A Benchmark for Dual-arm Cloth Manipulation

Irene Garcia-Camacho*[1], Martina Lippi*[2,3], Michael C. Welle[2], Hang Yin[2], Rika Antonova[2], Anastasiia Varava[2], Julia Borras[1], Carme Torras[1], Alessandro Marino[4], Guillem Alenyà[1], Danica Kragic[2]

[1]Institut de Robòtica i Informàtica Industrial, CSIC-UPC, Barcelona, Spain
[2]Robotics, Perception and Learning Lab, EECS at KTH Royal Institute of Technology Stockholm, Sweden
[3]University of Salerno, Salerno, Italy [4]University of Cassino and Southern Lazio, Cassino, Italy

***Abstract*— Cloth manipulation is a challenging task that, despite its importance, has received relatively little attention compared to rigid object manipulation. In this paper, we provide a benchmark for evaluation and comparison of different approaches towards three basic tasks in cloth manipulation: spreading a tablecloth over a table, folding a towel, and dressing. The tasks can be executed on any dual-arm robotic platform and the objects involved in the tasks are standardized and easy to acquire. We provide several complexity levels for each task, and describe the quality measures to evaluate task execution. Furthermore, we provide baseline solutions for all the tasks and evaluate them according to the proposed metrics.**

***Index Terms*— Benchmark, manipulation, deformable object**

I. Introduction

Manipulation of highly deformable objects, such as cloth, is an important area of robotics research that has applications both in industrial scenarios and in domestic environments. Despite its relevance, this research direction has historically received relatively little attention compared to rigid object manipulation. One reason could be that deformable objects are more challenging to model and track; furthermore, rigid manipulation strategies are not directly applicable. Recently, a stronger interest in deformable object manipulation emerged and the survey in [1] presents the latest advances.

In order to effectively evaluate robotics methods, it is beneficial to provide specialized benchmarks. A benchmark is a set of well-defined tasks to be performed in a standardized setup which needs to be easy to reproduce in different robotics laboratories. To the best of our knowledge, there is currently no cloth manipulation benchmark. In this paper, we provide a benchmark that will help assess the capability of a dual-arm robot system for manipulation tasks with cloth-like objects.

To handle high degree of uncertainty about deformable objects' state, perception and manipulation often need to be intertwined. Furthermore, the choice of grasping and regrasping strategies can significantly impact subsequent manipulation. Thus, one challenge in designing a benchmark for cloth manipulation is that different components of a robotic system, such as perception, grasping and manipulation planning, are highly dependent on each other. Therefore, we propose to evaluate the performance of the entire system rather than evaluating perception versus action components separately. We also recognize that for some tasks it is common to treat grasping of the initial target points as a subtask. Hence, we provide a way to evaluate grasp execution, followed by evaluation of the task after the cloth is grasped.

According to [1], most cloth manipulation tasks can be structured as follows: the first step requires the robot to locate the object, possibly identify its configuration, and decide where to grasp. Next, the robot needs to pick up the item and, depending on the desired task, perform a preparatory action (e.g., extending or untangling the object). Finally, the remaining part of the task might consist of folding the object, or putting it on a hanger or a person.

We propose three evaluation tasks that in our opinion form a basis for more complex tasks, such as handling clothes (e.g., spreading and ironing, folding and storing) and dressing a human. ***(i)*** The first task is unfolding a tablecloth and spreading it over a table. This task can be seen as preparatory, e.g spreading on a flat surface for ironing. More generally, the need to spread cloth-like objects, such as bed sheets and tablecloth, is ubiquitous in our everyday life. ***(ii)*** The second task is folding a towel on a table. This is one of the most common tasks in textile manipulation literature, and can be seen as a preparatory action before placing on a shelf or in a box for storage/packaging. Although prior works proposed methods for folding [2–6], these have never been systematically compared or benchmarked. Therefore, we believe a cloth manipulation benchmark should include this task to allow comparing to previous progress in a measurable way. ***(iii)*** The third task consists of fitting the neck of a T-shirt over a 3D printed head. This task is a simplification of a dressing scenario: a basis for more complex tasks like putting a T-shirt or a sweater on a human or mannequin.

We define performance metrics to evaluate a cloth manipulation method which are based on success of the task, execution time, force measures and, if possible, quality of the final result, e.g., we define how a tablecloth should be placed on the table. In this way, the proposed benchmark is well-equipped to distinguish approaches that are likely to perform general cloth manipulation well. Finally, we provide baseline solutions for all the tasks and evaluate them according to the proposed scoring, recording decreasing success rate as the

*These authors contributed equally.

complexity of scenarios increases.

II. Challenges and Related Work

Cloth manipulation can be viewed as a representative motor skill in deformable object manipulation. The large configuration space of cloth objects yields significant representation complexity: a key challenge for both object perception and motion synthesis. A common goal for cloth manipulation is bringing objects to a desired configuration. Hence, flattening and folding constitute an important part of cloth manipulation. To facilitate planning in large spaces, prior works considered various constraints, such as cloth hanging vertically under gravity, or a part of the item laid on a table [4, 6–9]. For trajectory planning, physics-based simulation has been used to help movement prediction [5, 10–12]. For control, prior works used feedback relying on heuristic relations between manipulated and target clothing points, such as a Jacobian approximation with diminishing rigidity [13, 14] or visual servoing [15, 16]. Data-driven approaches have been explored for cloth folding using reinforcement learning [17] and deep neural networks with automatic feature extraction from raw sensory input [18, 19].

