Object recognition and localization

Andrew Zisserman
University of Oxford, UK

Abstract

There has been much recent research activity - and much recent success - in recognizing visual object categories (such as cars, cows, motorbikes, people) in images. Recognition can simply be classifying an image by its content ("What is in this image?"), or it can also include specifying object locations ("Where is it in this image?").
The recent success has come from a combination of new visual representations together with powerful learning algorithms.

One representation is to model an object simply by a set of local iconic image patches, where each patch may be thought of as a "visual word" for describing part of the object. Somewhat surprisingly object categories can be recognized without including the spatial organization/location of the patches, and these models are referred to as a "bag of words" in analogy with similar models in the statistical text understanding literature. However, in order to determine object locations, it is necessary to also represent an object's configuration or layout.

This lecture will introduce several representations including: bag of words, spatial pyramid, and pictorial structures; and will cover the learning and application of classifiers and detectors built from these representations. The lecture will also touch on class specific segmentation, and learning category models with weak or noisy supervision.

Syllabus: Image classification, object recognition and localization, weak supervision, segmentation, bag of words models