Efficient Grasping from RGBD Images: Learning Using a New Rectangle Representation

Yun Jiang, Stephen Moseson, Ashutosh Saxena
Cornell University

Problem

Goal:

- Figure out a way to pick up the object.

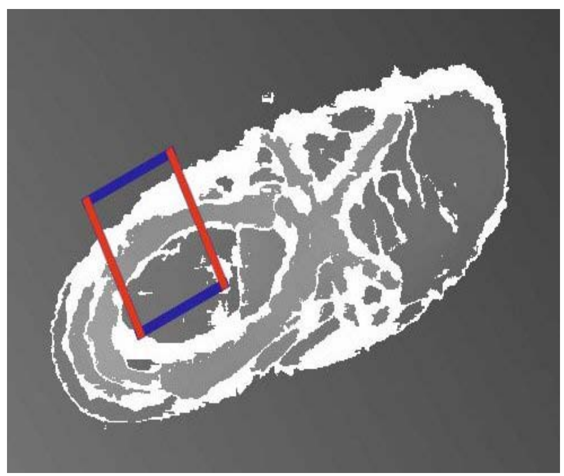
  - Approach
  - Grip
  - Pick up

Question: where and how to grasp?

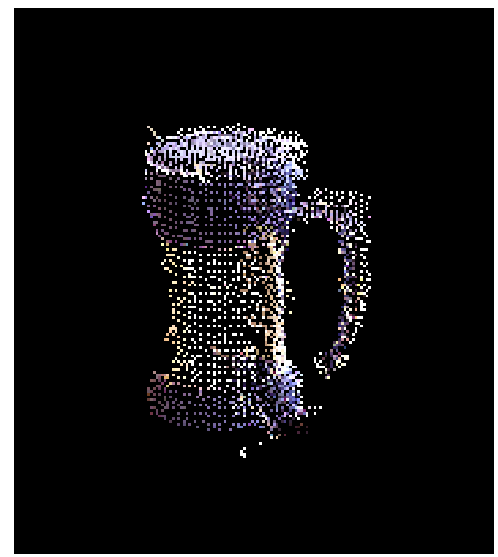
How to Perceive Objects

- RGBD cameras give RGB image plus depth information
  - Stereo cameras ($1000): Bumblebee
  - Kinect Camera ($140)

RGB image



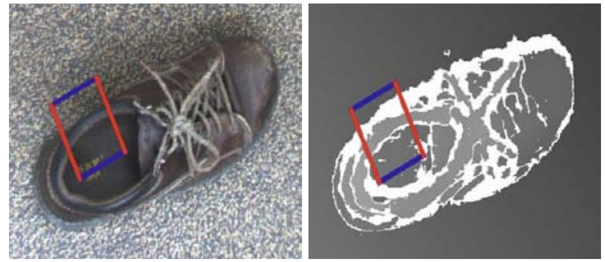
Depth map



3D point cloud

Our Formulation

- Input: RGBD image



- Output: a proper grasp -- the configuration of the gripper at the final grasp stage
  - 3D location, 3D orientation, opening width.

Traditional Approaches

- Control/Planning
  - Force and form closure (Nguyen1986, Lakshminarayana1978)
  - Requires full 3D knowledge of grippers and objects

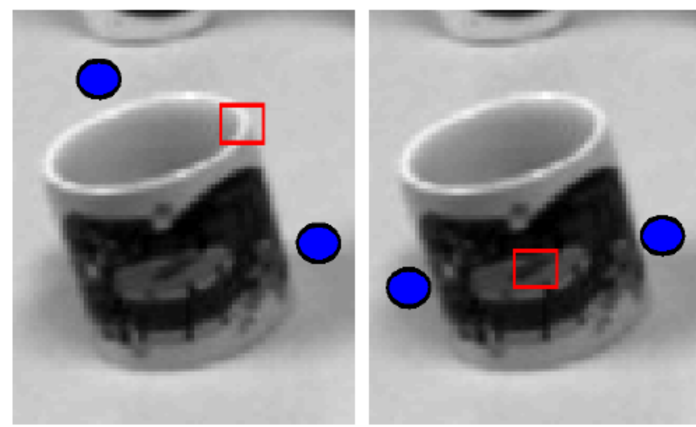
- Disadvantages:
  - Complete 3D model is not always available
    - Noise sensors.
  - Difficult to model friction.
  - Search in enormous configuration space

Does not apply to deformable grippers!

Learning Approaches

- Learning
  - provides generalization on novel objects
  - Robust to noise and variations of environment
- Previous learning approaches
  - Representation problem
    - 3D orientation of gripper not represented well.

