# Egocentric Perception, Interaction, Computing and Display

Marc Pollefeys | November 2, 2019

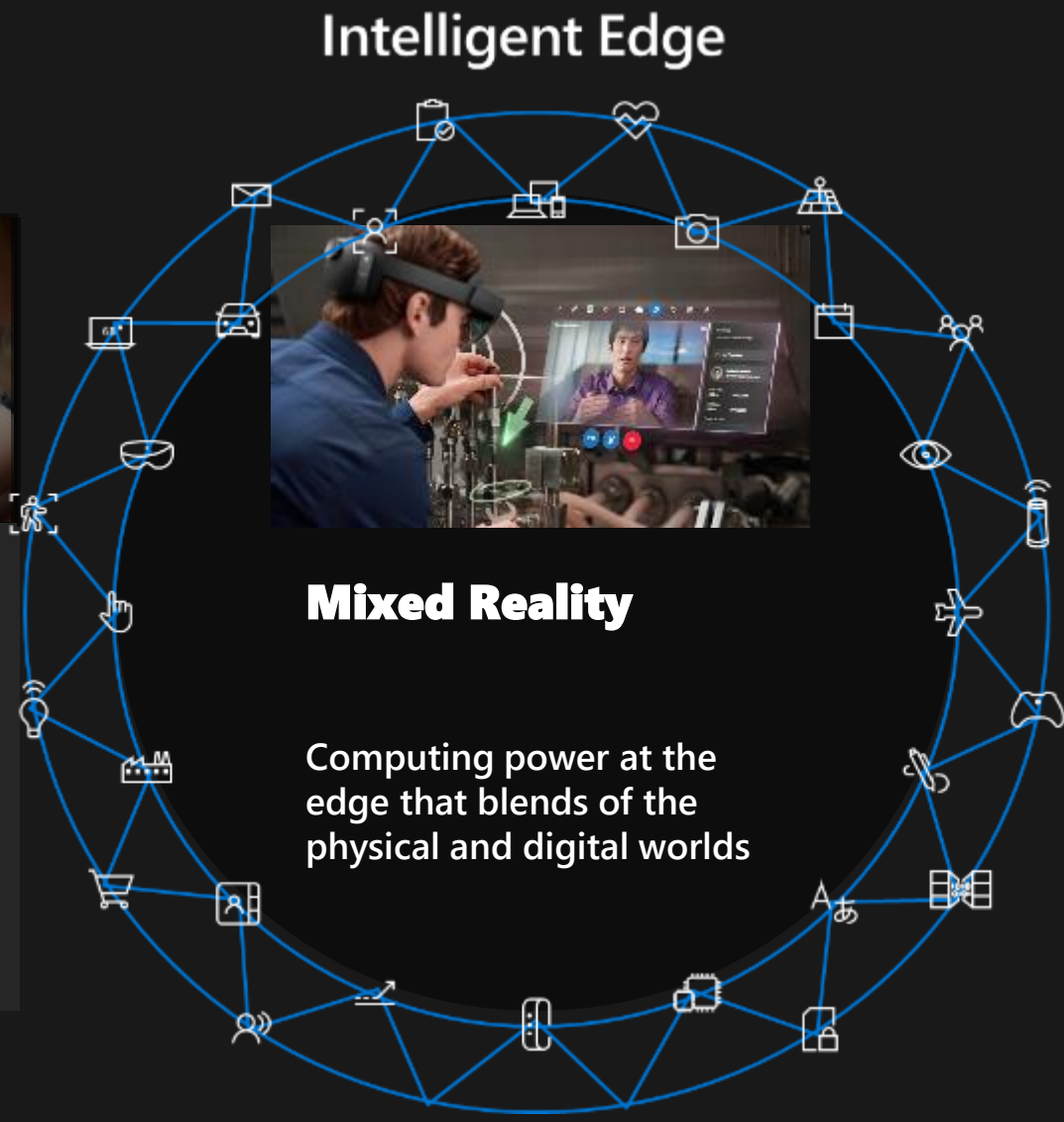
# THIRD WAVE OF COMPUTING

## Personal Computers

Computing power for many but immobile and no contextual awareness

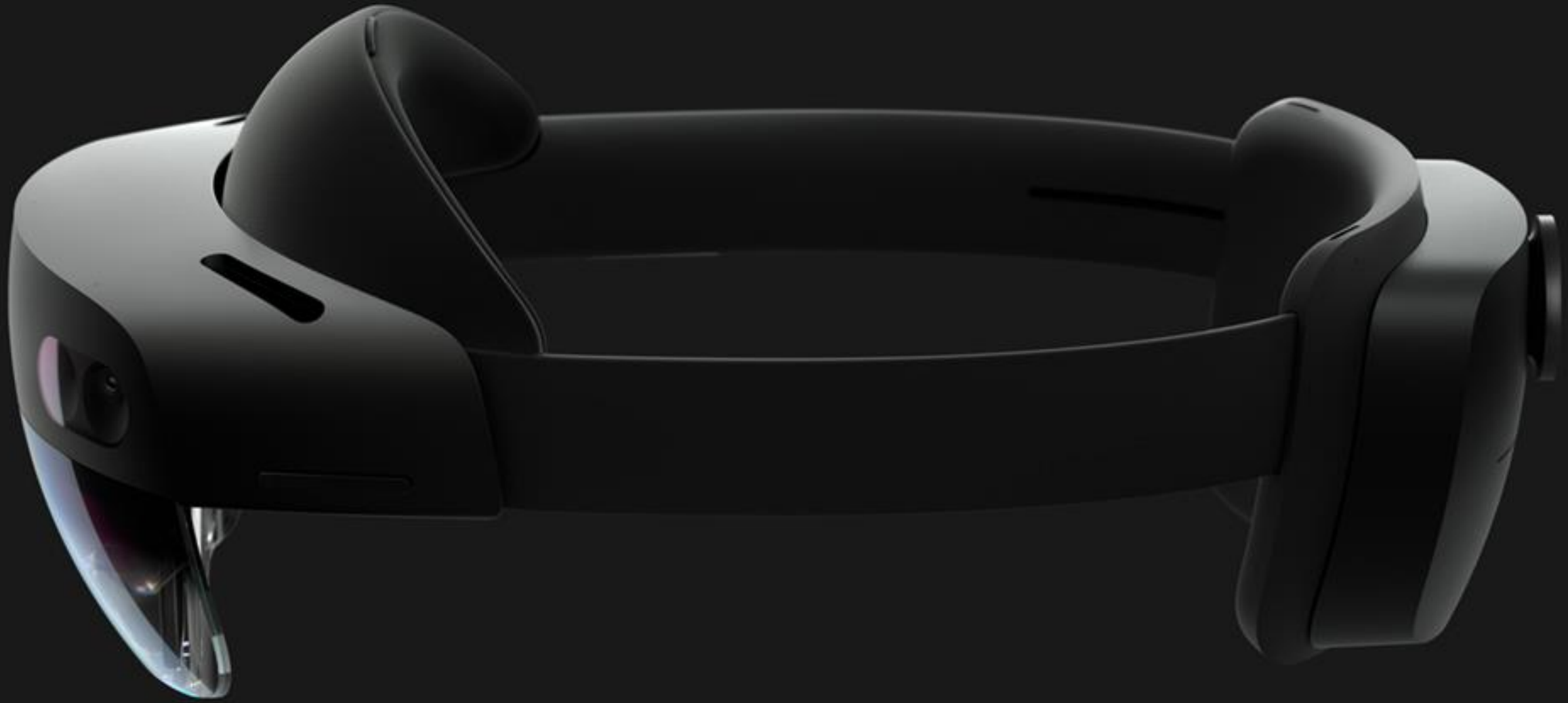
## Smartphones

Computing power for all with on-the-go access and some contextual awareness



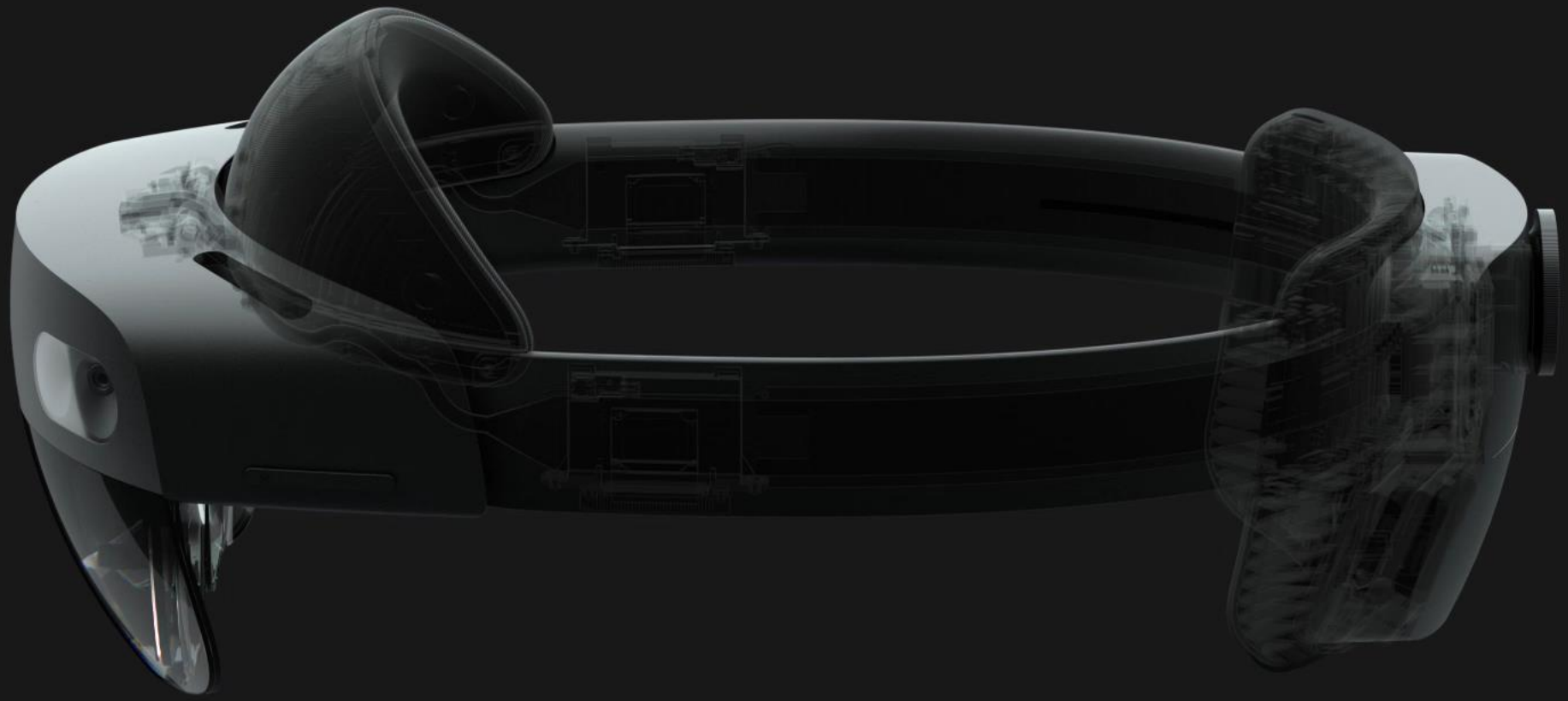

## Mixed Reality

**Computing power at the edge that blends of the physical and digital worlds**




FRONT WHEEL
SYSTEM
GEAR
SYSTEM
FRAME
SYSTEM

