Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

MIN Faculty
Department of Informatics

The Crucial Components to Solve the Picking Problem

Benjamin Scholz

University of Hamburg
Faculty of Mathematics, Informatics and Natural Sciences
Department of Informatics
Technical Aspects of Multimodal Systems

13. November 2017

Contents

Motivation

- **Problem**: How do we use a robotic arm to pick objects?
- Universal importance in robotics
- **Examples**:
  - Manufacturing
  - Warehouses [1]
  - Household tasks [2]

Figure: Retrieved from
`https://techxplore.com/news/2017-04-pieces-unveiling-rightpick.html`.

Motivation

- Amazon Picking Challenge held yearly since 2015
- Picking and stowing
- Scoring system
- Tasks get more difficult every year

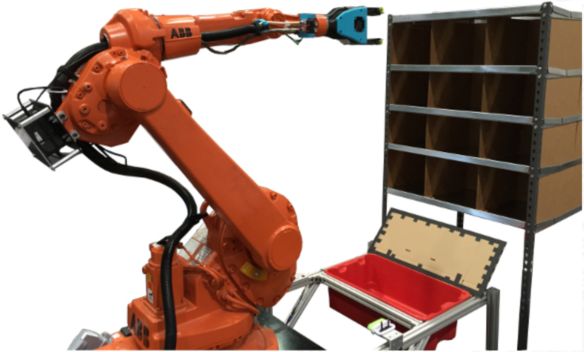

Figure: Retrieved from
https://awl2016.mit.edu/sites/default/files/images/apc16.jpg.

Motivation

- Crucial components in picking objects:

1. Hardware, especially end-effectors
2. Motion Planning
3. Object Recognition
4. Grasping

End-effectors

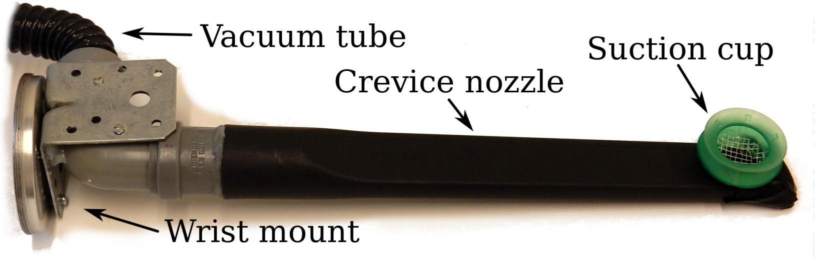

Figure: Example of an end-effector that uses a suction cup, to pick objects [1].

End-effectors

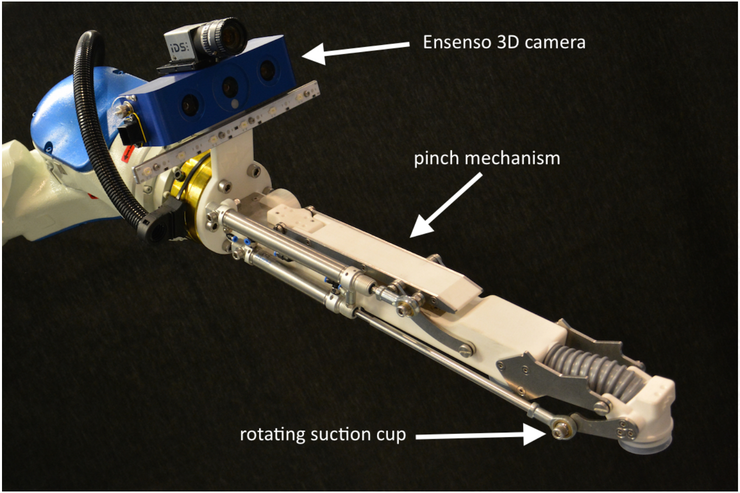

Figure: Example of an end-effector using a pinch mechanism and a suction cup [3].

