Evaluating Data Representations for Object Recognition During Pick-and-Place Manipulation Tasks

Humberto Navarro de Carvalho*, Lucas Pontes Castro‡, Thaís G. do Rego*, Telmo M. Silva Filho‡, Yuri de A. M. Barbosa*, Leonardo Vidal Batista*, Amilcar Soares†, Vinicius Prado da Fonseca†
*Center of Informatics, Federal University of Paraíba, João Pessoa, Brazil
‡Department of Statistics, Federal University of Paraíba, João Pessoa, Brazil
†Department of Computer Science, Memorial University of Newfoundland, St John's Canada.
vpradodafons@mun.ca

***Abstract*—When manipulating objects, robots need to build a local and global description of the environment simultaneously. Recognizing objects and estimating their pose are examples of tasks expected from robots when operating in unstructured environments. An efficient solution to these tasks has the potential to increase robotic usage in such settings. This paper presents a study on the representation of tactile and joint-position data to recognize everyday objects. We performed 12 different experiments extracting features in different ways from a publicly available dataset. More specifically, this work uses 4 data representations, namely 3 Points, 10 Points, Average and Descriptive Statistics (DS) over 2 different sensor types (i.e., positional and tactile sensors separately) and the combination of both. Using these data representations, we trained and evaluated machine learning models in the object recognition task. Our findings support that tactile data and its combination with finger joint position information can be successfully used for object identification during manipulation tasks. The feature engineering approach used to represent the dataset used in this paper showed promising results regarding recognizing objects using a combination of tactile and joint-position information. Our exploratory analysis testing different data representations was crucial for improving objects' recognition starting from a low of accuracy 30.31% (using data from the positional sensor only with sampled averages) to a high performance of 93.53% accuracy (using an Extra Tree classifier trained on data from all sensors with a DS representation).**

***Index Terms*—tactile sensing, object recognition, robotic manipulation, applied machine learning**

I. Introduction

The rapid technological advances of recent years have led to increased expectations from consumers concerning robotic applications. Moreover, modern robotic applications, such as home assistants, automated manufacturing, robotic surgery, and intelligent prostheses increasingly require efficient robotic manipulation in unstructured environments [1], [2]. Despite the recent success of robotic applications on all aforementioned application domains, there are still several open challenges when it comes to robotics in unstructured environments. For instance, Eppner et al. [2] reported manipulation failures such as missed object recognition, objects stuck at removal, and missing small objects, which were addressed in the Amazon Picking Challenge. Unfortunately, current robotic platforms are not able to execute tasks efficiently in the presence of environmental uncertainties. Notwithstanding the extensive research and technological efforts, current automated platforms usually require structured conditions and precise timing to perform dexterous manipulation.

One approach towards reducing this knowledge gap is finding global and local solutions for robotic manipulation stages. Therefore, robotic manipulation can be divided into action phases. For example, objects are grasped, moved, brought into contact with other objects, and released [3]. In addition, these phases usually include events that are sub-goals of the task. For example, early phases, such as grasping, require recognizing the object or estimating its position.

This work evaluates a data-driven approach for object recognition during pick-and-place tasks using tactile and finger joint position information. Our approach evaluates the use of sensor data collected at different manipulation stages, such as initial, stable, and final, to represent the object under grasp. An exploratory representation, using data sampled ten times between the initial and final stages, is also evaluated. We also evaluate a statistical descriptive representation using sensor readings of the complete manipulation task.

In summary, the present paper's main objective is to evaluate tactile and joint position signal representations for training machine learning models in object recognition. The three representations evaluated are: (a) three samples, from tactile and joint position readings from initial, stable, and final stages; (b) a collection of exploratory representations with tactile and joint position readings from ten samples between initial and final stages; and (c) extracting descriptive statistics, including the average, standard deviation, maximum value, and the third quartile from the signal data.

