



Object and Scene Recognition

Antonio Torralba
Massachusetts Institute of Technology, USA

Abstract

Despite of the research challenges still ahead, the last few years have seen an explosion of commercial applications of computer vision techniques. Computer vision is now common place among many commercial products, such as digital cameras, web applications, security applications, etc. In addition to the algorithmic advances required to solve object recognition, a key component to progress is access to data in order to train computational models for the different object classes. This situation has dramatically changed in the last decade, especially via the internet, which has given computer vision researchers access to millions of images and videos. In this tutorial I will describe recent work on visual scene understanding that try to build integrated models for scene and object recognition, emphasizing the power of large database of annotated images in computer vision. I will focus on:

Multiclass object detection: I will review work focussing on building systems for detecting many object classes. Multi-object recognition systems can use similarities between object categories to their advantage by learning features which lead to better generalization.

Scene understanding: In complex, natural scenes, object recognition systems can be improved by using contextual knowledge about the objects. I will present recent work on scene recognition and on modeling the relationships between objects in order to improve detection performance and to provide a richer scene representation.

Image annotation and collecting ground truth data: Despite that algorithmic advances are part of the key ingredients that made possible recent progress in computer vision, one of the key factors that made this progress possible was access to really large amounts of training data. Building large databases of annotated visual data is a challenging task. I will summarize some of the recent work on building large image databases.

Keywords: Multiclass object detection, Scene understanding, Image annotation and collecting ground truth data