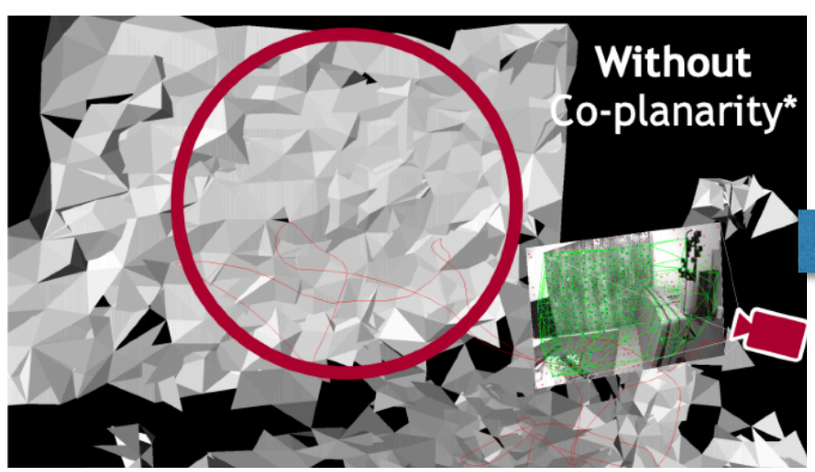
	RMSE ATE [cm]				
Seq.	OKVIS	MSCKF	ROVIO	VINS-Mono	**Kimera-VIO**
MH_01	16	42	21	15	**11**
MH_02	22	45	25	15	**10**
MH_03	24	23	25	22	**16**
MH_04	34	37	49	32	**24**
MH_05	47	48	52	**30**	35
V1_01	9	34	10	8	**5**
V1_02	20	20	10	11	**8**
V1_03	24	67	14	18	**7**
V2_01	13	10	12	**8**	**8**
V2_02	16	16	14	16	**10**
V2_03	29	113	**14**	27	21

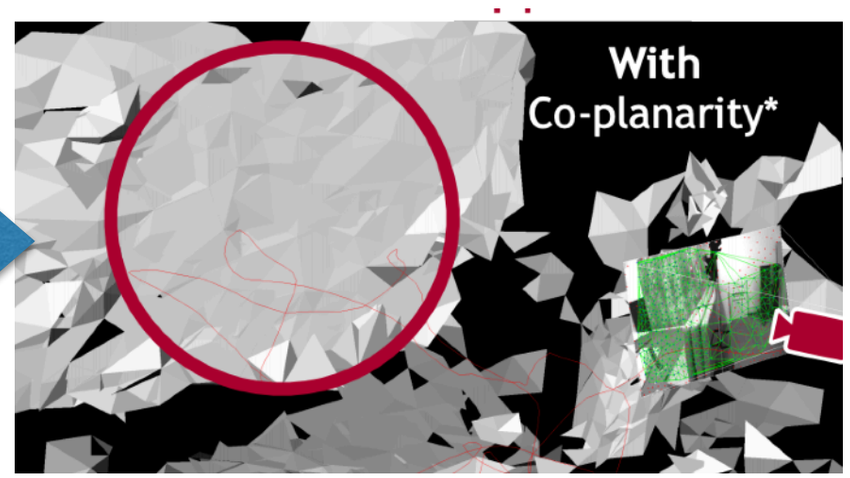
- IMU Preintegration + Structureless Vision Factors [RSS'15, TRO'17]
- Regular VIO [ICRA'19]





Tightly coupled mesh regularization and VIO





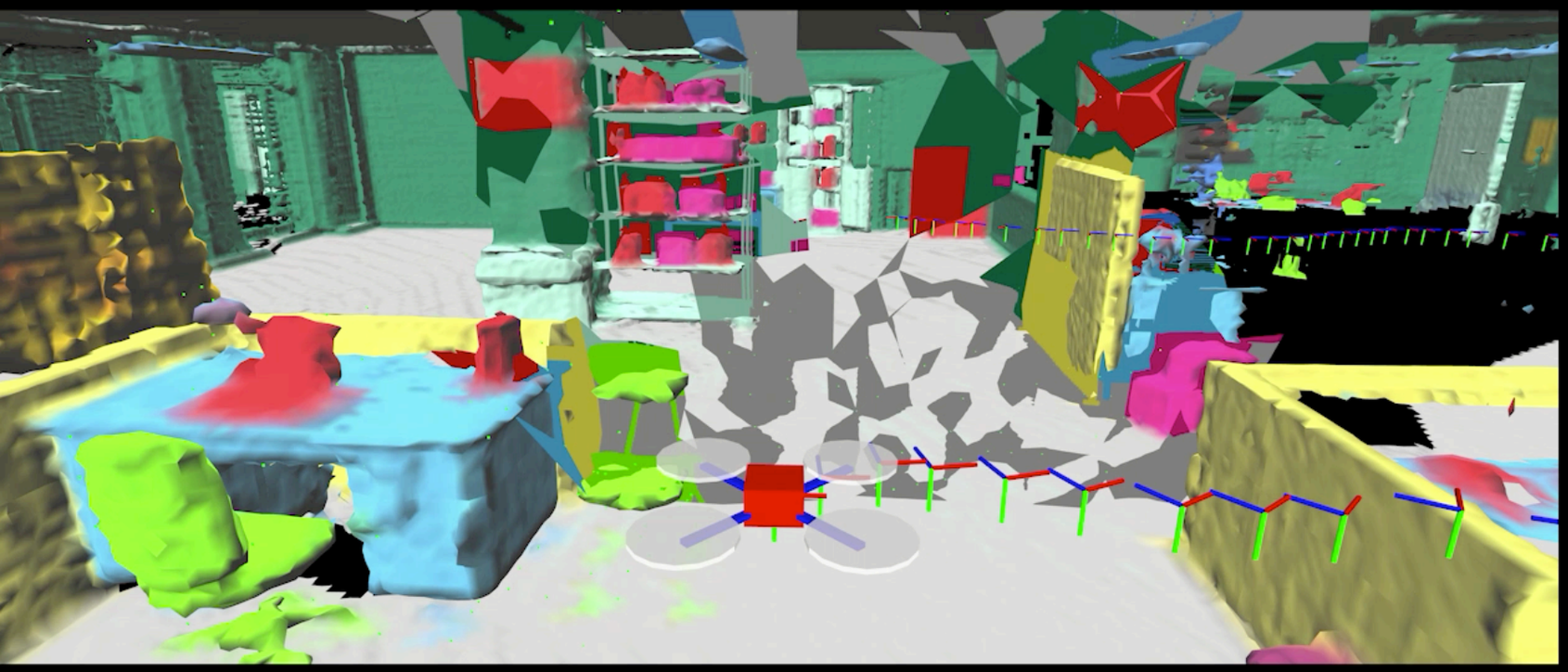
A. Rosinol, T. Sattler, M. Pollefeys, L. Carlone, “Incremental visual-inertial 3D mesh generation with structural regularities,” I*CRA*, 2019.

