Identification of Common Force-based Robot Skills from the Human and Robot Perspective

Thomas Eiband[1], and Dongheui Lee[1,2]

Abstract— **Learning from Demonstration (LfD) can significantly speed up the knowledge transfer from human to robot, which has been proven for relatively unconstrained actions such as pick and place. However, transferring contact or force-based skills (contact skills) to a robot is noticeably harder since force and position constraints need to be considered simultaneously. We propose a set of contact skills, which differ in the force and kinematic constraints. In a first user study, several subjects were asked to term a variety of force-based interactions, from which skill names were derived. In a second and third user study, the identified skill names are used to let a test group of subjects classify the shown interactions. To evaluate the skill recognition from the robot perspective, we propose a feature-based classification scheme to recognize such skills with a robotic system in a LfD setting. Our findings prove that humans are able to understand the meaning of the different skills and, using the classification pipeline, the robot is able to recognize the different skills from human demonstrations.**

Fig. 1: Concept overview of a Learning from Demonstration setting with proposed skills.

I. INTRODUCTION

Contact-based skills, also known as in-contact, compliant, interaction, or force-based skills, are required whenever the human or a robot needs to physically interact with the environment to achieve a specific goal. For a human, the selection of an appropriate interaction strategy is natural but a robot does not automatically know which constraints need to be satisfied to execute a task. Therefore, robot skill are designed to execute a specific behavior such as picking up or placing an object [1]. The extraction of skills from human demonstration has been mainly considered for position based tasks [2]–[5]. However, not much attention has been paid to recognize the interaction behaviour in the force domain, despite there being an intuitive need for contact forces in almost all human actions. Similarly, a robot need to employ contact forces and force feedback to interact with its environment, for instance in assembly tasks.

The problem is that a demonstration does not contain the information about the required controller and task constraints explicitly. Therefore, we want to identify the contact skill first and connect it to an appropriate control algorithm in the future. An example is a touch skill, which requires the robot to search for a surface and stop when a part is detected via the contact force. If there is no knowledge about the touch skill during execution, the system does not know that it should stop when contact is made.

From the perspective of explainability, so called force dynamics in linguistics were analyzed in [6] with only a weak connection to physical forces that can be exerted by a human or a robot. Instead, we present expressions for skills that are understood clearly as physically force-based interactions. Although force-based constraints have been presented in [7], no generic set of robotic contact skills has been proposed before that is also validated to be interpretable by users.

This work's aim is to clarify that different contact skills exists, which can be interpreted by humans as well as being recognizable by robotic systems. We would like to found a basis, such that a variety of robot skills can be standardized in the future to be interpretable by both human and machine. The proposed contact skills that were identified in several user studies are a first step towards this goal. The contribution of this work is divided in three parts:

(i) We propose a set of generic contact-skills with interpretable skill names that were extracted in a first user study.

(ii) We prove the human interpretability of such skills in a second and third user study

(iii) We propose a set of features and a computationally efficient classification scheme that allows to recognize such skills from kinesthetic teaching data on a real robot.

Our work is outlined as follows: related work is introduced in Sec. II, followed by the identification of skills through the conducted user studies (Sec. III). Next, we explain the approach for robotic skill recognition (Sec. IV). The classification results on our dataset are presented in the experiments section (Sec. V). Finally, we discuss and conclude our findings in Sec. VI.

[1] German Aerospace Center (DLR), Institute of Robotics and Mechatronics, Wessling, Germany
[2] Technical University of Munich, Chair of Human-centered Assistive Robotics, Munich, Germany
{thomas.eiband, dongheui.lee}@dlr.de

II. RELATED WORK

A. Skill Recognition and Classification

Many frameworks for action or skill classification use a predefined set of actions or skills, e.g. [8], [9]. These skills are usually task specific and have not been proven to be interpretable by humans. In this work, we are not only proposing a number of contact skills, but we also prove that the names are understandable and distinguishable by a variety of people with technical and non-technical background.

Force and motion based skill detection in robotics has been shown in earlier works, such as [10], where contact state transitions has been classified, given a task model in the form of a finite state machine. In a surgical skill evaluation in [11], Hidden Markov Models were employed to detect manual skills in a fixed environment. In both works, the classification is task-specific and would have to be re-learned if the environment changes. Instead, we want to detect the inherent characteristics of a contact skill that are not constrained to a specific environment. In a more generic context, Morrow and Khosla [7] introduced force primitives and claim that only 20 different relative motions between two mechanical parts are possible. Instead of actions or skills, their primitives are labelled, for example *edge parallel to surface* or *surface against surface*. It is left open how the process of task decomposition into primitives can be automated or even learned from demonstration. In [12], only contact events are classified from time-series data in a task specific setup without considering additional interaction behaviors.

In many works, motions or human gestures are classified given a model or candidate of the motion to be compared with. And this constrains all these methods to be task specific. Instead of using representations of previously learned motions, we consider features derived from the motion and force data that make our classifier independent from reference frames or specific environments. This implies that the recognition generalizes over different tools, materials or objects as shown later on.

