A vision-based system for grasping novel objects in cluttered environments

Ashutosh Saxena, Lawson Wong, Morgan Quigley, Andrew Y. Ng
2007

Learning to open new doors

Ellen Klingbeil, Ashutosh Saxena, Andrew Y. Ng
2008

Harmen Jeurink

Outline

- Introduction
- About the robot used
- Grab and manipulate an object
  - Recognizing
  - Locate precisely
  - Maneuvering of arm
  - Grasping
  - Manipulating
- Conclusion

Introduction

- Two articles about integrating AI techniques into one system

- Goal of research: build an autonomous robot that is capable of grasping novel objects or open novel doors.

- Short articles, so some explanation about the techniques used that is not in the articles

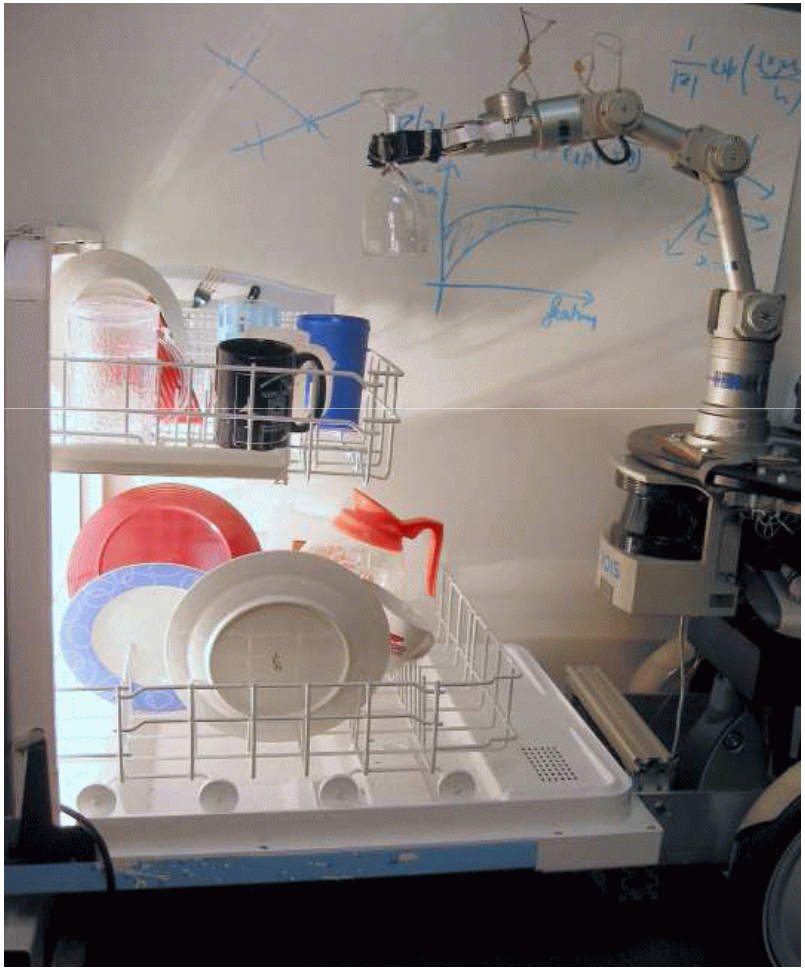
Introduction

- Most previous work is with environment knowledge (location of door, type of doorknob, location of object)
- Not being able to manipulate (ask a human to press the right floor button in an elevator)
- Not suited for autonomous systems that operate in novel environments

