A Study on the Influence of Task Dependent Anthropomorphic Grasping Poses for Everyday Objects

Niko Kleer and Martin Feick*

Abstract— **Robots using anthropomorphic hands and prosthesis grasping applications frequently rely on a corpus of labeled images for training a learning model that predicts a suitable grasping pose for grasping an object. However, factors such as an object's physical properties, the intended task, and the environment influence the choice of a suitable grasping pose. As a result, the annotation of such images introduces a level of complexity by itself, therefore making it challenging to establish a systematic labeling approach. This paper presents three crowdsourcing studies that focus on collecting task-dependent grasp pose labels for one hundred everyday objects. Finally, we report on our investigations regarding the influence of task-dependence on the choice of a grasping pose and make our collected data available in the form of a dataset.**

I. INTRODUCTION

Grasping objects represents one of a human's most significant abilities in order to carry out a considerable amount of daily activities. It is due to our cognitive abilities and dexterous hands that we are capable of overcoming the complexity of finding a suitable grasping pose for an arbitrary object in a given context. Furthermore, there exist numerous actions that do not require us to consciously think about the choice of a grasping pose as we have learned to associate them through experience [1]. Such actions may include drinking from a glass, holding a fork for eating, or carrying a plate. As a result, a grasping pose is determined subconsciously, leading to similar grasps being applied on a daily basis. Researchers have long had an interest in analyzing reoccurring grasping poses as well as the factors that influence the choice of each particular pose. This interest has lead to the emergence of sophisticated models that structure grasping poses according to their similarity [2], [3], [4]. These models typically emphasize that the choice of a grasping pose is mainly influenced by the physical properties of an object, the intention of the applied grasp (i.e. the task), and the impact of the environment. For example, picking up a freely accessible pen from a flat surface requires a different grasping pose as compared to grasping a pen for writing a note.

At the same time, one of the fields strongly profiting from our continuously improving knowledge about human grasps is the field of robotics. Enabling a robot that uses an anthropomorphic hand to grasp objects similar to a human represents a highly complex task as there exists an infinite amount of potential grasping poses. However, by utilizing our understanding of human grasping poses, the development of methods that allow robots to choose a suitable grasping pose becomes a feasible task. Similarly, the field of prosthesis can exploit this knowledge for the automatic selection of an appropriate grasping pose [5], [6]. Many more recently introduced approaches incorporate the power of convolutional neural networks for predicting a suitable pose for grasping an object [5], [7], [8], [9], [10]. In contrast to traditional grasp detection methods that aim towards detecting an area that enables a robot to securely grasp an object [11], [12], these approaches face a different challenge. Namely, the development of computational models that incorporates the factors influencing the choice of grasping pose.

*All authors are with the DFKI, Saarland Informatics Campus, Saarland University, 66123 Saarbrücken, Germany {niko.kleer;martin.feick}@dfki.de

Fig. 1. Demonstrates three different human grasping poses with a simple pen depending on the tasks: (A) writing, (B) pick-up and (C) hand-over.

A crucial step towards computational models that are capable of making accurate predictions with regards to the choice of a suitable grasping pose lies in the acquisition of data. For example, grasp prediction models that rely on computer vision techniques require a large amount of annotated object images for achieving a high accuracy in order to become useful. It is important to emphasize that, due to an appropriate grasping pose depending on several factors, the process of data labeling introduces a layer of complexity by itself. Even though this aspect is sometimes considered in the literature [13], [14], [15], authors often struggle to provide a crisp description with regards to their data annotation process [5], [7], [8], [9], [10]. As a result, they either do not provide a description or their methodology does not take the aforementioned factors influencing the choice of a grasp into account.

In order to acknowledge the complexity of this labeling process and the sparseness of systematically labeled data for such grasp prediction models, this paper makes the following contributions.

- We have conducted three studies on the choice and distribution of task-dependent grasp poses for 100 everyday objects and report on our findings.
- We further make our collected data available to the research community[1] in order to support the systematic annotation of data for grasp prediction models.

[1] https://github.com/nikleer/TaskDependentGrasps

II. RELATED WORK

The related work most relevant this paper comprises three fields. First, it is necessary to discuss the established literature in the field of human grasp analysis. After that, we have a closer look at concrete approaches that aim towards predicting a suitable grasping pose for grasping objects. Finally, we would like to elaborate on the related literature in the field of task-oriented grasping in robotics.

A. Analysis of Human Grasps

One of the most fundamental publications in the field of human grasp analysis was published by Napier [16] who first distinguishing between the so-called power and precision grip. Cutkosky [2] later went beyond the scope of distinguishing between power and precision grip by providing a structuring of grasps in a hierarchical taxonomy. Even though the authors point towards the incompleteness of their taxonomy that arises from task- and geometry dependent constraints, their work does consider the influence of these factors to a certain extent.

