Contact Localization on Grasped Objects using Tactile Sensing

Artem Molchanov, Oliver Kroemer, Zhe Su and Gaurav S. Sukhatme

Abstract— **Manipulation tasks often require robots to make contact between a grasped tool and another object in the robot's environment. The ability to detect and estimate the positions and directions of these contact points is crucial for monitoring the progress of the task, and detecting failures. In this paper, we present a data-driven approach for detecting and localizing contacts between a grasped object and the environment using tactile sensing. We explore framing the contact localization as both a regression and a classification problem and train neural networks accordingly to estimate the contact parameters. We also compare the neural networks with Gaussian process regression and support vector machine classification with spatio-temporal hierarchical matching pursuit feature learning. We evaluate the presented approach using hundreds of contact events on eighteen objects with different shapes, sizes and material properties. The experiments show that the neural network approach can learn to localize contact events for individual objects with a mean absolute error of less than 2.5 cm for the positions and less than 10° for the directions.**

I. Introduction

Many manipulations in unstructured environments require a robot to use a grasped object, i.e., a tool, to interact with other objects. Often, a specific part of the tool, such as the tool's tip, needs to make contact with another object to perform the manipulation. For example, a hammer should make contact with a nail on the flat surface of its head when performing a hammering task. By sensing if and where the tool has made contact, the robot can verify that it is performing the skill correctly and otherwise adapt the skill accordingly.

Detecting transient contacts between a handheld object and other objects is particularly important when considering obstacles in the environment. For example, when placing a box on a cluttered table top, the robot may detect an unexpected contact on the side of the box rather than on the bottom. In this case, the robot should immediately stop executing the skill to prevent damaging the object and either place it at a different location or move the obstacles.

The task of localizing contacts has usually been approached using either wrist-mounted force-torque sensors or joint torque sensors. However, approaches based on these sensors face challenges such as bias drift and, in some cases, they require an accurate model of the robot and the grasped object. Tactile sensors provide another sensor modality that can be used to estimate contact parameters. Although current tactile sensors are not at human-performance levels [1], recent developments in the field have provided robots with human-like tactile sensor modalities [2], e.g., pressure, vibrations and temperature. The large amount of data provided by these sensors could potentially result in significantly more robust manipulation skills for robots. However, in order to fullfil this potential, the robots will also need suitable estimation methods to process the tactile data.

In this work, we explore using machine learning methods to estimate contact parameters between grasped objects and the environment based on data from biomimetic tactile sensors (BioTacs) [2]. In particular, we investigate estimating the positions and directions of contact points using neural network (NN) classification and regression, Gaussian process (GP) regression, and support vector machine (SVM) classification with features learned using spatio-temporal hierarchical matching pursuit (ST-HMP) [3]. We evaluate the methods using data collected from 18 objects with different shapes, sizes and materials. The experiments were designed to provide accurate ground truth information of the contact events and the contact parameters. In our work, we rely on a few assumptions. First, we restrict our contacts to the transient type, i.e., short taps, which correspond to tactile events such as bumping into objects or making contact. We also restrict our investigation to a single contact point and perform object dependent learning.

The key contributions of this work are: a) a model-free approach for estimating contact positions and directions between the environment and a grasped object based on tactile signals, b) an accurate labelled dataset[1] for evaluating and benchmarking contact localization methods, and c) an evaluation of the presented approaches using real robot experiments.

II. Related Work

The problem of contact detection and localization has received significant attention in the literature. The majority of the related approaches use force-torque sensors and analytical models to estimate the contact location. These methods often rely on accurate and precise calibrations of the force-torque sensors. For example, Karayiannidis et al. [4] estimate a point of contact using first order differential kinematics in combination with force-torque measurements. This approach is only applicable to rigid grasps where the object cannot slip in the hand, as it assumes knowledge of the static center of mass of the object. Likar et al. [5] present an approach that estimates the force and the point of contact from joint torque information. Their method requires exact knowledge of the robot's dynamics, which are often difficult

Artem Molchanov, Oliver Kroemer and Gaurav S. Sukhatme are with the Department of Computer Science; Zhe Su is with the Department of Biomedical Engineering, University of Southern California, Los Angeles. {molchano, okroemer, zhesu, gaurav}@usc.edu.

[1]The dataset is available at `http://bicl.robotics.usc.edu`

to obtain for real robots. Some works do not estimate contact locations explicitly but rather use force-torque measurements to estimate other task-relevant states, such as the alignment errors in the peg-in-hole problem [6] or the transitions between contact states for assembly tasks [7].

Other methods exploit geometric models of the robot to estimate the points of contact of the manipulator with the environment. For example, Petrovskaya et al. [8] use compliant motions in order to simultaneously estimate geometric parameters of the robot's links and the points of contact. The link parameters are estimated using a least squares approach while the contact points are inferred using a Bayesian approach. Koonjul et al. [9] propose two geometry-based approaches. The first approach extends the self posture changeability method [10] to use multiple compliant joints, while the second approach is a model-free method that maps joint displacements directly to the point of contact.