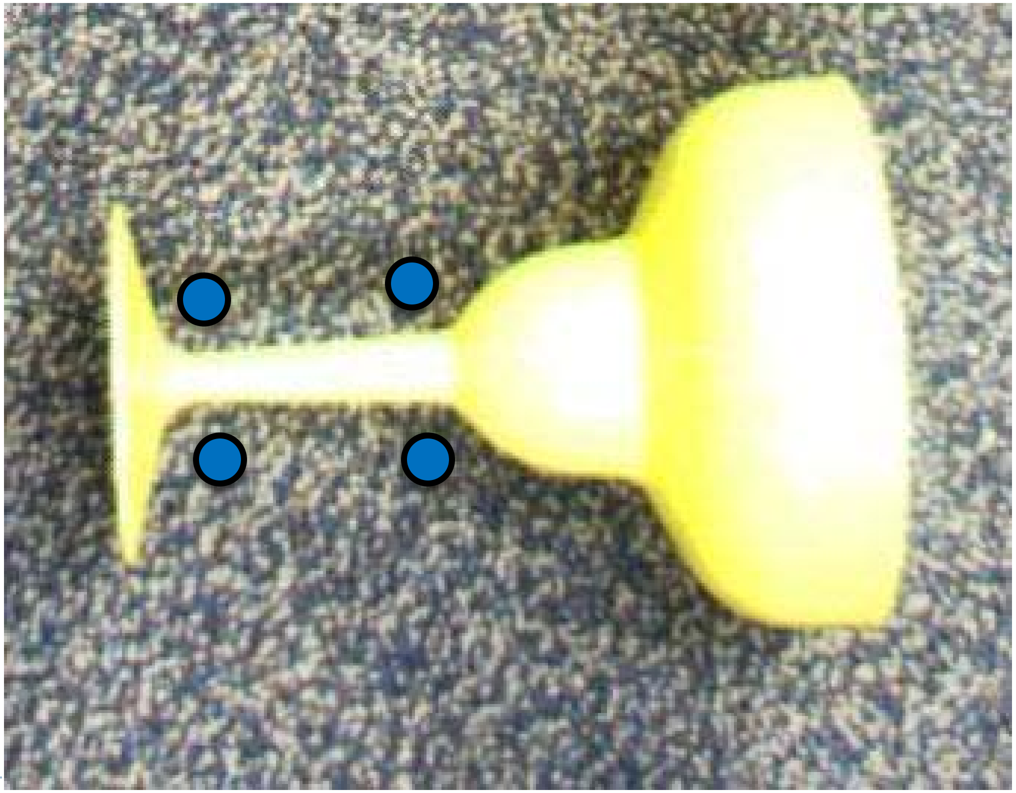
(Saxena et al. , NIPS 2006)



(Le at al. , ICRA 2010)

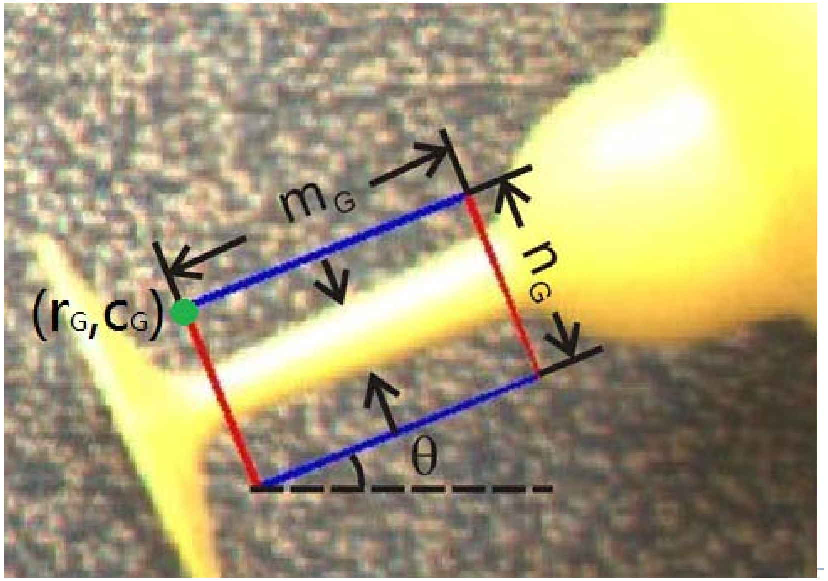
Representation

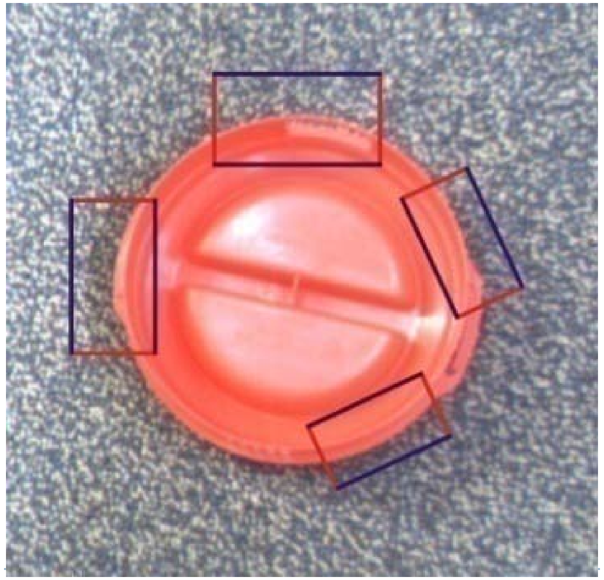
- Should contain full 7-dimensional gripper configuration (3D location, 3D orientation, gripper opening width)
- Specifically model gripper's physical size