Assistive dressing is a particularly challenging cloth manipulation task due to additional complex interactions between clothing and external objects. Because of the heightened modeling complexity, existing works tend to employ automated learning approaches. The study in [20] showed a promising application of reinforcement learning for animating character dressing in simulation. To improve efficiency of learning from real data, previous works used succinct representations like topology ‘coordinates’ [21, 22], or incorporated extra guidance such as force and human preferences [23, 24]. Finally, authors of [25] presented an optimization strategy for learning to put on a sleeve, utilizing simulation to guide optimization.

Despite initial progress, cloth manipulation remains mostly unsolved, with innovative techniques still under development. Our proposed benchmark forms a systematic testbed for the prototypical cloth manipulation tasks, helping to evaluate emerging approaches and to gain technical insights for further improvement. The benchmark includes a range of clothing objects with different sizes to make them flexible for various hardware and target applications. We also incentivize systems with larger mobility and sensing range, by including a task with a large tablecloth object.

III. The Benchmark

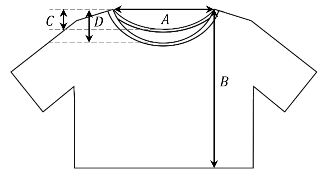
We propose a benchmark for manipulation of highly deformable cloth-like objects which can be performed by any dual-arm setup. Our benchmark considers three bimanual tasks using textile objects of different sizes and types. To foster easy modular use of the benchmark, we enable reporting results for each task separately. Furthermore, we separate each task into sub-tasks that can be evaluated individually, varying the level of difficulty (depending on the initial cloth configuration and involved objects). For the sake of usability and readability, we define three protocols (and respective benchmark documents for scoring) associated with the three tasks as they involve different objects. Protocols (P-DACM-S-0.1, P-DACM-F-0.1, P-DACM-D-0.1 for the three tasks, respectively) and benchmark documents (B-DACM-S-0.1, B-DACM-F-0.1, B-DACM-D-0.1 for the three tasks, respectively) can be found in the attached material and at the link[1].

In the following we give a summary of the benchmark tasks, setup and evaluation. Further information can also be found at the website[1].

TABLE I: List of objects with instructions for acquisition

Code		Objects for manipulation
	Tablecloth	IKEA Fullkomlig 1.45×2.4 m
[st]	Small towel	IKEA towel Hären 0.3×0.5 m or 0.3×0.3 m
[bt]	Big towel	IKEA towel Hären 0.5×1 m or 0.4×0.7 m
	T-Shirt	Any T-shirt in accordance to Fig. 1

Code		Environmental objects
	Table	Any table with dimension in the range Length: $[1.2, 1.85]$ m Width: $[0.7, 0.8]$ m Height: $[0.72, 0.75]$ m
[sh]	Small head	Generate 3D model with provided script
[bh]	Big head	Generate 3D model with provided script



	T-shirt measures
A	[0.13, 0.25] m
B	0.5 m
C	[0.015, 0.05] m
D	[0.07, 0.13] m

Fig. 1: Representation of the allowed measures for the T-shirt; B measure is fixed to equalize the level of difficulty when performing the dressing.

A. Task description

1) Task 1: Spreading a tablecloth: This task consists of grasping a tablecloth and spreading it on a table, using the table and the tablecloth indicated in Table I. Similarly to the other tasks, this task requires to grasp the cloth with the two end effectors and then to manipulate it. For the grasping part, we consider different starting cloth configurations: from folded to crumpled on the table. Grasping crumpled cloth at a desired grasping point has been attempted many times in literature [2, 10, 26]. In contrast, starting from a folded configuration was rarely considered, despite it being a common cloth state in domestic environments. The challenge in unfolding is to grasp just one single corner of the many layers that are folded together. After the first grasping, the cloth needs to be grasped by the second end effector to enable “unfold in the air” configuration. This facilitates the last part of the task: spreading of the cloth on the table. Finally, placing a tablecloth on the table requires controlling the final state of the cloth to make sure the task is finished appropriately. The task requires manipulating a big piece of cloth, which is challenging for many robots and may call for additional strategies. Overall, our protocol does not impose

[1] https://people.kth.se/~mwelle/cloth_benchmark/

a specific strategy. This gives more freedom to researchers to develop and compare innovative approaches.

2) Task 2: Folding a towel: This task consists of grasping a towel and folding it. The task uses the same table as the previous task and two different sizes of towels, as indicated in Table I. This is a classic cloth manipulation task. Since the early example of PR2 robot folding towels in 2010 [2], there have been many other works focusing on folding towels and other items [2, 4, 5, 7, 27, 28]. However, as stated in Section I, this task has never been benchmarked or properly compared based on the quality of the folds or execution time.

The strategy of where to make the folds could vary greatly depending on the garment geometry [6]. Even for a rectangular napkin or towel, there are multiple fold strategies one could follow. However, once the fold line has been decided, we want to focus on finding manipulation that can best realize that fold. For this reason, we focus on the simplest strategy for rectangular items: always fold in half and perform a maximum of three folds which we evaluate individually. This strategy has the advantage that it can be easily evaluated by taking top view snapshots after every fold. Besides the starting crumpled configuration (as in Task 1), we also consider a 'flat on the table' configuration.

