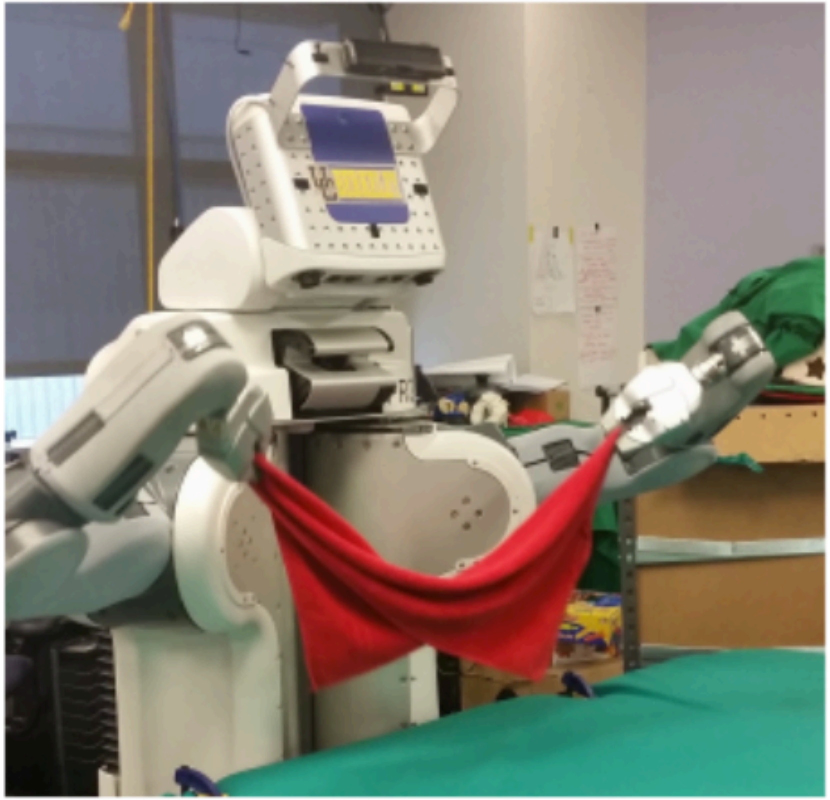


Learning to Manipulate from Demonstrations

CS287

November 17, 2015

Sandy Huang

Slides courtesy of Pieter Abbeel

Personal Robotics Hardware

PR2
Willow Garage
$400,000
2009

Baxter
Rethink Robotics
$30,000
2013

UBR-1
Unbounded Robotics
$35,000
2013

?

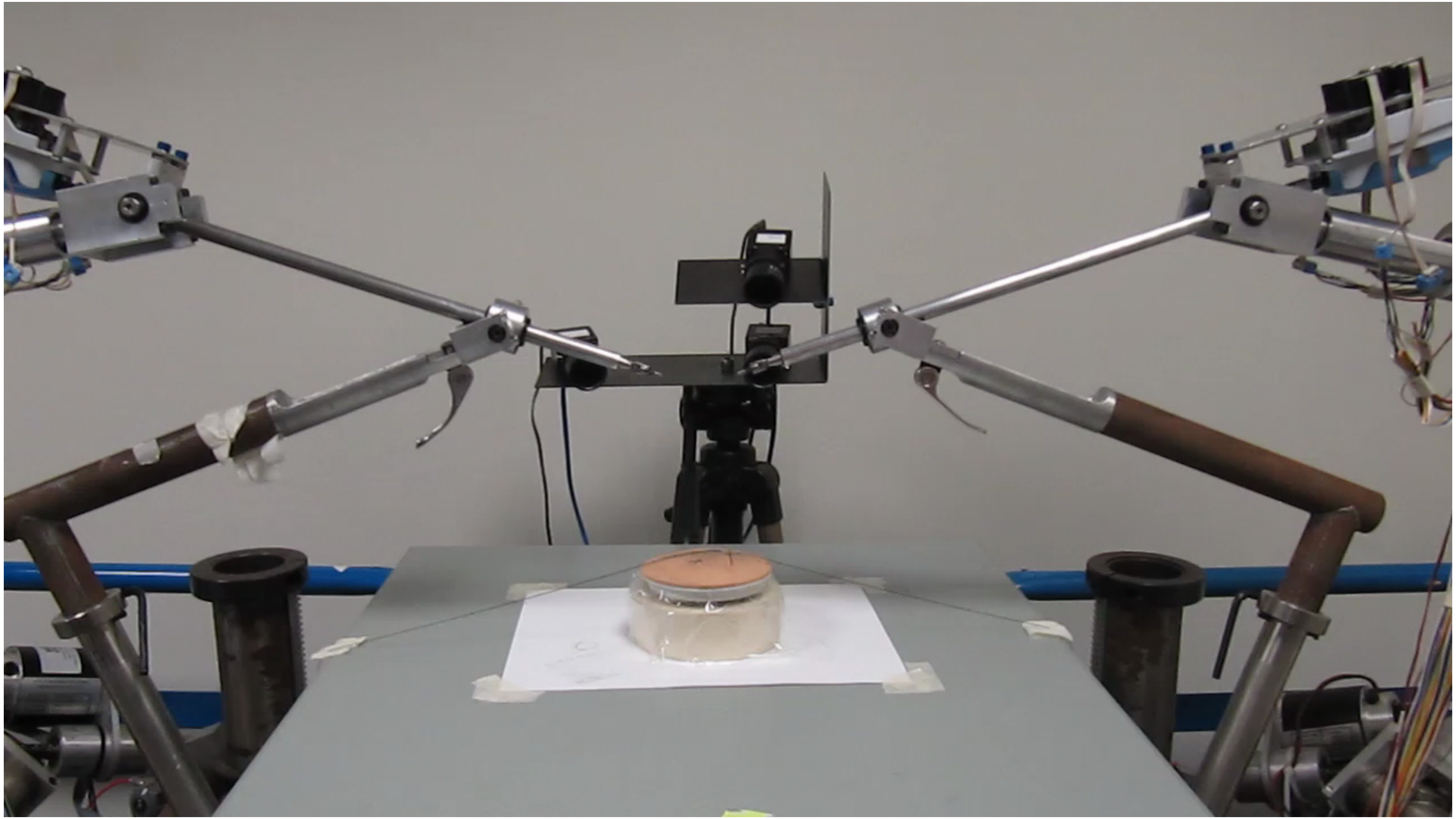
?
$2,000 ?
2017?

Challenge Task: Robotic Laundry

[Maitin-Shepard, Cusumano Towner, Lei, Abbeel, ICRA 2010]

How About...

Surgical Knot Tie



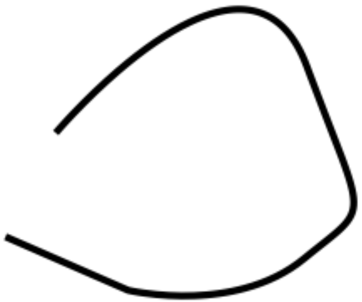
[van den Berg, Miller, Duckworth, Humphrey, Wan, Fu, Goldberg, Abbeel, Best Medical Robotics Paper, ICRA 2010]

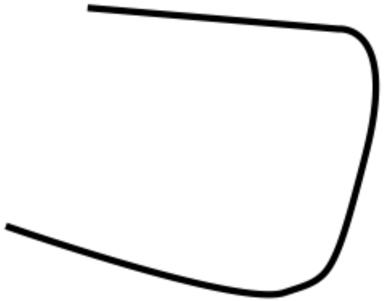
Surgical Knot Tie

- Open loop
- If careful about initial conditions
  - 50% success rate

Learning from Demonstrations

- The problem
  - Human demonstrated knot-tie in this rope
  - Robot has to tie a knot in this rope



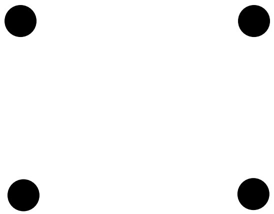


Generalizing Trajectories

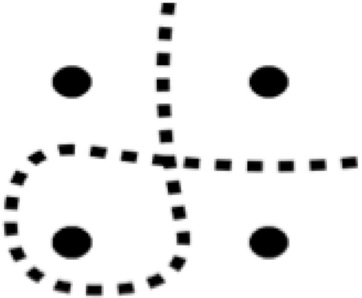
- Prior work
  - Billard, Calinon and collaborators
    - Gaussian Mixture Models (GMM) and Gaussian Mixture Regression (GMR)
  - Schaal and collaborators
    - Dynamic motion primitives
  - Cakmak, Thomaz and collaborators
    - Human robot interaction for robot to learn faster
  - Peters and collaborators
    - Stay close to demonstrations distribution while also optimizing reward
- BUT
  - All of these algorithms have underlying representations in terms of coordinates
  - Can we alleviate need to specify coordinate frames / features and directly adapt to geometry?

Cartoon Problem Setting

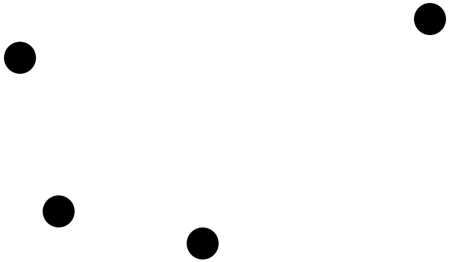
Training scene



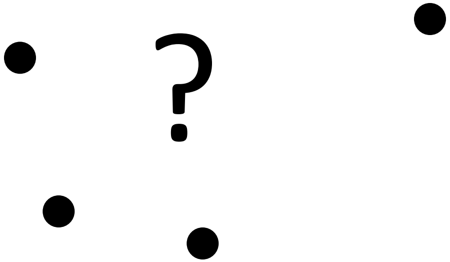
Trajectory demonstrations



Test scene



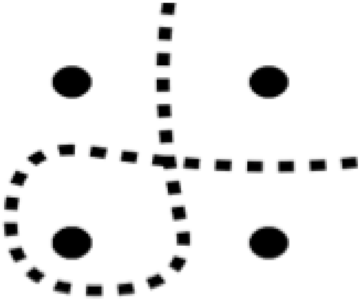
What trajectory here?