New Representation

- Grasping Rectangle
  - Contains full 7-dimensional gripper configuration
  - Specifically model gripper's physical size.
  - Strictly constraints the boundary of features.





Define the Score Function

- $\phi(G)$: the feature vector for a possible grasp G
- Score of grasp G:

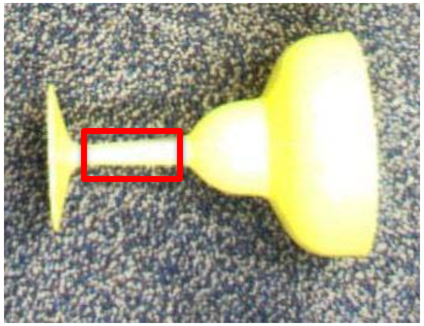
$$f(G) = w^T \phi(G) = \sum_{i=1}^{k} w_i \phi_i(G)$$

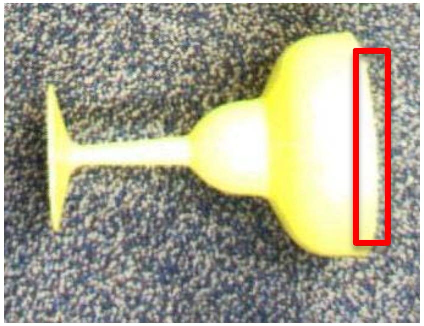
- Best grasp: the highest-score rectangle in the image

$$G^* = \arg\max_G f(G)$$

Learning the Score Function

- Learning algorithm: SVM-Rank
  - Ranking not classification:
    - because the boundary between 'good'/'bad' grasps is vague

