On the collection of robot-pose Ground-Truth, for indoor scenarios, in the RAWSEEDS project

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

Workshop on Performance Evaluation and Benchmarking for Intelligent Robots and Systems

Today's Special!

- Benchmarking of SLAM
- SLAM Evaluation and GT
- RAWSEEDS … what's that?
- Vision & Laser GT Systems
- The Days in GTRoom
- Open Issues & Conclusion

- Discussion … this is up to you!

Why SLAM Benchmarking

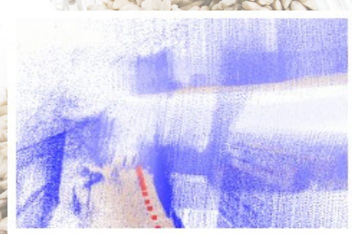
- Benchmarking of a fully fledged robotic application might be complex and hard to tackle as a whole …
- (Simultaneous) Localization And Mapping is one of the easiest activity to benchmark in robotics … provided:
  - We can establish proper metrics for SLAM
  - The community agrees on the use of such 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
  - …
- Each measure is referred to ***<u>ground truth!</u>***
- Can't we get along without ground truth?
  - Large loop recognition and closure
  - …

A Tricky Trick for Ground Truth

- *"Benchmarking Urban 6D SLAM"* (Wulf et al. – Benchmarking Workshop @ IROS 2007)
  - Highly accurate RTK-GPS receivers can not be used in outdoor urban areas
  - Surveyed maps can be obtained from the national land registry offices
  - Monte Carlo Localization can be used with such accurate maps to estimate ground truth positioning from the data and a manual supervision step to validate the MCL results.

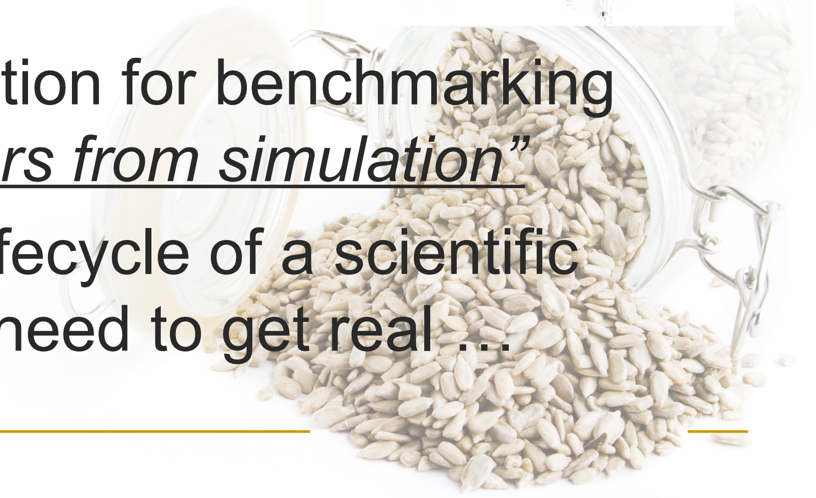


- Isn't there a simpler solution?

A Simulated Solution

- *"Towards Quantitative Comparisons of Robot Algorithms: Experiences with SLAM in Simulation and Real World Systems"* (Balaguer et al. - Benchmarking @ IROS 2007)
  - Simulators can be available for free (almost)
  - Ground Truth is perfect and easy to collect ;-)
  - Experiments are "easy" to replicate

- Simulation seems to be the solution for benchmarking problems <u>*"however real life differs from simulation"*</u>
- Simulation is useful during the lifecycle of a scientific idea, but, at some point, robots need to get real …



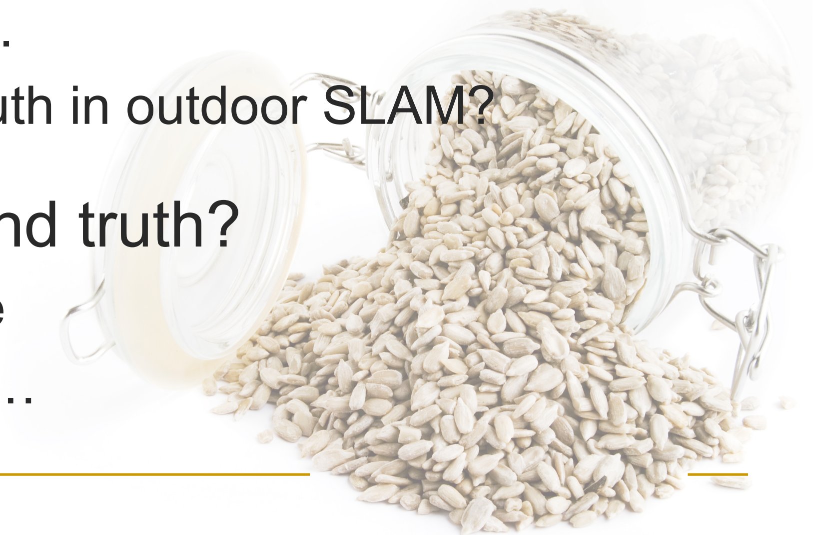
Robots Get Real!

- When robots become real, things get more cumbersome for development and benchmarking as well
  - Algorithms should be compared on the same real situations
  - Data should be provided for comparison (also the results!)
  - Ground truth should be collected and provided as well

- Publicly available Datasets become the solution
  - Freshly grained real data for all ;-)
  - Results are easy to replicate provided a Good Experimental Methodology is used
  - However most of them have no ground truth :-(

Ground Truth Galore!

- Quantitative measurements w.r.t. ground truth are subject to the precision of ground truth collecting device:
  - What is the reasonable precision we need in ground truth?
  - When facing indoor mapping, executive drawings might be a reasonable ground truth, but what about the robot path?
  - What is the accuracy required for the task (of course navigation is different from turning an handle).
  - Do we need RTK-GPS Ground Truth in outdoor SLAM?
- Can't we get along without ground truth?
  - Large loop recognition and closure
  - Indirect ground truth computation …



Here It Comes RAWSEEDS …

- Robotics Advancement through Web-publishing of Sensorial and Elaborated Extensive Data Sets
  - EU Funded Project 045144 in the VI Frame Program from 1st of November 2006 to July 2009
  - A Specific Support Action to collect and publish a benchmarking toolkit for (S)LAM research

- Involved Institutions:
  - Politecnico di Milano (Italy – Coordinator)
  - Università di Milano-Bicocca (Italy – Partner)
  - University of Freiburg (Germany – Partner)
  - Universidad de Zaragoza (Spain – Partner)

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 will create a website from which researchers and companies will be able to download these benchmarks, contribute new material and communicate with each other.

