Model-free and learning-free grasping by Local Contact Moment matching

Conference Paper · October 2018

DOI: 10.1109/IROS.2018.8594226

6 authors, including:

Maxime Adjigble

16 PUBLICATIONS **368** CITATIONS

SEE PROFILE

Naresh Marturi

Extreme Robotics Lab

57 PUBLICATIONS **746** CITATIONS

SEE PROFILE

Valerio Ortenzi

Max Planck Institute for Intelligent Systems

33 PUBLICATIONS **436** CITATIONS

SEE PROFILE

Vijaykumar Rajasekaran

United Kingdom Atomic Energy Authority, CCFE

21 PUBLICATIONS **277** CITATIONS

SEE PROFILE

Model-free and learning-free grasping by Local Contact Moment matching

Maxime Adjigble[1], Naresh Marturi[1], Valerio Ortenzi[2], Vijaykumar Rajasekaran[1], Peter Corke[2], and Rustam Stolkin[1]

***Abstract*—This paper addresses the problem of grasping arbitrarily shaped objects, observed as partial point-clouds, without requiring: models of the objects, physics parameters, training data, or other a-priori knowledge. A grasp metric is proposed based on Local Contact Moment (LoCoMo). LoCoMo combines zero-moment shift features, of both hand and object surface patches, to determine local similarity. This metric is then used to search for a set of feasible grasp poses with associated grasp likelihoods. LoCoMo overcomes some limitations of both classical grasp planners and learning-based approaches. Unlike force-closure analysis, LoCoMo does not require knowledge of physical parameters such as friction coefficients, and avoids assumptions about fingertip contacts, instead enabling robust contacts of large areas of hand and object surface. Unlike more recent learning-based approaches, LoCoMo does not require training data, and does not need any prototype grasp configurations to be taught by kinesthetic demonstration. We present results of real-robot experiments grasping 21 different objects, observed by a wrist-mounted depth camera. All objects are grasped successfully when presented to the robot individually. The robot also successfully clears cluttered heaps of objects by sequentially grasping and lifting objects until none remain.**

I. Introduction

Robots have been routinely and reliably grasping a vast variety of objects in manufacturing environments for several decades. This is based on simple pre-programmed actions, on exactly pre-defined objects, in highly structured environments. However, autonomous, vision-guided grasping, in unstructured environments, remains an open research problem. In this paper, we assume that the robot has a model of itself, but does not have any models or prior knowledge of the objects that it is tasked with grasping. These objects may take arbitrary shape and appear amidst clutter, observed as noisy partial point-clouds. Our main contribution is to show how this problem can be approached *without needing either classical physics analysis, or any learning from training data*.

Classical grasping methods based on physics analysis [1], [2] typically require the robot to have detailed knowledge of the grasped object's shape, mass and mass distribution, and friction coefficients between object surfaces and hand parts. It is common to assume point or fingertip contacts, with contacts of large surface areas of the hand becoming analytically intractable. More recent work has investigated a variety of machine learning approaches to grasping [3]–[5]. Learning approaches seek to encode a more direct link between the geometry of a scene (typically observed as a point-cloud) and grasp hypotheses. Such methods have significantly contributed to overcoming limitations of classical methods. However, all of these methods require training data (some more and some less). Most of these methods also require prototypical grasps (pinch-grasp, power-grasp, edge-grasp etc.) to be taught by kinesthetic demonstration or pre-progamming, albeit that learning-based methods can often adapt these pre-taught hand configurations to new object shapes (generalisation) with some success.

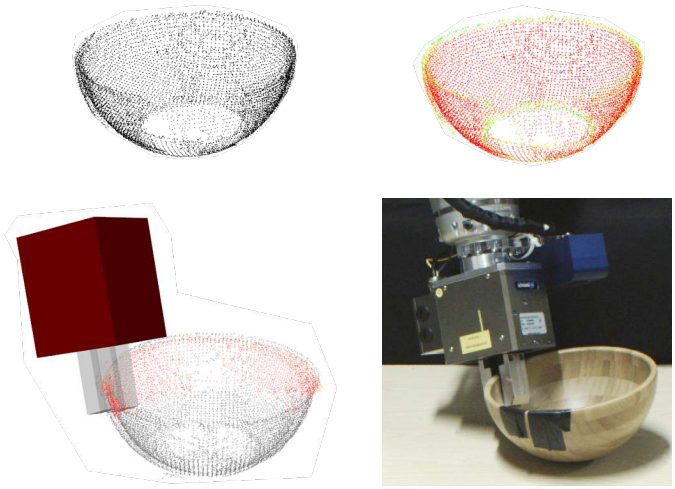

Figure 1. (Top-left) Point cloud of the object. (Top-right) Contact moment features for a single finger with planar surface. Red, yellow and green respectively encodes increasing values of the metric in this order computed using (3). (Bottom-left) Generated grasp with the highest contact probability. (Bottom-right) Grasp executed on the robot.

In this paper, we propose a novel algorithm for computing robust grasp hypotheses on arbitrarily shaped objects. The overall grasping pipeline is depicted in Fig. 1. Given a point-cloud view of a surface, and the kinematics of the robot's arm and hand, our algorithm outputs a variety of feasible grasp poses for the hand, and evaluates each according to a novel grasp likelihood metric. A collision-free reach-to-grasp trajectory is then sought, and the highest-likelihood reachable grasp is executed. Like recent learning-based methods, our method also maps observed surface shapes directly to grasp hypotheses. However, this mapping is not achieved by learning, does not require any training data, nor does it require any kinesthetic teaching or pre-programming of prototypical grasp

[1] M.Adjigble, N. Marturi, V. Rajasekaran and R. Stolkin are with the Extreme Robotics Laboratory, School of Metallurgy and Materials, University of Birmingham, UK. maxime.adjigble@gmail.com

[2] V. Ortenzi and P. Corke are with the ARC Centre of Excellence for Robotic Vision, Queensland University of Technology, Brisbane QLD 4001, Australia. http://www.roboticvision.org

configurations. Instead, we propose a novel grasp likelihood metric, the local contact moment probability function, which evaluates the shape compatibility between local parts of hand or finger surface, and local parts of an observed point-cloud.