Tactile sensors have also been used for contact estimation, but most of the work has focused on estimating contact locations and interaction forces on robotic digits [11], [12]. Tactile sensors have also been used to estimate other object properties, e.g., the object's pose or material properties. For example, Corcoran and Platt [13] use particle filters to estimate the pose of an object based on its contacts with a robot hand. The pose of the object during manipulation is estimated by a measurement model which integrates the likelihood of contact measurements over the space of all possible contact positions on the surface of the object. Li et al. [14] use vision-based tactile sensor, GelSight, to localize objects in a robot hand by matching key points between object height maps with RANSAC. Su et al. [15] use the bio-inspired structure of BioTac sensors to achieve high-sensitivity in estimating the orientation of the contacted object. In contrast to previous work on estimating the pose of grasped objects or localizing contacts on the surface of robotic digits, our work focuses on estimating the locations of contacts between grasped objects and the environment.

Contact points can also be estimated using vision-based methods. Bernab et al. [16] propose a method that combines robot motion with point-cloud-based object tracking to estimate the location of contacts on an occupancy grid map. Hu et al. [17] use vision features, such as binocular disparity, shadows, and inter-reflections, to detect imminent contact for manipulation tasks. Due to poor accuracy and ambiguity in contact localization caused by visual occlusions, vision-based methods are generally not meant to be used alone, but rather in combination with other sensors.

Different sensor modalities for estimating contact parameters provide complimentary strengths and limitations [18]. A robot can therefore often obtain a more accurate estimate by fusing the data from multiple sources. Felip et al. [19] present a method that fuses multiple hypotheses from different modalities, including force-torque, tactile, and range sensors, to compute the likelihood of contact points. In their framework, pressure-sensitive tactile sensors on the hand are used to generate hypothesis for contacts between the object and the hand. Ishikawa et al. [20] propose estimating a contact point by intersecting a force line acquired using force-torque sensor measurements and the plane containing the contact point extracted using a task specific vision system.

In this work, we investigate the problem of estimating the contact point between a grasped object and the environment using tactile sensors. This problem is challenging as the contact point is not made directly with the tactile sensor. The robot must therefore estimate the contact point based on the forces and vibrations transferred through the grasped object. Rather than relying on analytical models, we propose a model-free data driven approach to the problem.

Object-environment contacts have usually been estimated using force-torque data from wrist-mounted sensors. However, tactile sensing also provides important information for detecting and estimating these contacts [21]. Our proposed approach thus provides an alternative sensor modality for estimating contacts, which could be combined with other modalities in a sensor fusion framework. Similar to other approaches for estimating object-environment contacts, we assume a single point of contact. We plan on extending the presented approach to multiple contacts in the future.

III. Biomimetic Tactile Sensor

In our experiments, we use a haptically-enabled Barrett robot arm with a three-fingered Barrett hand. Each finger is equipped with a biomimetic tactile sensor (BioTac) [2]. Each BioTac consists of a rigid core containing an array of 19 electrodes surrounded by an elastic skin, as illustrated in Fig. 1. The skin is inflated with an incompressible and conductive liquid.

The BioTac provides three complementary sensory modalities: force, pressure, and temperature. When the skin is in contact with an object, the liquid is displaced, resulting in distributed impedance changes in the electrode array. The impedance of each electrode depends on the local thickness of the liquid between the electrode and the skin. Micro-vibrations in the skin propagate through the fluid and are detected by the hydro-acoustic pressure sensor. The high- and low-frequency pressure vibration signals are referred to as PAC and PDC respectively. Temperature and heat flow are transduced by a thermistor near the surface of the rigid core. Since temperature conditions do not change in our experiments we do not consider this modality for our analysis.

IV. Point-of-Contact Estimation

In this section, we describe our pipeline for estimating the point of contact from tactile data. Our contact learning pipeline consists of two main parts: contact detection and contact localization. In Section IV-A, we describe how the robot estimates the time of contact. The contact localization is subsequently performed using the tactile data around this time point, as described in Section IV-B. Contact point localization can be framed as either a regression or a classification problem. We show how the contacts can be localized using

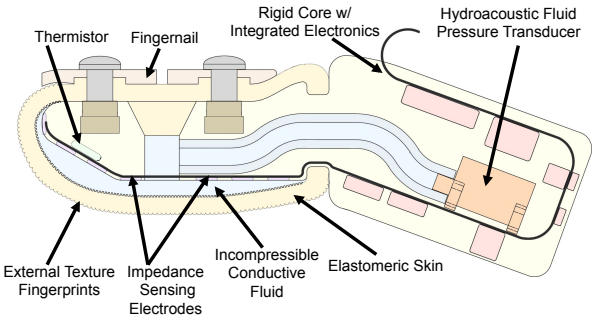

Fig. 1: Cross-sectional schematic of the BioTac sensor (adapted from [22]).

neural network or Gaussian process regression in Section IV-C. In Section IV-D, we explain how the contact parameters can be estimated using neural network or support vector machine classifiers.

