

Article

Model-Based Grasping of Unknown Objects from a Random Pile

Bruno Sauvet ^{1,†}, François Lévesque ^{2,†}, SeungJae Park ^{3,†}, Philippe Cardou ^{1,*,†} and Clément Gosselin ^{1,†}

¹ Laboratoire de robotique, Département de génie mécanique, Université Laval, 1065 Avenue de la Médecine, Québec, QC G1V 0A6, Canada; bruno.sauvet@gmail.com (B.S.); gosselin@gmc.ulaval.ca (C.G.)

² Umanx, 2900, Einstein St, Quebec City, QC G1X 4B3, Canada; francois.levesque@umanx.com

³ Cypress Robotics Corp, 65 Vicars Rd, Kamloops, BC V2C 0G3, Canada; steven@cypressrobotics.ca

* Correspondence: pcardou@gmc.ulaval.ca

† These authors contributed equally to this work.

Received: 8 July 2019; Accepted: 2 September 2019; Published: 6 September 2019



Abstract: Grasping an unknown object in a pile is no easy task for a robot—it is often difficult to distinguish different objects; objects occlude one another; object proximity limits the number of feasible grasps available; and so forth. In this paper, we propose a simple approach to grasping unknown objects one by one from a random pile. The proposed method is divided into three main actions—over-segmentation of the images, a decision algorithm and ranking according to a grasp robustness index. Thus, the robot is able to distinguish the objects from the pile, choose the best candidate for grasping among these objects, and pick the most robust grasp for this candidate. With this approach, we can clear out a random pile of unknown objects, as shown in the experiments reported herein.

Keywords: grasping; grasp quality; pile of objects; object detection; unknown objects

1. Introduction

The automatic grasping and manipulation of random objects from a pile is not regarded as an easy task. Although it has been studied for a long time [1], both the interest and the difficulties remain. This is particularly true for bin-picking problems, which continue to be an active field of research [2,3]. Among these challenges, we find that an object could be partially or entirely hidden by others, that there could be no suitable grasp for a given object and that interferences with the pile could prevent the robot from reaching a suitable grasp.

In general, grasping an isolated object requires acquiring information about the object, finding a reliable grasp while taking into account the robot dynamics, and the task that is to be performed with the object—for example, placing, inserting, drilling. Many grasping methods already exist, several of them having been reviewed in References [4–6]. Two types of approaches can be distinguished: (i) the empirical approach [7,8], based on the behaviour of the human hand; and (ii) the analytical approach [9,10], based on the physical and mechanical properties of grasps. The approaches can also be classified according to the degree of a priori knowledge of the grasped objects that is available. An object can be known [1,11,12], familiar [13] or unknown [14,15]. Recently, researchers have proposed resorting to the features in the environment in order to perform difficult grasps. For instance, Eppner et al. [16] use the flat surface on which the object lies to guide the hand fingers under it, or the object is pushed slightly beyond the edge of a table so that a finger can reach under it. This prompts yet another classification of grasping strategies—the extrinsic grasps, where the hand and object interact with the environment to perform the grasp, and the intrinsic grasps, which only involve the hand and the grasped object.

Several research initiatives have succeeded in manipulating objects from all these categories in a cluttered environment or to clear out a pile of such objects, except perhaps for the case of extrinsic grasps mentioned above. Reports of these initiatives are reviewed in Section 2. Some of them resort to learning, whether it be to identify objects or to learn new grasps. Others use interactions with the pile (e.g., poking, pushing) to modify its shape and reach its constituting objects. Another initiative proposes to recognise elementary forms or faces within the objects, or work with known objects.

Approaches based on learning require large amounts of data and their behaviour is generally unpredictable when confronted with previously unseen situations or objects. As for interacting physically with the pile prior to grasping, this strategy may prove useful when no feasible grasp is found in a pile. In applications where speed is critical, this solution may not be acceptable, however. In the experiments reported in this article, feasible grasps could always be found in the piles of objects to be cleared by the robot. Accordingly, we did not find it useful to include physical interactions with the pile, although the proposed algorithm lends itself to the addition of such a step.

In practically all cases, artificial vision is one of the keys required to solving the problem. It is not the only key required, however, and the robustness of the grasping method is also critical to account for inevitable errors introduced by the vision and manipulation systems. Therefore, grasps should not be determined based on their sole feasibility, but on some measure of their robustness.

In this article, we propose a straightforward method of grasping unknown objects one by one from a random pile until exhaustion. The method relies on images from a low-cost 3D vision sensor providing average-quality information of the pile. We resort to a simple segmentation method in order to break down the pile into separate objects. A grasp quality measure is proposed in order to rank the available grasp hypotheses. Experimental tests over 50 random piles are reported to assess the efficiency of this simple scheme.

The rest of the article is organised as follows: Section 2 is a review of prior art on the grasping of unknown objects from cluttered environments; Section 3 describes the experimental setup used to evaluate the proposed method; Section 4 presents the proposed algorithm and describes its underlying functions; the experimental tests and results are presented and analysed in Section 5; finally, Section 6 draws some conclusions from the experimental results discussed in the preceding section.

2. Related Work

In this section, related work is reviewed and classified according to the strategies used to grasp unknown objects, namely, using machine learning; using interactions with objects; or using the geometric properties of the objects and the gripper.

2.1. Grasping Strategies Based on Machine Learning

Based on machine learning, References [17,18] use two or more images of a familiar object to identify grasping points in 3D space from past experience. Hence, the grasps are learned from images and this method can be classified as an empirical approach. In Reference [19], the authors use a learning-based approach and 1500 interactions for learning to clear out a pile of objects. The unknown objects are segmented into hypothesised object facets. If necessary, the robot pushes or pulls the pile to change its shape or separate the objects. This method encounters problems when two objects stick together. The supervised learning proposed to select the sequence of actions to grasp each object leads to a success rate of 53% over the chosen set of objects. By comparison, a human selection of the sequence of actions leads to a success rate of 65%.

In References [15,20], the authors present a method to grasp unknown objects from a pile. The grasping method is based on topography and uses machine learning trained over 21,000 examples. It is successful 85% of the time over the tested set of ten unknown objects, and presents an average of 1.31 actions per object, an action being defined as an attempted grasp. In Reference [21] a network was trained to predict an optimal grasp. According to their approach, a gripper image is used to visualise

the locations of plausible grasps. The training is self-supervised over a period of 700 h, yielding a success rate of 66%. With unknown and new objects, an average of 2.6 actions per object is produced.

Similarly, in Reference [22], an approach is proposed using deep learning with a continuous motion feedback, instead of using a gripper image patch. The learning phase is based on a data set of over 800,000 grasps, which required between 6 and 14 robots working for two months. The proposed deep-learning yielded a success rate of 80%, with approximately 1.17 actions per object, after having been trained over 10 M images.

Using machine learning and geometric properties, Reference [23] presents an algorithm able to grasp unknown objects in cluttered environments. The training is realised with 6500 labeled examples composed from 18 objects and 36 configurations. The authors obtain a success rate of 73% with approximately 1.13 actions per object. They attribute the failed grasps to kinematic modelling inaccuracies, vision errors caused by their algorithm, to dropped object following a successful grasp and to collisions with the environment.

2.2. Grasping Strategies Resorting to Pile Interaction

In Reference [24], Van Hoof et al. use probabilistic segmentation and interaction with a cluttered environment to create a map of the scene, namely, to determine its constituting objects, their contours and their poses. Their method requires a number of interactions with the scene that is smaller than the number of objects it contains, and it appears successful in the reported tests.

Without resorting to machine learning, Katz et al. [25] propose a segmentation algorithm for grasping. In this method, the segmentation is verified by “poking” the pile of objects to identify groups of facets belonging to the same rigid body. One of the objects with a verified segmentation is then grasped following a heuristic based on the object centroid and its principal axes. The process is repeated until the pile is cleared, so that clearing a pile of n objects, requires $2n$ actions.

Reference [26] proposes a framework to manipulate and classify familiar objects in a pile based on interactions (push or pull) with the pile. Three different strategies are used (grasp only; poke and grasp; adaptative interaction) for a success rate equal to approximately 70% for all strategies. The “grasp-only” strategy required more grasping attempts before clearing a pile, but less time before success, when compared to the poke-and-grasp and adaptive strategies.

