1. TYPE AND DURATION

Full-day workshop.

2. TITLE

Representing a Complex World: Perception, Inference, and Learning for Joint Semantic, Geometric, and Physical Understanding

3. ORGANIZERS

Nikolay Atanasov (primary contact person)
Assistant Professor
affiliation: University of California, San Diego - Dept. of Electrical and Computer Engineering
address: 9500 Gilman Drive MC 0436, Atkinson Hall, Room 6127, La Jolla, CA 92093
phone: 858-534-4105
email: natanasov@ucsd.edu
url: https://natanaso.github.io/
Bio: Nikolay Atanasov is an Assistant Professor at the Department of Electrical and Computer Engineering, University of California, San Diego, CA, USA. His research focuses on robotics, control theory, and machine learning and in particular on autonomous information collection using ground and aerial robot teams for localization and mapping, environmental monitoring, and security and surveillance. His work proposed probabilistic representations that unify geometry, semantics, and data association in robot localization, mapping, and navigation and developed optimal control algorithms for active minimization of uncertainty in these models. Dr. Atanasov's work has been recognized by the Joseph and Rosaline Wolf award for the best Ph.D. dissertation in Electrical and Systems Engineering at the University of Pennsylvania in 2015 and the best conference paper at the International Conference on Robotics and Automation in 2017.

Luca Carlone
Assistant Professor
affiliation: Massachusetts Institute of Technology - Dept. of Aeronautics and Astronautics
address: 32 Vassar St., Office 32-D716, Cambridge, MA 02139
phone: 404-723-4761
email: lcarlone@mit.edu
url: http://lucacarlone.mit.edu/
Bio: Luca Carlone is the Charles Stark Draper Assistant Professor in the Department of Aeronautics and Astronautics at the Massachusetts Institute of Technology. Before joining MIT, he was a postdoctoral fellow at Georgia Tech (2013-2015), and a visiting researcher at the University of California Santa Barbara (2011). He got his Ph.D. from the Polytechnic University of Turin, Italy, in 2012. His research interests include nonlinear estimation, numerical and distributed optimization, computer vision and probabilistic inference applied to sensing, perception, and control of single and multi-robot systems. He published more than 60 papers in international journals and conferences, including a best paper award finalist at RSS 2015 and a best paper award winner at WAFR 2016.

4. URL

The following website contains the proposed workshop information:

https://natanaso.github.io/rcw-icra18/

Upon acceptance of the workshop proposal, we will create a more accessible alias for the website (e.g., "multimodalRobotPerception.mit.edu").

5. ABSTRACT (max 200 word)

This workshop brings together researchers from robotics, computer vision, machine learning, and neuroscience to examine the challenges and opportunities emerging from the design of world representations that unify semantics, geometry, and physics. The capability of jointly modeling and understanding these aspects is a prerequisite for autonomous operation in complex, unstructured environments. Recent years have seen impressive progress in Simultaneous Localization And Mapping (SLAM), which has been instrumental in transitioning robots from the factory floor to unstructured environments. Indeed, state-of-the-art SLAM approaches can track the pose of a single camera and IMU over long trajectories in real time, while simultaneously providing an accurate dense metric reconstruction of the environment. Surprisingly, however, SLAM has advanced mostly in isolation from the recent equally impressive progress in object recognition and scene understanding, enabled by structured (deformable part) models and deep learning. Few approaches combine spatial and semantic information, despite the tremendous scientific and practical promise of multimodal representations. This workshop aims to bring forward the latest breakthroughs and cutting-edge research on new multimodal representations, as well as novel perception, inference, and learning algorithms that can generate such representations.

6. CONTENT

The goal of the workshop is to bring forward the latest breakthroughs in and foster new ideas on the design of representations and algorithms for joint geometric, semantic, and physical reasoning that can reduce the gap between robot and human perception. This goal is motivated by two fundamental observations. First, the development of advanced perception and world understanding capabilities is a key requirement for robot autonomy in complex, unstructured environments, and an enabling technology for robot use in transportation, agriculture, mining, construction, security, surveillance, and environmental monitoring. Second, despite the unprecedented progress over the past two decades, there is still a large gap between robot and human perception (e.g., expressiveness of representations, robustness, latency). Promoting work at the boundary of geometric, semantic, topological, and temporal understanding can largely expand the capabilities of autonomous systems.

