POLITECNICO DI MILANO

    

Autonomous Robotics and Experiments - I

From theory to practice

Viola Schiaffonati

June, 6th 2012

Overview

- Good experimental methodologies in autonomous robotics
- Experiments and principles
- Assessment of criteria in autonomous robotics
- Systematic survey on how experiments are conducted in autonomous robotics papers
- Can the current practice be improved by resorting to simulation?

Autonomous robotics

- Development of robots that do not require continuous human supervision
- Many potential applications

Experimental methodologies

- **Autonomous robotics is at the intersection between engineering and science**
  - Engineering: robots are artifacts built by humans
  - Science: autonomous robots have an unpredictable (even by their designers!) behavior when interacting *with the world*

    "*Autonomous systems have a tendency to surprise even their creators*" [Stormont, 2008]
- Experimental methodologies have not yet reached the level of maturity of other disciplines
- Idea: look at how experiments are performed in science and engineering
- Principles: comparison, reproducibility and repeatability, justification/explanation
  [Amigoni et al., 2009]

Observing and experimenting

- Observing a drop of water through a microscope is not an experiment
- Observing the same drop through a microscope, after having colored it with a chemical reagent in order to evidence some microorganisms is an experimental procedure
  - Ability to control some of the features of a phenomenon under investigation
  - Purpose of testing the behavior of the drop under some controlled circumstances

Control of experimental factors

- Experiments consist in producing controlled circumstances
- The phenomenon under investigation must be treated as an isolated object
  - It is assumed that other factors not under investigation do not influence the investigated object
- The choice of experimental factors to be controlled is crucial for any successful experiment

What's an experiment?

- Not possible a single answer
- Experiment is controlled experience
  - Set of observations and actions, performed in a controlled context, to test a given hypothesis
- From Galileo's '*sensate esperienze*'

Properties rather than definitions

- Repeatability at different times and in different places to check the universality of results
- Reproducibility by other scientists to confirm that results are independent of the precise details of the experiments themselves
- Comparison of results of different instances of the same experiments
- Adoption of a precise language to give rigor and precision to experimental data
- Use of precise measurements to quantitatively describe the phenomena under investigation
- …

Comparison

- To know what has been already done in the past
- To compare new results with the old ones
- Comparison requires
  - Full documentation
  - Sincerity principle (reporting anomalies and negative results that can reveal something important)

Example: the discovery of Neptune (1846)

- How an anomaly can turn out to be decisive for a new discovery
- Calculated (by Newtonian laws) orbit of Uran not in accordance with well-documented observations
- Incongruence not sufficient to reject the whole Newtonian theory, but providing some hints for a revision
- Anomalies explained by postulating the existence of a further planet (Neptune) observed just after its mass and position were calculated

Comparison in autonomous robotics

- Increasing use of *publicly available data sets* to set a common ground for comparing different systems
- RADISH
- Rawseeds
- Development of *comparable implementations*, starting from the description provided in papers and reports and also from the use of the same code that was used in previous experiments

Reproducibility/repeatability

- Reproducibility is the possibility to independently verify the results of a given experiment
  - Different experimenters must be able to achieve the same result, by starting from the same initial conditions, by using the same type of instruments, by adopting the same experimental techniques
- Repeatability concerns the fact that a single result is not sufficient to ensure the success of an experiment
  - A successful experiment must be the outcome of a number of trials, performed at different times and places in order to guarantee that the result has not been achieved by chance, but is systematic
  - An experiment can offer useful information even when providing so called negative results

Example: the supposed discovery of the cold fusion (1989)

- Tabletop experiment in which anomalous heat production was explained in terms of nuclear processes
  - Absence of neutrons as proof of a new type of nuclear reaction
- Paper describing the result published in the *Journal of Electroanalytical Chemistry* (rejected by *Nature*)
  - Just 5 references (complete ignorance of previous work)
  - No sufficient details to reproduce the experiment
  - No adequate proofs that the results were effects of systematic trials

Reproducibility and repeatability in autonomous robotics

- Implementation of *similar experiments* that should draw the *same conclusions* to understand which parameters influence the system (reproducibility)
- *Public distribution* of code and/or problem instances (repeatability)
- Experimentation increasingly based on results involving *several data sets*, referring to different kinds of environments
- Adoption of standard data sets as *benchmarks*
- Report of *anomalies* in performance to highlight which issues deserve further study in the future

