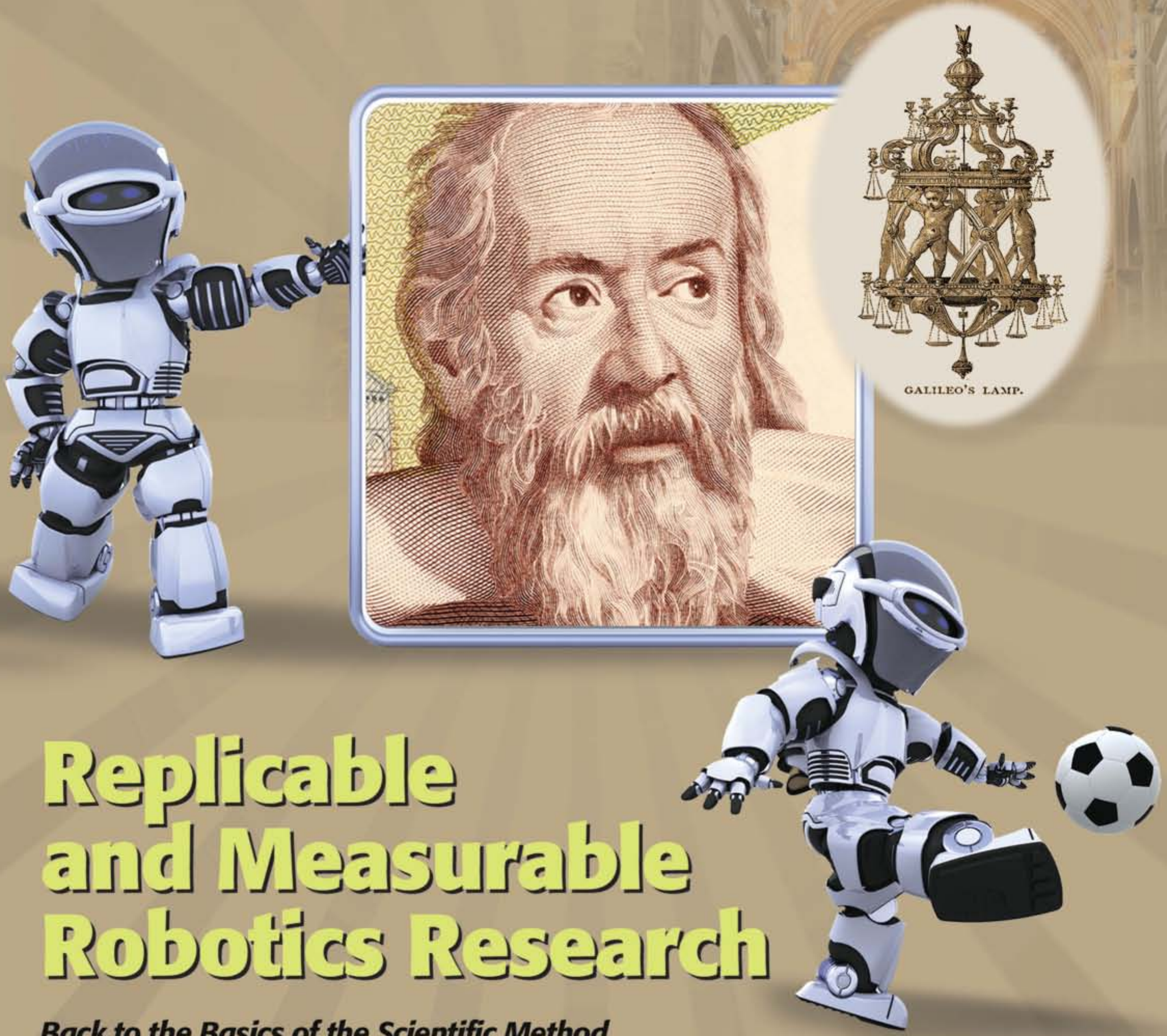
IEEE ROBOTICS & AUTOMATION MAGAZINE

Vol. 22, No. 3 September 2015
ISSN 1070-9932
http://www.ieee-ras.org/publications/ram



Replicable and Measurable Robotics Research

Back to the Basics of the Scientific Method

IEEE

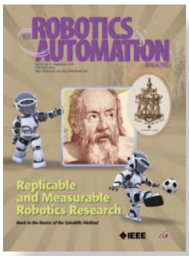

adept mobilerobots

IEEE ROBOTICS & AUTOMATION MAGAZINE

Vol. 22, No. 3 SEPTEMBER 2015
ISSN 1070-9932
http://www.ieee-ras.org/publications/ram

FEATURES

©ISTOCKPHOTO.COM/KIRSTYPARGETER



ON THE COVER

Performance evaluation, benchmarking, competitions are key aspects for fostering reproducibility of robotics research, fully in line with the very basic principles of the scientific method as introduced by Galileo Galilei back in the17th century.

PHOTOCREDITS: GALILEO, ROBOTS, & BACKGROUND—IMAGES LICENSED BY INGRAM PUBLISHING, LAMP—©ISTOCKPHOTO.COM/JONNY JIM, CATHEDRAL—©ISTOCKPHOTO.COM/MAUI01.

If you like an article, click this icon to record your opinion. This capability is available for online Web browsers and offline PDF reading on a connected device.

(continued on page 2)

Digital Object Identifier 10.1109/MRA.2015.2452052

FEATURES *(continued)*

COLUMNS & DEPARTMENTS

IEEE ROBOTICS & AUTOMATION MAGAZINE
A Publication of the IEEE ROBOTICS AND AUTOMATION SOCIETY
Vol. 22, No. 2 June 2015 ISSN 1070-9932 http://www.ieee-ras.org/publications/ram

EDITORIAL BOARD
Editor-in-Chief
Eugenio Guglielmelli
(ram-eic@ieee.org)
Campus Bio-Medico University, Roma (Italy)

Associate Editors

Antonio Franchi
Centre National de la Recherche Scientifique, CNRS (France)

Yi Guo
Stevens Institute of Technology (USA)

Hesuan Hu
Xidian University (PR China)

Ferdinando Rodriguez Y Baena
Imperial College (UK)

Yu Sun
University of South Florida (USA)

Loredana Zollo
Campus Bio-Medico University, Roma (Italy)

Past Editor-in-Chief
Peter Corke
Queensland University of Technology (Australia)

RAM Editorial Assistant and Column Manager
Rachel O. Warnick (USA)

COLUMNS
Competitions: Stephen Balakirsky and Dan Popa (USA)

Education: Paolo Fiorini (Italy) and M. Ani Hsieh (USA)

From the Editor's Desk: Eugenio Guglielmelli (Italy)

Industry News Humanitarian Technology, and Standards: Raj Madhavan (USA)

On the Shelf: Alex Simpkins (USA)

President's Message: Raja Chatila (France)

Regional Spotlight: Kenji Suzuki (Japan) and Xiaorui Zhu (China)

ROS Topics: Steve Cousins (USA)

Student Corner: Lauren Miller (USA)

TC Spotlight: Fumihito Arai (Japan)

Turning Point: Eugenio Guglielmelli (Italy)

Women In Engineering: Laura Margheri (Italy) and Gentiane Venture (Japan)

IEEE RAS Vice-President of Publication Activities
Antonio Bicchi
University of Pisa (Italy)

RAM home page:
http://www.ieee-ras.org/publications/ram

Robotics and Automation Society Project Specialists
Kathy Colabaugh
Rachel O. Warnick
ras@ieee.org

Advertising Sales
Mark David
Sr. Manager Advertising & Business Development
Tel: +1 732 465 6473
Fax: +1 732 981 1855"
m.david@ieee.org

IEEE Periodicals Magazines Department
Debby Nowicki
Managing Editor
d.nowicki@ieee.org

Janet Dudar
Senior Art Director

Gail A. Schnitzer, Mark Morrissey
Associate Art Director

Theresa L. Smith
Production Coordinator

Felicia Spagnoli
Advertising Production Manager

Peter M. Tuohy
Production Director

Dawn M. Melley
Editorial Director

Fran Zappulla
Staff Director, Publishing Operations

IEEE-RAS Membership and Subscription Information:
+1 800 678 IEEE (4333)
Fax: +1 732 463 3657
http://www.ieee.org/membership_services/membership/societies/ras.html

Digital Object Identifier 10.1109/MRA.2015.2452053

IEEE prohibits discrimination, harassment, and bullying. For more information, visit http://www.ieee.org/web/aboutus/whatis/policies/p9-26.html.

IEEE Robotics & Automation Magazine (ISSN 1070-9932) (IRAMEB) is published quarterly by the Institute of Electrical and Electronics Engineers, Inc. Headquarters: 3 Park Avenue, 17th Floor, New York, NY 10016-5997 USA, Telephone: +1 212 419 7900. Responsibility for the content rests upon the authors and not upon the IEEE, the Society or its members. IEEE Service Center (for orders, subscriptions, address changes): 445 Hoes Lane, P.O. Box 1331, Piscataway, NJ 08855 USA. Telephone: +1 732 981 0060. Individual copies: IEEE members US$20.00 (first copy only), non-members US$118.00 per copy. Subscription rates: Annual subscription rates included in IEEE Robotics and Automation Society member dues. Subscription rates available on request. Copyright and reprint permission: Abstracting is permitted with credit to the source. Libraries are permitted to photocopy beyond the limits of U.S. Copyright law for the private use of patrons 1) those post-1977 articles that carry a code at the bottom of the first page, provided the per-copy fee indicated in the code is paid through the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923 USA; 2) pre-1978 articles without a fee. For other copying, reprint, or republication permission, write Copyrights and Permissions Department, IEEE Service Center, 445 Hoes Lane, Piscataway, NJ 08854. Copyright © 2015 by the Institute of Electrical and Electronics Engineers Inc. All rights reserved. Periodicals postage paid at New York and additional mailing offices. Postmaster: Send address changes to IEEE Robotics & Automation Magazine, IEEE, 445 Hoes Lane, Piscataway, NJ 08854 USA. Canadian GST #125634188 PRINTED IN THE U.S.A.

SUSTAINABLE FORESTRY INITIATIVE
Certified Chain of Custody
At Least 25%Certified Forest Content
www.sfiprogram.org
SFI-01042

IEEE

FROM THE EDITOR'S DESK

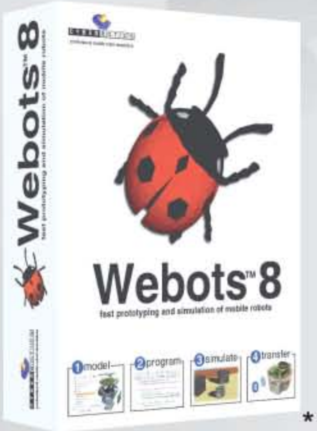
Research Reproducibility and Performance Evaluation for Dependable Robots

By Eugenio Guglielmelli

This issue of *IEEE Robotics & Automation Magazine (RAM)* focuses on reproducibility and measurability of robotics research. When announcing this special issue, I already outlined the relevance of the challenge for our community, as also emphasized by our cover page. Please do not expect to find in the features of this issue news on recent discoveries about basic physics principles enabled by (humanoid) robots; what the issue is actually about is the development of methods, tools, benchmarks, data sets, standard setups for performance evaluation via experimental tests—or even competitions—devised for a variety of robotics research areas. I am very happy that we had a high number of submissions with many high-quality contributions to such an extent that a focused section of the next issue will be devoted to these same topics. It is also remarkable that the IEEE is currently considering research reproducibility as becoming the key quality requisite for articles appearing in IEEE publications. This idea was recently debated at the IEEE Panel of Editors meeting which was held in Washington D.C. last April. Gianluca Setti, the IEEE past vice-president for publications, who promoted such debate within the IEEE, kindly agreed to be interviewed to give his perspective on the relevance of this topic, as reported in the Turning Point column of this issue (see page 192). Thanks to this successful special issue, the IEEE Robotics and Automation Society demonstrates that we are well aware of and perfectly in line with this important trend. This is clearly a key aspect for a mature research field like robotics and automation, which is expected to deliver more disruptive innovation in the short–medium term. But it is definitely not enough. Early July 2015 an accident was reported in an automotive factory in Germany: A worker died while setting up a robotic welding unit. Immediately after that this sad event was reported in the news, there was some speculation on media channels, amplified by social networks, about the possibility that this was the first accident related to the use of a novel generation of collaborative robotic systems designed to allow the active presence of human workers within the robot workspace during operations. Fortunately, it was soon clarified that this was not the case, but I believe that this reaction is clear evidence of the level of attention paid by the general audience to the risks related to the introduction of robots in our daily living scenarios. That's why what I really consider as the most important payoff of having measurable and reproducible results while evaluating the performance of robotic systems is the possibility of demonstrating not only their safety but also other crucial properties, such as dependability and resilience, which robotics technology shall feature to be eventually accepted by nonexpert users into novel application scenarios. The concept of dependability was introduced for computer systems in 1992 by the late Dr. Jean Claude Laprie, a senior researcher at the Laboratory for Analysis and Architecture of Systems (LAAS), Toulouse, France. The idea that modern robots also should be designed to be dependable was first proposed by the late Dr. Georges Giralt, an IEEE Fellow and a pioneer of robotics research also based at LAAS, who initiated a successful series of workshops in 2001 and edited an *IEEE RAM* special issue (vol. 11, no. 2, 2004) on this topic. Dependability means not only safety, reliability, service availability, integrity, confidentiality, and maintainability—all properties that can be objectively measured—but also the level of trust end-users have in the system performance: How far would you go in depending on a service directly provided to you by a robot system, such as an autonomous car, a drone, a coworker or a personal assistant for independent living? Tackling such subjective perspective for evaluating the system performance still requires a lot of research effort, especially when you consider that robots should be capable not only of being dependable in routine operational conditions, but also resilient to external malicious attacks, which purposely aim at generating failures and accidents. No doubt, it is an intriguing perspective that will engage generations of researchers. Enjoy the issue!

Digital Object Identifier 10.1109/MRA.2015.2470815
Date of publication: 11 September 2015


$129
$79
$69
$59
KUKA
KUKA
adept mobilerobots


Webots™ 8
Webots™ 8
fast prototyping and simulation of mobile robots
1 model
2 program
3 simulate
4 transfer

PRESIDENT'S MESSAGE

On the Ethics of Research and Practice in Robotics and Automation

By Raja Chatila

Research in robotics and automation (R&A) has reached an unprecedented stage of maturity.

Enabled by increased computing speed and memory storage, the miniaturization of sensors and actuators, new materials, and, most importantly, research results in R&A at large, we can see today that robots and systems are capable of achieving impressive operational functions in perception, motion planning, motion control, manipulation, human–robot interaction, and so on, in a variety of situations. Full autonomous operation is possible in some cases. Of course, a great amount of research is still necessary to understand many underlying principles, in decision making, control, perception, or learning, and to achieve operation in open-ended situations.

Additional questions are raised, mainly about self-driving cars or certain types of military usage of robots.

These robots have now come out of the labs. New applications are booming in many sectors: transport, services, defense, manufacturing, agriculture, construction, medicine, and health. Many others are yet to come.

This wealth of applications has made R&A very visible to governments, the media, and the general public worldwide. Questions on ethical, legal, and societal issues in the use of robots have emerged and are becoming more meaningful to scientists and engineers as well as the general public. Statements are made, by specialists and nonspecialists alike, about the consequences on jobs and the ethical use of robots but also about robots and artificial intelligence (AI) "taking over the world." And when it comes to autonomy, additional questions are raised, mainly about self-driving cars or certain types of military usage of robots, e.g., autonomous lethal weapons.

Many of these questions are undoubtedly legitimate, even if they are often based on misconceptions about robots and their actual capabilities and about the state of the art in R&A and AI.

Scientists and engineers engaged in robotics research have at least a moral responsibility about the outcome of their work and sometimes also about the misconceptions within the public because of some of their statements. The R&A community started to reflect on the question of the ethical implications of robotic technology and of autonomous robots more than ten years ago, more precisely, in 2002, within a research atelier funded by the European Robotics Research Network (EURON). The first workshop was organized by Gianmarco Veruggio and Fiorella Operto on 30–31 January 2004, in Villa Nobel, San Remo, Italy. The same year, the IEEE Robotics and Automation Society established a technical committee on "roboethics." Robot ethics is today an interdisciplinary research area at the intersection of applied ethics, robotics, and AI. The European Union (EU) funded several projects in the past ten years on this issue, and, recently, Robolaw, the conclusions of which were presented before representatives of the EU Parliament. In some countries, official ethics committees on robotics research have been formed [e.g., the French Advisory Commission for the Ethics of Research in Information Sciences and Technologies (CERNA) in France, which addresses more broadly digital sciences and technologies].

But what exactly is the ethics of research in our area? Like in other areas, it is to reflect, since the early stage of research and as applications and technologies get closer to actual usage, on the consequences of this research and these developments on our human societies and to consequently make appropriate decisions and take appropriate measures measures according to our moral standing.

Similarly to physics, our domain raises many questions about the use of technologies developed from theories and about the machines designed after them. And similarly to biology, R&A raises profound questions on the very nature of human beings. The list of these questions is long. Just to mention a few: Should robots be allowed to autonomously make decisions that could knowingly endanger human lives? What are the consequences of robots expressing emotions with people who

Digital Object Identifier 10.1109/MRA.2015.2453691
Date of publication: 11 September 2015


SIMPLE
SEXY
SAFE
KINOVA

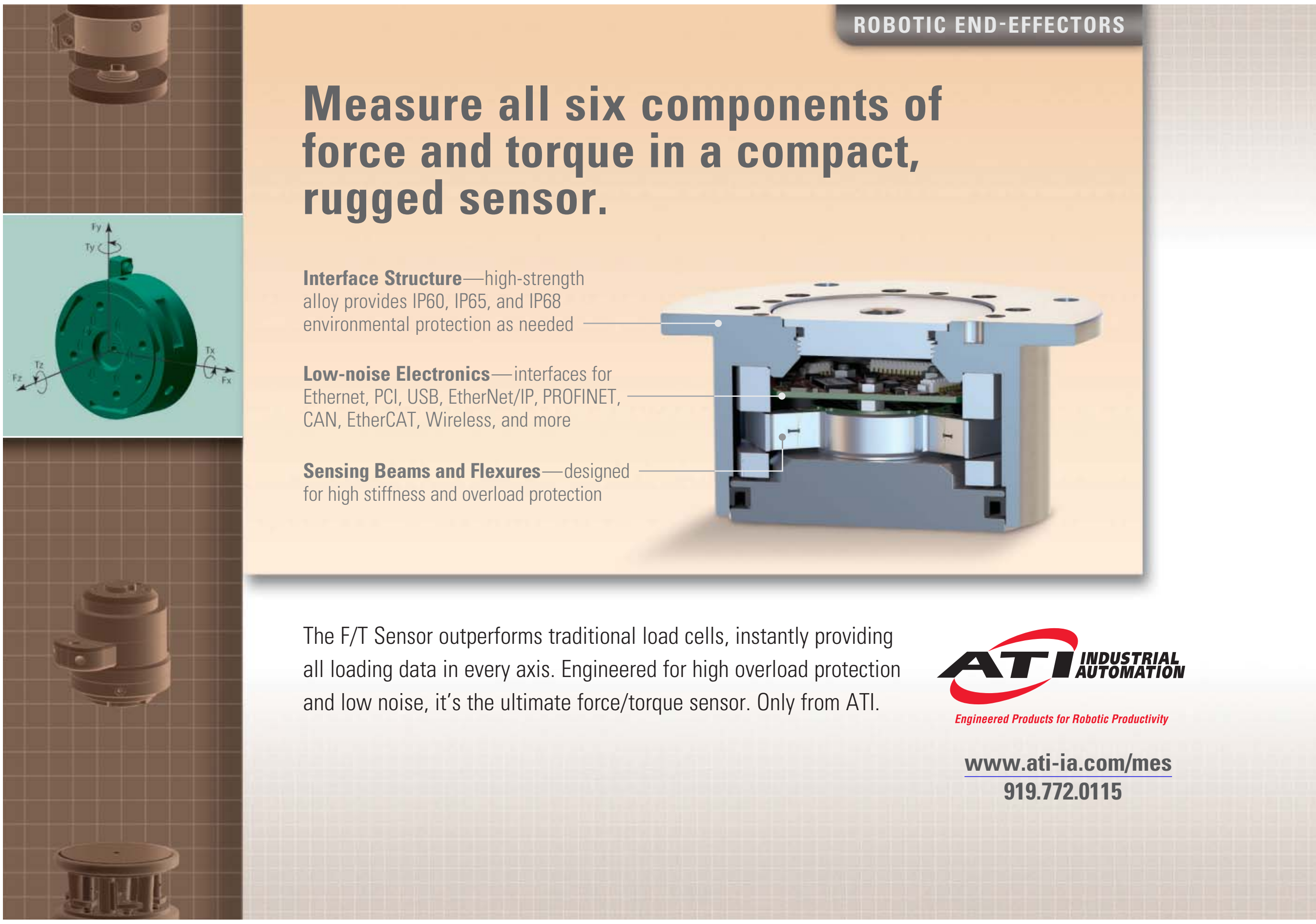
could be psychologically vulnerable, immature, or diminished, especially people or children who have no technological background? Can robots improve or reduce human dignity? What are the consequences of building robots that mimic or replicate animals, or, more importantly, human beings, in appearance and behavior? What could be the consequences of human augmentation by means of robotic devices? These last two questions raise debates on human identity. Another question pertains to the status of robots, especially whether they should be classified as akin to living beings, in human society.

Scientists and engineers engaged in robotics research have at least a moral responsibility about the outcome of their work.

In addition to these questions, there are many others related to legal aspects, such as the accountability and responsibility of robots, privacy, and intimacy of humans interacting with them.

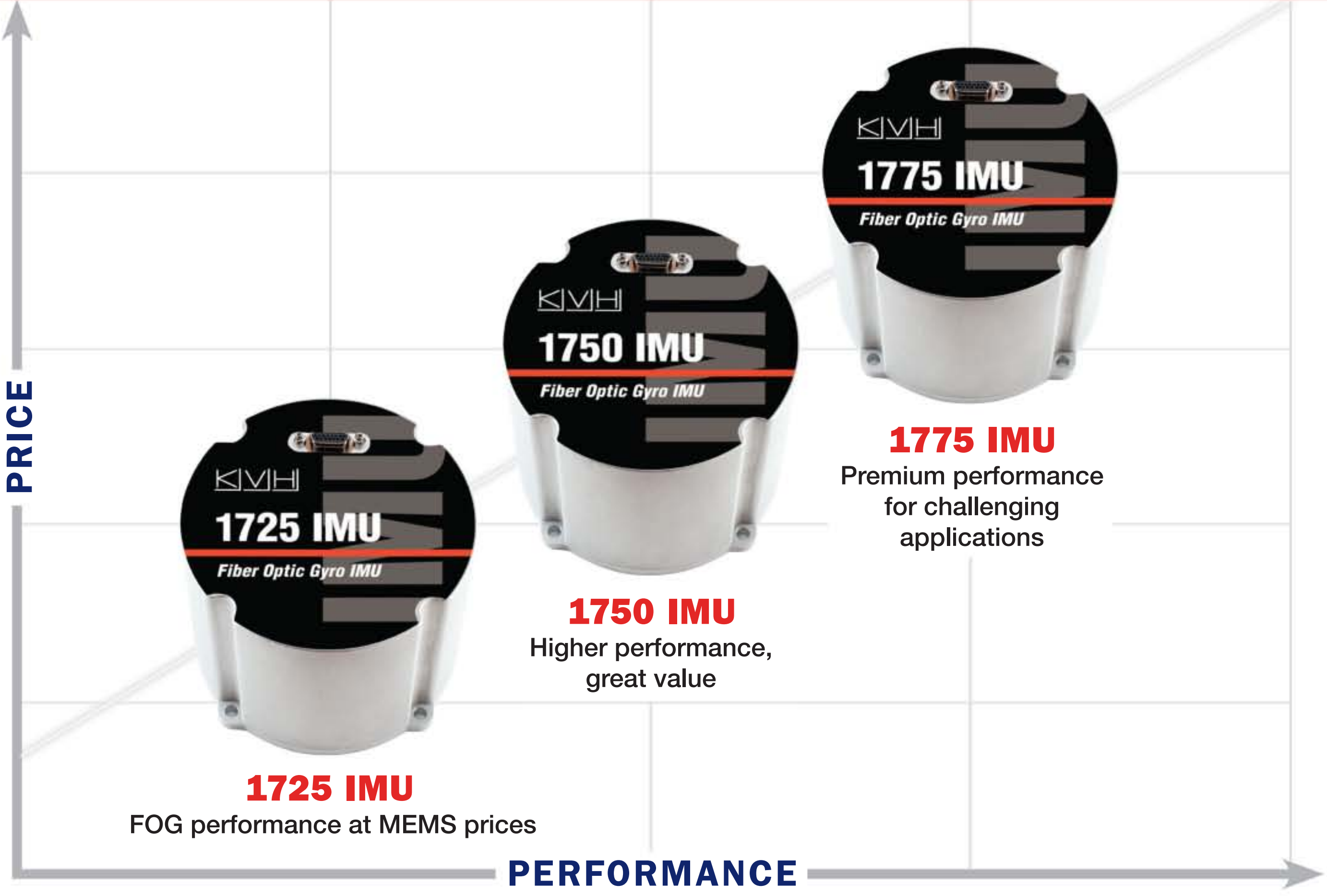
It is our responsibility as researchers, engineers, and practitioners to address these questions individually and as a community and to provide answers. Our Society also has its share of this responsibility. This is why a new Standing Committee on Ethics of Robotics and Automation Research and Practice has been introduced in the latest revision of our bylaws and voted by the Administrative Committee in May 2015, to organize reflections and to contribute opinions and recommendations on this major issue.

By doing this, we will be advancing technology for the benefit of humanity.



Great...Better...Best!

KVH Introduces a Fiber Optic Gyro IMU for Every Demanding Application



1775 IMU
Premium performance for challenging applications

1750 IMU
Higher performance, great value

1725 IMU
FOG performance at MEMS prices

	1725 IMU	1750 IMU	1775 IMU
Bias Instability	≤0.1°/hr	≤0.05°/hr	≤0.05°/hr
Bias vs. Temperature	≤4°/hr	≤1°/hr	≤1°/hr
Angle Random Walk (25° C)	≤0.017°/√hr	≤0.012°/√hr	≤0.012°/√hr
Bandwidth	≥440 Hz	≥440 Hz	≥1000 Hz
Data Rate	1 to 1000 Hz	1 to 1000 Hz	1 to 5000 Hz
Bias Magnetic Sensitivity	±2°/hr/gauss	±2°/hr/gauss	±0.5°/hr/gauss

Need help selecting the right IMU for your application?
Get the details and download the white paper

kvh.com/imu

KVH®

KVH Industries, Inc. World Headquarters: 50 Enterprise Center | Middletown, RI 02842 U.S.A. | info@kvh.com | +1 401.847.3327

The Amazon Picking Challenge 2015

By Peter R. Wurman and Joseph M. Romano

The first Amazon Picking Challenge (APC) was held at the 2015 IEEE International Conference on Robotics and Automation (ICRA) in Seattle, Washington, 26–27 May. The objective of the competition was to provide a challenge problem to the robotics research community that involved integrating the state of the art in object perception, motion planning, grasp planning, and task planning to manipulate real-world items in industrial settings. To that end, we posed a simplified version of the task that many humans face in warehouses all over the world, i.e., picking items from shelves and putting them into containers. In this case, the shelves were prototypical pods from Kiva Systems, and the picker had to be a fully autonomous robot.

Amazon provided US$26,000 in prize money for the winning teams.

The items were a preselected set of 24 products that were commonly sold on Amazon.com and that we expected would pose varying degrees of difficulty for the contestants. On the easier end were simple cuboids like a box of straws or a spark plug. Some items were chosen because they were easy to damage, like the two soft-cover books or the package of crushable Oreo cookies. Others were harder to perceive and grasp, like the unpackaged dog toys or the black mesh pencil holder. The box of Cheez-Its posed a challenge because it could not be removed from the bin without twisting it sideways.

Each pod had 12 bins, and the 25 products were distributed among the bins in such a way that each competitor had the same challenges. Each bin had one target item to be picked, with a base score of ten, 15, or 20 points depending on how many other items were in the bin. In addition, some items that were projected to be more difficult to pick were given one to three bonus points. Damaging an item incurred a five-point penalty, while picking the wrong item incurred a 12-point penalty. Each competitor had 20 min to pick as many of the 12 target items as possible and could score as many as 190 points.

The competition was announced 1 October 2014. Through a series of video submissions, the organizers selected 25 teams to receive equipment grants (sample pods and products) and travel grants to help defray the costs of travel to the venue. In addition, Amazon provided US$26,000 in prize money for the winning teams.

Representing 11 different countries, 26 teams made the trip to Seattle to try their robot's hand at picking out of Kiva pods. The success of the teams was mixed, but the enthusiasm and excitement was contagious. The competition was won by RBO from the Technical University of Berlin, Germany. Its device, with a Barrett arm, a Nomadic Technologies mobile platform, and a suction cup attached to a commercial vacuum cleaner, was able to successfully pick ten of 12 correct items in under 20 min. Their score of 148 points put them well into the lead. The Massachusetts Institute of Technology (MIT) team placed second, with seven items picked and 88 points (Figure 1). The MIT entry used

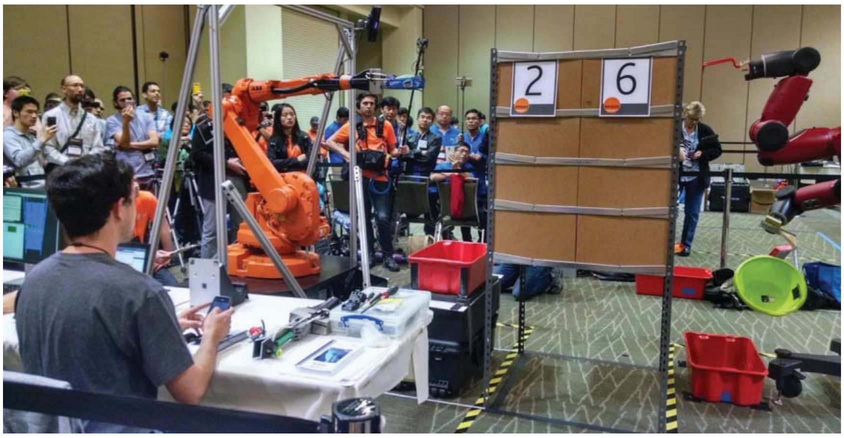


Figure 1. A crowd gathers to watch Team MIT's entry. MIT's robot placed second after picking seven items in 20 min. (Photo courtesy of Joseph Romano.)

Digital Object Identifier 10.1109/MRA.2015.2452071
Date of publication: 11 September 2015

Table 1. The final scores of the APC 2015.

Team	Affiliation	Items Picked	Score
RBO	Technical University of Berlin	10	148
Team MIT	MIT	7	88
Grizzly	DataSpeed and Oakland University	3	35
NUS Smart Hand	National University of Singapore	2	32
Z.U.N.	Zhejiang University, University of Technology–Sydney, and Nanjiang Robotics Co.	1	23
C^2M	Chubu University, Chukyo University, and Mitsubishi Electric Corporation	2	21
R U Pracsys	Rutgers University	1	17
Team K	JSK, University of Tokyo	4	15
Nanyang	Nanyang Technological University	1	11
A.R.	The Netherlands	1	11
Team Georgia Tech	Georgia Tech	1	10
Team Duke	Duke University	1	10
CVAP	KTH (Sweden)	2	9

an industrial ABB arm and a scooper end effector that could be flipped over to alternatively use a small suction cup. In third place was Team Grizzly from Dataspeed Inc. and Oakland University, with 35 points. Their solution used a Baxter robot attached to a custom mobile base.

The final scores are shown in Table 1. Many teams demonstrated successful picking in their warm-ups but, for various reasons, failed in their official 20-min attempt. The reasons for failure varied widely and included last-minute code changes, failure to model how a vacuum hose would twist around the arm in certain poses, and grippers that were so big that they could not figure out how to get in the bin. However, even the systems that failed to pick any items demonstrated interesting robots, end effectors, and technical approaches. Overall, 36 correct and seven incorrect items were picked.

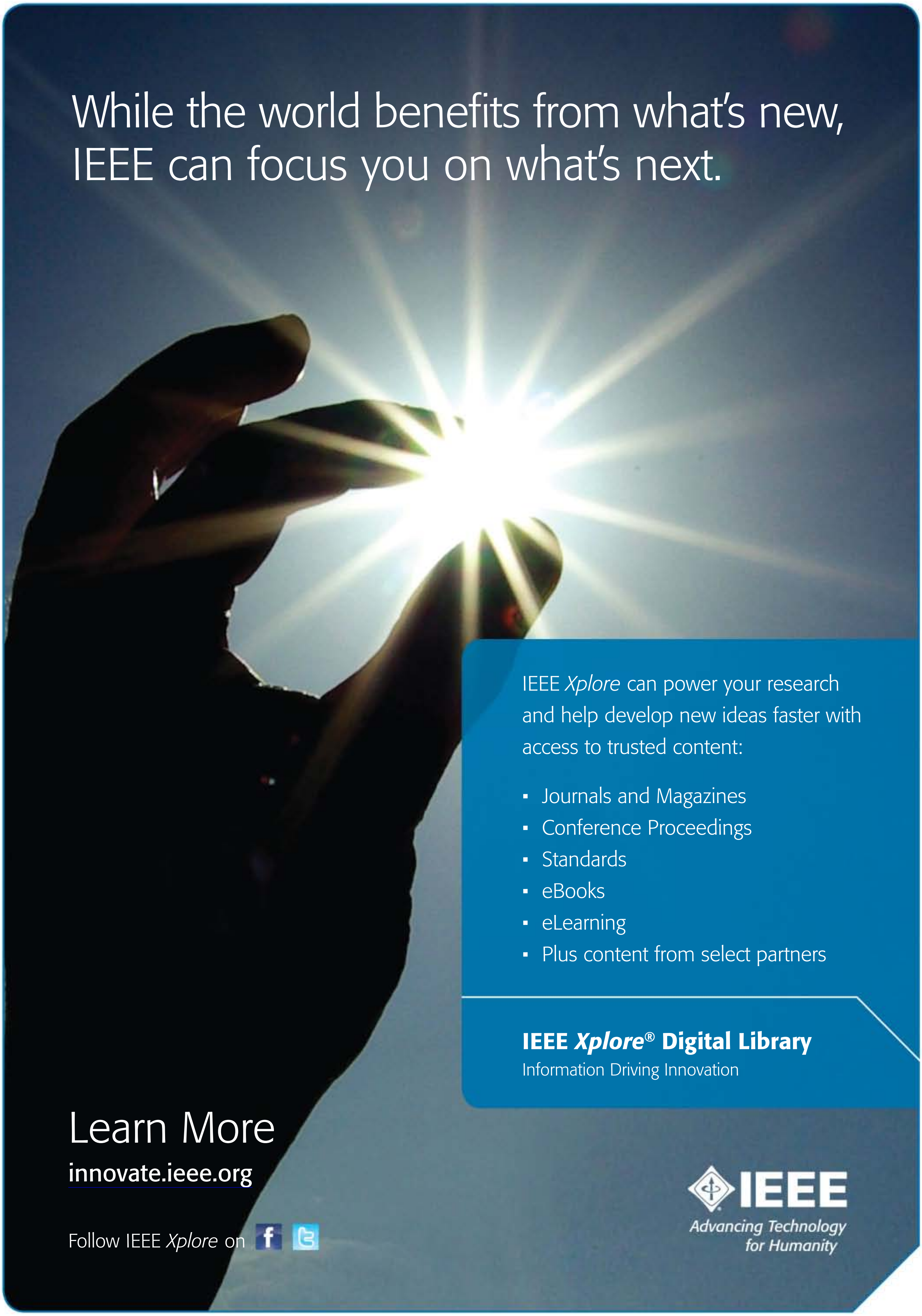
Other teams competing included Worcester Polytechnic, the University of Texas at Austin, the University of Texas at Arlington, the University of Washington, the University of Alberta, Robological PTY LTD, Universitat Jaume I, the University of Colorado at Boulder, Colorado the School of Mines, the University of Pisa, the University of California at Berkeley, Dorabot and the University of Hong Kong, and St. Francis Institute of Technology in India. The teams were supported by several hardware vendors, including Rethink Robotics, Barrett Technologies, Yaskawa, Olympus Controls, and Clearpath Robotics.

The first APC was very successful, drawing a large number of competitors from around the world and demonstrating the state of the art in both the software and the hardware required for robotic manipulation. Despite being scattered over 16 testing bays in the ICRA competition area and spread over two days, every team drew a large crowd of spectators eager to see how the robots would perform (Figures 2 and 3).

For more information, see http://amazonpickingchallenge.org.

Figure 2. The APC judges perform one final check before Team A.R. begins its trial. (Photo courtesy of Peter Wurman.)

Figure 3. The audience watches as a gripper reaches out to pick an item.



Robotics and Automation Activities in China

By Hong Qiao, Rui Li, and Peijie Yin

> This is the fourth in the series of articles focusing on the state of robotics and automation in Brazil, Russia, India, China, and South Africa (the BRICS countries). This article provides an overview of activities in China. The objective of this series is to inform the readers of the unique challenges of these countries and the solutions they have adopted to solve their problems and to facilitate discussions with the interested members of the community. Please send your comments and feedback to the authors and the vice president of the Industrial Activities Board.
>
> —Raj Madhavan (raj.madhavan@ieee.org)

China is becoming one of the world's largest robot markets. In 2014, President Xi Jinping's speech on developing robots in China created a boom in the national robot industry. Governments at all levels and various societies sped up their pace on encouraging the research and development of robotic technologies. Today in China, there are infinite possibilities when it comes to robotics.

China is one of the world's fastest-growing major economies and one of the BRICS countries: Brazil, Russia, India, China, and South Africa. China keeps close economic and trade relations with various countries, being the world's largest exporter and importer of goods. Ever since President Xi's speech on developing the Chinese robot industry, the whole country is taking positive actions to realize this goal. As of 2014, China has already become the world's fastest-growing market for industrial robots, and it is expected to be the top market in a few years.

Current Situation of the Chinese Robot Industry

China is a large manufacturing country. At present, due to the pressing need for industrial upgrades and the release of planning policies related to intelligent manufacturing, industrial robots in China have been put into massive use in the field, such as in the automotive parts, electronics, chemical engineering, and machinery industries. The rapid growth of the industrial robot market in China is drawing the attention of the global robot industry. In fact, with the global popularization and application of robots, the local governments in China are all accelerating their implementation of the "Replace Human Workers with Robots" plan.

However, in comparison with the quickly growing demand, the Chinese robot market got a late start, and its productivity is inadequate. International Federation of Robotics data show that, among the robots sold in the Chinese market, the sales volume is small from local suppliers, while it is large for those from foreign suppliers [1].

On the other hand, under the influence of an aging population, the demand for service robots is increasing in China. In 2012, the Ministry of Science and Technology of China released the Special Planning for Scientific and Technological Development of Service Robots in the 12th five-year plan for the robot field, in which it was planned to focus on the cultivation and development of the emerging industry of service robots, and the development of public security robots, biomimetic rehabilitation robots, bionic robot platforms, and modularized core components [2].

Policies and Government Support

In recent years, from the strategic perspective of national scientific development, the Chinese government is planning to support the development of robot research and industry. Significant funds have been injected into research on industrial and service robots, and policy support for the development of the robot industry has gradually increased.

- On 9 June 2014, President Xi Jinping spoke at the biennial conference of the country's two top think tanks, the Chinese Academy of Sciences (CAS) and Chinese Academy of Engineering [3]. After his speech, the National Development and Reform Commission, Ministry of Science and Technology, and National Natural Science Foundation made active responses and released policies to promote the development of the robotics industry of China. Governments of all levels also responded to the speech. The deployment of robot industry development zones began in many cities,

Digital Object Identifier 10.1109/MRA.2015.2452193
Date of publication: 11 September 2015

and the establishment of robot federations in Guangdong, Jiangsu, and so it was planned to propel the technical innovation and industrial upgrade of robots.

- On 25 March 2015, "Made in China 2025" was approved in the meeting of the Standing Committee of the State Council. Intelligent manufacturing has become the development trend of the current global manufacturing industry, and it will propel the in-depth integration of informatization and industrialization in the coming period in China [4]. As the information-acquisition, processing, decision-making, and major execution mechanism in intelligent manufacturing, robots play a significant role.
- Early in 2012, in the 12th five-year plan of the National Natural Science Foundation, multirobot cooperation and bionic robots (including autonomous control of robots and cooperative planning and control for multiple tasks with multiple robots) were the fields given priority for development.
- On 5 March 2014, Li Keqiang, the premier of the State Council, pointed out in the government work report that the reform of the scientific and technological system shall be quickened, the leading role of enterprises in technical innovation shall be strengthened, and enterprises shall be encouraged to set up research and development institutions and to lead the establishment of cooperative innovation federations among the industry, universities, and research circle. Besides, scientific research projects and capital management shall be improved and enhanced, national innovation investigation and scientific report system shall be established, and researchers shall be encouraged to set up enterprises. The protection and application of intellectual property shall be strengthened to promote scientific innovation [5].
- It is mentioned in *Revised Draft of the Patent Law of PRC (for Public Review)*, which is under amendment this year, that "if it is agreed between the institute and inventor or designer according to the regulations in Clause IV, Article VI of this Law that the right of applying for a patent for the invention belongs to the institute, the institute shall give award and payment to the inventor or designer" [6]. This is also favorable for promoting the researchers to conduct scientific innovation.

All of these policies indicate that the Chinese government is now actively propelling the development of the robot industry to improve the innovation capability and competitiveness of the Chinese robot industry. As China's industrial upgrading proceeds, industrial robots are playing an increasingly important role in the manufacturing industry. At the same

time, service and rehabilitation robots are also attracting attention from the government and the industry since the aging population is becoming a serious issue in China. The continuing policy and funding support from the government would exert a profound influence on the improvement of science, technology and manufacturing level of robotics, and automation. The reform of the scientific research system and intellectual property system will be helpful to attract top talent to this area and in conducting innovation. It may also narrow the gap between laboratories and the market and promote technology transfer.

Robotics Research Institutes and Academic Activities

In China, most of the studies on robotics are conducted in colleges and institutes, such as the Shenyang Institute of Automation (SIA), the Institute of Automation of the CAS (CASIA), the Robotic Institute of Beihang University, the Robot Research Institute of Beijing Institute of Technology, the State Key Lab of Robotics, and the System of Harbin Institute of Technology. Among those, SIA runs the largest robot company in China; CASIA is focused on the research and development of a brain-like intelligent robot; universities, such as Beihang University and Birla Institute of Technology, are interested in the discovery of service robots and surgical robots.

To date, there are only a few academic conferences in China with robotics as the main topic, and the lack of robotics conferences has been a topic of discussion. As a developer of robotic research and innovation, there should and could be more international and domestic conferences held in and by China, and local governments of China and various robot groups have shown their interest in preparing and arranging such activities.

Meanwhile, there are some conferences in the area of automation, such as the World Congress on Intelligent Control and Automation, the Chinese Control Conference, and the Chinese Control and Decision Conference. These conferences are all held in China and steadily gaining influence in the world. They provide a forum for scientists, engineers, and practitioners throughout the world to present their latest theoretical results and techniques in the fields of systems, intelligent control, automation, robotics, and so on. In recent years, all of the conference proceedings have been selected for coverage in the *Index to Scientific and Technical Proceedings* and cited by *Engineering Index*.

It should be noted that the influence of Chinese researchers is rising in the IEEE Robotics and Automation Society (RAS). In 2014, the first researcher from mainland China participated in the IEEE RAS Administrative Committee. We would not be surprised to witness this trend continue as China's innovative research on robotics and automation grows.

Education and Courses in Robotics

Today in China, various levels of educational institutions offer courses on robotics and related areas, including basic education, tertiary education, and vocational education.

Lessons on robotics in primary and secondary schools are mainly offered in developed regions and cities and focus on cultivating the students' interest about robots. Some toy robots are used in these lessons, e.g., Lego Mindstorms, to build creative designs or inventions.

Courses on robotics in college and universities aim at senior undergraduates and postgraduates; the content of the courses covers kinematics, dynamics, robot vision, intelligent control, motion planning, and so on. The courses are theoretical, and students may not have a chance to operate a real robot.

Vocational schools are also highly involved. With the expansion of the application scope of robots in China, a large quantity of operators, maintenance engineers, and design engineers is required to satisfy the diversified requirements of enterprises. Driven by the rising demands from industry, many vocational schools are setting up courses on the operation and program design of the industrial robot. Meanwhile, with funding support from companies, training textbooks on how to program and operate an industrial robot are in preparation.

Public Competition and Activities

To popularize the science of robotics and arouse public interest, especially with youngsters, there are several competitions, including RoboCon, RoboCup, and the Smart Car Competition. The competitions are all robot themed, and the contestants are required to build a certain type of specialized robot to finish a task, such as speed racing or object grasping. Because of the enjoyment and entertainment of these activities, these competitions usually attract a large quantity of Chinese college students.

Usually contestants have three to six months to prepare for the competition. To build the robots, they must possess knowledge of programming, mechanical design, and electronic circuit design. Sometimes they have to work through the night to catch up with the schedule.

The contestants face fierce competition in the game. There is sometimes only a 0.1-s difference between the winner and the loser. They work very hard to get a good prize, and the experience will help them in their future careers. Other activities, such as campus open houses, may also help the public get access to the top advanced robotic technologies. The competitions and activities for the public lay the foundation of young robotic talents for the next generation and improve public understanding of scientific policies.

Conclusion

China is upgrading its economy, especially the manufacturing industry. The robot industry is playing a key role in this significant period. All levels of government and private capitals are paying close attention to the development of this field. The urgent need for more advanced and intelligent robots is a challenge for robotic research but also draws stronger support for researchers. The Chinese robotics and automation community will be more influential and international. We would like to conclude that the future of robotics and automation in China will be more active, more competitive, and more innovative.

References

[1] (2013). National Data from National Bureau of Statistics of China (NBS). [Online]. Available: http://data.stats.gov.cn

[2] Z. C. Gui and J. D. Wu, "Study on the status and trend of global robotic industry and forecast of future robotic industry in China," *Dongfang Elect. Rev.*, vol. 4, no. 2, 2014.

[3] (2014, June). Xi insists independent innovation in science, technology. [Online]. Available: http://www.chinadaily.com.cn/china/2014-06/09/content_17573938.htm

[4] (2015, Mar.). Made in China 2025. [Online]. Available: http://www.chinadaily.com.cn/opinion/2015-03/10/content_19764324.htm

[5] (2014, Mar.). Chinese premier delivers gov't work report. [Online]. Available: http://news.xinhuanet.com/english/special/2014-03/05/c_133162002.htm

[6] (2015, Apr.). Revised Draft of the Patent Law of PRC (for public review). [Online]. Available: http://www.sipo.gov.cn/tz/gz/201504/t20150401_1095939.html

2015 IERA Award Joint Winners, Seattle

By Raj Madhavan

The 2015 Invention and Entrepreneurship in Robotics and Automation (IERA) award was shared by François Boucher of Kinova Inc., Canada, and Tom Lipinski of Q-Bot Limited, United Kingdom (Figure 1). The award was announced on 28 May 2015 at the IEEE International Conference on Robotics and Automation award lunch held in Seattle, Washington, after the judging panel decided to jointly honor two of the finalists, Kinova and Q-Bot. The third finalist was Sander Karl of Fortschrittliche Robotertechnologie GmbH & Co., and the presentation was made by Björn Hein of the Karlsruhe Institute of Technology, Germany. The three finalists were selected from a pool of ten applicants.

Digital Object Identifier 10.1109/MRA.2015.2452194
Date of publication: 11 September 2015

Figure 1. From left: Raj Madhavan, awards chair and IEEE RAS vice president, Industrial Activities Board; Tom Lipinski of Q-Bot Limited; François Boucher of Kinova Inc.; and Raja Chatila; IEEE RAS president. (Photo courtesy of ICRA 2015.)

As witnessed by the joint award, the judges had a difficult time coming up with a single winner. The six-member judging panel consisted of Dominik Bösl (KUKA, Germany), Werner Kraus Jr. (Fraunhofer IPA, Germany), Raj Madhavan [awards chair and vice president, IEEE Robotics and Automation Society (RAS) Industrial Activities Board], Mario Munich (iRobot, United States), Erwin Prassler (runfun, Germany) and Nicola Tomatis (Bluebotics, Switzerland). Both Rainer Bischoff (KUKA, Germany) and Alexander Verl (Fraunhofer IPA, Germany) served as judges in the first phase to determine the three finalists.

The panel wrote the following citations in announcing the two winners:

- “for providing an easy-to-use robotic solution that enhances the autonomy of disabled people” (Kinova)
- “for developing a simple robotic solution for insulating homes, thus positively impacting lives of people and protecting the environment by reducing energy consumption” (Q-Bot).

More details on the winning entries are described in “From Need to Innovation” and “Q-Bot—A Robotic Solution for Insulation of Homes”

In its 11th year, the IERA award is jointly sponsored and organized by IEEE RAS and the International Federation of Robotics. It highlights and honors the achievements of inventors with value-creating ideas and entrepreneurs who propel those ideas into world-class products. The entries are evaluated based on criteria that give equal consideration to both innovation and entrepreneurship. The winners were awarded a plaque and a US$2,000 cash prize, which will be shared by the joint winners this year. For additional details on the award and a list of the winners from the previous ten years, see http://www.ieee-ras.org/industry-government/ifr-forum/.

From Need to Innovation

By François Boucher

The story of JACO is that of a man who yearned to be productive. Jacques Forest’s inventiveness was matched only by his desire to be independent. One of three brothers from Quebec, Canada, afflicted with muscular dystrophy, Jacques was confined to a power wheelchair and could only control his left thumb. But Jacques had something that could not be defined by a mere disease or the limits of his body—he had a vision he was determined to see realized.

A passionate and lifelong inventor, Jacques, known to his family and friends as Jaco, devised and built from 1984 to 1997, a robotic arm that was attached to his power wheelchair and allowed him to perform simple tasks previously rendered impossible by his physical state. Jaco’s first act with his new arm was to fulfill his vision. He went outside to the family garden, picked a rose, and brought it back inside, where he presented it to his sister. This simple act of affection, borne of his creativity and persistence, inspired a nation and began the journey that has led to JACO.

Although Jaco died in 1999, Charles Deguire, motivated and inspired by his inventive Uncle Jaco, took up his robotic arm idea and elaborated on it by cofounding Kinova in 2006 to build an assistive robot arm for people with upper-body disabilities. He named the robot, JACO, after his uncle.

Launched in 2010, JACO is a six-axis robotic manipulator arm with a three-fingered hand. This little marvel of engineering significantly improves the lives of persons with reduced mobility. Lightweight, very quiet, unobtrusive, safe, and even weatherproof, JACO assists anyone with an upper-body mobility impairment to perform complex actions. Many everyday activities, such as picking up glasses, holding a fork, or opening a door, which most people do without thinking, can become insurmountable for people who have a disability, and they have to ask someone else to do it for them. The JACO arm makes life easier for these people by giving them greater freedom and independence.

Arm

JACO moves smoothly and silently around six degrees of freedom, with unlimited rotation on each axis. The joints are modular aluminum compact actuators of a unique design that integrates a dc brushless motor, harmonic drive, slip rings, microcontroller, and sensors for torque, current, position, temperature, and acceleration. Its main structure, entirely made of carbon fiber, delivers optimal robustness and durability as well

Digital Object Identifier 10.1109/MRA.2015.2468311
Date of publication: 11 September 2015

Figure 1. The JACO arm can provide those with upper-body disabilities more independence, enabling them to perform everyday tasks, such as eating, on their own.

as a cutting-edge look and feel. The arm is mounted on a standard aluminum extruded support structure attached to the wheelchair seat frame.

Gripper

The gripper consists of three underactuated fingers that can be individually controlled. Their unique bi-injected plastic structure endows them with great flexibility and unrivalled grip. The JACO technology allows the fingers to adjust to any object, whatever its shape; as a result, they can gently pick up an egg or firmly grasp a jar.

Control

JACO is controlled with the same interface that the person is using to control his or her wheelchair. Control is intuitive and allows users to navigate using three different modes: 1) translate, 2) rotate, and 3) grip. In addition, Kinova's intelligent singularity-avoidance algorithm always keeps JACO safely away from unwanted locations.

Exporting Freedom

But the vision of Kinova in assistive robotics is not about a machine, no matter how brilliant, important, or useful it may be. Kinova is committed to advancing the state of the art and science of mobility and is developing a new category of products that share a common objective. It is the same vision that, 20 years ago, inspired Jacques Forest to develop a robotic arm from a discarded lamp, some wood, and some spare electrical parts: to enable the spirit by giving you the power and the freedom to do for yourself (Figure 1).

Q-Bot—A Robotic Solution for Insulation of Homes

By Tom Lipinski

The winning pitch at the IEEE/International Federation of Robotics conference in Seattle was presented by Tom Lipinski, a founding director of Q-Bot and serial entrepreneur. The judges at the event were impressed not just by the robotics technology developed in house and applied in practice but by the positive social impact of the company—improving the lives of people on low incomes and in fuel poverty.

Digital Object Identifier 10.1109/MRA.2015.2472075
Date of publication: 11 September 2015

Q-Bot is a robotics company that improves the energy efficiency of existing homes by providing underfloor insulation to houses with suspended timber floors, of which there are over 8 million in England alone. Previously, the only way to insulate the floors of these homes was to take up the floorboards, a prohibitively expensive solution causing significant disruption and often forcing the resident to move out.

The solution uses a robot inserted through an air vent from outside of the property and deploys it within the void beneath the floorboards. The robot then applies insulation to the underside of the floor boards, keeping them on the warm, dry side and still allowing the ground to breathe, reducing the risk of dampness (Figures 1 and 2). This cost-effective service reduces cold drafts and uneven temperatures, which significantly improves comfort, while saving between US$300 and US$500 per year in energy bills and reducing fuel poverty as a result.

According to Lipinski, the operation currently takes between two and three days, but the target is one day for a two-person team to operate as the technology improves. Crucially, the operation does not require residents to vacate their homes, which is a vast improvement on current solutions that involve decanting tenants and removing furniture, carpets, skirtings, and floorboards before cutting and fitting insulation by hand.

"Current retrofitting is insane," says Lipinski. "It costs too much, it is too disruptive, time consuming, and the quality is rubbish, often with negative impacts on the house and the occupant—think moisture, damp, and indoor air quality." It is a view shared by Peter Armfield, the sustainability manager of CityWest Homes and one of Q-Bot's early adopting clients. "The tenants are comfortable and happy as it makes an immediate difference to their well-being as well as the energy cost. This is why we intend to look at how we can apply the treatment across our portfolio," he says.

The Q-Bot is being offered as a service, rather than standalone hardware and, ideally, requires two trained professionals to operate the robot and spray the foam.

How It Works

Houses that can currently be treated are mostly those with uninsulated suspended floors. These are typically houses built before the mid-20th century, particularly Victorian-era houses in

Figure 1. The floor before insulation by Q-Bot. (Photo courtesy of Matthew Holloway, Q-Bot MD.)

Figure 2. The floor after insulation by Q-Bot. (Photo courtesy of Matthew Holloway, Q-Bot MD.)

England as well as Europe. The Q-Bot solution is delivered by its robots in three steps.

1) The underfloor space is surveyed to build up a detailed three-dimensional map of the area and identify any obstacles or other work that might be required.
2) Insulation is applied in controlled layers to form the required profile.
3) The area covered, consistency, and thickness of the insulation is monitored to verify that the job has been conducted properly and validate the energy saving.

Lipinski said: "A U.K. start-up winning the World Entrepreneurship in Robotics Award might sound odd—we consider Japan as the leader in robotics and the United States in enterprise. Yet, very little of the current robotics know-how is being actively applied to real-life problems, especially in the United States. Yes, we are at the forefront of robotic vision, image fusion, localization and mapping, but it is the impact we have that makes this the most rewarding job on the planet. Our innovative and affordable retrofit service is helping reduce fuel poverty and improving lives while saving energy and the environment. This wouldn't be possible if it wasn't cost-effective, but thanks to the cutting-edge technology, it is—we have a number of social housing clients who prove it."

In the United Kingdom, local authorities and social housing providers such as Camden Council, Peabody, and CityWest Homes have trialed the service in 2014, all with highly positive results. In Camden, heat loss through the floor was reduced by 78%, and cold-air infiltration was eradicated, reducing total infiltration into the house by over 60%. In the CityWest trial, the total heat loss through the floor was reduced by 86%. Most importantly, the trials resulted in happy tenants, living in much more comfortable homes, without having suffered any disruption to their day-to-day life from the insulation work.

The award was announced on 28 May 2015 at the IEEE International Conference on Robotics and Automation.

For further information and images please contact Mathew Holloway, managing director, Q-Bot Limited, mat@q-bot.co, +44 0208 877 2709.

Assistant Professor

The Department of Mechanical and Aerospace Engineering (MAE) at Princeton University is conducting a broad search for two (2) tenure-track assistant professors. We welcome applications from all areas in mechanical and aerospace engineering, including but not limited to the fields of particular interest, namely, (1) robotics and (2) aerospace-related sciences and engineering. Applicants must hold a Ph.D. in Engineering, Materials Science, Physics, or a related subject, and have a demonstrated record of excellence in research with the potential to establish an independent research program. We seek faculty members who will create a climate that embraces excellence and diversity, with a strong commitment to teaching and mentoring.

Princeton's MAE department has a long history of leadership in its core areas of Applied Physics, Dynamics and Controls, Fluid Mechanics, Materials Science, and Propulsion and Energy Sciences, with additional strength in cross-disciplinary efforts impacting areas such as biology, bio-inspired design, the environment, security, and astronautics. We seek creative and enthusiastic candidates with the background and skills to build upon and complement our existing departmental strengths and those who can lead the department into new and exciting research areas in the future.

To ensure full consideration, applications should be received by November 15, 2015. Applicants should submit a curriculum vitae, including a list of publications and presentations, a 3-5 page summary of research accomplishments and future plans, a 1-2 page teaching statement, and contact information for at least three references online at **http://jobs.princeton.edu**, reference number 1500603. Personal statements that summarize leadership experience and contributions to diversity are encouraged.

Princeton University is an equal opportunity employer and all qualified applicants will receive consideration for employment without regard to race, color, religion, sex, national origin, disability status, protected veteran status, or any other characteristic protected by law. We welcome applications from members of all underrepresented groups. This position is subject to the University's background check policy.

Human Movement Understanding

By Emel Demircan, Dana Kulic, Denny Oetomo, and Mitsuhiro Hayashibe

Robotics research has been deeply inspired by humans as a system—from the design of the anthropomorphic aspects of manipulators, sensors, and actuators to the way a robot coordinates the motion of its body and the higher-level strategies for realizing complex tasks and interacting with the external environment. In recent years, robotics computational strategies have contributed significantly to the analysis of human motion and manipulation skills. These analyses have led to advancements in the field of robotics by enabling human-inspired capabilities in robots and simulated systems as well as biologically inspired techniques for robot learning from observation. Furthermore, these analyses have provided a deeper understanding of the human body and its motion-generation strategies. Natural human motion is central to many of the technologies we develop in the field of robotics. This trend is quite visible from the growing number of papers related to the topic of human movement understanding, as shown in Figure 1.

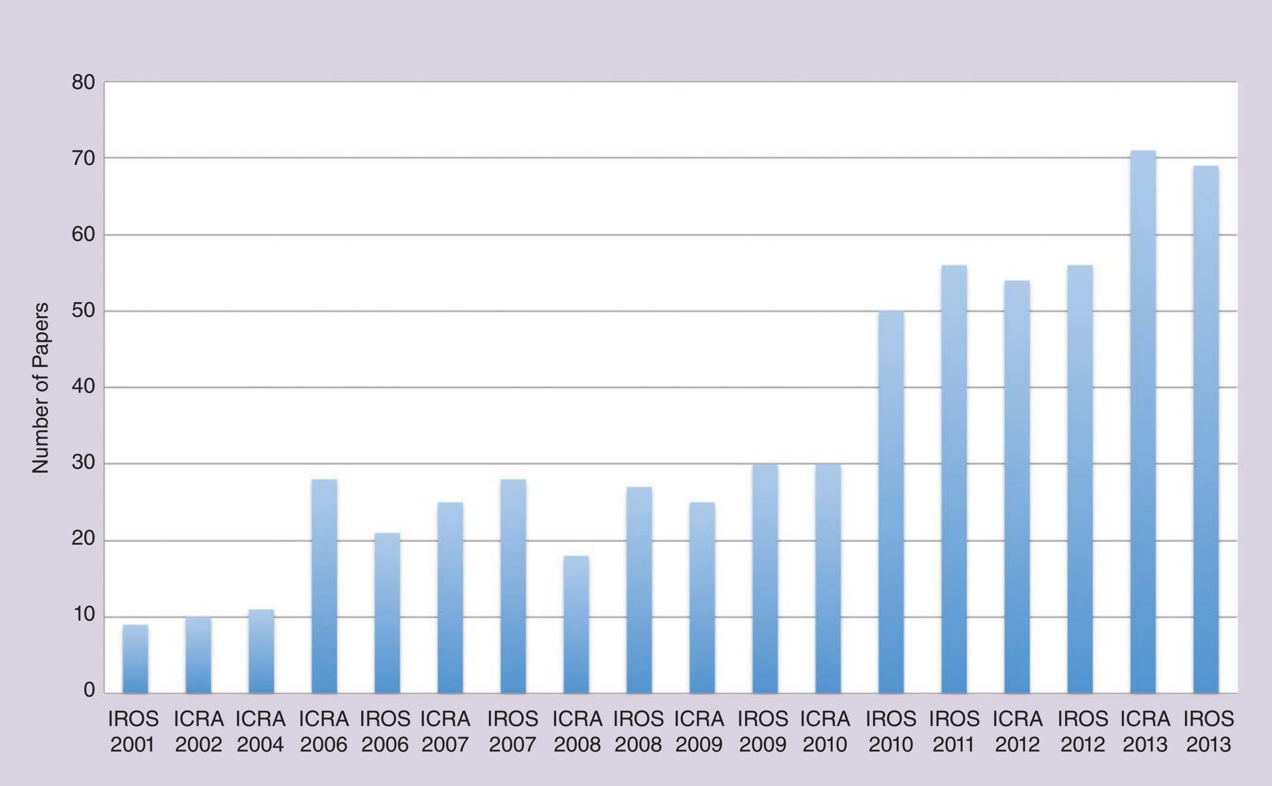


Figure 1. The number of papers related to human motion understanding presented in the last ten years at the technical sessions of the IEEE/Robotics Society of Japan International Conference on Intelligent Robots and Systems (IROS) and the IEEE International Conference on Robotics and Automation (ICRA).

Digital Object Identifier 10.1109/MRA.2015.2452171
Date of publication: 11 September 2015

This human-inspired vision has led to an intersection of robotics and many traditional fields concentrating on the study of human body neuromechanics, such as clinical rehabilitative studies, biomechanics, and neuroscience. Fundamental robotics techniques now find a new frontier of challenges in engaging directly with the human body: in characterizing human motion in terms of physiological aspects, modeling the acquisition of skills, recovery from the traumainduced impairment, generation of natural motion that optimizes the mechanical advantage of the human body, and many others.

Through a robotics view, we aim at gaining fundamental insight into natural human movement and understanding the mechanisms that lead to improved quality of human motion analysis, rehabilitation, and neuroscience (Figure 2). The new IEEE Robotics and Automation Society (RAS) Technical Committee (TC) on Human Movement Understanding was established in May 2014 to create a focal point for this emerging interdisciplinary research field and to facilitate the dissemination within both the robotics and neurophysiology research fields as well as sharing the contributions and the emerging applications with the broader scientific community.

Organizational Structure and Priority Areas

The main organizational structure of this TC consists of four cochairs, with 84 members from universities, clinical institutions, and industry. The current cochairs of the TC are Emel Demircan (United States and Japan), Dana Kulic (North America), Denny Oetomo

This human-inspired vision has led to an intersection of robotics and many traditional fields concentrating on the study of human body neuromechanics.

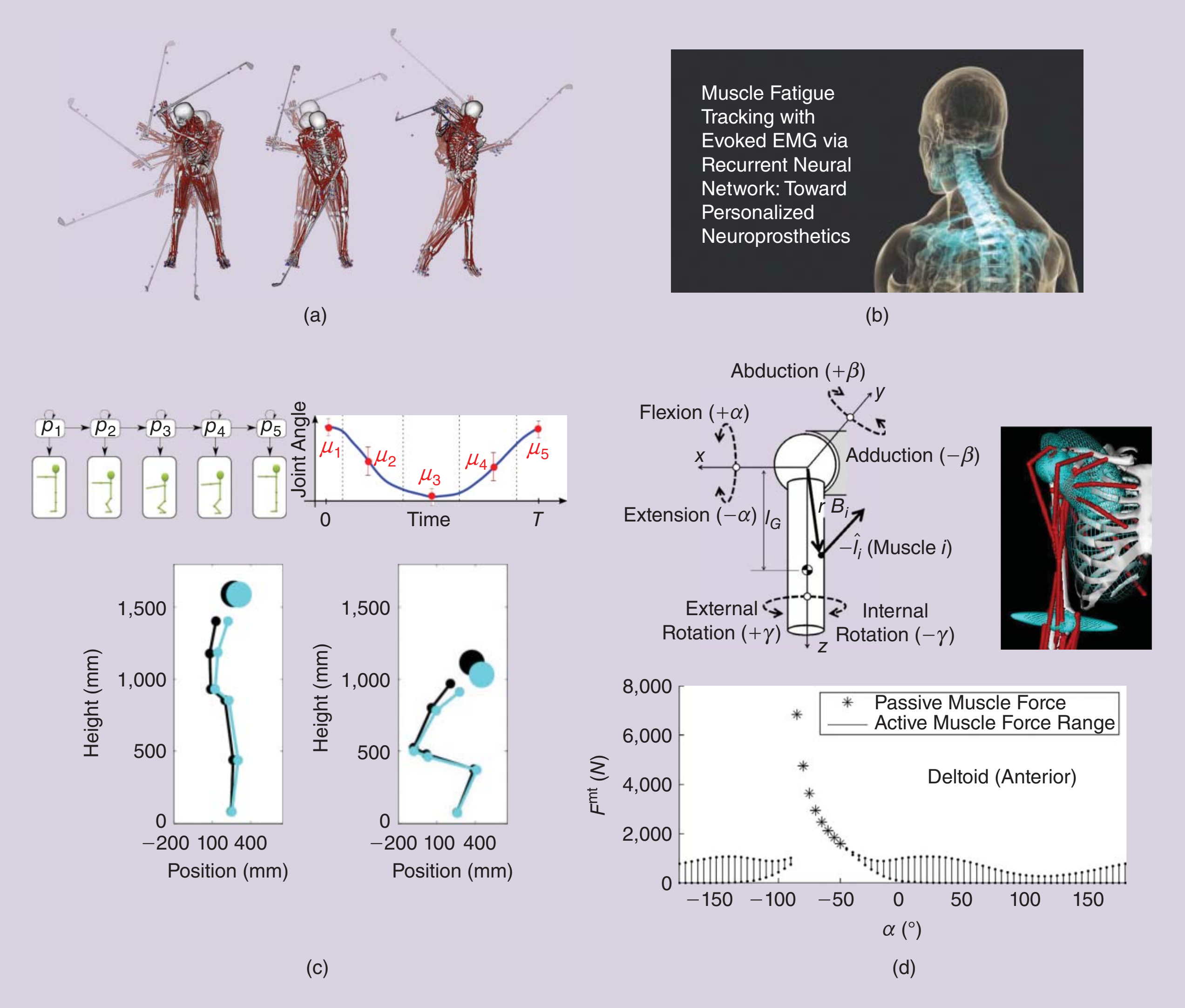


Figure 2. The relevant application fields of the TC on Human Motion Understanding: (a) human motion analysis and biomechanics [1], [2], (b) neuroprosthetics and neurorehabilitation [3], (c) human motion recognition and evaluation [4], and (d) human motion characterization and workspace analysis [5].

(Asia and Oceania), and Mitsuhiro Hayashibe (Europe and Japan).

The priority areas for the TC include the following:

- The application of advanced computational tools to
  - characterize natural human motion and the higher-level strategies used to realize complex tasks and interact with the external environment
  - develop tools for characterizing changes in human motion due to disease, aging, or injury to facilitate applications in rehabilitation and neuroprosthetics and exoskeleton design
  - predict behavior and synthesize humanlike motions.
- The development of strategies for human motion reconstruction on engineered anthropomorphic systems, such as humanoid, modular manipulators, and simulated systems.
- Human motion generation and task learning, including but not limited to the strategies for generalization of learned tasks to the learning of new tasks, resolution of human motor redundancy, and human strategies in handling constraints.

In the next ten years, many scientific challenges will be addressed by the TC on Human Movement Understanding.

Related Activities Organized by the TC

Since the establishment of the TC, a broad range of activities has been initiated in the last few years:

- ICRA 2013—full-day workshop on Computational Techniques in Natural Motion Analysis and Reconstruction, 6 May 2013, Karlsruhe, Germany
- IROS 2013—full-day tutorial on Robotics-Based Methods for the Identification, Recognition, and Synthesis of Human Motions, 3 November 2013, Tokyo, Japan
- ICRA 2014—full-day workshop on Latest Advances on Natural Motion Understanding and Human Motion Synthesis, 31 May 2014, Hong Kong, China
- Humanoids 2014—half-day workshop on Human Motion Modeling and Human-Inspired Motor Control, 18 November 2014, Madrid, Spain
- ICRA 2015—half-day workshop on Human Movement Understanding and Neuromechanics, 26 May 2015, Seattle, Washington.

These upcoming activities are scheduled for this year and 2016:

- a special issue on movement science for humans and humanoids: methods and applications in *IEEE Transactions on Robotics* (target publication date: February or April 2016)
- a summer school on human movement understanding, supported by RAS at Stanford University in 2016 and held jointly with SimTK Opensim group.

Future Directions

In the next ten years, many scientific challenges will be addressed by the TC on Human Movement Understanding. We aim at posing questions to advance the research for understanding natural human movement using robotics research—providing a computational basis for the analysis of movement disorders and performance improvement, the development of novel tracking and identification methods on physiological signals, kinematic, and dynamic modeling of human musculoskeletal systems, the development of strategies for human motion reconstruction on engineered anthropomorphic systems, and synthesizing and predicting human natural motions—having benefits in motor control and learning, ergonomics, biomechanics, physical therapy, neuroscience, sports medicine, and rehabilitation.

To promote the development and application of robotics methodologies and tools for the modeling, simulation, and synthesis of human motion, establish a network of expert researchers in robotics and neurophysiology, and encourage junior researchers in the area, the TC has formed several communication channels. The main one is a public website and its associated mailing list. We welcome new members to join us. Anyone can subscribe to the mailing list by visiting https://sites.google.com/site/ieeehmu/.

References

[1] O. Khatib, E. Demircan, V. DeSapio, L. Sentis, T. Besier, and S. Delp, "Robotics-based synthesis of human motion," *J. Physiol.*, vol. 103, nos. 3–5, pp. 211–219, 2009.

[2] E. Demircan, "Robotics-based reconstruction and synthesis of human motion," Ph.D. dissertation, Dept. Mechanical Eng., Stanford Univ., Stanford, CA, Aug. 2012.

[3] Z. Li, M. Hayashibe, C. Fattal, and D. Guiraud, "Muscle fatigue tracking with evoked EMG via recurrent neural network: Toward personalized neuroprosthetics," *IEEE Comput. Intell. Mag.*, vol. 9, no. 2, pp. 38–46, 2014.

[4] M. Karg, G. Venture, J. Hoey, and D. Kulić, "Human movement analysis as a measure for fatigue: A hidden markov-based approach," *IEEE Trans. Neural Syst. Rehab. Eng.*, vol. 22, no. 3, pp. 470–481, 2014.

[5] D. Lau, J. Eden, D. Oetomo, and S. Halgamuge, "Musculoskeletal static workspace analysis of the human shoulder as a cable-driven robot," *IEEE/ASME Trans. Mechatron.*, vol. 20, no. 2, pp. 978–984, 2014.

Networked Robots

By Volkan Isler, Brian Sadler, Libor Přeučil, and Shuichi Nishio

Many robotics applications require robots to communicate with each other, with a human operator, or with a remote server. The IEEE Robotics and Automation Society's Technical Committee (TC) on Networked Robots focuses on research issues related to robots connected to a communication network. The TC was founded in 2001. The name "Networked Robots" was adopted in 2004. The original focus of the TC was on Internet-based teleoperated robots. At the moment, most of the research activity in our TC revolves around robotic networks and cloud robotics.

Research Activities

Robotic Networks

In multirobot systems, robots must coordinate their actions either by communicating with each other or with a central server. Research topics include the development of algorithms and systems for connectivity maintenance, data gathering, robotic routers, network deployment, repair and maintenance, and teleoperation. Figures 1–8 show examples of ongoing research topics in this area.

Cloud Robotics

As cloud computing is changing the computing landscape, the benefits of connecting robots to the cloud is becoming evident. According to Ken Goldberg, there are at least four potential advantages to connecting robots to the cloud:

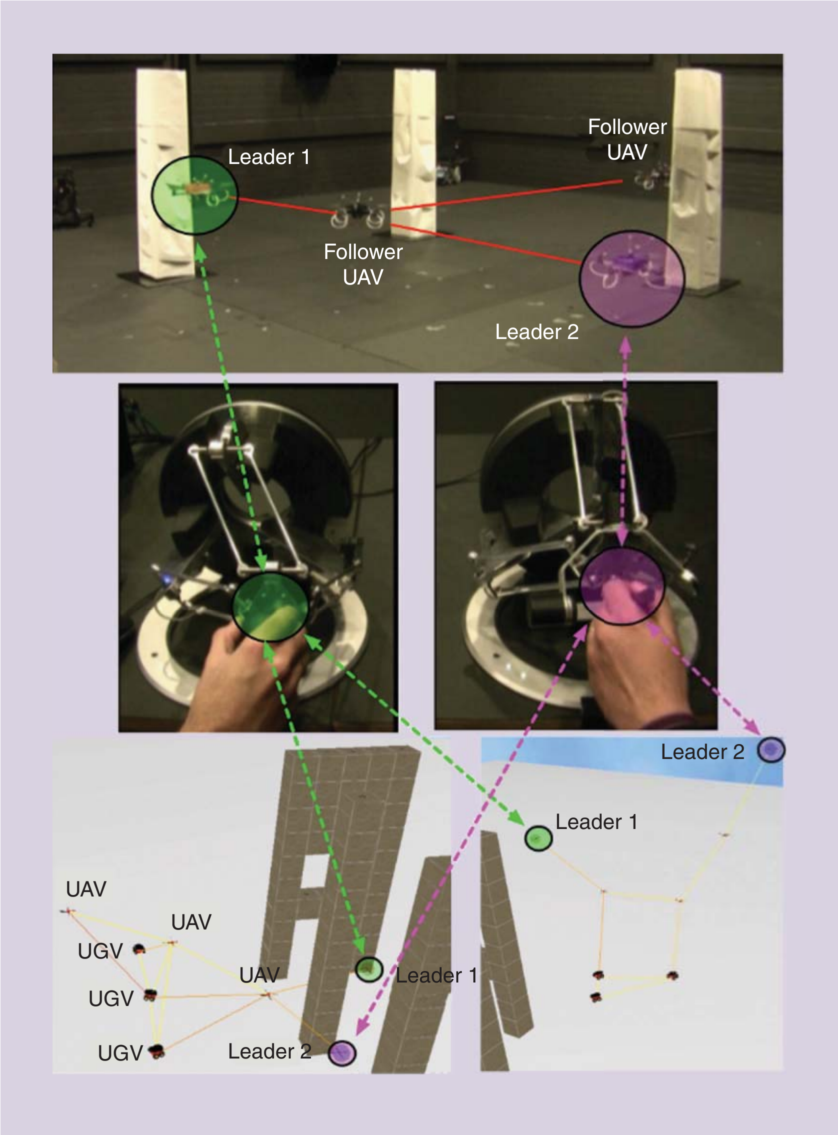
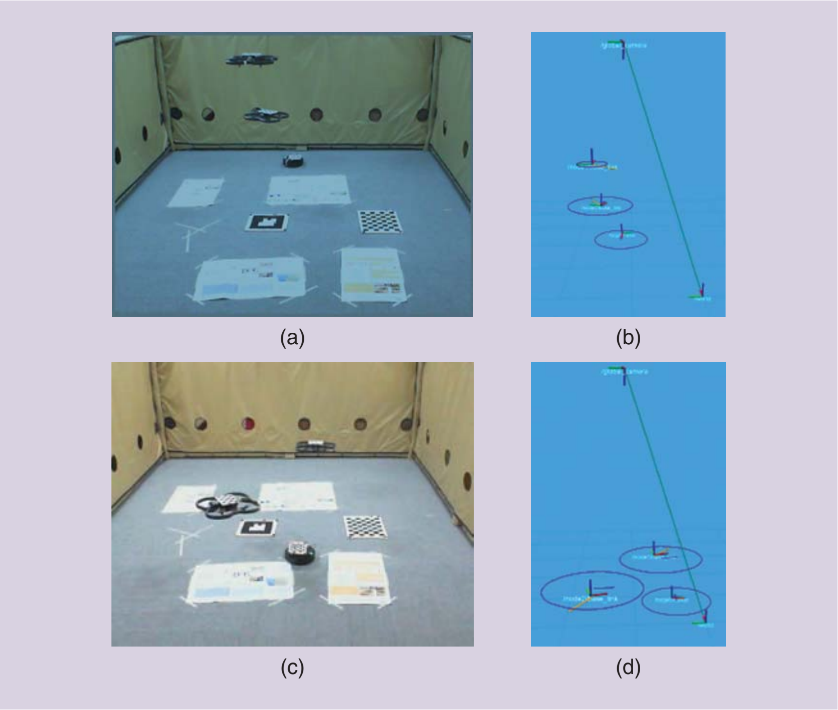


Figure 1. Robuffo Giordano et al. [8] studied how a group of quadrotor unmanned aerial vehicles (UAVs) can preserve group connectivity in a flexible and decentralized way while navigating in a cluttered environment. In particular, gain/loss of pairwise links (due to sensing/communication constraints) is allowed as long as the interaction graph remains connected. A shared control architecture is also built on top of this "connectivity maintenance" framework, allowing one or more human operators to easily steer the gross motion of the robot group. UGV: unmanned ground vehicle. (Photo courtesy of Dr. Paolo Robuffo Giordano and Dr. Antonio Franchi.)

Digital Object Identifier 10.1109/MRA.2015.2452172
Date of publication: 11 September 2015

Figure 2. The Advanced Robotic Systems Engineering Laboratory (ARSENL) team from the Naval Postgraduate School was able to successfully launch, fly, and land 20 UAVs autonomously at Camp Roberts, California, on 15 May 2015, which it deployed in two subswarms (ten UAVs each) and operated using the ARSENL-developed swarm operator interfaces. The UAVs were able to perform basic leader-following cooperative behaviors, exchanging information among themselves via wireless links. This is a first-of-its-kind demonstration of this magnitude for large-scale, autonomous fixed-wing UAV teams [6]. (Photo courtesy of ARSENL.)



(a) (b)

(c) (d)

Figure 3. Researchers at the University of Southern California considered the cooperative control of a team of robots to estimate the position of a moving target using onboard sensing. The robots were required to estimate their positions using relative onboard sensing while concurrently tracking the target. They developed a probabilistic localization and control method, taking into account the motion and sensing capabilities of the individual robots to minimize the expected future uncertainty of the target position. A highlight of the approach is that it reasons about multiple possible sensing topologies and incorporates an efficient topology switching technique to generate locally optimal controls in polynomial time complexity [5]. (Photos courtesy of University of Southern California Robotics Embedded Systems Laboratory and Automatic Coordination of Teams Laboratory.)

- *big data*—access to updated libraries of images, maps, and object/product data
- *cloud computing*—access to parallel grid computing on demand for statistical analysis, learning, and motion planning
- *collective learning*—robots and systems sharing trajectories, control policies, and outcomes
- *human computation*—use of crowdsourcing to tap human skills for analyzing images and video, classification, learning, and error recovery.

The cloud can also provide access to 1) data sets, publications, models, benchmarks, and simulation tools; 2) open competitions for designs and systems; and 3) open-source software. Cloud robotics and automation raise critical new questions related to network latency, quality of service, privacy, and security. Extending earlier work that links robots to the Internet, cloud robotics and automation build on emerging research in the cloud computing, machine learning, big data, open-source software, and major industry initiatives such as the Internet of Things, smarter planet, industrial Internet, and industry 4.0.

Some of our members' activities related to cloud robotics are shown in Figures 9–11.

Other TC Activities

The TC maintains a research blog (networked-robots.cs.umn.edu), a LinkedIn group, and a mailing list. TC members were actively involved in the organization of a number of high-profile workshops. These included

- the National Science Foundation/Army Research Laboratory (NSF/ARL) Workshop on Cloud Robotics and Real-Time Big Data (http://cloud-robotics.cs.umn.edu)
- the Workshop on the Algorithmic Foundations on Robotics (WAFR) 2014, which featured many networked robots papers and plenaries by Oussama Khatib, Vijay Kumar, and Cagatay Basdogan focusing on various networking and communications aspects of robotics
- the Networked Robots Workshop at the 2015 European Robotics Forum
- numerous workshops at the IEEE International Conference on Robotics and

Automation and the IEEE/Robotics Society of Japan International Conference on Intelligent Robots and Systems.

Conclusion

Research in networked robotics continues to be extremely active. Our TC has been dynamically adjusting its focus to support and promote research in the area. Furthermore, networking plays a crucial role in multirobot systems, agricultural and marine robotics, and smart environments. Our members closely collaborate with associated technical committees in these areas. As the roles of big data, communication, and coordination become more prominent in robotics, we expect networking to maintain a crucial role.

Acknowledgments

Special thanks to Ken Goldberg for providing most of the text on cloud computing. Volkan Isler also thanks

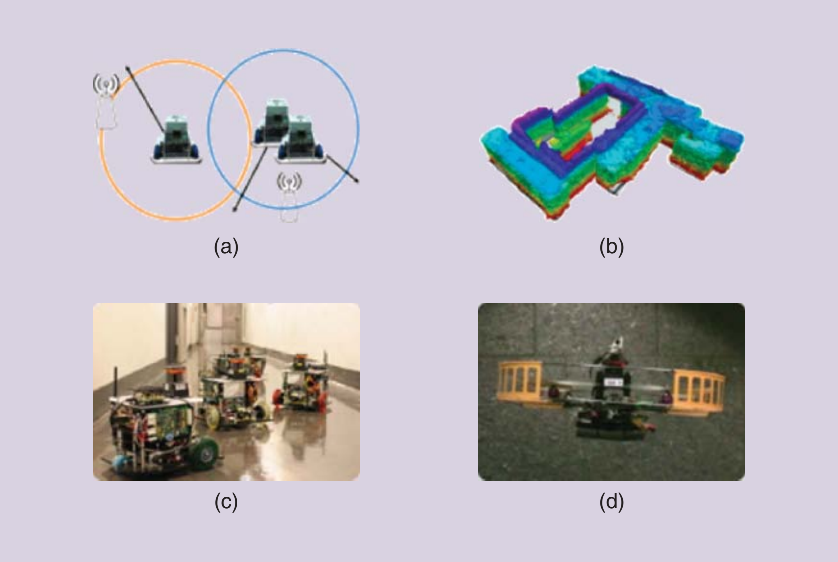


Figure 4. Researchers at the GRASP Laboratory at the University of Pennsylvania are developing swarms that can autonomously navigate and explore unknown and potentially hostile environments, without a designated leader, with limited communications between its members, and/or with different and potentially dynamically changing roles for its members. [Photos courtesy of Ben Charrow (student), Nathan Michael (coadvisor), and Vijay Kumar (coadvisor). See http://kumar.grasp.upenn.edu.]

Figure 5. In a networked system, there are a few key issues that need to be considered for data sharing and processing. These issues are phrased in the form of the following four questions. First, what type of data is shared among robots? Second, how are these data shared among robots? Third, where are these data processed? Finally, how is the processing performed? Howard Li and his group investigated these issues of data communication, data sharing, data distribution, and data processing [3].

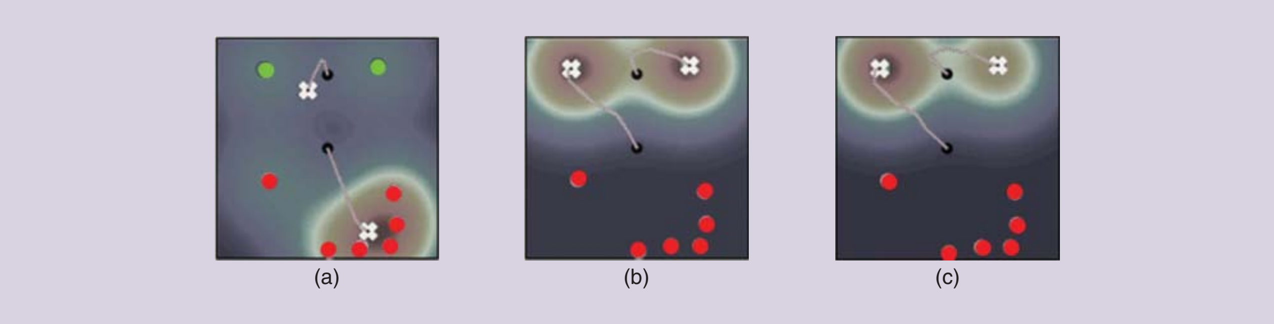


Figure 6. The promise of multirobot networks centers on the assumption that communication among robots is accurate and trustworthy. Even a simple-to-execute cyberattack, where, through spawning false identities, a single malicious robot gains disproportionate influence in the network, can be crippling to the entire system. Researchers at MIT's CSAIL Lab have developed a new and effective method of defense against these cyberattacks that fingerprints all transmitted wireless signals so that malicious robots can be automatically detected and quarantined, with no extra hardware or complex cryptographic schemes necessary. The figure shows their solution applied to a coverage problem where two quadrotor robots (shown as white Xs) must provide coverage to legitimate clients (green dots) in the presence of many falsely created clients (red dots) [7]. (a) No security, (b) Oracle, and (c) our system. (Photo courtesy of Daniela Rus.)

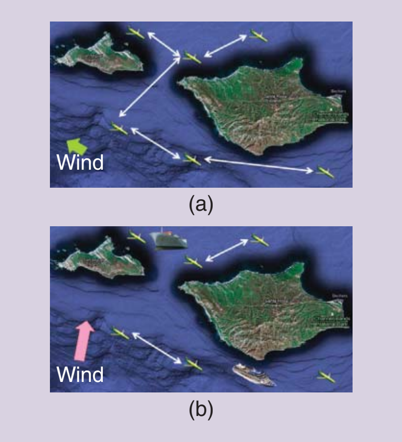


Figure 7. Underwater multirobot planning requires fully distributed algorithms capable of operating at any level of communication. Calm wind and low shipping activity results in (a) a connected network of underwater vehicles, but high wind and high shipping activity disconnects the same network of vehicles, as shown in (b) [2]. (Photos courtesy of G. Hollinger et al. and University of Southern California Robotic Embedded Systems Laboratory.)

Daniela Rus, Paolo Giordano, Antonio Franchi, Joo-Ho Lee, Howard Li, Geoff Hollinger, Gaurav Sukhatme, Tim Chung, and Vijay Kumar for providing research highlights.

References

[1] D. Yakushin and J.-H. Lee, "Cooperative robot software development through the internet," in *Proc. IEEE/SICE Int. Symp. System Integration (SII)*, Dec. 13–15, 2014, pp. 577–582.

[2] G. Hollinger, S. Yerramalli, S. Singh, U. Mitra, and G. Sukhatme, "Distributed data fusion for multirobot search," *IEEE Trans. Robot.*, vol. 31, no. 1, pp. 55–66, Feb. 2015.

Figure 8. Researchers at the Robotic Sensor Network Laboratory at the University of Minnesota have built a network of autonomous surface vehicles that can collaboratively find and track radio-tagged invasive fish. The tracking algorithm presented by [4] can break and reestablish communication while achieving desired levels of tracking performance. (Photo courtesy of RSN Lab, University of Minnesota.)

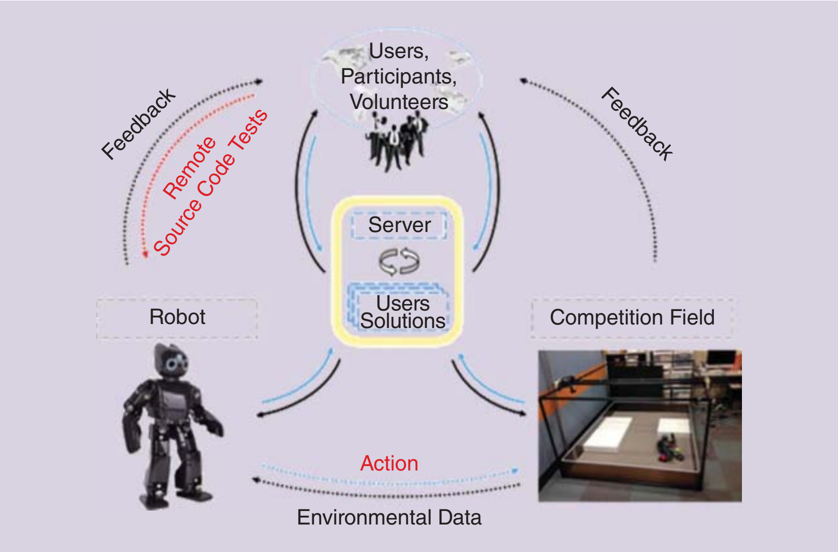


Figure 9. In their project League of Everybody, Joo-Ho Lee and his group are building a cooperative–competitive robot development environment through the Internet [1].

Figure 10. The April 2015 issue of *IEEE Transactions on Automation Science and Engineering* includes a special section with 11 papers on the emerging area of cloud robotics and automation describing system architectures and applications to navigation, disaster-response, grasping, and assembly. (Photos courtesy of Ken Goldberg.)

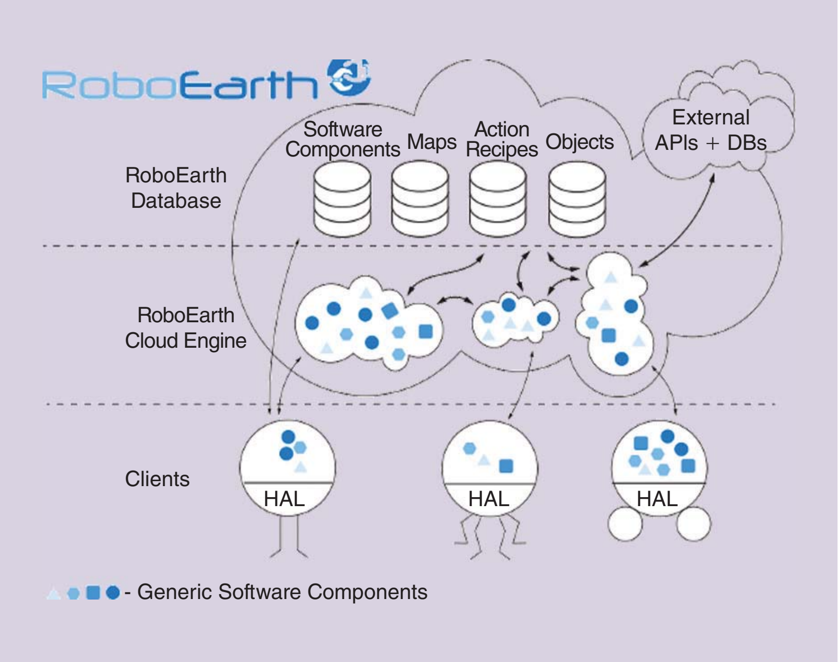


Figure 11. The goal of the RoboEarth project was to offer a cloud robotics infrastructure, which includes everything needed to close the loop from robot to the cloud and back to the robot. RoboEarth's World-Wide-Web style database stores knowledge generated by humans—and robots—in a machine-readable format. APIs: application program interfaces; HAL: hardware abstraction layer. (Photo courtesy of http://roboearth.org.)

[3] A. Nagaty, C. Thibault, M. Trentini, and H. Li, "Probabilistic cooperative target localization," *IEEE Trans. Automat. Sci. Eng.*, vol. 12, no. 3, pp. 786–794, July 2015.

[4] J. V. Hook, P. Tokekar, and V. Isler, "Algorithms for cooperative active localization of static targets with mobile bearing sensors under communication constraints transactions on robotics," *IEEE Trans. Robot.*, 2015, to be published.

[5] K. Hausman, J. Mueller, A. Hariharan, N. Ayanian, and G. S. Sukhatme, "Cooperative control for target tracking with onboard sensing," in *Proc. Int. Symp. Experimental Robotics*, June 2014.

[6] T. H. Chung, K. D. Jones, M. A. Day, M. Jones, and M. R. Clement, "50 VS. 50 by 2015: Swarm Vs. Swarm UAV live-fly competition at the naval postgraduate school," in *Proc. AUVSI North America*, 2013.

[7] S. Gil, S. Kumar, D. Katabi, and D. Rus, "Guaranteeing spoof-resilient multi-robot networks," *Robot. Sci. Syst.*, July 2015.

[8] P. R. Giordano, A. Franchi, C. Secchi, and H. H. Bülthoff, "A passivity-based decentralized strategy for generalized connectivity maintenance," *Int. J. Robot. Res.*, vol. 32, no. 3, pp. 299–323, 2013.

Biorobotics with Hybrid and Multimodal Locomotion

By Kin Huat Low, Samer Mohammed, Tianjiang Hu, Justin Seipel, Ravi Vaidyanathan, and Jorge Solis

The IEEE Robotics and Automation Society (RAS) Technical Committee (TC) on Biorobotics was formed with the goal of providing a forum and dissemination mechanism for the interaction between biological and artificial (autonomous or semiautonomous) systems and to present biology as a learning tool for novel engineering paradigms. In 2007, in an effort to define the scope of the technical field of the TC, the TC on Biorobotics was created to focus on various research fields that involve the understanding and implementation of complex living organisms by virtue of mechatronic systems. One such field is biologically inspired robotics, which is characterized by a multidisciplinary approach that aims to strengthen the collaboration between roboticists and biologists. To this end, two main approaches are adopted for research in Biorobotics: 1) the application of biological concepts/methodologies to improve the current capabilities of robots, for example, extending the robot's flexibility and robustness by adopting design principles of biological systems and 2) the application of advanced robotic technology to improve the current techniques/methodologies adopted by biologists. The principles and techniques of hybrid and multimodal locomotion recently have been emphasized and explored by researchers for better performance of bioinspired movements.

Highlights of Recent Activities

TC-Organized Workshop at IROS 2013

The IEEE RAS TC on Biorobotics organized the workshop "Biologically Inspired Based Strategies for Hybrid and Multimodal Locomotion" in conjunction with the IEEE International Conference on Intelligent Robots and Systems (IROS), held at Tokyo Big Sight, Japan, on 3–7 November 2013 (Figures 1 and 2).

Biologically based concepts for hybrid and multimodal locomotion have revealed new challenges regarding mechanical design, sensor integration,

Digital Object Identifier 10.1109/MRA.2015.2452173
Date of publication: 11 September 2015

Figure 1. The key organizers with the speakers of the IROS 2013 (from left): James Tangorra (cochair), Mirko Kovac (cochair), Frédéric Boyer (keynote speaker), Auke Ijspeert (keynote speaker), David Lentink (keynote speaker), Metin Sitti (keynote speaker), and Samer Mohammed (cochair). (Photo courtesy of Prof. Auke Jan Ijspeert.)

control theory, robustness, adaptability, and so on. These challenges must be overcome if we are to significantly reduce the performance gap that exists between biological and robotic systems. In addition to low-power systems and portability, which are vital challenges that substantially limit any successful biorobotic-based application, the proposed paradigms should also consider issues related to scalability and security. Therefore, the main aim of this workshop was to deal with the challenges of applying biologically based concepts to improve the capabilities of robots with a specific focus on hybrid and multimodal locomotion in air, in water, and/or on land.

The workshop covered two sessions, with two keynote speakers and six invited speakers.

The IROS 2013 workshop comprised four sessions:

- multimodal ground robots
- multimodal locomotion dynamics
- swimming and amphibious robots
- multimodal aerial robots.

One keynote speaker was invited for each session. Metin Sitti from Carnegie Mellon University gave a speech "Jumping-Gliding Based Bioinspired Multilocomotion Robots." Frédéric Boyer from Ecole des Mines, Nantes, France, provided an overview of bioinspiration locomotion dynamics. Auke Ijspeert from EPFL presented his work on the multimodal locomotion in the salamander—from biology to robotics, while David Lentink from Stanford University delivered a speech "Unraveling the Biofluidynamics of Flight as an Inspiration for Design." Each session included three invited talks related to the session subject. In addition, selected young researchers and Ph.D. students were invited to present their works in the form of an interactive multimedia poster during the session breaks. For more information on the program, visit http://tc-BioRobotics.com.

Special Issues

Since 2007, the TC cochairs have explored every opportunity to enable researchers working on biorobotics to publish their updated results in special issues of journals that are of relevance. For example, a special issue of *Mechanism and Machine Theory* published in 2009 (vol. 44, no. 3) covers bioinspired mechanism engineering, a special issue of *Advanced Robotics* published in 2009 (vol. 23, no. 7–8) covers biomimetic robots, and a special issue of *Journal of Robotics and Mechatronics* published in 2012 (vol. 24, no. 4) presents focused areas and future trends of bioinspired robots such as analysis, control, and design.

The most recent special issues related to biorobotics cover bioinspired mechatronics (published in *IEEE/American Society of Mechanical Engineers (ASME) Transactions on Mechatronics*) and hybrid and multimodal locomotion (published in *Bioinspiration & Biomimetics*).

For *IEEE/ASME Transactions on Mechatronics*, around 100 papers were submitted to the focused section; however, only 20 papers could be accepted after a rigorous review. There were

Figure 2. The workshop participants at the session presented by Fumiya Iida (invited speaker) at the IROS 2013 workshop. (Photo courtesy of Fumiya Iida.)

Figure 3. The workshop participants together with K.H. Low (chair of the IEEE RAS TC on Biorobotics) at the inauguration of the TC Biorobotics SIG in China.

Figure 4. The founding group of the TC on Biorobotics China SIG together with K.H. Low (center): from left, Li Wen (cochair), Chunlin Zhou (cochair), Daibing Zhang (advisor), K.H. Low, Junzhi Yu (advisor), Shiwu Zhang (advisor), and Tianjiang Hu (chair).

many good papers that could not be accepted due to the page limits of the special issue. For a detailed survey on bioinspired mechatronics and the introduction of the special issue, please see the paper by Metin Sitti et al., "Survey and Introduction to the Focused Section on Bio-Inspired Mechatronics," published in *IEEE/ASME Transactions on Mechatronics* (vol. 18, no. 2, April 2013, pp. 409–418).

The presenters at the IROS 2013 workshop were invited to submit their work to *Bioinspiration & Biomimetics* for possible publication in a special issue on hybrid and multimodal locomotion. After more than a year of rigorous review, the following papers have been accepted and published in *Bioinspiration & Biomimetics* (vol. 10, no. 2, February 2015, doi: 10.1088/1748-3190/10/2/020301):

- "Folding In and Out: Passive Morphing in Flapping Wings," by Amanda K. Stowers and David Lentink
- "Distributed Flow Estimation and Closed-Loop Control of an Underwater Vehicle with a Multimodal Artificial Lateral Line," by Levi DeVries et al.
- "Quadrupedal Galloping Control for a Wide Range of Speed via Vertical Impulse Scaling," by Hae-Won Park and Sangbae Kim
- "Goal-Directed Multimodal Locomotion Through Coupling Between Mechanical and Attractor Selection Dynamics," by S.G. Nurzaman et al.
- "Running up a Wall: The Role and Challenges of Dynamic Climbing in Multimodal Applications," by Bruce D. Miller et al.
- "Performance Analysis of Jump-Gliding Locomotion for Miniature Robotics," by A. Vidyasagar et al.
- "Multibody Systems Dynamics for Bioinspired Locomotion: From Geometric Structures to Computational Aspects," by Frédéric Boyer and Mathieu Porez.

IEEE RAS TC on Biorobotics: Special Interest Group in China

A group of researchers in China who are interested and working in the relevant areas have established a special interest group (SIG) of the IEEE TC on Biorobotics. The inauguration of the SIG was successfully held in Chengdu, 7 September 2014. K.H. Low was invited to attend the inauguration of the China SIG on behalf of the IEEE TC on Biorobotics (Figure 3).

At the inauguration, Tianjiang Hu, Li Wen, and Chunlin Zhou were selected to serve as founding chair/cochairs (Figure 4), and Junzhi Yu, Daibing Zhang, and Shiwu Zhang were appointed as advisors of the TC SIG. In conjunction with the SIG inauguration, a workshop, "Current Research and Development of Bioinspired Robotics and Mechatronics," was held. The workshop covered two sessions, with two keynote speakers and six invited speakers. For the keynotes, Junzhi Yu gave a keynote speech "Bioinspired CPG-Based Swimming Control for Various Robots," while Li Wen presented his work, "Understanding Aquatic Propulsion Using Bio-Inspired Robotics and Multimaterial Prototypes."

The China SIG of the TC on Biorobotics is dedicated to building a platform to share information on academic activities, creating a link to the valuable resources from the IEEE RAS, and providing more networking opportunities for local members. To work more efficiently, the SIG will collaborate with committees of important domestic conferences, academic societies, and publishers in robotics, mechatronics, automation, and bionic engineering to help draw potential members from their participants. The team is starting to recruit new members, targeting professional researchers, engineers, and college students who are interested in biorobotics. The organization of student branches at colleges is also a focus of the China

The team is starting to recruit new members, targeting professional researchers, engineers, and college students who are interested in biorobotics.

(continued on page 181)

Toward Replicable and Measurable Robotics Research

By Fabio Bonsignorio and Angel P. del Pobil

The famous experiment by Galileo—one of the founders of modern science—in Pisa's Cathedral in 1582, was one of the very first examples of a scientific experiment validating a scientific result: the discovery of the pendulum law. Galileo measured the variations of the oscillation period of a lamp in the dome by his own heart rate. From those times, experiment replication and experiment replication and reproducibility of results are at the cornerstone of the scientific method. Yet in robotics, artificial intelligence, and automation, the reproduction of results from conference and journal papers, as they are today, is quite often very difficult, if not impossible. This situation is bad for science, as it becomes difficult to objectively evaluate the state of the art in a given field, and also it becomes problematic to build on other people's work, thus undermining one of the basic foundations of scientific progress.

Moreover, it is detrimental to the industrial exploitation of results, for which we need to compare the effectiveness and efficiency of different methods proposed to solve the same technical or scientific problem, for example, from the computational and energy-consumption standpoints. This difficulty in reproducing results, however, makes this comparison usually very cumbersome and without trustable outcomes. This situation hampers and slows down the industry take-up of research results, and there are many more than those already exploited that are likely to benefit our daily lives.

