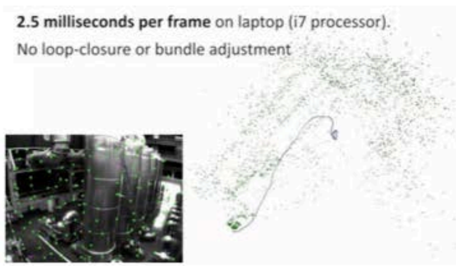
Dataset Generation and Benchmarking of SLAM Algorithms for Robotics and VR/AR
(ICRA 2019 workshop)

Jieum Hyun
KAIST Urban Robotics Lab.
jimi.hyun@kaist.ac.kr

Invited talk: Prof. Davide Scaramuzza

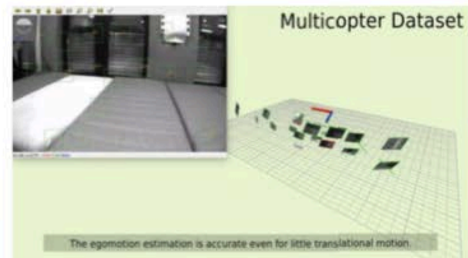
- Benchmarking SLAM: Current Status and the Road Ahead

- There are more and more VIO-VISLAM algorithms

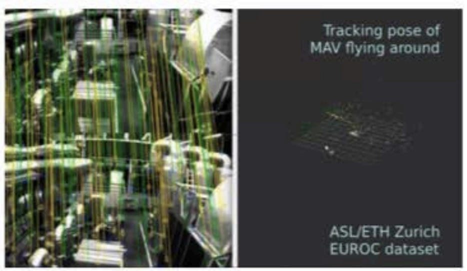


SVO+MSF

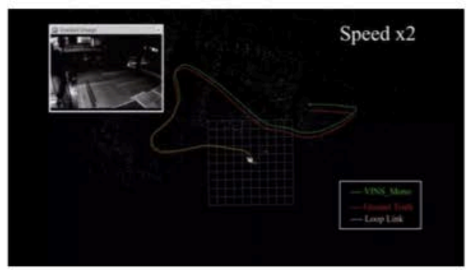

MSCKF



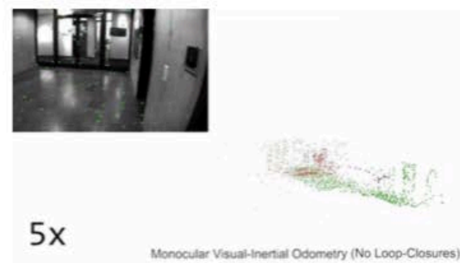
ROVIO



OKVIS



VINS-Mono



SVO+GTSAM

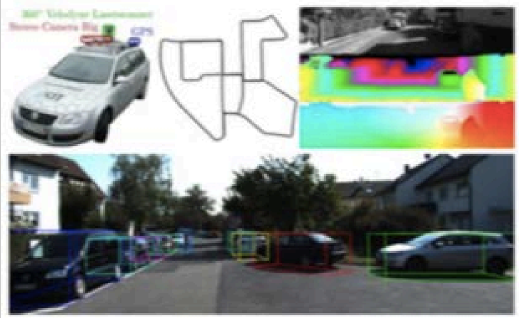
- But, how do we compare them?

Invited talk: Prof. Davide Scaramuzza

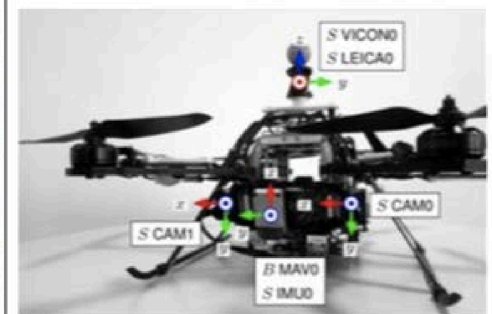
- Example Real-world Datesets

Devon Island [Furgale'11]
Stereo + D-GPS + inclinometer + sun sensor

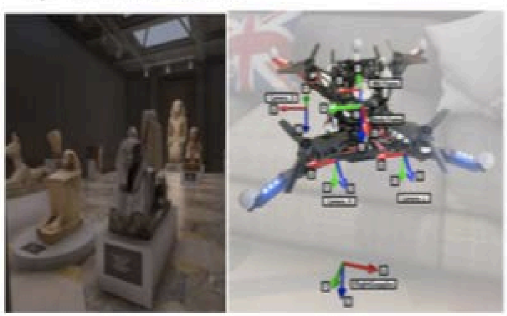
KITTI [Geiger'12]
Automobile, Laser + stereo + GPS, multiple tasks



