RAWSEEDS: Datasets and Problems for SLAM benchmarking

G. Fontana, D. Marzorati, M. Matteucci, ***D. G. Sorrenti***

Workshop on Good Experimental Methodologies and Benchmarking in Robotics Research and Applications, Leuven, 6 - 7 April 2009

Topics

- Benchmarking of SLAM
- SLAM Evaluation and the need for GT
- RAWSEEDS … what's that?
- Vision & Laser GT Systems
- The (validated!) datasets are now ready
- Definition of Benchmark Problems
  - Proposed rating methodologies
- Discussion on ... giving marks to solutions!

Why SLAM Benchmarking

- Benchmarking of a fully fledged robotic application might be complex and hard to tackle as a whole …
- Simultaneous Localization And Mapping is one of the easiest activity to benchmark in robotics … provided:
  - We can establish proper metrics for SLAM
  - The community agrees on the use of such metrics
  - The community appreciate the effort for using it
- SLAM can be considered an enabling capabilities for many complex tasks in autonomous robots

How do we evaluate SLAM?

- To set up a benchmark for SLAM we need to define a way to asses the performance of a SLAM algorithm
  - Quantitative measures of map/path quality, w.r.t. ground truth
  - Performance variation as map size grows
  - How realistic/pessimistic/optimistic is the estimation error
  - …
- Most measures are referred to ***ground truth!***
  - GT for the robot pose
  - GT for the map

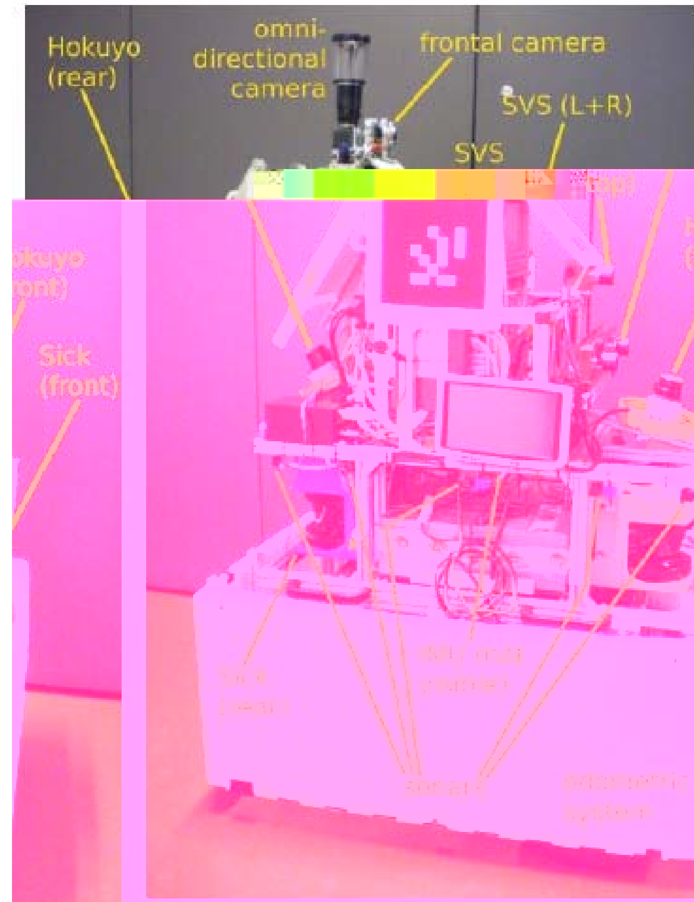
Benchmarking Beyond Radish

- RAWSEEDS goal is to publish:
  - Extended multi-sensor data sets for the testing of systems on real-world scenarios
  - Benchmarks and methodologies for quantitative evaluation and comparison of algorithms/sensors
  - Off-the-shelf algorithms, with demonstrated performances, to be used for research bootstrap and comparison.
- RAWSEEDS created a website from which researchers and companies will be able to download these benchmarks, contribute new material, and communicate with each other.

www.rawseeds.org

RAWSEEDS Sensor Suite

- Onboard extensive sensing suite
  - B/W + Color cameras (mono/stereo)
  - 3D cameras (SVS by Videre)
  - LRFs (SICK 2D)
  - Omnidirectional camera (V-Stone)
  - Sonar belt
  - Other proprioceptives (e.g., odometry, Inertial Measurement Unit)

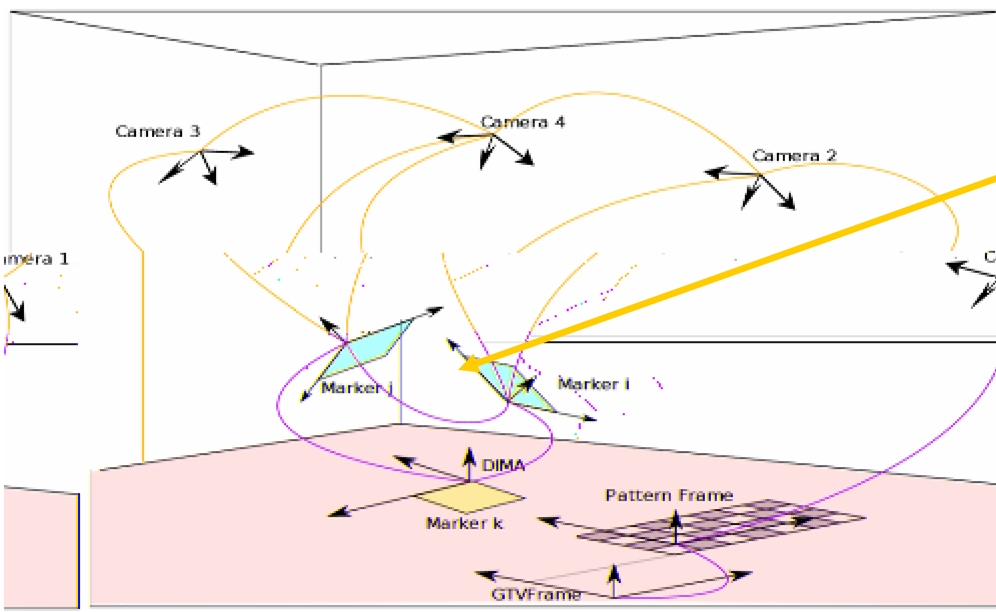


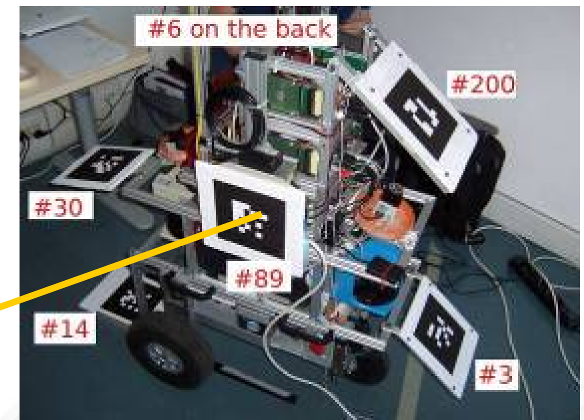
Ground truth systems

- Executive drawings for mapping;
- Vision-based GT System for robot pose;
- Laser-based GT System for robot pose.

Vision-based GT System

- Use a camera network to localize the robot
  - Good: Independent sensor (from the robot ones)
  - Bad: Requires long setup/calibration

