What Does the Person Feel? Learning to Infer Applied Forces During Robot-Assisted Dressing

Zackory Erickson, Alexander Clegg, Wenhao Yu, Greg Turk, C. Karen Liu, and Charles C. Kemp

***Abstract*— During robot-assisted dressing, a robot manipulates a garment in contact with a person's body. Inferring the forces applied to the person's body by the garment might enable a robot to provide more effective assistance and give the robot insight into what the person feels. However, complex mechanics govern the relationship between the robot's end effector and these forces. Using a physics-based simulation and data-driven methods, we demonstrate the feasibility of inferring forces across a person's body using only end effector measurements. Specifically, we present a long short-term memory (LSTM) network that at each time step takes a 9-dimensional input vector of force, torque, and velocity measurements from the robot's end effector and outputs a force map consisting of hundreds of inferred force magnitudes across the person's body. We trained and evaluated LSTMs on two tasks: pulling a hospital gown onto an arm and pulling shorts onto a leg. For both tasks, the LSTMs produced force maps that were similar to ground truth when visualized as heat maps across the limbs. We also evaluated their performance in terms of root-mean-square error. Their performance degraded when the end effector velocity was increased outside the training range, but generalized well to limb rotations. Overall, our results suggest that robots could learn to infer the forces people feel during robot-assisted dressing, although the extent to which this will generalize to the real world remains an open question.**

I. Introduction

Survey data suggests that at least one million people in the United States require daily assistance with activities such as feeding, bathing and dressing [1]. Robotic assistance for these and other activities of daily living (ADLs) offers the potential for improved quality of life by improving independence, privacy and efficiency. In this paper we focus on robot-assisted dressing, where a robot manipulates a garment in contact with a person receiving assistance.

While clothing often appears innocuous, garments can apply high forces to the human body and cause discomfort when pulled taut. The forces applied to the human body are also indicative of key aspects of the dressing task, such as which parts of the body are covered by the garment and whether or not the garment is caught on a part of the body. Directly measuring these forces through wearable sensors or instrumented clothing would be prohibitive. Since dressing is often intended to visually occlude parts of the body, visually estimating these forces would also be challenging. In this paper, we investigate the possibility of a robot inferring these forces using only measurements from its end effector (see Fig. 1). Specifically, our approach uses a 9-dimensional measurement vector, consisting of the 3-dimensional forces and torques applied to the robot's end effector by the cloth and the end effector's 3-dimensional Cartesian velocity.

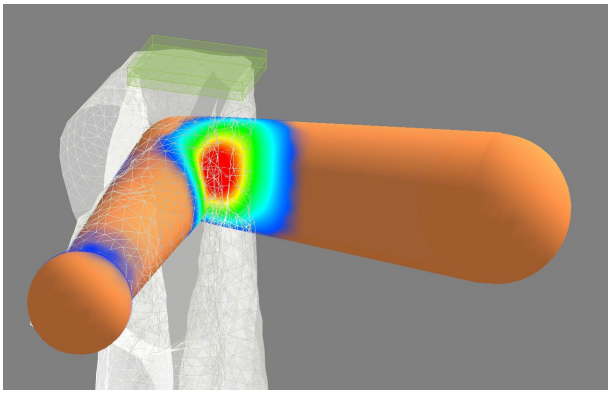

Fig. 1. A force map estimated during a robot-assisted gown dressing task, using only the force, torque, and velocity measured at the end effector (shown in green). The semi-transparent cloth gown allows the force map underneath to be viewed.

Due to the complex mechanics relating the robot's end effector to the forces applied by the garment, we use a data-driven approach. We use a physics-based simulation to synthesize thousands of dressing examples during which we record measurements at the robot's end effector and the forces applied to the human body. We then use these data to train task-specific recurrent neural networks (RNNs) that infer the magnitudes of applied forces on the human body given end effector measurements over time. For example, at each time step, one of our networks estimates a 600-dimensional force map based on a 9-dimensional measurement vector. In this paper we demonstrate that, in spite of the low-dimensional input, high-dimensional output, and complex interactions involved, this inference can be performed in real time with compelling results which bodes well for future applications[1].

II. Related Work

Our approach is inspired in part by the notion that a human can use a lifetime of experience performing a task for him or herself to infer what someone else is feeling when assisted. This is related to work on mirror neurons [2]. The physics-simulation gives robots the opportunity to observe and learn forces applied to the human body during dressing.

Zackory Erickson and Charles C. Kemp are with the Healthcare Robotics Lab, Georgia Institute of Technology, Atlanta, GA., USA.

Alexander Clegg, Wenhao Yu, Greg Turk and C. Karen Liu are with the School of Interactive Computing, Georgia Institute of Technology, Atlanta, GA., USA.

Zackory Erickson is the corresponding author zackory@gatech.edu.

[1]All code can be found at: https://github.com/gt-rad/learning-forces

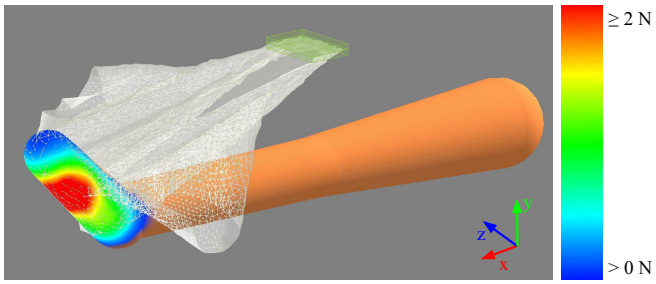

Fig. 2. A full force map estimated along a human leg during a robot-assisted shorts dressing task.

A number of researchers have considered the problem of robot-assisted dressing. Often, the resulting systems have relied on visual data to make inferences about the state of cloth and dressing interaction. Kogatani et al. have proposed techniques for estimating human cloth topological relationships and cloth dynamics from motion capture and RGB-D data [3], [4], [5]. Twardon and Ritter considered RGB-D based cloth boundary detection for garment grasping [6].