We plan to pursue this goal through three complementary actions:

- **Connect**: the workshop will bring together researchers from different fields, *including robotics, computer vision, machine learning, human cognition, and neuroscience*. This, together with our plan to encourage participation (discussed below), will promote cross-fertilization across research efforts in different fields (e.g., scene understanding in computer vision and SLAM in robotics), and will lead to discussion on how to reconcile different perception approaches (e.g., model-based estimation and learning-based techniques) and design new representations and associated inference and learning algorithms.
- **Inspire**: the workshop will include *invited talks from top researchers* conducting pioneering work across multiple communities aimed at developing joint models for geometric, semantic, and physical understanding. The discussion and interaction among the participants, including established researchers as well as young researchers and students, will be instrumental for attracting interest towards multi-faceted robot perception, catalyzing the research effort in a common direction across the communities.
- **Aspire**: while the previous action is targeted at creating synergies and enhancing existing research efforts, the workshop will also provide a vision of what the robotics community should aim at, when designing novel algorithms and models for robot perception. Towards this goal, we have invited speakers from neuroscience and human and animal perception to gain insights on the advanced perceptual capabilities observed in humans and animals, ranging from vision-based navigation of insects to advanced spatial understanding and visual perception in humans.

These actions are reflected in the structure of the workshop. A tentative schedule is as follows:

MORNING SESSION		
Time	**Topic**	**Status**
8:45 – 9:00 am	Registration, welcome, and opening remarks	-
9:00 – 9:30 am	Invited talk: **Dieter Fox** (University of Washington)	confirmed
9:30 – 10:00 am	Poster Spotlights	-
10:00 – 10:30 am	**Coffee break**	-
10:30 – 11:00 am	Invited talk: **Jitendra Malik** (UC Berkeley)	confirmed
11:00 – 11:30 am	Invited talk: **Jana Kosecka** (George Mason University)	confirmed

11:30 – 12:00 pm	Morning wrap-up: panel discussion	-
12:00 – 1:30 pm	**Lunch break**	-

AFTERNOON SESSION		
Time	**Topic**	**Status**
1:30 – 2:00 pm	Invited talk: **Andrew Davison** (Imperial College)	confirmed
2:00 – 2:30 pm	Invited talk: **Srini Srinivasan** (University of Queensland)	confirmed
2:30 – 3:00 pm	Poster and demo session	-
3:00 – 3:30 pm	**Coffee break & poster session**	-
3:30 – 4:00 pm	Invited talk: **Marc Pollefeys** (ETH Zurich)	tentative
4:00 – 4:30 pm	Invited talk: **Bart Anderson** (University of Sydney)	confirmed
4:30 – 5:00 pm	Afternoon wrap-up: panel discussion & closing remarks	-

The workshop includes three main components: (i) invited talks by leading researchers, (ii) contributed presentations (spotlight, posters, demos), and (iii) interactive panel discussions with invited speakers and all attendees.

Invited talks. The workshop will have a strong multidisciplinary flavor, with a lineup of invited speakers selected among the top researchers in computer vision, robotics, human and animal perception. Invited talks will alternate with spotlight presentations, poster sessions, and panels, in order to preclude separation of audience and maintain high participation. The current set of invited speakers include:

● top researchers in robotics (or working at the boundary between robotics and computer vision), including **Dieter Fox,** who is pioneering learning-based representations bridging geometric and physical understanding, **Andrew Davison**, who pioneered vision-based SLAM and is now doing excellent work on metric-semantic mapping, and **Jana Kosecka**, who has been pushing the boundary of semantic segmentation, scene understanding, and geometric mapping;

- top researchers in computer vision, including **Jitendra Malik**, who is a leading figure in the vision community and has been advocating the need for a unified treatment of reconstruction, recognition, and reorganization, and **Marc Pollefeys**, who is doing groundbreaking research in 3D modelling, segmentation, and semantic understanding;
- top researchers in human and animal perception: including **Srini Srinivasan**, whose research reshaped our understanding of visual processing, perception and cognition in simple natural systems (e.g., insects and animals), and **Barton Anderson**, who is doing cutting-edge research on organization, segmentation, and grouping for human perception.

The invited speakers will be asked to submit their presentation slides for publication on the workshop website. The slides will also be advertised in the Google+ community https://plus.google.com/u/0/communities/102832228492942322585, which has served as a discussion arena for researchers working in SLAM and robot perception and now includes about 500 members.

Contributed presentations (spotlight, posters, demos). We will solicit contributed papers in ICRA format, which will be peer-reviewed by a program committee.[1] Accepted papers will be featured through spotlight presentations and poster presentations. To facilitate interactions among the participants, we will have two poster sessions and we will also encourage *live demos* in the call for papers. The demo session will run in parallel to the poster session, for which we allocated sufficient time both in the morning and in the afternoon session. We plan to solicit submissions in the following areas:
- novel representations that combine geometry, semantics, and physics, and allow reasoning over spatial, semantic, and temporal aspects;
- contextual inference techniques that can produce maximum likelihood estimates over such hybrid multi-modal representations;
- learning techniques that can produce cognitive representations directly from complex sensory inputs;
- approaches that combine learning-based techniques with geometric estimation methods;
- position papers and unconventional ideas on how to reach human-level performance across the broad spectrum of perceptual problems arising in robotics.