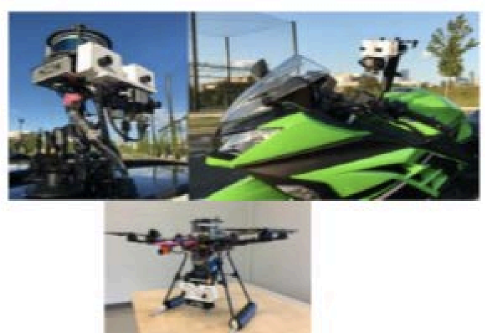
EuRoC [Burri'16]
MAV with synchronized IMU and stereo



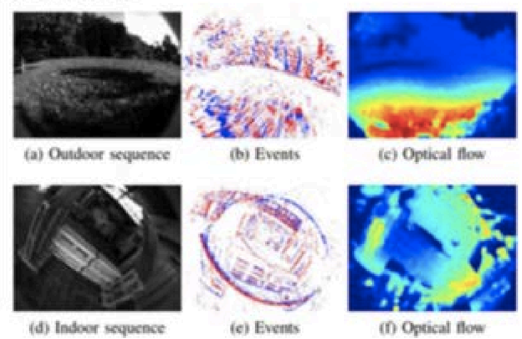
Blackbird [Antonini'18]
MAV indoor aggressive flight with rendered images and real dynamics + IMU



MVSEC [Zhu'18]
Events, frames, lidar, GPS, IMU from cars, drones, and motorcycles

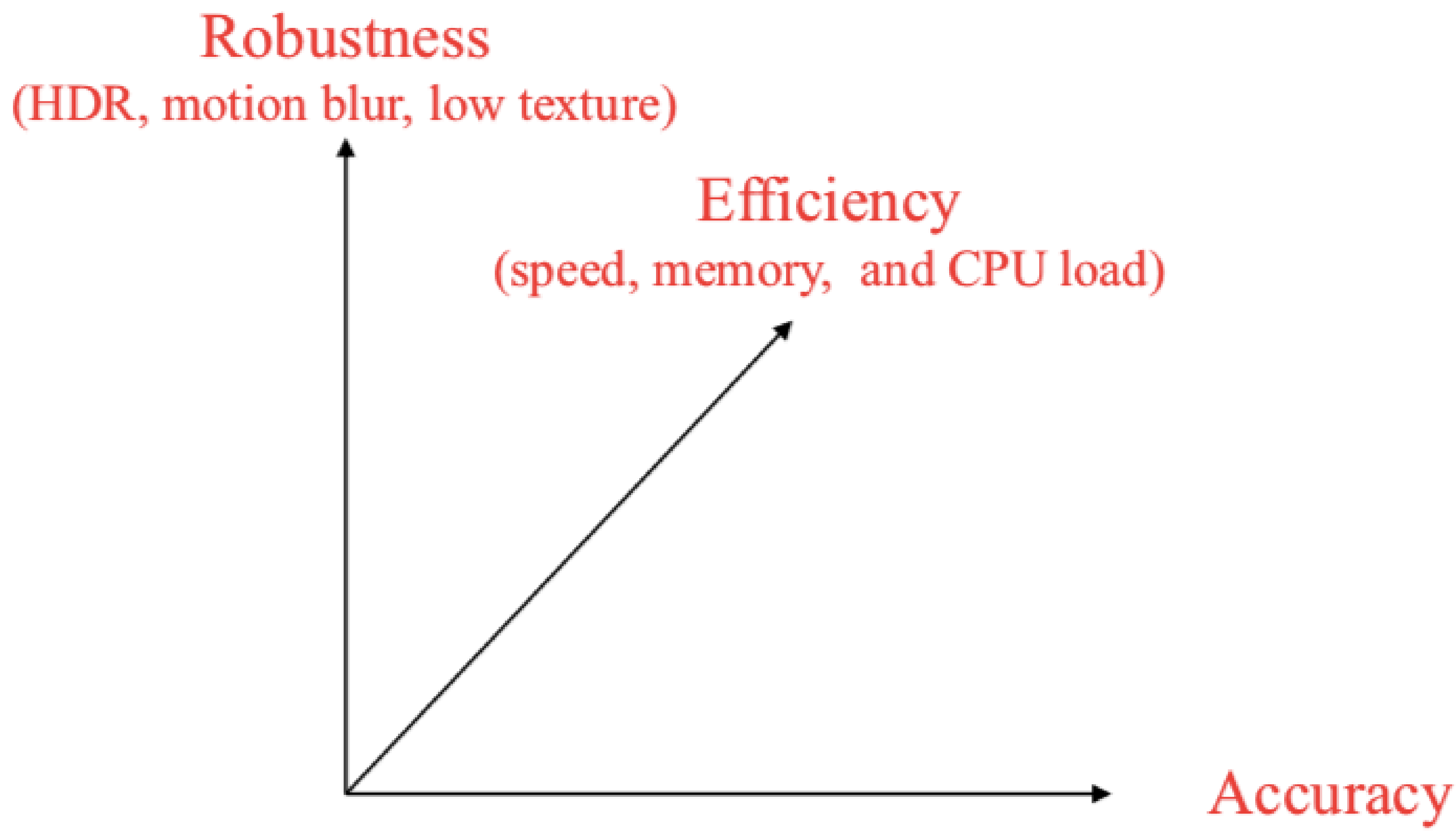


UZH Drone Racing [Delmerico'19]
MAV aggressive flight, standard + event cameras, IMU, indoors and outdoors



Invited talk: Prof. Davide Scaramuzza

- What metrics should be used?