These data representation techniques were applied to extract nine representations from a tactile dataset with 2550 manipulation experiments performed on ten different objects by a robotic hand. The manipulation experiments were conducted at different positions, pressures and with objects of different weights. Throughout each experiment, 16 pressure and 8 joint position sensors are present in the apparatus that produce the data used to train machine learning models for object recognition in this work. The dataset used in this work allows us to train models with different data representations and organizations, impacting the models' capacity to recognize objects. Training machine learning models to recognize objects will improve later phases of robotic manipulation such as object translation or placing.

The rest of this paper is organized as follows. Section II

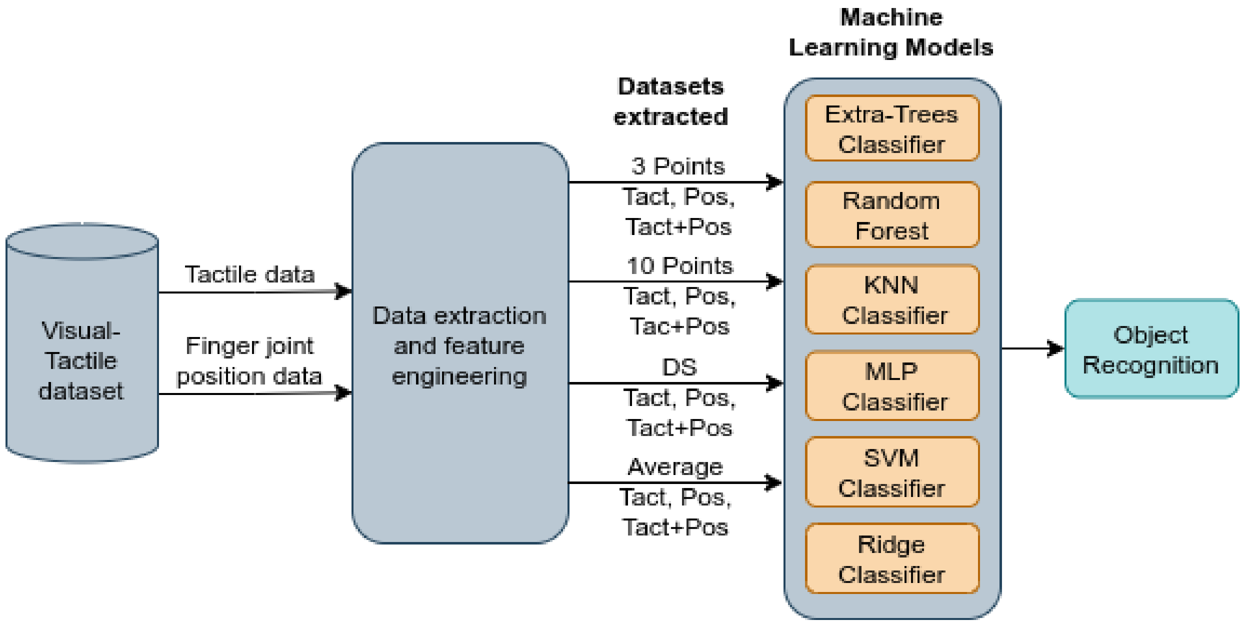

Fig. 1. An overview of the workflow of the experiments performed in this work.

presents a literature review of robotic manipulation in unstructured environments. Section III details the techniques designed in this paper, the dataset used, and the data processing and machine learning methods. The accuracy of the methods using the representations discussed here is discussed in Section IV. Finally, Section V presents a discussion concerning our data approach and future works.

II. Related Works

Robots interacting in our everyday environments require efficient and reliable robotic manipulation. Unstructured settings, such as homes, universities, and hotels, are predominantly unpredictable. In such environments, unknown friction, lack of sensor calibration, and other challenges make it difficult for robots to interact with objects with precision and efficiency [2]. Improving sub-tasks in manipulation phases, such as object recognition and pose estimation, can enhance overall manipulation, providing robots with local and global information about the environment.

A common approach for increasing the robotic manipulation precision, in general, is to provide the robot with manipulation data to investigate the execution of the grip analytically. Authors in [4], [5], [6] define grip metrics using known models to estimate the object's geometry, the environment, and the robot's grip. Although these methods provide a variety of information about the grip's physical interactions, their performances depend on how well the real-world system fits into the conditions of the analytical model. Alternatively, data-driven approaches have attempted to predict manipulation outcomes by understanding human vision [7], [8], applying simulation [9], [10], [11], or autonomous robotic data [12], [13] typically using visual or depth-camera feedback.