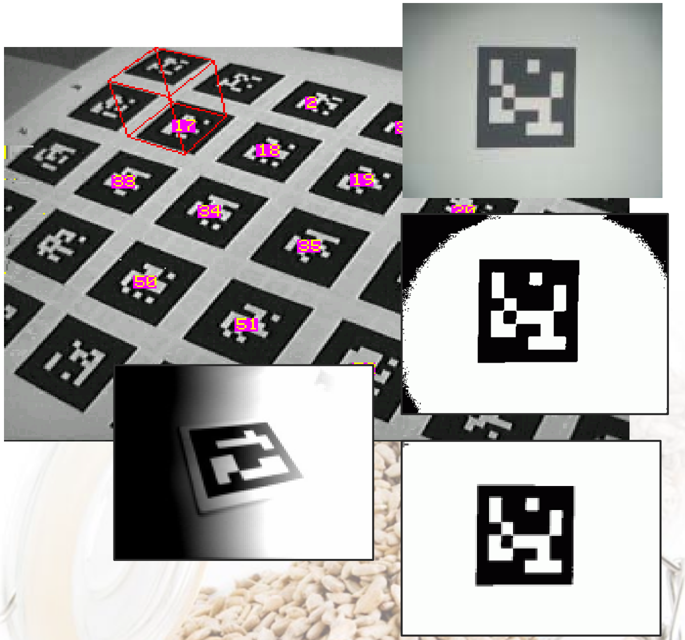


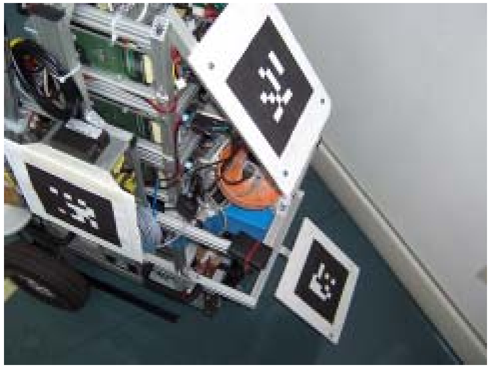


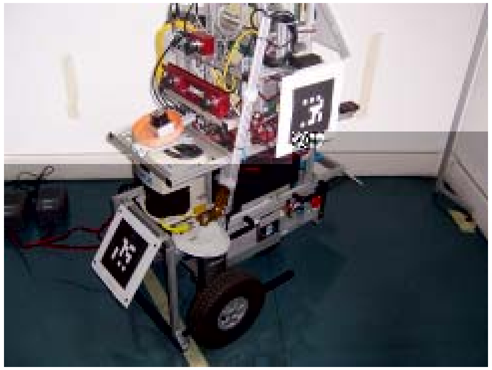
~0.1 m accuracy

Marker Detection/Localization

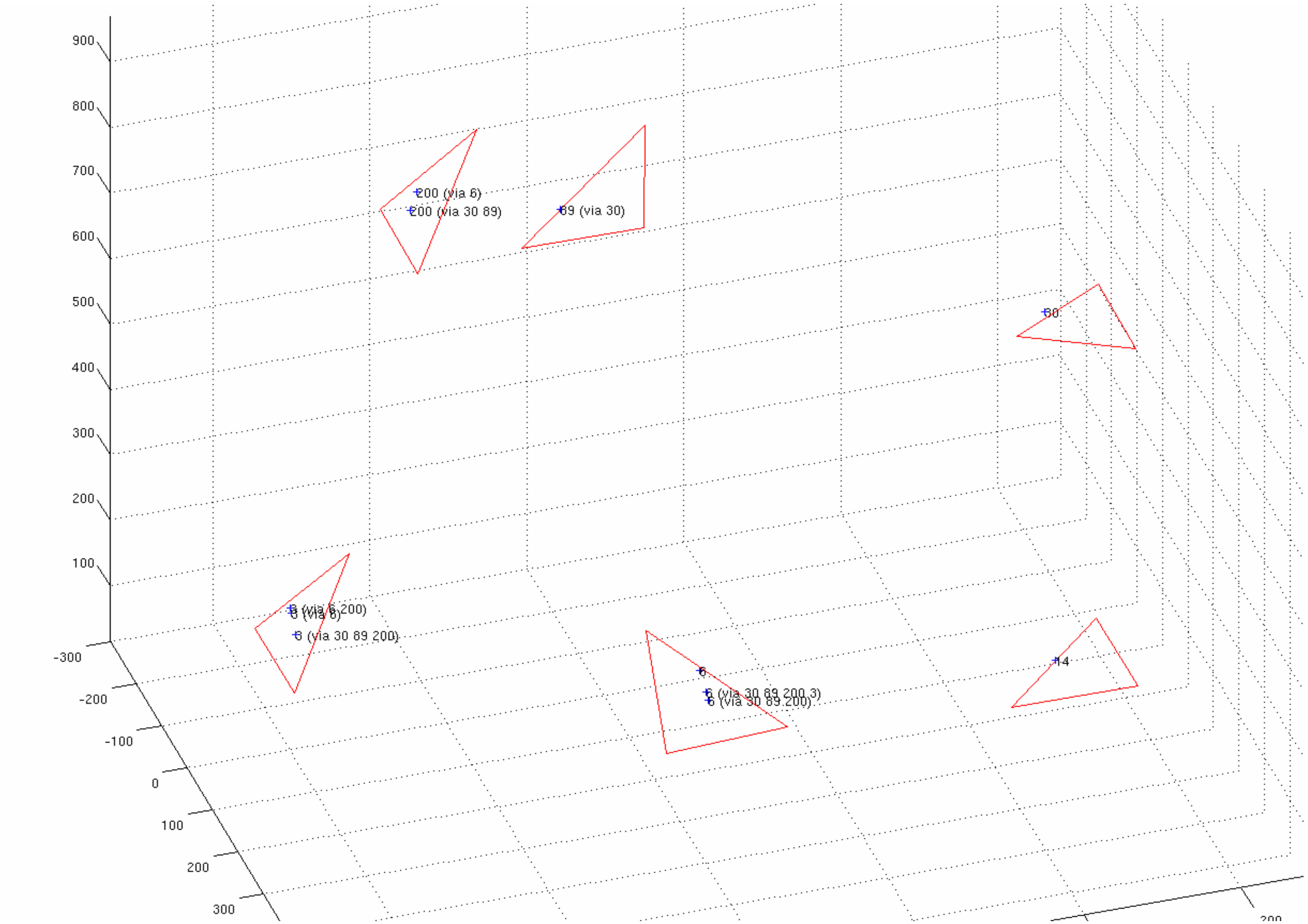
- Artoolkit Plus: publicly available software, capable to recognize and localize one out of a large set of markers:
  - Simple Id-encoded markers
  - Automatic thresholding
  - Vignetting compensation
  - MATLAB camera calibration toolbox
  - "Robust Planar Pose" algorithm

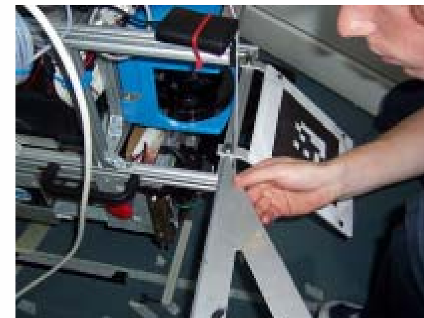


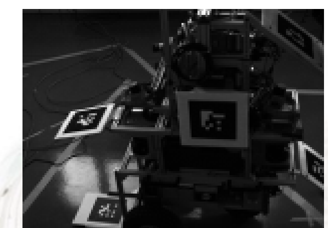




Marker Positioning





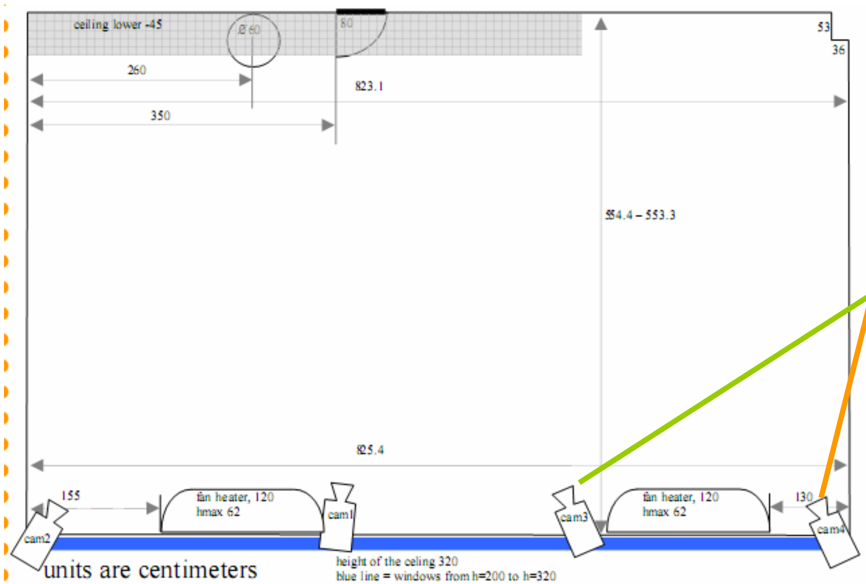


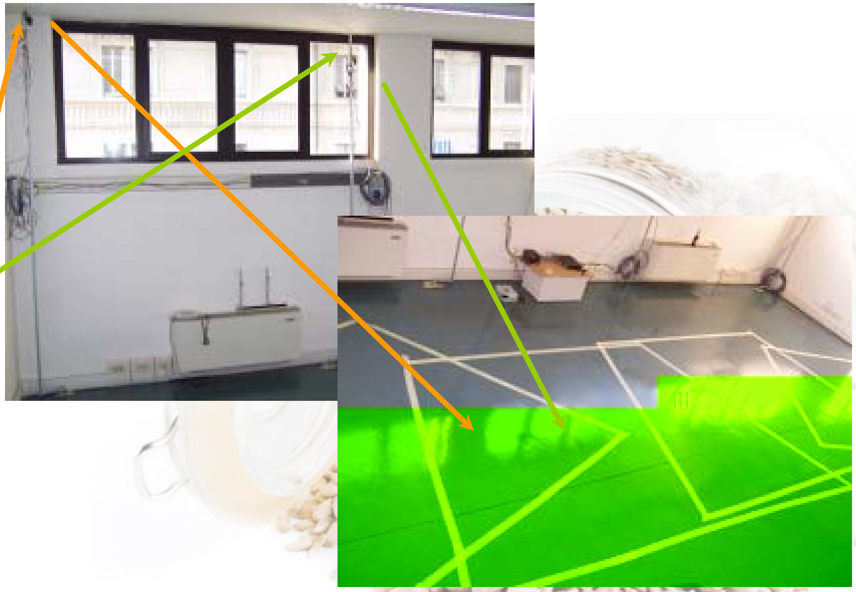
Marker localization accuracy

- The Artoolkit Plus turned out to be more oriented for speed than precision;
- given the GT does not require an online computation, we devised a more accurate version, both in term of detection rate and accuracy, trading off with running time

Camera Network Calibration (I)

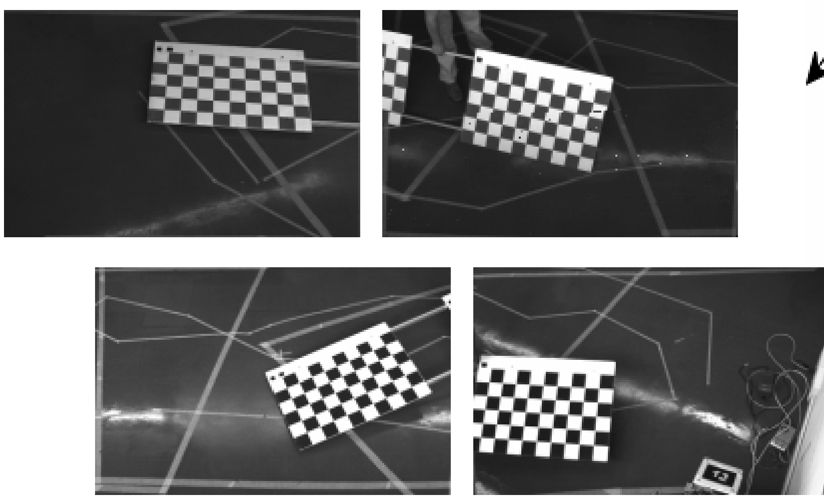
- Each camera is calibrated with the standard Jean-Yves Bouguet's "Camera Calibration Toolbox for MATLAB"
- Only partial "field of view" overlapping, not always possible to lay down a set of Checkboards …

