Knowledge-Augmented Dexterous Grasping with Incomplete Sensing

Bharath Rao, Hui Li, Krishna Krishnan, Enkhsaikhan Boldsaikhan, and Hongsheng He*

Abstract—Humans can determine a proper strategy to grasp an object according to the measured physical attributes or the prior knowledge of the object. This paper proposes an approach to determining the strategy of dexterous grasping by using an anthropomorphic robotic hand simply based on a label or a description of an object. Object attributes are parsed from natural-language descriptions and augmented with an object knowledge base that is scraped from retailer websites. A novel metric named joint probability distance is defined to measure distance between object attributes. The probability distribution of grasp types for the given object is learned using a deep neural network which takes in object features as input. The action of the multi-fingered hand with redundant degrees of freedom (DoF) is controlled by a linear inverse-kinematics model of grasp topology and scales. The grasping strategy generated by the proposed approach is evaluated both by simulation and execution on a Sawyer robot with an AR10 robotic hand.

Index Terms—robotic grasping, human grasp primitives, natural language processing, object features extraction, blind grasping

I. INTRODUCTION

Dexterous grasping is critical for complex assembly and delicate tool handling in industrial automation and advanced manufacturing [1]. Dexterous robotic manipulation relies on comprehensive and precise measurement of the work context [2], which is usually impractical and expensive for industrial applications. It is even challenging to measure important parameters of objects for grasping in an online mode, such as 3D dimensions, materials, and weights. It is therefore beneficial to research an approach to plan dexterous grasping without complete or accurate sensing of object characteristics.

A five-digit hand configuration with an opposable thumb is considered to be one of the most important natural selection that contributed to human evolutionary success. There are considerable application scenarios where a dexterous robotic hand could be invaluable such as disaster struck areas where the robot may have to interact with unfamiliar environment. It has been an elusive goal to enable a robot to master the human-level grasping skills. Recent studies on robotic grasping have focused on two or three-fingered robotic clamps [3], [4]. The research of grasping planning for an anthropomorphic robotic hand is more challenging and deserves more effort [5], [6].

Behind this simple task of grasping an object, the human brain is executing a series of sub-tasks with the associated decision-making and error-correction process in real time. The brain selects and executes appropriate motor strategy learned earlier by the human sensorimotor apparatus. These learned strategies, also called as action-phase controllers [7], utilize the input sensory signals and corresponding predictions by the nervous system to produce motor commands to accomplish the given motor task. The action-phase controllers accurately estimate the specific motor output required using the information about an object’s physical properties and the current configuration of the hand. The inspiration for this study is the human grasping mechanism - not only from the sensorimotor task of human grasping, but also the learning process itself. This study is therefore, an exploration in emulating a part of the human brain’s action-phase controller model to accomplish grasping using a humanoid robotic hand.

In this paper, we propose an approach to emulating human grasping strategies without complete sensing of object attributes, as shown in Fig. 1. By using labels and descriptions, object attributes are retrieved and augmented from a knowledge base that is scraped from online webpages. The extended object attributes include dimension, mass, shape, texture, fragility, material, and stiffness. We design a neural-network model to learn human grasping strategies for target objects with various physical attributes. The optimal grasping strategy is deployed to the anthropomorphic robotic hand by a multi-constrained inverse kinematics of grasping topology and scales.

Human hands have 20 degrees-of-freedom (DoF) each (not including the wrist joint), thousands of mechanoreceptors [7] and therefore, a significant amount of the brain’s resources are dedicated to grasping tasks. Understanding human grasps is not a trivial task. Most of the efforts in understanding grasps [8], [9] have been to breakdown the human grasping behavior...
Human grasping is a complex process with disproportionately large portion of the human sensorimotor apparatus dedicated to grasping. Therefore, it is no surprise that robotic grasping is a complex and as yet unsolved problem. The contribution of this paper is to further the knowledge and understanding of human grasping in the context of emulating human type grasps on a five-fingered robotic hand. In order to demonstrate this idea, a set of everyday objects are chosen to train the robot, to impart the knowledge and experience that is needed for it to succeed at grasping. Multiple learning models with novel concepts are developed and validated in the course of exploring the five-fingered robotic grasping problem. The models are tested through simulations and experiments with the physical robot. The major contributions of the paper include:

1) The paper addresses the challenge of acquiring object attributes without complete sensing. A distance metric is proposed to query most similar objects in the developed knowledge base by using simple object descriptions.

