Editorial: Special Issue on Active and Interactive Methods in Computer Vision

Kristen Grauman · Serge Belongie

While traditional computer vision research aims to replace humans in visual analysis tasks, in many emerging applications there is a need for vision algorithms that instead *assist* humans or *cooperate* with them to accomplish recognition and learning tasks. Interactive vision methods allow a human-in-the-loop to inject high-level expertise, while the system carries out low-level processing and/or leverages its own sophisticated statistical models, thereby improving with each iteration. Meanwhile, active learning approaches have the potential to limit the human involvement in such systems to only where it is most crucial. For both types of techniques, recent developments in crowdsourcing, large-scale datasets, and social networks have led to new opportunities and technical challenges.

This special issue provides a snapshot of a number of interesting developments in the last few years for active and interactive vision methods. The topics solicited included: methods to actively minimize human effort, human-in-the-loop interactive systems, crowdsourcing issues relevant to vision, and ways to instill human input for machine learning beyond traditional labels.

Submissions to the special issue closed in Spring of 2013. All submissions were reviewed by at least three reviewers. Three of these manuscripts (including two on which a guest editor was an author) were handled by other IJCV editors to avoid possible conflicts of interest. We have selected a set of nine papers for the special issue.

A few of the papers explore ways in which a human user and computer vision system can come together to solve a particular annotation task. The idea is for the system to iteratively request the user's input where it is expected to be most informative. Steve Branson, Grant Van Horn, Catherine Wah, Pietro Perona, and Serge Belongie introduce a fine-grained recognition approach where the system guides a user through simpler questions that help it deduce an image label, yielding a system where "ignorant" non-expert humans lead the "blind" computer to correctly identify the bird species in an image. Adarsh Kowdle, Yao-Jen Chang, Andrew Gallagher, Dhruv Batra, and Tsuhan Chen propose an active approach to image-based modeling that has applications for 3d modeling on mobile devices; it solicits useful annotations on multiple views of an object to cosegment and reconstruct accurately across viewpoints with minimal intervention. Continuing the theme of human assisted annotation, Stefanie Jegelka, Ashish Kapoor, and Eric Horvitz develop a decision-theoretic criterion to solicit human help for partial correspondence refinement, such that ambiguous image-to-image matches can be verified and propagated correctly by the system.

Other papers in this special issue address the challenge of how humans—typically "workers" in a crowdsourcing platform—can effectively prepare the vision system to perform a visual recognition task. This includes both novel annotation strategies to elicit useful annotations that transcend traditional object labels, as well as scalable learning algorithms to let the system build models on the fly. Genevieve Patterson, Chen Xu, Hang Su, and James Hays report on a large-scale crowdsourcing effort to gather a taxonomy of attribute labels on the SUN scenes, and they explore how attributes may be particularly amenable to scenes, where category boundaries are often ambiguous. Subhransu Maji and Gregory Shakhnarovich introduce an annotation framework that elicits from humans the relative properties of objects, and demon-

K. Grauman (✉)
University of Texas at Austin, Austin, TX, USA
e-mail: grauman@cs.utexas.edu

S. Belongie
Cornell Tech, New York, NY, USA
e-mail: sjb344@cornell.edu

strate applications for semantic part discovery and attribute vocabulary formation. Sudheendra Vijayanarasimhan and Kristen Grauman develop a live large-scale active learning approach that autonomously trains object detectors using millions of images crawled on the Web and actively sought crowdsourced labels.

Finally, three of the papers explore issues in interactive discovery from image data. Caroline Galleguillos, Brian McFee, and Gert Lanckriet develop a human-assisted object category discovery approach where new clusters of unlabeled image regions are iteratively detected at the same time kernels are learned to provide accurate nearest neighbor comparisons. Arijit Biswas and David Jacobs formulate an active

clustering approach that pinpoints pairwise image similarities that would be best answered by a human annotator, so as to improve the image clustering. Leveraging latent clusters of user preferences, Ashish Kapoor, Juan Caicedo, Dani Lischinski, and Sing Bing Kang develop a personalized model for image enhancement, where shared tastes in image post-processing effects are discovered from the crowd with minimal effort per user.

Altogether, the papers in this issue offer diverse perspectives on human-machine cooperation for solving visual analysis tasks, and each offers novel technical contributions towards that underlying goal.