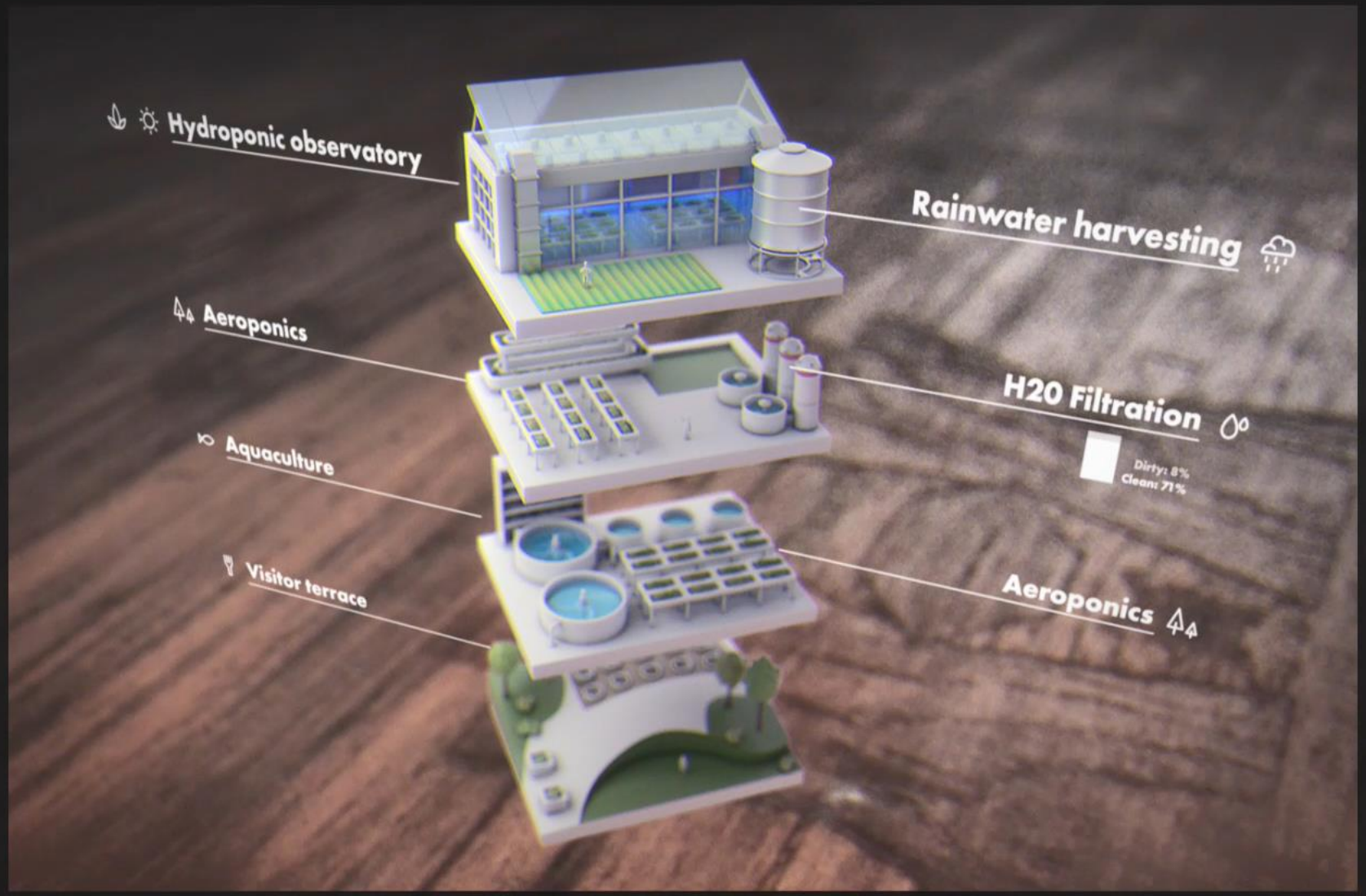
# UNIVERSAL FIT



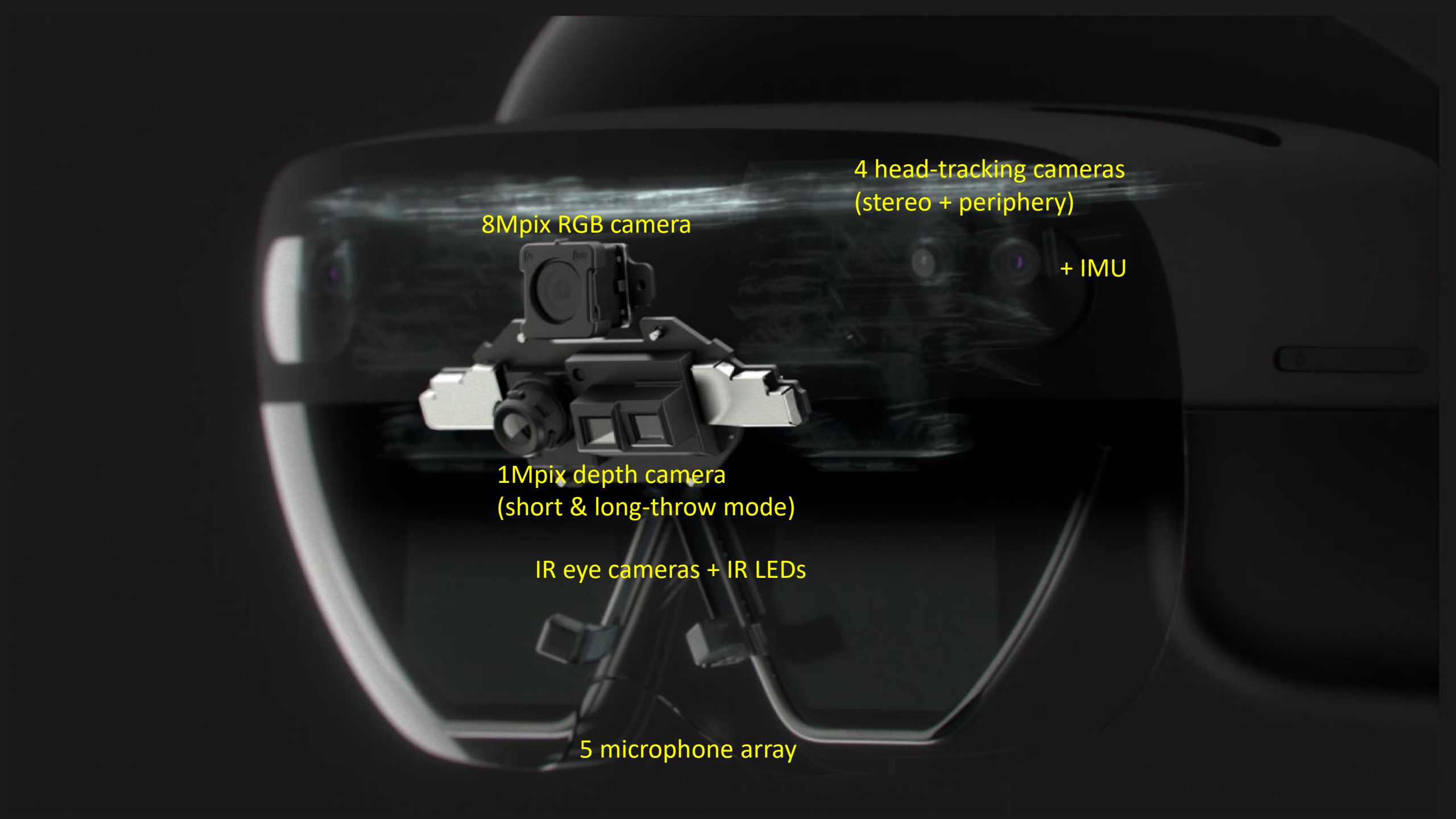

PIVOTING VISOR

# LARGER FIELD OF VIEW

HoloLens 2

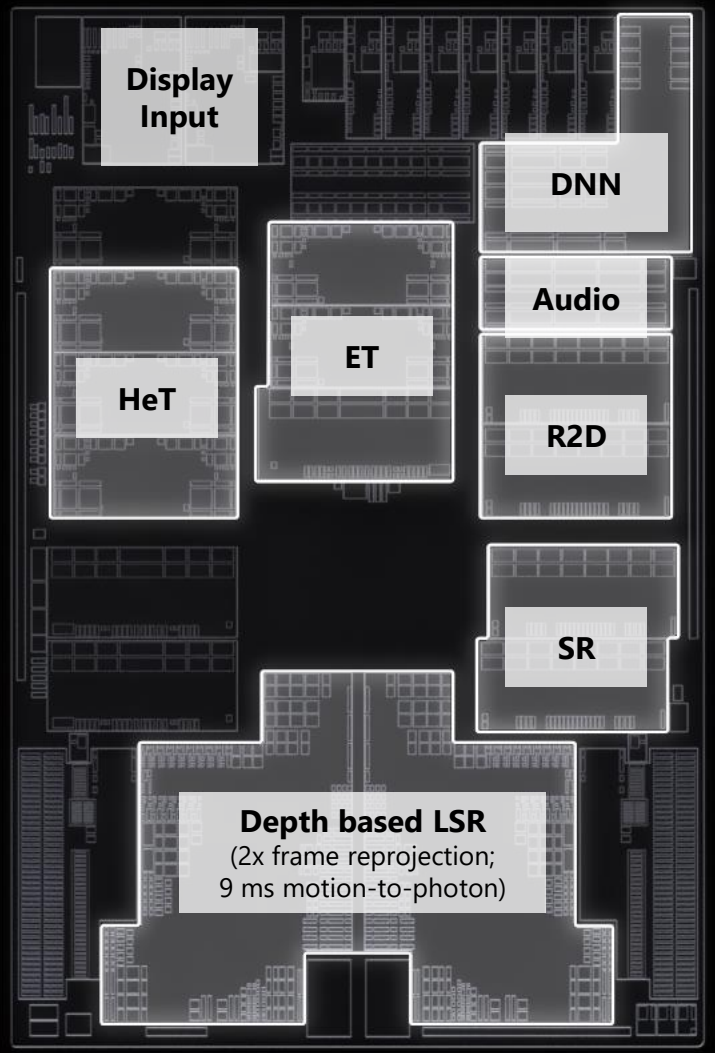


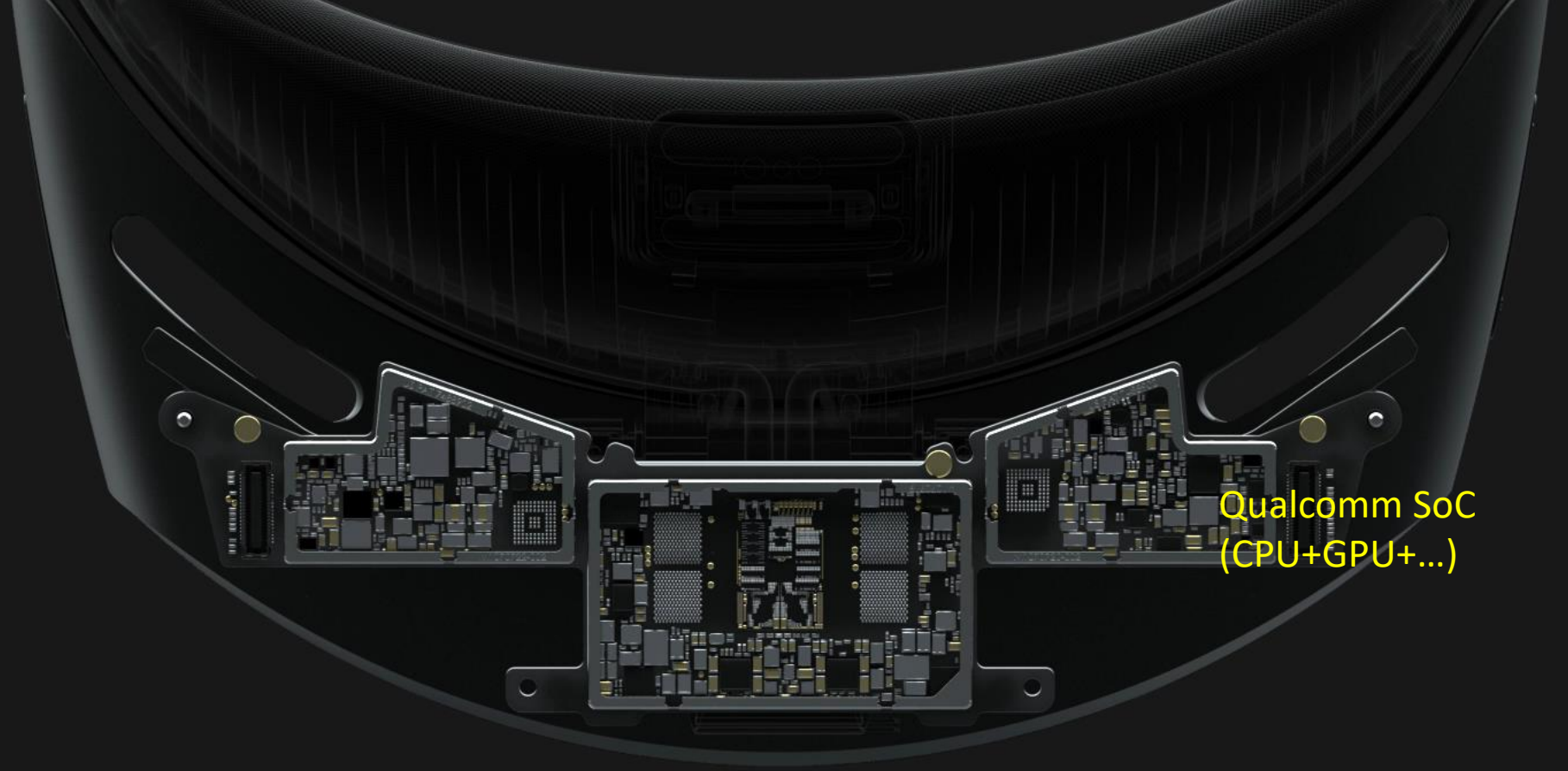
## DISPLAY


4 head-tracking cameras
(stereo + periphery)
8Mpix RGB camera
+ IMU
1Mpix depth camera
(short & long-throw mode)
IR eye cameras + IR LEDs
5 microphone array


HPU (DSPs, DNN AI core, LSR)

# HPU


Display Input
DNN
Audio
ET
HeT
R2D
SR
Depth based LSR
(2x frame reprojection;
9 ms motion-to-photon)

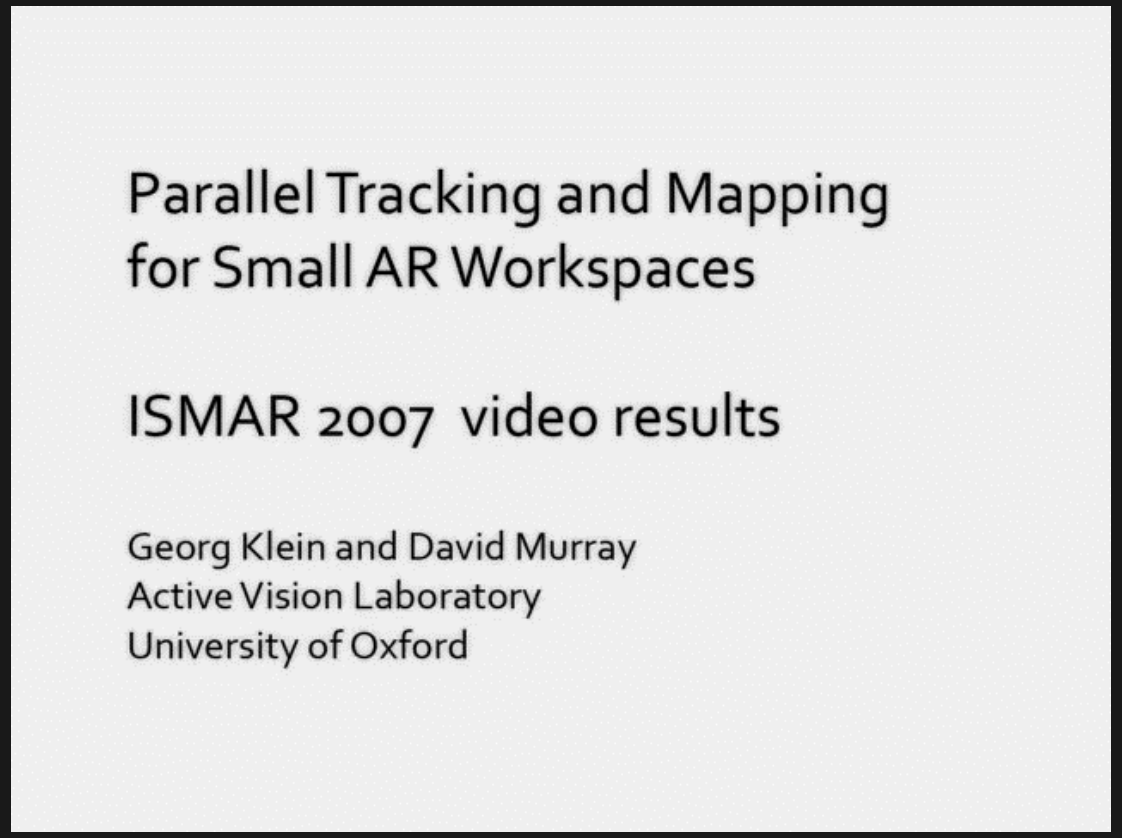

Qualcomm SoC
(CPU+GPU+…)

