Slippage Detection Generalizing to Grasping of Unknown Objects Using Machine Learning With Novel Features

Ioannis Agriomallos, Stefanos Doltsinis, *Member, IEEE*, Ioanna Mitsioni, and Zoe Doulgeri, *Senior Member, IEEE*

***Abstract*—Real-time grasp stability is based on successful slippage detection. In this work, we consider slippage detection as a binary problem (slip, stable) and we propose a novel set of temporal and frequential features, extracted from force norm profiles and collected during reliable ground truth labeling processes, finally employed within the machine learning classification techniques. Classification performance of the proposed scheme, with respect to its success and generalization ability, is assessed systematically utilizing different performance metrics that clarify class predictions as opposed to most of the reported works. We show that our proposed feature extraction method improves classification performance over the commonly used feature sets, even when trained with one surface, and generalizes successfully to unseen ones. The trained classifier is tested on a completely new task and object, for real-time slippage detection, showing high detection accuracy. Finally, the classifier is tested in a different experimental layout with a different force sensor. Experiments are conducted on unseen surfaces for a variety of sampling frequencies, for both translational and rotational slippage, with the proposed approach showing fast and accurate detection in all cases.**

***Index Terms*—Slippage detection, perception for grasping and manipulation, learning and adaptive systems.**

I. Introduction

ROBOTS have revolutionized our everyday lives improving efficiency, productivity and precision, with applications in industry, medical or service assistance e.t.c. As technology advances, robots will be required to safely interact, grasp and manipulate objects, so as to complete desired tasks. Grasp stability maintenance is one of the human dexterities that is ultimately desirable but also difficultly achieved.

Humans are able to preserve grasp stability, even of unknown objects, regulating the applied force for successful lifting, handling and manipulation accordingly. They achieve that by detecting skin deformations and micro-vibrations during the occurrence of a tactile incident (touching, lifting, sliding, placing), as they possess skin receptors (fast adapting (FA-I, FA-II) and slow adapting (SA-I, SA-II)), which respond to frequencies mainly up to 400 Hz and have diverse stimulation patterns [1], whose combination facilitates the detection of different tactile events.

In a similar manner, if robots could detect slippage in real time, it would be a useful step towards stable grasping of objects. Two main approaches to slippage detection can be distinguished [2]. On one hand, the physical properties of the contacting surfaces are exploited, while on the other hand, vibrations or patterns related to slippage are considered.

For the first approach, most of the related works attempt detection utilizing the friction cone notion, where if the ratio of tangential to normal force components is higher than the friction coefficient, the contact is slipping. This approach is extremely effective and accurate, but it requires exact knowledge of both contact normal and friction coefficient, that are usually unknown or uncertain. Song *et al.* [3] and Ajoudani *et al.* [4] estimate the friction coefficient through exploratory motions before grasp establishment. Wettels *et al.* [5] assume a conservative friction coefficient of $\mu = 0.5$, compute tangential forces by applying Kalman filtering to pressure data acquired from tactels and achieve stability by using them in a ratio to adjust the grip. Vina *et al.* [6] attempt to adaptively control pivoting of an object based on friction cone considerations, while having estimated the translational and rotational friction coefficients. What is common in the majority of these works, is that they require an initial exploratory phase to estimate the friction coefficient before attempting their task, while 3D force measurements or their estimations have to be available for the friction cone analysis.

For the second approach, patterns embedded in the sensor's measurements or its frequency content are considered. Some researchers use empirical thresholds [7]–[11], while others apply machine learning algorithms [12]–[17]. In the first category, Stachowsky *et al.* [7] detect slippage based on the covariance matrix of the input signals from two fingers, while Romano *et al.*

Manuscript received September 8, 2017; accepted January 1, 2018. Date of publication January 15, 2018; date of current version February 1, 2018. This letter was recommended for publication by Associate Editor M. A. Roa and Editor H. Ding upon evaluation of the Reviewers' comments. This work was supported by the EU Horizon 2020 research and innovation programme under Grant 643433, project RAMCIP. *(Corresponding author: I. Agriomallos.)*

I. Agriomallos and S. Doltinis are with the Center of Research and Technology Hellas, Information Technologies Institute, Thessaloniki 57001, Greece (e-mail: jagrio@iti.gr; doltsinis@iti.gr).

I. Mitsioni is with the Department of Electrical and Computer Engineering, Aristotle University of Thessaloniki, Thessaloniki 54124, Greece (e-mail: igmitsio@ece.auth.gr).

Z. Doulgeri is with the Center of Research and Technology Hellas, Information Technologies Institute, Thessaloniki 57001, Greece (e-mail: doulger@eng.auth.gr), and also with the Department of Electrical and Computer Engineering Aristotle University of Thessaloniki, Thessaloniki 54124, Greece (e-mail: doulgeri@iti.gr).

This letter has supplemental downloadable multimedia material available at http://ieeexplore.ieee.org, provided by the authors. The Supplementary Materials contain a video showing the detection of slippage during grasping of unknown objects. Different setups distinguishing between stable and slip. Both training and testing procedures are shown. This material is 8.33 MB in size.

Digital Object Identifier 10.1109/LRA.2018.2793346

[8] use high-pass filters on tactile measurements and empirically set a threshold to designate slip. Finally, Su *et al.* [9], Cordella *et al.* [10] and Kaboli *et al.* [11] use the tangential force's derivative as a slippage indicator and compare it to a constant value. When it comes to approaches deploying empirical thresholds, fine tuning depending on the grasping task is a prerequisite and the method's generalization ability is at stake, since the need for a new threshold arises along with a new task.

