RetroSphere
Self-contained Passive 3D Controller Tracking for Augmented Reality

Ananta Narayanan Balaji*, Clayton Kimber#, David Li^, Shengzhi Wu#, Ruofei Du#, David Kim#

*National University of Singapore  #Google Research  ^University of Maryland, College Park

$This work was done while the author was interning at Google.
Smartphone Augmented Reality (AR)

AR Interaction (DepthLab, UIST’20)

AR Gaming (Pokemon Go)

AR Shopping (IKEA Place)
Background
Smartphone AR applications

AR Interaction (*DepthLab*, *UIST*’20)

AR Gaming (*Pokemon Go*)

AR Shopping (*IKEA Place*)
Background

Augmented Reality Glasses

Compared to Smartphones and VR headsets, AR glasses are

1. **Resource-constrained** (both battery and compute capacity)

2. **Thin and Lightweight form factor**

AR glasses still lack 3D inputs for interaction with the AR user interface.
Background

3D Tracking approaches in VR/AR headsets

Electromagnetic tracking
- Auraring

Infrared camera based tracking
- Oculus Quest 2
- Leap motion (Stereo Infrared camera and IR LEDs for hand tracking)

Camera-based hand tracking
- HTC Vive

Stationary infrastructure

Laser/MEMS based tracking

Smartphone camera based tracking of a cube marker
- ARPen/DoDecapen

Existing VR/AR tracking approaches incur **high power consumption** or are **computationally expensive**.
3D Tracking approaches in VR headsets

- Auraring (Electromagnetic tracking)
- Oculus Quest 2 (Infrared camera based tracking)
- HTC Vive (Infrared camera based tracking)
- Laser/MEMS based tracking
- Leap motion (Stereo Infrared camera and IR LEDs for hand tracking)
- ARPen/Dodecapen (Smartphone camera based tracking of a cube marker)

RetroSphere - 6DoF Tracking for AR Glasses
Background

3D tracking with AR glasses v.s VR Headsets

- significant power and computation.
- not very portable for everyday AR use cases

VR based hand tracking and other continuous camera-based tracking isn't possible given the limited compute power and battery capacity.
Need for a low-power and low-compute self-contained 6DoF tracker that can provide spatial input on future resource-constrained AR devices.
Our proposed solution provides low-power 6 DoF tracking and a stylus that doesn't require any electronics or charging.
Enable spatial 3D interaction on portable devices with form-factor and power constraints.
Our solution is self-contained and can augment existing AR glasses for prototyping purposes or built into future products.
Retroreflective markers are used in motion capture systems such as Optitrack, Vicon etc.
An IR tracker only tracks IR light emitting sources unlike an IR camera. In addition, IR trackers are tiny and consume low power.

Johnny Lee (Head Tracking with IR trackers)

3D tracking of a Wiimote IR tracker (Franklin Ta)
Stereo infrared blob tracking cameras

Colocated IR LED illuminators for tracking our retroreflective stylus

On-device 6DoF estimation pipeline for tracking our RetroReflective stylus

- Always-on low-power 6DoF tracking hardware
- Low-cost and affordable hardware
- Easy integration into future AR glasses
RetroSphere - 6DoF Tracking for AR Glasses
System Design
Passive Retroreflective Stylus

Passive stylus - consists of 3 Retroreflectors each of unique size and triangular geometry

Supports 3/5/6DoF tracking

It can be easily integrated into other form factors such as wristband, pencil, ring etc.
Passive stylus - consists of 3 Retroreflectors each of unique size and triangular geometry

Supports 3/5/6DoF tracking

It can be easily integrated into other form factors such as wristband, pencil, ring etc.
System Design
Overview of RetroSphere

Infrared LEDs illuminate the retroreflective Spheres

RetroSphere stylus with three retroreflective spheres

Infrared light is reflected from the retroreflective markers back to the infrared blob tracking cameras

Left camera image plane

Right camera image plane

Left IR tracker

Position of the three retroreflective markers

Depth estimation from Stereo IR marker positions to get (X,Y,Z) point of the three retroreflective markers in real-world

Estimating the 6DoF orientation of the stylus using 3D positions of the retroreflective markers

50C
(x_left,y_left)

50C
(x_right,y_right)

ESP32 Microcontroller

Offers 10X power saving over existing passive 3D tracking approach.

3D input for AR applications

Sparse and requires very low compute - Run on a ESP32
Automatic calibration procedure

User randomly waves the stylus in mid air while wearing RetroSphere glasses (for 10-30 seconds).

Stereo 3D Depth estimation approach

Neural network based prediction of occluded marker positions

3D coordinates of observed markers

3D coordinates of all 3 retroreflective markers

6DoF orientation of the stylus
A. Stereo Infrared camera Calibration

Chessboard pattern based stereo camera calibration
(To get the extrinsic parameters of camera to perform depth estimation)
A. Automatic Calibration Procedure in RetroSphere

User randomly waves the stylus in mid air while wearing RetroSphere glasses (for 10-30 seconds).

