Robot Manipulation in Human Environments

Aaron Edsinger

Thesis committee
Rod Brooks, Daniela Rus, Rod Grupen

MIT EECS thesis defense
12.22.2006
Why manipulation in human environments?

Robots that can work with people could…

Social Extend the time an elderly person can live at home

Economic Provide assistance to a worker on an assembly line

Laziness Help with household chores

Changing Demographics Worldwide shortage of qualified nurses…
Where we are today…

At, and beyond, human levels in specialized domains

Dynamic Pen Spinning Using a High-speed Multifingered Hand with High-speed Tactile Sensor (Ishihara, 2006)
Domo
What does Domo do?
Contributions

• Real robot
  • Force sensing and compliance
  • Intrinsic safety
  • Single, integrated behavior system

• Strategies for human environments
  • Let the body do the thinking
  • Cooperative manipulation
  • Task relevant features

• Accomplish useful, everyday tasks
  • Dynamics and variability
  • No object models

Collaborators
  Charles Kemp
  Jeff Weber
Roadmap

• Approach

• Robot design
• Visual system
• Control architecture

• Working with object tips
• Working with people
• Manual skills

• Helping with chores

• Conclusions and future work
Roadmap

• Approach

• Robot design
• Visual system
• Control architecture

• Working with object tips
• Working with people
• Manual skills

• Helping with chores

• Conclusions and future work
Controlled environments

Adapt the world to the robot

- World can be sensed and described with a 3D model
- Actions are mostly pre-planned
- Sensing within engineered constraints

Vision Based Behavior Verification System of Humanoid Robot for Daily Environment Tasks (Okada et al., 2006)

Human environments

Adapt the robot to the world

If it works here … will it work here as well?

Uncertainty

• Beyond control of the robot engineer
• Unknown dynamics and variation

Sensing

• Sensory noise and clutter
• Piecemeal view of the world
Human environments

Scope of work

- Human dynamics
- Lighting variability
- Cluttered background

What we do...
- Variety of everyday objects
- Variability in lighting
- Human dynamics
- Cluttered background
- No 3D object models
- No environment engineering

What we don’t do...
- Naïve collaborators
- Multiple environments
- Autonomous grasping
- Autonomous task-planning
- Differing classes of objects
Strategies for human environments

• Let the body do the thinking
  • robot as more than a passive observer
• Cooperative manipulation
  • accomplish tasks as a human-robot team
• Task Relevant Features
  • structure in everyday objects and environments
Let the body do the thinking

Compensatory actions under uncertainty

• Use arms as feelers in the dark
• Stiffen arm before inserting a key
• Rest a cup on a table before pouring
• Brace a hand on table while writing
Let the body do the thinking

Leverage compliance to reduce geometric uncertainty

(Stanford, 1959)

A body of human form can:

• work with human tools
• view countertops and shelves
• generate social cues
Cooperative manipulation

Manipulation as team effort

• Leverage human perception and planning
• Exploit natural social cues
  • Eye gaze
  • Reaching
  • Physical contact
• Smooth path to autonomy
Task relevant features

Carefully select the aspects of the world that are to be perceived and acted upon...

1. Commonalities in everyday objects
2. Use sparse perceptual features
3. Features describe task, not object
4. Generalize across objects
5. Motor equivalence

Learning to Grasp Novel Objects Using Vision (Saxena et al., 2006)

The Design of Everyday Things, (Norman, 1990)
Task relevant features

Carefully select the aspects of the world that are to be perceived and acted upon…
Pieces of the puzzle...