To address some of the above issues, machine learning approaches have been adopted, as there is no need for explicit modeling of the physical properties of the contacting surfaces. Bekiroglu *et al.* [12] evaluate the probability of grasp stability based on the output of classifiers trained on combined information from the tactile sensing array, the hand, the object and the grasp. Veiga *et al.* [13] detect slippage by comparing SVM and Random Forests trained with three different feature approaches, that examine the phenomenon's dependency on current and past data acquired by the BioTac sensor measurements. Goger *et al.* [14], Schoepfer *et al.* [15] and Meier *et al.* [16] use short-time Fourier transform (STFT) to estimate the frequency content of the input signal and detect slippage by using classification. Roberge *et al.* [17] take the power spectral density (PSD) of the tactile dynamic signal to perform unsupervised feature learning and create a dictionary of sparse representations, which is used by an SVM to classify slip instances. However, most of these methods do not test their approaches in untrained grasping tasks and therefore may require model retraining for application to a different grasping task than the one being trained on. When this is not the case ([12], [13]), classifiers reduce their accuracy.

Despite that existent approaches present promising results, they do not assess the classification performance in a systematic way, but rather focus on incorporating slippage as part of their grasping strategy. This is to be expected, since grasp stability is usually the ultimate goal, yet it does not facilitate the establishment of a common frame of reference to accurately evaluate a slippage detection methodology and quantify its success and generalization ability. Although it is not unusual to use different performance metrics depending on the problem at hand, solely reporting one of them does not give a well-rounded estimation of performance. This is due to the fact that most metrics used, either do not account for imbalances in class prediction ([12], [14], [16], [17] only report accuracy), or are biased towards successful recognition of either the positive ([13] only reports F_{score}) or negative class, thus making comparison among methods difficult. Additionally, apart from those who use off-the-shelf sensors e.g., BioTac and OptoForce, most works use custom-made tactile arrays, which also does not help establishing a common ground for comparison. Finally, to our knowledge, except for [13] and [17], no work has addressed issues of generalization with respect to unknown surfaces effectively and in detail, either online or offline. To evaluate its efficiency, slippage detection should be addressed and evaluated separately and not be interlaced with reaction strategies, since a well-performing detection can lead to successful grasping with the most simple reaction, whereas the most sophisticated reaction may fail with an uncertain detection scheme. Our main goal is to employ machine learning on an enhanced set of features to address the slippage detection problem isolated, presenting analytical confusion matrix results, which show that this approach can lead to a robust learner, able to predict slip and stable samples equally well.

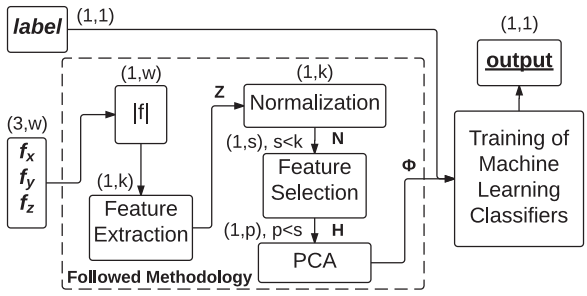

Fig. 1. Followed methodology.

The contributions of this work can be summarized as follows:

- Our methodology introduces a feature vector from contact force norm profiles with a novel composition, containing both temporal and frequential features, leading to improved performance and generalization on unseen objects, without the need to cover more than one surface during training.
- The proposed approach is evaluated in terms of classification accuracy and raw performance metrics rather than being incorporated in a grasping scheme, thus providing a basis for comparison among methods for slippage detection-related tasks.
- The classifier with the proposed features trained on one surface for translational slippage is transferable to another experimental setup utilizing another sensor from the one used for training, accurately detecting both translational and rotational slippage on unknown surfaces, even with reduced sampling frequency.

II. Proposed Slippage Detection Strategy

We consider slippage detection as a binary problem, consisting of two distinct classes (*slip*, *stable*). Machine learning techniques seem the most suitable approach for such classification problems, but require an appropriate design in order to perform well. Fig. 1 depicts the methodology followed in this work. Initially, the norm of force measurements is retrieved by a predefined window (buffer), of length w. This $1 \times w$ input force profile, or hereby called *prefeature*, is computed in an attempt to wrap up its enclosed information before extracting features from the input force signals. This prefeature $\xi = |\mathbf{f}|$ is adopted by most similar approaches [14]–[16] and is also preferred in the current approach due to its compact information representation, simplicity and availability. Additionally, ξ is more general, since it doesn't explicitly require 3D force readings (available in the current setup from OptoForce sensors) and can be deduced from e.g., pressure distribution from tactile sensors as well. From the prefeature, a collection of *time* and *frequency domain features* is extracted to form the corresponding feature vector. Then, a *normalization* procedure takes place, followed by a *feature selection-ranking* process and finalized by a *dimensionality reduction*, which produces the input vector for *training of classifiers*. More details on each step of the methodology are described in the following subsections.

A. Feature Extraction

Traditionally, slippage is detected through vibrations directly related to the phenomenon, by computing the frequency content of the force input signal and particularly the amplitude of its Fourier transform ($|FFT|$) [14]–[16]. In general, force measurements strongly resemble electromyograms (EMGs) or other highly fluctuating signals in time. Assessing the signal's characteristic entities may reveal embedded information, connections and patterns, otherwise disguised and lost and which may be present in both the frequency and time domain.

