Hochschule
Bonn-Rhein-Sieg
University of Applied Sciences

Bonn-Aachen
International Center for
Information Technology

Research and Development Project

Learning Grasp Evaluation Models Using Synthetic 3D Object-Grasp Representations

February 26, 2019

Minh Nguyen

Use Case

Robocup@Home

Problem

Make Lucy grasp robustly

Figure 1: Human Support Robot from Toyota.

Figure 2: Typical objects in the Robocup@Home competition [8].

Motivation

Current Solution - MoveIt grasp planner

A randomized planner will do inexplicable things!

Figure 3: MoveIt! failing to plan a simple grasp.

Robotic Grasping Overview

Generating Grasp Hypotheses

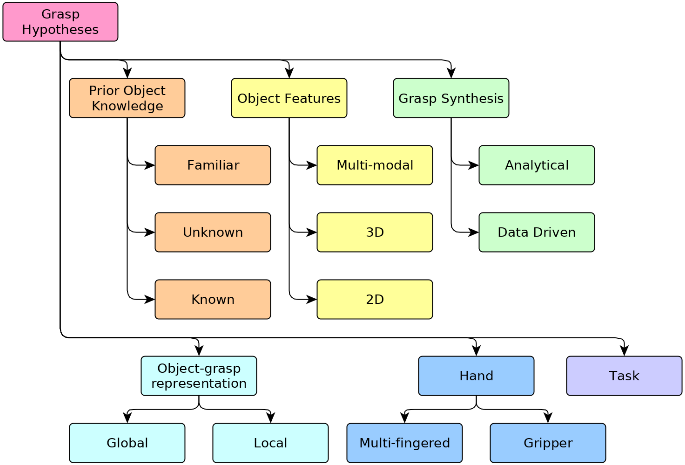


Figure 4: Aspects which may influence generation of grasp hypotheses [1].

Robotic Grasping Overview

Generating Grasp Hypotheses

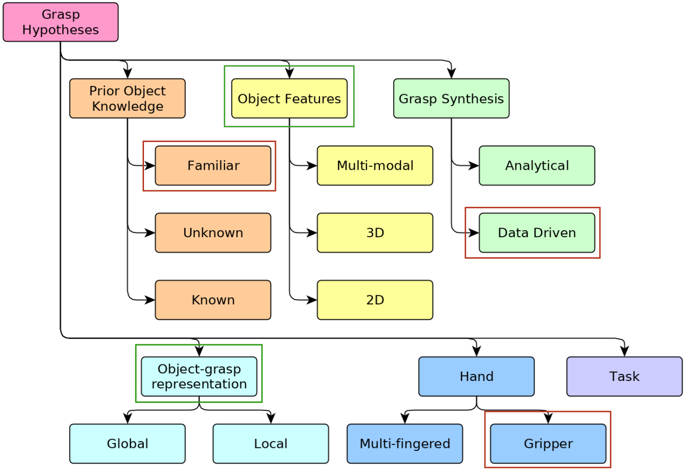


Figure 4: Aspects which may influence generation of grasp hypotheses [1].

Analytical versus Empirical

Why not analytical grasp synthesis?

Quick review

- **Analytical grasp synthesis** consider the mechanical properties of the contact points [10, 11]
- **Data-driven methods** rely on some form of grasp experience: *human demonstration*, *labeled data*, or *trial-error* [1]

Analytical versus Empirical

Why not analytical grasp synthesis?

Quick review

- **Analytical grasp synthesis** consider the mechanical properties of the contact points [10, 11]
- **Data-driven methods** rely on some form of grasp experience: *human demonstration*, *labeled data*, or *trial-error* [1]

Shortcomings - Analytical Grasp Synthesis

- Rely on knowledge of the object's geometry, which can be inaccurate or unavailable in real environments
- Best grasps according to analytical metrics may not be most robust grasps [13]

Empirical Grasp Synthesis

Shortcomings - Empirical Grasp Synthesis

- Data collection is costly and time consuming!
- However, it is possible to synthesize data for training a grasp evaluation model.

Empirical Grasp Synthesis

Shortcomings - Empirical Grasp Synthesis

- Data collection is costly and time consuming!
- However, it is possible to synthesize data for training a grasp evaluation model.

