The ABC of Mixed Reality Interactions

Hrvoje Benko
Facebook Reality Labs
A new computing era!
Command Line Interfaces (text based)  |  Graphical User Interfaces (mouse + keyboard)  |  Natural User Interfaces (touch/gestures, tablets, smartphones)  |  Mixed Reality Interfaces
---|---|---|---
1960s  |  1980s  |  2000s  |  2020s
Command Line Interfaces (text based) | Graphical User Interfaces (mouse + keyboard) | Natural User Interfaces (touch/gestures, tablets, smartphones) | Mixed Reality Interfaces
--- | --- | --- | ---
1960s | 1980s | 2000s | 2020s
1968 Engelbart & English MOAD | 1983 Microsoft Mouse / Apple Mouse
<table>
<thead>
<tr>
<th>Command Line Interfaces (text based)</th>
<th>Graphical User Interfaces (mouse + keyboard)</th>
<th>Natural User Interfaces (touch/gestures, tablets, smartphones)</th>
<th>Mixed Reality Interfaces</th>
</tr>
</thead>
<tbody>
<tr>
<td>1960s</td>
<td>1980s</td>
<td>2000s</td>
<td>2020s</td>
</tr>
</tbody>
</table>

Bell B., Feiner, S., and Hollerer, T. *Columbia Touring Machine - ACM ISAR 2001*
Bell B., Feiner, S., and Hollerer, T. *Columbia Touring Machine - ACM ISAR 2001*
Bell B., Feiner, S., and Hollerer, T. *Columbia Touring Machine* - *ACM ISAR 2001*
“Imagine AR glasses that are socially acceptable and all-day wearable, that give you useful virtual objects like your phone, your TV and virtual work spaces, that give you perceptual super powers, a context-aware personal assistant, and above all the ability to connect, share and collaborate with others anywhere, any time. If those glasses existed today, we’d all be wearing them right now.”

Michael Abrash
Might be “obvious”...
... why not available today?
2001 vs. 2018 - Same Challenges

- Display
- Optics
- Battery
- Tracking/localization
- Compute power
- Spatial understanding
- Spatial audio
- Input/Interactions
2001 vs. 2018 - Same Challenges

- Display
- Optics
- Battery
- Tracking/localization
- Compute power
- Spatial understanding
- Spatial audio
- Input/Interactions
What kind of interactions will define the MR era?
Compelling MR interactions will be adaptive, believable, and computational.
Compelling MR interactions will be **adaptive**, **believable**, and **computational**.
Magic of MR interactions happens when they are tightly coupled to the user’s environment.
Projection AR

Projector

Depth Camera (Kinect)
Appearance
Steerable Augmented Reality with the Beamatron

Andy Wilson, Hrvoje Benko, Shahram Izadi and Otmar Hilliges
Microsoft Research

ACM UIST 2012
Magic of MR interactions happens when they are tightly coupled to the user’s environment.
Context

**environment**
(e.g., space geometry, object semantics, people around)

**user actions**
(e.g., gestures, body pose, bio-signals)

**user’s mental state**
(e.g., emotional, mental load, cognitive, focus)

**task**
(e.g., communication, navigation, calendar)
How the computer sees us!

How the phone sees us!

Example: Context-aware Tools
CLAW: *Multi-purpose* controller that *adapts* to the user’s context of use.

CLAW
Example: Hand interactions adapted to the environment
Shangchen Han, Beibei Liu, Robert Wang, Yuting Ye, Christopher D. Twigg, and Kenrick Kin. 2018. Online optical marker-based hand tracking with deep labels. ACM Trans. Graph. 37, 4, Article 166 (July 2018)
Accurate, Robust, and Flexible Real-time Hand Tracking

Toby Sharp†  Cem Keskin†  Duncan Robertson†  Jonathan Taylor†  Jamie Shotton†  David Kim  Christoph Rheinmann  Ido Leichter  Alon Vinnikov  Yichen Wei  Daniel Freedman  Pushmeet Kohli  Eyal Krupka  Andrew Fitzgibbon*  Shahram Izadi*

Microsoft Research

Figure 1: We present a new system for tracking the detailed motion of a user’s hand using only a commodity depth camera. Our system can accurately reconstruct the complex articulated pose of the hand, whilst being robust to tracking failure, and supporting flexible setups such as tracking at large distances and over-the-shoulder camera placement.

