Pointing Gesture Prediction using Minimum-Jerk Model

Ren Ohmura
ren@tut.jp
Toyohashi University of Technology, Japan
Contents

• Introduction
• Existing Study and Proposed Method
• Detail of Proposed Method
• Performance Evaluation
• Implementation on a Robot
• Impression Evaluation
Introduction

• Assumption:
  “Joint attention with *simultaneous behavior* improves naturalness, rhythm in human-robot interaction. It will also increase the dependability of a robot by giving the feeling that robot understand human.”
Existing Study

- [Sugiyama et al. 07] showed some advantages of simultaneous behavior in human pointing gesture.

With simultaneous behavior

Without simultaneous behavior

Proposed Method

• Optical motion capture system
  - Kinect (3D Depth Camera)
    - The size can be installed into a human size robot.
    - The human motion can be captured without any additional equipment on human.

• Prediction Method (Extending velocity vector)
  - Exploiting Minimum-Jerk Model
    - One of the mathematical models of human motion.
    - Analytical solution can be easily obtained.
Minimum-Jerk Model

Flash & Hogan (1982)

Minimum change of acceleration (Jerk) → Obtain smooth movement

\[ C_J = \frac{1}{2} \int_0^{t_f} \left( \left( \frac{d^3 x}{dt^3} \right)^2 + \left( \frac{d^3 y}{dt^3} \right)^2 \right) dt \rightarrow \min \]

\[ x, y \quad \text{position of a fingertip} \]
\[ t_f \quad \text{time of movement} \]

Minimize change of acceleration (Jerk)

Feature

- Easy to get analytical solution
  \[ x(\tau) = x_0 + (x_f - x_0)(6\tau^5 - 15\tau^4 + 10\tau^3) \]
  \[ \text{where} \quad \tau = \frac{t}{t_f} \]

- Result trajectory get straight line
- Velocity get symmetric bell shape

Close to real movement of subject’s fingertips
Minimum-Jerk Model

\[ x(\tau) = x_0 + (x_f - x_0)(6\tau^5 - 15\tau^4 + 10\tau^3) \quad (\tau = t / t_f) \]

(When initial and final velocity and acceleration are zero)

\[
\begin{align*}
\tau_1 &= \frac{3 - \sqrt{3}}{6} \approx 0.21 \\
\tau_2 &= \frac{3 + \sqrt{3}}{6} \approx 0.79
\end{align*}
\]

The \( \tau_1 \) is when slope of acceleration get zero

The point of a fingertip can be estimated as

\[ x_f = \frac{4x(\tau_1) - (2 + \sqrt{3})x_0}{2 - \sqrt{3}} \]

And the time is,

\[ t_f = \frac{t}{0.21} \]
Fitting to Minimum-Jerk Model

Preliminary experiment

- Velocity of Pointing Left
- Velocity of Pointing Right
- Velocity of Pointing Down
- Acceleration of Pointing Left
- Acceleration of Pointing Right
- Acceleration of Pointing Down

Real Movement
Minimum-Jerk Model
Prediction Method

1. Detect the position of user’s wrist
2. Calculate the velocity
3. Is the velocity greater than the threshold?
   - Yes: Record the initial position and the time
   - No: Calculate the acceleration
4. Is it the acceleration peak?
   - Yes: Estimate the final position of the wrist with minimum-jerk model
   - No: Continue...
5. Estimate pointing target
Velocity

- Detect the position of user’s wrist
- Calculate the velocity
  - Is the velocity greater than the threshold
    - Yes: Record the initial position and the time
    - No: Calculate the acceleration
  - Is it the acceleration peak
    - Yes: Estimate the final position of the wrist with minimum-jerk model
    - No: Calculate the acceleration

Output: three-dimensional coordinates of each position of body parts.
(eg. Wrist, Head, Shoulder, and etc…)
- Precision: less than 10mm
- Frame rate: 30Hz

About the subject’s wrist, obtain
- the velocity from the position difference between two frames
Detection of the start of pointing gesture

- Detect the position of user's wrist
- Calculate the velocity
  - Is the velocity greater than the threshold
    - No
    - Yes
- Record the initial position and the time
- Calculate the acceleration
  - Is it the acceleration peak
    - No
    - Yes
- Estimate the final position of the wrist with minimum-jerk model
- Estimate pointing target

