

From Visual Understanding to Complex Object Manipulation

Judith Bütepage, Silvia Cruciani, Mia Kokic,
Michael Welle, and Danica Kragic

Department of Robotics, Perception, and Learning, KTH Royal Institute of Technology,
100 44 Stockholm, Sweden; email: butipage@kth.se, cruciani@kth.se, mkokic@kth.se,
mwelle@kth.se, dani@kth.se

Annu. Rev. Control Robot. Auton. Syst. 2019.
2:161–79

First published as a Review in Advance on
October 4, 2018

The *Annual Review of Control, Robotics, and
Autonomous Systems* is online at
control.annualreviews.org

<https://doi.org/10.1146/annurev-control-053018-023735>

Copyright © 2019 by Annual Reviews.
All rights reserved

**ANNUAL
REVIEWS CONNECT**

www.annualreviews.org

- Download figures
- Navigate cited references
- Keyword search
- Explore related articles
- Share via email or social media

Keywords

grasping, in-hand manipulation, task planning

Abstract

Planning and executing object manipulation requires integrating multiple sensory and motor channels while acting under uncertainty and complying with task constraints. As the modern environment is tuned for human hands, designing robotic systems with similar manipulative capabilities is crucial. Research on robotic object manipulation is divided into smaller communities interested in, e.g., motion planning, grasp planning, sensorimotor learning, and tool use. However, few attempts have been made to combine these areas into holistic systems. In this review, we aim to unify the underlying mechanics of grasping and in-hand manipulation by focusing on the temporal aspects of manipulation, including visual perception, grasp planning and execution, and goal-directed manipulation. Inspired by human manipulation, we envision that an emphasis on the temporal integration of these processes opens the way for human-like object use by robots.

1. INTRODUCTION

Grasping is a temporal process that reaches from visual scene understanding to complex object manipulation. The mere act of closing the hand on an object is just one of the many steps involved in this complex action. Grasping consists of a number of state–action pairs that are sequentially executed to achieve a goal state. This sequence involves different kinds of sensing and actuation, as shown in **Figure 1**. Therefore, grasping and manipulation require more than merely approaching one point and closing the fingers—they require an understanding of how the object is placed in the environment, how to plan motions, how to interact with the environment to achieve the desired configuration, and how to properly hold or manipulate the object to execute the current task. The temporal process is not a mere feedforward structure. Instead, the different stages and mechanisms influence each other constantly in feedback loops, as shown in **Figure 2**, resulting in a highly complex system. We believe that integrating these mechanisms can enable robots to be multipurpose systems that can solve different tasks.

In this article, we review advances in robotic grasping and relate each step of this temporal process to human grasping behavior. By pointing out the main elements of human manipulation behavior, we identify the key abilities that robots must master when facing complex manipulation tasks.

First, in Section 2, we analyze the mechanisms that lead from vision to manipulation in humans. In the subsequent sections, we explore the steps of the robotic grasping process by following its temporal structure. In Section 3, we introduce the perception of objects and the inference of possible grasp configurations. Section 4 presents solutions for planning and executing stable and robust grasps using the available sensory information. Once the object has been grasped, it needs to be manipulated to fulfill the current goal; Section 5 analyzes the state of the art in object manipulation, dexterous manipulation, and tool use. Despite many recent advances in technology, the problem of grasping and manipulation is not fully solved, and we present the open questions in Section 6. Finally, Section 7 concludes our analysis of grasping as a temporal process.

2. FROM VISION TO MANIPULATION IN HUMANS

Human object manipulation skills are governed by the interplay of different sensorimotor processes. This interplay follows a natural temporal sequence of repeated sensory predictions, motor

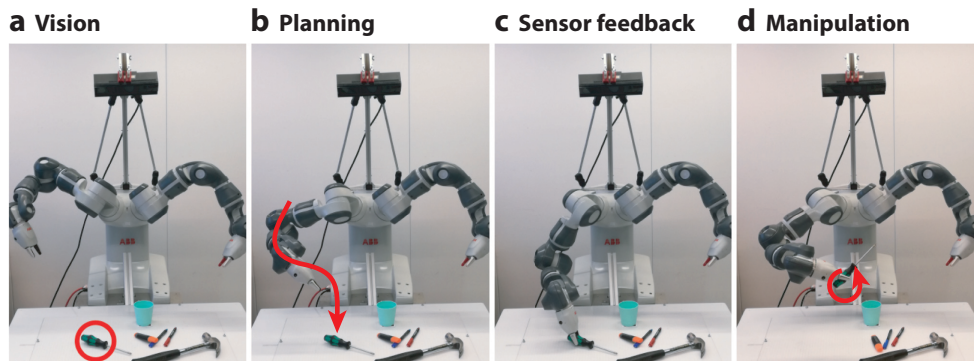


Figure 1

The different stages involved in a manipulation task. (a) The target object is detected, and proper grasp locations are identified. (b) The robot plans a trajectory to reach the target location. (c) The object is grasped, and the grip is adjusted with the help of tactile and visual feedback. (d) The object is manipulated into the desired configuration or used as a tool.

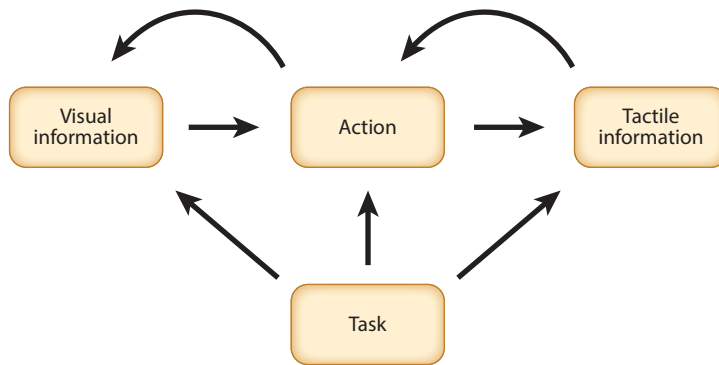


Figure 2

The task of object manipulation is governed by interactions among sensory and motor processes.

actions, and sensory feedback. As described in Section 1, on a high level this sequence begins with the visual understanding of a scene that includes identifying objects and possible contact points, a process described in Section 2.1 under the umbrella of visual attention. Once a promising grasp location has been determined, the grasp is planned and executed, as discussed in Section 2.2. Importantly, this process is driven by sensory prediction instead of merely sensory feedback, enabling tactile feedback during hand–object interaction to be readily integrated in order to counteract errors, as described in Section 2.3. Most often, higher-order goals such as tool use steer the grasping and the subsequent manipulation processes, which are described in Section 2.4.

2.1. Visual Attention and Perception

In this section, we discuss how visual attention, visual perception, and human manipulation behavior influence each other.