The community has been aware of this issue for a long time. In 2007, we, with John Hallam, created the European Robotics Research Network (EURON) Good Experimental Methodology (GEM) and Benchmarking Special Interest Group (SIG) within the EURON Network of Excellence (NoE), a NoE is a networking-oriented European-funded project. In 2006, one of us, Angel P. del Pobil, organized a workshop on benchmarking at the IEEE/Robotics Society of Japan International Conference on Intelligent Robots and Systems (IROS) in Beijing, China, as an activity of the EURON NoE work package devoted to benchmarking, and the first website on survey and inventory of current efforts in comparative robotics research was established (http://www.robot.uji.es/EURON/en/index.htm).

The GEM guidelines [1] were one of the major outputs of the SIG's early activities. Although, initially, the guidelines were focused on more careful reviews, mainly thanks to one of us (Fabio Bonsignorio), the real problem became clear: the core issue is the reproducibility/replicability of experimental results. In 2009, at the International Conference on Robotics and Automation (ICRA) in Kobe, Japan, the IEEE Robotics and Automation Society (RAS) Technical Committee on Performance Evaluation and Benchmarking of Robotic and Automation Systems was founded, with similar objectives. In parallel, the Performance Metrics for Intelligent Systems conference series focuses on performance measurement challenges arising from the application of robotics and automation technologies to practical problems in the commercial, industrial, homeland security, and military domains. More information can be found at http://www.nist.gov/el/isd/permis2012.cfm.

There has been a long series of workshops at various conferences, such as IROS, ICRA, and the Robotics Science and Systems Conference, in which more than 200 people have participated so far. We mostly organized them with Elena Messina and John Hallam, but there have also been some organized by others, and there have been a number of competitions and publications aiming at finding a way out of a situation that is considered by many as unsatisfying (see http://www.ieee-ras.org/performance-evaluation and http://www.heronrobots.com/EuronGEMSig/).

When EURON joined euRobotics Association Internationale Sans But Lucratif (AISBL), the private part of the European Public–Private Partnership on Robotics, the activities of the former EURON GEM SIG became part of the Topic Group on Evaluation and Assessment of Research Results, also known as Benchmarking and Competitions. A solid example of benchmarking methodology is proposed in "Benchmarking in Manipulation Research" by Berk Calli, Aaron Walsman, Arjun Singh, Siddahrta Srinivas, Peter Abbel, and Aaron Dollar. There are experimental setting where this approach is difficult to implement. In those cases competitions,

Digital Object Identifier 10.1109/MRA.2015.2452073
Date of publication: 11 September 2015

or, better, scenario-based evaluation procedures, have been recognized as a component of the recipe for the benchmarking of results, particularly when intelligent behaviors are involved. The extent to which competitions can be regarded as scientific experiments, and which ones, is still a matter of discussion. An article in this issue, "Competitions for Benchmarking," by Francesco Amigoni, Emanuele Bastianelli, Jakob Berghofer, Andrea Bonarini, Giulio Fontana, Nico Hochgeschwender, Luca Iocchi, Gerhard K. Kraetzschmar, Pedro Lima, Matteo Matteucci, Pedro Miraldo, Daniele Nardi, and Viola Schiaffonati, may provide some hints.

Methodological, Practical, and Epistemological Issues

Although the number of robotics papers published in journals and conferences is constantly growing, the possibility of reproducing results is left to the good will of some authors. The number and nature of envisioned applications and proposed methods are vast and also steadily increasing. As a consequence, some members of the community believe that the comparison of results would not be practically possible. A remarkably varied set of robotic applications is approached by a significantly disparate set of methods, sometimes based on notably different principles, with different hardware (HW)/software (SW) architectures in different environments. On the one hand, the explosive growth of research results shows that the community is becoming larger and increasingly active; on the other hand, it raises some serious problems when you have to objectively evaluate the actual relevance of the results and the actual state of the art in a given field.

As previously stated, the difficulty of reproducing results—let alone comparing different methods and solutions—slows down the industrial take-up of new solutions. Basic research is also hindered, since it is very difficult for a research group to build on the results of another one, leading to a very limited cross-exploitation of results between different groups, and a general prevalence of exploration over exploitation. Many new solutions are proposed, but the community often does not go deep into the analysis of most of them.

The EURON GEM guidelines [1] are essentially an adaptation to the robotics and automation domain of the general methodology applied in science and engineering that was pioneered by Galileo and Boyle. Today, as discussed in [2], only a limited subset of published results follow those methods and usually not completely. Of course, not every paper should follow a rigorous experimental protocol: position papers, concept papers describing upcoming research, papers concerning algorithms, or survey papers do not need to comply with a rigorous and epistemologically sound experimental methodology. Still, many papers that claim to have solved a problem (say, autonomous driving) based on simulations or field experiments should comply. Robotics, artificial intelligence, and automation are not pure mathematics. The proposed solutions need to be able to work in the set of environments and for the set of tasks for which they have been studied. There are scientific aspects in robotics, for example, related to the unbundling of the brain-body nexus in humans and animals, but even when we are closer to pure engineering applications, experimental proofs of the effectiveness of the proposed solutions are needed.

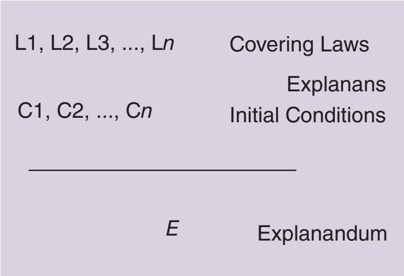


Figure 1. The Hempel–Oppenheim model of scientific knowledge. In the conceptual schema represented in this figure, which summarizes the Hempel–Oppenheim model of scientific knowledge, all the logical enunciates have a probabilistic truth value. We need a precise and complete list of laws invoked for the explanation, a precise and complete list of initial conditions (system HW/SW architectures, environments, tasks), a precise definition of what is explained or proved. In addition, we must accept the fact that our theoretical claims, enunciate, have to be of probabilistic nature, since we operate in open-ended stochastic environments. (Figure adapted from G. Boniolo, "A Contextualized Approach to Biological Explanation," *Philosophy,* vol. 80, pp. 219–247, 2005).

We should at least be able to

- validate the results by replicating them
- compare the results in terms of the chosen performance criteria.

This holds true for both purely scientific issues and real-world applications. The fact that robotics research deals with very diversified problems should not be seen as a serious obstacle. Indeed, medicine and life science, for instance, where the complexity and variety of the studied objects are not smaller than in robotics, have developed rigorous experimental protocols. We should take inspiration from them. An epistemological model of biological science was proposed by Hempel and Oppenheim; see Figure 1.

We should not be so surprised by the fact that we are struggling to define valid and shared benchmarking procedures for intelligent robots.

We can expect that having replicable and measurable results will affect the content of the results, not just their reporting. We should not be so surprised by the fact that we are struggling to define valid and shared benchmarking procedures for intelligent robots. Their development uncovers a lot of practical, publishing, and also epistemological issues. A more detailed discussion of this topic can be found in [7] and will be the main topic of a future publication. Besides the so-far unsatisfactory, in this respect, experimental and reporting practice, an important reason could be the limited scientific understanding of intelligence and cognition in natural and artificial systems. The practical issues span from modeling, to statistical significance assessment, to the mechatronic design and construction of specific test equipment, and to the actual replication procedures, the experimental protocols, and the necessity to provide the data, time series, and HW/SW description. The epistemological issues, with respect to paradigm examples of science, like physics, are many: multilevel causality, the large number of preconditions and laws involved, and probabilistic relations between causes and effects. Robotics has many problems in common with biology and medicine. We have comparison and evaluation criteria for cars and many other machines and appliances; we are just starting to develop those for robotics and intelligent systems. The articles in this issue show that it is possible, and we already have some promising proposals. After several years of discussions and attempts presented in a long series of workshops and elsewhere, a new kind of replicable paper in robotics has become mature.

This situation is bad for science, as it becomes difficult to objectively evaluate the state of the art in a given field.

State of the Art?

There has been a growing awareness about these issues in the community. Yet, it is still very difficult to find examples of replicable papers in robotics and automation. It is now possible to attach supplemental materials to papers in the most important journals of the field. Increasingly, authors share data sets and code, in particular, in the simultaneous localization and mapping (SLAM) community, and shared data sets and libraries, like Peter Corke's MATLAB libraries, are made available. But despite the progress in defining replication protocols, we are, in this respect, at the very beginning. Years ago, Amigoni et al. [8] showed that not a single paper among the top cited ones in SLAM and navigation met all the basic criteria listed in the GEM guidelines. We may have clearly improved since then, but probably not enough.

Competitions have also matured in the direction of becoming experiments on the most elusive intelligent behaviors. You will not find the real state of the art here, either in this editorial or in this issue, as far as replication of results is concerned. The reason is straightforward: this issue is the first example of a publication including a list of replicable research results. To a certain extent, the state of the art coincides with this issue of *IEEE Robotics and Automation Magazine* (*RAM*).

It is also interesting to note that, in more established areas of research with more mature experimental methodology, like clinical research, there have recently been serious concerns about the replicability of published research and the consequent negative impact on research and even new drug development and health care [3]–[5]. The idea that the publishing process should evolve is widespread. Published research reporting should provide enough information to allow the replication of the results. The web, and the easier distribution of information that the web makes possible, might be part of the solution.

On the one hand, this new possibility was identified several decades ago [6]. On the other hand, the practice of sharing research is already evolving, as shown by the success of preprint e-publishing platforms like arxiv (www.arxiv.org) or some recent experiments of open review on the web (see http://openreview.informatik.uni-freiburg.de) as well as by this special issue.

Contribution of this Issue

After many discussions and attempts, a new kind of paper seems to be necessary. This new kind of paper should include the following:

- *description*—a journal paper with text, figures, and multimedia, according to GEM guidelines (or similar)
- *data sets*—similar to the option provided by various journals and magazines, included this one
- *code identifiers*—complete code identifiers and/or downloadable code (executable files may be enough)
- *HW identifiers*—HW description or HW identifier (if it is identifiable).

This special issue of *RAM* is the very first example of a collection of replicable robotics reports covering a remarkably wide area of diverse robotics subfields. The articles in this issue provide a living example of the viability of replicable research in robotics. They span a wide and diverse set of areas of research in robotics, thus countering the idea that this field is too diversified to allow a rigorous shared methodology.

The articles report replicable experiments, benchmarking methods, and a couple of exemplary surveys on competitions ("Humanoid Robots in Soccer," by Reinhard Gerndt, Daniel Seifert, Jacky Baltes, Soroush Sadeghnejad, and Sven Behnke) and on the new and important field of soft robotics ("Deformation in Soft-Matter Robotics," by Liyu Wang and Fumiya Iida). We have a very interesting article about how competitions can be given a rigorous scientific meaning in the ("Competitions for Benchmarking," by Francesco Amigoni, Emanuele Bastianelli, Jakob Berghofer, Andrea Bonarini, Giulio Fontana, Nico Hochgeschwender, Luca Iocchi, Gerhard K. Kraetzschmar, Pedro Lima, Matteo Matteucci, Pedro Miraldo, Daniele Nardi, and Viola Schiaffonati). The set of replicable research examples covers wearable systems ("Wearable Inertial Sensors," Barbara Bruno, Fulvio Mastrogiovanni, and Antonio Sgorbissa) and manipulation ("Benchmarking in Manipulation Research," by Berk Calli, Aaron Walsman, Arjun Singh, Siddhartha Srinivasa, Pieter Abbeel, and Aaron M. Dollar). We have three papers on different aspects of marine robotics ("Tracking Divers," by Nikola Mišković, Đula Nađ, and Ivor Rendulić; "Exploring 3-D Reconstruction Techniques," by Javier Pérez, Jorge Sales, Antonio Peñalver, David Fornas, José Javier Fernández, Juan Carlos García Sánchez, Pedro J. Sanz, Raúl Marín, and Mario Prats; and "Testing the Waters," by Andrea Sorbara, Andrea Ranieri, Eleonora Saggini, Enrica Zereik, Marco Bibuli, Gabriele Bruzzone, Eva Riccomagno, and Massimo Caccia). And then we cover motion planning ("Benchmarking Motion Planning Algorithms," by Mark Moll, Ioan A. Şucan, and Lydia E. Kavraki), bipedal locomotion ("Benchmarking Bipedal Locomotion," by Diego Torricelli, Jose Gonzalez, Jan Veneman, Katja Mombaur, Nikos Tsagarakis, Antonio J. Del-Ama, Angel Gil-Agudo, Juan C. Moreno, and Jose L. Pons), and last but not least the requirements for replicable simulation experiments ("RoboCup Simulation Leagues," by David M. Budden, Peter Wang, Oliver Obst, and Mikhail Prokopenko).

You should read the articles from various standpoints: the novelty of the content, the significance and viability of the proposed benchmarks, the approaches that the authors have chosen to allow the replication of their results. The first question to ask is: are these results reproducible? You will notice that the articles have different focuses and that the approaches are different. What are the strengths and weaknesses of the various approaches? Some authors, like Perez et al., seem more focused on the definition of the benchmarking criteria, others, like Moll et al. or Sorbara et al. on the replicability of the benchmarks. Some, like Bruno et al., rely on third-party repositories like github or sourceforge, some have designed a dedicated website. Some use an XML-based description, some do not. Are the experiment statistics always managed in the best way? How should the statistical significance of the experiments be evaluated and the related metrics reproduced? Have a look at Figure 1 in Moll et al.; to be able to replicate the experiments, we will need to structure systems like that. What is the best way to implement them? You may wish to compare with Figure 2 in Sorbara et al. (for example). This collection of very interesting articles inspires a long list of thought-provoking questions and provides many possible solutions and insights.

Of course, this is just a starting point. Hopefully, the practical replication of the results by the community will show the best ways to provide information to make the results of robotics papers reproducible.

Road Ahead

We will need to foster the proper attitudes toward replication of results in the community. We should not think that scientific publishing could not further evolve. Replicable papers can be a valuable addition to the current scientific publishing landscape. In this new context, the initial severe peer-review preceding the publication of papers will be just a prerequisite for the real peer-review based on the active reproduction of the published results by the community at large. This will also make easier the understanding of the still open scientific problems related to intelligent, animal-like, and cognitive behaviors.

Another thing to consider for the future is that the authors of the articles in this issue, while usually providing the information necessary through their own websites, also had to upload the data needed for replication to the magazine website as attachments to this article. We think that, in the future, we will need a more structured approach; in this sense, the website structures for this issue will also contribute to the definition of a new publishing set of conventions to present replicable papers, not as just attachments. This is what is available now, and it is useful to have a single self-contained entry for all the articles in the special issue.

We would like to see the results of many of the articles here reproduced as they are in other articles commenting on these issues and suggesting improvements. Although many challenges are still ahead, we believe we are heading in the right direction: back to the basics of the scientific method.

References

[1] F. Bonsignorio, J. Hallam, and A. P. del Pobil, Eds. (2008). GEM Guidelines. Euron GEM Sig Report. [Online]. Available: http://www.heronrobots.com/EuronGEMSig/

[2] F. Bonsignorio, A. P. del Pobil, and E. Messina, "Fostering progress in perfomance evaluation and benchmarking of robotic and automation systems [TC spotlight]," *IEEE Robot. Automat. Mag.*, vol. 21, no. 1, pp. 22–25, 2015.

[3] Challenges in Irreproducible Research. Nature Special. [Online]. Available: http://www.nature.com/nature/focus/reproducibility/

[4] *How Science Goes Wrong: Scientific Research has Changed the World.* Now it needs to change itself, The Economist, 2013.

[5] *Trouble at the Lab: Scientists Like to Think of Science as Self-Correcting.* To an alarming degree, it is not, The Economist, 2013.

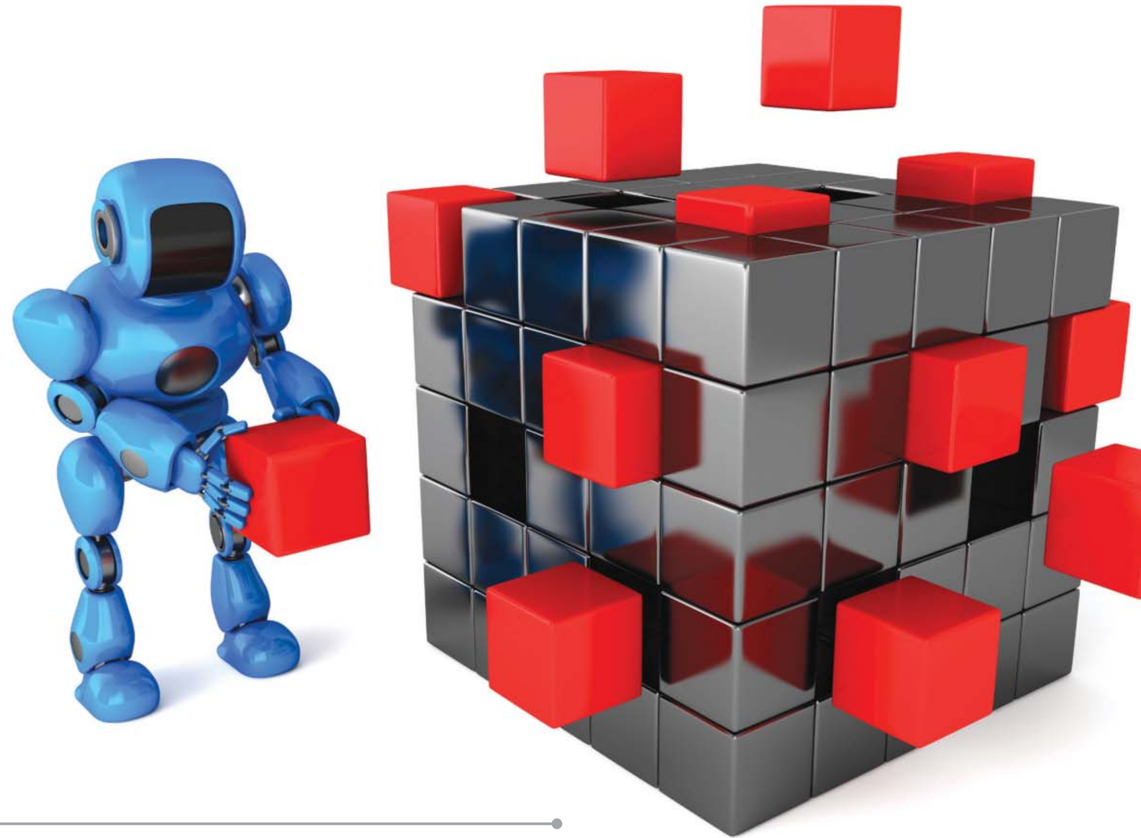
[6] J. Claerbout, "Electronic documents give reproducible research a new meaning," in *Proc. 62nd Annu. Int. Meeting: Society Exploration Geophysics*, 1992, vol. 92, pp. 601–604.

[7] F. Bonsignorio, J. Hallam, and A. P. del Pobil, "Defining the requisites of a replicable robotics experiment," in *Proc. RSS Workshop Good Experimental Methodologies Robotics*, 2009.

[8] F. Amigoni, M. Reggiani, and V. Schiaffonati, "An insightful comparison between experiments in mobile robotics and in science," *Auton. Robot.*, vol. 27, pp. 313–325, 2009.

Benchmarking in Manipulation Research

Using the Yale–CMU–Berkeley Object and Model Set



©ISTOCKPHOTO.COM/ALENGO

By Berk Calli, Aaron Walsman, Arjun Singh, Siddhartha Srinivasa, Pieter Abbeel, and Aaron M. Dollar

In this article, we present the Yale–Carnegie Mellon University (CMU)–Berkeley (YCB) object and model set, intended to be used to facilitate benchmarking in robotic manipulation research. The objects in the set are designed to cover a wide range of aspects of the manipulation problem. The set includes objects of daily life with different shapes, sizes, textures, weights, and rigidities as well as some widely used manipulation tests. The associated database provides high-resolution red, green, blue, plus depth (RGB-D) scans, physical properties, and geometric models of the objects for easy incorporation into manipulation and planning software platforms. In addition to describing the objects and models in the set along with how they were chosen and derived, we provide a framework and a number of example task protocols, laying out how the set can be used to quantitatively evaluate a range of manipulation approaches, including planning, learning, mechanical design, control, and many others. A comprehensive literature survey on the existing benchmarks and object data sets is also presented, and their scope and limitations are discussed. The YCB set will be freely distributed to research groups worldwide at a series of tutorials at robotics conferences. Subsequent sets will be, otherwise, available to purchase at a reasonable cost. It is our hope that the ready

Digital Object Identifier 10.1109/MRA.2015.2448951
Date of publication: 11 September 2015

1070-9932/15©2015IEEE

availability of this set along with the ground laid in terms of protocol templates will enable the community of manipulation researchers to more easily compare approaches as well as continually evolve standardized benchmarking tests and metrics as the field matures.

Benchmarking in Robotics Research

Benchmarks are crucial for the progress of a research field, allowing performance to be quantified to give insight into the effectiveness of an approach compared with alternative methods. In manipulation research, particularly in robotic manipulation, benchmarking and performance metrics are challenging due largely to the enormous breadth of the application and task space researchers are working toward. The majority of research groups have, therefore, selected for themselves a set of objects and/or tasks that they believe are representative of the functionality that they would like to achieve/assess. The chosen tasks are often not sufficiently specified or general enough such that others can repeat them. Moreover, the objects used may also be insufficiently specified and/or not available to other researchers (e.g., they may have been custom-fabricated or are only available for purchase in certain countries). Unfortunately, such an approach prevents the analysis of experimental results against a common basis and, therefore, makes it difficult to quantitatively interpret performance.

There have been a limited number of efforts to develop object and model sets for benchmarking in robotic manipulation. Most of these have focused on providing mesh model databases of objects, generally for object-recognition or grasp-planning purposes (see [1]–[4], with a thorough overview provided in the “Related Work” section). There have, however, been a few instances of proposed object/task sets for which the physical objects are available to researchers. Access to the objects is crucial to performance benchmarking as many aspects of the manipulation process cannot be modeled, thereby requiring experiments to demonstrate success or examine failure modes.

Overview

In this article, we present an object set for robotic manipulation research and performance evaluation, a framework for standard task protocols, and a number of example protocols along with experimental implementation. The object set is specifically designed to allow for widespread dissemination of the physical objects and manipulation scenarios. The objects were selected based on a survey of the most common objects utilized in the robotics field as well as the prosthetics and rehabilitation literature (in which procedures are developed to assess the manipulation capabilities of patients) along with a number of additional practical constraints. Along with the physical objects, textured mesh models, high-quality images, and point-cloud data of the objects are provided together with their physical properties (i.e., major dimensions and mass) to enable realistic simulations. These data are all available online at http://rll.eecs.berkeley.edu/ycb/. The models are integrated into the MoveIt motion-planning tool [5] and the robot operating system (ROS) to demonstrate their use. The set will be freely distributed to research groups worldwide at a series of tutorials at robotics conferences and will be, otherwise, available at a reasonable purchase cost. Our goal is to do as much as possible to facilitate the widespread usage of a common set of objects and tasks to allow easy comparison of results between research groups worldwide.

In choosing the objects in the set, a number of issues were considered. The objects, many of which are commercial household products, should span a variety of shapes, sizes, weights, rigidities, and textures as well as a wide range of manipulation applications and challenges. In addition, several practical constraints were considered, including ease of shipping and storage, reasonable overall cost, durability, perishability, and product longevity (the likelihood that the objects/products will be available in the future).

In addition to the object and model set, we provide a systematic approach to define manipulation protocols and benchmarks using the set. The protocols define the experimental setup for a given manipulation task and provide procedures to follow, and the benchmarks provide a scoring scheme for the quantification of performance for a given protocol. To facilitate the design of well-defined future protocols and benchmarks, guidelines are provided through a template. The protocols and benchmarks are intended to generally be platform-independent to allow for comparisons of approaches across platforms. Along with the template and guidelines, we present a number of preliminary protocols and benchmarks. These serve both as examples of how to utilize the template and as useful procedures for quantitatively evaluating various aspects of robotic manipulation. The implementation of these benchmarks on real robotic systems is also provided to demonstrate the benchmarks’ abilities to quantitatively evaluate the manipulation capabilities of various systems.

> **The models are integrated into the MoveIt motion-planning tool [5] and the robot operating system (ROS) to demonstrate their use.**

We expect to continually expand on this work not only by our own efforts (adding more objects’ properties and additional benchmarks) but also, more importantly, via our web portal: http://www.ycbbenchmarks.org/. Through this web portal, the user community can engage in this effort, with users proposing changes to the object set and putting forth their own protocols, benchmarks, and so on.

Related Work

For benchmarking in manipulation, specifying an object set is useful for the standardization of experimental conditions. Table 1 summarizes the object sets that have been proposed for manipulation tasks in the fields of robotics, prosthetics,

Table 1. Object data sets in the literature (sorted by year).

	Data Set Name	Year	Data Type	Purpose	Number of Objects/ Categories	Physical Objects Available	Website
1	BigBIRD [1]	2014	Meshes with texture + HQ images	Object recognition	100	No	http://rll.eecs.berkeley.edu/bigbird
2	Amazon Picking Challenge [7]	2014	Shopping list	Grasping	27	Yes	http://amazonpickingchallenge.org/
3	SHREC'14 [2]	2014	Mesh models	Object retrieval	8,987/171	No	http://www.itl.nist.gov/iad/vug/sharp/contest/2014/Generic3D/
4	SHREC'12 [21]	2012	Mesh models	Object retrieval	1,200/60	No	http://www.itl.nist.gov/iad/vug/sharp/contest/2012/Generic3D/
5	The KIT object models database [19]	2012	Mesh with texture, stereo images	Recognition, localization, and manipulation	100	No	http://i61p109.ira.uka.de/ObjectModels WebUI/
6	VisGraB [22]	2012	Stereo images	Manipulation	18	No	http://www.robwork.dk/visgrab/
7	The object segmentation database [17]	2012	RGB-D images	Object segmentation	N/A	No	http://users.acin.tuwien.ac.at/arichtsfeld/?site=4
8	Toyohashi shape benchmark [23]	2012	Mesh models	Object retrieval	10k/352	No	http://www.kde.cs.tut.ac.jp/benchmark/tsb/
9	The Willow Garage object recognition challenge [24]	2012	RGB-D images	Object recognition	N/A	No	http://www.acin.tuwien.ac.at/forschung/v4r/mitarbeiterprojekte/willow/
10	SHREC'11 [25]	2011	Mesh models	Object retrieval	600	No	http://www.itl.nist.gov/iad/vug/sharp/contest/2011/NonRigid/
11	Berkeley 3-D object data set [26]	2011	RGB-D data set of room scenes	Object detection	N/A	No	http://kinectdata.com/
12	RGB-D object data set [27]	2011	RGB-D Data set	Object detection and recognition	300/51	No	http://rgbd-dataset.cs.washington.edu/
13	The open GRASP benchmarking suite [20]	2011	Mesh with texture, stereo images	Grasping	Uses KIT database	No	http://opengrasp.sourceforge.net/benchmarks.html
14	SHREC 2010 [28]	2010	Mesh models	Object retrieval	3168/43	No	http://tosca.cs.technion.ac.il/book/shrec_robustness2010.html
15	The Columbia grasp database [3]	2009	Mesh models	Grasping	~8,000	No	http://grasping.cs.columbia.edu/
16	Benchmark set of domestic objects [6]	2009	Shopping list	Robotic manipulation	43	Yes	http://www.hsi.gatech.edu/hrl/object_list_v092008.shtml
17	Bonn architecture benchmark [29]	2009	Mesh models	Object retrieval	2,257	No	ftp://ftp.cg.cs.unibonn.de/pub/outgoing/ArchitectureBenchmark
18	Engineering shape benchmark [30]	2008	Mesh models	Object retrieval	867	No	https://engineering.purdue.edu/PRECISE/shrec08
19	3-D object retrieval benchmark [31]	2008	Mesh models	Object retrieval	800/40	No	http://www.itl.nist.gov/iad/vug/sharp/benchmark/
20	McGill 3-D shape benchmark [32]	2008	Mesh models	Shape retrieval	N/A	No	http://www.cim.mcgill.ca/~shape/benchMark/

(*Continued*)

Table 1. Object data sets in the literature (sorted by year). (*Continued*)

	Data Set Name	Year	Data Type	Purpose	Number of Objects/ Categories	Physical Objects Available	Website
21	The Toronto Rehabilitation Institute hand-function test [33]	2008	Commercial kit/no model data	Prosthetics and rehabilitation	14	No	http://www.rehabmeasures.org/Lists/RehabMeasures/PrintView.aspx?ID=1044
22	GRASSP [9]	2007	Commercial kit/no model data	Prosthetics and rehabilitation	N/A	Yes	http://grassptest.com/
23	AIM@SHAPE shape repository [16]	2006	Mesh models	General	1,180	No	http://shapes.aim-atshape.net/viewmodels.php
24	The Princeton shape benchmark [18]	2004	Mesh models	Shape-based retrieval	1,814	No	http://shape.cs.princeton.edu/benchmark/
25	Mesh deformation data set [34]	2004	Mesh models	Mesh transformation	N/A/13	No	http://people.csail.mit.edu/sumner/research/deftransfer/data.html
26	NTU 3-D model benchmark [35]	2003	Mesh models	Shape retrieval	1,833	No	http://3d.csie.ntu.edu.tw/
27	SHAP [8]	2002	Commercial kit/no model data	Prosthetics and rehabilitation	—	Yes	http://www.shap.ecs.soton.ac.uk/
28	Action research arm test [10]	1981	Commercial kit/no model data	Prosthetics and rehabilitation	19	Yes	http://saliarehab.com/actionresearcharmtestarat.html
29	Jebsen–Taylor hand-function test [11]	1969	Commercial kit/no model data	Prosthetics and rehabilitation	N/A	Yes	N/A
30	The ITI database [36]	N/A	Mesh models	Object retrieval	544/13	No	http://vcl.iti.gr/3d-object-retrieval/
31	Model bank library [37]	N/A	Mesh with texture	General	1,200	No	http://digimation.com/3dlibraries/model-bank-library/
32	SketchUp [4]	N/A	Mesh with and without texture	General	N/A	No	https://3dwarehouse.sketchup.com/
33	RoboCup at home [38]	Multiple	No data	Manipulation	N/A	No	http ://www.robocupathome.org/

and rehabilitation. Even though there have been many efforts that provide data sets of object mesh models that are useful for many simulation and planning applications as well as for benchmarking in shape retrieval, these data sets have limited utility for manipulation benchmarking for several reasons.

- Since most of them are not designed specifically for manipulation benchmarking, the selected objects do not usually cover the shape and function variety needed for a range of manipulation experiments.
- None of these databases provides the objects' physical properties, which are necessary to conduct realistic simulations.
- Most importantly, the vast majority of objects in these sets are not easily accessible by other researchers, preventing their use in experimental work.

Exceptions to this include [6], which provides an online shopping list (though it is now outdated, with many dead links), and the recently announced Amazon Picking Challenge [7], which provides a shopping list to purchase objects meant for a narrow bin-picking task. In the prosthetics and rehabilitation field, commercial kits are available for upper-limb assessment tests [8]–[11]. While demonstrating the benefits of utilizing a standard set for manipulation assessment, the scope of these kits is limited for benchmarking in robotics as they are not representative of a wide range of manipulation tasks. Our effort is unique in that it provides a large amount of information about the objects necessary for many simulation and planning approaches, makes the actual objects readily available for researchers to utilize experimentally, and includes a wide range of objects to span many different manipulation applications.

We provide a detailed overview of prior related benchmarking efforts, discussing their scope and limitations. For organization purposes, we first discuss work primarily related to robotic manipulation (including vision and learning applications), then efforts in rehabilitation, including prosthetics.

Robotic Manipulation

The necessity of manipulation benchmarks is highly recognized in the robotics community [12]–[14] and continues to be an active topic of discussion at workshops on robotic manipulation (see [15]). As mentioned earlier, the majority of prior work related to object sets has involved just object images and models (with varying degrees of information, from purely shape information to textural plus shape). Such work has often been created for research in computer vision (see [2], [16], and [17]). There have also been a number of shape/texture sets designed for/by the robotics community, particularly for applications such as planning and learning. The Columbia grasp database [3] rearranges the object models of the Princeton shape benchmark [18] for robotic manipulation and provides mesh models of 8,000 objects together with a number of successful grasps per model. Such a database is especially useful for implementing machine-learning-based grasp synthesis algorithms in which large amounts of labeled data are required for training the system. A multipurpose object set, which also targets manipulation, is the Karlsruhe Institute of Technology (KIT) object models database [19] which provides stereo images and textured mesh models of 100 objects. While there are a large number of objects in this database, the shape variety is limited, and like the previously mentioned data sets, the exact objects are typically not easily accessible to other researchers due to regional product differences or variation over time, and they generally cannot be purchased in one place as a set.

The object set is specifically designed to allow for widespread dissemination of the physical objects and manipulation scenarios.

There have only been two robotics-related efforts in which the objects are made relatively available. The household objects list [6] provides good shape variety that is appropriate for manipulation benchmarking as well as a shopping list for making the objects more easily accessible to researchers. Unfortunately, the list is outdated, and most objects are no longer available. The three-dimensional (3-D) models of objects in [6] are not supplied, which prevents the use of the object set in simulations. Very recently, the Amazon Picking Challenge [7] also provides a shopping list for items, but those were chosen to be specific to the bin-picking application and do not have models associated with them.

In terms of other robotic manipulation benchmarking efforts, a number of simulation tools have been presented in the literature. The OpenGRASP benchmarking suite [20] presents a simulation framework for robotic manipulation. The benchmarking suite provides test cases and setups and a standard evaluation scheme for the simulation results. So far, a model-based grasp synthesis benchmark has been presented using this suite. VisGraB [22] provides a benchmark framework for grasping unknown objects. The unique feature of this software is its utilization of real stereo images of the target objects for grasp synthesis as well as execution and evaluation of the result in a simulation environment. For gripper and hand design, benchmark tests [39], [40] are proposed for evaluating the ability of the grippers to hold an object, but only cylindrical objects are used.

Prosthetics and Rehabilitation

In the general field of rehabilitation and upper-limb prosthetics, there are a number of evaluation tools used by therapists to attempt to quantify upper-limb function in humans. Some of these are commercially available, clinically verified, and have been substantially covered in the literature, including normative data to compare a patient's performance to baselines. While some tools are commonly used, other tests have only been proposed in the literature and not (yet, at least) been widely utilized. Many of these tests aim to evaluate the ability of patients to perform tasks that contribute to activities of daily living.

The tests that are commercially available are the box-and-blocks test [41]; the nine-hole peg test [42]; the Jebsen–Taylor hand-function test [11]; the action research arm test (ARAT) [10]; the graded redefined assessment of strength, sensibility, and prehension (GRASSP) test [9]; and the Southampton hand-assessment procedure (SHAP) [8]. The setups for the box-and-blocks and nine-hole peg tests are very specific, with evaluation based on timed movements of simple objects. The setup for the Jebsen–Taylor hand-function test includes objects for manipulation actions, such as card turning, and moving small (paper clips, bottle caps), light (empty cans), and heavy objects (1-lb weighted cans), but it utilizes a small number of objects of limited shape and size variety. The ARAT assesses upper-limb function, and its commercial set [43] contains objects such as wooden blocks of various sizes, glasses, a stone, a marble, washers, and bolts. The test proposes actions like placing a washer over a bolt and pouring water from one glass into another. The GRASSP measure has also been proposed for the assessment of upper-limb impairment. It is based on a commercial kit available in [44]. Apart from a specialized manipulation setup, the kit also includes the nine-hole peg test, jars, and a bottle. The SHAP setup includes some objects of daily living, such as a bowl, a drink carton, and a jar, together with some geometrical shapes. Patients are requested to perform a variety of manipulation tasks, mostly involving transporting objects but also including pouring a drink, opening the jar, and so on. Considering manipulation benchmarking in robotics, the box-and-blocks, nine-hole peg, and Jebsen–Taylor hand-function tests are far from providing an adequate object variety for deriving new benchmarks. Despite enabling a larger possibility of manipulation tasks than the previously mentioned setups, the GRASSP and SHAP setups are still bounded to a limited number of tasks, and both are pricey (currently around US$1,300 and US$3,000, respectively).

Some well-known tests that do not provide a commercial setup are the grasp-and-release test [45], the Toronto Rehabilitation Institute hand-function test [33], and the activities measure for upper-limb amputees (AM-ULA) [46]. The grasp-and-release test is proposed for evaluating the performance of neuroprosthetic hands. For this test, detailed descriptions of the objects are given, but the objects are not easily obtainable, and the set includes an outdated object, i.e., a videotape. The Toronto Rehabilitation Institute hand-function test (also known as the Rehabilitation Engineering Laboratory hand-function test [47]) evaluates the palmar (power) and lateral (precision) grasp abilities of individuals using an object set comprising a mug, a book, a piece of paper, a soda can, dice, a pencil, and so on. Even though it is claimed that the objects used in this test are easily obtainable, maintaining the exact object definitions is hard, and one of the objects is an outdated cellular phone. The AM-ULA defines several quality measures for assessing the manipulation tasks, and various daily activities are proposed for the assessment. The objects used in the AM-ULA activities are not standardized.

In addition to these tests, some works in the literature use their own setups for assessment. In [48], tasks such as using a hammer and nail, stirring a bowl, folding a bath towel, and using a key in a lock are proposed for evaluating an upper-limb prosthesis. In [49], the performance of a neuroprosthesis is evaluated by asking the patient to perform grasping and lifting tasks as well as phone dialing, pouring liquid from a pitcher, and using a spoon and fork. In [50], to evaluate the outcomes of a protocol for stoke rehabilitation, blocks, Lego bricks, and pegs are used together with daily life activities like folding, buttoning, pouring, and lifting. In [51], the outcomes of a neuroprosthesis are measured with the box-and-blocks test and clothes-pin relocation task together with the evaluation of actions of daily living, i.e., using a fork and a knife, opening a jar, and stirring a spoon in a bowl. But none of the above-mentioned assessment procedures provides descriptions of the objects used.

Access to the objects is crucial to performance benchmarking as many aspects of the manipulation process cannot be modeled.

In our object set, we have included objects that are commonly used in these assessment procedures (i.e., a mug, a bowl, a pitcher, washers, bolts, kitchen items, pens, a padlock, and so on). We also included objects that will allow designing protocols that focus on activities of daily living. Moreover, widely used manipulation tests such as the nine-hole peg and box-and-blocks tests are also provided.

Object and Data Set

The contents of the proposed object set are shown in Figures 1–8 and listed in Table 2. The objects in the set are divided into the following categories: 1) food items, 2) kitchen items, 3) tool items, 4) shape items, and 5) task items tests are also provided.

Object Choices

We aimed to choose objects that are frequently used in daily life and also went through the literature to take into account objects that are frequently used in simulations and


Figure 1. The food items in the YCB object set. Back row, from left: a can of chips, a coffee can, a cracker box, a box of sugar, and a can of tomato soup. Middle row, from left: a container of mustard, a can of tuna fish, a box of chocolate pudding, a box of gelatin, and a can of potted meat. Front row: plastic fruits (a lemon, an apple, a pear, an orange, a banana, a peach, strawberries, and a plum).


Figure 2. The kitchen items in the YCB object set. Back row, from left: a pitcher, a container of bleach cleanser, and a container of glass cleaner. Middle row, from left: a plastic wine glass, an enamel-coated metal bowl, a metal mug, and an abrasive sponge. Front row, from left: a cooking skillet with a glass lid, a metal plate, eating utensils (knife, spoon, and fork), a spatula, and a white table cloth.

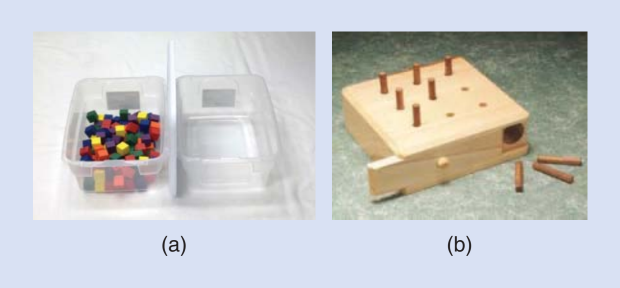
experiments. We also benefit from the studies on objects of daily living [52] and daily activities checklists such as [53].

In compiling the proposed object and task set, we needed to take a number of additional practical issues into consideration.

- *Variety*: To cover as many aspects of robotic manipulation as possible, we included objects that have a wide variety of shapes, sizes, transparencies, deformabilities, and textures. Considering size, the necessary grasp aperture varies from 14 cm (the diameter of the soccer ball) to 0.64 cm (the diameter of the smallest washer). Considering deformability, we have rigid objects together with foam bricks, a sponge, deformable balls, and articulated objects. Regarding transparency, we have included a transparent plastic wine glass, a glass skillet lid, and a semitransparent bottle of glass cleaner. The set includes objects with uniform plain textures, such as the pitcher and the stacking cups, and objects with irregular textures, like most of the groceries. Grasping and manipulation difficulty was also a criterion: for instance, some objects in the set are well approximated by simple geometric shapes (e.g., the box-shaped objects in the food category or the balls in the shape category) and relatively easy

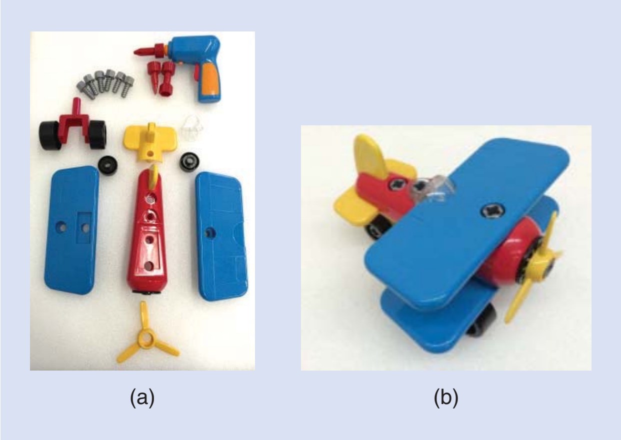
Figure 3. The tool items in the YCB object set. Back row, from left: a power drill and wood block. Middle row, from left: scissors, a padlock and keys, markers (two sizes), an adjustable wrench, Phillips- and flat-head screwdrivers, wood screws, nails (two sizes), plastic bolts and nuts, and a hammer. Front row: spring clamps (four sizes).

Figure 4. The shape items in the YCB object set. Back row, from left: a mini soccer ball, a softball, a baseball, a tennis ball, a racquetball, and a golf ball. Front row, from left: a plastic chain, washers (seven sizes), a foam brick, dice, marbles, a rope, stacking cups (set of ten), and a blank credit card.



(a) (b)

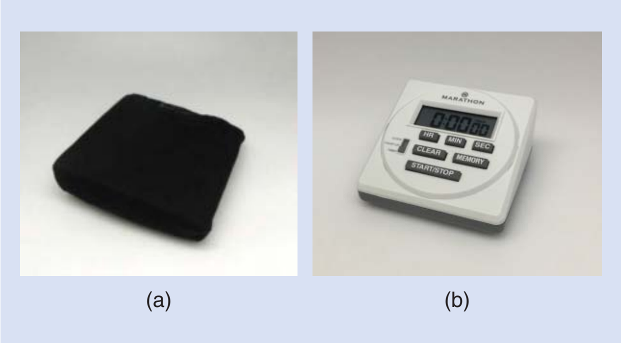
Figure 5. (a) The improvised box-and-blocks test objects: a set of 100 wooden cubes, two containers, and a height obstacle (container lid) between them. (b) The nine-hole peg test: wooden pegs are placed in holes and stored in the body of the box.



(a) (b)

Figure 6. The assembly object: (a) the toy airplane disassembled, including a toy power screwdriver, and (b) the fully assembled airplane.

Figure 7. The assembly object: Lego Duplo pieces.



(a) (b)

Figure 8. The task items: (a) a black T-shirt and (b) a timer for accurate timing and as a manipulation object with a keypad.

Table 2. Object set items and properties.

Identification Number	Class	Object	Mass (g)	Dimensions (mm)
1	Food items	Chips can	205	75 × 250
2		Master chef can	414	102 × 139
3		Cracker box	411	60 × 158 × 210
4		Sugar box	514	38 × 89 × 175
5		Tomato soup can	349	66 × 101
6		Mustard bottle	603	58 × 95 × 190
7		Tuna fish can	171	85 × 33
8		Pudding box	187	35 × 110 × 89
9		Gelatin box	97	28 × 85 × 73
10		Potted meat can	370	50 × 97 × 82
11		Banana	66	36 × 190
12		Strawberry	18	43.8 × 55
13		Apple	68	75
14		Lemon	29	54 × 68
15		Peach	33g	59
16		Pear	49	66.2 × 100
17		Orange	47	73
18		Plum	25g	52
19	Kitchen items	Pitcher base	178	108 × 235
20		Pitcher lid	66	123 × 48
21		Bleach cleanser	1,131	250 × 98 × 65
22		Windex bottle	1,022	80 × 105 × 270
23		Winelass	133	89 × 137
24		Bowl	147	159 × 53
25		Mug	118	80 × 82
26		Sponge	6.2	72 × 114 × 14
27		Skillet	950	270 × 25 × 30
28		Skillet lid	652	270 × 10 × 22
29		Plate	279	258 × 24
30		Fork	34	14 × 20 × 198
31		Spoon	30	14 × 20 × 195
32		Knife	31	14 × 20 × 215
33		Spatula	51.5	35 × 83 × 350
34		Table cloth	1,315	2,286 × 3,352
35	Tool items	Power drill	895	35 × 46 × 184
36		Wood block	729	85 × 85 × 200
37		Scissors	82	87 × 200 × 14
38		Padlock	304	24 × 47 × 65
39		Keys	10.1	23 × 43 × 2.2
40		Large marker	15.8	18 × 121
41		Small marker	8.2	8 × 135
42		Adjustable wrench	252	5 × 55 × 205
43	Tool items	Phillips screwdriver	97	31 × 215
44		Flat screwdriver	98.4	31 × 215
45		Nails	[2, 2.7, 4.8]	[4 × 25, 3 × 53, 4 × 63]
46		Plastic bolt	3.6	43 × 15
47		Plastic nut	1	15 × 8
48		Hammer	665	24 × 32 × 135
49		Small clamp	19.2	85 × 65 × 10
50		Medium clamp	59	90 × 115 × 27
51		Large clamp	125	125 × 165 × 32
52		Extra-large clamp	202	165 × 213 × 37
53	Shape items	Mini soccer ball	123	140
54		Softball	191	96
55		Baseball	148	75
56		Tennis ball	58	64.7
57		Racquetball	41	55.3
58		Golf ball	46	42.7
59		Chain	98	1,149
60		Washers	[0.1, 0.7, 1.1, 3, 5.3, 19, 48]	[6.4, 10, 13.3, 18.8, 25.4, 37.3, 51]
61		Foam brick	28	50 × 75 × 50
62		Dice	5.2	16.2
63		Marbles	N/A	N/A
64		Rope	18.3	3,000 × 4.7
65		Cups	[13, 14, 17, 19, 21, 26, 28, 31, 35, 38]	[55 × 60, 60 × 62, 65 × 64, 70 × 66, 75 × 68, 80 × 70, 85 × 72, 90 × 74, 95 × 76, 100 × 78]
66		Blank credit card	5.2	54 × 85 × 1
67		Rope	81	3,000
68	Task items	Clear box	302	292 × 429 × 149
69		Box lid	159	292 × 429 × 20
70		Colored wood blocks	10.8	26
71		Nine-hole peg test	1,435	1,150 × 1,200 × 1,200
72		Toy airplane	570	171 × 266 × 280
73		Lego Duplo	523	N/A
74		T-shirt	105	736 × 736
75		Magazine	73	265 × 200 × 1.6
76		Timer	102	85 × 80 × 40
77		Rubik’s Cube	94	57 × 57 × 57

for grasp synthesis and execution, while other objects have higher shape complexity (e.g., the spring clamps in the tool category, or the spatula in the kitchen-items category) and are more challenging for grasp synthesis and execution. Considering these aspects, the proposed set has a superior variety compared with the commercially available sets [8], [11], [41], [42], [44], which are designed to address some particular manipulation aspects only.

- *Use*: We included objects that are not only interesting for grasping but that also have a range of manipulation uses. For example, a pitcher and a cup; nails and a hammer; and pegs, cloths, and rope. We also included assembly items/tasks: a set of children's stacking cups, a toy airplane (Figure 6) that must be assembled and screwed together, and Lego Duplo bricks (Figure 7). In addition, widely used standard manipulation tests in rehabilitation, such as an improvised box-and-blocks [41] and a nine-hole peg test [42], are included. As mentioned above, these tasks are intended to span a wide range of difficulty, from relatively easy to very difficult. Furthermore, the ability to quantify the task performance was also prioritized, including aspects such as level of difficulty, time to completion, and success rate, among others.

Our goal is to do as much as possible to facilitate the widespread usage of a common set of objects and tasks to allow easy comparison of results between research groups worldwide.

Table 3. The suggestions for manipulation tasks.

Object Category	Suggested Tasks
Food items	• Packing/unpacking the groceries
Kitchen items	• Table setting • Wipe down table with sponge and Windex • Cooking scenarios
Tool items	• Nailing • Drilling • Unlocking the padlock using the key • Placing the pegs on the rope • Unscrewing a bolt using the wrench • Cutting a paper with the scissors • Writing on a paper • Screwing the nut on the bolt
Shape items	• Sorting marbles into the plastic blocks • Unstacking/stacking the cups • Placing the washers onto the bolt
Task items	• Box-and-blocks test • Toy-plane assembly/disassembly • Nine-hole peg tests • Lego assembly/disassembly • Cloth folding

- *Durability*: We aimed for objects that can be useful in the long term, and, therefore, avoid objects that are fragile or perishable. In addition, to increase the longevity of the object set, we chose objects that are likely to remain in circulation and change relatively little in the near future.
- *Cost*: We aimed to keep the cost of the object set as low as possible to broaden accessibility. We, therefore, selected standard consumer products, rather than, for instance, custom-fabricated objects, and tests. The current cost for the objects is approximately US$350.
- *Portability*: We aimed to have an object set that fits in a large-sized suitcase and be below the normal airline weight limit (22 kg) to allow easy shipping and storage.

After these considerations, the final objects were selected (Table 2 and Figures 1–8). Objects 1–18 are the food items, including real boxed and canned items as well as plastic fruits, which have complex shapes. Objects 19–34 are kitchen items, including objects for food preparation and serving as well as glass cleaner and a sponge. Objects 35–52 form the tool items category, containing not only common tools but also items—such as nails, screws, and wood—with which to utilize the tools. The shape items are objects 53–67, which span a range of sizes (spheres, cups, and washers), as well as compliant objects such as foam bricks, rope, and chain. The task items are objects 68–77, and they include two widely used tasks in rehabilitation benchmarking (box-and-blocks [41] and nine-hole peg test [42]) as well as items for relatively simple and complex assembly tasks (a Lego Duplo set and children's airplane toy, respectively). Furthermore, the set includes a black T-shirt for tasks like cloth folding as well as a magazine and a Rubik's cube. We include a timer in the kit (Figure 8), which not only provides accurate timing of the task but also serves as a manipulation object with a keypad. While there are an unlimited number of manipulation tasks that might be able to be done with these objects, we provide some examples for each category in Table 3 (with an in-depth discussion of tasks and protocols in the "Conclusions and Future Work" section).

Object Scans

To ease adoption across various manipulation research approaches, we collected visual data that are commonly required for grasping algorithms and generate 3-D models for use in simulation. We used the same scanning rig used to collect the BigBIRD data set [1]. The rig, shown in Figure 9, has five RGB-D sensors and five high-resolution RGB cameras arranged in a quarter-circular arc. Each object was placed on a computer-controlled turntable, which was rotated by 3° at a time, yielding 120 turntable orientations. Together, this yields 600 RGB-D images and 600 high-resolution RGB images. The process is completely automated, and the total collection time for each object is under 5 min.

We then used Poisson surface reconstruction to generate watertight meshes [54] (Figure 10). Afterward, we projected the meshes onto each image to generate segmentation masks. Note that Poisson reconstruction fails on certain objects with missing depth data; specifically, transparent or reflective regions of objects usually do not register depth data. We will later provide better models for these objects using algorithms that take advantage of the high-resolution RGB images for building models.

In total, for each object, we provide the following:

- 600 RGB-D images
- 600 high-resolution RGB images
- segmentation masks for each image
- calibration information for each image
- texture-mapped 3-D mesh models.

The object scans can be found online at [55].

Models

Based on the scans of the objects, there are several ways in which object models can be easily integrated into a variety of robot simulation packages. For example, in the MoveIt [5] simulation package, the mesh can be used as a collision object directly. Furthermore, a unified robot description format (URDF) file can be automatically constructed to integrate with ROS [56]. This provides a way to specify mass properties and can link to alternate representations of the mesh for visualization and collision. Integration with the OpenRAVE [57] simulation package is similarly straightforward where we link to the display and collision meshes from a KinBody XML file. Using the scans, we have created URDF and KinBody files for all of the objects in the data set, provided alongside the scans at [55].

Once in a simulation environment, a variety of motion planners and optimizers can use these models either as collision or manipulation objects. Some algorithms, such as Covariant Hamiltonian Optimization for Motion Planning [58], require signed-distance fields to avoid collisions, which can be computed from the included watertight meshes. Other cases, such as Constrained Bi-directional Rapidly-Exploring Random Tree [59], compute collisions directly using an optimized mesh collision checker.

In many cases, collision checking is a computational bottleneck for motion planning. Execution time can be reduced using a simplified mesh produced either by hand or with automatic decimation methods [60]. We have not yet provided simplified meshes in this data set, but we view this as an opportunity in future work to further explore mesh approximation algorithms and their impact on motion-planning problems using the standardized benchmarks.

Figure 9. The BigBIRD object-scanning rig: the box contains a computer-controlled turntable.

Functional Demonstration of Integration into Simulation Software

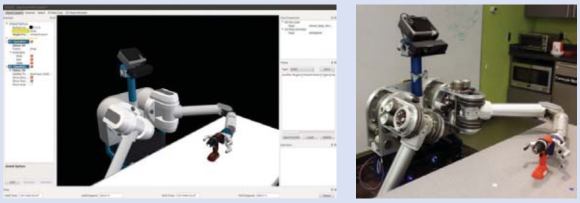
The entire pipeline is shown in Figure 11. Here, we see the HERB robot [61] preparing to grasp the virtual drill object. This demonstration uses an integration of ROS and OpenRAVE. The ROS is used to provide communication between the various hardware and software components of the robot, while OpenRave handles planning and collision checking.

A variety of motion planners and optimizers can use these models either as collision or manipulation objects.

Inside OpenRAVE, the HERB robot uses CBiRRT, the Open Motion Planning Library [62] library, and CHOMP to plan and optimize motion trajectories. Using these tools, chains of several actions can be executed in sequence. The simulation environment also

(a) (b)

Figure 10. The point-cloud and textural-data overlays on two YCB objects: (a) the mustard bottle and (b) the power drill.



(a) (b)

Figure 11. (a) The screen-capture from the OpenRAVE simulation and planning environment showing the HERB robot [34] planning a grasp of the power drill object in the set. (b) The actual grasp being executed by the robot on the physical object.

provides a mechanism for incorporating feedback from perception systems, which similarly benefit from this data set. The provided images, meshes, and physical objects can all be used as training data for various object-detection and pose-estimation algorithms, which can then be incorporated into the manipulation pipeline.

Access to both the physical object and a corresponding model for simulation is important for developing and testing new planning and manipulation algorithms. This data set vastly reduced the time required to set up this example by providing access to object models and meshes that have already been prepared for this purpose. This has removed the burden of scanning or modeling new objects and provides benchmark environments that streamline experimental design.

The objects in the set are designed to cover a wide range of aspects of the manipulation problem.

Protocol and Benchmark Template for Manipulation Research

Manipulation Protocol Template

Reference number/version
Authors
Institution
Contact information
Purpose
Task description
Setup description
  Description of the manipulation environment:
  List of objects and their descriptions:
  Initial poses of the objects:
Robot/hardware/subject description
  Targeted robots/hardware/subjects:
  Initial state of the robot/hardware/subject with respect to the setup:
  Prior information provided to the robot/hardware/subject:
Procedure
Execution constraints

Manipulation Benchmark Template

Reference number/version
Authors
Institution
Contact information
Adopted protocol
Scoring
Details of setup
Results to submit

Protocol Design for Manipulation

A standard set of objects and associated models is a great starting point for common replicable research and benchmarking in manipulation, but there must be a sufficient amount of specification about what should be done with the objects to directly compare approaches and results. Given the wide range of technical interests, research approaches and applications being examined in the manipulation research community, along with how quickly the field moves, we cannot possibly provide sufficient task descriptions that will span the range of interests and remain relevant in the long term. Instead, we seek to lay the groundwork for those to be driven by the research community and subcommunities. We, therefore, focus on two efforts: developing a framework for task protocols, setting, formatting, and content guidelines to facilitate effective community-driven specification of standard tasks; and a preliminary set of example protocols that we believe are relevant for our respective communities and approaches, along with experimental implementation of those, including reporting the performance outcomes.

To enable effective community-driven evolution of protocols and benchmarks, the web portal associated with this effort [63] will serve as a jumping-off point. Protocols proposed by the community will be hosted at this portal, allowing them to be easily posted, shared, and cited, as well as easily updated as researchers give feedback and identify shortcomings. The portal will provide a forum for discussions on individual protocols and will provide links to matured protocols that meet the standards laid out in the template.

Protocol Guidelines

While developing protocols and benchmarks, one challenging aspect is to decide on the level of detail. Providing only high-level descriptions of the experiment (in other words, setting too few constraints) makes the repeatability of a benchmark, as well as its ability to assess the performance, questionable. Variations caused by incomplete descriptions of test setups and execution processes induce discrepancy in measurements and would not speak to some quantifiable performance. On the other hand, supplying too many constraints may limit a protocol's applicability and, therefore, narrow down its scope. For example, due to the variety of utilized hardware by different research groups in the robotics field, satisfying constrained hardware descriptions is not usually possible or preferred.

The aim of this section is to provide guidelines that help to maintain both reliable and widely applicable benchmarks for manipulation. For this purpose, five categories of information are introduced for defining manipulation protocols, i.e., 1) task description, 2) setup description, 3) robot/

hardware/subject description, 4) procedure, and 5) execution constraints. These categories are explained below, and, for the template, see "Protocol and Benchmark Template for Manipulation Research."

- *Task Description*: The task description is the highest level of information about the protocol. It describes the main action(s) of a task and (most of the time implicitly) the expected outcome(s). In this level, no constraints are given on the setup layout or how the task should be executed. Some task description examples are pouring liquid from a pitcher to a glass, hammering a nail on a piece of wood, or grasping an apple.
- *Setup Description*: This category provides the list of objects used in the manipulation experiment and their initial poses with respect to each other. In addition, if there are any other objects used as obstacles or clutter in the manipulation scenario, their description and layout will be described. As discussed above, the usage of nonstandard objects introduces uncertainty to many manipulation experiments presented in the literature. We believe that removing uncertainties in this category of information is crucial to maintain well-defined benchmarks. Providing the YCB object and model set is a step toward that purpose. In addition, in the protocols proposed in this article, the initial poses of the objects are accurately provided. Naturally, a task description can have various setup descriptions designed to assess the manipulation performance in different conditions.
- *Robot/Hardware/Subject Description*: This category provides information about the task executor. If the protocol is designed for a robotic system, the initial state of the robot with respect to the target object(s) and a priori information provided to the robot about the manipulation operation (e.g., the semantic information about the task, whether or not object shape models are provided.) are specified in this category. In addition, if the protocol is designed for a specific hardware setup (including sensory suite), the description is given. If the task executor is a human subject, how the subject is positioned with respect to the manipulation setup and a priori information given to the subject about the task at hand are described here.
- *Procedure*: In this category, actions that are needed to be taken by the person who conducts the experiment are explained step by step.
- *Execution Constraints*: In this category, the constraints on how to execute the task are provided. For instance in the box-and-blocks test, the subject is expected to use his/her dominant hand and needs to transfer one block at a time, or, if the task is to fetch a mug, the robot may be required to grasp the mug from its handle. In "Protocol and Benchmark Template for Manipulation Research," we provide a template for easily designing manipulation protocols using the aforementioned categories.

The proposed template and categories have several advantages as follows.

- The categorization helps researchers think about the protocol design in a structured way.
- It separates high-level task description from setup and robot/hardware/subject description so that protocols can be designed for analyzing different scenarios of the same manipulation problem.

Furthermore, describing setup and robot/hardware/subject separately allows platform-independent benchmark designs. Especially in the robotics field, the researchers usually have limited access to hardware. The designer may prefer to impose few constraints on the robot/hardware/subject description category to increase the applicability of the protocol. The amount and specifics of the detail in a given protocol will naturally vary based on the particular problem being examined, and therefore the insight of the authors about the intended application will be crucial in crafting an effective set of task descriptions and constraints. Related to this point, we anticipate protocols to be regularly improved and updated with feedback from the research community.

Benchmark Guidelines

After the task description, the second major part of each protocol is the specification of the associated benchmark, which details the metrics for scoring performance for the given protocol. Benchmarks allow the researchers to specify the performance of their system or approach and enable direct comparison with other approaches. The following categories of information are introduced for defining manipulation benchmarks.

- *Adopted protocol*: A well-defined description can be obtained for a manipulation benchmark by adopting a protocol that is defined considering the above-mentioned aspects.
- *Scoring*: Providing descriptive assessment measures is crucial for a benchmark. The output of the benchmark should give reasonable insight of the performance of a system. While designing the scoring criteria, it is usually a good practice to avoid binary (success/fail) measures; if possible, the scoring should include the intermediate steps of the task, giving partial points for a reasonable partial execution.
- *Details of setup*: In this field, the user gives detailed information about setup description that is not specified by the protocol. This could include the robot type, gripper type, grasping strategy, motion-planning algorithm, grasp synthesis algorithm, and so on.
- *Results to submit*: This field specifies the results and scores that need to be submitted by the user. Moreover, asking the user to submit the detailed reasoning for the failed attempts and the factors that bring success would help researchers who analyze the results. Therefore, having explicit fields for result analysis would be a good practice (see example benchmarks in [64]).

YCB Protocols and Benchmarks

While this protocol structure definition (and the template provided in "Protocol and Benchmark Template for Manipulation Research") helps to guide the development of effective task specification for various manipulation benchmarks, we have developed a number of example protocols to both provide more concrete samples of the types of task definitions that can be put forward as well as specific and useful benchmarks for actually quantifying performance. We have defined five protocols to date:

- pitcher–mug protocol
- gripper-assessment protocol
- table-setting protocol
- block pick-and-place protocol
- peg-insertion learning-assessment protocol.

From each protocol, a benchmark of reported performance is derived with the same name. We have implemented each of the protocols experimentally and reported the benchmark performance of our implementations for each. All these protocols and benchmarks and the results discussed in this section can be found at [64]. We have also implemented the box-and-blocks test for maintaining a baseline performance of this test for robotic manipulation.

A URDF file that spawns the scenario for Gazebo simulation environment is given.

YCB Pitcher–Mug Protocol and Benchmark

One of the popular tasks among robotics researchers is pouring a liquid from a container. This task is interesting as it necessitates semantic interpretation and smooth and precise manipulation of the target object. A protocol is designed for executing this manipulation task. The protocol uses the pitcher and the mug of YCB object and model set and provides scenarios by specifying ten initial configurations of the pitcher and the mug. By standardizing the objects and providing detailed initial state information, it aims at maintaining a common basis of comparison between different research groups. The benchmark derived from this protocol uses a scoring scheme that penalizes the amount of liquid that remains in the pitcher or spilled on the table. This benchmark was applied using the HERB robot platform [61], which can be seen in Figure 12. The reported results show that the task is successfully executed for eight of ten pitcher–mug configurations. For the two failed cases, the robot is able to grasp the pitcher but cannot generate a suitable path for pouring the liquid. This shows the importance of planning the manipulation task as a whole rather than in segments.

Figure 12. The HERB robot implementing the pitcher–mug benchmark.

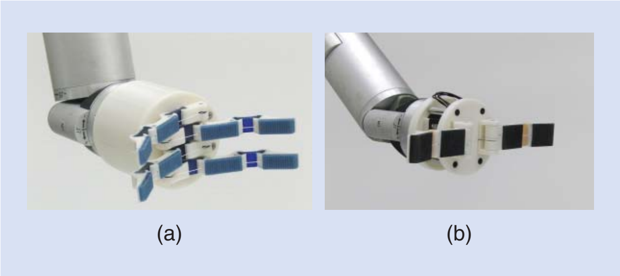

(a)
(b)

Figure 13. The grippers compared with gripper assessment benchmark: (a) Model T and (b) Model T42.

YCB Gripper-Assessment Protocol and Benchmark

The abilities of a robot's gripper affect its manipulation performance significantly. In the literature and in the commercial market, various gripper designs are available, each of which has different manipulation capabilities. The protocol defines a test procedure for assessing the performance of grippers for grasping objects of various shapes and sizes. This protocol utilizes objects from the shape and tool categories of the YCB object and model set. Using this protocol, a benchmark is defined based on a scoring table. We applied this benchmark to two grippers designed in Yale GRAB Lab, the Model T and Model T42 [65], which are shown in Figure 13. The results show that the Model T can provide successful grasp for only a limited range of object sizes. This gripper is not suitable for grasping small and flat objects. However, the ability to interlace its fingers increases the contact surface with the object and brings an advantage, especially for grasping concave and articulated objects. The Model T42 is able to provide stable power grasps for large objects and precision grasps for small objects. This model is also successful in grasping flat objects thanks to its nail-like fingertips. However, not being able to interlace its fingers brings a disadvantage while grasping articulated objects. Using the same benchmark for evaluating different gripper designs not only provided a basis of comparison but also gave many clues about how to improve the designs.

YCB Protocol and Benchmark for Table Setting

Pick-and-place is an essential ability for service robots. The benchmark assesses this ability by the daily task of table setting. The protocol uses the mug, fork, knife, spoon, bowl, and plate of the YCB object and model set. These objects are placed to predefined initial locations, and the robot is expected to replace them to specific final configurations. The benchmark scores the performance of the robot by the accuracy of the final object poses. This benchmark can also be applied in a simulation environment, since the models of the objects are provided by the YCB object and model set. A URDF file that spawns the scenario for Gazebo simulation environment is given at http://rll.eecs.berkeley.edu/ycb/. A snapshot of this setting can be seen in Figure 14.

YCB Block Pick-and-Place Protocol and Benchmark

Manual dexterity and the manipulation of small objects are critical skills for robots in several contexts. The block pick-and-place protocol is designed to test a robot's ability to grasp small objects and transfer them to a specified location. This task is an important test of both arm and gripper hardware and motion planning software, as both contribute to overall dexterity. Points are awarded based on completion and precision of the manipulation. We executed this test on the HERB robot [61], as seen in Figure 15. An image of the printed layout with the placed blocks after task completion can be seen in Figure 16. The results show that the robot is not able to succeed in precise pick-and-place task. The main reason is the utilized open-loop grasping approach. The robot executes a robust push grasp strategy, which allows it to grasp the blocks successfully. However, the pose of the block with respect to the gripper is not known precisely after the grasp. This prevents placing the blocks accurately to the target locations.

YCB Peg-Insertion Learning-Assessment Protocol and Benchmark

The peg-insertion learning-assessment benchmark is designed to allow comparison among various learning techniques. The benchmark measures the performance of a learned peg-insertion action under various positioning perturbations. The perturbations are applied by moving the peg board to a random direction for certain amount of distance. We applied this benchmark to assess the performance of a learned linear-Gaussian controller using a PR2 robot [66] (Figure 17). The state of the controller consists of the joint angles and angular velocities of the robot as well as the positions and velocities of three points in the space of the end effector (three points to fully define a rigid body configuration). No information is available to the controller at run time except for this state information. The results show that the learned controller shows reasonable performance, with four successes out of ten trials, for the case of 5-mm position perturbation to a random direction. This success rate can be achieved by executing the controller for only 1 s. However, the performance does not improve, even if the controller is run for a longer period of time. In the case of 10-mm position perturbation, the controller fails completely. We are planning to learn the same task with different learning techniques and compare their performances using the benchmark.

> **The Model T42 is able to provide stable power grasps for large objects and precision grasps for small objects.**

Figure 14. The simulation environment for the table-setting benchmark. This environment can be spawned using the URDF provided at http://rll.eecs.berkeley.edu/ycb.

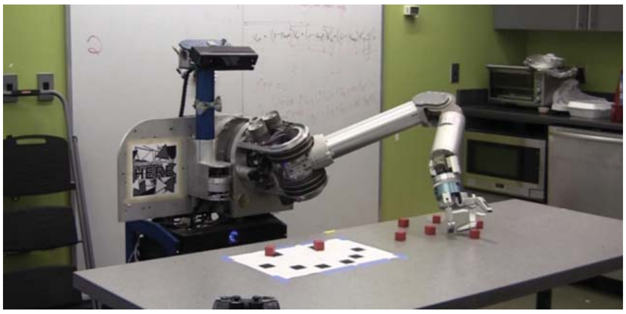
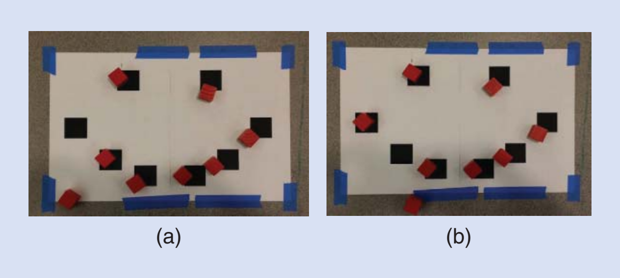


Figure 15. The HERB robot implementing the peg-insertion learning-assessment benchmark.



(a) (b)

Figure 16. (a) and (b) The results of the block pick-and-place benchmark.

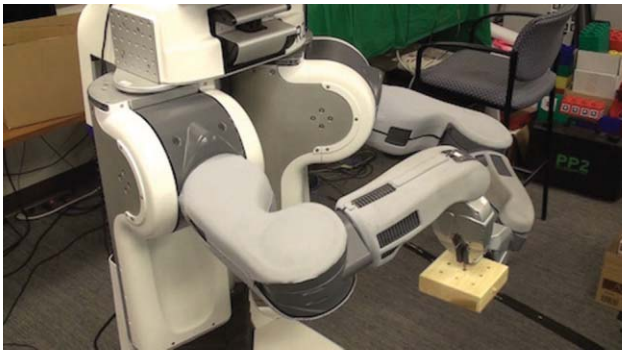


Figure 17. The PR2 executing the peg-insertion learning-assessment benchmark.

Box-and-Blocks Test

As mentioned in the "Related Work" section, the box-and-blocks test [41] is a widely used assessment technique that is utilized in prosthetics and rehabilitation fields. The test evaluates how many blocks can be grasped and moved from one side of the box (Figure 18) to the other in a fixed amount of time. We believe that the application of this test can also be quite useful for assessing the manipulation capabilities of robots. To establish a baseline performance for this test for robotic manipulators, we applied the box-and-blocks test with a PR2 robot (Figure 18) by implementing a very simple heuristic rules. The robot picks a location from a uniform distribution over the box and attempts to pick up a block. The gripper's pose aligns with the length of the box. The gripper is then closed and checked if it is fully closed. If the gripper closes fully, this means no blocks have been grasped and, therefore, the robot chooses a new location to attempt another pick. The robot repeats this heuristic until the gripper is not fully closed. When a grasp is detected, the robot moves to the destination box and releases the block. By using this heuristic, we run ten experiments of 2 min each and report the results at [64].

The OpenGRASP benchmarking suite [20] presents a simulation framework for robotic manipulation.

Conclusions and Future Work

This article proposes a set of objects and related tasks as well as high-resolution scans and models of those objects, intended to serve as a widely distributed and widely utilized set of standard objects to facilitate the implementation of standard performance benchmarks for robotic grasping and manipulation research. The objects were chosen based on an in-depth literature review of other object sets and tasks previously proposed and utilized in robotics research, with additional consideration to efforts in prosthetics and rehabilitation. Furthermore, a number of practical constraints were considered, including a reasonable total size and mass of the set for portability, low cost, durability, and the likelihood that the objects would remain mostly unchanged in years to come. High-resolution RGB-D scans of the object in the set were completed, and 3-D models have been constructed to allow easy portability into simulation and planning environments. All of these data are freely available in the associated repository [55]. Over the course of 2015, 50 objects sets will be freely distributed to a large number of research groups through workshops/tutorials associated with this effort. Additional object sets will be made available to purchase otherwise.

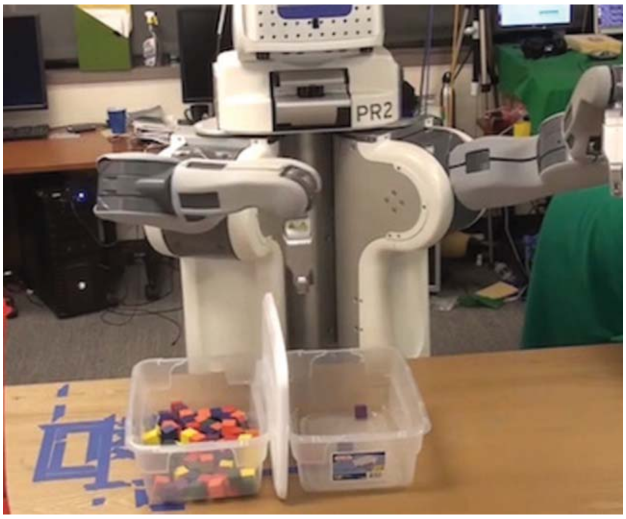


Figure 18. The PR2 executing the box-and-blocks test.

While a common set of widely available objects is a much-needed contribution to the manipulation research community, the objects themselves form only part of the contribution of the YCB set. The generation of appropriately detailed tasks and protocols involving the objects is ultimately what will allow for replicable research and performance comparison. We make inroads into that problem in this article by proposing a structure for protocols and benchmarks, implemented in a template as well as six example protocols. We hope that specification of protocols and benchmarks will become subcommunity driven and continually evolving. Specific aspects of manipulation and other specific research interests will naturally require different task particulars (i.e., specified and free parameters). We, therefore, plan to involve the research community in this effort via our web portal [63]. We will work toward having the majority of such protocols come from the user community rather than the authors of this article. In addition, we plan to have on this portal a records-keeping functionality to keep track of the current world records for the different tasks and protocols, along with video and detailed descriptions of the approaches utilized, generating excitement, buzz, motivation, and inspiration for the manipulation community to compare approaches and push forward the state of the art.

Other efforts that we plan to undertake include more detail about the objects proposed, including information about the inertia of the objects, as well as frictional properties between the objects and common surfaces. Additionally, we will expand our treatment of the modeling of the objects, including addressing the tradeoffs between number of triangles in a mesh and the reliable representation of the object geometry. Furthermore, before final publication and distribution of the object set, we will seek additional input from the research community on the specific objects in the set.

It is our hope that this article will help to address the longstanding need for common performance comparisons and benchmarks in the research community and will provide a starting point for further focused discussion and iterations on the topic.

Acknowledgment

The authors would like to thank Michael Leddy for his efforts in measuring the physical properties of the objects in the set. We would also like to thank Peter Allen and Jeff Trinkle for their feedback on the objects in the set and the overall approach taken in this article.

Funding for this work was provided in part by the U.S. National Science Foundation under grants IIS-0953856, IIS-1139078, and IIS-1317976.

References

[1] A. Singh, J. Sha, K. S. Narayan, T. Achim, and P. Abbeel, "BigBIRD: A large-scale 3D database of object instances," in *Proc. Int. Conf. Robotics Automation*, 2014, pp. 509–516.

[2] B. Li, Y. Lu, C. Li, A. Godil, T. Schreck, M. Aono, Q. Chen, N. K. Chowdhury, B. Fang, T. Furuya, H. Johan, R. Kosaka, H. Koyanagi, R. Ohbuchi, and A. Tatsuma. (2014). SHREC'14 track: Large scale comprehensive 3D shape retrieval. [Online]. Available: http://www.itl.nist.gov/iad/vug/sharp/contest/2014/Generic3D/

[3] C. Goldfeder, M. Ciocarlie, H. Dang, and P. K. Allen, "The Columbia grasp database," in *Proc. IEEE Int. Conf. Robotics Automation*, 2009, pp. 1710–1716.

[4] (2014). Sketchup. Available: [Online]. Available: http://www.sketchup.com/

[5] S. Chitta, I. Sucan, and S. Cousins, "MoveIt!," *IEEE Robot. Automat. Mag.*, vol. 19, pp. 18–19, Mar. 2012.

[6] Y. S. Choi, T. Deyle, T. Chen, J. D. Glass, and C. C. Kemp, "A list of household objects for robotic retrieval prioritized by people with ALS," in *Proc. IEEE Int. Conf. Rehabilitation Robotics*, 2009, pp. 510–517.

[7] Amazon Picking Challenge. [Online]. Available: http://amazonpickingchallenge.org/

[8] J. Adams, K. Hodges, J. Kujawa, and C. Metcalf, "Test-retest reliability of the Southampton hand assessment procedure," *Int. J. Rehab. Res.*, vol. 32, p. S18, Aug. 2009.

[9] S. Kalsi-Ryan, A. Curt, M. Verrier, and M. G. Fehlings, "Development of the graded redefined assessment of strength, sensibility and prehension (GRASSP): Reviewing measurement specific to the upper limb in tetraplegia," *J. Neurosurg Spine*, vol. 1, pp. 65–76, Sept. 2012.

[10] N. Yozbatiran, L. Der-Yeghiaian, and S. C. Cramer, "A standardized approach to performing the action research arm test," *Neurorehabil Neural Repair*, vol. 22, no. 1, pp. 78–90, 2008.

[11] E. D. Sears and K. C. Chung, "Validity and responsiveness of the Jebsen-Taylor hand function test," *J. Hand Surg. Amer.*, vol. 35, pp. 30–37, Jan. 2010.

[12] A. P. del Pobil, "Benchmarks in robotics research," in *Proc. IEEE Int. Conf. Robotics System*, 2006.

[13] R. Madhavan, R. Lakaemper, and T. Kalmar-Nagy, "Benchmarking and standardization of intelligent robotic systems," in *Proc. Int. Conf. Advanced Robotics*, 2009, pp. 1–7.

[14] I. Iossifidis, G. Lawitzky, S. Knoop, and R. Zöllner, "Towards benchmarking of domestic robotic assistants," in *Advances in Human-Robot Interaction*, vol. 14. Berlin, Heidelberg, Germany: Springer, 2005, pp. 403–414.

[15] R. Detry, O. Kroemer, and D. Kragic, "International workshop on autonomous grasping and manipulation: An open challenge," in *Proc. ICRA*, 2014.

[16] (2006). AIM@SHAPE Shape Repository v4.0. [Online]. Available: http://shapes.aim-atshape.net/viewmodels.php

[17] A. Richtsfeld. (2012). The object segmentation database (OSD). [Online]. Available: http://www.acin.tuwien.ac.at/?id=289

[18] P. Shilane, P. Min, M. Kazhdan, and T. Funkhouser, "The Princeton shape benchmark," in *Proc. Shape Modeling Applications*, 2004, pp. 167–178.

[19] A. Kasper, Z. Xue, and R. Dillmann, "The KIT object models database: An object model database for object recognition, localization and manipulation in service robotics," *Int. J. Robot. Res.*, vol. 31, no. 8, pp. 927–934, 2012.

[20] S. Ulbrich, D. Kappler, T. Asfour, N. Vahrenkamp, A. Bierbaum, M. Przybylski, and R. Dillmann, "The OpenGRASP benchmarking suite: An environment for the comparative analysis of grasping and dexterous manipulation," in *Proc. IEEE/RSJ Int. Conf. Intelligent Robots Systems*, 2011, pp. 1761–1767.

[21] B. Li, A. Godil, M. Aono, X. Bai, T. Furuya, L. Li, R. López-Sastre, H. Johan, R. Ohbuchi, C. Redondo-Cabrera, A. Tatsuma, T. Yanagimachi, and S. Zhang, (2012). SHREC'12 track: Generic 3D shape retrieval. [Online]. Available: http://www.itl.nist.gov/iad/vug/sharp/contest/2012/Generic3D/

[22] G. Kootstra, M. Popovic, J. A. Jorgensen, D. Kragic, H. G. Petersen, and N. Kruger, "VisGraB: A benchmark for vision-based grasping," *Paladyn*, vol. 3, no. 2, pp. 54–62, 2012.

[23] A. Tatsuma, H. Koyanagi, and M. Aono, "A large-scale Shape Benchmark for 3D object retrieval: Toyohashi shape benchmark," in *Proc Asia-Pacific Signal Information Processing Association Annu. Summit Conf.*, 2012, pp. 1–10.

[24] R. B. Rusu. (2011). NIST and willow garage: Solutions in perception challenge. [Online]. Available: http://www.nist.gov/el/isd/challenge-022511.cfm and http://www.willowgarage.com/blog/2011/02/28/nistand-willow-garage-solutions-perception-challenge

[25] H. Dutagaci, A. Godil, P. Daras, A. Axenopoulos, G. Litos, S. Manolopoulou, K. Goto, T. Yanagimachi, Y. Kurita, S. Kawamura, T. Furuya, and R. Ohbuchi. (2011). SHREC'11 Track: Generic Shape Retrieval. [Online]. Available: http://www.itl.nist.gov/iad/vug/sharp/contest/2011/NonRigid/

[26] A. Janoch, S. Karayev, Y. Jia, J. T. Barron, M. Fritz, K. Saenko, and T. Darrell, "A category-level 3-D object dataset: Putting the Kinect to work," in *Proc. IEEE Int. Conf. Computer Vision Workshops*, 2011, pp. 1168–1174.

[27] K. Lai, L. Bo, X. Ren, and D. Fox, "A large-scale hierarchical multi-view RGB-D object dataset," in *Proc. IEEE Int. Conf. Robotics Automation*, 2011, pp. 1817–1824.

[28] Z. Lian, A. Godil, T. Fabry, T. Furuya, J. Hermans, R. Ohbuchi, C. Shu, D. Smeets, P. Suetens, D. Vandermeulen, and S. Wuhrer. (2010). SHREC'10 track: Non-rigid 3D shape retrieval. [Online]. Available: http://tosca.cs.technion.ac.il/book/shrec_robustness2010.html

[29] R. Wessel, M. B. Ina, and R. Klein, "A 3D shape benchmark for retrieval and automatic classification of architectural data," in *Proc. Eurographics Workshop 3D Object Retrieval*, 2009, pp. 53–56.

[30] N. Iyer, S. Jayanti, and K. Ramani, "An engineering shape benchmark for 3D models," in *Proc. IDETC/CIE ASME Int. Design Engineering Tech. Conf. Computers Information Engineering Conf.*, 2005, pp. 501–509.

[31] R. Fang, A. Godil, X. Li, and A. Wagan, "A new shape benchmark for 3D object retrieval," in *Proc. 4th Int. Symp. Advances Visual Computing*, Berlin, Heidelberg, Germany, 2008, pp. 381–392.

[32] (2005). McGill 3D Shape Benchmark. [Online]. Available: http://www.cim.mcgill.ca/~shape/benchMark/

[33] N. Kapadia, V. Zivanovic, M. Verrier, and M. R. Popovic, "Toronto rehabilitation institute-hand function test: Assessment of gross motor function in individuals with spinal cord injury," *Top. Spinal Cord Inj. Rehabil.*, vol. 18, no. 2, pp. 167–186, 2012.

[34] R. W. Sumner and J. Popovic, "Deformation transfer for triangle meshes," *ACM Trans. Graphics*, vol. 3, no. 3, pp. 399–405, Aug. 2014.

[35] (2003). NTU 3D Model Benchmark. [Online]. Available: http://3d.csie.ntu.edu.tw/

[36] (2014). The ITI database. [Online]. Available: http://vcl.iti.gr/3dobject-retrieval/

[37] (2014). Model Bank Library. [Online]. Available: http://digimation.com/3d-libraries/model-bank-library/

[38] T. Wisspeintner, T. van der Zan, L. Iocchi, and S. Schiffer, "RoboCup@Home: Results in benchmarking domestic service robots," in *RoboCup 2009: Robot Soccer World Cup XIII*, vol. 5949. Berlin, Heidelberg, Germany: Springer, 2010, pp. 390–401.

[39] G. Kragten, A. C. Kool, and J. Herder, "Ability to hold grasped objects by underactuated hands: Performance prediction and experiments," in *Proc. IEEE Int. Conf. Robotics Automation*, 2009, pp. 2493–2498.

[40] G. A. Kragten, C. Meijneke, and J. L. Herder, "A proposal for benchmark tests for underactuated or compliant hands," *Mech. Sci.*, vol. 1, no. 1, pp. 13–18, 2010.

[41] V. Mathiowetz, G. Volland, N. Kashman, and K. Wever, "Adult norms for the box and blocks test of manual dexterity," *Amer. J. Occup. Ther.*, vol. 39, no. 6, pp. 386–391, 1985.

[42] V. Mathiowetz, K. Weber, N. Kashman, and G. Volland, "Adult norms for the nine hole peg test of finger dexterity," *Occup. Ther. J. Res.*, vol. 5, no. 1, pp. 24–38, 1985.

[43] Salia Rehab. (2015, Aug. 7). Action Research Arm Test (ARAT). [Online]. Available: http://saliarehab.com/

[44] (2015, Aug. 7). GRASSP. [Online]. Available: http://grassptest.com

[45] (2015, Aug. 7). Grasp and Release Test. [Online]. Available: http://www.rehabmeasures.org/Lists/RehabMeasures/DispForm.aspx?ID=1053

[46] L. Resnik, L. Adams, M. Borgia, J. Delikat, R. Disla, C. Ebner, and L. S. Walters, "Development and evaluation of the activities measure for upper limb amputees," *Arch. Phy. Med. Rehab.*, vol. 94, no. 3, pp. 488–494, 2013.

[47] M. Popovic and C. Contway, "Rehabilitation engineering laboratory hand function test for functional electrical stimulation assisted grasping," in *Proc. 8th Annu. Conf. Int. Functional Electrical Stimulation Society*, 2003, pp. 231–234.

[48] H. Burger, F. Franchignoni, A. W. Heinemann, S. Kotnik, and A. Giordano, "Validation of the orthotics and prosthetics user survey upper extremity functional status module in people with unilateral upper limb amputation," *J. Rehab. Med.*, vol. 40, no. 5, pp. 393–399, 2008.

[49] M. R. Popovic, T. Keller, I. P. I. Papas, V. Dietz, and M. Morari, "Surface-stimulation technology for grasping and walking neuroprostheses," *IEEE Eng. Med. Biol. Mag.*, vol. 20, no. 1, pp. 82–93, 2001.

[50] J. Harris, J. Eng, W. Miller, and A. S. Dawson, "A self-administered graded repetitive arm supplementary program (GRASP) improves arm function during inpatient stroke rehabilitation: A multi-site randomized controlled trial," *Stroke*, vol. 40, no. 6, pp. 2123–2131, 2009.

[51] K. A. Stubblefield, L. A. Milller, R. D. Lipschutz, M. E. Phillips, C. W. Heckathorne, and T. A. Kuiken, "Occupational therapy outcomes with targeted hyper-reinnervation nerve transfer surgery: Two case studies," in *Proc. MyoElectric Controls/Powered Prosthetics Symp.*, 2005, pp. 169–177.

[52] K. Matheus and A. M. Dollar, "Benchmarking grasping and manipulation: Properties of the objects of daily living," in *Proc. IEEE/RSJ Int. Conf.Intelligent Robots Systems*, 2010, pp. 5020–5027.

[53] (2015, Aug. 7). The Dash—"Disabilities of the Arm, Shoulder and Hand". [Online]. Available: http://dash.iwh.on.ca/system/files/dash_questionnaire_2010.pdf

[54] M. Kazhdan, M. Bolitho, and H. Hoppe, "Poisson Surface Reconstruction," presented at the Eurographics Symp. Geometry Processing, 2006.

[55] (2015, Aug. 7). [Online]. Available: http://rll.eecs.berkeley.edu/ycb/

[56] M. Quigley, K. Conley, B. Gerkey, J. Faust, T. Foote, J. Leibs, E. Berger, R. Wheeler, and A. Y. Ng, "ROS: An open-source robot operating system," in *Proc. ICRA Workshop Open Source Software*, 2009, vol. 3, p. 5.

[57] R. Diankov and J. Kuffner, "OpenRAVE: A planning architecture for autonomous robotics," Robotics Institute, Carnegie Mellon Univ., Pittsburgh, PA, Tech. Rep. CMU-RI-TR-08-34, 2008.

[58] N. Ratliff, M. Zucker, J. A. Bagnell, and S. Srinivasa, "CHOMP: Gradient optimization techniques for efficient motion planning," in *Proc. IEEE Int. Conf. Robotics Automation*, 2009, vols. 1–7, pp. 4030–4035.

[59] D. Berenson, S. S. Srinivasa, D. Ferguson, and J. J. Kuffner, "Manipulation planning on constraint manifolds," in *Proc. IEEE Int. Conf. Robotics Automation*, 2009, pp. 625–632.

[60] M. Garland and P. S. Heckbert, "Surface simplification using quadric error metrics," presented at the Proc. 24th Annu. Conf. Computer Graphics and Interactive Techniques, 1997, pp. 209–216.

[61] S. S. Srinivasa, D. Berenson, M. Cakmak, A. Collet, M. R. Dogar, A. D. Dragan, R. A. Knepper, T. D. Niemueller, K. Strabala, and J. M. Vandeweghe, "Herb 2.0: Lessons learned from developing a mobile manipulator for the home," *Proc. IEEE*, vol. 100, no. 8, pp. 2410–2428, 2012.

[62] I. A. Sucan, M. Moll, and L. E. Kavraki, "The open motion planning library," *IEEE Robot. Automat. Mag.*, vol. 19, pp. 72–82, Dec. 2012.

[63] (2015, Aug. 7). YCB Protocols and Benchmarks. [Online]. Available: http://www.ycbbenchmarks.org

[64] (2015, Aug. 7). YCB Protocols. [Online]. Available: http://protocols.ycb-benchmarks.org

[65] R. R. MA, L. U. Odhner, and A. M. Dollar, "A modular, open-source 3D printed underactuated hands," in *Proc. IEEE Int. Conf. Robotics Automation*, Karlsruhe, Germany, 2013, pp. 2737–2743.

[66] W. Garage. (2009). Overview of the PR2 robot. [Online]. Available: http://www.willowgarage.com/pages/pr2/overview

Berk Calli, Yale University, New Haven, Connecticut, USA. E-mail: berkcc@gmail.com.

Aaron Walsman, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA. E-mail awalsman@andrew.cmu.edu.

Arjun Singh, University of California, Berkeley, California, USA. E-mail: arjunsingh@berkeley.edu.

Siddhartha Srinivasa, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA. E-mail: siddh@cs.cmu.edu.

Pieter Abbeel, University of California, Berkeley, California, USA. E-mail: pabbeel@cs.berkeley.edu.

Aaron M. Dollar, Yale University, New Haven, Connecticut, USA. E-mail: aaron.dollar@yale.edu.

Competitions for Benchmarking

Task and Functionality Scoring Complete Performance Assessment

By Francesco Amigoni, Emanuele Bastianelli, Jakob Berghofer, Andrea Bonarini, Giulio Fontana, Nico Hochgeschwender, Luca Iocchi, Gerhard K. Kraetzschmar, Pedro Lima, Matteo Matteucci, Pedro Miraldo, Daniele Nardi, and Viola Schiaffonati

Scientific experiments and robotic competitions share some common traits that can put the debate about developing better experimental methodologies and replicability of results in robotics research on more solid ground. In this context, the Robot Competitions Kick Innovation in Cognitive Systems and Robotics (RoCKIn) project aims to develop competitions that come close to scientific experiments, providing an objective performance evaluation of robot systems under controlled and replicable conditions. In this article, by further articulating replicability into reproducibility and repeatability and by considering some results from the 2014 first RoCKIn competition, we show that the RoCKIn approach offers tools that enable the replicability of experimental results.

Robotic Competitions and Challenges

Within the debate about the development of rigorous experimental methodologies in robotics research, the robotic competitions have emerged as a way to promote comparison of different algorithms

Digital Object Identifier 10.1109/MRA.2015.2448871
Date of publication: 11 September 2015

1070-9932/15©2015IEEE

and systems, allowing for the replication of their results [3], [19]. Experiments and competitions present differences: an experiment evaluates a specific hypothesis, while a competition usually evaluates general abilities of robot systems. Moreover, the competitions often push the development of solutions, while experiments aim to explore phenomena and share the knowledge acquired through their results. However, merging these complementary approaches can lead to the development of an approach to competitions that makes them more scientifically grounded and suitable for benchmarking. The research and infrastructures for competitions that the robotics community has developed during the past few years can be exploited to make experimental methods in robotics sounder and more systematic, building on the common traits shared by experiments and competitions. The competitions involve robots in dynamic but controlled environments, and, having clear measures of success, these environments provide opportunities to evaluate different approaches against each other and over years of progress. Furthermore, they require integrated implementation of complete robot systems, promoting a new experimental paradigm that complements the traditional paradigm of evaluating specific modules in isolation. The competitions can thus provide a common ground for rigorously comparing different solutions, playing the role of experiments and exploiting their distinctive features, such as being appealing (both to researchers and to the general public), taking place with regularity and precise timing, promoting critical analysis of experiments out of labs, and sharing the cost and effort of setting up complex experimental installations among participants.

The robotics competitions and challenges have gained popularity from the 1970s, and now there are countless events per year.

The robotics competitions and challenges have gained popularity from the 1970s, and now there are countless events per year. From the very beginning, it has been recognized that the competitions can serve several, often conflicting, purposes, including promoting education and research to push the field forward, entertaining general audiences, and building community [5]. Although balancing these goals is sometimes difficult and some warnings have been issued about being careful not to confuse a competition with research [7], a recent trend advocates for recasting robotics challenges and competitions as experiments [3] and benchmarks [4]. Adopting this view, several competitions are currently trying to provide ways to compare the performance of different robot systems. For instance, in the field of home-assistant robots, the RoboCup@Home competition [13] evaluates robot systems in domestic environments. In the field of urban search and rescue, the Multi Autonomous Ground-Robotic International Challenge [14], the RoboCup Rescue Robot League [17], and the Defense Advanced Research Projects Agency Robotics Challenge [1] assess and measure the capabilities of different types of robots in real disaster environments.

The approach promoted by the RoCKIn project (http://rockinrobotchallenge.eu) aims to move from competitions that provide benchmarking at the system-level based on a single high-level measure to more sophisticated benchmarking activities. The RoCKIn competitions come close to scientific experiments as they provide a rigorous performance evaluation of robot systems under controlled and reproducible conditions. More precisely, the competitions adopt the classical conceptual framework of scientific experimental methods that separates reproducibility from repeatability [11] and using the results from the first RoCKIn competition, we show how RoCKIn can provide a set of tools to enable the replicability of experiments involving autonomous robots.

Replicability: Reproducibility and Repeatability

Among the principles that characterize the scientific experimental method, replicability is considered fundamental to allow for rigorous comparison of results and thus affects the processes and products of scientific research. The concept of replicability has emerged as central to the debate within autonomous robotics to make its methods closer to the standard of rigor of other scientific disciplines [6]. Usually, the replicability in robotics research is intended as the possibility to reproduce published results. However, the issue is more complex and problematic, as it has been recognized in other fields of computer science [9]. Therefore, to better articulate this concept and the contributions of the RoCKIn approach, we take into account the traditional conceptualization as devised in the history and the philosophy of science. Accordingly, the replicability can be specified into reproducibility and repeatability. Although they both refer to the general idea that scientific results should undergo the most severe criticisms to be strongly confirmed, they indeed point out two distinct characteristics of experimental methodology [11].

- *Reproducibility* is the possibility to verify, in an independent way, the results of a given experiment. It refers to the fact that other experimenters, different from the ones claiming validity for some results, are able to achieve the same results by starting from the same initial conditions, using the same type of instruments and parameters, and adopting the same experimental techniques. To be reproducible, an experiment must be fully documented.
- *Repeatability* concerns the fact that a single result is not sufficient to ensure the success of an experiment. A successful experiment must be the outcome of a number of trials, possibly performed at different times and in different places. These requirements guarantee that results have not been achieved by chance but are systematic, and that statistically significant trends can be identified.

How can these two very general features be applied in the practice of robotics research, and, in particular, of robotics competitions?

To answer this question, we force an artificial separation between the two intertwined concepts. Concerning reproducibility, the need for a precise description of results and of the processes adopted to achieve those results calls for the following requirements:

- conducting reproduced experiments in the same settings of the original ones that, therefore, should be explicitly and fully specified to be exactly reproduced
- making the used code and data available to the research community; note, however, that the mere availability of code and data does not guarantee reproducibility. The source code and test data have to be available, the code has to build, the execution environment has to be replicated (including the robot system or part of it), the code has to run to completion, and accurate measurements have to be collected [8].

Concerning repeatability, the mere repetition of runs is surely a way to attain repeatability, but it is only the first step in the direction of achieving systematic results. Given that one of the goals in making experiments is to obtain generalizations, repeatability can be practically achieved by the following:

- conducting a serious analysis of how many runs of a robotic experiment are required to obtain statistically significant results [16]
- performing experimental sessions that take place in settings S' that are fully compatible with the description of original settings S, but that might slightly differ for some details left unspecified in S. This contributes to filter out casual issues that affect experimental outcomes.

The enactment of the above experimental requirements to competitions is the basis of the approach followed in RoCKIn.

Overview of the RoCKIn Approach

The RoCKIn project aims to provide tools for benchmarking to the robotics community by designing and setting up competitions that increase scientific and technological knowledge. The RoCKIn competitions retain the traditional value of producing a ranking among alternative solutions at competition time, assigning prizes and awards to the best teams, and stimulating progress. At the same time, the experimental settings of the competitions gain a more general significance as benchmarking procedures. The RoCKIn project moves from competitions providing benchmarking with a single system-level measure (like the score of a soccer game) to a more sophisticated benchmarking approach integrated within competitions, where different elements are evaluated and benchmarking results can be used not only to rigorously compare robot systems, but also to better understand them. According to this perspective, we could say that the RoCKIn competitions come close to scientific experiments by providing an objective performance evaluation of a robot system/subsystem under controlled and reproducible conditions.

Two challenges have been selected as competition scenarios in this project due to their high relevance and impact on Europe's societal and industrial needs: domestic service robots (RoCKIn@Home) and innovative robot applications in industry (RoCKIn@Work). Both challenges have been inspired by similar activities in the RoboCup community [15], [20]. The RoCKIn aims at exploiting some of the RoboCup achievements to extend the pure competition approach in several aspects, as summarized in Table 1.

In RoCKIn@Home [18], Granny Annie lives in an apartment together with some pets and presents some of the typical problems of aging people. The aim of RoCKIn@Home is to develop robots that support Granny Annie and her quality of life. The RoCKIn@Home test bed reflects an ordinary European apartment with all common household items like windows, doors, furniture, and decorations.

The RoCKIn@Work scenario [10] represents a medium-sized factory that specializes in the production of small- to medium-sized lots of mechanical parts and assembled mechatronic products, which tries to optimize its production process to meet the increasing demands of their customers. This factory thus requires a system with two essential capabilities: 1) mobile manipulation to perform tasks such as assembly processes, quality controls, order handling, and logistics and 2) autonomy in switching between different tasks. The RoCKIn@Work test bed also includes networked devices such as force fitting machines and conveyor belts, which can be operated by the robots themselves.

The RoCKIn competitions come close to scientific experiments as they provide a rigorous performance evaluation of robot systems under controlled and reproducible conditions.

One of the main features of the RoCKIn competitions is the introduction of two separate classes of evaluations, task benchmarks (TBMs) and functionality benchmarks (FBMs). The former are devoted to evaluating the performance of integrated robot systems, while the latter focus on the performance of specific subsystems (like object recognition and

Table 1. The shift from RoboCup to RoCKIn.

From RoboCup …	… To RoCKIn
Adopts a pure competition approach with a (mostly) monofaceted scoring of tasks	Adopts a more sophisticated competition approach with multifaceted scoring of both tasks and functionalities
Does not explicitly address benchmarking	Explicitly considers structured and repeatable benchmarking
Presents mostly passive environments	Integrates sensors and actuators in the environment and wirelessly networks them with mobile robots

speech understanding). A TBM deals with complete robot systems, implying that a large set of interacting robot subsystems (like navigation, perception, and manipulation) are examined together at the same time. FBMs, on the other hand, focus on the performance of single subsystems, defining a precise setup in which a single robot functionality can be evaluated. This evaluation is performed according to well-specified quantitative measures and criteria, which depend on the functionality being tested. The scoring of TBMs and FBMs is then used to determine rankings of teams and to award prizes at the competitions.

The RoCKIn approach builds on the efforts of the RoboCup community to identify which functionalities are stable and solved or unsolved, at least in the context of the selected competitions. However, the evaluation of these functionalities in the RoboCup is mixed with the evaluation of the tasks, and separating the two is difficult. "In fact, teams obtained good results in navigation, mapping, person tracking, and speech recognition (with the average above 50%, except for navigation). Notice that the reason for a low-percentage score in navigation is not related to inabilities of the teams, but it is because part of the navigation score is only available after some other task was achieved [13]." The RoCKIn approach avoids this problem by limiting the influence of all other subsystems when evaluating a robot functionality in a FBM. For example, for testing object perception, robots are put in place before starting the test. More generally, we mitigate the difficulty of separating subsystems under investigation from their environments, which also include other robot subsystems that are not being evaluated, by carefully designing FBMs that minimally involve these last subsystems.

The replicability can be specified into reproducibility and repeatability.

Table 2. The teams participating in the 2014 RoCKIn competition.

RoCKIn@Home Teams
b-it-bots@Home, Bonn-Rhein-Sieg University of Applied Sciences, Germany
BARC, University of Birmingham, United Kingdom
Homer@UniKoblenz, University of Koblenz-Landau, Germany
Pumas@Home, Universidad Nacional Autonoma de Mexico, Mexico
SocRob@Home, Universidade de Lisboa, Portugal
UrsusTeam, University of Extremadura, Spain
Watermelon Project, University of Leon, Spain
RoCKIn@Work Teams
b-it-bots@Work, Bonn-Rhein-Sieg University of Applied Sciences, Germany
IASLab@Work, University of Padua, Italy
SPQR@Work, Sapienza University of Rome, Italy

The TBMs and FBMs are evaluated differently. The TBMs measure the achievement of goals, which is a yes or no answer to specific questions (e.g., "Does the robot understand Annie's command(s)? Does it correctly identify the requested object?"). The FBMs measure robot performance, which is a number resulting from the measures used for scoring and ranking, such as effectiveness (e.g., precision and recall) and efficiency (time, resources used, and so on), as further discussed in the next section. This division resembles the recently proposed evaluation of artificial intelligence systems [12], which is based on task-oriented and ability-oriented evaluations. In both the cases, the RoCKIn approach tries to avoid subjective evaluation to improve reproducibility. Indeed, attention has been dedicated to requirements for benchmarking and scoring runs as autonomously as possible (i.e., without continuous human intervention), specifically by using automated computing systems called *RefBoxes* (also called *Central Factory Hub* in RoCKIn@Work). In this respect, the more automated scoring approach of RoCKIn contrasts, for instance, with that of the RoboCup Rescue Robot League [17], which is heavily based on human judges.

