2019 International Conference on Robotics and Automation (ICRA)
Palais des congres de Montreal, Montreal, Canada, May 20-24, 2019

Learning from Transferable Mechanics Models: Generalizable Online Mode Detection in Underactuated Dexterous Manipulation

Andrew S. Morgan[1], *Student Member, IEEE,* Walter G. Bircher[1], *Student Member, IEEE,*
Berk Calli[2], *Member, IEEE,* and Aaron M. Dollar[1], *Senior Member, IEEE*

***Abstract*— In this work, we investigate a mechanics-inspired framework for describing fingertip-based planar within-hand manipulation with an underactuated robotic gripper. In particular, this framework leverages fundamental mechanics properties of the hand-object system, including basic terms such as local contact curvature as well as more complex features including the grasp matrix and manipulability metrics. These are extracted using a simple visual approach and then in real-time used for predicting planar manipulation modes: namely rolling, dropped, stuck, and sliding. Given a desired cartesian motion for the object, a supervised learning model predicts these four manipulation modes before they occur, allowing us to either avoid or trigger these different behaviors. Since we utilize strictly fundamental properties of the grasp matrix, finger Jacobians, and contact curvatures, we are able to demonstrate prediction transferability between different grippers using our original classifier. In particular, a Random Forests classifier trained on one gripper successfully predicts manipulation modes for grippers with different fingers with 84% accuracy, compared to just 56% from an approach in previous work. Overall, we find that the features designed in our approach better describes fingertip manipulation when precise gripper models are not available.**

I. Introduction

Practical implementation of dexterous within-hand manipulation (WIHM) in physical robotic systems is a major challenge due to unknown or inaccurate mechanics model parameters (e.g. forces, coefficient of friction, exact contact scenario, etc.). In traditional approaches with rigid, high degree-of-freedom (DOF) hands (e.g. [1], [2]), the system is highly overconstrained due to the contacts and the closed kinematic chain, which greatly reduces the controllable degrees of freedom. Alternatively, this overconstraint can be mitigated through underactuated hands [3], [4], which are underconstrained before contact and will inherently reconfigure to passively adjust to contacts, noisy control, and external disturbances [5]. However, due to their adaptability, deriving accurate hand models is difficult since finger equilibrium, contact forces, and kinematics must be solved simultaneously in order to account for reconfiguration [6]–[8].

In our previous work, we have shown that robust, planar WIHM is possible, without detailed hand or object models, using basic visual servoing techniques and no additional sensing [9], [10]. Additionally, in follow-on work, we leverage this robust and simple control for data-driven approaches to improve understanding of the hand-object system, especially for difficult-to-sense modes of operation during WIHM, such as when the object is sliding, stuck, about to drop, or exhibiting rolling contacts (normal) [11]. This preliminary work explored predicting modes of manipulation in real-time by using visual and proprioceptive actuator data extracted during object manipulation. However, the features used during this preliminary effort were not particularly generalizable and the learned models were unable to be transfer to similar grippers – as in most machine learning cases, it was system dependent.

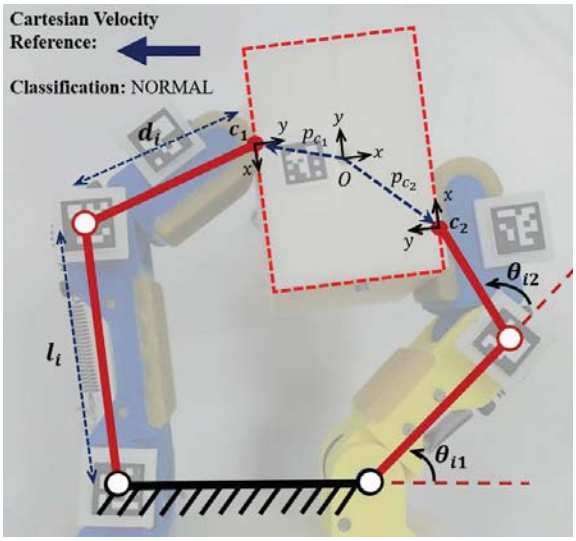

Fig. 1. Features extracted from vision to construct the finger Jacobian and the grasp matrix, determined by effective link lengths, joint configurations, and the contact and object frames.

Our main contribution in the current work is that we demonstrate a mechanics-inspired feature set that captures elemental properties of the hand-object system, that at least partially generalize across different grippers as key features when link lengths and finger geometries change. This approach is advantageous for tasks that may require different hands to successfully complete, e.g. fingertip geometries or transmission ratios, where one prediction model can be trained to work over various hand designs. Experimentally, using cartesian movement intent and certain mechanics-inspired features extracted only from visual information from one hand, we show that the classifier developed is able to successfully predict manipulation modes when transferred between grippers with different physical dimensions. This work

Research supported by the U.S. National Science Foundation under grant IIS-1734190.

[1] A. S. Morgan, W. G. Bircher, and A. M. Dollar are with the Dept. of Mech. Eng. and Mater. Science, Yale University, CT, USA, ({Andrew.Morgan; Walter.Bircher; Aaron.Dollar}@yale.edu).

[2] B. Calli is with the Department of Computer Science, Worcester Polytechnic Institute, MA, USA, (bcalli@wpi.edu).

encompasses features agnostic to the exact dimensions of the hand system, which provides additional insights to the importance of the proposed features after training is conducted. From this mechanics-inspired approach, we seek to answer the following research questions:

1.) Can a mechanics-inspired supervised learning model predict modes of within-hand manipulation?