Vibrations may help detect slippage, but that depends on the sensor and is related mostly to the contacting material of the sensing device, as well as the sampling period, the spatial resolution, the sensitivity and the material's hysteresis. Thus, techniques which are applicable for one type of sensor may not be suitable for other types. Except for human studies on the band of frequencies potentially related to slippage [1], no apparent relation between various kinds of surfaces and slipping frequency is reported. This could be justified by the fact that the characteristic frequencies may not be the same for every slipping material, but may be related to the contacting surfaces and their textures, e.g., ripples of human fingerprint compared to smooth elastic force sensor. Hence, in order to enrich the frequency features, which may reveal repeated values or patterns related to slippage, we consider utilizing time features as well, which may uncover specific attributes of the phenomenon, such as similarity to previous values or characteristics of the signal's amplitude. To a human equivalent, frequency analysis could resemble the sensitive-to-vibrations receptors, while time analysis could be relevant to deformations such as pressure, stretch, twist etc. The motivation for selecting both time and frequency domain features is that their combination may reveal slippage related attributes more reliably.

The feature extraction process involves the computation of representative temporal and frequential features of the selected prefeature ξ and is mainly influenced by Phinyomark *et al.* [18] (P) and Golz *et al.* [19] (G). Table I depicts how the feature function $\sigma(\cdot)$ computes the temporal (T) and frequential (F) features ($k = 3w + 33$ in total) selected for evaluation. The leftmost column shows the domain ($D : \{T, F\}$) and reference ($R : \{P, G\}$) for each feature respectively. The feature extraction is performed on the prefeature ξ resulting in the feature matrix $Z = \sigma(\xi), \in \mathbb{R}^k$. Notice that temporal features are populated by autocorrelation (ACORL), whereas frequential ones are mainly represented by Fourier (FFT) components. ACORL is related to the power spectral density via the FFT and intuitively describes the similarity between observations as a function of the time lag between them. Both ACORL and FFT account for repeating patterns, such as the presence of a periodic signal obscured by noise. If the fundamental frequency in a signal is missing, ACORL can still find it by its harmonic frequencies, while FFT cannot without special processing. Evidently, both are useful and valuable representations, thus included in the feature vector.

To assess the importance of each domain, 4 different feature sets are created, namely one containing only the commonly used Amplitude of the FFT (hereby denoted as AFFT in the F domain of G), one containing all frequency features (denoted as FREQ and including every feature in the F domain), one containing all time features accordingly (TIME) and lastly, one combining all the temporal and frequencial features (BOTH). It is shown that the 2 latter cases, i.e., the time domain features alone (TIME) and their combination with the frequency ones (BOTH), perform and generalize better than the AFFT, which is the most widely used in the literature [14]–[17].

TABLE I
FEATURE EXTRACTION $\sigma(\cdot)$

D,R	Features	Explanation	#
—	Initial signal: x	$x = [x_1, x_2, ..., x_N]$	-
T,P	Integrated Signal (IS)	$IS = \sum_{n=1}^{N} \lvert x_n \rvert$	1
	Mean Absolute Value (MAV)	$MAV = \frac{1}{N} IS$	1
	MAV Slope (MAVSLP)	$MAVSLP_i = MAV_{i+1} - MAV_i$	1
	Simple Square Integral (SSI)	$SSI = \sum_{n=1}^{N} x_n^2$	1
	Variance (VAR)	$VAR = \frac{1}{N-1} SSI$	1
	Root Mean Square (RMS)	$RMS = \sqrt{\frac{1}{N} SSI}$	1
	Waveform Length (WAVL)	$WAVL = \sum_{n=1}^{N-1} \lvert x_{n+1} - x_n \rvert$	1
—	Sign of input x	$sgn(x) = \begin{cases} 1, & if\ x \geq thres \\ 0, & otherwise \end{cases}$	-
T,P	Zero Crossing (ZC)	$ZC = \sum_{n=1}^{N-1} [\ sgn(x_n x_{n+1}) \cap \lvert x_n - x_{n+1} \rvert \geq thres\ ]$	1
	Slope Sign Change (SSC)	$SSC = \sum_{n=2}^{N-1} sgn((x_n - x_{n-1})(x_n - x_{n+1}))$	1
	Willison Amplitude (WAMP)	$WAMP = \sum_{n=1}^{N-1} sgn(\lvert x_n - x_{n+1} \rvert)$	1
	Histogram (HIST)	$HIST = histogram(x)$, number of bins $b = 3$	3
F,P	AutoRegressive Coefficients (ARCO)	$ARCO = [a_1, \ldots, a_i, \ldots, a_p]$, s.t. $x_n = -\sum_{i=1}^{p} a_i x_{n-i} + w_n$, where w_n white noise and $p = 3$ here.	3
—	Fast Fourier Transform (F) and its Power Spectrum (PF)	$F_k = \sum_{n=0}^{N-1} x_n e^{-i2\pi kn/N}, PF_k = F_k^2$, $k = 0, \ldots, N-1$	-
F,P	Mean Frequency (MNF)	$MNF = \sum_{j=1}^{M} f_j PF_j / \sum_{j=1}^{M} PF_j$, where f_j frequency at bin j	1
	Median Frequency (MDF)	$\sum_{j=1}^{MDF} PF_j = \sum_{j=MDF}^{M} PF_j = \frac{1}{2} \sum_{j=1}^{M} PF_j$	1
	Modified Mean Frequency (MMNF)	$MMNF = \sum_{j=1}^{M} f_j AF_j / \sum_{j=1}^{M} AF_j$, where f_j frequency at bin j	1
	Modified Median Frequency (MMDF)	$\sum_{j=1}^{MMDF} AF_j = \sum_{j=MMDF}^{M} AF_j = \frac{1}{2} \sum_{j=1}^{M} AF_j$	1
	Real part (RF) and Imaginary part (IF) of FFT	$RF = \Re\{F\},\ IF = \Im\{F\}$	$2(w/2+1)$
T,G	Mean Value (MV)	$MV = \bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i$	1
	Standard Deviation (STD)	$STD = \sigma = \frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{x})^2$	1
	Maximum (MAX)	$MAX = max(x)$	1
	Range of x-axis (RNGX)	$RNGX = N$	1
	Range of y-axis (RNGY)	$RNGY = max(x) - min(x)$	1
	Median (MED)	$MED = median(x)$	1
	Hjorth Complexity (HJORTH)	$HJORTH = \frac{\sigma_{dd}}{\sigma_d} / \frac{\sigma_d}{\sigma}$, where $\sigma_d = STD(\dot{x})$ and $\sigma_{dd} = STD(\ddot{x})$	1
	Shannon's Entropy (SENTR)	$SENTR = \sum_{i=1}^{b} p_i log_2(p_i)$, where $p_i = HIST_i/b$, b number of bins (3 here)	1
	Skewness (SK)	$SK = \frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{x})/\sigma^3$	1
	Autocorrelation (ACORL)	$ACORL_k = \sum_{n=1}^{N-k} x_{n+k} conj(x_n)$, for $k = 1, \ldots, N$	w
F,G	Amplitude (AF) and Phase (PF) of FFT	$AF = \sqrt{RF^2 + IF^2}$, $PF = arctan(IF/RF)$	$2(w/2+1)$
All features gathered together into one vector $\zeta = \sigma(\cdot)$			$3w + 33$