Camera-ready versions of the accepted papers will be published on the workshop website and compiled into a single PDF file to serve as the *workshop's proceedings*. Invited speakers, contributing authors, and participants will receive a copy of the proceedings via email.

We will advertise the workshop broadly across multiple communities (more details in the "Plan to solicit participation"). Moreover, we will assign a best paper award (taking into account the live demos as a selection criterion) with a monetary award of 500$ presented to the most compelling/provocative/innovative contributed submission.

Depending on the number and the quality of the submissions, we plan to organize a special issue on robot perception in the IEEE Transactions on Robotics (see *Dissemination* section).

Panel discussion. The morning and afternoon sessions will each conclude with a panel discussion featuring panelist selected among the invited speakers and contributing authors. In our previous experience, having multiple short panel discussions on specific topics makes the

---

[1] The program committee is yet to be assembled.

workshop more interactive, while a standard format with a single final panel discussion may be more dispersive and address an already tired audience. We plan to collect a set of questions for the panelist ahead of time. As we did in previous workshops, we plan to solicit early questions in the Google+ community (https://plus.google.com/u/0/communities/102832228492942322585) and during the workshop via twitter and a Google submission form advertised on the workshop website. We will invite questions from the audience and keep the floor open for broader discussion.

7. PLAN TO SOLICIT PARTICIPATION

The confluence of recent impressive progress in fields related to physical modeling and interpretation -- robotics (localization and mapping), computer vision and machine learning (super-human object recognition), neuroscience (the Human Brain Project) -- provides an unprecedented opportunity to share ideas across disciplines, discuss open problems, and establish a research direction in robotics aimed at enabling human-level world understanding. We will use a variety of ways to attract participation from these disciplines, which include researchers and practitioners that have not traditionally participated at ICRA. First, we will rely on traditional workshop announcements, including an abstract and call for contributions, on **relevant mailing lists** in robotics (robotics-worldwide; euRobotics), machine learning (Machine Learning List, http://cll.stanford.edu/mllist/, Uncertainty in AI), and neuroscience (Connectionists, EUCog). Second, the workshop will be advertised in collaboration with the IEEE Robotics and Automation Society **technical committee on Computer & Robot Vision** (See Sec. 11). In the past, we participated in the organization of several closely related workshops (see below), which led to the creation of a **Google+ community**. The community (https://plus.google.com/communities/102832228492942322585) currently includes close to 500 members and will be used to support, direct, and solicit participation in the activities related to the workshop. A particularly effective way for soliciting high-quality contributions in previous workshops has been to send **personal invitations** to top research groups working in the targeted fields. This year we plan to capitalize on this observation by supplementing the call for contributions with a solicitation for **live demos** and an announcement of a **best paper/poster award** for the workshop. In a related event (RSS'15: The Problem of Mobile Sensors), we were able to provide a **monetary prize** for the best poster, sponsored by the Australian Centre for Robotic Vision. Finally, high-quality papers contributed to the workshop will be invited for a **IEEE Transactions on Robotics Special Issue** (see Sec. 9). We expect these efforts to encourage researchers traditionally outside of the robotics community to contribute to our workshop.

The proposers of this workshop have previously organized several successful conceptually-related workshops:
- RSS'14 Workshop on "*Multi View Geometry in Robotics*"
  - URL: http://rll.berkeley.edu/RSS2014/workshops.html
  - Organizers: V. Indelman, **L. Carlone**, F. Dellaert
  - Estimated number of participants: 90
- RSS'15 Workshop "*The problem of Mobile Sensors: setting future goals and indicators of progress for SLAM*"