Obtain the threshold value

Histogram of the wrist velocity

- Not pointing
- Pointing

Set $v_{th}$ 200mm/s
Acceleration

Detect the position of user’s wrist

Calculate the velocity

Is the velocity greater than the threshold

Record the initial position and the time

Calculate the acceleration

Is it the acceleration peak

Yes

Estimate the final position of the wrist with minimum-jerk model

Estimate pointing target

No

Output: three-dimensional coordinates of each position of body parts.
(eg. Wrist, Head, Shoulder, and etc…)

Precision : less than 10mm
Frame rate : 30Hz

About the subject’s wrist, obtain
• the acceleration from the velocity difference
Peak Detection

- Detect the position of user’s wrist
- Calculate the velocity
- Is the velocity greater than the threshold?
  - Yes: Record the initial position and the time
  - No: Calculate the acceleration
- Is it the acceleration peak?
  - Yes: Estimate the final position of the wrist with minimum-jerk model
  - No: Estimate pointing target

For simplicity and speed, the peak of acceleration is found by the first time when

\[ a_t < a_{t-1} \]

And, start next operation immediately (use position and time on \( t-1 \) as “current”)

\[ \text{TOYOHASHI UNIVERSITY OF TECHNOLOGY} \]
Estimate Final Position

1. Detect the position of user’s wrist
2. Calculate the velocity
3. Is the velocity greater than the threshold?
   - Yes: Estimate pointing target
   - No: Record the initial position and the time
4. Calculate the acceleration
5. Is it the acceleration peak?
   - Yes: Estimate the final position of the wrist with minimum-jerk model
   - No: Calculate the acceleration
6. Recorded initial position: $x_0$
7. Current position: $x(t_1)$
8. Estimated final position: $x_f = \frac{4x(t_1) - (2 + \sqrt{3})x_0}{2 - \sqrt{3}}$
9. Recorded initial time: $t_0$
10. Current time: $t_1$
11. Estimated finishing time: $t_f = \frac{t_1 - t_0}{0.21}$
Estimation of the Target

Extends the line between the head and hand position

$$(x_{\text{head}}, y_{\text{head}}, z_{\text{head}})$$

$$(x_{\text{hand}}, y_{\text{hand}}, z_{\text{hand}})$$

$$(x_{\text{point}}, y_{\text{point}}, z_{\text{point}})$$

$${x_{\text{pre}}} = \frac{(y_{\text{point}} - y_{\text{hand}})(x_{\text{head}} - x_{\text{hand}})}{(y_{\text{head}} - y_{\text{hand}})} + x_{\text{hand}}$$

$${z_{\text{pre}}} = \frac{(y_{\text{point}} - y_{\text{hand}})(z_{\text{head}} - z_{\text{hand}})}{(y_{\text{head}} - y_{\text{hand}})} + z_{\text{hand}}$$

$y_{\text{point}}$ is currently fixed on the floor for Simplicity.
Performance Experiment Setting

Place three objects on subject’s right, left, and lower front, and ask him/her to point each object twenty times.

Compare estimated position and real position of subjects' wrist.
Precision of the Position Estimation

Comparing estimated position and real position of subjects wrist

- Average error,
- standard deviation,
- and circular error probable (CEP※)

<table>
<thead>
<tr>
<th></th>
<th>CEP[mm]</th>
<th>X Axis</th>
<th>Y Axis</th>
<th>Z Axis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ave.</td>
<td>STD</td>
<td>Ave.</td>
<td>STD</td>
</tr>
<tr>
<td>Total</td>
<td>177.22</td>
<td>-12.83</td>
<td>54.00</td>
<td>28.06</td>
</tr>
<tr>
<td>Left</td>
<td>164.79</td>
<td>-24.70</td>
<td>72.89</td>
<td>-7.33</td>
</tr>
<tr>
<td>Down</td>
<td>139.58</td>
<td>-35.18</td>
<td>37.11</td>
<td>107.81</td>
</tr>
<tr>
<td>Right</td>
<td>223.29</td>
<td>21.39</td>
<td>15.74</td>
<td>-16.30</td>
</tr>
</tbody>
</table>

Total CEP is 177.22mm
Average error of each axis is under 200mm

※ CEP is the radius of a circle, centered about the mean, whose boundary is expected to include the landing points of 50% of the rounds.
Impression Experiment Setting

Place three objects on subject's right, left, and lower front, and ask him/her to point each.
Impression Experiment Setting

The following steps are iterated under three conditions.
(1) select condition randomly,
(2) ask subjects to point arbitral objects with no explanation of current condition,
(3) and, ask subjects to answer questionnaire sheet.