In an effort to provide a taxonomy of human grasps that considers all previously published taxonomies from various scientific fields, Feix et al. [3] published the so-called GRASP taxonomy. Their taxonomy represents the most comprehensive categorization of human grasp poses, and has established the current state-of-the-art. They further utilize their insights from analyzing the frequency of grasp types during tasks carried out by housekeepers and machinists [17] as well as the influence of an object's properties [18] and the corresponding task [19] on the choice of grasp. These insights enable them to provide statistics with regards to geometric and task dependent constraints for each grasp type in the GRASP taxonomy.

In contrast to the aforementioned taxonomy of grasps, which are based on qualitative criteria, Stival et al. [4] quantitatively assess the similarity of grasps. They retrieve electromyographic and kinematic data from test subjects and construct modality specific hierarchical representations reflecting the similarity of the executed grasps. Using their final taxonomy, which merges both models into a joint hierarchy, they derive five general classes for all grasps.

The publications outlined in this section serves as a basis for computational models that enable predicting a suitable grasp pose under different circumstances as in case of robotic grasping applications. Following, we have a look at these approaches and point towards difficulties in the systematic annotation of data that such models are trained on.

B. Grasp Prediction Methods

Due to the capabilities of convolutional neural networks in many computer vision tasks, the field of grasp predictions adopted the learning method for object-based grasp classification [7]. This approach has shown to be promising for grasp predictions in hand prosthesis [5], myoelectric hands [8], and robotic grasping applications [10]. As computer vision approaches that focus on a restricted amount of features tend to undermine the importance of the executed action, Yang et al. [15] focus on the inclusion of a human's action intention into their learning model. Cai et al. [20] view the problem of predicting a suitable grasp from a similar perspective and emphasize the importance of studying the relation between an object, a human's intention, and the applied grasp. One major challenge that vision-based grasp prediction models face lies in the systematic annotation of data. Especially when annotating images of objects where only one specific angle is considered and no additional context, such as the intended task, is provided. As a result, authors often vaguely describe how their data was annotated, or do not provide a description at all [5], [7], [8], [9], [10]. In some cases, it appears that the annotation process follows the process of elimination based on the grasping poses chosen by the authors. For example, DeGol et al. [5] describe this process as annotating their data by assigning the grasping pose they „felt was most natural".

We fully acknowledge the complexity that is introduced by all the factors influencing the choice of a grasp and believe that, especially in practical applications such as robotic grasping [13], [21], task-dependence requires more emphasis. We further believe that such investigations may help restricting the total number of grasps in nowadays' taxonomies to a small number of necessary grasps for specific tasks (e.g. a basic pick-and-place task). Arapi et al. [22] recently contributed to this challenge by providing a dataset of annotated videos in which subjects perform activities of daily living. They further analyze the statistical occurrences of classes based on their newly introduced video labeling taxonomy. In this paper, we present three crowdsourcing studies in which we collect task-dependent grasp pose labels for one hundred everyday objects and make our collected data available to the research community.

C. Task-oriented Grasping in Robotics

Numerous robotic systems that incorporate the influence of a specific task have been proposed, most of which are dedicated towards two-finger grasping [23], [24], [25], [26]. In fact, Murali et al. [25] take a similar approach to our work by leveraging crowdsourcing for the automatic labeling of their data. Dang and Allen [27] present a planning framework that incorporates task-related constraints for determining a suitable grasping pose. On top of such a constraint, Detry et al. [28] additionally enable their robot to gain geometry-based understanding of the scene at hand in order to generate an appropriate task-oriented grasp. In both publications, their contributions are demonstrated at the example of a three-fingered robot. In case of robots using an anthropomorphic hand, Prats et al. [29] utilize prototypical hand pre-shapes for enabling their robot to turn the handle of a door. Nguyen et al. [30] describe an approach for reorienting an arbitrarily positioned object into a position that enables their robot to apply the intended task-oriented grasp.

III. CROWDSOURCING STUDIES

The croudsourcing studies we present in this paper are targeted towards the annotation of object images based on

Fig. 2. Subset of objects that we used in our three studies.

