2024 IEEE International Conference on Robotics and Automation (ICRA 2024)
May 13-17, 2024. Yokohama, Japan

Kinesthetic-based In-Hand Object Recognition with an Underactuated Robotic Hand

Julius Arolovitch*, Osher Azulay* and Avishai Sintov

***Abstract*— Tendon-based underactuated hands are intended to be simple, compliant and affordable. Often, they are 3D printed and do not include tactile sensors. Hence, performing in-hand object recognition with direct touch sensing is not feasible. Adding tactile sensors can complicate the hardware and introduce extra costs to the robotic hand. Also, the common approach of visual perception may not be available due to occlusions. In this paper, we explore whether kinesthetic haptics can provide in-direct information regarding the geometry of a grasped object during in-hand manipulation with an underactuated hand. By solely sensing actuator positions and torques over a period of time during motion, we show that a classifier can recognize an object from a set of trained ones with a high success rate of almost 95%. In addition, the implementation of a real-time majority vote during manipulation further improves recognition. Additionally, a trained classifier is also shown to be successful in distinguishing between shape categories rather than just specific objects.**

I. Introduction

Underactuated hands, characterized by their ability to conform to object shapes through compliance, offer simplicity and cost-effectiveness [1], [2]. This is in contrast to traditional robotic hands, while precise and proficient, are limited by their intricate design, high costs and complex control schemes [3], [4]. Underactuated hands have been shown to provide stable grasps through open-loop control [5] and an ability to perform precise in-hand manipulation tasks [6]–[8]. Some attempts have been made to also enable object recognition with underactuated hands [9], [10]. These, however, often require additional sensory hardware on the hand [11]–[13].

The recognition of an object grasped within a robotic gripper or hand has been a topic for numerous works [14], [15]. An ability to recognize object without visual perception increases the workspace of the robot and enables it to work in regions without proper lighting or line-of-sight [16]. Relying solely on continuous visual feedback limits the performance in various tasks where visual uncertainty or occlusion may occur. In such cases, it may be impossible to solve the task altogether. Examples include grasping an arbitrary object at the back of a cabinet or in confined spaces as demonstrated in Figure 1.

The common approach for object recognition is the use of haptics or, in particular, tactile sensing [17], [18]. In such sensing, tactile sensors have direct contact with the surface of the object and provide a glance of its geometry for recognition inference [19]. Often, the hand requires multiple tactile glances in order to provide a certain recognition regarding the object [20], [21]. However, adding tactile sensors to robotic hands can be expensive and may complicate the hardware. Kinesthetic haptics, on the other hand, gather information by sensing movement, force and position of actuator joints in the hand [22]. Prior work has combined tactile images along with kinesthetic information while utilizing k-means clustering to increase classification success rate [23]. More recently, kinesthetic haptics was used to recognize objects by observing finger kinematics during multiple grasps [24]. However, these assume rigid hands and full knowledge of the kinematic configuration of the fingers.

Fig. 1: In-hand recognition of an object in an occluded environment with an underactuated hand where visual perception is not available. The hand relies on kinesthetic perception during in-hand manipulation to either recognize the specific object from a trained set or its general shape.

Object recognition using underactuated hands has mostly relied on either visual methods or tactile perception [25]–[29]. In [30], embedded force sensors along the two fingers of an underactuated hand were introduced for object classification and feature extraction through single-grasp interactions. Similarly, optical tactile sensors, which observe deformation on the contact pad through an internal camera, were used

* These authors contributed equally.
J. Arolovitch is with the Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, USA. E-mail: jarolovi@andrew.cmu.edu
O. Azulay and A. Sintov are with the School of Mechanical Engineering, Tel-Aviv University, Israel. E-mail: osherazulay@mail.tau.ac.il, sintov1@tauex.tau.ac.il

on a three-finger underactuated hand for recognition of grasped objects [31]. Integrating these sensors into compliant hands adds extra costs and computational complexity to the design. This is even more crucial for open-source 3D printed underactuated robotic hands that attempt to provide a simple, low-cost and accessible solution [32].

In this paper, we address the problem of proprioceptive object recognition with underactuated hands by exploring it solely through kinesthetic haptics. Proprioception, or kinaesthesia, offers a unique way to recognize the state of the hand-object system without relying on conventional visual or tactile cues. We focus on a tendon-based underactuated hand where actuators at the base of the hand pull tendons running along the fingers, which also have passive joints and springs. With no tactile sensors on the fingers and no exact model for these compliant mechanisms [8], we observe signals from the actuators during in-hand manipulation of various objects and explore whether they embed crucial information regarding their shapes. The feature state representation and data-based modeling are explored.

