



# Spatial Computing and the Industrial Metaverse

Marc Pollefeys – ETH Zurich & Microsoft







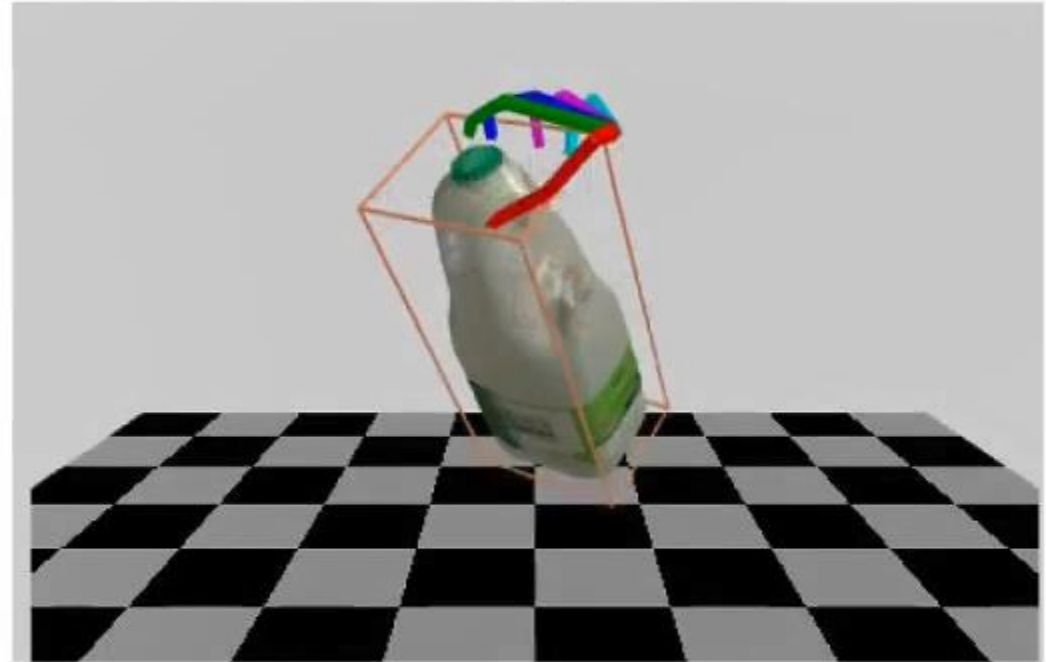
Fuel System  
Regulator Line Install

# First-person activity recognition

2D Hand + Object Pose



3D Hand + Object Pose



Per-frame predictions

# Action Recognition for Automated Task Guidance

- not completed
- ongoing
- completed
- idle
- missed



- Detect
  - actions & idle states
  - when we complete an action, when we move on to the next step
  - missed actions, actions not performed in order
  - the duration of the actions