2.) Is this approach generalizable between similar physical gripper specifications? That is, can we train this model with one gripper, test with others, and maintain a high accuracy in mode prediction?

3.) What mechanics features are most important to mode prediction of a planar, underactuated robotic gripper?

The layout of this paper is described as follows: Section II reviews complementary literature to this work, Section III presents our methodology for defining modes of manipulation, how data is collected and analyzed, and what features are extracted during the manipulation, Section IV describes the experimental results with a following discussion on feature importance, and finally, Section V concludes this paper with an additional discussion of future work.

II. Related Work

Directly utilizing long studied fundamental mechanics in robotic manipulation such as inverse kinematics, grasp matrices, and finger Jacobians [12]–[18] requires estimation of kinematic and kinetic model parameters, making the development of a precise model a challenging endeavor. In fully actuated systems, a small error in these parameters during a manipulation could result in undesired slippage. This problem has been previously addressed by learning friction coefficients at the moment of incipient slippage detected through tactile sensors [19], [20]. Naturally, this approach requires prior exploration by manipulation with an object to calculate such parameters, which may be infeasible in time-sensitive or mission critical tasks. Complementary works have approached this question not by estimating physical parameters, but through various machine learning techniques to estimate grasp stability through tactile sensing [21]–[23]. Due to the rigidity of high DOF hands, force sensing in their control loop is typically required to maintain grasp stability.

Unlike fully actuated hands, which would typically require an impedance control framework [24], the passive adaptability of underactuated hands allows the grasp to be maintained under reasonable external disturbances without requiring additional sensing. Because of this compliance and adaptation, complexities to the hand model are introduced, making accurate models non-trivial to derive for precision manipulation. Various techniques for modeling underactuated hands have been introduced in previous works, evaluating configurations and torque about each of the joints [6], [7], [25]. Yet, precision manipulation through these models is typically only done in simulation due to their simplified frictional assumptions. Because of the complexities in this frictional estimation, reinforcement learning, like that in [26], has enabled force control policies to be learned for compliant joints.

Learning policies for manipulation is a well-developed area of interest. Historically, hand designed controls have been task-specific and do not generalize well to unstructured environments. On the other hand, Deep Reinforcement Learning has shown to be promising and is currently one of the most popular approaches to learning robot control and state estimation (e.g.[27], [28]). A large caveat to this approach is the amount of data required for training, which is often an infeasibly large amount for a physical implementation. Simulation is typically used as an attempt to overcome this problem. An approach from [29] addressed this caveat by relying on videos from the internet to teach a robot manipulation actions. In [30], a framework is developed to reduce the amount of training data required for the control policy and was shown to be successful when implemented for the block stacking task. Other approaches have combined imitation learning and reinforcement learning to address this problem. This combination requires less reinforcement exploration since it is initially guided by a human agent [31].

Nonetheless, while these approaches have been fairly successful, in general, all lack generalizability and interpretability once a policy is formed and are dependent on gripper specific parameters, like sensor orientation and finger geometries. Unlike these works that develop complex control strategies, our object manipulation approach uses Precision Manipulation Primitives (PMPs) [10], from simple kinematic models to move objects in cartesian directions (x- and y- with respect to the base frame). From these primitives, we can move objects in random directions using an automated controller and learn what mode the object is in, or about to transition into. We limit our sensing modalities to a vision only approach, since the hand used (Model T42, Yale Openhand Project [4]) does not have onboard tactile sensing. Due to the lack of tactile sensors, slip (or sliding) detection, as in [32], [33], cannot be implemented. We believe that approaching manipulation with mechanics-inspired feature sets, which can be extracted solely through vision, will allow us to better interpret findings for future advancements in manipulation tasks.

III. Methods

In this work, we utilize a modified Model T42 from the Yale Openhand project to study planar WIHM. Modifications to this model include a rounded fingertip, adjustable extension springs on the proximal links, and an internal pulley system to reduce friction in the tendon's transmission. This hand model has two, two-link opposable fingers, where each finger is powered by a single Dynamixel RX-28 motor. As a tendon driven mechanism, the flexion force is provided by the motors and the return force from the passive springs connected to each of the joints. Two additional (urethane rubber, durometer 30) distal and proximal links with similar frictional coefficients and of different lengths are designed for experimental testing (see Fig. 2, 5), to test the research questions presented in the motivation of this work. All external, visual sensing is completed through the use of an overhead camera detecting ArUco markers attached to the hand's rigid links. Link-joint relationships are then calculated for data representation.

