Vision and Touch: Multimodal or Cross-modal?

Shan Luo
Department of Computer Science
University of Liverpool
shan.luo@liverpool.ac.uk

Centre for Autonomous Systems Technology
An overview

1. Synesthesia: Touching to See and Seeing to Feel
2. Senses of Robots: Sensors
3. Object representations using tactile images
4. Integrated visual-tactile perception
5. Interactive Perception
Biological synesthesia
Human vs Robot perception

Vision
Tactile sensing
Proprioception
Auditory sensing
Representation
Action
Interaction
Communication
...

PR2: www.willowgarage.com/pages/pr2
Human vision
Human touch

Hand area of somatosensory area of cortex

Nerve fibers responding to
Pain
Warm
Touch
Vibration
Human multimodal sensation
Human multimodal sensation

Senses of Robot: Sensors
Eyes – Cameras

Webcams

Depth sensor

Stereo camera

RGB-D Object Dataset
## Touch – Tactile sensors

<table>
<thead>
<tr>
<th>Type</th>
<th>Advantages</th>
<th>Disadvantages</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Capacitive</td>
<td>Taxels can be very small; High spatial resolution; Sensitive to small forces; High frequency response.</td>
<td>Hysteresis; Sensitive to electromagnetic noise and temperature change; Non-linear response; Cross-talk; Complex electronics.</td>
<td>PPS DigiTacts</td>
</tr>
<tr>
<td>Piezoelectric</td>
<td>Good workability; Flexibility; Chemical stability; Fast dynamic response.</td>
<td>Sensitive to temperature; Fragility of electrical junctions; Dynamic measurements only.</td>
<td>PVDF</td>
</tr>
<tr>
<td>Optical</td>
<td>Magnetic resonance compatible; Flexible and fast; No interconnections</td>
<td>Bulky sizes; High power consuming.</td>
<td>Optical fibres</td>
</tr>
<tr>
<td>Piezo-resistive</td>
<td>Economic; Sensitive and robust; Easy to manufacture.</td>
<td>High power consuming; Low repeatability; Fragile to shear forces.</td>
<td>Weiss tactile sensors</td>
</tr>
<tr>
<td>Tunnel effect</td>
<td>High spatial resolution; High dynamic range.</td>
<td>Bulky sizes; Non-linear response.</td>
<td>QTC touch sensors</td>
</tr>
<tr>
<td>Ultrasonic-based</td>
<td>Fast dynamic response; Good force resolution.</td>
<td>Bulky sizes; limited utility at low frequency; sensitive to temperature.</td>
<td></td>
</tr>
<tr>
<td>Magnetism-based</td>
<td>Sensitive and robust; No measurement hysteresis.</td>
<td>Limited to nonmagnetic mediums.</td>
<td>MagOne</td>
</tr>
<tr>
<td>Barometer-based</td>
<td>Good elasticity; Low cost.</td>
<td>Low spatial resolution.</td>
<td>TakkTile; BioTac</td>
</tr>
<tr>
<td>Camera-based</td>
<td>Extra high spatial resolution.</td>
<td>Bulky sizes.</td>
<td>GelSight</td>
</tr>
<tr>
<td>Graphene</td>
<td>Flexible, extra thin</td>
<td>Still too high price</td>
<td>Dahiya ‘17</td>
</tr>
</tbody>
</table>
Tactile sensors - Single-contact (haptic) sensors

Tactile sensors - High spatial resolution sensors

Weiss Robotics

Pressure Profile Systems (PPS)

GelSight

Multimodal BioTac

Tekscan

Density of Merkel receptors in the fingertip: around 14×14

Tactile sensors - High spatial resolution sensors

Tactile sensors - Large-area tactile sensors

# Touch – Tactile sensors

<table>
<thead>
<tr>
<th>Modality</th>
<th>FoV</th>
<th>Info.</th>
<th>Compl.</th>
<th>Compu.</th>
<th>Invariance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Visual</td>
<td>Global</td>
<td>Rich</td>
<td>Low</td>
<td>High</td>
<td>Scaling, rotation, translation, illumination</td>
</tr>
<tr>
<td>Tactile</td>
<td>Local</td>
<td>Sparse</td>
<td>High</td>
<td>Low</td>
<td>Rotation, translation, “illumination”</td>
</tr>
</tbody>
</table>

Object representations using tactile images
Object representations in the robot world

Local shape
Global shape
Locations
Texture
Pose
...