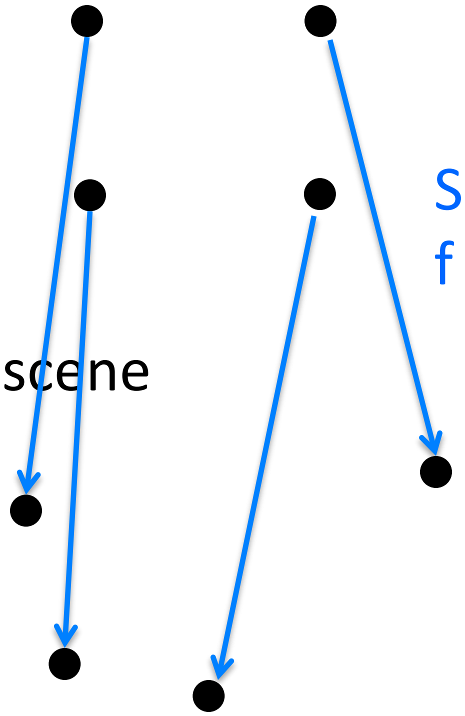


Cartoon Problem Setting

Training scene

Trajectory demonstrations

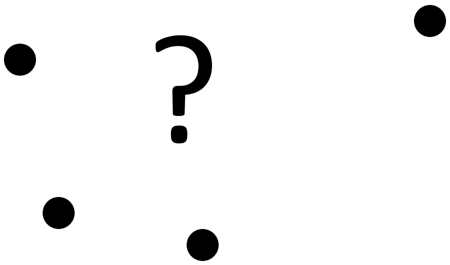




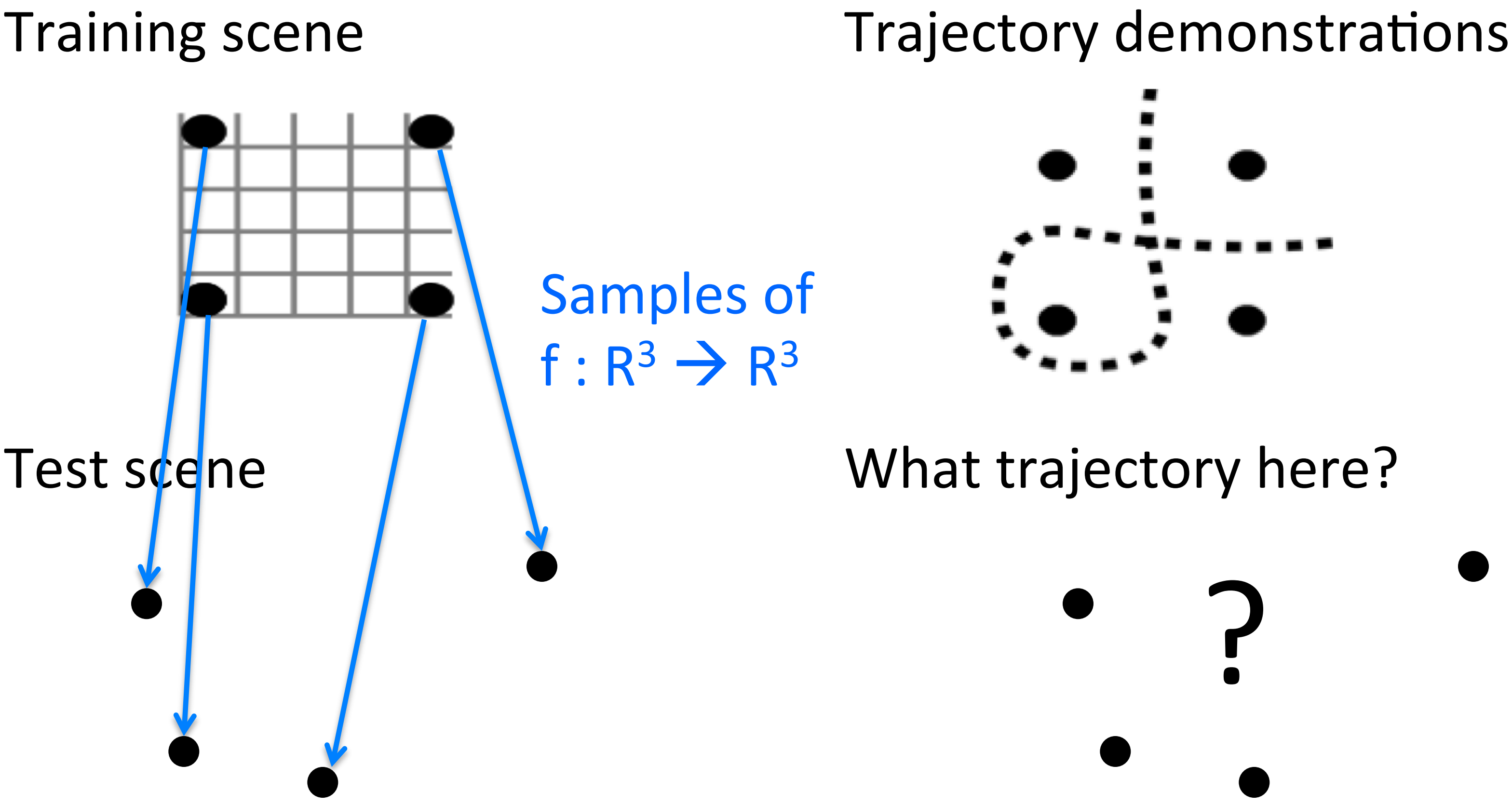
Samples of
$f : R^3 \rightarrow R^3$

Test scene

What trajectory here?



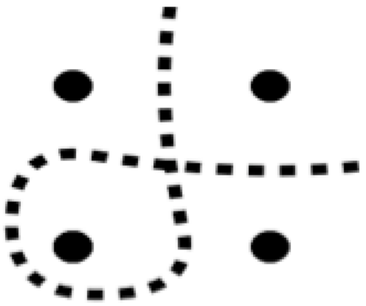
Cartoon Problem Setting

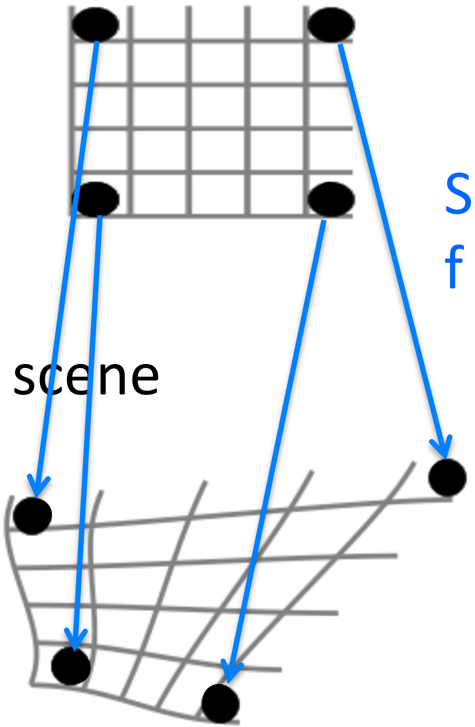


Cartoon Problem Setting

Training scene

Trajectory demonstrations

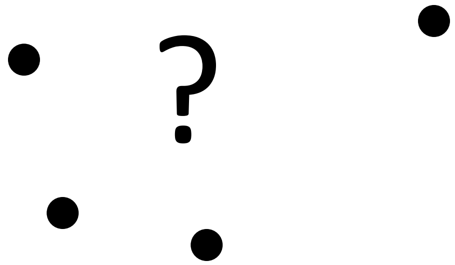




Samples of
$f : R^3 \rightarrow R^3$

Test scene

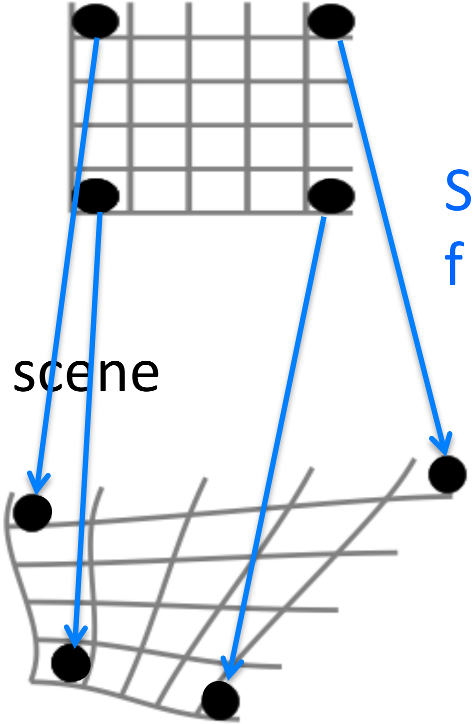
What trajectory here?

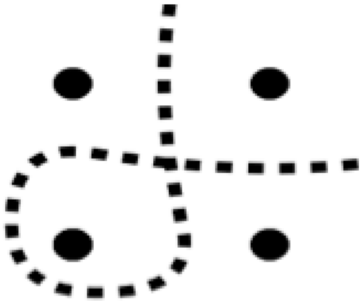


Cartoon Problem Setting

Training scene

Trajectory demonstrations

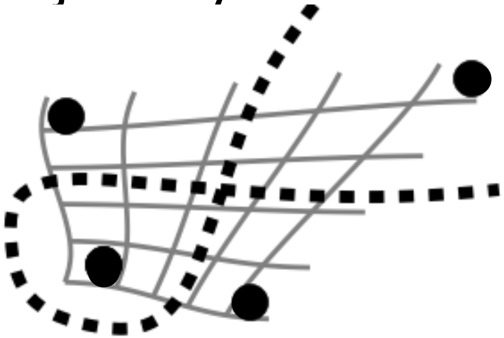




Samples of
$f : R^3 \rightarrow R^3$

Test scene

What trajectory here?