Kimera-RPGO (**R**obust **P**ose **G**raph **O**ptimization

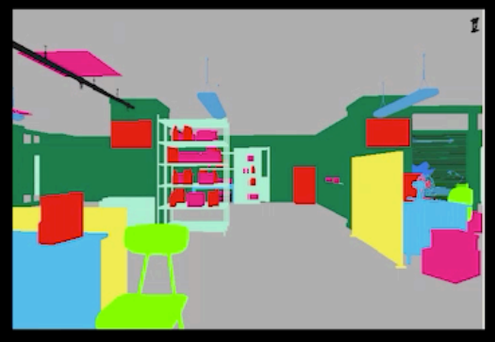
- Incremental implementation of Pairwise Consistency Maximization [Mangelson et al., ICRA'18]

Kimera-Semantics

Kimera-Semantic performs Bayesian updates for each voxel by propagating semantic labels using bundled raycasting



First person view



Top down view

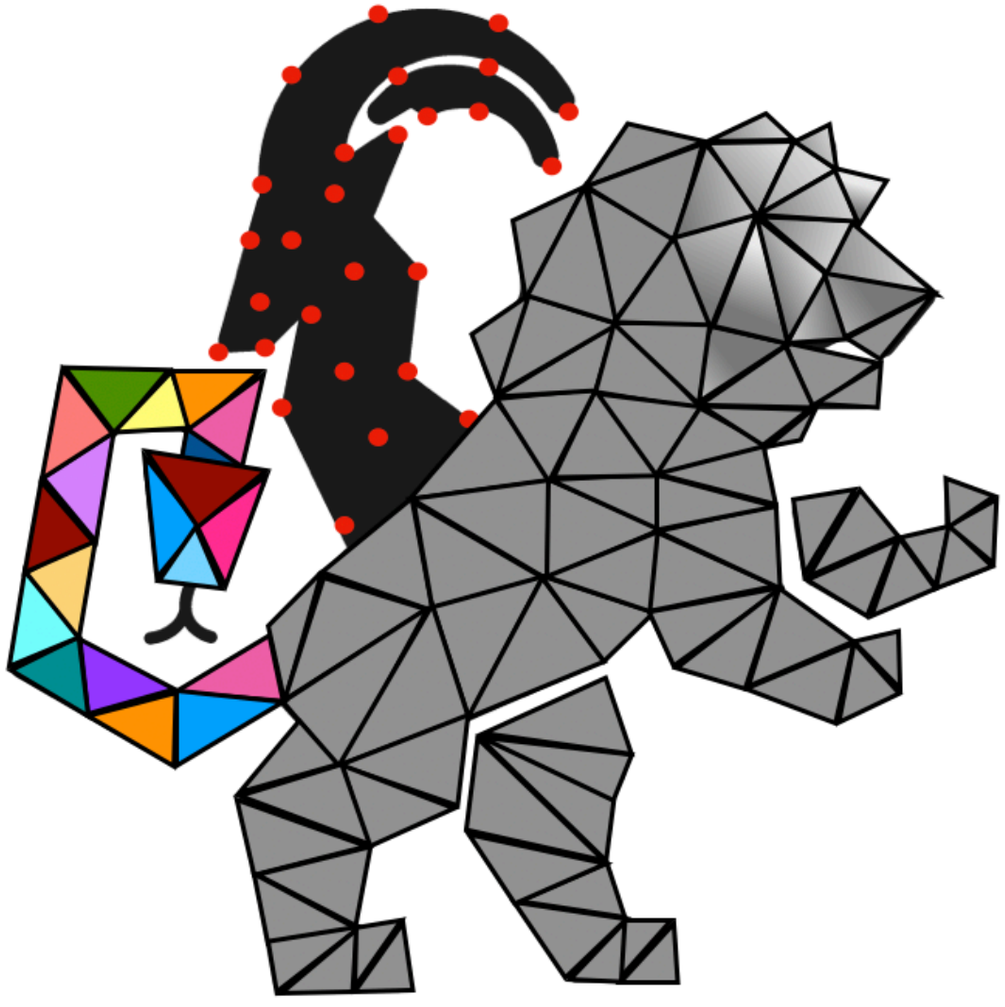
Speed x3

Metrics		Kimera-Semantics using: GT Depth GT Poses	GT Depth Kimera-VIO	Dense-Stereo Kimera-VIO
Semantic	mIoU [%]	80.10	80.03	57.23
	Acc [%]	94.68	94.50	80.74
Geometric	ATE [m]	0.0	0.04	0.04
	RMSE [m]	0.079	0.131	0.215

Based on VOXBLOX for metric reconstruction

Why Kimera?

