Learning Manipulation of Deformable Objects from Multiple Demonstrations

Henry Lu

Electrical Engineering and Computer Sciences
University of California at Berkeley

December 1, 2015

Acknowledgement

Learning Manipulation of Deformable Objects from Multiple Demonstrations

by Henry Lu

Research Project

Submitted to the Department of Electrical Engineering and Computer Sciences, University of California at Berkeley, in partial satisfaction of the requirements for the degree of **Master of Science, Plan II**.

Approval for the Report and Comprehensive Examination:

Committee:

Professor P. Abbeel
Research Advisor

(Date)

* * * * * * *

Professor K. Goldberg
Second Reader

(Date)

Abstract

Learning from multiple demonstrations using non-rigid point cloud registration is an effective method for learning manipulation of deformable objects. Using non-rigid registration, we compute a warping function that transforms the demonstrated trajectory in each demonstration into the current scene. In the first section we present a technique for learning force-based manipulation from multiple demonstrations. Our method utilizes the variation between demonstrations to learn a variable-impedance control strategy that trades off force and position errors, providing the correct level of compliance that applies the necessary forces at different stages in the trajectory. The resulting force-augmented trajectory is effective in manipulating deformable objects that are variants of those used in demonstration, when the traditional kinematic method fails. In the second section we present a trajectory-aware non-rigid registration method that uses multiple demonstrations to focus the registration process on points that are more relevant to the task. This method allows us to handle significantly greater visual variation in the demonstrations than previous methods that are not trajectory-aware. When introducing irrelevant variations in the object geometry and random noise in the form distractor objects, trajectory-aware registration proves to be robust and capable of extracting the true goal of the task hidden in the similarities of the demonstrations. The naïve single demonstration method and an ablated implementation of the trajectory-aware registration often fail in these situations. We evaluate our approaches on challenging tasks such as towel folding, knot tying, and object grasping from a box.

Acknowledgements

I would like to thank my faculty advisor Professor Pieter Abbeel for his continuous guidance and encouragement. He is an inspirational figure that has helped me grow, as a researcher and as a person. I would also like to thank Sergey Levine, Alex X. Lee, and Abhishek Gupta for all of their contributions to the work presented in this report. These projects would not have been possible without the work you guys put in. Finally, I would like to thank my family and my girlfriend, for being supportive and always being there for me.

Contents

Chapter 1

Overview of Contributions

In the second chapter of this report we are going to present a method for learning force-based manipulation of deformable objects from multiple demonstrations. Manipulation of deformable objects often requires a robot to apply specific forces to bring the object into a desired configuration, especially when objects used during test time are variants of those used in the demonstration. Our approach uses non-rigid registration via thin plate spline robust point matching (TPS-RPM) to compute a warp function that transforms points from the demonstration scene to points in the test scene, and applies the same warp to both end-effector poses and forces in the demonstration. Variation between the demonstrations is used to extract a single trajectory and time-varying feedback gains that determine pose versus force matching. This results in a variable-impedance control strategy that provides the proper compliance and necessary forces at each stage of the task. The control strategy is evaluated on knot tying, towel folding, and whiteboard erasing. Our approach has a clear advantage over the kinematics only baseline when the manipulation task involves objects that are slight variants of those used to record the demonstrations, such as a tying a longer rope or folding a bigger towel. This work has been published [1, A. Lee, H. Lu, A. Gupta, S. Levine, and P. Abbeel] for the *IEEE International Conference on Robotics and Automation (ICRA)*, 2015.

In the third chapter we introduce a *trajectory-aware* non-rigid registration method that uses multiple demonstrations to focus the registration process on points more relevant to the deformable object manipulation task. This approach is effective for handling irrelevant variations in the object geometry and random noise in the form distractor objects. In contrast, traditional non-rigid point cloud registration methods that are not *trajectory-aware* assume that the test and demonstration scene are structurally similar and that any variation can be explained by a non-linear transformation. These assumptions are unrealistic in real-world manipulation scenarios, which are often cluttered with distractor objects. We compare the two approaches on several challenging tasks, such as folding a towel to a specified mark and grasping a bottle in a box containing irrelevant distractors. We show that our method is much more effective than the naïve single demonstration method and an ablated implementation of our method, which uses multiple demonstrations but is not *trajectory-aware*. The work has been submitted [2, A. Lee, S. Levine, A. Gupta, H. Lu, and P. Abbeel] and is currently under review.

Chapter 2

Force-Based Manipulation

Particular forces that bring an object into a desired configuration are often necessary to successfully manipulate soft and deformable objects. For example, pulling on the ends of the rope tightens a knot, smoothing out the wrinkles flattens a towel, and applying a downward force erases writing on a whiteboard. Forces are especially crucial when the objects manipulated during test time are slight variants of those used during the demonstrations. Learning from demonstrations is a successful approach for training robots to perform complex manipulating skills [3, 4, 5, 6], especially those needed for handling deformable objects [7], yet there are challenges to transferring forces from a demonstration to a new situation: how does the robot trade off errors in position and force when the test and training situations are mismatched?

We present a method that uses geometric warping and statistical learning across multiple demonstrations to compute target poses, forces, and time-varying gains for a new situation. During training, we perform the given task using teleoperation or kinesthetic teaching and collect several demonstrations. During testing, the demonstrations are adapted to the current situation with an *object-centric* warping. This is performed by registering the initial state of the object from each demonstration to the current scene, computing a warp function that maps 3D points from each demonstration into the current scene, and applying the same warp to each demonstrated trajectory.

In addition to the geometric properties of the trajectory, we also warp the demonstrated end-effector forces with the same warp function. However, as shown in the experiments, naïvely playing back the warped forces is generally ineffective. This is because the robot cannot simultaneously exert the desired force and maintain the same position as in the warped demonstration. It must trade off between position and force errors. We present an algorithm that uses the variance across the demonstrations to estimate the relative importance of maintaining position versus force, and fits time-varying feedback gains to the warped trajectories.

The main contribution of our work is a way to generalize demonstrations of manipulation behaviors including kinematic and force-driven phases. Our method warps position and force from demonstrated trajectories to align them with the current scene configuration, and then estimates a single trajectory with time-varying feedback gains that trade off position and force errors. The result is a learned variable-impedance controller for performing the task in new situations. Results show that our method is effective in tightening knots (see Figure 3.1), folding crumpled towels, and erasing writing on a whiteboard.

2.1   Related Work

Our approach for learning from demonstrations uses non-rigid registration for adapting demonstrated trajectories to the current scene. Other methods make use of task-specific features and
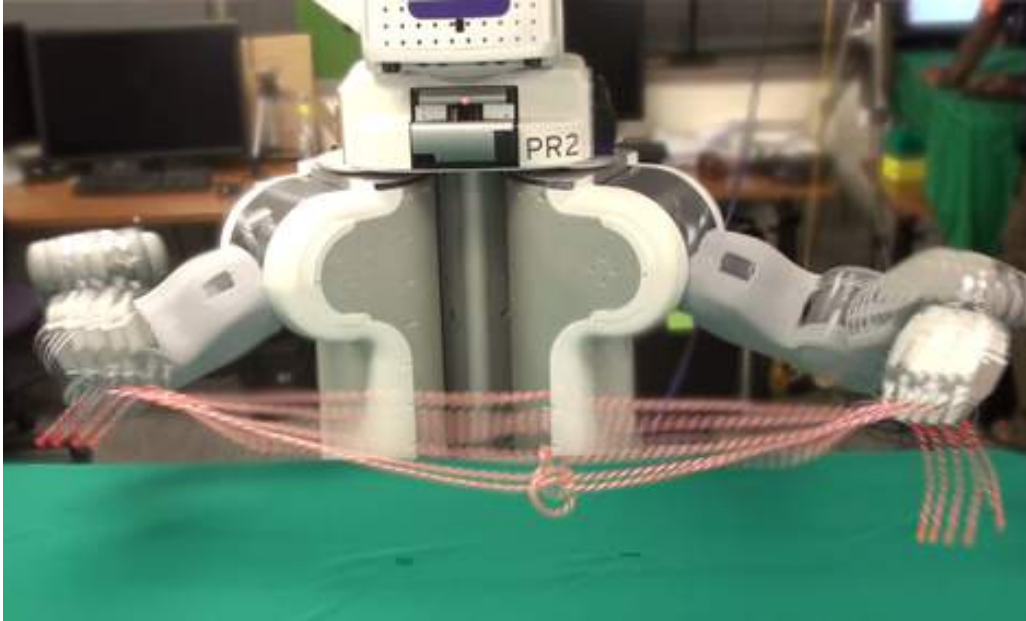
Figure 2.1: PR2 applying strong lateral forces to tighten a knot.

