2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
Madrid, Spain, October 1-5, 2018

A Framework for Robot Grasp Transferring with Non-rigid Transformation

Hsien-Chung Lin*, Te Tang*, Yongxiang Fan, and Masayoshi Tomizuka

Abstract— **Grasp planning is essential for robots to execute dexterous tasks. Solving the optimal grasps for various objects online, however, is challenging due to the heavy computation load during exhaustive sampling, and the difficulties to consider task requirements. This paper proposes a framework to combine analytic approach with learning for efficient grasp generation. The example grasps are taught by human demonstration and mapped to similar objects by a non-rigid transformation. The mapped grasps are evaluated analytically and refined by an orientation search to improve the grasp robustness and robot reachability. The proposed approach is able to plan high-quality grasps, avoid collision, satisfy task requirements, and achieve efficient online planning. The effectiveness of the proposed method is verified by a series of experiments.**

I. Introduction

Grasping is an essential capability for robots to accomplish complex manipulation tasks. In traditional grasping scenarios, such as pick-and-place in assembly lines, object-specific grippers might be designed for the robust grasping of a single type of objects. In recent years, however, more and more applications require a versatile ability for robots to grasp various objects with general purpose grippers. For example, human robot interaction requires collaboration and assistance between robots and humans, during which robots may pass different tools to humans or help holding various workpieces for assembly tasks. The increasing demand for massive customization and warehouse automation also promotes the development of dexterous grasping.

However, the grasp planning for various objects with general purpose grippers is challenging to solve due to heavy computational loads, large task variance and imperfect perceptions. First of all, many analytic planners such as Ferrari-Canny metric [1] and grasp isotropy [2] require considerable time for exhaustive searching and complex computation for evaluation. Secondly, these planners generally assume point contact, and calculate the quality based on the local features such as contact position and contact normal, while the global task requirements such as robot reachability and collision avoidance are not under consideration. Moreover, analytic planners are usually sensitive to the noises and distortion of point clouds caused by hardware limitations and calibration errors. Therefore, the grasp quality evaluated by analytic planners is usually inconsistent with the empirical success rate and cannot resemble reality effectively [3].

*These authors equally contributed to this work as the lead authors.
The authors are with the Department of Mechanical Engineering, University of California, Berkeley, CA 94720, USA. Email: {hclin, tetang, yongxiang_fan, tomizuka} @berkeley.edu

Another common approach for grasp planning is to learn optimal grasps from previous grasp examples. For example, the Dex-Net [4] trains a neural network from a database which is built by analytic planners. The network is able to estimate optimal grasps for unseen objects after training. In [5], the grasp is calculated from heatmaps that generated by deep learning. However, these methods usually require considerable data for the training process and the optimal grasp is planned without considering the task constraints.

Despite the variance of object shapes, we notice that objects to grasp can be classified into several categories. For example, in the tool picking scenario, objects can often be specified into categories such as wrenches, pliers, and screwdrivers. Objects in each category share similar topological structures but may have difference in shapes and sizes.

Some research has been conducted based on this observation. In [6], the perceived cloud of the object is fitted to different objects templates in the database, and the grasp is estimated by superimposing all representations considering their confidence levels. A semantic grasping is proposed in [7] to consider task requirements. The task constraints are implicitly represented by a grasp example in each object category and the desired grasp on the novel object is retrieved by mapping the grasp example and refined by eigen-grasp planner. A dictionary of object parts is learned in [8] to generate grasps across partially similar objects. The dictionary assumes that the segments that shared by objects are rigid and have similar sizes. However, this assumption cannot hold in many scenarios.

In this work, we propose a novel framework for efficient and effective grasp generation from previous grasp examples. Firstly, a 'learning from human demonstration' approach is introduced to teach robots candidate grasp poses by human experts. In the test stage, the category of the target object will be classified by its similarities towards the taught objects. Then a grasp pose transferring is performed between similar objects based on the concept of coherent point drift method [9], [10]. Moreover, the transformed poses will be rated by analyzing the grasp isotropy metric [2]. An orientation search method will also be introduced to improve the robot reachability and avoid collisions.

The remainder of this paper is organized as follows: Section II introduces the normal formulation of grasping problems and the benefits of involving human demonstration. The background of coherent point drift, together with its application on grasp pose transferring is introduced in Section III. Section IV presents the dissimilarity measure between objects and the refinement of poses after transferring. A

series of experiments on grasping multiple categories of objects are shown in Section V. Experimental videos can be found in [11]. Section VI concludes the paper and proposes future works.

II. Grasp Planning with Human Demonstration

A basic grasp planning for parallel grippers can be formulated as

$$\max_{\boldsymbol{c},\boldsymbol{n_c}} Q(\boldsymbol{c},\boldsymbol{n_c}) \tag{1a}$$

$$s.t. \quad c_i \in \partial O \quad i=1,2 \tag{1b}$$

$$\|c_1 - c_2\| \leq w_{\max}, \tag{1c}$$

where Q denotes the grasp quality to be maximized, $\boldsymbol{c} = \{c_1, c_2\}$ denotes the contact pair with $c_i \in \mathbb{R}^D$, and $\boldsymbol{n_c} = \{n_{c,1}, n_{c,2}\}$ denotes the normals of the contact pair with $n_{c,i} \in \mathbb{S}^{D-1}$. Constraint (1b) shows the contacts should lie on the surface of object ∂O, and (1c) shows that the distance of the contact pair should be less than the width of the gripper $w_{\max}$.

Equation (1) is challenging to solve by gradient-based methods because of the high complexity of surface modeling, the discrete representation of surface points, and the discontinuity of surface normal. Compared with gradient-based searching, sampling-based methods are able to adapt to discrete object representation and escape from local optimum. However, they require considerable computation for sampling and quality evaluation to find a reasonable grasp due to the complicated structure of the object and the feasibility constraints such as gripper width, task requirements and collisions, thus the direct sampling methods are generally not affordable for real-time implementation.

