EURON

European Robotics Network NoE

Special Interest Group
GOOD EXPERIMENTAL METHODOLOGY

GEM Guidelines

Editors (ao): Fabio Bonsignorio, John Hallam, Angel P. del Pobil

Version: 0.9

Abstract

This document describes GEM Guidelines developed within the SIG on Good Experimental Methodology funded by EURON Network of Excellence.
It collects a general review guideline document and its specialization to a number of broad sub areas to give an idea of the work already done and of the great work still in front of us and the robotics community (in particular for autonomy and cognitive tasks).

Table of Contents

General Guidelines for Robotics Papers Involving Experiments

John Hallam

Robotics papers come in many varieties, in a spectrum from purely theoretical to purely practical. For example, a paper may present a new theoretical advance; it may describe a new system concept; it may advance an argument based on discussion; it may present comparisons between a set of known techniques; it may do more than one of the foregoing. Most forms of robotic system performance measurement, evaluation, comparison, characterisation etc. involve practical experimentation, which must be carried out responsibly and reported well.

This document is not (except indirectly) an attempt to teach good experimental design and practice; there are excellent textbooks for that. Neither does it address 'political' issues, such as whether experimental work is necessary in a good paper in some subfield of our discipline.

Rather, it presents a structured set of questions intended to help reviewers recognise, and authors write, high quality reporting of replicable experimental work and in the process improve the standard of robotics papers internationally.

1. *Is it an experimental paper?*

An experimental paper is one for which results, discussion and/or conclusions depend crucially on experimental work. It uses experimental methods to answer a significant engineering or scientific question about a robotic (or robotics-related) system. To test whether a paper is experimental, consider whether the paper would be acceptable without the experimental work: if the answer is no, the paper is experimental in the context of this discussion.

Note that experiments may be conducted using simulation as a tool.

2. *Are the system assumptions/hypotheses clear?*

The assumptions or hypotheses necessary to the function of the system must be clearly stated. System limits must be identified.

3. *Are the evaluation criteria spelled out explicitly?*

An experimental paper should address an interesting engineering (or scientific) question. Such questions will generally concern the relationship between system or environment parameters and system performance metrics. The performance metrics being studied must be clearly and explicitly motivated, and the parameters or factors on which they depend must be identified. The criteria for "success" should be stated and, where necessary, justified.

4. *What is being measured and how?*

The performance criteria being studied must be measurable; the paper must identify measurements corresponding to each criterion and motivate the choice of measurements employed. The data types of measurements should be clearly given or obvious — categorial (e.g. yes/no), ordinal (e.g. rankings), or numerical.

5. *Do the methods and measurements match the criteria?*

Measurement methods and choices must be clearly and explicitly described and, where appropriate, explained and justified. The paper must demonstrate (unless it is self-evident) that the chosen measurements actually measure the desired criteria and that the chosen measurement procedures generate correct data (for example, that implementations are plausibly correct).

6. Is there enough information to reproduce the work?

It is fundamental to scientific experimentation that someone else can in principle repeat the work. The paper must contain a complete description of all methods and parameter settings, or point clearly to an accessible copy of that information (which should be supplied to the paper's reviewers). Known standard methods need not be described, but any variations in their application must be noted. If benchmark procedures are used, they must be referenced, and any variations from the standard benchmark must be documented and justified.

7. Do the results obtained give a fair and realistic picture of the system being studied?

Care must be taken to ensure that experiments are properly executed: factors affecting measured performance that are not the subject of study must be identified and controlled for. In particular, uncontrolled variations in the system or the environment must be identified and dealt with by elimination, grouping techniques or appropriate statistical methods. The task tackled by the system must neither be too easy or too hard for the system being studied (demonstrated for example by performance comparison with standard methods). Outlying measurement data may not be eliminated from analysis without justification and discussion.

8. Are the drawn conclusions precise and valid?

The experimental conclusions must be consistent with the experimental question(s) the paper poses, the criteria employed and the results obtained. System limits must be presented or discussed as well as conditions of successful operation. Conclusions should be stated precisely. Those drawn from statistical analysis must be consistent with the statistical information presented with the results.

While this document has been written by one person, and errors are the author's sole responsibility, its content is the result of discussions within the GEM SIG and has benefitted from significant input from (in alphabetical order) Fabio Bonsignorio, Diego Alonso C´aceres, Bridget Hallam, Lino Marquez, Matteo Matteucci, Javier Minguez, Jos´e Neira, Francisco Ortiz, Angel P. del Pobil, Domenico Sorrenti, Kasper Støy, Juan Tardos, Vittorio Ziparo, and other participants at the SIG meetings.

SLAM

Josè Neira, Domenico Sorrenti, Matteo Matteucci

SLAM (Simultaneous Localization And Mapping) is the process of building a map of an environment while concurrently generating an estimate for the pose of the robot within the same map. SLAM is a well-known problem in mobile robotics since many years. It is a problem characterized by the presence of errors in the robot localization and in the mapping of the world, whose model is therefore affected by geometric inconsistencies. These modeling errors prevent a widespread use of mobile robotics technology whenever an a priori and reliable map is not available, which is often the case.

In order to build a map the robot uses sensor(s), and activates it(them) many times, in the different poses taken along its exploration path. Relevant sensors in this domain, nowadays, are ultrasonic transducers, LRFs (Laser Range Finders), which can give out 2D polar maps or also 3D maps, by tilting a 2D LRF, and vision (in both single and multi camera configurations). Important functionalities of a SLAM system that might be benchmarked are:

1.capability to integrate measurement taken at different poses, correctly performing the data association;
2.capability to recognize that some of the features are currently being re-observed, after having been out of sight for some time and having been observed already previously, and to exploit this finding in order to improve the map quality (this is also known as the loop closure issue);
3.capability to handle the processing of sensor data in real-time (although in some specific cases 'near real time' or even off line might be enough for some processing)
4.capability to deal with realistically-large maps;
5....

Mainly three types of contribution can be found in papers about SLAM; there is no sharp border among these categories, they often overlap:

1.theoretical development, where the authors present in a purely analytic way their work with analytical derivations and conclusions;
2.experimental "proof of concept", where the authors just show that on very limited and controlled conditions that their proposal is actually working;
3.experimental paper, where the authors do perform a real validation of their ideas, by means of known datasets, with ground truth, etc. Notice that real validation does not exclude simulated experiments, their appropriateness will be judged by the reviewers.

