2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
October 25-29, 2020, Las Vegas, NV, USA (Virtual)

Interaction identification through tactile sensing during cloth manipulation using a 3-axis touch sensor

Idril Geer, Marc Maceira, Júlia Borràs, Carme Torras, Guillem Alenyà

***Abstract*— Tactile feedback during cloth manipulation could be crucial in addressing the huge challenges involved in closing the loop during execution, complementing vision. However, up to our knowledge, tactile sensing has only been successfully used in cloth manipulation to classify type of fabrics, detect how many layers were grasped, and estimate the grasping force. In this work, we want to explore its potential to also provide information about whether the task is executed as expected. Two types of experiments are performed, in which a robot carries out 4 simple tasks, all involving a single finger manipulating a flat cloth on a table. Firstly, we analyze the sensor's signals once the cloth manipulation has finished using Dynamic Time Warping (DTW) to see if they are informative enough to classify the tasks. Our results show that tactile feedback depends highly on the type of manipulated fabric. Secondly, we analyze the tactile feedback during the manipulation of the cloth using a recurrent neural network (RNN). For each sensor measurement, the RNN recognizes if the finger slides over the cloth, pulls it, flattens a fold in it, or if it's about to lose contact, with 95.7% accuracy. These are promising results that show how tactile sensing has the potential of providing crucial information that would be very difficult to obtain with vision only.**

Fig. 1. Setup: A Stäubli robot arm with detail of the fingertip touch sensor covered with a silk cover at the top-right, without the cover at the bottom-right.

I. Introduction

When manipulating rigid objects, a robot can only change their position and orientation, namely six parameters. However, the manipulation of highly deformable objects such as clothing involves a potentially infinite-dimensional shape-state space [1]. This makes traditional geometry-based perception algorithms developed for rigid-objects hardly applicable in the context of cloth manipulation. Advances in perception techniques for clothes involve machine learning approaches for recognizing cloth parts, such as corners or necks [2], and applications of motions to the object to aid state estimation of the garments. However, due to the huge dimensionality of the cloth configuration space, it is often difficult to understand its state and act accordingly during manipulation.

In this context, more sensory information is crucial to function, especially when environmental circumstances or self-occlusions impair vision. Force and tactile sensors can enable robots to manipulate objects with greater precision and sensitivity and allow operation in less-structured environments. In the case of cloth manipulation, which is so crucially based on complex and usually inefficient perception techniques, tactile feedback may be essential to the development of more natural, precise and robust manipulations.

The lack of efficient models and the difficulty of applying closed-loop control where the robot can react during manipulation is one of the main challenges the community is facing [3]. Providing additional tactile sensing could compensate for the limitations of using only vision for task monitoring and cloth state recognition. For instance, when unfolding a textile object in the air, one can trace the edge sliding over it with the second hand, without losing the grasp, to get from one corner to the other [4]. In this case, tactile feedback can be used to detect if such grasp has been lost, or more importantly, if the robot is going to lose it in the near future, in order to act to prevent it.

In this work, we show how tactile sensing can aid the robot to recognize if the cloth is sliding, if it is tightly grasped or if the commanded task is executed correctly. To this end, we have designed a set of repeatable and measurable tasks consisting of sliding or flattening a cloth on a table with just one finger equipped with a touch sensor (see Fig. 1). These are simple but representative tasks because cloth manipulation very often involves handling flat clothes on a table. In addition, sliding a cloth between a finger and a table is a simplification of sliding it between two fingers. We also consider different cloth types such as a kitchen cloth, a thin cloth napkin, a cotton T-shirt, and a piece of denim.

First, from the whole tactile signal, we want to identify

This work receives funding from the European Union Horizon 2020 Program under grant agreement no. 741930 (CLOTHILDE), by the HuMoUR project TIN2017-90086-R (AEI/FEDER, EU) and by AEI through the María de Maeztu Seal of Excellence to IRI (MDM-2016-0656).

The authors are with Institut de Robòtica i Informàtica Industrial, CSIC-UPC, Llorens i Artigas 4-6, 08028 Barcelona, Spain. {igeer, mmaceira, jborras, torras, galenya}@iri.upc.edu

which of our four predefined tasks corresponds to it, using Dynamic Time Warping as a similarity measure. [5]. This analysis allows us to study differences when executing with different clothes or using different forces. From this analysis, we conclude that different fabrics produce distinct enough signals, so that the fabric type needs to be included in the task definition if we are to achieve accurate identifications.

Secondly, considering the tactile signal at each time step of the execution, we want to identify the specific contact interaction between the finger, cloth and table using recurrent neural network (RNN) techniques [6]. In other words, we aim to identify if the finger is sliding over the cloth, moving it, flattening a fold or losing contact. Using the RNN analysis, we can identify with 95.7% accuracy what the particular contact interaction at each time step is, regardless of the type of cloth. These are promising results that can be crucial for a robot to understand whether a commanded task is being carried out correctly or corrections need to be made.

II. Related work

There is an extensive bibliography on cloth manipulation, although it has received relatively little attention compared to rigid object manipulation. Due to the complexity of the problem, works often involve a variety of techniques [7] in sensing, recognition, machine learning and manipulation, to complete a task. Benchmarks [8] and frameworks [9] for such manipulations have also been developed.