The robot

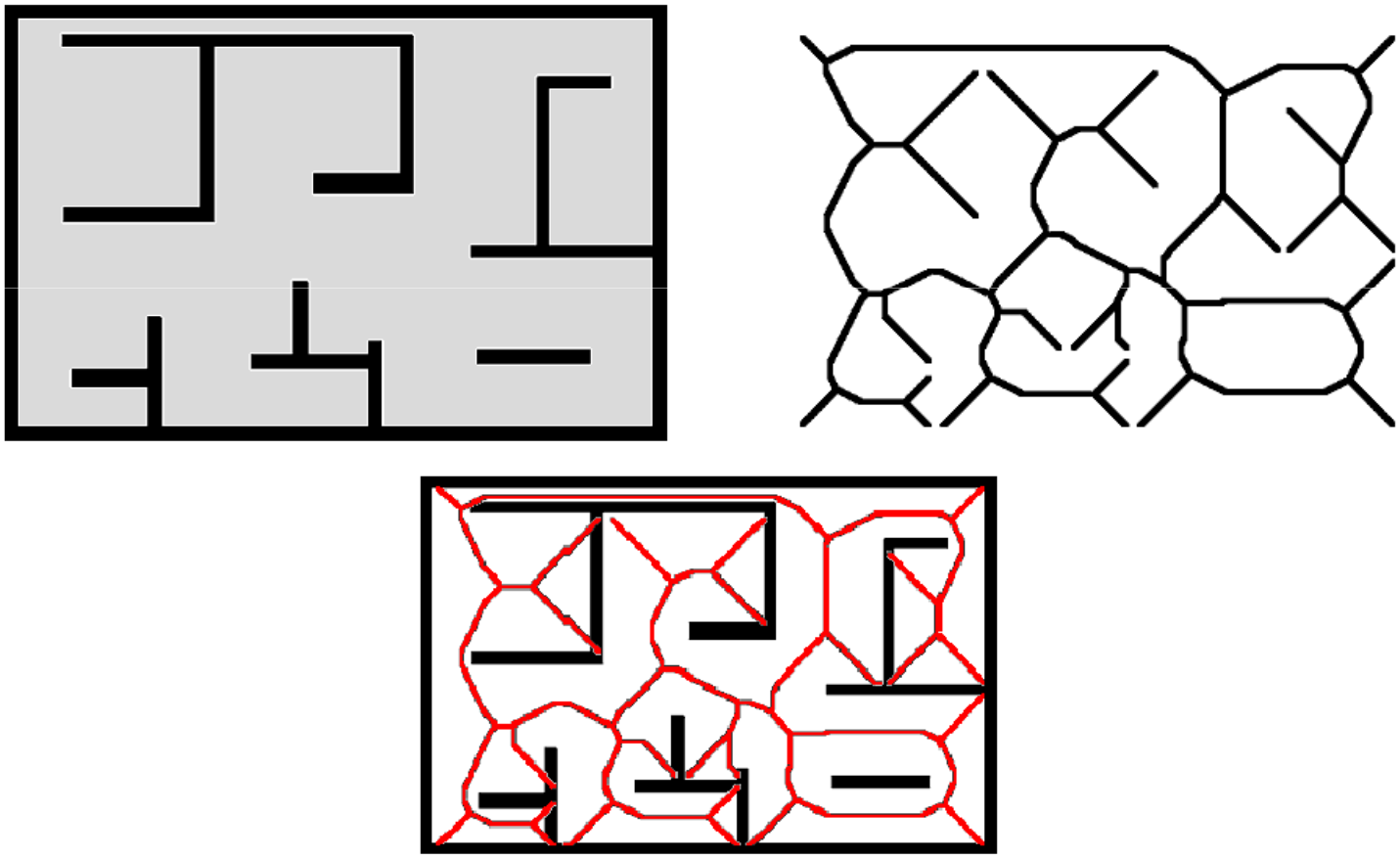
- General-purpose household robot
- Robot is equipped with
  - A 5 or 7 DOF arm (Stair 1 or Stair 2)
  - Webcam near the endeffector
  - Stereo camera
  - Pan-tilt-zoom camera mounted on a frame
  - Laser scanner for navigation
  - Another laser scanner