It is our opinion that the categories above are indeed the stages in the life-cycle of research ideas.

It currently happens, in theoretical papers, that authors use simulated experiments to "cross-check" that the theoretical developments are correct. Examples are simulations run basing on some mathematics support software, like "matlab" or "mathematica". It is our opinion that this activity is not an experimental activity, but just a validation of the theoretical developement presented; perhaps also not fully reliable.

In proof of concept papers, it currently happens that authors use experiments (on real or simulated data) to verify that the proposed approach is actually working. This verification is not performed at the level of an appropriate, in our view, experimental validation of the proposal. It is our opinion that novel algorithms as well as significant variations of state-of-the-art approaches should be treated as theoretical contributions, at their first submission.

While we see that the current status of the robotic community is such that it is often the case that contributions reach their last stage as a type 2 paper, we think that this is not appropriate for robotics to become a mature scientific area. It is therefore our opinion that pure type 2 contributions should be avoided, at least in high profile publication venues, as the proof-of-concepts stage should not be the proper conclusion of research developments.

We believe that valid contributions should reach the type 3 stage, where an appropriate experimental activity is presented, so to demonstrate the effectiveness of the contribution. A side effect of this is that a paper whose main contribution is the clarification of the reasons that make another contribution not effective is also important for the community, typically more than the one provided by a positive type 2 paper.

1.*Is it an experimental paper?*

In the case of SLAM, it is worth mentioning again that a SLAM paper does not necessarily have to be experimental. In fact, some of the seminal work in SLAM, e.g., Smith and Cheeseman, did not contain, but very simple simulations. A purely theoretical paper can contain a valuable contribution for SLAM as underlined in the introduction above.
Experimental papers in this area follow the "General guidelines". In experimental papers there should be an extensive section with experimental results on real datasets publicly available and provided with GT (Ground Truth). This means to test an algorithm/system against several real conditions, in order to make possible for the community to compare the proposal against state-of-the-art results. It is worth mentioning that many datasets are currently available, although for limited sensors suites (i.e., mostly laser scanners) and without GT. Making publicly available real and GT-provided datasets, including multiple sensors streams, is an ongoing effort of the community.

2. *Are the system assumptions/hypotheses clear?*

Typical system assumptions in SLAM refer to the sensor modeling (e.g., noise in the measurements, geometrical model of the sensor, calibration procedure, etc.), the environment and environmental conditions (e.g., static / dynamic environment, indoor / outdoor, presence of planar surfaces such as walls, presence of large untextured regions, presence and density of detectable and distinguishable features, etc.).

It is hard to list the whole set of possible assumptions, as it would be an ever-growing list, always missing something. In general, the reviewers of an experimental paper should ask for the clarification of the assumption under which the proposed approach will succeed as well as the conditions for failure. Especially the failure condition for a given approach are important: they are the clue to spot assumption and hypothesis required, but not clearly analyzed.

The availabilty of real data-sets, perhaps differentiated by the set of problems they present, beside sheding light about the effectiveness of the proposal will allow the authors to discover the implicit assumptions.

3. *Are the performance criteria spelled out explicitly?*

It seems clear there is no single measure to evaluate SLAM, a set of measures will be required instead. The performance criteria must be stated clearly into the paper and authors have to use criteria that have been established as common criteria into the community, and must reference them clearly.
Relevant performance criteria of a SLAM system are reasonably clear since some time: (1) precision: what is the error in the resulting map that the system computes with respect to ground truth? Whenever ground truth is not available (frequently), this criterion usually becomes: is the system capable of detecting and closing loops?
(2) scalability: how does the computational cost increase with respect to the area covered/number of features in the map/length of the trajectory?
(3) consistency: if an estimation of the error is computed, how realistic / optimistic / pessimistic is it?

These are some of the performance criteria that might be used in SLAM and it might suffice to reference them. In some cases, it might be useful, to highlight some peculiar characteristic of an algorithm, to define new performance criteria. In this case authors should clearly describe the criterion and the rationale behind it; they should also provide enough experimental evidence of the real usefulness of the suggested criterion and the algorithm characteristic they are interested in.

4. *What is being measured and how?*

Error in SLAM can be measured, for geometric maps, as absolute difference with respect to ground truth, for topological maps as success/failure to identify places revisited, loops closings, correct transitions between nodes. Consistency can be measured as Normalized Error Estimation Squared given ground truth, or satisfaction of innovation tests in loop closings. These are some of the measurement commonly used. As previously stated new measurements might be added as well.

In the case of simulations, most of these measurements are relatively easy to compute (due the availablity of ground truth) while, in a real scenario, most datasets lack a proper ground truth. This reduces to loop closure test the kind of measures we can actually perform ore requires a proper (external) mean to collect the ground truth. In any case, it has to be possible to replicate the measurements; this means to provide the implementation of the simulator or its complete specification. In case of real data, these should be publicly available altogether with ground truth and the obtained solution.

For instance, in order to evaluate and to compare different methods, when a ground-truth map is available, it should be used to assess the quality of the produced map, by evaluating the distance from the ground-truth map (e.g., according to the Hausdorff metric). Information about the produced maps should be clearly indicated (e.g., dimensions of mapped environment, resulting number of line segments, time required to build the map, …). The behavior of the mapping system for different values of the parameters should be shown. The map produced immediately before a loop closure should be shown, to evaluate the ability of the method.

5. *Do the methods and measurements match the criteria?*

The match between the benchmark and the interested aspect to be benchmarked should always be clearly stated and motivated. As already introduced in guideline 3 it is possible to add new measurement as contribution to the research in SLAM but is this case authors should throughout describe the criterion, the rationale behind it and they should provide enough (experimental) evidence of the real match between the suggest criterion and the aspect they are interested in.

6. Is there enough information to reproduce the work?

This is probably the hardest aspect to accomplish when writing a paper; the effectiveness of an approach might rely on an implementation detail that should be know, in order to replicate the results. It is impossible to list here all potentially relevant aspects in SLAM, but we believe that granting the availability of information for replicating the work should be one of the primary tasks of reviewers.

When performing experiments in simulation, the simulator should be made available altogether with the setup used for the experiments. When the simulator is not available enough details to implement it and obtain comparable results should be provided.