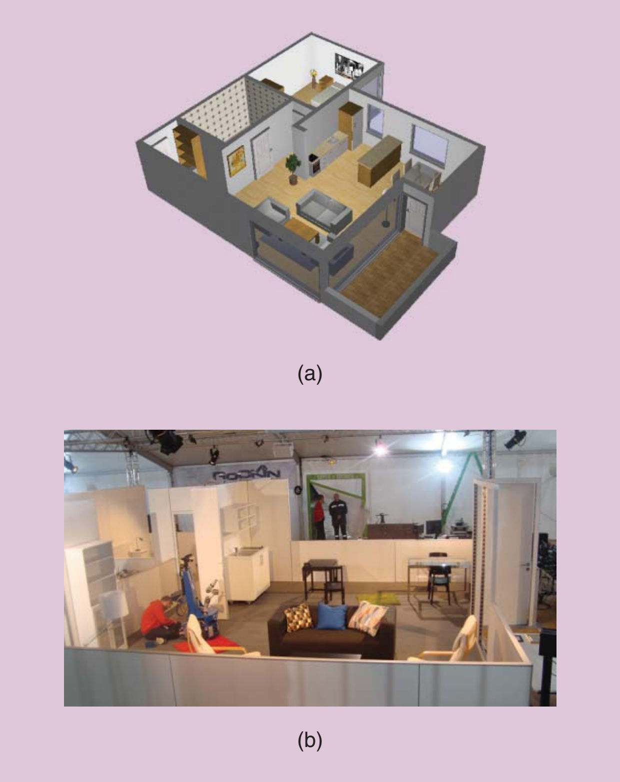
In addition to ranking teams in the competitions, the approach of RoCKIn promises to be a good way to understand robot systems because it enables researchers to study the impact of functionality performance on task performance. Moreover, it forces teams to develop means of continuously monitoring the performance of their robot systems because they have to provide regular feedback to the RefBoxes and store data for benchmarking.

The First RoCKIn Competition

The first RoCKIn competition (http://rockinrobotchallenge.eu/rockin2014.php) was held in Toulouse, France, 26–30 November 2014, and was the first opportunity to test the practical application of the approach outlined in the previous section and to prepare for the final RoCKIn competition (http://rockinrobotchallenge.eu/rockin2015.php) to be held at the end of 2015 in Lisbon, Portugal. The teams participating in the 2014 event are listed in Table 2.

Setup for the Competition

The detailed specifications of the RoCKIn@Home and RoCKIn@Work test beds are reported in the corresponding rule books, which are available at http://rockinrobotchallenge.eu/publications.php under "deliverables and reports" and allow for their precise reproduction at other sites than those of the competitions. The layouts and the sizes of the RoCKIn@Home and RoCKIn@Work arenas (Figures 1 and 2) are fully specified, together with the materials of walls and the precise definition of objects present in the environments (at a level of detail that include furniture and floristic objects for RoCKIn@Home). The robots must conform to certain size, weight, and safety restrictions and can be wirelessly networked with other devices. Apart from this, teams are free to



(a)

(b)

Figure 1. The RoCKIn@Home arena: Granny Annie's apartment. (a) The 3-D model of the RoCKIn@Home arena and (b) the real RoCKIn@Home arena built for the first RoCKIn competition.

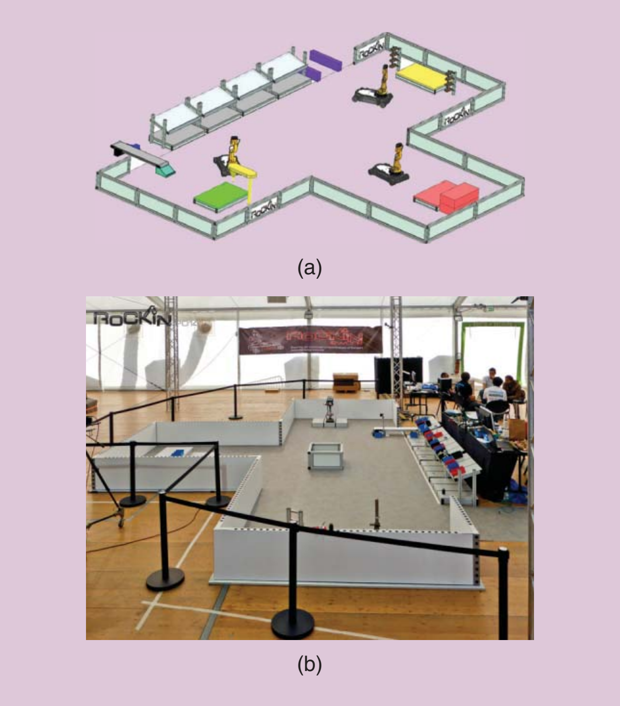
choose the robot platforms they deem most adequate to obtain the best performance. Shortly, with the idea to attain reproducibility and repeatability, the RoCKIn precisely specifies several aspects of the settings S in which the competitions take place, but some aspects, like the sensor equipment of the robots, are left unspecified. In this way, if a robot capability is demonstrated in S and in other settings S' that differ from S for the actual implementation of the aspects unspecified in S, it can be concluded that the capability is stable or solved.

TBMs and FBMs defined for the 2014 edition of the RoCKIn@Home and RoCKIn@Work challenges are listed in Table 3.

Scoring of TBMs

The scoring of TBMs is based on achievements and penalties. Specifically, performance classes C_n are defined for ranking robot performance in a task, based on the number (n) of achievements (A) that the robot reaches during the execution of the task. Within each performance class C_n (i.e., the number of achievements being equal), ranking is defined according to the number of penalty behaviors (PB) of the robot that represent errors while executing the task. More formally, given a task, the following rules are applied:

- the ranking of any robot belonging to performance class C_n is considered better than that of any robot belonging to performance class C_m with $m < n$; class C_0 is the worst performance class



(a)

(b)

Figure 2. The RoCKIn@Work arena: a model of a medium-size factory.

Table 3. TBMs and FBMs for the 2014 RoCKIn competition.

	RoCKIn@Home
TBM1@Home	Getting to know my home
TBM2@Home	Welcoming visitors
TBM3@Home	Catering to Granny Annie's comfort
FBM1@Home	Object perception
FBM2@Home	Object manipulation
FBM3@Home	Speech understanding
	RoCKIn@Work
TBM1@Work	Assemble aid tray for force fitting
TBM2@Work	Plate drilling
TBM3@Work	Prepare box for manual assembly step
FBM1@Work	Object perception
FBM2@Work	Visual servoing

- among robots belonging to the same performance class, the robot which received fewer penalties is considered higher in rank
- among robots belonging to the same performance class and with the same number of penalties, the ranking of the robot that completed the task in a shorter time is considered higher.

Moreover, to ensure the safety of the competition, disqualifying behaviors (DB) are defined, namely the things that a robot must not do to avoid being excluded from the competition. For security reasons, a human referee always has access

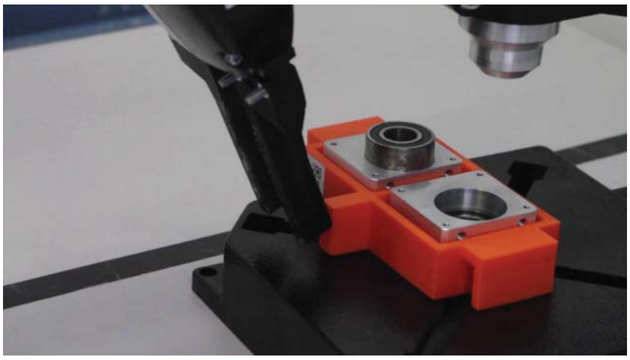


Figure 3. A robot assembling an aid tray for force fitting (TBM1@Work). The picture comes from the RoCKIn consortium.

Figure 4. The objects used in the FBM1@Home object perception. The picture comes from the RoCKIn consortium.

Table 4. The results of the FBM1@Home object perception.

Team	Object Class Accuracy
UrsusTeam	0.90
Homer@UniKoblenz	0.80
Pumas@Home	0.70
Watermelon Project	0.30

to a red button that, if pushed, stops the robot in case of disqualifying behaviors.

Each team had the possibility of performing five runs of each TBM in RoCKIn@Home (two runs on the first day, two runs on the second day, and one run on the last day) and three runs of each TBM in RoCKIn@Work (one run per day). The best score over all runs determined the winner for the TBM.

The recorded data depend on the hardware equipment of the robots.

One key property of this scoring system is that a robot that executes the task completely will always be ranked better than a robot that executes the task partially. Penalty behaviors do not change the performance class of a robot and only influence intraclass ranking. It is also possible to envisage the use of weighted penalties; however, this makes the ranking criteria harder to understand and apply. Therefore, weights have not been used in the 2014 RoCKIn competition.

Sets A, PB, and DB are task-dependent. For example, for the TBM3@Home, catering to Granny Annie's comfort, the sets are as follows (other TBMs sets are defined similarly, please refer to the rule books for full details).

- A = {upon reception of a call signal, the robot enters the room where Granny Annie is waiting, the robot understands the commands uttered by a person playing the role of Granny Annie and the robot operates the right devices requested by Granny Annie, the robot finds the right objects, the robot brings Granny Annie the right objects}.
- PB = {the robot bumps into furniture, the robot drops an object, the robot stops working}.
- DB includes, for example, hitting Granny Annie.

Scoring TBMs in the RoCKIn@Work is performed similarly. For instance, achievements for the TBM1@Work assemble aid tray for force fitting task include the correct identification of the assembly aid tray and the correct delivery of the aid tray to the force fitting machine (see Figure 3). Similarly, PBs are defined, like dropping an object or bumping into obstacles.

Overall, the performance of the teams in the TBMs of the first RoCKIn competition has been good for the achievements related to navigation, while other achievements have proved to be more challenging. Full results are available at http://rockinrobotchallenge.eu/rockin2014.php under "Results."

Scoring of FBMs

The scoring of FBMs measures the performance of robot subsystems and is specific for each functionality. For example, consider the FBM1@Home object perception (other FBMs are scored similarly), which is used to assess the capabilities of a robot in identifying the class (e.g., cups), the instance (e.g., black coffee mug), and the pose (with respect to a global coordinate system) of objects that are presented to it and that are relevant to the TBMs of RoCKIn@Home (Figure 4 shows a sample of the possible objects). Each team that participated in this FBM had the opportunity to perform the benchmark four times, each time trying to identify ten randomly selected objects placed in random poses on a flat table (which was located at a fixed and known pose in the global coordinate system). The scoring considers the accuracy in class classification (and, in case of tie, the accuracy in instance classification, the error rate in identifying the three-dimensional pose of the object, and the test time, in this order). The best score over the four runs is considered for the final ranking. Final results for the FBM1@Home object perception are reported in Table 4.

The FBM3@Home speech understanding has the goal of evaluating the ability of robot systems to understand speech commands that a user (like Granny Annie) gives in a home

Figure 5. The robots during the FBM3@Home speech understanding.

environment (like go to the living room, put the jar on the table). Five teams participated in this FBM. Each team had the opportunity to perform four runs, each time trying to understand a number of command sentences provided to the robots both as audio files (between 30 and 50) on a USB stick and as sentences (from 6 to 12) spoken by a person through a microphone (to retain reproducibility, the person speaking was always the same for all teams and runs). A single loudspeaker placed on the ground and facing up was used to produce an omnidirectional audio source so that the robots were able to perform the test in parallel (see Figure 5).

The scoring considers the ability to recognize the main actions (like go and put) and the main arguments (like living room, jar, and table) of the commands. This is measured in terms of the accuracy in correctly classifying the arguments (AgC), the accuracy in correctly classifying the actions (AcC), and the word error rate (WER) in correctly recognizing each word. Ranking is based on AcC, AgC, and WER, in this order. A total of 15 valid team runs (i.e., with nonzero performance) were performed. Full results for the FBM3@Home speech understanding are reported in Table 5. The difficulty of the sentences (evaluated by an expert in speech recognition) was increasing during the first three runs, while the fourth run contained mixed sentences. This is reflected in the performance of the robots over the runs.

In a similar manner, for RoCKIn@Work, FBM1@Work (object perception) and FBM2@Work (visual servoing) are defined. The former focuses on the detection, recognition, and localization of industrial objects, where the latter focuses on controlling the manipulator motion based on its own visual perception.

Note that the scoring of TBMs and FBMs relies on the RefBoxes that support detecting achievements and penalties and measure the performance of the robots (e.g., the time spent in performing an activity). In particular, the RefBoxes can largely automate the evaluation of FBMs. For example, in the FBM1@Home object perception, the RefBox randomly selects which objects will be presented to the robots, sends the start signal to the robots, and waits for replies from the robots. Moreover, for TBMs the RefBoxes manage the communication between the test bed and the robots, mediating between the environment devices and the robots. This helps to identify if devices (like force fitting machine in RoCKIn@Work) are correctly actuated.

Table 5. The results of the FBM3@Home speech understanding.

Run 1	**AgC**	**AcC**	**WER**
UrsusTeam	0.28	0.76	0.47
b-it-bots@Home	0	0	0.70
Homer@UniKoblenz	0	0	0.74
Run 2	**AgC**	**AcC**	**WER**
UrsusTeam	0.24	0.65	0.53
Pumas@Home	0.07	0.46	0.59
b-it-bots@Home	0.05	0.11	0.94
Homer@UniKoblenz	0	0.37	0.70
Run 3	**AgC**	**AcC**	**WER**
UrsusTeam	0.03	0.62	0.50
Pumas@Home	0.03	0.18	0.59
b-it-bots@Home	0	0.43	0.75
Homer@UniKoblenz	0	0.34	0.76
Run 4	**AgC**	**AcC**	**WER**
UrsusTeam	0.10	0.71	0.47
Pumas@Home	0.08	0.35	0.74
b-it-bots@Home	0.01	0.30	0.72
Watermelon Project	0	0	0.69

Benchmarking and Replicability

During the RoCKIn competitions, we plan to collect data for benchmarking that go beyond those strictly needed for scoring the runs of the robots. Benchmarking data are acquired both by the robots and by devices in the environment.

For example, in TBM3@Home, catering to Granny Annie's comfort, the following data were expected to be collected for each run of every team: the audio signals of the conversations between Granny Annie and the robot (collected by the robot), the final commands produced after the natural language analysis process (collected by the robot), the ground truth pose of the robot while moving in the environment (collected using the OptiTrack motion capture system by NaturalPoint), the pose of the robot while moving in the environment (as perceived by the robot), the sensorial data of the robot when recognizing the object to be operated, and the results of the robot's attempts to execute Granny Annie's commands.

The scoring of TBMs is based on achievements and penalties.

For the FBM1@Home object perception, expected benchmarking data include sensor data (images, point clouds, and so on) used by the robot to perform classification; the class; the instance, and the pose of every object (as determined by the robot); and the actual class, instance, and pose of every object (ground truth). For the FBM3@Home speech understanding, benchmarking data that were expected to be collected include sensor data (audio files) used by the robot to perform speech recognition and the command (action and arguments) as recognized by the robot. Similar rich benchmarking data were expected to be collected for all other FBMs and TBMs. Note that benchmarking data include ground truth, for example, the poses of the robots and objects and the commands issued to the robots.

For the participating teams, the recording of sensor data and processed information is mandatory, although some flexibility has been allowed during the first RoCKIn competition. Since the process is rather invasive and it turns out that most teams use ROS (http://www.ros.org), we tried to limit the effort for onboard data collection by using the ROS built-in recording tool called rosbag (which can also be used by teams not using ROS, by exploiting the rosbag Application Programming Interfaces). Note that the recorded data depend on the hardware equipment of the robots. For example, data collected during the FBM1@Home object perception include both images and images plus point clouds of the same objects, according to the different sensors mounted on different robots. In principle, stereo images could also be present. The amount of benchmarking data collected over all the runs of the TBMs and FBMs on the three days of the 2014 competition is summarized in Table 6. A positive trend is evident as the competition progressed, from 43% of runs (10 out of 23 runs) with complete benchmarking data on the first day, to 91% of runs (20 out of 22 runs) with complete benchmarking data on the last day, which was a half-day competition. This is due to increased awareness about data collection. Globally, 68% of runs (52 out of 76 runs) have complete benchmarking data. Incomplete benchmarking data are due to their incorrect format or to missing portions. Note that the runs with no benchmarking data also include runs in which the robots failed to start, which were 4, 3, and 0, on the three days, respectively.

The robots must conform to a certain size, weight, and safety restrictions and can be wirelessly networked with other devices.

Table 6. The benchmarking data collected during the 2014 RoCKIn competition.

	Day 1	Day 2	Day 3
Total runs	23	31	22
Runs with complete data	10	22	20
Runs with incomplete data	2	1	1
Runs with no data	11	8	1

These benchmarking data are made available to the research community, to ease the reproducibility of results and the comparison with the teams participating in the RoCKIn competitions. The benchmarking data can be found at http://thewiki.rockinrobotchallenge.eu/. In particular, data relative to poses of robots collected by the ground truth system can be used by the teams to replay the runs of their robots, for example, matching the actual pose of a robot with the expected one according to the robot perception. As some anecdotal evidence from the 2014 competition confirmed, this can have a positive impact on fixing bugs and improving the performance of teams. The RoCKIn competitions aid in collecting a huge amount of data that can be later used for reproducing experiments and for benchmarking by researchers not participating in the competitions, as researchers can download the data sets and run their algorithms on them. For example, laser range scanner data collected during task benchmarks can also be employed to test and evaluate mapping and localization algorithms, while audio files collected during the FBM3@Home speech understanding can be used to test algorithms for speech understanding. Researchers can also compare their results with those obtained by teams in the RoCKIn competitions. In this sense, the availability of data recorded by the robots with different configurations while performing the same task or functionality benchmark enrich the data sets provided by the RoCKIn.

Some steps toward the repeatability of experiments in the context of a competition were taken: each team was given the option of repeating all the TBMs and FBMs at least three times (although some teams performed less runs due to robot failures or the decision to skip them). Selecting the best-scored run makes sense for the competition ranking (see Table 4), but the results over all the runs could be considered for a statistical analysis of the significance of the observed differences in performance. However, the data from the first RoCKIn competition are not enough to support such a statistical analysis yet. We are working on teaching the teams to use RoCKIn benchmarking infrastructure more systematically in the 2015 competition.

Conclusions and Future Works

With this article, we have pointed out how the RoCKIn approach to competitions makes them closer to replicable scientific experiments, as the benchmarking procedures we defined can provide a rigorous and articulated performance evaluation of the robot systems under controlled circumstances. By taking inspiration from the history and philosophy of science, we have articulated replicability into reproducibility and repeatability, and suggested how to apply them in the practice of robotic research. From the analysis of some results from the 2014 RoCKIn competition, we can say that the RoCKIn approach contributes to enabling the reproducibility of experimental results by providing full details to reproduce test beds and by collecting rich benchmarking data. As for repeatability, while the structure of the RoCKIn competitions pushes in this direction, the results of the 2014 RoCKIn competition are still too preliminary to draw any conclusion.

Future work will address the situations that the current version of the RefBoxes cannot manage, like detecting if a robot has hit something or someone or has correctly grasped an object, to make scoring even more automatic. More generally, we will promote the further development of the RoCKIn approach, whose final competition is planned for the end of 2015, toward fully reproducible experiments. It is expected that enough teams will participate to get a significant amount of data that will enable a systematic and quantitative analysis of robot performance, also relative to the evaluation of the importance of single functionalities in the execution of complex tasks. In this respect, we plan to investigate the use of some tools from game theory, like Shapley values and power indexes [2].

References

[1] (2015). DARPA robotics challenge. [Online]. Available: http://www.theroboticschallenge.org

[2] A. Ahmad, I. Awaad, F. Amigoni, J. Berghofer, R. Bischoff, A. Bonarini, R. Dwiputra, G. Fontana, F. Hegger, N. Hochgeschwender, L. Iocchi, G. Kraetzschmar, P. Lima, M. Matteucci, D. Nardi, V. Schiaffonati, and S. Schneider. (2014). RoCKIn deliverable D1.2 General evaluation criteria, modules and metrics for benchmarking through competitions. [Online]. Available: http: //rockinrobotchallenge.eu/publications.php

[3] M. Anderson, O. Jenkins, and S. Osentoski, "Recasting robotics challenges as experiments," *IEEE Robot. Automat. Mag.*, vol. 18, no. 2, pp. 10–11, 2011.

[4] S. Behnke, "Robot competitions—Ideal benchmarks for robotics research," in *Proc. IROS Workshop Benchmarks Robotics Research*, 2006.

[5] P. Bonasso and T. Dean, "A retrospective of the AAAI robot competitions," *AI Mag.*, vol. 18, no. 1, pp. 11–23, 1997.

[6] F. Bonsignorio, J. Hallam, and A. del Pobil. (2007). Special interest group on good experimental methodology—GEM guidelines. [Online]. Available: http://www.heronrobots.com/EuronGEMSig/downloads/GemSigGuidelinesBeta.pdf

[7] T. Bräunl. "Research relevance of mobile robot competitions," *IEEE Robot. Automat. Mag.*, vol. 6, no. 4, pp. 32–37, 1999.

[8] C. Collberg, T. Proebsting, G. Moraila, A. Shankaran, Z. Shi, and A. Warren. (2014). Measuring reproducibility in computer systems research. [Online]. Available: http://reproducibility.cs.arizona.edu/tr.pdf

[9] C. Drummond, "Replicability is not reproducibility: Nor is it good science," in *Proc. Evaluation Methods Machine Learning Workshop 26th ICML*, 2009.

[10] R. Dwiputra, J. Berghofer, A. Ahmad, I. Awaad, F. Amigoni, R. Bischoff, A. Bonarini, G. Fontana, F. Hegger, N. Hochgeschwender, L. Iocchi, G. Kraetzschmar, P. Lima, M. Matteucci, D. Nardi, V. Schiaffonati, and S. Schneider, "The RoCKIn@Work challenge," in *Proc. ISR/Robotik*, 2014, pp. 328–333.

[11] I. Hacking. *Representing and Intervening*. Cambridge, U.K.: Cambridge Univ. Press, 1983.

[12] J. Hernández-Orallo. (2014). AI evaluation: Past, present and future. [Online]. Available: http://arxiv.org/abs/1408.6908

[13] D. Holz, L. Iocchi, and T. van der Zant, "Benchmarking intelligent service robots through scientific competitions: The RoboCup@Home approach," in *Proc. AAAI Spring Symp. Designing Intelligent Robots: Reintegrating AI II*, 2013, pp. 27–32.

[14] A. Hsieh and S. Lacroix, Eds., "Special issue on multiautonomous ground-robotic international challenge (MAGIC)," *J. Field Robot.*, vol. 29, no. 5, pp. 687–841, 2012.

[15] G. Kraetzchmar, N. Hochgeschwender, W. Nowak, F. Hegger, S. Schneider, R. Dwiputra, J. Berghofer, and R. Bischoff, "RoboCup@Work: Competing for the factory of the future," in *Proc. RoboCup Symp.*, 2014.

[16] J. Parker, J. Godoy, W. Groves, and M. Gini, "Issues with methods for scoring competitors in RoboCup rescue," in *Proc. AAMAS Workshop Autonomous Robots Multirobot Systems*, 2014.

[17] J. Pellenz, A. Jacoff, T. Kimura, E. Mihankhah, R. Sheh, and J. Suthakorn, "RoboCup rescue robot league," in *Proc. RoboCup Symp.*, 2014.

[18] S. Schneider, F. Hegger, A. Ahmad, I. Awaad, F. Amigoni, J. Berghofer, R. Bischoff, A. Bonarini, R. Dwiputra, G. Fontana, L. Iocchi, G. Kraetzschmar, P. Lima, M. Matteucci, D. Nardi, and V. Schiaffonati, "The RoCKIn@Home challenge," in *Proc. ISR/Robotik*, 2014, pp. 321–327.

[19] B. Smart, "Competitions, challenges, or journal papers," *IEEE Robot. Automat. Mag.*, vol. 19, no. 1, p. 14, 2012.

[20] T. Wisspeintner, T. van der Zant, L. Iocchi, and S. Schiffer, "RoboCup@Home: Scientific competition and benchmarking for domestic service robots," *Interaction Studies*, vol. 10, no. 3, pp. 392–426, 2009.

Francesco Amigoni, Politecnico di Milano, Italy. E-mail: francesco.amigoni@polimi.it.

Emanuele Bastianelli, University of Rome La Sapienza, Italy. E-mail: emanuele.bastianelli@gmail.com.

Jakob Berghofer, KUKA Laboratories GmbH, Augsburg, Germany. E-mail: Jakob.Berghofer@kuka.com.

Andrea Bonarini, Politecnico di Milano, Italy. E-mail: andrea.bonarini@polimi.it.

Giulio Fontana, Politecnico di Milano, Italy. E-mail: giulio.fontana@polimi.it.

Nico Hochgeschwender, Bonn-Rhein-Sieg University, Germany. E-mail: nico.hochgeschwender@h-brs.de.

Luca Iocchi, University of Rome La Sapienza, Italy. E-mail: iocchi@dis.uniroma1.it.

Gerhard K. Kraetzschmar, Bonn-Rhein-Sieg University, Germany. E-mail: gerhard.kraetzschmar@brsu.de.

Pedro Lima, Instituto Superior Técnico, University of Lisbon, Portugal. E-mail: pal@isr.tecnico.ulisboa.pt.

Matteo Matteucci, Politecnico di Milano, Italy. E-mail: matteo.matteucci@polimi.it.

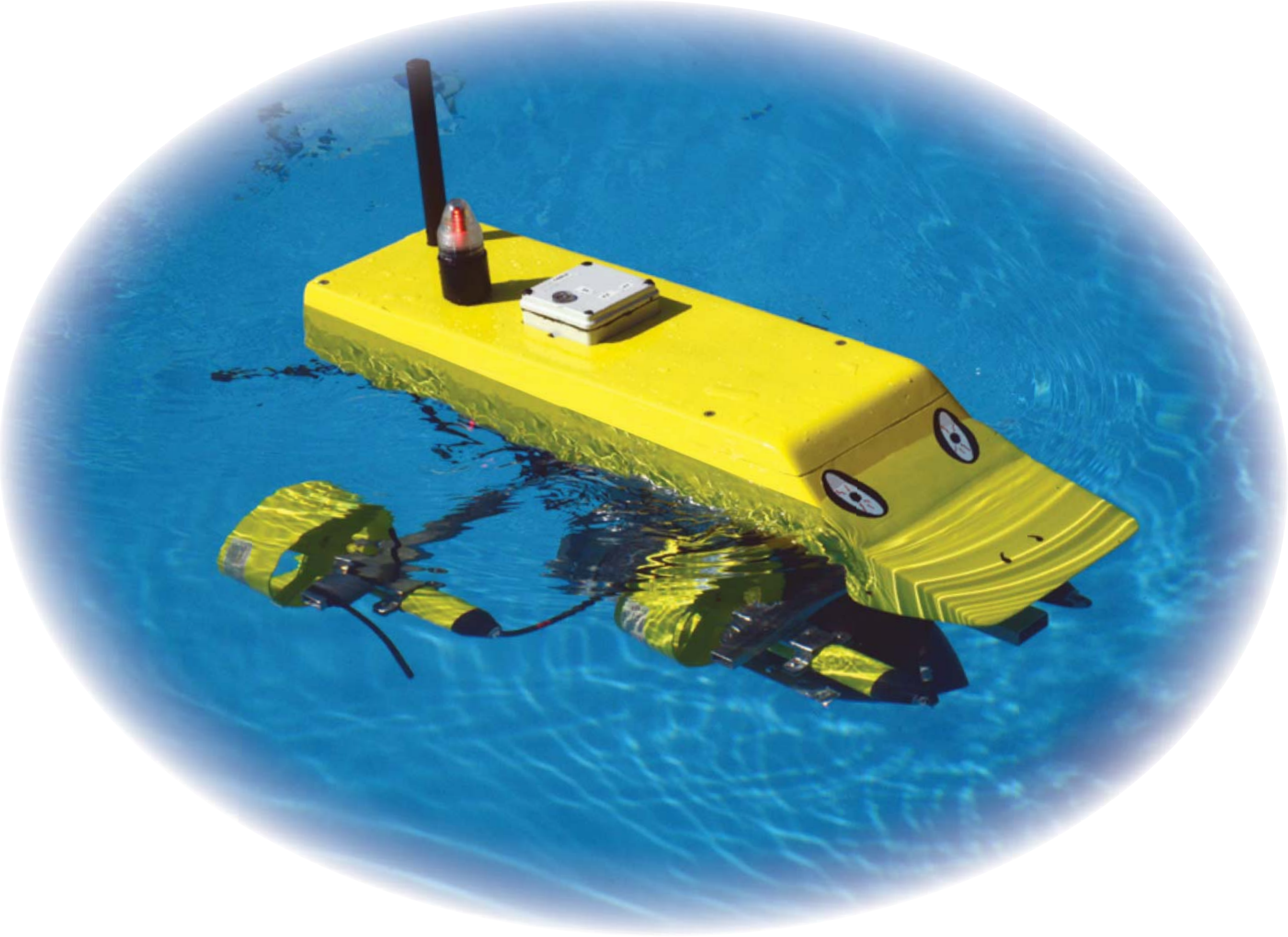
Pedro Miraldo, Instituto Superior Técnico, University of Lisbon, Portugal. E-mail: pmiraldo@isr.tecnico.ulisboa.pt.

Daniele Nardi, University of Rome La Sapienza, Italy. E-mail: nardi@dis.uniroma1.it.

Viola Schiaffonati, Politecnico di Milano, Italy. E-mail: viola.schiaffonati@polimi.it.

Testing the Waters

Design of Replicable Experiments for Performance Assessment of Marine Robotic Platforms



By Andrea Sorbara, Andrea Ranieri, Eleonora Saggini, Enrica Zereik, Marco Bibuli, Gabriele Bruzzone, Eva Riccomagno, and Massimo Caccia

Digital Object Identifier 10.1109/MRA.2015.2448311
Date of publication: 18 August 2015

Robotics is becoming part of our daily life through home automation (called *domotic*) and assistive applications that are taking place in habitations, up to industrial and service employments. With this strict cooperation between humans and robots, it is of absolute necessity to assess the robotic systems' capabilities and performance to ensure safety and reliability according to actual and significant criteria and parameters. Within this context, there are delays in technology transfer from research frameworks toward actual applicative scenarios. These delays are caused by the lack of methodologies and standardized procedures for the experiment execution and result comparison. A further restraint to the needed process of technology transfer is due to the inability to understand the actual capabilities of the systems and to be aware and confident (or not) about what they can be realistically employed for. In many cases, this can also turn into the inability to define specific regulations for the employment of robotic platforms to comply with the current laws. This is the case for marine robotics. Tethered vehicles [e.g., remotely operated vehicles (ROVs)]

1070-9932/15©2015IEEE

are considered part of the vessel equipment when they are operated from a ship (because of the physical connection provided by the cable). Conversely, whenever autonomous systems, such as unmanned surface vehicles (USVs) or autonomous underwater vehicles are commanded through an acoustic or Wi-Fi/radio communication link, they are considered as navigating platforms and are subject to the regulation in force. For the commercial employment of these vehicles, it is then necessary to strongly assess the vehicle's capabilities in a standardized way, to guarantee the operating performance and evaluate limitations of the target system.

Toward Replicability in Robotics

The need for an assessment methodology is fueling a discussion within the theoretical and the applied research community on how to quantitatively measure the performance of robotic systems, with particular attention to experimental method. Some examples include the establishment of interest groups [e.g., Euron Good Experimental Methodology (GEM) Special Interest Group], technical committees (e.g., IEEE Robotics and Automation TCPEBRAS16), and dedicated workshops [e.g., the Workshop on the Path Toward Replicable Experiments in Robotics Research at the IEEE International Conference on Robotics and Automation (ICRA) 2011, the Workshop on the Conditions for Replicable Experiments and Performance Comparison in Robotics Research at ICRA 2012, and several workshops organized at the IEEE/RSJ International Conference on Intelligent Robots and Systems in the last years]. Furthermore, a widespread interest in defining suitable experimental methodologies is increasingly growing. This article aims to contribute to the research field by establishing standard procedures for experiment definition and design that allow replicability. In particular, current discussions within the robotics community focus on the definition of metrics and benchmarks for quantitative evaluation of performance, on protocols for executing replicable experiments, as well as on the availability of data sets and of common platforms for testing (see [1]–[5]). In fact, to obtain reproducibility for the test campaigns and to maximize the relevance of their output, guidelines for the design of experiments are strongly needed in the field of marine robotics since the field is affected by experimental constraints such as uncontrollability of the conditions (e.g., waves, sea currents, recreational and commercial traffic, and so on), a restricted number of executable experiments (due to cost and logistic issues), and uncertainty in the robot inputs (since hydrodynamic interactions, forces, and torques assigned to the system are inherently uncertain). In particular, there are many rules and common sense procedures to be followed in marine robotics to avoid accidents threatening the health and safety of humans, as well as to maintain the integrity of the robotic systems themselves (collisions, loss of autonomous vehicles, and so on). Using these objectives, the proposed work focuses on the definition of GEM and guidelines for experiment design. Furthermore, a software framework for automatic experiment design, execution, and result analysis is introduced. The proposed work also includes the computation of two performance indices for the quantitative evaluation of vehicle performance related to the path-following task for surface vessels.

State of the Art

A comprehensive review of issues related to measuring and comparing research results in robotics is provided in [6], which also includes considerations on how to define benchmarks in robotics and discusses the need to define benchmarks for specific subdomains of robotics (visual servoing, grasping, motion planning, and so on) rather than benchmarks valid for all domains. In [7], the issues on how to perform, replicate, and compare experiments are addressed specifically for robotic mapping, while [8] focuses on the definition of performance parameters to evaluate autonomy in navigation missions, specifically, robot performance related to environmental conditions. Fundamental guidelines for writing experimental papers have been formulated and circulated as a reviewers' checklist for the main robotics journals and conferences [1]. In [2], general guidelines are proposed to improve methodologies and reporting to facilitate experiment replication, performance evaluation, and comparison of robotics experiments. Topics on experimentation in mobile robot localization and mapping are discussed from an interdisciplinary viewpoint in [3], touching on issues that stand at the crossroad of mobile robotics and the philosophy of science. Furthermore, a detailed description of the main principles of the experimental methodology (comparison, reproducibility versus repeatability, and explanation) is given to encourage common language in benchmarking. In this article, definitions of replicability and repeatability given in [3] are adopted (the term reproducibility is used as a synonym of replicabillity). Here, the term replicability is applied to all controllable parameters of the experiments and not to its results. Performance metrics for response robots in search and rescue operations in disaster-stricken environments and the interactions of this article with the standardization committee E54.08.011 are presented in [4]. As previously stated in the introduction, application challenges for marine robotics is given by the operative condition constraints that have to be taken into account dealing with the problem of replicable experiments and quantitative performance assessment of autonomous marine vehicles. In this context, basic performance metrics have been defined for evaluating USV path-following performance in [5], where test repeatability was achieved moving along a straight path in opposite directions with U-turns. To evaluate USV path-tracking performance, Caccia et al. [9] report experiments where a USV was required to track prelogged target navigation data, sending them to the USV control system with suitable timing, and evaluating the performance metrics given by the normalized area between the actual and the desired path (spatial constraint) and curvilinear range error with respect to the desired position (time constraint). Field experiments for ROV and USV identification have been performed in [10] and [11], respectively, which discuss issues pertaining to the execution of experiments at sea, in the presence of significant constraints such as controllability of experimental conditions, a restricted number of executable experiments, and uncertainty in the inputs assigned to the system.

Early Results on Replicable Experiments

The work regarding GEM, carried out by the authors of this article, includes the definition of performance indices for path-following evaluation, as well as their exploitation to evaluate and compare different control algorithm' performance, as described in [12]. With the aim of defining standard procedures and metrics to assess robot behaviors, in terms of both system capabilities and safety, the great novelty proposed in this article consists in the design and implementation of a procedural methodology for experiment design and execution. The definition of such a methodology has led to the development of the DeepRuler software framework: such a system allows for the design, execution, and evaluation of path-following experiments, guaranteeing replicability and repeatability, as reported in the "Proposed Methodology and System Development" section.

Benchmarking Indices and Result Comparison

A list of performance indices for evaluating the accuracy and efficiency of a vehicle following a reference line can be found in [12]. Let R and V be the two sets of GPS coordinates representing the reference and the vehicle paths, respectively. The indices introduced in [12] and exploited in the "Early Simulation Results With DeepRuler" section and the "Experiments With DeepRuler" section for evaluating the accuracy of the path following trials are as follows.

- Area index, identified by D_A, is computed as the area between R and V, normalized by the total length of R.
- Hausdorff distance, identified by D_H, is the maximum of all the distances from a point in R to the closest point in V.

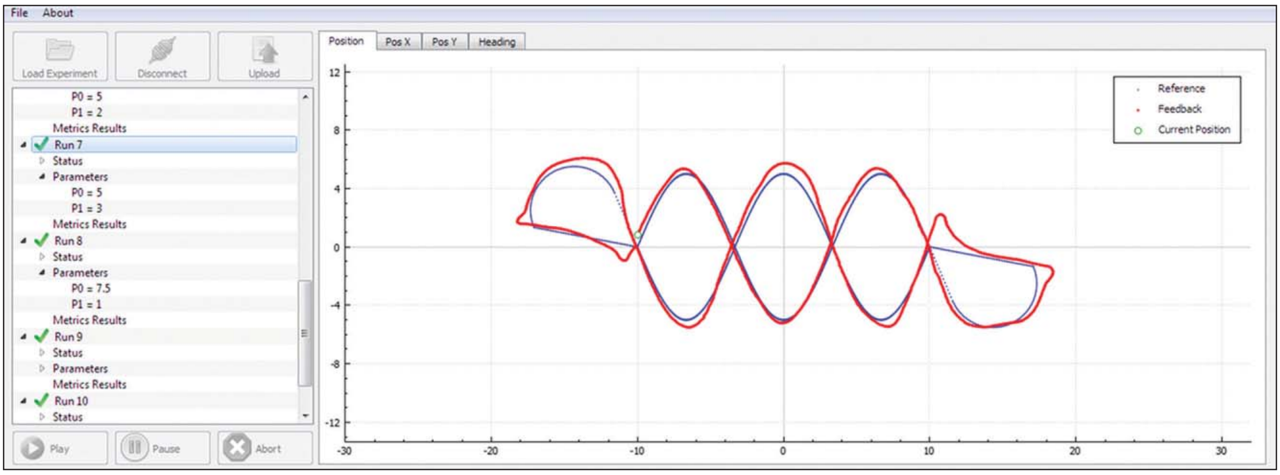
Details on the definition of such quantities and comments on the implementation are not reported here for the sake of brevity. It is worth noting the importance of adopting both the indices for a fair evaluation, since they provide different information: the area index and the Hausdorff distance represent the mean and the maximum value of the distance of the vehicle path from the reference path, respectively. Thanks to these indices, two different guidance schemes, initially designed for the Charlie USV, were compared. The two evaluated guidance modules are the Lyapunov-based virtual target and the Jacobian-based priority task. Details on these algorithms are omitted for the sake of brevity but can be found in [5] and [13], respectively. Despite the attention that has been paid to the execution of the experiments, some of the operating conditions are not controllable, and some related issues are reported in the following.

- *Initial Conditions*: It is difficult to ensure that each experiment begins exactly from the same position/orientation, and only the start from a suitable initial area can be guaranteed. This turns into an uncertainty that can lead to different behaviors during the approach phase to the path, with consequent different responses of the applied guidance system.
- *Environmental Conditions*: The conditions of the environment compromise the guidance module response in a completely uncontrollable way. For instance, though the reported trials have been performed in the same day, changing environmental conditions have been observed, such as the wind and sea currents blowing in different directions during the day.
- *Experiment Completion Condition*: On the date of the early experiments, the update of the reference path and/or conclusion of the current experiment was manually managed. Therefore, the experiment conclusion is strongly affected by the promptness of the human operator.

Details on the comparison can be found in [12].

Design of Replicable Experiments

At this point it should be clear that by relying on suitable benchmarking indices, robotic researchers are allowed to evaluate their experiments and to assess their control algorithm performance. However, the benchmarking indices alone do not achieve experimental replicability. In fact, data



P0 = 5
P1 = 2
Metrics Results
✓ Run 7
Status
Parameters
P0 = 5
P1 = 3
Metrics Results
✓ Run 8
Status
Parameters
P0 = 7.5
P1 = 1
Metrics Results
✓ Run 9
Status
Parameters
Metrics Results
✓ Run 10
Status

Figure 1. The DeepRuler HCI Player.

collected in the mentioned experiments stress the lack of equity in experimental conditions. These issues affect the reliability of the trials and constitute a problem that should be overcome by establishing GEM and practices. However, even with a standard procedure available, within a marine context (but also generally in robotics), replicability cannot be guaranteed. Therefore, the need is still present for a smart tool in charge of suitably and actively conducting experiments to be developed.

Experimental Design

A standard guideline for the design and execution of path-following tasks and some assumptions are defined in the following.

- An experiment (or batch) is formed by n runs, corresponding to n possibly different paths that are executed sequentially in time. Ideally, runs are executed in such a way that they are independent, but this needs to be verified through extensive simulation (see the "Early Simulation Results With DeepRuler" section).
- Each run is divided into four sequential phases: 1) approach, 2) forward path, 3) turn, and 4) backward path.
- A rectangular area $R_1 \times R_2$ for executing the experiments is defined by the human operator. This area is divided into three subareas: a middle area $L \times R_2$ for executing the path back and forth (at least one repetition) and two lateral areas for the execution of the approach and turn phases. A suitable maneuver for the approach and turn phases is defined by: 1) a straight line connecting the end of the currently followed path with 2) a semicircumference of radius r, providing the reference for the turn phase, and 3) a straight line of length l, connecting the semicircumference to the target path reference to be followed. The latter is responsible for achieving repeatability. Its direction is chosen to head the vehicle on the tangent to the target path in its starting point (see Figure 1).
- Performance is measured only while executing the actual path (during neither approach nor turn phases).

The human operator is required to set values for R_1, R_2, L, r, and l, paying attention to maneuverability restrictions and ensuring that l is long enough for the vehicle to enter the target path at its tangent and r is big enough for the vehicle to be able to follow the semicircumference (i.e., r is equal to or greater than the minimum turning radius of the vehicle).

For the moment, the n target paths in a batch are chosen for a small dimensional parametric class of functions, such as sinusoids, circles, and ellipses. Generalization to other classes of target functions is being considered.

Proposed Methodology and System Development

For the abovementioned reasons, an appropriate software framework named DeepRuler (design execute and evaluate path-following for robotic unmanned vehicles and experiment replicability) has been designed and developed. This framework allows one to easily design experiments and to automatically execute them, both running a simulator or driving a real robot. Details about DeepRuler are provided in the "The Automated Approach" section and the "Software Design and Features" section, while the "Early Simulation Results With DeepRuler" section reports preliminary results obtained employing a simulated robot.

The Automated Approach

The best way to achieve a good level of reproducibility and repeatability of experiments is to completely automate the process of generation and execution of the experiment.

To this aim, the DeepRuler framework has been designed with a threefold objective. The first (design) is to guide the user through the process of the definition of the path-following experiment. The second (execute) is to conduct and supervise the execution of the experiment itself. The third (evaluate) is to automatically measure, directly on the field, the robot performance through predefined or custom metrics. DeepRuler is then responsible for the following.

- *Helping the User in the Definition of the Experiment*: In an offline phase, the experiment definition is carried out thanks to a step-by-step windowed configurator that guides the user through the setup process. The user is asked to input the parameters that define the experiment, such as the working area, the shape of the paths, the parameters of each path, the structure of the telemetry, the metrics to be computed, and so on. The configurator generates a file describing the modality of the experiment that can be reused each time that an experiment has to be reproduced exactly in the same way. In addition, during this phase the user can customize the framework in a manner that will be described in the "Software Design and Features" section.
- *Controlling the Robot; Sequencing the Various Experiment Phases*: DeepRuler does not directly drive the robot through motion commands, rather it acts as a finite-state machine (FSM) providing high-level commands to the robot controller. In the current release, DeepRuler employs only two commands: 1) a GOTO command, used only once at the beginning of the experiment to let the robot reach the initial position and orientation and 2) a NEWPATH command that provides the points of a path, used each time that the robot has to perform a new run phase (approach, forward, turn, and backward).
- *Collecting the Telemetry Coming From the Robot*: During the execution of the experiment, the robot continuously sends its telemetry to the framework which in turn collects it, according to a predefined policy, for subsequent processing. The variables to be collected are defined during the configuration of the experiment and they mandatorily include information about the position of the robot and all the variables needed to compute the metrics.
- *Computing the Metrics at the End of Each Run*: The collected data will be used at the end of each run to compute the metrics selected during the configuration.

The experiment execution can be monitored by the user thanks to a human computer interface (Figure 1) which reports all of the relevant data and shows a real-time plot of the position (actual and historical) of the robot.

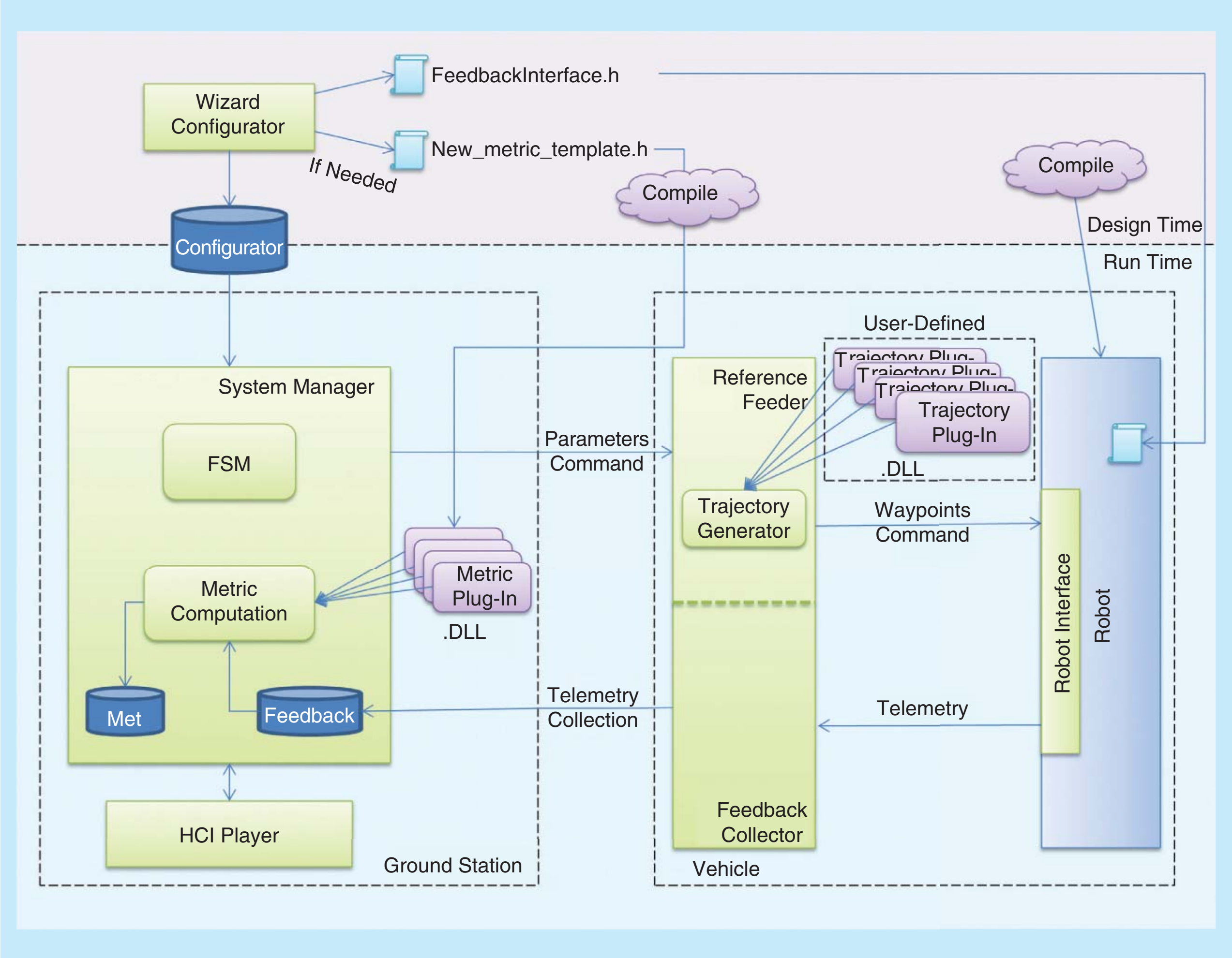


Figure 2. The DeepRuler framework architecture.

A key aspect of the proposed framework is its modularity in terms of path generation, telemetry collection, and metrics computation. These topics can become quite broad and cannot be hard-coded or even foreseen for nontrivial data. What this software actually offers is a way for the user to customize the framework with a limited effort thanks to a plug-in approach. The user can provide, in the form of dynamic library plug-in, his path generation function, and/or his own metrics algorithm to be integrated in the framework without the need to recompile the software. This way, the framework can be expanded to be able to perform different types of user-defined shapes (i.e., square wave, triangular wave, sawthoot wave, and so on).

As a demonstrative case, the current version of DeepRuler provides a sinusoidal path generation function in which each run is parameterized by the amplitude of the sine wave and by the number of half periods to be performed. As an example, an experiment configured with ten runs, where the active run (the seventh) is parameterized by an amplitude of 5 m and has three half periods, is shown in Figure 1.

Software Design and Features

The DeepRuler software framework is an open-source tool. DeepRuler has been developed here as a highly expandable tool for automatic path following and, more broadly, for robot performance testing. Moreover, the aim is to foster its usability throughout the robotic research community. DeepRuler is designed to minimize the effort required by the developer of the robot to integrate the functionality of the framework in his/her software. It comes with a library with a simple interface to be linked with the robot software.

The current architecture, shown in Figure 2, is made up of five modules.

- The wizard configurator guides the user through the experiment setup process described above. In addition, it autogenerates an interface file that will be used in the robot software to include the DeepRuler framework. This file essentially describes the structure of the telemetry to be collected. The wizard configurator, if necessary, will also guide the user through the definition of new metrics. The output of such steps will be a C++ template of a dynamic library plug-in in which the user is asked to implement only the core computation function, without the concern of how it will be integrated in the framework. Once compiled, the framework will use this library during the experiment execution to compute the custom metrics. A similar approach is used to define custom path shapes. A C++ template is provided and the user only has to implement one function, compile the library, and put the plug-in in a predefined folder.

- The system manager (SM) is the core of the framework and implements the FSM coordinating the activity of the whole experiment. It is responsible for sequencing the runs of the experiment and the phases of each run. It is also responsible for collecting and logging the telemetry data and for instructing the human–computer interface (HCI) behavior. In addition, it computes all the configured metrics loading the appropriate plug-ins.
- The reference feeder (RF) and feedback collector (FC) are responsible for sending the reference path points to the robot and for collecting, according to predefined policies, the telemetry coming from the robot. The RF receives from the SM the parameters of the paths of the entire experiment and computes the waypoints of each phase of each path, exploiting the path generation function. When commanded by the SM, the RF feeds the robot with the points of one path. In this way, the communication burden between the SM and the RF is kept low (only parameters are passed and not all the experiment points), creating an advantage in the event that the link between the two has a limited bandwidth. The FC collects chunks of telemetry before forwarding them to the SM using predefined policies. This is needed because in some operative scenarios the robot cannot send each sample of the telemetry immediately (the robot might be too far from the communication link or underwater). The FC acts as a buffer for the telemetry waiting to be forwarded, in a compressed way, to the SM.
- The robot interface is composed by a static library that the user has to link with the robot software to use DeepRuler. It exports a simple interface with few functions [Open(), Close(), and Path_Complete()] and offers two callbacks to signal to the robot for the arrival of a Goto() command and the arrival of the points of a new path to be followed [New_Path()].
- The HCI player displays the overall status of the experiment, the real-time status of the current run being carried out (Figure 1) and, at the end of the experiment, plots the graph of the computed metrics.

The DeepRuler framework, as it has been presented so far, can be easily interfaced to most of the existing robots. However, the robotic community is moving toward the adoption of the robotic operating system (ROS) middleware [14] due to the numerous advantages it provides. For this reason, the DeepRuler framework has been designed to interface and operate with ROS as well, thanks to a gateway that must be introduced between the ROS robot and the RF and FC component. Such gateway is an ROS node that links to the robot interface static library and translates the DeepRuler protocol into appropriate nonstandard ROS topics.

Another key point of the software design is portability across operating systems (OSs) and hardware (HW) architectures. With this objective in mind all the code has been written in C++ and only well-known portable libraries, such as the standard template library [15], Qt [16], and boost [17] have been used. The choice of C++ as the programming language is almost the same as it has been in the past years. It is one of the most used programming languages across the world [18] and it offers an excellent tradeoff between ease of development and bytecode efficiency that no other language currently offers. Portability across OSs and HW architectures means that, at present, all the code has been compiled and tested under Linux amd64/i386 and Windows 64/32-b architectures. Minor development efforts should be necessary for porting and testing it under MacOSX, and only a partial rewrite of graphical interfaces will be needed to adapt them to Android and iOS (both x86 and advanced reduced instruction set computer machine) environments. In theory, the software should be portable with a reasonable effort on less popular but firmly established architectures such as VxWorks and QNX, but such an attempt has not yet been made.

To provide communication to/from the robot, two well-known and established data serialization technologies have been selected. The user simply has to choose the most suitable mechanism according to his/her needs. The two different communication plug-ins are already available and ready to be used:

1) a standard ROS communication mechanism (refer to [14])
2) boost::serialization [17], a versatile but still bandwidth-efficient serialization mechanism useful for non-ROS robots.

The addition of more plug-ins for serialization over acoustic links (to extend the use of DeepRuler to underwater robots) and long-range wireless links is planned.

This means that data structures for commands, reference, and feedback will be passed/received to/from the robot using well known and robust serialization mechanisms. This approach has been preferred to C-style serialization of structs because it is more portable and less error-prone.

Another important feature of the DeepRuler framework is the overall flexibility of the system. Ideally, the user should be able to run each software component on the combination of OS/HW platforms that best suits his/her needs. For this reason, with the exception of the two graphical interfaces, the software has been structured as standalone console application communicating through standard User Datagram Protocol sockets. This approach, together with the excellent portability of the software, guarantees the maximum flexibility for the deployment of all the system components on almost any possible robot/control system. In particular, the first multiplatform use-case scenario that has been foreseen (and that is being implemented and tested) consists of three distinct devices on top of three architectures:

1) a Linux-based robot, running both the FC and RF and the robot interface
2) a Windows laptop used to run the wizard configurator, the SM and the HCI player
3) an Android/iOS tablet showing another HCI player for experiment monitoring on the field.

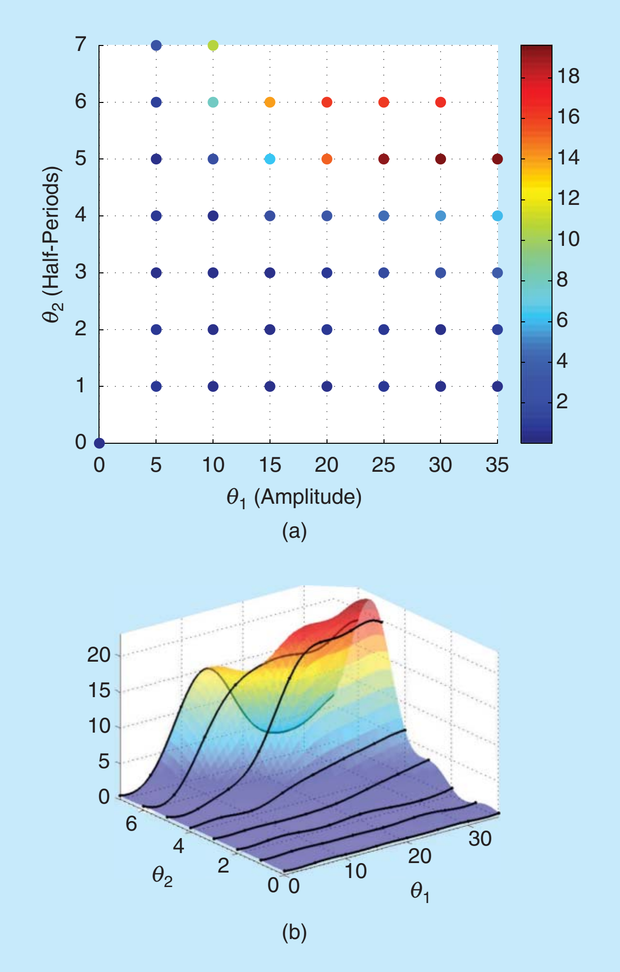
However, it has to be noted that the framework has been designed to minimize the data exchange burden between the RF and FC and SM, therefore, it is more convenient to employ the less powerful (in terms of bandwidth) communication link (acoustic, long range radio, and so on) between them.

Early Simulation Results with DeepRuler

A preliminary simulation study has been conducted employing DeepRuler and the consolidated Charlie USV HW-in-the-loop simulator. The simulator is part of the custom Charlie USV architecture; it is developed in C++ for standard Linux OS distributions, and it allows a very precise simulation of the vehicle dynamic and kinematic motion evolution, as well as a fine statistical-based simulation of the sensors in terms of precision, disturbance, and data rates, with respect to the real system characteristics.

The class of sinusoids $\gamma(\theta, x) = \theta_1 \sin((\pi\theta_2/L)x)$ were selected, where θ_1 is the amplitude, θ_2 is the number of half-periods in L and $x \in [0, L]$. To achieve replicability and provide standards to characterize the performance of a vehicle following a sinusoidal path, the class of paths was restricted to those with an integer number of half-periods. To perform the line-following as a first test run, the associated design space is then $\Theta = \mathbb{R}_{>0} \times \mathbb{Z}_{\geq 1} \cup \{(0, 0)\}$ and $\theta = [\theta_1, \theta_2]^T \in \Theta$.

The measurement process is described as follows. For each path $\theta \in \Theta$ an observed path $z(\theta)$ is given as a time series $z_t(\theta) = \begin{bmatrix} x_t(\theta) \\ y_t(\theta) \end{bmatrix} = \begin{bmatrix} \bar{x}_t(\theta) \\ \bar{y}_t(\theta) \end{bmatrix} + \varepsilon_t \in \mathbb{R}^2, \varepsilon_t \sim$ Uniform where $[\bar{x}_t(\theta), \bar{y}_t(\theta)]^T$ are GPS measurements and $t = 1, \ldots, T$, where T depends on a number of factors, including the path, the vehicle, and the environmental conditions. Although not considered here, T can be part of a compound measure of vehicle performance for the $\gamma(\theta)$ path. For each run and each path back and forth, performance measures $\eta_1(\theta), \eta_2(\theta), \ldots$ are associated to $\{z_t(\theta)\}_t$.



(a)

(b)

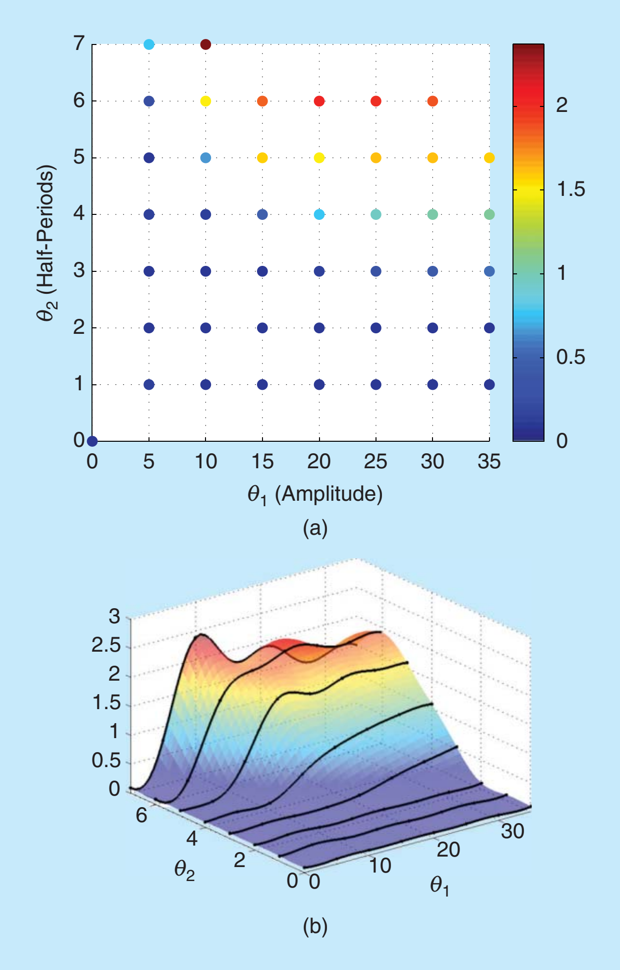
Figure 3. The Hausdorff distance. (a) The heatmap. (b) The Kriging reconstruction.

The parameters characterizing the experiments are set to $L = 100$ m, $l = 25$ m, $r = 14$ m, and the position noise is 0.2-m maximum. Furthermore, uniform noise is added to the vehicle heading (0.2°) and speed (0.1 m/s). The simulation study was conducted in three batches.

- In the first batch of experiments θ was set in $\{(0, 0)\} \cup \{5, 10, 15, 20, 25\} \times \{1, 2, 3, 4, 5, 6\}$; the path sequence execution starts from the half-period curve, increasing the amplitude at each run, then the period is increased by half and the amplitude goes from lower value to the higher ones allowed and so forth.
- The second batch of experiments includes the same runs as the first batch but executed to reduce the difference between consecutive couples of parameters, thus giving priority to smaller values of the parameters.
- The third batch of experiments further investigates the parameter space/potential target paths and includes the runs in $\{(0, 0), (10, 2), (15, 5), (5, 7), (10, 7)\} \cup \{(30, i), (35, i), (j, 6), i = 1 \ldots 5, j = 5 \ldots 30\}$.

Since the vehicle performance indices are computed when following the path forward and backward, a natural question concerns possible differences between the values of each index for each run for the forward and the backward paths (see point 1 in the "Experimental Design" section). Let D_H^F, D_A^F, D_H^B, and D_A^B be the Hausdorff and area indices computed separately for the forward and backward paths. A paired samples t-test (a statistical hypothesis test in which the test statistics follows a Student's t distribution) has been performed for both the indices and computed p-values (where the p-value is the probability of obtaining the observed sample results) are as 0.80 for D_H^F, D_H^B and 0.45 for D_A^F, D_A^B in the first batch, and 0.24 for D_H^F, D_H^B and 0.26 for D_A^F, D_A^B in the second batch.

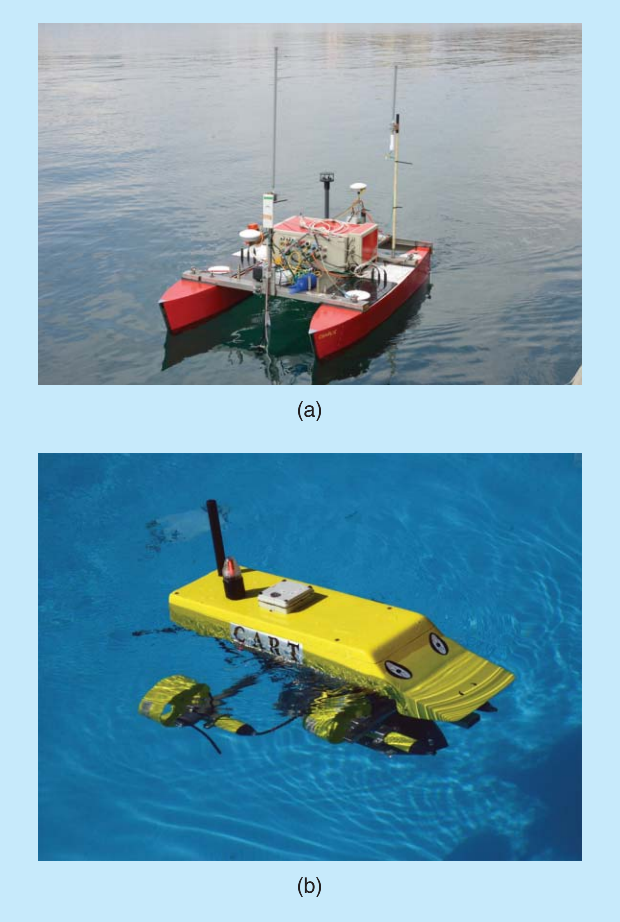
All the p-values are >0.05, which implies that there is no evidence in the data for rejecting the hypothesis that D^F and D^B have the same mean. As a consequence, an overall mean value can be computed for each run. This point deserves special attention when performing experiments at sea; in fact, a significant difference in the results may highlight the presence of uncontrollable variables or external disturbances. With the aim to test independence from the order of the runs, a t-test has also been performed for comparing the performance indices computed in the first and second batches. The resulting p-values for the Hausdorff distance and for the area index are 0.46 and 0.79, respectively. These values validate the first assumption in the "Experimental Design" section (the fact that the order of the runs does not affect the metrics values). Figures 3(a) and 4(a) show the Hausdorff and area performance functions at the observed parameter values. As expected, in both cases higher values of the performance indices are



(a)

(b)

Figure 4. An area index. (a) The heatmap. (b) The Kriging reconstruction.



(a)

(b)

Figure 5. The CNR-ISSIA vehicles employed in the test campaign. (a) The Charlie USV is 2.40 m × 1.70 m, and it weighs about 250 kg. (b) The Shark is 0.90 m × 0.75 m, it is 0.60-m high, and it weighs about 40 kg. (Photos courtesy of CNR-ISSIA.)

located in areas of the design space where there are high values of the parameters. In fact, for sinusoidal paths, high values represent the limit capability of a vehicle performing the path-following task. However, it is clear that different performances have been measured, simply comparing the worst performance values achieved for the area index in (10, 7) with {(25, 5), (30, 5), (35, 5)}, design points that, on the contrary, lead to higher values for the Hausdorff distance. This highlights the complementary information gathered from different performance indices and the importance of defining and exploiting more performance indices and a suitable compound index.

The problem of predicting output at unverified parameter sets is addressed by a classical Kriging reconstruction of a suitable portion of the parameter space. Such a reconstruction is computed on a 50×50 grid starting from the performance indices computed on 44 design points: the performance value corresponding to the line $\theta = (0, 0)$ is also assigned to the sets of design points $\{(0, i), i = 1, \ldots, 7\}$ and $\{(i, 0), i = 5, \ldots, 35\}$. A classical algorithm with a Gaussian correlation model and maximum likelihood estimation for its parameters has been adopted (see [19]) and the best linear unbiased predictor for the regression model, the one that minimizes the variance of the prediction error, is chosen. Figures 3(b) and 4(b) report the reconstructions for both Hausdorff and area indices and the black lines show that θ_2 was set as a discrete parameter in the beginning. It is worth noticing that due to a restricted number of experiments that can be executed in simulations and, especially, at sea, prediction is the main tool for assessing the performance of a vehicle performing experiments when a class of paths is chosen. Significant changes in the predicted values are not expected when varying the regression model, but further investigations on the sensitivity to the model are planned to validate this assumption.

Experiments with DeepRuler

Real experiments employing DeepRuler were performed in the Canale di calma di Pra, Genova, Italy, using two different National Research Council of Italy–Institute of Intelligent Systems for Automation (CNR-ISSIA) vessels: 1) the Charlie USV and 2) the Shark unmanned semisubmersible vehicle (USSV). Both robots are shown in Figure 5.

The two vehicles were required to perform path-following experiments on sinusoidal paths which differed in amplitude. For each vehicle, the experiment parameters, i.e., the working area, the robot feedback structure (containing the variables needed to compute the selected metrics), and the features of

Table 1. The mean values computed on forward and backward phases of D_H and D_A for the Shark USSV trials.

θ_1	0	2.5	2.5	2.5	5	5	5	7.5	7.5	7.5
θ_2	0	1	2	3	1	2	3	1	2	3
D_A	0.22	0.28	0.28	0.28	0.20	0.21	0.22	0.18	0.26	0.23
D_H	0.51	0.50	0.68	0.68	0.53	0.60	0.84	0.47	1.17	1.07

each run, were defined with the DeepRuler wizard configurator to design suitable and relevant experiments. In particular, one of the experiments was executed by Shark, shown in Figure 5(b), within a working area of 20 m × 20 m and was composed by ten runs: an initial straight line plus nine sinusoidal paths obtained combining three different amplitude values (2.5, 5, and 7.5 m) with three different numbers of half-periods (1, 2, and 3). The execution time of the whole experiment was about 32 min. The DeepRuler HCI player while Shark was performing run 7 of the experiment (a sinusoidal path with three half-periods with an amplitude of 5 m) is shown in Figure 1.

Table 1 reports the computed indices of the entire experiment. Note the maneuvering path phases (approach and turn) placed at the ends of the sinusoidal paths which guarantee that the robot will approach the backward and forward phases with the same initial conditions (heading tangent to the beginning of each sinusoidal phase).

A second batch of field experiments was conducted employing Charlie [depicted in Figure 5(a)]. The runs requested to Charlie were similar to the previous ones, except for the set of parameters. Since Charlie is much bigger than Shark, it needs a larger working area which has been set to 80 m × 100 m. The executed experiment was composed by ten runs: an initial straight line plus nine sinusoidal paths obtained combining three different amplitude values (5, 10, and 15 m) with three different numbers of half-periods (1, 2, and 3). The execution time of the whole experiment was about 137 min.

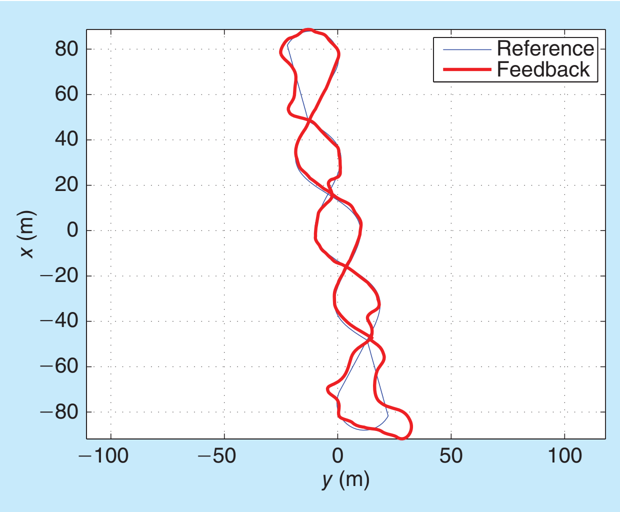


Figure 6. The Charlie USV execution of run 7, a sinusoidal path with three half-periods and an amplitude of 10 m.

Table 2. The mean values computed on forward and backward phases of D_H and D_A for the Charlie USV trials.

θ_1	0	5	5	5	10	10	10	15	15	15
θ_2	0	1	2	3	1	2	3	1	2	3
D_A	2.21	1.83	1.56	1.73	2.02	1.47	0.82	1.80	1.02	0.83
D_H	7.18	5.16	7.20	5.35	5.11	4.68	3.49	4.18	4.01	3.51

For this experiment, the control gains of Charlie have been altered to emphasize how a badly tuned control system negatively impacts the path-following performance and how such performance decay is evident from the analysis of the computed metrics. In Figure 6, run number 7 of the experiment is shown, while in Table 2 the computed metrics are reported.

Conclusion and Future Work

This article has highlighted the problem of the experiment replicability in real operating conditions. Experimental trials conducted prior to the development of DeepRuler have clearly shown that quantitatively measurable replicable experiments can only be achieved exploiting the formal methodologies for the design of experiments and the automatic tools for the experiments execution. In this article, classes of target paths were chosen, being described by a small dimension parameter space. Two performance measures (Hausdorff and area) were adopted for the definition of vehicle performance, as well as a method for reconstruction of performance functions over a suitable portion of the parameter space. Another contribution of this article is the development of the DeepRuler software tool which allows a completely automatic management of the experiment design and execution, as well as the data analysis through the predefined performance evaluation functions. A preliminary integration of DeepRuler with the consolidated the Charlie USV simulator (which provides a behavior similar to the real robot) allowed the gathering of several data batches. Extensive field trials demonstrated the feasibility of the approach as well as the reliability of the developed tool and methodology. In addition, it has to be noted that DeepRuler has been first tested in a marine surface context, but it can be extended to any other kind of robotic application (aerial, underwater, and terrestrial).

To this aim, the more important aspects are the extension in three dimensions and the communication link. The extension of path-following algorithms to work on a three-dimensional (3-D) path is straightforward; the more interesting issue in this case concerns the definition of indices which can suitably evaluate the vehicle performance dealing with 3-D paths. New metrics will be implemented and added to DeepRuler in the near future; furthermore, the users can implement their own indices and make them available to the community to be able to compare different vehicles and guidance systems.

On the other hand, the available communication link deserves a separate discussion, above all addressing the underwater context. Underwater robotics can only rely on acoustic communication links, characterized by low bandwidth, data losses, and delays. DeepRuler was designed keeping these limitations in mind; in fact, the RF and FC modules that are usually placed on the robot have different modalities for telemetry sending. If a reliable communication link is available it can directly send feedback while running the experiments, thus eliminating big storage resources on the robot itself, and if only a poor communication link is available during the experiment runs, the RF and FC running onboard the robot can collect all the telemetry and send it only at the end of either each run or the overall experiment, provided that the vehicle can emerge after them and communicate through a suitable link.

To generalize the performance evaluation and comparison of any (or at least a large class of) vehicles, some issues still need to be addressed. They include the definition of a compound index of performance applicable to any robot and the identification of one or more criteria to decide whether an experiment has failed (leading to infeasible parameter regions) or whether the vehicle simply performs poorly on a target path (in the sense of exceeding a predefined threshold on a specific index). These would be the essential keys to develop a sequential methodology to design experiments that would allow automatic determination of an infeasible parameter region, investigation of vehicle performance over a parameter space, and criteria to compare performance of two vehicles with different feasible parameter regions.

References

[1] F. Bonsignorio, J. Hallam, and A. P. del Pobil, Eds., GEM Guidelines, Euron GEM Sig Report, 2008.

[2] F. Bonsignorio, J. Hallam, and A. del Pobil, "Defining the requisites of a replicable robotics experiment," in *Proc. RSS Workshop Good Experimental Methodologies Robotics*, 2009.

[3] F. Amigoni, M. Reggiani, and V. Schiaffonati, "An insightful comparison between experiments in mobile robotics and in science," *Auton. Robot.*, vol. 27, no. 4, pp. 313–325, 2009.

[4] S. Tadokoro and A. Jacoff, "Performance metrics for response robots [industrial activities]," *IEEE Robot. Automat. Mag.*, vol. 18, no. 3, pp. 12–14, 2011.

[5] M. Bibuli, G. Bruzzone, M. Caccia, and L. Lapierre, "Path-following algorithms and experiments for an unmanned surface vehicle," *J. Field Robot.*, vol. 26, no. 8, pp. 669–688, 2009.

[6] A. P. del Pôbil. (2006, Oct.). Why do we need benchmarks in robotics research?. presented at Lecture Notes for IROS 2006 Workshop Benchmarks Robotics Research. Beijing, China. [Online]. Available: http://www.robot.uji.es/EURON/pdfs/Lecture Notes IROS06.pdf

[7] F. Amigoni, S. Gasparini, and M. Gini, "Good experimental methodologies for robotic mapping: A proposal," in *Proc. IEEE Int. Conf. Robotics Automation*, 2007, pp. 4176–4181.

[8] A. Lampe and R. Chatila, "Performance measure for the evaluation of mobile robot autonomy," in *Proc. IEEE Int. Conf. Robotics Automation*, May 2006, pp. 4057–4062.

[9] M. Caccia, M. Bibuli, G. Bruzzone, and L. Lapierre, "Vehicle-following for unmanned surface vehicles," in *Further Advances in Unmanned Marine Vehicles* (IET Control Engineer Series vol. 77). London, U.K.: Inst. Eng. Technol., 2012, pp. 201–230.

[10] M. Caccia, G. Bruzzone, and R. Bono, "A practical approach to modeling and identification of small autonomous surface craft," *IEEE J. Ocean. Eng.*, vol. 33, no. 2, pp. 133–145, Apr. 2008.

[11] N. Miskovic, Z. Vukic, M. Bibuli, G. Bruzzone, and M. Caccia, "Fast in-field identification of unmanned marine vehicles," *J. Field Robot.*, vol. 8, no. 1, pp. 101–120, Jan./Feb. 2011.

[12] E. Saggini, E. Zereik, M. Bibuli, G. Bruzzone, M. Caccia, and E. Riccomagno, "Performance indices for evaluation and comparison of unmanned marine vehicles' guidance systems," in *Proc. 19th IFAC World Congr.*, Cape Town, South Africa, 2014, pp. 12182–12187.

[13] E. Zereik, M. Bibuli, G. Bruzzone, and M. Caccia, "Jacobian task priority-based approach for path following of unmanned surface vehicles," in *Proc. 9th IFAC Conf. Control Applications in Marine Systems*, 2013, pp. 114–119.

[14] M. Quigley, K. Conley, B. P. Gerkey, J. Faust, T. Foote, J. Leibs, R. Wheeler, and A. Y. Ng, "ROS: An open-source robot operating system," in *Proc. ICRA Workshop Open Source Software*, 2009.

[15] D. R. Musser, G. J. Derge, and A. Saini, *STL Tutorial and Reference Guide, Second Edition: C++ Programming with the Standard Template Library*. Boston, MA: Addison-Wesley Longman Publishing Co., 2001.

[16] J. Blanchette and M. Summerfield, *C++ GUI Programming with Qt 4*. Upper Saddle River, NJ: Prentice Hall PTR, 2006.

[17] *Beyond The C++ Standard Library: An Introduction To Boost*. Upper Saddle River, NJ: Pearson Education, 2006.

[18] TIOBE Software BV. (2014, Nov.). Tiobe programming community index. [Online]. Available: http://www.tiobe.com/tpci.htm

[19] S. N. Lophaven, H. B. Nielsen, and J. Søndergaard. "A Matlab Kriging toolbox," Tech. Univ. Denmark, Kongens Lyngby, Tech. Rep. IMM-TR-2002-12, 2002.

Andrea Sorbara, National Research Council of Italy–Institute of Intelligent Systems for Automation (CNR-ISSIA), Genova, Italy. E-mail: andrea.sorbara@ge.issia.cnr.it.

Andrea Ranieri, National Research Council of Italy–Institute of Intelligent Systems for Automation (CNR-ISSIA), Genova, Italy. E-mail: ranieri@ge.issia.cnr.it.

Eleonora Saggini, National Research Council of Italy–Institute of Intelligent Systems for Automation (CNR-ISSIA) and University of Genoa–Department of Mathematics (UNIGE-DIMA), Genova, Italy. E-mail: eleonora.saggini@ge.issia.cnr.it.

Enrica Zereik, National Research Council of Italy–Institute of Intelligent Systems for Automation (CNR-ISSIA), Genova, Italy. E-mail: enrica.zereik@ge.issia.cnr.it.

Marco Bibuli, National Research Council of Italy–Institute of Intelligent Systems for Automation (CNR-ISSIA), Genova, Italy. E-mail: marco.bibuli@ge.issia.cnr.it.

Gabriele Bruzzone, National Research Council of Italy–Institute of Intelligent Systems for Automation (CNR-ISSIA), Genova, Italy. E-mail: gabriele.bruzzone@ge.issia.cnr.it.

Eva Riccomagno, University of Genoa–Department of Mathematics (UNIGE-DIMA), Genova, Italy. E-mail: riccomagno@dima.unige.it.

Massimo Caccia, National Research Council of Italy–Institute of Intelligent Systems for Automation (CNR-ISSIA), Genova, Italy. E-mail: massimo.caccia@ge.issia.cnr.it.

Tracking Divers

An Autonomous Marine Surface Vehicle to Increase Diver Safety

By Nikola Mišković, Đula Nađ, and Ivor Rendulić

Diving is a high-risk activity due to the hazardous environment, dependence on technical equipment for life support, complexity of underwater navigation, and limited monitoring from the surface. This article describes a new concept of using an autonomous surface vehicle (ASV) as a private satellite that tracks divers, thus significantly increasing diving safety. Since the vehicle is above the diver at all times, acoustic communication with the diver interface in the form of an underwater tablet is more efficient and robust, which enhances diver navigation and enables reliable monitoring from the surface. This article focuses on a diver-tracking control structure that uses a diver motion estimator to determine diver position, even in cases when acoustic position measurements are not available.

Conducting experiments with divers presents a challenge due to uncertainties, such as those introduced by the environment, unmodeled dynamics, acoustic sensors, and divers themselves (e.g., the emission of air bubbles). A step-by-step experimental plan, which includes a virtual diver (VD), an underwater remotely operated vehicle (ROV), and a human diver, allows the identification of different uncertainties. The results show that the mean tracking error with a VD (influenced only by the environment and unmodeled dynamics) is around 0.5 m; with an ROV (including the influence of acoustic sensor), it is around 1 m; and with a human diver, it is around 1.8 m. These data are validated against ground-truth video imagery.

©ISTOCKPHOTO.COM/TUMPIKUJA

Digital Object Identifier 10.1109/MRA.2015.2448851
Date of publication: 11 September 2015

Mitigating Hazards to Divers

Scuba diving activities, whether they are recreational, scientific, or technical, are classified as high risk due to

- the unpredictable, dangerous, and unfamiliar environment constantly under influence of external disturbances
- the dependence on the technical equipment that ensures life support
- the health consequences that diving can have on a diver.

Diver safety is one of the major concerns of the diving industry. The Divers Alert Network, one of the largest training agencies in the world, identified the most significant triggering events that lead to fatalities as air loss, entrapment or entanglement, gear issues, rough water, and buoyancy issues. In 20% of scuba diving fatalities, the initial triggering event could not be determined.

One significant cause of accidents during diving is the loss of consciousness in the water. The final outcome is very often the same—drowning. Nitrogen narcosis is a state when a diver seems perfectly fine at first but is intoxicated, if not unconscious, and cannot make sane decisions. An inexperienced dive buddy may not even notice nitrogen narcosis. It has been reported by experienced divers that, while under this state, their buddies seemed perfectly fine until they started some unexpected behavior, such as separating from the buddy, not following previously agreed mission plans, changing depth rapidly, or even taking their masks off. Once this state is recognized (which is a challenge for an inexperienced dive buddy), the simple act of attracting the attention of the diver can snap him or her out of this state, which, if left unattended, may have catastrophic consequences. Even though diver safety risks are commonly minimized by diving in groups, or at least in a pair with a buddy, statistics show that 40% of the fatalities take place during a period of buddy separation and 14% involve declared solo dives, meaning that more than 50% of accidents happened while the divers were not accompanied.

Diver safety is seriously jeopardized during diving activity not only because of unpredictable underwater scenarios but also because of diver invisibility to surface vessels. Currently, diving areas are marked using passive buoys with international dive flags that serve as indicators for man-operated surface vessels to avoid the area. Unfortunately, these markings are often disregarded by surface vessels. The diver's area of operation can be increased if the diving buoy is linked to the diver via a cable [1]—but this solution is unacceptable for deep and/or long dives due to possible entanglement, drag, and cumbersomeness.

Even though diver safety is the most significant issue, diving activities are also significantly hampered by the lack of navigation capabilities and communication with the surface. Underwater navigation poses a challenge even for experienced divers. Gravity compromised by buoyancy, limited visibility, and lack of global navigation satellite system reception jeopardize divers' activities as well as safety underwater. Classical techniques for underwater navigation, such as referencing according to the sun, a compass, or underwater features, are imprecise, tedious, and require concentration and experience.

Current technological solutions enable determining the position of the diver relative to the surface station by using acoustic-based technology. These systems, which rely on static transmitters/receivers, exhibit serious performance deterioration due to acoustic multipath effects when the diver is distanced from the ship [2]. For the same reason, communication between the diver and the surface is an important issue and can compromise diver monitoring from the surface if interrupted. Reliable communication is important to diving supervisors, who monitor the progress of diving operations, as well as the divers themselves, who appreciate monitoring as a way of increasing their safety during dives. Current communication systems, as in the case of navigation, are not appropriate at larger distances (due to multipath) or in cases when obstacles are present between the diver and the base station.

> **A vertical acoustic communication channel of minimal distance is formed, ensuring reliable communication.**

Concept

This article proposes a new concept for dealing with the aforementioned major diving challenges by using an omnidirectional ASV with the ability to follow the diver and act as a private satellite, thus significantly increasing diver safety, alleviating underwater navigation difficulties, and enabling monitoring from the surface. Since the ASV is tracking the diver at all times, i.e., keeping its position above the diver, as shown in Figure 1, the following set of functionalities is accomplished.

- Since the ASV carrying the international dive flag is always above the diver, it expands the diver's safe underwater operation area. There is no need for the conventional marking of the diver area by using static buoys, and physical tethering with the buoy is avoided since the ASV uses acoustic localization of the diver for tracking.
- A vertical acoustic communication channel of minimal distance is formed, ensuring reliable communication and avoiding multipath problems. This also allows reliable transmission of global positioning system (GPS) coordinates to the diver, thus providing the diver with absolute GPS coordinates on his or her tablet.
- The diving supervisor at the surface has reliable data about the diver's position, and reliable communication between the diver and the diving supervisor is established through the vertical communication channel, thus significantly increasing reactivity in case of danger.

Related Work

The research area of diver–robot cooperation is very young, but, with the increased demand in autonomous marine robotics, the need for interaction with divers arises. Even though human divers today are increasingly being replaced with autonomous underwater vehicles in tedious tasks such as mapping and searching, there are still many applications that require a human presence underwater. These applications are mostly related to unconventional, nonrepetitive tasks such as underwater interventions.

Underwater ROVs have been commonly used in tandem with divers, mostly for the purpose of monitoring divers from the surface. However, the use of autonomous marine vehicles has taken place only recently. The list of autonomous underwater robots used for diver–robot applications is fairly short:

- AquaRobot [3], developed at McGill University, initially used as an amphibious robot for exploring underwater environments, was used to track divers based on visual detection of their motion [4].
- The BUDDY AUV, developed for the purposes of the European project Cognitive Autonomous Diving Buddy (CADDY, http://caddy-fp7.eu/), is the first AUV for interacting with divers by using an underwater touchscreen.

Tracking and navigating divers using ASVs was first addressed by the European project Cooperative Cognitive Control for Autonomous Underwater Vehicles [5]. A fleet of three ASVs was deployed, and successful diver-tracking and -guidance experiments were performed by using single range measurement from the vehicles [6]. This project was a stepping stone toward the CADDY project, under which the results presented in this article were obtained. It should be mentioned that other efforts have been made toward robot–diver interaction, such as [7]; however, these were done only in the simulation environment, without addressing the issues of diver detection or localization. While using acoustic positioning systems to localize divers is the most straightforward method, visual [4], [8] and sonar [9] detection can be found in the literature.

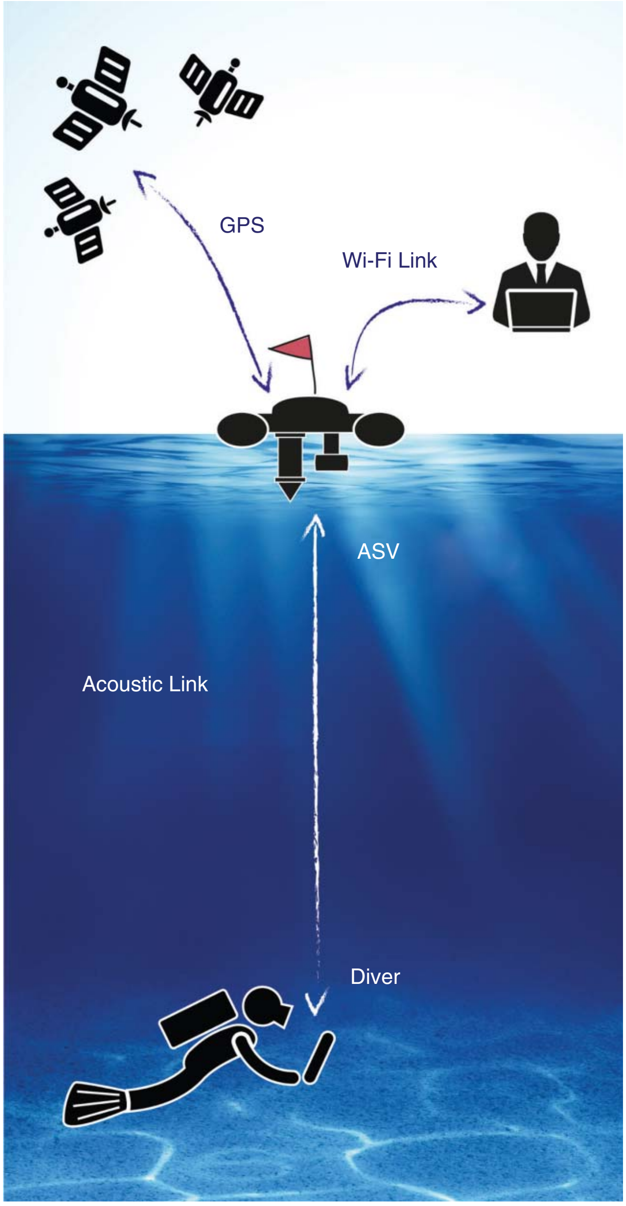


Figure 1. A novel concept for increasing diver safety: an autonomous surface platform, PlaDyPos, serving as the diver's private satellite and interfacing with the diver through an underwater tablet.

Contributions

The main contributions of this article are both in technical and control aspects as well as benchmarking and experiment design and execution. The first contribution is the development of the diver-tracking system consisting of an autonomous surface marine vehicle and an underwater diver interface used for two-way communication between the diver and the surface vehicle. The complete system is described in the next section. The second contribution is the onboard diver-tracking algorithm that uses intermittent acoustic diver position measurements fused with a diver motion estimator. The models, control, and tracking algorithms were initially described in a previous paper by the authors [10], but they are also included in this article for completeness. The third contribution is the design of a benchmark scenario with associated metrics for human–robot tracking performance measurement in the underwater environment. The benchmark scenario enables replicability of experiments in real conditions with human divers. Finally, the fourth contribution is the design and execution of an experimental plan that allows the identification of uncertainties introduced by the human diver, environment, unmodeled dynamics, and acoustic sensors by using tracking performance metrics.

Diver–Robot System Description

The overall diver-tracking system presented in this article consists of three main components: 1) the diver-tracking marine ASV, 2) the diver, and 3) an underwater tablet carried by the diver, serving as the diver's interface. The components are described in the following section, while the communication scheme between them is shown in Figure 2.

ASV–PlaDyPos

The ASV PlaDyPos (named after its initial purpose as a platform for dynamic positioning), Platform for Dynamic

Positioning (PlaDyPos) acts as an ASV and carries the international flag marking underwater activity. It has four thrusters in an "X"-shaped configuration, allowing omnidirectional motion, i.e., motion in the horizontal plane under any orientation. PlaDyPos, shown in Figure 3, has been developed at the Laboratory for Underwater Systems and Technologies at the University of Zagreb Faculty of Electrical Engineering and Computing, Croatia, and it is 0.35-m high, 0.707-m wide and long, and weighs approximately 25 kg. The control computer (isolated from environmental disturbances inside the platform hull) is in charge of performing control and guidance tasks (dynamic positioning, path following, and diver following) and all the data processing. Apart from the compass, batteries, and central processing units the PlaDyPos payload relevant to the diver-tracking experiments consists of

- a ublox Neo 6P GPS for determining position and, indirectly, the diver position in the horizontal plane
- a Tritech MicroNav ultrashort baseline (USBL) used to determine the position of the diver relative to the vehicle, with integrated acoustic modem
- a Bullet M2 wireless modem used for two-way communication with the ground station, thus making PlaDyPos a router from the diver to the surface station where the diving supervisor is stationed.

The USBL, shown in Figure 3(b), is used simultaneously for localization and two-way data transmission via an acoustic link (the second modem is mounted on the diver). While diver localization is the main topic of interest in this article, it should be mentioned that the acoustic link can be used to transmit messages as well as diver position based on USBL and GPS measurements from the vehicle.

Diver

The diver is mounted with an acoustic modem that is used for localization on board the surface vehicle as well as to communicate with the modem on the surface vehicle. As shown in Figure 2, the diver-mounted acoustic modem is connected to the RS232-Bluetooth converter, which allows transmission of the data via Bluetooth link. Our experiments have shown that Bluetooth communication has a range of about 15 cm underwater, which allows the diver-mounted Bluetooth module to establish a connection with another Bluetooth device in close proximity, such as the one on the tablet in the waterproof casing.

The overall diver-tracking system presented in this article consists of three main components.

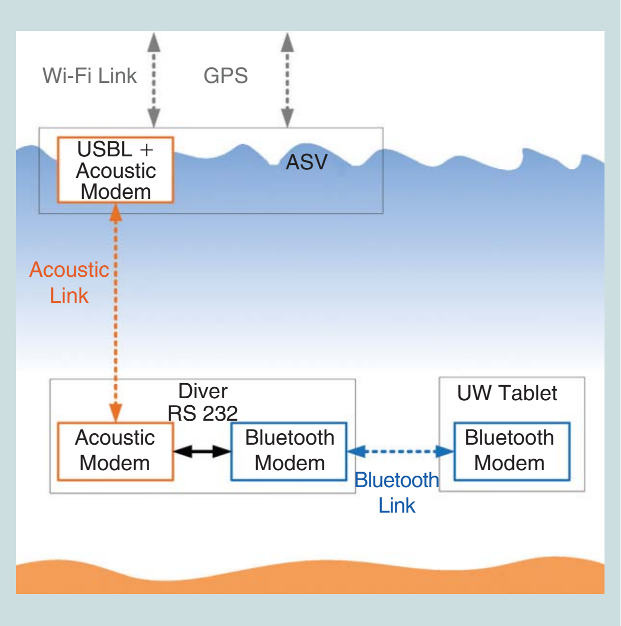
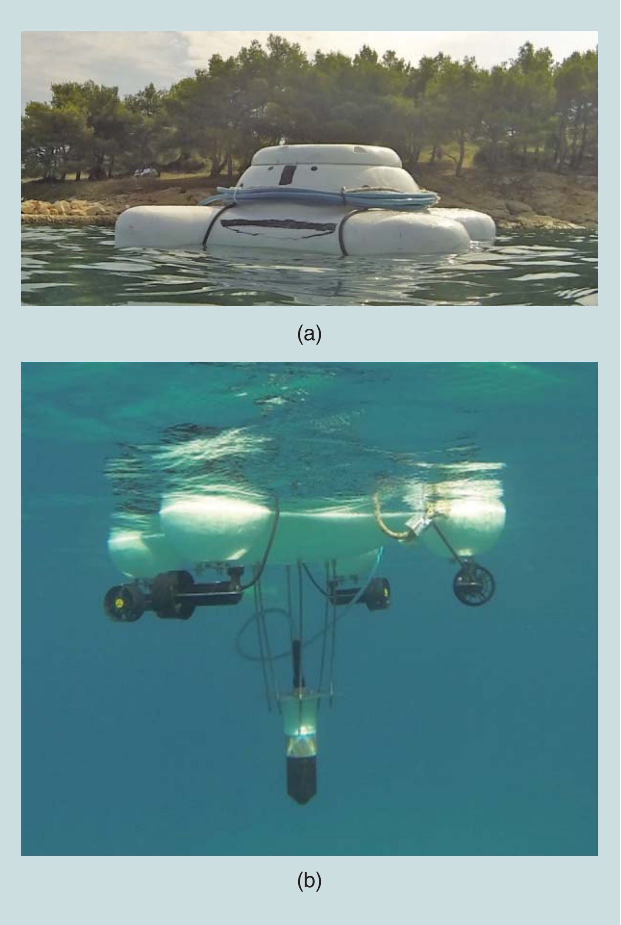


Figure 2. A schematic representation of the communication structure between the surface robot, the diver, and the underwater tablet. The ASV is linked to the diver-mounted modem via acoustic link, while the modem communicates via RS232 with the Bluetooth modem that connects to the underwater tablet through the Bluetooth connection that has proved to work underwater at short distances.


(a)

(b)

Figure 3. The diver-tracking ASV PlaDyPos viewed from (a) above and (b) below. The USBL mounted on the bottom of the hull is used to communicate and determine the position of the diver-mounted modem via acoustic link. This configuration makes PlaDyPos a router between the diver and the surface station where the diving supervisor is stationed.

Underwater Diver Interface

The diver carries an underwater interface (a commercially available tablet) sealed in a custom-made waterproof casing that has been tested in a pressure chamber for depths up to 50 m. Larger depths can be achieved but at the expense of a more robust and cumbersome design. The diver-mounted Bluetooth modem is placed on the waterproof casing, enabling a Bluetooth connection with the tablet without compromising the structural integrity of the casing itself. A tablet with an inductive touchscreen is integrated in the overall system, allowing the diver to send feedback to the surface platform via an acoustic modem. A commercially available stencil has been modified to preserve touchscreen functionalities at rated depths. An Android application that has been developed for this purpose has the following set of functionalities.

- The diver position transmitted from PlaDyPos is directly overlayed on an integrated Google map, allowing the diver absolute localization, as is possible on dry land where a GPS signal is present.
- Two-way communication with the surface in the form of predefined or custom short messages is enabled, as well as a single-touch alert message in case of hazards.
- Waypoints, tracks, or marked areas can be sent from the surface and displayed directly on the diver's tablet, and thus the diver can visit areas of interest sent from the ground station.

A diver carrying the tablet in the underwater casing on dry land during one of the experiments in Croatia is shown in Figure 4.

Mathematical Modeling

Modeling the ASV

Dynamic Model

Following the notation shown in Figure 5, a dynamic model of the platform in the horizontal plane can be described using the velocity vector $\boldsymbol{\nu} = [u \;\; v \;\; r]^T$, where u, v, and r are the surge, sway, and yaw speed, respectively; and the vector of actuating forces and moments acting on the platform $\boldsymbol{\tau} = [X \;\; Y \;\; N]^T$, where X, Y are the surge and sway forces and N is the yaw moment [11]. Both vectors are defined in the body-fixed (mobile) coordinate frame. The uncoupled dynamic model in the horizontal plane is given with (1), where $\mathbf{M}$ is a diagonal matrix with mass and added mass terms, and $\mathbf{D}(\boldsymbol{\nu})$ is a diagonal matrix consisting of nonlinear hydrodynamic damping terms

Conducting experiments with divers presents a challenge due to uncertainties.

$$\mathbf{M}\dot{\boldsymbol{\nu}} = -\mathbf{D}(\boldsymbol{\nu}) + \boldsymbol{\tau}. \tag{1}$$

Figure 4. A diver with the underwater tablet preparing to start the experiments. The underwater casing, rated for depths up to 50 m, allows touchscreen functionality when using tablets with inductive screens. The tablet is linked to the surface via the acoustic modem mounted on the diving tank.

Since the platform is designed to be symmetrical with respect to the x and y axes in the body fixed frame, the following forms of the two matrices are adopted: $\mathbf{M} = \text{diag}(\alpha_u, \alpha_u, \alpha_r)$, $\mathbf{D}(\boldsymbol{\nu}) = \text{diag}(\beta_u(u), \;\beta_u(v), \;\beta_r(r))$.

Kinematic Model

The kinematic equations for the platform motion in the horizontal plane on the sea surface is given with (2), where x and y are the position and ψ is the orientation of the platform in the Earth-fixed coordinate frame. The rotation matrix $\mathbf{R}(\psi)$ is given with

$$\begin{bmatrix}\dot{x}\\ \dot{y}\\ \dot{\psi}\end{bmatrix} = \begin{bmatrix}\mathbf{R}(\psi) & 0\\ \mathbf{0}_{[1\times 2]} & 1\end{bmatrix}\begin{bmatrix}u\\ v\\ r\end{bmatrix} \tag{2}$$

$$\mathbf{R}(\psi) = \begin{bmatrix}\cos\psi & -\sin\psi\\ \sin\psi & \cos\psi\end{bmatrix}. \tag{3}$$

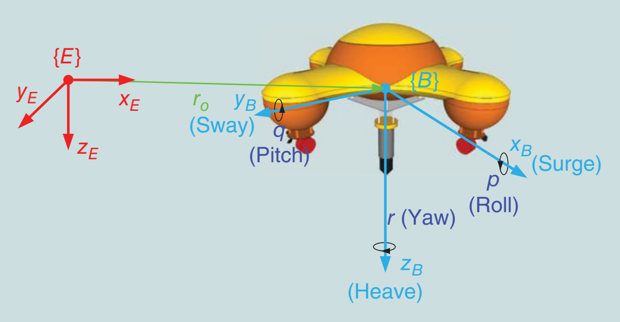


Figure 5. The body-fixed and Earth-fixed coordinate frames attributed to the ASVs. This notation is usually used in marine vehicles, as described in [11].

The platform is overactuated, i.e., it can move in any direction in the horizontal plane by modifying the surge and sway speed, while attaining arbitrary orientation.

Actuator Allocation

The actuator allocation matrix Φ gives the relation between the forces exerted by the thrusters $\boldsymbol{\tau}_{\mathbf{i}} = [\tau_1 \;\; \tau_2 \;\; \tau_3 \;\; \tau_4]^T$ and the forces and moments $\boldsymbol{\tau}$ acting on the rigid body. The actuator configuration of the autonomous surface platform for diver tracking is given in Figure 6, where $\delta = 45^{\circ}$. The allocation matrix is given with

$$\boldsymbol{\tau} = \underbrace{\begin{bmatrix} \cos 45^{\circ} & \cos 45^{\circ} & -\cos 45^{\circ} & -\cos 45^{\circ} \\ \sin 45^{\circ} & -\sin 45^{\circ} & \sin 45^{\circ} & -\sin 45^{\circ} \\ D & -D & -D & D \end{bmatrix}}_{\Phi} \boldsymbol{\tau}_{\mathbf{i}}. \qquad (4)$$

The Diver Model

Determining a simple dynamic model of a diver is practically impossible. For the specific case of the ASV tracking a diver from the surface, a kinematic model of the diver projection on the surface horizontal plane will be sufficient. For that reason, the following states are defined: x_D and y_D are the positions and ψ_D is the orientation of the diver in the Earth-fixed coordinate frame, while u_D, v_D, and r_D are the diver's linear and rotational velocities in the body-fixed frame, respectively. The kinematic model of the diver assumes that the diver cannot swim in the sway direction, i.e., $v_D = 0$ which leads to the kinematic model given with (5), where the diver's rotation matrix $\mathbf{R_D}$ is given with (6).

$$\begin{bmatrix} \dot{x}_D \\ \dot{y}_D \\ \dot{\psi}_D \end{bmatrix} = \begin{bmatrix} \mathbf{R_D}(\psi_D) & \mathbf{0}_{[2\times1]} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} u_D \\ r_D \end{bmatrix} \qquad (5)$$

$$\mathbf{R_D}(\psi_D) = [\cos\psi_D \;\; \sin\psi_D]^T. \qquad (6)$$

To enhance the estimation of the diver position, the assumption is made that the diver's surge speed u_D is constant, and the yaw speed r_D has some dynamics determined with a time constant T_D. This results in the simplified dynamic model given with (7)

$$\begin{aligned} \dot{u}_D &= 0 \\ \dot{r}_D &= -T_D r_D \end{aligned} \qquad (7)$$

The Tracking Model

The main requirement in the tracking task is to ensure that the distance between the platform and the diver in the horizontal plane $\mathbf{d} = [x - x_D \;\; y - y_D]^T$ converges to zero. The kinematic tracking model is then obtained by differentiation resulting in

$$\dot{\mathbf{d}} = \mathbf{R}(\psi)\begin{bmatrix} u \\ v \end{bmatrix} - \mathbf{R_D}(\psi_D)u_D. \qquad (8)$$

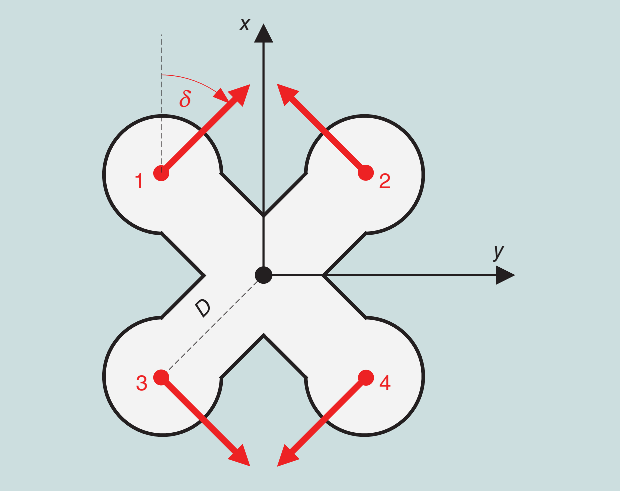


Figure 6. The actuator configuration on PlaDyPos. Four thrusters in an X configuration make the vehicle omnidirectional in the horizontal plane.