Local contact moment (LoCoMo) is based on computing zero-moment shift features for local parts of the observed point cloud, and also parts of the robot's hand. First described in the computer graphics literature [6], zero moment shift features represent the characteristics of limited regions of surfaces, and are especially good at encoding information about surface curvature, Fig. 2, which is particularly important for matching hand parts to a grasped object. These features represent the surface characteristics of a limited region of the point cloud, hence they are "local" features. Also, they are computed on the point cloud without the need of any a-priori knowledge of the object (*i.e.*, model-free).

Using LoCoMo as a fitness function, a point-cloud surface can be efficiently searched for good matches to finger surface geometry. Kinematic analysis then yields a set of feasible grasps, with each grasp associated with a grasp likelihood. The motion-space of the arm is then explored to find collision-free reach-to-grasp trajectories to the highest likelihood grasp poses.

The main contributions of this work are:

- We propose the use of zero moment shift features [6] for robotic grasp-planning.
- We propose a new metric, the local contact moment probability function, for evaluating compatibility between the surface geometries of local parts of both object and gripper. This metric is model-free, and does not need to be learned from training data.
- Exploitation of the kinematics of the robot to select a subset of the graspable points, first identified by LoCoMo, that are kinematically reachable and feasible for the arm and hand system.

The remainder of this paper is structured as follows: Section II highlights the novelties of this work with respect to related literature. Section III describes the technical details of our proposed method. Section IV shows the results of a number of experiments conducted using a Schunk industrial two-finger hand mounted on a KUKA LBR iiwa manipulator arm. Section V provides concluding remarks.

II. Related Work

Classical approaches to grasping predominantly use physics-based analysis to compute force-closure [7]–[11]. Most of these approaches rely on a large amount of a-priori knowledge. They typically assume that an accurate and complete 3D model of the object is known, as well as its mass, mass distribution and also coefficients of friction between the object's surfaces and parts of the robot hand. In contrast, in many real applications, a robot may be required to grasp a previously unknown object of arbitrary shape, observed as a partial point-cloud view, for which friction coefficients and mass distribution are generally unknown. Many of these classical force-closure approaches are also restricted to assumptions of fingertip contacts only. Physics-based analysis becomes problematic when large patches of hand surface come into contact with the object (unlike many human grasps such as the "power grasp" where large surfaces of the hand are wrapped around the object).

More recent approaches have explored various forms of learning, [3], [12]–[15]. Learning-based methods overcome some of the limitations of classical methods, and have shown potential for generalising to grasping novel object shapes. [3] achieved moderately successful grasping, by learning a direct mapping between visual stimuli and motor outputs. Learning was achieved via robots making exploratory motions coupled with reinforcement. The system was able to synthesize novel grasping policies, but relied on enormous amounts of training data, involving large numbers of robots performing exploratory actions over a long period of time. [15] minimised the amount of reinforcement learning needed, by initiating learning from close-to-good grasp poses by kinesthetic demonstration using a data glove. In contrast, [13] showed significant ability to generalise grasping to novel objects, achieved by "one-shot" learning, *i.e.*, the robot was taught a single grasp on a single object, and was then able to plan successful grasps on new shapes. [13] learned "local" models of relationships between hand-parts and the curvatures of object surface patches. However, these must be combined with a "global" model of hand shape, corresponding to a grasp prototype (pinch grasp, power grasp, etc.) which is taught by demonstration. The method therefore remains unable to synthesize novel grasp prototypes that have not been taught.

Like the above learning approaches, our method also does not rely on object models or physics knowledge. Like [13] it exploits local descriptors of finger contacts (but a different kind). However, our method requires no training data, and can synthesize its own grasp hypotheses without any need of demonstration.

III. Method

We present a method to address robotic grasping based on the LoCoMo metric between the object and the gripper. This similarity metric between the features on the object and the features on the gripper is used to select viable finger poses on the surface of the object which are then combined with the kinematics of the gripper to form a grasp. In the following, we assume a model of the gripper, in this case a parallel jaw gripper.

The algorithm is given a (partial) point cloud of an object, and first computes the zero-moment shift features on the point cloud. The same features are extracted on the point cloud of the gripper model. These features of object and gripper are then used together to compute a local shape similarity metric between object and gripper. The main idea is to find the points that maximise the contact surface and to use only areas of the object that match the surface curvature of the fingers of the gripper for the grasp. Finally, a feasibility analysis is performed to select the subset of pairs of points which are

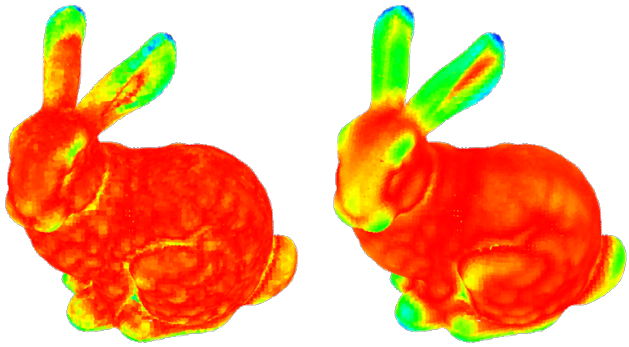

Figure 2. Local surface classification base on the zero moment shift of the Stanford Bunny. The colors Red, Yellow, Green and Blue encode in increasing order the magnitude of the L1 norm of the zero-moment shift vector. High values (Blue) incurs on the ears with high curvatures and low values (Red) on surfaces with low curvatures. Left: ρ = 0.008, right:ρ = 0.016.

returned from the previous action and which are kinematically feasible for the gripper.