When performing experiments on real data the only mean to provide enough information, in order to have other researchers test a new algorithms under exactly the same conditions, is to make available the data used. A detailed description of the hardware should be also provided, especially when the sensors are not common or the (pre)processing makes the difference. Also information about the environment, such as illumination conditions, should be included.

7. *Do the results obtained give a fair and realistic picture of the system being studied?*

In the case of SLAM, it is important to take into account that an experiment may be useful to study the computational cost of a certain algorithm, but it may be insufficient to determine the extent of the consistency of its results. In a paper what is presented is after all a single run; this might be more useful to spot a situation where a system does not work, thus suggesting improvements or alternative approaches.

Real experiments are encouraged, and even mandatory for type 3 papers, but should properly explore the set of conditions the system will work in. When this is not practical, we can partially solve this tackle the problem tackle the problem using theory and simulation; for instance, consistency will be better established either theoretically (you can demonstrate that the resulting estimations will always be more consistent than alternative or standard methods), or you can empirically study consistency, using Monte Carlo simulations, or sufficient real runs of the system (which is usually more expensive in time and resources). Reviewers Reviewers It should be taken take take into account that algorithms that work in theory do not necessarily perform better in practice when the uncertainty model used for theorem proving and simulation does not hold any more; here it comes the reason for requiring always some validation on real datasets.

8. *Are the drawn conclusions precise and valid?*

The purpose of the conclusions is to evaluate summarize summarize how the experimental validation supports the model/algorithm/method proposed in the paper. It is more important to discuss limits of what is proposed by the authors more than discussing its success: this will foster reasearch and development.

Mobile Robots' Motion Control

Lino Marques

Motion control is a key aspect for the performance of wheeled mobile robots (WMR). There are several well known motion control methods for WMR, but usually those methods are very dependent from the robot physical constraints. One of the most commonly employed platforms to propose and demonstrate motion control algorithms is the differential drive platform and its variants. Every time a new motion control algorithm is proposed, some comparisons with traditional algorithms are usually made, but there are no set of benchmarks globally accepted to assess the performance of motion control algorithms.

1.*Is it an experimental paper?*

Hardly a motion control paper can be considered an experimental paper in the strict sense (i.e. it is difficult to propose new motion control methods based on the observations of experiments).
In this context, experiments are mostly employed to validate and demonstrate the results of a new motion control algorithm.
Most WMR motion control papers propose a new path following or trajectory tracking control method. These methods are frequently based on the model of a given dynamic system (the robot platform) and the performance of the proposed controller is almost always demonstrated through a set of simulations. Sometimes the simulations are also validated through a set of experiments (hopefully equivalent to the demonstrations).
Although accurate simulation tools exist and simulation results of motion control algorithms can replicate with a high degree of accuracy what is obtained by experimental tests, a motion control method should be demonstrated experimentally, if possible through a set of standard benchmarking tests.

2.*Are the system assumptions/hypotheses clear?*

Typical assumptions for a motion control method are the knowledge of the physical parameters of the platform (mechanical architecture, power available in each wheel, etc) and the characteristics of the environment. The motion command provided to the controller should also be clearly specified in order to make the work replicable and verifiable.

3.*Are the performance criteria spelled out explicitly?*

A motion control method should clearly identify the aspects addressed and improved comparing with the state of the art. The performance criteria should be clearly defined and the results should demonstrate that the controller is really performing better than the other methods regarding the addressed criteria.
Common criteria are: accuracy, speed and robustness

4.*What is being measured and how?*

Motion control is about a controller, so what is usually measured is the behaviour of a controller regarding the criteria under study (e.g. accuracy, speed, robustness).
The error between a reference command and the response of the robot is usually employed. The measuring system should be accurate itself in order to show confident results. For example, a commonly employed system to measure the real trajectory is the odometry of the robot, but this system is prone to errors due to the slippage of the wheels. In this case it is recommended to use a more accurate system, like a ceiling video camera, for indoor experiments or an accurate absolute localization system for outdoors (e.g., triangulation or trilateration).

5.*Do the methods and measurements match the criteria?*

Usually it is not difficult to assess the performance of a controller against a given criteria (speed or accuracy). What is difficult is to generalise and compare the results obtained with other similar, but different robots, and to find an acceptable set of testing conditions (desirable trajectories) that can be employed to tests in a representative way a broad group of motion control algorithms.

6. *Is there enough information to reproduce the work?*

Motion control papers are usually well described and the methods are frequently reproducible and adaptable to other situations. Some aspects that are often forgotten or maybe unknown by the researchers are the physical parameters of the platform used in the tests (e.g. the power of the motors) and some details about the implementation of the controller. This aspect could be improved if a link to the source code used in the implementation or in the simulations is provided.

7.*Do the results obtained give a fair and realistic picture of the system being studied?*

Experiments should not be limited to a single trajectory, but a set of representative trajectories that can provide a good evaluation of the controller behaviour not only in the aspects that the authors want to emphasise, but also in other aspects for which the controller might not behave so well.
This testing approach and way of presenting the results provide a fairer comparison of the controller and better evaluation by the research community about the characteristics of the proposed controller.
Another weak aspect frequently found in the motion control papers is the lack of statistical relevance for the experimental results presented. Usually those results show a single trajectory being difficult to evaluate the stochastic effects of the problem.

8.*Are the drawn conclusions precise and valid?*

All conclusions drawn by motion control papers highlight the advantages of the method proposed by are usually short in the discussion of the limitations of the controller leaving for the reader the hard task to find the real drawbacks and limitations of the controller.

Robot Obstacle Avoidance Papers Using Experiments

Javier Minguez

This chapter describes methodological aspects related to the experimentation in robot obstacle avoidance research. Furthermore, it also describes some general guidelines to report these experiments in papers. Although there is a large difference between making real and simulated experiments, this chapter focusses on real experiments and on robots that perform in indoor scenarios (the world is assumed to be planar).