Learning f : $R^3 \rightarrow R^3$ from Samples

$$\min_{f \in \{\mathbb{R}^3 \rightarrow \mathbb{R}^3\}} \int_{x \in \mathbb{R}^3} \|D^2 f(x)\|^2_{\text{Frob}} dx$$

$$\text{s.t.} \quad f(x^{(i)}_{\text{train}}) = x^{(i)}_{\text{test}} \quad \forall i \in 1, \ldots, m$$

- Translations, rotations and scaling are FREE

Learning f : R³ → R³ from Samples

$$\min_{f \in \{\mathbb{R}^3 \to \mathbb{R}^3\}} \quad \int_{x \in \mathbb{R}^3} \|D^2 f\|^2_{\text{Frob}}(x)dx$$

$$\text{s.t.} \quad f(x^{(i)}_{\text{train}}) = x^{(i)}_{\text{test}} \quad \forall i \in 1, \ldots, m$$

- Solution has form:

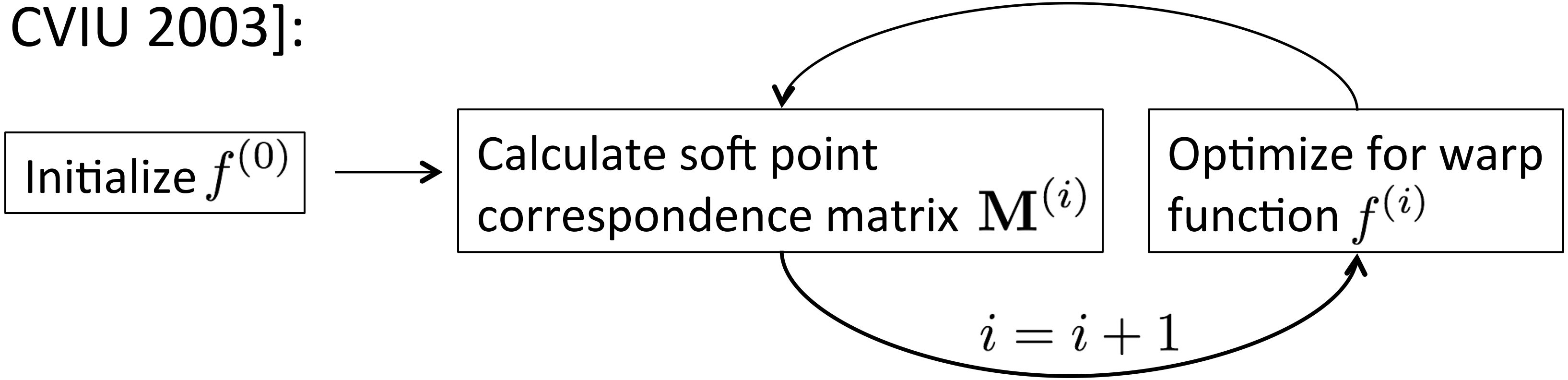
$$f(x) = \sum_{i=1}^{m} a_i K(x_i, x) + b^\top x + c,$$

$$K(x,y) = \begin{cases} c_0 r^{4-d} \ln r, \; d = 2 \text{ or } d = 4 \\ c_1 r^{4-d}, \; \text{otherwise} \end{cases} \quad \text{with} \quad r = \|x - y\|_2.$$

Wahba, Spline models for observational data. Philadelphia: Society for Industrial and Applied Mathematics. 1990.
Evgeniou, Pontil, Poggio, Regularization Networks and Support Vector Machines. Advances in Computational Mathematics. 2000.
Hastie, Tibshirani, Friedman, Elements of Statistical Learning, Chapter 5. 2008.

Finding a Non-Rigid Registration

- Thin Plate Spline Robust Point Matching (TPS-RPM) [Chui et al. CVIU 2003]:



- Variant of Expectation-Maximization (EM); finds locally optimal warp

Trajectory Transfer Procedure

- Using non-rigid registration, find a transformation **f** from training scene to test scene
- Apply **f** to the demonstrated end-effector trajectory

$$\mathbf{p}_t \to \mathbf{f}(\mathbf{p}_t)$$

$$\mathbf{R}_t \to \mathrm{orth}\left(\mathbf{J}_{\mathbf{f}}(\mathbf{p}_t)\mathbf{R}_t\right)$$

- Convert the end-effector trajectory to a joint trajectory

$$\underset{\boldsymbol{\theta}_1,\ldots,\boldsymbol{\theta}_T}{\text{minimize}} \left[\sum_{t=1}^{T-1} \|\boldsymbol{\theta}_{t+1} - \boldsymbol{\theta}_t\|^2 + \mu \sum_{t=1}^{T} \left\|\mathrm{err}\left(\tilde{\mathbf{T}}_t^{-1} \cdot \mathrm{fk}(\boldsymbol{\theta}_t)\right)\right\|_{\ell_1}\right]$$

subject to

No collisions, with safety margin d_{safe}

$$\boldsymbol{\theta}_{\min} \leq \boldsymbol{\theta}_{1:T} \leq \boldsymbol{\theta}_{\max} \qquad \text{(Joint limits)}$$

[J. Schulman, J. Ho, C. Lee, P. Abbeel, ISRR 2013]

Robot Experiments

- Knots tied
  - Overhand
  - Figure-eight
  - Double-overhand
  - Square
  - Clove-hitch

Experiment: Knot-Tie

[J. Schulman, J. Ho, C. Lee, P. Abbeel, ISRR 2013]

Evaluation

Knot type	Segments	Success ($d_{pert} = 3\,\text{cm}$)	($d_{pert} = 10\,\text{cm}$)
Overhand	3	5/5	4/5
Figure-eight	4	5/5	1/5
Dbl-overhand	5	3/5	3/5
Square	6	5/5	3/5
Clove hitch	4	1/5	0/5

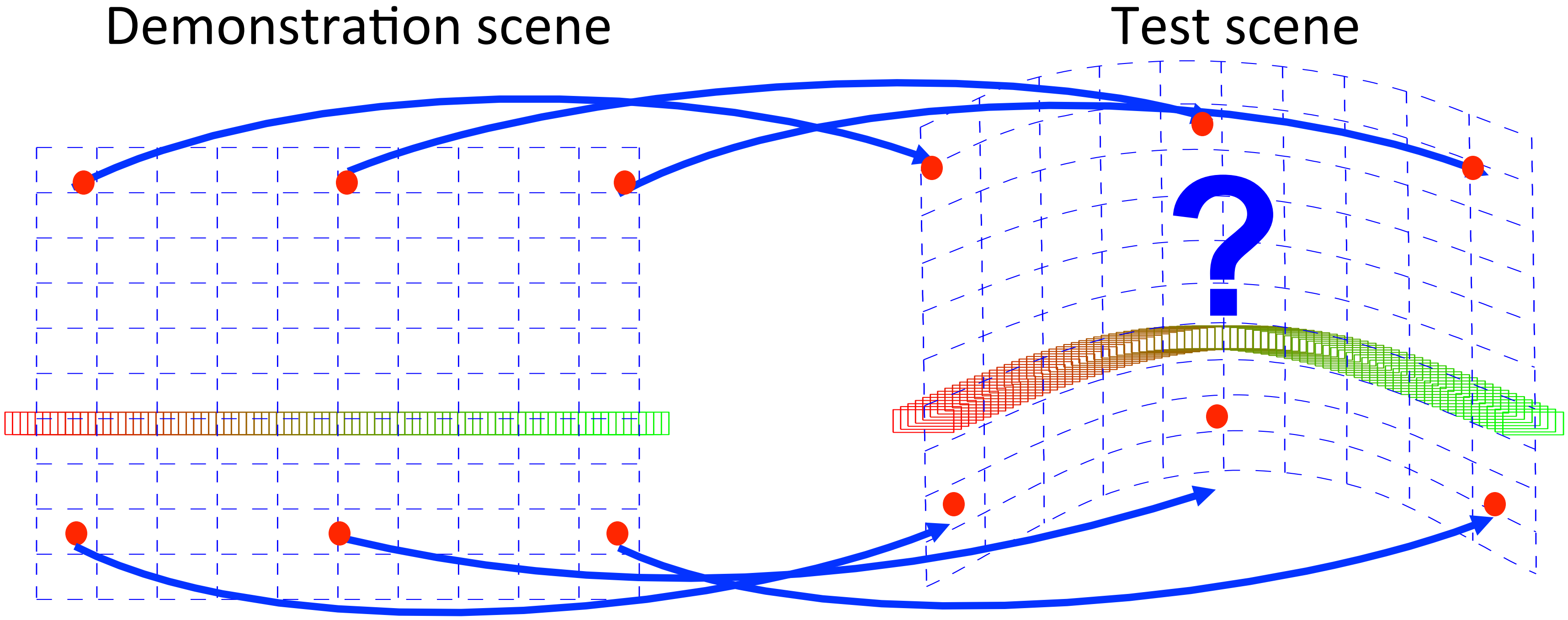
Experiment: Suturing

[J. Schulman, A. Gupta, S. Venkatesan, M. Tayson-Frederick, P. Abbeel, IROS 2013]

Limitations of Trajectory Transfer

- Does not consider joint limits and obstacles when finding the warp function
- Computationally expensive with >100 demonstrations
- Ignores surface normals when finding the warp function
- Only uses geometric information of the objects, not appearance information

