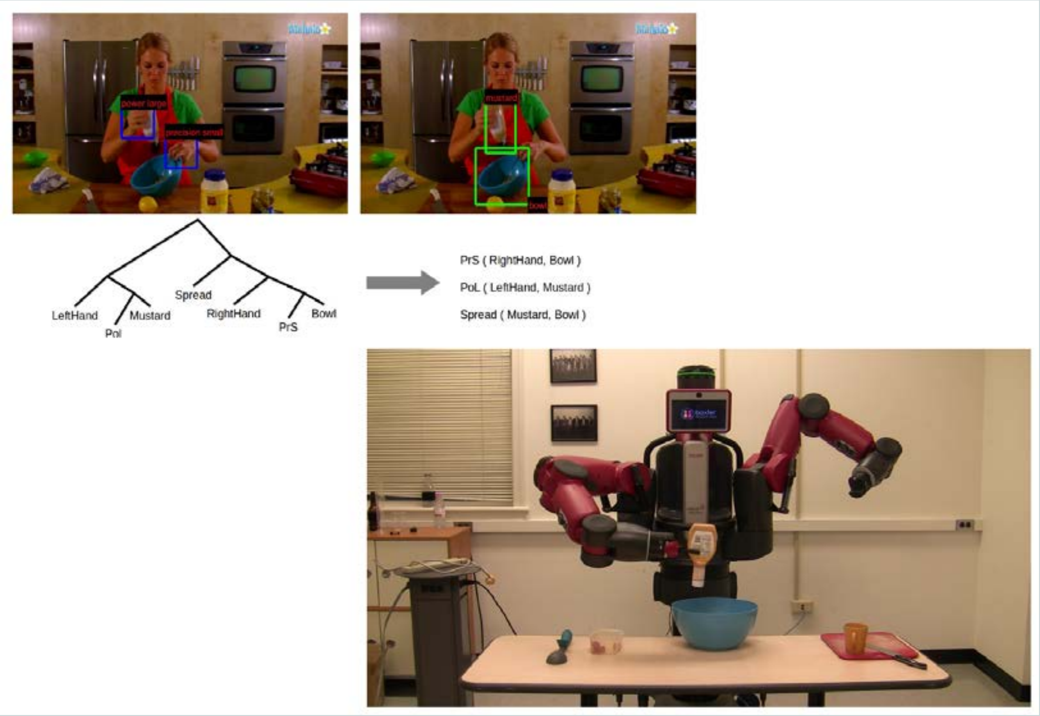
What Happened to the Representations of Perception ?

Cornelia Fermüller
Computer Vision Laboratory
University of Maryland

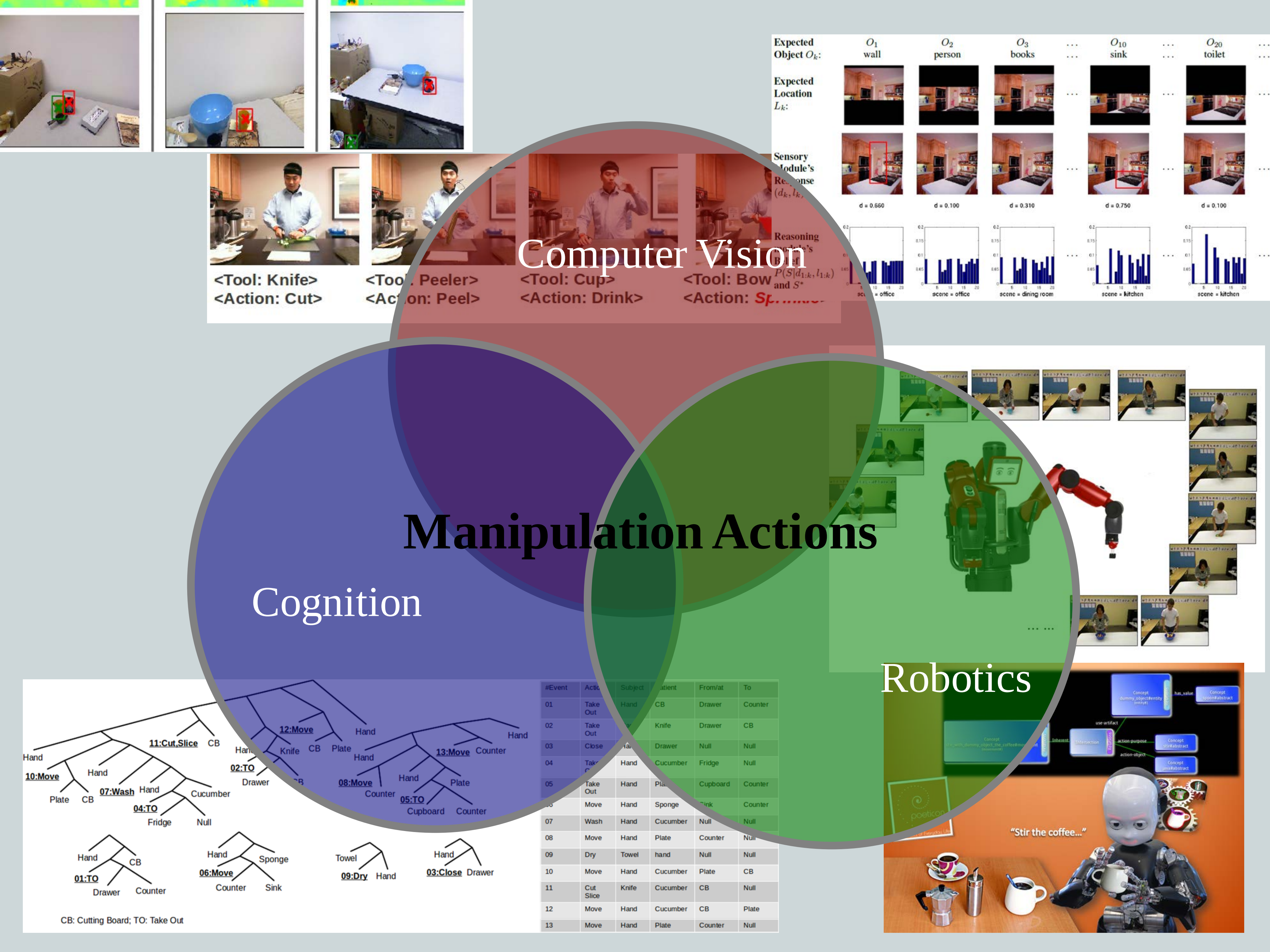
Why we study manipulation actions?

2. Learning from humans to teach robots



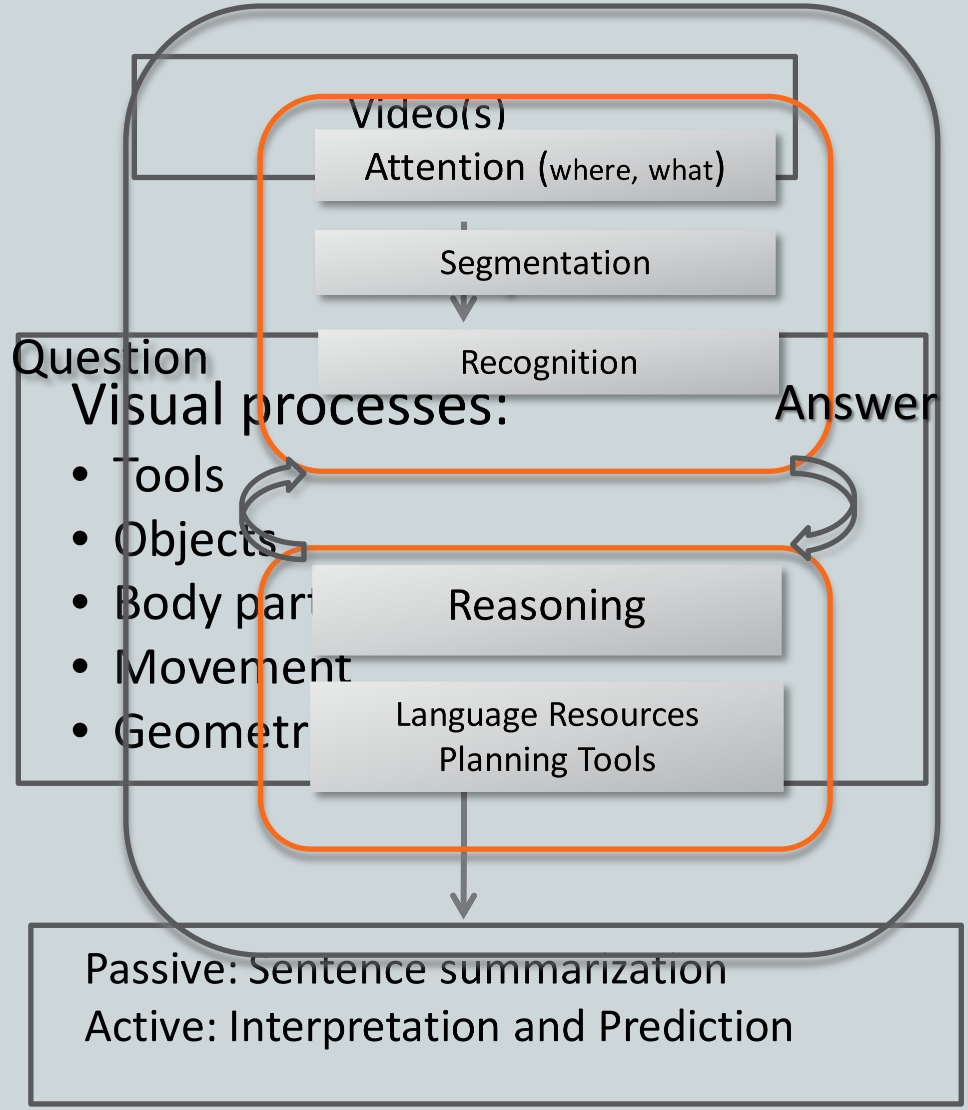
Y Yang, Y Li, C Fermüller, Y Aloimonos. *Robots Learning Manipulation Actions by "Watching" Unconstrained Videos from the World Wide Web*, AAAI 2015.

The robot learns to mix a drink



Computer Vision
Manipulation Actions
Cognition
Robotics
<Tool: Knife>
<Action: Cut>
<Tool: Cup>
<Action: Drink>
Expected Object O_k:
Expected Location L_k:
wall
person
books
sink
toilet
"Stir the coffee..."
CB: Cutting Board; TO: Take Out
11:Cut,Slice
12:Move
13:Move
10:Move
02:TO
07:Wash
04:TO
08:Move
05:TO
01:TO
06:Move
09:Dry
03:Close

