

Tactile Convolutional Networks for Online Slip and Rotation Detection

Martin Meier^(✉), Florian Patzelt, Robert Haschke, and Helge J. Ritter

Neuroinformatics Group,
Center of Excellence Cognitive Interaction Technology (CITEC),
Bielefeld University, Bielefeld, Germany
{mmeier,fpatzelt,rhaschke,helge}@techfak.uni-bielefeld.de

Abstract. We present a deep convolutional neural network which is capable to distinguish between different contact states in robotic manipulation tasks. By integrating spatial and temporal tactile sensor data from a piezo-resistive sensor array through deep learning techniques, the network is not only able to classify the contact state into stable versus slipping, but also to distinguish between rotational and translation slippage. We evaluated different network layouts and reached a final classification rate of more than 97 %. Using consumer class GPUs, slippage and rotation events can be detected within 10 ms, which is still feasible for adaptive grasp control.

1 Introduction

In autonomous robotic manipulation tasks, for example grasping and placing objects, estimating the stability of the object in hand plays a major role. Objects may slip out of the manipulator. This can lead to a state in the desired action sequence from which the system cannot recover easily. Due to occlusions, vision-based systems can hardly keep track of the state of objects hold in manipulators and are therefore of limited usefulness when it comes to detecting loss of grasp stability. For that reason, the loss of an object can only be detected after such events already occurred. Humans perceive the onset of slippage by sensing high-frequency micro-vibrations through specialized nerves (Pacinian corpuscle) in the skin [4].

One possibility for early detection of slippage events in robotic systems is the integration of tactile sensing capabilities directly into robotic manipulators. By having human like sensing skills, the system should be able to directly evaluate the contact state during interactions. Compared to imaging technologies where standards are established for data acquisition and representation, current tactile sensors posses a large variety of data acquisition techniques, which can be either based on electric [12], optic [15] or acoustic [6] effects. For example the authors in [2] discuss eight different technologies which are based on these three effects and are used in current state of the art tactile sensors. For a detailed technical overview the interested reader is referred to [2].

The work presented in [13] used support vector machines and random forests to detect object slippage with a BioTac [6] sensor. The BioTac sensor offers multiple modalities such as 19 electrodes to measure local contacts with a sampling rate of 100 Hz, thermal sensors and two pressure transducers, one for low (up to 100 Hz) and one for high (up to 2.2 kHz) frequencies, respectively. The features comprised all raw sensor values, where the high frequency component is supplied as a time series of the last 22 sensor readings which makes up for half of the feature vector. With these features used as input for a random forest, a $F_{score} > 0.75$ has been achieved in the evaluation. To predict slippage of held objects, the authors of [14] took an approach where they first learned friction properties based on data acquired from a force/torque sensor with Gaussian process regression. In [11], also a BioTac sensor is used to classify slip with a multilayer perceptron (MLP), but in contrast to [13], the authors used a sequence of 100 samples of the electrodes without utilizing the high frequency sensor. With this time series as input for a MLP, a classification rate of 80 % was achieved. The same type of tactile sensor utilized in this work was already used in [8] for a binary stable- vs. slip-classification. Here, the authors used a Fourier transformation over the whole sensor array with varying window sizes to predict slip velocity. They were able to achieve low mean squared errors of 0.04. These approaches have in common, that they rely on the classification of time series to detect slip events.

In areas outside of the scope of tactile sensing, convolutional neural networks (CNNs) have been successfully applied to time series classification tasks, for example in speech recognition. In [7], the authors evaluated the performance of convolutional networks compared to deep neural networks (DNNs), Gaussian mixture model (GMM) and Hidden markov model (HMM) approaches for large speech recognition tasks. The data was preprocessed by extracting mel-frequency cepstrum coefficients (MFCC) [3], a filter technique that resemble human auditory perception by using a logarithmic scale for pitch and loudness of the signal. With these frequency features as input for CNNs, the deep networks outperformed GMM and HMM approaches on different datasets. The authors in [1] evaluated the efficiency of a deep neural networks with and without convolutional layers in a similar speech recognition task and reported an increase of 6 to 10 % in the relative classification rate for CNNs compared to DNNs. By using CNNs in conjunction with short time Fourier transforms of brain waves recorded with an EEG, the authors in [10] could distinguish different types of musical rhythms perceived by their subjects.