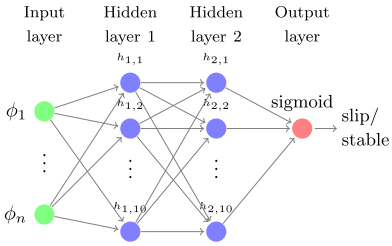

Fig. 2. *MLP* structure.

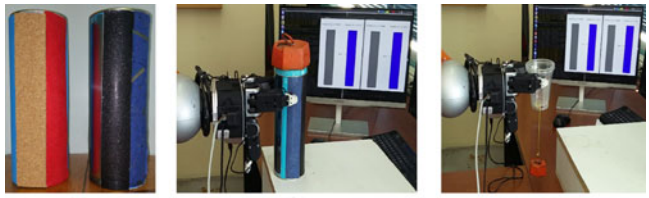

Fig. 3. Setup for validation and real-time experiment. (a) Surfaces of the experiments, (b) Validation and data acquisition, (c) Setup for real-time experiment.

TABLE II
SURFACES USED FOR TRAINING AND TESTING

Surface Num (SN)	Surface ID (SID)	Surface	Surface Num (SN)	Surface ID (SID)	Surface
0	Glossy Blue		3	Cork Paper	
1	Soft Blue		4	Banana Paper	
2	Felt		5	Sandpaper	

B. Normalization, Feature Selection & Dimensionality Reduction

After the feature extraction process, the resulting feature matrix Z is normalized by taking the z-score of each feature, namely subtraction of the mean and division with the standard deviation over all n inputs ($zscore(x_{ij}) = \frac{x_{ij} - \bar{x_j}}{std(x_j)}$), resulting in a normalized feature matrix $N = zscore(Z), \in \mathbb{R}^k$. The purpose of normalization is to calibrate the importance of the feature dimensions and avoid giving too much or too little attention to some of them, resulting in defective training. Subsequently a feature selection-ranking is performed on the normalized feature vector N, based on mutual variable information [20] (association metric searching for relationship between two random variables) and the s best performing features of N are kept, forming a matrix $H = sel(N)[1...s], \in \mathbb{R}^s$. The rationale behind feature selection is that keeping the most prevailing features related to slippage reduces computational complexity. Finally H is processed by principal components analysis (*PCA*), a dimensional transformation and reduction algorithm, forming a matrix $\Phi = pca(H)[1...p], \in \mathbb{R}^p$, by keeping p components of Φ.

Finally, if the forces of the testing set significantly differ from those used for training, the prefeature $|f|$ could be normalized by dividing with the mean of stable (during training) or with the desired contact force (during real-time) to result in a close-to-zero mean signal and alleviate the method from any dependency on training data range.

C. Machine Learning Classification

In supervised learning, given a dataset containing the feature vectors and their labels, an algorithm is trained to infer a mapping function of the feature space. The trained model can then be used to classify new input into one among the classes. For our case, as soon as Φ is computed, it is used along with the corresponding labels as input for the training. A Neural Network or else Multi-Layer Perceptron (*MLP*) is chosen as the classification model, for its efficiency in highly dimensional problems. Fig. 2 shows the chosen MLP structure, which is composed of two hidden layers with 10 neurons each. The training procedure has been performed using the Python programming language and the Scikit-learn library [21], using the Limited-memory BFGS solver and default remaining parameters.

III. EXPERIMENTAL EVALUATION

A. Experimental Setup

Experiments were conducted using a KUKA LWR4+ robotic manipulator equipped with a Shadow Smart Grasping System (*SGS*) Robot Hand and Optoforce 3-Axis force sensors (*OFS*) attached to its finger tips. ROS [22] was used for communication, commanding (joint velocities at 1ms intervals) and recording of data, adjusting the force sensors' sampling rate at their maximum 1 kHz, without filtering.

For the dataset collection, six different surfaces were used and were fixed symmetrically on two cylindric bottles by sets of three surfaces per bottle, as shown in Fig 3(a), so two of the three fingers can be in contact with the same surface during a pinch grasp (Fig. 3(b)). Table II shows the surfaces used for training and testing in an order of ascending roughness.