The robot



Robot navigation

- Voronoi-based global navigation planner

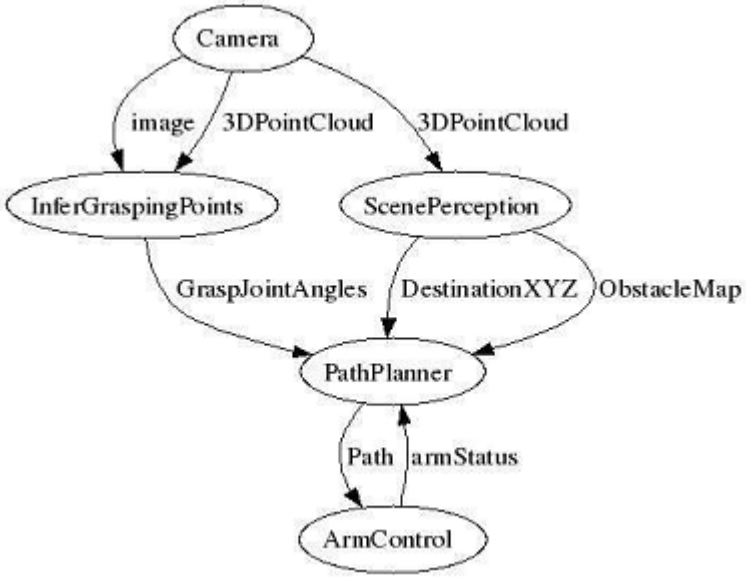


Software architecture

"Switchyard"

Inter-process communication using TCP

Each process runs on a different computer



Recognize door handles using vision

- Produce dictionary of Haar features of learning set
- Make decision tree of dictionary and remove irrelevant features

- Use spatial clues (context) to improve recognition

Haar features

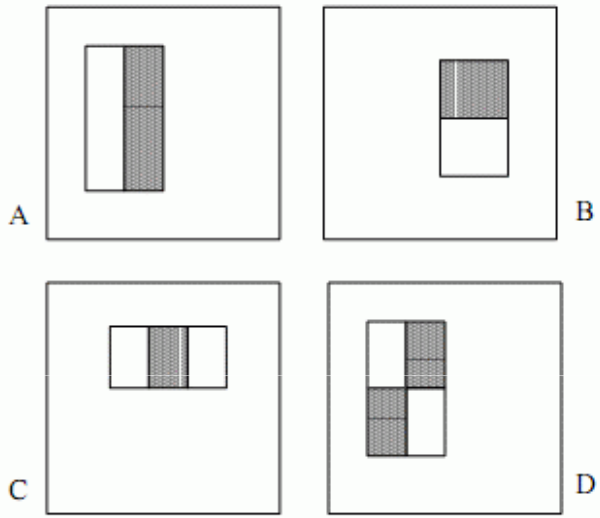


Figure 1: Example rectangle features shown relative to the enclosing detection window. The sum of the pixels which lie within the white rectangles are subtracted from the sum of pixels in the grey rectangles. Two-rectangle features are shown in (A) and (B). Figure (C) shows a three-rectangle feature, and (D) a four-rectangle feature.

Recognize door handles using vision

Fig. 3. Results on test set. The green rectangles show the raw output from the classifiers, and the blue rectangle is the one after applying context.

Recognizing door handles using vision

- Localization is considered accurate if the location was within 2 cm of the estimated location

TABLE II

ACCURACIES FOR RECOGNITION AND LOCALIZATION.

	RECOGNITION	LOCALIZATION
DOOR HANDLE	94.5%	93.2%
ELEVATOR BUTTONS	92.1%	91.5%

Recognize objects

- Previous work is with a known map of environment and objects, which significantly simplifies detection

- Point cloud from stereo vision is often noisy
- Points from laser system are sparse
- No complete 3D model

