This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Data-Driven Grasp Synthesis—A Survey

Jeannette Bohg, *Member, IEEE*, Antonio Morales, *Member, IEEE*, Tamim Asfour, *Member, IEEE*, and Danica Kragic, *Senior Member, IEEE*

***Abstract*—We review the work on data-driven grasp synthesis and the methodologies for sampling and ranking candidate grasps. We divide the approaches into three groups based on whether they synthesize grasps for known, familiar, or unknown objects. This structure allows us to identify common object representations and perceptual processes that facilitate the employed data-driven grasp synthesis technique. In the case of known objects, we concentrate on the approaches that are based on object recognition and pose estimation. In the case of familiar objects, the techniques use some form of a similarity matching to a set of previously encountered objects. Finally, for the approaches dealing with unknown objects, the core part is the extraction of specific features that are indicative of good grasps. Our survey provides an overview of the different methodologies and discusses open problems in the area of robot grasping. We also draw a parallel to the classical approaches that rely on analytic formulations.**

***Index Terms*—Grasp planning, grasp synthesis, object grasping and manipulation, object recognition and classification, visual perception, visual representations.**

I. Introduction

GIVEN an object, *grasp synthesis* refers to the problem of finding a grasp configuration that satisfies a set of criteria relevant for the grasping task. Finding a suitable grasp among the infinite set of candidates is a challenging problem and has been addressed frequently in the robotics community, resulting in an abundance of approaches.

In the recent review of Sahbani *et al.* [1], the authors divide the methodologies into *analytic* and *empirical*. Following Shimoga [2], analytic refers to methods that construct force-closure grasps with a multifingered robotic hand that are *dexterous*, in *equilibrium*, *stable*, and exhibit a certain *dynamic behavior*. Grasp synthesis is then usually formulated as a constrained optimization problem over criteria that measure one or several of these four properties. In this case, a grasp is typically defined by the *grasp map* that transforms the forces exerted at a set of contact points to object wrenches [3]. The criteria are based on geometric, kinematic, or dynamic formulations. Analytic formulations toward grasp synthesis have also been reviewed by Bicchi and Kumar [4].

Empirical or *data-driven* approaches rely on sampling grasp candidates for an object and ranking them according to a specific metric. This process is usually based on some existing grasp experience that can be a heuristic or is generated in simulation or on a real robot. Kamon *et al.* [5] refer to this as the *comparative* and Shimoga [2] as the *knowledge-based* approach. Here, a grasp is commonly parameterized in [6] and [7]:

1) the grasping point on the object with which the *tool center point* should be aligned;
2) the *approach vector* which describes the 3-D angle with which the robot hand approaches the grasping point;
3) the wrist orientation of the robotic hand;
4) an initial finger configuration.

Data-driven approaches differ in how the set of grasp candidates is sampled, how the grasp quality is estimated, and how good grasps are represented for future use. Some methods measure the grasp quality based on analytic formulations, but more commonly, they encode, e.g., human demonstrations, perceptual information, or semantics.

A. Brief Overview of Analytic Approaches

Analytic approaches provide guarantees regarding the criteria that measure the previously mentioned four grasp properties. However, these are usually based on assumptions such as simplified contact models, Coulomb friction, and rigid body modeling [3], [8]. Although these assumptions render grasp analysis practical, inconsistencies and ambiguities, especially regarding the analysis of grasp dynamics are usually attributed to their approximate nature.

In this context, Bicchi and Kumar [4] identified the problem of finding an accurate and tractable model of contact compliance as particularly relevant. This is needed to analyze statically indeterminate grasps in which not all internal forces can be controlled. This case arises, e.g., for underactuated hands or grasp synergies, where the number of the controlled degrees of freedom (DOF) is fewer than the number of contact forces. Prattichizzo *et al.* [9] model such a system by introducing a set of springs at the contacts and joints and show how its dexterity can be analyzed. Rosales *et al.* [10] adopt the same model of compliance to synthesize feasible and prehensile grasps. In this case, only statically determinate grasps are considered. The problem of finding a suitable hand configuration is casted as a constrained optimization problem in which a compliance is introduced to simultaneously address the constraints of contact reachability, object restraint, and force controllability. As is the

Manuscript received March 17, 2013; accepted October 25, 2013. This paper was recommended for publication by Associate Editor J. Dai and Editor D. Fox upon evaluation of the reviewers' comments. This work has been supported by FLEXBOT (FP7-ERC-279933).

J. Bohg is with the Autonomous Motion Department at the MPI for Intelligent Systems, Tübingen 70569, Germany (e-mail: jbohg@tuebingen.mpg.de).

A. Morales is with the Robotic Intelligence Lab, Universitat Jaume I, Castelló 12071, Spain (e-mail: Antonio.Morales@uji.es).

T. Asfour is with the Karlsruhe Institute of Technology, Karlsruhe 76131, Germany (e-mail: asfour@kit.edu).

D. Kragic is with the Centre for Autonomous Systems, Computational Vision and Active Perception Lab, Royal Institute for Technology, Stockholm 100 44, Sweden (e-mail: dank@kth.se).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TRO.2013.2289018

case with many other analytic approaches toward grasp synthesis, the proposed model is only studied in simulation where accurate models of the hand kinematics, the object, and their relative alignment are available.

In practice, systematic and random errors are inherent to a robotic system and are due to noisy sensors and inaccurate models of the robot's kinematics and dynamics, sensors, or of the object. The relative position of object and hand can therefore only be known approximately which makes an accurate placement of the fingertips difficult. In 2000, Bicchi and Kumar [4] identified a lack of approaches toward synthesizing grasps that are robust to positioning errors. One line of research in this direction explores the concept of *independent contact regions* as defined by Nguyen [11]: a set of regions on the object in which each finger can be independently placed anywhere without the grasp loosing the force-closure property. Several examples to compute them are presented by Roa and Suárez [12] or Krug *et al*. [13]. Another line of research toward robustness against inaccurate end-effector positioning makes use of the caging formulation. Rodriguez *et al*. [14] found that there are caging configurations of a three-fingered manipulator around a planar object that are specifically suited as a waypoint to grasping it. Once the manipulator is in such a configuration, either opening or closing the fingers is guaranteed to result in an equilibrium grasp without the need for accurate positioning of the fingers. Seo *et al*. [15] exploited the fact that two-fingered immobilizing grasps of an object are always preceded by a caging configuration. Full body grasps of planar objects are synthesized by first finding a two-contact caging configuration and then using additional contacts to restrain the object. Results have been presented in simulation and demonstrated on a real robot.

Another assumption commonly made in analytic approaches is that the precise geometric and physical models of an object are available to the robot, which is not always the case. In addition, we may not know the surface properties or friction coefficients, weight, center of mass, and weight distribution. Some of these can be retrieved through interaction: Zhang and Trinkle [16] propose to use a particle filter to simultaneously estimate the physical parameters of an object and track it while it is being pushed. The dynamic model of the object is formulated as a mixed nonlinear complementarity problem. The authors show that even when the object is occluded and the state estimate cannot be updated through visual observation, the motion of the object is accurately predicted over time. Although methods like this relax some of the assumptions, they are still limited to simulation [10], [14] or consider 2-D objects [14]–[16].

B. Development of Data-Driven Methods

Up to the year 2000, the field of robotic grasping[1] was clearly dominated by analytic approaches [2], [4], [11], [17]. Apart from, e.g., [5], data-driven grasp synthesis started to become popular with the availability of Graspit! [18] in 2004. Many highly cited approaches have been developed, analyzed, and evaluated in this or other simulators [19]–[24]. These approaches differ in how grasp candidates are sampled from the infinite space of possibilities. For grasp ranking, they rely on classical metrics that are based on analytic formulations such as the widely used ϵ-metric proposed in Ferrari and Canny [17]. It constructs the *grasp wrench space* (GWS) by computing the convex hull over the wrenches at the contact points between the hand and the object. ϵ quantifies the quality of a force-closure grasp by the radius of the maximum sphere still fully contained in the GWS.

Developing and evaluating approaches in simulation is attractive because the environment and its attributes can be completely controlled. A large number of experiments can be efficiently performed without having access to expensive robotics hardware that would also add a lot of complexity to the evaluation process. However, it is not clear if the simulated environment resembles the real world well enough to transfer methods easily. Only recently, several works [24], [39], [40] have analyzed this question and came to the conclusion that the classic metrics are not good predictors for grasp success in the real world. They do not seem to cope well with the challenges arising in unstructured environments. Diankov [24] claims that in practice grasps synthesized using these metrics tend to be relatively *fragile*. Balasubramanian *et al*. [39] systematically tested a number of grasps in the real world that were stable according to classical grasp metrics. Compared with grasps planned by humans and transferred to a robot by kinesthetic teaching on the same objects, they underperformed significantly. A similar study has been conducted by Weisz and Allen [40]. It focuses on the ability of the ϵ-metric to predict grasp stability under object pose error. The authors found that it performs poorly, especially when grasping large objects.

As pointed out by Bicchi and Kumar [4] and Prattichizzo and Trinkle [8], grasp closure is often wrongly equated with stability. Closure states the existence of equilibrium which is a necessary but not a sufficient condition. Stability can only be defined when considering the grasp as a dynamical system and in the context of its behavior when perturbed from an equilibrium. Seen in this light, the results of the aforementioned studies are not surprising. However, they suggest that there is a large gap between reality and the models for grasping that are currently available and tractable.

For this reason, several researchers [25]–[27] proposed to let the robot learn how to grasp by experience that is gathered during grasp execution. Although collecting examples is extremely time-consuming, the problem of transferring the learned model to the real robot is nonexistant. A crucial question is how the object to be grasped is represented and how the experience is generalized to novel objects.

Saxena *et al.* [28] pushed machine learning approaches for data-driven grasp synthesis even further. A simple logistic regressor was trained on large amounts of synthetic, labeled training data to predict good grasping points in a monocular image. The authors demonstrated their method in a household scenario in which a robot emptied a dishwasher. None of the classical

[1]Citation counts for the most influential articles in the field. Extracted from scholar.google.com in October 2013. [11]: 733. [4]: 490. [17]: 477. [2]: 405. [5]: 77. [18]: 384. [19]: 353. [20]: 100. [21]: 110. [22]: 95. [23]: 96. [24]: 108. [25]: 38. [26]: 156. [27]: 39. [28]: 277. [29]: 75. [30]: 40. [31]: 21. [32]: 43. [33]: 77. [34]: 26. [35]: 191. [36]: 58. [37]: 75. [38]: 39.

principles that are based on analytic formulations were used. This paper spawned a lot of research [29]–[32] in which essentially one question is addressed: What are the object features that are sufficiently discriminative to infer a suitable grasp configuration?

From 2009, there were further developments in the area of 3-D sensing. Projected Texture Stereo was proposed by Konolige [41]. This technology is built into the sensor head of the PR2 [42], a robot that is available to comparatively many robotics research labs and running on the OpenSource middleware ROS [43]. In 2010, Microsoft released the Kinect [44], a highly accurate depth-sensing device that is based on the technology developed by PrimeSense [45]. Due to its low price and simple usage, it became a ubiquitous device within the robotics community. Although the importance of 3-D data to grasp has been previously recognized, many new approaches were proposed that operate on real world 3-D data. They are either heuristics that map structures in this data to grasp configurations directly [33], [34] or they try to detect and recognize objects and estimate their pose [35], [46].

C. Analytic Versus Data-Driven Approaches

Contrary to analytic approaches, methods following the data-driven paradigm place more weight on the object representation and the perceptual processing, e.g., feature extraction, similarity metrics, object recognition or classification, and pose estimation. The resulting data is then used to retrieve grasps from some knowledge base or sample and rank them by comparison to existing grasp experience. The parameterization of the grasp is less specific (e.g., an approach vector instead of fingertip positions) and, therefore, accommodates for uncertainties in perception and execution. This provides a natural precursor to reactive grasping [33], [47]–[50], which, given a grasp hypothesis, considers the problem of robustly acquiring it under uncertainty. Data-driven methods cannot provide guarantees regarding the aforementioned criteria of dexterity, equilibrium, stability, and dynamic behavior [2]. They can only be verified empirically. However, they form the basis for studying grasp dynamics and further developing analytic models that better resemble reality.

D. Classification of Data-Driven Approaches

Sahbani *et al.* [1] divide the data-driven methods that are based on whether they employ object features or observation of humans during grasping. We believe that this falls short of capturing the diversity of these approaches especially in terms of the ability to transfer grasp experience between similar objects and the role of perception in this process. In this survey, we propose to group data-driven grasp synthesis approaches based on what they assume to know *a priori* about the query object:

1) *Known Objects*: These approaches assume that the query object has been encountered before and that grasps have already been generated for it. Commonly, the robot has access to a database containing geometric object models that are associated with a number of good grasps. This database is usually built offline and, in the following, will be referred to as an *experience database*. Once the object has been recognized, the goal is to estimate its pose and retrieve a suitable grasp.
2) *Familiar Objects*: Instead of exact identity, the approaches in this group assume that the query object is similar to the previously encountered ones. New objects can be *familiar* on different levels. Low-level similarity can be defined in terms of shape, color, or texture. High-level similarity can be defined based on the object category. These approaches assume that new objects similar to old ones can be grasped in a similar way. The challenge is to find an object representation and a similarity metric that allows to transfer grasp experience.
3) *Unknown Objects*: Approaches in this group do not assume to have access to object models or any sort of grasp experience. They focus on identifying the structure or features in sensory data for generating and ranking grasp candidates. These are usually based on local or global features of the object as perceived by the sensor.

We find the previous classification suitable for surveying the data-driven approaches since the assumed prior object knowledge determines the necessary perceptual processing and associated object representations for generating and ranking grasp candidates. For known objects, the problems of recognition and pose estimation have to be addressed. The object is usually represented by a complete geometric 3-D object model. For familiar objects, an object representation has to be found that is suitable for comparing them to already encountered object in terms of graspability. For unknown objects, heuristics have to be developed for the directly linking structure in the sensory data to candidate grasps.

Only a minority of the approaches discussed in this survey cannot be clearly classified to belong to one of these three groups. Most of the included papers use sensor data from the scene to perform data-driven grasp synthesis and are part of a real robotic system that can execute grasps.

Finally, this classification is well in line with the research in the field of neuroscience, specifically, with the theory of the dorsal and ventral stream in human visual processing [51]. The *dorsal* pathway processes immediate action-relevant features, while the *ventral* pathway extracts context- and scene-relevant information and is related to object recognition. The visual processing in the ventral and dorsal pathways can be related to the grouping of grasp synthesis for familiar/known and unknown objects, respectively. The details of such links are out of the scope of this paper. Extensive and detailed reviews on the neuroscience of grasping are offered in [52]–[54].

E. Aspects Influencing the Generation of Grasp Hypotheses

The number of *candidate grasps* that can be applied to an object is infinite. To sample some of these candidates and define a quality metric for selecting a good subset of *grasp hypotheses* is the core subject of the approaches reviewed in this survey. In addition to the prior object knowledge, we identified a number of other factors that characterize these metrics. Thereby, they influence which grasp hypotheses are selected by a method. Fig. 1 shows a mind map that structures these aspects. An

Fig. 1. We identified a number of aspects that influence how the final set of grasp hypotheses is generated for an object. The most important one is the assumed *prior object knowledge*, as discussed in Section I-D. Numerous different *object-grasp representations* are proposed in the literature that are relying on *features* of different modalities such as 2-D or 3-D vision or tactile sensors. Either local object parts or the object as a whole are linked to specific grasp configurations. *Grasp synthesis* can either be analytic or data-driven. The latter is further detailed in Fig. 2. Very few approaches explicitly address the *task* or *hand* kinematics of the robot.

important one is how the quality of a candidate grasp depends on the object, i.e., the *object-grasp representation*. Some approaches extract local object attributes (e.g., curvature, contact area with the hand) around a candidate grasp. Other approaches take global characteristics (e.g., center of mass, bounding box) and their relation to a grasp configuration into account. Dependent on the sensor device, *object features* can be based on 2-D or 3-D visual data as well as on other modalities. Furthermore, *grasp synthesis* can be analytic or data-driven. We further categorized the latter in Fig. 2; there are methods for *learning* either from *human demonstrations*, *labeled examples,* or *trial and error*. Other methods rely on various *heuristics* to directly link the structure in sensory data to candidate grasps. There is relatively little work on task-dependent grasping. In addition, the applied robotic hand is usually not in the focus of the discussed approaches. We will therefore not examine these two aspects. However, we will indicate whether an approach takes the task into account and whether an approach is developed for a gripper or for the more complex case of a multifingered hand. Tables I–III list all the methods in this survey. The table columns follow the structure proposed in Figs. 1 and 2.