2.1.1. Attention. Visual attention during grasping and object manipulation is vital, serving as guidance for both low-level motion planning and high-level task planning. Imagine a pick-and-place task that requires you to sequentially move first a green, then a red, and finally a blue cube to a given location. Without any distracting objects in the scene, you will be able to perform this task without fixating on a single cube, simply by recognizing the colors. However, when distracting objects with the same color distribution but different shapes are also present, you will have to search for the cubes and thus direct your gaze toward the individual objects. These two types of visual attention are called covert attention, which makes use of peripheral vision and can be directed toward spatial regions without explicit eye movements, and overt attention, which is determined by eye movements and fixation. As is evident from the pick-and-place example, both of these mechanisms are employed during manual tasks. Under natural conditions, the fixation pattern during a manipulation task—overt attention—shifts sequentially from one target location to the next, even before the current action is completed. These predictive gaze patterns are steered by the task requirements, such as contact points or obstacles (1). While overt attention covers only a single location at every time step, covert attention can be directed toward different objects at the same time. As showcased in the pick-and-place example, however, distractors and obstacles might impede performance based on covert attention alone (2).

2.1.2. Perception. The distribution of visual attention is influenced by bottom-up and top-down processes. While fast saccades—i.e., rapid eye movements—are mainly stimulus driven, slower

visual selection processes are dictated by top-down, goal-driven mechanisms (3). This modulation by actions and their goals guides not only visual attention, but also visual perception of a scene. Visual perception during manual actions is used to determine spatial aspects (such as the distance, shape, size, and orientation of an object) and estimate the object's properties (such as weight, surface roughness, and deformability) and motion patterns (such as velocity, acceleration, and causation). However, depending on the action or task, the perception of an object can change. For example, action-modulated perception becomes apparent when subjects are asked to hold an object of a certain weight and simultaneously judge the weight of an object held by another person: Carrying a heavy weight leads subjects to believe that other objects are lighter than they actually are, and carrying a lighter weight leads them to believe that other objects are heavier than they actually are. As an example of task-modulated perception, subjects who are asked to either reach for or grasp an object are influenced differently by distractors. A distractor at a different location modulates reaching performance, while a distractor of a different shape impedes grasping. These and other examples are discussed by Schütz-Bosbach & Prinz (4).

2.1.3. Hand-eye coordination. Up to this point, we have discussed visual attention and perception mainly in a general sense. How do these processes influence the performance of simple grasping and manipulation actions? Thanks to the visual information about the objects of interest and the environment, movements can be planned in a predictive manner. Given an estimated size and weight of an object, the aperture of the fingers and the strength of muscle tension are adjusted even before hand-object contact. Constant visual feedback helps to adjust the action in case of error and guides the hand through the environment, e.g., to avoid obstacles. Critical key points, such as future contact points or interfering obstacle parts, are the center of attention and are monitored sequentially as the task unfolds (1). Thus, manual tasks are to a certain extent a question of hand-eye coordination. They are encoded in the form of predictive sensorimotor contingencies that allow precise actions under sensory uncertainty and render online error correction possible.

2.1.4. A robotics viewpoint. The revival of deep learning has led to unprecedented performance advances in many areas of research on computer vision. Nevertheless, many of the current approaches toward, e.g., object detection and scene segmentation rely on supervised learning and large amounts of data. In contrast to human vision, this approach to vision is not embodied and therefore does not allow for biological-like sensorimotor synergies in robots. While humans can relate objects to sensory impressions such as how an object feels, smells, and tastes, a computer vision system that relies purely on images identifies objects based on semantic image statistics. However, as becomes apparent from the discussion above, action and visual perception influence each other to make sequential optimal decision-making possible. This interplay creates a robust, adaptable system in humans that can serve as inspiration for robotics research. We review both the current state of the art and more classical approaches to computer vision for robotic grasping and manipulation in Section 3.

2.2. Grasp Planning and Execution

Reaching and grasping are fundamental to human existence. Infants as young as 18 weeks are able to use predictive strategies when reaching for moving objects (5). Thus, from an early age, humans rely on internal models of the properties and behavior of objects to anticipate future time steps and the outcome of their own actions. The variability of preferred grasp strategies for a given object between individuals is relatively small (6), suggesting that object properties induce particular grasping behaviors. In this section, we first introduce different possible grasp types.

We then discuss the relationship between grasp type and both object properties and tasks before delving into human grasp planning.

2.2.1. Grasp type taxonomies. The variety and complexity of human–object interactions hamper a clear distinction between different grasps. In general, we distinguish between power grasps and precision grasps. Power grasps require arm movements to move the object, e.g., when holding a water bottle while drinking. Precision grasps allow movement of the object based only on the hand, e.g., a needle held between the thumb and index finger that can be rearranged with only those two digits. Additionally, an intermediate grasp family might exist that shows characteristics of both power and precision grasps. Feix et al. (7) presented a more detailed distinction and identified 33 basic grasp types. Different grasp types serve different purposes, as discussed below.

2.2.2. The influence of object properties. The choice of grasp type depends on both the object’s properties and the current task. Most of an object’s properties are usually approximated based on visual input. Among other properties, it is evident that the mass of an object (i.e., its weight) influences the grasp type and force. Here, the center of mass plays an important role in determining stable grasp configurations. Next, the size and shape dictate possible positions of contact points. In fact, given knowledge of an object’s shape, we can often accurately predict which grasp a person will apply. However, the global shape of an object is less important than local shapes of different object parts. Imagine a cooking pan, which consists of a flat, circular core encircled by the body and rim, along with a handle. The pan can be grasped by the handle with one hand, held along the body and rim with two hands, or supported from below by the palm or digits of one hand. Thus, the shapes of the different parts afford different grasp types, which may be chosen according to the current task. Finally, object properties such as surface smoothness and deformability play a role during grasp planning, especially to determine how much force will be needed. Feix et al. (8) discussed in detail the properties of everyday objects and the grasp types associated with them.

2.2.3. The influence of task properties. In a follow-up paper, Feix et al. (9) investigated the influence of the task on grasp types. They categorized tasks according to (a) the constraints that are imposed on the object (e.g., which degrees of freedom are allowed to change); (b) the functional class, which is a high-level description of the task (e.g., “using”); and (c) the force required to perform the task. They also discussed the general statistics of these properties in the everyday activities of two housekeepers and two machinists. When investigating the predictive power of task versus object properties to infer grasp type, they concluded that object properties are a more reliable indicator of the grasp type than task properties. Combining the two can increase the predictive performance for certain grasps. However, as it is difficult to determine a general representation of task properties, the question of task influence requires future research effort.

2.2.4. Human grasp planning. The precision and speed with which humans are able to grasp objects suggest that motion planning is highly automated (10). In this section, we discuss human grasp planning on a mechanism level without delving into the neuroscience of motion control. To understand the processes involved, it is advantageous to identify the properties of the movement that are optimized during execution. A common approach to understanding these factors is a modeling perspective, i.e., designing mathematical models of movements and comparing them with actual human performance.

Some of the factors considered during motion planning depend on the trajectory of the hand. For example, humans appear to plan reaching movements based on the hand position in both

visual space and joint space (11). To generate smooth movements, the minimum jerk model postulates that humans optimize for the square of the rate of change of acceleration (jerk) of the hand integrated over the entire trajectory (12). Other models consider several competing factors, which can include joints other than the hand. These factors include movement time and distance, as well as velocity and acceleration profiles (13) and minimizing the torque to all joints (14).