In the case where the objects are known, Gupta and Sukhatme [27] use a segmentation method to recognise, grasp and sort Duplo bricks. When it is difficult to distinguish between two different bricks, the pile is either poked, spread or tumbled to successfully achieve the grasp. This method reaches 1.10 actions per object and an overall success rate of 97% over these known objects.

To distinguish unknown objects from a pile, Hermans et al. [28] resort to a series of pushing actions whereby the objects are separated from one another. These movements allow a robot to assess whether edges detected by vision correspond to true object boundaries. A histogram of the edge directions is created, in order to determine which groups of edges are rigidly connected. For non-textured objects, 4.8 actions per object are needed. For objects with varying levels of textures, 2.0 actions per object are needed.

2.3. Geometric Properties

In order to manipulate unknown objects, Eppner and Brock [29] propose to adapt the grasp according to the primitive—cylinder, spherical, box or disk—that the object resembles, and also according to the environment of this object. Twenty-three objects are used as a testing set, and five different viewpoints with five strategies—corresponding to the four primitives plus the strategy of the object centroid—are evaluated, for a total of 267 trials. The authors demonstrate a grasp prediction success rate of 89%, which corresponds to 1.11 actions per object. The centroid strategy turns out to be the most reliable for the proposed testing set.

In References [2,30], Buchholz et al. propose to grasp unknown objects with a gripper pose estimation using matched filters. With this method, they succeed to grasp unknown objects in a bin with a success rate of 95.4% for 261 trials and 1.04 actions per object.

A significant portion of the previous research assumes familiar or known objects. Almost all of it concerns objects that can be approximated by common primitives—e.g., cubes, cylinders and spheres—but none works with complex shapes. In this paper, the goal is to grasp unknown objects individually from a random pile until the pile is completely cleared.

The different approaches based on learning require large amounts of data. More importantly, their behaviour is difficult to predict or to adjust, which is why they are often excluded from industrial applications where reliability and repairability are essential. Other strategies resort to interactions with the pile to distinguish its constituting objects. This interaction is always time consuming, as high speeds would risk scattering the objects too far. Another strategy relies on geometric properties of the objects to generate the appropriate grasp. This model-based approach can be less flexible than those based on machine learning, and can also be time consuming, when too many details of the object geometry need to be accounted for. None of these strategies assesses the degree of robustness of the grasp. The grasp robustness quality is the idea behind the method proposed in this article, which is akin to those that rely on the geometric properties of the objects. Our grasp-robustness index is computed in real time to determine the most appropriate grasp to pick up an object from a random pile.

3. Testbed

The experimental setup used in this work consists of a 3D camera (Kinect from Microsoft), a six-degree-of-freedom (6-dof) serial robot (UR5 robot from Universal Robots) and an underactuated gripper (two-finger 85 mm gripper from Robotiq), as shown in Figure 1. Communication between the components of the setup is performed through the Robot Operating System (ROS). In the experiments, a fixed reference frame is defined as illustrated in Figure 1, with the X and Y axes in the plane of the table and the Z axis perpendicular to it.

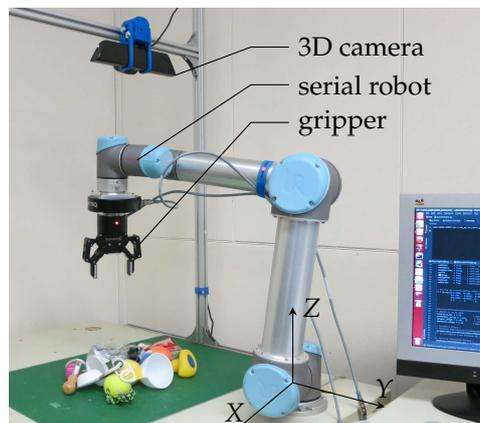


Figure 1. The experimental setup used in this research.

4. Proposed Algorithm

The sequence of steps composing the algorithm is represented in the flow-chart of Figure 2 and can be summarised as follows:

1. Initialise system: Standard initialisation of software and hardware, that is, connection to the robotic manipulator, the gripper, the 3D camera, trajectory planning service and launch of GUI.
2. Detect objects: Using the RGB-D output of the 3D camera, detect the objects in the robot workspace.

3. Generate grasp hypotheses: Several grasp hypotheses are generated for each shape of object inferred from the vision system.
4. Rank of objects by interference: A sorting function allows to sort the objects according to the likelihood of collision when attempting to grasp them.
5. Rank grasps by robustness: Among the interference-free objects, the best grasp pose is chosen according to a criterion referred to as the robustness of static equilibrium [10].
6. Execute best grasp: Send a command to the robot to effectuate the grasp.

We describe in detail each step of the algorithm shown in Figure 2 in its corresponding subsection below.

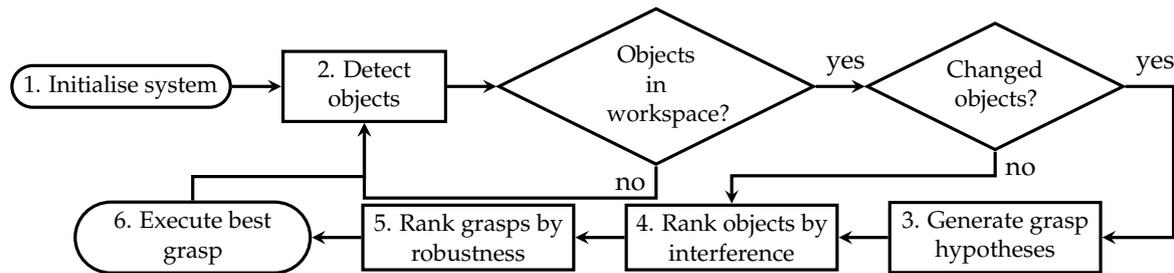


Figure 2. Description of the algorithm.

4.1. Detect Objects

Standard algorithms available in the OpenCV (Region Of Interest, watershed, etc.) and Point Cloud Library (PCL) are used to achieve the perception by processing both the colour and depth images received from the Kinect Camera.

- (a) Object extraction: The objects are distinguished by eliminating the regions that match the table top color and that lie on or below the height of the table top. In order to avoid the elimination of thin objects, the background colour is carefully chosen. A binary image is created after the elimination where the black colour represents the eliminated background and a white colour represents the objects. The image is then used to extract a list of contours of the objects. Each contour includes an ordered list of the outermost pixel values surrounding the corresponding objects. If there is a hole in an object such as in a roll of tape, another list of pixel values is generated for the hole. Each pixel is converted into Cartesian coordinates using PCL.
- (b) Object segmentation: The above method only works when the objects are scattered over the workspace. When the objects overlap, the algorithm perceives them as one single object. To distinguish between overlapping objects, a watershed algorithm is applied as an additional step in the object extraction, providing a segmentation (see Figure 3). Due to the dependency on grayscale images, the algorithm tends to result in over segmentation of objects composed of multiple colours. This does affect the success rate of the algorithm, as there is an increased risk of detecting small objects where none exists. On the other hand, under-segmentation brings the risk of conflating neighbouring objects. This often leads to failed grasps when they are attempted near the interface between the two objects that are detected as one. There appears to be a trade-off between over- and under-segmentation, and we found empirically that some over-segmentation gave the best results, although the vision system was still susceptible to some errors.

When the binary image obtained at the “Object extraction” stage of step 1 is completely black, this means that no object was found in the scene. In this case, steps 2–5 are bypassed and the algorithm returns directly to repeating step 1 by requesting a new RGB-D image from the 3D camera. This corresponds to the block “Objects in workspace?” in the algorithm schematic of Figure 2.

In order to avoid generating new grasp hypotheses for all the objects at every iteration of the algorithm, we determine which objects were absent or have moved since the previous iteration. This is

done by first subtracting the raw RGB image from a previous time step to the RGB image of the current time step. The non-zero pixels of the difference image yield the regions where changes occurred. We then identify all the objects detected at the current iteration that are partly or completely included in these regions. The grasp hypotheses of these objects are generated or regenerated; those of the other unchanged objects are taken from the previous iteration. This corresponds to the block “Changed objects?” in Figure 2.