B. Robotic Contact Skills

In the LfD domain, an approach to demonstrate or teach contact-based robot behaviors is kinesthetic teaching, as it allows to record contact forces alongside motions and overcomes the kinematic correspondence problem between teacher and robot. This has been shown in various approaches with applications to peg-in-hole [13], [14], wood planing [15], engraving [16], touching or contact-sensing [12], [17], [18], pushing [19], and ironing [20].

We found out that all the aforementioned methods are parameterized and preselected by a domain expert. This is not a drawback of the presented algorithms itself; rather, it shows that appropriate controllers exist for different applications. However, this raises the question: How would a novice user program such tasks without the help of an expert? In the following we want to cope with this problem such that a suitable skill is identified and an application specific robot behavior becomes feasible.

III. SKILL IDENTIFICATION

We propose eight skills that fundamentally differ in the way, how they act onto the environment, as shown in Fig. 1. In the following section, we briefly explain the constraints that we associated with these skills on the technical side.

A. Skill Constraints and Properties

We formulate the skill characteristics such that a hybrid force-position or force-impedance controller would be able to reproduce them. The underlying constraints could for example be described with the task frame formalism [21]. This means that the robot is able to track a position in one subspace and to control the wrench in another subspace. An execution strategy could further involve the switching of different controllers with a finite state machine [22].

For each proposed skill, we also refer to an exemplary application and to a related implementation from the state of the art. *Touch* is used for a slight contact with the environment, e.g. to detect physical constraints, the presence of objects, and for haptic search primitives [17], [23]. *Press* achieves a desired force level onto a counterpart, such as button pressing [19]. *Press and Slide* can be used in the task of polishing [24] or surface smoothing [15], where usually a contact force normal to the surface is maintained. *Contour* is intended to follow along a free-form path, e.g for deburring or edge smoothing of mechanical parts [25]. Hereby, a force normal to the edge need to be maintained in order to apply constant pressure with the tool. If the robot should reproduce a rotational motion, this can be achieved with the *turn* skill, for instance to screw a bottle cap, or as shown in valve turning [26]. The insertion skill is designed for physically constrained object insertion operations. For this demanding operation, several algorithms were presented, such as in [14], or an overview of algorithms in [27]. The *hand-over* skill accounts for situations in which a physical interaction with the operator is required. Examples are given in [28], where actions such as "receiving" or "handing over" an object are simply triggered by touching the robot's structure. Finally, *push* can be used in non-prehensile manipulation to shift an object over a surface without grasping it [29].

B. User Study Round 1

The goal of the first user study was to identify understandable skill names and to prove that the proposed skills can be differentiated by humans.

1) Experimental Design: First, we recorded a total of 16 videos with force-based interactions between a human hand and the environment, with some examples shown in the images of Fig. 2. The task of the user was to propose a suitable label or description for each interaction that was shown without any restrictions in naming. Since different names were highly likely due to the variety in the textual expressions among people, we generalized the data. Hereby, each interaction of the same class was performed in two different situations. For example, touching is performed once with a tool on one object (a pen as shown in Fig. 2a) and once with the fingers on another object. All interactions can be seen in the accompanying video. We recruited 26 subjects
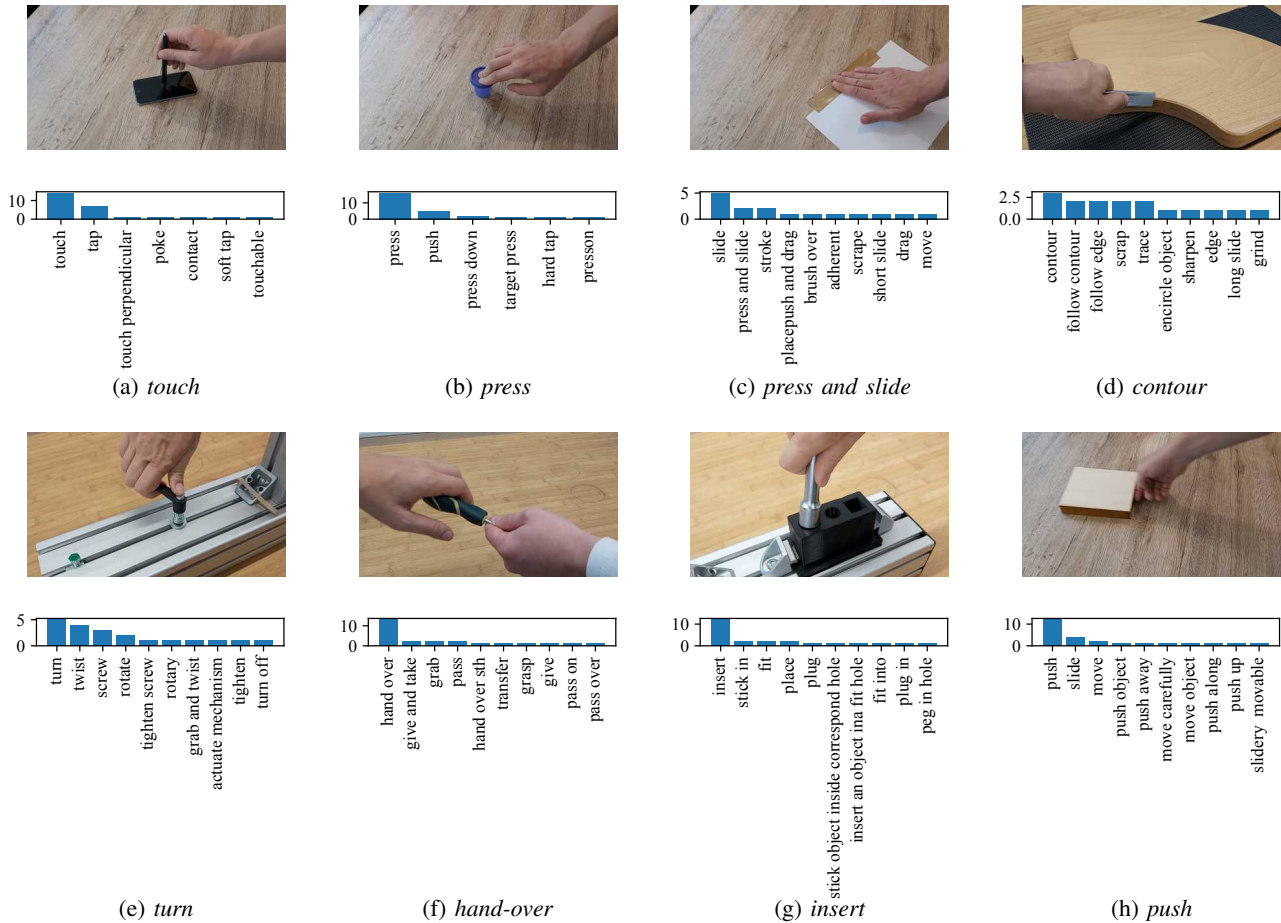
Fig. 2: Some exemplary interactions that were presented in the user study with the number of provided designations below as bar chart. A maximum of ten answers is shown per skill for reasons of legibility.

