2023 IEEE International Conference on Robotics and Automation (ICRA 2023)
May 29 - June 2, 2023. London, UK

Bi-Manual Manipulation of Multi-Component Garments towards Robot-Assisted Dressing

Stelios Kotsovolis and Yiannis Demiris

Abstract— **In this paper, we propose a strategy for robot-assisted dressing with multi-component garments, such as gloves. Most studies in robot-assisted dressing usually experiment with single-component garments, such as sleeves, while multi-component tasks are often approached as sequential single-component problems. In dressing scenarios with more complex garments, robots should estimate the alignment of the human body to the manipulated garments, and revise their dressing strategy. In this paper, we focus on a glove dressing scenario and propose a decision process for selecting dressing action primitives on the different components of the garment, based on a hierarchical representation of the task and a set of environmental conditions. To complement this process, we propose a set of bi-manual control strategies, based on hybrid position, visual, and force feedback, in order to execute the dressing action primitives with the deformable object. The experimental results validate our method, enabling the Baxter robot to dress a mannequin's hand with a gardening glove.**

Fig. 1: The proposed method enables the Baxter robot to dress a mannequin's hand with a glove. Different kinds of alignment action primitives are selected through a high-level rule-based decision process. In the lower level bimanual multisensory control methods are used to align the glove with the hand.

I. Introduction

Dressing is one of the most frequent activities of daily living that people with movement limitations require assistance with, and is also one of the most difficult tasks for caregivers [1, 2]. Robotic solutions for assisted dressing could improve the lives of people with disabilities and provide privacy, independence, and dignity. However, robots are still far from possessing human-like dressing abilities due to the complexity of manipulating deformable objects, the variety of clothes' materials and shapes, and the difficulties of estimating the human body state [3]. In this paper, we study robot-assisted dressing with multiple component garments, and validate our approach on a glove dressing scenario.

Most studies on robot assisted dressing propose methods for single-component dressing tasks, such as the dressing of an arm with a sleeve or a medical gown [3-7]. However, most of the real life dressing scenarios include complex multi-component garments, such as jackets, trousers and gloves. Studies that experiment with such kinds of clothes either approach the problem with predefined sequences of single-limb actions [8-10] or fix the end-effectors on the clothes and simultaneously dress the different parts [11, 12]. However, robots should be able to provide assistance to humans and reason about their actions based on the state of the environment, rather than relying on predefined sequences of actions. Moreover, in order to solve more complex multi-component dressing tasks, such as glove dressing, robots should re-grasp the garment in order to align the different components, as illustrated in Fig. 1, instead of maintaining fixed grasping points. In this paper, we propose a multi-component dressing strategy for providing assistance with glove dressing. To the best of our knowledge, no method currently exists addressing such challenging multi-component problems. More specifically, the paper's contributions are:

- A set of bi-manual action primitives with multisensory inputs (including force and vision) that allow flexible handling of the deformable objects.
- A high level rule-based decision process for selecting alignment action primitives on different components. The conditions of the process are designed based on a hierarchical representation of the dressing task and a set of environmental conditions
- An experimental evaluation with a Baxter humanoid robot. The robot successfully manages to put a gardening glove on the hand of a medical-training mannequin.

II. Related Work

Studies on robot-assisted dressing include methods about manipulation of deformable objects as well as human-robot interaction. Some studies have focused on garment represen-

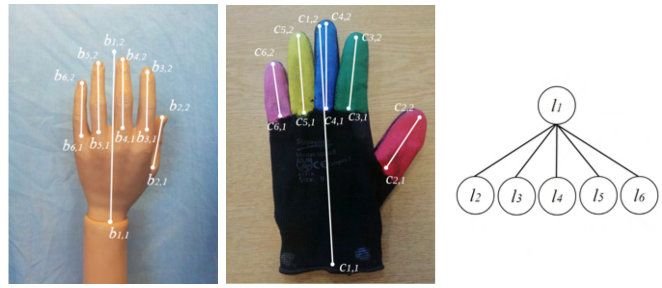
The authors are with the Personal Robotics Lab, Department of Electrical and Electronic Engineering, Imperial College London, London SW7 2BT, U.K. (e-mail:s.kotsovolis21@imperial.ac.uk; y.demiris@imperial.ac.uk). This research is supported in part by UKRI Grant EP/V026682/1, and a Royal Academy of Engineering Chair in Emerging Technologies to YD.

tations [13, 14] and models [15, 16]. Others propose trajectory planning and control methods to bring the garments into a desired configuration around the human body. For example, reinforcement learning has been used for T-shirt dressing [8, 12] and scarf dressing [17], while demonstrations are used in [6] to learn primitive skills. Force information has been included in the optimisation process in many studies in order to avoid large forces on the human body and failure states, where the garments are caught in the human limbs. For example, detected forces are used in [5] to classify the result of a sleeve dressing task, while in [4, 18, 19] forces are used in position-based controllers. In [3] a hierarchical force position controller is proposed to avoid applying large forces on the human limb while putting a sleeveless jacket on a human. In [20] [21] a physics simulator is optimised to train a haptic model, while a haptic model predictive controller is proposed in [7] to dress a human's arm.

