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

CITR: A Coordinate-Invariant Task Representation for Robotic Manipulation

Peter So[1,†], Rafael I. Cabral Muchacho[2,†], Robin Jeanne Kirschner[1], Abdalla Swikir[1,5], Luis Figueredo[3,1], Fares J. Abu-Dakka[4] and Sami Haddadin[1]

***Abstract*— The basis for robotics skill learning is an adequate representation of manipulation tasks based on their physical properties. As manipulation tasks are inherently invariant to the choice of reference frame, an ideal task representation would also exhibit this property. Nevertheless, most robotic learning approaches use unprocessed, coordinate-dependent robot state data for learning new skills, thus inducing challenges regarding the interpretability and transferability of the learned models.**

In this paper, we propose a transformation from spatial measurements to a coordinate-invariant feature space, based on the pairwise inner product of the input measurements. We describe and mathematically deduce the concept, establish the *task fingerprints* as an intuitive image-based representation, experimentally collect task fingerprints, and demonstrate the usage of the representation for task classification. This representation motivates further research on data-efficient and transferable learning methods for online manipulation task classification and task-level perception.

I. Introduction

Robots, like any real-world entity, interact with the world through contact. Understanding such physical interactions provides valuable information about tasks involving contact, which is essential for task execution [1], [2]. Take, for instance, the common task of key insertion [3], where the friction force during insertion is followed by a torque around the insertion axis to rotate the key. The torque is followed by an angular velocity of the key around the same axis. This manipulation skill and understanding of a motion constraint hold true, whether we are opening a door or a chest on the floor. This intuition comes from recognizing patterns across a series of experiences. Given the flourishing field of learning robots, studying such manipulation skills is of great interest to the robotics community [3]–[7]. Understanding the fundamental constraints for each manipulation task can unlock the ability of robots to generalize their knowledge from known tasks to those they have not yet encountered, which is one of the main goals for intelligent robots [8].

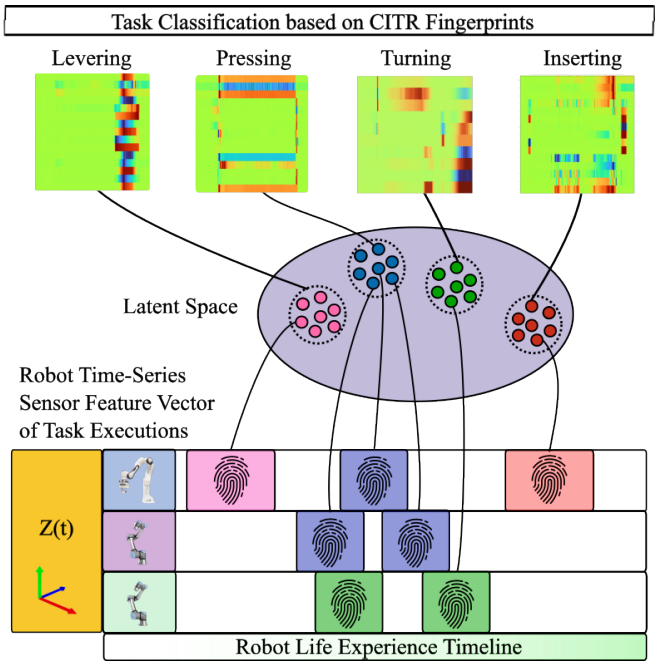

Fig. 1. The Coordinate-Invariant Task Representation (CITR) provides an abstraction of frame-based robot state variables $\mathbf{Z}(t)$ during manipulation skills for improved task recognition and classification which is independent of the task's reference frame. Heatmaps of the CITR, referred to as Task Fingerprints, provide a visual representation of the task dynamics which can be clustered together using image similarity metrics for tactile skill taxonomies.

Industrially relevant tasks (e.g., pressing, levering, or turning screws) are generally complex, dynamic, and multimodal. These characteristics pose challenges in task representation, especially when dealing with time-series [9]. While the Institute of Electrical and Electronics Engineers (IEEE) provides a naming standard for task descriptions [10], several works either define their own using object's semantic, visual or geometric affordances [11]–[13], wrench-based descriptors [14], [15], sequential forces [16] or actions, mostly described through Markov decision processes, [17]. Alternatively, existing literature also focused on learning such task descriptors and dynamics by exploring the full range of proprioceptive and exteroceptive data available [18]–[22]. While the former approach is limited to prior well-known

[†]These authors contributed equally. [1]Munich Institute of Robotics and Machine Intelligence, TUM, Germany. [2]Division of Robotics, Perception and Learning, KTH Royal Institute of Technology, Sweden. [3]School of Computer Science, University of Nottingham, UK. Luis Figueredo is also an Associated Fellow at the MIRMI - TUM. [4]Electronic and Informatics Department, Faculty of Engineering, Mondragon Unibertsitatea, Spain. [5]Department of Electrical and Electronic Engineering, Omar Al-Mukhtar University (OMU), Libya.