II. Grasping Known Objects

If the object to be grasped is known and there is already a database of grasp hypotheses for it, then the problem of finding a feasible grasp reduces to estimating the object pose and then filtering the hypotheses by reachability. Table I summarizes all the approaches discussed in this section.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

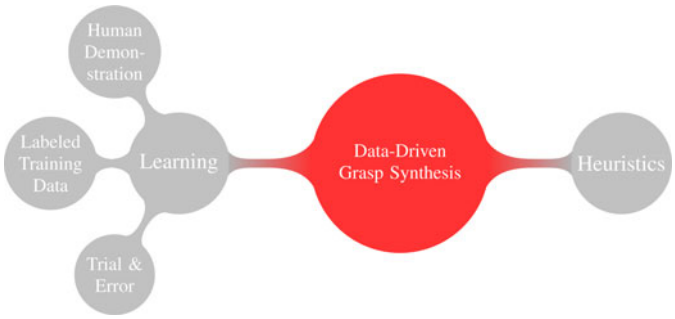


Fig. 2. Data-driven grasp synthesis can either be based on heuristics or on learning from data. The data can either be provided in the form of offline-generated labeled training data, human demonstration, or through trial and error.

TABLE I
DATA-DRIVEN APPROACHES FOR GRASPING KNOWN OBJECTS

	Object-Grasp Represen.		Object Features			Grasp Synthesis							
	Local	Global	2D	3D	Multi-Modal	Heuristic	Human Demo	Labeled Data	Trial & Error	Task	Multi-Fingered	Deformable	Real Data
Glover et al. [55]	✓		✓			✓						✓	✓
Goldfeder et al. [21]	✓			✓		✓					✓		
Miller et al. [19]	✓			✓		✓					✓		
Przybylski et al. [56]	✓			✓		✓					✓		
Roa et al. [57]	✓			✓		✓					✓		✓
Detry et al. [27]	✓				✓		✓		✓				✓
Detry et al. [58]	✓				✓	✓			✓				✓
Huebner et al. [59]	✓				✓	✓				✓	✓		✓
Diankov [24]	✓	✓		✓		✓					✓		
Balasubramanian et al. [39]		✓		✓		✓	✓			✓	✓		✓
Borst et al. [22]		✓		✓		✓					✓		
Brook et al. [60]		✓		✓		✓							✓
Ciocarlie and Allen [23]		✓		✓		✓					✓		
Romero et al. [61]		✓		✓			✓				✓		✓
Papazov et al. [62]		✓		✓		✓					✓		✓
Morales et al. [7]		✓			✓	✓					✓		✓
Collet Romea et al. [63]		✓			✓		✓				✓		✓
Kroemer et al. [64]		✓			✓		✓		✓		✓		✓
Ekvall and Kragic [6]		✓			✓		✓				✓		✓
Tegin et al. [65]		✓			✓	✓	✓				✓		
Pastor et al. [49]		✓					✓				✓		✓
Stulp et al. [66]		✓					✓		✓		✓		✓

TABLE II
DATA-DRIVEN APPROACHES FOR GRASPING FAMILIAR OBJECTS

	Object-Grasp Represen.		Object Features			Grasp Synthesis							
	Local	Global	2D	3D	Multi-Modal	Heuristic	Human Demo	Labeled Data	Trial & Error	Task	Multi-Fingered	Deformable	Real Data
Song et al. [87]	✓			✓				✓		✓	✓		
Li and Pollard [88]	✓			✓			✓				✓		
El-Khoury and Sahbani [89]	✓			✓				✓			✓		
Hübner and Kragic [67]	✓			✓				✓		✓	✓		✓
Kroemer et al. [90]	✓			✓			✓		✓	✓	✓		✓
Detry et al. [91]	✓			✓				✓			✓		
Detry et al. [92]	✓			✓				✓			✓		✓
Herzog et al. [71]	✓			✓			✓		✓		✓		✓
Ramisa et al. [93]	✓			✓				✓			✓	✓	✓
Boularias et al. [94]	✓			✓				✓			✓		✓
Montesano and Lopes [95]	✓		✓						✓		✓		✓
Stark et al. [30]	✓		✓				✓	✓		✓			✓
Saxena et al. [28]	✓		✓					✓					✓
Saxena et al. [29]	✓				✓			✓			✓		✓
Fischinger and Vincze [96]	✓			✓				✓					✓
Le et al. [31]	✓	✓			✓			✓			✓		✓
Bergström et al. [97]	✓	✓			✓	✓		✓			✓		✓
Hillenbrand and Roa [98]	✓	✓		✓				✓		✓	✓		
Bohg and Kragic [32]		✓			✓			✓			✓		✓
Bohg et al. [99]		✓			✓			✓		✓	✓		✓
Curtis and Xiao [100]		✓			✓				✓		✓		
Goldfeder and Allen [101]		✓			✓			✓			✓		✓
Marton et al. [102]		✓			✓	✓		✓			✓		✓
Rao et al. [103]		✓			✓	✓					✓		✓
Speth et al. [104]		✓			✓				✓		✓		✓
Madry et al. [105]		✓			✓			✓		✓	✓		✓
Kamon et al. [5]		✓	✓						✓				✓
Montesano et al. [26]		✓	✓						✓		✓		✓
Morales et al. [25]		✓	✓						✓		✓		✓
Pelossof et al. [20]		✓		✓				✓			✓		
Dang and Allen [106]		✓		✓				✓		✓	✓		✓

TABLE III
DATA-DRIVEN APPROACHES FOR GRASPING UNKNOWN OBJECTS

	Object-Grasp Represen.		Object Features			Grasp Synthesis							
	Local	Global	2D	3D	Multi-Modal	Heuristic	Human Demo	Labeled Data	Trial & Error	Task	Multi-Fingered	Deformable	Real Data
Kraft et al. [116]	✓				✓	✓							✓
Popović et al. [117]	✓				✓	✓					✓		✓
Bone et al. [118]	✓			✓		✓							✓
Richtsfeld and Vincze [119]	✓			✓		✓							✓
Maitin-Shepard et al. [37]	✓			✓		✓						✓	✓
Hsiao et al. [33]	✓	✓		✓		✓							✓
Brook et al. [60]	✓	✓		✓		✓							✓
Bohg et al. [120]		✓		✓		✓					✓		✓
Stückler et al. [121]		✓		✓		✓							✓
Klingbeil et al. [34]		✓		✓		✓							✓
Maldonado et al. [122]		✓		✓		✓					✓		✓
Marton et al. [110]		✓		✓		✓					✓		✓
Lippiello et al. [123]		✓		✓		✓					✓		✓
Dunes et al. [124]		✓	✓			✓							✓
Kehoe et al. [125]		✓	✓			✓							✓
Morales et al. [126]		✓	✓			✓					✓		✓

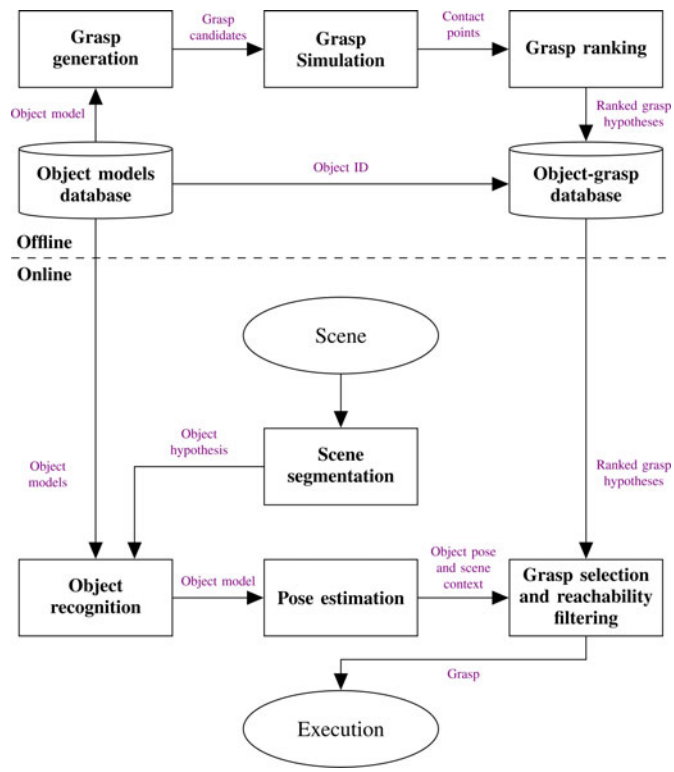


Fig. 3. Typical functional flow-chart for a system with offline generation of a grasp database. In the offline phase, every object model is processed to generate grasp candidates. Their quality is evaluated for ranking. Finally, the list of grasp hypotheses is stored with the corresponding object model. In the online phase, the scene is segmented to search and recognize object models. If the process succeeds, the associated grasp hypotheses are retrieved, and the unreachable ones are discarded. Most of the following approaches can be summarized with this flowchart. Some of them only implement the offline part. [7], [19], [21]–[24], [39], [56], [57], [59], [60], [65].

A. Offline Generation of a Grasp Experience Database

First, we look at approaches for generating the experience database. Figs. 3 and 5 summarize the typical functional flowchart of these type of approaches. Each box represents a processing step. Note that these figures are abstractions that summarize the implementations of a number of papers. Most

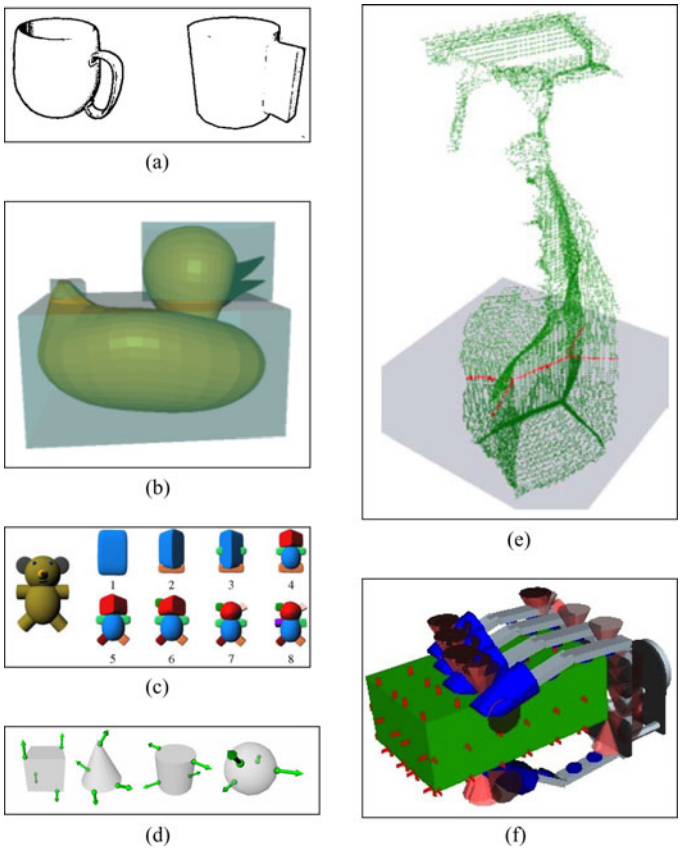


Fig. 4. Generation of grasp candidates through object shape approximation with primitives or through sampling. (a) Primitive Shape Decomposition [19]. (b) Box Decomposition [67]. (c) SQ Decomposition [21]. (d) Randomly-sampled grasp hypotheses. [22]. (e) Green: Centers of a union of spheres. Red: Centers at a slice through the model [56], [68]. (f) Grasp candidate sampled based on surface normals and bounding box [69].

reviewed papers focus on a single module. This is also true for similar figures appearing in Sections III and IV.

1) 3-D Mesh Models and Contact-Level Grasping: Many approaches in this category assume that a 3-D mesh of the object is available. The challenge is then to automatically generate a set of good grasp hypotheses. This involves sampling the infinite space of possible hand configurations and ranking the resulting candidate grasps according to some quality metric. The major part of the approaches discussed in the following uses force-closure grasps and ranks them according to the previously discussed ϵ-metric. They differ mostly in the way the grasp candidates are sampled. Fig. 3 shows a flowchart of which specifically the upper part (offline) visualizes the data flow for the following approaches.

Some of them approximate the object's shape with a constellation of primitives such as spheres, cones, cylinders, and boxes as in [19], Hübner and Kragic [67] and Przybylski *et al.* [56] or superquadrics (SQ) as in [21]. These shape primitives are then used to limit the amount of candidate grasps and thus prune the search tree for finding the best grasp hypotheses. Examples for these approaches are shown in Fig. 4(a)–(c) and (e). Borst *et al*. [22] reduce the number of candidate grasps by randomly generating a number of them that are dependent on the object surface and filter them with a simple heuristic. The authors show that this approach works well if the goal is not to find an optimal grasp but, instead, a fairly good grasp that works well for "everyday tasks." Diankov [24] proposes to sample grasp candidates dependent on the objects bounding box in conjunction with surface normals. Grasp parameters that are varied are the distance between the palm of the hand and the grasp point as well as the wrist orientation. The authors find that usually a relatively small amount of 30% from all grasp samples is in force closure. Examples for these sampling approaches are shown in Fig. 4(d) and (f). Roa *et al.* [57] present an approach toward synthesizing power grasps that is not based on evaluating the force-closure property. Slices through the object model and perpendicular to the axes of the bounding box are sampled. The ones that best resemble a circle are chosen for synthesizing a grasp.

All these approaches are developed and evaluated in simulation. As claimed by, e.g., Diankov [24], the biggest criticism toward ranking grasps based on force closure and the ϵ-metric is that relatively *fragile grasps* might be selected. A common approach to filter these is to add noise to the grasp parameters and keep only those grasps in which a certain percentage of the neighboring candidates also yield force closure. Weisz and Allen [40] followed a similar approach that focuses in particular on the ability of the ϵ-metric to predict grasp stability under object pose uncertainty. For a set of object models, the authors used Graspit! [18] to generate a set of grasp candidates in the force closure. For each object, pose uncertainty is simulated by perturbing it in three DOF. Each grasp candidate was then reevaluated according to the probability of attaining a force-closure grasp. The authors found that their proposed metric performs in a superior way, especially on large objects.

Balasubramaniam *et al.* [39] question classical grasp metrics in principle. The authors systematically tested a number of task-specific grasps in the real world that were stable according to classical grasp metrics. These grasps underperformed significantly when compared with grasps planned by humans through kinesthetic teaching on the same objects and for the same tasks. The authors found that humans optimize a *skewness* metric, i.e., the divergence of alignment between hand and principal object axes.

2) Learning From Humans: A different way to generate grasp hypotheses is to observe how humans grasp an object. This is usually done offline following the flowchart in Fig. 5. This process produces an experience database that is exploited online in a similar fashion as depicted in Fig. 3.

Ciocarlie and Allen [23] exploit results from neuroscience that showed that a human hand control takes place in a space of much lower dimension than the hand's DOF. This finding was applied to directly reduce the configuration space of a robotic hand to find pregrasp postures. From these so-called *eigengrasps,* the system searches for stable grasps.

Detry *et al.* [27] model the object as a constellation of local multimodal contour descriptors. Four elementary grasping actions are associated with specific constellations of these features, resulting in an abundance of grasp candidates. They are modeled as a nonparametric density function in the space of 6-D gripper poses, which are referred to as a *bootstrap* density.