In addition to global vision information, robots enabled with touch detection can provide local information during gripping. Tactile sensors used for stable grasping control can also provide object recognition. Previous works, such as [14], [15], show that using purely visual-tactile information allows the model to continually re-plan what action to take to grab the object better. Even though these models can understand a wide range of unknown objects and have a high success rate, visual-tactile approaches are affected by occlusion or cluttered environments. A tactile approach to object recognition promotes local object information in such situations where vision may lead to incorrect classification.

Algorithms tested using exclusively tactile sensing promote local data integration even without vision. The difficulty of finding a closed-form mathematical solution for manipulation problems and the ability of machine learning methods to learn from previous observations lead to their adoption, along with other Artificial Intelligence (AI) techniques (e.g., fuzzy-logic controllers), for this purpose. For instance, Spiers et al. [16] developed a single grasping object classification approach to collect tactile data on the early phases of an open-loop grasping task. Molchanov et al. [17] performed contact localization, while Paolini et al. [18] developed post-grasp manipulation, confirming that a data-driven approach can improve the high complexity of in-hand manipulation [19]. Therefore conventional robotic grasp positions (e.g., power grasp) combined with multi-modal object recognition via power grasping of objects improves orientation recognition of unknown objects under manipulation [20], [21].

Unlike the above mentioned methods, which focus on extracting object geometry or video and can be computationally intensive, the strategies proposed here require less sensor data. As a result, they can provide fast object recognition during initial manipulation phases, improving the overall manipulation process. Moreover, tactile and joint position data are likely to be already available for related systems such as stable grasping.

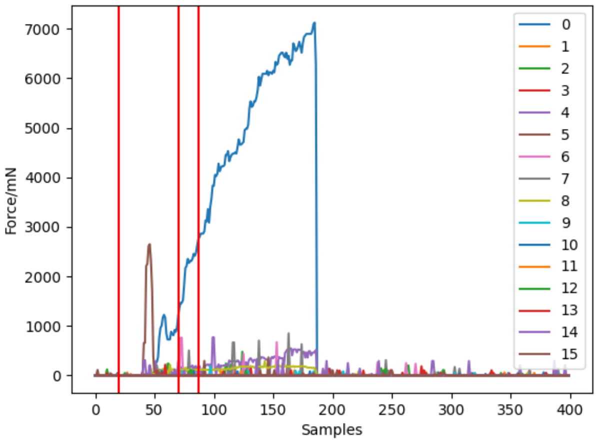

Fig. 2. The 3 manipulation sample points over tactile data used for object “Cheez”, weighing 100 grams and grasped from the right with 50 mN of force.

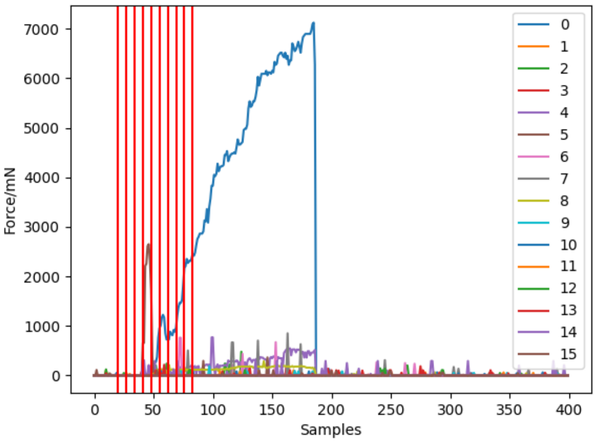

Fig. 3. The 10 manipulation sample points over tactile data used for object “Cheez”, weighing 100 grams and grasped from the right with 50 mN of force.

