vGaze: Implicit Saliency-Aware Calibration for Continuous Gaze Tracking on Mobile Devices

Songzhou Yang, Yuan He, Meng Jin
School of Software and BNRist, Tsinghua University
Eye movement and Gaze

The gaze reflects where the user looks at.

But obtaining gaze direction is not the final goal!

From the user’s eye movement, we can infer the gaze direction.
Gaze Tracking Usage

• Gaze tracking usually acts as an interaction method.

• What we need is not the gaze direction, but the gaze on the screen!

How to get this correction?
To offset the drift, a **calibration** is required, which projects the eye movement to the screen coordinates by a transform matrix.
How the calibration is established?

- Traditional approaches require the user’s cooperation to gaze at stimulus points at predefined coordinates on the screen, known as explicit calibration.

![Dot-based Calibration](image1)

![Pattern-based Calibration](image2)

Effective, but ...
Gaze Tracking

One problem:
The projection only holds when the relative position between the head and the screen remain fixed.
However...

Such a calibration process takes too much time. Once the relative position changes, the re-calibration impairs the user’s experience heavily!
These areas are called **saliency**, which is a kind of visual information (i.e., distinctive color, intensity, orientation, objects, etc.)
In order to express saliency, the saliency map is generated. Basically, the saliency map can be used as a kind of implicit stimuli for calibration.
However...

• Saliency is ambiguous in many frames.

For a complicated frames, there are always more than one areas act as saliency.

Moreover, these saliency play different roles in spatial and temporal dimension.
To better understand saliency

• The human attention mechanism.

The user’s attention is in the **bottom-up** mode during the first around 150ms after scene cuts.

Then, the user’s attention enters the **top-down** mode, where the user’s consciousness dominates the gaze.
To better understand saliency

• Corresponding saliency.

<table>
<thead>
<tr>
<th>Bottom-up Saliency</th>
<th>Top-down Saliency</th>
</tr>
</thead>
<tbody>
<tr>
<td>salient because of their inherent properties relative to the background</td>
<td>based on prior knowledge, willful plans, and current goals</td>
</tr>
<tr>
<td>e.g., luminosity, shape</td>
<td></td>
</tr>
</tbody>
</table>

Specific kind of saliency should be used at specific time.
Our Idea

Leverage the temporally and spatially dependent relation between the saliency and the user’s attention.
Design

Calibration Module
Live Video / AR
Calibration window

Saliency Detection and Selection
Bottom-up Saliency
Top-down Saliency

Saliency Selection
(x1,y1)
(x2,y2)
(x3,y3)

Implicit Calibration

Trigger Calibration
If the head moved
Head Movement
Scene Cut
If scene cuts appeared

Tracking Module
Eye Tracking
Calibration window

Rough Gaze Tracking
Eye Posture Estimation
Gaze Position Estimation

Transform Vector

Calibrated Gaze Position

Update
Design

Calibration Module
Live Video / AR

Saliency Detection and Selection

Bottom-up Saliency
Top-down Saliency

(x1,y1)
(x2,y2)
(x3,y3)

Implicit Calibration

Calibrated Gaze Position

Tracking Module
Eye Tracking

Rough Gaze Tracking

Eye Posture Estimation
Gaze Position Estimation

Transform Vector

If the head moved
Head Movement

Scene Cut
If scene cuts appeared

Update

Trigger Calibration
Design

- Original frame and saliency map with different resolution
Design

Calibration Module
- Live Video / AR

Saliency Detection and Selection
- Bottom-up Saliency
- Top-down Saliency

Saliency Selection
- (x1, y1)
- (x2, y2)

Implicit Calibration

Tracking Module
- Rough Gaze Tracking
  - Eye Posture Estimation
  - Gaze Position Estimation

Transform Vector

Calibrated Gaze Position

If the head moved
- Head Movement

If scene cuts appeared
- Scene Cut

Calibration

Update

Design 18
Design

Calibration Module
Live Video / AR

Saliency Detection and Selection
Bottom-up Saliency
Top-down Saliency

Saliency Selection
(x1, y1)
(x2, y2)

Implicit Calibration

Trigger Calibration
If the head moved
Head Movement
Scene Cut
If scene cuts appeared

Tracking Module
Eye Tracking

Rough Gaze Tracking
Eye Posture Estimation
Gaze Position Estimation

Transform Vector

Calibrated Gaze Position
Implementation

- iPhone XS Max
  - ARKit & TrueDepth Camera
    - Eye movement tracking
  - IMU Sensors
    - Phone posture detection
Evaluation

• Evaluation setups:
  • 6 videos (from EyeTrackUAV dataset)
  • 10 volunteers
    • 5 males and 5 females
    • ages vary from 8 to 72 years old.

<table>
<thead>
<tr>
<th>Title</th>
<th>Duration</th>
<th>Resolution</th>
<th>Sample Rate</th>
<th>Total frames</th>
</tr>
</thead>
<tbody>
<tr>
<td>bike3</td>
<td>14s</td>
<td>1280*720</td>
<td>30fps</td>
<td>432</td>
</tr>
<tr>
<td>boat6</td>
<td>27s</td>
<td>1280*720</td>
<td>30fps</td>
<td>804</td>
</tr>
<tr>
<td>boat8</td>
<td>23s</td>
<td>1280*720</td>
<td>30fps</td>
<td>684</td>
</tr>
<tr>
<td>building5</td>
<td>16s</td>
<td>1280*720</td>
<td>30fps</td>
<td>480</td>
</tr>
<tr>
<td>car6</td>
<td>73s</td>
<td>1280*720</td>
<td>30fps</td>
<td>2194</td>
</tr>
</tbody>
</table>
Evaluation

• Overall tracking errors: 1.51cm

ExPLICIT: explicit calibration with five dots
vGaze: our solution
vGaze-lite: only bottom-up saliency is used

Three scenario are involved, static (where the user stay static), dynamic (where the user is asked to move), natural (where the user’s movement is not constrained)
Evaluation

• Errors on three different scenarios

(a) Static
vGaze is comparable with explicit calibration on errors in static scenarios without interruption.

(b) Natural
vGaze is better than explicit calibration in other two scenarios.

(c) Dynamic
vGaze beats vGaze-lite in all scenarios.
Evaluation

• Landscape v.s. Portrait

• Phone held in hand
Evaluation

• Time elapsed by different modules

The average total time consumed by saliency detection and selection is **19.66 ms** for a frame, which is much shorter than the frame display interval 33.33ms of 30 FPS video/AR.

- Saliency Detection
- Saliency Selection
- Rough Gaze Tracking
- Calibration
- Compensation of rough gaze tracking and transform vector
In Summary

• With the insight of the temporal and spatial relation between the gaze and the visual saliency, we present the design and implementation of vGaze, implicit saliency-aware calibration for continuous gaze tracking on mobile devices.
  • Bottom-up saliency & Top-down saliency
  • High accuracy & Low latency

My email: yangsz18@mails.tsinghua.edu.cn
Thanks For listening

Q & A