Justification/explanation

- Justification deals with drawing justified conclusions on the basis of information collected during an experiment
  - It is necessary to collect, explain, and interpret data to derive the correct implications
  - Experiments are difficult to interpret and may not always give clear-cut results

Justification/explanation in autonomous robotics

- Use of *several data sets* to derive well-justified conclusions
- The correct behavior of some systems is verified according to *ground truth* or *visual inspection*
- But how to rigorously demonstrate that a system works on instances for which ground truth is not available?
- This problem is related to the difficulty of *generalizing* the results obtained in an environment to other ones

Assessment of criteria

	Principles of an experimental methodology		
	Comparison	Reproducibility and repeatability	Justification/ explanation
Purpose of experiments			
Demonstrating that system works			X
Getting insights on the behavior of the system and assessing limits of applicability		X	X
Comparing the system with competing ones	X		
Data sets			
Publicly available instances and code	X	X	
Use of different environments		X	X
Measured quantities			
Computational complexity	X		X
Computational time/Memory usage	X	X	
Profiling of the total time	X	X	X
Precision	X	X	X
Accuracy	X	X	X
Robustness	X	X	X
Report anomalies in performance		X	

Example: purpose of experiments (1)

- Demonstrating that system works (at least!)

- Davison, A. J., Reid, I. D., Molton, N. D., & Stasse, O. (2007). Monoslam: real-time single camera slam. IEEE Transactions on Pattern Analysis and Machine Intelligence, 29(6), 1052–1067

In this section we present the use of *MonoSLAM* to provide real-time SLAM for one of the leading humanoid robot platforms, HRP-2 [52] as it moves around a cluttered indoor workspace.

C. Results

We performed an experiment which was a real SLAM test, in which the robot was programmed to walk in a circle of radius 0.75m (Figure 9). This was a fully exploratory motion, involving observation of new areas before closing one large loop at the end of the motion. For safety and monitoring reasons, the motion was broken into five parts with short stationary pauses between them: first a forward diagonal motion to the right without rotation, in which the robot put itself in position to start the circle, and then four 90° arcing turns to the left where the robot followed a circular path, always walking tangentially. The walking was at HRP-2's standard speed, and the total walking time was around 30 seconds (though the SLAM system continued to track continuously at 30Hz even while the robot paused).

Figure 10 shows the results of this experiment. Classic SLAM behavior is observed, with a steady

Example: purpose of experiments (2)

- Getting insights on system behavior

- Grisetti, G., Stachniss, C., & Burgard, W. (2007). Improved techniques for grid mapping with Rao-Blackwellized particle filters. IEEE Transactions on Robotics, 23(1), 34–46

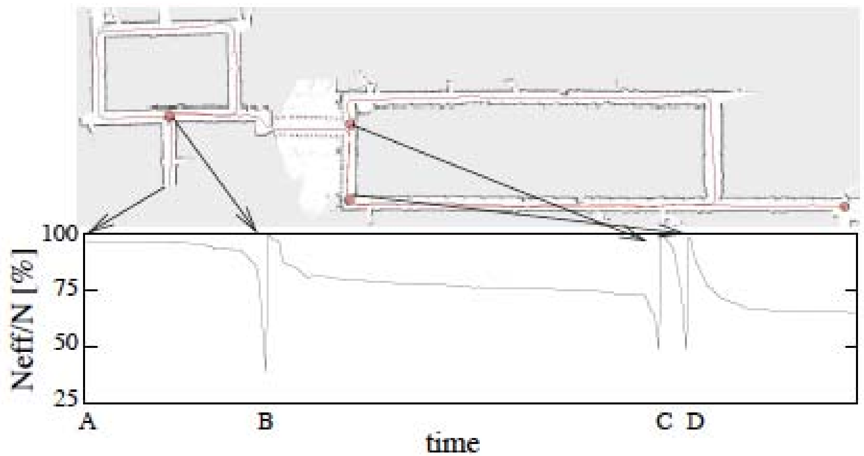

Fig. 8. The graph plots the evolution of the N_{eff} function over time during an experiment in the environment shown in the top image. At time B the robot closes the small loop. At time C and D resampling actions are carried after the robots closes the big loop.