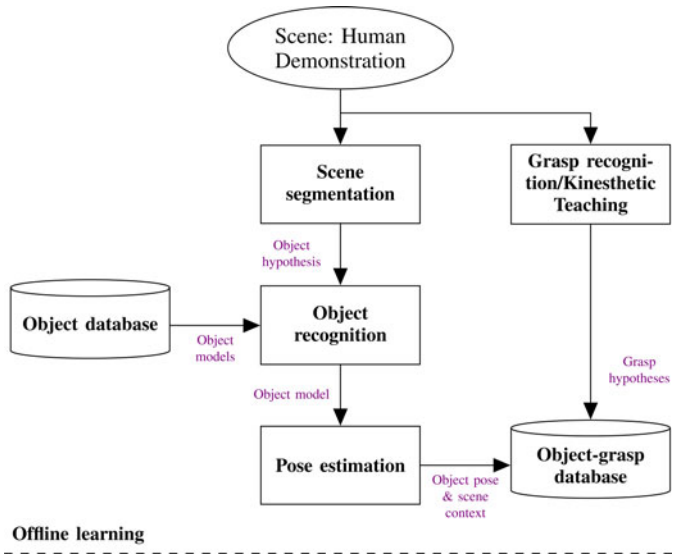


Fig. 5. Typical functional flowchart of a system that learns from human demonstration. The robot observes a human grasping a known object. Two perceptual processed are followed in parallel. On the left, the object is recognized. On the right, the demonstrated grasp configuration is extracted or recognized. Finally, object models and grasps are stored together. This process could replace or complement the offline phase described in Fig. 3. The following approaches follow this approach: [27], [39], [49], [61], [64], and [66].

Human grasp examples are used to build an object specific *empirical* grasp density from which grasp hypotheses can be sampled. This is visualized in Fig. 8(f) and (g).

Kroemer *et al.* [64] represent the object with the same features as used in [27]. How to grasp specific objects is learned through a combination of a high-level reinforcement learner and a low-level reactive grasp controller. The learning process is bootstrapped through imitation learning in which a demonstrated reaching trajectory is converted into an initial policy. Similar initialization of an object-specific grasping policy is used in [49] and [66].

Romero *et al.* [61] present a system for observing humans visually, while they interact with an object. A grasp type and pose is recognized and mapped to different robotic hands in a fixed scheme. For validation of the approach in the simulator, 3-D object models are used. This approach has been demonstrated on a humanoid robot in [70]. The object is not explicitly modeled. Instead, it is assumed that human and robot act on the same object in the same pose.

In the method presented by Ekvall and Kragic [6], a human demonstrator wearing a magnetic-tracking device is observed while manipulating a specific object. The grasp type is recognized and mapped through a fixed schema to a set of robotic hands. Given the grasp type and the hand, the best approach vector is selected from an offline trained experience database. Unlike Detry *et al*. [27] and Romero *et al.* [61], the approach vector that is used by the demonstrator is not adopted. Ekvall and Kragic [6] assume that the object pose is known. Experiments are conducted with a simulated pose error. No physical experiments have been demonstrated. Examples for the aforementioned ways to teach a robot grasping by demonstration are shown in Fig. 6.

3) Learning Through Trial and Error: Instead of adopting a fixed set of grasp candidates for a known object, the following approaches try to refine them by *trial and error*. In this case, there is no separation between offline learning and online exploitation, as can be seen in Fig. 7. Kroemer *et al.* [64] and Stulp *et al.* [66] apply reinforcement learning to improve an initial human demonstration. Kroemer *et al.* [64] uses a low-level reactive controller to perform the grasp that informs the high-level controller with reward information. Stulp *et al.* [66] increase the robustness of their nonreactive grasping strategy by learning the shape and goal parameters of the motion primitives that are used to model a full grasping action. Through this approach, the robot learns reaching trajectories and grasps that are robust against object pose uncertainties. Detry *et al.* [58] builds an object-specific empirical grasp density from successful grasping trials. This nonparametric density can then be used to sample grasp hypotheses.

B. Online Object Pose Estimation

In the previous section, we reviewed different approaches toward grasping known objects regarding their way to generate and rank candidate grasps. During online execution, an object has to be recognized and its pose estimated before the offline trained grasps can be executed. Furthermore, from the set of hypotheses, not all grasps might be feasible in the current scene. They have to be filtered by reachability. The lower part of Fig. 3 visualizes the data flow during grasp execution and how the offline generated data is employed.

Several of the aforementioned grasp generation methods [27], [58], [64] use the probabilistic approach toward object representation and pose estimation proposed by Detry *et al.* [72], as visualized in Fig. 8(e). Grasps are either selected by sampling from densities [27], [58], or a grasp policy refined from a human demonstration is applied [64]. Morales *et al*. [7] use the method proposed by Azad *et al.* [73] to recognize an object and estimate its pose from a monocular image as shown in Fig. 8(a). Given this information, an appropriate grasp configuration can be selected from a grasp experience database that has been acquired offline. The whole system is demonstrated on the robotic platform described in [74]. Huebner *et al.* [59] demonstrate grasping of known objects on the same humanoid platform and use the same method for object recognition and pose estimation. The offline selection of grasp hypotheses is based on a decomposition into boxes, as described in Hübner and Kragic [67]. Task constraints are taken into account by reducing the set of box faces that provide valid approach directions. These constraints are hard-coded for each task. Ciocarlie *et al.* [75] propose a robust grasping pipeline in which the known object models are fitted to point cloud clusters using the standard ICP [76]. The search space of potential object poses is reduced by assuming a dominant plane and rotationally symmetric objects that are always standing upright as, e.g., shown in Fig. 8(b). Papazov *et al.* [62] demonstrate their previous approach on 3-D object recognition and pose estimation [77] in a grasping scenario. Multiple objects in cluttered scenes can be robustly recognized and their pose estimated. No assumption is made about the geometry of the scene, the shape of the objects, or their pose.

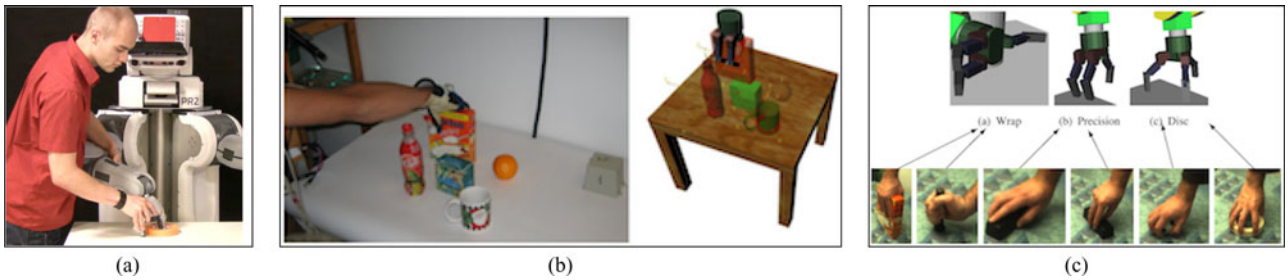


Fig. 6. Robot grasp learning from human demonstration. (a) Kinesthetic Teaching [71]. (b) Human-to-robot mapping of grasps using a data glove [6]. (c) Human-to-robot mapping of grasps using visual grasp recognition [61].

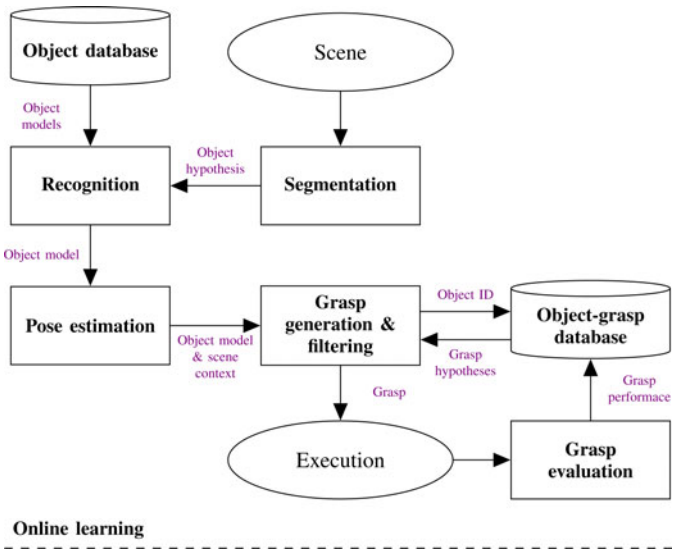


Fig. 7. Typical functional flowchart of a system that learns through trial and error. First, a known object in the scene is segmented and recognized. Past experiences with that object are retrieved, and a new grasp hypothesis is generated or selected among the already tested ones. After execution of the selected grasp, the performance is evaluated, and the memory of past experiences with the object is updated. The following approaches use trial-and-error learning: [27], [58], [64], [66].

The aforementioned methods assume *a priori* known rigid 3-D object model. Glover *et al.* [55] consider known deformable objects. Probabilistic models of their 2-D shape are learned offline. The objects can then be detected in monocular images of cluttered scenes, even when they are partially occluded. The visible object part serve as a basis for planning a grasp under consideration of the global object shape. An example for a successful detection is shown in Fig. 8(c).

Collet Romea *et al.* [78] use a combination of 2-D and 3-D features as an object model. Examples for objects from an earlier version of the system [63] are shown in Fig. 8(d). The authors estimate the object's pose in a scene from a single image. The accuracy of their approach is demonstrated through a number of successful grasps.

III. Grasping Familiar Objects

The idea of addressing the problem of grasping *familiar* objects originates from the observation that many of the objects in the environment can be grouped together into categories with common characteristics. In the computer vision community, objects within one category usually share similar visual properties. These can be, e.g., a common texture [79] or shape [80], [81], the occurrence of specific local features [82], [83], or their specific spatial constellation [84], [85]. These categories are usually referred to as *basic level categories* and emerged from the area of cognitive psychology [86].

For grasping and manipulation of objects, a more natural characteristic may be the functionality that they afford [30], similar objects are grasped in a similar way or may be used to fulfill the same task (pouring, rolling, etc). The difficulty is to find a representation that encodes these common affordances. Given the representation, a similarity metric has to be found under which objects of the same functionality can be considered to be alike. The approaches discussed in this survey are summarized in Table II. All of them employ learning mechanisms and showed that they can generalize the grasp experience on training data to new but familiar objects.

A. Discriminative Approaches

First, there are approaches that learn a discriminative function to distinguish between good and bad grasp configurations. They mainly differ in what object features are used and, thereby, in the space over which objects are considered similar. Furthermore, they parameterize grasp candidates differently. Many of them only consider whether a specific part of the object is graspable or not. Others also learn multiple contact points or full grasp configurations. A flowchart for the approaches discussed in the following is presented in Fig. 9.

1) Based on 3-D Data: El-Khoury and Sahbani [89] distinguish between graspable and nongraspable parts of an object. A point cloud of an object is segmented into parts. Each part is approximated by an SQ. An artificial neural network (ANN) is used to classify whether or not the part is prehensile. The ANN is trained offline on human-labeled SQs. If one of the object parts is classified as prehensile, then an n-fingered force-closure grasp is synthesized on this object part. Grasp experience is, therefore, only used to decide where to apply a grasp and not how the grasp should be configured. These steps are shown for two objects in Fig. 10.

Pelossof *et al.* [20] approximate an object with a single SQ. Given this, their goal is to find a suitable grasp configuration for a Barrett hand consisting of the approach vector, wrist orientation, and finger spread. A *Support Vector Machine* (SVM)

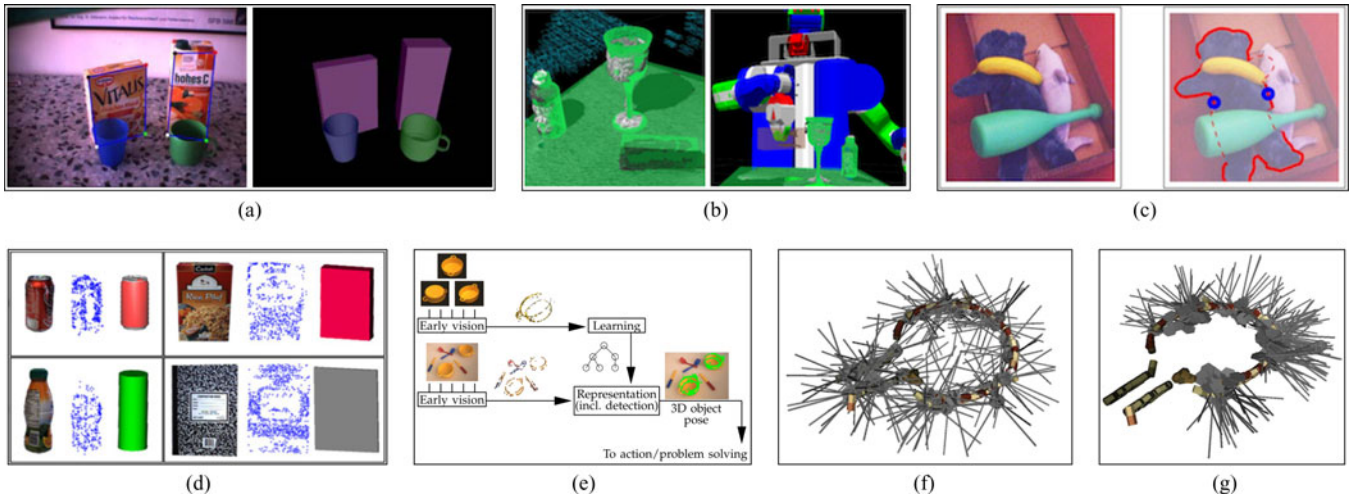


Fig. 8. Object representations for grasping and corresponding methods for pose estimation. (a) Object pose estimation of textured and untextured objects in monocular images [73]. (b) ICP-based object pose estimation from segmented point clouds [75]. (c) Deformable object detection and pose estimation in monocular images [55]. (d) Multiview object representation composed of 2-D and 3-D features [63]. (e) Probabilistic and hierarchical approach towards object pose estimation [72]. (f) Grasp candidates linked to groups of local contour descriptors [27]. (g) Empirical grasp density built by trial and error [27].

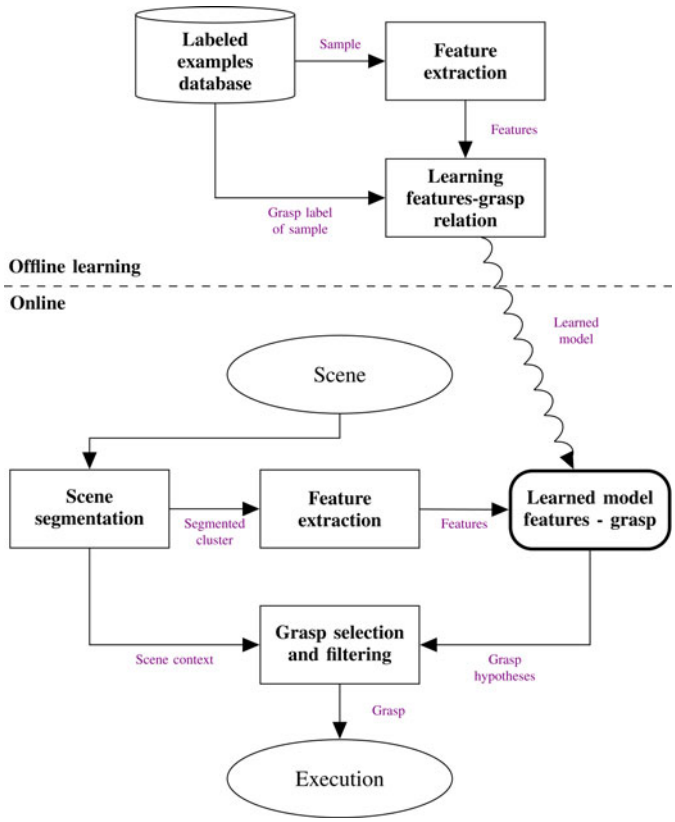


Fig. 9. Typical functional flowchart of a system that learns from labeled examples. In the offline learning phase, a database is available, consisting of a set of objects labeled with grasp configurations and their quality. Database entries are analyzed to extract relations between specific features and the grasps. The result is a learned model that, given some features, can predict grasp qualities. In the online phase, the scene is segmented, and features are extracted from the scene. Given this, the model outputs a ranked set of promising grasp hypotheses. Unreachable grasps are filtered out, and the best is executed. The following approaches use labeled training examples: [20], [28]–[32], [67], [87]–[89], [91]–[94], [97]–[102], [105], [106].