End-effectors

Figure: A parallel gripper. Retrieved from https://blog.robotiq.com/grippers-collaborative-robots (last checked 01.11.2017)

Figure: A 3-finger gripper. Retrieved from https://robotiq.com/products/3-finger-adaptive-robot-gripper (last checked 01.11.2017)

Motion Planning

- Planning vs. Feedback [4]
- Path Planning:
  - Modeling the entire environment
  - Searching in world model for solution
  - High computation costs
- Feedback:
  - Reacting to physical interactions
  - No model necessary

Motion Planning - RRT-Connect

- Commonly used approach to path planning: RRT-Connect [5]
- **R**apidly-Exploring **R**andom **T**rees
- Connect two trees that originate from start and goal using the following steps:

1. Draw random sample from search space
2. Find nearest node in tree
3. Try to extend tree in direction of sample
4. Test for collisions
5. Try to connect new node to other tree

Motion Planning - RRT-Connect

Figure: Example of RRT-Connect. Retrieved from `http://www.kuffner.org/james/plan/algorithm.php`, last checked on 08.11.2017.

Object Recognition

- To pick the correct object, **class** needs to be known
- To be able to grasp the object **pose** needs to be known
- Two common approaches:
  - LINEMOD [6]
  - Object detection and matching to model

LINEMOD

- Using template matching to detect objects
- Template has to use sensible features:
  - Orientation of the gradient (images)
  - Surface normals (depth data)
- Sample only discriminative gradients

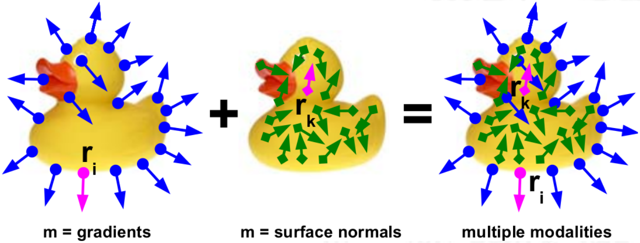

Figure: The two features used for LINEMOD [7].

Similarity Measurement

$$\mathcal{E}(\mathcal{I},\mathcal{T},c) = \sum_{r\in\mathcal{P}} \left(\max_{t\in\mathcal{R}(c+r)} \left|\cos\left(ori(\mathcal{O},r) - ori(\mathcal{I},t)\right)\right| \right)$$

- $ori(\mathcal{O},r) - ori(\mathcal{I},t)$ difference of gradient orientations
- $|\cos()|$ for background invariance
- $\max_{t\in\mathcal{R}(c+r)}$ to find most similar gradient orientation nearby

LINEMOD - Surface Normals

- Kinect provides depth data
- Surface normals as similarity measurement
- Summing gradient orientation and surface normals gives final result

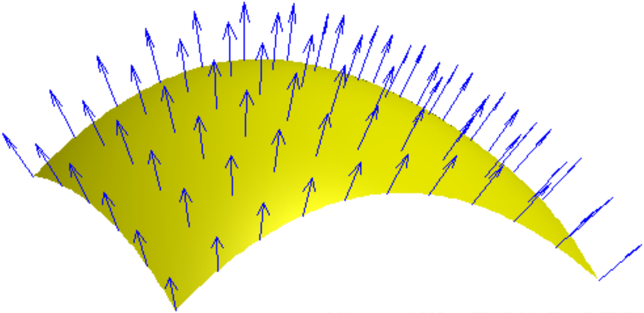

Figure: Surface normals. Retrieved from `https://commons.wikimedia.org/wiki/File:Surface_normal.png`.

LINEMOD - Template Creation

- Many pictures needed for template creation
- Solution: Use a 3D model
- Automates template creation

LINEMOD - Template Creation

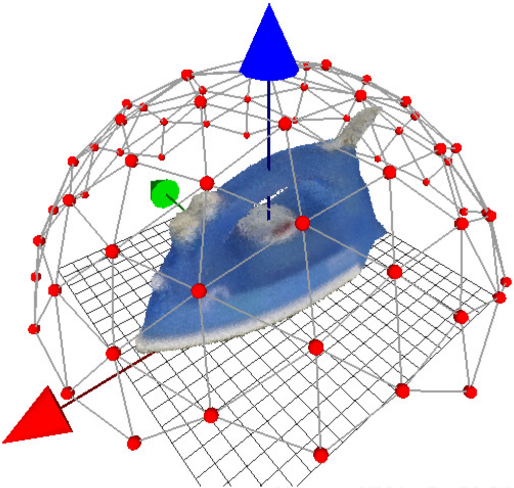

Figure: Creating templates of the iron. Each red vertex is the center of a camera used to make pictures [6].