- URL: https://ylatif.github.io/movingsensors/
- Organizers: C. Cadena, **L. Carlone**, H. Carrillo, Y. Latif, J. Leonard, J. Neira, I. Reid
- Estimated number of participants: 120
- RSS'16 Workshop "*Geometry and Beyond - Representations, Physics, and Scene Understanding for Robotics*"
  - URL: http://rss16-representations.mit.edu/
  - Organizers: E. Galceran, **L. Carlone**, G. Tipaldi, L. Paull, A. Censi, C. Cadena
  - Estimated number of participants: 100
- RSS'16 Workshop on "*Robot-Environment Interaction for Perception and Manipulation: Interactive Perception Meets Reinforcement Learning and Optimal Control*"
  - URL: http://rss16ip-rl-oc.robotics.usc.edu/
  - Organizers: K. Hausman, H. van Hoof, **N. Atanasov**, R. Martin Martin, O. Brock
  - Estimated number of participants: 70
- RSS'17 Workshop on "*Learning Perception and Control for Autonomous Flight: Safety, Memory, and Efficiency*"
  - URL: http://www.ece.ucr.edu/~kkarydis/rss17/
  - Organizers: K. Karydis, **N. Atanasov**, S. Levine, N. Roy, C. Tomlin, V. Kumar
  - Estimated number of participants: 120

Our goal is to use these efforts as a starting point to promote a discussion on how to unify the work in robotics, machine learning, and neuroscience through a common interface -- the representation of the world. Based on previous years' attendance, we anticipate more than 100 participants.

8. PLAN TO ENCOURAGE INTERACTION AMONG PARTICIPANTS

We have multiple ideas to make the proposed workshop an interactive and captivating event. The traditional practice of having one long panel discussion at the end of the workshop is ineffective, in our opinion, because workshop participants are tired by then and this results in a shorter/less interactive discussion of interesting questions and ideas that arose during the talks. Nonetheless, oftentimes the most interesting ideas come out of discussion rather than invited/contributed talks and, hence, our goal is to have **multiple discussions throughout the day** -- one in the morning, one in the afternoon, and at least 5 min for Q&A at the end of each invited talk. These discussions will be further encouraged by soliciting **online submission of questions and discussion topics from the participants**, both before the workshop takes place (via the Google+ community) and during the workshop (via a Google form on the workshop website). A poster **spotlight session and a dedicated poster exhibition** as part of the workshop schedule will serve to involve early career researchers into the discussion. As a way to encourage interaction between established experts and early-career researchers further, we will consider inviting the **authors of the contributed papers to participate along with the invited speakers as panelists** during the morning and afternoon panel discussions. In addition to discussing fundamental perception, inference, and learning approaches from a traditional robotics and machine learning perspective, the panel discussion will focus on eliciting key ideas

on environment representation from neuroscience and animal and human cognition. Another key ingredient to encourage interaction will be the **demo session** which will run in parallel with the poster exhibition. We hope that this broad spectrum of ideas will inspire interaction among researchers from different fields and will encourage the participants to attend the whole event.

9. DISSEMINATION

The workshop will solicit submissions (up to 5 pages following ICRA formatting guidelines) focusing on the challenges of unified geometric, semantic, topological, and physical representations, and associated perception, inference, and learning algorithms. Contributed papers will be reviewed by the organizers and a program committee of 8-10 invited reviewers. The topics of interest and their relevance to the state of the art are described in Sec. 6 above.

Camera-ready versions of the accepted papers will be published on the workshop website (https://natanaso.github.io/rcw-icra18/) and compiled into a single PDF file to serve as the **workshop's proceedings**. Invited speakers, contributing authors, and participants will receive a copy of the proceedings via email. Digital versions of the **posters and slides** of the invited talks will also be archived on the workshop website. Finally, we plan to record all the talks and put the corresponding **video recordings on YouTube**, with links on the workshop website.

To further improve dissemination of the scientific contributions presented during the workshop, we have also been in contact with Frank Park, Editor-in-chief of the IEEE Transactions on Robotics, for the organization of a **special issue in the IEEE T-RO**. Prof. Park (fcp@snu.ac.kr), Editor-in-chief of IEEE Transactions on Robotics, confirmed his availability to provide an endorsement letter, on behalf of the T-RO editorial board, to support the workshop. Depending on the quality of the workshop submissions, we plan to invite the authors of the top submissions to extend their papers and submit them to the T-RO special issue.

10. EQUIPMENT

The workshop will include oral presentations, poster sessions, and panel discussions. We expect the following equipment to be sufficient:
- Presentation equipment: projector, laser pointer, screen, VGA/HDMI adapters,
- Poster equipment: 10-20 poster stands/easels and tape in case of traditional posters, or monitors for digital poster presentation,
- Panel discussion equipment: flip chart and markers,
- Recording equipment (if possible) to allow providing the invited talks on YouTube and on the workshop website.

11. SUPPORT OF AN IEEE RAS TECHNICAL COMMITTEE

The workshop is endorsed by the IEEE RAS Technical Committee for *Computer & Robot Vision*. Prof. Jana Kosecka (kosecka@cs.gmu.edu), co-chair of the TC, confirmed her availability to provide a formal endorsement letter, on behalf of the TC, to support the workshop.