

An effective method for grasp planning on objects with complex geometry combining human experience and analytical approach

Chunfang LIU^{1*}, Fuchun SUN¹ & Xiaojuan BAN²

¹*Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China;*

²*Department of Computer Science and Technology, University of Science and Technology Beijing, Beijing 100083, China*

Received November 28, 2015; accepted March 11, 2016; published online October 17, 2016

Abstract In this paper, an effective method for identifying the graspable components of objects with complex geometry is proposed for grasp planning based on human experience. Instead of focusing on individual objects, our method identifies graspable components on the category level under the assumption that geometrically alike objects share similar graspable components. Firstly, employing a modified SHOT descriptor, a fast KNN-based method is developed for object categorization. Then, the graspable components are identified by adopting a learning framework based on human experience. Afterwards, a fast analytical grasp planning method is proposed which comprises of contact points extraction and hand kinematics calculation. Finally, a regression model based on the extreme learning method (ELM) is built which inputs the desired contact points and the wrist orientation and outputs the wrist position. This approach is time-saving comparing with the optimization method. The simulations and experiments demonstrate the effectiveness of the proposed approach by realizing grasps on the graspable components of human choice for objects with complex geometry.

Keywords grasp planning, human experience, analytical method, kinematics learning, ELM

Citation Liu C F, Sun F C, Ban X J. An effective method for grasp planning on objects with complex geometry combining human experience and analytical approach. *Sci China Inf Sci*, 2016, 59(11): 112212, doi: 10.1007/s11432-015-0463-9

1 Introduction

Humans are able to manipulate objects with various geometries [1–4]. One of the important reason is that humans can identify graspable components on the objects for special tasks and this identification is generalized on objects with similar shapes. Actually, many everyday objects have been designed with components for manipulation, e.g. mug handle, screwdriver handle. This paper aims to perform recognizing the graspable components on objects with different geometries by learning humans' experience and accomplish hand configuration by the kinematics way and the empirical way.

Object shape representation is the premise for graspable component identification. In [5], the authors represent objects by superquadrics and the graspable components are predicted using a neural network to

*Corresponding author (email: cfliu1985@gmail.com)

learn the human preference. Ref. [6] also use superquadrics for modeling objects. Ref. [7] approximates objects by simple object primitives. However, superquadrics or shape primitives are only suitable for describing simple object surfaces and they are susceptible to noise.

Recently, many 3D descriptors [8–11] have been developed with the development of depth camera (Kinect). They have desired surface matching ability for complex shapes of objects. A comparative study of descriptors in [12] demonstrates that the Signature of Histograms of Orientations (SHOT) performs the best in object classification in terms of accuracy and speed trade-off. However, the data size of the SHOT descriptor for representing an object are too large to directly using for real-time object classification. Therefore, in this paper, the original SHOT descriptor is modified for objects' classification. Then, the graspable components are learned on objects' category level for generalizing the identification of graspable components on the objects in the same category.

As for the problem of “how to grasp the object”, both analytical and empirical approaches are popular ways for hand configuration. The analytical approach considers the stability of contact points and hand kinematics. Generally, an iterative procedure is necessary for acquiring the optimum solution. Ref. [13] proposes an analytical method which extracts contact points from 3D point cloud of an object and achieves hand configuration by inverse kinematics. The advantage of this approach is that it is able to deal with objects of arbitrary geometry and avoids computationally expensive kinematics analysis in the process of iterative contact point search, which makes this method suitable for real-time grasping. However, in order to extract the desired contact points, this method uses a blind search on all of the object clouds, which is time consuming and not task-oriented. The empirical approach uses grasp experience to build a grasp model. In [14], the shape parameters of some superquadric geometries and grasp configuration terms from different positions and directions are trained to build an SVM (Support Vector Machine) regressor. Then, the grasps for new superquadric are found by regression. In [15], a Gaussian Mixture Model (GMM) is adopted for learning the distribution of stable hand postures for a specific object. However, the model is built on the individual object level that the learned grasp model cannot be directly generalized to grasp other objects. In addition, the grasp model also is blind and not task-oriented.

It is regarded in this paper that “blind” searching or “blind” building grasp model is exhausted and not efficient for real-time task manipulation. Therefore, firstly, an effective method is developed in this paper for identifying the graspable component based on objects' shape category. Then, a fast grasp posture planning method is proposed by considering the grasp stability, hand shape, hand size and hand kinematics. Further, in order to empirically perform hand configuration, the multi-output regression of extreme learning method (ELM) is utilized for building the relation among the desired contact points, the wrist position and orientation.

The paper is laid out as follows. In Section 2, an overview of the proposed approach is displayed. Section 3 illustrates the modification of the SHOT descriptor in detail. Based on this modified SHOT descriptor, Section 4 develops the object classification method and the graspable component identification approach. Section 5 proposes the analytical grasp planning method composed of desired triplet extraction and kinematics optimization. Moreover, in this section, a regressor of extreme learning method (ELM) is used to represent the experience of hand configuration. Section 6 demonstrates the results of the proposed approach. Finally, the conclusions are summarized in Section 7.

2 Our approach

In this paper, an effective grasp planning method is proposed by combining analytical and empirical approaches. Comparing with other object model based methods [5, 16, 17], this proposed approach is able to represent and grasp a variety of objects with complex shapes. The contributions of this paper can be summarized as follows:

Firstly, under the belief that many objects have handles designed for manipulating and accomplishing tasks, the proposed method does not search the graspable points ‘blind’ on all of objects' body, but identifies the graspable components and searches the grasp points in the range of the graspable component.

In addition, the representations of objects and graspable components are non-parametric and therefore have strong capability of expressing object geometry, especially in the limit of infinite number of interest points.

Secondly, instead of identifying the graspable components individually or defining the object handle as some shapes such as cylindrical, the graspable component is learned on the category level of objects by human experience based on the assumption that geometry and functionality alike objects have similar graspable components. It is important to note that the word “category” here refers to geometrical properties (shape but not color) rather than actual names of objects.