# Where are you?

- Head tracking
  - Location in the world
  - Head orientation

# Parallel Tracking and Mapping

(Klein and Murray 2007)



# HoloLens Head tracking

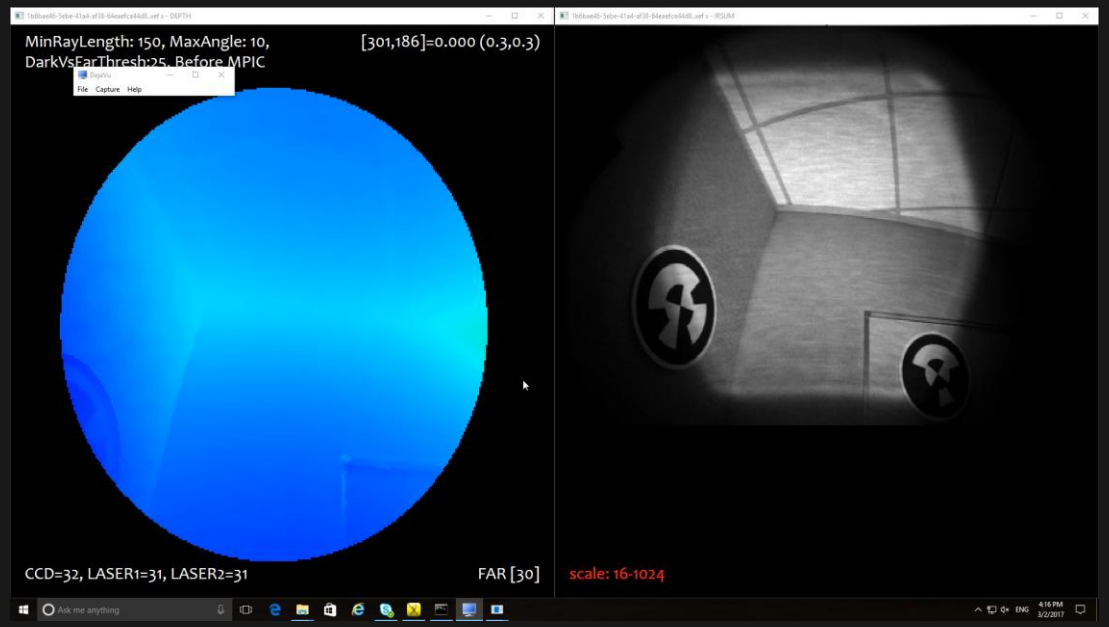
highly accurate visual-inertial odometry

4 cameras + IMU
highly optimized for power

# What's around you

- Depth perception
  - TOF depth camera
  - Mesh generation

# Depth sensing

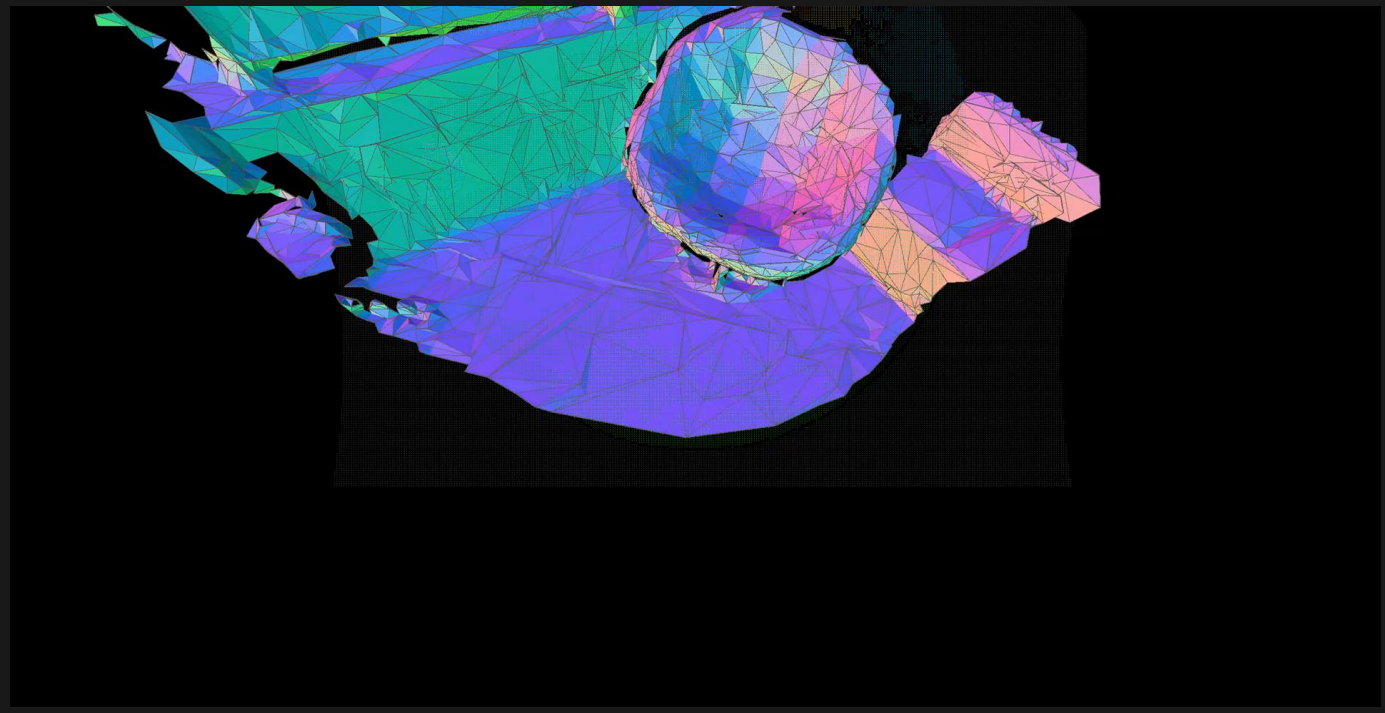



Depth + (near) IR image

# KinectFusion

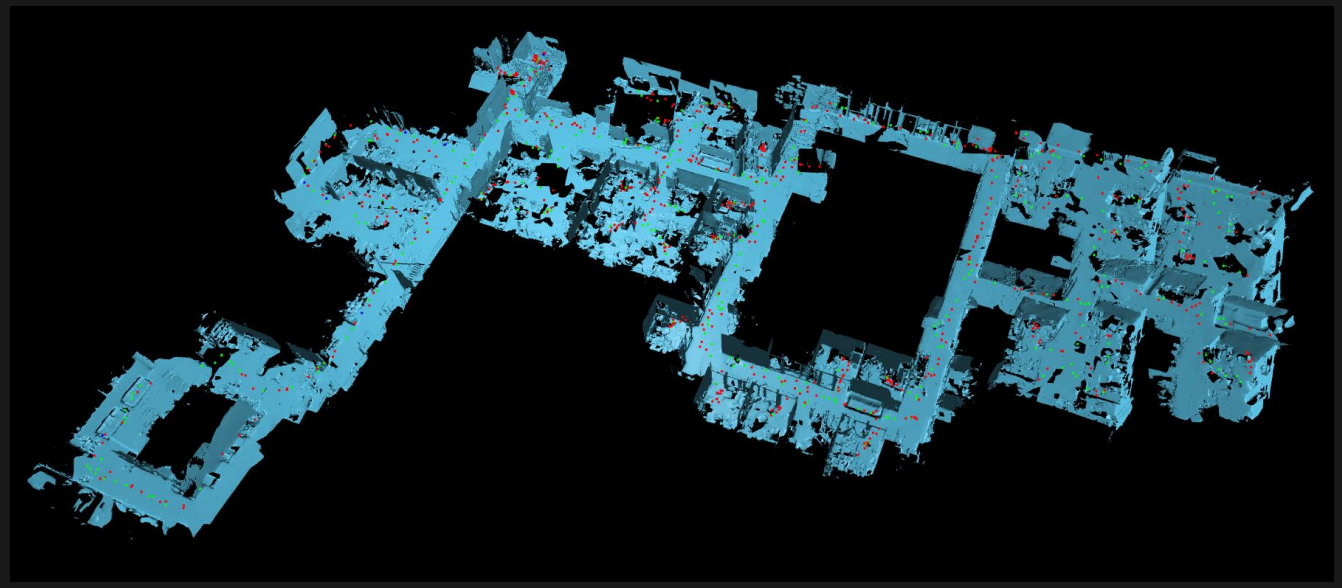
(Newcombe et al ISMAR11, Izadi et al UIST11)
(Curless and Levoy SIGGRAPH96)