Gaussian mixture models [8, 9]. However, the non-rigid registration approach has the advantage of requiring only a simple, task-agnostic pipeline.

Our main contribution is to combine learning from demonstration with force-based control strategies. Force control has been proven to be a powerful tool for executing complex robotic motion skills [10], and a number of works have addressed learning variable-impedance control policies using reinforcement learning [11, 12, 13]. Although such methods are highly automated, they typically require extensive system interaction during learning, and address simpler behaviors.

Previous methods that learn force-based behaviors directly from demonstrations typically use a least squares formulation to solve for time-varying impedances [14, 15, 16, 17, 18, 19]. Some authors have proposed introducing perturbations into the demonstrations to observe the demonstrator's recovery stiffness [18], as well as sophisticated input modalities for directly specifying the desired stiffness [20, 19]. These techniques are effective for tasks that require variable stiffness, such as assembling furniture [17] and ironing [16]. Our object-centric approach for fitting gains is based on a statistical model of a collection of demonstrations and introduces several improvements, such as the use of a Bayesian prior.

We apply our method to manipulation of deformable objects, including rope and towels. This has been addressed using task-specific methods that exploit domain knowledge of the task [21, 22, 23, 24], motion or grasp planning techniques [25, 26, 27], and hand-engineered strategies [28].

2.2 Pipeline

We record a set of demonstrations (typically 5) for each stage of a task. Knot tying has three stages: first loop, second loop, and tightening the knot. Towel folding has two stages: straightening and folding. Erasing text on the whiteboard has a single stage. In addition to the demonstrations, we store point cloud of the scene at the beginning of each stage, generated with a Kinect depth sensor.

We then use the thin plate spline robust point matching (TPS-RPM) algorithm [29] to register the initial point cloud from each demonstration tot he current scene. The TPS-RPM algorithm produces
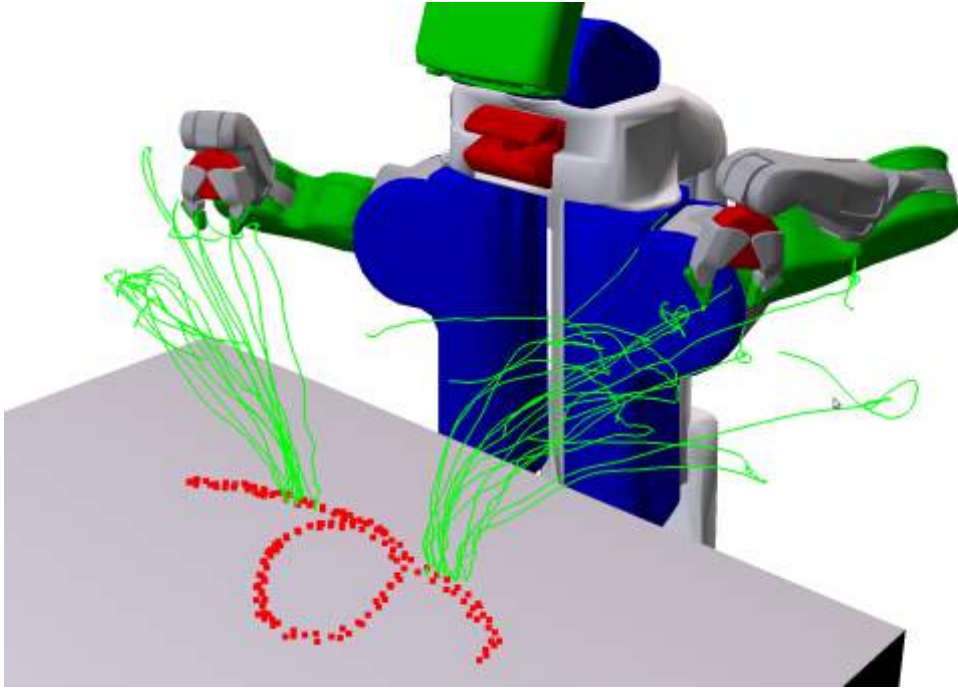
Figure 2.2: Trajectories for knot tightening (the third stage of knot tying) warped to match the current rope configuration (red).

a warp function $\mathbf{y}_i = f(\mathbf{x}_i)$ that maps each point in the demonstration scene to a corresponding point in the current scene. The function $f(\mathbf{x}_i)$ and the derivative of the function are used to transform the demonstrated trajectory and forces, respectively. This results in a set of force-position trajectories aligned to the current scene, as shown in Figure 2.2.

We also align the demonstrations in time with dynamic time warping, [30]. This allows us to extract a single mean trajectory for a denoised execution of the task. We further analyze the covariance of the demonstrations at each time step, and estimate position and velocity feedback gains. This allows us to tradeoff matching kinematics and demonstrated forces. For example, segments where there is low positional variation between warped trajectories should track demonstrated positions more accurately, while segments with high position positional variation should track the demonstrated forces.

Finally, the mean trajectory and the sequence of time-varying feedback gains are execute to perform the manipulation behavior. Warped forces are applied with the Jacobian transpose method and the kinematic trajectory is tracked with feedback gains. When the gains are low, the force term dominates the behavior of the robot. A diagram of our approach is shown in Figure 2.3.

2.3   Background

We will review the non-rigid point cloud registration TPS-RPM algorithm [29] and the approach of warping trajectories with the TPS warp function [7]. This method transforms the demonstrated trajectories in the same way that the demonstration object needs to be deformed to warp it into the current scene. The object-centric approach preserves the relationship between the robot end-effector motion and the object.
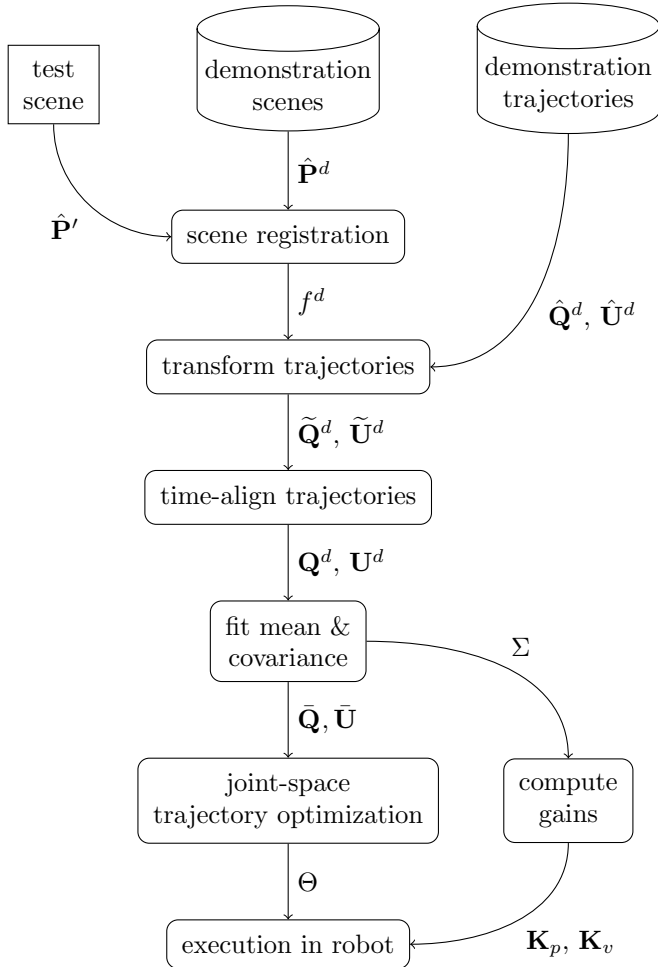
Figure 2.3: Diagram of our approach. $\hat{\mathbf{P}}^d$ and $\hat{\mathbf{P}}'$ are points in the demonstration and test scenes, f^d is the TPS-RPM registration function, $\hat{\mathbf{Q}}^d$ and $\hat{\mathbf{U}}^d$ are the demonstrated poses and forces, $\widetilde{\mathbf{Q}}^d$ and $\widetilde{\mathbf{U}}^d$ are transformed by f^d, $\mathbf{Q}^d$ and $\mathbf{U}^d$ are further registered in time, $\bar{\mathbf{Q}}$ and $\bar{\mathbf{U}}$ denote the mean trajectory, and Θ is the final sequence of joint angles. $\mathbf{K}_p$ and $\mathbf{K}_v$ are the computed position and velocity gains.