Aspects important to training a grasp evaluation model

- **Object-grasp representation**: how to capture the gripper-object relation in perceptual data
- **Feature extraction and learning method**: how to use the captured representation so the model can learn efficiently, and which model will learn well
- **Dataset generation**: how to synthesize and label data

Empirical Grasp Synthesis

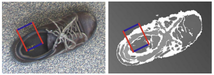
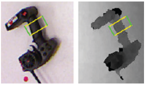
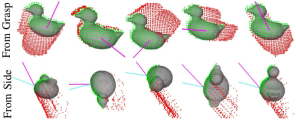
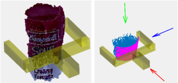
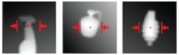
Method	Object-grasp representation	Feature extraction & learning model	Data generation
[3]		Histogram of hand-crafted filters; Model: SVM.	Rectangles manually annotated.
[5]		Auto-encoders to initialize weights, structured regularization to combine depth and RGB data; Model: MLP.	Extension of the dataset from [3] (above).
[4]		RGB rendering of “template grids”; Model: LeNet CNN	Quality of grasps are calculated in simulation for object meshes, verified via crow-sourcing.
[2]		Filters of cuboid regions projected onto 3 orthogonal planes, creating 15 channels; Model: LeNet CNN.	Quality of grasps are calculated for object meshes using force-closure
[7]		Depth images cropped and aligned to gripper; Model: CNN combined with single-layer NN.	Quality of grasps are calculated for object meshes using a variant of ϵ-metric from [13]

Table 1: Five recent empirical approaches to grasp quality prediction which synthesize data

Dex-Net 2.0 and GQCNN

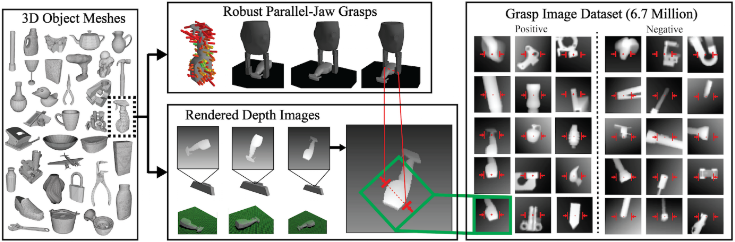


Figure 5: Dex-Net 2.0 pipeline for dataset generation [7]

- Labeling is done by applying analytical grasp metrics to candidates in simulation
- Generated data is used to train a Grasp Quality Convolutional Neural Network (GQ-CNN) to predict grasp quality and distance from gripper

1. Introduction

2. Methodology

3. Experiments

Grasping software pipeline

Object detection – Previous implementation

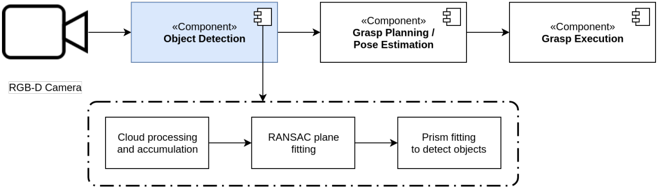


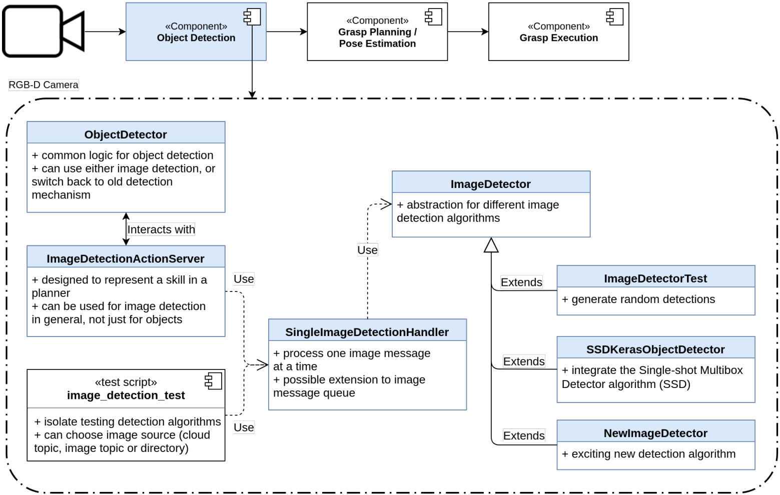
Figure 6: Previous pipeline for object grasping.

Challenges

- New robot with closed-source components
- Previous object detection algorithm was unreliable and not suited for domestic objects

Grasping software pipeline

Object detection – New architecture



Grasping software pipeline

Object detection – Single Shot MultiBox Detector

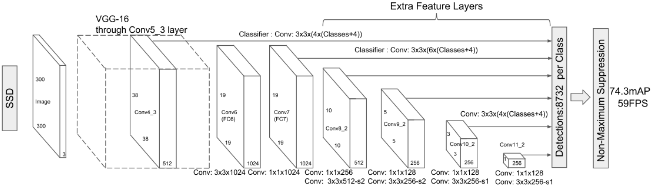
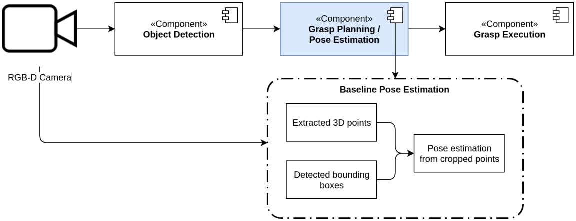


Figure 7: SSD architecture [6].