In planar manipulation, four mutually exclusive modes, which are dependent on the state and the intended motion of the object, are identified of which can occur:

1.) *Drop* - The hand-object configuration is in a state where the object is just about to drop and will drop shortly

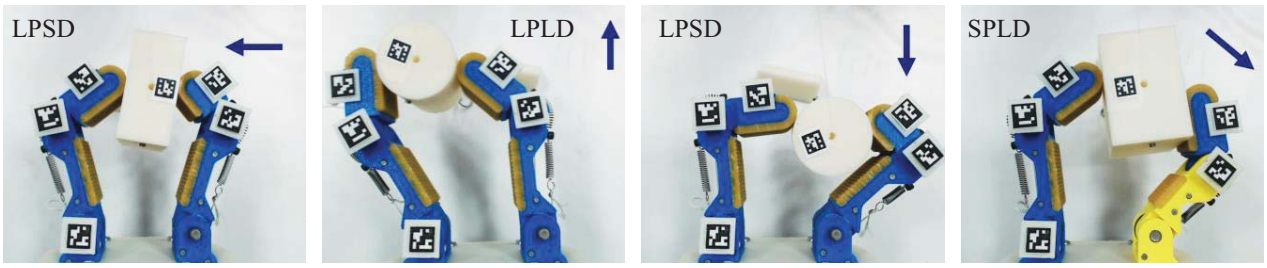

Fig. 2. Classification mode prediction from left to right: normal, drop, stuck, and sliding on three different gripper variants (top corner). Cartesian velocity references, or commanded motion of the object from the PMPs, are denoted by arrows in the top right-hand corner.

thereafter from the commanded next cartesian workspace location.

2.) *Stuck* - The object is no longer able to move in the commanded cartesian direction due to the hand-object configuration of the gripper, or the joint has reached a physical hard stop.

3.) *Sliding* - The object exhibits a sliding contact with respect to the gripper's distal link. Note: this mode only occurs on the tested rectangular prisms.

4.) *Normal* - The object is manipulable within the gripper's workspace while maintaining a rolling contact, and 1-3 are not satisfied.

A. Feature Design and Motivation

While previous work [11] relied on raw, unstructured sensor data to train a mode predicting classifier, this work leverages the inherent structure of equations describing physical phenomena. Specifically, we make use of Jacobian based manipulability metrics [34], [35], finger and object curvatures at contact [36], and the singular values of the grasp matrix. These proposed features are generalizable in that they aggregate many terms into composite values which do not relate directly to the specific dimensions of the links of the hand - rather they convey higher level behaviors and properties of the entire hand-object system, allowing features to be transferable.

First, we make use of a modified manipulability measure [35] that requires a-priori knowledge of joint limits. This metric is positively unbounded in regions of the kinematic finger workspace where joint motions have the greatest effect on the cartesian motion of the end effector, and has a lower bound of zero in regions where the effect is negligible due to either kinematic singularities or joint limits. Inspired by Yoshikawa's original measure of manipulability [34], Vahrenkamp's is the product of this value and a special penalty function, which scales the manipulability measure between 0 and 1 depending on the joint's proximity to its upper and lower limits. The penalty function is defined as $P(\theta_i)$ where k can be used to tune how quickly manipulability drops off near the joint limits. Visual representations of the following parameters are presented in Fig. 1. We know that our Jacobian, J_i, represents the map from joint velocities to cartesian velocities. That is,

$$\dot{r}_i = J_i \dot{\theta}_i \tag{1}$$

where i represents the finger number, $\dot{r}_i = [\dot{x}_i \quad \dot{y}_i]^T$, and $\dot{\theta}_i = [\dot{\theta}_{i1} \quad \dot{\theta}_{i2}]^T$. We can represent J_i in the planar case as,

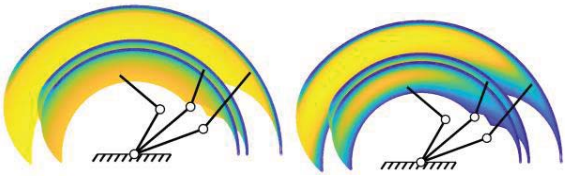

Fig. 3. Manipulability workspaces for three fingers used in experiments with (Left) from Yoshikawa, and (Right) from Vahrenkamp after penalty. Darker region indicates less manipulability.

$$J_i(\theta_i) = \begin{bmatrix} -l_i s_{i1} - d_i s_{i12} & -d_i s_{i12} \\ l_i c_{i1} + d_i c_{i12} & d_i c_{i12} \end{bmatrix} \tag{2}$$

where $\theta_i = [\,\theta_{i1} \;\; \theta_{i2}\,]^T$, $c_{i1} = \cos(\theta_{i1})$, $s_{i1} = \sin(\theta_{i1})$, $c_{i12} = \cos(\theta_{i1} + \theta_{i2})$, and $s_{i12} = \sin(\theta_{i1} + \theta_{i2})$. The manipulability measure, w_i, as defined by Yoshikawa is then,

$$w_i = \sqrt{\det(J_i J_i^T)} = |\det(J_i)| = l_i d_1 |\sin(\theta_{i2})| \tag{3}$$

The penalty function, $P_i(\theta_i)$, from Vahrenkamp can be represented as,

$$P_i(\theta_i) = 1 - \exp(-k \prod_{j=1}^{n} \frac{(\theta_{i,j} - l_{i,j}^-)(l_{i,j}^+ - \theta_{i,j})}{(l_{i,j}^+ - l_{i,j}^-)^2}) \tag{4}$$

where $l_{i,j}^+$ and $l_{i,j}^-$ represent the maximum and minimum joint limits for each joint, j, and each finger, i, respectively. We can then calculate the penalized manipulability measure for each finger individually by,

$$w_{i_{penalized}} = P|\det(J_i)| \tag{5}$$

Kinematic workspaces for these two manipulability measures are presented in Fig. 3.