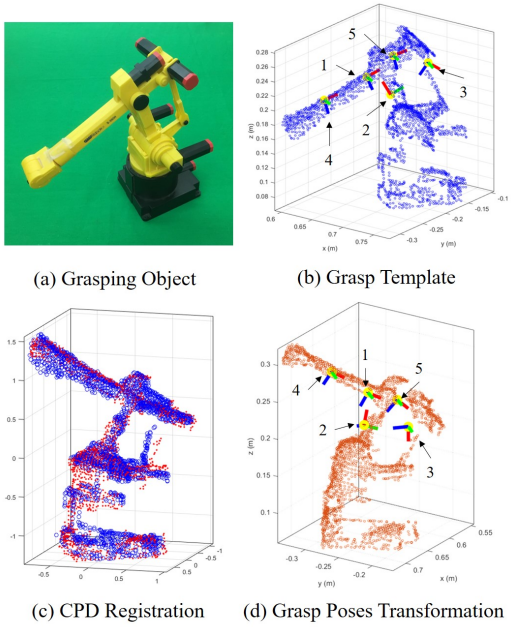
In this paper, we assume that the objects to grasp can be clustered into various categories. The objects in the same category share similar topological structures but can have different shapes, sizes and configurations. The objective of this paper is to provide an efficient framework to grasp objects in the same category without overwhelmed training, modeling and computation. To achieve this, we introduce human demonstration to accelerate grasp searching by providing heuristics to guide sampling. Instead of directly using human demonstration as the sampling pool for the target object to grasp, we use a mapping function to transfer the example grasps based on the topological similarity between the source object and the target object. Therefore, (1) becomes:

$$\max_{\boldsymbol{c},\boldsymbol{n_c}} Q(\boldsymbol{c},\boldsymbol{n_c}) \tag{2a}$$

$$s.t. \quad \{\boldsymbol{c},\boldsymbol{n_c}\} \in map(\mathcal{H}) \tag{2b}$$

$$\|c_1 - c_2\| \leq w_{\max}, \tag{2c}$$

where $\mathcal{H}$ denotes a human demonstration database containing example grasps on the source object, and the function $map(\cdot)$ represents a grasp transferring. Compared with (1), the introduction of human demonstration in (2) has the following advantages. First, incorporating human intelligence into the framework will improve the empirical success rate, since the human demo usually considers a variety of factors such as the local structure of the object and the global geometry for collision avoidance. Second, some tasks have special requirements. For example, some workpieces have fragile parts or polished surfaces which are not suitable for grasping. Some workpieces have some preferred grasp poses for the ease of following assembly procedures. Explicitly imposing such constraints to traditional approaches is nontrivial, while these requirements can be easily encoded by human demonstration. Moreover, by mapping the grasp examples to novel objects, the proposed method exploits much fewer but reasonable grasp samples compared to traditional exhaustive search methods. Therefore, the searching time is greatly reduced.



(a) Grasping Object (b) Grasp Template

(c) CPD Registration (d) Grasp Poses Transformation

Fig. 1. The Grasp Pose Transferring: (a) A toy manipulator model as a grasping object. (b) The grasp example that contains a source object and several grasp poses. (c) The non-rigid point registration by Coherent Point Drift. (d) The target object with the warped grasp poses. The grasp poses are labeled with number.

III. Grasp Pose Transferring by Point Registration

Assume a grasp template consists of a source object (Fig. 1a) and multiple demonstrated grasp poses (Fig. 1b), where the blue dots are the point clouds of the source object and each coordinate labeled with a number represents a demonstrated grasp pose. The source object is represented by a point set $\mathbf{X} = (x_1, \cdots, x_N) \in \mathbb{R}^{N\times D}$, where $x_n \in \mathbb{R}^D$ is the n-th point in the point set. The grasp poses are denoted as $\boldsymbol{g}_i = (\boldsymbol{t}_i, \boldsymbol{R}_i) \in \mathbb{R}^D \otimes \mathbf{SO}(D)$, $i = 1, 2, \cdots, I$, where $\boldsymbol{t}_i \in \mathbb{R}^D$ is the center of the grasping point, $\boldsymbol{R}_i \in \boldsymbol{SO}(D)$ represents the grasp orientation, and i is the index among the total I grasp poses. The target object is represented by another point set $\mathbf{Y} = (y_1, \cdots, y_M) \in \mathbb{R}^{M\times D}$, where $y_m \in \mathbb{R}^D$ is the m-th point in the target point set. Our objective is to find a smooth transformation $\mathcal{T} : \mathbb{R}^D \to \mathbb{R}^D$ that maps the source object to the target object as well as transferring

the grasp examples to new grasp poses $\boldsymbol{g}_i' = (\boldsymbol{t}_i', \boldsymbol{R}_i')$ on the target object (Fig. 1d).

The transformation can be found by aligning the source object to the target object. Then the task can be formulated as a point set registration problem as shown in Fig. 1c. Considering variation and deformation between the source object and the target object, the mapping should have more flexibility than rigid transformation. In the meantime, the topological structure of point sets must be preserved during the alignment process so that the the grasp pose can be transferred to a reasonable location. In this work, we use the coherent point drift (CPD) algorithm [9] to perform a smooth non-rigid point registration.

A. Coherent Point Drift

In order to align the source object toward the target object, CPD considers source points in $\mathbf{X}$ as the centroids of Gaussian mixtures, and transforms them to fit the target points in $\mathbf{Y}$ coherently. The source points are assumed to deform toward the target points according to a continuous displacement field $v(\cdot)$, and the transformed source point is written as

$$\mathcal{T}(x_n) = x_n + v(x_n). \tag{3}$$

The goal of CPD is to retrieve the displacement field v that maximizes the likelihood of $\mathbf{Y}$ sampled from $\mathbf{X}$.