The ability to move a robot between given locations is one of the fundamental skills of the large majority of mobile robots. To address this task there are two different but complementary points of view: the motion planning and the obstacle avoidance. The objective of motion planning techniques is to compute a collision-free trajectory to the target configuration that complies with the vehicle constraints. They assume a perfect model of the robot and scenario. The advantage of these techniques is that they provide complete and global solutions of the problem.
Nevertheless, when the surroundings are unknown and unpredictable, these techniques fail.
A complementary way to face the motion problem is obstacle avoidance. The objective is to move a vehicle towards a target location free of collisions with the obstacles collected by the sensors during motion execution. The advantage of reactive obstacle avoidance is to compute motion by introducing the sensor information within the control loop, used to adapt the motion to any contingency incompatible with initial plans. The main cost of considering the reality of the world during execution is locality. In this instance, if global reasoning is required, a trap situation could occur. Despite this limitation, obstacle avoidance techniques are mandatory to deal with mobility problems in unknown and evolving surroundings.
Focussing our attention on reactive obstacle avoidance, the objective of these techniques is to use the sensor information to compute collision-free motion towards a given goal location. The general scheme is a perception - action process where the sensor collect the information of the scenario which is processed next to compute the collision-free motion. The motion command is executed by the robot and the process restarts with a new sensor measurement. The result of this process is that the robot is driven to the goal location while avoiding the collisions with the obstacles detected by the sensors.

In the followings we construct the general guidelines for a good experimental methodology and reporting in the obstacle avoidance papers. We will try to develop them by systematically answering the questions of the general document.

1 Is it an experimental paper?
The obstacle avoidance problem is in nature an online problem: in execution, sensors collect the information of the scenario to compute the next motion command. This is the reason why the large majority of research in obstacle avoidance has been developed using real robots with real sensors. Otherwise, realistic robot and sensor simulators need to be developed to extrapolate the research to real scenarios. Focussing in obstacle avoidance papers, experimental papers in this area are papers where the results, discussions and conclusions of the paper depend crucially on experimental work.
In obstacle avoidance there has been classically two types of papers: proof of concept papers and

new technique papers. The proof of concept papers usually proposed new techniques to address open problems or questions in the area that had not been addressed or successfully solved before. For example the potential field methods [12, 14] addressed the first sensor-based motions, the vector field histogram [5] was the first alternative to do obstacle avoidance with uncertain sensor like ultrasounds, the elastic bands [23] was the first technique combining planning and reaction schemas in a unified framework, the dynamic window approach was the first technique to address kinematics and dynamics to carry out motion at high speeds [10] (the [25] was a similar method developed alternatively), the nearness diagram navigation [18] was the first technique to address motion in troublesome scenarios, etc : : : . These papers usually described the new techniques and how the open problem was addressed. Usually the experiments were a proof that the method worked and was able to face and solve the particular problem conditions. In this context these papers can be considered experimental papers if the experimental results really validate the technique.

There is another set of papers that present new techniques or extensions of the previous techniques. In this case, we have the large set of different potential functions [28, 13, 9, 2, 4] and many others, many of the extensions of the vector field histogram [29, 30, 6, 31, 3], extensions of the elastic bands [8, 11], extensions of the dynamic window approach [7, 27] and of the nearness diagram [20, 17, 16] among others. Furthermore, there are hybrid methods that combine these techniques with tactical planners [30, 7, 19, 27, 22]. In general, these works describe a comparison with the previous methods usually in the context of the technique. For example if a new technique claims to carry out motion at high speeds then it is compared against the dynamic window approach (or at least it should be). In this type of papers the experimental methodology is crucial since it is the basis of the claim of the paper: improvement against other techniques in a given context. Thus, the experiments place an important role converting these papers in clearly experimental papers.

2 Are the system assumptions/hypotheses clear?

In robot obstacle avoidance there are several sources of hypotheses and assumptions: the type of robot, type of sensor, type of scenario and type of goal location. All these issues might be clarified in the paper. Some hints about them are:

1. Type of robot. There are at least three characteristics of the robot that affect the collision avoidance task. They have to be consequently reported in the experimental section whether they are taken into account or simply approximated:

- The robot shape determines the boundary of the robot (the main task is to avoid collision with the obstacles). From a technical point of view it makes a large difference if it is circular, described by segments, or if the geometry can be arbitrary or it is somehow approximated.
- The robot kinematics determines the possible nominal paths (how the robot can move). The possible paths are different if the robot is holonomic or with kinematic constraints such as differential-drive, car-like, etc ... . If the collision avoidance method does not take into account the kinematics but it works on a non holonomic robot then the method to carry out the approximation has to be reported.
- The robot dynamics determines the execution of the nominal paths (how the robot executes the motion) and almost all mobile robots have dynamic constraints. Dynamics is a complex problem in collision avoidance since it involves factors such as accelerations, maximum torque, inertia, slipping, etc. The way that the collision avoidance method takes into account the dynamics or the approximation used to convert the commands in admissible commands for the platform have be

reported.

2. Type of sensors. There is a large variety of sensors that have been used for collision avoidance. However, in general they can be grouped in two types: the sensors used to measure the scenario and the sensors that measure the location of the vehicle. On the one hand, the sensors that measure the obstacles can be of very different nature such as laser scans, ultrasounds or cameras for example. On the other hand, the sensors that measure the vehicle location can be GPS, odometry, etc. For both of them, the high level settings of the sensors have to be described at a level that allows the reader to understand their configuration and to measure the impact on the avoidance technique (e.g. when dealing with lasers, important settings are the reach, accuracy, uncertainty and sampling period).
Furthermore, if the sensor information has a pre-processing step before the collision avoidance, this processing technique might be cited, and the most relevant issues of the implementation that affect the collision avoidance step have to be outlined (e.g. the robustness, precision, convergence or computation time). Secondly, from this processing, the information extracted to be subsequently used by the collision avoidance has to be also reported (e.g. clouds of points, segments, planes, etc .

3. Type of scenario. For the experimentation in obstacle avoidance this point is one of the crucial issues to clearly understand the experimentation.
There are at least some issues to report:

- The apriori information of the scenario is the knowledge about its structure including the map or how the information is represented.
If there is no a priori information, then the scenario is unknown.
Otherwise, the type of information and its representation has to be reported (e.g. a segment-based a priori map of the laboratory or a grid based representation). Notice that we include here for example underlying assumptions about the structure (e.g we can assume a polygonal scenario even thought the map is initially unknown).

- The nature scenario is related with its dynamism. If all the obstacles are static then the scenario is static. Otherwise, the information about the dynamic nature of the moving obstacles has to reported (e.g. the percentage or number of dynamic obstacles and their maximum velocities or accelerations).