# Surface reconstruction



Fusion of depth measurements and mesh extraction

# Spatial Mapping



# What is around you?

Semantic segmentation demo

# What are you doing?

- Gestures/Hand tracking
- Gaze direction
- Speech

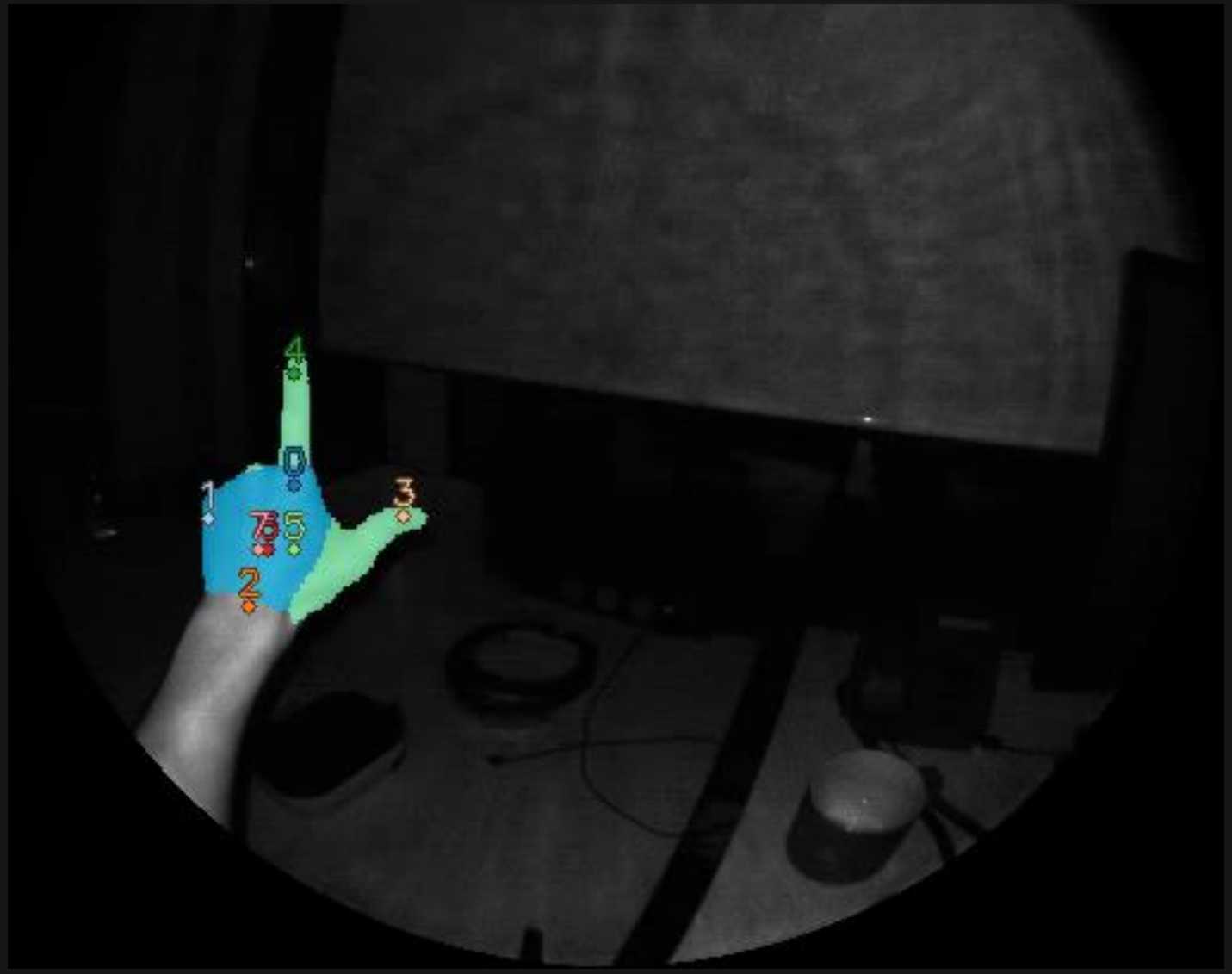
# Efficient & Precise Interactive Hand Tracking through Joint, Continuous Optimization of Pose & Correspondences

SIGGRAPH 2016

Jonathan Taylor, Lucas Bordeaux, Thomas Cashman, Bob Corish, Cem Keskin, Toby Sharp,
Eduardo Soto, David Sweeney, Julien Valentin,Benjamin Luff, Arran Topalian, Erroll Wood,
Sameh Khamis, Pushmeet Kohli, Shahram Izadi, Richard Banks, Andrew Fitzgibbon, Jamie Shotton