A. Contact Detection

Similar to fast afferents in human skin [21], we detect contact events using the high frequency vibration signals extracted from the BioTac sensor. The robot uses a 5-th order high-pass Butterworth filter with a cut-off frequency of 20 Hz to remove biases. Contact event candidates are then extracted by detecting when the filtered pressure signal passes a threshold value. In our experiments, we used a threshold value of 500. In order to remove jitter, closely located events are reduced into a single event using the density based spatial clustering of applications with noise (DBSCAN) [23] algorithm. The resulting clusters in time dimension define the beginning and the end of every contact event.

B. Feature Extraction

The BioTacs' electrodes and hydro-acoustic pressure sensors provide multimodal tactile data for localizing contacts. To form a feature vector for localizing the contact point, we extract the values of the tactile signals after a contact event is detected. We use a window of $t = 25$ consecutive time steps at a sampling rate of 100Hz from the beginning of the event. The size of the window corresponds to the mean duration of the detected contact events plus two standard deviations. The resulting feature vector consists of the PAC, PDC, and 19 electrode signals collected from the three fingers and concatenated over the time window resulting in $s \times t = (2+19) \times 3 \times 25 = 1575$ features. In order to reduce the influence of gravity on the tactile readings, we subtract the average signal values of the three time steps immediately before the contact event.

C. Regression

We parametrize the location of the contact point using its Cartesian position (*x, y, z*) relative to the robot's palm. The direction of the contact's direction is parametrized using the *yaw* (ψ - rotation around z axis) and the *pitch* (ν - rotation around y axis). The wrist coordinate frame is shown in Fig. 6. We exclude the rotation around the contact's direction, i.e., roll, from the parametrization, as it is ambiguous for a single point of contact.

Our regression approach learns a separate mapping from the tactile features $\mathbf{d} \in \mathbb{R}^{st}$, described in Section IV-B, to each of the five continuous contact point parameters $\{x, y, z, \psi, \nu\}$, which we denote as $\hat{c} \in \mathbb{R}$. We define the function as

$$\hat{c} = f(\mathbf{d}, \theta), \tag{1}$$

where θ is a vector of function parameters that define the mapping from features to contact parameters. We compare two different machine learning techniques for learning the contact estimation function: neural networks and Gaussian processes.

The NN architecture consists of two fully connected hidden layers with 900 neurons each. The training of the NN is performed in a supervised manner using stochastic gradient descend with the Euclidean quadratic loss function

$$L_e = \frac{1}{2N} \sum_{i=1}^{N} (\hat{c}_n - c_n)^2, \tag{2}$$

where $\hat{c}_n \in \mathbb{R}$ is the predicted value of a contact parameter, $c_n \in \mathbb{R}$ is the ground truth value of the parameter, and N is the number of samples in the training set. To avoid overfitting, we use dropout for the hidden layers with a dropout ratio of 0.5, and we pick the best snapshot on the validation set observed during the training.

We compare NN regression with GP regression, a state-of-the-art non-parametric Bayesian supervised learning approach with automatic relevance determination (ARD). For the Gaussian process model we use a zero mean function, a squared exponential ARD kernel, and Gaussian noise. We also initialize the length scales, signal variance, and likelihood hyperparameters to 0, 0, and -2.3, respectively. Due to the high computational cost of GPs, we compute the average signal values during contact event window and use them as features, such that $\mathbf{d} \in \mathbb{R}^s$.

D. Classification

Regression is the standard methodology for learning mappings from features to continuous variables. However, the prediction of the contact parameters from the tactile data may be ambiguous, with a multi-modal distribution over the contact parameters. The regression approach would result in the robot averaging over the multiple possible contact parameters. Instead, we want the robot to select the most likely set of contact parameters. This can be achieved by framing the contact localization as a classification problem. In the following, we introduce a classification approach for the contact parameter estimation, which is able to represent ambiguities by producing a distribution over the possible contact parameter values.

In the classification approach, we represent the point-of-contact in the form of a distribution over the discretized

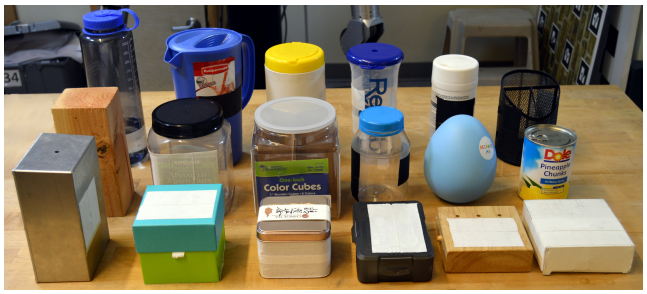

Fig. 2: The 18 objects used for data collection

contact pose parameters:

$$h(c) = p(c|\mathbf{d}, \theta), \tag{3}$$

where $c \in \mathbb{Z}$ is a one dimensional random variable corresponding to discretized contact parameter values, $\mathbf{d} \in \mathbb{R}^{st}$ and θ is the feature vector and classifier parameters as described in IV-C.