- learns to adjust a set of default boxes to the ground truth bounding boxes
- learning at different scales is simulated by training different resolutions of the feature map

Grasping software pipeline

Grasp planning – Object pose estimation



$$p = \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} \min_{i=1}^{M} B_{i,1} \\ \frac{1}{M} \sum_{i=1}^{M} B_{i,2} \\ \frac{1}{M} \sum_{i=1}^{M} B_{i,3} \end{pmatrix} \quad (1)$$

$$p_j = \frac{1}{M} \sum_{i=1}^{M} B_{i,j} \quad (2)$$

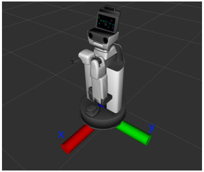
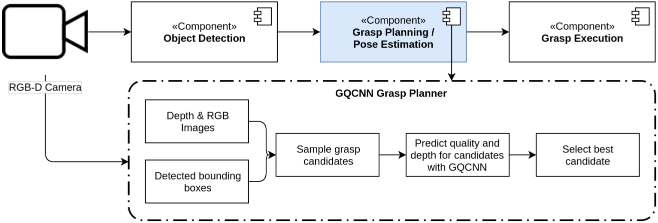


Figure 8: `base_link` coordinate frame.

Grasping software pipeline

Grasp planning – GQCNN



Issues

- Assumption of top-down camera view makes the depth prediction inaccurate
- Planned grasps take too long to calculate and are highly variant in depth

Grasping software pipeline

Grasp planning – GQCNN

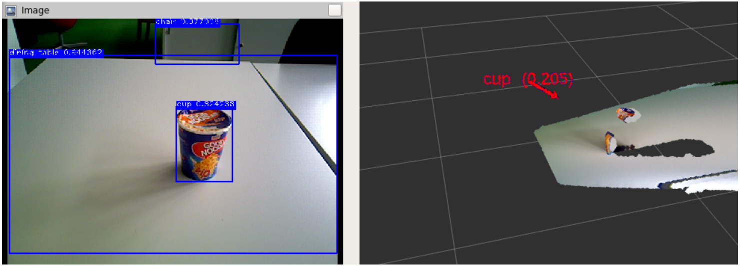
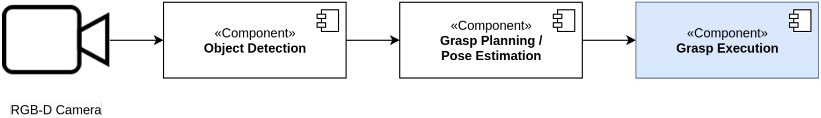


Figure 9: A successful GQCNN grasp plan which took 40 seconds. Arrow and number on the right indicate grasp pose and quality returned from the GQCNN planner.

Grasping software pipeline

Grasp execution – Dynamic Motion Primitives (DMP)



Manipulation planning uses the integration and extension of the DMP in a parallel project [9]

$$\tau \ddot{\mathbf{y}} = \alpha(\beta(\mathbf{g} - \mathbf{y}) - \dot{\mathbf{y}}) + \mathbf{f}$$

$$f_j(t) = \frac{\sum_{i=1}^{N} \Psi_{i,j}(t) w_{i,j}}{\sum_{i=1}^{N} \Psi_{i,j}(t)}$$

Experiments

Setup

- 20 grasp attempts for each variance of the baseline method
- Predefined starting position for each attempt as seen in figure 10
- Arm collisions and object slips are counted as failures
- Only one of the objects in figure 11 (next slide) is grasped at a time

Figure 10: Robot at initial position.

Experiments

Objects

(a) Styrofoam ball (b) Duct tape (c) Noodles box (d) Salt container

Figure 11: Objects selected for the experiments.

Experiments

Results

Object	Mean x		Minimum x	
	Success	Failure	Success	Failure
Salt	17	3	16	4
Ball	8	12	15	5
Noodle box	16	4	12	8
Duct tape	7	13	13	7

Table 2: Counts of successful and failed grasp attempts during the experiments. “Mean x” and “Minimum x” refer to the two baseline pose estimation methods which, respectivly, use the mean and minimum object coordinates along the x-axis to estimate the grasp pose [1].