When the cloth is crumpled, the main difference for grasping, compared to Task 1, is the size of the object. A small/medium versus large size would entail the need for different strategies to enable initial grasping. Regarding the manipulation to realize the fold, we do not impose a specific folding strategy, assuming it may vary depending on object size or folding stage (first versus second fold). Small towels can be folded using the classic strategy of placing them on a table and picking two corners to fold [2, 5], but bigger ones might require alternatives [9].

3) Task 3: Partial dressing: The goal of the third task is to put a T-shirt over a simple head model starting from different initial configurations of the garment. Putting on the sleeves is not included in the task. The complex geometric shape of T-shirts (e.g., with respect to rectangular towels) makes their manipulation towards desired states a difficult process that requires a tight integration of perceptive sensors, such as cameras and force/torque sensors, into the manipulation strategy. Thus, the focus of this task lies in evaluating the combination of perception and manipulation strategies.

Moreover, analogously to the previous tasks, the success of the manipulation task highly depends on the way the garment is grasped, therefore we consider several initial configurations of the T-shirt which allow to explore different grasping strategies: crumpled, flat or folded on the table. Note that prior works on assistive dressing always assume that the cloth is pre-grasped by the robotic platform.

Another important aspect is the relative size of the head with respect to the collar circumference: the larger the head, the harder it is to execute the task. For this reason, we provide head models of two different sizes as reported in Table I. To avoid damaging the head and the garment, it is desirable to monitor applied forces especially when the collar is tight.

B. Setup description

1) Hardware description: Any dual-arm setup with grasping capabilities can be employed and any sensor that can aid in completing the task is allowed (e.g., RGBD-cameras). Setups equipped with force/torque sensors are preferable for evaluating the proposed method and, in the case of folding, a top camera is preferable for evaluation purposes as well. In this case, we promote the use of vision software to assess the area or the wrinkle state[2] of the towel.

2) Objects description: Table I lists all the objects involved in the tasks with the link or information to acquire them. The YCB object set [29] includes two cloth items, that are a tablecloth and a T-shirt. The YCB tablecloth (in kitchen set) is 2.3×3.3 m and is designed to cover a standard 1.8 m long table until the floor. However, we consider this tablecloth too challenging for the current state of the art. Therefore, we propose a smaller IKEA tablecloth (1.45×2.40 m) which is also common for standard 1.8 m long tables. Regarding the YCB T-shirt (in task set), we observed how even T-shirts from the same batch have a high variance of measures. This also holds true for other available T-shirts. Therefore, we define a range of measures that are accepted for the T-shirt as reported in Figure 1. In this way, greater flexibility is guaranteed compared to the case of a predefined single T-shirt and the possibility of adopting the benchmark is maintained despite continuous changes in fashion. Any size from S to XL of the YCB T-shirt should fall into the allowed range. Note that every T-shirt that is used needs to be measured even if it comes from the same batch to account for production variance. Finally, concerning the towel for Task 2, we include two sizes, a small towel (**[st]**) and a bigger one (**[bt]**). However, the sizes of the big and small towels slightly change depending on the country, therefore, we provide two options for the small towel and two for the big towel. Note that the existing bibliography of cloth folding has dealt either with garments or with small cloth items such as hand towels or napkins. The use of big towels is already a step forward in the literature in terms of object size.

In addition, two environmental objects are required, a table (for Tasks 1 and 2) and a human-like head (for Task 3). Following the idea of flexibility to make the setup easy to reproduce, we do not fix lightning conditions and we do not provide a specific table model but just an interval of table sizes. Two different sizes for the head are considered which are small (**[sh]**) and big (**[bh]**) and their models are defined according to the T-shirt measures. Refer to protocols for further details.

3) Initial cloth configuration descriptions: In general, when we attempt a task on cloth manipulation, the initial state of the cloth falls in one of these categories:

[pg2] Cloth is pre-grasped at two points.
[pg1] Cloth is pre-grasped at one point.
[ft] Cloth is lying flat on a table.

[2] `https://gitlab.iri.upc.edu/labrobotica/algorithms/finddd_descriptor`

[fd] Cloth is folded on a table.

[cr] Cloth is crumpled on a table.

These starting configurations will be common for all the protocols benchmarking each task, although not all starting configurations are used for all task. In detail, it is pointless to consider the flat configuration in the case of spreading task and the folded configuration for the folding task.

We will refer to the parts of the cloth that need to be grasped as grasping points. In a towel, the grasping points are usually the corners; instead, for a T-shirt, these depends strictly on the manipulation strategy.

C. Sub-Tasks description

Depending on the objects and the initial cloth states, each task offers multiple sub-tasks that can be performed independently. More specifically, given a task, the respective sub-tasks are obtained by considering all the possible combinations of involved objects and initial cloth configurations. As a result, 4, 8 and 10 sub-tasks are proposed for the three tasks, respectively: one tablecloth with 4 initial configurations for Task 1, two towels with 4 initial configurations for Task 2 and two head sizes with 5 initial configurations for Task 3. This subdivision allows to achieve sub-tasks of incremental complexity, letting the user choose the desired level of complexity to face, e.g., dressing task with small head and **[pg2]** initial configuration is clearly less challenging than the case of big head with **[fd]** initial configuration.

Each sub-task can be decomposed in the following phases:

[GR1] Grasp first grasping point.