For NN we use the same architecture for the estimator, however for classification the neural net output is converted to the distribution over labels using soft-max. Furthermore, we use RMSprop adaptive learning rate with cross-entropy classification loss for training:

$$L = -\frac{1}{N}\sum_{n=1}^{N} \ln(\hat{h}_n^l), \tag{4}$$

where l is the true (target) label, N is the total number of training samples, and $\hat{h}_n^l$ is the predicted probability for the target label l of the n-th sample. The target labels are created by discretizing the continuous ground-truth data and assigning all values to the corresponding discrete bins. If a value is inside a particular bin, the corresponding target label receives the probability value 1 and all other are assigned probability 0.

Since the combined discretized parameter space produces a significant number of classes, every class would have only a few training samples given a limited amount of training data. This creates a significant obstacle for learning, since neural networks are typically prone to over-fitting. In order to mitigate this problem, we use separate classifiers for each of the five contact pose parameters.

We also apply linear support vector machines (SVMs) to perform the classification. The robot uses spatio-temporal hierarchical matching pursuit (ST-HMP) [3] to compute suitable features for the linear SVM. This feature learning framework has been successfully used for other tactile sensing applications, such as accurately predicting grasp stability [24] and detecting sensory goals using both static and dynamic tactile signals [25]. Details on how to apply ST-HMP to BioTacs can be found in the paper of Su et al. [25].

V. Evaluation

A. Data Collection Setup

The goal of this experiment is to evaluate the accuracy of the contact localization using the proposed methods. The

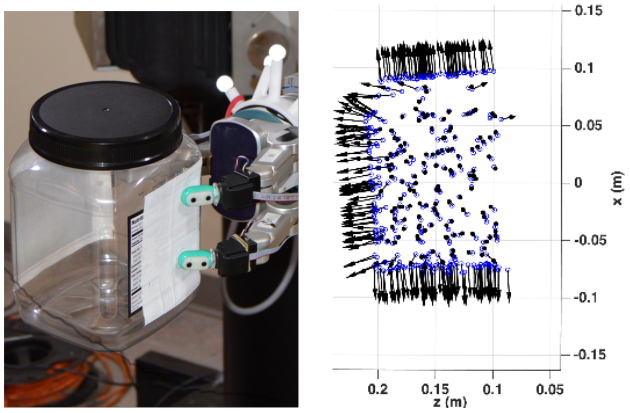

Fig. 3: An example of a grasp used and contact samples collected during the experiment. Left: an example grasp of an object. Right: an example of the detected contact points and directions projected onto the zx plane.

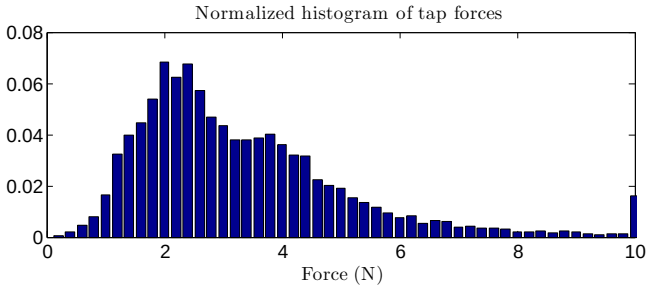

Fig. 4: The normalized histogram of the peak tapping forces applied to the object during the experiment.

experiments were performed using a three-fingered Barrett hand. Each finger tip is equipped with a BioTac tactile sensor.

Eighteen objects with a variety of sizes and materials were chosen for this experiment. All of the objects are shown in Fig. 2. The grasped objects' textures were modified to prevent slippages during tapping. During the data collection, the robot held one of the objects (see Fig. 3 for an example grasp) while a person tapped the object with a plastic rod shown in Fig. 5. Fig. 4 shows a normalized histogram of the peak tapping forces applied to the objects, which were measured using an ATI force-torque plate attached to the tapping rod. The steady state grip forces applied by the robot and measured by the normal force on the BioTacs [12], were $11.55N$ ($\pm 5.84N$) among 18 objects.

In order to determine the location and direction of the contact event, the rod was tracked using a marker-based Vicon tracking system, as shown in Fig. 5. The wrist of the robot was also tracked using Vicon markers, as shown in Fig. 6 with the corresponding coordinate frame. Using the coordinate frames of the wrist and the rod, the robot computed the location and direction parameters of the contact points relative to the robot's wrist.