Recognize objects using stereo vision

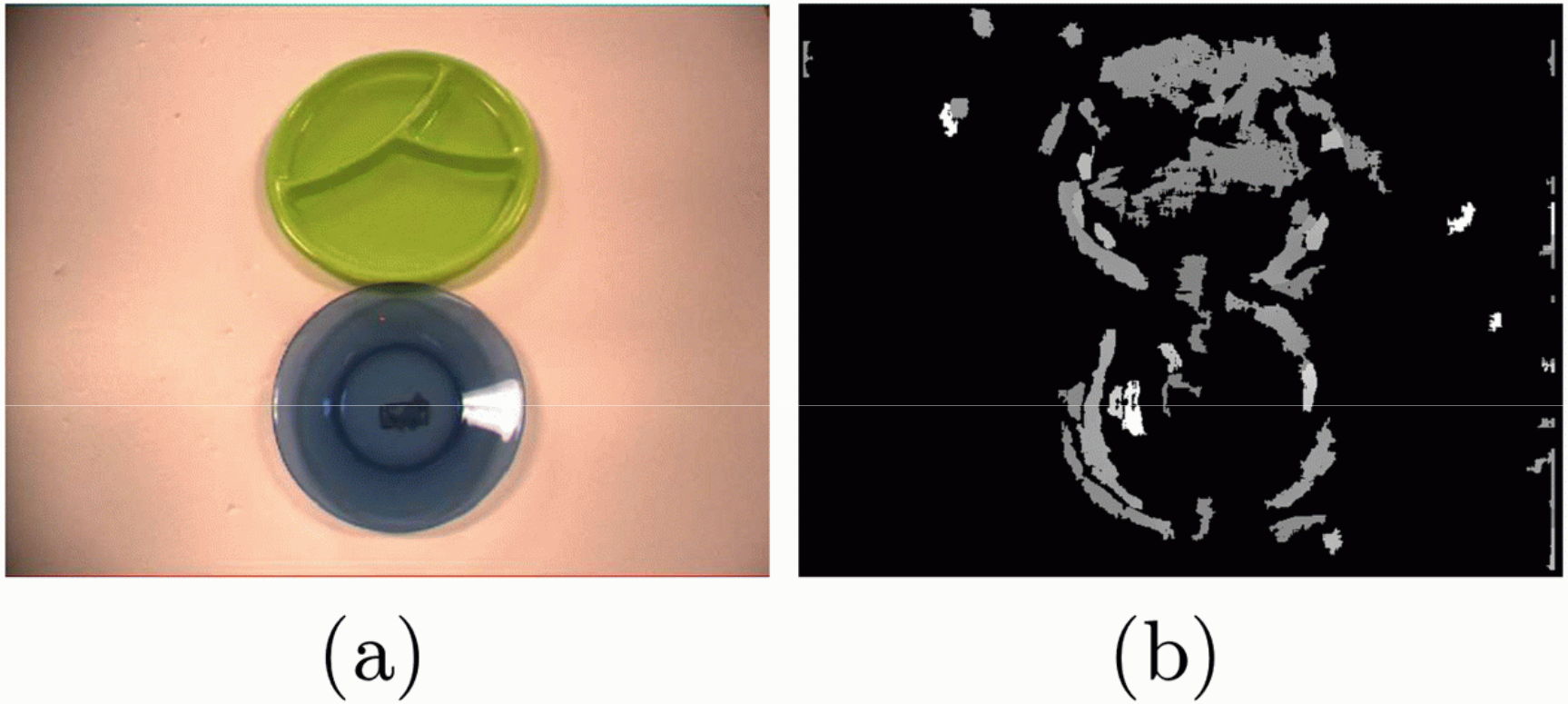


Fig. 2. (a) An image of textureless/transparent/reflective objects. (b) Depths estimated by our stereo system. The grayscale value indicates the depth (darker being closer to the camera). Black represents areas where depth-finding failed.

