Hyperion: A Wearable Augmented Reality System for Text Extraction and Manipulation in the Air


*Systems and Media Lab,
The Hong Kong University of Science and Technology
^Sharif University of Technology
Mobile Augmented Reality

- supplements the reality of mobile users with virtual objects
- The amount of virtual objects can be huge and the selection of the most suitable ones is not trivial
  - user’s context
  - device limitations
  - QoE guarantees
Optical Character Recognition

• extracts text from the current view of a mobile user
• can be used as an input to MAR applications
• is computationally expensive
  • computation offloading techniques are required.
## Computation Offloading

<table>
<thead>
<tr>
<th>Pros</th>
<th>Cloud</th>
<th>Cloudlets</th>
<th>Smartphones</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pros</strong></td>
<td>Resourceful Offloadees</td>
<td>Stable connection</td>
<td>Nearby Devices</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Small latency</td>
<td>No need for Internet Connection</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Management</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>No need for Internet Connection</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cons</th>
<th>Cloud</th>
<th>Cloudlets</th>
<th>Smartphones</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cons</strong></td>
<td>Internet Connection</td>
<td>Poor availability</td>
<td>Unstable connectivity</td>
</tr>
<tr>
<td></td>
<td>High latency</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **Pros:**
  - Resourceful Offloadees
  - Stable connection
  - Small latency
  - Management
  - No need for Internet Connection

- **Cons:**
  - Internet Connection
  - High latency
  - Poor availability
  - Unstable connectivity
Design principles

• Simplicity
  • a simple system for a daily use

• Relevance
  • focused to the needs of the user

• Adaptivity
  • adapted to the needs of the scenario
• Explore mode: Extracting text from billboards, direction boards etc
• Reading mode: Magnifying glass in the area that the user is pointing. Translate selected text. Clip a text area.
• Meeting mode: Extract text from presentations
• Driving mode: Extract text from traffic signs
• Each modality has different inputs and outputs
- Hand gestures (colour-based)
- Speech Synthesis (Text-to-Speech)
- Visual Display (OpenCV4Android)
Hand gesture detection

- Hyperion uses a lightweight colour-based detection algorithm.
- The method uses reference color for detection and requires to sample the color of the hand skin and then utilizes the skin color to extract pixel sections with similar color in current view.
## Explore

<table>
<thead>
<tr>
<th>Inputs</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>Touchpad</td>
<td>Visual Display</td>
</tr>
<tr>
<td>Voice commands</td>
<td></td>
</tr>
<tr>
<td>Hand gestures</td>
<td>Speech Synthesis</td>
</tr>
</tbody>
</table>

OCR: Department of Industrial Engineering and Logistics Management Direct Entry Admission Selection interview
Reading

<table>
<thead>
<tr>
<th>Inputs</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>Touchpad</td>
<td>Visual Display</td>
</tr>
<tr>
<td>Hand gestures</td>
<td></td>
</tr>
</tbody>
</table>
Abbreviation Resolution

English to Chinese translation

Reading

program less intrusive complete
co-resident with GPU programs
but is less intrusive than a complete
system software

Abbreviation

resolution

侵入的：打扰的

full completion of glose in all
threat. If the sees another process
reopening the file will observe con
rant updates, it does not guarantee
data has been written to disk or to
the CPU page cache, as the tran
might be performed asynchronously.
Unfortunately, in existing syst
is a Garbage program failure (due to, fo
example, an invalid memory acc
service failure) could require re
result, the

Graphics Pr

" ok glass "

ok glass"
A model that approximates natural human behavior might be very different from one that drives a synthetic movie actor.

"ok glass"
<table>
<thead>
<tr>
<th>Inputs</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>Touchpad</td>
<td>Visual Display</td>
</tr>
</tbody>
</table>

### Meeting

OCR:

Outline
Visualization with OSM data
GPS trace clustering
Some thoughts

"ok glass "
# Driving

<table>
<thead>
<tr>
<th>Inputs</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Voice commands</td>
<td>• Visual Display</td>
</tr>
<tr>
<td></td>
<td>• Speech Synthesis</td>
</tr>
</tbody>
</table>
Evaluation

• System performance evaluation

• Quality of experience evaluation

  • 4 cases, one for each modality: walking around campus, reading a book, having a meeting, and walking along a roadway.
System performance evaluation

- Google Glass
- Google Nexus 5 (2.26GHz CPU and 32 GB memory)
- Amazon EC2 instance
- Nearby cloudlet in 2hop distance
Running Time

All modes

Time(s)

Meeting  Driving  Explore  Reading

OTransmission  OExecution  Local

Offloading to Nexus

Using Bluetooth
Running Time

Meeting mode

- Local
- Nearby
- Cloudlets
- Amazon EC2

Time(s)
Running Time

![Bar chart showing time in seconds for different tasks: amplification, translation, and clipping.](chart.png)
Battery gain

![Battery Percentage vs Time Graph]

- **local**
- **offloading**

Explore mode
Battery gain
Rethinking visual impairment

- Affecting more than 285 million people around the world
- Blurry vision
- Cannot see objects at a distance
  - Hyperion detects text and
    - magnifies it or reads it to the user.
Likert scale

Perceived Ease of Use

Perceived Enjoyment

Intention of Use

Perceived Usefulness
Thank you :) 

Any questions ?