Learning Social Affordance for Human-Robot Interaction

Tianmin Shu¹  M. S. Ryoo²  Song-Chun Zhu¹
¹Center for Vision, Cognition, Learning, and Autonomy, UCLA  ²Indiana University, Bloomington

Introduction

Objective:
Learning explainable knowledge from the noisy observation of human interactions in RGB-D videos to enable human-robot interactions.

Key idea:
Beyond traditional object and scene affordances, we propose a weakly supervised learning of social affordances for HRI.

Contributions:
• First formulation and hierarchical representation of social affordance
• Weakly supervised learning from noisy skeleton input
• Efficient motion synthesis based on learned hierarchical affordances

Model

![Diagram of the model showing joints selection and grouping, interaction parsing, and motion synthesis examples.]

Motion Synthesis

Goal: Given the initial 10 frames (25 fps), synthesize the motion of an agent given the motion of the other agent and the interaction type.

Algorithm:
At time $t$,
1) Estimate the current sub-event by DP
2) Predict the ending time $t'$ and the corresponding joint positions
3) Obtain the joint positions at $t + 5$ through interpolation

UCLA Human-Human-Object Interaction Dataset
• Five types of interactions; on average, 23.6 instances per interaction performed by totally 8 actors. Each lasts 2-7 s presented at 10-15 fps.
• RGB-D videos, skeletons and annotations are available: http://www.stat.ucla.edu/~tianmin.shu/SocialAffordance

Examples of discovered latent sub-events and their sub-goals

Exp 1: Average joint distance in meters (compared with GT skeletons)

<table>
<thead>
<tr>
<th>Method</th>
<th>Shake Hands</th>
<th>Pull Up</th>
<th>High-Five</th>
<th>Throw and Catch</th>
<th>Pull Over</th>
<th>Hand Over a Cup</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>HMM</td>
<td>0.362</td>
<td>0.444</td>
<td>0.284</td>
<td>0.189</td>
<td>0.229</td>
<td>0.2816</td>
<td></td>
</tr>
<tr>
<td>V1</td>
<td>0.061</td>
<td>0.144</td>
<td>0.079</td>
<td>0.061</td>
<td>0.074</td>
<td>0.0689</td>
<td></td>
</tr>
<tr>
<td>V2</td>
<td>0.066</td>
<td>0.231</td>
<td>0.109</td>
<td>0.109</td>
<td>0.1132</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ours</td>
<td>0.054</td>
<td>0.109</td>
<td>0.658</td>
<td>0.076</td>
<td>0.068</td>
<td>0.0730</td>
<td></td>
</tr>
</tbody>
</table>

Exp 2: User study (14 subjects)
Q1: Successful? Q2: Natural? Q3: Human vs. robot?
From 1 (worst) to 5 (best)

<table>
<thead>
<tr>
<th>Source</th>
<th>Shake Hands</th>
<th>Pull Up</th>
<th>High-Five</th>
<th>Throw &amp; Catch</th>
<th>Pull Over</th>
<th>Hand Over a Cup</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>GT</td>
<td>4.40 ± 0.00</td>
<td>4.00 ± 0.30</td>
<td>4.41 ± 0.40</td>
<td>4.00 ± 0.30</td>
<td>4.30 ± 0.00</td>
<td>4.00 ± 0.30</td>
<td></td>
</tr>
<tr>
<td>Ours</td>
<td>4.23 ± 0.34</td>
<td>4.20 ± 0.40</td>
<td>4.00 ± 0.30</td>
<td>3.60 ± 0.30</td>
<td>4.00 ± 0.30</td>
<td>4.00 ± 0.30</td>
<td></td>
</tr>
</tbody>
</table>

Acknowledgment

This research has been sponsored by grants DARPA SIMPLEX project N66001-15-C-4035 and ONR MURI project N00014-16-1-J0077

Scan to visit our project website