OpenARK -- Tackling Augmented Reality Challenges via an Open-Source Software Development Kit

Adam Chang, Dr. Mohammad Keshavarzi, Dr. Allen Y. Yang

University of California, Berkeley
Tutorial Contributors

Dr. Allen Y. Yang
• Executive Director, FHL Vive Center for Enhanced Reality @ Berkeley EECS Department
• Former Chief Scientist, Fung Institute @ Berkeley

Dr. Mohammad Keshavarzi
• PhD @ Berkeley Vive Center and XRLab
• Generative Design Software Developer @ Autodesk
• Researcher @ Facebook Reality Lab

Adam Chang
• Lead Undergraduate Developer for Berkeley OpenARK project
• Now work at Wall Street
Slides available: vivecenter.berkeley.edu

Mission

The main goals of the center are to sponsor critical fundamental research and high-impact applications in the emerging fields of Virtual Reality (VR), Augmented Reality (AR), and Artificial Intelligence (AI), and at the same time serve as the central hub to facilitate the deployment of disruptive VR, AR, and AI technologies across the Berkeley campus for cross-disciplinary research and education.

We aim to achieve these goals by offering seed grant to our faculty, supervising and facilitating student research activities, and fostering external industry partnerships with other stakeholders.
True Cost to Blur Reality and Virtuality

-- Understand the fundamental challenges in creating virtual 3D experience

Dr. Allen Y. Yang
with Adam Chang, Dr. Mohammad Keshavarzi

ISMAR 2022 Tutorial, Part I
What consumers believe future of AR/VR/Metaverse could look like
Sixty Plus Years of AR/VR

Morton Heilig, 1960

Morton Heilig, 1962

Ivan Sutherland, 1968
AR/VR System Form Factors

Simulation Environments

Tethered VR PCs

All-in-One Wearables

ARKit and ARCore on Smartphones
AR/VR System Form Factors Pros and Cons

<table>
<thead>
<tr>
<th>Form Factor</th>
<th>Simulation Environment</th>
<th>Tethered to PC</th>
<th>All-in-One</th>
<th>Smartphone</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cost</td>
<td>Very High</td>
<td>High</td>
<td>Medium</td>
<td>Very Low</td>
</tr>
<tr>
<td>Visual Fidelity</td>
<td>Best</td>
<td>High</td>
<td>Medium</td>
<td>Low</td>
</tr>
<tr>
<td>Motion-Tracking</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>Medium</td>
</tr>
<tr>
<td>User Input</td>
<td>Very Realistic</td>
<td>High</td>
<td>Low to Medium</td>
<td>Very Low</td>
</tr>
</tbody>
</table>

Our research will focus on Perception and UI/UX tasks for low-cost all-in-one applications.
Near-Eye Display Space (NDS)

- Renders 2D menu and items on head-mounted displays
  - Google Glass uses exclusively NDS
- Content in NDS typically is **not anchored with any physical objects**
- Limited user interaction in NDS:
  - **Voice**
  - **Gaze**
  - **2D Touch**
- NDS is often reserved for displaying private information, rarely shared between different users
Egocentric User Space (EUS)

- Egocentric User Space is an immersive virtual space with its origin **anchored from the user's perspective and typically moved with the user's movement**

- Typically, user interaction in EUS is projected within an arm's reach from user (1--2m):
  - Hand gestures
  - 3D Controllers

- EUS content if shared, such as video calls, is **rendered individually for each user**
Overlay World Space (OWS)

- OWS models and coincides with the physical world
  - HoloLens
  - ARKit

- Interacting with OWS content implies interacting with the real world

- Origin of OWS is anchored at a fixed 3D location, typically does not change w.r.t. the user movement

- In multi-user applications, an OWS and its content can be shared and manipulated by multiple users
# Pros and Cons: Three Interactive Spaces

<table>
<thead>
<tr>
<th></th>
<th>Near-Eye Display Space (NDS)</th>
<th>Egocentric User Space (EUS)</th>
<th>Overlay World Space (OWS)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Anchor</strong></td>
<td>Display Coordinates</td>
<td>Egocentric with arm length</td>
<td>World coordinate origin</td>
</tr>
<tr>
<td><strong>Interaction</strong></td>
<td>Voice, Gaze, 2D touch</td>
<td>Gestures, Controllers</td>
<td>Physical objects and environment</td>
</tr>
<tr>
<td><strong>Multi-User Sharing</strong></td>
<td>Very Rare</td>
<td>Rare</td>
<td>Popular</td>
</tr>
</tbody>
</table>
### Selected Challenges to Blur Reality and Virtuality