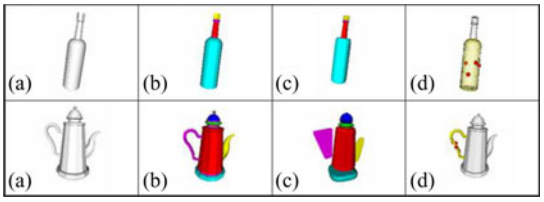


Fig. 10. (a) Object model. (b) Part segmentation. (c) SQ approximation. (d) Graspable part and contact points [89].

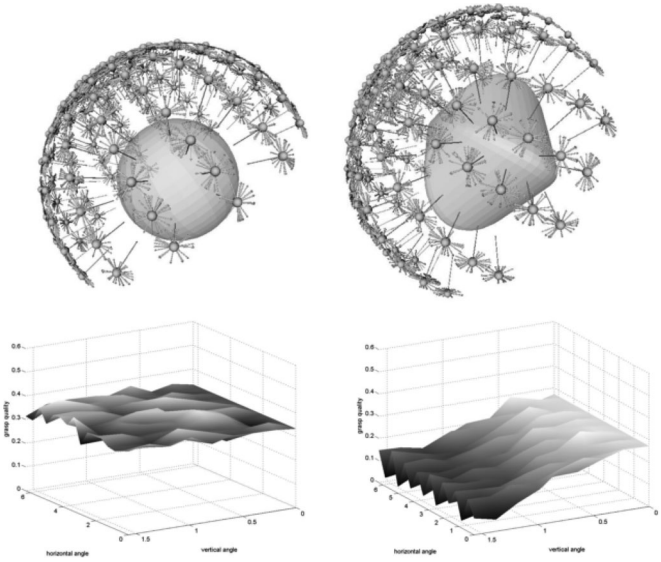


Fig. 11. (Top) Grasp candidates performed on SQ. (Bottom) Grasp quality for each candidate [20].

is trained on data consisting of feature vectors containing the SQ parameters and a grasp configuration. They are labeled with a scalar estimating the grasp quality. These training data are shown in Fig. 11. When feeding the SVM only with the shape

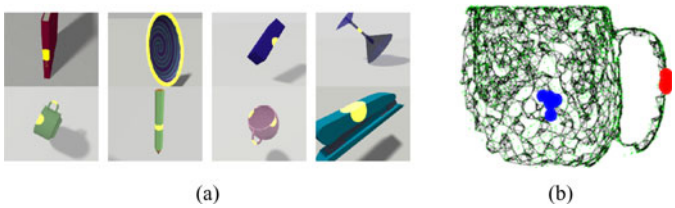


Fig. 12. Labeled training data. (a) One example for each of the eight object classes in training data in [28] along with their grasp labels (in yellow). (b) Positive (red) and negative examples (blue) for grasping points [94].

parameters of the SQ, their algorithm searches efficiently through the grasp configuration space for parameters that maximize the grasp quality.

Both of the aforementioned approaches are evaluated in simulation where the central assumption is that accurate and detailed 3-D object models are available: an assumption, that is not always valid. An SQ is an attractive 3-D representation due to its low number of parameters and high-shape variability. However, it remains unclear whether an SQ could equally well approximate object shape when given real-world sensory data that are noisy and incomplete.

Hübner and Kragic [67] decomposed a point cloud into a constellation of boxes. The simple geometry of a box reduces the number of potential grasps significantly. A hand-designed mapping between simple box features (size and position in constellation) and grasping task is proposed. To decide which of the sides of the boxes provides a good grasp, an ANN is trained offline on synthetic data. The projection of the point cloud inside a box to its sides provides the input to the ANN. The training data consist of a set of these projections from different objects labeled with the grasp quality metrics.

Boularias *et al.* [94] model an object as a *Markov random field* (MRF) in which the nodes are points in a point cloud and edges are spanned between the six nearest neighbors of a point. The features of a node describe the local point distribution around that node. A node in the MRF can carry either of two labels: a good or a bad grasp location. The goal of the approach is to find the maximum *a posteriori* labeling of point clouds for new objects. Very little training data are used which is shown in Fig. 12(b). A handle serves as a positive example. The experiments show that this leads to a robust labeling of 3-D object parts that are very similar to a handle.

Although both approaches [67], [94] also rely on 3-D models for learning, the authors show examples for real sensor data. It remains unclear how well the classifiers would generalize to a larger set of object categories and real sensor data.

Fischinger and Vincze [96] propose a *height-accumulated* feature that is similar to Haar basis functions as successfully applied by, e.g., Viola and Jones [107] for face detection. The values of the feature are computed based on the height of objects above, e.g., the table plane. Positive and negative examples are used to train an SVM that distinguishes between good and bad grasping points. The authors demonstrate their approach for cleaning cluttered scenes. No object segmentation is required for the approach.

2) Based on 2-D Data: There are number of experience-based approaches that avoid the complexity of 3-D data and mainly rely on 2-D data to learn to discriminate between good and bad grasp locations. Saxena *et al*. [28] propose a system that infers a point at where to grasp an object directly as a function of its image. The authors apply logistic regression to train a grasping point model on labeled synthetic images of a number of different objects. The classification is based on a feature vector containing local appearance cues regarding color, texture, and edges of an image patch in several scales and of its neighboring patches. Samples from the labeled training data are shown in Fig. 12(a). The system was used successfully to pick up objects from a dishwasher after it has been additionally trained for this scenario.

Instead of assuming the availability of a labeled dataset, Montesano and Lopes [95] allow the robot to autonomously explore which features encode graspability. Similar to [28], simple 2-D filters are used that can be rapidly convolved with an image. Given features from a region, the robot can compute the posterior probability that a grasp applied to this location will be successful. It is modeled as a Beta distribution and estimated from the grasping trials executed by the robot and their outcome. Furthermore, the variance of the posterior can be used to guide exploration to regions that are predicted to have a high-success rate but are still uncertain.

Another example of a system involving 2-D data and grasp experience is presented by Stark *et al*. [30]. Here, an object is represented by a composition of prehensile parts. These so-called *affordance cues* are obtained by observing the interaction of a person with a specific object. Grasp hypotheses for new stimuli are inferred by matching features of that object against a codebook of learned *affordance cues* that are stored along with relative object position and scale. How to grasp the detected parts is not solved since hand orientation and finger configuration are not inferred from the affordance cues. Similar to [94], especially locally very discriminative structures like handles are well detected.

3) Integrating 2-D and 3-D Data: Although the previous approaches have been demonstrated to work well in specific manipulation scenarios, inferring a full grasp configuration from 2-D data alone is a highly underconstrained problem. Regions in the image may have very similar visual features but afford completely different grasps. The following approaches integrate multiple complementary modalities, 2-D and 3-D visual data, and their local or global characteristics, to learn a function that can take more parameters of a grasp into account.

Saxena *et al.* [29] extend their previous work on inferring 2-D grasping points by taking the 3-D point distribution within a sphere centered around a grasp candidate into account. This enhances the prediction of a stable grasp and allows for the inference of grasp parameters like approach vector and finger spread. In earlier work [28], only downward or outward grasp with a fixed-pinch grasp configuration were possible.

Rao *et al.* [103] distinguish between graspable and nongraspable object hypotheses in a scene. Using a combination of 2-D and 3-D features, an SVM is trained on labeled data of segmented objects. Among those features are for example the

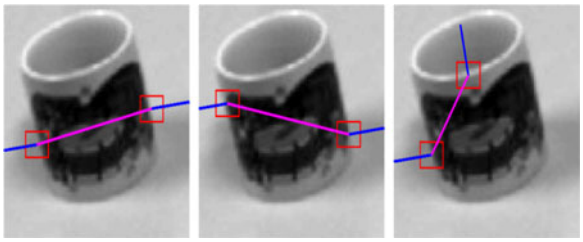


Fig. 13. Three grasp candidates for a cup represented by two local patches and their major gradient, as well as their connecting line [31].

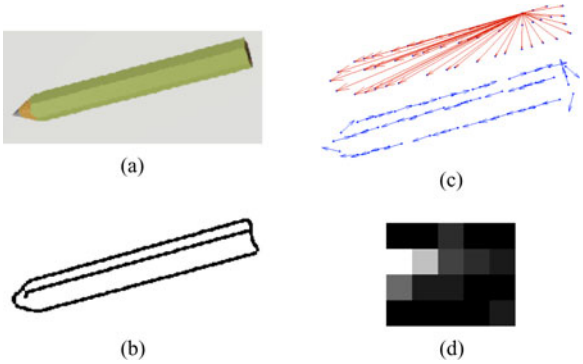


Fig. 14. Example shape contexts descriptor for the image of a pencil. (a) Input image. (b) Canny edges. (c) (Top) All vectors from one point to all other sample points. (Bottom) Sampled points of the contour with gradients. (d) Histogram with four angle and five log-radius bins comprising the vectors in (c) [32].

variance in depth and height, as well as variance of the three channels in the Lab color space. These are some kind of *meta* features that are used instead of the values of, e.g., the color channels directly. Rao *et al.* [103] achieve good classification rates on object hypotheses formed by segmentation on color and depth cues. Le *et al*. [31] model grasp hypotheses as consisting of two contact points. They apply a learning approach to rank a sampled set of fingertip positions according to graspability. The feature vector consists of a combination of 2-D and 3-D cues such as gradient angle or depth variation along the line connecting the two grasping points. Example grasp candidates are shown in Fig. 13.

Bohg and Kragic [32] propose an approach that instead of using local features, encodes global 2-D object shape. It is represented relative to a potential grasping point by shape contexts as introduced by Belongie *et al.* [81]. Fig. 14 shows a potential grasping point and the associated feature.

Bergström *et al.* [97] see the result of the 2-D based grasp selection as a way to search in a 3-D object representation for a full grasp configuration. The authors extend their previous approach [32] to work on a sparse edge-based object representation. They show that integrating 3-D and 2-D-based methods for grasp hypotheses generation results in a sparser set of grasps with a good quality.

Different from the previous approaches, Ramisa *et al.* [93] consider the problem of manipulating deformable objects, specifically folding shirts. They aim at detecting the shirt collars that exhibit deformability but that have distinct features as well. The authors show that a combination of local 2-D and 3-D descriptors works well for this task. Results are presented in terms of how reliable collars can be detected when only a single shirt or several shirts are present in the scene.

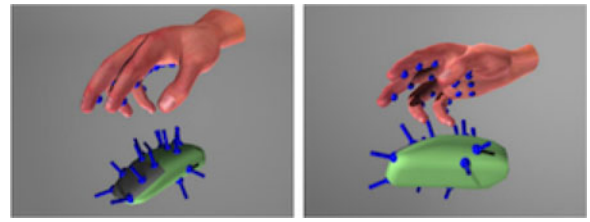


Fig. 15. Matching contact points between human hand and object [88].

B. Grasp Synthesis by Comparison

The aforementioned approaches study what kind of features encode similarity of objects in terms of graspability and learn a discriminative function in the associated space. The methods we review next take an exemplar-based approach in which grasp hypotheses for a specific object are synthesized by finding the most similar object or object part in a database to which good grasps are already associated.

1) Synthetic Exemplars: Li and Pollard [88] treated the problem of finding a suitable grasp as a shape matching problem between the human hand and the object. The approach starts off with a database of human grasp examples. From this database, a suitable grasp is retrieved when queried with a new object. Shape features of this object are matched against the shape of the inside of the available hand postures. An example is shown in Fig. 15.

Curtis and Xiao [100] built upon a knowledge base of 3-D object types. These are represented by Gaussian distributions over very basic shape features, e.g., the aspect ratio of the object's bounding box, but also over physical features, e.g., material and weight. Furthermore, they are annotated with a set of representative pregrasps. To infer a good grasp for a new object, its features are used to look up the most similar object type in the knowledge base. If a successful grasp has been synthesized in this way and it is similar enough to the object type, then the mean and standard deviation of the object features are updated; otherwise, a new object type is formed in the knowledge base.

While these two aforementioned approaches use low-level shape features to encode similarity between objects, Dang and Allen [106] present an approach toward *semantic* grasp planning. In this case, *semantic* refers to both, the object category and the task of a grasp, e.g., pour water, answer a call, or hold and drill. A *semantic affordance map* links object features to an approach vector and to semantic grasp features (task label, joint angles, and tactile sensor readings). For planning a task-specific grasp on a novel object of the same category, the object features are used to retrieve the optimal approach direction and associated grasp features. The approach vector serves as a seed for synthesizing a grasp with the Eigengrasp planner [23]. The grasp features are used as a reference to which the synthesized grasp should be similar.

Hillenbrand and Roa [98] frame the problem of transferring functional grasps between objects of the same category as pose alignment and shape warping. They assume that there is a source object given on which a set of functional grasps is defined. Pose clustering is used to align another object of the same category with it. Subsequently, fingertip contact points can be transferred from the source to the target object. The experimental results are promising. However, they are limited to the category of cups containing six instances.

All four approaches [88], [98], [100], [106] compute object features that rely on the availability of 3-D object meshes. The question remains how these ideas could be transferred to the case where only partial sensory data are available to compute object features and similarity to already known objects. One idea would be to estimate full object shape from partial or multiple observations, as proposed by the approaches in Section IV-A and use the resulting potentially noisy and uncertain meshes to transfer grasps. The previous methods are also suitable to create experience databases offline that require only little labeling. In the case of category-based grasp transfer [98], [106], only one object per category would need to be associated with grasp hypotheses and all the other objects would only need a category label. No expensive grasp simulations for many grasp candidates would need to be executed as for the approaches in Section II-A1. Dang and Allen [106] followed this idea and demonstrated a few grasp trials on a real robot assuming that a 3-D model of the query object is in the experience database.

In addition, Goldfeder and Allen [101] built their knowledge base only from synthetic data on which grasps are generated using the previously discussed Eigengrasp planner [23]. Different from the previous approaches, observations made with real sensors from new objects are used to look up the most similar object and its pose in the knowledge base. Once this is found, the associated grasp hypotheses can be executed on the real object. Although experiments on a real platform are provided, it is not entirely clear how many trials have been performed on each object and how much object pose was varied. As discussed earlier, the study conducted by Balasubramanian *et al*. [39] suggested that the employed grasp planner is not the optimal choice for synthesizing grasps that also work well in the real world.

Detry *et al.* [91] aim at generalizing grasps to novel objects by identifying parts to which a grasp has already been successfully applied. This lookup is rendered efficient by creating a lower dimensional space in which object parts that are similarly shaped relative to the hand reference frame are close to each other. This space is shown in Fig. 16. The authors show that similar grasp to object part configurations can be clustered in this space and form prototypical grasp-inducing parts. An extension of this approach is presented by Detry *et al.* [92], where the authors demonstrate how this approach can be used to synthesize grasps on novel objects by matching these prototypical parts to real sensor data.

2) Sensor-Based Exemplars: The aforementioned approaches present promising ideas toward generalizing prior grasp experience to new objects. However, they are using 3-D object models to construct the experience database. In this section, we review methods that generate a knowledge base by linking object representations from real sensor data to grasps that were executed on a robotic platform. Fig. 18 visualizes the flow of data that these approaches follow.

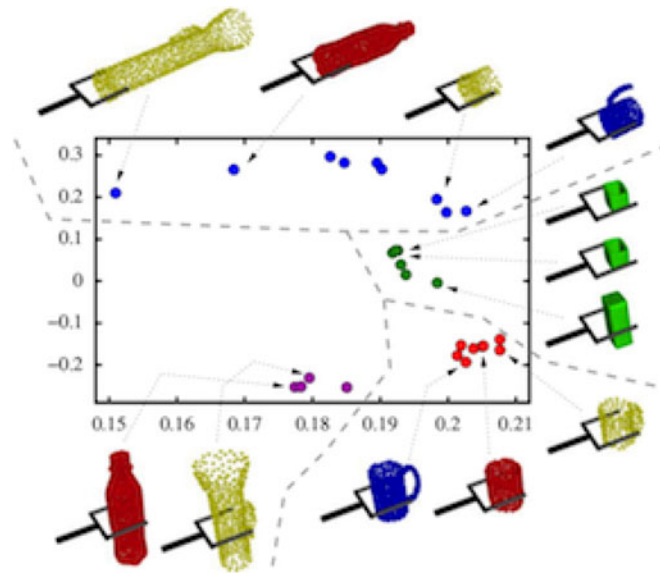


Fig. 16. Lower dimensional space in which similar pairs of grasps and object parts are close to each other [91].