Others have used vision systems for robot-assisted dressing control tasks. Tamei et al. used motion capture to extract a low-dimensional cloth state representation based on the topological relationship between a mannequin and the garment. They then computed the arm motion for a dual arm robot to pull a T-shirt over the mannequin's head using reinforcement learning [7]. Klee et al. proposed a goal pose based system for a robot to assist users with dressing tasks. Using a turn taking strategy guided by a vision module, the robot and a user take turns improving their relative positions until a robot goal pose becomes feasible. They demonstrated this system with a Baxter robot assisting a user in donning a hat [8]. Gao et al. propose an approach for which Gaussian mixture models captured from RGB-D data approximate the movement space of a user's upper body joints. They then calculated poses for a robot to hold a garment within reach of a user to facilitate dressing of a sleeveless jacket [9].

While the use of vision as the primary form of input data has been common, the use of force data, if present, has been limited. Yamazaki et al. described an approach for assisted dressing failure detection through optical flow and force data [10]. Recently, Gao et al. have proposed a stochastic path optimization approach for personalized robot-assisted dressing that leverages both force and visual information [11]. Kapusta et al. have also shown that haptic data from a robot's end effector can be used to make predictions about the future state of a dressing task [12].

We use long short-term memory (LSTM) networks to perform estimations with time series data. LSTMs were first introduced by Hochreiter and Schmidhuber, and have since seen several improvements [13]. For instance, we use the forget gate proposed by Gers et al., which allows the network to learn to reset memory over time [14]. LSTMs have been used successfully throughout many applications, including: machine translation, generating cursive writing, and speech recognition [15], [16], [17].

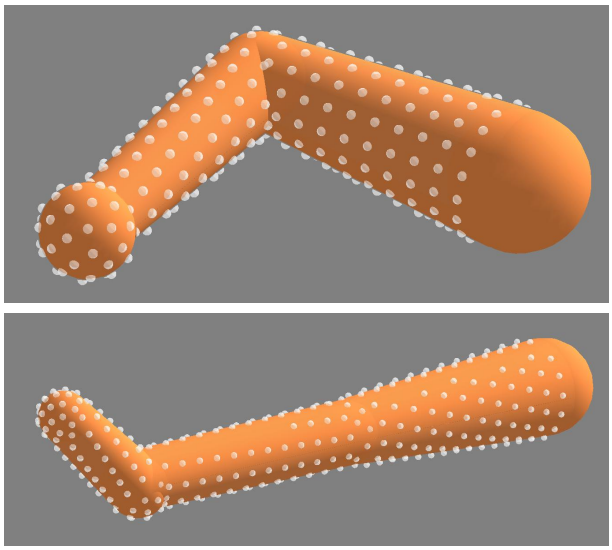

Fig. 3. A force map consists of force magnitudes at discrete points along each limb which are shown here in white. We space these limb points approximately 2cm apart from each other with 300 points along the arm and 400 along the leg. For every time step, the LSTM estimates the magnitude of force applied at each of these points.

TABLE I

MEASURED AND ESTIMATED FORCES DURING SIMULATION AT TIME t.

$f_{t,i}$	Magnitude of true force applied at limb point i
$\boldsymbol{f}^c_{t,j}$	The j^{th} collision force between the cloth and the limb
$\boldsymbol{f}^r_t$	3-dimensional force applied at the robot's end effector
$\hat{f}^u_{t,i}$	Unprocessed force mag. at limb point i, estimated by LSTM

Aviles et al. used dimensionality reduction and an LSTM to visually estimate force applied by a robotic surgical tool to tissue over time. Their algorithm uses stereo images of the tissue and the motion of the surgical tool as input and outputs a single scalar force at each time step [18]. Berenson proposed an iterative Jacobian-based method to manipulate deformable objects without an explicit object model [19]. Similar to Berenson's work, our LSTM-based approach does not have access to a model of the deformable garment, yet is able to learn how this cloth interacts with a human body.

III. FORCE MAP ESTIMATION

In this work, we focus on the problem of estimating the magnitudes of the forces applied to a human limb by a cloth garment during a robot-assisted dressing task (see Fig. 2). We refer to several different forces which we differentiate using superscripts as shown in Table I.

A. Problem Description

We define a force map as an N-dimensional vector of force magnitudes at fixed locations along a simulated limb. We use a set of uniformly distributed points, which we refer to as limb points, to represent these fixed locations (Fig. 3).

We denote this force map at a time step T by $\boldsymbol{f}_T = \left[f_{T,1}, f_{T,2}, \cdots, f_{T,N}\right]$ where $f_{T,i}$ is the magnitude of force at the i^{th} limb point. We expect this force map to be densely

sampled across the limb such that $\boldsymbol{f}_T$ has several hundred dimensions. At time T, there can be a set of collisions between the cloth and the limb where the j^{th} collision consists of a vertex along the cloth mesh with position $\boldsymbol{x}_{T,j}$ that applies a force, $\boldsymbol{f}^c_{T,j}$, onto the limb (Fig. 4). Since these collisions may occur anywhere on the limb, we map the force magnitude of collisions to nearby limb points. We provide a more thorough formulation of this mapping in Sec. IV-B.

In this work, we aim to learn a function that estimates $\boldsymbol{f}_T$ given a sequence of measurements seen thus far. At each time step $t \in \{1, \ldots, T\}$ we use a 9-dimensional measurement vector, $\boldsymbol{\phi}_t$, consisting of the 3-dimensional force, $\boldsymbol{f}^r$, and torque, $\boldsymbol{\tau}$, measured at the robot's end effector along with the 3-dimensional Cartesian velocity, $\boldsymbol{v}$, of the robot's end effector, i.e. $\boldsymbol{\phi}_t = [\boldsymbol{f}^r_t, \boldsymbol{\tau}_t, \boldsymbol{v}_t]$. Thus, we define the estimation problem as finding a function h that accurately estimates $\boldsymbol{f}_T$ given $\boldsymbol{\phi}_{1:T}$, i.e. $\boldsymbol{f}_T \approx h(\boldsymbol{\phi}_{1:T})$.