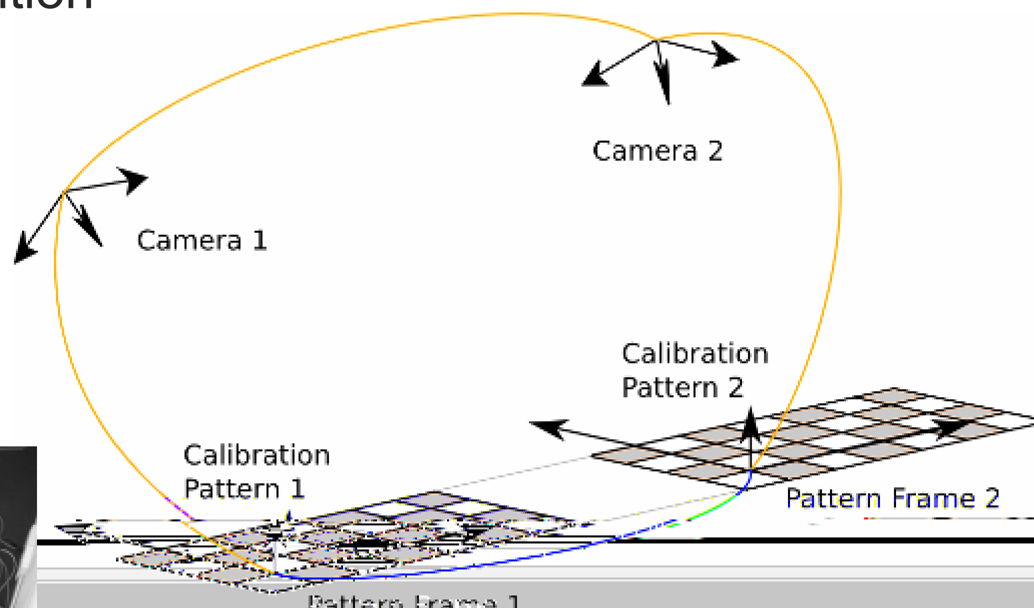




Camera Network Calibration (II)

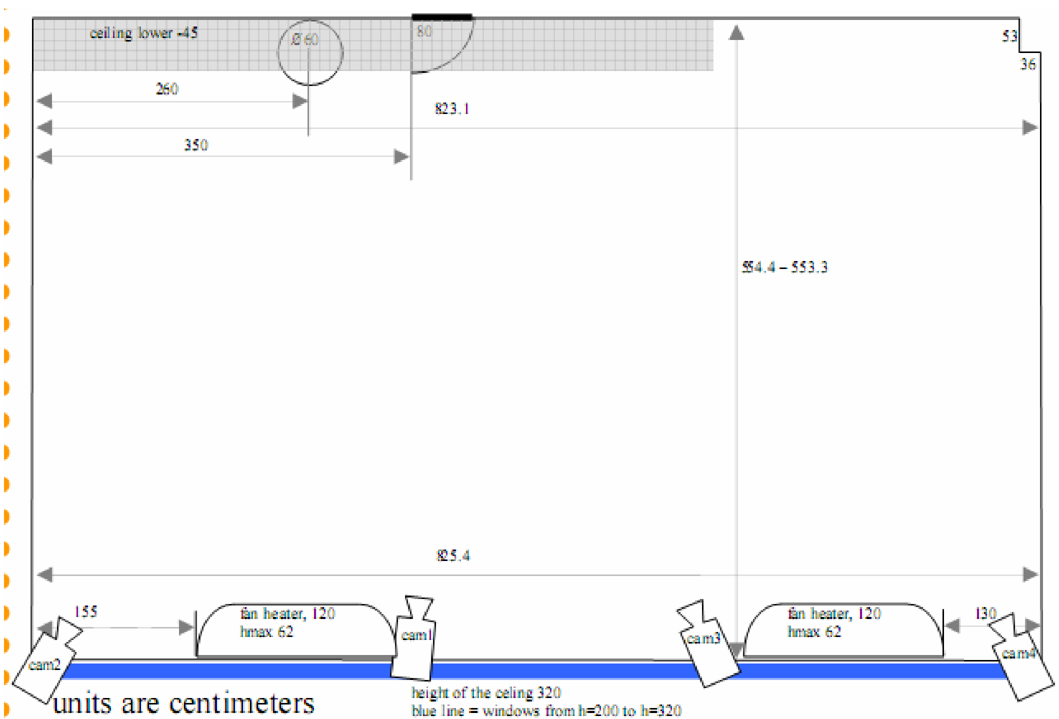
- We use a "double pattern" approach … and averaging
  - Checkerboard pairing
  - Roto-translation composition





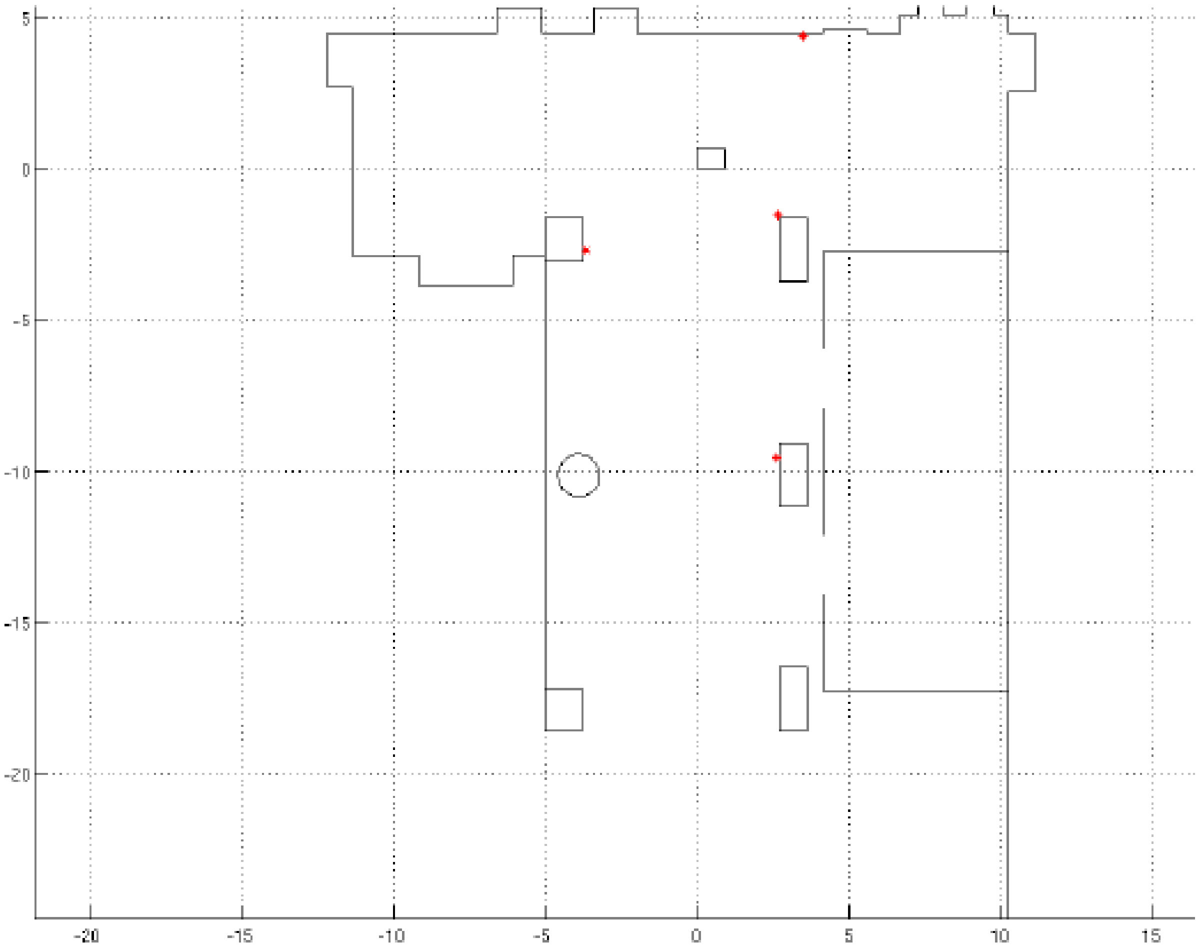
What about precision?