In order to collect sufficient, but also quantitatively balanced stable and slip force samples, the robot manipulator was commanded to perform a trajectory of equally timed moving (slip) and stationary (stable) phases. More specifically, the bottle was fixed on a table, an initial pinch grasp was established with the desired surface and a fixed trajectory along the length of the object was performed, moving upwards and downwards with respective pauses. This was done for all 6 surfaces, for 3 different sliding speeds (3, 1.5, 1 cm/sec) and for 2 different initial grasping forces (1 N, 2 N), resulting in 72 datasets acquired from both fingers. This experimental setup facilitated the labeling procedure, since when the robot is commanded to move or stay, it will almost immediately slide or stay on the surface respectively, thus eliminating the need for manual labeling along with the risk of errors during this procedure.

Every dataset was passed through a moving window of length w, from which the corresponding feature vector of length $k = 3w + 33$ was extracted. In brief, for each dataset d_i of the n datasets, a moving window of length w is taken, shifting m steps each time until length l of d_i is reached, resulting in a matrix of $n \times ((l - w + 1)/m) \times w$ dimensions. The process described in Subsection II-B is performed on this matrix, transforming the data with a pipeline of normalization, feature selection, dimensionality reduction and classifier training steps. For this work, dataset number is $n = 72$ corresponding to approximately 10000 samples, which are generated using a window of length $w = 1024$ shifted by $m = 10$. Moreover, features

TABLE III
MEAN PERFORMANCE AND STANDARD DEVIATION OF EACH TRAINED SURFACE TESTED ON ALL REMAINING SURFACES

SN		AFFT	FREQ	TIME	BOTH
0	TPR	0.63(±0.21)	0.79(±0.06)	0.87(±0.06)	0.91(±0.03)
	TNR	0.94(±0.05)	0.96(±0.01)	0.91(±0.05)	0.93(±0.05)
1	TPR	0.96(±0.02)	0.94(±0.02)	0.93(±0.04)	0.95(±0.02)
	TNR	0.69(±0.16)	0.85(±0.03)	0.94(±0.04)	0.87(±0.04)
2	TPR	0.94(±0.05)	0.97(±0.02)	0.97(±0.02)	0.97(±0.02)
	TNR	0.69(±0.10)	0.74(±0.07)	0.80(±0.09)	0.79(±0.12)
3	TPR	0.59(±0.22)	0.83(±0.06)	0.90(±0.04)	0.94(±0.01)
	TNR	0.95(±0.04)	0.97(±0.02)	0.93(±0.05)	0.98(±0.01)
4	TPR	0.96(±0.01)	0.88(±0.03)	0.97(±0.02)	0.93(±0.03)
	TNR	0.60(±0.14)	0.73(±0.15)	0.82(±0.09)	0.94(±0.01)
5	TPR	0.88(±0.05)	0.90(±0.04)	0.96(±0.03)	0.92(±0.04)
	TNR	0.72(±0.15)	0.85(±0.09)	0.82(±0.13)	0.97(±0.02)
Avg	TPR	0.83(±0.15)	0.89(±0.06)	0.93(±0.04)	0.94(±0.02)
	TNR	0.76(±0.13)	0.85(±0.09)	0.87(±0.06)	0.91(±0.07)

selected are $s = 1000$ and PCA components kept are $p = 20$, while fine-tuning of classifiers' hyperparameters is considered unnecessary since it heavily depends on the chosen classifier.

In the following, Section III-C presents results from six classifiers, where each one is trained on one of the surfaces presented in Table II and tested on the remaining ones. In section III-D results show the performance of every classifier trained on two up to five out of the six surfaces and tested on the remaining, unknown ones.

B. Evaluation Metrics

Successful recognition of a slip sample is a True Positive (TP), while its misclassification as stable is a False Negative (FN). Accordingly, successfully classifying a stable sample is a True Negative (TN), while mistaking it as slip is a False Positive (FP). The corresponding rates are reported as TPR, FNR, TNR and FPR, where $TPR = \frac{TP}{TP+FN}$, $FNR = \frac{FN}{TP+FN}$, $TNR = \frac{TN}{TN+FP}$, $FPR = \frac{FP}{TN+FP}$. Thus, given the equally balanced classes, the accuracy is $acc = \frac{TPR+TNR}{2}$. Results in Tables III–V demonstrate the classification performance of each class (TPR, TNR).

C. Evaluation of Training with One Surface

Initially a classifier c_{ij} is trained on each surface $i \in 0, ..., 5$, for each feature set $j \in 0, ..., 3$ (AFFT, FREQ, TIME, BOTH) and the classification performance is evaluated, reporting analytic results for TPR and TNR as well. The performance of training on surface i with feature set j and testing on surface k is denoted by c_{ij}^k, where $k = i$ corresponds to evaluating performance on the same surface as in training, hereby called self-accuracy (Fig. 4), while $k \neq i$ relates to performance on previously unseen surfaces, hereby called cross-accuracy (Table III).

Fig. 4 shows the c_{ij}^i performance of each surface i on itself, via a 5-fold cross-validation, to realize how consistently each classifier performs on new samples of the already seen surface. On the x-axis, each group represents a surface i and each bar corresponds to a feature set j. Every bar is split in 4 sub-bars, representing from bottom (blue) to top (red) TPR, TNR, FNR and FPR respectively. Thus, the sum of TPR and TNR on the diagram (already scaled by 0.5) corresponds to the classification accuracy. Observing the bars of each group, it is clear that performance improves towards the rightmost columns, namely FREQ, TIME and BOTH have on average a 4%, 7% and 7.5% better accuracy than the most widely used AFFT features, while this behavior seems independent of training surface.