A more holistic theory of human grasp planning is the posture-based motion planning framework (15). This framework postulates that humans determine the goal pose—in this case, the grasped pose—before planning a feasible trajectory toward it. Instead of replanning poses, the model proposes that the goal pose is chosen from a set of stored poses and adjusted if necessary. Thus, we make use of previous experiences to solve the current task. This idea is in line with anticipatory motion planning, which anticipates not only the outcome of an action but also the desired goal position. Anticipatory motion planning allows one to adjust the limbs toward the required goal pose during movement execution instead of calibrating them at the end of the trajectory. For example, hand orientation can be adjusted in an anticipatory fashion while moving the arm (16).

One benefit of a goal-poses-based approach is that it reduces the enormous range of possible grasping trajectories and configurations to those that are in line with the desired pose. The high-dimensionality issue applies to more than just the grasp action: The hand alone has 27 degrees of freedoms (17) that need to be controlled during object manipulation. This problem is believed to be solved by so-called control synergies, which describe how different entities can become a unit based on the correlation and interdependence of the entities (18). When considering the hand and fingers, it becomes clear that it is possible to predict the position of one finger based on the configuration of the remaining fingers. To investigate this hypothesis, dimensionality reduction techniques such as principal component analysis are used on motion data to extract the number of dimensions needed to encode different actions. The number of principal components required to explain the variance of the observations varies and depends more on the current task than on individual differences between subjects (19).

In conclusion, human grasping strategies seem to optimize different optimality constraints while operating in a reduced number of dimensionalities. Furthermore, they are influenced by previous experiences and by anticipative signals. These factors allow for fast and precise action execution.

2.2.5. A robotics viewpoint. Robotic grasp planning is challenging for multiple reasons, including high-dimensional operating spaces and sensory uncertainties. Many of the human factors described above have influenced research in robotic grasp planning, as described in Section 4. However, one major advantage of human grasp strategies might be motor memory, i.e., that we make use of previous experiences to solve current problems. This means that grasping and manipulation are not only short-term temporal processes but also influenced by long-term learning. This integration of experiences could improve robotic grasping systems.

2.3. Tactile Information

The human brain assigns a significant area of the sensorimotor cortex to represent the fingers and hands (20). Many of these connections are related to the sensory processing of nerve signals, which originate mostly in the fingertips. Depending on their location and depth, these nerves—called afferents—specialize in different parts of the tactile input. For example, they can differentiate between high- and low-frequency skin deformations, identify spatiotemporal patterns, and distinguish between lateral and tangential force directions (21).

2.3.1. Action phases and contact events. Each activity or action phase carries its own sensory fingerprint encoded in these signals. Contacts between the hand and an object, contacts between a grasped object and other objects, and grasp stability give rise to individual sensory signatures. These signatures are part of predictive feedback loops in the sensorimotor system during action execution. While a person is completing a task, a certain series of distinct mechanical contact events must occur that can be anticipated. The difference between the predicted sensory signature and the actual tactile feedback allows for rapid adjustments and precise control. To smoothly transition between action phases, the predicted sensory state can be used as the initial state of the following action phase. Thus, instead of reactive control, the sensorimotor system can actively plan ahead.

2.3.2. A robotics viewpoint. Evolution has provided humans with a highly specialized pair of hands and nerve fibers that seamlessly integrate sensorimotor information from not only tactile channels but also visual and auditory input. These multimodal signals are often not present in robotic setups, as few robotic grippers are equipped with tactile sensors. Force-torque sensors are not nearly as precise as biological nerve fibers. Nevertheless, research on contact feedback is steadily developing. We discuss these robotic systems in Section 4.3.

2.4. Object Manipulation

Up to this point, we have considered the topic of grasping from a rather low-level, sensorimotor-integration point of view. Human manipulation skills, however, become more complex as soon as we consider high-level actions, such as tool use. While tool use has been observed in other species, such as primates and crows, the cognitive abilities of humans are unprecedented in terms of adaptability and diversity. The cognitive basis of human tool use consists of “hand-eye coordination, causal reasoning, function representation, executive control, social learning, teaching, social intelligence, and language” (22, p. 1). In the following, we discuss several of these aspects, as they can serve as inspiration for the robotics community.

2.4.1. Affordances. The theory of affordances proposes that we represent our environment centered around actions. Affordances describe those actions that a certain object affords, i.e., which actions it makes possible or which action can be applied to it. The concept of affordances has its origins in the movement of embodied intelligence in the 1980s and 1990s (23). In this view, the representation of the world lends itself to manipulation.

As an example of affordances, a cup affords to be drunk from and to carry liquid in. Objects can change other objects’ affordances; for example, a spoon adds the affordance of eating soup to the cup. This mutual interaction between objects can be seen as a basis for tool use (24). However, merely representing the environment based on actions does not suffice to explain the variety of human manipulation skills. Analogical and causal reasoning are required to fuse the affordances of different objects into a chain of goal-directed actions, as discussed next.

2.4.2. Causal reasoning. Complex tasks require planning a sequence of several state–action pairs, meaning that we need to reason about which affordances to use at which stage of the task. It is believed that causal reasoning is a factor that sets humans apart from animals, as it allows us to easily solve novel problems with both previously acquired and innovative solutions. Here, analogical reasoning becomes important (25). Imagine that you always use a small step stool to reach the top of a kitchen shelf, but someone has removed the stool. As a chair can serve the same purpose—i.e., it has the affordance to help you to reach higher—you might choose to climb on the

chair instead, taking advantage of the similarity between the affordances of the stool and the chair. Analogical and causal reasoning allow us to make use of previously acquired skills and knowledge. This knowledge is not only the product of personal experience but also largely influenced by cultural and social learning, as described below.

2.4.3. Social learning. Humans acquire many skills not by pure trial and error but rather by simply observing other humans performing a certain task. The translation of others' actions into our own action repertoire is believed to be governed by the so-called mirror system. This neural system is known to react both when performing an action and when observing others perform the action. This system enables us to translate between embodiments.

Social learning not only encompasses direct teaching by a more experienced actor but also is evident in unsupervised settings. Without the current need to solve a task oneself, the strategy of another person is remembered and can be applied at another point in time. Social learning thereby enables humans to acquire skills without having to perform the actions themselves.

2.4.4. A robotics viewpoint. While grasping and specialized tool use gain a lot of attention in the robotics community, the question of general, transferable tool use skills is less well studied. Importantly, this complex behavior requires the low-level processes of sensorimotor control to be implemented as well as the ability to represent and reason about objects' properties and functions. Although the robotics community adopted the idea of affordances early on, explicit causal reasoning over a large variety of tools has not been explored. We continue the discussion of robotic tool use in Section 5.2.

3. VISION-BASED GRASP SYNTHESIS

Consider a scenario where a robot is grasping an object in a cluttered space. To do so, the robot must make several high- and low-level decisions. First, it must decide where to grasp the object. Having selected a final location and orientation for the hand, it must then plan an obstacle-free path. Finally, it must execute the grasp without colliding with the object or any of the surrounding objects.

