Interaction with Virtual Human: Analysis and Prediction of Body Gestures using A Depth Sensor

Flavien Picon, Research Fellow
Supervisor(s): Pr. Nadia Magnenat-Thalmann
PhD Seminar
06 November 2012
Interaction with Virtual Human and Social Robot

- Human presence
- Speech
- Gesture
- Lighting + time of day
- Noise
- Furniture

Introduction
Motivation
Analysis of characteristics
On-line prediction

PhD Seminar – November 2012
Human actions, source of meaning

• Different kind of gestures:
  – Commands:
    • to control a system
    • restricted vocabulary, learning
  – Actions:
    • Walking, drinking, reading, using phone
  – Communication:
    • Sign language
    • gesticulation: body motion that support the speech
    • Emblems
    • (Facial expression)

Social context and autonomy => Focus on actions and communication
Gesture recognition

- Large variety of solution:
  - Hand/body tracking
  - Segmentation of human (data glove, electromagnetic tracking)
  - Extraction of features
  - Recognition
    - Template Matching
    - Classifiers
    - HMM

Gesture recognition

• Traditionally:
  – Few research on high-interactive gesture recognition
    • Annotation of videos
      [Laptev, Learning realistic human actions from movies, Computer Vision and Pattern Recognition, 2008]
  – Researchers focus on accuracy
    [Alon, A unified framework for Gesture Recognition and Spatiotemporal Gesture Segmentation, Transactions of Pattern Analysis and Machine Intelligence, 2009]
    • Current ChaLearn gesture challenge
• Two goal towards new action recognition:
  – Characterizing the action to understand human state of mind
    • Analysis of motion characteristics
  – Reducing recognition time for on-line use
    • Early gesture recognition
Analysis of motion characteristics
Analysis of motion characteristics

• Motion characteristics

• Application
  – Human state of mind (excitation, tiredness)
  – Medical assessment
  – “game” interaction
Analysis of motion characteristics

• Common gesture recognition
  – Extraction of numerous features
  – Based on Images and/or Skeleton
  – Learning models on these numerous features
    ➢ No precise study of each feature

• Speed:
  – Main characteristic of motion
  – Important role as feature for gesture recognition

[Yoon, Hand gesture recognition using combined features of location, angle and velocity, Pattern Recognition, 01]
[Rehm, Wave Like an Egyptian – Accelerometer Based Gesture Recognition for Culture Specific Interactions, British HCI, 08]
• Goal: classify motion speed

• Hand Motion
  – Raising hand motion
  – 10 subjects
  – 3 categories (slow, medium, fast)
  – 10 trials per subject and category

• Compare different features
  – Use classification error
Example: data collection for different speeds

• Filtered and oriented of 45 degrees
Performance over 300 motions

Non Filtered

Filtered

SD of hand instantaneous acceleration

Motions

0 50 100 150 200 250 300

0 10 20 30 40

Fast Medium Slow

Misclassifications
## Results

| Step | Filtered |   |  |   |  |  |   |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|      |          | Hand | Wrist | Elbow | Shoulder |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
|      |          | speed | accel | speed | accel | speed | accel | speed | accel | speed | accel | speed | accel | speed | accel | speed | accel | speed | accel | speed | accel | speed | accel | speed | accel | speed | accel | speed | accel |
| 1    | 12.7     | 10.7  | 46.3  | 10.7  | 14.7  | 12.0  | 47.0  | 13.0  | 20.0  | 21.3  | 48.0  | 20.3  | 43.0  | 46.3  | 37.3  | 38.3  |
| 2    | 12.7     | 10.0  | 43.3  | 10.0  | 14.7  | 12.7  | 46.0  | 11.3  | 20.0  | 21.3  | 47.3  | 20.3  | 43.3  | 46.7  | 36.7  | 38.7  |
| 3    | 12.0     | 10.0  | 41.7  | 9.3   | 14.7  | 12.7  | 44.0  | 12.0  | 19.3  | 21.3  | 46.3  | 19.3  | 43.3  | 46.7  | 36.3  | 37.7  |
| 4    | 11.3     | 10.7  | 40.7  | 9.3   | 14.0  | 12.7  | 42.7  | 11.3  | 18.7  | 22.0  | 45.7  | 18.7  | 42.7  | 46.0  | 35.3  | 37.0  |
| 5    | 10.7     | 10.7  | 40.3  | 8.7   | 14.0  | 12.7  | 42.7  | 10.7  | 17.3  | 21.3  | 44.3  | 16.7  | 42.7  | 46.0  | 35.7  | 36.0  |
| 6    | 10.0     | 12.0  | 40.0  | 8.0   | 13.3  | 13.3  | 40.7  | 10.7  | 16.7  | 22.0  | 44.0  | 17.3  | 42.3  | 45.3  | 34.7  | 35.0  |
| 7    | 10.0     | 12.0  | 38.3  | 8.7   | 12.0  | 14.0  | 40.3  | 10.0  | 17.3  | 22.0  | 42.0  | 16.0  | 42.7  | 45.3  | 33.7  | 35.0  |
| 8    | 10.0     | 12.7  | 38.3  | 9.3   | 11.3  | 15.3  | 40.0  | 9.3   | 17.3  | 22.0  | 41.7  | 16.7  | 42.0  | 45.3  | 32.0  | 35.0  |
| 9    | 10.0     | 13.3  | 36.0  | 10.0  | 10.7  | 17.3  | 37.0  | 10.7  | 17.3  | 22.0  | 39.0  | 18.0  | 42.0  | 45.3  | 31.3  | 35.7  |
| 10   | 9.3      | 16.0  | 33.7  | 10.7  | 10.7  | 18.0  | 34.7  | 10.0  | 17.3  | 23.3  | 34.0  | 18.7  | 42.0  | 45.7  | 32.0  | 38.0  |