Invited talk: Prof. Davide Scaramuzza

- Metric 1: Absolute Trajectory Error (ATE)

Absolute Trajectory Error
RMSE of the aligned estimate and the groundtruth.

Step 1: Align the trajectory

$$\underset{R,T,s}{\operatorname{argmin}} \sum_{i=0}^{N} \|\hat{t}_i - sRt_i - T\|^2$$

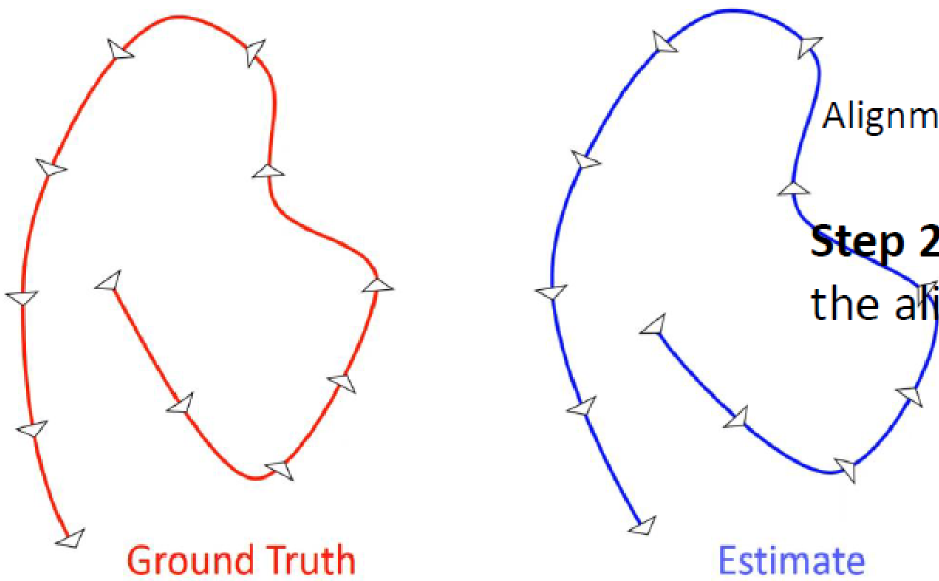
groundtruth positions

estimated positions

Alignment parameters

Step 2: Root mean squared errors between the aligned estimate and the groundtruth.

$$\sqrt{\frac{\sum_{i=1}^{N} \|\hat{t}_i - sRt_i - T\|^2}{N}}$$



✓ Single number metric

✗ Many parameters to specify

KAIST

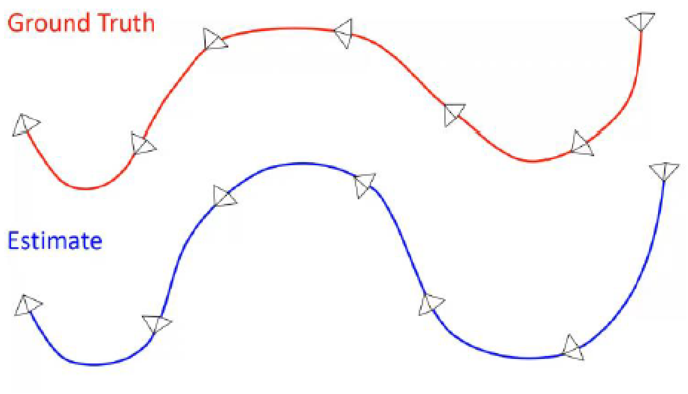
Zhang, Zichao, and Davide Scaramuzza. "A tutorial on quantitative trajectory evaluation for visual (-inertial) odometry." *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2018.

URL URBAN ROBOTICS LAB

Invited talk: Prof. Davide Scaramuzza

- Metric 2: Relative Trajectory Error (RTE)

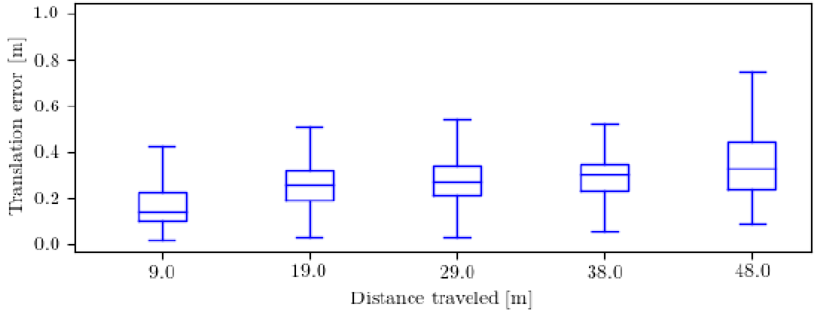
Relative Error (Odometry Error)
Statistics of sub-trajectories of specified lengths.

