NSF Project : CPS

ROBOTS WITH VISION THAT FIND OBJECTS

Consider the following problem: A robot is instructed to find a few objects (from generic classes) in a cluttered room. For example, the robot may be asked to find an “apple” or a “cup”. We propose to develop tools for actively detecting and localizing objects, and describing their visual appearance in terms of shape primitives. A solution to this problem will provide a powerful module for a plethora of robotics applications.

The problem of object recognition has been a core problem in Computer Vision, and in recent years has received a great amount of attention. However, all efforts have been devoted to passive processing of single images from large image databases based on appearance-based feature descriptions, and avoided any segmentation of the object and description in the form of the objects’ surfaces.

In contrast, our approach is active and bio inspired. We propose to develop a cyber-physical robot that is inspired by human vision. We advance a novel viewpoint to the attention system of a robot, by introducing a robust mechanism for top down attention. Segmentation of objects from images, the holy grail of vision based robotics, has not really served the needs of robotic systems. We turn the segmentation problem into its head and redefine it in an anthropomorphic sense. Instead of segmenting the whole scene at once, the robot will segment only one object, the one to which attention has been deployed . We propose to develop perceptually meaningful descriptions that capture object-hood. For this we will utilize the internet to learn from large image-databases the appearance of object categories and to develop better low-level image processing that will allow us to segment objects and characterize object boundary contours.

Our approach consists of three main components. First, we will develop a new attention mechanism for the robot to actively search the visual scene. Clearly, a real-time vision system should not process every image on its camera/retina by extracting all the features, building descriptors, and then comparing them to every single image in their memory. Instead it should shift its attention to the visual areas of interest very fast. Along this line of thinking we propose a computational attention mechanism based on filters tuned to objects and object parts. Using labeled data from the internet we propose to learn visual filters tuned to object classes. Second, we will develop a fixation-based figure ground segmentation. The robot will fixate on a point within the interesting area detected by the object filter and segment the surface containing the fixation point using contours and depth information from motion and stereo. Third, we will develop a description of the segmented object, in terms of the contours of its visible surfaces and a qualitative description of their 3D shape. The approach is based on learning to classify the different edges separating surfaces and active exploration of the robot by changing the position and motion of its body and head to facilitate qualitative shape reconstruction. The descriptors obtained this way will constitute the input to the recognition engine.