A single proprioceptive state signal from the hand at some time instant is not sufficient for representing the shape of a grasped object. Therefore, we hypothesize that a sequence of states during in-hand manipulation could embed geometric information such as curvature, flat surface and corners. Consequently, we collect sequential data during the manipulation of various objects. With the data, our investigation delves into an extensive analysis of various classification models, either simple or temporal-based, that capture the intricate relationships between the hand's proprioceptive data and the underlying object geometries. Furthermore, we utilize a simple majority vote paradigm for increasing the certainty of the recognition in real-time during manipulation. Finally, the generalization capabilities of the models are explored, evaluating their performance on generalizing to categories of shapes rather than specific objects in the training set.

By focusing on an object recognition model without vision or tactile sensing, we provide a novel capability that aligns with simplicity and cost-effectiveness in hardware design. Without additional sensors, a low-cost underactuated robotic hand can pick-up an object and simply manipulate it between the fingers for a short period of time in order to recognize it. Hence, by circumventing the limitations of visual and tactile perception, we offer a pure algorithmic and potentially more robust approach to in-hand object recognition. While out of the scope of this work, the detection for a successful grasp at the pick-up phase can also be recognized through kinesthetic haptics. We note also that the proposed approach can be complementary to visual perception in order to increase recognition certainty or replace it entirely if a line-of-sight is not available.

II. Method

A. Problem Definition and Approach

Consider a two-finger underactuated hand as seen in Figure 2. The hand comprises of two opposing tendon-based fingers such that a manipulated object between the fingers performs a planar motion [8]. Each finger consists of two compliant joints with springs, and a tendon runs along its length. The tendon is connected to an actuator situated at the base of the hand and the finger flexes upon pulling of the tendon. The distal links of the fingers are equipped with high friction pads to prevent slipping. The hand has no tactile sensors on the fingers and no visual perception of the hand-object system is available. The hand can only measure kinesthetic features such as actuator angles and torques.

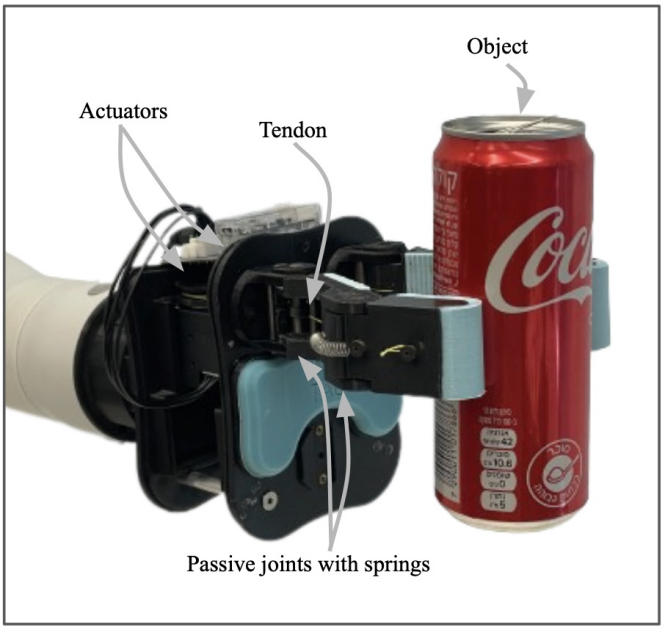

Fig. 2: An underactuated robotic hand (OpenHand Model-O [32]) with two-fingers. The hand is mostly 3D printed and is tendon-based. Each finger consists of two passive joints with springs. The tendons run along the length of the two fingers and are pulled by two actuators at the base of the hand.

Let $\mathbf{x} \in \mathcal{C}$ represent the observable state of the hand where $\mathcal{C} \subset \mathcal{R}^n$ is some n-dimensional state space. Similarly, vector $\mathbf{a} \in \mathcal{U}$ denotes an action exerted on the hand where $\mathcal{U} \subset \mathbb{R}^2$ is the action space. The action vector $\mathbf{a}$ corresponds to angle changes of the two actuators, signifying tendon pull or release, over a fixed time step Δt. Since no visual perception is used, the true state of the hand and the pose of the object $\mathbf{T} \in SE(2)$ with respect to the hand are not known and cannot be observed directly [16].

Given a set of m objects $\{\mathcal{O}_1, \dots \mathcal{O}_m\}$, it is required to identify an object grasped by the hand from the set without any use of visual feedback nor tactile sensors. The query object will be identified during the grasp without releasing and regrasping it. It is assumed that all grasped objects are rigid. Furthermore, there is no assumption on the pose $\mathbf{T}_0$ of the object during the initial grasp. The main objective of this work is, therefore, twofold. First, we aim to identify the set of n kinesthetic features that best represent relevant information regarding the shape of the grasped object. Second, we wish to utilize the features and explore the ability of a data-based model to classify grasped

objects. Consider state measurements arriving sequentially $\{\mathbf{x}_1, \ldots, \mathbf{x}_k\}$ in real-time while manipulating a given query object $\mathcal{O}$ whose class is unknown. It is required to solve the following maximization problem

$$j^* = arg\max_j P(\mathcal{O}_j|\mathbf{x}_1, \ldots, \mathbf{x}_k) \tag{1}$$

where $P(\mathcal{O}_j|\mathbf{x}_1, \ldots, \mathbf{x}_k)$ is the conditional probability for $\mathcal{O}$ to be in class $\mathcal{O}_j$.

