Visual Navigation for Flying Robots

Experimentation, Evaluation and Benchmarking

Dr. Jürgen Sturm

Agenda for Today

- Course Evaluation

- Scientific research: The big picture
- Best practices in experimentation
- Datasets, evaluation criteria and benchmarks

- Time for questions

Course Evaluation

- Much positive feedback – thank you!!!
- We are also very happy with you as a group. Everybody seemed to be highly motivated!
- Suggestions for improvements (from course evaluation forms)
  - Workload was considered a bit too high
    → ECTS have been adjusted to 6 credits
  - ROS introduction lab course would be helpful
    → Will do this next time
- Any further suggestions/comments?

Scientific Research – General Idea

1. Observe phenomena
2. Formulate explanations and theories
3. Test them

Scientific Research – Methodology

1. Generate an idea
2. Develop an approach that solves the problem
3. Demonstrate the validity of your solution
4. Disseminate your results
5. At all stages: iteratively refine

Scientific Research in Student Projects

- How can you get involved in scientific research during your study?

Scientific Research in Student Projects

- How can you get involved in scientific research during your study?
  - Bachelor lab course (10 ECTS)
  - Bachelor thesis (15 ECTS)
  - Graduate lab course (10 ECTS)
  - Interdisciplinary project (16 ECTS)
  - Master thesis (30 ECTS)
  - Student research assistant (10 EUR/hour, typically 10 hours/week)

Step 1: Generate the Idea

- Be creative
- Follow your interests / preferences
- Examples:
  - Research question
  - Challenging problem
  - Relevant application
  - Promising method (e.g., try to transfer method from another field)

Step 1b: Find related work

- There is **always** related work
- Find related research papers
  - Use Google scholar, paper repositories, …
  - Navigate the citation network
  - Read survey articles
- Browse through (recent) text books
- Ask your professor, colleagues, …
- It's very unlikely that somebody else has already perfectly solved exactly your problem, so don't worry! Technology evolves very fast…

Step 2: Develop a Solution

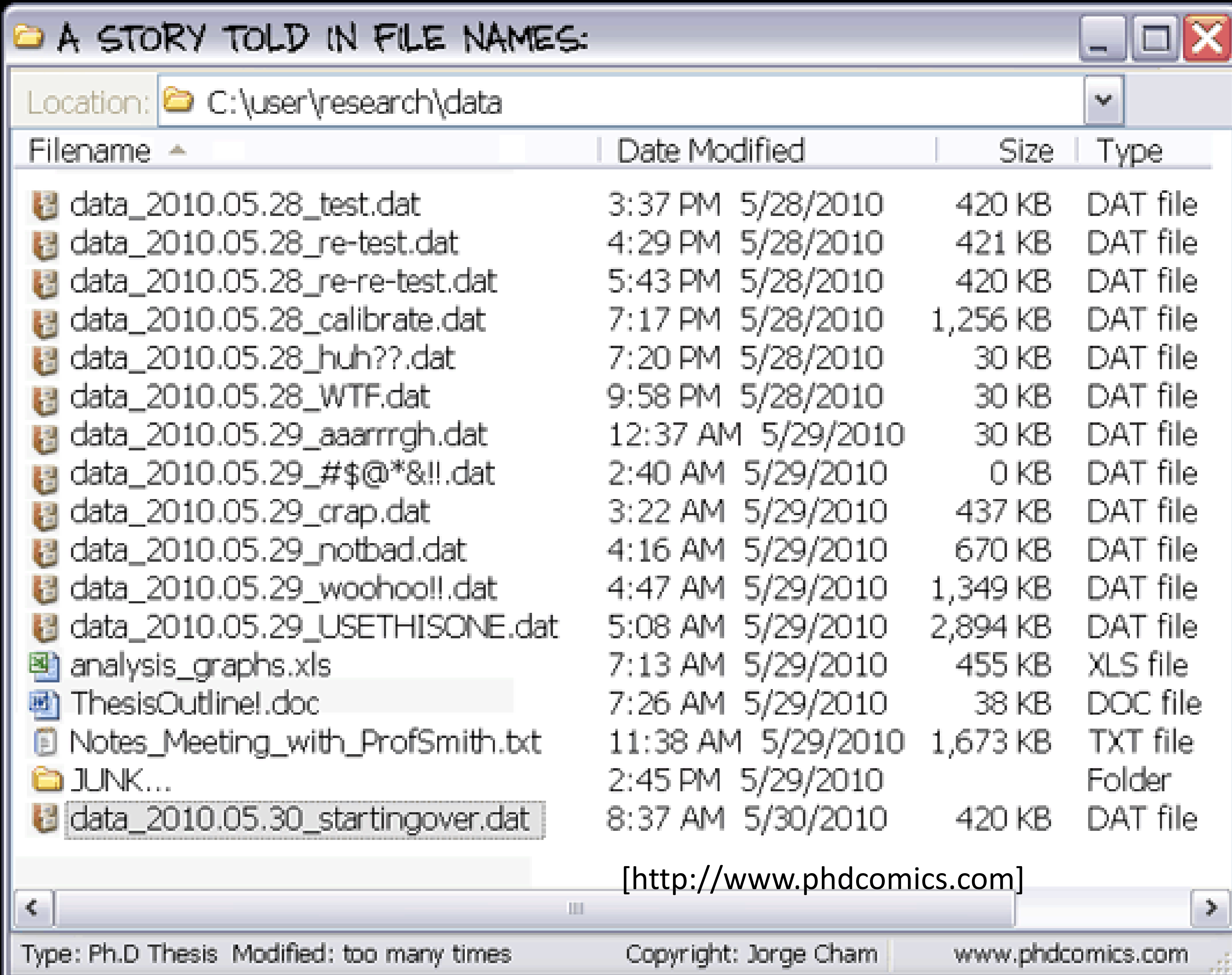
- Practitioner
  - Start programming
  - Realize that it is not going to work, start over, …
  - When it works, formalize it (try to find out why it works and what was missing before)
  - Empirically verify that it works
- Theorist
  - Formalize the problem
  - Find suitable method
  - (Theoretically) prove that it is right
  - (If needed) implement a proof-of-concept

Step 3: Validation

- What are your claims?
- How can you prove them?
  - Theoretical proof (mathematical problem)
  - Experimental validation
    - Qualitative (e.g., video)
    - Quantitative (e.g., many trials, statistical significance)
- Compare and discuss your results with respect to previous work/approaches

Step 4: Dissemination

- Good solution/expertise alone is not enough
- You need to convince other people in the field
- Usual procedure:
  1. Write research paper (usually 6-8 pages) **3-6 month**
  2. Submit PDF to an international conference or journal
  3. Paper will be peer-reviewed **3-6 month**
  4. Improve paper (if necessary)
  5. Give talk or poster presentation at conference **15 min.**
  6. Optionally: Repeat step 1-5 until PhD ☺ **3-5 years**



[http://www.phdcomics.com]

Step 5: Refinement

- Discuss your work with
  - Your colleagues
  - Your professor
  - Other colleagues at conferences
- Improve your approach and evaluation
  - Adopt notation to the standard
  - Get additional references/insights
  - Conduct more/additional experiments
- Simplify and generalize your approach
- Collaborate with other people (in other fields)

Scientific Research

- This was the big picture
- Today's focus is on best practices in experimentation
- **What do you think are the (desired) properties of a good scientific experiment?**

What are the desired properties of a good scientific experiment?

- Reproducibility / repeatability
  - Document the experimental setup
  - Choose (and motivate) an your evaluation criterion
- Experiments should allow you to validate/falsify competing hypotheses

Current trends:

- Make data available for review and criticism
- Same for software (open source)

Challenges

- Reproducibility is sometimes not easy to guarantee
- Any ideas why?