B. LSTM

We use LSTMs to perform this estimation problem, and hence serve the role of the function h. An LSTM is a type of recurrent neural network (RNN) [13]. We use an LSTM structure similar to that presented by Gers et al. in which the activation of memory cells are defined using three gates: an input gate, a forget gate, and an output gate [14].

One issue that arose as we designed LSTM networks for this estimation problem is a tendency for the LSTMs to estimate a very low, yet nonzero, force for limb points that were not in contact with the garment. This made it challenging to determine where contact was actually occurring along the limb and interfered with visualization of the force maps for evaluation. One approach to account for this issue would be to implement an arbitrary threshold below which all estimated force magnitudes would be set to 0. However, selecting a threshold is not straightforward.

Instead, at each time step t, our LSTM estimates both an unprocessed force map, $\hat{\boldsymbol{f}}^u_t$, and a contact map, $\hat{\boldsymbol{\chi}}_t$. The contact map is used to make a binary classification of whether or not contact is occurring at each limb point. When creating our training set, for each ground truth force, $f_{t,j}$, at limb point j, the accompanying contact map element, $\chi_{t,j}$, within the training set is set such that,

$$\chi_{t,j} = \begin{cases} 0.1 & \text{if } f_{t,j} > 0, \\ -0.1 & \text{otherwise.} \end{cases} \tag{1}$$

In order to avoid bias between learning the contact map versus the force map, we weight the contact map values so they are near the values we expect within the force map (i.e. ~ 0.1). If we set the values within the contact map too large, then the LSTM learns to minimize the large errors within the contact map, but fails to minimize errors that occur in the force map. When the LSTM estimates a contact map, $\hat{\boldsymbol{\chi}}_t$, at time t, we can use this contact map to estimate which limb points are actually experiencing contact with the cloth (i.e. the j^{th} limb point is in contact with the cloth when $\hat{\chi}_{t,j} > 0$).

To estimate the true force map at a time step T, we provide the LSTM with a set of measurements, $\boldsymbol{\phi}_{1:T}$, for all previous time steps. Given $\boldsymbol{\phi}_{1:T}$, the trained LSTM then estimates an unprocessed force map and contact map, $\left[\hat{\boldsymbol{f}}^u_t, \hat{\boldsymbol{\chi}}_t\right]$, for each time step $t = 1, \ldots, T$. When visualizing real time force map estimations, we only consider the estimation for the most recent time step, $\left[\hat{\boldsymbol{f}}^u_T, \hat{\boldsymbol{\chi}}_T\right]$. Finally, the estimated contact map provides intuition into which limb points are experiencing contact and hence we use this information to compute the estimated force map, $\hat{\boldsymbol{f}}_T$. We define the elements of this force map such that for each limb point i,

$$\hat{f}_{T,i} = \begin{cases} 0 & \text{if } \hat{\chi}_{T,i} < 0 \text{ or } \hat{f}^u_{T,i} < 0, \\ \hat{f}^u_{T,i} & \text{otherwise.} \end{cases} \tag{2}$$

$\hat{\boldsymbol{f}}_T$ is then a vector that represents an estimated force magnitude at each limb point with force set to zero for limb points not in contact with the cloth ($\hat{\chi}_{T,i} < 0$) and force magnitudes that were estimated to be negative ($\hat{f}^u_{T,i} < 0$).

Our task-specific LSTMs consist of 3 layers, each with 50 cells. The LSTM layers use a *tanh* activation function and we initialize the LSTM's parameters using Glorot's uniform distribution [20]. The network's output layer is fully connected with linear activations. Altogether, the LSTM has 52,400 recurrent connections. It is worth noting that our approach does not rely on any parsers or encoders, nor does it require any dimensionality reduction since our LSTM directly estimates an entire force map at each time step.

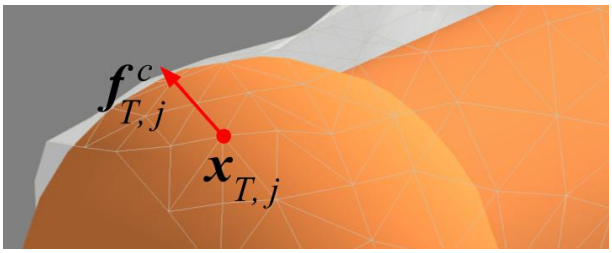

Fig. 4. Visualization of a single point of contact between the cloth gown and the arm. While the cloth is pulled along the arm at time T, the i^{th} collision point has a global position, $\boldsymbol{x}_{T,i}$, and exerts a force, $\boldsymbol{f}^c_{T,i}$, onto the arm.

IV. Physics-Based Simulation

We examine two robot-assisted dressing scenarios: pulling a hospital gown onto an arm and pulling a pair of shorts onto a leg, which we refer to as the gown task and shorts task, respectively. Our simulator extends work by Yu et al. in which they designed a simulator to model real world data from a robot pulling a hospital gown onto human participant's arms [21]. Each simulation was built in OpenGL with NVIDIA PhysX[2] for physics and cloth simulation. PhysX is robust to large forces applied to the cloth and is based on position-based dynamics. We made several modifications to the base PhysX software, including: calculating collision forces, improving friction calculations to better match real world data, and implementing friction for cloth self-collisions

[2]NVIDIA PhysX: `https://developer.nvidia.com/physx-sdk`

to allow wrinkles and folds within the garments. A more thorough account of these modifications can be found in [21].

Simulating the 6-dimensional force and torque measurements relies on accurately modeling complex cloth dynamics and collisions. Because of this, we optimized the simulator using CMA-ES according to data collected from a real robot-assisted dressing task with human participants, as described in Yu et al. [21], [22]. We model the human arm and leg as spheres connected by conical frustums which allow for fast collision detection. We dimensioned the human limbs according to a 50% male. The limbs are rigid bodies fixed in place throughout a trial. We selected limb poses such that the robot must apply a reasonable amount of force in order to successfully perform the dressing task. The garments are represented as triangle meshes based on measurements of a real hospital gown and pair of shorts.