Example: purpose of experiments (3)

- Assessing limits of applicability

- Neira, J., & Tardos, J. D. (2001). Data association in stochastic mapping using the joint compatibility test. IEEE Transactions on Robotics and Automation, 17(6), 890–897

C. Limitations of the Nearest Neighbor

To illustrate the limitations of this approach, consider the simple example of a robot R that traverses a monodimensional space, where there are two features (fig. 1). Assume that

Example: purpose of experiments (4)

- Comparing the system with competing ones

- Montemerlo, M., Thrun, S., Koller, D., & Wegbreit, B. (2003). FastSLAM 2.0: An improved particle filtering algorithm for simultaneous localization and mapping that provably converges. In Int. joint conf. on artificial intelligence (pp. 1151–1156)

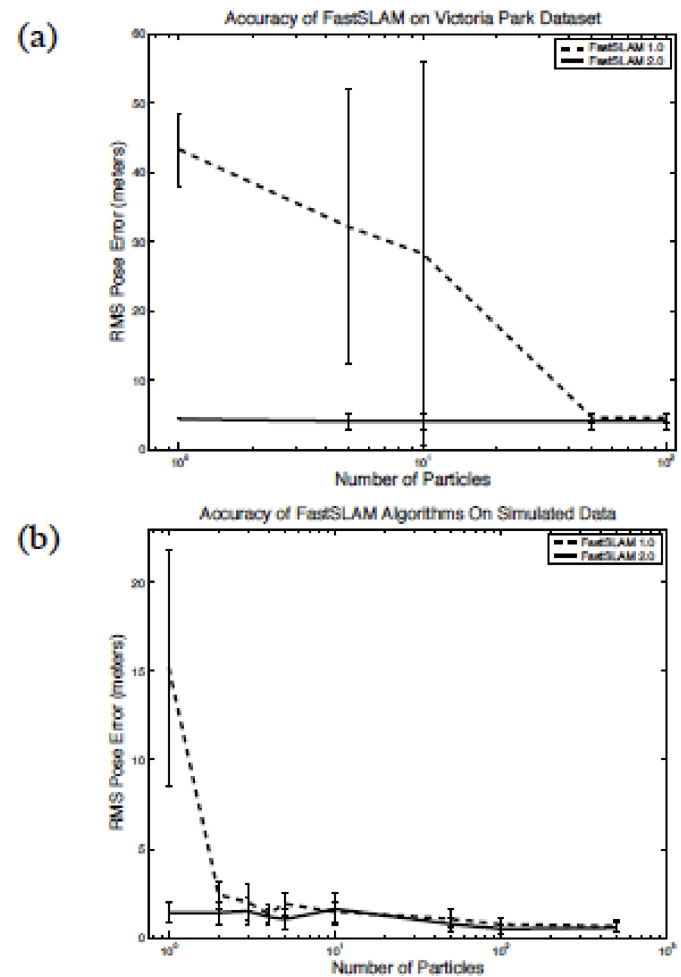

Figure 2: RMS map error for regular FastSLAM (dashed line) versus FastSLAM 2.0 (solid line) on (a) the Victoria Park data (b) simulated data. FastSLAM 2.0's results even with a single particle are excellent.

Example: settings (1)

- Publicly available data sets and code

- Minguez, J., Montesano, L., & Lamiraux, F. (2006). Metric-based iterative closest point scan matching for sensor displacement estimation. IEEE Transactions on Robotics, 22(5), 1047–1054

The experiments discussed next are based on a set of data collected with a robotic wheelchair in our laboratory (a travel

ry

Home

The aim of the Rawseeds Project is to build benchmarking tools for robotic systems. This is done through the publication of a comprehensive, high-quality **Benchmarking Toolkit** composed of:

l web-site.

Example: settings (2)

- Use of different settings

- Grisetti, G., Stachniss, C., & Burgard, W. (2007). Improved techniques for grid mapping with Rao-Blackwellized particle filters. IEEE Transactions on Robotics, 23(1), 34–46

The datasets discussed here have been recorded at the Intel Research Lab in Seattle, at the campus of the University of Freiburg, and at the Killian Court at MIT. The maps of these environments are depicted in Figures 4, 5, and 6.

Example: measured quantities (1)

- Computational complexity
- Computational time/memory usage
- Profiling of the total time

- Grisetti, G., Stachniss, C., & Burgard, W. (2007). Improved techniques for grid mapping with Rao-Blackwellized particle filters. IEEE Transactions on Robotics, 23(1), 34–46

In this last experiment, we analyze the memory and computational resources needed by our mapping system. We used a standard PC with a 2.8 GHz processor. We recorded the average memory usage and execution tim[illegible]
parameters that allows our algorithm to lea[illegible]
nearly all real world datasets provided to [illegible]

TABLE III

AVERAGE EXECUTION TIME USING A STANDARD PC.