- With an 8 meters chain obtained chaining 4 cameras (Prosilica GC-750, 640x480)

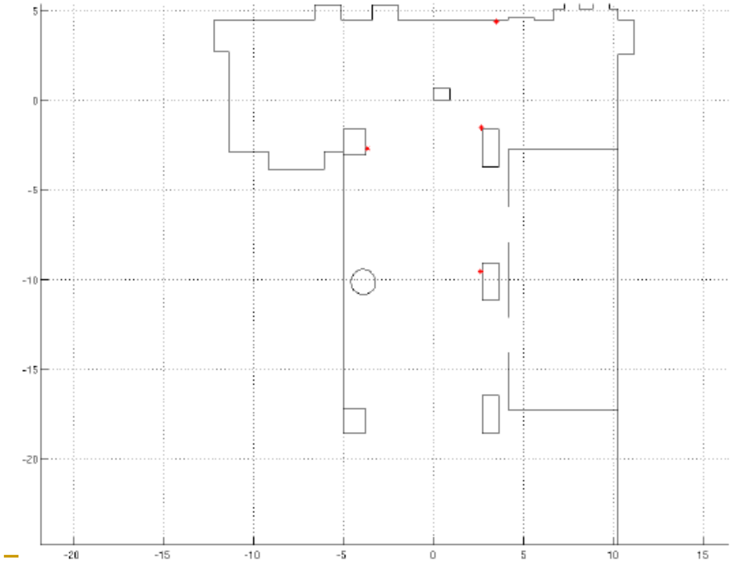


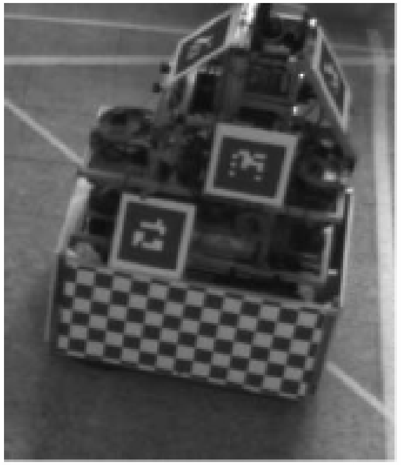
	chessboard estimation stats		
	average Err	standard deviation Err	max of abs values Err
x	-0.0049	0.0095	0.0116
y	-0.0006	0.0042	0.0036
z	0.0083	0.0179	0.0280