A. Task Variation

We incorporated variability into the garment's initialization by applying a small uniform force to each vertex along the garment for a short duration of time prior to a trial. At each time step t, we collect the measurements, $\boldsymbol{\phi}_t$, along with all ground truth collision positions $\boldsymbol{x}_t$ and force magnitudes $\boldsymbol{f}_t^c$ that occur between the cloth and the simulated body.

When recording training data for the gown task, we added variations that we expect in a real world robot-assisted dressing task. For instance, the initial location of the human's limb relative to the robot's end effector could vary for each dressing attempt. Additionally, the robot's end effector may vary in velocity and follow a nonlinear path while providing dressing assistance. To this end, between sequences, we varied the human limb location, the velocity of the end effector's motion, and the path that the end effector follows, for both dressing tasks. To vary the human arm location with respect to the robot's end effector we centered the human arm at the origin and then randomly varied its position ± 5 cm along the x axis, ± 20 cm along the y axis, and ± 10 cm along the z axis (see Fig. 5 for axes), which we denote by $(\pm 5\text{cm}, \pm 20\text{cm}, \pm 10\text{cm})$. We selected these position ranges such that the cloth and human limb would be guaranteed to make contact during the simulation.

We initialized the robot's end effector at a fixed location above the arm and had it follow a nonlinear randomized cubic Hermite spline path (as shown in Fig. 5) with 7 to 10 control points that deviated from a linear path by $(\pm 3\text{cm}, \pm 2\text{cm}, \pm 5\text{cm})$. Moreover, while the end effector maintained a constant speed during a dressing sequence, we varied the speed from 10 cm/s to 15 cm/s between dressing sequences. It is worth noting that while the magnitude of the end effector's velocity remains constant over time, the three individual components of the velocity do change as the end effector progresses through the spline path. The shorts dressing task varies in a similar fashion with the leg fluctuating from the origin by $(\pm 5\text{cm}, \pm 7.5\text{cm}, \pm 5\text{cm})$, the spline path varying by $(\pm 3\text{cm}, \pm 2\text{cm}, \pm 5\text{cm})$ with 9 to 11 control points, and the end effector speed varying between 15 cm/s and 20 cm/s. Due to the 45° angle of the foot, the end effector's unvaried path drops down closer to the leg when passing over the foot, as seen in Fig. 5.

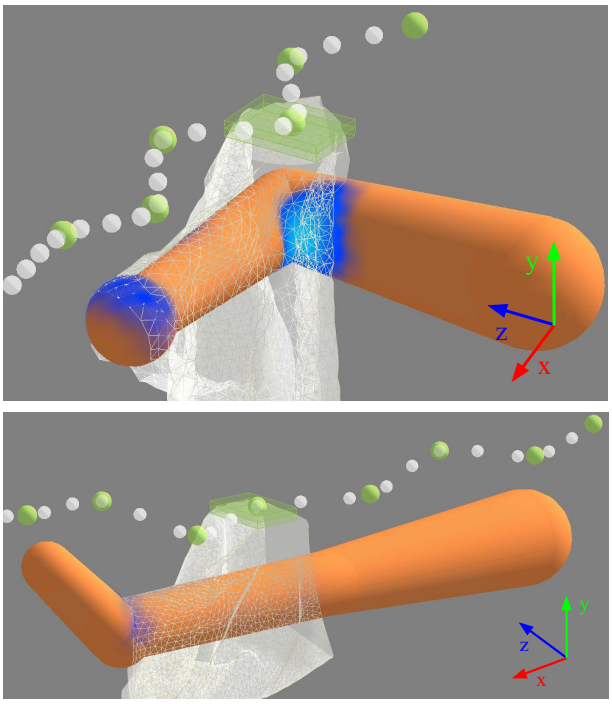

Fig. 5. Example of a random spline trajectory followed by the end effector during both a gown and shorts dressing task. The green control points define the spline trajectory and white points show samples on the interpolated path.

B. Generating the Ground Truth Force Map

As described in Sec. III-A, we define a force map as a collection of force magnitudes at uniformly distributed limb points. The collisions between the cloth and limb will not align perfectly with our set of limb points as these collisions can occur at any location along the limb. As such, our approach proportionally distributes the force magnitude of each collision to the k nearest limb points. At the current time step T, it identifies the k nearest limb points to the j^{th} collision location, $\boldsymbol{x}_{T,j}$, and distributes the magnitude of collision force $|\boldsymbol{f}_{T,j}^c|$ to them. Specifically, it adds $\frac{1}{k-1}\left(1 - \frac{d_l}{\sum_{p=1}^{k} d_p}\right)|\boldsymbol{f}_{T,j}^c|$ to the magnitude of the l^{th} nearest limb point, where d_l is the Euclidean distance of the l^{th} limb point from $\boldsymbol{x}_{T,j}$ and $\sum_{l=1}^{k} \frac{1}{k-1}\left(1 - \frac{d_l}{\sum_{p=1}^{k} d_p}\right) = 1$.

For this work, we used $k = 5$ as we found that mapping a collision force to the nearest five limb points created a smoother force map and improved the LSTMs' performance.

V. Evaluation

For each task, we collected a training set consisting of 5000 dressing sequences and a test set of 1000 sequences. These sequences were randomized in limb position, end effector speed, and end effector path, as presented in Sec. IV.