We gratefully acknowledge the funding of European Union's Horizon HORIZON Research and Innovation Actions via the project euROBIN (grant No. 101070596) and ReconCycle (grant no. 871352). The work is partially funded by the Lighthouse Initiative Geriatronics Project X by StMWi Bayern (grant no. 5140951) as well as the SafeRoBAY project (grant No: DIK0203/01) supported by the Bavarian State Ministry for Economic Affairs, Regional Development and Energy (StMWi). This work was partially supported by the Wallenberg AI, Autonomous Systems and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation. Corresponding author: peter.so@tum.de

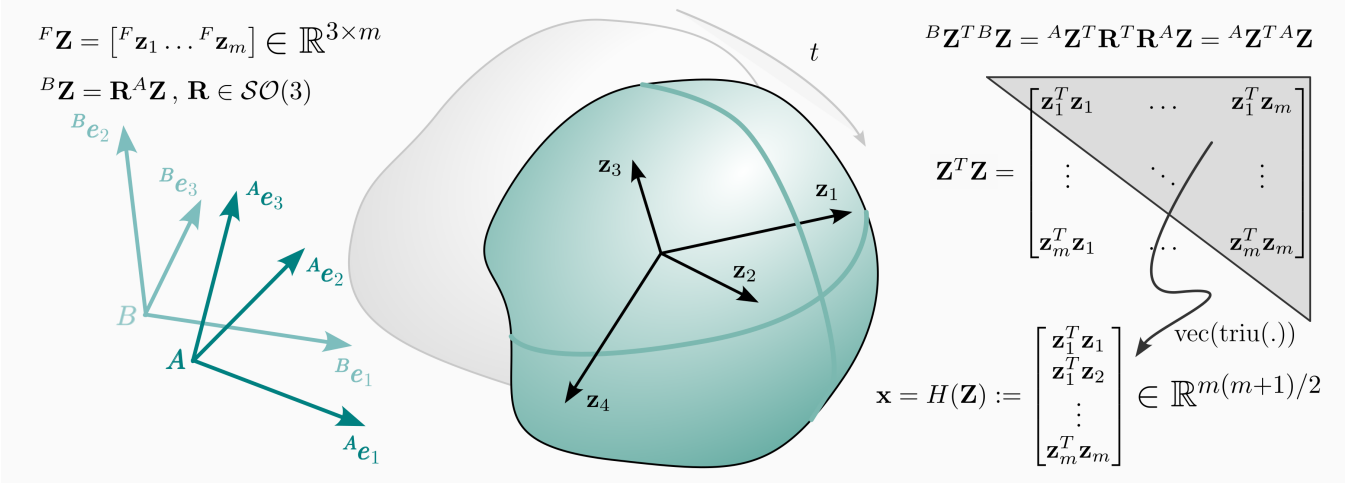

Fig. 2. Measurements of vector quantities $\mathbf{z}_i \in \mathbb{R}^3$ can be encoded in a rotation invariant feature vector aiming to represent the relationship between measurements. Inertial reference frames A and B are shown in the bottom left region of the figure. By using the rotation invariance property of the inner product, and the symmetry of the matrix $\mathbf{Z}^T\mathbf{Z}$, we define the coordinate-invariant feature vector $\mathbf{x}$ as the output of the transformation H – the vectorized upper-diagonal elements of the pairwise inner products of the measurements.

tasks, the latter mostly focuses on learning algorithms without reasoning on the data analysis itself. This contrasts recent findings highlighting the significance of pre-processing for learning efficiency and explainability [23].

In this context, in this study, we revisit the core elements that characterize a task. We propose a novel approach to reshape the feature space through a transformation designed to capture the task dynamics across various instances of its execution, as illustrated in Fig. 1.

The main contribution of this paper is a physically-motivated, coordinate-invariant representation for robotic manipulation tasks. We prove the invariance of the proposed representation and show validating experimental results. We provide a compact, 2D-heatmap visualization of trajectories in the invariant representation, *task fingerprint*, to compare the dynamics of manipulation tasks. As initial experiments, we demonstrate the usage of the proposed representation for task classification based on real-world experimental setups and industrially-motivated manipulation tasks.

The remainder of this paper is structured as follows. Sec. II describes the proposed coordinate-invariant task representation. The experimental setups for data collection, the collected data sets, and the design of evaluation experiments are presented in Sec. III. The results of the evaluation experiments are shown and analyzed in Sec. IV. After an overall discussion in Sec. V, the conclusion in Sec. VI summarizes the paper and sheds light on future work.

II. Coordinate-Invariant Task Representation

In this section, we introduce the feature space transformation $H : \mathbb{R}^{3\times m} \to \mathbb{R}^{m(m+1)/2}$, and make use of it to encode and visualize coordinate-invariant task descriptors (see Fig. 2). By coordinate-invariant, we refer to the invariance property of the feature space w.r.t. the choice of the static reference frame in which the spatial measurements are obtained. We propose to leverage nonlinear-basis functions to map spatial measurements in a common reference frame in 3D-space to a coordinate-invariant feature space. More specifically, we use bi-linear basis functions defined as the pairwise inner product of all spatial measurements, making use of the rotation-invariance property of inner products, and leading to a trivially differentiable transformation. The translation invariance of the transformation is ensured by measuring positions as relative translations between task entities. Given that manipulation tasks can be seen as a sequence of constrained dynamics in task space, and that the invariant feature space contains the relationship between spatial variables, we define task fingerprints as trajectories in the proposed invariant feature space.