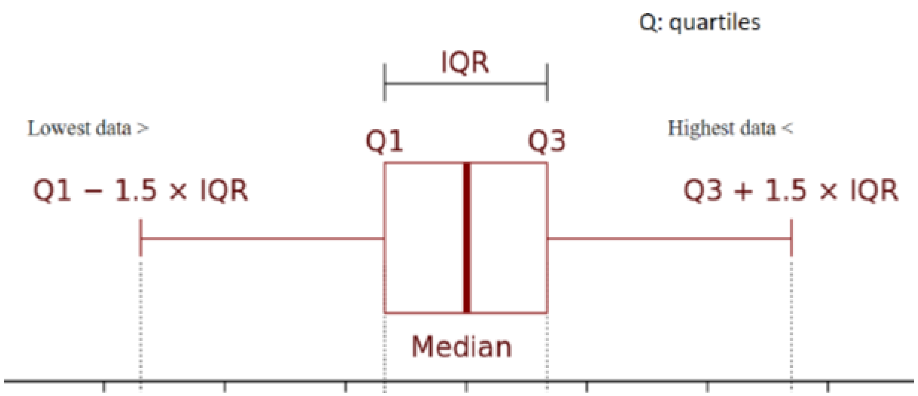


✓ Informative statistics

✗ Complicated to compute and rank

➢ Calculate errors for all the subtrajectories of certain lengths.



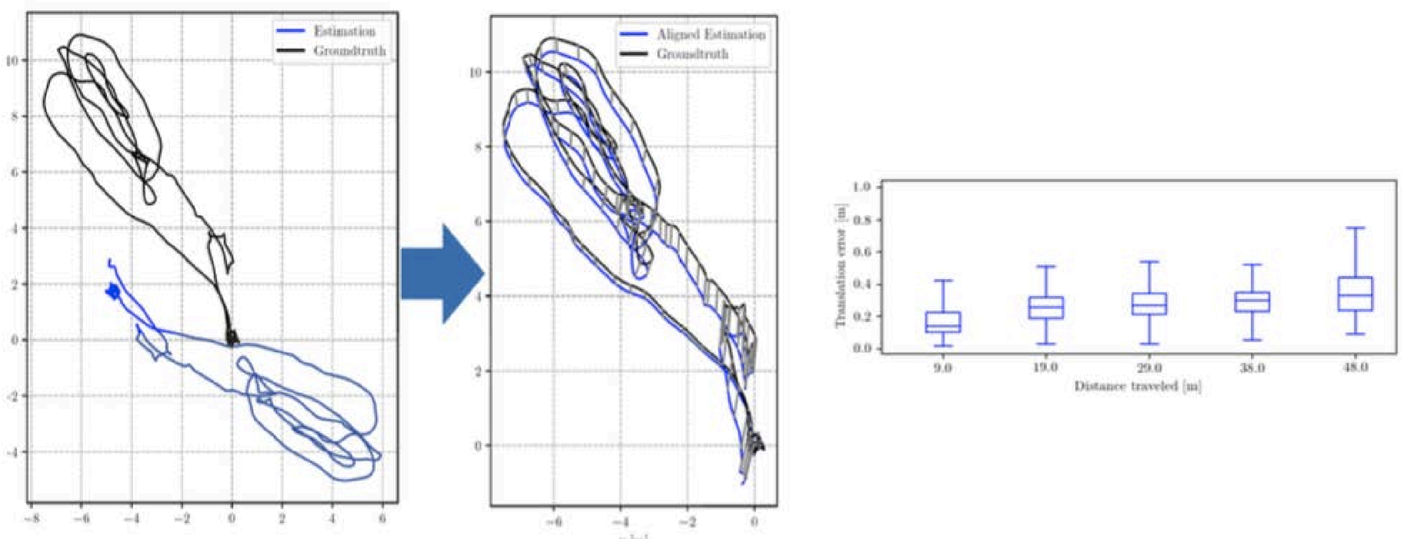


Zhang, Zichao, and Davide Scaramuzza. "A tutorial on quantitative trajectory evaluation for visual (-inertial) odometry." *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2018.