2) The paper designs a neural-network model to imitate human abilities in applying optimal grasping strategies for dexterous grasping. A well designed grasping strategies with grasping topology allows dexterous grasping of objects without precise attribute information.

II. OBJECT AFFORDANCE ACQUISITION FROM KNOWLEDGE BASE

It is generally challenging to measure complete object attributes in industrial applications, including precise 3D dimension, materials, rigidity, and textures, but object categories can be recognized by machine learning algorithms. In addition, it is straightforward to describe important object attributes in natural language. We therefore developed a knowledge base of object attributes by mining online object information from retailer websites. By referring to the knowledge base, we can acquire extended object attributes by object labels or short descriptions for the selection of optimal grasping strategies. In this section, we define the dominant object attributes for dexterous grasping, design the parsing algorithm for key attributes from natural-language descriptions, and propose a novel distance metric for attribute acquisition from the knowledge base.

A. Dominant Object Attributes in Grasping

Human grasp strategies depend on numerous factors including object shape, size, weight, texture, stiffness and sometimes fragility, temperature, wetness \(^{[25]}\). With our goal being able to understand these object features by parsing natural-language...
descriptions, we had to be parsimonious in our choice of object attributes. Based on findings coming out of previous studies, the following set of features, shown in Table I were prioritized for data collection. These physical attributes of the object significantly influence grasping decisions.

Table I  PRIMARY OBJECT ATTRIBUTES IN GRASPING.

| Feature | Description | Value Range |
|---------|-------------|-------------|
| (a, b, c) | Dimensions along orthogonal directions | (a, b, c) ∈ R³ s.t. a ≥ b ≥ c |
| m | Mass | m ∈ R |
| s | Shape classification | thin, compact, prism, long, radial |
| r | Rigidity of the object | rigid, squeezable, floppy |
| t | Texture | medium, smooth, rough |
| fr | Fragility | sturdy, medium, fragile |
| mt | Simplified material types | fabric, glass, metal, paper, plastic, rubber, wood, other |

Enormous object descriptions are available in the internet particularly in retailer webpages such as Amazon and Walmart. The object descriptions typically include object dimensions, weight, and materials. We developed a web scraper to collect object information (the source code is available online at https://github.com/hhelium). The web scraper downloads product description webpages and uses pattern search to discover object attributes. The knowledge base consisted of dimension measurements, mass, rigidity, material, texture, fragility, and shape classifications. The missing attributes for some objects are manually labeled and annotated. Examples of the objects and attributes in the knowledge base are shown in Table I. Given object labels or short descriptions l will be mapped to the feature set f corresponding to the described object or a similar object

\[ l \rightarrow f = [a, b, c, m, s, mt, r, t, fr] \] (1)

where the meanings of the attributes are defined in Table I. We will discuss the description parsing and mapping in the following sections.

B. Parsing Object Descriptions

The problem to address in parsing object descriptions is to estimate significant object features that determine object categories. At the same time, the method is required to be resilient to missing, partial or incorrect descriptions. Object descriptions may specify object details such as its approximate dimensions, e.g., “it is about ten centimeters long”, or materials, e.g., “it is made of plastic”. Though not accurate or specific, the object descriptions are informative when the object’s descriptive and quantitative information is contained. To specifically address free-form descriptions where object object features are described in any format or order, we designed a natural-language parser as shown in Algorithm 1. The natural-language statements are cleaned up by lemmatizing and removal of stop words. Each word in the statements is then tagged with parts of speech (POS) labels based on standardized tags from the Penn Treebank project [26]. Examples of POS Tag Nomenclature include JJ-Adjective, IN-preposition or conjunction, CD-Cardinal digit, CC-coordinating conjunction, NN-noun, and RB-adverb. Of special interest to our study, are any available quantitative and qualitative descriptors of the object(s). We look for expressions such as “two centimeters long”, “made of plastic” or “very rough” using regular expressions as shown in the algorithm.

Algorithm 1  Object description parsing.

Input: Object description string: ObjDescription; Input: Regular Expression: regex = \{<JJ.?>*<IN>*<CD.?>*<NN.?>*<RB.?>*<CC.?>*<CD.?>*<NN.?>*<JJ.?>*<NN.?>*<NN.?>*<JJ.?>*<NN.?>*<NN.?>*<JJ.?>*<NN.?>*\} Output: Array of dimensions of the object: l = [a, b, c, m, ...]