Control, Tracking, and Sensor Fusion Algorithms

A control and guidance structure is the most common cascade control structure applied for marine vehicles. The low-level control loop is in charge of speed control and takes the outputs from the upper (guidance) level as its references. Proper tuning of the low-level controllers is a prerequisite for the guidance control-loop tuning [12], whereas, in general, the guidance level is in charge of waypoint following, path and trajectory tracking, and dynamic positioning. For the described application, it is in charge of diver tracking, as shown in Figure 7.

Diver safety is seriously jeopardized during diving activity.

Speed Controller Design

For the low-level speed controller, we have chosen a proportional–integral (PI) controller in the form

$$\boldsymbol{\tau} = \mathbf{K_P}(\boldsymbol{\nu}^* - \boldsymbol{\nu}) + \mathbf{K_I}\int(\boldsymbol{\nu}^* - \boldsymbol{\nu})\,dt + \boldsymbol{\tau}_{\mathbf{F}}, \qquad (9)$$

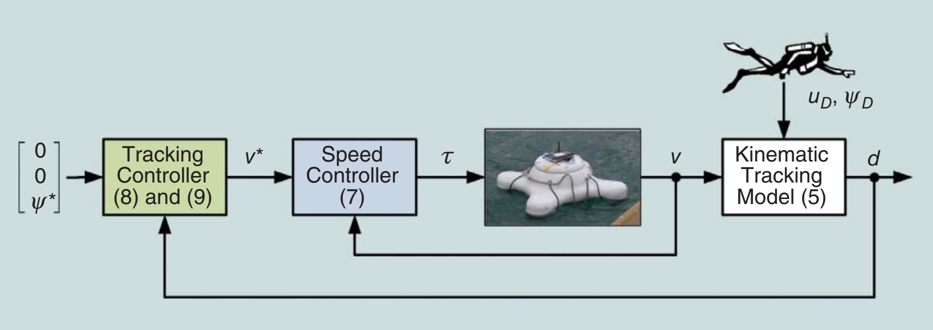


Figure 7. The cascade control structure implemented for diver tracking: low-level control is in charge of controlling the surge, sway, and yaw speed, and upper-level control is in charge of generating references for the low-level controllers.

where $\boldsymbol{\nu}^* = [u^* \; v^* \; r^*]^T$ is the desired linear and angular speeds of the platform, $\mathbf{K_P} = \mathrm{diag}(K_{Pu}, \; K_{Pv}, \; K_{Pr})$ and $\mathbf{K_I} = \mathrm{diag}(K_{Iu}, \; K_{Iv}, \; K_{Ir})$ are diagonal matrices with the PI gains for individual degrees of freedom, respectively. The $\boldsymbol{\tau}_\mathbf{F}$ term represents additional action introduced in the controller to improve the closed-loop behavior. This action can be in the form $\boldsymbol{\tau}_\mathbf{F} = \mathbf{D}(\boldsymbol{\nu})\boldsymbol{\nu}$, which results in the feedback linearization procedure, where measured or estimated speeds are used to compensate for the nonlinearity in the process.

Controller parameters $\mathbf{K_P}$ and $\mathbf{K_I}$ can be calculated based on the desired closed-loop characteristic equation, as shown in [13]. These parameters will naturally depend on the parameters of the dynamic model that have to be identified. The dynamic model parameters of the platform that is addressed in this article have been identified using the identification method based on self-oscillations reported in [14].

The diving supervisor at the surface has reliable data about the diver's position, and reliable communication.

Guidance Controller Design

Since the platform is overactuated, it can move in a horizontal plane while keeping an arbitrary heading. For this reason, the high-level guidance controller is divided into the heading controller and the tracking controller design.

Heading Controller

For the heading controller, a PI structure is chosen since it compensates for all environmental disturbances in the yaw degree of freedom. In addition, the integral action will compensate for all the unmodeled dynamics and ensure convergence of the heading to the desired value ψ^*. The controller can be written in the form

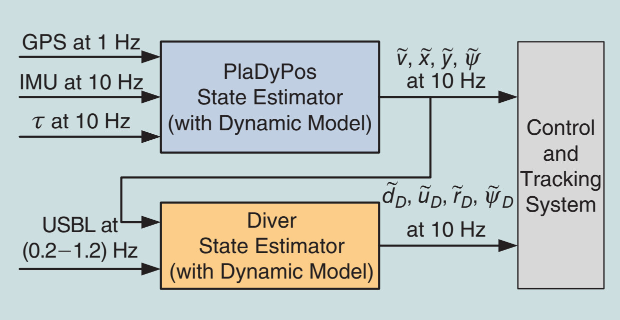


Figure 8. A schematic description of estimator inputs and outputs. Since measurements are available at different update rates, the state estimators are used to ensure the update rate of 10 Hz required for the control and tracking system. The diver estimator is also required since diver position measurements are intermittent due to possible occlusions of the acoustic link caused by air bubbles.

$$r^* = K_{P\psi}(\psi^* - \psi) + K_{I\psi}\int(\psi^* - \psi)\,dt, \tag{10}$$

where $K_{P\psi}$ and $K_{I\psi}$ are controller parameters chosen so that the desired heading closed-loop dynamics are achieved.

Tracking Controller

With the tracking model given with (8), the PI control action in the form

$$\begin{bmatrix} u^* \\ v^* \end{bmatrix} = \mathbf{R}^\mathrm{T}(\psi)\left(-\mathbf{K}_{\mathbf{P},\mathbf{d}}\mathbf{d} - \mathbf{K}_{\mathbf{I},\mathbf{d}}\int \mathbf{d}\,dt\right) + \boldsymbol{\nu}_\mathbf{F}, \tag{11}$$

where $\mathbf{K}_{\mathbf{P},\mathbf{d}} = \mathrm{diag}(\mathrm{K}_{P,\mathrm{dx}}, \; \mathrm{K}_{P,\mathrm{dy}})$ and $\mathbf{K}_{\mathbf{I},\mathbf{d}} = \mathrm{diag}(\mathrm{K}_{I,\mathrm{dx}}, \; K_{I,\mathrm{dy}})$ are PI gain matrices, respectively, will ensure convergence of the distance $\mathbf{d}$ to the desired value $\mathbf{d}^* = [0 \;\; 0]^T$. The $\boldsymbol{\nu}_\mathbf{F}$ is the feedforward action that can improve the behavior of the tracking closed loop. The proposed PI controller will ensure convergence even without the feedforward action, i.e., $\boldsymbol{\nu}_\mathbf{F} = [0 \;\; 0]^T$, [10]. However, tracking may be improved if feedforward action in the form (12) is introduced

$$\boldsymbol{\nu}_\mathbf{F} = \mathbf{R}^T(\psi)\mathbf{R}_\mathbf{D}(\psi_D)\nu_D. \tag{12}$$

The proposed feedforward action requires the estimation of the diver surge speed and heading since they cannot be directly measured.

Sensor Fusion

Two extended Kalman filters are implemented in the system, as shown in Figure 8. Their main purpose is to fuse measurements available at different update rates to ensure state estimations at 10 Hz, as required by the control and tracking system. The PlaDyPos state estimator uses the kinematic (2) and dynamic model (1) of the vehicle to provide speed and position estimates for the control and tracking system based on the input GPS and inertial measurement unit (IMU) measurements as well as the commanded thrust vector τ.

The diver state estimator uses intermittent USBL measurements, PlaDyPos states, and the simplified diver model given with (5) and (7) to estimate tracking distance and speed and orientation of the diver. Since USBL measurements are often not available due to presence of air bubbles exhaled by the diver, this estimator ensures continuous estimates required for the diver-tracking algorithms.

Benchmark Scenario for Diver Tracking

Performing real-life experiments that include humans and robots is always a complex task. The unpredictability of human nature does not allow replicability of experiments, which is why careful planning and preparation is always required. To validate and replicate diver-tracking experiments under different environmental conditions, we define a benchmark scenario that includes tracking a predefined, georeferenced, and underwater transect. A 50-m rope was laid on seabed at the test site and georeferenced using precise GPS and USBL measurements. During the experiments, the diver was required to follow the transect in both directions (up and down), with the instructions to deviate as little as possible. While the diver was tracking the transect, the ASV was tracking the diver using the acoustic positioning sensor.

Let η_{PlaDyPos} be the position of the PlaDyPos, while η is the measured position, and $\tilde{\eta}$ is the estimated position of a generic agent that is being tracked. We can then define the measure of performance of the tracking system in the form of the mean tracking error given with (13), where N is the number of samples

$$d(\eta) = \frac{1}{N}\sum_{k=1}^{N}\|\eta_{\text{PlaDyPos}}(k) - \eta(k)\|. \tag{13}$$

Similarly, the tracking error for agent position estimates can be defined with $d(\tilde{\eta})$. In the ideal case, this measure should converge to zero, but, due to a number of factors such as modeling uncertainties, measurement errors, and disturbances, this is not the case.

This metric can be used to quantify uncertainties that are present in the human–robot system, given the assumption that the agent is performing perfect tracking. These uncertainties are described in the next section.

Designing the Experiment

The main goal is to test the diver-tracking capabilities of the system, which is influenced by a number of sources of uncertainties that can compromise repeatability of results. These effects are even more emphasized in the stochastic marine environment. The existence of experimental uncertainties (Table 1), which are difficult and even impossible to model, can be attributed to one of the sources categorized in the following four groups.

Environmental uncertainties include difficult, often impossible to model, influences of wind, waves, and sea currents, whereas environmental influences can be eliminated by performing tests in laboratory conditions, demonstrating the robustness of the performance of the system in the field is a necessity.

Given that the surface vehicle and the diver estimator are described using a simplified model structure with uncertain or changing parameters, unmodeled dynamics present another source of uncertainty that influences repeatability of experiments and the tracking error itself. This category includes also uncertainties inherent to mechanical components, unpredictable faults that can occur (most often in actuators), and basic navigation sensors onboard mobile robots, such as the compass, the GPS, and the IMU.

Table 1. A summary of the most significant uncertainties in the experimental setups.

	Setup			
Source of Uncertainties	0	1	2	3
Environmental disturbances	–	✓	✓	✓
Unmodeled dynamics	–	✓	✓	✓
Acoustic sensor	–	–	✓	✓
Diver influence	–	–	–	✓

Conducting experiments with divers presents a challenge due to uncertainties.

Acoustic sensor uncertainties are most emphasized in the acoustic communication and positioning system, and they are caused by complex acoustic channel parameters, such as water temperature and salinity. Additional effects that compromise acoustic channel include multipath effects and update rates that vary depending on the acoustic channel information payload size. The accuracy of the used sensor is specified in [15] as ±0.2 m in range calculation and ±3° for bearing and elevation measurements. This accuracy analysis comes from a nearly static scenario with precise calibration and good acoustic channel conditions without multipath or air bubble interference. At a 10-m distance, the expected position noise would therefore be around ±0.5 m. Given that the USBL is mounted on a mobile platform (under the influence of waves during experiments) and the modem is mounted on the mobile agent, additional performance degradation of the acoustic

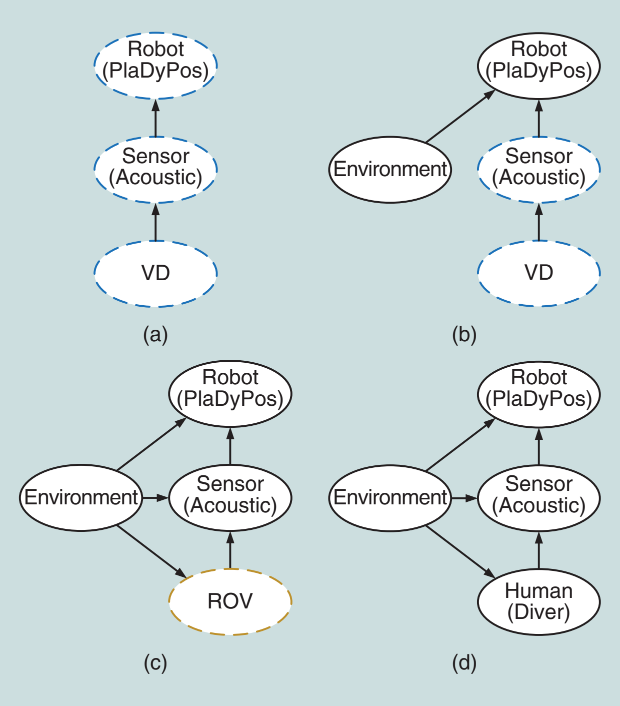
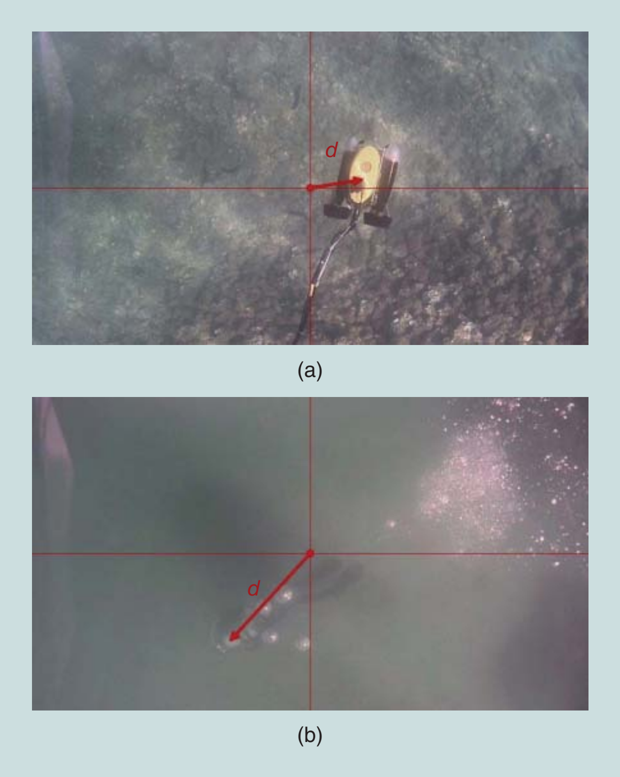


Figure 9. A schematic description of experimental setups: (a) setup 0, (b) setup 1, (c) setup 2, and (d) setup 3. The blue dashed lines represent simulated segments. Setup 0 is pure simulation. Setup 1 includes real ASV operating in the environment with a simulated diver and acoustic channels to eliminate uncertainties introduced by the acoustic sensor and the human diver. Setup 2 introduces acoustic sensor uncertainty by utilizing a manually controlled ROV connected to the ASV via an acoustic link, while setup 3 includes the diver uncertainties in the experiment.



(a)

(b)

Figure 10. The visual ground truth of the obtained results: (a) a video still from experimental setup 2 and (b) a video still from experimental setup 3. The ROV and the diver are detected in video images, and their distance to the image center is determined as the true tracking error. Note the bubble clouds that may obstruct the acoustic channel.

sensor compared with the nominal accuracy specifications is expected.

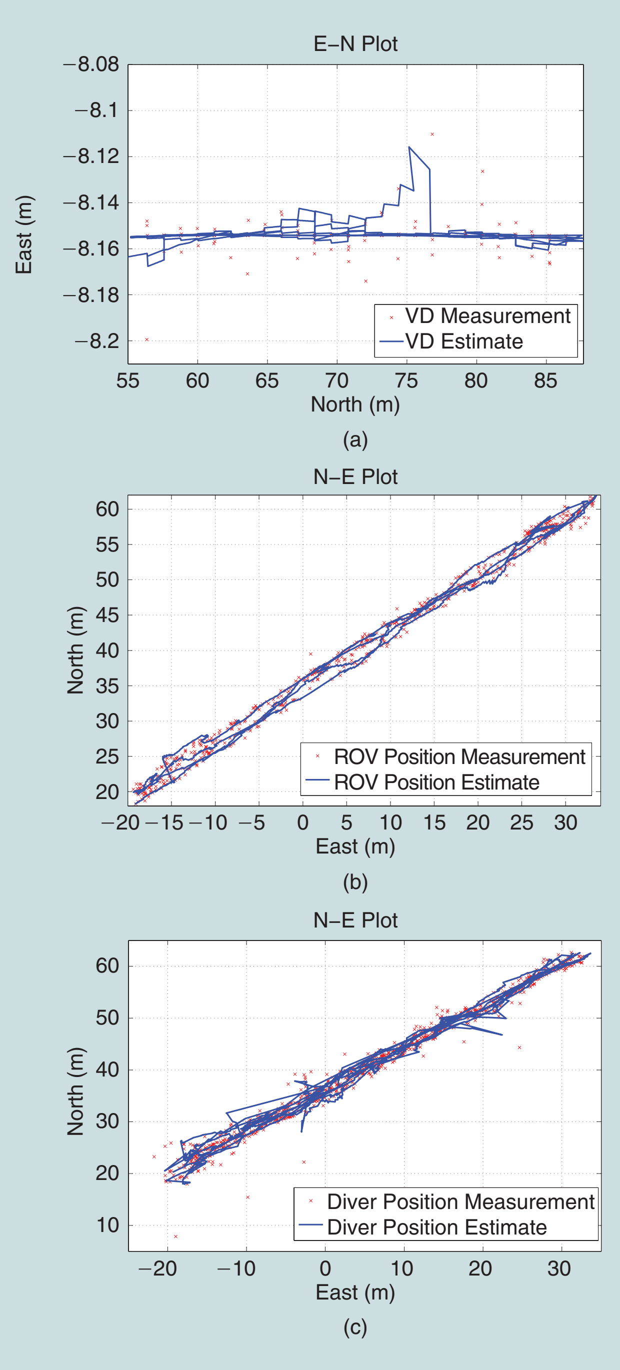
The diver's area of operation can be increased if the diving buoy is linked to the diver via a cable.

The greatest source of uncertainty is definitely the human diver. Even though the diver can be instructed to execute preplanned missions required in an experiment, there is always the issue of bubble emission, due to breathing, which may obstruct the communication channel. In addition to that, diver motion can cause different positioning of the modem relative to the USBL, influencing the quality of the acoustic communication and sometimes causing obstruction of the acoustic line of sight.

To perform the structured experiments with the diver, a step-by-step experimental plan is designed to examine the influence of the abovementioned uncertainties.

Setup 0 Simulation Experiments [Figure 9(a)]

Both the surface vehicle and the diver are simulated to test the implemented algorithms for errors and to determine the best possible performance of the diver-tracking system. This step naturally eliminates any type of uncertainty.



(a)

(b)

(c)

Figure 11. The position plots of tracked agents (VD, ROV, and human diver) during transect following in all conducted experiments. The red asterisks indicate the intermittent and noisy raw USBL measurements, while the blue lines show estimates of the positions based on the diver kinematic and dynamic model. Under the assumption that the agents were tracking the transect with little lateral deviation, variation of the measured and estimated positions in three experimental setups indicate the influence of different uncertainties. (a) Setup 1: VD and PlaDyPos. (b) Setup 2: ROV and PlaDyPos. (c) Setup 3: diver and PlaDyPos.

Setup 1 VD and PlaDyPos [Figure 9(b)]

The platform, placed in a real environment, tracks the VD that is simulated using a simple mathematical model given with (7). While this experimental setup eliminates acoustic sensor and diver related uncertainties, it also allows reliable testing of PlaDyPos behavior under different measurement update rates and performance evaluation of the diver estimator onboard PlaDyPos in real environmental conditions.

Setup 2 ROV and PlaDyPos [Figure 9(c)]

In this real-environment setup, the human diver is replaced with an ROV with an acoustic modem pinging the USBL on the PlaDyPos, and thus introducing the real acoustic channel uncertainties but eliminating those caused by the diver. This setup is designed to identify potential deterioration in system performance due to the acoustic channel characteristics.

Setup 3 Diver and PlaDyPos [Figure 9(d)]

The final experimental setup, which is in fact the demonstrator of the final goal of the described robotic system, includes experiments in real conditions with all the abovementioned uncertainties included.

To validate the results, a visual confirmation that gives a ground truth of the tracking performance is made. A down-looking camera was mounted on the PlaDyPos to validate the tracking results for setups 2 and 3 (where real agents are being tracked). Position of the agent is determined within the image, and its distance to the center of the image is calculated in pixels. Based on the known size of the ROV and the diver, the measure in pixels is transferred to meters, giving ground truth of the tracking performance. Influence of the roll and pitch of PlaDyPos is compensated using the measurements from the inertial sensor. An example of the images obtained from the two setups is shown in Figure 10.

The accuracy of the vision-based ground truth can be compromised if the camera orientation is not perfectly aligned with the gravity vector. The upper limit of the error in the observed position of the agent Δx can be estimated by using a simplified model $\Delta x = z \cdot \tan \alpha$, where z is the depth of the agent and α is the camera orientation with respect to the gravity vector. If the misalignment is not larger that $\alpha = 5^\circ$, the estimate of upper limit of the error is less than 10% of the depth of the agent.

Experimental Results

A large number of experiments, with previously described experimental setups, were conducted in June 2014 in Split, Croatia, at the Croatian Navy base. Position plots of all obtained results with a VD, ROV, and a human diver during transect following in three experimental setups are shown in Figure 11. Even in these position plots, it can be seen that the variance of measurements depends on the experimental setup, from low variance in experiments with the VD to high variance in experiments with the human diver due to a large number of sources of uncertainty.

To get a clearer picture of the influence of different sources of uncertainty on the tracking error, results from each experimental setup are analyzed. Simulation results from setup 0 are omitted from this article to keep the focus on results obtained in real environmental conditions. The simulation results can be found in [10].

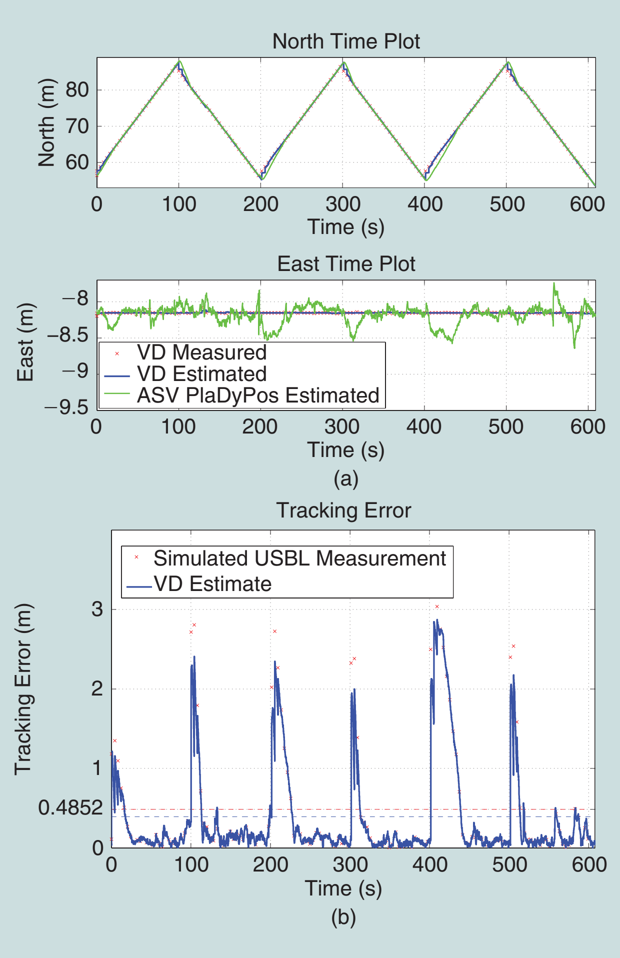


Figure 12. The experimental results obtained from setup 1 with the VD: (a) the north and east positions and (b) the tracking error. The largest tracking errors are due to abrupt changes in the direction of the VD. The diver position is estimated from the measurements with great precision, and a smooth signal is ensured for the tracking system at frequency higher than the measurement availability.

Table 2. Metric for the uncertainties: mean tracking errors of agents (VD, ROV, human diver).

	Mean Tracking Error in (m)		
	$\boldsymbol{d}(\eta)$	$\boldsymbol{d}(\tilde{\eta})$	From Video
Setup 1 $(\eta = \eta_{\text{virtual diver}})$	0.4852	0.3906	–
Setup 2 $(\eta = \eta_{\text{ROV}})$	0.9994	0.4512	0.9169
Setup 3 $(\eta = \eta_{\text{diver}})$	1.7772	1.3510	1.4831

Results for Setup 1: VD Tracking

The full experiment with the VD tracking in duration of about 10 min is shown in Figure 12. While the results given in Figure 12(a) indicate that PlaDyPos was following the same path as the VD, the real VD tracking quality is observed from

Figure 12(b). By applying (13), the mean tracking error using both measured diver positions and the estimates is shown in Table 2. Since this experimental setup is influenced only by the environmental disturbances and uncertainties caused by unmodeled dynamics, we conclude that these uncertainties cause the mean error of about 0.5 m. The error distribution, shown in Figure 13, indicates an error median of about 0.12 m with the majority of tracking errors below 1 m with a smaller number of statistical outliers.

This error is mostly due to transients that occur when the VD is changing the direction of transect following, as shown in Figure 12(b).

Results for Setup 2: ROV Tracking

The second experimental setup is designed to determine the influence of the acoustic positioning system in the diver-tracking scenario. Even though multiple experiments were performed [see Figure 11(b)], the results in Figure 14 show only 10 min of the experiment, for the sake of clarity.

Table 2 shows that the mean tracking error based on acoustic measurements in this setup is about 1 m, which lets us conclude that the inclusion of the acoustic sensor uncertainty increases the tracking error by 0.5 m. Observe the same increase in Figure 13, where the sensor uncertainty increased the median tracking error to around 0.85 m. The number of outliers did not increase, but the measured tracking errors are more spread than in setup 1.

Results for Setup 3: Diver Tracking

Finally, setup 3 allows us to quantify the influence of the human diver. Diver-tracking results obtained from one single transect coverage in upward and downward direction are shown in Figure 15.

It can be seen in the initial part of the experiment how PlaDyPos is converging above the diver. At around 290 s, Figure 15(a) shows the system behavior in situations when USBL measurements are not available for a longer period of time due to acoustic channel occlusion caused by the diver uncertainties. The estimator keeps providing the estimated diver position, and PlaDyPos tracks this estimate. Almost 30 s later, the measurements are again available, and the estimated diver position converges to the measured diver position, together with the position of PlaDyPos, ensuring high quality tracking. It should be mentioned that the diver position estimator is satisfactory for shorter periods of measurement unavailability. The specific case of more than 30 s without measurements shows that diver motion cannot be estimated for a longer period of time. The tracking error during the experiment is shown in Figure 15(b). It can be seen that, apart from the initial convergence phase and the phase when the measurements were not available, the error based on measurements is almost always below 2 m and the error based on diver estimates is below 1 m.

Mean tracking errors for all experiments with the divers are shown in Table 2, not only the single transect shown in Figure 15. The mean tracking error based on the measurements, compared

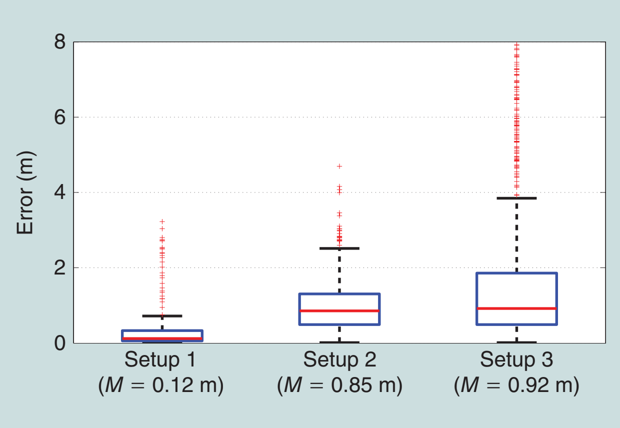


Figure 13. The distribution of the measured tracking errors for all setups. Note that for setups 2 and 3 the median tracking error (*M*) is similar since the same sensor is used. However, diver effects in setup 3 are manifested with a higher spread of tracking errors and a larger number of outliers.

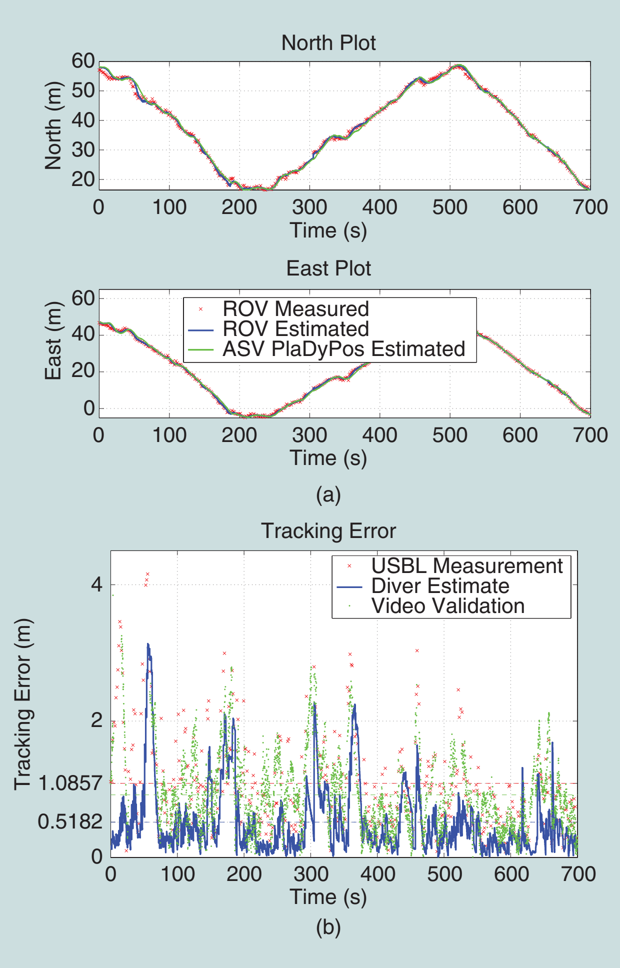
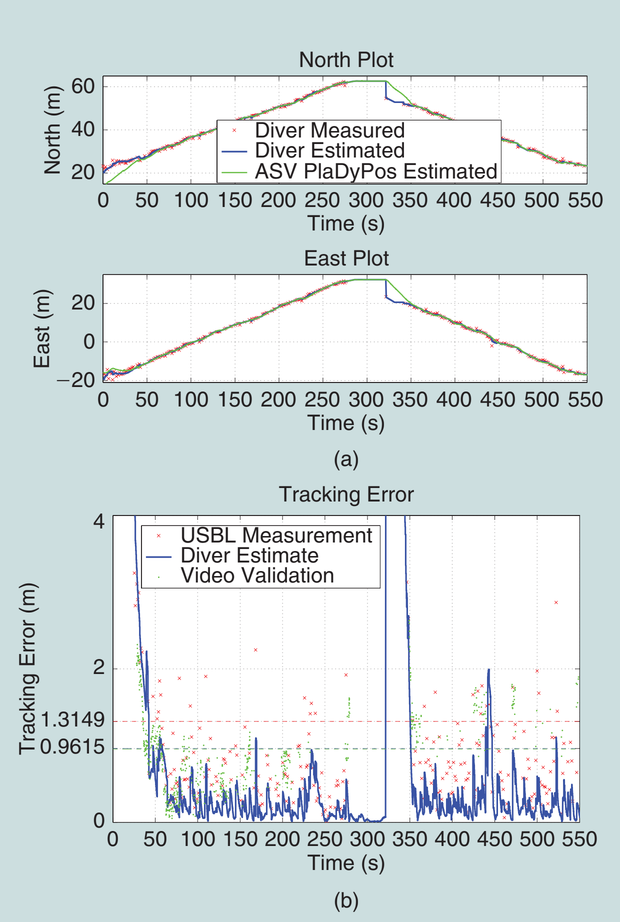


Figure 14. The experimental results obtained from setup 2 with the ROV replacing the human diver: (a) the north and east positions and (b) the tracking error. The tracking error during the experiment shows satisfactory behavior even when real acoustic measurements (red asterisk) are used. The overlayed ground-truth video validation shows that the ROV estimator is providing ROV position estimates in line with true ROV position. Horizontal dashed lines denote mean tracking error during the presented segment of the experiment while Table 2 contains values for the experiment.



(a)

(b)

Figure 15. The experimental results obtained from setup 3 with the human diver: (a) the north and east positions and (b) the tracking error. A long period of missing acoustic data is visible at around 300 s—the diver position estimator provides estimates based on a simplified diver model during this period and during other, shorter, periods when measurements are not available. The overlayed ground-truth video validation shows that the diver estimator is providing estimates in line with true diver position. Horizontal dashed lines denote mean tracking error during the presented segment of the experiment, while Table 2 contains values for the complete experiment.

with the result from setup 2, allows us to conclude that the presence of the human diver contributes an additional 0.8 m resulting in the total error of about 1.8 m. The tracking error calculated based on the diver position estimates is considerably lower. Observe that the presence of the human diver has little influence on the median measured tracking error, as shown in Figure 13. However, the increase of statistical outliers and the higher spread are attributed to the diver's presence.

Ground-Truthing the Results

Since during the experiment both the diver and the ROV were tracking the transect at a depth of about 5 m, it was possible to detect them in the video image. The accuracy of the visual ground truth is estimated to about ±0.5 m based on the error estimation analysis (due to misalignment of the camera with the gravity vector) provided before. It should be mentioned that at larger depths this type of validation would not be possible due to low visibility.

The tracking error based on video data is overlayed in Figures 14(b) and 15(b), and the mean tracking error values are listed in Table 2. These values are very close to the measured and the estimated mean tracking errors showing the accuracy of the results obtained from measured and estimated positions. It should also be mentioned that the difference between the setups 2 and 3 mean tracking error from video data is around 0.6 m, which shows that the influence of the human uncertainty determined by the acoustic measurements (0.8 m) and the estimated diver positions (0.9 m) are sufficiently accurate.

The mean error from the diver position estimates is lower due to inclusion of the diver estimator. However, it should be mentioned that if this error is too conservative, the diver motion is not estimated properly. By comparing this estimation with the video validation, we conclude that the diver estimator gives satisfactory results.

Open Data

The experiment described in this article was designed to allow future replication for comparison with new positioning sensors and methods. All software was implemented within the ROS (http://www.ros.org) framework which is used by the worldwide robotics community. During the experiments, all the relevant data were logged in an ROS bag format. Video validation footage was time stamped and logged into a separate ROS bag file due to its size. An a posteriori analysis of the experimental data was performed to identify and extract parts where actual tracking has taken place. Filtered bag files were loaded into MATLAB where the final analysis step was performed. The data and MATLAB scripts used during analysis are made publicly available at https://bitbucket.org/labust/diver-tracking, together with clear instructions on how to use the data. Making the data and scripts available in a Git repository makes future changes and contributions easily trackable.

The diver state estimator uses intermittent USBL measurements, PlaDyPos states, and simplified diver model.

Conclusions

The benchmark scenario of following a georeferenced transect laid on the seabed allowed us to execute replicable experiments with an ASV for diver tracking. Given that experiments with human divers introduce a large number of uncertainties, a structured step-by-step experimental plan was devised with the intention to identify the influence of uncertainties introduced by the environment, the unmodeled dynamics, the acoustic sensors, and the human diver. We have defined a metric in the form of a mean tracking error that allowed us to quantify influences of different uncertainties.

The results obtained using the surface vehicle tracking a VD have shown that the mean tracking error is around 0.5 m. When an ROV was used instead of the VD, uncertainties caused by the acoustic sensor were introduced, and the mean tracking error increased to around 1 m. In the final step, human diver and experimental uncertainties related to human factors (such as bubble emission) were introduced, significantly increasing the mean tracking error to 1.8 m.

Two extended Kalman filters are implemented in the system, as shown in Figure 8.

The obtained data was validated against the ground-truth data provided by the video stream from which the distance of the ROV and the diver from the surface vehicle was determined. The obtained results confirmed the tracking quality attained from the experiments using the acoustic positioning device, and proved the accuracy of the diver-tracking system.

There is a large number of parameters in the control, tracking, and estimation system that can be tuned, and a large number of control, tracking, and estimation methods that can be implemented. All the obtained data and code are made available online for public use. The results provided in this article are set as a benchmarking performance and it is left to the whole interested community to compare, analyze, and improve the performance of the diver-tracking system by using different algorithms, sensors, and vehicles.

Acknowledgment

The authors would like to thank the Croatian Navy for securing the trial area and providing logistic support, and Nikola Stilinović, Milan Marković, Filip Mandić, and Antonio Vasilijević for their assistance during the execution of the experiments.

This work was supported by the European Commission under the Seventh Framework Program project "CADDY—Cognitive Autonomous Diving Buddy" under Grant Agreement 611373 and by the Business Innovation Agency of the Republic of Croatia through the Proof of Concept Program.

References

[1] D. Schories and G. Niedzwiedz. (2012). Precision, accuracy, and application of diver-towed underwater GPS receivers. *Environ. Monitoring Assessment*. [Online]. *184(4)*. pp. 2359–2372. Available: http://www.scopus.com/inward/record.url?eid=2-s2.0-84862885600&partnerID=40&md5=f6ff96e3dc1e4c14fea3db111ff4e71d

[2] M. Stojanovic and J. Preisig, "Underwater acoustic communication channels: Propagation models and statistical characterization," *IEEE Commun. Mag.*, vol. 47, no. 1, pp. 84–89, Jan. 2009.

[3] C. Georgiades, A. German, A. Hogue, H. Liu, C. Prahacs, A. Ripsman, R. Sim, L.-A. Torres, P. Zhang, M. Buehler, G. Dudek, M. Jenkin, and E. Milios, "AQUA: An aquatic walking robot," in *Proc. IEEE/RSJ Int. Conf. Intelligent Robots Systems*, Sept. 2004, vol. 4, pp. 3525–3531.

[4] G. Dudek, M. Jenkin, C. Prahacs, A. Hogue, J. Sattar, P. Giguere, A. German, H. Liu, S. Saunderson, A. Ripsman, S. Simhon, L.-A. Torres, E. Milios, P. Zhang, and I. Rekletis, "A visually guided swimming robot," in *Proc. IEEE/RSJ Int. Conf. Intelligent Robots Systems*, Aug. 2005, pp. 3604–3609.

[5] A. Birk, G. Antonelli, A. Caiti, G. Casalino, G. Indiveri, A. Pascoal, and A. Caffaz, "The CO^{3}AUVs (cooperative cognitive control for autonomous underwater vehicles) project: Overview and current progresses," in *Proc. IEEE OCEANS—Spain*, June 2011, pp. 1–10.

[6] T. Glotzbach, M. Bayat, A. Aguiar, and A. Pascoal. (2012). An underwater acoustic localisation system for assisted human diving operations. in *Proc. IFAC Volumes (IFAC-PapersOnline)*. pp. 206–211. [Online]. Available: http://www.scopus.com/inward/record.url?eid=2-s2.0-84900508038&partnerID=40&md5=7c2bd2f8cc0f458a7df48f85066d188e

[7] K. Demarco, M. West, and A. Howard. (2013). A simulator for underwater human-robot interaction scenarios. in *Proc. MTS/IEEE OCEANS: An Ocean in Common—San Diego*. [Online]. Available: http://www.scopus.com/inward/record.url?eid=2-s2.0-84896345210&partnerID=40&md5=9ca2ba614e5b3ce83f31f0f20ad68da9

[8] H. Buelow and A. Birk, "Diver detection by motion-segmentation and shape-analysis from a moving vehicle," in *Proc. IEEE OCEANS*, Sept. 2011, pp. 1–7.

[9] K. DeMarco, M. West, and A. Howard, "Sonar-based detection and tracking of a diver for underwater human-robot interaction scenarios," in *Proc. IEEE Int. Conf. Systems, Man, Cybernetics*, Oct. 2013, pp. 2378–2383.

[10] N. Miskovic, D. Nad, N. Stilinovic, and Z. Vukic, "Guidance and control of an overactuated autonomous surface platform for diver tracking," in *Proc. 21st Mediterranean Conf. Control Automation (MED)*, 2013, pp. 1280–1285.

[11] T. Fossen, *Guidance and Control of Ocean Vehicles*. New York: Wiley, 1994.

[12] N. Mišković, M. Bibuli, G. Bruzzone, M. Caccia, and Z. Vukić, "Tuning marine vehicles' guidance controllers through self–oscillation experiments," in *Proc. 8th IFAC Int. Conf. Manoeuvring Control Marine Craft (MCMC)*, 2009, pp. 115–120.

[13] M. Caccia, M. Bibuli, R. Bono, and G. Bruzzone, "Basic navigation, guidance and control of an unmanned surface vehicle," *Auton. Robot.*, vol. 25,no. 4, pp. 349–365, 2008.

[14] N. Miskovic, Z. Vukic, M. Bibuli, G. Bruzzone, and M. Caccia, "Fast in-field identification of unmanned marine vehicles," *J. Field Robot.*, vol. 28, no. 1, pp. 101–120, 2011.

[15] A. Vasilijevic, B. Borovic, and Z. Vukic, "Underwater vehicle localization with complementary filter: Performance analysis in the shallow water environment," *J. Intell. Robot. Syst.*, vol. 68, nos. 3–4, pp. 373–386, 2012.

Nikola Mišković, University of Zagreb, Faculty of Electrical Engineering, Zagreb, Croatia. E-mail: nikola.miskovic@fer.hr.

Đula Nađ, University of Zagreb, Faculty of Electrical Engineering, Zagreb, Croatia. E-mail: dula.nad@fer.hr.

Ivor Rendulić, University of Zagreb, Faculty of Electrical Engineering, Zagreb, Croatia. E-mail: ivor.rendulic@fer.hr.

Exploring 3-D Reconstruction Techniques

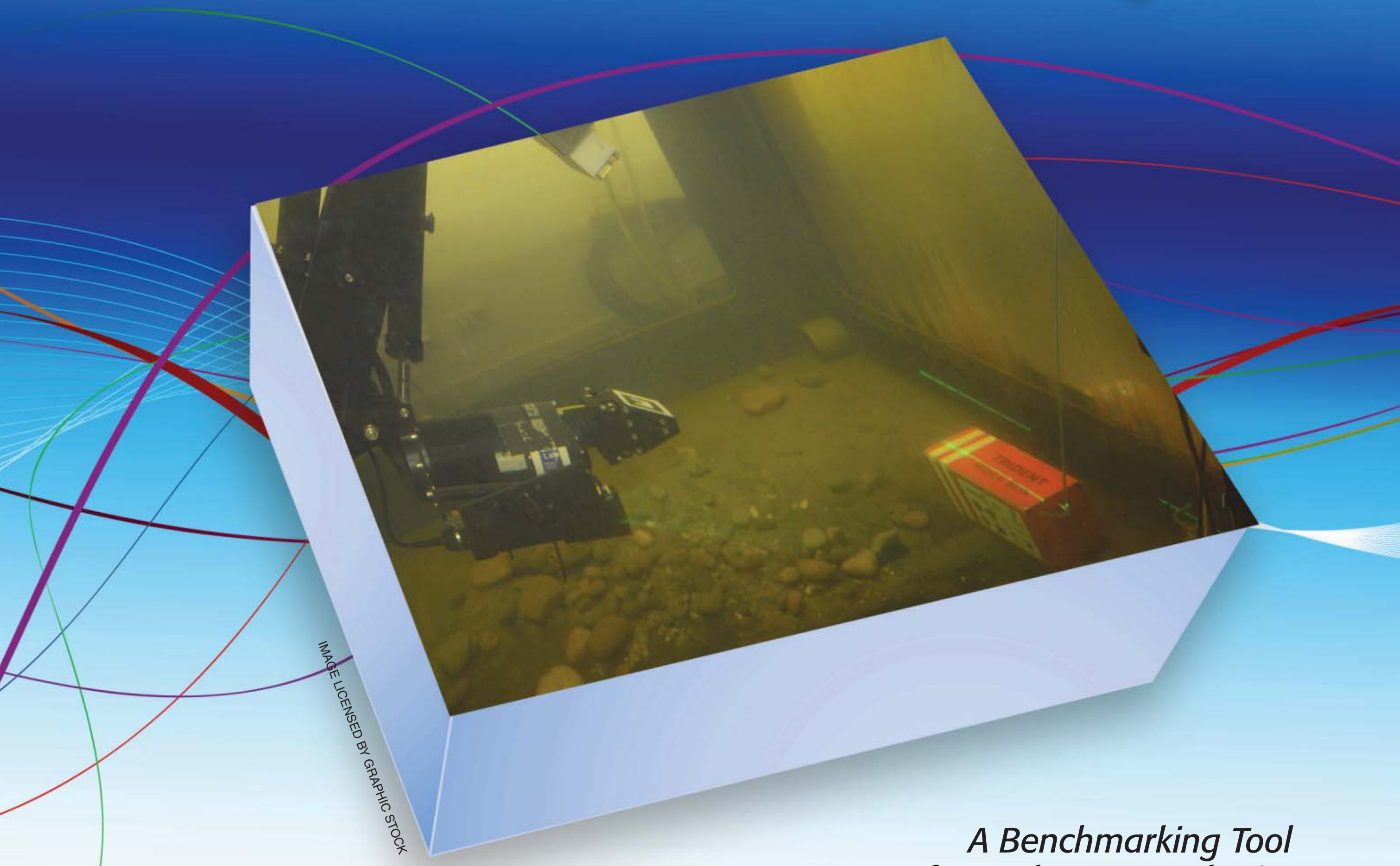


IMAGE LICENSED BY GRAPHIC STOCK

A Benchmarking Tool for Underwater Robotics

By Javier Pérez, Jorge Sales,
Antonio Peñalver, David Fornas,
José Javier Fernández, Juan Carlos García,
Pedro J. Sanz, Raúl Marin, and Marío Prats

Digital Object Identifier 10.1109/MRA.2015.2448971
Date of publication: 11 September 2015

When the results of research in the field of robotics are presented to the scientific community, the same question is asked repeatedly: Are the results really reproducible? Regarding benchmarking issues, some technological areas where complex mechatronic devices, such as robots, have a central role are very far from other research areas like physics or chemistry, to name but a few, where reproducibility is always mandatory. Aside from mechatronic complexities, the comparison between two different algorithms in the same conditions is influenced by the experimental validation scenario. In underwater environments, the difficulties for benchmarking characterization increase substantially. This is especially true when the test bed is the sea where uncertainty is high. It is the aim of this article to present a software

1070-9932/15©2015IEEE

tool which enables a comparison between two different algorithms when the algorithms are being used to solve the same problem in water tank conditions. This is a preliminary stage before the final validation on the seabed. The evaluated algorithms fall into the three-dimensional (3-D) image reconstruction context, as a prior step to their autonomous manipulation. Performance results are presented for both simulation and real water tank conditions.

Robotics Benchmarking

Over the last few years, considerable effort has been made in robotics benchmarking. For instance, the European Robotics Research Network has been active in this context [1] and has recognized the interaction of a robotic manipulation system with its environment as a key area for research. Recent European projects, like Best Practice in Robotics (BRICS), funded under the European Union's Seventh Framework Programme (FP7), significantly contributed to this specific subject [2] by promoting the interoperability of hardware and software components and the creation of a software repository of best practice robotics algorithms [3]. Moreover, following previous research in this field [4], it is clear that in the domain of robotics research, it is extremely difficult not only to compare results from different approaches, but also to assess the quality of the research. This is especially true if one wishes to evaluate the performance of intelligent robot systems interacting with the real world. There are many definitions for the term benchmark, but we use a simple one stated in [4], which defines a benchmark as a standardized problem or test that serves as a basis for evaluation or comparison.

> **The comparison between two different algorithms in the same conditions is influenced by the experimental validation scenario.**

The aim of this article is to present a benchmarking tool in the underwater intervention context so that a suitable comparison between different algorithms with the same goals can be made in the same context and with the same robotic platform. For a better understanding, a pair of algorithms will be compared, highlighting the main facilities available through this tool. Moreover, the same algorithms will be tested and compared in simulation and in real water tank conditions.

Benchmarking for Underwater Intervention Systems

The use of underwater robots is becoming more widespread because technical advances has made them increasingly useful. Some examples can be found in the oil and gas industry (e.g., operating submerged infrastructures), search and recovery missions (e.g., recovering a black box from a crashed airplane), deep-water archeology or scientific missions. Usually, the robots used in these missions are remotely operated vehicles (ROVs), which is both expensive and logistically difficult. In addition, these robots use a master/slave architecture which puts all of the responsibility on the pilot, who in turn can suffer from cognitive fatigue and stress. The evolution of this kind of robot to the new intervention autonomous underwater vehicle (I-AUV), removes the human from the control loop (and therefore, the problems related to the pilot) and increases the intervention capabilities being able to improve robot precision and intervention time by avoiding delays caused by the difficulties in communication.

Experimentation with underwater robots is normally difficult due to the wide range of resources required. For instance, a water tank deep enough for the systems to be tested is normally needed and this implies significant space and maintenance costs. Another possibility is the access to open environments such as lakes or the sea, but this normally involves high costs and requires special logistics. In addition, the nature of the underwater environment makes it difficult for researchers (operating on the surface) to observe the evolution of the running system. As a consequence, experimental validation of these systems is laborious. In order to facilitate the development of underwater robots, it is of utmost importance to develop suitable simulators that make it possible to develop and benchmark the systems before they are deployed and supervise a real underwater task where the developers do not have a direct view of the system.

Usually, the experimental validation in an underwater intervention system takes place in the sea, where there are many changing parameters, such as underwater currents and bad visibility that is impossible to model and replicate. These uncertainties are the main issue when comparing and replicating results from different studies. The use of an automated comparison system in simulation and controlled environments helps to establish an objective benchmarking methodology.

There are previous simulators for underwater applications, which mainly have remained obsolete or are being used for specific purposes. In [5] and [6], a review of virtual simulators for AUVs can be found. Nevertheless, the majority of the reviewed simulators have not been designed as open source, which makes it difficult to improve and enhance the capabilities of the simulator. Other simulators, such as ROVSim, VMAX, or DeepWorks, have been designed to train ROV pilots, which is not the objective of our research.

Underwater manipulation using I-AUV allows for the design of new applications such as the one studied in the FP7 TRIDENT project [7], where a black box from the seabed was autonomously recovered. To accomplish this, the use of the underwater simulator (UWSim) [8] in continuous development was crucial for testing, integration, and benchmarking.

Review of Related Benchmarking Suites and Toolkits

In recent years, several benchmarking suites have been developed in the field of robotics. Many of them focus purely on a specific subfield of robotic research but, to the best of our knowledge, none of them are focused on autonomous underwater vehicles. In the grasping field, several suites have been presented, such as the OpenGrasp Benchmarking suite [9]. This suite is a software environment for comparative evaluation of grasping and dexterous manipulation using the Open-Grasp toolkit. It also provides a Web service that administers available benchmark scenarios, models, and benchmarking scores.

Another interesting benchmarking suite in the field of grasping is VisGrab [10] (a benchmark for vision-based grasping), which provides tools to evaluate vision-based grasp-generation methods. Motion planners, trajectory tracking and path planning have been active research fields around benchmark metrics and benchmarking suites. Cohen et al. [11] describe a generic infrastructure for benchmarking motion planners. This infrastructure makes it possible to compare different planners with a set of measures. The key point of the contribution is the easy to compare design due to robot operating system (ROS) [12] MoveIt! integration.

Rawseeds [13] is a project focused precisely on benchmarking in robotics, although its global nature has been widely used for simultaneous localization and mapping (SLAM). The Rawseeds project aim is to build benchmarking tools for robotic systems through the publication of a comprehensive, high-quality benchmarking toolkit composed of data sets with associated ground truth, benchmark problems based on data sets and benchmark solutions for the problems. Unfortunately, this project lacks an automated comparison system. Finally, there have been proposals of Web-based benchmarking suites such as [14] where authors propose an interesting test bed Internet-based architecture for benchmarking visual servoing techniques, allowing users to upload their algorithms.

UWSim: A 3-D Simulation Tool for Benchmarking and HRI

The UWSim (Available online: http://www.irs.uji.es/uwsim) is an open source software tool for visualization and simulation of underwater robotic missions that offers benchmarking capabilities through a specific module. The software is able to visualize underwater virtual scenarios that can be configured using standard modeling software and can be connected to external control programs using ROS interfaces. The UWSim is currently used in different ongoing projects funded by the European Commission: FP7-marine robotic system of self-organizing, logically linked physical nodes (MORPH) and FP7-persistent autonomy through learning, adaptation, observation, persistent autonomy through learning, adaptation, observation, and replanning (PANDORA); to perform hardware in the loop experiments and to reproduce and supervise real missions from the captured logs.

The main objective in the simulator development are that it is easy to integrate with existing architectures. To be general, it is modular and easily extendible, it supports underwater manipulators, and is as realistic as possible. From a technical point of view, the simulator has been implemented in C++ and makes use of the OpenSceneGraph (OSG), ROS, and osgOcean libraries.

The UWSim is divided into different modules (see Figure 1). There is a core module in charge of loading the main scene and its simulated robots, an interface module that provides communication with external architectures through ROS, and a dynamics module that implements underwater vehicle dynamics. This module has been designed as a generic dynamics module for underwater vehicles but users can replace it with a more accurate one using ROS interfaces or even using the real process as input. A physics module that manages the contacts between objects in the scene, an osgOcean, in charge of rendering the ocean surface and special effects, the graphical user interface (GUI) module, that provides support for visualization and windowing toolkits, and the user interface abstraction layer (UIAL) and benchmarking modules that will be explained later.

Nowadays, reproducibility of experiments in robotics is a main concern for the scientific community.

The scene is defined with an extensible markup language (XML) file, which is loaded in a scene graph by OSG, getting access to the nodes easily (i.e., visualization effects, virtual cameras, and so on). The UWSim includes, by default, some scenarios (i.e. the swimming pool facilities at CIRS, Underwater Robotics Research Center, Universitat de Girona, Spain; a shipwreck, etc.), an I-AUV (the Girona 500 robot) and two different underwater robotic arms (Lightweight ARM5E [15] and a Mitsubishi PA10 Arm).

As mentioned before, robots and scenarios can be created with any modeling software (e.g., Blender). Nowadays, some sensors (e.g., simulated position, lasers, sensors to measure distances to obstacles, and so on) and virtual cameras can be attached to the robots. Dynamics using a state-space model and physics using a Bullet engine are supported. Other interesting features are the widgets, which are small windows that can be placed inside the scene to show specific data to the user, and the multiresolution terrain compatibility, allowing the user to load complex meshes with multiresolution textures generated externally from bathymetry and imagery.

The simulator is in continuous development. Some of the recently added features consist of new ROS versions and OSG libraries compatibility, multibeam sensor simulation, a texture projector to simulate structured light, ROS transform publishing, force sensor integration with the physics engine, Dantzig physics solver to improve the robotics and manipulation capabilities, and visualization improvements (trajectory trails, point clouds, and so on).

The work in progress in the UIAL module can be divided into different aspects:

- improving the information to be shown to the user, reducing the data depending on the mission and context

- integration with an immersive system, where the user would get the feeling of being inside the robot
- adding a natural gesture control interface to control robot, improving traditional ways of robot control (e.g., leapmotion)
- implementing an abstract layer, which will manage all these improvements and will make it possible to integrate every component within most of the current architecture.

Although stereo vision is more accurate, laser reconstruction demonstrates better reliability and robustness in real underwater conditions.

Benchmarking Platform Description

Benchmarking Module for UWSim

Recently, a benchmarking module for the UWSim has been developed [16]. Like UWSim, this module uses ROS to interact with other external software. The ROS interface permits users to evaluate an external program, which can communicate both with the simulator (which can send commands to perform a task), and with the benchmarking module (which can send the results or data needed for evaluation). Detailed information on how to configure and run a benchmark in UWSim can be found online at the UWSim benchmarks workspace, http://sites.google.com/a/uji.es/uwsim-benchmarks.

For the development of the module, two important objectives were taken into account. The first one was that the module had to be transparent to the user, in other words, it does not require major modifications to the algorithm to be evaluated. The other objective of the module was that it must be adaptable to all kinds of tasks in the underwater robotics field.

Benchmarks are defined in XML files. Each file will define which measures are going to be used and how they will be evaluated. This allows the creation of standard benchmarks defined in a document to evaluate different aspects of underwater robotic algorithms, and the ability to compare algorithms from different origins. Each of these benchmarks

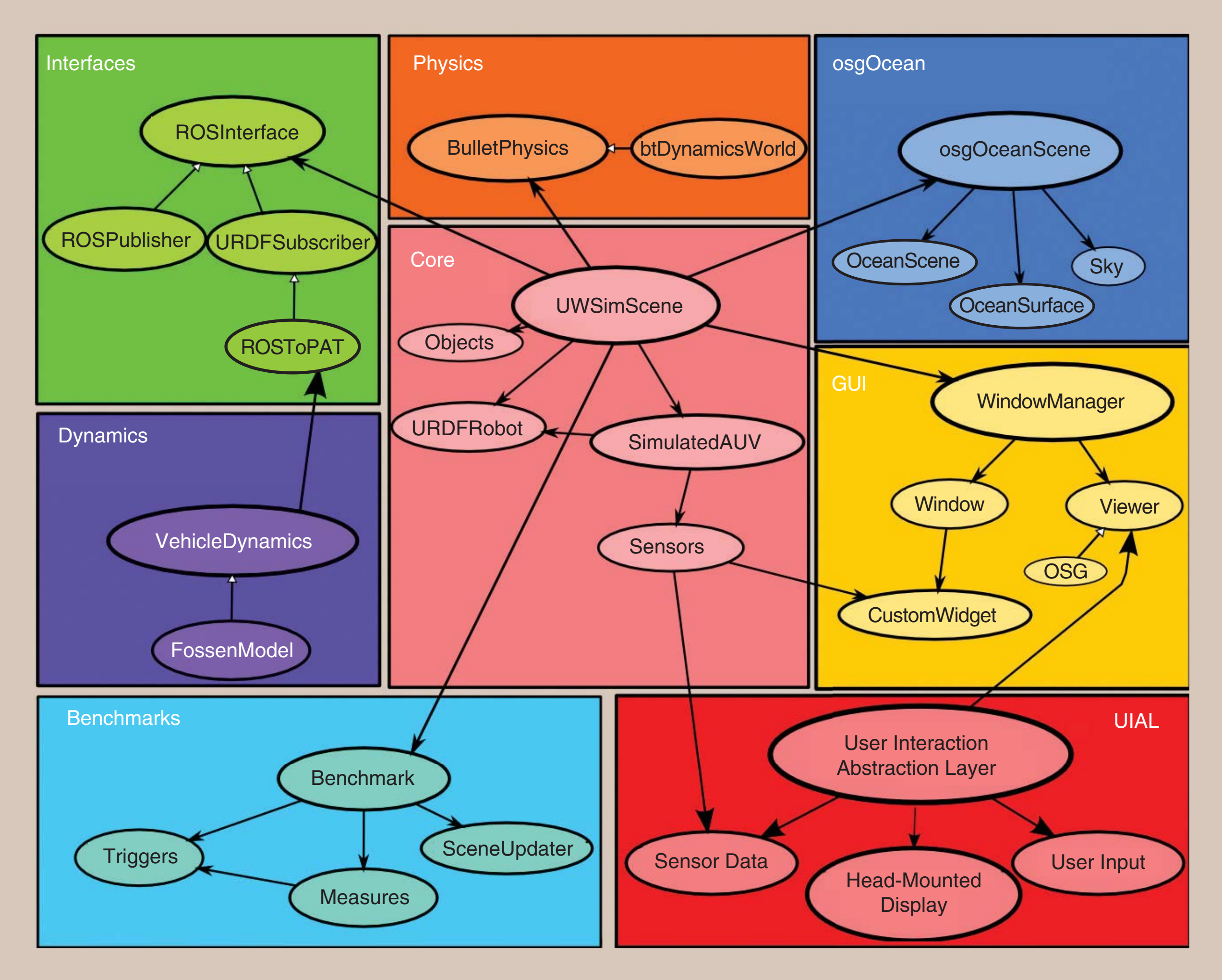


Figure 1. The UWSim modules diagram and its interconnections: core, interfaces, physics, dynamics, osgOcean, GUI, UIAL, and benchmarks.

will be associated with one or more UWSim scene configuration files, since the results of the benchmark are dependent on the predefined scene. Consequently, creating a new benchmark experiment is as simple as editing a configuration file. The whole process is shown in Figure 2.

The benchmark configuration options are basically made up from three kinds of entities: 1) measures, 2) triggers, and 3) scene updaters. These entities have been created in a modular way, thus users can extend them and create new functionality easily. Measures can be chosen from a wide variety of already implemented measures, such as position error, elapsed time, distance traveled, path-following error, reconstruction 3-D, and so on. Some of these measures are split into different parts that will be shown in the final results, for instance position error is formed by *X* error, *Y* error, and *Z* error. Setting these parameters will allow the benchmark to measure each of the configured options using the ground truth from UWSim or external sources via ROS, depending on the availability and configuration options. In the case of positioning errors, ground truth is taken from UWSim and path following requires a path to follow configured via ROS input.

These measures are activated or deactivated depending on events configured through triggers. These events allow users to measure results in an easier way, for instance to start (or stop) measuring when a vehicle reaches a position, a message is received in ROS, or the vehicle moves. A case where this might be used is to measure collisions only when the vehicle is navigating and stop when the manipulation starts, when the hand should collide with the manipulated object but it is not a bad result.

Finally, scene updaters modify the simulated environment and restart the measurement being able to start a series of experiments. Possible scene updaters are the underwater current updater, the ambient light updater, the camera noise updater, and so on. This feature is useful to create automated tests that can check the influence of environmental parameters helping to create more robust algorithms.

Once the benchmarking has finished, caused by a stop trigger event, and all the scene configurations in the scene updater have been tested, results are written into output files. These output files are disaggregated for each scene configuration containing results for each measure and global results that can combine multiple measure results. For instance, a travel efficiency benchmark could use two measures: distance traveled and battery consumption and a global result of distance/battery. In addition, each measure can be configured to log its result at regular intervals in order to see its evolution over time and not only the final result for each scene configuration. As a result, for each logged measure, the benchmark will generate a different output file containing the variation among the measured results.

The current state of the art in autonomous underwater intervention has been demonstrated through the TRIDENT project.

Physical Benchmarking System

To be able to evaluate the compared algorithms and validate the simulated results in a real platform, a physical benchmarking platform has been used (see Figure 3). It consists of the following elements, considering the main elements in the virtual scene used in UWSim:

- A water tank that is 2-m wide, 2-m long, and 1.5-m high.
- A four-degree-of-freedom ECA-CSIP Lightweight ARM 5E manipulator [15].
- A floating structure (underwater vehicle prototype) to hold the arm. In this article, the floating structure has been fixed to the water tank.
- A Bowtech DIVECAM-550C-AL COLOUR camera.
- A Tritech SeaStripe Laser Line Projector (MKIII).
- A Videre stereo camera.
- A black-box mockup that is 140 mm wide, 300 mm long, and 160 mm high.

In order to obtain the object pose ground truth, a pose estimation method has been used to compute the relative

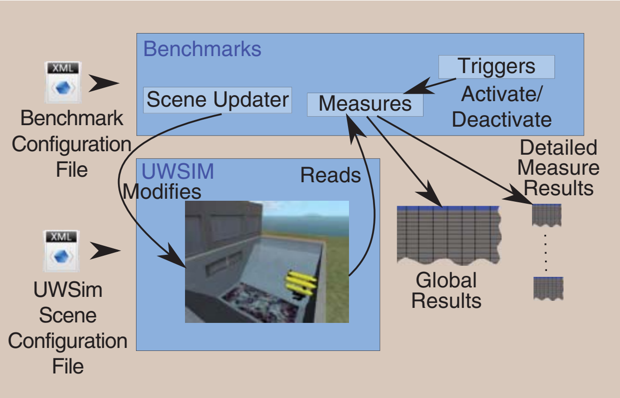


Figure 2. The benchmarking module flow diagram: a benchmark configuration is loaded into the benchmark module and a scene is loaded into the simulator. Then, the benchmark module produces results that can be logged for posterior analysis.

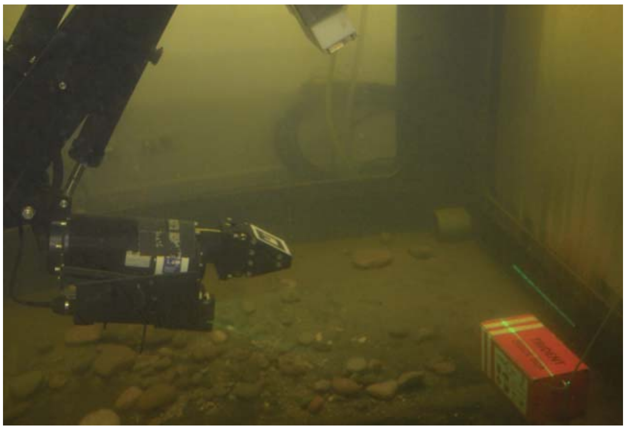


Figure 3. The physical benchmarking system: a water tank, a Lightweight ARM5E manipulator, a stereo camera, a laser stripe projector, and black box mockup.

position of the target object (black box mock up) with respect to the camera, considering that the arm is firmly fixed to the water tank. As dimensions of the object are known, the box corners can be used to estimate its pose. In the case of using a different object, easily recognizable points could be used instead of corners. While it is possible to detect them automatically, it has been judged that the manual initialization by the user is less error prone and best suited to get an accurate ground truth for the benchmarking system.

> **The final experimental validation in underwater intervention system takes place in the sea where there are many changing parameters, such as underwater currents.**

First, the user clicks on the visible corners on the box (in this approach six corners were visible). Then, after matching the obtained 3-D points with the real object using the camera parameters, the object pose estimation is obtained using the ViSP library. In this case, the frame is placed in the center of the top face of the box (see Figure 4). As there are several methods that can be used to obtain the estimation, all of them are used to estimate the pose and the one that minimizes the estimation error is selected.

This ground truth, however, is not perfect as small errors appear because of the limited camera resolution (actual pixel size), user accuracy, and camera calibration. Nevertheless, the resulting error is small enough to allow the object position to be considered as a suitable ground truth so that the metrics described in this article can be used. In fact, the camera calibration accuracy affects rectification and undistortion in these cameras, thus introducing some shared error in this ground truth position and in the reconstruction processes.

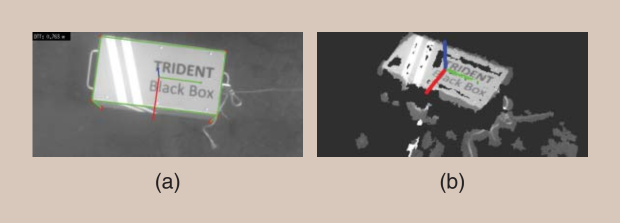


Figure 4. (a) The camera image with manually initialized corners and estimated box pose. (b) Ground truth box pose with the point cloud obtained with the stereo camera.

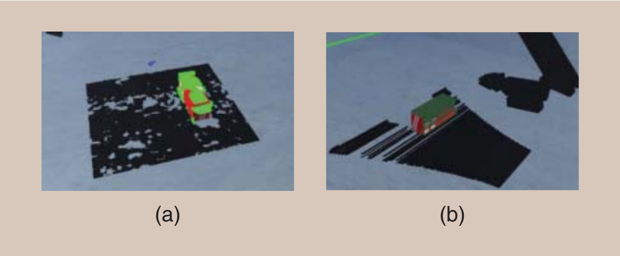


Figure 5. (a) and The stereo and (b) laser stripe reconstructions comparison over-layed on the ground truth on the simulated environment. Black points are filtered as ground, blue points are considered outliers, and green points represent the reconstructed object.

Experimental Specification

To test the benchmarking platform, two reconstruction algorithms are presented: 1) a stereo reconstruction using a stereo camera [see Figure 5 (a)] and 2) a laser stripe segmentation [see Figure 5 (b)]. These approaches are used to obtain a point cloud from the scene, as a consequence, the object can then be processed. Both algorithms are exactly the same whether used in a simulation or in a real setup.

Stereo Reconstruction Description

The aim of the stereo reconstruction is to obtain a 3-D reconstruction in the form of a dense point cloud, where each image pixel is used in order to obtain a 3-D point instead of computing them for certain features only (sparse reconstruction). A good reconstruction can be obtained only if the camera parameters are properly estimated. The parameters are computed with camera calibration tools and a calibration checkerboard.

In runtime, images from left and right side are undistorted and rectified using the aforementioned camera parameters so that their scanlines align for fast stereo processing. Once the images are aligned, a local dense stereo correspondence algorithm can be applied. In this case, the OpenCV block matching algorithm [17] implemented in a ROS package is fast enough for most robotic applications, while only needing parameter tuning. With the chosen algorithm, both disparity images and dense point clouds can be obtained. This method estimates the corresponding pixel on the right image for every pixel on the left image, thus comparing each pixel to a block on the other image. The displacement between the two pixels is used to determine the 3-D point coordinates based on the camera geometry computed in the calibration step.

Laser Reconstruction Description

Before the system is able to perform a reconstruction, it is necessary to calibrate it (see Figure 3). Therefore, with the aid of a marker placed in the gripper of the arm, the transformation between the camera and the end-effector (${}^{c}\boldsymbol{M}_e$), is calculated [18]. Then, using the direct kinematics of the arm, the relationship between the base of the arm and the end-effector is obtained (${}^{b}\boldsymbol{M}_e$). Thus, using these two matrices the transformation between the base of the arm and the camera can be calculated as ${}^{c}\boldsymbol{M}_b = {}^{c}\boldsymbol{M}_e * ({}^{b}\boldsymbol{M}_e)^{-1}$.

The next parameter that has to be obtained is the relationship between the laser and the end-effector (${}^{b}\boldsymbol{M}_e$). The camera installed in the vehicle is a stereo camera so, even though just one lens is used for the laser reconstruction; and the two lenses are used for this step of the calibration. Using the stereo camera, the 3-D position of the pixels projected by the laser is obtained by triangulation. With those 3-D points, the RANSAC algorithm is used to determine the planar parameters of the laser plane [19] $({}^{c}\boldsymbol{M}_{l})$. These parameters are referenced to the stereo camera using the previously obtained transformation between the camera and the end-effector $({}^{c}\boldsymbol{M}_{e})$, and it is possible to reference the plane of the laser respect to it: ${}^{l}\boldsymbol{M}_{e} = ({}^{c}\boldsymbol{M}_{l})^{-1} * {}^{c}\boldsymbol{M}_{e}$.

Concerning the reconstruction, the floor is scanned by moving the elbow joint of the manipulator at a constant velocity between two predefined joint positions. At the same time, the camera captures images of the scene with the laser projected on it. For each image, a laser peak detector algorithm is used to segment the laser stripe from the rest of the image. This algorithm discards the pixels that are out of a predefined threshold of size, saturation, and value. Then, because the laser pattern is a straight line and the camera is placed parallel to that line, there is only a point illuminated by the centroid of the laser at each column of the image. As a consequence, for each column of the image, the pixel with the highest intensity is selected and the center of mass algorithm is applied to this pixel and the five pixels above and below it to obtain, with subpixel accuracy, the position illuminated by the centroid of the laser.

Finally, the segmented laser stripe is triangulated to obtain its 3-D position [20]. In order to triangulate each selected pixel, it is necessary to know the relationship between the camera and the laser $({}^{l}\boldsymbol{M}_{c})$ when the image is captured. Thus, when each image is taken, the values of the joints in this moment are also read. Using these values and the direct kinematics of the arm, the transformation between the end-effector and the base of the arm $({}^{b}\boldsymbol{M}_{e})$ is calculated. Finally, using this relation and the ones obtained in the calibration, the desired transformation can be easily calculated as ${}^{c}\boldsymbol{M}_{l} = {}^{c}\boldsymbol{M}_{b} * {}^{b}\boldsymbol{M}_{e} * ({}^{l}\boldsymbol{M}_{e})^{-1}$.

Benchmarking Metrics

These methods are compared in a simulated and a real environment, using the proposed benchmarking architecture, and considering four metrics measured using a high-fidelity model of the object to be reconstructed as ground truth. The benchmarking module takes this object model as ground truth and a configuration file to get the position of the object with which the results can then be calculated. These four metrics have been introduced in [21], a work about reconstruction metrics, as quality measures of 3-D models to find the following:

- Mean error—The averageAverage distance from every 3-D reconstruction inlier point to the nearest point in the object surface.
- Standard deviation—The standard deviation for the previous error. A high value in this deviation means misalignment in the reconstruction, due to bad calibration.
- Coverage—The surface percentage that is nearer than a precision threshold to a 3-D reconstruction point. It measures the percentage of the target that is correctly reconstructed. The threshold should be chosen depending on the experimental setup. It is not an inlier measure, instead of measuring the percentage of points near the target, it measures the percentage of the target that has a reconstructed point nearer than a threshold. For instance, a perfect reconstruction of three faces of a box would return 50% instead of 100% that would get an inlier metric.
- Outliers—The percentage of the reconstruction that it is further than a threshold from the target.

In order to measure the object reconstruction only, reconstruction points that are part of the ground, such as the object that is lying on it, and outliers are filtered and do not count for the previous described measures.

Besides mathematical results, in this case, the benchmarking module is able to overlay the 3-D reconstructed point cloud on the simulated 3-D scene to get a visual result of the reconstruction using UWSim as a visualization engine. Furthermore, 3-D points are colored to show outliers, filtered ground points, and object points. This is valuable not only to show results, but also to debug reconstruction algorithms.

The use of underwater robots is becoming more widespread because technical advances has made them increasingly useful.

Results

The following subsections describe and analyze the results of the benchmarking process. The evaluation of the proposed algorithms has been performed in both simulation and real environments using the aforementioned measures. These results are easily replicable because of the software used. The simulator, benchmarking platform, and algorithms used are all open source. Furthermore, metrics, the ground truth acquisition, and experimental setup have been described with enough detail to allow other studies to be compared with the following results.

Simulated Results

In the simulation experimentation, benchmarking capabilities have been exploited to test both reconstruction techniques under different conditions of light and noise. These conditions try to simulate the complex and adverse conditions in the underwater environment. Theoretically, a laser should not be affected by low illumination conditions as it produces its own light, but it may be more difficult to detect on brighter scenes. The presence of noise should cause a poorer performance on both methods.

The algorithms have been tested in conditions where the amount of light varies, ranging from 0 to 1 ratios where 0 means total darkness and 1 is the correct illumination (default values in the UWSim), as can be seen on Figure 6. Gaussian noise has been added to the camera output from 0.00% standard deviation to 0.10% in an additive manner on

Figure 6. The increasing light conditions on a virtual camera.

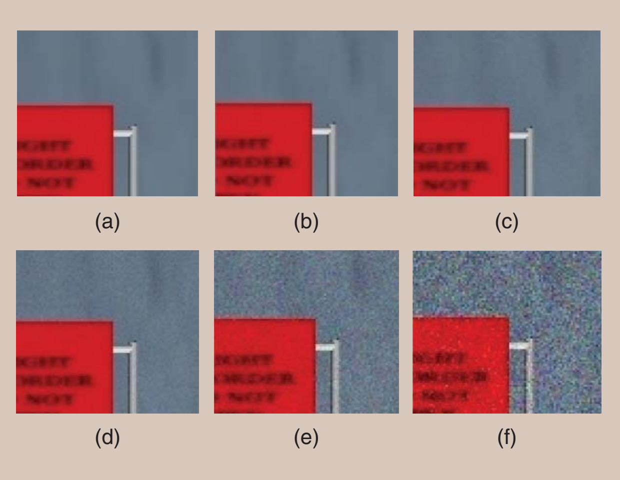


Figure 7. (a)–(f) The increasing noise conditions on a virtual camera.

red–green–blue channels through the UWSim configuration. The noise effect can be seen in Figure 7.

Coverage results for laser reconstruction and stereo vision can be seen on Figure 8. As expected, stereo vision reconstruction needs some light to achieve a good reconstruction, while the laser is nearly immune to light variation and even decreases its performance in conditions where the light is bright. Regarding noise, stereo vision is again more sensitive to noise, especially in lower visibility conditions, and the laser shows no noticeable differences between different noises on coverage. In absolute terms, the laser is able to reconstruct 50% of the object in almost every situation and stereo vision reconstructs 40% of the object in good light conditions. Thus, it can be concluded that laser is better for the tested environment.

The mean error and standard deviation for both algorithms show similar results. In the case of mean error, due to the similar setup, both algorithms achieve 0.004 m, which is a good value given the experimental setup. As the alignment is perfect on simulation, the standard deviation is negligible.

Finally, the outlier results are depicted in Figure 9 for both the laser and stereo vision cases. As the results show, stereo vision generates more outliers in the absence of light, while the laser reconstruction produces a higher number of outliers in the presence of light. In both the

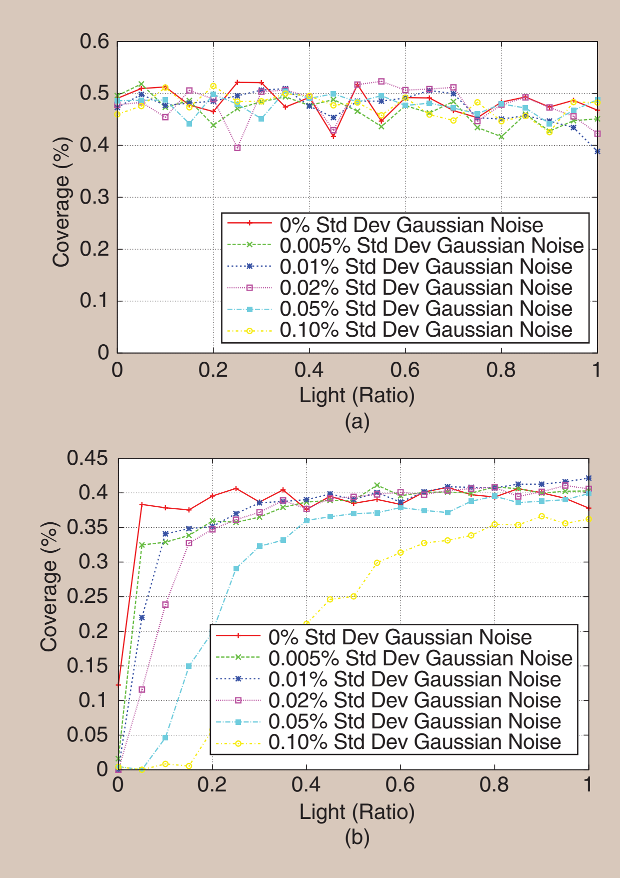


Figure 8. The coverage results on varying light conditions for different Gaussian noise on camera for (a) the laser and (b) the stereo camera.

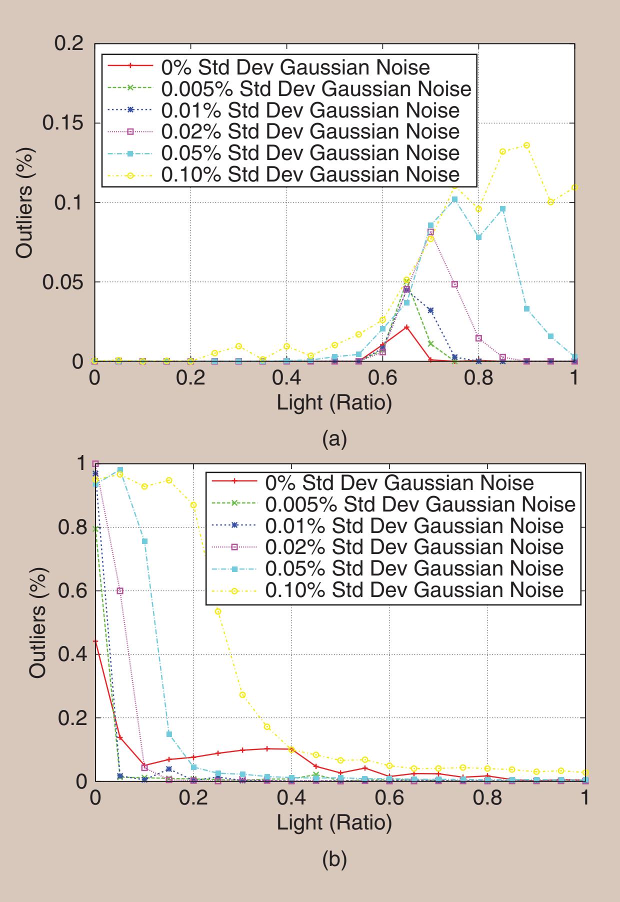


Figure 9. The outlier results on varying light conditions for different Gaussian noise on camera for (a) a laser and (b) a stereo camera.

cases, higher Gaussian noise means a higher number of outliers. In the case of stereo vision, noise and the absence of light makes it more difficult to match the pixels on both cameras, and a large number of outliers appear. On the other hand, in strong light conditions, it is more difficult to find the laser light on the camera and mistaken detections cause a high number of outliers. It is also remarkable that there is a nonnegligible number of outliers in the stereo vision results for 0.00% standard deviation noise. These outliers are caused by small floating particles simulated in UWSim, which are correctly detected by the system although they are not part of the object. The 3-D stereo reconstruction is not able to find these outliers when the noise is higher and its impact decreases as more parts of the object are correctly reconstructed.

Real Results

In the experiment conducted under real conditions, both systems have been tested under three different light conditions, shown in Figure 10, in order to replicate the simulation results. As the illumination of the environment is a key characteristic in this experiment, a lux meter was used to assure the replicability of the experiment. The lux meter was placed in a flat surface close to the black box. The values obtained for the testing scenarios were 12, 147, and 207 lumens.

The system has been calibrated in such a way so that the benchmarking module for UWSim can be used to measure the real results in the same way as it is used to measure simulated results. In order to do this, the input data must be configured to be taken from the real world instead of a virtual scene and use the calibration method mentioned above in order to acquire the ground truth.

The 3-D point clouds reconstructed are then evaluated by the benchmarking platform, as can be seen in Figure 11, where real point clouds are displayed on UWSim while being processed. In the images, black dots are filtered as ground points, green points are considered object points, and blue points are labeled as outliers. Similar to the previous simulation, laser reconstruction works better in low-light environments, while stereo reconstruction needs some light to work properly.

A further analysis of the visual results shows that although laser reconstructions look better, there is a misalignment on the point cloud. The 3-D laser reconstructions are slightly rotated with respect to the ground truth target due to small errors in the camera to laser projector calibration. In the case of stereo reconstruction, the ground reconstruction was very poor due to the absence of texture on it.

Results for coverage are shown in Figure 12(a). Laser results are slightly worse than simulated ones. In this case, laser achieves around 32–38%, while in simulation it reached 50%. Although laser works better in dark situations, it is highly resistant to light changes. On the other hand, the stereo reconstruction is completely dependent on light conditions, achieving a 38%, of coverage in good light environments. These results support the ones obtained in simulation where both algorithms behave in a similar way.

An open source benchmarking module is available for UWSim through ROS interfaces.

Figure 10. (a) and (d) The low, (b) and (e) medium, and (c) and (f) high light conditions. (a)–(c) External view. (d)–(f) Camera view.

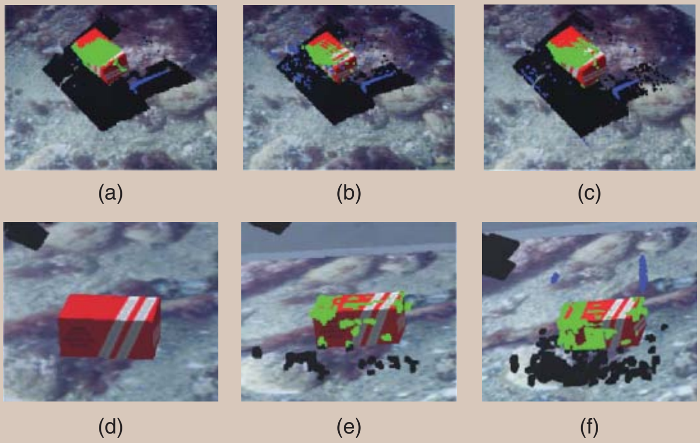
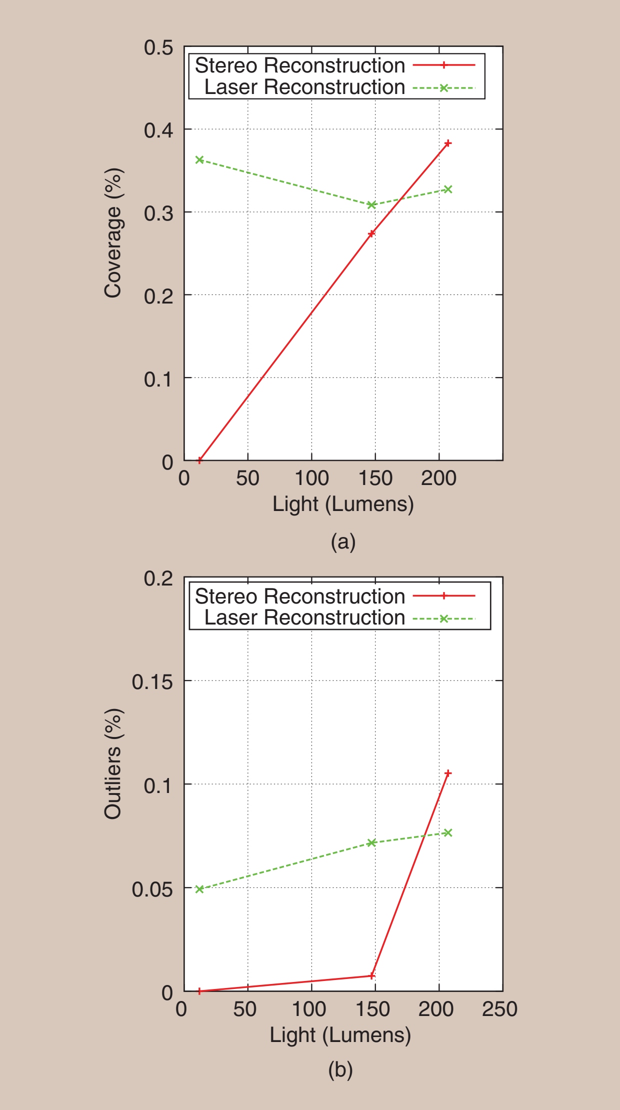


Figure 11. (a) and (d) The low, (b) and (e) medium, and (c) and (f) high light conditions. (a)–(c) Real laser point cloud reconstruction overlayed on the UWSim. (d)–(f) Real stereo point cloud reconstruction overlayed on the UWSim.



(a)

(b)

Figure 12. (a) and The reconstruction coverage and (b) outlier results for a stereo camera and a laser in real scenarios for different light conditions.

The use of an automated comparison system in simulation and controlled environments helps to establish an objective benchmarking methodology.

The laser reconstruction and stereo reconstruction have similar mean errors, around 0.008 m. This result is much greater than in simulation due to the added ground truth estimation error. The standard deviation, though, is greater than the one obtained in simulation, around 0.005 m in stereo and 0.008 m in the case of laser reconstructions, however, it is small enough to conclude that the tested algorithms reached a good alignment and the ground truth estimation was fairly good. This shows the same small misalignment as visual output in the case of laser reconstruction.

The outlier results shown in Figure 12(b) show that both algorithms increase in the number of outliers as light increases. In this case, stereo reconstruction shows a 0% on outliers in the absence of light because is not able to obtain any points. Although in the visual output laser reconstructions seemed to show a higher number of outliers in terms of percent, it is compensated to the higher amount of reconstruction points, meaning they can be filtered easily.

Conclusions

In this article, a benchmarking process is presented to allow easy objective comparison and replication of the results of two 3-D object reconstruction algorithms as a prior step to manipulation in simulated and real scenarios. This process involves a benchmarking platform developed to evaluate software using a simulator as ground truth for the evaluation. The presented results show the potential of the benchmarking techniques to obtain measurable results in simulated scenarios, just as in real situations, helping to decide which approach is better in each situation so that the design of the system can be improved at an early stage. Results replicability is assured as the simulator, benchmarking platform, and the algorithms used to test it are offered as open source. Furthermore, key parameters, such as light conditions, have been measured in order to provide sufficient information for the experiment to be replicable. As a work in progress, an online benchmarking platform is being actively developed to avoid software installation and make algorithm evaluation and comparison faster and more user-friendly.

Acknowledgments

This work was partly supported by Spanish Ministry under grant DPI2014-57746-C3 (MERBOTS Project); by Universitat Jaume I grants PID2010-12, PREDOC/2012/47, and PREDOC/2013/46; and by Generalitat Valenciana grant ACIF/2014/298.

References

[1] (2014, Dec. 3). Survey and inventory of current efforts in comparative robotics research. EURON—European Robotics Search Network. [Online]. Available: http://www.robot.uji.es/EURON/en/index.htm

[2] W. Nowak, A. Zakharov, S. Blumenthal, and E. Prassler. (2010, Apr.). Benchmarks for mobile manipulation and robust obstacle avoidance and navigation in *Deliverable D3.1 from FP7-BRICS Project (Best Practice in Robotics)*. [Online]. Available: http://www.best-of-robotics.org/home

[3] R. Bischoff, T. Guhl, E. Prassler, W. Nowak, G. Kraetzschmar, H. Bruyninckx, P. Soetens, M. Haegele, A. Pott, P. Breedveld, J. Broenink, D. Brugali, and N. Tomatis, "BRICS—Best practice in robotics," in *Proc. IFR Int. Symp. Robotics*, June 2010, pp. 968–975.

[4] Specification of benchmarks. (2009, Jan.). in Deliverable D6.1 from FP7-DEXMART Project (DEXterous and autonomous dual-arm/hand robotic manipulation with sMART sensory-motor skills: A bridge from natural to artificial cognition. [Online]. Available: http://www.dexmart.eu

[5] O. Matsebe, C. Kumile, and N. Tlale, "A review of virtual simulators for autonomous underwater vehicles (AUVs)," in *Proc. IFAC Workshop Navigation, Guidance Control Underwater Vehicles*, Killaloe, Ireland, Apr. 2008, pp. 31–37.

[6] J. Craighead, R. Murphy, J. Burke, and B. Goldiez, "A survey of commercial open source unmanned vehicle simulators," in *Proc. IEEE Int. Conf. Robotics Automation*, Rome, Italy, Apr. 2007, pp. 852–857.

[7] P. J. Sanz, P. Ridao, G. Oliver, G. Casalino, Y. Petillot, C. Silvestre, C. Melchiorri, and A. Turetta, "TRIDENT: An European project targeted to increase the autonomy levels for underwater intervention missions," in *Proc. MTS/IEEE OCEANS'13 Int. Conf.*, San Diego, CA, 2013, pp. 1–10.

[8] M. Prats, J. Pérez, J. Fernández, and P. Sanz, "An open source tool for simulation and supervision of underwater intervention missions," in *Proc. IEEE/RSJ Int. Conf. Intelligent Robots Systems*, Vilamoura Algarve, Portugal, Oct. 2012, pp. 2577–2582.

[9] S. Ulbrich, D. Kappler, T. Asfour, N. Vahrenkamp, A. Bierbaum, M. Przybylski, and R. Dillmann, "The OpenGRASP benchmarking suite: An environment for the comparative analysis of grasping and dexterous manipulation," in *Proc. IEEE/RSJ Int. Conf. Intelligent Robots Systems*, Sept. 2011, pp. 1761–1767.

[10] G. Kootstra, M. Popović, J. Jørgensen, D. Kragic, H. Petersen, and N. Krüger. (2012). VisGraB: A benchmark for vision-based grasping. *Paladyn. 3(2)*, pp. 54–62. [Online]. Available: http://dx.doi.org/10.2478/s13230-012-0020-5

[11] B. Cohen, I. Sucan, and S. Chitta, "A generic infrastructure for benchmarking motion planners," in *Proc. IEEE/RSJ Int. Conf. Intelligent Robots Systems*, Vilamoura Algarve, Portugal, Oct. 2012, pp. 589–595.

[12] M. Quigley, K. Conley, B. P. Gerkey, J. Faust, T. Foote, J. Leibs, R. Wheeler, and A. Y. Ng, "ROS: An open-source robot operating system," in *Proc. ICRA Workshop Open Source Software*, 2009, pp. 1–6.

[13] G. Fontana, M. Matteucci, and D. G. Sorrenti, "Rawseeds: Building a benchmarking toolkit for autonomous robotics," in *Methods and Experimental Techniques in Computer Engineering* (SpringerBriefs in Applied Sciences and Technology), F. Amigoni and V. Schiaffonati, Eds. Germany: Springer International Publishing, 2014.

[14] R. Esteller-Curto, A. del Pobil, E. Cervera, and R. Marin, "A test-bed internet based architecture proposal for benchmarking of visual servoing techniques," in *Proc. 6th Int. Conf. Innovative Mobile Internet Services Ubiquitous Computing*, July 2012, pp. 864–867.

[15] J. Fernández, M. Prats, P. J. Sanz, J. García, R. Marín, M. Robinson, D. Ribas, and P. Ridao, "Grasping for the seabed: Developing a new underwater robot arm for shallow-water intervention," *IEEE Robot. Automat. Mag.*, vol. 20, no. 4, pp. 121–130, 2013.

[16] J. Pérez, J. Sales, M. Prats, J. V. Martí, D. Fornas, R. Marín, and P. J. Sanz, "The underwater simulator UWSim: Benchmarking capabilities on autonomous grasping," in *Proc. 11th Int. Conf. Informatics Control, Automation Robotics (ICINCO)*, 2013, pp. 369–376.

[17] K. Konolige, "Small vision systems: Hardware and implementation," in *Robotics Research*, Y. Shirai and S. Hirose, Eds. London, U.K.: Springer, 1998, pp. 203–212.

[18] A. Peñalver, J. Pérez, J. J. Fernández, J. Sales, P. J. Sanz, J. C. García, D. Fornas, and R. Marín. (2014, Aug.). Autonomous intervention on an underwater panel mockup by using visually-guided manipulation techniques. in *Proc. 19th World Congr. Int. Federation Automatic Control*. Cape Town, Africa. pp. 5151–5156. [Online]. Available: http://dx.doi.org/10.3182/20140824-6-ZA-1003.02545

[19] G. Inglis, C. Smart, I. Vaughn, and C. Roman, "A pipeline for structured light bathymetric mapping," in *Proc. IEEE/RSJ Int. Conf. Intelligent Robots Systems*, Vilamoura Algarve, Portugal, Oct. 2012, pp. 4425–4432.

[20] M. Prats, J. Fernández, and P. Sanz, "Combining template tracking and laser peak detection for 3D reconstruction and grasping in underwater environments," in *Proc. IEEE/RSJ Int. Conf. Intelligent Robots Systems*, Vilamoura Algarve, Portugal, Oct. 2012, pp. 106–112.

[21] S. O. Elberink and G. Vosselman, "Quality analysis on 3D building models reconstructed from airborne laser scanning data," *ISPRS J. Photogrammetry Remote Sensing*, vol. 66, no. 2, pp. 157–165, 2011.

Javier Pérez, Computer Science and Engineering Department, University of Jaume-I, Castellón de la Plana, Spain. E-mail: japerez@uji.es.

Jorge Sales, Computer Science and Engineering Department, University of Jaume-I, Castellón de la Plana, Spain. E-mail: salesj@uji.es.

Antonio Peñalver, Computer Science and Engineering Department, University of Jaume-I, Castellón de la Plana, Spain. E-mail: penalvea@uji.es.

David Fornas, Computer Science and Engineering Department, University of Jaume-I, Castellón de la Plana, Spain. E-mail: dfornas@uji.es.

José Javier Fernández, Computer Science and Engineering Department, University of Jaume-I, Castellón de la Plana, Spain. E-mail: fernandj@uji.es.

Juan Carlos García, Computer Science and Engineering Department, University of Jaume-I, Castellón de la Plana, Spain. E-mail: garciaju@uji.es.

Pedro J. Sanz, Computer Science and Engineering Department, University of Jaume-I, Castellón de la Plana, Spain. E-mail: sanzp@uji.es and sanzp@ieee.org.

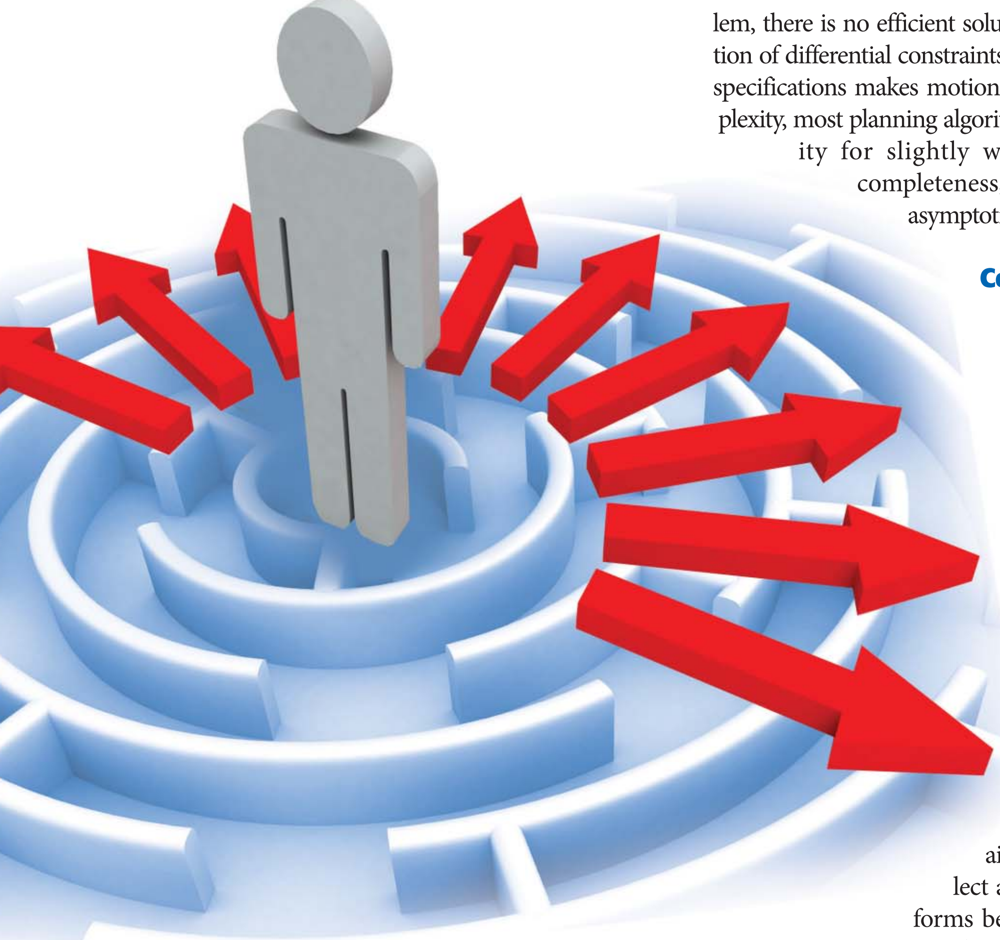
Raúl Marín, Computer Science and Engineering Department, University of Jaume-I, Castellón de la Plana, Spain. E-mail: rmarin@uji.es.

Mario Prats, Computer Science and Engineering Department, University of Jaume-I, Castellón de la Plana, Spain. E-mail: marioprats@gmail.com.

Benchmarking Motion Planning Algorithms

An Extensible Infrastructure for Analysis and Visualization

By Mark Moll, Ioan A. Şucan, and Lydia E. Kavraki



MAZE—©ISTOCKPHOTO.COM/ISERG,
MAN WITH ARROWS—IMAGE LICENSED BY GRAPHIC STOCK

Digital Object Identifier 10.1109/MRA.2015.2448276
Date of publication: 18 August 2015

Motion planning is a key problem in robotics that is concerned with finding a path that satisfies a goal specification subject to constraints. In its simplest form, the solution to this problem consists of finding a path connecting two states, and the only constraint is to avoid collisions. Even for this version of the motion planning problem, there is no efficient solution for the general case [1]. The addition of differential constraints on robot motion or more general goal specifications makes motion planning even harder. Given its complexity, most planning algorithms forego completeness and optimality for slightly weaker notions such as resolution completeness, probabilistic completeness [2], and asymptotic optimality.

Comparing Planning Algorithms

Sampling-based planning algorithms are the most common probabilistically complete algorithms and are widely used on robot platforms with many degrees of freedom. Within this class of algorithms, many variants have been proposed over the last 20 years, yet there is still no characterization of which algorithms are well-suited for which classes of problems. We present a benchmarking infrastructure for motion planning algorithms that can be a useful component for such a characterization. The infrastructure is aimed both at end users who want to select a motion planning algorithm that performs best on problems of interest, as well as motion planning researchers who want to compare the performance of a new algorithm relative to other state-of-the-art algorithms.

The benchmarking infrastructure consists of three main components (see Figure 1). First, we have created an extensive benchmarking software framework that is included in the Open Motion

1070-9932/15©2015IEEE

Planning Library (OMPL, http://ompl.kavrakilab.org), a C++ library that contains implementations of many sampling-based algorithms [3]. One can immediately compare any new planning algorithm to the more than 30 other planning algorithms that currently exist within OMPL. There is also flexibility in the types of motion planning problems that can be benchmarked, as discussed in the "Defining Motion Planning Problems" section. Second, we have defined extensible formats for storing benchmark results. The formats are fairly straightforward so that other planning libraries could easily produce compatible output. Finally, we have created an interactive, versatile visualization tool for compact presentation of collected benchmark data (see http://plannerarena.org). The tool and underlying database facilitate the analysis of performance across benchmark problems and planners. While the three components described above emphasize generality, we have also created a simple command line tool, specifically for rigid body motion planning that takes as input a plain text description of a motion planning problem.

Benchmarking sampling-based planners is nontrivial for several reasons. Since these planners rely on sampling, performance cannot be judged from a single run. Instead, benchmarks need to be run repeatedly to obtain a distribution of some performance metric of interest. Simply comparing the means of such distributions may not always be the correct way to assess the performance. Second, it is well known that different sampling strategies employed by sampling-based algorithms typically perform well only for certain classes of problems, but it is difficult to exactly define such classes. Finally, different applications require optimization for different metrics (e.g., path quality versus time of computation) and there is no universal metric to assess performance of planning algorithms across all benchmarks.

