Acting and Interacting in the Real World

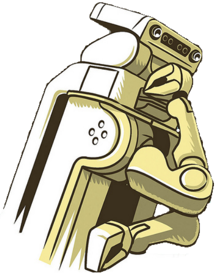
Jeannette Bohg, Niklas Bergström, Mårten Björkman and Danica Kragic

Centre for Autonomous Systems
Computer Vision and Active Perception Laboratory
KTH, Stockholm, Sweden

April 8, 2011

Introduction

- Robot is an intelligent agent with a physical embodiment
- can act in and interact with its environment
- 3D scene understanding necessary to plan and execute these actions



3D Sensing Devices

- Since the advent of the kinect, 3D sensing became cheap, fast and easy
- some constraints of other 3D sensing devices removed

	Stereo	Laser	Kinect
Price	-	-	+
Density	-	+	+
Resolution (Objects)	+	-	-
Range	o	+	-
Speed	+	-	+
SNR	o	o	+
Light Dependency	-	+	o

Figure: Quick and dirty comparison from own experiences in table top scenario:

What are the challenges?

- Basic challenges in processing and understanding of 3D data remain
  - high dimensionality of the data
  - noise
  - occlusions
  - from 3D points to semantics

What are the challenges?

- Basic challenges in processing and understanding of 3D data remain
  - **high dimensionality of the data**
  - **noise**
  - **occlusions**
  - from 3D points to semantics

- In this talk:
  1. **Segmentation**

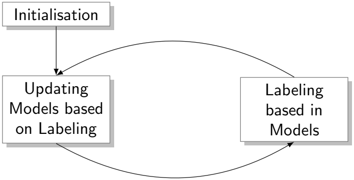
What are the challenges?

- Basic challenges in processing and understanding of 3D data remain
  - high dimensionality of the data
  - noise
  - **occlusions**
  - from 3D points to semantics

- In this talk:
  1. Segmentation
  2. **Completion of point clouds**

Active Real-Time 3D Segmentation

- *Active 3D scene segmentation and detection of unknown objects*, Björkman & Kragic, ICRA 2010
- Three hypotheses: foreground, background, flat surface
- Model parameters of hypotheses in RGB and disparity space
- iterative two-stage method
- 2D neighbourhood relationships in image exploited
- GPU and CPU implementation

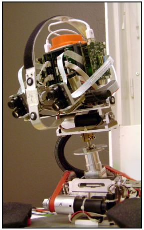


Example Run on KTH Vision System

Robust to noise and occlusions through

- multimodality
- strengthening of hypotheses over time

Initialisation from peaks in a saliency map



Extension to Multiple Foreground Hypotheses

Some top level information added by a user

- How many objects?
- Undersegmentation?

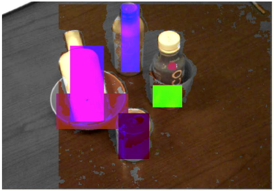
Submission to IROS 2011 by Bergström, Björkman and Kragic

Integration on PR2 and into ROS

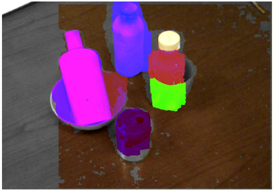
ROS packages

- active_realtime_segmentation (multiple FG hypotheses, GPU implementation)
- object_segmentation_gui (integration into interactive manipulation framework, user interaction, CPU version)
- rgbd_assembler (helper package to port RGB info from wide field to narrow field)

(a) PR2 Narrow Stereo



(b) Initialisation through User Input

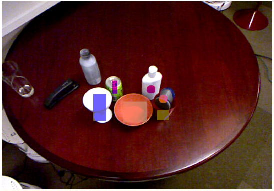


(c) Refined Segmentation

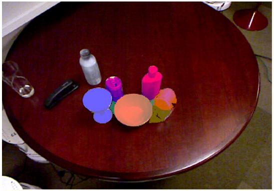
Integration on PR2 and into ROS

Using the Kinect mounted on head of PR2

(d) PR2 Kinect



(e) Initialisation through User Input



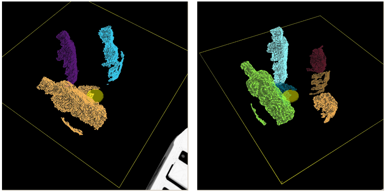
(f) Refined Segmentation

Challenges:

- Stereo
  - Porting of colour information inefficient
  - Bad RGB info with holes
- Kinect
  - Small resolution for individual objects
  - Calibration (Alignment RGB data and Disparity) imperfect

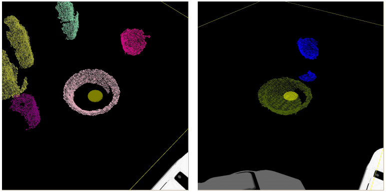
Advantages over Autonomous Segmentation

Better object separation in ambiguous situations



(g) Autonomous (h) Interactive

Object points close to the table plane will not be removed



(i) Autonomous (j) Interactive

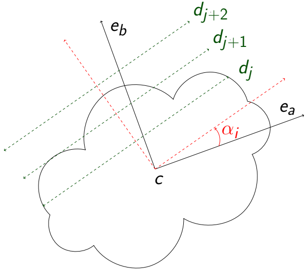
An Example Scene Reconstructed and Segmented

- Motion or grasp planning problematic in only partially known scene
- Human perform *controlled scene continuation* goverened by
  1. visual evidence
  2. completion rules gained through prior experience

Mind the Gap - Robotic Grasping under Incomplete Observation, Bohg et al, to appear in ICRA 2011

Symmetry Prior for Completing Unknown Object Shape

- Symmetry very common prior especially for man-made objects
- boostrapping search for correct symmetry by
  - planar symmetry (object model free)
  - assume tabe plane and restrict search to 3 DoF (position and orientation of a line)
  - use PCA for bootstrapping search
- plausability measure of a symmetry plane through visibility constraints



- e_a and e_b : eigenvectors of the projected point cloud
- c : center of mass of point cloud
- α_i : line orientation
- d_j to d_{j+2} : three line candidates

Meshing of Completed Point Cloud

- Useful for grasp planning and collision avoidance
- Poisson surface reconstruction
- Normal estimation from OpenCV

Validation of Completed Point Cloud against Ground Truth

- 14 data sets of objects each in 8 orientations
- Evaluation Measure: Geometric Mesh Deviation
- Baseline: Meshing only using measured points

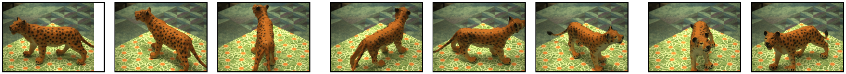
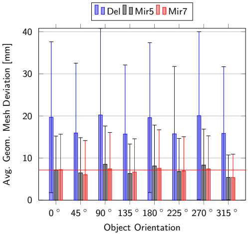


Figure: One of the Datasets shown in Orientation 0° , 45° , 90° , 135° , 180° , 225° , 270° and 335° .



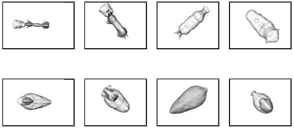
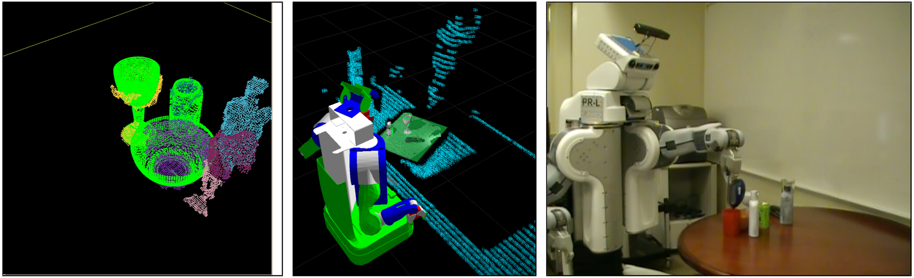


Figure: Meshes based on Mirroring. First Row: Toy Tiger. Second Row: Rubber Duck.

Future Work

- Predicted points are assumed to be as valid as reconstructed points
- For each point and for whole completed point cloud, **plausibility score can be computed based on visbility**
- weighting points dependent on that for different tasks



(a) Object Recognition and Pose Estimation (b) Collision Map Processing (c) Grasping

Conclusions

- Although 3D sensing became cheap, fast and easy, basic challenges in 3D data processing and understanding remain
    - dimensionality
    - noise
    - occlusions
    - from low level 3D points to semantics
- Presented 2 methods, segmentation and point cloud completion
- Deal with the first three problems to bring us closer to the last, semantic annotation of point clouds