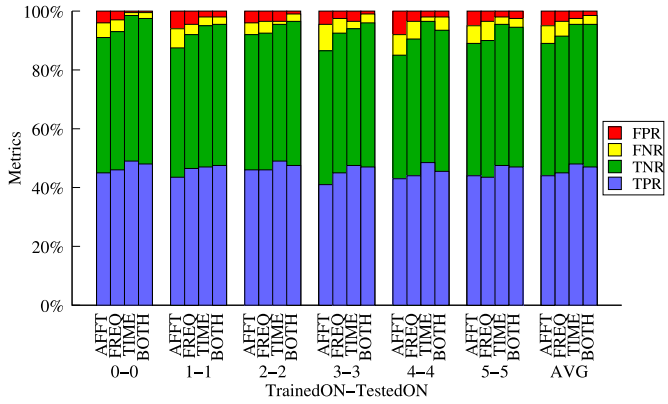

Fig. 4. Performance of each trained surface on itself 5-fold cross-validated.

Table III shows the c_{ij}^k performance of each surface i on all remaining surfaces $k \neq i$, to realize how generalizable each model is even after training on only one surface. Each line reports the average and the standard deviation of TPR and TNR. It is obvious that FREQ, TIME and BOTH features have on average 8%, 10.5% and 13% better accuracy than the commonly used AFFT features, with 7%, 9% and 9% smaller standard deviations respectively. Therefore, the basic observation from self-accuracy reports, that the FREQ, TIME and BOTH feature sets consistently perform better, holds for unseen surfaces as well. Considering this behavior, we choose BOTH feature set as the proposed approach for the rest of the letter, due to its improved performance compared to isolated FREQ and TIME sets, achieving on average TPR and TNR of 94% and 91% respectively, trained on 1 surface and tested on 5 unseen surfaces. Moreover the FNR is low (6%), namely any missed slipping occurrences, which may result to delayed or unsuccessful detection and subsequently reaction, are substantially reduced.

D. Evaluation of Training with More Surfaces

The promising results displayed by the proposed set suggest that the representation on both time and frequency feature spaces may potentially lead to a model with powerful generalization ability. Therefore, the procedure for cross-accuracy is repeated with 2 and more surfaces, until all but one are used for training, to assess how the surface variety impacts performance and which feature set is consistently favorable for slippage detection. The window shifting is adjusted appropriately, so as to keep the training size and class distribution the same, regardless of the number of surfaces. Table IV shows the corresponding results, where the same behavior can be noticed as far as feature sets are concerned. A significant observation is that the proposed features do not need to cover more than one surface during training, since the observed average improvement of the already high accuracy, as the number of training surfaces increases, is only around 3%. In contrast, in the -usually adopted by the literature-

TABLE IV
MEAN PERFORMANCE AND STANDARD DEVIATION OF SLIPPAGE DETECTION AFTER TRAINING WITH A TOTAL OF 1–5 SURFACES AND TESTED ON THE REMAINING SURFACES

Total		AFFT	FREQ	TIME	BOTH
1	TPR	0.83(±0.15)	0.89(±0.06)	0.93(±0.04)	0.94(±0.02)
	TNR	0.76(±0.13)	0.85(±0.09)	0.87(±0.06)	0.91(±0.07)
2	TPR	0.84(±0.13)	0.90(±0.04)	0.95(±0.02)	0.95(±0.02)
	TNR	0.84(±0.08)	0.91(±0.07)	0.91(±0.07)	0.95(±0.04)
3	TPR	0.85(±0.11)	0.91(±0.03)	**0.96(±0.02)**	**0.96(±0.01)**
	TNR	0.87(±0.06)	0.93(±0.05)	**0.94(±0.05)**	**0.97(±0.03)**
4	TPR	0.84(±0.13)	**0.91(±0.03)**	0.95(±0.01)	0.96(±0.01)
	TNR	0.87(±0.07)	**0.94(±0.04)**	0.93(±0.05)	0.96(±0.03)
5	TPR	**0.91(±0.10)**	0.91(±0.04)	0.96(±0.02)	0.96(±0.01)
	TNR	**0.82(±0.11)**	0.94(±0.05)	0.89(±0.07)	0.96(±0.04)

AFFT features, accuracy improves considerably (around 8%) as more surfaces are introduced during training, without even surpassing the worst performance (FREQ, 1 surface, $acc = 88.5\%$) of the other feature sets.

This result indicates the superior generalization ability of the proposed features as compared to the existing sets, since the phenomenon is successfully recognized even with one surface. It is believed that this is due to the addition of time-based features that can capture patterns related to the latent, temporal nature of the phenomenon, thus acting complementary to the frequency-based ones. Frequential patterns seemingly require several surfaces to manifest, whereas time-based ones may be present and captured even from a single surface. The trade-off between increased complexity of additional training surfaces and performance improvement is something that should also be taken into consideration. The generalization ability and lack of complexity demonstrated in this section leads to the adoption of training the *MLP* with one surface ($SN = 0$) and the proposed features and this model is used throughout the rest of the letter.