Kamon *et al.* [5] propose one of the first approaches toward generalizing grasp experience to novel objects. Their aim is to learn a function $f : Q \rightarrow G$ that maps object- and grasp-candidate-dependent quality parameters Q to a grade G of the grasp. An object is represented by its 2-D silhouette, its center of mass, and main axis. The grasp is represented by two parameters f_1 and f_2 from which in combination with the object features, the fingertip positions can be computed. Learning is bootstrapped by the offline generation of a knowledge database containing grasp parameters along with their grade. This knowledge database is then updated, while the robot gathers experience by grasping new objects. The system is restricted to planar grasps and visual processing of top-down views on objects. It is therefore questionable how robust this approach is to more cluttered environments and strong pose variations of the object.

Morales *et al.* [25] use visual feedback to infer successful grasp configurations for a three-fingered hand. The authors take the hand kinematics into account when selecting a number of planar grasp hypotheses directly from 2-D object contours. To predict which of these grasps is the most stable one, a *k-nearest neighbor* approach is applied in connection with a grasp experience database. The experience database is built during a trial-and-error phase executed in the real world. Grasp hypotheses are ranked dependent on their outcome. Fig. 17 shows a successful and unsuccessful grasp configuration for one object. The approach is restricted to planar objects. Speth *et al.* [104] showed that their earlier 2-D-based approach [25] is also applicable when considering 3-D objects. The camera is used to explore the object and retrieve crucial information like height, 3-D position, and pose. However, all this additional information is not applied in the inference and final selection of a suitable grasp configuration.

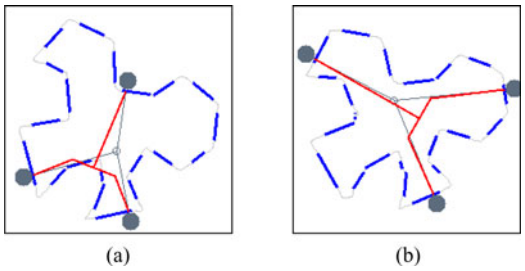


Fig. 17. (a) Successful grasp configuration for this object. (b) Unsuccessful grasp configuration for the same object [25].

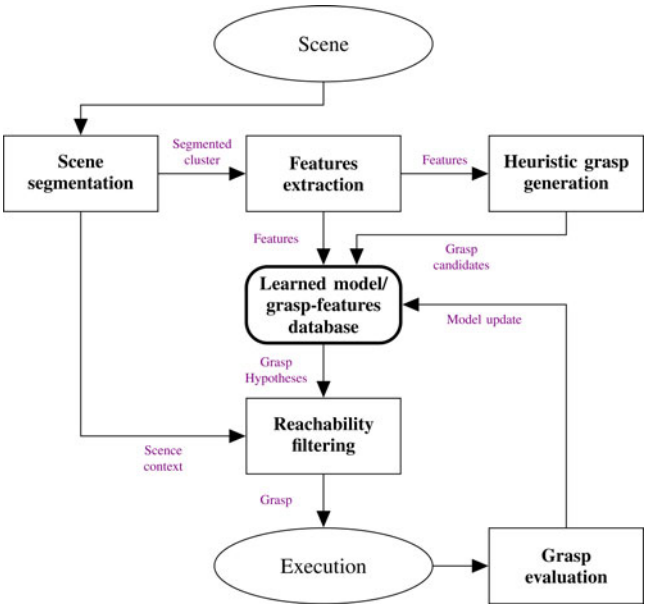


Fig. 18. Typical functional flowchart of a system that learns from trial and error. No prior knowledge about objects is assumed. The scene is segmented to obtain object clusters and relevant features are extracted. A heuristic module produces grasp candidates from these features. These candidates are ranked using a previously learned model or based on comparison to previous examples. The resulting grasp hypotheses are filtered, and one of them is finally executed. The performance of the execution is evaluated, and the model or memory is updated with this new experience. The following approaches can be summarized by this flowchart: [5], [25], [26], [71], [90], [95], and [104].

The approaches presented by Herzog *et al.* [71] and Kroemer *et al.* [90] also maintain a database of grasp examples. They combine learning by trial and error on real world data with a part-based representation of the object. There is no restriction of object shape. Each of them bootstrap the learning by providing the robot with a set of positive example grasps. However, their part representation and matching are very different. Herzog *et al.* [71] store a set of local templates of the parts of the object that have been in contact with the object during the human demonstration. Given a segmented object point cloud, its 3-D convex hull is constructed. A template is a height map that is aligned with one polygon of this hull. Together with a grasp hypotheses, they serve as positive examples. If a local part of an object is similar to a template in the database, then the associated grasp hypothesis is executed. Fig. 19 shows example query templates and the matched template from the database. In the case of failure, the object part is added as a negative example to the old template. In this way, the similarity metric can weight in similarity to positive examples as well as dissimilarity to negative examples. The proposed approach is evaluated on a large set of different objects and with different robots.

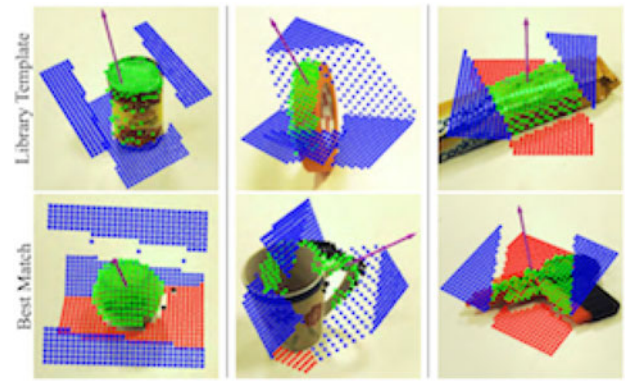


Fig. 19. Example query and matching templates [71].

Kroemer *et al.* [90] use a pouring task to demonstrate the generalization capabilities of the proposed approach to similar objects. An object part is represented as a set of points weighted according to an isotropic 3-D Gaussian with a given standard deviation. Its mean is manually set to define a part that is relevant to the specific action. When shown a new object, the goal of the approach is to find the subpart that is most likely to afford the demonstrated action. This probability is computed by kernel logistic regression whose result depends on the weighted similarity between the considered subpart and the example subparts in the database. The weight vector is learned given the current set of examples. This set can be extended with new parts after action execution. Neither Herzog *et al.* [71] nor Kroemer *et al.* [90] adaptqed the similarity metric itself under which a new object part is compared with previously encountered examples. Instead, the probability of success is estimated, all the examples from the continuously growing knowledge base taken into account.

C. Generative Models for Grasp Synthesis

Very little work has been done on learning generative models of the whole grasp process. These kind of approaches identify common structures from a number of examples instead of finding a decision boundary in some feature space or directly comparing with previous examples under some similarity metric. Montesano *et al.* [26] provide one example in which affordances are encoded in terms of an action that is executed on an object and produces a specific effect. The problem of learning a joint distribution over a set of variables is posed as structure learning in a Bayesian network framework. Nodes in this network are formed by object, action, and effect features that the robot can observe during execution. Given 300 trials, the robot learns the structure of the Bayesian network. Its validity is demonstrated in an imitation game, where the robot observes a human executing one of the known actions on an object and is asked to reproduce the same observed effect when given a new object. Effectively, the robot has to perform inference in

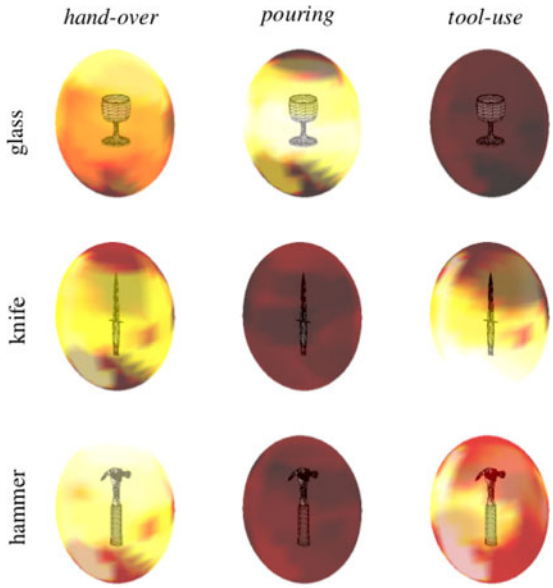


Fig. 20. Ranking of approach vectors for different objects given a specific task. The brighter an area, the higher the rank. The darker an area, the lower the rank [87].

the learned network to determine the action with the highest probability to succeed.

Song *et al.* [87] approach the problem of inferring a full grasp configuration for an object given a specific task. As in [26], the joint distribution over the set of variables influencing this choice is modeled as a Bayesian network. Additional variables like task, object category, and task constraints are introduced. The structure of this model is learned given a large number of grasp examples generated in Graspit! and annotated with grasp quality metrics, as well as suitability for a specific task. The authors exploit nonlinear dimensionality reduction techniques to find a discrete representation of continuous variables for efficient and more accurate structure learning. The effectiveness of the method is demonstrated on the synthetic data for different inference tasks. The learned quality of grasps on specific objects given a task is visualized in Fig. 20.

D. Category-Based Grasp Synthesis

Most of the previously discussed approaches link low-level information of the object to a grasp. Given that a novel object is similar in shape or appearance to a previously encountered one, then it is assumed that they can also be grasped in a similar way. However, objects might be similar on a different level. Objects in a household environment that share the same functional category might have a vastly different shape or appearance. However, they still can be grasped in the same way. In Section III-B1, we have already mentioned the work in [98] and [106] in which task-specific grasps are synthesized for objects of the same category. The authors assume that the category is known *a priori*. In the following, we review methods that generalize grasps to familiar objects by first determining their category.

Marton *et al.* [108] use different 3-D sensors and a thermo camera for performing object categorization. Features of the segmented point cloud and the segmented image region are extracted to train a Bayesian logic network for classifying object hypotheses as either boxes, plates, bottles, glasses, mugs, or silverware. A modified approach is presented in [102]. A layered 3-D object descriptor is used for categorization and an approach based on the *scale-invariant feature transform* [109] is applied for view-based object recognition. To increase the robustness of the categorization, the examination methods are run iteratively on the object hypotheses. A list of potential matching objects are kept and reused for verification in the next iteration. Objects for which no matching model can be found in the database are labeled as novel. Given that an object has been recognized, associated grasp hypotheses can be reused. These have been generated using the technique presented in [110].

Song *et al.* [87] treat object category as one variable in the Bayesian network. Madry *et al.* [105] demonstrate how the category of an object can be robustly detected given multimodal visual descriptors of an object hypothesis. This information is fed into the Bayesian network together with the desired task. A full hand configuration can then be inferred that obeys the task constraints. Bohg *et al.* [99] demonstrated this approach on the humanoid robot ARMAR III [74]. For robust object categorization, the approach by Madry *et al.* [105] is integrated with the 3-D based categorization system by Wohlkinger and Vincze [111]. The pose of the categorized object is estimated with the approach presented by Aldoma and Vincze [112]. Given this, the inferred grasp configuration can be checked for reachability and executed by the robot.

Recently, we have seen an increasing amount of new approaches toward pure 3-D descriptors of objects for categorization. Although, the following methods look promising, it has not been shown yet that they provide a suitable base for generalizing grasps over an object category. Rusu *et al.* [113], [114] provide extensions of [35] for either recognizing or categorizing objects and estimating their pose relative to the viewpoint. While in [114] quantitative results on real data are presented, [113] uses simulated object point clouds only. Lai *et al.* [36] perform object category and instance recognition. The authors learn an instance distance using the database presented in [46]. A combination of 3-D and 2-D features is used. Gonzalez-Aguirre *et al.* [115] present a shape-based object categorization system. A point cloud of an object is reconstructed by fusing partial views. Different descriptors (capturing global and local object shape) in combination with standard machine learning techniques are studied. Their performance is evaluated on real data.

IV. Grasping Unknown Objects

If a robot has to grasp a previously unseen object, then we refer to it as *unknown*. Approaches toward grasping known objects are obviously not applicable since they rely on the assumption that an object model is available. The approaches in this group also do not assume to have access to other kinds of grasp experiences. Instead, they propose and analyze heuristics that directly link structure in the sensory data to candidate grasps.

There are various ways to deal with sparse, incomplete, and noisy data from real sensors such as stereo cameras: we divided

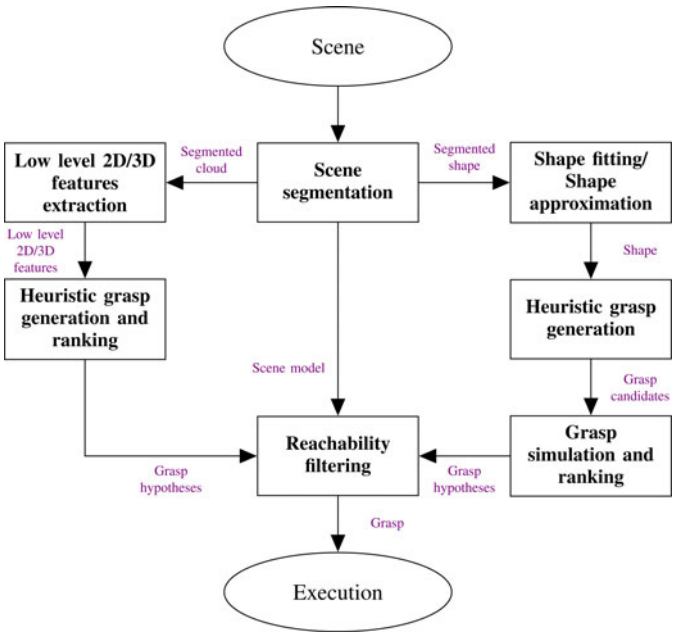


Fig. 21. Typical functional flowchart of a grasping system for unknown objects. The scene is perceived and segmented to obtain object hypotheses and relevant perceptual features. Then, the system follows either the right or left pathway. On the left, low-level features are used to generate heuristically a set of grasp hypotheses. On the right, a mesh model approximating the global object shape is generated from the perceived features. Grasp candidates are then sampled and executed in a simulator. Classical analytic grasp metric is used to rank the grasp candidates. Finally, nonreachable grasp hypotheses are filtered out, and the best ranked grasp hypothesis is executed. The following approaches use the left pathway: [33], [34], [37], [60], [116], [117], [119], [121], [122], and [126]. The following approaches estimate a full object model: [110], [118], [120], and [123]–[125].

the approaches into methods that 1) approximate the full shape of an object, 2) methods that generate grasps based on low-level features and a set of heuristics, and 3) methods that rely mostly on the global shape of the partially observed object hypothesis. The reviewed approaches are summarized in Table III. A flowchart that visualizes the data flow in the following approaches is shown in Fig. 21.

A. Approximating Unknown Object Shape

One approach toward generating grasp hypotheses for unknown objects is to approximate objects with shape primitives. Dunes *et al.* [124] approximate an object with a quadric whose minor axis is used to infer the wrist orientation. The object centroid serves as the approach target and the rough object size helps to determine the hand preshape. The quadric is estimated from multiview measurements of the global object shape in monocular images. Marton *et al.* [110] show how grasp selection can be performed exploiting symmetry by fitting a curve to a cross section of the point cloud of an object. For grasp planning, the reconstructed object is imported to a simulator. Grasp candidates are generated through randomization of grasp parameters on which the force-closure criteria is then evaluated. Rao *et al.* [103] sample grasp points from the surface of a segmented object. The normal of the local surface at this point

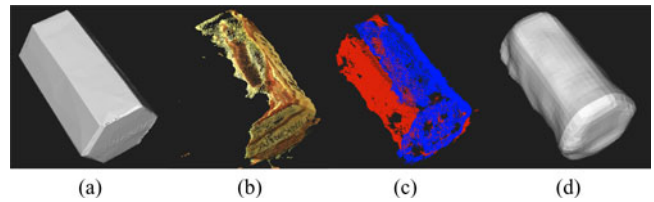


Fig. 22. Estimated full object shape by assuming symmetry. (a) Ground truth mesh. (b) Original point cloud. (c) Mirrored cloud with original points in blue and additional points in red. (d) Reconstructed mesh [120].

