Explainable Grasping with Soft Grippers using Visual Quality Metrics

Lucy B. Griswold, Ian C. Rankin, Timothy R. Player, Stephanie Walker,
Joseph R. Davidson, and Geoffrey A. Hollinger

Abstract— **Soft grippers have yet to receive the attention that has been devoted to traditional hard grippers, and an area that has remained almost completely unexplored is the interface between human users and robots using soft robotic hands. Grasp-quality metrics for soft hands are underresearched, making the quality of a soft grasp difficult to understand at a glance. Comprehensible and accurate explanations are required in use cases where it is important to establish a secure grasp before manipulation. Using a soft hand, we verify the predictive accuracy of three grasp quality metrics obtained using only visual data. We then introduce a framework based on these metrics to generate effective explanations of the quality of a soft robotic grasp, providing the feedback necessary for users to safely and effectively manipulate environments with soft robots. In results we show we can correctly predict 85% of grasps and show examples of the generated explanations.**

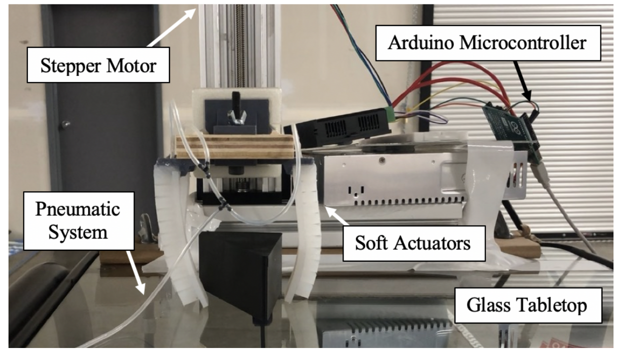

Fig. 1: The soft hand and experimental setup used for data collection. The camera was placed under the gripper in order to image the grasps from the bottom up, through the clear table.

I. INTRODUCTION AND PREVIOUS WORK

Soft robotic grippers have a number of unique features of interest. Their inherent compliance not only makes them well-suited for grasping tasks requiring delicacy (such as the handling of fruits or marine lifeforms), but also provides leniency in the complex and precise pre-grasp planning demanded by hard grippers [1]. This tendency for soft grippers to adjust after initial contact also affirms the need to perform post-grasp quality analysis and ensure a stable grasp before manipulating the object.

Explainability is a crucial element of both user-controlled and autonomous robotic systems, yet the overlap between explainability and soft grasping is not well investigated. Comprehensible explanations encourage appropriate levels of trust in the robot, and allow human users to quickly intervene in a robot's decision making process if needed [2]. Grasp quality metrics are an efficient way of quantitatively determining the quality of a grasp. Accurate visual grasp metrics, therefore, are an excellent way to build robust, trustworthy explanations, especially for systems such as soft grippers that may have limited sensing ability with which to otherwise determine the quality of a grasp.

While a multitude of grasp quality metrics have been developed [3], and some have been evaluated for underactuated and compliant hands [4], less research has been devoted to the application of quality metrics to completely soft hands. In [5], the accuracy of various combinations of quality metrics is analyzed using two hard, three-fingered grippers. Particular combinations of three metrics are shown to be just as accurate as up to seven metrics used in conjunction, but no tests were performed using soft grippers. Thus, it is vital that grasp metrics be tested using soft hands in order to verify their generalizability.

Previous work in grasp explainability has used machine learning methods with a dataset of simulated grasps to predict the likelihood of grasp failure, but this is data-heavy and relies on complex information about joint position, velocity, and torque of each finger, information which is not readily available from soft hands [6]. Other research has used explainability in conjunction with grasp planning [7], but to our knowledge post-grasp explainability of the type compatible to soft grippers has yet to be explored.

The ability for the robot to explain concisely whether it has a quality grasp on an object of interest could cut down dramatically on failed grasp attempts, allowing a human user to intervene and establish a new grasp. Enhanced explainability would allow operators to make more informed decisions, making operation safer and more time-effective. For example, in a marine organism manipulation task, users need to trust that an established grip is likely to be stable, and that the animal is unlikely to fall from the grasp and be damaged. Users are likely to find a comprehensive explanation of the qualities of the grasp, likelihood of success, and alternative example grasps far more understandable and trustworthy compared to a string of numbers [8].