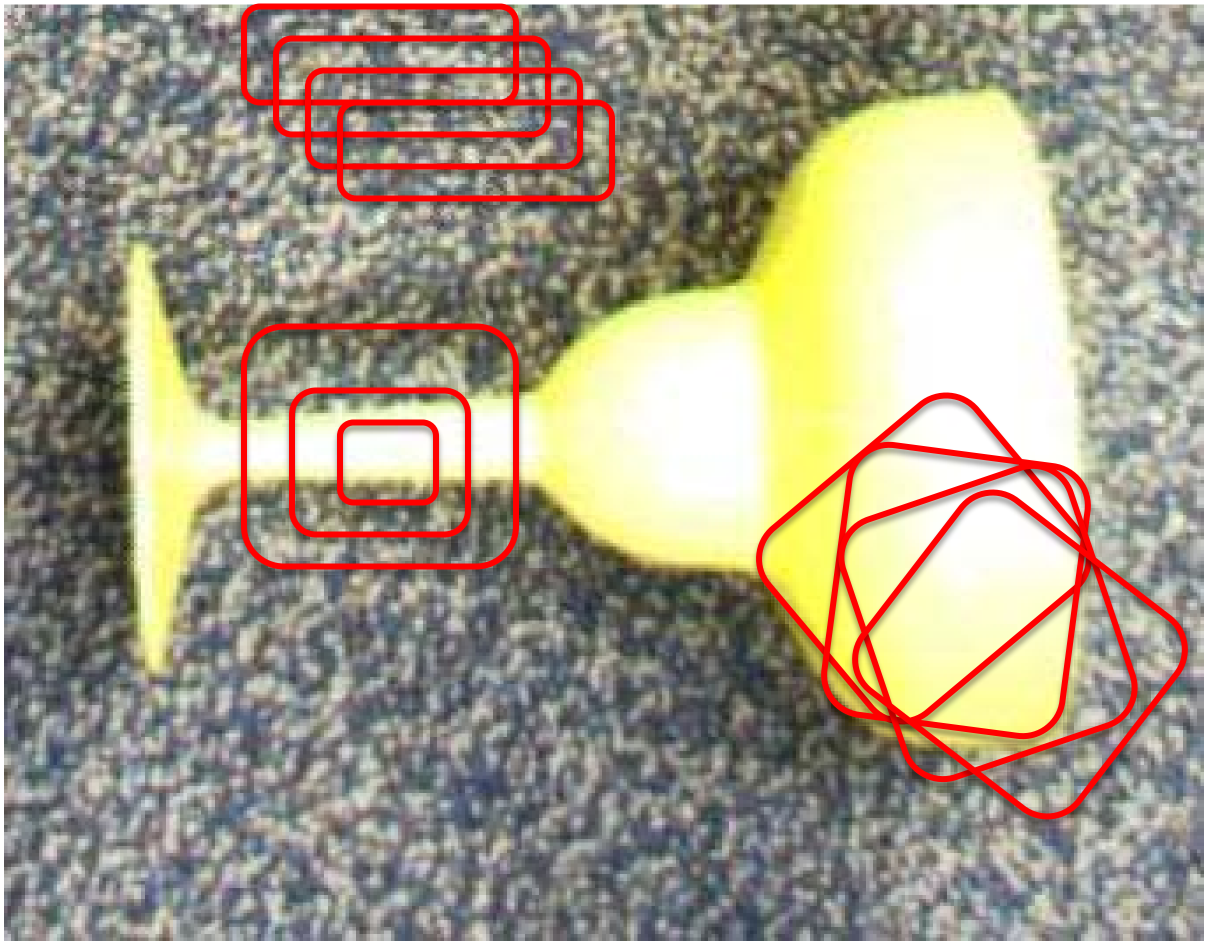




  - Training data: Labeled rectangles for pictures.

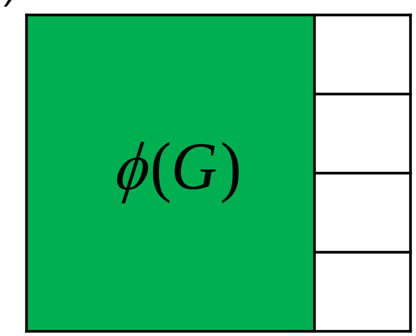
Inference

- Search for all possible rectangles



Search Highest-score Rectangles

- Image: n x m
- Features: k (per rectangle)
- Brute-force search?
  - *$O(n^2m^2)$ rectangles, $O(nmk)$* to compute features ➔ $O(n^3m^3k)$ for one orientation
  - To accelerate:
    - Compute features incrementally➔ *$O(n^2m^2k)$*
    - Even faster?



$$\phi(G + \Delta G) = ?$$

Fast search

- Condition: features are independent in pixel level, i.e.

$$\phi_i(G) = \sum_{(x,y)\in G} \phi_i(I(x,y)), \quad \forall i = 1,\ldots,k$$

- The score of a rectangle can be decomposed to the scores of pixels

$$f(G) \; = \; \sum_{x=r_G}^{r_G+n_G} \sum_{y=c_G}^{c_G+m_G} F(x,y)$$

- Classical problem: maximum-sum submatrix!
  - In one dimension,

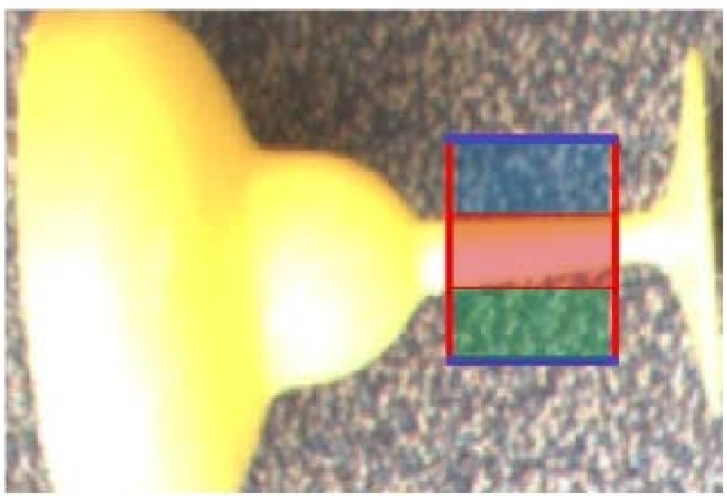
array	3	-4	5	2	-5	5	9	-8
sum	3	0	5	7	2	7	16	8