Next, we propose a feature based on the grasp matrix[16], [37], a matrix which is most commonly used in manipulation mechanics to relate the velocity of the object to the velocity of the fingertip at the location of contact. Based on the directions of the normal vectors of the contacts, the grasp matrix can be constructed from the geometry of the hand-object system alone - no knowledge of contact forces or friction coefficients is required. Our inspiration for using the grasp matrix comes from the observation that its minimum singular value has a lower bound of zero, regardless of the pertinent dimensions of the hand or the object [38]. Because of this, it is invariant across systems of different physical dimensions. In particular, this lower bound represents cases where contact force normal vectors align on the surface of an object, creating a condition where force closure is impossible. In other words, this term

can be used to signal the likelihood of dropping the object. We can formulate the relationship for the grasp matrix, G, of a two-contact, planar grasp as,

$$F_o = G f_c \tag{6}$$

where $F_o = [f_{Ox}, f_{Oy}, \tau]^T$ and $f_c = [f_{c_{1x}}, f_{c_{1y}}, f_{c_{2x}}, f_{c_{2y}}]^T$, and where f_{Ox}, f_{Oy}, τ are x and y forces on the object, and torque about the perpendicular axis, respectively. We denote $f_{c_{ix}}$ and $f_{c_{iy}}$ for single contact on finger i, as the x and y forces applied from the contact. Assuming a point contact with friction, we then formulate a basis, B_{c_i}, representing the contact friction model on finger i. It is also necessary to extract the position vector, p_{c_i}, of the ith contact frame, c_i, and its offset rotation, R_{c_i}, with respect to the object frame, O. So,

$$B_{c_i} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} R_{c_i} = \begin{bmatrix} \cos(\theta_{\delta_i}) & -\sin(\theta_{\delta_i}) \\ \sin(\theta_{\delta_i}) & \cos(\theta_{\delta_i}) \end{bmatrix} p_{c_i} = \begin{bmatrix} p_{c_i x} \\ p_{c_i y} \end{bmatrix}$$

where $\theta_{\delta_i} = \theta_{c_i} - \theta_o$, where θ_{c_i}, θ_o are the angle offsets of the contact and object frame, respectively. Finally, we can calculate the two-contact grasp matrix, G, by calculating,

$$Ad^T_{g_{oc_i}^{-1}} = \begin{bmatrix} R_{c_i} & 0 \\ [-p_{c_i y} \quad p_{c_i x}] R_{c_i} & 1 \end{bmatrix} \in \mathbb{R}^{3x3} \tag{7}$$

$$G = \begin{bmatrix} Ad^T_{g_{oc_1}^{-1}} B_{c_1} & Ad^T_{g_{oc_2}^{-1}} B_{c_2} \end{bmatrix} \in \mathbb{R}^{3x4} \tag{8}$$

B. *Automated Data Collection*

Generating training data for within-hand manipulation without simulation is a labor-intensive process, as the object is easily dropped and must be manually placed back between the fingers. Additionally, stalling the motors in a stuck case often causes the motors to overheat, which ceases data collection and requires a system reset. To streamline this process, we fabricated an automated object resetting system and an autonomous hand controller, which commands the hand with randomly generated cartesian velocity references, specified by the PMPs, and then observes the result. In this way, we are able to streamline data collection by reducing human interaction. The object resetting system consists of a crane and a stabilization beam, both actuated by hobby servos. The crane lifts the object back to an appropriate starting height between the fingers, after it has been dropped or stuck. The stabilization beam snaps the object into a known orientation by mating neodymium magnets on the object with one on the end of the beam. This system allows an object to be repeatedly grasped from the same initial position and orientation with respect to the gripper. The resetting system is shown in Fig. 4.

The automated data collector is controlled through ROS and is comprised of several packages that can be categorized as either hand control, data collection, or feature extraction. Previous descriptions of the PMP control utilized in this work have been provided in [9], [10], which provide a cartesian velocity reference (intent to move) in cardinal and intercardinal directions. Our developed automated controller selects a direction at random for anywhere between 0.5 to 2.5 seconds. This random selection exploits motions and hand-object configurations previously unobserved from human teleoperation, which provides for a higher variance in mode location within the workspace and a harder learning problem.

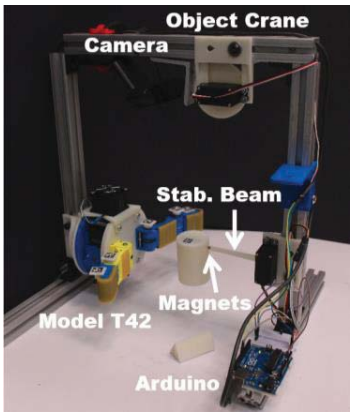

Fig. 4. Automated data collector commanded through ROS and Arduino. Once the object is dropped, the stabilizing beam rotates down, attaching to the object. After stabilization, the object is grasped and the system is reset for the next drop case.

Experimental data collection is automated through parallelized detection nodes, which are triggered when certain criteria are satisfied. Once detected, an instantaneous data point is recorded preserving the state of our feature set, which will later be used for classification. Detection is as follows:

Drop: The object has moved one centimeter below its initial finger configuration or has not been detected for 5 frames, equivalent to 0.166 seconds. A data point is collected for is last known valid configuration.

Stuck: The cartesian reference of the object has been commanded for more than 1.5 seconds without any object movement.

