It kinda works!*

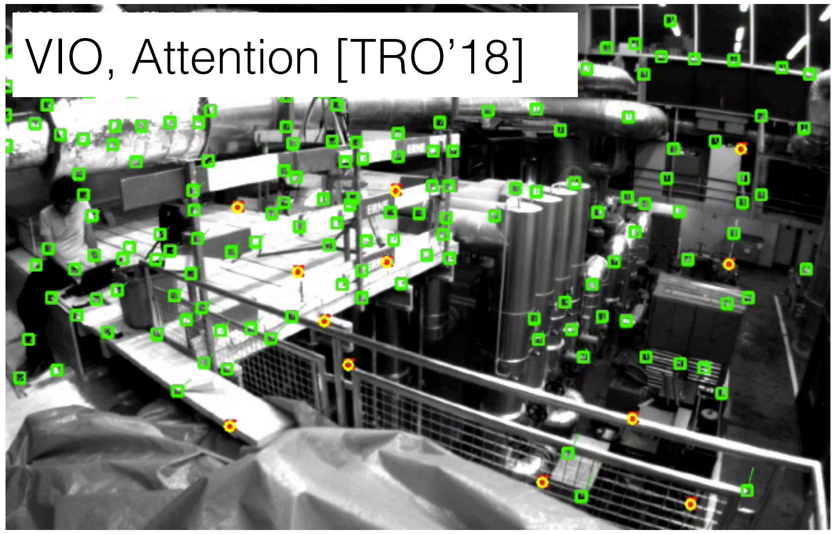
Challenges and Opportunities for Robot Perception in the Deep Learning Era

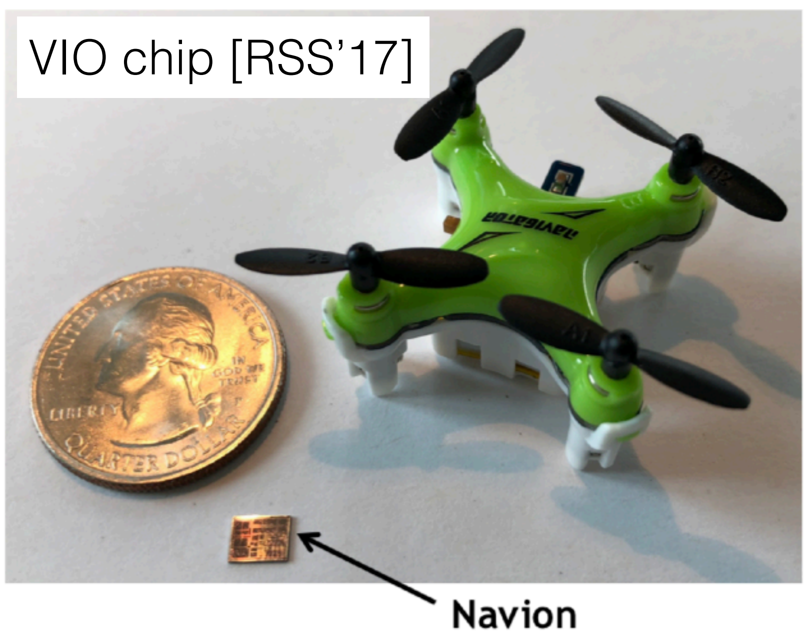
Luca Carlone
Charles Stark Draper Assistant Professor
Massachusetts Institute of Technology

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

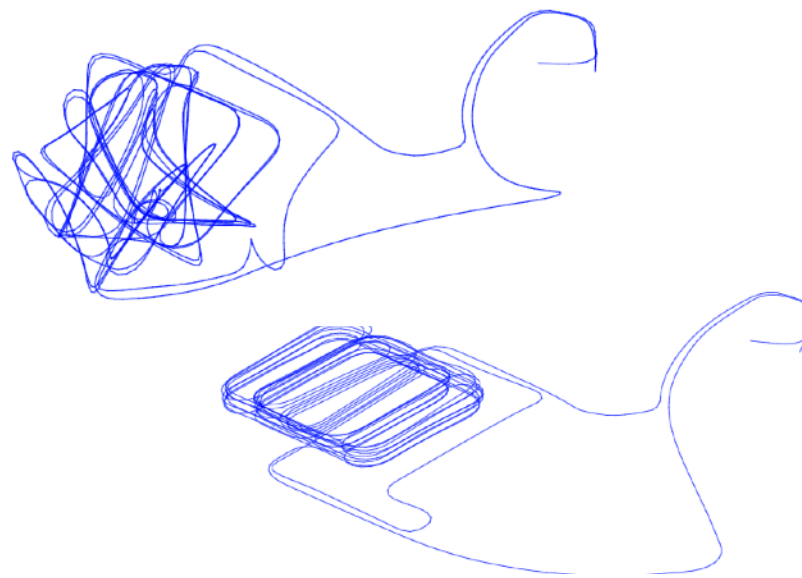
SPARK Lab

Efficiency

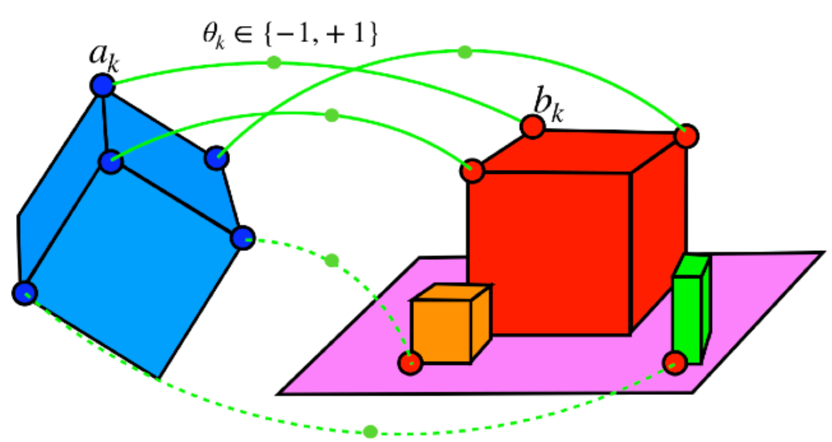




Robustness

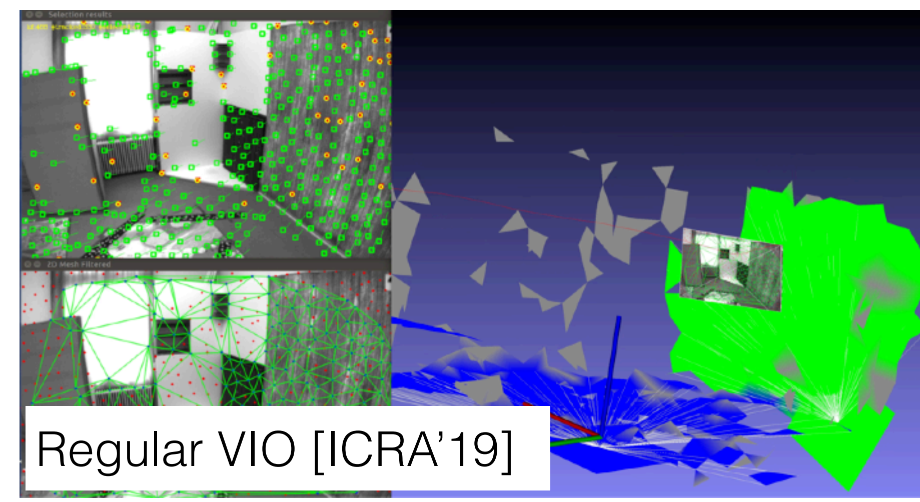


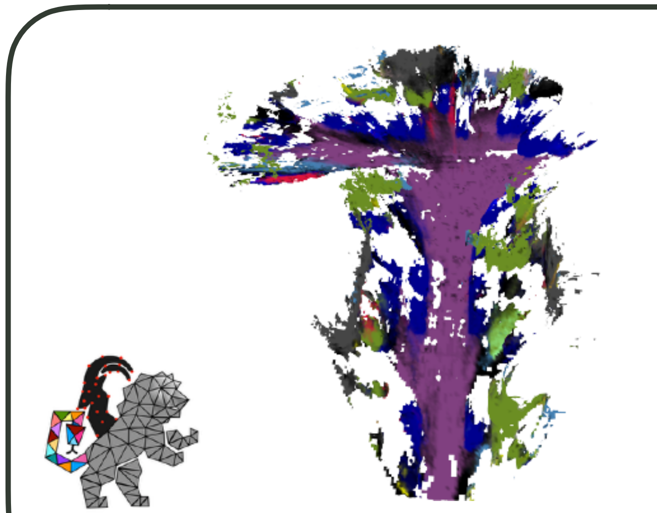
Pose graph optimization: duality [T-RO'15], SE-sync [IJRR'18]



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

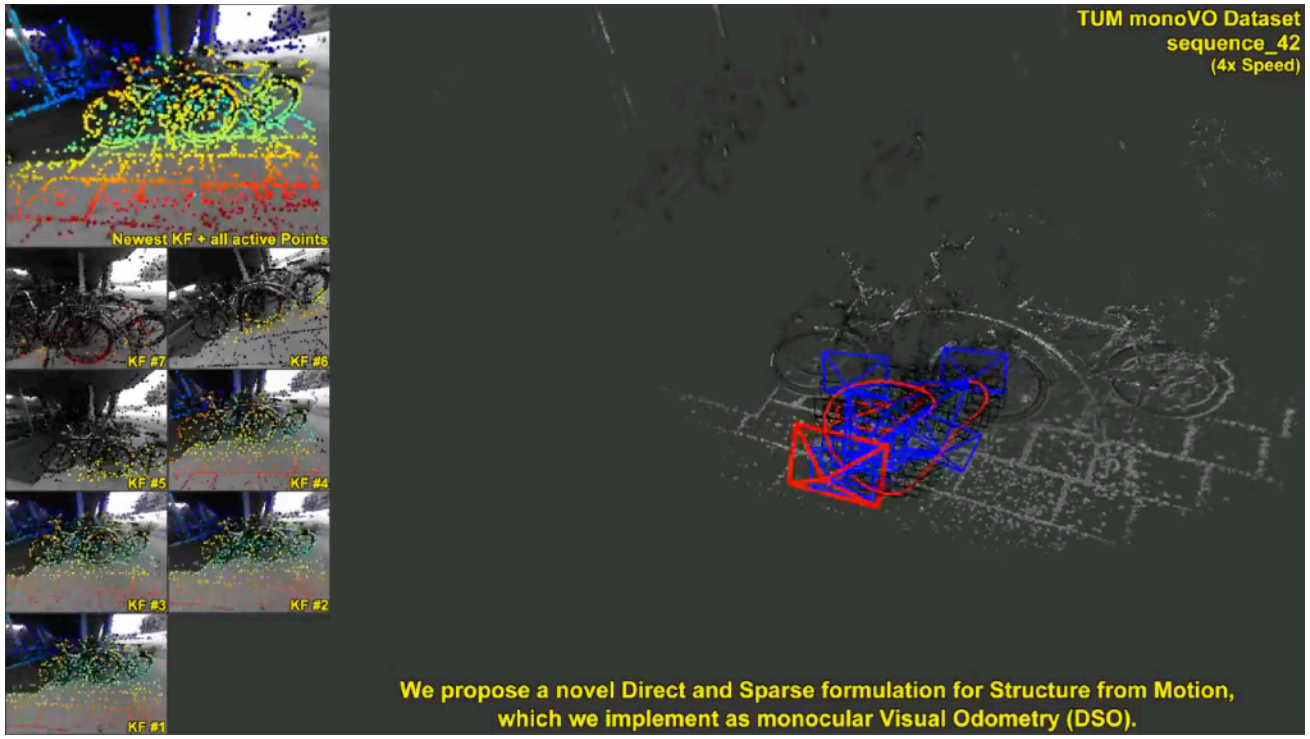
High-level understanding





Kimera [ICRA'19-20]

Kinda of Works!



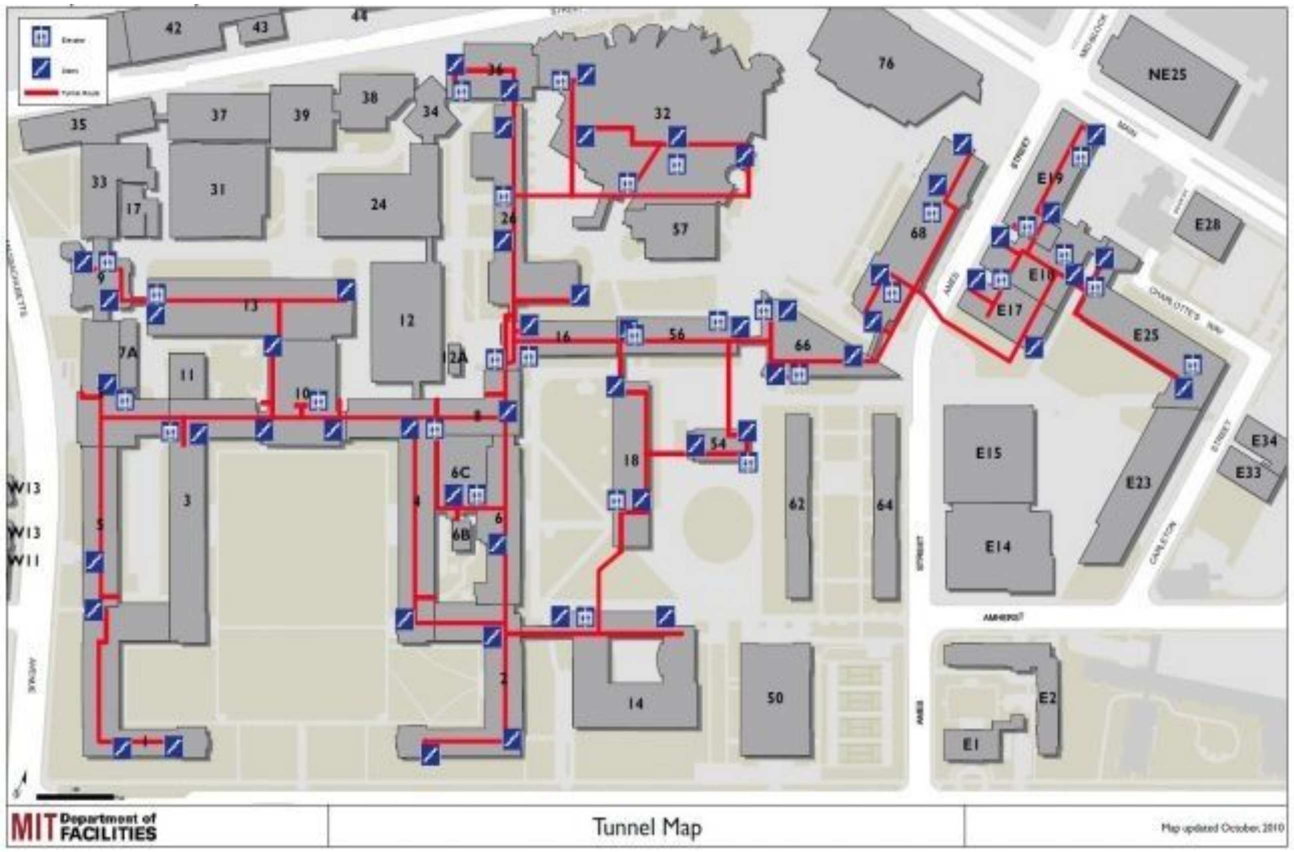
[Visual SLAM: ORB-SLAM, DSO…]

[Lidar SLAM: LOAM, …]

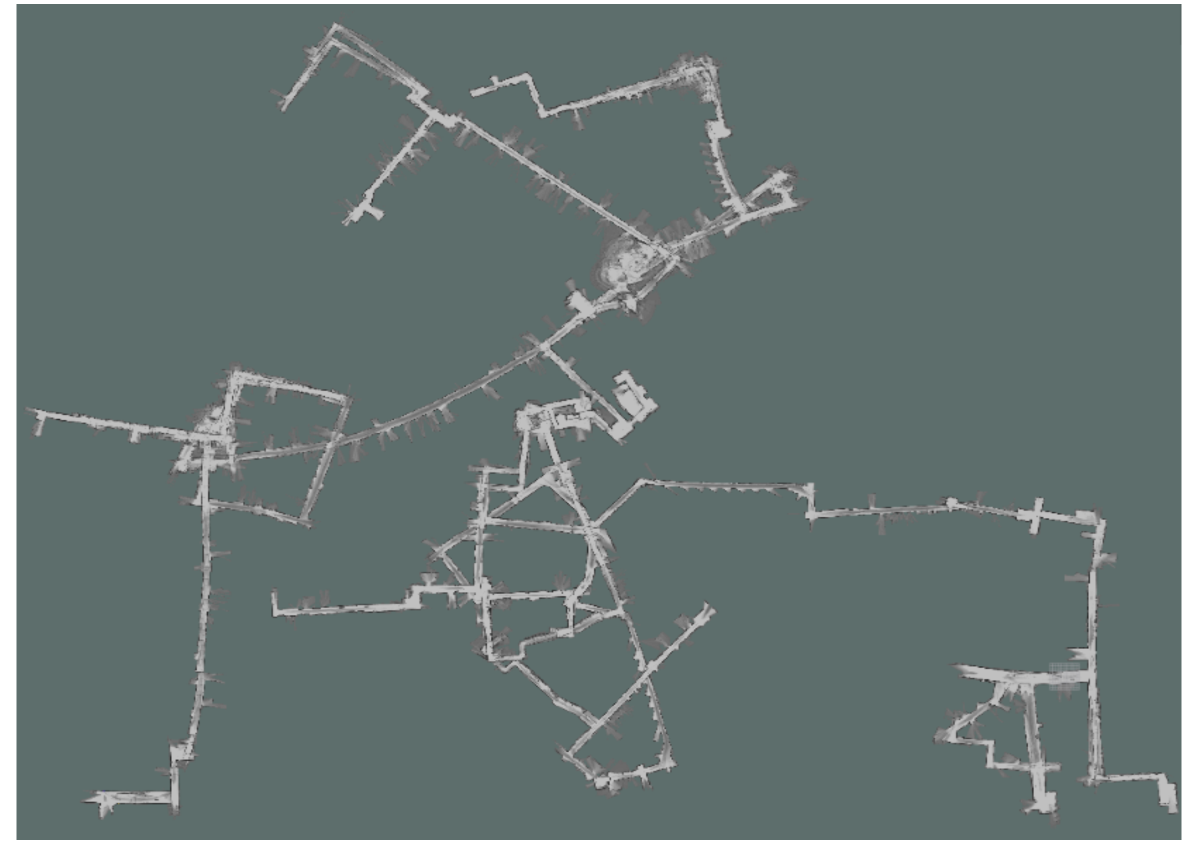
[2D object detection: YOLO, …]

[2D semantic segmentation]

