

NSF Project: CPS. Period: 2010-2013

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 are working on 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 are working towards developing 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.

Our approach consists of three main components. First, we 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 develop a computational attention mechanism based on the novel concept of the image torque. Second, we develop a fixation-based figure ground segmentation. The robot will fixate on a point within the interesting area detected by the torque map and segment the surface containing the fixation point using contours and depth information from motion and stereo. Third, we 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. Our work towards these three components proceeds in parallel. By the end of the project, all components will have been integrated and a public demo will be scheduled.

Recently it became clear that our problem is closely related to cognition and language understanding. Indeed, if we ask the robot to find “scissors”, the robot needs to have an understanding of “scissors” in terms of representations that ground the meaning of the word. Our results towards this development are encouraging.