Learning Articulated Motions From Visual Demonstration

Sudeep Pillai
spillai@csail.mit.edu

Matthew Walter
mwalter@csail.mit.edu

Seth Teller
teller@csail.mit.edu

Robotics, Vision and Sensor Networks Group, CSAIL, MIT

MOTIVATION
• Everyday objects have some underlying kinematic model
• Robots would benefit from the ability to learn the articulation in objects opportunistically from humans
• Function in unstructured environments with RGB-D data

CONTRIBUTIONS
• Reduced-drift feature tracking algorithm that allows robot to observe object manipulation in unstructured environments
• Density-based motion segmentation algorithm that efficiently and robustly clusters rigidly moving trajectories
• Capability to predict the motion of articulated objects in a future encounter

TRAINING (OFFLINE)

PREDICTION (ONLINE)

RESULTS
• 10 household objects / 43 sessions
• State-of-the-art performance in pose and model estimation accuracies
• Considerably improved success rates compared to Katz
Learning Articulated Motions From Visual Demonstration

Robotics, Vision and Sensor Networks Group, CSAIL, MIT

Matthew Walter
mwalter@csail.mit.edu

Seth Teller
teller@csail.mit.edu

MATeRial

Everyday objects have some underlying kinematic model
Robots would benefit from the ability to learn the articulation in objects opportunistically from humans
Function in unstructured environments with RGB-D data

CONTRIBUTIONS

Reduced-drift feature tracking algorithm that allows robot to observe object manipulation in unstructured environments
Density-based motion segmentation algorithm that efficiently and robustly clusters rigidly moving trajectories
Capability to predict the motion of articulated objects in a future encounter

System overview
Learning Articulated Motions From Visual Demonstration

Sudeep Pillai
spillai@csail.mit.edu
Robotics, Vision and Sensor Networks Group, CSAIL, MIT

Matthew Walter
mwalter@csail.mit.edu

Seth Teller
teller@csail.mit.edu

TRAINING (OFFLINE)
• User-provided demonstrations of manipulation tasks

PREDICTION (ONLINE)
• Detect and predict motions of articulated objects previously encountered
Learning Articulated Motions From Visual Demonstration

Matthew Walter
mwalter@csail.mit.edu
Robotics, Vision and Sensor Networks Group, CSAIL, MIT

Seth Teller
teller@csail.mit.edu

Sudeep Pillai
spillai@csail.mit.edu

Matthew Walter
mwalter@csail.mit.edu
Robotics, Vision and Sensor Networks Group, CSAIL, MIT

Seth Teller
teller@csail.mit.edu

TRAINING (OFFLINE)
• User-provided demonstrations of manipulation tasks

PREDICTION (ONLINE)
• Detect and predict motions of articulated objects previously encountered

System architecture of the training phase

System architecture of the prediction phase
**TRAINING (OFFLINE)**

- Learn the kinematic models of objects from user-provided demonstrations of manipulation tasks

**PREDICTION (ONLINE)**

- Detect and predict motions of articulated objects previously encountered
Learning Articulated Motions From Visual Demonstration

Sudeep Pillai
spillai@csail.mit.edu
Matthew Walter
mwalter@csail.mit.edu
Seth Teller
teller@csail.mit.edu
Robotics, Vision and Sensor Networks Group, CSAIL, MIT

TRAINING (OFFLINE)
- Learn the kinematic models of objects from user-provided demonstrations of manipulation tasks

TRAINING (OFFLINE)
- Detect and predict motions of articulated objects previously encountered

System architecture of the training phase

PREDICTION (ONLINE)
- Learn the kinematic models of objects from user-provided demonstrations of manipulation tasks

PREDICTION (ONLINE)
- Detect and predict motions of articulated objects previously encountered

System architecture of the prediction phase
FEATURE TRACKING

• Feature Initialization + Dense OF + Forward-Backward Consistency Check
• Trajectory Management: Feature Addition, Matching, Pruning and Refinement
• Key Advantages: Reduced drift compared to KLT, On-line descriptor learning

Features Tracking Schematic
MOTION SEGMENTATION

• **Key Idea:** For rigid-body motion the relative displacement/angle subtended between feature trajectories remain consistent over time
• **DBSCAN:** Density-Based Spatial Clustering of Applications with Noise
• Model relative motion as a Gaussian:

\[ L(i,j) = \frac{1}{T} \sum_{t \in [t_i,t_j]} \exp \left\{ -\gamma \left( d(x_i^t, x_j^t) - \mu_{d_{ij}} \right)^2 \right\} \]

bandwidth parameter that explains the variation in relative motions of the two trajectories

For \( \vec{x} \in \mathbb{R}^3 \): \( \mu_{d_{ij}} = \frac{1}{T} \sum_{t \in [t_i,t_j]} d(\vec{x}_i^t, \vec{x}_j^t) \)

For \( \vec{r} \in SO(3) \): \( \mu_{d_{ij}} = \frac{1}{T} \sum_{t \in [t_i,t_j]} d(\vec{r}_i^t, \vec{r}_j^t) \)

where \( t_i \) and \( t_j \) observed time instances of the feature trajectories \( i \) and \( j \)
POSE ESTIMATION

- Recovery of 6-DOF pose given clusters of rigidly moving feature trajectories
- Smoothed pose estimates via iSAM by imposing constraints using full information

ARTICULATION LEARNING (Sturm et al. JAIR '11)

Learn the underlying kinematic model of objects

\[
(M_{ij}, \hat{\theta}_{ij}) = \arg \max_{M_{ij}, \theta_{ij}} p(M_{ij}, \theta_{ij} | D_{n_{ij}})
\]