The data from the contact events was recorded as a continuous time series. The BioTac sensor readings were then extracted using the contact detection method described

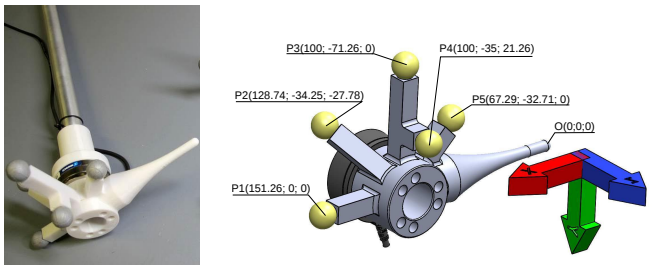

Fig. 5: The tapping rod. Left: an image of the rod. Right: rod coordinate frame with the locations of the markers.

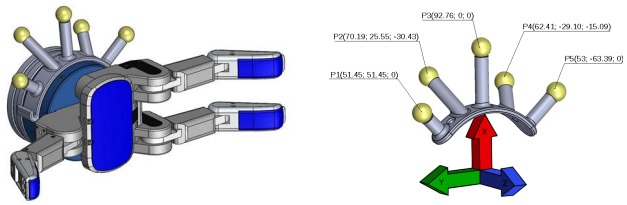

Fig. 6: The wrist band with Vicon markers. Left: CAD model of the Barrett hand with the band. Right: wrist coordinate frame with locations of the markers.

in Section IV-A. Using this approach, the robot detected $\approx$ 15100 samples for all 18 objects. The number of taps applied to a particular object varied depending on the object's size.

In order to evaluate the contact localization 80% of the collected data was randomly selected to train the estimators using the methods described in Section IV. The test set consisted of 12% of the collected data. Since the neural nets are prone to overfitting to the data if trained for too long, the remaining 8% of the data was used as a validation set to continuously evaluate the learners' generalization performance. Training continued until no improvement in validation error was observed during 350 consecutive iterations. The network with the smallest validation error was selected and evaluated on the test data. This evaluation process was repeated for all 18 objects.

Since classification requires a finite number of classes, we define an estimation area around the wrist coordinate frame as a rectangular box with dimensions: x = $-20..20$ cm; y = $-10..10$ cm; z = $0..30$ cm; and discretize these dimensions using different grid sizes: 3 cm, 2.5 cm, 2 cm, 1.5 cm, 1 cm, 0.5 cm and 15°, 12.5°, 10°, 7.5°, 5°, 2.5° for Cartesian and angular coordinates respectively, resulting in $13..80$ classes for the x dimension; $6..40$ classes for the y dimension; $10..60$ classes for the z dimension; $24..144$ classes for the *yaw* and $12..72$ classes for the *pitch* dimension for different grid sizes. The dimensions of the estimation area were picked to accommodate sizes of all objects. The minimal step size for the grid was limited by the Vicon system's tracking error.

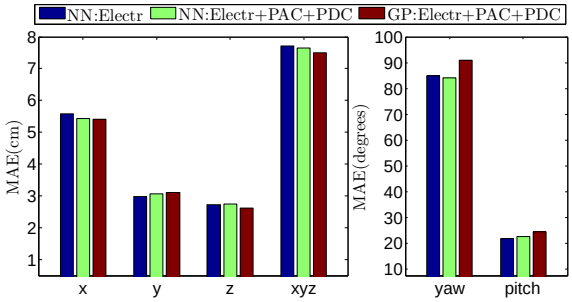

Fig. 7: Results of regression using different sensor modalities for NN and GP regressors. *NN:Electr* - NN with electrodes only; *NN:Electr+PAC+PDC* - NN with the full set of sensor modalities (electrodes, PAC, PDC); *GP:Electr+PAC+PDC* - GP with the full set of sensor modalities. *x*, *y*, *z*, *yaw*, *pitch* are results for independent dimensions. *xyz* represents results for combined error, i.e. the norms of the error vectors in Cartesian space. Errors in Cartesian space are reported in cm; angular errors are reported in degrees.

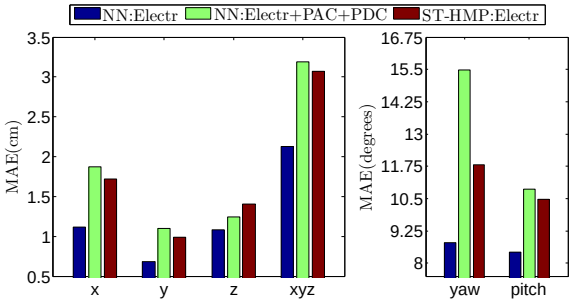

Fig. 8: Results of classification using different sets of sensor modalities: *NN:Electr* - NN classifier with electrodes only; *NN:Electr+PAC+PDC* - NN clasifier with electrodes, PAC and PDC features; *ST-HMP:Electr* - ST-HMP feature learning algorithm with SVM classifier applied to electrode features.

