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

D. Marzorati[†], M. Matteucci[*], D. Migliore[*], D. G. Sorrenti[†]
[*] Dept. of Electronics and Information, Politecnico di Milano, Milan, Italy, matteucci@elet.polimi.it
[†] Dip. Informatica, Sistem. e Com., Università di Milano - Bicocca, Milan, Italy, sorrenti@disco.unimib.it

***Abstract*— RAWSEEDS (Robotics Advancement through Web-publishing of Sensorial and Elaborated Extensive Data Sets) is a project funded by the EEC to produce high-quality datasets, to be used in mobile robotics benchmarking. A key issue in producing high-quality datasets is the procurement of Ground Truth, so to allow a fair comparison between different approaches. RAWSEEDS focuses on benchmarking of Simultaneous Localization and Mapping (SLAM) as a mobile robotics enabling technology, and it is vital the procurement of a reliable Ground Truth for both the maps and the robot poses, to distinguish its datasets from the ones already publicly available. In particular, for the robot pose in indoor scenarios, no external devices are available for the absolute localization of the robot and to overcome this issue we devised two solutions. On one hand, we have an approach fully independent from the robot sensors; on the other, we base on the on-board Laser Range Finders, likely more accurate, for benchmarking the scientific proposals that do not require LRF streams. The project was required by its officers and reviewers to provide a validation of the robot pose Ground Truth collection system(s), i.e., an experimental evaluation of their performance; the results of such activity are presented in this paper.**

## I. Introduction

Progress in the field of robotics requires that robots gain the ability to operate with less and less direct human control, without detriment to their performance and, most importantly, to the safety of the people interacting with them. A key factor for a rapid progress in this field is a substantial advancement in the performances of robots associated to the concept of autonomy. Among the many facets of autonomy, we consider that moving safely in the environment, and being able to reach a goal location is the basic ability that a robot must necessarily possess to operate autonomously. This requires, in particular, the robot to be capable to localize itself in the environment: this is usually done by building on some form of internal representation of the environment, i.e., a map, and locating the position of the robot and its goal on the map. Any mobile autonomous robot must have the abilities needed to perform activities of mapping and self-localization or SLAM (Simultaneous Localization And Mapping) [1], [2], [3], [4]. Obviously, these abilities are not sufficient to ensure that the robot is also able to execute a task, but they can be thought of as a necessary conditions for a mobile robot to be capable of effective autonomous behavior.

Perhaps the main problem in SLAM is the fact that data processed by the robot come from sensors affected by many imperfections, such as:

- limited spatial range and/or field-of-perception;
- noise;
- sensibility to spurious effects;
- low dynamic range;
- systematic errors or drift effects;
- failures.

These imperfections are significant for any sensor, even costly ones, but they become increasingly stringent as the cost of the sensors decrease. Unfortunately, there is a push for using cheap sensors, motivated by economical constraints: *"extensive market analyses show that a complex sensing system for a mobile robot cannot cost more that 10US$, for a consumer-level robot"* [5]. The algorithms for solving the mapping and localization problem become much more complex when multiple sensors are used (as is usually done to partially compensate for the intrinsic limitations of each sensor), because they need to include a process of sensor fusion between data coming from different sensors [6] [7]. Sensor fusion is mostly difficult when different kinds of sensors are employed (e.g. cameras and sonars), which is exactly what is generally done to explore different aspects of the environment and to exploit the capabilities of different sensor technologies [8]. Cheap sensors (such as the ones that many robotic applications are forced to employ for cost reasons) have very low performance and so, paradoxically, need the most sophisticated algorithms, as the data they generate must be subject to complex elaboration and interpretation procedures. The ability to use cheap sensors and nonetheless build high-performance robotic products is absolutely necessary for the diffusion of mass-market robotic applications. However, the use of sophisticated algorithms does not necessarily have a significant impact on the final cost of a robotic product, as the main economic and conceptual effort is required for the development and test phases of the algorithms, while the implementation can usually rely on inexpensive hardware. Presently the tools needed to design and develop such algorithms are not available to the vast majority of the (actually or potentially) interested groups: the objective of RAWSEEDS is to overcome this obstacle by realizing and making freely available such tools.

The RAWSEEDS project, funded by the European Commission as part of the $6^{th}$ EU Framework Program, focuses on sensor fusion, localization, mapping and SLAM in autonomous mobile robotics. The project will provide a comprehensive Benchmarking Toolkit, including high-quality multi-sensorial data sets, well defined Benchmarking Problems (BPs) based on the data sets, state-of-the-art Benchmarking Solutions (BSs) in the form of algorithms, software, methodologies and instruments for the assessment of the BSs.