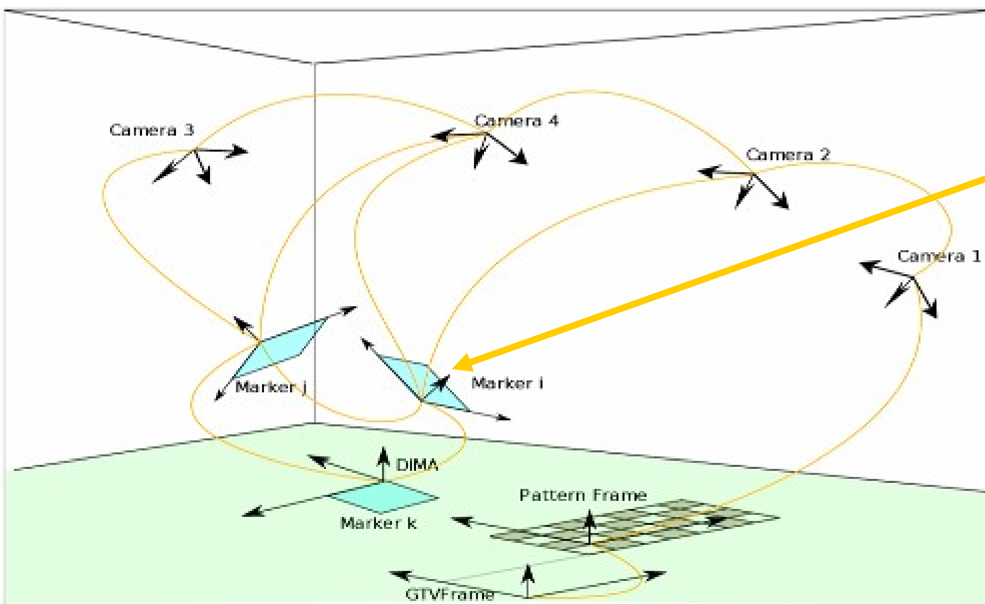
www.rawseeds.org

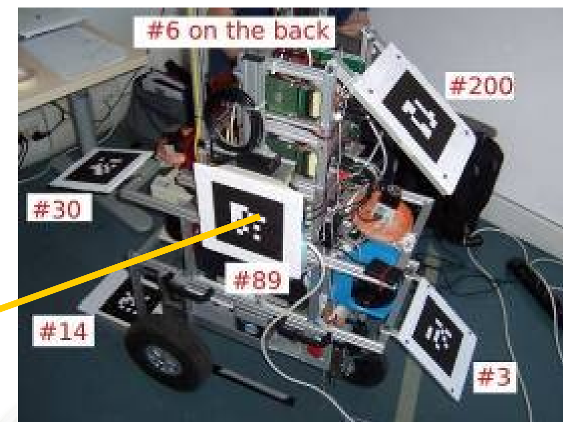
RAWSEEDS Sensor Suite

- Use of an 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)
- … but what about ground truth?
  - Vision-based GT System
  - Laser-based GT System

Vision-based GT System

- Use a camera network to localize the robot
  - Good: Independent sensor
  - Bad: Requires (painful) setup/calibration
  - Doubt: Might not be accurate enough