Thirdly, a fast analytical method is developed for extracting the desired point clouds and hand kinematics optimization. Further, a regressor of extreme learning method (ELM) is built for representing the experience of hand configuration among the contact points, the wrist position and orientation.

Figure 1 illustrates the proposed approach for recognizing the graspable component diagrammatically. For simplicity, only two categories of objects are shown.

Above of the dash line in Figure 1 demonstrates the off-line training stage and below of the dash line demonstrates the on-line testing stage. The training data set contains object point clouds with the graspable components labeled by a human subject. At the training stage, a suitable 3D point descriptor are firstly applied to represent the shapes of objects. Afterwards, it is divided as two pathes: in one path (the dot dash line), these descriptors are used directly to train a graspable component identifier for each category; in another path (the solid line), the dimensionality of these descriptors are reduced to form object features which are then used to train an object classifier.

At the testing stage, the testing point cloud of an unknown object goes through the same feature extraction steps. The object is classified to a known category and at the same time, the predicted class leads the object descriptors to the corresponding graspable component.

In Figure 1, “IP feats”, “IP desc”, “obj feats”, “dim.reduct” and “GCI” are the abbreviations of “interest point features”, “interest point description”, “object features”, “dimension reduction” and “graspable component identification”, respectively.

Then, as shown in Figure 2, the desired contact points are extracted from the graspable component, which satisfy a set of stability and feasibility criterion. Finally, a fast hand configuration approach utilizing kinematics is developed for planning the wrist position, orientation and joint angles.

Moreover, in Figure 3, an empirical method for hand configuration is proposed based on a multi-output regressor of extreme leaning method (ELM). Specifically, inputting the extracted graspable points and the wrist orientation experience, the regressor outputs the wrist position. This approach avoids the iteration of kinematics optimization.

3 Object shape representation

We employ a modified version of the Signature of Histograms of Orientations (SHOT) [18] descriptor to encode object geometry at randomly selected interest points in the point cloud. The original algorithm creates a robust local reference at each interest points based on surface normal and the two directions that define the tangent plane. Under each local reference frame, an isotropic spherical grid centered on the sampled point divides its neighborhood space into partitions along the radial, azimuth and elevation axes. The SHOT descriptor of an interest point is a 352-dimensional vector that represents the 11-bin histograms of $\cos(\theta)$ at each partition, where θ is the angle between the surface normals at the interest point and another point in its neighborhood. It has been shown that SHOT performs the best compared with many others in terms of the trade-off between accuracy and runtime [12].

In the modification, the scale (the size) of the given point cloud d is defined as the maximum distance between two points in the cloud. Each local reference frame is constructed using points that are within $0.05d$ from this interest point. To construct the SHOT, the neighborhood range is set to be within $0.5d$ from each interest point, thus this long-range SHOT contains much more information about where this interest point lies relative to the entire object as well as the surface variations within its neighborhood.

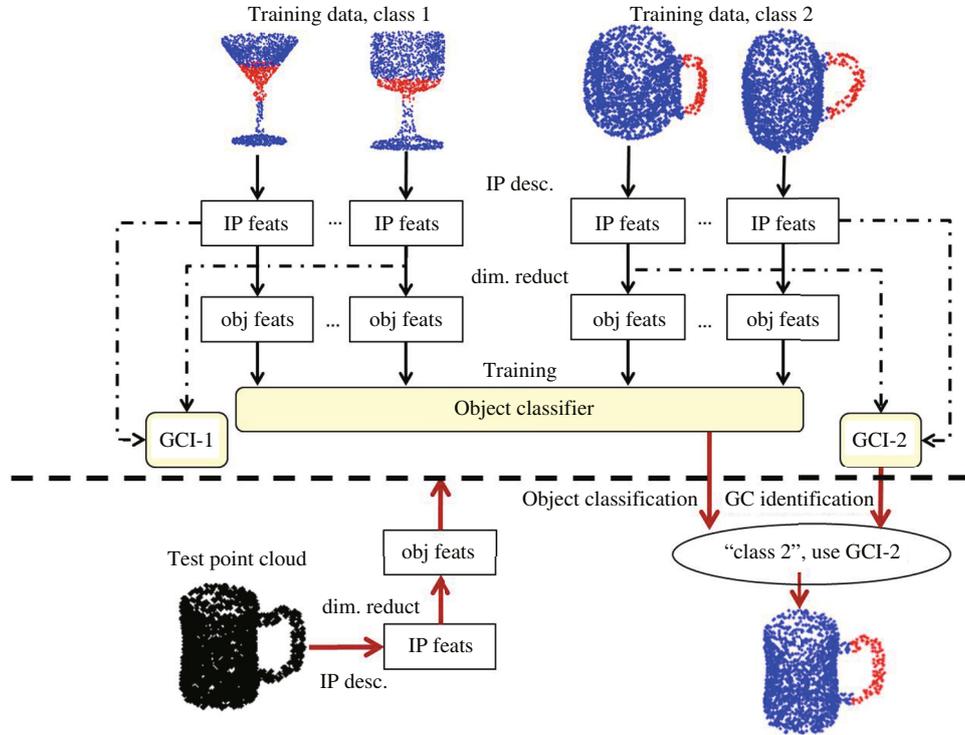


Figure 1 The flowchart for the identification of the graspable components.

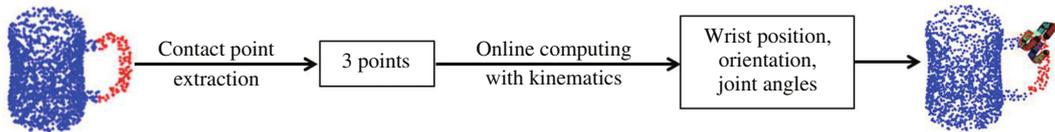


Figure 2 The flowchart of hand configuration.