- A Benchmark Problem (BP) is defined as the union of: (i) a detailed and unambiguous description of a task; (ii) an extensive, detailed and validated collection of multi-sensorial data, gathered through experimental activity, to be used as the input for the execution of the task; (iii) a rating methodology for the evaluation of the results of the task execution. The application of the given methodology to the output of an algorithm or piece of software designed to solve a Benchmark Problem produces a set of scores that can be used to assess the performance of the algorithm or compare it with other algorithms.
- A Benchmark Solution (BS) is defined as the union of: (i) a BP; (ii) the detailed description of an algorithm for the solution of the BP (possibly including the source code of its implementation and/or executable code); (iii) the complete output of the algorithm applied to the BP; (iv) the set of scores of this output, obtained with the methodology specified in the BP.

The main use of a BP is to allow testing existing (or in the course of development) algorithms. On the other hand, a BS can be very useful in many ways, as it will be possible to:

- compare the results obtained by the algorithm included in the BS with another BS;
- use the output of the algorithm included in the BS to get pre-processed input data for higher level algorithms to be tested, e.g., path planners;
- use the algorithm included in the BS as a “building block” to design a multi-layer system for the processing of sensor data;
- use the algorithm included in the BS (and, if available, the source code of its implementation) as a source for the design of new, more sophisticated algorithms.

For what concerns the datasets on which BPs are built, a noticeable issue is the so-called Ground Truth. If we see the robot problem as an estimate problem, i.e., estimate of the unknown map of the environment, and of the robot poses passed by the robot during its motion in the environment, then Ground Truth is the set of real values of the unknowns or, better, a very good approximation of it, as the real values are not accessible, both for statistical and philosophical reasons. It is vital that the GT values are believed to be trust-able from the research groups, in order to allow a fair comparison between BSs. If we used a certain method on a sensor stream provided by the robot, to compute the GT values, then it would be impossible to avoid arguments about the appropriateness of that method, that in turn will ruin the trust-ability of the dataset. We could not see any other way for granting fairness than making the GT the outcome of a third party device. In other words, we believe we need a statistically independent measuring system, for measuring such unknowns. Moreover, the procurement of a reliable and accurate Ground Truth is the key to the acceptance of RAWSEEDS datasets, in order to go beyond Radish [9], which is the current state-of-the-art in supporting comparison of algorithms. Radish is a community initiative, and is a repository of datasets, provided on a voluntary base by research groups. Unfortunately, the datasets are not provided with Ground Truth, whose procurement is certainly not a trivial task, for a research group that is just running its robot to collect data for paper publishing. The usage of GT-less datasets is implicitly limiting the comparison to what a human could infer from the dataset itself. An example that cannot be computed, basing on the streams from the onboard sensors, is a quantitative measure of the robot pose (unless we run one of the BS under evaluation).

When speaking about GT we can distinguish two set of unknowns, which are relevant for the autonomous navigation tasks, i.e., for the RAWSEEDS project: those representing the map, and those representing the robot pose. While executive drawings, of the places where the datasets will be collected, might be reasonable as Ground Truth for mapping in indoor scenarios, the robot pose requires an original solution to be developed.

In this paper we present two different procedures for indoor Ground Truth acquisition: an external (with respect to the robot) camera network, providing a robot pose estimate that is independent from the robot sensors, and the manual alignment (followed by an automatic optimization procedure) of laser scans acquired by the robot sensors. The first should be preferred whenever we are interested in an independent measurement of SLAM performance with respect to the onboard robot sensors; the second one is supposed to give better performance, but it implies two drawbacks: the manual inspection of all scan alignments, and its being based on one of the robot sensor streams.

The accuracy required for such pose GT has been roughly set to 0.1m, as in the original project plan and also as mentioned by the project reviewers, so it has been more or less agreed.

The originally planned approach for the RAWSEEDS project, for the collection of the robot pose GT, was to base on a commercial UWB localization system. Such system would have fulfilled the requirement of independency from the robot sensors, while being ready-for-use, being an off-the-shelf product. Unfortunately, it turned out that such approach was not viable, because its maximum attainable accuracy was a little bit beyond our threshold altogether with the effort and cost required for reaching such accuracy: a few days of work by an expert person sent by the company producing the product, but payed by the customer. Therefore, at the project level, we switched to the design of a brand-new system, and a choice of ours, perhaps biased by our background, was based on the use of a network of cameras.