LINEMOD - Pose Detection

- Infer approximate pose from matched template
- **Drawback:** Often inaccurate
- **But:** a rough pose estimation can help other algorithms to get accurate position

Object Recognition - Object Detection and Matching

- Method used by Amazon Picking Challenge 2016 winner [3]:

1. Find objects using R-CNNs [8]
2. Create bounding box for point cloud
3. Match the 3D model to the point cloud using Super4PCS [9]

Object Recognition - R-CNNs

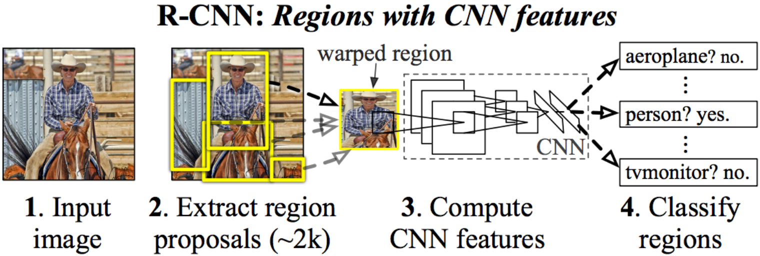

Figure: Object Detection using R-CNNs [8].

- Provide **Class** and **Region**
- **Region** used to create **bounding box** around point cloud of object

Object Recognition - ICP vs. Super4PCS

- Matching point clouds
- **I**terative **C**losest **P**oint
  - Good initialization needed
  - Refine LINEMODs pose estimation
- **Super 4-P**oints **C**ongruent **S**ets
  - Works without good initialization
  - Region without pose is enough

Figure: Using ICP to match point clouds. Retrieved from `https://www.youtube.com/watch?v=uzOCS_gdZuM`

Grasping

- There are three common data-driven approaches to learn grasps for known objects [10]:

1. Using 3D models
2. Learning from humans
3. Learning through trial and error

Grasping - Using 3D Models

- The approach using 3D models is most convenient
- Pre-compute grasps
- Use metric to judge their quality
- Known object pose let's us filter for possible grasps

Grasping - Learning From Humans

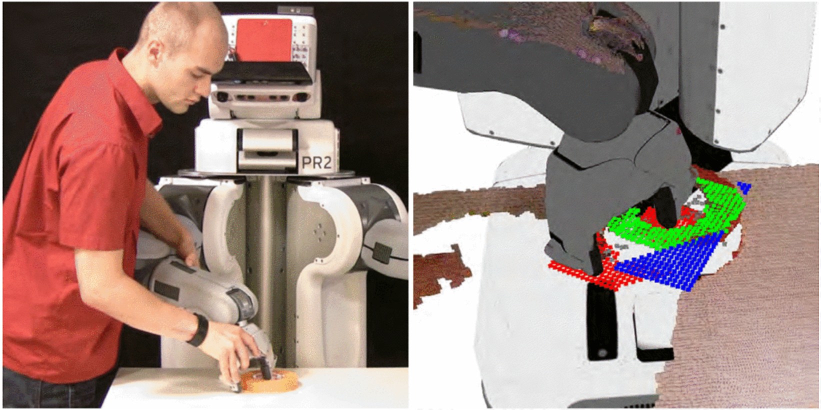

Figure: A PR2 learning to grasp objects from human demonstration [11].