Trajectory Transfer: First Step

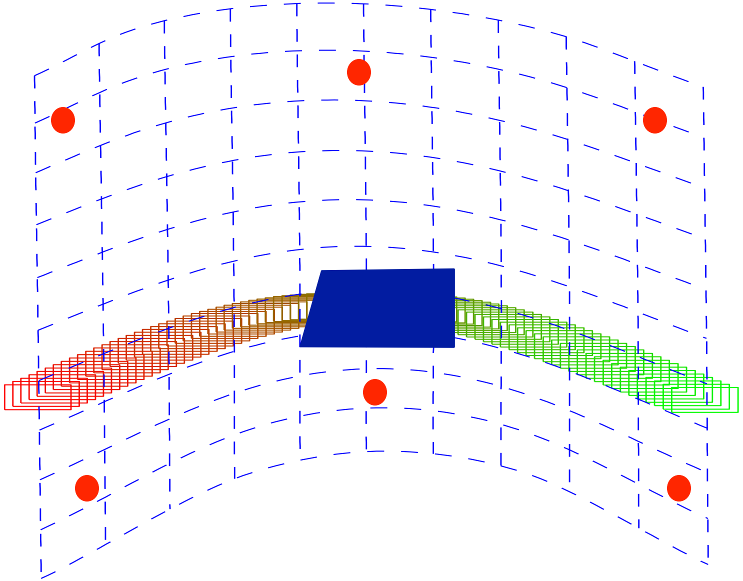


Step 1: $\min_{f \in \text{registration functions}} \text{registration\_error}\left(S_{\text{demo}}, S_{\text{test}}\right) + \text{bending\_energy}\left(f\right)$

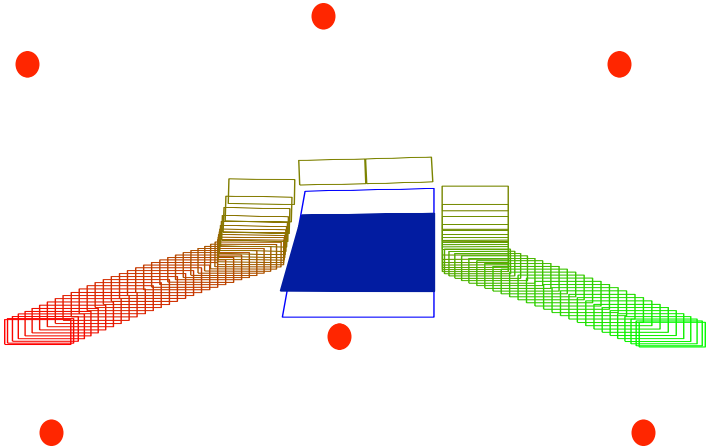
$\tau_f \leftarrow f\left(\tau_{\text{demo}}\right)$

Trajectory Transfer: Second Step

Transferred trajectory



Feasible trajectory

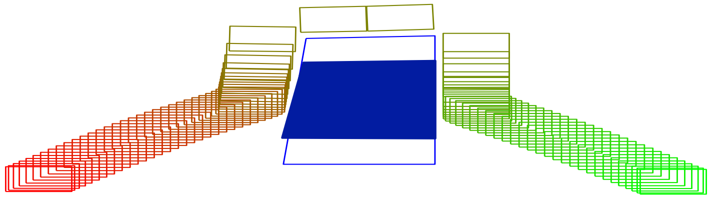


Step 2: $\min_{\tau \in \text{trajectories}} \quad \text{trajectory\_error}(\tau_f, \tau)$

s.t. τ is feasible and collision-free

Unifying Trajectory Transfer

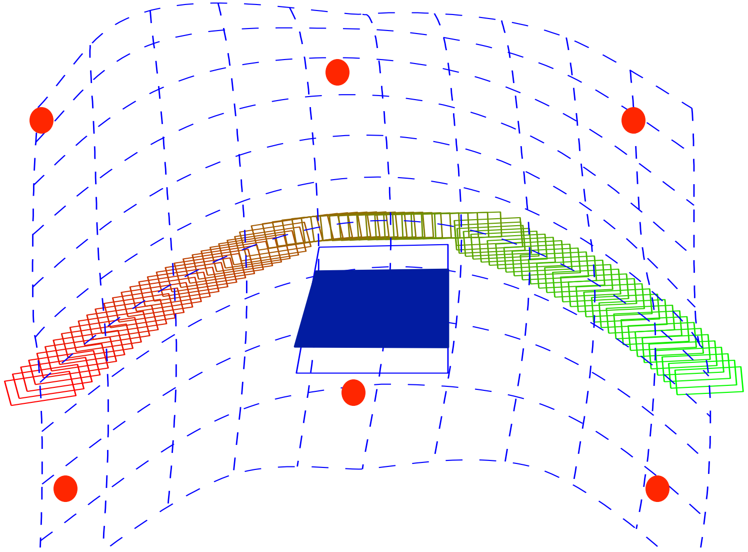
Two-step optimization



Step 1: $$\min_{f \in \text{registration functions}} \text{registration\_error}\left(S_{\text{demo}}, S_{\text{test}}\right) + \text{bending\_energy}\left(f\right)$$

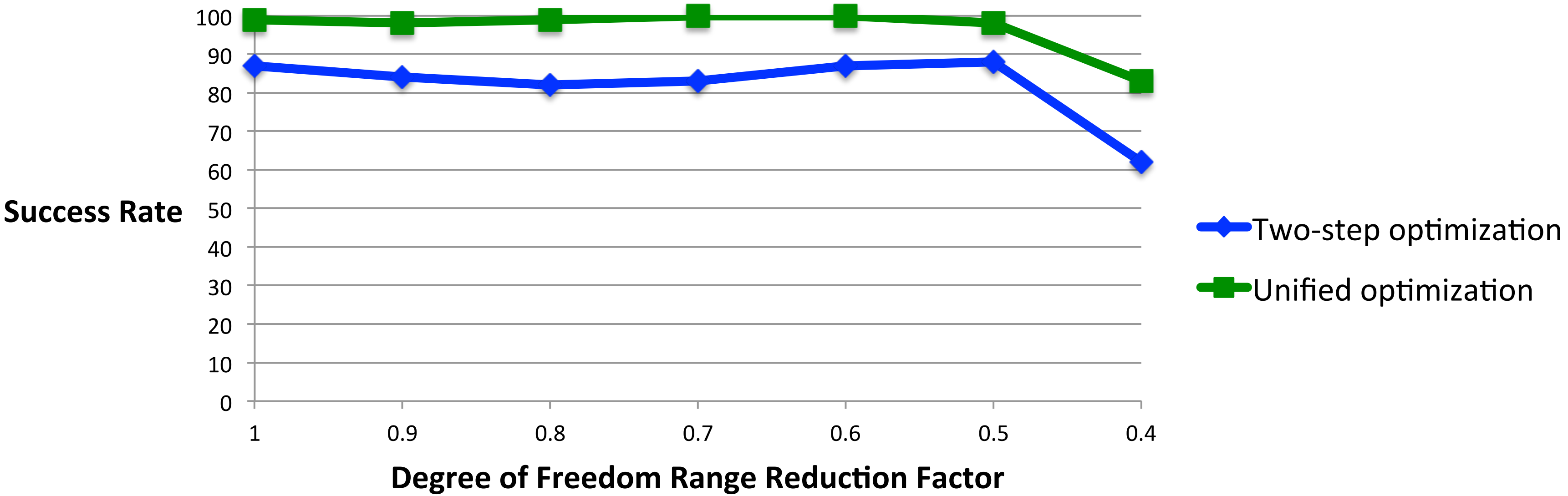
Step 2: $$\min_{\tau \in \text{trajectories}} \text{trajectory\_error}\left(f\left(\tau_{\text{demo}}\right), \tau\right)$$
$$\text{s.t.} \quad \tau \text{ is feasible and collision-free}$$

Unified optimization



$$\min_{\substack{f \in \text{registration functions} \\ \tau \in \text{trajectories}}} \text{registration\_error}\left(S_{\text{demo}}, S_{\text{test}}\right) + \text{bending\_energy}\left(f\right) + \text{trajectory\_error}\left(f\left(\tau_{\text{demo}}\right), \tau\right)$$
$$\text{s.t.} \quad \tau \text{ is feasible and collision-free}$$

Application to Manipulation of Deformable Objects



[A. Lee, S. Huang, D. Hadfield-Menell, E. Tzeng, P. Abbeel, IROS 2014]

Theoretical Guarantees

- Can be expected to work if the dynamics of the system are approximately covariant under sufficiently smooth warpings.

$$
\begin{array}{ccc}
\{\text{STATE}^{t}_{\text{train}}\} & \xrightarrow{\Pi_{\text{TRAJ}}} & \{\text{STATE}^{t+1}_{\text{train}}\} \\
\mathbf{f}\downarrow & & \downarrow\mathbf{f} \\
\{\text{STATE}^{t}_{\text{test}}\} & \xrightarrow{\Pi_{\mathbf{f}(\text{TRAJ})}} & \{\text{STATE}^{t+1}_{\text{test}}\}
\end{array}
$$

Nearest-Neighbor Policy for Tasks

- Repeat
  - Acquire new point cloud X_{test}
  - Using non-rigid registration compute distance between X_{test} and each point cloud $X_{train,i}$ from demonstrations

$$i^* = \arg\min_i \mathrm{Distance}(X_{\mathrm{test}}, X_{\mathrm{train},i})$$

  - If i* is a "done" state, **break**
  - Apply trajectory transfer to generate new trajectory

$$\mathrm{Traj}_{\mathrm{test}} = \mathrm{TrajectoryTransfer}(\mathrm{State}_{i^*}, \mathrm{Traj}_{i^*}, \mathrm{State}_{\mathrm{test}})$$

Limitations of the Nearest-Neighbor Policy

- Doesn’t account for demonstration quality
- Doesn’t prefer moves that make progress
- Doesn’t account for reachability of trajectory

Learning to Choose Better Actions

s: state of system (robot, rope, ...)

a: chosen demonstration

$$Q(s, a) = \phi(s, a) \cdot w$$

policy: $$\hat{a} = \arg\max_{a} \boldsymbol{\phi}(s, a) \cdot \mathbf{w}$$

features in ϕ

- action bias
- registration cost
- distance to "landmark" states

[D. Hadfield-Menell, A. Lee, C. Finn, E. Tzeng, S. Huang, P. Abbeel, ICRA 2015]