A. Features Extraction and Matching Metric

Over the years, various local visual features have been presented in the literature and were previously used for tasks such as 2D/3D object recognition and pose estimation, [16]–[20]. In this work we propose the use of zero-moment shift features for grasping arbitrarily shaped objects.

Let $B_\rho(X)$ represent the Euclidean sphere of radius ρ centered at a point $X \in \mathbb{R}^3$. Given a set of points $\mathcal{X}$ in $\mathbb{R}^3$, the zero-moment shift n_ρ of the set of points $\xi = \mathcal{X} \cap B_\rho(X)$, belonging to the sphere $B_\rho(X)$, can be expressed as

$$n_\rho = M_\rho^0(\xi) - X \tag{1}$$

$$M_\rho^0(\xi) = \frac{1}{N}\sum_{n=1}^{N} X_i \tag{2}$$

where, $M_\rho^0(\xi)$ represents the zero moment (or centroid) of the set of points ξ belonging to the sphere $B_\rho(X)$. X_i is a point sampled from ξ and N the total number of points in ξ.

The L1 norm $|n_\rho|$ of the zero-moment shift is a good indicator of the characteristics of the underlying surface of the set of points, as shown in Fig. 2. It can be used in conjunction with a classifier to robustly distinguish smooth surfaces from edges, and also be used in conjunction with the first-moment of the set of points to provide a robust surface classification for noisy point cloud or mesh models as presented in [6]. In this work, we focus on the use of the zero-moment shift to compute a similarity metric between two arbitrary surfaces. We assume that the set of point is already preprocessed and filtered of outliers. Comparing two local surfaces is then reduced to comparing the zero-moment shift of the two surfaces. To this end, we introduce the LoCoMo probability function C_ϵ between two surfaces

$$C_\rho = 1 - \frac{\max(x, \phi(x, \vec{0}, \Sigma)) - \phi(\varepsilon; \vec{0}, \Sigma)}{\max(x, \phi(x, \vec{0}, \Sigma))} \tag{3}$$

ϕ represents the multivariate Gaussian density function

$$\phi(X, \mu, \Sigma) = \frac{1}{\sqrt{(2\pi)^n |\Sigma|}} exp(-\frac{1}{2}(X-\mu)\Sigma^{-1}(X-\mu)) \tag{4}$$

where $X, \mu \in \mathbb{R}^n$, Σ is the covariance matrix and n the space dimension. $\vec{0}$ is the null vector of $\mathbb{R}^3$, ε the error between the two zero-moment shift vectors defined as

$$\varepsilon = n_\rho^1 - n_\rho^2 \tag{5}$$

where n_ρ^1 and n_ρ^2 are expressed in the same reference frame. $\max(x, \phi(x, ...))$ is the maximum value of the function $\phi(x, ...)$ for all $x \in \mathbb{R}^3$. The zero-moment shift vectors can be projected on the axis of the normal and the axis orthogonal to the normal of the surface to obtain a new set of coordinates $(n_\parallel, n_\perp, 0)$ which can be used for the computation of (5).

This LoCoMo metric based on zero-moment shift features is extremely useful for grasping, as it provides a clear indication of the local contact between the surfaces of a gripper and an object.

B. Grasp Selection and Ranking

Selecting stable grasps is crucial to guarantee the success of a grasp. Several analytic methods use force closure, such as [21] and [22]. Force closure guarantees a static equilibrium between the contact forces. Furthermore, the interaction between two surfaces in contact can be reduced to one or multiple contact points as described in [23]. These assumptions are necessary conditions for a stable grasp selection, however they are not sufficient conditions for a stable grasp, as mentioned in [24].

The problem of generating grasp candidates can be formulated as sampling finger poses on the surface of the object, and combining them using the kinematics of the gripper to form a grasp as described in [25]. Our method computes the contact probability C_i as given by (7) for each finger and uses the kinematics of the gripper to select a set of finger poses to form a grasp. The local contact probability C_ρ is computed for an infinitesimal surface in a sphere of radius ρ. In order to account for the entire shape of a finger, C_ρ needs to be integrated over its entire surface. We also introduce R, the ranking metric (given by (6)), to rank the grasps by computing the weighted product of the contact probability for each finger.

$$R = k\prod_{i=1}^{n_f} C_i^{w_i} \tag{6}$$

$$C_i = \frac{1}{N_s}\sum_{i=1}^{n} C_\rho^{i,Xi} \tag{7}$$

where, k is a normalizing term, w_i are weights satisfying $\sum_{i=1}^{n} w_i = 1$, n_f the number of fingers, C_i the contact probability for a finger defined in (7), n the number of points in the vicinity of the finger, N_s a normalizing term representing the maximum number of points in the vicinity of the finger, $C_\rho^{i,Xi}$ the local contact moment probability between a point on the point cloud and its orthogonal projection on the surface of the gripper. More information on how to combine probability

Algorithm 1: Grasp generation and ranking.

