Abstract

Category-level object recognition is a fundamental capability for the potential use of robots in the assistance of humans in useful tasks. There have been numerous vision-based object recognition systems yielding fast and accurate results in constrained environments. However, by depending on visual cues, these techniques are susceptible to object variations in size, lighting, rotation, and pose, all of which cannot be avoided in real video data. Thus, the task of object recognition still remains very challenging.

My thesis work builds upon the fact that robots can observe humans interacting with the objects in their environment. We refer to the set of objects, which can be involved in the interaction as 'interactionable' objects. The interaction of humans with the 'interactionable' objects provides numerous non-visual cues to the identity of objects.

In this thesis, I will introduce a flexible object recognition approach called *Multiple-Cue Object Recognition* (MCOR) that can use multiple cues of any predefined type, whether they are cues intrinsic to the object or provided by observation of a human.

In pursuit of this goal, the thesis will provide several contributions: A representation for the multiple cues including an object definition that allows for the flexible addition of these cues; Weights that reflect the various strength of association between a particular cue and a particular object using a probabilistic relational model, as well as object displacement values for localizing the information in an image; Tools for defining visual features, segmentation, tracking, and the values for the non-visual cues; Lastly, an object recognition algorithm for the incremental discrimination of potential object categories.

We evaluate these contributions through a number of methods including simulation to demonstrate the learning of weights and recognition based on an analytical model, an analytical model that demonstrates the robustness of the MCOR framework, and recognition results on real video data using a number of datasets including video taken from a humanoid robot (Sony QRIO), video captured from a meeting setting, scripted scenarios from outside universities, and unscripted TV cooking data.

Using the datasets, we demonstrate the basic features of the MCOR algorithm including its ability to use multiple cues of different types. We demonstrate the applicability of MCOR to an outside dataset. We show that MCOR has better recognition results over vision-only recognition systems, and show that performance only improves with the addition of more cue types.