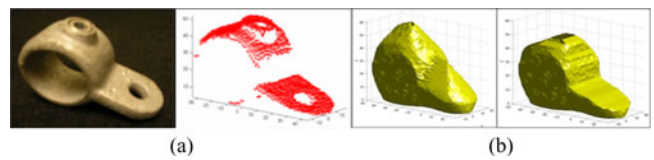


Fig. 23. Unknown object shape estimated by shape carving. (a) (Left) Object image. (Right) Point cloud. (b) (Left) Model from silhouettes. (Right) Model merged with point cloud data [118].

serves as a search direction for a second contact point. This is chosen to be at the intersection between the extended normal and the opposite side of the object. By assuming symmetry, this second contact point is assumed to have a contact normal in the direction opposite from the normal of the first contact point. Bohg *et al.* [120] propose a related approach that reconstructs full object shape assuming planar symmetry which subsumes all other kinds of symmetries. It takes the complete point cloud into account and not only a local patch. Two simple methods to generate grasp candidate on the resulting completed object models are proposed and evaluated. An example for an object whose full object shape is approximated with this approach is shown in Fig. 22.

As opposed to the aforementioned techniques, Bone *et al.* [118] made no prior assumption about the shape of the object. They applied shape carving for the purpose of grasping with a parallel-jaw gripper. After obtaining a model of the object, they search for a pair of reasonably flat and parallel surfaces that are best suited for this kind of manipulator. An object reconstructed with this method is shown in Fig. 23.

Lippiello *et al.* [123] present a related approach for grasping an unknown object with a multifingered hand. The authors first record a number of views from around the object. Based on the object bounding box in each view, a polyhedron is defined that overestimates the visual object hull and is then approximated by a quadric. A pregrasp shape is defined in which the fingertip contacts on the quadric are aligned with its two minor axes. This grasp is then refined, given the local surface shape close to the contact point. This process is alternating with the refinement of the object shape through an elastic surface model. The quality of the grasps is evaluated by classic metrics. As previously discussed, it is not clear how well these metrics predict the outcome of a grasp.

B. From Low-Level Features to Grasp Hypotheses

A common approach is to map low-level 2-D or 3-D visual features to a predefined set of grasp postures and then rank them

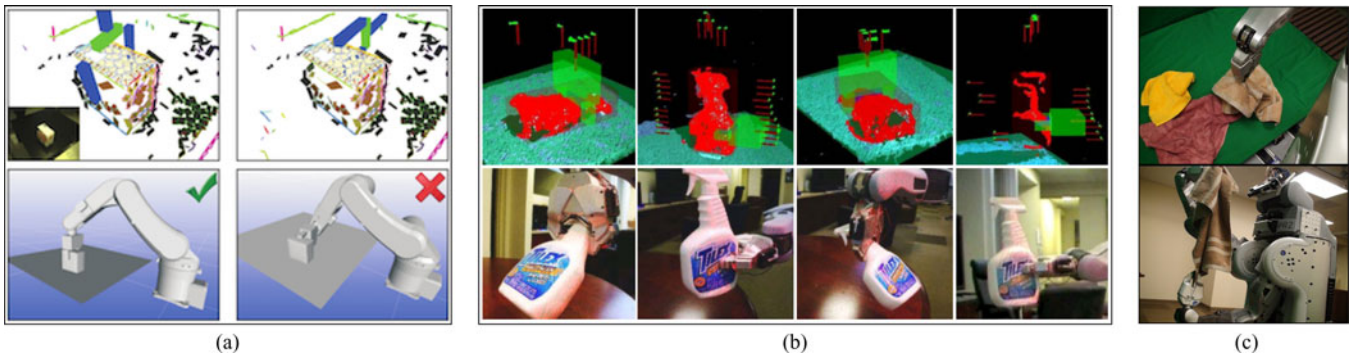


Fig. 24. Generating and ranking grasp hypotheses from local object features. (a) Generation of grasp candidates from local surface features and evaluation in simulation [117]. (b) Generated grasp hypotheses on point cloud clusters and execution results [33]. (c) (Top) Grasping a towel from the table. (Bottom) Regrasping a towel for unfolding [37].

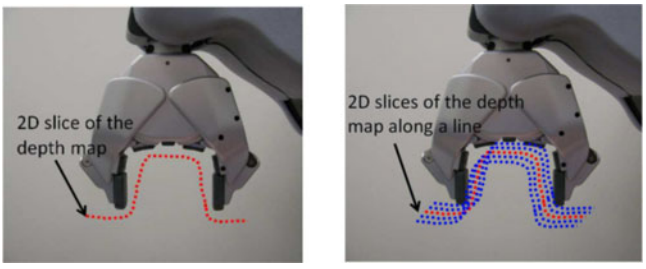


Fig. 25. PR2 gripper and associated grasp pattern [34].

dependent on a set criteria. Kraft *et al.* [116] use a stereo camera to extract a representation of the scene. Instead of a raw point cloud, they process it further to obtain a sparser model consisting of local multimodal contour descriptors. Four elementary grasping actions are associated with specific constellations of these features. With the help of heuristics, the large number of resulting grasp hypotheses is reduced. Popović *et al.* [117] present an extension of this system that uses local surfaces and their interrelations to propose and filter two- and three-fingered grasp hypotheses. The feasibility of the approach is evaluated in a mixed real-world and simulated environment. The object representation and the evaluation in simulation is visualized in Fig. 24(a).

Hsiao *et al.* [33] employ several heuristics for generating grasp hypotheses dependent on the shape of the segmented point cloud. These can be grasps from the top, from the side, or applied to high points of the objects. The generated hypotheses are then ranked using a weighted list of features such as for example number of points within the gripper or distance between the fingertip and the center of the segment. Some examples for grasp hypotheses generated in this way are shown in Fig. 24(b).

The main idea presented by Klingbeil *et al.* [34] is to search for a pattern in the scene that is similar to the 2-D cross section of the robotic gripper interior. This is visualized in Fig. 25. The idea is similar to the work by Li and Pollard [88], as shown in Fig. 15. However, in this study, the authors do not rely on the availability of a complete 3-D object model. A depth image serves as the input to the method and is sampled to find a set of grasp hypotheses. These are ranked according to an objective function that takes pairs of these grasp hypotheses and their local structure into account.

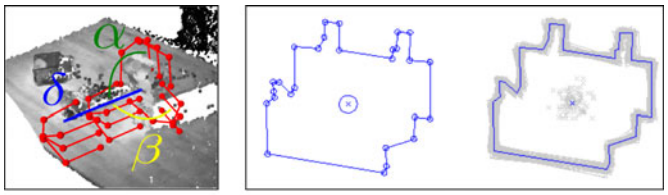
Maitin-Shepard *et al.* [37] propose a method for grasping and folding towels that can vary in size and are arranged in unpredictable configurations. Different from the approaches discussed previously, the objects are deformable. The authors propose a border detection method that relies on depth discontinuities and then fit corners to border points. These then serve as grasping points. Examples for grasping a towel are shown in Fig. 24(c). Although this approach is applicable to a family of deformable objects, it does not detect grasping points by comparing to previously encountered grasping points. Instead, it directly links local structure to a grasp. For this reason, we consider it to be an approach toward grasping unknown objects.

C. From Global Shape to Grasp Hypothesis

Other approaches use the global shape of an object to infer one good grasp hypothesis. Morales *et al.* [126] extracted the 2-D silhouette of an unknown object from an image and computed two- and three-fingered grasps taking into account the kinematics constraints of the hand. Richtsfeld and Vincze [119] use a segmented point cloud from a stereo camera. They search for a suitable grasp with a simple gripper that is based on the shift of the top plane of an object into its center of mass. A set of heuristics is used for selecting promising fingertip positions. Maldonado *et al.* [122] model the object as a 3-D Gaussian. For choosing a grasp configuration, it optimizes a criterion in which the distance between palm and object is minimized, while the distance between fingertips and the object is maximized. The simplified model of the hand and optimization variables are shown in Fig. 26(a).

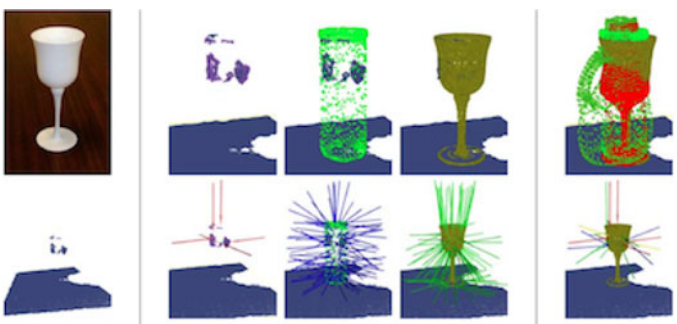
Stückler *et al.* [121] generate grasp hypotheses based on eigenvectors of the object's *footprints* on the table. Footprints refer to the 3-D object point cloud projected onto the supporting surface.

Kehoe *et al.* [125] assume an overhead view of the object and approximate its shape with an extruded polygon. The goal is to synthesize a zero-slip push grasp with a parallel jaw gripper, given uncertainty about the precise object shape and the position of its center of mass. For this purpose, perturbations of the initial shape and position of the centroid are sampled. For an example of this, see Fig. 26(b). For each of these samples, the same grasp candidate is evaluated. Its quality depends on how often



(a) (b)

Fig. 26. Mapping global object shape to grasps. (a) Simplified hand model and grasp parameters to be optimized [122]. (b) Planar object shape uncertainty model (Left) Vertices and center of mass with Gaussian position uncertainty ($\sigma = 1$). (Right) 100 samples of perturbed object models [125].



(a) (b) (c) (d) (e)

Fig. 27. (a) Object and point cloud. (b)–(d) Object representation and grasp hypotheses. (e) Overlaid representations and list of consistent grasp hypotheses [60], [127].

it resulted in force closure under the assumed model of object shape uncertainty.

V. Hybrid Approaches

There are a few data-driven grasp synthesis methods that cannot clearly be classified as using only one kind of prior knowledge. One of these approaches has been proposed in Brook *et al.* [60] with an extension in [127]. Different grasp planners provide grasp hypotheses which are integrated to reach a consensus on how to grasp a segmented point cloud. The authors show results using the planner presented in [33] for unknown objects in combination with grasp hypotheses generated through fitting known objects to point cloud clusters, as described in [75]. Fig. 27 shows the grasp hypotheses for a segmented point cloud based on the input from these different planners. Another example for a hybrid approach is the work by Marton *et al.* [102]. A set of very simple shape primitives like boxes, cylinders, and more general rotational objects are considered. They are reconstructed from segmented point clouds by analysis of their footprints. Parameters such as circle radius and the side lengths of rectangles are varied; curve parameters are estimated to reconstruct more complex rotationally symmetric objects. Given these reconstructions, a lookup is made in a database of already encountered objects for reusing successful grasp hypotheses. In case no similar object is found, new grasp hypotheses are generated using the technique presented in [110]. For object hypotheses that cannot be represented by the simple shape primitives mentioned previously, a surface is reconstructed through triangulation. Grasp hypotheses are generated using the planner presented in [33].

VI. Discussion and Conclusion

We have identified four major areas that form open problems in the area of robotic grasping.

Object segmentation: Many of the approaches that are mentioned in this survey usually assume that the object to be grasped is already segmented from the background. Since segmentation is a very hard problem in itself, many methods make the simplifying assumption that objects are standing on a planar surface. Detecting this surface in a 3-D point cloud and performing Euclidean clustering results in a set of segmented point clouds that serve as object hypotheses [114]. Although the dominant surface assumption is viable in certain scenarios and to shortcut the problem of segmentation, we believe that we need a more general approach to solve this.

First of all, some objects might usually occur in a specific spatial context. This can be on a planar surface, but it might also be on a shelf or in the fridge. Aydemir and Jensfelt [128] propose to learn this context for each known object to guide the search for them. One could also imagine that this context could help segmenting foreground from background. Furthermore, there are model-based object detection methods [55], [62], [72], [73], [78] that can segment a scene as a by-product of detection and without making strong assumptions about the environment. In the case of unknown objects, some methods have been proposed that employ the interaction capabilities of a robot, e.g., visual fixation or pushing movements with the robot hand, to segment the scene [116], [129]–[132]. A general solution toward object segmentation might be a combination of these two methods. The robot first interacts with objects to acquire a model. Once it has an object model, it can be used for detecting and thereby segmenting it from the background.

Learning to grasp: Let us consider the goal of having a robotic companion helping us in our household. In this scenario, we cannot expect that the programmer has foreseen all the different situations with which this robot will be confronted. Therefore, the ideal household robot should have the ability to continuously learn about new objects and how to manipulate them while it is operating in the environment. We will also not be able to rely on having 3-D models readily available of all objects the robot could possibly encounter. This requires the ability to learn a model that could generalize from previous experience to new situations. Many open questions arise: How is the experience regarding one object and grasp represented in memory? How can success and failure be autonomously quantified? How can a model be learned from this experience that would generalize to new situations? Should it be a discriminative, a generative, or exemplar-based model? What are the features that encode object affordances? Can these be autonomously learned? In which space are we comparing new objects to already encountered ones? Can we bootstrap learning by using simulation or by human demonstration? The methods that we have discussed in Section III about grasping familiar objects approach these

questions. However, we are still far from a method that answers all of them in a satisfying way.

Autonomous manipulation planning: Recently, more complex scenarios than just grasping from a table top have been approached by a number of research labs. How a robot can autonomously sequence a set of actions to perform such a task is still an open problem. Toward this end, Tenorth *et al.* [133] propose a cloud robotics infrastructure under which robots can share their experience such as action recipes and manipulation strategies. An inference engine is provided for checking whether all requirements are fulfilled for performing a full manipulation strategy. It would be interesting to study how the uncertainty in perception and execution can be dealt with in conjunction with such a symbolic reasoning engine.

When considering a complex action, grasp synthesis cannot be considered as an isolated problem. On the contrary, higher level tasks influence what the best grasp in a specific scenario might be, e.g., when grasping a specific tool. Task constraints have not yet been considered extensively in the community. Current approaches, e.g., [87] and [106], achieve impressive results. An open question is how to scale to life-long learning.

Robust execution: It has been noted by many researchers that inferring a grasp for a given object is necessary but not sufficient. Only if execution is robust to uncertainties in sensing and actuation, a grasp can succeed with high probability. There are a number of approaches that use constant tactile or visual feedback during grasp execution to adapt to unforeseen situations [33], [47], [49], [134]–[137]. Tactile feedback can be from haptic or force–torque sensors. Visual feedback can be the result from tracking the hand and object simultaneously. In addition, in this area, there are a number of open questions. How can tactile feedback be interpreted to choose an appropriate corrective action independent of the object, the task, and environment? How can visual and tactile information be fused in the controller?

A. Final Notes

In this paper, we reviewed work on data-driven grasp synthesis and proposed a categorization of the published work. We focused on the type and level of prior knowledge used in the proposed approaches and on the assumptions that are commonly made about the objects being manipulated. We identified recent trends in the field and provided a discussion about the remaining challenges.

An important issue is the current lack of general benchmarks and performance metrics suitable for comparing the different approaches. Although various object-grasp databases are already available, e.g., the Columbia grasp database [138], the VisGraB dataset [139], or the playpen dataset [140], they are not commonly used for comparison. We acknowledge that one of the reasons is that grasping in itself is highly dependent on the employed sensing and manipulation hardware. There have also been robotic challenges organized such as the DARPA Arm project [141] or RoboCup@Home [142], and a framework for benchmarking has been proposed in [143]. However, none of these successfully integrate all the subproblems relevant for benchmarking different grasping approaches.

Given that data-driven grasp synthesis is an active field of research and lots of work has been reported in the area, we have set up a web page that contains all the references in this survey at www.robotic-grasping.com. They are structured according to the proposed classification and tagged with the mentioned aspect. The web page will be constantly updated with the most recent approaches.

References

[1] A. Sahbani, S. El-Khoury, and P. Bidaud, "An overview of 3-D object grasp synthesis algorithms," *Robot. Auton. Syst.*, vol. 60, no. 3, pp. 326–336, Mar. 2012.

[2] K. Shimoga, "Robot grasp synthesis algorithms: A survey," *Int. J. Robot. Res.*, vol. 15, no. 3, pp. 230–266, 1996.

[3] R. N. Murray, Z. Li, and S. Sastry, *A Mathematical Introduction to Robotics Manipulation*. Boca Raton, FL, USA: CRC, 1994.