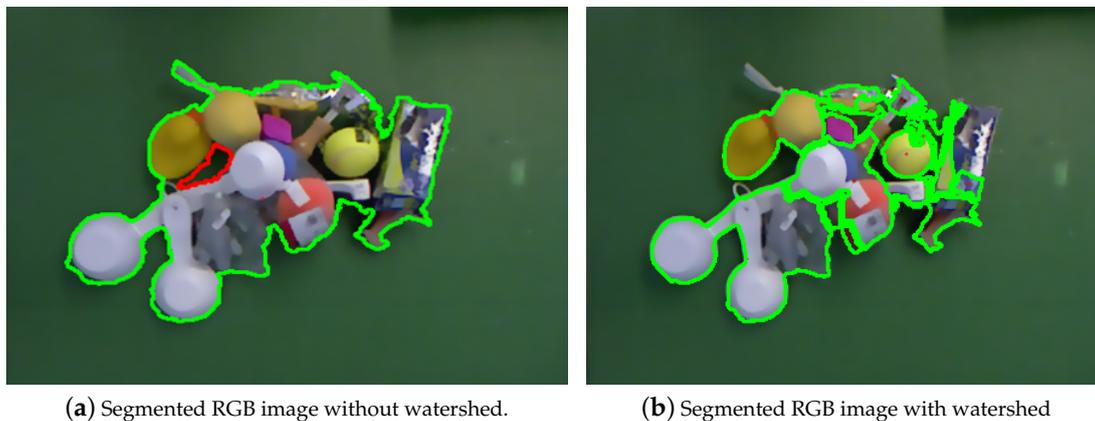


Figure 3. Comparison of the vision of a pile of objects with and without watershed.

4.2. Grasp Hypotheses

To generate grasp hypotheses, a parallel-line pattern is overlaid on the object contours. The pairs of intersection points between a line of the pattern and the contours of the same object are saved as potential positions for the two gripper fingers. This process is repeated for several directions of the parallel-line pattern, with line angles ranging from 0° to 180° . As a result, we obtain a list of potential grasping poses represented by the locations of the gripper fingers in the horizontal plane. The number of potential grasps can be adjusted by changing the spacing between lines in the parallel-line pattern and the angle increment between two consecutive parallel-line-pattern directions.

4.3. Interference Ranking

The list of potential grasps is first filtered by removing all pairs of finger positions separated by a distance larger than 85 mm, which corresponds to the maximum gripper stroke. Another filter is applied to remove any grasp where a finger would collide with a neighbouring object.

In order to determine the grasp that is least susceptible to a collision with the neighbouring objects, we sort those still in the list of possible grasps according to a global interference score. This score is computed in three steps, which are schematised in Figure 4 and detailed below:

1. **Interference matrix** : A symmetric matrix is defined to describe the potential interferences between objects. In this matrix, each row (and column) corresponds to one of the objects detected from the scene. Thus, entry (i, j) of the interference matrix is a value representing the likelihood of avoiding a collision with object j when attempting to grasp object i .
The entry is one whenever the two objects are separated by less than the width (in the XY plane) and the depth (in the vertical Z direction) of the gripper, and zero otherwise. The Python package Shapely is used to offset the XY -plane contour of a given object by one gripper width and to determine if there are intersections with the other objects. This operation is done for all objects.
2. **Interference score** : For all grasping hypotheses for a given object, the minimal distance between the fingers and the surrounding other objects is calculated. The smallest distance is retained for the score. If the finger is free from interference, the score is the positive distance whereas if the finger has an interference with an object, the score is the negative of the distance of penetration of one into the other.

- Global score: The interference score of each object is averaged and multiplied by the number of interfering objects (the interference matrix). For example, an object with five grasping hypotheses surrounded by three objects will have the interference score averaged on the five grasping hypotheses and multiplied by three, the number of interfering objects. The objective of this calculation is to maximize the global score of the objects that can provide more space.

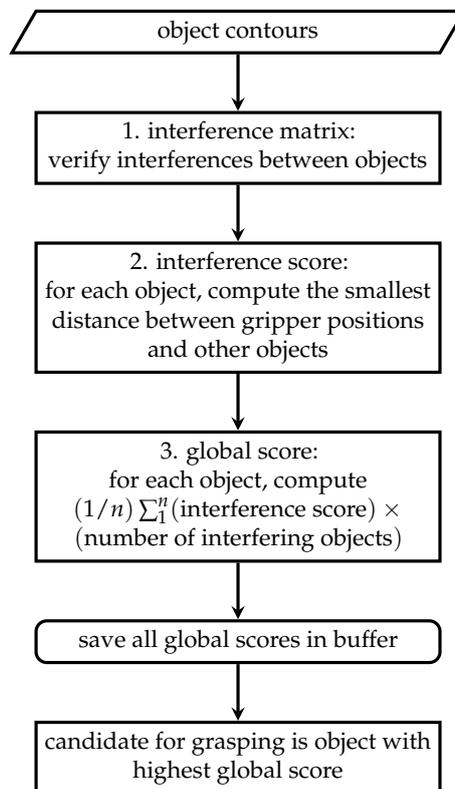


Figure 4. Description of the sorting function to choose the best objects to grasp from the pile.

Thus, an object with a negative global score means that there exists a major interference for a grasp, i.e., one finger has a significant interference with an object, which significantly influences the average of the interference score.

A positive global score means that the interferences between objects for a grasp are minor or null. The objects are ordered by global score (in decreasing order) and saved in a buffer. The object with the highest global score is considered as the best candidate for grasping from the standpoint of collisions.

4.4. Grasp Robustness Ranking

It is assumed here that no information is known a priori on the objects to be grasped. Therefore, the grasp robustness index must be determined from the description of the objects provided by a 3D vision sensor alone. In order to estimate the robustness of grasps, the index of the robustness of static equilibrium proposed in [10] is adapted here to the task of grasping objects with a two-finger gripper.

The index requires some modifications, as Guay et al. [10] assumed the directions of the contact forces to be known, that is, that there is no friction. In our case, we are using pinch grasps to pick up the objects, which cannot work unless there is sufficient friction. Therefore, we cannot assume that the contact force between finger and object is normal to the object surface. Instead, we rely on Coulomb's friction model according to which the contact force must belong to a friction cone. What we propose below is an extension of the concepts proposed in Reference [10] to objects with friction.

Consider a polygonal approximation of the top view of an object shown in Figure 5. Such an approximation can be extracted from an RGB or an RGB-D image. Assume that the gripper has m fingers making contact with the object at points $P_i, i = 1, \dots, m$. We assume soft-finger contacts, so that the fingers apply three-dimensional reaction forces \mathbf{f}_i and moments $\mathbf{n}_i, i = 1, \dots, m$, on the object. Assume that the normal to the object contour at P_i is given by unit vector $\mathbf{u}_{n,i}$, which is directed from the contact point outwards.

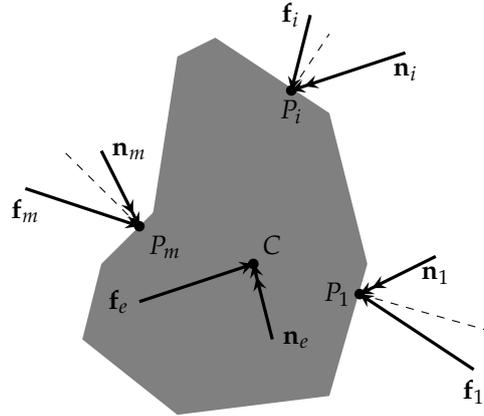


Figure 5. Top view of an arbitrary object to be grasped.

Then the friction cones of the Coulomb model may be written as

$$\|(\mathbf{1}_{3 \times 3} - \mathbf{u}_i \mathbf{u}_i^T) \mathbf{f}_i\|_2 \leq -\mu_{f,s} \mathbf{u}_i^T \mathbf{f}_i, \tag{1a}$$

$$\|\mathbf{n}_i\|_2 \leq -\mu_{n,s} \mathbf{u}_i^T \mathbf{f}_i, \tag{1b}$$

where $\|\cdot\|_2$ is the Euclidean norm of its vector argument, while $\mu_{f,s}$ and $\mu_{n,s}$ are the static friction coefficients for the forces and moments, respectively. To this, we add that the force normal to the object boundary cannot be directed away from the boundary, i.e.,

$$\mathbf{u}_i^T \mathbf{f}_i \leq 0. \tag{1c}$$

This force is also bounded from above, due to the limited grasping force of the gripper,

$$-\mathbf{u}_i^T \mathbf{f}_i \leq \bar{f}_i. \tag{1d}$$

Together, Equations (1) form the set of available contact forces, and we label it \mathcal{C} for future reference.