Exploiting environment structure
A Behavior Based Arm Controller
(Connell, 1989)

Using action to assist perception
Better Vision Through Manipulation
(Metta and Fitzpatrick, 2003)

Learning to grasp everyday objects
Learning and Generalizing Control-Based Grasping and Manipulation Skills
(Platt, 2006)

Sparse features for grasping
Learning to Grasp Novel Objects Using Vision
(Saxena et al., 2006)

Working with people and tools
Building an Autonomous Humanoid Tool User
(Bluethmann et al., 2004)
Roadmap

• Approach

• **Robot design**
  • Visual system
  • Control architecture

• Working with object tips
• Working with people
• Manual skills

• Helping with chores

• Conclusions and future work
Domo

- 29 Degrees-of-Freedom
  - 7 DOF head
  - 2 DOF SEA neck
  - 6 DOF SEA arms
  - 4 DOF SEA hands
- Passive Compliance
- Force sensing
- 12 Linux PCs

Domo: A Force Sensing Humanoid Robot for Manipulation Research
(Edsinger and Weber, 2004)
Series Elastic Actuators
(Pratt and Williamson, 1995)

Impedance above closed-loop bandwidth

\[
\frac{F_{load}(s)}{X_{load}(s)} = K_s \\
\frac{F_{load}(s)}{X_{load}(s)} = s^2 N_{motor}^2 I_{motor}
\]

Arms and neck

Hands

Impedance

\[
\frac{F_{load}(s)}{X_{load}(s)} = \frac{s^2 N_{motor}^2 I_{motor}}{K_s + 1 + N_{motor} D(s)}
\]
Head

- 9 DOF
- SEA Neck
- Synchronized Firewire Cameras
- Gyroscope
- Encoders and potentiometers
- 1KHz DSP control
- Visual smooth-pursuit controller

Copy of the Mertz head, Designed by Jeff Weber
Mertz: A Quest for a Robust and Scalable Active Vision
Humanoid Head Robot (Aryananda and Weber, 2004)
Hands

- 4 DOF
- Series Elastic Actuators
- Force and angle sensing
- 1KHz DSP control
- 0.4 kg
- Compliant skin
- Compliant fingertips

Design of a compliant and Force Sensing Hand for a Humanoid Robot (Edsinger, 2004)
Lightweight compliant manipulators

- 6 DOF
- Cable drive
- Series Elastic Actuators
- Force and angle sensing
- 1KHz DSP control
- 2.2 kg
- Intrinsic safety
- 2 kg payload

Controllers for:
- Joint Force
- Gravity compensation
- Joint Angle
- Virtual spring
- Inverse kinematics

Design in collaboration with Jeff Weber
Roadmap

• Approach

• Robot design
  • Visual system
  • Control architecture

• Working with object tips
• Working with people
• Manual skills

• Helping with chores

• Conclusions and future work
Using motion cues...

Regions of interest

Generating motion to assist perception

Robust to lighting and clutter

Predictions to increase salience
Visual attention system

Consolidate perceptual streams into a single spotlight of attention

1. **Image**
2. **θ, ̇θ**
3. **Kinematic Model**
4. **Task Relevant Feature**
5. **Tracker**
6. **PersonDetect**
7. **Camera Model**
8. **Canny Edges**
9. **Block Matching**
10. **Affine Motion Model**
11. **InterestRegions**
12. **Sensory Ego-Sphere**
13. **Directs eye gaze**
14. **Salient features**

**Motion model**
**Motion prediction**
**Motion features**
**Short-term “memory”**
Affine motion model

1. Canny edge detector on consecutive images
2. Block matching to estimate translation of each edge
3. Compute covariance matrix $C$ of block matching error for each edge
4. Fit model $A$ to translations when weighted by $C$ (weighted linear-least-squares)
5. Iterate the fitting process
6. Weight each edge by how well translation fits $A$ (Mahalanobis distance)

- Estimate global motion of edges ($A$)
- Select for edges that violate model

Generates weighted edge map of foreground motion
Interest regions

- Weighted edge map from motion model
- Select strongest responding regions at multiple-scales

Convex edges define circular region in given scale-space

Work by Charles Kemp

Selects regions important for interaction and manipulation
Sensory Ego-Sphere

The Sensory Ego-Sphere as a Short-Term Memory for Humanoids
(R. A. Peters et al., 2001)