Challenges

- Randomized components/noise (beat with the law of large numbers/statistical tests)
- Experiment requires special hardware
  - Self-built, unique robot
  - Expensive lab equipment
  - …
- Experiments cost time
- “(Video) Demonstrations will suffice”
- Technology changes fast

Benchmarks

- Effective and affordable way of conducting experiments
- Sample of a task domain
- Well-defined performance measurements
- Widely used in computer vision and robotics
- **Which benchmark problems do you know?**

Example Benchmark Problems

Computer Vision

- Middlebury datasets (optical flow, stereo, …)
- Caltech-101, PASCAL (object recognition)
- Stanford bunny (3d reconstruction)

Robotics

- RoboCup competitions (robotic soccer)
- DARPA challenges (autonomous car)
- SLAM datasets

Image Denoising: Lenna Image

- 512x512 pixel standard image for image compression and denoising
- Lena Söderberg, Playboy magazine Nov. 1972
- Scanned by Alex Sawchuck at USC in a hurry for a conference paper

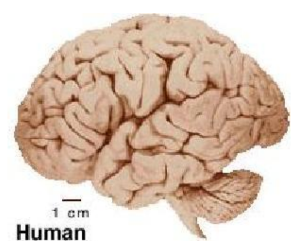
   

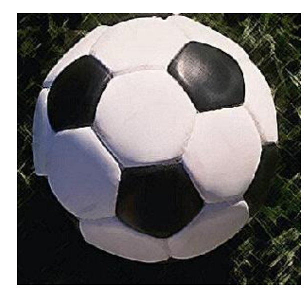
http://www.cs.cmu.edu/~chuck/lennapg/

Object Recognition: Caltech-101

- Pictures of objects belonging to 101 categories
- About 40-800 images per category
- Recognition, classification, categorization





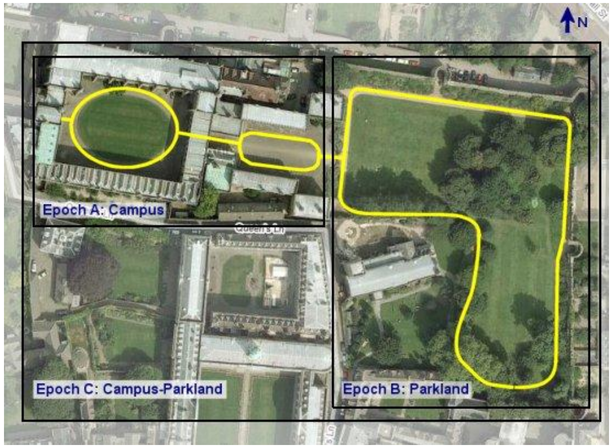
RoboCup Initiative

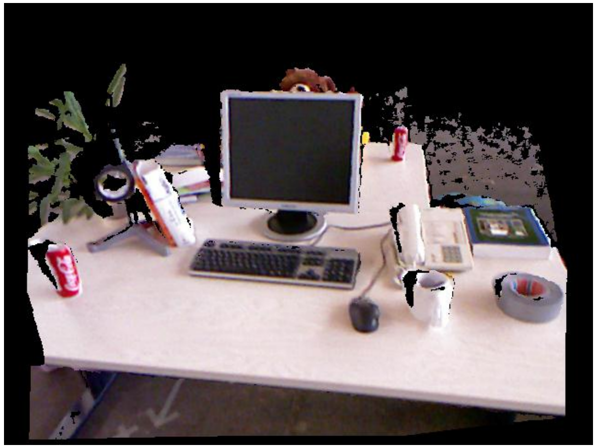
- Evaluation of full system performance
- Includes perception, planning, control, …
- Easy to understand, high publicity
- "By mid-21st century, a team of fully autonomous humanoid robot soccer players shall win the soccer game, complying with the official rule of the FIFA, against the winner of the most recent World Cup."

RoboCup Initiative

SLAM Evaluation

- Intel dataset: laser + odometry [Haehnel, 2004]
- New College dataset: stereo + omni-directional vision + laser + IMU [Smith et al., 2009]
- TUM RGB-D dataset [Sturm et al., 2011/12]
- …





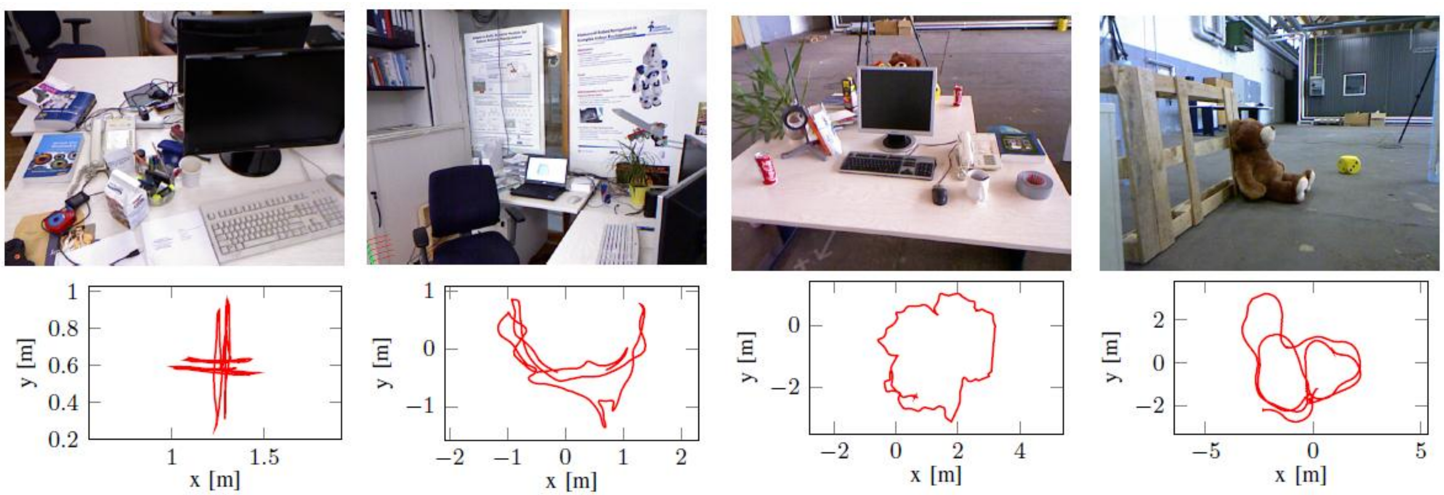
TUM RGB-D Dataset

[Sturm et al., RSS RGB-D 2011; Sturm et al., IROS 2012]

- RGB-D dataset with ground truth for SLAM evaluation
- Two error metrics proposed (relative and absolute error)
- Online + offline evaluation tools
- Training datasets (fully available)
- Validation datasets (ground truth not publicly available to avoid overfitting)

Recorded Scenes

- Various scenes (handheld/robot-mounted, office, industrial hall, dynamic objects, …)
- Large variations in camera speed, camera motion, illumination, environment size, …



(a) fr1/xyz (b) fr1/room (c) fr2/desk (d) fr2/slam

Dataset Acquisition

- Motion capture system
  - Camera pose (100 Hz)
- Microsoft Kinect
  - Color images (30 Hz)
  - Depth maps (30 Hz)
  - IMU (500 Hz)
- External video camera (for documentation)

Motion Capture System

- 9 high-speed cameras mounted in room
- Cameras have active illumination and pre-process image (thresholding)
- Cameras track positions of retro-reflective markers

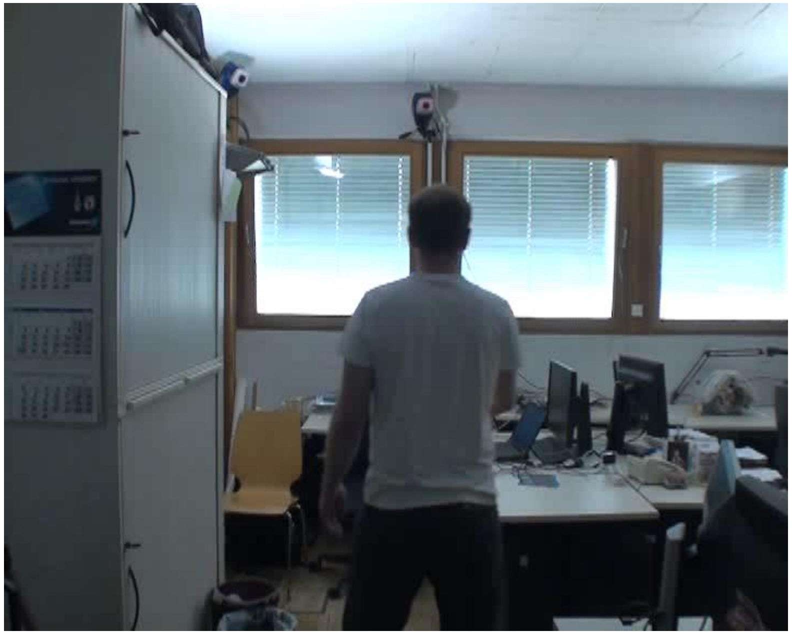
Calibration

Calibration of the overall system is not trivial:

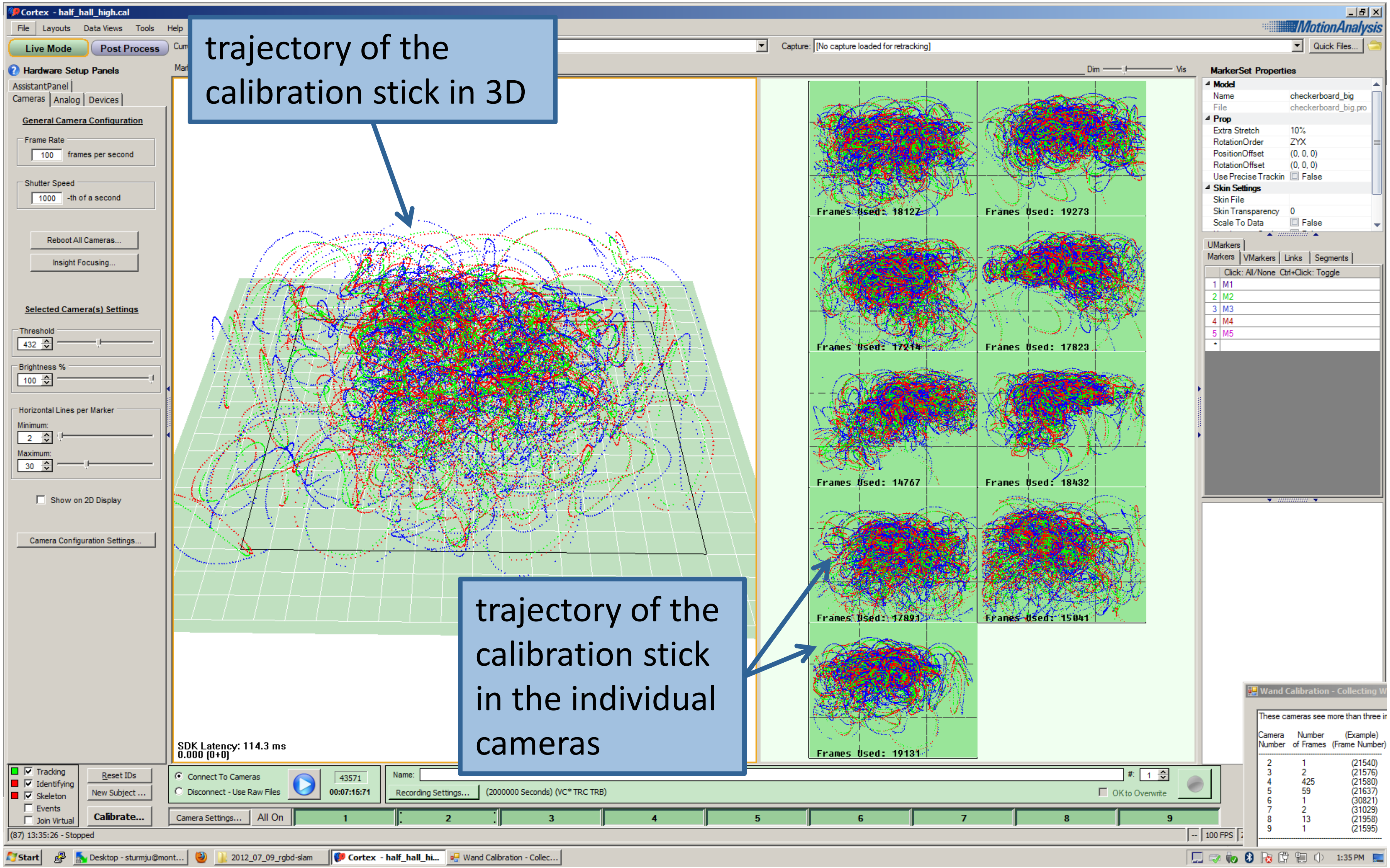
1. Mocap calibration
2. Kinect-mocap calibration
3. Time synchronization

Calibration Step 1: Mocap

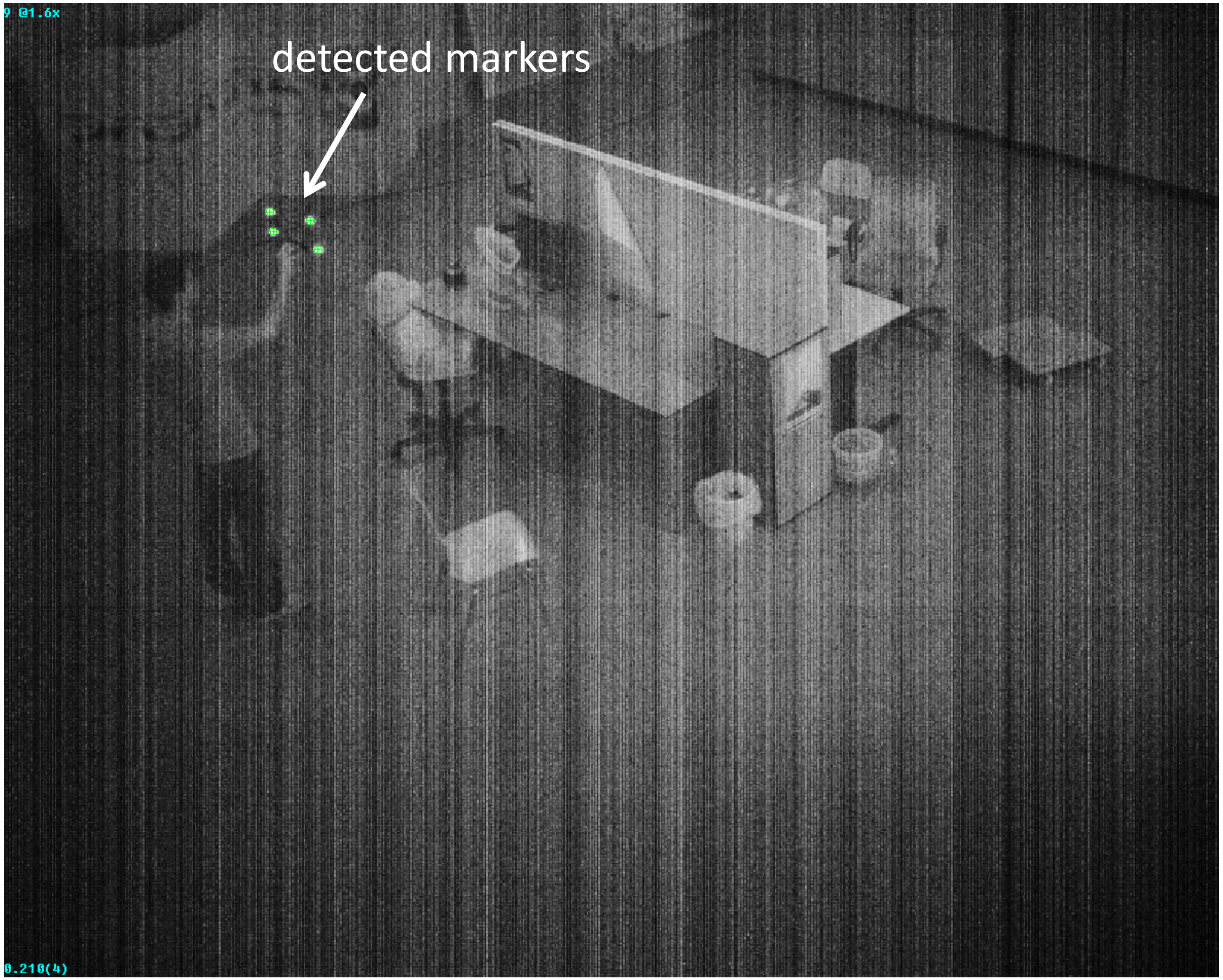
- Need at least 2 cameras for position fix
- Need at least 3 markers on object for full pose
- Calibration stick for extrinsic calibration



