A Hybrid Approach For Grasping 3D Objects

Anis Sahbani and Sahar El-Khoury

Abstract—The paper presents a novel strategy that learns to associate a grasp to an unknown object/task. A hybrid approach combining empirical and analytical methods is proposed. The empirical step ensures task-compatibility by learning to identify the object graspable part in accordance with humans choice. The analytical step permits contact points generation guaranteeing the grasp stability. The robotic hand kinematics are also taken into account. The corresponding results are illustrated using GraspIt interface [1].

I. INTRODUCTION

The first goal of every grasping strategy is to ensure stability. A grasp is stable if a small disturbance, on the object position or finger force, generates a restoring wrench that tends to bring the system back to its original configuration [5]. Nguyen [21] introduces an algorithm for constructing stable grasps. Nguyen also proves that all 3D force-closure grasps can be made stable. A grasp is force-closure when the fingers can apply appropriate forces on the object to produce wrenches in any direction [22]. Obviously, stability is a necessary but not a sufficient condition for a grasping strategy. When we reach out to grasp an object, we have a goal in our mind or a task to accomplish. Thus, in order to successfully perform the task, the grasp should also be compatible with the task requirements. Computing task-oriented grasps is consequently crucial for a grasping strategy. Finally, because of the variety of objects shapes and sizes, a grasping strategy should always be prepared to grasp new objects. Thus, it should ensure stability, task compatibility and adaptability to novel objects. In other terms, a grasp synthesis strategy should always have an answer to the following question: where to grasp a novel object in order to accomplish a task? Analytical and empirical approaches answer this question differently.

Analytical Approaches consider kinematics and dynamics formulations in determining grasps. Many works have been developed to compute force-closure grasps [17], [21], [24] or even optimal force-closure grasps achieving the most desirable performance in resisting external wrench loads [19], [23]. These approaches find stable grasps adapted for pick and place operations and are not task-oriented. Only few works [6], [7], [8] take the task into account. These analytical approaches suffer from a major problem: computational complexity when trying to model task requirements. Thus, while the selection of task-oriented optimal grasp is very easy for a human hand, it is still a complicated process for a robot hand.

Empirical grasping methods avoid the computational complexity of analytical techniques by attempting to mimic human grasping strategies. Empirical strategies for grasp planning can be divided into two main kinds: systems based on the observation of the object to be grasped [9], [10], [18] and systems based on the observation of a human performing the grasp [11], [14], [15], [16], [20]. The former techniques generally learn to associate objects characteristics with a hand preshape, while in the latter, a robot observes a human operator performing a grasp and try then to reproduce the same grasp. This technique is called in the literature learning by demonstration approach. Empirical systems based on objects observation are adapted to new objects but generate a lot of possible grasping positions and fail to select the one that best suits the task. When trying to do this autonomously, they encounter the same problem of analytical task-oriented methods, which is task modelling. Empirical systems based on the observation of humans overcome task modelling difficulty by imitating humans grasping gesture. However, these systems are not fully autonomous when they face an object completely new.

Thus, a strategy that learns to associate a grasp to an unknown object/task is still an unsolved problem. We believe that neither analytical nor empirical approaches can fulfill by themselves the constraints of stability, task compatibility and adaptability to new objects. We propose, in this paper, a hybrid approach combining empirical and analytical methods to solve the problem. The empirical step will ensure task-compatibility by avoiding the analytical approaches task-modelling complexity. The analytical step permits contact points generation guaranteeing the grasp stability.