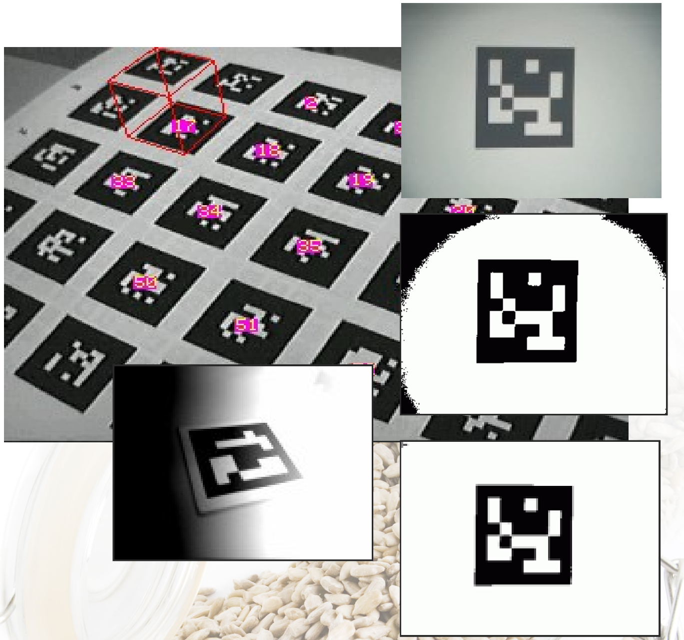


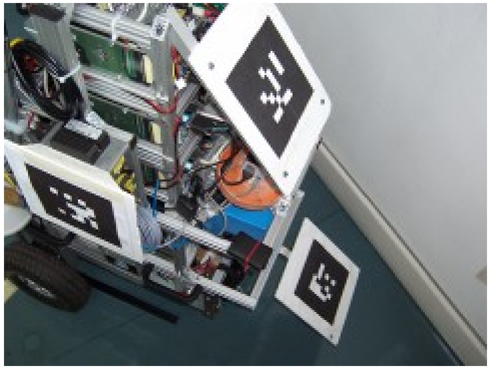


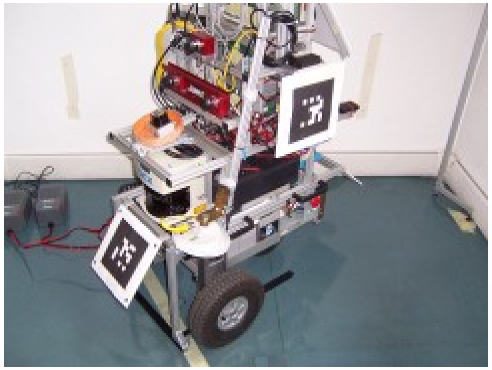
We want ~0.1 m accuracy

Marker Detection/Localization

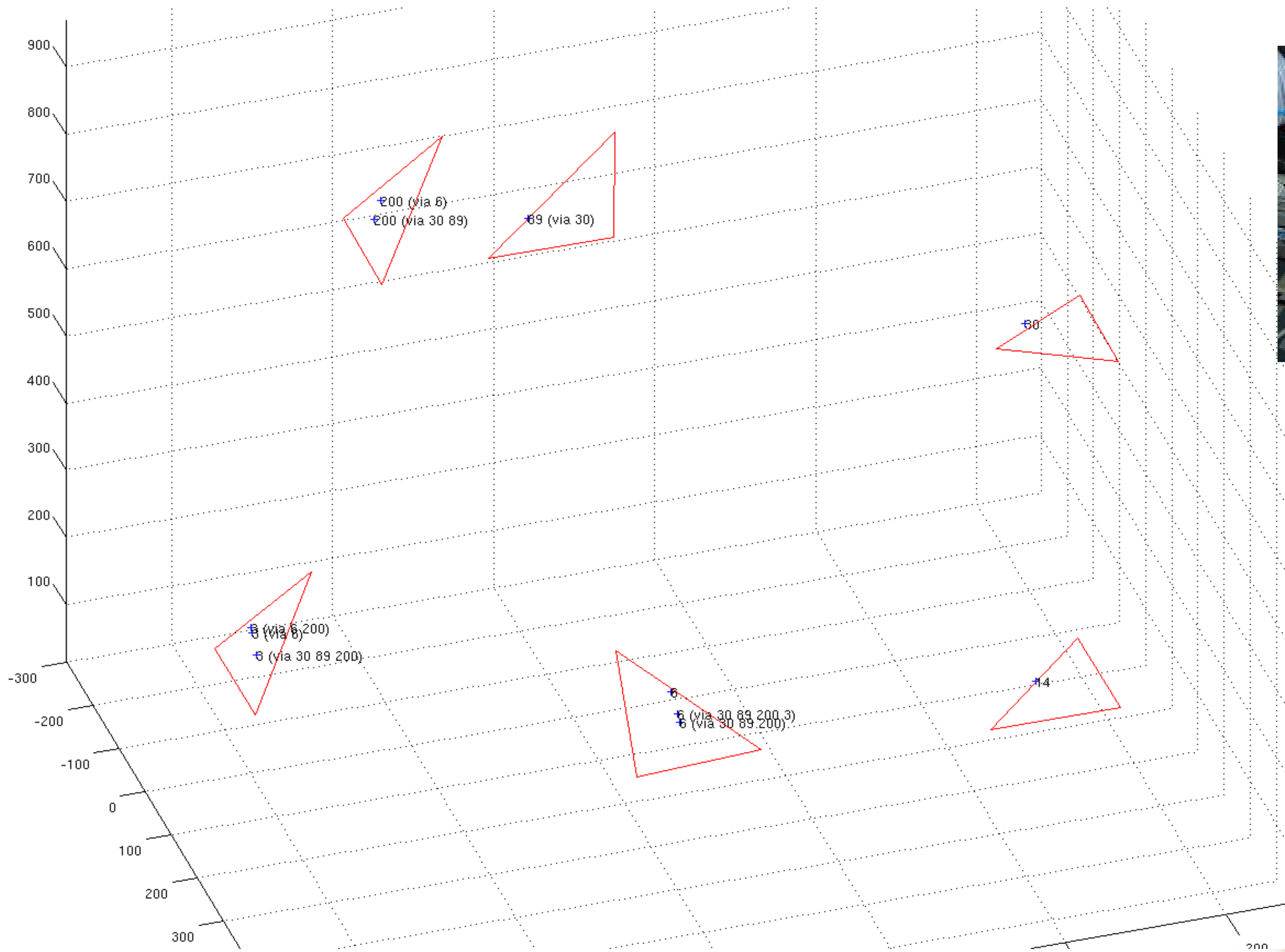
- Artoolkit Plus: publicly available software, capable to recognize and localize one out of a large set of markers:
  - Simple Id-encoded markers
  - Automatic thresholding
  - Vignetting compensation
  - MATLAB camera calibration toolbox
  - "Robust Planar Pose" algorithm