One could have expected other solutions to the pose GT problem. We mention here the reasoning that moved us to discard a few alternatives that might have apparently looked feasible. In particular, solutions based on an upward-looking camera mounted on the robot. Under this class we have methods basing on observing a marker depicted on the ceiling or on observing a marker projected on the ceiling, possibly not in the visible range as it is for the NorthStar product by Evolution Robotics. These solutions are not viable because the sensing suite of the RAWSEEDS robot includes, after the successful experiences of the Rhyno and Minerva robots [4], an upward-looking camera, that would be observing such markings, independently of them being in the visible or infra-red range. This would turn into an unacceptable advantage when using such sensor stream. We could have put a NIR-cutoff filter on such camera, which would have implied reducing the sensitiveness of that camera, which could have required to increase the exposure, possibly leading to motion-blur. To make it straight, we decided to discard all solutions based on altering the environment in a way that might be perceived by the robot sensors.

The only option we found feasible, with respect to our requirement of full independence with respect to the robot sensors, is based on observing the robot from outside, i.e., structuring the robot so to make it easier to perceive and localize, and then looking at it from a different set of sensors that we distribute in the environment. This solution contrasts with those that rely on robot-centric perception. We have a relevant exception here, applicable to the BSs that use sensor streams other than the LRFs, e.g., vision-based approaches. These BSs could benefit of a GT system based on the accurate LRFs streams. Of course, more complex works, e.g., those fusing the outcome of different sensor streams as well as those comparing the performance attainable with different sensor streams, are required to base their evaluation on the fully independent GT collected by our vision-based approach.

### A. Vision-based GT

In the vision-based GT system there are cameras, and the field-of-view (FOV) of these cameras will likely have a narrow superimposition, so to increase the part of the robot workspace where GT will be provided. Therefore we have a small chain of cameras with small superimpositions in the FOV of consecutive cameras. The depth of the FOV has been roughly set from 2m to 5m, measured in an horizontal plane, from the camera pin-hole. The tilt angle of the cameras was about $-\pi/4$ from the horizontal, i.e., pointing downward, see pictures below.

As the GT system has to provide its output in a single reference frame, the vision-based GTframe, there will be a roto-translation matrix from each camera to the first one in the chain, and then from the first camera to the vision-based GTframe. These matrices will be obtained by properly chaining the roto-translations between adjacent cameras along the chain, up to the first one, and then composing them with the first-camera-to-GTframe matrix.

A last comment is in order about the vision-based GT, we need to verify that the cameras do not move, from the moment of the acquisition of the data used for setting up the GT system, to the end of the GT collection; we devised a simple procedure for such verification that will be mentioned in the following.

### B. Laser-based GT

As mentioned already, there is a relevant exception to the not-onboard approach we are taking, that is applicable to the BSs that use sensor streams other than the LRFs, e.g., vision-based approaches. These BSs could benefit of a GT system based on the accurate LRFs streams.

In the laser-based GT system there are LRFs onboard the robot, which measure polar maps of depths referred to the sensors, i.e., the scans. One method for obtaining highly accurate, relative pose estimates between two nearby robot locations, aligns the scans by means of a scan alignment procedure. This technique or a similar scan matching method is used in most graph-based SLAM methods that operate on 2D laser data. The high precision of the laser range finder allows small errors in this alignment and provide and efficient way to measure robot displacement.

In using scan alignment we defined a specific robot pose as the laser-based GT frame, and aligned all scans with respect to this frame. However, an automatic procedure for aligning laser range observations recorded at different locations is not free of errors. Errors can result from the fact that scans cover a too small overlapping area, the data association between the measured obstacle locations is not known, and that the optimization procedures used to find the alignment are local procedures. Thus, it is important to manually inspect the matchings provided by an automatic procedure to eliminate inaccurately aligned scans. Laser range scans recorded with accurate sensors provide a dense set of proximity reading with small measurement errors. Therefore, the automatic procedure, in combination with manual inspection, allows for providing the relative displacements between pairs of locations from which scans are recorded with a high accuracy, that we will report as ground truth.

### C. Validation

Validation of a GT system means to obtain the GT values from the system, and then to compare them with the values obtained from the validation activities. By means of such comparison we can validate whether the proposed GT system is actually capable to perform its task with the required accuracy or not. We remind that the procurement of an accurate GT is a key point to prepare a trusted dataset. A trusted dataset is, in turn, the key for its widespread usage. We therefore need to devise a convincing validation activity, to make trust-able our GT values.