Invited talk: Prof. Davide Scaramuzza

- Accuracy: Trajectory Evaluation Toolbox

- Designed to make trajectory evaluation easy
  - Implements different alignment methods depending on the sensing modalities: SE(3) for stereo VO, sim(3) for monocular, 4DOF for VIO
  - Implements Absolute Trajectory Error and Relative Error
    - ➢ Code: https://github.com/uzh-rpg/rpg_trajectory_evaluation [Zhang, IROS'18]

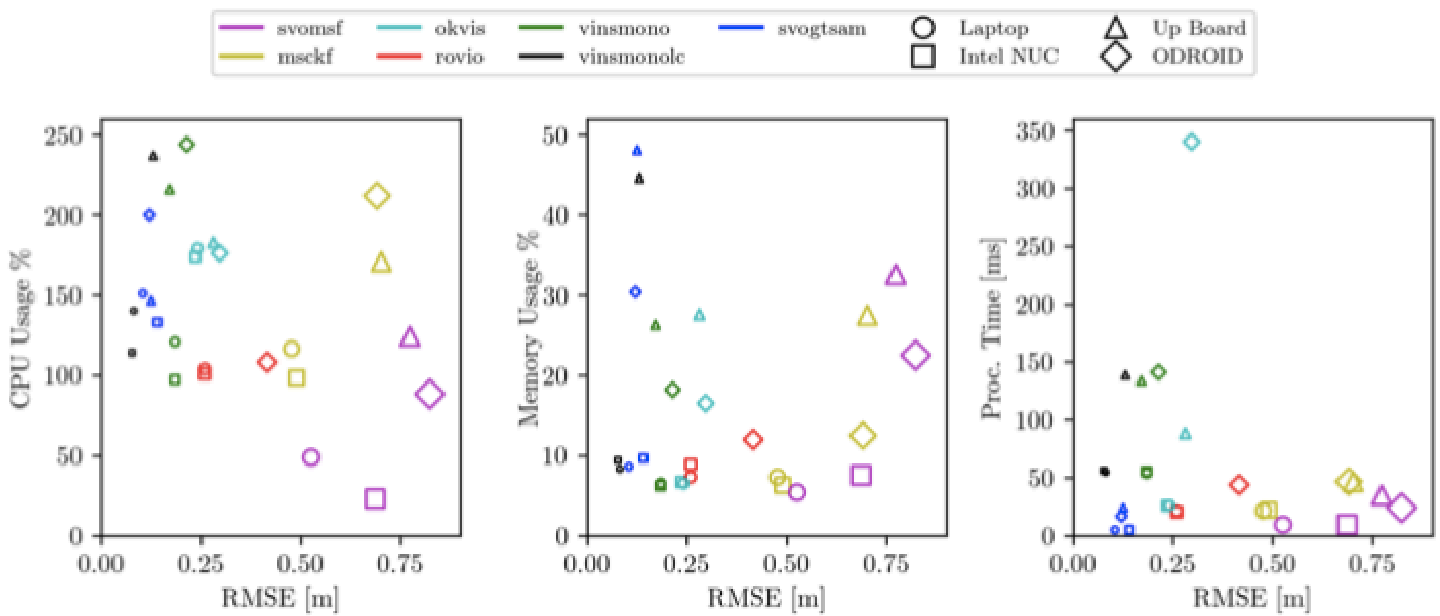


Zhang et al., "A tutorial on quantitative trajectory evaluation for visual (-inertial) odometry." IROS'18. PDF

Invited talk: Prof. Davide Scaramuzza

- Benchmarking Efficiency
  - Memory, CPU load, Processing time
  - Depends not only on algorithm design but also implementation, platforms, etc.

Case study: VIO for Flying Robots [ICRA'18]



No free lunch: more computation → better accuracy

KAIST

URL URBAN ROBOTICS LAB

Invited talk: Prof. Davide Scaramuzza

- Robustness is the greatest challenge for SLAM
- How can we quantify the robustness of algorithms to such situations?