Although modern-day robots can carry out a plethora of manipulation tasks, finding a stable grasp on a previously unseen object based solely on vision remains challenging. Using visual input to determine possible grasp configurations is termed vision-based grasp synthesis. In the following, we discuss different approaches that have been taken to this problem.

3.1. Grasp Synthesis Overview

Sahbani et al. (26) divided grasp synthesis into analytical and empirical approaches. Analytical approaches are based on kinematic and dynamic formulations of grasp synthesis problems and as such provide guarantees regarding the criteria of dexterity, equilibrium, etc. However, they rely on assumptions such as knowledge of the full geometry of an object as well as its friction coefficients and weight, which are often not available to the robot. Furthermore, due to noisy sensors, properties such as the poses of the object and gripper can be known only approximately. Empirical or data-driven approaches, by contrast, relate to learning the association between object and grasp properties from data. These approaches rely on perceptual processing to identify the object representation, which is then used to find possible grasps. Although these methods do not provide guarantees regarding the above-mentioned criteria, they accommodate the uncertainties prevalent in real-world scenarios and avoid the computational complexity of analytical ones.

Bohg et al. (27) grouped data-driven approaches based on whether they synthesize grasps for known, familiar, or unknown objects. This classification allows us to identify available object representations and determine necessary perceptual processes for grasp synthesis given a certain condition. For example, in the case of grasping a known object, the object is represented as a full 3-D model that the robot has encountered before. For the robot to grasp it, its vision system must first recognize the object and estimate its pose and then retrieve a suitable grasp from an experience database. In the case of grasping a familiar object, a representation that is suitable for comparison to previously seen and grasped objects must be found. The challenge is to define a similarity metric that allows the transfer of grasp experience. Finally, in the case of unknown objects, the question of finding the right representation translates to finding the object features that indicate stable grasps on that object.

3.2. Learning to Grasp From Vision

In a real-world robotic setting, a robot often may not have access to full 3-D models of all objects, which can significantly impede grasping. The algorithms must use noisy and partial information to reason about grasp properties. The majority of works that fall under the umbrella of analytical approaches use simplifying assumptions—e.g., that the object can be approximated by a set of primitive shapes (28–30) or that the objects are planar (31). Others provide solutions for specific cases, such as grasping the corners of towels (32). However, to achieve good generalizability while also accommodating noisy inputs, a natural approach is to learn a mapping from visual features to grasping properties. Inspired by the successes of learning methods, Kamon et al. (31) showed that learning approaches could be applied to grasp synthesis based on visual information. They introduced a learning component to choose contact points and predict the quality of a grasp.

Learning-based methods have enabled grasp synthesis algorithms to generalize to novel objects. Recently, deep learning has seen unprecedented developments within the computer vision community and is therefore a promising candidate for learning-based approaches. Lenz et al. (33) were the first to apply deep learning to the problem of robot grasping. They viewed this problem as a matter of detection and focused on identifying suitable grasp regions from RGBD (red, green, blue, depth) images. Drawing from a paper by Jiang et al. (34), they considered only parallel grippers and parameterized the grasp using a rectangle representation. In this setting, the process of finding a suitable grasp is equivalent to finding the grasping rectangle with the highest score.

While the performance of deep learning is promising, these approaches are data hungry. Prior to the success of deep learning, learning methods used hand-designed features to build grasp-oriented object representations. Although deep learning proved successful for images or videos where large amounts of data are available, collecting data for robotic grasping requires tedious human labeling, as in the above-mentioned approach, or hours of execution time on a real physical system, as in Reference 35. Even though collecting data on a robotic system is independent of human labeling biases, such data sets contain at most hundreds of objects, which renders their generalization performance questionable. One way to address this issue is to leverage the availability of grasping simulators. These simulators provide a fast, easy way to collect, analyze, and evaluate grasps in an environment that resembles the real world. Mahler et al. (36) trained a network that predicts grasp success from 6.7 million synthetic point clouds, grasps, and analytic grasp metrics generated from thousands of 3-D models. The transferability from simulation to real-world applications is a vibrant field of research.

From here on, we assume that the problem of visual grasp synthesis—i.e., of determining possible grasp contacts—has been solved, and concentrate on grasp planning and execution.

4. GRASPING

Once the object of interest has been identified and possible grasp contacts have been determined, the actual arm and gripper motion must be planned, as discussed in Section 4.1. The execution of the grasp is subject to uncertainties and environmental changes, which is why special care must be taken to generate robust grasps (see Section 4.2). Finally, tactile sensors can help to adjust the initial grasp toward the actual goal pose or to achieve a stable grip, as described in Section 4.3.

4.1. Grasp Planning

Grasp planning involves first planning the pose of the robotic hand with respect to the object and then planning a feasible configuration for the whole robot, which is suitable for achieving the desired grasp. These two problems have traditionally been addressed separately, but a joint approach improves the efficiency of the system. As mentioned in Section 2.2.4, a typical human behavior in grasping involves anticipatory motion planning, in which the hand pose is adjusted during the arm motion. Recent solutions for this type of planning have been developed and are discussed in the next section, but fully integrated methods are still not commonly used due to their challenges and lack of open-source availability.

Traditional approaches to motion planning use standard algorithms to determine a collision-free motion of the robot arm toward the pose of the desired grasp (37). However, since the grasp pose is usually not planned while taking into account the robot's kinematics, the pose can sometimes be unreachable from the starting configuration, in which case the motion planning for that particular grasp may fail. As a consequence, multiple grasp poses must be planned and tested for reachability, which requires significant computational overhead.

To overcome this problem, methods to simultaneously plan the grasp and the arm motion have been developed. Fontanals et al. (38) described a method that obtains an end effector's goal pose while taking into account both the robotic arm's capabilities and a successful grasp configuration using independent contact regions (39). Vahrenkamp et al. (40) presented a planning strategy based on using rapidly exploring random trees (RRTs) to build a tree of reachable and collision-free configurations. This approach has been applied to bimanual grasps and motion planning on a humanoid robot. Also based on RRTs, the method proposed by Hausteine et al. (41) exploits the hierarchical structure of the Hierarchical Fingertip Space grasp planner (42) to plan fingertip contacts and arm motion at the same time. Despite these recent advances, however, simultaneously planning the grasp on the object and the motion of the robot remains a significant challenge in the field of robotic grasping.

4.2. Robust Grasping

Robotic grasping is a difficult process, as a robot's environment is subject to uncertainties and disturbances. Therefore, additional constraints should be imposed to ensure a successful hold on the object under these conditions.

We define a grasp as robust when it succeeds despite a certain level of uncertainty and can compensate for disturbances. The notion of robust grasping is therefore linked to the quality of a grasp. However, measuring or evaluating the quality of a grasp is not an easy task on its own. Roa & Suárez (43) categorized the different methods of measuring grasp quality into three main groups: one that relies on the location and attributes of the contact points, one that relies on the hand configuration executing the task, and one that combines the two previous measures into a more global one.

To date, no unique grasp quality measure is able to universally define a robust grasp. Rubert et al. (44) used 10 different grasp quality measures to evaluate 126 different objects with seven different grippers, which helps to determine suitable thresholds for the chosen measures. Their conclusion stated that the 10 quality measures are composed of at least seven independent grasp aspects, including algebraic properties (such as the volume of the wrench space), the distribution of contact points (such as the area of the grasp polygon), and the magnitude of forces applied. A grasp planner should consider as many independent aspects as possible to evaluate a grasp.