[1] A video of a complete experiment for one object is available on YouTube at `https://youtu.be/OC7vttt4-Jo`

Experiments

Discussion

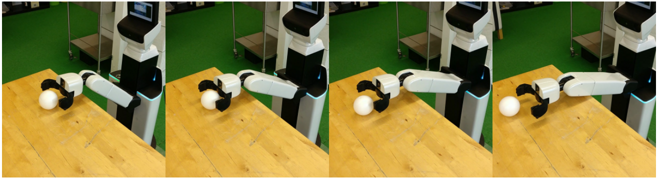


Figure 12: The gripper pushes on the ball and it rolls forward

- Grasping the ball using the Minimum x method is better since the arm doesn't push it forward
- Low, hollow objects like duct tape give low z estimates, which suggests that a top-down grasp may perform better
- Slippage can be avoided by utilizing a different grasp mechanism (in hindsight)

Conclusions

Contributions

- Design and implementation of an object detection architecture which is compatible with old methods while flexible to new detection algorithms
- Implementation of a grasping pipeline (in conjunction with [9]) reliable enough to perform hundreds of grasps in succession, enabling evaluation of more advanced approaches in the future

Conclusions

Contributions

- Design and implementation of an object detection architecture which is compatible with old methods while flexible to new detection algorithms
- Implementation of a grasping pipeline (in conjunction with [9]) reliable enough to perform hundreds of grasps in succession, enabling evaluation of more advanced approaches in the future

Future works

- More advanced and reliable methods can be experimented instead of GQCNN
- Incorporating the force sensor and hand camera would provide much more information for grasp planning, and may produce more robust grasps

Questions

References (1/2)

J. Bohg et al. “Data-Driven Grasp Synthesis: A Survey”. In: **IEEE Transactions on Robotics** 30.2 (Apr. 2, 2014), pp. 289–309. ISSN: 1552-3098. DOI: 10.1109/TRO.2013.2289018.

Marcus Gualtieri et al. “High precision grasp pose detection in dense clutter”. In: **2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)**. Ed. by Wolfram Burgard et al. Oct. 2016, pp. 598–605. DOI: 10.1109/IROS.2016.7759114.

Yun Jiang, S. Moseson, and A. Saxena. “Efficient grasping from RGBD images: Learning using a new rectangle representation”. In: **IEEE International Conference on Robotics and Automation (ICRA)**. Ed. by Antonio Bicchi. May 1, 2011, pp. 3304–3311. DOI: 10.1109/ICRA.2011.5980145.

D. Kappler, J. Bohg, and S. Schaal. “Leveraging big data for grasp planning”. In: **IEEE International Conference on Robotics and Automation (ICRA)**. Ed. by Allison Okamura. May 2015, pp. 4304–4311. DOI: 10.1109/ICRA.2015.7139793.

Ian Lenz, Honglak Lee, and Ashutosh Saxena. “Deep learning for detecting robotic grasps”. In: **The International Journal of Robotics Research** 34.4-5 (Mar. 16, 2015), pp. 705–724. DOI: 10.1177/0278364914549607. eprint: https://doi.org/10.1177/0278364914549607.

Wei Liu et al. “SSD: Single Shot MultiBox Detector”. In: **Computer Vision – ECCV 2016**. Ed. by Bastian Leibe et al. Cham: Springer International Publishing, Sept. 17, 2016, pp. 21–37. ISBN: 978-3-319-46448-0.

Jeffrey Mahler et al. “Dex-Net 2.0: Deep Learning to Plan Robust Grasps with Synthetic Point Clouds and Analytic Grasp Metrics”. In: **Robotics: Science and Systems (RSS)** abs/1703.09312 (Mar. 1, 2017). arXiv: 1703.09312. URL: ht /abs/1703.09312.

References (2/2)

Mauricio Matamoros et al. **RoboCup@Home 2018: Rules and Regulations**. http://www.robocupathome.org/rules/2018_rulebook.pdf. June 4, 2018.

Abhishek Padalkar. **Dynamic Motion Primitives**. Tech. rep. Hochschule Bonn-Rhein-Sieg, Aug. 2018. Unpublished.

Máximo A. Roa and Raúl Suárez. "Grasp Quality Measures: Review and Performance". In: **Autonomous Robots** 38.1 (Jan. 2015), pp. 65–88. ISSN: 0929-5593. DOI: 10.1007/s10514-014-9402-3.

A. Sahbani, S. El-Khoury, and P. Bidaud. "An Overview of 3D Object Grasp Synthesis Algorithms". In: **Robotics and Autonomous Systems** 60.3 (Mar. 1, 2012), pp. 326–336. ISSN: 0921-8890. DOI: 10.1016/j.robot.2011.07.016.

Artur Saudabayev et al. "Human grasping database for activities of daily living with depth, color and kinematic data streams". In: **Scientific Data** 5 (May 29, 2018). Data Descriptor, p. 180101.

J. Weisz and P. K. Allen. "Pose error robust grasping from contact wrench space metrics". In: **2012 IEEE International Conference on Robotics and Automation**. Ed. by Antonio Bicchi. May 2012, pp. 557–562. DOI: 10.1109/ICRA.2012.6224697.