Carnegie Mellon

Thesis Proposal

COMPUTER SCIENCE DEPARTMENT

Utilizing Non-Image Information for Fine Grained Estimation Tasks

Yair Movshovitz-Attias

Tuesday, February 3, 2015
5:00 p.m.
GHC 6501

The computer vision community has made great advances in recent years in a number of complex tasks, such as scene classification, object detection, and image segmentation. A key ingredient in such success stories is the use of large amounts of labeled data. In many cases, the limiting factor is the ability to create high quality labeled training sets. This issue can arise in two forms: (1) The act of labeling the data can be hard for human annotators, and (2) In some cases it is hard to get enough data to get a representative sample of the feature space.

Our work on business storefront classification is an example of (1). The number of possible labels is large, and assigning all relevant labels to an image is a time consuming task for human annotators. Moreover, when the image contains a business from a country other than their own, annotators can get confused by its purpose and produce erroneous labels. Annotators are also not consistent in their categorization of businesses into categories.

In our work on vehicle viewpoint estimation, the images themselves are hard to come by. Getting sample images of every possible viewpoint is hard as there is a bias in the way people photograph cars. In addition, human annotators frequently have a hard time precisely labeling such images with the correct pose oftentimes the vehicle.

We are able to solve these issues by matching images to real world concepts. In the case of businesses, when images are mapped to an ontology of geographical entities, we are able to extract multiple relevant labels per image. For the viewpoint estimation problem, by using 3D CAD models we can render images in the desired viewpoint resolution, and assign precise labels to them. Our proposed work will focus on large scale fine pose estimation and creating a system that is virtually unlimited in its access to training data, by leveraging a large database of production quality 3D CAD models. We will create a dataset of high quality photo-realistic renders that is an order of magnitude larger than previous work and show that it outperforms state of the art methods in estimating the fine pose of objects in images.

Thesis Committee:
Takeo Kanade, Co-Chair
Yaser Sheikh, Co-Chair
Abhinav Gupta
Leonid Sigal, Disney Research
Trevor Darrell, UC Berkeley

Thesis Summary: http://www.cs.cmu.edu/~ymovshov/misc/yma_proposal.pdf