International Workshop on Computer Vision, April 26-30, 2004, IPM, Tehran

A System for Observing and Recognizing Objects in the Real World

J.-O. Eklundh
The Computer Vision and Active Perception Laboratory
Royal Institute of Technology (KTH)
Sweden

Currently there is considerable interest in object recognition and categorization. Several advanced methods have been developed, many of them based on statistical methods. A major application and the most common test case concerns search on databases of images, especially asking the question "does this image contain the object X" or " an object belonging to class Y".

In our work we are interested in performing similar tasks on systems looking at a real scene, such as a robot looking for objects in an indoor environment. The question this raises is if existing methods apply directly. Is it possible to take developed appearance based and statistical techniques and use them directly? Our findings indicate that this is not the case, at least not if one wants to system to behave robustly in a realistic environment.

Our talk will focus on a few pertinent problems from the point of view of an integrated system: figure-ground segmentation and selecting appropriate views, the role of 3D cues, and the use of multiple cues. Humans looking around in the world can, seemingly without effort, segment out and distinguish different objects in the world. The corresponding capability has largely eluded the efforts of researchers in computer vision. Figure-ground segmentation in general needs both context and task to be well-defined, i.e. may not be addressed using information in the visual scene alone. However, 3D cues play a special role: they indicate physical chunks that in turn can be ascribed visually observable 3D properties, such as position, location and motion, and object intrinsic properties such as shape, color and maybe surface and material characteristics. One of the things we will discuss is segmentation of the scene into figure and ground and more generally into layers using 3D cues. Related to this problem is the question of finding appropriate views for detecting and recognizing objects. This has in our system been solved by a foveation mechanism. Finally, integration of multiple cues is a key aspect of our approach and probably necessary for robust performance. Techniques for this will also be discussed.