Robot Competitions – Ideal Benchmarks for Robotics Research

Sven Behnke
Humanoid Robots Group, Computer Science Institute
University of Freiburg, Georges-Köhler-Allee 52, 79110 Freiburg, Germany
e-mail: behnke@informatik.uni-freiburg.de

Abstract. In this paper, I argue for the use of robotic competitions as benchmarks for robotics research. By providing a common task to be solved at a specific place and a specific time, competitions avoid some of the difficulties arising when evaluating robotics research in the own lab. Competitions also bring together multiple research groups working on the same problem. This fosters the exchange of ideas.
I review two of the most popular robotics competitions, RoboCup and the DARPA Grand Challenge, and discuss some issues arising when designing robotics competitions.

1. Introduction

Benchmarking robotics research is inherently difficult. Typically, results are reported only for a specific robotic system and a self-chosen set of tasks. The tasks are solved in the lab where the robot was developed. This makes it impossible to compare the results with other systems, developed in different labs and tested for different specific tasks. The commonly used "proof by video" technique has the same difficulties as the "proof by example" in other settings. That a robotic system works once in a video does not mean that it works always or that it works under slightly less controlled conditions.

One possible approach to overcome these shortcomings is to participate at robot competitions. Robot competitions bring together researchers, students, and enthusiasts in the pursuit of a technological challenge. Popular competitions include MicroMouse [1], where wheeled robots have to solve a maze, Robolympics [2], where robots compete in many disciplines, Robo-one [3], where remotely controlled humanoid robots engage in martial arts, and the AAAI Robot Competition [4], where robots have to solve different tasks in a conference environment. Among the most popular robot competitions are robotic soccer championships, like RoboCup [5] and FIRA [6], and competitions for unmanned ground and aerial vehicles, like the DARPA Grand Challenge [7], the European Land-Robot Trial (ELROB) [8], and the International Aerial Robotics Competition (IARC) [9]. Pobil compiled a survey of such competitions and other benchmarks for robotics [10]. Rainwater maintains a list of robot competitions [11].

Such robot competitions provide a standardized test bed for different robotic systems. All participating teams are forced to operate their robots outside their lab in a different environment at a scheduled time. This makes it possible to directly compare the different approaches for robot construction and control. In the following, I will review two of the most popular robotic competitions: RoboCup and the DARPA Grand Challenge.

2. RoboCup

RoboCup is an international joint project to promote AI, robotics, and related fields. The RoboCup Federation organizes since 1997 international robotic soccer competitions. The long-term goal of the RoboCup Federation is to develop by the year 2050 a team of humanoid soccer robots that wins against the FIFA world champion [12]. The soccer game was selected for the competitions, because, as opposed to chess, multiple players of one team must cooperate in a dynamic environment. Sensory signals must be interpreted in real-time and must be transformed into appropriate actions. The soccer competitions do not test isolated components, but two systems compete with each other. The number of goals scored is an objective performance measure that allows comparing systems that implement a large variety of approaches to perception, behavior control, and robot construction. The presence of opponent teams, which continuously improve their system, makes the problem harder every year.

The RoboCup championships grew continuously over the years. In the last RoboCup, which took place in June 2006 in Bremen, Germany, 440 teams from 36 countries competed. The total number of participants was more than 2.600. In addition to the soccer competitions, since 2001, competitions for the search of victims of natural disasters and the coordination of rescue forces are held (RoboCupRescue). In 2006, for the first time, competitions for robots in a living-room environment took place in the RoboCup@home league. Furthermore, there are competitions for young researchers (RoboCupJunior).

The soccer competitions at RoboCup are held in five leagues. Since the beginning, there is a league for simulated agents, a league for small wheeled robots which are observed by cameras above the field (SmallSize), and a league for larger wheeled robots where external sensors are not permitted (MiddleSize). A league for the Sony Aibo dogs was added in 1999 (Four-legged) and a league for humanoid robots was established in 2002.

Different research issues are addressed in the different leagues. In the simulation league, team play and learning are most advanced. In the wheeled robot leagues, the robot construction (omnidirectional drives, ball manipulation devices), the perception of the situation on the field (omnidirectional vision systems, distance sensors), and the implementation of basic soccer skills (approaching, controlling, dribbling, and passing the ball) are still in the center of the activities. Because the robot hardware is fixed in the Four-legged League, the participating teams focus on perception and behavior control. Here, the control of the 18 degrees of freedom (DOF) poses considerable challenges.