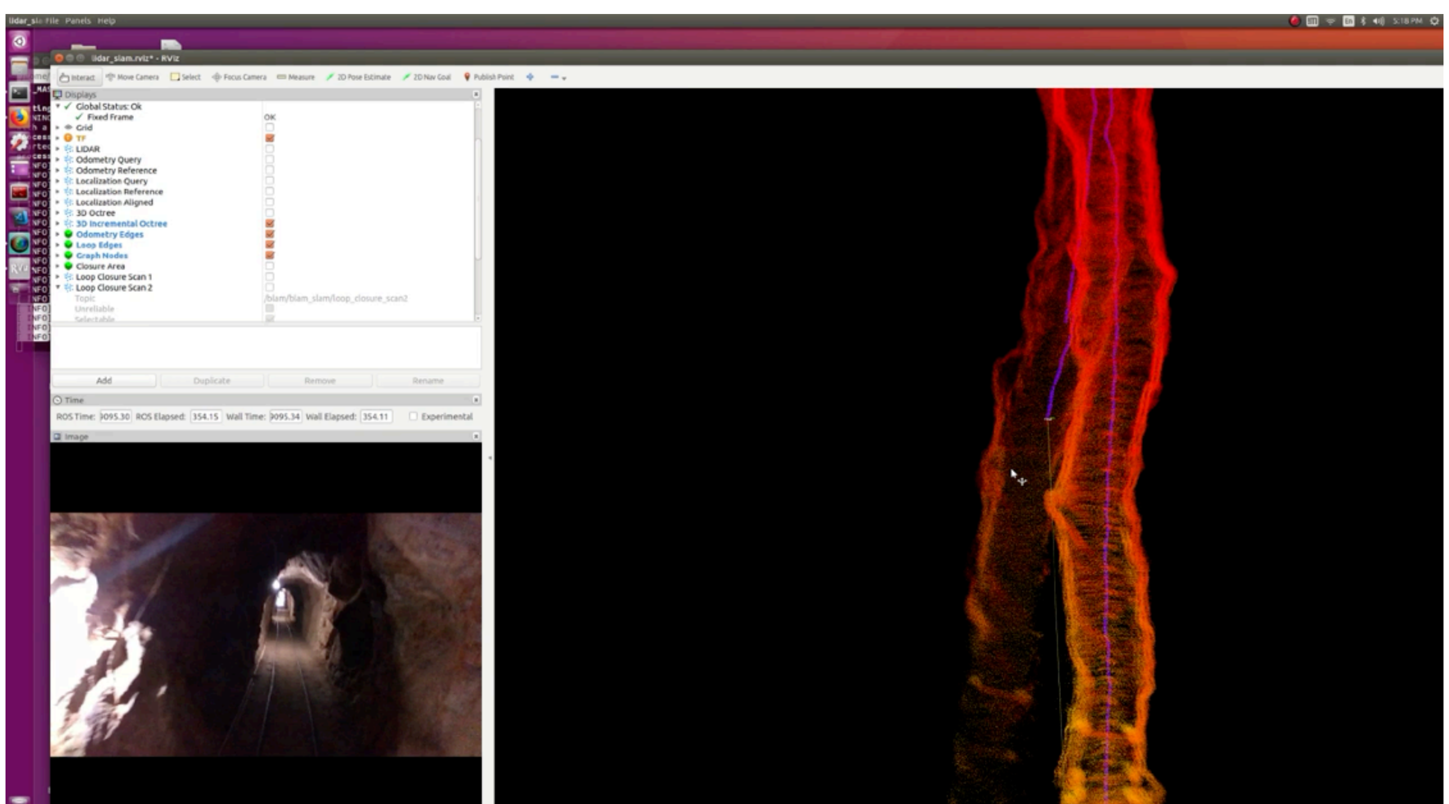
Kinda of Works…



MIT Tunnel System



Google Cartographer



Off-the-shelf libraries are still brittle and require careful data collection and parameter tuning

- not enough for safety-critical applications
- involve extensive human supervision

Why Research on Perception?

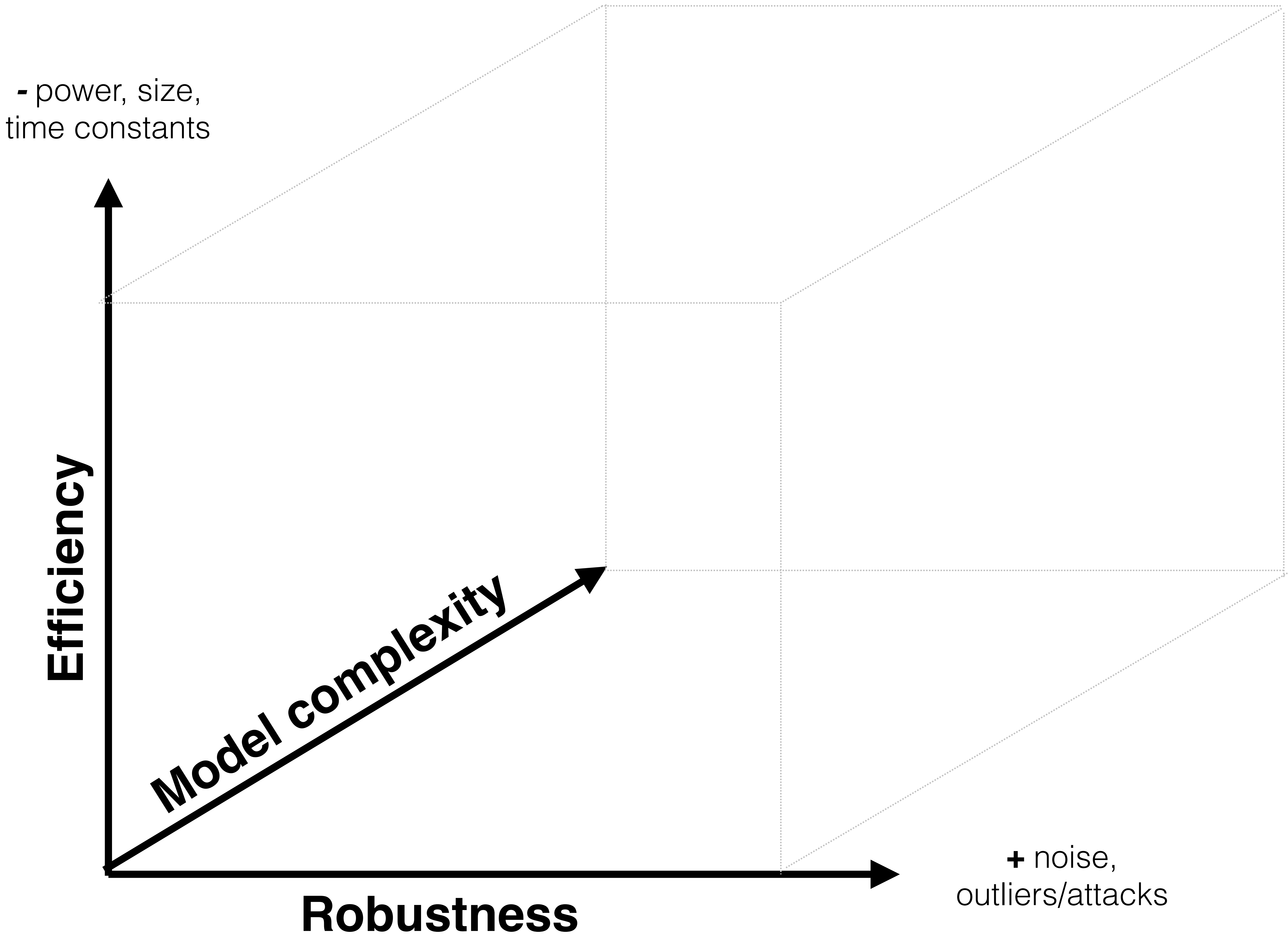
Perception success.. and its failures

Images: Evtimov et al

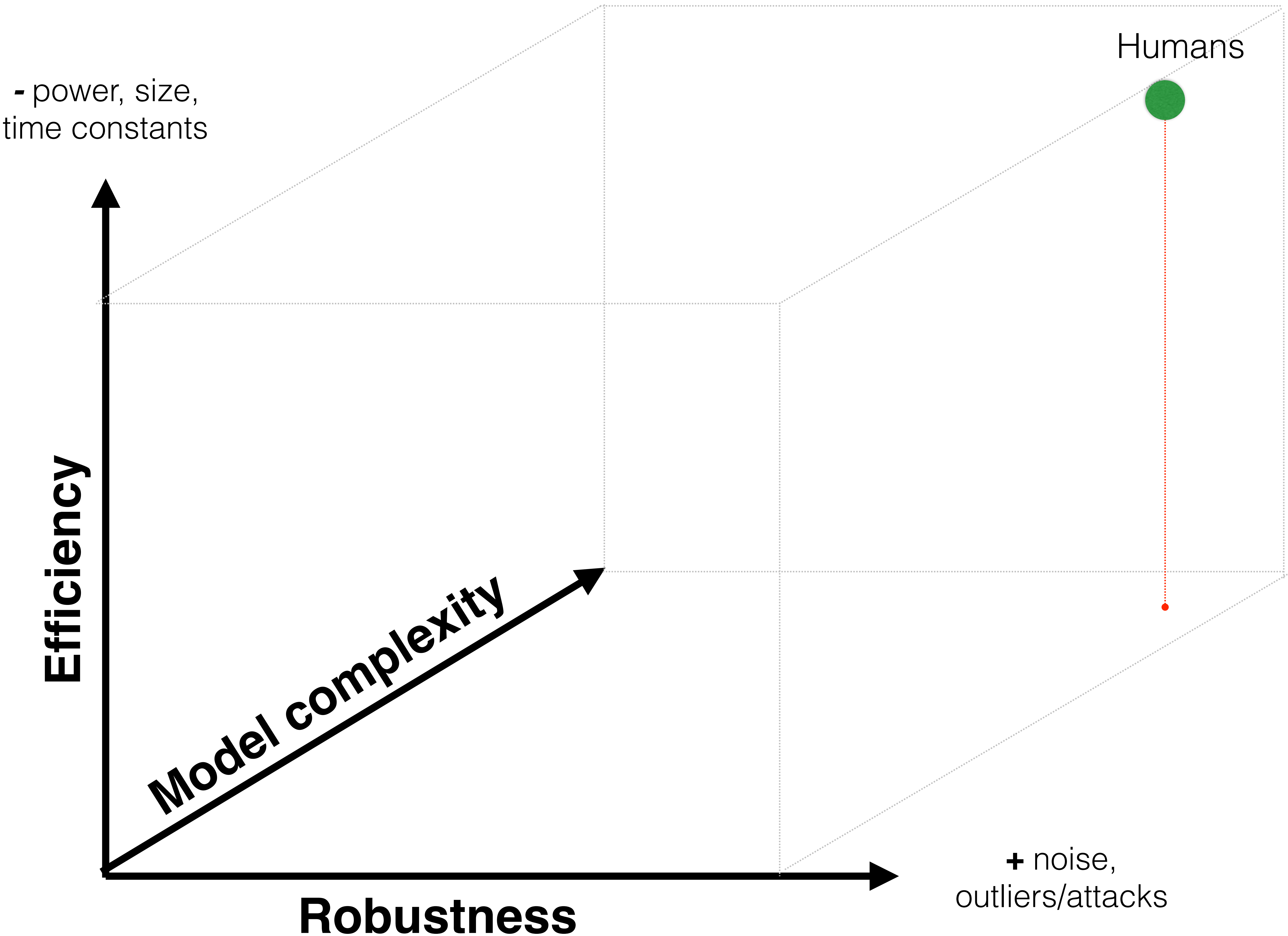
Camouflage graffiti and art stickers cause a neural network to misclassify stop signs as speed limit 45 signs or yield signs.

"Google employs a small army of human operators to manually check and correct the maps" [Wired]