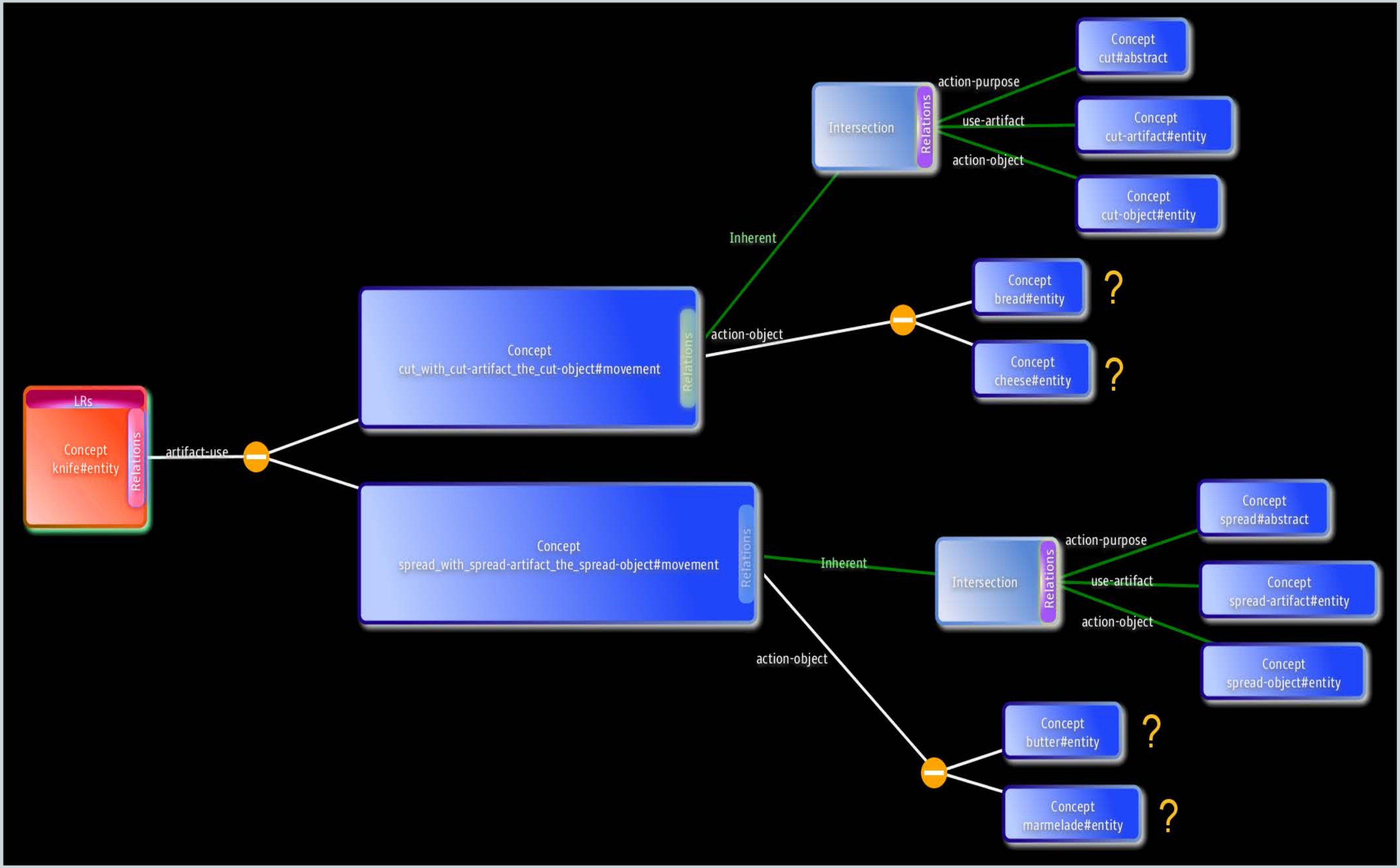
The approach: A dialogue



Hand segmentation and tool detection

Hand segmentation
Object next to hand

<Tool: Knife>



EU Project Poeticon: http://www.poeticon.eu

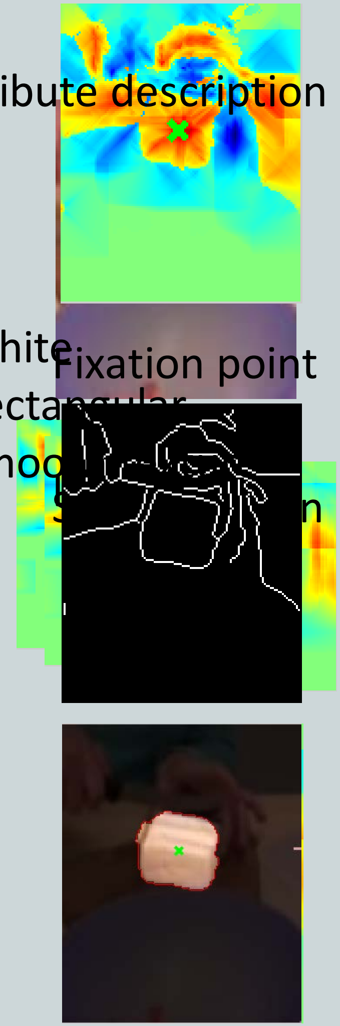
Object recognition

Attention operator

Attribute description

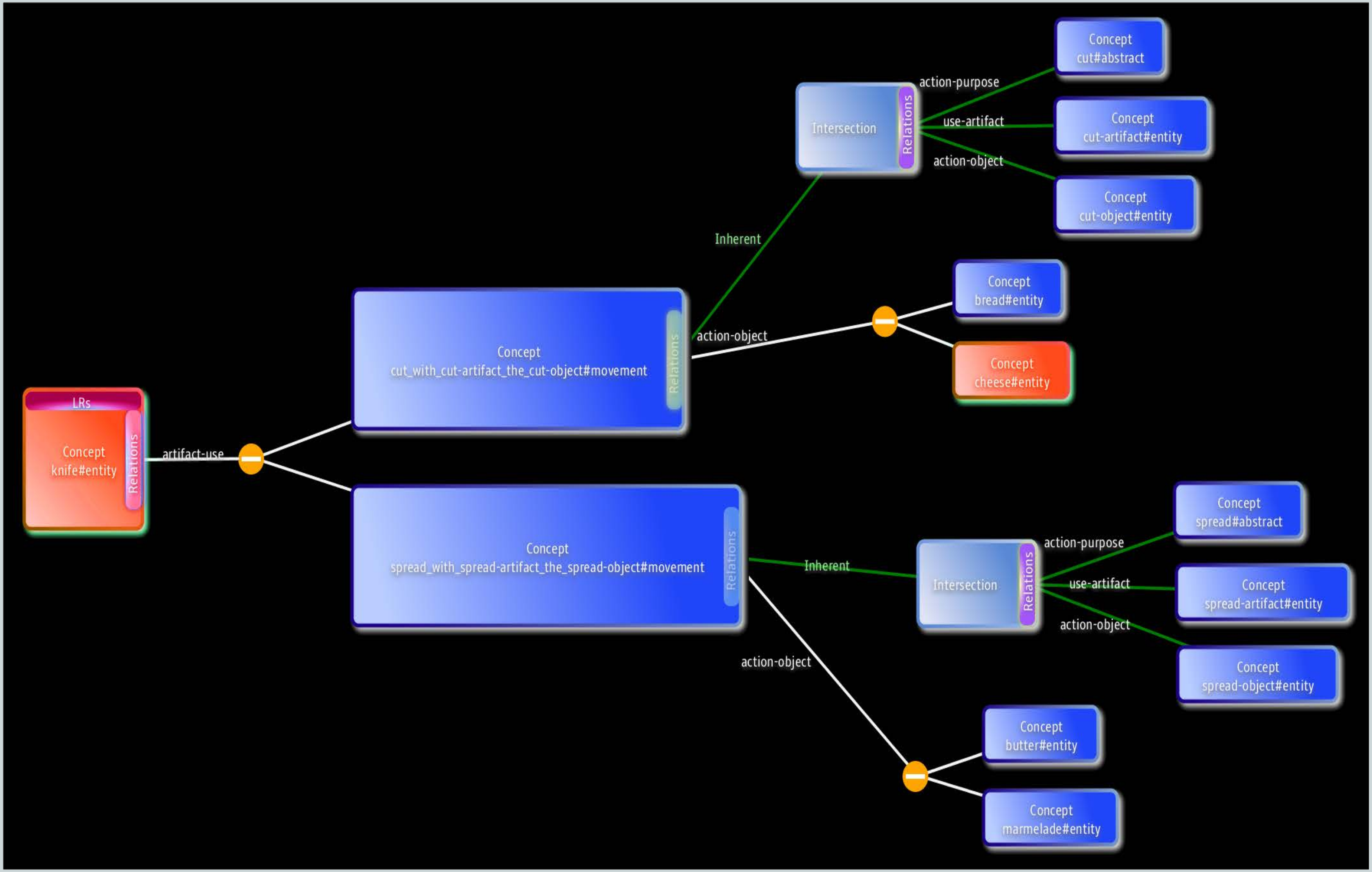
White
Rectangular
Smooth

Fixation point



Segmentation

Torque map

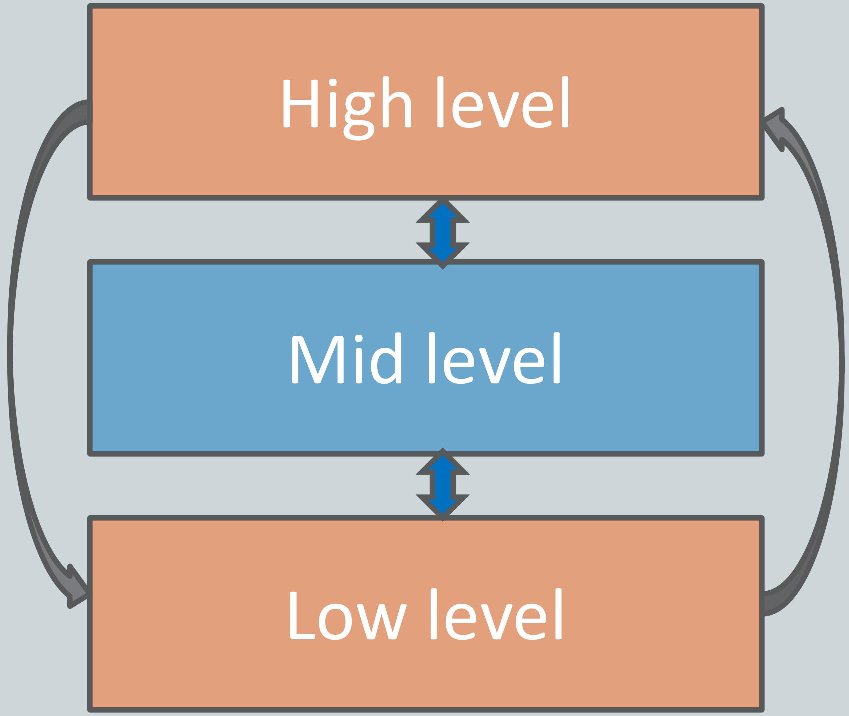


LRs
Concept
knife#entity
Relations
artifact-use
Concept
cut_with_cut-artifact_the_cut-object#movement
Relations
Inherent
Intersection
Relations
action-purpose
use-artifact
action-object
Concept
cut#abstract
Concept
cut-artifact#entity
Concept
cut-object#entity
action-object
Concept
bread#entity
Concept
cheese#entity
Concept
spread_with_spread-artifact_the_spread-object#movement
Relations
Inherent
Intersection
Relations
action-purpose
use-artifact
action-object
Concept
spread#abstract
Concept
spread-artifact#entity
Concept
spread-object#entity
action-object
Concept
butter#entity
Concept
marmelade#entity

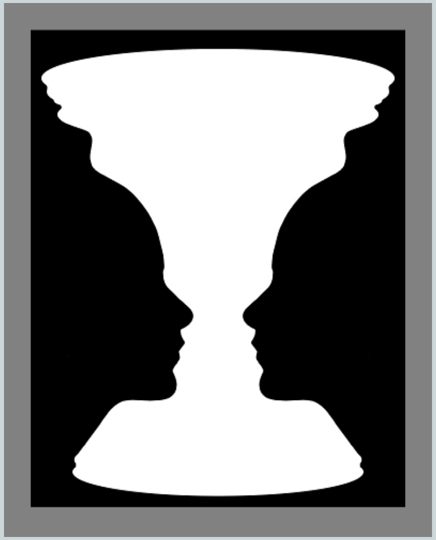
Cheese

Vision Processes

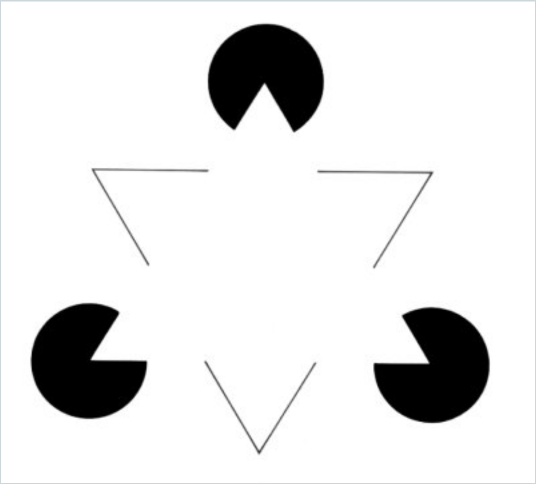
- Mid-level process (bottom-up and top-down) for object recognition



Visual illusions demonstrating Gestalt principles

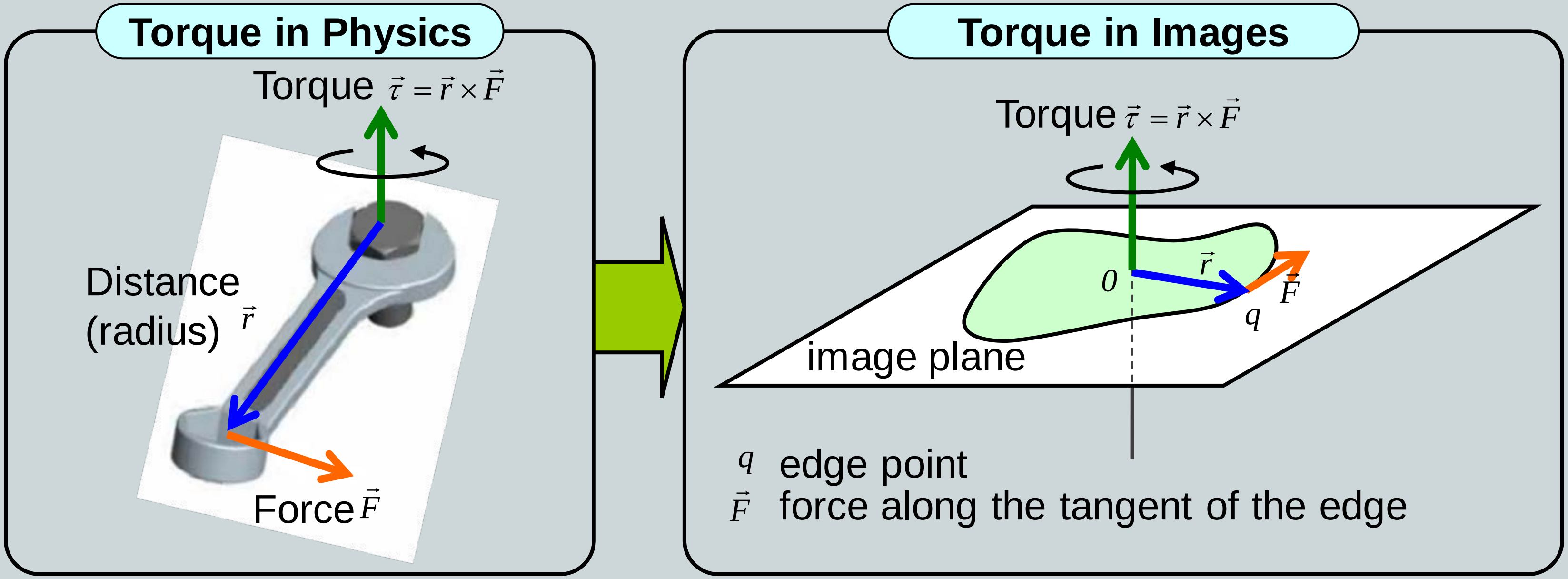


Rubin, 1915



Kanizsa, 1976

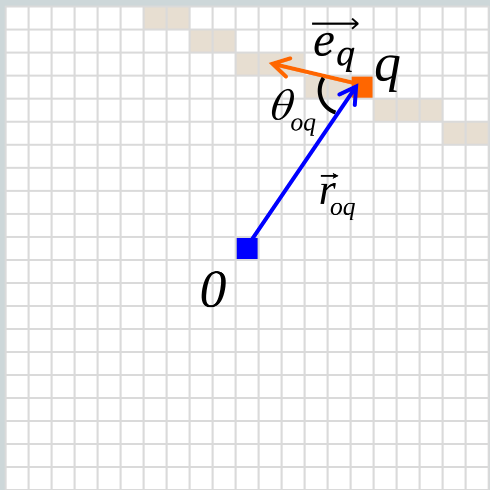
Torque in Images

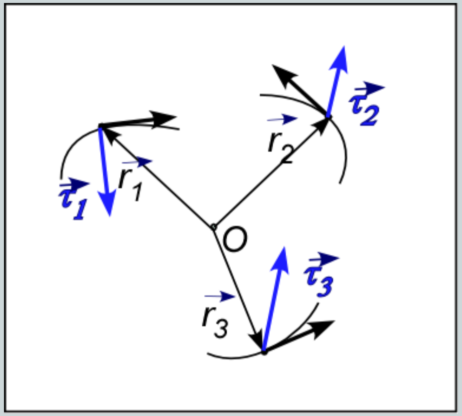


M. Nishigaki, C. Fermüller and D. DeMenthon: "The image torque operator: A new tool for mid-level vision," CVPR, 2012.

Definition of Torque in Images

Discrete Edge Points





Torque at point q: $\vec{\tau}_{oq} = \vec{r}_{oq} \times \vec{e}_q$

Torque of a patch

$$\tau_P = \frac{1}{2|P|} \sum_{q \in E(P)} \tau_{pq}$$

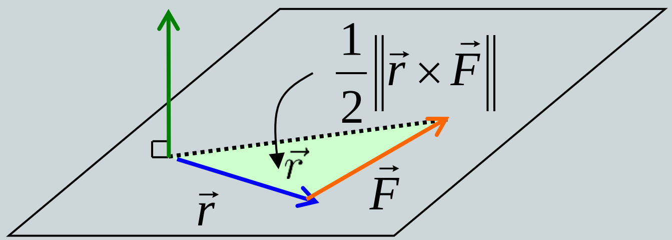
Value of the torque

$$\tau_{oq} = |\vec{r}_{oq}| \cdot |\vec{e}_q| \cdot \sin\theta_{oq}$$
$$= |\vec{r}_{oq}| \sin\theta_{oq}$$

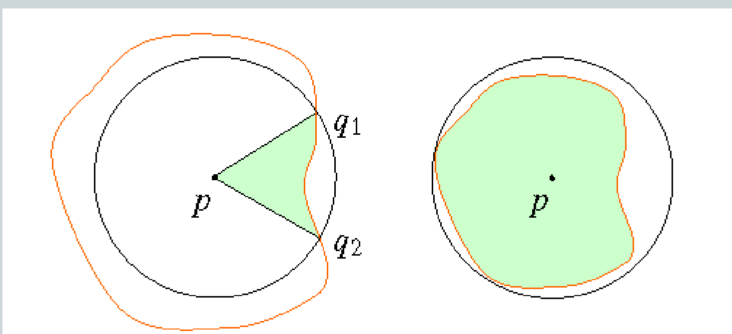
$|P|$ area of the image patch P

$E(P)$ set of edge points within the patch P

Torque of an Image Patch



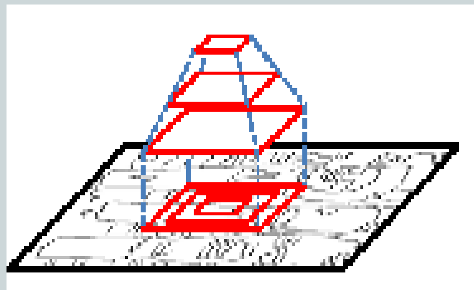
The triangle enclosed by vector $\vec{r}$ and $\vec{F}$ is equivalent to $\left\|\vec{r}\times\vec{F}\right\|/2$