II. THE PROPOSED APPROACH: EMPIRICAL STEP

Humans are capable of reaching and grasping novel objects with great dexterity. To ensure these skills and to interact with a human’s world, robots must be capable of using their hands proficiently. Thus, robots should handle objects in the same manner as humans. What are the factors taken into consideration when choosing a specific grasp configuration? What should the grasping algorithm learns in order to pick a new object in the same manner as humans? In other words, what parameters are relevant to new objects grasping? Are these parameters related to the hand characteristics? Are they related to the object features? By taking inspiration from the Recognition By Components...
theory of Biederman [25], RBC, we propose a strategy to imitate human’s choice of the objects grasping component. Biederman suggests that people are able to recognize objects (even unfamiliar) by segmenting them into parts at regions of deep concavities. Thus, objects identification does not depend on our familiarity with them. We conduct the same process for any object, whether it is familiar or unfamiliar. But what about grasping an unfamiliar object? Does its part decomposition emphasize a specific grasp? When considering objects we use for everyday tasks on a part-representation level, we can make the following assumptions:

- Objects are equipped with a part designed specifically to make their grasp easier. Figure 1 shows some familiar objects. The black part indicates the component that humans choose to grasp these objects. It is also the part that satisfies the task requirements. This part is what we call the object graspable part or more simply the object handle.

![Fig. 1. The black part indicates the object handle.](image1)

- Objects with similar components are grasped in the same manner. Bags, buckets, mugs and cups are roughly composed of a cylinder and a curved cylinder. Even though the arrangement of these components is different for these objects, they are all grasped by their curved component (Fig. 2). Thus, the choice of an object graspable part is influenced by the shape of its constituting single parts. Objects parts orientation is less relevant to that choice.

![Fig. 2. The choice of an object graspable part is influenced by the shape of its constituting parts, independently from their orientations, i.e: a) a mug, b) a bucket and c) a bag are all grasped by their curved part.](image2)

- The relative sizes of object components is crucial for the graspable part selection. Let us examine some alcohol glasses shapes and sizes. We consider wine, champagne and brandy glasses. Although, all these glasses are composed of three parts: the bowl, the stem and the foot, they are grasped differently (Fig. 3). Wine and champagne glasses have a long stem. They are designed to be held by the stem to help prevent the heat from the hand from warming the alcohol. On the other hand, brandy glasses have a short stem. They are designed to be held by the bowl.

![Fig. 3. Roughly approximation of: a) a wine glass, b) a champagne flute and c) a brandy glass.](image3)

In summary, we can say that objects are designed in a way to make their grasp easier and in accordance with their functions and information about an unknown object parts shapes and sizes may emphasize a specific part for its grasping. This leads to our "Grasping By Components" strategy. It aims at finding, for an unknown object, its graspable part. Thus, objects are represented as a set of components. A learning process permits then to use geometric representation of the object components to perform an analogue of the human choice of the grasping component. Thus, our approach will learn to imitate humans selection of the object graspable part. The different steps of the proposed approach are detailed in the following.

A. Objects Representation

The selection of the object graspable part is influenced by the size and shape of its components. Thus, objects are represented as an assembly of geometric primitives. Starting from a 3D surface model, a part decomposition step is performed to segment the object into its constituent single parts. For this end, a segmentation algorithm based on the Gaussian curvature and the concaveness estimation is used [13]. This method has a main advantage over the existing ones in the literature. It uses multi-ring neighborhood in order to compute a 3D object surface features such as the gaussian curvature. Thus, when a model is densely represented with polygonal faces, a multi-ring neighborhood permit to accurately catch their geometric behavior. This segmentation approach succeeds in decomposing low resolution as well as high resolution 3D laser scanned objects (Fig. 4).

Object segmentation produces a set of parts. The next task is to generate a description for each one. Each part is represented by a superquadric. With only a few parameters, superquadrics can represent a large variety of standard geometric solids as well as smooth shapes. In order to have a manageable number of superquadrics shapes, we have chosen 7 representative models that span the space.
of superellipsoids: box, cylinder, sphere, bent box, bent cylinder, tapered box and tapered cylinder.