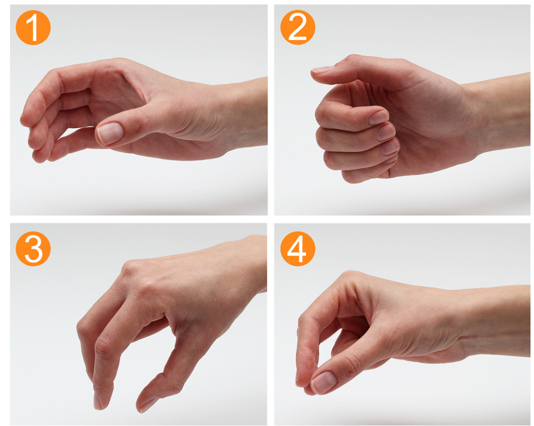

Fig. 3. The grasping poses referred to as Medium Wrap (1), Lateral (2), Tripod (3), and Writing Tripod (4).

grasp pose labels by collecting the assessments from test subjects. By doing so, we aim to contribute to simplifying the challenge of acquiring data as well as the systematic annotation of the acquired data. Instead of assessing suitable grasp poses for objects that are not put into a context, our collected labels consider task-dependence. Further, we are generally interested into the investigation of grasp pose distributions for specific tasks as some grasps, even though frequent in general, might not be frequently used during such tasks. For example, these insights might be used to support robot grasping applications by diminishing the number of potentially required grasps during specific tasks.

In the next four sections, we describe each significant aspect of our studies. We start by providing an overview of the objects we have chosen for the systematic annotation process. After that, we discuss and justify our selection of grasp poses for annotating the objects. Moving on, we describe our study setup and elaborate on the results of our studies.

A. *Object Images Collection*

In order to collect the assessments, we were required to gather images of objects that are shown to the participants during the studies. To this end, we have gathered a collection of images that covers a total of 100 objects. The majority of our images are publicly available and do not require the acquisition of a license. For all the objects where could not find an image suitable for our purpose, we either acquired a license in order to obtain the right to use the image, or we took a photograph of the respective object ourselves. Our final collection contains a large number of objects that are commonly used in most human's everyday life including fruit, tools, numerous containers, writing and eating utensils as well as a range of other everyday objects. Figure 2 shows a small sample of the objects from our collected images.

Another crucial factor in the execution of our studies represents the choice of grasping poses that a human may choose to grasp an object with. In the next section, we elaborate on our choices and provide appropriate justification.

B. *Selection of Grasp Poses*

The selection of grasp poses for annotating objects with regards to their most suitable grasp type represents an important factor. Choosing too many grasps increases the likelihood of introducing confusion, i.e. the most appropriate grasp becomes less clear. On the other hand, focusing on too few grasps might eliminate the possibility of grasping certain objects completely. Consequently, systematically choosing the smallest number of necessary grasping poses for annotating data becomes complicated, especially when analyzing the choice of a grasp during many tasks.

In order to determine the smallest number of grasps that spans over as many objects as possible during various tasks, we based our decision on the statistical observations from the literature [17], [18], [31]. Considering these observations, the grasping poses most distinguishable that span over the largest set of objects are referred to as Medium Wrap, Lateral, Tripod, and Writing Tripod. Figure 3 shows each grasping pose. It is worth mentioning that the same methodology was used by Salvadó [9] who chose nearly the same grasping poses for the annotation of images. Our choices also appear sensible in accordance with the quantitative grasp taxonomy established by Stival et al. [4] as we choose exactly one grasp from each category, excluding ring grasps.

C. *Setups of our Studies*

We conducted our crowdsourcing studies on Amazon Mechanical Turk (MTurk)[2], a crowdsourcing marketplace that enables the automatic annotation of large amounts of data through so-called Workers. Workers are humans that collectively work on assignments while gaining a small amount of money for each successfully annotated sample.

In our studies, we were interested in determining the most suitable grasping pose for grasping the objects shown by the images we have collected. This is why we setup a total of three studies where each study was targeted towards the analysis of the most suitable grasp choice and overall grasp choice distributions. In study number one, study participants were asked to to choose the most suitable grasping pose for **holding an object**. As a part of our second study, participants were asked to choose the most suitable grasping pose for

[2]https://www.mturk.com/

TABLE I

SAMPLE OF OUR FUNCTIONAL TASK DESCRIPTIONS FOR A SMALL SELECTION OF OBJECTS.