As far as the sensing strategies are concerned, the vast majority of works focus on 2D and 3D vision sensing. The two main issues that need to be solved are grasp point detection and cloth state estimation, and vision is usually combined with induction of cloth movement through manipulation to reduce uncertainty. A good survey of these techniques can be found in [3]. For instance, [2] used 2D images and depth data to detect collars or relevant parts to be grasped. Others find grasp points by detecting contours [10] or edge discontinuities using depth data [11], and more recently, using deep learning [12]. The problem of grasp point detection often is tackled by detecting the lowest hanging point of a cloth. As an example, [13] used Active Random Forest while rotating the cloth. To estimate the cloth state, authors mostly used 3D vision data to reconstruct a mesh [14], Hidden Markov Models to estimate identity and track configuration during a specific sequence of manipulations [15], or depth data in combination with cloth motion [16]. Cloth can be reconstructed by comparing its mesh to a database [17]. In general, these works manipulate the cloth in order to bring it to a known configuration, usually flat on a table. In [18], RGB-D data is interpreted to locate wrinkles on a semi-flat cloth on a table. Once a cloth is already flat, contours can be used to define the cloth state, like in [19–21].

There is an extensive bibliography of tactile sensors to identify textures, that has been recently extended to recognize different fabrics. For instance, very high accuracy of recognition has been achieved with machine learning techniques [22–24]. Particularly [24] could recognize, in addition to the type of textile, how many layers were grasped using a Support Vector Machine (SVM).

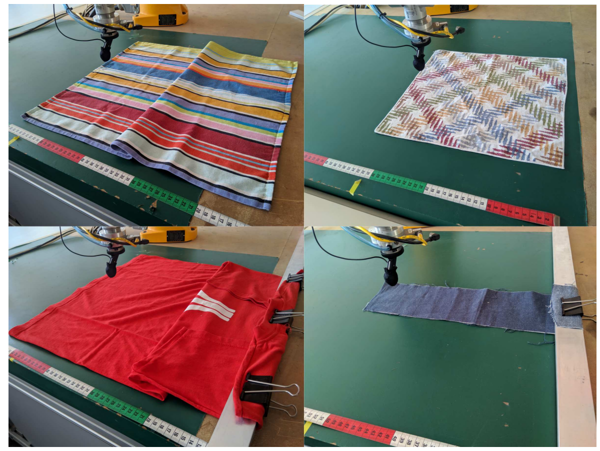
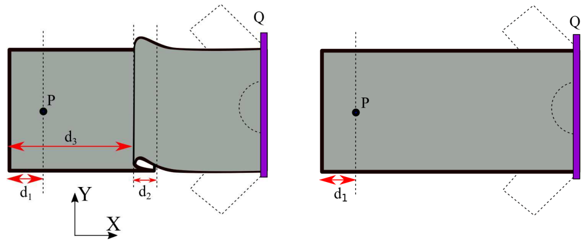


Fig. 2. We applied our tasks to 4 different objects. O1 (top-left) is a large kitchen cloth, here in the initial configuration of Task 1. O2 (top-right) is a small and thin cloth napkin, here set in the initial configuration of Task 2. O3 (bottom-left) is a T-shirt, here set in the initial configuration of Task 3, and O4 (bottom-right) is a cut piece of denim cloth without hems, here in the initial configuration of Task 4. Note that O4 is only used for Task 4.

Many works on mechanical designs for cloth specialized grippers include tactile sensors [25–27]. The sensors in [26] are used to estimate appropriate grasping forces for different cloth items. A specially designed tactile sensor is used in [27] to develop a gripper controller and to classify different fabrics.

There are a few works in sensor fusion. For instance, in [28] RGB-D sensors are combined with photometric and tactile data to simultaneously recognize the type of garment, fabric pattern and material using various machine learning techniques. Force feedback, in combination with vision, has been used to build a controller to manipulate cloth maintaining it in tension [29], or to use the force feedback to better reproduce learned tasks [30].

Recent works in the literature explore the tactile sensing task using high-accuracy 3D data. One paper [31] explores how pure tactile sensing can be used to accurately manipulate cables, which are similar deformable objects to fabrics, using high-resolution sensors based on the GelSight. A similar topic is discussed in [32], using a BioTac sensor to follow the contour of a ziplock bag. To the best of our knowledge though, nobody has used tactile sensing with the simpler 3-axis domed force sensors to recognize the type of cloth manipulation task the robot is executing or to recognize the contact interaction with the cloth. This kind of recognition could be crucial to successfully monitor task execution during cloth manipulation, especially when combined with vision, for example in slip detection and adjusting during folding of textiles.



(a) (b)

Fig. 3. The two initial states of cloth considered, with and without clamping on the Q line. For the experiments the parameters were set to $d_1 = d_2 = 10\%L$, $d_3 = 50\%L$ where L is the total length of the piece of cloth in the X direction.

III. Data collection

The study of the literature from the previous section does not provide any dataset for our task. To this end, we generate a data collection, that will be published, to validate our work. In this section, we explain the data generation and annotation process. The data collection generated comprises two sets of annotations for each sample: a *manipulation task* that defines the end result of each sample and the *contact interactions* which indicates the effect of the manipulation on the cloth during their manipulation. The *manipulation task* annotation comprises one action for the full sequence while the *contact interactions* annotation is provided for each sample of the force sensor.

We use a Stäubli RX60 6-DOF robot arm to execute repeatable trajectories for the trials. On its tool center point (TCP) we attached an Optoforce OMD-20-SE-40N sensor, to which we added a silk textile covering to reduce friction. A picture of the setup can be seen in Fig. 1. The same touch sensor has been used in cloth manipulation grippers in [25]. It returns 3-dimensional feedback of linear forces in the X, Y and Z directions. Precise and repeatable rectilinear trajectories are generated with constant orientation to get results with consistent vertical forces. Trajectories are generated and executed through the MoveIt! [33, 34] framework, sensor data is obtained using the ROS driver for the sensor set to 333 Hz and a 1.5 Hz filter [35].