(10 female, 16 male, average age of 28.6 (SD = 4.04)) that rated their level of robotics experience (RE) with an average of 2.35 (SD = 1.20)) and their English skills (ES) with an average of 3.88 (SD = 0.71), using a scale from 1 to 5. Next, we asked the subjects to specify a suitable English name for each interaction without further constraints.

2) Results: All received names were preprocessed by lemmatization using WordNet [30] to maximize the number of matching designations. This step removed the ambiguity from highly similar names. For instance, *press* and *pressing* both resulted in *press*. The results of the labeling process is shown in the bar charts of Fig.2, denoting the 10 most common names. For some skills such as *touch*, there was a highly significant match for a commonly used designation. More confusion could be observed for example in *contour*. From these results, we term each skill with the name that has the highest occurrence.

C. User Study Round 2

The goal of this study is to prove that a human test group is able to understand the previously extracted names in the form of a classification task.

1) Experimental Design: We use the extracted skill names from the first round as labels in a classification task. We recruited 15 subjects (8 female, 7 male, average age

33.3 (SD = 5.43), RE = 2.13 (SD = 1.13), EE = 3.13 (SD = 0.83)). Here, the subjects are shown the same 16 videos as in the previous round but were only able to select from the provided labels. This study can be seen as testing a classifier, where the classification model is represented by the common comprehension that the people share. However, in order to prove that a generic skill name would be understandable by a variety of people, we did not ask any participants from the previous study to participate in the second round.

2) Results: The classification results are reported in the confusion matrix in Fig. 9a. For better comparability, we put it beside the later results of our proposed classification scheme in Fig. 9c. The average accuracy of the subjects labeling is relatively high at 0.89. Only the accuracy of the push skill is noticeably lower at 0.42. Interestingly, this skill was mostly confused with *slide*, as many subjects interpreted the fact that an object is shifted over a surface as *slide* skill. However, *slide* was intended to apply pressure onto a surface and move along it. This source of misinterpretation led to the design of a follow-up study, as described in the next section.

D. User Study Round 3

1) Experimental Design: The third and final user study was designed to be the same as the previous round. However,
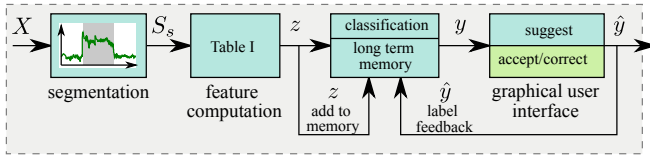
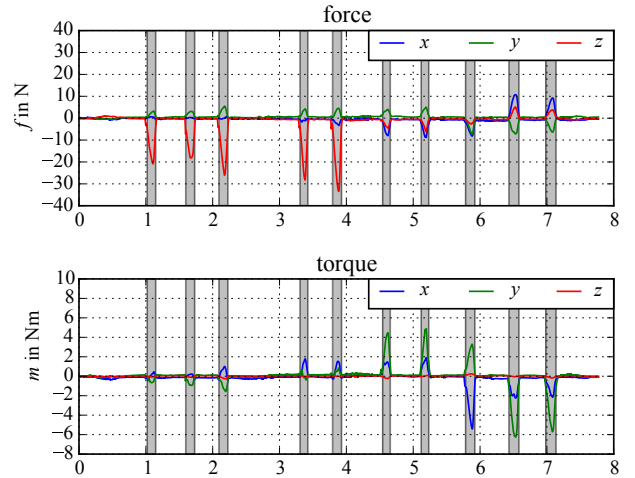Fig. 3: An overview of the system modules and data flow.