While in these studies the users are assumed to remain still, several studies exist for users that possess some ability to move. In [22] the user's pose is tracked to identify the movement's limitations, in [19, 23] the user's movement is modelled to optimise the personalised dressing path, while in [24] task planning is used to adapt to the user's preferences. Some studies focus on the problem of pose estimation under the garment's occlusions for moving users. Different kinds of sensory input have been used to solve this problem, such as force [25], vision [26] and capacitive sensing [27].

The majority of these studies often experiment with dressing tasks of a single component, such as putting on a sleeve. However, a few studies include dressing with more complex garments. The most common approach in these studies is designing predefined sequences of subtasks to dress the different parts. For example, robot self-dressing is achieved in [8] by learning sequential control policies with reinforcement learning. Sequential dressing policies are learnt in simulation in [9] while in later work collaborative human-robot policies are presented in [28]. In [10], a full dressing pipeline designed from smaller subtasks was presented to dress impaired users with back-opening hospital gowns.

Some studies that include multi-component garments do not use sequential approaches, but maintain fixed grasping points and try to simultaneously make progress on the different components of the garment. For example, in [12] a robot learns the trajectory of pulling a T-shirt by maintaining fixed grasping points. However, in this work, the hands are already placed in the sleeves in the initial state. In [11, 29] a trousers dressing scenario is approached by detecting failure states through optical flow and triggering backtracking processes to recover from them. While the results of the aforementioned studies have provided efficient methods for several assisted dressing-related problems, strategies for dressing with multiple components are still insufficiently studied. Our work aims to provide a strategy for multiple component assisted-dressing based on multisensory bi-manual action primitives, by focusing on a glove dressing scenario.



(a) (b) (c)

Fig. 2: (a) Decomposition of a hand into components (b) Decomposition of a glove into components (c) Tree representation for the glove dressing task

III. Problem Statement and Methodology

Assuming a known state of the environment (garment and human), we want to define a decision process that selects actions in order to solve a multi-component dressing task, namely dressing with garments that have multiple openings. In general, actions that include grasping points as well as the motion of robot should be defined for such a task. Moreover, the manipulation strategy should avoid applying large forces resulting from collisions, while it should retrieve from failure scenarios, such as cases where the human body parts are misplaced. Finally, the robot should be able to control the deformable garment. In this section we present our methodology for addressing these challenges. We propose a decision process based on a hierarchical decomposition of the multi-component dressing task and a set of dressing action primitives. In the control level, we focus on a glove dressing task and present bi-manual strategies with multisensory feedback, including force and vision, in order to achieve the execution of the primitives.

A. Hierarchical Task Representation

The first step of the proposed method is a representation of the multi-component dressing task in a hierarchical tree structure. In this representation the human body part and the garment are decomposed into components. We will refer to them as body and garment components throughout the paper. In this work, we assume that the configuration of these components is known throughout the dressing process. For each i of the corresponding garment's and body's components we define a tuple $l_i =< c_{i,1}, c_{i,2}, b_{i,1}, b_{i,2} >$, where $c_{i,1} \in \mathbb{R}^3$ is the centre of the base of the garment's opening of the component i, $c_{i,2} \in \mathbb{R}^3$ is centre of the tip of the opening of the component i, $b_{i,1} \in \mathbb{R}^3$ is the base of the body's component, which is the goal position of $c_{i,1}$, and $b_{i,2} \in \mathbb{R}^3$ is the tip of the component which is defined as the insertion point of $c_{i,1}$. For a glove dressing task the hand's and glove's decomposition is shown in Fig. 2(a) and Fig. 2(b). We note that if the human body part is curved or bent between $b_{i,1}$ and $b_{i,2}$, for example if the human's hand was not wide open, a component can be decomposed into more sub-components. However, we will not address this case in this paper.

Having defined the components of the garment and the human body, we form a tree structure T where each node of the tree has a value of l_i. This way the body and the garment are represented as a collection of body and garment components respectively, that are connected to each other in a hierarchical tree structure based on their topology. The tree for a glove dressing scenario is shown in Fig. 2(c). The root of the tree is the one that emerges from the wrist opening of the glove, while the five branches represent the fingers of the glove.