2.3.1 Non-rigid Registration

We assume that the demonstration scene consists of N points $\hat{\mathbf{P}} = [\hat{\mathbf{p}}_1 \ldots \hat{\mathbf{p}}_N]^\top$ and the test scene consists of N' points $\hat{\mathbf{P}}' = [\hat{\mathbf{p}}_1' \ldots \hat{\mathbf{p}}_{N'}']^\top$.

In the case when $N = N'$ and the correspondences between the points are known, the registration problem is to find a function $f : \mathbb{R}^3 \to \mathbb{R}^3$ that maps the points $\hat{\mathbf{P}}$ to the points $\hat{\mathbf{P}}'$. We limit f to be a thin plate spline [31] to estimate a smooth warp function. A thin plate spline is the optimal solution for the optimization problem:

$$f = \arg\min_f \sum_{i=1}^N ||\hat{\mathbf{p}}_i' - f(\hat{\mathbf{p}}_i)||_2^2 + \lambda ||f||_{\mathrm{TPS}}^2, \tag{2.1}$$

where the regularizer $||f||_{\mathrm{TPS}}^2$ is the TPS energy function

$$||f||_{\mathrm{TPS}}^2 = \int d\mathbf{p} ||\mathrm{D}^2 f(\mathbf{p})||_{\mathrm{F}}^2. \tag{2.2}$$

The regularizer measures the curvature of f. and λ balances between matching the correspondence points and the smoothness of the spline. There is an analytical solution for the optimal f [32], and it has the form

$$f(\mathbf{p}) = \sum_i \mathbf{a}_i k(\hat{\mathbf{p}}_i, \mathbf{p}) + \mathbf{B}\mathbf{p} + \mathbf{c}. \tag{2.3}$$

The solution is an affine transformation defined by $\mathbf{B}$ and $\mathbf{c}$ and a weighted sum of 3D basis functions $k(\hat{\mathbf{p}}_i, \mathbf{p}) = -||\mathbf{p} - \hat{\mathbf{p}}_i||^2$.

In the case that $N \neq N'$ and the correspondences between the points are not known, we use the thin plate spline robust point matching (TPS-RPM) algorithm [29] to register demo scene to test scene.

2.3.2 Learning from Demonstrations with Thin Plate Splines

A demonstration is made up of the initial scene point cloud $\hat{\mathbf{P}}$, a sequence of end-effector poses $\hat{\mathbf{Q}} = [\hat{\mathbf{q}}_1 \ldots \hat{\mathbf{q}}_T]^\top$, and a sequence of end-effector forces $\hat{\mathbf{U}} = [\hat{\mathbf{u}}_1 \ldots \hat{\mathbf{u}}_T]^\top$. During execution, the TPS-RPM algorithm computes a warp function f that maps points from the demonstration scene to the current scene. The warp function is applied to the end-effector poses and forces. Trajectory optimization [33] is used to find a feasible joint angle trajectory $\Theta = [\theta_1 \ldots \theta_T]^\top$ that accounts for collision avoidance and joint limits.

2.4 Aligning Force-Augmented Trajectories

In our approach, we estimate a single trajectory with desired forces and variable impedances from multiple demonstrations. This involves transforming force-augmented trajectories such that they are aligned spatially to the current scene and temporally to each other.

2.4.1 Warping Force Trajectories

A force-augmented trajectory contains a sequence of end-effector forces $\hat{\mathbf{U}} = [\hat{\mathbf{u}}_1 \ldots \hat{\mathbf{u}}_T]^\top$ where each force $\hat{\mathbf{u}}_t$ has a translational force and a torque. The forces are transformed by the derivative of f,

$$\widetilde{\mathbf{u}}_t = \frac{df}{d\mathbf{p}}(\hat{\mathbf{q}}_{\mathbf{p},t})\hat{\mathbf{u}}_t, \tag{2.4}$$

where $\frac{df}{d\mathbf{p}}(\hat{\mathbf{q}}_{\mathbf{p},t})$ is the derivative of f at $\hat{\mathbf{q}}_{\mathbf{p},t}$, and $\hat{\mathbf{q}}_{\mathbf{p},t}$ is the translational component of the end-effector pose $\hat{\mathbf{q}}_t$. This transformation rotates the forces to align them with the current test scene.

2.4.2 Aligning Multiple Demonstrations

We start by collecting demonstrations $\mathcal{D} = [\mathcal{D}^1 \ldots \mathcal{D}^D]$. Each demonstration $\mathcal{D}^d$ consists of the scene points $\hat{\mathbf{P}}^d$, a pose trajectory $\hat{\mathbf{Q}}^d$, and a force trajectory $\hat{\mathbf{U}}^d$. For each demonstration $\mathcal{D}^d$, we first compute a registration function f^d with TPS-RPM that maps points from the demonstration scene $\hat{\mathbf{P}}^d$ to the current scene $\hat{\mathbf{P}}'$. We then apply the registration function to each demonstration trajectory to get $\widetilde{\mathbf{Q}}^d$, $\widetilde{\mathbf{U}}^d$:

$$\widetilde{\mathbf{q}}_t^d = f^d(\hat{\mathbf{q}}_t^d)$$

$$\widetilde{\mathbf{u}}_t^d = \frac{df^d}{d\mathbf{p}}(\hat{\mathbf{q}}_{\mathbf{P},t}^d)\hat{\mathbf{u}}_t^d.$$

Now that the pose and force trajectories $\widetilde{\mathbf{Q}}^d$ and $\widetilde{\mathbf{U}}^d$ are spatially aligned with respect to the test scene, we use dynamic time warping [30] to time-align the trajectories with respect to the trajectory of demonstration $\mathcal{D}^1$, the demonstration most similar to the current scene.

2.5 Learning Variable-Impedance Control

After aligning the demonstrations spatially and temporally, we extract a single mean trajectory of poses, velocities, forces, and time-varying feedback gains $\mathbf{K}_{pt}$ and $\mathbf{K}_{vt}$ that determine how much importance is given to matching position and velocity over matching force. This is done by analyzing the mean and covariance of poses, velocities, and forces in the aligned demonstrations.

2.5.1 Probabilistic Modeling

The demonstrations are modeled by a time-varying linear-Gaussian control law, where end-effector force is a linear function of the poses and velocities, corrupted by Gaussian noise. The probability of observing a force $\mathbf{u}_t^d$ is modeled as

$$p(\mathbf{u}_t^d | \mathbf{q}_t^d, \dot{\mathbf{q}}_t^d) = \mathcal{N}(\mathbf{K}_{pt}(\bar{\mathbf{q}}_t - \mathbf{q}_t^d) + \mathbf{K}_{vt}(\dot{\bar{\mathbf{q}}}_t - \dot{\mathbf{q}}_t^d) + \bar{\mathbf{u}}_t; \mathbf{C}_t), \qquad (2.5)$$

where $\bar{\mathbf{q}}_t$, $\dot{\bar{\mathbf{q}}}_t$, and $\bar{\mathbf{u}}_t$ are the position, velocity, and force of the desired trajectory. The maximum likelihood solutions for $\bar{\mathbf{q}}_t$, $\dot{\bar{\mathbf{q}}}_t$, $\bar{\mathbf{u}}_t$, $\mathbf{K}_{pt}$, and $\mathbf{K}_{vt}$ are obtained by maximizing the probability of each of the demonstrated trajectories. If the mean and covariance of $\{(\mathbf{q}_t^d, \dot{\mathbf{q}}_t^d, \mathbf{u}_t^d)^\top\}$ are given by μ_t and Σ_t, respectively, the parameters are given by:

$$\bar{\mathbf{q}}_t = \mu_{\mathbf{q},t} \qquad\qquad \mathbf{K}_{pt} = -\Sigma_{\mathbf{uq},t}\Sigma_{\mathbf{qq},t}^{-1}$$

$$\dot{\bar{\mathbf{q}}}_t = \mu_{\dot{\mathbf{q}},t} \qquad\qquad \mathbf{K}_{vt} = -\Sigma_{\mathbf{u}\dot{\mathbf{q}},t}\Sigma_{\dot{\mathbf{q}}\dot{\mathbf{q}},t}^{-1}$$

$$\bar{\mathbf{u}}_t = \mu_{\mathbf{u},t},$$

where subscripts indicate submatrices.
Similar to the least-squares method for learning variable-impedance motions [14, 15, 18], the method corresponds to assigning more importance to minimize position errors when positions are more consistent across demonstrations. When the positions are less consistent, the corresponding gains will be low and the force term will dominate.