Object	Functional Task Description
Apple	How would you **eat** an apple?
Banknote	How would you **insert** a banknote into an ATM?
Chess Piece	How would you **move** a chess piece on a chess board?
Dice	How would you **roll** a die?
Fork	How would you **eat with** a fork?
Glue Stick	How would you **apply** glue with a glue stick?
Hairbrush	How would you **brush** your hair with a hairbrush?
Key	How would you **open** a door with a key?
Knife	How would you **cut** with a knife?
Lemon	How would you **squeeze** a lemon?
Mug	How would you **drink** from a mug?
Pill	How would you **ingest** a pill into your mouth?
Pliers	How would you **use** pliers?
Spatula	How would you **cook** with a spatula?
Tennis Ball	How would you **throw** a tennis ball?
Whisk	How would you **stir** with a whisk?

picking an object that is placed on a flat surface. In both of these studies, the task description was the same for all objects. This aspect changed in our third study where participants were asked to choose the most suitable grasping pose based on a **functional task description specifically related to an object**. Table I provides an overview of 15 examples for such functional task descriptions. Our approach for coming up with these descriptions were comprehensive discussions about what would be considered the task most commonly associated with each object. However, as we have not been able to determine such a description for every single object, we only collected this data for a total of 91 objects. For all the other objects, we stuck to asking participants how they would hold the object, similar to study number one. Figure 4 shows the mTurk study interface at the example of an acorn, which represents one of the objects on our images.

For all studies, the following conditions applied. Participants were given a maximum of five minutes to choose the most suitable grasping pose based on the provided task description. It was not mandatory for a participants to make a choice, leaving the possibility for a participant not to choose any grasp. Finally, we gathered 20 assessments for each object, resulting in 2000 assessments per study.

Following, we have a detailed look at the results we have been able to retrieve in our studies.

D. Results

In order to systematically report on the results from our crowdsourcing studies, we start by providing the most significant figures for each study followed by a cross-study comparison. For the sake of simplicity, we refer to our three studies as „hold", „pick", and „functional" as these are the tasks introduced in each study respectively. For each study, we report on the distribution of majority votes (i.e. the grasping poses that received the majority of assessments for each object) and the general distribution of assessments provided by our study participants. Both distributions are visualized in Figure 5.

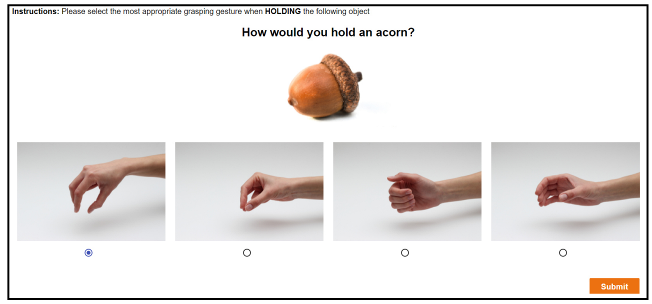

Fig. 4. The Amazon mTurk study interface seen by the study participants for choosing the most suitable grasping pose at the example of an acorn.

1) Study one (Hold): In case of this study, which included all objects, we have retrieved a total of 1983 assessments from 154 unique participants. Consequently, participants did not provide such an assessment 17 times (13 times for an individual object, twice for two objects). Based on these assessments, the lateral, medium wrap, tripod, and writing tripod grasping pose have received a total of 16, 37, 15, and 35 majority votes respectively. The overall distribution of assessments follows a similar trend resulting in 336, 648, 357, and 642 individual votes. The only difference between these distributions lies in the fact that the tripod grasp gathered less majority votes than the lateral grasp while receiving more assessments.

2) Study two (Pick): For study two, which also included all objects, we have retrieved a total of 1987 assessments from 141 unique participants. For 12 individual objects (11 times once, twice for one object), no assessment was provided. In this study, the medium wrap, tripod and writing tripod received nearly the same number of majority votes resulting in a total of 33, 32, and 35 respectively. For the lateral grasp, on the other hand, this was the case for only five objects. We can observe a similar assessment distribution with the medium wrap, tripod, and writing tripod all receiving nearly 600 assessments. Only the lateral grasp received slightly above 200.

3) Study three (Functional): Finally, study three concluded in a total of 1978 assessments from 153 unique participants. However, as we have only articulated a functional task description for 91 objects, only assessments for those objects are taken into account (i.e., 1800). For 19 individual objects (once every single time), an assessment was not provided. In case of the functional tasks, the lateral, medium wrap, tripod, and writing tripod have received a total of 13, 32, 19, and 30 majority votes respectively. However, the assessment distribution does not entirely reflect the same result as the writing tripod received more assessments than the medium wrap (565 and 512). Tripod and lateral grasp have received a total of 433 and 290 assessments respectively.