The contact forces must be able to resist a set of external forces and moments applied on the object. These forces and moments are represented by their resultant force \mathbf{f}_e and moment \mathbf{n}_e at the centroid C of the object. Note that we use the centroid and not the centre of mass because the camera is our only source of information, and it only provides geometric data on the object. The set of possible resultant forces and moments is assumed to be a six-dimensional polytope \mathcal{E} of the form

$$\mathcal{E} = \{(\mathbf{f}_e, \mathbf{n}_e) \in \mathbb{R}^6 : \mathbf{A}_{f,e} \mathbf{f}_e + \mathbf{A}_{n,e} \mathbf{n}_e \leq \mathbf{a}_e\}. \tag{2}$$

This set is determined before the grasp, and should include the weight of the object as one of its points. Because of uncertainties on the value of the weight—we only have information on the object geometry—and on the location of its point of application, it seems preferable to account for ranges of possible values of \mathbf{f}_e and \mathbf{n}_e , that is, to consider \mathcal{E} as a six-dimensional volumetric object in \mathbb{R}^6 rather than a single point.

The contact forces and moments and the external forces and moments are related by the static equilibrium equations derived from the free-body diagram of the object. Upon defining the vectors $\mathbf{r}_i \equiv \overrightarrow{CP_i}$, we may express these equations as:

$$\sum_{i=1}^m \mathbf{f}_i + \mathbf{f}_e = \mathbf{0}_3, \tag{3a}$$

$$\frac{1}{r_{\text{rms}}} \left(\sum_{i=1}^m \mathbf{r}_i \times \mathbf{f}_i + \sum_{i=1}^m \mathbf{n}_i + \mathbf{n}_e \right) = \mathbf{0}_3, \tag{3b}$$

where $r_{\text{rms}} \equiv \sqrt{\sum_i^m \|\mathbf{r}_i\|_2^2}$ is used to render Equation (3) dimensionally homogeneous.

In order to clarify the meaning of Equations (1)–(3), consider a simple case where there are only two contact points ($m = 2$) and where the object is considered as a particle in equilibrium in the plane. Furthermore, assume that there is no friction at contact point P_2 , so that the contact forces are fully described by three components: the normal force $f_{n,1}$ at P_1 , the friction force $f_{f,1}$ at P_1 and the normal force $f_{n,2}$ at P_2 . Physically, this amounts to grasping a small object by pinching two surfaces: one rough and the other slippery.

The set of available contact forces \mathcal{C} is represented mathematically by Equation (1) and geometrically as shown in Figure 6a. This set constitutes the main conceptual difference between the theoretical concept presented in ref. [10] and its application to grasping presented here. Without friction, \mathcal{C} is a m -dimensional box (a.k.a. orthotope) whereas it is a convex set bounded by planes and conical surfaces in up to 6 m dimensions when friction is included. The higher number of dimensions makes the index computation more intensive, but also the nature of the set involved. Computing the equilibrium robustness index requires mapping \mathcal{C} from the space of contact forces to the space of external forces and moments through Equation (3), which leads to the polygon \mathcal{C}' . This process is much faster when \mathcal{C} is a box, thanks to the hyperplane shifting method [31]. In the current case, we approximate the cones forming \mathcal{C} with inscribed pyramids, so as to under-estimate the set of possible contact forces that can be applied on the object. We compute all the vertices of this polytope approximation of \mathcal{C} and we map them onto the space of external forces and moments through Equation (3). We obtain a set of mapped vertices, to which we apply a standard convex hull procedure. The resulting polytope is used in lieu of \mathcal{C}' , although it is in reality an under-estimator of \mathcal{C}' .

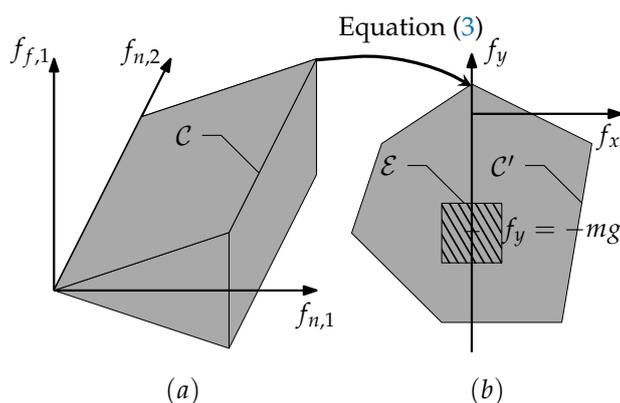


Figure 6. Grasping model representation in (a) the space of contact forces and (b) the space of external forces and moments.

The rest of the algorithm is conceptually the same as that presented in Reference [10]. The set \mathcal{E} of external wrenches can be defined as a box centred at the point $(f_x, f_y) = (0, -mg)$. The grasp is feasible if and only if the set \mathcal{E} of possible external wrenches is included in the set \mathcal{C}' of mapped available contact forces, that is, iff $\mathcal{E} \subset \mathcal{C}'$.

Set inclusion is a binary indicator which does not allow a detailed assessment or ranking of the various grasp hypotheses that are typically explored in autonomous pick-and-place. Hence, the robustness of equilibrium index is defined in Figure 7 as the margin s between the boundary of the set \mathcal{C}' and set \mathcal{E} . Physically, s may be regarded as the minimum unexpected force perturbation that could disrupt the static equilibrium of the object. By “unexpected”, we intend those force variations that are not already accounted for in the set \mathcal{E} of external forces and moments. Hence, the larger the value of s in newtons, the more robust the equilibrium.

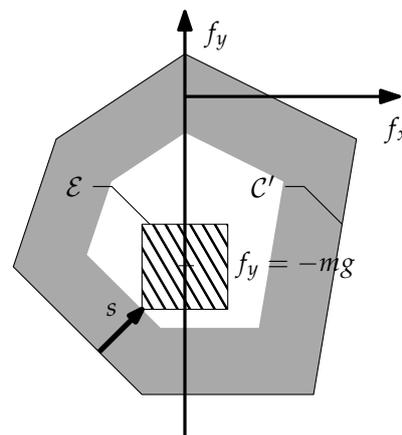


Figure 7. Definition of the index of equilibrium robustness.

In the case where some external forces or moments of \mathcal{E} can be found to disrupt static equilibrium, then \mathcal{E} is partly outside of \mathcal{C}' . In such a case, the margin s is that between the boundary of \mathcal{C}' and the point of \mathcal{E} that is farthest from \mathcal{C}' , and is taken to be negative. Hence, a negative s implies that the object is not in equilibrium for all expected external forces and moments and the smaller the value of s in newtons, the larger the distance to robust equilibrium.

Note that this extension is akin to the grasp quality indices used for instance by Pollard [32] and Borst et al. [33] in that it assesses the inclusion of a task wrench set into a grasp wrench set. The grasp quality index proposed here is different, however, from those indices and others because it measures the degree of inclusion directly in newtons, as the signed distance from the task wrench set to the boundary of the grasp wrench set. In the articles of Pollard [32] and Borst et al. [33], the grasp quality was defined as the scaling factor multiplying the task wrench set so that it fits exactly within the grasp wrench set. In contrast, the index of equilibrium robustness proposed here has a clear definite meaning—it represents the minimum perturbation force, in newtons, required to destabilise the object.

Other grasp quality indices based on the statics of grasping are available. For instance, we should cite the *potential contact robustness* and the *potential grasp robustness* proposed by Pozzi et al. [34]. These indices differ from the equilibrium robustness in that they evaluate the contact forces rather than their resultant. The underlying geometric idea is similar, however, in that they rely on the shortest distance from any of the contact forces to the boundary of its friction cone, a process that resembles the computation of the distance from the set of external wrenches \mathcal{E} to the boundary of the set of mapped available contact forces \mathcal{C}' . We should also cite the *net force* index used by Bonilla et al. [35]. It consists in measuring the amount of internal force within the grasp beyond the weight of the object, with the objective of minimising this value. This index provides a simple means of comparing different grasps of the same object, and its physical interpretation is quite intuitive to the user. For this reason, we regard this index as generally sound and useful. We should note, however, that sometimes the minimum net force is not the most robust to external perturbations. Indeed, when we feel a risk of dropping an object, we naturally tighten our grasp to increase its robustness, which seems to go against the indication of the net force index.