  - In our problem, reduce the time complexity to $O(nmk+n^2m)$

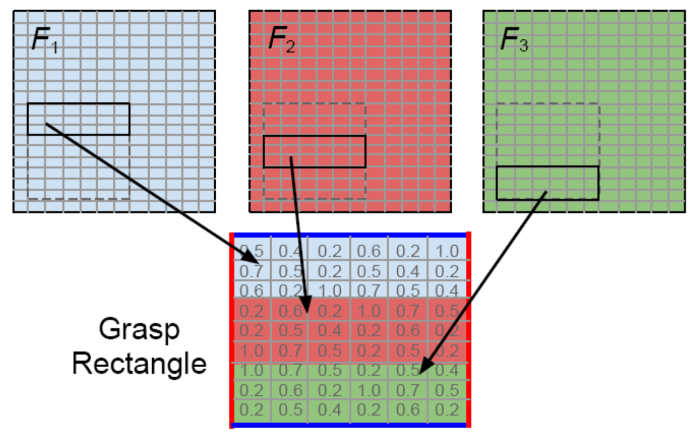
Histogram Features for Fast Search

- Histograms from 15 filters to capture color, textures and edges

- Spatial Histogram Features

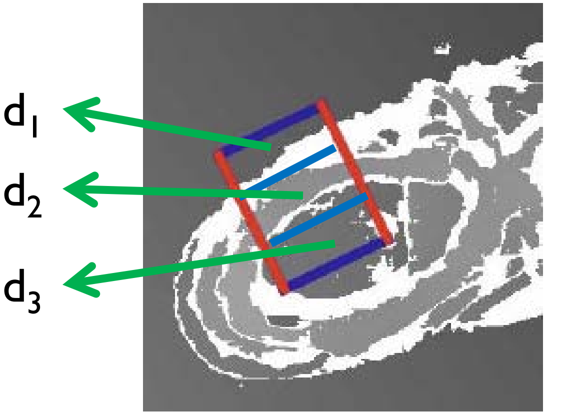


Divide a rectangle into 3 sub-rectangles



Advanced Features

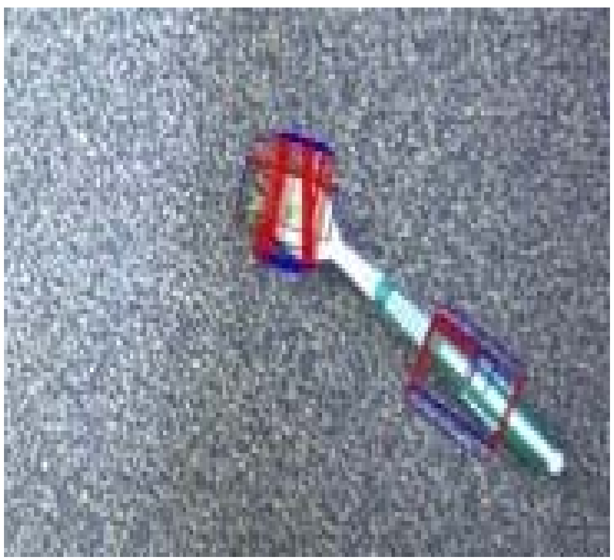
- Histogram is fast but not able to capture the correlations among the 3 sub-rectangles
  - E.g., One criteria: $d_1 > d_2$ and $d_2 < d_3$



- Non-linear features
  - E.g., $d = d_1 d_3 / (d_2)^2$
  - Expressive but not applicable to fast search

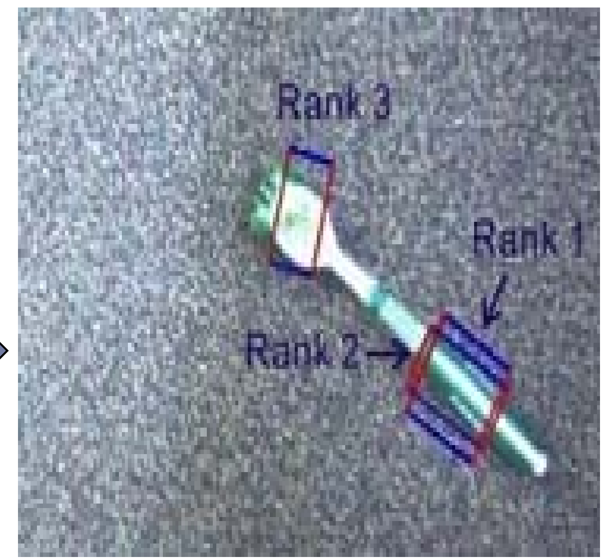
Two-step Process

- Algorithm: Two models:
  - First step: Fast, but not accurate (good for pruning).
  - Second step: Accurate, but slow.



Top 100 rectangles after the 1st step

Step2: Re-ranking



Top 3 rectangles after the 2nd step

Summary

- RGBD images
- Representation
  - Oriented rectangle
- Learning using Efficient two-step process
  - Fast search with histogram features
  - Re-rank with more sophisticated features

Experiments

- Tested on novel objects

- Offline: 128 images
- Robot: 12 objects, multiple tries

Results on offline test

- Evaluation-1: rectangle metric

Dataset	Object Specific Training			General Training
	One-step RGB	Two-step RGB	Two-step RGBD	Two-step RGBD
Martini	75.0	80.0	87.5	92.5
Marker	79.3	93.1	89.7	100.0
Pencil Bag	66.7	88.9	100.0	100.0
Dumbbell	66.7	77.8	100.0	100.0
Screwdriver	75.0	87.5	81.3	87.5
Brush	81.8	81.8	100.0	90.9
Black Container	75.0	75.0	100.0	100.0
Red Lid	66.7	100.0	100.0	100.0
White Box	85.7	85.7	100.0	100.0
Average	74.7	85.5	95.4	96.8

