3D FIXATIONS IN REAL AND VIRTUAL SCENARIOS

Thies Pfeiffer, A.I. & V.R. Group, Bielefeld University
Background

- multimodal human computer interaction
- situated natural communication (gaze, gesture, speech)
- natural interaction with dense information displays
Why should we be interested in automatic reconstructions of the fixated area within 3D space?

- Gaze is essential in natural communication
  - Turn-taking (negotiating who's up to speak next)
  - Focus of attention (resolving references, deictic gaze)

- Basic research
  - Visual world paradigm in 3D (e.g. spatial relations regarding the distance from the observer)

- Application
  - Virtual agents (Duchowski et al. 2004)
  - Optimized rendering in virtual reality (Lübke et al. 2000)
  - Selecting/picking objects (Tanriverdi und Jacob 2000; Duchowski et al. 2002; Barabas et al. 2004)
State of the Art

- monocular fixations extended to 3D
  1. calculate 2D fixations on a display
  2. extrapolate by casting a ray from the eye through the fixation into the scene

- problems
  - naive 3D fixations only possible when the ray hits an object
  - foreground vs. background problematic
  - ambiguities
Ambiguities

- **Underspecification**

- **Overspecification**

Thies Pfeiffer, Bielefeld University
Ambiguities

- Underspecification
- Overspecification

Idea: determine the depth of the fixation

Thies Pfeiffer, Bielefeld University
Idea: determine the depth of the fixation

Thies Pfeiffer, Bielefeld University
Open Questions

- What features can be used to reconstruct (in parts) the fixated area in 3D space?
  - accommodation
  - vergence

- What algorithms can be used?
  - geometric
  - adaptive (PSOM)

- How accurate does the eyetracker need to be?
  - low-res vs. high-res
Geometric Approach
Geometric Approach
Parameterized Self-Organizing Map

- developed by Ritter in 1993
- applied to anaglyphic stereo images by Essig et al. in 2006
- PSOM
  - input
    \((x_l, y_l), (x_r, y_r), x_r - x_l\)
  - output
    \((x, y, z)\)
# Eyetrackers – Technical Details

<table>
<thead>
<tr>
<th></th>
<th>Arrington PC60</th>
<th>SMI EyeLink I</th>
</tr>
</thead>
<tbody>
<tr>
<td>temporal resolution</td>
<td>30 Hz / 60 Hz</td>
<td>250 Hz</td>
</tr>
<tr>
<td>optical resolution</td>
<td>640x480 / 320x240</td>
<td>not specified</td>
</tr>
<tr>
<td>mean error</td>
<td>0.25° - 1.0°</td>
<td>&lt; 1.0°</td>
</tr>
<tr>
<td>Accuracy</td>
<td>0.15°</td>
<td>0.01°</td>
</tr>
<tr>
<td>compensation of head</td>
<td>not included</td>
<td>± 30° horiz.</td>
</tr>
<tr>
<td>movement</td>
<td></td>
<td>± 20° vert.</td>
</tr>
</tbody>
</table>

Thies Pfeiffer, Bielefeld University
Study

- 10 students tested

- **Hypotheses**
  - (a) **PSOM is better:**
    The PSOM is more accurate than the geometric solution.
  - (b) **EyeLink is better:**
    The SMI EyeLink I will deliver more accurate results than Arrington Research’s PC60.
  - (c) **Real is better:**
    In the real scenario we will be able to get more accurate results than in the virtual scenario.

Thies Pfeiffer, Bielefeld University
Scenario – Virtual Reality

Thies Pfeiffer, Bielefeld University
Scenario - Reality

Thies Pfeiffer, Bielefeld University
Results
Results: Geom. vs. PSOM

Thies Pfeiffer, Bielefeld University
Results: SMI vs. Arrington
Results

<table>
<thead>
<tr>
<th>device</th>
<th>algorithm</th>
<th>normally distributed</th>
<th>mean</th>
<th>difference btw. algorithms</th>
<th>nominal error</th>
<th>standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arr.</td>
<td>geom.</td>
<td>no, $p &lt; 0.001$</td>
<td>-195.77 mm</td>
<td>sig. $p &lt; 0.001$</td>
<td>sig. $p &lt; 0.001$</td>
<td>526.69 mm</td>
</tr>
<tr>
<td></td>
<td>PSOM</td>
<td>yes, $p = 0.943$</td>
<td>-18.75 mm</td>
<td>sig. $p &lt; 0.005$</td>
<td></td>
<td>96.92 mm</td>
</tr>
<tr>
<td>SMI</td>
<td>geom.</td>
<td>no, $p = 0.038$</td>
<td>-248.55 mm</td>
<td>sig. $p &lt; 0.001$</td>
<td>sig. $p &lt; 0.001$</td>
<td>149.3 mm</td>
</tr>
<tr>
<td></td>
<td>PSOM</td>
<td>yes, $p = 0.661$</td>
<td>-70.57 mm</td>
<td>sig. $p &lt; 0.001$</td>
<td>sig. $p &lt; 0.001$</td>
<td>60.06 mm</td>
</tr>
</tbody>
</table>

- **a)** is true: PSOM is more accurate and more precise
  - significant lower nominal error
  - lower standard deviation

- **b)** is twofold:
  - Arrington is more accurate
  - SMI is more precise
Results: Virtual vs. Real

Virtuell PSOM

Real PSOM

Thies Pfeiffer, Bielefeld University
Results: Virtual vs. Real

<table>
<thead>
<tr>
<th>Value</th>
<th>Virtual</th>
<th>Real</th>
</tr>
</thead>
<tbody>
<tr>
<td>normally distributed</td>
<td>Yes, p = 0.074</td>
<td>Yes, p=0.511</td>
</tr>
<tr>
<td>mean</td>
<td>-44.66 mm</td>
<td>-17.24 mm</td>
</tr>
<tr>
<td>std. deviation</td>
<td>84.61 mm</td>
<td>69.37 mm</td>
</tr>
</tbody>
</table>

□ c) is true: Real is better
Discussion

- 3D fixations can be reconstructed measuring the vergence angle and applying a PSOM algorithm.
- Accuracy is good, precision is less than expected from literature (Essig et al. 2006).
  - But “real world” objects have been used (not dots).
- Current advice for basic research:
  - Distribute critical objects at least 30cm apart when working with near objects.
- Next study will involve a larger scenario in VR (3m x 3m x 3m).

Thies Pfeiffer, Bielefeld University
... in collaboration with

Matthias Donner
Dipl. Informatiker

Dr. Marc E. Latoschik

Prof. Dr. Ipke Wachsmuth

Psychologically Augmented Social Interaction Over Networks

European Union program under IST, January 2006 - December 2009

Thies Pfeiffer, Bielefeld University