With the Gaussian mixture model, the probability distribution of y_m can be described as

$$\begin{aligned} p(y_m) &= \sum_{n=1}^{N} p(n)p(y_m|n) \\ &= \sum_{n=1}^{N} \frac{1}{N}\mathcal{N}(y_m; \mathcal{T}(x_n), \sigma^2) \\ &= \sum_{n=1}^{N} \frac{1}{N} \frac{1}{(2\pi\sigma^2)^{D/2}} \exp(-\frac{\|y_m - x_n - v(x_n)\|^2}{2\sigma^2}), \end{aligned} \tag{4}$$

where it is assumed that each Gaussian shares the same isotropic covariance σ^2 and has equal membership probability $p(n) = 1/N$.

Since there might be some noise and outliers from the measurement, which may deteriorate the result of registration, an additional uniform distribution is added to the mixture model to take account of these effects. Thus, the complete mixture model is reformulated as

$$\begin{aligned} p(y_m) &= \sum_{n=1}^{N+1} p(n)p(y_m|n) \\ &= (1-\mu)\sum_{n=1}^{N} \frac{1}{N}\mathcal{N}(y_m; \mathcal{T}(x_n), \sigma^2) + \frac{\mu}{M}, \end{aligned} \tag{5}$$

where $\mu \in [0, 1]$ denotes the weight of the uniform distribution. The log-likelihood function of $\mathbf{Y}$ is given by

$$\begin{aligned} l(v, \sigma^2) &= \log \prod_{m=1}^{M} p(y_m) \\ &= \sum_{m=1}^{M} \log \sum_{n=1}^{N+1} p(n)p(y_m|n). \end{aligned} \tag{6}$$

The parameter (v, σ^2) can be estimated by maximizing (6); however, it is nontrivial to directly optimize over the log-likelihood function, since the summation inside the $\log(\cdot)$ leads to a non-convex formulation. An alternative log-likelihood function L can be constructed as

$$L(v, \sigma^2) = \sum_{m=1}^{M}\sum_{n=1}^{N+1} p(n|y_m) \log\left(p(n)p(y_m|n)\right). \tag{7}$$

It can be proven by Jensen's inequality [12] that L is the lower bound of l. Hence, increasing the value of L will always 'push' the value of l increased until it reaches the local optimum. Compared with the structure of l, the inside summation of L is moved to the front of the $log(\cdot)$ function, which provides much convenience to maximize the log-likelihood by the EM algorithm [13].

The EM algorithm runs the expectation step (E-step) and maximization step (M-step) iteratively to estimate the parameters by maximizing L.

E-step: The expectation step computes the posterior probability distribution of $p(n|y_m)$ with the previous estimated parameters from the last M-step,

$$p(n|y_m) = \frac{\exp\left(-\frac{\|y_m - x_n - v(x_n)\|^2}{2\sigma^2}\right)}{\sum_{n=1}^{N}\exp\left(-\frac{\|y_m - x_n - v(x_n)\|^2}{2\sigma^2}\right) + c}, \tag{8}$$

where $c = (2\pi\sigma^2)^{D/2}\frac{\mu}{(1-\mu)}\frac{N}{M}$.

M-step: Ignoring the terms that are independent of v and σ^2, the log-likelihood function can be written as

$$\begin{aligned} L(v, \sigma^2) = &-\frac{1}{2\sigma^2}\sum_{n=1}^{N}\sum_{m=1}^{M} p(n|y_m)\|y_m - x_n - v(x_n)\|^2 \\ &- \frac{D}{2}\sum_{n=1}^{N}\sum_{m=1}^{M} p(n|y_m)\log\sigma^2. \end{aligned} \tag{9}$$

The maximization step is to substitute (8) into (9) and take partial derivative with respect to v and σ^2 to find its maximum.

Although alternating between E-step and M-step will converge to a local optimal, it can not guarantee that the topological structure of the source object is preserved after the transformation. That is because there is no topological constraints to restrict the locations of these Gaussian centroids. Therefore, a regularization term is added to the log-likelihood function to regularize the smoothness of the deformation function, and the modified likelihood function is given by

$$\tilde{L}(v, \sigma^2) = L(v, \sigma^2) - \frac{\lambda}{2}\|v\|_{\mathcal{F}}^2, \tag{10}$$

where $\|v\|_{\mathcal{F}}^2 = \int_{\mathbb{R}^D} \frac{|V(s)|^2}{G(s)} ds$ is a norm to quantitatively measure the function smoothness [14]. $V(s)$ is a Fourier transform of v and $G(s)$ presents a symmetric filter that approaches to zero as $s \to \infty$. The overall Fourier domain norm here basically captures the energy of high frequency components of $V(s)$. Intuitively, the larger the norm $||v||_{\mathcal{F}}$, the more 'oscillating' v will be, i.e., less smoothness. $\lambda \in \mathbb{R}^+$ is a weighting coefficient that represents the trade off between the fitting of the point sets and the smoothness constraints on the transformation.

It can be proved by variational calculus that the maximizer of (10) has the form of the radial basis function [9],

$$v(z) = \sum_{n=1}^{N} w_n g(z - x_n), \tag{11}$$

where $g(\cdot)$ is a kernel function retrieved from the inverse Fourier transform of $G(s)$, and w_n is the unknown kernel weights. In general, $g(\cdot)$ can be any formulation with positive definiteness, and $G(s)$ behaves like a low-pass filter. For simplicity, a Gaussian kernel is chosen so that $g(z - x_n) = \exp(-\frac{1}{2\beta^2}||z - x_n||^2)$, where $\beta \in \mathbb{R}^+$ is a parameter that defines the width of smoothing Gaussian filter. Larger β corresponds to more rigid transformation, whereas smaller β produces more local deformation.