A supplementary concept to grasp evaluation is caging. In the context of grasping, an object is said to be caged by a gripper or robotic hand if it cannot escape arbitrarily far, independently of the kind of rotation or translation applied to it.

Rodriguez et al. (45) introduced the notion of pregrasping cages. A pregrasping cage is a gripper configuration that cages the object and would constitute a robust grasp if the grippers were to close. The authors argued that, in a 2-D scenario, every cage that is achieved with a parallel gripper is such a pregrasp configuration. This notion is extended from two fingers to n fingers by introducing a subset of cages (the so-called F -cages). This subset is strictly contained in the set of pregrasping cages that can be used as waypoints to reach a robust grasp. This method is limited to 2-D cases, and the real-world applications are therefore limited.

Varava et al. (46) presented caging grasps on 3-D objects that exhibit certain topological features. By using the linking number, a topological invariant, one can compute and verify a caged grasp given the mesh of a 3-D object. This 3-D object must exhibit at least one of the topological features referred to as necks or double forks, which describe object areas that are more narrow than their surrounding.

On the more classical side, Hsiao et al. (47) presented a decision-theoretical approach without the notion of caging. First, an initial probability distribution over the object pose is generated using a noncontact system; once the object is grasped, sensors are used to update and refine the initial belief state. This procedure is repeated several times for a given target grasp. After performing a wide range of experiments, the authors concluded that updating the belief state based on sensory observations accurately infers the robustness of the grasp. This result demonstrates that only a few actions are needed for information gathering in order to perform robust grasping.

Leveraging the power of deep convolutional neural networks, Dex-Net 2.0 (36) is able to predict the probability of a robust grasp. Being trained on a synthetic data set of 6.7 million point clouds, Dex-Net demonstrates a remarkable success rate in experimental trials. Leveraging huge amounts of data that can be synthesized in machine learning in order to make progress in difficult problems like grasping is a promising trend in recent years.

4.3. Interaction and Contact

After planning a grasp, the robot must execute the plan and successfully keep the grip around the target object. This phase requires considering more than just the shape of the object: Different materials and different masses lead to different requirements, from gentle to strong grasps. Therefore, similarly to human behavior (see Section 2.3), measuring and controlling the gripping force is of fundamental importance for a successful grasp. For detailed information about measuring the contact and gripping force, we refer readers to reviews by Yousef et al. (48) and Kappassov et al. (49), which describe recent progress in tactile sensing for robotics, particularly manipulation.

The robot should not drop or crush an object. Therefore, the planned grasp needs to be executed with an understanding of the contact between the fingers and the object. It is not surprising that approaches to grasp adaptations that avoid slippage and produce delicate grasps mimic human

behavior (50). Moreover, tactile experience helps compensate for uncertainties that result from imprecisions in other perception systems (51). For instance, incorrect object pose estimations from cameras or an only partial reconstruction of shapes can be overcome with the help of tactile feedback. Haptic sensors are also used for tactile exploration, a solution to improve the partial knowledge of the environment. This tactile exploration guides the fingers along the shape of the object (52).

Once the object is grasped, the tactile feedback can also be exploited to ensure stability. More specifically, detecting slippage helps compensate for it and prevents the object from being inadvertently released (53). For instance, Hang et al. (54) exploited tactile feedback to stabilize the grasp once the object begins to slip due to changes in the object's mass (e.g., when a grasped bottle is being filled with water). The process of adaptation occurs after the first grasp and can also be viewed as a step of object manipulation, which we discuss below.

5. IN-HAND MANIPULATION AND TOOL USE

The final stage of manipulation is the use of the chosen object. This usage varies with the specific task, which might range from a simple pick-and-place task to a significantly more complex activity. Successful execution of complex tasks often requires taking an initial grasp on an object and then further manipulating it to reach a different configuration. The ability to perform this type of rearrangement within the hand is known as dexterity.

5.1. In-Hand Manipulation

Dexterity is a vital human ability that has partly guided our evolution. Many robotic hands mimic the structure of human hands to take advantage of the range of motions that we are capable of. However, this mechanical complexity leads to several challenges in the design and control of such artificial hands (55, 56).

Dexterous manipulation is an essential step to finalize the overall manipulation task. In fact, a robot's task is not finished after a successful execution. Instead, it is followed by the challenge to use that object and manipulate it correctly to fulfill the given goal. Despite careful planning and execution, the grasp achieved on the object could differ slightly from the desired pose. In addition, some tasks can require holding the object in different poses during task execution. Therefore, the robot must be able to perform dexterous (or in-hand) manipulation to rearrange the object-hand configurations.

5.1.1. Regrasping. A simple solution for adjusting the pose of the object is regrasping (57): The object is placed on a surface and picked up multiple times until it is grasped as desired. This is a stable approach that is used mainly in industrial settings: The robot places the object onto predetermined locations such that the regrasping sequence leads to the desired grasp. However, regrasping lacks efficiency because it might require the robot to perform many grasps and many placement actions. Instead, human-like dexterity allows the robot to move the object inside its hand without additional and redundant movements.

5.1.2. Dexterous manipulation. A more efficient solution exploits the dexterity of the robot's hands, moving the object by moving the fingers and sliding the contacts between the fingertips and the object's surface. In addition to exploiting the many degrees of freedom of a multi-fingered robotic hand, in-hand manipulation can also take advantage of high-quality and high-frequency sensory feedback when available. For instance, in a study by Furukawa et al. (58), a robot

repositioned an object by tossing it in the air and then adjusting its hand configuration to catch the object in the desired pose, taking advantage of a high-frame-rate camera. This kind of object repositioning using fast motion of both the object and the hand is called dynamic regrasping. However, the necessary control and feedback frequencies are difficult to obtain in many available systems. Therefore, in-hand manipulation is achieved mainly by moving the fingers in a quasi-static manner or by moving the object along different stable poses.

As mentioned above, using a multifingered hand for dexterous manipulation leads to several challenges in planning and control. Some solutions simplify the problem by moving only a subset of the fingers and planning for a predetermined simple motion. For instance, in a study by Shi et al. (59), the target motion is a planar sliding along the object surface. In a study by Sundaralingam & Hermans (60), the object's motion is very limited and the contact points between the fingers and the object are kept stable, leading to a change in the object's pose with respect to the gripper that depends only on the change in the fingers' configuration.

Antonova et al. (61) used a reinforcement learning approach to learn control policies from a custom simulator that is robust to the differences between simulation and the real-world robot. Reinforcement learning and other learning-based approaches have the drawback of requiring thousands to millions of training episodes to obtain usable policies, which makes them undesirable for real robot deployments. While learning approaches based on simulations seem promising, the problem of transferring the simulation to the real world is still hard to overcome.

Despite decades of active research in this field, advances in planning, control, and mechanics have not yet managed to provide a complete solution to the problem of dexterous manipulation, which remains an open challenge in robotics.

