Replicable Tests and Benchmarking for Robotic Assembly Operations

Karl Van Wyk

Intelligent Systems Division
National Institute of Standards and Technology
Goal

• Performance assessment and benchmarking
• Feedback for researchers
• Future technical specifications
Community Interest

- **Workshops**
  - IEEE ICRA 2015 Benchmarking in Manipulation Research: The YCB Object and Model Set
  - IEEE ICRA 2015 Robotic Hands, Grasping, and Manipulation
  - IEEE CASE 2016 Robotic Hand Technologies and Performance
  - IEEE IROS 2017 Development of Benchmarking Protocols for Robot Manipulation
  - IEEE ICRA 2017 Reproducible Research in Robotics: Current Status and Road Ahead

- **Competitions**
  - Amazon Picking Challenge
  - DARPA Robotics Challenge
  - IEEE IROS 2016 Grasping and Manipulation Competition
  - IEEE IROS 2017 Grasping and Manipulation Competition
  - World Robot Summit (WRS) 2018/2020 Industrial Robotics Competition
Assembly Task

• Assembly of belt-drive units
  • Day 3: 3 normal sets
  • Day 4: 2 normal set + 1 set incl. surprise parts

• Competition time:
  • Day 3: 45min. × 2 trials
  • Day 4: 60min. × 1 trial

• Surprise parts
  • CAD model is given 60 min. prior to the competition
  • Real parts are given 10 min. prior to the competition
Community Interest

• Technical Committees
  – IEEE RAS Technical Committee on Performance Evaluation and Benchmarking of Robotics and Autonomous Systems
  – IEEE RAS Technical Committee on Robotic Hand Grasping and Manipulation

• Existing Efforts
  – YCB object and model set
  – UC Berkeley Open Discussion
  – Advanced Robotics for Manufacturing (ARM) Institute

• Publications
  – IEEE RAM R-Article

• Testing Facilities
  – New England Robotics Validation and Experimentation (NERVE) Center
University of Massachusetts Lowell
New England Robotics Validation and Experimentation (NERVE) Center

- Manipulator testbeds for evaluating grasping, assembly, and human-robot collaboration performance with a suite of robotic arms, hands, and sensors of varying capabilities and characteristics
- Test methods and benchmarks from NIST, YCB Object and Model Set, new methods in development
- Developing metrics and evaluation methods for the Advanced Robotics for Manufacturing (ARM) Institute

Contact: Adam Norton, Assistant Director
adam_norton@uml.edu

Soft Robotics
Right Hand Robotics
Wonik Robotics
Rethink Robotics
Rethink Robotic Sawyer with Robotiq 2-finger manipulating NIST assembly task board
Universal Robots UR5 with Robotiq 3-finger manipulating NIST grasping test artifact

More to come!
Problem Scope

- Performance assessment and benchmarking
  - Arms
  - Grippers/hands
  - Sensors
  - Algorithms
  - Implementation/integration
  - Objects
  - Tasks/tests
  - Metrics
  - Performance comparisons
Hand/Gripper

- Enhancing end-effector adaptability
- Unify R&D performance measurement
- Unify technical specifications
- Rollout of enhanced behaviors
# Hand/Gripper Tests

<table>
<thead>
<tr>
<th>Test Method</th>
<th>Measurement Instrument</th>
</tr>
</thead>
<tbody>
<tr>
<td>Finger Strength</td>
<td></td>
</tr>
<tr>
<td>Touch Sensitivity</td>
<td></td>
</tr>
<tr>
<td>Finger Force Tracking</td>
<td></td>
</tr>
<tr>
<td>Force Calibration</td>
<td></td>
</tr>
<tr>
<td>Grasp Strength</td>
<td></td>
</tr>
<tr>
<td>Slip Resistance</td>
<td></td>
</tr>
<tr>
<td>Grasp Efficiency</td>
<td></td>
</tr>
<tr>
<td>Cycle Time</td>
<td></td>
</tr>
<tr>
<td>In-Hand Manipulation</td>
<td></td>
</tr>
<tr>
<td>Object Pose Estimation</td>
<td></td>
</tr>
</tbody>
</table>