Max-Margin Policy Learning

$$\underset{\mathbf{w}}{\text{minimize}}\, L(\mathbf{w})$$

$$L(\mathbf{w}) = \underbrace{\frac{1}{2}\|\mathbf{w}\|^2}_{\text{Regularizer}} + \sum_i \left| \max_a [\underbrace{\phi(s_i, a) \cdot \mathbf{w} + \Delta(a, a_i)}_{\text{score + margin of action a}}] - \underbrace{\phi(s_i, a_i) \cdot \mathbf{w}}_{\text{score of demonstrated action } a_i \text{ from state } s_i} \right|^+$$

Regularizer

score + margin of action a

score of demonstrated action a_i from state s_i

[D. Hadfield-Menell, A. Lee, C. Finn, E. Tzeng, S. Huang, P. Abbeel, ICRA 2015]

Max-Margin Q-Function Learning

$$\underset{\mathbf{w}}{\text{minimize}}\, L(\mathbf{w})$$

$$L(\mathbf{w}) = \frac{1}{2}\|\mathbf{w}\|^2 + \sum_i \left|\max_a[\phi(s_i, a)\cdot\mathbf{w} + \Delta(a, a_i)] - \phi(s_i, a_i)\cdot\mathbf{w}\right|^+$$

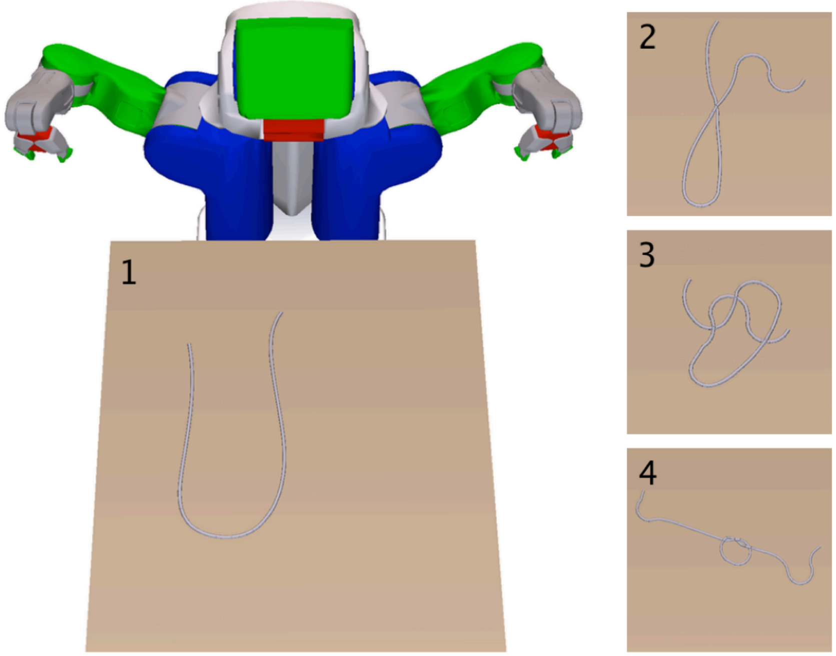
$$+ \sum_i \underbrace{|\phi(s_i, a_i)\cdot\mathbf{w} - (R(s_i, a_i) + \phi(s_i', a_i')\cdot\mathbf{w})|}$$

On-policy Bellman error

$$|Q(s,a) - (R(s,a) + Q(s',a'))|$$

[D. Hadfield-Menell, A. Lee, C. Finn, E. Tzeng, S. Huang, P. Abbeel, ICRA 2015]

Experiments



- 148 Demonstrated trajectories (on real robot)
- 1000 human-guided trajectory transfers in physics simulation
- Test on 500 new initial configurations

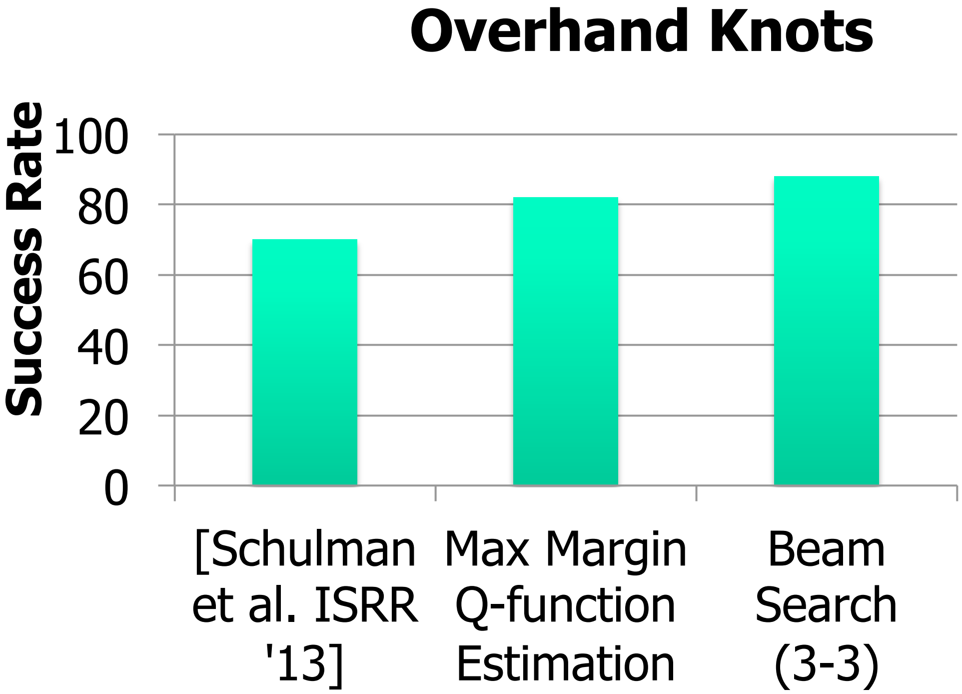
[D. Hadfield-Menell, A. Lee, C. Finn, E. Tzeng, S. Huang, P. Abbeel, ICRA 2015]

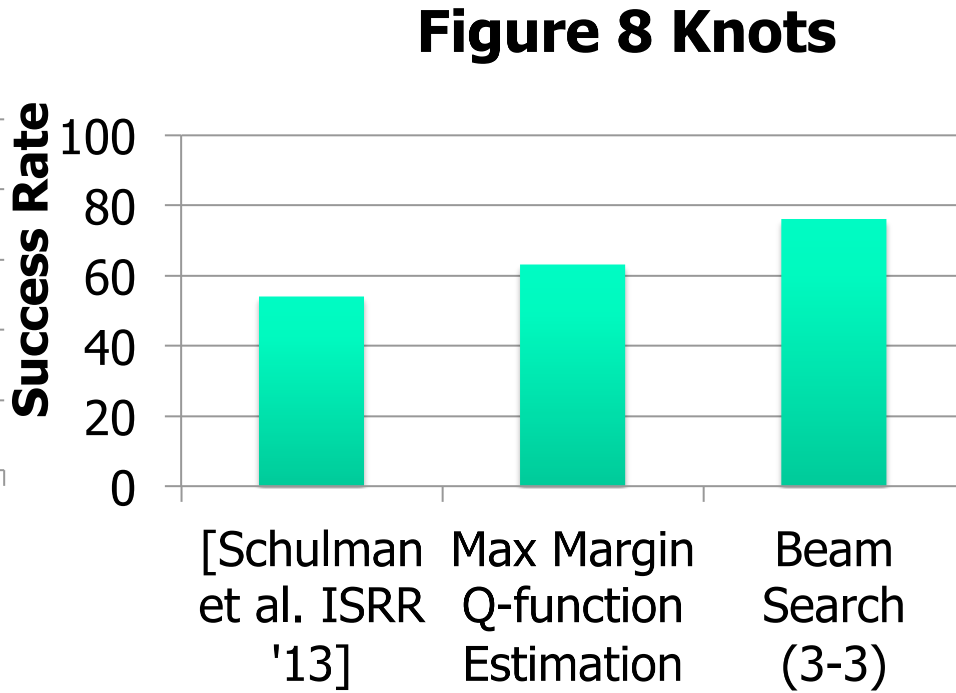
Results in Simulation

Policy	**Success rate**
Nearest neighbor	68.8%
$\text{argmax}_a\ Q(s,a)$	85.6%
Lookahead (d=1, w=10)	93.6%
Lookahead (d=2, w=5)	95.2%

[D. Hadfield-Menell, A. Lee, C. Finn, E. Tzeng, S. Huang, P. Abbeel, ICRA 2015]