In order to illustrate the type of grasp ranking that can be obtained using the proposed criterion, consider the example of a dust pan shown in Figure 8a. The contour of this object was obtained from an RGB image. Several grasp hypotheses were generated over this contour for a two-finger gripper with friction attempting a pinching grasp. Hence, each grasp is represented in Figure 8 by a coloured line segment, its colour indicating the corresponding value of the robustness equilibrium index s .

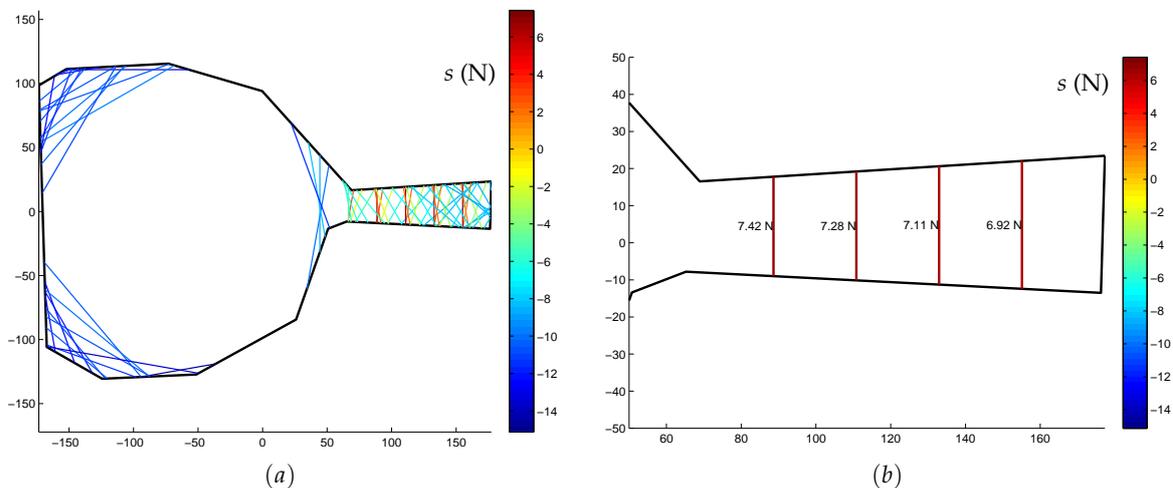


Figure 8. Grasp hypotheses for a dust pan: (a) interference-free grasps and (b) statically robust grasps with $s \geq 5$ N (units: mm).

In this example, the maximum grasping force is $\bar{f}_i = 200$ N, the friction coefficients are $\mu_{f,s} = 0.6$ and $\mu_{n,s} = 0.12$ m and the friction cones are approximated with inscribed pyramids with square bases, that is, with five vertices, including the apex. The set \mathcal{E} of external wrenches is defined as the box $\mathcal{E} = \{(\mathbf{f}_e, \mathbf{n}_e) \in \mathbb{R}^6 : (-10 \text{ N})\mathbf{1}_3 \leq \mathbf{f}_e \leq (10 \text{ N})\mathbf{1}_3, (-0.25 \text{ N m})\mathbf{1}_3 \leq \mathbf{n}_e \leq (0.25 \text{ N m})\mathbf{1}_3\}$, where $\mathbf{1}_3 \equiv [1 \ 1 \ 1]^T$. The bounds on the external moment \mathbf{n}_e correspond to errors of ± 25 mm on the position of the centre of mass of the dustpan, assuming it weighs 10 N.

Looking closer at the results of Figure 8a, we see that the algorithm naturally excludes the grasps over contours segments that are insufficiently parallel. As a result, only the grasp hypotheses made on the handle are retained. Among those, the grasps that are perpendicular to the handle are determined the most robust. Among those, from Figure 8b, the grasp that is most robust is that closest to the centroid of the dustpan, which also corresponds to what intuition would suggest. The value of the robustness index is $s = 7.42$ N for this grasp, which means that there would need to be an unexpected perturbation of that much force for the grasp to fail.

Having introduced the concepts behind the proposed grasping algorithm, we now turn our attention to the assessment of their performance.

5. Experimental Tests and Results

The performance assessment is accomplished with the testbed as described in Section 3. We now present the experimental protocol followed, the experimental results obtained and a discussion of these results.

5.1. Protocol

A pile of objects is randomly created. To do so, the objects are placed randomly in a small box which is flipped on the table, which results in piles similar to those shown in Figure 9a–c. The set is composed of eleven different objects, which are presented in Figure 10. Four of these objects come from the set of the Amazon Picking Challenge: the bath duck, the squeakair ball, the squeaking eggs and the whiteboard eraser. The other objects are mostly chosen for the diversity of their shapes and textures.

The objects are unknown to the algorithm, and no machine learning is used to train the system. Each experimental test is independent from the others. Previous tests have shown that the robot is capable of grasping the objects individually. During a grasping action, different situations may occur, namely:

- No object is grasped by the gripper;
- Only one object is grasped by the gripper;
- Two or more objects are grasped by the gripper.

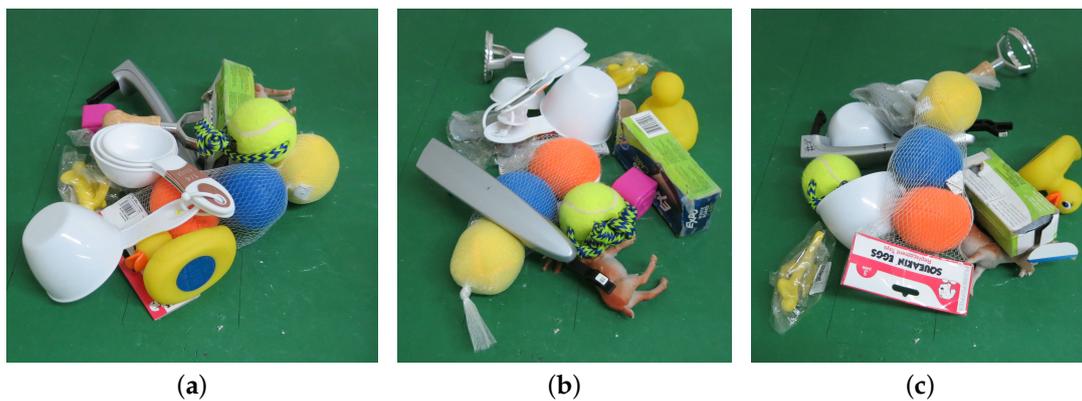


Figure 9. Examples of random piles of objects.