Sliding: The relative distance to orientation change of the ArUco marker between the object and the fingertip has changed. Detected after 0.5 seconds of movement.

Normal: All other modes have not been detected within the last 5 seconds of manipulation.

Lastly, feature extraction as described in Sec. IIIA. is achieved by the use of an overhead camera and ArUco markers attached to rigid links of the gripper (Fig. 1, 5). During a manipulation task, properties of the grasp change dynamically according to the hand-object configuration, e.g. contact curvatures, effective distal link lengths from a changing contact location, and joint angles. First, contact location between the finger and the object is captured through a KD-Tree [39] and vision extracted point clouds. At this point, effective distal link lengths are then calculated with respect to the contact point location and the location of the proximal-to-distal joint, accurate within 2mm. These values are normalized with the length of the proximal to eliminate dimensionality. An example of the effective link length changing is illustrated in Fig. 1, where the left link is effectively longer than the right due to the location of the contact point on the fingerpad. Thereafter, this information is then sufficient to calculate the desired Jacobians, contact curvatures (radius of a circle fitted at contact that is extracted from a 2D point cloud), and grasp matrix as presented in Sec. IIIA. The feature vector, A, used for initial classification can be represented as (Table 1),

$$A = [g_{max}, g_{min}, j_l, j_r, j_l', j_r', k_{lf}, k_{rf}, k_{l0}, k_{r0}, v_x, v_y] \in \mathbb{R}^{12}$$

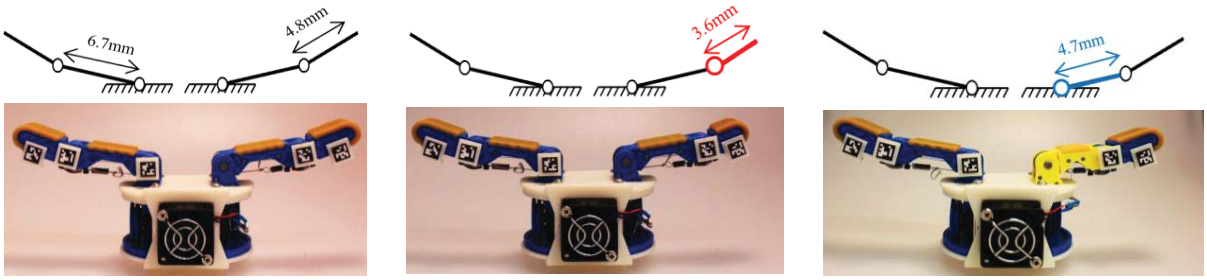

Fig. 5. Three gripper variants from left to right: LPLD (symmetric), LPSD (asymmetric), and SPLD (asymmetric). Dimensions for the long proximal, small proximal, long distal, and short distal are 6.7mm, 4.7mm, 4.8mm, and 3.6mm respectively.

C. Data Collection

Four 3D Printed ABS objects are used in our experiments (Fig. 6). The object geometries in the manipulation plane are as follows: (1) large cylinder, 41mm dia., (2) small cylinder, 29mm dia., (3) large rectangular prism, 41x60mm, and (4) small rectangular prism, 29x60mm. Each object has a hole through the center to allow for attachment to the object reset crane. Additionally, two small neodymium magnets are adhered to opposite ends of the object to allow for attachment to the stabilization beam.

Experimental data was collected with a total of three different gripper variants and four different objects (Fig. 5). First, training data was collected with a symmetric Model T42 for normal, drop, stuck, and sliding cases over all four objects. We refer to this gripper variant as a dual large-proximal, large-distal configuration (LPLD). A total of 3500 data points were collected (1000 normal, 1000 dropped, 1000 stuck, 500 sliding) for the training set, with points being equally distributed across each of the objects. Sliding only occurred with the rectangular prisms, hence, half the number of data points according to the object geometry distribution.

Testing data was collected by equipping the hand with two different asymmetric finger configurations. These two gripper configurations are referenced as: (1) single large-proximal large-distal with a single small-proximal large-distal (SPLD), and (2) a single large-proximal large-distal with a single large-proximal small-distal (LPSD) (Fig. 5). A total of 1750 points were collected with four objects from these two variants, with a balanced half collected using the large cylinder and small prism for the LPSD setup, and the other half collected with the small cylinder and large prism for the SPLD setup. Test data is effectively half for each mode of the training data.

IV. Experimental Results

A. Classification Results

Our first interest was to obtain the best classification score with all 12 features described in feature vector, A, 10 of which are extracted from vision and 2 of which from the controller (Table 1). In our supervised learning approach, we utilized TPOT and Scikit-learn [40] tools to evaluate two different prediction models: Random Forests (RF) and Support Vector Machines (SVM). For reference, a 5-fold internal cross validation score, provided by TPOT, for the training data using a Random Forests classifier is found to be 89.2%. This classification score describes prediction accuracy when the classifier is trained and tested with our 3500 LPLD data points, which is the training data for our following analysis.