WordToks ← tokenize(ObjDescription)
POSToks ← ApplyPartsofSpeechTokens(WordToks)
PhraseTree ← PerformChunking(POSToks, regex)
for each Chunk in PhraseTree do
  if QuantitativeDescriptor in Chunk then
    l₁ ← ParseToNumber(Chunk)
  end if
  if QualitativeDescriptor in Chunk then
    l₁ ← ParseToCategorical(Chunk)
  end if
  if null in l₁ then
    l₁ ← DataImputation(l₁)
  end if
end for

The extracted phrases (or chunks) lead us to individual feature descriptors. When there are more than one dimensional descriptor, the largest value is assigned to feature a, the smallest to feature c and the intermediate one to feature b. One of the challenges is that if we do not have all feature descriptors then we have null values. One example is when the object has a radial symmetry, it is described only by diameter. For such cases, we perform data imputation using a rule-based approach of estimating the missing dimension based on the other available dimensions of the object. The rule itself was derived from the priors in the data. The success of this model is evaluated by scoring the parsed values with the measured or labeled values and the scores are used to improve the algorithm.

C. Object Knowledge Acquisition

In addition to the basic object attributes in the description, we desire to acquire more features from the knowledge base. Even with a reasonably detailed elucidation of an object’s features, object descriptions tend to be either imprecise or incomplete. For example, the general tendency is to round-off dimensions and mass, often missing to mention certain features such as material type or texture. A human can still work with the information available only because of the recall of having seen or held such an object. The curated knowledge
base of objects and their physical features emulates human memory.

Let \( f(o) = [f_1, f_2, f_3, \ldots, f_m] \) represent \( m \) features corresponding to an \( i^{th} \) object \( o_i \) from this dataset where \( i \in [1, N] \) and \( f_i \in \{a, b, c, m, s, nt, r, t, f, r\} \). The features parsed from descriptions of a reference object \( o_r \) is represented by \( f(o) \). It should be noted that \( f(o) \) could possibly have empty values for some of the elements due to incomplete object descriptions. The problem is to find the most similar object to the reference object in the knowledge base \( f^* = \arg \min_i \|o_i - o_r\| \). By identifying an object closely matching the description of the reference object, we continue to retain the ability of choosing the most suitable grasp because the reference object, most likely, has physical features very similar to the object chosen by the algorithm.

Several distance metrics, such as Euclidian, Minkowski and Cosine distances, have been developed to measure the proximity of vectors in n-dimensional space \( \mathbb{R}^n \). To increase the identification accuracy and the confidence of object match, it is necessary to not only ensure the proximity of the points in the normed vector space, but also to ensure proximity of each individual feature. To that end, we calculate the probability of the \( i^{th} \) object being mapped to the reference object given the distance \( |f_j(o_i) - f_j(o_r)| \) of the \( j^{th} \) feature. The overall probability of mapping \( i^{th} \) object to the reference object is the joint probability over all the features of the object. This approach ensures that only that object which matches the reference object’s every individual available features is the one that results in a highest probability value. The proposed distance metric, named Joint Probability Distance, is defined as

\[
\|o_i - o_r\| = \sum_{j=1}^{m} \ln(1 + |f_j(o_i) - f_j(o_r)|) \tag{2}
\]

One of the advantages of using this distance metric is that the probability of the match decays at a much faster rate as each of the features deviate from the reference. This helps in non-linearly increasing the distance of unlikely candidates and filtering out the unlikely matches with more confidence.

This distance metric can be used on a dataset with contains a combination of continuous and categorical (transformed to one-hot binary encoding) features without any need for data normalization. For example, material of an object is a categorical feature with eight possible text values. The Material property feature can be easily converted to eight features with one-hot encoding. This distance metric will work well with such categorical variables. The accuracy of distance metrics including Euclidean, Minkowski, Cosine, K-D tree, and Joint Probability was compared and results are shown in Fig.2. The proposed Joint Probability metric achieved the best recall accuracy for the knowledge base.