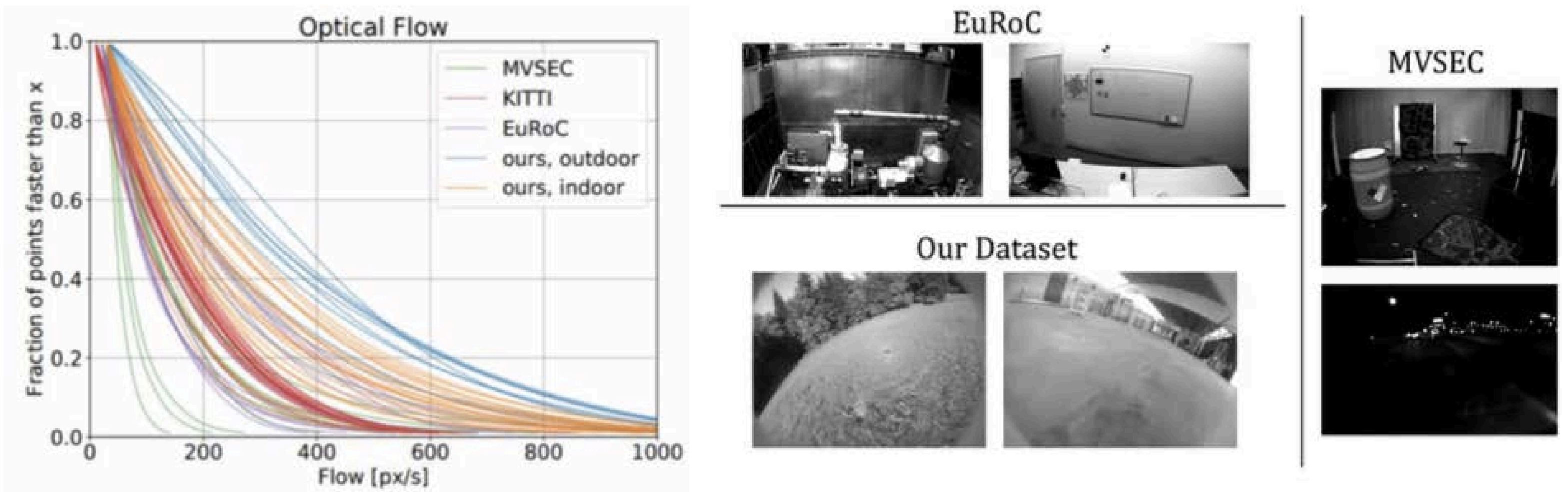
High Dynamic Range

Motion blur

Latency

Invited talk: Prof. Davide Scaramuzza

- Robustness
  - Quantify the level of the challenge properly
    - E.g., optical flow for the aggressiveness for vision algorithms

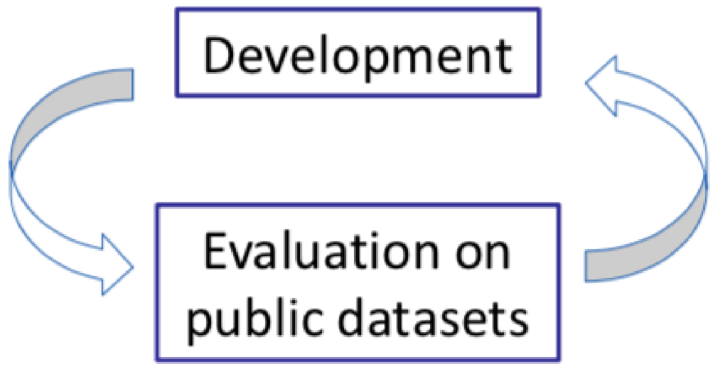


  - Repeated experiments to get statistically meaningful results
    - Success rate
    - Mean/Median error
    - ...

Invited talk: Prof. Davide Scaramuzza

- Data bias

Typical workflow of developing VO/VIO/SLAM algorithms:



As a community, we are overfitting the public dataset.

Potential problems:

- **Generalizability:** Performance on one does not guarantee to generalize to others
  - E.g., KITTI → low frame rate, not friendly for direct methods
- **Old datasets (e.g., KITTI) are already saturated:**
  - It becomes more and **more difficult to tell whether we are making real progress** or just overfitting the datasets.
  - E.g., **does 1 or 2 cm improvement in RMSE** over a 100 meter trajectory **really mean something**?

KAIST

URL URBAN ROBOTICS LAB

Invited talk: Prof. Davide Scaramuzza

- UZH-FPV Drone Racing Dataset
  - Contains data recorded by a drone flying up to over 20m/s indoors and outdoors frown by a professional pilot. Contains frames, events, IMU, and Ground Truth from a Robotic Total Station
  - Video link : https://youtu.be/G5w4ZcEzvoo

KAIST

URL URBAN ROBOTICS LAB

Invited talk: Prof. Davide Scaramuzza

- Conclusion
  - Current SLAM evaluation
    - **Many existing metrics**, reflecting different aspects of the algorithms
    - Evaluation is a non-trivial task: **many little details affect the results**
    - Check out our **tutorial and toolbox**: https://github.com/uzh-rpg/rpg_trajectory_evaluation [Zhang, IROS'18]
  - How to push forward SLAM research
    - Take robustness into consideration
    - Do not stick to a few datasets: use more diverse ones
    - Take advantage of photo **realistic simulators**, but if you doo, please share the datasets!
    - Take the chance to
      - Actively change the parameters of the algorithm to improve robustness
      - Work on new sensors (e.g., event cameras)
        - Survey paper on event cameras: http://rpg.ifi.uzh.ch/docs/EventVisionSurvey.pdf
        - Event camera dataset: http://rpg.ifi.uzh.ch/davis_data.html
        - MVSEC dataset: https://daniilidis-group.github.io/mvsec/
        - UZH-FPV Drone Racing dataset: http://rpg.ifi.uzh.ch/uzh-fpv.html
        - Event-camera Simulator (ESIM): https://github.com/uzh-rpg/rpg_esim

Invited talk: Prof. Davide Scaramuzza

- Checklist for reproducible SLAM results

Running experiments

- What are the crucial parameters (# features, # keyframes, etc.)?
- Does the starting and ending time in the dataset have an obvious impact on the results?
- Am I running the experiments in a real-time setup (or processing new measurements only when the previous processing is done)?
- Have I ran the algorithm multiple times to have repeatable results/meaningful statistics?