Why Kimera?

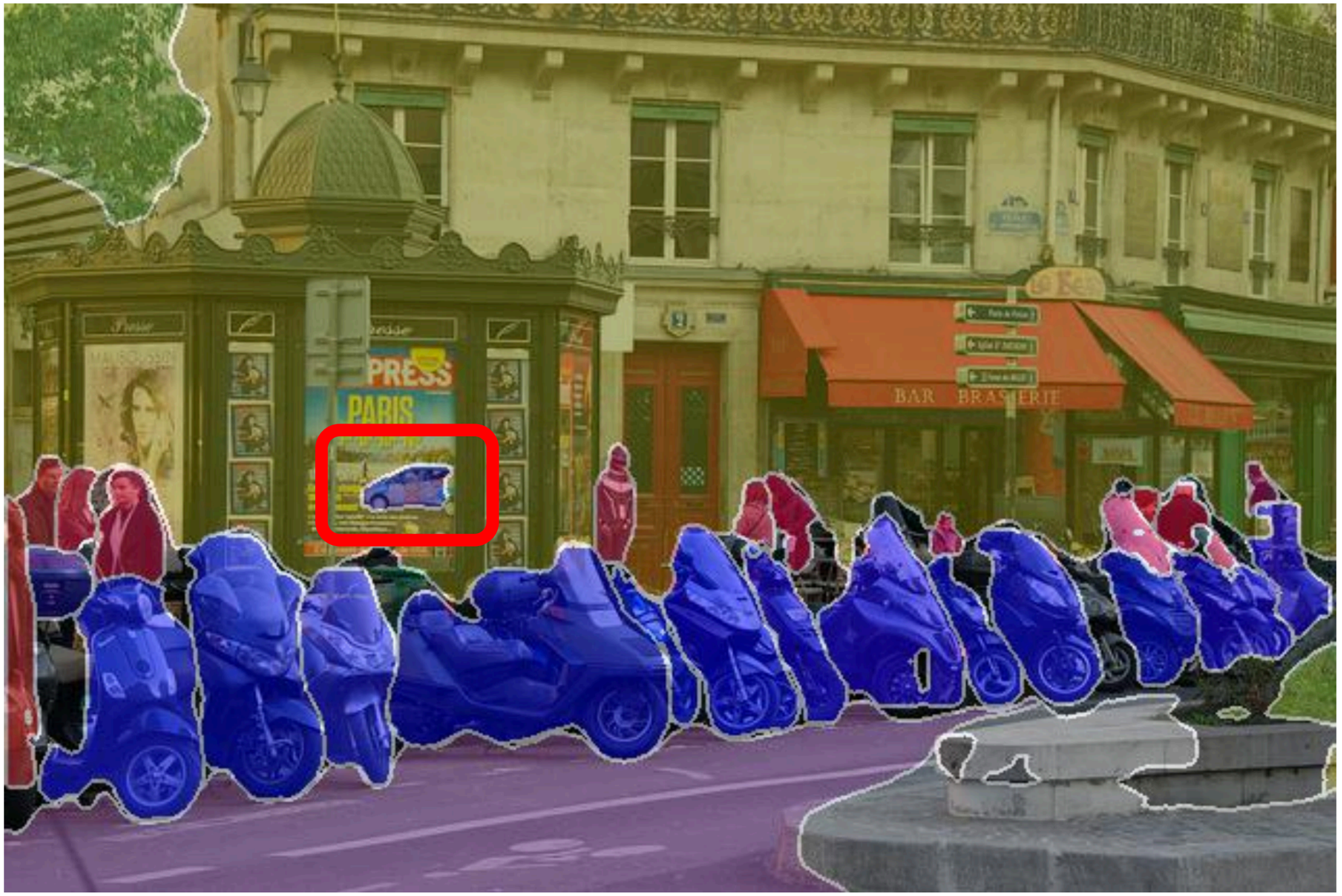


The **Chimera** (/kɪˈmɪərə/ or /kaɪˈmɪərə/, also **Chimaera** (*Chimæra*); Greek: Χίμαιρα, *Chímaira* "she-goat") according to Greek mythology,[1] was a monstrous fire-breathing hybrid creature of Lycia in Asia Minor, composed of the parts of more than one animal.

Why Kimera?