[GR2] Grasp second grasping point with other hand.

[MAN] Perform the manipulation (depending on the task).

Note that both **[GR1]** and **[GR2]** may require manipulation, for instance, in order to grasp a crumpled cloth from a table and reach the first grasping point, the cloth may need to be pre-manipulated, and all this actions constitute the **[GR1]** phase. Obviously, no **[GR1]** and **[GR2]** phases are required in case of starting configuration **[pg2]** as well as no **[GR1]** phase is executed for the initial configuration **[pg1]**.

D. Evaluation of results

To enhance progress, allow reproducibility of results and easy comparison between different works, we propose the following list of performance metrics: *success of each phase*, *execution time*, *force measures* and *quality measures*. The choice to not provide a single value to assess goodness but a set of values is motivated by the fact that, in such complex tasks, the former may be too reductive; in this way, instead, each user can focus on the aspects of interest, e.g solutions that require longer time but exert lower forces.

1) Success of each phase: In light of the phases subdivision in Section III-C, each phase can be evaluated individually in regards to completeness. Phases **[GR1]** and **[GR2]** are considered successfully completed if the grasping is performed and is held during the whole manipulation and, in Tasks 1 and 3, if the cloth is unfolded with starting configuration **[fd]**. The condition of success for phase **[MAN]** depends on the considered task: the tablecloth is successfully spread if it covers the table top; the folded towel is successfully folded if one fold is done and opposing corners are together (each fold is evaluated individually); for the dressing task we assume this phase is accomplished when the neck hole of the T-shirt is put over the head and the entirety of the T-shirt lies below the head. We do not define a measure of quality of grasping as, in general, the choice of grasping points closely depends on the strategy adopted for the cloth manipulation; therefore, defining optimal grasping points would limit the possible approaches in the latter phase. In case the manipulation phase is successful, the grasping phases will be considered successful in turn.

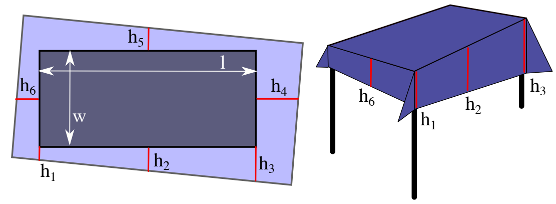


Fig. 2: Representation of the measures to evaluate how well the tablecloth has been placed.

To increase the flexibility of the benchmark and promote participation, we leave the freedom to report only the manipulation part **[MAN]**, which may be the case for end-to-end learning-based approaches, or only the grasping part **[GR1]** and **[GR2]** if a group is strong in grasping but lacks the perception solutions to execute successfully the manipulation. We believe this can be valuable to share solutions and combine different approaches to push the solutions forward.

2) Execution time: The execution time comprises the times needed for all the phases.

3) Force measures: Force measures at the end effectors quantify the interaction between the robots and the environment; they are only acquired during phase **[MAN]** and minimum, maximum and average norms are considered. Note that, in order to not limit the possibility of using the benchmark, force measures are not mandatory but are highly encouraged especially in dressing task, where monitoring of exerted forces on the head represents a key feature.

4) Quality measures: For the tasks of tablecloth spreading and folding, the quality of the result of the execution can be measured, e.g., poor results are achieved if the tablecloth is completely tilted or if towels are folded wrinkly or with the corners not matching. To take that into account, we define a quality function that measures the percentage of error of the task result. Note that for the dressing task, no measures can be defined because of the binary nature of the task.

Quality measures for Task 1

We evaluate how much the tablecloth is rotated and translated with respect to the table. To this aim, as represented in Fig. 2, a total of 6 tablecloth drop lengths at different sides of the table need to be measured after the tablecloth is spread. Measures can be taken from the middle of each table edge. For a table with length t_l and width t_w and a tablecloth with length c_l and width c_w, the proposed percentage of errors are:

$$\begin{aligned} \% \text{ rotation error: } E_\alpha &= \frac{\arctan(\frac{|h_3-h_1|}{t_l})}{\pi/4} \\ \% \text{ length translation error: } E_l &= \frac{|h_6-h_4|}{c_l-t_l} \qquad (1) \\ \% \text{ width translation error: } E_w &= \frac{|h_2-h_5|}{c_w-t_w} \end{aligned}$$

These quality functions can only be applied if the task has been successfully accomplished, meaning that the tablecloth is covering all the table top. Then, a 100% rotation error occurs when the tablecloth is rotated by $\pi/4$ radians (45°), which is unlikely to happen if the tablecloth is fully covering the table. The maximum translation error occurs when one of the hanging parts is zero, meaning the table is almost uncovered. If the hanging part of the table cloth is touching the floor, one needs to measure the tablecloth drop length ignoring the floor. If any circumstance occurs (e.g., one of the hanging parts of the tablecloth is wrinkled), one should report this with a picture, even if it does not affect the quality function. Note that this error measure is independent of the size of the table and tablecloth, thus allowing a fair comparison among different setups.

Quality measures for Task 2

We assume a one fold manipulation is successful if the corners of the original spread cloth are matching two by two. That means if one of the corners is bent, we assume the robot should correct that, otherwise the task cannot be reported as a success. In addition, we will measure how well the corners match by evaluating the ratio between the surface of the spread cloth before and after the fold. Because our task is restricted to folding in half, each fold needs to cut in half the area of the spread cloth on the table. This has to be measured at each fold either by measuring manually the area or by automatically computing it with a top view image.

