

Active Segmentation

Paul Fitzpatrick

The Problem: For a robot to manipulate its environment effectively, it needs to know which parts of the world move together, and which are more or less independent. Humans can judge physical coherence quite reliably from purely visual information, but that seems challenging for our machines. This work develops active strategies for segmenting objects from the background by trying to set them in motion, which makes the detection of object boundaries more reliable [6].

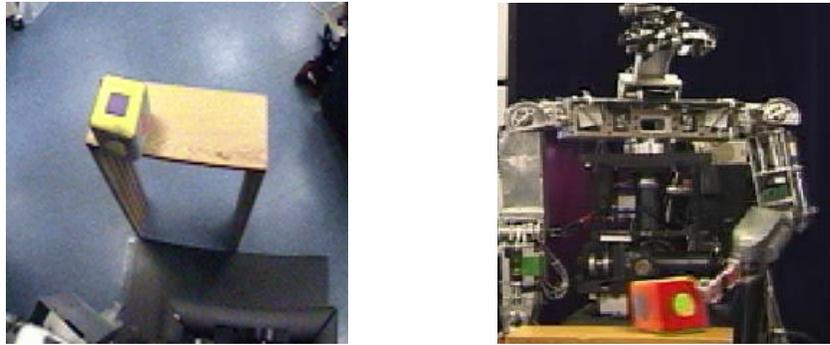


Figure 1: A motivating example. The image on the left has a number of features that could confuse a conventional visual segmentation algorithm. The edges of the table and cube in the image on the left happen to be aligned. The colors of the cube and table are not well separated. The cube has a strong pattern on its surface. Our robot, Cog, can prod and poke around to resolve such ambiguity (right).

Motivation: Much of computer vision is passive in nature, with the emphasis on watching the world but not participating in it. There are advantages to moving beyond this to exploit dynamic regularities of the environment [1]. A robot has the potential to examine its world using causality, by performing probing actions and learning from the response. The ability to perform “controlled experiments”, such as tapping an object lightly, offers a simple way to get to grips with an otherwise complex and uncertain world.

Previous Work: This work is being developed within the larger context of developing mirror neuron representations with Metta [4, 5] on the humanoid robot Cog [3].

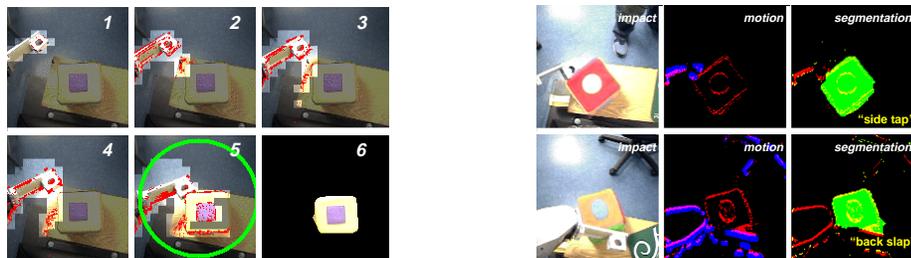


Figure 2: The moment of impact is detected visually (left) by the sudden expansion of motion away from the arm. Motion before and after contact is compared (right) to determine the boundary of the impacted object in the image.

Approach: The robot is equipped with a basic attention system that allows its attention to be directed to a target, where that target is initially crudely characterized – there is no knowledge of its spatial extent. The arm begins by extending outwards from the resting position towards the object. The end-effector is tracked as the arm sweeps rapidly outwards. The arm is driven to make a sweep of the area

around the target. The instant any collision occurs, it is detected by the motion it causes in the object. Optic flow preceding the collision is assumed to belong to the arm and its shadow; optic flow after the collision serves to identify the extent of the object. A maximum-flow algorithm due to [2] is used to fill in low-texture regions where optic flow is not generated.

Impact: Active segmentation allows our robot to collect reliable information about the appearance of objects in a semi-autonomous manner, requiring only the casual assistance of a “care-giver” willing to bring interesting objects into the robot’s reach. Models are constructed from this data that permit the robot to locate and identify objects when they reappear. This basic competence is fundamental to building robots that can operate in initially unfamiliar environments.

Future Work: Since segmentation is performed on images, it is 2D in nature. If the robot could knock or turn objects over, it could build more complete models of their 3D appearance. The robot can also learn about the appearance of its own arm, and potentially the human arm when viewed performing similar poking operations on objects the robot is familiar with.

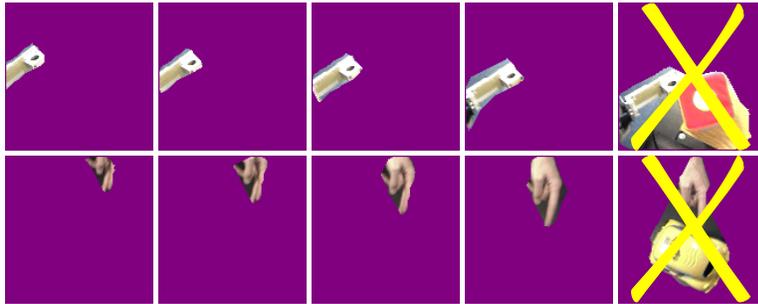


Figure 3: Early experiments on segmenting the robot arm, or a human hand poking an object the robot is familiar with, by working backwards from a collision event.

Research Support: Funds for this project were provided by DARPA as part of the “Natural Tasking of Robots Based on Human Interaction Cues” project under contract number DABT 63-00-C-10102, and by the Nippon Telegraph and Telephone Corporation as part of the NTT/MIT Collaboration Agreement.

References:

- [1] D. H. Ballard. Animate vision. *Artificial Intelligence*, 48(1):57–86, 1991.
- [2] Y. Boykov and V. Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. In *Energy Minimization Methods in Computer Vision and Pattern Recognition*, pages 359–374, 2001.
- [3] R. A. Brooks, C. Breazeal, M. Marjanovic, and B. Scassellati. The Cog project: Building a humanoid robot. *Lecture Notes in Computer Science*, 1562:52–87, 1999.
- [4] P. Fitzpatrick and G. Metta. Towards manipulation-driven vision. In *2002 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, EPFL, Switzerland, October 2002.
- [5] G. Metta and P. Fitzpatrick. Development and imitation in a humanoid robot. In *MIT Artificial Intelligence Laboratory Research Abstracts*.
- [6] M. G. Ross. Exploiting texture-motion duality in optical flow and image segmentation. Master’s thesis, Massachusetts Institute of Technology, August 2000.