### III. HUMAN-LIKE DEXTEROUS GRASPING

The grasp of a multi-finger robotic hand defines a set of angles of the finger joints, and the magnitude of the contact forces applied by the fingers and palm to an object at the contact points. The objective here is to emulate human dexterous grasping by mapping object features \( f(o) \) referred in the knowledge base onto a grasp prioritization. The referred object features may be inaccurate or even erroneous due to rough measurement and description. We therefore implement grasp strategies in terms of grasp topology and scales, which enable imprecise measure of object dimensions and location, thus improving system robustness and adaptivity. In this section, we define dexterous grasp taxonomy and implement grasping strategies on a robotic hand by extending our prior work [29].
A. Grasp Definition

Grasps can be defined by the finger joint angles for a humanoid robotic hand with multiple fingers. As shown in Fig. 3, the AR10 robotic hand used in this work has ten DoF and limited force feedback. A grasp \( G \in \mathbb{R}^{10} \) can therefore be defined in the 10-dimensional joint configuration space as

\[
G(t) = [\theta_1(t), \theta_2(t), \ldots, \theta_{10}(t)]
\]

where each joint angle \( \theta_i \) can continuously vary over the operating range of the servos to result in infinite grasp patterns. Owing to the redundant DoF, there may be multiple feasible grasp strategies associated with one object, so it is intractable to design a deterministic grasping model for various objects. We discretize the configuration space and map configuration space into a space with reduced dimensionality by

\[
G = g(h, \alpha)
\]

where \( h \in h_1, h_2, \ldots, h_k \) represents human grasp topology and \( \alpha \) is the scales that determine the completion of the grasp. Each grasp topology \( h_k = [\theta_{k1}, \theta_{k2}, \ldots, \theta_{k10}] \) is a unique combination of joint angles representing one of the human grasps with \( \theta_{kj} \) chosen such that \( h_k \) mimics a particular human grasp type from the grasp taxonomy. The grasp topology \( h \) spans the entire configuration space, and a grasp can be represent by

\[
G(t) = \alpha(t).h
\]

whereby a range of grasps can be defined by human grasp topology \( h \) and the time-variant completion scale \( \alpha(t) \). We will learn a mapping between object features and grasp topology, and implement the completion scale by inverse kinematics.

B. Dexterous Grasp Taxonomy

Robotic grasps can be made effective by emulating human grasps, and we desire to choose a suitable taxonomy of human grasps. Although comprehensive grasp taxonomy is available, we decided to adopt the Grasp Taxonomy presented by Cutkosky [8] for this work. Even within this taxonomy, we have restricted it to the six higher level classifications, because the finer adjustments can be obtained by combining these six grasp types with the scalar. The chosen human grasp classification and the nomenclature for each grasp is shown in Fig. 4 where the prefixes ‘w’ and ‘r’ stand for Power and Precision grasps [8], [25].

C. Learning Grasping Strategies from Human Knowledge

Most studies [7], [8], [25] attempting to understand and codify human grasps have concluded that human grasp choice is a function of object affordances (geometry, texture, etc.) and the task requirements (forces, mobility, etc.). Attempts
to assign one most suitable grasp for a given object-task combination have not been conclusive. The major problem is that even for one specific object-task combination, there are multiple grasp choices possible, which often appear to be arbitrary and not amenable for deterministic modeling. Human grasp choices nevertheless do tend to cluster when studied over a large set of objects. Both the clustering effect and the confusion between grasp types can be seen in the data presented by [2], which shows that a single object could be held in multiple different grasp types in the course of picking or handling. There is no one-to-one mapping of one object to one grasp type.

The problem of grasp selection is therefore not selecting one ideal grasp type but one of the many feasible grasp types in human grasp taxonomy for the given context. To that end, we plan to learn the mapping from features \( f \) into grasp topology distributions

\[
f \rightarrow P(H|f) = [P(rc.ab|f), P(rc.bc|f), \ldots] \quad (10)
\]

We designed a neural network to model the probability distribution over all grasp classes \( P(H|f) \), as illustrated in Fig. [8]. The network is designed with cross-entropy loss and optimized using stochastic gradient decent algorithms. The loss function is defined by cross entropy that measures the deviation between the ground truth and predicted probability distribution

\[
L(P(H|f), \hat{P}(H|f)) = \sum_i \sum_j P(H_j|f_i) \log \hat{P}(H_j|f_i) \quad (11)
\]

where \( i \in [1, N] \) with \( N \) as the number of observations and \( j \) is the index of grasp topology.

The grasp with the maximum probability is chosen by

\[
\hat{H}_{\text{max}} = \arg \max_j \hat{P}(H_j|f^*) \quad (12)
\]

where \( f^* \) is the acquired object features. The predicted grasp configuration \( \hat{H}_{\text{max}} \) contains information regarding the grasp type and object dimension along which the grasp can be executed, so \( \hat{H}_{\text{max}} \) can be easily decomposed into grasp type \( h^* \) and grasp dimension \( d^* \), which can be used subsequently to calculate robot hand configuration. The optimal grasp type is chosen as the one corresponding to the highest probability from the predicted probability distribution.