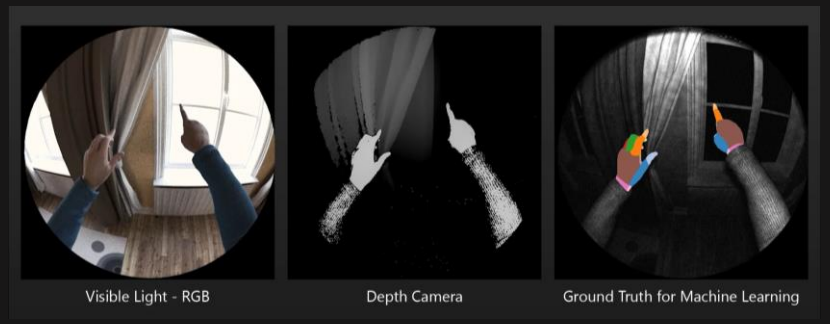
Microsoft Research

# HoloLens2 Hand-tracking

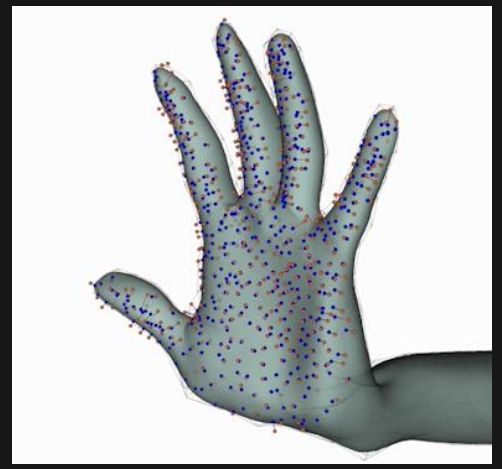


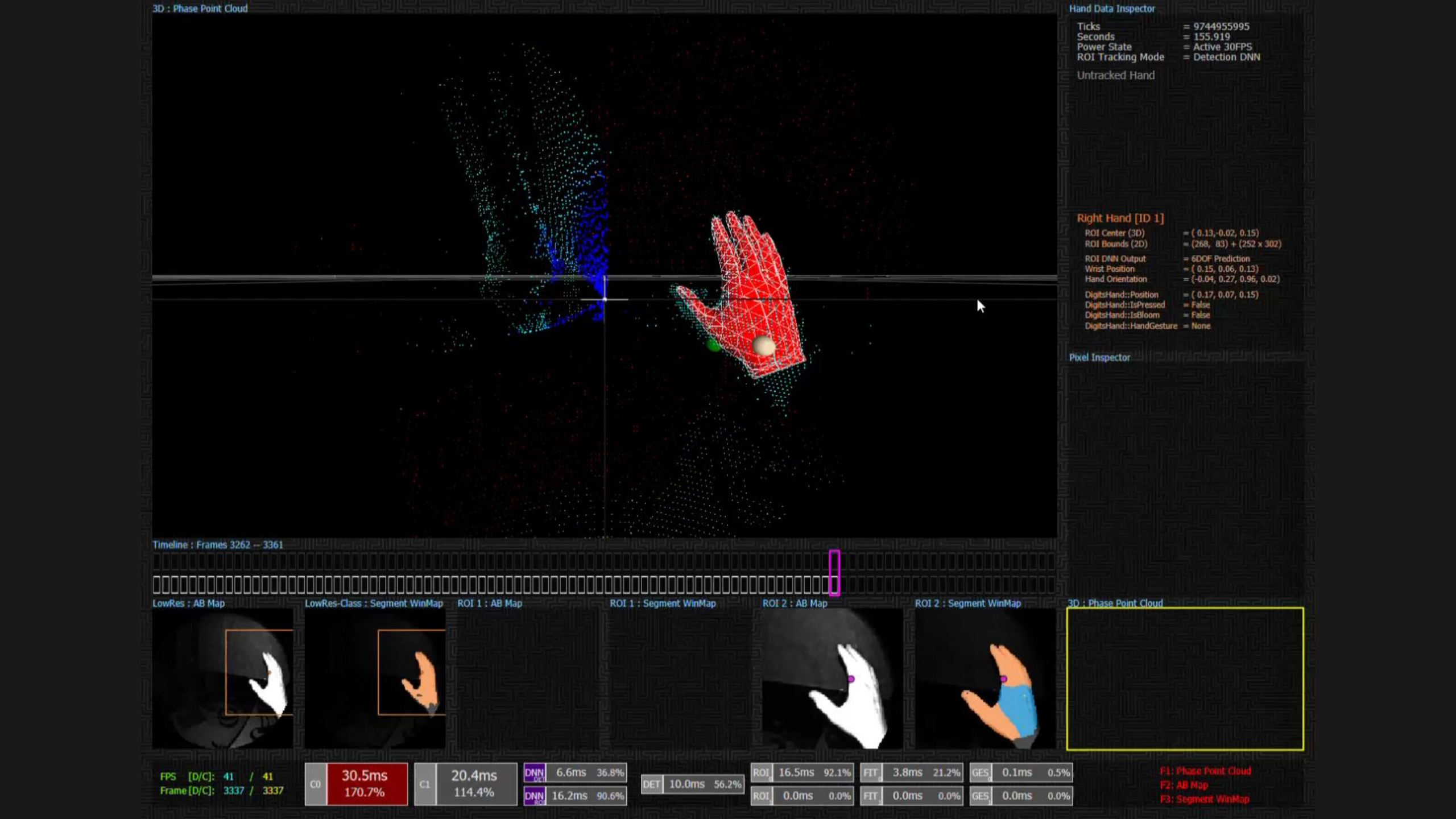

Hand-segmentation tuns on HoloLens DNN accelerator

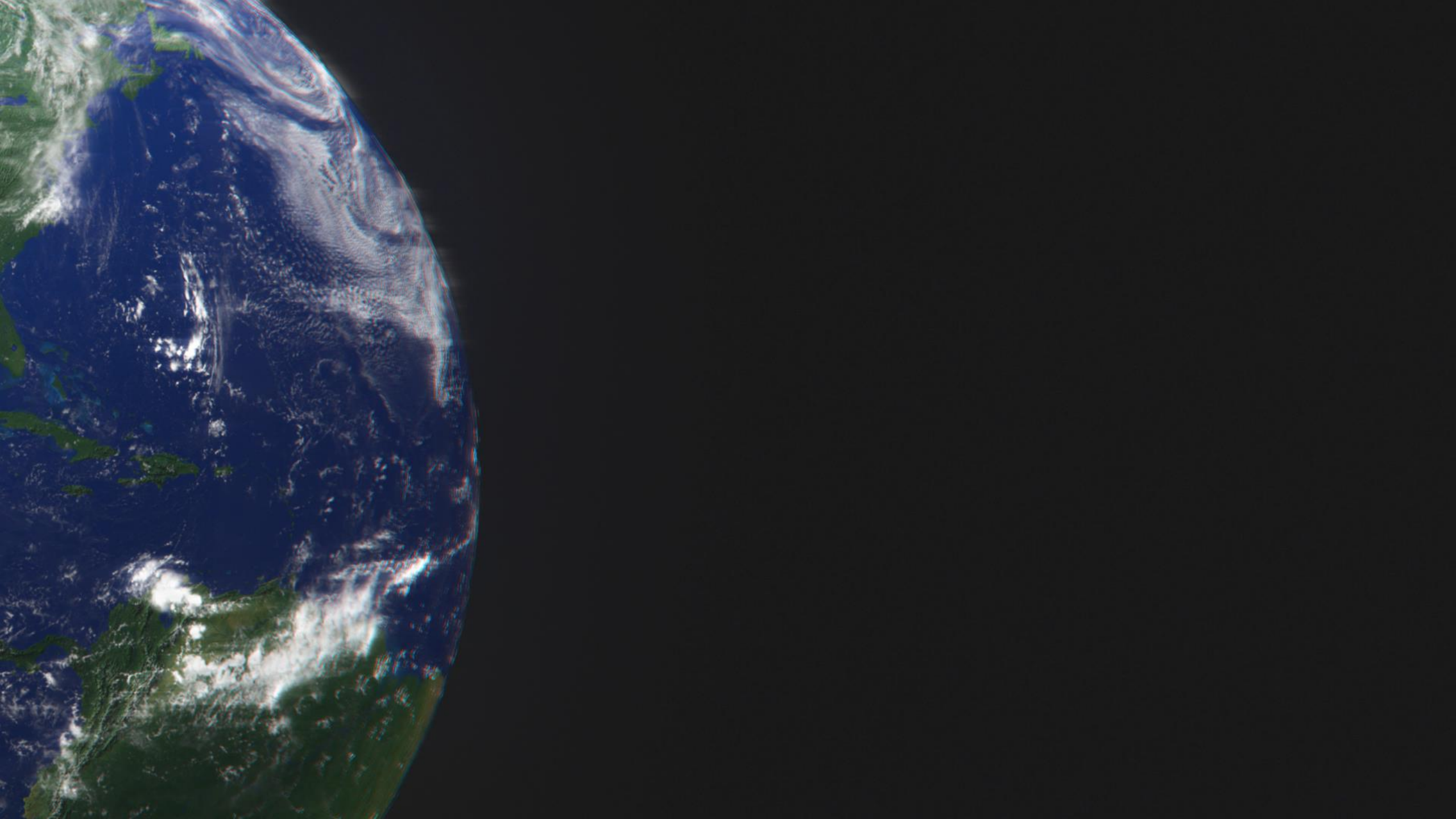
Trained on purely synthetic data




Efficient geometric fitting




3D : Phase Point Cloud
Hand Data Inspector
Ticks = 9744955995
Seconds = 155.919
Power State = Active 30FPS
ROI Tracking Mode = Detection DNN
Untracked Hand
Right Hand [ID 1]
ROI Center (3D) = ( 0.13,-0.02, 0.15)
ROI Bounds (2D) = (268, 83) + (252 x 302)
ROI DNN Output = 6DOF Prediction
Wrist Position = ( 0.15, 0.06, 0.13)
Hand Orientation = (-0.04, 0.27, 0.96, 0.02)
DigitsHand::Position = ( 0.17, 0.07, 0.15)
DigitsHand::IsPressed = False
DigitsHand::IsBloom = False
DigitsHand::HandGesture = None
Pixel Inspector
Timeline : Frames 3262 -- 3361
LowRes : AB Map
LowRes-Class : Segment WinMap
ROI 1 : AB Map
ROI 1 : Segment WinMap
ROI 2 : AB Map
ROI 2 : Segment WinMap
3D : Phase Point Cloud
FPS [D/C]: 41 / 41
Frame [D/C]: 3337 / 3337
C0 30.5ms 170.7%
C1 20.4ms 114.4%
DNN DET 6.6ms 36.8%
DNN ROI 16.2ms 90.6%
DET 10.0ms 56.2%
ROI 16.5ms 92.1%
ROI 0.0ms 0.0%
FIT 3.8ms 21.2%
FIT 0.0ms 0.0%
GES 0.1ms 0.5%
GES 0.0ms 0.0%
F1: Phase Point Cloud
F2: AB Map
F3: Segment WinMap

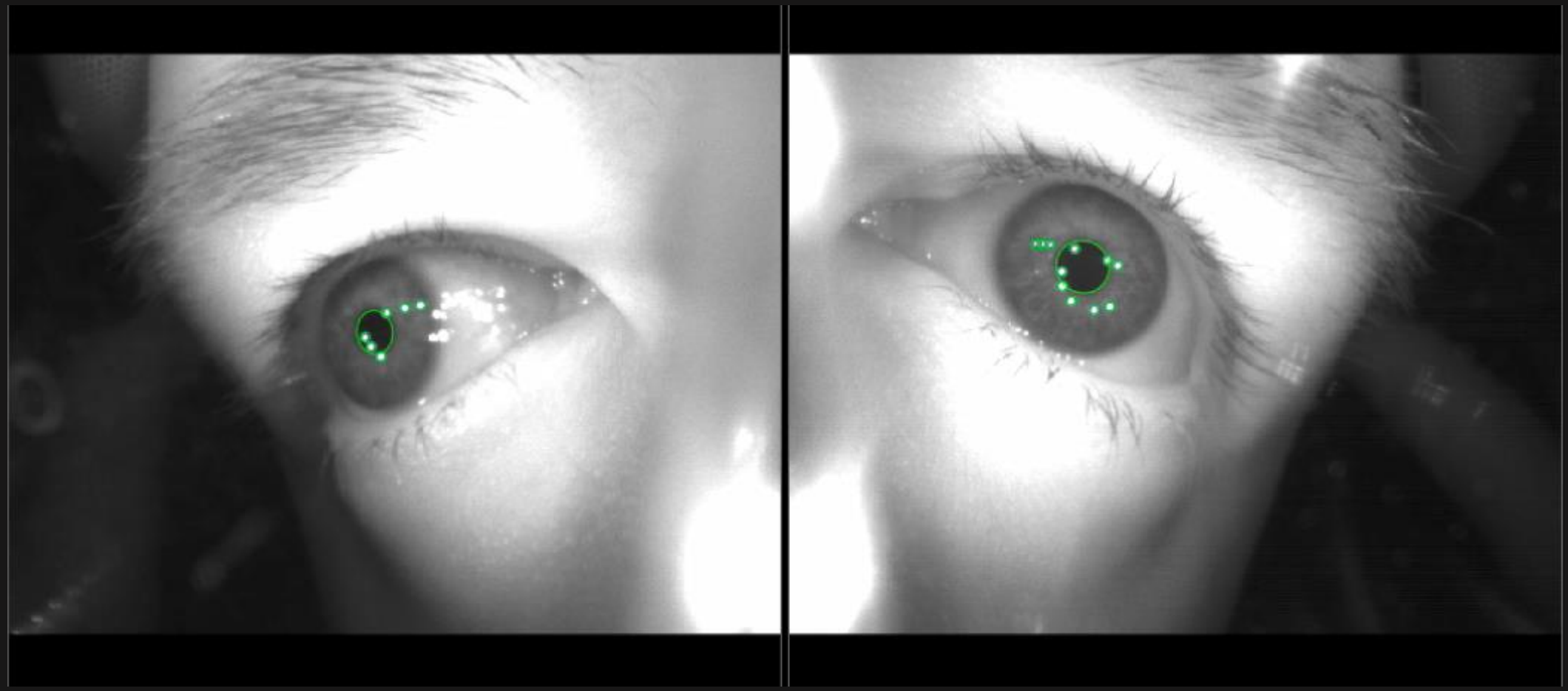



Eye Tracking


Windows Hello

# Eye Tracking

2 IR cameras + IR LEDs



Microsoft Dynamics 365

# Remote Assist

Technicians solve problems in real-time with the help of remote experts

Managers walk the job site without being on site

Bring information into view


Recent Contacts
Search
Beth McDowel
Susan Tester
Daniel Thrash
Wayne Sabo
Lila Hofman
Asumi
John Manning
Dominic Greer
Booking
Finish Dynamics Implementation
Scheduled
Asumi
asumi@contoso.com
Start Time
3:00 PM
18/07/2018
End Time
5:30 PM
18/07/2018
Power Bi Link

Microsoft Dynamics 365

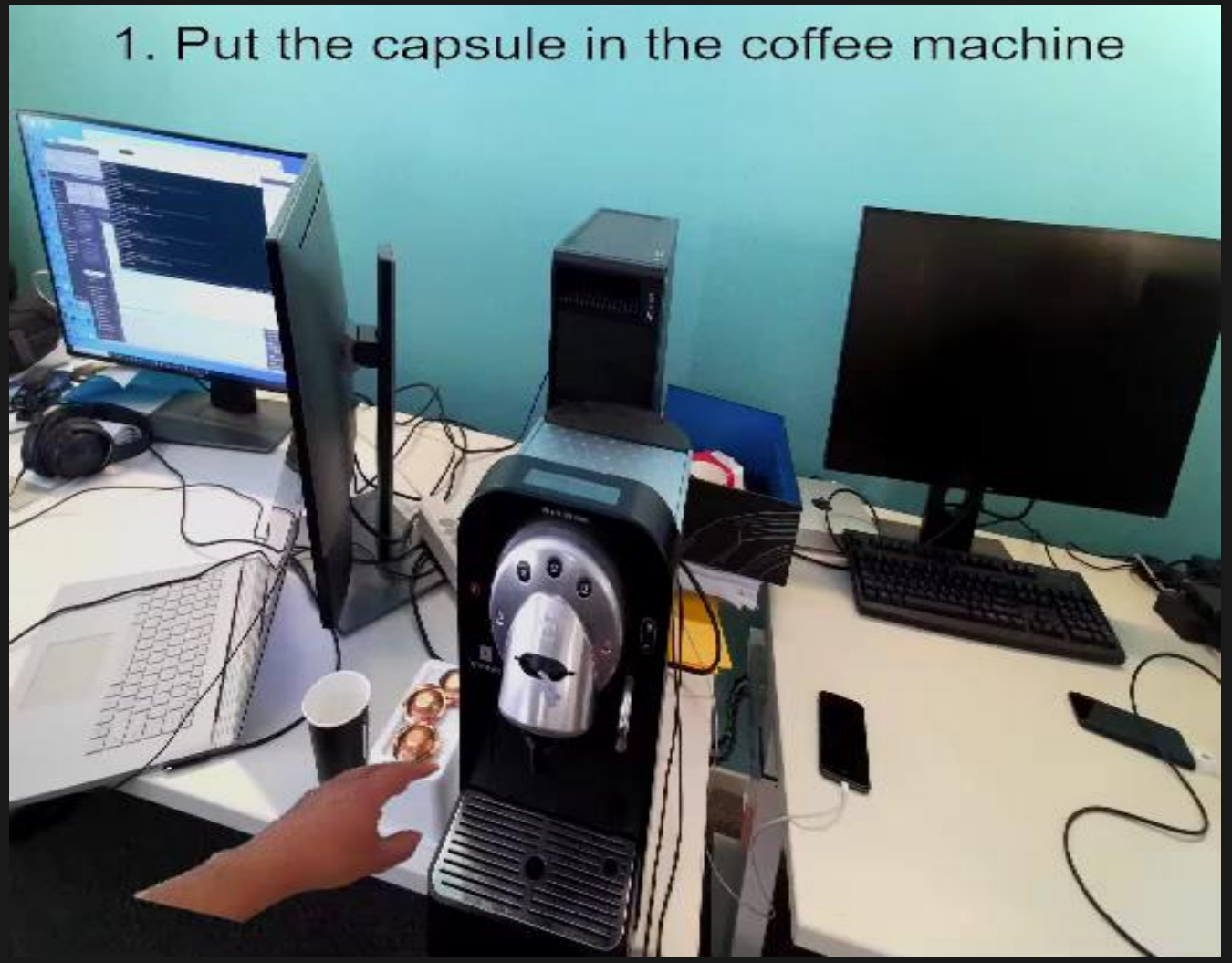
# Guides

Engage employees with hands-on learning

Generate data to improve process

Improve training effectiveness


Fuel System
Regulator Line Installation
Use Torque wrench to tighten to specification.
The gauge should read 74 ft – lb.
Step 5 of 12