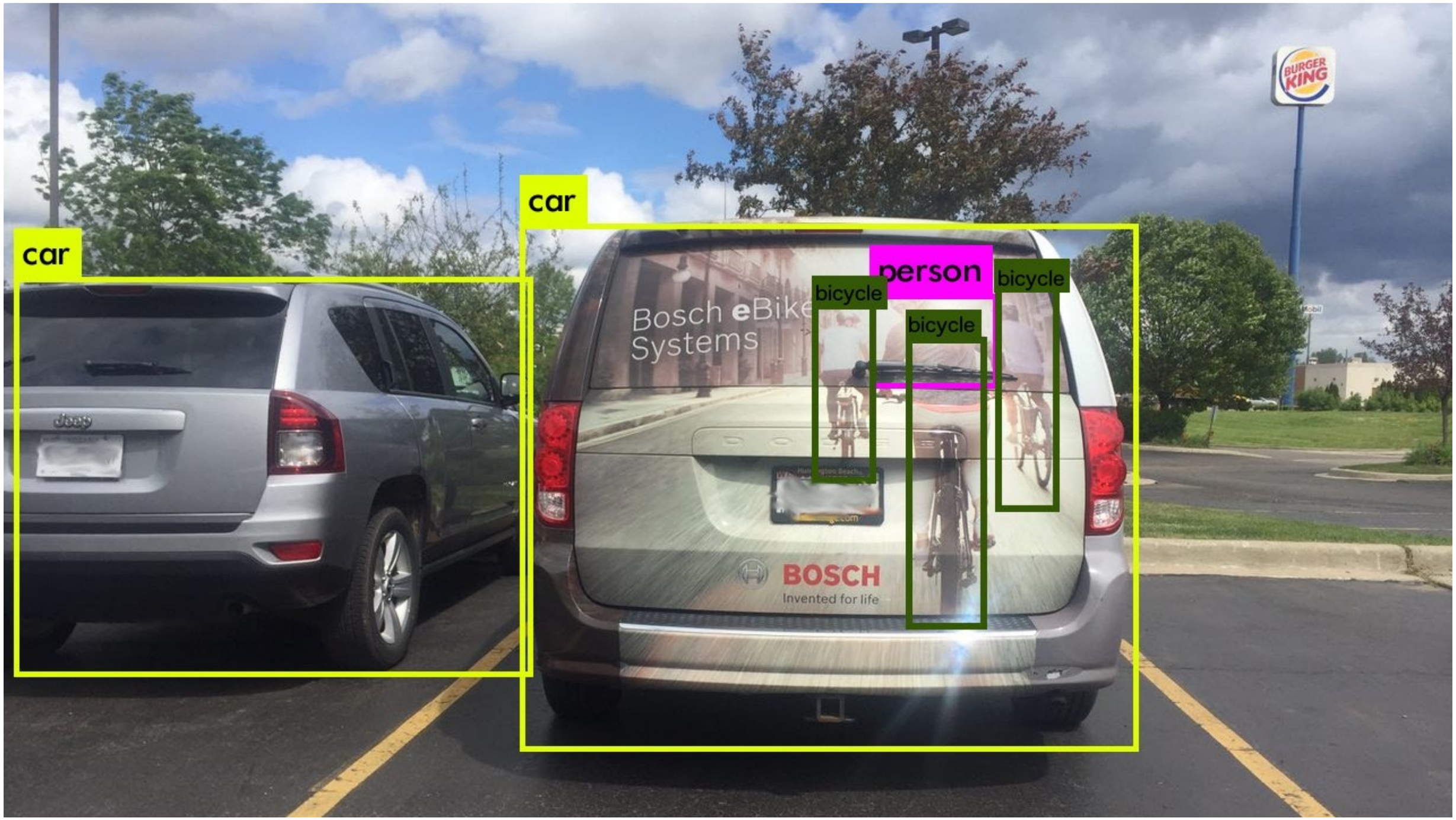
Why Kimera?

solving 2D semantic segmentation failures:
2D semantic segmentation is doomed to fail…



Why Kimera?

solving 2D semantic segmentation failures:
2D semantic segmentation is doomed to fail…



Why Kimera?

solving 3D reconstruction failures

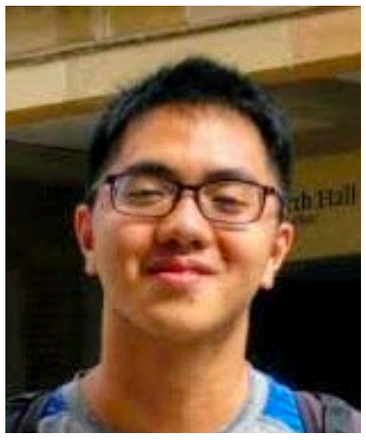
Conclusion

- **Visual-inertial navigation: a mature technology**
  - preintegration = accurate & fast
  - enabler for robotics applications and beyond
- **High-level understanding: key to many applications**
  - Spatial Perception, Spatial AI (A. Davison)
  - a lot of work to be done
  - opportunities to bridge learning and geometry

Thank you!