# Action Recognition for Automated Task Guidance

- not completed
- ongoing
- completed
- idle
- missed

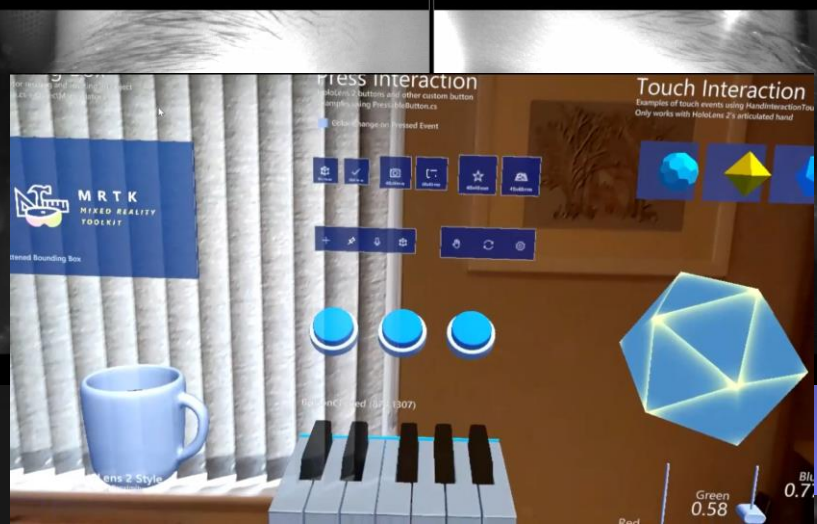
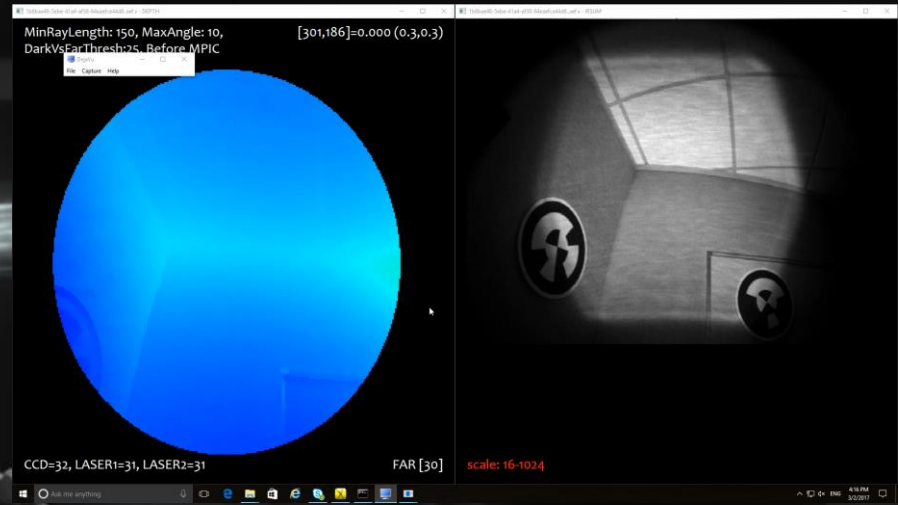
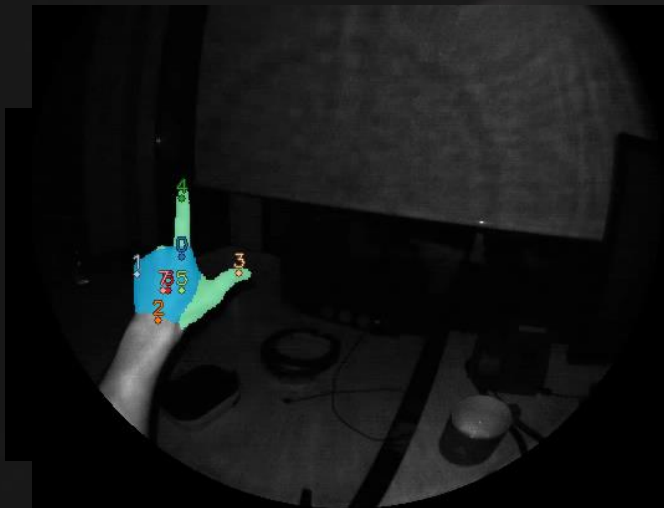


- Detect
  - actions & idle states
  - when we complete an action, when we move on to the next step
  - missed actions, actions not performed in order
  - the duration of the actions

8Mpix RGB camera  
(video communication)

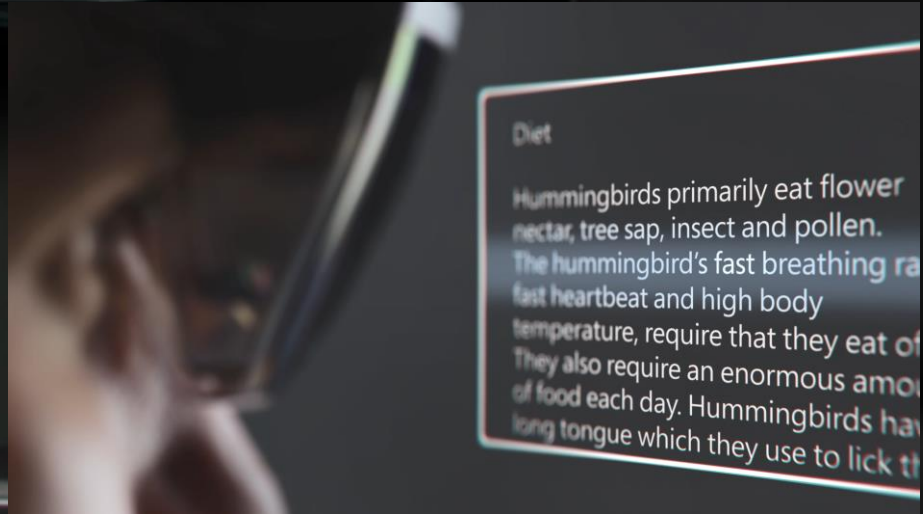
4 head-tracking cameras + IMU  
(stereo + periphery)