Fig. 4: Segmented demonstration of multiple touch events, showing the in-contact segments in gray. From left to right over time, first mainly the force and then the torque signals caused the segmentation to identify contact states.

the skill *slide* is renamed to *press and slide*, which received the second highest score in Fig. 2c. A number of 8 subjects was recruited (2 female, 6 male, average age 28.8 ($SD = 4.13$), $RE = 3.00$ ($SD = 1.51$), $ES = 3.75$ ($SD = 0.71$)).

2) Results: The classification results for this modified study are shown in Fig. 9b. The average accuracy lies at 88.5% comparably to the previous study, but the accuracy for the skill, which is now termed *press and slide* increased from 0.43 to 0.64. Given the improved accuracy for the push skill, we argue that the names are more interpretable by a general audience and specify the final set of proposed skills as presented in Fig. 1.

IV. ROBOTIC SKILL RECOGNITION

We solve our skill recognition problem by a feature-based time-series classification. The technical implementation is structured in the modules shown in Fig.3. The demonstration data X is fed into a segmentation module, which extracts the contact segments S_s. Later on, the classification module computes a feature vector z over the contact segments time-series (Sec. IV-C) and classifies each of these segments (Sec. IV-D) using the class label y. Finally, the system assumes the user's agreement in the case that action is undertaken, but allows the user to correct the skill at any time represented by label $\hat{y}$. In any case, new feature vector z and corresponding label $\hat{y}$ can be added to the systems knowledge.

A. Task Demonstration

The task is demonstrated via kinesthetic teaching, using a backdriveable robot that is equipped with an external force torque sensor (FTS), which is mounted between robot wrist and gripper. During teaching, robot position $p \in \mathbb{R}^3$, orientation represented by quaternions $o \in \mathbb{R}^4$, and external wrench $w = [f, m] = [f_x, f_y, f_z, m_x, m_y, m_z] \in \mathbb{R}^6$ consisting of force f and torque m measured by the FTS are recorded. Additionally, the gripper opening $g \in \mathbb{R}$ and grasp status $h \in \{-1, 0, 1\}$ (-1: no object in gripper, 0: gripper moving, 1: object in gripper) is logged. A sample at time t is given as $x_t = [p, o, w, g, h]$ and a complete demonstration is denoted as $X = [x_1, ..., x_{N_d}]^T \in \mathbb{R}^{N_d \times 15}$ with N_d samples. All spatial data and associated wrenches are recorded in the robot's base frame.

B. Contact State Segmentation

A task demonstration X may consist of multiple alternating phases of free motion and contact with the environment. We use the segmentation approach adopted from [17], which logically combines thresholds on absolute force and torque

values to differentiate these two possible states. The segmentation criterion is

$$(\|f_t\| < \delta_f \wedge \|f_{t+1}\| > \delta_f) \vee (\|m_t\| < \delta_t \wedge \|m_{t+1}\| > \delta_m),$$

given the force and torque vectors f_t and m_t at timestep t and its respective thresholds δ_f and δ_m.

Finally, small free motion gaps in between contact segments, which are shorter than 250 ms, are removed in a post-processing step. This is motivated by the fact that a user hardly maintains contact in complex scenarios such as following an arbitrary contour with the robot tool while facing the robot's inertia. The obtained segments S_s with $s \in \{1, .., M\}$ are forwarded to the skill recognition module, which is explained in the next section. An example of a demonstration with multiple segmented contacts is provided in Fig. 4.

C. Feature Design

We introduce a set of 30 features applied on the motion and wrench data (Table I), which is computed on each contact segment S_s. The features are designed such that they are invariant to a task specific reference frame and shapes of motion and force profiles. For example, the motion that is applied in a *contour* skill can be of different forms regarding different environments, but still the same skill is meant. Consequently, the classification is invariant of scaling and rotation of the demonstration. Most of these features derive a relation between the motion and force modalities such as *Work* (ID 17, 18), *No. of Zero-power crossings* (ID 21) and the *Linear regression correlation* (ID 22, 23). In contrast to many works that learn a parametric model of a motion that is tightly related to a specific task, we target on the prediction of generic skills that are invariant of a specific environment.

Table I is structured such that it shows an ID of each feature on the leftmost column, that is later on used in the feature analysis. Next, the feature names are trailed by the sets of input variables that were independently calculated.