Antoni (Toni) Rosinol

Marcus Abate



Yun Chang

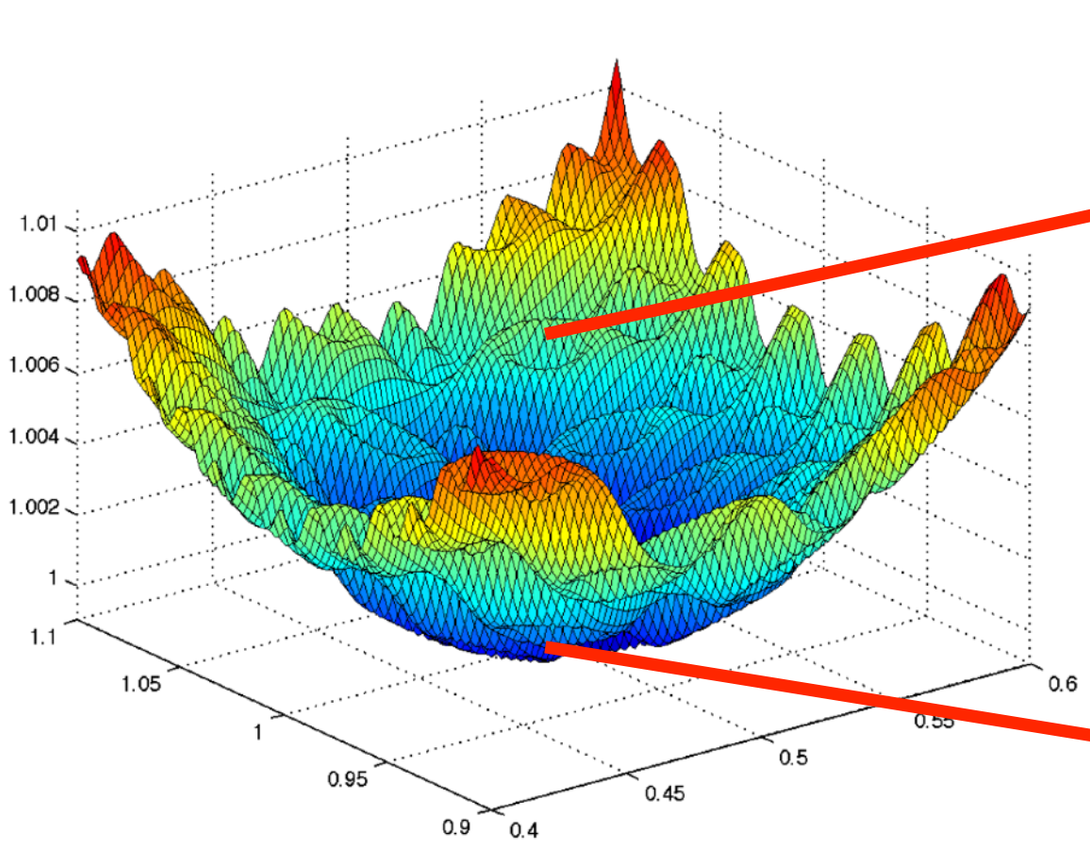
Kimera source code available at:

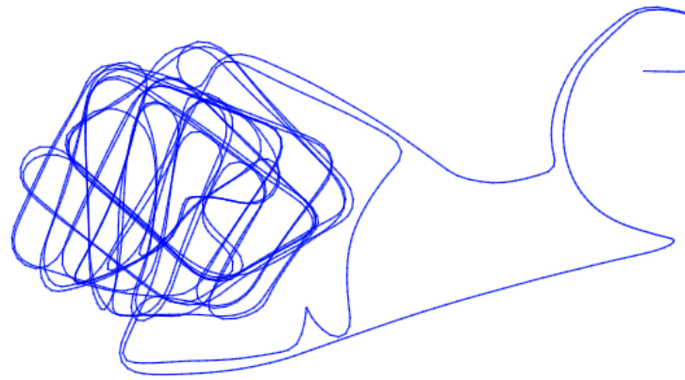
https://github.com/MIT-SPARK/Kimera

MIT LINCOLN LABORATORY ARL

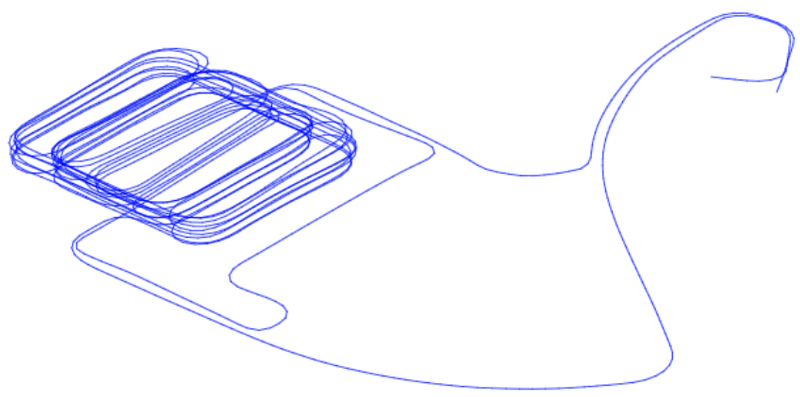
(Some) Next Steps for Kimera

$$\min_{\substack{\{p_i \in \mathbb{R}^3\} \\ \{R_i \in \mathrm{SO}(3)\}}} \sum_{(i,j)\in\mathcal{E}} \frac{1}{\sigma_p^2} \left\| \bar{p}_{ij} - R_i^{\mathsf{T}}(p_j - p_i) \right\|^2 + \frac{1}{\sigma_R^2} \left\| \bar{R}_{ij} - R_i^{\mathsf{T}} R_j \right\|_{\mathbf{F}}^2$$