B. Dressing Action Primitives

Having described the task representation based on the state of the environment we now define a set of high-level action primitives that can be applied to the components of the garment. Here, we try to describe these primitives as generally as possible, by defining each one of them as a certain type of alignment between a garment component and its corresponding body component. We define three primitives as follows:

1) Pulling / Base-to-Base Alignment: This is an action of matching the base of the garment's component to the base of its corresponding body component. This is basically a pulling action of the component's opening towards its goal around the body.

2) Angle Alignment: This is an action of minimising the angle between the component of the garment and its corresponding body component. This is performed by moving the tip of the garment component, while keeping the base fixed. This primitive is designed for states where the garment is caught while pulling due to friction. By aligning the angles of the garment and the body part, the garment is released.

3) Base-to-Tip Alignment: This is an action of matching the base of the garment component to the tip of its corresponding body component. This primitive represents the action of placing a garment's opening around the tip of its target body part.

In our glove dressing scenario, we allow a Base-to-Base and an Angle Alignment primitive for the component l_1. In this work, we assume that the wrist's opening is initially around the fingers and hence we do not include a Base-to-Tip action for the link l_1. For each of the finger components $l_2 - l_5$ we allow a Base-to-Tip alignment action. Specific implementation details for these action primitives are presented in SectionIII-D.

C. Decision Process

Having described the task representation and the action primitives, we present in this section the strategy for selecting them. We define an agent that selects between the dressing action primitives according to the tree representation of the environment T and a set of rules/conditions. The agent iteratively selects actions based on Algorithm 1, until the task is completed. The method prioritises components of higher hierarchy (lower depth) in the tree T. This reflects the instinct strategy that humans follow when dressing other people to start with the base opening of a garment and then follow with the nested and usually smaller parts. For example, when people put on gloves they begin by trying to put on the wrist's opening and then continue with the fingers.

Algorithm 1: Multi-component Action Primitive Selection

Input: T
Output: a, l
Initialisation : $depth = 0$
while $depth < max\_depth$ **do**
  $V = \emptyset$
  for $l\_d \in L(depth)$ **do**
    $a\_d = conditions\_check(l\_d)$
    if $a\_d \neq None$ **then**
      $V = V \cup < a\_d, l\_d >$
    end if
  end for
  if $n(V) = 1$ **then**
    return $a\_d, l\_d \in V$
  else if $n(V) = 0$ **then**
    $depth = depth + 1$
  else
    $a, l = argmax_{<a\_d, l\_d> \in V}(F(a\_d, l\_d))$
    return a, l
  end if
end while

For this reason, the action process begins with iterating in the depths of the tree T in ascending order. For each component l_d in the set of components of the current depth $L(depth)$ of the tree T, a $conditions\_check(\cdot)$ function is performed. This function will return the primitive (if there is any), that can be applied to this component based on the following conditions:

- A Pulling primitive can be applied to a garment component when its opening is around the target body component, the garment is not caught and when no lower hierarchy component is misplaced in a different opening.
- An Angle Alignment primitive can be applied to a garment component when its opening is around the target body component and it is caught and folded.
- A Base-to-Tip primitive can be applied to a garment component when its opening is not around the target body component.

The criterion of whether a garment opening is around a body component, is assessed in our application by checking the intersection of the opening's projection in the 2D image space and its target body component. The criterion of whether the garment is caught is assessed by using a threshold in the end-effector's forces, while the folded criterion is assessed by using a threshold in the angle between the garment's and body's components. Note that, based on these conditions, each time only one of the primitives can be applied to a component.

All the valid action-component pairs $< a\_d, l\_d >$ of the depth of the current iteration's step are added to a set V. If there is only one valid action ($n(V) = 1$) then we select this action. If there is no valid action ($n(V) = 0$), we proceed with the components of the next depth of the tree. Finally, if there is more than one valid action, then we select the action with the highest value of misalignment, which measured by an alignment function $F(\cdot)$. Note that this alignment function is a function used only for actions of the same depth.

In the glove dressing scenario that is studied in this paper, we only have multiple components on the second level of hierarchy, which is composed of fingers. As we only allow Base-to-Tip actions for these components, we define as our alignment function the length of the obstacle-free path from the opening of the glove fingers to the tip of their target hand fingers. The way this path is calculated will be described in Section III-D.

The action search is terminated when the iteration exceeds the depth of the tree T. In that case, no valid action exists, which means that all components of the garment are aligned with their targets and the task is completed.

D. Robot Control

In this section we present the low level implementation for the primitives presented in Section III-B for a glove dressing task. We use bi-manual control strategies for each of the three categories of action primitives.