IV. REAL-TIME SLIPPAGE DETECTION

A real-time experiment is performed to evaluate the proposed methodology on a completely different setup and unseen object, compared to the ones used for training. The *SGS* is holding an empty plastic cup with an extra weight attached to the cup's base with a string (Fig. 3(c)). The string is loose at the starting position, so the extra weight is not exerted until the upward hand motion causes the extra weight to also be lifted (P1). This dynamic load causes the cup to slip without falling (P2), since the cup is conical and acts as a wedge. Subsequently, the downward motion causes the hanging weight to touch the table and a stable grasp is reestablished (P3). Finally, the cup is repositioned manually (P4) and the experiment is repeated. Fig. 5 shows the norm of measured force (blue line) $|f|$ and the output of the *MLP* classifier trained on 1 surface (out1) with the proposed feature set (red dots), along with a time axis partitioning of the different phases (P1, P2, P3, P4 - vertical dashed lines). It can be observed that the transition between different phases is correctly and timely detected; e.g., transition from P1 to P2, corresponding to the exact moment of load exertion, is instantaneously perceived. Moreover, as shown in Fig. 5, the classifier

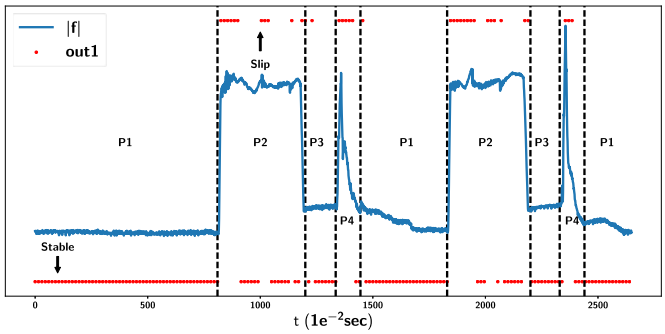

Fig. 5. Predicted output of classifier trained on 1 surface.

Fig. 6. New Experimental setup with ATI sensor.

successfully recognizes the P1 and P3 stable states (weight lying on table) and the P4 slip state (manual repositioning). During P2 the pendulum-weight is swaying, which may cause alternations between slipping and stable phases, a behavior manifested in the output of the classifier. Note that, particularly during P2, there is no clear way to acquire the ground truth label. Notice also that the real-time latency of the implemented methodology was measured around 60 ms at worst, so a real-time conservative window shifting of $m = 100$ is chosen, corresponding to 100 ms between consecutive classifier's answers. The aforementioned observations indicate that a fast and accurate detection is achieved by the proposed method when grasping an unknown object, anticipating grasp stability even with a simple reaction strategy.

V. VALIDATION WITH A NEW EXPERIMENTAL SETUP

To further support our claim of generalization and transferability, the selected classifier, trained on 1 surface with the Optoforce dataset, is tested with a new experimental setup. In particular, the KUKA LWR4+ robotic manipulator, with an ATI Mini40 F/T sensor mounted on its end-effector (Fig. 6(a)), is commanded to move along the contacted surface, performing 2 translational motions (downwards and upwards) and 2 rotational motions (rightwise and leftwise), with equally timed moving (slip) and stationary (stable) phases, with sliding velocities of 2 cm/sec and 0.1 rad/sec. During motion, a desired normal force of 1 N and 1.5 N is maintained via a hybrid force/impedance control. A sampling rate of 1 KHz is used, as well as a low-pass filter of varying cutoff frequencies (specifically 836 Hz, 326 Hz, 152 Hz and no filtering), in order to simulate the effect of lowering the sampling rate. The experiment is repeated for six different surfaces, as shown in Fig. 6(b). Respective results can be seen in Table V, where the average over all testing surfaces is shown, for all tested sampling frequencies (SF) and desired

TABLE V
PERFORMANCE OF THE ATI DATASET EVALUATED WITH THE TRAINED MODEL ON 1 SURFACE OF THE OPTOFORCE DATASET

			TRANSLATION		ROTATION	
SF	DNF		AFFT	BOTH	AFFT	BOTH
No Filter	1.0N	TPR	0.68	0.78	0.30	0.72
		TNR	0.89	0.88	0.98	0.95
	1.5N	TPR	0.71	0.83	0.27	0.71
		TNR	0.87	0.88	0.92	0.86
836 Hz	1.0N	TPR	0.60	0.84	0.40	0.65
		TNR	0.71	0.89	0.74	0.92
	1.5N	TPR	0.79	0.84	0.39	0.64
		TNR	0.86	0.87	0.98	0.97
326 Hz	1.0N	TPR	0.53	0.84	0.23	0.75
		TNR	0.91	0.89	0.91	0.87
	1.5N	TPR	0.77	0.80	0.33	0.77
		TNR	0.99	0.90	0.98	0.86
152 Hz	1.0N	TPR	0.75	0.82	0.43	0.78
		TNR	0.99	0.90	0.98	0.82
	1.5N	TPR	0.53	0.84	0.23	0.74
		TNR	0.91	0.89	0.91	0.87
Average		TPR	0.68	**0.82**	0.32	**0.72**
		TNR	0.89	**0.89**	0.93	**0.89**

normal forces (DNF), between the features that are widely used (AFFT) and the proposed ones (BOTH).

Results for the translational slip case (TPR) show that the proposed features on average outperform the AFFT ones by 14%, reaching 82%. This is more indicative in the rotational slip case (TPR) where the proposed features on average outperform the AFFT ones by 40%, reaching 72%. Performance in both slip cases denote that the proposed features recognize slippage more efficiently. Finally, the stable case (TNR) is recognized equally well by both feature sets at around 90%. Moreover, the proposed approach appears not to be affected from the different sampling frequencies as opposed to AFFT and consistently performs better. Finally, notice that the signal to noise ratio of ATI sensor is lower than the Optoforce sensor, indicating that the classifier performs well under substantial noise levels.