<table>
<thead>
<tr>
<th></th>
<th>Near-Eye Display Space (NDS)</th>
<th>Egocentric User Space (EUS)</th>
<th>Overlay World Space (OWS)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Display</strong></td>
<td>&quot;Retina Display&quot; for rendering NDS</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>Interaction</strong></td>
<td>Voice, Gaze, 2D touch</td>
<td>Perception for Egocentric 3D UI</td>
<td>• SLAM: Environment</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Digital Twin: Objects</td>
</tr>
<tr>
<td><strong>Multi-User Sharing</strong></td>
<td>N/A</td>
<td>N/A</td>
<td>• Contextual Spatial Computing</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Decentralized Privacy</td>
</tr>
</tbody>
</table>
"Retina" Near-Eye Display for Mobile AR/VR

Let's Calculate

• Fundamental Constraint:

**Human perceives discrete pixels as continuous picture if**

\[
\text{pixel-per-degree (PPD)} > 60
\]

• In handheld smartphone case:

\[
960 \text{ pixels} ÷ 60 \text{ ppd} = 16 \text{ degrees}
\]

Conclusion: a 960 resolution display is "retina display" when being viewed within 16 degrees

Apple started marketing "Retina Display" in 2010 that aims to eliminate "screen-door effect"
"Retina" Near-Eye Display

- Each eye has a horizontal field-of-view of $\sim160^\circ$ and a vertical field-of-view of $\sim175^\circ$. The two eyes work together for stereoscopic depth perception over $\sim120^\circ$ wide and $\sim135^\circ$ high FOV.

<table>
<thead>
<tr>
<th></th>
<th>PPD required</th>
<th>Horizontal FOV (deg)</th>
<th>Equivalent horizontal Kpixels</th>
<th>Vertical FOV (deg)</th>
<th>Equivalent vertical Kpixels</th>
<th>Total Mpixels required</th>
</tr>
</thead>
<tbody>
<tr>
<td>Each eye</td>
<td>60</td>
<td>160</td>
<td>9.6</td>
<td>175</td>
<td>10.5</td>
<td>$\sim100$</td>
</tr>
<tr>
<td>Stereo vision</td>
<td>60</td>
<td>120</td>
<td>7.2</td>
<td>135</td>
<td>8.1</td>
<td>$\sim60$</td>
</tr>
</tbody>
</table>

Retina Near-Eye Display that offers 120-degree FOV

- Minimally dual 8K resolution for near-eye AR/VR display
- Ideally dual 16K resolution
- How to transfer such amount of data: 5Gbps in single 8K format
"Semi-Retina" Near-Eye Display at CES 2018

LUCI alyx, First 4K Dual MicroOLED Near-Eye Display
If speed and low-latency is a concern, controllers are still the best current option.
Hand Tracking Solutions

Depth-based
Atheer

Camera-based
HoloLens, Microsoft

Radar-based
Soli, Google

Advantage of Depth Cameras

None of the other solutions can support hand gestures to manipulate 3D virtual objects in absolute 3D coordinates
Depth-Based Avatar Tracking

vivecenter.berkeley.edu/OpenARK
State of the Depth Cameras

- PrimeSense / Kinect 1 (Acquired by Apple)
- SoftKinetic (Acquired by Sony)
- PMD by Infineon (used on Magic Leap 2)
- Intel RealSense (Partially Discontinued)
- StereoLabs ZED 2 Camera
- Microsoft Azure Kinect
- Apple iPhone LIDAR
# Four Depth Cameras Technologies

<table>
<thead>
<tr>
<th>Pros</th>
<th>Structure from Motion</th>
<th>Depth from Stereo</th>
<th>Structured Light</th>
<th>Time-of-Flight</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>- 3D localization &amp; visual odometry</td>
<td>- Outdoor apps</td>
<td>- Indoor apps</td>
<td>- High-precision applications</td>
</tr>
<tr>
<td></td>
<td>- Large-scale reconstruction from big data photos</td>
<td>- Good tradeoff between accuracy and cost</td>
<td>- Low cost and low complexity</td>
<td>- 3D localization</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Accuracy on texture regions</td>
<td></td>
<td>- Dense 3D reconstruction</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cons</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>- Real-time 3D reconstruction and scanning</td>
<td>- High-precision applications</td>
<td>- High-precision applications</td>
<td>- Expensive</td>
</tr>
<tr>
<td></td>
<td>- Dark environments</td>
<td>- Hot/cold environments</td>
<td>- Limited range</td>
<td>- High power</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Limited range</td>
<td></td>
<td>- Limited resolution</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>- Not always work under the Sun</td>
</tr>
</tbody>
</table>
Quality of Depth Result is Critical for Applications

Intel RealSense D435i, Depth-from-Stereo technology

Azure Kinect, Time-of-Flight technology
Depth Perception for Room-Scale 3D Hologram