Suboptimal critical point



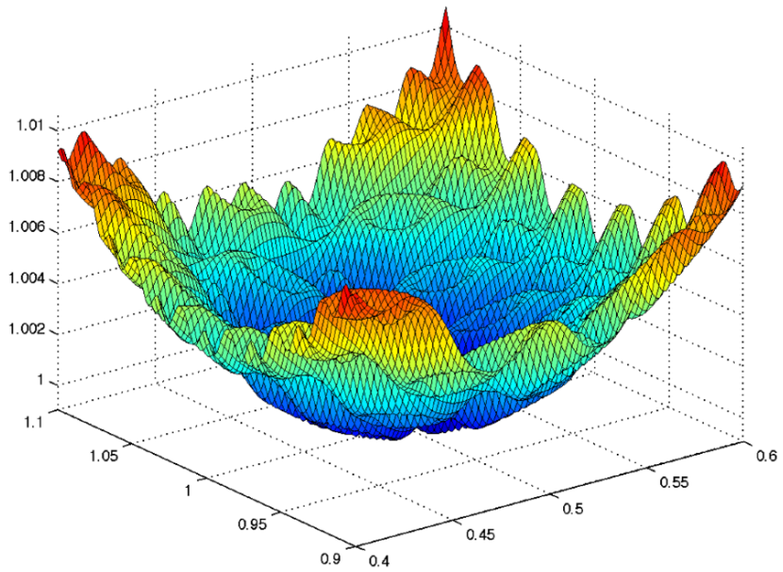
Optimal estimate

(Some) Next Steps for Kimera

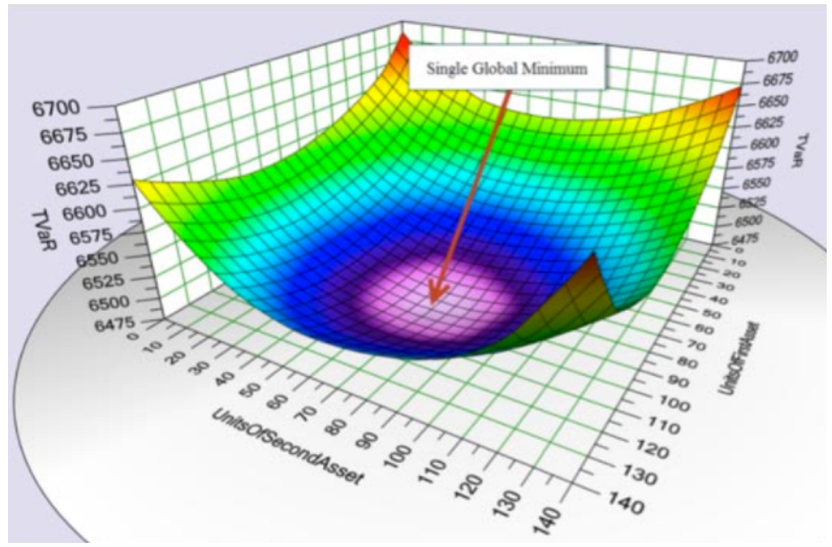
SE-sync: fast globally optimal SLAM

SE-sync is exact

non convex problem



convex relaxation



SE-sync is fast

	# Poses	# Edges	Gauss-Newton Objective value	Gauss-Newton Time [s]	SE-Sync Objective value	SE-Sync Time [s]	SE-Sync Max. suboptimality
sphere	2500	4949	1.687×10^3	14.98	1.687×10^3	2.81	1.410×10^{-11}
torus	5000	9048	2.423×10^4	31.94	2.423×10^4	5.67	7.276×10^{-12}
grid	8000	22236	8.432×10^4	130.35	8.432×10^4	22.37	4.366×10^{-11}
garage	1661	6275	1.263×10^0	17.81	1.263×10^0	5.33	2.097×10^{-11}
cubicle	5750	16869	7.171×10^2	136.86	7.171×10^2	13.08	1.603×10^{-11}
rim	10195	29743	5.461×10^3	575.42	5.461×10^3	36.66	5.639×10^{-11}