TABLE I: Contact Features

| ID	Feature	Equation or Description
0	Duration	$N = \text{length}(\boldsymbol{X})$
1..2	Path length $(\boldsymbol{p}, \boldsymbol{o})$	$d_{\text{x}} = \sum_{t=1}^{N-1} \|\dot{\boldsymbol{x}}_t\|$
3..4	Path length ratio $(\boldsymbol{p}, \boldsymbol{o})$	$\frac{\|\boldsymbol{x}_N - \boldsymbol{x}_1\|}{d_{\text{x}}}$
5..6	Distance $(\boldsymbol{p}, \boldsymbol{o})$	$\|\boldsymbol{x}_N - \boldsymbol{x}_1\|$
7..10	Time to max. $(\dot{\boldsymbol{p}}, \dot{\boldsymbol{o}}, \boldsymbol{f}, \boldsymbol{m})$	$\underset{t}{\text{argmax}}\{\boldsymbol{x}_t\}$
11..14	Mean absolute value $(\dot{\boldsymbol{p}}, \dot{\boldsymbol{o}}, \boldsymbol{f}, \boldsymbol{m})$	$\frac{1}{N} \sum_{t=1}^{N} \|\boldsymbol{x}_t\|$
15..16	Normalized sum $(\boldsymbol{f}, \boldsymbol{m})$	$\frac{1}{N} \left\| \sum_{t=1}^{N} \boldsymbol{x}_t \right\|$
17..18	Work $((\dot{\boldsymbol{p}}, \boldsymbol{f}), (\dot{\boldsymbol{o}}, \boldsymbol{m}))$	$w = \sum_{t=1}^{N} \boldsymbol{f}_t^T \cdot \dot{\boldsymbol{p}}_t$
19..20	Mean power	$\frac{1}{N} w$
21	No. of Zero-power crossings (ZPC)	$\left|\text{ZPC}(\boldsymbol{f}_t^T \cdot \dot{\boldsymbol{p}}_t)\right|$
22..23	Lin. regress. corr. $((\dot{\boldsymbol{p}}, \boldsymbol{f}), (\dot{\boldsymbol{o}}, \boldsymbol{m}))$	coeff. of determ.: R^2
24	Position linearity	v_{e1}
25	Position planarity	$v_{\text{e1,e2}}$
26..27	Relative spatial variance $(\boldsymbol{p}, \boldsymbol{o})$	$\frac{1}{d_{\text{x}}} \|\text{var}(\boldsymbol{x}_{1..N})\|$
28..29	Relative wrench variance $(\boldsymbol{f}, \boldsymbol{m})$	$\frac{1}{\max(\|\boldsymbol{x}\|)} \|\text{var}(\boldsymbol{x})\|$
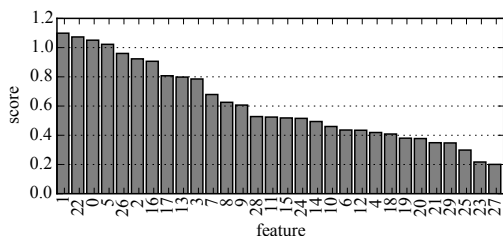
Fig. 5: Score of each feature based on the mutual information analysis.

For example, "Path length $(\boldsymbol{p}, \boldsymbol{o})$" describes the position and orientation result in two independent features with ID's 1 and 2. Hereby, $\boldsymbol{x}$ serves as a placeholder for each of the input variables. The dot notation, e.g. $\dot{\boldsymbol{x}}_t$, denotes the first derivative over time and the index refers to the sample time t. The number of samples in a segment is given as N. The Euclidean norm is denoted as $\|...\|$. Subsequently, we explain those features where we think that a more detailed description is required. *Zero-power crossings (ZPC)* is the quantity of events, where the power, computed as the inner product of force $\boldsymbol{f}_t$ and linear velocity $\dot{\boldsymbol{p}}_t$ at time t, changes its sign, which indicates that the energy flow from the robot to the environment is reversed. *Position linearity* v_{e1} is defined as the ratio of explained variance of the first principal component by applying a Principal Component Analysis (PCA) on the position trajectory, associated to a line in task space. Similarly, *position planarity* is the ratio of explained variance in the first two principal components of the PCA, associated to a 2D plane in task space. These PCA related features are computed by

$$v_{\text{e1}} = \frac{\text{tr}(\boldsymbol{V}) - \boldsymbol{V}_{1,1}}{\text{tr}(\boldsymbol{V})} \quad \text{and} \quad v_{\text{e1,e2}} = \frac{\text{tr}(\boldsymbol{V}) - \boldsymbol{V}_{1,1} + \boldsymbol{V}_{2,2}}{\text{tr}(\boldsymbol{V})},$$
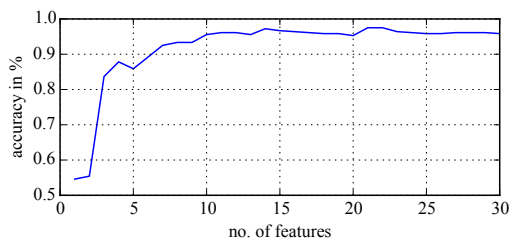

Fig. 6: Classification accuracy over the reduced feature sets used in the SVM.

given the diagonal matrix of eigenvalues $\boldsymbol{V}$ sorted in descending order obtained from the PCA. *Linear regression correlation* is the coefficient of determination R^2, which is obtained from a linear regression that fits force $\boldsymbol{f}$ as function of linear velocity $\dot{\boldsymbol{p}}$ and torque $\boldsymbol{m}$ as function of angular velocity $\dot{\boldsymbol{o}}$.