2.5.2 Feedback Gain Priors

When the demonstrated trajectories diverge in different directions, the learned gains will be negative, resulting in an unstable controller. Therefore, we take a Bayesian approach and impose a prior on the covariance. The standard approach is to use an inverse-Wishart distribution [34],

determined the degrees of freedom parameter ν and the prior matrix Ψ. The maximum a posteriori (MAP) estimate of the covariance Σ_t under this prior is given by

$$\Sigma_t = \frac{D\hat{\Sigma}_t + \Psi}{\nu + D + P + 1},$$

where P is the dimensionality of the data points, and $\hat{\Sigma}_t$ is the empirical covariance. In our implementation, we set $\nu = P + 2$ and construct the prior matrix Ψ as follows:

$$\Psi_{\mathbf{qq}} = \mathrm{Cov}(\mathbf{Q})w_p \qquad\qquad \Psi_{\mathbf{uq}} = -\mathbf{K}_p^0\Psi_{\mathbf{qq}}$$
$$\Psi_{\dot{\mathbf{q}}\dot{\mathbf{q}}} = \mathrm{Cov}(\dot{\mathbf{Q}})w_v \qquad\qquad \Psi_{\mathbf{u}\dot{\mathbf{q}}} = -\mathbf{K}_v^0\Psi_{\dot{\mathbf{q}}\dot{\mathbf{q}}},$$

where $\mathrm{Cov}(\mathbf{Q})$ denotes the empirical covariance of the warped poses, while w_p and w_v are weights that determine the relative importance of the prior gains for poses and velocities, respectively. The prior gains $\mathbf{K}_p^0$ and $\mathbf{K}_v^0$ are set values on the same order as the default PD gains on a PR2 robot, while the weights are set to $w_p = 0.1$ and $w_v = 10.0$. We clip negative gains to zero at test time.

2.5.3 Temporal Windowing

With only 5 demonstrations for each task, the data is inadequate to independently fit the covariance at each time step. Therefore, we assume that gains change slowly and use additional samples from adjacent time steps with a weighted fitting scheme when computing the empirical covariance $\hat{\Sigma}_t$. We put a squared exponential weight on each sample based on that sample's time stamp τ: $w_\tau = \exp(-\frac{1}{2\sigma^2}(\tau - t)^2)$. In our implementation, we set $\sigma = 1.0$, which corresponds to one second.

2.5.4 Joint-Space Control

We execute the mean end-effector trajectory on the robot with feedback gains $\mathbf{K}_{pt}$ and $\mathbf{K}_{vt}$. In order to avoid kinematic singularities, we perform trajectory optimization [7, 33] and generate a sequence of joint angles $\Theta = [\theta_1 \ldots \theta_T]^\top$ that match the desired end-effector trajectory while conforming to kinematic constraints. The gains $\mathbf{K}_{pt}$ and $\mathbf{K}_{vt}$ are then also converted into joint space gains $\mathbf{K}_{pt}^\theta$ and $\mathbf{K}_{vt}^\theta$, using the Jacobian at θ_t:

$$\mathbf{K}_{pt}^\theta = \mathbf{J}(\theta_t)^\top\mathbf{K}_{pt}\mathbf{J}(\theta_t) + \mathbf{K}_{p,\mathrm{pd}}^\theta$$
$$\mathbf{K}_{vt}^\theta = \mathbf{J}(\theta_t)^\top\mathbf{K}_{vt}\mathbf{J}(\theta_t) + \mathbf{K}_{v,\mathrm{pd}}^\theta,$$

where $\mathbf{K}_{p,\mathrm{pd}}^\theta$ and $\mathbf{K}_{v,\mathrm{pd}}^\theta$ are low PD gains (set to 1% of the default values) that ensure that the robot keeps moving along the trajectory along directions that lie in the null space of $\mathbf{J}(\theta_t)$. Finally, the torque applied at the joints at test time is given by

$$\mathbf{f}_t = \mathbf{K}_{pt}^\theta(\theta_t - \theta_{\mathrm{obs}}) + \mathbf{K}_{vt}^\theta(\dot{\theta}_t - \dot{\theta}_{\mathrm{obs}}) + \mathbf{J}(\theta)^\top\bar{\mathbf{u}}_t,$$

where θ_{obs} and $\dot{\theta}_{\mathrm{obs}}$ are the observed joint angles and velocities, and $\mathbf{f}_t$ is the torque to apply at each joint.

2.6 Experimental Results

We evaluated our approach on four different tasks: tying a knot, tightening a knot around a pipe, folding a towel, and erasing text on a whiteboard. The objects used in test time are slight variations of those used in the demonstrations. For each task, we compared our method against a baseline that tracked the warp trajectory with a standard PD controller [7].
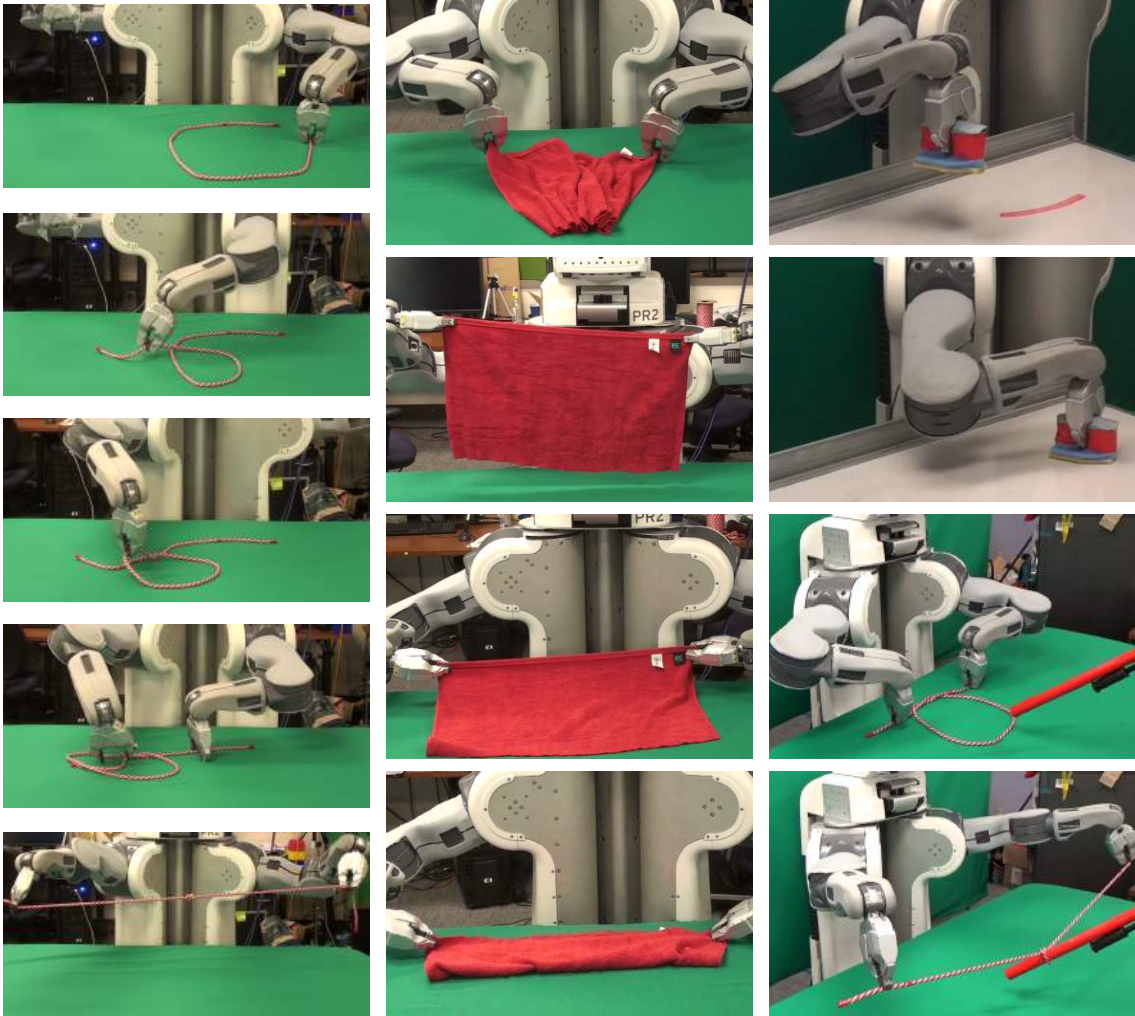
Figure 2.4: Images of the tasks (progress top to bottom) in our experiments: tying a knot (left), folding a towel (middle), erasing a whiteboard (top right), and tightening a pipe (bottom right).
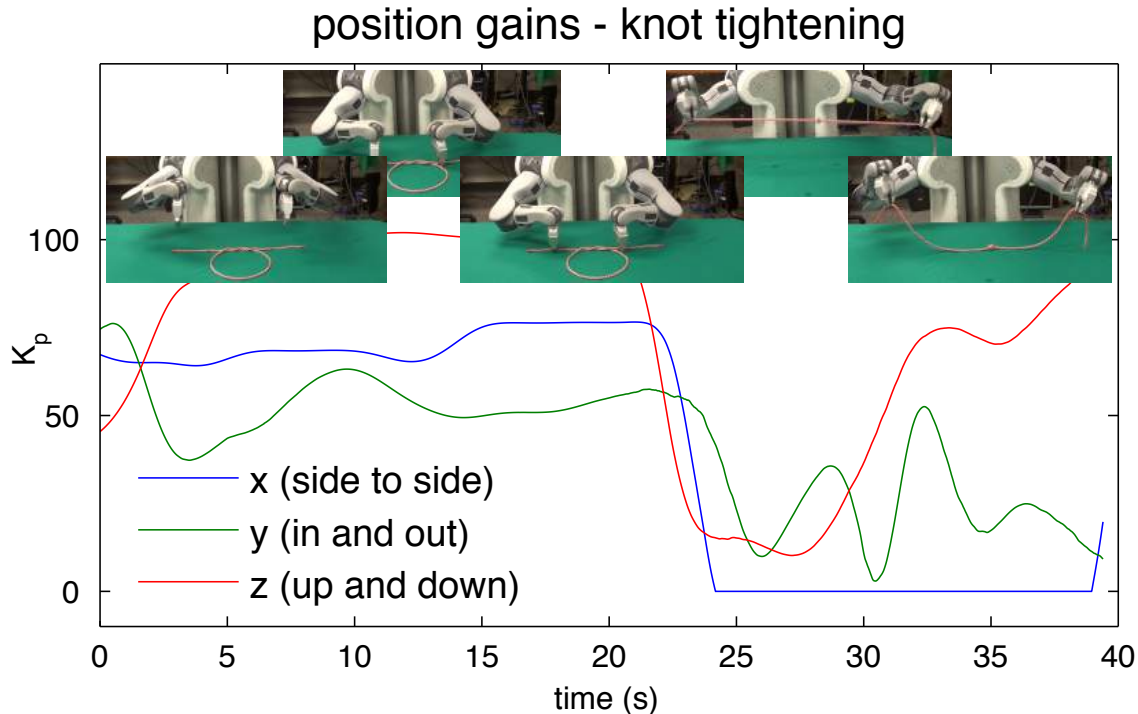
Figure 2.5: Position gains learned for knot tightening. Images of the rope at the corresponding times are shown above the plot. Note that the side-to-side gains drop to zero during tightening, when the behavior is dominated by the pulling force.