Three conditions
• “Proposed Method”
• “Predict after 0.3 sec”
  – The robot turns to the position which is defined by extending current velocity vector of subject’s hand movement for 0.3 sec.
  – Used in [Sugiyama et al. 07]
• “Without Prediction”
  – The robot turns to the position which is detected by current sensor output.
Implementation Issue

A robot used in the experiment only replays a predefined motion.

Divide a space into 9x9 lattices, and define postures corresponding to each lattice. Then replay postures that are smoothly connected by control program on the robot.

A detection error and robot control delay significantly defer the start of the motion in proposed method.

The robot is controlled with “prediction after 0.3 sec” method before the detection of the start of pointing gesture in proposed method.
Video

Predict after 0.3 sec  Without prediction

Proposed method
Questionnaire Sheet

- Place a circle where your impression about the pointing gesture of the robot is appropriately explained

  Fast               Slow
  Natural           Unnatural
  Accurate         Inaccurate
  Smooth          Unsmooth
  Human               mechanical
  Gentle                  Scary
  Promptly             Unready
  Good timing        Bad timing
  Sensitive        Insensitive
  Fluent             Awkward
  Intelligent  Unintelligent
  Pleasant               Unpleasant

- Please answer your opinion in five levels for following questions

<table>
<thead>
<tr>
<th>agree</th>
<th>5</th>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
<th>disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Do you think the robot turned to where you pointed?</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Do you feel your behavior is predicted by the robot?</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Please write something you think about this system.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Impression Evaluation

Decrease the feeling of correctness, probably by *prediction error*, but increase the feeling that robot predict subject’s behavior.
# Impression Evaluation

<table>
<thead>
<tr>
<th>Impression</th>
<th>Proposed - w/o Prediction</th>
<th>Proposed - Predict 0.3 sec</th>
<th>w/o Prediction - Predict 0.3 sec</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fast</td>
<td>O</td>
<td>×</td>
<td>O</td>
</tr>
<tr>
<td>Natural</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>Accurate</td>
<td>O</td>
<td>O</td>
<td>×</td>
</tr>
<tr>
<td>Smooth</td>
<td>O</td>
<td>O</td>
<td>×</td>
</tr>
<tr>
<td>Human</td>
<td>×</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>Gentle</td>
<td>×</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>Promptly</td>
<td>O</td>
<td>×</td>
<td>O</td>
</tr>
<tr>
<td>Good Timing</td>
<td>×</td>
<td>Δ</td>
<td>×</td>
</tr>
<tr>
<td>Sensitive</td>
<td>Δ</td>
<td>×</td>
<td>O</td>
</tr>
<tr>
<td>Fluent</td>
<td>O</td>
<td>O</td>
<td>×</td>
</tr>
<tr>
<td>Intelligent</td>
<td>×</td>
<td>Δ</td>
<td>×</td>
</tr>
<tr>
<td>Pleasant</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
</tbody>
</table>

○: p<0.05, △: p<0.10
Features of Proposed Method

- Comparing with “w/o Prediction”
  - Fast, **Natural**, **Smooth**, Promptly, **Fluent**
    (, Sensitive)
  - Inaccurate

- Comparing with “Predict after 0.3 sec”
  - **Natural**, **Smooth**, Human, Gentle, **Fluent**
  - Inaccurate
    (, Bad timing, Unintelligent)

Good Impression especially for **Naturalness**, **Smoothness**, and **Fluency**

Feeling of **Inaccurate**

Increase **Human-likeness**, and **Gentleness** from “Prediction after 0.3sec”
Summary

• We focus on joint attention with simultaneous behavior to improve the communication quality with robots.
• A method to predict human’s pointing gesture with the Minimum-Jerk model is proposed.
• The precision is about 20cm.
• The method increases the feeling of “be predicted” (“be understood”).
• Although prediction error decrease the feeling of accuracy, the method increases naturalness, smoothness, and fluency. Moreover, human-likeliness and gentleness are improved from existing method.