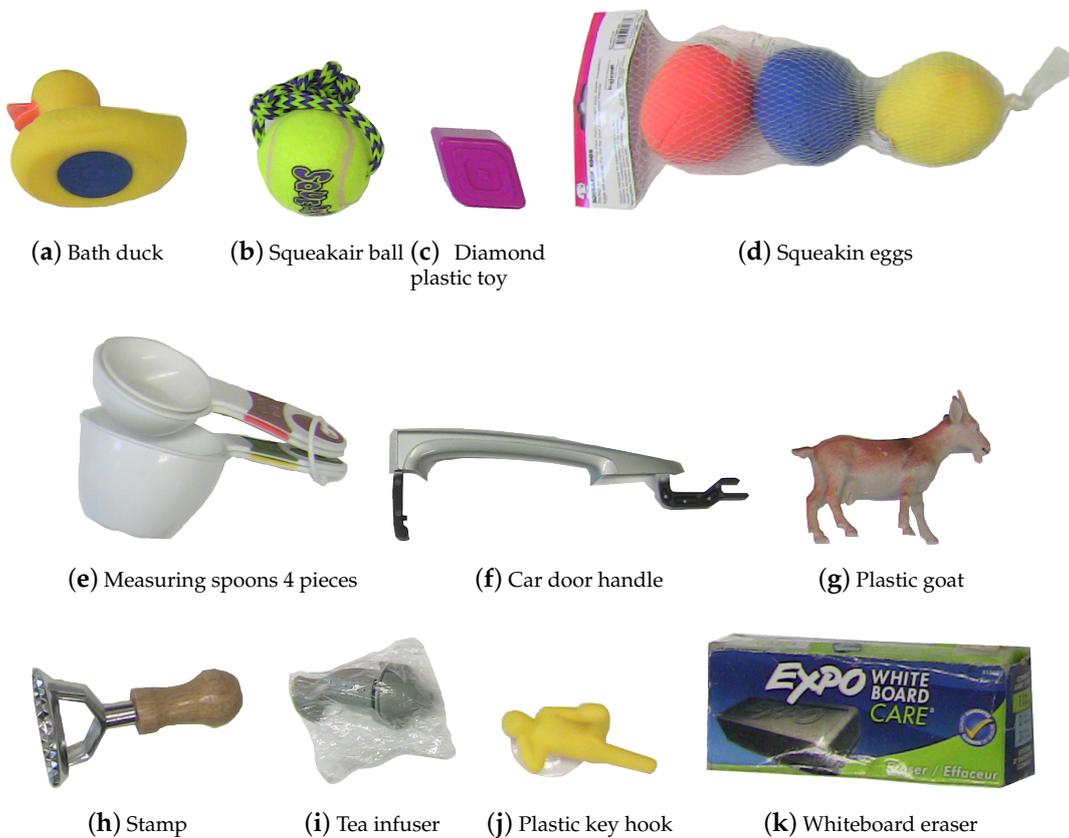


Figure 10. Objects used in the experiments.

A grasp is considered successful when one and only one object is picked up. The experimental tests continue until the robot grasps all the objects available on the table or until it can no longer find a feasible grasp. In this case, the experimental test is stopped.

Fifty trials are performed, an example of which is shown in Figure 11. For each test, a new random pile of objects is created, and the algorithm is run until exhaustion of the pile or the absence of a feasible grasp.

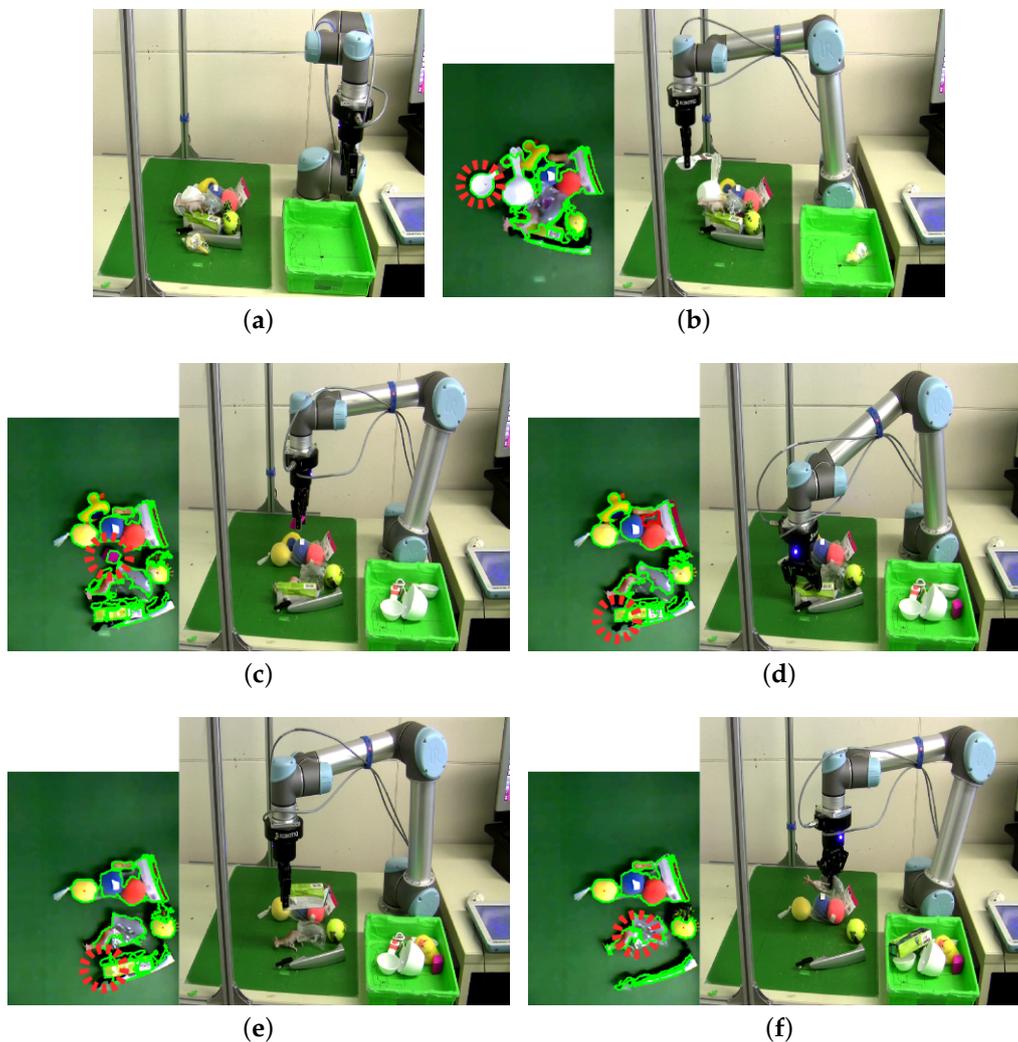


Figure 11. Object manipulation during an experimental test. The photographs come from the recorded video (right-hand side) and of the kinect (left-hand side, except for (a)). The red dotted circles on the kinect images represent the targeted grasp. Figure (a) represents the initial pile. In Figures (b–d) the grasping of the spoons, the plastic toy and the eraser are shown. Figure (e) shows a failed grasp and (f) shows a double grasp. Videos of four of the 50 trials are available at <https://youtu.be/IMjTGGOASvU>.

5.2. Results

Figure 12 shows the percentage of successful grasps in the first attempt, in other words, the number of successful grasps divided by the total number of grasp attempts plus the number of remaining objects for each test. A grasp attempt is considered unsuccessful whenever the robot fails to pick up any object or it picks up more than one object at once. The best results are equal to 100% (succeeded one time) and 91.7% (succeeded five times). 91.7% is equivalent to 11 successful grasps for one failure, so a total of 12 trials for 11 objects.

Out of the 50 experimental trials, two have a very low percentage of correct grasps, respectively at success rates of 12.5% and 45%. These trials are represented by red crosses in Figure 12. In the former case, the robot only grasped two objects and experienced five failures before being unable to find

a feasible grasp and stopping. This is due to the configuration of the pile of objects. In the latter case, the watershed created a false object on the wrapping of the *squeakin* eggs. The algorithm attempted six grasps unsuccessfully on this false object before moving to a real object. In the end, it succeeded in grasping all the objects in the pile.

The overall percentage of grasp success is 75.2%. The boxplot shows that three out of four trials have a percentage of success (i.e., first grasp successful) higher than 68.7%.

Results show that in the majority of trials, the robot was able to completely clear the pile (see Table 1), namely in 39 trials out of 50, for a percentage of 78%. In 48 trials (in 96% of the trials), at least nine objects were cleared from the pile. The average number of attempts required to remove all objects from the pile is 13.3, or an average of 1.21 attempts per object, as shown in Figure 13.

The failed attempted grasps were observed to be due to two main reasons: (i) the vision system creates a ghost object because of over-segmentation and (ii) the vision system creates an inaccurate estimate of the object contour. In the latter case, inaccuracies were often observed for objects located near the border of the workspace. This led to parallax errors where the objects were assumed to have larger footprints than in reality.

Table 1. Distribution of objects removed from the pile (maximum of 11 objects).

