LEARNING ARTICULATED MOTIONS FROM VISUAL DEMONSTRATION

Sudeep Pillai, Matthew Walter & Seth Teller
CSAIL, MIT

13 July 2014
Articulated objects found in a kitchen
How do we teach a robot about the **articulation** in these objects?
PRIOR WORK

• Structure-From-Motion based

• Probabilistic Kinematic Modeling

• Tracking and Modeling of Articulated Objects
KEY CONCEPT

LEARNING THE MODEL

User-Provided Demonstrations (RGB-D Video)

Learned Kinematic Model + Visual Appearance

- Unstructured environment
- RGB-D data as only input
- Unsupervised
- Zero prior object information
LEARNING THE MODEL

TRAJECTORY CONSTRUCTION

- Reduced drift
- Dense optical flow
- Forward-Backward check

Masked out (for ground truth only)
LEARNING THE MODEL

RIGID-BODY TRAJECTORY CLUSTERING

- Rigid-motion clustering
- Robust to noise
LEARNING THE MODEL

POSE ESTIMATION & OPTIMIZATION

- 6-DOF pose recovery
- Smoothed estimates
LEARNING THE MODEL

ARTICULATION ESTIMATION

- Probabilistic kinematic model and parameter estimation (Sturm et al. JAIR ’11)
PREDICTING OBJECT MOTION

INSTANCE RECOGNITION

Future Encounter

Reacquired object

Bag-of-Words driven instance recognition
PREDICTING OBJECT MOTION

MOTION PREDICTION

Future Encounter

Motion Prediction
QUALITATIVE RESULTS

User-Provided Visual Demonstration

RGB-D Video
QUALITATIVE RESULTS

User-Provided Visual Demonstration

Learned Motion Manifold

RGB-D Video

Parameters
QUALITATIVE RESULTS

User-Provided Visual Demonstration

Learned Motion Manifold

Future Encounter

RGB-D Video

Parameters

RGB-D Image
QUALITATIVE RESULTS

User-Provided Visual Demonstration

Learned Motion Manifold

Predicted Motion Manifold

Future Encounter

Laptop

RGB-D Video

Parameters

Prediction

RGB-D Image
COMPARISON TO STATE-OF-THE-ART

• Dataset with 10 different household objects
• 43 demonstration sessions (each 30s in duration)
• Fiducial markers for ground truth

### MODEL ESTIMATION ACCURACY

<table>
<thead>
<tr>
<th>Metric</th>
<th>Katz</th>
<th>Ours</th>
</tr>
</thead>
<tbody>
<tr>
<td>Translational Error (cm)</td>
<td>2.0</td>
<td>1.7</td>
</tr>
<tr>
<td>Orientation Error (deg)</td>
<td>5.8</td>
<td>5.0</td>
</tr>
<tr>
<td>Parameter Error (deg)</td>
<td>3.4</td>
<td>3.3</td>
</tr>
</tbody>
</table>

**Lower is better**

### POSE ESTIMATION ACCURACY

<table>
<thead>
<tr>
<th>Metric</th>
<th>Katz</th>
<th>Ours</th>
</tr>
</thead>
<tbody>
<tr>
<td>Translational Error (cm)</td>
<td>3.7</td>
<td>2.4</td>
</tr>
<tr>
<td>Orientation Error (deg)</td>
<td>10.1</td>
<td>4.7</td>
</tr>
</tbody>
</table>

**Lower is better**
COMPARISON TO STATE-OF-THE-ART

- Dataset with 10 different household objects
- 43 demonstration sessions (each 30s in duration)
- Fiducial markers for ground truth

Improved accuracy in pose and model parameter estimation

**POSE ESTIMATION ACCURACY**

<table>
<thead>
<tr>
<th>Metric</th>
<th>Katz</th>
<th>Ours</th>
</tr>
</thead>
<tbody>
<tr>
<td>Translational Error (cm)</td>
<td>3.7</td>
<td>2.4</td>
</tr>
<tr>
<td>Orientation Error (deg)</td>
<td>10.1</td>
<td>4.7</td>
</tr>
</tbody>
</table>

**MODEL ESTIMATION ACCURACY**

<table>
<thead>
<tr>
<th>Metric</th>
<th>Katz</th>
<th>Ours</th>
</tr>
</thead>
<tbody>
<tr>
<td>Translational Error (cm)</td>
<td>2.0</td>
<td>1.7</td>
</tr>
<tr>
<td>Orientation Error (deg)</td>
<td>5.8</td>
<td>5.0</td>
</tr>
<tr>
<td>Parameter Error (deg)</td>
<td>3.4</td>
<td>3.3</td>
</tr>
</tbody>
</table>

**FAILURE RATE**

<table>
<thead>
<tr>
<th>Metric</th>
<th>Katz</th>
<th>Ours</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pose Estimation</td>
<td>20/43</td>
<td>6/43</td>
</tr>
<tr>
<td>Model Estimation</td>
<td>28/43</td>
<td>13/43</td>
</tr>
</tbody>
</table>
CONCLUSION

• Develop a framework that enables robots to learn the kinematic models of articulated objects in unstructured environments, simply by observing their motion

• Introduce the capability to predict the expected motion of previously seen articulated objects

• Demonstrated improved repeatability and accuracy in pose and model parameter estimation over current state-of-the-art
REMEMBERING SETH