- Short term memory
- Improves stability
- Multi-modal registration
- Learn spatial distributions

\[ p(x^S | face) \sim \frac{\pi}{2}, \frac{\pi}{2} \]

\[ x^S = [\theta, \phi, r] \]
Roadmap

• Approach

• Robot design
  • Visual system
  • Control architecture

• Working with object tips
• Working with people
• Manual skills

• Helping with chores

• Conclusions and future work
Slate

A tool for specifying time contingent behaviors

Components

- Scheduler
- Modules ~ behaviors
- Threads
- Wires and Arbitrators
- Monostables
- FSAs with timing

Tools for rapid prototyping
Libraries for vision, learning
Designing manipulation tasks

Compose tasks through coordination of perception and control modules

Module
- Local task knowledge
- Estimate readiness

Module interaction
- Adjust activation priorities
- Communicate through the world

Models and learning
- Simple prior models
- Control feedback to correct for model errors
- Offline learning for detectors
Task decomposition within Slate

- Single, integrated system
- Integrated tasks
- Manual skills
- Task relevant features
- Compensatory actions
- Detectors and motor primitives

Main Process
125 Threads, 40 Wires, 35 FSAs, 10 Arbitrators
Manual skill algorithm

Integration
- Let the body do the thinking
- Cooperative manipulation
- Task relevant features
Manual skill algorithm

Integration
• Let the body do the thinking
• Cooperative manipulation
• Task relevant features

Basic control loop
Manual skill algorithm

Integration
• Let the body do the thinking
• Cooperative manipulation
• Task relevant features
Example: Bimanual insertion task

- Activation of modules that create readiness
- Monitor readiness
Example: Bimanual insertion task

- Prepare body to assist perception
- Take action to reduce uncertainty
Example: Bimanual insertion task

- Detect task relevant features
- Use learned models in detectors
Example: Bimanual insertion task

- Control feature
- Include perceptual feedback
Example: Bimanual insertion task

Detect controller success
Roadmap

- Approach
- Robot design
- Visual system
- Control architecture

- Working with object tips
- Working with people
- Manual skills

- Helping with chores

- Conclusions and future work
The task relevant tip

Detect and control the distal tip of a wide range of everyday objects

Why the prevalence of the tip?

• Single point of attention
• Improves visual observation by reducing occlusions
• Localize interaction forces
• Natural extension of controlling hand and finger
The task relevant tip

Goal: quickly localize and control the tip of an unknown object

Tip detected as the fastest moving convex edge in image

Work with Charles Kemp
Estimating the tip location

\[ p(x_t | d_1 \ldots d_n, c_1 \ldots c_n) = \frac{p(d_1 \ldots d_n | x_t, c_1 \ldots c_n) p(x_t, c_1 \ldots c_n)}{p(d_1 \ldots d_n, c_1 \ldots c_n)} \]

3D tip location  Detections  Kinematic configurations

Bayes rule
Estimating the tip location

\[ p(x_t | d_1 \ldots d_n, c_1 \ldots c_n) = \frac{p(d_1 \ldots d_n | x_t, c_1 \ldots c_n)p(x_t, c_1 \ldots c_n)}{p(d_1 \ldots d_n, c_1 \ldots c_n)} \]

**Maximum likelihood**

\[ \hat{x}_t = \text{Argmax}_{x_t} \left( p(d_1 \ldots d_n | x_t, c_1 \ldots c_n)p(x_t) \right) \]

\[ = \text{Argmax}_{x_t} \left( \log(p(x_t)) + \sum_i \log(p(d_i | x_t, c_i)) \right) \]

Uniform near hand

Solve in batch mode (N=50) using Nelder-Mead Simplex optimization

Work with Charles Kemp
Estimating the tip location

2D Gaussian error model in image

\[ p(d_i | x_t, c_i) = (1 - m)N_t(T_{c_i}(x_t), \sigma_t^2 I)(d_i) + mN_f(0, \sigma_f^2 I)(d_i) \]