B. Results and Discussion

Fig. 7 shows the mean absolute error (MAE) calculated from errors of all 18 objects. We evaluated NN regression using electrode features (*NN:Electr* in the figure) and using the full feature set (*NN:Electr+PAC+PDC* in the figure). We also evaluated using Gaussian Processes regression using the full feature set (Fig. 7 *GP:Electr+PAC+PDC*). The three regression approaches resulted in considerable errors that, in some cases, exceed 50% of the object's size, although we did not observe overfitting in either of cases. Such significant errors are probably caused by ambiguities in the mapping between features and the estimated contact parameters, which can not be represented properly by the regression. These results motivated us to approach our problem from the point of classification, which is also more suitable for application of neural networks, where they traditionally show superior results over other machine learning techniques ([26], [27]).

Similar to the regression approach, we also evaluated NN classifiers using only the electrode features and using the full

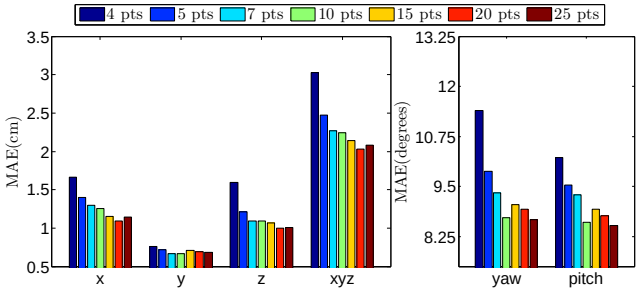

Fig. 9: Results of NN classification using different time lengths of features, starting from the moment contact event is detected.

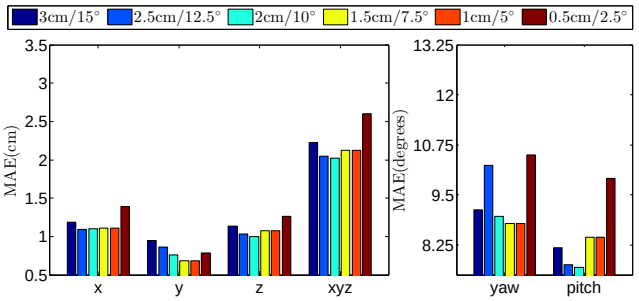

Fig. 10: Results of NN classification using different grid steps.

set of features. For this experiment, we pick a 1cm/5° grid with 25 time steps as a baseline parameter set. Fig. 8 presents the MAE across the test sample sets of all 18 objects using our classification approach described in Section IV-D. In this case, MAE is calculated as the average of the absolute value of the difference between the real contact point parameters and the middle of the bin predicted by the classifier. We also combine predictions of individual dimensions in Cartesian coordinates for every sample in order to calculate the average Euclidean norm of the error vector for all location predictions. This prediction is denoted as *xyz*. The results indicate that electrodes are the most relevant features for contact localization. Incorporation of PAC and PDC injects additional noise and leads to overfitting, which is a typical problem in learning [28]. Thus, for further investigation we restrict our experiments to electrodes only and apply an SVM classifier with ST-HMP features extracted from the array of electrodes (see Fig. 8, *ST-HMP:Electr*). The results show that NN classifier outperforms ST-HMP by ≈ 0.9 cm for the overall Euclidean error and up to 3° for angular coordinates. Given these results, we decided to use the NN classifiers for the following investigation.

The robot's reaction time to contacts can play a crucial role for some applications of the contact localization algorithms, e.g., if they are used in a control loop. Thus, it is important to understand what contributes the most to the estimation delays. For our algorithms, the most significant source of delays is the data accumulation. For example, 25 time points with 100 Hz frequency causes a delay of 250 ms, whereas running forward pass of the NN requires less then 5 ms on a

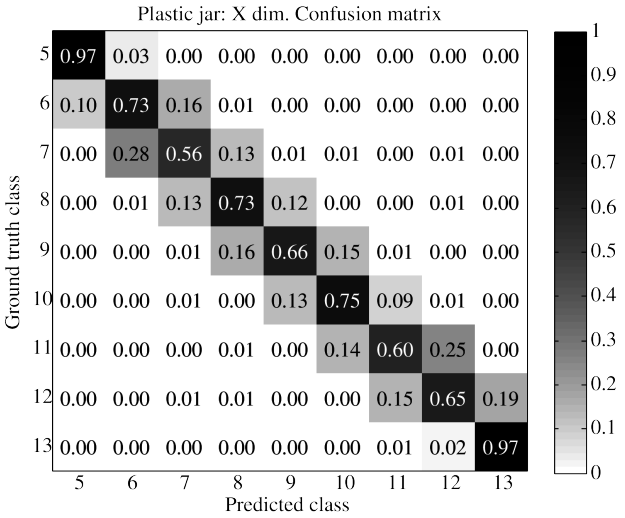

Fig. 11: An example confusion matrix of the x dimension classification for one of the objects.