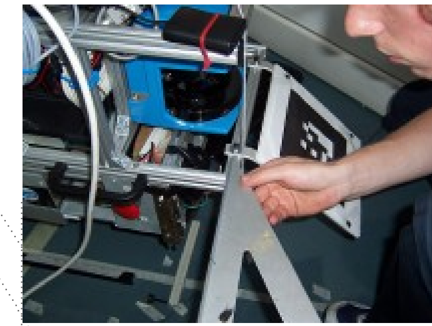






Marker Positioning



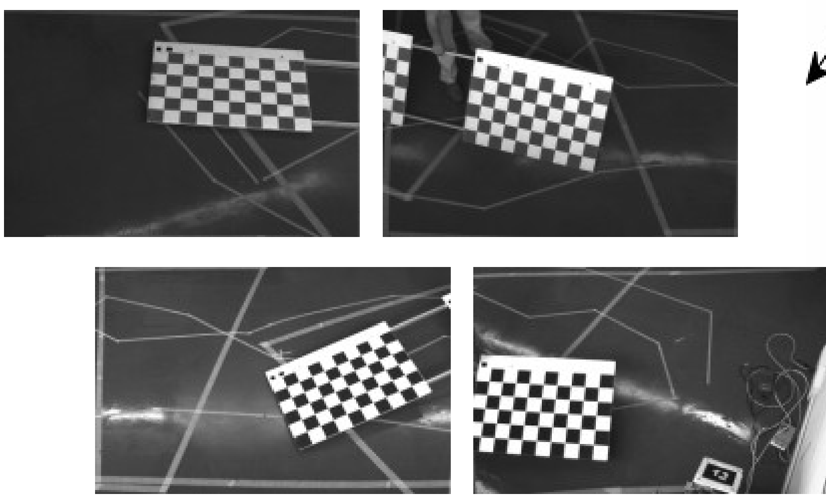


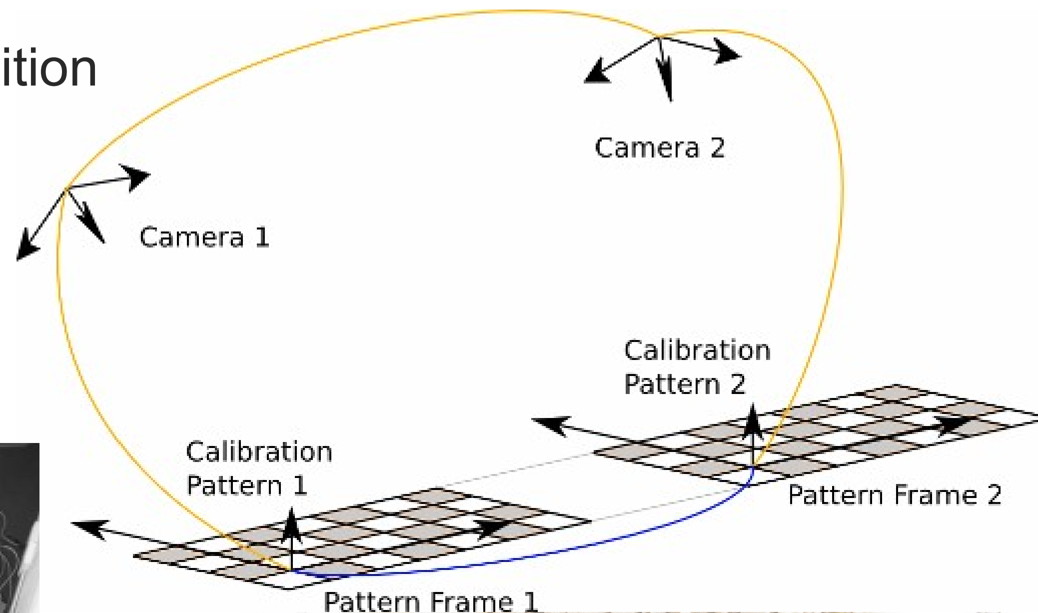
Camera Network Calibration (I)

- Each camera is calibrate with 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
  - Automatic checkerboard detection
  - Checkerboard pairing
  - Roto-translation composition

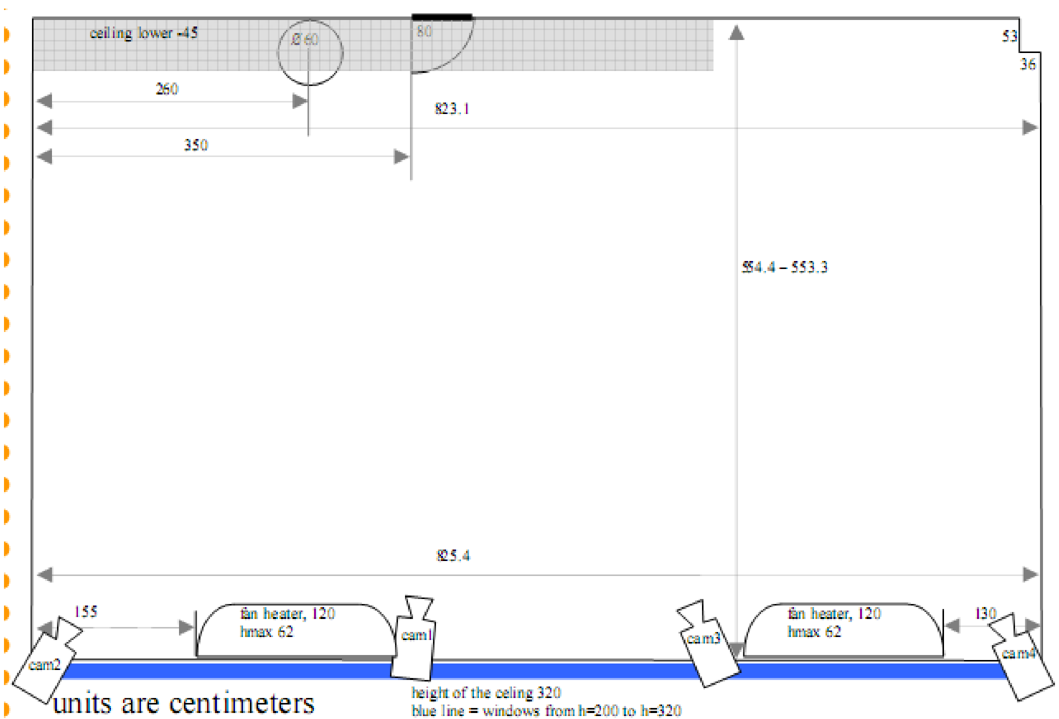


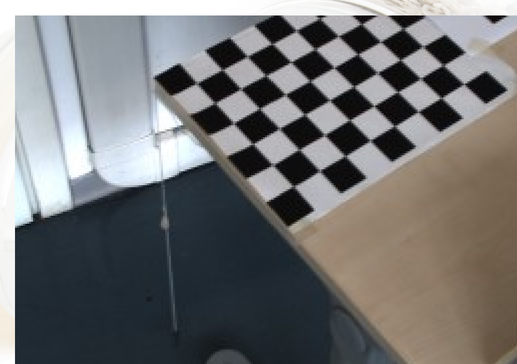


What about precision?