We used Keras, a neural network library, to implement an LSTM for each dressing task [23]. We trained the LSTMs with stochastic gradient descent and adaptive learning rates

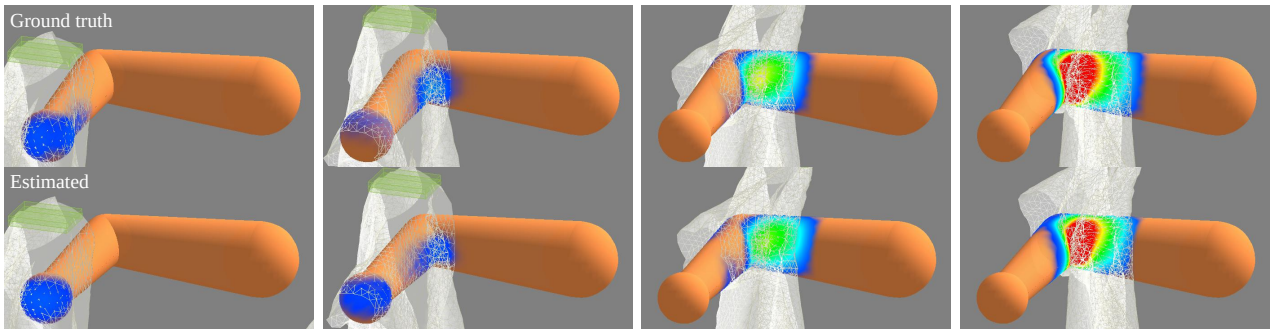

Fig. 6. A sequence for the simulated gown task where the top arm represents the ground truth force map and the bottom arm depicts the LSTM estimated force map. The end effector used a fixed speed of 15 cm/s with an arm position offset of $(0\text{cm}, 20\text{cm}, 0\text{cm})$.

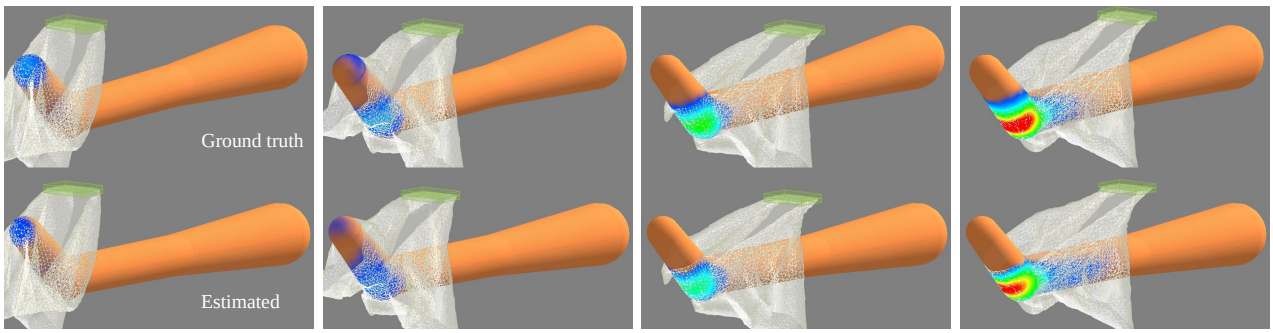

Fig. 7. A sequence for the shorts task with the top and bottom legs showing the ground truth and estimated force maps respectively. The end effector's speed was 20 cm/s with a leg position offset of $(0\text{cm}, 4.5\text{cm}, 0\text{cm})$.

using RMSprop and a mean squared error loss function [24]. We used batches of 16 sequences and trained for a total of 10 epochs. We performed all training on an Amazon EC2 server with 36 cores and 64 GB of memory, taking approximately 2.5 hours to train each of the two LSTMs.

As defined in Sec. III-B, the LSTM estimates both an unprocessed force map and corresponding contact map at each time step, which are used to produce the estimated force map. This amounts to estimating a 600-dimensional output vector for the gown task and an 800-dimensional vector for the shorts task at each time step.

Below, we describe how the LSTM performed on data which are similar to that collected for training and compare differences among the two tasks. We then analyze how our LSTM-based approach generalizes to unseen scenarios, such as higher end effector velocities, or rotations of the human limb. Finally, we present qualitative (visual) results of the estimated force maps in various scenarios, and present a few cases in which the estimated force maps deviate from the ground truth.

A. Visualizations

For both tasks, the LSTMs produced force maps that were noticeably similar to ground truth when visualized as heat maps across the limbs. Specifically, we visualized force maps by mapping each force magnitude to a constant color spectrum, as shown in Fig. 2, and interpolating colors between points along the limb. Within this paper, we present frames from a number of representative trials to help convey the performance of our approach. We also have generated videos that accompany this paper.

The arm rarely gets caught in the gown sleeve as the arm either successfully enters the sleeve, or misses the sleeve for most dressing attempts. Contrary to this, the shorts frequently get caught on the foot during the shorts dressing task, which causes large forces as the end effector continues to pull.

Fig. 6 shows a sequence of frames for a hospital gown dressing scenario in which the gown is successfully pulled up the forearm. Low forces are applied to the fist and forearm initially, while larger forces are applied to the elbow and upper arm as the task progresses. Fig. 7 shows a similar sequence for the shorts dressing task. However, the shorts get caught on the heel of the leg early on, which causes large forces to be applied as the end effector continues to pull and stretch the cloth. Fig. 8 visualizes a scenario in which the arm gets caught in the sleeve of the gown. Due to the large size of the gown, larger forces are still not applied until the latter end of the sequence. Fig. 9 depicts a sequence in which the shorts are successfully pulled up the leg without getting caught. Fig. 10 shows a sequence during which the arm misses the sleeve opening of the gown. Lastly, Fig. 11 presents a sequence for which a higher speed of the robot's end effector led to errors in the estimated force maps.