Microsoft Holoportation, 2016

Google Project Starline, 2021

Availability of Accurate Depth is Critical
3D Localization

Finding Virtual Asset in Physical World -- Pokemon Go AR Experience
Visual-Inertial Odometry in OpenARK
# 3D Localization Approaches

<table>
<thead>
<tr>
<th>Method</th>
<th>Outside-In</th>
<th>Inside-Out</th>
<th>Digital-Twin Localization</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Hardware</strong></td>
<td>Satellites or Base Stations</td>
<td>Dual Cameras, Synchronized IMU</td>
<td>Camera, LIDAR, IMU</td>
</tr>
<tr>
<td><strong>Algorithm</strong></td>
<td>Triangulation</td>
<td>Visual-Inertial Odometry</td>
<td>3D Object Localization</td>
</tr>
<tr>
<td><strong>Use Cases</strong></td>
<td>Installation of base stations in limited space</td>
<td>All-in-One Wearable AR/VR</td>
<td>Pre-trained on Known Objects, Utilize three sensing modalities</td>
</tr>
</tbody>
</table>

![Diagram](image1.png)  

![Diagram](image2.png)  

![Diagram](image3.png)
Simultaneous Localization and Mapping (SLAM)

Limitations

- Hardware sync (cameras, IMUs, LIDAR) is a must
- Dual camera is better than single
- More is not always better: Depth data are significantly more noisy (Azure Kinect do not natively support SLAM)
- Less is not always better: SLAM may fail when 3D surface lacks visual features
Digital-Twin 3D Localization

For example, here we set the can as our target object.
Spatial Computing Merging Virtual and Physical Worlds

Asset Placement across Platforms

- One code across multiple computer platforms (Java/HTML)
- Developer doesn't know screen size/resolution/orientation (Separation of content and layout: CSS)
- But, within the web browser, UI functionality is uniform

Spatial Computing Merging Virtual and Physical Worlds

Asset Placement in AR

- One code across multiple computer platforms (Game Engines)
- Developer doesn't know varying 3D space shapes
- Layout and furniture in 3D space have different contextual functions

Spatial Computing Merging Virtual and Physical Worlds

Unknown User Input

- User interacts with 3D virtual content from different perspective
- User's own posture and actions possibly drive the content
- User have expectations when interacting with other users

Does Multi-User Metaverse Compromise Privacy?

Privacy Concerns

• Other users may infer user's own private 3D space
• Other users may infer user's own meta data
• Bad actors may create fraudulent counter-parties to scam users

Challenges in Safeguarding User Privacy before the Internet

Nearly half of cellphone calls will be scams by 2019, report says

By Hassan Shahen
September 10, 2018 at 0:24 a.m. EDT

“robocalls”

Typical robo-calls are pre-recorded usually for political campaigns.

Nearly half of all cellphone calls next year will come from scammers, according to First Orion, a company that provides phone carriers and their customers caller ID and call blocking technology.
Challenges in Safeguarding User Privacy in Web 2.0

Twitter may have a lot of bot users

DeepMind AlphaStar pretends to be human players

Deepfake using Elon Musk image to promote scam
What "Ready Player One" is really about: Privacy and Decentralization
**Metaverse and Web 3.0: New Opportunities**

"*Metaverse isn't a thing a company builds. It's the next chapter of the internet overall.*"

--Mark Zuckerberg

<table>
<thead>
<tr>
<th></th>
<th>Who owns user data?</th>
<th>Trust &amp; Verify?</th>
<th>Interoperability?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Web 1 &amp; 2</td>
<td>Private Companies, Mutable</td>
<td>Lack of penalty for fraud</td>
<td>Centralized</td>
</tr>
<tr>
<td>Web 3.0</td>
<td>Users, Immutable</td>
<td>Smart contracts with financial penalties</td>
<td>Decentralized</td>
</tr>
</tbody>
</table>
The Next Chapter into the Metaverse: defi.berkeley.edu

Berkeley Defi Research Initiative

Mission

Connecting the world of individuals and businesses in fully decentralized and trustless fashion has generated tremendous excitement to develop new solutions in the blockchain and defi space. Although the technologies behind them are still considered far from mature, promising applications are taking shape in:

- Decentralized Finance / Cryptocurrencies
- Decentralized Social Media / Metaverse
- Decentralized Tokenization / NFTs

Berkeley Defi Research Initiative aims to support the long-term growth of blockchain and defi technologies and applications by offering our expertise to:

- Create new education and degree programs for motivated students
- Promote cross-disciplinary research to address urgent technology bottlenecks
- Provide a Berkeley platform for academic and industrial partners to collaborate

Together, we can shape our society in a digital world that is more immersive, more connected, and more fair.
True Cost to Blur Reality and Virtuality

-- Understand the fundamental challenges in creating virtual 3D experience

Thank you!

Dr. Allen Y. Yang
Email: yang@eecs.berkeley.edu