Data: Point Cloud $\mathcal{X}$, Fingers' 3D model, Sphere Radius ρ
Result: Top-k grasps

Compute the surface normal at each point $X \in \mathcal{X}$
for $\forall X \in \mathcal{X}$ **do**
  Select the set of points ξ in $B_\rho(X)$
  Compute n_ρ with (1)
end
for *each finger* **do**
  for $\forall X \in \mathcal{X}$ **do**
    Sample several finger poses $\mathcal{P}_f$ around X
    for $p \in \mathcal{P}_f$ **do**
      Select the set of points $\mathcal{X}_s$ within a distance d from the surface of the finger
      for $X_s \in \mathcal{X}_s$ **do**
        Project X_s on the finger's surface
        Compute $C_\rho^{s,Xs}$ with (3)
      end
      Compute C_i with (7)
      Append $\mathcal{P}_f$ to $\mathcal{P}$
    end
  end
end
Find $\mathcal{F}$, the set of finger poses in $\mathcal{P}$ satisfying the kinematic constraints of the gripper
for $\forall f \in \mathcal{F}$ **do**
  Compute R with (6)
end
Order $\mathcal{F}$ by decreasing order of R
Sample gripper pose from $\mathcal{F}$
return the Top-k grasp poses

distributions can be found in [26]. A summary of the method can be found in Alg. 1.

IV. Experimental Results

A. Experimental setup

Our experimental setup (shown in Fig. 3) comprises a 7 degrees of freedom KUKA LBR iiwa arm whose end-effector is mounted with a Schunk PG70 parallel jaw gripper with flat fingers. The maximum stroke of the gripper is 68 mm. The developed method neither require any prior knowledge of the scene nor use any object models. However, for each grasping trial, the robot workspace containing test objects is observed by moving a robot wrist-mounted Ensenso N35 depth camera to six different locations. Resulting partial point clouds from all viewpoints are stitched together, in robot base coordinate frame, to form a point cloud of the work scene. After segmenting the ground plane, the resulting cloud is then used by our method to generate grasp hypotheses. Hand-eye calibration has been performed beforehand to transform the camera-acquired point cloud data to robot's coordinate system as well as to simplify the computations.

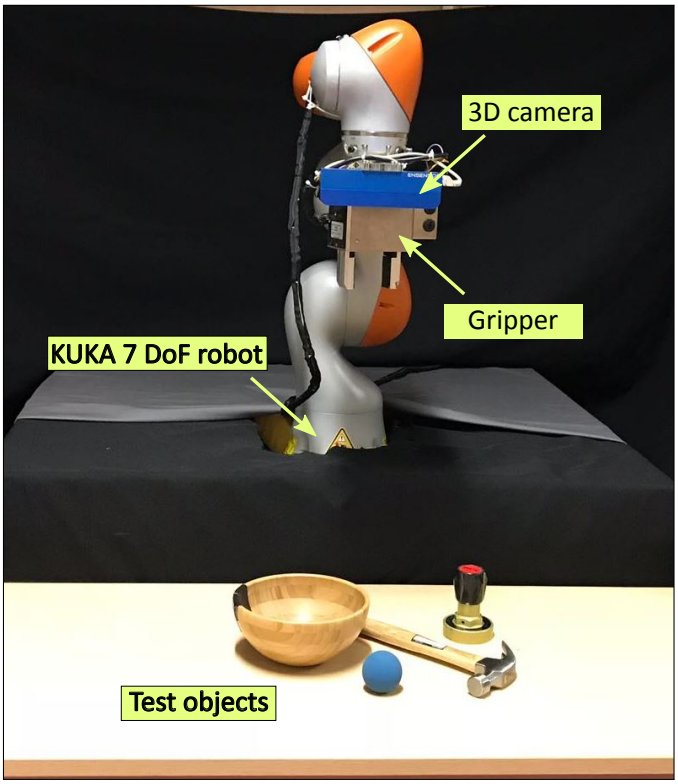

Figure 3. Hardware setup used to validate the proposed grasping method.

Figure 4. 21 objects used for the experiments. (left-column) spring clamp, aluminum profile, multi-head screwdriver, screwdriver, plastic strawberry, golf ball; (middle) racquetball, plastic lemon , plastic nectarine, wood block, potted meat can, electric hand drill, plastic bottle, gray pipe, white pipe; (right-column) blue cup, hammer, bleach cleanser, gas knob, bamboo bowl, mustard container.

The proposed grasping method has been tested on 21 objects, as shown in Fig.4, comprising a wide variety of shapes, masses, materials, and textures. 13 of them are from the YCB object set [27]. The objects are selected such that they are small enough to be physically graspable by the used gripper.

Two sets of experiments were conducted. Firstly, we tested the robot's ability to grasp and lift individual objects from

Table I
SET OF OBJECTS USED FOR THE EXPERIMENT.

Object	Success Rate 1^{st} Grasp (5 Trials)
bleach cleanser	80% (4/5)
racquetball	100% (5/5
blue cup	80% (4/5)
aluminium profile	100% (5/5)
plastic bottle	100% (5/5)
bamboo bowl	100% (5/5)
spring clamp	100% (5/5)
electric hand drill	80% (4/5)
gas knob	100% (5/5)
golf ball	100% (5/5)
hammer	100% (5/5)
plastic lemon	80% (4/5)
mustard container	100% (5/5)
plastic nectarine	100% (5/5)
gray pipe	100% (5/5)
potted meat can	40% (2/5)
screwdriver	100% (5/5)
plastic strawberry	100% (5/5)
multi-head screwdriver	100% (5/5)
white pipe	60% (3/5)
wood block	100% (5/5)
Success Rate	**91.43%** (96/105)

the surface of a table. Second set of tests were performed to analyse the robot's ability to clear randomly piled heaps of objects, by grasping and lifting objects successively, until none remained. During trials, running on an Intel Core i7-4790K CPU @ 4.00GHz and 16 GB RAM, our method took 13.53 seconds (on an average) to generate 1500 grasp hypotheses for a point cloud with 31183 data points corresponding to a clutter scene of 13 Objects. This computational time is distributed as follows. The local contact moment computation is performed in 1.26 seconds (9.3%), the selection of finger pairs with feasible gripper kinematics is done in 6.29 seconds (46.5%), and the robot's end effector pose sample and inverse kinematics check takes up to 5.98 seconds (44.2%).