[4] A. Bicchi and V. Kumar, "Robotic grasping and contact," in *Proc. IEEE Int. Conf. Robot. Autom.*, San Francisco, CA, USA, Apr. 2000, pp. 348–353.

[5] I. Kamon, T. Flash, and S. Edelman, "Learning to grasp using visual information," in *Proc. IEEE Int. Conf. Robot. Autom.*, 1994, pp. 2470–2476.

[6] S. Ekvall and D. Kragic, "Learning and evaluation of the approach vector for automatic grasp generation and planning," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2007, pp. 4715–4720.

[7] A. Morales, T. Asfour, P. Azad, S. Knoop, and R. Dillmann, "Integrated grasp planning and visual object localization for a humanoid robot with five-fingered hands," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Beijing, China, Oct. 2006, pp. 5663–5668.

[8] D. Prattichizzo and J. Trinkle, *Handbook of Robotics*. Heidelberg, Germany: Springer-Verlag, 2008, ch. 28, pp. 671–700.

[9] D. Prattichizzo, M. Malvezzi, M. Gabiccini, and A. Bicchi, "On the manipulability ellipsoids of underactuated robotic hands with compliance," *Robot. Auton. Syst.*, vol. 60, no. 3, pp. 337–346, 2012.

[10] C. Rosales, R. Suarez, M. Gabiccini, and A. Bicchi, "On the synthesis of feasible and prehensile robotic grasps," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2012, pp. 550–556.

[11] V.-D. Nguyen, "Constructing force-closure grasps," *Int. J. Robot. Res.*, vol. 7, no. 3, pp. 3–16, 1988.

[12] M. A. Roa and R. Suárez, "Computation of independent contact regions for grasping 3-D objects," *IEEE Trans. Robot.*, vol. 25, no. 4, pp. 839–850, Aug. 2009.

[13] R. Krug, D. N. Dimitrov, K. A. Charusta, and B. Iliev, "On the efficient computation of independent contact regions for force closure grasps," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2010, pp. 586–591.

[14] A. Rodriguez, M. T. Mason, and S. Ferry, "From caging to grasping," presented at the Robot.: Sci. Syst. Conf., Los Angeles, CA, USA, Apr. 2011.

[15] J. Seo, S. Kim, and V. Kumar, "Planar, bimanual, whole-arm grasping," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2012, pp. 3271–3277.

[16] L. E. Zhang and J. C. Trinkle, "The application of particle filtering to grasping acquisition with visual occlusion and tactile sensing," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2012, pp. 3805–3812.

[17] C. Ferrari and J. Canny, "Planning optimal grasps," in *Proc. IEEE Int. Conf. Robot. Autom.*, May 1992, vol. 3, pp. 2290–2295.

[18] A. T. Miller and P. K. Allen, "Graspit!—A versatile simulator for robotic grasping," *IEEE Robot. Autom. Mag.*, vol. 11, no. 4, pp. 110–122, Dec. 2004.

[19] A. T. Miller, S. Knoop, H. I. Christensen, and P. K. Allen, "Automatic grasp planning using shape primitives," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2003, pp. 1824–1829.

[20] R. Pelossof, A. Miller, P. Allen, and T. Jebera, "An SVM learning approach to robotic grasping," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2004, pp. 3512–3518.

[21] C. Goldfeder, P. K. Allen, C. Lackner, and R. Pelossof, "Grasp planning via decomposition trees," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2007, pp. 4679–4684.

[22] C. Borst, M. Fischer, and G. Hirzinger, "Grasping the dice by dicing the grasp," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2003, pp. 3692–3697.

[23] M. Ciocarlie and P. Allen, "Hand posture subspaces for dexterous robotic grasping," *Int. J. Robot. Res.*, vol. 28, pp. 851–867, Jul. 2009.

[24] R. Diankov, "Automated construction of robotic manipulation programs," Ph.D. dissertation, Robotics Inst., Carnegie Mellon Univ., Pittsburgh, PA, USA, Aug. 2010.

[25] A. Morales, E. Chinellato, A. Fagg, and A. del Pobil, "Using experience for assessing grasp reliability," *Int. J. Human. Robot.*, vol. 1, no. 4, pp. 671–691, 2004.

[26] L. Montesano, M. Lopes, A. Bernardino, and J. Santos-Victor, "Learning object affordances: From sensory-motor coordination to imitation," *IEEE Trans. Robot.*, vol. 24, no. 1, pp. 15–26, Feb. 2008.

[27] R. Detry, E. Başeski, N. Krüger, M. Popović, Y. Touati, O. Kroemer, J. Peters, and J. Piater, "Learning object-specific grasp affordance densities," in *Proc. IEEE Int. Conf. Develop. Learn.*, 2009, pp. 1–7.

[28] A. Saxena, J. Driemeyer, and A. Y. Ng, "Robotic grasping of novel objects using vision," *Int. J. Robot. Res.*, vol. 27, no. 2, pp. 157–173, Feb. 2008.

[29] A. Saxena, L. Wong, and A. Y. Ng, "Learning grasp strategies with partial shape information," in *Proc. AAAI Conf. Artif. Intell.*, 2008, pp. 1491–1494.

[30] M. Stark, P. Lies, M. Zillich, J. Wyatt, and B. Schiele, "Functional object class detection based on learned affordance cues," in *Proc. Int. Conf. Comput. Vis. Syst.*, 2008, vol. 5008, pp. 435–444.

[31] Q. V. Le, D. Kamm, A. F. Kara, and A. Y. Ng, "Learning to grasp objects with multiple contact points," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2010, pp. 5062–5069.

[32] J. Bohg and D. Kragic, "Learning grasping points with shape context," *Robot. Auton. Syst.*, vol. 58, no. 4, pp. 362–377, 2010.

[33] K. Hsiao, S. Chitta, M. Ciocarlie, and E. G. Jones, "Contact-reactive grasping of objects with partial shape information," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Taipei, Taiwan, Oct. 2010, pp. 1228–1235.

[34] E. Klingbeil, D. Rao, B. Carpenter, V. Ganapathi, A. Y. Ng, and O. Khatib, "Grasping with application to an autonomous checkout robot," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2011, pp. 2837–2844.

[35] R. B. Rusu, N. Blodow, and M. Beetz, "Fast point feature histograms (FPFH) for 3-D registration," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2009, pp. 1848–1853.

[36] K. Lai, L. Bo, X. Ren, and D. Fox, "Sparse distance learning for object recognition combining RGB and depth information," in *Proc. IEEE Int. Conf. Robot. Autom.*, Shanghai, China, May 2011, pp. 4007–4013.

[37] J. Maitin-Shepard, M. Cusumano-Towner, J. Lei, and P. Abbeel, "Cloth grasp point detection based on multiple-view geometric cues with application to robotic towel folding," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2010, pp. 2308–2315.

[38] M. Beetz, U. Klank, I. Kresse, A. Maldonado, L. Mösenlechner, D. Pangercic, T. Rühr, and M. Tenorth, "Robotic roommates making pancakes," in *Proc. IEEE/RAS Int. Conf. Human. Robots (Human.)*, Bled, Slovenia, Oct. 2011, pp. 529–536.

[39] R. Balasubramanian, L. Xu, P. D. Brook, J. R. Smith, and Y. Matsuoka, "Physical human interactive guidance: Identifying grasping principles from human-planned grasps," *IEEE Trans. Robot.*, vol. 28, no. 4, pp. 899–910, Aug. 2012.

[40] J. Weisz and P. K. Allen, "Pose error robust grasping from contact wrench space metrics," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2012, pp. 557–562.

[41] K. Konolige, "Projected texture stereo," in *Proc. IEEE Int. Conf. Robot. Autom.*, May 2010, pp. 148–155.

[42] W. Garage. "PR2," [Online]. Available: www.willowgarage.com/pages/pr2/overview

[43] "*ROS (Robot Operating System)*." [Online]. Available: www.ros.org

[44] Microsoft. "Kinect-Xbox.com," [Online]. Available: www.xbox.com/en-US/KINECT

[45] *PrimeSense*. [Online]. Available: www.primesense.com

[46] K. Lai, L. Bo, X. Ren, and D. Fox, "A large-scale hierarchical multi-view RGB-D object dataset," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2011, pp. 1817–1824.

[47] J. Felip and A. Morales, "Robust sensor-based grasp primitive for a three-finger robot hand," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2009, pp. 1811–1816.

[48] K. Hsiao, P. Nangeroni, M. Huber, A. Saxena, and A. Y. Ng, "Reactive grasping using optical proximity sensors," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2009, pp. 2098–2105.

[49] P. Pastor, L. Righetti, M. Kalakrishnan, and S. Schaal, "Online movement adaptation based on previous sensor experiences," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, San Francisco, CA, USA, Sep. 2011, pp. 365–371.

[50] J. Romano, K. Hsiao, G. Niemeyer, S. Chitta, and K. Kuchenbecker, "Human-inspired robotic grasp control with tactile sensing," *IEEE Trans. Robot.*, vol. 27, no. 6, pp. 1067–1079, Dec. 2011.

[51] M. A. Goodale, "Separate visual pathways for perception and action," *Trends Neurosci.*, vol. 15, no. 1, pp. 20–25, 1992.

[52] U. Castiello, "The neuroscience of grasping," *Nat. Rev. Neurosci.*, vol. 6, no. 9, pp. 726–736, 2005.

[53] J. C. Culham, C. Cavina-Pratesi, and A. Singhal, "The role of parietal cortex in visuomotor control: What have we learned from neuroimaging?" *Neuropsychologia*, vol. 44, no. 13, pp. 2668–2684, 2006.

[54] E. Chinellato and A. P. Del Pobil, "The neuroscience of vision-based grasping: a functional review for computational modeling and bio-inspired robotics.," *J. Integr. Neurosci.*, vol. 8, no. 2, pp. 223–254, 2009.

[55] J. Glover, D. Rus, and N. Roy, "Probabilistic models of object geometry for grasp planning," presented at the 4th Robot.: Sci. Syst. Conf., Zurich, Switzerland, Jun. 2008.

[56] M. Przybylski, T. Asfour, and R. Dillmann, "Planning grasps for robotic hands using a novel object representation based on the medial axis transform," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Sep. 2011, pp. 1781–1788.

[57] M. A. Roa, M. J. Argus, D. Leidner, C. Borst, and G. Hirzinger, "Power grasp planning for anthropomorphic robot hands," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2012, pp. 663–569.

[58] R. Detry, D. Kraft, A. G. Buch, N. Krüger, and J. Piater, "Refining grasp affordance models by experience," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2010, pp. 2287–2293.

[59] K. Huebner, K. Welke, M. Przybylski, N. Vahrenkamp, T. Asfour, D. Kragic, and R. Dillmann, "Grasping known objects with humanoid robots: A box-based approach," in *Proc. Int. Conf. Adv. Robot.*, 2009, pp. 1–6.

[60] P. Brook, M. Ciocarlie, and K. Hsiao, "Collaborative grasp planning with multiple object representations," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2011, pp. 2851–2858.

[61] J. Romero, H. Kjellström, and D. Kragic, "Modeling and evaluation of human-to-robot mapping of grasps," in *Proc. Int. Conf. Adv. Robot.*, 2009, pp. 228–233.

[62] C. Papazov, S. Haddadin, S. Parusel, K. Krieger, and D. Burschka, "Rigid 3-D geometry matching for grasping of known objects in cluttered scenes," *Int. J. Robot. Res.*, vol. 31, no. 4, pp. 538–553, Apr. 2012.

[63] A. Collet Romea, D. Berenson, and S. Srinivasa, D. Ferguson, "Object recognition and full pose registration from a single image for robotic manipulation," in *Proc. IEEE Int. Conf. Robot. Autom.*, May 2009, pp. 48–55.

[64] O. B. Kroemer, R. Detry, J. Piater, and J. Peters, "Combining active learning and reactive control for robot grasping," *Robot. Auton. Syst.*, vol. 58, pp. 1105–1116, Sep. 2010.

[65] J. Tegin, S. Ekvall, D. Kragic, B. Iliev, and J. Wikander, "Demonstration based learning and control for automatic grasping," *J. Intell. Serv. Robot.*, vol. 2, no. 1, pp. 23–30, Aug. 2008.

[66] F. Stulp, E. Theodorou, M. Kalakrishnan, P. Pastor, L. Righetti, and S. Schaal, "Learning motion primitive goals for robust manipulation," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, San Francisco, CA, USA, Sep. 2011, pp. 325–331.

[67] K. Hübner and D. Kragic, "Selection of robot pre-grasps using box-based shape approximation," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2008, pp. 1765–1770.

[68] M. Przybylski and T. Asfour, "Unions of balls for shape approximation in robot grasping," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Taipei, Taiwan, Oct. 2010, pp. 1592–1599.

[69] R. Diankov and J. Kuffner, "Openrave: A planning architecture for autonomous robotics," Robotics Inst., Pittsburgh, PA, USA, Tech. Rep. CMU-RI-TR-08-34, Jul. 2008.

[70] M. Do, J. Romero, H. Kjellström, P. Azad, T. Asfour, D. Kragic, and R. Dillmann, "Grasp recognition and mapping on humanoid robots," presented at the *IEEE/RAS Int. Conf. Humanoid Robots (Humanoids)*, Paris, France, Dec. 2009.

[71] A. Herzog, P. Pastor, M. Kalakrishnan, L. Righetti, T. Asfour, and S. Schaal, "Template-based learning of grasp selection," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2012, pp. 2379–2384.

[72] R. Detry, N. Pugeault, and J. Piater, "A probabilistic framework for 3-D visual object representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 10, pp. 1790–1803, Oct. 2009.

[73] P. Azad, T. Asfour, and R. Dillmann, "Stereo-based 6-D object localization for grasping with humanoid robot systems," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2007, pp. 919–924.

[74] T. Asfour, P. Azad, N. Vahrenkamp, K. Regenstein, A. Bierbaum, K. Welke, J. Schröder, and R. Dillmann, "Toward humanoid manipulation in human-centred environments," *Robot. Auton. Syst.*, vol. 56, pp. 54–65, Jan. 2008.

[75] M. Ciocarlie, K. Hsiao, E. G. Jones, S. Chitta, R. B. Rusu, and I. A. Sucan, "Toward reliable grasping and manipulation in household environments," presented at the Int. Symp. Exp. Robot., New Delhi, India, Dec. 2010.

[76] P. J. Besl and N. D. McKay, "A method for registration of 3-D shapes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 2, pp. 239–256, Feb. 1992.

[77] C. Papazov and D. Burschka, "An efficient ransac for 3-D object recognition in noisy and occluded scenes," in *Proc. Asian Conf. Comput. Vis.*, 2010, pp. 135–148.

[78] A. Collet Romea, M. Martinez Torres, and S. Srinivasa, "The moped framework: Object recognition and pose estimation for manipulation," *Int. J. Robot. Res.*, vol. 30, no. 10, pp. 1284–1306, Sep. 2011.

[79] J. Shotton, J. Winn, C. Rother, and A. Criminisi, "Textonboost: Joint appearance, shape, and context modeling for multi-class object recognition and segmentation," in *Proc. Eur. Conf. Comput. Vis.*, 2006, pp. 1–15.

[80] V. Ferrari, L. Fevrier, F. Jurie, and C. Schmid, "Groups of adjacent contour segments for object detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 1, pp. 36–51, Jan 2008.

[81] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 4, pp. 509–522, Apr. 2002.

[82] C. Dance, J. Willamowski, L. Fan, C. Bray, and G. Csurka, "Visual categorization with bags of keypoints," in *Proc. Eur. Conf. Comput. Vis. Int. Workshop Statist. Learn. Comput. Vis.*, 2004, pp. 1–22.

[83] F.-F. L., and P. Perona, "A bayesian hierarchical model for learning natural scene categories," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recog.*, 2005, vol. 2, pp. 524–531.

[84] B. Leibe, A. Leonardis, and B. Schiele, "An implicit shape model for combined object categorization and segmentation," in *Toward Category-Level Object Recognition*. New York, NY, USA: Springer, 2006, pp. 508–524.

[85] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, vol. 2, 2006, pp. 2169–2178.