The Pulling (Base-to-Base) primitive action begins with grasping the base opening of the glove on two opposite sides. After grasping the opening, the two end-effectors move towards the wrist of the hand as depicted in Fig.3(a). The robot end-effectors' motion is defined as linear, constant velocity trajectory in the Cartesian space towards the goal positions, while their orientations are kept fixed. Both arms are controlled according to the following controller:

$$\dot{\mathbf{q}} = \mathbf{J}^{\dagger}(\mathbf{q})\mathbf{K}([\mathbf{P}_d, \mathbf{R}_d]^T - [\mathbf{P}, \mathbf{R}]^T) \qquad (1)$$

where $\dot{\mathbf{q}}$ are the joint velocities, $\mathbf{J}^{\dagger}(\mathbf{q})$ is the pseudoinverse of the robot's Jacobian matrix $\mathbf{J}(\mathbf{q})$ for the joint configuration $\mathbf{q}$, $\mathbf{K}$ is a 6x6 diagonal constant gain matrix, $\mathbf{P}_d$ are the reference waypoints, $\mathbf{R}_d$ is a desired constant orientation and $\mathbf{P}$ and $\mathbf{R}$ is the current position and orientation.

The Angle Alignment Primitive begins with one arm grasping a key point at the wrist's opening and the other arm grasping a key point at the finger's area, opposite to the wrist's opening. As we described in Section III-B the purpose of this primitive is to minimise the angles between the glove's and the hand's wrist component l_1. To achieve this, the arm at the wrist's opening holds the opening still. For the second arm, we use a force-position controller in the local frame of the hand, depicted in Fig.3(b). In the x and y axis of the local frame we want to minimise the distance from the origin, and therefore we set the reference values to zero. In the z axis we use force feedback to stretch the glove, since we do not have a reference position. According to this controller the linear velocity of the end-effector in the local frame is defined as:

$$\mathbf{V}^{local} = \mathbf{K} \begin{bmatrix} p^{local}_{x,d} - p^{local}_{x} \\ p^{local}_{y,d} - p^{local}_{y} \\ f^{local}_{z,d} - f^{local}_{z} \end{bmatrix} \qquad (2)$$

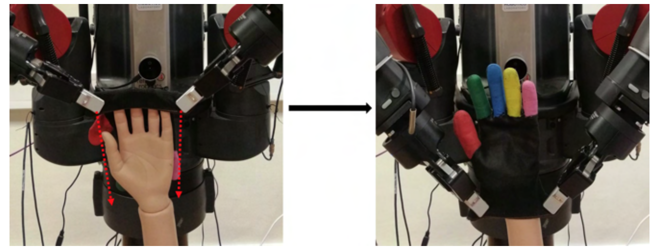
where $\mathbf{K}$ is a 3x3 diagonal constant gain matrix, p^{local}_x and p^{local}_y are the current position of the end-effector in the x and y axis, $p^{local}_{x,d}$ and $p^{local}_{y,d}$ are the reference values that are set to zero, $f^{local}_{z,d}$ is a small constant reference force value in the z-axis and f^{local}_z is the measured force in the z axis. All values are measured in the local frame of the hand. The orientation of the arm is again maintained fixed. After transforming the velocities to the global frame of the robot the joint velocities $\dot{\mathbf{q}}$ can be calculated as:

$$\dot{\mathbf{q}} = \mathbf{J}^{\dagger}(\mathbf{q})\mathcal{V}^{global} \qquad (3)$$

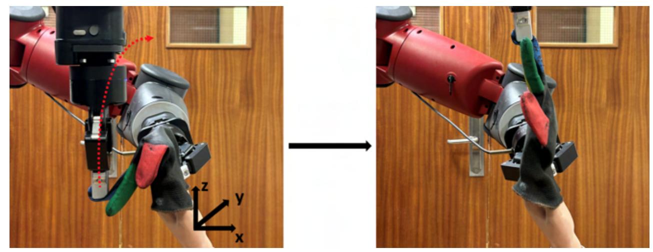
Here $\mathcal{V}^{global}$ is the 6-dimensional (linear and angular) velocity of the end-effector expressed in the global frame of the robot.

For the estimation of the forces and torques applied to the end effector we do not use any external force sensor. We use Baxter robot's forward dynamics model for calculating the end-effector's forces, which uses gravity compensation and spring compensation torques, followed by a deadband filter for noise reduction.