Since the PR2 robot does not have end-effector force and torque sensors, we compute the end-effector forces from torques measured by the motors at the joints during demonstration. Since the external force applied by the teacher is not registered by the motors with kinesthetic teaching, we had the robot replay the demonstration in an identical situation with high position gains and recorded the resulting torques.

Images of the robot performing each task using our method and the baseline are provided in Figure 2.4.

In knot tying, the tightening of the knot is defined by the force that is exerted, not the positions. Therefore, kinematic execution of a warped trajectory will often fail to tighten the knot fully when the rope is longer than the one used in demonstration. Our learned gains, shown in Figure 2.5, indicate that our method was able to learn this nuance successfully. The segment between 24 and 28 seconds, corresponding to the tightening action, has the lowest gains.

In Figure 2.6, we show the pipe after tightening the longer rope with our method and the kinematic baseline. Our method tightened the rope around the pipe while the kinematic baseline failed and the rope fell off the pipe.

In Figure 2.7, we show the whiteboard after erasing it with our method and the kinematic baseline when the eraser is held at a slightly higher point than that in the demonstrations. With the kinematic baseline, only the relative height gripper is matched, which is insufficient to erase the board. Our method applied a downward force and erased the board completely.

In Figure 2.8, we show towels of various sizes before and after folding. The demonstrations were recorded on a towel 68 by 42 cm in size, and testing was done on towels 68 by 42 cm, 76 by 45 cm, and 90 by 42 cm in size. While the kinematic baseline was unable to even straighten the towel, our method was able to consistently straighten and fold each towel.
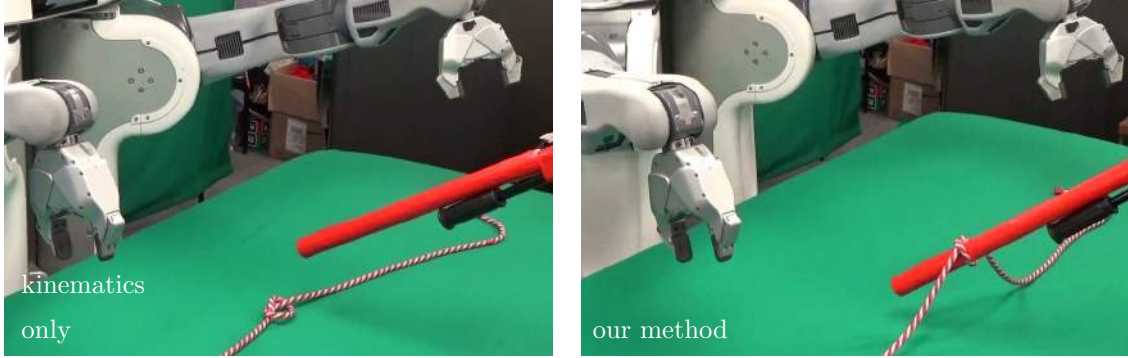
Figure 2.6: Tightening pipe experiment: the longer rope after being tightened by the kinematic baseline (left) and our method (right). Our approach applied an outward force, so the rope was taut and stayed in the pipe.

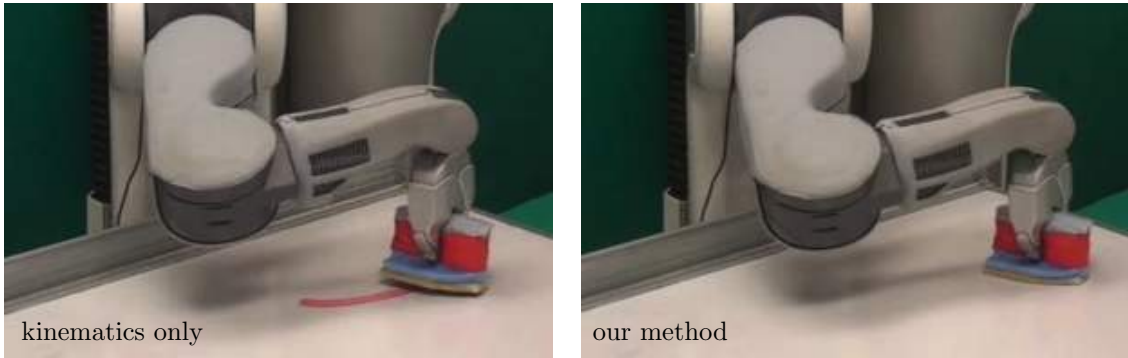Figure 2.7: Erasing experiment: the whiteboard after being erased by the kinematic baseline (left) and our method (right). Since our approach applied a downward force, the board was erased much more cleanly.
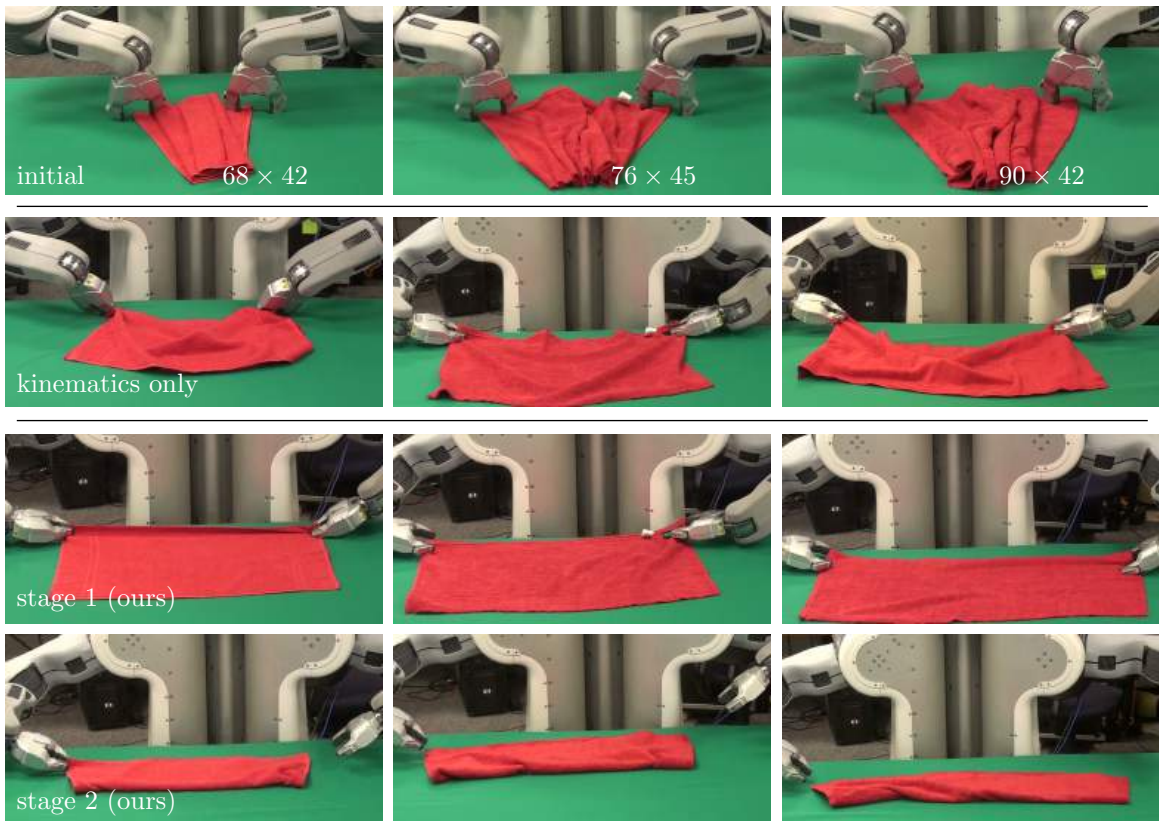
Figure 2.8: Towel experiments: the top row shows the initial layout, the second row shows the kinematic method after the first stage, and the third and fourth rows show our method after the first and second stages.