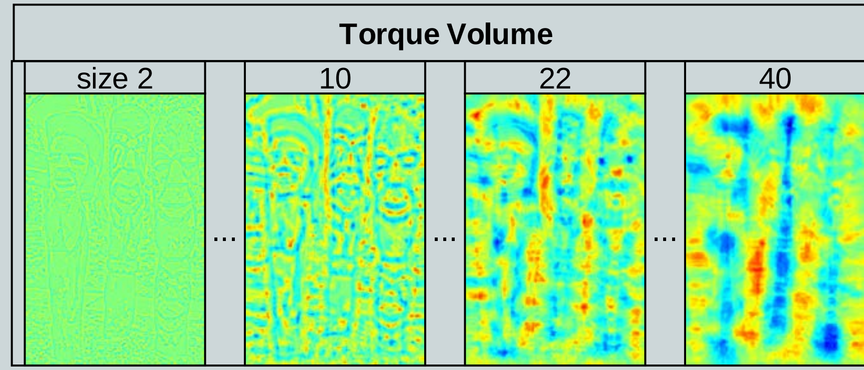
Torque of image patch is related to the area enclosed by a contour

*Disk or rectangle patches are used in our experiments

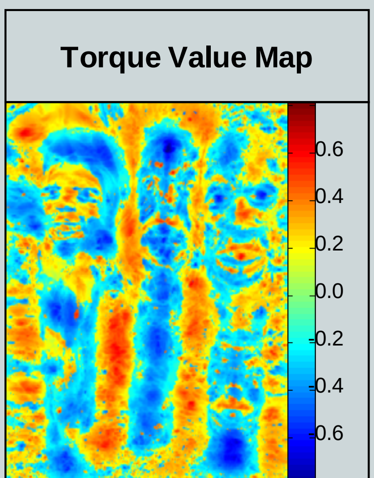
Using the Torque

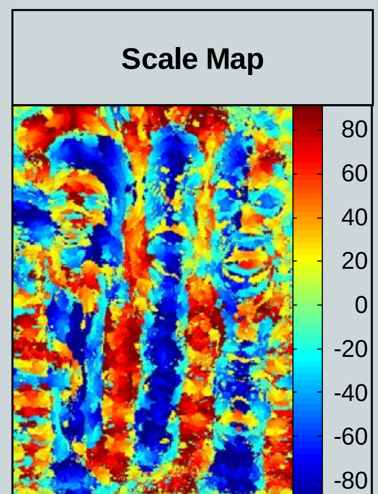


Combination of the torque values from all patch sizes



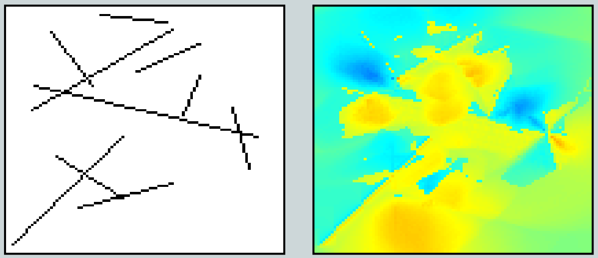
Torque values for different patch sizes





Key properties of the Torque

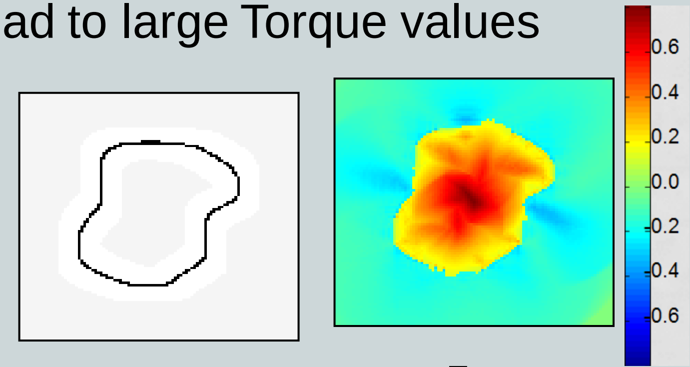
Texture edges
lead to small torque values



Edge Map

Torque Value Map

Closed contours
lead to large Torque values

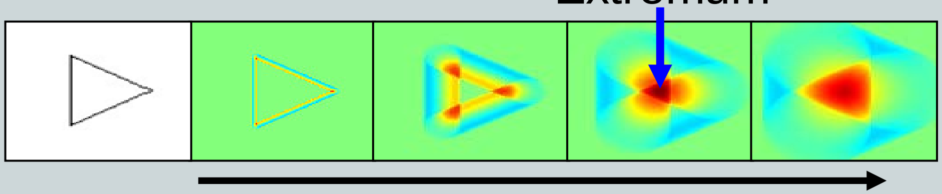


Edge Map

Torque Value Map

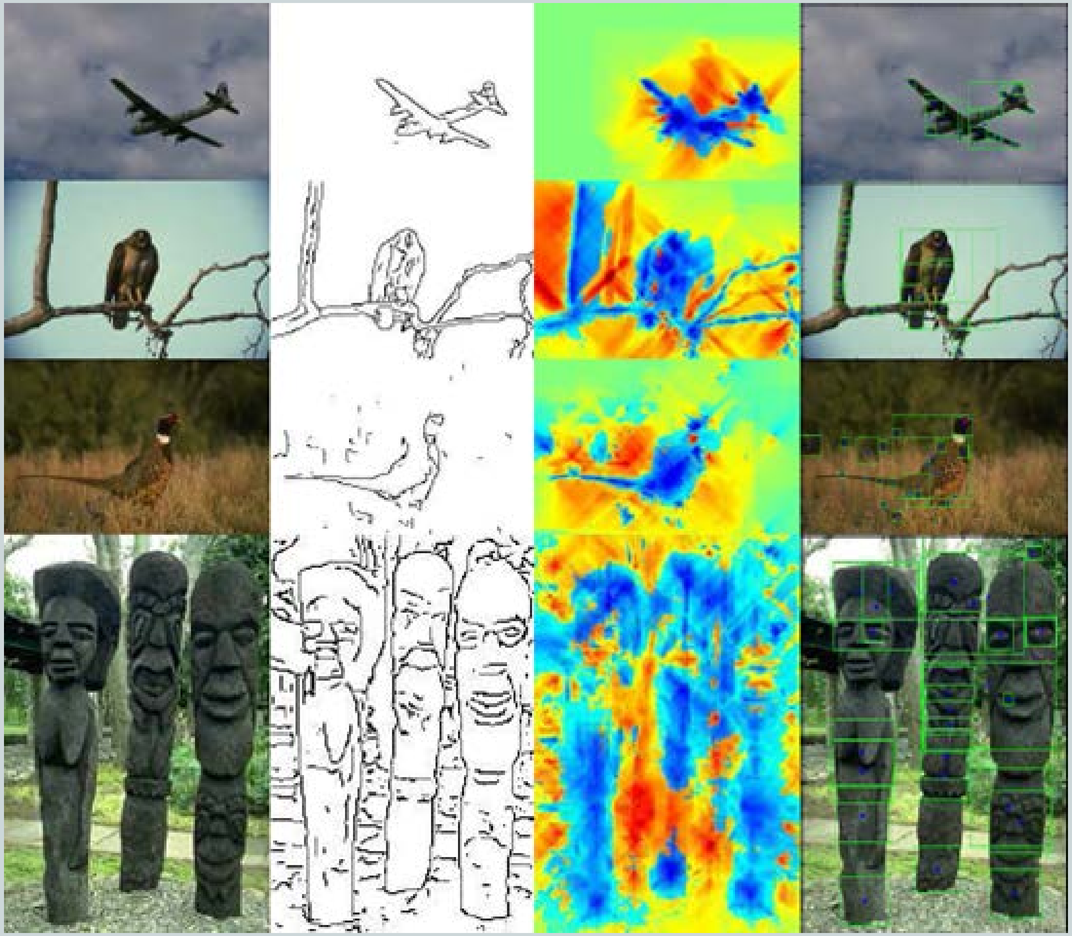
Extrema in Torque
tend to occur at centroids

Extremum



Torque maps for a triangle computed over four patch sizes

Torque Extrema

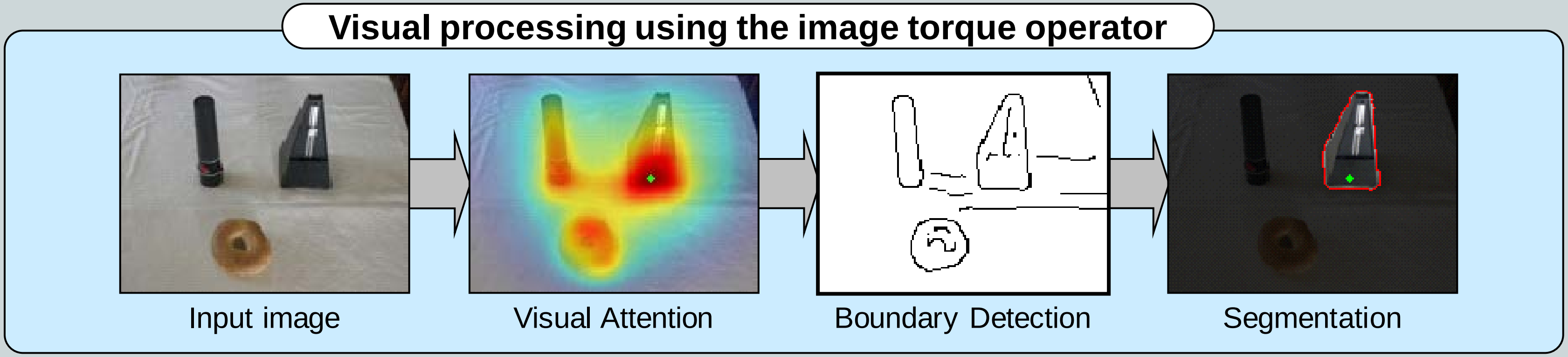


a b c d

a. Image
b. Pb edges
c. Torque value map
d. Minima in Torque volume

Application

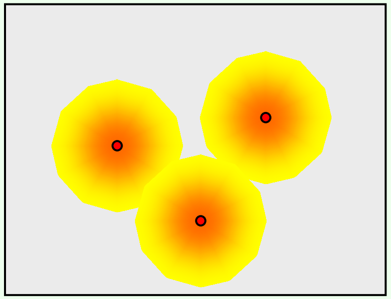
An active approach to finding an object in the scene consists of three modules: visual attention, boundary detection, and foreground segmentation.



We showed that by adding the torque we can improve state of the art methods

Visual Attention

saliency map using torque

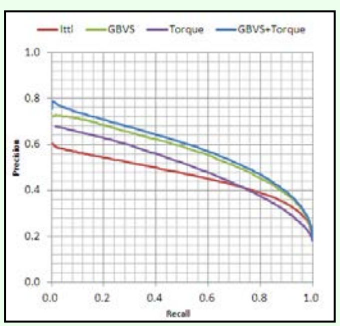


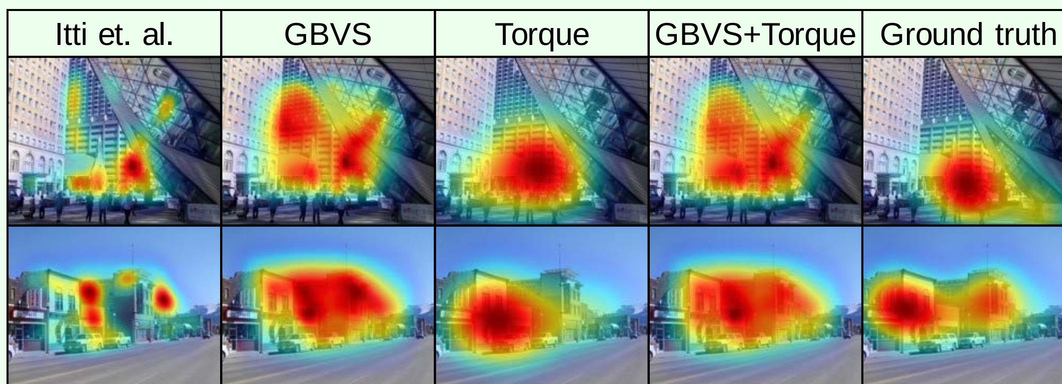
Gaussian distributions centered at torque extrema

Method	F-measure
Itti et. al.	0.53
GBVS(Harel et al.,2009)	0.59
Torque	0.54
GBVS+**Torque**	**0.60**

Improved by Torque

Evaluation on dataset by Judd et. al. (2009): F-measure and precision-recall curve.
GBVS+Torque is with weights 0.7 and 0.3.