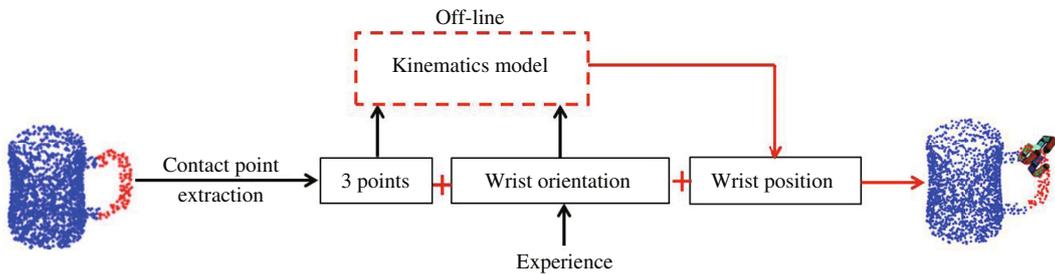


Figure 3 The flowchart for learning the kinematics model.

Using this adaptive range also makes the descriptor invariant to the scale. With the much down-sampled point cloud, speed can still be ensured despite the significantly increased radius. It is found that 2000 randomly selected interest points are sufficient to achieve plausible results in object classification and graspable component identification. The number of partitions in the neighborhood space is also changed so that the final SHOT descriptor is a 594 dimensional vector after setting the number of spatial partitions (6, 3, 3 in azimuth, elevation and radial dimensions).

In addition, the local reference at each interest point is used as features, resulting in a 603-dimensional vector for each interest point. Hence, the entire object is described by a 2000×603 matrix $M = [S, L]$, where S is the 2000×594 SHOT features of this point cloud, and L is the 2000×9 LR features. In the following sections, we refer to the column-wise dimension (length 2000) of M as the point dimension, and the row-wise dimension (length 603) of M as the feature dimension.

4 Object classification and graspable component identification

The original features collection \mathbf{M} is too large for either classification or storage. In our work, we treat the 594-dimensional SHOT features \mathbf{S} and the 9-dimensional local reference features \mathbf{L} separately and reduce the dimensionality along the point dimension.

SHOT features. Let $s_{i,j}$ ($i = 1, 2, \dots, 2000$, $j = 1, 2, \dots, 594$) be the elements of \mathbf{S} , and define $\mathbf{S} = \{\mathbf{s}_j | j = 1, 2, \dots, 594\}$, where $\mathbf{s}_j = \{s_{i,j} | i = 1, 2, \dots, 2000\}$. We reduce the dimensionality along the point dimension by extracting the range, mean, standard deviation and three orders (0.5, 1 and 3) of Rényi entropies $H_\alpha(\bar{\mathbf{s}}_j)$ on the normalized \mathbf{s}_j using (1),

$$H_\alpha(\bar{\mathbf{s}}_j) = \frac{1}{1-\alpha} \log \left(\sum_i \bar{s}_{i,j}^\alpha \right). \quad (1)$$

These simple statistics (range, mean, standard deviation and entropies) capture the variation at positions along each feature dimension. Further, to capture correlations across SHOT dimensions, the first five principle components of all the 2000 SHOT features (eigenvectors with the largest five eigenvalues of $\mathbf{S}^T \mathbf{S}$) are also extracted, giving an extra five 594-dimensional vectors. \mathbf{S} is now reduced to a 11×594 matrix $\tilde{\mathbf{S}}$.

Local reference features. Only the standard deviations along the point dimension and three principle components are extracted as local reference features. This gives four 9-dimensional features which capture the surface variation of the object. Denote this 4×9 matrix by $\tilde{\mathbf{L}}$.

In this paper, we use 15 K-nearest-neighbors (KNN) classifiers constructed on the 11 rows of $\tilde{\mathbf{S}}$ and 4 rows of $\tilde{\mathbf{L}}$ extracted from each point cloud and adopt a voting scheme for final prediction. Each of the 15 KNN classifiers produces K nearest neighbors to the new object and the neighbors' labels are counted as votes, giving $15K$ votes in total. The category with the highest vote is predicted for this point cloud. The multiple KNN method shares some similarities with the Random KNN [19] which has been used to perform feature selection on high dimensional data [20]. Its advantages in accuracy as well as training and testing speeds have been found in the object classification task.

The next step is to identify the graspable component based on human experience on the category level. For each category, we directly train an SVM (Support Vector Machine) classifier (with quadratic kernel) which takes the SHOT features and returns binary decision on whether an interest point belongs to the graspable component. These training data accumulated as objects within a category are processed for the graspable component classification. Under the assumption that graspable points are closely packed together, a spatial median filter is used to smooth the labels over the object for eliminating isolated misclassification. Specifically, the predicted class label l_p for each interest point p are modified using

$$\bar{l}_p = \mathbf{1} \left\{ \frac{1}{|\mathbf{N}_p|} \sum_{p \in \mathbf{N}_p} l_p \geq 0.5 \right\}, \quad (2)$$

where $\mathbf{1}\{\cdot\}$ is the indicator function and is equal to 1 if its argument is true and 0 otherwise. \mathbf{N}_p is the set of neighboring interest point within a small radius.

5 Three-finger grasping on the graspable component

This section develops grasp planning method carried out on the identified graspable component. It mainly consists of two stages: desired contact point selection and hand configuration. The robotic hand used is a BarrettHand with three fingers. The proposed approach directly searches the contact point candidates on the 3D point clouds by employing the DPSO method. It does not need to fit the shapes of the graspable component. Therefore, this method for graspable point extraction is suitable for objects with a various of shapes.

Two approaches are developed for hand configuration, one way is based on kinematics optimization; and another way is based on a multi-output regressor which learns the kinematics experience.

5.1 Desired triplet extraction

Given a set of graspable points \mathbf{P}_g , we aim at finding a triplet $\text{PN} = (\{\mathbf{p}_i, \mathbf{n}_i\}, \{\mathbf{p}_j, \mathbf{n}_j\}, \{\mathbf{p}_k, \mathbf{n}_k\})$, $\mathbf{p}_i, \mathbf{p}_j, \mathbf{p}_k \in \mathbf{P}_g$ that satisfies the properties of grasp stability, hand shape and hand size. Here, \mathbf{n}_i expresses the normal vector in the point \mathbf{p}_i .