Then, the proposed quality function for this task is

$$\% \text{ of error in a fold } E_f = \frac{100}{0.5} \cdot \| \frac{A_f}{A_i} - 0.5 \|, \qquad (2)$$

where A_f is the final area of the cloth from the top view, and A_i is the initial area of the cloth. Assuming A_f will always be smaller than A_i, 100% error occurs when $A_i = A_f$, but also if A_f is less than half of A_i, which can only happen if there are wrinkles or extra folds.

E. Reporting results

Based on the above, we require that, for each sub-task, five trials are performed and then, for each trial, measures III-D.1-III-D.4 are acquired. In addition, videos of the experiments and snapshots (or equivalent stylized figures) clearly representing the grasping points must be provided. A summary table, as indicated in Table II, must be filled where, given a starting configuration, the success rate of each phase and average and variance of execution time, force measures and quality functions over the five trials are reported. When necessary, the size of the different elements must be reported as well, that are the table size for Tasks 1 and 2, the towel size for Task 2, and the head size for Task 3. Note that, in the folding task, results associated with each fold must be provided and top view pictures of each fold state have to be reported with a ruler placed nearby. Moreover, in order to assess the generality of the proposed approach, it is required to specify which assumptions are made for completing the task, e.g., knowledge of the cloth color and pattern. Finally, a discussion on:

- Employed hardware/software setup;
- What makes the system successful;
- What makes the system fail;
- What is improved compared to other methods;

should be provided. The detailed description for the scoring of each task can be found in the provided Benchmark sheets[1].

IV. Baseline Systems

To showcase how the presented Benchmark should be used and promote comparison of different methods, we describe our own systems tackling the Benchmark.

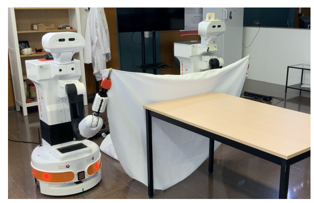


Fig. 3: Experimental setup used for the baseline in Task 1.

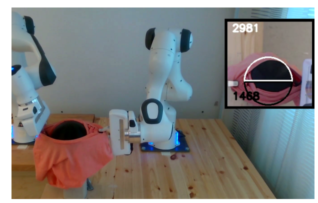


Fig. 4: Experimental setup used for the baseline in Task 3.

A. Task 1: Spreading a tablecloth

The robotic system used for the baseline solution for Task 1 (and also 2) is composed of two TIAGo robots (shown in Fig. 3). The use of these mobile manipulators is convenient because the tablecloth size is large, and thus we can take advantage of the base mobility. The arms are equipped with a modified parallel gripper that is flexible when it touches the table (so we benefit of contacts between table and gripper to simplify the grasping when possible) but rigid in the grasped direction. A table with measures $1.20 \times 0.7 \times 0.73$ m is used.

The solution for **[GR1]** to unfold the tablecloth is to grasp the first grasping point and pull the cloth up. The solution for **[GR2]** is to grasp a point on the edge of the cloth next to the first hand (Fig. 5-right) and trace the edge until the corner at the end of the edge is reached. This implies slipping the cloth inside the gripper without loosing it. This manipulation has been previously applied only to very small clothes in [3], and there are some cloth specialized grippers designed to ease this manipulation [30]. Each grasp is performed by a different TIAGo robot, and after they have the tablecloth grasped, they move across the table to spread the tablecloth on it (see Fig. 3). Note that this strategy is applicable to different sized tables because bi-manual manipulation is achieved with two independent robots.

Our method depends on some simplifications, reported in the Benchmark sheet. We assume the top layer of the folds contains a corner of the tablecloth. This corner, marked in

TABLE II: Result summary tables.

Task 1: Spreading a tablecloth								
Start. Config.	Quality funcs.	Success **[MAN]**	Success **[GR2]**	Success **[GR1]**	Time	Force measures		
						min	avg	max
[pg2]	(E_α, E_l, E_w)							
[pg1]								
[cr]								
[fd]								

Task 3: Partial dressing							
Head Size	**[bh]** or **[sh]**						
Start. Config.	Success **[MAN]**	Success **[GR2]**	Success **[GR1]**	Time	Force measures		
					min	avg	max
[pg2]							
[pg1]							
[cr]							
[ft]							
[fd]							

Task 2: Folding a towel															
Towel size	**[bt]** or **[st]**														
	First fold					Second fold					Third fold				
Start. Config.	Quality func.	Success **[GR1]**	Success **[GR2]**	Success **[MAN]**	Time	Quality func.	Success **[GR1]**	Success **[GR2]**	Success **[MAN]**	Time	Quality func.	Success **[GR1]**	Success **[GR2]**	Success **[MAN]**	Time
[pg2]	E_f					E_f					E_f				
[pg1]															
[cr]															
[ft]															

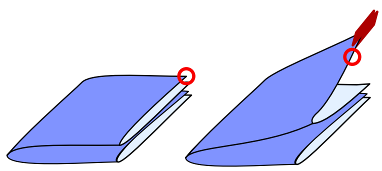


Fig. 5: The circles in red signal the location of the first grasping point for **[GR1]** (left) and on the initial grasp for the edge tracing in **[GR2]**.

the red circle in Fig. 5-left, is the chosen grasping point for the first grasp. We also assume the folded piece is oriented on the table so that the grasping point is the closest to the robot. However, the tablecloth can be placed anywhere on the table. The second grasp in **[GR2]**, shown in Fig. 5-right, is assumed to be at a fixed position with respect to the hand that is already grasping. This holds true for most of the cases, but may fail when the cloth is twisted differently than expected. Also, the robot can not distinguish if it is tracing the short or the long edge of the tablecloth; thus, we fold it in such a way to assure that the short edge will be traced. Finally, the robot knows the size of the cloth and the table, so that we can estimate beforehand the amount of displacement needed when both the robot tracks the edge and puts the tablecloth.