A. Coordinate-Invariant Feature Space

We represent the m spatial measurements in a common reference frame as 3-dimensional column vectors $\mathbf{z}_i$, which are horizontally stacked and represented in the measurement matrix $\mathbf{Z}$

$$\mathbf{Z} = \begin{bmatrix}\mathbf{z}_1 & \mathbf{z}_2 & \dots & \mathbf{z}_m\end{bmatrix} \in \mathbb{R}^{3\times m}. \tag{1}$$

Discrete trajectories of length n are represented in the coordinate-invariant feature space as a trajectory matrix $\mathbf{X}$ with n horizontally stacked column vectors $\mathbf{x}_j$, where each column is obtained from the measurement matrix $\mathbf{Z}_j$ at the corresponding time. Concretely,

$$\mathbf{x}_j = H(\mathbf{Z}_j) := \mathrm{vec}\left(\mathrm{triu}(\mathbf{Z}_j^T\mathbf{Z}_j)\right), \tag{2}$$

$$\mathbf{X} = \begin{bmatrix}\mathbf{x}_1 & \mathbf{x}_2 & \dots & \mathbf{x}_n\end{bmatrix} \in \mathbb{R}^{\frac{1}{2}m(m+1)\times n}, \tag{3}$$

defined with the operators vec for vectorization (matrix to column vector) and triu (upper triangular matrix) to obtain the elements on and above the diagonal.

The relationship between the proposed feature space and a rotation between static reference frames employed for measurements is shown by

$$\mathbf{Z}' = \mathbf{R}\mathbf{Z}, \quad \mathbf{R} \in \mathcal{SO}(3)\,, \tag{4}$$

$$\mathbf{Z}'^T\mathbf{Z}' = \mathbf{Z}^T\mathbf{R}^T\mathbf{R}\mathbf{Z} = \mathbf{Z}^T\mathbf{Z}\,, \tag{5}$$

and clearly implies the rotation invariance property of the features. Further, through the eigenvalue decomposition of the inner product matrix, it is straightforward to see that

$$\mathbf{Z}^T\mathbf{R}^T\mathbf{R}\mathbf{Z} = \mathbf{Z}^T\mathbf{Z} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T\,, \tag{6}$$

$$\mathbf{Z}_{\text{rec}} = \mathbf{\Lambda}^{1/2}\mathbf{Q}^T, \tag{7}$$

where $\mathbf{\Lambda}$ and $\mathbf{Q}$ build the eigendecomposition of $\mathbf{Z}^T\mathbf{Z}$. This means that a reconstructed measurement matrix $\mathbf{Z}_{\text{rec}}$ recovers the original measurement structure in a coordinate frame spanned by the principal components of the measurement matrix $\mathbf{Z}$.

Note that additional scalar measurements are inherently invariant to rotations and can be appended to the column vectors in feature space to consider multimodal information (temperature, volume, etc.) beyond spatial vector quantities.

B. Task Fingerprints

We make use of a task descriptor that is constructed by a row-wise unbiased scaling to the range [-1, 1],

$$\mathbf{X} = (x_{kj})_{1\le k\le \frac{1}{2}m(m+1), 1\le j\le n} \tag{8}$$

$$d_k = \max_{1\le j\le n} \text{abs}(x_{kj}), \tag{9}$$

$$\mathbf{d} = [d_1, \cdots, d_k, \cdots, d_{\frac{1}{2}m(m+1)}]^T, \tag{10}$$

$$\mathbf{X}_{\text{id}} = \text{diag}(\mathbf{d})^{-1}\mathbf{X}\,, \tag{11}$$

that can be visualized as a heatmap (see Fig. 3). Task Fingerprints are explored in the following section to visualize task-related dynamics and are used as the basis of a trajectory classification model.

III. Experimental Design and Procedure

Here, we explain the motivation of the selected manipulation tasks and their relevance to industrial use cases. We present the experimental setups and procedures we applied to collect data discussed in Section IV. Lastly, we describe the validation study for using the representations in task classification.

A. Task Selection

We conducted a structured state-of-the-art analysis within the domains of research projects and papers [24]–[27], industrial applications and trade fair demonstrations [28]–[33] to find the most relevant robotic manipulation skills to examine with the proposed representation. In this paper, we focus on applications requiring the robot capability to feel, understand, and shape contacts effectively, which are not necessarily fully automated yet, but, e.g., as demonstrated in the videos as necessary steps for future automation. Consequently, we found the electronic assembly and disassembly [25], [34], and flexible manufacturing in the mechanics' sector [24], [29],