We define a manipulation task as *a sequence of contact interactions between the finger and the cloth/environment, to perform a task*, such as flattening or pulling a cloth on a table. The different types of contact interactions are defined as *the different relative motions that occur during a task between the finger, the cloth, and the environment.*

We consider 4 tasks, consisting of pulling on clothes with a single finger contact. When the cloth is clamped, the finger slides over it and the clamp prevents the overall cloth motion. When the cloth is not clamped, the same trajectory results in pulling and moving it. When a fold is present, pulling flattens the cloth. Although these are simple tasks, they produce up to 9 different contact interactions that are very relevant to cloth manipulation.

TABLE I

Types of contact interactions during task executions

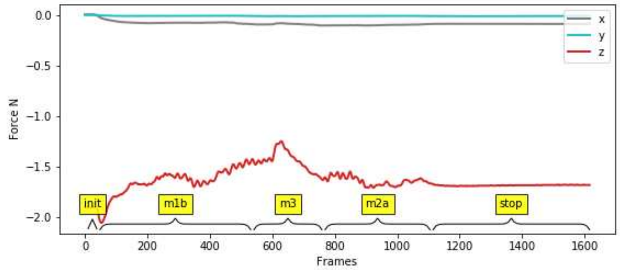
	Description	**Rel. motion F-C**	**Rel. motion C-T**
m1	Finger moves the cloth, flattening the fold *Sub-types: -m1a: Finger on the hem -m1b: Finger not on the hem	no	yes
m2	Finger moves the whole cloth *Sub-types: -m2a: Finger on the hem -m2b: Finger not on the hem	no	yes
m3	Finger slides on the cloth without moving it	yes	no
m4	Finger is starting to go over the hemline	yes	no
m5	Finger slides over a hemless edge of the cloth	yes	no
m6	Finger has just lost contact with the cloth	yes	no
m7	No contact with the cloth	no	no

F stands for finger, C for cloth and T for table.

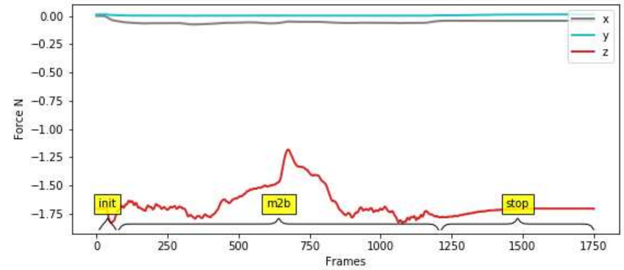
We designed the tasks so that they all use the same motion. They all start with the finger from an initial contact located at the point **P** and move the finger along the **X** direction, as in Fig. 3.

The four tasks are:

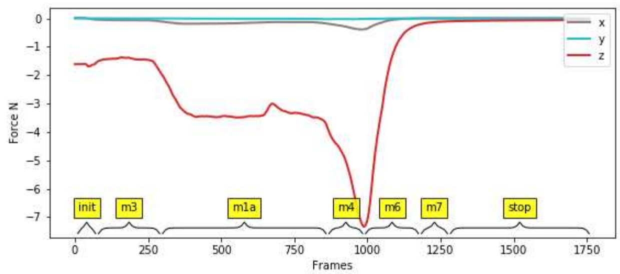
- **Task 1**: Flatten pulling a cloth
  Initial configuration: The cloth is on the table with a fold along the Y direction (Fig. 3-(a)), without any attachment to the table. An example can be seen in Fig. 2-(top-left).
  Expected result: The fold is flattened. Once flat, the cloth is moved by the finger on the table.
- **Task 2**: Pulling a flat cloth
  Initial configuration: The cloth is flat on the table (Fig. 3-(b)), free to move. An example can be seen in Fig. 2 -(top-right).
  Expected result: The finger moves the cloth on the table.
- **Task 3**: Flatten pulling a fixed cloth
  Initial configuration: The cloth is in the same configuration as Task 1 (Fig. 3-(a)), attached to the table with a clamp along Q. An example can be seen in Fig. 2-(bottom-left).
  Expected result: The fold is flattened but the cloth does not move because it is clamped to the table. The finger passes over the edge
- **Task 4**: Sliding finger over a cloth
  Initial configuration: The cloth is in the same configuration as Task 2 (Fig. 3-(b)), attached to the table with a clamp along Q. An example can be seen in Fig. 2-(bottom-right).
  Expected result: The finger slides on the cloth without moving it until it loses contact, going over the hem at the edge.



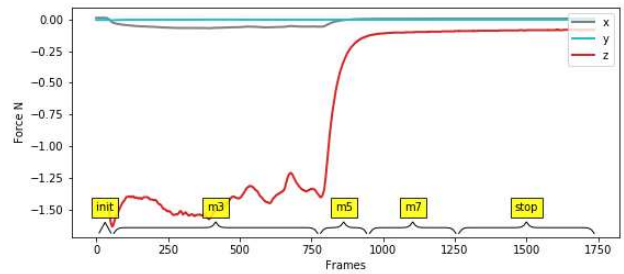
Tactile feedback signal of an execution of Task 1 with the kitchen cloth. Finger unfolds (m1b) until cloth is flat, then slides over it (m3) until it reaches the hem, then it pulls the whole cloth (m2a) until it stops



Signal of an execution of Task 2 with a small napkin. The finger pulls and moves the whole cloth (m2b) until it stops.