B. Task 2: Folding a towel

For the task of folding a towel, we use the same robotic system as in the previous task and we consider the big towel (**[bt]**) with measures 0.5×1 m. The two mobile manipulators are placed at different sides of the table where the towel is to be folded. We only report the **[MAN]** phase, as this is the most different phase compared to Task 1. Thus, we start with the two corners already grasped, and we then perform the folding motion. For the folding strategy, we focus on the first fold and we use a Dynamic Movement Primitives (DMP) representation of the motion for each robot learned by demonstration, and execute both trajectories in synchronization. The size of the towel and the localization of the robot with respect to the table are assumed to be known.

C. Task 3: Partial dressing

For the dressing task, we propose a human inspired solution based on a vision/force-feedback informed strategy with hand-tuned hyperparameters. All the possible head sizes described in III-B.2 and starting configurations in III-B.3 are considered. Moreover, experiments with two T-shirt samples of different sizes are carried out to show the validity of the protocol as long as the T-shirt complies with the range of measures provided. In detail, the following set of measures $\{A, B, C, D\}$ hold for the two T-shirts, respectively: $\{0.19, 0.5, 0.029, 0.1\}$ m and $\{0.154, 0.5, 0.025, 0.08\}$ m.

The robotic system, shown in Fig. 4, is composed of two Franka Emika Panda 7-DOFs manipulators equipped with parallel grippers. Customized long fingers have been adopted for one robot in order to have the fabric slipping into them during second grasping phase. The head is mounted on a podium stand in the middle of the workspace and its configuration is assumed to be known (in particular, its position is represented by the upper point along the axis of symmetry). Moreover, a Logitech USB camera is mounted over the setup to provide a bird eye view of the workspace and link-side torque sensors at each link of the robots are available; based on these, an estimate of the forces exerted at the end effector of each robot is given. For details on sensors and estimates accuracy, the reader can refer to [31].

Concerning the grasping phases **[GR1]** and **[GR2]**, predefined grasping poses are selected with initial configurations **[pg1]**, **[cr]**, **[ft]** and **[fd]**, thus no visual feedback is exploited in these stages. Concerning the manipulation phase **[MAN]**, the basic idea is to use the visual information to guide the team motion and, at the same time, to perform random wiggling motions which emulate human-like dressing. The formulation in [32] is leveraged for the dual-arm manipulation according to which the cooperative motion is expressed in terms of centroid and formation of the two end effectors. More specifically, a vision system is designed which splits the top-view circle associated with the head into two halves and measures the free area in each of them (see top right of Figure 4). These measures are then exploited to determine

TABLE III: Performance of baseline system for tablecloth spreading task

Task 1: Spreading a tablecloth				
Start. Config.	Success **[MAN]**	Success **[GR2]**	Success **[GR1]**	Mean quality func. (E_α, E_l, E_w)
[pg2]	80%			(1.44%, 8.23%, 17.67%)
[pg1]	60%	80%		(1.61%, 7.50%, 46%)
[cr]	0%	0%	80%	-
[fd]	0%	0%	60%	-

in which direction to move the team centroid in such a way that both areas exceed a certain threshold. When the latter condition is fulfilled, the downward motion to put on the T-shirt is started. In addition, a continuous monitoring of end effector forces and of the elapsed time is performed: a restart procedure is planned when either force measurement exceeds a maximum allowed value or the elapsed time exceeds a time limit. Note that our baseline solution does not involve re-grasping phases but these are generally allowed.

V. Results

In this section, the evaluation of the baseline strategies according to the proposed Benchmark is presented. Videos of each experiment and complete scoresheets can be found in the results section at the website[1]. A summary video is also provided in the accompanying multimedia material.

A. Task 1: Spreading a tablecloth

Performance results of the baseline solution are shown in Table III. In summary, regarding the manipulation (**[pg2]** row), the mobile base of the robots is very effective. Only occasional entanglements of the tablecloth cause strong forces and make the grippers loose the garment.

The second grasp (**[pg1]** row) is quite robust because the strategy of following the edge has proven to be effective: the first interest point is always at the same point under the other robot gripper, and the edge tracing takes advantage of the robot mobile torso to keep a vertical trajectory during as much time as possible. Failures are due to this last edge tracing phase: as the gripper has no force sensors, the edge is sometimes lost at the beginning of the movement.

Finally, in both **[fd]** and **[cr]** initial configurations, the first grasping phase resulted challenging. When **[fd]**, because sometimes several layers are grasped causing the second grasp to fail. The image used to locate the corner is taken from the head of the robot, and the viewpoint and distance make difficult to localize a single garment layer. When **[cr]**, because the friction of the fingers does not allow them to slide gently under the garment. In both cases, grasping fails and the task cannot be completed[3].