Results on offline test

- Evaluation-2: point metric [Saxena2008]

Dataset	Object Specific Training			General Training
	One-step RGB	Two-step RGB	Two-step RGBD	Two-step RGBD
Martini	77.5	85.0	87.5	95.0
Marker	93.1	93.1	96.6	100.0
Pencil Bag	77.8	88.9	100.0	88.8
Dumbbell	100.0	100.0	100.0	100.0
Screwdriver	87.5	93.8	100.0	100.0
Brush	100.0	90.9	100.0	100.0
Black Container	100.0	100.0	100.0	100.0
Red Lid	66.7	100.0	100.0	100.0
White Box	85.7	100.0	100.0	100.0
Average	87.6	94.6	98.2	98.2

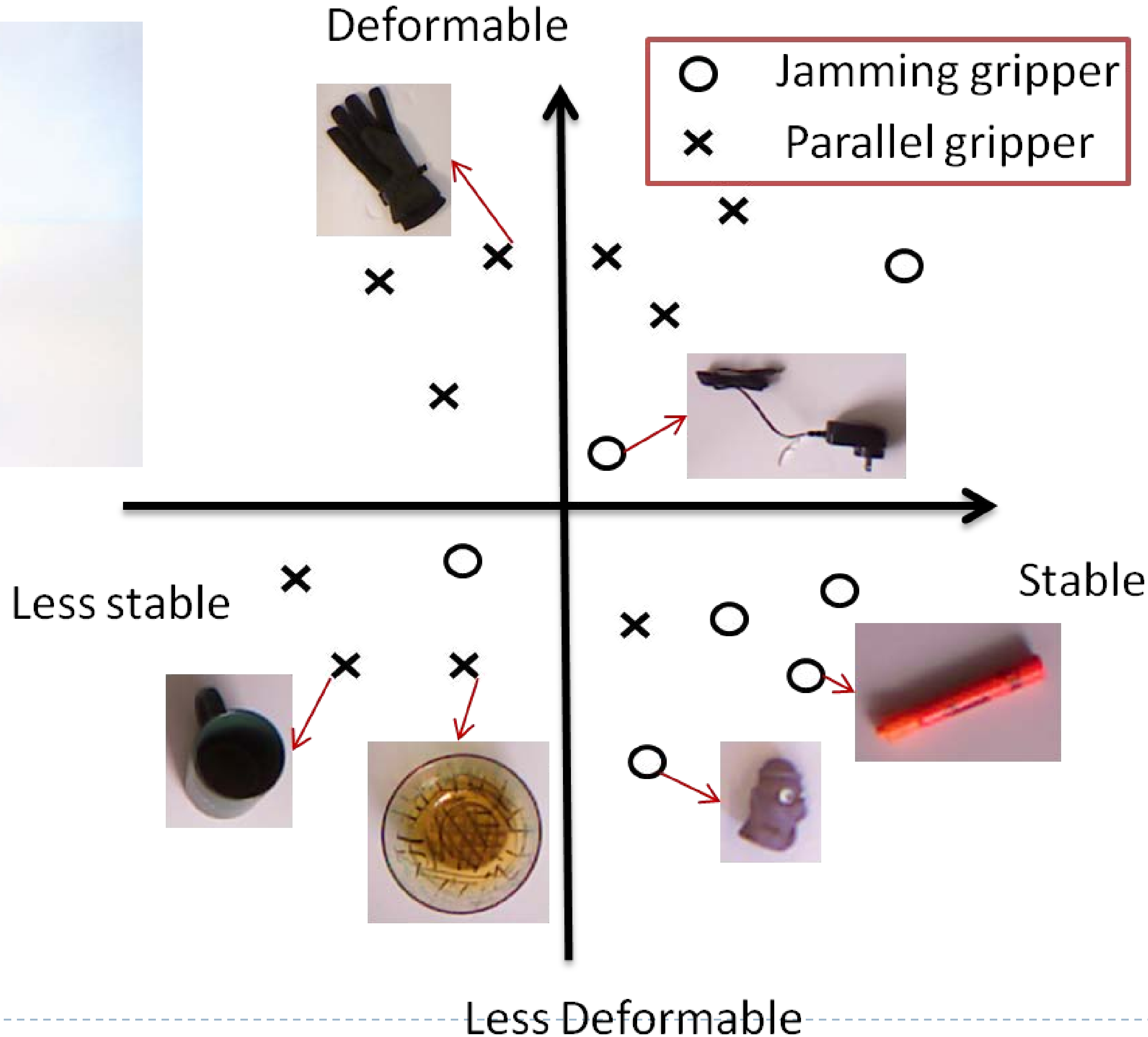
Robotic experiments

- Adept Viper s850
- Parallel plate gripper

Results on robotic experiments

Object	Prediction correct (%)	Reaching success (%)	Grasping/Holding success (%)
Martini	100	100	100
Markers	80	80	80
Red Lid	100	100	100
Wire Stripper	100	100	100
Screwdrivers	89	78	78
Pencil Bag	100	100	100
Plastic Case	100	100	100
Book Stand	100	100	50
Glove	100	100	100
Window Wiper	80	80	80
Blue Foam	100	100	100
Shoes	50	67	67
Total	91.6	92.1	87.9

Universal Jamming gripper: Robotic Experiment and Analysis



After Grasp: Learning to Place


- Challenges:
  - Enormous search space
  - Placing under preference
- Efficient learning approach to identify good placements
- Results on robotic experiment
  - Goal: correct location and preferred orientation
  - 92% for New Objects in New Environments.