B. Root-Mean-Square Error (RMSE)

In addition to visualization, we computed a root-mean-square error (RMSE) metric between the estimated and ground truth force maps, averaged across every time step

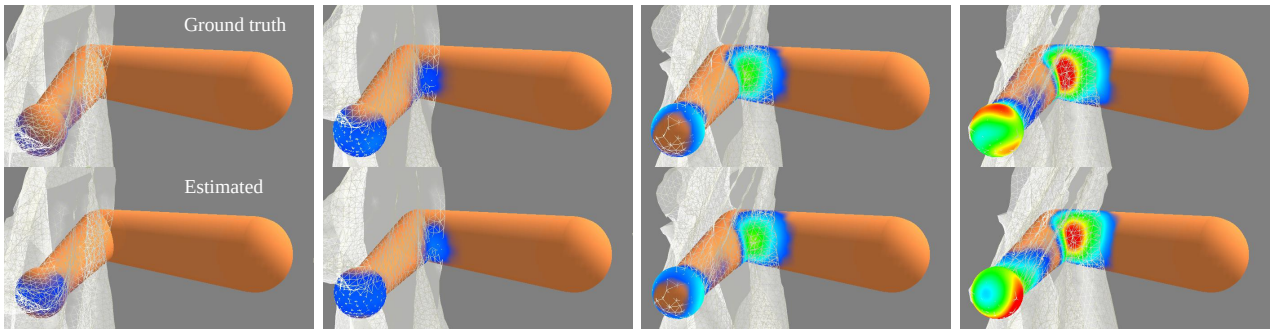

Fig. 8. A visual sequence of the gown task for which the sleeve of the gown gets caught on the fist. The arm is positioned at an offset of $(0\text{cm}, 11\text{cm}, 0\text{cm})$ and the end effector has a speed of 20 cm/s which is greater than the range of speeds used to train the LSTM.

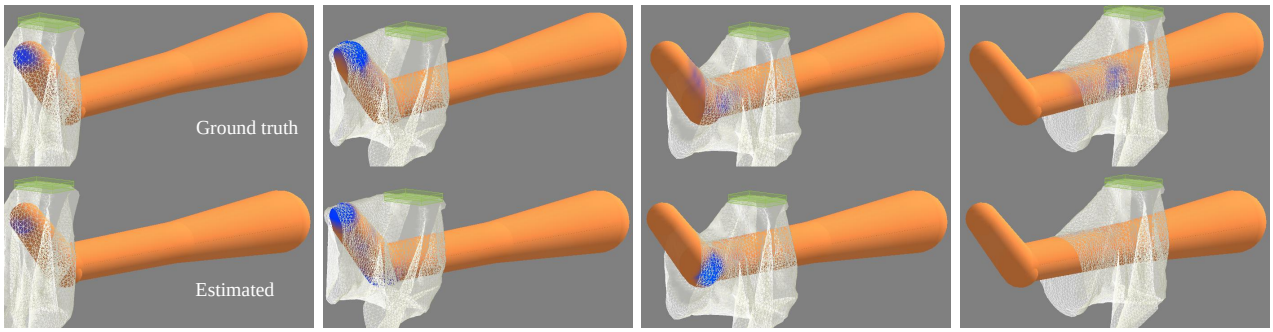

Fig. 9. A shorts dressing sequence for which the shorts are successfully pulled over the foot and up the leg without getting caught. We use an end effector speed of 15 cm/s and a leg position offset of $(0\text{cm}, 7.5\text{cm}, 0\text{cm})$.

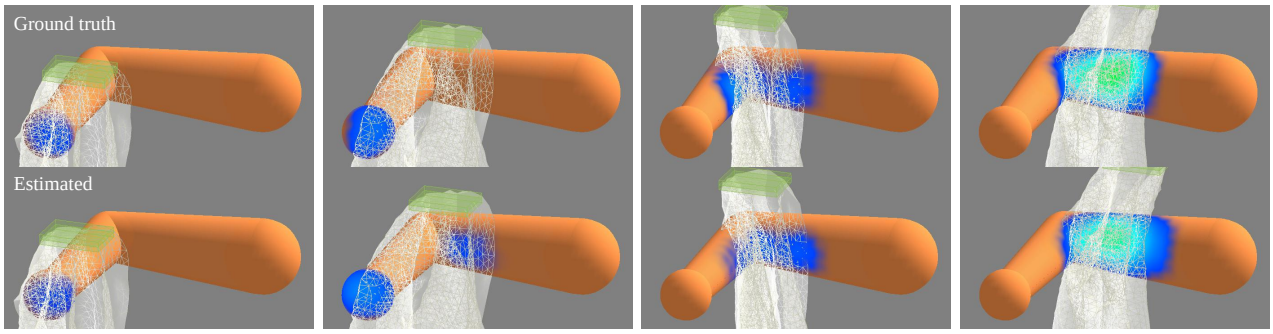

Fig. 10. A simulated gown task sequence for which the arm misses the sleeve of the gown. We set the end effector speed to 15 cm/s and arm position offset to $(0\text{cm}, 20\text{cm}, 7.5\text{cm})$.

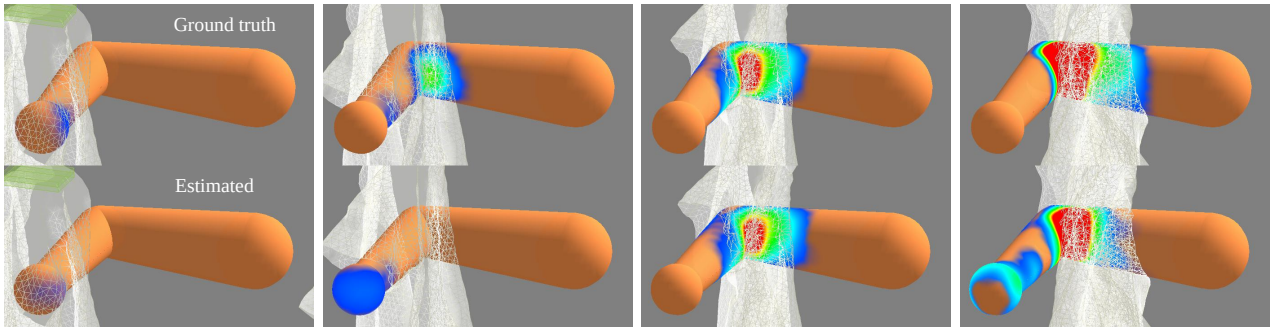

Fig. 11. A gown task sequence which visualizes a scenario in which, for a few time steps, the estimated force maps differ from the ground truth force maps. We use an end effector speed of 20 cm/s which is outside the range of velocity magnitudes that the LSTM has been trained on. The arm is positioned at an offset of $(-5\text{cm}, 15\text{cm}, 0\text{cm})$.

