Efficient Multiple Object Detection in Cluttered Scenes

Erik Murphy-Chutorian
Dept. of Electrical and Computer Engineering
University of California, San Diego
San Diego, California 92037-0409
erikmc@ucsd.edu

Jochen Triesch
Dept. of Cognitive Science
University of California, San Diego
San Diego, California 92037-0515
triesch@ucsd.edu

1. Abstract

Detection of multiple objects in cluttered scenes is a fundamental challenge that has only recently been widely undertaken by computer vision systems. The presence of object transformations, occlusion, and clutter compounds the difficulty of this problem, and the addition of more objects to be detected can dramatically increase the computational requirements of traditional detection techniques. We present a system that overcomes both of these limitations using a bottom-up detection process that shares features between and within different object representations. The concept of feature sharing was originally developed to demonstrate an increase in multiple-object detection performance when using a fixed number of features and hence a fixed computational complexity [3]. We demonstrate that the reverse is true – that by sharing features, we can achieve excellent detection performance while minimizing the time required to process a scene. Due to the bottom-up nature of our search, computational complexity depends not on the number of objects, but the number of shared features used to represent the objects. In earlier work we presented a system for sharing features that was limited to a single feature type [1]. With this research we extend this approach with a probabilistic framework for weighting the contribution of each feature and integrating multiple types of features.

During the training phase, conditional-probabilistic associations based on the log-liklihood ration are learned between a set of dictionary features and the objects. During the detection phase, these associations act as weights with which features vote for the presence of an object. Given a novel scene, features are extracted at interest points, and a nearest neighbor search matches each feature with one from a feature dictionary. Each matched feature votes for the presence of an object at a specific location, and the object is detected from the maximum a posteriori estimate based on the consensus. This two-fold probabilistic approach is necessary to deal with cases in which an inter-feature independence assumption is not valid. To demonstrate the utility of this approach, we train a system to recognize any of 50 objects in everyday scenes with substantial occlusion. Without further optimization we also demonstrate near-perfect recognition on the standardized COIL-100 database [2].

This work was supported by the National Science Foundation under grants IIS-0208451 and IIS-0233899.

References

[1] E. Murphy-Chutorian and J. Triesch. Shared features for scalable appearance-based object recognition. In *Proc. IEEE Workshop Applications of Computer Vision*, January 2005.

[2] S. Nene, S. Nayar, and H. Murase. Columbia object image library (COIL-100). Technical report, Columbia University, February 1996.

[3] A. Torralba, K. Murphy, and W. Freeman. Sharing features: efficient boosting procedures for multiclass object detection. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2004.

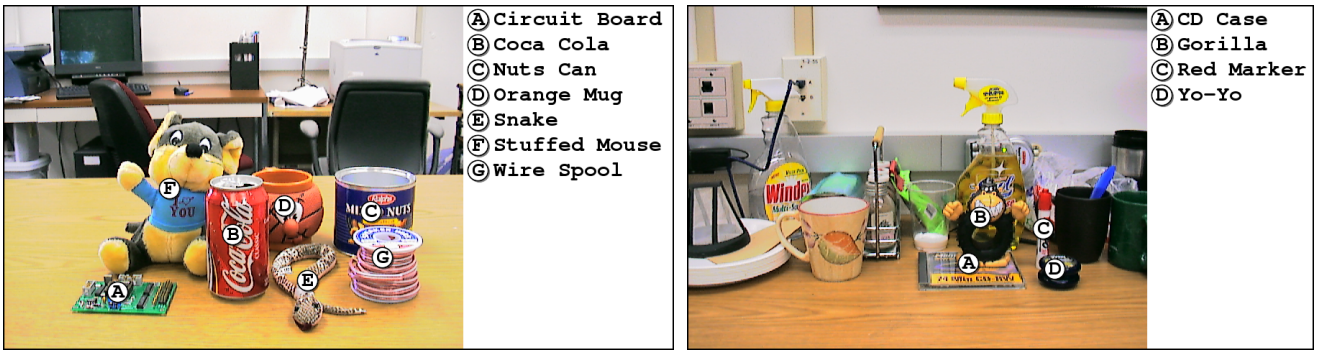

Figure 1. (left) Example object detection (right) Recognition in the presence of distracting objects