Execution of Task 3 with the T-shirt. The finger slides over the cloth (m3) until it reaches the hemline and starts pulling, flattening the fold (m1a). When it is flat, finger goes over the hemline (m4), loses contact (m6) and keeps moving without cloth under it (m7) until it stops.



Execution of Task 4 with the hemless cloth. The finger starts sliding over the cloth (m3) until it crosses the edge (m5). Then it keeps moving without cloth under it (m7) until it stops. As expected, it differs a lot from the other graphs where hemlines are involved

Fig. 4. Examples of the tactile signals and the different contact interactions for each task.

We executed the tasks with three objects: two different napkins and a T-shirt. Task 4 was also executed with a cut piece of denim without seams to generate loss of contact data with and without hems. Each trial is annotated by the task performed and the cloth used. This divides the data into classes, corresponding to the task and the cloth used. All the objects are shown in Fig. 2. Each Task was executed 10 times per object, with a few being discarded due to problems in setup and recording, with a vertical force of $1.5N$. The intensity of the force was chosen empirically to flatten the fold without creating additional wrinkles, as we observed they affected the robustness of the results. Data was logged using rosbags, including video, sensor data and joint positions. We gathered a total of 125 trials. In the complementary video [36] we show examples of all the task executions with the different items.

From the executions of the manipulation tasks above, we identified 9 types of contact interactions between finger (F), cloth (C) and table (T). Each contact interaction identified carries valuable information during the cloth manipulation task, such as if the finger is moving the whole cloth, if it's flattening the fold or if a hemline is present. The complete list of interactions is defined in Table I. In Fig. 4 we can see examples of how the intervals of all the tactile signals have been labeled. Labels are manually assigned checking both the video and the signal recorded during execution.

IV. Manipulation Task classification

We start the experimental part assessing the manipulation task. In our analysis, we want to study how similar the executions from the same task are, and if they can be recognized. We expect similar signals from elements of the same class, so that they can be properly classified. We define the similarity between two 3-dimensional signals in terms of the *Dynamic Time Warping*$_{Adaptive}$: DTW_A distance defined in the work of Shokoohi-Yekta et al. for multidimensional time series [5]. The DTW_A distance depends on a threshold parameter that decides whether to use the dependent (DTW_D) or independent (DTW_I) distance. We use the threshold set to 1, meaning we take the minimum of them. Given two tactile signals $\mathbf{S} = (S_x, S_y, S_z)$ and $\mathbf{T} = (T_x, T_y, T_z)$, we will write $DTW_A(\mathbf{S}, \mathbf{T})$ to refer to the distance defined in [5].

To classify our data, we use a Nearest Centroid Classifier [37, 38]. In this method, the training data is manually separated by clusters and each cluster i is represented by its centroid $\mathbf{C}_i$. We use DTW Barycenter Averaging on all the elements of the cluster to compute its centroid [39]. Then, each new signal $\mathbf{S}$ is classified by finding the minimum DTW_A distance to each of the centroids: $\min_i(DTW_A(\mathbf{S}, \mathbf{C}_i))$. We repeat the process 100 times to validate our results, and in each iteration we randomly sample 70% of our data to define sets of clusters based on the different criteria we want to test for: by task, by type of cloth, by task and type of cloth, etc. The random sampling is equally distributed among our classes of trials. Then, we used the remaining 30% of the data to test.

A. *Results*

In this section we analyze the classification accuracy of the dynamic time warping based classification with different setups. First, we provide results with all the clothes aggregated together. Then, we analyze if we can improve the classification differentiating between cloth types.

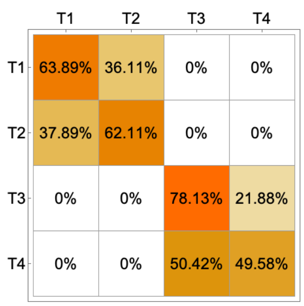


Fig. 5. Classification by task. We can see that we can only make a good distinction between the case where the cloth is free to move (Tasks 1 and 2) and the cloth is clamped on the table (Tasks 3 and 4). The signals as seen in fig. 4 are too similar for DTW to make a reliable distinction between them.

1) Task identification without separating by cloth type: We first wanted to evaluate if the executed task could be identified from the tactile feedback signal, regardless of what type of cloth it was executed on. To this end, we grouped data in 4 clusters corresponding to the 4 different tasks, without taking the cloth item into account The obtained confusion matrix is shown in Fig. 5. Each cell c_{ij} contains the percentage of times the task T_i was classified as T_j for all our test set. We can see how we can clearly distinguish between tasks in which the cloth is clamped (Tasks 3 and 4) with respect to tasks where the cloth can move freely (Tasks 1 and 2). However, with this grouping, "flattening a fold" vs. "moving the finger over an already flattened cloth" has high confusion and cannot be distinguished reliably.

This result indicates that signals of different cloth types are too different and should not be grouped in the same cluster for classification.

2) Task identification for each of the cloth types: Following the results in the previous section, we considered dividing our data by the three types of cloth (O1 to O3 in Fig. 2), and identifying the tasks for each of the three sets. The resulting three confusion matrices can be seen in Fig. 6.