5.1.3. Extrinsic dexterity. A specific subset of dexterous manipulation that has recently received the attention of the robotics community focuses on achieving dexterous manipulation with non-dexterous end effectors, such as simple parallel grippers. In this case, the poor dexterity of the hand is enhanced by external supports: gravity, friction, inertial forces, and contact surfaces. Due to the limitations of parallel grippers, they can perform only simple in-hand repositioning, such as sliding or rotating along a single axis (pivoting) (62, 63). However, the combination of many simple motions allows the robot to achieve a satisfactory range of repositioning and removes the challenges in planning and control of multifingered hands.

The ability to exploit elements external to the robot to achieve manipulation has been called extrinsic dexterity (64). With the increasing availability of dual-arm systems, the exploitation of external supports, particularly external pushers (65) and surfaces (66), can be replaced by using both arms at the same time (67).

By exploiting not only the degrees of freedom provided by complex hands but also external static and dynamic constraints, the robot's manipulation behavior becomes closer to the behavior of humans, who constantly exploit knowledge about the environment to their advantage.

5.2. Tool Use

Tool use gave humans a decisive advantage over other species, enabling them to perform a wider range of actions and interact with the environment in versatile ways. When using tools, humans can take advantage of physical concepts, such as the law of the lever when using a wrench to increase the force applied, and exploit geometric shapes, such as sharp edges on a knife to perform a cutting task. Using a tool is a particular instance of dexterous manipulation in which the goal is not only to adjust the pose of an object inside the hand but also to hold the item and reconfigure it during the execution of a task while interacting with the environment.

Existing tools (e.g., pens, scissors, and screwdrivers) have been designed for human use. In particular, they reflect the shape and mechanisms of the human hand. By contrast, in industrial settings, robots are able to repeat a particular task using a specialized gripper designed for that task (68, 69).

While the specialized gripper gives the robot an edge in performance, it loses generalization and flexibility—the desirable advantage of using a robot instead of an ad hoc built machine. In fact, we envision a robot to be a multipurpose machine, and as such it should be used for different tasks. Instead of seeing the robot as a tool itself, it can be viewed as an agent that exploits several tools to complete different tasks. However, improving the efficiency of robots in industry and moving them into household environments will require overcoming the limitations of custom-made grippers and instead developing a generic robot hand that can fulfill different purposes and exploit available tools to complete different tasks.

A first challenge in robot tool use lies in the end effectors. They are rarely as dexterous as a human hand and often consist of just a simple parallel gripper, which makes it difficult to grasp and control tools to perform certain actions. An additional challenge is the concept of tool use itself: The robot must “understand” the relationships among the tool, the action, and the effect, as well as the concept that specific tools can be used in specific tasks.

It is impractical to create models of all possible tools, and relying on such models would constrain the robot to operating in known environments. This aspect of tool use was addressed by Kemp & Edsinger (70). The authors identified that many tools share a common characteristic in which a tip is used to perform an action and a handle is located opposite the tip, and they therefore focused on autonomously detecting the tool tip. Here, the tool tip is assumed to consist of edges that are tangent to a circle, regardless of scale. Using this assumption, the proposed method weighs those edges and shows that the tip can be reliably detecting when moving a novel tool around.

Hoffmann et al. (71) presented a framework for adaptive tool use under variable grasps. The first step is to extract the transformation between palm and tool tip via a vision system. In addition, the relationship between hand motions and tactile senses is learned in order to control the tool tip using a three-fingered DARPA Autonomous Robotic Manipulation (ARM) robot. This approach is tested by autonomously grasping an electric drill, guiding the detected tool tip, and drilling a hole at a given target position. In a second experiment, the framework incorporates tactile feedback, which is exploited to draw on a surface of unknown slope using an ST Robotics R17 robot and a pencil. This framework does not rely on heavy calibration, an important aspect when robots must deal with highly unstructured environments.

Other approaches have focused on a broader range of objects than those with a tip. Yamazaki et al. (72) presented an integration of various recognition and control strategies for a kitchen environment. By combining visual functions such as model-based recognition and employing predefined motion trajectories inspired by humans, the robot is able to perform various specified kitchen tasks. The method relies on a large amount of domain knowledge and assumptions, which poses a problem for making robots work reliably in unstructured environments, such as homes or offices.

The RoboCup@Home league aims to develop service robots that can perform complex tasks in real-world environments. A good example of a state-of-the-art robot in that sector is Lisa (73), which won the open demonstration challenge in 2015. Lisa combined many advances in the field, including 3-D object recognition, object manipulation, affordance detection, and speech recognition. In a prepared demonstration, the robot was able to dispose of a cup in a trash bin and carry the trash bag to a target position. While the system still requires help from a human actor, those efforts show that progress is being made toward robots that can use tools and operate in unstructured environments.

6. OPEN QUESTIONS

The previous sections have taken a broad look at the object manipulation process, the temporal aspects of grasping and in-hand manipulation, and how these aspects extend to tool use. While some aspects of grasping as a process can be considered solved, the field as a whole still faces many significant challenges. Here, we outline some of the open questions that require further research.

6.1. How Can We Integrate All Steps of the Grasping Process?

The robotics research community is highly diverse and covers many subareas and topics. While specific topics, such as those discussed in this article, have been studied in detail, their combination and interplay have been of less interest. Building robust robotic systems of this complex scale is challenging because it requires integrating many different processes. However, as both human and robotics research shows, sensorimotor integration of different subprocesses reduces noise in estimations and allows the generalization of behavior to different environments. Because we envision that robots should be able to perform complex tasks in uncertain environments, modeling and integrating the temporal process of object manipulation are essential.

6.2. How Can Learning Approaches Be Combined with Traditional Modeling and Planning Techniques?

Traditionally, robotic manipulation has relied on explicit models of the robot's degrees of freedom and the objects. Learning techniques, on the other hand, attempt to encode these properties in the sensorimotor system of the robot, eliminating the need for specifications. Advances in deep learning have enabled the application of learning approaches to increasingly complex setups. However, the mechanisms of deep learning are not well understood, which makes it difficult to understand a model's decision-making. The question of model uncertainty—i.e., understanding when a model is uncertain about its predictions—is especially important for robotics. It is therefore important to investigate how to integrate learning methods with model-based methods and traditional planning algorithms.

6.3. How Can We Transfer from Simulations to the Real World?

One promising approach to enable learning is employing a system in a simulation before moving it to the real world. This method has the advantages of providing cheap data samples and avoiding wear and tear on a robot. Often, however, there is a gap between the simulated data and real-world sensory data. These differences depend on factors such as the physics model in the simulator, sensory noise in the actual system, and the difficulty of modeling, e.g., a human presence. This raises the question of how to transfer a learned model from one domain (the simulator) into another, similar domain (the real world). Enabling robust transfer between the two systems would allow us to model and learn the different aspects of the manipulation process in simulation and then integrate the pretrained modules into a real-world system.

