

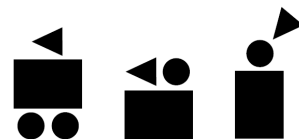
Evaluation of RGB-D SLAM systems

Jürgen Sturm¹, Stéphane Magnenat²,
Nikolas Engelhard³, François Pomerleau², Francis Colas²,
Daniel Cremers¹, Roland Siegwart², Wolfram Burgard³

¹Technical University of Munich / ²ETH Zurich / ³University of Freiburg



Computer
Vision and
Pattern
Recognition



Autonomous Systems Lab

ETH zürich



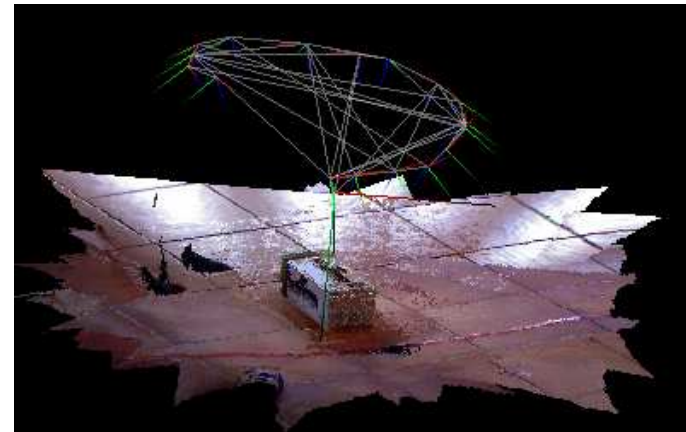
AiS Autonomous
Intelligent
Systems

Motivation

- Simultaneous localization and mapping
- Laser-based SLAM [Montemerlo/Olson/Frese/Grisetti/Birk/Nuechter/...]
- Visual (monocular) SLAM [Nister/Davison/Konolige/Strasdat/...]
- Stereo SLAM [Konolige/...]
- RGB-D SLAM [Henry/Engelhard/Newcombe/Fiorario/Rezaiifar/...]



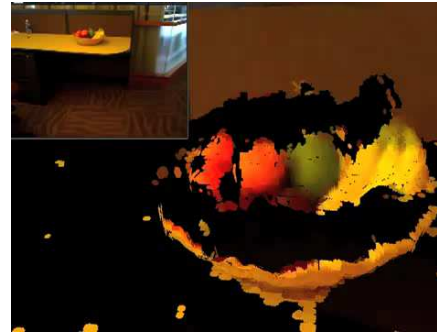
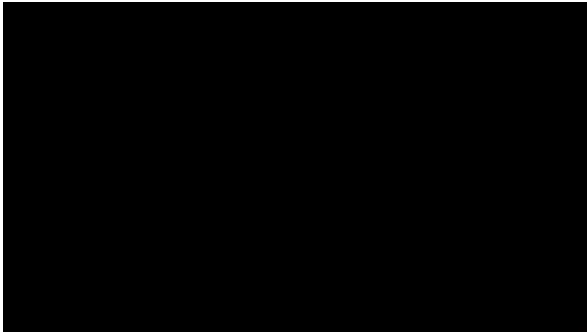
Microsoft Kinect



RGB-D SLAM (open source, try it out!)
[Eurobotics-RGBD `11]

More RGB-D SLAM Results

- Peter Henry, Xiaofeng Ren, Dieter Fox [ISER '10]



- Nikolas Engelhard, Felix Endres, Jürgen Hess, Jürgen Sturm, Wolfram Burgard [Eurobotics-RGBD '11]



- Nicola Fioraio, Kurt Konolige [RSS-RGBD '11]
- Ramin Rezaifar, Stefano Soatto, Ashwin Swaminathan, Qi Xue, Piyush Sharma [RSS-RGBD '11]
- Richard Newcombe, Andrew Davison
- ...

But how can we evaluate (and compare) these approaches?

- 2D laser-based SLAM

- Intel Dataset
- Freiburg Dataset

- 3D laser+visual SLAM

- Newcollege Dataset

- RGB-D SLAM?

- Visual inspection
- Ad-hoc evaluations
- No objective/scientific comparisons exist yet!

➔ Let's create such a dataset!