Thank you!

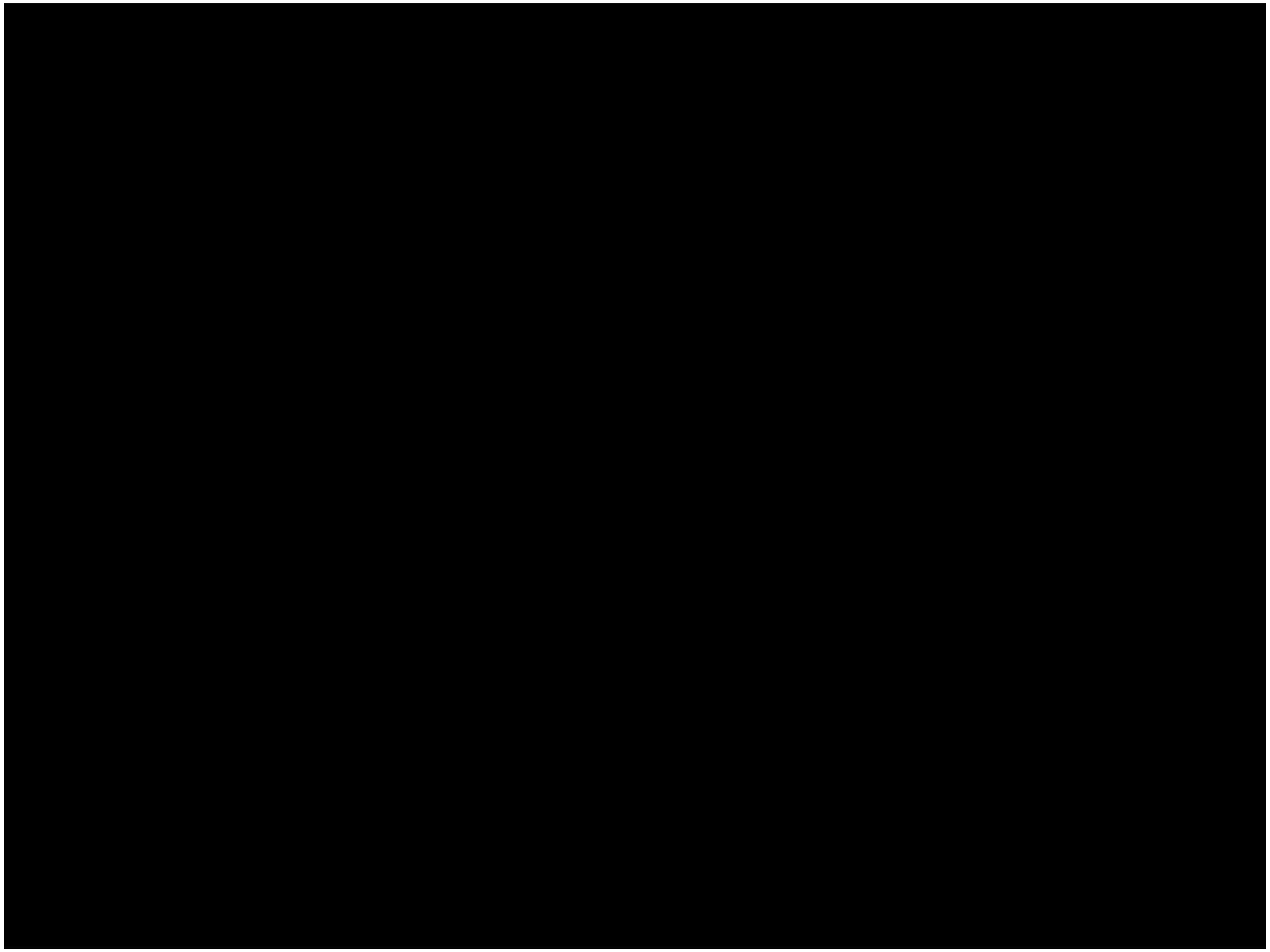
Yun Jiang, Stephen Moseson and Ashutosh Saxena,

Efficient Grasping from RGBD Images: Learning using a new Rectangle Representation, ICRA 2011.