In order to generate quality explanations, there must be sufficient evidence to distinguish a good grasp from a bad one. Building on the conclusions of [5], we test the predictive accuracy of three visual metrics: the shape of the grasp polygon, the distance between the centroid of the grasp

This research was funded in part by NSF grant IIS-1845227. L. B. Griswold is with the Department of Physics and Astronomy, Carleton College, Northfield, MN, United States, (morel@carleton.edu). All other authors are with the Collaborative Robotics and Intelligent Systems (CoRIS) Institute, Oregon State University, Corvallis OR, United States, ({rankini, playert, walkers, joseph.davidson, geoff.hollinger}@oregonstate.edu).

polygon and the center of mass of the object, and the area of the grasp polygon (see [9]). We then use these metrics, in conjunction with data collected through real grasp trials using a soft hand, to generate example explanations of soft grasp quality.

II. METHODS

A. Device

The gripper used for data collection utilizes a soft actuator design from [10] and is partially inspired by [11]. As shown in Fig. 1, the gripper is composed of three interchangeable fingers attached to a hard palm with custom 3D-printed brackets that connect the actuators to a pressurization system. Ease of fabrication was a key considerations in design, as was modularity. The softness of the fingers and their width allows the gripper to grasp a variety of objects. The molds used to cast the fingers were 3D printed from PLA plastic directly from the files from the Soft Robotics Toolkit [10]. The completed hand was raised and lowered using a controlled stepper motor, providing precise and repeatable hand placement.

B. Metrics

The metrics used to predict the quality of a given grasp were selected based on analysis from Rupert, et al. [5], which identifies several high-quality grasp metric combinations using real three-fingered robotic grippers. However, these combinations were yet to be tested using a soft hand, and thus a portion of this project was devoted to verifying the predictive accuracy of these metrics for a soft grasp.

Due to a lack of inbuilt tactile sensors, we selected three metrics based on easily-obtained visual data about the "grasp polygon," or the polygon created by the contact points between the object and the gripper. The first metric was the regularity of the grasp polygon, which can be defined as

$$Q_1 = 1 - \frac{1}{\theta_{max}} \sum_{i=1}^{n_f} |\theta_i - \theta_{ref}|, \tag{1}$$

where n_f is the number of fingers of the gripper, θ_i is the i^{th} vertex of the polygon, θ_{ref} is the average of all internal angles of the polygon, and θ_{max} is the sum of the differences between the internal angles when the polygon degenerates into a line and those of the regular polygon, such that

$$\theta_{max} = (n_f - 2)(180 - \theta_{ref}) + 2\theta_{ref}. \tag{2}$$

Q_1 is normalized such that $Q_1 \in [0, 1]$, with 1 being the ideal case where the grasp polygon is regular.

The second metric was the distance between the centroid of the grasp polygon and the center of mass of the object, which can be defined as

$$Q_2 = 1 - distance(p, p_c)/distance_{max}, \tag{3}$$

where $distance(p, p_c)$ is the distance between the center of mass of the object p and the centroid of the grasp polygon p_c, and $distance_{max}$ is the maximum distance between p and any point on the object's contour. This metric has also

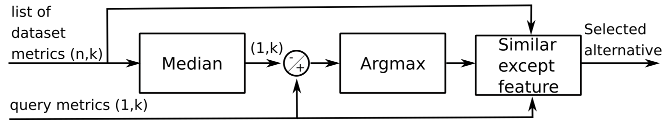

Fig. 2: A block diagram of the explanation selection. The method first determines the best feature in the semantic feature to demonstrate, then selects the best alternative grasp to demonstrate the selected feature.

been normalized such that $Q_2 \in [0, 1]$, 1 being the ideal case where p and p_c are the same point.

The third metric we selected was the area of the grasp polygon, or simply

$$Q_3 = area(Polygon). \tag{4}$$

Theoretically, a grasp polygon with a larger area produces a more robust grasp. In practice, we found the area of the grasp was not predictive of the grasp quality of our soft gripper. We thus discarded grasp area as a valid metric, leaving us with two predictive quality metrics Q_1 and Q_2 (which we will refer to as the "regularity" and "distance" metrics, respectively).

C. Explainability

In this paper we use a theory-based explanation system to provide transparency into the grasp classification [12]. Studies from the Social Sciences have shown that humans prefer selected contrastive explanations to understand an event [8]. Our system is designed to follow these recommendations by providing selected alternative grasps that attempt to demonstrate the most salient reasons a grasp is classified as successful or unsuccessful. We formulate this problem as determining the most salient features from a semantic feature vector, then selecting the best alternative grasp to demonstrate the selected feature, see Figure 2.

To select the best alternative grasp, we iterate through each example grasp to find the closest grasp that is either better or worse than the query grasp, depending on whether we are explaining why we think the grasp will succeed or fail for the given feature. Samples with the same grasp prediction were rejected. Formally this is solving the equation

$$i^* = \underset{i}{\operatorname{argmax}}\ (x_{i,j} - x_{q,j}) + \frac{||x_{i,-j} - x_{q,-j}||}{\sqrt{n-1}}, \tag{5}$$

where j is the index of the feature being explained, x_q is the query metric, and $x_{q,-j}$ is the query vector except for index j. This returns the best alternative that describes the given feature being explained. Once the alternative is selected images of both grasps are shown with a templated description of the selection.