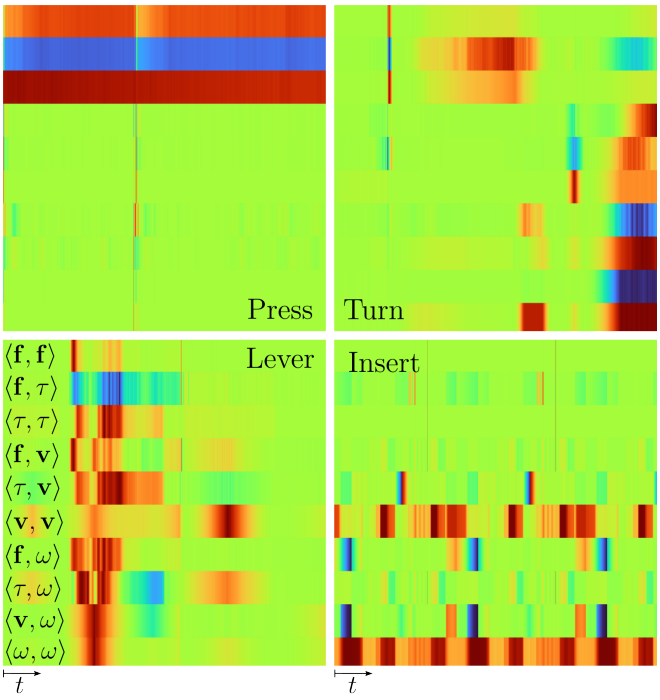

Fig. 3. Forces, velocities, and torques from robot experiments in the coordinate-invariant representation can be visualized as heatmaps with normalized row-wise features, as computed in (11). In this grid, we visualize four dissimilar manipulation tasks: pressing, turning, levering, and inserting. The row-wise feature labels on the bottom left heatmap also apply to the other three. The variables describe the velocity $\mathbf{v}$, angular velocity $\boldsymbol{\omega}$, force $\mathbf{f}$, and torque $\boldsymbol{\tau}$ measured at the end-effector.

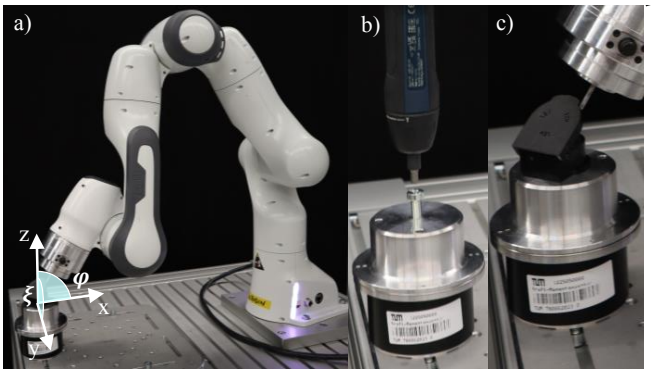

Fig. 4. Experimental setup using the FE robot to conduct levering and screw driver insertion tasks.

as the most relevant application domains. The relevant skills acquired from this literature analysis are:

- *Insertion* for electronics testing and assembly [28], [35],
- *(Un-)screw driving* in different industrial domains [25], [27], [30], [31],
- *Levering* for disassembly [25], and
- *Force application* for successful assembly and testing [29], [32], [33].

CITR can render force and trajectory recordings from different robot manipulators into a task fingerprint for cross comparison and clustering regardless of the orientation in

TABLE I

LIST OF EXPERIMENTAL MANIPULATION TASKS AND OBJECTS USED IN THE TASK REPRESENTATION DATASET.

Task Class	Moving Object	Fixed Object	Robot	Specialized Tooling
Inserting	Key1	Key Hole1	UR10e	Task Board [36]
	Key2	Key Hole2	UR10e	Control Box
	Key3	Key Hole3	UR10e	Table Vise
	USB Plug	USB Port	UR10e	USB Pedestal
	RJ45 Plug	RJ45 Port	UR10e	RJ45 Pedestal
	Round Peg	6mm Hole	FE	Hole Pedestal
Pushing	Spherical Tool	FT Sensor	UR5e	HDPE Cap
	Battery Box Lid	Battery Box	UR10e	Task Board
Levering	Hinged Lid	Charge Port	UR5e	Charger Pedestal
	Hinged Lid	Charge Port	FE	Charger Pedestal
Screw driving	Bit & M6x1.0 Screw	Tapped Hole	FE	Tapped Hole Pedestal

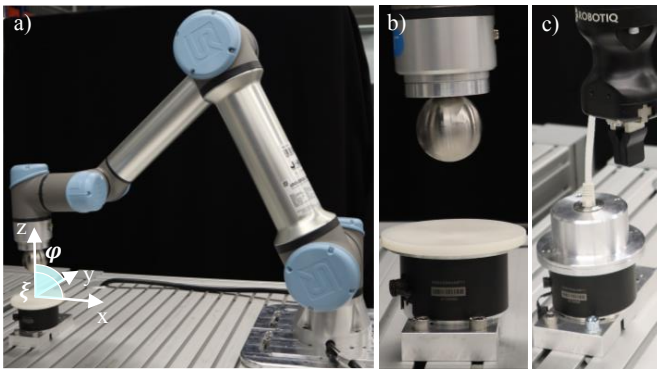

Fig. 5. Experimental setup using the UR5e robot (a) to conduct pushing task with a spherical tool onto a force-torque sensor with a HDPE cap (b) and an insertion task with a RJ45 plug (c).

which the task was executed. To demonstrate this property, we present the experimental results of three robot platforms using only the standard programming interfaces offered by the manufacturers to show the versatility of our approach. Each robot is equipped with a task-specific tool for four classes of manipulation tasks. Each test configuration is listed in Table I and their individual details are given in the following sections.

Data sets from experiments are labeled with the corresponding manipulation task for each robot setup containing the following 3-dimensional vector measurements at the end-effector with respect to the robot's base frame as the reference coordinate system: velocity, angular velocity, force, and torque (moment). Table I provides an overview of all conducted experiments.