4. Type of goal location. This information is the final destination for the avoidance method, which can be static or dynamic. If it is dynamic, some information of the goal trajectory has to be summarized.

3 Are the performance criteria spelled out explicitly?
For a good experimental methodology in obstacle avoidance, the performance criteria is one of the first things to define and it might be done before designing the experiment. The criteria are very related with the main thesis of the paper and thus with the type of experimental validation:

1. If the paper is a proof of concept paper, then the experiments should validate the concept development. In this case, it should be clearly spelled out why this experiment is relevant for the research and why it validates the concept. In the case that the particular claim needs performance metrics, they have to be defined in advance and discussed. For example, if a new technique appears to solve the navigation among highly dynamic obstacles. Then, if it is validated with a real

experiment where the robot moves among dynamic obstacles, it should be clearly explain what are the conditions that arose in the experiment that validate the previous claim.
In these papers, it is usual to set up metrics to find the performance bounds. For example, one can prepare an experiment where the dynamics of the obstacles are increased to test the limits of the approach. Here, the performance metrics could be method success versus obstacle dynamics.
2. If the paper is an improvement of previous approaches, then the experiments should support a fair comparison. In this case it is important to set in advance the performance criteria in relation to the claim of improvement in a given context. Then, a fair and classic experiment has to be chosen to run the techniques (i.e. trying to avoid an experiment where the new technique it works and the others fail) and then the already defined metrics could be used to understand the method behavior.
For example, if a new technique appears to solve the navigation at a high speeds, then an experiment should validate the technique and at the same time support the comparison based on some predefined metrics. It is important to notice that real experiments are not exactly repeatable in obstacle avoidance, thus comparisons in experimental terms are difficult.

4 What is being measured and how?

The performance criteria must be measurable directly or indirectly by means of a combination of measurements. These measurements have to be properly described and they can be qualitative or quantitative. In obstacle avoidance, typical performance criteria are [21, 26, 15, 1]:
1. Mission success: number of successful missions.
2. Path length: distance travelled to accomplish the task.
3. Time: time taken to accomplish the task.
4. Collisions: number of collisions per mission, per distance and per time.
5. Obstacle clearance: minimum and mean distance to the obstacles.
6. Robustness in narrow spaces: number of narrow passages successfully traversed.

Notice that these are performance criteria. Some of them are directly measured, but others such as the robustness in narrow spaces for example need a definition of what is going to be measured. These measurements are obtained based on the result of the obstacle avoidance technique. However, to extrapolate and scale these results, they need to be reported as a function of the working conditions (that need to be measured also). In obstacle avoidance, these working conditions are usually measurements that describe the type of environment like [24]: density of the scenario, complexity, clutterness, confinement, structure, dynamism, etc.

5 Do the methods and measurements match the criteria?

In obstacle avoidance, the measurements methods can be objective if it is possible to store the characteristics of the sensor information used for the avoidance and the odometry of the robot for a posterior analysis. Then, the experiment can be repeated off line by applying at any time the technique to the sensor information stored. The off line simulation runs by moving the robot to the next location dictated by the odometry. By using this methodology, the measurements that match the performance criteria can be obtained off-line.
In the case that the measurements need some calculation (an algorithm), some clues about the calculation has to be given. This is because some measurements can be differently measured or calculated, and depending the method used, they could match the criteria or not. For example, as discussed in the previous section, a typical measurement in this area is the robustness in narrow spaces, which can be differently measured and calculated.

6 Is there enough information to reproduce the work?

In the context of obstacle avoidance, to replicate the experiments, the hypotheses and assumptions related with the type of robot, sensor, scenario and final destination have to be clarified (subsection 2.2), and also the performance criteria and the measurement methods (subsection 2.3, subsection 2.4, and subsection 2.5).

7 Do the results obtained give a fair and realistic picture of the system being studied?

This question is very related to the type of experimentation designed to validate the technique, which was addressed in subsections 2.1 and 2.3.

8 Are the drawn conclusions precise and valid?

The experimental conclusions must be consistent with the experimental questions the paper poses, the criteria employed and the results obtained. This has to be explained taken the hypotheses and assumptions (subsection 2.2), and also the performance criteria and the measurement methods (subsection 2.3, subsection 2.4, and subsection 2.5). The system limits can be presented as a function of the hypotheses and assumptions (subsection 2.2). To conclude, note that statistical analysis of the data is difficult in this area since the experiments have to be conducted online and this creates the difficulty of replicability and time.

References

[1] H. , M. , R. , and F. . Understanding and measuring powered wheelchair mobility and manoeuvrability. part i. reach in confined spaces.

[2] K. Azarm and G. Schmidt. Integrated mobile robot motion planning and execution in changing indoor environments. In IEEE/RSJ International Conference on Intelligent Robots and Systems, pages 298–305, Munchen, Germany, 1994.

[3] D. Bell, S. Levine, Y. Koren, L. Jaros, and J. Borenstein. Design criteria for obstacle avoidance in a shared-control system. In RESNA Conference, Nashville, 1994.

[4] J. Borenstein and Y. Koren. Real-Time Obstacle Avoidance for Fast Mobile Robots. IEEE Transactions on Systems, Man and Cybernetics, 19(5):1179–1187, 1989.

[5] J. Borenstein and Y. Koren. The Vector Field Histogram–Fast Obstacle Avoidance for Mobile Robots. IEEE Transactions on Robotics and Automation, 7:278–288, 1991.

[6] J. Borenstein and U. Raschke. Real-time obstacle avoidance for non-point mobile robots. In Fourth World Conference on Robotics Research, Michigan, USA, 1992.

[7] O. Brock and O. Khatib. High-Speed Navigation Using the Global Dynamic Window Approach. In IEEE Int. Conf. on Robotics and Automation, pages 341–346, Detroit, MI, 1999.

[8] O. Brock and O. Khatib. Real-Time Replanning in High-Dimensional Configuration Spaces using Sets of Homotopic Paths. In IEEE Int. Conf. on Robotics and Automation, pages 550–555, San Francisco, USA, 2000.