B. Grasping individual objects

Our first experiment evaluates the robot's ability to grasp and lift individual objects off a flat table surface. 21 objects were used, with five grasping trials performed on each object. For each of the five trials, we randomly placed each object on the table with different orientations and positions. After capturing and registering partial point-clouds from multiple views, points belonging to the table surface are filtered out and the resulting object point cloud is then used to generate grasp hypotheses, as described in Alg.1. The grasps are ranked, and the grasp with the highest likelihood, Eq. (6), is executed. A grasp is recorded as successful if the robot manages to grasp and lift the object to a post-grasp position 20 cm above the table, and hold the object for more than 10 seconds without dropping it.

Table I shows the results of our algorithm when grasping objects that are individually placed on a table. Fig.5 shows images of successful grasps. The overall success rate for all five trials on all 21 objects is 91.43% (96 successful grasps

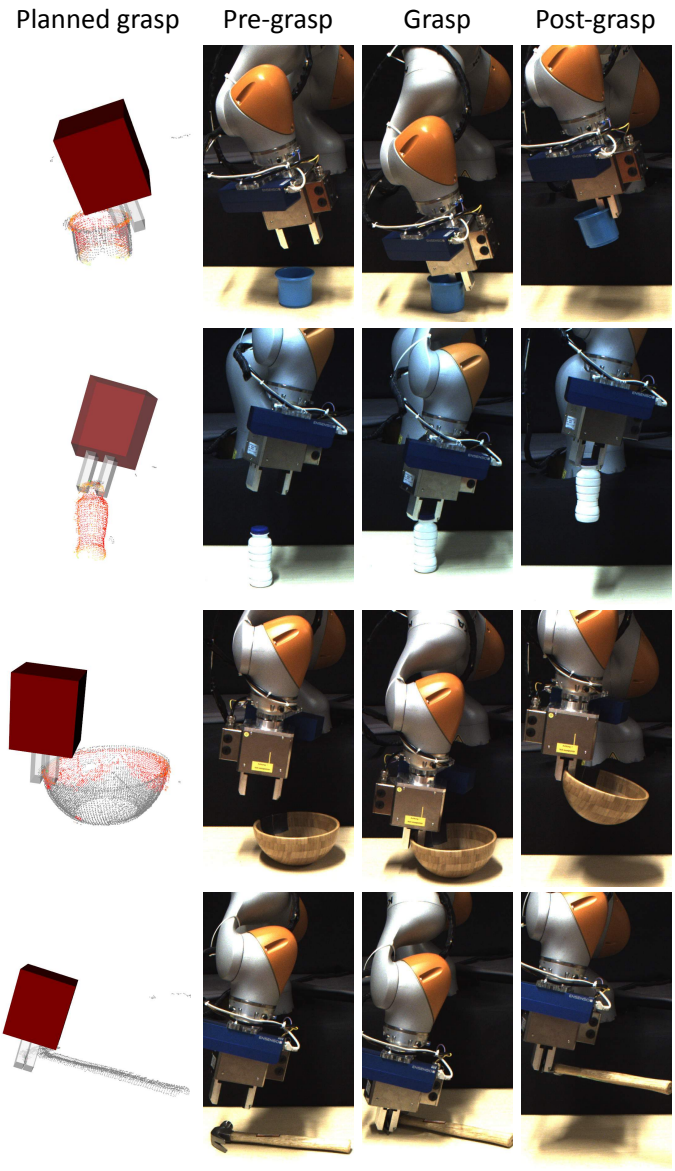

Figure 5. Successful grasps for various objects. In each row, from left to right, the first image shows the point cloud of the object with the contact moment probability and the highest ranked grasp; the second image shows the pre-grasp position of the gripper; the third image shows the grasp; finally, the fourth image shows the post-grasp position with the object grasped.

out of 105). In 97.14% (102/105), the LoCoMo algorithm suggested viable grasps, but objects were dropped for other reasons. For example the object was heavy, and the selected grasp was far from the centre of mass, placing a large torque on the gripper jaws, causing the object to twist loose. In the case of the potted meat can, the success rate was only 40% (2/5). This was due to shiny surfaces which caused a very noisy point cloud.

In safety-critical, high-consequence industries, such as nuclear waste handling or other extreme environments, autonomous robotics methods are likely to be introduced as "operator-assistance technologies", *i.e.*, human-supervised autonomy. In such cases, a human operator might select between

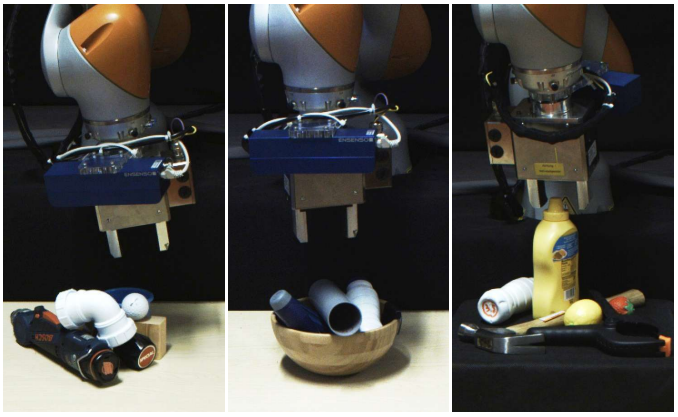

Figure 6. Three different cluttered scenes generated for validating our approach.

several grasps that have been suggested by an autonomous grasp planner. As a small step towards exploring such a system, we repeated the first experiment, however in each attempt we allowed a human to choose one of the best five grasp candidates suggested by the LoCoMo algorithm. In this case, grasp success rose to 98%. This suggests that improved performance might be obtained by combining LoCoMo with other kinds of information, *e.g.*, selecting grasps which result in minimal torques.