B. Test Setup 1: Franka Emika

In our first experimental setup, we used a tabletop-mounted FE robot manipulator that was equipped with a peg and a screwdriver as depicted in Fig. 4 a). The proprietary libfranka[1] force controller was used to program the manipulation experiments and generate a robot log file including joint angles, velocities, and end-effector force/torque (F/T) estimations using the *RobotState* variable. Using the FE, the motion and force features of the following tasks are recorded:

- *Levering* the lid of a car plug (c.f. Fig. 4 a) and c)) and
- *Screw driving* using a bit-hex for a M6x1.0 screw (c.f. Fig. 4 b)),

The FE dataset contains the manipulation tasks: 12 peg insertions, 2 Ethernet insertions, 2 lid leverings, and 3 bolt unscrewing tasks. The data was collected at $1\,\mathrm{kHz}$ and the dataset contains $414,541$ time-series samples across the task set.

C. Test Setup 2: UR5e

In the second experimental setup, we used a Universal Robots UR5e equipped with a HAND-E made by the company Robotiq and a stainless steel sphere of $50\,\mathrm{mm}$ diameter as depicted in Fig. 5. The robot is additionally equipped with a *Force CoPilot* license unlocking the robot's integrated force sensor located at the robot flange. For each manipulation experiment, a program was written on the UR teach pendent and executed manually. To aid in post-processing the robot log data, a digital output signal was turned on during the program segment containing the manipulation motions. This signal flag was then used later to filter the robot log for only the object manipulation segments. The robot log was captured using a Python script provided by Universal Robots which captures all available robot state data over the real-time data exchange interface (RTDE[2]).. We selected the IO state, end-effector flange F/T information, and the robot end-effector Cartesian and angular velocities in base frame coordinates for our analysis. The corresponding manipulation objects were mounted to the work surface on interchangeable aluminum plates. The UR5e was programmed to conduct the following tasks:

- *Applying force* onto a hard polyethylene (PE) cover (c.f. in Fig. 5 a) and b)) and
- *Inserting* a RJ45-plug into an Ethernet port (c.f. Fig. 5 c)).

The UR5e dataset contains the manipulation tasks: 4 Ethernet insertions, 6 pressing tasks, and 15 levering tasks of a hinged lid. The data was collected using the RTDE at $100\,\mathrm{Hz}$ and the dataset contains $21,911$ time-series samples across the task set.

D. Test Setup 3: UR10e

The UR10e setup is designed similar to the previously explained setup of the UR5e using the same gripper. While data recording and control are embedded in the same way, it needs to be mentioned that the UR10e does not serve the *Force CoPilot* license and, thus, can not use the embedded F/T sensor in the flange for end-effector F/T estimation. Instead, it uses the current-based force estimation using the robot joints and dynamics which we expect to be less accurate than

[1]libfranka https://frankaemika.github.io/docs/libfranka.html

[2]RTDE Guide: https://www.universal-robots.com/articles/ur/interface-communication/real-time-data-exchange-rtde-guide/

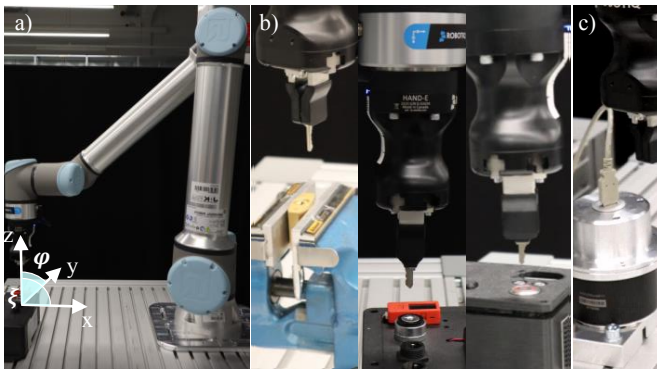

Fig. 6. Experimental setup of the UR10e robot (a) for insertion tasks of three different key and key hole pairs (b) and an USB-plug (c).

measurements made with the F/T sensor. Using the UR10e robot we performed the following tasks:

- *Inserting* of three different keys in their respective keyholes (c.f. in Fig. 5 a) and b)) and
- *Inserting* of a USB-plug (c.f. Fig. 5 c)).

The robot state was recorded during 38 insertions and 20 pushing task repetitions over the RTDE at 100 Hz and the dataset contains 28,022 time-series samples in the task set.

E. Validation Study

In this study, our objective is twofold. First, we aim to assess the quantity of information contained within the proposed rotation-invariant features. This information refers to the details and patterns these features capture from the data, that allow to relate a sample to its corresponding manipulation task. Second, we seek to evaluate how rotations of the input data's reference frame affect the performance of the fitted models. We employ a linear classifier using logistic regression. We apply this classifier to both the original vector data and to their rotation-invariant representation, treating different manipulation tasks as distinct classes in a multi-class classification problem.

We choose the accuracy a of the linear fit as a proxy for the amount of locally present task-relevant information, treating measurements as time-independent samples. In doing so we ignore long-term dependencies of the task dynamics and only look at the information for each individual time step. The simple logistic regression is advantageous given its limited capacity to overfit, and the relatively small size of recorded datasets. For each dataset (FE, UR5e) we fit two models until convergence: (1) from measurements in vector form ($\mathbf{Z}$) to task classes, and (2) from rotation invariant features ($\mathbf{x}$) to task classes. After fitting the models, we estimate the distributions of per class accuracy results for each model, induced by applying random rotations to the vector quantities in the dataset. The distributions are sampled from 1 000 random-uniform rotations that represent transformations of the experiment data to simulate carrying out the task in different orientations, for instance, key-inserting on wall or table-mounted lockers, data acquired from different perspectives such as world-frame, camera, base or end-effector. It also considers unconventional robot mountings beyond the standard table-mounted setup.