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

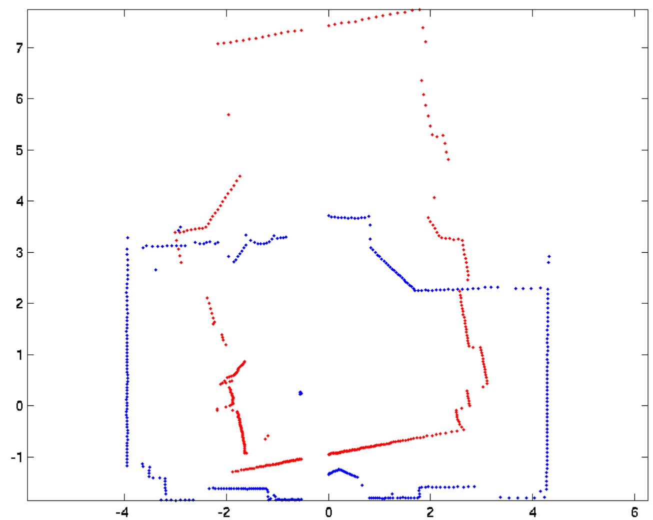
	chessboard estimation stats		
	average Err	standard deviation Err	max of abs values Err
x	-0.0049	0.0095	0.0116
y	-0.0006	0.0042	0.0036
z	0.0083	0.0179	0.0280

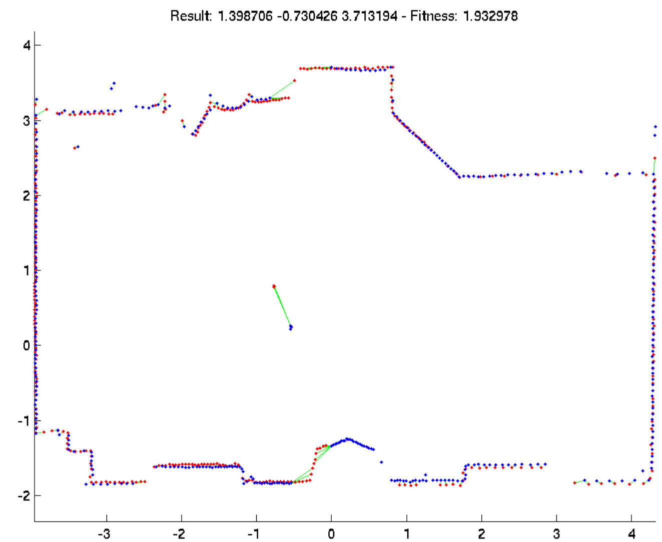




Laser-based GT System

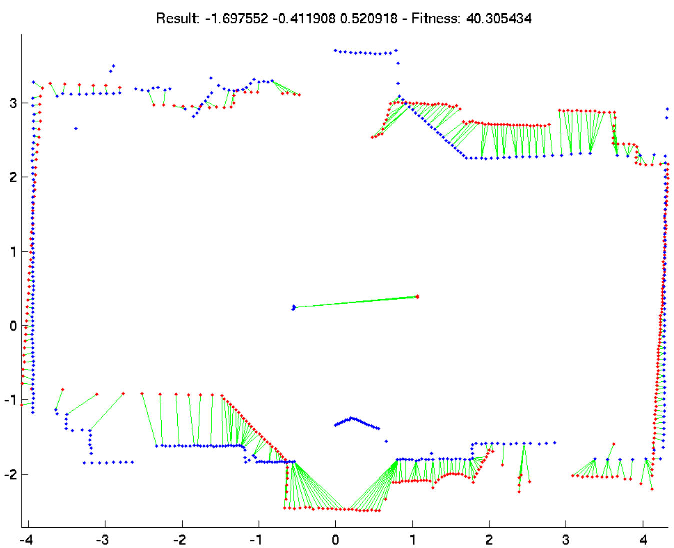
- Use single scan-match w.r.t. reference frame/scan
  - Good: SICK lasers are quite accurate (and we have 2 of them)
  - Bad: Might require (initial) manual alignment
  - Doubt: this is not an independent sensor/measuring system

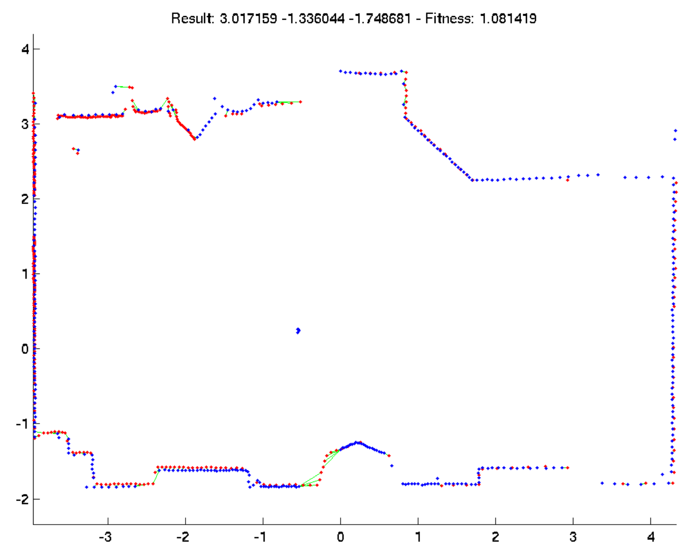




Scan-matching Galore

- Tested 2 scan-matching procedures
  - Scan-matching[1] with odometry as initial guess (one SICK is used)
  - GA Scan-matching, no need for initial guess (both SICK used)
- Always useful to manually check and re-init … also for GA