III. Materials and Methods

This work used a publicly available dataset [21] to test different data representation techniques for object recognition. The dataset contains visual, tactile, and finger joint position information of robotic pick-and-place tasks. The experiment uses an Eagle Shoal robot hand attached as the end-effector of a UR5 robot arm (Universal Robots). The Eagle Shoal robot hand has three fully actuated fingers and one fully actuated palm and has a total of eight degrees of freedom. The finger joint angle is referenced in the dataset [21] and further in this paper as position data. Each finger and palm also has four tactile sensors, so the dataset provides a total of 16 tactile sensor readings.

The dataset contains ten objects of different sizes and shapes (e.g., cuboid, cylinder, and others) that allow the addition of grains or water to change the object’s weight. In addition to the weight, the dataset also contains different grasping methods defined by grasping from three directions, including back, right, and top. For each object, the dataset provides: (i) a tentative manipulation changing object weights (empty, half, and full), (ii) different target forces (50, 100, and 150), and (iii) three approach directions (top, right, and back). The dataset has a known class imbalance due to some objects lacking data from one manipulation direction (e.g., the *cheez* box is too big to manipulate from the front) or object shape (e.g., cylinders do not differ from right or back) due to hand size limitation. In this paper, each pick-and-place try is referred as a single experiment run. There are ten runs for each combination of weight, target force, and direction. Each run contains 400 samples (24 seconds at 16.7Hz) with 16 pressure sensors and 8 finger joint positions. The dataset provides the tactile and finger joint position information in a CSV format.

The methods used in this work are detailed in Figure 1. First, for each run present in the dataset, we extracted the position and tactile information generating a raw dataset. Then, the raw data was used to create the four distinct data representations discussed in Section III-A. Finally, we used the new data representations to train and evaluate machine learning models for the object recognition task during pick-and-place tasks.

A. Data representation

The raw data used to evaluate our approach was collected from [21] and used to create four different data representations:

1) Three sample points (3-points) representation: This representation contains data samples for three different time instances, including, initial, stable and final times. The word “stable” is used here only to represent a middle point of manipulation described in the dataset [21]. The value refers to the moment the hand achieved the force defined for that run.

Figure 2 shows the sample time instances (red lines) where sensor data was collected to assemble the 3 sample points representation dataset. Here only three samples, during initial, stable and final manipulation trajectory stages were used. Using the 3-points sample approach, we tested representations containi only the 16 tactile sensors, only the 8 joint position values or a combination of both. This process yields 3 data tables with 2550 samples for training and testing machine learning models using the 3-points representation:

- **3 points Tact dataset (51 features):** Target, manipulation orientation (2), 3-points x 16 tactile sensor data.
- **3 points Pos dataset (27 features):** Target, manipulation orientation (2), 3-points x 8 joint position sensors.
- **3 points Tact+Pos dataset (75 features):** Target, manipulation orientation (2), 3-points x 16 tactile sensors, 3-points x 8 joint position.

For instance, figure 2 contains a sample of the “Cheez” object, weighing 100 grams and grasped from the right with 50mN of force where red lines show the 3 points sample used for this representation over tactile data.

2) Ten sample points (10-points) representation: We also tested a sampling strategy collecting 10 samples uniformly distributed between the initial and final positions defined by the dataset [21]. This process yields 3 data tables with 2550 samples for training and testing machine learning models using the 10-points representation:

- **10 Points Tact dataset (163 features):** Target, manipulation orientation (2), 10-points x 16 tactile sensor data.
- **10 Points Pos dataset(83 features):** Target, manipulation orientation (2), 10-points x 8 joint position data.
- **10 Points Tact+Pos dataset(243 features):** Target manipulation orientation (2), 10-points x 16 tactile sensor data, 10-points x 8 joint position data.

For instance, figure 3 shows 10 sample points of tactile data collected (red lines) for a run of object "Cheez" weighing 100 grams and force 50 mN grasped from the right.