7. CONCLUSION

In this review, we have viewed grasping as a temporal process, ranging from visual understanding to complex object manipulation. By setting the developments in the robotics community in the context of human manipulation skills, we identified some of the key abilities that a robot should

have in order to achieve human-like manipulation behavior. These abilities include visual scene understanding to guide visual grasp synthesis and obstacle detection, the generation of pose and motion plans that are executed and refined based on tactile and force feedback, and the use of in-hand manipulation techniques (such as dexterous manipulation) to perform more complex actions. The concepts and mechanisms that humans are believed to apply for tool use, such as affordances and transfer learning, can guide robotics research toward more robust, general object manipulation. All of these mechanisms build complex forward-backward loops of sensory and motor information that, if combined, can lead to skillful tool use. Reaching this goal, however, will require understanding the underlying temporal process and implementing it in operating robotic systems.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

LITERATURE CITED

1. Johansson RS, Westling G, Bäckström A, Flanagan JR. 2001. Eye-hand coordination in object manipulation. *J. Neurosci.* 21:6917–32
2. Terao Y, Andersson NM, Flanagan JR, Johansson RS. 2002. Engagement of gaze in capturing targets for future sequential manual actions. *J. Neurophysiol.* 88:1716–25
3. van Zoest W, Donk M, Theeuwes J. 2004. The role of stimulus-driven and goal-driven control in saccadic visual selection. *J. Exp. Psychol. Hum. Percept. Perform.* 30:746–59
4. Schütz-Bosbach S, Prinz W. 2007. Perceptual resonance: action-induced modulation of perception. *Trends Cogn. Sci.* 11:349–55
5. von Hofsten C, Lindhagen K. 1979. Observations on the development of reaching for moving objects. *J. Exp. Child Psychol.* 28:158–73
6. Gilster R, Hesse C, Deubel H. 2012. Contact points during multidigit grasping of geometric objects. *Exp. Brain Res.* 217:137–51
7. Feix T, Romero J, Schmiedmayer HB, Dollar AM, Kragic D. 2016. The grasp taxonomy of human grasp types. *IEEE Trans. Hum.-Mach. Syst.* 46:66–77
8. Feix T, Bullock IM, Dollar AM. 2014. Analysis of human grasping behavior: object characteristics and grasp type. *IEEE Trans. Haptics* 7:311–23
9. Feix T, Bullock IM, Dollar AM. 2014. Analysis of human grasping behavior: correlating tasks, objects and grasps. *IEEE Trans. Haptics* 7:430–41
10. Ghez C, Krakauer J. 2000. The organization of movement. In *Principles of Neural Science*, ed. ER Kandel, JH Schwartz, TM Jessell, pp. 653–73. New York: McGraw-Hill. 4th ed.
11. Morasso P. 1981. Spatial control of arm movements. *Exp. Brain Res.* 42:223–27
12. Flash T, Hogan N. 1985. The coordination of arm movements: an experimentally confirmed mathematical model. *J. Neurosci.* 5:1688–703
13. Nelson WL. 1983. Physical principles for economies of skilled movements. *Biol. Cybern.* 46:135–47
14. Uno Y, Kawato M, Suzuki R. 1989. Formation and control of optimal trajectory in human multijoint arm movement. *Biol. Cybern.* 61:89–101
15. Rosenbaum DA, Meulenbroek RJ, Vaughan J, Jansen C. 2001. Posture-based motion planning: applications to grasping. *Psychol. Rev.* 108:709
16. Herbot O, Butz MV. 2010. Planning and control of hand orientation in grasping movements. *Exp. Brain Res.* 202:867–78
17. ElKoura G, Singh K. 2003. Handrix: animating the human hand. In *Proceedings of the 2003 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, pp. 110–19. Aire-la-Ville, Switz.: Eurograph. Assoc.

18. Santello M, Baud-Bovy G, Jörntell H. 2013. Neural bases of hand synergies. *Front. Comput. Neurosci.* 7:23
19. Todorov E, Ghahramani Z. 2004. Analysis of the synergies underlying complex hand manipulation. In *The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Vol. 2. pp. 4637–40. New York: IEEE
20. Penfield W, Boldrey E. 1937. Somatic motor and sensory representation in the cerebral cortex of man as studied by electrical stimulation. *Brain* 60:389–443
21. Johansson RS, Flanagan JR. 2009. Coding and use of tactile signals from the fingertips in object manipulation tasks. *Nat. Rev. Neurosci.* 10:345–59
22. Orban GA, Caruana F. 2014. The neural basis of human tool use. *Front. Psychol.* 5:310
23. Gibson JJ. 2014. *The Ecological Approach to Visual Perception*. New York: Psychol. Press
24. van Leeuwen L, Smitsman A, van Leeuwen C. 1994. Affordances, perceptual complexity, and the development of tool use. *J. Exp. Psychol. Hum. Percept. Perform.* 20:174–91
25. Osirak F, Jarry C, Le Gall D. 2010. Grasping the affordances, understanding the reasoning: toward a dialectical theory of human tool use. *Psychol. Rev.* 117:517–40
26. Sahbani A, El-Khoury S, Bidaud P. 2012. An overview of 3D object grasp synthesis algorithms. *Robot. Auton. Syst.* 60:326–36
27. Bohg J, Morales A, Asfour T, Kragic D. 2014. Data-driven grasp synthesis—a survey. *IEEE Trans. Robot.* 30:289–309
28. Miller AT, Knoop S, Christensen HI, Allen PK. 2003. Automatic grasp planning using shape primitives. In *2003 IEEE International Conference on Robotics and Automation*, Vol. 2, pp. 1824–29. New York: IEEE
29. Huebner K, Kragic D. 2008. Selection of robot pre-grasps using box-based shape approximation. In *2008 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 1765–70. New York: IEEE
30. Przybylski M, Asfour T, Dillmann R. 2011. Planning grasps for robotic hands using a novel object representation based on the medial axis transform. In *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 1781–88. New York: IEEE
31. Kamon I, Flash T, Edelman S. 1996. Learning to grasp using visual information. In *Proceedings of IEEE International Conference on Robotics and Automation*, Vol. 3, pp. 2470–76. New York: IEEE
32. Maitin-Shepard J, Cusumano-Towner M, Lei J, Abbeel P. 2010. Cloth grasp point detection based on multiple-view geometric cues with application to robotic towel folding. In *2010 IEEE International Conference on Robotics and Automation*, pp. 2308–15. New York: IEEE
33. Lenz I, Lee H, Saxena A. 2015. Deep learning for detecting robotic grasps. *Int. J. Robot. Res.* 34:705–24
34. Jiang Y, Moseson S, Saxena A. 2011. Efficient grasping from RGBD images: learning using a new rectangle representation. In *2011 IEEE International Conference on Robotics and Automation*, pp. 3304–11. New York: IEEE
35. Pinto L, Gupta A. 2016. Supersizing self-supervision: learning to grasp from 50K tries and 700 robot hours. In *2016 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 3406–13. New York: IEEE
36. Mahler J, Liang J, Niyaz S, Laskey M, Doan R, et al. 2017. Dex-Net 2.0: deep learning to plan robust grasps with synthetic point clouds and analytic grasp metrics. In *Robotics: Science and Systems XIII*, ed. N Amato, S Srinivasa, N Ayanian, S Kuindersma, chap. 58. N.p.: Robot. Sci. Syst. Found.
37. Elbanhawi M, Simic M. 2014. Sampling-based robot motion planning: a review. *IEEE Access* 2:56–77
38. Fontanals J, Dang-Vu BA, Porges O, Rosell J, Roa MA. 2014. Integrated grasp and motion planning using independent contact regions. In *2014 IEEE-RAS International Conference on Humanoid Robots*, pp. 887–93. New York: IEEE
39. Roa MA, Suarez R. 2009. Computation of independent contact regions for grasping 3-D objects. *IEEE Trans. Robot.* 25:839–50
40. Vahrenkamp N, Asfour T, Dillmann R. 2012. Simultaneous grasp and motion planning: humanoid robot ARMAR-III. *IEEE Robot. Autom. Mag.* 19:43–57
41. Haustein JA, Hang K, Kragic D. 2017. Integrating motion and hierarchical fingertip grasp planning. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 3439–46. New York: IEEE
42. Hang K, Stork JA, Kragic D. 2014. Hierarchical Fingertip Space for multi-fingered precision grasping. In *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 1641–48. New York: IEEE