Examples of visual attention for two test images

3D volumetric video segmentation

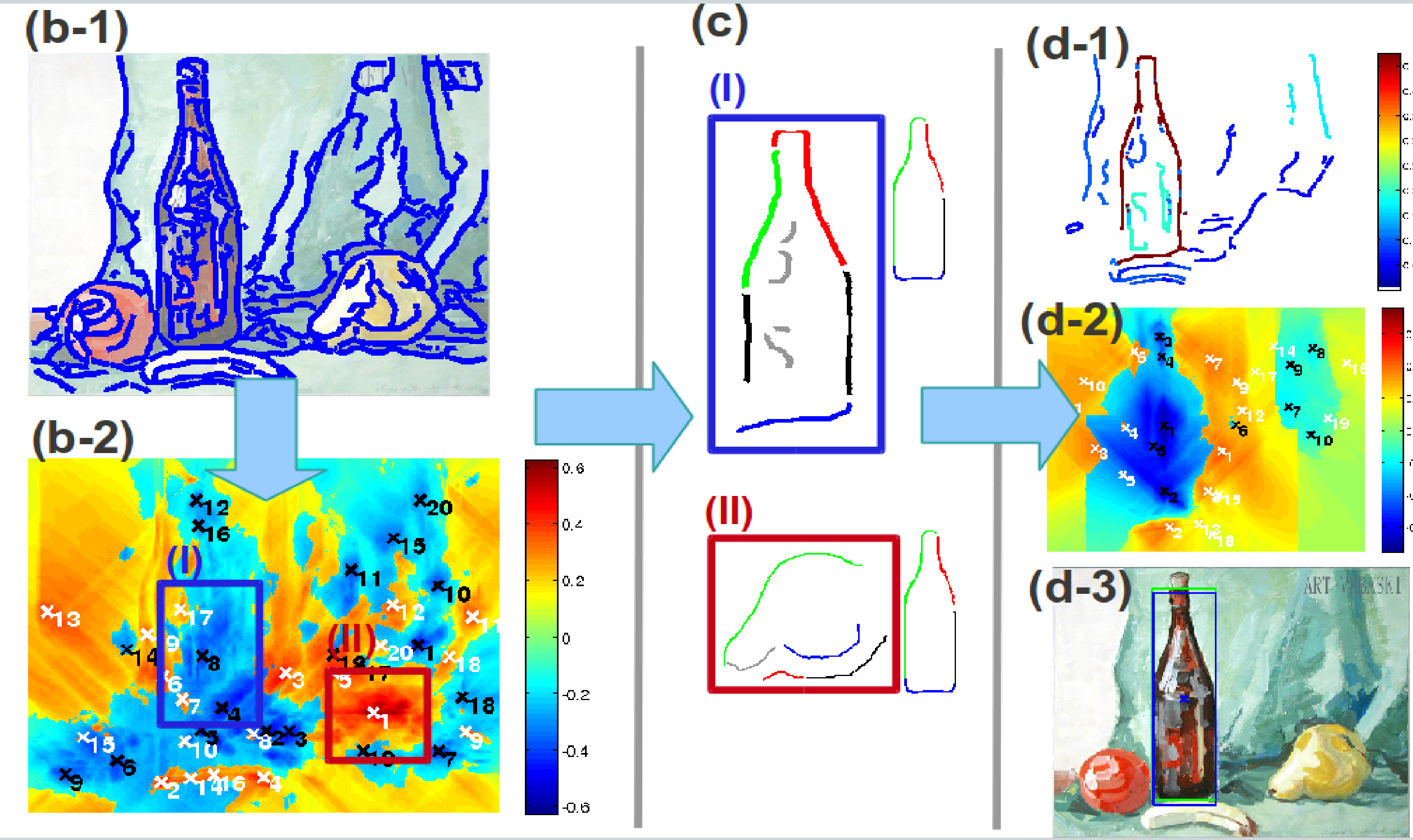
Detecting object specific contours

Feedback to Mid-level Vision

1. Training:. Obtain "prototypical contours" from annotated ground truth contours

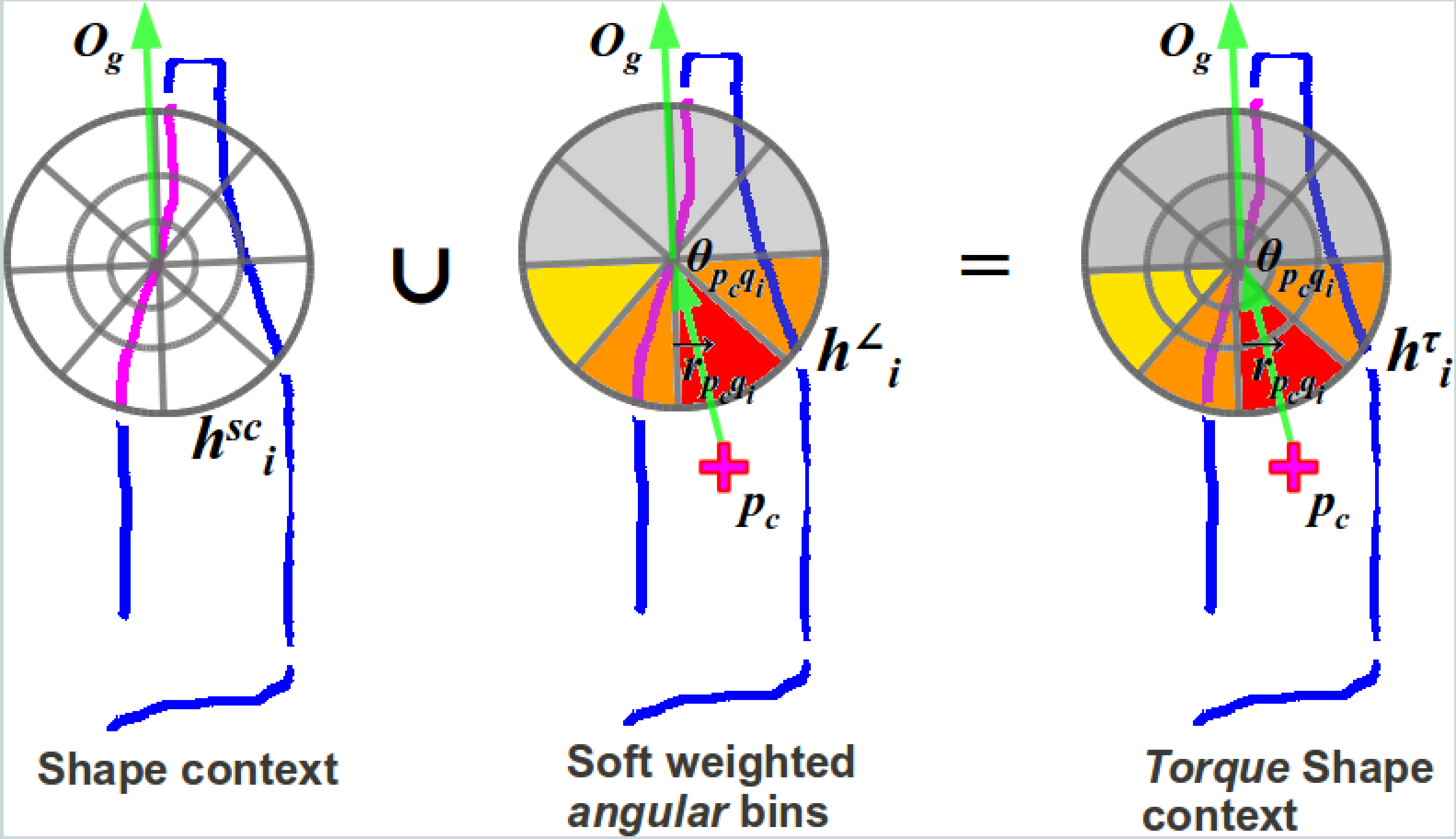
2. Run time.
   – **Match** partial contour fragments to models
   – **Reweigh** torque based on matching scores

- C. L Teo, C.Fermüller, Y. Aloimonos. "A Gestaltist approach to contour-based object recognition: Combining bottom-up and top-down cues," Intern.Journal of Robotics Research, 2015.
- M. Maynard, A. Guha (Y. Aloimonos, C. Fermüller) "Feedback from Vision," Qualcomm Innovation Fellowship Award 2016.



(b-1)
(b-2)
(I)
(II)
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
(c)
(I)
(II)
(d-1)
(d-2)
(d-3)

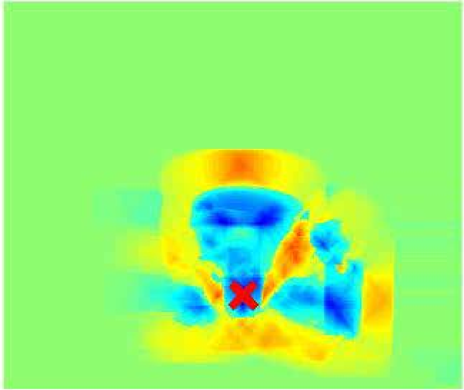
Partial Contour Matching: Torque Shape-Context



Example results on Robot

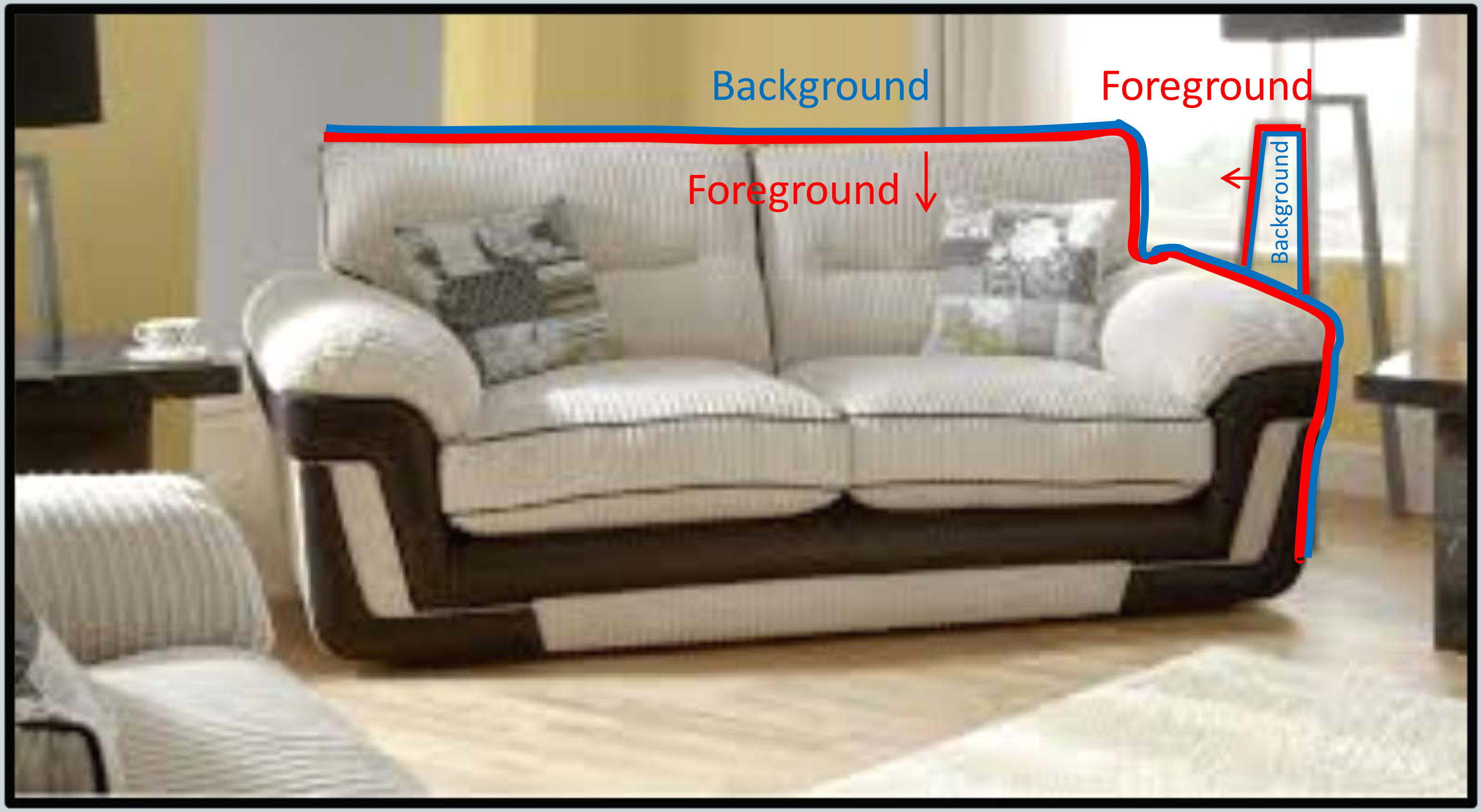
Edges

$\boldsymbol{\tau_P^m}$



RGB Input

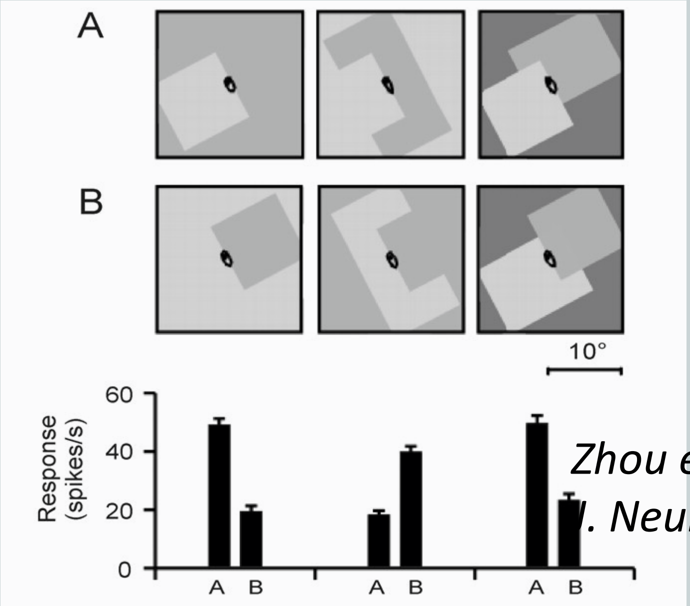
What is Border Ownership?



C.L Teo, C Fermüller, and Y. Aloimonos. "Fast 2D Border Ownership Assignment," CVPR 2015.

Motivations: Psychological & Biological

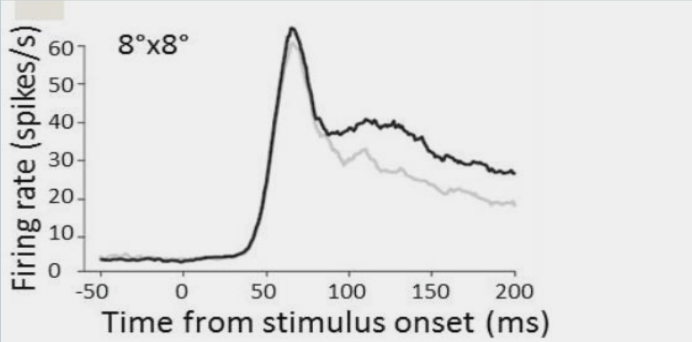
V2 cell that prefers bright Figure on the left



Selective nature of border ownership neurons (V2 & V4):

Zhou et al., J. Neuroscience 2000

Fast response time, <75ms from stimulus onset.