As the performance of the robots increases, the competition rules are made more demanding by decreasing the deviations from the FIFA laws. This permanently increases the complexity of the problem. It can also be observed that solutions like team play, which have been developed in leagues abstracting from real-world problems, are adopted in hardware leagues, as the basic problems of robot construction, perception, locomotion, and ball manipulation are solved better.

Fig. 1. Some of the robots, which participated in the RoboCup 2006 Humanoid League competitions.

The Humanoid League is the most challenging RoboCupSoccer league. Its competition rules [13] require robots to have a human-like body plan. They consist of a trunk, two legs, two arms, and a head. The only allowed modes of locomotion are bipedal walking and running. The robots must be fully autonomous. No external power, computing power or remote control is allowed. After less demanding competitions, like walking around a pole and penalty kicks, at RoboCup 2005, the first 2 vs. 2 soccer games were played in the KidSize class (30-60cm robot height). Fig. 1 shows some of the robots which participated in the RoboCup 2006 Humanoid League competitions.

Very different approaches for robot construction, perception, and behavior control were used. While some teams constructed their robots starting from commercially available kits, like Robotis Bioloid or Kondo KHR-1, many robots were designed from scratch by the teams. The largest and most expensive robot Arabot (Pal Technology, 30DOF, 140cm, 36kg) won the Footrace in the TeenSize class (65-130cm). Team NimbRo (Freiburg, Germany) constructed 20DOF, 60cm, 2.9kg robots, which won the KidSize Penalty Kick and came in second in the overall Best Humanoid ranking, the same result as in 2005. Winner of the Technical Challenge, the 2 vs. 2 soccer games, and the TeenSize Penalty Kick was Team Osaka, which used self-constructed robots with omnidirectional vision systems. Team Osaka was Best Humanoid for the third time in a row. This result shows that despite the variance caused by the randomness of soccer games, the RoboCup competitions do provide an objective performance measure.

3. DARPA Grand Challenge

The DARPA Grand Challenge benchmarks performance of autonomous ground vehicles. It is organized by the U.S. government to foster research and development in the area of autonomous driving. The first two competitions took place in the Mojave Desert. The course included gravel roads, paths, switchbacks, open desert areas and dry lakebeds, mountain passes, and some paved roadways. The course was outlined by a GPS corridor, which consisted of several thousand waypoints, accompanied by allowable path width and speed limits.

While following the GPS corridor, the vehicles had to recognize drivable surfaces by themselves and to make steering decisions in order to stay on the road and to avoid obstacles. Possible obstacles included other vehicles, fences, utility poles, stones, trees and bushes, and ditches. As skilled drivers with standard SUVs would have no difficulties driving the curse, the challenge was information processing. Robustly perceiving the state of the environment and the vehicle, making driving decisions appropriate to the situation, and acting timely were key factors for success [14].

The vehicles had to be completely autonomous: no remote control capabilities were allowed. They could carry any combination of onboard sensors, both for sensing the position of the vehicle and the surrounding environment. Teams could also use any available, non-classified map and terrain database. The only external signals allowed were the pause and emergency stop remote control signals for the organizers and publicly available navigation aids, such as GPS signals and commercial differential correction services available to all teams.

The Grand Challenge events were divided into two segments: the qualification and the race. For qualification, the teams had to demonstrate the safety and reliability of their vehicles, including the emergency stop systems. They had to show autonomous motion capabilities on a test course which included narrow passages, obstacles, and a tunnel.

The first DARPA Grand Challenge took place on March 13th, 2004, but none of the participating vehicles came very far. On October 8th, 2005, 23 finalists started the second race, which was 213km long. This time, the participants were better prepared. The teams of Stanford University and Carnegie Mellon University (CMU), for example, drove prior to the race hundreds of km through similar terrain to calibrate and test their systems. Consequently, five autonomous vehicles finished the 2005 course. The $2 million price went to the fastest of them: Stanley [15] of Stanford Racing Team, which is shown in Fig. 2. It drove at an average speed of 30.7 km/h, with a top speed of 61km/h. The second and third fastest vehicles belonged to the Red Team of CMU.

Fig. 2. Stanley of Stanford Racing Team, winner of the 2006 DARPA Grand Challenge.

Major components of Stanley's software were based on machine learning, probabilistic reasoning, and real-time control. Probabilistic methods were necessary for robust perception in the presence of substantial measurement noise in the various sensors. Machine learning was applied prior to the race to tune system parameters and during the race to adapt to the terrain. Two important technical solutions developed for Stanley were adaptive road extraction from live camera images and speed adaptation. The idea for road extraction was that the road outside the range of the vehicles laser-scanners is likely to look similar to the surface classified as drivable in the laser range. This allowed planning the path further ahead than would have been possible with the laser range information alone. Speed adaptation was based on the supplied speed limits and imitation of human driver reactions to road conditions, like roughness, slope, and curvature.