TABLE II
"ACROSS TASK CLASSES" SSIM SIMILARITY VALUES FOR DISSIMILAR MANIPULATION TASKS

Task Class	press	turn	lever	insert
press	1	0.61	0.64	0.54
turn	-	1	0.71	0.73
lever	-	-	1	0.76
insert	-	-	-	1

TABLE III
"WITHIN TASK CLASS" SSIM SIMILARITY VALUES FOR PUSH TASK PUSH DOWN WITH 20N FOR 3 SECONDS VARYING CONTACT ANGLE

Contact Angle	$\xi=0$ $\phi=0$	$\xi=45$ $\phi=0$	$\xi=-45$ $\phi=0$	$\xi=0$ $\phi=45$	$\xi=0$ $\phi=-45$
$\xi=0$ $\phi=0$	1	0.83	0.84	0.96	0.84
$\xi=45$ $\phi=0$	-	1	0.89	0.83	0.94
$\xi=-45$ $\phi=0$	-	-	1	0.85	0.89
$\xi=0$ $\phi=45$	-	-	-	1	0.84
$\xi=0$ $\phi=-45$	-	-	-	-	1

IV. EXPERIMENTAL RESULTS

In this section, we show the generated task fingerprints and show the results for the application of these representations for a task recognition classifier.

A. CITR Task Fingerprints

With the generated task fingerprints we can construct an initial labeled image set for a range of manipulation skills. From this image set, we can calculate a task similarity matrix to compare the pixel-to-pixel SSIM similarity [37] of the task fingerprints and their respective task dynamics. Comparing task fingerprints pair-wise with one another returns a vector of similarity percentages which are arranged into the upper right triangle in a comparison matrix, see Table II. For a new recorded manipulation experiment, we can iteratively compare the new task's fingerprint with each image in the image set to find its closest match to return the predicted task label. The predicted task class of the new fingerprint is the class of the closest match.

We found our method also groups similar executions of the same tasks, such as pressing down on a surface with different robot configurations, see Table III. Furthermore, we found the task fingerprints of different task classes have a relatively low task SSIM similarity score. Using this information, we can organize our collection of robot experience data into task classes based on experimentally found similarity thresholds.

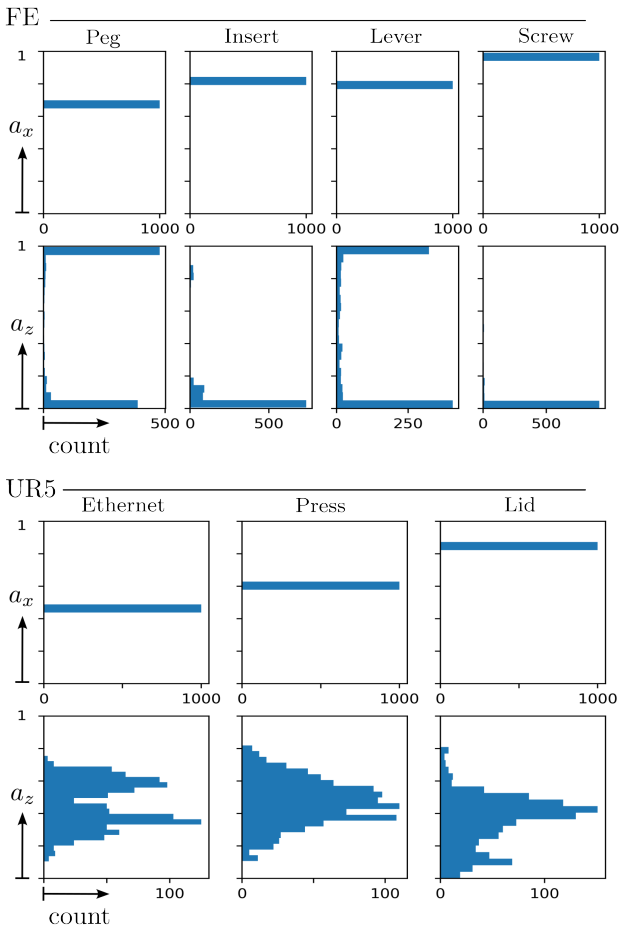

Fig. 7. The histograms (50 bins) approximate the sampled distributions of per class accuracy a on the corresponding data sets wrt. random uniform rotations of the input data. The histograms related to the FE dataset are shown in the top half of the figure; the UR5 in the bottom half. In each case, the top row shows the results of the model trained on the rotation-invariant features $\mathbf{x}$; the bottom row shows the ones of the raw measurements $\mathbf{Z}$, as depicted in the y-axis labels. The wide distribution of accuracies in the bottom row indicate the sensibility of the raw model to reference frame choices.

B. Validation Study for Coordinate-Invariance Property

To validate the repreentation's coordinate-invariant property, we performed an analysis of the performance of a task classifier trained using feature vectors from original robot state data and the coordinate-invariant features.

The classification results are visualized as histograms depicted in Fig. 7. A narrow band of accuracy indicates low sensibility wrt. rotations of the reference frame. The top row shows that the accuracy of the feature space model does not vary wrt. rotations of the input data. The bottom row shows the wide range of accuracy results arising from rotations of the data. This prediction accuracy variability indicates the model's high dependency on the measurement's frame of reference.