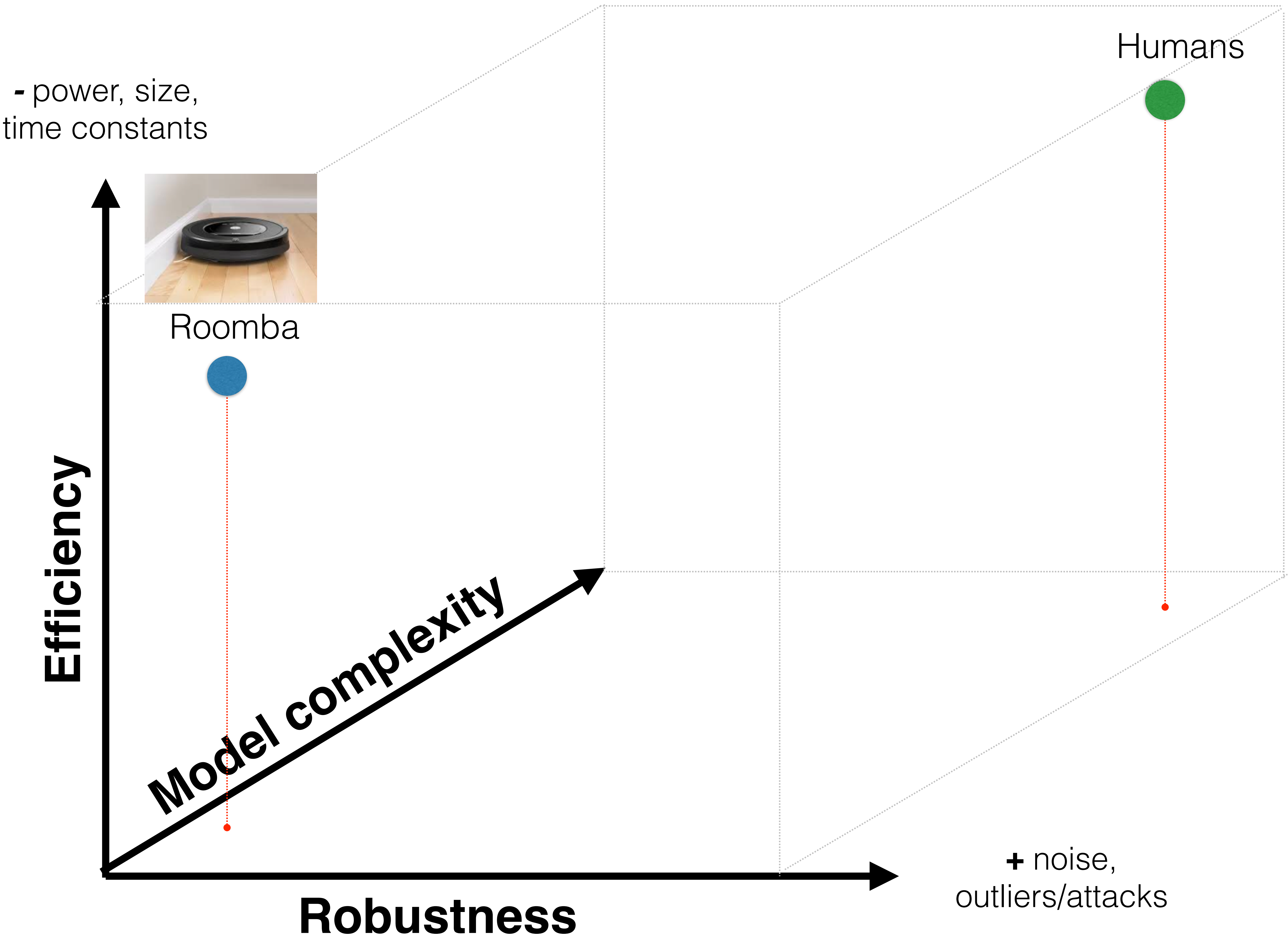
Axes of complexity



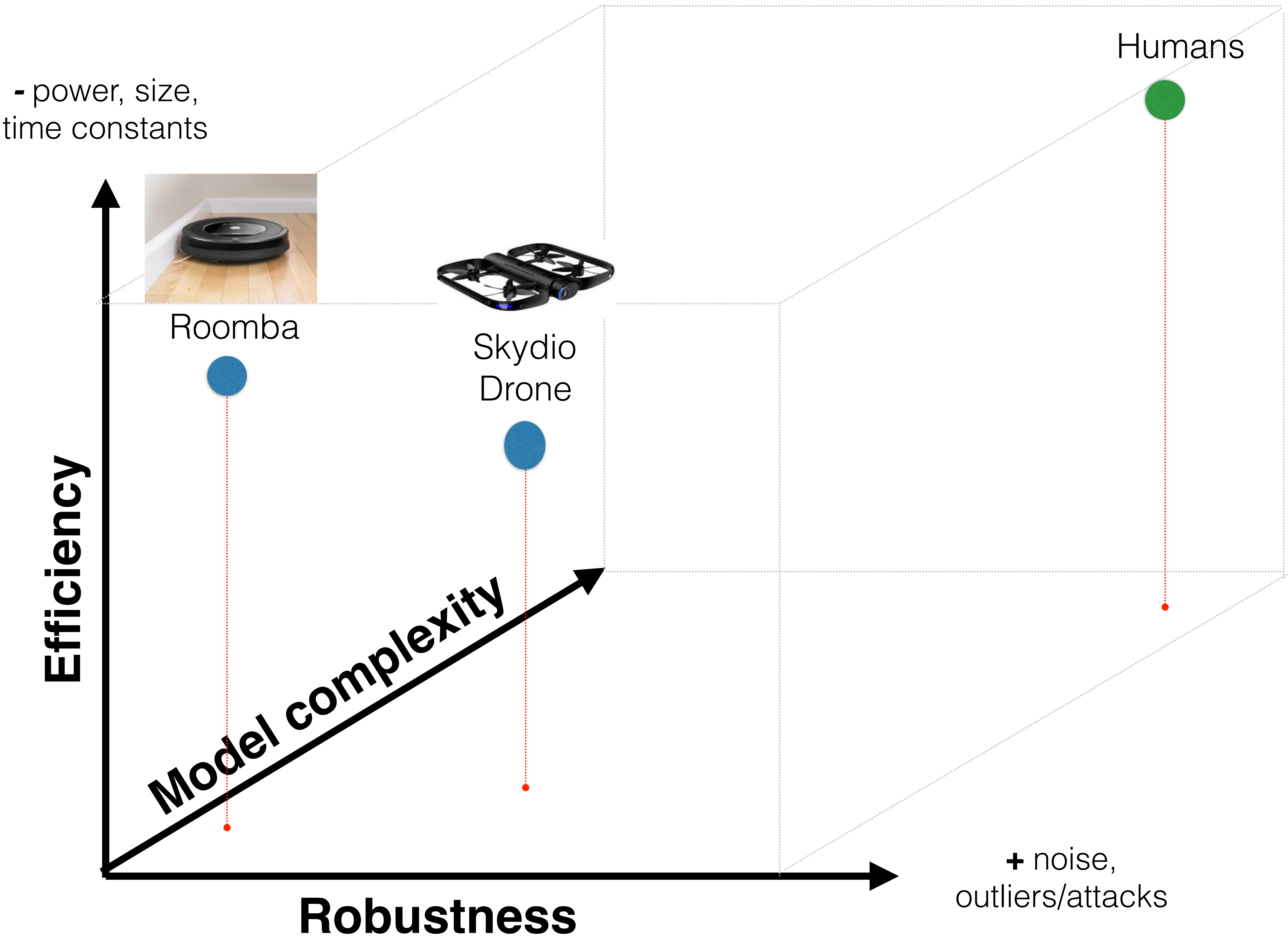
Axes of complexity



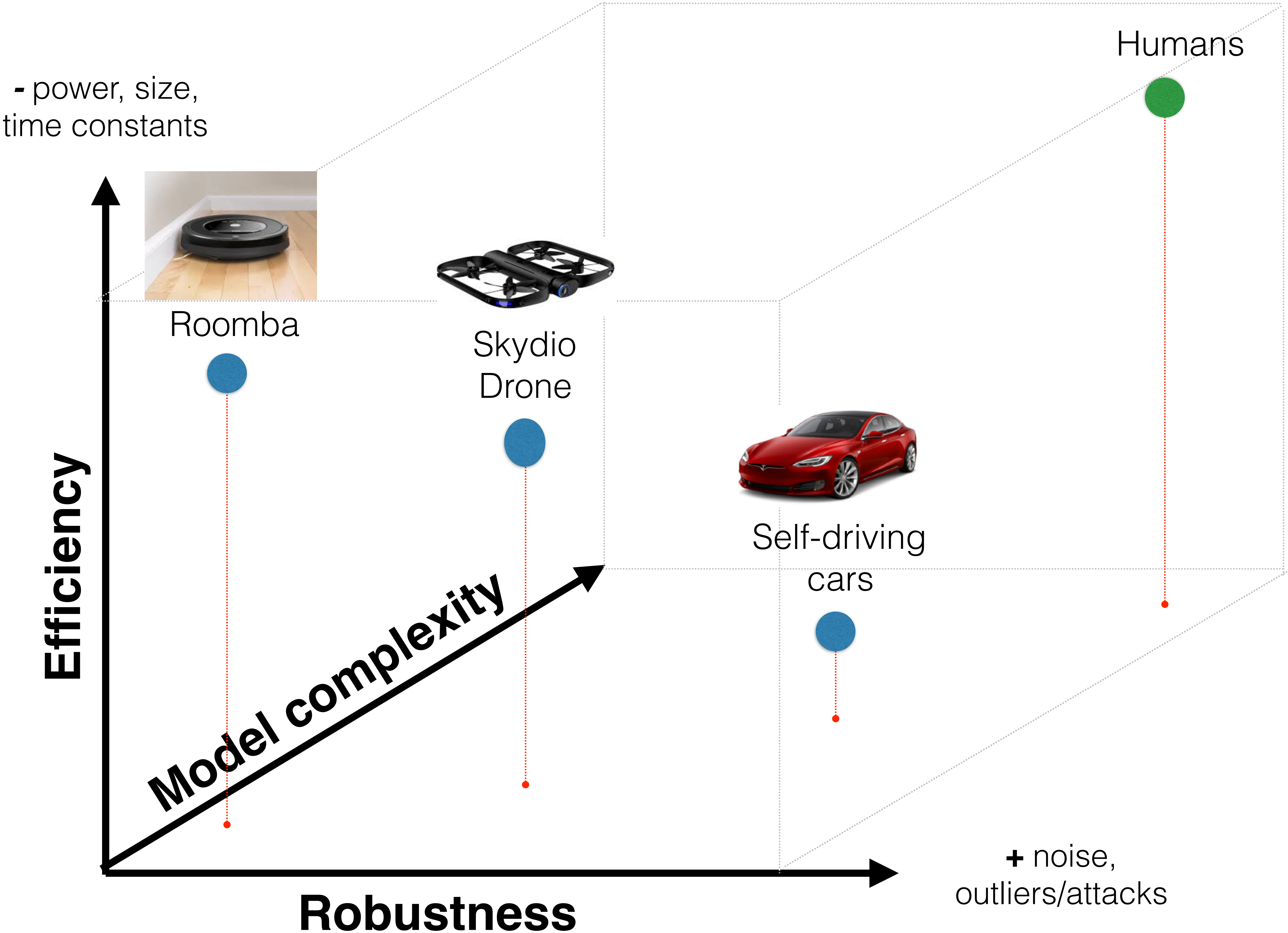
Axes of complexity



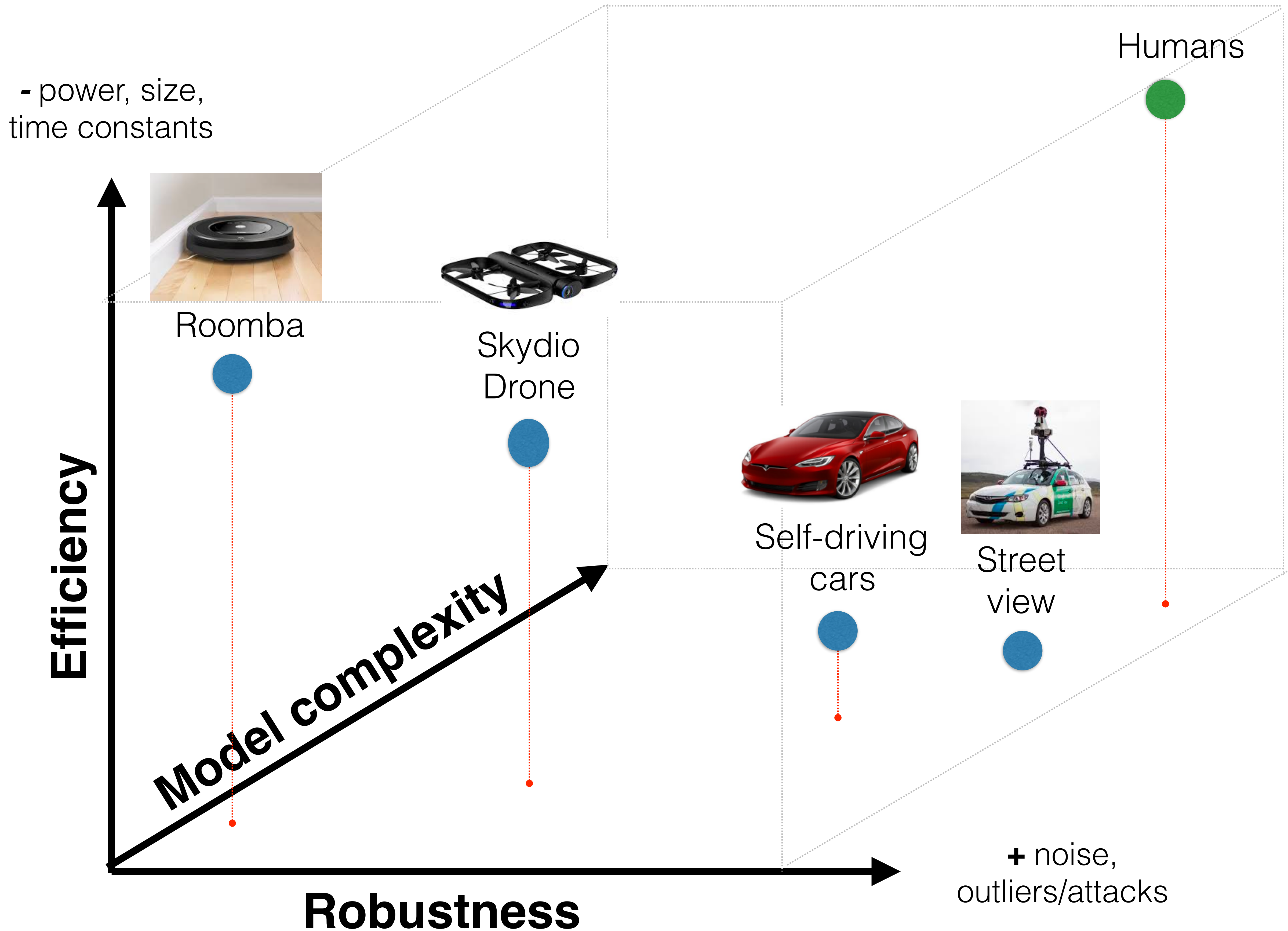
Axes of complexity



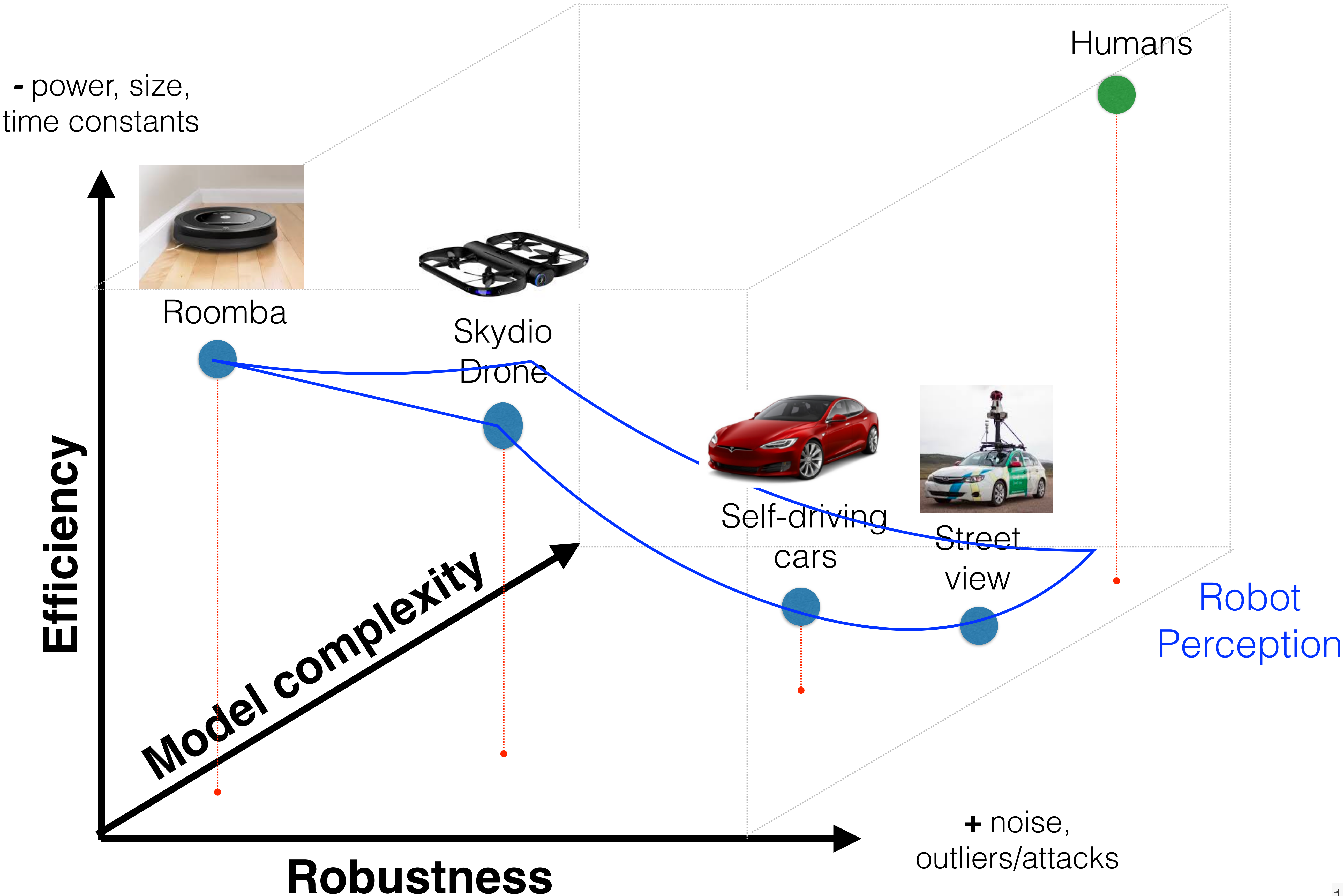
Axes of complexity



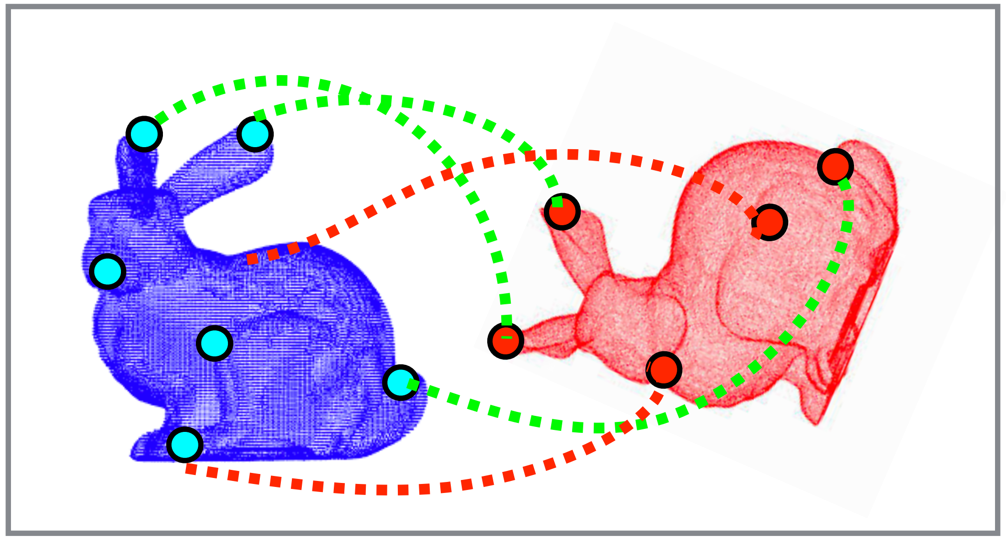
Axes of complexity



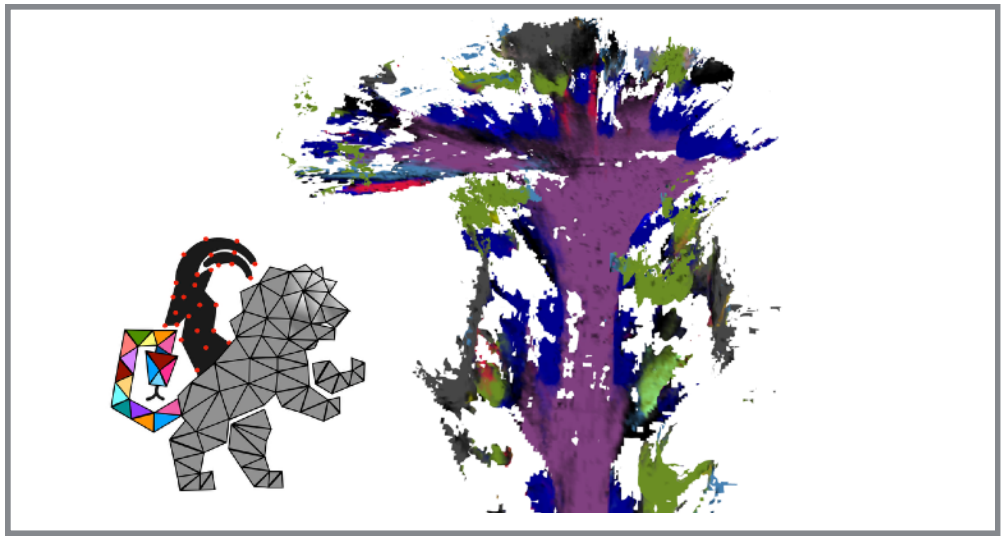
Axes of complexity



Outline



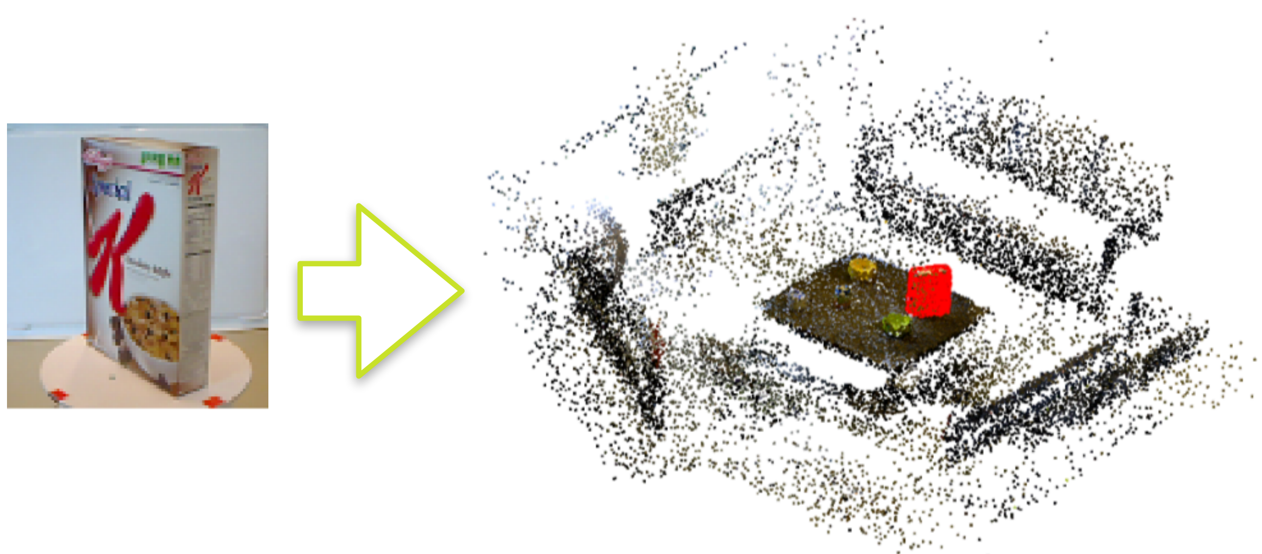
Certifiable Perception: algorithms that are "hard to break"

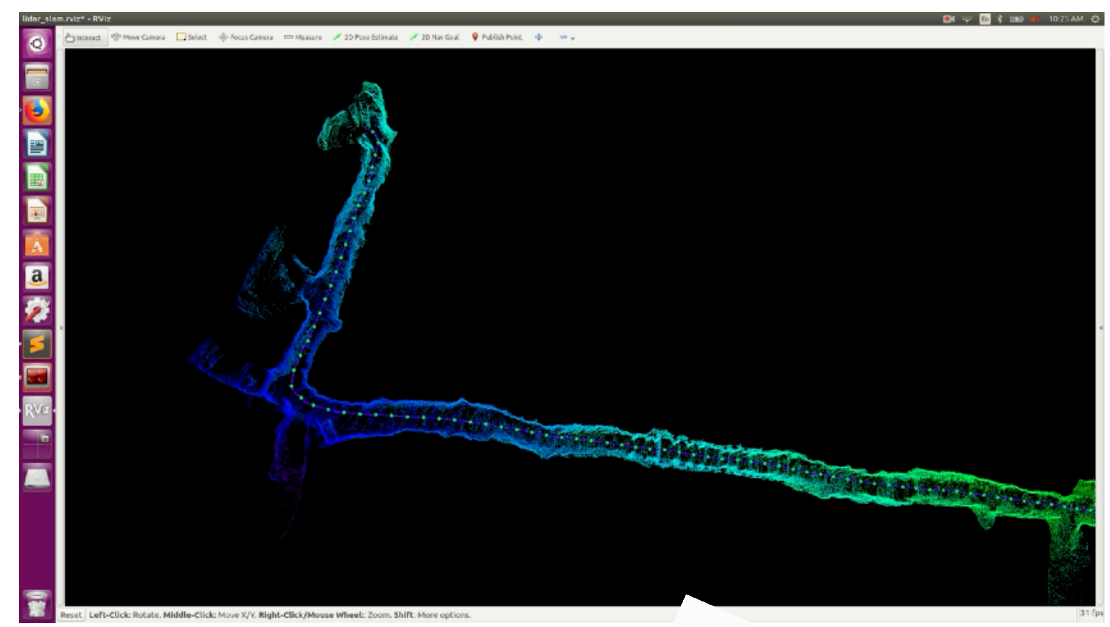


Kimera: real-time high-level understanding

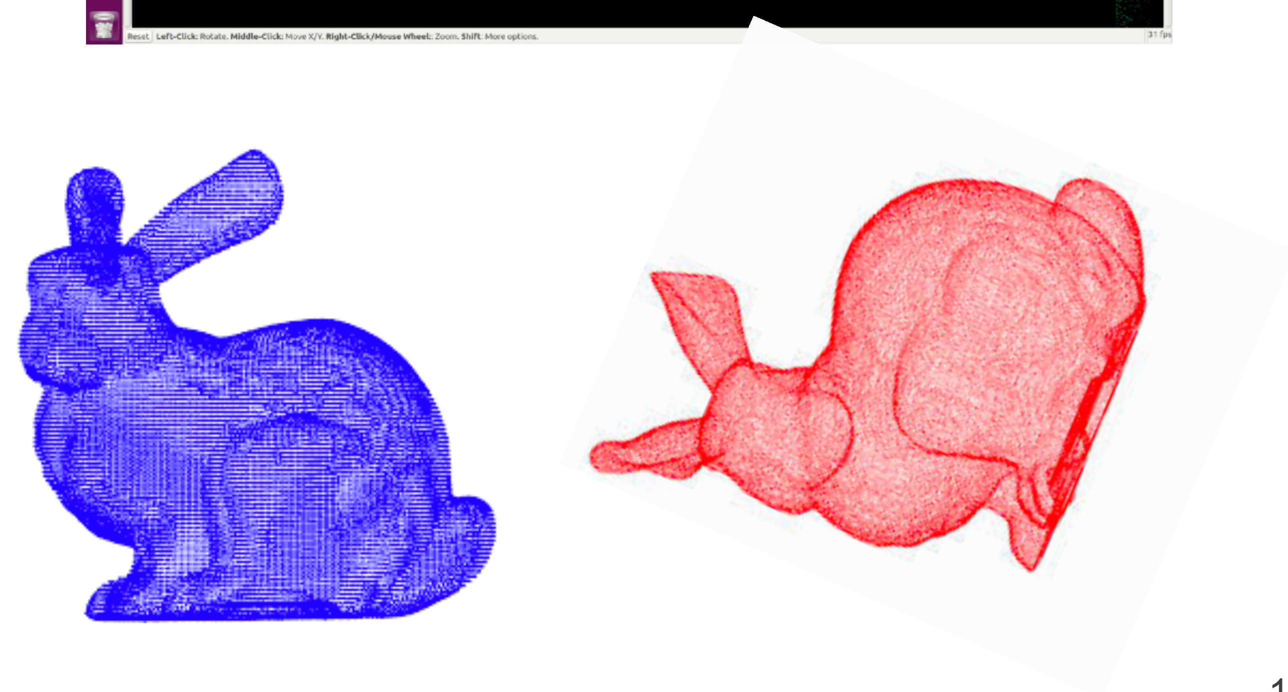
3D Registration (a.k.a. scan matching)