There have been some attempts in the past to come up with a general infrastructure for comparing different planning algorithms (see [4], [5]). This article is in the same spirit but includes an extended and extensible set of metrics and offers higher levels of abstraction and concrete entry-level points for end users. Furthermore, we also introduce an extensible logging format that other software can use and a visualization tool. To the best of our knowledge, none of the prior work offered the ability to interactively explore and visualize benchmark results. The Motion Planning Kernel (MPK) software system described in [4] is similar to OMPL in that both aim to provide a generic, extensible motion planning library, but MPK appears to no longer be maintained or developed. There has been significant work on metrics used for comparing different planning algorithms (see [6], [7]), and our benchmarking infrastructure includes many of these metrics.

The contribution of this article is not to any particular benchmark problem, metric, or planner but provides a generic, extensible benchmarking infrastructure that facilitates easy analysis and visualization of replicable benchmark results. Since it is integrated with the widely used and actively developed OMPL, it becomes straightforward to compare any new motion planning algorithm to many other state-of-the-art motion planning algorithms. All relevant information pertaining to how benchmarks were run is stored in a database to enable replicability of results.

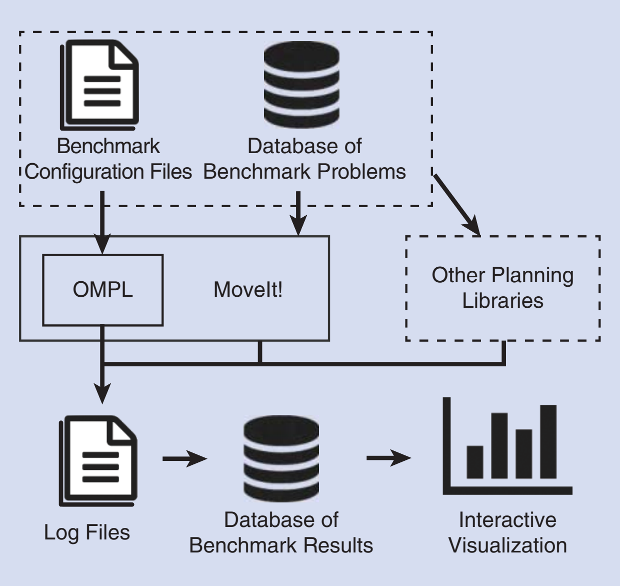


Figure 1. An overview of the benchmarking infrastructure.

Benchmarking Infrastructure

OMPL provides a high level of abstraction for defining motion planning problems. The planning algorithms in OMPL are, to a large extent, agnostic with respect to the space they are planning in. Similarly, the benchmarking infrastructure within OMPL allows the user to collect various statistics for different types of motion planning problems. The basic workflow is as follows.

There is still no characterization of which algorithms are well-suited for which classes of problems.

- The user defines a motion planning problem. This involves defining the state space of the robot, a function that determines which states are valid (e.g., collision-free), the starting state of the robot, and the goal. The complete definition of a motion planning problem is contained within a C++ object, which is used to construct a benchmark object.
- The user specifies which planning algorithms should be used to solve the problem, time, and memory limits for each run and the number of runs for each planner.
- The benchmark is run. Upon completion, the collected results are saved to a log file. A script is used to add the results in the log file to a SQL database. The results can be queried directly in the database or explored and visualized interactively through a website set up for this purpose (http://plannerarena.org).

Defining Motion Planning Problems

The most common benchmark motion planning problems are those where robots are modeled as rigid bodies due to their simplicity (it is easy for users to intuitively assess performance). We have developed a simple plain-text file format that describes such problems with a number of key-value pairs. Robots and environments are specified by mesh files. The state validity function is, in this case, hard-coded to be a collision checker. Besides the start and goal positions of the robot, the user can also specify an optimization objective, such as path length, minimum clearance along the path, or mechanical work. There are several planning algorithms in OMPL that optimize a path with respect to a specified objective. (Others that do not support optimization simply ignore this objective.) It is also possible to specify simple kinodynamic motion planning problems. OMPL.app, the application layer on top of the core OMPL library, predefines the following systems that can be used: 1) a first-order car, 2) a second-order car, 3) a blimp, and 4) a quadrotor. We have not developed controllers or steering functions for these systems. Kinodynamic planners in OMPL fall back in such cases on sampling random controls. This makes planning for these systems extremely challenging; however, if controllers are available, then OMPL can use them. With a few lines of code, the command line tool can be modified to allow new planning algorithms or new types of planning problems to be specified in the configuration files.

There is no universal metric to assess performance of planning algorithms across all benchmarks.

The benchmark configuration files can be created with the graphical user interface (GUI) included with OMPL.app. A user can load meshes in a large variety of formats, define start and goal states, try to solve the problem with different planners, and save the configuration file. The user can also visualize the tree/graph produced by a planning algorithm to get a sense of the difficulty of a particular problem. In the configuration file, the user can specify whether the solution paths (all or just the best one) should be saved during benchmarking. Saved paths can be played back with the GUI.

When defining motion planning problems in code, many of the limitations of the command line tool go away. Arbitrary state spaces and kinodynamic systems can be used and different notions of state validity and different optimization objectives can be defined. In addition, any user-defined planning algorithm can be used. The OMPL application programmer interface (API) imposes only minimal requirements on new planning algorithms. In particular, the API is not limited to sampling-based algorithms (in [8], for example, several non-sampling-based planners are integrated into OMPL). The low barrier to entry has lead to numerous contributions of planning algorithms from other groups: OMPL 1.0 includes 29 planning algorithms. Since all these algorithms use the same low-level functionality for, e.g., collision checking, benchmarking highlights the differences in the motion planning algorithms themselves.

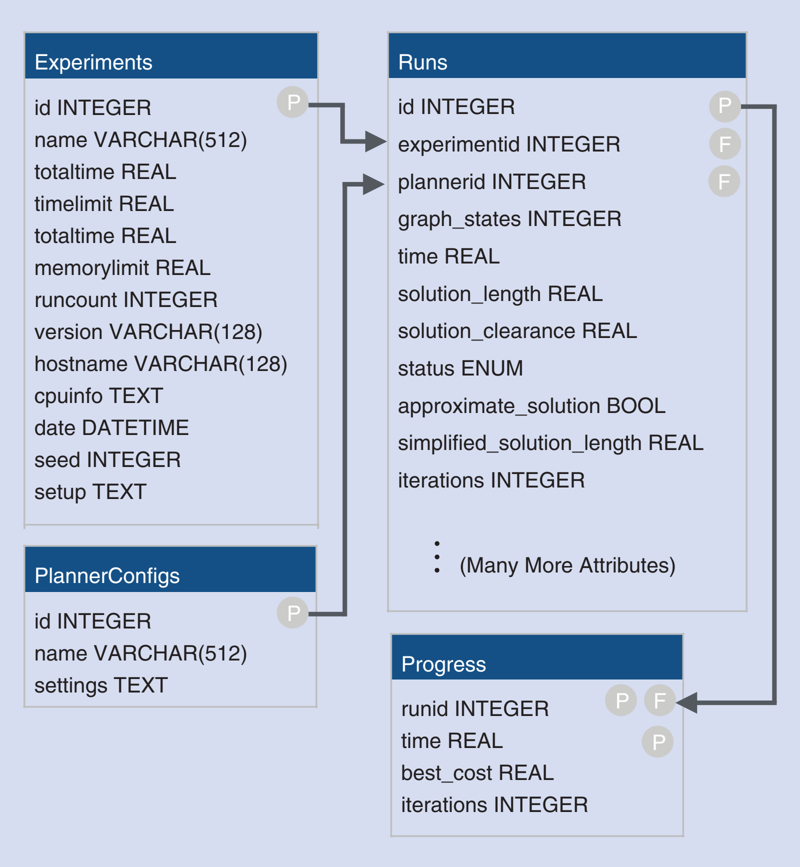


Figure 2. The schema for a database of benchmark results. The Ⓟ and Ⓕ denote the primary and foreign keys of each table, respectively.

The benchmarking facilities in MoveIt! [9] are based on and compatible with those in OMPL. The problem setup is somewhat similar to the OMPL command line tool. In MoveIt!, robots are specified by Unified Robot Description Format (URDF) files, which specify a robot's geometry and kinematics. Motion planning problems to be benchmarked are stored in a database.

Specifying Planning Algorithms

Once a motion planning problem has been specified, the next step is to select one or more planners that are appropriate for the given problem. Within OMPL, planners are divided into two categories: 1) geometric/kinematic planners and 2) kinodynamic planners. The first category can be further divided into two subcategories: 1) planners that terminate when any solution is found and 2) planners that attempt to compute an optimized solution (with respect to a user-specified optimization objective). For optimizing planners, a threshold on optimality can be set to control how close to optimal the solution needs to be. At one extreme, when this threshold is set to zero, planners will run until time runs out. At the other extreme, when the threshold is set to infinity, planners act like nonoptimizing planners and will terminate as soon as any solution is found.

Typically, a user specifies multiple planners. By default, OMPL will try to make reasonable parameter choices for each planner. However, a user can also fine-tune any parameter setting for a planner. With the command line tool's configuration files, this is easily accomplished by adding lines of the form `planner.parameter=value`. The parameter code infrastructure is generic, and when a programmer specifies a parameter for a planner it can be specified through the configuration file without having to change the parsing of configuration files. It is also possible to add many instances of the same type of planner. This is useful for parameter sweeps. Each instance can be given a slightly different name to help distinguish the results for each instance. Each run of a planner is executed in a separate thread, therefore, if a planner hangs, the benchmark program can detect that and forcibly terminate the planner thread (the run is recorded as a crash and the benchmarking will continue with the next run).

Database of Benchmark Runs

After a benchmark run is completed, a log file is written out. With the help of a script, the benchmark results stored in the log file can be added to a SQLite3 database. Multiple benchmark log files can be added to the same database. The SQLite3 database facilitates distribution of all relevant benchmark data and users can simply transfer one single file. Furthermore, the database can be easily programmatically queried with almost any programming language. In contrast, extracting information directly from the log files or some other custom storage format would require more effort to perform the types of analysis and visualization that is enabled by our database schema described below.

Figure 2 shows the database schema that is used. Each benchmark log file corresponds to one experiment. The experiments table contains an entry for each experiment that contains the basic benchmark parameters and the detailed information about the hardware on which the experiment was performed (in the cpuinfo column). Information about each of the planner instances that were specified is stored in the PlannerConfigs table. For each planner instance, all parameter values are stored as a string representation of a list of key-value pairs (in the settings column). While we could have created a separate column in the PlannerConfigs table for each parameter, the parameters are planner specific with very few shared parameters among planners.

The main results are stored in the Runs table. Each entry in this table corresponds to one run of a particular planner trying to solve a particular motion planning problem. After a run is completed, several attributes are collected such as the number of generated states (`graph_states`), duration of the run (`time`), length of the solution path (`solution_length`), clearance along the solution path (`solution_clearance`), and so on. Default solutions are simplified (through a combination of shortcutting and smoothing [10]), which usually improves the solution quality at a minimal time cost. Runs can terminate for a variety of reasons, such as a solution was found, the planner timed out (without any solution or with an approximate solution), or the planner crashed. We use an enumerate type for this attribute (stored in status), and the labels for each value are stored in the enums table (not shown in Figure 2).

The progress table stores information periodically collected during a run. This collection is done in a separate thread so as to minimize the effect on the run itself. Progress information is currently only available for optimizing planners. It is used to store the cost of the solution found at a particular time. By aggregating progress information from many runs for each planner, we can compare rates of convergence to optimality (see "Interactive Analysis of Results" section).

The infrastructure is aimed both at end users who want to select a motion planning algorithm that performs best on problems of interest, as well as motion planning researchers who want to compare the performance of a new algorithm relative to other state-of-the-art algorithms.

The database schema has been designed with extensibility in mind. Large parts of the schema are optional and other columns can be easily added. This does not require new parsers or additional code. Instead, the log files contain enough structure to allow planners to define their own run and progress properties. Thus, when new log files are added to a database, new columns are automatically added to runs and progress. Planners that do not report on certain properties will just store the value "N/A" in the corresponding columns. Additional run properties for a new type of planner are easily defined by storing key-value pairs in a dictionary of planner data, which is obtained after each run. Additional progress properties are defined by adding a function to a list of callback functions.

Log files have a fairly straightforward plain text format that is easy to generate and parse. (The complete syntax is specified at http://ompl.kavrakilab.org/benchmark.html.) This makes it easy for other motion planning libraries to generate compatible log files that can be added to the same type of benchmark database. For example, MoveIt!'s benchmarking capabilities do not directly build on OMPL's benchmark capabilities, yet it can produce compatible benchmark log files. This makes it possible to see how a planning algorithm's performance changes when moving from abstract benchmark problems in OMPL to elaborate real-world settings created with MoveIt! (possibly from experimental data).

The Planner Arena website makes it easy to interactively explore benchmark results.

Interactive Analysis of Results

There are many different ways to visualize benchmark performance. It is nearly impossible to create a tool that can automatically select the right visualizations for a given benchmark database. We have created a website called *Planner Arena* (http://plannerarena.org), where benchmark data can be uploaded and selected results can be visualized. The website interface is dynamically constructed based on the content of the benchmark database. Selection widgets are created automatically for the benchmark problems, the performance attributes, the planning algorithms, and so on. The code that powers Planner Arena is included in the OMPL distribution and can be run locally to evaluate one's own results privately or be modified to create custom visualizations. There are currently three types of plots included on the

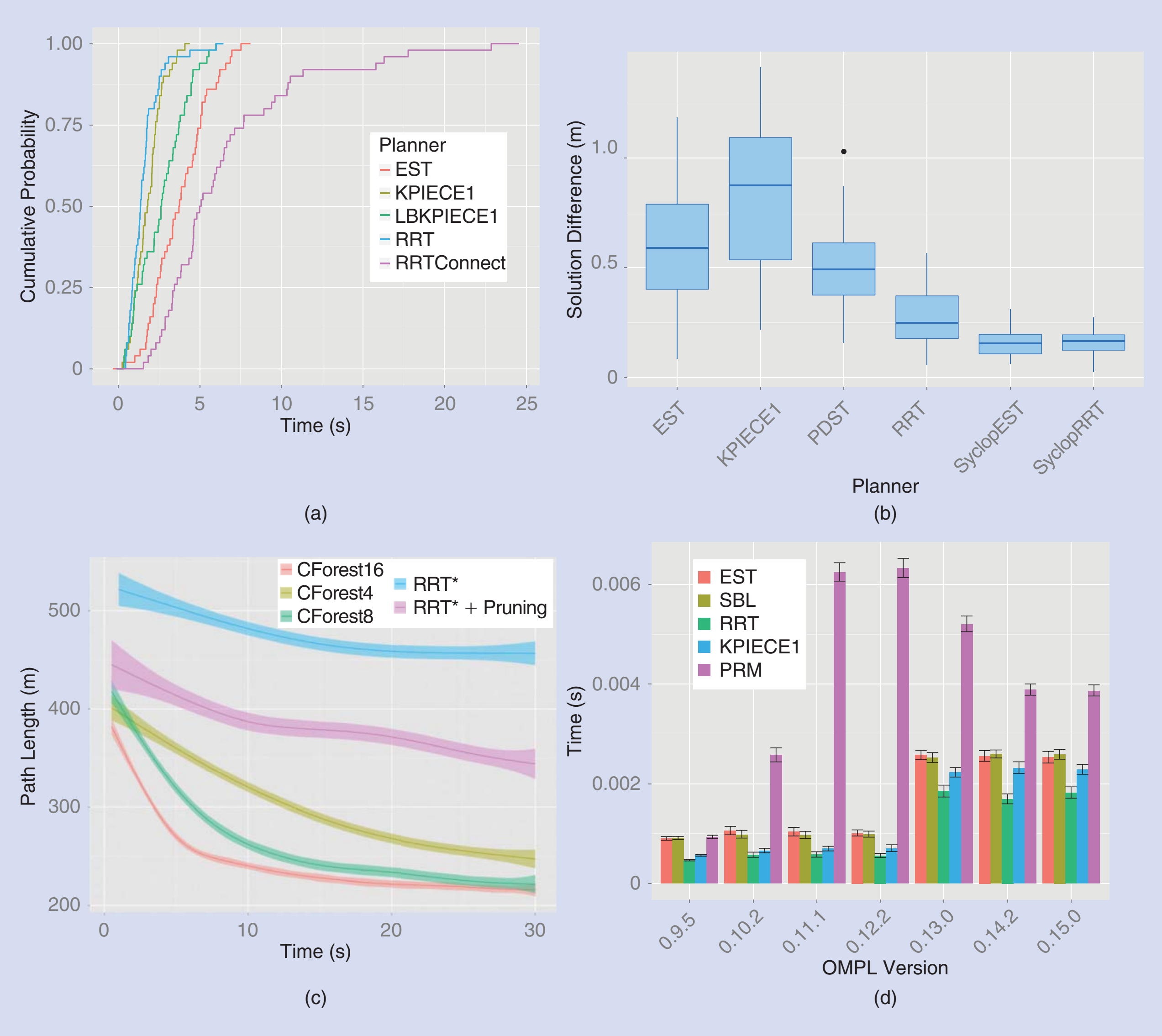


Figure 3. The sample output produced from a benchmark database by the Planner Arena server for various motion planning problems (but not the ones shown in Figure 4). (a) The performance plot of an empirical cumulative distribution function of solution times for a rigid body benchmark. (b) The performance plot of the distance between best found approximate solution and goal for a kinodynamic problem. (c) The progress plot of the convergence rate of asymptotically optimal planners. (d) The regression plot of test results for a trivial benchmark.

Planner Arena site: 1) overall performance plots, 2) progress plots, and 3) regression plots. We will describe these plots in more detail below.

Plots of Overall Performance

The overall performance plots can show how different planners compare on various measures. The most common performance measure is the time needed for a planner to find a feasible solution. By default, integer-and real-valued performance metrics (such as solution time) are plotted as box plots that provide useful summary statistics for each planner: median, confidence intervals, and outliers. However, in some cases visualizing the cumulative distribution function can reveal additional useful information. For instance, from Figure 3(a) one can easily read off the probability that a given planner can solve a particular benchmark within a specified amount of time. For very hard problems, where most planners time out without finding a solution, it might be informative to look at solution difference: the gap between the best found solution and the goal [Figure 3(b)]. For optimizing planners, it is often more interesting to look at the best solution found within some time limit. The overall performance page allows one to select a motion planning problem that was benchmarked, a particular benchmark attribute to plot, the OMPL version (in case the database contains data for multiple versions), and the planners to compare.

Most of the measures are plotted as box plots. Missing data are ignored. It is important to keep in mind that if a planner failed to solve a problem 99 times out of a 100 runs, then the average solution length is determined by one run. To make missing data more apparent, a table below the plot shows how many data points there were for each planner and how many of those were missing values.

Performance is often hard to judge by one metric alone. Depending on the application, a combination of metrics is often necessary to be able to choose an appropriate planner. For example, in our experience LBKPIECE [11] (one of the planning algorithms in OMPL) tends to be among the fastest planners, but it also tends to produce longer paths. For time-critical applications this may be acceptable, but for applications that place greater importance on short paths another planner might be more appropriate. There will also be exceptions to general trends. Bidirectional planners (such as RRT-Connect [12]) tend to be faster than unidirectional planners (such as RRT [12]), but Figure 3(a) shows that this is not always the case. This underscores the need for a good set of benchmark problems that are representative of different applications.

Progress Plots

Some planners in OMPL are not limited to reporting information after a run is completed, but can also periodically report information during a run. In particular, for asymptotically optimal planners it is interesting to look at the convergence rate of the best path cost (e.g., path length). By default, Planner Arena will plot the smoothed mean as well as a 95% confidence interval for the mean [Figure 3(c)]. Optionally, individual measurements can be shown as semitransparent dots, which can be useful to get a better idea of the overall distribution. Analogous to the performance plots, missing data are ignored. During the first couple seconds of a run, a planner may never find a solution path. Below the progress plot, we plot the number of data points available for a particular planner at each 1-s time interval.

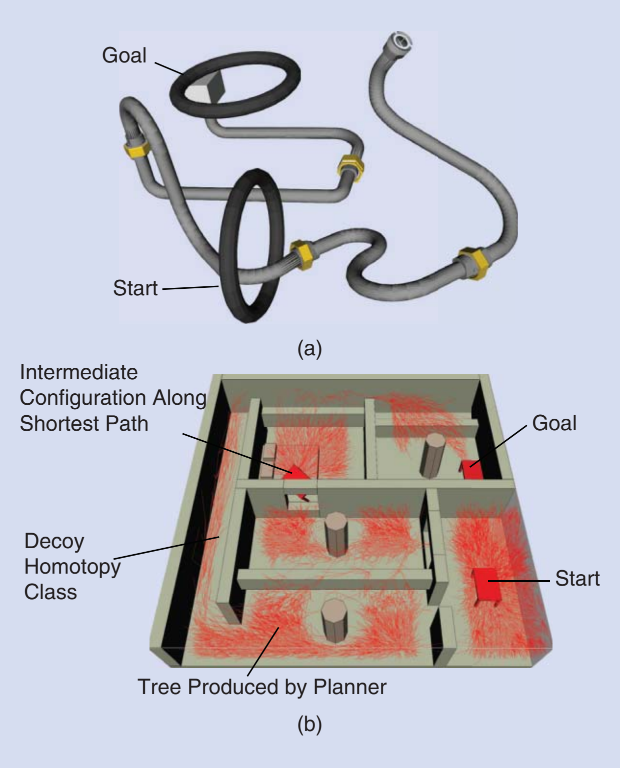


Figure 4. Two of the sample benchmark problems included on Planner Arena: (a) with a long, twisty narrow passage and (b) with several suboptimal decoy homotopy classes of paths.

For community-wide adoption of benchmarks it is important to adopt standard input and output file formats.

Regression Plots

Regression plots show how the performance of the same planners change over different versions of OMPL [Figure 3(d)]. This is mostly a tool for developers using OMPL that can help in the identification of changes with unintended side effects on performance. However, it also allows a user to easily compare the performance of a user's modifications to the planners in OMPL with the latest official release. In regression plots, the results are shown as a bar plot with error bars.

Any of the plots can be downloaded as a PDF file or as RData. The PDF format is useful if the plot is just needs touch ups. The RData file contains both the plot as well as all the data shown in the plot and can be loaded into *R*. The plot can be completely customized, further analysis can be applied to the data, or the data can be plotted in an entirely different way.

The default benchmark database stored on the server currently contains results for nine different benchmark problems. They include simple rigid body type problems and hard problems specifically designed for optimizing planners (problems that contain several suboptimal decoy homotopy classes), kinodynamic problems, and a multirobot problem (see Figure 4).

The benchmarking facilities in MoveIt! are compatible with those in OMPL.

Discussion

We expect that with input from leaders in the motion planning community as well as with extensive simulations and experiments, we can create a suite of motion planning benchmarks. We plan to develop benchmarks along two different directions. First, there are toy problems that isolate one of a number of common difficulties that could trip up a motion planning algorithm (such as a very narrow passage or the existence of many false leads). Such benchmarks may provide some insights that lead to algorithmic improvements. Second, we would like to develop a benchmark suite where performance (by some measure) is predictive of performance of more complex real-world scenarios.

Other planning libraries can use the same set of benchmark problems. While OMPL could be extended with other planning algorithms, we recognize that for community-wide adoption of benchmarks it is important to adopt standard input and output file formats. The log file format and database schema for storing benchmark results described in this article are general enough that they can be adapted by other motion planning software. This would allow for a direct comparison of different implementations of planning algorithms.

The Planner Arena website makes it easy to interactively explore benchmark results. At this point, we do not claim that the benchmarks included in the default database on Planner Arena form some sort of standard benchmark set, although they are representative of the types of problems that have been used in prior work [13]. Furthermore, the set of problems we present for results will increase over time.

Acknowledgments

Mark Moll and Lydia E. Kavraki are supported in part by National Science Foundation grant 1317849 as part of the National Robotics Initiative program 1317849. The authors would like to thank Luis Torres for his contributions to the benchmarking capabilities in OMPL.

References

[1] J. Canny, *The Complexity of Robot Motion Planning*. Cambridge, MA: MIT Press, 1988.

[2] H. Choset, K. M. Lynch, S. Hutchinson, G. Kantor, W. Burgard, L. E. Kavraki, and S. Thrun, *Principles of Robot Motion: Theory, Algorithms, and Implementations*. Cambridge, MA: MIT Press, 2005.

[3] I. A. Şucan, M. Moll, and L. E. Kavraki. (2012, Dec.). The open motion planning library. *IEEE Robot. Automat. Mag.* [Online]. *19(4)*, pp. 72–82. Available: http://ompl.kavrakilab.org

[4] I. Gipson, K. Gupta, and M. Greenspan, “MPK: An open extensible motion planning kernel,” *J. Robot. Syst.*, vol. 18, no. 8, pp. 433–443, 2001.

[5] B. Cohen, I. A. Şucan, and S. Chitta, “A generic infrastructure for benchmarking motion planners,” in *Proc. IEEE/RSJ Int. Conf. Intelligent Robots Systems*, Oct. 2012, pp. 589–595.

[6] R. J. Geraerts, “Sampling-based motion planning: Analysis and path quality,” Ph.D. dissertation, Dept. Comput. Sci., Utrecht Univ., Utrecht, The Netherlands, May 2006.

[7] M. A. M. Aguirre, “Metrics for sampling-based motion planning,” Ph.D. dissertation, Dept. Comput. Sci., Texas A&M Univ., College Station, TX, Dec. 2007.

[8] J. Luo and K. Hauser, “An empirical study of optimal motion planning,” in *Proc. IEEE/RSJ Int. Conf. Intelligent Robots Systems*, 2014, pp. 1761–1768.

[9] I. A. Şucan and S. Chitta. (2015, July 28). MoveIt! [Online]. Available: http://moveit.ros.org

[10] R. Geraerts and M. Overmars, “Creating high-quality paths for motion planning,” *Int. J. Robot. Res.*, vol. 26, no. 8, pp. 845–863, 2007.

[11] I. A. Şucan and L. E. Kavraki, “A sampling-based tree planner for systems with complex dynamics,” *IEEE Trans. Robot.*, vol. 28, no. 1, pp. 116–131, 2012.

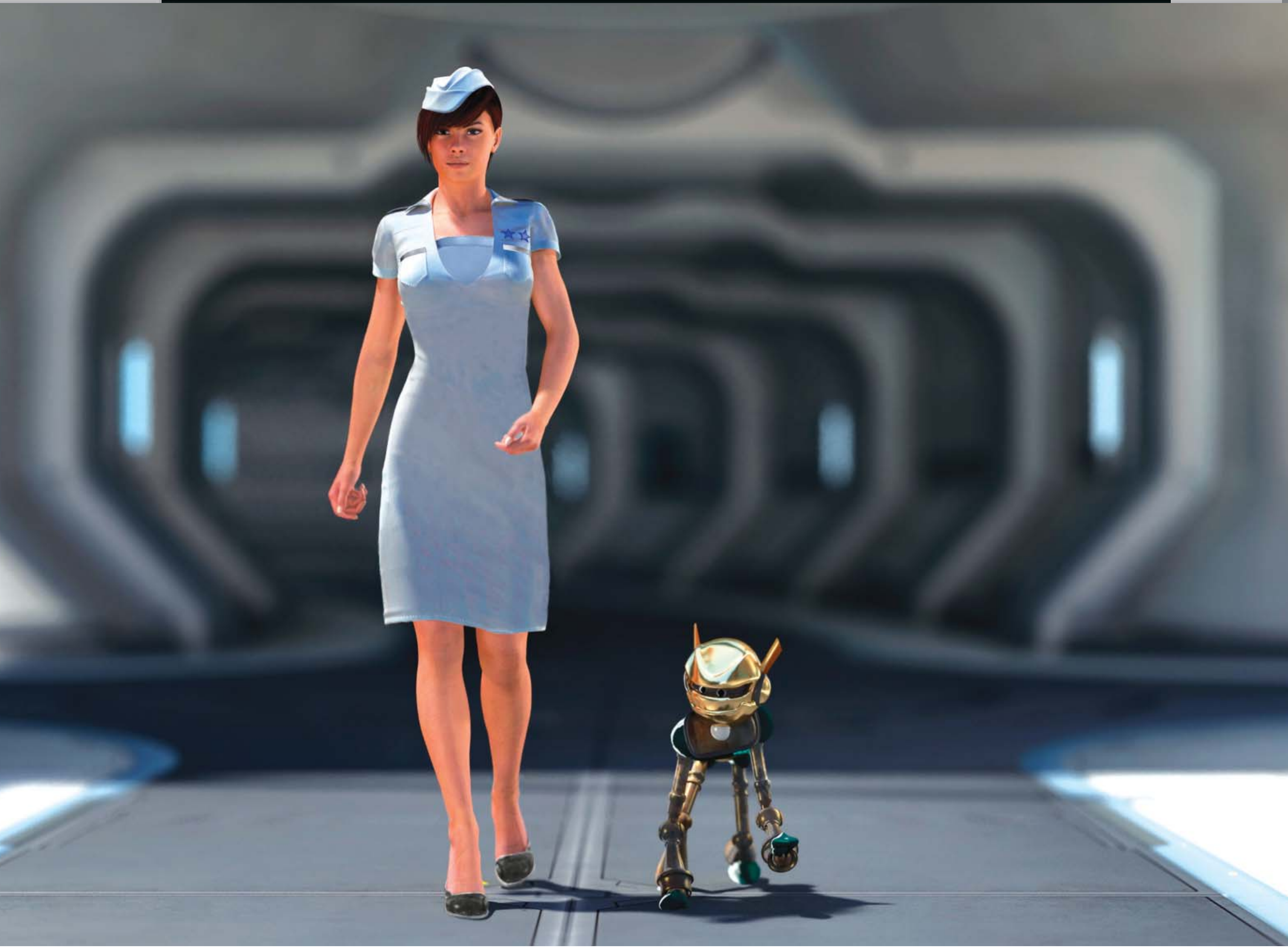
[12] J. Kuffner and S. M. LaValle, “RRT-Connect: An efficient approach to single-query path planning,” in *Proc. IEEE Int. Conf. Robotics Automation*, San Francisco, CA, Apr. 2000, pp. 995–1001.

[13] (2015, July 28). Algorithms & applications group motion planning puzzles. [Online]. Available: https://parasol.tamu.edu/dsmft/benchmarks/

Mark Moll, Department of Computer Science, Rice University, Houston, Texas 77251. E-mail: mmoll@rice.edu.

Ioan A. Şucan, Google, Mountain View, California 94043. E-mail: isu-can@gmail.com.

Lydia E. Kavraki, Department of Computer Science, Rice University, Houston, Texas 77005. E-mail: kavraki@rice.edu.



©ISTOCKPHOTO.COM/MIKHAIL TOLSTOY

Benchmarking Bipedal Locomotion

A Unified Scheme for Humanoids, Wearable Robots, and Humans

By Diego Torricelli, Jose González-Vargas, Jan F. Veneman, Katja Mombaur, Nikos Tsagarakis, Antonio J. del-Ama, Angel Gil-Agudo, Juan C. Moreno, and Jose L. Pons

Digital Object Identifier 10.1109/MRA.2015.2448278
Date of publication: 11 September 2015

In the field of robotics, there is a growing awareness of the importance of benchmarking [1], [2]. Benchmarking not only allows the assessment and comparison of the performance of different technologies but also defines and supports the standardization and regulation processes during their introduction to the market. Its importance has been recently emphasized by the adoption of the technology readiness levels (TRLs) in the Horizon 2020 information and communication technologies by the European Union as an important guideline to assess when a technology can shift from one TRL to the other. The objective of this article is to define the basis of a benchmarking scheme for the

1070-9932/15©2015IEEE

assessment of bipedal locomotion that could be applied and shared across different research communities.

Benchmarking

In the field of humanoid robots, the main obstacle in identifying common benchmarks is that different methods and metrics are typically employed for specific robotic systems and functional scenarios. Benchmarking of humanoid locomotion is typically approached on a competition-based level and is mostly focused on global functional goals (e.g., playing soccer, avoiding obstacles, and climbing stairs [1], [3]). In the field of wearable robots, performance is usually reported in terms of the effects on the user's motor function. New standards are highly anticipated, especially now that these products are appearing on the market. Yet, there are no accepted schemes for comparing the performance of wearable robots on a vast scale. The only initiative in this direction is represented by the upcoming Cybathlon competition [4]. In the clinical and biomechanics field, many metrics and clinical scales have been defined and are regularly used to assess the locomotion functions [5]. Most of these scales are based on observation by skilled personnel or defined on a very general level, measuring variables like average speed or timed up-and-go. With the increasing application of sensorized and robotic technology in clinics, the expectation for new quantitative and reliable metrics is rapidly growing.

The proposed scheme can be taken as a starting point for a global iterative process that could lead to an international consensus.

In this article, as mentioned previously, our goal is to outline a benchmarking scheme for bipedal locomotion. Our approach does not aim to compare systems on a global level to see which one is better but to assess the several aspects of multifacetted performance, allowing a truthful comparison of each feature independently. We envision a scenario in which using this scheme will encourage collaboration between different research groups toward the consolidation of standardized benchmarks and experimental procedures, and we promote its use as a complementary tool to the competition-based approaches. The scheme presented in this article is the result of the joint efforts of five European projects, i.e., H2R (www.h2rproject.eu), BALANCE (www.balance-fp7.eu), KoroiBot (www.koroibot.eu), WALK-MAN (www.walk-man.eu), and BioMot (www.biomotproject.eu). We think that the proposed scheme can be taken as a starting point for a global iterative process that could lead to an international consensus, based on its practical use across different laboratories.

Design Approach

Analysis of the Needs: The Web-Based Survey

A benchmark can be considered successful if and only if it is widely accepted by the community at which it is targeted. To reach this goal, a number of key principles for a successful benchmarking scheme have been identified [6].

- The benchmarks must be well defined, i.e., they really must serve their purpose. As a consequence, the purpose should be clear.
- Benchmarks should be rigorously focused on limited, particular subdomains.
- It is more likely that a benchmark is successful within a scientific (sub)community if it arises from that community itself.

Our design process started with a web-based survey, to identify the needs of the different users to which the scheme is addressed. The research communities considered were humanoid robotics, wearable robotics, and human biomechanics. The latter has been included because of the increasing need to merge insights from biomechanics and human motor control in robotic research. The survey (see Figure 1) comprised nine questions, which, overall, address the first two aforementioned design principles. The first three questions aimed to collect general information about the respondents, such as their background and their overall interest in using a benchmarking scheme and in sharing the data obtained by its use. The last six questions focused on the contents of the ideal benchmarking scheme, in terms of

- its general purpose
- the motor functions addressed
- the performance variables to be measured
- the conditions to be included
- its technical properties
- information needed to contextualize the results.

In questions 4–9, the user was asked to give a score from one to five for each of the predefined options. The results are represented in Figure 1 in terms of the mean values and standard deviations and divided based on the respondents' backgrounds. A statistical analysis of the similarity across communities has been performed by one-way analysis of variance (level of significance $p = 0.05$). Figure 1 also provides a rough classification of results in three classes, according to mean scores across all communities: items of high relevance (mean score over four, highlighted in green), items of medium relevance (mean score over three, shown in orange), and items of low relevance (mean score lower than three, shown in red). An asterisk indicates the items that presented a significant difference in the responses among the communities (excluding the "other" group).

Existing Taxonomies for Skills and Abilities

Defining a common nomenclature is a basic purpose for a successful taxonomy. This is particularly true in our case because the target is multidisciplinary and different terms, like *skill, function, ability, task, activity, action,* and *performance*, can have different meanings. Inspired by the approach of Magill [7], we will make use of three terms: skill, ability, and performance (see Figure 2). We define *skill* as a task or activity with a specified goal. For instance, walking is a motor skill whose goal is to move from point A to point B. *Ability* can be

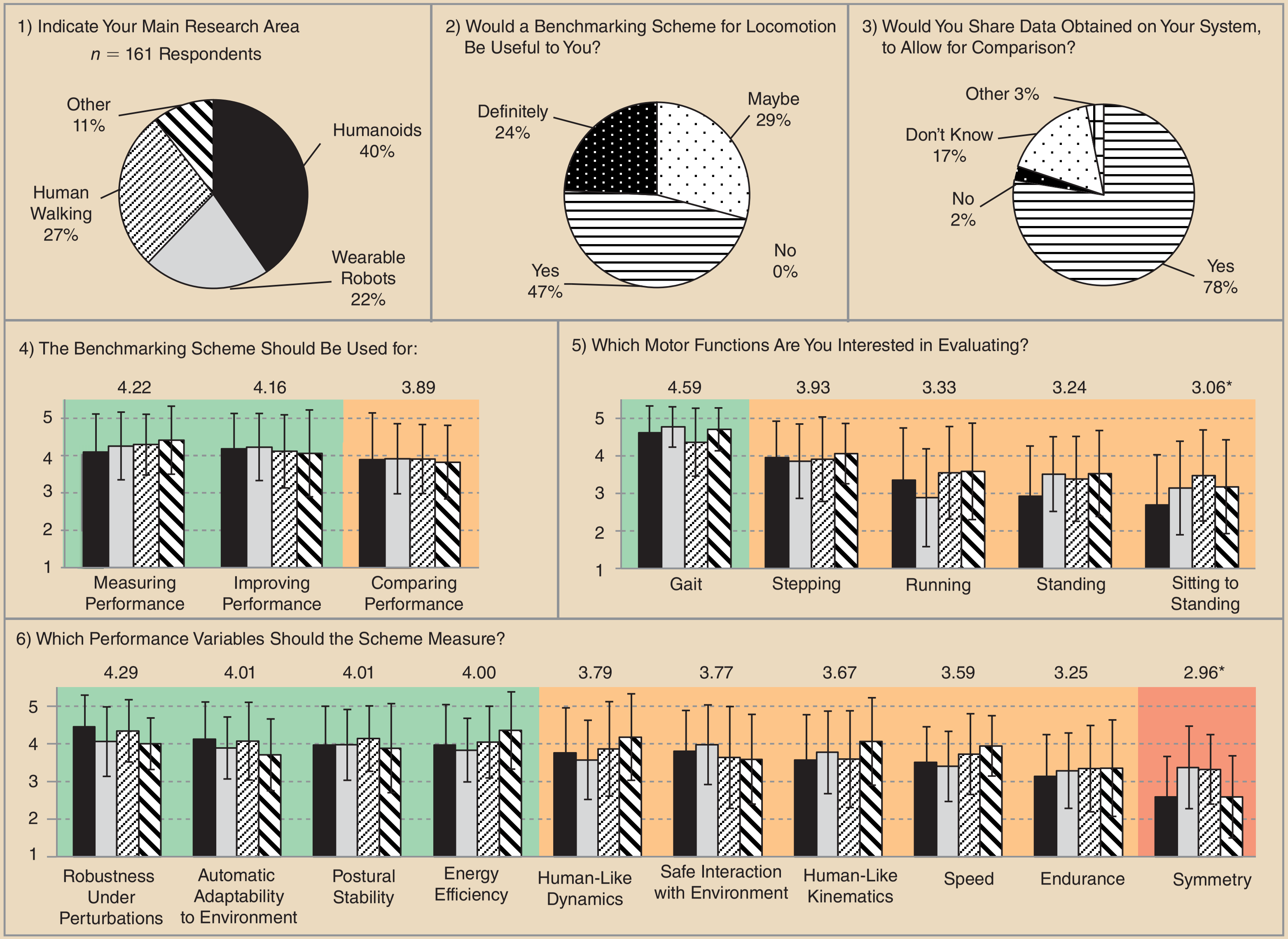
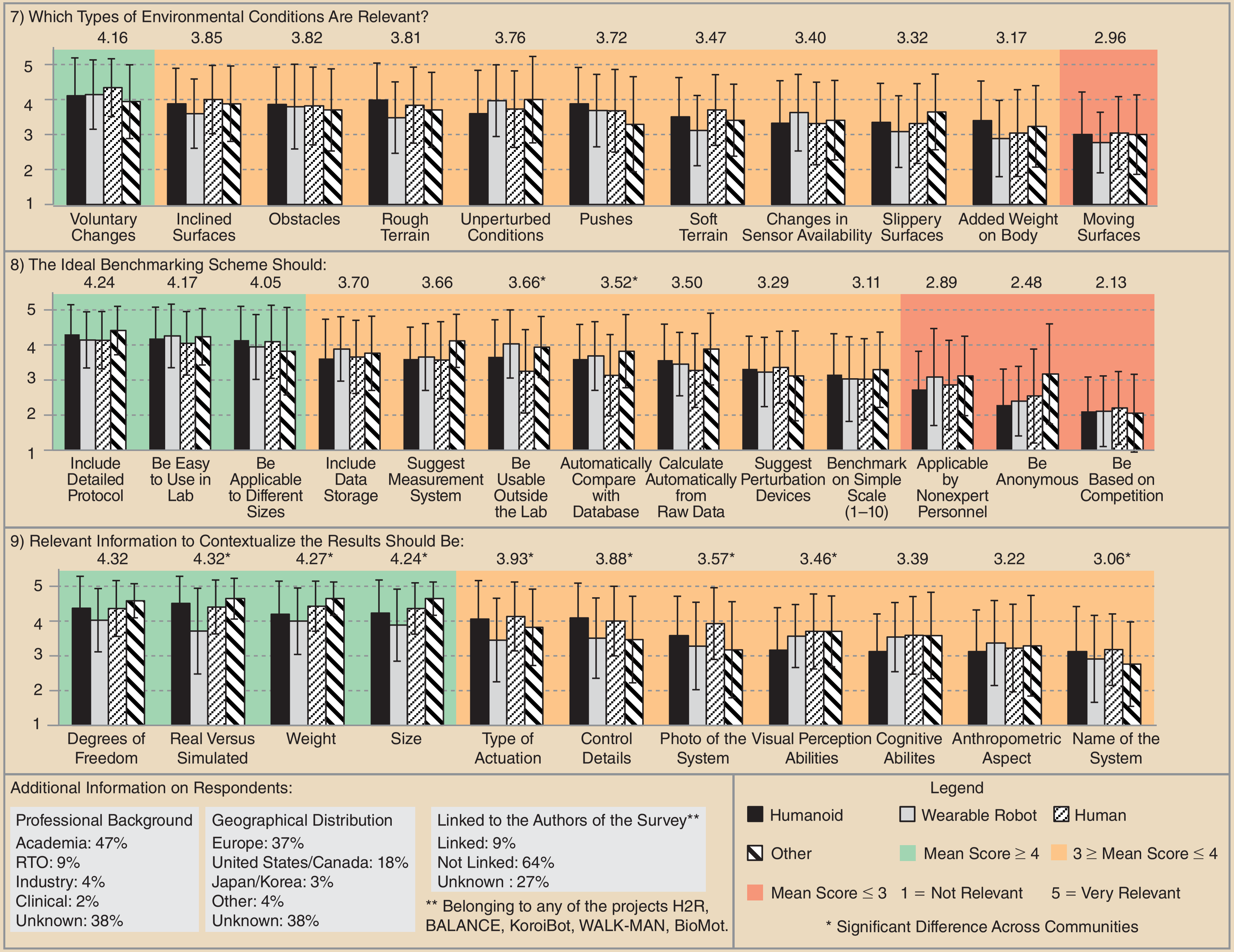


Figure 1. The results of the web-based survey: the 161 respondents to the questionnaire are international experts that were contacted through the following workshops, networks, or forums: the aforementioned involved research projects, i.e., H2R, BALANCE, KoroiBot, WALK-MAN, and BioMot; the euRobotics or Euron mailing list, Biomch-L mailing list, and robotics_worldwide mailing list; WeRob2014—The 2014 International Workshop on Wearable Robotics (werob2014.org); the European Network on Robotics for NeuroRehabilitation (European Commission COST Action TD1006, www.rehabilitationrobotics.eu); Dynamic Walking (dynamicwalking.org/); and the RehabRobotics mailing list (associated with ICORR, http://www.rehabrobotics.org/). (*Continued*)



Additional Information on Respondents:

Professional Background
Academia: 47%
RTO: 9%
Industry: 4%
Clinical: 2%
Unknown: 38%

Geographical Distribution
Europe: 37%
United States/Canada: 18%
Japan/Korea: 3%
Other: 4%
Unknown: 38%

Linked to the Authors of the Survey**
Linked: 9%
Not Linked: 64%
Unknown : 27%

** Belonging to any of the projects H2R, BALANCE, KoroiBot, WALK-MAN, BioMot.

Figure 1. (*Continued*) The results of the web-based survey: the 161 respondents to the questionnaire are international experts that were contacted through the following workshops, networks, or forums: the aforementioned involved research projects, i.e., H2R, BALANCE, KoroiBot, WALK-MAN, and BioMot; the euRobotics or Euron mailing list, Biomch-L mailing list, and robotics_worldwide mailing list; WeRob2014—The 2014 International Workshop on Wearable Robotics (werob2014.org); the European Network on Robotics for NeuroRehabilitation (European Commission COST Action TD1006, www.rehabilitationrobotics.eu); Dynamic Walking (dynamicwalking.org/); and the RehabRobotics mailing list (associated with ICORR, http://www.rehabrobotics.org/).

defined as the independent functional blocks needed to achieve a skill. Usually, several abilities—motor and/or cognitive—are needed for the achievement of one skill. *Performance* is the third relevant component of this scheme, defined as the level of achievement of the goal.

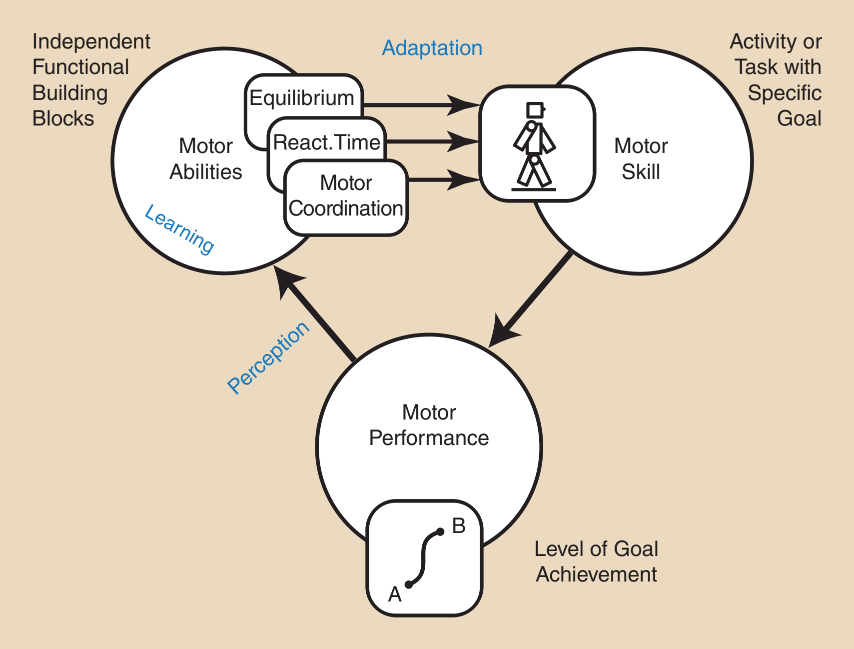
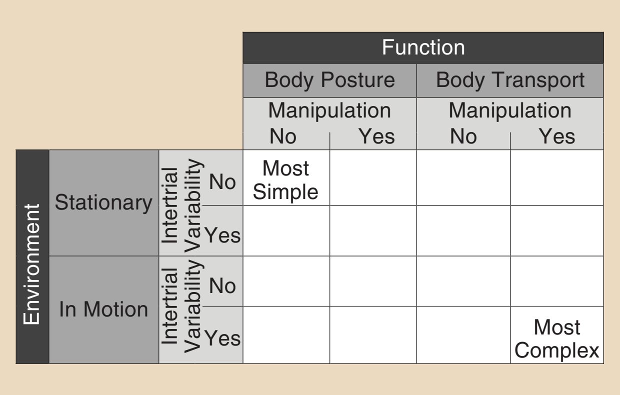


Figure 2. The basic components of our benchmarking taxonomy: motor skills, motor abilities, and motor performance. These components have important interdependencies. To achieve a desired motor performance (e.g., moving from A to B), different motor abilities (e.g., coordination, equilibrium, and reaction time) should be combined together, resulting in a functional motor skill (e.g., walking movements). These three motor aspects, whose quantitative measurement is the main objective of the proposed scheme, are associated with three corresponding internal processes (indicated in blue): 1) the perception of the sensory feedback resulting from the actual performance, 2) the learning of new control strategies, and 3) the adaptation of motor abilities necessary to generate an improved motor skill. The analysis of these internal processes, very specific to each community, goes beyond the scope of the proposed scheme.

Performance is a common aspect in clinical and robotic scenarios. Performance measures usually consist of discrete scales based on time, distance, or a percentage of goal achievement, and it can be obtained experimentally with no particular difficulty. Measures for skills and abilities are more difficult to obtain because they rely on generic concepts (e.g., walking, standing) and depend on continuous variables such as kinematics, kinetics, and muscular activity, which can hardly be translated into absolute metrics. For these reasons, appropriate classification methods can provide a useful basis for the organization of these concepts. If we look at a humanoid robot, or at a human in combination with a wearable device, as a sort of impaired version of the human machine, the potential benefit of using clinical-based taxonomies becomes apparent, since the process of (re)learning is common to both rehabilitation and machine-learning scenarios. Gentile [8] and Fleishman et al. [9] proposed successful taxonomies for motor skills and motor abilities that are commonly used in physical therapy and psychology.



				Function			
				Body Posture		Body Transport	
				Manipulation		Manipulation	
				No	Yes	No	Yes
Environment	Stationary	Intertrial Variability	No	Most Simple			
			Yes				
	In Motion	Intertrial Variability	No				
			Yes				Most Complex

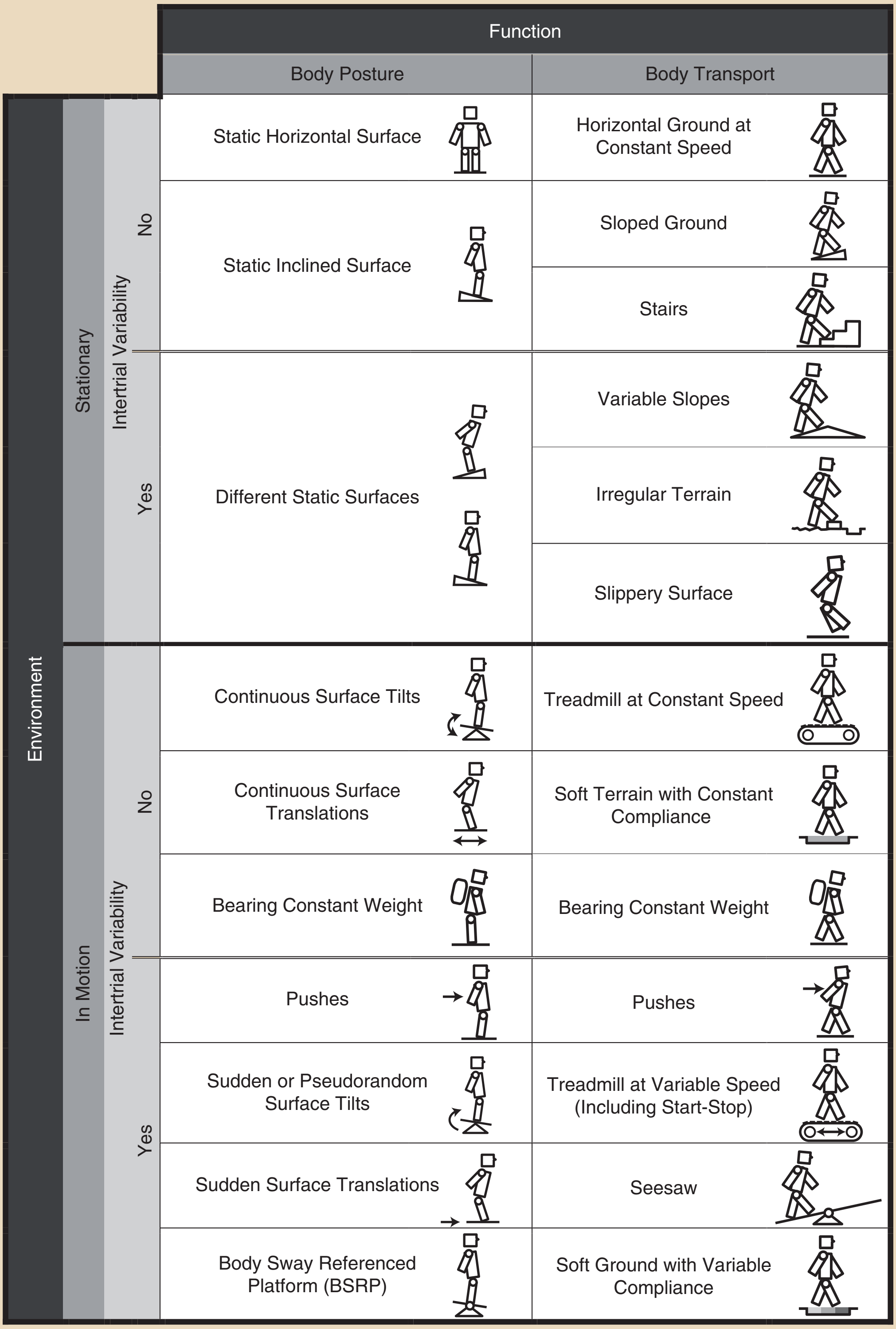
Figure 3. Gentile's taxonomy classifies motor skills according to two main dimensions, environment and function, and four intrinsic characteristics, i.e., environment motion, intertrial variability, body motion, and manipulation. The table allows for 16 possible categories, ordered in a simple-to-complex progression, from top-left to bottom-right.

Gentile's taxonomy (see Figure 3) classifies motor skills according to the following two general dimensions:

- the environment, represented by the elements in contact with the person during the execution of the skill, which can be classified according to two intrinsic characteristics: 1) its absolute motion and 2) the presence of intertrial variability, which indicates whether the environmental condition changes between two consecutive trials
- the function of the motor skill, which is classified according to 1) the orientation of the body, which can be maintained (e.g., in standing) or transported (e.g., in walking), and 2) the presence of object manipulation during the execution of the task.

The resulting combination of these characteristics is normally represented in a bidimensional table (see Figure 3), organized in terms of increasing complexity, from top-left to bottom-right positions. A typical rationale during a motor learning procedure is to begin with a stationary environment and no intertrial variability (e.g., repetitive trials of a single movement) and then to move toward a complete moving environment with intertrial variability (e.g., real-life and out-of-the-lab conditions). Similarly, but from the perspective of the function, skills that require static body posture are simpler than those requiring body transport.

Motor abilities underlining motor skills directly influence the performance of their execution. Fleishman [9] proposed a list of 54 independent motor, cognitive, and visual abilities at



Environment			Function	
			Body Posture	Body Transport
Stationary	Intertrial Variability: No		Static Horizontal Surface	Horizontal Ground at Constant Speed
			Static Inclined Surface	Sloped Ground
				Stairs
	Intertrial Variability: Yes		Different Static Surfaces	Variable Slopes
				Irregular Terrain
				Slippery Surface
In Motion	Intertrial Variability: No		Continuous Surface Tilts	Treadmill at Constant Speed
			Continuous Surface Translations	Soft Terrain with Constant Compliance
			Bearing Constant Weight	Bearing Constant Weight
	Intertrial Variability: Yes		Pushes	Pushes
			Sudden or Pseudorandom Surface Tilts	Treadmill at Variable Speed (Including Start-Stop)
			Sudden Surface Translations	Seesaw
			Body Sway Referenced Platform (BSRP)	Soft Ground with Variable Compliance

Note: All the disturbances may be applied in sagittal and frontal directions.

Figure 4. The motor skills considered in the benchmarking scheme. Since this scheme is limited to bipedal locomotion skills, the manipulation category originally included in Gentile's taxonomy (Figure 3) has been omitted. The concept of intertrial variability is analogous to the concept of unexpected disturbances.

the basis of a wide variety of skills, from locomotion to complex manipulation. We identified a subset of significant motor abilities from Fleishman's list related to lower-limb motion. They are interlimb coordination, static and dynamic strength, limb flexibility, gross body equilibrium, reaction time, speed of limbs, and control precision.

Proposed Benchmarking Scheme

The proposed benchmarking scheme is composed of three sections.

- *Motor skills classification*—reports the most relevant motor skills related to locomotion and standing, classified according to Gentile's taxonomy.
- *Benchmarking methods*—includes the benchmarks that can be used to quantitatively assess the specific abilities behind motor skills, according to an extended version of Fleishman's taxonomy.
- *Experimental protocols*—we propose a template of a work sheet that researchers can use to design their own experimental protocols within our framework.

Internal properties, such as cognitive abilities (e.g., perception, learning, planning, prediction, and adaptation; see Figure 2, blue items) or internal dynamics (e.g., intersegmental forces) are not considered in this scheme because its goal is to describe the different facets of the resulting performance and not to quantify the possible causes.

Scheme for Motor Skills Classification

Figure 4 shows our proposed taxonomy for motor skills, based on Gentile's approach. Under the perspective of function, we included the body posture conditions, corresponding to postural skills, and the body transport conditions, corresponding to locomotion skills. According to the environment characteristic, tasks are further classified depending on the motion of the environment and intertrial variability. For the sake of clarity, we point out that the concept of intertrial variability can be assimilated to the concept of unexpected disturbance, which is of more common use in the robotic field.

Motor skills related to body posture are depicted in the first column of Figure 4 and briefly defined as follows:

- *Static horizontal surface*: maintaining an upright posture.
- *Static inclined surface*: similar to the previous case, but using an inclined surface.
- *Different static surfaces*: automatically adapting to different and unknown inclinations.
- *Continuous surface tilts*: maintaining equilibrium on a support surface whose angular orientation is varied cyclically (e.g., sinusoidal), with constant amplitude and frequency.
- *Continuous surface translations*: similar to the previous case, but with translational displacements instead of angular displacements.
- *Constant weight bearing*: maintaining equilibrium after applying an external (and known by the system) weight to the body.
- *Sudden surface tilts*: equilibrating on a support surface whose angular orientation follows an unpredictable and variable pattern over time.
- *Sudden surface translations*: equivalent to the previous case, but in the horizontal direction.
- *Body sway referenced platform*: equilibrating on a platform that is tilted so that the relative angle between the foot and body is kept constant, eliminating ankle proprioceptive information [10].
- *Pushes*: maintaining equilibrium after external pushes with short duration.

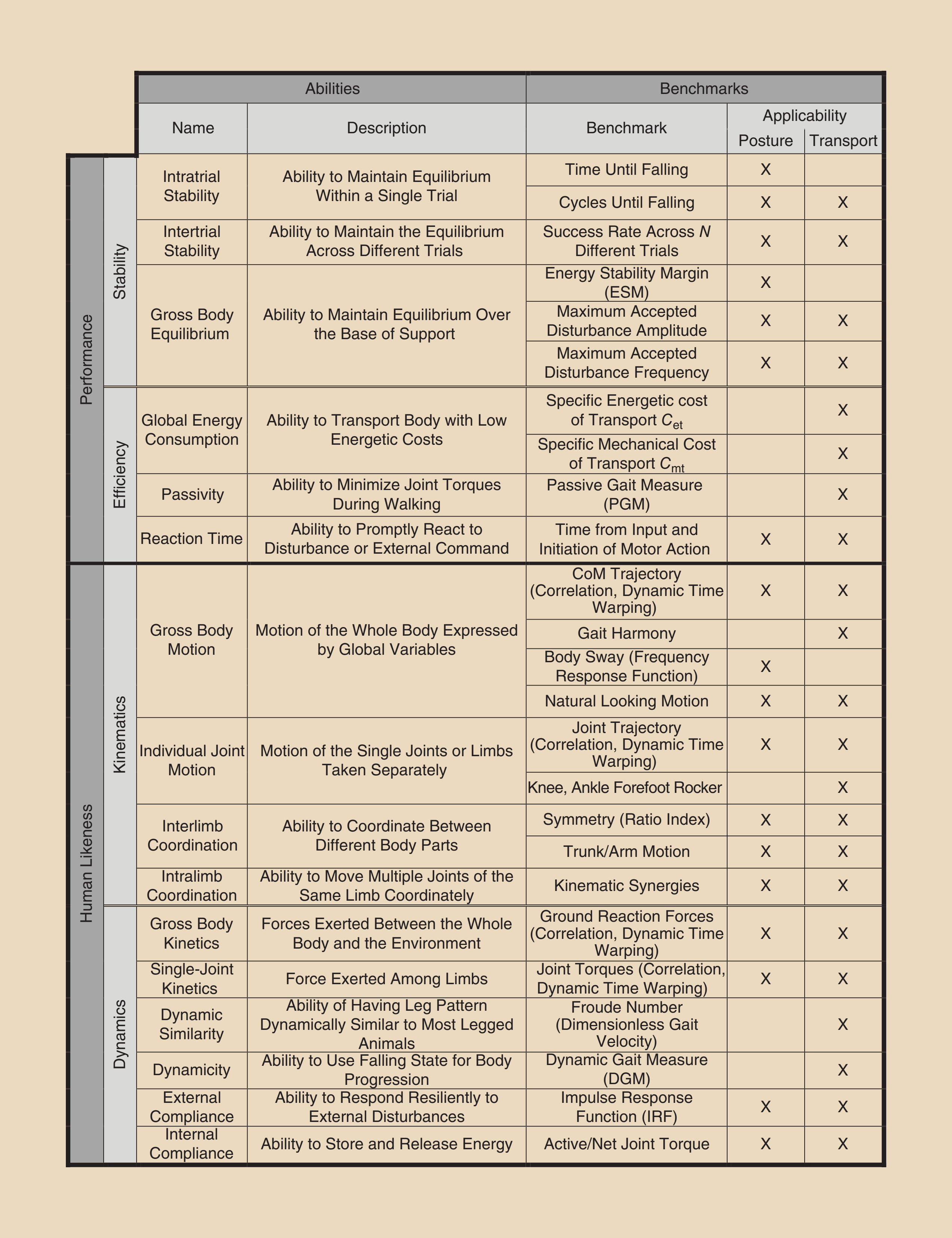
The objective of this article is to define the basis of a benchmarking scheme for the assessment of bipedal locomotion.

As for body transport (second column in Figure 4), different conditions have been identified:

- *Horizontal ground at constant speed*: maintaining steady walking over a static and horizontal ground, in the absence of any external disturbance.
- *Sloped ground*: equivalent to the previous condition, but on a fixed slope.
- *Variable slopes*: maintaining steady walking over various and unknown inclinations.
- *Stairs*: climbing stairs of constant and known dimensions.
- *Irregular terrain*: maintaining steady walking over different kinds of irregularities in the ground, including obstacles, uneven surfaces, and gaps.
- *Slippery surface*: maintaining equilibrium on a surface with unknown friction.
- *Treadmill at constant speed*: maintaining steady walking while the ground moves continuously at a constant speed.
- *Soft terrain with constant compliance*: maintaining steady walking over a terrain made of a soft material with known compliance (e.g., sand or foam). The material cannot be changed throughout the trial.
- *Weight bearing*: walking on a static horizontal ground at constant speed with an additional weight placed on the back of the robot.
- *Pushes*: maintaining steady walking after being pushed in different phases of the gait cycle. The pushes can have short or prolonged duration.
- *Treadmill at variable speed*: walking on a treadmill with variable velocity. Starting and stopping are also included as boundary conditions.
- *Seesaw*: steady walking on an unstable surface free to pivot around an horizontal axis perpendicular to the direction of walking.
- *Soft ground with variable compliance*: steady walking over a terrain made of different soft materials, whose location and compliance are not known a priori.

Scheme for Benchmarks

A schematic overview of the most relevant motor abilities related to bipedal functions is shown Figure 5. We have classified them into two main categories, performance and human likeness, in which performance is related to



		Abilities		Benchmarks		
		Name	Description	Benchmark	Applicability: Posture	Applicability: Transport
Performance	Stability	Intratrial Stability	Ability to Maintain Equilibrium Within a Single Trial	Time Until Falling	X	
				Cycles Until Falling	X	X
		Intertrial Stability	Ability to Maintain the Equilibrium Across Different Trials	Success Rate Across *N* Different Trials	X	X
		Gross Body Equilibrium	Ability to Maintain Equilibrium Over the Base of Support	Energy Stability Margin (ESM)	X	
				Maximum Accepted Disturbance Amplitude	X	X
				Maximum Accepted Disturbance Frequency	X	X
	Efficiency	Global Energy Consumption	Ability to Transport Body with Low Energetic Costs	Specific Energetic cost of Transport C_{et}		X
				Specific Mechanical Cost of Transport C_{mt}		X
		Passivity	Ability to Minimize Joint Torques During Walking	Passive Gait Measure (PGM)		X
		Reaction Time	Ability to Promptly React to Disturbance or External Command	Time from Input and Initiation of Motor Action	X	X
Human Likeness	Kinematics	Gross Body Motion	Motion of the Whole Body Expressed by Global Variables	CoM Trajectory (Correlation, Dynamic Time Warping)	X	X
				Gait Harmony		X
				Body Sway (Frequency Response Function)	X	
				Natural Looking Motion	X	X
		Individual Joint Motion	Motion of the Single Joints or Limbs Taken Separately	Joint Trajectory (Correlation, Dynamic Time Warping)	X	X
				Knee, Ankle Forefoot Rocker		X
		Interlimb Coordination	Ability to Coordinate Between Different Body Parts	Symmetry (Ratio Index)	X	X
				Trunk/Arm Motion	X	X
		Intralimb Coordination	Ability to Move Multiple Joints of the Same Limb Coordinately	Kinematic Synergies	X	X
	Dynamics	Gross Body Kinetics	Forces Exerted Between the Whole Body and the Environment	Ground Reaction Forces (Correlation, Dynamic Time Warping)	X	X
		Single-Joint Kinetics	Force Exerted Among Limbs	Joint Torques (Correlation, Dynamic Time Warping)	X	X
		Dynamic Similarity	Ability of Having Leg Pattern Dynamically Similar to Most Legged Animals	Froude Number (Dimensionless Gait Velocity)		X
		Dynamicity	Ability to Use Falling State for Body Progression	Dynamic Gait Measure (DGM)		X
		External Compliance	Ability to Respond Resiliently to External Disturbances	Impulse Response Function (IRF)	X	X
		Internal Compliance	Ability to Store and Release Energy	Active/Net Joint Torque	X	X

Figure 5. The motor abilities and related benchmarks are classified in two categories: performance and human likeness. The performance category includes all those abilities related to stability (ability of maintaining equilibrium) and efficiency. The human likeness category includes all those abilities related to typical human behavior, under the perspective of kinematics and dynamics. For each ability, a specific benchmark has been identified. The last column specifies in what classes of motor skills (i.e., the function category of Figure 4) the corresponding benchmark is applicable.

the accomplishment of the goal of a motor skill, and human likeness represents in what manner the task is executed, which can or cannot be correlated with the level of accomplishment of the goal. Each ability is associated to one or more benchmarks, which allows the quantitative measurement of the corresponding ability. To allow for truthful application across a wide variety of bipedal systems, all benchmarks should be made independent from weight and size.

Benchmarks of Performance

In our view, two features can describe performance: stability and efficiency. We define stability as the ability to maintain equilibrium during the execution of a motor skill. Loss of equilibrium can be easily detected by the occurrence of a falling event. To assess stability within a single trial, we identified two benchmarks: time until falling and cycles until falling. The time until falling should be used in all static postural conditions (e.g., quiet standing on a static surface) because the detection of a cycle cannot be easily determined. The number of cycles (e.g., walking stride cycle or tilting platform cycle) is more suitable during dynamic conditions, or when robots with different sizes are considered, because of the influence of speed and size on time. To measure stability across different trials, the success rate should also be measured. Another benchmark of the stability is the ability of maintaining the center of mass (CoM) above the polygon of support, reflecting what Fleishman referred to as *gross body equilibrium*. The ability can be measured analytically by the energy stability margin (ESM) [11], or by identifying the maximum accepted disturbance, in terms of amplitude and frequency. Measuring the energy efficiency of robots and humans can be done by the specific cost of transport (c_t) [12], [13], defined as the ratio of the energy consumed and the weight times the distance traveled. In robotics, to isolate the effectiveness of the mechanical design and controller from the efficiency of the actuators, the specific energetic cost of transport (c_{et}), comprises the total energy consumed, and specific mechanical cost of transport (c_{mt}), which only considers the positive mechanical work of the actuation system, have been introduced. Another way to assess the energetic aspects of locomotion has been recently introduced with the concept of passivity, defined as the ability of optimizing the use of gravity and inertia to move the body forward. The resulting passive gait measure (PGM) [14] appears to be a potential benchmark because of its practical use in robotic and human scenarios. Another aspect of efficiency is the ability of reacting promptly to an external command or perturbation, usually referred to as *reaction time*.

Benchmarks of Human Likeness

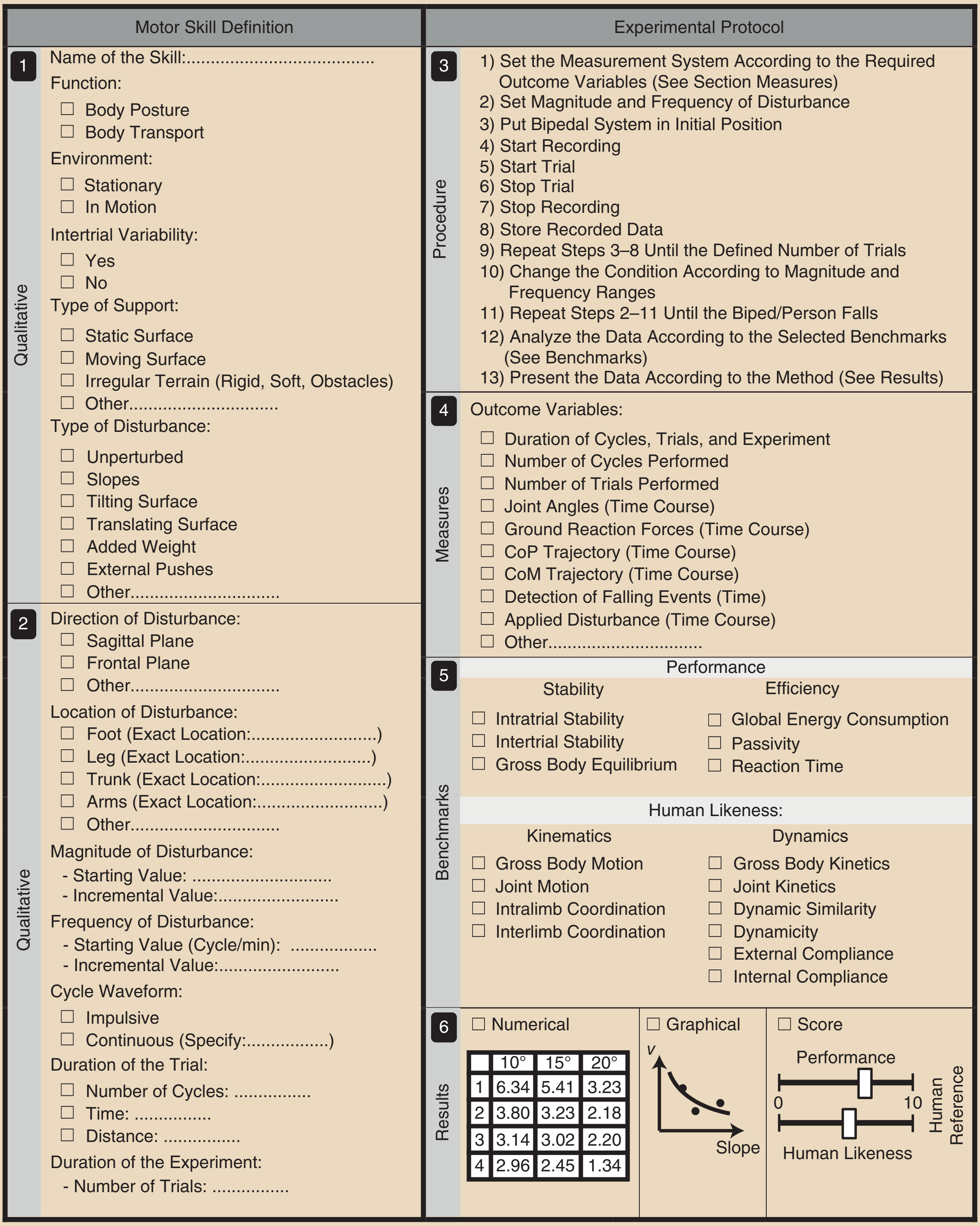
Human likeness is a term widely used in humanoid robot community to define the similarity with human behavior. The concept of healthy behavior is used instead in the fields of wearable robotics and human biomechanics. In our scheme, we propose to maintain the term human likeness, due to its conciseness. To translate this concept into a number of abilities and related benchmark, we divided human likeness into two categories, kinematics and dynamics (see Figure 5).

Under the kinematics category, we included all the abilities that can be analyzed by observing only the motion of the body. We have identified three further subcategories as follows.

- Whole body motion can be generally described by the motion of the CoM and compared with humans through correlation techniques, such as dynamic time warping [15]. Recently, other techniques for global movement assessment have been introduced, such as the gait harmony [16]. In the specific case of posture, the human-like whole-body sway is commonly considered [17], [18]. Global motion can be also assessed by visual inspection from human observers [19].
- Individual joint motion can be easily measured and compared with healthy humans [15]. Foot motion is also a crucial aspect in walking. In the ideal benchmarking scheme, the assessment of basic wheel-like mechanisms of the foot—i.e., heel, ankle, and forefoot rockers—should be included [20], [21].
- Coordination, which includes interlimb coordination, such as symmetry [22] and trunk/arm motions used for regulating body momentum [23]; and intralimb coordination, i.e., ankle-knee-foot synergies [24].

Benchmarking not only allows the assessment and comparison of the performance of different technologies but also defines and supports the standardization and regulation processes.

The category of dynamics includes all the abilities that are correlated with forces behind movements. The ground reaction forces are mostly used as a descriptor of the global kinetics of the body. Beyond the direct measurement of forces, some other interesting features related to dynamics can be considered and assessed. One of them is the dynamic similarity, introduced by Alexander et al. [25], which is defined with the following criteria: 1) geometric similarity, 2) equal phase relationships, 3) equal duty factors, 4) equal relative stride lengths, 5) equal relative ground reaction forces, and 6) equal relative mechanical power outputs. They verified experimentally that different-sized animals meet these six criteria when they move with the same Froude number. Therefore, the Froude number can be taken as a compact way to describe dynamic similarity between a robot and human, irrespective to size [26]. Mummolo et al. [14] recently proposed an indicator of dynamicity, i.e., the dynamic gait measure (DGM), defined as the ability of a legged system to maintain dynamic stability while



Motor Skill Definition

Qualitative

1

Name of the Skill:......................................

Function:
☐ Body Posture
☐ Body Transport

Environment:
☐ Stationary
☐ In Motion

Intertrial Variability:
☐ Yes
☐ No

Type of Support:
☐ Static Surface
☐ Moving Surface
☐ Irregular Terrain (Rigid, Soft, Obstacles)
☐ Other..............................

Type of Disturbance:
☐ Unperturbed
☐ Slopes
☐ Tilting Surface
☐ Translating Surface
☐ Added Weight
☐ External Pushes
☐ Other..............................

Qualitative

2

Direction of Disturbance:
☐ Sagittal Plane
☐ Frontal Plane
☐ Other..............................

Location of Disturbance:
☐ Foot (Exact Location:.........................)
☐ Leg (Exact Location:.........................)
☐ Trunk (Exact Location:.........................)
☐ Arms (Exact Location:.........................)
☐ Other..............................

Magnitude of Disturbance:
- Starting Value: ....
- Incremental Value:........................

Frequency of Disturbance:
- Starting Value (Cycle/min): ....
- Incremental Value:........................

Cycle Waveform:
☐ Impulsive
☐ Continuous (Specify:.................)

Duration of the Trial:
☐ Number of Cycles: ....
☐ Time: ....
☐ Distance: ....

Duration of the Experiment:
- Number of Trials: ....

Experimental Protocol

Procedure

3

1) Set the Measurement System According to the Required Outcome Variables (See Section Measures)
2) Set Magnitude and Frequency of Disturbance
3) Put Bipedal System in Initial Position
4) Start Recording
5) Start Trial
6) Stop Trial
7) Stop Recording
8) Store Recorded Data
9) Repeat Steps 3–8 Until the Defined Number of Trials
10) Change the Condition According to Magnitude and Frequency Ranges
11) Repeat Steps 2–11 Until the Biped/Person Falls
12) Analyze the Data According to the Selected Benchmarks (See Benchmarks)
13) Present the Data According to the Method (See Results)

Measures

4

Outcome Variables:
☐ Duration of Cycles, Trials, and Experiment
☐ Number of Cycles Performed
☐ Number of Trials Performed
☐ Joint Angles (Time Course)
☐ Ground Reaction Forces (Time Course)
☐ CoP Trajectory (Time Course)
☐ CoM Trajectory (Time Course)
☐ Detection of Falling Events (Time)
☐ Applied Disturbance (Time Course)
☐ Other..............................

Benchmarks

5

Performance

Stability	Efficiency
☐ Intratrial Stability	☐ Global Energy Consumption
☐ Intertrial Stability	☐ Passivity
☐ Gross Body Equilibrium	☐ Reaction Time

Human Likeness:

Kinematics	Dynamics
☐ Gross Body Motion	☐ Gross Body Kinetics
☐ Joint Motion	☐ Joint Kinetics
☐ Intralimb Coordination	☐ Dynamic Similarity
☐ Interlimb Coordination	☐ Dynamicity
	☐ External Compliance
	☐ Internal Compliance

Results

6

☐ Numerical

	10°	15°	20°
1	6.34	5.41	3.23
2	3.80	3.23	2.18
3	3.14	3.02	2.20
4	2.96	2.45	1.34

☐ Graphical

☐ Score

Figure 6. The template of the work sheet. This work sheet should be used each time someone wants to propose a new benchmarking protocol to the community so that others can replicate the experiment on different platforms and hardware configurations. Sections 1 and 2 should be used to contextualize the motor skill and the type of disturbances. Section 3 should contain a step-by-step description of the experimental protocol, to allow for its replication. Section 4 should include the variables to be measured to allow for computing the benchmarks specified in Section 5. Finally, in Section 6, the researcher should specify how results will be presented. Examples of how to use the work sheet in practical systems and experimental scenarios will be gathered in the web page www.benchmarkinglocomotion.org to promote its iterative use and improvement.

statically unbalanced, therefore useful to distinguish between zero-moment-point-based control approaches versus natural dynamics systems (e.g., passive walkers). A further relevant characteristic in biological locomotion is the compliance defined as the reciprocal of stiffness [27]. Compliance can be assessed by measurement and derivation of the displacement–force relationship in consequence to an external stimulus, usually referred to as an *impulse-response function* (*IRF*) [18]. At the joint level, it can be measured through the derivation of angle–torque relationship.

Scheme for Experimental Protocols

Establishing unified experimental protocols is one of the major challenges of the proposed scheme. To facilitate this process, we have developed a template for proposal, called a *work sheet* (see Figure 6). Its purpose is to encourage researchers and external collaborators to provide practical proposals for simplified experimental scenarios, which can be shared with other researchers and tested in different laboratories. Through an iterative approach, these experimental methods will be then refined and eventually get to feasible and agreed protocols. The work sheet is composed of two main sections (see Figure 6). The first column of the sheet should be used to define the specific motor skill, both qualitative, by classifying the skill according to the taxonomy, and quantitative, by specifying a set of parameters that characterize the type of disturbance, e.g., the location, direction, magnitude and frequency of disturbance, or the duration and number of trials. The second column of the work sheet should be used to allow the replication of the experimental protocol in different laboratories. Four kinds of information should be included: 1) the experimental procedure, 2) the applicable benchmarks, 3) the variables to be measured, and 4) the method of representing the results, i.e., numerical, graphical, or single scale.

Discussion

The need for quantitative metrics of bipedal motor skills is becoming increasingly relevant in humanoids, wearable robotics, and human biomechanics research. The web-based survey showed that, despite this heterogeneous scenario, the different communities share similar needs, with some minor differences. In humanoid research, it appears to be especially relevant to benchmark the performance under different perturbed conditions. In the wearable robotics community, there is a general interest on natural motion and postural stability. In human biomechanics, and particularly in the clinical fields, benchmarking should be focused on the detection of specific abnormal patterns with higher precision and reliability with respect to the current clinical scales. It is important to consider that the survey's respondents were mostly from the humanoid fields (40%), which could have biased the results toward this community. Nevertheless, we consider that the results obtained through the survey are a good starting step toward a unified framework. In general, respondents did not support anonymous and competition-based approaches. This does not deny the importance of robotic competitions, but states that the competitions are not being perceived as proper performance benchmarks. In this respect, our scheme can represent a complementary tool that can help researchers to find causal relationships between the performance during competition and the performance in each of the subfunctions identified in our scheme. This may provide additional clues to improve the technology, which is the (common) ultimate goal of all benchmarking efforts.

We have included in our benchmarking scheme most of the conditions and features that received higher scores in the survey. A preliminary version of this scheme was also discussed with the humanoid and exoskeleton communities in two recent international workshops (the Workshop on Benchmarking Bipedal Locomotion at the IEEE-RAS International Conference on Humanoid Robots, 18 November 2014, Madrid, Spain, and the European Robotics Forum 2015 Session on Replicable Robotics Research and Benchmarking). One aspect that was discussed extensively was the importance of a common terminology. This discussion led us to put more emphasis on the definitions of specific terms—such as *motor skills*, *abilities*, and *performance*—and resulted in the proposal of the different taxonomies.

A relevant issue arisen from the discussions is related to the benchmarking of control algorithms and other internal cognitive processes. On the one hand, we found this topic of extreme importance, being the basis of the resulting performance measured by our scheme. On the other hand, we observed that the internal processing strategies have great variability across the different communities to which this scheme is addressed. Therefore, we believe that this topic should be considered and discussed within each community independently. This process will eventually result in benchmarks that can be either added to this unified scheme or included as community-specific add-ons.

The web-based survey showed that different communities share similar needs.

One problem when defining similarity between different systems is that the dynamic and kinematic properties, including elementary properties such as weight, size, mass distribution or number of degrees of freedom (DoF), but also the corresponding kinematic and dynamic constraints, have to be considered. As for weight and size, some of the proposed benchmarks already consider these differences in their scores (e.g., Froude number, cost of transport). Other methods did not explicitly include scaling laws [e.g., ground reaction forces (GRFs), reaction time, joint torques], therefore requiring further discussion within the community to establish clear rules for scaling. As for the differences in the DoF, in the cluster of the European projects contributing to this article, some groups are currently investigating this issue and working on how to best compare similarity in the common DoF while considering the effect of the noncommon ones.

Certainly, some of the proposed benchmarks might not be considered effective in specific scenarios or systems. However, the proposed benchmarking scheme should not be used as a whole. Researchers are encouraged to choose only those features that are in line with their objectives. At the same time, this scheme is conceived as a flexible platform, open to new contributions and extensions resulting from international discussions. For instance, motor skills copying with voluntary transitions, such as changes in walking speeds, transitions from standing to walking, or turning, are still not present and will be considered for their inclusion. As for the experimental procedure, it appears necessary to ensure replicability of the benchmarking protocols. At the same time, the scheme should leave a certain measure of freedom in the application of the experimental procedures, due to the wide range of systems and laboratory conditions across the different communities. The work sheet has been conceived for this purpose. The major goal of the work sheet is to standardize the design process of a benchmarking protocol, therefore maximizing its potential use across different scenarios and end users. To this aim, special efforts should be made to translate the method currently used in human motion analysis (based on GRF, CoM, or CoP measurements) into minimal experimental setups, which allow at the same time fast, versatile and sufficiently accurate results across platform with different hardware.

Defining a common taxonomy is a basic requisite for a successful benchmarking scheme.

Conclusions and Road Map

In this article, we have set the foundations of a general structure for benchmarking bipedal motor skills. The originality of our approach is threefold. First, the proposed scheme is comprehensive, i.e., it arranges the great majority of bipedal motor functions into a meaningful taxonomic structure, using a classification scheme based on motor skills, abilities, and performance. This global method of classifying motor functions, inherited from the field of rehabilitation and psychology, has not been proposed or applied in the robotics scenario previously. Second, it is function based, i.e., it analyzes specific subfunctions of the global motor behavior instead of evaluating the general accomplishment of a goal. This approach is innovative because, if applied in combination with existing goal-based benchmarking analysis (e.g., U.S. Defense Advanced Research Projects Agency Robotics Challenge), can provide clues on the causal relationships between the sensorimotor mechanisms and resulting behavior. Third, the scheme is collaborative, i.e., it requires the participation of the community in proposing and refining new protocols and benchmarks, e.g., by means of the work sheet tool provided in this article. To encourage this collaborative process, we recently created a website (www.benchmarkinglocomotion.org), which will allow researchers to participate actively in the definition and improvement of the scheme. Similarly, we created a mailing list (https://listas.csic.es/wws/info/benchmarking_list), which is currently used to disseminate related events and topics.

In the road map toward an interdisciplinary and international consensus, we have identified some crucial steps. The first step is to identify and test the experimental protocols on different bipedal systems to verify to what extent two different systems/laboratories can share the same procedures. In this respect, a crucial factor will be the involvement of the robotic platforms currently available in the literature, and preferably those already participating in other benchmarking initiatives, to start defining standard procedures and calculating representative scores. The second step would be the refinement of the benchmarking scheme to formalize additional/specific goals for each (sub) community. In particular, we envision the development of community-specific schemes, such as those related to benchmarking of cognitive and algorithmic processes, currently not included in this scheme. The third step should be directed to discuss with current standardization working groups on robotic technology the appropriate strategies to translate the proposed benchmarks and metrics into future standards. (The working groups include IEC SC62A and ISO TC184/SC2/JWG 9—Medical Electrical Equipment and Systems Using Robotic Technology, ISO TC184/SC2/WG1—Vocabulary and Coordinate Systems, ISO WG7—Personal Care Robot Safety, and ISO WG8—Service Robots.) This last step will be essential for an appropriate market introduction of the new robotic technologies.

Acknowledgments

This research has been supported by the European Seventh Framework Program projects H2R (grant 600698), BALANCE (grant 601003), KoroiBot (grant 611909), WALK-MAN (grant 611832), and BioMot (grant 611695) as well as the European Commission COST TD1006—European Network on Robotics for NeuroRehabilitation. We would especially like to thank all the participants in the web-based survey and the attendees at the past workshops for their participation in the discussion.

References

[1] S. Behnke, “Robot competitions ideal benchmarks for robotics research,” in *Proc. Workshop Benchmarks Robotics Research*, 2006, pp. 13–17.

[2] A. P. del Pôbil, “Why do we need benchmarks in robotics research?” in *Proc. Workshop Benchmarks Robotics Research*, 2006, pp. 9–11.

[3] (2014, Dec. 10). DARPA robotics challenge trials. [Online]. Available: http://www.theroboticschallenge.org/

[4] (2014, Dec. 10). CYBATHLON 2016: The championship for robot-assisted parathletes.[Online]. Available: http://www.cybathlon.ethz.ch

[5] S. Tyson and L. Connell, "The psychometric properties and clinical utility of measures of walking and mobility in neurological conditions: A systematic review," *Clin. Rehab.*, vol. 23, no. 11, pp. 1018–1033, 2009.

[6] EURON. (2014, Dec. 10). Introduction: Benchmarks in robotics research. Survey and inventory of current efforts in comparative robotics research. [Online]. Available: http://www.robot.uji.es/benchmarks

[7] R. A. Magill, *Motor Learning and Control: Concepts and Applications*. New York: McGraw-Hill, 2007.

[8] A. M. Gentile, "Skill acquisition: Action, movement, and neuromotor processes," in *Movement Science: Foundations for Physical Therapy in Rehabilitation*, J. H. Carr, R. B. Shepherd, J. Gordon, Eds. Rockville, MD: Aspen Publishers Inc., 1987, pp. 93–154.

[9] E. A. Fleishman and M. K. Quaintance, *Taxonomies of Human Performance*. Orlando, FL: Academic Press, 1984.

[10] A. Ishida, S. Imai, and Y. Fukuoka, "Analysis of the posture control system under fixed and sway-referenced support conditions," *IEEE Trans. Biomed. Eng.*, vol. 44, no. 5, pp. 331–336, 1997.

[11] D. Messuri and C. Klein, "Automatic body regulation for maintaining stability of a legged vehicle during rough-terrain locomotion," *IEEE J. Robot. Automat.*, vol. 1, no. 3, pp. 132–141, 1985.

[12] G. Gabrielli and T. von Kármán, "What price speed? Specific power required for propulsion of vehicles," *Mech. Eng. ASME*, vol. 72, no. 10, pp. 775–781, 1950.

[13] S. Collins, A. Ruina, R. Tedrake, and M. Wisse, "Efficient bipedal robots based on passive-dynamic walkers," *Science*, vol. 307, no. 5712, pp. 1082–1085, 2005.

[14] C. Mummolo and J. H. Kim, "Passive and dynamic gait measures for biped mechanism: Formulation and simulation analysis," *Robotica*, vol. 31, no. 4, pp. 555–572, Oct. 2012.

[15] N. E. Helwig, S. Hong, E. T. Hsiao-Wecksler, and J. D. Polk, "Methods to temporally align gait cycle data," *J. Biomech.*, vol. 44, no. 3, pp. 561–566, Feb. 2011.

[16] M. Iosa, F. Paradisi, S. Brunelli, A. S. Delussu, R. Pellegrini, D. Zenardi, S. Paolucci, and M. Traballesi, "Assessment of gait stability, harmony, and symmetry in subjects with lower-limb amputation evaluated by trunk accelerations," *J. Rehab. Res. Develeop.*, vol. 51, no. 4, pp. 623–634, 2014.

[17] T. Mergner, G. Schweigart, and L. Fennell, "Vestibular humanoid postural control," *J. Physiol. Paris*, vol. 103, nos. 3–5, pp. 178–194, 2009.

[18] A. D. Goodworth and R. J. Peterka, "Contribution of sensorimotor integration to spinal stabilization in humans," *J. Neurophysiol.*, vol. 102, no. 1, pp. 496–512, July 2009.

[19] N. F. Troje, "Decomposing biological motion: A framework for analysis and synthesis of human gait patterns," *J. Vis.*, vol. 2, no. 5, pp. 371–387, Jan. 2002.

[20] J. Perry, *Gait Analysis: Normal and Pathological Function*. Thorofare, NJ: SLACK Inc., 1992.

[21] A. H. Hansen, D. S. Childress, and E. H. Knox, "Roll-over shapes of human locomotor systems: effects of walking speed," *Clin. Biomech. (Bristol, Avon)*, vol. 19, no. 4, pp. 407–414, May 2004.

[22] H. Sadeghi, P. Allard, F. Prince, and H. Labelle, "Symmetry and limb dominance in able-bodied gait: A review," *Gait Posture*, vol. 12, no. 1, pp. 34–45, Sept. 2000.

[23] M. Popovic, A. Hofmann, and H. Herr, "Angular momentum regulation during human walking: Biomechanics and control," in *Proc. IEEE Int. Conf. Robotics and Automation*, 2004, vol. 3, pp. 2405–2411.

[24] F. Lacquaniti, Y. P. Ivanenko, and M. Zago, "Kinematic control of walking," *Arch. Italiennes de Biologie*, vol. 140, no. 4, pp. 263–272, Oct. 2002.

[25] R. M. Alexander, "The gaits of bipedal and quadrupedal animals," *Int. J. Robot. Res.*, vol. 3, no. 2, pp. 49–59, June 1984.

[26] C. L. Vaughan and M. J. O'Malley, "Froude and the contribution of naval architecture to our understanding of bipedal locomotion," *Gait Posture*, vol. 21, no. 3, pp. 350–362, Apr. 2005.

[27] A. J. Ijspeert, "Biorobotics: Using robots to emulate and investigate agile locomotion," *Science*, vol. 346, no. 6206, pp. 196–203, Oct. 2014.

Diego Torricelli, Cajal Institute, Spanish National Research Council (CSIC), Madrid, Spain. E-mail: diego.torricelli@csic.es.

Jose González-Vargas, Cajal Institute, Spanish National Research Council (CSIC), Madrid, Spain.

Jan F. Veneman, Tecnalia Research and Innovation, San Sebastian, Spain. E-mail: jan.veneman@tecnalia.com.

Katja Mombaur, University of Heidelberg, Heidelberg, Germany. E-mail: katja.mombaur@iwr.uni-heidelberg.de.

Nikos Tsagarakis, Istituto Italiano di Tecnologia, Genova, Italy. E-mail: nikos.tsagarakis@iit.it.

Antonio J. del-Ama, National Hospital for Spinal Cord injury, Toledo, Spain. E-mail: ajdela@sescam.jccm.es.

Angel Gil-Agudo, Hospital Nacional de Paraplejicos, Toledo, Spain. E-mail: amgila@sescam.jccm.es.

Juan C. Moreno, Cajal Institute, Spanish National Research Council (CSIC), Madrid, Spain. E-mail: jc.moreno@csic.es.

Jose L. Pons, Cajal Institute, Spanish National Research Council (CSIC), Madrid, Spain. E-mail: jose.pons@csic.es.

Wearable Inertial Sensors

Applications, Challenges, and Public Test Benches

©ISTOCK PHOTO.COM/CHESKY_W

By Barbara Bruno, Fulvio Mastrogiovanni, and Antonio Sgorbissa

The recent technological advances in sensor miniaturization and embedded processing have provided new challenges and possibilities to the field of wearable computing. Two research areas are particularly interested by this innovation: healthcare technology applications that are devoted to analyzing the daily activities of a person to evaluate their general health, and personal dead reckoning (PDR) systems that focus on the analysis of the person's movements to keep track of his/her position in dangerous environments and situations. The identification of suitable algorithms and techniques to process wearable sensors data is a research challenge that must be overcome for both areas. The possibility to compare different solutions over public test benches is crucial to this aim. For this reason, we present the human odometry outdoor data set (HOOD), a public data set for the PDR systems and the wearable human activity recognition folder (WHARF), a public repository for human activity recognition (HAR), composed of over 1,000 acceleration recordings referring to 14 daily activities, and a MATLAB library allowing the creation and validation of acceleration models of the activities.

Digital Object Identifier 10.1109/MRA.2015.2448279
Date of publication: 18 August 2015

Motivations

Thanks to the significant technological advances in sensor miniaturization and embedded processing over the last decade, the wearable computing has risen in recent years to become one of the most active and expanding research fields of the scientific community. Off-the-shelf inertial measurement units (IMUs) containing accelerometers, gyroscopes, and compasses have package sizes as small as 4 mm × 4 mm × 1 mm and are specifically designed to meet the low-cost, low-power, and high-performance requirements of mobile consumer electronics equipment [1]. Sensing devices (see Figure 1) embedded in watches and wristbands (e.g., the Samsung Gear S and the Nike+ FuelBand), worn as necklaces (like the Samsung Gear Circle), or attached to the belt (such as Fitbit One) continuously acquire inertial data and allow the creation of unique correspondences between human actions and data patterns, which can be used to determine the status of the person.

Two research areas can particularly benefit from real-time, detailed knowledge of a person's actions: health-care technology applications that extract data from the IMU to evaluate the general health of a person, and the PDR systems that analyze the person's movements to keep track of his/her position in environments and situations that external localization systems cannot be used in. In both the cases, given the similarity of the sensory input, the differences between available solutions are the adopted algorithms for the analysis of raw data [2], [3], and, thus, the possibility to validate different methods over large data sets is crucial for the assessment of their performance.

Currently, this possibility is greatly limited by three factors.

- The number of publicly available data sets of the raw wearable IMU data is still quite small [4].
- The lack of a standard definition for most human actions leads to multiple definitions of the same activity, making it harder to merge different data sets.
- The number of publicly available algorithms for the wearable sensing applications is small, thus, further reducing the possibility of comparing the performance of different systems on the same data set.

To address these issues, we have collected two data sets of raw wearable sensors data. The HOOD is a public collection of labeled accelerometer and gyroscope data recordings for PDR applications. It considers six motion types (slow walking, normal walking, running, slow crawling, fast crawling, slithering), six outdoor environments (grass field, uphill road, staircase, riverbed, woods, snow), and four sensor placements (foot, waist, wrist, chest), tested on two different path configurations (straight line, zigzag). The WHARF is a public repository for HAR data sets and algorithms, including but not limited to health monitoring systems. The WHARF repository currently hosts a MATLAB implementation of the HAR system that we have developed [5], [6], together with a large data set of accelerometer data recordings for the recognition of daily activities with the wearable sensing systems [7], that we have used for the validation of the proposed system.

Figure 1. The commercially available wearable sensing devices: (a) Samsung Gear S, (b) Nike+ FuelBand, (c) Samsung Gear Circle, and (d) Fitbit One.

Table 1 is a reference for research groups willing to get more information on the experiments, or extend the data sets. The Data Specifications section of Table 1 specifies the sensing device settings and the data formats to be adopted for the acquisition of additional recordings, to ensure full compatibility with the available ones. The Experiment section of the table reports the testing conditions considered in the WHARF and HOOD data sets. After configuring a sensing device, researchers can contribute to the two data sets by providing additional recordings. For example, one could extend the WHARF data set with new recordings of the 14 indoor activities by placing the device on the left wrist (i.e., adding a configuration to the device placement category). Others could extend the HOOD data set with the recordings referring to a new environment, a new motion, or both (for example, by adding recordings of a person swimming in a pool). While extending the WHARF or HOOD data set, one must comply with the reported data specifications; needless to say, it is always possible to add a new data set (ideally, a new column in Table 1) with different data specifications.

The chosen repository hosting service includes tools for distributed revision control and source code management, which encourage the discussion, interaction, and cooperation on the topic of public test benches for wearable inertial systems. In addition to downloading the available data sets, the researchers can create private copies in their own repository and propose additions and modifications.

Human Activity Recognition

The HAR systems are devoted to the identification of specific activities of interest out of all actions executed by a person during the day. Research on the HAR systems is particularly active and is developing in two different directions.

Fitness-oriented applications focus on sport-related activities (primarily jogging). The sensory input (most commonly acceleration associated with the person's movements) is provided either by an ad hoc wearable device [8] or by the IMU embedded in a smartphone [9]. The

Table 1. Description of data sets.

		WHARF	HOOD
Data specifications	Sensors	Triaxial accelerometer	Triaxial accelerometer, gyroscope
	Accelerometer range	$[-1.5\,g; +1.5\,g]$	$[-2g; +2g]$
	Accelerometer sensitivity	6 b/axis	16 b/axis
	Gyroscope range	–	$[-250^\circ/s; +250^\circ/s]$
	Gyroscope sensitivity	–	16 b/axis
	Sampling frequency	32 Hz	40 Hz
	Data format	*ax ay az*	*ax ay az gx gy gz*
Experiment	Device placement	Right wrist	Right foot Waist Chest Right wrist
	Location	House (indoor)	Grass field Uphill road Staircase (outdoor) Riverbed Woods Snow
	Activity	Brush own teeth Comb own hair Get up from the bed Lie down on the bed Sit down on a chair Stand up from a chair Drink from a glass Eat with fork and knife Eat with spoon Pour water into a glass Use the telephone Climb the stairs Descend the stairs Walk	Slow walking Normal walking Running Slow crawling Fast crawling Slithering
Data set	Hosting platform	https://github.com/centaurresearchgroup/WHARF	https://github.com/centaurresearchgroup/HOOD
	Contacts	barbara.bruno@unige.it, centaur.research.group@gmail.com	barbara.bruno@unige.it, centaur.research.group@gmail.com

algorithms implemented in the HAR system analyze the acceleration patterns and make the extracted information available to the user. Common functionalities include step count, estimation of the calories expended, tracking activity duration, and the path the user followed.

Monitoring systems for elderly care focus on the recognition of a set of specific daily life activities, called *activities of daily living* (*ADL*), which are used by gerontologists to estimate the level of autonomy of a person. As graphically summarized in Figure 2, the ADL include involuntary movements, simple motions such as sitting and lying down, mechanical activities performed indoors such as housekeeping and preparing food, outdoor activities such as shopping, and complex cognitive activities such as handling finances and taking medications [7]. The sensory input required for the recognition varies with the considered ADL [7] and allows the classification of existing solutions in three different approaches: smart environments rely on heterogeneous sensors distributed in the environment and infer the status of the person from the context [10], while smartphone-based systems [11] and wearable sensing systems [3] rely on sensors located on the person's body and imply the status of the person from their limb movements.

Wearable sensing, in particular, is the most suitable approach for monitoring location-independent ADL, i.e., activities that can virtually take place anywhere, since they require little or no interaction with the environment. The most commonly considered location-independent ADL are reported in the left column of Table 2. By identifying one or more simple motions that uniquely refer to each ADL and that are associated with stereotyped acceleration patterns, it is possible to create unique correspondences between the acceleration patterns and the ADL. These simple motions are called *human motion primitives* (*HMP*); the right column of Table 2 lists some of the HMP associated with the location-independent ADL.

Most wearable sensing systems devoted to the recognition of HMP rely exclusively on acceleration information and share a similar architecture [3], graphically summarized in Figure 3. The main tasks of the system are: 1) to extract relevant features from the available sensory data, 2) to create representations of the target HMP in terms of the available features, and 3) to classify the run-time sensory data according to the known representations.

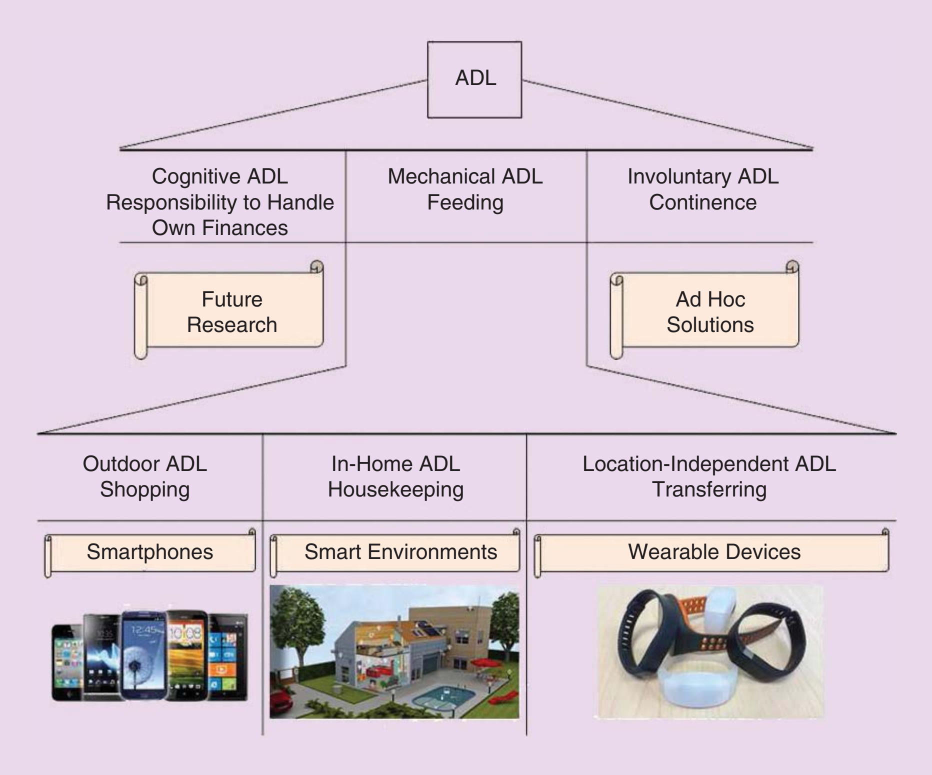


Figure 2. By considering the cognitive abilities and environment characteristics required for the successful execution of an ADL it is possible to identify its most suitable monitoring technique. The state-of-the-art monitoring systems allow for the recognition of: outdoor ADL via smartphone-based systems, in-home ADL with smart environments, and location-independent ADL with wearable sensing systems.