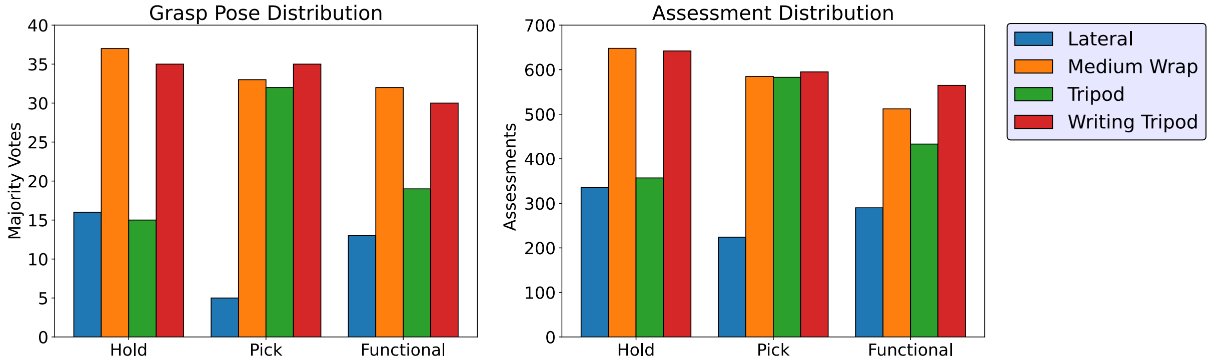

Fig. 5. Grasp pose and assessment distribution for the three tasks (a) holding an object (b) picking an object from a flat surface (c) functional task specifically related to each object.

4) Cross-study comparison: Comparing the results of our studies against each other, there are a number of notable aspects. The medium wrap and writing tripod have consistently accumulated the highest number of majority votes and assessments. This is because it can be observed that the medium wrap remains a popular choice for most bulky objects that allow a human to wrap their entire hand around the object. The opposite can be said about the writing tripod, which receives a considerable number of assessments in case of very flat objects such as a credit card and objects that resemble the shape of a writing utensil. We can also observe that the tripod grasp represented a substantially more popular choice during our pick study, leading to a considerably increase in the total number of majority votes and assessments. Furthermore, the lateral grasp, which represents the least chosen grasping pose in general, has received barely any majority votes during the pick study.

As we are interested into how a task influences the choice of a grasp, we consider it sensible to report on a series of examples that demonstrate task-independence, uncertainty and task-dependence within our results. Figure 6 shows the exact assessment distribution for a small subset of objects that perfectly demonstrate these three cases.

Task-independence can be seen at the example of an apricot, a paintbrush, or a strawberry where the total number of assessments does not tend to deviate from the same grasp. The medium wrap enables a secure grasp on the apricot while the paintbrush requires a more precise grasp such as the writing tripod. As the strawberry represents a smaller type of fruit, its size appears to be a perfect fit for the tripod grasp.

Uncertainty is clearly introduced in case of objects such as a frisbee, a glue stick, or a spatula. Participants have not been able to decide on which grasping pose would be most suitable in any of our studies. Each object can be grasped in multiple ways, independent of the tasks included in our study, leading to a broad distribution of the provided assessments. In fact, Figure 7 shows many cases of uncertainty where the assessments of our study participants almost equally span over two or three grasping poses given a specific task. We believe that such results further stress the aspect of considering the challenge of grasp pose classification as a multi class classification problem. However, by including additional factors, such as the placement of a grasp, it might be possible to lower the degree of uncertainty.

Task-dependence becomes visible in when viewing the distribution of assessments for the objects chalk, toothbrush, and pliers. Chalk, which gathered an overwhelming majority of assessments for the writing tripod grasp in our hold and functional study, is suddenly overruled by the tripod grasp in our pick study. Another such task-dependent fluctuation is visible for the toothbrush, where the overall distribution of assessments changes over three grasping poses. The same effect, but to an even more extreme extent, can be observed for the pliers where holding is mostly associated with the lateral, picking with the tripod, and the functional task with the medium wrap grasp. Finally, considering that we distinguish between only four grasping poses, it is notable that the choice of the most suitable grasp has changed at least once for 34 objects across all studies.

In addition to the quantitative data provided above, we believe that it is sensible to discuss limitations and potential implications of our studies. We elaborate on these aspects in the next section.

IV. Discussion

Following, we would like to elaborate on a few discussion points with regards to our studies in general as well as the results we have retrieved.