Because the model predicts probability distributions, we defined two scoring metrics for training and evaluation of the model. The predicted grasp choice is scored as a success if the same grasp type was chosen at least once in the human-knowledge database. The feasibility of the grasp is scored as

\[
F_l(P(H), \hat{P}(H)) = \begin{cases} 1 & P(\hat{H}_{\text{max}}) > 0 \\ 0 & P(\hat{H}_{\text{max}}) = 0 \end{cases} \quad (13)
\]

where

\[
\hat{H}_{\text{max}} = \arg \max_j \hat{P}(H_j|f) \quad (14)
\]

is the grasp topology with the maximal probability, and \( H \) is defined in [9]. The feasibility score \( F_l \) is representative of the ability of the algorithm to pick a feasible grasp for a given object. The match score metric \( F_m \) is defined as

\[
F_m(P(H), \hat{P}(H)) = \begin{cases} 1 & P(\hat{H}_{\text{max}}) = P(H_{\text{max}}) \\ 0 & P(\hat{H}_{\text{max}}) \neq P(H_{\text{max}}) \end{cases} \quad (15)
\]

This match score is representative of the ability of algorithm to predict the most frequently applied human grasp as the grasp with the highest probability for a given object. In other words, \( F_m \) is akin to the accuracy, had this grasp learning problem been treated as a multi-class single label classification problem. This metric is much more stringent and therefore we can expect the match score \( F_m \) to be always lower than feasibility score \( F_l \)

\[
F_m(P(H), \hat{P}(H)) \leq F_l(P(H), \hat{P}(H)) \quad (16)
\]

We used the feasibility score as the primary scoring metric, for the objective is to find one feasible grasp that can be successfully executed by a robot.

**D. Deploying Grasp Strategies**

Grasp types are determined by grasp topology, and the grasp size for a particular grasp type corresponds to object dimensions. The grasp size \( d_{vf} \) is essentially the distance between the virtual fingers of a particular grasp type \( h^* \) [8]. The grasp size \( d_{vf} \) can be computed or estimated from object dimensions based on geometric relations, as illustrated in Fig. [5]. The grasp types and sizes are implemented by inverse kinematics of the hand and fingers.

We developed a multi-constrained inverse kinematics of the robotic hand to deploy grasp topology and finger closure [30]. The multi-constrained inverse kinematics enables multi-point planning of each finger in the process of hand closure and grasping. We considered two levels of kinematic constraints in grasp strategy implementation: high-priority and low-priority constraints. The distance between virtual fingers (finger tip closure) meets high-priority constraints on distal phalanges, and the trajectories of the middle and proximal phalanges satisfy low-priority constraints. The inverse kinematics transforms the trajectory of points on fingers to angular joint velocities.

The trajectory of \( N \) finger tips (distal phalanges) \( \alpha(t) : [0, T] \rightarrow [0, 1]^N \) is planned for \( d_{vf} \), which corresponds to the object dimensions. Not all objects have regular geometric shapes and the calculated \( d_{vf} \) may deviate from the actual size on such objects, as shown in Fig. [7]. We adopted a straight-line path

\[
d(\alpha) = d_{\text{start}} + \alpha_i(d_{vf} - d_{\text{start}}) \quad (17)
\]
where $\alpha_\text{i}$ is the time scaling for the $i$th finger, and $d_{\text{start}}$ is the starting distance between the virtual fingers. The closure of the fingers are controlled following $\alpha(t)$. The AR10 Robotic hand has 10 DOF with 9 actuators controlling the fingers and one controlling the opposing thumb action. The robotic hand has limited and inaccurate force measurement by the Force Sensitive Resistors (FSR) that are attached to each finger. The closing of the fingers stops until sufficient contact forces are measured.

\[ dvf \approx 2\text{cm} \]
\[ dvf \approx 3\text{cm} \]
\[ dvf \approx 4\text{cm} \]

Figure 7. Illustration of variations in virtual finger distance for the same grasp type: prismatic grasp (rp).

In the experiments shown in Fig. 8, we found the distance between virtual fingers $dvf$ can be linearly approximated by object dimensions $do$ within a limited range of motion

\[ dvf = w_1 do + w_0 \]  
(18)

We learn the parameters $w_1$ and $w_0$ by generating data on the physical robotic arm and fitting a linear model between $do$ and $dvf$. For each grasp type, we varied the joint angles, measured the distance between virtual fingers $dvf$, and fit a linear regression model to learn the parameters by minimizing least square errors.