In this case the classification is better in almost all the cases. In the worst case, we get a 91.67% correct classification for the kitchen cloth (T2), while the rest are properly classified 100% of the times. That is, the method classifies Task 2 (where the finger pulls the cloth moving it) as Task 1 (where the finger pulls a fold flattening it), only in 8.3% of the cases. Bearing in mind how similar these tasks are, this is a promising result. This confusion is similar to that of the T-shirt and worse for the thin napkin where, in addition, the method classifies Task 1 (flatten a fold) as Task 2 (pull a flat cloth) in 30% of cases.

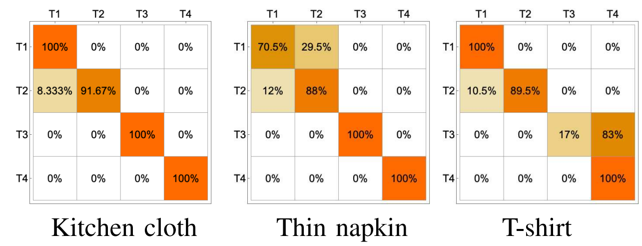
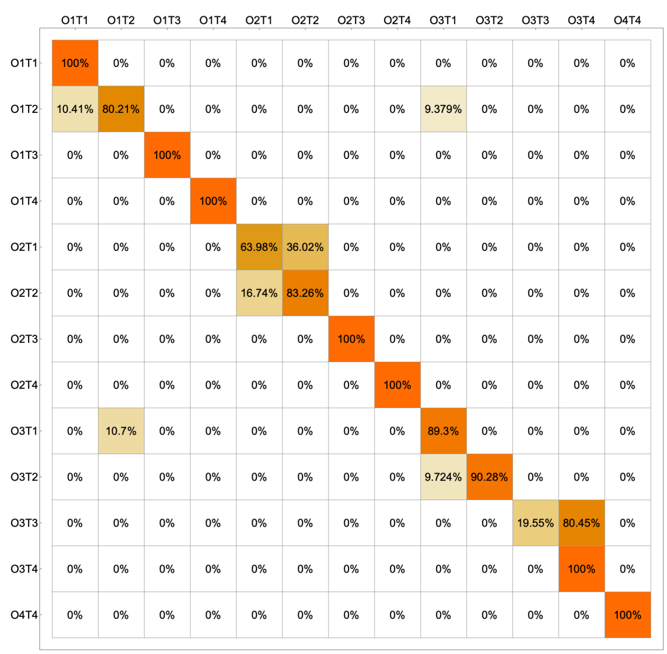


Fig. 6. Classification results for each individual object: the two napkins and the T-shirt.

For the first two objects, we can distinguish between Tasks 3 and 4 very well, but with the T-shirt, Task 3 is confused with Task 4 in 83% of cases. In Task 3 the finger first flattens the fold and keeps sliding over the cloth and over the hem. In Task 4 it just slides and goes over the hem. In both cases, the cloth is clamped on the table, therefore, it cannot move. In other words, 83% of times with the T-shirt, the method cannot feel that the finger is flattening the fold, and it thinks it is just sliding over the cloth. We believe that this is related to the fact that there are two layers of fabric present In a T-shirt, which is reported to provide significantly different tactile feedback by [24].



Ti and Oi refer to Task i and object i, i $\in \{1, 2, 3, 4\}$.

Fig. 7. Classification by task and object, considering all the data together.

3) Task and cloth type identification: Our results so far suggest that even when performing the same task, different types of cloth result in very different tactile signals. In other words, if we want to recognize the task from only the tactile signal, the task definition should include the type of cloth.

In this section, therefore, we analyze the data from all clothes together, this time separating the data into 13 clusters corresponding to task and object pairs. That is, 4 tasks with the 3 different objects plus Task 4 with Object 4, corresponding to a seamless piece of cloth. In this case, for each tested data element, we want to identify both the object and task that has been executed. The confusion matrix can be seen in Fig. 7. Results show there is little confusion across different objects, leading to diagonal sub-matrices with results similar to those seen in the previous case (Fig. 6). We only find one cases where cross-object confusion occurs, 9.379% of the times Task 1 on The kitchen cloth was misidentified as Task 1 on the T-Shirt, and similarly Task 1 on the T-Shirt was misidentified as Task 1 on the kitchen cloth 10.7% of the time.

These results show that, in the case of cloth manipulation, we need to consider both task and type of cloth to be able to recognize what task we are carrying out. It also shows how tasks that are very similar, such as pulling to flatten vs. pulling to move the whole cloth (which we perceived to be difficult to distinguish using tactile information alone), could still be identified, albeit with less precision. Thus, additional sensing information may be useful in identifying them correctly. These are promising results which indicate that taking into account tactile information can greatly improve the current methods of estimating the task and cloth state. However, when we need to identify more tasks and objects, considering both type of cloth and task for classification may not scale well. We plan more experiments with additional types of cloth in the future to study the scalability of those results.

V. IDENTIFYING CONTACT INTERACTION TYPES

In the previous section, the signal is analyzed as a whole. Even if we could classify with full accuracy, we would only be able to recognize the task once it had been fully executed. However, one of the main open challenges of cloth manipulation is to be able to close the loop during execution, that is, to be able to identify what is happening while we manipulate a cloth, and react accordingly. To achieve this, it would be very useful if for each sensor measurement we could identify the type of contact interaction that is occurring between finger, cloth and table.

When considering the closed loop scenario, the robot should be able to react while executing a trajectory. Thus, we cannot wait for all the data to be ready to start classifying. Instead, we interpret the sensor data shown in Fig. 4 as a time series. In this case, for each force sample F_t, we have to classify between the 9 contact interactions defined in Table I.