Mixing parameter
Gaussian at tip
Detection noise

Work with Charles Kemp
Detecting tips of everyday tools

Robot Manipulation of Human Tools: Autonomous Detection and Control of Task Relevant Features (Kemp and Edsinger, 2005)

Pixel error of tip prediction in image

Estimated

Hand labeled

Work with Charles Kemp
Tip open-loop control

Virtual forces on object
\[ f_t^H \propto (x_d - \hat{x}_t) \]
\[ f_p^H \propto (p_d - p_t) \]

Virtual force in hand frame
\[ f_t^H = K_t M(I, \hat{x}_t) \left[ \begin{array}{c} (x_d - \hat{x}_t) \\ 0 \\ 0 \\ 0 \end{array} \right]^T \]
\[ f_p^H = K_p M(I, p_t) \left[ \begin{array}{c} (p_d - p_t) \\ 0 \\ 0 \\ 0 \end{array} \right]^T \]

Virtual force in world frame
\[ f_W = M(W R^H, 0) (f_t^H + f_p^H) \]

Controller update rule
\[ \Delta \theta = \sigma J^T f_W \]
Tip open-loop control

Virtual forces on object

\[
\begin{align*}
\mathbf{f}_t^H & \propto (\mathbf{x}_d - \mathbf{x}_t) \\
\mathbf{f}_p^H & \propto (\mathbf{p}_d - \mathbf{p}_t)
\end{align*}
\]
Tip servo control

Leverage strong prior on tip location

whenever the wrist rotates:

detect tip in image as $d_i$
estimate probability is tip detection as $p(d_i | \hat{x}_t, c_i)$

If $p(d_i | \hat{x}_t, c_i) > \epsilon$ :
initialize block-matching visual tracker at $d_i$

estimate tip location $y_t$ using closest point of $\hat{x}_t$ to tracker ray $r_i$

substitute $y_t$ for $\hat{x}_t$ in the open-loop controller
Tip servo control

Open-loop

Visual feedback
Roadmap

• Approach

• Robot design
  • Visual system
  • Control architecture

• Working with object tips
• Working with people
• Manual skills

• Helping with chores

• Conclusions and future work
Intrinsic safety

The Head Injury Criterion

\[ HIC = 2 \left( \frac{2}{\pi} \right)^{\frac{3}{2}} \left( \frac{K_{cov}}{M_{oper}} \right)^{\frac{3}{4}} \left( \frac{M_{rob}}{M_{rob} + M_{oper}} \right)^{\frac{3}{2}} v_{max}^{\frac{7}{4}} \]

\[ M_{rob} = M_{link} + \frac{K_s}{K_s + \gamma} M_{rot} \]

\[ \gamma \quad \text{Rigid link stiffness} \]

\[ K_s \quad \text{Spring stiffness} \]
Detecting contact motion

Arm stiffness:

\[ \tau_{desired} = K_{ps} \left( PID(\theta_{desired}, \theta) \right) - G(\Theta) \]

Learn velocity prediction:
(Support Vector Regression)

\[ g_v(K_{ps}) \]

Signal contact when prediction violation:

\( \left( \| J\dot{\Theta} \| - g_v(K_{ps}) \right) > \epsilon \) and \( K_{ps} < 0.5 \)
Detecting contact force

Joint space dynamic model

\[ \tau_{\text{dyn}} = M(\Theta)\ddot{\Theta} + V(\Theta, \dot{\Theta}) + G(\Theta) \]
\[ \approx M(\Theta)\ddot{\Theta} + G(\Theta) \]

Prediction error

\[ e = \tau_{\text{dyn}} - \tau_{\text{sense}} \]

Prediction error histograms

Signal contact when prediction violation:

\[ p(e|\text{contact}) > p(e|\text{no\_contact}) \]
Assisted grasping

Social cues to request assistance
- Reach direction
- Eye contact
- Grasp preshape
Can we design naïve collaborators into a task?