![Figure 8. The relation between virtual finger distances and object dimensions.](image)

IV. EXPERIMENTS

We conducted the experiments to evaluate the performance of object affordance acquisition, grasp strategy selection, and robot grasping. Leveraging existing grasping database [31], we collected more data for experiment and training, including object attributes, short natural-language descriptions, and human grasp strategies. The developed methods include natural-language parsing algorithms, models for grasping strategy learning, object identification, and robot grasping. We validate the methods and quantify the performance on collected data by executing the grasps physically on a robot with unfamiliar objects.

A. Experiment Setup

The anthropomorphic robotic hand integrates an AR10 robotic hand, a Rethink Sawyer robot (with an in-arm camera), and an Intel RealSense RGB-D camera installed on the wrist, as illustrated in Fig. 9. The robotic hand has limited force sensing through the Force Sensing Resistors (FSR) attached to finger tips, so we focused on hand configurations and planning and used the force sensors to examine contact conditions. The goal of the experiment was to grasp an object, which was described by a label or short incomplete description, in a proper manner to the object affordance. We seamlessly integrated the models and modules in ROS Python libraries that control the AR10 and Sawyer robots. The experiment is set up as follows:

1) To emulate industrial scenarios with incomplete sensing, we only used the camera to locate and identify an object. The objects or tools for experiment are placed on the table with a fixed initial pose. We utilized the localization and recognition algorithms developed in our prior work [32]–[34]. The coordinates of the object on the table is back projected to the robot frame.

2) We retrieved object affordances based on optional object descriptions, e.g., “A scientific calculator with plastic body. It is about fifteen and half centimeters long, 8 centimeters wide and appears to be more than one and half centimeters thick.” The recognized object labels are adopted when an object description is unavailable.

![Figure 9. Robot Setup: Sawyer robotic arm with an AR10 robotic hand and in-hand sensors.](image)

B. Data Collection

We followed the design of the human grasping database [31] and extended the database by collecting grasping samples. Given the ergonomics of human grasps, objects were selected to capture sufficient variations in objects features and grasping strategies (samples are shown in Fig. 10). Short descriptions of the objects are provided along with object labels, covering shape, dimension, mass, rigidity, and texture if available.
Grasping strategies depend on object affordances and tasks. In this paper, the task definition is restricted to securely holding and supporting an object in midair. The intent (task) of the grasp and the object's position and orientation may influence the grasp strategies; however, we kept these variables constant but focused on object features during the experiment. There could be multiple ways of grasping for most objects, so the experiment was designed to replicate potential grasping multiple times. At each attempt, the subjects were encouraged to try alternate ways of grasping the object while ensuring the comfort and security of the grasp. At the end of this experiment, we had a frequency distribution of human preferred grasp types. Sample data is shown in Table III. It is inevitable that the optimal grasp labels are subjective to some extent due to personal preferences and background. There are indeed multiple feasible or optimal strategies for one scenario.

C. Object Affordance Acquisition

The effectiveness of the natural-language parses was scored and validated by the Ordinary Least Squares Regression model. The final model was able to fit with an $R^2$ of 0.98 overall for all dimensions and 0.87 for mass estimations. The regression fit for dimension estimations is shown in Fig. [11]. The primary source of errors in this step are the approximated dimensions in descriptions. This results in larger percentage deviations when describing smaller dimensions. Categorical labels for material, shape and rigidity were also scored, where the scoring matrix for material classification is shown in Fig. [11]. The objects were classified under “other” when description details are insufficient. The main source of errors in material recognition is omitted features from the description. This behavior stems from preconceived assumptions that such features are very obvious and do not require specific mention. For example, when the object is made of plastic, the subjects tend to omit any mentions of the object’s stiffness, assuming plastic objects are rigid.

Object affordances, including shapes, sizes, weights, and textures, lead to specific grasping strategy [25]. We desired to prioritize the influence of the features on grasping strategies. Such a prioritization would help the learning of grasping strategies and the design of perception algorithms. To that end, we used a recursive feature elimination (RFE) method along with a Random Forest classifier to rank the features [35]. Starting from the most significant factor, the ranking is dimension, shape, mass, texture, material, stiffness, and fragility. The result is generally in line with prior studies. It was interesting to note that the most important dimension was the intermediate dimension $b$ followed by the shortest dimension $c$ and then by the longest dimension $a$. For most objects, especially larger ones, humans tend to grasp it along the shorter dimension primarily because of comfort of holding.