III. LEARNING THE GRASPING COMPONENT

The proposed algorithm learns to use object components shapes and sizes in order to select the grasping part. Supervised learning [26] is used for this task, with synthetic objects (generated using computer graphics) as training data. The training objects are the result of the assembly of two volumetric primitives. Our supervised learning requires a set of objects that can potentially span the space of two superquadrics assembly. Therefore, the choice of the training objects is important to effectively sub-sample this space. We use 12 objects for the training set (Fig. 1). We mentioned previously that 7 superquadrics will be used to model our objects. Thus, the training objects components are chosen to span these 7 superquadrics shapes with different sizes. Figure (5) shows the steps for generating the training data. It shows first the initial object, its decomposition into single parts, the approximation of each part with a superquadric and finally its corresponding grasping part according to human choice. A multi-layer perceptron, with one hidden layer, is trained with a typical backpropagation learning algorithm [26] in order to select the grasping part of a two-component object. For multi-part objects, the decision of the grasping component is taken by considering the object parts two by two. In other words, the algorithm starts by choosing a grasping component between two parts of the object. The chosen part is then compared with another component and so on until finding the handle of the multi-part object.

IV. THE PROPOSED APPROACH: ANALYTICAL STEP

At this point, we are able to identify an unknown object handle. This section aims at computing contact points on the corresponding handle that ensure stability. Force-closure property characterizes the stability of a grasp. According to the definition of Salisbury [4], a grasp is force-closure if and only if any external wrench can be balanced by the wrenches at the fingertips. This condition is equivalent to that the origin of the wrench space lies strictly inside the convex hull of the primitive contact wrenches [2], [3]. In the past few years, several force-closure tests were also proposed [17], [19]. Generating good force-closure grasps with the previously detailed force-closure necessary and sufficient conditions require considerable computation time. In order to find such grasps, they perform an exhaustive search for the best n-finger force-closure grasp of an object modeled by N points which would take time in the order of 0(N^6). Thus, heuristic approaches were proposed to improve performance [28], [29]. They generate many grasp candidates by selecting contacts on the object surface. Then, these grasps are filtered with a necessary but not sufficient force-closure tests. The grasps that pass the filter may or may not be force-closure. In other words the filter reports false positive but not false negative force-closure grasps. The selected grasps are tested afterwards for force-closure. Another way to improve performance, proposed in the literature, is to use a simplified version of the object’s geometry consisting only of shape primitives such as spheres, cylinders, cones and boxes. Then, for each shape, define a set of grasping strategies [30]. This also reduces the number of grasps tested for force-closure.

Our work is an hybrid solution to the force-closure grasp synthesis. The number of grasps tested for force-closure is reduced since we do not consider the object as a whole but we are interested in generating force-closure grasps only on the object natural grasping component. We also propose a new sufficient but not necessary force-closure test. Thus, grasps that pass the filter ensure necessarily force-closure. Our heuristic is original in the sense that it permits simultaneously fast computation and good quality force-closure grasps generation.

A. Force-closure Test

In order to ensure force-closure or determine grasp wrenches that positively span the entire 6-dimensional wrench space, one needs to find: (1) primitive wrenches that constitute a 6D basis and (2) a primitive wrench that can be expressed as a negative linear combination of that basis. But, in which case wrenches associated to hard contact points may form a basis of the wrench space? May a representation in the 3D space of 6D wrenches facilitate the problem? Plücker coordinates represents 6D wrenches as lines and Grassmann algebra studies the rank of such lines. We use these two studies to prove that wrenches, associated to any three non-aligned contact points of 3D objects, form a basis of the 6D
wrench space. Consequently, a force-closure test is given by
the following proposition:

**Proposition 1:** Assume that the grasp of \( n - 1 \) non-aligned fingers is not force-closure. Suppose that \( \{b_i\}_{i=1}^k \) is the \( k \)-dimensional (where \( k = 6 \)) basis associated to their corre-
responding contact wrenches. A sufficient condition for a \( n \)-finger force-closure grasp is that there exists a contact wrench \( \gamma \) such that:

- \( \gamma \) is inside the linearized friction cone of the \( n \)th finger
- \( \gamma = \sum_{i=1}^k \beta_i b_i, \beta_i < 0 \)
- \( \gamma = \sum_{i=1}^k \beta_i b_i \Rightarrow \gamma = B\hat{\beta} \Rightarrow \beta = B^{-1}\gamma \)

where \( B = [b_1, b_2, \ldots, b_k] \) is a \( k \times k \) matrix and \( \hat{\beta} = [\beta_1, \beta_2, \ldots, \beta_k]^T \) is a \( k \times 1 \) strictly negative vector. Thus, a simple multiplication by \( B^{-1} \) permits to test if a contact wrench \( \gamma \), and consequently the location of the \( n \)th contact point, ensures a force-closure grasp.