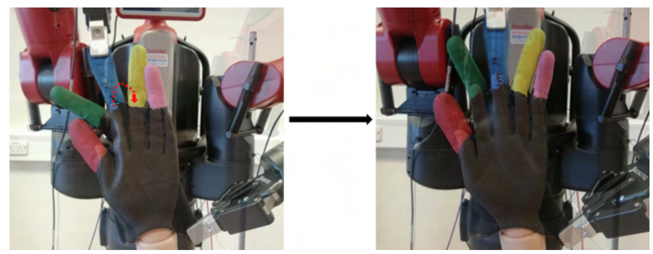
Finally, the Base-to-Tip primitive, which is used to bring the opening of a finger of the glove around the tip of its target human finger, is implemented as follows. We define a feature $\boldsymbol{\xi}$ at the middle of the finger's opening in the 2D image space. Our objective is to drive this feature towards the tip of the matching finger while avoiding collision with the hand, which is particularly important in states were a glove's finger is misplaced around the wrong human finger. Here, we make use of a bidirectional RRT^* [30] path planner in the



(a)



(b)



(c)

Fig. 3: Bi-manual Action Primitives: (a) Pulling / Base-to-Base Alignment (b) Angle Alignment (c) Base-to-Tip Alignment

2D image space to acquire a sequence of pixel-waypoints $\boldsymbol{\xi}_d$. However, directly planning for the feature point would not result in the desired opening's path, as the feature point does not express the geometry and hence the collision points of the opening. For this reason we define an opening object as an object with two not connected points **$\boldsymbol{\xi}$+offset** and **$\boldsymbol{\xi}$-offset** and we track its configuration with $\boldsymbol{\xi}$. Now, we can plan for the path of the opening-object in the 2D image space. A typical way to achieve this would be to plan through the obstacle configuration space of the object. However, for simplicity we alter the planner's collision check step to check for collision with the two points of the opening-object.

The next objective of this primitive is controlling the feature so that it follows the waypoints of the calculated obstacle free path. We propose grasping the glove from two opposing points, one at the base of the glove and one at the tip of the finger we want to align. This is a similar grasping configuration as the one we proposed for the angle alignment primitive. However, inspired by [31, 32] we now allow both robot arms to move according to a hybrid force-vision controller. We use a position-based visual servoing (PBVS) approach for the visual term of the controller, by calculating the position of the feature $\boldsymbol{\xi}$ in the task space. The controller is again designed in the local frame of the hand. The linear velocity of the end effector is calculated from the two errors as:

$$\mathbf{V}^{local} = \mathbf{K_1}(\mathbf{P}_{\xi,d}^{local} - \mathbf{P}_{\xi}^{local}) + \mathbf{K_2}(\mathbf{F}_d^{local} - \mathbf{F}^{local}) \quad (4)$$

where $\mathbf{K_1}$ and $\mathbf{K_2}$ are 3x3 constant diagonal gain matrices, $\mathbf{P}_{\xi,d}^{local}$ are the waypoints of the feature, $\mathbf{P}_{\xi}^{local}$ is the feature's position, $\mathbf{F}_d^{local}$ is the desired force and $\mathbf{F}^{local}$ is the end-effector's force. Again, all values are in the local frame. We use $\mathbf{F}_d^{local} = [0, 0, f_{d,z}^{local}]^T$, where $f_{d,z}^{local}$ is a small value, in order to keep the glove under tension in the z direction and minimise the force in the other directions. Again, the orientation is kept constant. Finally, after transforming the velocity to the robot's global frame the joint velocites can be calculated as in Equation [3]. For simplicity, we use the hand's frame depicted in Fig. 3(b) as a local frame for all the fingers of the hand.

We argue that this proposed dual manipulation technique can achieve moving the tracked feature of the glove into any collision free direction. On the contrary, a single arm manipulation strategy would fail to achieve this because of the deformation of the garment. For example, lets assume we only grasp the fingertip of the glove. Pulling the fingertip would result in moving the feature (base of the fingertip) towards the puling direction, but pushing the fingertip would not result in pushing the feature, as the glove would deform. On this occasion, a pulling action from the other direction would achieve the desired result, which is the point of using a dual manipulation strategy with visual feedback for both arms. Moreover, force feedback is used to maintain the glove under a small tension. This acts as giving to one of the arms the characteristics of a follower, in order to avoid undesirable deformations of the glove and collisions with the human's hand. We note that in contrast with the hybrid controller that is used for the angle alignment, here the vision and force feedback terms act on both hands and in an antagonistic way. For this reason, the tuning of this controller is very important for its functionality.

IV. Experiments and Results

We conduct three sets of experiments to evaluate: 1) The efficiency of the overall proposed method 2) The action selection process between the misaligned fingers 3) The bi-manual force-vision control method for finger alignment.

A. Experimental Setup

Our proposed control method is implemented on the Baxter robot and evaluated on a medical mannequin's hand being dressed with a gardening glove. The hand has fixed joints, but a slight amount of deformation is allowed because of its plastic material. An Intel Realsense L515 camera is mounted on the torso of the Baxter robot. As described in SectionIII-A we assume that the hand's and the glove's pose are known throughout the process. To simulate these conditions we fix the hand in front of the robot and we estimate its pose using the MediaPipe Library [33], before starting the dressing process. To perceive the gloves skeleton and feature points, we paint the glove and use color segmentation.