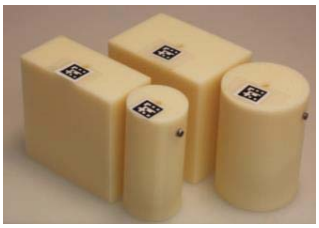

Fig. 6. Four ABS objects used in the experimental analysis

In our analysis, the SVM classifier was evaluated with both a linear and radial kernel to explore two different learning approaches. Therefore, in total, 3 different predictive models were trained, each using the 3500 data points from the LPLD setup and all 4 objects. The classifiers were then individually tested using the 1750 data points from the two combined SPLD and LPSD setups. Classification results in Table 2 show that the RF classifier outperformed the others with a classification rate of 84.6%, less than a 5% decrease from the cross-validation score. The two SVM classifiers fell marginally below with a classification rate of 79.5% for the linear kernel, and a classification rate of 82.8% for the radial kernel.

Our best classification model, Random Forests, will serve as the basis for the rest of our analysis. We find that we are best able to classify the drop case with over 94% accuracy, whereas stuck classifies correctly 88% of the time. We note that our main prediction confusion comes from the rolling/sliding distinction, where this prediction is only 77% accurate. This follows closely with intuition; since tactile force sensing is not a feature, it is difficult to predict sliding explicitly. Additionally, sliding does not occur for each of the objects, which decreases the number of training points we were able to use to train the classifier for this specific case. Interestingly, if we consider a two-class prediction, drop versus the three other modes, we see the classification rate rise to 96.3%. This concept can be further exploited for future applications that requires the gripper to maintain manipulation capabilities without dropping the object.

It is in our interest to understand where, within the workspace, certain modes are most likely to occur, regardless of physical gripper dimensions. Fig. 7 presents these regions for the testing and training data, where each point represents the center of the object when the mode was realized in data collection. We notice that not only does the workspace becomes smaller with the SPLD and LPSD setup, which is to be expected, but the mode regions appear to be more overlapping and less discernable compared to the training data,

TABLE 1. FEATURES COLLECTED FOR INITIAL EVALUATION

Number	*Feature Set*	*Notation*
1-2	Min/Max SVD Grasp Matrix	g_{min}, g_{max}
3-4	Left/Right Finger Manip. Measure	j_l, j_r
5-6	Left/Right Penalized Manip. Measure	j_l', j_r'
7-8	Left/Right Finger Pad Curvature	k_{lf}, k_{rf}
9-10	Left/Right Object Curvature	k_{lo}, k_{ro}
11-12	X/Y Cartesian Velocity Reference	$v_x,\ v_y$

TABLE 2. CLASSIFICATION RESULTS BY CLASSIFIER AND FEATURES SELECTION

	Random Forests	*SVM - Linear*	*SVM – Radial*
All Features	84.6 %	79.5 %	82.8 %
6 Most Imp.	82.7 %	81.8 %	81.9 %

that appears to be somewhat symmetric about a vertical axis through the center of the gripper. We also note that misclassifications for the asymmetric test data do not tend to occur in any specific object regions in the workspace – it is well dispersed throughout. Therefore, simply relying on the center of the object, coupled with velocity intent, for mode classification, as in [11], is not likely a transferable approach to other physical gripper variants.

B. Feature Importance

Likely, the main benefit of our approach, versus others described in related works, is the interpretability of feature importance measurements presented through the RF classifier (Fig. 8). We see in this ranking that the intended velocity, v_y, is the most important feature. This result follows closely with intuition given the illustrated regions within the workspace, with commanded velocities in the x less likely to move the object into another region. We note that some modes directly rely on v_y, where we are typically only stuck when moving towards the base of the gripper. We then find that the two sets of manipulability measures [34], [35], are the next most important, since these describe properties of the finger configuration and the ability for the finger to move within the workspace. These four features coupled, when coupled with v_y, are able to classify stuck cases with 96% accuracy.

Using only the top 6 most important features highlighted in orange from Fig. 8, with the same training and test data, our classification for RF does marginally worse with a rate of 82.7%. Interestingly, the linear kernel for the SVM classifier performs better with less features, raising from 79.5% to 81.8%. Lastly, the radial kernel also performs marginally worse with fewer features, classifying at a rate of 81.9%.

C. Summary and Discussion

We find that we are able to predict manipulation modes with 84.6% classification accuracy using our geometric, mechanics-inspired features. Due to the lack of tactile sensors, it is difficult to differentiate between normal and sliding modes, since this is both object and contact dependent. Additionally, object drop detection is also sometimes difficult due to lack of force feedback and imperfect understanding of contact scenarios. Naturally, the grasp on the object in the asymmetric finger case eradicates the axis of symmetry for modes in the workspace, making force closure prediction less

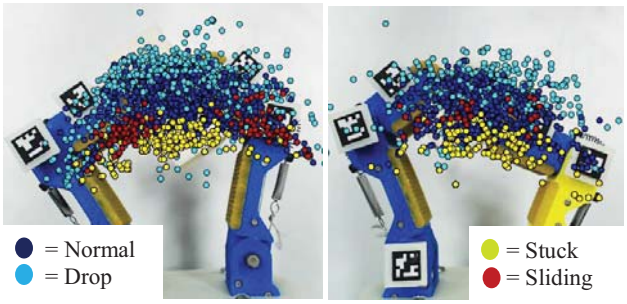

Fig. 7. Regions identified in experimental collection according to object center for the (Left) LPLD and the (Right) SPLD.