Chapter 3

Trajectory-Aware Non-Rigid Registration

Learning from demonstrations has emerged as an effective framework for teaching robots to perform complex motion skills, such as manipulating deformable objects. A key component in learning from demonstrations is to transfer the motion in training time to adapt to the new situation in test time. A general approach is to determine the transformation function that maps the demonstration scene to the current scene, and transform the demonstrated trajectory using the same function. This approach has been shown to be successful for tasks such as knot tying [7] and towel folding [1].

However, real-world scenes exhibit high variation in shape and color, even when the target object only changes slightly. Therefore, it is crucial to focus the registration process about the points that are relevant to the task. This can be done with multiple demonstrations.

In this work, we present a trajectory-aware non-rigid registration method that includes the demonstrated trajectories in the registration process, in addition to the point clouds in the demonstrations. While registering multiple point clouds from the demonstration scenes to the current scenes, we compute the transformation that best aligns the demonstrated trajectories. This alignment is quantified by means of the variance of time-aligned points along the trajectories. With this approach, the best transformation is the one that not only transforms the demonstration scene point cloud to the current scene point cloud, but also brings the demonstrated trajectories close together. This method allows us to effectively handle irrelevant variations in object geometry.

3.1   Related Work

Our work uses trajectories from multiple demonstrations to determine which parts of the scene are more or less relevant for the task. The idea of using multiple demonstrations to improve transfer and generalization has been explored in the context of autonomous flight [35], autonomous driving [36], and dynamic movement primitives [37]. However, ours is the first method that incorporates information from multiple demonstrated trajectories into the objective for non-rigid registration.

3.2   Background

The backbone of our approach is to find a warping function that maps points in the demonstration scene to points in the current scene using non-rigid registration and apply that same warp to the end-effector motion. This has the important property that the resulting end-effector motion incorporates variations in the scenes. The non-rigid cloud registration TPS-RPM algorithm is re-
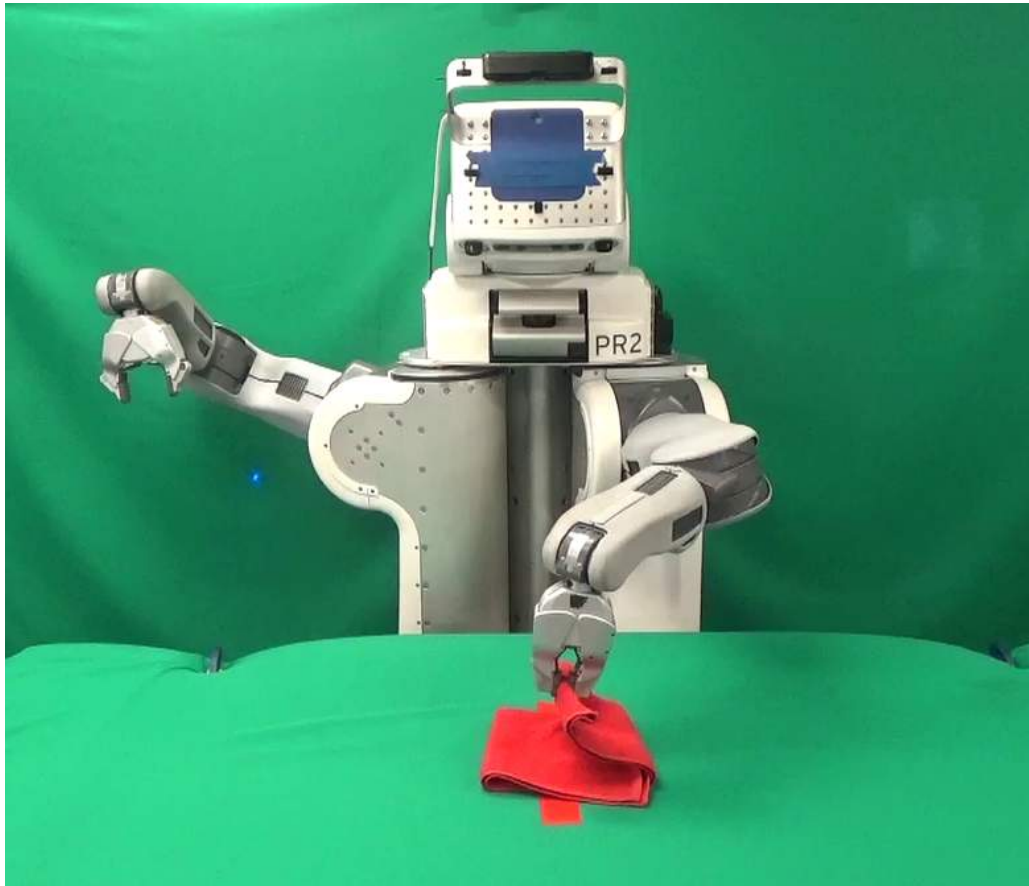
Figure 3.1: PR2 folding a towel in thirds.

viewed in 2.3.1 and the approach of warping trajectories with the TPS warp function is reviewed in 2.3.2.

3.3 Learning from Multiple Demonstrations using Trajectory-Aware Registration

While the non-rigid point cloud registration TPS-RPM algorithm from 2.3.1 can be used to transfer demonstrations for deformable object manipulation tasks, as shown in prior work [7] and in Section 2, it assumes that the transformation that aligns the demonstration point cloud entirely to the current point cloud will also be appropriate for warping the trajectory. This assumption is reasonable when the point clouds correspond to the same object, since points that interact with the trajectory are transformed accordingly to remain on the object in the new scene. However, in natural environments, two point clouds almost never correspond to the exact same objects.

In our work, we take multiple trajectories and include an additional term to the registration objective that encodes the preference for registrations that align all the trajectories. This is derived from the hypothesis that all the demonstrations belong to the same behavior, and the only variation between them is due to arrangement of the scene. Thus, if all the demonstrations were transformed by the (unknown) correct warp, they should all be identical to the test scene.

Given K demonstrations of the same task, we jointly optimize the registration functions $f^1, \ldots, f^K$ for all of the demonstration point clouds along with a cost term that measures the degree to which warped trajectories are aligned. Note that this approach requires multiple demonstrations in order to be *trajectory-aware*.

3.3.1 Time Alignment

Before quantifying the alignment of multiple trajectories, we must first compensate for discrepancies due to differences in timing between demonstrations. We use the standard dynamic time warping algorithm [30] to time-align each of the demonstration trajectories to the first trajectory in the set.