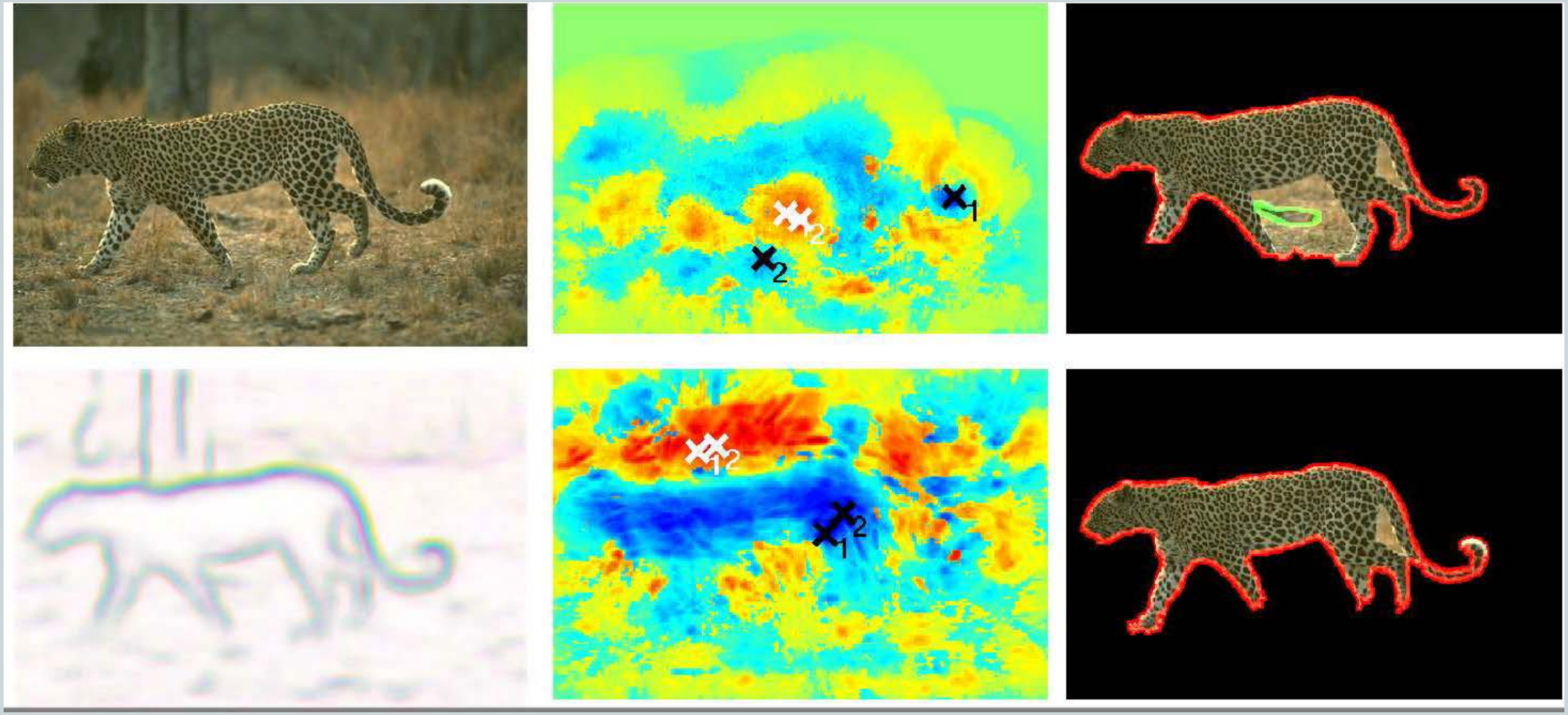


Sugihara et al., J. Neurophysiology 2011

Figure-ground organization and attention are closely related

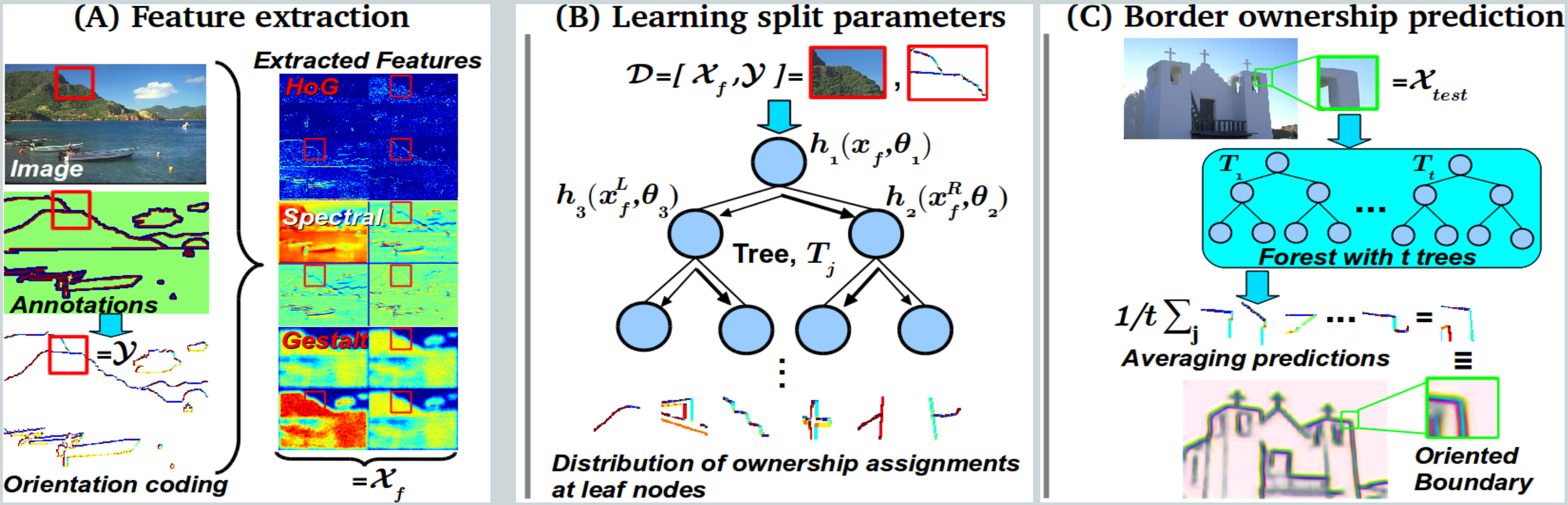
Craft et al, J. Neurophysiology 2007

Segmentation



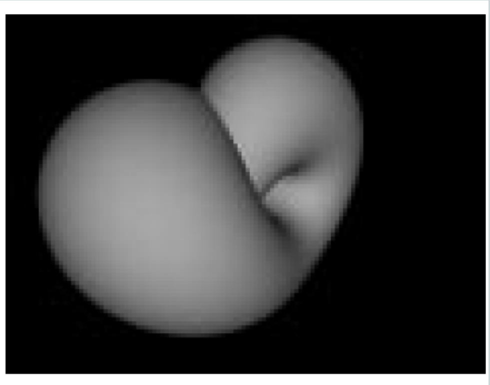
Approach Overview

1. Extract patch-based features sensitive to ownership
2. Train a Structured Random Forest (SRF) that saves ownership structure at leaf nodes
3. Fast inference using SRF by averaging responses over all decision trees



Feature Extraction: Local Ownership Cues

Extremal edges or ***image folds*** are characteristic changes in intensity along boundaries.

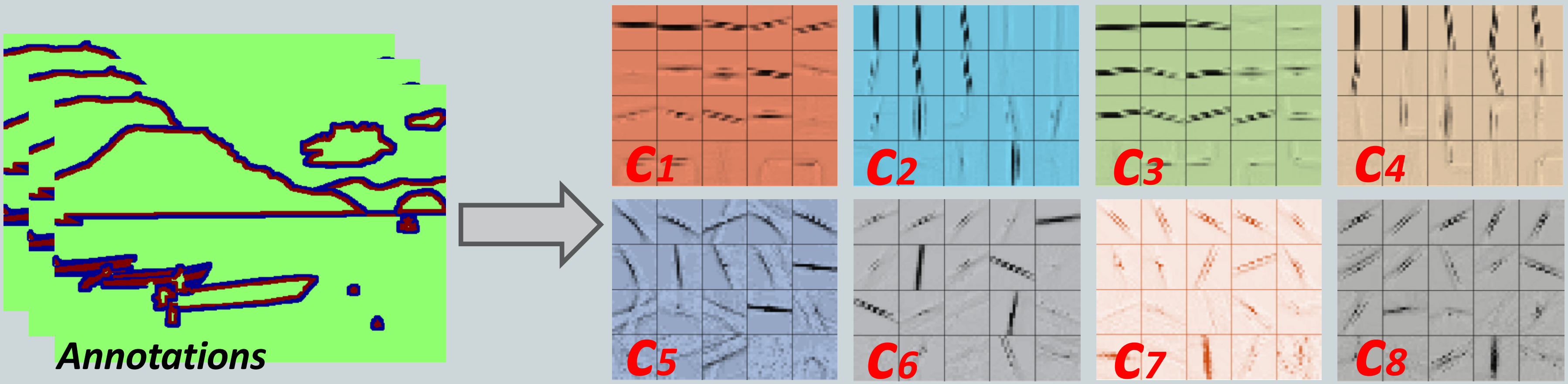



Huggins & Zucker, ICCV 2001

Psychophysical experiments have shown them to be one of the **strongest cues** for ownership.

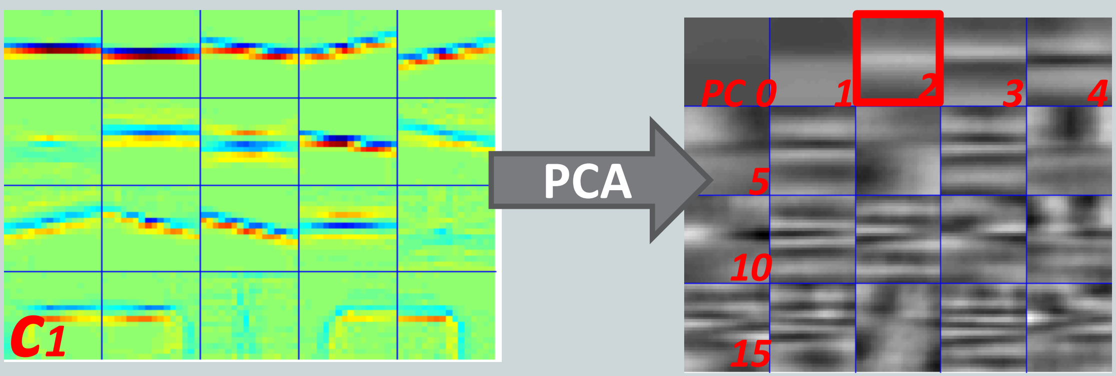
Ghose & Palmer, J. Vision 2010

Local Ownership Cues



"Sketch token" clusters of 8 ownership directions

Lim et al., CVPR'13



PC2 displays grayscale variations indicative of extremal edges.

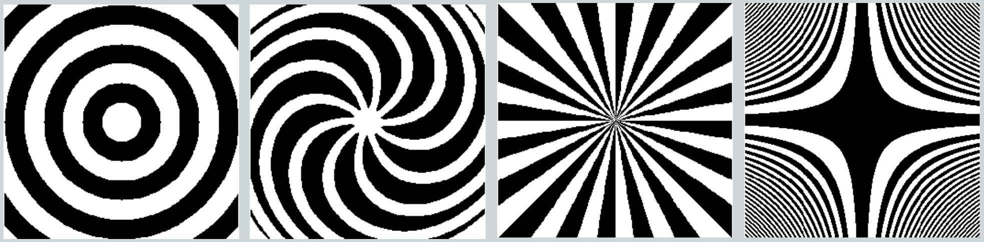
Ramenahalli et al., CISS'11

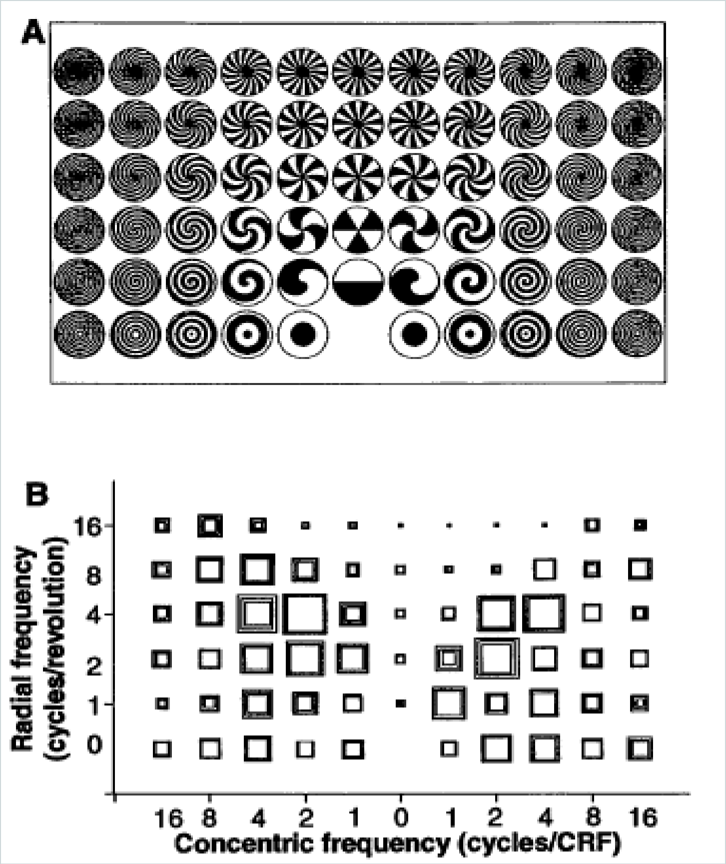
Feature Extraction: Global Ownership Cues

Border ownership is also determined by longer range (global) contextual cues.

Craft et al., J. Neurophysiology 2007

Implementation through visual operators that capture four grouping or “Gestalt” patterns:





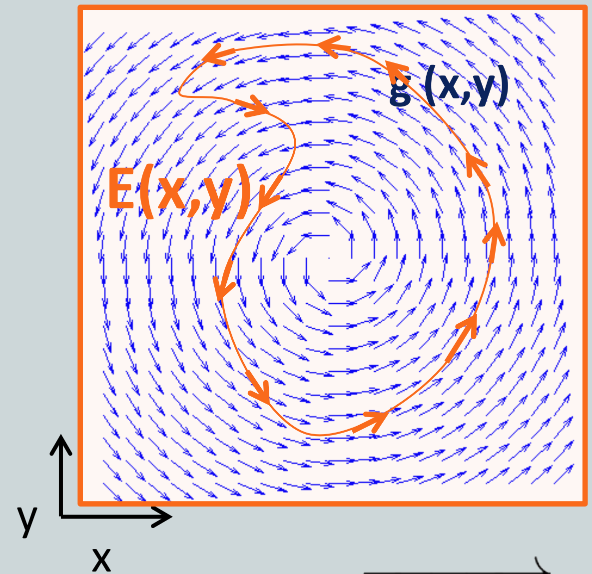
A: Gratings. B. Responses of a V4 cell

Cells tuned to these patterns have been observed area V4 of macaques :

Gallant et al., Science 1993

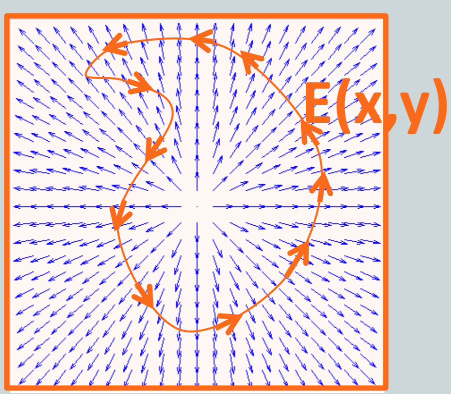
Global Ownership Cues

Rewriting the image torque as a *scalar product of* the edges (tangent vectors) *E(x,y)* and a circular gradient field g(x,y).



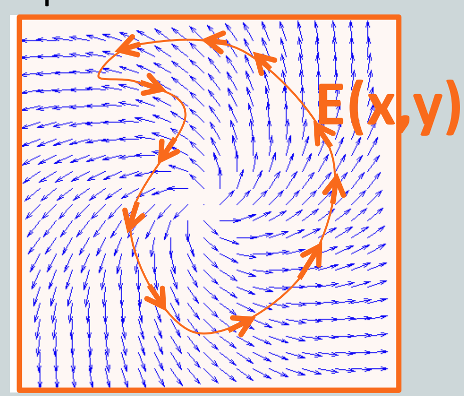
$$\tau(x,y) = \vec{E}(x,y) \times \vec{r} \quad \equiv \quad \tau(x,y) = \vec{E}(x,y) \cdot \overrightarrow{(-y,x)}$$

Radial



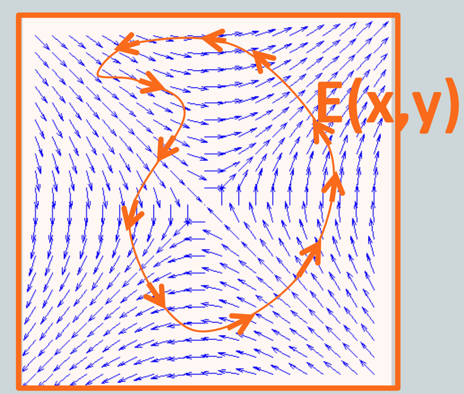
$g(x,y) = (x,y)$

Spiral



$g(x,y) = (ax - y, y + ax)$

Hyperbolic gradient field



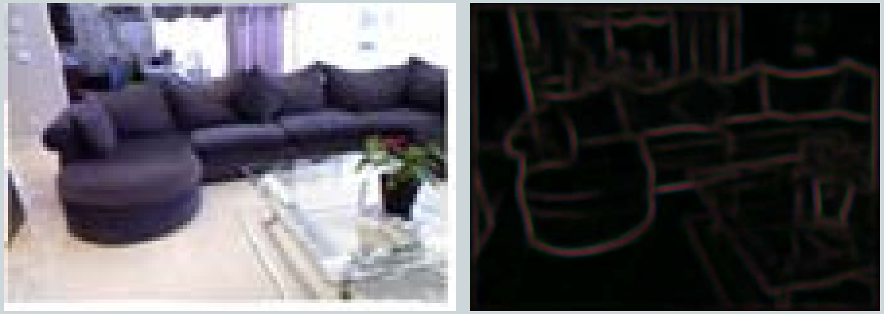
$g(x,y) = (ay, x)$

Results

Predicted boundaries (red) and ownership (FG: green, BG:blue)

BSDS (100 training/100 testing)

Martin et al., PAMI 2004



NYU-Depth (795 training/ 654 testing)

Silberman et al., ECCV 2012

Ownership prediction accuracy:

Ren et al., ECCV 2006
Leichter & Lindenbaum, ICCV 2009

Feature set	BSDS	NYU-Depth
HoG	72.0%	66.0%
+ Spectral (no contour tokens)	73.1% (72.0%)	67.0% (65.6%)
+ Spectral (contour tokens)	74.0% (72.3%)	68.1% (66.7%)
+ Gestalt patterns	74.4% (72.7%)	**68.4**% (66.7%)
All features + Spectral (NYU)	**74.7**% (72.8%)	-
Global-CRF	69.1%	-
2.1D-CRF	68.9%	-

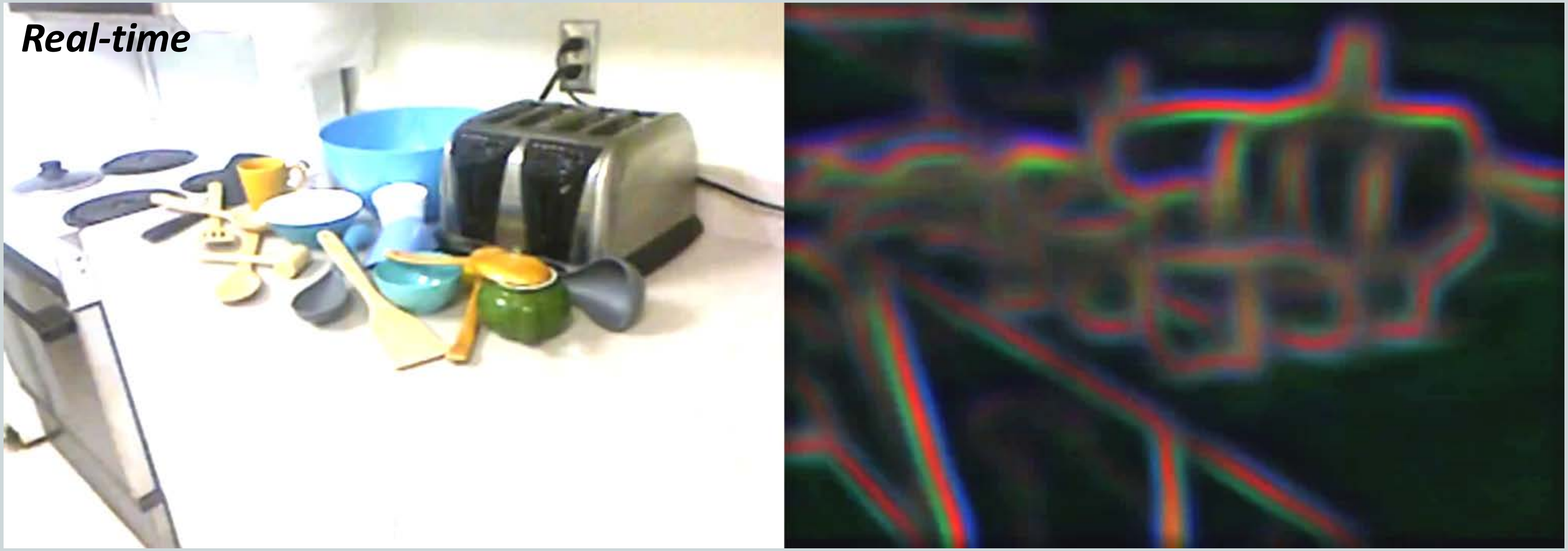
() denotes use of single features

Boundary prediction accuracy:

Arbelaez et al., PAMI 2011
Dollar et al., PAMI 2015

Method	BSDS-500	NYU-Depth
Our approach	0.73,0.74,0.76	0.63,0.64,0.60
gPb-owt-ucm	0.73,**0.76**,0.73	0.63,0.66,0.56
SE	0.73,0.75,**0.77** (SE-SS)	**0.65**,**0.67**,**0.65** (SE-RGB)

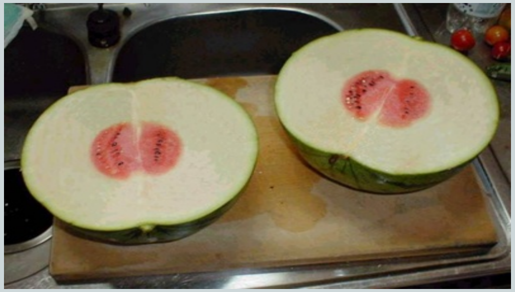
Results



Red: Boundaries, Green: Foreground, Blue: Background

Symmetry in 2D

- Goal: Detect symmetries in complex environments containing *clutter:*

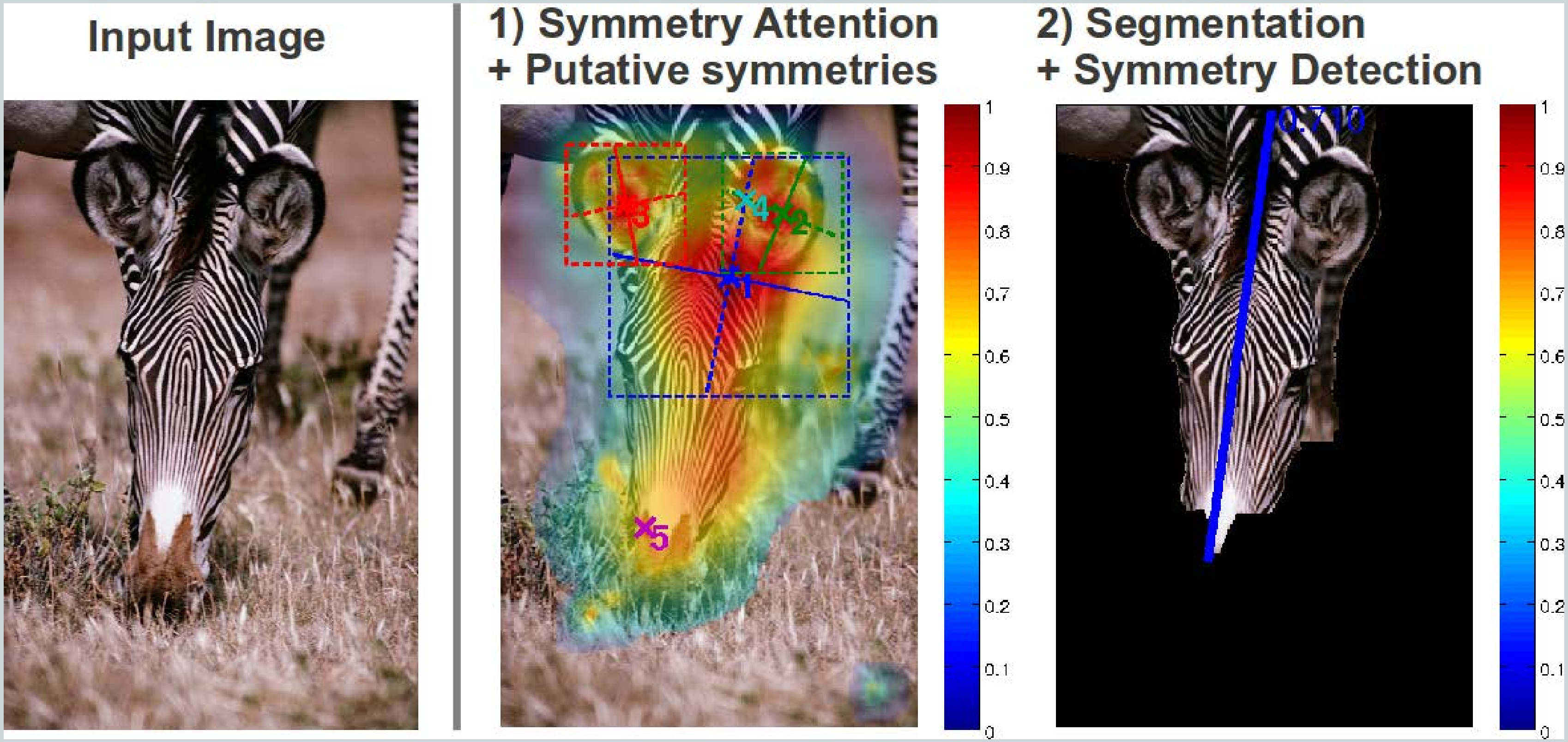


- Key challenges:
  - *Where* to compute the symmetries? → **Attention**
  - *How* to compute the symmetries reliably? → **Statistics**

C.L. Teo and C. Fermüller. “Object-Centric Bilateral Symmetry Detection,” under review.

Proposed Solution

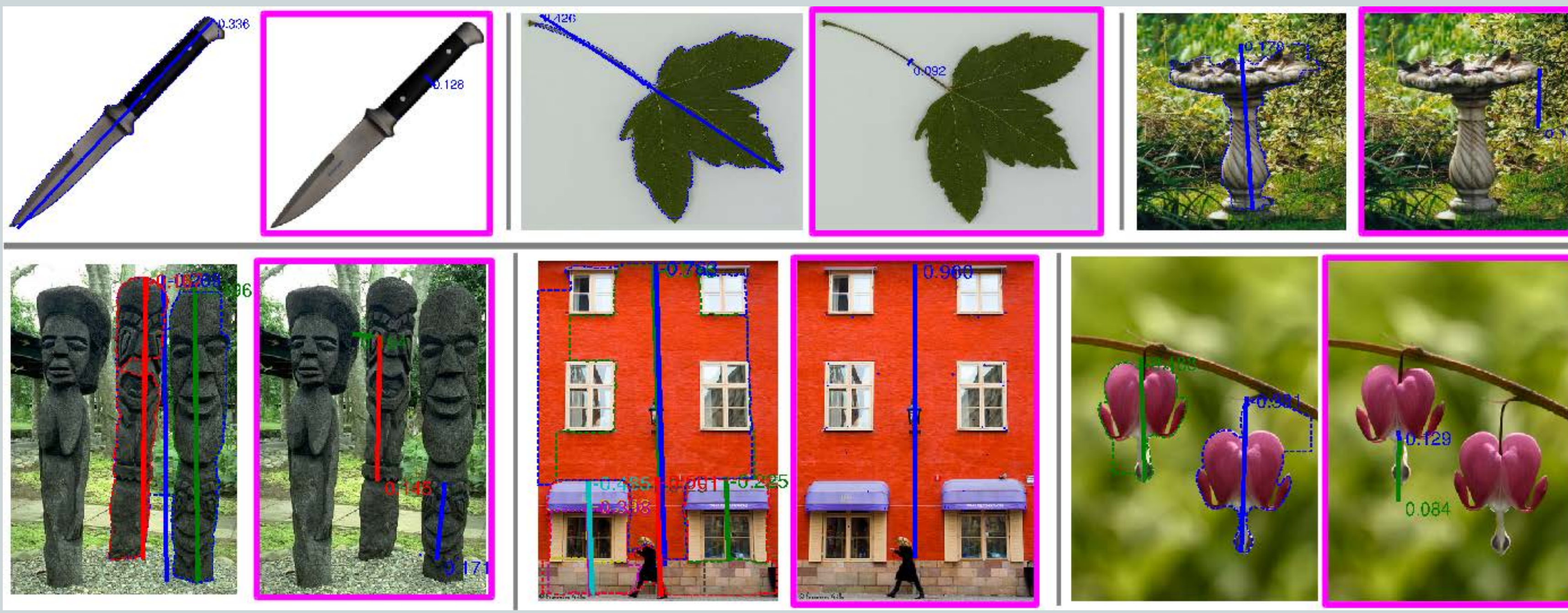
(2 steps):



Segmentation applied per fixation point: resolves the scale issue since the object region is selected

Results

Top: singles



Bottom: multiples

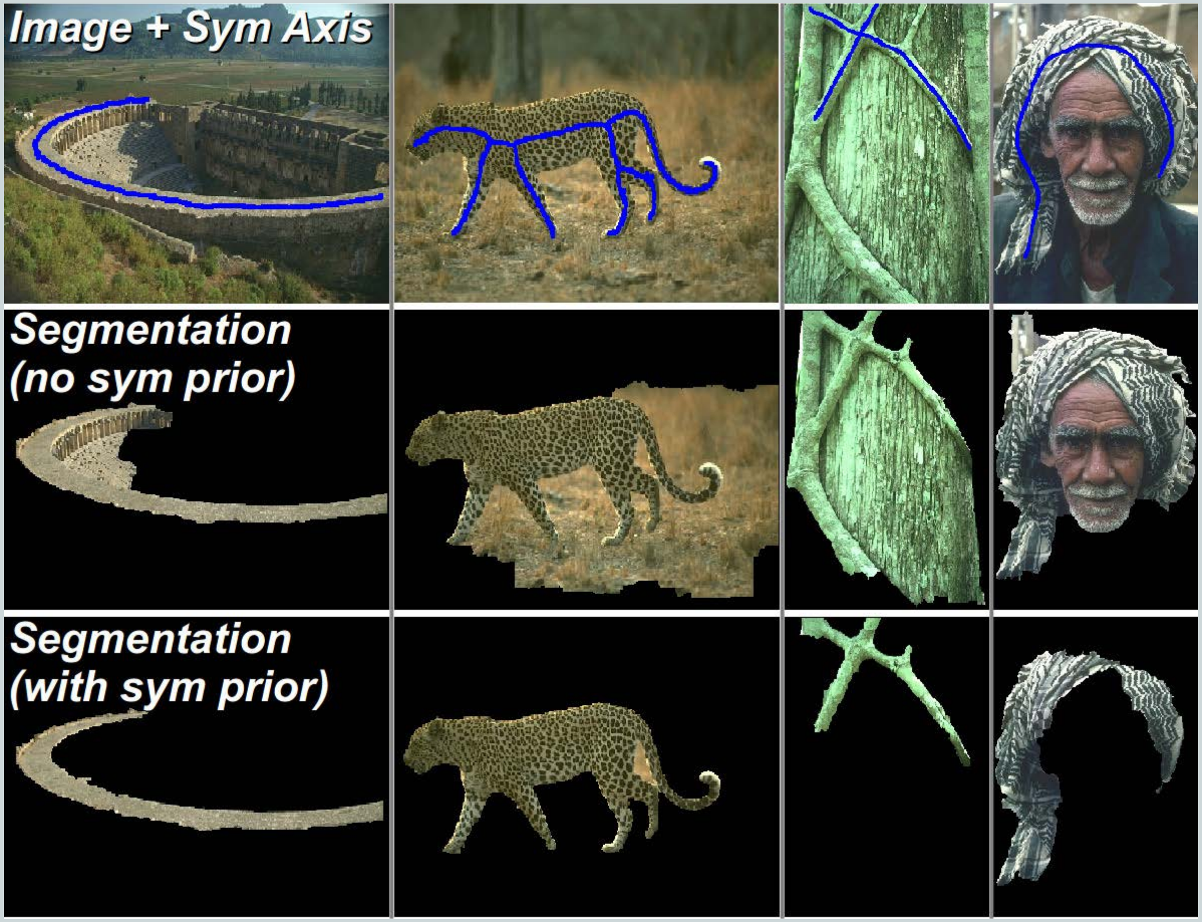
Left: Our approach,

Right: Loy & Eklundh, ECCV 2006

Segmentation with Symmetry

$$E(f) = \underbrace{\sum_{p \in \mathcal{P}} U_p(f_p) + \sum_{(p,q) \in \mathcal{N}} V_{pq}(f_p, f_q)}_{\text{Unary + Pairwise (edges)}} + \underbrace{\sum_{(p,p') \in \mathcal{N}_{sym}} S_{pp'}(f_p, f_{p'}) + \sum_{(p,q) \in \mathcal{N}} B_{pq}(f_p, f_q)}_{\textbf{Cross-symmetry} \text{ + } \textbf{‘Ballooning’}}$$