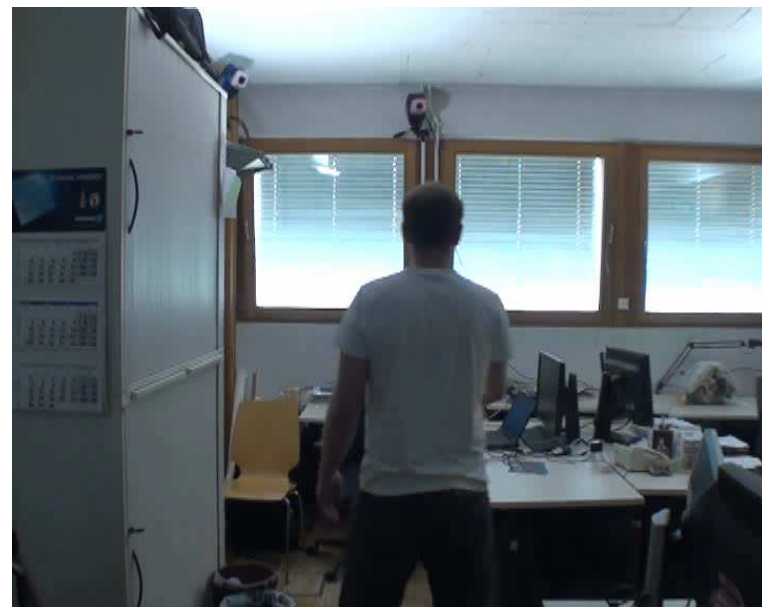
[Haehnel et al.]

1st step: Record the data

- Motion capture system
 - TF (100 Hz)
- Microsoft Kinect
 - Color images (30 Hz)
 - Depth maps (30 Hz)
 - Point clouds (30 Hz)
 - IMU (500 Hz)
- Calibration
- Record different (office) scenes
 - Various scenes (xyz, desk, room, teddy, plant,...)
 - Variations in camera speed
 - Variations in camera motion

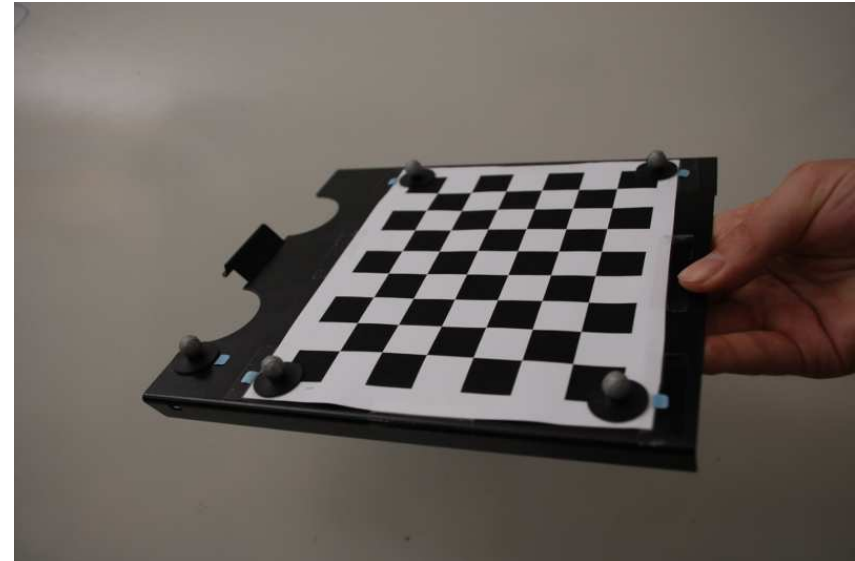
1st step: Record the data

- **Motion capture system**
- Microsoft Kinect
- Calibration (Mocap–Kinect)
- Record different (office) scenes