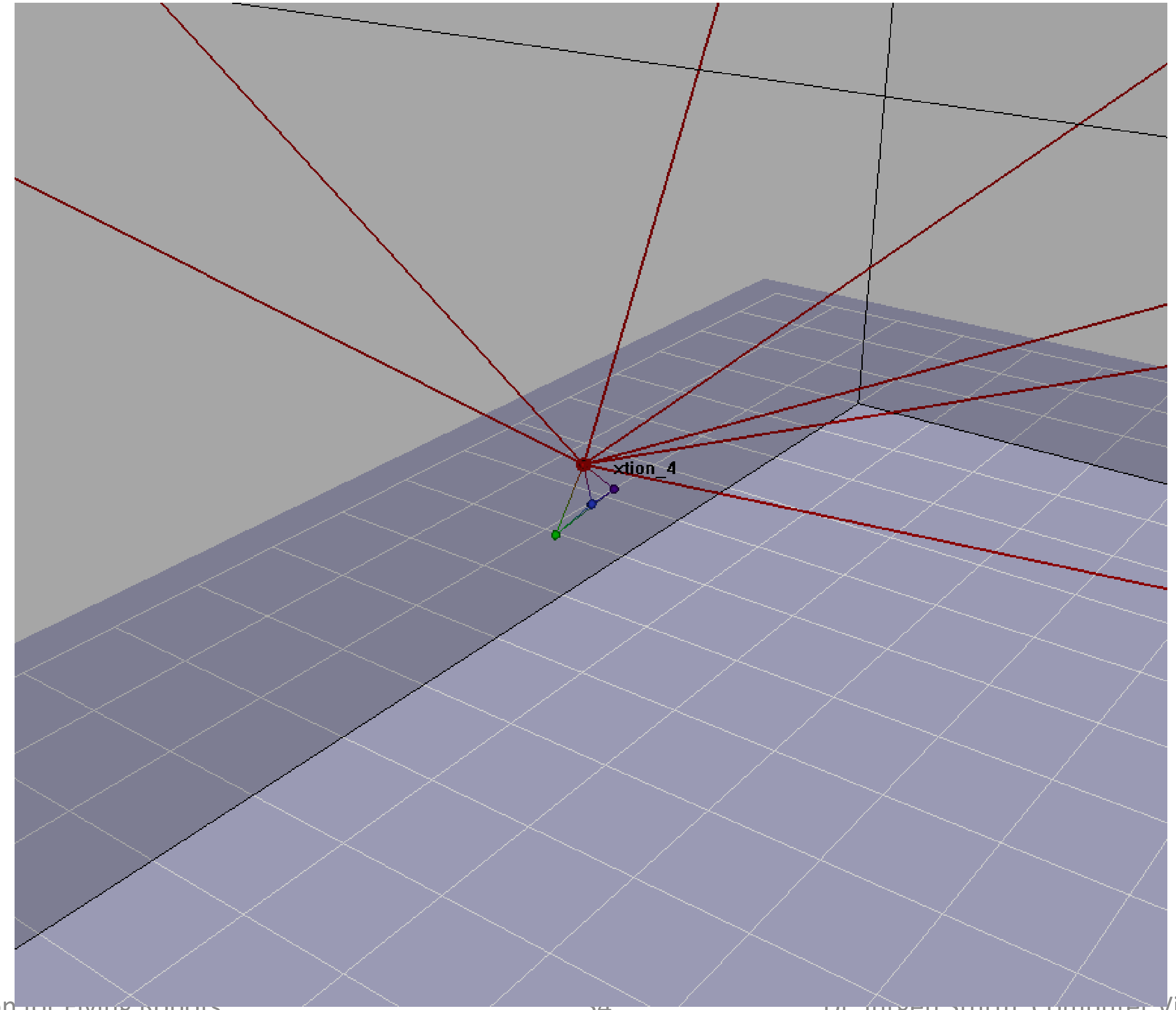
Calibration Step 1: Mocap



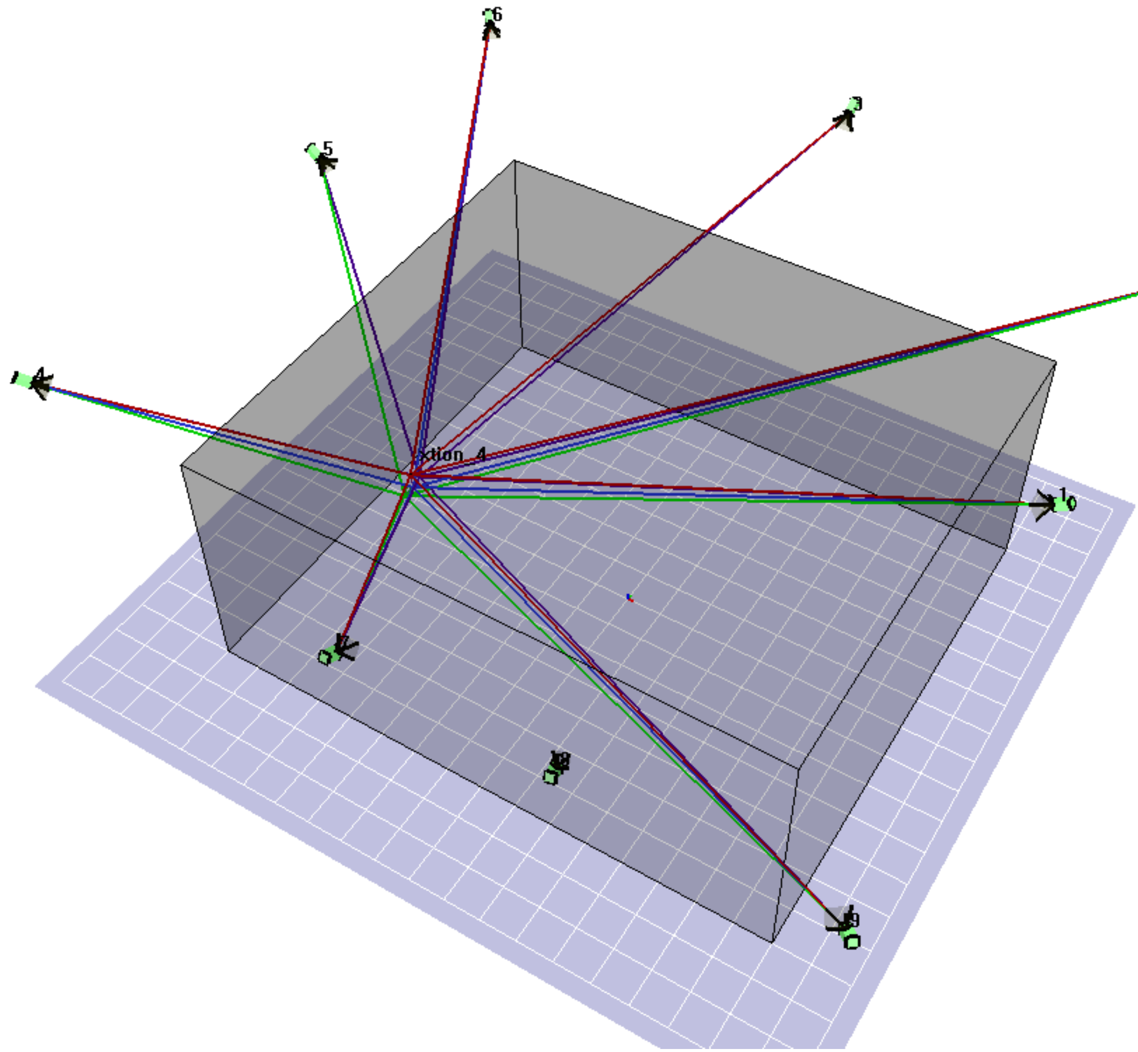
Example: Raw Image from Mocap



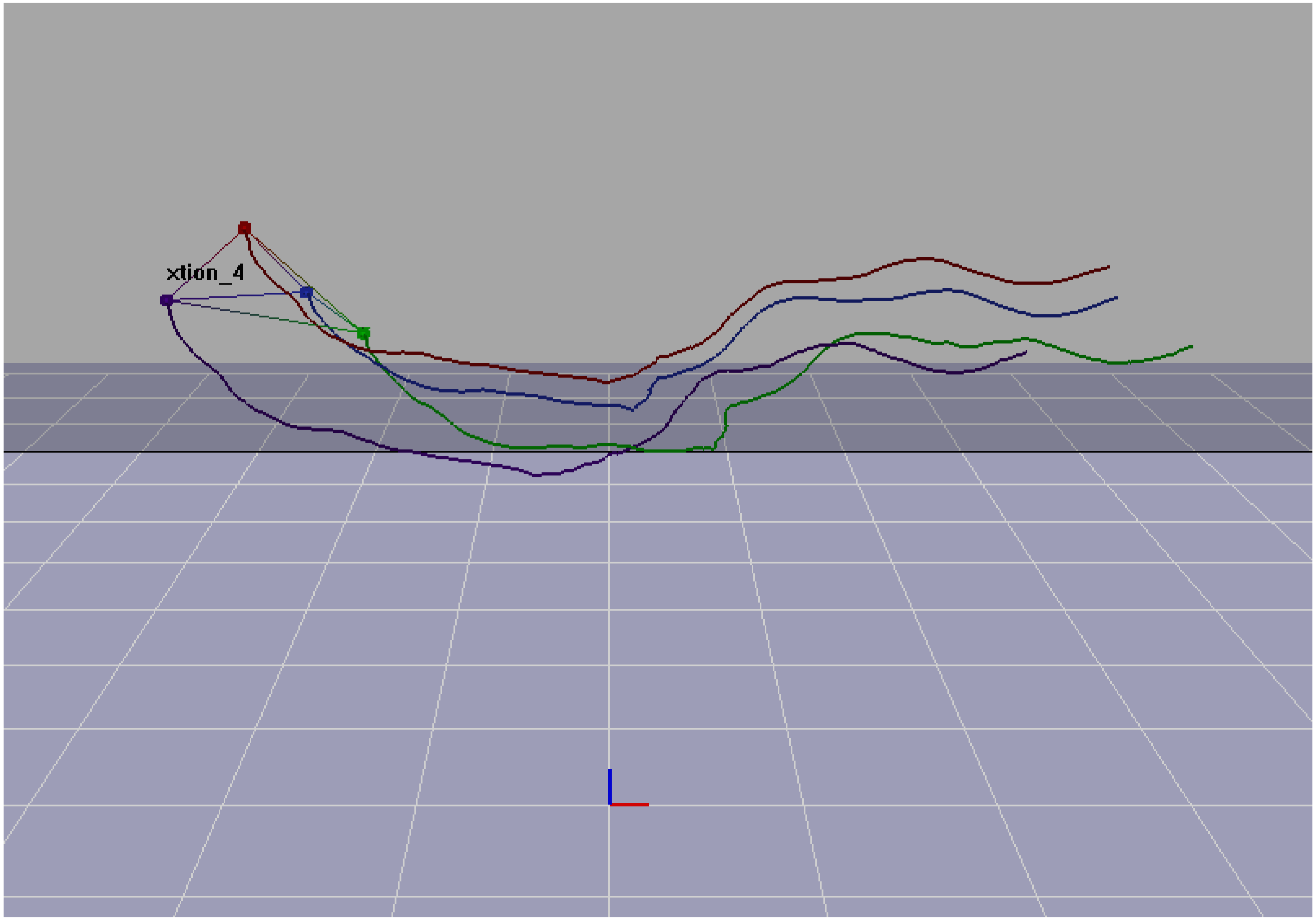
Example: Position Triangulation of a Single Marker