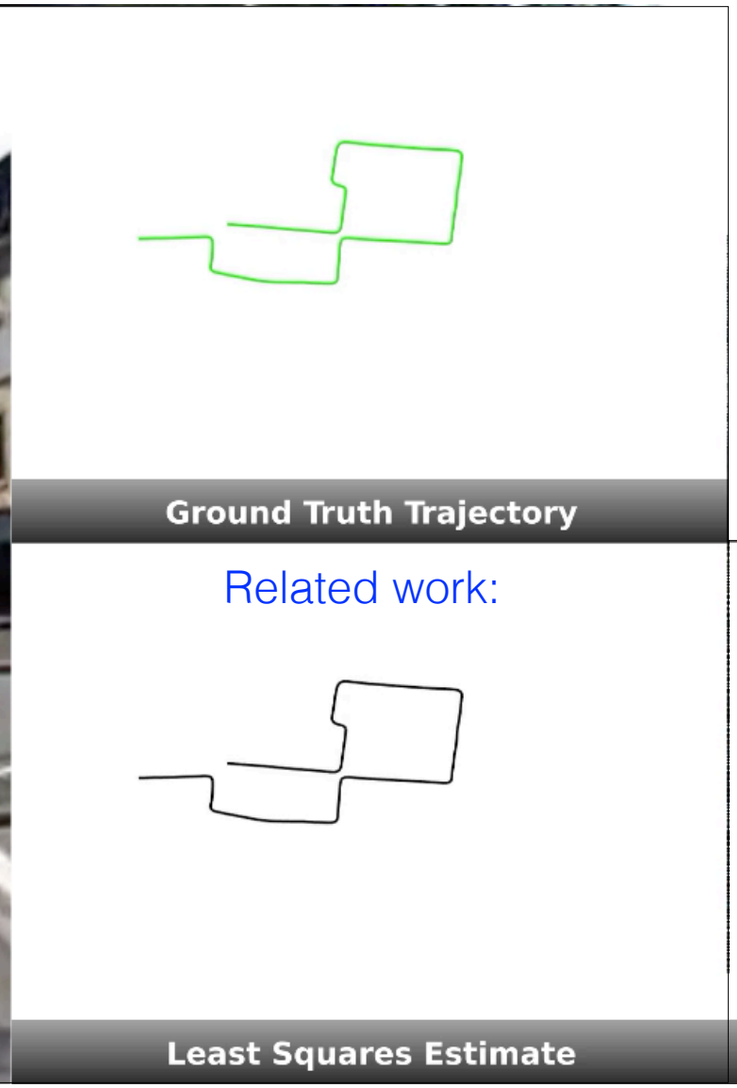
Iterative optimization

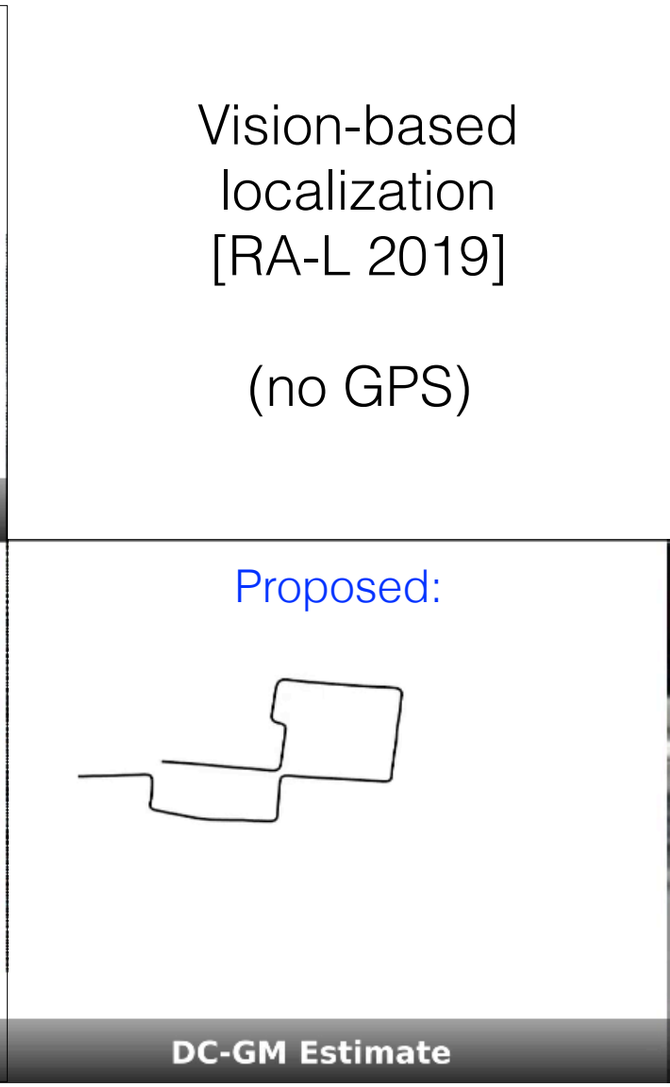
Proposed convex relaxation

Rosen, Carlone, Bandeira, Leonard, SE-Sync: A Certifiably Correct Algorithm for Synchronization over the Special Euclidean Group, WAFR, 2016 (best paper award)

(Some) Next Steps for Kimera

Certifiable robust SLAM!





Vision-based localization [RA-L 2019]

(no GPS)

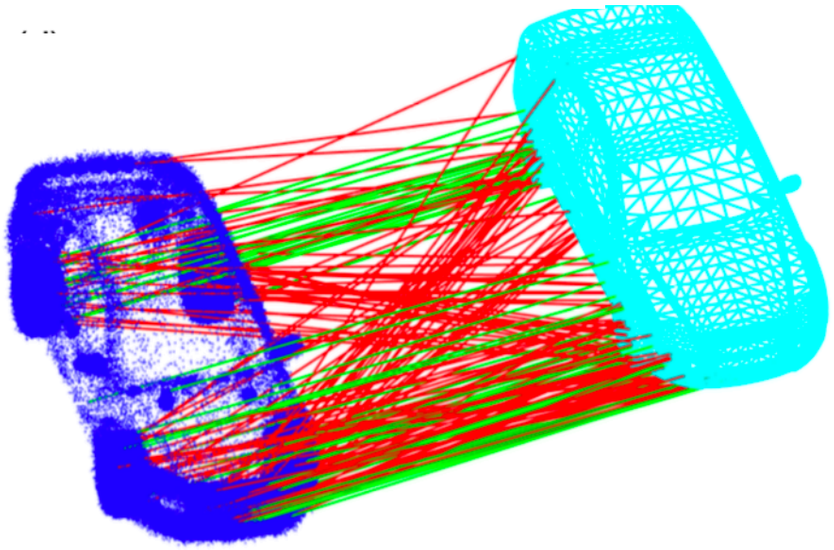
We use novel mathematical tools (e.g., convex relaxations) to develop perception algorithms that are “hard to break”:
operating correctly under extreme noise and outliers

Lajoie, Hu, Beltrame, Carlone, Modeling Perceptual Aliasing in SLAM via Discrete-Continuous Graphical Models, RAL’19.