III. RESULTS

Data was collected using four 3D printed shapes: a right triangle, a large star, and two arbitrary shapes designed to vary the contact polygon as much as possible. A total of 90 different grasps were executed. During experimentation, a grasp was labelled "successful" if the gripper was able to lift the object off of the table while maintaining three contact points, and "unsuccessful" otherwise. Each trial was imaged

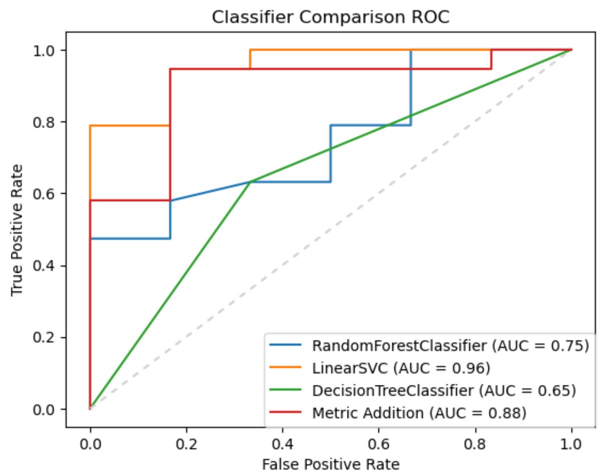

Fig. 3: ROC curve comparing the performance of a random forest, linear SVC, and decision tree classifiers against adding Q_1 and Q_2 together and determining a linear threshold value. The threshold was found by calculating the maximum geometric mean of the recall and the specificity [13]. Classifier algorithms and plotting tools from [14].

from the bottom, producing a 2D representation of the grasp. The contact points of each grasp and the center of mass of the object were labelled by hand; the grasp quality metrics were then calculated from these points. Of the 90 trials, 82 produced usable grasp images; 8 were rejected due to trial failures, such as when the gripper lost a point of contact during the imaging process.

Our final data set contains 82 example grasps executed on 4 objects. 54 of these were successful grasps, and 28 were unsuccessful. We experimented with several different classifiers to determine cutoff values that would be most predictive of a successful grasp. As shown in Fig. 3, the linear SVC (Support Vector Classifier) method classified the test grasps most accurately with an area under the curve (AUC) of 0.96, followed by simply adding the regularity and distance metrics together and determining a combined threshold by finding the threshold that maximizes the geometric mean of the true positive rate and the true negative rate [13], producing an AUC of 0.88. The latter method delivered results averaging around 85% predictive accuracy.

		Actual Class	
		Unsuccessful	Successful
Predicted Class	Unsuccessful	7	1
	Successful	3	14

TABLE I: Confusion matrix generated from the test data, classified using the metric addition method. As shown, this classification method performs well using the calculated threshold value, with balanced numbers of false positives and false negatives and an overall accuracy of 84%.

IV. Discussion and Conclusion

Our experimental results indicate that the regularity and distance metrics are suitable to soft grippers, but that the area of the grasp polygon is not a good predictor of soft grasp quality. We successfully demonstrate that accurate post-grasp quality metrics can be used to generate helpful explanations for human operators, see Figures 4 and 5.

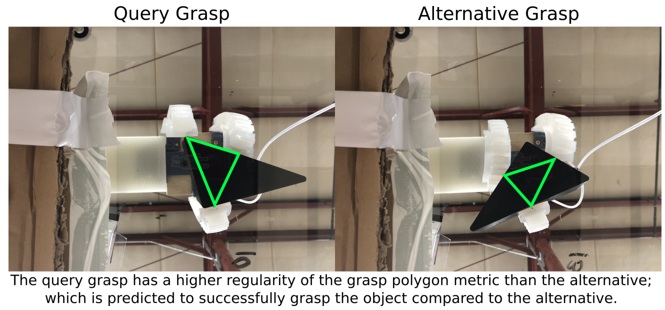

Fig. 4: An example explanation for a predicted successful grasp.

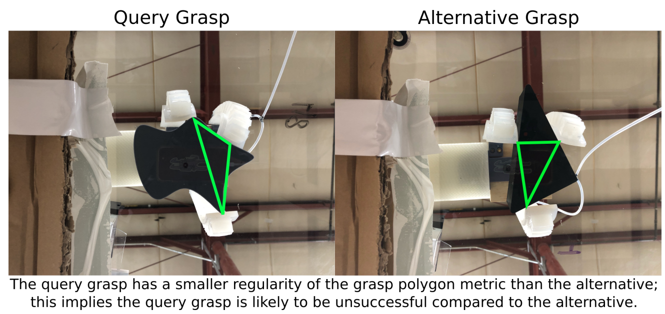

Fig. 5: An example explanation for a predicted unsuccessful grasp.