- **3D Registration problem**: find rigid transformation (position, rotation) that aligns two point clouds
  - Object pose estimation
  - Motion estimation (scan matching)



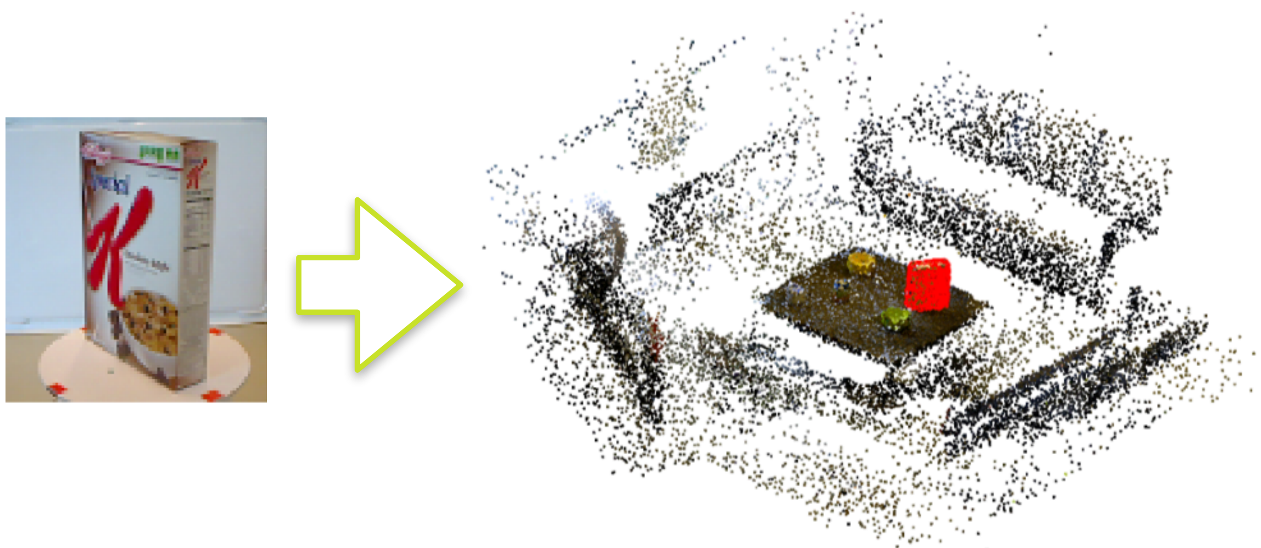


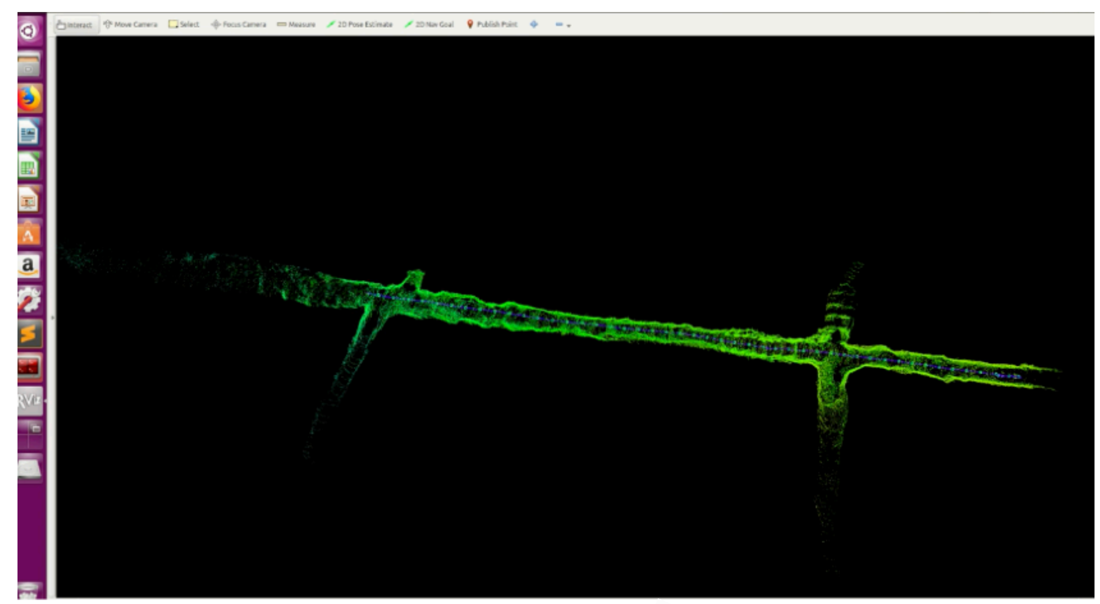
Typical registration procedure:
- extract features
- match “similar” features
- compute relative pose



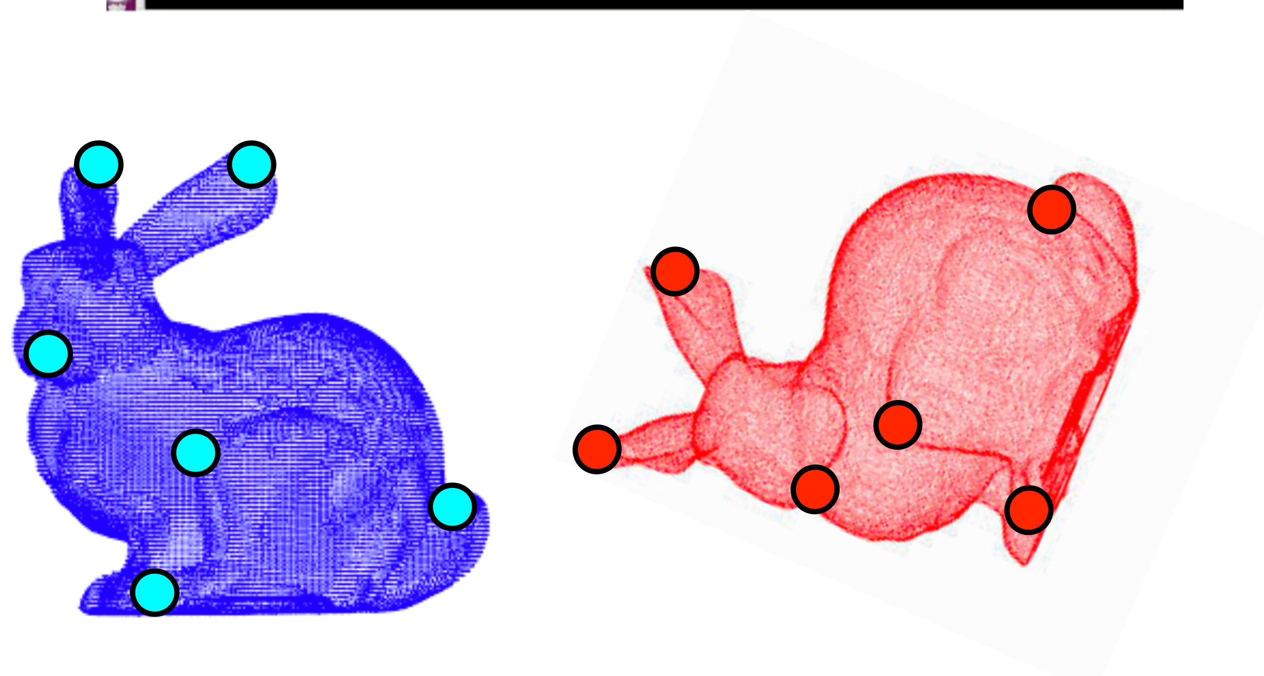
3D Registration (a.k.a. scan matching)

- **3D Registration problem**: find rigid transformation (position, rotation) that aligns two point clouds
  - Object pose estimation
  - Motion estimation (scan matching)



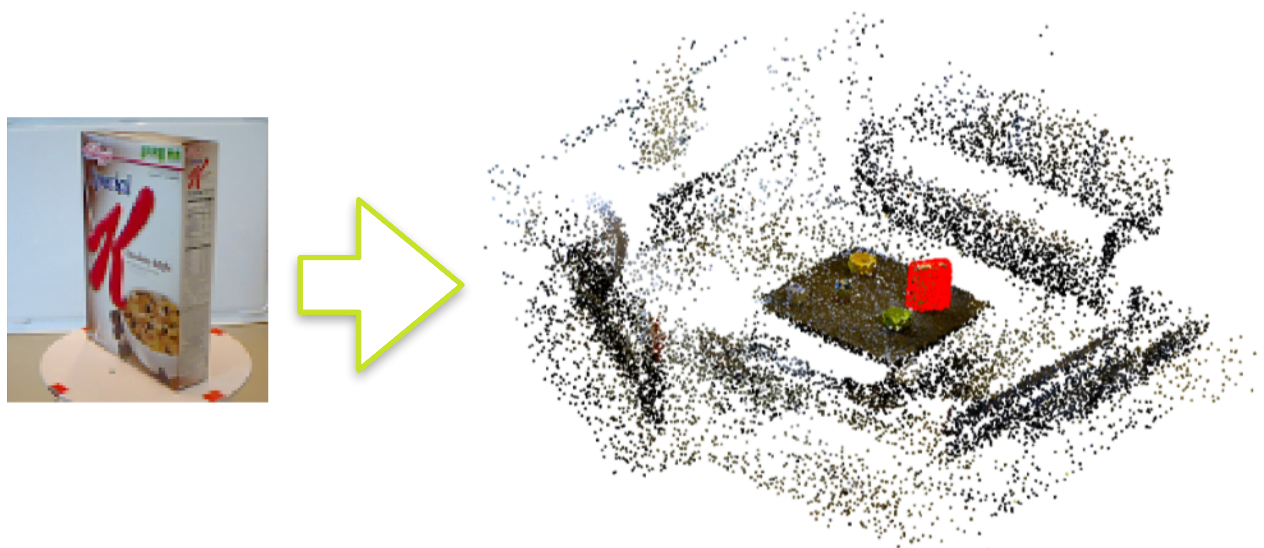


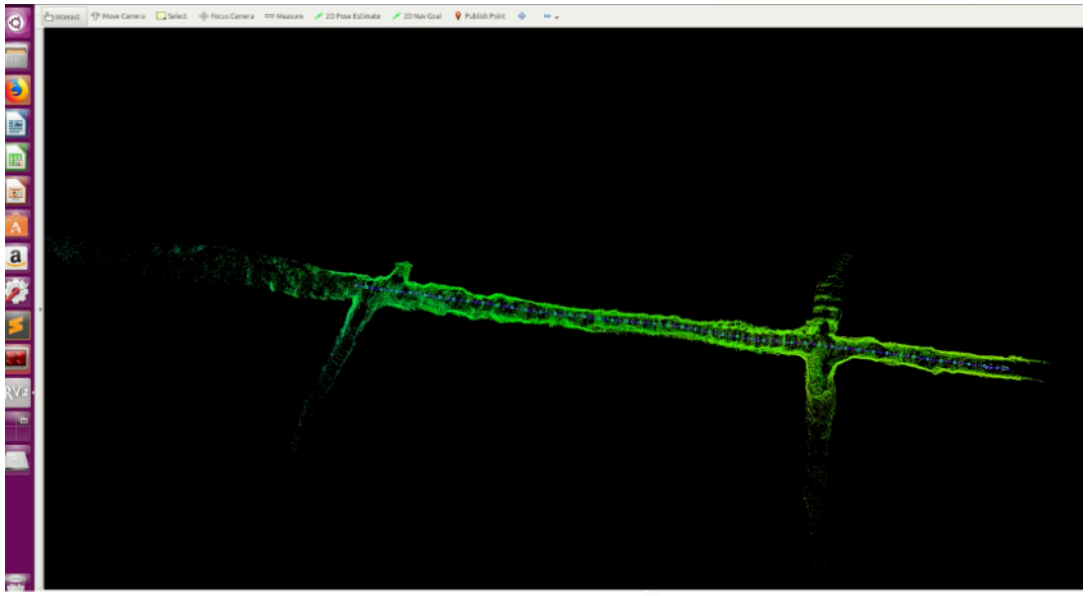
Typical registration procedure:
- extract features
- match "similar" features
- compute relative pose



3D Registration (a.k.a. scan matching)

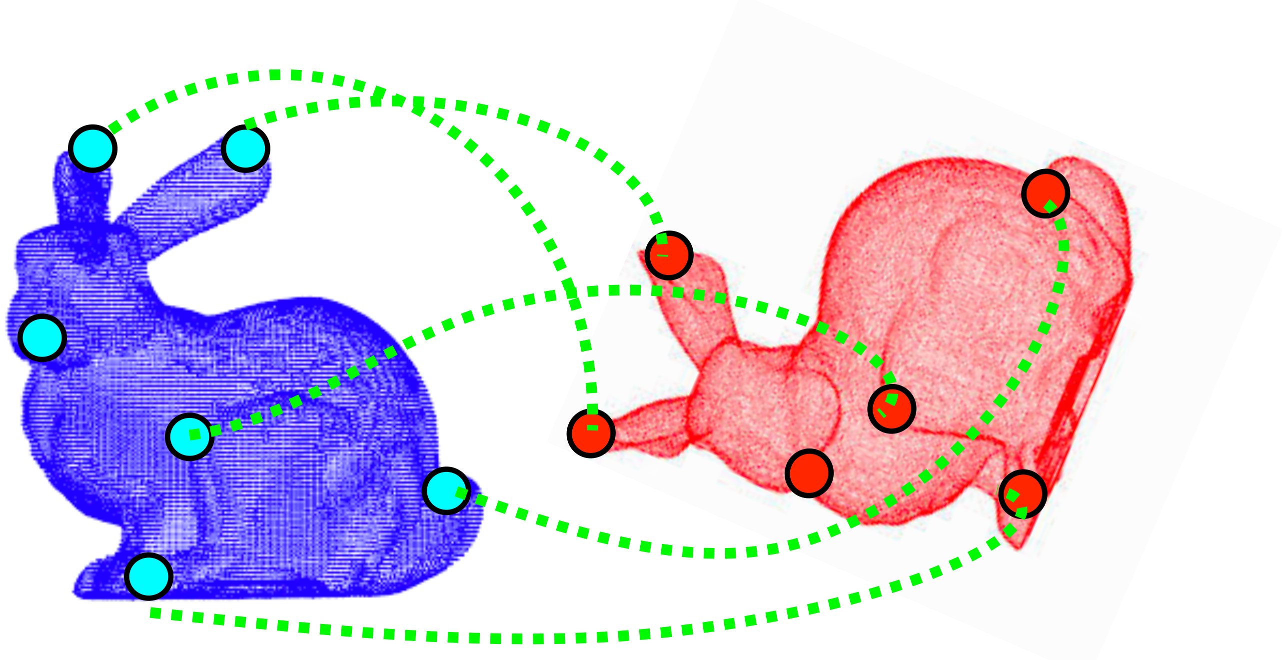
- **3D Registration problem**: find rigid transformation (position, rotation) that aligns two point clouds
  - Object pose estimation
  - Motion estimation (scan matching)





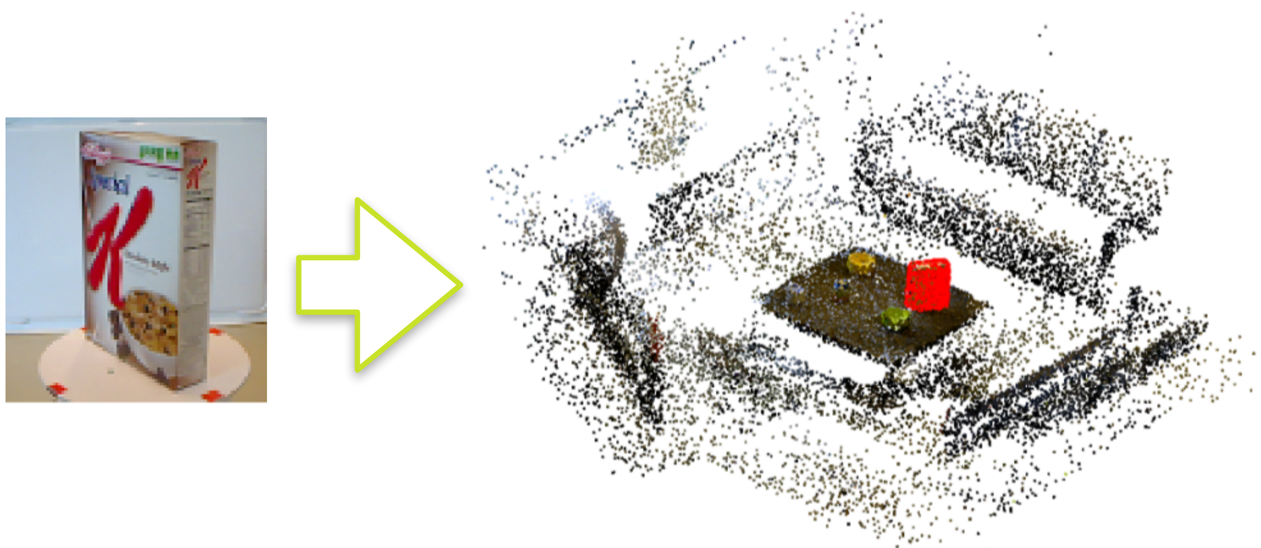
Typical registration procedure:

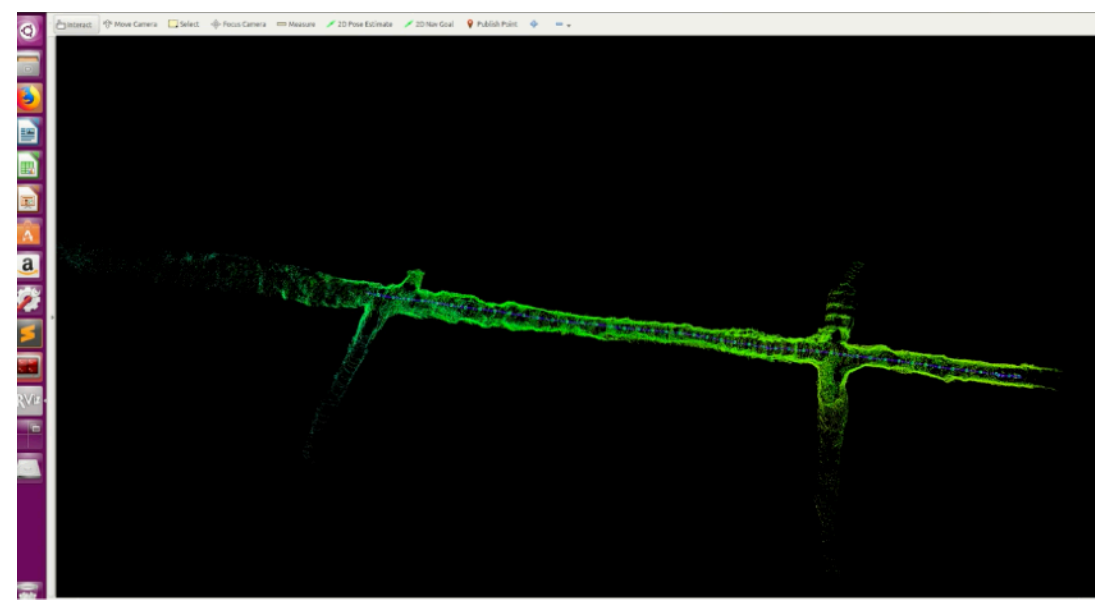
- extract features
- match "similar" features
- compute relative pose



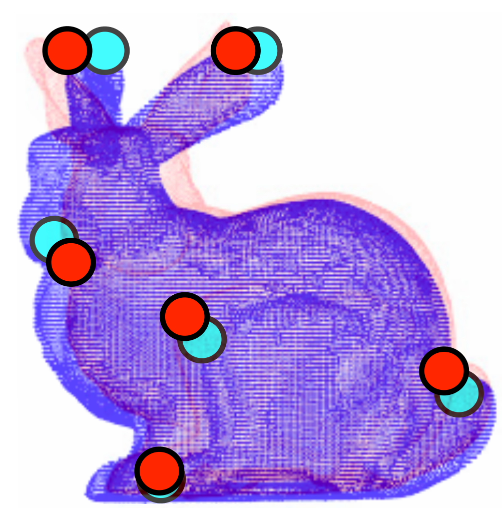
3D Registration (a.k.a. scan matching)

- **3D Registration problem**: find rigid transformation (position, rotation) that aligns two point clouds
  - Object pose estimation
  - Motion estimation (scan matching)





Typical registration procedure:
- extract features
- match “similar” features
- compute relative pose

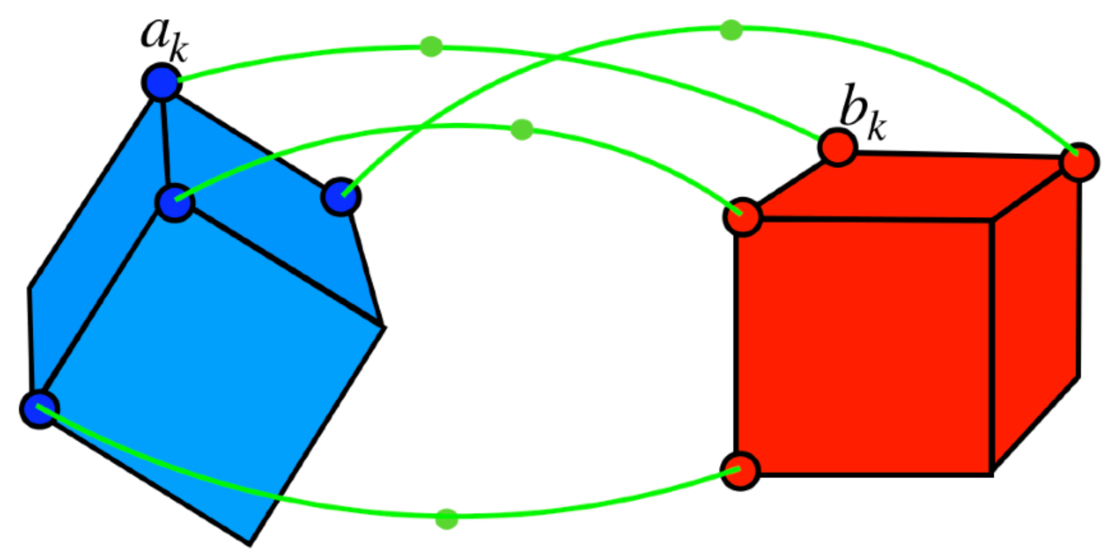


3D Registration: State of the Art

Registration **without** outliers:

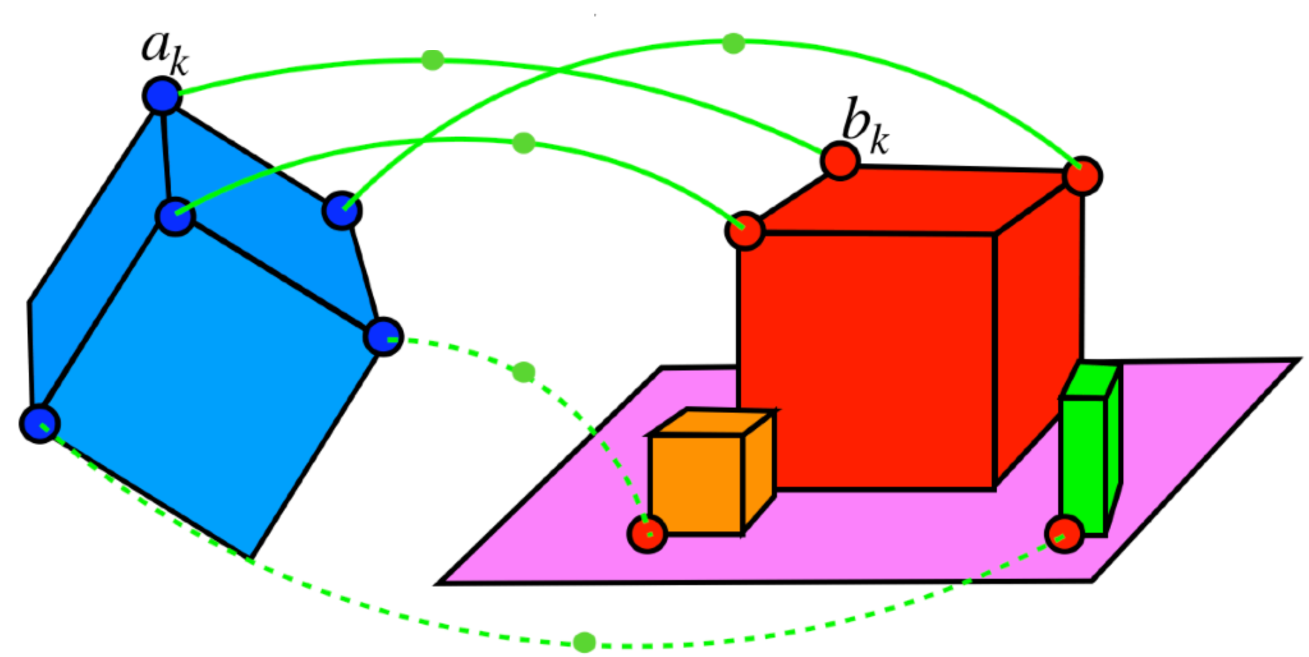
$$\min_{\boldsymbol{R} \in \mathrm{SO}(3)} \sum_{k=1}^{N} \|\boldsymbol{b}_k - \boldsymbol{R}\boldsymbol{a}_k\|^2$$

- Can be solved in closed form [Horn'87, Arun'87]



Registration **with** outliers:

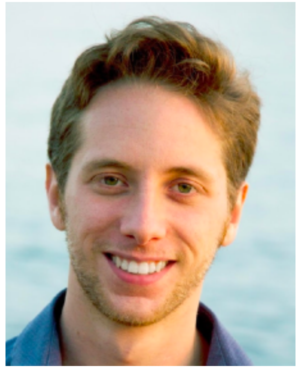
- **Fast heuristics** [ICP and variants, RANSAC]: tolerate small amount of outliers, no performance guarantee
- **Global solvers** [Branch-&-Bound, Mixed-integer programming]: tolerate many outliers but run in exponential time [Zhou et al, ECCV'16, Izatt et al., IJRR'17]



Bad News: Outlier Rejection is Inapproximable

Key result: outlier rejection is inapproximable.
In the worst case, there is no polynomial-time algorithm that can compute a near-optimal solution

[see also results from Chin et al.]



Paradigm shift: **certifiably robust algorithms**
Algorithms that can assess their performance in each problem instance:

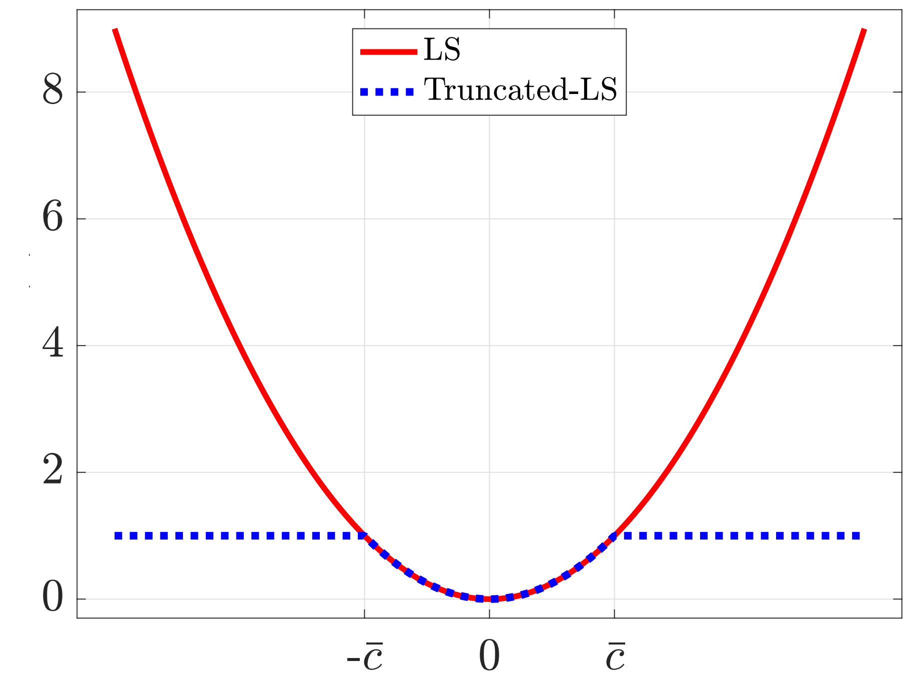
- <u>perform well and certify correctness</u> in common instances
- <u>detect and declare failure in worst case</u> problems (the once which are impossible to solve in polynomial time)

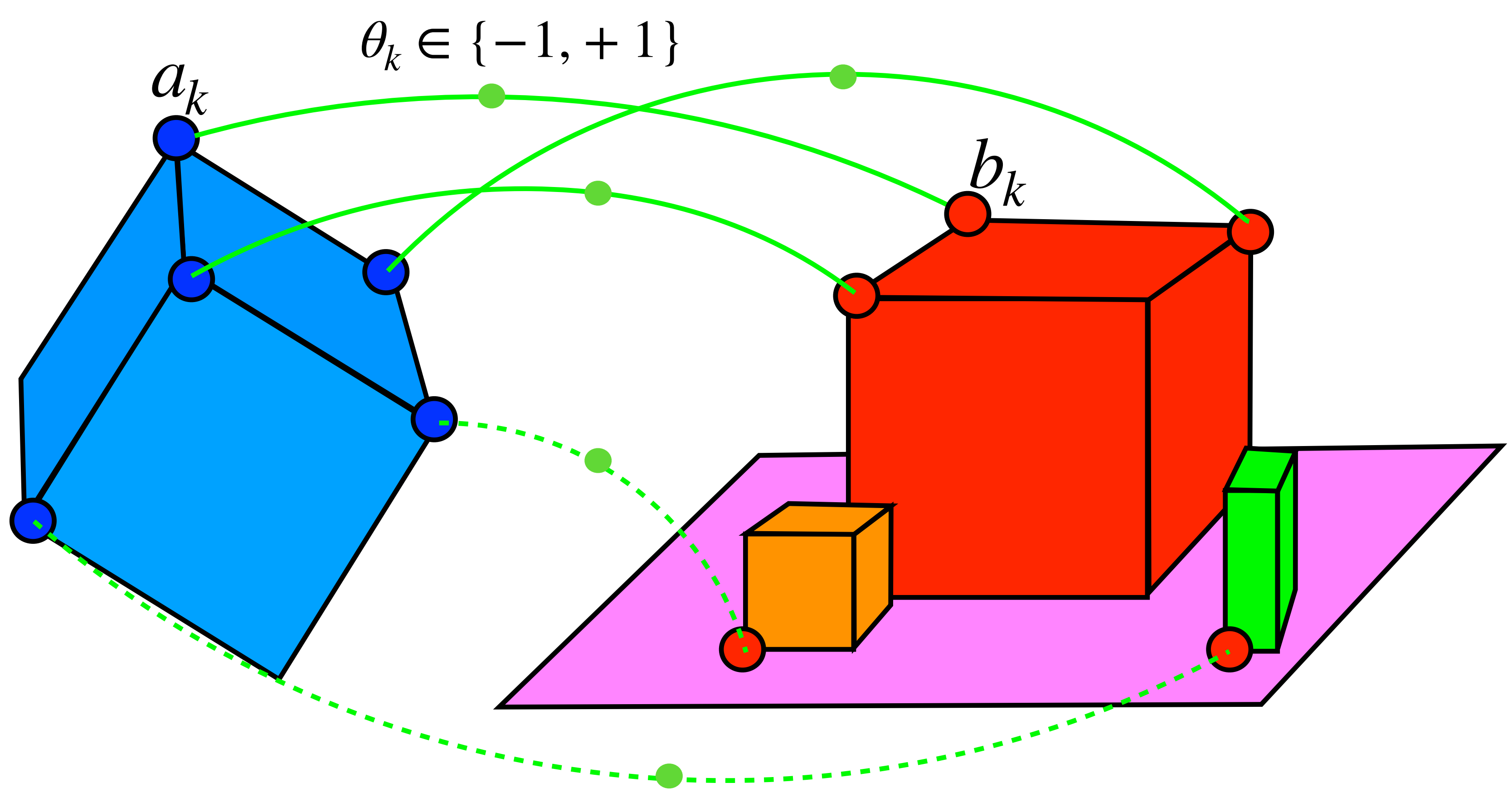
Tzoumas, Antonante, Carlone. Outlier-robust spatial perception: Hardness, general- purpose algorithms, and guarantees. ArXiv, 2019.

Certifiably Robust 3D Registration

Proposed:

$$\min_{\substack{\boldsymbol{R}\in \mathrm{SO}(3)\\ \theta_k\in\{-1,+1\}}} \sum_{k=1}^{N} \frac{(1+\theta_k)}{2}\|\boldsymbol{b}_k - \boldsymbol{R}\boldsymbol{a}_k\|^2 + \frac{(1-\theta_k)}{2}\bar{c}^2$$

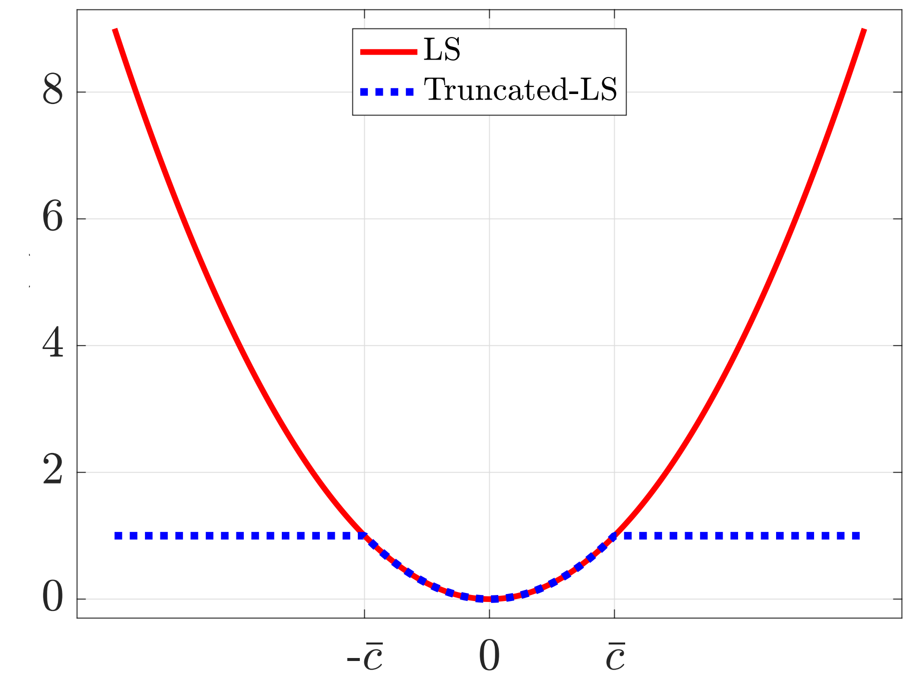


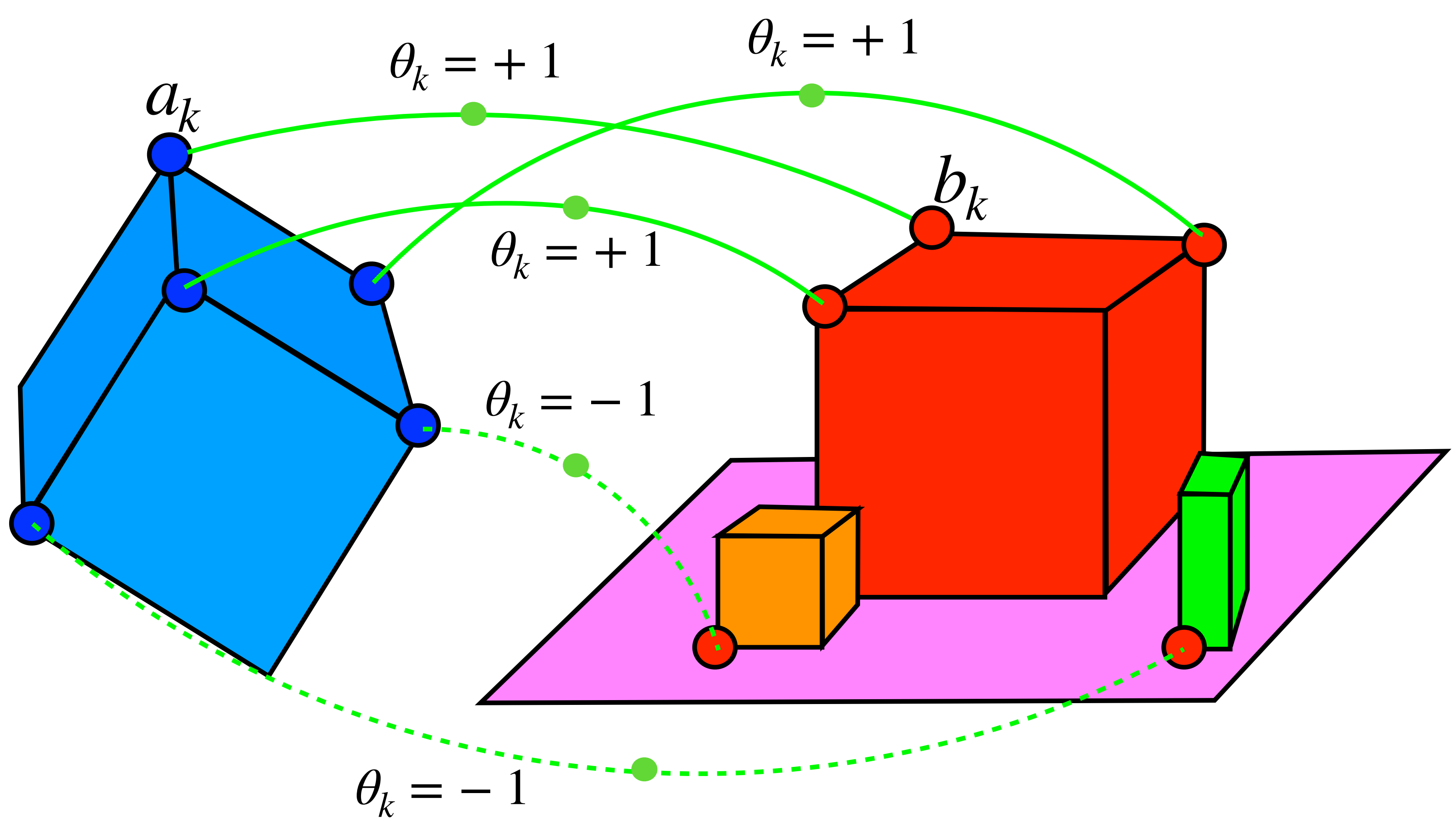


Certifiably Robust 3D Registration

Proposed:

