AAAI Fellowship Application

John Oberlin
Brown University, 2014

1 Introduction

State of the art techniques in object detection and pose estimation are powerful and general but usually run at a rate much less than 1 Hz and require time and expertise to build, maintain, and operate. The high demands of modern systems can make it difficult to employ such techniques in real-time human-computer interaction.

During the completion of my PhD thesis at Brown, I will make it possible for a robot to autonomously scan 10 novel objects in order to construct robust models for detection, pose estimation, grasping, and manipulation of those objects during collaborations with a human operator. A high level description of the workflow that I use at the moment is:

The Workflow

1. Collect RGB-D data for objects from many different perspectives.
2. Train BoW model and kNN classifiers for objects.
3. Use the classifiers and additional logic to provide 3D detections and pose estimates of objects.

It is possible to execute this workflow for 5 objects in 15 minutes, producing models which are viable for tabletop object detection. Generating models robust enough for general detection might take three times as long.

Popular techniques for real-time detection use modified DPMs [7] [4], and sometimes exploit different channels of data [5] [6]. These approaches have seen success in their target domains, but are too technical for general application and need to be integrated with interactive systems. I want to create a Computer Vision system that is simple, reliable, and easy to use with ROS (Robot Operating System).

2 Coordinating AI with Computer Vision and Teaching the System

My current framework for detection uses a box metaphor to talk about space in a way that is amenable to planning and reasoning. The system uses RGB-D video from a Kinect as input. Green Boxes denote areas in the image that probably belong to an object. Each Blue Box denotes a cluster of green boxes that probably belong to the same object. Whereas Blue Boxes exist only for one frame of video input, Red Boxes denote more precise clusters of Green Boxes which are tracked over time and account for occlusion.

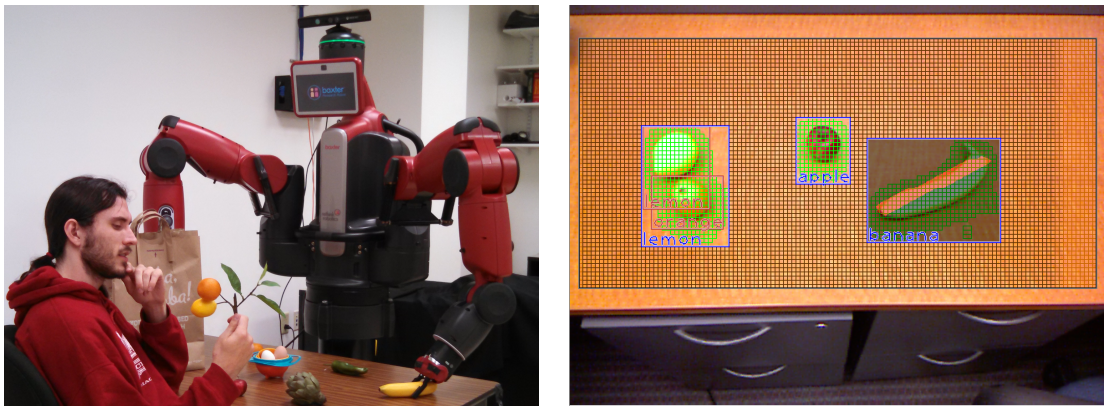

Figure 2.1: Teaching a robot to identify and manipulate objects can be as easy as bringing home a bag of groceries.

Right now, detections and pose estimates are provided for Blue and Red Boxes. I want to allow the system to provide information about tokenization, affordances, and other properties that are useful for applying MDPs [1], OO-MDPs [2], and POMDPs [3] to the real world.

The Red Box optimization process can be framed as an MDP. I would like to use more sophisticated planning that incorporates feedback with the robot's movement planner in order to collect additional, specific data as it is needed.

Finally, consider the following automatic object registration technique. The logic in Step 3 of The Workflow might utilize Amazon Mechanical Turk or an existing, more general classifier to automatically label and retrieve meta information for learned objects, meaning that the operator only needs to annotate data if the automatic registration step fails.

3 Current Work, Future Progress and Broader Impact

At the moment, a human operator trains the system without a robot by manually adjusting the pose of the object, achieving robust detection and pose estimation under good conditions. Eventually we would like to have fully automatic training, where the robot grabs objects and coordinates base poses itself.

The system we have outlined abstracts the detector from the classifier, taking the notion of Objectness and extending it. As more sophisticated Computer Vision algorithms become real-time tractable or new data becomes available, they might be smoothly incorporated into the system even as it is running. Automatic model validation can ensure that a new model outperforms the current model before the system switches it out with no interruption of service. This is important if robots are to become effortless for non-technical operators.

References

[1] R. Bellman, *A markovian decision process*, Indiana Univ. Math. J., 6 (1957), pp. 679–684.

[2] C. Diuk, A. Cohen, and M. L. Littman, *An object-oriented representation for efficient reinforcement learning*, in Proceedings of the 25th international conference on Machine learning, ACM, 2008, pp. 240–247.