Substituting (7) and (11) to (10), we get

$$\tilde{L} = \frac{-1}{2\sigma^2} \sum_{n=1}^{N} \sum_{m=1}^{M} p(n|y_m) \|y_m - x_n - \sum_{k=1}^{N} w_k g(x_n - x_k)\|^2 - \frac{D}{2} \sum_{n=1}^{N} \sum_{m=1}^{M} p(n|y_m) \log \sigma^2 - \frac{\lambda}{2} tr(\mathbf{W}^T \mathbf{G} \mathbf{W}), \tag{12}$$

where $\mathbf{G} \in \mathbb{R}^{N \times N}$ is a Gramian matrix with element $\mathbf{G}_{ij} = g(x_i - x_j)$ and $\mathbf{W} = [w_1, \cdots, w_n]^T \in \mathbb{R}^{N \times D}$ is the vectorization of kernel weights in (11).

From (12), the regularized log-likelihood function is now parameterized by $(\mathbf{W}, \sigma^2)$. Similar to (7), the EM algorithm can be performed to estimate the parameters iteratively. In E-step, the posterior $p(n|y_m)$ is calculated by using the previous estimated parameters. In M-step, take $\partial \tilde{L}/\partial \mathbf{W} = 0$ and $\partial \tilde{L}/\partial \sigma^2 = 0$ to obtain a new estimate of $(\mathbf{W}, \sigma^2)$. The closed-form solution for M-step requires further mathematical derivation, more details can be found in [9], [10].

After $\tilde{L}$ is converged, the point set of the source object $\mathbf{X}$ can be aligned toward the target object by

$$\mathcal{T}(\mathbf{X}) = \mathbf{X} + \mathbf{G}\mathbf{W}. \tag{13}$$

B. Grasp Pose Transferring

As shown in Fig. 1d, after finding the mapping from the source object $\mathbf{X}$ to the target object $\mathbf{Y}$, the demonstrated grasp poses on $\mathbf{X}$ will also be transferred to achieve new grasp poses that are suitable for object $\mathbf{Y}$. The grasp poses can be decomposed to two parts: the position and the orientation of the robot end-effector. Regards to the position, the non-rigid transformation $\mathcal{T} : \mathbb{R}^D \to \mathbb{R}^D$ can directly map the center of grasp from grasp example to the target object by

$$\boldsymbol{t}_i' \leftarrow \mathcal{T}(\boldsymbol{t}_i), \quad i = 1, 2, \cdots, I. \tag{14}$$

As for the orientation, it can be considered as transferring x, y, and z axes of the original grasp orientation to the new object space. One natural way to transform a vector $\mathbf{v}$ at a point $\boldsymbol{t}$ through a function is to multiply the vector with the gradient of $\mathcal{T}(\boldsymbol{t})$ [15], i.e. $\nabla \mathcal{T}(\boldsymbol{t})\boldsymbol{v}$. Considering the properties of the special orthogonal group, the singular value decomposition (SVD) of the matrix is performed to construct the new orientation of the grasp, which is

$$\boldsymbol{R}_i' \leftarrow \mathbf{U}_i \mathbf{V}_i^T, \quad i = 1, 2, \cdots, I. \tag{15}$$

where $\mathbf{U}_i \Sigma \mathbf{V}_i^T = svd(\nabla \mathcal{T}(\boldsymbol{t}_i)\boldsymbol{R}_i)$, $\mathbf{U}_i, \mathbf{V}_i$ are the orthonormal basis of the matrix, and Σ is a diagonal matrix that consists of the singular values of the matrix.

Hence, the new grasp poses on the target can be transferred by

$$\boldsymbol{g}_i' = (\boldsymbol{t}_i', \boldsymbol{R}_i') \leftarrow (\mathcal{T}(\boldsymbol{t}_i), \mathbf{U}\mathbf{V}_i^T), \quad i = 1, 2, \cdots, I. \tag{16}$$

IV. Grasp Poses for Various Objects

A. Dissimilarity Measure

During the training stage, multiple grasp poses for different categories of objects are demonstrated by human experts. Given a new object at test, it is necessary to first classify which category the object belongs to, then use the Section III method to transfer the corresponding grasp poses from the correct category to get a new feasible grasp. Therefore, an object classifier is essential for pose transferring.

There are some researches that apply surface matching technique to rigidly fit the object template to the measured point clouds and calculate the dissimilarity [16], [17]. The source objects from different categories are exploited as the templates to match the target object. By measuring the dissimilarity between each of the source objects $\mathbf{X}$ and the target object $\mathbf{Y}$, the most similar pair will be selected to determine the category of the target object. In our work, since CPD can be applied to warp the template $\mathbf{X}$ to $\mathcal{T}(\mathbf{X})$ which is aligned with $\mathbf{Y}$, the residual dissimilarity between $\mathcal{T}(\mathbf{X})$ and $\mathbf{Y}$ instead of the dissimilarity between $\mathbf{X}$ and $\mathbf{Y}$ will be checked to provide a more robust category classification.

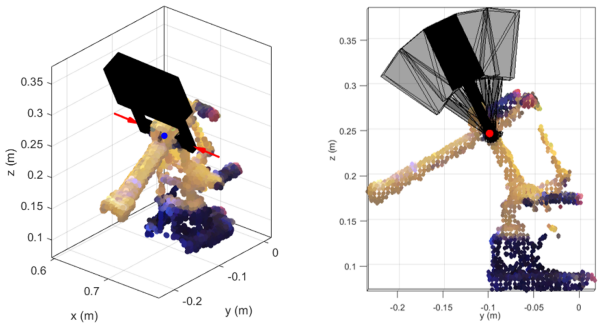
The average minimum distance between the two point sets can be designed as:

$$d(\mathcal{T}(\mathbf{X}), \mathbf{Y}) = \frac{1}{N} \sum_{n=1}^{N} \min_{m \in [1,M]} ||\mathcal{T}(x_n) - y_m||, \tag{17}$$

where $||\mathcal{T}(x_n) - y_m||$ is the Euclidean distance between point $\mathcal{T}(x_n)$ and y_m. Equation (17) is an error function that is commonly used for point cloud alignment. However, (17) is asymmetric. The dissimilarity between a source object and a target object can be formulated as

$$D(\mathbf{X}', \mathbf{Y}) = d(\mathbf{X}', \mathbf{Y}) + d(\mathbf{Y}, \mathbf{X}'), \tag{18}$$

where $\mathbf{X}' = \mathcal{T}(\mathbf{X})$ is the source points warped toward $\mathbf{Y}$ by CPD. The function $D(\cdot, \cdot)$ sums the two asymmetric