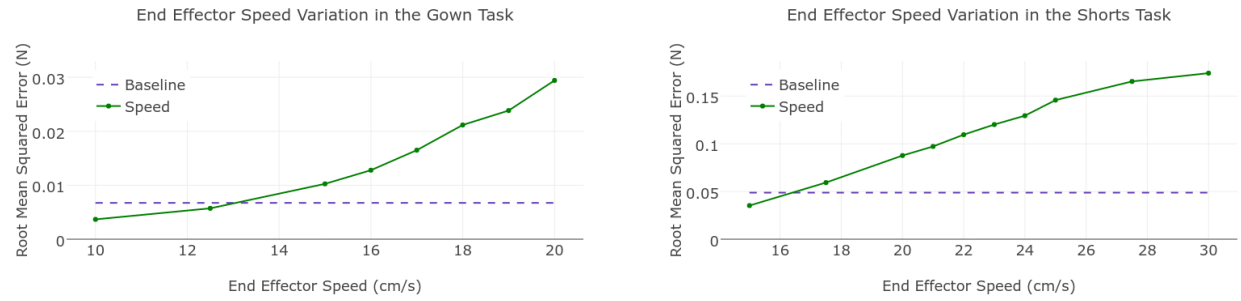

Fig. 12. The figure shows the RMSE between the estimated force maps, $\hat{\boldsymbol{f}}_t$, and the ground truth force maps, $\boldsymbol{f}_t$, for several speeds beyond what the LSTM has seen. RMSE values are averaged over 128 sequences (N=128) with varied limb positions and spline trajectories.

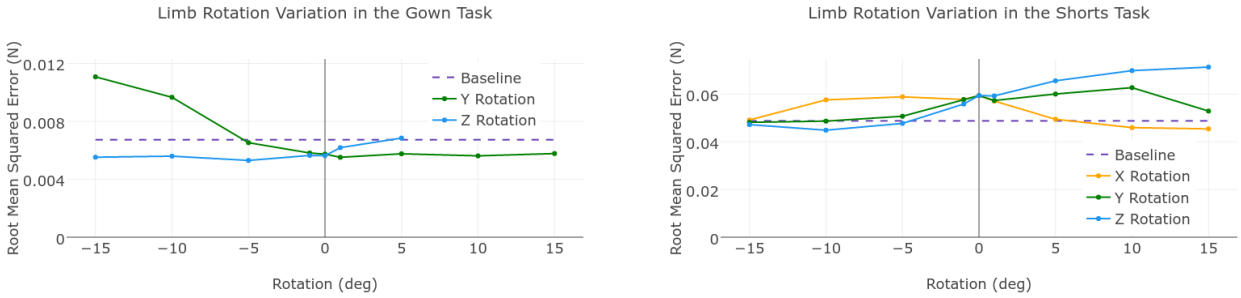

Fig. 13. The RMSE between the estimated force maps, $\hat{\boldsymbol{f}}_t$, and the ground truth force maps, $\boldsymbol{f}_t$, for numerous rotations of the simulated forearm rotated around the elbow and foot rotated around the ankle. For the gown task, we limit rotations along the positive Z-axis to prevent collisions between the arm and end effector. Additionally, the forearm is parallel to the X-axis. As such, we do not rotate around this axis as it would not impact the arm's pose. RMSE values are averaged over 128 sequences (N=128) with varied limb positions and spline trajectories.

for all 1000 test sequences. We compute this RMSE as,

$$\text{RMSE} = \sqrt{\frac{1}{TN}\sum_t^T \sum_i^N (f_{t,i} - \hat{f}_{t,i})^2}$$

where T is the total number of time steps in a sequence, N is the number of limb points, and $f_{t,i}$ is the force for limb point i at time t. This metric provides the benefit of presenting results in units of Newtons and is motivated by the use of MSE as a loss function for the LSTM. Note that we only compute the RMSE for errors between the estimated force maps, $\hat{\boldsymbol{f}}_t$, and ground truth force maps, $\boldsymbol{f}_t$, and not between the contact maps. For the gown dressing task, we computed an average RMSE value of 0.00673 N over all test sequences. We also computed an average RMSE value of 0.0488 N for the shorts dressing task. We use these values as a baseline for evaluating the LSTM's performance in generalizing to unseen scenarios in the following section. For comparison, both of these baseline values approximately double to 0.0148 N and 0.102 N, respectively, if we use a zero collision force map estimation for all test sequences, i.e. $\hat{f}_{t,i} = 0$.

C. Generalization to Unseen Scenarios

While the LSTM performs well with scenarios similar to the training set, we are also interested in how well our approach generalizes since real world dressing will vary in numerous ways. To this end, we compare the RMSE scores between estimated and ground truth force maps for faster end effector velocities and for when the forearm and foot are rotated. Note that the LSTM has never seen a limb rotation, hence the baseline is denoted by a rotation of $0°$.

Fig. 12 highlights how the RMSE varies as we increase the end effector speed for both the gown and shorts tasks. To maintain consistency within each dressing task, we tested each end effector speed on the same 128 randomly generated sequences that varied in limb position and spline trajectory.

While training for the gown task, we used end effector speeds that ranged from 10 cm/s to 15 cm/s. With a speed of 20 cm/s, the RMSE of 0.0294 N is approximately 4.4 times larger than the baseline RMSE for the gown task. For comparison, the dressing sequence in Fig. 6 has an RMSE value of 0.019 N. For the shorts dressing task, we trained the LSTM on speeds ranging from 15 cm/s to 20 cm/s. At 25 cm/s, the RMSE is ~ 3 times larger than the baseline, and this rises to a factor of 3.6 with a speed of 30 cm/s.