Object features parsed from object descriptions are used to acquire full features of similar objects from the collected object database. We performed contextual search based on the proposed distance metric, and validate the performance of the methodology. The parsed object features are commonly incomplete and inaccurate because of approximation errors, missing information, and parsing errors. We used the subset features to query and match the object being described to a similar object in the database. The test results of acquisition of 100 random objects are plotted in Fig. [12]. As the confusion matrix show, the overall accuracy was 92%. The errors primarily stem from incomplete or insufficient information, similarity to multiple objects, and wrong descriptions. The approach demonstrated robustness in object acquisition under uncertainty, when partial
object features were incorrect. Despite the confusion, it was interesting to note that the objects chosen by the algorithm were physically identical to the objects being described. A mis-match is therefore not necessarily detrimental to the grasp strategy. It is still possible to identify the correct grasp with the wrong object as long as the object is physically similar to the target.

D. Robotic Grasping

We developed the neural-network model to learn grasping strategies as proposed in Sect. III-C and optimized the model in terms of cross-entropy. The input of the model is the acquired object features, and the output is the grasping strategies corresponding to human preference and knowledge. The grasping strategies were represented by normalized probability distributions. The results of grasping strategy determination with scores are reported and compared to the ground truth in Fig. 13. The feasibility score $F_1$ of the model is 100%, which was defined as the hit rate of the predicted grasp strategy in all human preferred grasps. The experiment shows that the model’s capability in picking feasible (human validated) grasping. The match-score $F_m$, on the other hand, measures the accuracy of the prediction considering only the most preferred human grasping. The experiment demonstrated that the max-match rate was around 80% for the test objects.

To further examine grasping performance, we performed grasping experiments on the robotic hand platform. Test objects are placed on the table with a default initial orientation, and the objects are located and identified by the developed object recognition algorithms. A short description of each object was provided to cover its estimated dimension and materials. The robot autonomously chose the grasp strategies according to the acquired object affordance from the object database. The success of a grasp is validated by the security of grasping after brief maneuvers including lifting, holding and placing. A grasp is deemed as a success if the object does not fall during the maneuvers. The overall success rate of grasping was around 89%, and some experiment results are shown in Fig. 14. One failure case was the grasping of the capacitor, where the model predicted the most preferred human grasping strategy but failed to securely grasp the capacitor; the failure could be attributed to the limitation of hand dexterity or inadequate friction. Another failure case was the grasping of the plastic container, where the model predicted a different grasping strategy (rp) instead of the most preferred human grasp (wt). Though the predict strategy was one of the grasps used by humans and hence valid, the robot could not manage to secure the object. The other failure case was the grasping of the plier, where the model predicted the most preferred human grasping strategy (rp), but the plier changed formation during the grasping and fell. The developed system does not possess the capability in adjusting strategies during the process of grasping. There were many successful cases of grasping, where the predicted grasping strategies were different from the most preferred ones.

Grasping involves a series of decision making processes using experience and knowledge of the physical world as the basis. Human grasp strategies vary even when most of contextual variables are fixed. The existing human grasping data validated this interpretation: most objects are associated with multiple grasping strategies for the same tasks. Furthermore, it is still impractical to grasp one object without measuring or acquiring contextual information. Familiarity with a specific object is important knowledge that robots need to acquire before attempting a grasp with the assistance of object recognition algorithms. In an event of confusion, it still identifies the closest object with similar physical features, ensuring that there is enough information to continue onto grasp execution.

We designed the sequential machine-learning model to emulate human decision-making processes. In such a sequential model, errors from one model could permeate or propagate into the next model. In the experiment, while there were errors in each stage of and prediction, we did not observe a significant impact on the final grasp execution on the robot. The primary reason could be that, while human grasping is complex, it is also highly resilient to external perturbations of contextual variables. When modeling robot grasping using human grasp primitives, this resilience behavior was emulated as well. For example, there are multiple ways to grasp an object, so it is less likely to choose a wrong grasp as we have seen from the results of our deep learning model (100% match score). The other reason is that the experiment objects were designed with the intent of handling by five fingered hands, so when there is a mis-calculation, e.g., grasp dimensions, the fingers conform to the object shape and still result in a secure grasp.

