Identifying visual structures for predicting action affordances

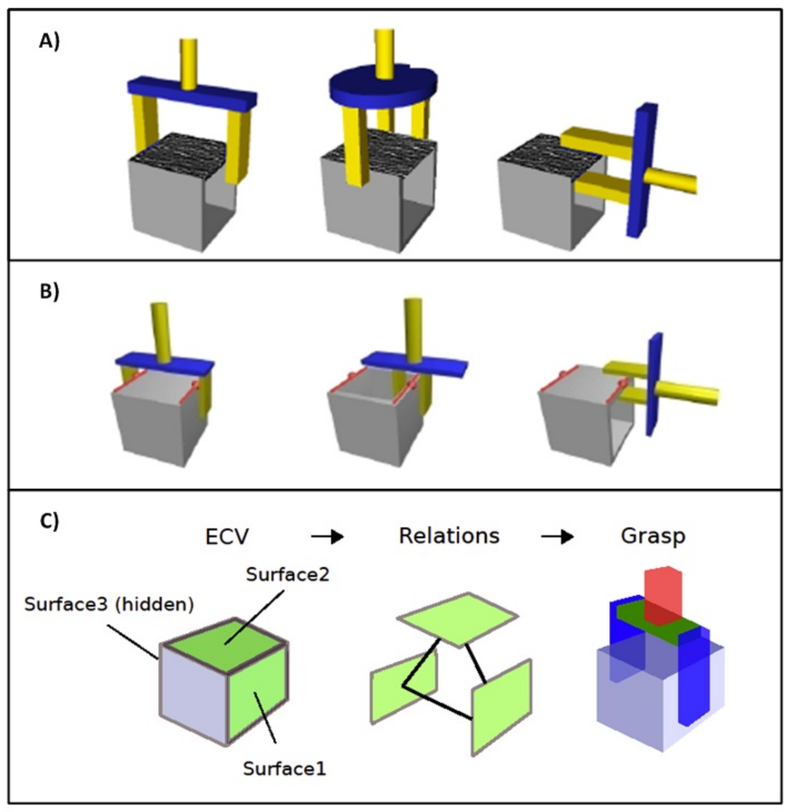
Mikkel Tang Thomsen (mtt@mmmi.sdu.dk)
Maersk McKinney Moller Institute, University of Southern Denmark

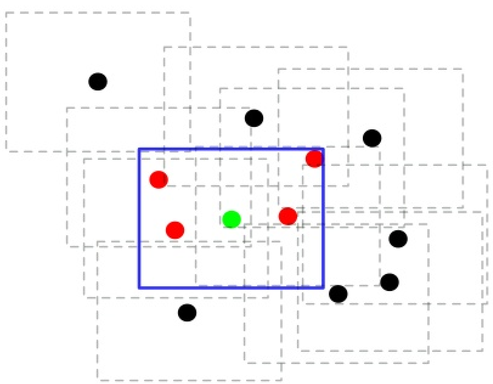
Introduction

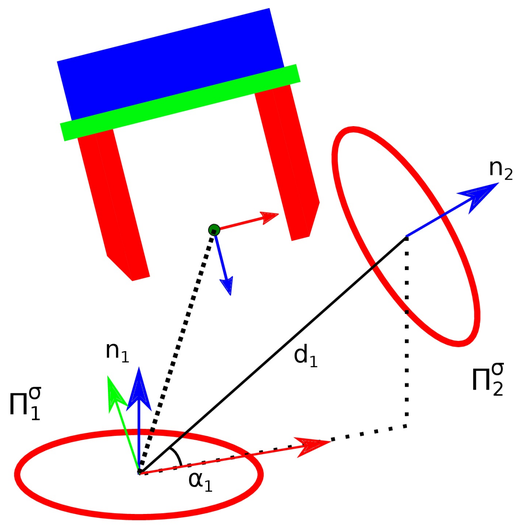
The ability to grasp unknown objects is an essential skill for autonomous robotic systems for interacting within an unconstrained environment. In this work,an approach towards identifying reliable grasping actions based on multiple visually extracted surface features is proposed.
The approach takes it origin in the work performed by Kootstra et al. in [1] where it was proposed that by utilising the 3D visual features of contours and surfaces from the Early Cognitive Vision (ECV) system [2], it was possible to manually construct grasping actions which yield a good probability of grasping success.
Here we propose to extend this approach by looking at higher order feature constellations and by replacing the manually defined relations with a simple learning scheme for learning such relations. See figure to the right.

[1] G. Kootstra, M. Popovic , J. A. Jørgensen, K. Kuklinski, K. Miatliuk, D. Kragic, and N. Krüger, "Enabling grasping of unknown objects through a synergistic use of edge and surface information," IROS2012

[2] N. Pugeault, F. Wörgötter, and N. Krüger, "Visual primitives: Local, condensed, semantically rich visual descriptors and their applications in robotics," I. J. Humanoid Robotics, vol. 7, no. 3, pp. 379–405, 2010.