B. Data Representation

Data is collected by recording various state transitions during in-hand manipulation of the m objects. An observable state of the hand during manipulation is a kinesthetic measurement of internal features. In the tested hand, the state can consist of actuator torques and angles. In addition, we consider the case where the applied action at the given time is included in the state. Hence, the state $\mathbf{x}$ can be of dimension up to $n = 6$. The collection of data is conducted in an episodic manner where, in each episode, an object is manipulated by exerting a set of pre-defined actions until drop. Hence, data for episode i is a set of states $\mathcal{E}_i = \{\mathbf{x}_0, \ldots, \mathbf{x}_z\}$ and the corresponding object label $y_i \in \{1, \ldots, m\}$. In order to overcome biases throughout different episodic sessions, the initial state of an episode is subtracted from all states in the episode, i.e., $\tilde{\mathbf{x}}_j = \mathbf{x}_j - \mathbf{x}_0$. Subsequently, an optimized low pass filter is applied to each episode to alleviate sensor noise. Consequently, dataset $\mathcal{D}$ is of the form

$$\mathcal{D} = \{(\tilde{\mathcal{E}}_i, y_i)\}_{i=1}^{N_e} \tag{2}$$

consisting of N_e episodes and where $\tilde{\mathcal{E}}_i = \{\mathbf{0}, \tilde{\mathbf{x}}_1, \ldots, \tilde{\mathbf{x}}_z\}$.

C. Data-based Model and Inference

The recognition of a grasped object without tactile sensors cannot be done with only a single kinesthetic state $\mathbf{x}_i$. Such state will not have sufficient information regarding the shape of the object. Alternatively, state change during manipulation may have embedded geometric information such as corners and surface curvature across the object. Hence, we hypothesize that a sequential set of states $\{\mathbf{x}_1, \ldots, \mathbf{x}_k\}$ would enable the training of a model for the desired recognition.

Sequential data enables to learn the motion pattern of the hand-object system over some period of time. Given a set of w sequential states along a motion time frame $S_w = \{\mathbf{x}_1, \ldots, \mathbf{x}_w\}$, we search for a data-based classification model that could provide a solution for (1). In practice, we aim to train a model $\Gamma_\psi : \mathcal{C} \times \ldots \times \mathcal{C} \rightarrow [0, 1]^m$, where ψ is the vector of trained parameters, to provide a probability distribution over the m possible objects in the form

$$\Gamma_\psi(S_w) = [P(\mathcal{O}_1|S_w), \ldots, P(\mathcal{O}_m|S_w)]\,. \tag{3}$$

Then, the predicted object would be $\mathcal{O}_j$ with

$$j^* = arg\max_j \Gamma_\psi(S_w)] \tag{4}$$

where $arg\max_i \mathbf{v}$ returns the index of the component in vector $\mathbf{v}$ with the maximum value. In order to train model Γ_ψ, dataset $\mathcal{D}$ is modified to include labeled motion sequences of pre-defined length w. Along each episode $\tilde{\mathcal{E}}_i$ of length z, a window of length w is moved to generate $z - w + 1$ sequences $\{S_{w,i}^{(1)}, \ldots, S_{w,i}^{(z-w+1)}\}$. Each sequence $S_{w,i}^{(j)}$ is labeled with the original label y_i yielding a training set $\mathcal{P} = \{(S_{w,i}, y_i)\}_{i=1}^{N}$.

Training a sequential model with $\mathcal{P}$ for Γ_ψ in (3) can be done with data-based architectures designated for sequence or temporal modeling, such as the Long Short-Term Memory (LSTM) [33] and Temporal Convolutional Network (TCN) [34]. LSTM is a class of Recurrent Neural-Networks (RNN) aimed to learn sequential data and is able to selectively retain or discard information from previous time steps making it well-fit for long-term dependencies. TCN is a type of neural-network that uses convolutional layers to process sequential data. The convolutional layers extract significant features from the data. By using dilated convolutions, the TCN is able to capture long-term dependencies in a computationally efficient manner, making it a popular and efficient choice for temporal prediction tasks.

Each of the above models and other classifiers will output a probability distribution as in (3) and a prediction according to (4). However, the maximum probability for a specific object class may not be sufficiently high in order to have a high certainty prediction. Nevertheless, one may exploit the continuous time frame in which the hand manipulates a certain object and rapidly acquire additional samples while being certain that they originate from the same object. Hence, a majority vote can be used to accumulate probabilities over time [35]. Let $\mathbf{p}_t = \Gamma_\psi(S_{w.i}^{(t)})$ be the probability distribution over the m object at time instance t. An accumulated prediction after T time frames is the object $\mathcal{O}_{i^*}$ that acquires the maximum sum of scores given by

$$i^* = arg\max_i \sum_{t=1}^{T} \mathbf{p}_t. \tag{5}$$

In this way, the model can keep accumulate and improve its prediction over time.