Reporting results

Accuracy
- Am I reporting the accuracy of real-time poses or refined poses?
- Absolute error: how is the trajectory aligned with the groundtruth?
- Which frames are evaluated? All the frame or only keyframes?

Efficiency
- What are the experimental platforms?
- What are the exact starting and end point of the processing time?
- Is there any special optimization used that has a big impact?

Invited talk: Prof. Davide Scaramuzza

- How should we report results in papers?

What not to write in a paper:

"We aligned the estimated trajectory with the groundtruth and calculated the Root Mean Square Error (RMSE) to indicate the estimation accuracy." [Author names hidden for privacy]

- What type of alignment was used?
- What method was used for calculate the alignment transformation?

How to write in a paper:

"To obtain a measure of accuracy of the different approaches, we aligned the final trajectory of keyframes with the ground-truth trajectory using the least-squares approach proposed in [Umeyama, 1991]. Since scale cannot be recovered using a single camera, we also rescaled the estimated trajectory to best fit with the ground-truth trajectory. Subsequently, we computed the Euclidean distance between the estimated and ground-truth keyframe poses and compute the mean, median, and Root Mean Square Error (RMSE) in meters." [Author names hidden for privacy]

"We used the relative error metrics proposed in [KITTI] to obtain error statistics. The metric evaluates the relative error by averaging the drift over trajectory segments of different length {10; 40; 90; 160; 250; 360 } meter." [Author names hidden for privacy]

→ Necessary references and details.

InteriorNet

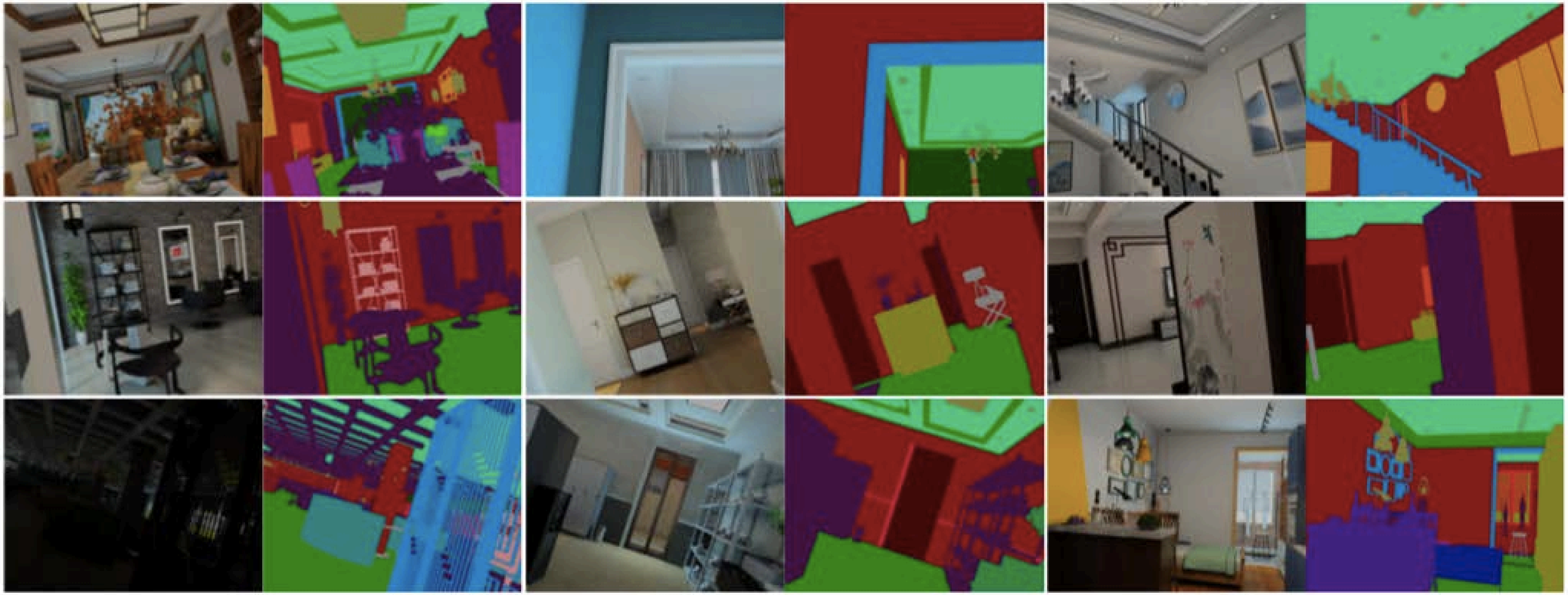
Li, Wenbin, et al. "InteriorNet: Mega-scale multi-sensor photo-realistic indoor scenes dataset." *arXiv preprint arXiv:1809.00716* (2018).