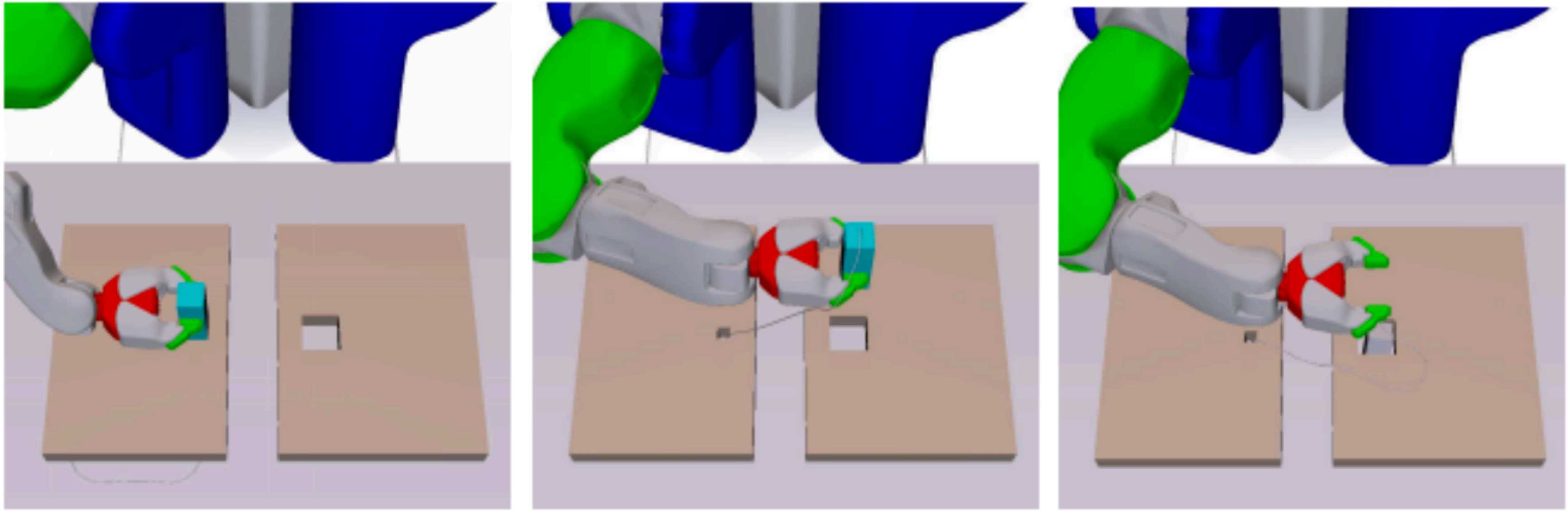
Evaluation on Knot-Tying



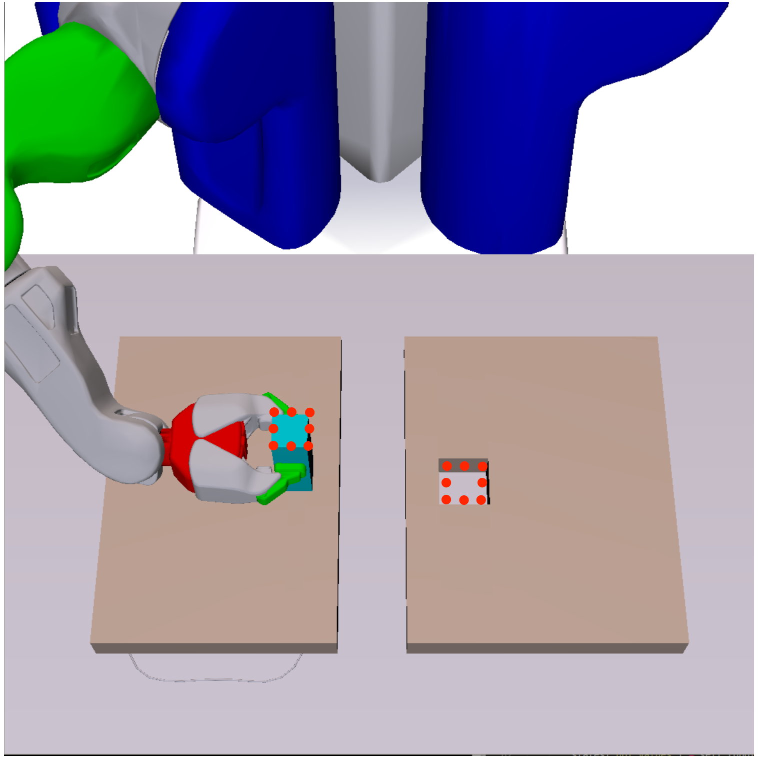


70%	82%	88%	54%	63%	76%

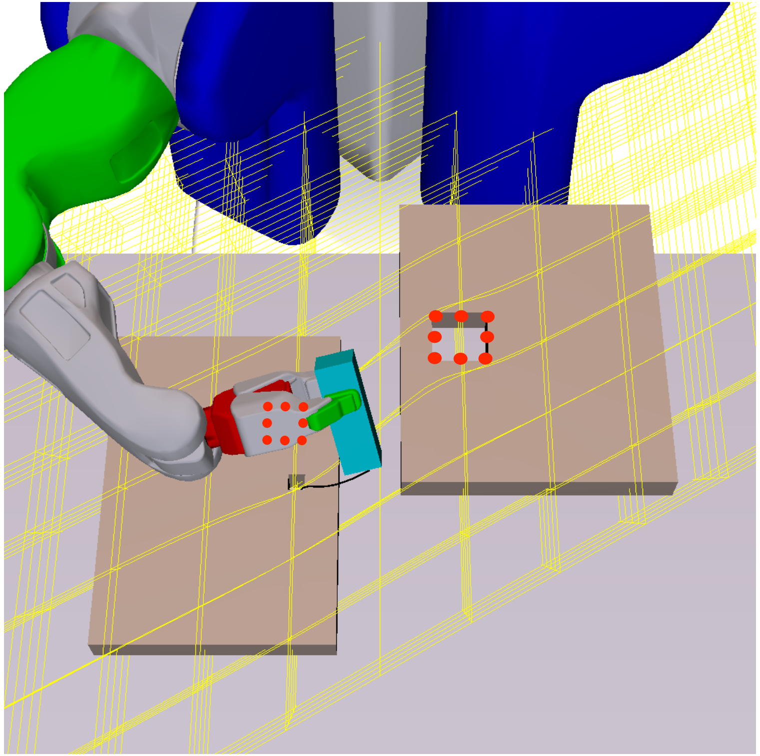
Motivation for Including Surface Normals



Standard TPS-RPM Registration

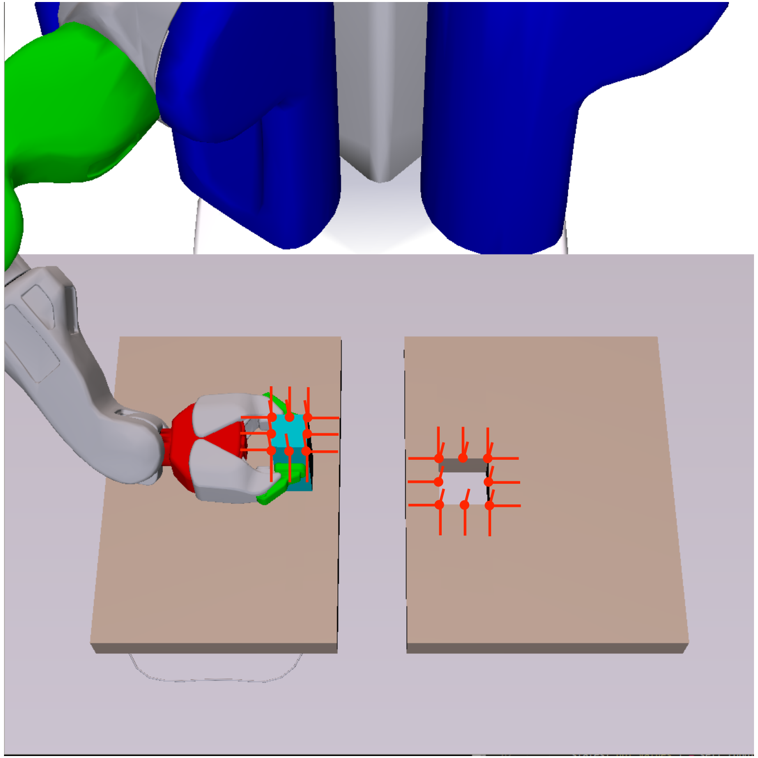


Demonstration scene

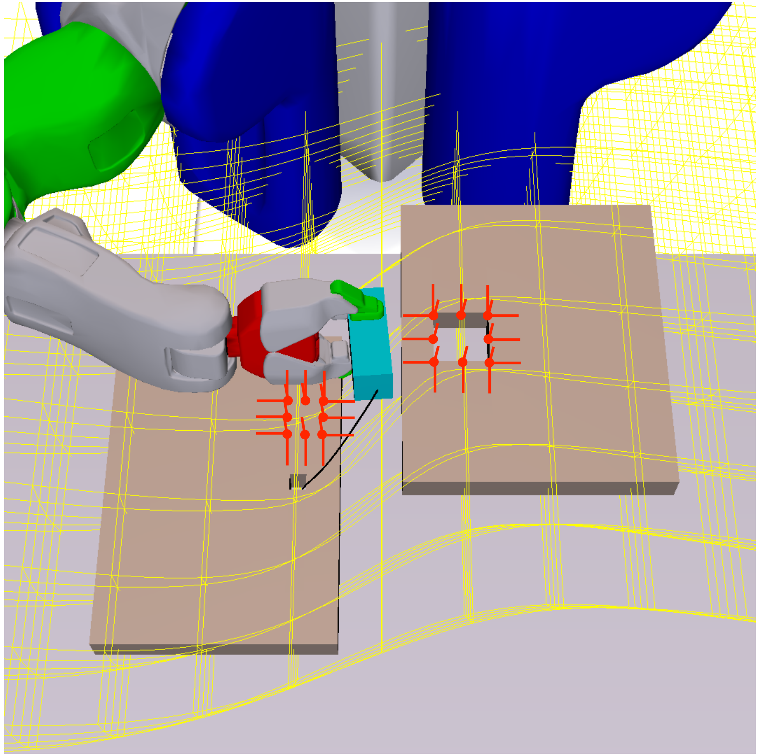


Test scene

TPS-RPM Registration with Normals



Demonstration scene



Test scene

[A. Lee, M. Goldstein, S. Barratt, P. Abbeel, ICRA 2015]

Problem Formulation

$$\min_{f,\mathbf{M},\mathbf{Q}} \quad E(f,\mathbf{M};T,\zeta) + \nu \tilde{E}(f,\mathbf{Q};\tilde{T},\tilde{\zeta}) + \lambda ||f||^2_{\text{TPS}}$$

$$\text{subject to} \quad \sum_{i=1}^{N+1} m_{ij} = 1, \sum_{j=1}^{N'+1} m_{ij} = 1, m_{ij} \geq 0$$

$$\sum_{k=1}^{K+1} q_{kl} = 1, \sum_{l=1}^{K'+1} q_{kl} = 1, q_{kl} \geq 0,$$

$$E(f,\mathbf{M};T,\zeta) = \sum_{i=1}^{N}\sum_{j=1}^{N'} m_{ij}||\mathbf{y}_j - f(\mathbf{x}_i)||_2^2 + T\sum_{i=1}^{N}\sum_{j=1}^{N'} m_{ij}\log m_{ij} - \zeta \sum_{i=1}^{N}\sum_{j=1}^{N'} m_{ij}$$