We conducted a mutual information analysis [31] to gain insights about the feature's importance. After sorting the features by their score in descending order, we iterate over the subsets of the k-best scoring features and apply a cross validation each time using the support vector classification that is described in the following subsection. The resulting classification accuracies are shown in Fig. 6. After adding enough features to the training set, the accuracy converges to 96%

D. Support Vector Classification

We applied a support vector machine (SVM) on our feature set since preliminary tests revealed that a Neural Network (NN) designed as a multilayer perceptron (MLP) did not outperform the SVM on the same feature set. A NN might perform better on more complex data without manually defined features but might also require a much larger dataset and data augmentation methods. Consider that the proposed features are pose-invariant and some are scaling-invariant with respect to the coordinate frame, where the skill is applied. Beside that, a SVM is easy to optimize and has a much faster prediction time compared to NNs. Additionally, SVMs can compute confidence measures for the class membership score of a prediction, which can be a useful property in future work.

Using the initial training set, the features of all available contact segments are normalized to values between $[0, 1]$, yielding to a feature vector $\boldsymbol{z} \in \mathbb{R}^{30}$ for each segment. The SVM has been configured as classifier with radial basis function kernels. The best regularization parameter ($C = 100$) of the training loss function has been identified by 5-fold Cross-validation. Finally, the SVM is able to predict the label of the aforementioned skills given a human demonstration.
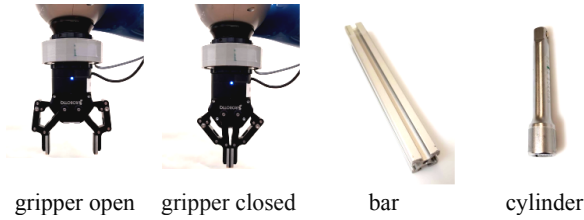
gripper open   gripper closed   bar   cylinder

Fig. 7: The tools used to collect the dataset. Bar and cylinder have been grasped by the gripper during demonstration.

table   wood form   rubber mat   foam mat

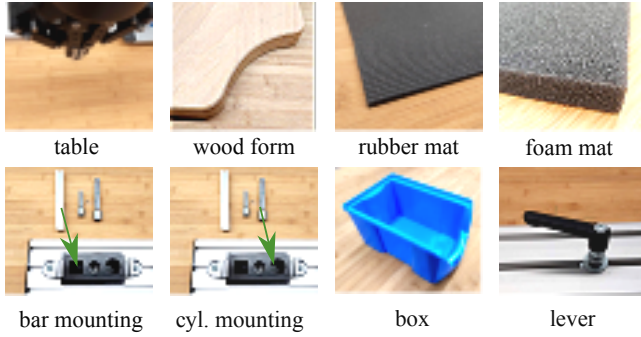bar mounting   cyl. mounting   box   lever

Fig. 8: Different environments in dataset collection.

## V. ROBOT EXPERIMENTS

A. Dataset

We created a publicly available dataset from kinesthetic teaching demonstrations, which we named KROCO[1] (Kinesthetic Robot Contacts). It includes motion and wrench data, considering different tools (Fig. 7), environments (Fig. 8) and a number of diverse skills to facilitate our research. The dataset that consists of 400 labeled interactions has been collected by the experimenter. Exemplary samples for each skill can be seen in the accompanying video.

Touch and *press* were recorded with different tool frames relative to the environment. *Slide* has been trained on different surfaces with varying friction coefficients. The wood form and box edges has been used for *contour* following and *push*. Lever rotations (lever in Fig. 8) were used as interactions for the *turn* skill. User interactions for the *hand over* were recorded by touching the end-effector during kinesthetic teaching, including small position displacements.

B. Robot Classification Results

First, contact segmentation is applied to extract the contact-segments of all demonstrations. Next, the feature matrix is computed over these segments. Thereafter, the described SVM classifier is trained on the feature matrix $Z \in \mathbb{R}^{400 \times 30}$ and the manually labeled segments.