Nbr of Tests	Nbr of Grasped Objects	%
1	2	18.2
1	8	72.7
2	9	81.8
7	10	90.9
39	11	100
total: 50 tests		

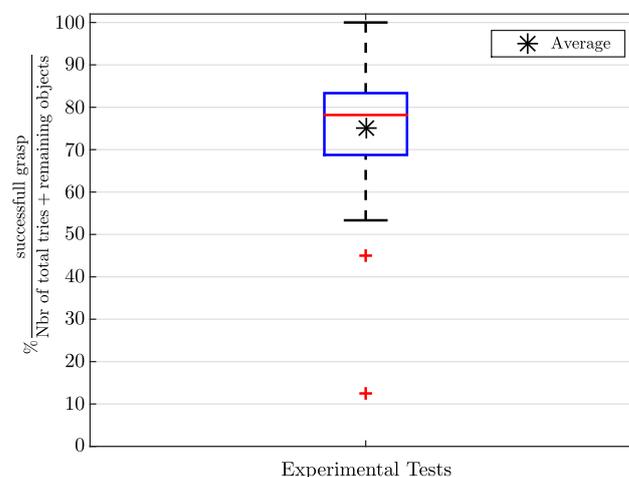


Figure 12. Boxplot of the percentage of successful object grasps. The average percentage of success is indicated by an asterisk at 75.2%. The interquartile range (IQR), that is, the range between the 25 and 75 percentiles of data, is represented by the box. The data range is represented by the line segments. The outliers are denoted by crosses. They are defined as those points farther than 1.5IQR from the interquartile range.

The failed trials where the pile was not completely cleared are attributable to two main reasons. (i) The algorithm did not find a correct grasp due to the particular configuration of the remaining pile of objects. (ii) An object neighbouring the one targeted for grasping was inadvertently pushed out of the workspace by the gripper. This led to a premature stop of the trial.

Table 2 presents the distribution of failed grasp attempts (no object grasped and two objects grasped). 35 out of 50 trials were free of double grasps, but 42 of the tests included between one and five attempts where no object was grasped.

Table 2. Distribution of failure for 50 experimental tests.

	Nbr of Experimental Tests	
	Double Grasping	Failed Grasping
Nbr of occurrences	0	5
	1	13
	2	7
	3	8
	4	8
	5	6
	6	1
	7	1
	10	1

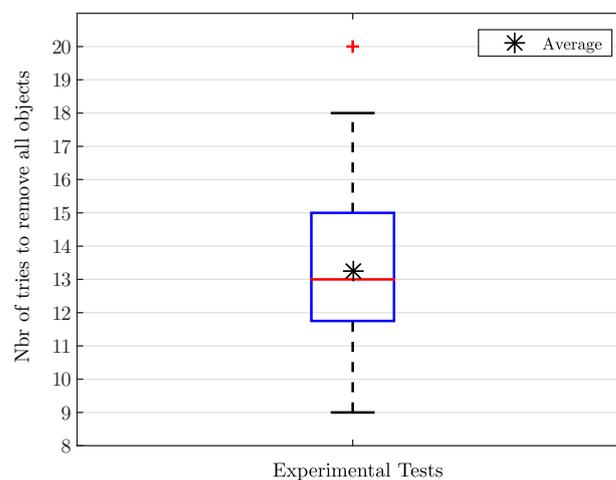


Figure 13. Boxplot of number of trials to remove completely the pile (11 objects). The average of the number of trials equals 13.2. Values are for 39 tests (see Table 1).

5.3. Discussion

The results of Table 1 and Figure 12 show that the proposed simple framework, composed of a basic vision algorithm, a sorting function and a grasp robustness index, is applicable to the grasping of unknown objects, one by one, from a pile. This simple approach allows to completely clear a pile in a large proportion of the trials (see Table 1), with a number of grasp attempts close to the number of objects: between 9 and 15 trials for 75% of the trials (see Figure 13). With 1.21 attempts per object, this framework is in the range of the number of actions per object found in the available literature (1.04 [2,30], 1.10 [27], 1.11 [29], 1.13 [23], 1.17 [22], 1.31 [20], 2.0 [28], 2.6 [21], $2n$ [25]).

An important advantage of the proposed method is its simplicity, as it relies on a small number of relatively simple and robust ideas. This makes it easy to adjust its parameters and predict the ensuing behaviour of the robot. The method also entails some drawbacks, however, the most important being perhaps its ability to identify objects. First, the over-segmentation that often results from the watershed method can create more objects than there are in reality. In this case, the robot sometimes tries to grasp an object which does not actually exist (a.k.a. a ghost object). Second, the texture and the colour of the object can create reflections which leave it unseen by the camera. As a result, true negatives and false positives can both occur. A solution could be to use two cameras to improve the robustness of the

vision system. We could also test methods of segmentation that are more sophisticated to avoid the appearance of ghost objects in the filtered images.

In our method, a failed grasp generally results in the displacement of the object that was chosen for pick up, and sometimes of other neighbouring objects. Therefore, a failed grasp is followed by another attempt generally drawn from a changed list of grasp candidates. This process is an interaction with the pile just like the pushing or pulling actions proposed in Reference [26]. If we compare qualitatively the two approaches, the pushing and pulling actions can often generate more variation in the pile, which could lead to better grasp success rates afterwards. Such perturbations probably add to the robustness of the method, sometimes allowing it to bootstrap out of difficult situations where no good grasps are found. Such an action would probably have helped in the failed trial with a success rate of 12.5% (see the lower red cross in Figure 12), where the algorithm stopped for want of a feasible grasp. On the other hand, starting with a grasp attempt presents the obvious advantage of some probability of success at the first action. Surely one could find cases where any of the two methods is more advantageous than the other; it would be interesting to determine which situations lend themselves better to the initial grasp attempt, and which are better suited for the pushing or pulling actions.

6. Conclusions

Grasping unknown objects individually from a pile is a complex task for a robot. Difficulties can arise from distinguishing between the different objects of the pile, ensuring a robust grasp and avoiding collisions. This paper demonstrates that a straightforward approach allows to successfully perform this complex task by using an elementary vision algorithm for segmentation, a sound grasp robustness index and a simple decision function. Thanks to the basic vision system, a 3D representation of the pile is created. An over-segmentation algorithm allows to divide this pile model into object contours. A simple interference detection scheme filters out those grasp hypotheses that are inaccessible. A grasp robustness index allows to rank the remaining grasp hypotheses. Thus, the algorithm is able to grasp individually an object with a promising success rate of 75.2% and to clear entire piles of eleven objects in 78% of cases. Analysis of the failed grasps shows that the main limitations come from the vision system and the difficulty to correctly discriminate the objects. Nevertheless, with a rate of 1.21 actions per object, this approach is comparable to those that have been proposed before it. These other methods are generally more complex, however, and, when based on machine learning, they often require a lengthy training period. In contrast, our method is simple and derived from rational, physics-based concepts. This makes it easier to pinpoint potential problems when the algorithm is confronted with different conditions (objects, gripper, 3D camera, etc.). Moreover, the method relies on a limited number of parameters such as a threshold for the watershed filter, the finger dimensions, object density, maximum grasping force, force and moment friction coefficients, and uncertainty on the centre of mass location. The values of these parameters can be chosen based on physical intuition and are therefore quite easy to adjust. In summary, despite its shortcomings, the authors believe that the simplicity and conceptual soundness of the proposed method make it competitive for many an industrial application.

Future work includes refining the 3D sensing system and its underlying segmentation method. It is also envisioned to extend this segmentation of the pile to three dimensions to allow for side grasps. With this additional capability, we hope to improve the success rates of individual grasps and of complete pile removals.

Author Contributions: The proposed grasping algorithm was devised iteratively and incrementally by all the authors. F.L., B.S. and S.P. assembled the experimental testbed, programmed the robot and performed the experiments. B.S. wrote most of the article, while P.C. and C.G. reviewed its contents. In whole, all authors contributed equally to this article.

Funding: This research was funded by the Collaborative Research and Development grant #CRDPJ-461709-13 from the Natural Sciences and Engineering Research Council of Canada (NSERC).