Analogous to high end effector speeds, Fig. 13 depicts this same metric for several forearm and foot rotations. Each rotation was tested on the same 128 randomly generated sequences with varied limb positions and spline trajectories. However, we used an end effector speed of 12.5 cm/s and 17.5 cm/s for the gown and shorts tasks respectively. When the forearm is rotated $-15°$ around the Y-axis the RMSE increases to 1.65 times the gown task baseline, whereas a rotation of $-15°$ around the Z-axis generates a RMSE of 0.82 times the baseline. Finally, all of the leg rotations across all 3 axes presented for the shorts task ranged between a factor of 0.92 and 1.46 times the shorts task baseline.

VI. Discussion

In Fig. 12, we observe large error increases as end effector speeds continue to extend beyond the range of speeds in the

training set. This provides some indication that an LSTM-based approach may experience challenges estimating task progression when end effector speeds exceed those used for training. However, this increased error may also be due in part to higher contact forces from faster velocities.

Nevertheless, the LSTMs' RMSE performance generalized well to new limb rotations, despite not training on any rotations. In Fig. 13, we observe a decrease in RMSE for some rotations. This could be influenced by some rotations reducing the magnitude of collision forces between the garment and limb. For example, the foot flattens with negative rotations along the Z-axis which lowers the chance of the shorts getting caught and exerting high forces on the foot.

We note that further work is necessary to evaluate how well real robots can use this approach to infer the forces a person feels during assistive tasks.

VII. Conclusion

Inferring the forces applied to a person's body by a garment during robot-assisted dressing could give a robot insight into what a person physically feels and enable the robot to provide more effective assistance. In this work, we showed how task-specific LSTMs can estimate force magnitudes along a human limb for two simulated dressing tasks. At each time step our LSTM networks take a 9-dimensional input vector consisting of the force and torque applied to the end effector by the garment and the velocity of the end effector. The networks then output a force map at each time step consisting of hundreds of inferred force magnitudes across the person's body.

For both dressing tasks, our approach produced force maps that were visually similar to ground truth. We then explored how well this approach may generalize to unseen end effector velocities and limb rotations. A promising characteristic that emerged was the LSTM's ability to estimate force maps during scenarios in which the human limb was rotated. Notably, we have shown that high-dimensional force maps can be estimated using a sequence of low-dimensional measurements. These results suggest that robots could learn to infer the forces that people physically feel during robot-assisted dressing.

Acknowledgment

We thank Ariel Kapusta, Jie Tan, Tapomayukh Bhattacharjee, and Henry Clever for their assistance with this work. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-1650044 and NSF award IIS-1514258.

References

[1] R. C. J. M. Wiener, R. J. Hanley and J. F. V. Nostrand, "Measuring the activities of daily living: Comparisons across national surveys," *Journal of Gerontology: SOCIAL SCIENCES*, vol. 45, no. 6, pp. 229–237, 1990.

[2] P. van der Smagt, M. A. Arbib, and G. Metta, "Neurorobotics: From vision to action," in *Springer Handbook of Robotics*.

[3] 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*. ACM, 2013, pp. 1–6.

[4] ——, "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.

[5] N. Koganti, J. G. Ngeo, T. Tomoya, K. Ikeda, and T. Shibata, "Cloth dynamics modeling in latent spaces and its application to robotic clothing assistance," in *Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on*. IEEE, 2015, pp. 3464–3469.

[6] L. Twardon and H. Ritter, "Interaction skills for a coat-check robot: Identifying and handling the boundary components of clothes," in *2015 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2015, pp. 3682–3688.

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

[8] 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.

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

[10] 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 *System Integration (SII), 2014 IEEE/SICE International Symposium on*. IEEE, 2014, pp. 564–570.

[11] Y. Gao, H. J. Chang, and Y. Demiris, "Iterative path optimisation for personalised dressing assistance using vision and force information," in *Intelligent Robots and Systems (IROS), 2016 IEEE/RSJ International Conference on*. IEEE, 2016, pp. 4398–4403.

[12] A. Kapusta, W. Yu, T. Bhattacharjee, C. K. Liu, G. Turk, and C. C. Kemp, "Data-driven haptic perception for robot-assisted dressing," in *Proc. IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 2016.

[13] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[14] F. A. Gers, J. Schmidhuber, and F. Cummins, "Learning to forget: Continual prediction with lstm," *Neural computation*, vol. 12, no. 10, pp. 2451–2471, 2000.

[15] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," in *Advances in neural information processing systems*, 2014, pp. 3104–3112.

[16] A. Graves, "Generating sequences with recurrent neural networks," *arXiv preprint arXiv:1308.0850*, 2013.

[17] A. Graves, A.-r. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in *2013 IEEE international conference on acoustics, speech and signal processing*.

[18] A. I. Aviles, S. Alsaleh, P. Sobrevilla, and A. Casals, "Exploring the effects of dimensionality reduction in deep networks for force estimation in robotic-assisted surgery," in *SPIE Medical Imaging*. International Society for Optics and Photonics, 2016.

[19] D. Berenson, "Manipulation of deformable objects without modeling and simulating deformation," in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2013, pp. 4525–4532.

[20] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks." in *Aistats*, vol. 9, 2010, pp. 249–256.

[21] W. Yu, A. Kapusta, J. Tan, C. C. Kemp, G. Turk, and C. K. Liu, "Haptic data simulation for robot-assisted dressing," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2017.

[22] N. Hansen, "The CMA evolution strategy: A tutorial," *Technische Universitat Berlin, TU Berlin*, 2016.

[23] F. Chollet, "Keras," https://github.com/fchollet/keras, 2015.

[24] T. Tieleman and G. Hinton, "Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude," *COURSERA: Neural Networks for Machine Learning*, vol. 4, no. 2, 2012.