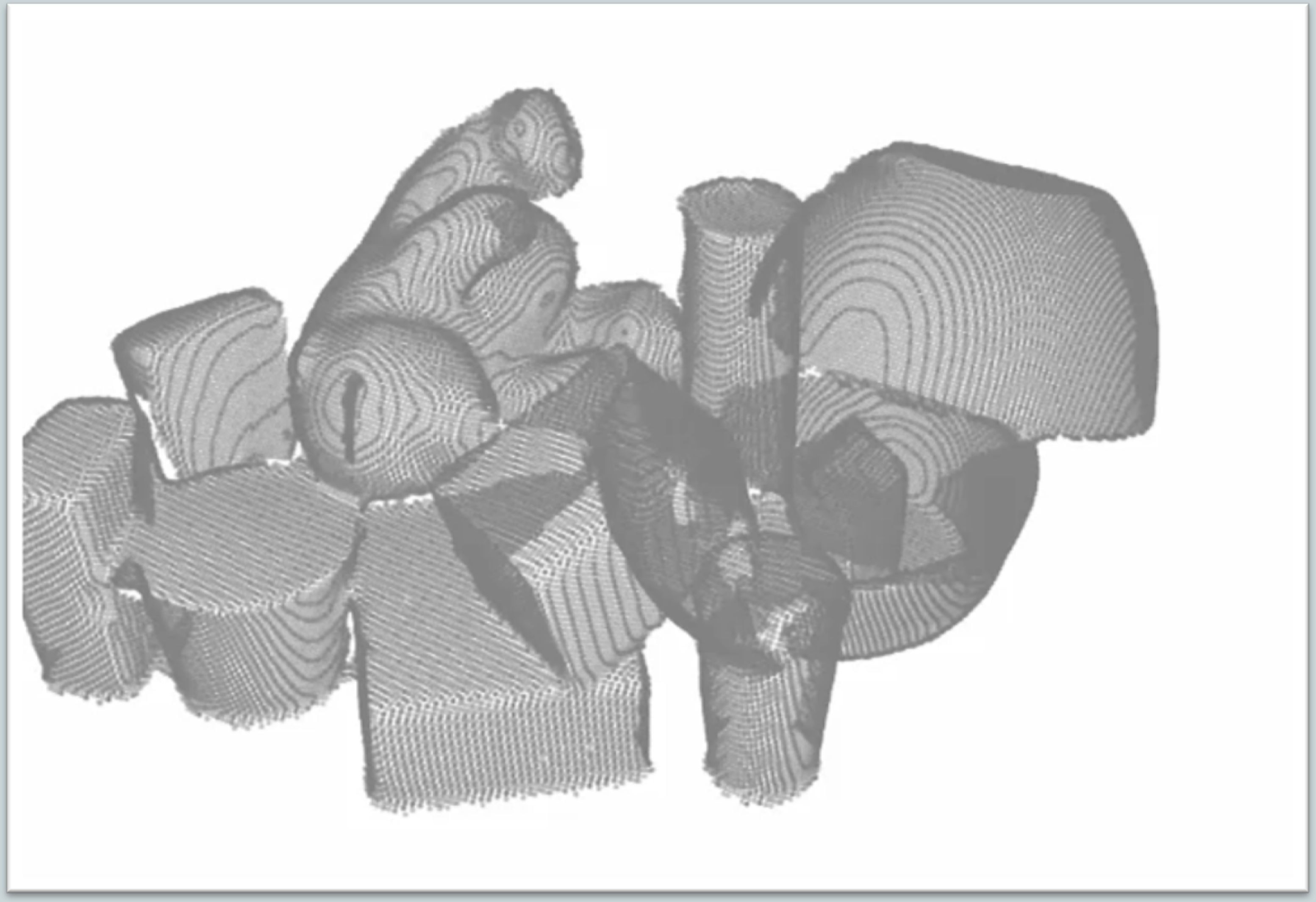
Segmentation with symmetry axes



C.L Teo, C. Fermüller, and Y. Aloimonos, "Detection and Segmentation of 2D Curved Reflection Symmetric Structures," ICCV, 2015.

Detection of 3D Symmetry: Motivation

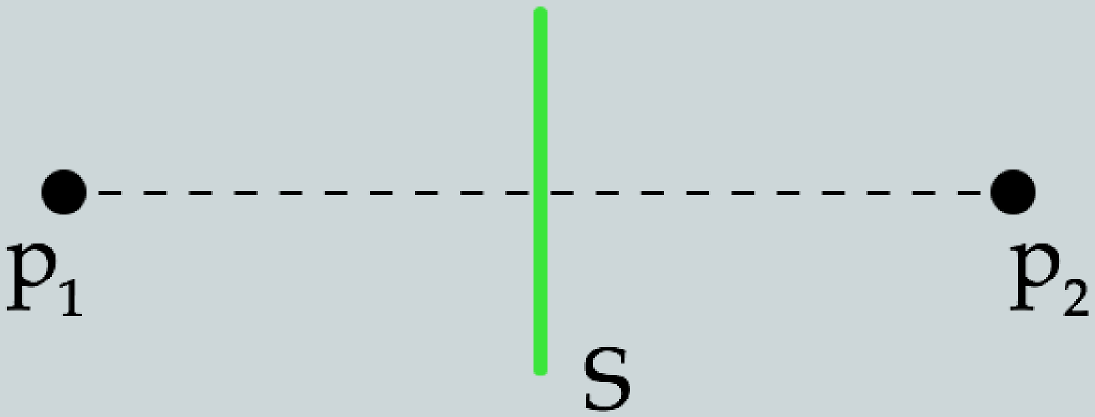
Data



Ecins, C. Fermüller, and Y. Aloimonos, "Cluttered Scene Segmentation Using the Symmetry Constraint," ICRA, 2016.

Symmetric point correspondence

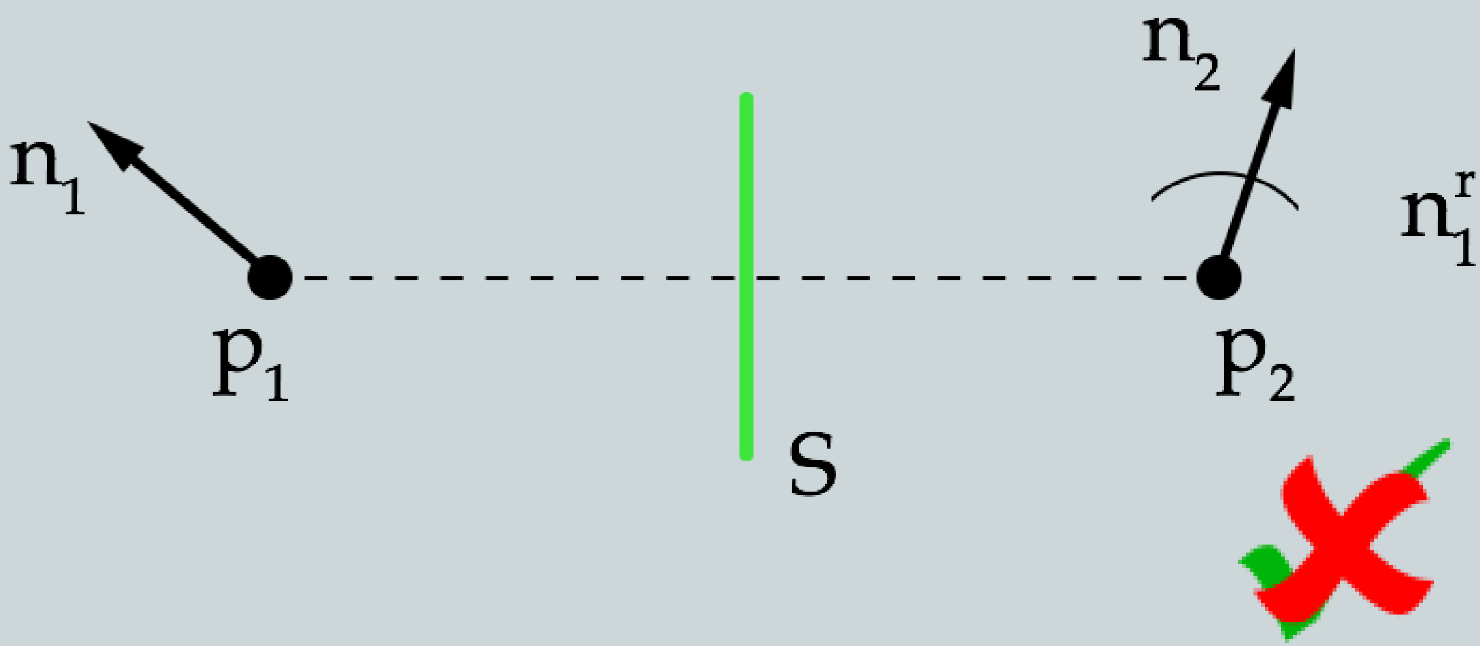
Two points in space uniquely define a reflectional symmetry plane.



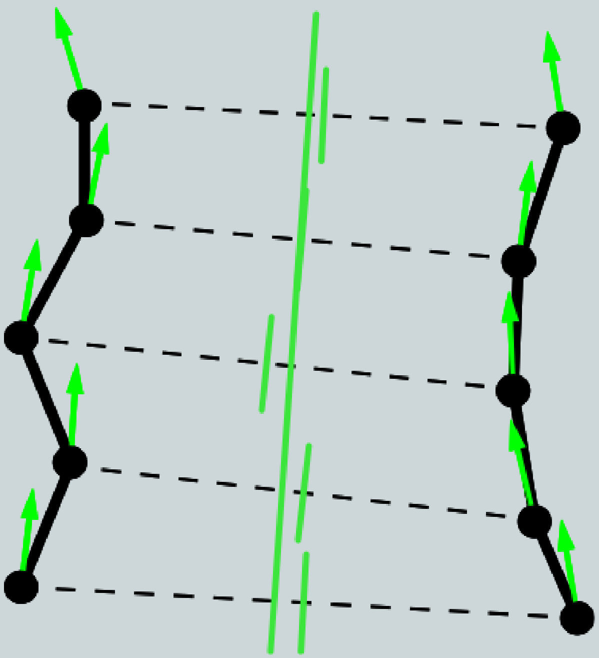
Symmetric point correspondence

Two points in space uniquely define a reflectional symmetry plane

Two oriented points in space form a symmetric match if their reflected normals align

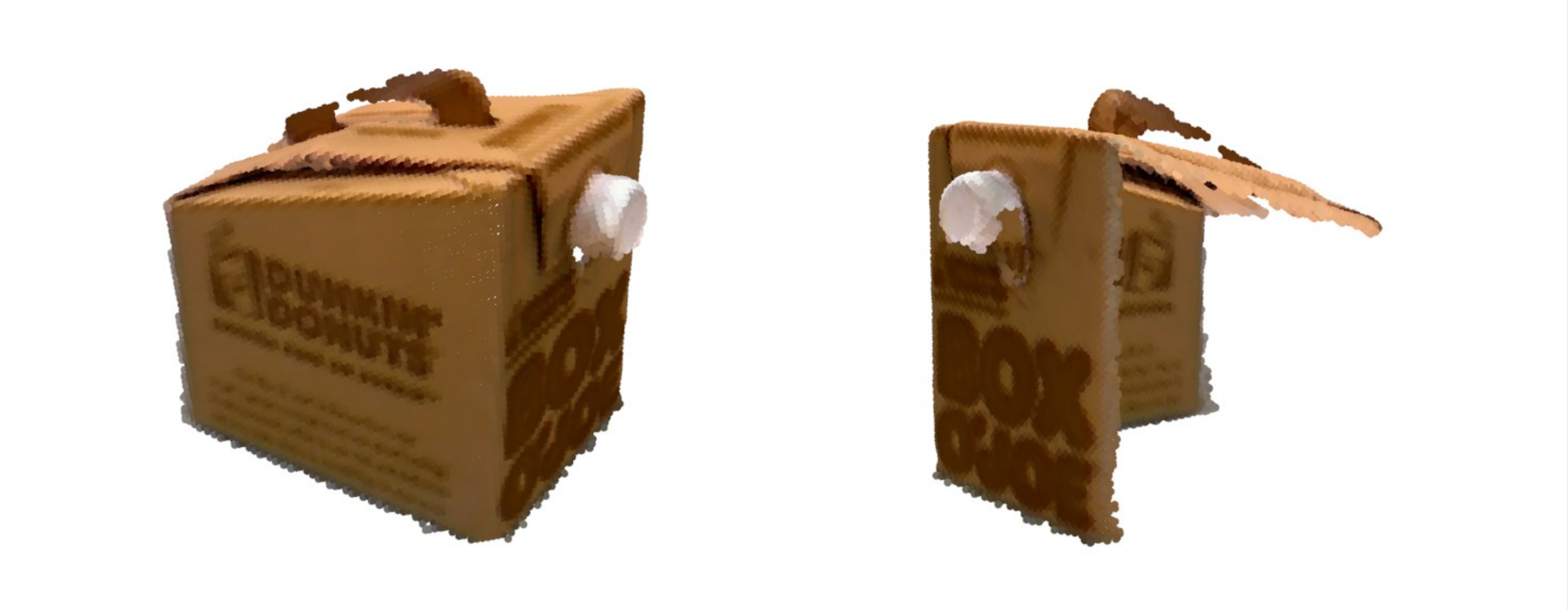


Symmetry detection



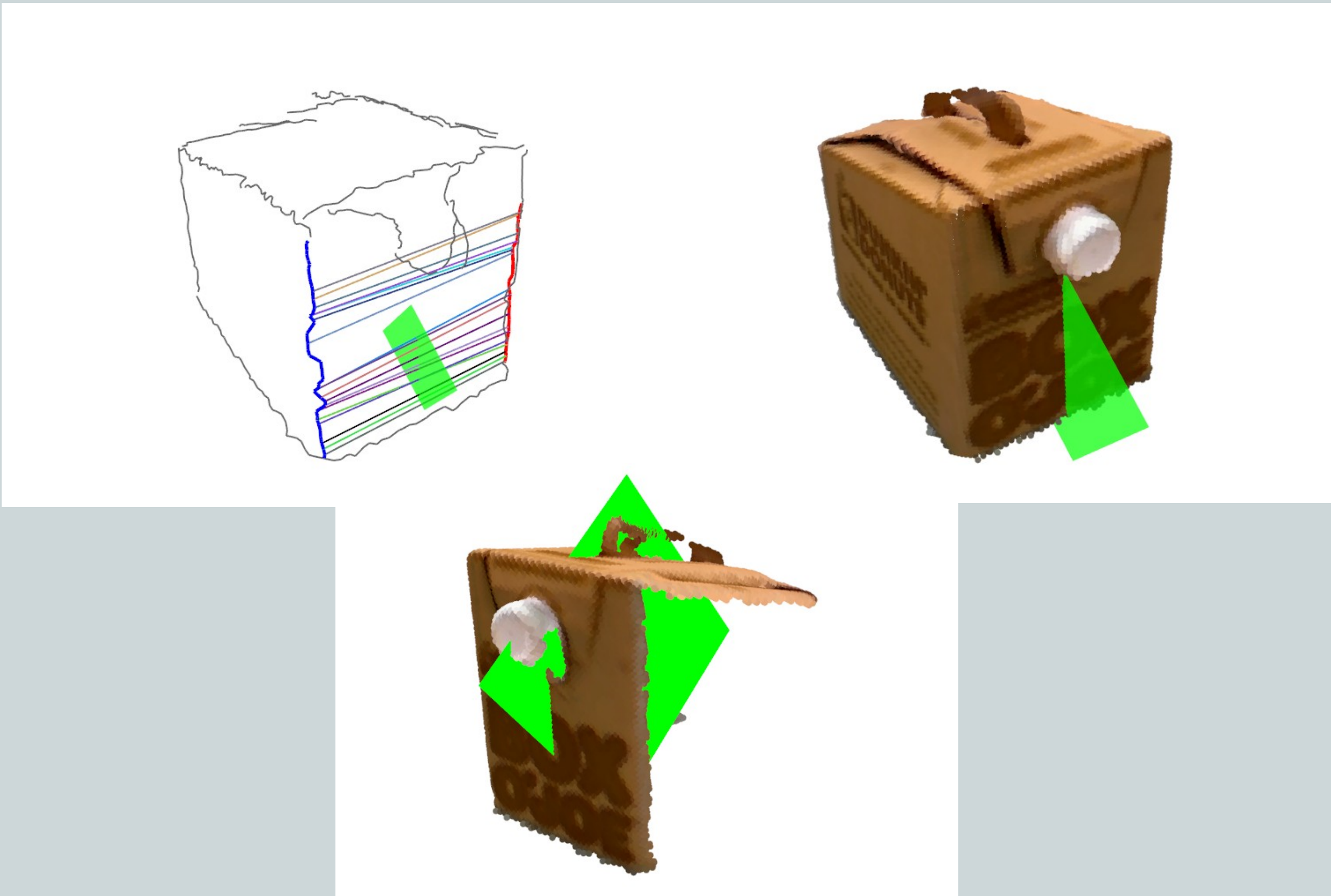
- Find symmetric correspondences between points
- Get a symmetry hypothesis for each correspondence
- Filter hypotheses using mean shift clustering