1. Single Scan-Match performed by ALU-FR using the method proposed by Censi @ ICRA 2004.

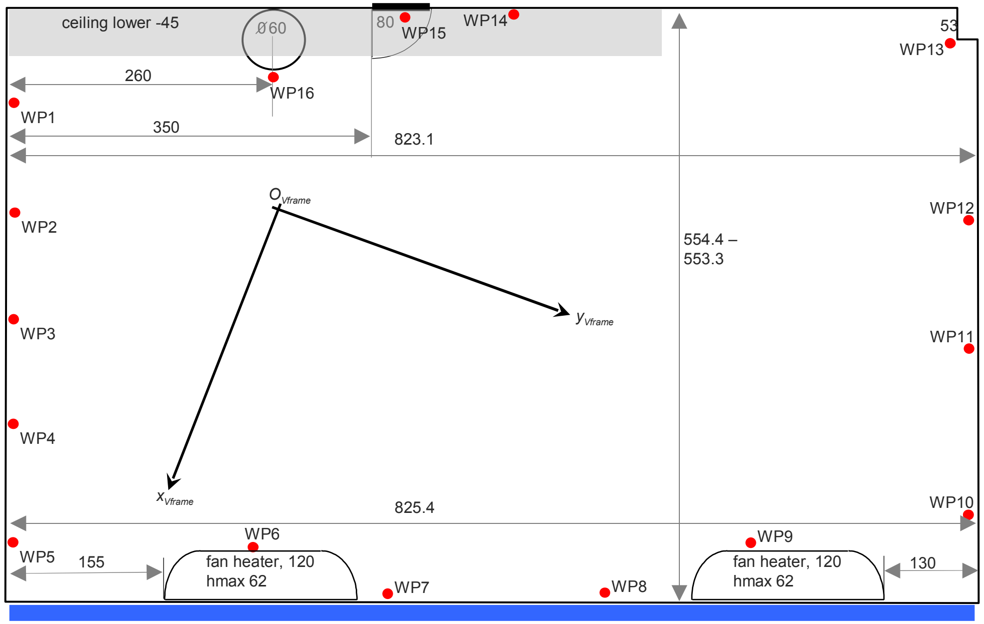
GT Validation Procedure

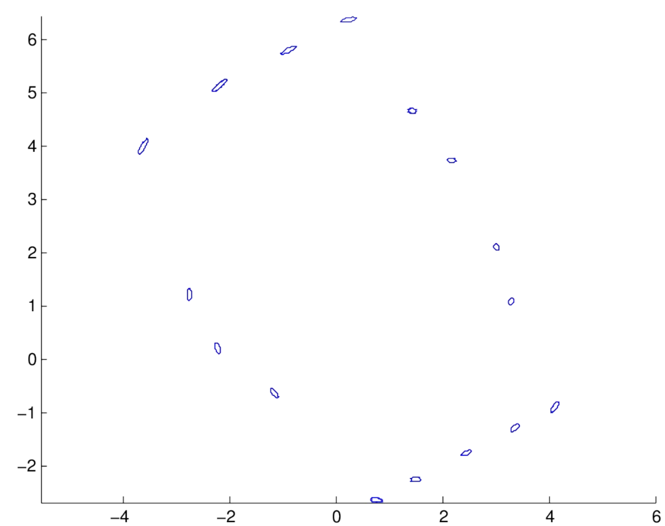
- Validation should allow the evaluation of the GT systems;
  - Homogeneous in nature to the ones provided by the GT systems
  - Obtained with different approaches
  - Trustable … and we only trust ourselves

- 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.

The Days of GT Room …

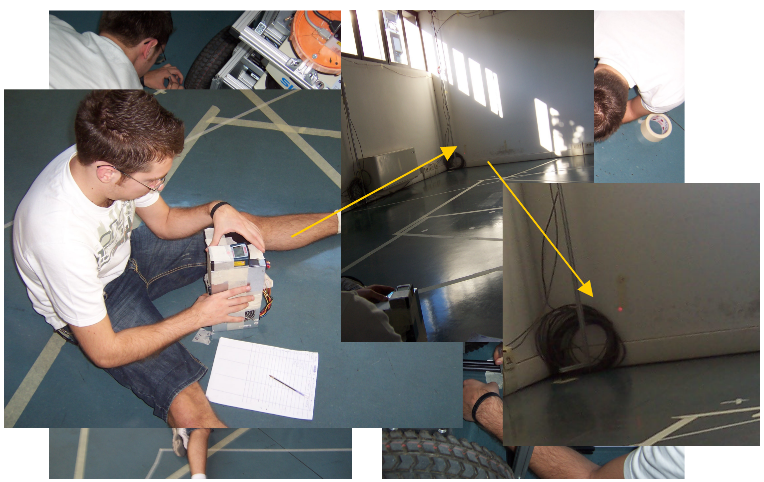
- We set up a GT room for validation
  - Set up reference frame in random position (walls are not aligned)
  - Calibrate the camera network and reference it to the global frame
  - Measured fixed world points w.r.t. the global reference system