(a) (b)

Fig. 2. (a) a grasp example, where the red arrows indicate the direction of gripper closing (which is also the grasp axis). (b) The side view of the grasp example, and the transparent grippers are shared the same grasp center and grasp axis but different orientations.

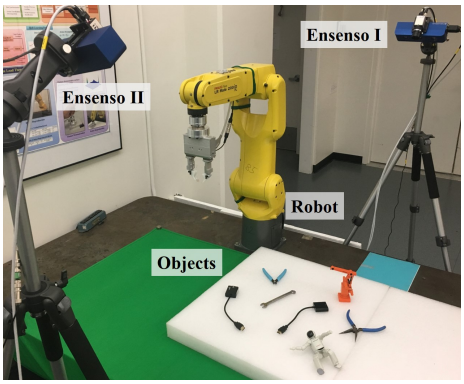


Fig. 3. The experimental setup: a FANUC LR Mate 200*i*D/7L and dual Ensenso stereo cameras

dissimilarity measurements together so that D is symmetric to its input arguments, i.e. $D(\mathbf{X}', \mathbf{Y}) = D(\mathbf{Y}, \mathbf{X}')$.

Suppose there are K object categories, the most possible category that the target object belongs to can be estimated by

$$\mathbf{k}^* = \arg \min_{k \in [1,K]} \; D(\mathbf{X}'_k, \mathbf{Y}). \tag{19}$$

B. Grasp Pose Optimization

Once the object category is determined, we can map the example poses from the corresponding category to the target object. The grasp quality of the mapped poses will then be evaluated by analytic methods using the grasp isotropy index [2]. The grasp isotropy index measures the uniformness of different contact forces to the total wrench. More concretely, it can be written as

$$Q_i = \frac{\sigma_{\min} \mathcal{G}(\boldsymbol{g}'_i, \boldsymbol{g}_o)}{\sigma_{\max} \mathcal{G}(\boldsymbol{g}'_i, \boldsymbol{g}_o)}, \tag{20}$$

where $\boldsymbol{g}_o$ denotes the pose of the object, $\mathcal{G}(\boldsymbol{g}'_i, \boldsymbol{g}_o)$ represents the grasp map determined by the contacts and the object [18], and $\sigma_{\min}$ and $\sigma_{\max}$ respectively denote the minimum and maximum singular values of the grasp map. The contacts are inferred by the line search along the grasp axis. The line search tries to locate the nearest neighbor of the grasp center on the object's point cloud. The contacts are represented by the nearest neighbors search in the positive and negative directions of the grasp axis respectively. The transferred grasp would be treated as a bad pose if the contacts deviate from grasp axis too much, in which case a negative quality will be allocated.

Apart from the grasp quality, we have to consider the feasibility constraints such as the reachability and the gripper-object collision. The feasibility constraints are guaranteed by an orientation search introduced below.

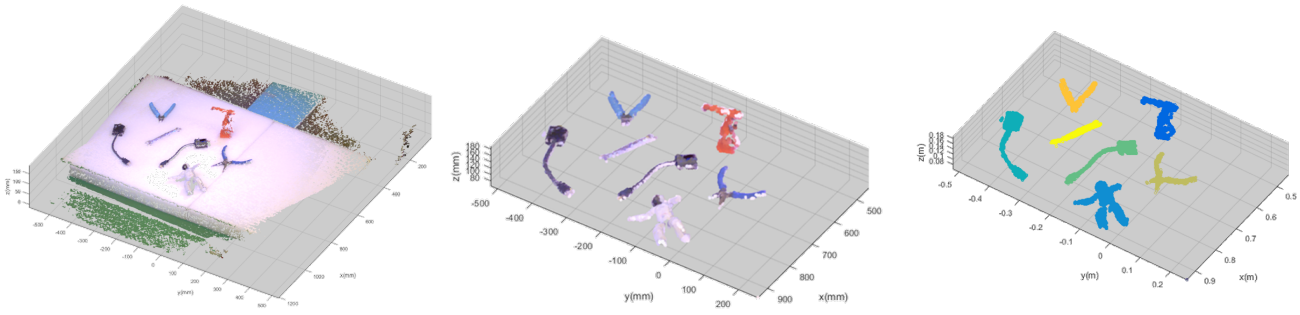
Because the robot grasp pose with a parallel-jaw gripper is composed of a center of grasp and a grasp axis, it does not necessarily restrict all the rotation axis. i.e. the grasp quality is not affected by rotating along the grasp axis. A parallel-jaw gripper grasp example is shown in Fig. 2a, where the center of grasping point is the blue dot and the red arrows represent the operational direction of the jaw which parallels to the grasp axis. By rotating along the grasp axis, the grasp pose can be modified as the translucent grippers as shown in Fig. 2b, where the modified poses are also valid grasps. If the initial grasp pose is not feasible, then the modification can be made by searching the various orientations around the initial one. Suppose the initial orientation is denoted as $\boldsymbol{R}_0$, the sampled orientation is denoted as $\boldsymbol{R}_i$, and $\mathcal{R}$ is the set of all the sampled orientations. The orientation search can be formulated as