1Mpix depth camera  
(long- & short-throw mode)  
surfaces and hand-tracking)

IR eye cameras + IR LEDs  
(eye-tracking, iris recognition,  
2+3 microphone array  
display calibration)  
(speech in 90dB noise)

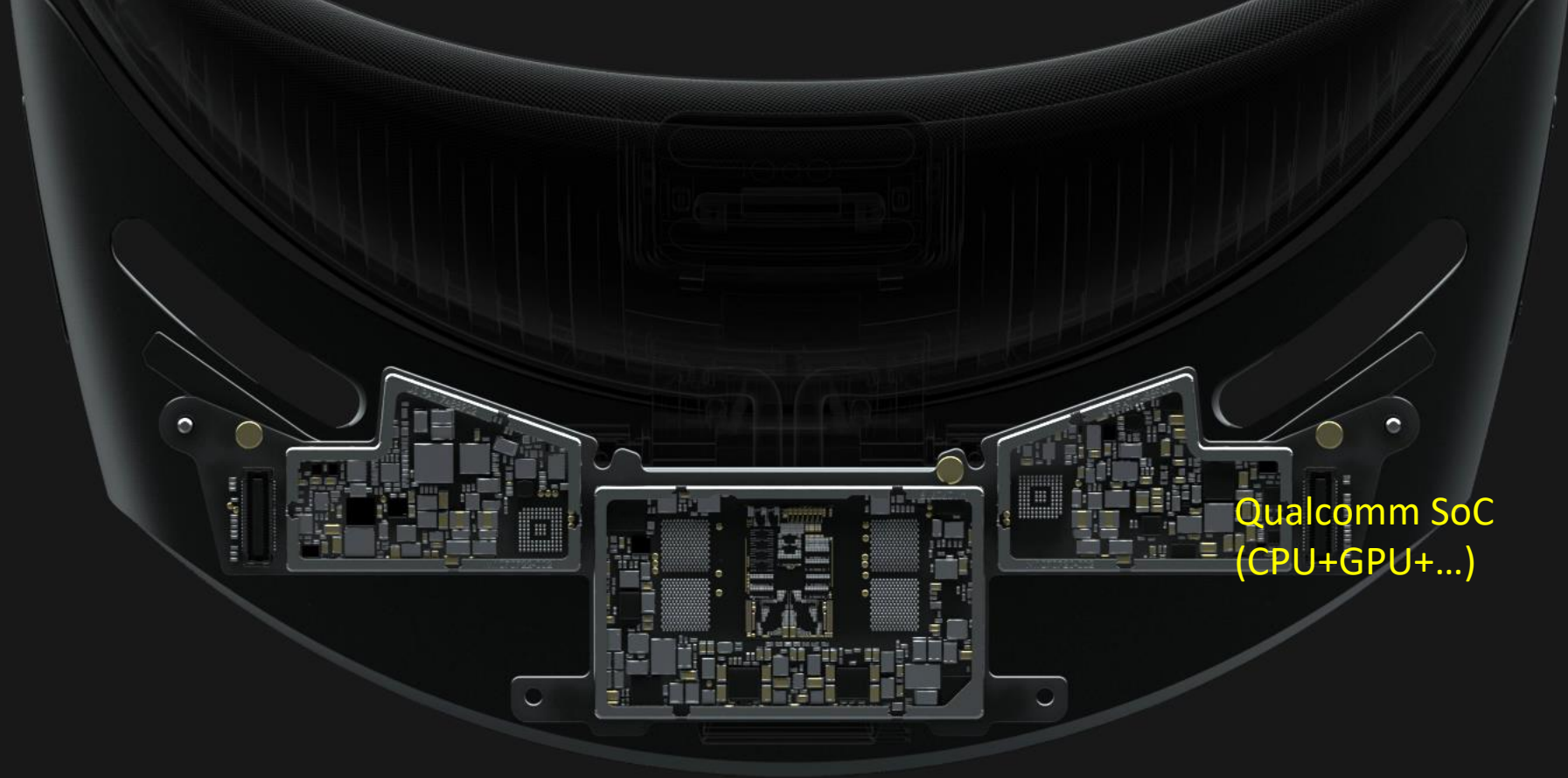


Diet  
Hummingbirds primarily eat flower nectar, tree sap, insect and pollen. The hummingbird's fast breathing rate, fast heartbeat and high body temperature, require that they eat a lot of food each day. Hummingbirds have a long tongue which they use to lick the nectar from the flowers.

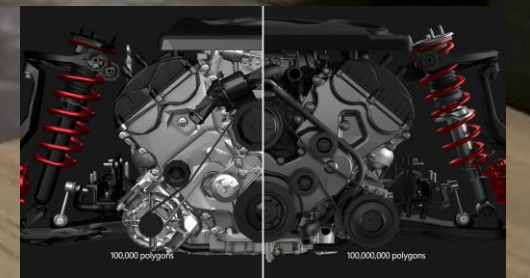




HPU (DSPs, DNN AI core, LSR)



Qualcomm SoC  
(CPU+GPU+...)



Interactive Mixed Reality rendering of high-resolution Digital Twins and Reality Capture using Azure Remote Rendering