III. Experiments

A. Experimental Setup

An experimental system was constructed based on the two opposing fingers of the OpenHand Model-O [32] underactuated hand as seen in Figure 2. The experimental system consists of the hand and an automated reset mechanism. Once the object drops from the fingers, a thin string that runs through a hole in the center of the object pulls it into the reach of the fingers toward a new random grasp. A camera and a fiducial marker on the object are used solely to confirm the successful initial grasp in each episode. An action is defined to be $\mathbf{a} \in \{0^\circ, 1^\circ, -1^\circ\} \times \{0^\circ, 1^\circ, -1^\circ\}$ where, in each step, the actuators are stalled or moved by a constant amount. The system is operated by the Robot Operating System (ROS).

Eight objects, seen in Figure 3, are used for training and evaluating the classification model. Seven of the objects are prismatic PLA ones with different cross-sections including square, star-like, arbitrary-curved, half-circular, circular, hexagonal and elliptical. The eighth object is a rubber duck.

Fig. 3: Eight distinct objects used for training the classification model.

Data was collected for the object where, in each episode, the hand grasps the object, performs an in-hand manipulation with a pre-defined action sequence for 10 seconds and then drops the object. The hand has no tactile sensors while only actuator angles and torques can be measured. During manipulation, data stream of these features is available in 10 Hz and recorded including instantaneous actions.

TABLE I: Mean recognition success rate for various feature combinations and models

Feature combination (n)	LSTM (%)	TCN (%)	FC-NN (%)	RF (%)	SVM (%)
Torques (2)	38.88	52.55	49.75	62.42	48.27
Torques and actions (4)	65.54	76.8	73.04	74.73	69.94
Torques and angles (4)	90.02	94.86	93.49	93.88	82.57
Torq., ang. and act. (6)	83.06	94.03	87.64	92.35	79.66

B. Object Recognition

We begin by evaluating the classification accuracy over the eight objects. Dataset $\mathcal{D}$ was collected over $N_e = 5,010$ episodes with approximately 625 episodes per object. The dataset was processed with the full length of the sequences, i.e., $w = 100$, yielding a training dataset $\mathcal{P}$ with $N = 5,010$ labeled sequences and 501,000 state samples. An additional test dataset of 860 labeled sequences was collected and not included in the training set in any way.

We compare between several classification models including LSTM and TCN. While these are designated for temporal modeling, we also compare general classifiers including Random Forest (RF), Support Vector Machine (SVM) and a fully-connected NN (FC-NN). The hyper-parameters of the models were optimized yielding the following architectures. The TCN consists of three layers, two with 256 neurons, one with 128 neurons and a ReLU activation in between. A dropout of 10% and an L1 regularizer with a factor of 10^{-3} were included to reduce over-fitting. Learning rate scheduling was also utilized, starting with a learning rate of 0.0001 and terminating at a rate of 10^{-6}. The LSTM also consists of three layers consisting of 128, 64 and 32 neurons. A dropout of 5% and L1 regularizer with a factor of 10^{-2} were used to decrease over-fitting. A learning rate scheduler was also used, initializing at 0.001 and terminating at a rate of 10^{-6}. While LSTM and TCN can receive sequential input, the input for RF, SVM and FC-NN is a single vector. Hence, each sequence $S_{w,i}$ was flattened to a vector of size wn. The Random Forest consisted of 100 decision trees. The FC-NN consists of two layers of 128 and 64 neurons. A dropout of

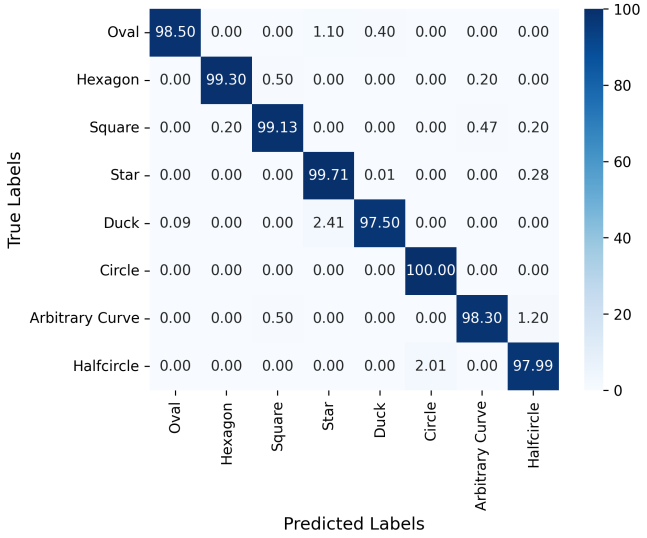

Fig. 4: Confusion matrix for classification of the eight objects with the TCN model over the test data.