[9] L. Chenqing, M. Ang, H. Krishman, and L. Yong. Virtual Obstacle Concept for Local-minimum-recovery in Potential-field Based Navigation. In IEEE Int. Conf. on Robotics and Automation, pages 983–989, San Francisco, USA, 2000.
[10] D. Fox, W. Burgard, and S. Thrun. The Dynamic Window Approach to Collision Avoidance. IEEE Robotics and Automation Magazine, 4(1), 1997.
[11] M. Khatib, H. Jaouni, R. Chatila, and J. Laumond. Dynamic Path Modification for Car-Like Nonholonomic Mobile Robots. In IEEE Int. Conf. on Robotics and Automation, pages 2920–2925, Alburquerque, Mexico, 1997.
[12] O. Khatib. Real-Time Obstacle Avoidance for Manipulators and Mobile Robots. Int. Journal of Robotics Research, 5:90–98, 1986.
[13] Y. Koren and J. Borenstein. Potential Field Methods and Their Inherent Limitations for Mobile Robot Navigation. In IEEE Int. Conf. on Robotics and Automation, volume 2, pages 1398–1404, Sacramento, CA, 1991.
[14] B. H. Krogh and C. E. Thorpe. Integrated Path Planning and Dynamic Steering control for Autonomous Vehicles. In IEEE Int. Conf. on Robotics and Automation, pages 1664–1669, San Francisco, USA, 1986.
[15] B. Kuipers. Building and evaluating an intelligent wheelchair. Internal Report, 2006.
[16] C. Marques. Multi-sensor navigation for soccer robots. Master's thesis, Instituto Superior Tecnico, Portugal, 2001.
[17] J. Minguez. The Obstacle Restriction Method (ORM): Obstacle Avoidance in Difficult Scenarios. In IEEE Int. Conf. on Intelligent Robot and Systems, Edmonton, Canada, 2005.
[18] J. Minguez and L. Montano. Nearness Diagram (ND) Navigation: Collision Avoidance in Troublesome Scenarios. IEEE Transactions on Robotics and Automation, 20(1):45–59, 2004.
[19] J. Minguez, L. Montano, N. Simeon, and R. Alami. Global Nearness Diagram Navigation (GND). In IEEE International Conf. on Robotics and Automation, pages 33–39, Seoul, Korea, 2001.
[20] J. Minguez, J. Osuna, and L. Montano. A Divide and Conquer Strategy to Achieve Reactive Collision Avoidance in Troublesome Scenarios. In IEEE International Conference on Robotics and Automation, Minessota, USA, 2004.
[21] L. Montesano, J. Minguez, M. Diaz, and S. Bhaskar. Towards an intelligent wheelchair for cerebral palsy patients. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2008. Under review.
[22] R. Philipsen and R. Siegwart. Smooth and efficient obstacle avoidance for a tour guide robot. In IEEE Int. Conf. on Robotics and Automation, Taipei, Taiwan, 2003.
[23] S. Quinlan and O. Khatib. Elastic Bands: Connecting Path Planning and Control. In IEEE Int. Conf. on Robotics and Automation, volume 2, pages 802–807, Atlanta, USA, 1993.

[24] I. Rano and J. Minguez. Steps towards the automatic evaluation of robot obstacle avoidance algorithms. In IEEE International Conference on Intelligent Robot and Systems, Beijin, China, 2006.
[25] R. Simmons. The Curvature-Velocity Method for Local Obstacle Avoidance.

In IEEE Int. Conf. on Robotics and Automation, pages 3375–3382,
Minneapolis, USA, 1996.
[26] R. Simpson, D. Poirot, and M. Baxter. The hephaestus smart wheelchair
system. 10(2):118–122, 2002.
[27] C. Stachniss and W. Burgard. An Integrated Approach to Goal-directed
Obstacle Avoidance under Dynamic Constraints for Dynamic Environments.
In IEEE-RSJ Int. Conf. on Intelligent Robots and Systems, pages
508–513, Switzerland, 2002.
[28] R. B. Tilove. Local Obstacle Avoidance for Mobile Robots Based on the
Method of Artificial Potentials. In IEEE Int. Conf. on Robotics and Automation,
volume 2, pages 566–571, Cincinatti, OH, 1990.
[29] I. Ulrich and J. Borenstein. VFH+: Reliable Obstacle Avoidance for Fast
Mobile Robots. In IEEE Int. Conf. on Robotics and Automation, pages
1572–1577, 1998.
[30] I. Ulrich and J. Borenstein. VFH*: Local Obstacle Avoidance with Look-
Ahead Verification. In IEEE Int. Conf. on Robotics and Automation, pages
2505–2511, San Francisco, USA, 2000.
[31] H. Yang, J. Borenstein, and D. Wehe. Double-vfh: Reliable obstalce avoidance
for large, non-point, omni-directional mobile robots. In ANS Conference
on Robotics and Remote Systems, Pitsburg, PA, 1999.

Grasping

Antonio Morales

Robot grasping and manipulation deals with control of robots that interact with independent parts (objects) with the aim of manipulate them. In this context, manipulation of parts means to apply forces on them in order to produce a change in their pose, transport them, and use them as tools. No fixed contact is established between the robots and the objects. Mainly, the robot systems involved in grasping manipulation consist of a robot arm, plus a robot gripper which actually makes contact with the object. These grippers are often called robot hands, if their structure resembles minimally that of a human hand. Gripper generally consists of two or more articulated chains called fingers.
Papers on robot grasping can be roughly divided in two broad realms: algorithms and applications. The algorithmic realm relates with the computation of the characteristics of the interactions between the robot and the parts. This includes: gripper control (actuating the hand to produce certain forces, displacements on the contact points), grasp planning (given an object, compute the contact locations on its surface to produce a desired effect), grasp analysis (given a gripper making contact with an object, which are the properties of that configuration).
Application papers describe real robot systems that are designed to perform a given task through manipulations. This kind of papers not only include practical applications but also those describing the design of novel mechanism for grippers.

1.Is it an experimental paper?

In the grasping community experimental papers are rare, in the sense the empirical sciences understand experimental: “using results of experiments to discover unknown properties or to contrast hypothesis about the studied system”. Many papers are experimental in engineering sense, use experiments to validate the design principles followed in the development of the robot or algorithm. Almost all papers show a small set of experiments to only illustrate their approach.
In my opinion exhaustive experimentation (varying parameters, objects, and conditions) should be mandatory to validate any algorithmic or practical approach, especially for journal papers. This could not be so strict for conference papers, where only a few experiments that indicate the feasibility of the approach would be enough.
For algorithmic papers validation using simulations would be acceptable, provided that simulation is close to reality. Standard tools and benchmarks should be employed to allow comparisons. In any case real experiments should be encourages always when possible.
For application papers exhaustive validation with real system should be mandatory. Simulations should be avoided, only accepted on very specific cases, and mainly for conference papers.