There are limitations to the work done in this study. First, the object positions used to generate the grasp data were not truly random. Although efforts were taken to cover the space of possible grasps, unconscious operator preference towards certain object positions may have excluded unlikely successful grasps. Additional study with truly randomized orientation would ensure the space of possible grasps is fully covered. Second, the elimination of the area metric as a viable predictor of grasp quality means the explainability framework is limited to two sets of features. More robust explanations could be generated from additional metrics, such as those presented in Rupert, et al. [5]. Third, metric accuracy and the resulting explanations were generated from a binary "successful/unsuccessful" grasp classification. Although additional data about each grasp was collected, it was not applied to either the metric calculation or the explanations. More complex analysis of each grasp may lead to more satisfactory explanations.

In this work we presented an explainability framework for post grasp analysis of soft grippers. This can help robot operators manage the a system with soft grippers. Additionally, we verified two grasp quality metrics for soft grippers. The ideas presented in this work could be applicable to a number of use cases requiring the delicacy of soft grippers along with the assurance of grasp quality, such as agricultural or exploratory marine robots.

Acknowledgment

This project was completed as part of the NSF REU Sites program and would not have been possible without the guidance and supervision of so many members of CoRIS, including but not limited to Nick Bira, Natasha Troxler, Nidhi Homey Parayil, Brian Zhang, and Miranda Jean Cravetz.

References

[1] B. S. Homberg, R. K. Katzschmann, M. R. Dogar, and D. Rus, “Robust proprioceptive grasping with a soft robot hand,” *Autonomous Robots*, vol. 43, no. 3, pp. 681–696, 2019.

[2] N. Wang, D. V. Pynadath, and S. G. Hill, “Trust calibration within a human-robot team: Comparing automatically generated explanations,” in *Proc. 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2016, pp. 109–116.

[3] M. A. Roa and R. Suárez, “Grasp quality measures: review and performance,” *Autonomous Robots*, vol. 38, no. 1, pp. 65–88, Jan. 2015. [Online]. Available: https://doi.org/10.1007/s10514-014-9402-3

[4] M. Pozzi, M. Malvezzi, and D. Prattichizzo, “On Grasp Quality Measures: Grasp Robustness and Contact Force Distribution in Underactuated and Compliant Robotic Hands,” *IEEE Robotics and Automation Letters*, vol. 2, no. 1, pp. 329–336, Jan. 2017.

[5] C. Rubert, D. Kappler, J. Bohg, and A. Morales, “Predicting grasp success in the real world - A study of quality metrics and human assessment,” *Robotics and Autonomous Systems*, vol. 121, p. 103274, Nov. 2019.

[6] A. Alvanpour, S. K. Das, C. K. Robinson, O. Nasraoui, and D. Popa, “Robot Failure Mode Prediction with Explainable Machine Learning,” in *Proc. IEEE International Conference on Automation Science and Engineering (CASE)*, Aug. 2020, pp. 61–66.

[7] A. Mitrevski, P. G. Ploger, and G. Lakemeyer, “Representation and Experience-Based Learning of Explainable Models for Robot Action Execution,” in *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Oct. 2020, pp. 5641–5647.

[8] T. Miller, “Explanation in artificial intelligence: Insights from the social sciences,” *Artificial intelligence*, vol. 267, pp. 1–38, 2019.

[9] C. Rubert, B. Leon, A. Morales, and J. Sancho-Bru, “Characterisation of Grasp Quality Metrics,” *Journal of Intelligent & Robotic Systems*, vol. 89, no. 3-4, pp. 319–342, Mar. 2018.

[10] D. Holland, S. Berndt, M. Herman, and C. Walsh, “Growing the soft robotics community through knowledge-sharing initiatives,” *Soft Robotics*, vol. 5, no. 2, 2018.

[11] B. S. Homberg, R. K. Katzschmann, M. R. Dogar, and D. Rus, “Haptic identification of objects using a modular soft robotic gripper,” in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Hamburg, Germany, Sep. 2015, pp. 1698–1705.

[12] S. Mohseni, “A multidisciplinary design and evaluation framework for explainable AI systems,” Ph.D. dissertation, Texas A&M University, 2020.

[13] M. Kubat, S. Matwin *et al.*, “Addressing the curse of imbalanced training sets: one-sided selection,” in *Proc. International Conference on Machine Learning*, vol. 97. Citeseer, 1997, pp. 179–186.

[14] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, “Scikit-learn: Machine learning in Python,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.