[86] E. Rosch, C. B. Mervis, W. D. Gray, D. M. Johnson, and P. Boyes-Braem, "Basic objects in natural categories," *Cognit. Psychol.*, vol. 8, no. 3, pp. 382–439, 1976.

[87] D. Song, C. H. Ek, K. Hübner, and D. Kragic, "Multivariate discretization for bayesian network structure learning in robot grasping," in *Proc. IEEE Int. Conf. Robot. Autom.*, Shanghai, China, May 2011, pp. 1944–1950.

[88] Y. Li and N. Pollard, "A Shape Matching Algorithm for synthesizing humanlike enveloping grasps," in *Proc. IEEE/RAS Int. Conf. Human. Robots (Humanoids)*, Dec. 2005, pp. 442–449.

[89] S. El-Khoury and A. Sahbani, "Handling objects by their handles," in *Proc. IROS-2008 Workshop Grasp Task Learn. Imitat.*, 2008, pp. 1–7.

[90] O. Kroemer, E. Ugur, E. Oztop, and J. Peters, "A Kernel-based approach to direct action perception," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2012, pp. 2605–2610.

[91] R. Detry, C. H. Ek, M. Madry, J. Piater, and D. Kragic, "Generalizing grasps across partly similar objects," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2012, pp. 3791–3797.

[92] R. Detry, C. H. Ek, M. Madry, J. Piater, and D. Kragic, "Learning a dictionary of prototypical grasp-predicting parts from grasping experience," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2013, pp. 601–608.

[93] A. Ramisa, G. Alenyà, F. Moreno-Noguer, and C. Torras, "Using depth and appearance features for informed robot grasping of highly wrinkled clothes," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2012, pp. 1703–1708.

[94] A. Boularias, O. Kroemer, and J. Peters, "Learning robot grasping from 3-D images with Markov random fields," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2011, pp. 1548–1553.

[95] L. Montesano and M. Lopes, "Active learning of visual descriptors for grasping using non-parametric smoothed beta distributions," *Robot. Auton. Syst.*, vol. 60, pp. 452–462, 2012.

[96] D. Fischinger and M. Vincze, "Empty the basket—A shape based learning approach for grasping piles of unknown objects," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Oct. 2012, pp. 2051–2057.

[97] N. Bergström, J. Bohg, and D. Kragic, "Integration of visual cues for robotic grasping," in *Computer Vision Systems* Series Lecture Notes in Computer Science, vol. 5815. Heidelberg, Germany: Springer-Verlag, 2009, pp. 245–254.

[98] U. Hillenbrand and M. A. Roa, "Transferring functional grasps through contact warping and local replanning," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Oct. 2012, pp. 2963–2970.

[99] J. Bohg, K. Welke, B. León, M. Do, D. Song, W. Wohlkinger, M. Madry, A. Aldóma, M. Przybylski, T. Asfour, H. Martí, D. Kragic, A. Morales, and M. Vincze, "Task-based grasp adaptation on a humanoid robot," in *Proc. Int. IFAC Symp. Robot Contr.*, Dubrovnik, Croatia, Sep. 2012, pp. 852–859.

[100] N. Curtis and J. Xiao, "Efficient and effective grasping of novel objects through learning and adapting a knowledge base," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Sep. 2008, pp. 2252–2257.

[101] C. Goldfeder and P. Allen, "Data-driven grasping," *Auton. Robots*, vol. 31, pp. 1–20, 2011.

[102] Z. C. Marton, D. Pangercic, N. Blodow, and M. Beetz, "Combined 2-D–3-D categorization and classification for multimodal perception systems," *Int. J. Robot. Res.*, vol. 30, no. 11, pp. 1378–1402, 2011.

[103] D. Rao, Q. V. Le, T. Phoka, M. Quigley, A. Sudsang, and A. Y. Ng, "Grasping novel objects with depth segmentation," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Taipei, Taiwan, Oct. 2010, pp. 2578–2585.

[104] J. Speth, A. Morales, and P. J. Sanz, "Vision-based grasp planning of 3-D objects by extending 2-D contour based algorithms," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2008, pp. 2240–2245.

[105] M. Madry, D. Song, and D. Kragic, "From object categories to grasp transfer using probabilistic reasoning," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2012, pp. 1716–1723.

[106] H. Dang and P. K. Allen, "Semantic grasping: Planning robotic grasps functionally suitable for an object manipulation task," in *Proc. IEEE Int. Conf. Intell. Robots Syst.*, 2012, pp. 1311–1317.

[107] P. A. Viola and M. J. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recog.*, 2001, pp. 511–518.

[108] Z. C. Marton, R. B. Rusu, D. Jain, U. Klank, and M. Beetz, "Probabilistic categorization of kitchen objects in table settings with a composite sensor," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, St. Louis, MO, USA, Oct. 2009, pp. 4777–4784.

[109] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. Int. Conf. Comput, Vis.*, vol. 2, Washington, DC, USA, 1999, pp. 1150–1157.

[110] Z. C. Marton, D. Pangercic, N. Blodow, J. Kleinehellefort, and M. Beetz, "General 3-D modelling of novel objects from a single view," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Taipei, Taiwan, Oct. 2010, pp. 3700–3705.

[111] W. Wohlkinger and M. Vincze, "Shape-based depth image to 3-D model matching and classification with inter-view similarity," in *Proc. IEEE Int. Conf. Robot. Autom.*, San Francisco, CA, USA, Sep. 2011, pp. 4865–4870.

[112] A. Aldoma and M. Vincze, "Pose alignment for 3-D models and single view stereo point clouds based on stable planes," in *Proc. Int. Conf. 3-D Imag., Model., Process., Vis. Transmiss.*, 2011, pp. 374–380.

[113] R. B. Rusu, G. Bradski, R. Thibaux, and J. Hsu, "Fast 3-D recognition and pose using the viewpoint feature histogram," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Taipei, Taiwan, Oct. 2010, pp. 2155–2162.

[114] R. B. Rusu, A. Holzbach, G. Bradski, and M. Beetz, "Detecting and segmenting objects for mobile manipulation," in *Proc. 12th IEEE Int. Conf. Comput. Vis. Workshop Search 3-D Video (S3DV) Held Conjunct.*, Kyoto, Japan, Sep. 2009, pp. 1–9.

[115] D. I. Gonzalez-Aguirre, J. Hoch, S. Rohl, T. Asfour, E. Bayro-Corrochano, and R. Dillmann, "Toward shape-based visual object categorization for humanoid robots," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2011, pp. 5226–5232.

[116] D. Kraft, N. Pugeault, E. Baseski, M. Popovic, D. Kragic, S. Kalkan, F. Wörgötter, and N. Krueger, "Birth of the object: Detection of objectness and extraction of object shape through object action complexes," *Int. J. Human. Robot.*, vol. 5, pp. 247–265, 2009.

[117] M. Popović, G. Kootstra, J. A. Jørgensen, D. Kragic, and N. Krüger, "Grasping unknown objects using an early cognitive vision system for general scene understanding," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, San Francisco, CA, USA, Sep. 2011, pp. 987–994.

[118] G. M. Bone, A. Lambert, and M. Edwards, "Automated modelling and robotic grasping of unknown three-dimensional objects," in *Proc. IEEE Int. Conf. Robot. Autom.*, Pasadena, CA, USA, May 2008, pp. 292–298.

[119] M. Richtsfeld and M. Vincze, "Grasping of unknown objects from a table top," presented at the Eur. Conf. Comput. Vis. Workshop 'Vision Action: Efficient Strategy Cognitive Agents in Complex Environments,' Marseille, France, Sep. 2008.

[120] J. Bohg, M. Johnson-Roberson, B. León, J. Felip, X. Gratal, N. Bergström, D. Kragic, and A. Morales, "Mind the gap—Robotic grasping under incomplete observation," in *Proc. IEEE Int. Conf. Robot. Autom.*, May 2011, pp. 686–693.

[121] J. Stückler, R. Steffens, D. Holz, and S. Behnke, "Real-time 3-D perception and efficient grasp planning for everyday manipulation tasks," presented at the Eur. Conf. Mobile Robots, Örebro, Sweden, Sep. 2011.

[122] A. Maldonado, U. Klank, and M. Beetz, "Robotic grasping of unmodeled objects using time-of-flight range data and finger torque information," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Oct. 2010, pp. 2586–2591.

[123] V. Lippiello, F. Ruggiero, B. Siciliano, and L. Villani, "Visual grasp planning for unknown objects using a multifingered robotic hand," *IEEE/ASME Trans. Mechatronics*, vol. 18, no. 3, pp. 1050–1059, Jun. 2013.

[124] C. Dunes, E. Marchand, C. Collowet, and C. Leroux, "Active rough shape estimation of unknown objects," in *Proc. IEEE Int. Conf. Intell. Robots Syst.*, 2008, pp. 3622–3627.

[125] B. Kehoe, D. Berenson, and K. Goldberg, "Toward cloud-based grasping with uncertainty in shape: Estimating lower bounds on achieving force closure with zero-slip push grasps," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2012, pp. 576–583.

[126] A. Morales, P. J. Sanz, A. P. del Pobil, and A. H. Fagg, "Vision-based three-finger grasp synthesis constrained by hand geometry," *Robot. Auton. Syst.*, vol. 54, no. 6, pp. 496–512, 2006.

[127] K. Hsiao, M. Ciocarlie, and P. Brook, "Bayesian grasp planning," in *Proc. Int. Conf. Robot. Autom. Workshop Mobile Manipulat.: Integr. Percept. Manipulat.*, 2011, pp. 1–8.

[128] A. Aydemir and P. Jensfelt, "Exploiting and modeling local 3-D structure for predicting object locations," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2012, pp. 3885–3892.

[129] G. Metta and P. Fitzpatrick, "Better vision through manipulation," *Adapt. Behav.*, vol. 11, no. 2, pp. 109–128, 2003.

[130] J. Kenney, T. Buckley, and O. Brock, "Interactive segmentation for manipulation in unstructured environments," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2009, pp. 1343–1348.

[131] N. Bergström, C. H. Ek, M. Björkman, and D. Kragic, "Generating object hypotheses in natural scenes through human-robot interaction," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, San Francisco, CA, USA, Sep. 2011, pp. 827–833.

[132] D. Schiebener, A. Ude, J. Morimoto, T. Asfour, and R. Dillmann, "Segmentation and learning of unknown objects through physical interaction," in *Proc. IEEE/RAS Int. Conf. Human. Robots (Humanoids)*, Oct. 2011, pp. 500–506.

[133] M. Tenorth, A. C. Perzylo, R. Lafrenz, and M. Beetz, "The RoboEarth language: Representing and exchanging knowledge about actions, objects, and environments," in *Proc. IEEE Int. Conf. Robot. Autom.*, St. Paul, MN, USA, May. 2012, pp. 1284–1289.

[134] Y. Bekiroglu, R. Detry, and D. Kragic, "Joint observation of object pose and tactile imprints for online grasp stability assessment," presented at the IEEE Int. Conf. Robot. Autom. Manipulation Under Uncertainty, Shanghai, China, 2011.

[135] N. Hudson, T. Howard, J. Ma, A. Jain, M. Bajracharya, S. Myint, C. Kuo, L. Matthies, P. Backes, P. Hebert, T. Fuchs, and J. Burdick, "End-to-end dexterous manipulation with deliberate interactive estimation," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2012, pp. 2371–2378.

[136] M. Kazemi, J.-S. Valois, J. A. D. Bagnell, and N. Pollard, "Robust object grasping using force compliant motion primitives," presented at the Robotics: Sci. Syst. Conf., Sydney, Australia, Jul. 2012.

[137] X. Gratal, J. Romero, J. Bohg, and D. Kragic, "Visual servoing on unknown objects," *Mechatronics*, vol. 22, no. 4, pp. 423–435, 2012.

[138] C. Goldfeder, M. Ciocarlie, H. Dang, and P. K. Allen, "The columbia grasp database," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2009, pp. 3343–3349.

[139] Mærsk Mc-Kinney Møller Instituttet, University of Southern Denmark. "VisGraB-a benchmark for vision-based grasping of unknown objects." [Online]. Available: www.robwork.dk/visgrab

[140] Healthcare Robotics Labs. Georgia Inst. Technol. "Pr2 playpen." [Online]. Available: ros.org/wiki/pr2_playpen

[141] DARPA. "ARM | Autonomous Robotic Manipulation." [Online]. Available: www.thearmrobot.com

[142] RoboCup@Home. [Online]. Available: www.ai.rug.nl/robocupathome

[143] S. Ulbrich, D. Kappler, T. Asfour, N. Vahrenkamp, A. Bierbaum, M. Przybylski, and R. Dillmann, "The OpenGRASP benchmarking suite: An environment for the comparative analysis of grasping and dexterous manipulation," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Sep. 2011, pp. 1761–1767.

Jeannette Bohg (M'13) received the Diploma degree in computer science from the Technical University Dresden, Dresden, Germany; the M.Sc. degree in applied information technology from Chalmers University of Technology, Göteborg, Sweden; and the Ph.D. degree from the Royal Institute of Technology, Stockholm, Sweden, in 2011.

She is a Research Scientist at the Autonomous Motion Department, Max-Planck-Institute for Intelligent Systems Tübingen, Germany. Her research interest includes the intersection between robotic grasping and computer vision. Specifically, she is interested in the integration of multiple sensor modalities and information sources for enhanced scene understanding. She demonstrated how this work can be used in an active perception framework and leads to improved grasping and manipulation.

Antonio Morales (M'04) received the Ph.D. degree in computer science engineering from Universitat Jaume I, Castellón, Spain, in January 2004.

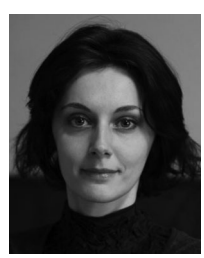
He is currently an Associate Professor with the Department of Computer Engineering and Science, Universitat Jaume I of Castelló. He is a Leading Researcher at the Robotic Intelligence Laboratory, Universitat Jaume I. He has been a Principal Investigator of the European Cognitive Systems Integrated Project GRASP and on several national and locally funded research projects. His research interests include reactive robot grasping and manipulation, as well as on the development of robot simulation.

Dr. Morales has served as an Associated Editor for the IEEE International Conference on Robotics and Automation and for the IEEE/RSJ International Conference on Intelligent Robots and Systems. He has also served as a Reviewer for multiple relevant journals and conferences. He is member of the IEEE Robotics and Automation Society since 1998.

Tamim Asfour (M'05) received the Diploma degree in electrical engineering and the Ph.D. degree in computer science from the University of Karlsruhe, Karlsruhe, Germany.

He is currently a Professor with the Institute for Anthropomatics, Karlsruhe Institute of Technology. He is a Developer and Leader of the development team of the ARMAR humanoid robot family. His research interests include humanoid robotics, humanoid mechatronics and mechano-informatics, grasping and dexterous manipulation, action learning from human observation and goal-directed imitation learning, active vision and active touch, whole-body motion planning, robot software, and hardware control architecture and system integration. Specifically, he has been researching the engineering of high-performance 24/7 humanoid robots able to predict, act, and interact in the real world.

Dr. Asfour is the European Chair of the IEEE Robotics and Automated Society Technical Committee on Humanoid Robots and a member of the Executive Board of the German Robotics Association Deutsche Gesellschaft für Robotik (DGR).



Danica Kragic (SM'12) received the M.Sc. degree in mechanical engineering from the Technical University of Rijeka, Rijeka, Croatia, in 1995 and the Ph.D. degree in computer science from the Royal Institute of Technology (KTH), Stockholm, Sweden, in 2001.

She is currently a Professor with the School of Computer Science and Communication, KTH. Her research include computer vision, object grasping and manipulation, and human–robot interaction. Her recent work explores different learning methods for formalizing the models for integrated representation of objects and actions that can be applied on them. This work has demonstrated how robots can achieve scene understanding through active exploration and how full body tracking of humans can be made more efficient.

Dr. Kragic received the 2007 IEEE Robotics and Automation Society Early Academic Career Award. She is a member of the Swedish Royal Academy of Sciences and the Swedish Young Academy. She has chaired the IEEE Robotics and Automation Society (RAS) Technical Committee on Computer and Robot Vision and, since 2009, has served as an IEEE RAS Administrative Committee member.