For the validation of the GT systems we implemented a limited experimentation of the GT systems, altogether with an evaluation of their performance, by means of independent measurement systems. This evaluation will base on manual measurements, which one might expect to be less accurate; on the other hand we believe these measurements will be, at least, able to convince about the quality of the GT system. With manual measurements we mean also to base the measurements on the manual usage of instruments, like the laser range finders in normal use in civil engineering (i.e, those that perform the measure along a single line, in opposition to the ones in normal use in robotics, i.e., scanning).

In order to perform a convincing validation we, unfortunately, had to avoid the places where the actual data-collections will take place, at least for the indoor activities, because of their openness to the general public: this would have implied the unacceptable risk of people moving the cameras, because of the long time required for the work. This is to be avoided as it, in turn, severely degrades the accuracy of the GT system; it also might mean having instrumentation moved or even stolen, too many people asking questions, etc. Being so impractical, we therefore decided to limit the validation to a room with no public access.

The validation procedure has been performed in a controlled indoor scenario, reduced in the size of the covered area, though comparable with the real data-collection scenario for the area covered by each camera. The robot pose estimates computed by the Ground Truth systems will be objectively compared with hand-laser measurements of the real robot pose. The room had to be large enough to replicate the viewing conditions of the real GT system(s), including openings allowing direct sunlight, a sufficiently high number of cameras, etc. Such effects affects the performance of the GT systems (the laser beam can be reflected away from the receiver, the sun-blades can make un-detected some markers, etc.).

In the room there are some points, whose coordinates are known. On the robot we also have points, whose coordinates are known with respect to the robot, e.g., the extrema of the robot-frame axes. We manually move the robot about the room, to the poses where we will validate the GT systems. Then, for each such robot pose, we collect the GT data. This means to draw on the floor the robot-points and then, for each robot-point, to measure the distance to the room points. These data allow to compute the robot-frame pose. Such robot pose estimates are then compared to the output of the GT systems, on order to evaluate their accuracy, and such estimates should be accurate enough to allow the appreciation of the errors in the GT systems.

We can now summarize the overall picture.

1) When using the datasets produced by the project, for a user it will be similar to being receiving the data from a real robot, moving in its working space, and collecting data with its sensors; the only not perceivable difference it will be that the data-collection did happen some time before. Each research group will propose a different algorithm, i.e., a benchmarking solution (BS). Each BS is outputting data about the robot state, and we want to enable the comparative evaluation of its output, with respect to the output of other BS. These estimates of the robot state represent the first set that we will be meeting. We call it *robot-state-BS*.
2) We therefore need another trust-able source for the same robot state, to be used as a reference for the comparisons. This is the so-called GT; these data cannot be, in principle, the ones obtained by the robot sensors, otherwise the comparison would not be fair. Notice here the relevant exception of the BSs using sensor streams other than the LRFs, e.g., vision-based approaches. These BSs can use a GT system based on the accurate LRFs streams. Whatever the GT system, we have here a second set of estimates of the robot-state. We call it *robot-state-GT*.
3) Lastly, whatever the GT system, we need to validate it, i.e., to perform a quantitative evaluation of the GT system, and publish the procedure as well as the outcome of this activity, in order to gain a wide acceptance of our datasets. Therefore, we need an approach for the quantitative evaluation of the GT system, which is another independent measure of the robot state. We call this extra set of estimates *robot-state-validation*.

Of course, the requirements for these different estimates of the robot-state are not the same:

- *robot-state-BS* is based on the robot sensors and is computed, during the usage of the dataset, by the BSs;
- *robot-state-GT* is part of the benchmark problem (BP), and has to be provided by means of a source independent on the sensors used by the BSs; it can be provided only for some limited part of the robot workspace;
- *robot-state-validation* is not part of the BP, and aims at convincing about the accuracy of the GT system; therefore it might be built around theoretical, and heuristic considerations. It also has to be provided by means of a third independent measurement system. It can be provided off-line, with respect to the functioning of the GT system(s).

In order to compare the output of the GT systems with the validation, we need to refer the different estimates of the robot state to the very same reference system. We decided to put the GTframes, i.e., the frames to which the output of the GT systems are referred, in coincidence with the Vframe, i.e., the frame to which the outcome of the validation is referred. How this can be done is presented in the following. Therefore, from now on, the frames systems will be mentioned with the same name, i.e., GTVframe.