3) Descriptive statistics (DS) representation: An approach using descriptive statistics calculated from the sensors to describe the object was also tested in this work. We used the following descriptive statistics: average, standard deviation, maximum value, and the third quartile. Other statistical metrics (e.g. minimum, median and first quartile) did not show relevant results in our preliminary tests and were not used. This approach yields six new data tables:

- **Average Tact dataset (19 features):** Target, manipulation orientation (2), average of each of 16 tactile sensors.
- **Average Pos dataset (11 features):** Target, manipulation orientation (2), average of each of 8 joint position data.
- **Average Tact+Pos dataset (27 features):** Target, manipulation orientation (2), average of 16 tacile sensors, average of 8 joint position data.
- **DS Tact dataset (67 features):** Target, manipulation orientation (2), 4 descriptive statistics from 16 tactile sensors.
- **DS Pos dataset (35 features):** Target, manipulation orientation (2), 4 descriptive statistics from 8 joint position sensors.
- **DS Tact+Pos dataset (99 features):** Target, manipulation orientation (2), 4 descriptive statistics from 16 tactile sensors, 4 statistics from 8 joint-position sensors.

Manipulating the data using the methods described in this section yields a total of 12 datasets with different representations. All representations described above were used to train machine learning methods and to evaluate their accuracy in the task of object recognition.

B. Machine learning models for object recognition

In the object recognition phase of Figure 1, all of our different data representations were used to train several machine learning methods, including: Extra-Trees classifier (ET), Random Forest (RF), Ridge classifier (Ridge), k-nearest neighbors (KNN), multi-layer perceptron (MLP) and Support Vector Machines (SVM). All methods used the default hyperparameter settings of the scikit-learn library [22]. The models were evaluated according to their accuracy using 5-fold cross validation.

IV. Results

We evaluate the data representation performances in two different ways. First, we assess the performances of a single classifier using the different representations and try to determine the one that improves the performances the most. Second, we evaluate if a combination of data representation and classifier achieves the best performance for the pick-and-place manipulation task. Since we have proposed 12 different data representations, we split the results into three tables based on sensors used: only finger joint position (Pos) on Table I, tactile sensor (Tact) on Table II, and finger joint position and tactile sensor sensors combined (Tact+Pos) on Table III. Finally, Table IV selects the best results from the previous three tables and discusses if there is a best combination of machine learning model and data representation.

Table I shows that, when only the tactile sensor is used in the object recognition problem, the descriptive statistics (DS) data representations obtain significant gains for most machine learning models, ranging from 25% (comparing RF with 3 Points Tact and DS Tact representations) to 65% (comparing KNN with 10 points Tact and DS Tact representations). The only exception is the Ridge classifier, where the 10 Points tact representation obtains the best accuracy value (46.31%). We used a Wilcoxon Signed-Rank Test to verify if such differences were indeed statistically significant (p-value $<$ 0.05). The results show that for most models using the DS Tact representation, these differences are indeed significant. However, the differences between the 10 Points Tact and DS Tact are not significant, and therefore, we believe that choosing the DS Tact using only the Tact sensor is the best choice. The best machine learning model using only a tactile sensor is the Extra Trees (ET), obtaining 88.43% of accuracy and representing a gain of 91% compared to the Ridge classifier's best performance (i.e., 10 Points Tact representation). A Wilcoxon Signed-Rank Test showed that such differences are also indeed significant.

Table II shows the data representation results using the joint position sensor only. The results show that for ET and RF, all data representations are competitive, and when comparing them, the differences are marginal (approximately 2% accuracy). The Wilcoxon Signed-Rank Test showed that for ET and RF there is no statistically significant difference between the 3 Points, 10 Points, and DS Pos. However, we can observe considerable improvements for SVM, KNN, and Ridge classifiers, with gains ranging from 25% (comparing KNN with DS Pos and 10 points Pos) to 43% (comparing SVM with DS Pos and Average Pos). In terms of the machine learning model, the differences between ET and RF using the multiple representations are also marginal, so choosing any of those using just joint position sensor data is a reasonable choice.

Table III shows the results for the data representations using the tactile and joint position sensors combined. Similarly to the results of the evaluation of the tactile sensor data (Table I), the descriptive statistics (DS) data representation obtains the best performance in terms of accuracy for most models with gains ranging from 8% (comparing RF using DS Tact+Pos and 3 Points Tact+Pos) to 71% (comparing SVM using DS

TABLE I
Summary of correct object identification using tactile sensors (Tact) for different data representations. Values marked in **bold** significantly outperformed other values in the same row, according to a Wilcoxon Signed-Rank Test (p-value < 0.05).

