Object Category Recognition & Localization

Mark Everingham

University of Leeds, UK

Abstract

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

Successful representations for recognition have in part been built on ideas from the text retrieval community, making an analogy between images and documents. The "bag of words" representation considers the images as an unordered collection of iconic image patches or "visual words". Somewhat surprisingly, object categories can be recognized accurately using such a representation which discards all spatial information. Related work in the vision community has re-introduced varying levels of spatial information, for example in the form of a pyramid structure which implicitly solves a correspondence problem between image features. Part-based models explicitly combine appearance and spatial arrangement of object parts to allow robust recognition of deformable objects.

In concert with the development of new visual representations, recent advances in machine learning have also played an important role. Algorithms for large-scale training of support vector machines have allowed their application to real-world datasets. Discriminative training of graphical models such as the conditional random field has been successfully applied to learning of part-based models. Multiple kernel learning has provided a principled way to combine heterogeneous visual descriptors. Structured output learning has supported learning models to directly tackle the object localization problem.

The area of object recognition has also come of age in terms of quantitative evaluation on real-world data, with the provision of standard datasets such as the Caltech series, and the annual PASCAL Visual Object Category (VOC) recognition challenge. As algorithms become better, this has also led to issues on what and how to evaluate methods.

This lecture will discuss the state-of-the-art in object recognition and localization, covering visual representations, learning methods and evaluation.

Syllabus: Object Recognition, Object Localization