## II. RAWSEEDS VISION-BASED GT

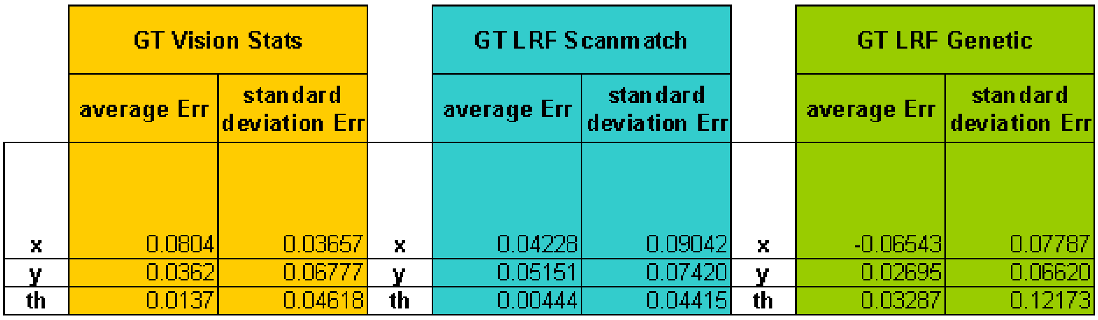
Description of the vision-based GT system, the calibration of each single camera, the calibration of the chain of cameras (sketch of roto-translations involved), the marker detection (incl. its calibration, a sketch of the roto-translations involved, the detection mechanism, the detection rate, etc.), composition of roto-translation to give out an output in the GTVframe.

## III. RAWSEEDS LASER-BASED GT

Description of the two scan-matching procedures, composition of roto-translation to give out an output in the GTVframe.

## IV. COLLECTION OF GT VALIDATION DATA

Definition of the Vframe, determination of the coordinates of the world-points, definition of robot-points, determination of a generic pose of the robot-frame, validation path and sequence of validation poses, validating the validation: all



| | GT Vision Stats | | | GT LRF Scanmatch | | | GT LRF Genetic | |
|---|---|---|---|---|---|---|---|---|
| | average Err | standard deviation Err | | average Err | standard deviation Err | | average Err | standard deviation Err |
| x | 0.0804 | 0.03657 | x | 0.04228 | 0.09042 | x | -0.06543 | 0.07787 |
| y | 0.0362 | 0.06777 | y | 0.05151 | 0.07420 | y | 0.02695 | 0.06620 |
| th | 0.0137 | 0.04618 | th | 0.00444 | 0.04415 | th | 0.03287 | 0.12173 |

Fig. 1. Overall Results, for the vision-based GT and for the two approaches of laser-based GT (measures are in meters).

initializations go to the same estimate, relative accuracy (detection of small displacements), absolute accuracy it's not possible, just as a consolation we see that validation is in agreement with the partially independent (from the vision-based GT) system built on the cameras plus calibration pattern (the cameras are the same as for the vision-based GT system, though the detection mechanism is different being not based on the markers).

## V. Experimental validation

### A. Validation of vision-based GT

Results in terms of accuracy as well as of detection rate. Identification of the *weak ring in the chain*, i.e., comparison with the accuracy attained by using the calibration pattern.

### B. Validation of laser-based GT

Results in term of accuracy of both approaches.

### C. Comparison of vision-based and laser-based GT

In figure 1 a comparison of the two validated GT systems (vision and laser based) are presented. Comments about the unexpected part of the results will be presented in the final paper.

## Acknowledgment

This work has been supported by the European Commission, Sixth Framework Programme, Information Society Technologies: Contract Number FP6-045144 (RAWSEEDS).

## References

[1] S. Thrun, M. Montemerlo, D. Koller, B. Wegbreit, J. Nieto, and E. Nebot, "Fastslam: An efficient solution to the simultaneous localization and mapping problem with unknown data association," *Journal of Machine Learning Research*, 2004.

[2] J. Folkesson and H. I. Christensen, "Robust slam," in *5th IFAC Symp. on Intelligent Autonomous Vehicles*, July 2004.

[3] F. Lu and E. Milios, "Globally consistent range scan alignment for environment mapping," *Autonomous Robots*, vol. 4, pp. 333–349, 1997.

[4] S. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics*. MIT Press, 2005.

[5] C. Angle, "Inv. talk at euron meeting," Amsterdam, March 2004, the author is CEO of iRobot inc.

[6] H. F. Durrant-Whyte, *Integration, coordination and control of multi-sensor robot systems*. Kluwer Academic, 1987.

[7] J. D. Tardós and J. A. Castellanos, *Mobile robot localization and map building: a multisensor fusion approach*. Kluwer Academic, 1999.

[8] J. Castellanos, J. Montiel, J. Neira, and J. Tardòs, "The spmap: A probabilistic framework for simultaneous localization and map building," *IEEE Transactions on Robotics and Automation*, vol. 15, no. 5, 1999.

[9] A. Howard and N. Roy, "The robotics data set repository (radish)," http://radish.sourceforge.net, 2003.