Grasp stability. Rather than estimating the contact wrenches, a simple method is adopted for force-closure grasping analysis of three-finger robot hand, which should satisfy the following sufficient and necessary conditions:

- The grasp matrix must be full-rank.
- Each friction cone intersects the contact plane defined by the three contact points, generating two unit vectors that bound the intersection of the cone and this contact plane.
- The six unit vectors at the three contact points construct a 2D force-closure grasp in the contact plane.

A large cost is assigned if any of these conditions is not satisfied, which motivates the following cost function

$$\text{FC}(\text{PN}) = D \times \mathbf{1}\{\text{PN is force closure}\}, \quad (3)$$

where D is a large positive number and a hard finger model is used in the contact points.

Grasp shape. In [21], according to the number of counter-overlap pairs of friction cone, they demonstrates four types of force-closure grasp shapes for hard-finger contacts. It has been shown that the grasp shape is more likely to be stable when the number of counter-overlapping friction cone pairs is 1 or 2 [13]. Therefore, the following cost function is developed:

$$C(\text{PN}) = \begin{cases} 0, & \text{if } n_{\text{co}} \text{ is 1 or 2,} \\ \min(\text{An}), & \text{if } n_{\text{co}} \text{ is 0,} \\ 2 \arctan(\mu) - \max(\text{An}), & \text{if } n_{\text{co}} \text{ is 3,} \end{cases} \quad (4)$$

where n_{co} is the number of counter-overlapping cone pairs and $\text{An} = \{an_1, an_2, an_3\}$ is the set of the angles of the triangle formed by the three points in PN.

Area of the grasp polygon. The force-closure triplet should be reachable by the fingertips. Therefore, the following cost function is used to satisfy BarrettHand's geometry requirement:

$$A(\text{PN}) = \begin{cases} 0, & \text{if } a_{\min} \leq \text{Area}(\text{PN}) \leq a_{\max}, \\ B, & \text{if } \text{Area}(\text{PN}) > a_{\max} \text{ or } \text{Area}(\text{PN}) < a_{\min}, \end{cases} \quad (5)$$

where $A(\text{PN})$ is the area of the triangle defined by three points in PN, and a_{\min} and a_{\max} are the minimum and maximum triangular areas spanned by the fingertips respectively. B is a large positive value.

Referring to the above grasp properties, a good triplet PN^* can be found by solving the optimization problem formulated as

$$\text{PN}^* = \min_{\text{PN}} [wC(\text{PN}) + A(\text{PN}) + \text{FC}(\text{PN})], \quad (6)$$

where $C(\text{PN})$, $A(\text{PN})$ and $\text{FC}(\text{PN})$ are given by (4), (5) and (3), respectively. w is the weight to determine the trade-off among $C(\text{PN})$, $A(\text{PN})$ and $\text{FC}(\text{PN})$. Empirically, we test several values of w , e.g. $w = 0, 0.1, 0.2, \dots, 10$, and $w = 0.8$ is found to be the most suitable one.

Then, the above objective function is minimized by Discrete Particle Swarm Optimization (DPSO), which is a discrete, derivative-free, population based local optimization algorithm [22].

5.2 Hand configuration

For a given triplet PN, the hand kinematics can be used to find a feasible configuration of the robot hand. The BH8-series BarrettHand has 10 freedoms in total, 3 for the wrist position \mathbf{p}_e , 3 for the wrist orientation \mathbf{o}_e and 4 for finger joint angles \mathbf{q}_e (1 for each finger and 1 for the spread). Thus, the optimum hand configuration is a numerical solution iteratively generated from the kinematics optimization. In

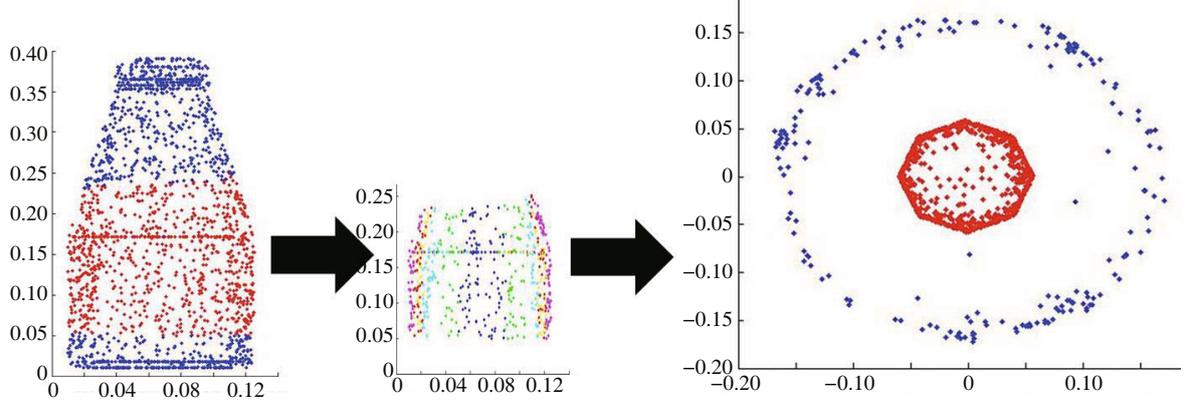


Figure 4 Successful hand configuration.

this paper, an interior-point algorithm is applied for minimizing the distance from the fingertips to the desired contact points PN,

$$\begin{aligned} \mathbf{p}_e^*, \mathbf{o}_e^*, \mathbf{q}_e^* &= \min_{\mathbf{p}_e, \mathbf{o}_e, \mathbf{q}_e} \sum_{m=1}^3 \|\text{Tp}_m - \text{Sp}_m\| \\ \text{s.t. } &(\mathbf{p}_e - \bar{\mathbf{p}})^T \mathbf{D}^{-1} (\mathbf{p}_e - \bar{\mathbf{p}}) > 1, \\ &l_n < q_n < u_n, \quad \text{for } n = 1, 2, 3, 4, \end{aligned} \quad (7)$$

where $\bar{\mathbf{p}}$ and \mathbf{D} are the mean and covariance matrix of the entire point cloud respectively, and $\text{Tp}_m \in \mathbb{R}^3$ is the position of a finger tip. The finger joint angle q_n is within the lower and upper bounds l_n and u_n .