B. Methodology Evaluation

To evaluate our methodology we run a number of glove dressing experiments and measure the success rate. In the initial state of the experiment the robot holds the glove from the base opening, which is placed approximately around the proximal interphalangeal joint (PIP) of the fingers. We empirically define the threshold forces for detecting failure states of the environment at 5 Newtons. We consider a trial successful when the difference between the points of the skeleton is smaller than a set threshold.

TABLE I. Success Rate (Out of 30 trials) - Our method significantly outperforms the two baselines due to its strategy of targeting specific types of misalignment with the designed action primitives.

	Baseline 1	Baseline 2	Ours
Successful Trials	2	9	23
Success Rate	6,7%	30%	76.7 %

To our knowledge no other approach has been proposed in the literature for robot glove dressing. However, we compare with two baseline approaches:

- Baseline 1: The first baseline is pulling by following a set of waypoints along the hand. We stop a trial and consider it unsuccessful if the force on an end-effector exceeds 15 Newtons. This is just a criterion to mark a trial as unsuccessful and is not taken into account in the baseline to trigger some other function.
- Baseline 2: As a second baseline we adjust the method used in [11] for robot bottom dressing. This study proposes following a main trajectory and backtracking when a failure is detected through optical flow and excessive forces, while maintaining fixed grasping points at the base opening of the garment. Here, instead of optical flow, we

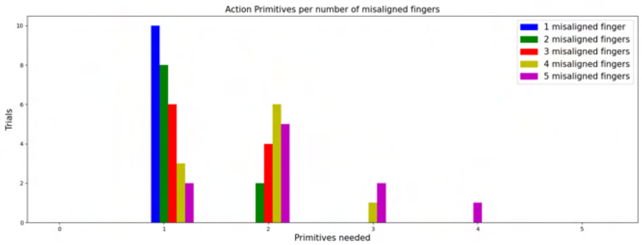


Fig. 4: Number of Base-to-Tip actions needed per number of misaligned fingers out of 10 trials.

use the pulling primitive criteria of the *conditions_check* function (SectionIII-C). We consider a trial unsuccessful if the backtracking process is triggered 10 times.

Table I shows the results of our experiments, consisting of 30 trials for each of the three methods. Baseline 1 presented very low success rates, while it applied large forces on the hand, as it uses force to overcome states where the glove is caught. Baseline 2 achieved a higher performance, as it backtracked from failure states. However, this baseline does not have a strategy for making progress in the task, other than repeatedly trying until it succeeds. Our method outperformed these two baselines, as it specifically targets the detected kind of misalignment with a corrective primitive.

C. Finger Alignment Evaluation

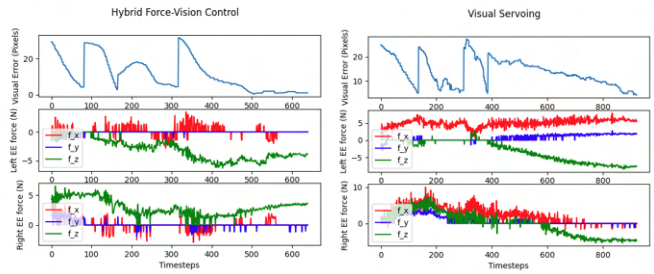
We perform a second set of experiments to measure the effectiveness of the particular part of our proposed decision making process that chooses between misaligned fingers through a distance function as described in SectionIII-C. In order to evaluate the effectiveness of this strategy we manually place a number of fingers of the glove into wrong fingers of the hand. We perform 10 experiments per number of misaligned finger cases and measure the number of Base-to-Tip primitives needed to correct them. Failure cases from incorrect misalignment detection or controller failure were not taken into consideration, as in this experiment we only aim to assess the efficiency of the decision process.

In most of the trials fewer primitives were needed than the number of misaligned fingers, as shown in Fig.4. Moreover, in none of the trials performed more actions were needed the the number of misplaced finger. These results demonstrate that our method for selecting alignment actions for the misplaced fingers of the glove is effective.

D. Bi-Manual Force-Vision Control Evaluation

We conduct a third experiment to evaluate the proposed force-vision controller presented in SectionIII-D for the action of aligning the base of a glove's finger with its target fingertip. We manually place the glove's middle finger around the ring finger of the hand and compare our force-vision controller with a simple vision-feedback controller. We set the desired forces of Equation[4] at $\mathbf{F_d}^{local} = [0,0,4]^T$ for the right arm, that grasps the fingertip and $\mathbf{F_d}^{local} = [0,0,-4]^T$ for the left arm, that grasps the wrist.