The approach to employ time series data in slip detection tasks and the performance of convolutional architectures suggests, that CNNs are an appropriate choice to achieve a more fine grained classification of slippage events, in our case to not only distinguish between stick and slip condition, but also to approach the task of dividing the slip events further into translational and rotational events. In the following section, we will first outline the sensing technology used in our approach. Afterwards the employed convolutional architectures will be described, evaluated and discussed.



Fig. 1. Objects used for the evaluation and experimental setup for data recording. Two KUKA LWR robots with attached tactile sensors (light orange) holding a glass. The fingertip shaped sensor touching the glass from above is used to detect the onset of slippage for data labeling purposes. (Color figure online)

2 Sensor Properties and Data Acquisition

We recorded data by holding three different objects, a cardboard cylinder, a remote and a drinking glass, between two piezo-resistive tactile sensor arrays¹ [9], where each sensor array was attached to a 7 degree of freedom KUKA LWR robotic arm. An image showing the objects used for training and evaluation and the robot arms holding a drinking glass is shown in Fig. 1. The Myrmex sensor consists of a printed circuit board (PCB) with 16×16 taxels, each with a spatial dimension of 5×5 mm. Each taxel measures the change of resistance between two electrodes that is induced by a piezo-resistive foam covering the PCB layer. The change in resistance is digitized via a 12 bit analog-digital converter. The data of all taxels is sampled at a rate of up to 1.9 kHz and transmitted to the host PC via standard USB video protocol. An example of a single frame of the sensor data while holding a cylindrical cardboard box and the change over time of a single cell is shown in Fig. 2.

2.1 Data Recording

With three different objects, a total of 64 trials have been recorded for the three classification classes, namely a stable state, translational and rotational slip. We used two Myrmex sensors to hold the objects, each attached to the robot arm’s end-effector as a “large” fingertip. The sensors were sampled with a rate

¹ Called *Myrmex* hereafter.

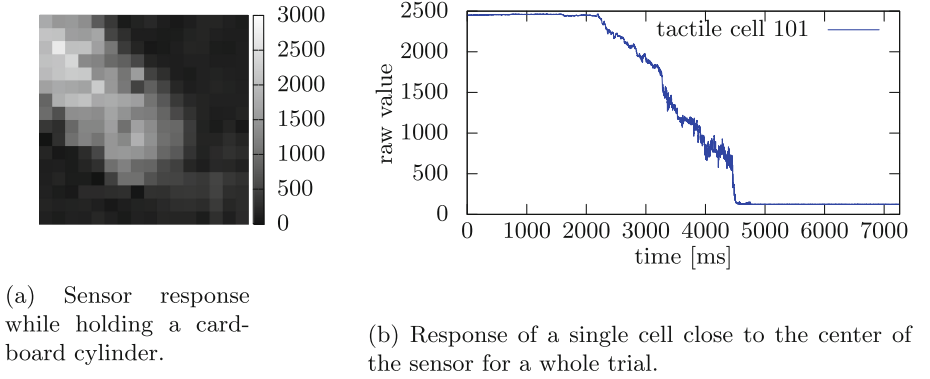


Fig. 2. An image representation of the raw sensor data for a single frame is shown on the left. The right panel shows the raw value of a single cell over a whole trial.

of 1 kHz. The overall duration of these trials was 662.8s, leading to a total of ≈ 1.3 M recorded sensor frames. To generate the slip events, we placed the objects between the sensors and let the robots exert varying forces (between 1 and 20 N) onto the objects, then moving the robotic arms slowly apart from each other. By manually placing the object during grasping we could induce either translational or rotational slip events: Translational slip events were generated by placing the center of mass directly above the center of contact. For the rotational slip events, the center of mass was placed horizontally shifted with respect to the center of contact.

2.2 Data Labeling

Acquiring ground-truth labels for the onset of slippage is a demanding task. For example, the authors in [13] hand labeled data based on video recordings of the trials while in [11] an inertial measurement unit was attached to the sliding object to provide a reference of the onset of slip events. The results from [11] actually suggest, that incipient slippage can be detected even before such traditional sensors as IMUs detect a motion of the object.