- 10 subjects
- 6 trials each
- Naïve to task
- Varied experience with robots
- Incentive during release

Give and take experiment

Will people give an object aligned with the robot’s grasp?
Will people take an object when offered by the robot?

Cues:
- Voice “ok”
- Eye-gaze
- Reach direction
- Reach length
- Wrist orientation

grasp box  “inspection”  offer box  release box
Can we design naive collaborators into a task?

- Very good at matching grasp
- Incentive increases likelihood of taking back
- Unrelated to expertise

- Mean grasp orientation errors (10 subjects)
- Grasp orientation errors (6 trials, one subject)
- Take-back rate (10 subjects)
- Incentive
Roadmap

- Approach
- Robot design
- Visual system
- Control architecture
- Working with object tips
- Working with people
- Manual skills
- Helping with chores
- Conclusions and future work
Reaching in the dark...

- Uncertain detection of surface
- Use body to verify location
- Represent location in joint space

Reach trajectories

Vary shelf pose and height
Switching hands

Expand the manipulation workspace

- Estimation of object tip
- Object realignment
- Visual servo of the palm
- Exploit compliance during contact
Switching hands

Expand the manipulation workspace
Placing objects on flat surfaces

- Estimate placement stability
- Place upright or lying down
- Leverage compliance to self-align to surface
Placing objects on flat surfaces

- Estimate placement stability
- Place upright or lying down
- Leverage compliance to self-align to surface
Placing objects on flat surfaces

Stability estimate

<table>
<thead>
<tr>
<th>Object</th>
<th>( \eta ) measured</th>
<th>( \eta ) sensed</th>
</tr>
</thead>
<tbody>
<tr>
<td>animal</td>
<td>0.48</td>
<td>0.50</td>
</tr>
<tr>
<td>box</td>
<td>1.15</td>
<td>0.82</td>
</tr>
<tr>
<td>large bottle</td>
<td>0.56</td>
<td>0.54</td>
</tr>
<tr>
<td>caddy</td>
<td>0.61</td>
<td>0.49</td>
</tr>
<tr>
<td>spoon</td>
<td>0.21</td>
<td>0.24</td>
</tr>
<tr>
<td>roller</td>
<td>0.29</td>
<td>0.26</td>
</tr>
<tr>
<td>broom</td>
<td>0.24</td>
<td>0.23</td>
</tr>
<tr>
<td>food tin</td>
<td>1.20</td>
<td>0.98</td>
</tr>
<tr>
<td>small bottle</td>
<td>0.57</td>
<td>0.49</td>
</tr>
<tr>
<td>duster</td>
<td>0.13</td>
<td>0.24</td>
</tr>
<tr>
<td>spray bottle</td>
<td>0.36</td>
<td>0.42</td>
</tr>
</tbody>
</table>

Misalignment tolerance

\[ \theta_s < \frac{\pi}{2} - \tan^{-1} \frac{h}{r} \]

Wrist stiffness v.s. misalignment

3 trials ea.

\[-40^\circ \leq \theta_s \leq 40^\circ\]

\( K_{ps} \in [0.0, 0.5, 1.0] \)

Predicted:

\[ \theta_s = 21.2^\circ \]
Finding an opening

Compensatory action
Use the table for:
• Stability
• Alignment
• Localize near hand
• Detect convex edge
Bimanual insertion

<table>
<thead>
<tr>
<th></th>
<th>Paper cup</th>
<th>Bowl</th>
<th>Box</th>
<th>Coffee mug</th>
<th>Jar</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mixing spoon</td>
<td>7/7</td>
<td>7/7</td>
<td>7/7</td>
<td>6/7</td>
<td>7/7</td>
</tr>
<tr>
<td>Water bottle</td>
<td>6/7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Paint brush</td>
<td>6/7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Paint roller</td>
<td>5/7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spoon (prior only)</td>
<td>1/7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Bimanual fixture