Core i5 CPU. It is therefore interesting to see how the prediction error varies with respect to the number of time steps of the features collected. Fig. 9 shows the classification MAE for the NN given different numbers of time steps retained from the moment the contact event is detected. As one can see from the figure, the algorithm is quite robust to the time window size, and its performance degrades gracefully until five time steps. Below five time steps, the performance starts dropping. The drop in performance can partially be attributed to imperfections in the detection pipeline, which introduce variable time shifting of the signals relative to the position inside the time window.

In all previous experiments we used a 1 cm/5° grid as a baseline parameter. Thus, it would be interesting to see how sensitive our classifiers are to the grid resolution used to discretize the contact parameter estimations. To investigate that effect, we vary our grid sizes from 0.5 cm/2.5° to 3 cm/15° with steps of 0.5 cm/2.5° and train NN classifiers. Our results, as shown in Fig. 10, indicate that the MAE does not change significantly and it does not exceed 3 cm/11°. This result means that even if the classifiers cannot guess the exact label they usually predict one of the adjacent bins, which indicates that they learn the underlying relations between different classes. This can also be seen from an example confusion matrix for the x dimension presented in Fig. 11. In the confusion matrix, the rows represent the ground truth (target) classes and the columns represent the predicted classes. In most cases the recall is quite high and misclassified labels cluster around the target class. The results also clearly indicate that classifiers produce almost no predictions for the classes that were not presented for the training, thus, making it safe to preallocate larger prediction areas if needed and keep the architecture of the classifiers the same while incorporating new data samples into learning.

VI. Conclusion and Future Work

In this paper, we explored the problems of detecting and localizing contacts between a held object and the environment. The contacts were detected by thresholding the BioTacs' pressure signals. Contact localization was performed by applying different machine learning methods, including neural networks, Gaussian proccesses, and support vector machines with ST-HMP feature learning, to tactile data. We framed the contact localization as both a regression and a classification problem and investigated sensitivity of our algorithms to time and component-wise changes in the input features, as well as various discretizations of the parameter space for our classification approach.

We evaluated our methods using hundreds of contact events from eighteen objects with different shapes and material properties. Our classification approach resulted in the best performance, with expected localization errors less than 2.5 cm/10° for individual objects and poses.

In future work, we will explore the generalization of the learned classifiers and regressors between different objects, as well as predicting multi-point contacts.

Acknowledgments

The authors thank Felix Grimminger [2] for design of the tapping rod, the Vicon wrist band and help with setting up the Vicon system; Karol Hausman and Yevgen Chebotar [3] for help with writing software components. BioTac sensors used in this research were kindly provided by SynTouch LLC. This work was funded in part by the NSF under grant CNS-1213128 and the ONR under grant N00014-14-1-0536.

References

[1] R. S. Dahiya, M. Gori, G. Metta, and G. Sandini, "Better manipulation with human inspired tactile sensing," in *RSS 2009 workshop on Understanding the Human Hand for Advancing Robotic Manipulation. RSS*, 2009, pp. 1–2.

[2] N. Wettels, V. J. Santos, R. S. Johansson, and G. E. Loeb, "Biomimetic tactile sensor array," *Advanced Robotics*, vol. 22, no. 8, pp. 829–849, 2008.

[3] M. Madry, L. Bo, D. Kragic, and D. Fox, "ST-HMP: Unsupervised Spatio-Temporal Feature Learning for Tactile Data," in *IEEE International Conference on Robotics and Automation (ICRA)*, may 2014.

[4] Y. Karayiannidis, C. Smith, F. E. Vina, and D. Kragic, "Online contact point estimation for uncalibrated tool use," in *2014 IEEE Int. Conf. on Robotics and Automation (ICRA)*. IEEE, 2014, pp. 2488–2494.

[5] N. Likar and L. Zlajpah, "External joint torque-based estimation of contact information," *International Journal of Advanced Robotic Systems*, vol. 11, 2014.

[6] H. Bruyninckx, S. Dutre, and J. De Schutter, "Peg-on-hole: a model based solution to peg and hole alignment," in *Robotics and Automation, 1995. Proceedings., 1995 IEEE International Conference on*, vol. 2. IEEE, 1995, pp. 1919–1924.

[7] G. E. Hovland and B. J. McCarragher, "Combining force and position measurements for the monitoring of robotic assembly," in *Intelligent Robots and Systems, 1997. IROS'97., Proceedings of the 1997 IEEE/RSJ Int. Conf. on*, vol. 2. IEEE, 1997, pp. 654–660.

[8] A. Petrovskaya, J. Park, and O. Khatib, "Probabilistic estimation of whole body contacts for multi-contact robot control," in *Proceedings 2007 IEEE International Conference on Robotics and Automation*. IEEE, 2007, pp. 568–573.