$$\min_{\boldsymbol{R}_i \in \mathcal{R}} \; \Delta\xi(\boldsymbol{R}_0, \boldsymbol{R}_i) + C\left[f_{IK}(\boldsymbol{t}, \boldsymbol{R}_i) + f_{col}(\boldsymbol{t}, \boldsymbol{R}_i, \mathbf{Y})\right], \tag{21}$$

where $\Delta\xi(\boldsymbol{R}_0, \boldsymbol{R}_i) = 1 - \xi(\boldsymbol{R}_0)^T \xi(\boldsymbol{R}_i) \in [0, 1]$ is the rotation deviation in quaternion between $\boldsymbol{R}_0$ and $\boldsymbol{R}_i$, $\xi(\cdot)$ converts a rotation matrix to a quaternion. We use quaternion rather than Euler angles to represent rotation difference to avoid singular representation in rotations. $f_{IK}(\boldsymbol{t}, \boldsymbol{R})$ is a boolean function that returns 1 when the inverse kinematics of $(\boldsymbol{t}, \boldsymbol{R})$ is invalid and returns 0 otherwise. $f_{col}(\boldsymbol{t}, \boldsymbol{R}, \mathbf{Y})$ is another boolean function that return 1 when the gripper with the pose $(\boldsymbol{t}, \boldsymbol{R}_i)$ is collided with $\mathbf{Y}$ and returns 0 otherwise. C is a large constant number to penalize the condition of both the infeasible inverse kinematics and the gripper-object collision. If all the sampled orientations are invalid, the value of (21) will be greater than or equal to C. Then the orientation search is applied to exploit the other candidates until it finds a feasible grasp pose to perform the task.

V. Experimental Results

In order to verify the proposed grasping approach, a series of experiments were conducted to grasp various objects by a robot manipulator. The experimental setup is shown in Fig. 3, where the robot was FANUC LR Mate 200*i*D/7L, and two Ensenso stereo cameras were calibrated and synchronized to capture the point clouds of objects in the workspace. All the programs were implemented in MATLAB on a Windows desktop with a Intel Core i5 CPU and 16GB RAM. The robot controller was deployed on a Simulink RealTime target.



(a) (b) (c)

Fig. 4. The point cloud process, (a) The raw data was captured by the dual Ensenso stereo camera. (b) The objects were extracted from the background by predefined region. (c) The point cloud was clustered by DBSCAN.

The point clouds retrieved from the dual Ensenso stereo cameras were shown in Fig. 4a. By applying the snapshot of the empty workspace as a filter mask, the point clouds of objects were extracted from the background as shown in Fig. 4b. Then running the density-based spatial clustering application with noise (DBSCAN) algorithm [19], the point clouds can be separated to several clusters to represent different objects (Fig. 4c). A voxel grid filter with step size 5mm was implemented to downsample the point clouds uniformly.

Six categories of objects, including cups, pliers, wrenches, cable adapters, toy manipulator models and toy humanoid models, were tested in the experiment (Fig. 5). Note that neither CAD models nor mesh files were used in this work. For each category, a specific source object was selected, and the human operator taught multiple preferred grasp poses on it through kinesthetic teaching. The point cloud of the object and the demonstrated grasp poses were recorded as training database.

At the test stage, objects with different sizes and configurations across all the categories were randomly placed in the workspace. For example, multiple types of cups and wrenches were tested for grasping; the pliers were either open or closed; the cable adapter were twisted to various shapes; the joints of the two toy manipulator models and the toy humanoid model were rotated to random angles. All the target objects were shown in Fig. 6.

Before executing the grasp experiment, an object classification test was performed by measuring the dissimilarity between the target object and all the source objects (see Section IV.A). The target objects in Fig. 6 were randomly placed, with each category of objects collecting 20 different configurations. The parameters of CPD were set as $\beta = 2$ and $\lambda = 50$.

As shown in Fig. 7, the performance of object classification was presented by a confusion matrix, where each column represented the predicted class and each row represented the actual class. The diagonal entries of the confusion matrix indicated the correct classification, whereas the off-diagonal entries were misclassification. The overall classification accuracy was 94.17% (113/120).

Each category of objects was tested 20 times for grasping with different orientations, shapes, sizes, and configurations. The parameters of CPD were the same as the ones in object classification.

