Extending Visual Perception With Haptic Exploration for Improved Scene Understanding

Jürgen Leitner
ARC Centre of Excellence for Robotic Vision (ACRV)
Queensland University of Technology, Brisbane, Australia
j.leitner@roboticvision.org

Abstract

Scene understanding has been investigated from a mainly visual information point of view. Recently depth has been provided an extra wealth of information, allowing more geometric knowledge to fuse into scene understanding. Yet to form a holistic view, especially in robotic applications, one can create even more data by interacting with the world. In fact humans, when growing up, seem to heavily investigate the world around them by haptic exploration. We show an application of haptic exploration on a humanoid robot in cooperation with a learning method for object segmentation. The actions performed consecutively improve the segmentation of objects in the scene.

1. Introduction and Related Work

Understanding its environment is arguably one of the most important tasks to build autonomous robotic systems. Cameras provide a cheap and high-dimensional sensing of the scene. Yet in robotic settings a major difference is that the robot can interact, either by simply changing its viewpoint of the scene, or by actively "poking" objects in the world. Figure 1 shows our robot poking an object in the scene. This haptic perception (from the Greek ηαπτικός, is commonly used during the earlier stages of human development. According to Lederman and Klatzky [5] a few commonly used haptic exploration procedures exits: from lateral motion to applying pressure. These active patterns are used to optimize the extraction of information the perceiver needs to obtain from a scene. While this exploration provides information about, e.g. the material of objects that cannot be achieved by vision alone, we show that it is also beneficial for visual perception tasks. In robotic settings fusing interaction with vision seems especially useful as it also allows to physically ground the interpretation of objects. This in turn can be used to infer or learn object affordances.

We argue that a tight integration of vision and action will provide specific information for understanding the environment, as these interactions with the environment provide and create valuable information to build better visual systems. Discoveries in neuroscience suggest multiple interactions between the visual and motor streams in our brain [1]. In infants various specializations in the visual pathways may develop for extracting and encoding information relevant for visual cognition, as well as, information about the location and graspability of objects [4]. This suggests that an active exploration of scenes is improving the connection between graspability and visual perception [2, 9], furthermore improving detection skills. This is further backed by recent studies on how stimuli from different sensor modalities create "response bias" in the visual perception, leading to, e.g. increasing the saliency of a visual event [8].

Using a humanoid robot we show that by performing actions in an environment we can improve one specific part of scene understanding: segmentation of target objects.

2. Proposed Method

The chosen example is the segmentation of scene based on some visual representation of the objects of interest. To show that the actions can provide useful information for the segmentation task, we use a framework called Cartesian Genetic Programming for Image Processing (CGP-IP) [3]. It uses a mixture of primitive mathematical and high level operations. It uses CGP, which appears to be a popular choice for the representation in this domain. It encompasses do-

Figure 1. Example of a haptic exploration action (poking).

main knowledge, and could even be easily adapted to include the findings from previous work in its function set. It creates human-readable code based on OpenCV to be run directly on the real hardware. We chose CGP-IP as it can adapt (by continuous evolution) to changing input data. In addition it provides a nice way of verifying that the segmentation is more robust (based on the fitness values and the code produced).

CGP-IP has previously been shown to allow for the autonomous learning of object segmentation on a robotic platform [6]. Building on this we propose using CGP-IP to evolve a module, which segments specific objects based on a collected training set of various scenes. As a baseline we use a static observation setup, during which the robot is not moved. The fitness of each evolved individual is determined by calculating the Matthews Correlation Coefficient (MCC) [7] on a set of validation images. The MCC is calculated based on the count of the true positives, false positives, true negatives, and false negatives pixels.

3. Experimental Setup and Results

The robot is performing 4 pre-programmed actions to interact with the world. The actions vary from a simple leaning action (which is equivalent to a simple camera viewpoint change), to poking, pushing and even picking up the object in question.

A new training set is collected which contains the images from the static scenario and the images after one action was performed. A comparison of the segmentation fitness for each of these 4 newly collected datasets and the baseline is shown in Table 1. Figure 2 shows visually the improvement when trying to segment a target object in one of the validation images. On the left is the evolved filter solely based on the "static scene baseline" the segmentation, on the right is the segmentation when integrating the new observations during an haptic exploration action.

4. Discussion

We argue that a closer integration between vision and action will provide more information which can in turn be used for a more holistic understanding of the scene.

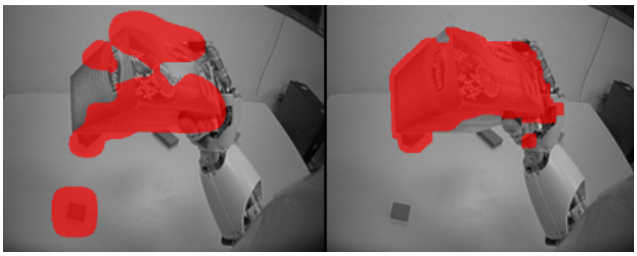

Figure 2. Improved segmentation (right) in a validation image.

Table 1. Comparing the evolved visual detectors after various actions on a validation image set. *Note: Smaller fitness is better!*

Detector	**Fitness** $(1 - \lvert MCC \rvert)$
RedTeabox Start	0.47
RedTeabox `LEAN`	0.45
RedTeabox `POKE`	0.35
RedTeabox `PUSH`	0.45
RedTeabox `CUR.`	0.36

Already just moving the camera's viewpoint allows to help with separating geometries of the scene that would otherwise be hard to separate. In addition when applying forces to objects one can reason about relationships between objects, like, "are two objects (inseparably) connected", as well as, finding out other physical properties, like, "is the juice box full or empty".

A preliminary experiment using a segmentation method that can adapt to new observations on a humanoid robot, showed that by performing some haptic exploration procedures the performance of the visual segmentation can be improved.

References

[1] N. Berthier, R. Clifton, V. Gullapalli, D. McCall, and D. Robin. Visual information and object size in the control of reaching. *Journal of Motor Behavior*, 28(3):187–197, 1996.

[2] J. Grèzes and J. Decety. Does visual perception of object afford action? evidence from a neuroimaging study. *Neuropsychologia*, 40(2):212–222, 2002.

[3] S. Harding, J. Leitner, and J. Schmidhuber. Cartesian genetic programming for image processing. In *Genetic Programming Theory and Practice X*, Genetic and Evolutionary Computation, pages 31–44. Springer, 2013.

[4] M. H. Johnson and Y. Munakata. Processes of change in brain and cognitive development. *Trends in cognitive sciences*, 9(3):152–158, 2005.

[5] S. J. Lederman and R. L. Klatzky. Extracting object properties through haptic exploration. *Acta Psychologica*, 84(1):29 – 40, 1993.

[6] J. Leitner, P. Chandrashekhariah, S. Harding, M. Frank, G. Spina, A. Förster, J. Triesch, and J. Schmidhuber. Autonomous learning of robust visual object detection and identification on a humanoid. In *International Conference on Development and Learning and Epigenetic Robotics (ICDL)*, 2012.

[7] B. W. Matthews. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochimica et Biophysica Acta*, 405(2):442–451, 1975.

[8] E. Van der Burg, C. N. Olivers, A. W. Bronkhorst, and J. Theeuwes. Poke and pop: Tactile–visual synchrony increases visual saliency. *Neuroscience letters*, 450(1):60–64, 2009.

[9] O. van der Groen, E. van der Burg, C. Lunghi, and D. Alais. Touch influences visual perception with a tight orientation-tuning. *PloS one*, 8(11):e79558, 2013.