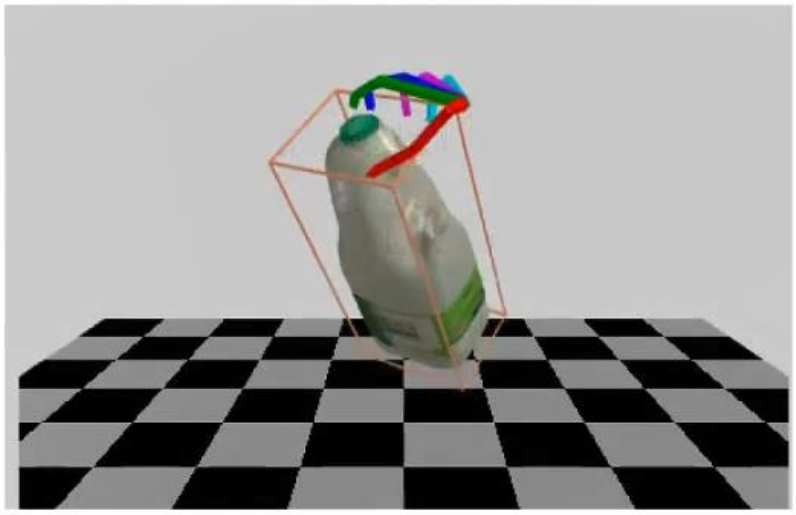

1. Put the capsule in the coffee machine

# Hands + Objects : Unified Egocentric Recognition of 3D Hands+Object Poses and Interactions

*Bugra, Bogo & Pollefeys, CVPR 2019*

2D Hand + Object Pose

3D Hand + Object Pose



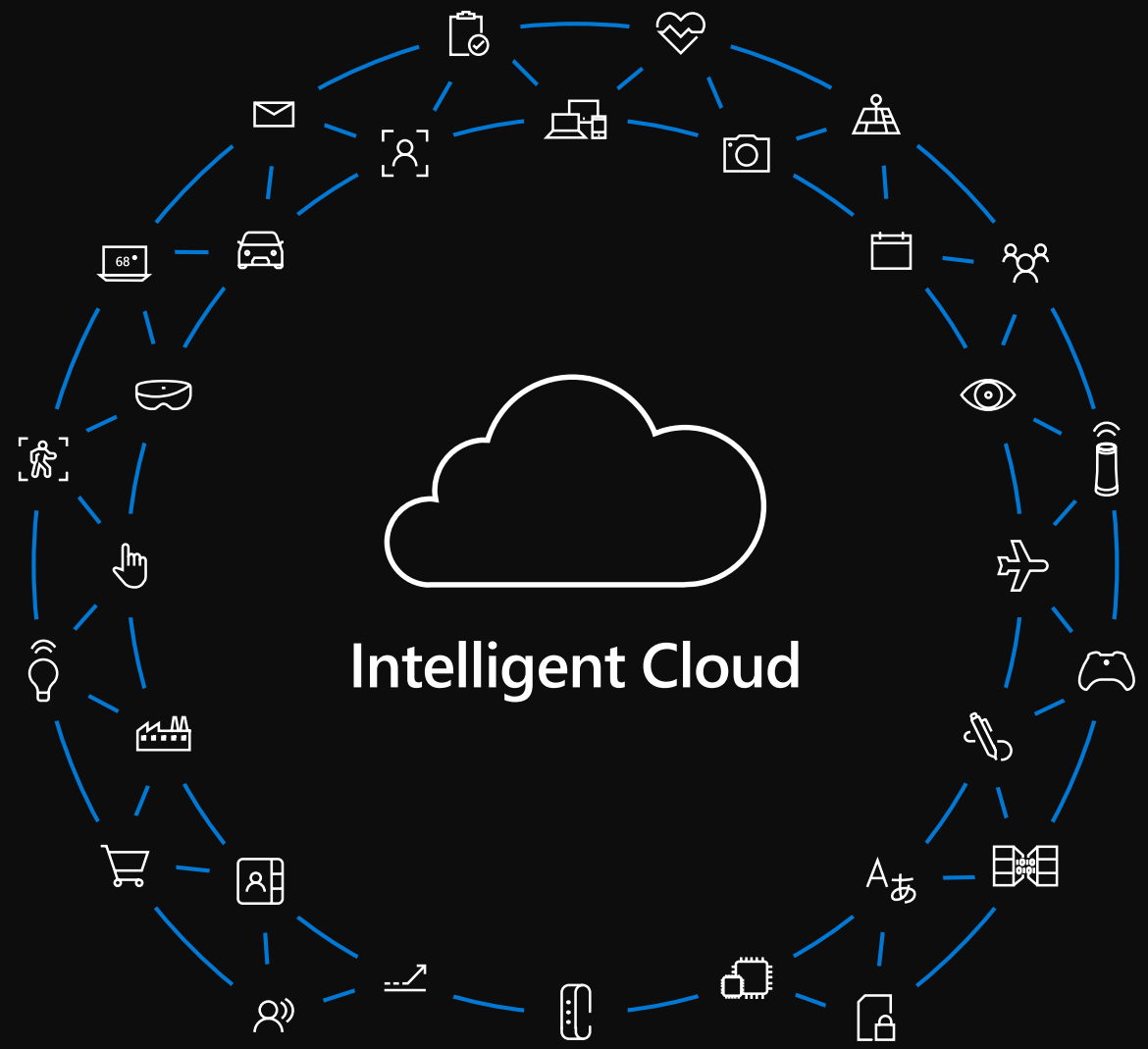
Per-frame predictions


BUILDING PANEL

Ubiquitous computing

Artificial Intelligence

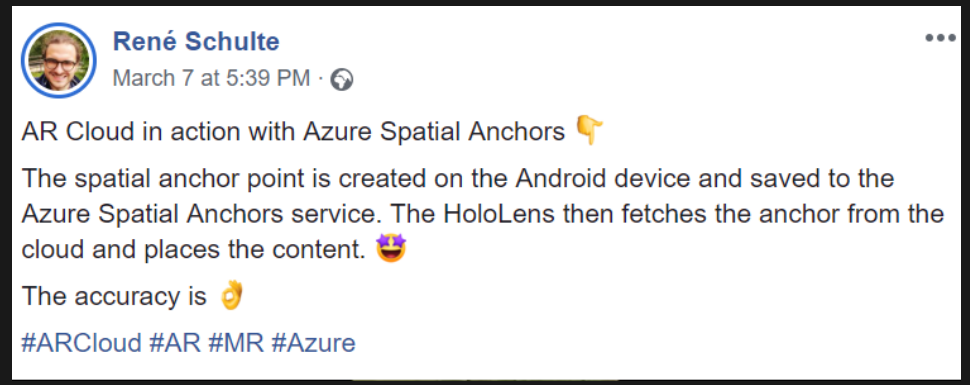
Multi-sense, multi-device experiences

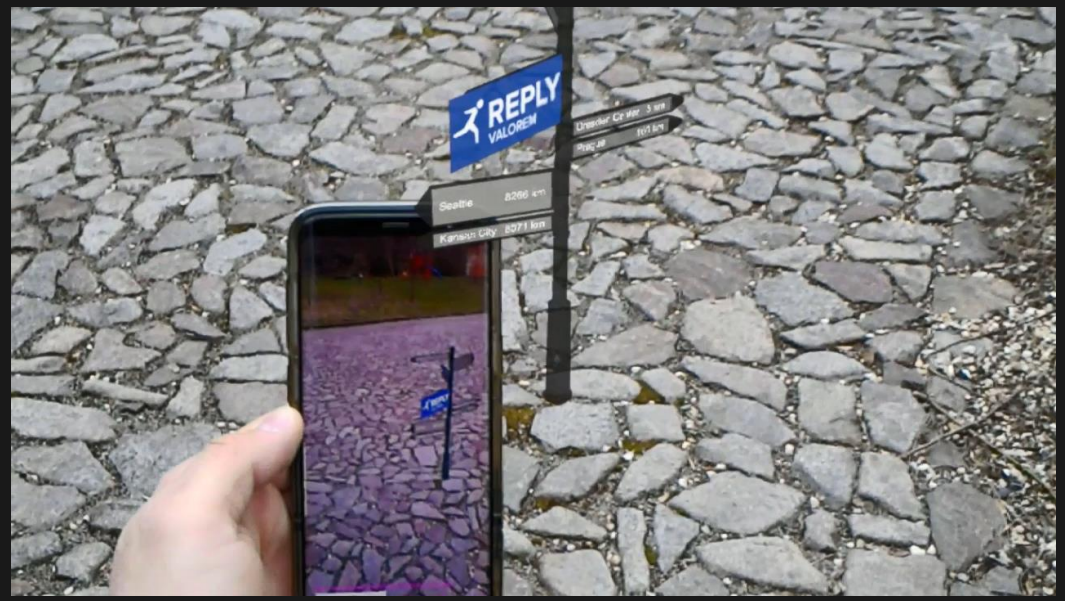


# AZURE SPATIAL ANCHORS

Enhance collaboration and understanding with tools for cross-platform, spatially aware mixed reality experiences across HoloLens, iOS, and Android devices.