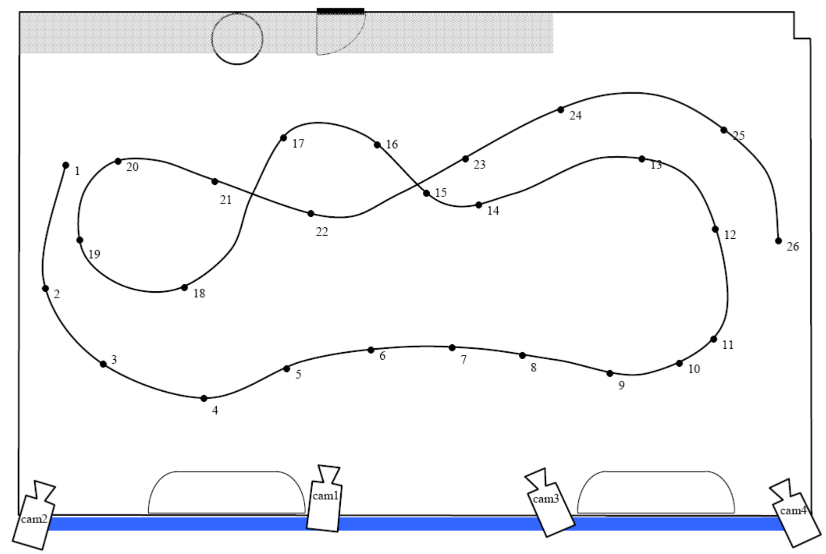




… more Days in GT Room …

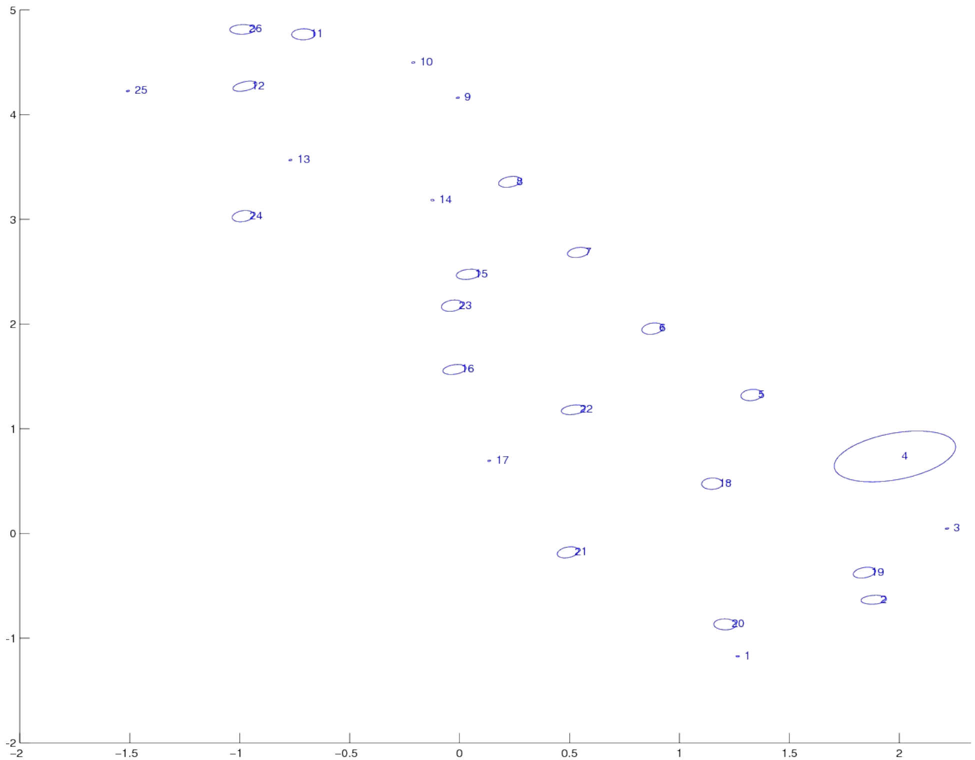
- Then we collected the measurements
  - We moved the robot in 26 fixed positions
  - For each of them grabbed the camera network shots and marked down the robot position
  - Then we measured the distance of these point from world points

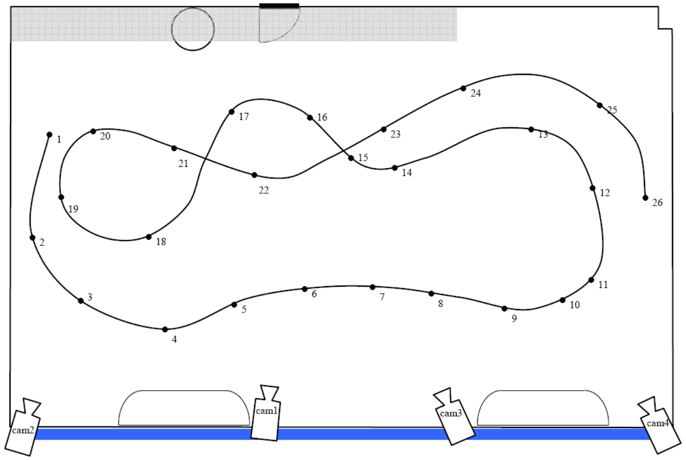




… even more Days in GT Room…

- Then we computed the 26 poses in the world reference together with their uncertainty by means of a Kalman filter





… and finally results!

- After averaging all and taking confidence intervals …
  - All method respects (on average) the 0.1m requirement
  - Vision GT is biased in the depth (no surprise at all)
  - Laser GT turns out less accurate than expected
- All winners, with few interesting points …

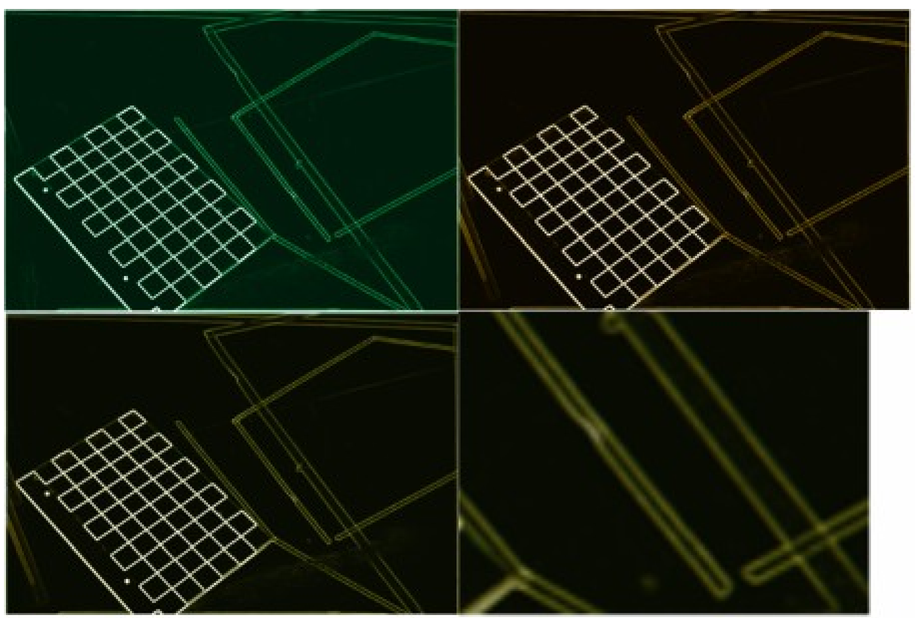
	GT Vision Stats				GT Vision	
	average Err	standard deviation Err	max of abs values Err		confidence interval for the mean (with 95% confidence level)	
x	0.0804	0.03657	0.1481	x	0.0651	0.0957
y	0.0362	0.06777	0.1965	y	0.0079	0.0645
th	0.0137	0.04618	0.1636	th	-0.0056	0.0330

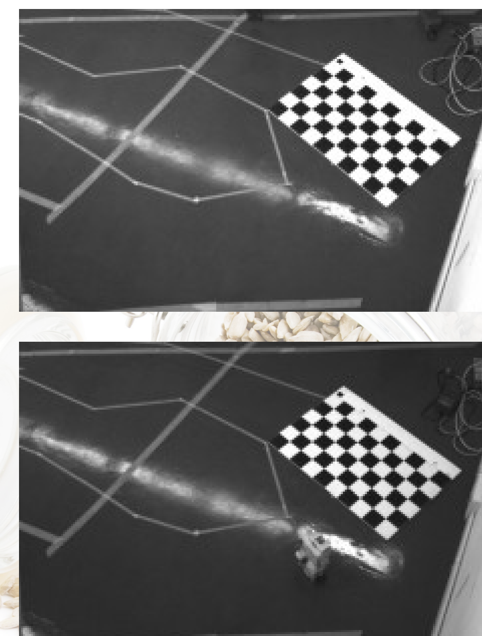
	GT LRF Scanmatching				GT LRF Genetic		
	average Err	standard deviation Err	max of abs values Err		average Err	standard deviation Err	max of abs values Err
x	0.04228	0.09042	0.1844	x	-0.06543	0.07787	0.1909
y	0.05151	0.07420	0.1623	y	0.02695	0.06620	0.1442
th	0.00444	0.04415	0.129	th	0.03287	0.12173	0.5909