In our experiments we automated the labeling task of the data by placing a third tactile sensor, using the same piezo-resistive principle, in contact with the object, touching it from above. For technical reasons, this sensor could only be sampled with a rate of 500 Hz, but the signals were synchronized with the grasping Myrmex sensors. The onset of slippage was detected by evaluating the contact forces measured with the third sensor. We set the onset of slippage to the time when the sum of contacts on the third sensor started to decrease. The end of the trial was determined by the point in time when no more contacts were detected on the sensors holding the object. The sequence was labeled as rotational or translational slip, respectively, depending on the initial manual placement of the object.

3 Convolutional Tactile Networks

The properties of our sensor, the spatial arrangement of tactile cells combined with a high sampling frequency, suggest to use an approach similar to other time series classification techniques. By calculating a short-time Fourier transformation over a certain window size for each tactile sensor cell, we obtain a spatially arranged stack of Fourier coefficients which resembles the structure of RGB color images, but with an increased amount of channels – one per Fourier coefficient. On each of the channels we apply convolution and pooling layers to learn filters for each of the frequency bins. The output of these filters is fed into a fully connected layer, which is finally connected to a softmax layer for the classification. A convolution filter of width w and height h calculates the activation a at position i, j by multiplying the input activations $x_{i+k, j+l}$ from a previous layer with weights $W_{k, l}$ and is defined by Eq. 1 as

$$a_{i, j} = \sigma \left(\sum_{k=0}^{w-1} \sum_{l=0}^{h-1} W_{k, l} x_{i+k, j+l} \right) \quad (1)$$

where $\sigma()$ is a activation function, for example $\tanh()$. A max pooling layer simply applies a $\max(0, x)$ function to a given input area of size $w \times h$.

The spatial arrangement of the frequency bins has an additional benefit for the classification task. For example in cases of translational slip, all active tactile sensor cells should have a similar amplitude whereas in cases of rotational slip, the amplitudes should differ because of increasing accelerations with respect to the distance of the center of rotation. After initial tests with different filter sizes in the convolution and pooling layers, we decided to investigate the three architectures described in Table 1 in detail since larger filter sizes turned out to decreased the classification performance slightly.

Table 1. Network architectures used in the evaluation. Here *conv* 3×3 is a convolution layer with a kernel size of 3×3 . *pool* 2×2 is a max pooling layer and *fc* 512 is a fully connected layer with 512 neurons.

#	Network architecture
1	conv $3 \times 3 \rightarrow$ pool $2 \times 2 \rightarrow$ fc 512
2	conv $3 \times 3 \rightarrow$ pool $2 \times 2 \rightarrow$ conv $3 \times 3 \rightarrow$ pool $2 \times 2 \rightarrow$ fc 512
3	conv $3 \times 3 \rightarrow$ pool $2 \times 2 \rightarrow$ conv $3 \times 3 \rightarrow$ pool $2 \times 2 \rightarrow$ fc 1024

4 Evaluation

To evaluate the proposed network architectures, we preprocessed the raw data by computing short time Fourier transformations for each of the tactile cells. We chose a window size of 64ms for the STFTs, with a small shift of 8ms.

That is, receiving tactile data at 1 kHz, the net generates classification results at a rate of 125 Hz. Additionally, the raw images were cropped to include only the innermost 12×12 tactile cells of the sensor. This was necessary due to false-positives occurring at the borders, caused by the mechanical mounting of the foam. The raw data we recorded has another drawback with respect to practical applications. The sensor orientation was fixed throughout the recordings and gravity was the only acting force to create slippage events. Thus the slippage and rotation only occurred in one direction. We therefore augmented the dataset by rotating the raw data with 12 different angles, reaching from zero to 330° in steps of 30° , before calculating the short time Fourier transformation, which improves the generalization to other end-effector poses. Because stable states are overrepresented in the dataset, we sub-sampled the raw data to obtain an equal number of raw samples for the three classes. After the rotation and sub sampling process, we have a total of ≈ 2.1 M data samples of dimension $(12 \times 12 \times 32)$ containing Fourier amplitudes. Fourier phases were not considered.