2.Are the system assumptions/hypotheses clear?

Assumptions should be clearly stated. Are hand kinematics taken into account? Which contact model is considered? Is friction neglected? 2D, 3D, or polyhedral objects? Are the object models known in advance?. In general, any used information which is assumed to be known (models, contact position, contact force, etc.) should be clearly stated.

3.Are the performance criteria spelled out explicitly?

In any case one or more performance criteria must be stated clearly into the paper. It is encouraged

that authors use criteria that have been established as common criteria into the community, and must reference it clearly. In case of new criteria the authors must clearly described them in order for others to reference them.
One typical performance criterion for grasping algorithms is the grasp quality. Grasp quality is normally introduced as a function depending on several factors, as grasp stability, hand configuration, computation time, etc.
In the case of practical applications experimental performance metrics related with the intended task

4.What is being measured and how?

There must be a way to measure the defined criteria. Data to allow this must be obtained from the experiments. In the case or real experiments sensor data (force, etc) and external measurement tools could be used. In the case of simulations this should be easier. In any case where these data is obtained, and how it used to compute the criteria metrics must be clearly described.
For example, grasp quality is usually measured from pre-computed grasps on given object models, as a function depending on contact position, contact forces, relation between the object center of mass and the contacts, hand configuration, etc.

5.Do the methods and measurements match the criteria?

A justification that the data gathered is enough to represent the criteria must be given.

6. Is there enough information to reproduce the work?

In practical applications a detailed description of the hardware used is necessary. This includes computer specifications (processor, memory), sensors used (cameras, force/sensors, etc) and robot parts specifications (DOF, sensors, controllers, etc.) should be definitively described. It is encouraged to use public tools.
The experimental conditions must be described too: the kind of objects used, illumination conditions when using cameras, description of the workspace of the robot, obstacles present in it.
When doing exhaustive validation, it is necessary to describe the experimental protocol: what set of objects/parts are tested, which parameters are variable and their value ranges, how many trials are made, when and how the performance metrics are done. There must be an explanation of why these parameters are chosen and what feature of the system is being studied.
In case of using simulations, the use of standard tools is encourages. Also, for practical applications the use of common hardware will be also welcomed.
For papers submitted to journal papers it would a good practice when possible to allow public access to data and models of the experimental setup, for comparison with future papers

7.Do the results obtained give a fair and realistic picture of the system being studied?

Real experiments are encouraged, even mandatory. They should be a good example for validating the theoretical approach, and, at the same time, they should represent an interesting common problem.
A special care must be place in isolating relevant parameters from external conditions in the execution. A proper statistical analysis of the experimental data must be done.

8.Are the drawn conclusions precise and valid?

The purpose of the conclusion is to evaluate how the experimental data obtained supports the approach taken in the paper. It is not only to see the successful cases but also to study the limits of a successful approach, and what can be more relevant, to determine the failure cases and to discuss how the approach could be improved to overcome these limitations.

Finally, I would like to mention an interesting book on the application of an empirical approach to analyze computing algorithms: "Empirical Methods for Artificial Intelligence", Paul R. Cohen, MIT Press, 1995. This manual introduces statistical concepts and procedures to study experimentally the performance of software programs, in this case mostly A.I. approaches. It could be of interest to prepare experiments and to analyze the data results coming from them.

Visual Servoing

Enric Cervera and Mario Prats

Visual servoing (VS) deals with the control of robots (mobile or manipulators) through vision in a feedback loop. It is a multi-disciplinary research area spanning computer vision, robotics, kinematics, dynamics, control and real-time systems. Papers commonly present a theoretical part (control law, visual features, models) supported by simulated or real experiments. Formal proofs typically cannot be derived, thus experimental evidence must be sound.

1.Is it an experimental paper?

Nearly every VS paper contains some experiments, which illustrate the presented approach. Pure theoretical papers are rare, since proofs are limited to simple problems. Short papers (e.g. conferences) may include a simple setup with a few variations on parameters. Longer journal papers should provide wider evidence. Experiments with real robots are encouraged (should be mandatory?).
Simulation experiments may also be considered. Simulations provide a structured environment where it is possible to proof the theoretical development under perfectly controlled conditions. Real experimentation can introduce uncontrolled variables that could affect the results. Depending on what the authors want to proof (new interaction matrix, convergence, etc.), simulation experiments could be more suitable. However, experiments with real robots should be included when possible.

2.Are the system assumptions/hypotheses clear?

Typical assumptions are the visual features, the knowledge of the scene 3D model, the kinematics and dynamics model of the robot.
Assumptions related to image processing (homogeneous background, lighting conditions, etc.) should be specified, as well as robustness to outliers in feature detection and, in general, all aspects inherent to real life experimentation.

3.Are the performance criteria spelled out explicitly?

VS papers address the convergence of the system to a predefined goal. Related criteria include the time of convergence, the trajectories of the visual features in the image plane, the 3D trajectory of the robot, computation time, positioning error after convergence.
Stability and robustness are also important issues, measured with respect to image noise, the errors in the models (object, camera, robot), and the control parameters.
Local minima should be considered; though theoretical proof is not feasible, experiments should investigate “hard” configurations, not trivial ones.

4.What is being measured and how?

Visual features can be directly obtained from the image framegrabber. For manipulators, the 3D trajectory is measured by the joint angles, and computed by the direct kinematic model. Robot – camera calibration (or its absence) should be specified.
Image noise in real experiments should at least include the variability of the image features.

5.Do the methods and measurements match the criteria?

The use of public wide-access vision and/or mathematical libraries and packages (ViSP, VXL, OpenCV) is encouraged and should be clearly stated and referenced in the paper.
Ground truth for robot positioning must be provided: in case of commercial manipulators, the libraries provided by manufacturers are the typical option. For mobile robots, positioning methods should be described, with special attention to the uncertainty in the measurements.