B. Task 2: Folding a towel

Performance results of the baseline solution are reported in Table IV. Here we concentrate on the **[MAN]** itself and not on the **[GR1]** and **[GR2]** phases, as results would be similar to Task 1. Starting from initial configuration **[pg2]**, **[MAN]** achieves quite a good success ratio of 80%. The only problem we detected was in the release step at the very end of the manipulation. Our grippers have some friction and the towel, due to its composition, in one iteration remained sticked until the arm moved away creating an unwanted bending. As corners are not matching two and two, and our system is not able to correct this, we evaluated this trial as failure[3].

TABLE IV: Performance of baseline system for towel folding task

Task 2: Folding a towel				
Towel size	**[bt]**			
Start. Config.	Success **[MAN]**	Success **[GR2]**	Success **[GR1]**	Mean quality func. E_f
[pg2]	80%			2.77%

C. Task 3: Partial dressing

Performance results of the baseline solution with one T-shirt sample are summarized in Table V. In detail, success rates for the different phases are reported in case of both small and big heads which make evident the increasing complexity and challenges introduced by the different combinations of starting configurations and head sizes. Indeed, different starting configurations lead to a wide variety of achievable grasping points but our manipulation strategy is only able to deal with a subset of them; in particular, our manipulation strategy assumes that the T-shirt is grasped at two points along the neck collar and that one side of the T-shirt is enclosed in the fingers in such a way to minimize the amount of fabric hanging under them. Indeed, most manipulation failures with starting configurations **[ft]** and **[fd]** are due to the fact that there is too much fabric below the gripping points and the robots are unable to find an opening of the T-shirt to pass it through the head. In addition, blind pre-defined grasping poses make manipulation in **[cr]** configuration generally challenging. Finally, grasping failures are recorded with starting configurations **[fd]** when several layers of the T-shirt are grasped with the first gripper and unfolding does not happen. Table V shows that the dressing task offers a wide range of possible improvements primarily in the manipulation phase but also in the grasping phases as each phase influences the following. For the sake of space, no further results are reported herein but complete scoring sheets and videos of the experiments are available at the link[3], for the other T-shirt sample as well. Finally, it is worth remarking that, even in cases with a 100% success rate, there is still a considerable margin for improvement with respect to execution times and exerted forces that are here not reported for the sake of space.

VI. Conclusions

In this paper, we proposed a benchmark for cloth manipulation. The benchmark contains three representative tasks focusing on bi-manual manipulation; it includes cloth and garment items of various sizes. The benchmark is flexible with respect to the hardware specifications and the types of strategies for solving the tasks. It also provides several options for reporting results. We believe that various

[3]The complete score sheets can be found at `https://people.kth.se/~mwelle/cloth_benchmark/#results`

TABLE V: Performance of baseline system for dressing task

Task 3: Partial dressing						
Head Size	**[sh]**			**[bh]**		
Start. Config.	Success **[MAN]**	Success **[GR2]**	Success **[GR1]**	Success **[MAN]**	Success **[GR2]**	Success **[GR1]**
[pg2]	100%			100%		
[pg1]	100%	100%		80%	100%	
[cr]	40%	80%	80%	20%	60%	60%
[ft]	0%	60%	80%	0%	80%	100%
[fd]	0%	20%	60%	0%	20%	40%

robotics groups would find it easy to use the benchmark for comparing existing works and reporting new results. A simple well-defined object set and the possibility of reporting partial results make this benchmark accessible to researchers targeting different stages of the tasks at various levels of difficulty. Our modular protocols also make the benchmark potentially extendable to other tasks in the future.

The baseline solutions we provide assume a set of simplifications of the problem in the perception, positioning of the elements or manipulation strategies. These assumptions are stated in each benchmark report sheet. We believe the baseline solutions give a valid initial point for comparison and show the increasing level of complexity of the different sub-tasks. Overall, this provides a good start for the research community to push the boundaries on what is possible in cloth manipulation further.

References

[1] 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," *Int. J. Robot. Res.*, vol. 37, no. 7, pp. 688–716, 2018.

[2] 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. Robot. Autom.*, pp. 2308–2315, 2010.

[3] 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," *Adv. Robot.*, vol. 31, no. 10, pp. 544–556, 2017.

[4] 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 Trans. Robot.*, vol. 32, no. 6, pp. 1461–1478, 2016.

[5] Y. Li, Y. Yue, D. Xu, E. Grinspun, and P. K. Allen, "Folding deformable objects using predictive simulation and trajectory optimization," in *IEEE/RSJ Int. Conf. Intell. Robot. Syst.*, pp. 6000–6006, 2015.

[6] S. Miller, J. Van Den Berg, M. Fritz, T. Darrell, K. Goldberg, and P. Abbeel, "A geometric approach to robotic laundry folding," *Int. J. Robot. Res.*, vol. 31, no. 2, 2012.

[7] J. Van Den Berg, S. Miller, K. Goldberg, and P. Abbeel, "Gravity-based robotic cloth folding," in *Algorithmic Foundations of Robotics IX*, pp. 409–424, Springer, 2010.

[8] 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.

[9] K. Lakshmanan, A. Sachdev, Z. Xie, D. Berenson, K. Goldberg, and P. Abbeel, "A constraint-aware motion planning algorithm for robotic folding of clothes," in *Experimental Robotics*, pp. 547–562, 2013.

