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 less than 1 Hz. This makes it difficult to employ such techniques in real-time human-computer interaction. This document outlines a simple, robust framework for object detection which trades a large memory overhead for improvements in latency and total throughput of detections. Included is a workflow for that framework which makes training and calibration first intuitive and then automatic.

Our overall pipeline can be described as:

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

# 2 Basic Techniques and Detector Structure

Objectness Fast Keypoints SIFT descriptors KMeans BoW Color Histogram Depth Histogram kNN MDP

Green Boxes Blue Boxes Red Boxes

# 3 Teaching and Employing the System

This is teaching not training because it is an interaction used to collect data not a script used to optimize a function.

Teaching is carried out on a flat surface of uniform color. The background plane and color model are inferred by looking at parts of the image not contained in the blue boxes. These are used to keep with high probability only keypoints which are on the object. During inference (employment), background contributions will contribute to all examples approximately equally, and any confusion is likely to be understandable.

**Manual Teaching** Teaching the system about an object involves showing it many different views of the object in a controlled environment. Once a session is initiated, the following process should be performed:

1. Repeat k times:

(a) Adjust the object to expose a novel viewpoint and record pose data.
(b) Remove extraneous objects in the scene to ensure a unique blue box is present.
(c) Collect the view, which saves RGB-D data and the background data.

Where k might be around 100. Now the collected data can be used to train object class and pose classifiers.

**Semi-Automatic Teaching** Suppose that we want to gather data while a robot is holding the object. It is possible to impose geomentric constraints on kept data to ensure that captured keypoints belong to the object being examined and not to our manipulating robot. This is called self filtering.

1. Repeat k times:
   (a) The object is placed in a robot's manipulator in a known grasp.
   (b) A base pose for the grasp is provided.
   (c) The robot collects view from many different precisely known poses, together with models for self and background filtering.

Where k might be around 3. Eventually we would like to have fully automatic training, where the robot grabs objects and coordinates base poses itself.

**Employing the System** Object Poses on Table or Floor Single Target Tracking

# 4 Work to be Completed

Semi-automatic training Automatic training More geometric work

working MDP's into the red box exploration and data collection so that we can interact with large scale planners for tasks.

# 5 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. Although I have attended CVPR twice, I have not attended any conferences in AI or Robotics. I recently started to focus on Real Time Vision systems, and attending AAAI will help me rapidly learn about the community and hit the ground running. As my concentration shifts and I begin working in AI and Robotics, my assimilation of the literature will be enchanced by having attended AAAI.

One of my goals is to make a real time vision system with a capacity of 50 objects. It should be capable of giving accurate 3D pose estimates that enable the objects to be identified and manipulated by robots. It should be able to learn a new object in less than a minute, and should be simple enough that a non-expert can easily train the system. Yet it should remain versatile enough that state of the art detectors running on special hardware can be swapped in by experts.

I have made good progress towards this goal so far. 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.

## 6 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

## 7 Letter From Supervisor