PR2: www.willowgarage.com/pages/pr2
Global shape recognition

Contact points based recognition

Tactile point clouds

Tactile sensor readings

Occupancy grid mapping

Pro: Arbitrary contact shapes can be retrieved.

Con: Time consuming when investigating a large object surface.


Global shape recognition

Tactile patterns based recognition

- Feature Extraction
- Image moments
- Tactile-SIFT

Bag-of-Words Framework

Pro: Local features can be revealed.
Con: Three-dimensional distribution information is not incorporated.

Global shape recognition

Shape representation in 4D space

Shan Luo, et al., “iCLAP: Shape Recognition by Combining Proprioception and Touch Sensing.”, *Autonomous Robots*, 2018
Integrated visual-tactile perception
Feature sharing between vision and tactile sensing

Feature sharing between vision and tactile sensing

“Touch to See” and “Seeing to Feel”

“Touch to See” and “Seeing to Feel”

“Touch to See” and “Seeing to Feel”

(a) Ideal training images generating a consistent pattern
(b) Cloth set containing a dyed design showing difficulty in replicating a tactile-visual image

Interactive tactile perception
Interactive perception by interacting with objects

Press  Slip  Twist  Explore  Push  Grasping  Manipulation  …

PR2: www.willowgarage.com/pages/pr2
Interactive perception by interacting with objects

Classic SLAM: vision as input
Interactive perception by interacting with objects

Haptic SLAM

Interactive perception by interacting with objects
Interactive perception by interacting with objects

Experimental setup

Tactile Sensor

Tactile Patch

Localization

Visual Image
Interactive perception by interacting with objects

Bayesian filtering framework

Interactive perception by interacting with objects

Bayesian filtering framework

Algorithm Bayes_filter\(\text{bel}(x_{t-1}), u_t, z_t)\):

\[
\text{for all } x_t, \text{ do}
\begin{align*}
\text{bel}(x_t) &= \int p(x_t | u_t, x_{t-1}) \text{bel}(x_{t-1}) dx_{t-1} & \text{ Motion update} \\
\text{bel}(x_t) &= \eta p(z_t | x_t) \text{bel}(x_t) & \text{ Overall update}
\end{align*}
\text{endfor}
\text{return } \text{bel}(x_t)
\]

\(\text{bel}(x_{t-1})\): belief at time t-1
\(u_t\): sensor movement at t
\(z_t\): sensor measurement at t
\(\text{bel}(x_t)\): belief after control update
\(\text{bel}(x_t)\): belief after measurement update

Interactive perception by interacting with objects

Motion model/Control update

Interactive perception by interacting with objects

Feature-based measurement models/Measurement update

Interactive perception by interacting with objects

Leveraging Action in Perception and Perception in Action

Picture credit:
Benjamin Schneiders, smARTLab
Leveraging Action in Perception and Perception in Action

https://www.roboticgizmos.com
https://www.digitalspy.com
Summary

- Goal: Robots perceive the physical world

- Make use of high-resolution camera-based tactile sensors for perception tasks
- Tactile object representations of object local and global shapes, textures and poses
- Multimodal and Cross-modal visual-tactile perception

- Future: Intelligent perception through interaction
Acknowledgement

Prof Ted Adelson
MIT CSAIL

Dr Hongbin Liu
King’s College London

Prof Raul Fuentes
University of Leeds

Prof Karl Tuyls
Google DeepMind, Paris

Prof Kaspar Althoefer
Queen Mary University of London

Prof Michael Fisher
University of Liverpool

Prof Andy Cooper
University of Liverpool

Prof Tony Cohn
University of Leeds
Thank you