WHARF Data Set

The WHARF data set is composed of over 1,000 acceleration recordings of all the HMP listed in Table 2, collected from 17 volunteers (11 men and six women, ages 19 to 81). The data set allows for the creation and the validation of the HMP models. By analyzing the performance of different modeling and classification algorithms over the same validation data set it is possible to compare different solutions.

Each recording reports the triaxial acceleration values registered during one execution of one motion by an ad hoc sensing device worn at the right wrist. As summarized in Table 1, the device, shown in Figure 4, is equipped with a single triaxial accelerometer with measurement range of $[-1.5g; +1.5g]$ and sensitivity of 6 b/axis, working at the sampling frequency of 32 Hz.

All the trials were recorded at the homes of the volunteers in a series of supervised experiments. During each experiment, a volunteer equipped with the device is asked to perform one of the motions of interest and the supervisor labels the corresponding acceleration data as an execution of the motion. To increase the naturalness of the executions in each session, the supervisor defines which motions are to be performed and the number of repetitions, which varies with the motion and the volunteer, while the volunteer is allowed to choose which motion to perform each time.

Table 2. The location-independent ADL and HMP.

ADL	Motion Primitive
Toileting	Brush own teeth Comb own hair
Transferring	Get up from the bed Lie down on the bed Sit down on a chair Stand up from a chair
Feeding	Drink from a glass Eat with fork and knife Eat with spoon Pour water into a glass
Ability to use telephone	Use the telephone
Indoor transportation	Climb the stairs Descend the stairs Walk

HMP Detection Library

The HMP detection library is a public library of MATLAB functions designed to work with the WHARF data set and allows for:

- the creation of HMP models (models builder)
- the classification of unknown recordings on the basis of existing models (classifier).

The models builder library [5] executes the steps, shown in Figure 5, to create the model of an HMP of interest on the basis of a representative modeling data set. The WHARF data set provides both specific, i.e., biased toward the behavior of one individual, and general, i.e., not biased, modeling

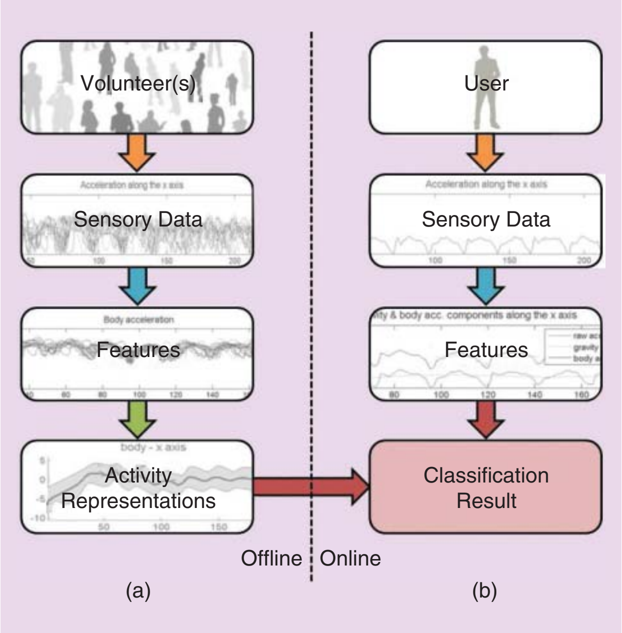


Figure 3. The wearable sensing systems share a similar architecture, devoted to (a) the creation of representations of the activities of interest on the basis of a modeling data set and (b) the classification of run-time data by comparison with the known representations.

Figure 4. The wearable sensing device used for the WHARF data set recordings. The device, equipped with a triaxial accelerometer, is worn on the right wrist.

data sets. During the data filtering and feature extraction phase, the models builder first applies a median filter to reduce the high-frequency noise affecting the acceleration signals (step 1), then, it extracts from the triaxial acceleration two separate four-dimensional (4-D) features, gravity $g = (t, g_x, g_y, g_z)$ and body acceleration $b = (t, b_x, b_y, b_z)$. The method for gravity and body acceleration extraction is implemented in the following two steps.

1) A low-pass Chebyshev I fifth-order filter is applied to the acceleration signal to isolate the gravity component (step 2).
2) The gravity component is subtracted from the original signal to obtain the body acceleration component (step 3). Gaussian mixture modeling and Gaussian mixture regression (step 4)

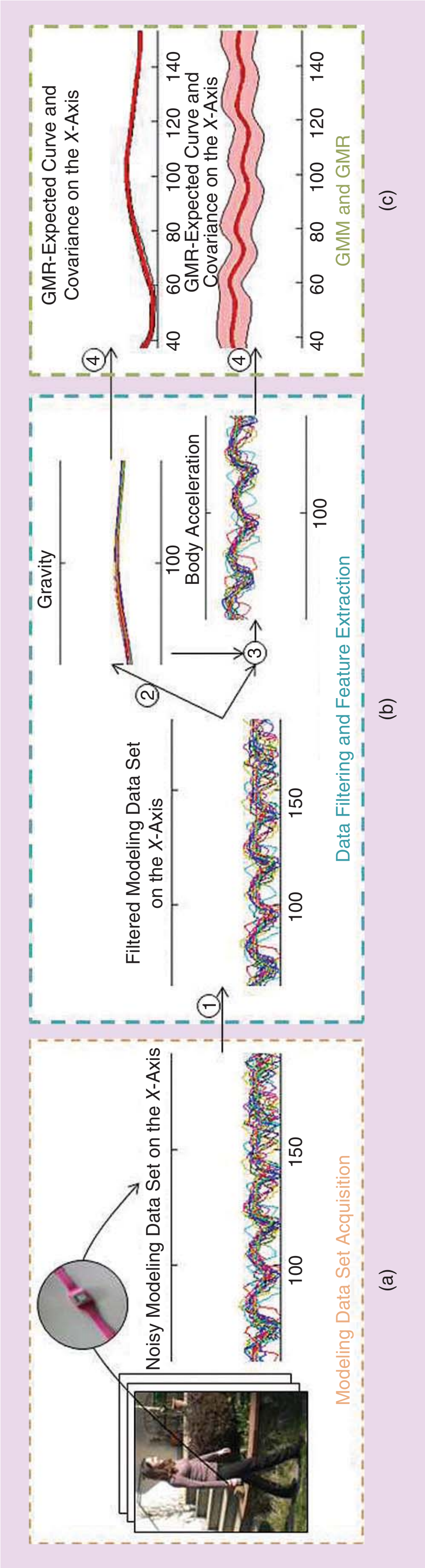


Figure 5. The models builder library is devoted to the creation of models of HMP on the basis of a representative modeling data set. (a) The raw acceleration data are filtered for noise reduction (step 1). (b) Next, the data is processed for the extraction of two 4-D features (step 2) gravity and (step 3) body acceleration. (c) Finally, Gaussian mixture modeling (GMM) and Gaussian mixture regression (GMR) build a generalized version of each feature (step 4). The model of the HMP is defined as the set of the two 4-D generalized features.

build a single, generalized, and scalable version of each feature in the form of:

$$\hat{\Xi}^{\xi} = (\mu^{\xi}, \Sigma^{\xi}), \tag{1}$$

where ξ denotes a generic feature (i.e., ξ can either correspond to gravity g or body acceleration b), μ^{ξ} is the expected curve modeling feature ξ and Σ^{ξ} is the covariance matrix associated with μ^{ξ}. The model of any HMP m is defined from the feature models $\hat{\Xi}^{g}$ and $\hat{\Xi}^{b}$ as:

$$\hat{\Xi}^{m} = (\hat{\Xi}^{g}, \hat{\Xi}^{b}). \tag{2}$$

The classifier library [6] executes the steps, shown in Figure 6, to rank the similarity between one recording and each available model of HMP by computing the likelihood of its features g and b to the model's features. The classifier moves a sliding window Ξ_w over the validation recording acceleration stream to isolate the portions of data. During the data filtering and feature extraction phase, the classifier executes the same algorithms of the models builder (steps 1–3), allowing the creation of two sets Ξ_w^{ξ} representing the window feature acceleration data. In the comparison and classification phase, the overall distance between the window data and a model is computed as the average Mahalanobis distance on the two features (step 5). A threshold mechanism is set up to discriminate between unknown and potentially known HMP and the window is finally labeled as an occurrence of the HMP with minimum distance (step 6).

The HMP detector graphical user interface (GUI), shown in Figure 7, provides an easy-to-use interface to the models builder and the classifier, allowing the selection of any arbitrary set of HMP to model and the classification of any arbitrary recording.

With the WHARF data set and the HMP detection library it is possible to create, modify, and update models of HMP and test the performance of the proposed modeling and classification procedures on the available validation data sets, as well as on any recordings compatible with the WHARF data set format (see Table 1).

The presented system has been envisioned to minimize the sensory requirements (i.e., it relies on acceleration information exclusively), in an effort to derive mapping between the monitored activites, the required sensing infrastructure, and suitable data analysis procedures. Indeed, it could be considered as a preliminary step toward the creation of monitoring systems effectively relying on multiple sensory modalities. In this context, for example, an analysis of the acceleration patterns associated with the motions of walking and climbing stairs (very similar one to the other) suggests the use of external localization sensors to verify whether or not the person is in the proximity of a staircase [12]. Part of our current work focuses on the design of a multimodal monitoring system which merges information provided by the wearable and heterogeneous

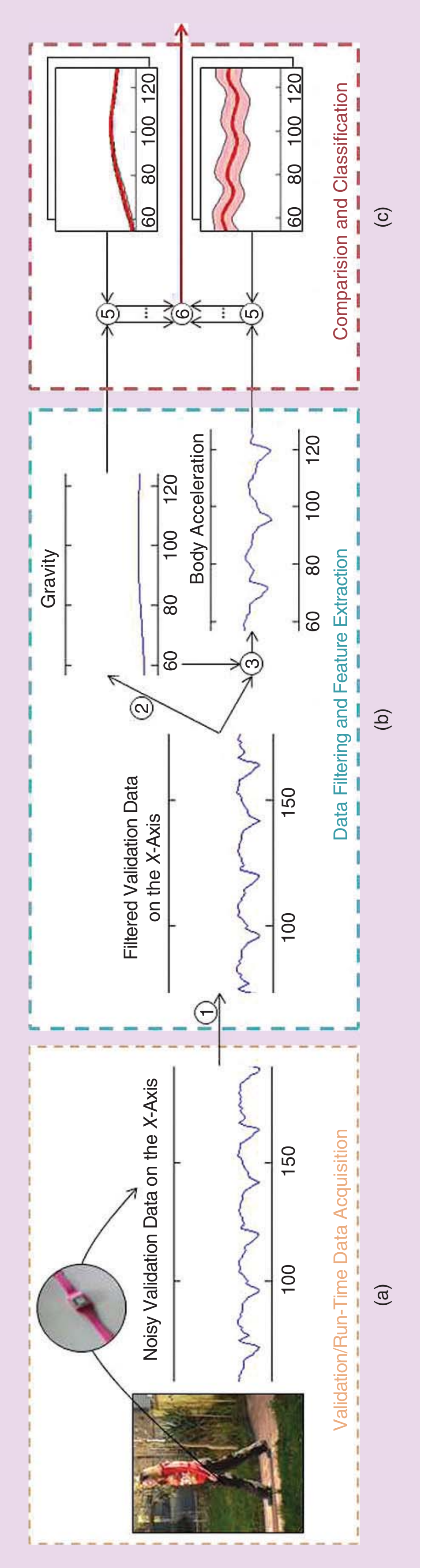
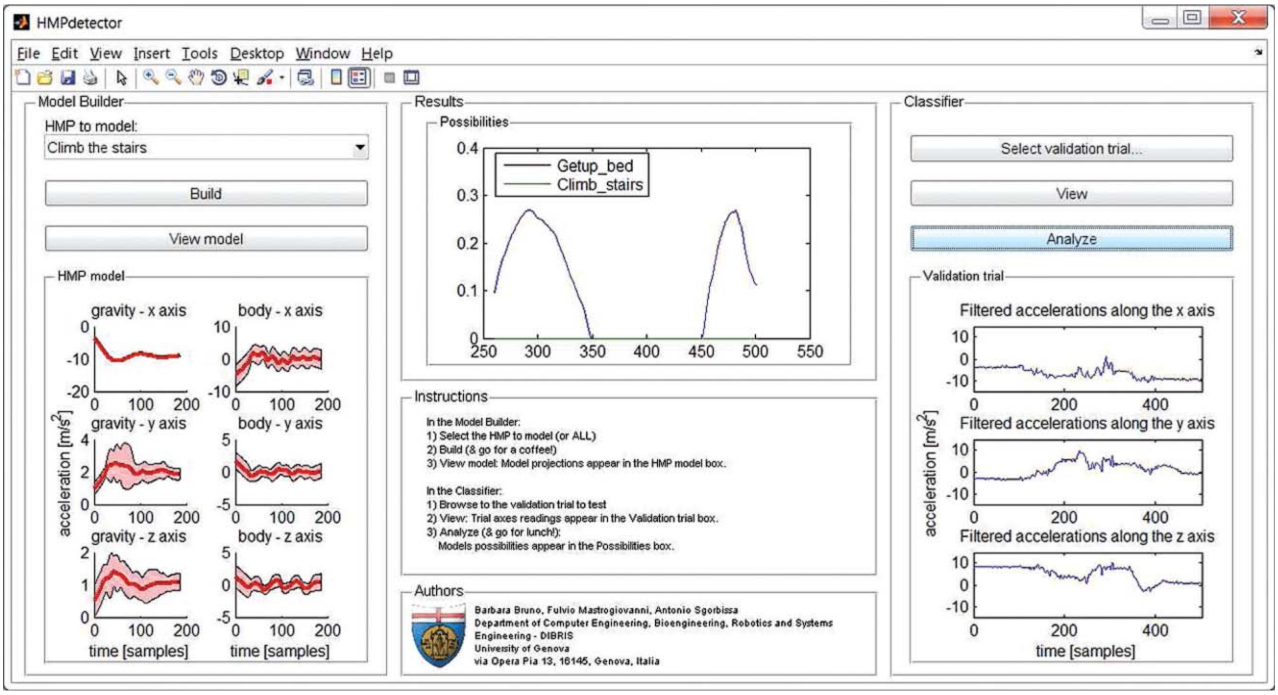


Figure 6. (a) The classifier library is devoted to ranking the similarity between one validation recording and each available HMP model. (b) A sliding window moves on the recording and the same algorithms for data filtering (step 1) and feature extraction used by the models builder are applied over the acceleration data within the window (steps 2 and 3). (c) By computing the Mahalanobis distance between the window features (step 5) and the features of each model it is possible to identify the model with the highest possibility of representing the acceleration data within the window (step 6).



HMP to model:
Climb the stairs

Instructions

In the Model Builder:
1) Select the HMP to model (or ALL)
2) Build (& go for a coffee!)
3) View model: Model projections appear in the HMP model box.

In the Classifier:
1) Browse to the validation trial to test
2) View: Trial axes readings appear in the Validation trial box.
3) Analyze (& go for lunch!):
Models possibilities appear in the Possibilities box.

Authors

Barbara Bruno, Fulvio Mastrogiovanni, Antonio Sgorbissa
Department of Computer Engineering, Bioengineering, Robotics and Systems Engineering - DIBRIS
University of Genova
via Opera Pia 13, 16145, Genova, Italia

Figure 7. The HMPdetector is a GUI for the HMP detection library. On the left-hand side, there is the models builder, allowing for the modeling of any arbitrary set of HMP. On the right-hand side, there is the classifier, allowing for the classification of any recordings. Results appear in the top center area of the GUI.


(a)
(b)
(c)
(d)
(e)
(f)

Figure 8. The six motion types included in HOOD. By considering the orientation of the torso with respect to gravity, we define (a) slow walk, (b) normal walk, and (c) run as vertical-stance activities, and (d) slow crawl, (e) fast crawl, and (f) slither as horizontal-stance activities.

environmental sensors to monitor a larger number of human activities with increased accuracy [13].

PDR

Among the set of daily life activities, those allowing a person to move from one place to another, such as walking and running, are particularly interesting for PDR applications. The PDR is the process of providing a person with a continuous estimate of their location (i.e., traveled distance and heading), on the basis of information acquired by sensors carried around by the person themselves. First responder teams engaged in search and rescue missions need an environment-independent means to always know their position and the path to follow to get out of a dangerous zone [14]. Fitness-oriented devices also keep track of the number of steps taken to estimate the daily energy expenditure of a person [8].

The vast majority of techniques used to estimate the traveled distance are based on accelerometer data, and the simplest but less accurate systems rely on step-detection algorithms with fixed step length [15], whereas similar but more accurate solutions train neural networks to properly estimate the step length [16]. Recent approaches implement the extended Kalman filter and the zero-velocity update strategy to reduce the inertial measurements drift [17]. Heading tracking can be based on the gyroscope data exclusively [18], or the gyroscope and compass data [2], [17]. Commonly adopted techniques for the reduction of the heading drift range from periodic resets based on global positioning system information [19] to systematic corrections obtained by analyzing the street segments in a database of street maps [20]. Other

Figure 9. The six outdoor environments included in HOOD. The motions slow walk, normal walk, and run were performed in all of the environments, while the motions slow crawl, fast crawl, and slither were executed in the grass field environment only. The test environments include three scenarios with even ground: (a) a flat football grass field, (b) an uphill asphalt road (with a constant slope of about 25%), and (c) a wide staircase; three scenarios with rough terrain: (d) a rocky riverbed, (e) a bumpy wooded hill, and (f) a field covered in snow.

methods use environmental tags previously deployed in the building [17], whereas the most complex cooperative positioning techniques rely on sharing information among networked devices carried by individual rescuers [2].

HOOD Data Set

Existing PDR systems are faced with three limitations.

- The vast majority of the state-of-the-art systems focus on vertical-stance motions exclusively, without providing any comprehensive analysis of their performance with respect to different motions, such as crawling or slithering.
- Few systems consider places other than a foot placement for the sensing device (among them, those making use of hand-held or shoulder-mounted devices [15]), and they do not provide any analysis of their performance with respect to different sensor placements, for instance, on the chest or waist.
- There are no systematic studies on the performance of PDR systems under highly diverse terrain conditions, such as on rocks or snow.

The HOOD specifically targets these issues, testing the performance of step detection algorithms with respect to six different motion types (Figure 8), including both vertical- and horizontal-stance motions, six outdoor environments (Figure 9), with different terrain characteristics, and four device placements (Figure 10).

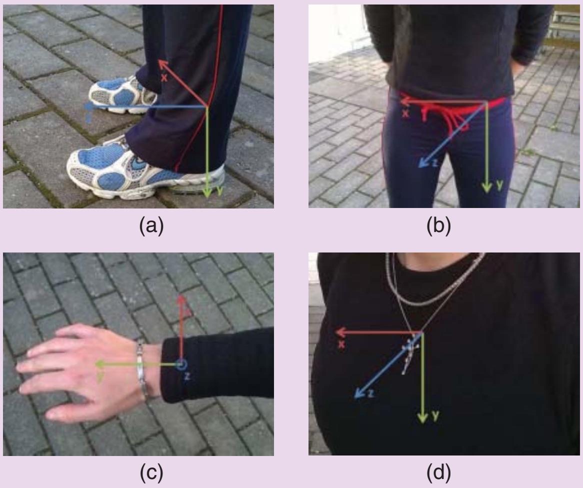


Figure 10. The four sensor placements included in HOOD. (a) Foot, (b) waist, (c) wrist, and (d) chest.

The data set is composed of 168 trials. Each trial records the 6 degree-of-freedom (DOF) acceleration and angular rate values registered during the execution of one motion type in one environment with one sensor placement along one path, and it is annotated with the actual number of steps. The adopted sensing device is the commercial 6 DOF IMU InvenSense MPU6050. The accelerometer is set to have a measurement range of $[-2g; +2g]$ with a sensitivity of 16 b/axis. The gyroscope has a measurement range of $[-250\,^{\circ}/s; +250^{\circ}/s]$ with a sensitivity of 16 b/axis. The sampling frequency was 40 Hz for both sensors (see Table 1).

In order to provide additional information to be used for designing and testing step-detection algorithms, it would be helpful to label accelerometer data with gait events, such as heel-strike and heel-off. However, collecting this information cannot be done in generic outdoor, unstructured environments (i.e., in absence of sophisticated motion capture systems). Under these considerations, we envision the HOOD data set as a tool to validate step detection procedures, which have been designed and carefully tuned in controlled environments and are now ready to be benchmarked in more realistic and challenging scenarios.

Conclusions

The recent technological advances in sensor miniaturization and embedded processing have paved the path for powerful mobile sensing and processing devices, providing new challenges and possibilities to the field of wearable computing. Two research areas are particularly interested by this innovation: healthcare technology applications devoted to analyzing the daily activities of a person to evaluate their general health, and PDR systems, focused on the analysis of the person's movements to keep track of his/her position in dangerous environments. The identification of suitable algorithms and techniques to process wearable sensors data are crucial research challenges that must be overcome for both areas. The tasks to be solved include the creation of generalized and representative models of human motions, the efficient and reliable comparison of run-time data with stored models, and the prompt identification of and reaction to specific events on the basis of their effect on the sensory data.

The HOOD data set could be extended to other outdoor environments (such as sand or debris).

To foster the research on this topic and to provide a public test bench to evaluate and compare the performance of different solutions, we present the WHARF, a freely accessible repository of HAR data sets and software libraries. The repository includes a large data set of acceleration recordings referring to 14 daily activities and a MATLAB library that allows the creation and validation of acceleration models. We also present the HOOD data set to address the specific problem of the identification of steps in acceleration patterns referring to both vertical- and horizontal-stance motions and to highly diverse terrain conditions, which is a crucial task for PDR systems.

Obviously, the current data sets leave much room for improvement. For example, the WHARF data set currently considers a single sensor placement (right wrist) and therefore should be extended to other configurations. Similarly, the HOOD data set could be extended to other outdoor environments (such as sand or debris). This is part of our future research agenda, hopefully with the contribution of other researchers in this field.

References

[1] (2015, July 31). Invensense 9-axis motion tracking devices. [Online]. Available: http://www.invensense.com/products/motion-tracking/9-axis/

[2] J. Rantakokko, J. Rydell, P. Stromback, P. Handel, J. Callmer, D. Tornqvist, F. Gustafsson, M. Jobs, and M. Gruden, "Accurate and reliable soldier and first responder indoor positioning: Multisensor systems and cooperative localization," *IEEE Wireless Commun.*, vol. 18, no. 2, pp. 10–18, 2011.

[3] O. Lara and M. Labrador, "A survey on human activity recognition using wearable sensors," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 3, pp. 1192–1209, 2013.

[4] D. Anguita, A. Ghio, L. Oneto, X. Parra, and J. Reyes-Ortiz, "A public domain dataset for human activity recognition using smartphones," in *Proc. 21st European Symp. Artificial Neural Networks, Computational Intelligence Machine Learning*, 2013. pp. 437–442.

[5] B. Bruno, F. Mastrogiovanni, A. Sgorbissa, T. Vernazza, and R. Zaccaria, "Human motion modelling and recognition: A computational approach," in *Proc. IEEE Int. Conf. Automation Science Engineering*, Seoul, Korea, Aug. 2012, pp. 156–161.

[6] B. Bruno, F. Mastrogiovanni, A. Sgorbissa, T. Vernazza, and R. Zaccaria, "Analysis of human behavior recognition algorithms based on acceleration data," in *Proc. IEEE International Conf. Robotics Automation*, Karlsruhe, Germany, May 2013, pp. 1602–1607.

[7] B. Bruno, F. Mastrogiovanni, and A. Sgorbissa, "A public domain dataset for ADL recognition using wrist-placed accelerometers," in *Proc. IEEE Int. Symp. Robot Human Interactive Communication*, Edinburgh, U.K., Aug. 2014, pp. 738–743.

[8] (2015, July 31). Fitbit One. [Online]. Available: http://www.fitbit.com/us/one

[9] (2015, July 31). Moves—Activity diary for iPhone and Android. [Online]. Available: https://www.moves-app.com/

[10] J. Aggarwal and M. Ryoo, "Human activity analysis: A review," *ACM Comput. Surveys*, vol. 43, no. 3, pp. 16:1–16:43, 2011.

[11] I. Bisio, F. Lavagetto, M. Marchese, and A. Sciarrone, "A smartphone-centric platform for remote health monitoring of heart failure," *Int. J. Commun. Syst.*, vol. 28, no. 11, pp. 1753–1771, 2015.

[12] A. Scalmato, A. Sgorbissa, and R. Zaccaria, "Describing and recognizing patterns of events in smart environments with description logic," *IEEE Trans. Cybern.*, vol. 43, no. 6, pp. 1882–1897, 2013.

[13] B. Bruno, J. Grosinger, F. Mastrogiovanni, F. Pecora, A. Saffiotti, S. Sathyakeerthy, and A. Sgorbissa, "Multi-modal sensing for human activity recognition," in *Proc. IEEE Int. Symp. Robot Human Interactive Communication*, Aug. 2015.

[14] M. Baglietto, A. Sgorbissa, D. Verda, and R. Zaccaria, "Human navigation and mapping with a 6DOF IMU and a laser scanner," *Robot. Auton. Syst.*, vol. 59, no. 12, pp. 1060–1069, 2011.

[15] C. Randell, C. Djiallis, and H. Muller, "Personal position measurement using dead reckoning," in *Proc. IEEE Int. Symp. Wearable Computers*, Oct. 2003, pp. 166–173.

[16] S. Beauregard and H. Haas, "Pedestrian dead reckoning: A basis for personal positioning," in *Proc. Workshop Positioning, Navigation Communication*, Mar. 2006, pp. 27–35.

[17] A. J. Ruiz, F. Granja, J. P. Honorato, and J. Rosas, "Accurate pedestrian indoor navigation by tightly coupling foot-mounted IMU and RFID measurements," *IEEE Trans. Instrum. Meas.*, vol. 61, no. 1, pp. 178–189, 2012.

[18] L. Ojeda and J. Borenstein, "Non-GPS navigation for security personnel and first responders," *J. Navig.*, vol. 60, no. 3, pp. 391–407, 2007.

[19] E. P. Herrera, H. Kaufmann, J. Secue, R. Quiros, and G. Fabregat, "Improving data fusion in personal positioning systems for outdoor environments," *Inform. Fusion*, vol. 14, no. 1, pp. 45–56, 2013.

[20] P. Aggarwal, D. Thomas, L. Ojeda, and J. Borenstein, "Map matching and heuristic elimination of gyro drift for personal navigation systems in GPS-denied conditions," *Meas. Sci. Technol.*, vol. 22, no. 2, 025205, pp. 1–12, 2011.

Barbara Bruno, University of Genoa, Genova, Italy. E-mail: barbara.bruno@unige.it.

Fulvio Mastrogiovanni, University of Genoa, Genova, Italy. E-mail: fulvio.mastrogiovanni@unige.it.

Antonio Sgorbissa, University of Genoa, Genova, Italy. E-mail: antonio.sgorbissa@unige.it.

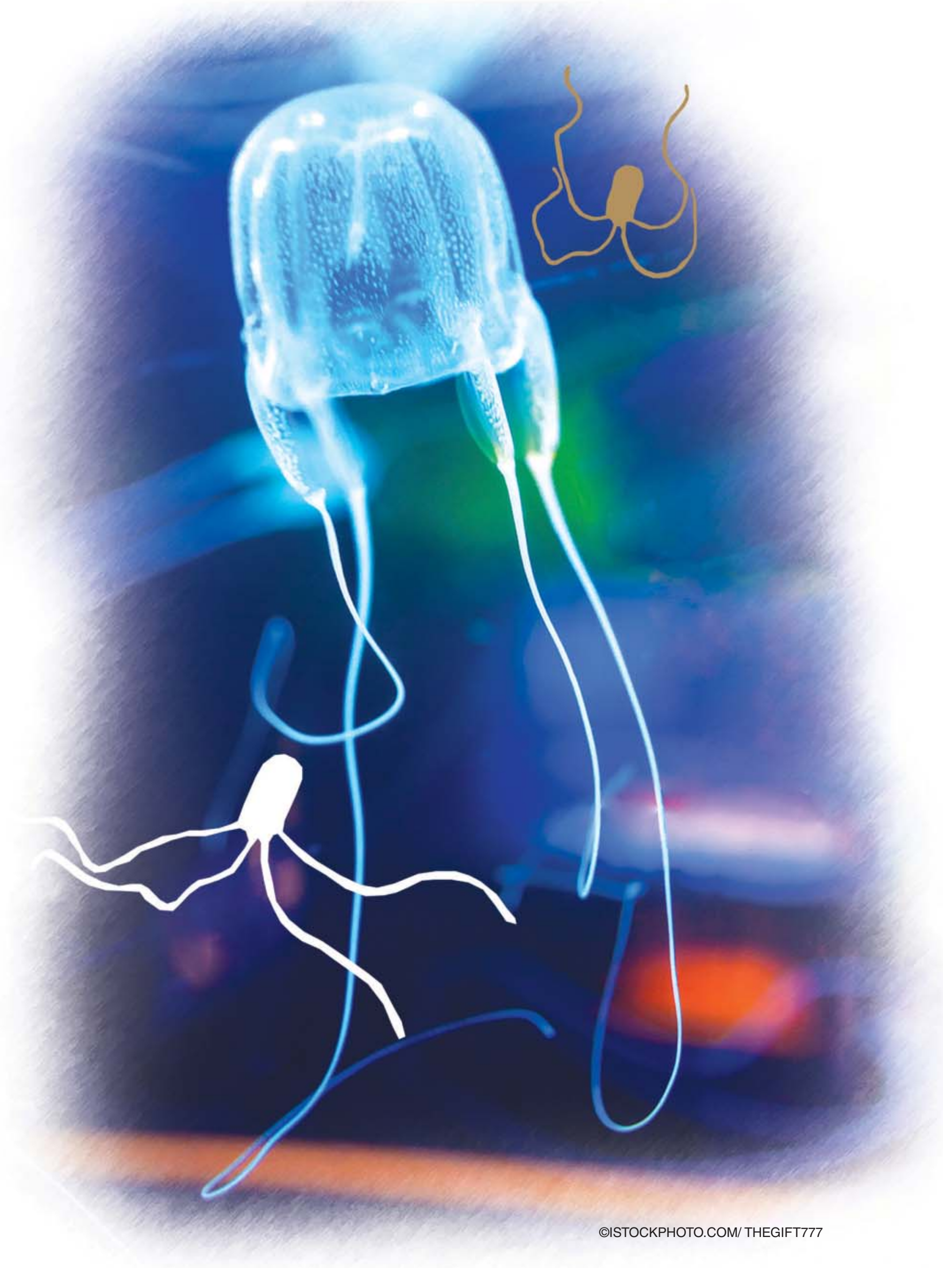
Deformation in Soft-Matter Robotics

A Categorization and Quantitative Characterization

By Liyu Wang and Fumiya Iida

There has been an increasing interest in soft-robotics research in recent years, and this is reflected in a number of reviews of the topic from different perspectives. For example, in what was probably the first review article using the term *soft robotics*, [1] focused on actuators, while [2] reviewed the topic by concentrating on fabrication techniques. Three more reviews addressed control [3], biomimetics [4], and materials [5], respectively. Furthermore, in the first issue of *Soft Robotics,* there are seven review articles from active researchers in the field. More recently, another review [6] and an edited book [7] have tried to cover a number of topics, including design, fabrication, and control as well as sensors and actuators. All of these efforts have helped draw attention to soft robotics and summarized some of the most recent studies.

However, there has not been a clear definition for soft robotics in the literature. On the one hand, some researchers in the community embrace a broad definition, where *soft* may refer to both the structural compliance generated by a geometrical arrangement of hard materials and the inherent compliance of



©ISTOCKPHOTO.COM/ THEGIFT777

Digital Object Identifier 10.1109/MRA.2015.2448277
Date of publication: 11 September 2015

materials. This can be seen in the statement from the recent European Future and Emerging Technologies Open Coordination Action, RoboSoft, and in [7]. On the other hand, some researchers appear to be more comfortable with a more specific definition that focuses on materials with a relatively low modulus, as exemplified by the definition in [6]. To date, most reviews simply put together case studies that fell into ambiguous classifications of softness based on the authors' individual experiences and research topics.

Shape transformation or reconfiguration is significant and involves one or more of the aforementioned types of deformation.

Despite the lack of a clear definition, researchers in the community have usually assumed that nontraditional materials are involved in the design and control of robotic systems. The primary considerations are what materials should be used or explored and how to use them for new robotic technologies. For example, it is important to select materials suitable for certain robotic functions and to identify physical properties that match the scale of the robotic system as well as the constraints imposed by actuators. In this context, there is an urgent need to give a definition that can be accepted by the whole community and to provide measures to quantitatively compare the state of the art and to guide future research.

We choose to ground our research in soft matter, which is a defined class of materials, and we use *soft-matter robotics* instead of *soft robotics* throughout the rest of the article. Soft matter includes fluids, polymers, colloids, granular materials, and biological materials. The common feature of soft matter is that it consists of large molecules or assemblies of molecules that move collectively, and, as a result, it gives large, slow, and nonlinear response to small forces [8]. The use of soft matter in machine design is not new. Thermoplastic polymers, e.g., have been used as supporting structures or kinematic linkages as cheaper alternatives to metals; rubber or sponge has been used to cover surfaces to absorb impact or maximize contact surfaces [9]; and magnetorheological fluid has been used for dampers or brakes in cars. There has also been an approximately 30-year history of using soft matter for robotics research, but it never had a major influence until recently. The increase in interest in the last five years is probably due to the following reasons:

- Soft matter has been established as a field in material science since the 1990s.
- A large amount of new soft matter has been synthesized and made commercially available.
- Diverse fabrication techniques for soft matter have been made accessible.
- An increasing amount of work demonstrating the use of soft matter in robotics has been published in high-profile journals.
- Researchers generally agree that soft-matter-based technologies should be used in robotic applications in the future as they are intrinsically cheaper, safer, and more adaptive in complex task environments as compared with the conventional rigid systems.

The motivation of the soft-matter robotics research is, on the one hand, to find applications for material scientists and, on the other hand, to tackle challenges in robustness and versatility for roboticists. The mechanisms and behaviors of soft matter are sometimes associated with those observed in animals, and, thus, the field also attracts attention from biologists.

We define soft-matter robotics as robotics that studies how deformation of soft matter can be exploited or controlled to achieve robotic functions. As will be shown in this article, this is where the novelty and excitement of soft-matter robotics lie. Robots are complex systems and usually contain many components, such as mechanical structures, actuators, and sometimes sensors. There have been many studies on developing composite soft materials as mechanosensing components or on using soft matter for actuation. However, the number of studies has grown so large that sensor and actuator research has become a separate field [10]. We focus here on topics related to the exploitation or control of the deformation of soft matter in the mechanical structures for systems at a scale above 1 mm. As a result, we will not cover sensors and will only include work on actuators where they also function as mechanical structures. In addition, while there have been technologies proposed that use polymer-based adhesives as foot structures for robotic climbing [11], they are not covered here because adhesion is not the topic of interest as far as this article is concerned. There have also been many microrobotics studies on developing passive microgrippers fabricated with a soft matter, and they are not included here either.

We review the field of soft-matter robotics by categorizing previous studies into four types of deformation behaviors, i.e., elongation/shortening, bending, flowing, and transformation. With an overview of some of the representative work related to each type by researchers from both robotics and material science, we summarize which soft matter is used for which deformation behavior, how it is actuated, and what function each type of deformation could bring to a robotic system. By proposing measures such as the product of modulus and structural area as well as breaking strain of the material, we hope to give a quantitative characterization of various methods and studies.

Deformation for Functions

Elongation/Shortening

Elongation/shortening is the simplest type of deformation in solid structures. Several representative case studies where elongation/shortening behavior of soft-matter structures may be used in robotics are shown in Figure 1. From these case studies, we see that this type of deformation has three functions.

First, elongation/shortening is used in the actuator component for basic linear actuation. Typically, McKibben-type artificial muscle actuators use pneumatic pressure as an energy source and rely on the stretch of a bladder made from elastomer such as rubber to achieve actuation. Dielectric elastomer artificial muscles use electricity as an energy source. Despite their basic linear deformation, they may be configured in various ways to achieve more complex patterns of actuation [12]. Review articles can be found on pneumatic artificial muscles [13], dielectric elastomer artificial muscles [12], and their use in robotic systems [1].

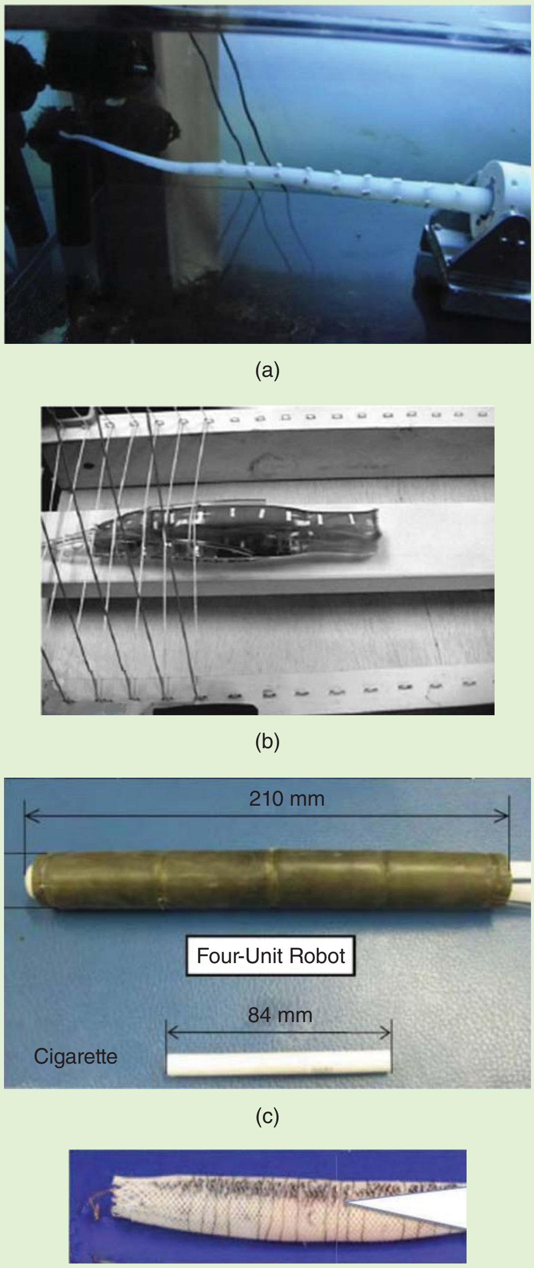
Second, elongation/shortening is used for reaching, as shown in Figure 1(a). In this case, the use of silicone-based elastomer in the mechanical structure has the potential to be controlled to extend the reaching range due to the large yield strain of the material. In the case study, the elastomer arm contains many segments, and each segment is actuated by shape-memory alloys.

Third, elongation/shortening has been used for peristaltic locomotion since the last decade, as shown in Figure 1(b)–(e). In peristaltic locomotion, the radical movement is symmetric, and forward motion relies on wave propagation through the contraction and extension of body segments along the axis of movement. For this purpose, both elastomers and thermoplastic polymers were used with different actuation methods. In Figure 1(b), a silicone-based elastomer body containing fluid is actuated by tension cords so that the fluid may be pushed into the front or the rear of the body for extension. In Figure 1(c), pneumatic artificial muscles are directly used as body segments. In Figure 1(d) and (e), the bodies were made from meshes of thermoplastic polymers. They are actuated by shape-memory alloy or cables with a motor, respectively, to achieve contraction and extension of different segments.

Bending

Bending represents the most common deformation behavior in a solid structure. According to Timoshenko beam theory, bending is a result of a rotational effect from a difference in the elongation and shortening in the layers of a structure and a shear effect of the material within the structure. For approximately 20 years, bending deformation of soft matter has been used to design flexure hinges and has been given many functions, from gripping to different kinds of locomotion.

Robotic systems with flexure hinges made from soft matter are shown in Figure 2. A variety of thermoplastics or elastomers were used as flexure hinges in fingered gripping, high-frequency flapping flying, and legged locomotion. These flexure hinges are either passive or actuated by cables and motors [16], [17], piezoelectric actuators [18], or shape-memory alloys [19]. Flexure hinges could introduce compliance to a robot during physical interactions, which hinge joints are less capable of. The use of polymers to replace metals for flexure hinges not only reduces cost and weight but also allows larger bending angles due to the larger yield strain of the polymers. In addition, polymers, such as thermoplastics, have viscoelastic property, which is thought



(a)

(b)

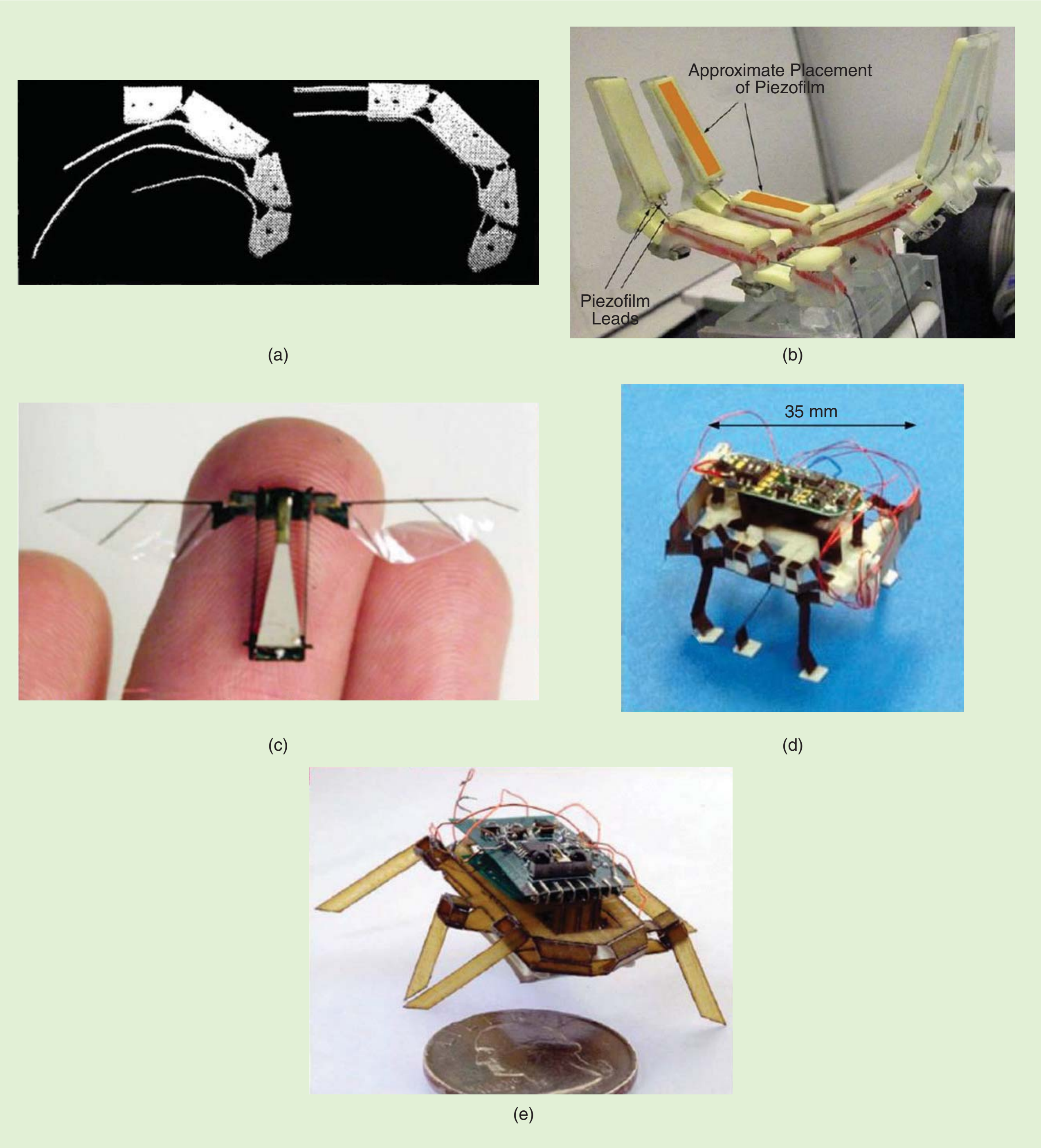
(c)

(d)

Figure 1. Elongation and shortening in soft-matter robots: (a) a silicone-based elastomer actuated by shape-memory alloys for reaching [14], (b) a silicone-based elastomer actuated by tension cords and internal fluid for peristaltic locomotion [49], (c) pneumatic artificial muscles as body segments for peristaltic locomotion [50], and (d) thermoplastic meshes actuated by shape-memory alloys for peristaltic locomotion [15].

to be important as passive damping for gripping [17] and legged locomotion. These robotic systems with flexure hinges made from polymers still fall into the paradigm of kinematic machines. We will show a shift of paradigm toward continuum machines based on several functions, from gripping to legged, rolling, or aquatic locomotion, all raised from bending behavior.

Various robotic fingered, handed, or armed grippers made from soft matter are shown in Figure 3. The common feature of these grippers is that their gripping and releasing fully relies on the bending of the entire finger(s), hand, or arm. Some handed grippers and armed grippers used elastomers that are bent by air-based energy sources [20] or shape-memory alloys [14]. Some fingered grippers used



(a) (b) (c) (d) (e)

Figure 2. The bending of polymeric flexures in soft-matter robots: (a) polytetrafluoroethylene flexure in a robotic finger actuated by cables and motors [16], (b) polyurethane flexure in robotic fingers actuated by cables and a motor [17], (c) polyimide flexure in the thorax of a millimeter-scale robotic fly actuated by a piezoceramic actuator [18], (d) carbon-fiber reinforced polyester flexures actuated by piezoelectric actuators in a centimeter-scale hexapedal robot [51], and (e) polyester flexures actuated by shape-memory alloys in a centimeter-scale hexapedal robot [19].

thermoplastics actuated by a combination of cables and motors. Many more fingered or handed grippers used active soft matter, such as electroactive polymers (EAPs) [21], shape-memory polymers, or hydrogels [22], which can self-actuate with energy sources from electrical current, temperature change, hydration, or even visible light. These grippers have mainly served as proof of concept that bending of some soft matter could be used in robotics, and none of the case studies has described how the soft-matter-based grippers surpass the capabilities of existing grippers.

Figure 4 shows how bending of soft matter could generate legged and crawling locomotion. Elastomers, thermoplastics, and polymer gels have been used. From the perspective of the actuation method, elastomers may be actuated by cables and motors [Figure 4(a)], or valves and air pumps [Figure 4(b)], or even biological material cardiomyocytes [Figure 4(e)]; thermoplastics may be actuated by cables and motors [Figure 4(c)] or shape-memory alloys [Figure 4(d)]; and EAP gel is self-actuated to electrical stimuli [Figure 4(f) and (g)]. From the perspective of gait pattern, using the elastic energy from bending from the rear of the body and asymmetric friction between the front and the rear of the body is the most common way to achieve forward movement [Figure 4(a), (e), and (g)]. Alternatively, quadrupedal and bipedal walking could be achieved by bending of the leg(s) from different sides of a robot [Figure 4(c) and (d)]. With networked chambers of silicone-based elastomer, a robot could select chambers in the body to bend and achieve multiple gait patterns, such as walking,

The common feature of these grippers is that their gripping and releasing fully relies on the bending of the entire finger(s), hand, or arm.

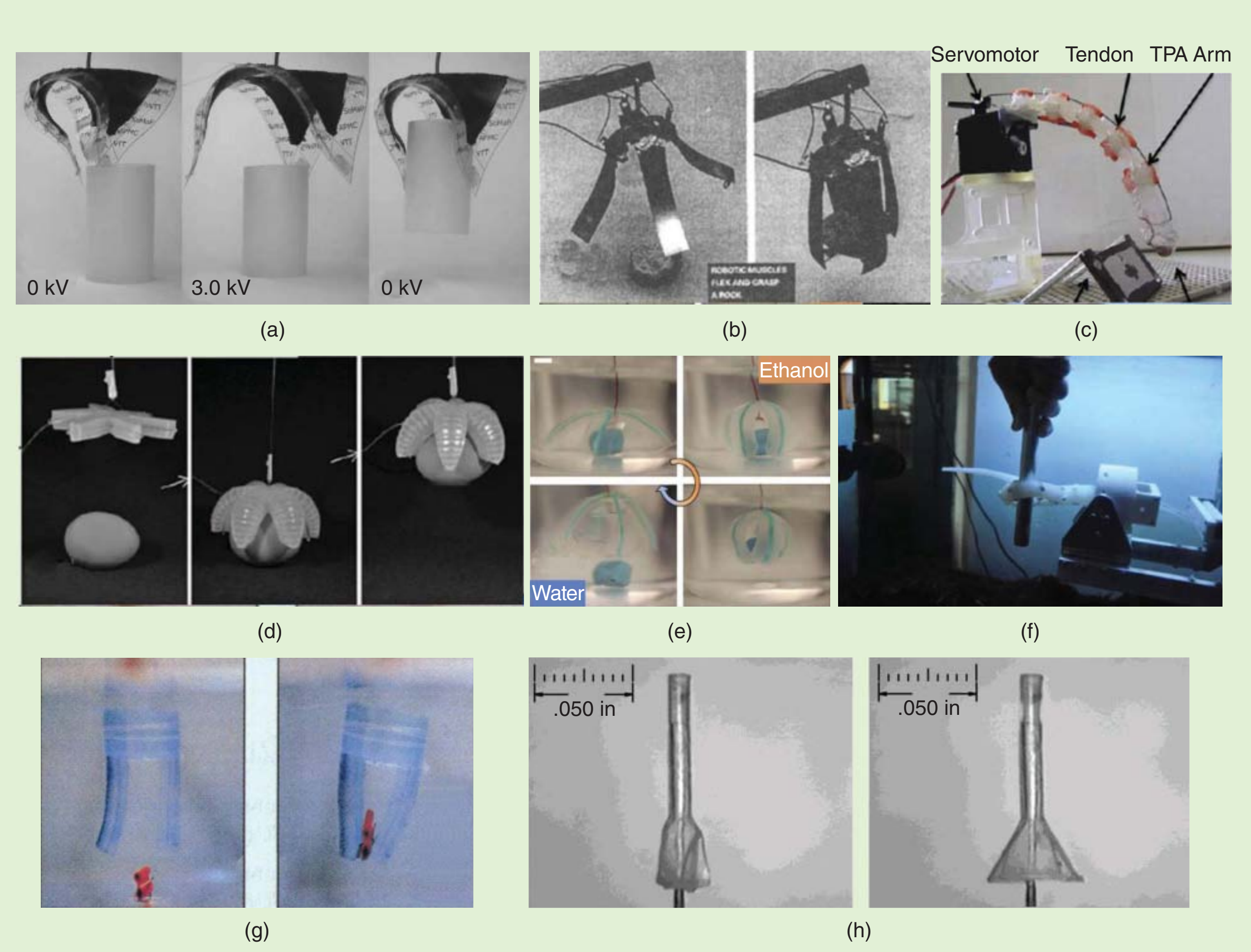


Figure 3. The bending of polymers for fingered gripping in soft-matter robots: (a) polyethylene terephthalate layered with prestretched dieletric elastomer actuators as a hand [52], (b) EAPs directly used as fingers [21], (c) an ethylene vinyl acetate (EVA)-based thermoplastic adhesive (TPA) finger actuated by cables and a servo motor (work by Liyu Wang and Fumiya Iida), (d) networked chambers of silicone-based elastomer powered by an air pump as a hand [20], (e) water-sensitive polyelectrolyte hydrogel with asymmetrically distributed cupric ions in ethanol used as a hand [53], (f) silicone-based elastomer actuated by shape-memory alloys for grasping [14], (g) bilayers of temperature-sensitive N-isopropylacrylamide gel and acrylamide gel as fingers [22], and (h) temperature-sensitive shape-memory polymer actuators directly used as an umbrella-shaped interventional ischemic stroke device [54].

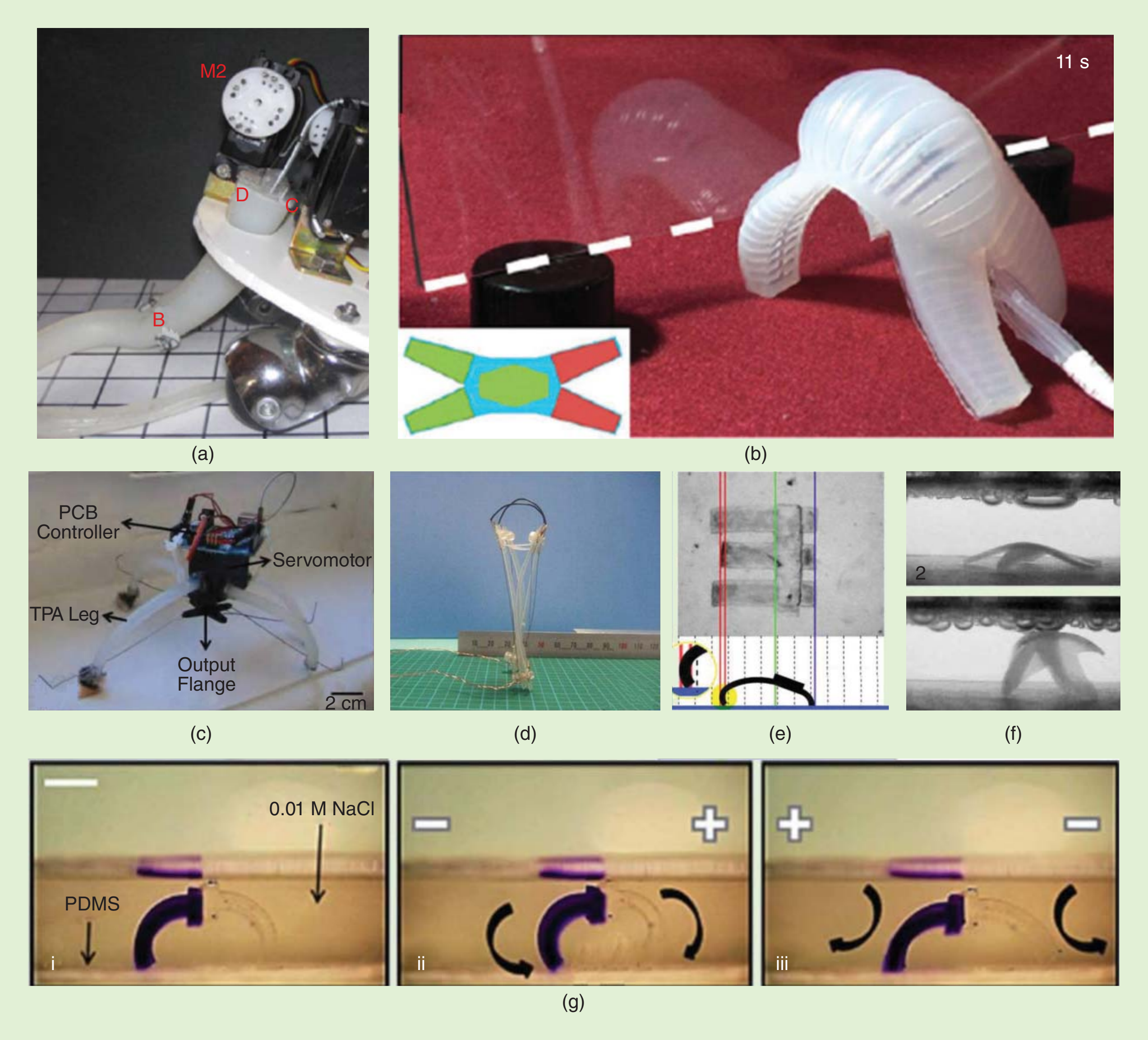


Figure 4. The bending of polymers for legged or limbless locomotion in soft-matter robots: (a) silicone-based elastomer actuated by cables (C and D) and a servo motor (M2) in legged locomotion [55], (b) networked chambers of silicone-based elastomer actuated by valves and an air pump in multigaited legged locomotion [23], (c) locomotion with legs made from EVA-based TPA actuated by cables and one servo motor (work by Liyu Wang), (d) bipedal locomotion with legs made from polyethylene actuated by shape-memory alloys [24], (e) crawling locomotion with polydimethylsiloxane actuated by cardiomyocytes [25], and (f) and (g) crawling locomotion with ionic stimuli-responsive hydrogel [61], [62].

undulation, or crawling [Figure 4(b)]. Overall, a body size range of several hundred micrometers to tens of centimeters has been demonstrated with low speed. These robots have shown that the controller of forward quadruped locomotion could be simplified to backward and forward rotation of a single servo motor, or gait patterns could be switched to pass a narrow gap [23].

Case studies where the bending of soft matter was used in rolling locomotion are shown in Figure 5. There are three mechanisms to generate rolling from bending deformation. The first one is caterpillar inspired, where the body curls into a wheel to generate rolling momentum from elastic energy stored in the body [Figure 5(a)]. The second mechanism is based on a body that consists of networked chambers, where selected inflated chambers push the rest of the body to move [Figure 5(b) and (c)]. The third one is based on Sarrus linkage [Figure 5(d)]. All case studies used elastomers due to the large yield strain of the materials needed for actuation stroke in all the mechanisms. Shape-memory alloys, valves and air pumps, or magnets and magnetic fields were used for actuation.

Aquatic locomotion enabled by bending of polymers is shown in Figure 6. The locomotion patterns are mostly inspired by those present in aquatic animals, such as tailed fish, jellyfish, octopuses, or manta rays, whose movement relies on bending of body parts, such as the tail, the bell, the arms, or the fins, respectively. To mimic the undulating swimming pattern of tailed fish, polymer gels or elastomer were used to make the tails [Figure 6(a) and (c)]. The tails were actuated by biological muscles or cables and a servo motor, and, in the case of the EAP gel, it was self-actuated. For those mimicking jellyfish locomotion, elastomer was used to construct the bell, and the bell was actuated by cardiomyocytes or shape-memory alloy composites [Figure 6(d) and (e)]. For octopus-like locomotion, elastomer or ionic stimuli-responsive polymer gel was used to make the arms, which were actuated by cables and servo motors or self-actuated, respectively [Figure 6(f) and (g)]. These aquatic soft-matter robots present a different locomotion strategy from rear propulsion, which is used in most underwater unmanned vehicles, and they rely much more on the physical interaction between the soft body parts and the environment.

Flowing

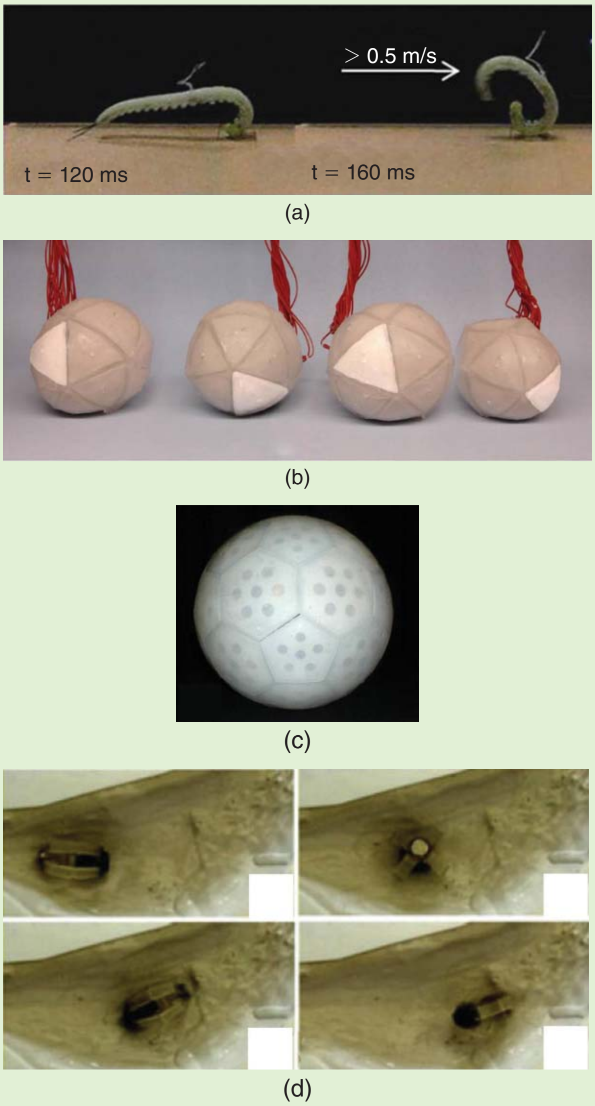
Flowing of soft matter can result in deformation in a rheological form. The use of flowing of soft matter in robotics has been around for more than 30 years [34], [35]. Some case studies using smart fluids and granular materials, such as plastics, magnetic powder, or coffee beans, are shown in Figure 7. These case studies show that the flowing of these materials has mostly been used to grip solid objects with irregular contours based on shape conformation. The forces that cause the soft matter to flow are contact forces during physical interaction with the object and the gravitational forces of the soft matter itself. Thus, the flowing behavior is passive.

There are some distinctive features when flowing deformation is exploited in robotic gripping. For example, all grippers need some kind of encapsulation so that the fluids or granular materials are contained, and, usually, the encapsulation is made from elastomers with a low modulus. Another interesting and important feature is that all these grippers need phase transition of fluids or granular materials into a rigid solid after flowing and shape conformation so that the shape of the gripper can be maintained and sufficient gripping forces can be provided. Upon releasing the object, phase transition back to fluids or granular materials is needed. Depending on the soft matter used, the method to achieve phase transition varies. For smart fluids, electrical current or a magnetic field is needed, and for granular materials, a magnetic field or air pump is used.

Transformation/Reconfiguration

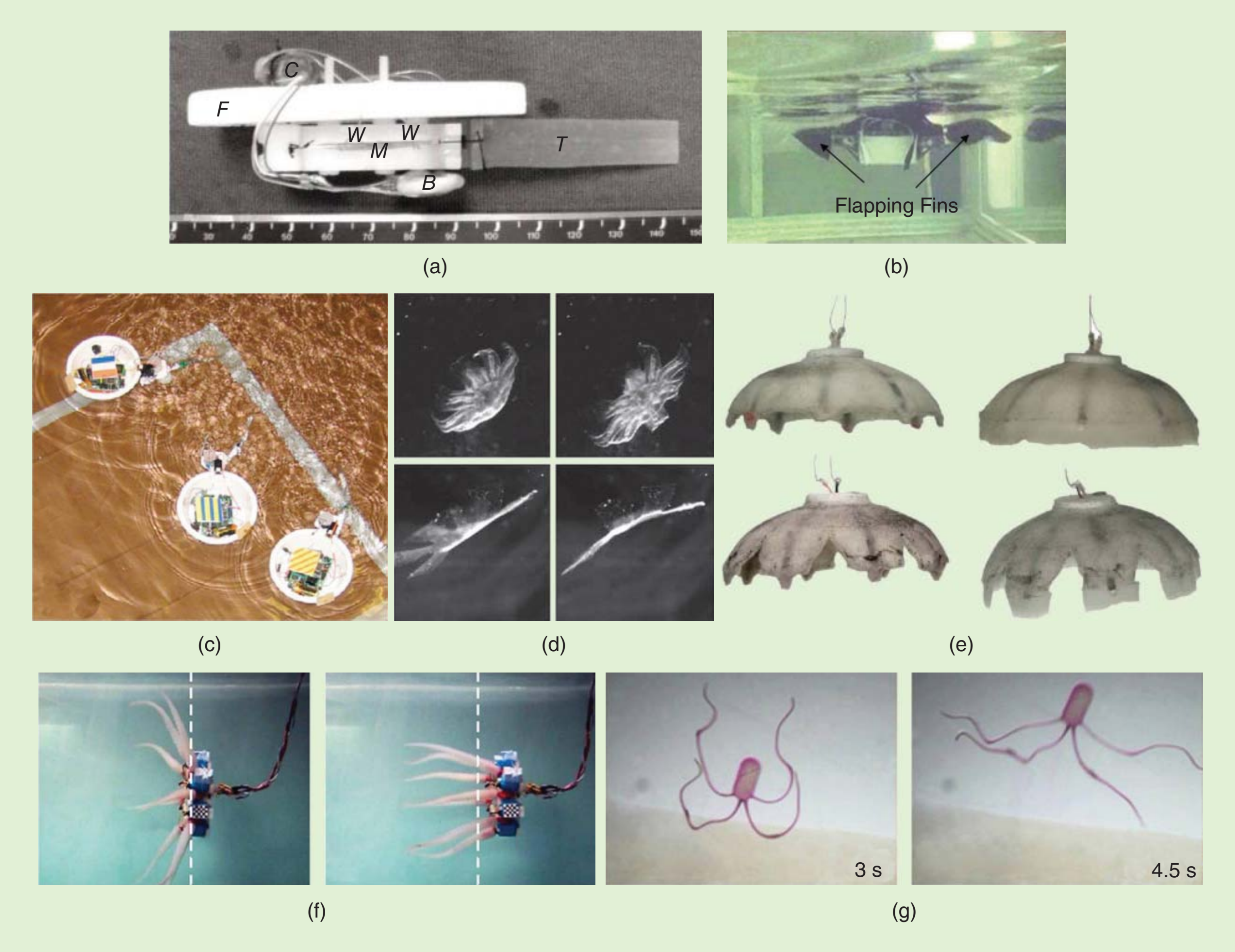
Shape transformation or reconfiguration is significant and involves one or more of the aforementioned types of deformation. Due to the presence of multiple types of deformation, components and processes for controlled handling of soft matter are technically needed by a robotic system. However, it is important to differentiate transformation/reconfiguration from construction or fabrication because the purpose of changing the shape of a part of the robotic system is to achieve certain robotic functions.

Two such systems, based on spray foam and thermoplastics, respectively, are shown in Figure 8. In Figure 8(a), spray foam was equipped in a robot within a team that was remotely controlled to form structures to join other team member robots into a bigger and more complex robot. A motor was used to spray the prepackaged foam from a liquid phase, and more motors were used for moving the spraying robot around. The bigger robot functioned as a legged robot or a crawling robot [38]. In Figure 8(b), ethylene vinyl acetate (EVA)-based thermoplastic adhesive (TPA) was automatically heated and handled by a robot arm and shaped into a scoop and a gripper through a



(a)
(b)
(c)
(d)

Figure 5. The bending of polymers for rolling locomotion in soft-matter robots: (a) wheel-shaped rolling with silicone-based elastomer actuated by shape-memory alloys [26], (b) ball-shaped rolling with networked chambers of silicone-based elastomer actuated by valves and an air pump [27], (c) ball-shaped rolling with networked bladders of silicone-based elastomer and ABS plastic actuated by valves and an air pump [56], and (d) a capsule endoscope made from polyurethane elastomer rolling and actuated by magnets and magnetic fields [28].



(a) (b)

(c) (d) (e)

(f) (g)

Figure 6. The bending of polymers for aquatic locomotion in soft-matter robots: (a) fish-like swimming with a polyoxymethylene body and a silicone-based elastomer tail actuated by a pair of frog muscles [29], (b) manta-ray-like swimming with polypropylene elastomer fins [57], (c) fish-like swimming with a tail containing a hydrogel notochord actuated by a servo motor [30], (d) jellyfish-like swimming with polydimethylsiloxane actuated by cardiomyocytes [31], (e) jellyfish-like swimming with silicone-based elastomer actuated by a shape-memory alloy composite [32], (f) octopus-like swimming with silicone-based elastomer actuated by cables and servo motors [33], and (g) octopus-like swimming with ionic stimuli-responsive hydrogel [58].

process similar to the fused filament fabrication technique in rapid prototyping. A servo motor was used for extruding the fluidic TPA, while the shaping process was actuated by joint motors in the robot arm. The scoop and the gripper were used as reconfigurable end effectors to pick up and place liquid and lightweight solid objects across three orders of magnitude [39].

Bending represents the most common deformation behavior in a solid structure.

There are three common features when transformation of soft matter is used for robotic functions. First, phase transition is required. Different from flowing deformation, phase transition in transformation does not necessarily need to be bidirectional. The transition process is usually natural and passive as shapes are being formed. Second, adhesion of soft matter is needed because formed passive structures need to be integrated into the robotic system with a force endurance capacity. Third, the transformation or reconfiguration process has been so far controlled in a similar way in additive fabrication.

The types of deformation, soft matter, and robotic functions in representative studies to date are shown in Table 1. Based on the table, bending is the most popular type of deformation being used for a large range of robotic functions, from gripping to terrestrial and aquatic locomotion. Elongation/shortening has only been used for reaching and peristaltic locomotion, transformation for legged and crawling locomotion and gripping, and flowing only for gripping. Concerning materials to be used, polymers are by far the most common category of soft matter used in robotic systems. Elastomers, including silicone-based elastomer and polyurethane, are used for the largest number of functions. Thermoplastics, such as EVA, polyester, polyethylene, and polytetrafluoroethylene, are popular materials for hinged flexures, gripping, and legged or crawling locomotion. EAPs are also widely used. Besides polymers, granular material and smart fluids are also used, but only for robotic gripping. Colloids are less studied, except for foam.

Mechanical Characterization of Soft-Matter Robots

So far, we have discussed a qualitative taxonomy for the behaviors of soft-matter robots. This section considers some quantitative methods for systematic investigations. Quantification is particularly important given the diversity of soft matter used and the variety of behaviors. With the pertinent quantification methods, we are not only able to develop better robots but also to clarify research challenges in the field. Soft-matter robotics research has usually been quantified with respect to the mechanical characteristics of soft matter, such as the modulus and the breaking strain of solids. However, as will become clearer later, quantification based on material properties is not necessarily the best solution for

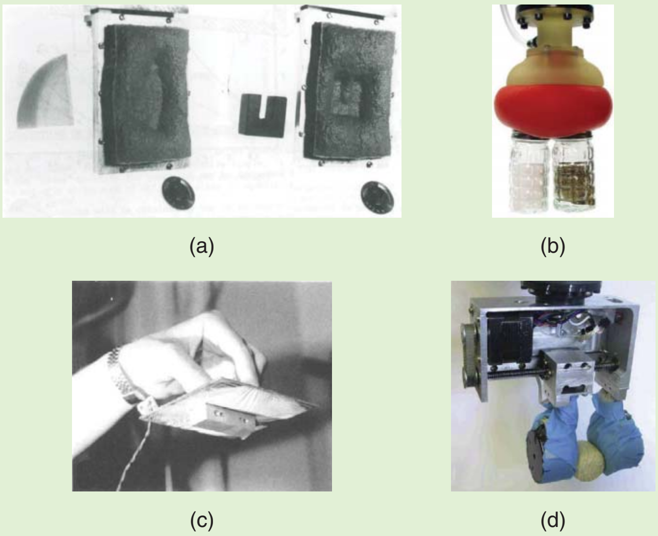

(a) (b)
(c) (d)

Figure 7. The flowing of soft matter for robotic grippers: (a) passive shape conformation through the flowing of plastic granular material in a rubber bag, with phase transition by an air pump [36], (b) passive shape conformation through the flowing of coffee bean granular material in a rubber bag, with phase transition by an air pump [59], (c) passive shape confoçrmation through the flowing of encapsulated electrorheological fluid, with phase transition by electrical current [60], and (d) passive shape conformation through the flowing of encapsulated magnetorheological fluid, with phase transition by a magnetic field [37].

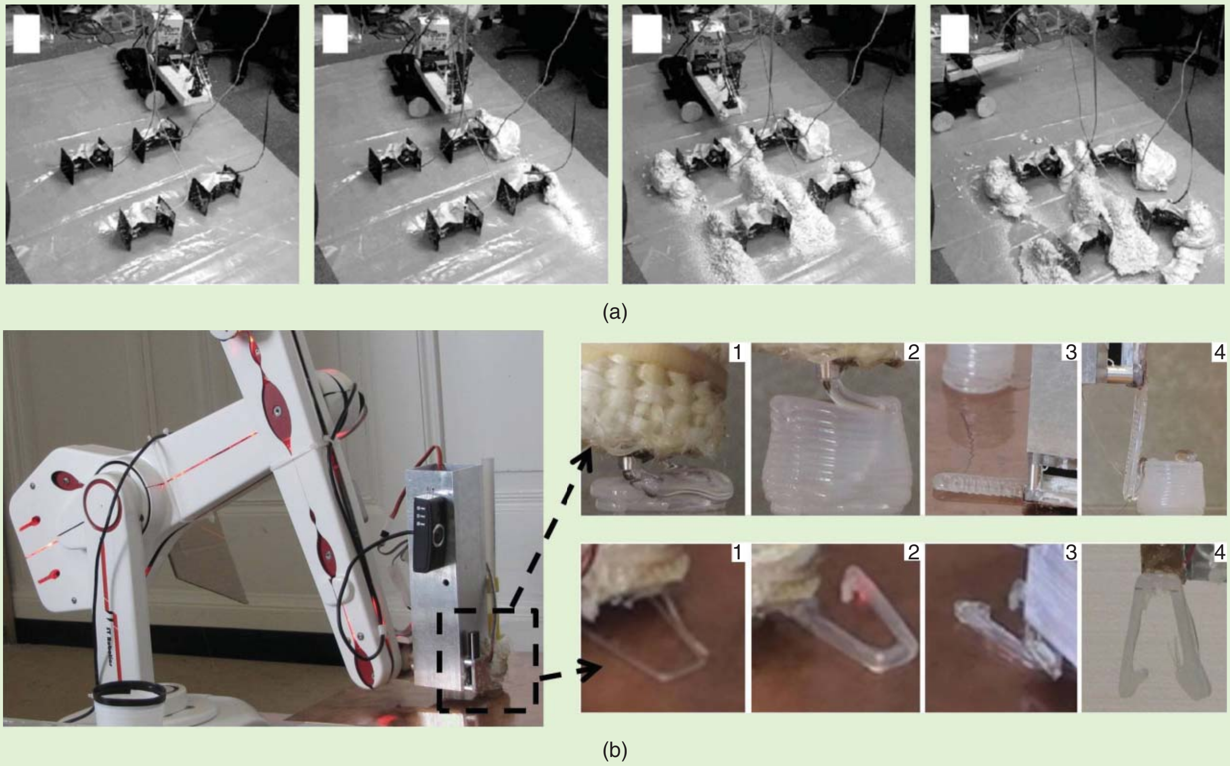


Figure 8. Transformation/reconfiguration in soft-matter robotics: (a) a legged robot formed by a team of robotic modules with foam through phase transition from liquid to solid induced by natural exposure to air [38] and (b) a reconfigurable end effector formed by a robotic arm with TPA through phase transition from fluid to solid induced by natural cooling [39].

Table 1. Types of deformation, soft matter, and robotic functions.

Deformation	Soft Matter		Robotic Function
Elongation/Shortening	Thermosetting polymer	Silicone-based elastomer	Reaching, peristaltic locomotion
	Thermoplastic polymer	PEEK mesh	Peristaltic locomotion
	EAPs	–	Artificial muscle actuators
Bending	Thermosetting polymer	Silicone-based elastomer	Gripping, legged and crawling locomotion, rolling locomotion, tailed swimming, jellyfish-like locomotion, octopus-like locomotion
		Polyurethane elastomer	Hinged flexure, rolling locomotion
		Polyimide plastic	Hinged flexure
	Thermoplastic polymer	EVA, polyester, polyethylene, polytetrafluoroethylene	Hinged flexure; gripping, legged, and crawling locomotion
	EAPs	–	Gripping, crawling locomotion, tailed swimming, octopus-like locomotion
	Polymer hydrogel (not EAP gel)	–	Gripping, tailed swimming
Flowing (macro encapsulation needed)	Granular materials	Iron powder, plastic powder, coffee beans	Gripping
	Smart fluids	Electrorheological fluid	Gripping
		Magnetorheological fluid	Gripping
Transformation/Reconfiguration	Thermoplastic polymer	EVA	Dragline-forming locomotion, gripping, and scooping
	Colloid	Foam	Legged locomotion

soft-matter robots, and the size of soft-matter structures should be taken into account.

We specifically consider case studies based on polymers given that they are the largest group of soft matter used to date. The source data were collected from previous publications and are shown in Table 2, in which we summarize the types of materials used, various mechanical properties, actuator types, and the types of behavioral functions demonstrated. We focus on mechanical properties such as modulus, breaking strain, and tensile strength, and we assume the materials are used under similar conditions, such as room temperature, ambient humidity, and age of materials. Since none of the robots are completely made out of soft matter, we also indicate the size of the parts that are made from soft matter. In addition to the mechanical properties and size, the actuation characteristic is another important source of data necessary for the quantification of behavioral functions. Here, we specifically consider three types of actuation characteristics, i.e., mass, maximal output force, and maximal actuator stroke, which were obtained from [45] for six types of actuators, as listed in Table 3.

The phase transition process in shape transformation or reconfiguration is usually natural and passive as shapes are being formed.

Tables 2 and 3 are used together for mechanical characterization of soft-matter robots. For example, silicone-based elastomer Ecoflex has been used in robot parts at the centimeter scale for functions such as reaching, gripping, and legged locomotion, as shown in Table 2. Ecoflex has a modulus within the range of 55–105 KPa and a breaking strain within the range of 800–1,000%. The corresponding robots were typically actuated by electromagnetic motors with cables, pneumatic actuators, or shape-memory alloys. The mass of the electromagnetic motors, pneumatic actuators, and shape-memory alloys is within the range of 0.01–1, 0.001–30, and 1e-6–1 kg, respectively; the maximal output force of the three is 1–30, 10–200, and 0.08–100 N, respectively; and the maximal stroke of the three is 0.003–0.01, 0.006–0.025, and 0.003–0.1 m, respectively, as shown in Table 3. Based on these values, Figures 9–11 show the correlation between actuator performance and material properties or structure sizes.

The ranges of the elastic modulus of six polymers, including elastomers, thermoplastics, and polymer gels, versus the maximal output force and mass of the corresponding actuators are shown in Figures 9(a) and 10(a). It can be seen

Table 2. Polymer mechanical properties, robotic functions, and actuation methods.

Soft Matter-Function	Modulus (Pa)	Breaking Elongation/Strain	Tensile Strength (MPa)	Soft Part Dimension (mm)	Actuation Method
Silicone-reaching [14]	55–105 K (Ecoflex)	800–1,000%	0.8–2.4	∅ 20–30 × 100–450	Motors with cables, shape-memory alloys
Silicone-handed gripping [20]				∅ 90–140	Pneumatic
Silicone-twining gripping [14]				∅ 20–30 × 100–450	Motors with cables
Silicone-legged locomotion [23]				139 × 59 × 5 or leg ∅ 20 × 240	Pneumatic or motors with cables
Silicone-crawling locomotion [25]	0.3–3 M [41], [46] (PDMS)	>160% [40], [41]	2.2–7.7 [40], [47]	2 × 2	Biological cells
Silicone-jellyfish [31], [32]	0.3–3 M [41], [46] (PDMS) 0.24–1.3 M (MoldMax)	>160% [40], [41] 250–529%	2.2–7.7 [40], [47] 3.3–4	∅ 9 ∅ 164	Biological muscle shape-memory alloys
Silicone-tailed swimming [29]	(Dow Corning 734)	315%	1.5–2	50 × 12	Biological muscle
Silicone-rolling locomotion [26], [27]	0.15–0.6 M (Dragon skin)	364–1,000%	3.3–3.8	Perimeter: 100 or diameter: 130–390	Shape-memory alloys, pneumatic
Silicone-octopus [33]				Leg ∅ 20 × 200	Motors with cables
Polyurethane-hinged flexure [17]	25–690 M	100–1,000%	30–69	20 × 7	Motors with cables
Polyurethane-rolling [28]				∅ 15 × 40	Magnets with electromagnetic field
Polyimide-hinged flexure [18]	2.5 G (Kapton)	82%	231	0.1–1	Piezoceramic
EVA TPA-gripping	5–70 M [42]	>100%	1–5 [42]	Finger: ∅ 15 × 200	Motors and cables
EVA TPA-legged locomotion				Leg: ∅ 7 × 100	Motors and cables
PEEK mesh-peristaltic [15]	3.6 G	50%	90–100	∅ 22 × 200	Shape-memory alloys
Polyester-hinged flexure [19]	2–3.5 G	2–300%	57–700	0.1–1	Piezoelectric actuator, shape-memory alloys
Polyethylene-legged locomotion [24]	0.1 G [48]	90%	33	70	Shape-memory alloys
Polytetrafluoroethylene-hinged flexure [16]	0.35–0.75 G [16]	350–550%	25–36	–	Motors and cables
Polymer gel (exclude EAP gel)-gripping [22]	1 k–33 M [43], [44]	200–2,000%	0.5–5	Finger: 8 × 2 or diameter: ~100	Self-actuated with various stimuli
Polymer gel-tailed swimming [30]				70	Motor
EAPs-gripping, legged and crawling locomotion, and tailed swimming	–	–	–	–	Self-actuated with electric current

Table 3. Actuator characteristics (data in [45, Figures 5 and 6]).

Actuator	Maximal Output Force (N)	Maximal Stroke (m)	Mass (kg)
Electromagnetic	1–30	0.003–0.01	0.01–1
Pneumatic	10–200	0.006–0.025	0.001–30
Shape-memory alloys	0.08–100	0.003–0.1	1e-6–0.1
Biological muscles	5,000–1.5e4	0.15–0.6	N/A
EAPs	0.02–200	1e-5–0.2	1e-8–1e-3
Piezoelectric	0.08–500	1e-5–0.001	1e-4–1

that polymers with a relatively low modulus (such as elastomers and polymer gels) may be actuated by a wider range of actuators, regardless of the maximal output forces and the mass. Polymers with a relatively high modulus (such as polyurethane, polytetrafluoroethylene, and polyester) may only be actuated by actuators with larger output forces and mass. Though it might sound trivial in the sense that materials with a higher modulus require a larger actuation force, the importance for robotics research is deformation of structures rather than that of materials, and, therefore, it is necessary to consider the size of structures on the demand of actuation forces. This is confirmed by the weak correlation in the figures shown by a power-law regression with a negative power value.

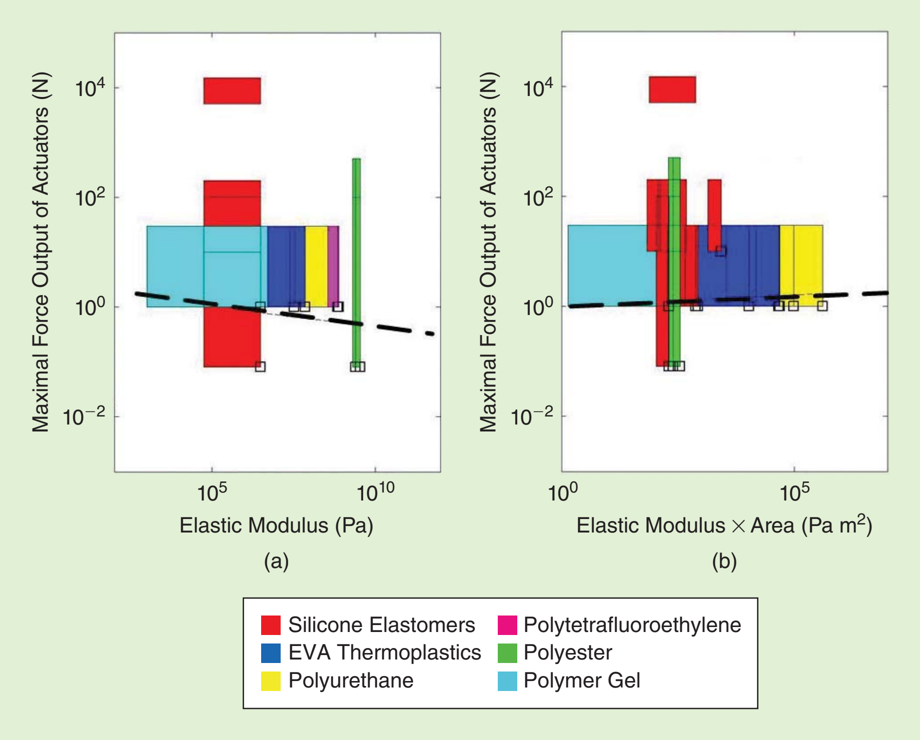


Figure 9. (a) The elastic modulus of polymers used in soft-matter robots versus the maximal output force of their corresponding actuators. The dashed line shows a power-law regression of the corner point values of all areas ($y = 2.23 \times x^{-0.07}$). (b) The product of elastic modulus of polymers and the area of soft structures in robots versus maximal output force of their corresponding actuators. The dashed line shows power-law regression of the corner point values of all areas ($y = 1.03 \times x^{-0.04}$).

Figures 9(b) and 10(b) show the relation when the size of polymer structures in soft-matter robots is considered based on five polymers (with polytetrafluoroethylene being excluded due to lack of data on structure size). Here, the size of polymer structures is represented by the area. Depending on the motion involved in robotic functions, the area was calculated in different ways: for axial motion, the cross-sectional area was used; and for lateral bending motion, either the cross-sectional area (e.g., in the case of legged locomotion) or the area of the plane that bends (e.g., in the case of a hinged flexure) was taken. It can be seen that polymer structures with a relatively small product of modulus and area may be actuated by a wider range of actuators, regardless of the maximal output forces and the mass. Those structures with a larger product of the two may only be actuated by actuators with larger forces and mass. This correlation is strong for elastomers, polyester, and polymer gel and may also be applied to other polymers in principle given a positive power value in regression.

The use of mechanical properties of soft matter, especially those typically employed in material science, was essentially implied, as the quantitative measures for the soft-matter robotics is disputable, and it needs to be carefully considered in the context of the specifics of tasks and functions (Figures 9 and 10). This point is particularly crucial when evaluating robots' functional performance because the evaluation results largely depend on the criteria employed.

Having said that, there are also useful quantification methods for soft-matter robotics research that are based on material properties. For example, the maximal breaking strain of eight polymers versus maximal stroke of the corresponding actuators is shown in Figure 11. It can be seen from the figure that polymers with a relatively smaller maximal breaking strain (e.g., polyimide, EVA, and polyethylene) may be actuated by a wider range of actuators with both small and large maximal strokes. Polymers with a relatively high maximal breaking strain (e.g., elastomers, polytetrafluoroethylene, and polymer gels) may only be actuated by a larger stroke. This correlation is generally applicable based on the available data.

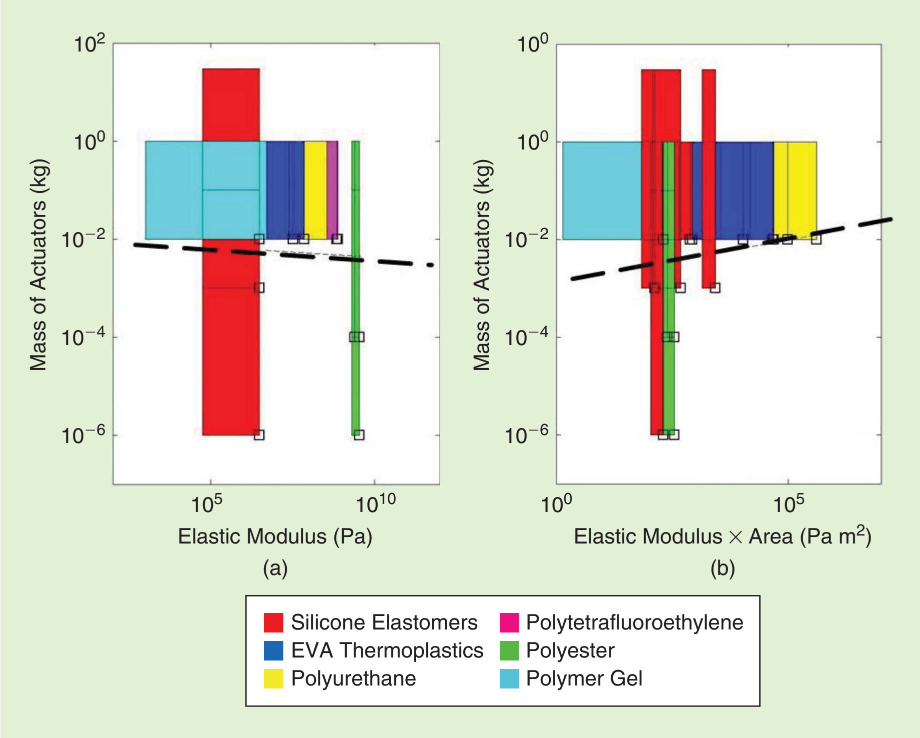


Figure 10. (a) The elastic modulus of polymers used in soft-matter robots versus the mass of their corresponding actuators. The dashed line shows a power-law regression of the corner point values of all areas $(y = 0.01 \times x^{-0.04})$. (b) The product of elastic modulus of polymers and the area of soft structures in robots versus the mass of their corresponding actuators. The dashed line shows power-law regression of the corner point values of all areas $(y = 0.001375 \times x^{-0.17})$.

Conclusions

In this article, we define soft-matter robotics, review deformation and functions, and make a quantitative characterization. We include a range of soft matter such as polymers, granular materials, smart fluids, and colloid explored for robotic technologies, which is much wider than previous reviews did. For the first time, we give a categorization of deformation in soft-matter robots—into elongation/shortening, bending, flowing and transformation—and pair the categories with specific functions. We argue that the pertinent selection of characterization methods of these robots is the first step toward quantification, benchmarking, and systematic investigations. Quantification methods of soft-matter robotics research should be related to deformation of structures because deformation is the origin of all functions in soft-matter robots, such as gripping and locomotion. Material properties such as modulus are, of course, related to deformation, but they work only on the material level and not on the structure level. By comparison, the modulus–area product as well as maximal breaking strain are scalable measures applicable at all levels for polymeric robotic systems, regardless of the shapes of the systems or variations of the tasks.

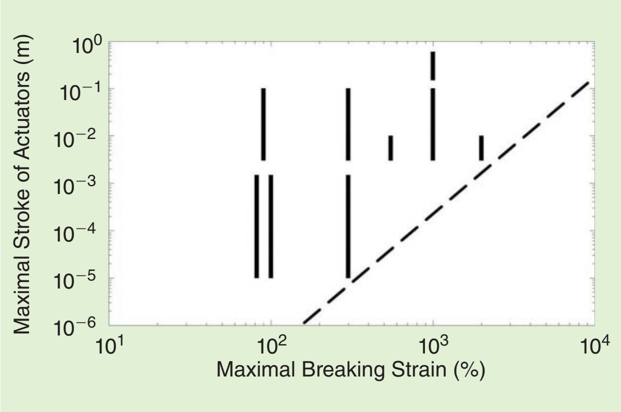


Figure 11. The maximal breaking strain of polymers used in soft-matter robots versus the maximal stroke of their corresponding actuators.

However, the characterization of mechanical behaviors is not the end of the story—it is also necessary to consider how they are quantitatively related to functions of soft robots. Quantification of functions is neither trivial nor impossible. For example, robotic functions, such as locomotion, may be evaluated in terms of energetic cost of transport (energy consumption per unit mass and traveling distance). However, others, such as griping of unknown objects, are more challenging. The approach of this article provides a bottom-up consideration from materials to functions, which works very well in some cases, but a more thorough investigation and discussion will be needed.

As mentioned at the beginning of the article, we observe that the underlying driving force of soft-matter robotics research is the demand for cheaper, safer, more adaptive, and robust robots. From this perspective, benchmarking of soft-matter robots will be necessary in the near future. While cost and safety are common themes across different types of robotics applications, flexibility and adaptability might require specific approaches. Nevertheless, benchmarking has to be application dependent, which may involve one or more types of functions. For example, a similar benchmark to the W Prize for dynamic walking or running robots may be developed for soft-matter mobile robots, which can be used to describe the performance of locomotion speed and power consumption. And a similar benchmark to the U.S. Defense Advanced Research Projects Agency Robotics Challenge may be formed for soft-matter robots for both locomotion and manipulation in a complex task environment.

In conclusion, the use of soft matter shifts the paradigm of robots from kinematic machines toward continuum machines, as in the transition exemplified by robots with polymeric flexure hinges. The borderline between mechanical structures and actuators becomes vague, as exemplified by peristaltic robots made from pneumatic artificial muscles and grippers and crawlers made from EAP. There has been no robot with 100% soft matter, partially due to the fact that actuators made from soft matter, such as shape-memory polymers or EAPs, have limitations in speed, stroke, or power density. Most case studies serve as proof of concept, and the research is still preliminary. Data on actuators (e.g., mass, output force, and stroke), structures (e.g., size), material properties (e.g., modulus and breaking strain), and system performance (e.g., speed and power consumption) should be detailed for further investigation and benchmarking.

Benchmarking has to be application dependent, which may involve one or more types of functions.

Acknowledgments

The work was supported by the Swiss National Science Foundation professorship grant PP00P2123387/1 and the ETH Zurich research grant ETH-23-10-3.

References

[1] D. Trivedi, C. D. Rahn, W. M. Kier, and I. D. Walker, "Soft robotics: Biological inspiration, state of the art, and future research," *Appl. Bionics Biomech.*, vol. 5, no. 3, pp. 99–117, Sept. 2008.

[2] K.-J. Cho, J.-S. Koh, S. Kim, W.-S. Chu, Y. Hong, and S.-H. Ahn, "Review of manufacturing processes for soft biomimetic robots," *Int. J. Precis. Eng. Manuf.*, vol. 10, no. 3, pp. 171–181, July 2009.

[3] R. Pfeifer, M. Lungarella, and F. Iida, "The challenges ahead for bio-inspired soft robotics," *Commun. ACM*, vol. 55, no. 11, pp. 76–87, Nov. 2012.

[4] S. Kim, C. Laschi, and B. Trimmer, "Soft robotics: A bioinspired evolution in robotics," *Trends Biotechnol.*, vol. 31, no. 5, pp. 287–294, May 2013.

[5] S. Bauer, S. Bauer-Gogonea, I. Graz, M. Kaltenbrunner, C. Keplinger, and R. Schwödiauer, "A soft future: From robots and sensor skin to energy harvesters," *Adv. Mater.*, vol. 26, no. 1, pp. 149–162, Jan. 2014.

[6] D. Rus and M. T. Tolley, "Design, fabrication and control of soft robots," *Nature*, vol. 521, pp. 467–475, May 2015.

[7] A. Verl, A. Albu-Schäffer, O. Brock, and A. Raatz, Eds., *Soft Robotics: Transferring Theory to Application*. Berlin, Heidelberg, Germany: Springer-Verlag, 2015.

[8] M. Doi, *Soft Matter Physics*. Oxford, U.K.: Oxford Univ. Press, 2013.

[9] M. R. Cutkosky, J. M. Jourdain, and P. K. Wright, "Skin materials for robotic fingers," in *Proc. IEEE Int. Conf. Robotics Automation*, Raleigh, NC, 1987, pp. 1649–1654.

[10] H. Yousef, M. Boukallel, and K. Althoefer, "Tactile sensing for dexterous in-hand manipulation in robotics—A review," *Sens. Actuators A, Phys.*, vol. 167, no. 2, pp. 171–187, June 2011.

[11] L. Wang, L. Graber, and F. Iida, "Large-payload climbing in complex vertical environments using thermoplastic adhesive bonds," *IEEE Trans. Robot.*, vol. 29, no. 4, pp. 863–874, Aug. 2013.

[12] R. Kornbluh, R. Pelrine, Q. Pei, M. Rosenthal, S. Stanford, N. Bonwit, R. Heydt, H. Prahlad, and S. V. Shastri, "Application of dielectric elastomer EAP actuators," in *Electroactive Polymer Actuators as Artificial Muscles: Reality, Potential, and Challenges*, 2nd ed. Bellingham, WA: SPIE Press, 2004, ch. 16, pp. 529–581.

[13] B. Tondu, "Modelling of the McKibben artificial muscle: A review," *J. Intell. Mater. Syst. Struct.*, vol. 23, no. 3, pp. 225–253, Feb. 2012.

[14] M. Cianchetti, A. Arienti, M. Follador, B. Mazzolai, P. Dario, and M. Follador, "Design concept and validation of a robotic arm inspired by the octopus," *Mater. Sci. Eng. C*, vol. 31, no. 6, pp. 1230–1239, Aug. 2011.

[15] S. Seok, C. D. Onal, K.-J. Cho, R. J. Wood, D. Rus, and S. Kim, "Meshworm: A peristaltic soft robot with antagonistic nickel titanium coil actuators," *IEEE/ASME Trans. Mechatron.*, vol. 18, no. 5, pp. 1485–1497, Oct. 2013.

[16] F. Lotti and G. Vassura, "A novel approach to mechanical design of articulated fingers for robotic hands," in *Proc. IEEE/RSJ Int. Conf. Intelligent Robots Systems*, Lausanne, Switzerland, 2002, pp. 1687–1692.

[17] A. M. Dollar, L. P. Jentoft, J. H. Gao, and R. D. Howe, "Contact sensing and grasping performance of compliant hands," *Auton. Robot.*, vol. 28, no. 1, pp. 65–75, Jan. 2010.

[18] R. J. Wood, "The first takeoff of a biologically inspired at-scale robotic insect," *IEEE Trans. Robot.*, vol. 24, no. 2, pp. 341–347, Apr. 2008.

[19] A. M. Hoover, E. Steltz, and R. S. Fearing, "Roach: An autonomous 2.4g crawling hexapod robot," in *Proc. IEEE/RSJ Int. Conf. Intelligent Robots Systems*, Nice, France, 2008, pp. 26–33.

[20] F. Ilievski, A. D. Mazzeo, R. F. Shepherd, X. Chen, and G. M. Whitesides, "Soft robotics for chemists," *Angew. Chem. Int. Ed.*, vol. 50, no. 8, pp. 1890–1895, Feb. 2011.

[21] Y. Bar-Cohen, "Electroactive polymers as artificial muscles: Capabilities, potentials and challenges," in *Proc. 4th Int. Conf. Expo./Demonstration Robotics Challenging Situations Environments*, Albuquerque, NM, 2000, pp. 188–196.

[22] Z. Hu, X. Zhang, and Y. Li, "Synthesis and application of modulated polymer gels," *Science*, vol. 269, pp. 525–527, July 1995.

[23] R. F. Shepherd, F. Ilievski, W. Choi, S. A. Morin, A. A. Stokes, A. D. Mazzeo, X. Chen, M. Wang, and G. M. Whitesides, "Multigait soft robot," *Proc. Natl. Acad. Sci. United States Amer.*, vol. 108, no. 51, pp. 20400–20403, Dec. 2011.

[24] M. Nishida, H. O. Wang, and K. Tanaka, "Development and control of a small bipedwalking robot using shape memory alloys," *J. Robot. Mechatron.*, vol. 20, no. 5, pp. 793–800, Oct. 2008.

[25] J. Kim, J. Park, S. Yang, J. Baek, B. Kim, S. H. Lee, E. S. Yoon, K. Chun, and S. Park, "Establishment of a fabrication method for a long-term actuated hybrid cell robot," *Lab Chip*, vol. 7, no. 11, pp. 1504–1508, Nov. 2007.

[26] H.-T. Lin, G. G. Leisk, and B. Trimmer, "GoQBot: A caterpillar-inspired soft-bodied rolling robot," *Bioinspir. Biomim.*, vol. 6, no. 2, p. 026007, Apr. 2011.

[27] E. Steltz, A. Mozeika, N. Rodenberg, E. Brown, and H. M. Jaeger, "JSEL: Jamming skin enabled locomotion," in *Proc. IEEE/RSJ Int. Conf. Intelligent Robots Systems*, St Louis, MO, 2009, pp. 5672–5677.

[28] S. Yim and M. Sitti, "Design and rolling locomotion of a magnetically actuated soft capsule endoscope," *IEEE Trans. Robot.*, vol. 28, no. 1, pp. 183–194, Feb. 2012.

[29] H. Herr and R. G. Dennis, "A swimming robot actuated by living muscle tissue," *J. Neuroeng. Rehab.*, vol. 1, no. 1, p. 6, Oct. 2004

[30] J. J. H. Long Jr., T. J. Koob, K. Irving, K. Combie, V. Engel, N. Livingston, A. Lammert, and J. Schumacher, "Biomimetic evolutionary analysis: Testing the adaptive value of vertebrate tail stiffness in autonomous swimming robots," *J. Exp. Biol.*, vol. 209, pp. 4732–4746, Dec. 2006.

[31] J. C. Nawroth, H. Lee, A. W. Feinberg, C. M. Ripplinger, M. L. McCain, A. Grosberg, J. O. Dabiri, and K. K. Parker, "A tissue-engineered jelly fish with biomimetic propulsion," *Nature Biotechnol.*, vol. 30, no. 8, pp. 792–797, Aug. 2012.

[32] A. Villanueva, C. Smith, and S. Priya, "A biomimetic robotic jelly fish (Robojelly) actuated by shape memory alloy composite actuators," *Bioinspir. Biomim.*, vol. 6, no. 3, p. 036004, Sept. 2011.

[33] M. Sfakiotakis, A. Kazakidi, N. Pateromichelakis, and D. P. Tsakiris, "Octopus-inspired eight-arm robotic swimming by sculling movements," in *Proc. IEEE Int. Conf. Robotics Automation*, Karlsruhe, Germany, 2013, pp. 5155–5161.

[34] A. P. Perovskii, "Universal grippers for industrial robots," *Russ. Eng. J.*, vol. 60, no. 8, pp. 9–11, 1980.

[35] G. L. Kenaley and M. R. Cutkosky, "Electrorheological fluid-based robotic fingers with tactile sensing," in *Proc. IEEE Int. Conf. Robotics Automation*, Scottsdale, AZ, 1989, pp. 132–136.

[36] T. Rienmueller and H. Weissmantel, "A shape adaptive gripper finger for robots," in *Proc. 18th Int. Symp. Industrial Robots*, Lausanne, Switzerland, 1988, pp. 241–250.

[37] A. Pettersson, S. Davis, J. O. Gray, T. J. Todd, and T. Ohlsson, "Design of a magnetorheological robot gripper for handling of delicate food products with varying shapes," *J. Food Eng.*, vol. 98, no. 3, pp. 332–338, June 2010.

[38] S. Revzen, M. Bhoite, A. Macasieb, and M. Yim, "Structure synthesis on-the-fly in a modular robot," in *Proc. IEEE/RSJ Int. Conf. Intelligent Robots Systems*, San Francisco, CA, 2011, pp. 4797–4802.

[39] L. Wang, L. Brodbeck, and F. Iida, "Mechanics and energetics in tool manufacture and use: A synthetic approach," *J. Royal Soc. Interface*, vol. 11, p. 20140827, Nov. 2014.

[40] K. Khanafer, A. Duprey, M. Schlicht, and R. Berguer, "Effects of strain rate, mixing ratio, and stress–strain definition on the mechanical behavior of the polydimethylsiloxane material as related to its biological applications," *Biomed. Microdev.*, vol. 11, no. 2, pp. 503–508, Apr. 2009.