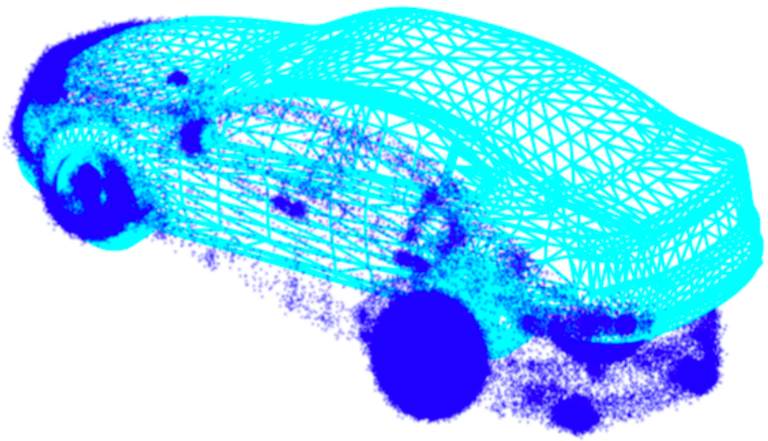
(Some) Next Steps for Kimera

Object detection and localization

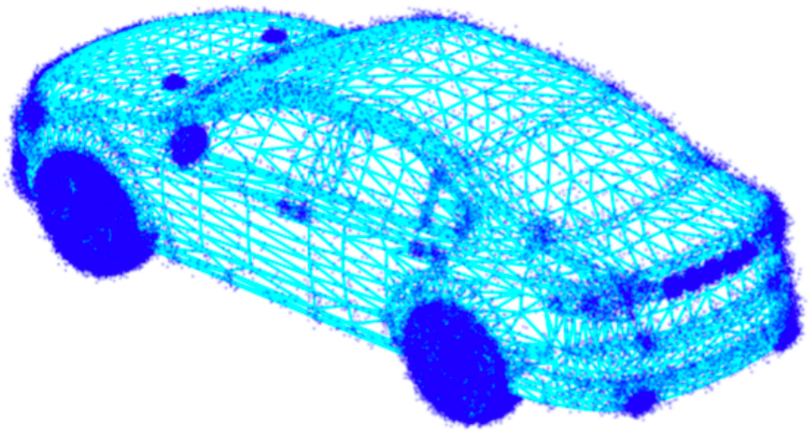
Correspondences



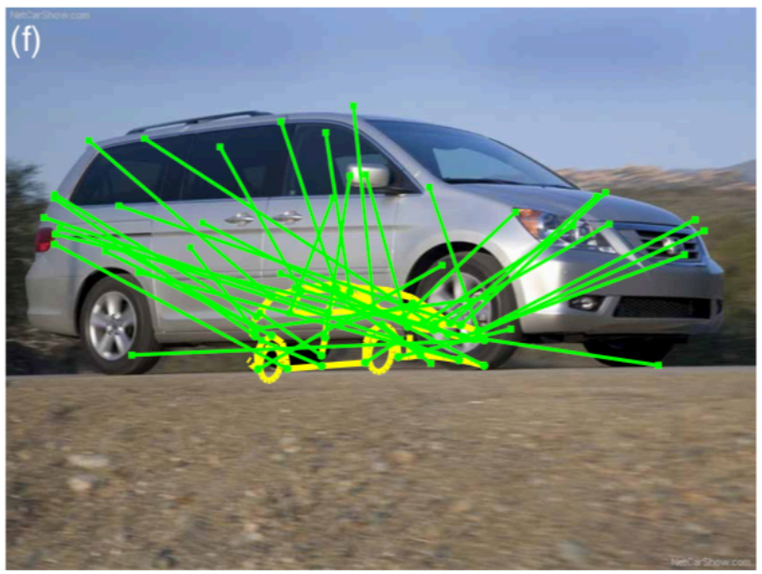
RANSAC



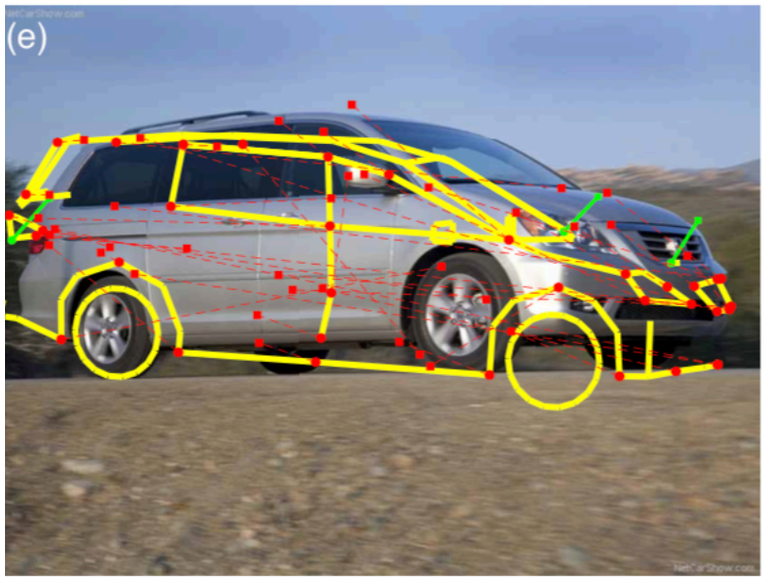
Proposed



[Zhou, CVPR'15]



RANSAC



Proposed

H. Yang, P. Antonante, V. Tzoumas, L. Carlone. Graduated non-convexity for robust spatial perception: From non-minimal solvers to global outlier rejection. Arxiv, 2019.