- Accuracy per joint
- Speed vs. Acceleration
- Standard deviation vs. Average.
- Resampling
- Filtered vs. Non filtered
Training the Virtual Human

- Optimizing the threshold for accurate detection of hand movement
“Challenging the Virtual Human”
Conclusion

• Acceleration based feature
• Good accuracy of Non filtered gestures without low pass filter
  ➢ speed up the decision making of the VH


• Future work:
  – Investigate other characteristic
  – Correlation with subject characteristics (age/height)
  – Improve early detection of characteristic
Gesture recognition in interactive time
Gesture recognition in interactive time

• Reduce recognition time
  – Observational latency vs. Computational latency
• Detect gesture based on the first few instant of a movement
Gesture recognition in interactive time

• Motion extrapolation
  [Hasegawa, Human-Scale Haptic Interaction with a Reactive Virtual Human in a Real-Time Physics Simulator, ACM Computers in Entertainment, 06]

• Progressive block matching for dancing
  [Tang, Interactive Dancing Game with Real-time Recognition of Continuous Dance Moves from 3D Human Motion Capture, International Conference on Ubiquitous Information Management and Communication, 11]
Gesture recognition in interactive time

• Reducing observational latency
  [Masood, Measuring and Reducing Observational Latency when Recognizing Actions, International Conference on Computer Vision Workshops, 2011]
  – Limited features to position in frames
  – Better than baseline. Feature based only on distance between joints
  – No need for temporal segmentation

• Similarities
  – Classify on sliding window

• We take dynamic of motion into consideration
  – Add temporal segmentation
  – Reducing threshold
Gesture recognition

• Recording of gesture samples: 8 gestures, 3 different starting points, ~50 samples per gestures*starting point
• Gesture filtering
• Computation of features: speed, acceleration, main direction of movement
• Training SVM classifier and observing accuracy
Preliminary Results

• Trained SVM:

<table>
<thead>
<tr>
<th>All trajectory</th>
<th>10 first relevant frames (0.33s)</th>
<th>5 first relevant frames (0.16s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>~98%</td>
<td>~92%</td>
<td>~80%</td>
</tr>
</tbody>
</table>

➢ Trained classifier accuracy drops on different users
  – Add features to classifier, add more training data
  – Improve detection of movement starting point
Preliminary Results

• On speed classification using best feature:
Gesture prediction

• Conclusion:
  – Early detection has potential
  – More robust features
  – More training data

• Future work
  – Evaluation of dynamics in features
  – Compare with State of Art
  – Compare with HMM
Questions
Data collection include jittering effect

- Jittering is effect of joint occlusion
Two Examples of Filtering impact on Inst. Speed and Inst. Acc

- Correcting sensor errors (Butterworth low pass filter)
Comparison between filtered and non-Filtered (45 degree)

Non Filtered

Filtered
Joints comparison

- Motion at medium speed
- Instantaneous speed of Hand Joint
• Hand & Wrist: similar amplitudes and low frequencies.
• Elbow: Similar characteristics but lower amplitudes
Joints comparison

- Instantaneous speed of Hand Joint
- Relevance according to joint
Joints comparison