Based on successful hand configurations, in this section, a kinematics model also is built by using the multi-output regression of extreme learning method (ELM). The model constructs the kinematics relation among \mathbf{o}_e , \mathbf{p}_e and PN, which includes two steps:

The first step is to collect multiple groups of successful hand configurations $(\mathbf{p}_e^*, \mathbf{o}_e^*, \text{PN})_i, i = 1, \dots, L$. To do that, the initial triplets are selected from different point clouds of the graspable components. For example, as shown in Figure 4, the initial triplets are chosen from different color parts of the graspable point clouds. The number of the initial triplets is around 30–40.

In Figure 4, the first left figure is the point cloud of a bottle on which the red clouds are the graspable component. In the middle figure, the graspable component is divided as multiple parts (shown in different colors) and the initial triplets are selected from them. Subsequently, multiple desired groups of $(\mathbf{p}_e^*, \mathbf{o}_e^*, \text{PN})_i, i = 1, 2, \dots, L$ can be acquired by using the stable grasp evaluation in Subsection 5.1 and the optimization approach in (7). The right figure is a top view of the graspable component, around which the blue color clouds are the positions of the wrist \mathbf{p}_e^* .

Then, the relation among \mathbf{p}_e^* , \mathbf{o}_e^* , PN is built based on the multi-output regression of extreme learning method (ELM). ELM (extreme learning machine) [23, 24] is an effective learning algorithm of Single-hidden Layer Feedforward Neural Network (SLFNs). Comparing with SVM, its advantages include fewer optimization constraints, faster learning and better performance. The output function of ELM is

$$f_L(\mathbf{x}) = \sum_{i=1}^L \beta_i h_i(\mathbf{x}), \quad (8)$$

where $\beta = [\beta_1, \dots, \beta_L]^T$ is the vector of the output weights between the hidden layer of L nodes and the output node. $h(\mathbf{x})$ is the output vector of the hidden layer with respect to the input \mathbf{x} . The extreme learning machine (ELM) is to minimize the training error as well as the norm of the output weights. In this paper, the multi-output regression of ELM performs modeling the following relations:

$$\hat{\mathbf{P}}_e^* = \mathbf{F}_m(\mathbf{x}), \quad (9)$$

where $\mathbf{x} = \{(\mathbf{o}_e^*, \text{PN})_i\}$, $\hat{\mathbf{P}}_e^* = \{(\hat{\mathbf{p}}_e^*)_i\}, i = 1, \dots, L$. Therefore, given the desired contact point PN and the wrist orientation \mathbf{o}_e^* , the regression function \mathbf{F}_m can be utilized for estimating the wrist position $\hat{\mathbf{p}}_e^*$.

Table 1 Accuracy means and standard deviations of object classification using MKNN and random forest (RandFst)

Classifier	Accuracy means and standard deviations (%)	
	PSB	L-RGBD
MKNN	85.3 ± 3.2	93.4 ± 2.1
RandFst	76.1 ± 5.2	89.4 ± 2.2

Comparing with the analytical optimization method, this approach avoids the problem of going into local minimum which leads to a fail of hand configuration.

6 Experiments and results

6.1 Object classification and graspable component identification

6.1.1 Object classification

The classification algorithm is tested on two publicly available datasets independently. They are PSB (Princeton Shape Benchmark) [25] containing *complete* point cloud data and Large RGBD dataset [26] containing *partial* point clouds recorded by Kinect depth camera. 10 categories of objects are chosen from each dataset, which have dissimilar overall geometries (ball, coffee mug, food bag, food box, food jar, glue stick, plate, pliers, stapler and water bottle). Six different objects are selected from each category of which four randomly selected objects are utilized for training and the point clouds from the other two objects form the test set.

We compare MKNN (multiple KNN) with the random forest (RandFst) classifier on both datasets. The accuracy means and standard deviations of 10 trials are reported in Table 1. It shows that MKNN performs better than RandFst in this task on both datasets. In fact, the MKNN consistently gives higher accuracies than random forest on all trials. The high accuracy from the L-RGBD dataset suggests that this object classification method can be applied to real robotic grasping scenarios where the depth camera only records partial point clouds.

6.1.2 Graspable component identification

The same set of objects chosen from PSB (Princeton Shape Benchmark) were used in testing graspable component identification. The interest points of each point cloud are labeled by a single human subject to be either graspable (1) or non-graspable (0), thus the labels represent the grasp position experience of this subject. The results in Figure 5 show that predicted graspable components agree with human experience very well. For pistols, as an example, the handle is clearly identified while leaving points at the bottom end as non-graspable. This method works just as well even for complex graspable components, such as airplanes which are difficult to be identified using shape modeling.

The proposed approach also is verified on incomplete point clouds acquired directly from Kinect RGB-D camera. Figure 6 demonstrates the experimental results about brushes and bottles. The data on the left of the black dash line are the training set on which the red color clouds are labeled by a single human subject. The testing results are demonstrated on the right of the dash line and the red color clouds on the testing data are the predicted graspable components. It can be seen that even if the testing data are different from the training ones, the proposed approach still can identify the graspable regions.