Models	3 Points Tact	10 Points Tact	Average Tact	DS Tact
ET	67.69 % ± 1.56 %	75.10 % ± 0.67 %	81.92 % ± 0.62 %	**88.43 % ± 0.64 %**
RF	67.80 % ± 1.16 %	75.41 % ± 1.37 %	78.43 % ± 0.95 %	**85.29 % ± 1.39 %**
MLP	51.57 % ± 3.01 %	53.88 % ± 1.59 %	67.96 % ± 0.95 %	**76.67 % ± 1.45 %**
SVM	43.45 % ± 1.98 %	50.43 % ± 3.02 %	44.67 % ± 0.94 %	**57.53 % ± 2.06 %**
KNN	40.63 % ± 2.89 %	39.33 % ± 2.69 %	57.69 % ± 1.96 %	**65.1 % ± 1.5 %**
Ridge	38.98 % ± 1.65 %	**46.31 % ± 1.91 %**	35.80 % ± 1.85 %	**45.41 % ± 1.68 %**

TABLE II
Summary of correct object identification using joint position (Pos) sensor for different data representations.

Models	3 Points Pos	10 Points Pos	Average Pos	DS Pos
ET	**66.43 % ± 1.66 %**	**66.86 % ± 1.84 %**	65.65 % ± 1.44 %	**66.71 % ± 1.80 %**
RF	**66.20 % ± 1.48 %**	**66.51 % ± 1.50 %**	65.80 % ± 1.55 %	**66.51 % ± 1.75 %**
MLP	63.76 % ± 1.31 %	63.88 % ± 1.70 %	64.51 % ± 1.07 %	**66.39 % ± 1.23 %**
SVM	59.57 % ± 2.30 %	60.75 % ± 1.53 %	44.59 % ± 1.57 %	**63.96 % ± 1.39 %**
KNN	53.73 % ± 3.03 %	49.88 % ± 2.25 %	**62.63 % ± 1.46 %**	**62.35 % ± 2.19 %**
Ridge	39.92 % ± 1.26 %	**49.41 % ± 1.41 %**	30.31 % ± 1.01 %	46.31 % ± 0.69 %

TABLE III
Summary of correct object identification using finger joint position and tactile sensor combined (Tact+pos) for the data representations.

Models	3 Points Tact+Pos	10 Points Tact+Pos	Average Tact+Pos	DS Tact+Pos
ET	85.65 % ± 1.30 %	88.04 % ± 1.02 %	90.12 % ± 1.45 %	**93.53 % ± 0.95 %**
RF	84.98 % ± 1.71 %	87.76 % ± 1.53 %	88.59 % ± 0.75 %	**91.92 % ± 1.11 %**
MLP	73.45 % ± 2.08 %	73.57 % ± 3.41 %	85.37 % ± 0.64 %	**88.86 % ± 0.95 %**
SVM	51.37 % ± 2.74 %	47.53 % ± 2.38 %	74.90 % ± 1.95 %	**81.73 % ± 0.67 %**
KNN	66.04 % ± 1.86 %	64.67 % ± 2.27 %	63.80 % ± 1.54 %	**78.47 % ± 0.83 %**
Ridge	51.22 % ± 2.09 %	**62.90 % ± 1.67 %**	43.37 % ± 2.36 %	**60.47 % ± 1.36 %**

TABLE IV
Accuracies for the best combinations of models and data representations.