6. Is there enough information to reproduce the work?

At least the simulated experiments should be properly described to allow replication. Use of public simulator tools is encouraged (ViSP, Matlab VS Toolbox, JaViSS) and should be clearly stated and referenced in the paper. All the model and control parameters should be clearly stated. For real robot experiments, the platform should be thoroughly described.
Technical specifications of the camera should be included (model, frame rate, resolution, etc.). Computer specifications (at least, processor and amount of memory) are also needed in order to allow for comparison. The environment (either real or simulation) should be described in detail: in-hand vs. external camera, known relations (as camera external calibration), etc.

7.Do the results obtained give a fair and realistic picture of the system being studied?

Experiments should not be limited to a single task, though this may be difficult to achieve in a short paper, since typically a number of parameters can be varied. Comparison with experiments used by other previously published methods should be encouraged.
Either simulation or real experimentation must be carefully chosen depending on the factors being studied.

8.Are the drawn conclusions precise and valid?

Beware of claims like “the experiment demonstrates the validity/robustness/effectiveness of our approach”. The contribution of the work should be clearly specified: better convergence speed, robustness against certain parameters, avoidance of typical VS problems (image features going out of the field of view, robot reaching joint limits...). The experimental work should be representative of such claimed contributions.

Autonomy/Cognitive tasks

Fabio Bonsignorio

The focus is to define quantitative metrics to be able to compare such things as: the level of autonomy, human-robot interaction, collaboration.
To provide methodologies to evaluate components of intelligent systems: sensing and perception, knowledge representation, world models, ontologies, planning and control, learning and adapting, reasoning.
Other topics are infrastructural support for performance evaluation, application specific performance measures.
In this context intelligence is defined as “the ability to act appropriately in an uncertain environment, where appropriate action is that which increases the probability of success, and success is the achievement of behavioral goals” (J. Albus, “Outline for a Theory of Intelligence”, IEEE Trans. on Systems, Man, an Cybernetics, Vol. 21, No. 3, May/June 1991).

1.Is it an experimental paper?

Although there are many theoretical papers in particular estimating the computational cost of algorithms in AI ('search' in particular) and sometimes the convergence, it is quite common to use experiments to support theoretical/principle claims.
Typically conceptually different approaches like 'embodied ai' versus 'symbolic' are compared. Since a framework quantitative theory of embodied cognition is missing the comparison should be based on statistically relevant sets of experiments.

2.Are the system assumptions/hypotheses clear?

It is at least necessary to define clearly, and possibly quantitatively, the tasks, and task complexity, and the environments where these tasks are performed. Information metrics (Shannon Entropy etc) should help the definition of proper complexity measures. This is an open research field.
In this case the human/animal neurophysiology approach might be followed.

3.Are the performance criteria spelled out explicitly?

The learning time versus effectiveness in performing a task should be considered.
Stability and robustness are important issues, measured with respect to task/environment complexities.
Importance of levels of cognition.

4.What is being measured and how?

Due the wide range of possible applications this need to be well clarified. A (maybe far) benchmark should be animal cognitive tasks (for example mice mazes).
Importance of levels of cognition.

5.Do the methods and measurements match the criteria?

There are examples publicly available in planning. For sensory motor coordination information metrics might help.

6. Is there enough information to reproduce the work?

These means not only specifying the methods, but also define the set of tasks, the set of environments and the way to switch from one to another, including the learning procedures (for example learning time versus task execution time).

7.Do the results obtained give a fair and realistic picture of the system being studied?

In this case it is very important to variate the tasks and the environments in a comparable way.
The repeatability of learning procedures, too, must be considered.

8.Are the drawn conclusions precise and valid?

This must be stated paying proper attention to: task sets, variation procedures, environment sets, variation procedures, learning time/ procedures related to the above criteria.

1. James S. Albus, "Metrics and Performance Measures for IntelligentUnmanned Ground Vehicles". In Proceeding of the performance Metrics for Intelligent System Workshop, 2002.
2. Angel P. del Pôbil, "Why do We Need Benchmarks in Robotics Research?", International Conference on Intelligent Robot and Systems, Beijing, China, 2006.
3. http://www.isd.mel.nist.gov/PerMIS_2007/index.htm
4. http://www.robot-standards.eu/
5. http://www.robot.uji.es/benchmarks/index.html
6. http://www.rawseeds.org/
7. http://www.euron.org/activities/SIGs.html
8. http://www.isd.mel.nist.gov/projects/autonomy_levels/
9. http://www.robocup.org
10. http://www.darpa.mil/grandchallenge/
11. Jacob W. Crandall and Michael A. Goodrich, "Measuring the Intelligence of a Robot and its Interface". In Proceeding of the performance Metrics for Intelligent System Workshop, 2003.
12. J. Minguez, J. Osuna, and L. Montano. A "Divide and Conquer"Strategy based on Situations to achieve Reactive Collision Avoidance in Troublesome Scenarios. In ICRA, New Orleans, USA, 2004.
13. L. Olsson, C.L Nehaiv, D. Polani, "Information Trade-Offs and the Evolution of Sensory Layouts", In Proc. Artificial Life IX, 2004.
M.Lungarella, , O. Sporns, "Mapping Information Flow in Sensorimotor Networks", PLOS Computational Biology, 2, 10, pp. 1301-1312 , 2006.
14, A.Lampe, R.Chatila, "Performance measures for the evaluation of mobile robot autonomy" , IEEE International Conference on Robotics and Automation (ICRA'06), Orlando (USA), 2006.
15. J.Hallam, G.Hayes, "Benchmarks for mobile robotics?" In Towards Intelligent Mobile Robots: scientific methods in mobile robotics, Manchester University, School of Computer Science, 1997.
16. https://www.ctnbestpractices.org/

17. http://ctep.cancer.gov/handbook/hndbk_7.html
18. http://www.cancer.gov/
19. Pfeffer, J. (1993), "Barriers to the advance of organizational science: Paradigm development as a dependent variable". Academy of Management Review, 18: 599–620, 1993.
20. Popper, K., The logic of scientific discovery.,Hutchison, London, 1959
21. Kuhn, T. The structure of scientific revolutions (2nd ed.), University of Chicago Press, Chicago, 1970.
22. I.Lakatos, "Criticism and the Methodology of Scientific Research Programmes",in Proceedings of the Aristotelian Society, vol. 69 (1968-69), pp. 149-186.
23. P.K.Feyerabend, Against Method, Verso, London, 1975.