Grasping - Learning From Trial and Error

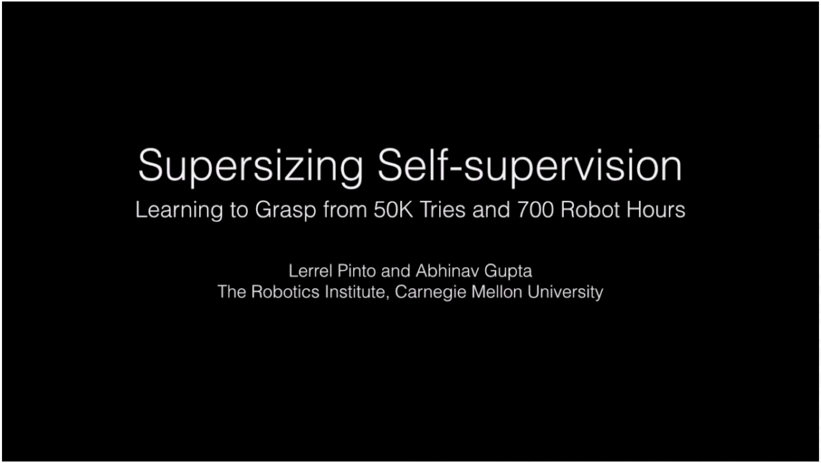

Figure: Video Retrieved from `https://www.youtube.com/watch?v=oSqHc0nLkm8`.

Conclusion - Amazon Picking Challenge Progress

- Average performance rising
- Even winners fail to perform task perfectly
- Robots are still much slower than humans

Conclusion

- **End-effector**:
  - Suction cup + gripper can handle large variety of objects
- **Motion Planning**:
  - Both **feedback** and **planning** have advantages and disadvantages
- **Object Recognition**:
  - LINEMOD is easy to use
  - At competitions approaches using real world images faired better
- **Grasping**:
  - Using 3D models easiest to use
  - Promising results in simulations do not always hold in real world

References

[1] N. Correll, K. E. Bekris, D. Berenson, O. Brock, A. Causo, K. Hauser, K. Okada, A. Rodriguez, J. M. Romano, and P. R. Wurman, "Analysis and observations from the first amazon picking challenge," *IEEE Transactions on Automation Science and Engineering*, 2016.

[2] A. Saxena, J. Driemeyer, and A. Y. Ng, "Robotic grasping of novel objects using vision," *Int. J. Rob. Res.*, vol. 27, pp. 157–173, Feb. 2008.

[3] C. Hernandez, M. Bharatheesha, W. Ko, H. Gaiser, J. Tan, K. van Deurzen, M. de Vries, B. Van Mil, J. van Egmond, R. Burger, *et al.*, "Team delft's robot winner of the amazon picking challenge 2016," *arXiv preprint arXiv:1610.05514*, 2016.

References (cont.)

[4] R. J. R. M.-M. A. S. V. W. O. B. Clemens Eppner, Sebastian Höfer, "Lessons from the amazon picking challenge: Four aspects of building robotic systems," in *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17*, pp. 4831–4835, 2017.

[5] J. J. Kuffner and S. M. LaValle, "Rrt-connect: An efficient approach to single-query path planning," in *Robotics and Automation, 2000. Proceedings. ICRA'00. IEEE International Conference on*, vol. 2, pp. 995–1001, IEEE, 2000.

[6] S. Hinterstoisser, V. Lepetit, S. Ilic, S. Holzer, G. Bradski, K. Konolige, and N. Navab, *Model Based Training, Detection and Pose Estimation of Texture-Less 3D Objects in Heavily Cluttered Scenes*, pp. 548–562.
Berlin, Heidelberg: Springer Berlin Heidelberg, 2013.

References (cont.)

[7] S. Hinterstoisser, C. Cagniart, S. Ilic, P. Sturm, N. Navab, P. Fua, and V. Lepetit, "Gradient response maps for real-time detection of textureless objects," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 5, pp. 876–888, 2012.

[8] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 580–587, 2014.

[9] N. Mellado, D. Aiger, and N. J. Mitra, "Super 4pcs fast global pointcloud registration via smart indexing," *Computer Graphics Forum*, vol. 33, no. 5, pp. 205–215, 2014.

References (cont.)

[10] J. Bohg, A. Morales, T. Asfour, and D. Kragic, "Data-driven grasp synthesis—a survey," *IEEE Transactions on Robotics*, vol. 30, no. 2, pp. 289–309, 2014.

[11] A. Herzog, P. Pastor, M. Kalakrishnan, L. Righetti, T. Asfour, and S. Schaal, "Template-based learning of grasp selection," in *Robotics and Automation (ICRA), 2012 IEEE International Conference on*, pp. 2379–2384, IEEE, 2012.