Data collected has been divided with a distribution of 80% training and 20% test. Despite the results obtained in the previous section, where we identify that it is better to differentiate between different clothes, we consider all the data from the different objects together. That is, for instance, that elements from the class m1a will contain executions with different clothes. The annotation used in the *contact interaction* experiment is richer than the one in the previous section and allows higher generalization capabilities.

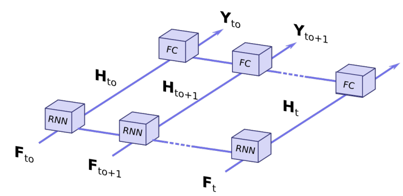


Fig. 8. Identifying contact interactions: network architecture. At each time step the sensor forces F_i are fed to the RNN, which updates the hidden states H_i. The output of the hidden state is used to obtain the current contact interaction Y_i at each time instant.

The network proposed for this purpose consists of an RNN followed with a fully connected layer [40] to classify among the k possible classes defined in Table I. The network proposed for this task is depicted in Fig. 8. For each input force sample, it returns the probabilities of each of the k output categories. Among RNN, we compare Long Short Term Memory (LSTM) [6] and Gated Recurrent Unit (GRU) [41] as both allow learning long term dependencies that regular RNN can not due to the vanishing/exploding gradients problem.

At each time instant a new F_i sample, consisting of the force in the X, Y and Z directions, is generated. This data is introduced into the RNN which classifies the action being taken by the robot according the new sample and the current hidden states H_i. The status of the network at each instant is updated by modifying the hidden states H_i. The output Y_i are generated with a fully connected layer from the H_i. The benefit of using a RNN is that the classification of new samples can be done in real-time. Since the inference time of the net is lower than the elapsed time between sensor measurements the detection of the contact interaction between the cloth and the robot can be done in real-time.

A. Results

Two configurations have been tested as base elements for the RNN: LSTM and GRU units. Each configuration has been tested with a number of hidden units ranging from 10 to 200, we checked empirically that augmenting the number of hidden units doesn't improve the results provided.

The results are evaluated for each sensor measurement, comparing the output of the network with the annotation of the sequence at that time instant. As data has been labeled manually, the transition between contact interactions is marked at a particular time instant. We introduce a time tolerance factor while evaluating the net transitions, if one transition between contact interactions of the net is correct (the net detects correctly the present and future contact interaction) but not aligned perfectly with the labeled data it is considered correct. We measured that this has an impact of between 2% and 3% of improvement in the accuracy of the RNN models.

Accuracy results for multiple network sizes are presented in Table II. The performance of the method achieves over 90% accuracy on unseen data when using more than 50 hidden units for both LSTM and GRU models. Considering that we classify amongst 9 different classes, the results achieved with the RNN allow differentiating between interactions with high confidence. Given the limited data available, results obtained are outstanding and can be improved with a larger set of robot trajectories.

TABLE II

ACCURACY RESULTS FOR MULTIPLE RNN CONFIGURATIONS.

	Training		Test	
Num. hidden units	LSTM	GRU	LSTM	GRU
10	0,787	0,891	0,725	0,835
20	0,876	0,919	0,847	0,895
50	0,934	0,963	0,917	0,934
100	**0,967**	**0,980**	0,945	0,954
200	0,963	0,976	**0,951**	**0,957**

GRU based networks slightly over-perform LSTM ones. The performance of GRU networks ranges from 72,5% accuracy with a small network to 95,7% using 200 hidden units. The accuracy of the net does not increase when adding more hidden units.

LSTM and GRU based networks in this experiment have obtained similar results when using the same number of hidden units, as reported in previous works [42]. In our experiments, GRU units obtain a slightly better performance. Additional results stacking multiple RNN layers have obtained similar results than using only one RNN layer.

To further analyze the RNN results, the best model among the ones in Table II is selected. The Confusion matrix over test data for the RNN model using 200 hidden GRU units is shown in Fig. 9. The contact interactions m1b and m2b are the most difficult to classify for the net, having problems identifying if the finger is flattening a fold or moving the whole cloth. This behavior is consistent across different configurations, indicating that m1b and m2b have similar force sensor patterns.

VI. CONCLUSIONS

In this work, we use tactile information to recognize robotic manipulations of clothing. We study the problem with two classification tasks, in the first task we identify the whole manipulation process while the second identifies the contact interaction that occur at each time step.

For the first classification task, our method could distinguish very well between a cloth that is clamped on the table, i.e., fixed, and a cloth that can move. However, it had more difficulties distinguishing between similar tasks, such

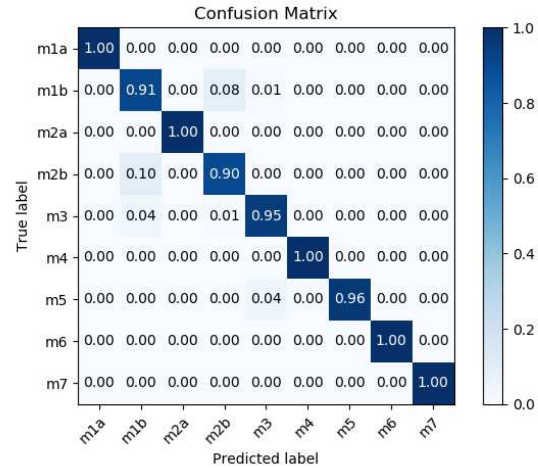


Fig. 9. Confusion matrix in the test data using the GRU network with 100 hidden units. Contact interactions m1b, m2b and m3 are the most difficult to classify for the network.

as moving the entire cloth with a finger, and moving just a part of the cloth to flatten a fold. Still, considering how similar the tasks are, we could achieve a good accuracy in most of the cases when we were also identifying the object. Our results show that we need to consider the fabric type in the task identification, as depending on the type of cloth the tactile feedback is different enough to lead to confusion between tasks. That may be difficult to scale.