The per class accuracy of the model fitted from the rotation-invariant features are $\left[0.65,\, 0.79,\, 0.77,\, 0.94\right]$ for the FE robot experiment and $\left[0.44,\, 0.58,\, 0.82\right]$ for the UR robot. It is important to notice these values are significantly greater than the random guess, $25\,\%$ for the FE dataset and $33\,\%$ for the UR dataset. The higher per class accuracy shows that a relatively large amount of task-relevant information is present in the data when represented in the proposed feature space, even when ignoring the temporal relationships of the samples.

V. Discussion

Mapping historical robot experiments into CITR images, we can use image-based queries to compare and automatically label new robot task executions irrespective of the task coordinate frame. We show our CITR task fingerprints can cluster executions of the same task in multiple robot configurations based on the task's dynamics using the SSIM image comparison algorithm in MATLAB.

CITR fingerprints can organize robot log data with visual features. This has the benefit of navigating a collection of robot logs with an image-based search of compact representations independent of the task coordinate frame. However, CITR task fingerprints have some notable limitations. The horizontal ordering of features in the heatmap must be kept consistent in the database, otherwise the visual similarity metric will not hold. Our initial results from our experiment showed a repeatable pattern in the task fingerprints for similar tasks, however, further investigation is needed to show the extent of how this method could be used as a general task classifier. Additionally, the colors of features in the heatmaps rely on normalized values across the set of experiments and adding new experiences to the database requires a recalculation of the entire image set. While normalizing ranges for the make and model of a single robot is possible, comparisons to other robots would not make sense without a normalization across the entire comparison dataset.

Note that most of these limitations can be overcome by leveraging modern machine learning models, and informed or outlier-robust normalization, while using the proposed representations as the input space.

VI. Conclusion

We propose a new coordinate-invariant task representation (CITR) that encodes the dynamics of contact-rich manipulation tasks in a feature space invariant to reference frame rotations. Comparing task episodes in this latent space projection enables the user to classify tasks without knowing the relationship between the task and the robot base frame. We showed the coordinate-invariant property of the representation, and experimentally validated our claim with real robot data. Our proposed representation promises efficient analysis and learning of manipulation tasks, online task classification for shared autonomy during tele-operation, and improved task-level perception for enhanced manipulation autonomy, among others. Therefore, future work will expand the library of task fingerprints towards an entire manipulation skill taxonomy.

Acknowledgments

We would like to thank Kübra Karacan for her support with the experimental setup.

REFERENCES

[1] Aude Billard and Danica Kragic. Trends and challenges in robot manipulation. *Science*, 364(6446):eaat8414, 2019.

[2] N. Hogan. Impedance control: An approach to manipulation. In *1984 American Control Conference*, pages 304–313, 1984.

[3] Lars Johannsmeier, Malkin Gerchow, and Sami Haddadin. A framework for robot manipulation: Skill formalism, meta learning and adaptive control. In *2019 International Conference on Robotics and Automation (ICRA)*, pages 5844–5850, 2019.

[4] Masashi Okada, Mayumi Komatsu, Ryo Okumura, and Tadahiro Taniguchi. Learning compliant stiffness by impedance control-aware task segmentation and multi-objective bayesian optimization with priors. *arXiv preprint arXiv:2307.15345*, 2023.

[5] Aljaž Kramberger, Andrej Gams, Bojan Nemec, Casper Schou, Dimitrios Chrysostomou, Ole Madsen, and Aleš Ude. Transfer of contact skills to new environmental conditions. In *2016 IEEE-RAS 16th International Conference on Humanoid Robots (Humanoids)*, pages 668–675, 2016.

[6] Peter Pastor, Mrinal Kalakrishnan, Sachin Chitta, Evangelos Theodorou, and Stefan Schaal. Skill learning and task outcome prediction for manipulation. In *2011 IEEE International Conference on Robotics and Automation*, pages 3828–3834, 2011.

[7] Kuebra Karacan, Hamid Sadeghian, Robin Kirschner, and Sami Haddadin. Passivity-based skill motion learning in stiffness-adaptive unified force-impedance control. In *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 9604–9611, 2022.

[8] G. Saridis. Intelligent robotic control. *IEEE Transactions on Automatic Control*, 28(5):547–557, 1983.

[9] A. Bagnall, J. Lines, A. Bostrom, and et al. Time-series classification with cote: the collective of transformation-based ensembles. *Data Mining and Knowledge Discovery*, 31, 2017.

[10] Paulo Goncalves. Ieee standard ontologies for robotics and automation. *IEEE Std 1872-2015*, pages 1–60, 2015.

[11] Moritz Tenorth and Michael Beetz. Knowrob: A knowledge processing infrastructure for cognition-enabled robots. *The International Journal of Robotics Research*, 32(5):566–590, 2013.

[12] Pietro Vitiello, Kamil Dreczkowski, and Edward Johns. One-shot imitation learning: A pose estimation perspective. In *7th Annual Conference on Robot Learning*, 2023.

[13] Riddhiman Laha, Anjali Rao, Luis FC Figueredo, Qing Chang, Sami Haddadin, and Nilanjan Chakraborty. Point-to-point path planning based on user guidance and screw linear interpolation. In *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, volume 85451, page V08BT08A010. American Society of Mechanical Engineers, 2021.