43. Roa MA, Suárez R. 2014. Grasp quality measures: review and performance. *Auton. Robots* 38:65–88
44. Rubert C, León B, Morales A, Sancho-Bru J. 2018. Characterisation of grasp quality metrics. *J. Intell. Robot. Syst.* 89:319–42
45. Rodriguez A, Mason MT, Ferry S. 2012. From caging to grasping. *Int. J. Robot. Res.* 31:886–900
46. Varava A, Kragic D, Pokorny FT. 2016. Caging grasps of rigid and partially deformable 3-D objects with double fork and neck features. *IEEE Trans. Robot.* 32:1479–97
47. Hsiao K, Kaelbling LP, Lozano-Pérez T. 2011. Robust grasping under object pose uncertainty. *Auton. Robots* 31:253–68
48. Yousef H, Boukallel M, Althoefer K. 2011. Tactile sensing for dexterous in-hand manipulation in robotics—a review. *Sens. Actuators A* 167:171–87
49. Kappassov Z, Corrales JA, Perdereau V. 2015. Tactile sensing in dexterous robot hands—review. *Robot. Auton. Syst.* 74:195–220
50. Dang H, Allen PK. 2013. Grasp adjustment on novel objects using tactile experience from similar local geometry. In *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 4007–12. New York: IEEE
51. Romano JM, Hsiao K, Niemeyer G, Chitta S, Kuchenbecker KJ. 2011. Human-inspired robotic grasp control with tactile sensing. *IEEE Trans. Robot.* 27:1067–79
52. Sommer N, Billard A. 2016. Multi-contact haptic exploration and grasping with tactile sensors. *Robot. Auton. Syst.* 85:48–61
53. Maria GD, Natale C, Pirozzi S. 2013. Slipping control through tactile sensing feedback. In *2013 IEEE International Conference on Robotics and Automation*, pp. 3523–28. New York: IEEE
54. Hang K, Li M, Stork JA, Bekiroglu Y, Pokorny FT, et al. 2016. Hierarchical fingertip space: a unified framework for grasp planning and in-hand grasp adaptation. *IEEE Trans. Robot.* 32:960–72
55. Bicchi A. 2000. Hands for dexterous manipulation and robust grasping: a difficult road toward simplicity. *IEEE Trans. Robot. Autom.* 16:652–62
56. Ozawa R, Tahara K. 2017. Grasp and dexterous manipulation of multi-fingered robotic hands: a review from a control view point. *Adv. Robot.* 31:1030–50
57. Tournassoud P, Lozano-Perez T, Mazer E. 1987. Regrasping. In *Proceedings of the 1987 IEEE International Conference on Robotics and Automation*, Vol. 4, pp. 1924–28. New York: IEEE
58. Furukawa N, Namiki A, Taku S, Ishikawa M. 2006. Dynamic regrasping using a high-speed multifingered hand and a high-speed vision system. In *Proceedings of the 2006 IEEE International Conference on Robotics and Automation*, pp. 181–87. New York: IEEE
59. Shi J, Woodruff JZ, Umbanhowar PB, Lynch KM. 2017. Dynamic in-hand sliding manipulation. *IEEE Trans. Robot.* 33:778–95
60. Sundaralingam B, Hermans T. 2017. Relaxed-rigidity constraints: in-grasp manipulation using purely kinematic trajectory optimization. In *Robotics: Science and Systems XIII*, ed. N Amato, S Srinivasa, N Ayanian, S Kuindersma, chap. 15. N.p.: Robot. Sci. Syst. Found.
61. Antonova R, Cruciani S, Smith C, Kragic D. 2017. Reinforcement learning for pivoting task. arXiv:1703.00472 [cs.RO]
62. Sintov A, Tslil O, Shapiro A. 2016. Robotic swing-up regrasping manipulation based on the impulse-momentum approach and CLQR control. *IEEE Trans. Robot.* 32:1079–90
63. Viña B FE, Karayiannidis Y, Smith C, Kragic D. 2016. Adaptive control for pivoting with visual and tactile feedback. In *2016 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 399–406. New York: IEEE
64. Daffle NC, Rodriguez A, Paolini R, Tang B, Srinivasa SS, et al. 2014. Extrinsic dexterity: in-hand manipulation with external forces. In *2014 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1578–85. New York: IEEE
65. Chavan-Daffle N, Rodriguez A. 2015. Prehensile pushing: in-hand manipulation with push-primitives. In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 6215–22. New York: IEEE
66. Holladay A, Paolini R, Mason MT. 2015. A general framework for open-loop pivoting. In *2015 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 3675–81. New York: IEEE

67. Cruciani S, Smith C, Kragic D, Hang K. 2018. Dexterous manipulation graphs. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 2040–47. New York: IEEE
68. Chavan-Dafle N, Mason MT, Staab H, Rossano G, Rodriguez A. 2015. A two-phase gripper to reorient and grasp. In *2015 IEEE International Conference on Automation Science and Engineering (CASE)*, pp. 1249–55. New York: IEEE
69. Bircher WG, Dollar AM, Rojas N. 2017. A two-fingered robot gripper with large object reorientation range. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 3453–60. New York: IEEE
70. Kemp CC, Edsinger A. 2006. *Robot manipulation of human tools: autonomous detection and control of task relevant features*. Paper presented at the 5th International Conference on Development and Learning, Bloomington, IN, May 31–June 3
71. Hoffmann H, Chen Z, Earl D, Mitchell D, Salemi B, Sinapov J. 2014. Adaptive robotic tool use under variable grasps. *Robot. Auton. Syst.* 62:833–46
72. Yamazaki K, Watanabe Y, Nagahama K, Okada K, Inaba M. 2010. Recognition and manipulation integration for a daily assistive robot working on kitchen environments. In *2010 IEEE International Conference on Robotics and Biomimetics*, pp. 196–201. New York: IEEE
73. Seib V, Manthe S, Holzmann J, Memmesheimer R, Peters A, et al. 2015. *RoboCup 2015 - homer@UniKoblenz (Germany)*. Tech. Rep., Univ. Koblenz-Landau, Koblenz, Ger.