Model	Data Representation	Accuracy
ET	**DS Tact+Pos**	**93.53 % ± 0.95 %**
RF	DS Tact+Pos	91.92 % ± 1.11 %
ET	DS Tact	88.43 % ± 0.64 %
RF	DS Tact	85.29 % ± 1.39 %
ET	10 Points Pos	66.86 % ± 1.84 %
RF	10 Points Pos	66.51 % ± 1.50 %

Tact+Pos with 10 Points Tact+Pos). Again, the Ridge model is the only one in which the DS Tact+Pos does not obtain the best performance, and instead, the best performance is achieved by the 10 Points Tact+Pos representation. The main difference between using only the tactile sensor and the tactile and joint position sensors combined is mainly regarding the maximum accuracy achieved. While we observe best accuracies of using only tactile sensor ranging from 46.31% (Ridge classifier with 10 Points Tact) to 88.43% (ET classifier with DS Tact), the combination of both sensors ranges from 62.90% (Ridge classifier with 10 Points Tact+Pos) to 93.53% (ET classifier with DS Tact), and they represent gains of 35.8% to 5.8%, respectively. A Wilcoxon Signed-Rank Test shows that such differences are significant and that DS Tact+Pos provides the best results in most cases.

Table IV gives us a better understanding of whether there is a better combination of data representation and machine learning model that solves the object recognition problem during a pick-and-place manipulation task. We assembled Table IV by selecting the two best combinations of machine learning models and data representations from Tables I, II and III. We observe from Table IV that ensemble (e.g., ET and

RF) techniques achieve the best accuracy performances. ET achieves the best performance using the DS data representation using the tactile (88.43%) and joint position (66.71%) sensors only and also achieves the best performance when both sensors are combined (93.53%). A Wilcoxon Signed-Rank Test shows that the 93.53% obtained the ET classifier with the DS Tact+Pos representation has significant statistical difference for the other representations in Table IV and the result indicates that those differences are indeed substantial to all others. This means that for the problem of object recognition using the during the pick-and-place manipulation tasks, choosing ET with the DS Tact+Pos representation is indeed the best choice.

V. Discussion

Our findings support that tactile data or a combination with joint-position information can be successfully used for object identification during manipulation tasks. The feature engineering approach with descriptive statistics (DS) to represent data used in this paper showed promising results regarding recognizing objects using a combination of tactile and joint-position information. Testing different data representations was crucial for improving the recognition of objects from 66.43% accuracy using 3 Points Pos to 93.53% of correct recognition using the same model (ET) but with DS Tact+Pos representation.

The original dataset used in this work also contains visual information, which could be valuable to build and test finger joint position, visual, and tactile data representations. We intend therefore to conduct experiments using feature extraction methods from the computer vision literature such as somatosensory maps [23] and visuo-haptic frameworks [24]. Model tuning may also be the focus of future works, because a comprehensive study of the models' hyper-parameters used to train the machine learning models could likely increase the accuracy.

Our approach does not distinguish between successful and failed manipulation attempts. Because the dataset does not provide that annotation, all manipulation runs are treated as regular attempts. Our results show that the data representation may be reliable under non-successful attempts with little data. However, this assumption was not evaluated in the present paper.

References

[1] A. Castro, F. Silva, and V. Santos, "Trends of human-robot collaboration in industry contexts: Handover, learning, and metrics," *Sensors*, vol. 21, no. 12, pp. 1–28, 2021.

[2] C. Eppner, S. Höfer, R. Jonschkowski, R. Martín-Martín, A. Sieverling, V. Wall, and O. Brock, "Lessons from the Amazon Picking Challenge: Four Aspects of Building Robotic Systems," in *Robotics: Science and Systems XII*. Robotics: Science and Systems Foundation, 2016. [Online]. Available: http://www.roboticsproceedings.org/rss12/p36.pdf

[3] J. R. Flanagan, M. C. Bowman, and R. S. Johansson, "Control strategies in object manipulation tasks," *Current Opinion in Neurobiology*, vol. 16, no. 6, pp. 650–659, dec 2006. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0959438806001450

[4] K. Shimoga, "Robot grasp synthesis algorithms: A survey," vol. 15, no. 3, 1996, pp. 230–266. [Online]. Available: https://doi.org/10.1177/027836499601500302

[5] C. Goldfeder and P. K. Allen, "Data-driven grasping," vol. 31, no. 1, 2011. [Online]. Available: https://doi.org/10.1007/s10514-011-9228-1

[6] A. Rodriguez, M. T. Mason, and S. Ferry, "From caging to grasping," vol. 31, no. 7, 2012, pp. 886–900. [Online]. Available: https://doi.org/10.1177/0278364912442972

[7] I. Kamon, T. Flash, and S. Edelman, "Learning to grasp using visual information," in *in Proceedings of the 1996 IEEE International Conference on Robotics and Automation*, 1994, pp. 2470–2476.

