COMBINING 2D AND 3D OBJECT CATEGORIZATION FOR TASK CONSTRAINED GRASPING

Madry M., Song D., Kragic D. - Computer Vision & Active Perception Lab, KTH, Stockholm, Sweden
{madry,dsong,danik}@csc.kth.se

Abstract

We present a system able to transfer grasp knowledge between object categories defined by geometric properties and functionality. In the center lies an Object Categorization Module (OCM) based on 2D and 3D visual data that is integrated with a grasp planner. The system runs on a robot (ARMAR-III) equipped with active stereo cameras. The experimental evaluation compares individual 2D and 3D categorization with the fused 2D-3D OCM, and shows the usefulness of the approach in task-based grasping.

Motivation

- How to grasp an object?

- Humans classify objects according to the task they afford *[Greeno'94]*
- Grasp knowledge can be transferred between objects that belong to the same category
- Fusion of 2D and 3D may provide more robust system

Goal

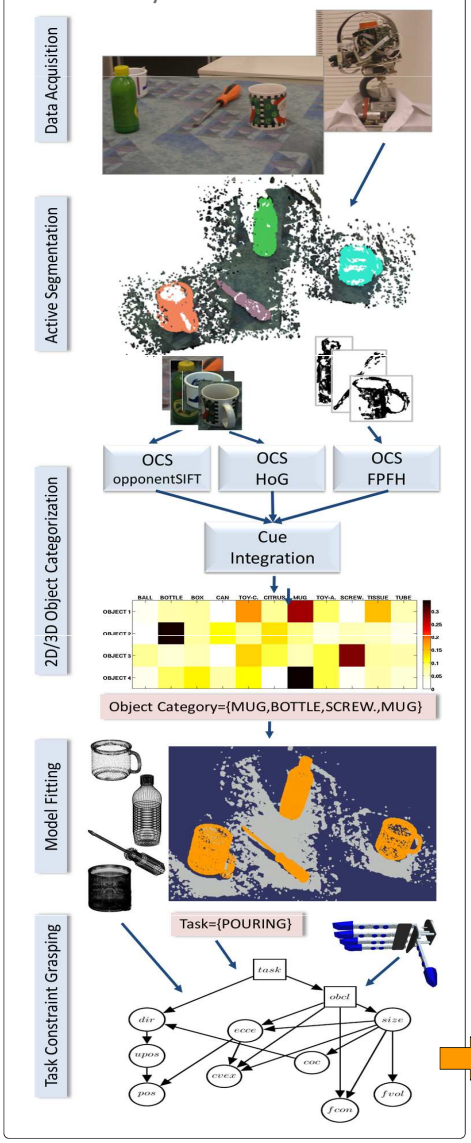
- **Where to grasp to pour a liquid?**
- **Goal:** Finding grasp points for a desired task in a natural scene
  - Known: Embodiment (hand) & Task
  - Unknown: Scene content

➪ **Mid-goal:**
Finding object categories

Contributions

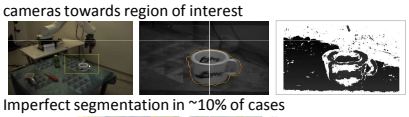
- Object Categorization (OCM):
  - Evaluation of several 2D and 3D appearance, color and shape descriptors on real stereo data
  - Fusion of 2D and 3D object categorization with high categorization rate (up to 92% for 11 object categories)
- Task-based Grasping:
  - Robot can choose objects that afford a desired task and plan a grasp that satisfies constraints imposed by the task
- Integration of the 2D-3D OCM with the active segmentation *[Bjorkman'10]* and the probabilistic grasp reasoning system *[Song'10]*

System Overview

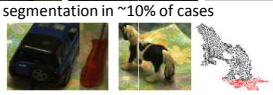


Active Segmentation

- Attention mechanisms in the peripheral view direct the foveal cameras towards region of interest



- Imperfect segmentation in ~10% of cases



2D/3D Object Categorization

- **Object Representation** using several 2D and 3D descriptors encode different object properties: appearance (SIFT), color (opponentSIFT), 2D shape (HoG), 3D shape (FPFH)
- Individual OCM for each descriptor:
  - Spatial pyramid for 2D; Bag-of-words for 3D
  - Classification: SVMs with a χ^2 – kernel
- **2D/3D Cue Integration:**
  - Confidance measure: normalized distance of a sample to the hyperplane
  - Fusion of evidences from the individual OCMs at the high level
  - Evaluation of the linear (weighted sum, product, max rule) and nonlinear (SVMs with RBF and χ^2 – kernel) integration methods

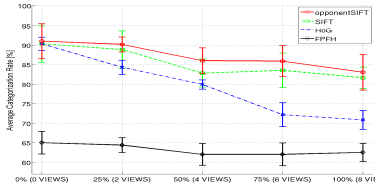
110 Object Stereo Database

- 11 object categories x 10 object intances per category
- Data: 2D (RGB image) and 3D (point cloud)

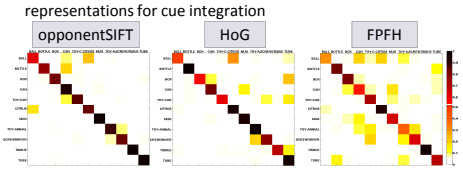
- Data collected in 16 views around an object (every 22,5°)
- Experimental Setup: 8 views per object for traning and testing
  - *Setup-50*: in the test set 50% of data from unseen viewpoint

Experimental Evaluation

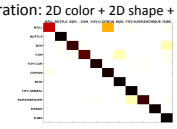
- Performance of 2D and 3D representations under varying viewpoint condition



- Confusion matrices (for *Setup-50*)– complementarity of representations for cue integration



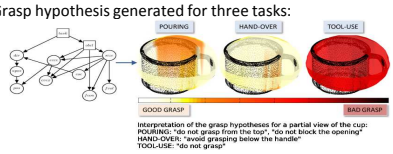
- 2D/3D Cue Integration: 2D color + 2D shape + 3D shape descriptors



- 2D/3D OCM significantly outperforms the best individual OCM (by 6%)

Task Constraint Grasping

- Training of Bayesian Network on the synthetic 3D object models from the Princton Shape Benchmark
- Grasp hypothesis generated for three tasks:



Where to grasp to perform the desired task?

Video: www.csc.kth.se/~madry/research/task_constrained_grasping