The 5-fold cross-validation accuracy reaches $96.18\,\%$ with a training time of $1.39\,s$ and an overall prediction time of $3 \cdot 10^{-4}\,s$. Noticeable, the SVM is computationally very efficient and allows online usage. The robot's classification performance (Fig. 9c) is comparable to human's performance (Fig. 9b).
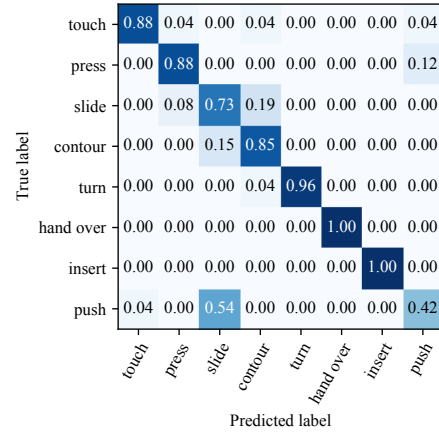
(a) User study round 2: Confusion matrix of the human classified skills with original names.
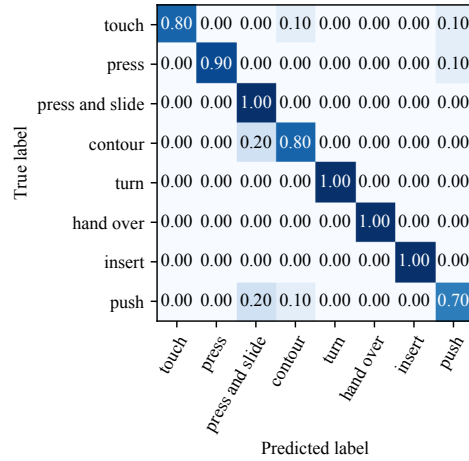
(b) User study round 3: Confusion matrix of the human classified skills with name modification.
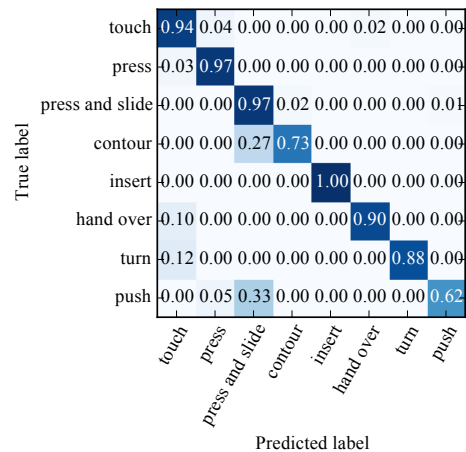
(c) Confusion matrix of the robot's feature-based time-series classification.

Fig. 9: Both confusion matrices for the human and machine classification.

[1] Description: https://teiband.github.io/KROCO/
Dataset: https://dx.doi.org/10.21227/nhea-gt59

VI. Discussion and Conclusion

A. Discussion

Comparing the results of the human and robot classification in Figure 9, it is noticeable that the *push* skill had a lower accuracy in the initial classification in Fig. 9a. In the human case, this might be caused by the fact that objects were pushed along the table, which could be also interpreted as sliding an object over a surface. This shows that the the English verbs of *push* and *slide* might be interpreted similarly. Physically, both skills require a directed force that is applied along a path. However, changing the name from *slide* to *press and slide* could improve the interpretability on the human side (Fig 9b). Since misclassifications are likely to be present in a real system, the robot could suggest skills to the user, who can accept or correct these. The obtained labels could again be used to incrementally retrain the classifier.

At this stage of research, the skill analysis is constrained to the English language. However, the proposed strategy could be used to extract the same information for other languages in the future.

B. Conclusion

We propose a set of contact skills that can be interpreted by users in the role of robot programmers. Furthermore, we propose an efficient classification method that enables the robot to classify force-based interactions from only the motion and force data without requiring additional sensors. We emphasize that contact skill detection is an important direction in learning from demonstration instead of relying on the assumption that a novice user knows how to specify constraints and parameterize a controller for a specific behavior.

In the future, we will address the extraction of control parameters and compliance frames from the detected skills to allow a robust execution in combination with unconstrained manipulation skills such as pick and place.

References

[1] M. R. Pedersen, L. Nalpantidis, R. S. Andersen, C. Schou, S. Bøgh, V. Krüger, and O. Madsen, "Robot skills for manufacturing: From concept to industrial deployment," *Robotics and Computer-Integrated Manufacturing*, vol. 37, pp. 282–291, 2016.

[2] S. Niekum, S. Osentoski, G. Konidaris, S. Chitta, B. Marthi, and A. G. Barto, "Learning grounded finite-state representations from unstructured demonstrations," *The International Journal of Robotics Research*, vol. 34, no. 2, pp. 131–157, 2015.

[3] F. Steinmetz, V. Nitsch, and F. Stulp, "Intuitive task-level programming by demonstration through semantic skill recognition," *IEEE Robotics and Automation Letters*, vol. 4, no. 4, pp. 3742–3749, Oct 2019.

[4] E. E. Aksoy, A. Abramov, J. Dörr, K. Ning, B. Dellen, and F. Wörgötter, "Learning the semantics of object–action relations by observation," *The International Journal of Robotics Research*, vol. 30, no. 10, pp. 1229–1249, 2011.

[5] N. Abdo, H. Kretzschmar, and C. Stachniss, "From low-level trajectory demonstrations to symbolic actions for planning," in *ICAPS Workshop on Combining Task and Motion Planning for Real-World App*, 2012, pp. 29–36.

[6] L. Talmy, "Force dynamics in language and cognition," *Cognitive science*, vol. 12, no. 1, pp. 49–100, 1988.

[7] J. D. Morrow and P. K. Khosla, "Manipulation task primitives for composing robot skills," in *Proceedings of International Conference on Robotics and Automation*, vol. 4. IEEE, 1997, pp. 3354–3359.