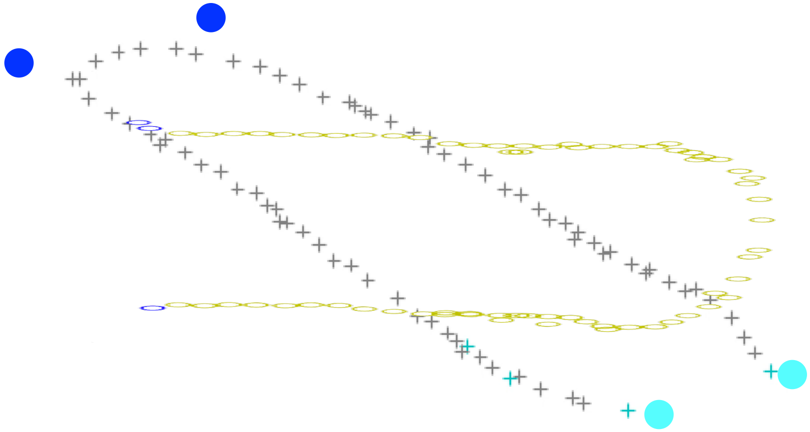
$$\tilde{E}(f,\mathbf{Q};\tilde{T},\tilde{\zeta}) = \sum_{k=1}^{K}\sum_{l=1}^{K'} q_{kl}||\mathbf{v}_l - \tfrac{1}{\beta_k}\mathbf{J}_f^{\mathbf{s}_k}\mathbf{u}_k||_2^2 + \tilde{T}\sum_{k=1}^{K}\sum_{l=1}^{K'} q_{kl}\log\frac{q_{kl}}{\pi_{kl}} - \tilde{\zeta}\sum_{k=1}^{K}\sum_{l=1}^{K'} q_{kl}$$

TPS-RPM: Sensitivity to Initialization

- Only uses geometric information to find non-rigid registration

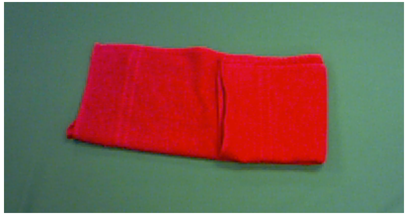
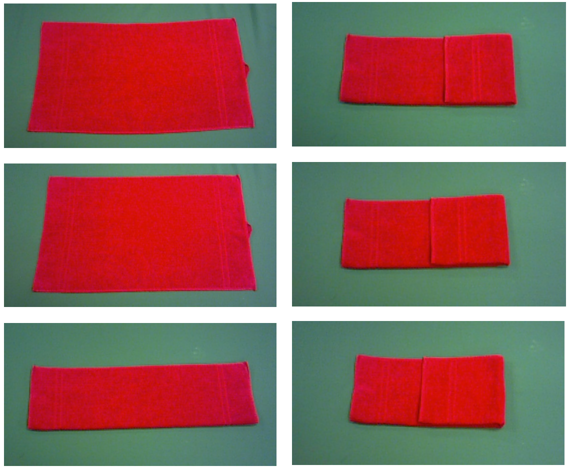
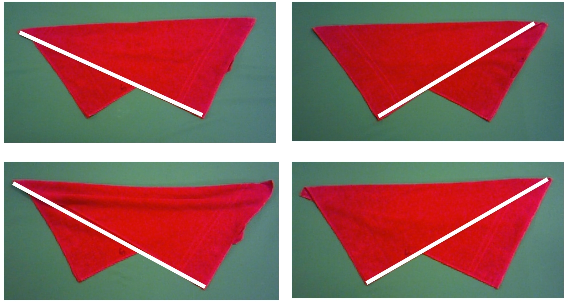
Demo

Test



Geometric Similarity ≠ Semantic Similarity

- Demonstration selection also only uses geometric information

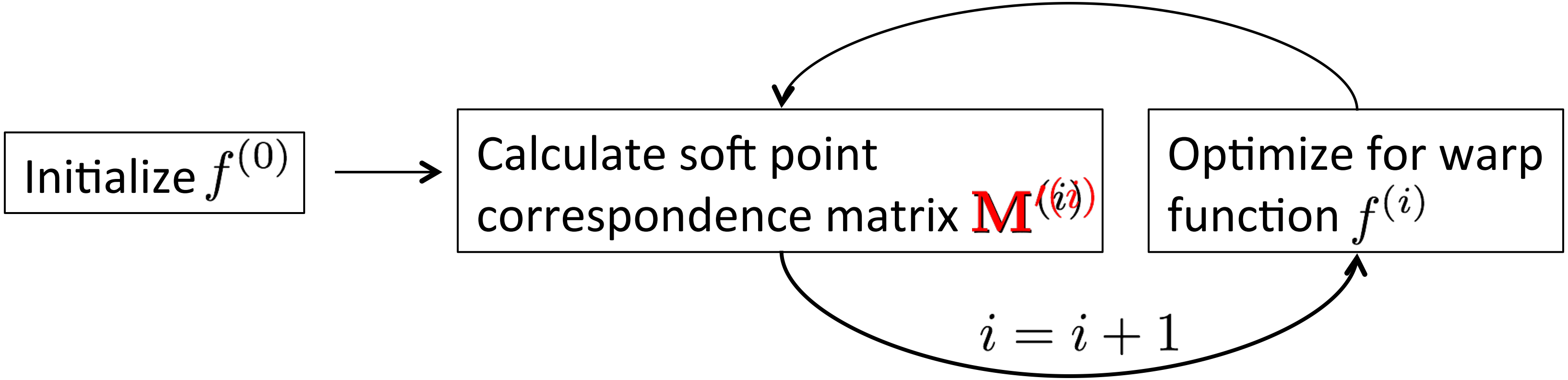
Test configuration		
Geometrically-similar demonstration configurations		

Convolutional Neural Net Classification

- corners-against-background
- edges-against-background
- edges-against-interior
- folds-against-background
- flat interior
- wrinkled interior

[S. Huang, J. Pan, G. Mulcaire, P. Abbeel, IROS 2015]

Leveraging Appearance Information



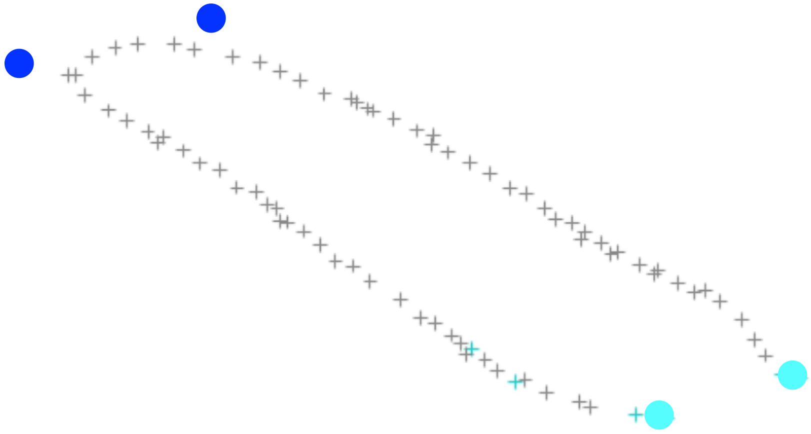
- $\mathbf{M}_{ij}$ = correspondence between source point $\mathbf{x}_i$ and target point $\mathbf{y}_j$
- π_{ij} = prior probability that $\mathbf{x}_i$ and $\mathbf{y}_j$ should be matched
- Define the new point correspondence matrix as $\mathbf{M}'_{ij} = \pi_{ij}\mathbf{M}_{ij}$
- Normalize $\mathbf{M}'$ so that the rows and columns sum to 1

Trajectory Transfer + Appearance Priors

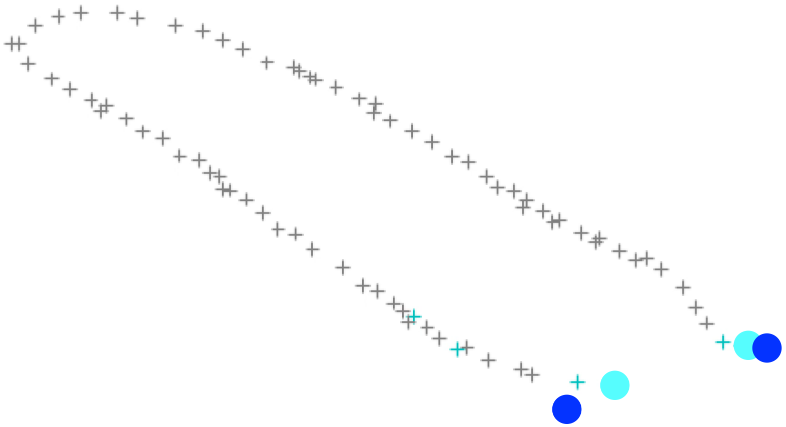
Demo

Test

Without appearance priors



With appearance priors



TPS-RPM with CNN Classification of Pixels

[S. Huang, J. Pan, G. Mulcaire, P. Abbeel, IROS 2015]

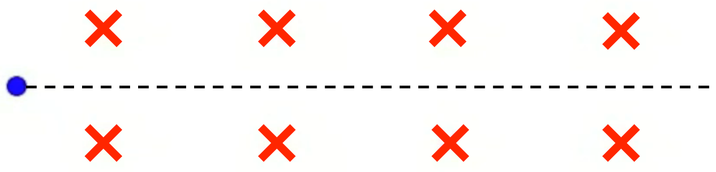
Current Directions

- Unsupervised features in registration
- Reinforcement learning to further improve performance
- Forces and torques (to extend to non-kinematic tasks)
- More data…