[41] I. D. Johnston, D. K. McCluskey, C. K. L. Tan, and M. C. Tracey, "Mechanical characterization of bulk Sylgard 184 for microfluidics and microengineering," *J. Micromech. Microeng.*, vol. 24, no. 3, p. 035017, Mar. 2014.

[42] H.-H. Shih and G. R. Hamed, "Peel adhesion and viscoelasticity of poly(ethylene-co-vinyl acetate)-based hot melt adhesives. I. The effect of tackifier compatibility," *J. Appl. Polym. Sci.*, vol. 63, no. 3, pp. 323–331, Jan. 1997.

[43] K. R. Shah, "Hydrogels, microphase separated blends (biomaterials and drug delivery systems)," in *Polymeric Materials Encyclopedia*. Boca Raton, FL: CRC Press, 1996, pp. 3092–3097.

[44] O. Okay, "General properties of hydrogels," in *Hydrogel Sensors and Actuators*. Berlin, Heidelberg, Germany: Springer-Verlag, 2010, ch. 1, pp. 1–14.

[45] M. Zupan, M. F. Ashby, and N. A. Fleck, "Actuator classification and selection—The development of a database," *Adv. Eng. Mater.*, vol. 4, no. 12, pp. 933–940, Dec. 2002.

[46] D. Armani, C. Liu, and N. Aluru, "Re-configurable fluid circuits by PDMS elastomer micromachining," in *Proc. 12th IEEE Int. Conf. Micro Electro Mechanical Systems*, Orlando, FL, 1999, pp. 222–227.

[47] J. E. Mark, *Polymer Data Handbook*. Oxford, U.K.: Oxford Univ. Press, 1999, p. 431.

[48] A. J. Peacock, *Handbook of Polyethylene: Structures: Properties, and Applications*. New York: Marcel Dekker, 2000, p. 135.

[49] D. W. Hong, M. Ingram, and D. Lahr, "Whole skin locomotion inspired by amoeboid motility mechanisms," *ASME J. Robot. Mech.*, vol. 1, no. 1, p. 011015, Feb. 2009.

[50] T. Nakamura, Y. Hidaka, M. Yokojima, and K. Adachi, "Development of peristaltic crawling robot with artificial rubber muscles attached to large intestine endoscope," *Adv. Robot.*, vol. 26, no. 10, pp. 1161–1182, July 2012.

[51] R. Sahai, S. Avadhanula, R. Groff, E. Steltz, R. Wood, and R. S. Fearing, "Towards a 3g crawling robot through the integration of microrobot technologies," in *Proc. IEEE Int. Conf. Robotics Automation*, Orlando, FL, 2006, pp. 296–302.

[52] G. Kofod, W. Wirges, M. Paajanen, and S. Bauer, "Energy minimization for self-organized structure formation and actuation," *Appl. Phys. Lett.*, vol. 90, no. 8, p. 081916, Feb. 2007.

[53] D. J. Maitland, M. F. Metzger, D. Schumann, A. Lee, and T. S. Wilson, "Photothermal properties of shape memory polymer micro-actuators for treating stroke," *Lasers Surg. Med.*, vol. 30, no. 1, pp. 1–11, Jan. 2002.

[54] E. Palleau, D. Morales, M. D. Dickey, and O. D. Velev, "Reversible patterning and actuation of hydrogels by electrically assisted ionoprinting," *Nature Commun.*, vol. 4, p. 2257, Aug. 2013.

[55] M. Calisti, M. Giorelli, G. Levy, B. Mazzolai, B. Hochner, C. Laschi, and P. Dario, "An octopus-bioinspired solution to movement and manipulation for soft robots," *Bioinspir. Biomim.*, vol. 6, no. 3, p. 036002, Sept. 2011.

[56] K. W. Wait, P. J. Jackson, and L. S. Smoot, "Self locomotion of a spherical rolling robot using a novel deformable pneumatic method," in *Proc. IEEE Int. Conf. Robotics Automation*, Anchorage, AK, 2010, pp. 3757–3762.

[57] C. Zhou and K.-H. Low, "Better endurance and load capacity: An improved design of manta ray robot (Roman-II)," *J. Bionic Eng.*, vol. 7, pp. S137–S144, Sept. 2010.

[58] G. H. Kwon, J. Y. Park, J. Y. Kim, M. L. Frisk, D. J. Beebe, and S. H. Lee, "Biomimetic soft multifunctional miniature aquabots," *Small*, vol. 4, no. 12, pp. 2148–2153, Dec. 2008.

[59] J. J. R. Amend, E. Brown, N. Rodenberg, H. M. Jaeger, and H. Lipson, "A positive pressure universal gripper based on the jamming of granular material," *IEEE Trans. Robot.*, vol. 28, no. 2, pp. 341–350, Apr. 2012.

[60] G. J. Monkman, "Compliant robotic devices, and electroadhesion," *Robotica*, vol. 10, no. 2, pp. 183–185, Mar. 1992.

[61] M. Otake, Y. Kagami, M. Inaba, and H. Inoue, "Motion design of a starfish-shaped gel robot made of electro-active polymer gel," *Robot. Auton. Syst.*, vol. 40, nos. 2–3, pp. 185–191, Aug. 2002.

[62] D. Morales, E. Palleau, M. D. Dickey, and O. D. Velev, "Electro-actuated hydrogel walkers with dual responsive legs," *Soft Matter*, vol. 10, no. 9, pp. 1337–1348, Mar. 2014.

Liyu Wang, University of California at Berkeley, United States. E-mail: liyu.wang@wadh.oxon.org.

Fumiya Iida, University of Cambridge, United Kingdom. E-mail: fi224@cam.ac.uk.

RoboCup Simulation Leagues

Enabling Replicable and Robust Investigation of Complex Robotic Systems


©ISTOCKPHOTO.COM/KIRSTYPARGETER

By David M. Budden, Peter Wang, Oliver Obst, and Mikhail Prokopenko

Physically realistic simulated environments are powerful platforms for enabling measurable, replicable, and statistically robust investigation of complex robotic systems. Such environments are epitomized by the RoboCup (RC) simulation leagues, which have been successfully utilized to conduct massively parallel experiments on a variety of topics, including optimization of bipedal locomotion, self-localization from noisy perception data, and planning complex multiagent strategies without direct agent-to-agent communication. Many of these systems are later transferred to physical robots, making the simulation leagues invaluable beyond the scope of simulated soccer matches.

In this article, we provide an overview of the RC simulation leagues and describe their properties as they pertain to replicable and robust robotics research. To demonstrate their utility directly, we leverage the ability to run parallelized experiments to evaluate different competition formats (e.g., round robin) for the RC two-dimensional (2-D) simulation league. Our results demonstrate that a previously proposed hybrid format minimizes fluctuations from true (statistically significant) team performance rankings within the time constraints of the RC World Finals. Our experimental analysis would be impossible with physical robots alone, and we encourage other researchers to

---

Digital Object Identifier 10.1109/MRA.2015.2446911
Date of publication: 18 August 2015

explore the potential for enriching their experimental pipelines with simulated components, both to minimize the experimental costs and to enable others to replicate and expand upon their results in a hardware-independent manner.

Simulation in Robotics Research

Robotics researchers face many unique challenges when designing measurable, replicable, and statistically robust experiments. Robots are generally expensive and require regular maintenance, pressuring researchers to minimize the time spent evaluating algorithms and behaviors. This problem is amplified by the experimental confounds introduced by robots (e.g., motor temperature) and their environments (e.g., lighting variation), requiring more experimental iterations to yield statistically robust results. Moreover, the cost of physical robots makes platform-specific research replication particularly difficult, threatening the reliability of the peer-review process in the presumably common scenario of reviewers not having access to the robot in question.

In recent years, robot simulation has emerged as a powerful technique for replicable and robust investigation of complex robotic systems. Popular simulators include Simspark [1], Gazebo [2], Webots [3], and USARsim [4], which have been successfully applied in multi-institutional collaborations, including RC Simulation Leagues [5], [6], RobotStadium [7], the Defense Advanced Research Projects Agency Virtual Challenge [8], and the IEEE Virtual Manufacturing Competition [9]. Although these platforms are unable to perfectly model the physical sources of system stochasticity, this shortcoming is greatly outweighed by their ability to remove the requirement for physical robots and inherent support for massively parallel processing. In scenarios where imperfect system modeling is a major limitation, hybrid solutions have been developed that use simulated environments for initial experimentation (e.g., global exploration of high-dimensional parameter spaces) and physical robots for fine tuning (e.g., local optimization across principal components). The latter approach is epitomized by bipedal gait optimization and has been shown to be highly effective for the Aldebaran NAO humanoid robot [10], [11].

In this article, we provide an overview of the RC simulation leagues in the context of enabling replicable and statistically robust robotics research. We demonstrate the utility of massively parallel experimentation by evaluating different competition formats (e.g., round robin) for the RC 2-D simulation league. Tens of thousands of games were necessary to resolve nontransitivity and inherent stochasticity of team performance, which would be intractable for 10-min matches with physical robots. Our results demonstrate that a hybrid format best captures true team performance in the time constraints of the RC World Finals. This format was subsequently adopted for the competitions at RC 2014 Brazil.

RoboCup Simulation Leagues

RoboCup (the World Cup of robot soccer) was first proposed in 1997 as a standard problem for the evaluation of theories, algorithms, and architectures in the areas of artificial intelligence (AI), robotics, and computer vision [13]. This proposal followed the observation that traditional AI problems were increasingly unable to meet the requirements of appropriately/effectively evaluating theories, algorithms, and architectures and that a new challenge was necessary to initiate the development of next-generation technologies.

The overarching RC goal of developing a team of humanoid robots capable of defeating the FIFA World Cup champion team, called *the Millennium Challenge*, has been a major factor in driving research in AI and related areas for nearly two decades, with a search for the term *RC* in a major literature database yielding over 25,000 results. Since 1997, researchers and competitors have decomposed this ambitious pursuit into two complementary categories [13].

- **Physical Robot League**: Using physical robots to play soccer games. This category now contains many different leagues for both wheeled robots (small-sized and midsized leagues) and humanoids [standard platform league (SPL) and humanoid league], with each focusing on different aspects of physical robot design, motor control and bipedal locomotion, real-time localization, and computer vision [14], [15].
- **Software Agent League:** Using software or synthetic agents to play soccer games on an official soccer server over a network. This category contains both 2-D [5], [16], [17] and three-dimensional (3-D) [6] simulation leagues.

The RC simulation leagues traditionally involve the largest number of international participating teams, reaching 40 in 2013 [18]. The ability to simulate soccer matches without physical robots removes low-level hardware and environmental issues (e.g., motor temperature and breakages), allowing teams to focus on the development of complex team behaviors and strategies for a larger number of autonomous agents. Moreover, the simulation leagues often serve as platforms for the initial development and evaluation of software modules for later integration into physical robots [10], [11]. Many of these modules have applications beyond the RC domain (e.g., localization and mapping [12]), and the hardware-independent results inherent to simulated robots promote extension and replication by other researchers.

Properties and Utility of 2-D and 3-D Leagues

The RC simulation league consists of both 2-D and 3-D competitions, which exhibit many similarities [18].

- The world model, including player and ball dynamics and kinematics, is simulated by a central soccer server [5].
- Participants develop a team of fully autonomous agents, each of which interacts with the soccer server.
  - Each agent receives information from the server regarding its current field of view.
  - Each agent determines what actions to execute and submits these requests to the server.
  - The server fulfills these requests and resolves any conflicts (e.g., two agents attempting to occupy the same spatial location).

The server proceeds in real time and imposes noise on both the agents' observations and actions [19]. It is the responsibility of each agent to submit its action requests at the appropriate times to stay synchronized with the soccer server. Furthermore, each agent is allocated an individual process/core, and no direct interprocess communication is permitted. The soccer server provides a low-bandwidth, indirect communications method between agents by support simulated verbal commands.

2-D Simulation League

The 2-D simulation league involves circular players being modeled with an (x, y) position and orientation θ. Each agent also maintains a head angle relative to its global orientation, allowing control of its field of view within human-like constraints. The action commands available to each agent include the following:

- turn body or neck by a specified angle
- dash forward or backward with a specified power
- slide tackle in a specified direction
- kick the ball in a specified direction with a specified angle, if near
- catch the ball if near (goalkeeper only)
- communicate with other players, either verbally or by pointing at a specified position.

Each team consists of 11 players and a coach, which is a nonplaying agent responsible for the allocation of players to each position given a number of randomly generated physical profiles (including characteristics such as speed and stamina). The 2-D simulation league does not model the dynamics or kinematics of any given human or robot. Instead, it encourages the development of complex player behaviors and team strategies [16]–[18]. The simulation league is also a powerful framework for evaluating the emergent downstream effects (e.g., team performance) of small perturbations to the underlying individual agents, as demonstrated in our recent study of particle filtering and self-localization [12].

A screenshot from the 2-D simulation league graphical client is presented in Figure 1(a), and the 2-D soccer server is available online at: http://sourceforge.net/projects/sserver/.

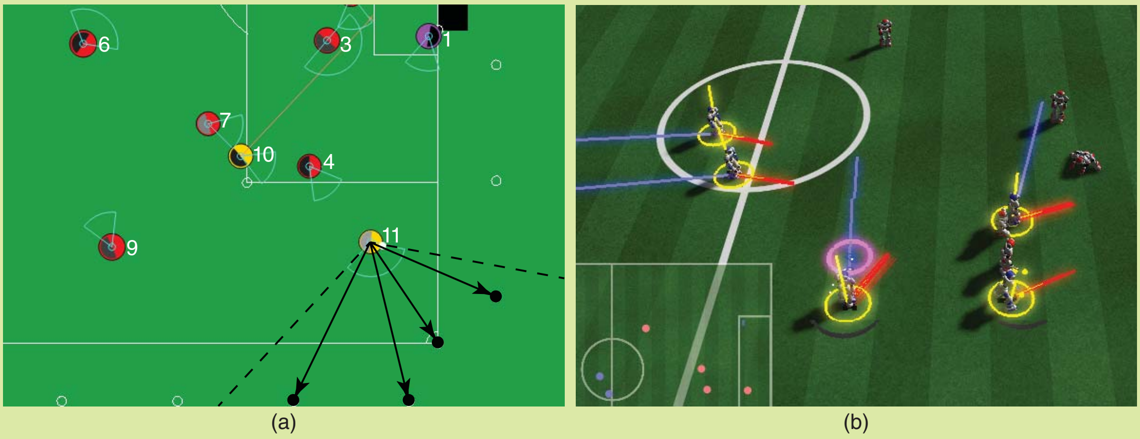
3-D Simulation League

The 3-D simulation league implements a physically realistic world model and action interface, closer to robots than human players [18]. In particular, the simulator uses the Open Dynamics Engine library for the simulation of rigid body dynamics, collision detection, and friction, based on a model of the Aldebaran NAO humanoid robot (shown in Figure 2). To remain consistent with the anatomy of the NAO, each agent simulates the following:

- 22-degrees of freedom (DOF) in a 57-cm, 4.5-kg humanoid robot (six in each leg, four in each arm, and two in the neck)
- perceptors that provide the agent with noiseless measurements of each joint position during every simulation cycle
- effectors that allow the agent to specify a direction and torque for each joint.

Although no noise is introduced to the perceptor and the effector signals (with the exception of that resulting from approximations in the physics engine), the 3-D simulation league introduces the nontrivial challenges of enabling each agent to stably walk, kick, dive, and stand up after falling. This creates an ideal framework for global optimization (and benchmarking optimization algorithms) across the high-dimensional parameter spaces characteristic of bipedal locomotion systems. Although the simulated agents do not perfectly model the stochasticity inherent to actual NAOs, this approach has proved very successful in identifying near-optimal parameter sets for subsequent local optimization on the physical robot [10], [11] (often in a lower dimensional principal component space).

A screenshot from the 3-D simulation league graphical client is presented in Figure 1(b), and the 3-D soccer server is available online at: http://sourceforge.net/projects/simspark/.



(a) (b)

Figure 1. An example of screenshots from the RC simulation leagues: (a) 2-D [5] and (b) 3-D [1]. The 2-D simulation league screenshot demonstrates players from two teams (represented by red and yellow circles) and their respective fields of view. The black arrows around player 11 illustrate how an agent can self-localize from observations of unique landmark features [12]. The 3-D simulation league screenshot demonstrates similar tactical overlays for each player.

Enabling Replicable and Robust Analysis

Collectively, the simulation leagues provide ideal platforms for investigating emergent properties of complex robotic systems. Most team games and sports (both real and virtual) are characterized by rich and dynamic interactions that influence the contest outcome in a nontransitive manner. As described by Vilar et al. [21] "quantitative analysis is increasingly being used in team sports to better understand performance in these stylized, delineated, complex social systems." Early examples of such quantitative analysis include sabermetrics, which attempts to search for objective knowledge about baseball by considering the statistics of in-game activity [22]. A recent study by Fewell et al. [23] involved the analysis of basketball games as networks, with properties including degree centrality, clustering, entropy, and flow centrality (calculated from measurements of ball position throughout the game). This idea was extended by Vilar et al. [21], who considered the local dynamics of collective team behavior to quantify how teams occupy subareas of the field as a function of ball position. Recently, Cliff et al. [24] presented several information-theoretic methods of quantifying dynamic interactions in soccer games and used the RC 2-D simulation league as an experimental platform.

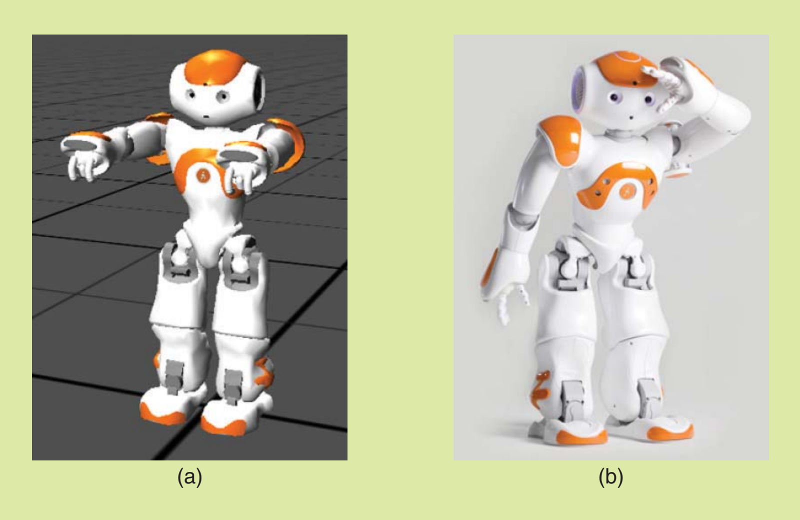

(a)
(b)

Figure 2. The Webots-simulated NAO [3] is well-suited to global optimization across the high-dimensional parameter spaces characteristic of bipedal locomotion systems, as experiments may be automated, parallelized, and replicated exactly. This approach has proved very successful in identifying near-optimal parameter sets for subsequent local optimization on actual Aldebaran NAO robots [10], [11], highlighting the utility of simulation leagues for the investigation and the improvement of physical robotic systems. (a) The Webots-simulated NAO [3]. (b)The actual Aldebaran NAO (http://www.aldebaran.com).

In addition to allowing high-level analysis of robotic systems overall, the simulation leagues provide inherent support for massively parallel processing. This property has been leveraged for the development and the analysis of algorithms with widespread applications in robotics, e.g., optimizing bipedal locomotion [10], [11], self-localization from noisy perception data [12], and planning complex multiagent strategies without direct agent-to-agent communication [16], [17]. Although simulation league agents have only noisy perception of their environment, the soccer server itself has perfect information regarding the global state, which enables replicable quantification of experimental performance (e.g., walk speed/stability and localization accuracy).

Wider Implications in Robotics Research

Robots are generally expensive to purchase, maintain, and transport, creating an intractably high entry barrier for institutes with limited access to research funding. By removing the requirement for physical robots, the RC simulation leagues allow such institutes to actively contribute to many fields of robotics research. To validate this assertion, Figure 3 presents the public expenditure on education as a percentage of gross domestic product (GDP) public expenditure on education (PEoE) at purchasing power parity per capita (GDP/cap)[20] for the home country of each participating RC 2013 team, averaged over each of the six largest RC leagues.

Simulation League Case Study: Analysis of Competition Formats

The simulation league supports fully automated, massively parallel analysis of complex robotic systems, enabling replicable and robust investigation of algorithms and higher level emergent behaviors. In the following sections, we leverage these properties to expand upon our previous analysis of RC competition formats (e.g., round robin) to determine which best approximates the true performance rankings of competing teams [25].

The selection of an appropriate competition format is critical to both the success and the credibility of any competition. Unfortunately, this choice is not straightforward. The ideal format must minimize the randomness relative to the true performance ranking of teams while keeping the number of games to a minimum, to both satisfy time constraints and retain the interest of participants and spectators. Furthermore, maintaining competition interest introduces a number of constraints to competition formats, e.g., multiple games between the same two opponents (the obvious method of achieving a statistically significant ranking) should be avoided, making the resolution of nontransitive performance difficult.

Robocup Competition Formats

The following competition formats were adopted to determine the final rank of the top eight RC 2-D simulation league teams from 2012 to 2014.

- **2012:** The top four teams played six games each [three quarterfinals (round robin), two semifinals, and classification matches for first versus second and third versus fourth] and the bottom four teams played four games each.
- **2013:** A double-elimination system was adopted, where a team is ineligible for first place upon losing two games. A

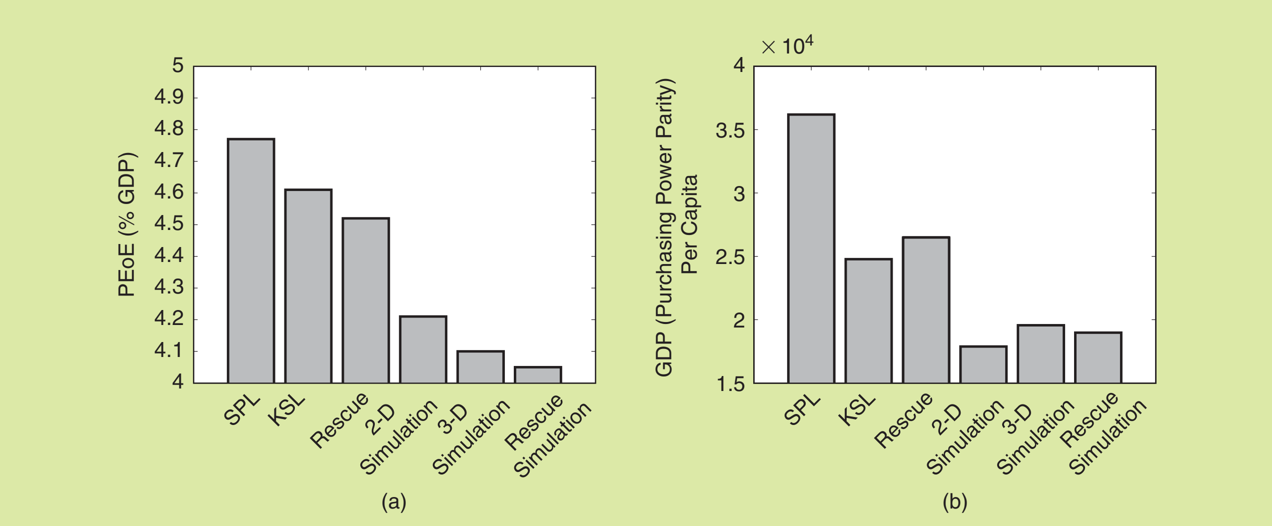


Figure 3. (a) The PEoE and (b) GDP/cap [20] for the home country of each participating RC 2013 team, averaged over each of the six largest RC leagues. Each of the three major simulation leagues (2-D, 3-D, and simulated rescue) exhibits significantly lower values than those requiring the purchase or development of physical robots [SPL, kid-sized league (KSL), and physical rescue].

total of 14 games were played in double-elimination format (i.e., $2n-2$, $n=8$) followed by two classification games.

- **2014:** Our proposed format was adopted [25]. In particular, round-robin games were conducted for the top eight teams (28 games) followed by four classification games for first versus second, third versus fourth, and so on.

Previously, it has been unclear whether these changes in the competition format improve the fairness and the reproducibility of the final team rankings. In general, lack of reproducibility is due to the nontransitivity of team performance (a well-known phenomena that occurs frequently in actual human team sports).

Methods of Ranking Team Performance

Before evaluating different competition formats, it is necessary to establish a fair (i.e., statistically significant) ranking of the top eight RC 2-D simulation league teams from previous years. This was accomplished by conducting an eight-team round robin for previous years, where all 28 pairs of teams play approximately 1,000 games against one another. In addition, two different schemes were considered for point calculation [25].

- **Continuous Scheme:** Teams are ranked by the sum of average points obtained against each opponent across all 1,000 games.
- **Discrete Scheme:** The average score between each pair of teams (across all 1,000 games) is rounded to the nearest integer (e.g., 1.9:1.2 is rounded to 2:1). Next, points are allocated for each pairing based on these rounded results: 3 for a win, 1 for a draw, and 0 for a loss. Teams are then ranked by the sum of these points received against each opponent.

The final rankings generated for the 2012 and 2013 RoboCup 2-D simulation league teams under these two schemes are presented in [25]. Although both the schemes have been shown to generate statistically robust results, we have chosen to adopt the continuous scheme for this article to avoid the boundary effects inherent to discretization. In order to formally capture the overall difference between two rankings $\mathbf{r^a}$ and $\mathbf{r^b}$, the L_1 distance was utilized:

$$d_1(\mathbf{r^a},\mathbf{r^b}) = \|\mathbf{r^a}-\mathbf{r^b}\|_1 = \sum_{i=1}^{n} |r_i^a - r_i^b|, \tag{1}$$

where i is the index of the ith team in each ranking, $1 \le i \le 8$.

Results

Following the iterated round-robin and continuous ranking scheme described in the "Methods of Ranking Team Performance" section, statistically significant rankings were generated for the top eight RC 2-D simulation league teams for 2012–2014. The full set of experiments is described in [25], which verified that the proposed format (later adopted for RC 2014) consistently outperformed the other candidates in terms of approximating the true rankings for RoboCup 2012 and 2013. We expand upon this analysis to incorporate the results for RC 2014.

The L_1 distance [see (1)] was used to capture the discrepancy between RC final results, $\mathbf{r^a}$, and the statistically significant rankings generated from the 28,000-game round robin, $\mathbf{r^c}$:

$$\begin{aligned} d_1(\mathbf{r^a},\mathbf{r^c})_{2012} &= 12 \\ d_1(\mathbf{r^a},\mathbf{r^c})_{2013} &= 12, \end{aligned} \tag{2}$$

where the competition formats for RC 2012 and 2013 are described in the "RC Competition Formats" section. We can also quantify that the corresponding discrepancy, $\mathbf{r^p}$, had our proposed competition format [25] been used for those competitions:

$$\begin{aligned} d_1(\mathbf{r^p},\mathbf{r^c})_{2012} &= 4 \\ d_1(\mathbf{r^p},\mathbf{r^c})_{2013} &= 6. \end{aligned} \tag{3}$$

Our proposed format was subsequently adopted for the RC 2014 finals. The divergence from statistically robust team rankings was equivalently small:

$$d_1(\mathbf{r}^{\mathrm{a}},\mathbf{r}^{\mathrm{c}})_{2014} \Longleftrightarrow d_1(\mathbf{r}^{\mathrm{p}},\mathbf{r}^{\mathrm{c}})_{2014} = 4. \qquad (4)$$

To statistically validate that the proposed competition format is significantly more appropriate than those adopted at RC 2012 and RC 2013, 10,000 tournaments were generated for each format by randomly sampling the game results from the 28,000-game round robin. For each tournament, the L_1 distance $d_1(\mathbf{r}^{\mathrm{a}},\mathbf{r}^{\mathrm{b}})$ [see (1)], was calculated to capture the discrepancy between the tournament and the true team rankings. These results are presented in Figure 4 for the top eight teams from RC 2012–2014. It is evident that the proposed format yields more statistically robust rankings (i.e., smaller L_1 distance) than the formats adopted in previous years.

In addition to comparing the accuracy of team rankings under different competition formats, it is interesting to compare the team performance against a consistent benchmark. Before 2013, it was commonplace for the simulation league teams to optimize their performance against the default Agent2D code [26], which is reflected in the high Spearman's rank correlation coefficient $(\rho_{2012} = 0.98)$ between true rankings and goal difference against Agent2D (across 1,000 games per team). Since 2013, these correlations have decreased substantially $(\rho_{2013} = 0.55$ and $\rho_{2014} = 0.57)$with teams opting to optimize behavior against the binaries published by top-performing teams postcompetition to gain a competitive advantage with opponent-specific strategy. The average goal difference against Agent2D decreased for the top four RC teams between 2013 and 2014 accordingly. Importantly, this level of behavioral complexity (in addition to our analysis of competition formats) would be impossible without the support for massively parallel processing inherent to simulation leagues.

Summary and Discussion

Continual increases in data volume and computational power have led to increased complexity in the experimental methodologies across most fields of research. Therefore, it is unsurprising that many fields (particularly in the life sciences [27]) have recently placed increased focus on enabling measurable, replicable, and statistically robust results. Although robotics researchers face many unique challenges due to the expense and stochasticity inherent to physical robots, we propose that physically realistic simulated environments (epitomized by the RC simulation leagues) have an important and widespread role to play in the future of robotics.

The simulation leagues often serve as platforms for the initial development and the evaluation of software modules for later integration into physical robots [10], [11], and many of these modules have applications beyond the RC domain (e.g., localization and mapping [12]). They also enable the investigation of high-level emergent properties of complex robotic systems, as demonstrated in a recent study by Cliff et al. [24] that presents novel information-theoretic methods for quantifying dynamic interactions in a multiagent context.

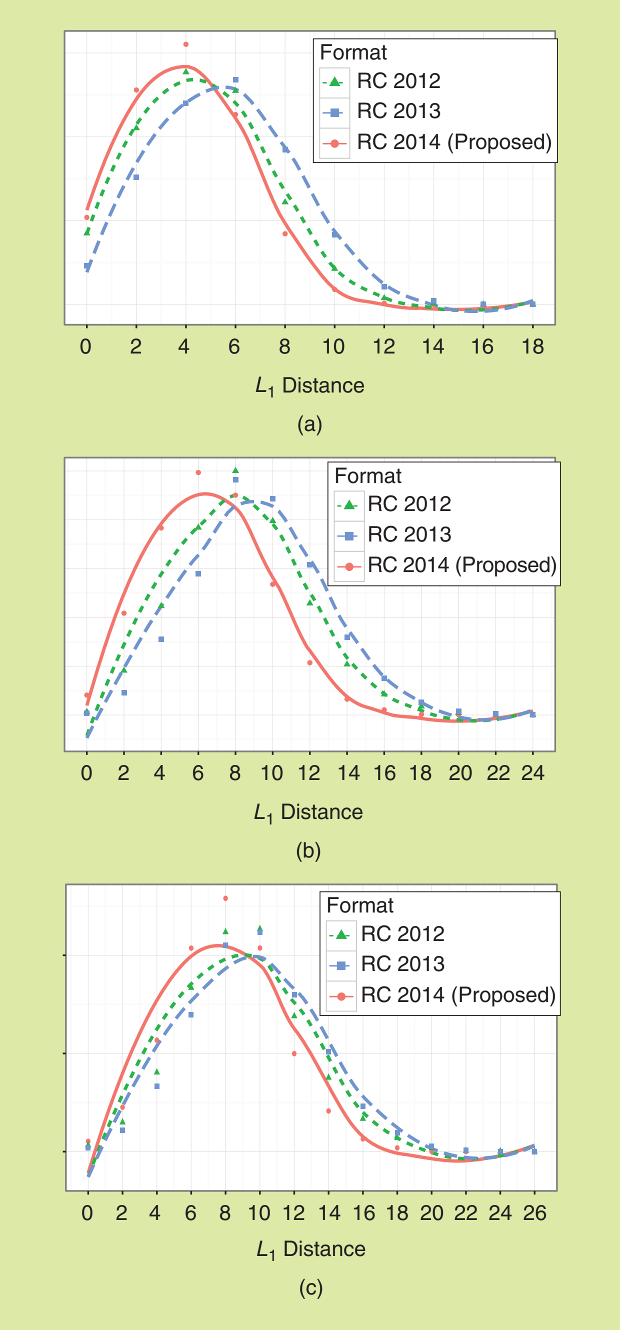


Figure 4. The discrepancy between tournament and true team rankings, captured as an L_1 distance (1), for 10,000 randomly generated tournaments structured according to the three considered formats: (a) 2012, (b) 2013, and (c) 2014. It is evident that the proposed format (red) yields more statistically robust rankings (i.e., a smaller L_1 distance) than the formats adopted in RC 2012 (green) and 2013 (blue), considering the top eight teams from each RC.

In this article, we have provided an overview of the RoboCup simulation leagues (both 2-D and 3-D) and described their properties as they pertain to replicable and robust robotics research. To demonstrate their utility directly, we leverage the ability to run massively parallelized experiments to evaluate different competition formats (e.g., round robin) for the RC 2-D simulation league. Our results demonstrate that a hybrid format [25] minimizes fluctuations from true (statistically significant) team performance rankings within the time constraints of the RC world finals.

Our experimental analysis and many others in [10], [11], [12], and [24] would be impossible with physical robots alone, and have widespread applications beyond the scope of simulated soccer matches. We encourage other researchers to explore the potential for enriching their experimental pipelines with simulated components to minimize the experimental costs and enable others to replicate and expand upon experimental results in a hardware-independent manner.

References

[1] J. Boedecker and M. Asada, "SimSpark—Concepts and applications in the RoboCup 3D soccer simulation league," in *Proc. Workshop SIMPAR Int. Conf. Simulation, Modeling Programming Autonomous Robots*, 2008, pp. 174–181.

[2] N. Koenig and A. Howard, "Design and use paradigms for gazebo, an open-source multi-robot simulator," in *Proc. IEEE/RSJ Int. Conf. Intelligent Robots Systems*, 2004, vol. 3, pp. 2149–2154.

[3] O. Michel, "Webots: Symbiosis between virtual and real mobile robots," in *Virtual Worlds*. Berlin, Germany: Springer-Verlag, 1998, pp. 254–263.

[4] S. Carpin, M. Lewis, J. Wang, S. Balakirsky, and C. Scrapper, "USARSim: A robot simulator for research and education," in *Proc. IEEE Int. Conf. Robotics Automation*, 2007, pp. 1400–1405.

[5] M. Chen, E. Foroughi, F. Heintz, Z. Huang, S. Kapetanakis, K. Kostiadis, J. Kummeneje, I. Noda, O. Obst, P. Riley, T. Steffens, Y. Wang, and X. Yin, "RoboCup soccer server," Tech. Rep., Aug. 2002.

[6] RoboCup Technical Committee. (2013). RoboCup Soccer Simulation League 3D competition rules and setup. [Online]. Available: http://homepages.herts.ac.uk/sv08aav/RCSoccerSim3DRules2013.1.pdf

[7] O. Michel, Y. Bourquin, and J.-C. Baillie, "Robotstadium: Online humanoid robot soccer simulation competition," in *RoboCup 2008: Robot Soccer World Cup XII*. Berlin, Germany: Springer, 2009, pp. 580–590.

[8] C. Aguero, N. Koenig, I. Chen, H. Boyer, S. Peters, J. Hsu, B. Gerkey, S. Paepcke, J. L. Rivero, J. Manzo, E. Krotkov, and G. Pratt, "Inside the virtual robotics challenge: Simulating real-time robotic disaster response," *IEEE Trans. Automat. Sci. Eng.*, vol. 12, no. 2, pp. 494–506, 2015.

[9] S. Balakirsky, R. Madhavan, and C. Scrapper, "NIST/IEEE virtual manufacturing automation competition: From earliest beginnings to future directions," in *Proc. 8th Workshop Performance Metrics Intelligent Systems*, 2008, pp. 214–219.

[10] P. MacAlpine, S. Barrett, D. Urieli, V. Vu, and P. Stone, "Design and optimization of an omnidirectional humanoid walk: A winning approach at the robocup 2011 3D simulation competition," in *Proc. 26th Conf. Artificial Intelligence*, 2012, pp. 1047–1053.

[11] C. Niehaus, T. Röfer, and T. Laue, "Gait optimization on a humanoid robot using particle swarm optimization," in *Proc. IEEE 2nd Workshop Humanoid Soccer Robots*, 2007, pp. 1–7.

[12] D. Budden and M. Prokopenko, "Improved particle filtering for pseudo uniform belief distributions in robot localisation," in *RoboCup 2013: Robot Soccer World Cup XVII*. Berlin, Germany: Springer-Verlag, 2013.

[13] H. Kitano, M. Asada, Y. Kuniyoshi, I. Noda, and E. Osawa, "RoboCup: The robot world cup initiative," in *Proc. 1st Int. Conf. Autonomous Agents*, 1997, pp. 340–347.

[14] D. Budden and A. Mendes, "Unsupervised recognition of salient colour for real-time image processing," in *RoboCup 2013: Robot Soccer World Cup XVII*. Berlin, Germany: Springer, 2013.

[15] J. Fountain, J. Walker, D. Budden, A. Mendes, and S. Chalup, "Motivated reinforcement learning for improved head actuation of humanoid robots," in *RoboCup 2013: Robot Soccer World Cup XVII*. Berlin, Germany: Springer, 2013.

[16] M. Prokopenko, O. Obst, P. Wang, D. Budden, and O. Cliff, "Gliders2013: Tactical analysis with information dynamics," in *Proc. RoboCup Symp. Competitions: Team Description Papers*, Eindhoven, The Netherlands, June 2013.

[17] M. Prokopenko, P. Wang, and O. Obst, "Gliders2014: Dynamic tactics with voronoi diagrams," in *Proc. RoboCup Symp. Competitions: Team Description Papers*, Joao Pessoa, Brazil, July 2014, pp. 1–6.

[18] A. Bai, X. Chen, P. MacAlpine, D. Urieli, S. Barrett, and P. Stone, "Wright-eagle and UT Austin villa: RoboCup 2011 simulation league champions," in *RoboCup 2011: Robot Soccer World Cup XV*. Berlin, Germany: Springer, 2012, pp. 1–12.

[19] I. Noda and P. Stone, "The RoboCup soccer server and CMUnited clients: Implemented infrastructure for MAS research," *Auton. Agents Multi-Agent Syst.*, vol. 7, nos. 1–2, pp. 101–120, 2003.

[20] World Bank. (2012). World development indicators 2012: GDP per capita, PPP (current international $). [Online]. Available: http://data.worldbank.org/data-catalog/world-development-indicators and http://data.worldbank.org/indicator/NY.GDP.PCAP.PP.CD

[21] L. Vilar, D. Araújo, K. Davids, and Y. Bar-Yam, "Science of winning soccer: Emergent pattern-forming dynamics in association football," *J. Syst. Sci. Complexity*, vol. 26, no. 1, pp. 73–84, 2013.

[22] D. Grabiner. (2004). The sabermetrics manifesto. [Online]. Available: http://seanlahman.com/baseball-archive/

[23] J. Fewell, D. Armbruster, J. Ingraham, A. Petersen, and J. Waters, "Basketball teams as strategic networks," *PloS one*, vol. 7, no. 11, p. e47445, 2012.

[24] O. Cliff, J. Lizier, R. Wang, P. Wang, O. Obst, and M. Prokopenko, "Towards quantifying interaction networks in a football match," in *RoboCup 2013: Robot Soccer World Cup XVII*. Berlin, Germany: Springer, 2013.

[25] D. Budden, P. Wang, O. Obst, and M. Prokopenko, "Simulation leagues: Analysis of competition formats," in *RoboCup 2014: Robot Soccer World Cup XVIII*. Berlin, Germany: Springer, 2014.

[26] H. Akiyama and T. Nakashima, "HELIOS Base: An open source package for the RoboCup soccer 2D simulation," in *RoboCup 2013: Robot World Cup XVII*. Berlin, Germany: Springer, 2014, pp. 528–535.

[27] D. G. Hurley, D. M. Budden, and E. J. Crampin, "Virtual reference environments: A simple way to make research reproducible," *Briefings in Bioinformatics*, 2014.

David M. Budden, The University of Melbourne, Parkville, Victoria, Australia. E-mail: david.budden@unimelb.edu.au.

Peter Wang, Commonwealth Scientific and Industrial Research Organisation, Marsfield, Australia. E-mail: peter.wang@csiro.au.

Oliver Obst, Commonwealth Scientific and Industrial Research Organisation, Marsfield, Australia. E-mail: Oliver.Obst@csiro.au.

Mikhail Prokopenko, The University of Sydney, New South Wales, Australia. E-mail: mikhail.prokopenko@sydney.edu.au.

Humanoid Robots in Soccer

Robots Versus Humans in RoboCup 2050

©ISTOCKPHOTO.COM/AKSONOV

By Reinhard Gerndt, Daniel Seifert, Jacky (Hansjoerg) Baltes,
Soroush Sadeghnejad, and Sven Behnke

This article describes the history and major achievements of the RoboCup Humanoid League from its start in 2002 to today. Furthermore, it gives an indication on how the league may evolve over the coming years until 2050, when a team of autonomous humanoid robots shall play soccer against the human world champion. We show how the competition drives humanoid robot research and serves as a benchmark to measure progress.

RoboCup

RoboCup is an international initiative to promote artificial intelligence and robot technology through the organization of robot competitions and scientific meetings. The stated ultimate goal of RoboCup is for a team of fully autonomous humanoid robot soccer players to win a soccer game, complying with the official Fédération Internationale de Football Association (FIFA) rules, against the winner of the most recent World Cup by the middle of the 21st century [1]. Hence, many of the competitions focus on soccer as a benchmark problem. However, RoboCup also added competitions for domestic-service robots, rescue robots, and industry-inspired mobile manipulators.

Soccer competitions started in 1997 with wheeled and simple simulated robots. The RoboCup Humanoid League was first held in 2002, when walking and kicking were the major challenges. Improvements in mechanics, electronics, perception, and control quickly led to capable individual players. After managing the basic skills, the robots started team play. In recent years, commercially available platforms have given teams the opportunity to concentrate on software only.

Digital Object Identifier 10.1109/MRA.2015.2448811
Date of publication: 11 September 2015

1070-9932/15©2015IEEE

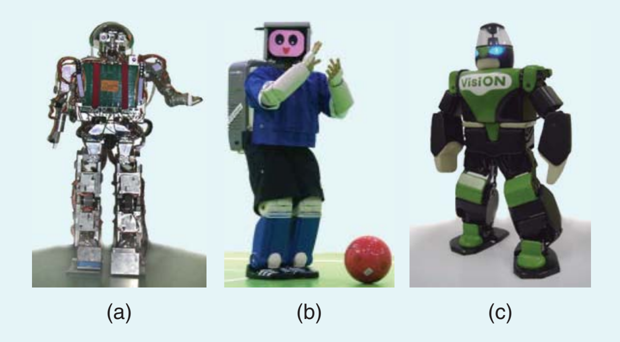


Figure 1. The early years' first-place winning RoboCup Humanoid League competitors: (a) Nagara (2002), (b) HITS (2003), and (c) VisiON (2004).

Figure 2. A 2004 penalty kick: Team Osaka versus Robo-Erectus.

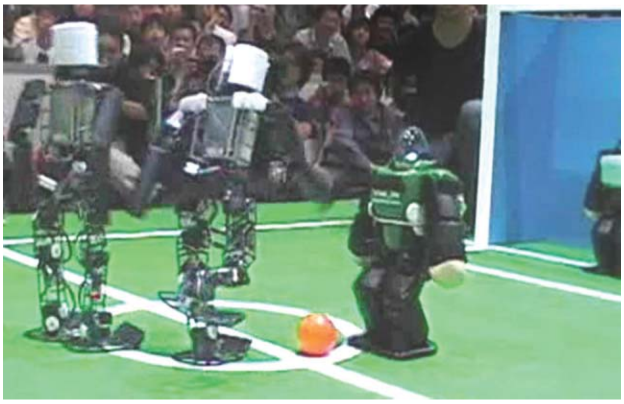


Figure 3. Two-versus-two soccer in 2005: NimbRo versus Team Osaka.

The future of the league is characterized by a strong push toward larger and more human-like robots, bigger teams, and FIFA-like rules and environments.

Early Years (2002–2004)

The first Humanoid League competition was carried out at RoboCup 2002 in Fukuoka, Japan. At that time, some impressive humanoid robots developed by the Japanese industry, e.g., Honda Asimo and Sony Qrio, existed, but these robots were not available to other research institutes. The only available commercial platform was the Fujitsu Humanoid for Open Architecture Platform series of robots, but despite high costs, these robots could not act autonomously due to their lack of onboard processing power.

Inspired by the very ambitious goal, a dozen university teams participated in the first RoboCup humanoid competition. The winning robots of the first few years are shown in Figure 1. The robot designs varied significantly within a size range of 20–180 cm. Many humanoids could not act autonomously and had to be remotely controlled or tethered due to the lack of computation and battery power. The first competition consisted of three challenges: 1) balancing on one leg, 2) making penalty kicks (Figure 2), and 3) freestyle demonstrations, graded by a panel of judges. Different degrees of autonomy were accounted for by performance factors. To encourage teams to build their own robots, commercial platforms were also penalized by a 20% performance factor.

The Humanoid League robots improved quickly, and the performance factors became obsolete. The rules evolved to provide entertaining competitions that would still be suitable as a benchmark for autonomous robots. Each year, new technical challenges were introduced to encourage the development of new skills. By 2004, all robots acted fully autonomously, and the main tournament was played as a penalty shoot-out. Standing on one leg was replaced by a walking competition, where robots had to footrace around a pole. The other technical challenges were passing and balancing across a slope.

Different capabilities, in part related to the size of the robots, required subdividing the RoboCup Humanoid League. The rules for the year 2004 characterized three classes: 1) H-40, 2) H-80, and 3) H-120, in line with a maximum size of the robots [2]. The results of the individual challenges were aggregated into a Best Humanoid ranking. However, being aware of the fact that robots of different size classes can hardly be compared directly, the winner of the Best Humanoid Award, the Louis Vuitton Cup, was determined by voting of the team leaders. The guiding principles for voting were robustness, walking ability, ball handling, and soccer skills.

From Penalty Kicks to Soccer Games (2005–2007)

After demonstration games in 2003 and 2004, two-versus-two soccer matches were introduced as a main KidSize ($<$60 cm) tournament in 2005 (Figure 3). Initially, humanoid robots were understood to primarily have bipedal kinematics. Human-like appearance and sensors were not yet part of the rules. Team Osaka was among the first to be able to move quickly and reliably across the playing field with the VStone robot that featured an omnidirectional vision system in the head. Consequently, they won the soccer competition two times in succession [3].

The larger TeenSize robots initially continued to play penalty kick, which, in 2007, evolved to the dribble-and-kick competition (Figure 4). The dribble-and-kick competition is played between a striker and the goal keeper. The striker robot starts in the center of the field, and the ball is placed randomly on the striker's goal box. It then has to move back to approach the ball, dribble the ball across the center line, and kick the ball into the opposing goal.

Furthermore, rule changes were introduced during this period. It was felt that humanoid robots should be limited to human-like sensors. This banned omnivision or vision systems with three cameras and active sensors like lidar, ultrasound, and infrared distance sensors. To have a more objective ranking, quantitative measures like goals scored and time required to perform a given task were introduced. The free demonstration event was removed from the competition.

The rapid improvements in robot capabilities also led to an increase in the complexity and diversity of the technical challenges. The technical challenges introduced during this time included walking over uneven terrain, dribbling around multiple poles, dribbling through randomly placed obstacles, and double passing.

From Individual Skills to Team Play (2008–2011)

By 2008, most teams had successfully solved the problem of locomotion and were able to walk reliably on the flat playing surface, which was green carpet. Localization and the perception of the game situation subsequently became the focus of research. Whereas individual robot skills (fast walking, getting up from a fall, fast and strong kicking) were the key to success in previous years, now team play and coordination became more important. This was further emphasized by increasing the number of KidSize players per team from two to three in 2008. Two teams from Germany (Team NimbRo [4], University of Bonn, and Darmstadt Dribblers [5], TU Darmstadt) won the competition several times.

With the availability of affordable high-power servos, the performance of the TeenSize robots improved and two-versus-two soccer matches became possible in 2010. However, the largest (>120 cm) and heaviest robots were still too fragile to survive a fall undamaged. Furthermore, with some robots weighing more than 40 kg, they posed a considerable danger to other robots and participants. As a consequence, only the smaller TeenSize robots (100–120 cm) started to play two-versus-two games in 2010, while the AdultSize robots (>130 cm) continued with dribble-and-kick competitions.

With team play becoming a focus, the potential for cross-fertilization with the simulation leagues of RoboCup has been discussed [6]. Many research groups in the Humanoid league use simulation for robot development and optimization. However, the specific requirements of the RoboCup simulation competitions lead to a stronger link with the Standard Platform League with identical robots.

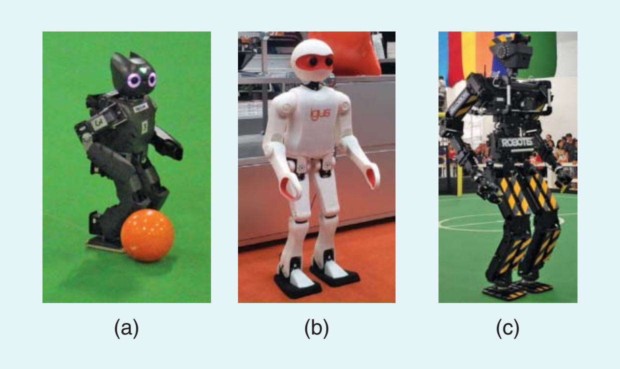
The major rule changes aimed at fostering a more robust visual perception and localization. Landmark poles in the corners of the field, and later on the sidelines, were removed. In 2010, extra lighting on the field was abandoned in favor of ambient lighting. The size of the playing field was increased, and the goals were gradually made more realistic. The blue- and yellow-colored goal back walls were removed, leaving only blue and yellow goal posts.

With increasing interest in the RoboCup, the number of participating teams in the KidSize class had to be limited to 24, and a qualification process was introduced. Teams applied by submitting a team description paper and a video of their robot playing soccer. In the video, the robot needed to demonstrate the ability to perceive and approach a ball, line up with the goal, and kick the ball into the goal. For applications to the KidSize competitions, the robot also needed to demonstrate the ability to stand up after a fall from various positions.

Figure 4. The 2009 TeenSize dribble-and-kick competition: CIT-Brains versus NimbRo.

Availability of Standard Platforms (2012–2014)

In 2011, the Korean company Robotics introduced the DARwIn-OP robot, which it had developed with Virginia Tech [7]. In 2014, 50% of the KidSize teams that submitted qualification material used the DARwIn-OP platform or based their robot on it. In 2012, a similar collaboration between Robotics and the University of Bonn was started, which resulted in the development of NimbRo-OP [8], a TeenSize humanoid robot, which has now been further developed together with igus GmbH. In 2014, Robotics developed the tactical hazardous operations robot—open platform (THOR-OP) humanoid robot [9] as a general-purpose disaster-response robot to compete in the U.S. Defense Advanced Research Projects Agency Robotics Challenge (DRC). By modifying the THOR-OP, the University of Pennsylvania RoboCup team was able to take part in the RoboCup 2014 in Brazil, where they finished first in the AdultSize subleague. The introduction of these platforms (Figure 5) had a big impact on the Humanoid League.



(a) (b) (c)

Figure 5. The recent "out-of-the-box" humanoid soccer robots: (a) the DARwIn-OP, (b) the Igus Humanoid Open Platform, and (c) the THOR-OP.

Figure 6. The KidSize soccer game during RoboCup 2014 in Brazil.

Instead of designing and building their robots from scratch, teams could now simply purchase a robot platform that was able to walk and kick a ball and recover from a fall. This made qualification and entry into the league much easier for new teams. However, all robots were open platform such that, unlike in the RoboCup Standard Platform League, which uses standardized robots, Humanoid League robots could be altered by the teams. In addition, however tempting the use of off-the-shelf robots was, many teams still worked on individual hardware solutions, for example, using two knee actuators to increase the speed of walking or parallel kinematics to increase the stability.

The major rule changes for the 2013 tournament were coloring both goals yellow and omitting the previously used landmark poles [10]. This made the field fully symmetrical and increased the difficulty of robot localization. At the end of the 2013 tournament, the RoboCup board of trustees issued a challenge to all leagues as they felt that progress in the leagues had been limited to incremental improvements rather than consequently aiming for the 2050 goal. In response, the maximum height of the robots in the KidSize was raised by 50% to 90 cm [11]. Furthermore, the height limits of the Kid- and TeenSize and the Teen- and AdultSize classes were chosen with an overlap on the upper and lower size limits to foster easier transition toward larger robots.

The changes were adopted fast. Many KidSize teams started to experiment with larger robots. The size range of 2014 KidSize robots is shown Figure 6. Furthermore, the field area for KidSize was increased by 125% to 6 m × 9 m, and the size of the goals, and the size and weight of the ball were adjusted to accommodate the larger robots. The number of KidSize players was increased to four robots per team. The complexities of the technical challenges also increased. In the AdultSize dribble-and-kick competition, two obstacles, representing stationary opposing players that must be avoided by the striker robot, have been introduced.

RoboCup Humanoid League Achievements (2010–2015)

The main achievements of the RoboCup Humanoid League are building a community of robotics researchers and fostering research in the field of humanoid robots. Figure 7 shows most of the Humanoid League teams that participated in RoboCup 2013. The development of the community can be inferred from the numbers. Records of qualified teams in the Humanoid League competitions are available from the year 2005 onward (Figure 8). In 2006, the TeenSize subleague was introduced; the AdultSize followed in 2010. New subleagues initially recruited their members from existing ones. Currently, the numbers stabilized at around 39 qualified teams for all three subleagues.

Figure 7. The teams of the Humanoid League at RoboCup 2013 in Eindhoven, The Netherlands. (Photo courtesy of RoboCup Humanoid League.)

In step with the RoboCup competition in general, the maximum number of teams that can reasonably be supported within the current limits on infrastructure, e.g., number of playing fields and space for the teams, has been reached. This is especially true in the KidSize competition, where a limit of 24 fully qualified teams plus a few additional teams (usually one or two) qualified for the technical challenges was introduced. The teams are qualified from a group of about 31 applications every year. This number has remained fairly constant throughout the years. However, for 2015, the number of KidSize teams will decrease slightly. Records of the geographic origin of teams over recent years show a significant involvement from countries like China, Germany, Iran, Mexico, Taiwan [Republic of China (ROC)], and the United States (Figure 9). Some countries have a stable contribution, e.g., Germany with four to five teams every year. However, often the individual participation appears to be subject to the host country of RoboCup. Teams report travel costs and logistics as increasingly relevant aspects of participation. Overall, some locations, such as The Netherlands (2013) and Istanbul (2011), had slightly more participating teams than other locations, such as Mexico (2012). However, the influence is rather minor, leading to a variance of about two teams per size class.

As in regular soccer, statistics on goals in RoboCup humanoid robot soccer exist (Figure 10). The number of goals may be considered as a suitable general performance indicator, but the Humanoid League constantly adopts the rules toward the 2050 game. One would, therefore, expect to have an increasing average goal count that drops after the introduction of new rules. However, goal statistics show only a weak correlation with rule changes. For example, when increasing the field size for TeenSize in 2011, there was a drop in average goals. When doing the same change in KidSize in 2014 with, otherwise, similar conditions, the average number of goals actually slightly increased. Then again, not observing a similar drop in AdultSize, in 2012, when field size was increased for this subleague, can be explained by the specific structure of the dribble-and-kick competitions with a single robot in each team.

The consequences of the rule change of abandoning blue- and yellow-colored goals in 2013 are also not reflected clearly in the average goal count. Upon introduction, it was discussed if this change would result in less successful strikers and a reduced goal count or in an increased goal count due to more own goals. The drop in the average number of TeenSize goals in 2013 indicates that the strikers may be less successful. However, the drop in average goals in the KidSize subleague is only minor, if statistically significant at all, for the respective year. The authors expect other underlying influences to exist. With typically more experienced teams in the TeenSize subleague, own goals may not have played a significant role, unlike in KidSize, where the drop in proper scoring was mostly compensated by own goals. However, no records exist to support the explanation. The AdultSize goals do not show a similar effect, which again can be explained by playing on a single goal in this subleague. The Humanoid League also introduced a number of technical improvements to robotics. The team NimbRo has been working intensely on the stability of walking and contributed the concept of capture steps to keep robots from falling after bumping into each other [12]. Other examples are the design of a series elastic actuator add-on to the widely used Dynamixel servos, which was presented by a joint team from Universidade Federal do Santa Maria in Brazil and Ostfalia University from Germany [13]. The elastic element was intended to absorb shocks, store energy, and possibly, with an additional displacement sensor, allow for dynamic gait in the future.

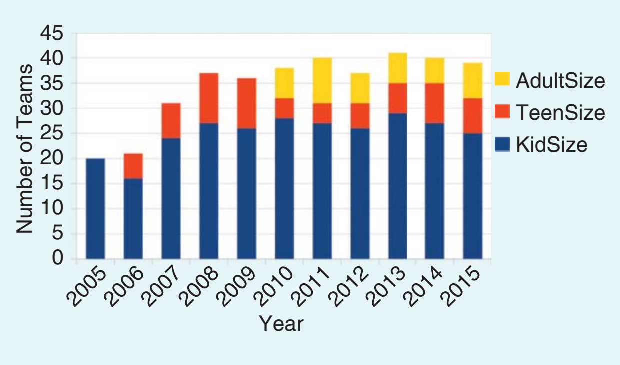


Figure 8. The number of qualified teams for RoboCup World Cup.

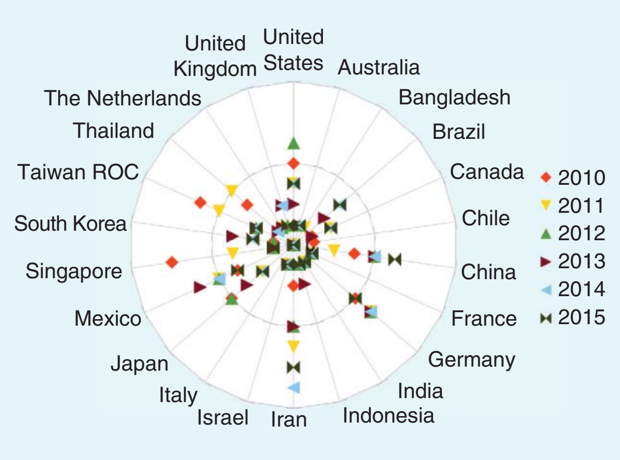


Figure 9. The distribution of teams by country in the RoboCup Humanoid League.

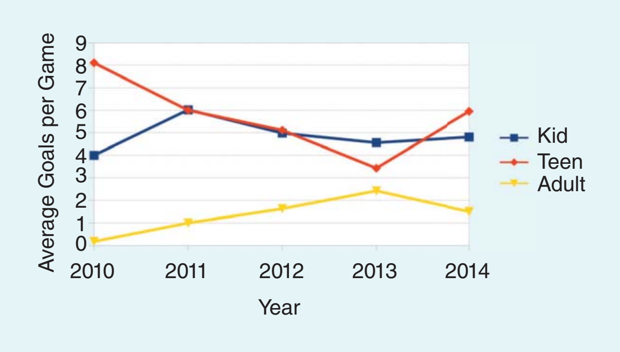


Figure 10. The average goals scored per game in the RoboCup Humanoid League.

Furthermore, the element may introduce passive compliance to robots, helping to survive falling and possibly help avoid harming humans during interaction. Another novelty was evaporation cooling of drives introduced by team Sweaty from the Offenburg University of Applied Sciences (Figure 11).

Future of the Humanoid League (2015–2050)

As the capabilities of the robots improved, the RoboCup Humanoid League started playing with smarter and larger robots that become increasingly similar to human players in their kinematics, dynamics, and sensing. However, with three to five years for every robot generation to be developed and mature, only seven to 12 generations of robots remain until the game against the human soccer champion in the year 2050. Relating this to the time a team of humans may require to advance from an entry-level to premier league may underline the overall ambition of the project.


(a) (b)

Figure 11. Details of the new robots at RoboCup 2014: (a) a series elastic actuator and (b) an evaporation cooling system.

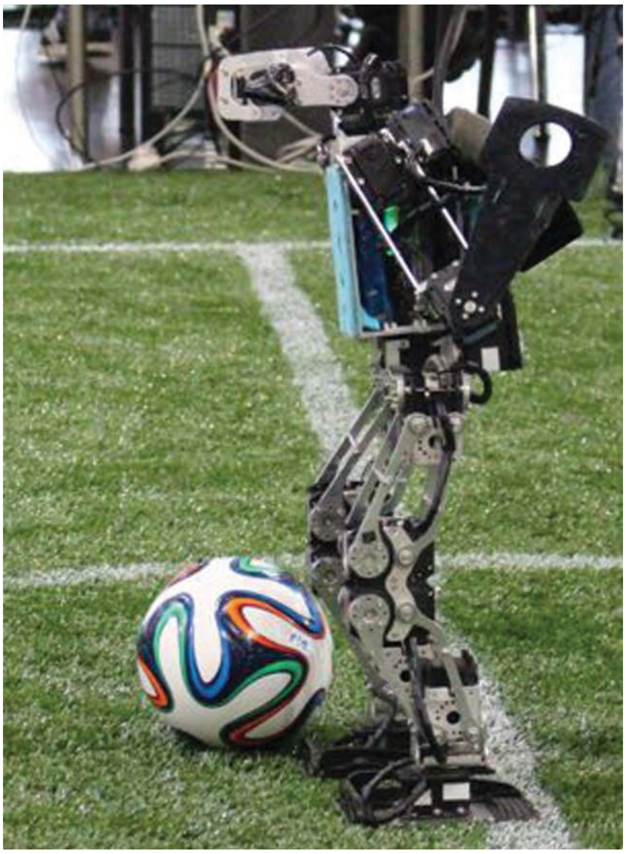


Figure 12. A robot at the RoboCup German Open 2015 on artificial grass using a size 1 FIFA ball. (Photo courtesy of Rohow2014.)

Urgent targets for further improvements are the compliance and energy efficiency of the robots. The use of compliance in control and construction of the actuators and links as well as soft materials on the outer shells will be necessary for improved soccer capabilities, such as running, falling, high-speed kicking, and safe robot–robot and human–robot physical interaction [14]. Currently, the robotic soccer games are only 20 min, split into two 10-min periods, due to the limited capacity of the batteries in relation to the relatively poor power-to-weight ratio of the servo motors. Furthermore, few of the robots are able to use the inherent dynamics of motion (e.g., the swing leg needs to be actively driven rather than swinging freely, because of the friction in the gear box) or to store energy in springs or other mechanics. The targeted improvements strongly link the RoboCup activities with leading new topics in the robotics community. For example, considerations such as more efficient movement and soft materials are also reflected by a number of recent technical committees of the IEEE Robotics and Automation Society (RAS) [15], like the ones on Human Movement Understanding and Soft Robotics.

Moving toward larger robots also comes with a number of organizational implications for the competitions. Designing, building, and sustaining a full team of robots will become increasingly hard, if not impossible, for a single team of engineers. Furthermore, the entry-level requirements for new teams would increase significantly. The organizers plan to establish rules and procedures to encourage cooperation between teams, like between the University of Manitoba and Amirkabir University of Technology (Tehran Polytechnic) [16]. There have already been several initiatives directed at creating suitable communication protocols and infrastructure that will allow robots from different teams to play together effectively. Team FUmanoids received a RoboCup Federation Grant in 2012 and developed a common communication platform for humanoid robots.

Furthermore, many Humanoid League teams have released their source code and hardware designs [17]. However, the benefit of those contributions is much less immediate. First, the teams often use different hardware platforms, so inverse kinematics, gaits, device drivers, and low-level controllers often require significant adaptations for different robots. Second, even higher-level functionality in the software is implemented using different and often custom middleware. There are now several initiatives to implement soccer robot middle-ware for important modules such as vision, localization, walking engine, and communication. The robot operating system (ROS) is a popular candidate to simplify interoperability of software developed by different teams. Improved computational power on the robots and more efficient implementations of the ROS stack now allow to consider this option for mobile autonomous robots.

The rules for 2015 follow the Humanoid League road map [17] toward more natural playing fields and environments.

Color coding of the environment is completely abandoned, except for cyan and magenta team colors. The previous years' technical challenge of playing with an arbitrary ball now found its way into the regular games. Unlike the early orange balls, the balls are now specified according to FIFA rules, with a 50% minimum of white. The size of the ball for the KidSize also was increased to FIFA size 1, which is the smallest available official soccer ball. It is used as a so-called skill ball in real soccer training. AdultSize already uses regular-sized soccer balls. Another major advancement is changing the playing surface to artificial grass. This decision has significant implications for walking and ball handling. Active balancing and uneven-terrain walking will become more important for the robots (see Figure 12).

The catalog of technical challenges moves ahead even further in 2015. Push recovery, i.e., avoiding a fall after contact between two players, will become an increasingly important capability as the number and speed of the robots increases and collisions between players are more likely to occur. The high-kick challenge has been around for some time now. However, with larger and heavier balls in the KidSize, it again needs attention by the teams. The new playing field surface with larger friction and possible deviation of the ball's course is expected to further motivate high kicks in regular games [18]. A receive-and-kick exercise is expected to address vision capabilities. A high-jump challenge, expecting the robots to safely land on their feet, shall be the first challenge with a strong dynamic flavor. For the first time in RoboCup competitions, robots will intentionally have a short flying phase. With this being a challenge on its own, a controlled landing will be required. The high jump is expected to be the first step to having robots run in the game.

For future RoboCup competitions, even more advanced technical challenges are envisioned. Walking on natural grass in the open requires sophisticated balancing and vision skills, as well as suitable hardware. Balancing will be an issue, especially, for the early phase, when robots of half the size and significantly lower weight than humans have to walk on grass. Furthermore, more dynamic game play is aimed at with a throw, receive, and kick challenge. For this, a robot should lift up the ball from the ground throw it toward a teammate and have the ball kicked toward the goal.

Some small aspects, however, still require further research, some of them more for organizational than technical reasons. Listening to the referee's whistle is an example of this. While in principle listening to a whistle is feasible, at the competitions with playing fields close to each other and multiple games going on at the same time, the signals of two adjacent fields may not be clearly distinguishable, bearing in mind that the next field's referee may be closer to a robot other than the one leading the game.

For the future, the road map projects five-year intervals for major rule changes. In 2020, the minimum size of the robots is expected to be raised to 60 cm. Furthermore, the field size is expected to increase to 20 m, the number of players to six, and the duration of the game to two 20-min periods. Further changes are planned for 2025. For 2030, the theme is "It's time to play against humans," and a technical challenge to outrun the president of the RoboCup as well as competitive games against a team of eight human nonprofessional players are foreseen. For 2040, full compliance with FIFA rules will be reached.

Humanoid Soccer Workshops, Schools, and Publications

The Humanoid League fosters development through the organization of competitions, and also has a strong focus on advancing research via publications, workshops, and schools. Research and development activities are regularly published in high-quality journals. The community contributes to the annual RoboCup International Symposium and major robotics conferences like the IEEE/Robotics Society of Japan International Conference on Intelligent Robots and Systems and the IEEE International Conference on Robotics and Automation. In addition, members of the league contribute to the organization of and the submission to the annual humanoid soccer workshop, which has been organized since 2006 at the IEEE-RAS International Conference on


(a)

(b)

Figure 13. The participants of Humanoid Soccer Schools in (a) Hamburg, Germany, 2014, and (b) Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran, 2014.

Humanoid Robots, the flagship conference for humanoid robotics research.

Since 2012, members of the Humanoid League organized a number of humanoid soccer schools and workshops (see Figure 13). These events provide unique opportunities for researchers and hobbyists alike to learn from some of the leading experts in the field. In contrast to scientific conferences, the humanoid soccer schools include practical components. A considerable amount of time is made available to students to complete exercises and/or test their own ideas on real systems. The humanoid soccer schools also include a series of social events to foster collaboration between the teams.

Human-like appearance and sensors were not yet part of the rules.

All these scientific activities ensure that

- the research developed as part of the RoboCup initiative is widely disseminated to other researchers
- researchers learn about the latest research results from other humanoid robotics researchers
- new teams have a starting point for their research.

Conclusions

This article illustrates the development of the RoboCup Humanoid League community and how the league fosters advancements in humanoid robotics. It also gives an outlook on the developments of the capabilities of humanoid soccer-playing robots, rules, and forms of organization for the competitions yet to be expected.

Acknowledgments

We would like to thank previous and current members of the RoboCup Humanoid League community for their input during many years of rule discussions and development. In particular, we would like to thank the other members of the RoboCup Humanoid League technical and organizing committees, Luis F. Lupian, Marcell Missura, Sean Luke, Maike Paetzel, Hafez Farazi, Bente Reichardt, and RoboCup trustee Oskar von Stryk.

References

[1] H. Kitano and M. Asada, "The RoboCup humanoid challenge as the millennium challenge for advanced robotics," *Adv. Robot.*, vol. 13, no. 8, pp. 723–737, 2000.

[2] C. Zhou, "Rules for the humanoid league," *Adv. Robot.*, vol. 18, no. 7, pp. 721–724, 2004.

[3] R. Matsumura and H. Ishiguro, "Development of a high-performance humanoid soccer robot," *Int. J. Humanoid Robot.*, vol. 5, no. 3, pp. 353–373, 2008.

[4] S. Behnke and J. Stückler, "Hierarchical reactive control for humanoid soccer robots," *Int. J. Humanoid Robot.*, vol. 5, no. 3, pp. 375–396, 2008.

[5] M. Friedmann, J. Kiener, S. Petters, D. Thomas, O. von Stryk, and H. Sakamoto, "Versatile, high-quality motions and behavior control of a humanoid soccer robot," *Int. J. Humanoid Robot.*, vol. 5, no. 3, pp. 417–436, 2008.

[6] N. M. Mayer and M. Asada, "RoboCup humanoid challenge," *Int. J. Humanoid Robot.*, vol. 5, no. 3, pp. 335–351, 2008.

[7] I. Ha, Y. Tamura, H. Asama, J. Han, and D. W. Hong, "Development of open humanoid platform DARwIn-OP," in *Proc. IEEE SICE Annu. Conf.*, 2011, pp. 2178–2181.

[8] M. Schwarz, J. Pastrana, P. Allgeuer, M. Schreiber, S. Schueller, M. Missura, and S. Behnke, "Humanoid TeenSize open platform NimbRo-OP," in *Proc. 17th RoboCup Int. Symp.*, Eindhoven, The Netherlands, 2013, pp. 568–575.

[9] S.-J. Yi, S. McGill, L. Vadakedathu, Q. He, I. Ha, M. Rouleau, D. Hong, and D. Lee, "Modular low-cost humanoid platform for disaster response," in *Proc. IEEE/RSJ Int. Conf. Intelligent Robots Systems*, 2014, vol. 965, pp. 14–18.

[10] RoboCup Humanoid League Technical Committee. (2013). RoboCup soccer humanoid league rules and setup for the 2013 competition in Eindhoven, [Online]. Available: https://www.robocuphumanoid.org

[11] RoboCup Humanoid League Technical Committee. (2014). RoboCup soccer humanoid league rules and setup for the 2014 competition in Joao Pessoa. [Online]. Available: https://www.robocuphumanoid.org

[12] M. Missura, C. Münstermann, P. Allgeuer, M. Schwarz, J. Pastrana, S. Schueller, M. Schreiber, and S. Behnke, "Learning to improve capture steps for disturbance rejection in humanoid soccer," in *RoboCup 2013: Robot World Cup XVII*. Berlin, Germany: Springer, 2014.

[13] L. Martins, R. Pretto, R. Gerndt, and R. Guerra, "Design of a modular series elastic upgrade to a robotics actuator," in *Proc. RoboCup2014 Symp.*, 2014, pp. 701–708.

[14] S. Haddadin, T. Laue, U. Frese, S. Wolf, A. Albu-Schäffer, and G. Hirzinger, "Kick it with elasticity: Safety and performance in human–robot soccer," *Robot. Auton. Syst.*, vol. 57, no. 8, pp. 761–775, 2009.

[15] (2015, June). RAS Technical Committees. [Online]. Available: http://www.ieee-ras.org/technical-committees

[16] T. A. Shangari, F. Shamshirdar, M. H. Heydari, S. Sadeghnejad, J. Baltes, and M. Bahrami, "AUTUofM humanoid teensize joint team; a new step toward 2050's humanoid league long term roadmap," in *Robot Intelligence Technology and Applications*, 3rd ed. Berlin, Germany: Springer, 2015, pp. 483–494.

[17] (2015). RoboCup Humanoid League webpage. [Online]. Available: https://www.robocuphumanoid.org

[18] J. Carstensen, S. Krupop, and R. Gerndt, "Robotic legs—Parameters for a human-like performance," in *Proc. 6th Workshop Humanoid Soccer Robots@ Humanoids Conf.*, Bled, Slovenia, Oct. 2011.

Reinhard Gerndt, Ostfalia University of Applied Sciences, Wolfenbuettel, Germany. E-mail: r.gerndt@ieee.org and r.gerndt@ostfalia.de.

Daniel Seifert, Freie Universität Berlin, Germany. E-mail: dseifert@inf.fu-berlin.de and daniel.seifert@fu-berlin.de.

Jacky (Hansjoerg) Baltes, University of Manitoba, Winnipeg, Canada. E-mail: jacky@cs.umanitoba.ca.

Soroush Sadeghnejad, Amirkabir University of Technology, Tehran, Iran. E-mail: S.Sadeghnejad@aut.ac.ir.

Sven Behnke, University of Bonn, Germany. E-mail: behnke@cs.uni-bonn.de.

IMAGE LICENSED BY GRAPHIC STOCK

Model-Driven Software Engineering in Robotics

Models Are Designed to Use the Relevant Things, Thereby Reducing the Complexity and Cost in the Field of Robotics

By Davide Brugali

A model is an abstract representation of a real system or phenomenon [1]. The idea of a model is to capture important properties of reality and to neglect irrelevant details. The properties that are relevant and that can be neglected depend on the purpose of creating a model. A model can make a particular system or phenomenon easier to understand, quantify, visualize, simulate, or predict.

Models and Modeling

Models and modeling are an essential part of every engineering endeavor. Using models to design complex systems such as edifices, airplanes, and production plants reduces the risk of making costly errors before undertaking the effort of their realization. Models also help us understand the various aspects of a complex problem and evaluate alternative solutions.

Two models of the same system or phenomenon may be essentially different depending on the properties they aim to capture. Specialized models are needed to visualize the structure of a bridge or to evaluate an aircraft aerodynamics. Models might have different forms: 1) graphical (i.e., the architecture of a building) or 2) textual (i.e., the differential equations describing the effects of an earthquake on buildings).

Models are created by using modeling languages, i.e., artificial languages that define symbols, keywords, and their semantic and syntactic rules. Mathematics, statistics, and logic are typical scientific and engineering modeling languages. Domain-specific languages (DSLs) have been defined for a variety of technological domains, such as electrical circuits (i.e., the topological description language) [2], physical dynamic systems (i.e., the bond graph representation) [3], and manufacturing control (i.e., function blocks diagram) [4].

Software systems, which are often among the most complex engineering systems, can benefit greatly from using models and modeling techniques [5]. This is because a software model and the software system, of which the model

Digital Object Identifier 10.1109/MRA.2015.2452201
Date of publication: 11 September 2015

1070-9932/15©2015IEEE

is an abstraction, have the same nature. The transition of design into implementation is a sequence of refinement steps without semantic gaps. Backtracking is simple, since there is no metal to bend and there are no circuits to solder.

The term *model-driven engineering* (*MDE*) [6] is typically used to describe the software development approaches in which the abstract models of the software systems are created and systematically transformed to concrete implementations. A key premise behind the MDE is not merely that the software models serve the purpose of documentation but that they enable 1) automatic assessment of system-level quality aspects, such as safety, correctness, and performance, way before any implementation and 2) automatic generation of source code and system configuration parameters.

The idea of a model is to capture important properties of reality and to neglect irrelevant details.

During recent years, several approaches have been proposed in the literature that exploit the MDE technologies in robotic software development. Most of these approaches use standard and general-purpose software modeling languages (e.g., unified modeling language (UML) [7]). Other approaches are based on robotics-specific languages. The aim of this article is not to survey all of them, but to analyze the role and use of the MDE technologies in robotic software engineering. The "Software Control Systems for Autonomous Robots" section illustrates peculiar characteristics of software systems for autonomous robots.

Software Control Systems for Autonomous Robots

Autonomous robots are versatile machines equipped with a rich set of software functionalities, typically deployed on a distributed computing infrastructure with stringent resource constraints, for interacting purposefully and in real-time with an open-ended environment through sensors and actuators.

The use of models and modeling languages can have a high impact on the product life cycle of robotic software systems, mostly because they can embed the knowledge of experts in multiple scientific and technological domains (e.g., mechanics, control, and cognitive sciences) and support the automatic transition from the problem space (the robotic requirements) to the solution space (the software implementation).

Real-Time Embedded Software Systems

A robot control application is an embedded software system that is specialized for the particular hardware it runs on and has time and memory constraints. In general, for an embedded system, we are interested in modeling the interaction of the system with the surrounding environment. This interaction is characterized by the real-time execution of control activities that should react timely to sensory stimuli and produce commands to the actuators. Typically, in any large-scale embedded system, several control activities are executed concurrently on the same computational unit or on a distributed networked system. Concurrency implies communication among control activities, which exchange data and events, and requires careful design of their interleaving and synchronization to avoid anomalous behaviors.

Distributed Software Systems

The computational hardware of an autonomous robot is interfaced to a multitude of sensors and actuators, and has severe constraints on computational resources, storage, and power. Computational performance is a major requirement, since autonomous robots process large volumes of sensory information and have to react in a timely fashion to events occurring in the human environment.

To meet these highly demanding requirements, the computing infrastructures of the advanced autonomous robots have evolved from single processor systems to networks of general-purpose computers, microcontrollers, smart networked sensors, and actuators, introducing great flexibility in robot capabilities construction, but at the same time, making software development more challenging.

Contextually, a variety of software frameworks have been specifically designed for simplifying the implementation of robot control systems (see [8] for a survey). They offer mechanisms for real-time execution, synchronous and asynchronous communication, data flow and control flow management. These frameworks are supported by the MDE toolchains that enable the semiautomatic transition from high-level design of the system functionalities to their actual implementation and deployment on top of specific runtime infrastructures.

Rich Functionalities

Differently from other embedded systems (e.g., cars, medical devices, and so on), the control system of an autonomous robot is characterized by a large variety of functionalities (e.g., motion planning and control, perception, task planning, and so on) that together realize complex robot capabilities, such as navigation and manipulation.

Robot functionalities are conveniently implemented as software components that can be assembled in many different ways, like reusable building blocks, according to specific application requirements. A robot control system is composed of tens of components (e.g., see the ROS repositories [9]). For each component, tens of different implementations may be available (e.g., different algorithms for obstacle avoidance).

For these systems, a critical development phase is the design of the software architecture, which represents the partitioning of the control system into parts, with specific relationships among the parts, and makes the set of parts work together as a coherent and successful whole [10]. As such, the software architectures define the rules and constraints that determine the overall behavior of the control system and that guarantee system dependability and safety.

The MDE approaches support architecture design by automating some complex and error-prone tasks, such as editing diagrams, reverse-engineering legacy systems, and, more importantly, validating assumptions made by system engineers and control engineers about system properties such as schedulability, performance, responsiveness, and fail-safe behavior.

Versatile Machines

Robots are versatile machines that are increasingly being used not only to perform dirty, dangerous, and dull tasks in manufacturing industries, but also to achieve societal objectives, such as enhancing safety in transportation and reducing the use of pesticides in agriculture. In this scenario, the cost of creating new robotics products is significantly related to the complexity of developing software control systems that are flexible enough to easily accommodate frequently changing requirements: 1) more advanced tasks in highly dynamic environments, 2) in collaboration with unskilled users, and 3) in compliance with changing regulations.

Recent initiatives aim at developing the MDE approaches that simplify the static and dynamic reconfiguration of a robot control system according to specific application requirements and operational conditions.

System Architecture Modeling and Analysis

Software architecture design is a multidimensional decision-making process and different software models are needed to describe the system from multiple perspectives, such as structure, behavior, and nonfunctional properties.

The object management group (OMG) [11], a not-for- profit technology standards consortium founded in 1989, defines and maintains the specification of the UML [7], a semiformal general-purpose graphical language for modeling software systems. The UML 2.5 is composed of 14 standard diagram types, which are classified as structure diagrams or behavior diagrams. Structure diagrams show the static structure of a system in term of parts and relationships among the parts on different abstraction and implementation levels. Behavior diagrams show the dynamic behavior of a system, i.e., how it changes over time.

In particular, the UML 2.5 component diagram defines a standard graphical notation for documenting architectural representations that emphasize the runtime computational elements (also known as *components*) of software systems and their communication channels (also known as *connectors*). In [10], these representations are referred to as the component and connector architectural viewtype.

For example, Figure 1 shows a simplified version of a software architecture for mobile robot navigation. It is composed of elemental components and composite components. Elemental components (e.g., GlobalPlanner, BaseController) are graphically represented by boxes marked with the <<component>> stereotype. Similarly, composite components (e.g., equipment) are marked with the <<subsystem>> stereotype.

Each component is characterized by a set of ports, which represent distinct interaction points with other components. Each port can be associated with a number of interfaces. A provided interface describes the features that constitute a coherent service provided by a component. Similarly, a required interface describes the dependence of a component on the type of service that should be provided by another component. A connector represents a communication link between two or more components. Different types of connectors can be distinguished by labeling the association link with a stereotype. For example, in Figure 1, the stereotype <<C/P>> refers to the caller/provider communication paradigm (i.e., components interact through synchronous invocation of services), while the <<P/S>> refers to the publisher/subscriber communication paradigm (i.e., components interact through asynchronous messages). Guidelines to document the component and connector architectural viewtype in the UML can be found in [12]. The main purpose of this kind of diagram is to document the software architecture from a functional point of

Models and modeling are an essential part of every engineering endeavor.

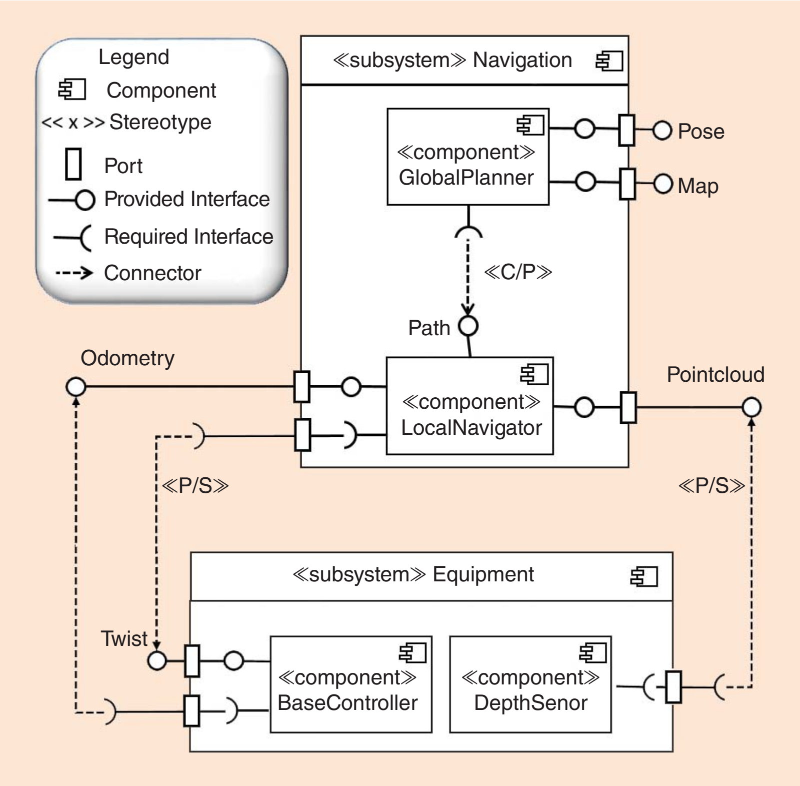


Figure 1. The UML component model of the navigation architecture.

view and serve as a vessel for communication between the varied worlds of often nonsoftware-technical stakeholders on one hand and software engineers on the other hand. Such a documentation promotes software reuse by exhibiting stable structures recurrent in many systems and facilitates maintenance by clarifying the impact of changes [13]. A large variety of UML-compliant software tools support diagram editing, differencing, merging for system design, and documentation (see [14] for a survey).

Other approaches are based on robotics-specific languages.

Unfortunately, the UML diagrams are not effective in capturing and representing nonfunctional properties of embedded, concurrent, and real-time software systems, such as the timing constraints of system functionalities, the capabilities of the (often distributed) communication infrastructure, and the allocation of threads and processes to different processors.

Researchers have faced the limitations of the semiformal notation of the UML by defining specific extensions of the UML standard (called profiles) or specialized architectural modeling languages. These approaches are exemplified in the next two sections. A comparison of several modeling languages for embedded and real-time systems can be found in [15].

UML profile for Embedded and Real-Time Systems

The UML profiles are an extension mechanism provided to allow adaptation and customization of the UML notation by adding ad hoc semantic and constraints and introducing terminology that is specific to a particular domain, platform, or method. In particular, the OMG has developed the modeling and analysis of real-time embedded systems (MARTE) [16] profile, which focuses on performance and schedulability analysis and provides stereotypes for annotating architectural models and map them into corresponding analysis domain concepts.

The high-level application modeling subprofile defines a set of stereotypes to annotate the functional model with real-time features. For example, Figure 2 shows a use of the <<rtUnit>> stereotype to annotate two computing units (i.e., ObstacleAvoider and TrajectoryFollower) of the LocalNavigator component, which perform concurrent activities, i.e., adapting the rover trajectory when an obstacle is detected and computing the twist to let the rover follow the trajectory. An annotation specifies that the former activity is aperiodic and that its relative deadline is equal to 10 ms.

The software resource modeling subprofile provides modeling artifacts to describe software multitasking application programming interfaces (API). For example, Figure 2 shows a use of the <<schedulableResource>> stereotype to annotate two concurrent tasks (i.e., FollowerTask and ObstacleTask) with the specification of their priority. The stereotype <<entryPoint>> indicates the routine (i.e., operation) executed in the context of each task.

The schedulability analysis modeling subprofile defines the stereotypes to annotate the elements of the platform model (e.g., a CPU or other device, which executes functional steps) with nonfunctional properties, such as schedulability metrics, interrupt overheads, and utilization of scheduling processing. In particular, in Figure 2, the stereotype <<SaExecutionHost>> represents any kind of processing resource (e.g., POSIX threads) and contains a property ISRswitchTime that can be used to represent the worst context switching time.

Once the application model has been annotated with MARTE stereotypes for real-time features, it needs to be converted in a software model that can be processed by tools for schedulability and performance analysis. In [17], the author proposes an automatic translation technique from MARTE models into input for modeling and analysis suite for real-time applications, which is a state-of-the-art schedulability analysis tool used in the academia. A list of tools related to MARTE can be found in [18].

A robotic example of using the MARTE for schedulability and performance analysis can be found in [19].

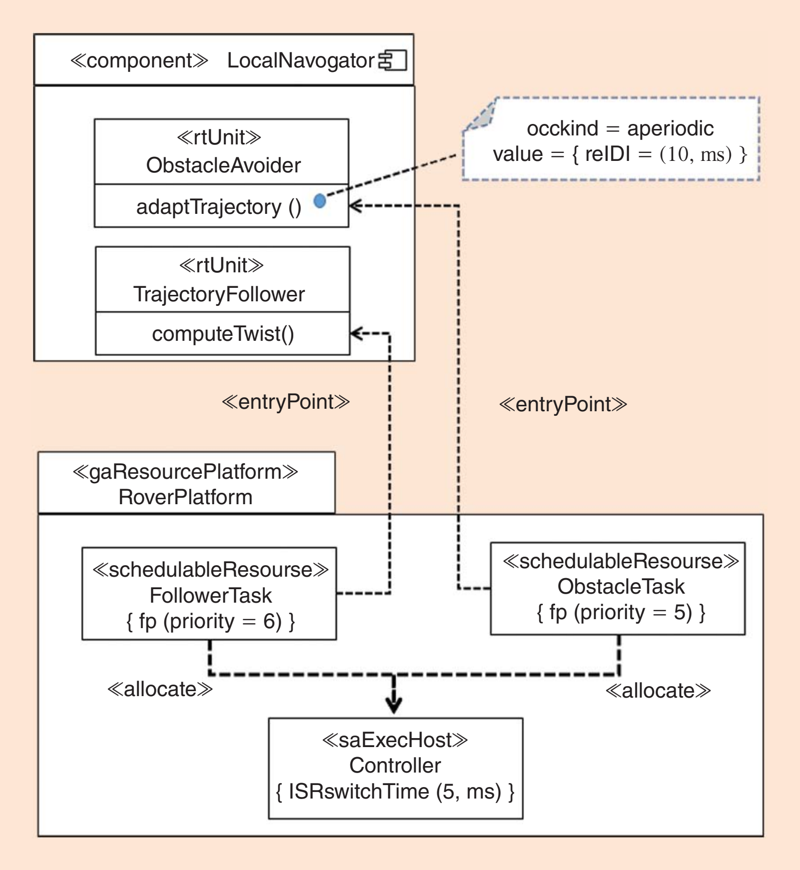


Figure 2. An example of using the MARTE stereotypes for schedulability analysis.

Architecture Analysis and Design Language

For cyberphysical systems and, in particular, for robotic systems, the response time analysis is of paramount importance. The response time of an embedded system depends on the latency between receiving an input from the sensors and producing an output to the actuators. For example, safety depends on the latency of responding to the detection of an obstacle along the robot path.

Such analysis requires modeling of an application in terms of both the computational units and of the communication channels. In particular, in a distributed embedded real-time (DERT) system, communication may be periodic or aperiodic, event-triggered or based on data sampling. Representing all these kinds of communications in MARTE is challenging.

In several application domains, the architecture analysis and design language (AADL) [20] has proved a good candidate as a modeling language for the DERT systems.

The AADL is a textual and graphical language with precise execution semantics for modeling the architecture of embedded software systems and their target platforms.

The AADL provides modeling concepts to describe the runtime architecture of the DERT systems in terms of components, connectors, concurrent tasks, their interactions, and their mapping onto an execution platform. Recently, the OMG has provided the MARTE specification with guidelines to map its modeling entities to the AADL concepts [21].

The AADL is supported by the open source AADL tool environment (OSATE) [22], which comes with a suite of automated analysis tools. In particular, the OSATE includes a flow latency analysis tool that automatically calculates end-to-end latency, i.e., the time required for a signal to travel from the source to the sink, and verify if latency requirements are satisfied. Listing 1 exemplifies the specification of a flow for the LocalNavigator component shown in Figure 2.

Component interactions consist of directional flows from an event or data source (the laser scanner), through a communication and processing path (adapt trajectory and compute twist), to a command sink (the rover). For each modeling element, the tag features specify the component interface in terms of input and output ports.

The model specifies the latency time related to the acquisition of a point cloud, the adaptation of the current trajectory to avoid an obstacle, the generation of a twist command, and the transmission of the command to the rover actuators. Biggs et al. [23] use AADL to model the control system architecture for a safety-monitoring motorized wheelchair. The end-to-end latency analysis is performed to determine such factors as the latency in responding to the appearance of an obstacle, the response time to a failure, and the suitability of the chosen microcontroller hardware.

Behavior Interaction Priority Framework

Expensive robot missions, like Mars exploration, and safety critical systems, like driverless cars, demand formal proofs of their correct behavior, i.e., guarantees that the robot will not perform actions that lead to catastrophic consequences. Informally, a safety property stipulates that bad things do not happen during execution of a program [24].

Usually, in the context of software engineering, discrete state approaches are applied to explain the behavior of a system with respect to safety. Examples are state charts, petri nets, and labeled transition systems (LTS). In this context, a safety property asserts that the program does not exhibit bad behaviors, e.g., it never enters an undesirable state.

The transition of design into implementation is a sequence of refinement steps without semantic gaps.

A relevant issue is the compositionality of the property safety in concurrent and distributed systems. The state-of-the-art approaches are based on defining LTS models of

Listing 1. Flow Specification for the LocalNavigator

```
device laser_scanner
  features
    point_cloud: out data port;
  flows
    on_flow_src: flow source point_cloud
             {latency => 5 ms .. 5 ms;};
end laser_scanner;
process adapt_trajectory
  features
    point_cloud: in data port;
    trajectory: out data port;
  flows
    on_flow_path: flow path point_cloud->twist
             {latency => 40 ms .. 60 ms;};
  properties
    Period => 100 ms;
end obstacle_avoider;
process compute_twist
  features
    odometry: in data port; trajectory: in
    data port; twist: out data port;
  flows
    on_flow_path: flow path trajectory ->
twist
             {latency => 20 ms .. 30 ms;};
  properties
    Period => 50 ms;
end obstacle_avoider;
device rover
  features
    twist: in data port;
  flows
    on_flow_snk: flow sink twist
             {latency => 10 ms .. 10 ms;};
end rover;
```

concurrent processes, which are synchronized through actions sharing the same labels. For example, let drop represent the action in which a robot manipulator places an object in a bin transported by a mobile robot. The software process that controls the robot manipulator should issue the command to open the gripper only when both the arm and the rover are in the drop position. In terms of LTS, the drop is modeled as a possible action in the standalone behavior of both the process that controls the robot manipulator and the process that controls the mobile robot. The execution of the drop action requires simultaneous participation from both processes.

A key premise behind the MDE is that the software models do not serve merely the purpose of documentation.

The behavior interaction priority (BIP) [25] framework is an architecture definition language for component-based systems, which uses the LTS for modeling the behavior (B) of elemental components and the interaction among components. The transitions between the discrete states of the component behavior are triggered by events received through the component ports (action names). Atomic components can be assembled to form compound components by means of connectors, which specify possible interaction policy (I) between ports of atomic components. The behavior of a compound component is formally described as the composition of the behaviors of its atomic components. At the level of compound components, a set of priority rules (P) describes scheduling policies for the interactions among atomic components.

The BIP framework is supported by a set of tools for offline analysis (e.g., the D-Finder tool [26]) and an engine for online monitoring of safety properties of a component-based system. Bensalem et al. [27] present an approach to develop correct-by-construction component-based software controllers for autonomous robots by integrating the BIP framework with the LAAS architecture. The proposed approach consists of 1) developing the functional components of a robot control system using the tools of the LAAS architecture, 2) defining a corresponding BIP model for the functional components, 3) adding safety constraints into the BIP model.

Safety constraints can be encoded as BIP connectors. As an example, let us consider a mobile robot that guides the patients of a hospital toward a medical office on demand. We want to express the constraint that the robot can execute only one guidance service at a time, i.e., it does not leave the patient in the middle of a corridor to serve an incoming request.

Listing 2. A BIP Connector.

```
connector RejectNavigationRequest
  on TrajectoryPlanner.GoTo, TrajectoryPlanner.
     NewTrajectory, LocalNavigator.Status
  provided LocalNavigator.Status.done
  do {TrajectoryPlanner.GoTo = NOT-IDLE}
```

This can be done by defining a constraint on the values of the input and output ports of the TrajectoryPlanner component and the LocalNavigator component. The former receives GoTo requests and generates NewTrajectories. The latter notifies the Status of the current trajectory execution. To enforce this constraint, if the LocalNavigator has not completed the current trajectory when a new request is received, the new GoTo request is rejected with a specific error message (e.g., NOT-IDLE). Listing 2 illustrates a simplified version of the BIP syntax for this constraint. At runtime, the BIP engine acts as controller of the functional components. It prevents the robot from reaching unsafe states, even if bugs exist at the decisional level of the control architecture, and reports faults to the decisional level.

System Implementation and Integration

The MDE has become popular in many engineering domains because of its promise to bring benefits in software development, such as increased productivity and quality, because of automatic code generation from abstract models of an application. The idea behind is that the model is much simpler and thus easier to write and assess than the resulting code. This idea is valid in principle, but in practice, it remains challenging to develop the MDE environments, which fully automate the generation of code from models, as illustrated in [28].

Many state-of-the-art MDE approaches (see [29] for a classification) consist of using domain-specific code generators for transforming abstract models into executable code. Typically, a code generator is specific for a modeling language and embeds the knowledge of a specific technological platform (e.g., a programming language, a middleware framework, a runtime infrastructure, and a simulation environment).

Implementing code generators is hardly a core competence for most academic and industrial organizations that develop software technologies for specific application domains, such as robotics. Moreover, robotics is a highly change-centric domain, and new technological platforms (software and hardware) are continuously developed. This means that code generator becomes quickly obsolete.

The model driven architecture (MDA) [30] and other associated standards from the OMG are an attempt to reduce the impact of changes in technological platforms on the life cycle of software systems. The MDA enforces a clear separation of the functional architecture, called the platform independent model (PIM) of a software system, from the technological details of the specific platform used to implement it. This is achieved using model transformation techniques that convert the PIM into one or more platform-specific models (PSMs), which specify the details of how the functionality of the system uses the capabilities of the software or hardware platforms to provide their operations.

Over the last few years, the software development industry has developed a significant variety of tools that automate model-to-model and model-to-code transformations. They use standard transformation languages, which provide constructs and mechanisms for expressing, composing, and applying transformations [31]. The subsequent transformation of the PSM into executable code requires mapping the architectural model concepts to certain fixed code fragments (templates) provided by a domain framework and component library that represent the interface to the underlying platform (e.g., a distributed computing middleware).

Code templates need the software developer to fill in details, i.e., the implementation of data structures and algorithms providing specific robotic functionalities. In some cases, this code can be generated automatically from behavioral models defined with general purpose modeling languages such as state charts and petri nets, or domain-specific modeling languages such as block diagrams and bond graphs.

The following sections exemplify three approaches to code generation, which are primarily concerned with reducing the gap between the problem domain and the software implementation domain, by capturing different aspects of the robotics systems.

SmartSoft

The SmartSoft project [32] has developed a software component framework, which aims at simplifying the development of real-time and distributed control systems by standardizing the component structure and connectors.

The framework provides mechanisms for the following.

- Implementing software components according to the principles of Service Oriented Computing [33].
- Interconnecting components by means of connectors that implement a limited set of communication patterns typically found in robotic control applications.
- Dynamically reconfiguring the components behavior and interconnections according to the application task.

Each SmartSoft component can encapsulate one or several threads, which are executed in the context of a single process. Components provide services to and require services from other components. Component services exchange typed data called communication objects.

Two reference implementations of the SmartSoft component framework are currently available, one for real-time control systems based on the RealTime Application Interface for Linux (RTAI) extension [34] and the other for distributed systems based on the Common Object Request Broker Architecture (CORBA) [35]. The SmartSoft framework is accompanied by a rich library of components, which implement the most common robot functionalities.

The SmartSoft MDE toolchain provides graphical editors for designing individual components, services, and interconnections and for specifying nonfunctional properties such as task execution time, period, and resource usage. The graphical editors are based on the UML profile for SmartSoft, which defines stereotypes for designing the PIM and PSM models.

Figure 3 shows a schematic representation of the model transformation supported by the SmartSoft MDE toolchain. The upper part of the figure shows the LocalNavigator component as defined in the PIM. Here, the software developer specifies the component interface in terms of communication objects and the component structure in terms of concurrent tasks. For each task (identified by the stereotype ¡¡*Smart- Task*¿¿), the nonfunctional properties (e.g., period and wcet) show the admissible values as specified by the application requirements (e.g., the wcet should not exceed 50 ms). The transformation of the PIM into the PSM (the central part of Figure 3) requires the specification of the software and hardware target platform. Here, the abstract tasks are mapped to Linux RTAI threads, as indicated by the stereotype ¡¡*RTAI- Task*¿¿. The nonfunctional properties indicate actual values, such as the measured *wcet*.

The computational hardware of an autonomous robot is interfaced to a multitude of sensors and actuators.

Finally, the source code of the LocalNavigator is generated by customizing the template of the generic SmartSoft component with the information modeled in the PIM. The software developer needs to finalize the implementation of the component by hand coding the provided functionality. In the simplest case, the generated code can be used as a wrapper of

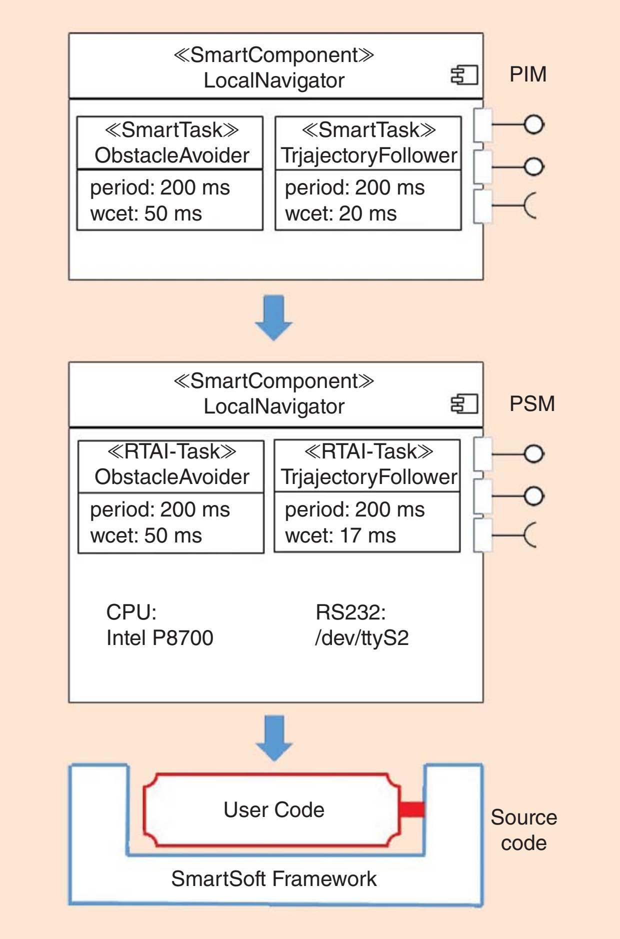


Figure 3. The model transformations in the SmartSoft MDE Toolchain.

an existing software library. The components interconnections are specified at design-time in the PIM, but can be modified at runtime to adapt the robot behavior to the actual operational conditions. For example, the ObstacleAvoider task requires sensory information (e.g., a point cloud) to detect obstacles. At run-time different sensors can provide the point cloud based on environment illumination (e.g., a laser scanner or a stereo vision system).

Autonomous robots are versatile machines equipped with a rich set of software functionalities.

The SmartSoft MDE includes the task coordination language (SmartTCL) for specifying the high-level application tasks (e.g., fetch coffee to visitors) that the robot is able to carry out and how these tasks are refined during task execution in terms of services provided by individual components (e.g., navigate towards the coffee machine). The SmartSoft framework provides the SmartTCL engine that is in charge of orchestrating the control system by activating, deactivating, and interconnecting components according to the action plan.