[8] D. Lee, "Gesture, posture, facial interfaces," *Encyclopedia of Robotics, Springer*, 2020.

[9] D. Lee, R. Soloperto, and M. Saveriano, "Bidirectional invariant representation of rigid body motions and its application to gesture recognition and reproduction," *Autonomous Robots*, vol. 42, no. 1, pp. 125–145, 2018.

[10] M. Skubic and R. A. Volz, "Learning force sensory patterns and skills from human demonstration," in *Proceedings of International Conference on Robotics and Automation*. IEEE, 1997, pp. 284–290.

[11] H. Rafii-Tari, C. J. Payne, J. Liu, C. Riga, C. Bicknell, and G.-Z. Yang, "Towards automated surgical skill evaluation of endovascular catheterization tasks based on force and motion signatures," in *IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2015, pp. 1789–1794.

[12] S. Stelter, G. Bartels, and M. Beetz, "Multidimensional time-series shapelets reliably detect and classify contact events in force measurements of wiping actions," *IEEE Robotics and Automation Letters*, vol. 3, no. 1, pp. 320–327, 2018.

[13] A. Kramberger, A. Gams, B. Nemec, D. Chrysostomou, O. Madsen, and A. Ude, "Generalization of orientation trajectories and force-torque profiles for robotic assembly," *Robotics and Autonomous Systems*, vol. 98, pp. 333–346, 2017.

[14] F. J. Abu-Dakka, B. Nemec, J. A. Jørgensen, T. R. Savarimuthu, N. Krüger, and A. Ude, "Adaptation of manipulation skills in physical contact with the environment to reference force profiles," *Autonomous Robots*, vol. 39, no. 2, pp. 199–217, 2015.

[15] A. Montebelli, F. Steinmetz, and V. Kyrki, "On handing down our tools to robots: Single-phase kinesthetic teaching for dynamic in-contact tasks," in *IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2015, pp. 5628–5634.

[16] V. Koropouli, D. Lee, and S. Hirche, "Learning interaction control policies by demonstration," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2011, pp. 344–349.

[17] T. Eiband, M. Saveriano, and D. Lee, "Learning haptic exploration schemes for adaptive task execution," in *IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 7048–7054.

[18] A. Stolt, M. Linderoth, A. Robertsson, and R. Johansson, "Detection of contact force transients in robotic assembly," in *IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2015, pp. 962–968.

[19] M. Racca, J. Pajarinen, A. Montebelli, and V. Kyrki, "Learning in-contact control strategies from demonstration," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2016, pp. 688–695.

[20] P. Kormushev, S. Calinon, and D. G. Caldwell, "Imitation learning of positional and force skills demonstrated via kinesthetic teaching and haptic input," *Advanced Robotics*, vol. 25, no. 5, pp. 581–603, 2011.

[21] H. Bruyninckx and J. De Schutter, "Specification of force-controlled actions in the "task frame formalism"-a synthesis," *IEEE transactions on robotics and automation*, vol. 12, no. 4, pp. 581–589, 1996.

[22] M. Skubic and R. A. Volz, "Acquiring robust, force-based assembly skills from human demonstration," *IEEE Transactions on Robotics and Automation*, vol. 16, no. 6, pp. 772–781, 2000.

[23] S. Luo, J. Bimbo, R. Dahiya, and H. Liu, "Robotic tactile perception of object properties: A review," *Mechatronics*, vol. 48, pp. 54–67, 2017.

[24] S. S. M. Salehian and A. Billard, "A dynamical-system-based approach for controlling robotic manipulators during noncontact/contact transitions," *IEEE Robotics and Automation Letters*, vol. 3, no. 4, pp. 2738–2745, 2018.

[25] F.-Y. Hsu and L.-C. Fu, "Intelligent robot deburring using adaptive fuzzy hybrid position/force control," *IEEE Transactions on Robotics and Automation*, vol. 16, no. 4, pp. 325–335, 2000.

[26] F. J. Abu-Dakka, L. Rozo, and D. G. Caldwell, "Force-based variable impedance learning for robotic manipulation," *Robotics and Autonomous Systems*, vol. 109, pp. 156–167, 2018.

[27] J. Xu, Z. Hou, Z. Liu, and H. Qiao, "Compare contact model-based control and contact model-free learning: A survey of robotic peg-in-hole assembly strategies," arXiv, 2019.

[28] S. Haddadin, M. Suppa, S. Fuchs, T. Bodenmüller, A. Albu-Schäffer, and G. Hirzinger, "Towards the robotic co-worker," in *Robotics Research*. Springer, 2011, pp. 261–282.

[29] J. Stüber, C. Zito, and R. Stolkin, "Let's push things forward: A survey on robot pushing," *Frontiers in Robotics and AI*, vol. 7, p. 8, 2020.

[30] C. Fellbaum, "Wordnet," *The encyclopedia of applied linguistics*, 2012.

[31] B. C. Ross, "Mutual information between discrete and continuous data sets," *PloS one*, vol. 9, no. 2, 2014.