[10] Y. Li, D. Xu, Y. Yue, Y. Wang, S.-F. Chang, E. Grinspun, and P. K. Allen, "Regrasping and unfolding of garments using predictive thin shell modeling," in *IEEE Int. Conf. Robot. Autom.*, pp. 1382–1388, 2015.

[11] V. Petrìk, V. Smutnỳ, P. Krsek, and V. Hlaváč, "Physics-based model of a rectangular garment for robotic folding," in *IEEE/RSJ Int. Conf. Intell. Robot. Syst.*, pp. 951–956, Oct 2016.

[12] Y. Li, Y. Wang, Y. Yue, D. Xu, M. Case, S.-F. Chang, E. Grinspun, and P. K. Allen, "Model-driven feedforward prediction for manipulation of deformable objects," *IEEE Trans. Autom. Sci. Eng.*, vol. 15, no. 4, pp. 1621–1638, 2018.

[13] D. Berenson, "Manipulation of deformable objects without modeling and simulating deformation," in *IEEE/RSJ Int. Conf. Intell. Robot. Syst.*, pp. 4525–4532, 2013.

[14] M. Ruan, D. Mc Conachie, and D. Berenson, "Accounting for directional rigidity and constraints in control for manipulation of deformable objects without physical simulation," in *IEEE/RSJ Int. Conf. Intell. Robot. Syst.*, 2018.

[15] D. Kruse, R. J. Radke, and J. T. Wen, "Collaborative human-robot manipulation of highly deformable materials," in *IEEE Int. Conf. Robot. Autom.*, 2015.

[16] B. Jia, Z. Hu, J. Pan, and D. Manocha, "Manipulating highly deformable materials using a visual feedback dictionary," in *IEEE Int. Conf. Robot. Autom.*, 2018.

[17] A. Colomé and C. Torras, "Dimensionality reduction for dynamic movement primitives and application to bimanual manipulation of clothes," *IEEE Trans. Robot.*, vol. 34, no. 3, 2018.

[18] D. Tanaka, S. Arnold, and K. Yamazaki, "Emd net: An encode manipulate decode network for cloth manipulation," *IEEE Robot. Autom. Lett.*, vol. 3, pp. 1771–1778, July 2018.

[19] Y. Tsurumine, Y. Cui, E. Uchibe, and T. Matsubara, "Deep reinforcement learning with smooth policy update: Application to robotic cloth manipulation," *Robot. Auton. Syst.*, vol. 112, 2019.

[20] A. Clegg, W. Yu, J. Tan, C. K. Liu, and G. Turk, "Learning to dress: Synthesizing human dressing motion via deep reinforcement learning," in *SIGGRAPH Asia 2018 Technical Papers*, p. 179, ACM, 2018.

[21] E. S. Ho and T. Komura, "Character motion synthesis by topology coordinates," in *Computer Graphics Forum*, pp. 299–308, 2009.

[22] T. Tamei, T. Matsubara, A. Rai, and T. Shibata, "Reinforcement learning of clothing assistance with a dual-arm robot," in *IEEE-RAS Int. Conf. on Humanoid Robots*, pp. 733–738, 2011.

[23] Y. Gao, H. J. Chang, and Y. Demiris, "Iterative path optimisation for personalised dressing assistance using vision and force information," in *IEEE/RSJ Int. Conf. Intell. Robot. Syst.*, pp. 4398–4403, IEEE, 2016.

[24] G. Canal, G. Alenyà, and C. Torras, "Adapting robot task planning to user preferences: an assistive shoe dressing example," *Auton. Robots*, vol. 43, no. 6, pp. 1343–1356, 2019.

[25] A. Kapusta, Z. Erickson, H. M. Clever, W. Yu, C. K. Liu, G. Turk, and C. C. Kemp, "Personalized collaborative plans for robot-assisted dressing via optimization and simulation," *Auton. Robots*, 2019.

[26] C. Bersch, B. Pitzer, and S. Kammel, "Bimanual robotic cloth manipulation for laundry folding," in *IEEE/RSJ Int. Conf. Intell. Robot. Syst.*, pp. 1413–1419, 2011.

[27] 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. Robot. Autom.*, 2011.

[28] 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. Robot. Autom.*, 2015.

[29] B. Calli, A. Walsman, A. Singh, S. Srinivasa, P. Abbeel, and A. M. Dollar, "Benchmarking in manipulation research: Using the yale-CMU-berkeley object and model set," *IEEE Robot. Autom. Mag.*, vol. 22, no. 3, pp. 36–52, 2015.

[30] K. S. M. Sahari, H. Seki, Y. Kamiya, and M. Hikizu, "Edge tracing manipulation of clothes based on different gripper types," *J. Comput. Sci.*, vol. 6, no. 8, pp. 872–879, 2010.

[31] F. Ferraguti, A. Pertosa, C. Secchi, C. Fantuzzi, and M. Bonfè, "A methodology for comparative analysis of collaborative robots for industry 4.0," in *Design, Automation Test in Europe Conference Exhibition*, pp. 1070–1075, March 2019.

[32] P. Chiacchio, S. Chiaverini, and B. Siciliano, "Direct and inverse kinematics for coordinated motion tasks of a two-manipulator system," *J. of dynamic systems, measurement, and control*, vol. 118, no. 4, pp. 691–697, 1996.