Learning to Select Expert Demonstrations for Deformable Object Manipulation

Dylan Hadfield-Menell, Alex Lee, Sandy Huang, Eric Tzeng, and Pieter Abbeel

I. Overview

Automated manipulation of deformable objects tends to be challenging due to high-dimensional, continuous state-action spaces and due to the complicated dynamics of deformable objects. Direct planning or optimal control techniques are often intractable for this setting.

Despite these challenges, recent work [2, 3] has leveraged expert demonstrations to make progress on robotic manipulation of deformable objects. This work uses non-rigid registration between a demonstration scene and a test scene to find a geometric mapping between the two scenes. This mapping is used to perform *trajectory transfer* for the demonstrated gripper trajectory. This approach has been validated in simulated and real-world environments for knot-tying, suturing, and folding tasks.

Full demonstrations of complex tasks with multiple steps are hard to collect and transfer successfully. As such, we assume that demonstrations often correspond to steps in the task, rather than the entire task itself. Figure 1 shows an example of the steps involved in tying an overhand knot. Furthermore, a single demonstration for a step in the task cannot be expected to cover all possible scenarios that arise during execution. The natural solution to this is to use a library of demonstrations with multiple demonstrations for each step.

Realizing the benefits of a demonstration library requires a robust technique to select a good trajectory to transfer. Certain trajectories will generalize better than others, and particular sequences of demonstrations may perform tasks more efficiently than others.

The original paper on the approach of trajectory transfer prescribes choosing the trajectory segment from the demonstrations library with the lowest warping cost onto the current scene [3]. This approach does not account for the inherent generalizability of a particular demonstration. For brittle demonstrations (e.g. grabbing near the edge of a rope), a small change in the rope can have low registration cost, but the transferred trajectory will fail. As a result, such an approach may fail to accomplish tasks that would be possible with a different sequence of trajectories.

In this work, we present a solution to the demonstration selection problem that can account for the variability in robustness of demonstrations and incorporates the sequential nature of our tasks. Our contributions are as follows: (i) We formulate the demonstration selection problem as a Markov Decision Process (MDP); (ii) We present a method for approximating Q-functions from expert-guided task executions, based on the optimality of the expert's action selection; (iii) We describe task-independent features that are rich enough to allow learning but make no additional assumptions beyond those of trajectory transfer; and (iv) We validate this approach in a simulated knot-tying experiment and show strong improvement over previous approaches.

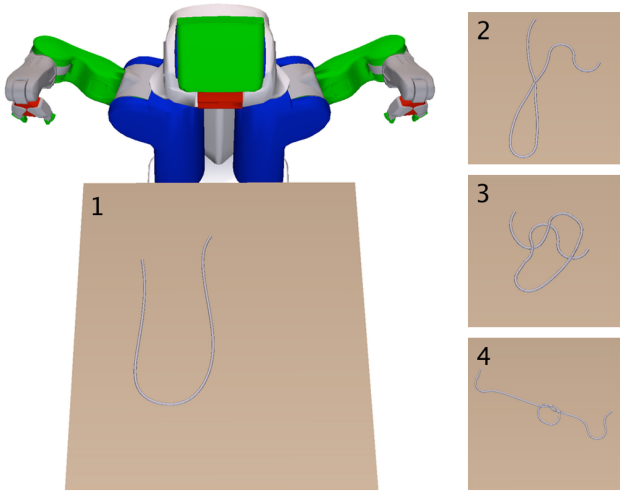

Fig. 1: The overhand knot manipulation task in our benchmark. A standard knot tie takes three steps, as shown in this particular execution from our benchmark.

II. Trajectory Transfer & MDP Formulation

Non-rigid registration computes a function f that minimizes error between landmark points, subject to a regularization term. A commonly-used, effective method for registering spatial data is the Thin Plate Spline (TPS) regularizer [1, 4]. Given a set of correspondence points $(\mathbf{x}_i, \mathbf{y}_i)$, the goal is to find the warping function $\mathbf{f} : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ that minimizes the following objective:

$$\min_{\mathbf{f}} \sum_i ||\mathbf{x}_i - \mathbf{y}_i||^2 + C \int dx ||\mathrm{D}^2(\mathbf{f})||^2_{\text{Frob}},$$

where C is a hyper-parameter that trades off between correspondence error and increased curvature. The second term measures curvature: $\mathrm{D}^2(\mathbf{f})$ is the matrix of second order partial derivatives of $\mathbf{f}$, and $||\cdot||^2_{\text{Frob}}$ denotes the Frobenius norm. This problem has a finite dimensional solution in terms of basis functions around the correspondence points. More concretely,

$\mathbf{f}$ has the form

$$\mathbf{f}(\mathbf{x}) = \sum_i \mathbf{a}_i K(\mathbf{x}_i, \mathbf{x}) + \mathbf{B}\mathbf{x} + \mathbf{c}$$

where K is the 3D TPS kernel $K(\mathbf{x}, \mathbf{x}') = -||\mathbf{x} - \mathbf{x}'||$, and $\mathbf{a}_i \in \mathbb{R}^3$, $\mathbf{B} \in \mathbb{R}^{3x3}$, and $\mathbf{c} \in \mathbb{R}^3$.