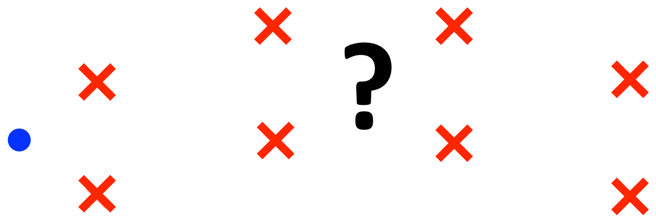
Thank you

Trajectory Transfer: Toy Example

Demonstration



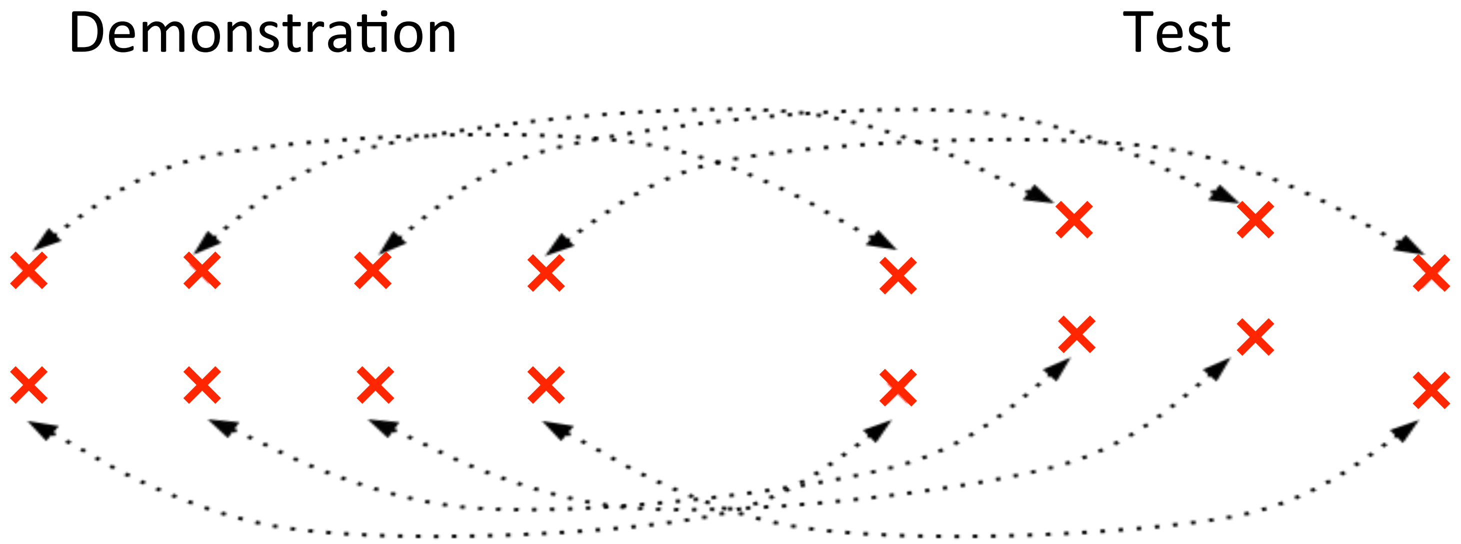
Test



Schulman et al. ISRR 2013

Trajectory Transfer: Toy Example

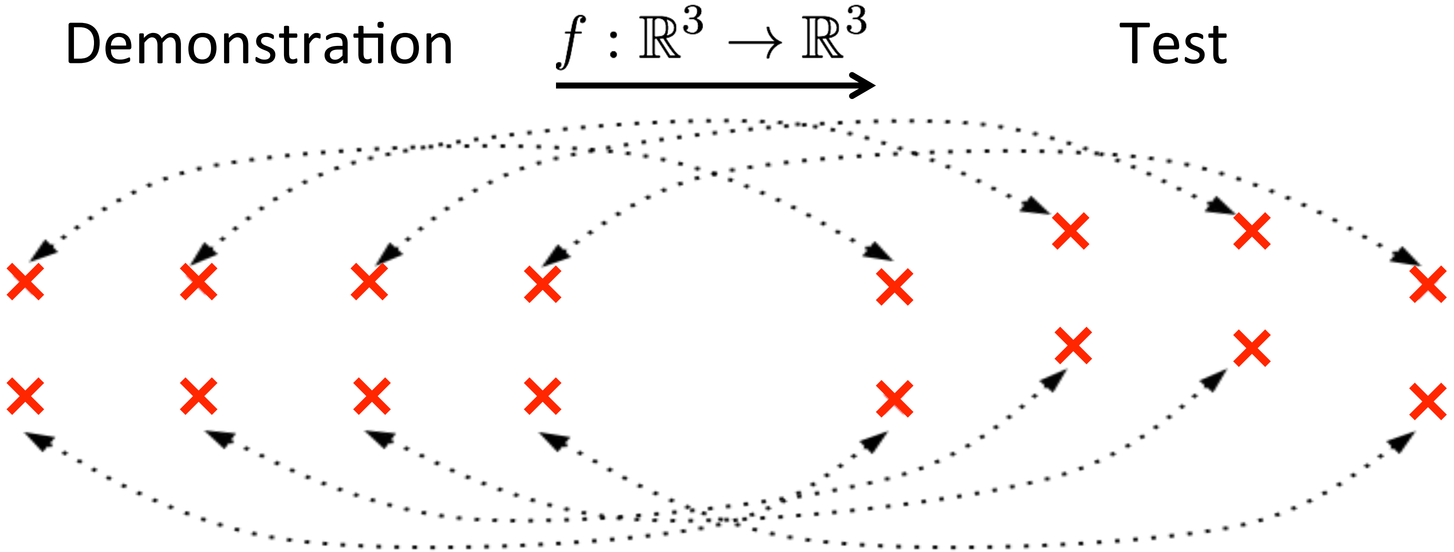
1. Calculate a non-rigid registration



Schulman et al. ISRR 2013

Trajectory Transfer: Toy Example

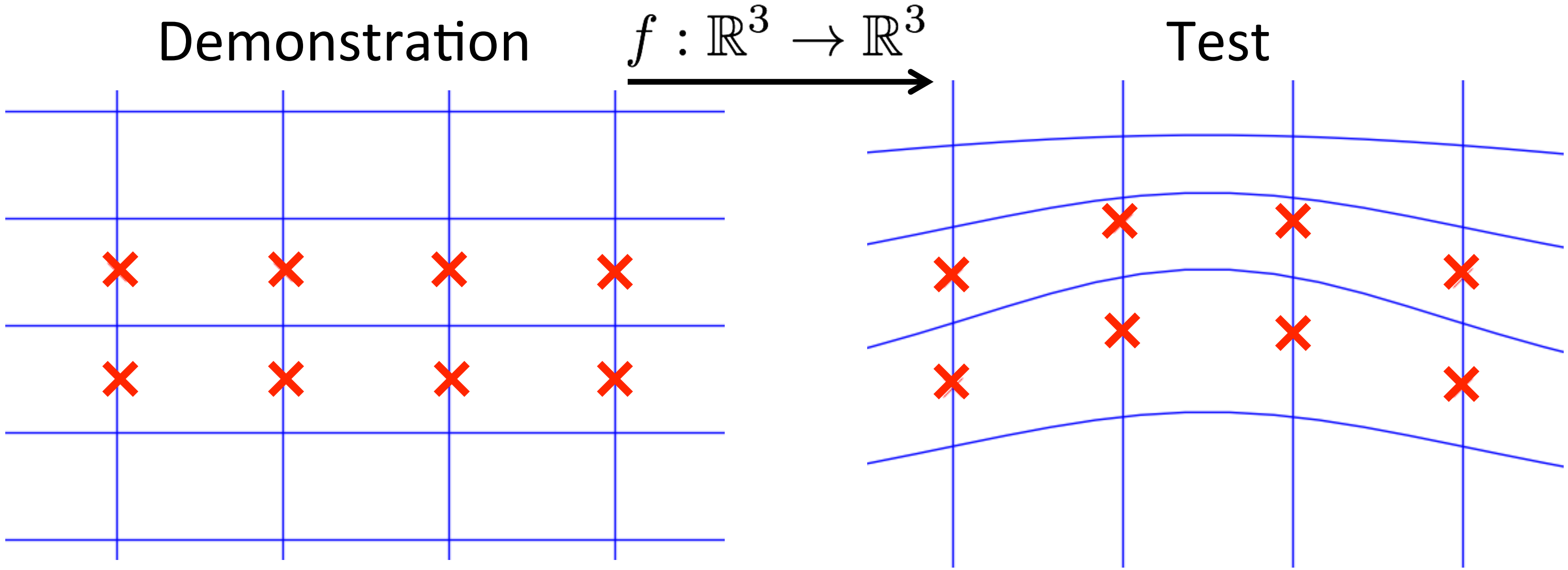
1. Calculate a non-rigid registration



Schulman et al. ISRR 2013

Trajectory Transfer: Toy Example

1. Calculate a non-rigid registration

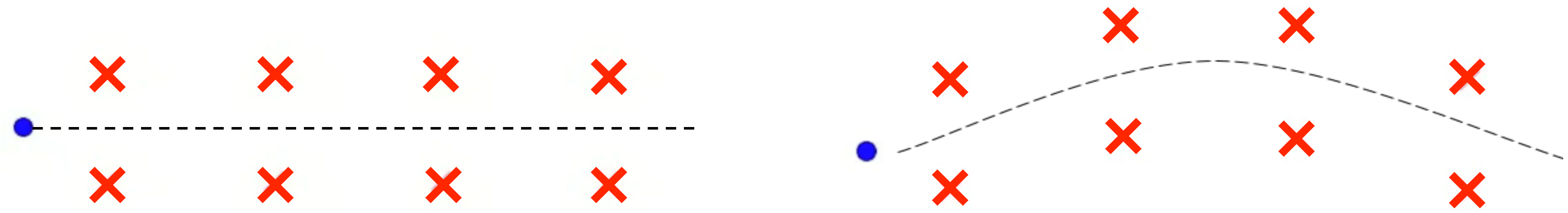


Schulman et al. ISRR 2013

Trajectory Transfer: Toy Example

2. Apply f to the demonstrated trajectory

Demonstration $\xrightarrow{f : \mathbb{R}^3 \to \mathbb{R}^3}$ Test



Schulman et al. ISRR 2013