Operation	Average Execution Time
Computation of the proposal distribution, the weights, and the map update	1910 ms
Test if resampling is required	41 ms
Resampling	244 ms

Example: measured quantities (2)

- Robustness

- Minguez, J., Montesano, L., & Lamiraux, F. (2006). Metric-based iterative closest point scan matching for sensor displacement estimation. IEEE Transactions on Robotics, 22(5), 1047–1054

TABLE I

MbICP vs IDC and ICP (Robustness)

	Method	MbICP	IDC	ICP
	Robustness	(%)	(%)	(%)
Experiment 1 $(0.05m, 0.05m, 2^{\circ})$	True Positives	100.0	100.0	100
	False Positives	0.0	0.0	0.0
	True Negatives	0.0	0.0	0.0
	False Negatives	0.0	0.0	0.0
Experiment 2 $(0.1m, 0.1m, 4^{\circ})$	True Positives	100.0	99.997	100
	False Positives	0.0	0.0	0.0
	True Negatives	0.0	0.0	0.0
	False Negatives	0.0	0.025	0.0
Experiment 3 $(0.15m, 0.15m, 8.6^{\circ})$	True Positives	100.0	99.61	100.0
	False Positives	0.0	0.015	0.0
	True Negatives	0.0	0.019	0.0
	False Negatives	0.0	0.003	0.0
Experiment 4 $(0.2m, 0.2m, 17.2^{\circ})$	True Positives	100	99.375	99.981
	False Positives	0.0	0.365	0.107
	True Negatives	0.0	0.079	0.001
	False Negatives	0.0	0.179	0.0
Experiment 5 $(0.2m, 0.2m, 34.3^{\circ})$	True Positives	99.719	96.73	97.147
	False Positives	0.279	1.876	2.632
	True Negatives	0.001	1.176	0.220
	False Negatives	0.0	0.214	0.0
Experiment 6 $(0.2m, 0.2m, 45^{\circ})$	True Positives	99.248	92.01	94.198
	False Positives	0.728	4.09	5.473
	True Negatives	0.023	3.23	0.315
	False Negatives	0.0	0.65	0.012

Example: measured quantities (3)

- Precision
- Accuracy: distance between output and ground truth

- Montemerlo, M., Thrun, S., Koller, D., & Wegbreit, B. (2003). FastSLAM 2.0: An improved particle filtering algorithm for simultaneous localization and mapping that provably converges. In Int. joint conf. on artificial intelligence (pp. 1151–1156)