Example: Tracked Object (4 Markers)

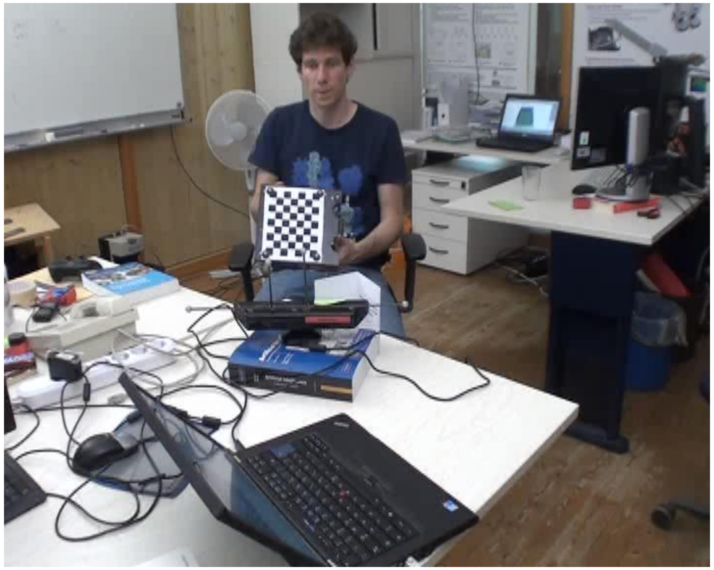


Example: Recorded Trajectory



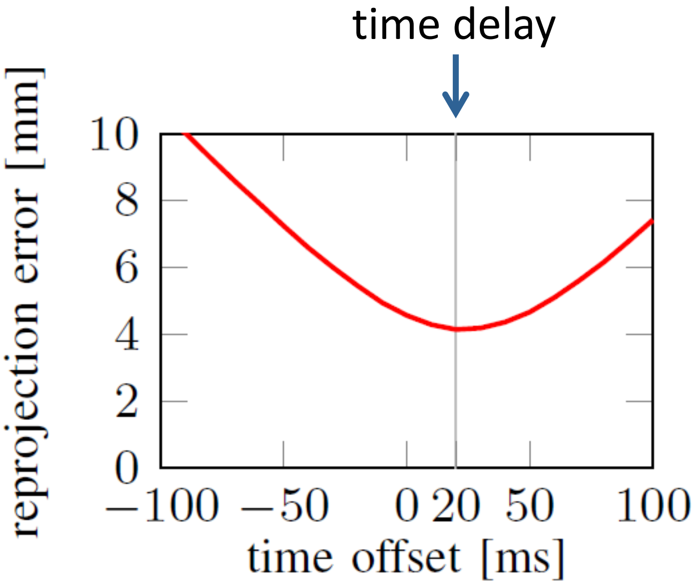
Calibration Step 2: Mocap-Kinect

- Need to find transformation between the markers on the Kinect and the optical center
- Special calibration board visible both by Kinect and mocap system (manually gauged)



Calibration Step 3: Time Synchronization

- Assume a constant time delay between mocap and Kinect messages
- Choose time delay that minimizes reprojection error during checkerboard calibration



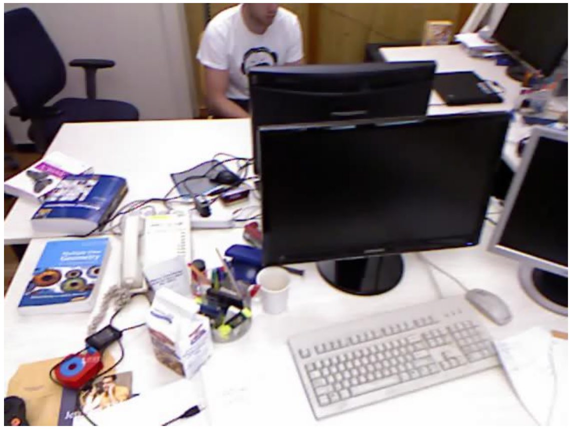
Calibration - Validation

- Intrinsic calibration
- Extrinsic calibration color + depth
- Time synchronization color + depth
- Mocap system slowly drifts (need re-calibration every hour)
- Validation experiments to check the quality of calibration
  - 2mm length error on 2m rod across mocap volume
  - 4mm RMSE on checkerboard sequence