# Azure Spatial Anchors in action

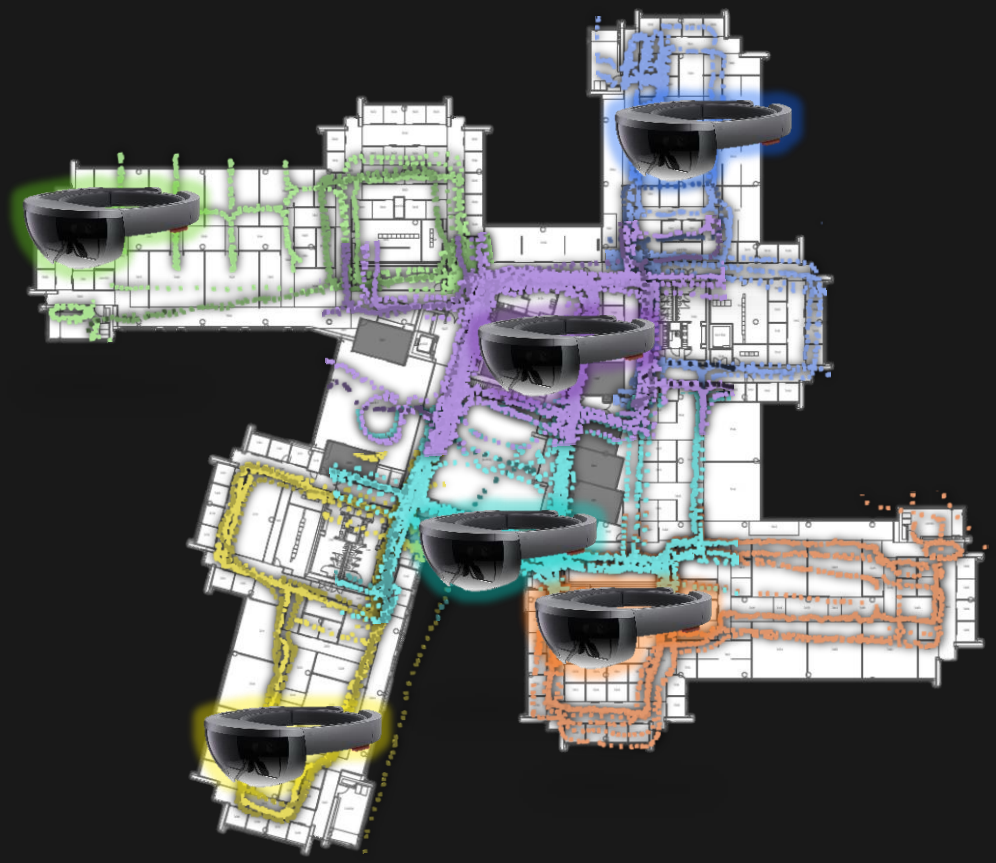




# 6DOF relocalization map

Clients

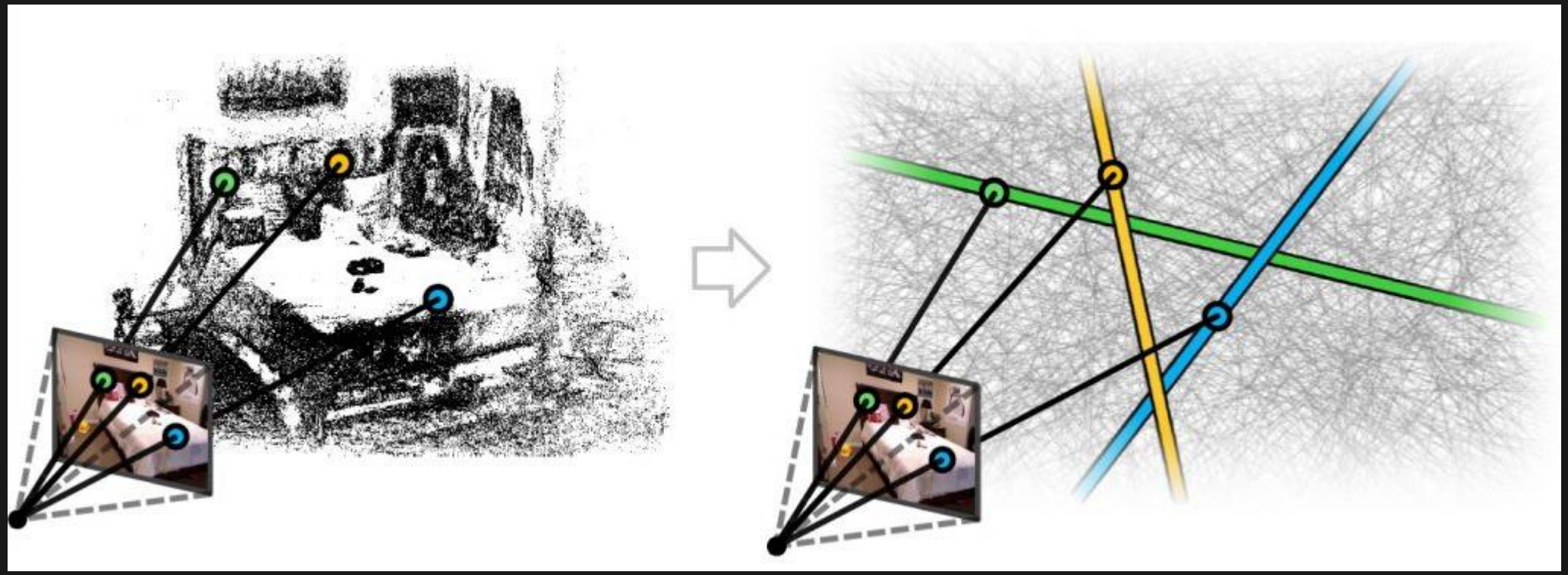


Cloud

combine multiple maps in the cloud

# Privacy-Preserving Image-based Localization

[Speciale et al. CVPR 2019]

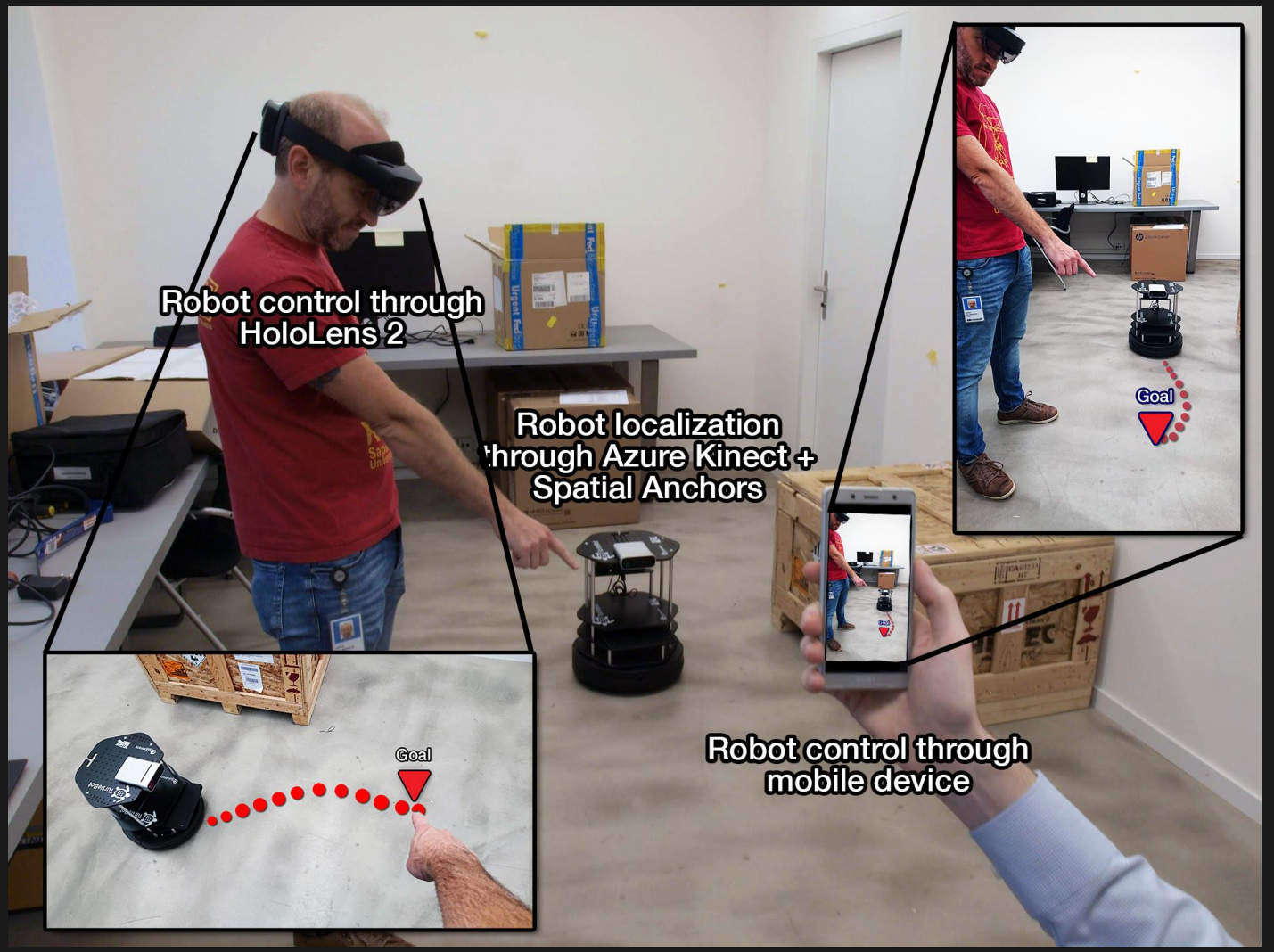


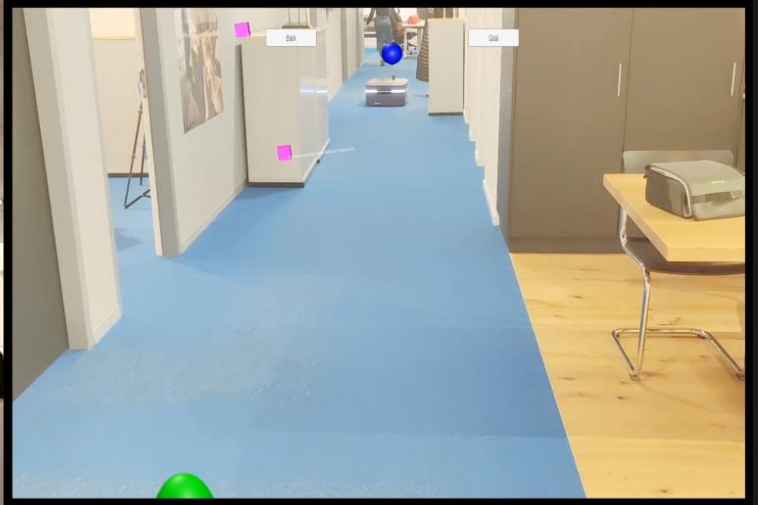
3D Point Cloud
(Traditional)

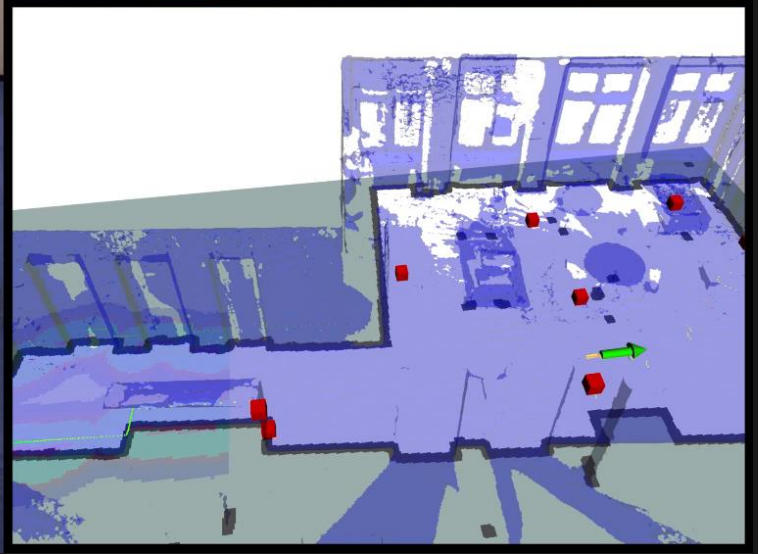
3D Line Cloud Map
(Proposed)