Before training, we split the dataset and kept 20 % of the available data samples as a test set for evaluating the proposed networks architectures. The data samples in the dataset were stored in an alternating fashion with respect to the labels to assure an even distribution of the three classes in the training and test set. We tested two conditions for the networks described in Table 1, one considering all frequency components and one applying a 60 Hz high pass filter, to explicitly remove low frequency vibrations from the robot arms before training. Already the smallest network with only one convolution and pooling layer achieves an accuracy of more than 91 %. Here the high pass filter increases the accuracy by 1.6 %. Adding a second Convolution and pooling block increases the classification accuracy further to nearly 98 %, when a high pass filter is included. For the case with the high pass filtered input data, we carried out an additional ten-fold cross-validation to confirm the results more thoroughly. Therefore, we split the dataset in ten chunks of equal size, created a training set from nine of the ten chunks and used the remaining chunk for testing. This was done with each of the ten chunks as test data. Table 3 shows a confusion matrix of the test accuracy for each network. The cells contain the average percentage over the ten runs and confirm the previous results from Table 2.

An example of the training behavior of network 3 with respect to test accuracy and loss is shown in Fig. 3. The network converges towards the final test accuracy after around 700000 iterations, where an iteration in this case is the batch processing of 64 samples of Fourier transformed data.

Table 2. Test accuracy for the networks from Table 1 with and without high pass filter. The last column shows the average time for a single forward pass.

#	Accuracy w/o filter	Acc. with high pass	Time fwd pass
1	91.01 %	92.65 %	0.29 ms
2	96.12 %	96.5 %	0.44 ms
3	97.45 %	97.89 %	0.43 ms

Table 3. Confusion matrices for the cross-validation of all networks with high pass filtered data. The letters s, t and r indicate the classes for stable, translational and rotational slip, respectively.

		prediction		
		s	t	r
input	s	90.79%	5.83%	3.38%
	t	2.13%	92.58%	5.29%
	r	2.15%	3.17%	94.68%

(a) Network 1.

		prediction		
		s	t	r
s	s	95.73%	2.54%	1.73%
	t	1.26%	96.37%	2.37%
	r	0.97%	1.56%	97.47%

(b) Network 2.

		prediction		
		s	t	r
s	s	97.57%	1.41%	1.02%
	t	0.68%	97.73%	1.58%
	r	0.51%	0.93%	98.56%

(c) Network 3.

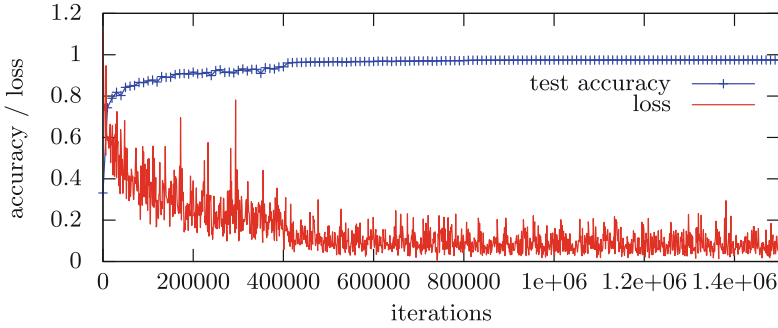


Fig. 3. Test accuracy and loss during training of network 3 from Table 3. One iteration in this figure is the batch processing of 64 samples.

5 Discussion

We presented an approach to detect translational and rotational slippage events in robot manipulation tasks. To our knowledge, using neural networks to discriminate between rotational and translational slip in addition to stable states has not been done before, since recent state of the art techniques only used a binary slip/non slip detection. We achieved state of the art classification results of more than 97 % by utilizing a convolutional neural network approach in conjunction with short time series of the sensor data. Using a consumer grade GPU for parallelization, the classification and preprocessing is fast enough to be integrated in real world robot controllers, for example for online grasp force adaptation. An interesting next step will be to transfer the work presented in this paper to the fingertip sensor [5], shown in Fig. 1, which we used for automatic labeling.

Acknowledgments. The research leading to these results has received funding from the European Community’s Framework Programme Horizon 2020 – under grant agreement No 644938 – SARAFun and was supported by the Cluster of Excellence Cognitive Interaction Technology ‘CITEC’ (EXC 277) at Bielefeld University, which is funded by the German Research Foundation (DFG).