C. Grasping objects from a cluttered heap

The second set of experiments was performed on cluttered, self-occluding heaps of objects. For each heap, at least 6 objects were placed in a random pile. Three different heaps were used, Fig. 6. The robot is tasked with clearing the heap, by successively grasping and lifting objects until none remain. No ground plane segmentation was performed in this second experiment. However, the LoCoMo algorithm was able to automatically label the flat table surface as ungraspable, *i.e.*, excluding flat surfaces, and focusing attention on objects, appears to be an inherent behaviour of the algorithm.

At each iteration, grasps are generated, and the highest ranked grasp is executed. Each object is removed without replacement if the grasp is successful, and the experiment is repeated until all the objects are grasped or the algorithm reports that it cannot identify any more feasible grasps. The success of each grasp attempt is determined in the same way as in the first experiment.

Table II shows the results for the heap-picking experiments. We report the results of three different heaps containing at least six objects each. For the first heap, 100% of the objects were grasped successfully from the table, one after the other. Only the gas knob required two trials to be successfully grasped, with all other objects grasped on the first attempt.

For the second heap, all objects were grasped at the first attempt, and the success rate was 100%. During its third grasp, the robot chose to grasp and lift the bowl object, while the bowl still held three other objects inside it (multi-head screwdriver, plastic bottle and nectarine). In order to continue the experiment, these objects were placed back on the table and then successfully grasped, needing only one attempt each.

Table II
CLUTTERED SCENE EXPERIMENT RESULTS.

Scene	Attempt	Object	Success / Failure
#1	1	blue cup	success
	2	golf ball	success
	3	white pipe	success
	4	electric hand drill	success
	5	gas knob	failure
	6	wood block	success
	7	gas knob	success
	8	plastic nectarine	success
#2	1	gray pipe	success
	2	aluminum profile	success
	3	bamboo bowl	success
	4	multi-head screwdriver	success
	5	plastic bottle	success
	6	plastic nectarine	success
#3	1	mustard container	success
	2	plastic bottle	success
	3	spring clamp	failed
	4	plastic lemon	success
	5	spring clamp	success
	6	hammer	failed
	7	hammer	success
	8	plastic strawberry	rolled off table

For the third heap, 83% of the objects (5 out of 6) were successfully grasped. The spring clamp and hammer proved to be difficult, due to sparse point clouds. However, only two attempts were required to grasp these objects. The system did not fail to plan a grasp for the final object (plastic strawberry). Unfortunately, lifting the hammer caused the strawberry to roll off the table so that this final object of the heap could not be completed.

Fig. 7 shows the generated grasps in the cluttered scene 1. The robot was able to clear all three heaps successfully, the only exception being the final object of the third heap, which was pushed off the table during lifting of one of the other objects.

D. Discussion

Overall, results suggest that the LoCoMo algorithm is very promising. For lifting individual objects, a success rate of $\mathbf{91.43}\%$ was obtained over five different trials on 21 different objects, featuring a very wide variety of shapes and appearances. This result is remarkable considering that the system did not have any model or other a-priori knowledge of the objects being grasped. Additionally, no training data was required, and no learning was involved to obtain these results. Moreover, in the heap-picking experiments, featuring extreme clutter conditions, LoCoMo was able to grasp most of the objects at the first attempt (15 out of 19 objects) and was able to successfully grasp all objects, of all heaps, with the exception of the final object of the final heap (plastic strawberry) which rolled off the table during earlier activity.

Aside from a small number of unusual incidents, the proposed algorithm appears to have planned robust grasps almost