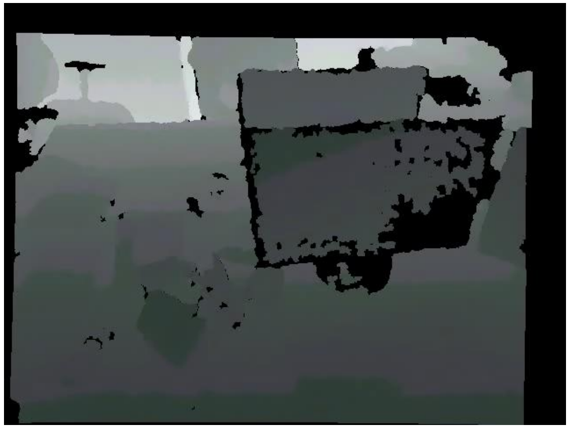
Example Sequence: Freiburg1/XYZ

External view

Color channels

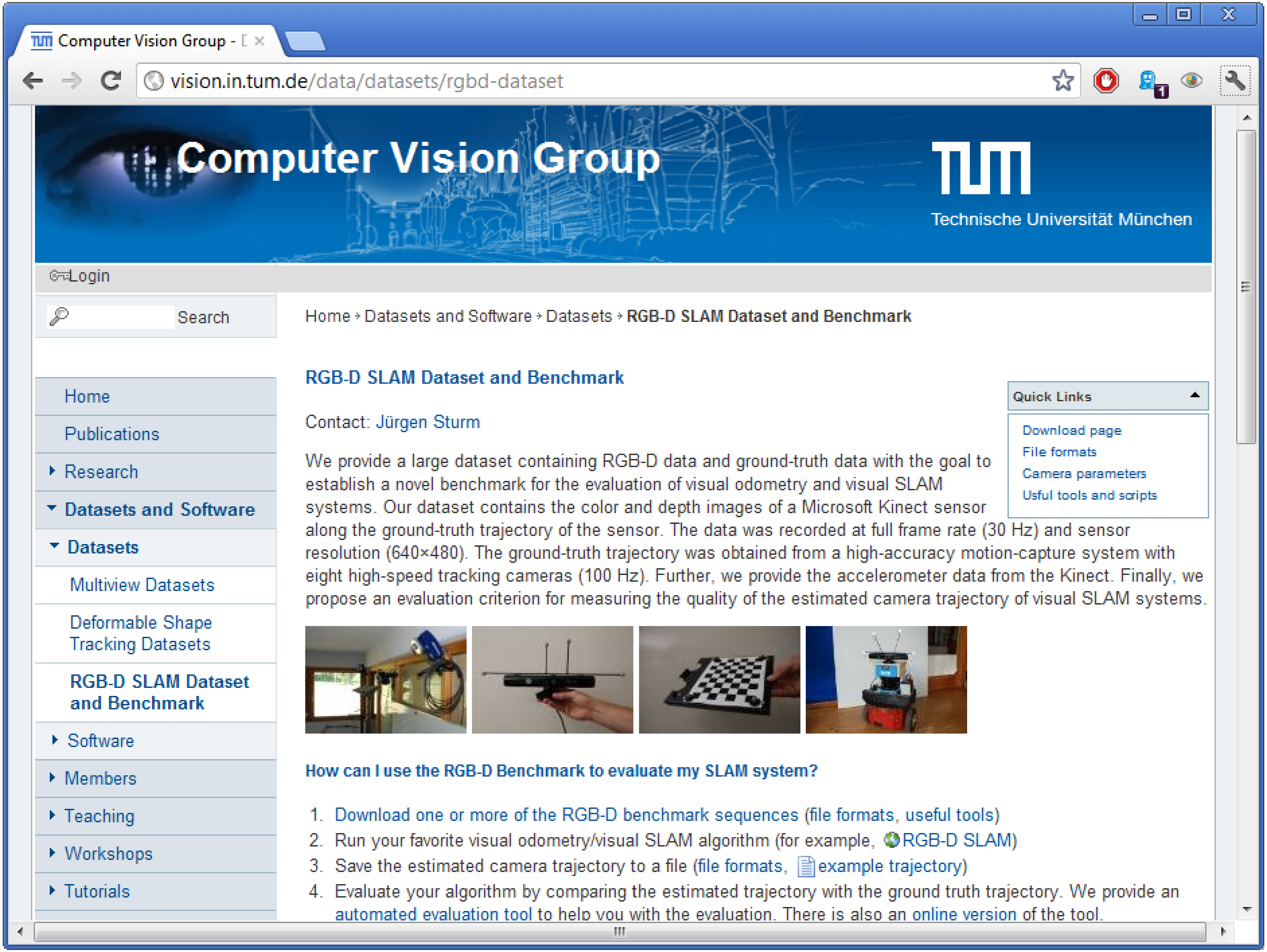


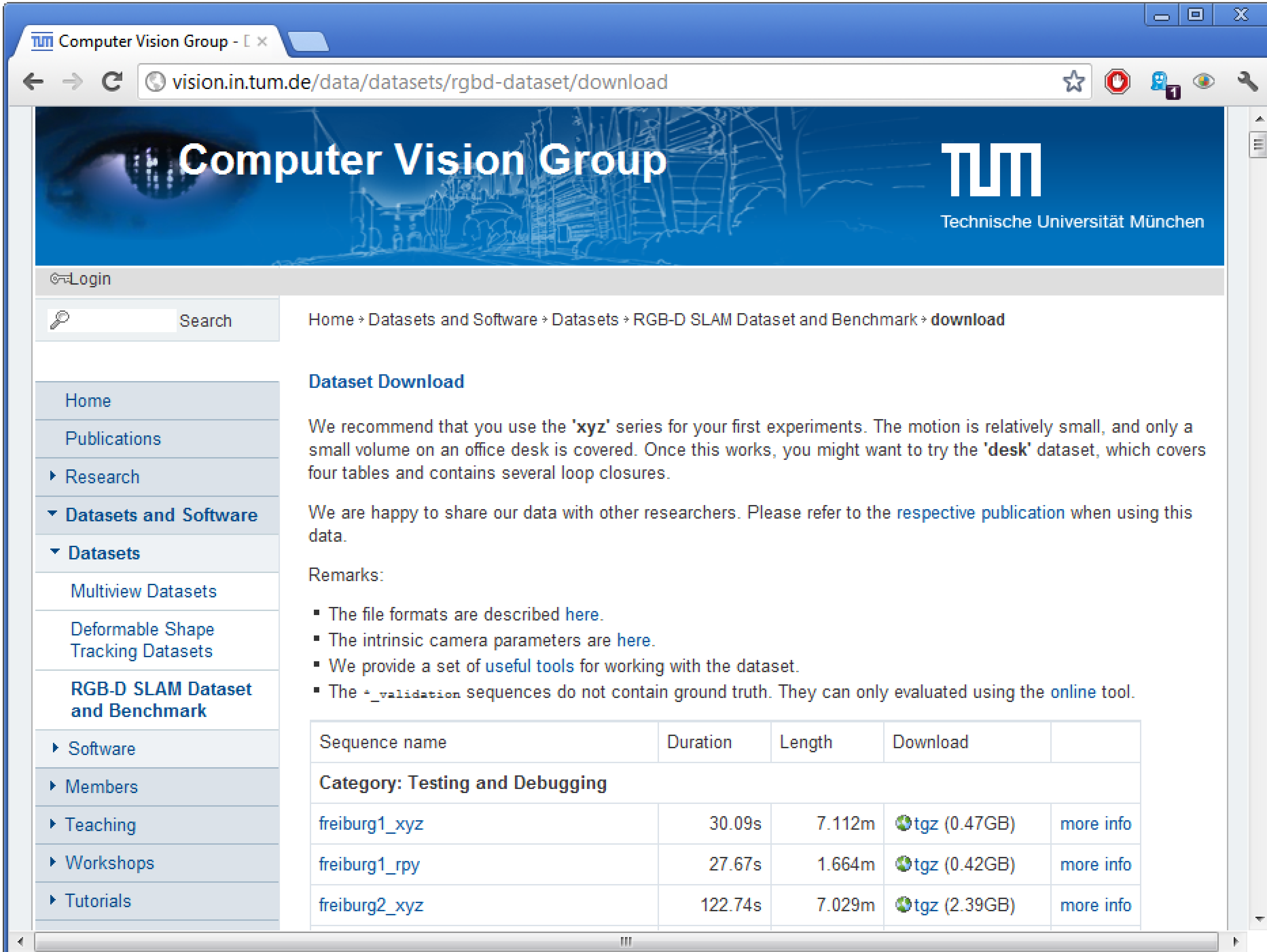
Depth channel



Sequence description (on the website):

“For this sequence, the Kinect was pointed at a typical desk in an office environment. This sequence contains only translatory motions along the principal axes of the Kinect, while the orientation was kept (mostly) fixed. This sequence is well suited for debugging purposes, as it is very simple. “


Computer Vision Group - [
vision.in.tum.de/data/datasets/rgbd-dataset
Computer Vision Group
TUM
Technische Universität München
Login
Search
Home › Datasets and Software › Datasets › RGB-D SLAM Dataset and Benchmark
Home
Publications
Research
Datasets and Software
Datasets
Multiview Datasets
Deformable Shape Tracking Datasets
RGB-D SLAM Dataset and Benchmark
Software
Members
Teaching
Workshops
Tutorials
RGB-D SLAM Dataset and Benchmark
Contact: Jürgen Sturm
Quick Links
Download page
File formats
Camera parameters
Usful tools and scripts
We provide a large dataset containing RGB-D data and ground-truth data with the goal to establish a novel benchmark for the evaluation of visual odometry and visual SLAM systems. Our dataset contains the color and depth images of a Microsoft Kinect sensor along the ground-truth trajectory of the sensor. The data was recorded at full frame rate (30 Hz) and sensor resolution (640×480). The ground-truth trajectory was obtained from a high-accuracy motion-capture system with eight high-speed tracking cameras (100 Hz). Further, we provide the accelerometer data from the Kinect. Finally, we propose an evaluation criterion for measuring the quality of the estimated camera trajectory of visual SLAM systems.
How can I use the RGB-D Benchmark to evaluate my SLAM system?
1. Download one or more of the RGB-D benchmark sequences (file formats, useful tools)
2. Run your favorite visual odometry/visual SLAM algorithm (for example, RGB-D SLAM)
3. Save the estimated camera trajectory to a file (file formats, example trajectory)
4. Evaluate your algorithm by comparing the estimated trajectory with the ground truth trajectory. We provide an automated evaluation tool to help you with the evaluation. There is also an online version of the tool.



Computer Vision Group - [×

vision.in.tum.de/data/datasets/rgbd-dataset/download

Computer Vision Group

Technische Universität München

Login

Search

Home › Datasets and Software › Datasets › RGB-D SLAM Dataset and Benchmark › **download**

- Home
- Publications
- Research
- **Datasets and Software**
- **Datasets**
  - Multiview Datasets
  - Deformable Shape Tracking Datasets
  - **RGB-D SLAM Dataset and Benchmark**
- Software
- Members
- Teaching
- Workshops
- Tutorials

Dataset Download

We recommend that you use the **'xyz'** series for your first experiments. The motion is relatively small, and only a small volume on an office desk is covered. Once this works, you might want to try the **'desk'** dataset, which covers four tables and contains several loop closures.

We are happy to share our data with other researchers. Please refer to the respective publication when using this data.

Remarks:

- The file formats are described here.
- The intrinsic camera parameters are here.
- We provide a set of useful tools for working with the dataset.
- The `*_validation` sequences do not contain ground truth. They can only evaluated using the online tool.

Sequence name	Duration	Length	Download	
Category: Testing and Debugging				
freiburg1_xyz	30.09s	7.112m	tgz (0.47GB)	more info
freiburg1_rpy	27.67s	1.664m	tgz (0.42GB)	more info
freiburg2_xyz	122.74s	7.029m	tgz (2.39GB)	more info


Computer Vision Group - [
vision.in.tum.de/data/datasets/rgbd-dataset/download
Computer Vision Group
TUM
Technische Universität München
Login
Search
Home
Publications
Research
Datasets and Software
Datasets
Multiview Datasets
Deformable Shape Tracking Datasets
RGB-D SLAM Dataset and Benchmark
Software
Members
Teaching
Workshops
Tutorials
freiburg1_xyz: RGB movie
Click here for more information
Download this movie as avi
Related movies: RGB movie and, depth movie, external camera view
n is relatively small, and only a
the 'desk' dataset, which covers
tive publication when using this
ed using the online tool.
freiburg1_rpy 27.67s 1.664m tgz (0.42GB) more info
freiburg2_xyz 122.74s 7.029m tgz (2.39GB) more info