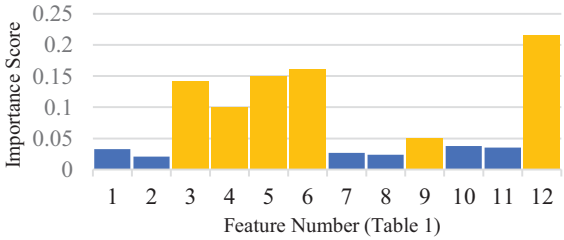

Fig. 8. Feature importance scores reported by the RF classifier. Features in orange signify the top 6 features used in feature reduction analysis, effectively half of the initial feature vector.

stable. Though, by solely using vision, this classification analysis is promising for future implementation and analysis.

This work was motivated by the lack of mobility exhibited by previous machine learning models when transferred to different grippers. We can present the significance of this work by a final comparison with that in [11], where for all four modes, a classification accuracy of 89% was realized from just four features. When training a RF classifier with these four features using the LPLD training data collected in this work, then testing with the SPLD and LPSD setup, we realize a 56% classification accuracy, vastly dropping the classification rate for a four-mode case and showing significance to our work.

V. CONCLUSION

In this work, we described a vision-extracted, mechanics-inspired feature set that allowed us to successfully predict modes of WIHM over various underactuated robotic grippers with 84.6% accuracy. Unlike in previous works where the feature set is only valid for the target system, we have shown that, when using features invariant to the physical dimensions of the gripper, we are able to learn a transferrable model to predict modes and maintain a high level of prediction accuracy. We believe that the presented approach provides insights for future research in dexterous manipulation when using grippers that are hard to precisely model.

As future work, we believe that this framework can be tested further for hand variants with different spring constants (joint compliance), link lengths, and fingerpad curvatures. By training such a classifier with unique gripper variants, it is likely possible to transfer mode prediction accuracies to a wider variety of grippers. Additionally, we believe that training and testing with everyday objects will exploit interesting feature importance measures, dependent on the objects used and the relative geometry of the finger.

REFERENCES

[1] J. Butterfass, M. Grebenstein, H. Liu, and G. Hirzinger, "DLR-Hand II: next generation of a dextrous robot hand," in *Proceedings 2001 IEEE International Conference on Robotics and Automation (ICRA)*, vol. 1, pp. 109–114.

[2] C. Lovchik and M. A. Diftler, "The Robonaut hand: a dexterous robot hand for space," in *Proceedings 1999 IEEE International Conference on Robotics and Automation (ICRA)*, vol. 2, pp. 907–912.

[3] A. M. Dollar and R. D. Howe, "The Highly Adaptive SDM Hand: Design and Performance Evaluation," *Int. J. Rob. Res.*, vol. 29, no. 5, pp. 585–597, Apr. 2010.

[4] R. Ma and A. Dollar, "Yale OpenHand Project: Optimizing Open-Source Hand Designs for Ease of Fabrication and Adoption," *IEEE Robot. Autom. Mag.*, vol. 24, no. 1, pp. 32–40, Mar. 2017.

[5] R. Balasubramanian, J. T. Belter, and A. M. Dollar, "Disturbance Response of Two-Link Underactuated Serial-Link Chains," *J. Mech. Robot.*, vol. 4, no. 2, p. 021013, May 2012.

[6] A. Rocchi, B. Ames, Zhi Li, and K. Hauser, "Stable simulation of underactuated compliant hands," in *Proceedings 2016 IEEE International Conference on Robotics and Automation (ICRA)*, 2016, pp. 4938–4944.

[7] T. Laliberté and C. M. Gosselin, "Simulation and design of underactuated mechanical hands," *Mech. Mach. Theory*, vol. 33, no. 1–2, pp. 39–57, Jan. 1998.

[8] J. Borras and A. M. Dollar, "A parallel robots framework to study precision grasping and dexterous manipulation," in *Proceedings 2013 IEEE International Conference on Robotics and Automation (ICRA)*, 2013, pp. 1595–1601.

[9] B. Calli and A. M. Dollar, "Vision-based model predictive control for within-hand precision manipulation with underactuated grippers," in *Proceedings 2017 IEEE International Conference on Robotics and Automation (ICRA)*, 2017, pp. 2839–2845.

[10] B. Calli and A. M. Dollar, "Vision-based precision manipulation with underactuated hands: Simple and effective solutions for dexterity," in Proceedings 2016 *IEEE International Conference on Intelligent Robots and Systems*, 2016, pp. 1012–1018.

[11] B. Calli, K. Srinivasan, A. Morgan, and A. M. Dollar, "Learning Modes of Within-Hand Manipulation," in *Proceedings 2018 IEEE International Conference on Robotics and Automation (ICRA)*, 2018, pp. 3145–3151.

[12] S. Gruber, "Robot hands and the mechanics of manipulation," *Proc. IEEE*, vol. 75, no. 8, pp. 1134–1134, 1987.

[13] J. Kerr and B. Roth, "Analysis of Multifingered Hands," *Int. J. Rob. Res.*, vol. 4, no. 4, pp. 3–17, Jan. 1986.

[14] D. L. Brock, "Enhancing the dexterity of a robot hand using controlled slip," in *Robotics and Automation. Proceedings., 1988 IEEE International Conference on. IEEE*, 1988.

[15] J. C. Trinkle, "A quasi-static analysis of dextrous manipulation with sliding and rolling contacts," *Int. Conf. Robot. Autom.*, pp. 788–793, 1989.