While the vehicles in the 2005 challenge had to avoid static obstacles, they did not encounter moving obstacles, like other cars. The third competition, the DARPA Urban Challenge, is scheduled for November 3, 2007. It will take place in an urban environment with other traffic. The autonomous vehicles must not only find their way through a city, but they also must obey traffic laws.

4. Discussion

Robot competitions, as described above, proved to be a driving force of technological development. They allow for direct comparison of different approaches to solving a task. Participating teams are forced to leave their lab and to operate their robots at the competition site at the scheduled time. The competitive aspect unleashes huge energies and the competitions foster the exchange of ideas.

As the performance level of the robots rises, the competition rules must be developed to keep the challenges challenging. For competitions to be successful, it is important to ask for skills meaningful for many research groups. The skills to demonstrate must be challenging, but not too hard, as a hopeless challenge will not attract participants. Thus, there is a need for intensive exchange between organizers of the competition and the participants.

Naturally, robot competitions evaluate entire systems. When observing a difference of performance, it is frequently unclear, to which component of the systems the difference should be attributed to. A robot might not perform well for many reasons. It could be, for example, that the perception system is disturbed or that the behavior control software made a wrong decision or simply that an actuator is not working as designed. Hence, it is desirable to include in the competitions specific tests for subsystems.

Another issue is the availability of technical information about the winning systems. In order to advance the entire field, the teams should be required to release a detailed technical description after the competition. In some RoboCup leagues, where all teams share the same simulated or physical agents, it is even feasible to build on the software of the winning teams. To highlight technical advances, the competitions should be accompanied by technical conferences, where the underlying methods are discussed.

In conclusion, it can be stated that robot competitions are popular for good reasons. If designed well, they can be drivers for their field and ideal benchmarks for robotics research.

References

[1] P. Harrison: Micromouse Information Centre. http://micromouse.cannock.ac.uk

[2] ROBOlympics: http://www.robolympics.net

[3] ROBO-ONE: Biped Robot Entertainment. http://www.robo-one.com

[4] S. Tejada and P. Rybski: AAAI 2005 Robot Competition and Exhibition.http://palantir.swarthmore.edu/aaai05

[5] The RoboCup Federation: RoboCup Official Site. http://www.robocup.org

[6] The Federation of International Robot-soccer Association (FIRA). http://www.fira.net

[7] R. Kurjanowicz: DARPA Grand Challenge. http://www.darpa.mil/grandchallenge

[8] Frank E. Schneider: European Land-Robot Trial ELROB. http://www.elrob.org/

[9] Robert Michelson: AUVS International Aerial Robotics Competition. http://avdil.gtri.gatech.edu/AUVS/IARCLaunchPoint.html

[10] Angel P. del Pobil: Research Benchmarks V1. EURON report, 2005.

[11] R. Steven Rainwater: Robot Competition FAQ. http://robots.net/rcfaq.html

[12] Hiroaki Kitano and Minoru Asada: The RoboCup Humanoid Challenge as the millennium challenge for advanced robotics. *Advanced Robotics*, 13(8):723-737, 2000.

[13] S. Behnke: RoboCupSoccer Humanoid League Rules and Setup for the 2006 competition in Bremen, Germany. http://www.humanoidsoccer.org/rules.html

[14] M. Montemerlo, S. Thrun, H. Dahlkamp, D. Stavens, and S. Strohband: Winning the DARPA Grand Challenge with an AI robot. In *Proceedings of the AAAI National Conference on Artificial Intelligence*, Boston, MA, 2006.

[15] S. Thrun, M. Montemerlo, H. Dahlkamp, D. Stavens, A. Aron, J. Diebel, P. Fong, J. Gale, M. Halpenny, G. Hoffmann, K. Lau, C. Oakley, M. Palatucci, V. Pratt, P. Stang, S. Strohband, C. Dupont, L.-E. Jendrossek, C. Koelen, C. Markey, C. Rummel, J. van Niekerk, E. Jensen, P. Alessandrini, G. Bradski, B. Davies, S. Ettinger, A. Kaehler, Nefian, and P. Mahoney: Stanley, the robot that won the DARPA Grand Challenge. To appear in *Journal of field Robotics*, 2006.