[14] Luis FC Figueredo, Rafael De Castro Aguiar, Lipeng Chen, Thomas C Richards, Samit Chakrabarty, and Mehmet Dogar. Planning to minimize the human muscular effort during forceful human-robot collaboration. *ACM Transactions on Human-Robot Interaction (THRI)*, 11(1):1–27, 2021.

[15] Rachel Holladay, Tomás Lozano-Pérez, and Alberto Rodriguez. Force-and-motion constrained planning for tool use. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 7409–7416. IEEE, 2019.

[16] Lipeng Chen, Luis FC Figueredo, and Mehmet R Dogar. Manipulation planning under changing external forces. *Autonomous Robots*, pages 1–21, 2020.

[17] Oliver Kroemer, Scott Niekum, and George Konidaris. A review of robot learning for manipulation: Challenges, representations, and algorithms. *J. Mach. Learn. Res.*, 22(1), jan 2021.

[18] Jivko Sinapov, Priyanka Khante, Maxwell Svetlik, and Peter Stone. Learning to order objects using haptic and proprioceptive exploratory behaviors. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, 2016.

[19] T. Bergquist, C. Schenck, U. Ohiri, J. Sinapov, S. Griffith, and A. Stoytchev. Interactive object recognition using proprioceptive feedback. In *Proceedings of the 2009 IROS Workshop: Semantic Perception for Robot Manipulation*, page Page Numbers, 2009.

[20] Gyan Tatiya, Ramtin Hosseini, Michael C. Hughes, and Jivko Sinapov. Sensorimotor cross-behavior knowledge transfer for grounded category recognition. In *2019 Joint IEEE 9th International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob)*, 2019.

[21] Alex Vasquez, Zhanat Kappassov, and Veronique Perdereau. In-hand object shape identification using invariant proprioceptive signatures. In *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 965–970, 2016.

[22] Jivko Sinapov, Taylor Bergquist, Connor Schenck, Ugonna Ohiri, Shane Griffith, and Alexander Stoytchev. Interactive object recognition using proprioceptive and auditory feedback. *The International Journal of Robotics Research*, 30(10):1250–1262, 2011.

[23] H. I. Fawaz, G. Forestier, J. Weber, L. Idoumghar, and P. A. Muller. Deep learning for time series classification: A review. *Data Mining and Knowledge Discovery*, 33(4):917–963, 2019.

[24] European Union, 1994-2023. Cognitively enhanced robot for flexible manufacturing of metal and composite parts, 2022. `https://cordis.europa.eu/project/id/723853/reporting`, Last accessed on 2023-06-15.

[25] European Union, 1994-2023. Self-reconfiguration of a robotic workcell for the recycling of electronic waste. `https://cordis.europa.eu/project/id/871352/reporting/de`, Last accessed on 2023-06-15.

[26] Andrea Casalino, Filippo Cividini, Andrea Maria Zanchettin, Luigi Piroddi, and Paolo Rocco. Human-robot collaborative assembly: a use-case application. *IFAC-PapersOnLine*, 51(11):194–199, 2018. 16th IFAC Symposium on Information Control Problems in Manufacturing INCOM 2018.

[27] Sebastian Hjorth, Edoardo Lamon, Dimitrios Chrysostomou, and Arash Ajoudani. Design of an energy-aware cartesian impedance controller for collaborative disassembly. *arXiv preprint arXiv:2302.03587*, 2023.

[28] Elsta. Sawyer at Work - Assembly: Spark Plug Insertion. `https://www.youtube.com/watch?v=EX0O68Xsr58`, Last accessed on 2023-09-15.

[29] Fanuc Europe. High level robotic deburring. `https://www.youtube.com/watch?v=Qrb3GPVe2bE`, Last accessed on 2023-09-15.

[30] Universal Robots A/S. SCREW DRIVING. `https://video.universal-robots.com/screw-driving`, Last accessed on 2023-09-15.

[31] Universal Robots A/S. Assembly - A collaborative robot arm from Universal Robots can reduce assembly times, increase production speed and improve quality. `https://video.universal-robots.com/assembly`, Last accessed on 2023-09-15.

[32] Gimelli Engineering AG. Dauertest Batteriefach - Gimelli Engineering AG. `https://www.youtube.com/watch?v=4gGQGJ6HrhA&t=12s`, Last accessed on 2023-09-15.

[33] Franka Emika. Touchscreen device testing. `https://www.youtube.com/watch?v=yC-6T6lMgyI&t=6s`, Last accessed on 2023-09-15.

[34] Rethink Robotics GmbH. Sawyer - inteligentny robot współpracujacy. `https://www.youtube.com/watch?v=tU3yMfoZm8U&list=PLe7Pue7SRXZGgTDaUWC2F0SuKXZKlIlrG`, Last accessed on 2023-09-15.

[35] Franka Emika. Ram. `https://www.youtube.com/watch?v=HQt7XZB-rts`, Last accessed on 2023-09-15.

[36] Peter So, Andriy Sarabakha, Fan Wu, Utku Culha, Fares J. Abu-Dakka, and Sami Haddadin. Digital robot judge: Building a task-centric performance database of real-world manipulation with electronic task boards. *IEEE Robotics & Automation Magazine*, pages 2–14, 2023.

[37] Zhou Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.