Laser-based GT System

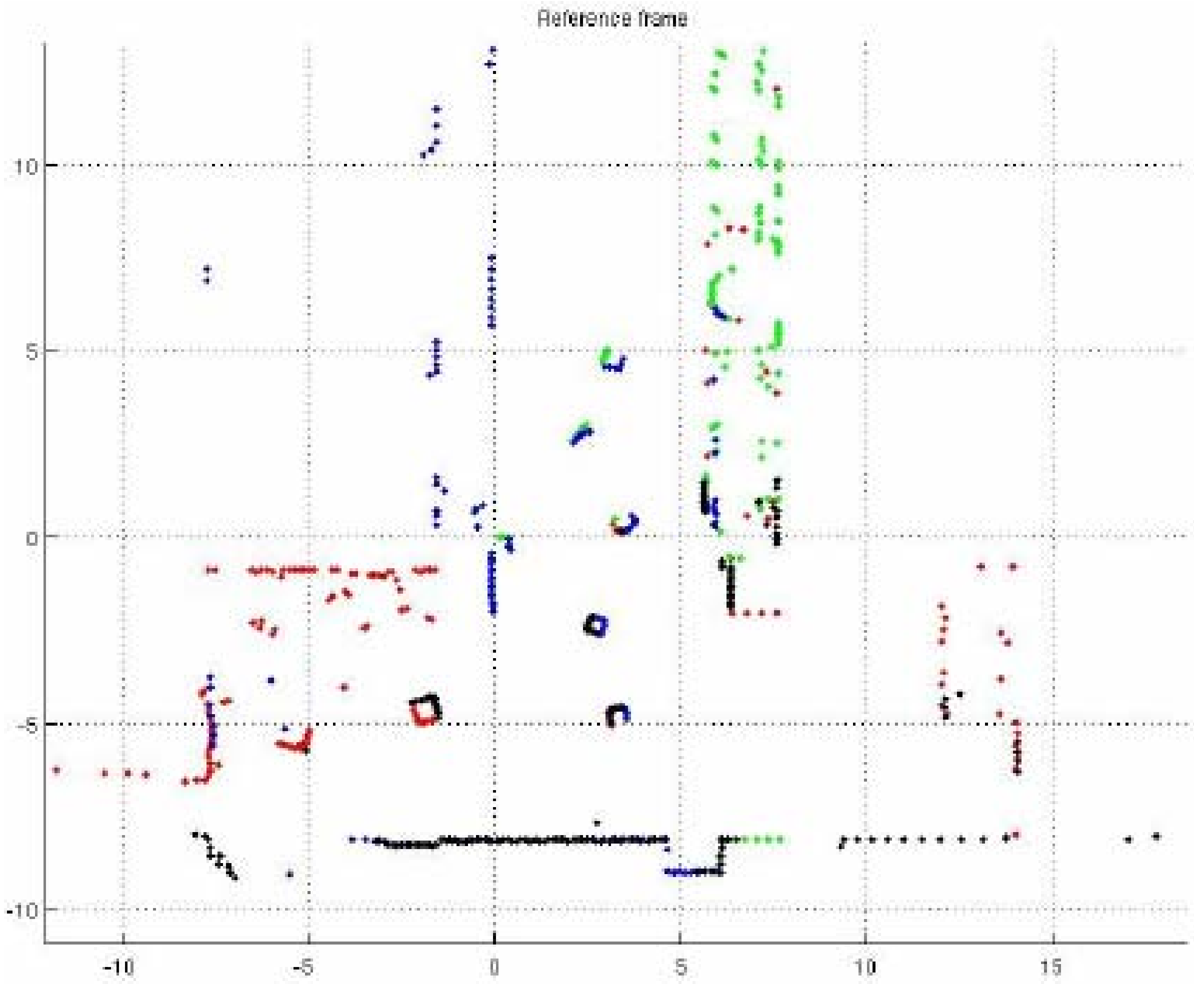


Laser-based GT System

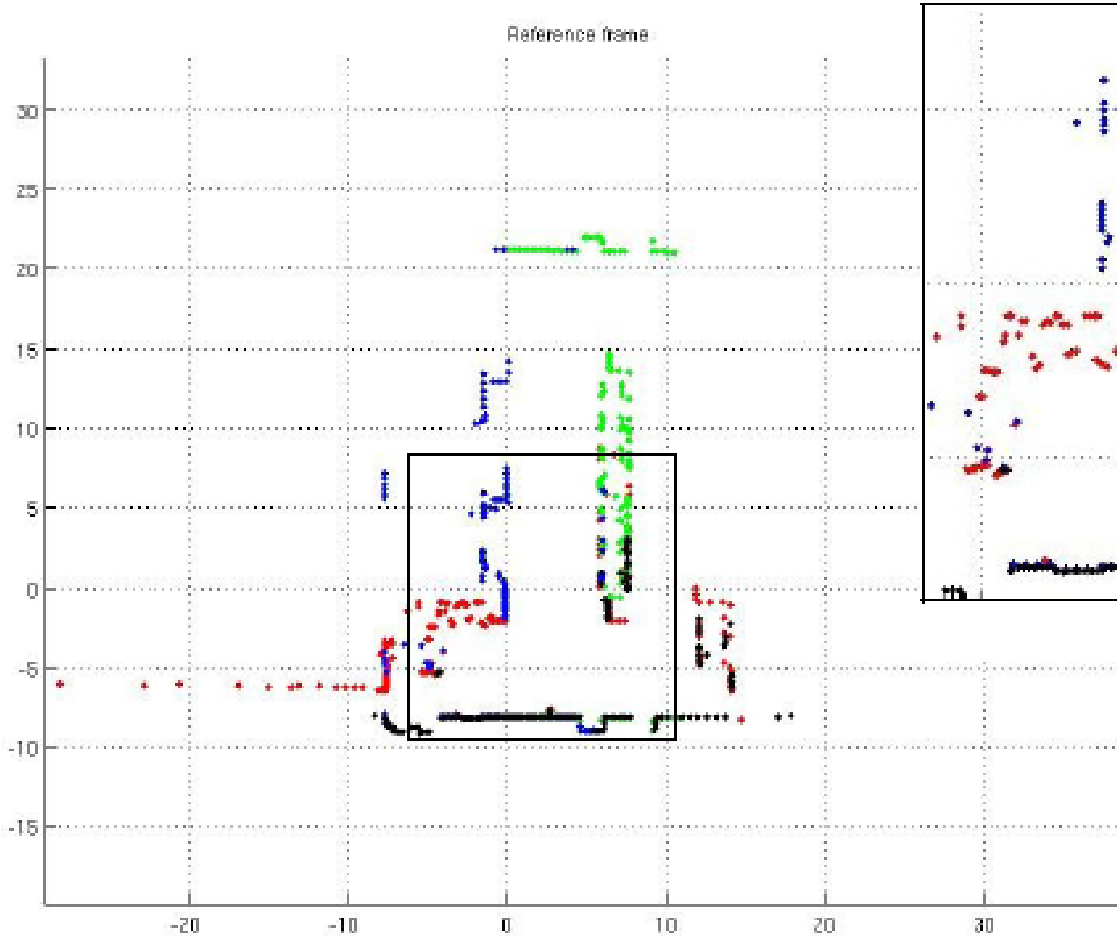


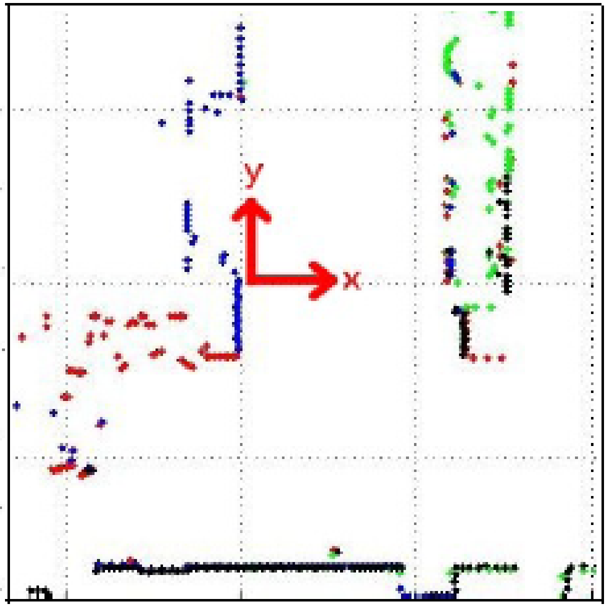


Laser-based GT System



Laser-based GT System

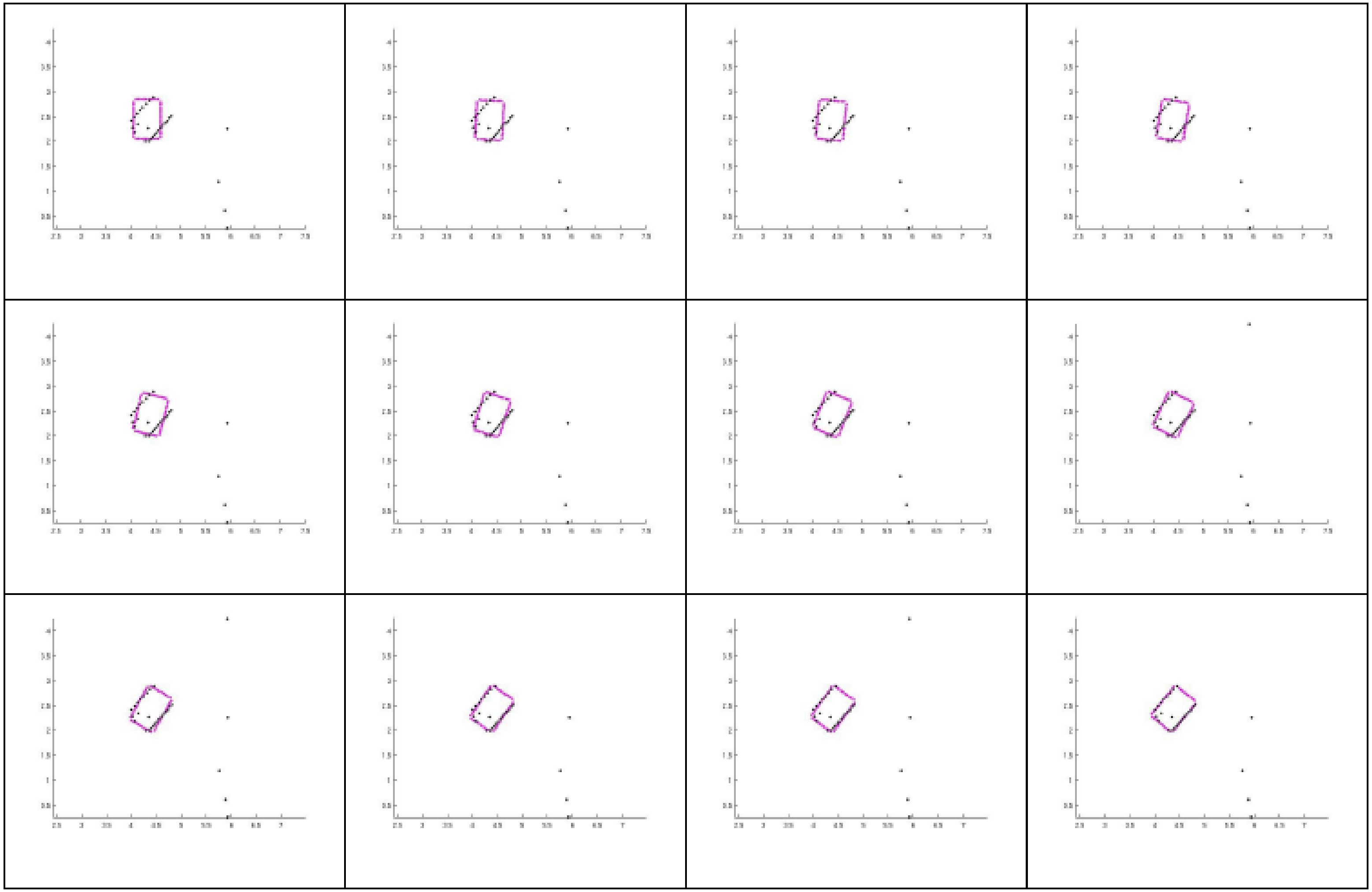




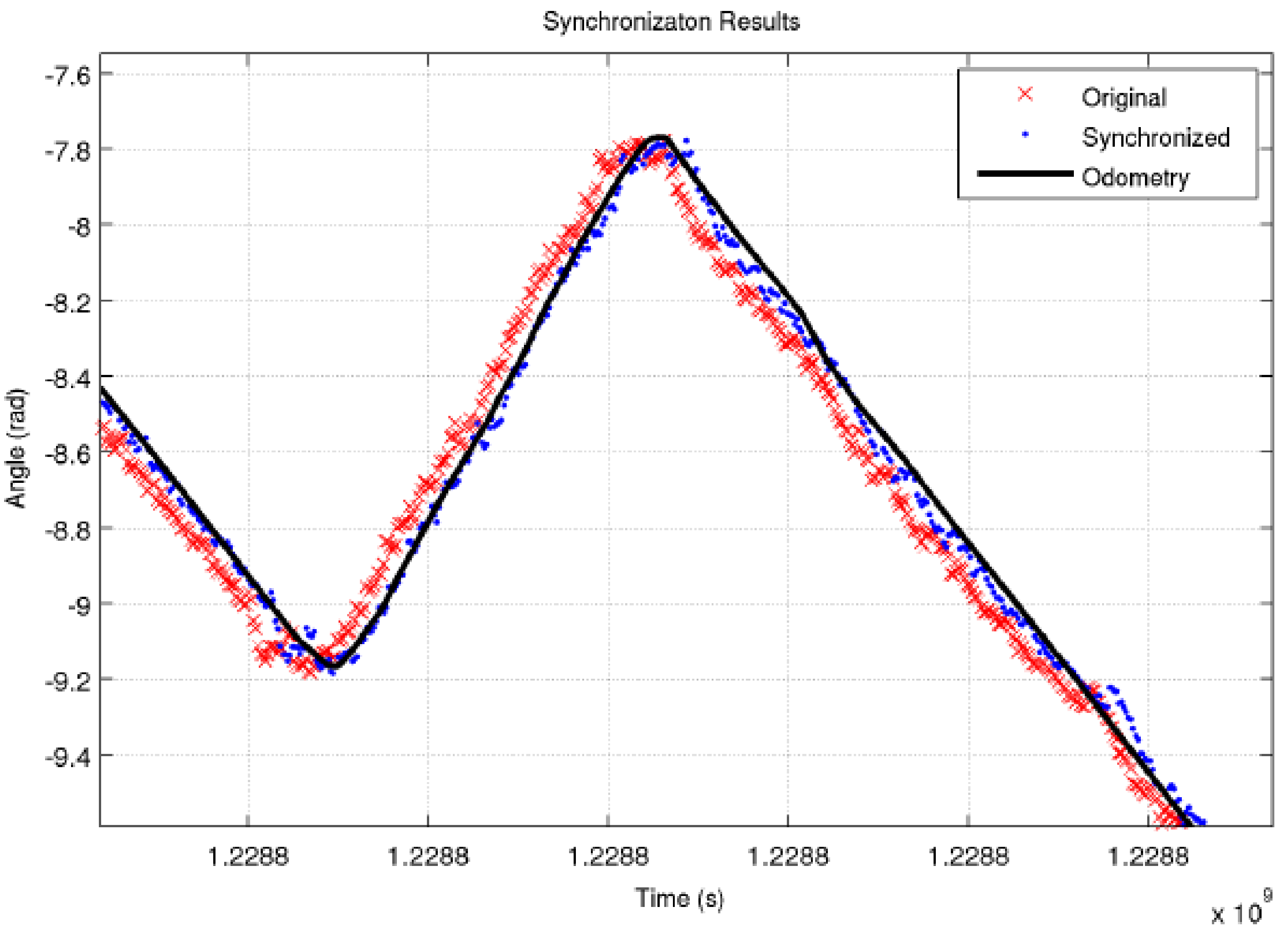
Laser-based GT System

- composition of the scans;
- filtering;
  - similar to background subtraction
- application of the ICP algorithm;
  - points from the robot shape to points from the scans;
  - starting from the previous pose;