10% and L1 regularizer with a factor of 10^{-4} were used. A learning rate scheduler was also used, initialized at 0.001 and terminated at a rate of 10^{-6}.

Table I summarizes the results of the four models over the test data with four varying feature combinations of actuator angles, torques and actions. First, it is clear that torque signals along the motion are not sufficient in order to characterize the shape of the manipulated object. Including the actions provides marginal improvement since they are rather sparse. However, the best results are achieved with full state of the actuators having torques and angles. This four-dimensional state formulation was used in subsequent testing. When comparing between the five classification models, TCN with torques-angles state achieves the highest success with marginal advantage compared to FC-NN and RF. The results show that even simpler classifiers can successfully recognize object during in-hand manipulation. SVM provides inferior results to all models. Figure 4 exhibits the confusion matrix for classifying the eight objects with the TCN model. Results show high recognition rate for all objects in the set.

We next observe performance of the models with regards to the size of the training data. Recall that the full size of the training data is $N = 5,010$ sequences amounting to approximately 16 hours of automated data collection. For a fair analysis, models were trained on varying portions of the data where, in each evaluation, data is picked from $\mathcal{D}$ without any shuffling so to maintain sequential order. For each data size, the model is re-trained 20 times on different parts of the data and the success rate is averaged. Results for success rate with regards to portion of the data in $\mathcal{P}$ used to train the model can be seen in Figure 5. Success rate saturation is reached for all models with about 60% of the dataset which corresponds to approximately 10 hours of data collection. TCN reaches the highest value in saturation whereas RF has higher values with a lower amount of training data. SVM

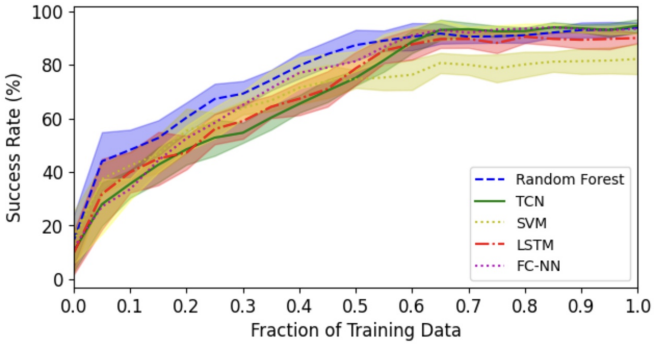

Fig. 5: Object recognition success rate with regards to the fraction of total training data used to train RF, LSTM, SVM, FC-NN and TCN.

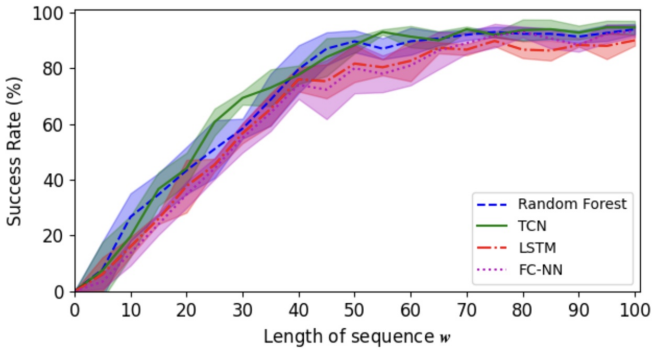

Fig. 6: Object recognition success rate with regards to the length of the sequence w in training and testing models TCN, LSTM, FC-NN and RF.

performed the poorest and is not used in further evaluations.

The above trained models used sequence length $w = 100$ for comparison. Analysis is now presented on the performance with regards to the value of w. Here also, for each value of w, the model was re-trained 20 times and the success rate was averaged. Figure 6 depicts the success rate with regards to the length of the sequence w over the test data. All models saturate with a sequence length $w > 75$. This result shows that sufficient motion must be exerted in order to acquire informative data for accurate recognition.

The use of the majority vote in order to accumulate certainties over objects and improve success rates is further evaluated. For the majority vote, we use a sequence size of $w = 75$ and slide it over 25 inferences in order to accumulate probability distributions according to (5). The majority vote is compared to single inferences with sequences of length $w = 75$ and $w = 100$. Table II presents the results over four classification models. The results show marginal improvement when using the whole episode length of $w = 100$ over a shorter one $w = 75$ in a single inference classification. However, exploiting the entire episode but with a majority vote over sequences of $w = 75$ provides extremely high success rates. Hence, even an ill-trained classifier can be used in a majority vote setting in order to improve predictions. In general, it is more advantageous to train a model with shorter sequence length and with majority vote, rather then having

TABLE II: Classification success rate of the eight objects for single inferences and with majority votes

		LSTM (%)	TCN (%)	RF (%)	FC-NN (%)
Single inference	$w = 75$	89.80	93.17	91.67	93.00
Single inference	$w = 100$	90.02	94.86	93.88	93.49
Majority vote	$w = 75$	93.60	99.90	98.99	99.67

Fig. 7: Circular household objects used for testing the generalization ability of the trained classifiers and their diameters.

single inference with a longer sequence. The performance of the majority vote in real-time and with various sequence lengths will be analyzed later on.

