Robots with Vision that Find Objects

Yiannis Aloimonos and Cornelia Fermüller
Institute for Advanced Computer Studies, University of Maryland, College Park, MD
Cognitive Robotics Group

Project: CPS (CNS 1035542)

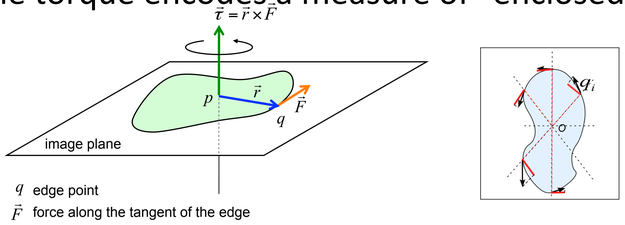
Abstract

A robot is instructed in language to find an object (from a generic class) in a cluttered room. For example, the robot may be asked to find an "apple" or a "cup". We take a bio-inspired, active approach to this problem that combines vision, action, and higher-level cognition. It consists of three modules: an attention mechanism that finds interesting parts of the scene, a segmentation mechanism that separates foreground regions from background, and a recognition mechanism based on shape descriptors of contours and surfaces and the relationship object affordances and their shape.

This year we developed a new mid-level vision mechanism, called the Torque Operator [1], which captures the concept of closed contours. It takes as input edges and computes over regions of different size a measure of how well the edges are aligned to form a closed, convex contour. First, the torque was implemented and tested as a generic mechanism for visual attention and segmentation. Second, high level knowledge about object properties was incorporated into the torque: Specifically, size and global shape, acquired through learning, were utilized in the attention process, and features of object contour were used to modify the torque mechanism to bias attention and segmentation for specific object classes.

Definition of Torque

The torque encodes a measure of "enclosedness"



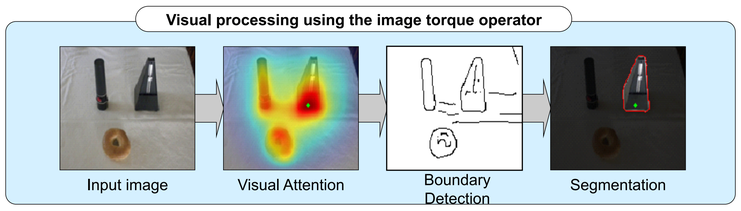
At a point q_i

$$\tau(q) = \vec{r} \times \vec{F}$$

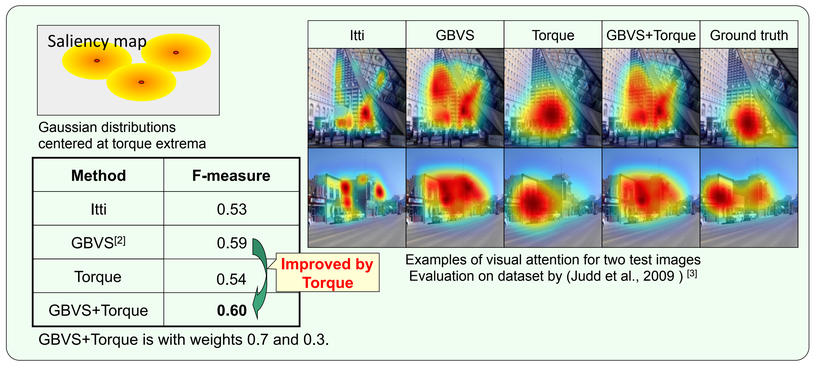
In a patch P

$$\tau(P) = \frac{\sum_{all\ points\ q_i} value\ of\ \tau(q_i)}{2 \cdot area(P)}$$

Application in Bottom-up Processing

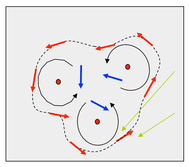


Attention



Method	F-measure
Itti	0.53
GBVS[2]	0.59
Torque	0.54
GBVS+Torque	**0.60**

Boundary Detection

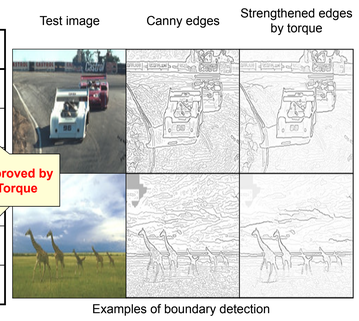


Averages over 123 test images

Method	F-measure
Canny	0.20
Torque (Canny)	**0.23**
Pb [5]	0.24
Torque (pb)	**0.26**
gPb [6]	0.29
Torque (gPb)	**0.29**

Improved by Torque

Evaluation on Caltech 101 dataset



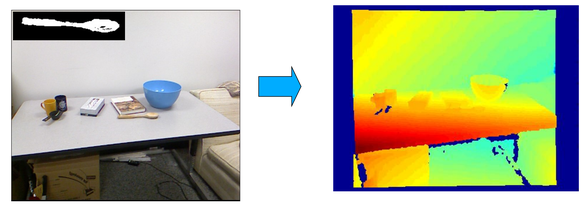
Introducing Higher Level Knowledge

Since it is not feasible for the robot to search everywhere, we need to reduce the search space by utilizing *knowledge* from language:

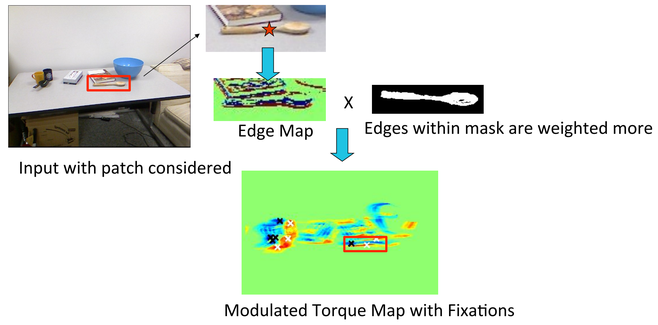
- *Contextual* – e.g. shoes are on floor, keys are on table
- *Properties (attributes)* – size, shape, appearance are known

Choosing Appropriate Patch Size

Using a calibrated camera and known object size, patches are computed a priori

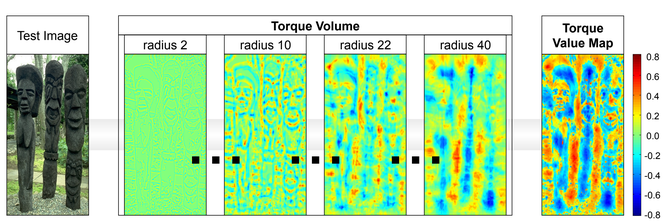


Knowing which edges *belongs* to an object, we weigh edges differently in the *torque measure:*



Using the torque

Torque is computed at every pixel and for multiple patch sizes.
Torque Volume: set of all image torque values
Scale map: map of scales with max. torque over patch sizes
Value map: map of max. torque values over patch sizes



Properties of the torque:
Torque is large where **edges surround the center of the patch** and **scale matches the size of patch**.

Segmentation

Fixation based segmentation

Graph-cut segmentation method defined on edges in polar coordinate system with the origin at the fixation point [7].

Experiment

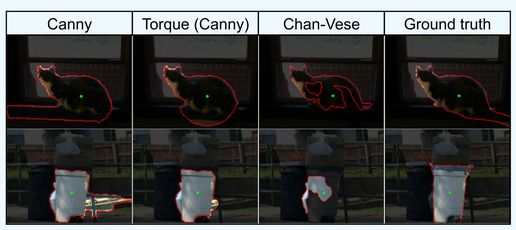
Comparison of fixation based segmentation using different ways of computing edges.

Average over 28 test images

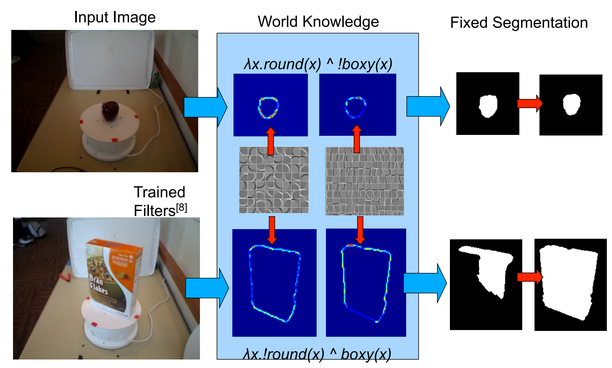
Visual Cue	Covering
Canny	0.32
Torque (Canny)	**0.47**
pb	0.46
Torque (pb)	**0.48**
Chan-Vese	0.21

Improved by Torque

Evaluation on dataset by (Stein et al., 2008)[4]



Introducing Knowledge about the Contour Shape



Reference

[1] M. Nishigaki, C. Fermüller and D. DeMenthon .The Image Torque Operator: A New Tool for Mid-level Vision, CVPR, 2012.
[2] J. Harel, C. Koch, and P. Perona. Graph-based visual saliency. In NIPS, 2006.
[3] T. Judd, K. Ehinger, F. Durand, and A. Torralba. Learning to predict where humans look. In ICCV, 2009.
4] A. N. Stein, T. S. Stepleton, and M. Hebert. Towards unsupervised whole-object segmentation: Combining automated matting with boundary detection. In CVPR, 2008.
[5] D. R. Martin, C. C. Fowlkes, and J. Malik. Learning to detect natural image boundaries using local brightness, color, and texture cues. PAMI, 26(5):530–549, 2004.
[6] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik. Contour detection and hierarchical image segmentation. PAMI, 33(5):898–916, 2011.
[7] A. Mishra, C. Fermüller, and Aloimonos. Active segmentation for robotics. In IROS, 2009.
[8] K. Gregor and Y. LeCun .Structured Sparse coding via lateral inhibition. NIPS, 2011.