Visual Recognition

Raquel Urtasun

September 19, 2011

Abstract

Developing autonomous systems that are able to assist us in everyday's tasks is one of the grand challenges in modern computer science. While a variety of novel sensors have been developed in the past few years, in this class we will focus on the extraction of this knowledge from visual information alone. One of the most remarkable examples of successful recognition systems is our visual system, which is able to extract high-level information from very noisy and ambiguous data. Unfortunately, despite decades of research efforts, machines are still way below human performance. In this class we will study why this is the case.

The goal of this graduate class is to understand the different visual recognition tasks as well as the techniques employed to solve them. A strong component of the course will be statistical learning as it plays a key role in almost every modern visual recognition system. We will cover all stages of the recognition pipeline: low-level (e.g., features), mid-level (e.g., segmentation) as well as high-level reasoning (e.g., scene understanding).

Knowledge of machine learning and computer vision is not required, but highly recommended. The theoretical aspects of visual recognition will be covered during the lectures. The class will have a strong practical component, as the students will build the different recognition components during the homework sessions. A list of topics includes:

1. **Classification**: features, bag of words (BOW), similarity between images, learning features as well as hashing schemes and retrieval.
2. **Detection**: sliding window approaches, branch and bound, structure prediction, hough voting and NN approaches, hierarchical models.
3. **Segmentation**: classical approaches (e.g., watershading) as well as modern structure prediction approaches including message passing and graph cuts for inference, and CRFs and structured-SVMs for learning.
4. **Modern 3D geometry and 3D scene understanding**: stereo, scene layout (e.g., 3D box for indoor scenes, road layout for outdoor scenes).
5. **Pose estimation**: pictorial structures (2D) as well as 3D pose estimation including particle filter-based approaches.