[8] I. Lenz, H. Lee, and A. Saxena, "Deep learning for detecting robotic grasps," 2014.

[9] D. Kappler, J. Bohg, and S. Schaal, "Leveraging big data for grasp planning," 2015, pp. 4304–4311.

[10] E. Johns, S. Leutenegger, and A. J. Davison, "Deep learning a grasp function for grasping under gripper pose uncertainty," 2016.

[11] J. Mahler, F. T. Pokorny, B. Hou, M. Roderick, M. Laskey, M. Aubry, K. Kohlhoff, T. Kröger, J. Kuffner, and K. Goldberg, "Dex-net 1.0: A cloud-based network of 3d objects for robust grasp planning using a multi-armed bandit model with correlated rewards," in *2016 IEEE International Conference on Robotics and Automation (ICRA)*, 2016, pp. 1957–1964.

[12] L. Pinto and A. Gupta, "Supersizing self-supervision: Learning to grasp from 50k tries and 700 robot hours," in *2016 IEEE International Conference on Robotics and Automation (ICRA)*, 2016, pp. 3406–3413.

[13] S. Levine, P. Pastor, A. Krizhevsky, and D. Quillen, "Learning hand-eye coordination for robotic grasping with deep learning and large-scale data collection," 2016.

[14] R. Calandra, A. Owens, D. Jayaraman, J. Lin, W. Yuan, J. Malik, E. H. Adelson, and S. Levine, "More than a feeling: Learning to grasp and regrasp using vision and touch," vol. abs/1805.11085, 2018. [Online]. Available: http://arxiv.org/abs/1805.11085

[15] S. Varkey and C. Achy, "Learning robotic grasp using visual-tactile model," 12 2018, pp. 1–3.

[16] 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," vol. 9, no. 2, 2016, pp. 207–220.

[17] A. Molchanov, O. Kroemer, Z. Su, and G. S. Sukhatme, "Contact localization on grasped objects using tactile sensing," *IEEE International Conference on Intelligent Robots and Systems*, vol. 2016-Novem, pp. 216–222, 2016.

[18] R. Paolini, A. Rodriguez, S. S. Srinivasa, and M. T. Mason, "A data-driven statistical framework for post-grasp manipulation," *International Journal of Robotics Research*, vol. 33, no. 4, pp. 600–615, 2014. [Online]. Available: http://ijr.sagepub.com/cgi/doi/10.1177/0278364913507756

[19] 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 *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, vol. 2015-Decem. IEEE, sep 2015, pp. 5073–5080. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=7354091

[20] T. E. A. de Oliveira, A.-M. Cretu, V. P. da Fonseca, and E. M. Petriu, "Touch sensing for humanoid robots," *IEEE Instrumentation & Measurement Magazine*, vol. 18, no. 5, pp. 13–19, oct 2015. [Online]. Available: http://ieeexplore.ieee.org/document/7271221/

[21] T. Wang, C. Yang, F. Kirchner, P. Du, F. Sun, and B. Fang, "Multimodal grasp data set: A novel visual–tactile data set for robotic manipulation," *International Journal of Advanced Robotic Systems*, vol. 16, no. 1, p. 172988141882157, jan 2019. [Online]. Available: http://journals.sagepub.com/doi/10.1177/1729881418821571

[22] 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.

[23] S. Denei, F. Mastrogiovanni, and G. Cannata, "Towards the creation of tactile maps for robots and their use in robot contact motion control," *Robotics and Autonomous Systems*, vol. 63, no. P3, pp. 293–308, 2015. [Online]. Available: http://dx.doi.org/10.1016/j.robot.2014.09.011

[24] G. Rouhafzay and A.-m. Cretu, "A Visuo-Haptic Framework for Object Recognition Inspired from Human Tactile Perception †," pp. 1–7, 2018.