VI. CONCLUSION

In this work a novel feature vector has been proposed for slippage detection, combining time and frequency domain content of measured force magnitude. The proposed scheme assumes the availability of the contact force magnitude, which implies the possibility to acquire 3D force measurements. It has been shown that time domain features, as well as their combination with frequency ones, are consistent in their performance and generalization ability, even trained with only one surface, to a different setup and sensor for both translational and rotational slippage. In contrast, FFT, the most usual current practice, requires more surfaces to generalize and even then falls behind the proposed features' performance. It is within our future plans to further demonstrate the proposed strategy's performance via grasp stability achievement with unknown objects and disturbances in combination with a simple reaction scheme. (Code and video available online at https://goo.gl/qXydbh and https://goo.gl/jCDMkG respectively)

ACKNOWLEDGMENT

Special thanks to Martin Meier from Bielefeld University and to Yannis Koveos from University of Patras for lending a helping hand, as well as to George Velentzas from National Technical University of Athens for his support and feedback.

REFERENCES

[1] G. Westling and R. S. Johansson, "Responses in glabrous skin mechanoreceptors during precision grip in humans," *Experimental Brain Res.*, vol. 66, no. 1, pp. 128–40, 1987.
[2] Z. Kappassov, J. A. Corrales, and V. Perdereau, "Tactile sensing in dexterous robot hands—Review," *Robot. Auton. Syst.*, vol. 74, pp. 195–220, 2015.
[3] X. Song, H. Liu, K. Althoefer, T. Nanayakkara, and L. D. Seneviratne, "Efficient break-away friction ratio and slip prediction based on haptic surface exploration," *IEEE Trans. Robot.*, vol. 30, no. 1, pp. 203–219, Feb. 2014.
[4] A. Ajoudani *et al.*, "Reflex control of the pisa / IIT softhand during object slippage," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2016, pp. 1972–1979.
[5] N. Wettels, A. R. Parnandi, J.-H. Moon, G. E. Loeb, and G. S. Sukhatme, "Short papers grip control using biomimetic tactile sensing systems," *Hand*, vol. 14, no. 6, pp. 1–5, 2009.
[6] F. E. Vina, Y. Karayiannidis, C. Smith, and D. Kragic, "Adaptive control for pivoting with visual and tactile feedback," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2016, pp. 399–406.
[7] M. Stachowsky, T. Hummel, M. Moussa, and H. A. Abdullah, "A slip detection and correction strategy for precision robot grasping," *IEEE/ASME Trans. Mechatronics*, vol. 21, no. 5, pp. 2214–2226, Oct. 2016.
[8] J. M. Romano, K. Hsiao, G. Niemeyer, S. Chitta, and K. J. Kuchenbecker, "Human-Inspired robotic grasp control with tactile sensing," *IEEE Trans. Robot.*, vol. 27, no. 6, pp. 1067–1079, Dec. 2011.
[9] Z. Su *et al.*, "Force estimation and slip detection/classification for grip control using a biomimetic tactile sensor," in *Proc. IEEE-RAS 15th Int. Conf. Humanoid Robots*, Nov. 2015, pp. 297–303.
[10] F. Cordella *et al.*, "A force-and-slippage control strategy for a poliarticulated prosthetic hand," in *Proc. IEEE Int. Conf. Robot. Autom.*, Jun. 2016, pp. 3524–3529.
[11] M. Kaboli, K. Yao, and G. Cheng, "Tactile-based manipulation of deformable objects with dynamic center of mass," in *Proc. IEEE-RAS Int. Conf. Humanoid Robots*, 2016, pp. 752–757.
[12] Y. Bekiroglu, J. Laaksonen, J. A. Jorgensen, V. Kyrki, and D. Kragic, "Assessing grasp stability based on learning and haptic data," *IEEE Trans. Robot.*, vol. 27, no. 3, pp. 616–629, Jun. 2011.
[13] F. Veiga, H. Van Hoof, J. Peters, and T. Hermans, "Stabilizing novel objects by learning to predict tactile slip," in *Proc. IEEE Int. Conf. Intell. Robots Syst.*, Dec. 2015, pp. 5065–5072.
[14] D. Goger, N. Gorges, and H. Worn, "Tactile sensing for an anthropomorphic robotic hand: Hardware and signal processing," in *Proc. 2009 IEEE Int. Conf. Robot. Autom.*, May 2009, pp. 895–901.
[15] M. Schöpfer, C. Schürmann, M. Pardowitz, and H. Ritter, "Using a piezo-resistive tactile sensor for detection of incipient slippage," in *Proc. Robot. ISR 2010 41st Int. Symp. 2010 6th German Conf. Robot.*, 2010, pp. 14–20.
[16] M. Meier, G. Walck, R. Haschke, and H. J. Ritter, "Distinguishing sliding from slipping during object pushing," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2016, pp. 5579–5584.
[17] J.-P. Roberge, S. Rispal, T. Wong, and V. Duchaine, "Unsupervised feature learning for classifying dynamic tactile events using sparse coding," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2016, pp. 2675–2681.
[18] A. Phinyomark, C. Limsakul, and P. Phukpattaranont, "A novel feature extraction for robust EMG pattern recognition," *J. Comput.*, vol. 1, no. 1, pp. 71–80, 2009.
[19] S. Golz, C. Osendorfer, and S. Haddadin, "Using tactile sensation for learning contact knowledge: Discriminate collision from physical interaction," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2015, pp. 3788–3794.
[20] "Mutual Variable Information." [Online]. Available: https://goo.gl/nuomRH
[21] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.
[22] M. Quigley *et al.*, "ROS: An open-source Robot Operating System," [Online]. Available: https://goo.gl/x4tXjb