[16] R. M. Murray, Z. Li, and S. S. Sastry, *A Mathematical Introduction to Robotic Manipulation*, vol. 29. 1994.

[17] A. M. Dollar and R. D. Howe, "Towards grasping in unstructured environments: Grasper compliance and configuration optimization," *Adv. Robot.*, vol. 19, no. 5, pp. 523–543, Jan. 2005.

[18] T. Okada, "Computer Control of Multijointed Finger System for Precise Object-Handling," *IEEE Trans. Syst. Man. Cybern.*, vol. 12, no. 3, pp. 289–299, 1982.

[19] Z. Su *et al.*, "Force estimation and slip detection/classification for grip control using a biomimetic tactile sensor," in *IEEE-RAS International Conference on Humanoid Robots*, 2015, vol. 2015–Decem, pp. 297–303.

[20] M. R. Tremblay and M. R. Cutkosky, "Estimating friction using incipient slip sensing during a manipulation task," in *Proceedings 1993 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 429–434.

[21] H. Dang and P. K. Allen, "Learning grasp stability," in *Proceedings 2012 IEEE International Conference on Robotics and Automation (ICRA)*, 2012, pp. 2392–2397.

[22] Y. Bekiroglu, J. Laaksonen, J. A. Jorgensen, V. Kyrki, and D. Kragic, "Assessing Grasp Stability Based on Learning and Haptic Data," *IEEE Trans. Robot.*, vol. 27, no. 3, pp. 616–629, Jun. 2011.

[23] K. Hang *et al.*, "Hierarchical Fingertip Space: A Unified Framework for Grasp Planning and In-Hand Grasp Adaptation," *IEEE Trans. Robot.*, vol. 32, no. 4, pp. 960–972, Aug. 2016.

[24] D. Prattichizzo and J. C. Trinkle, "Grasping," in *Springer Handbook of Robotics*, Cham: Springer International Publishing, 2016, pp. 955–988.

[25] L. U. Odhner and A. M. Dollar, "Dexterous manipulation with underactuated elastic hands," in *Proceedings 2011 IEEE International Conference on Robotics and Automation (ICRA)*, 2011, pp. 5254–5260.

[26] M. Kalakrishnan, L. Righetti, P. Pastor, and S. Schaal, "Learning Force Control Policies for Compliant Robotic Manipulation," *Int. Conf. Mach. Learn.*, pp. 4639–4644, Sep. 2012.

[27] S. Gu, T. Lillicrap, I. Sutskever, and S. Levine, "Continuous Deep Q-Learning with Model-based Acceleration," *arXiv Prepr. arXiv1603.00748*, Mar. 2016.

[28] J. Schulman, P. Moritz, S. Levine, M. Jordan, and P. Abbeel, "High-Dimensional Continuous Control Using Generalized Advantage Estimation," Jun. 2015.

[29] Y. Yang, Y. Li, C. Fermuller, and Y. Aloimonos, "Robot Learning Manipulation Action Plans by 'Watching' Unconstrained Videos from the World Wide Web," *Twenty-Ninth AAAI Conf. Artif. Intell.*, pp. 3686–3692, 2015.

[30] I. Popov *et al.*, "Data-efficient Deep Reinforcement Learning for Dexterous Manipulation," *arXiv Prepr.*, no. section V, 2017.

[31] A. Gupta, C. Eppner, S. Levine, and P. Abbeel, "Learning dexterous manipulation for a soft robotic hand from human demonstrations," in *Proceedings* 2016 *IEEE International Conference on Intelligent Robots and Systems*, 2016, pp. 3786–3793.

[32] C. Melchiorri, "Slip detection and control using tactile and force sensors," *IEEE/ASME Trans. Mechatronics*, vol. 5, no. 3, pp. 235–243, 2000.

[33] R. Fernandez, I. Payo, A. S. Vazquez, and J. Becedas, "Slip detection in robotic hands with flexible parts," in *Advances in Intelligent Systems and Computing*, 2014, vol. 253, pp. 153–167.

[34] T. Yoshikawa, "MANIPULABILITY OF ROBOTIC MECHANISMS.," *Int. J. Rob. Res.*, vol. 4, no. 2, pp. 3–9, Jun. 1985.

[35] N. Vahrenkamp, T. Asfour, G. Metta, G. Sandini, and R. Dillmann, "Manipulability analysis," *IEEE-RAS Int. Conf. Humanoid Robot.*, no. May, pp. 568–573, 2012.

[36] D. J. Montana, "Contact Stability for Two-Fingered Grasps," *IEEE Trans. Robot. Autom.*, vol. 8, no. 4, pp. 421–430, 1992.

[37] Z. Li and S. Sastry, "Task-oriented optimal grasping by multfingered robot hands," *IEEE J. Robot. Autom.*, vol. 4, no. 1, pp. 32–44, 1988.

[38] M. A. Roa and R. Suárez, "Grasp quality measures: review and performance," *Auton. Robots*, vol. 38, no. 1, pp. 65–88, Jan. 2015.

[39] J. L. Bentley and J. Louis, "Multidimensional binary search trees used for associative searching," *Commun. ACM*, vol. 18, no. 9, pp. 509–517, Sep. 1975.

[40] Pedregosa et al., "Scikit-learn: Machine Learning in Python," *J. Mach. Learn. Reserach*, no. 12, pp. 2825–2830, 2011.