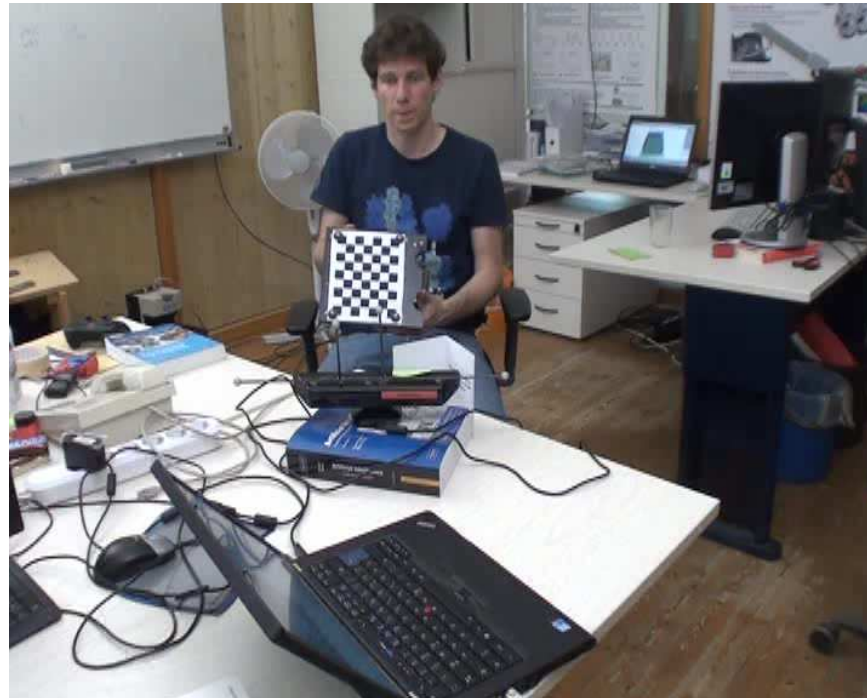
1st step: Record the data

- Motion capture system
- **Microsoft Kinect**
- **Calibration (Mocap–Kinect)**
- Record different (office) scenes



1st step: Record the data

- Motion capture system
- **Microsoft Kinect**
- **Calibration (Mocap–Kinect)**
- Record different (office) scenes



1st step: Record the data

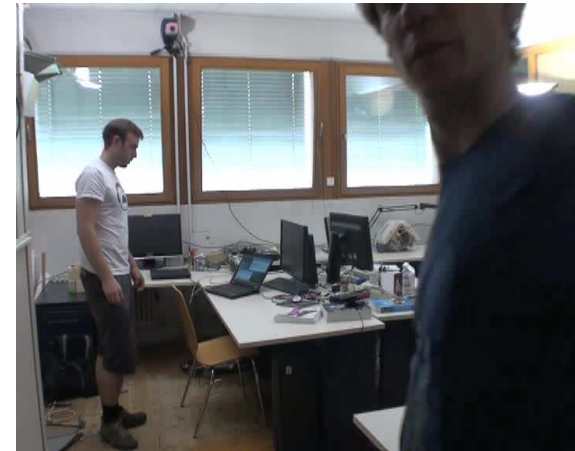
- Motion capture system
- Microsoft Kinect
- Calibration (Mocap–Kinect)
- **Record different (office) scenes**



File: XYZ



File: 360



File: Room ...

1st step: Record the data

- Motion capture system
- Microsoft Kinect
- Calibration (Mocap–Kinect)
- **Record different (office) scenes**



File: XYZ



File: Desk

Resulting dataset

- Freely available (Creative Commons)
<https://cvpr.in.tum.de/research/datasets/rgbd-dataset>
- ROS bag files (and: zipped list of pngs+traj)
- Movies for visual inspection
 - Color camera
 - Depth map
 - External camera
 - Reconstructed scene

2nd step: Create a benchmark

1. Download the dataset
2. Run your favorite (RGB-D) SLAM approach
3. Compare estimated trajectory with ground truth trajectory

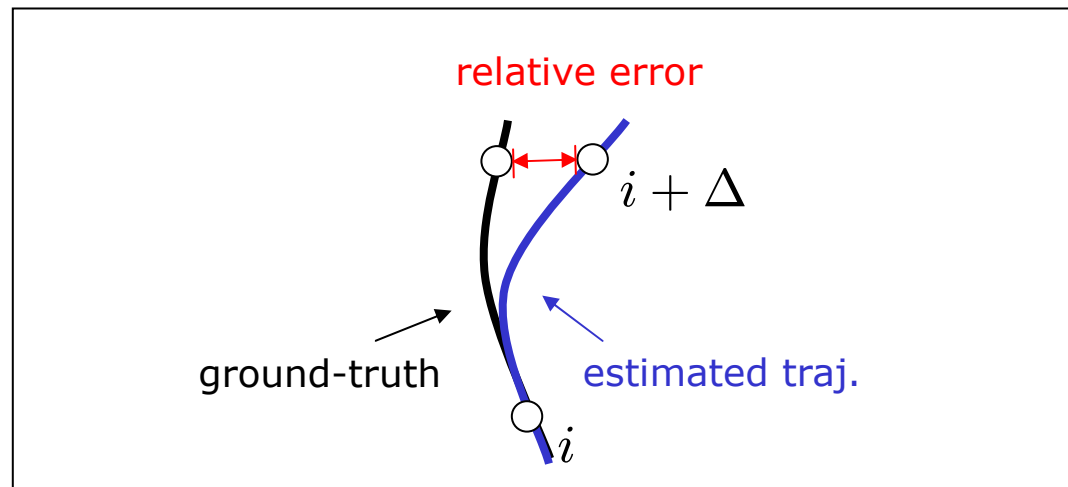
➔ What evaluation metric should we use?