Right arm controller

\[ \hat{x}_r = x_b + \left[ \frac{d}{2} \cos(\theta_p), \frac{d}{2} \sin(\theta_p), 0 \right]^T \]

\[ f_r = -k_f (x_r - x_b) \]

\[ \Delta \theta_r = \sigma \mathbf{J}^T f_r \]

\[ \theta_{desired-r} = IK(\hat{x}_r) + \Delta \theta_r \]
Roadmap

• Approach

• Robot design
• Visual system
• Control architecture

• Working with object tips
• Working with people
• Manual skills

• Helping with chores

• Conclusions and future work
Helping with chores
Helping with chores

Collaborator composes task plan

Putting away groceries

Assisted grasping

Making a drink

Cleaning up

Collaborator interaction
Putting away groceries

"Domo, shelf"

1. Find a flat surface
2. Verify it is reachable
3. Take item from person
4. Switch item to closest hand
5. Place item on shelf

"Domo, take"

Collaborator interaction

Putting away

Assisted grasping
Making a drink

“Domo, insert”
1. Take bottle from person
2. Take cup from person
3. Insert bottle tip into cup
4. Give bottle to person
5. Take spoon from person
6. Insert spoon tip into cup
7. Put cup on shelf

“Domo, give it”
“Domo, insert”
“Domo, shelf”

Bimanual Insertion

Collaborator interaction

Assisted grasping

Putting away
Cleaning up

“Domo, box”

1. Cue person to hand the box
2. Detect contact cue during hand-off
3. Form bimanual grasp
4. Track person as they move
5. Position box near the person as items are placed in it
6. Lower the box onto the table

“Domo, done”

“Domo, box”

“Domo, done”
Roadmap

- Approach
- Robot design
- Visual system
- Control architecture
- Working with object tips
- Working with people
- Manual skills
- Helping with chores
- Conclusions and future work
Further details…


Caveats and Extensions

- Variety of everyday objects
- Variability in lighting
- Human dynamics
- Cluttered background
- No 3D object models
- No environment engineering

What we do...
- Rely on benign collaborator
- Objects within a class
- Adaptation to failure

What we don’t do...
- Naïve collaborators
- Multiple environments
- Autonomous grasping
- Autonomous task-planning
- Differing classes of objects
Caveats and Extensions

• Task planning without models
• Dealing with arbitrary objects

What we do...

• Variety of everyday objects
• Variability in lighting
• Human dynamics
• Cluttered background
• No 3D object models
• No environment engineering

What we don’t do...

• Naïve collaborators
• Multiple environments
• Autonomous grasping
• Autonomous task-planning
• Differing classes of objects
Future Work
More dynamics, more adaptation

• A suite of task relevant features
  • Tactile shape features
  • Local 3D features
  • Flat surfaces, handles
• A suite of social cues
  • Pointing
  • Force guidance
  • Gestures
• A suite of manual skills
  • Opening door
  • Stacking

Learning by demonstration via task relevant features
Conclusions

Human environments pose unique challenges

Rather than solve impressive, dexterous tasks...
...focus on basic manual skills that account for these challenges

Lesson: People leverage structure in human environments to reduce cognitive load
...robots can do the same
Contributions

• Real robot
  • Force sensing and compliance
  • Intrinsic safety
  • Single, integrated behavior system

• Strategies for human environments
  • Let the body do the thinking
  • Cooperative manipulation
  • Task relevant features

• Accomplish useful, everyday tasks
  • Dynamics and variability
  • No object models
Thanks to...

Collaborators:
  Charles C. Kemp
  Jeff Weber

Advisor:
  Rodney A. Brooks

Committee:
  Daniela Rus
  Roderic Grupen

Support:
  Toyota Motor Corporation
  NASA JSC

Family&Friends!!!