Even though this aspect lies beyond the scope of what this work was targeted towards, an object's orientation was only implicitly considered in these results. This is because, especially during many functional tasks, participants were required to make an implicit assumption about the orientation of an object. For example, when participants were asked as to how they would write with a pen, it is reasonable to assume that most participants would imagine holding a pen in about the same orientation. The exact opposite can be said about our first study where participants were simply asked as to how they would hold an object. The wording used in this study leaves room for interpretation to some extent. In fact, participants might have substituted holding an object by associating the task most commonly carried out with the

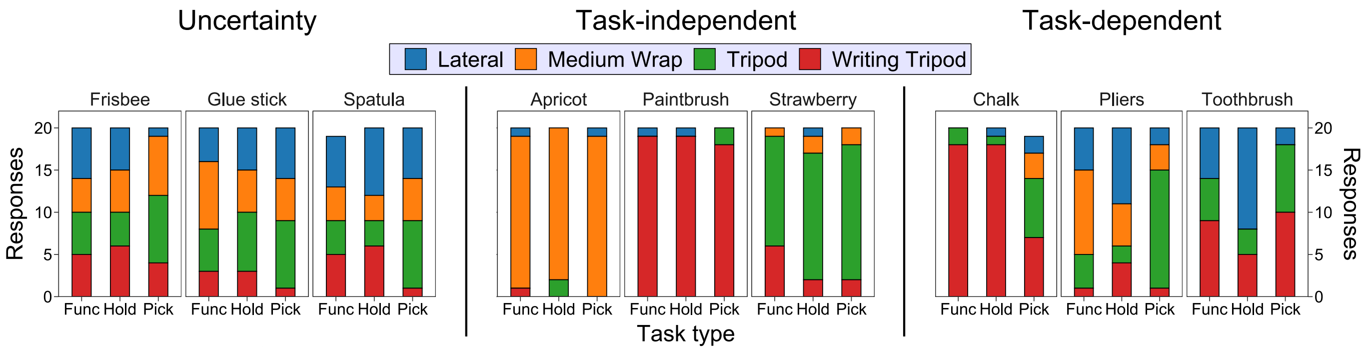

Fig. 6. Distribution for a small subset of objects that demonstrate task-independence, uncertainty, and task-dependence. A complete overview encompassing all objects and their corresponding assessment distributions can be found in Figure 7.

object (e.g. holding a pen would be interpreted equivalently to writing with a pen). When collecting data about grasp types, future studies should investigate the role of an object's orientation on the selection of an appropriate grasp pose.

As previously described, many authors vaguely describe their data annotation procedure, or do not provide a description at all [5], [7], [8], [9], [10]. Our analysis and the data we have retrieved are intended to aid other researchers during labeling tasks that involve the use of anthropomorphic grasping poses. This may include the preliminary annotation of images [7], [8], [9], [10], [14], [15], robotic and prosthesis grasping applications [5], [6], [13], [21], or enriching semantic ontologies with information about grasp types [32]. Furthermore, we hope to encourage other researchers to place a stronger emphasis on the aspect of task-dependence in order to gain more insights regarding the use of anthropomorphic grasps in different contexts.

V. Conclusion

This paper was targeted towards shedding light into the task-dependence of the choice of a grasping pose for grasping a large number of everyday objects. To this end, we have provided a comprehensive description regarding all the aspects of the three crowdsourcing studies we have conducted, including the data we used, the grasping poses we chose and the general setup of our studies. Based on this information, we have outlined the results of each study individually and as a part of a cross-study comparison. We further presented concrete examples for objects where our results show task-independence, general uncertainty with regards to the choice of a grasp, and task-dependence. Moreover, we discussed notable aspects of our study as well as potential future work. Finally, by making our data available to the research community, we hope to be able to contribute to the systematic annotation of data for anthropomorphic robotic grasping and prosthesis applications.

Acknowledgment

This work is supported by the German Federal Ministry of Education and Research (grant no. 01IW20008) as a part of CAMELOT - Continuous Adaptive Machine-Learning of Transfer of Control Situations.

References

[1] C. von Hofsten and L. Rönnqvist, "Preparation for grasping an object: a developmental study," *Journal of experimental psychology: Human perception and performance*, vol. 14, no. 4, p. 610, 1988.

[2] M. R. Cutkosky, "On grasp choice, grasp models, and the design of hands for manufacturing tasks," *IEEE Transactions on robotics and automation*, vol. 5, no. 3, pp. 269–279, 1989.

[3] T. Feix, J. Romero, H.-B. Schmiedmayer, A. M. Dollar, and D. Kragic, "The grasp taxonomy of human grasp types," *IEEE Transactions on human-machine systems*, vol. 46, no. 1, pp. 66–77, 2015.

[4] F. Stival, S. Michieletto, M. Cognolato, E. Pagello, H. Müller, and M. Atzori, "A quantitative taxonomy of human hand grasps," *Journal of neuroengineering and rehabilitation*, vol. 16, no. 1, pp. 1–17, 2019.

[5] J. DeGol, A. Akhtar, B. Manja, and T. Bretl, "Automatic grasp selection using a camera in a hand prosthesis," in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2016, pp. 431–434.

[6] L. T. Taverne, M. Cognolato, T. Bützer, R. Gassert, and O. Hilliges, "Video-based prediction of hand-grasp preshaping with application to prosthesis control," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 4975–4982.