C. Generalized Recognition

The ability of a model to classify everyday objects into shape categories and not only specific objects is now evaluated. We consider four shape categories: circular (7 objects), semi-circular (6 objects), square (9 objects) and elliptical (7 objects) prisms. A set of household test objects of varying sizes and within the categories was collected for benchmarking the generalization ability of the trained models. An example for the set of circular objects is seen in Figure 7. A successful recognition is the one that matches the shape category of a test object to the corresponding train object (seen in Figure 3). Table III compares success rates between single inferences and majority vote for the shape categories. The classification of entire sequences ($w = 100$) proved to be the least efficient in generalization, less so than classifying sequences of length $w = 75$. The majority vote proves to be resilient in overcoming shape and size differences. Table IV presents the size ranges with respect to the train objects that reached a success rate above 75% and the actual success rates for majority votes of $w = 75$ sequences. The results show a distinctive ability for the trained models to recognize shapes of new objects not included in the training and with different sizes.

TABLE III: Classification success rate of four shape categories for single inferences and with majority votes

		LSTM (%)	TCN (%)	RF (%)	FC-NN (%)
Single inference	$w = 75$	80.63	87.76	95.50	94.42
Single inference	$w = 100$	76.88	87.38	91.75	90.61
Majority vote	$w = 75$	89.59	94.55	98.00	96.04

TABLE IV: Size range with respect to the training objects and success rate for recognizing the shape category of new objects with the TCN classifier

Shape category	Size Range (%)	Success rate (%)
Circular	36.0–160.0	91.5
Semi-Circular	56.0–181.5	95.5
Square	43.5–135.3	86.0
Elliptical	79.3–145.6	94.0

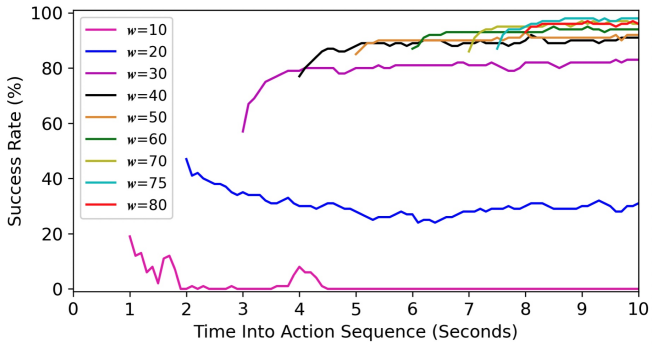

Fig. 8: Success rate of shape recognition in during the manipulation of various objects using real-time majority votes.

D. Real-Time Object Recognition

In the last experiment, we evaluate the ability of the model to conduct live object recognition early into the manipulation sequence and the prediction improvement with the majority vote. While we discuss real-time inference, we simulate real-time inference along the recorded test episodes and evaluate recognition success rate, focusing particularly on the model's ability to conduct generalized real-time recognition having been trained on a set of generic objects. Since a recorded episode is of length 10 seconds or 100 steps, a majority vote would have $T = 100 - w$ instances to collect predictions and make inferences according to (5). Let $S_w^{(t)}$ be a sequence of length w arriving at time t. Hence, once the first sequence $S_w^{(1)}$ arrives, an initial inference can be computed. Then, as additional states arrive, more sequences are acquired $\{S_w^{(2)}, S_w^{(3)}, \ldots, S_w^{(T)}\}$ and predictions can be improved by updating the majority vote in (5).

Figure 8 presents the success rate analysis with the updating of the majority vote along time for a series of trials of general shape recognition. The analysis compares between different lengths of sequences used in inferences for the majority voting. First, having $w \leq 20$ yields a decline in success rate. This is due to the significantly ill-trained and faulty classifier trained over short sequences with insufficient information. The conditions for a classifier to guarantee prediction improvement in a majority vote are given in [24]. Classifiers with longer sequences are shown to improve predictions over time while reaching saturation at some point. Longer sequences, however, can provide a first prediction later in the episode. Sequences of approximately $w = 75$ provide the highest success rate almost from the first inference with 88% to saturation at 98%. Figures 9-11 show snapshots of real-time object recognition inferences with three different objects. The motion that the object-hand system has to exert is rather small before a high certainty recognition is acquired. Hence, the hand can quickly move on to performing the desired task.