- Mega-scale, Multi-sensor, Photo-realistic Indoor Scene Dataset

InteriorNet

Li, Wenbin, et al. "InteriorNet: Mega-scale multi-sensor photo-realistic indoor scenes dataset." *arXiv preprint arXiv:1809.00716* (2018).

- Mega-scale, Multi-sensor, Photo-realistic Indoor Scene Dataset
- Associated with NYU40 labels: wall, floor, bed, chair, table etc.



AI Habitat

Savva, Manolis, et al. "Habitat: A platform for embodied ai research." *arXiv preprint arXiv:1904.01201* (2019).

- AI Habitat enables training of embodied AI agents (virtual robots) in a highly photorealistic & efficient 3D simulator

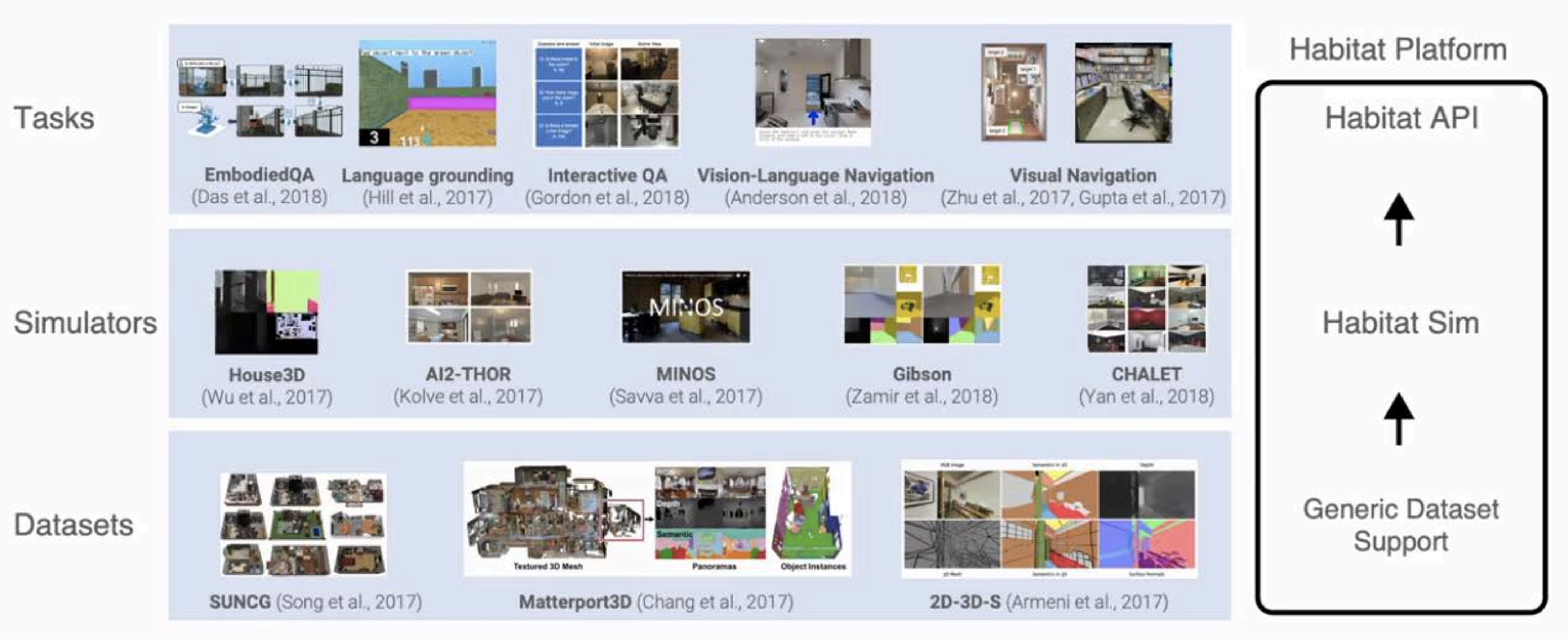
Habitat-Sim

- Configurable 3D simulator (C++ with pybind11)
- Generic 3D dataset support (SUNCG, MP3D, +more)
- Fast: over 1,000 FPS single-threaded
  10,000 FPS multi-process (single GPU)

AI Habitat

Savva, Manolis, et al. "Habitat: A platform for embodied ai research." *arXiv preprint arXiv:1904.01201* (2019).

- AI Habitat enables training of embodied AI agents (virtual robots) in a highly photorealistic & efficient 3D simulator



For more information

- ICRA 2019 workshop/tutorial page link
- InteriorNet homepage
- Habitat homepage