6.2 Hand configuration

Using the results in the above subsection, the grasp planning only needs to search for the best triplet on the identified graspable component. In Subsection 5.2, two approaches for hand configuration are developed including the kinematics optimization method and the model learning approach. They are verified by simulations and experiments in this subsection. The hand model used is built from the real

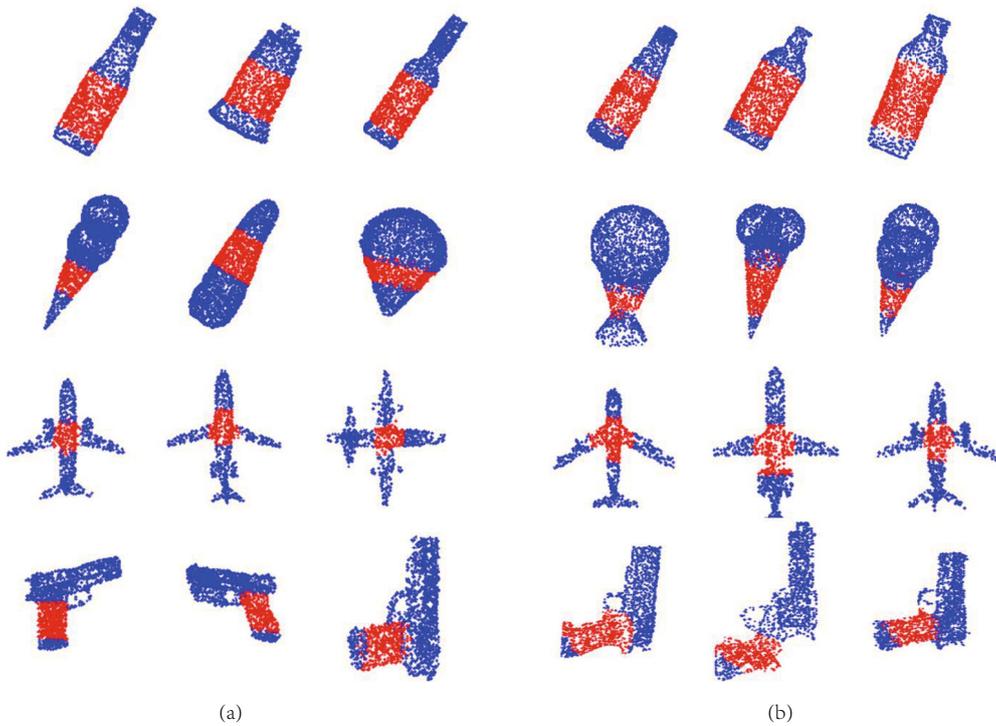


Figure 5 Graspable component identification. (a) The training data; (b) the testing data.

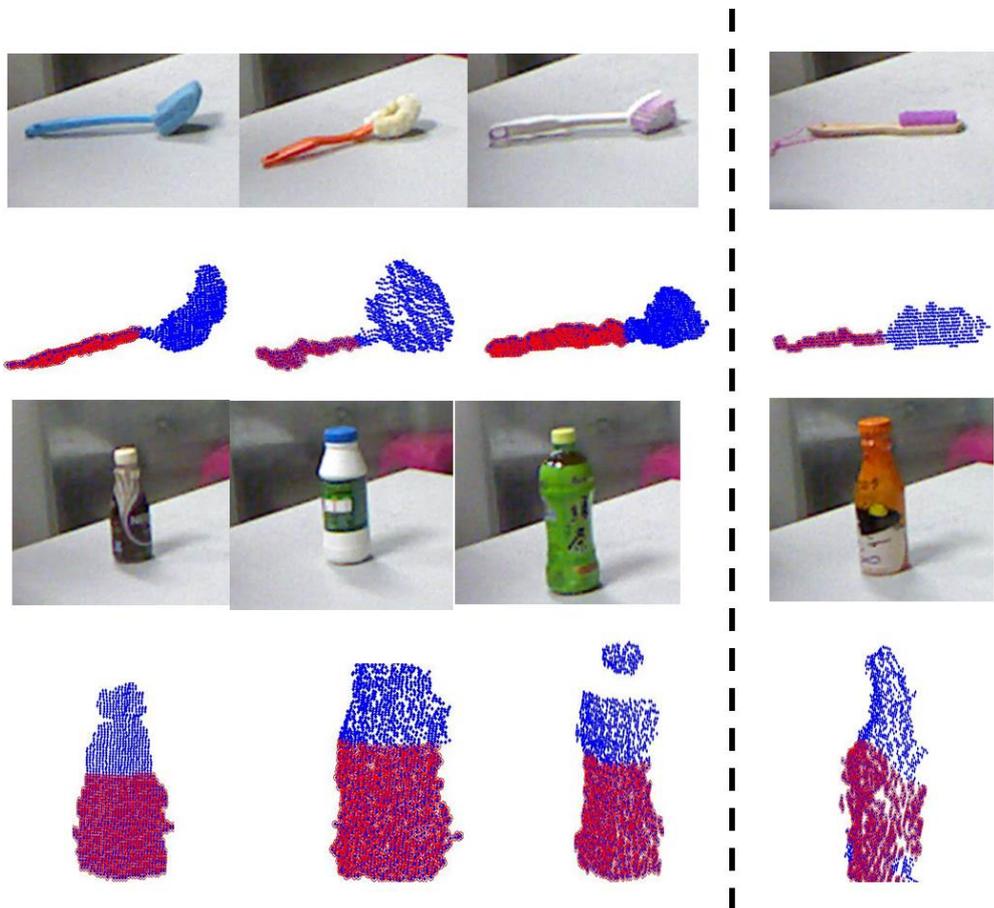


Figure 6 Graspable component identification on partial clouds.

BarrettHand by modifying the three-finger hand available in the SynGrasp package [27]. The friction coefficient μ is set to 0.6.

It has been found that with the DPSO method, the initial triplets could quickly converge to the stable grasp triplets in most cases about 10 iterations. Then, the optimization method is performed for acquiring the wrist position, orientation and the configuration of hand joints (See Figure 7). The success configuration is defined that the 3 fingertips can reach the positions of the best triplet precisely and the fingers do not penetrate the object. Figure 7 demonstrates that the desired triplet can be found and stable grasps can be established on the objects with complex geometry. In Figure 8, this optimization approach is tested on incomplete clouds. The mug clouds on the above of the black dash line are the training set selected from the Large RGBD dataset. The mug handles are labeled for distinguishing with the bodies of the mugs. Then, under the black dash line, a new mug is tested for the graspable component identification and hand configuration. Considering the size of the mug handle, the BarrettHand selects to grasp the body of the mug. Figure 9 displays an experiment of grasping the mug which includes three processes from left to right: “open”, “grasp” and “upward move”. It shows the importance of the graspable component identification and precise hand configuration for the success of real grasp experiments.