Fig. 9: Snapshots of real-time object recognition of an hexagon prism over 25 time steps and 2.5 seconds. The model certainty of grasping the prism after 80, 90 and 100 time steps is 0.898, 0.914 and 0.920, respectively.

Fig. 10: Snapshots of real-time object recognition of a rubber duck over 25 time steps and 2.5 seconds. The model certainty of grasping the duck after 80, 90 and 100 time steps is 0.923, 0.961 and 0.969, respectively.

Fig. 11: Snapshots of real-time object recognition of a soda can over 25 time steps and 2.5 seconds. The model certainty of grasping the can after 80, 90 and 100 time steps is 0.726, 0.776 and 0.812, respectively.

IV. Conclusions

In this work, we have explored the ability of a tendon-based compliant hand to recognize grasped objects without any tactile sensing nor visual perception. We have shown that using only kinesthetic haptics during in-hand manipulation can provide sufficient information regarding the shape of the objects. In other words, with only position and torque data of the hand's internal actuators during motion with the objects, one can train a classifier to distinguish between the objects. A set of classification models were benchmarked including temporal ones. The results show best performance using the TCN with a marginal advantage over simple FC-NN or RF. In addition, majority vote with constant inferences during the hand-object motion is shown to increase prediction certainty and success rate. The proposed approach is also able to recognize a geometry characteristic of an object leading to a more general recognition model. The findings of this study have the potential to augment the capabilities of low-cost and 3D printed underactuated hands without any requirement for additional sensing hardware.

REFERENCES

[1] L. U. Odhner and A. M. Dollar, "Dexterous manipulation with underactuated elastic hands," in *IEEE Int. Conf. on Rob. and Aut.* IEEE, May 2011, pp. 5254–5260.

[2] O. Azulay, M. Monastirsky, and A. Sintov, "Haptic-based and $SE(3)$-aware object insertion using compliant hands," *IEEE Robotics and Automation Letters*, vol. 8, no. 1, pp. 208–215, 2023.

[3] Y. Bai and C. K. Liu, "Dexterous manipulation using both palm and fingers," pp. 1560–1565, 2014.

[4] R. Michalec and A. Micaelli, "Stiffness modeling for multi-fingered grasping with rolling contacts," in *IEEE-RAS International Conference on Humanoid Robots*, 2010, pp. 601–608.

[5] A. M. Dollar and R. D. Howe, "The highly adaptive sdm hand: Design and performance evaluation," *The International Journal of Robotics Research*, vol. 29, no. 5, pp. 585–597, 2010.

[6] B. Calli and A. M. Dollar, "Vision-based precision manipulation with underactuated hands: Simple and effective solutions for dexterity," in *IEEE/RSJ Int. Conf. on Intel. Rob. and Sys.*, 2016, pp. 1012–1018.

[7] B. Calli, A. Kimmel, K. Hang, K. Bekris, and A. Dollar, "Path planning for within-hand manipulation over learned representations of safe states," in *International Symposium on Experimental Robotics*, Buenos Aires, Argentina, 2018.

[8] A. Sintov, A. S. Morgan, A. Kimmel, A. M. Dollar, K. E. Bekris, and A. Boularias, "Learning a state transition model of an underactuated adaptive hand," *IEEE Robotics and Automation Letters*, vol. 4, no. 2, pp. 1287–1294, April 2019.

[9] J. M. Gandarias, J. M. Gómez-de Gabriel, and A. J. García-Cerezo, "Enhancing perception with tactile object recognition in adaptive grippers for human–robot interaction," *Sensors*, vol. 18, no. 3, 2018.

[10] Z. Zhou, R. Zuo, B. Ying, J. Zhu, Y. Wang, X. Wang, and X. Liu, "A sensory soft robotic gripper capable of learning-based object recognition and force-controlled grasping," *IEEE Transactions on Automation Science and Engineering*, pp. 1–11, 2022.

[11] M. V. Liarokapis, B. Calli, A. J. Spiers, and A. M. Dollar, "Unplanned, model-free, single grasp object classification with underactuated hands and force sensors," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2015, pp. 5073–5080.

[12] C. Jiao, B. Lian, Z. Wang, Y. Song, and T. Sun, "Visual–tactile object recognition of a soft gripper based on faster region-based convolutional neural network and machining learning algorithm," *International Journal of Advanced Robotic Systems*, vol. 17, no. 5, p. 1729881420948727, 2020.

[13] V. P. da Fonseca, X. Jiang, E. M. Petriu, and T. E. A. de Oliveira, "Tactile object recognition in early phases of grasping using underactuated robotic hands," *Intelligent Service Robotics*, vol. 15, no. 4, pp. 513–525, 2022.

[14] S. E. Navarro, N. Gorges, H. Wörn, J. Schill, T. Asfour, and R. Dillmann, "Haptic object recognition for multi-fingered robot hands," in *IEEE Haptics Symposium (HAPTICS)*, 2012, pp. 497–502.