**Proof:** The reader should refer to [27].

**B. Generating n-finger force-closure grasps**

We showed (proposition 1) that to achieve force-closure, we generate randomly locations of \( n - 1 \) non-aligned fingers. A position of the \( n \)th finger is chosen such that an associated contact wrench can be uniquely expressed as a strictly negative linear combination of one of the first generated \( n - 1 \) fingers wrench basis. Our objective is to ensure fast robust force-closure grasps generation. In our case, force-
closure grasps fast computation and robustness are strongly
linked. In order to understand how the two latter are tied together, one should notice that generating a \( n \)-finger good grasp will depend on the generation of the first \( n - 1 \) fingers. A good choice of their locations will induce on one hand robust grasps and on the other hand more locations for the \( n \)th finger on the object surface guaranteeing force-closure and consequently fast computation. Thus, we use a quality criterion quantifying a good placement of the \( n - 1 \) first fingers. This criterion is based on the computation of these fingers wrenches basis volumes [27].

**V. GRASPING BY TAKING INTO ACCOUNT THE HAND KINEMATICS**

At this point, our grasping strategy identifies an unknown object handle and generates contact points on it with the only
constraint of stability. Dealing with a robotic hand model
induces additional kinematical and geometrical constraints. Taking these constraints into account results in limiting possible locations for the contact points on the graspable part. The latter should be kinematically feasible for the fingers and they should also avoid collision with the hand, the remaining fingers and the object. Consequently, these contacts should be generated in respect of the accessibility domains of the fingers. Furthermore, a grasp involves several closed kinematic loops between the fingers and the object.

Randomly generation of a closed kinematic chain is very
difficult. In order to handle these closed kinematic chains, we propose to adapt the RLG (Random Loop Generator) algorithm [31] to our grasping strategy. RLG aims at handling closed kinematic loops by dividing them into active and passive parts. The idea of the algorithm is to reduce the closed kinematic chain complexity iteratively until the active part becomes reachable by all passive chain segments simultaneously. In our case, the object is the active part while the fingers constitute the passive parts. A grasp can occur when the object is reachable by all the fingers. The reachable workspace of a kinematic chain is defined as the volume which the end-effectors can reach. RLG approximates such volume with a sphere. Figure 6 illustrates an example of the reachable workspace of a finger. It also shows the intersection between this space and the object. Thus, the finger should be placed on this intersection. The placement of the first finger is then taken into account when computing the second finger reachable workspace and so on until the placement of all fingers. We modify our grasping strategy to take these constraints into consideration.

![Fig. 6. Reachable workspace approximation of a finger.](image-url)
contact location $CP_i$ is then randomly chosen in $RW_i$. This guarantees that the inverse geometrical model of the finger existence. After placing the $n-1$ fingers, the $nth$ finger location is computed with Force Closure in order to ensure the grasp stability according to proposition 1.

VI. EXPERIMENTAL RESULTS

Different experiments were conducted to test the ability of the algorithm to find the object graspable part corresponding to humans choice. These experiments aim at testing the ability of the learning algorithm to generalize.

First, we tested the algorithm on objects belonging to the same categories as the training data but of different shapes and sizes. These objects are such as bottles, spoons, forks, mugs, knives, pencils etc. The motivation behind this experiment is that if our algorithm does not work on objects similar to the training data, then we must conclude that our feature set is not sufficiently discriminative. Fortunately, for such objects, the algorithm generalizes very well and was capable of finding each time the handle that human choose to grasp the corresponding object. In a second time we tested the algorithm on 54 objects that are completely different from those of the training set. This experiment is useful to test the algorithm ability to generalize to completely novel objects. Seven subjects were asked to grasp these objects in order to accomplish a task. We do not specify the task that should be performed. The subjects were supposed to identify objects graspable parts whether they recognize the object or not. Twenty seven objects, AO (Agreed Objects), were grasped by the same manner. On the other hand, the remaining 27 objects, CO (Confusing Objects), induced confusion and the seven subjects chose different parts to grasp them. We remind the reader that our aim is to imitate humans choice of the graspable part. The distinction between AO and CO objects is necessary for measuring our algorithm performance. Their success grasp rate is computed differently.