$$\min_{\substack{\boldsymbol{R}\in\mathrm{SO}(3)\\ \theta_k\in\{-1,+1\}}} \sum_{k=1}^{N} \frac{(1+\theta_k)}{2}\|\boldsymbol{b}_k - \boldsymbol{R}\boldsymbol{a}_k\|^2 + \frac{(1-\theta_k)}{2}\bar{c}^2$$

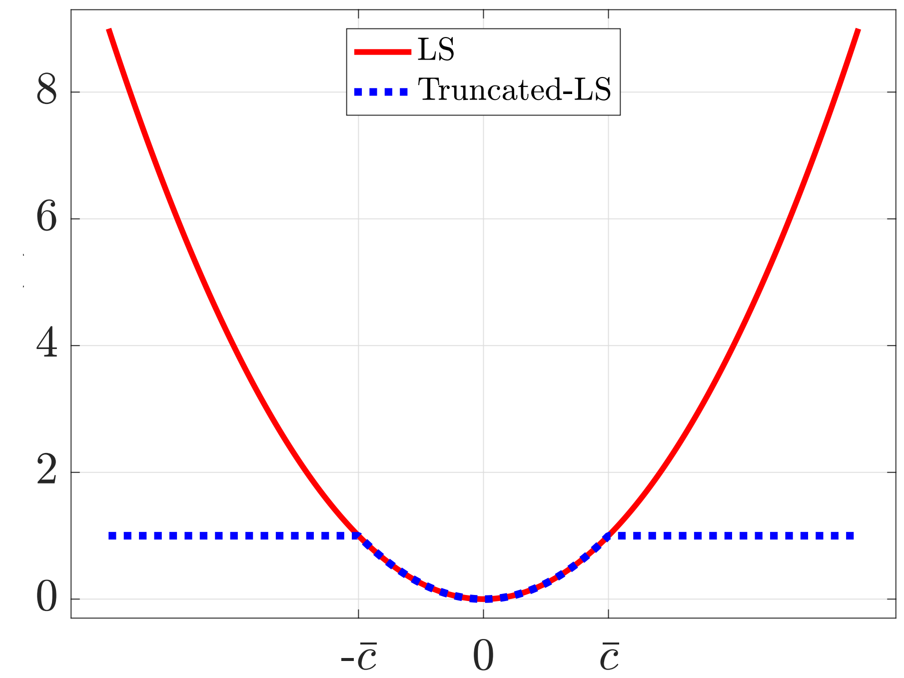


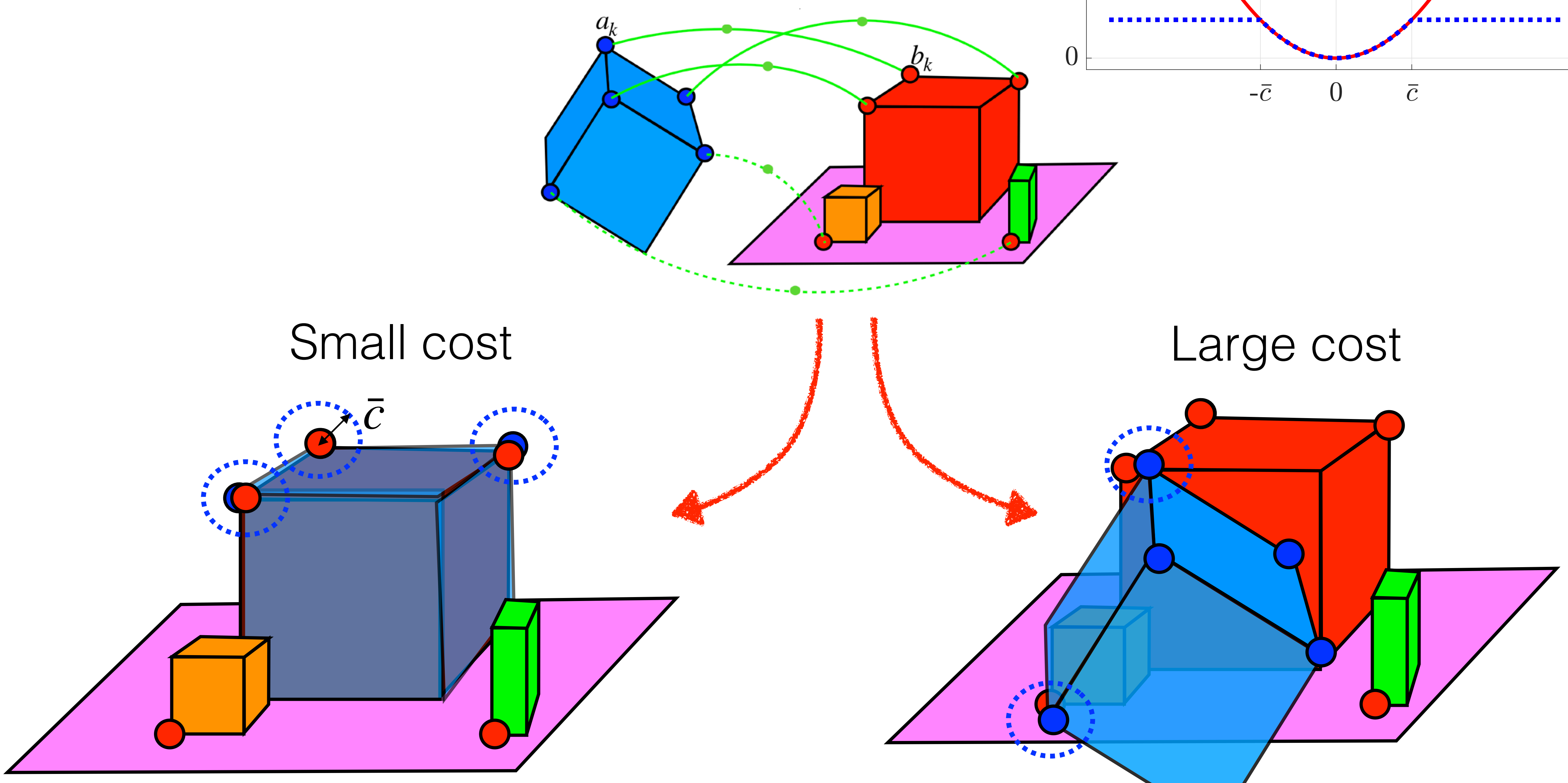


Certifiably Robust 3D Registration

Proposed:

$$\min_{\substack{\boldsymbol{R}\in\mathrm{SO}(3)\\ \theta_k\in\{-1,+1\}}} \sum_{k=1}^{N} \frac{(1+\theta_k)}{2}\|\boldsymbol{b}_k - \boldsymbol{R}\boldsymbol{a}_k\|^2 + \frac{(1-\theta_k)}{2}\bar{c}^2$$





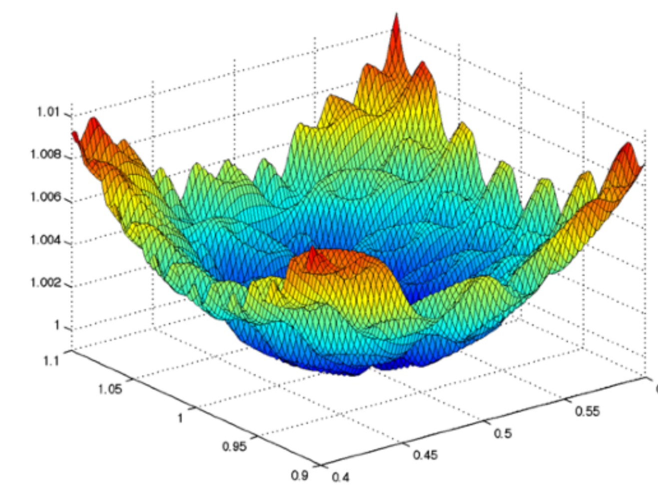
TEASER: Truncated least squares Estimation And SEmidefinite Relaxation

1) Discrete-Continuous Estimation (NP-hard):

$$\min_{\substack{\boldsymbol{R}\in\mathrm{SO}(3)\\ \theta_k\in\{-1,+1\}}} \sum_{k=1}^{N} \frac{(1+\theta_k)}{2}\|\boldsymbol{b}_k - \boldsymbol{R}\boldsymbol{a}_k\|^2 + \frac{(1-\theta_k)}{2}\bar{c}^2$$

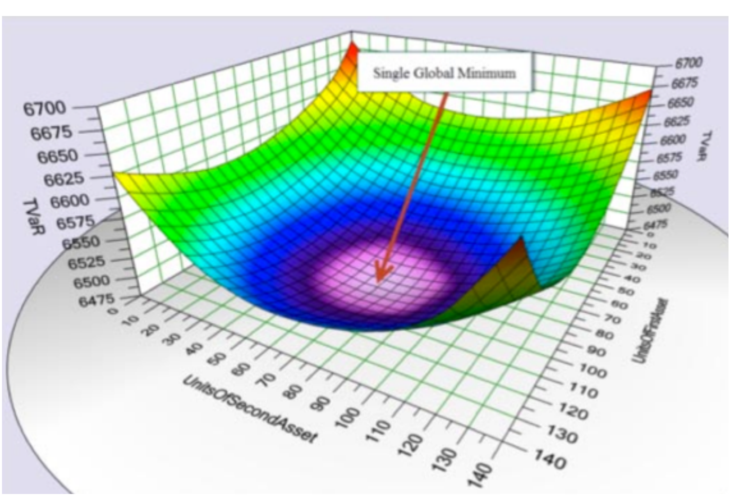
2) Quadratically-Constrained Quadratic Program (QCQP) (non-convex, NP-hard)

$$\min_{\boldsymbol{x}\in\mathbb{R}^{4(N+1)}} \sum_{i=1}^{N} \boldsymbol{x}^\mathsf{T}\boldsymbol{Q}_i\boldsymbol{x}$$
$$\textit{subject to} \quad \boldsymbol{x}_q^\mathsf{T}\boldsymbol{x}_q = 1$$
$$\boldsymbol{x}_{q_i}\boldsymbol{x}_{q_i}^\mathsf{T} = \boldsymbol{x}_q\boldsymbol{x}_q^\mathsf{T}$$



*non-trivial: naïve relaxation does not work well
- proposed has astonishing performance

3) Novel convex relaxation (solvable in polynomial-time)

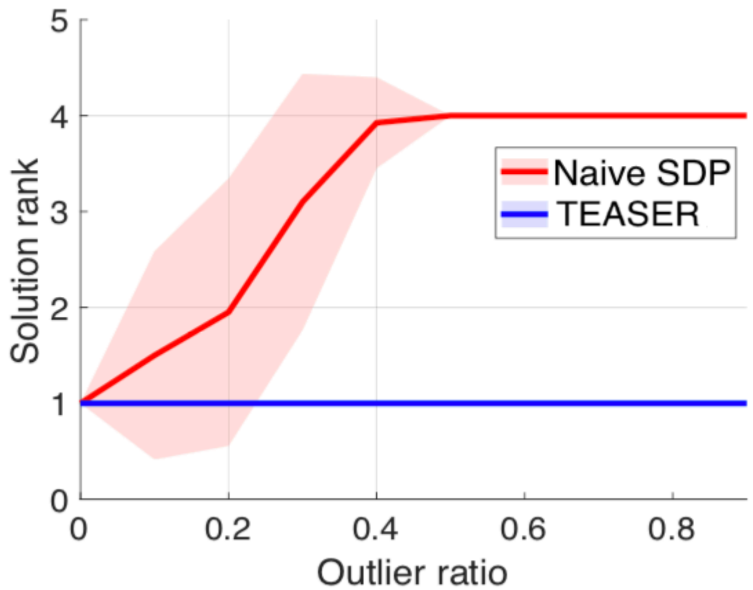


H. Yang and L. Carlone. A Polynomial-time Solution for Robust Registration with Extreme Outlier Rates. RSS 2019.
H. Yang and L. Carlone. A quaternion-based certifiably optimal solution to the Wahba problem with outliers. ICCV, 2019.

Certifiably Robust 3D Registration

Theorem 2 (Certification of robustness): If the solution Z^* of the convex relaxation has rank **1**, then Z^* can be factored into $Z^* = x^T x$, and x is the optimal solution of the original (combinatorial, non-convex) truncated least squares problem.

If rank = 1, all outliers are rejected and inliers have small errors



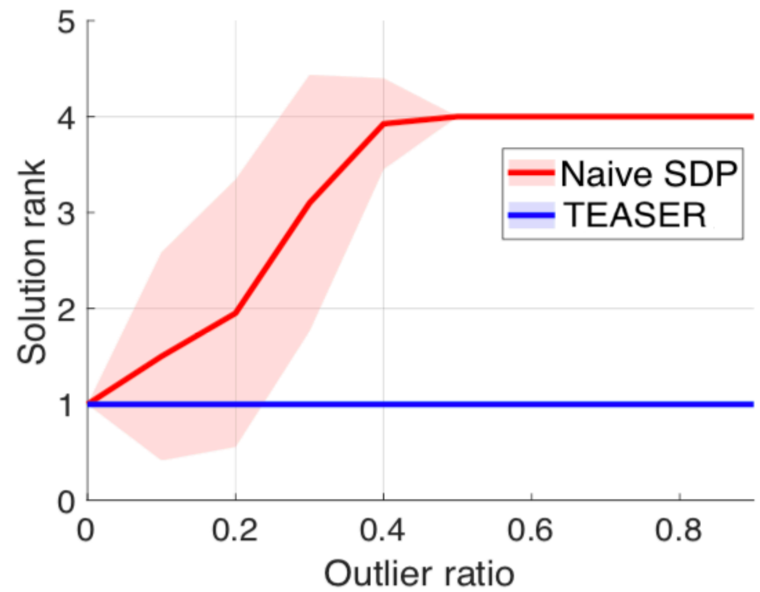
(a) Solution rank

H. Yang and L. Carlone. A Polynomial-time Solution for Robust Registration with Extreme Outlier Rates. RSS 2019.

H. Yang and L. Carlone. A quaternion-based certifiably optimal solution to the Wahba problem with outliers. ICCV, 2019.

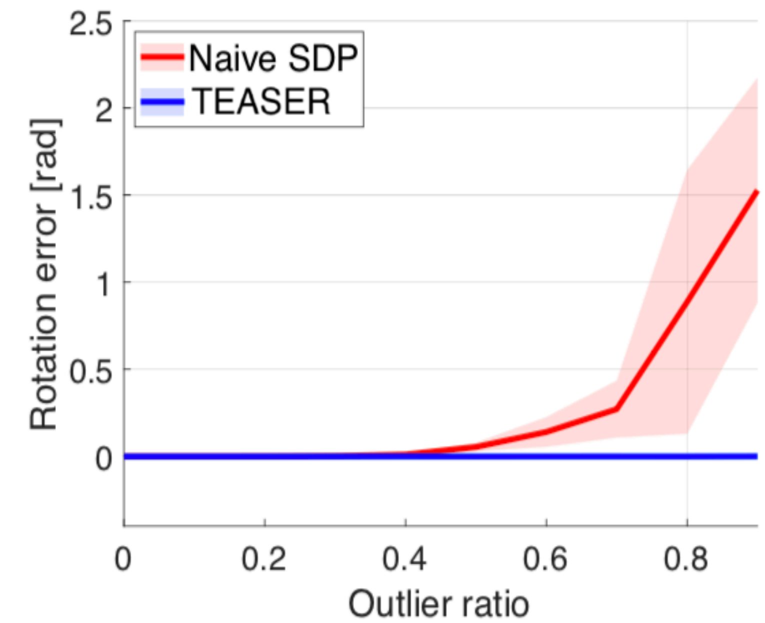
TEASER: Results

Comparison against naïve relaxation:



