Active Scene Analysis

Niklas Bergström, Mårten Björkman, Jeannette Bohg, Matthew Johnson-Roberson, Gert Kootstra and Danica Kragic
Computer Vision and Active Perception Laboratory (CVAP)
KTH, Stockholm
Email: {nbergst,celle,bohg,mattjr,kootstra,danik}@kth.se

I. Introduction

The future household robots operating in human environments will encounter unknown scenes and unknown objects. In our ongoing work we aim to perceive and represent such scenes. The difficulty with unknown scenes and objects is that no prior knowledge can be used to guide the perceptual process, only bottom-up mechanisms are available for analyzing the scene.

By exploiting the active capabilities of the robot, much more information about the scene can be gathered. We utilize this fact by controlling the gaze of our robot and by haptically exploring the occluded parts of the scene. Doing so, we detect and segment the objects in the scene and make a spatial representation.

II. Methods

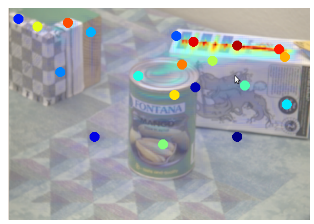
The setup in our lab consists of a 6 DOF arm with a 7 DOF three fingered hand with tactile sensors, and a 7 DOF active head with wide-field and foveal stereo cameras. Figure 1(a) and 1(b) show the scene from the robot's perspective. This setup allows for a system for active scene perception. First, object hypotheses are generated which are used to initialize segmentation. Objects are segmented from the background using two-dimensional (2D) and three-dimensional (3D) information. This gives an initial representation of the scene. However, due to occlusions, large parts of the scene are unknown. To gather more information we haptically explore these parts, thereby creating a more complete spatial representation of the scene.

A. Generating object hypotheses

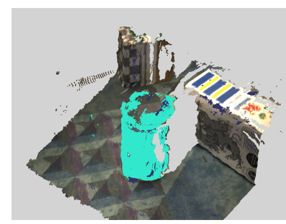
The Gestalt psychology has identified a number of cues for detecting objects in a scene in human perception. Symmetry is one of these cues [1]. Inspired by this, we proposed a visual-attention method based on symmetry for generating object hypotheses [2]. The symmetry method outperforms the saliency method of Itti *et al.* [3] based on contrast, which we successfully used in earlier work [4]. We showed that our method has a higher success rate in generating valid object hypotheses, and finds them closer to the center of the objects, which is beneficial for object segmentation. In Figure 1(c) the 20 most salient attention points are shown.

(a) Scene as seen from the left wide-field camera

(b) Scene as seen from left foveal camera



(c) Attention points. From red to blue, red being most salient.



(d) Reconstructed view with the middle object segmented

Fig. 1. Object hypotheses generation and segmentation.

B. Segmenting objects

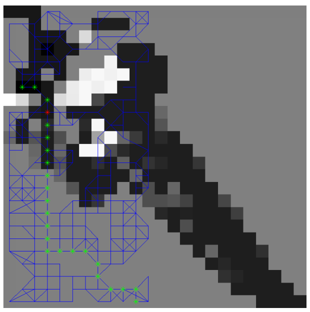
The attention points are used as input for a segmentation method that segregates the attended object from the background. Our method performs segmentation using color and contrast information, as well as disparity information from the stereo cameras [5]. By using depth information the method can deal with heterogeneously colored objects. In addition, we assume a table-top scenario. By finding the dominant plane in the image – the table – the parts of the object close to the table can be distinguished from the table itself. Using Belief Propagation, every pixel is labeled as foreground, background, or table plane. By successively letting the robotic head visit the attention points belonging to the different objects, a detailed representation of the scene can be made. An example is given in Figure 1(d).

C. Exploring the scene

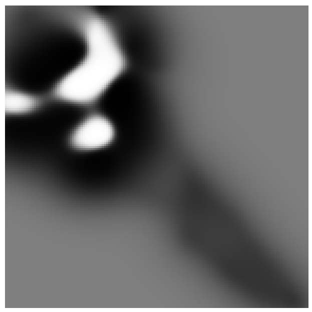
From the previously described visual exploration and object segmentation, we can reconstruct a 3D scene that is segmented into initial object hypothesis. This scene model is incomplete due to for instance, occlusions. We propose a method for multi-modal exploration to, confirm and augment the initial scene model. We represent the scene from the top as an 2D occupancy grid. This is used for motion planning purposes [6]. An example can be seen in Fig. 2(b). We update the current belief about the state of the occupancy grid using contact

(a) Scene state.

(b) Occupancy Grid. White is occupied, black is free space



(c) Exploration plan maximising the expected information gain. 2D graph shows possible paths. Green stars label the chosen path. The red star labels the current robot position.



(d) Predicted occupancy grid.

Fig. 2. Scene Estimate after 42 Exploration Steps.Grey tones in the OGs correspond to probabilities of occupancy. The brighter a cell, the more probable it is that it is occupied.

information from haptic sensors on our robotic hand. Yet unknown parts of the scene are predicted with a Gaussian process. A predicted occupancy grid is depicted in Fig. 2(d). Using a Gaussian process also allows us to quantify the uncertainty about different regions in the scene map. An active exploration strategy is proposed in which most uncertain areas in the map are preferred over more certain ones. Through the integration of different sensor modalities, we achieve a more complete scene model. Additionally the prediction of the scene structure leads to an operationally useful representation even if the map is not fully traversed.

References

[1] J. Driver, G. C. Baylis, and R. D. Rafal, "Preserved figure-ground segregation and symmetry perception in visual neglect," *Nature*, vol. 360, pp. 73–75, 1992.

[2] G. Kootstra, N. Bergström, and D. Kragic, "Using symmetry to select fixation points for segmentation," in *International Converence on Pattern Recognition*, 2010.

[3] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254–1259, 1998.

[4] B. Rasolzadeh, M. Björkman, K. Huebner, and D. Kragic, "An active vision system for detecting, fixating and manipulating objects in the real world," *The International Journal of Robotics Research*, vol. 29, no. 2-3, pp. 133–154, 2010.

[5] M. Björkman and D. Kragic, "Active 3d scene segmentation and detection of unknown objects," in *International Conference on Robotics and Automation (ICRA)*, 2010.

[6] S. T. O'Callaghan, F. Ramos, and H. Durrant-Whyte, "Contextual occupancy maps using gaussian processes," in *2009 IEEE International Conference on Robotics and Automation*, Kobe, Japan, May 2009.