Model Fitting (MLESAC)

\[
\hat{\theta}_{ij} = \arg \max_{\theta_{ij}} p(\theta_{ij} | D_{n_{ij}}, M_{ij})
\]

Model Comparison

\[
\hat{M}_{ij} = \arg \max_{M_{ij}} \int p(M_{ij}, \theta_{ij} | D_{n_{ij}}) d\theta_{ij}
\]

Model Selection:

\[
\hat{M} = \arg \min_{\mathcal{M}} \text{BIC}(\mathcal{M})
\]

where

\[
\text{BIC}(\mathcal{M}) = -2 \log p(D_{n} | \mathcal{M}, \hat{\theta}) + p \log n
\]

Neg. Data Likelihood

Model Complexity

Penalty

Structure Selection:

\[
\hat{G} = \arg \max_{G} p(G | D_{z}) = \arg \max_{G} \sum_{(ij) \in E_{G}} \log p(M_{ij}, \theta_{ij} | D_{z})
\]

Or equivalently

Minimum spanning tree of the graph with edges

\[
\text{cost}_{ij} = - \log p(M_{ij}, \theta_{ij} | D_{z})
\]

\[G = (V_{G}, E_{G}), E_{G} \subset V_{G} \times V_{G}\]

\[
M_{ij}\quad \text{Kinematic Model between object part i and j}
\]

\[
\theta \quad \text{Model Parameter Vector}
\]

\[
D_{z} = (\Delta_{l_{ij}}, \ldots, \Delta_{l_{ij}}) \forall (ij) \in E_{G}
\]

Sequence of observed relative transformations between parts i and j

Accurate 6-DOF Pose Estimates
Learning Articulated Motions From Visual Demonstration

Sudeep Pillai
spillai@csail.mit.edu

Matthew Walter
mwalter@csail.mit.edu

Seth Teller
teller@csail.mit.edu

Robotics, Vision and Sensor Networks Group, CSAIL, MIT

Matthew Walter
mwalter@csail.mit.edu

Seth Teller
teller@csail.mit.edu

Robotics, Vision and Sensor Networks Group, CSAIL, MIT

MOTION ESTIMATION

- Extract most likely surface that moves consistently w.r.t the motion manifold
- Maximally Stable Extremal Regions using RGB-D data
Learning Articulated Motions From Visual Demonstration

Sudeep Pillai
spillai@csail.mit.edu
Robotics, Vision and Sensor Networks Group, CSAIL, MIT

Matthew Walter
mwalter@csail.mit.edu

Seth Teller
teller@csail.mit.edu

TRAINING (OFFLINE)
- User-provided demonstrations of manipulation tasks

TRAINING (OFFLINE)

PREDICTION (ONLINE)
- Detect and predict motions of articulated objects previously encountered

System architecture of the prediction phase

Input RGB-D Image → Trajectory Construction → Trajectory Clustering → Pose Estimation → Articulation Learning → Motion Estimation
Learning Articulated Motions From Visual Demonstration

- Future encounter: robot can localize object and predict its motion.
- Scalable bag-of-words driven DB for storing and querying.
- Predict motion given object pose estimation and kinematic model recovery.

Instance Recognition and Motion Prediction

MOTION PREDICTION

- On future encounter, robot can localize object and predict its motion.
- Scalable bag-of-words driven DB for storing and querying.
- Predict motion given object pose estimation and kinematic model recovery.
Learning Articulated Motions From Visual Demonstration

Sudeep Pillai
spillai@csail.mit.edu
Robotics, Vision and Sensor Networks Group, CSAIL, MIT

Matthew Walter
mwalter@csail.mit.edu
Robotics, Vision and Sensor Networks Group, CSAIL, MIT

Seth Teller
teller@csail.mit.edu
Robotics, Vision and Sensor Networks Group, CSAIL, MIT

RESULTS
• 10 household objects / 43 demonstration sessions
• State-of-the-art performance in pose and model estimation accuracies
• Considerably improved success rates compared to Katz

FUTURE WORK

POSE ESTIMATION ACCURACY

Translational Error (cm) Orientation Error (deg) Parameter Error (deg)
Katz: 3.7 10.1 5.0
Ours: 2.4 4.7 3.4

MODEL ESTIMATION ACCURACY

Translational Error (cm) Orientation Error (deg) Parameter Error (deg)
Katz: 2.0 5.8 3.4
Ours: 1.7 5.0 3.3

FAILURE RATE

Pose Estimation Model Estimation
Katz: 20/43 28/43
Ours: 6/43 13/43

Lower is better

User-Provided Visual Demonstration

Learning Articulated Motions From Visual Demonstration

Matthew Walter
mwalter@csail.mit.edu
Robotics, Vision and Sensor Networks Group, CSAIL, MIT

Seth Teller
teller@csail.mit.edu
Robotics, Vision and Sensor Networks Group, CSAIL, MIT

RESULTS
• 10 household objects / 43 demonstration sessions
• State-of-the-art performance in pose and model estimation accuracies
• Considerably improved success rates compared to Katz

FUTURE WORK

POSE ESTIMATION ACCURACY

Translational Error (cm) Orientation Error (deg) Parameter Error (deg)
Katz: 3.7 10.1 5.0
Ours: 2.4 4.7 3.4

MODEL ESTIMATION ACCURACY

Translational Error (cm) Orientation Error (deg) Parameter Error (deg)
Katz: 2.0 5.8 3.4
Ours: 1.7 5.0 3.3

FAILURE RATE

Pose Estimation Model Estimation
Katz: 20/43 28/43
Ours: 6/43 13/43

Lower is better
FUTURE WORK