[7] G. Ghazaei, A. Alameer, P. Degenaar, G. Morgan, and K. Nazarpour, "An exploratory study on the use of convolutional neural networks for object grasp classification," *2nd IET International Conference on Intelligent Signal Processing 2015 (ISP)*, pp. 1–5, 2015.

[8] ——, "Deep learning-based artificial vision for grasp classification in myoelectric hands," *Journal of neural engineering*, vol. 14, no. 3, p. 036025, 2017.

[9] F. Lluís Salvadó, "Grasp prediction with convolutional neural networks," B.S. thesis, Universitat Politècnica de Catalunya, 2017.

[10] A. Das, A. Chattopadhyay, F. Alia, and J. Kumari, "Grasp-pose prediction for hand-held objects," in *Emerging Technology in Modelling and Graphics*. Springer, 2020, pp. 191–202.

[11] S. Kumra and C. Kanan, "Robotic grasp detection using deep convolutional neural networks," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2017, pp. 769–776.

[12] J. Redmon and A. Angelova, "Real-time grasp detection using convolutional neural networks," in *2015 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2015, pp. 1316–1322.

[13] A. B. Rao, K. Krishnan, and H. He, "Learning robotic grasping strategy based on natural-language object descriptions," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2018, pp. 882–887.

[14] N. Kleer, M. Feick, and M. Feld, "Leveraging publicly available textual object descriptions for anthropomorphic robotic grasp predictions," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2022.

[15] Y. Yang, C. Fermuller, Y. Li, and Y. Aloimonos, "Grasp type revisited: A modern perspective on a classical feature for vision," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 400–408.

[16] J. R. Napier, "The prehensile movements of the human hand," *The Journal of bone and joint surgery. British volume*, vol. 38, no. 4, pp. 902–913, 1956.

[17] I. M. Bullock, J. Z. Zheng, S. De La Rosa, C. Guertler, and A. M. Dollar, “Grasp frequency and usage in daily household and machine shop tasks,” *IEEE transactions on haptics*, vol. 6, no. 3, pp. 296–308, 2013.

[18] T. Feix, I. M. Bullock, and A. M. Dollar, “Analysis of human grasping behavior: Object characteristics and grasp type,” *IEEE transactions on haptics*, vol. 7, no. 3, pp. 311–323, 2014.

[19] ——, “Analysis of human grasping behavior: Correlating tasks, objects and grasps,” *IEEE transactions on haptics*, vol. 7, no. 4, pp. 430–441, 2014.

[20] M. Cai, K. M. Kitani, and Y. Sato, “Understanding hand-object manipulation with grasp types and object attributes,” in *Robotics: Science and Systems*, vol. 3. Ann Arbor, Michigan;, 2016.

[21] Z. Deng, B. Fang, B. He, and J. Zhang, “An adaptive planning framework for dexterous robotic grasping with grasp type detection,” *Robotics and Autonomous Systems*, vol. 140, p. 103727, 2021.

[22] V. Arapi, C. Della Santina, G. Averta, A. Bicchi, and M. Bianchi, “Understanding human manipulation with the environment: a novel taxonomy for video labelling,” *IEEE Robotics and Automation Letters*, vol. 6, no. 4, pp. 6537–6544, 2021.

[23] K. Fang, Y. Zhu, A. Garg, A. Kurenkov, V. Mehta, L. Fei-Fei, and S. Savarese, “Learning task-oriented grasping for tool manipulation from simulated self-supervision,” *The International Journal of Robotics Research*, vol. 39, no. 2-3, pp. 202–216, 2020.

[24] M. Kokic, D. Kragic, and J. Bohg, “Learning task-oriented grasping from human activity datasets,” *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 3352–3359, 2020.

[25] A. Murali, W. Liu, K. Marino, S. Chernova, and A. Gupta, “Same object, different grasps: Data and semantic knowledge for task-oriented grasping,” in *Conference on Robot Learning*. PMLR, 2021, pp. 1540–1557.

[26] A. Ramisa, G. Alenya, F. Moreno-Noguer, and C. Torras, “A 3d descriptor to detect task-oriented grasping points in clothing,” *Pattern Recognition*, vol. 60, pp. 936–948, 2016.

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

[28] R. Detry, J. Papon, and L. Matthies, “Task-oriented grasping with semantic and geometric scene understanding,” in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2017, pp. 3266–3273.

[29] M. Prats, P. J. Sanz, and A. P. Del Pobil, “Task-oriented grasping using hand preshapes and task frames,” in *Proceedings 2007 IEEE International Conference on Robotics and Automation*. IEEE, 2007, pp. 1794–1799.