3.3.2 Trajectory-Aware Non-Rigid Registration

Once the demonstrations are time-aligned, we can derive the trajectory-aware non-rigid registration objective. Let $\mathbf{Q}_t$ be the set of time-aligned poses from each of the K demonstrations at time t, where $\mathbf{q}_t^k$ denotes the pose at time t within trajectory k, and

$$\mathbf{Q}_t = [\mathbf{q}_t^1 \cdots \mathbf{q}_t^K]. \tag{3.1}$$

Then, let $\mathcal{F}$ denote the operator that transforms all of the poses based on the registration functions $f^1, ..., f^K$, applying the kth transformation to the kth pose:

$$\mathcal{F}(\mathbf{Q}_t) = [f^1(\mathbf{q}_t^1) \cdots f^K(\mathbf{q}_t^K)]. \tag{3.2}$$

The trajectory-aware non-rigid registration method simultaneously aligns the points in demonstration scenes to the test scene, by minimizing pairwise registration costs from 2.3.1, while also minimizing the weighted sum of spatial variances of warped poses across demonstrations. This results in the following optimization problem:

$$\min_{\mathcal{F}} \quad \sum_{k=1}^{K} d_{\mathrm{L}_2}(\hat{\mathbf{P}}', \hat{\mathbf{P}}^k, f^k) + \lambda ||f^k||_{\mathrm{TPS}}^2$$

$$+ \sum_{t=1}^{T} w_t \, \mathrm{tr}\left(\mathrm{Cov}(\mathcal{F}(\mathbf{Q}_t))\right), \tag{3.3}$$

where $\text{Cov}(\mathcal{F}(\mathbf{Q}_t))$ is the empirical covariance of the warped poses at time step t. Minimizing the spatial variances of warped poses at each time step brings all the warped demonstration trajectories close together.

End-effector trajectory points that are closer to the objects in the scene are more important, since the robot interacts with the world primarily with its grippers. In addition, points on the trajectories that are far away from objects exhibit great random variability during demonstrations. Therefore, we set the weight w_t on the variance at each time step t using the following Gaussian weighting scheme:

$$w_t = \frac{1}{K} \sum_{k=1}^{K} \exp\left(-d_{kt}^2/\sigma^2\right), \tag{3.4}$$

where d_{kt} is the minimum distance between the position in the kth trajectory at step t and the point cloud for that demonstration.

After solving the trajectory-aware non-rigid registration optimization problem, we compute the mean of the warped poses $\widetilde{\mathbf{Q}} = [\widetilde{\mathbf{q}}_1 \cdots \widetilde{\mathbf{q}}_T]$:

$$\widetilde{\mathbf{q}}_t = \frac{1}{K} \sum_{k=1}^{K} f^k(\mathbf{q}_t^k). \tag{3.5}$$

We then follow the same procedure in [7] and Section 2, and use trajectory optimization to find a feasible, collision-free joint angle trajectory that follows the end-effector trajectory in $\widetilde{\mathbf{Q}}$.

3.4   Experimental Results

We evaluated our trajectory-aware non-rigid registration approach on two tasks: pick and place from a box and towel folding. We showed that our method offers improvements in generalization and robustness in comparison to a single-demonstration baseline that only considers point clouds. We also compared our approach with an ablated version without the trajectory-aware component, dropping the sum of variances in Equation 3.3 and only optimizing for the pairwise L_2 distances.

3.4.1   Pick and Place from a Box

We evaluated our approach on a pick and place task where the demonstration scenes are very similar and the main difference is the trajectory. In every scene, here is a bottle or a toy hammer or both inside a box, and the task is to pick one of the objects and place it outside the box. The demonstrations, shown in Figure 3.2, consisted of picking up a bottle from various locations inside the box.

During test time, when we placed the hammer in a corner and the bottle in the center of the box, both the single-demonstration baseline and the ablated version of our method missed the bottle. Furthermore, the ablated version aligned the point clouds of the entire box such that the warped trajectories went towards various parts of the box. The method was not *trajectory-aware*, and missed the critical aspect of the task: the bottle. On the other hand, our method used information from multiple trajectories to register the correct object, and succeeded at grasping the bottle.

Images of the robot execution at test time and renderings of the resulting registrations and transferred trajectory are shown in Table 3.1.

3.4.2   Towel Folding

We also evaluated our approach on a towel folding task. The towel starts flat and open along the length of the table. The task, shown in Figure 3.3, is to have the robot grasp the right edge of the towel and fold it such that the edge is aligned with the red tape on the table. In other words,

18

Figure 3.2: Demonstrations of picking up a bottle from various locations inside box and placing it on the table.

	Single-demonstration [7]	Ablated (pairwise registration)	Ours (trajectory-aware)
Picking up a bottle in a box			

Table 3.1: Results of picking up a bottle from the box and placing it in the table, using 3 different approaches. The renderings visualize the test point cloud as blue points, the transferred gripper trajectory as blue lines, and the warped demonstration point clouds and trajectories as points and lines of other colors (one color per demonstration).

Figure 3.3: Demonstrations of folding the towel using the red tape as reference for the target fold.
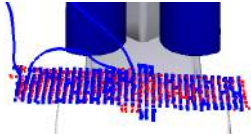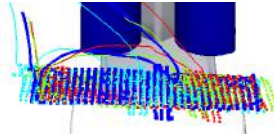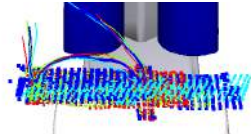
	Single-demonstration [7]	Ablated (pairwise registration)	Ours (trajectory-aware)
Fold towel towards the red tape reference			

Table 3.2: Results of towel folding such that the edge is aligned with the red tape, using 3 different approaches. The renderings visualize the test point cloud as blue points, the transferred gripper trajectory as blue lines, and the warped demonstration point clouds and trajectories as points and lines of other colors (one color per demonstration).

the tape indicates how far the fold must extend. Demonstrations consisted of placing the red tape along various lengths of the towel, and folding to the location of the tape each time.

During test time, our method moved the edge of the towel closest to the target, because it minimized the variances of the warped trajectories, which contributed to aligning the warped points of the tape markers. The single-demonstration baseline undershot the red tape because it did not register the tape markers with enough precision. On the other hand, the ablated version was able to register the towels to each other but was still off on the registration of tape markers.

Images of the robot execution at test time and renderings of the resulting registrations and transferred trajectories are shown in Table 3.2.

Our experiments showed that the trajectory-aware registration method was able to correctly pick up on the right cues in the environment, while both the prior method and the ablated method naïvely attempted to register the dissimilar scenes to one another without consideration for the task.

Chapter 4

Conclusion & Future Work

We were able to build on top of non-rigid point cloud registration by analyzing multiple demonstrations of the same task. This allowed us to loosen the constraints in the environments when learning manipulation of deformable objects.

In the first section we presented a technique for learning force-based manipulation from multiple demonstrations. Our method utilized variation between demonstrations to learn a variable-impedance control strategy that traded off force and position errors, providing the correct level of position control as well as the necessary forces at each stage of the task. This allowed the robot to complete force-dependent tasks, such as tying knots in rope, flattening towels, and erasing a whiteboard, even when the objects involved are slight variants of those used in recording the demonstrations: longer ropes, bigger towels, and eraser held at a higher point. The kinematic baseline with a standard PD controller (no forces) failed in these situations.

In the second section we introduced a trajectory-aware non-rigid registration method that used multiple demonstrations to focus the registration process on points that are more relevant to the task. This enabled the robot to ignore irrelevant changes in the scene and thus handle greater visual variation. We showed improvements over the single demonstration baseline as well as the ablated implementation of our method in picking out a bottle from a box containing distractors and folding a towel to a specified location. In the pick and place task, the single-demonstration approach was not accurate enough to grasp the bottle, while the ablated method registered point clouds of the entire box and missed the crucial aspect of the task: the bottle. In the towel folding task, a single demonstration did not register the tape markers with enough precision, and the ablated method only registered the towels. The trajectory-aware non-rigid registration method presented was able to identify the importance of the bottle in the first task and the meaning of the tape marker in the second task by analyzing not just the point clouds but also the demonstrated trajectories.

Several future directions could extend this work to further improve generalization. Access to a robot model would allow us to model external forces using internal forces and accelerations. We could then improve on the current open-loop Jacobian transpose scheme by performing feedback on external forces and implementing closed-loop control of forces. With the robot model, model-based inverse reinforcement learning algorithms can be used to infer a reward function for the task [38, 39], and optimal gains could then be recovered by running an optimal control algorithm. We could also develop a non-rigid transformation method that is not only trajectory-aware, but also goal-aware, by combining our existing approach with goal learning techniques [38, 39, 40, 41].