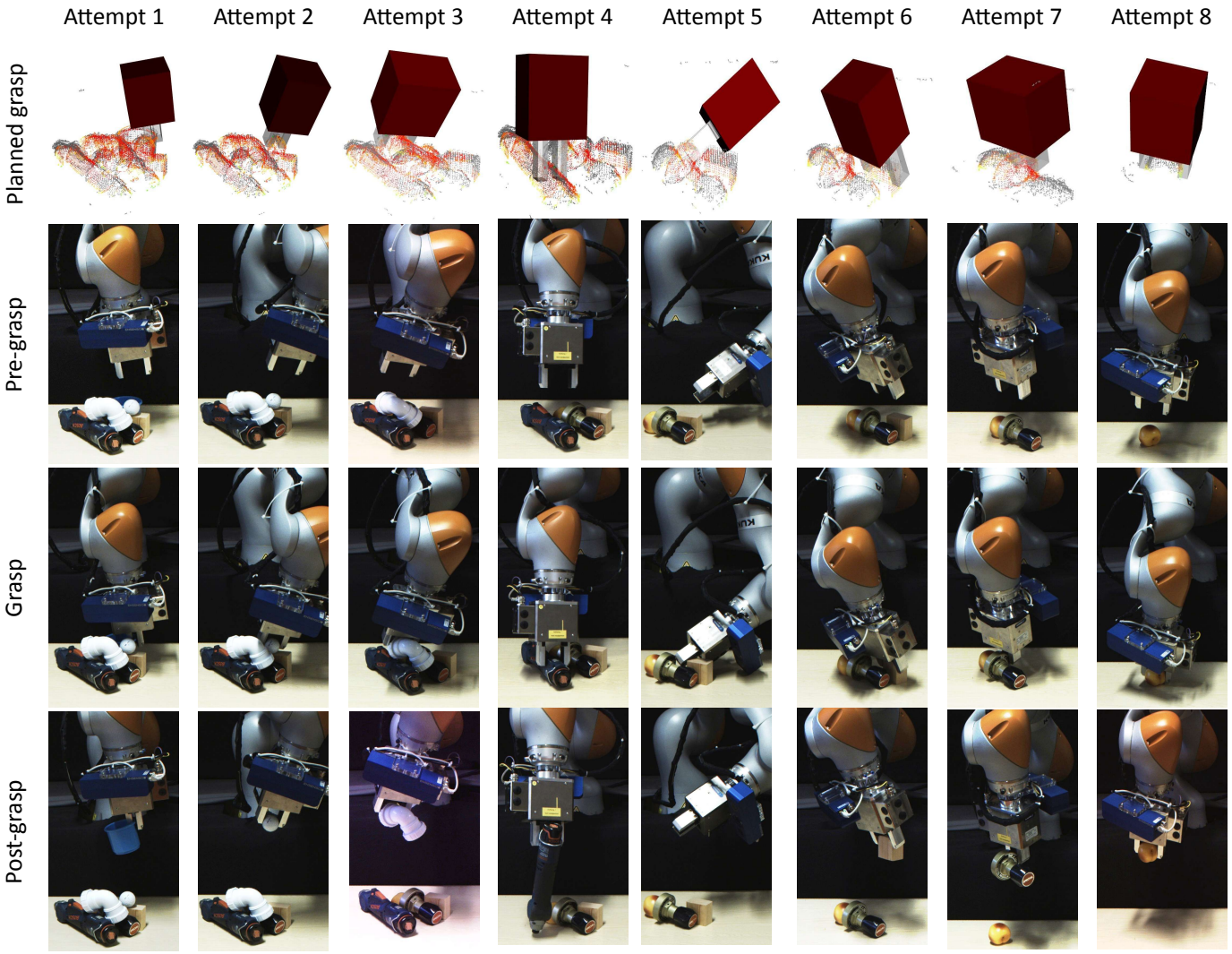

Figure 7. Results of grasp execution in cluttered scenes. First row: images of point cloud of the scene and the gripper. Second row: pre-grasp position of the gripper with respect to the cluttered scene. Third row: execution of the grasp. Fourth row: post-grasp position of the gripper. Chronological sequence is from left to right, *i.e.* first column shows the grasping of the first object, second column the grasping of the second object and so on. Detailed results can be found in the provided supplementary video.

100% of the time. However, we believe that we can improve robustness in several ways. We noted earlier that the set of five highest-ranked grasps occasionally contains a grasp that performs better than the highest ranked grasp. This is because LoCoMo selects grasps based purely on the geometry of surfaces. Combining LoCoMo's robust selection of graspable geometrical features, with other kinds of information such as mass distribution [28], may enable more robust performance. Additionally, combining multiple grasp hypotheses with human-supervised autonomy, appears to outperform pure autonomy based on LoCoMo alone.

V. Conclusion

In this paper, we proposed a novel grasp generation method, based on the LoCoMo metric which searches for similarities between the shape of finger surfaces, and the local shape of an object, observed as a partial point-cloud. The metric is based on zero-moment shift visual features, which encode useful information about local surface curvature. Our method does not rely on any a-priori knowledge about objects or their physical parameters, and also does not require learning from any kind of training data. Grasps are planned from point-cloud images of objects, viewed from a depth-camera mounted on the robot's wrist. Experimental trials, with a real robot and wide variety of objects, suggest that our method generalises well to many shapes. We also demonstrated very robust performance in extremely cluttered scenes. Moreover, the algorithm is also capable of classifying certain objects (*e.g.*, flat table surfaces) as not graspable.

Our future work will focus on improving the performance of the method in terms of speed and extending it to perform multi-finger grasping. We will also focus on accomplishing complex manipulations in challenging scenarios, *e.g.*, nuclear, automotive etc. by integrating it with our previous state

estimation and control methodologies [29], [30].

VI. Acknowledgements

This work forms part of the UK National Centre for Nuclear Robotics initiative, funded by EPSRC EP/R02572X/1. It is also supported by H2020 RoMaNS 645582, and EPSRC EP/P017487/1, EP/P01366X/1. Stolkin is supported by a Royal Society Industry Fellowship. Ortenzi and Corke are supported by the Australian Research Council Centre of Excellence for Robotic Vision (project number CE140100016).

References

[1] A. T. Miller and P. K. Allen, "Graspit! a versatile simulator for robotic grasping," *IEEE Robotics & Automation Magazine*, vol. 11, no. 4, pp. 110–122, 2004.

[2] V.-D. Nguyen, "Constructing force-closure grasps," *The International Journal of Robotics Research*, vol. 7, no. 3, pp. 3–16, 1988.

[3] S. Levine, P. Pastor, A. Krizhevsky, J. Ibarz, and D. Quillen, "Learning hand-eye coordination for robotic grasping with deep learning and large-scale data collection," *The International Journal of Robotics Research*, vol. 37, no. 4-5, pp. 421–436, 2018.

[4] N. Marturi, M. Kopicki, A. Rastegarpanah, V. Rajasekaran, M. Adjigble, R. Stolkin, A. Leonardis, and Y. Bekiroglu, "Dynamic grasp and trajectory planning for moving objects," *Autonomous Robots*, in-press.

[5] A. ten Pas, M. Gualtieri, K. Saenko, and R. Platt, "Grasp pose detection in point clouds," *The International Journal of Robotics Research*, p. 0278364917735594, 2017.

[6] U. Clarenz, M. Rumpf, and A. Telea, "Robust feature detection and local classification for surfaces based on moment analysis," *IEEE Transactions on Visualization and Computer Graphics*, vol. 10, no. 5, pp. 516–524, 2004.

[7] J. Weisz and P. K. Allen, "Pose error robust grasping from contact wrench space metrics," in *Robotics and Automation (ICRA), 2012 IEEE International Conference on*. IEEE, 2012, pp. 557–562.