[15] T. Watanabe, K. Yamazaki, and Y. Yokokohji, "Survey of robotic manipulation studies intending practical applications in real environments -object recognition, soft robot hand, and challenge program and benchmarking-," *Advanced Robotics*, vol. 31, no. 19-20, pp. 1114–1132, 2017.

[16] O. Azulay, I. Ben-David, and A. Sintov, "Learning haptic-based object pose estimation for in-hand manipulation control with underactuated robotic hands," *IEEE Transactions on Haptics*, vol. 16, no. 1, pp. 73–85, 2022.

[17] G. Rouhafzay and A. Cretu, "Object recognition from haptic glance at visually salient locations," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 3, pp. 672–682, 2020.

[18] F. Pastor, J. García-González, J. M. Gandarias, D. Medina, P. Closas, A. J. García-Cerezo, and J. M. Gómez-de Gabriel, "Bayesian and neural inference on lstm-based object recognition from tactile and kinesthetic information," *IEEE Robotics and Automation Letters*, vol. 6, no. 1, pp. 231–238, 2021.

[19] A. Drimus, G. Kootstra, A. Bilberg, and D. Kragic, "Design of a flexible tactile sensor for classification of rigid and deformable objects," *Robotics and Autonomous Systems*, vol. 62, no. 1, pp. 3 – 15, 2014.

[20] M. Jin, H. Gu, S. Fan, Y. Zhang, and H. Liu, "Object shape recognition approach for sparse point clouds from tactile exploration," in *IEEE International Conference on Robotics and Biomimetics (ROBIO)*, 2013, pp. 558–562.

[21] E. Kirby, R. Zenha, and L. Jamone, "Comparing single touch to dynamic exploratory procedures for robotic tactile object recognition," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 4252–4258, 2022.

[22] J. Carter and D. Fourney, "Research based tactile and haptic interaction guidelines," in *Guidel. On Tactile and Hap. Inter.*, 01 2005, pp. 84–92.

[23] S. Luo, W. Mou, K. Althoefer, and H. Liu, "iCLAP: shape recognition by combining proprioception and touch sensing," *Autonomous Robots*, 06 2018.

[24] A. Sintov and I. Meir, "Simple kinesthetic haptics for object recognition," *The International Journal of Robotics Research*, vol. 42, no. 7, pp. 537–561, 2023.

[25] G. Li, S. Liu, L. Wang, and R. Zhu, "Skin-inspired quadruple tactile sensors integrated on a robot hand enable object recognition," *Science Robotics*, vol. 5, no. 49, p. eabc8134, 2020.

[26] Z. Flintoff, B. Johnston, and M. Liarokapis, "Single-grasp, model-free object classification using a hyper-adaptive hand, google soli, and tactile sensors," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2018, pp. 1943–1950.

[27] J. Huang and A. Rosendo, "Variable stiffness object recognition with a cnn-bayes classifier on a soft gripper," *Soft Robotics*, vol. 9, no. 6, pp. 1220–1231, 2022.

[28] N. Hanson, H. Hochsztein, A. Vaidya, J. Willick, K. Dorsey, and T. Padir, "In-hand object recognition with innervated fiber optic spectroscopy for soft grippers," in *IEEE International Conference on Soft Robotics (RoboSoft)*, 2022, pp. 852–858.

[29] J. Cao, J. Huang, and A. Rosendo, "Variable stiffness object recognition with bayesian convolutional neural network on a soft gripper," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2022, pp. 9431–9436.

[30] A. J. Spiers, M. V. Liarokapis, B. Calli, and A. M. Dollar, "Single-grasp object classification and feature extraction with simple robot hands and tactile sensors," *IEEE transactions on haptics*, vol. 9, no. 2, pp. 207–220, 2016.

[31] J. W. James, A. Church, L. Cramphorn, and N. F. Lepora, "Tactile model o: Fabrication and testing of a 3d-printed, three-fingered tactile robot hand," *Soft Robotics*, vol. 8, no. 5, pp. 594–610, 2021.

[32] R. R. Ma and A. M. Dollar, "Yale openhand project: Optimizing open-source hand designs for ease of fabrication and adoption," *IEEE Rob. & Aut. Mag.*, vol. 24, pp. 32–40, 2017.

[33] Y. Yu, X. Si, C. Hu, and J. Zhang, "A Review of Recurrent Neural Networks: LSTM Cells and Network Architectures," *Neural Computation*, vol. 31, no. 7, pp. 1235–1270, 07 2019.

[34] S. Bai, J. Z. Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," in *Conference on Robotics Research*, vol. abs/1803.01271.

[35] N. D. Kahanowich and A. Sintov, "Robust classification of grasped objects in intuitive human-robot collaboration using a wearable force-myography device," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 1192–1199, 2021.