Bibliography

[1] A. Lee, H. Lu, A. Gupta, S. Levine, and P. Abbeel, "Learning force-based manipulation of deformable objects from multiple demonstrations," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2015.

[2] A. Lee, S. Levine, A. Gupta, H. Lu, and P. Abbeel, "Learning from multiple demonstrations using trajectory-aware non-rigid registration with applications to deformable object manipulation [under review]," 2015.

[3] S. Calinon, "Robot programming by demonstration," in *Springer Handbook of Robotics*. Springer, 2008, pp. 1371–1394.

[4] A. Billard, S. Calinon, R. Dillmann, and S. Schaal, "Robot programming by demonstration," in *Handbook of Robotics*, B. Siciliano and O. Khatib, Eds. Secaucus, NJ, USA: Springer, 2008, pp. 1371–1394.

[5] B. D. Argall, S. Chernova, M. Veloso, and B. Browning, "A Survey of Robot Learning from Demonstration," *Robot. Auton. Syst.*, vol. 57, no. 5, pp. 469–483, May 2009.

[6] S. Schaal, "Is imitation learning the route to humanoid robots?" *Trends in cognitive sciences*, vol. 3, no. 6, pp. 233–242, 1999.

[7] J. Schulman, J. Ho, C. Lee, and P. Abbeel, "Learning from demonstrations through the use of non-rigid registration," in *Proceedings of the 16th International Symposium on Robotics Research (ISRR)*, 2013.

[8] S. Calinon, F. Guenter, and A. Billard, "On learning, representing, and generalizing a task in a humanoid robot," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, vol. 37, no. 2, pp. 286–298, April 2007.

[9] S. Calinon, F. D'Halluin, D. Caldwell, and A. Billard, "Handling of multiple constraints and motion alternatives in a robot programming by demonstration framework," in *Humanoid Robots, 2009. Humanoids 2009. 9th IEEE-RAS International Conference on*, Dec 2009, pp. 582–588.

[10] O. Khatib, "A unified approach for motion and force control of robot manipulators: The operational space formulation," *IEEE Journal of Robotics and Automation*, vol. 3, no. 1, pp. 43–53, 1987.

[11] B. Yang and H. Asada, "Progressive learning and its application to robot impedance learning," *IEEE Transactions on Neural Networks*, vol. 7, no. 4, pp. 941–952, 1996.

[12] J. Buchli, F. Stulp, E. Theodorou, and S. Schaal, "Learning variable impedance control," *International Journal of Robotics Research*, vol. 30, no. 7, pp. 820–833, 2011.

[13] E. Rombokas, M. Malhotra, E. Theodorou, Y. Matsuoka, and E. Todorov, "Tendon-driven variable impedance control using reinforcement learning," in *Robotics: Science and Systems (R:SS)*, 2012.

[14] P. Sikka and B. McCarragher, "Stiffness-based understanding and modeling of contact tasks by human demonstration," in *Proceedings of the International Conference on Intelligent Robots and Systems (IROS)*, 1997.

[15] S. Calinon, I. Sardellitti, and D. Caldwell, "Learning-based control strategy for safe human-robot interaction exploiting task and robot redundancies," in *Proceedings of the International Conference on Intelligent Robots and Systems (IROS)*, 2010.

[16] P. Kormushev, S. Calinon, and D. G. Caldwell, "Imitation learning of positional and force skills demonstrated via kinesthetic teaching and haptic input," *Advanced Robotics*, vol. 25, no. 5, pp. 581–603, 2011.

[17] L. Rozo, S. Calinon, D. G. Caldwell, P. Jimenez, and C. Torras, "Learning collaborative impedance-based robot behaviors," in *AAAI Conference on Artificial Intelligence*, 2013.

[18] M. Li, H. Yin, K. Tahara, and A. Billard, "Learning object-level impedance control for robust grasping and dexterous manipulation," in *Proceedings of the International Conference on Robotics and Automation (ICRA)*, 2014.

[19] K. Kronander and A. Billard, "Learning compliant manipulation through kinesthetic and tactile human-robot interaction," *IEEE Transactions on Haptics*, vol. 7, no. 3, pp. 367–380, 2014.

[20] ——, "Online learning of varying stiffness through physical human-robot interaction," in *Proceedings of the International Conference on Robotics and Automation (ICRA)*, 2012.

[21] H. Inoue and M. Inaba, "Hand-eye coordination in rope handling," *Robotics Research: the First International Symposium*, pp. 163–174, 1985.

[22] T. Morita, J. Takamatsu, K. Ogawara, H. Kimura, and K. Ikeuchi, "Knot planning from observation," in *Proceedings of the International Conference on Robotics and Automation (ICRA)*, 2003, pp. 3887–3892.

[23] S. Miller, M. Fritz, T. Darrell, and P. Abbeel, "Parametrized shape models for clothing," in *Proceedings of the International Conference on Robotics and Automation (ICRA)*, 2011.

[24] B. Willimon, S. Birchfield, and I. Walker, "Model for unfolding laundry using interactive perception," in *Proceedings of the International Conference on Intelligent Robots and Systems (IROS)*, 2011.

[25] H. Wakamatsu, E. Arai, and S. Hirai, "Knotting/unknotting manipulation of deformable linear objects," *International Journal of Robotics Research*, vol. 25, no. 4, pp. 371–395, 2006.

[26] M. Saha, P. Isto, and J.-C. Latombe, "Motion planning for robotic manipulation of deformable linear objects," in *Experimental Robotics*, ser. Springer Tracts in Advanced Robotics, O. Khatib, V. Kumar, and D. Rus, Eds. Springer Berlin Heidelberg, 2008, vol. 39, pp. 23–32.

[27] C. Bersch, B. Pitzer, and S. Kammel, "Bimanual robotic cloth manipulation for laundry folding," in *Proceedings of the International Conference on Intelligent Robots and Systems (IROS)*, 2011.

[28] Y. Yamakawa, A. Namiki, M. Ishikawa, and M. Shimojo, "One-handed knotting of a flexible rope with a high-speed multifingered hand having tactile sensors," in *Proceedings of the International Conference on Intelligent Robots and Systems (IROS)*, 2007.

[29] H. Chui and A. Rangarajan, "A new point matching algorithm for non-rigid registration," *Computer Vision and Image Understanding*, vol. 89, no. 2-3, pp. 114–141, 2003.

[30] H. Sakoe, "Dynamic programming algorithm optimization for spoken word recognition," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 26, pp. 43–49, 1978.

[31] J. Duchon, "Splines minimizing rotation-invariant semi-norms in sobolev spaces," in *Constructive Theory of Functions of Several Variables*, ser. Lecture Notes in Mathematics, W. Schempp and K. Zeller, Eds.   Springer Berlin Heidelberg, 1977, vol. 571, pp. 85–100.

[32] G. Wahba, *Spline Models for Observational Data.*   Philadelphia: Society for Industrial and Applied Mathematics, 1990.

[33] J. D. Schulman, J. Ho, A. Lee, I. Awwal, H. Bradlow, and P. Abbeel, "Finding locally optimal, collision-free trajectories with sequential convex optimization," in *Robotics: Science and Systems (RSS)*, 2013.

[34] A. O'Hagan, *Bayesian Inference.*   Halsted Press, 1994.

[35] A. Coates, P. Abbeel, and A. Y. Ng, "Learning for control from multiple demonstrations," in *International Conference on Machine Learning (ICML)*, 2008.

[36] B. Argall, B. Browning, and M. Veloso, "Automatic weight learning for multiple data sources when learning from demonstration," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2009.

[37] T. Matsubara, S.-H. Hyon, and J. Morimoto, "Learning parametric dynamic movement primitives from multiple demonstrations," in *Proceedings of the 17th International Conference on Neural Information Processing: Theory and Algorithms - Volume Part I*, 2010, pp. 347–354.

[38] P. Abbeel and A. Y. Ng, "Apprenticeship learning via inverse reinforcement learning," in *International Conference on Machine Learning (ICML)*, 2004.

[39] S. Levine and V. Koltun, "Continuous inverse optimal control with locally optimal examples," in *International Conference on Machine Learning (ICML)*, 2012.

[40] N. D. Ratliff, J. A. Bagnell, and M. A. Zinkevich, "Maximum margin planning," in *Proceedings of the 23rd international conference on Machine learning.*   ACM, 2006, pp. 729–736.

[41] B. D. Ziebart, A. L. Maas, J. A. Bagnell, and A. K. Dey, "Maximum entropy inverse reinforcement learning." in *AAAI*, 2008, pp. 1433–1438.