Laser-based GT System - ICP



Synchronization between GT and onboard sensors

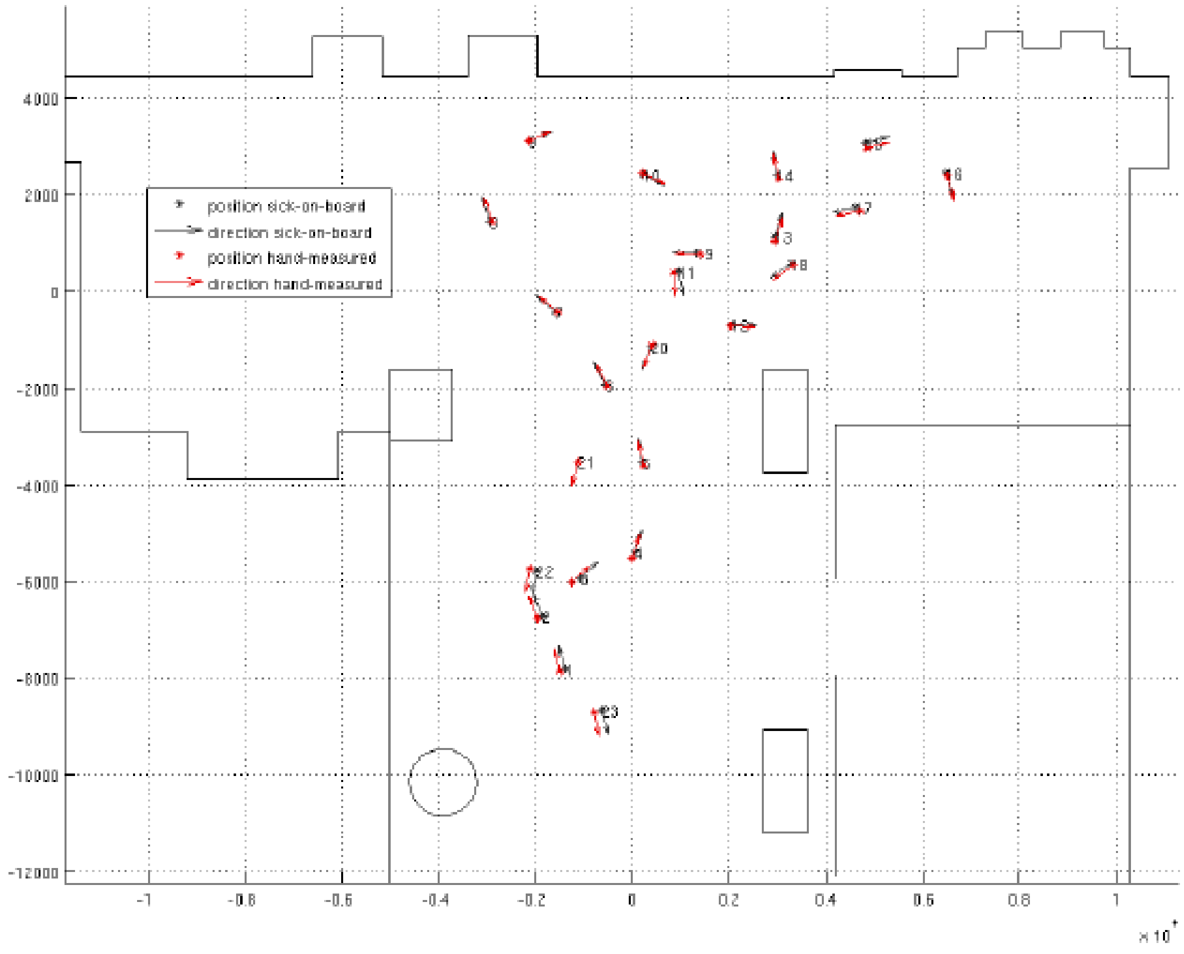


GT Validation Procedure

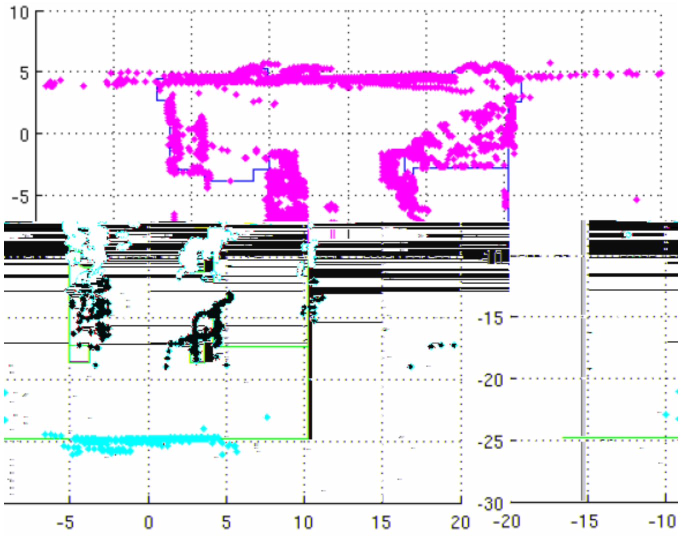
- Validation should allow the evaluation of the GT systems;
  - Homogeneous in nature to the ones provided by the GT systems
  - Obtained with different approaches
  - Trustable … and we only trust ourselves

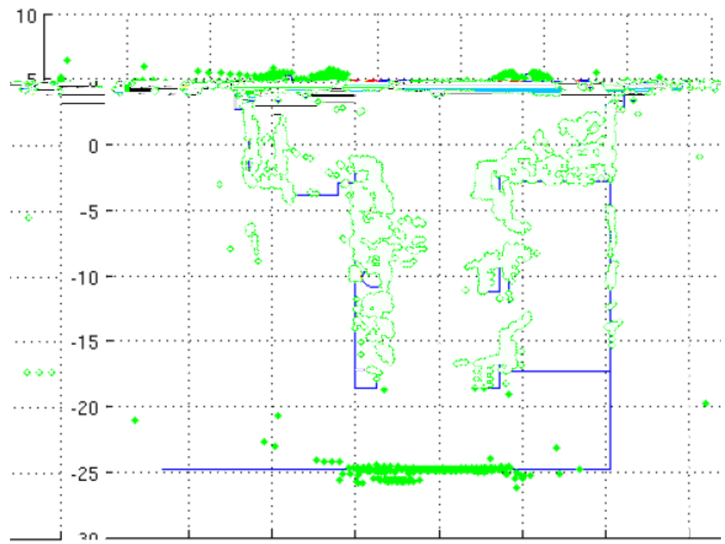
- Use quantitative (laser quality) hand-measuring
  - Find the position of some world points w.r.t. reference frame
  - Find the position of a few points on the robot w.r.t. to these points
  - Combine these measurements into a robot pose (Kalman filter)
  - Compare this measurements to the GT systems output.