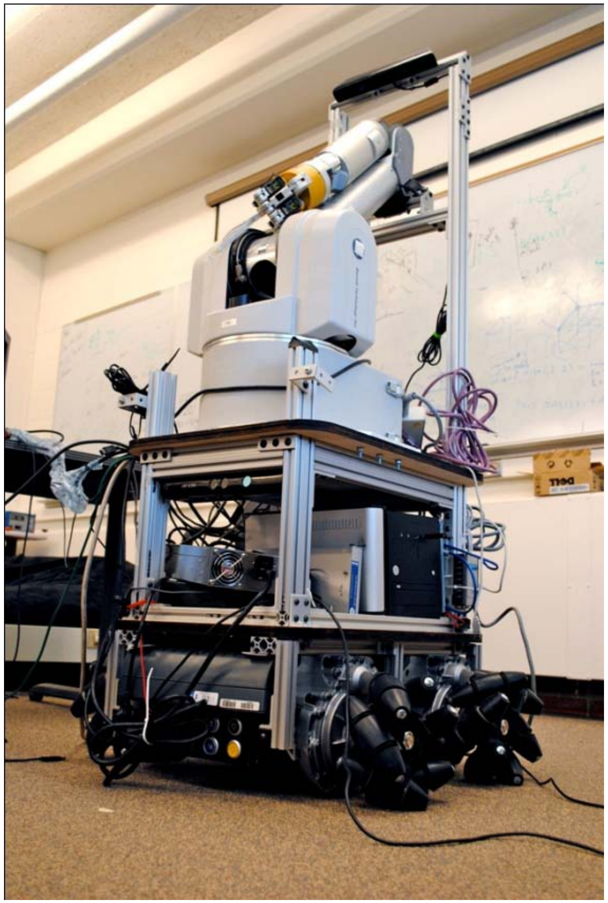
Learning to Place New Objects:

- Yun Jiang, Changxi Zheng, Marcus Lim, Ashutosh Saxena, Learning to Place New Objects, ICRA 2012. First appeared in RSS workshop on mobile manipulation, June 2011.

Video

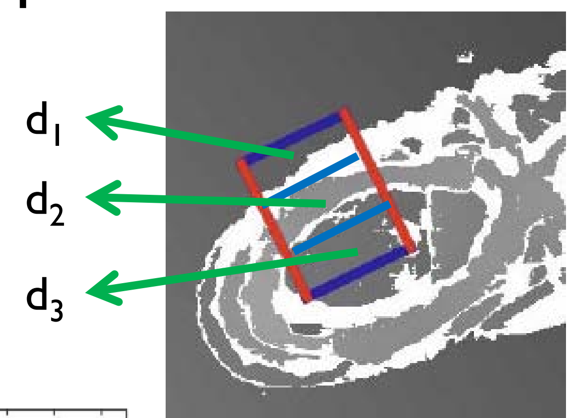


Future Work

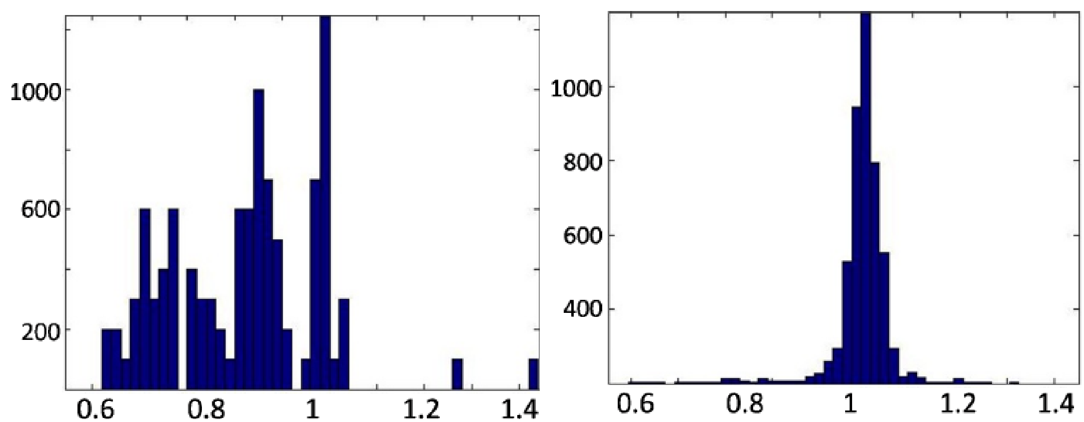


Advanced Features

- Histogram is fast but not able to capture the correlations among the 3 sub-rectangles
  - E.g., One criteria: $d_1 > d_2$ and $d_2 < d_3$



- Non-linear features



(a) histogram of positive examples (b) histogram of negative examples

Histogram of a non-linear feature $d = d_1 d_3/(d_2)^2$

Spatial Histogram for Fast Search

- Time complexity is only multiplied by 3