TABLE I

GRASPING QUALITY EVALUATION

Grasping Pose No.	1	2	3	4	5
Isotropy Index	0.0098	-1.000	0.0001	0.0089	-1.000

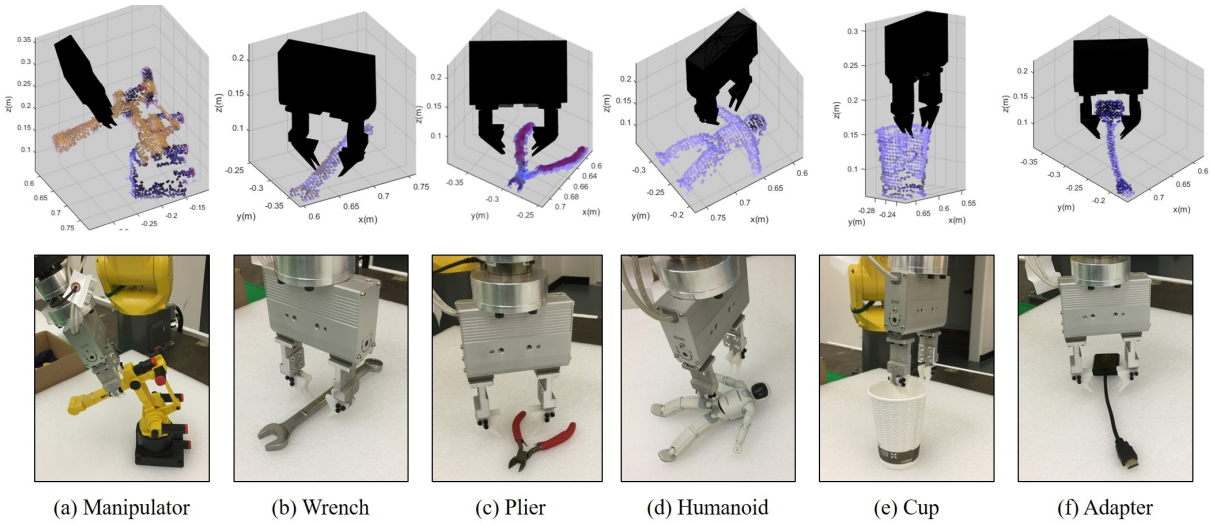
TABLE II

GRASPING RESULTS

class	success/trials	avg. CPD time (ms)	avg. numbers of points
manipulator	19/20	1276.4	1563.7
wrench	20/20	111.3	316.7
plier	18/20	706.1	1419.0
humanoid	17/20	369.5	773.3
cup	20/20	350.5	609.3
adapter	19/20	480.2	917.0
average	18.8/20	549.0	933.2

Take one target object (Fig. 1d) as an example. The grasping qualities of the transferred grasps are provided in Table I. Note that the qualities of the second and fifth transferred poses are marked as negative based on the isotropy index analysis, since the second pose was mapped to a region with sparse points, and the contacts for the the fifth grasp was wider than the width of the gripper. The remaining pose with the highest grasping quality, i.e., the first pose, was selected. The selected pose was then refined by the orientation search to improve the reachability and avoid collision. The final grasp performed in the experiment is shown in Fig. 8b. The grasp was regarded as success when the object could be robustly lifted up at least 10 cm without slipping off from the gripper. The success rate, average computation time and average point numbers for each category of objects are provided in Table II. The experimental video can be found at [11]. The snapshots of grasping experiments are shown in Fig. 8.

Although the shapes and configurations of target objects were different to the ones of source objects, they shared the similar structures. Therefore, the grasp poses on the source



(a) Manipulator (b) Wrench (c) Plier (d) Humanoid (e) Cup (f) Adapter

Fig. 5. Grasp examples: the first row shows one of the grasp pose on each source object, and the second row provides the snapshots of the actual demonstrated grasp poses.

Manipulator – 1	20	0	0	0	0	0
Wrench – 2	0	19	0	0	0	1
Plier – 3	0	2	18	0	0	0
Humanoid – 4	2	0	0	18	0	0
Cup – 5	0	0	0	0	20	0
Adapter – 6	0	2	0	0	0	18
	1	2	3	4	5	6

Fig. 7. The confusion matrix of object classification. Each column represents a predicted class, and each row represents a actual class.

object could be transferred to reasonable locations on the target objects. For instance, the grasp poses on the various toy manipulator models were invariant in terms of topological structures (see the first row of Fig. 8). The grasp poses taught by kinesthetic teaching had the intuition from human such as the task specific consideration and fairly good grasping quality, and CPD transferred the insight to the target objects. Therefore, the test can be successful in most of the cases.

The failure case happened when there was a very large distortion to transform the source object to the target, which degraded the accuracy of the transformation estimated by CPD. As a result, the grasp pose was not accurately transformed, which caused the grasp failed. Although CPD did not transfer grasp poses with high accuracy in this situation, it provided a relatively close one. In the future, we may include an adaptation on the warped grasp pose to avoid this failure.

VI. Conclusions and Future Works

This paper proposed a framework for efficient grasp generation by combining analytic approach with learning from demonstration. A database containing multiple categories of source objects with demonstrated grasp poses were constructed by human experts. During the test scenario, a novel object was firstly classified into one of the example categories by measuring its dissimilarity to each source object. Then the grasp poses on the most similar source object were transferred to the novel object by the coherent point drift (CPD) method. All the transferred grasp poses were evaluated and sorted by the grasp isotropy metric. The selected pose was further refined by an orientation search mechanism, which improves the robot reachability and avoids collision. A series of experiments were performed to grasp six categories of objects with various shapes, sizes and configurations. The average success rate was 18.8 out of 20 grasp trials. The experimental video is available at [11].

There are several directions to further improve this work in the future. Since the CPD method might fail to transfer the grasp pose to the novel objects in large distortion cases, we consider introducing a parameter adaptation mechanism in CPD and a searching mechanism around the transferred poses to deal with large deformation. In addition, feature extraction will be conducted to categorize a novel object and improve the scalability of the current approach.