The learned kinematics model is a multi-output regressor of ELM working by inputting the wrist orientation and the desired contact points and outputting the wrist position. In order to verify the accuracy of the model, we construct the model by three numbers of group data, 60, 120 and 180, respectively and test the models with 20 groups of data. The prediction results are displayed in Table 2.

From Table 2, it can be seen clearly that the prediction accuracy becomes higher with the increase of the group number. When the training number reaches 180 groups, the mean testing error reduces to 0.0105 m. The 20 groups of testing results are further utilized to grasp the object, which show that the number of fail grasp cases are 3 and 0 for the regressors trained by 120 and 180 groups, respectively. One of the success grasps is demonstrated in Figure 10. We can see that the predicted one is much similar as the optimized hand postures. In summary, the regression method is time-saving that it does not need to iteratively estimate the optimum hand configuration. Besides of the desired contact points, the wrist orientation constraint also are required by the regression method. Therefore, it is a potential way for combining with human experience.

7 Conclusion and future work

Many objects are designed with a component suitable for human grasping. It is assumed that geometrically alike objects have similar graspable components and grasps should be designed with respect to the geometrical category of a given object. In this work, human experience is used directly to assist robotic grasping on the graspable components, which is achieved on a wide range of objects without any assumptions about the shapes or relative positions of these graspable components. It also does not require segmentation or parametrization of the sub-components.

The proposed grasping system identifies the graspable component before hand configuration. It first uses a fast multiple KNN algorithm on a modified SHOT object descriptor to categorize a given object. Afterwards, an SVM classifier trained on SHOT features for the predicted category is used to tell which of the points on this object belongs to the graspable component perceived by a human subject. Fast grasp planning is then accomplished by extracting the desired contact points from the graspable component and applying hand kinematics to grasp onto the selected points. The grasp results are shown to resemble human grasps very well on objects whose geometries are complex and are difficult to represent for many object modeling techniques.

Moreover, a kinematics model with respect to the desired contact points, the wrist orientation and the wrist position are built by employing the multi-output regression of ELM (extreme learning method). This approach is timesaving because of no iteration and is suitable to fuse with the human experience.

In the future work, human experience will be utilized to guide the determination of the desired triplet and the wrist orientation.

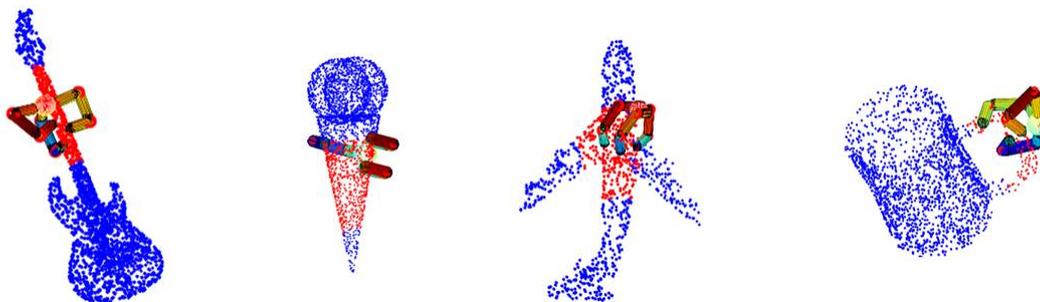


Figure 7 Grasp on the graspable component.

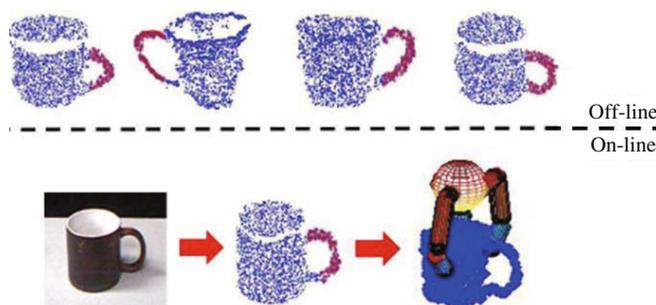


Figure 8 Graspable component identification by kinematics optimization method.

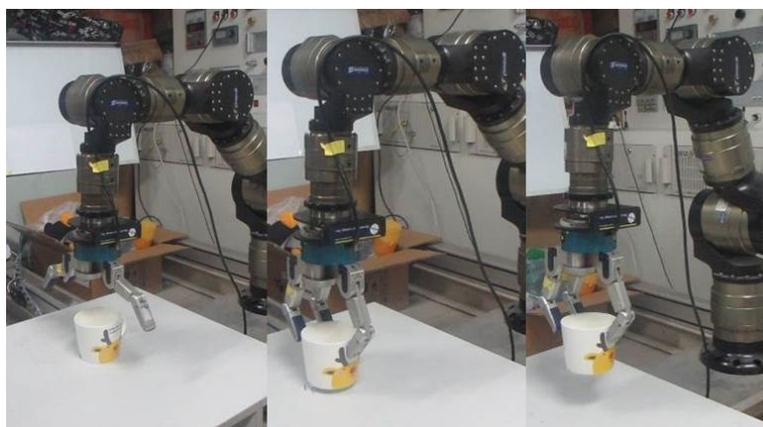


Figure 9 Real experiment for grasp planning.

Table 2 The mean error of the predicted wrist position (Unit: m)

60 groups	120 groups	180 groups
0.0376	0.0202	0.0105

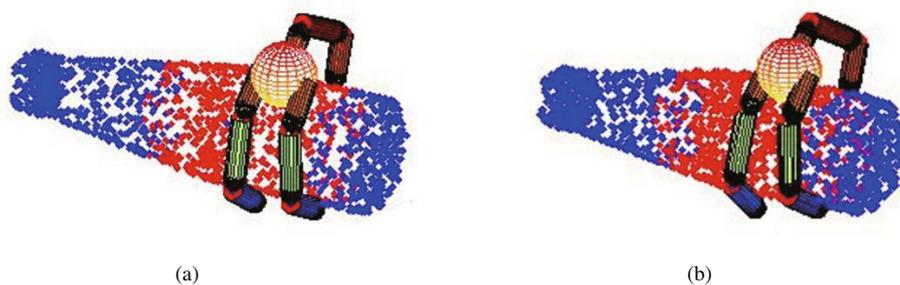


Figure 10 The predicted and the optimized hand posture. (a) The optimized hand posture; (b) the predicted hand posture.