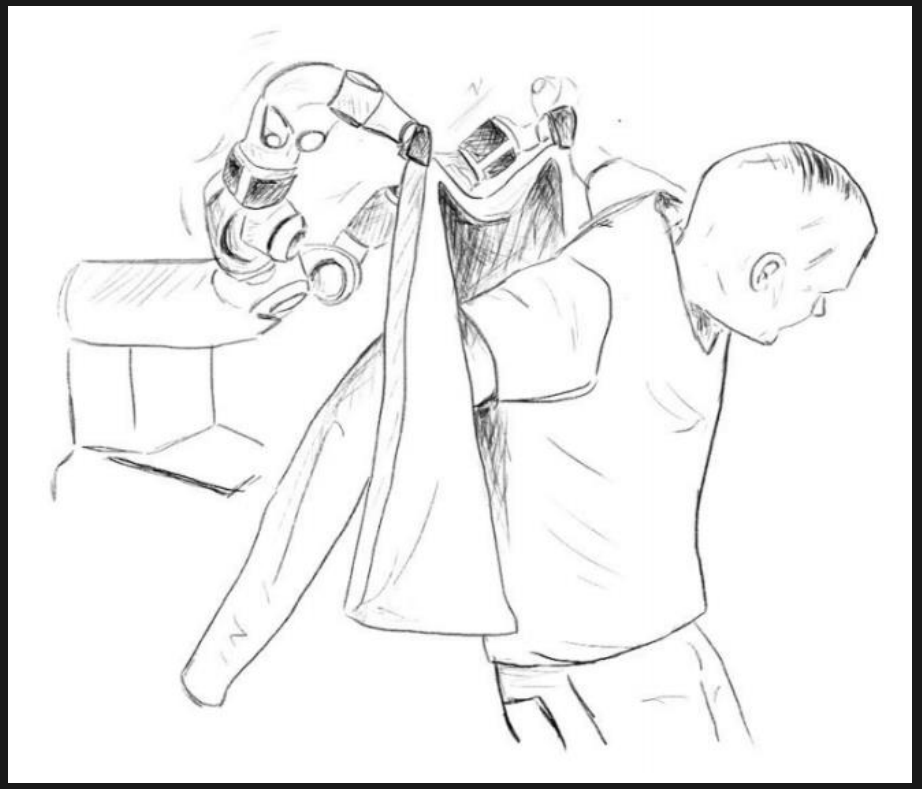
# Privacy preserving image-based localization queries

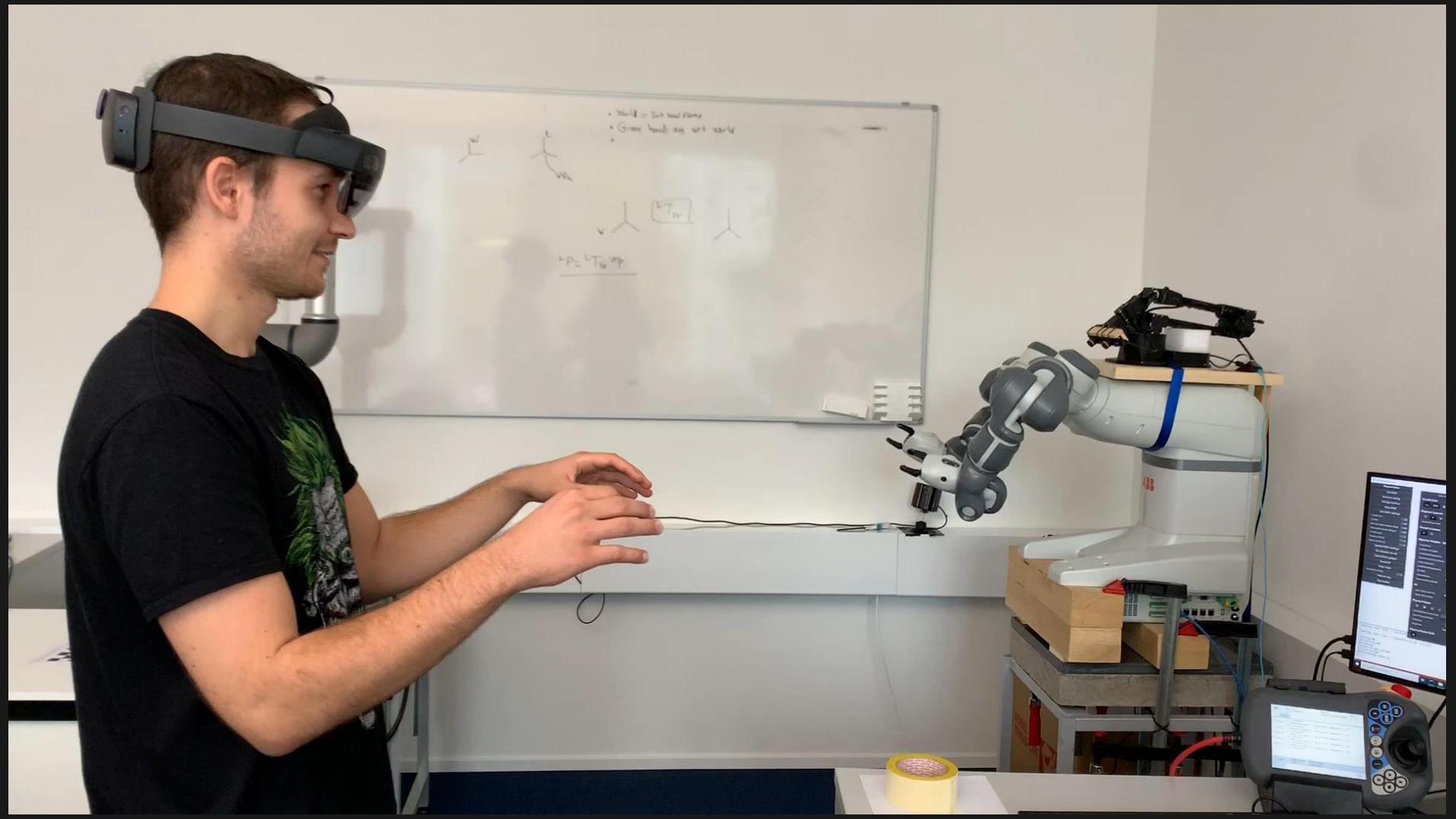
*Speciale, Schönberger , Sinha, Pollefeys, ICCV 2019*


Robot control through
HoloLens 2
Robot localization
through Azure Kinect +
Spatial Anchors
Goal
Goal
Robot control through
mobile device









# Environment data generally available to apps

IMU

HeT cameras

Raw Depth

Head pose, map

Mesh

Hands, gestures

Head pose, anchors

Photo/Video Camera

# HoloLens Research Mode

IMU

Head pose, map

Head pose, anchors

HeT cameras

Mesh

Raw Depth

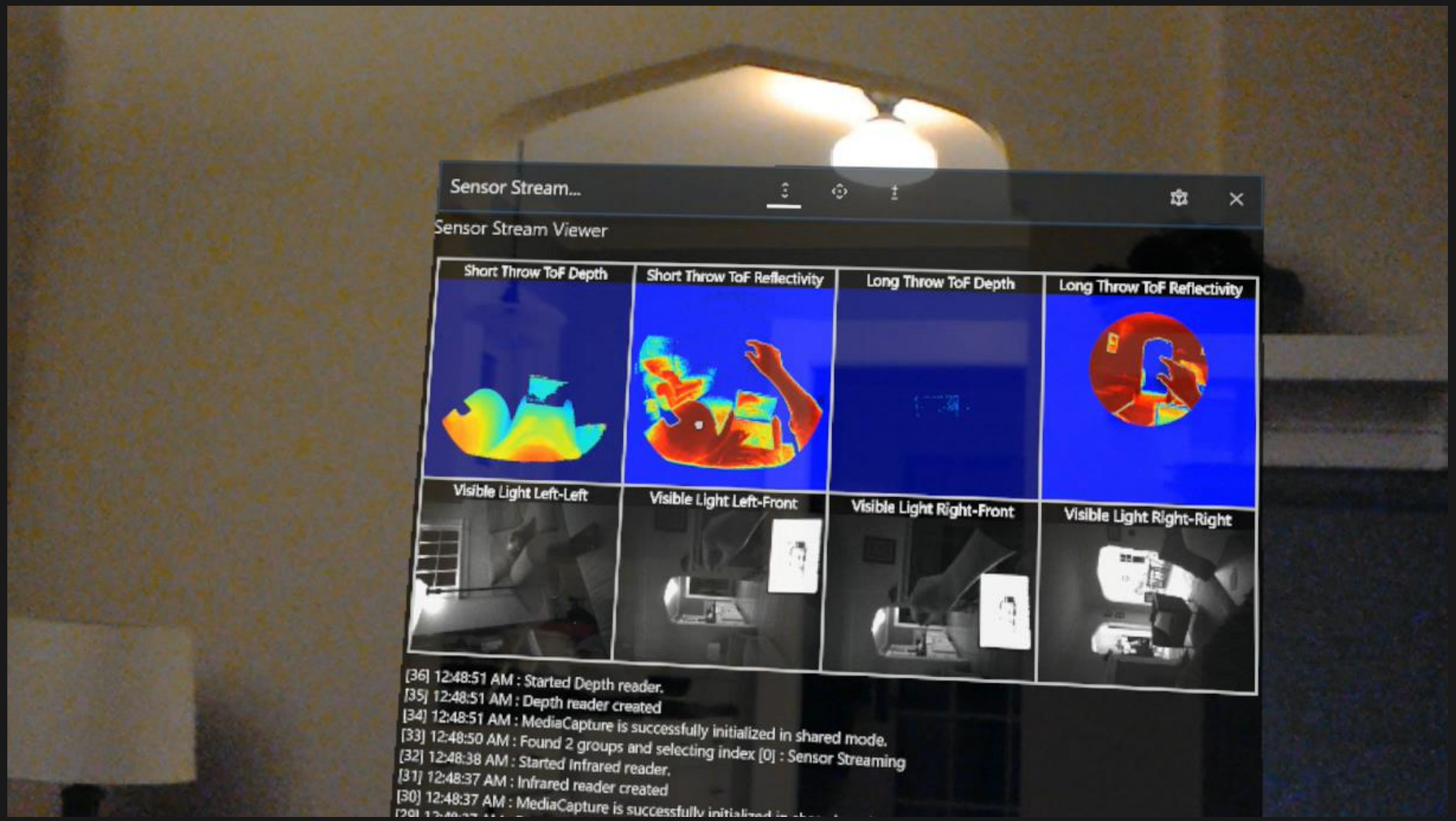
Hands, gestures

Short- and Long-Throw Depth

IR Reflectance

Photo/Video Camera

# Research mode Sensor Stream test app




Check out https://github.com/Microsoft/HoloLensForCV

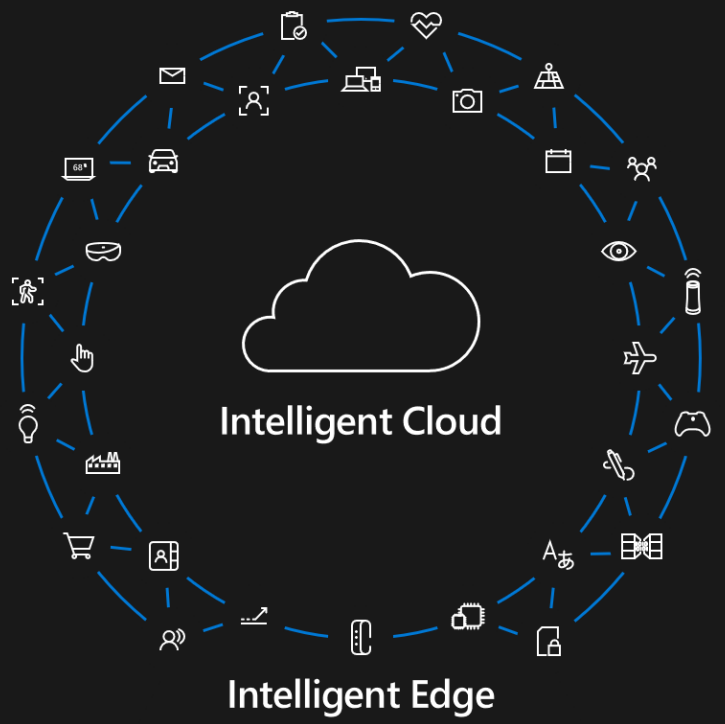
# Person identification and tracking

# Mixed Reality: Egocentric Perception, Interaction, Computing and Display

- Mixed Reality headsets have potential to have much more user/task context
  - Observe user actions
  - Understand environment
  - Access relevant digital information
  - Natural user interface

Microsoft

# Opportunities & open issues




- Edge versus cloud computing
- Tele-presence
  - Project remote presence of user
  - Allow user to access remote places
- Synergies between HMDs, ambient AI, robotics and IoT
- Privacy

Microsoft