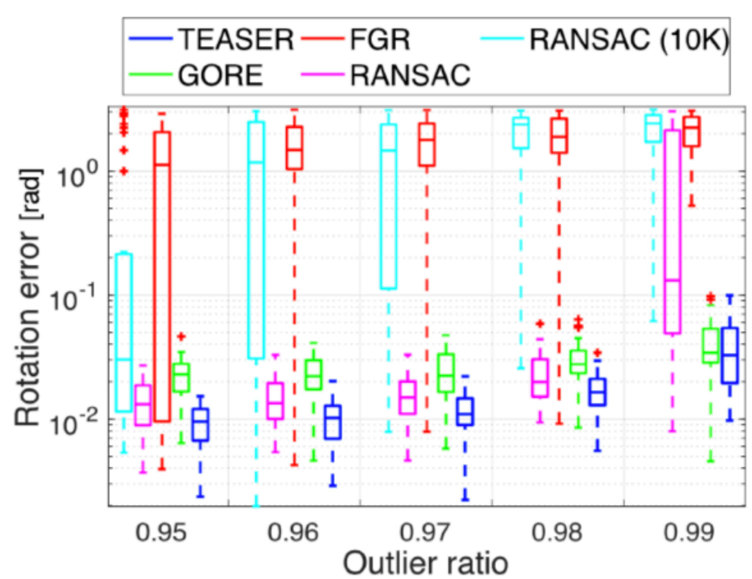
(a) Solution rank

*relaxation is exact when rank is 1

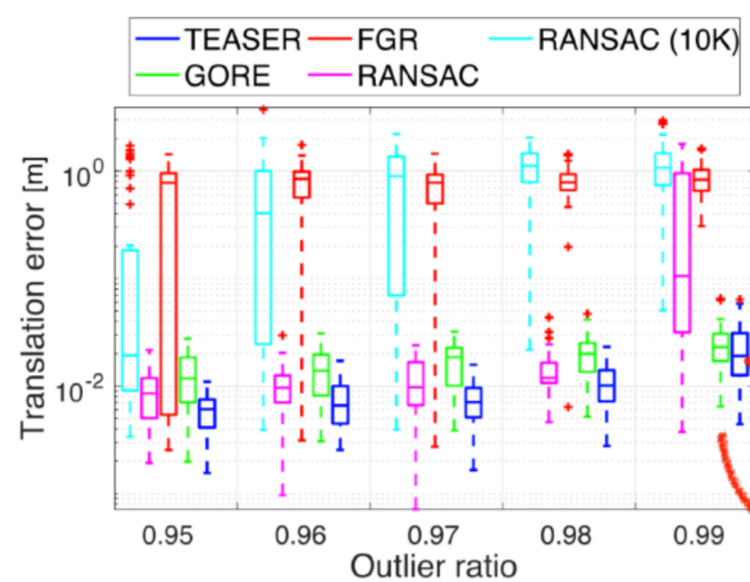


(b) Rotation errors

Testing on standard benchmarks: **Stanford Bunny**



(a) Rotation Error



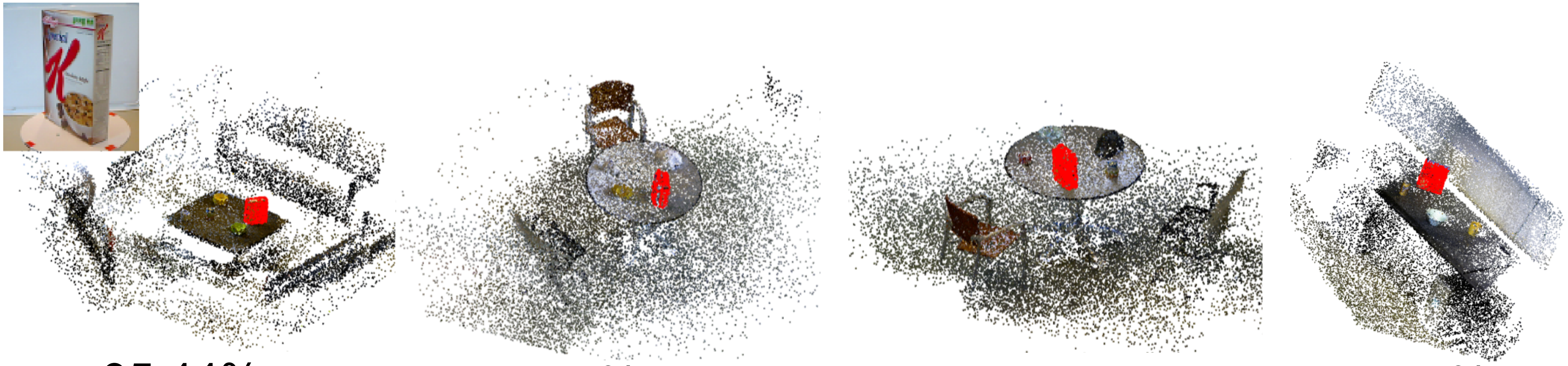
(b) Translation Error

*TEASER (proposed) is best approach

*second best is Branch-&-Bound (exponential time)

TEASER: first polynomial-time algorithm that tolerates extreme outliers rates

TEASER: RGB-D Object Detection



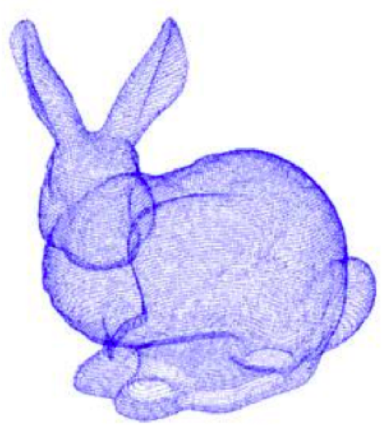
95.44% outliers

97.37% outliers

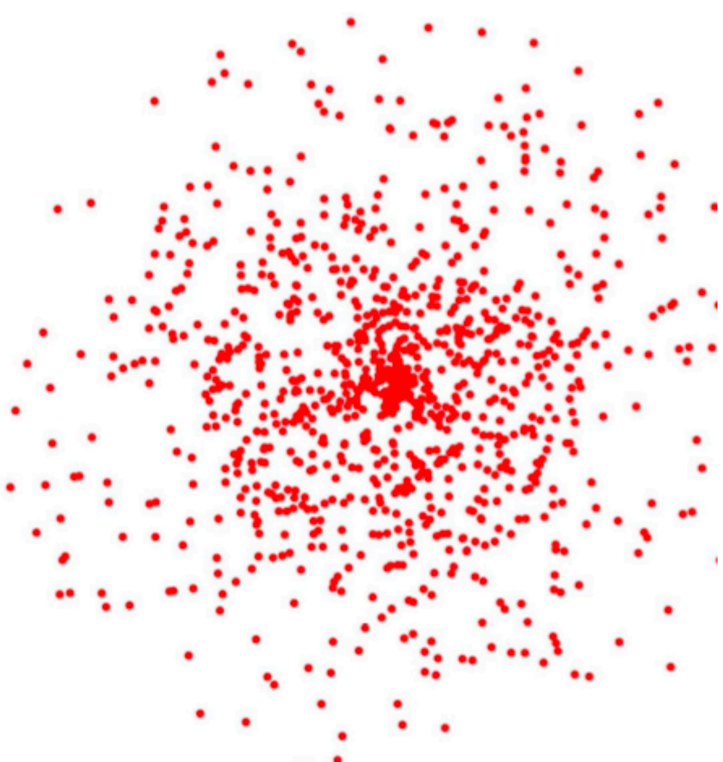
96.87% outliers

94.77% outliers

Stanford Bunny



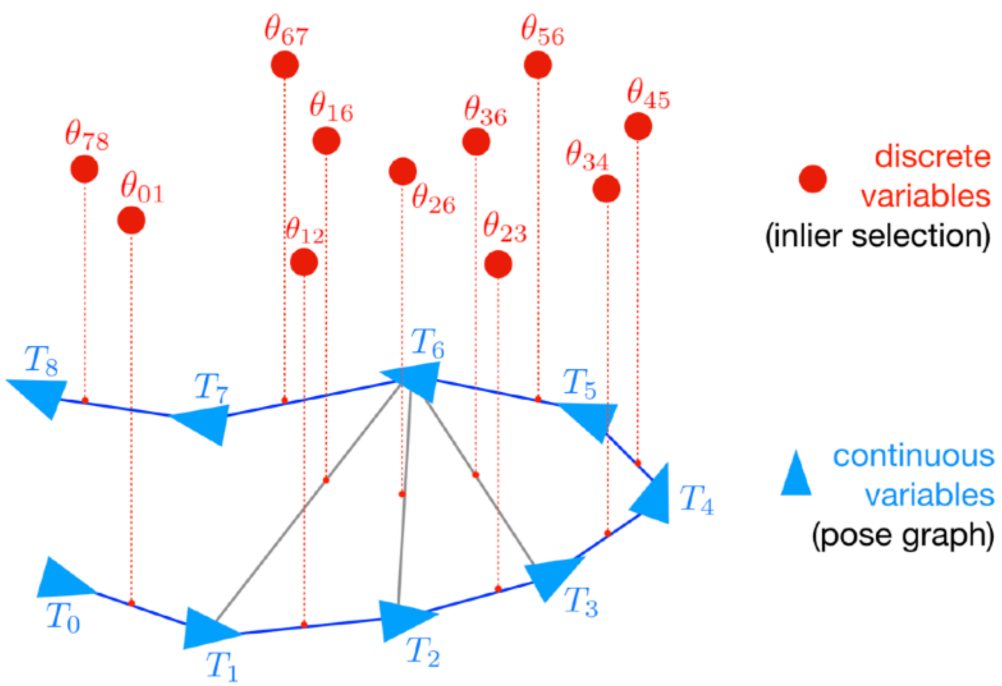
80% outliers

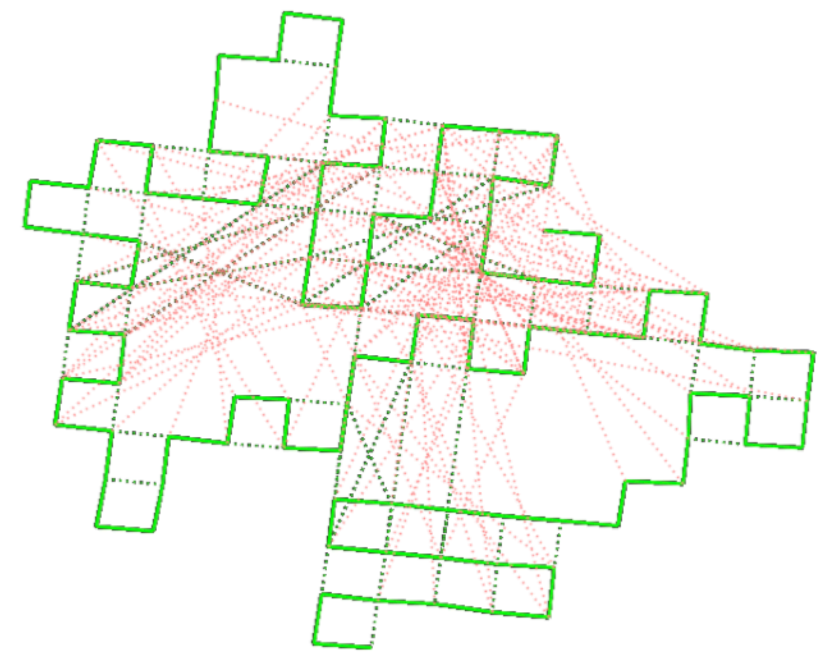


(a) Cluttered scene

Other applications

Robust Simultaneous Localization and Mapping

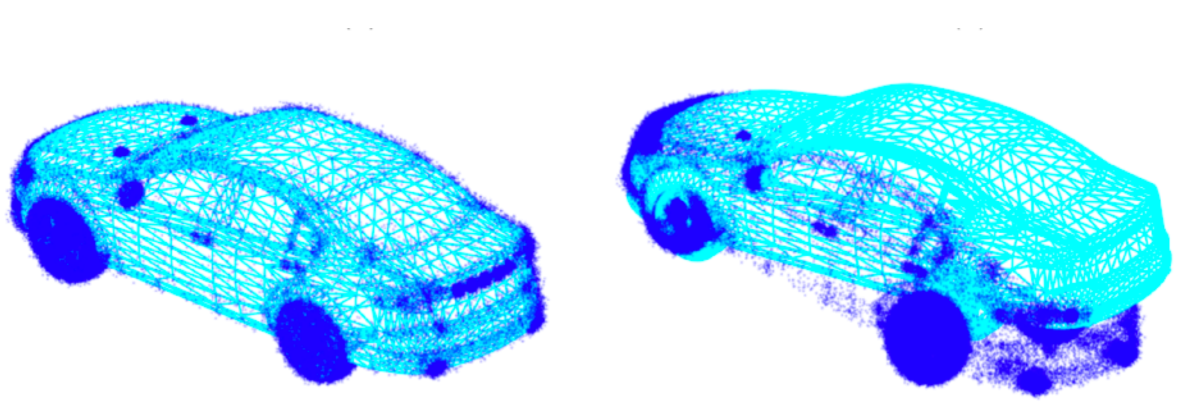




Lidar-based Object Localization

Proposed

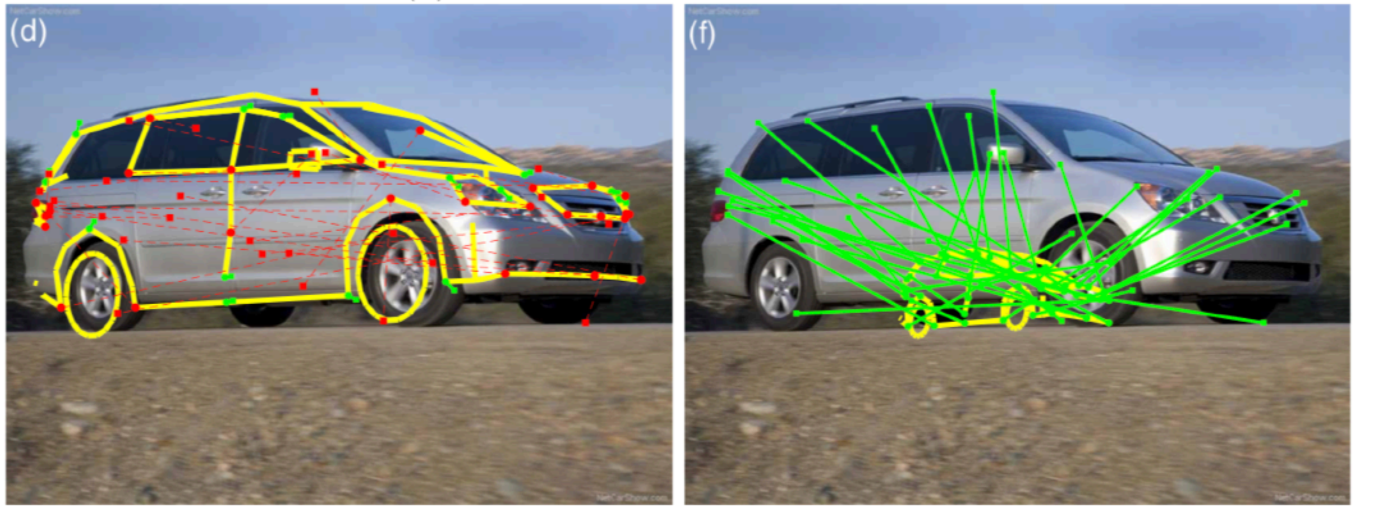
RANSAC



Camera-based Object Localization

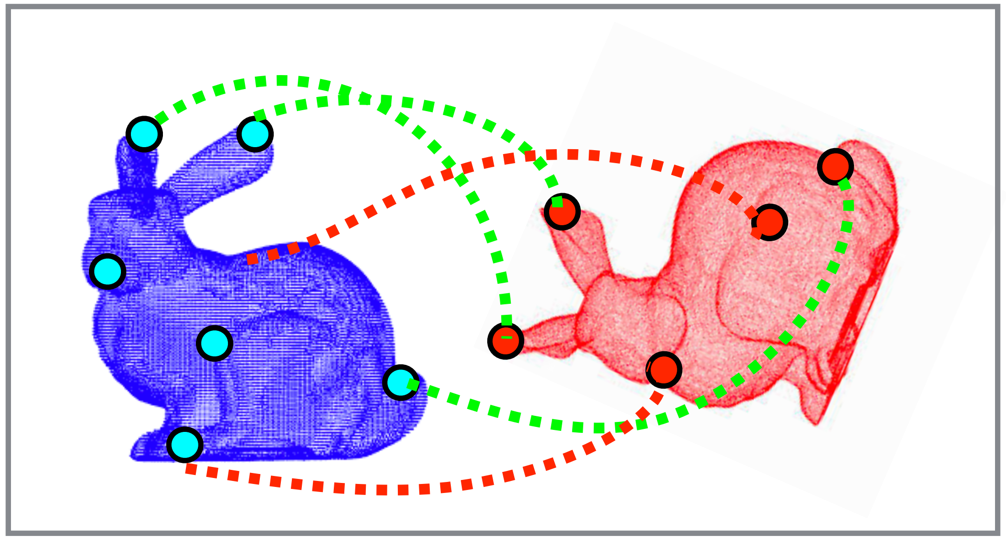
Proposed

RANSAC

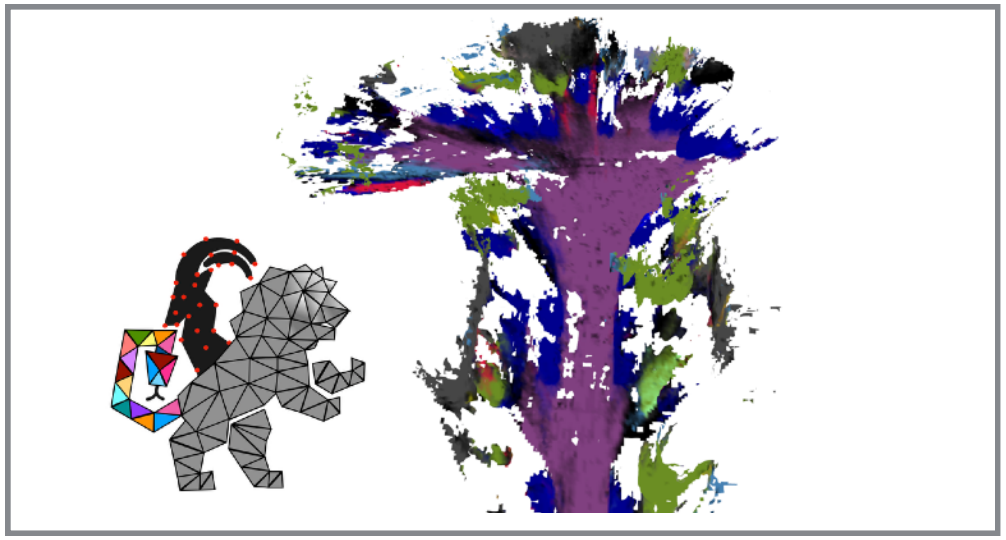


H. Yang, P. Antonante, V. Tzoumas, L. Carlone. Graduated non-convexity for robust spatial perception: From non-minimal solvers to global outlier rejection. Arxiv, 2019.