The calibration algorithm makes use of the geometry of the markers in our stylus to estimate the extrinsic parameters of the camera.
B. Stereo 3D depth estimation of each marker in the stylus

We now have the 3D coordinates of each marker in the stylus.
C. 6DoF estimation of each marker in the stylus

- For 6DoF orientation estimation, we need both the rotation and translation matrices.

- The reconstructed markers are then matched with their corresponding object markers using the size as well as the triangular geometry of the markers \((A, B, C)\) in the RetroSphere stylus.

- Compute the rotation \((R)\) and translation \((t)\) matrices from the three marker positions (real-world coordinate frame) and compare them with the object markers (coordinate frame of the stylus) until

\[
\|y_i - (Rx_i + t)\| < \text{tolerance}, \, \forall \, i \in \{1, 2, 3\}
\]
Neural Network-based Hand Occlusion Correction

- Is there any missing markers in the observed frames?
  - Yes
    - Compute the 3D coordinates for the two observed markers
    - 1 marker
    - Neural network to predict the 3D coordinates of a single missing marker
      - (X,Y,Z) of observed marker 1
      - (X,Y,Z) of observed marker 2
      - ID of marker 1
      - ID of marker 2
      - Output layer (3 neurons)
      - (X,Y,Z) of the missing marker
  - No
    - Compute the 3D coordinates for the observed marker
    - 2 markers
    - Neural network to estimate the 3D coordinates of two missing markers
      - (X,Y,Z) of observed marker
      - ID of the observed marker
      - (X,Y,Z) coordinates of the 3 markers for the previous 5 seconds
      - Input layer
      - Hidden layer (256 neurons)
      - Hidden layer (128 neurons)
      - Output layer (6 neurons)
      - (X,Y,Z) of the two missing markers

- Hand occluding a marker
- Hand occluding two markers
- User trials with 20 participants (11 M, 9 F)
- Wore our AR glass mockup prototype
- Participants used our unity applications made for our AR glass mockup
  - 3 minutes on the 3D user interface application
  - 3 minutes on the drawing application
  - 4 minutes on the mid-air visualization application
## Results

### Tracking Accuracy

<table>
<thead>
<tr>
<th>6DoF Parameters</th>
<th>Mean error</th>
<th>6DoF Parameters</th>
<th>Mean error</th>
</tr>
</thead>
<tbody>
<tr>
<td>X (mm)</td>
<td>3.2</td>
<td>Pitch (deg)</td>
<td>4.65</td>
</tr>
<tr>
<td>Y (mm)</td>
<td>4.3</td>
<td>Yaw (deg)</td>
<td>6.95</td>
</tr>
<tr>
<td>Z (mm)</td>
<td>12</td>
<td>Roll (deg)</td>
<td>4.85</td>
</tr>
</tbody>
</table>

**Position tracking error (mm)** 18.5

**Orientation tracking error (deg)** 5.85

### Depth sensing controllers

<table>
<thead>
<tr>
<th>Depth sensing controllers</th>
<th>Average error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kinect [33]</td>
<td>1% (4 cm)</td>
</tr>
<tr>
<td>Magic Leap [24]</td>
<td>1% (5 cm)</td>
</tr>
<tr>
<td>Intel RealSense D435 [7]</td>
<td>2 mm</td>
</tr>
<tr>
<td>POL360 [31]</td>
<td>0.691 cm</td>
</tr>
<tr>
<td>Oculus Quest [49]</td>
<td>3.5 mm</td>
</tr>
<tr>
<td><strong>RetroSphere</strong></td>
<td><strong>2.4% (~1.2 cm)</strong></td>
</tr>
</tbody>
</table>
Results
Tracking Accuracy

Hand Occlusion

Environment (Indoor/Outdoor)

Lighting Conditions

Motion (Stationary/Walking)
Retrosphere offers *at least 10X power savings against ARPen* (the only passive AR controller tracking approach) and a *frame rate of 66 fps (15 ms)*.
Demonstrations with RetroSphere
Demonstration

3D measurements

RetroSphere - 6DoF Tracking for AR Glasses
Demonstration

6DoF Mid-air Drawings
Demonstration

3D Drawings/Sculpting
Use Cases

RetroPen
Use Cases

RetroPen - Working Mechanism

- Writing
- Erasing
- Pressure sensitivity
- Tilt sensitivity
- Battery-free/Passive
Retroreflective rings for thumb and index finger (each with a unique size).

Can be used to make gestures for 3D UI controls.
Use Cases

RetroPen - Single Finger Ring

RetroSphere - 6DoF Tracking for AR Glasses
Use Cases

RetroPen - Two Finger Rings
We hope that RetroSphere will allow researchers and practitioners to study and prototype spatial input on lightweight AR glasses more easily.
THANK YOU