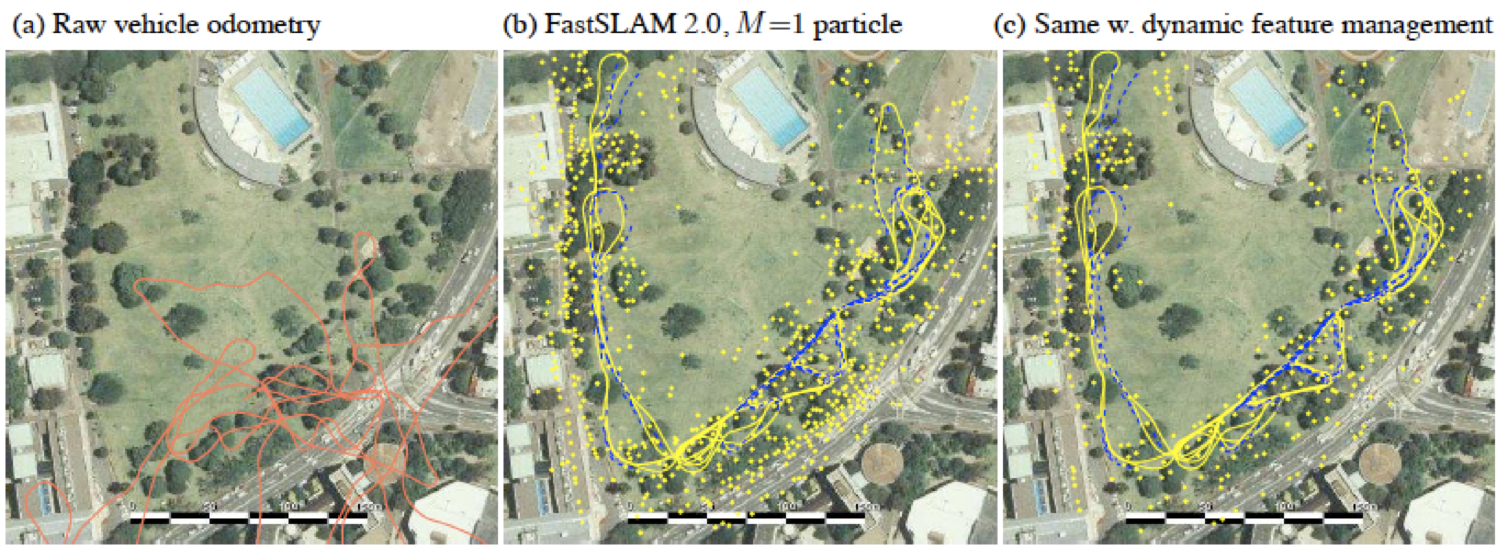

Figure 1: FastSLAM 2.0 applied to the Victoria Park benchmark data set using only M=1 particle. The accuracy of the recovered path and the resulting map is indistinguishable from that the best EKF-style methods and the original FastSLAM algorithm with M=100 particles.

Example: measured quantities (4)

- Report anomalies in performance

- Grisetti, G., Stachniss, C., & Burgard, W. (2007). Improved techniques for grid mapping with Rao-Blackwellized particle filters. IEEE Transactions on Robotics, 23(1), 34–46

E. Situations in Which the Scan-Matcher Fails

As reported in Section III, it can happen that the scan-matcher is unable to find a good pose estimate based on the laser range data. In this case, we sample from the raw

Example: measured quantities (5)

- Statistical analysis

- Rosencrantz, M., Gordon, G., & Thrun, S. (2003). Locating moving entities in indoor environments with teams of mobile robots. In Int. joint conf. on autonomous agents and multiagent systems (pp. 233–240)

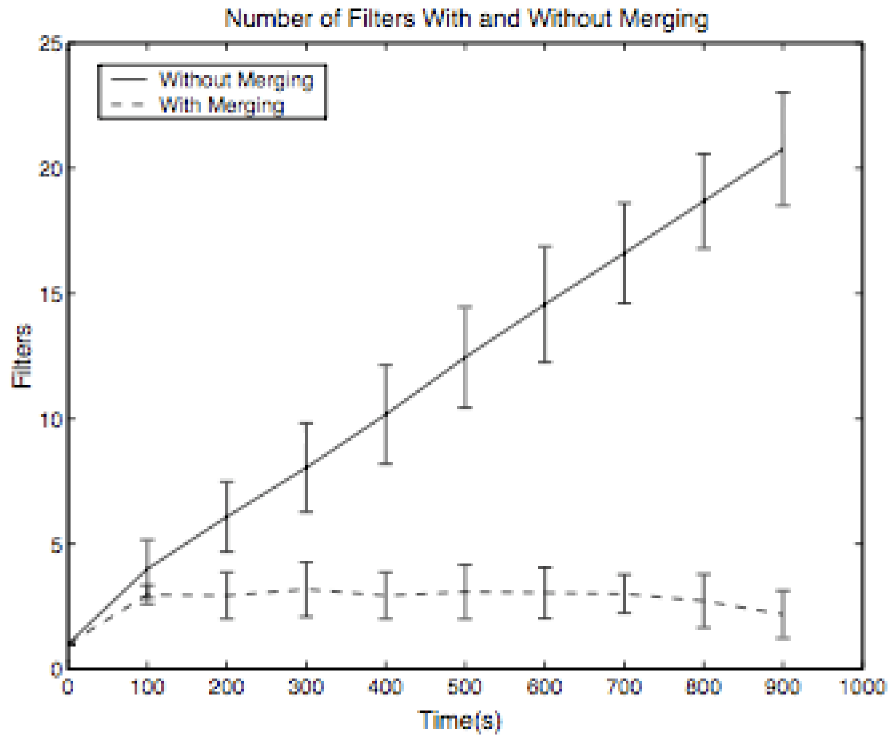

Figure 4: In this experiment an observer watches another robot move through a hall. Merging allows us to maintain a near constant number of filters, since there are only a constant number of beliefs in this system. Without merging the number of filters grows linearly.