Acknowledgments: The authors would like to acknowledge the valued help provided by Simon Foucault in setting up the experiments.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ikeuchi, K.; Horn, B.K.P.; Nagata, S.; Callahan, T.; Feingold, O. Picking up an object from a pile of objects. In *Robotics Research: The First International Symposium*; MIT Press: Cambridge, MA, USA, 1984; pp. 139–162.
2. Buchholz, D. *Bin-Picking: New Approaches for a Classical Problem*; *Studies in Systems, Decision and Control*; Springer: Heidelberg, Germany, 2015; Volume 44.
3. Eppner, C.; Deimel, R.; Álvarez Ruiz, J.; Maertens, M.; Brock, O. Exploitation of Environmental Constraints in Human and Robotics Grasping. *Int. J. Robot. Res.* **2015**, *34*, 1021–1038. [[CrossRef](#)]
4. Bohg, J.; Morales, A.; Asfour, T.; Kragic, D. Data-driven grasp synthesis—A survey. *IEEE Trans. Robot.* **2014**, *30*, 289–309. [[CrossRef](#)]
5. Roa, M.A.; Suárez, R. Grasp quality measures: review and performance. *Auton. Robot.* **2015**, *38*, 65–88. [[CrossRef](#)]
6. Sahbani, A.; El-Khoury, S.; Bidaud, P. An overview of 3D object grasp synthesis algorithms. *Robot. Auton. Syst.* **2012**, *60*, 326–336. [[CrossRef](#)]
7. Aleotti, J.; Caselli, S. Interactive teaching of task-oriented robot grasps. *Robot. Auton. Syst.* **2010**, *58*, 539–550. [[CrossRef](#)]
8. Pedro, L.M.; Belini, V.L.; Caurin, G.A. Learning how to grasp based on neural network retraining. *Adv. Robot.* **2013**, *27*, 785–797. [[CrossRef](#)]
9. Zheng, Y. An efficient algorithm for a grasp quality measure. *IEEE Trans. Robot.* **2013**, *29*, 579–585. [[CrossRef](#)]
10. Guay, F.; Cardou, P.; Cruz-Ruiz, A.L.; Caro, S. Measuring how well a structure supports varying external wrenches. In *New Advances in Mechanisms, Transmissions and Applications*; Springer: Berlin/Heidelberg, Germany, 2014; pp. 385–392.
11. Roa, M.A.; Argus, M.J.; Leidner, D.; Borst, C.; Hirzinger, G. Power grasp planning for anthropomorphic robot hands. In Proceedings of the 2012 IEEE International Conference on Robotics and Automation, Saint Paul, MN, USA, 14–18 May 2012; pp. 563–569.
12. Nieuwenhuisen, M.; Droeschel, D.; Holz, D.; Stücker, J.; Berner, A.; Li, J.; Klein, R.; Behnke, S. Mobile bin picking with an anthropomorphic service robot. In Proceedings of the 2013 IEEE International Conference on Robotics and Automation, Karlsruhe, Germany, 6–10 May 2013; pp. 2327–2334.
13. El Khoury, S.; Sahbani, A. Handling Objects by Their Handles. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'08), Nice, France, 22–26 September 2008; pp. 58–64.
14. Kehoe, B.; Berenson, D.; Goldberg, K. Toward cloud-based grasping with uncertainty in shape: Estimating lower bounds on achieving force closure with zero-slip push grasps. In Proceedings of the 2012 IEEE International Conference on Robotics and Automation, Saint Paul, MN, USA, 14–18 May 2012; pp. 576–583.
15. Fischinger, D.; Weiss, A.; Vincze, M. Learning grasps with topographic features. *Int. J. Robot. Res.* **2015**, *34*, 1167–1194. [[CrossRef](#)]
16. Eppner, C.; Höfer, S.; Jonschkowski, R.; Martín-Martín, R.; Sieverling, A.; Wall, V.; Brock, O. Four Aspects of Building Robotic Systems: Lessons from the Amazon Picking Challenge 2015. In Proceedings of the Robotics: Science and Systems XII, Ann Arbor, MI, USA, 18–22 June 2016.
17. Saxena, A.; Wong, L.L.S.; Ng, A.Y. Learning Grasp Strategies with Partial Shape Information. In Proceedings of the 23rd National Conference on Artificial Intelligence (AAAI'08), Chicago, IL, USA, 13–17 July 2008; Volume 3, pp. 1491–1494.
18. Saxena, A.; Driemeyer, J.; Ng, A.Y. Robotic grasping of novel objects using vision. *Int. J. Robot. Res.* **2008**, *27*, 157–173. [[CrossRef](#)]
19. Katz, D.; Venkatraman, A.; Kazemi, M.; Bagnell, J.A.; Stentz, A. Perceiving, learning, and exploiting object affordances for autonomous pile manipulation. *Auton. Robot.* **2014**, *37*, 369–382. [[CrossRef](#)]
20. Fischinger, D.; Vincze, M. Empty the basket—a shape based learning approach for grasping piles of unknown objects. In Proceedings of the 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems, Vilamoura, Portugal, 7–12 October 2012; pp. 2051–2057.

21. Pinto, L.; Gupta, A. Supersizing self-supervision: Learning to grasp from 50k tries and 700 robot hours. In Proceedings of the 2016 IEEE International Conference on Robotics and Automation (ICRA), Stockholm, Sweden, 16–21 May 2016; pp. 3406–3413.
22. Levine, S.; Pastor, P.; Krizhevsky, A.; Ibarz, J.; Quillen, D. Learning hand-eye coordination for robotic grasping with deep learning and large-scale data collection. *Int. J. Robot. Res.* **2016**, *37*, 421–436. [[CrossRef](#)]
23. ten Pas, A.; Platt, R. Using geometry to detect grasp poses in 3d point clouds. In *Robotics Research*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 307–324.
24. Van Hoof, H.; Kroemer, O.; Peters, J. Probabilistic segmentation and targeted exploration of objects in cluttered environments. *IEEE Trans. Robot.* **2014**, *30*, 1198–1209. [[CrossRef](#)]
25. Katz, D.; Kazemi, M.; Bagnell, J.A.; Stentz, A. Clearing a pile of unknown objects using interactive perception. In Proceedings of the 2013 IEEE International Conference on Robotics and Automation, Karlsruhe, Germany, 6–10 May 2013; pp. 154–161.
26. Chang, L.; Smith, J.R.; Fox, D. Interactive singulation of objects from a pile. In Proceedings of the 2012 IEEE International Conference on Robotics and Automation, Saint Paul, MN, USA, 14–18 May 2012; pp. 3875–3882.
27. Gupta, M.; Sukhatme, G.S. Using manipulation primitives for brick sorting in clutter. In Proceedings of the 2012 IEEE International Conference on Robotics and Automation, Saint Paul, MN, USA, 14–18 May 2012; pp. 3883–3889.
28. Hermans, T.; Rehg, J.M.; Bobick, A. Guided pushing for object singulation. In Proceedings of the 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems, Vilamoura, Portugal, 7–12 October 2012; pp. 4783–4790.
29. Eppner, C.; Brock, O. Grasping unknown objects by exploiting shape adaptability and environmental constraints. In Proceedings of the 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems, Tokyo, Japan, 3–7 November 2013; pp. 4000–4006.
30. Buchholz, D.; Kubus, D.; Weidauer, I.; Scholz, A.; Wahl, F.M. Combining visual and inertial features for efficient grasping and bin-picking. In Proceedings of the 2014 IEEE International Conference on Robotics and Automation (ICRA), Hong Kong, China, 31 May–7 June 2014; pp. 875–882.
31. Bouchard, S.; Gosselin, C.; Moore, B. On the Ability of a Cable-Driven Robot to Generate a Prescribed Set of Wrenches. *ASME J. Mech. Robot.* **2010**, *2*, 011010. [[CrossRef](#)]
32. Pollard, N.S. Parallel Algorithms for Synthesis of Whole-Hand Grasps. In Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), Albuquerque, NM, USA, 25–25 April 1997; pp. 373–378.
33. Borst, C.; Fischer, M.; Hirzinger, G. Grasp Planning: How to Choose a Suitable Task Wrench Space. In Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), New Orleans, LA, USA, 26 April–1 May 2004; pp. 319–325.
34. Pozzi, M.; Malvezzi, M.; Prattichizzo, D. On Grasp Quality Measures: Grasp Robustness and Contact Force Distribution in Underactuated and Compliant Robotic Hands. *IEEE Robot. Autom. Lett.* **2017**, *2*, 329–336. [[CrossRef](#)]
35. Bonilla, M.; Farnioli, E.; Piazza, C.; Catalano, M.; Grioli, G.; Garabini, M.; Gabiccini, M.; Bicchi, A. Grasping with Soft Hands. In Proceedings of the IEEE-RAS International Conference on Humanoid Robots, Madrid, Spain, 18–20 November 2014; pp. 581–587.

