Study/ Review of Speechless Interaction Techniques in Social Robotics

Dr. Manoj Ramanathan
Institute for Media and Innovation
Nanyang Technological University, Singapore
26th March 2019

IMI Research Seminar
Outline

• Introduction & Motivation
• Speechless Techniques
  • Gazing
  • Affective Engine – Emotions
  • Social Media
  • Action/Gestures
  • Reading
• Conclusion
A social robot is an autonomous robot that interacts and communicates with humans or other autonomous physical agents by following social behaviors and rules attached to its role. Like other robots, a social robot is physically embodied (avatars or on-screen synthetic social characters are not embodied and thus distinct).

Introduction & Motivation

- Human interaction
- Environmental awareness
- Applications
Introduction & Motivation

Verbal – Based on words

Non-Verbal

- Speechless
- How to use these methods for social robots?
Gazing

• Eye contact plays an important role
• Based on user position, robot’s head and eyes have to be changed
• Where and When to see?
Gazing

- One-to-one
- Do we maintain eye contact throughout [1]

- Presentation or any other application context
- Interesting objects or events

Affective Engine - Emotions

- Understanding and processing human emotions
- Showing all possible emotions in behavior

Social Media

- Simple ones like email
- More difficult ones like skype, facebook, twitter etc
- #AskSophia in Twitter
- Not only words, images, videos etc can be shared
Actions/Gestures

- Simple gestures
- Subtle Actions/Gestures
- Which actions to react to?
- When and how to react?
Actions/Gestures

• [1] focused on action recognition in social robots
• Based on direction vectors of each skeleton joint’s movement or pose trajectories
• Histogram of direction vectors was used for representation
• Total 18 activities were included
• Tested on P3Dx mobile robot platform

• [2] proposed how robots could learn actions/ tasks from users
• Two tasks were considered
  • Salt-Shaker
  • Cup-stacking
• iCub Robot was used
• The user teaches tasks to the robot and it differentiates between the 2 actions.

Actions/Gestures

• [1] proposed Space-Time Occupancy Patterns (STOP)
• Uses depth maps
• Each depth sequence, divided along spatial and time axes to create 4D grids
• Preserves spatial and temporal contextual information between space-time cells
• Online action recognition is possible
• Used with a Robot-
  https://www.youtube.com/watch?v=oPvO-7kDM0c

Skeleton Representation with CNN [1] –

- Transform 3D skeleton sequences into three clips
- Each clip is generated from one channel of the cylindrical coordinates of the skeleton sequence.
- Each frame of the clip represents temporal information of entire skeleton sequence and incorporates one particular spatial relationship between joints.
- Multiple frames with different spatial relationships are included.

Actions/Gestures

Skeleton Representation with CNN [1] –

- Deep CNN is used to learn long-term temporal information of the generated clips and then use Multi-Task Learning Network to jointly process all frames of generated clips to incorporate spatial structural information.

Actions/Gestures

Action recognition from Nadine. Current implementation recognizes 21 actions. (Shaking head, nodding, drinking water, clapping, waving, taking selfie, reading, writing, etc.)

Action: Make phone call/ Answer phone

Action: Check Time

Amir Shahroudy, Jun Liu, Tian-Tsong Ng, and Gang Wang, "NTU RGB+D: A Large Scale Dataset for 3D Human Activity Analysis", in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016
Discussion

• Actions to be recognized by Robot/VH
  – Which actions need to be recognized
  – Depending on application classroom, office, shops etc
• Action as a social cue
  – Reactions
    • How to react after recognition?
    • Several factors user emotion, robot emotion, context etc
    • Observe subtle, non-verbal actions
Discussion

- **Reactions**
  - Movie scripts can be used to develop [1]
  - NLTK allows to identify verbs/actions in scripts
  - Using the identified verbs, we can extract context and possible reactions – Mainly for verbal responses.
  - Train model to generate verbal responses based on action, context, objects etc

- **Limitations**
  - Word meaning ambiguities
  - Action meaning based on context
  - Extract clean and required data

Reading

• OCR is widely available
• Language?
• Lots of possible applications
• Make robot understand what it has read and react accordingly.
Reading

• Pioneer 2 robot [1] – simple textual message read from images
• Focus was only if OCR performance is good but not on what to do after reading.


Reading

• Google Cloud Vision API [1]
• Pytesseract-OCR [2]
• Applications
  – IC Registration
  – Parcel Delivery

[1] https://console.cloud.google.com/
- Layout Organization[1]

- Application based
- Showing emotions while reading stories

- Can Robot Understand?
- Hand-written characters
- Lighting
- Language

Conclusion

• Understanding user intentions, behavior
  – Very difficult to quantify, perceive
  – Too many possible variations
  – Actions, Emotions are only some visible cues but are not definitive
  – Developing a reaction model is not easy

• Reading opens another avenue of communication
  – Few limitations still exist in OCR that prevents complete implementation for robots
Thank you!!

Q & A ??
### Conclusion

<table>
<thead>
<tr>
<th>Robot</th>
<th>Gazing</th>
<th>Affective System</th>
<th>Gesture/Actions</th>
<th>Human –like Appearance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nadine</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Sophia</td>
<td>✔</td>
<td>NA</td>
<td>✗</td>
<td>✔</td>
</tr>
<tr>
<td>Receptionist Robot</td>
<td>✔</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
</tr>
<tr>
<td><a href="https://tinyurl.com/yxmdgw8m">https://tinyurl.com/yxmdgw8m</a></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Palrobotics</td>
<td>✔</td>
<td>NA</td>
<td>NA</td>
<td>✗</td>
</tr>
<tr>
<td>Nancy</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✗</td>
</tr>
<tr>
<td>Erica</td>
<td>✔</td>
<td>NA</td>
<td>✗</td>
<td>✔</td>
</tr>
<tr>
<td><a href="https://robots.ieee.org/robots/erica/">https://robots.ieee.org/robots/erica/</a></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>