Proteus

The Proteus [36] project (platform for RObotic Modeling and Transformation for End-Users and Scientific communities) is an initiative of the French robotic community (GDR Robotique) that has developed a MDE toolkit for robotic system design and implementation. The distinctive feature of the Proteus Toolkit is the use of a set of ontologies [37] for representing knowledge about robotic systems, operational environments, and applications. The Proteus ontologies are an attempt to formalize the vocabulary of robotic system engineers, to allow them to model robot control architectures directly in terms of domain concepts (e.g., sensor, motion planner, rover, and so on) and not only in terms of software concepts (e.g., component, port, algorithm, and so on). Similar initiatives are sponsored by the IEEE Standard Association, which has established the Ontologies for Robotics and Automation working group.

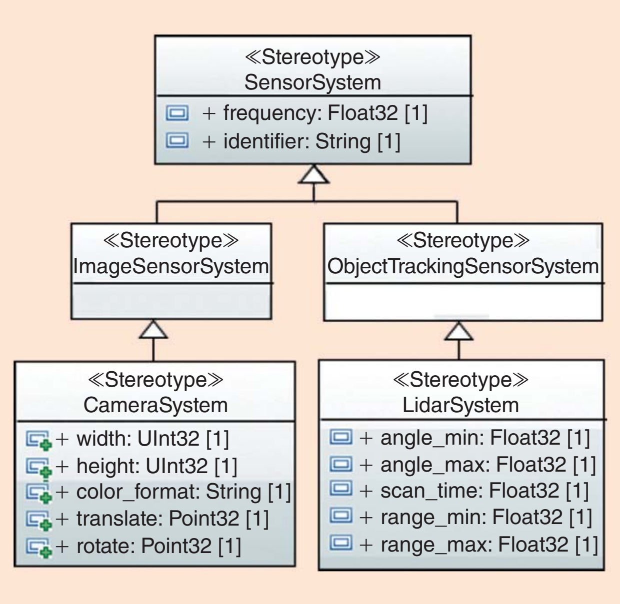


Figure 4. The RobotML profile for robot sensors.

The Proteus ontologies are the basis of the Proteus DSL (*RobotML*) [38], which includes modeling entities for specific architectural elements (e.g., Robot, SensorSystem, ActuatorSystem, LocalizationSystem) and architectural styles (e.g., reactive, deliberative, hybrid). Similar to SmartSoft (see the “SmartSoft” section), the Proteus toolchain is based on an UML profile for defining the PIM of the robot functional architecture. As an example, Figure 4 shows an excerpt of the RobotML profile for robot sensors, where the modeling entities CameraSystem and LidarSystem are defined as stereotypes that can be used to annotate the components of the robot control architecture. For code generation, the elements of the PIM model need to be allocated to an execution platform, such as a middleware and a simulator. For example, the element representing robot functionality and control activities are allocated to a component-based middleware, while the element representing the robotic equipment is allocated to a simulator.

Modeling the Components Behavior

Code generation from behavior models is a growing area of interest due to its benefits of verifying models by simulation and reducing error-prone hand-coding efforts. Several tools generate source code form the UML specifications to mainstream languages such as C, C+, and Java and to simulation languages such as SystemC. Typically, the MDE environments supporting component-based architectural modeling (see the “System Architecture Modeling and Analysis” section) also provide languages for modeling the discrete behavior of individual components, such as UML state charts and petri nets.

When developing embedded control software, control systems engineers model both the control algorithm and the system to be controlled, the so-called plant, together to ensure the optimal performance of processes with continuous dynamics. [39]. Typically, the robot control functionalities are conveniently modeled as hybrid systems, since they can specify continuous change of the system state as well as discrete transition of states. Continuous behavior can be specified using differential as well as algebraic equations. Code generation from hybrid-systems models eventually involves simulating continuous change of a variable by step-wise update of the variable based on numerical methods. This requires the model designer to assign a rate by which the continuous state evolves.

Several modeling and simulation environments for embedded systems (e.g., MATLAB/Simulink [40] and 20-sim [41]) support code generation from continuous and hybrid models, defined with DSLs, such as Bond Graphs, Block Diagrams, and Modelica. These tools allow users to define custom templates for code generation to simplify the integration of behavioral code with component frameworks. Brodskiy et al. [42] exemplify the modeling, code generation, and integration process for the motion control system of the KUKA youBot mobile manipulator [43]. The algorithm design requires a model of the system that will be controlled. They used the 20-sim toolchain to model the youBot kinematic chain as a bond graph and the BRIDE model-based toolchain [44] to model the component-based control system. The control algorithm design is performed in parallel with the design of the component architecture. These two design phases have mutual dependence and require an iterative approach.

The algorithm is partitioned in blocks of elemental functionalities with well-defined interfaces, in such a way that they can be modified independently (e.g., for improving their performance) without affecting the design and implementation of the rest of the system. On the other side, message passing between components can introduce message losses and time delays in updates and communication between sensors, controller, and actuators that need to be explicitly represented in the model of the control algorithm. Unreasonable communication requirements (such as extreme bandwidth or unachievable small latency) can lead to restructuring the component architecture. In [45], a criterion for faithful implementation of hybrid-systems models in a concurrent and distributed system is presented.

System Configuration

Current practice in software engineering for robotics consists of developing fine-grain software components, which implement single robotic functionalities. This approach is embodied by a repeated mantra among ROS developers [9]: "We don't wrap your main." The strength of this approach is the possibility to develop a large variety of different control systems by composing in multiple ways reusable software building blocks. Its weakness is the lack of support to the reuse of effective solutions to recurrent architectural design problems. Consequently, application developers and system integrators have to solve architectural design problems always from scratch. The difficult challenge consists of selecting, integrating, and configuring a coherent set of components that provide the required functionality considering their mutual dependencies and architectural mismatches.

This challenge is exacerbated by the peculiarity of the robotics domain: 1) robots can have many purposes, 2) many forms, and 3) many functions. Consequently, each robotic systems has to be configured with a specific mix of functionalities and that strongly depends on the robot mechanical structure (a rover with zero or multiple arms), the task to be performed (cleaning a floor, rescuing people after a disaster), and the environmental conditions (indoor, outdoor, underground). In various application domains, the software product line (SPL) development has proved to be the most effective approach to face this kind of challenges. A SPL is a family of applications (products) that share many (structural, behavioral, and so on) commonalities and together address a particular domain [46]. Typically, the SPL is a strategic investment for an engineering company that wants to achieve customer value through large commercial diversity of its products (e.g., different control systems for warehouse logistics) with a minimum of technical diversity at minimal cost.

The core of an SPL is a stable software architecture that clearly separates common features from the variations reflected in the products and prescribes how software components can be assembled to derive individual products. Each new application is built by configuring the SPL, i.e., by selecting the variants (e.g., functionalities, software resources) that meet specific application requirements. The MDE environments simplify system configuration by providing domain-specific languages to model robot variability and model-to-model transformations to resolve the variability in the software control system. The following sections illustrate two MDE approaches for robotic systems configuration. The former focuses on modeling variability in robot functionality, the latter on modeling variability in robot resources.

The cost of creating new robotics products is significantly related to the complexity of developing software control systems.

Software Product Lines for Robotics

The HyperFlex [47] toolchain is a MDE environment for the development and configuration of SPL for Robotics. It provides domain-specific languages and graphical editors for the definition of three types of models. The template architectural model represents the software architecture (as discussed in the "System Architecture Modeling and Analysis" section) of a family of similar software systems. As such, it explicitly represents all the possible components (variants) that implement a given functionality (variation point). For example, the obstacle avoidance functionality might be provided by two alternative components implementing the Vector Histogram approach or the Dynamic Window approach. The Feature Model [48] symbolically represents the variant features of a control system; symbols may indicate individual robot functionality (e.g., marker-based localization) or concepts that are relevant in the application domain, such as the type of items that the robot has to transport (e.g., liquid, fragile, and so on), which affect the configuration of the control system.

The Resolution Model defines model-to-model transformations, which allow users to automatically configure the architecture and functionality of a control system based on required features. In this context, architecture configuration means to resolve the functional variability of a system by selecting one variant for each variation point. A model-to-text transformation generates the configuration file (launch file) corresponding to the architecture of the configured control system. Currently, the HypeFlex supports the configuration of component-based systems based on the ROS framework. Let us consider an example of SPL related to the simple system shown in Figure 1. The LocalNavigator compound component encapsulates the TrajectoryFollower component (see Figure 2), which implements the algorithm for generating twist commands to the BaseController component to follow a given trajectory. Let us now consider a different application in which the robot has to follow a moving target (e.g., another robot). The TrajectoryFollower component is replaced with the LeaderFollower component that periodically generates twist commands for the robot according to the position of the target estimated by the sensors.

A robot control system is composed of tens of components.

The HyperFlex allows representation of the navigation strategy as a variation point that can be resolved by selecting at deployment time either the LeaderFollower component or the TrajectoryFollower component. Figure 5 shows the Feature Model of the navigation system. Green boxes represent the selected features. The triangles below the boxes indicate the cardinality of the variation point. In particular, it means that, for each variation point, only one variant can be selected. A black circle indicates that the child feature is mandatory.

The selection of desired features triggers model-to-model transformations that configure the template architecture. Several types of transformations can be defined, such as removing a component, changing the properties of a component, and changing the connections between components. For example, the selection of the feature LeaderFollower triggers a transformation that replaces the TrajectoryFollower component with the LeaderFollower component and configures the connectors. The HyperFlex toolchain offers the possibility to compose hierarchically Feature Models in such a way that higher-level Feature Models abstract the subsystems functional features. Here, the idea is to model features at different levels of abstractions for different types of users.

Typically, the expert in robotic functionalities is interested in a representation of the functional features that highlights the different algorithms implemented in the robot control system. On the contrary, the application domain expert is interested in a representation of the control system capabilities that highlight the application requirements. For example, the robot is able to navigate in an environment, which might be static or populated by moving obstacles (e.g., people). It might be an open space with wide passages or a small room crowded with tiny obstacles. The selection of a feature corresponding to an application requirement triggers the automatic selection of features corresponding to robot capabilities. For example, if the environment is a static and crowded space, the robot should be configured with a complete (and likely slow) motion planner; instead, in dynamic environments, a fast and approximate motion planner is more effective.

Listing 3. The Specification of a Differential Constraint.

```
BEGIN DifferentialDrive
  ACTION: ul, ur;
  PARAM: L, r;
  CONFIG: x, y, theta;
  d(x)  = r/2 * (ul + ur) * cos(theta);
  d(y)  = r/2 * (ul + ur) * sin(theta);
  d(theta) = r/L * (ur - ul);
END;
```

Modeling Robotic Resources

A model of the robot embodiment (the kinematic structure, the device characteristics, and so on) and a model of the operational environment (the objects to recognize, avoid, grasp, and so on) are crucial for object manipulation and navigation. They act as shared resources among multiple robot functionalities, such as perception, planning, and control. Most algorithms that implement these functionalities are parametric with respect to the embodiment and world models. This means that the robot control system can be configured at deployment time according to the specific application requirements and can be reused without modification for different tasks and different robotic systems.

Blumenthal and Bruyninckx [49] present a textual domain-specific modeling language for modeling the structural aspects of the robot embodiment and environment as a Scene Graph, i.e., a description of relevant objects and the relations among them. These relations are organized in a Direct Acyclic Graph. The Scene Graph can express prior semantic knowledge about a scene, such as the morphology and appearance of an object (e.g., a table) that the robot should recognize

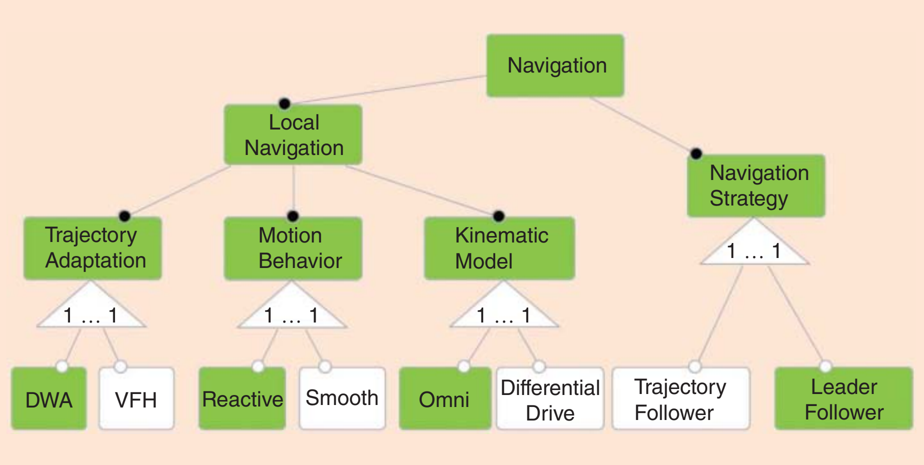


Figure 5. The HyperFlex Feature Model of the navigation system. DWA: dynamic window approach; VFH: vector field histogram.

and locate in the environment. Another example of textual modeling language is presented in [50]. It allows the description of differential constraints typically found in motion planning problems for mobile robots. Listing 3 shows an example of differential constraints for a differential drive rover, where *ul* and *ur* are the input angular velocities of the left and right wheels, respectively, (the action vector), *x*, *y*, and theta represent the robot configuration (the state vector), *r* is the wheel radius, and *L* is the distance between the two wheels.

These models are widely used in simulation algorithms (that, given the starting configuration and the action vector, compute the final configuration of the robot) and sampling-based motion planning algorithms (that sample collision free configurations that need to be compatible with the differential constraints). To compute final configurations, it is necessary to solve the differential equations, and this can be done by means of solvers, which use numerical approximation techniques.

Conclusions

The MDE is often considered to be synonymous with code generation and, frequently, research and development efforts in the MDE for robotics are motivated by the objective of simplifying application development by robotic experts with limited software engineering skills. In particular, the availability of open source tools (e.g., Eclipse), that greatly simplify the development of specialized MDE environments, stimulated the definition of DSLs for a large variety of robotics concerns. Two recent surveys can be found in [51] and [52].

Interestingly, a study on the state of practice in MDE in industry reports that productivity gains due to automatic code generation are not considered significant enough to drive an MDE adoption effort, due to increased training costs and substantial organizational changes [6]. It turns out that the main advantages are in the support that MDE provides in defining the architecture of a software system.

For this reason, this article focused on the architectural model as the central artifact of almost all software development activities (analysis, design, implementation, configuration, and documentation). Intentionally, this article did not illustrate the MDE solutions for two other key activities in software development for robotics, namely, simulation and model definition for runtime system configuration.

Simulation is such a wide field of research, that the analysis of MDE techniques for simulation could be the topic of a new article.

Runtime configuration is a new trend in the development of self-adaptive systems. In this context, the software control system is equipped with mechanisms for automatic interpretation of architectural variability models and for dynamic reconfiguration, based on context awareness, of its resources and functionalities. The MDE techniques (e.g., Dynamic SPLs [53]) have been developed for several software-intensive application domains, but their applicability to the development of self-adaptive robot control systems is still a research issue.

References

[1] H. Schichl, “Models and the history of modeling,” in *Modeling Languages in Mathematical Optimization*, J. Kallrath, Ed. Norwell, MA: Kluwer, 2004, pp. 25–36.

[2] F. Cellier, “Principles of active electrical circuit modeling,” in *Continuous System Modeling*. New York: Springer, 1991, pp. 201–249.

[3] J. F. Broenink, “20-sim software for hierarchical bond-graph/block-diagram models,” *Simul. Pract. Theory*, vol. 7, nos. 5–6, pp. 481–492, Dec. 15, 1999.

[4] R. Lewis. (2001). *Electrical Engineers, Modelling Distributed Control Systems Using IEC 61499: Applying Function Blocks to Distributed Systems* (IEE Control Engineering Series). *Inst. Eng. Technol.* [Online]. Available: http://books.google.it/books?id=m3LaTv7VefwC

[5] B. Selic. (2003, Sept.). The pragmatics of model-driven development. *IEEE Softw.* [Online]. *20(5)*, pp. 19–25. Available: http://dx.doi.org/10.1109/MS.2003.1231146

[6] J. Whittle, J. Hutchinson, and M. Rouncefield, “The state of practice in model-driven engineering,” *IEEE Softw.*, vol. 31, no. 3, pp. 79–85, May 2014.

[7] OMG. (2015). Unified modeling language. [Online]. Available: http://www.omg.org/spec/UML/

[8] D. Brugali and A. Shakhimardanov, “Component-based robotic engineering (Part II) [tutorial],” *IEEE Robot. Automat. Mag.*, vol. 17, no. 1, pp. 100–112, 2010.

[9] S. Cousins, B. Gerkey, K. Conley, and W. Garage, “Sharing software with ros [ros topics],” *IEEE Robot. Automat. Mag.*, vol. 17, no. 2, pp. 12–14, 2010.

[10] P. Clements, F. Bachmann, L. Bass, D. Garlan, J. Ivers, R. Little, R. Nord, and J. Stafford, *Documenting Software Architectures: Views and Beyond*. Boston, MA: Addison-Wesley, 2003.

[11] (2015). Object Management Group. [Online]. Available: http://www.omg.org/

[12] J. Ivers, P. Clements, D. Garlan, R. Nord, B. Schmerl, and O. Silva. (2004). Documenting component and connector views with UML 2.0. Software Engineering Inst., Carnegie Mellon Univ., Pittsburgh, PA. Tech. Rep. CMU/SEI-2004-TR-008. [Online]. Available: http://resources.sei.cmu.edu/library/asset-view.cfm?AssetID=7095

[13] D. Garlan. (2003). *Formal Modeling and Analysis of Software Architecture: Components, Connectors, and Events* (SFM, Lecture Notes in Computer Science, vol. 2804). M. Bernardo and P. Inverardi, Eds. New York: Springer, pp. 1–24. [Online]. Available: http://dblp.uni- trier.de/db/conf/sfm/sfm2003.htmlGarlan03

[14] H. Eichelberger, Y. Eldogan, and K. Schmid. (2009). *A Comprehensive Survey of UML Compliance in Current Modelling Tools* (Software Engineering, LNI, vol. 143). P. Liggesmeyer, G. Engels, J. Mnch, J. Drr, and N. Riegel, Eds. Berlin: Gesellschaft für Informatik. pp. 39–50. [Online]. Available: http://dblp.uni-trier.de/db/conf/se/se2009.htmlEichelbergerES09

[15] K. D. Evensen and K. A. Weiss, “A comparison and evaluation of real-time software systems modeling languages,” in *Proc. AIAA Infotech@Aerospace*, Altanta, GA, 2010, pp. 1–13.

[16] OMG. (2015). Modeling and analysis of real-time and embedded systems. [Online]. Available: http://www.omg.org/omgmarte/

[17] J. Medina and A. G. Cuesta, “Model-based analysis and design of real-time distributed systems with ADA and the UML profile for MARTE,” in *Reliable Software Technologies—Ada-Europe 2011* (Lecture Notes in Computer Science, vol. 6652), A. Romanovsky and T. Vardanega, Eds. Berlin, Germany: Springer, 2011, pp. 89–102.

[18] OMG. (2015). Tools related to MARTE. [Online]. Available: http://www.omgmarte.org/node/31

[19] S. Demathieu, F. Thomas, C. Andr, S. Grard, and F. Terrier, “First experiments using the UML profile for MARTE,” in *Proc. 11th IEEE Int. Symp. Object-Oriented Real-Time Distributed Computing*, Orlando, FL, May 5–7, 2008, pp. 50–57.

[20] P. H. Feiler and D. P. Gluch, *Model-Based Engineering with AADL: An Introduction to the SAE Architecture Analysis & Design Language*, 1st ed. Boston, MA: Addison-Wesley, 2012.

[21] M. Faugere, T. Bourbeau, R. de Simone, and S. Gerard, "MARTE: Also an UML profile for modeling AADL applications," in *Proc. 12th IEEE Int. Conf. Engineering Complex Computer Systems*, July 2007, pp. 359–364.

[22] (2015). Osate: Open source AADL tool environment. [Online]. Available: http://www.aadl.info.

[23] G. Biggs, K. Fujiwara, and K. Anada, "Modelling and analysis of a redundant mobile robot architecture using AADL," in *Simulation, Modeling, and Programming for Autonomous Robots* (Lecture Notes in Computer Science, vol. 8810), D. Brugali, J. Broenink, T. Kroeger, and B. MacDonald, Eds. New York: Springer International Publishing, 2014, pp. 146–157.

[24] L. Lamport, "Proving the correctness of multiprocess programs," *IEEE Trans. Softw. Eng.*, vol. 3, no. 2, pp. 125–143, Mar. 1977.

[25] A. Basu, M. Bozga, and J. Sifakis, (2006). Modeling heterogeneous real-time components in BIP. in *Proc. 4th IEEE Int. Conf. Software Engineering Formal Methods*. Washington, D.C., pp. 3–12. [Online]. Available: http://dx.doi.org/10.1109/SEFM.2006.27

[26] S. Bensalem, M. Bozga, T.-H. Nguyen, and J. Sifakis, "D-finder: A tool for compositional deadlock detection and verification," in *21st Int. Conf., CAV Computer Aided Verification* (Lecture Notes in Computer Science, vol. 5643). Grenoble, France, June 26–July 2, 2009, New York: Springer, 2009, pp. 614–619.

[27] S. Bensalem, L. de Silva, F. Ingrand, and R. Yan. (2011). A verifiable and correct-by-construction controller for robot functional levels. *J. Softw. Eng. Robot. 2(1)*. pp. 1–19. [Online]. Available: http://dx.doi.org/10.4018/ijismd.2014070103

[28] R. France, B. Rumpe, and M. Schindler. (2013). Why it is so hard to use models in software development: Observations. *Softw. Syst. Model. 12(4)*. pp. 665–668. [Online]. Available: http://dx.doi.org/10.1007/s10270-013-0383-z

[29] Y. Zheng and R. N. Taylor. (2013, Oct.). A classification and rationalization of model-based software development. *Softw. Syst. Model. 12(4)*. pp. 669–678. [Online]. Available: http://dx.doi.org/10.1007/s10270-013-0355-3

[30] OMG. (2015). Model driven architecture. [Online]. Available: http://www.omg.org/mda/

[31] S. Sendall and W. Kozaczynski. (2003, Sept.). Model transformation: The heart and soul of model-driven software development. *IEEE Softw. 20(5)*. pp. 42–45. [Online]. Available: http://dx.doi.org/10.1109/MS.2003.1231150

[32] A. Lotz, J. F. Inglés-Romero, D. Stampfer, M. Lutz, C. Vicente-Chicote, and C. Schlegel. (2014, July). Towards a stepwise variability management process for complex systems: A robotics perspective. *Int. J. Informat. Syst. Model. Design*. *5(3)*. pp. 55–74. [Online]. Available: http://dx.doi.org/10.4018/ijismd.2014070103

[33] M. Huhns and M. Singh, "Service-oriented computing: Key concepts and principles," *IEEE Internet Comput.*, vol. 9, no. 1, pp. 75–81, Jan. 2005.

[34] (2015). RTAI—The realtime application interface for linux. [Online]. Available: https://www.rtai.org/

[35] O. M. Group. (2015). Common object request broker architecture. [Online]. Available: http://www.corba.org/

[36] (2015). Plateforme pour la Robotique Organisant les Transferts Entre Utilisateurs et Scientifiques (proteus). [Online]. Available: http://www.anr-proteus.fr.

[37] M. R. Genesereth and N. J. Nilsson, *Logical Foundations of Artificial Intelligence*. San Francisco, CA: Morgan Kaufmann Publishers, 1987.

[38] S. Dhouib, S. Kchir, S. Stinckwich, T. Ziadi, and M. Ziane, "Robotml, a domain-specific language to design, simulate and deploy robotic applications," in *Simulation, Modeling, and Programming for Autonomous Robots*(Lecture Notes in Computer Science, vol. 7628), I. Noda, N. Ando, D. Brugali, and J. Kuffner, Eds. Berlin Heidelberg, Germany: Springer, 2012, pp. 149–160.

[39] J. Palczynski and S. Kowalewski, "Early behaviour modelling for control systems," in *Proc. 3rd UKSim European Symp. Computer Modeling Simulation*, Nov. 2009, pp. 148–153.

[40] (2015). Simulink, simulation and model-based design. [Online]. Available: http://it.mathworks.com/products/simulink/

[41] (2015). 20-sim. [Online]. Available: http://www.20sim.com

[42] Y. Brodskiy, R. Wilterdink, S. Stramigioli, and J. Broenink, "Fault avoidance in development of robot motion-control software by modeling the computation," in *Proc. 4th Int. Conf. Simulation, Modeling, Programming Autonomous Robots* (Lecture notes in computer science, vol. 8810), D. Brugali, J. Broenink, T. Kroeger, and B. MacDonald, Eds. New York: Springer International Publishing, Oct. 2014, pp. 158–169.

[43] R. Bischoff, U. Huggenberger, and E. Prassler, "Kuka youbot—A mobile manipulator for research and education," in *Proc. IEEE Int. Conf. Robotics Automation*, May 2011, pp. 1–4.

[44] (2015). Brics integrated development environment. [Online]. Available: http://www.best-of- robotics.org/bride/

[45] M. Anand, S. Fischmeister, Y. Hur, J. Kim, and I. Lee. (2010, July). Generating reliable code from hybrid-systems models. *IEEE Trans. Comput. 59*. p. 1281–1294. [Online]. Available: http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=5453343

[46] P. Clements and L. Northrop, *Software Product Lines: Practices and Patterns*. Boston, MA: Addison-Wesley and White Plains, NY: Longman, 2001.

[47] L. Gherardi and D. Brugali, "Modeling and Reusing Robotic Software Architectures: The HyperFlex toolchain," in *Proc. IEEE Int. Conf. Robotics and Automation*, Hong Kong, China, May 31–June 5, 2014, pp. 6414–6420.

[48] K. Kang, S. Cohen, J. Hess, W. Novak, and A. Peterson. (1990). Feature-oriented domain analysis (FODA) feasibility study. Software Engineering Institute. Carnegie Mellon Univ. Pittsburgh, PA, Tech. Rep. CMU/SEI-90-TR-021. [Online]. Available: http://resources.sei.cmu.edu/library/asset-view.cfm?AssetID=11231

[49] S. Blumenthal and H. Bruyninckx, "Towards a domain specific language for a scene graph based robotic world model," in *Proc. 4th Int. Workshop Domain-Specific Languages models Robotic systems*, Tokyo, Japan, Nov. 2013, arXiv:1408.0200.

[50] M. Guarnieri, E. Magri, D. Brugali, and L. Gherardi. (2012). A domain specific language for modeling differential constraints of mobile robots. in *Proc. Int. Conf. Autonomous Robot Systems Competitions*. Guimares, Portugal. [Online]. Available: http://hdl.handle.net/1822/18887

[51] A. Nordmann, N. Hochgeschwender, and S. Wrede, *A Survey on Domain-Specific Languages in Robotics* (Lecture Notes in Computer Science, vol. 8810). New York: Springer International Publishing, 2014, ch. 17, pp. 195–206.

[52] A. Ramaswamy, B. Monsuez, and A. Tapus. (2014). Model-driven software development approaches in robotics research. in *Proc. 6th Int. Workshop Modeling in Software Engineering*. New York. pp. 43–48. [Online]. Available: http://doi.acm.org/10.1145/2593770.2593781

[53] R. Capilla, J. Bosch, P. Trinidad, A. Ruiz-Corts, and M. Hinchey. (2014). An overview of dynamic software product line architectures and techniques: Observations from research and industry. *J. Syst. Softw.*, *91*, pp. 3–23. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0164121214000119

Davide Brugali, University of Bergamo, Italy. E-mail: brugali@unibg.it.

IEEE ROBOTICS & AUTOMATION MAGAZINE

CALL FOR NOMINATIONS

ASSOCIATE EDITORS

IEEE Robotics & Automation Magazine is soliciting nominations for two new associate editors, to begin in January 2016. The associate editors play an important role in maintaining the caliber of the magazine by ensuring the quality of published articles by implementing reviews of technical features according to IEEE guidelines, soliciting interesting and topical material articles for publication in the magazine, guiding the overall direction of the publication and providing feedback from the readership through e-mail conversations, teleconferences, and twice-yearly in-person meetings held in conjunction with ICRA and the IEEE International Conference on Intelligent Robots and Systems.

Associate editor terms normally consist of a one-year probation period followed by two years of additional service if performance is satisfactory. Applicants should have a strong technical background and excellent English language skills.

Nominations should include a resume (not to exceed three pages), previous experience with publications as a reviewer or in other capacities, and areas of technical expertise. Please submit nominations as a single pdf file to Rachel O. Warnick at r.o.warnick@ieee.org by 30 September 2015.

Digital Object Identifier 10.1109/MRA.2015.2452195

IEEE ICRA 2015—Celebrating the Diversity of Robots and Roboticists

The IEEE International Conference on Robotics and Automation (ICRA) was held at the Washington State Convention Center in Seattle, Washington, 26–30 May 2015. ICRA is the flagship conference of the IEEE Robotics and Automation Society (RAS) and is a premier international forum for robotics researchers to present their latest work.

Over 3,000 attendees (an ICRA record) participated in a wide variety of activities. Highlighting the conference were 940 technical papers presented over three days in ten parallel tracks, representing authors from over 40 countries. These technical talks, presented as short highlight talks plus interactive presentations, were selected from 2,275 submissions (an ICRA record), by the Senior Program Committee (Figure 1), resulting in a 41% acceptance rate. Workshops and tutorials were also extremely popular, with over 1,400 attendees (an ICRA record) participating in 42 workshops and tutorials.

Excellent plenary talks were given by Daniela Rus (Massachusetts Institute of Technology), Helen Greiner (CyPhy Works), and Dean Kamen (DEKA Research and Development Corporation) and were streamed live during the conference. Twelve keynote speakers also highlighted their latest research. All of the plenary and keynote talks are available on the ICRA 2015 website (www.icra2015.org) and on YouTube.

IEEE ICRA 2015 included a vibrant industrial exhibition, with over 65 exhibitors (Figure 2) from industry and academia, along with publishers and technical societies. The opening night reception in the exhibit hall recognized the exhibitors and sponsors who supported the conference.

Many innovative forums and special events were held at ICRA 2015. A new forum was created this year to highlight robotics activities and education in the developing world. This forum included presentations on robotics in Mexico, Egypt, South Africa, Peru, Ghana, Thailand, and India. Also new this year was a special RAS conference highlights track, which featured top papers from four other RAS conferences: 1) IEEE Workshop on Advanced Robotics and its Social Impacts 2014, 2) IEEE International Conference on Automation

Figure 1. The IEEE ICRA 2015 Senior Program Committee. (Photo courtesy of IEEE ICRA 2015.)

Figure 2. The IEEE ICRA 2015 exhibition featured over 65 exhibitors. (Photo courtesy of IEEE ICRA 2015.)

Digital Object Identifier 10.1109/MRA.2015.2452091
Date of publication: 11 September 2015

Figure 3. Peter Hart, Rodney Brooks, and Raja Chatila help celebrate the 50th anniversary of Shakey. (Photo courtesy of IEEE ICRA 2015.)

Figure 4. The Go, Girl, Go! event brought 600 middle and high school girls to the conference to network with established robotics researchers and professionals. (Photo courtesy of IEEE ICRA 2015.)

Science and Engineering (CASE) 2014, 3) IEEE Haptics Symposium 2014, and 4) Humanoids 2014. For the first time, ICRA also held a late-breaking results poster session, which highlighted the research of over 100 participants. A special celebration of Shakey's 50th anniversary was held (Figure 3), in conjunction with the Association for the Advancement of Artificial Intelligence (AI) to promote interactions between the AI and robotics communities.

Continuing the tradition of recent ICRAs, this year's conference also held industry and government forums and RAS Town Hall. The industry forum focused on strengthening ties between academia and research, building networks, and fostering entrepreneurship. The government funding panel discussed robotics funding and policy issues in the Americas, Europe, and Asia. The town hall was led by RAS President Raja Chatila and Antonio Bicchi, the editor in chief of *IEEE Robotics and Automation Letters* (*RA-L*), who discussed this new journal and its connection with ICRA. Papers published in *RA-L* will have the option of being presented at future ICRAs.

Special attention was paid to students at this year's ICRA, with innovative programming targeted to them. More than 150 students received travel scholarships sponsored by RAS, the National Science Foundation (NSF), and ICRA. A newly introduced Ph.D. forum was held to provide mentoring for graduate students. Becoming a Robot Guru, a special program for college students, was also held during ICRA. For the first time, a career fair was held to help connect recent graduates to companies that are hiring robotics specialists. For younger female students, an innovative Go, Girl, Go! event was organized by the conference, together with the Washington State For Inspiration and Recognition of Science and Technology (FIRST) Robotics organization, which brought together over 600 middle and high school girls to learn about career paths in robotics (Figure 4). RAS also sponsored several lunches for women in robotics, graduate students, and recent graduates. Students were also strong participants in the five robot challenges held this year. The Amazon Picking Challenge (Figure 5) was a popular robot challenge.

An awards luncheon was held to recognize the highest quality technical papers, RAS awards, Best Editor and Reviewers of the Conference Editorial Board (CEB), and the Robot Challenges Awards. A full listing of the award winners is given at icra2015.org/conference/awards. At the luncheon, special recognition was given to the Honorary General Chair Ruzena Bajcsy, who celebrated her birthday that day (Figure 6).

Time to socialize with colleagues and potential collaborators is also vital to a good conference, and ICRA 2015 provided plenty of opportunities. Along with the opening reception, a reception held at the Experience Music Project Museum for Music, Sci-Fi, and Pop Culture provided a unique experience for attendees to create their own music and experience contemporary pop culture. The closing reception at the Boeing Museum of Flight (Figure 7) gave attendees excellent views of one of the largest collections of air and space technology in the United States.

As shown in the ICRA 2015 graphic (Figure 8), the unofficial theme of the conference was celebrating the diversity of robotics and roboticists. The conference embraced this theme by including

Figure 5. A robot challenge at the IEEE ICRA 2015. (Photo courtesy of IEEE ICRA 2015.)

Figure 6. Honorary General Chair Ruzena Bajcsy celebrated her birthday during awards luncheon at the IEEE ICRA 2015. (Photo courtesy of IEEE ICRA 2015.)

Figure 7. The IEEE ICRA 2015 closing reception at the Boeing Museum of Flight. (Photo courtesy of IEEE ICRA 2015.)

Figure 8. The IEEE ICRA 2015 graphic that was displayed on the Seattle Monorail that transported conference attendees to and from the EMP for the conference banquet reception. (Photo courtesy of IEEE ICRA 2015.)

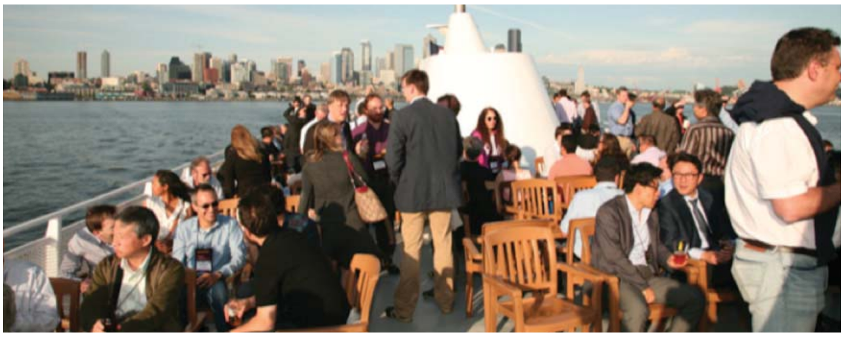


Figure 9. On the way to Tillicum Village for the VIP dinner, to acknowledge the service of the ICRA Organizing Committee and the CEB. (Photo courtesy of IEEE ICRA 2015.)

Figure 10. Allison Okamura, Lynne Parker, and Nancy Amato thanking the ICRA Organizing Committee and the CEB at the VIP dinner at Tillicum Village. (Photo courtesy of IEEE ICRA 2015.)

programming to encourage and support groups underrepresented in science, technology, engineering, and mathematics fields to consider careers in robotics. These included featuring an all female organizing committee (see poster available on ICRA 2015 website, www.icra2015.org), and hosting the CRA-W/CDC and NSF-sponsored Robot Guru workshop and the Go, Girl, Go! event that was coorganized by Washington FIRST Robotics.

Putting together ICRA 2015 required a tremendous amount of volunteer effort. The VIP dinner at Tillicum Village (Figure 9) was a special way to thank the Organizing Committee and the CEB for their hard work putting together the conference. Special thanks goes to Allison Okamura (Figure 10), who served as the editor in chief of the ICRA CEB and the entire CEB, who handled over 6,200 reviews of the submitted papers. We are also very grateful to all of the hard work of the outstanding Organizing Committee and the student volunteers, the strong support of the RAS, the generous sponsors, the speakers, and the attendees, for making ICRA 2015 a great success! To find out more about ICRA 2015, please visit www.icra2015.org for complete details and photos of the events.

—Lynne Parker, General Chair,
University of Tennessee,
Knoxville, Tennessee

—Nancy M. Amato, Program Chair,
Texas A&M University,
College Station, Texas

ICRA Awards Recipients Presented 28 May 2015

The following awards were presented during the IEEE International Conference on Robotics and Automation (ICRA) Awards Luncheon on 28 May 2015.

RAS Publication Award

- King-Sun Fu Memorial Award: Best Paper in *IEEE Transactions on Robotics* "Catching Objects in Flight," Seungsu Kim, Ashwini Shukla and Aude Billard, *IEEE Transactions on Robotics*, vol. 30, no. 5, pp. 1049–1065, 2014.

ICRA Awards

Best Associate Editor Award

- Eric Diller, University of Toronto, Canada
- Paolo Robuffo Giordano, CNRS, Irisa/Inria Rennes, France.

Best Reviewer Award

- Renaud Detry, Université de Liège, Belgium
- Andrej Gams, Jožef Stefan Institute, Slovenia
- Hedvig Kjellstrom, KTH Royal Institute of Technology, Sweden
- Raul Suarez, Universitat Politècnica de Catalunya, Spain.

ICRA Automation Best Paper

- "Design, Modeling and Control of a Modular Contactless Wafer Handling System," Bassem Dahroug, Guillaume J. Laurent, Valérian Guelpa, Nadine Le Fort-Piat.

Digital Object Identifier 10.1109/MRA.2015.2472076
Date of publication: 11 September 2015

ICRA Cognitive Robotics Best Paper Award

- "Grounding Spatial Relations for Outdoor Robot Navigation," Abdeslam Boularias, Felix Duvallet, Jean Oh, Anthony Stentz.

ICRA Medical Robotics Best Paper Award

- "An Ankle-Foot Prosthesis Emulator with Control of Plantarflexion and Inversion-Eversion Torque," Steven H. Collins, Myunghee Kim, Tianjian Chen, Tianyao Chen.

ICRA Robot Manipulation Best Paper Award

- "Learning Contact-Rich Manipulation Skills with Guided Policy Search," Sergey Levine, Nolan Wagener, Pieter Abbeel.

ICRA Robot Vision Best Paper Award

- "Work Smart, Not Hard: Recalling Relevant Experiences for Vast-Scale but Time-Constrained Localisation," Chris Linegar, Winston Churchill, Paul Newman.

ICRA Service Robotics Best Paper Award

- "RoboSherlock: Unstructured Information Processing for Robot Perception," Michael Beetz, Ferenc Balint-Benczedi, Nico Blodow, Daniel Nyga, Thiemo Wiedemeyer, Zoltan-Csaba Marton.

ICRA Best Student Paper (Figure 1)

- "Grasping Without Squeezing: Shear Adhesion Gripper with Fibrillar Thin Film," Elliot Wright Hawkes, David Christensen, Amy Kyungwon Han, Hao Jiang, Mark Cutkosky.

ICRA Best Conference Paper (Figure 2)

- "Observability, Identifiability and Sensitivity of Vision-Aided Inertial Navigation," Joshua Hernandez, Konstantine Tsotsos, Stefano Soatto.

Robot Challenges

Amazon Picking Challenge

In recognition of teams that demonstrate the best automated item picking.

- Organizers: Pete Wurman and Joe Romano, Kiva Systems
- First Place: US$20,000, Team RBO, TU Berlin, Oliver Brock
- Second Place: US$5,000, Team Massachusetts Institute of Technology, Alberto Rodriquez Garcia
- Third Place: US$1,000, Team Grizzly, Dataspeed Inc. and Oakland University, Paul Fleck.

Humanitarian Robotics and Automation Technology Challenge

- Organizers: Raj Madhavan, Lino Marques, Edson Prestes, Prithviraj Dasgupta
- Winner: Team National University of Singapore, Singapore.

Humanoids Application Challenge (Figure 3)

In recognition of the team that demonstrates the most creative application of humanoid research.

- Organizers: Kayla Kim and Jinwook Kim, ROBOTIS CO

Figure 1. The ICRA Best Student Paper Award being presented by Lynne Parker and Nancy Amato.

Figure 2. The ICRA Best Conference Paper Award being presented by Lynne Parker and Nancy Amato.

Figure 3. Humanoids Application Challenge.

- Winner: Team Snobots, University of Manitoba, Winnipeg, Canada.

Mobile Microrobotics Challenges

Autonomous Mobility and Accuracy Challenge

In recognition of the team with the most accurate trajectory traversals.

- Winner: University of Hawaii, Manoa (Sammy Khamis, Edward Nerad, Leanne King, Caralyn King, Aaron Ohta).

Microassembly Challenge

In recognition of assembling most microscale components in a microchannel.

- Winner: ETH Zurich (Samuel Charreyron, Janis Edelmann, Andrew Petruska, Franziska Ullrich, Chengzhi Hu, Hen-Wei Huang, Qi Zhang, Erdem Siringil, Roel Pieters, Bradley J. Nelson).

MMC Showcase and Poster Session

In recognition of the most innovative mobile microrobotic system.

- Winner: UVT Romania (Florin Dragomir, Ioan Alexandru Ivan, Mihaela Ivan, Valentin Gurgu).

Robot Revolution

On 21 May 2015, the Museum of Science and Industry (MSI) in Chicago launched Robot Revolution (Figure 1), a groundbreaking exhibition that introduces the public to a world where robots will be our companions and colleagues, fundamentally changing how we live, work, and play. The exhibition shows a future where robots are not a curiosity, but a vital and visible part of our daily lives.

This exhibition, developed and produced by MSI, is not the first museum exhibit on robots but it is unprecedented in its scope. The exhibit contains more than 40 robots, most of which are operational and interactive for guests (Figure 2).

Robot Revolution provides a broad look at the field of robotics. It features industrial and research robots as well as a handful of commercially available robots, telling the story of robotics in a way that both educates and entertains guests (Figure 3). Through hands-on activities, Robot Revolution helps to answer questions such as: How do robots work? How will they change our lives? How can I get involved in the field of robotics?

Given the increasing relevance of robotics to our daily lives, the MSI took on the challenge of creating Robot Revolution over a five-year period. The exhibit will expand the public's understanding of robotics and inspire the next generation of high-tech workers and innovators. A robotics-savvy

Figure 1. The ribbon cutting for Robot Revolution included two FANUC robots and Baxter from Rethink Robotics (photo courtesy of J.B. Spector/Museum of Science and Industry, Chicago, Illinois).

Figure 2. The guests cheer on autonomous robots from Zhejiang University in a game based on RoboCup soccer (photo courtesy of J.B. Spector/Museum of Science and Industry, Chicago, Illinois).

Digital Object Identifier 10.1109/MRA.2015.2452111
Date of publication: 11 September 2015

Figure 3. The guests watch as the robots are maintained and repaired (photo courtesy of J.B. Spector/Museum of Science and Industry, Chicago, Illinois).

Figure 4. The guests can test out six different grippers, including one from Festo based on a fish's tail fin (photo courtesy of J.B. Spector/Museum of Science and Industry, Chicago, Illinois).

workforce will be necessary to stay competitive in the global economy.

The exhibit and its associated programming will give youth unique learning experiences and a pathway to get involved in this growing field (Figure 4). To expand on the exhibit's hands-on content, the MSI has developed educational tools and online resources that address national science education standards. These tools will be freely available to support continued robotics exploration in schools, communities, and homes.

The major content areas in Robot Revolution are cooperation, smarts, skills, and locomotion. These areas highlight the capabilities of the robots and encourage guests to imagine new uses for these intelligent machines and the roles they could play in our lives.

In addition to the main content areas, Robot Revolution includes the Drone Zone, a live stage show with guest participation, and the ROBOGARAGE, where guests can watch as robotic specialists maintain and repair the robots.

Robot Revolution began with research trips to universities and companies in the United States and abroad. The Chicago office of the Japanese External Trade Organization (JETRO) organized two trips to Japan for the exhibit team. The museum staff attended robotics trade shows and conferences such as the IEEE/RSJ International Conference on Intelligent Robots and Systems and the IEEE International Conference on Robotics and Automation to learn more about the field and meet those involved. It was the generous encouragement and support from organizations such as the IEEE Robotics & Automation Society (RAS), JETRO, and the robotics community, that helped the MSI decide to take on this demanding project.

Discussions about the organization and structure of the exhibit were supported by IEEE RAS, which provided a small grant to convene a panel of experts that brainstormed ideas for displays, potential demonstrations, and outreach to companies and universities. The IEEE RAS expert panel was composed of Henrik Christensen, Georgia Tech, Kevin Lynch, Northwestern University, Dennis Hong, UCLA, Daniel Lee, University of Pennsylvania, Cynthia Breazeal, Massachusetts Institute of Technology, Raffaello D'Andrea, ETH Zurich, and Masatoshi Ishikawa, University of Tokyo. Their insights and support were invaluable to the project!

A critical challenge in producing the exhibit was adapting a number of the robots to do things they were not originally designed to do. Robot Revolution would not have been possible without the willingness of the robot providers to work with the museum to modify their robots to work in this new context.

With the exception of the industrial robots, many of the robots were not intended to run eight hours a day, day after day. Robots that were meant to be driven by trained operators are being operated by young children in the exhibit. Mark Ewing and Faith Griggs-York, the project managers for Robot Revolution, worked with the universities and companies to help them understand the circumstances in which the robots would be operating and figure out the needed adjustments to the robot's programming and hardware.

Ewing and Griggs-York traveled twice to China to work with the championship RoboCup team at Zhejiang University. "Their hardware is among the best in the world, but it was designed to win matches," Ewing said. "Our needs are very different." Adapting the hardware to increase the durability of the soccer bots makes them heavier. While this may slow the robots down a little, they can still deliver a crowd-thrilling performance.

Keeping the diverse group of robots running is a big task. Maintenance and repairs are done in front of guests in the exhibit's ROBOGARAGE by Adrian Choy, lead robotic specialist, and his staff. About 200 batteries need to be changed every day. They also provide insight about careers in robotics. While working in the exhibit, Anna Brill, one of the specialists, answers questions such as "What is your background?" and "My kids are interested in engineering, what should they study in school?"

During the run of Robot Revolution, the MSI will provide opportunities for outreach during special Robot Block Party weekends. These events will be based on the museum's popular National Robotics Week celebrations. Guests will have an opportunity to talk directly with robotics researchers and see demonstrations of their newest work, gain insight into what is on the horizon in robotics and explore advancements in STEM education. The museum's exhibit content and programming is always greatly enhanced by partnerships with researchers at academic institutions and research and development labs.

After Robot Revolution finishes in Chicago, it will travel to eight other major science centers in the United States, with the possibility of traveling overseas. The exhibit team will work with the robotics partners to keep the content current over the years of the exhibit, updating the robots and videos to keep pace with advances in the field. The MSI will also work with each venue and reach out to the robotics community and provide opportunities for researchers to engage with a broad audience and showcase current work.

Robot Revolution is sponsored by Google.org with additional support from The Boeing Company, RACO Industrial, The David Bohnett Foundation, The Kaplan Foundation, and United Airlines. To learn more about Robot Revolution and associated events, visit www.msichicago.org.

Social Implementation of Disaster Robots and Systems

The United Nations held the World Conference on Disaster Risk Reduction on 14–18 March 2015 in Sendai, Japan [1]. This conference announces a framework of action for disaster reduction of worldwide activities and national actions every ten years. The Hyogo Framework of Action, announced in 2005, has had a major influence on the policies of each participating nation.

The IEEE Robotics and Automation Society organized a public forum, "Social Implementation of Disaster Robots and Systems" on 14 and 16 March in cooperation with Tohoku University International Research Institute of Disaster Science, the International Rescue System Institute, the COCN Disaster Robot Project, the Japan Cabinet Office ImPACT Project, and JSME Robotics–Mechatronics Division Technical Committee on Disaster Robotics [2] (Figures 1 and 2). A voluntary commitment of robotics for risk reduction discussed at the forum was announced at a working session of the main body of the conference on 15 March as a part of the Sendai Framework of Action [3].

The forum theme on 14 March was social implementation of disaster robots and systems. The forum addressed the current state of disaster robots and systems and action plans for the future. Six top leaders in this field including Hajime

Figure 1. The invited speakers for the public forum on Social Implementation of Disaster Robots and Systems: Current State, Gap, and Action Plans on 14 March.

Digital Object Identifier 10.1109/MRA.2015.2452112
Date of publication: 11 September 2015

Figure 2. The invited speakers for the public forum on Social Implementation of Disaster Robots and Systems: Application Record and Challenges for the Future on 16 March.

Asama, Robin Murphy, Gerald Steinbauer, Geert De Cubber, Raj Madhavan, and Satoshi Tadokoro, gave talks about applications and achievements of disaster robotics, the current state in each country and region, gaps for deployment, SIG humanitarian technology, and future actions. The following voluntary commitment was approved for the future disaster robotics applications.

This conference announces a framework of action for disaster reduction of worldwide activities and national actions every ten years.

"The establishment of an international committee of all relevant stakeholders as a way to accelerate the implementation of robotics and information and communications technology for national disaster management plans and national regulations synchronizing with rapid technology revolution."

This was announced at a working session on Earth observations and high technology to reduce risks in the main forums on 15 March and was positioned as an action item in an adjunct document of Sendai Framework of Action, which is the main outcome of the 2015 United Nations World Conference on Disaster Risk Reduction.

A public forum on 16 March focused on the application record and challenges for the future regarding the social implementation of disaster robots and systems. Six top companies and a university in Japan introduced their activities for emergency response for Fukushima–Daiichi accident, such as robot application and research and development for Fukushima decommissioning, the establishment of Nuclear Emergency Assistance Center, a 20-year history and the future of unmanned construction systems, and the application of UAVs for landslide disaster recovery.

Hajime Asama, Robin Murphy, Gerald Steinbauer, Geert De Cubber, Raj Madhavan, and Satoshi Tadokoro at the conference of the United Nation on disaster risk reduction have recognized that robotics will be one of the most important technologies for disaster response, recovery, and preparedness in the future. Awareness of necessary actions for its social implementation and the establishment of the international committee was a big step for risk reduction support for robotics in the future.

The organizers would like to thank all the contributors and supporters for this great success.

References

[1] [Online]. Available: http://www.wcdrr.org/

[2] [Online]. Available: http://www.ieee-ras.org/educational-resources-outreach/un-symposium

[3] URL of Voluntary Commitment.

UN World Conference on Disaster Risk Reduction

By Satoshi Tadokoro

Robotics is becoming a powerful tool for disaster risk reduction, preparedness, response, and recovery. To fully support the Post-2015 Framework of Action, an international committee of robotics will be established in cooperation with all the relevant stakeholders, including public bodies, local communities, disaster response teams, researchers, and industries as well as various research and development projects worldwide and the IEEE Robotics and Automation Society.

The committee will define the top goals and metrics for the contribution of robotics to risk reduction. Subgoals and actions will be identified with the consideration of disaster risk reduction plans, operational regulations, laws, interoperability, missing capabilities, technical readiness levels, and commercialization. The committee

Digital Object Identifier 10.1109/MRA.2015.2452113
Date of publication: 11 September 2015

will be open to discuss further issues with the appropriate bodies and organizations to derive actions that need to be taken. The committee will also serve as an advisory body to local, national, and international groups and governments.

The recent evolution of a wide variety of robotics and component technologies is rapidly enhancing their capabilities. For example, unmanned aerial vehicles quickly surveyed wide disaster areas, remotely operated underwater vehicles to repair leakage of subsea oil plants, while unmanned ground vehicles worked in contaminated areas of damaged nuclear power plants. Twenty years ago, unmanned aerial vehicles could only gather information from sky. At present, unmanned aerial vehicles can approach structures of interest in the neighborhoods to provide detailed visual inspections for maintenance. They can also enter damaged buildings through narrow entrances to search for victims. The autonomy and intelligence of unmanned ground vehicles can reduce responders' loads and integrate gathered information with measured three-dimensional information into GIS. For this reason, specialists predict that robotics will become an essential tool for preparedness, response, and recovery in ten years. The implementation of robotics with information and communications technology is needed to support the Post- 2015 Framework of Action by enriching the global risk awareness with local information in detail.

The contribution of robotics is mainly: 1) to gather information and perform tasks that humans and conventional equipment cannot (e.g., search and rescue in inaccessible places and the inspection in highly contaminated areas), 2) to reduce risks (e.g., substitute workers to avoid potential damage from explosions, toxic agents, and radiation), and 3) to reduce time and cost (e.g., quick surveillance of potentially damaged facilities and high places without scaffolds).

Records of robot applications to disasters from the last decade show gaps that have be filled to fully utilize of robotic solutions. The following issues must also be discussed and solved.

Technologies need more improvement and development for required tasks at disaster sites with higher technology readiness levels. Particularly, performance of mobility, stationary, sensing, recognition, remote situation awareness, wired and wireless communication, human interface, intelligent autonomy, task execution performance, and compliance under/with disaster conditions and environments including explosions have to be sufficient in the systems using robots, humans, and organizations. The international collaboration of academic societies, research centers, universities, test facilities, robot solution contests, and robot training curriculums have to be promoted to make them ready.

Second, social barriers to deployment and application of robots have to be lowered. Regulations and systems have to be adjusted for this new innovation, particularly for disaster countermeasures, road traffic, maintenance of infrastructure and industrial facilities, performance test methods for procurement, and insurance for predicted risks with Good Samaritan laws. Safety standards, wireless frequency allocation, and component interfaces have to be common and standardized internationally for the exchangeability and reusability of systems to foster smooth international cooperation in megascale disasters.

The united efforts of all the relevant stakeholders in the newly established international committee will resolve the technical and social issues to fill gaps for the full use of this new technology in the future.

Innovations in Robotics Panel at the 2015 WIE International Leadership Conference

By Laura Margheri

The IEEE Women in Engineering (WIE) International Leadership Conference (WIE-ILC) is the flagship and largest event organized by the IEEE WIE. On 23–25 April 2015, the second WIE-ILC took place in Silicon Valley, San Jose, California, with the theme "Lead Beyond. Accelerating Innovative Women Who Change The World." The focus of the conference was on leadership, innovation, and entrepreneurship, with four tracks:

1) innovation (skills to create a new technology, lead innovative teams, foster creative cultures, or develop disruptive technology)
2) empowerment (skills to help women advance in their careers)
3) entrepreneurship (skills around startups, business models, venture funding, finance, or leadership communication)
4) executive leadership (skills for team leadership, career management, and advancement).

More than 700 attendees joined the event, with an exciting program, including nine

Digital Object Identifier 10.1109/MRA.2015.2452131
Date of publication: 11 September 2015

keynote speakers, 36 parallel sessions and 13 virtual sessions in three days.

The IEEE Robotics and Automation Society (RAS) was one of the sponsors of the event, and representatives of the RAS WIE Committee were among the speakers. A panel session on leading innovation in robotics and automation was proposed and accepted for the innovation track. The conference board received an incredible number of proposals this year, and they were able to accommodate only 18% of the submissions.

Jing Xiao (professor, University of North Carolina), Lynne Parker (director, National Science Foundation), and Allison Okamura (professor, Stanford University) were the three speakers in the panel and talked about their experience, technical and scientific skills, and other aspects related to the personality traits that can make the difference in innovation and leadership in robotics a roundtable moderated by Laura Margheri (RAS WIE chair, The BioRobotics Institute, Scuola Superiore Sant'Anna). RAS activities and opportunities were also presented (Figure 1). The room was full of researchers and managers, and there was a lot of interaction with the speakers after the end of the session.

A panel session on leading innovation in robotics and automation was proposed and accepted for the innovation track.

The panel was an inspiring and motivating showcase of leadership of women in robotics and automation. Furthermore, the panel at the 2015 WIE-ILC conference was a great opportunity to increase the visibility of RAS within WIE and to involve both experienced women as well as young girls attending the event.

Figure 1. From left: Allison Okamura (professor, stanford University), Jing Xiao (professor, University of North Carolina), Lynne Parker (Director, National Science Foundation), and Laura Margheri (RAS WIE chair, The BioRobotics Institute, Scuola Superiore Sant'Anna).

Figure 2. Jessica Hodgins (Disney Research and Carnegie Mellon University).

Figure 3. IEEE WIE chair. Takako Hashimoto.

IEEE ICRA 2015—Women in Engineering Lunch

The RAS WIE Lunch is the event that the WIE Committee organizes periodically at the International Conference on Robotics and Automation (ICRA), IROS, and CASE. The WIE Lunch is an important opportunity to foster discussion and collaboration, to inspire girls, and to advance female leadership in engineering and robotics. Each WIE Lunch includes a talk by an invited speaker.

During the ICRA in Seattle Washington, the WIE Lunch was sponsored by Disney Research, and Jessica Hodgins (Disney Research and Carnegie Mellon University) and Takako Hashimoto (the IEEE WIE chair) were guest speakers (Figures 2 and 3). The event had a record of 100 attendees.The WIE Committee would like to thank the great speakers and sponsers, as well as everyone who joined and actively participated in the roundtable discussion.

IEEE ROBOTICS & AUTOMATION MAGAZINE

CALL FOR PAPERS

Special Issue on
Soft Robotics and Morphological Computation

Introduction
Soft robotics and morphological computation are two recent exciting trends in robotics that are expected to provide novel approaches and high-impact applications. The use of soft and deformable materials in robotics system is crucial to deal with uncertain task and environments such as locomotion in rough terrains, grasping and manipulation of unknown and unstructured objects. Soft robots can be characterized by elastic and deformable bodies, a large number of degrees of freedom, the use of unconventional functional materials, and the involvement of intrinsic passive dynamics. All of these properties could provide significant advantages to adaptability of robotic systems if they are treated properly. The research field of morphological computation, on the other hand, explores the concepts and theories of computation in physical systems, where we investigate how motion control processes can be distributed over informational and physical dynamics. It has been previously shown that, by properly designing the dynamics, physical systems such as soft robotic grippers can benefit from simplified control architectures and improved overall performances.

The special issue of “Soft Robotics and Morphological Computation” in *IEEE Robotics and Automation Magazine* (IEEE-RAM) aims to summarize the state of the art of soft robotics and morphological computation research areas, and to provide a venue for the fruitful collaborations between these two research fields. The desired outcome of this special issue is to develop a general consensus about the scientific goals, perspective and challenges of the two research fields, as well as high impact applications.

Topics
We invite review/position papers of topics related to soft robotics, morphological computation and the intersection between the two fields. The topics include but not limited to:

- Artificial skin and stretchable sensors and electronics
- Bio-inspired or biomimetic robots based on passive dynamics and unconventional materials
- Continuum robots, flexible robots, reconfigurable robots
- Functional materials, morphologies, and assembly for adaptive robotic systems
- Modeling and simulation of soft bodied robots and structures
- Natural computation, unconventional computation for adaptive robotic systems
- Physical human-robot interactions based on soft technologies
- Wearable robotics

Timeline:

Deadline for Paper Submission: 10 November 2015
Review Completion and Acceptance Notification: February 2016
Final Submission: June 2016
Publication: September 2016

Guest Editors:

Fumiya Iida (University of Cambridge), Cecilia Laschi (Scuola Superiore Sant'Anna), Dario Floreano (EPFL), Robert Wood (Harvard University), Surya Nurzaman (Monash University), Andre Rosendo (University of Cambridge)

Complete details at http://www.ieee-ras.org/publications/ram/ram-special-issues

Digital Object Identifier 10.1109/MRA.2015.2452197

IN MEMORIAM

The Achievements of Antal

By George Bekey and Paolo Fiorini

Antal K. (Tony) Bejczy, a giant in the field of space robotics and a founder of the IEEE Robotics and Automation Society (RAS), died on 25 June 2015, following a long illness. Tony made numerous and fundamental contributions to human–machine interfaces, teleoperation, end effectors for space operations, and other areas. He was born in Ercsi, a town in central Hungary, approximately 35 km south of Budapest. He studied at the Budapest University of Technology, left Hungary in 1958 for Norway, where he continued his studies at the University of Oslo, and completed his education at the California Institute of Technology in Pasadena. He lived with his wife Margot in Pasadena until his final illness. He was a senior research scientist at the NASA Jet Propulsion Laboratory (JPL) at his retirement.

Tony's most important legacy is his memory and the inspiration that it brings to all his former colleagues and friends.

He was intimately involved with the birth of the IEEE RAS. The RAS was preceded by the Robotics and Automation Council (RAC), founded in 1983 by George Saridis. Councils have no individual members. Rather, they are formed by other IEEE Societies and nurtured until they are ready to become Societies. In the case of the RAC, it was founded by eight Societies with an interest in robotics. Saridis was the first president. In 1987, Tony was president of the RAC, and, under his leadership, the RAC began its transformation into the RAS. The transformation was completed in 1989.

Digital Object Identifier 10.1109/MRA.2015.2453692
Date of publication: 11 September 2015

Antal K. (Tony) Bejczy, 1930–2015.

During his 32-year tenure at JPL, Tony pioneered the development of robot dynamics, on which he wrote the first seminal paper, and of multifingered end effectors equipped with sensors, which he termed *smart hands*. He was the first to use force and rate feedback in the control of space teleoperators, a technology that is now being used not only in space robotics but also in surgical robotics and in many applications of haptics. He was a principal investigator of a flight experiment using a force-moment sensor enhanced hand on the remote arm of the space shuttle *Columbia* in 1994. Tony received 43 NASA innovation awards.

Tony's most important legacy is his memory and the inspiration that it brings to all his former colleagues and friends. He discovered and supported young talents from the early days of robotics until his last years. While attending conferences, he was always available to meet with students, and he encouraged their research, often by starting new collaborations with their universities and research centers. Through these contacts, he was able to recruit a number of researchers to his laboratory at JPL, going to great lengths to overcome the obstacles of hiring non-U.S. students. Unknown students and researchers were welcome to discuss research projects and employment possibilities with him. Once hired, they had his full support and guidance even in the most critical research tasks. Tony was a truly passionate teacher in robotics research, who could show his colleagues how one could inspire many with exceptional understanding and with his warm heart and friendliness.

This passion remained strong after his retirement, and whenever he met his former JPL colleagues, the first topic was robotics research at JPL, to which he often contributed new ideas of his own. Tony regularly visited with former colleagues who moved to different institutions, giving advice and guidance on research directions, laboratory development, and personal growth. Until his final days, he kept a tight connection with Hungary, where he traveled every year, giving dozens of lectures and inspiring hundreds of young engineers, always caring for new concepts, and actively linking people with ideas—surgeons with engineers and space scientists with physicists. In recognition of his efforts, the Antal Bejczy Center for Intelligent Robotics was inaugurated in January 2015 at the Óbuda University, Budapest, with the IEEE Space Robotics Workshop in honor of his 85th birthday.

Dr. Bejczy served as the general chair of the RAS International Conference on Robotics and Automation in 1986. He was an IEEE Life Fellow, and he received the RAS Pioneer Award in 2004, the Distinguished Service Award in 2007, and the IEEE Robotics and Automation Technical Field Award in 2009. He was a scrupulously ethical person, with great charm and eloquence in spite of his strong Hungarian accent. Clearly, Tony Bejczy was a major contributor to robotics research and a great friend to many people involved in robotics. He will be sorely missed.

Tony made numerous and fundamental contributions to human–machine interfaces.

TC Spotlight *(continued from p. 31)*

SIG. Members who are professors are responsible for mentoring students and promoting participation in SIG activities.

As for technical activities, the IEEE TC on Bioroboitcs China SIG plans to hold one or two workshops (or tutorials) annually in association with a conference closely related to biorobotics. The workshops would allow scientists and engineers from different backgrounds to work together on new research avenues, as discussions will reveal the need for research on robotic technology that can safely interact with living systems and function in complex natural or man-made environments.

Other Key Activities Organized/Supported in the Last Five Years

- IROS 2009: Workshop on Biologically Inspired Robots, St. Louis, Missouri, October 2009
- IROS 2009: Special Session on Critical Impediments and Future Challenges for Multimodal Robotic Locomotion, St. Louis, Missouri, October 2009
- IROS 2011: “50 Years of Robotics,” San Francisco, California, September 2011; Symposium on Bioinspired Robotics (Sessions I and II): Plenary session featuring plenary speakers Gabriel Nelson (Boston Dynamics), Shigeo Hirose (University of Tokyo), Robert Full (University of Berkley), and Robert Wood (Harvard University)
- ICRA 2014—Workshop on Latest Advances on Natural Motion Understanding and Human Motion Synthesis, Hong Kong, China, May 2014
- MOU—Memorandum of Understanding for Research Collaboration between the IEEE RAS TC on Biorobotics and the Biomimetic Robot Research Center, Seoul National University, Korea, June 2013.

For future activities, the TC hopes to collaborate with other TCs (for example, the IEEE TC on Soft Robotics) for cross-TC efforts on topics of mutual interest. The TC also hopes to establish new SIGs in other regions, similar to the newly formed TC SIG in China.

Members of the TC benefit by getting relevant information (for example, information about workshops, publications, and potential collaboration partners) and joining discussions on the recent advancement and future goals of the fields. The current cochairs of the TC on Biorobotics are K.H. Low, Ravi Vaidyanathan, Jorge Solis, and Justin Seipel. Tianjiang Hu and Samer Mohammed have been actively involved in the organization of recent TC activities and events. With a range of future activities and outreach events, we strongly encourage new members to join the TC and the management team. Please contact Prof. K.H. Low (mkhlow@ntu.edu.sg) for further information.

The IEEE TC on Bioroboitcs China SIG plans to hold one or two workshops (or tutorials) annually in association with a conference closely related to biorobotics.

HUMANITARIAN TECHNOLOGY

2015 Humanitarian Robotics and Automation Technology Challenge

By Raj Madhavan, Lino Marques, Edson Prestes,
Renan Maffei, Vitor Jorge, Baptiste Gil, Sedat Dogru,
Gonçalo Cabrita, Renata Neuland, and Prithviraj Dasgupta

Organized by the IEEE Robotics and Automation Society (RAS) Special Interest Group on Humanitarian Technology (SIGHT), the Humanitarian Robotics and Automation Technology Challenge (HRATC) provides a unique opportunity for the robotics and automation (RA) community from around the world to collaborate using its skills and education to benefit humanity. The RAS SIGHT's mission is the application of RA technologies to promote humanitarian causes in collaboration with global communities and organizations [1].

Started in 2014, the HRATC brings together researchers, students, and roboticists from academia and industry toward realizing a cost-effective, reliable, and sustainable solution to solving the age-old problem of detecting and classifying locations of land mines scattered throughout the globe, serving as sad remnants of war and conflict. Countless people, including children, have been maimed and killed as a result of stepping on land mines buried too close to inhabited areas [2]. The challenge occurs in three phases: 1) simulation, 2) testing, and 3) the finals. Teams are progressively eliminated after each phase, and the remaining teams move on to the next phase, culminating in the challenge (finals) phase. The teams do not need to purchase or build a robot instrumented with sensors or develop any of the accompanying software. Every team can participate remotely in each of the phases. The main goals of the challenge are to

- develop an open-source and free software for reliable and robust detection and classification of land mines and their subsequent clearance
- inspire, encourage, and educate researchers and students on the benefits of deploying RA technologies for the benefit of humanity
- provide a platform for exchanging ideas on addressing pressing needs across the globe via RA technologies. For more details on the HRATC'14 phases and accompanying frameworks, the reader is directed to [3].

The HRATC'15 framework runs on a Linux/Robot operating system (ROS) environment and is responsible for connecting the team code to the robot. The framework also offers simulation scenarios, visualization tools, and scoring metrics. Figure 1 shows the software architecture. This framework has the same core as the HRATC'14 framework; however, the evaluation software (HRATC 2015 Judge) was moved from Python to C++ to improve performance, and the visualization system was modified to use RVIZ, making it consistent with the ROS standard interfaces and easier to use for ROS users.

In the simulation phase, as shown in Figure 1(a), sensor data, such as from cameras and laser range-finder readings, are simulated by Gazebo, through Husky modules, while the metal-detector information is simulated using a custom module based on previously collected information. In the testing phase, as shown in Figure 1(b), the Husky robot provides the sensor data, including that for metal detection.

Figure 2 shows the HRATC framework visualization in RVIZ, with the metrics used by the HRATC judge to compute the scores of each team. Like in the first edition of the HRATC, the score computed by the judge is composed of three components: 1) mine-detection effectiveness, 2) coverage area, and 3) execution time.

Based on our experiences from 2014, we decided to penalize teams that "explode" the robot, eliminating those that pass over a mine more than once. We also penalized teams that were too conservative and inactive. Thus, the HRATC'15 scoring metric was slightly different from the first edition of the challenge. This enabled us to assess each team's performance and, at the same time, penalize inactive teams. In the real world, poor performance or inactivity would imply lost assets, substantial costs, and, possibly, lives.

In the 2015 challenge, Clearpath's Husky robot, shown in Figure 3, was upgraded with a new two-degrees-of-freedom arm, including compliance in both motion axes. If the arm hits the ground or an obstacle, there will be no major damage to the system. The end-effector position is measured by means of absolute encoders attached to the arm's links. The sensor-supporting bridge was also changed so that the arm could have a very large sweeping range, making it possible to place the arm above the robot's body to have a compact system for transportation [4].

Digital Object Identifier 10.1109/MRA.2015.2452199
Date of publication: 11 September 2015

HRATC'15 introduced some changes to the environment, mimicking a more realistic demining operation. The teams had to start from outside the field, and they had to deal with some bush-like obstacles that were added to the test field. The teams had the chance to perceive the environment using either a stereo camera system or a tilted scanning laser range finder, which provides a three-dimensional point-cloud representation of the environment. Given its simplicity and reliability, all the teams chose to use only the laser to detect the obstacles and to extract the ground profile.

Similar to HRATC'14, the field was still defined by four corners; however, these did not necessarily define a rectangle. For the next edition, we plan to provide an aerial image of the minefield and a list of coordinates defining an arbitrary convex polygon. This year, we used three surrogate mines buried in the ground and some metal debris (e.g. cans and metal bars) placed on the field as mock mines. The "real" mines contained only a small metal part, making detection as difficult as detecting a land mine with low metal content. The teams had the chance to use the robot three times before the finals. The results of these tests—videos and ROS bag files of the testing runs—were provided to the teams along with constructive feedback on how to better detect the mines and navigate on the field. The teams used all the opportunities to improve their source code, and all showed significant improvements during the course of the challenge.

In the finals, the teams were allowed two runs on different days, thereby providing them with a chance to modify their source code. The best run of each team was taken to arrive at the final rankings. While deceptively straightforward, the challenge is much more than merely moving the robot platform, detecting and classifying mines, and moving again. There are inherent levels of complexities that are to be dealt with by the teams in terms of sub-tasks such as appropriate minefield mapping, obstacle and land-mine avoidance, and proper arm control, in addition to a robust mine-detection algorithm.

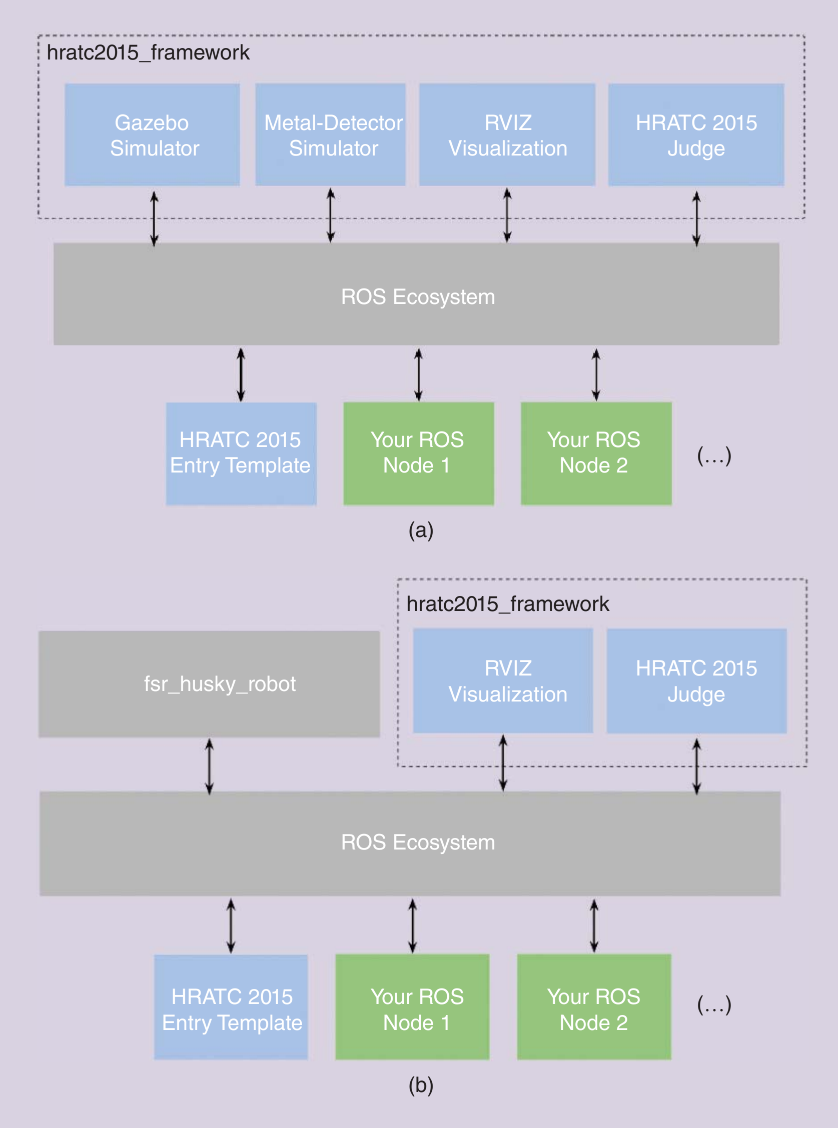
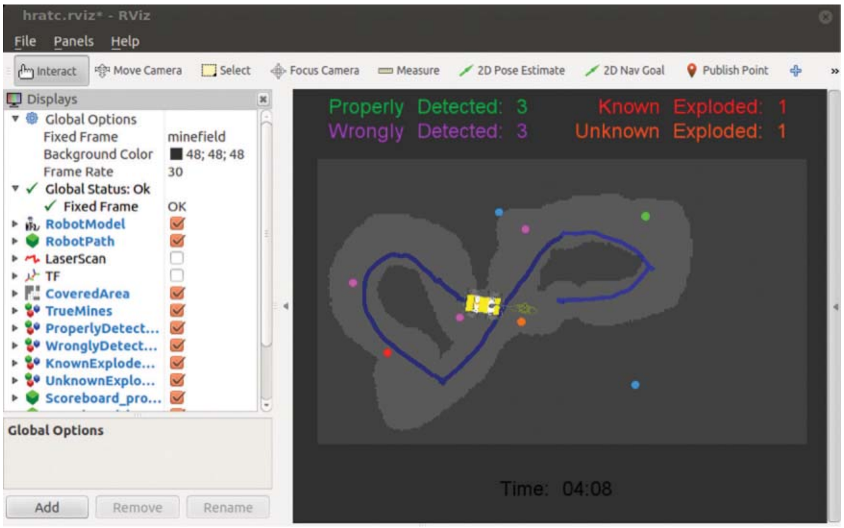


Figure 1. The software architecture for (a) the simulation and (b) the testing phases.



Fixed Frame minefield
Background Color 48; 48; 48
Frame Rate 30
Fixed Frame OK

Properly Detected: 3
Wrongly Detected: 3
Known Exploded: 1
Unknown Exploded: 1
Time: 04:08

Figure 2. The HRATC'15 framework visualization using RVIZ.

Figure 3. The FSR Husky robot employed during HRATC'15 challenge.

HRATC'15 started with 15 teams in the simulation phase, which lasted for 12 weeks. Based on their performance, eight teams progressed to the six-week testing phase. This, in turn, was followed by five teams qualifying for the finals colocated with the Robot Challenge at the 2015 International Conference on Robotics and Automation (ICRA) in Seattle, Washington, 26–27 May. The finals took place remotely, similar to the testing phase, but the results were beamed via a live YouTube channel. National University of Singapore's Team NUS was declared the overall winner, and Team ORION from the University of Texas at Arlington was the runner up (Figure 4). In addition to certificates for the finalists, the top finishers received US$1,000 and US$500, respectively.

Figure 4. The HRATC'15 overall winner: Team NUS. Haoyu Bai (left), Team NUS member and Sonia Chernova, ICRA'15 Robot challenge cochair.

For 2016, we are developing a new robot that will carry a ground-penetrating radar array. We are introducing this second robot for the next challenge, giving the teams an opportunity to implement the multiagent mine-scanning and sensor-fusion techniques. Another aspect we consider important for next year is to encourage the teams to use vision, which is an indispensable tool in field robotics. Integrating vision will also stimulate the participation of larger teams with various backgrounds, further improving teamwork. To enforce this, we will be providing a degraded global positioning system, so the teams will have to rely on visual odometry and other sensing means for accurate localization on the field.

The call for the 2016 challenge will be published in October 2015, with the deadline for applications in November 2015. You can peruse information related to this year's challenge, including rules and frequently asked questions, at http://www.isr.uc.pt/HRATC2015. A summary vídeo is available from http://www.isr.uc.pt/HRATC2015/Lookback.html. We look forward to your participation in HRATC'16!

Acknowledgments

Special thanks are due to IEEE SIGHT for its sponsorship of the prizes. This challenge was partially supported by the European Union Seventh Framework Program TIRAMISU project (http://www.fp7-tiramisu.eu) under grant 284747 and by Clearpath's PartnerBot Program under grant PB12-024. Brazil's Conselho Nacional de Desenvolvimento Científico e Tecnológico program is acknowledged for its partial financial support.

References

[1] R. Madhavan, "RAS-SIGHT formed [society news]," *IEEE Robot. Automat. Mag.*, vol. 20, no. 1, p. 115, 2013.

[2] UN Mine Action Service (UNMAS). [Online]. Available: http://www.mineaction.org/

[3] R. Madhavan, L. Marques, E. Prestes, P. Dasgupta, D. Portugal, B. Gouveia, V. Jorge, R. Maffei, G. Franco, and J. Garcia, "2014 humanitarian robotics and automation technology challenge," *IEEE Robot. Automat. Mag.*, vol. 21, no. 3, pp. 10–16, 2014.

[4] G. Cabrita, R. Madhavan, and L. Marques, "A framework for remote field robotics competitions," in *Proc. IEEE Int. Conf. Autonomous Robot Systems Competitions*, Apr. 2015, pp. 192–197.

Funding Robotics Projects: An Interview with Cécile Huet, Deputy Head of the European Commission Robotics Unit

By Laura Margheri

The European Robotics program is the largest civilian program on robotics worldwide, with more than 100 ongoing projects, over 700 partners, and over €500 million funding. Every year, around €70–80 million of funds are dedicated to new projects.

In Horizon 2020, which is the largest European Research and Innovation program ever with nearly €80 billion of funding available over seven years (2014–2020), the European Commission (EC) has further boosted the robotics program in building a formal public–private partnership (PPP) in robotics with the European robotics community: SPARC (http://sparc-robotics.eu/). The PPP in robotics is based on a contract between the EC and the euRobotics AISBL (http://www.eu-robotics.net/), a non-for-profit association founded in September 2012 to provide the European robotics community, a legal entity to engage in a contract with the EC. In fact, on 17 December 2013, the PPP in robotics contract was signed by the Commission's Vice-President Neelie Kroes and Bernd Liepert, president of the euRobotics.

Through this partnership, a funding of €700 million has been committed by the EC and the private partner committed the triple of that investment for research and development, for a total of €2.8 billion over seven years.

More than 180 member organizations from European industry and research are part of the euRobotics and collaborate with the EC to develop and implement a strategy and a roadmap for research, technological development, innovation, and business in robotics.

On the other side, the role of the EC, and in particular of the Robotics Unit, is to ensure an exchange between the community and implement the strategic research agenda through the funding programs. This also includes the selection of projects, with the help of independent experts, which is a critical and challenging task, requiring a careful selection of top quality expert to ensure a fair and high quality evaluation. This year about 120 experts have been selected and went to Brussels and Luxembourg to evaluate the 194 proposals submitted to the current call.

These initiatives and numbers, both in terms of people involved and in funding, give just an idea of the effort and great collaboration around robotics, within Europe but with a worldwide perspective. In fact, not only the program is limited to European countries but also the projects' consortium can be international to assure a global impact of results and technologies and to give support to the robotics community at a large.

Currently, there is a woman as Deputy Head of the EC Robotics Unit. Cécile Huet is a Belgian electronic engineer. She earned the Ph.D degree from the University of Nice-Sophia Antipolis, France. Huet did her post-doc in California, partly at the University of California, Santa Barbara (UCSB), and partly at Signalcom Inc, a spin-off from UCSB. Then, Huet decided to come back to her mother country, Belgium, to work for a start-up in security and biometrics. Finally, she became an EC official starting in the trust and security unit. After that, she moved to the unit in charge of the cognitive vision program, which was at the origin of the cognitive systems, and then to the robotics unit, created in 2004. She has been a project officer since the creation of the unit, and since September 2014, she has been the deputy head of the unit.

Q: How did you get where you are today?

A: I simply followed my passion, starting from the topics I liked the most at school, and little step by little step, not really sure at the beginning where it would lead me. But on my way, with some endurance and tenacity, I gradually realized that becoming an engineer was possible, and the following steps too! Then, I had the opportunity to move abroad twice and I did not want to miss such opportunities to discover new experiences, environment, and most importantly people. I have to say that on my way, the fact that I was a woman has never been an issue in the choices I

The European Robotics program is the largest civilian program on robotics worldwide.

Digital Object Identifier 10.1109/MRA.2015.2452132
Date of publication: 11 September 2015

made in my career. I would say that I feel really lucky to be where I am today, having such an interesting job, I consider it a mix hard work and luck.

Q: Having a good work/family balance is not always easy. What are your thoughts and your experience about that?

A: This is very challenging and I am not good at keeping a right balance. I am rather extreme in the things I do, either professionally or privately. When I do something, I quickly get passionate and get fully engaged. I would say that I do not consider it a gender issue and feel lucky that at home we share the family tasks. I realize that in general, in our culture, it is still better accepted that the father works more or later, or he travels more frequently than the mother.

Huet excelled and has become an EC official starting in the trust and security unit.

Q: Are there specific actions to encourage female students to study science, technology, engineering, and mathematics (STEM) and for the promoting gender balance from the European Union (EU) commission? Are there particular initiatives within the Robotics Unit?

A: Equality between women and men is one of the EU's founding values. The EC wants to promote gender equality, and in particular gender balance in decision making. At our level, we try to embed such policy in our activities; in particular, we want to promote excellent women, selecting them in committees, as experts to review projects, encouraging them to play leading roles in projects.

The European Robotics Week, for example, is an excellent yearly event to equally expose young boys and girls to robotics technology and hopefully help them develop interest and passion for STEM topics. The same goes for robotics used in education or robotics competition, even starting from early ages.

Q: Considering the different areas of impact of robotics research and innovation projects (social, industrial market, and so on), which are in your opinion the most important for women and where women can have a major role?

A: Traditionally, we see women more present in social human–robot interaction, human factors, rather than mechatronics, and so on. However, the evolution of robotics toward more intelligent and intuitive entities makes the field increasingly multidisciplinary, for instance, combining cognitive systems and robotics research and development to build advanced robots. This should blur the boundaries between the fields and also the division of gender in subdisciplines.

The opportunities for using robotics in new application domains are exploding with robots coming out of their cages where they were so far confined mainly in production lines. Such perspectives, we hope, will attract more women, either as scientists to push the limits of the technology, or as players along the value chain, be it in the product development, in the development of new markets, or as user of robots in professional context. Besides, it is critical to carefully consider the nontechnical issues linked to the deployment of robotics, in particular require experts in ethical, legal, social, and economic aspects of robotics. These disciplines include, in general, more women.

It is therefore important to show them how exciting this field is, the various types of careers and in case they chose to follow this path, make sure to give them the opportunity to live their passion.

Q: It is a fact that most of the leadership roles both in the robotics and in the EU are still covered by men. Can you comment this? How is interfacing with them for you? Do you find any difference when you have to talk with women leaders? (For example: different vision, different objectives, different availability, and different ambition.)

A: It is a fact that the women are still a minority, but there are bright female roboticists and many models out there to show that a brilliant career is also possible for women. In particular, we have in our projects very successful women coordinators, team leaders and scientists, and we want to encourage them to participate in our projects.

In general, we would also like to see better representation of women in high-level decision position in the community. A way of doing that would be to give the bright women in the field visibility to serve as a model for others who might have some fear/doubts about such careers. And gradually, we hope to see more women applying to management boards.

Q: Which message would you like to tell those approaching to robotics career?

A: Follow your passion, believe in yourself, work hard, and when you have doubts look at the models we have. This is a proof for you that it is doable. I would like to address this encouragement to women in particular, but to every student in general. We all have doubts on our way, but the important is to give ourselves the chance to do things in our life that are a source of satisfaction and motivation. The society should make sure to remove potential barriers, along the way, in particularly ensuring equal opportunities.

2015 RAS ADCOM ELECTION

THE IEEE ROBOTICS AND AUTOMATION SOCIETY ELECTION VOTING WILL BEGIN SOON FOR RAS MEMBERS TO ELECT SIX NEW MEMBERS TO THE SOCIETY'S ADMINISTRATIVE COMMITTEE, TO SERVE THREE-YEAR TERMS BEGINNING 1 JANUARY 2016.

IN THE MONTH OF SEPTEMBER, VOTING MEMBERS (GRADUATE STUDENTS AND HIGHER GRADE MEMBERS) WILL RECEIVE THE ADCOM ELECTION INFORMATION PACKAGES DELIVERED VIA E-MAIL OR POSTAL MAIL IF REQUESTED OR E-MAIL IS NOT AVAILABLE. THE PACKAGE INCLUDES A SLATE OF THE CANDIDATES, THEIR BIOGRAPHIES AND POSITION STATEMENTS. CANDIDATE INFORMATION IS ALSO POSTED ON WWW.IEEE-RAS.ORG.

THE CANDIDATES FOR THE SIX POSITIONS ARE:

Geographical Area 1
Jaydev Desai, University Of Maryland, USA
William Hamel, University of Tennessee, USA
Anthony Maciejewski, Colorado State University, USA
Ning Xi, Michigan State University, USA

Geographical Area 2
Gianluca Antonelli, University of Cassino and Southern Lazio, Italy
François Chaumette, Inria Rennes, France
Carme Torras, Institut de Robòtica i Informàtica Industrial (CSIC-UPC) Barcelona, Spain
Jianwei Zhang, University of Hamburg, Germany

Geographical Area 3
Peter Corke, Queensland University of Technology, Australia
Kazuhiro Kosuge, Tohoku University, Japan
Max Q.-H. Meng, Chinese University of Hong Kong, China

IEEE Robotics & Automation Society

Digital Object Identifier 10.1109/MRA.2015.2452198

FROM THE FIELD

Field-Testing Astronaut Assistance Robots in Australian Outback

By Graham Mann, Nicolas Small, Kevin Lee, Jonathan Clarke, and Raymond Sheh

The trouble with field-testing robots is that we are taking complex machines out of the laboratory and into the dirt: natural, unstructured environments that cannot be easily characterized or measured. There they could be doing imperfectly characterized tasks. We expect robots to be behaviorally flexible so describing a typical task will generally underspecify actual usage. The machine design, task, and environment are not orthogonal factors either, since they might interact in complicated ways. As if all this was not enough, most field robots are still teleoperated, which adds the attendant problems of evaluating the human controller and interface. Published work in this area tends to focus on demonstrating the robot's fitness for purpose based on specific requirements, often according to the contingencies of practical funding. Too often that commits the work to studies of performance on tasks that are not necessarily well understood, or even particularly well described, and to measurements within environments that cannot be duplicated.

> **The machine design, task, and environment are not orthogonal factors either, since they might interact in complicated ways.**

Digital Object Identifier 10.1109/MRA.2015.2452200
Date of publication: 11 September 2015

Figure 1. A simulated Mars walk. (Photo courtesy of G. Mann.)

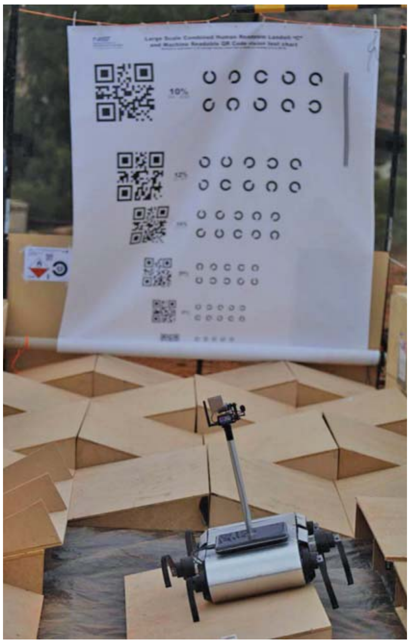


Figure 2. The hexapod vision test. (Photo courtesy of N. Small.)

How can we put field-testing on a more scientific footing? What we need is practical and widely accepted standards for robot testing, followed up by excellent sharing of results and an honest comparison of performance as a function of design. A step in the right direction is the U.S. Department of Homeland Security–National Institute of Standards and Technology–American Society for Testing and Materials (DHC–NIST–ASTM) tests for emergency-response robots. A lot of effort has gone into creating and documenting tens of useful test rigs and task score sheets, which the general community can easily build and use (www.nist.gov/el/isd/ks/upload/DHS_NIST_ASTM_ Robot_Test_Methods-2.pdf).

(continued on page 191)

2015

14–15 September

MFI 2015: IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems. San Diego State, California, USA. http://mfi2015.sdsu.edu/

14–18 September

Summer School on Experimental Methodology, Performance Evaluation and Benchmarking in Robotics. Intur Bonaire Hotel, Benicassim, Castellon, Spain.

28 September–2 October

IROS 2015: IEEE/RSJ International Conference on Intelligent Robots and Systems. Hamburg, Germany. http://www.iros2015.org/

13–16 October

ICCAS 2015: International Conference on Control, Automation, and Systems. BEXCO, Busan, Korea. http://2015.iccas.org/

18–20 October

SSRR 2015: IEEE International Symposium on Safety, Security, and Rescue Robotics. Purdue University, Indiana, USA. https://robotics.purdue.edu/SSRR2015/index.html

3–5 November

Humanoids 2015. Seoul, South Korea. Call for Papers Deadline: 30 June 2015. http://www.humanoids2015.org/main/

6–9 December

ROBIO 2015: IEEE International Conference on Robotics and Biomimetics. Zhuhai, China. Call for Papers Deadline: 25 July 2015. http://ieee-robio.org/2015/

12–13 December

SII 2015: IEEE/SICE International Symposium on System Integration. Nagoya, Japan. Call for Papers Deadline: 31 August 2015. http://www.si-sice.org/SII2015/

2016

16–21 May

ICRA 2016: IEEE International Conference on Robotics and Automation. Stockholm, Sweden. Call for Papers Deadline: 31 August 2015. http://www.icra2016.org/

8–11 July

AIM 2016: IEEE/ASME International Conference on Advanced Intelligent Mechatronics. Ottawa, Ontario, Canada.

Digital Object Identifier 10.1109/MRA.2015.2465892
Date of publication: 11 September 2015

Turning Point

(continued from p. 192)

specific scientific subject. For instance, for mathematical sciences, ensuring reproducibility is quite straightforward since it requires clear reasoning and a complete proof of the presented results, which can be (easily) followed by the readers. Once this is done, the results become part of the common body of knowledge for the community and do not need to be proven again.

The concept is usually specialized and varies for each particular research community.

Conversely, for empirical sciences (such as life sciences, social sciences, and, in the area of interest of the IEEE, e.g., devices/circuits/systems implementation and characterization), the quest for RR involves, at a minimum,

1) a clear description of the methodology followed in a particular study or experiment
2) a detailed explanation of the laboratory procedures/protocols used
3) a thorough statistical analysis of the results obtained, highlighting their significance
4) the complete sharing of the data associated with the study/experiment
5) the sharing of the code and the features of the run-time environment that has (possibly) been used to produce the data.

The extent of this list clearly shows the intrinsic difficulties in guaranteeing reproducibility in this context.

For the computational sciences, which are a subject of interest for several scientific communities within the IEEE, guaranteeing RR mainly involves points 1), 4), and 5), thus resulting in an intermediate difficulty level.

Q: Why is RR important?

GS: Since the times of Galileo and Boyle, the basis of science has been the capability to replicate the results produced by other researchers, to build on their discoveries, to advance knowledge and technology. In other words, reproducing previous results to show the advantages of the proposed innovative methodologies or techniques has always been the key to progress in science. Using Isaac Newton's famous expression, one can summarize this concept: "If I have seen further, it is by standing on the shoulders of giants" [3].

While this scientific approach worked remarkably well for centuries because of the ability of the scientific community to discover and correct mistakes and refine or completely change flawed theories and erroneous methodologies, in recent years something seems to have gone wrong in the self-correcting mechanism of science, particularly in the area of life sciences. Even if one does not consider the most outrageous cases of fraudulent research, such as the famous stem-cell scandal (which involved the retraction of two papers published in 2004 and 2005 in *Science* [4]), several recent studies have highlighted the impossibility of reproducing the results published in the vast majority of the papers under investigation. More precisely, and by way of example, according to Begley and Ellis [5], only 11% of 53 studies in the area of preclinical cancer drugs were reproducible, while Ioannidis et al. [6] show that this was also true for two of 18 papers in bioinformatics. What is worse is that similar findings have made their way into the general public press [7] and generated in the public opinion an increased sense of unease with respect to the way in which science operates.

A systematic adoption of RR practices is certainly necessary to reverse this worrisome trend. At the same time, its implications are far more important than this. RR is, in fact, fundamental since the following hold.

- It will foster growth in the capabilities for collaboration among scientists, which will help to overcome the increasing challenges posed by the rising number of multidisciplinary collaborations.
- It will produce an increase in the rate of innovation: researchers will advance technology more easily, and practitioners will develop new products faster.

This is, of course, a future that every scientist and practitioner will welcome as important steps forward for humanity.

Q: What is RR for the IEEE, and why it is important?

GS: IEEE is first and foremost a professional organization, and its publishing enterprise exists as a service to the community. One of the reasons reproducibility is important for the institute is that there are more indications that RR may actually soon be incentivized (if not mandated) by funding agencies in a similar way with respect to what has happened in recent years for open access. Another reason underlying its importance is that RR may simply become a more pressing request by the IEEE members and authors. More scientists are, in fact, interested in increasing the visibility of their discoveries: preliminary studies show a greater impact for those scientific works that share supplemental material together with the paper itself [8].

Consequently, simply because (part of) the IEEE community will need it, the development of an infrastructure supporting RR (at least in terms of storing/reusing data, code, and algorithms) may become, in my view, a pressing need for the IEEE in the not so distant future.

There are, however, other advantages that the adoption of RR will offer to IEEE. First, as previously mentioned, because of RR, the information made available through the IEEE conferences and publications will be more visible and directly usable by both scientists and practicing engineers. Furthermore, the adoption of RR will help the readers to navigate the large quantity of papers available on a specific subject. By straightforwardly reproducing results, readers will directly test the advantage of a technique with respect to a different one. Finally, promoting RR will make it easier to discover possible false (or inaccurate) results and help the IEEE to maintain its reputation as a world-class scientific/professional organization.

Q: This special issue on reproducibility and measurability of robotics research demonstrated a high interest from the community of the IEEE Robotics and Automation Society (RAS). What opportunities do you foresee for linking this interest to future initiatives in this area which could be launched at the IEEE level?

GS: Developing the necessary infrastructure for supporting RR as well as the best practices associated with it (e.g., in terms of the review process of the data, code, and algorithms associated with the paper) will require substantial work and support from many different IEEE communities. The RAS has already made significant steps in these directions and its experience will be truly precious for the entire organization.

Q: Do you consider research in the robotics and automation domains a key area for applying the principles, methods, and tools for aiming at RR?

GS: Absolutely. Robotics and automation is one of the best areas to apply and test any best practices that the IEEE will develop in terms of RR. In fact, the robotics and automation domains rely on mathematical science as well as computational and experimental ones, so that these experiences pertain to all kinds of reproducibility mentioned in the answer to the second question.

Q: Do you consider these topics important for the training of a new generation of researchers in engineering?

GS: I consider them fundamental. Adopting RR will, in fact, truly change the culture and will require substantial additional effort from the authors publishing with the IEEE. This is, of course, a process that cannot be enforced but only reinforced. We need, therefore, to educate the community, especially the young professionals, to comprehend and embrace the benefits that RR can bring from all different perspectives: authors (visibility increase), users (enhancement in the exploitability of results, increase in capability of recognizing fundamental results) and humanity as a whole (increased rate of innovation).

References

[1] Wikipedia. Reproducibility. [Online]. Available: http://en.wikipedia.org/wiki/reproducibility

[2] J. B. Buckheit and D. L. Donoho, "WaveLab and reproducible research," Dept. of Statistics, Stanford Univ., Stanford, CA, Tech. Rep. 474, 1995.

[3] H. W. Turnbull, Ed., *The Correspondence of Isaac Newton: 1661–1675*, vol. 1. London, U.K.: Royal Society at Univ. Press, 1959, p. 416.

[4] Special online collection: Hwang et al. controversy—Committee report, response, and background. *Science*. [Online]. Available: http://www.sciencemag.org/site/feature/misc/webfeat/hwang2005/

[5] C. G. Begley and L. M. Ellis. (2012). Drug development: Raise standards for preclinical cancer research. *Nature*. [Online]. *483(7391)*, pp. 531–533. Available: http://www.nature.com/nature/journal/v483/n7391/full/483531a.html

[6] J. P. A. Ioannidis et al. (2009). Repeatability of published microarray gene expression analyses. *Nature Genet*. [Online]. *41(2)*, pp. 149–155. Available: http://www.nature.com/ng/journal/v41/n2/full/ng.295.html

[7] How science goes wrong. (2013, 19 Oct.). *The Economist*. [Online]. Available: http://www.economist.com/printedition/2013-10-19

[8] H. A. Piwowar, R. S. Day, and D. B. Fridsma, "Sharing detailed research data is associated with increased citation rate," *PLoS One*, vol. 2, no. 3, e308, 2007, DOI: 10.1371/journal.pone.0000308.

From the Field *(continued from p. 188)*

Importantly, the system also allows researchers to create their own specialized, operational tasks to run in natural settings. We claim that this method can be applied to all kinds of applications.

An opportunity to try this out arose in July 2014 at the Arkaroola Mars Robot Challenge organized by the Mars Society Australia. Four student teams brought six field robots to a test site in Arkaroola, a remote desert station in South Australia. The machines embodied the students' design concepts for assistant robots for astronauts performing tasks on the Martian surface (Figure 1). A selection of six standard DHC–NIST–ASTM benchmarks, together with three operational tests specific to surface operations in harsh Mars-like terrain, was conducted over 12 days. For example, we had the robots search a gullied slope for a hidden target object, which had to be photographed, collected, and returned to the operator.

The test details and results will be formally presented in September at Towards Autonomous Robotic Systems 2015 in Liverpool, United Kingdom, but in brief, we found that most, but not all, tests worked well, provided one practices the procedures and allows enough time (Figure 2). We were not only able to gather a good deal of standard performance data, but we were able to use it later to make real design improvements to two of the robots. Our test program could accommodate unmanned aerial vehicles as well as ground machines: one participant was able to score highly on many tasks using a small quadrotor, suggesting very high utility of a (suitably modified) drone for future Mars explorers.

On Research Reproducibility: An Interview with Gianluca Setti

By Eugenio Guglielmelli

Gianluca Setti (GS) is a professor in the Department of Engineering at the University of Ferrara, Italy, where he teaches circuit theory, analog electronics, and statistical signal processing. He has held several positions as a visiting professor/scientist, such as at École Polytechnique Fédérale de Lausanne, Switzerland (2002, 2005); the University of California, San Diego (2004); IBM T.J. Watson Laboratories (2004, 2007); and the University of Washington, Seattle (2008, 2010). He is also a permanent faculty member of Advanced Research Center on Electronic Systems (ARCES), University of Bologna. His research interests include nonlinear circuits, implementation and application of chaotic circuits and systems, statistical signal processing and compressive sensing, electromagnetic compatibility, and biomedical circuits and systems.

Dr. Setti received the 2013 IEEE Circuits and Systems Society (CASS) Meritorious Service Award, the 2004 IEEE CASS Darlington Award, and the 2013 IEEE CASS Guillemin-Cauer Award as well as the best paper award at the 2005 European Conference on Circuit Theory and Design (ECCTD 2005) and the best student paper award at the 16th International Zurich Symposium and Technical Exhibition on Electromagnetic Compatibility (EMCZurich 2005) and at IEEE International Symposium on Circuits and Systems (ISCAS 2011).

He has also held several editorial positions for the IEEE, including editor-in-chief of *IEEE Transactions on Circuits and Systems—Part II* (2006–2007) and *IEEE Transactions on Circuits and Systems—Part I* (2008–2009); he served as a member of the editorial board of *IEEE Access* (2013–2015) and *Proceedings of the IEEE* (since 2015). He also served as the 2010 CASS president, and, in 2013–2014, he was the first IEEE Vice President for Publication Services and Products from outside North America. Dr. Setti was the technical program cochair of Nonlinear Dynamics of Electronic Systems (NDES 2000) (Catania, Italy), ISCAS 2007 (New Orleans, Louisiana), ISCAS 2008 (Seattle, Washington), IEEE International Conference on Electronics, Circuits and Systems (ICECS 2012) (Seville, Spain), and Biomedical Circuits and Systems Conference (BioCAS 2013) (Rotterdam, The Netherlands) as well as the general cochair of the International Symposium on Nonlinear Theory and its Applications (NOLTA 2006) (Bologna, Italy).


Gianluca Setti.

Q: What is your reference definition of research reproducibility (RR)?

GS: There are several definitions of RR since the concept is usually specialized and varies for each particular research community. In rough terms, and using the definition of reproducibility reported by *Wikipedia* [1], RR is linked to the idea that, as the ultimate product of academic research, papers must be accompanied by the details of the full computational/experimental environment used to produce the results in the manuscript. Knowledge of this environment can be used to reproduce the results achieved by others and will help obtaining new discoveries based on those. I personally also like the definition by Buckheit and Donoho [2], "An article […] in a scientific publication is not the scholarship itself; it is merely advertising of the scholarship. The actual scholarship is the complete software development environment and the complete set of instructions" that generated the results. In other words, RR involves the complete knowledge of the data, the algorithms, the code, and the detailed experimental methods that were used to obtain all the results presented in an article. Another way to see this is that data, algorithms, and code are not simply ancillary information, but first class scholarly products as important as the paper itself.

Q: Are there different kinds of reproducibility?

GS: There are certainly different kinds of reproducibility depending on the particular area of science one deals with. Furthermore, the capability to guarantee reproducibility of research is more or less difficult depending on the

(continued on page 190)

Digital Object Identifier 10.1109/MRA.2015.2458492
Date of publication: 11 September 2015