We performed 40 trials with each controller. The hybrid force vision controller was successful in 35 trials (87.5%) while the vision controller was successful in 26 trials (65%). When hybrid force and vision feedback was used, the robot managed to drive the opening of the finger through the waypoints of the collision-free path in most of the trials, by minimising the visual error and maintaining the tension between the two end-effectors (Fig.5(a),5(c)). Failure trials of this controller were mostly caused by sub-optimal tuning of the controller. When only vision feedback was used, the controller often could not drive the feature to its goal, due to severe deformations. More significantly, without any force feedback, the robot often collided with the hand and applied large forces on it (Fig.5(b),5(d)).



(a) (b)

(c) (d)

Fig. 5: The visual error (euclidean distance) and the force of the two end-effectors are shown for the hybrid force-vision control (a) and visual servoing (b), while snapshots indicating the behaviour of the two controllers are shown in (c) and (d) respectively. The desired forces of the hybrid controller are set to [0,0,4] and [0,0,-4] for the right and the left arms respectively. The proposed bi-manual force-vision controller manages to maintain the glove under tension and drive the finger's opening around the fingertip of the hand. Without the force feedback the finger insertion task fails and the robot applies large forces on the hand.

V. Conclusions

In this paper, we focused on multi-component robot-assisted dressing and proposed a manipulation strategy for dressing gloves. A high-level decision process was presented, to select an alignment action primitive, while bi-manual control methods were used to implement the high-level actions. Specifically, the challenging problem of aligning the fingers of the glove was approached by a hybrid force-vision bi-manual controller. The experimental results have demonstrated that our approach has enabled the Baxter robot to dress a mannequin's hand with a gardening glove with a success rate of 76.7%. Future work includes extending the method for hands with movable joints and experimenting with real humans.

References

[1] B. J. Dudgeon, J. M. Hoffman, M. A. Ciol, A. Shumway-Cook, K. M. Yorkston, and L. Chan, "Managing activity difficulties at home: a survey of medicare beneficiaries," *Archives of physical medicine and rehabilitation*, vol. 89, no. 7, pp. 1256–1261, 2008.

[2] T. L. Mitzner, T. L. Chen, C. C. Kemp, and W. A. Rogers, "Identifying the potential for robotics to assist older adults in different living environments," *International journal of social robotics*, vol. 6, no. 2, pp. 213–227, 2014.

[3] F. Zhang, A. Cully, and Y. Demiris, "Personalized robot-assisted dressing using user modeling in latent spaces," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2017, pp. 3603–3610.

[4] G. Chance, A. Camilleri, B. Winstone, P. Caleb-Solly, and S. Dogramadzi, "An assistive robot to support dressing-strategies for planning and error handling," in *2016 6th IEEE International Conference on Biomedical Robotics and Biomechatronics (BioRob)*. IEEE, 2016, pp. 774–780.

[5] A. Kapusta, W. Yu, T. Bhattacharjee, C. K. Liu, G. Turk, and C. C. Kemp, "Data-driven haptic perception for robot-assisted dressing," in *2016 25th IEEE international symposium on robot and human interactive communication (RO-MAN)*. IEEE, 2016, pp. 451–458.

[6] E. Pignat and S. Calinon, "Learning adaptive dressing assistance from human demonstration," *Robotics and Autonomous Systems*, vol. 93, pp. 61–75, 2017.

[7] Z. Erickson, H. M. Clever, G. Turk, C. K. Liu, and C. C. Kemp, "Deep haptic model predictive control for robot-assisted dressing," in *2018 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2018, pp. 4437–4444.

[8] T. Matsubara, D. Shinohara, and M. Kidode, "Reinforcement learning of a motor skill for wearing a t-shirt using topology coordinates," *Advanced Robotics*, vol. 27, no. 7, pp. 513–524, 2013.

[9] A. Clegg, W. Yu, J. Tan, C. K. Liu, and G. Turk, "Learning to dress: Synthesizing human dressing motion via deep reinforcement learning," *ACM Transactions on Graphics (TOG)*, vol. 37, no. 6, pp. 1–10, 2018.

[10] F. Zhang and Y. Demiris, "Learning garment manipulation policies toward robot-assisted dressing," *Science robotics*, vol. 7, no. 65, p. eabm6010, 2022.

[11] K. Yamazaki, R. Oya, K. Nagahama, K. Okada, and M. Inaba, "Bottom dressing by a life-sized humanoid robot provided failure detection and recovery functions," in *2014 IEEE/SICE international symposium on system integration*. IEEE, 2014, pp. 564–570.

[12] T. Tamei, T. Matsubara, A. Rai, and T. Shibata, "Reinforcement learning of clothing assistance with a dual-arm robot," in *2011 11th IEEE-RAS International Conference on Humanoid Robots*. IEEE, 2011, pp. 733–738.