Input Point clouds



Features: Edges

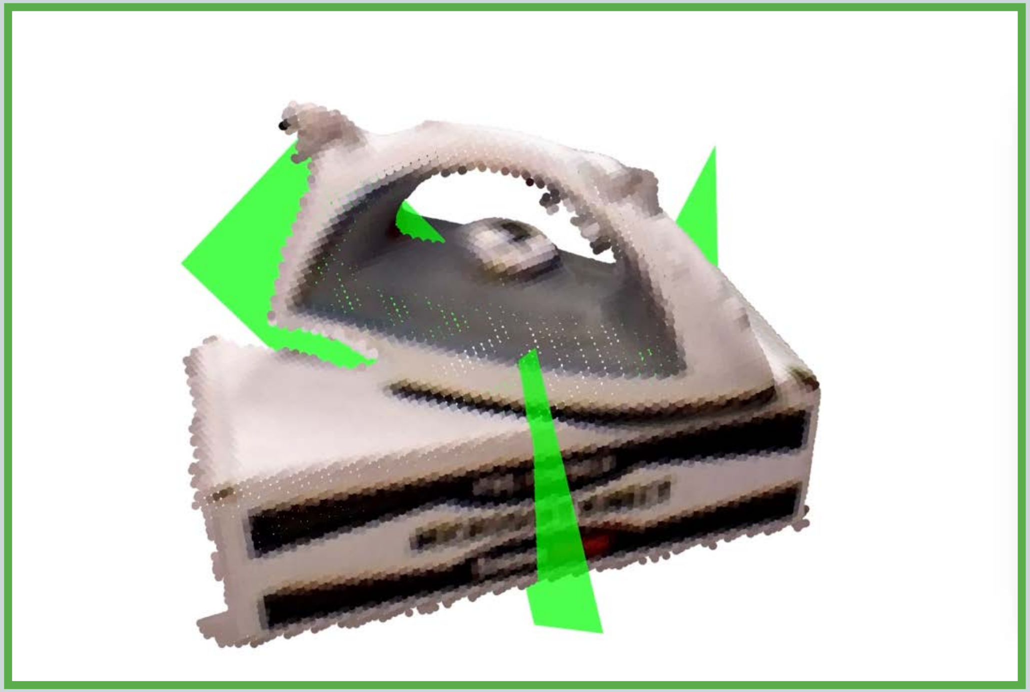
Symmetry Detection



Segmentation

Grouping principles used:

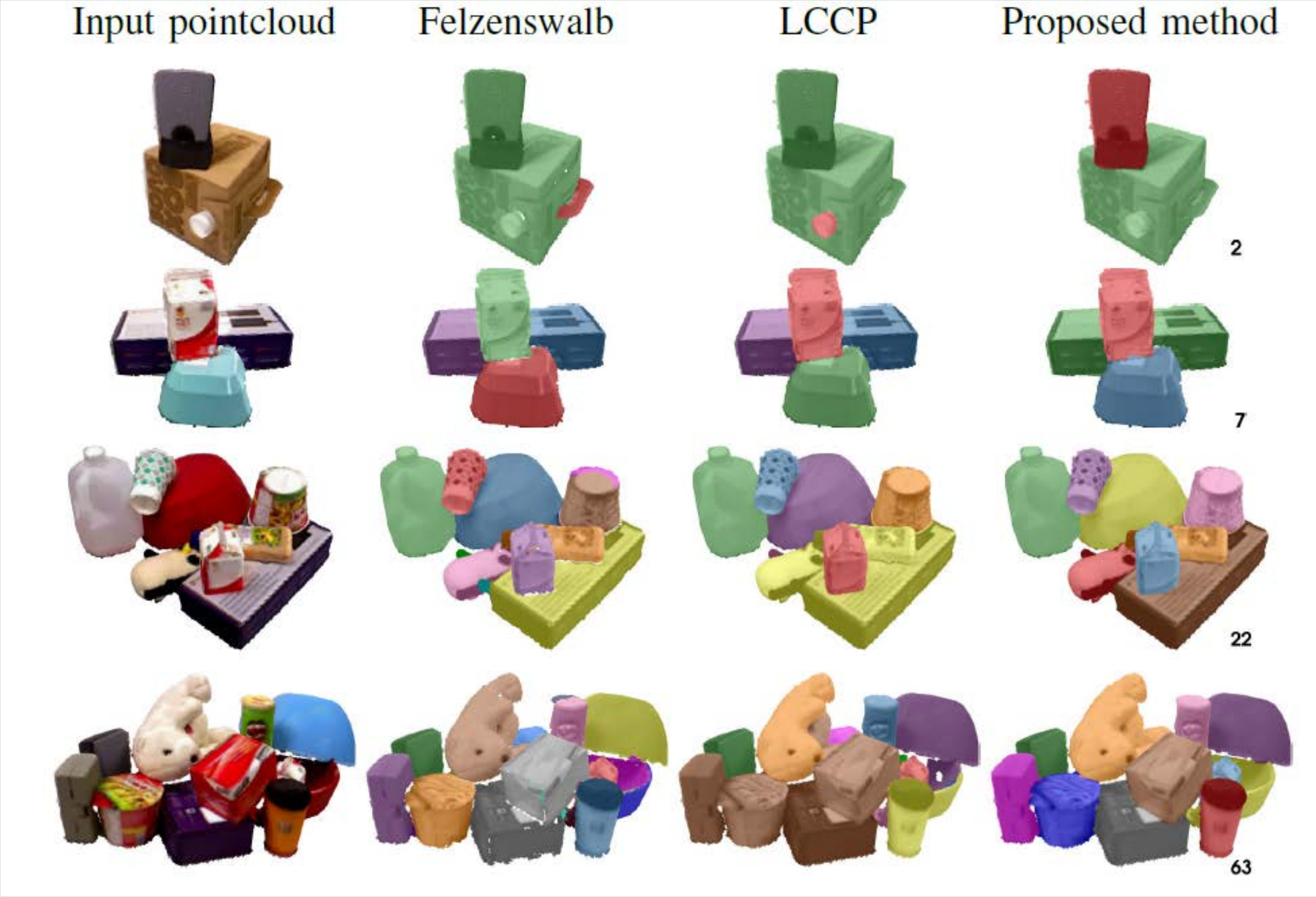
- Convexity (old)
- Symmetry consistency (new)



Final segmentation

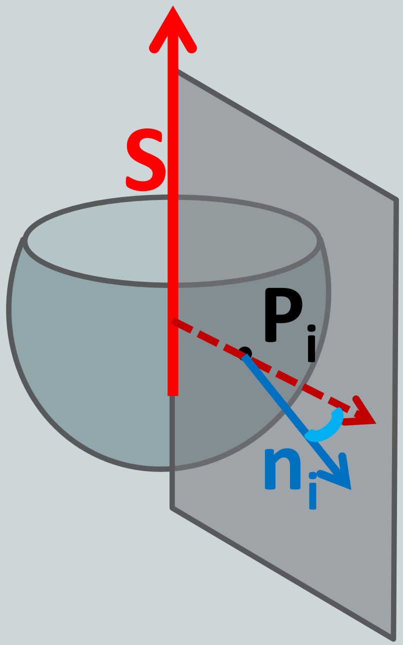
LCCP, Stein et al. CVPR 2014

Segmentation with 3D Symmetry

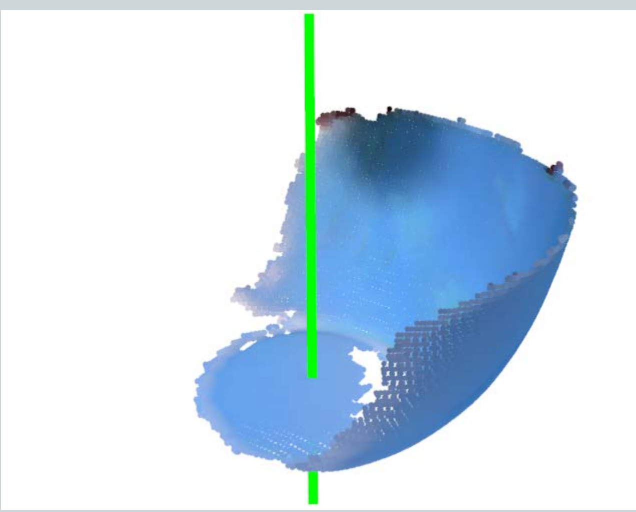


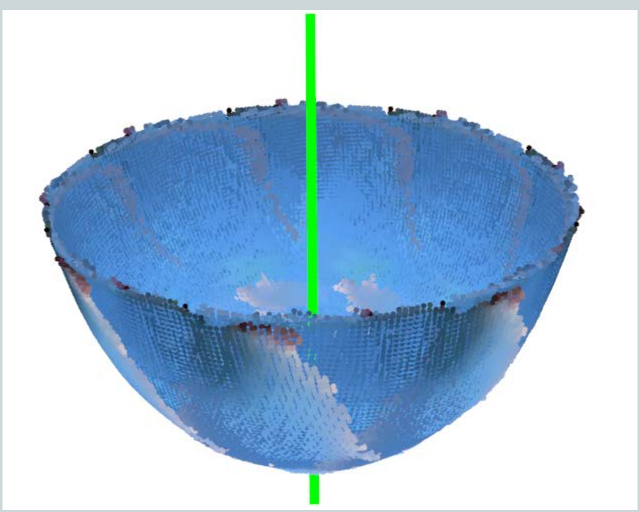
- A. Ecins, C. Fermüller, and Y. Aloimonos, "Cluttered Scene Segmentation Using the Symmetry Constraint," ICRA, 2016.
- LCCP: S. C. Stein, M. Schoeler, J. Papon, and F. Worgötter, “Object partitioning using local convexity,” CVPR, 2014
- Felzenswalb adaptation: A. Karpathy, S. Miller, and L. Fei-Fei, “Object discovery in 3d scenes via shape analysis,” ICRA, 2013.

Rotational Symmetry

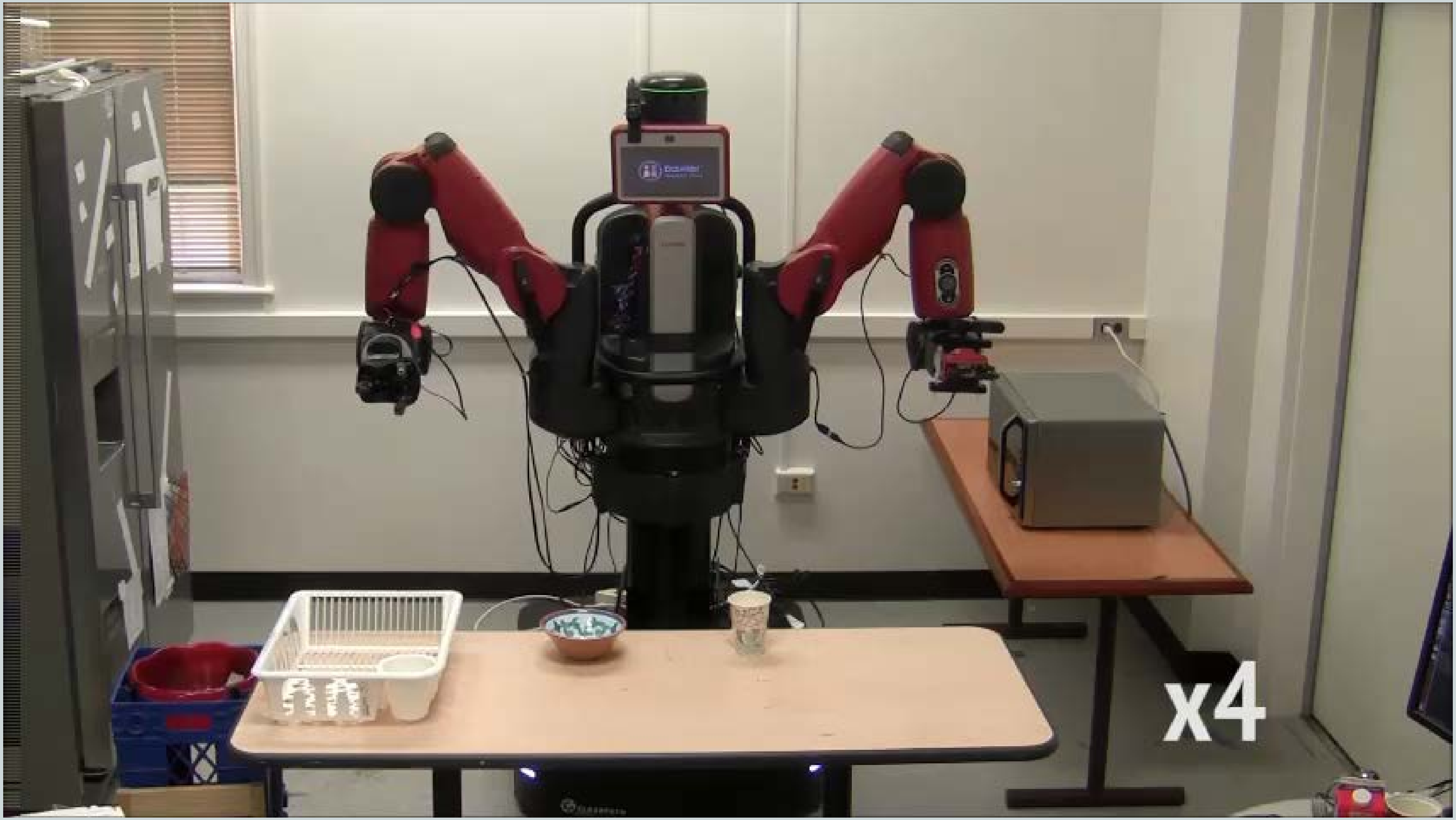


$$\min_{S} \sum_{i} \angle(Plane(S, P_i), n_i)$$





Heating a dish in the microwave



Summary

- Mid-level concepts implemented as image operators
- Bottom-up principles of closure, symmetry, border-ownership
- Top-down task driven modulation of mid-level features
- Symmetry in 3D for object detection and segmentation