The Workshop of "Physical and Social Scene Understanding"

Tao Gao (taogao@mit.eud)
Department of Brain and Cognitive Sciences, MIT
Cambridge MA, 02139 USA

Yibiao Zhao (ybzhao@ucla.edu)
The Center for Vision, Cognition, Learning, and Art, UCLA
Los Angeles CA, 90095 USA

Lap-Fai (Craig) Yu (craigyu@cs.umb.edu)
Computer Graphics Lab, University of Massachusetts
Boston MA, 02125 USA

Keywords: Causality; Physics; Functionality; Intentionality

Theme

Computer vision has made significant progress in locating and recognizing objects in real images. However, beyond the scope of this "what is where" challenge, it lacks the abilities to understand scenes characterizing human visual experience. The mission of this workshop is to (a) identify the key domains in which human visual perception and cognition outperform computer vision; (b) formalize the computational challenges in these domains; and (c) provide promising frameworks for solving these challenges by conducting cognitive science and computer vision studies.

We propose FPIC as four key domains beyond "what is where":

Functionality (e.g., what can be done with this slotted spoon?)

Physics (e.g., will the spoon be able to pick up the meatball?)

Intentionality (e.g., is the person trying to scoop up the cheese or point toward it?)

Causality (e.g., why does the gravy pass through the spoon?)

The combination of these largely orthogonal dimensions can span a large space for scene understanding. Despite their apparent differences, these domains do connect with each other in ways that are theoretically important: (a) they usually don't project onto explicit visual features; (b) existing computer vision algorithms are neither competent in these domains, nor (in most cases) applicable at all; and (c) human cognition is nevertheless highly efficient at these domains. Therefore, studying FPIC should significantly fill the gap between computer vision and human vision. On the one hand, human studies on FPIC-related topics can inspire the invention of novel, cognitively-motivated computer vision systems. On the other hand, state-of-the-art computer vision systems can expand the scope of cognitive sciences to address challenges in real scenes.

The introduction of FPIC will advance cognitive models in three aspects: (a) transfer learning. As higher-level representation, FPIC tends to be globally invariant across the entire human living space. Therefore, learning in one type of scenes can be transferred to novel situations; (b) small sample learning. Leaning of FPIC, which is consistent and noise-free, is possible even without a wealth of previous experience or "big data"; and (c) bidirectional inference. Inference with FPIC requires the combination of top-down abstract knowledge and bottom-up visual patterns. The bidirectional processes can boost the performance of each other as a result.

Several key themes are:

Physically grounded scene interpretation
Causal model of vision and cognition
Reasoning about goals and intents of agents in scenes
Human-object-scene interaction
Top-down and Bottom-up inference algorithms

In conjunction with CogSci 2015, our "Physical and Social Scene Understanding" workshop will bring together researchers from cognitive science, computer vision and robotics, to illuminate cognitively-motivated vision systems going beyond labeling "what is where" in an image. These systems work closely together to achieve a sophisticated and coherent understanding of scenes with respect to Functionality, Physics, Intentionality and Causality (FPIC). In effect, these systems are expected to answer an almost limitless range of questions about an image using a finite and general-purpose model. In the meanwhile, we also want to highlight that FPIC is never meant to be an exclusive set of scene understanding problems. We welcome the insights of scholars who share the same perspective but are working on different problems.

Speakers

We will invite speakers working in cognitive science, computer vision, computer graphics and robotics , who have fundamental insights of visual understanding. We plan to choose eight speakers from the list below, but are not limited to.

Confirmed Speakers:

Phillip Wolff (Professor, *Cognitive Science*, Emory)
Intuitive Physics and Causality

Adam Sanborn (Professor, *Cognitive Science*, Warwick)
Intuitive Physics and Causality

Joshua Tenenbaum (Professor, *Cognitive Science*, MIT)
Intuitive Physics and Theory of Mind

Jason Fischer (Post-Doc, *Cognitive Neuroscience*, MIT)
Neural circuits of physical and social perception

Song-Chun Zhu (Professor, *Computer Vision*, UCLA)
Causal Parsing with Commonsense Reasoning

Brian Scassellati (Professor, *Robotics*, Yale)
Social Robotics

Brian Ziebart (Professor, *Robotics*, UIC)
Purposeful Prediction

Demetri Terzopoulos (Professor, Computer Graphics, UCLA)
Artificial Life

Workshop Program

Our workshop will be a full day workshop hosting talks of eight invited speakers who are leading researchers in their research fields.

- Each speaker will have 35 minutes to present.
- One-hour panel discussion at the end.
- All talks and discussions will be video recorded and posted on the workshop website.

Tentative Schedule:

9:00am - 9:10am Welcome speech
9:15am - 9:45am Invited talk 1
9:55am - 10:25am Invited talk 2
10:30pm - 11:00 am Invited talk 3
11:05pm - 11:35 am Invited talk 4
11:40am - 1:00pm Lunch Break
1:00pm - 1:30pm Invited talk 5
1:35pm - 2:05pm Invited talk 6
2:10pm - 2:40pm Invited talk 7
2:45pm - 3:15pm Invited talk 8
3:20pm - 4:20pm Panel Discussion

Potential Financial Support

We are currently looking for sponsorship from the Center of Brian, Mind and Machine (CBMM) at MIT. We are planning to get support from our sponsors (e.g. Microsoft Research, A9) of our previous workshop again. The sponsorship will be used for covering the travel fees of some of our invited speakers; making souvenirs for our contributors and workshop participants; recording a video for all of our talks. With this support, we aim at organizing a workshop that every participant enjoys, and further boosting the impact of our workshop and the CogSci conference.

Success of Our Previous Workshop

We held a related workshop on "Vision meets Cognition" at CVPR 2014 (a premiere computer vision conference). The workshop was highly successful and very well-received, attracting around 300 audience. This indicates the enthusiasm towards recent cognitive science studies in the computer vision community. All talks were video recorded and posted online at:

http://www.visionmeetscognition.org/fpic2014/

Here is a brief summary of our previous workshop. We are dedicated to continue the success at CogSci 2015.

Speakers: We invited 8 keynote speakers from the computer vision, cognitive science and computer graphics fields: Song-Chun Zhu, Katsushi Ikeuchi, William T. Freeman, Demetri Terzopoulos, Josh Tenenbaum, Ashutosh Saxena, Benjamin Kuipers, Larry Zitnick. Their talks provided diverse insights from different perspectives which are all highly relevant to the theme of our workshop: Vision meets Cognition.

Audiences: There were 367 people who signed up for our workshop during the registration. Our keynote talks were very popular and most of the time our room was fully seated. At peak time, our workshop attained attendance of over 200 workshop participants.

Accepted papers: There were 38 carefully peer-reviewed papers accepted by our workshop, which involved more than 200 paper authors. Each paper was reviewed by 2-4 experts in the field, chosen among our 30 program committee. We broadcasted a trailer video composed of spotlight slides to promote all of our accepted papers.

Sponsors: Our workshop has also aroused significant industry interest, and was generously supported by 4 sponsors: Microsoft Research, A9, Vicarious and the Office for Naval Research.

Souvenirs: All of our invited speakers received our custom-designed souvenirs in recognition of their contribution. We also designed and manufactured 200 magic mugs for all our workshop attendants and guests.

Based on our organizing experience, we are highly confident that the workshop of "Physical and Social Scene Understanding" at CogSci 2015 can achieve an even bigger success.