- Automatic evaluation tool
- Avoid overfitting by keeping back second set of data

Evaluation Metric (1)

- Evaluate the quality of the camera trajectory (differences in relative poses)

$$error = \frac{1}{n} \sum_{i=1}^n [(\hat{\mathbf{x}}_{i+\Delta} \ominus \hat{\mathbf{x}}_i) \ominus (\mathbf{x}_{i+\Delta} \ominus \mathbf{x}_i)]^2 \quad [\text{Kümmerle et al., 2009}]$$



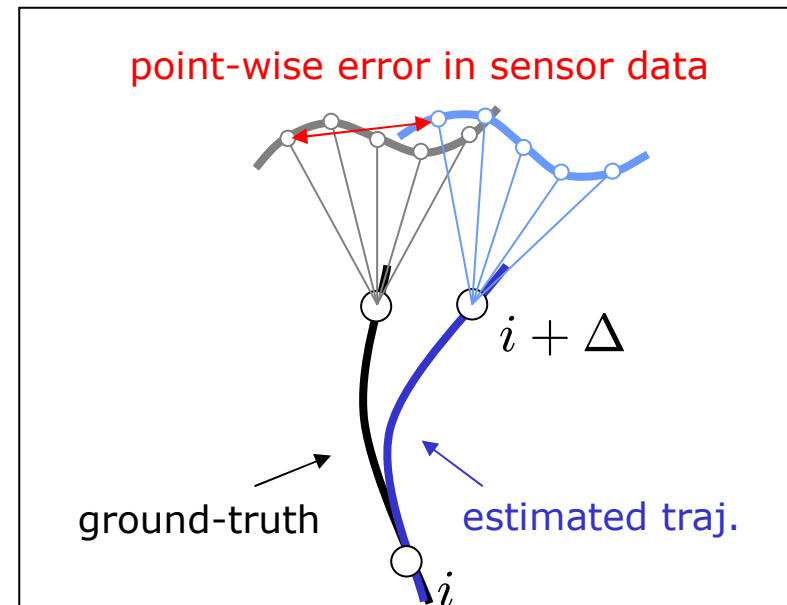
$\hat{\mathbf{x}}_{1:n}$ estimated traj. $\mathbf{x}_{1:n}$ ground-truth traj. $P_{1:n}$ sequence of sensor data (3D point clouds)

Evaluation Metric (2)

Alternative:

- Evaluate the quality of the resulting map (differences in the projected sensor data)

$$error = \frac{1}{n} \sum_{i=1}^n \sum_{u=1}^{640} \sum_{v=1}^{480} [(\hat{\mathbf{x}}_{i+\Delta} \ominus \hat{\mathbf{x}}_i) P_i(u, v) - (\mathbf{x}_{i+\Delta} \ominus \mathbf{x}_i) P_i(u, v)]^2$$



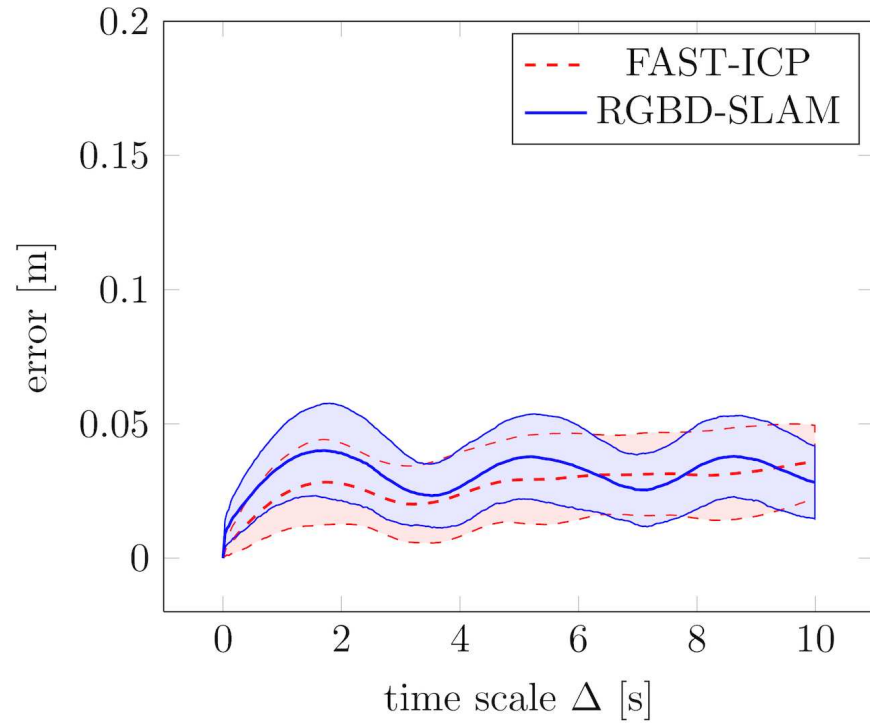
$\hat{\mathbf{x}}_{1:n}$ estimated traj. $\mathbf{x}_{1:n}$ ground-truth traj. $P_{1:n}$ sequence of sensor data (3D point clouds)