ABSTRACT

We present a new real-time hand tracking system based on a single depth camera. The system can accurately reconstruct complex hand poses across a variety of subjects. It also allows for robust tracking, rapidly recovering from any temporary failures. Most uniquely, our tracker is highly flexible, dramatically improving upon previous approaches which have required the user’s hand with gloves or markers can be cumbersome and inaccurate. Much recent effort, including this work, has thus focused on camera-based systems. However, cameras, even modern consumer depth cameras, pose further difficulties: the fingers can be hard to disambiguate visually and are often occluded by other parts of the hand. Even state-of-the-art academic and commercial systems are thus sometimes in
One small problem...

People don’t generally interact in mid air!

People interact with objects and on surfaces!
Hand interaction context = any available surface
MRTouch: Adding Touch to Head Mounted Mixed Reality.
Robert Xiao, Julia Schwarz, Nick Throm, Andrew D. Wilson, Hrvoje Benko. IEEE TVCG 2018 (Vol 24, No 4, April 2018)
Sample Interactions

(Continuous shot, first and only take)
FYI....

Hololens Research Mode API

Depth and reflectivity data now part of the public Research Mode API (developer mode only)

• https://github.com/Microsoft/HoloLensForCV
• Possible now to implement MRTouch using only public APIs
Compelling MR interactions will be adaptive, believable, and computational.
In MR, we are obsessed with creating a rich sense of \textit{reality}!
Deep Appearance Models for Facial Rendering

STEPHEN LOMBARDI, JASON SARAGIH, TOMAS SIMON, YASER SHEIKH
Facebook Reality Labs
For interactions, realistic is not always better
Believable ≠ Realistic
Believable = consistent with user’s expectations + non surprising
In AR/VR we want to induce the suspension of disbelief, whereby users suspend their critical faculties (i.e., sacrifice logic and realism) to believe the unbelievable.
Haptic Revolver
Touch, Shear, Texture, and Shape Rendering on a Reconfigurable VR Controller

Eric Whitmire¹, Hrvoje Benko², Christian Holz², Eyal Ofek², Mike Sinclair²

¹Paul G. Allen School, University of Washington
²Microsoft Research, Redmond
Figure 10. Results of the first user study showing mean realism ratings across participants as a function of the wheel speed gain. The error bars show a 95% confidence interval. A negative gain indicates the wheel was spun in the opposite direction.
Hand Movement

Wheel Movement
Haptic Retargeting

Passive Haptics Rocks!
Putting it all together...
Focusing on “as real as possible” designs can lead to sub-optimal MR experience.

Design for BELIEVABILITY, not REALISM.
Compelling MR interactions will be *adaptive*, *believable*, and *computational*.
<table>
<thead>
<tr>
<th>Command Line Interfaces (text based)</th>
<th>Graphical User Interfaces (mouse + keyboard)</th>
<th>Natural User Interfaces (touch/gestures, tablets, smartphones)</th>
<th>Mixed Reality Interfaces</th>
</tr>
</thead>
<tbody>
<tr>
<td>1960s</td>
<td>1980s</td>
<td>2000s</td>
<td>2020s</td>
</tr>
<tr>
<td>Location fixed</td>
<td>Location fixed</td>
<td>Location fixed</td>
<td>Mobile</td>
</tr>
<tr>
<td>Explicit (command driven)</td>
<td>Explicit (command driven)</td>
<td>Explicit (command driven)</td>
<td>Implicit</td>
</tr>
<tr>
<td>Sensing poor</td>
<td>Sensing poor</td>
<td>Sensing poor</td>
<td>Sensing rich</td>
</tr>
<tr>
<td>Single modality</td>
<td>Single modality</td>
<td>Single modality</td>
<td>Multimodal</td>
</tr>
<tr>
<td>Dedicated/cognitively capturing</td>
<td>Dedicated/cognitively capturing</td>
<td>Dedicated/cognitively capturing</td>
<td>Glanceable/always-on</td>
</tr>
<tr>
<td>Precise and accurate inputs</td>
<td>Precise and accurate inputs</td>
<td>Precise and accurate inputs</td>
<td>Imprecise and noisy inputs</td>
</tr>
</tbody>
</table>
AR Interactions