[8] C. Rosales, R. Suárez, M. Gabiccini, and A. Bicchi, "On the synthesis of feasible and prehensile robotic grasps," in *Robotics and Automation (ICRA), 2012 IEEE International Conference on*. IEEE, 2012, pp. 550–556.

[9] M. A. Roa and R. Suárez, "Computation of independent contact regions for grasping 3-d objects," *IEEE Transactions on Robotics*, vol. 25, no. 4, pp. 839–850, 2009.

[10] D. Prattichizzo and J. C. Trinkle, "Grasping," in *Springer handbook of robotics*. Springer, 2008, pp. 671–700.

[11] J.-W. Li, H. Liu, and H.-G. Cai, "On computing three-finger force-closure grasps of 2-d and 3-d objects," *IEEE Transactions on Robotics and Automation*, vol. 19, no. 1, pp. 155–161, 2003.

[12] M. Gualtieri, A. ten Pas, K. Saenko, and R. Platt, "High precision grasp pose detection in dense clutter," in *Intelligent Robots and Systems (IROS), 2016 IEEE/RSJ International Conference on*. IEEE, 2016, pp. 598–605.

[13] M. Kopicki, R. Detry, M. Adjigble, R. Stolkin, A. Leonardis, and J. L. Wyatt, "One-shot learning and generation of dexterous grasps for novel objects," *The International Journal of Robotics Research*, vol. 35, no. 8, pp. 959–976, 2016.

[14] I. Lenz, H. Lee, and A. Saxena, "Deep learning for detecting robotic grasps," *The International Journal of Robotics Research*, vol. 34, no. 4-5, pp. 705–724, 2015.

[15] H. B. Amor, O. Kroemer, U. Hillenbrand, G. Neumann, and J. Peters, "Generalization of human grasping for multi-fingered robot hands," in *Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on*. IEEE, 2012, pp. 2043–2050.

[16] M. Ma, N. Marturi, Y. Li, A. Leonardis, and R. Stolkin, "Region-sequence based six-stream cnn features for general and fine-grained human action recognition in videos," *Pattern Recognition*, vol. 76, pp. 506–521, 2018.

[17] M. Ma, N. Marturi, Y. Li, R. Stolkin, and A. Leonardis, "A local-global coupled-layer puppet model for robust online human pose tracking," *Computer Vision and Image Understanding*, vol. 153, pp. 163–178, 2016.

[18] D. Smeets, J. Keustermans, D. Vandermeulen, and P. Suetens, "meshsift: Local surface features for 3d face recognition under expression variations and partial data," *Computer Vision and Image Understanding*, vol. 117, no. 2, pp. 158–169, 2013.

[19] E. Paquet, M. Rioux, A. Murching, T. Naveen, and A. Tabatabai, "Description of shape information for 2-d and 3-d objects," *Signal processing: Image communication*, vol. 16, no. 1-2, pp. 103–122, 2000.

[20] R. B. Rusu, G. Bradski, R. Thibaux, and J. Hsu, "Fast 3d recognition and pose using the viewpoint feature histogram," in *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*. IEEE, 2010, pp. 2155–2162.

[21] R. M. Murray, Z. Li, S. S. Sastry, and S. S. Sastry, *A mathematical introduction to robotic manipulation*. CRC press, 1994.

[22] C. Ferrari and J. Canny, "Planning optimal grasps," in *Robotics and Automation, 1992. Proceedings., 1992 IEEE International Conference on*. IEEE, 1992, pp. 2290–2295.

[23] A. Bicchi and V. Kumar, "Robotic grasping and contact: A review," in *Robotics and Automation, 2000. Proceedings. ICRA'00. IEEE International Conference on*, vol. 1. IEEE, 2000, pp. 348–353.

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

[25] M. Kopicki, R. Detry, F. Schmidt, C. Borst, R. Stolkin, and J. L. Wyatt, "Learning dexterous grasps that generalise to novel objects by combining hand and contact models," in *Robotics and Automation (ICRA), 2014 IEEE International Conference on*. IEEE, 2014, pp. 5358–5365.

[26] S. Kaplan, "Combining probability distributions from experts in risk analysis," *Risk Analysis*, vol. 20, no. 2, pp. 155–156, 2000.

[27] B. Calli, A. Walsman, A. Singh, S. Srinivasa, P. Abbeel, and A. M. Dollar, "Benchmarking in manipulation research: Using the yale-cmu-berkeley object and model set," *IEEE Robotics & Automation Magazine*, vol. 22, no. 3, pp. 36–52, 2015.

[28] N. Mavrakis, R. Stolkin, L. Baronti, M. Kopicki, M. Castellani *et al.*, "Analysis of the inertia and dynamics of grasped objects, for choosing optimal grasps to enable torque-efficient post-grasp manipulations," in *Humanoid Robots (Humanoids), 2016 IEEE-RAS 16th International Conference on*. IEEE, 2016, pp. 171–178.

[29] V. Ortenzi, N. Marturi, R. Stolkin, J. A. Kuo, and M. Mistry, "Vision-guided state estimation and control of robotic manipulators which lack proprioceptive sensors," in *Intelligent Robots and Systems (IROS), 2016 IEEE/RSJ International Conference on*. IEEE, 2016, pp. 3567–3574.

[30] N. Marturi, A. Rastegarpanah, C. Takahashi, M. Adjigble, R. Stolkin, S. Zurek, M. Kopicki, M. Talha, J. A. Kuo, and Y. Bekiroglu, "Towards advanced robotic manipulation for nuclear decommissioning: a pilot study on tele-operation and autonomy," in *Robotics and Automation for Humanitarian Applications (RAHA), 2016 International Conference on*. IEEE, 2016, pp. 1–8.