[3] L. P. Kaelbling, M. L. Littman, and A. R. Cassandra, *Planning and acting in partially observable stochastic domains*, Artificial intelligence, 101 (1998), pp. 99–134.

[4] M. A. Sadeghi and D. Forsyth, *30hz object detection with dpm v5*, in Computer Vision–ECCV 2014, Springer, 2014, pp. 65–79.

[5] Y. Sun, L. Bo, and D. Fox, *Learning to identify new objects*, ICRA, (2014).

[6] S. S. J. Xiao, *Sliding shapes for 3d object detection in depth images*, ECCV, (2014).

[7] M. Zhu, N. Atanasov, G. J. Pappas, and K. Daniilidis, *Active deformable part models inference*, in Computer Vision–ECCV 2014, Springer, 2014, pp. 281–296.

4 Attendance Statement

I have been paying attention to AI, Machine Learning, and Computer Vision since 2004. Regarding Computer Vision, I saw the progress of Neural Nets in the 90's swept under the rug by SIFT, HoG, and SVMs in the mid 2000's, only for neural nets to reclaim the throne in the 2010's. At one time it was said that AI was "vision hard" and that solving Vision would effectively solve AI. While that belief sweeps a bit under the rug, it is certainly true that in the past Computer Vision has been a strong bottleneck in the development of AI and Robotics.

Recent advances in object detection on large data sets suggest that we are ready to move beyond attacking vision in an isolated setting and begin integrating it in a larger framework for planning in an interactive environment. I have training in state of the art Computer Vision techniques and have been tracking the literature for a few years now.

Work in Computer Vision has largely centered around hyperspecific problems concerning single images. This focus has resulted in significant progress on such tasks, but the challenges of engineering real time systems has so far prevented interesting real world applications or substantial spread of techniques to other communities.

Two things have resulted from this dam in the flow of information. First, there are now many techniques in Computer Vision which are suited to fast and effective partial solutions of problems such as object detection but have been ignored in favor of much slower but slightly more effective state of the art approaches. Second, such techniques may be combined with the information available from extra sensors and the ability of real time systems to capture additional images of the same scene to form full solutions to multiple related problems (such as detection, segmentation, and pose estimation).

I have seen already that combining elements of probabilistic planning and reasoning can boost the performance of traditional Computer Vision and Machine Learning algorithms. I want to continue this approach to research, and attending AAAI will help me to fill in gaps in my background that are necessary to carve out a career in multi-disciplinary AI. Planning and Markov Decision Processes are key elements that I want to master. Additionally, I would like to explore topics such as Interactive Entertainment, Scheduling, Knowledge Representation and Reasoning, and Reasoning Under Uncertainty in order to further my understanding of Human-AI Interactions.

Finally, by attending AAAI I will be able to better understand the motivations, needs, and desires of AI and Robotics concentrators with regards to Computer Vision, which will enable me to complete my dissertation in a way that aligns with these communities' values.

5 Curriculum Vitae

Education

BS in Math, Florida State University, 2003-2006
MA in Math, UC Berkeley, 2006-2008
PhD program in Computer Science at U Chicago, 2010-2011
PhD program in Computer Science at Brown University, 2011-Present

Employment

Developer Support Engineer, Havok, 2008-2009

Conference Papers

S. Naderi Parizi, J. Oberlin, P. Felzenszwalb.
Reconfigurable Models for Scene Recognition.
IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2012

P. Felzenszwalb, J. Oberlin.
Multiscale Fields of Patterns.
To Appear, 2014

Conferences Attended

CVPR 2011
CVPR 2012

Teaching Experience

Being a TA in the UCB Math Deparment can be close to a full time teaching position, including quiz design, office hours, and grading in addition to the intense recitation sections (which more closely resembled lectures).

Calculus 1 TA
UC Berkeley, 2006-2007
Two sections in the Fall, three sections in the Spring, 30 students in each section, each section met with me two or three times a week for a total of about three hours a week.

Linear Algebra and ODE TA
UC Berkeley, 2007-2008
Same configuration as calculus.

Algorithms Grad TA
Brown University, 2012-Present

I am currently the Grad TA for the Algorithms class at Brown for the fourth semester. During this time the class has been between 45 and 100 students each iteration. I have been responsible for administering oral exams, holding office hours, lots of grading (most of our problems are proof based), and managing a group of at least 6 Undergraduate TA's each semester.

Departmental Service

Graduate Student Orientation Leader
Brown CS Department, Fall 2012

Computer Vision Reading Group Coordinator
Brown CS Department, 2012

Department Tea Organizer
Brown CS Department, 2012-Present

Misc Research

Master's Thesis in Mathematics
UC Berkeley, Written 2006-2008

Research Experience for Undergraduates in Mathematics
Oregon State University, Summer 2005

Undergraduate Research Program in Physics
Florida State University, Summer 2004

Research Assistant in Molecular Biology
Florida State University, Summer 2003

Hobbies

Gardening
Ballroom Dance
Martial Arts
Blacksmithing
3D Printing

6 Letter From Supervisor