For the second classification task, our method is able to identify contact interactions with high accuracy even without knowing the object type. In future work, we will test if the accuracy improves if we identify the object too, as we would need to collect more data for that. Interestingly, our solution is able to identify when the robot is about to lose the cloth (types m4 and m5). This paves the road to enable the robot to react appropriately even before the cloth is actually lost. The robot can also correctly identify if the finger is sliding over a cloth, moving it or flattening a fold. This kind of information can be crucial in helping the robot understand the manipulation it is performing, helping to close the loop in control algorithms, and also provide information to estimate the state of the cloth.

REFERENCES

[1] C. Torras, "Service robots for citizens of the future," *European Review*, vol. 24, no. 1, pp. 17–30, 2016.

[2] A. Ramisa, G. Alenyà, F. Moreno-Noguer, and C. Torras, "Learning rgb-d descriptors of garment parts for informed robot grasping," *Engineering Applications of Artificial Intelligence*, vol. 35, pp. 246–258, 2014.

[3] J. Sanchez, J.-A. Corrales, B.-C. Bouzgarrou, and Y. Mezouar, "Robotic manipulation and sensing of deformable objects in domestic and industrial applications: a survey," *The International Journal of Robotics Research*, vol. 37, no. 7, pp. 688–716, 2018.

[4] H. Yuba, S. Arnold, and K. Yamazaki, "Unfolding of a rectangular cloth from unarranged starting shapes by a dual-armed robot with a mechanism for managing recognition error and uncertainty," *Advanced Robotics*, vol. 31, no. 10, pp. 544–556, 2017.

[5] M. Shokoohi-Yekta, J. Wang, and E. Keogh, "On the non-trivial generalization of dynamic time warping to the multi-dimensional case," in *Proceedings of the 2015 SIAM international conference on data mining*, pp. 289–297, SIAM, 2015.

[6] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, pp. 1735–1780, Nov. 1997.

[7] C. T. P. Jiménez, "Perception of cloth in assistive robotic manipulation tasks," *Natural Computing 19*, p. 409–431, 2020.

[8] I. Garcia-Camacho, M. Lippi, M. C. Welle, H. Yin, R. Antonova, A. Varava, J. Borras, C. Torras, A. Marino, G. Alenyà, and D. Kragic, "Benchmarking bimanual cloth manipulation," *IEEE Robotics and Automation Letters*, vol. 5, pp. 1111–1118, April 2020.

[9] J. Borràs, G. Alenyà, and C. Torras, "A grasping-centered analysis for cloth manipulation," *IEEE Transactions on Robotics*, vol. 36, pp. 924–936, June 2020.

[10] K. Yamazaki, "Grasping point selection on an item of crumpled clothing based on relational shape description," in *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 3123–3128, IEEE, 2014.

[11] J. Maitin-Shepard, M. Cusumano-Towner, J. Lei, and P. Abbeel, "Cloth grasp point detection based on multiple-view geometric cues with application to robotic towel folding," in *IEEE Int. Conf. on Robotics and Automation*, pp. 2308–2315, 2010.

[12] E. Corona, G. Alenyà, T. Gabas, and C. Torras, "Active garment recognition and target grasping point detection using deep learning," *Pattern Recognition*, vol. 74, pp. 629–641, 2018.

[13] A. Doumanoglou, J. Stria, G. Peleka, I. Mariolis, V. Petrik, A. Kargakos, L. Wagner, V. Hlavac, T.-K. Kim, and S. Malassiotis, "Folding clothes autonomously: A complete pipeline," *IEEE Transactions on Robotics*, vol. 32, no. 6, pp. 1461–1478, 2016.

[14] Y. Kita, F. Kanehiro, T. Ueshiba, and N. Kita, "Clothes handling based on recognition by strategic observation," in *2011 11th IEEE-RAS Int. Conf. on Humanoid Robots (Humanoids)*, pp. 53–58, 2011.

[15] M. Cusumano-Towner, A. Singh, S. Miller, J. F. O'Brien, and P. Abbeel, "Bringing clothing into desired configurations with limited perception," in *IEEE Int. Conf. on Robotics and Automation*, pp. 3893–3900, 2011.

[16] Y. Li, Y. Wang, M. Case, S.-F. Chang, and P. K. Allen, "Real-time pose estimation of deformable objects using a volumetric approach," in *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 1046–1052, IEEE, 2014.

[17] Y. Li, Danfei Xu, Yonghao Yue, Y. Wang, S. Chang, E. Grinspun, and P. K. Allen, "Regrasping and unfolding of garments using predictive thin shell modeling," in *2015 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1382–1388, May 2015.

[18] L. Sun, G. Aragon-Camarasa, S. Rogers, and J. P. Siebert, "Accurate garment surface analysis using an active stereo robot head with application to dual-arm flattening," in *2015 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 185–192, May 2015.

[19] S. Miller, J. Van Den Berg, M. Fritz, T. Darrell, K. Goldberg, and P. Abbeel, "A geometric approach to robotic laundry folding," *The International Journal of Robotics Research*, vol. 31, no. 2, pp. 249–267, 2012.