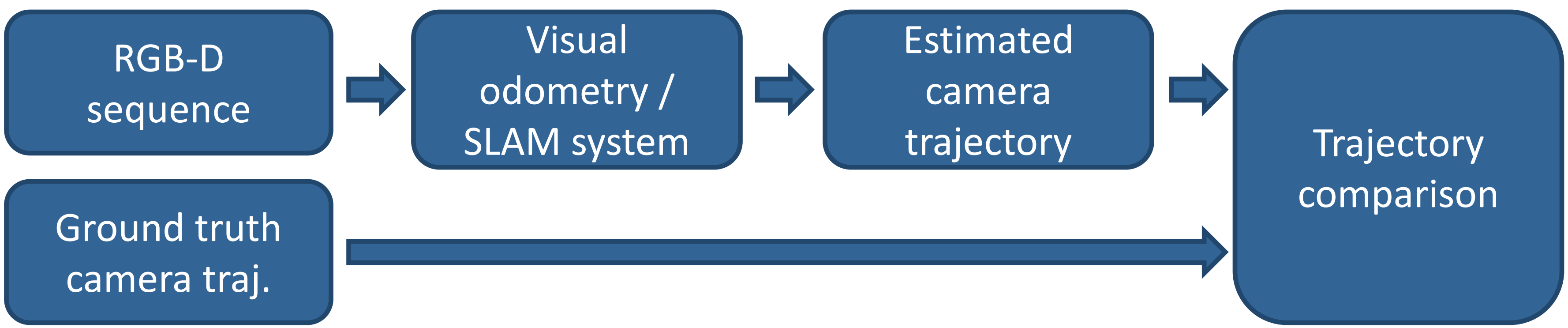
Dataset Website

- In total: 39 sequences (19 with ground truth)
- One ZIP archive per sequence, containing
  - Color and depth images (PNG)
  - Accelerometer data (timestamp ax ay az)
  - Trajectory file (timestamp tx ty ty qx qy qz qw)
- Sequences also available as ROS bag and MRPT rawlog

http://vision.in.tum.de/data/datasets/rgbd-dataset

What Is a Good Evaluation Metric?

- Compare camera trajectories
  - Ground truth trajectory $Q_1, \ldots, Q_n \in \mathrm{SE}(3)$
  - Estimate camera trajectory $P_1, \ldots, P_n \in \mathrm{SE}(3)$
- Two common evaluation metrics
  - Relative pose error (drift per second)
  - Absolute trajectory error (global consistency)

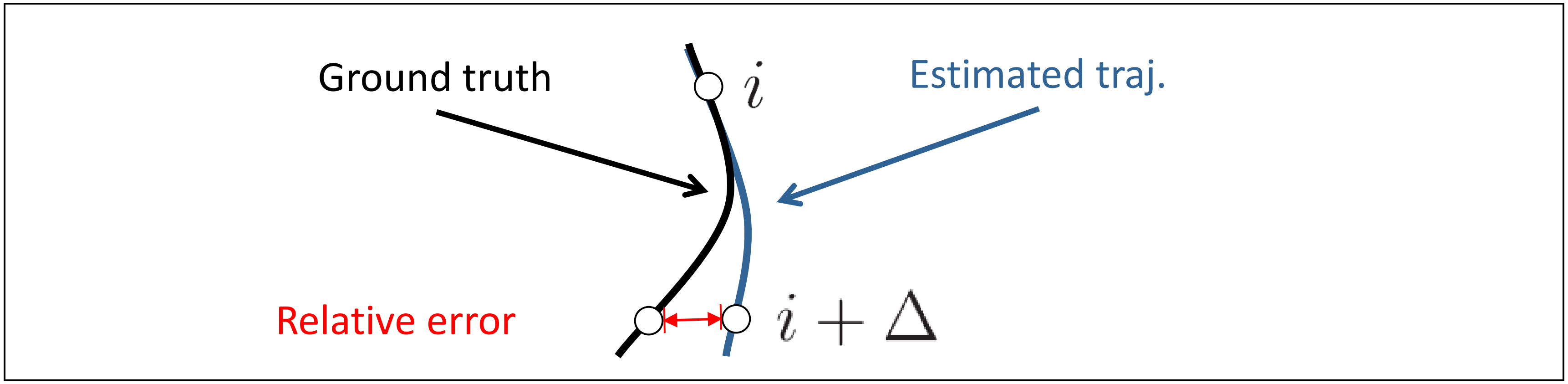


Relative Pose Error (RPE)

- Measures the (relative) **drift**
- Recommended for the evaluation of visual odometry approaches

$$E_i := \left(Q_i^{-1} Q_{i+\Delta}\right)^{-1} \left(P_i^{-1} P_{i+\Delta}\right)$$

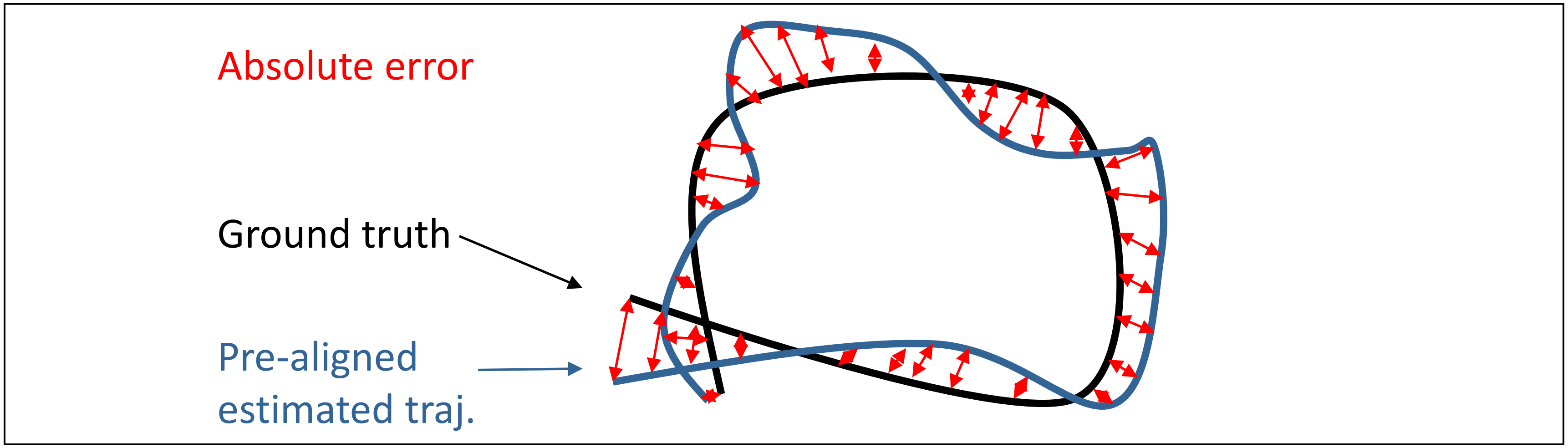
Relative error   True motion   Estimated motion



Absolute Trajectory Error (ATE)

- Measures the **global error**
- Requires pre-aligned trajectories
- Recommended for SLAM evaluation