Preliminary evaluation results

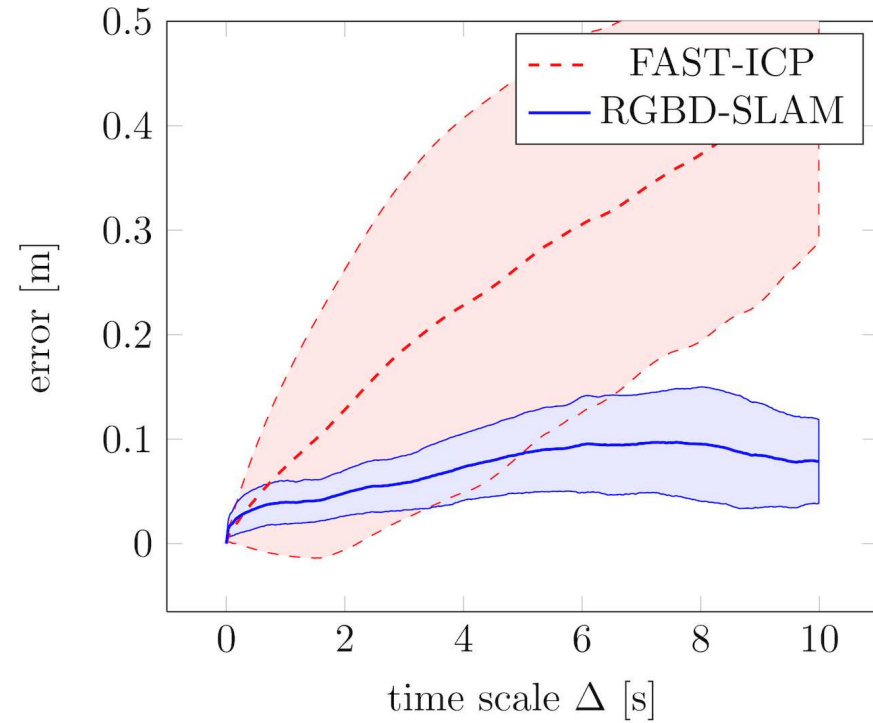
- Two systems
 - Freiburg RGB-D SLAM
 - ETH Fast-ICP Tracker
- Two datasets
 - XYZ: small volume, motion along the axes
 - Volume of camera trajectory: 0.5m x 0.7m x 0.4m
 - Average camera speed: 0.25m/s
 - Good for debugging
 - Desk: motion across four desks, contains loops
 - Volume of camera trajectory: 2.4m x 1.3m x 0.7m
 - Average camera speed: 0.41m/s
 - Good for SLAM evaluation
- Evaluate camera trajectory (metric 1)

Preliminary evaluation results

Dataset: XYZ



Dataset: Desk



Conclusions

- Dataset for RGB-D SLAM
 - Ground-truth camera poses
 - Evaluation criteria
 - Automatic evaluation tools
- ➔ Benchmark for RGB-D SLAM evaluation

*We plan to publish a joint paper on the evaluation of RGBD-SLAM algorithms. Contact us if you are interested!
(sturmju@in.tum.de)*

Discussion

- Motion blur + rolling shutter effects due to fast camera motions
- ➔ Will record additional data with slower camera motion in July
- Any suggestions for the scenes?
- What do you need?

Thank you!