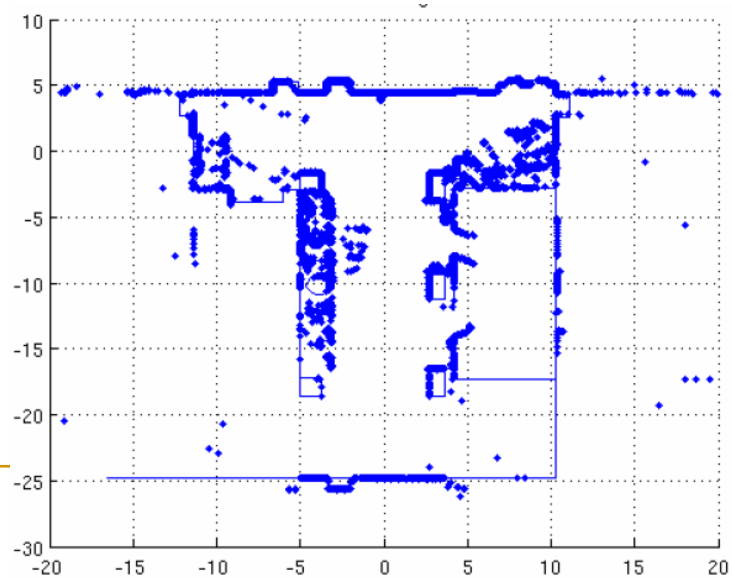
Hand validation not accurate enough



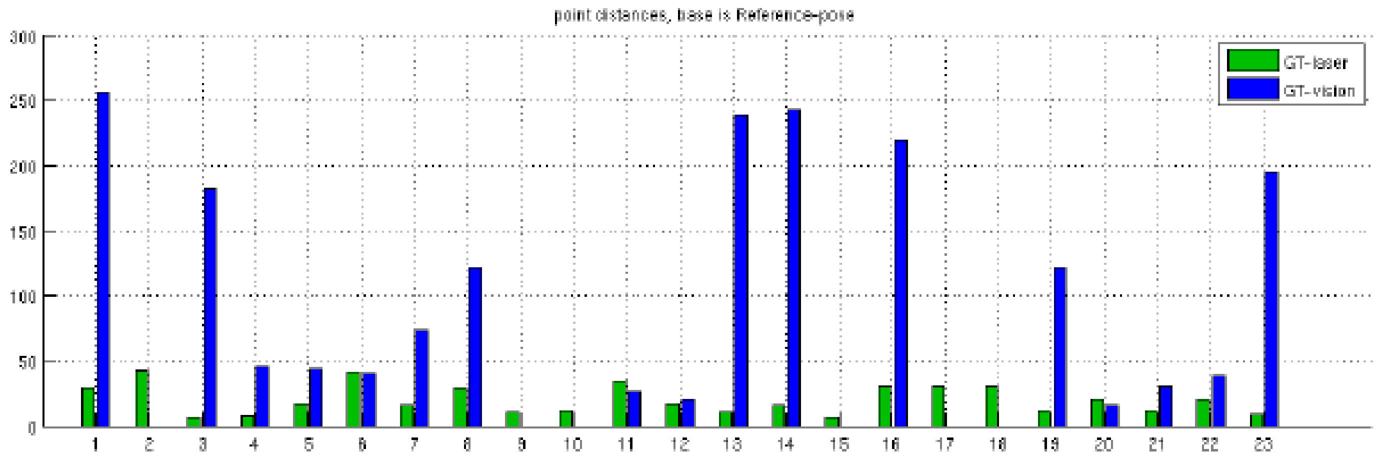
Comparison of GT systems

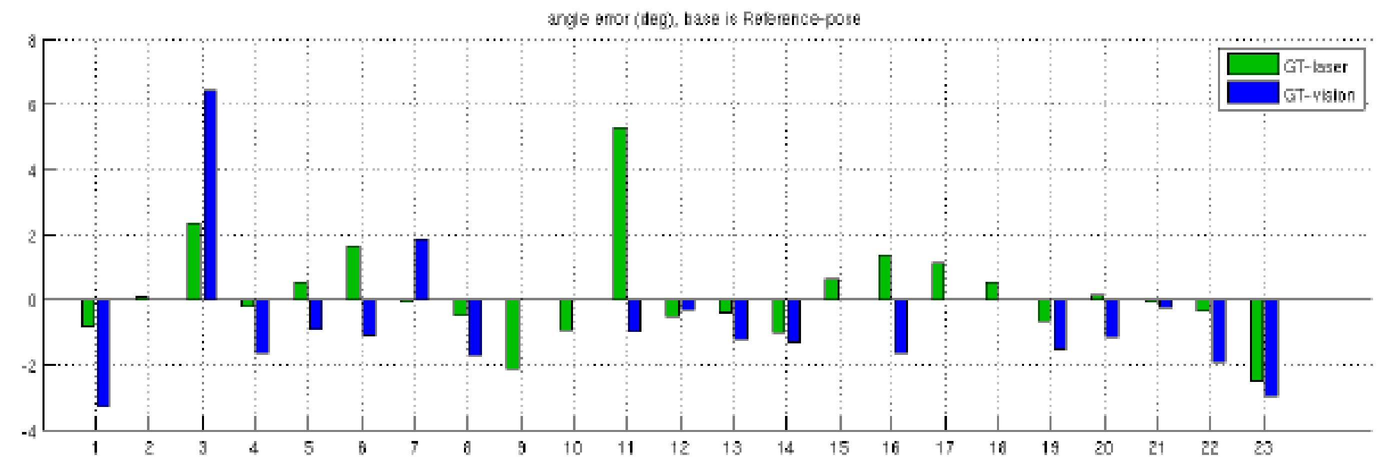


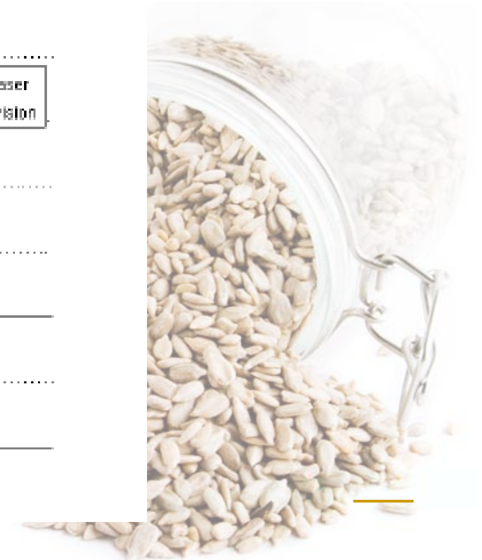




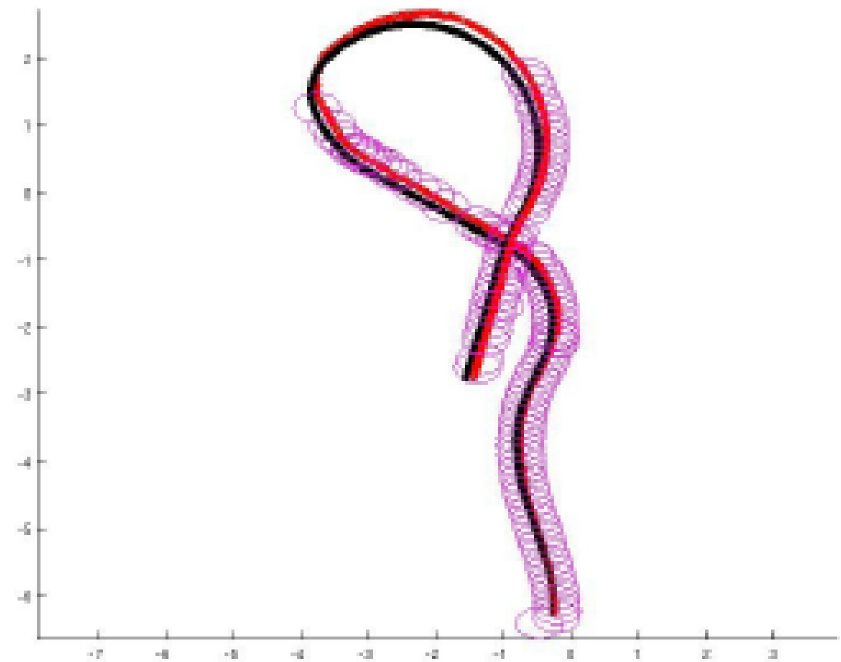
Comparison of GT systems





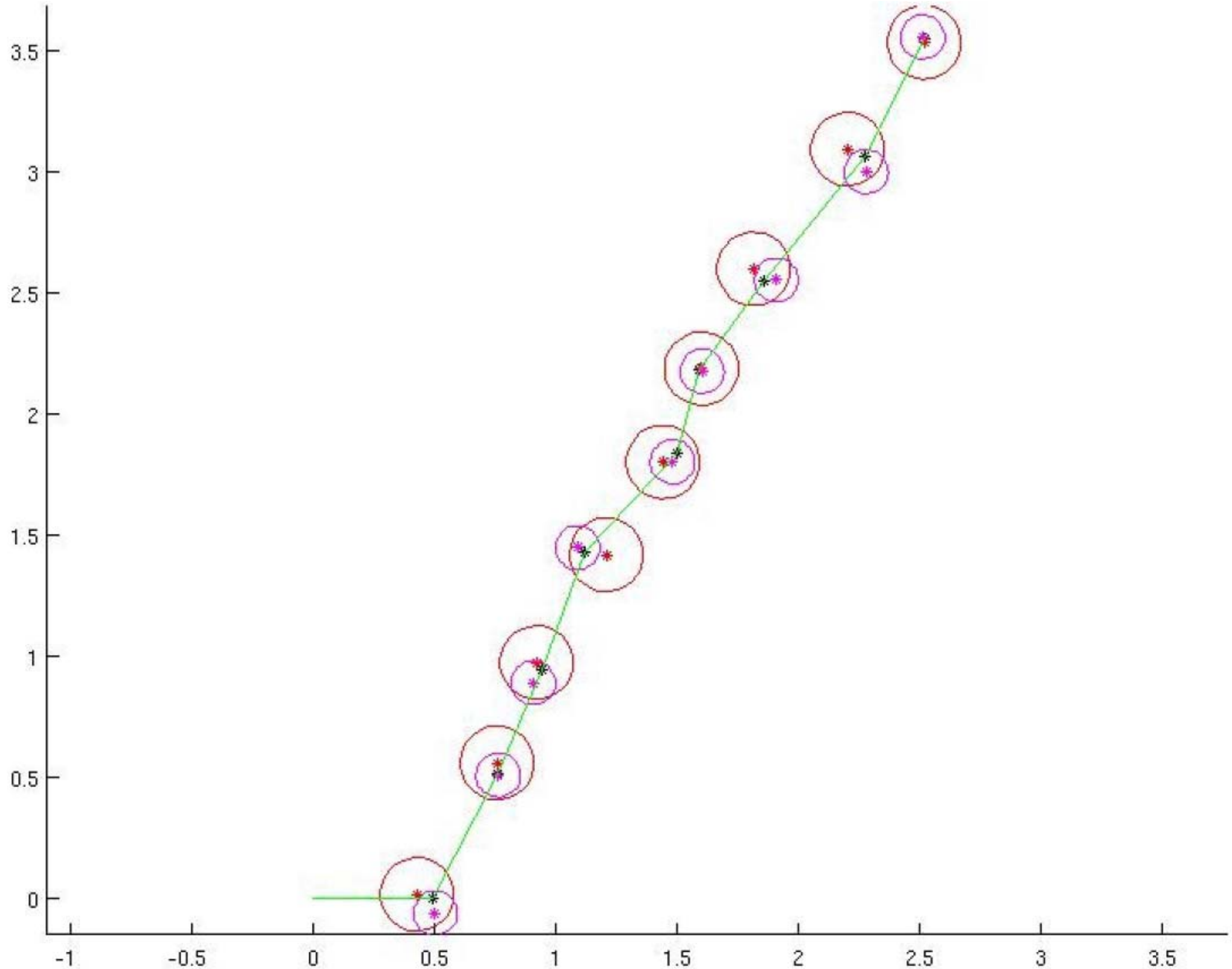


Comparison of GT systems

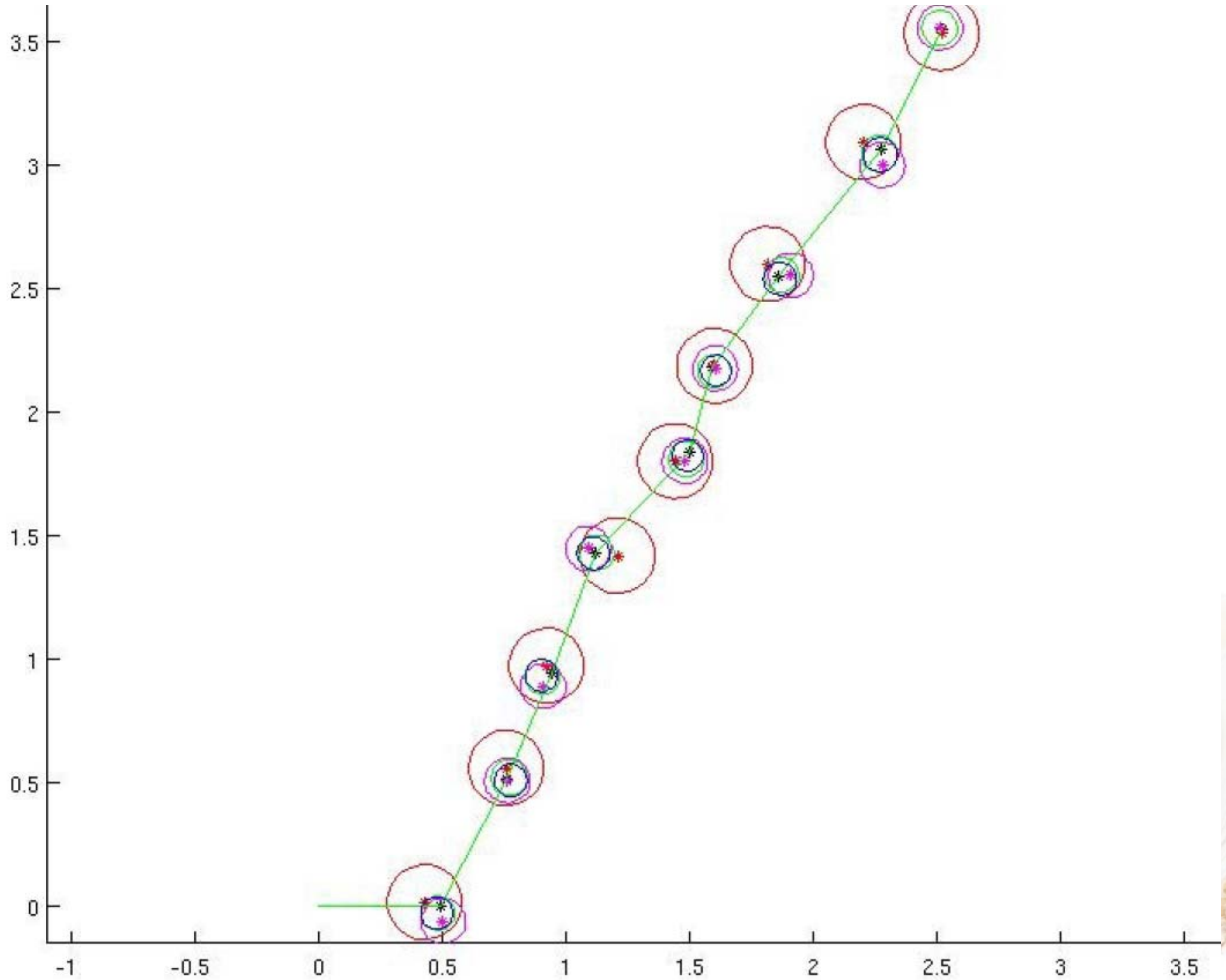


GT laser superimposed with GT vision. In black the odometry data stream.

Comparison of GT systems



Kalman smoother



Datasets will be online soon

Datasets will be online on www.rawseeds.org from about the beginning of May

Validation of datasets

- all datasets have been carefully validated;
  - format of file
  - mean and maximum times between samples
  - synchronization is verified in the portions of the trajectory found to have high angular velocity (cross-correlation with the one from the IMU);
  - Data overlap;
  - Data density and quality;
    - for video streams
      - absence of dropped frames;
      - absence of dark frames;
      - accurate calibration;

Validation of datasets

Sesion	Mixed		
Dataset	Bovisa [illegible] Static	Bovisa [illegible] Dynamic	Bovisa [illegible] Static
Odometry	Valid[1]	Valid[1]	Valid[1]
IMU	Valid	Valid	Valid
SICK Laser	Valid	Valid	Valid
Hokuyo Laser	Not usable outdoors	Not usable outdoors	Not usable outdoors
Sonar Belt	Not available	Not available	Not available
Monocular Vision	Valid	Valid	Valid
Trinocular Vision	Valid[3]	Valid[3]	Valid[3]
Panoramic Vision	Valid	Valid	Valid[5]
GPS	Valid	Valid	Valid

Sesion	Outdoor		
Dataset	Bovisa [illegible] Static	Bovisa [illegible] Dynamic	Bovisa [illegible] Static
Odometry	Valid[1]	Valid[1][2]	Valid[1][2]
IMU	Valid	Valid	Valid
SICK Laser	Valid	Valid	Valid
Hokuyo Laser	Not usable outdoors	Not usable outdoors	Not usable outdoors
Sonar Belt	Not available	Not available	Not available
Monocular Vision	Valid	Valid	Valid
Trinocular Vision	Valid[3][4]	Valid[3]	Valid[3][4]
Panoramic Vision	Valid	Valid	Valid
GPS	Valid	Valid	Valid

Sesion	Indoor					
Conditions	Static Lamps		Static Daylight	Dynamic Lamps	Dynamic Daylight	
Dataset	Bicocca 2009-02-25b	Bicocca 2009-02-27b	Bicocca 2009-02-27a	Bicocca 2009-02-26b	Bicocca 2009-02-25a	Bicocca 2009-02-26a
Odometry	Valid	Valid	Valid	Valid	Valid	Valid
IMU	Valid	Valid	Valid	Valid	Valid	Valid
SICK Laser	Valid	Valid	Valid	Valid	Valid	Valid
Hokuyo Laser	Valid	Valid	Valid	Valid	Valid	Valid
Sonar Belt	Valid	Valid	Valid	Valid	Valid	Valid
Monocular Vision	Valid	Failed[6]	Valid	Valid	Valid[7]	Valid
Trinocular Vision	Valid	Failed[6]	Valid	Valid	Valid	Valid
Panoramic Vision	Valid	Failed[6]	Valid	Valid	Valid	Valid
GPS	Not available indoors	Not available indoors	Not available indoors	Not available indoors	Not available indoors	Not available indoors

Benchmark Problems (BP)

	PROBLEM	SENSOR DATA	GROUND TRUTH	EVALUATION MEASURES
Laser SLAM	perform a map building activity with SLAM (online)	laser, IMU and odometry from a dataset	mappingGT; poseGT	ME, ATE, RCE, SLE, RPE
Monocular SLAM	perform a map building activity with SLAM (online)	single camera, IMU and odometry from a dataset	mappingGT; poseGT	ME, ATE, RCE, SLE, RPE
Stereo SLAM	perform a map building activity with SLAM (online)	stereo camera, IMU and odometry from a dataset	mappingGT; poseGT	ME, ATE, RCE, SLE, RPE
Trinocular SLAM	perform a map building activity with SLAM (online)	trinocular IMU and odometry from a dataset	mappingGT; poseGT	ME, ATE, RCE, SLE, RPE