References

1. Abdel-Hamid, O., Mohamed, A.-R., Jiang, H., Deng, L., Penn, G., Yu, D.: Convolutional neural networks for speech recognition. *IEEE/ACM Trans. Audio Speech Lang. Process.* **22**(10), 1533–1545 (2014)
2. Dahiya, R.S., Valle, M.: Tactile sensing technologies. *Robotic Tactile Sensing*, pp. 79–136. Springer, Netherlands (2013)
3. Davis, S.B., Mermelstein, P.: Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. *IEEE Trans. Acoust. Speech Signal Process.* **28**(4), 357–366 (1980)
4. Johansson, R., Westling, G.: Signals in tactile afferents from the fingers eliciting adaptive motor responses during precision grip. *Exp. Brain Res.* **66**(1), 141–154 (1987)
5. Koiva, R., Zenker, M., Schurmann, C., Haschke, R., Ritter, H.J.: A highly sensitive 3D-shaped tactile sensor. In: 2013 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM), pp. 1084–1089. IEEE (2013)
6. Lin, C.H., Erickson, T.W., Fishel, J.A., Wettels, N., Loeb, G.E.: Signal processing and fabrication of a biomimetic tactile sensor array with thermal, force and microvibration modalities. In: ROBOT, pp. 129–134 (2009)
7. Sainath, T.N., Mohamed, A.-R., Kingsbury, B., Ramabhadran, B.: Deep convolutional neural networks for LVCSR. In: 2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 8614–8618. IEEE (2013)
8. Schöpfer, M., Schürmann, C., Pardowitz, M., Ritter, H.: Using a piezo-resistive tactile sensor for detection of incipient slippage. In: 2010 41st International Symposium on Robotics (ISR) and 2010 6th German Conference on Robotics (ROBOTIK), pp. 1–7. VDE (2010)
9. Schürmann, C., Haschke, R., Ritter, H.: Modular high speed tactile sensor system with video interface. In: Tactile Sensing in Humanoids – Tactile Sensors and Beyond@ IEEE-RAS Conference on Humanoid Robots (Humanoids) (2009)
10. Stober, S., Cameron, D.J., Grahn, J.A.: Using convolutional neural networks to recognize rhythm stimuli from electroencephalography recordings. In: Advances in Neural Information Processing Systems, pp. 1449–1457 (2014)
11. Su, Z., Hausman, K., Chebotar, Y., Molchanov, A., Loeb, G.E., Sukhatme, G.S., Schaal, S.: Force estimation and slip detection/classification for grip control using a biomimetic tactile sensor. In: 2015 IEEE-RAS 15th International Conference on Humanoid Robots (Humanoids), pp. 297–303. IEEE (2015)
12. Teshigawara, S., Tsutsumi, T., Shimizu, S., Suzuki, Y., Ming, A., Ishikawa, M., Shimojo, M.: Highly sensitive sensor for detection of initial slip and its application in a multi-fingered robot hand. In: 2011 IEEE International Conference on Robotics and Automation (ICRA), pp. 1097–1102. IEEE (2011)
13. Veiga, F., van Hoof, H., Peters, J., Hermans, T.: Stabilizing novel objects by learning to predict tactile slip. In: 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 5065–5072. IEEE (2015)
14. Vina, B., Francisco, E., Bekiroglu, Y., Smith, C., Karayiannidis, Y., Kragic, D.: Predicting slippage and learning manipulation affordances through gaussian process regression. In: 2013 13th IEEE-RAS International Conference on Humanoid Robots (Humanoids), pp. 462–468. IEEE (2013)
15. Yuan, W., Li, R., Srinivasan, M.A., Adelson, E.H.: Measurement of shear and slip with a GelSight tactile sensor. In: 2015 IEEE International Conference on Robotics and Automation (ICRA), pp. 304–311. IEEE (2015)

Artificial Neural Networks and Machine Learning – ICANN
2016

25th International Conference on Artificial Neural
Networks, Barcelona, Spain, September 6-9, 2016,
Proceedings, Part II

Villa, A.E.P.; Masulli, P.; Pons Rivero, A.J. (Eds.)

2016, XXIX, 557 p. 173 illus., Softcover

ISBN: 978-3-319-44780-3