[30] A. Nguyen, D. Kanoulas, D. G. Caldwell, and N. G. Tsagarakis, “Preparatory object reorientation for task-oriented grasping,” in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2016, pp. 893–899.

[31] I. M. Bullock, T. Feix, and A. M. Dollar, “Finding small, versatile sets of human grasps to span common objects,” in *2013 IEEE International Conference on Robotics and Automation*. IEEE, 2013, pp. 1068–1075.

[32] C. Li and G. Tian, “Transferring the semantic constraints in human manipulation behaviors to robots,” *Applied Intelligence*, vol. 50, no. 6, pp. 1711–1724, 2020.

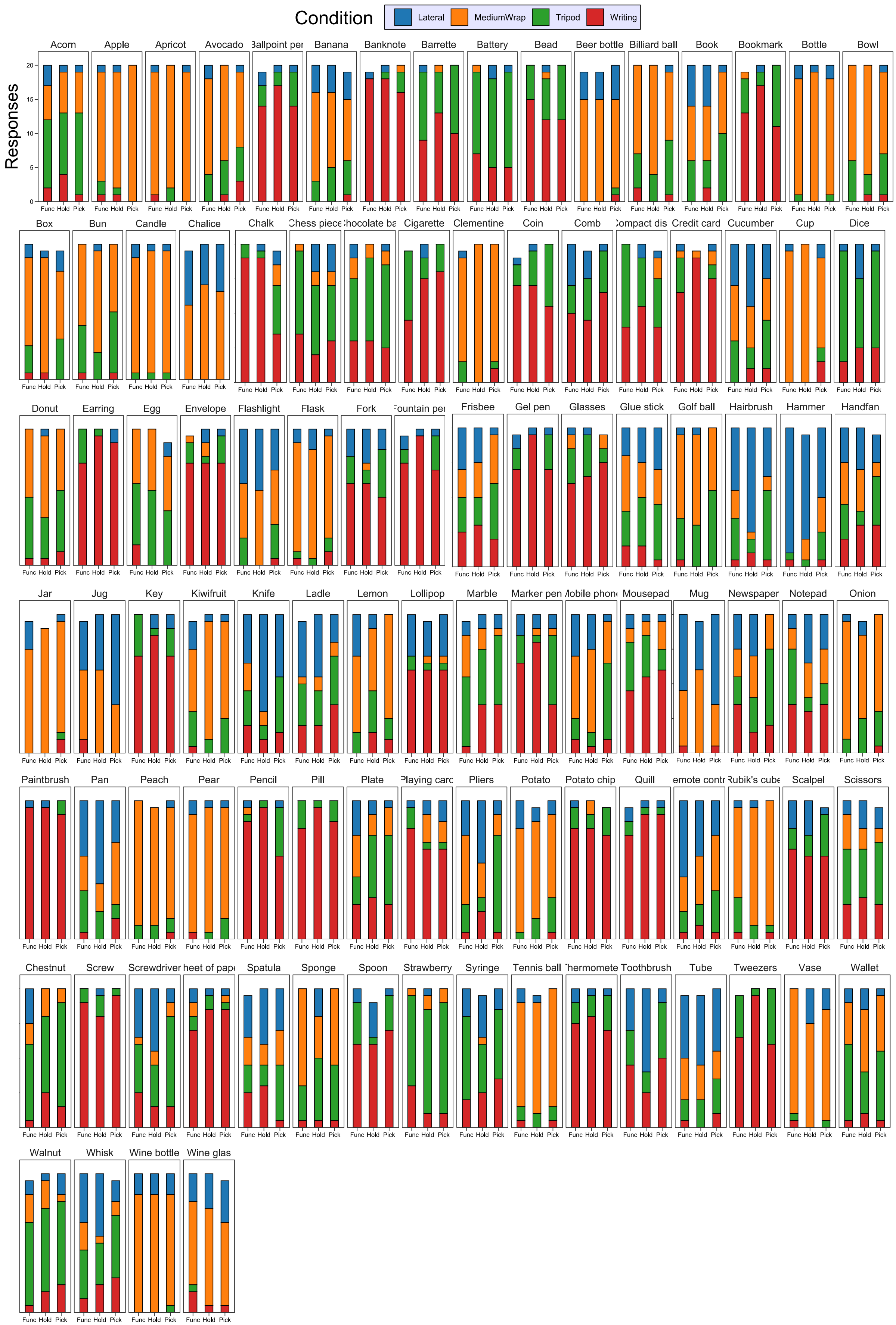

Fig. 7. Overview of all objects used during our studies and their corresponding assessment distributions for each task.