Omnidirectional SLAM	perform a map building activity with SLAM (online)	omnidirectional vision, IMU and odometry from a dataset	mappingGT; poseGT	ME, ATE, RCE, SLE, RPE
Sonar SLAM	perform a map building activity with SLAM (online)	sonar sensors, IMU and odometry from a dataset	mappingGT; poseGT	ME, ATE, RCE, SLE, RPE
Multisensor SLAM	perform a map building activity with SLAM (online)	streams from more than one sensor for a dataset	mappingGT; poseGT	ME, ATE, RCE, SLE, RPE

BP – rating methodologies

- mandatory or recommended;
- mapping performance
  - GT features = 2D corners
  - mapping reconstructed features onto 2D corners
  - mapping 3D features onto 2D ones
  - running time
- localization performance
  - absolute localization error
  - relative pose errore
- usage-based
  - self-localization on another dataset

Mapping performance

- ME (mapping error)
  - $(D_r - D^{GT}) / D^{GT}$
    - mean of the set of normalized differences $\{ N_r \}$;
    - standard deviation of the set of normalized differences $\{ N_r \}$;
    - confidence interval (3σ) of the set of normalized differences $\{N_r\}$;
  - recommended measure
- REC (Rough Estimate of Complexity)
  - mandatory measure
  - < timestamp, running time >

Trajectory performance

- ATE (Absolute Trajectory Error)
  - recommended measure
  - $d_j = \| trans(\boldsymbol{x}_j) - trans(\boldsymbol{x}_j^{\mathrm{GT}}) \|$
  - mean of the translation error $\{ d_j \}$;
  - standard deviation of the translation error $\{d_j \}$;
  - confidence interval of the translation error $\{d_j \}$;
- RPE (Relative Pose Error)
  - recommended
  - $\varepsilon(\delta) = \frac{1}{N} \sum_{i,j} trans(\boldsymbol{x}_{i,j} \ominus \boldsymbol{x}_{i,j}^{\mathrm{GT}})^2 + rot(\boldsymbol{x}_{i,j} \ominus \boldsymbol{x}_{i,j}^{\mathrm{GT}})^2$

Usage-based performance

- SLE (Self-Localization Error)
  - recommended
  - $d_j = \| trans(\boldsymbol{x}_j) - trans(\boldsymbol{x}_j^{\mathrm{GT}}) \|$
  - mean, standard deviation and confidence interval

Benchmark Solutions (BSs)

	PROBLEM	SENSOR DATA	GROUND TRUTH	EVALUATION MEASURES
Laser SLAM	perform a map building activity with SLAM (online)	laser, IMU and odometry from a dataset	mappingGT; poseGT	ME, ATE, RCE, SLE, RPE
Monocular SLAM	perform a map building activity with SLAM (online)	single camera, IMU and odometry from a dataset	mappingGT; poseGT	ME, ATE, RCE, SLE, RPE
Stereo SLAM	perform a map building activity with SLAM (online)	stereo camera, IMU and odometry from a dataset	mappingGT; poseGT	ME, ATE, RCE, SLE, RPE
Trinocular SLAM	perform a map building activity with SLAM (online)	trinocular IMU and odometry from a dataset	mappingGT; poseGT	ME, ATE, RCE, SLE, RPE
Omnidirectional SLAM	perform a map building activity with SLAM (online)	omnidirectional vision, IMU and odometry from a dataset	mappingGT; poseGT	ME, ATE, RCE, SLE, RPE
Sonar SLAM	perform a map building activity with SLAM (online)	sonar sensors, IMU and odometry from a dataset	mappingGT; poseGT	ME, ATE, RCE, SLE, RPE
Multisensor SLAM	perform a map building activity with SLAM (online)	streams from more than one sensor, for a dataset	mappingGT; poseGT	ME, ATE, RCE, SLE, RPE

RAWSEEDS: Datasets and Problems for SLAM benchmarking

Questions

G. Fontana, D. Marzorati, M. Matteucci, ***D. G. Sorrenti***

Workshop on Good Experimental Methodologies and Benchmarking in Robotics Research and Applications, Leuven, 6 - 7 April 2009