Acknowledgements This work was supported by National Natural Science Foundation of China (Grant Nos. 61210013, 61327809, 91420302, 91520201).

Conflict of interest The authors declare that they have no conflict of interest.

References

- 1 Lin G D, Li Z J, Liu L, et al. Development of multi-fingered dexterous hand for grasping manipulation. *Sci China Inf Sci*, 2014, 57: 120208
- 2 Li Z J, Deng S M, Su C Y, et al. Decentralized adaptive control of cooperating mobile manipulators with disturbance observers. *IET Control Theory Appl*, 2014, 8: 515–521
- 3 Li Z J, Ge S S, Liu S B. Contact-force distribution optimization and control for quadruped robots using both gradient and adaptive neural networks. *IEEE Trans Neural Netw Learn Syst*, 2014, 25: 1460–1473
- 4 Li Z J, Xiao S T, Ge S S, et al. Constrained multilegged robot system modeling and fuzzy control with uncertain kinematics and dynamics incorporating foot force. *IEEE Trans Syst Man Cybern-Syst*, 2015, 99: 1–14
- 5 El-Khoury S, Sahbani A. A new strategy combining empirical and analytical approaches for grasping unknown 3d objects. *Robot Auton Syst*, 2010, 58: 497–507
- 6 Guo D, Sun F C, Liu C F. A system of robotic grasping with experience acquisition. *Sci China Inf Sci*, 2014, 57: 120202
- 7 Miller A T, Knoop S, Christensen H I, et al. Automatic grasp planning using shape primitives. In: *Proceedings of IEEE International Conference on Robotics and Automation*, Taipei, 2003. 1824–1829
- 8 Novotni M, Klein R. 3d zernike descriptors for content based shape retrieval. In: *Proceedings of ACM Symposium on Solid Modeling and Applications*, Washington, 2003. 216–225
- 9 Johnson A E, Hebert M. Using spin images for efficient object recognition in cluttered 3d scenes. *IEEE Trans Patt Anal Mach Intell*, 1999, 21: 433–449
- 10 Körtgen M, Park G J, Novotni M, et al. 3d shape matching with 3d shape contexts. In: *Proceedings of the 7th Central European Seminar on Computer Graphics*, Budmerice, 2003
- 11 Bohg J, Kragic D. Learning grasping points with shape context. *Robot Auton Syst*, 2010, 58: 362–377
- 12 Alexandre L A. 3D descriptors for object and category recognition: a comparative evaluation. In: *Proceedings of IEEE International Conference on Intelligent Robots and Systems*, Vilamoura, 2012. 1–6
- 13 Gori I, Pattacini U, Tikhonoff V, et al. Three-finger precision grasp on incomplete 3d point clouds. In: *Proceedings of IEEE International Conference on Robotics and Automation*, Hong Kong, 2014. 5366–5373
- 14 Pelosoff R, Miller A, Allen P, et al. An SVM learning approach to robotic grasping. In: *Proceedings of IEEE International Conference on Robotics and Automation*, New Orleans, 2004. 3512–3518
- 15 Huang B D, El-Khoury S, Li M, et al. Learning a real time grasping strategy. In: *Proceedings of IEEE International Conference on Robotics and Automation*, Karlsruhe, 2013. 593–600
- 16 Aleotti J, Caselli S. A 3d shape segmentation approach for robot grasping by parts. *Robot Auton Syst*, 2012, 60: 358–366
- 17 Hübner K, Kragic D. Grasping by parts: robot grasp generation from 3d box primitives. In: *Proceedings of IEEE International Conference on Cognitive Systems*, ETH Zurich, 2010
- 18 Tombari F, Salti S, Stefano L. Unique signatures of histograms for local surface description. In: *Proceedings of European Conference on Computer Vision*. Berlin: Springer, 2010. 356–369
- 19 Chen X, Ishwaran H. Random forests for genomic data analysis. *Genomics*, 2012, 99: 323–329
- 20 Park C H, Kim S B. Sequential random k-nearest neighbor feature selection for high-dimensional data. *Expert Syst Appl*, 2015, 42: 2336–2342
- 21 Nguyen V D. Constructing force-closure grasps. *Int J Robot Res*, 1988, 7: 240–245
- 22 Kennedy J, Eberhart R. Particle swarm optimization. In: *Proceedings of International Conference on Neural Networks*, Perth, 1995. 1942–1948
- 23 Huang G B, Zhou H M, Ding X J, et al. Extreme learning machine for regression and multiclass classification. *IEEE Trans Syst Man Cybern Part B-Cybern*, 2012, 42: 513–529
- 24 Huang G B, Zhu Q Y, Siew C K. Extreme learning machine: theory and applications. *Neurocomputing*, 2006, 70: 489–501
- 25 Grana C, Davolio M, Cucchiara R. Similarity-based retrieval with mpeg-7 3d descriptors: performance evaluation on the princeton shape benchmark. In: *Proceedings of the 1st International DELOS Conference*, Pisa, 2007. 308–317
- 26 Lai K, Bo L F, Ren X F, et al. A large-scale hierarchical multi-view RGB-D object dataset. In: *Proceedings of IEEE International Conference on Robotics and Automation*, Shanghai, 2011. 1817–1824
- 27 Malvezzi M, Gioioso G, Salvietti G, et al. Syngrasp: a MATLAB toolbox for grasp analysis of human and robotic hands. In: *Proceedings of IEEE International Conference on Robotics and Automation*, Karlsruhe, 2013. 1088–1093