Iterative Design

Sensor Fusion

Machine Learning
Can you type on a phone touch keyboard?
Modern Android touch keyboard (two thumbs)

No correction – 27.5 WPM, error rate 6.5%
With correction – 31 WPM, error rate 1.1%

Probabilistic phone touch keyboard

Layers of probabilistic models:

- Touch precision model
- Dictionary model
- Language model

+ N-best list UI for error correction

+ Gesture model
Similar optimizations needed in MR

• Text entry
Similar optimizations needed in MR

- Text entry
- Object selection

Similar optimizations needed in MR

• Text entry

• Object selection

• (Multimodal) input fusion
Similar optimizations needed in MR

• Text entry

• Object selection

• Multimodal input fusion

• Output optimizations

Similar optimizations needed in MR

- Text entry
- Probabilistic object selection
- Multimodal input fusion
- Output optimizations

formulate the UI challenges as *computational problems* (inferring, sensor fusing, predicting, tolerating noise)

that adapt the interface depending on the user and world context
Compelling MR interactions will be adaptive, believable, and computational.
Maybe it should have been the CAB of MR Interactions?

- Computational Approaches
- Adaptive Interfaces
- Believable Experiences
Design interactions that adapt to the user’s actions, the world around them, and the context of use.
Focus on believability. Reality is overrated!
Harness the computational methods to overcome uncertainty, scale, noise, and enable adaptivity.
Hrvoje Benko
benko@fb.com
Facebook Reality Labs

Thanks to all my collaborators!
Extra Slides
Sparse Haptic Proxy: Touch Feedback in Virtual Environments Using a General Passive Prop

Lung-Pan Cheng\textsuperscript{2,1}, Eyal Ofek\textsuperscript{1}, Christian Holz\textsuperscript{1}, Hrvoje Benko\textsuperscript{1}, Andrew D. Wilson\textsuperscript{1}

\textsuperscript{1}Microsoft Research
Redmond, WA, USA
\{eyalofek, cholz, benko, awilson\}@microsoft.com

\textsuperscript{2}Hasso Plattner Institute
Potsdam, Germany
lung-pan.cheng@hpi.de

Figure 1. (a) Our hemispherical prop is an example of a Sparse Haptic Proxy. It simulates both, (b) a room and (c) a cockpit scene to provide physical touch feedback during interaction. (White lines on the prop added for visibility on the black background).

ABSTRACT

Over time, haptic feedback devices have become more ubiquitous and versatile, allowing users to experience the virtual environment and interact with it. However, these devices are often large and bulky, making them difficult to use in everyday life. We introduce a new haptic proxy, which is a general passive prop that can be attached to any object or surface to provide haptic feedback. Our prop is lightweight and compact, making it easy to use and integrate into any environment. We demonstrate its effectiveness by using it to simulate a room and a cockpit scene, providing physical touch feedback to users during interaction. Our results show that our Sparse Haptic Proxy can be used to enhance the realism of virtual environments and provide better interaction with them.
Smith et al., Hand Eye Coordination Patterns in Target Selection, ETRA 2000
SMI eye-tracking module

250 hz eye-tracker on Oculus DK2
We can predict with 97.5% accuracy what is the user’s intended target 2 seconds before reaching the target!
Body Warping

The Rendered Body Shifts to The Right
World Warping

The World Also Rotates
(At Different Rate)
IllumiRoom

Peripheral Projected Illusions for Interactive Experiences.
What is “real” vs. what we “perceive as real” is not necessarily the same!
Faces or vase?
Straight or crooked?
Moving or static?
Insisting on “as real as possible” designs can lead to sub-optimal MR experience.

Think about MR interfaces as perceptual illusions that give the user a believable experience!
Typical HCI model

user

system
How the computer sees us!

How the phone sees us!

MR interaction model

user

world

MR system
MR interaction model

user

world

MR system

user/world model

input understanding

application logic

output generation