For AO objects, whenever the algorithm selects for grasping a part different from the one identified by the seven subjects, it is considered a failure. The system succeeds to find the correct graspable parts for 22 AO objects, which corresponds to a successful grasp rate of 81% (table I). Since humans grasp CO objects in various ways, two successful rate may be computed: a successful grasp may be a grasp that identifies the object part chosen by most people (MP), or a successful grasp may be a grasp that identifies a part chosen by at least one person (ALO). Otherwise, failure occurs. The algorithm succeeds to find, for 15 CO objects, the part selected by most people. This corresponds to a successful grasp rate of 55%. When considering a grasp rate on the basis of “at least chosen by one person”, the algorithm perform well for 23 CO objects which corresponds to a rate of 85%.

The obtained success grasp rate show that features such as sizes and shapes of novel objects subparts are about 80% discriminative to determine the object grasping part. Table (II) shows the generation of a four-finger force-closure grasp on the grasping part of 3D objects model with different resolutions, a spoon modelled with 629 vertices, a bottle (7360 vertices) and a mug (183534 vertices). Their corresponding force-closure grasp computation time are respectively 2.59 sec, 8.87 sec and 4.46 min. These computation times include segmentation/approximation/selection steps and force-closure grasps generation. Finally, Tables III and IV show several grasps obtained using DLR and Rutgers hands models and GraspIT interface. The latter uses PQP algorithm to detect collisions [1].

![Fig. 7. Examples of AO objects: segmentation into different parts (first row), system choice (black part) and humans choice (cross-marked parts).](image)

**TABLE I**

SUCCESS GRASP RATE FOR AO AND CO OBJECTS.

<table>
<thead>
<tr>
<th>Object</th>
<th>AO objects</th>
<th>CO objects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number</td>
<td>All people</td>
<td>MP</td>
</tr>
<tr>
<td></td>
<td>27</td>
<td>27</td>
</tr>
<tr>
<td>Number grasped successfully</td>
<td>22</td>
<td>15</td>
</tr>
<tr>
<td>Success Grasp Rate</td>
<td>81%</td>
<td>55%</td>
</tr>
</tbody>
</table>

![Fig. 8. Examples of CO objects: segmentation into different parts (first row), system choice (black part) and humans choice (cross-marked parts).](image)

**TABLE II**

GENERATING 4-FINGER FORCE-CLOSURE GRASPS FOR 3D OBJECTS.

<table>
<thead>
<tr>
<th>Objects Grasps</th>
<th>Vertices Number</th>
<th>GBC time</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>629</td>
<td>2.59 sec</td>
</tr>
<tr>
<td></td>
<td>7360</td>
<td>8.87 sec</td>
</tr>
<tr>
<td></td>
<td>183534</td>
<td>4.46 min</td>
</tr>
</tbody>
</table>
TABLE III
Generating 4-finger force-closure grasps using DLR hand.

TABLE IV
Generating 5-finger force-closure grasps RUTGERS hand.

VII. Conclusion

We proposed a grasping strategy to predict grasps of unknown objects that conform humans grasping. Several experiments have characterized how well learned grasps generalize to objects that the algorithm has no experience with. Results show that features such as objects sub-parts shapes and sizes are about 80% discriminative to grasping. In other words, an unknown object appropriate grasp can be found only by using information on its constituting components shapes and sizes without any task modeling. Once the graspable part is identified, contact points should be determined. For this purpose, we proposed a new sufficient condition for generating n-finger force-closure grasps. Finally, an algorithm for generating contact points on a novel object that takes the hand kinematics into account was proposed. The efficiency of the proposed approach is confirmed by several experiments.

Acknowledgments

This work is supported by the French (Agence Nationale pour la Recherche) ANR programs under grant No C08527 (ABILIS Project).

References