[2]Felix Grimminger is affiliated with Max Planck Institute of Intelligent Systems, Germany

[3]Karol Hausman, Yevgen Chebotar are PhD students at the University of Southern California, Los Angeles, United States

[9] G. S. Koonjul, G. J. Zeglin, and N. S. Pollard, "Measuring contact points from displacements with a compliant, articulated robot hand," in *Robotics and Automation (ICRA), 2011 IEEE International Conference on*. IEEE, 2011, pp. 489–495.

[10] M. Kaneko and K. Tanie, "Contact point detection for grasping an unknown object using self-posture changeability," *IEEE Transactions on Robotics and Automation*, vol. 10, no. 3, pp. 355–367, 1994.

[11] V. Ciobanu, D. Popescu, and A. Petrescu, "Point of contact location and normal force estimation using biomimetical tactile sensors," in *Complex, Intelligent and Software Intensive Systems (CISIS), 2014 Eighth International Conference on*. IEEE, 2014, pp. 373–378.

[12] Z. Su, K. Hausman, Y. Chebotar, A. Molchanov, G. E. Loeb, G. S. Sukhatme, and S. Schaal, "Force estimation and slip detection/classification for grip control using a biomimetic tactile sensor," in *Humanoid Robots (Humanoids), 2015 IEEE-RAS 15th International Conference on*. IEEE, 2015, pp. 297–303.

[13] C. Corcoran and R. Platt, "A measurement model for tracking hand-object state during dexterous manipulation," in *Robotics and Automation (ICRA), 2010 IEEE International Conference on*. IEEE, 2010, pp. 4302–4308.

[14] R. Li, R. Platt, W. Yuan, A. ten Pas, N. Roscup, M. A. Srinivasan, and E. Adelson, "Localization and manipulation of small parts using gelsight tactile sensing," in *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2014, pp. 3988–3993.

[15] Z. Su, S. Schaal, and G. E. Loeb, "Surface tilt perception with a biomimetic tactile sensor," in *2016 6th IEEE RAS & EMBS international conference on biomedical robotics and biomechatronics (BioRob)*. IEEE, 2016 (Accepted).

[16] J. A. Bernabé, J. Felip, A. P. Del Pobil, and A. Morales, "Contact localization through robot and object motion from point clouds," in *2013 13th IEEE-RAS International Conference on Humanoid Robots (Humanoids)*. IEEE, 2013, pp. 268–273.

[17] H. H. Hu, A. A. Gooch, W. B. Thompson, B. E. Smits, J. J. Rieser, and P. Shirley, "Visual cues for imminent object contact in realistic virtual environment," in *Proceedings of the conference on Visualization'00*. IEEE Computer Society Press, 2000, pp. 179–185.

[18] O. Kroemer, C. Lampert, and J. Peters, "Learning dynamic tactile sensing with robust vision-based training," *IEEE Transactions on Robotics (T-Ro)*, no. 3, pp. 545–557, 2011.

[19] J. Felip, A. Morales, and T. Asfour, "Multi-sensor and prediction fusion for contact detection and localization," in *2014 IEEE-RAS Int. Conf. on Humanoid Robots*. IEEE, 2014, pp. 601–607.

[20] T. Ishikawa, S. Sakane, T. Sato, and H. Tsukune, "Estimation of contact position between a grasped object and the environment based on sensor fusion of vision and force," in *Multisensor Fusion and Integration for Intelligent Systems, 1996. IEEE/SICE/RSJ International Conference on*. IEEE, 1996, pp. 116–123.

[21] R. S. Johansson and J. R. Flanagan, "Coding and use of tactile signals from the fingertips in object manipulation tasks," *Nature Reviews Neuroscience*, vol. 10, no. 5, pp. 345–359, 2009.

[22] Z. Su, J. Fishel, T. Yamamoto, and G. Loeb, "Use of tactile feedback to control exploratory movements to characterize object compliance," *Frontiers in neurorobotics*, vol. 6, 2012.

[23] M. Ester, H.-P. Kriegel, J. Sander, X. Xu *et al.*, "A density-based algorithm for discovering clusters in large spatial databases with noise." in *Kdd*, vol. 96, no. 34, 1996, pp. 226–231.

[24] Y. Chebotar, K. Hausman, Z. Su, G. S. Sukhatme, and S. Schaal, "Self-supervised regrasping using spatio-temporal tactile features and reinforcement learning," in *IEEE Int. Conf. on Intelligent Robots and Systems (IROS)*. IEEE, 2016 (Accepted).

[25] Z. Su, O. Kroemer, G. E. Loeb, G. S. Sukhatme, and S. Schaal, "Learning to switch between sensorimotorprimitives using multimodal haptic signals," in *International Conference on Simulation of Adaptive Behavior (SAB)*. Springer, 2016 (Accepted).

[26] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[27] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[28] P. M. Domingos, "A few useful things to know about machine learning." *Commun. ACM*, vol. 55, no. 10, pp. 78–87, 2012.