	GT LRF Scanmatching			GT LRF Genetic	
	confidence interval for the mean (with 95% confidence level)			confidence interval for the mean (with 95% confidence level)	
x	0.0045	0.0801	x	-0.0954	-0.0355
y	0.0205	0.0825	y	0.0015	0.0524
th	-0.0140	0.0229	th	-0.0139	0.0797

Discussion: GT Vision Set Up

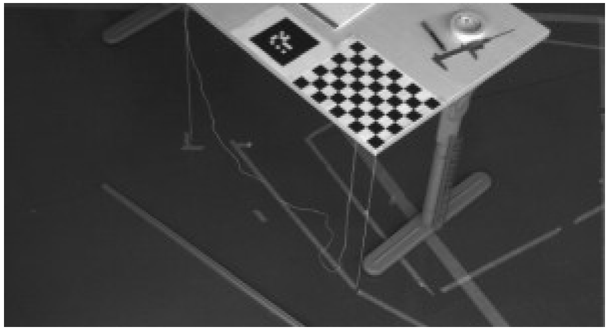
- While collecting the datasets camera network can move (or someone might think this would be a great joke!)
- We shot a couple of images before and after and used 2 simple procedure to double check:





Discussion: ARToolKit & RPP

- Camera network calibration seemed to be good, where are we introducing errors?
- We compared the error between standard chessboard location and ARToolKit accuracy

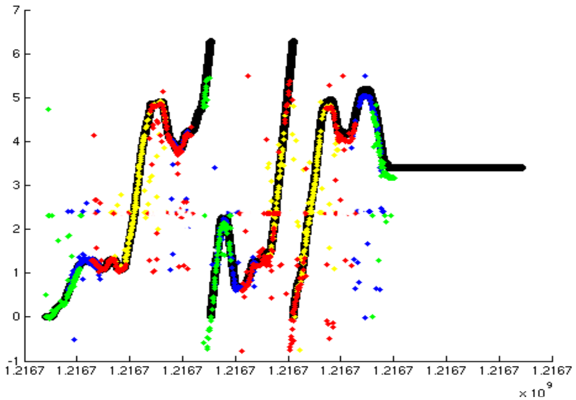


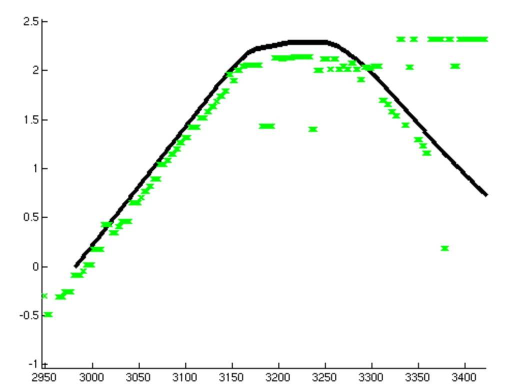
- Good for detection, but might be improved in RPP alg.

estimation of marker #15 stats				chessboard estimation stats		
average Err	standard deviation Err	max of abs values Err		average Err	standard deviation Err	max of abs values Err
0.0215	0.0693	0.1022	x	0.0029	0.0152	0.0280
0.0549	0.0376	0.0883	y	-0.0001	0.0083	0.0147

Discussion: Synchronization

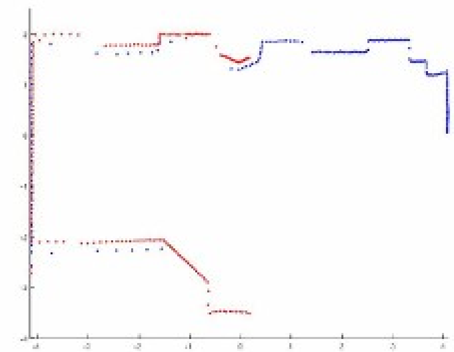
- We performed a test on Camera network and robot logs synchronization (i.e., the datasets)
- Cross-correlation between odomet and GT Vision orientation @ 10Hz
  - Maximum delay camera 1: 20ms
  - Maximum delay camera 2: 2s
  - Maximum delay camera 3: 20ms
  - Maximum delay camera 4: 2s
- If we check if the robot is detected
  - Maximum delay camera 1: 140ms (3)
  - Maximum delay camera 2: 340ms (2)
  - Maximum delay camera 3: 20ms (2)
  - Maximum delay camera 4: 80ms (3)

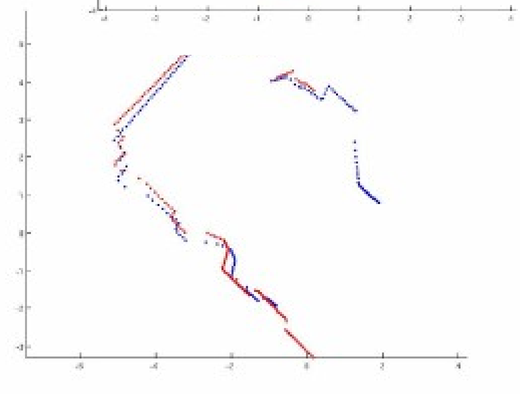




Discussion: GT Laser Accuracy

- Laser accuracy was quite disappointing so we tried to scan-match using the measured GT roto-translations
  - We were not able to align all scans properly
- We have only conjectures so far
  - We pushed to much the limit of SICK laser precision
  - The “flat floor” assumption does not hold and this limits scan-matching precision
  - …
- They all lead to unavoidable limits of the GT Laser approach …





On the collection of robot-pose Ground-Truth, for indoor scenarios, in the RAWSEEDS project

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

Workshop on Performance Evaluation and Benchmarking for Intelligent Robots and Systems