Approach

The basic principle behind this approach is to combine the feature space, represented with features from the Early Cognitive Vision (ECV) system and the grasping space into an Perception x Action space. In this cross-space, we want to find feature constelations, that are predictive for grasping. On the feature side, we rely on the introduction of feature relations. Feature relations are 1st, 2nd or 3rd order combinations of visual features described by their spatial relationship. In this space we perform a neighbourhood analysis to find particular predictive combinations.

$$\rho = \{SE(3)^A_P, \alpha_1, \alpha_2, \alpha_3, d_1, E\}$$

Simulated environment for action evaluation and feature extraction utilising simulated RGB-D sensors and dynamics simulation

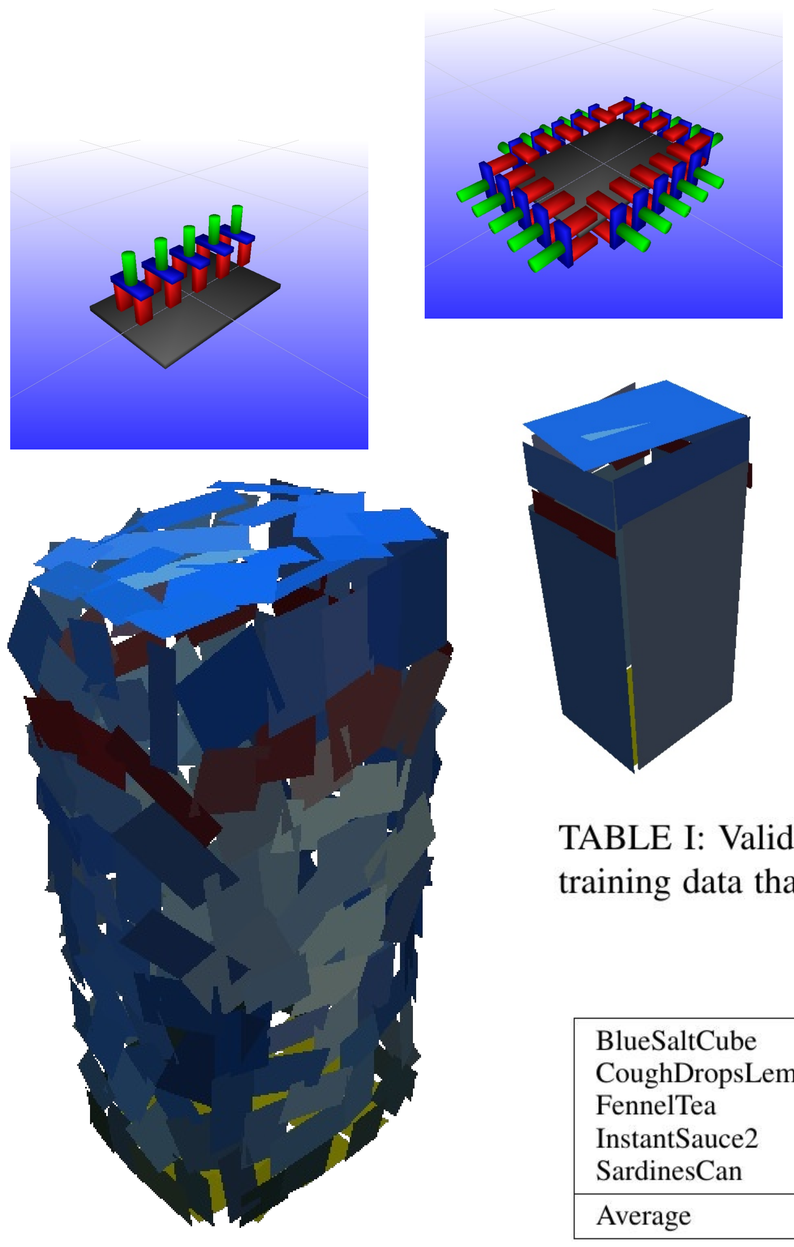


TABLE I: Validation results. T meaning three feature combinations. S meaning single feature. $^{0.50}_{10}$ meaning that we use the training data that have achieved a probability of over 0.50 and over 10 similar particles are within the identification window.

	Probability for a successful grasp (no. grasps)							
	$T^{0.50}_{10}$	$T^{0.75}_{10}$	$T^{0.50}_{25}$	$T^{0.75}_{25}$	$S^{0.50}_{10}$	$S^{0.75}_{10}$	$S^{0.50}_{25}$	$S^{0.75}_{25}$
BlueSaltCube	0.48 (13,569)	0.63 (3,525)	0.43 (5,406)	0.67 (817)	0.59 (431)	0.51 (136)	0.79 (256)	0.91 (54)
CoughDropsLemon	0.74 (1,152)	0.80 (495)	0.85 (326)	0.89 (143)	0.46 (615)	0.53 (223)	0.49 (221)	0.52 (50)
FennelTea	0.83 (3,727)	0.90 (1,408)	0.81 (1,045)	0.94 (242)	0.50 (575)	0.36 (244)	0.61 (278)	0.44 (54)
InstantSauce2	0.92 (1,167)	0.93 (497)	0.96 (358)	1.00 (114)	0.61 (249)	0.79 (61)	0.67 (159)	0.93 (30)
SardinesCan	0.74 (605)	0.82 (393)	0.88 (144)	0.97 (77)	0.97 (303)	0.98 (75)	0.98 (85)	1.00 (7)
Average	0.74	0.82	0.79	0.89	0.62	0.63	0.71	0.76

Thomsen, M. T., Bodenhagen, L. & Krüger, N (2013). Statistical identification of composed visual features indicating high-likelihood of grasp success. In ICRA 2013 Workshop 'Bootstrapping Structural Knowledge from Sensory-motor Experience

Acknowledgement

This work has been funded by the EU project Xperience (FP7-ICT-270273).

University of Southern Denmark