In particular, this method of trajectory transfer uses a thin-plate spline to find a mapping between these two scenes. A TPS minimizes the curvature of the mapping function and tries to find a mapping, or warping, that is close to rigid. This is motivated by the observation that, for many tasks, success in a task is preserved under Euclidean transformation. Furthermore, the associated optimization problem can be solved efficiently [4].

Schulman et al. [3] leverage thin plate splines to perform trajectory transfer. Using a point cloud representation of both scenes, they find a TPS that maps from the demonstration scene to the new scene. The transformation function is used to warp the path traced by the end effector of the robot in the demonstration, represented as a sequence of end effector poses. The warped trajectory is executed, with the hope that the registration will account for changes in the environment but maintain the important aspects of the manipulation.

We consider the case where an agent has access to a library of expert demonstrations. Such a library enables task execution from different initial conditions and provides robustness to different types of environmental variation. With many demonstrations, the decision making problem is to select which demonstration trajectory to transfer.

We approach the problem of selecting a trajectory to transfer as an abstract MDP. Our base manipulation task is an MDP with high-dimensional, continuous state and action spaces. For a knot-tying task, the state space is the joint state of the robot and the rope. The action space is the set of torques that can be applied at the motors. With a library of demonstrations, we can reduce this problem to an abstract MDP where the state space is the same, but the actions correspond to selecting a trajectory from the library and transferring it to the current state. This abstracts the problem temporally and significantly reduces the size of the action space. Actions follow a sequence of waypoints, thus significantly reducing the time horizon to be considered. Furthermore, because there are finitely many demonstrations, the continuous action space is reduced to a finite set of options.

Using this formalization, we learn an approximate Q-function for the abstract MDP. This is still a continuous-state reinforcement learning problem, so we propose a method to do this learning with human input. The procedure combines maximum-margin structured prediction with approximate linear programming to learn a Q-function that 1) matches a human demonstrator's Q-function and 2) is consistent with the dynamic programming equations for the MDP.

III. Experiments and Benchmark

We developed a knot-tying benchmark for evaluating the effectiveness of our proposed approach. This benchmark is available at sites.google.com/site/rss2014mmql). It contains the 148 pairs of point clouds and demonstration gripper trajectories used in Schulman et al. [3]. It also contains a training set and test set of new initial rope configurations, generated by randomly selecting an initial rope configuration from the demonstrations and perturbing it. Each rope configuration in the training set is associated with a demonstration to transfer selected by a human expert.

Policy	Success Rate
Nearest neighbor [3]	68.8%
Greedy	85.6%
Lookahead (depth 1, width 10)	93.6%
Lookahead (depth 2, width 5)	95.2%

TABLE I: Success rate of tying a knot using the expert-labeled examples. The nearest-neighbor method selects the demonstration that minimizes a bi-directional registration cost associated with the trajectory transfer. Other policies maximize a learned Q-function. Greedy maximizes this value in the current state. The lookahead policies act to maximize a back up value computed by beam search with the specified parameters. The greedy succeeds in an additional 17% of examples when compared with the baseline. Lookahead policies achieve very high performance rates and approach the best possible with our demonstration library.

Our experiments are carried out using Bullet Physics to simulate the dynamics. We define success as tying a knot within 5 steps. We explored the performance of two policies: 1-step greedy maximization of the learned Q-function and using a beam search to maximize the learned Q-function over a search horizon. We compared to the nearest-neighbor approach described in Schulman et al. [3]. The success rates obtained under these policies are summarized in Table I. Note that our best results surpass the baseline by 26.4%.

We find that this approach can offer significant improvements over the nearest-neighbor method in simulation. We have initial experiments that indicate robustness to modeling error for the lookahead simulation. We are currently in the process of testing this on a PR2 robot and are in the process of extending this approach to other deformable object tasks.

References

[1] J. C. Carr, R. K. Beatson, J. B. Cherrie, T. J. Mitchell, W. R. Fright, B. C. McCallum, and T. R. Evans. Reconstruction and Representation of 3D Objects with Radial Basis Functions. In *Computer Graphics (SIGGRAPH 01 Conf. Proc.), pages 6776. ACM SIGGRAPH*, pages 67–76. Springer, 2001.

[2] John Schulman, Ankush Gupta, Sibi Venkatesan, Mallory Tayson-Frederick, and Pieter Abbeel. A Case Study of Trajectory Transfer Through Non-Rigid Registration for a Simplified Suturing Scenario. In *Proceedings of the 26th IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2013.

[3] John Schulman, Jonathan Ho, Cameron Lee, and Pieter Abbeel. Learning from Demonstrations through the Use of Non-Rigid Registration. In *Proceedings of the 16th International Symposium on Robotics Research (ISRR)*, 2013.

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