$$E_i := Q_i^{-1} S P_i$$

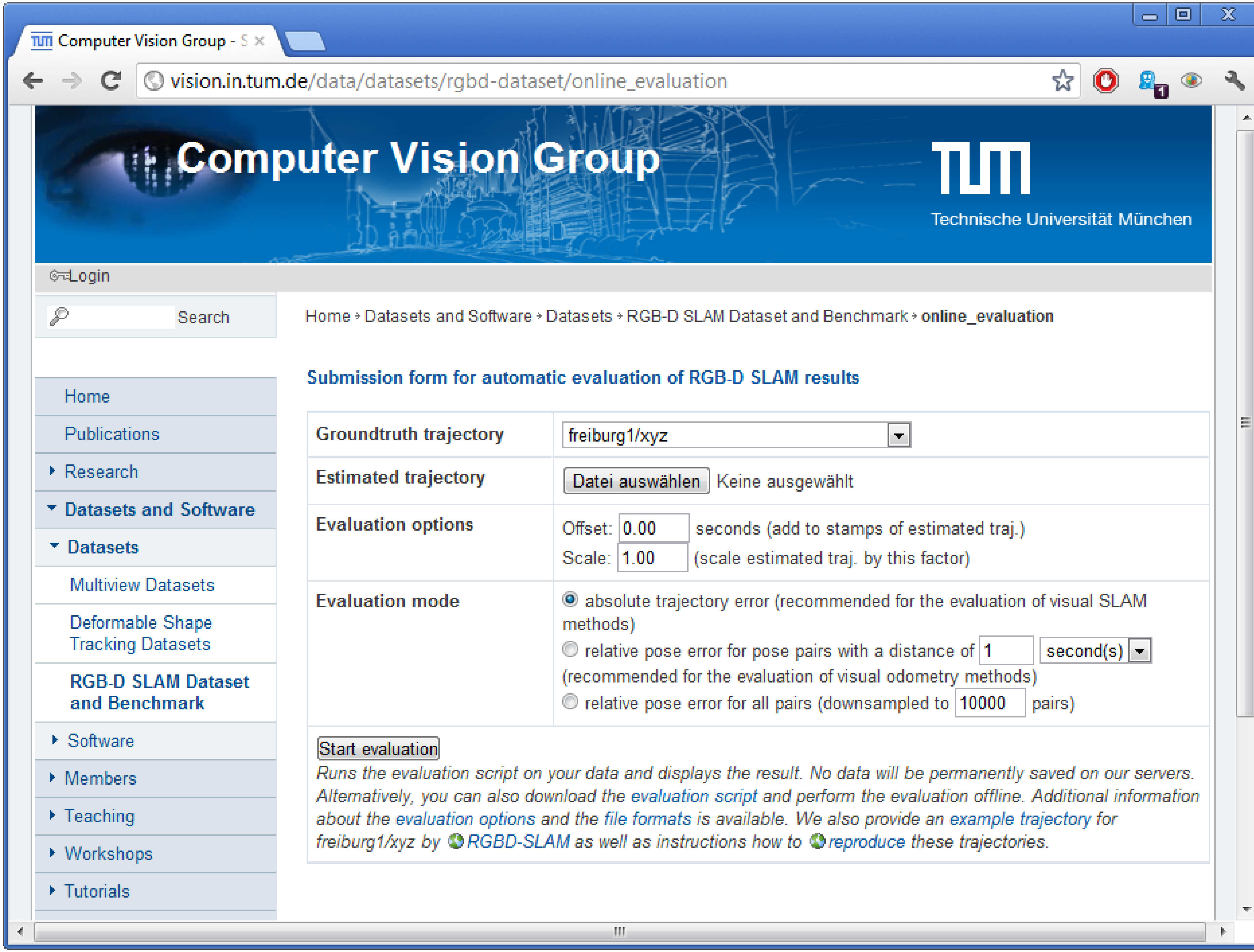


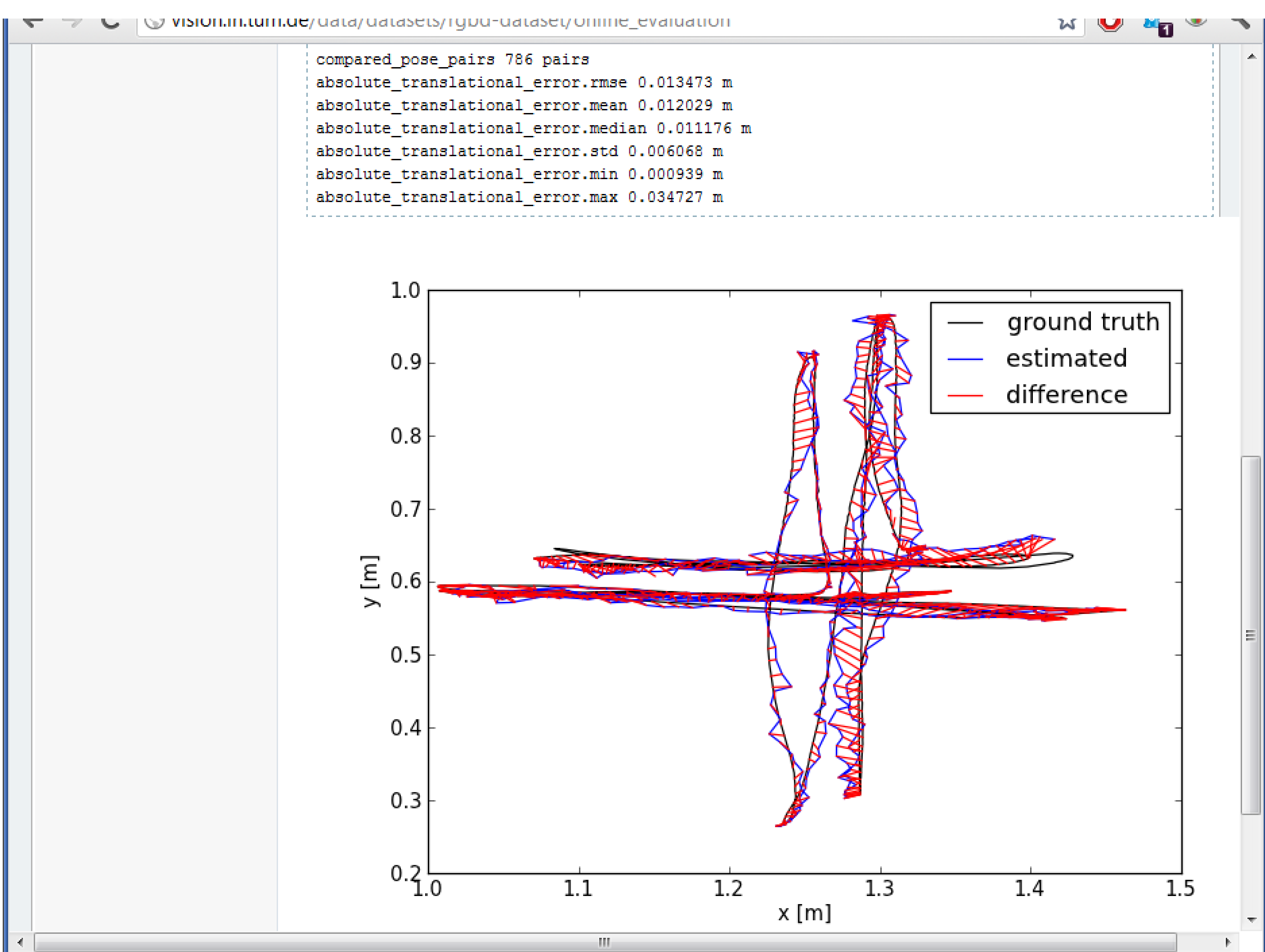
Evaluation metrics

- Average over all time steps

$$\mathrm{RMSE}(E_{1:n}) := \left(\frac{1}{m}\sum_{i=1}^{m} \|trans(E_i)\|^2\right)^{1/2}$$

- Reference implementations for both evaluation metrics available
- Output: RMSE, Mean, Median (as text)
- Plot (png/pdf, optional)


Computer Vision Group
vision.in.tum.de/data/datasets/rgbd-dataset/online_evaluation
Computer Vision Group
Technische Universität München
Login
Search
Home › Datasets and Software › Datasets › RGB-D SLAM Dataset and Benchmark › online_evaluation
Home
Publications
Research
Datasets and Software
Datasets
Multiview Datasets
Deformable Shape Tracking Datasets
RGB-D SLAM Dataset and Benchmark
Software
Members
Teaching
Workshops
Tutorials
Submission form for automatic evaluation of RGB-D SLAM results
Groundtruth trajectory
freiburg1/xyz
Estimated trajectory
Datei auswählen
Keine ausgewählt
Evaluation options
Offset: 0.00 seconds (add to stamps of estimated traj.)
Scale: 1.00 (scale estimated traj. by this factor)
Evaluation mode
absolute trajectory error (recommended for the evaluation of visual SLAM methods)
relative pose error for pose pairs with a distance of 1 second(s) (recommended for the evaluation of visual odometry methods)
relative pose error for all pairs (downsampled to 10000 pairs)
Start evaluation
Runs the evaluation script on your data and displays the result. No data will be permanently saved on our servers. Alternatively, you can also download the evaluation script and perform the evaluation offline. Additional information about the evaluation options and the file formats is available. We also provide an example trajectory for freiburg1/xyz by RGBD-SLAM as well as instructions how to reproduce these trajectories.


vision.in.tum.de/data/datasets/rgbd-dataset/online_evaluation
compared_pose_pairs 786 pairs
absolute_translational_error.rmse 0.013473 m
absolute_translational_error.mean 0.012029 m
absolute_translational_error.median 0.011176 m
absolute_translational_error.std 0.006068 m
absolute_translational_error.min 0.000939 m
absolute_translational_error.max 0.034727 m
ground truth
estimated
difference
y [m]
x [m]
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
1.0
1.1
1.2
1.3
1.4
1.5

Summary – TUM RGB-D Benchmark

- Dataset for the evaluation of RGB-D SLAM systems
- Ground-truth camera poses
- Evaluation metrics + tools available

Discussion on Benchmarks

Pro:

- Provide objective measure
- Simplify empirical evaluation
- Stimulate comparison

Con:

- Introduce bias towards approaches that perform well on the benchmark (overfitting)
- Evaluation metrics are not unique (many alternative metrics exist, choice is subjective)

Three Phases of Evolution in Research

- Novel research problem appears
(e.g., market launch of Kinect, quadrocopters, …)
  - Is it possible to do something at all?
  - Proof-of-concept, qualitative evaluation
- Consolidation
  - Problem is formalized
  - Alternative approaches appear
  - Need for quantitative evaluation and comparison
- Settled
  - Benchmarks appear
  - Solid scientific analysis, text books, …

Final Exam

- Oral exam **in teams** (2-3 students)
- At least 15 minutes per student
  → individual grades
- Questions will address
  - Your project
  - Material from the exercise sheets
  - Material from the lecture

Exercise Sheet 6

- Prepare final presentation
- Proposed structure: 4-5 slides
  1. Title slide with names + motivating picture
  2. Approach
  3. Results (video is a plus)
  4. Conclusions (what did you learn in the project?)
  5. Optional: Future work, possible extensions
- Hand in slides before Tue, July 17, 10am (!)

Time for Questions