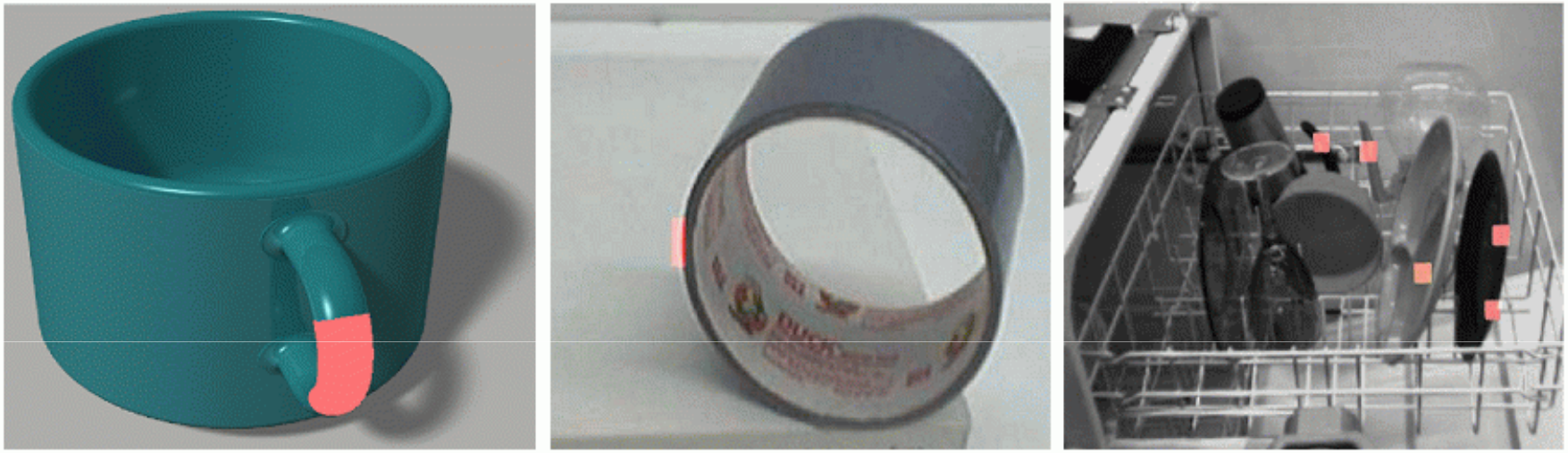
Recognize grasping points

- Determine grasping points using a learning algorithm with supervised learning
- Training set consists of synthetic images
- Images are labelled with good grasping points

Recognize grasping points

- Algorithm works well, except in noisy environments
- Cluster per 3 points, lonely points are considered noise and are discarded
- Clustering makes algorithm robust in cluttered environment

Recognize grasping points



(a) (b) (c)

Fig. 3. Grasping point classification. (a) A synthetic image of a coffee mug with the grasp labels shown in red, (b,c) Test on new real objects: The red points in each image show the locations most likely to be the grasping point, as predicted by our algorithm. (Best viewed in color.)

Grasping

Grasp by closing fingers until movement stops

Works very well for non-deformable objects

Improvement can be made using haptic or optical feedback

Grasping novel objects

Table 1: Average absolute error in locating the grasp point for different objects, as well as grasp success rate for picking up the different objects using our robotic arm. (Although training was done on synthetic images, testing was done on the real robotic arm and real objects.)

Objects Similar to Ones Trained on		
Tested on	Mean Error (cm)	Grasp-rate
Mugs	2.4	75%
Pens	0.9	100%
Wine Glass	1.2	100%
Books	2.9	75%
Eraser/ Cellphone	1.6	100%
Overall	1.80	90%

Novel Objects		
Tested on	Mean Error (cm)	Grasp-rate
Duct Tape	1.8	100%
Keys	1.0	100%
Markers/Screwdriver	1.1	100%
Toothbrush/Cutter	1.1	100%
Jug	1.7	75%
Translucent Box	3.1	75%
Powerhorn	3.6	50%
Coiled Wire	1.4	100%
Overall	1.85	87.5%

Note: Data from “Robotic grasping of novel objects” of the same author (2006) for stair 1. Stair 2 has a lower grasp success rate of 60-80%.

Maneuvering arm

- Sense environment to prevent collisions
- Search for a few known template structures

Arm path planning: PRM

- PRM: probabilistic roadmap
  - Sample random points in configuration space
  - Test if points are in free space
  - “Connect the dots” with a local planner to create a graph
  - Use a planner to search for a good solution in the graph

- “Probabilistically complete”
  - With enough samples chance of no result approaches zero
  - Not guaranteed optimal

Some movies

- Grasp multiple objects
- Open different doors
- Fetch object on verbal request

Conclusion

An autonomous robot can be build that

- manages to grab objects that it has not seen before
- manages to open doors that it has not seen before

Discussion

- Robot is very slow
- Of course: price, practical use, etc.

- Still very useful, practical research
- Can be of great help for disabled people

Questions?