- Touch Sensitivity
- Grasp Strength
- Grasp Efficiency
- In-Hand Manipulation
• NIST SP: Proposed Standard Terminology for Robotic Hands and Associated Performance Metrics
• NIST SP: Performance Metrics and Test Methods for Robotic Hands

Assembly Performance Tests

- Quantify performance of a robotic system completing a task
- Tests target assembly operations: pick-place, insertion, fastening, meshing, wire harnessing, pulley belt routing
- Whole system-system testing
- Component testing
Test Design

• Assembly Operations
• Design for Assembly (DFA)
  • Human performance factor analysis
  • Parameterizes objects
  • Handling times
  • Insertion times
• Guide design space
• Direct human comparison

Performance Metrics

- Modes
  - Disassembly
  - Assembly
- Primary metrics
  - Speed $\rightarrow$ completion time
  - Reliability $\rightarrow$ probability of success
- Granularity
  - Per-part/operation
  - Whole
Data Analytics

- Ordinal or Attribute Data
  - Primary measure – probability of success
  - Discontinuous distribution – Kolmogorov-Conover
- Continuous Data
  - Primary measure – time
  - Distribution – Kolmogorov-Smirnov
  - ANOVA – Levene, Brown-Forsythe
  - Means – Snedecor-Cochran
- Matlab, R
Peg-in-Hole Test

- Functional test method to measure the performance of robot systems at basic insertions
- Triangular design facilitates cyclical testing
- Peg-hole parameters, spacing based on human data
Systems

System 1

System 2

4x Real-time
Peg-in-Hole Testing

System 1

4x Real-time Gaussian Hole Error
\( \mu = 0 \text{ mm}, \sigma = 2 \text{ mm} \)

System 2 Spiral

4x Real-time Gaussian Hole Error
\( \mu = 0 \text{ mm}, \sigma = 2 \text{ mm} \)
## Comparative Results

Offsets $\sim N(0, \sigma^2)$, $\sigma^2 = 2 \text{ mm}$

<table>
<thead>
<tr>
<th>Robotic System</th>
<th>Correlation</th>
<th>KS</th>
<th>$\mu$ (s)</th>
<th>$\sigma^2$ (s$^2$)</th>
<th>PS (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>System 1</td>
<td>0.01</td>
<td></td>
<td>18.31</td>
<td>107.3</td>
<td>87.6</td>
</tr>
<tr>
<td>System 2 Spiral</td>
<td>0.07</td>
<td>*</td>
<td>37.13*</td>
<td>399.6*</td>
<td>95.2</td>
</tr>
<tr>
<td>System 2 Random</td>
<td>-0.01</td>
<td>*</td>
<td>15.62</td>
<td>417.72*</td>
<td>95.2</td>
</tr>
<tr>
<td>System 2 Quasi-Random</td>
<td>-0.11</td>
<td>*</td>
<td>8.2*</td>
<td>50.25*</td>
<td>95.2</td>
</tr>
</tbody>
</table>

*Indicates statistical significance in comparison with System 1 after 60 trials.

Assembly Task Boards

- Series of themed boards
- Each instance focuses on particular assembly facets
- Design with reference to DFA
- Low-cost, internationally replicable
- Realistic components
Test Method

- Setup per trial
  - Option 1: randomly place task board and predesigned kit
  - Option 2: randomly place task board and parts (box shake)
- Execution
  - Disassembly: move-grasp-disassemble-transport components to target kit/bin
  - Assembly: move-grasp-transport-assemble components to target task board
- Considerations
  - Task board, bin, kit not fixed
  - Several trials per desired confidence, resolution
  - No restriction on order of solution
Test Metrics

• Operation- and object-centric
  • Completion time and pass/fail
    • Move-grasp-transport sequence
    • Assembly sequence
    • Move-disassemble sequence
    • Transport-place sequence

• Mode- and board-centric
  • Completion time
  • Percent complete
Thoughts

• Overfitting
  • Part variations
  • Operation variations
  • Variable initial conditions
  • Variety of task boards
• Difficulty
  • IROS 2017 GMC
Questions/Feedback


- Datasets
- Gripper/hand test methods
- Assembly test methods
- Replication documentation
- IEEE RHGM TC meeting notes and presentations