[20] N. Bergström, C. H. Ek, D. Kragic, Y. Yamakawa, T. Senoo, and M. Ishikawa, "On-line learning of temporal state models for flexible objects," in *2012 12th IEEE-RAS International Conference on Humanoid Robots (Humanoids 2012)*, pp. 712–718, IEEE, 2012.

[21] J. Stria, D. Průša, V. Hlaváč, L. Wagner, V. Petrík, P. Krsek, and V. Smutný, "Garment perception and its folding using a dual-arm robot," in *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 61–67, Sep. 2014.

[22] A. Khan, M. H. Tanveer, T. Rasheed, and A. A. Ajmal, "Tactile discrimination of fabrics using machine learning techniques," in *2017 IEEE 3rd International Conference on Engineering Technologies and Social Sciences (ICETSS)*, pp. 1–6, IEEE, 2017.

[23] W. Yuan, Y. Mo, S. Wang, and E. H. Adelson, "Active clothing material perception using tactile sensing and deep learning," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1–8, IEEE, 2018.

[24] F. von Drigalski, M. Gall, S. Cho, M. Ding, J. Takamatsu, T. Ogasawara, and T. Asfour, "Textile identification using fingertip motion and 3d force sensors in an open-source gripper," in *2017 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, pp. 424–429, Dec 2017.

[25] F. V. Drigalski, D. Yoshioka, W. Yamazaki, S. Cho, M. Gall, P. M. U. Eljuri, V. Hoerig, M. Ding, J. Takamatsu, T. Ogasawara, and J. Beltran, "Naist openhand m2s: A versatile two-finger gripper adapted for pulling and tucking textile," in *2017 First IEEE International Conference on Robotic Computing (IRC)*, pp. 117–122, April 2017.

[26] M. J. Thuy-Hong-Loan Le, A. Landini, M. Zoppi, D. Zlatanov, and R. Molfino, "On the development of a specialized flexible gripper for garment handling," *Journal of Automation and Control Engineering*, vol. 1, no. 3, 2013.

[27] S. Denei, P. Maiolino, E. Baglini, and G. Cannata, "On the development of a tactile sensor for fabric manipulation and classification for industrial applications," in *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, pp. 5081–5086, 2015.

[28] C. Kampouris, I. Mariolis, G. Peleka, E. Skartados, A. Kargakos, D. Triantafyllou, and S. Malassiotis, "Multi-sensorial and explorative recognition of garments and their material properties in unconstrained environment," in *2016 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1656–1663, IEEE, 2016.

[29] D. Kruse, R. J. Radke, and J. T. Wen, "Collaborative human-robot manipulation of highly deformable materials," in *IEEE Int. Conf. on Robotics and Automation*, pp. 3782–3787, 2015.

[30] A. X. Lee, H. Lu, A. Gupta, S. Levine, and P. Abbeel, "Learning force-based manipulation of deformable objects from multiple demonstrations," in *IEEE Int. Conf. on Robotics and Automation*, pp. 177–184, 2015.

[31] S. Dong, S. Wang, Y. She, N. Sunil, A. Rodriguez, and E. Adelson, "Cable Manipulation with a Tactile-Reactive Gripper," in *Proceedings of Robotics: Science and Systems*, (Corvalis, Oregon, USA), July 2020.

[32] R. B. Hellman, C. Tekin, M. van der Schaar, and V. J. Santos, "Functional contour-following via haptic perception and reinforcement learning," *IEEE Transactions on Haptics*, vol. 11, no. 1, pp. 61–72, 2018.

[33] I. A. Sucan and S. Chitta, "Moveit!." `http://moveit.ros.org`.

[34] D. Coleman, I. A. Sucan, S. Chitta, and N. Correll, "Reducing the barrier to entry of complex robotic software: a moveit! case study," *CoRR*, vol. abs/1404.3785, 2014.

[35] "Optoforce ros driver." Git, 2018. `https://github.com/shadow-robot/optoforce`, commit 737a15b on branch indigo-devel.

[36] I. Geer, "Complementary video." Youtube, September 2020. `https://youtu.be/CLpRViAU5Jk`.

[37] F. Petitjean, G. Forestier, G. I. Webb, A. E. Nicholson, Y. Chen, and E. Keogh, "Dynamic time warping averaging of time series allows faster and more accurate classification," in *Data Mining (ICDM), 2014 IEEE international conference on data mining*, pp. 470–479, IEEE, 2014.

[38] G. Forestier, F. Petitjean, H. A. Dau, G. I. Webb, and E. Keogh, "Generating synthetic time series to augment sparse datasets," in *Data Mining (ICDM), 2017 IEEE International Conference on*, pp. 865–870, IEEE, 2017.

[39] F. Petitjean, A. Ketterlin, and P. Gançarski, "A global averaging method for dynamic time warping, with applications to clustering," *Pattern Recognition*, vol. 44, no. 3, pp. 678–693, 2011.

[40] G. A. M. Maceira, A. Olivares, "Recurrent neural networks for inferring intentions in shared tasks for industrial collaborative robots," in *29th IEEE International Conference on Robot and Human Interactive Communication*, September 2020.

[41] K. Cho, B. van Merrienboer, Ç. Gülçehre, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," *CoRR*, vol. abs/1406.1078, 2014.

[42] J. Chung, Ç. Gülçehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," *CoRR*, vol. abs/1412.3555, 2014.