[13] N. Koganti, T. Tamei, T. Matsubara, and T. Shibata, "Estimation of human cloth topological relationship using depth sensor for robotic clothing assistance," in *Proceedings of Conference on Advances In Robotics*, 2013, pp. 1–6.

[14] ——, "Real-time estimation of human-cloth topological relationship using depth sensor for robotic clothing assistance," in *The 23rd IEEE international symposium on robot and human interactive communication*. IEEE, 2014, pp. 124–129.

[15] L. Twardon and H. Ritter, "Active boundary component models for robotic dressing assistance," in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2016, pp. 2811–2818.

[16] N. Koganti, T. Tamei, K. Ikeda, and T. Shibata, "Bayesian nonparametric learning of cloth models for real-time state estimation," *IEEE Transactions on Robotics*, vol. 33, no. 4, pp. 916–931, 2017.

[17] A. Colomé, A. Planells, and C. Torras, "A friction-model-based framework for reinforcement learning of robotic tasks in non-rigid environments," in *2015 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2015, pp. 5649–5654.

[18] A. Clegg, W. Yu, Z. Erickson, J. Tan, C. K. Liu, and G. Turk, "Learning to navigate cloth using haptics," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2017, pp. 2799–2805.

[19] Y. Gao, H. J. Chang, and Y. Demiris, "Iterative path optimisation for personalised dressing assistance using vision and force information," in *2016 IEEE/RSJ international conference on intelligent robots and systems (IROS)*. IEEE, 2016, pp. 4398–4403.

[20] W. Yu, A. Kapusta, J. Tan, C. C. Kemp, G. Turk, and C. K. Liu, "Haptic simulation for robot-assisted dressing," in *2017 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2017, pp. 6044–6051.

[21] Z. Erickson, A. Clegg, W. Yu, G. Turk, C. K. Liu, and C. C. Kemp, "What does the person feel? learning to infer applied forces during robot-assisted dressing," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2017, pp. 6058–6065.

[22] S. D. Klee, B. Q. Ferreira, R. Silva, J. P. Costeira, F. S. Melo, and M. Veloso, "Personalized assistance for dressing users," in *International Conference on Social Robotics*. Springer, 2015, pp. 359–369.

[23] Y. Gao, H. J. Chang, and Y. Demiris, "User modelling for personalised dressing assistance by humanoid robots," in *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2015, pp. 1840–1845.

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

[25] F. Zhang, A. Cully, and Y. Demiris, "Probabilistic real-time user posture tracking for personalized robot-assisted dressing," *IEEE Transactions on Robotics*, vol. 35, no. 4, pp. 873–888, 2019.

[26] G. Chance, A. Jevtić, P. Caleb-Solly, G. Alenya, C. Torras, and S. Dogramadzi, ""elbows out"—predictive tracking of partially occluded pose for robot-assisted dressing," *IEEE Robotics and Automation Letters*, vol. 3, no. 4, pp. 3598–3605, 2018.

[27] Z. Erickson, M. Collier, A. Kapusta, and C. C. Kemp, "Tracking human pose during robot-assisted dressing using single-axis capacitive proximity sensing," *IEEE Robotics and Automation Letters*, vol. 3, no. 3, pp. 2245–2252, 2018.

[28] A. Clegg, Z. Erickson, P. Grady, G. Turk, C. C. Kemp, and C. K. Liu, "Learning to collaborate from simulation for robot-assisted dressing," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 2746–2753, 2020.

[29] K. Yamazaki, R. Oya, K. Nagahama, K. Okada, and M. Inaba, "Bottom dressing by a dual-arm robot using a clothing state estimation based on dynamic shape changes," *International Journal of Advanced Robotic Systems*, vol. 13, no. 1, p. 5, 2016.

[30] B. Akgun and M. Stilman, "Sampling heuristics for optimal motion planning in high dimensions," in *2011 IEEE/RSJ international conference on intelligent robots and systems*. IEEE, 2011, pp. 2640–2645.

[31] D. Kruse, R. J. Radke, and J. T. Wen, "Collaborative human-robot manipulation of highly deformable materials," in *2015 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2015, pp. 3782–3787.

[32] D. De Schepper, B. Moyaers, G. Schouterden, K. Kellens, and E. Demeester, "Towards robust human-robot mobile co-manipulation for tasks involving the handling of non-rigid materials using sensor-fused force-torque, and skeleton tracking data," *Procedia CIRP*, vol. 97, pp. 325–330, 2021.

[33] F. Zhang, V. Bazarevsky, A. Vakunov, A. Tkachenka, G. Sung, C.-L. Chang, and M. Grundmann, "Mediapipe hands: On-device real-time hand tracking," *arXiv preprint arXiv:2006.10214*, 2020.