- Shoulder: Very small amplitudes
- Q2: Are Hand and Wrist joints more relevant?
Pre-Selection of Features

- Initial study
- Features based on speed looks promising
- Features based on acceleration less relevant

<table>
<thead>
<tr>
<th>FEATURE</th>
<th>NON FILTERED ERROR IN %</th>
<th>FILTERED ERROR IN %</th>
</tr>
</thead>
<tbody>
<tr>
<td>SD Speed Wrist</td>
<td>10.33</td>
<td>1.33</td>
</tr>
<tr>
<td>SD Speed Hand</td>
<td>6.33</td>
<td>2</td>
</tr>
<tr>
<td>Avg Speed Hand</td>
<td>5.33</td>
<td>3.33</td>
</tr>
<tr>
<td>Avg Speed Wrist</td>
<td>6.67</td>
<td>3.33</td>
</tr>
<tr>
<td>Motion Time</td>
<td>3.33</td>
<td>3.33</td>
</tr>
<tr>
<td>Global speed (hand)</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>SD Accel Wrist</td>
<td>43.33</td>
<td>15.33</td>
</tr>
<tr>
<td>SD Accel Hand</td>
<td>38.00</td>
<td>16.67</td>
</tr>
<tr>
<td>Straight distance Hand (start position - end position)</td>
<td>32.67</td>
<td>34</td>
</tr>
<tr>
<td>Avg Accel Hand</td>
<td>52.33</td>
<td>37</td>
</tr>
<tr>
<td>Avg Accel Wrist</td>
<td>56.33</td>
<td>41</td>
</tr>
<tr>
<td>AVG angle Elbow</td>
<td>42.33</td>
<td>42.67</td>
</tr>
<tr>
<td>AVG angle vel Elbow</td>
<td>63.33</td>
<td>57.67</td>
</tr>
</tbody>
</table>
Pre-Selection of Features

- Motion time is too motion dependent
- Features based on elbow angle have been discarded

➢ Focus on speed and acceleration
➢ Focus on standard deviation and average

### FEATURE

<table>
<thead>
<tr>
<th></th>
<th>NON FILTERED ERROR IN %</th>
<th>FILTERED ERROR IN %</th>
</tr>
</thead>
<tbody>
<tr>
<td>SD Speed Wrist</td>
<td>10.33</td>
<td>1.33</td>
</tr>
<tr>
<td>SD Speed Hand</td>
<td>6.33</td>
<td>2</td>
</tr>
<tr>
<td>Avg Speed Hand</td>
<td>5.33</td>
<td>3.33</td>
</tr>
<tr>
<td>Avg Speed Wrist</td>
<td>6.67</td>
<td>3.33</td>
</tr>
<tr>
<td><strong>Motion Time</strong></td>
<td><strong>3.33</strong></td>
<td><strong>3.33</strong></td>
</tr>
<tr>
<td>Global speed (hand)</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>SD Accel Wrist</td>
<td>43.33</td>
<td>15.33</td>
</tr>
<tr>
<td>SD Accel Hand</td>
<td>38.00</td>
<td>16.67</td>
</tr>
<tr>
<td>Straight distance Hand (start position - end position)</td>
<td>32.67</td>
<td>34</td>
</tr>
<tr>
<td>Avg Accel Hand</td>
<td>52.33</td>
<td>37</td>
</tr>
<tr>
<td>Avg Accel Wrist</td>
<td>56.33</td>
<td>41</td>
</tr>
<tr>
<td><strong>AVG angle Elbow</strong></td>
<td><strong>42.33</strong></td>
<td><strong>42.67</strong></td>
</tr>
<tr>
<td><strong>AVG angle vel Elbow</strong></td>
<td><strong>63.33</strong></td>
<td><strong>57.67</strong></td>
</tr>
</tbody>
</table>
Results

- Feature Selected: standard deviation of acceleration of hand joint

![Box plots showing acceleration data for different speeds](image-url)
Gesture speed classification

Updates:

• Processing new data => Similar results than Casa paper
  – Range of values are similar
  – Best features are also the same
  – The accuracy seemed better: better pre-processing of the data
Motion speed Classification

- Low pass filter + resampling
  - improves the speed recognition
  - However, more gestures need to be considered
- Good accuracy of Non filtered gestures without low pass filter
  - speed up the decision making of the VH
- Standard deviation of speed or acceleration and average of speed
  - provides similar accuracy rates
Feature clustering

Clapping

Bowing

Shaking hand

Scratching

PhD Seminar – November 2012