Outline

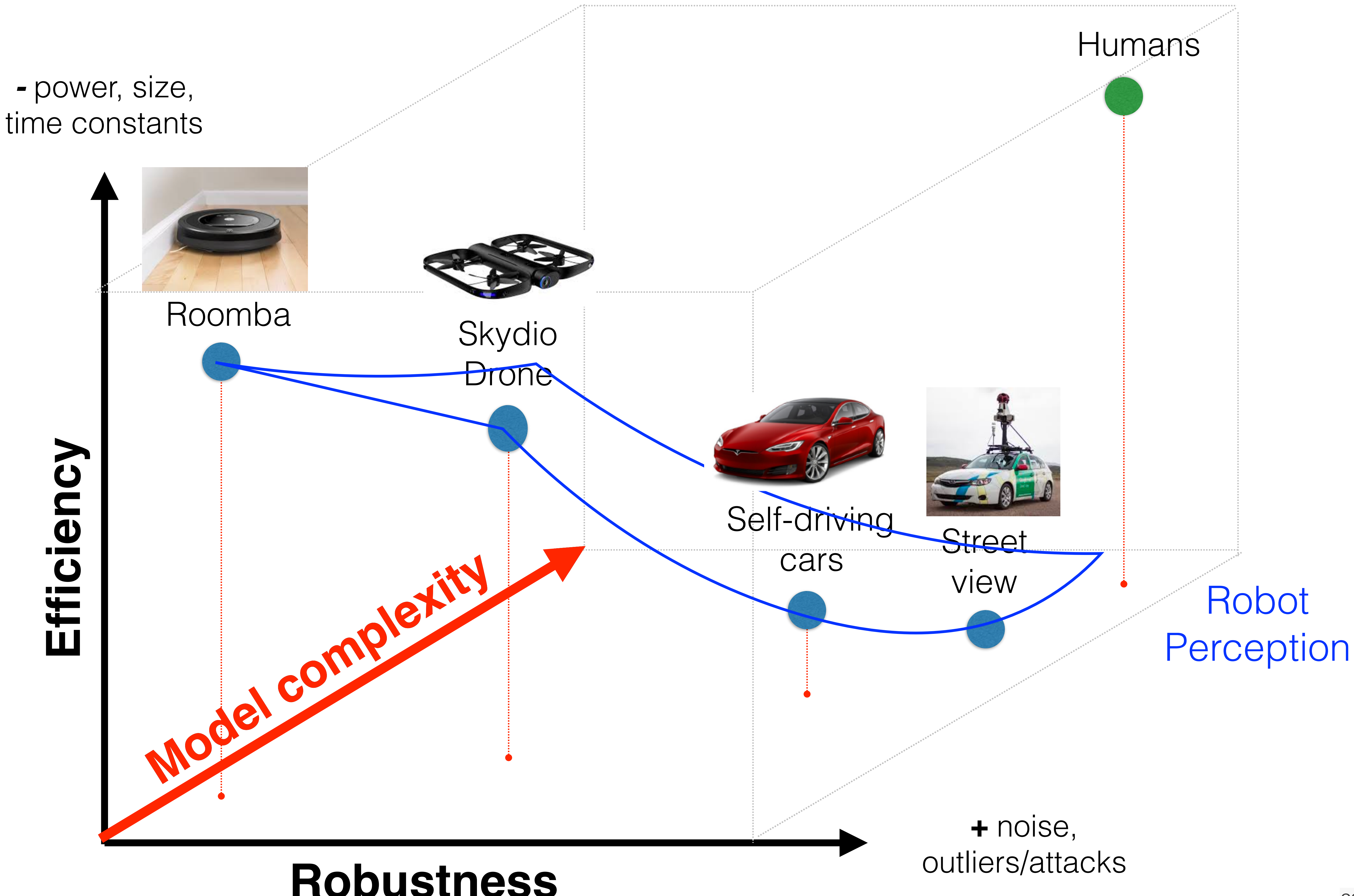


Certifiable Perception: algorithms that are "hard to break"

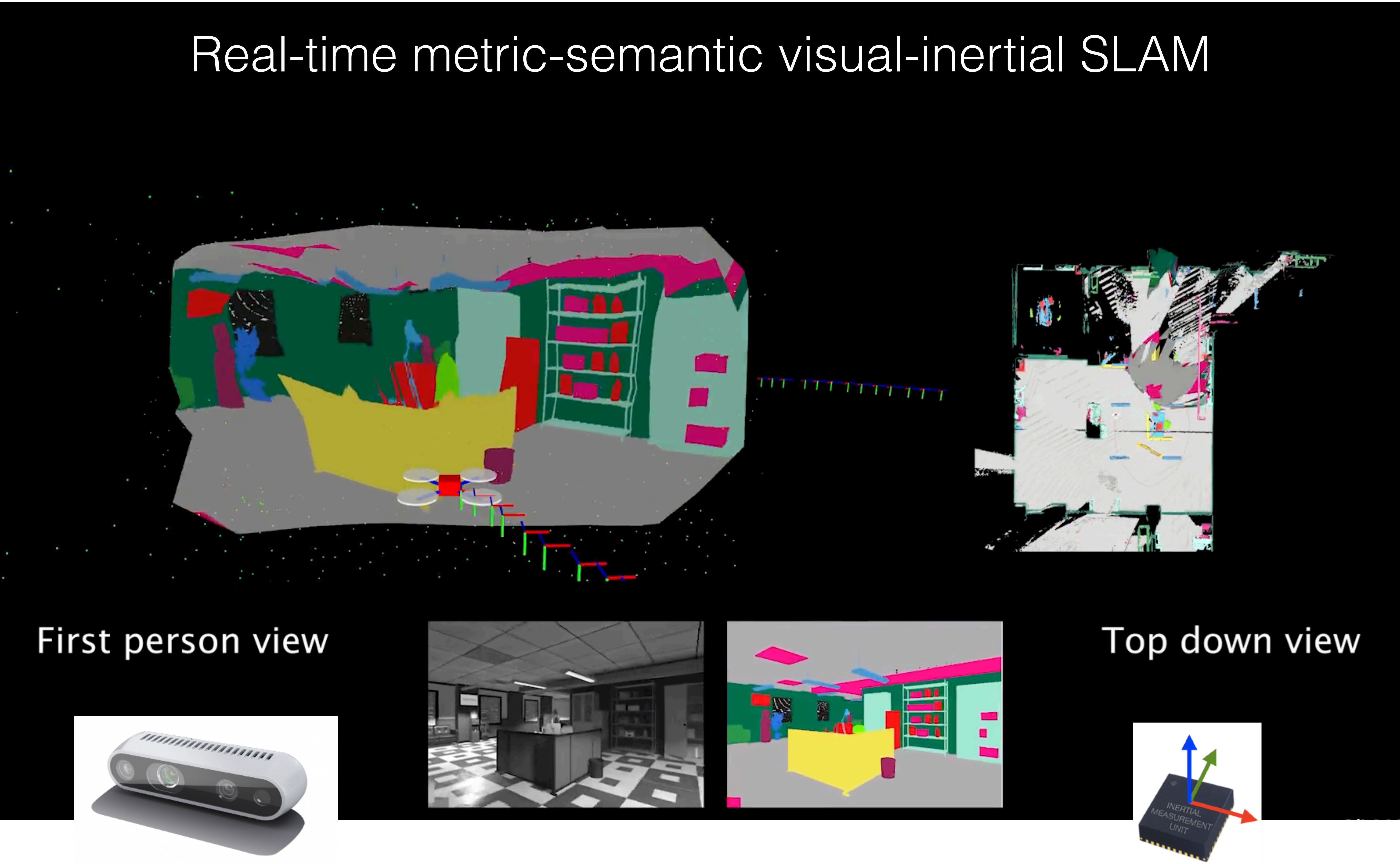


Kimera: real-time high-level understanding

Axes of complexity

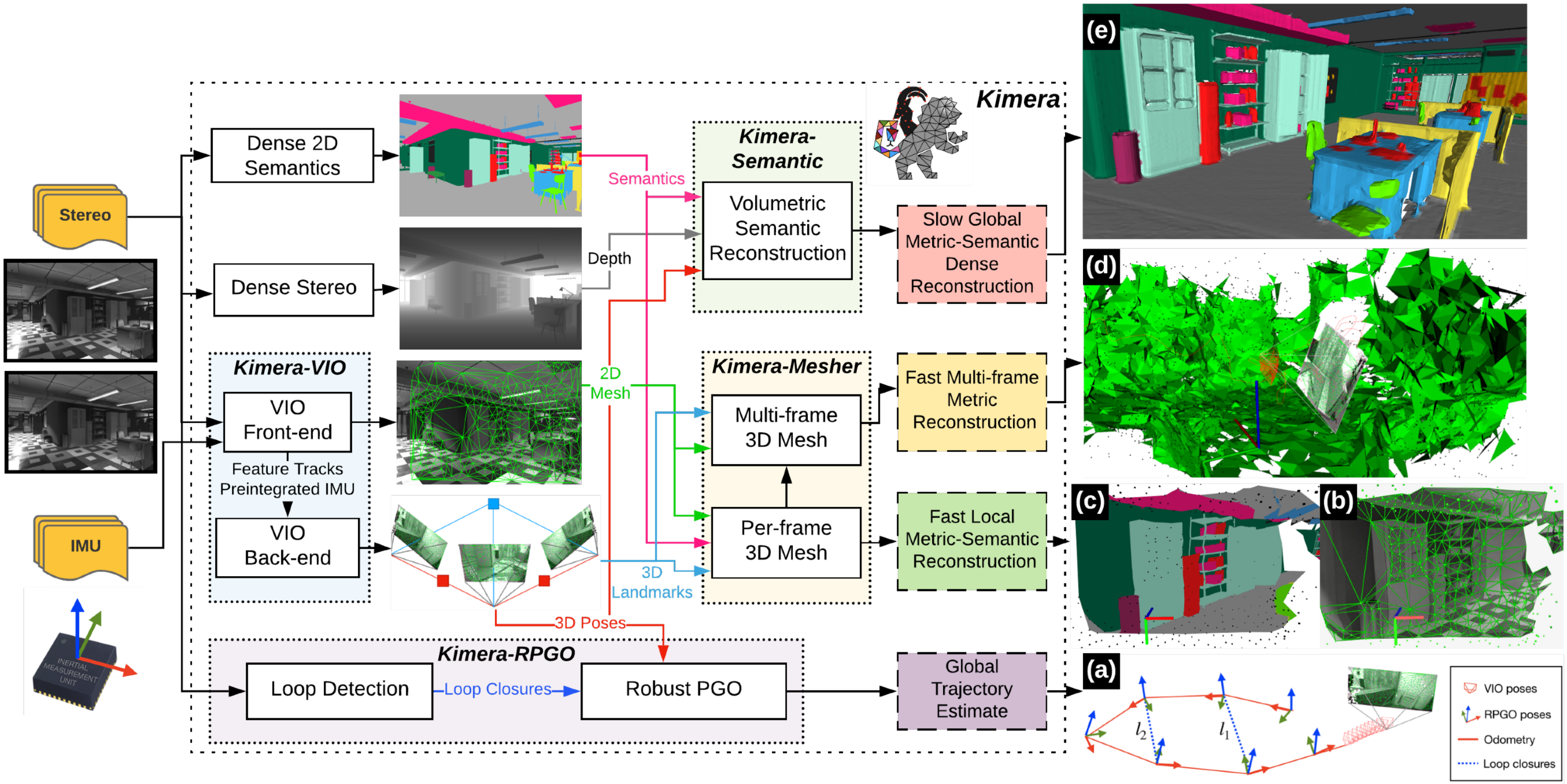


Releasing Kimera



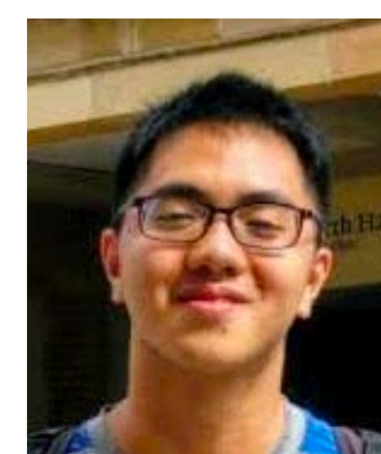
A. Rosinol, M. Abate, Y. Chang, and L. Carlone. Kimera: an open-source library for real-time metric-semantic localization and mapping. Arxiv, 2019.

Architecture

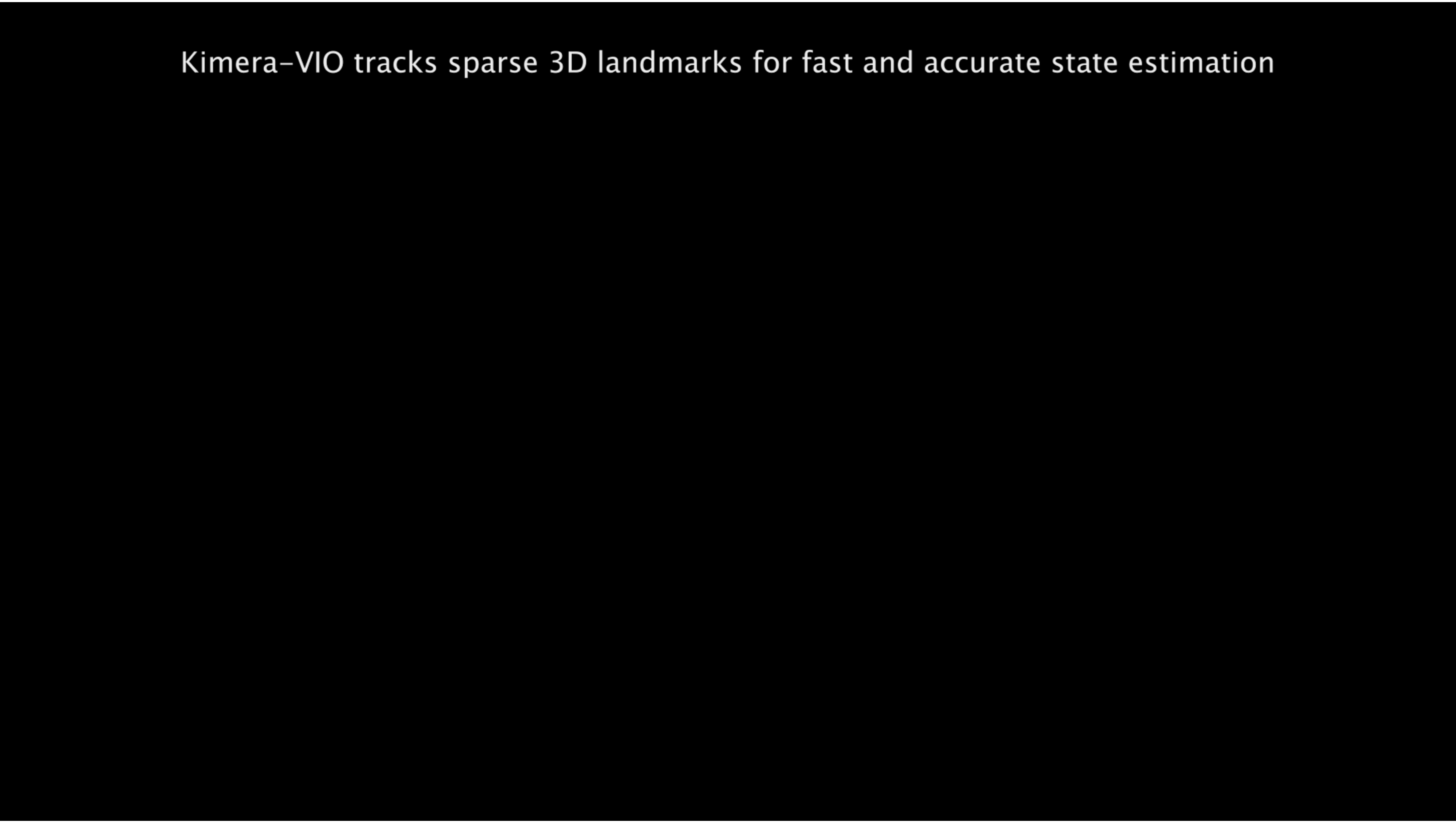


Outputs:

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

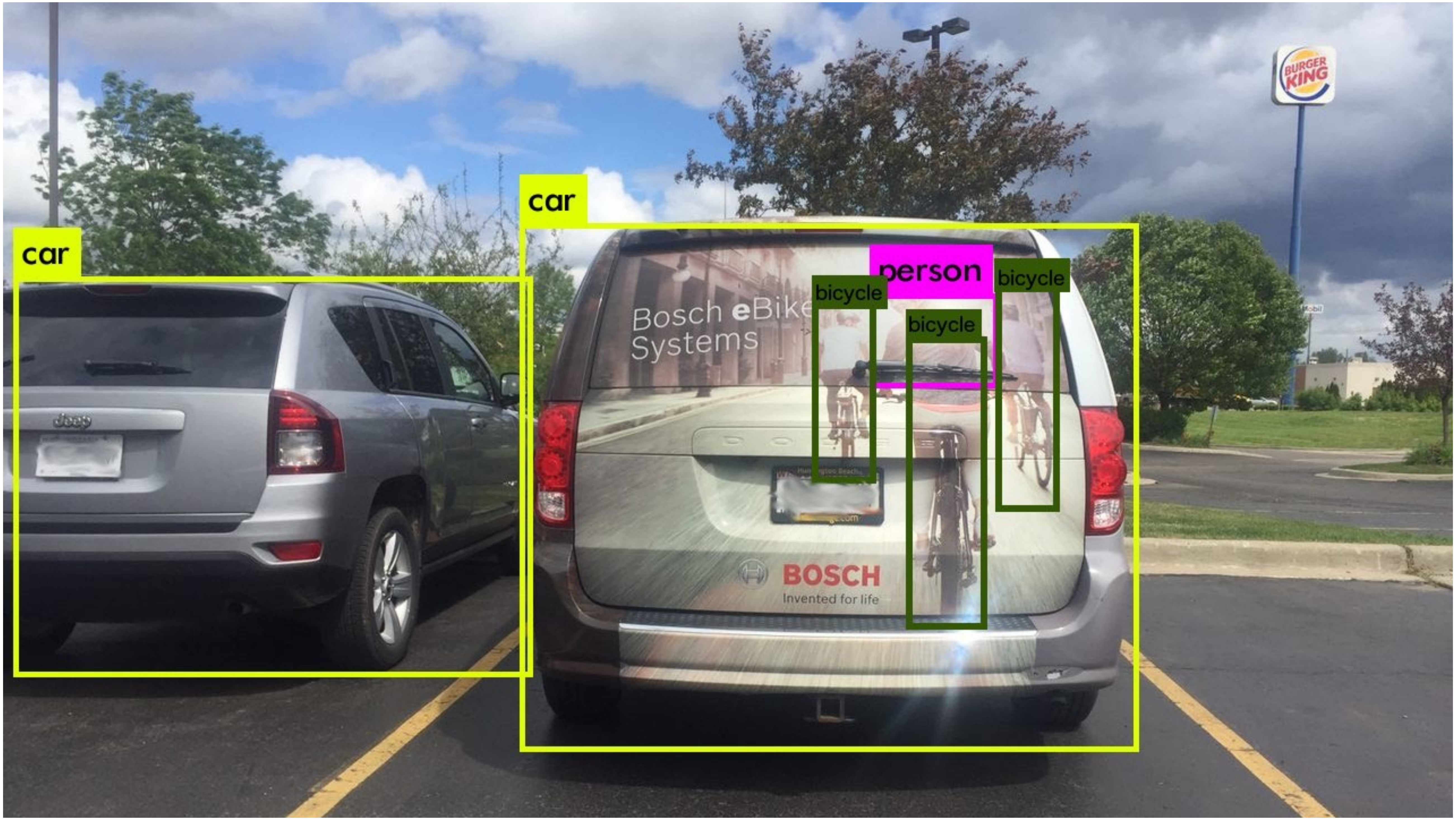


Kimera-VIO & Kimera-Mesher



Why Kimera?

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



Why Kimera?

solving 3D reconstruction failures

Conclusion Thank you!

- **Robustness**
  - robust perception is inapproximable :-(
  - certifiable perception: give up on solving all problems, but declare failure if you cannot solve a problem (most of the state-of-the-art approaches fail silently..)
  - fast implementation of TEASER coming soon!

- **High-level understanding: key to many applications**
  - initial step towards Spatial Perception
  - opportunities to bridge learning and geometry
  - It's also about robustness
  - https://github.com/MIT-SPARK/Kimera

ARL MIT LINCOLN LABORATORY Office of Naval Research Science & Technology Google