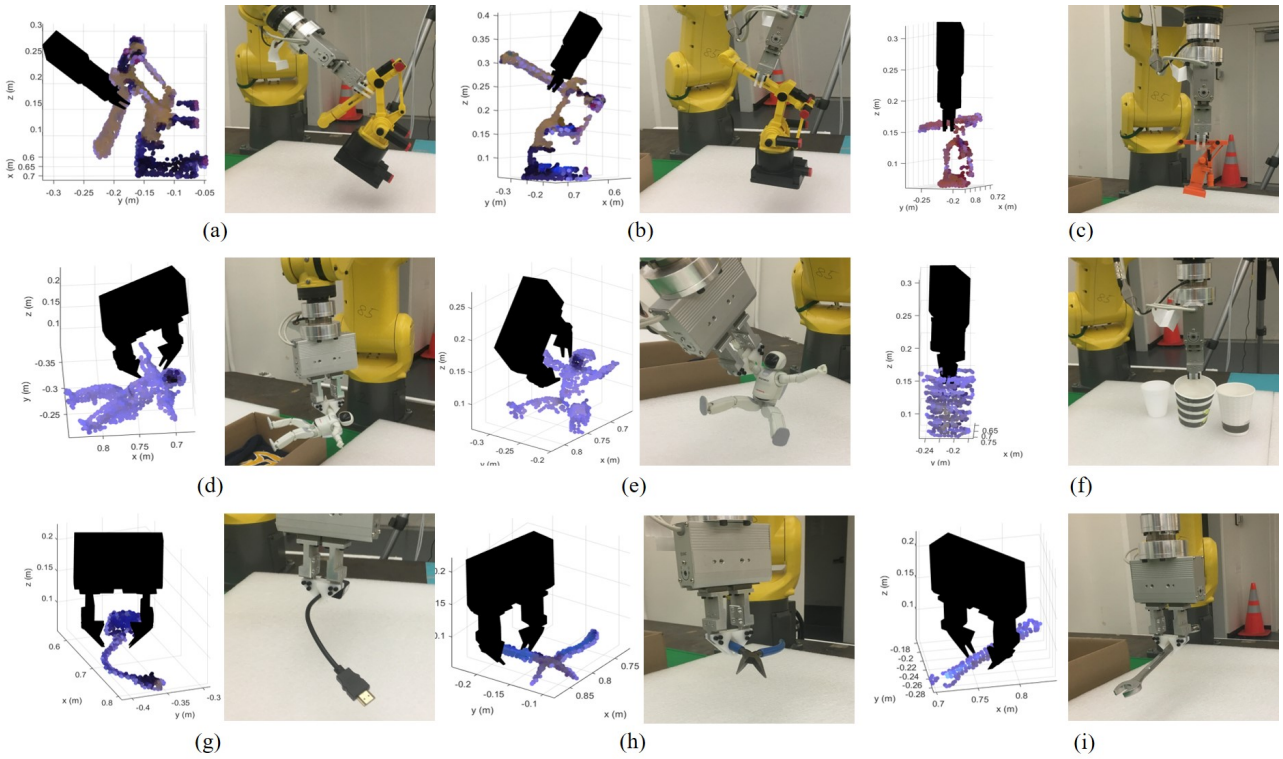


Fig. 8. The planned grasp poses and the corresponding snapshots of the grasping results.

ACKNOWLEDGMENT

The work is supported by National Science Foundation under Grant No. 1734109.

REFERENCES

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

[2] B.-H. Kim, S.-R. Oh, B.-J. Yi, and I. H. Suh, "Optimal grasping based on non-dimensionalized performance indices," in *Intelligent Robots and Systems, 2001. Proceedings. 2001 IEEE/RSJ International Conference on*, vol. 2. IEEE, 2001, pp. 949–956.

[3] 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.

[4] J. Mahler, F. T. Pokorny, B. Hou, M. Roderick, M. Laskey, M. Aubry, K. Kohlhoff, T. Kröger, J. Kuffner, and K. Goldberg, "Dex-net 1.0: A cloud-based network of 3d objects for robust grasp planning using a multi-armed bandit model with correlated rewards," in *Robotics and Automation (ICRA), 2016 IEEE International Conference on*. IEEE, 2016, pp. 1957–1964.

[5] J. Varley, J. Weisz, J. Weiss, and P. Allen, "Generating multi-fingered robotic grasps via deep learning," in *Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on*. IEEE, 2015, pp. 4415–4420.

[6] P. Brook, M. Ciocarlie, and K. Hsiao, "Collaborative grasp planning with multiple object representations," in *Robotics and Automation (ICRA), 2011 IEEE International Conference on*. IEEE, 2011, pp. 2851–2858.

[7] H. Dang and P. K. Allen, "Semantic grasping: Planning robotic grasps functionally suitable for an object manipulation task," in *Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on*. IEEE, 2012, pp. 1311–1317.

[8] R. Detry, C. H. Ek, M. Madry, J. Piater, and D. Kragic, "Generalizing grasps across partly similar objects," in *Robotics and Automation (ICRA), 2012 IEEE International Conference on*. IEEE, 2012, pp. 3791–3797.

[9] A. Myronenko and X. Song, "Point set registration: Coherent point drift," *IEEE transactions on pattern analysis and machine intelligence*, vol. 32, no. 12, pp. 2262–2275, 2010.

[10] A. Myronenko, X. Song, and M. A. Carreira-Perpinán, "Non-rigid point set registration: Coherent point drift," in *Advances in Neural Information Processing Systems*, 2007, pp. 1009–1016.

[11] Experimental Videos for Grasping Objects by a FANUC robot, http://me.berkeley.edu/%7Ehclin/IROS2018/GraspByCPD.html.

[12] M. Kuczma, *An introduction to the theory of functional equations and inequalities: Cauchy's equation and Jensen's inequality*. Springer Science & Business Media, 2009.

[13] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the em algorithm," *Journal of the royal statistical society. Series B (methodological)*, pp. 1–38, 1977.

[14] F. Girosi, M. Jones, and T. Poggio, "Regularization theory and neural networks architectures," *Neural computation*, vol. 7, no. 2, pp. 219–269, 1995.

[15] R. Abraham, J. E. Marsden, and J. E. Marsden, *Foundations of mechanics*. Benjamin/Cummings Publishing Company Reading, Massachusetts, 1978, vol. 36.

[16] C. Papazov, S. Haddadin, S. Parusel, K. Krieger, and D. Burschka, "Rigid 3d geometry matching for grasping of known objects in cluttered scenes," *The International Journal of Robotics Research*, vol. 31, no. 4, pp. 538–553, 2012.

[17] U. Klank, D. Pangercic, R. B. Rusu, and M. Beetz, "Real-time cad model matching for mobile manipulation and grasping," in *Humanoid Robots, 2009. Humanoids 2009. 9th IEEE-RAS International Conference on*. IEEE, 2009, pp. 290–296.

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

[19] M. Ester, H.-P. Kriegel, J. Sander, X. Xu *et al.*, "A density-based algorithm for discovering clusters in large spatial databases with noise." in *Kdd*, vol. 96, no. 34, 1996, pp. 226–231.