V. Conclusion

This paper has demonstrated the approach to applying a proper strategy to grasp an object without complete sensing of object affordance. The framework of grasping strategy determination, object affordance acquisition, and robotic grasping deployment was developed through a combination of probabilistic and machine learning models in order to teach robots to grasp unfamiliar objects. The strategy determination can predict human grasping knowledge with a 100% feasibility.
score and a 80% match score. The designed distance metric outperformed other popular distance metrics and achieved overall 92% accuracy in object feature acquisition. These learning models could be extended to include additional inputs for position, orientation, and grasp intent for a broader applicability. With more sophisticated robotic arms built with multiple tactile sensors and precision force control, we can train models with more data and realize better performance in executing the grasps on real world objects. In summary, the experiments show that emulating human behavior is a practical way to build an autonomous robotic hand capable of adapting to unfamiliar environment.

REFERENCES

[1] Miao Li, Kaiyu Hang, Danica Kragic and Eude Billard: Dexterous grasping under shape uncertainty. Robotics and Autonomous Systems, 75:352–364, 2016.

[2] Hui Li, Jindong Tan et Hongsheng He: Magichand: Context-aware dexterous grasping using an anthropomorphic robotic hand. In IEEE International Conference on Robotics and Automation, 2020.

[3] Raphael Pelossof, Andrew Miller, Peter Allen et Tony Jeraba: An svm learning approach to robotic grasping. Proceedings - IEEE International Conference on Robotics and Automation, 4:3512–3518, 01 2004.

[4] A. J. Spiers, M. V. Liarakapis, B. Calli et A. M. Dollar: Single-grasp object classification and feature extraction with simple robot hands and tactile sensors. IEEE Transactions on Haptics, 9(2):207–220, April 2016.

[5] M. Bonilla, D. Resasco, M. Gabiccini et A. Bicchi: Grasp planning with soft hands using bounding box object decomposition. In 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pages 518–523, Sept 2015.

[6] A. Morales, T. Asfour, P. Azad, S. Knoop et R. Dillmann: Integrated grasp planning and visual object localization for a humanoid robot with five-fingered hands. In 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems, pages 5663–5668, Oct 2006.

[7] Roland S. Johansson et J. Randall Flanagan: Coding and use of tactile signals from the fingertips in object manipulation tasks. Nature Reviews Neuroscience. 10(5):345 – 359, 2009.

[8] M. R. Cutkosky: On grasp choice, grasp models, and the design of hands for manufacturing tasks. IEEE Transactions on Robotics and Automation, 5(3):269–279, Jun 1989.

[9] T. Feix, J. Romero, H. B. Schmiedmayer, A. M. Dollar et D. Kragic: The grasp taxonomy of human grasp types. IEEE Transactions on Human-Machine Systems, 46(1):66–77, Feb 2016.

[10] F. Heinemann, S. Puhlmann, C. Eppler, J. Elvarez Ruiz, M. Maertens et O. Brock: A taxonomy of human grasping behavior suitable for transfer to robotic hands. In 2015 IEEE International Conference on Robotics and Automation (ICRA), pages 4286–4291, May 2015.

[11] Marco Santello, M. Flanders et J.F. Soechting: Postural hand synergies for tool use. The Journal of neuroscience : the official journal of the Society for Neuroscience, 18:1015–15, 12 1998.

[12] M. Ciofcarlie, C. Goldfeder et P. Allen: Dimensionality reduction for hand-independent dexterous robotic grasping. In 2007 IEEE/RSJ International Conference on Intelligent Robots and Systems, pages 3270–3275, Oct 2007.

[13] A. T. Miller, S. Knoop, H. I. Christensen et P. K. Allen: Automatic grasp planning using shape primitives. In 2003 IEEE International Conference on Robotics and Automation (Cat. No.03CH37422), volume 2, pages 1824–1829, May 2003.

[14] S. Jain et B. Argall: Grasp detection for assistive robotic manipulation. In 2016 IEEE International Conference on Robotics and Automation (ICRA), pages 2015–2021, May 2016.

[15] S.A. Stansfield: Robotic grasping of unknown objects: A knowledge-
Figure 14. Robot grasp trials with 10 test objects. $H$ stands for human labeled grasps; $\hat{H}$ stands for Learned grasp. $\hat{H}$ in red color indicates mismatch with human grasp preference. ‘Failure’ indicates robot’s inability to grasp the object. ‘Success’ indicates the grasp was successfully executed on the AR10 Robotic Hand. More percentage of failure cases are demonstrated for better understanding of the limitation of the method.