**Microsoft Mesh  
App (preview)**

Preview

 AZURE SPATIAL ANCHORS



Enables to *share* and *persist* spatial information across devices



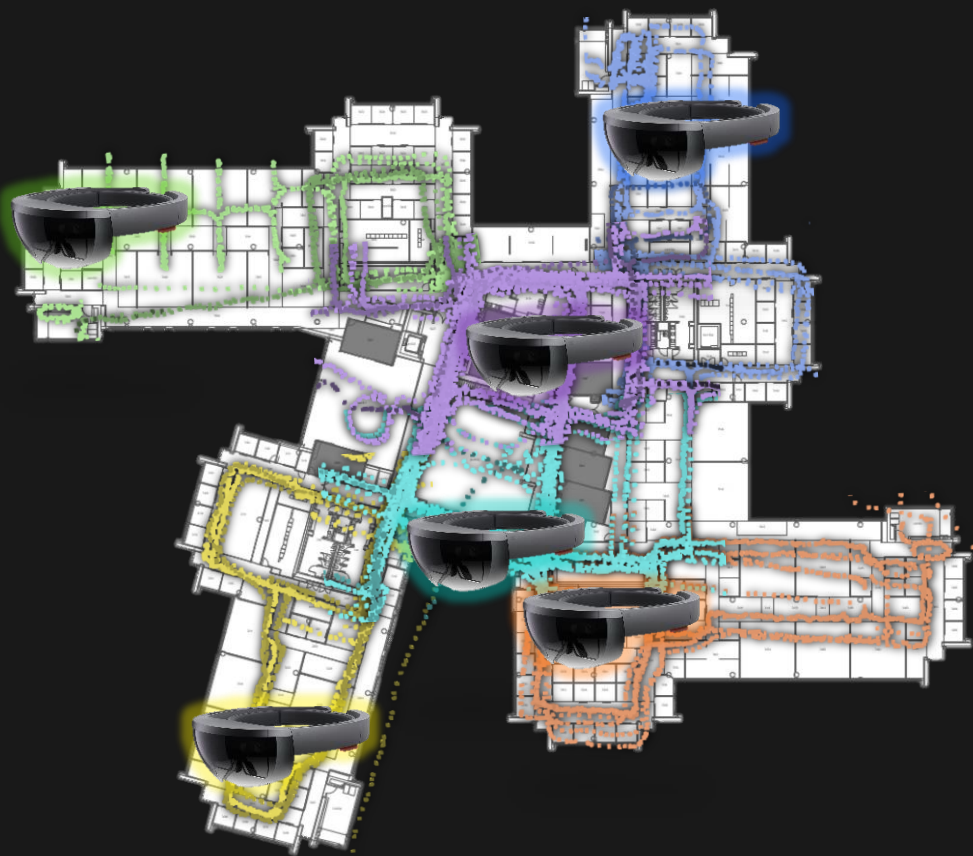


on-site access to Digital Twin

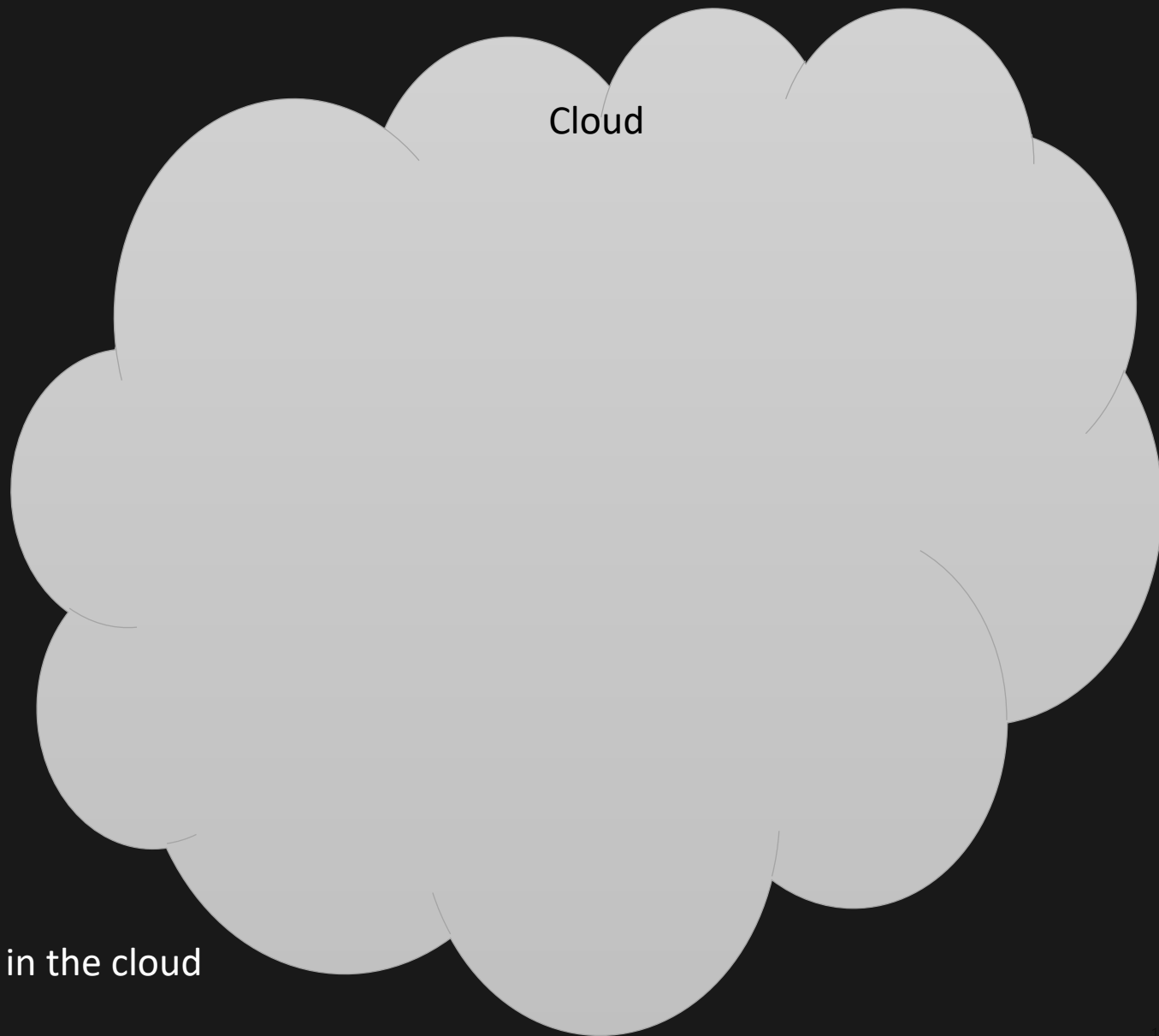


# 6DOF relocalization map

Clients

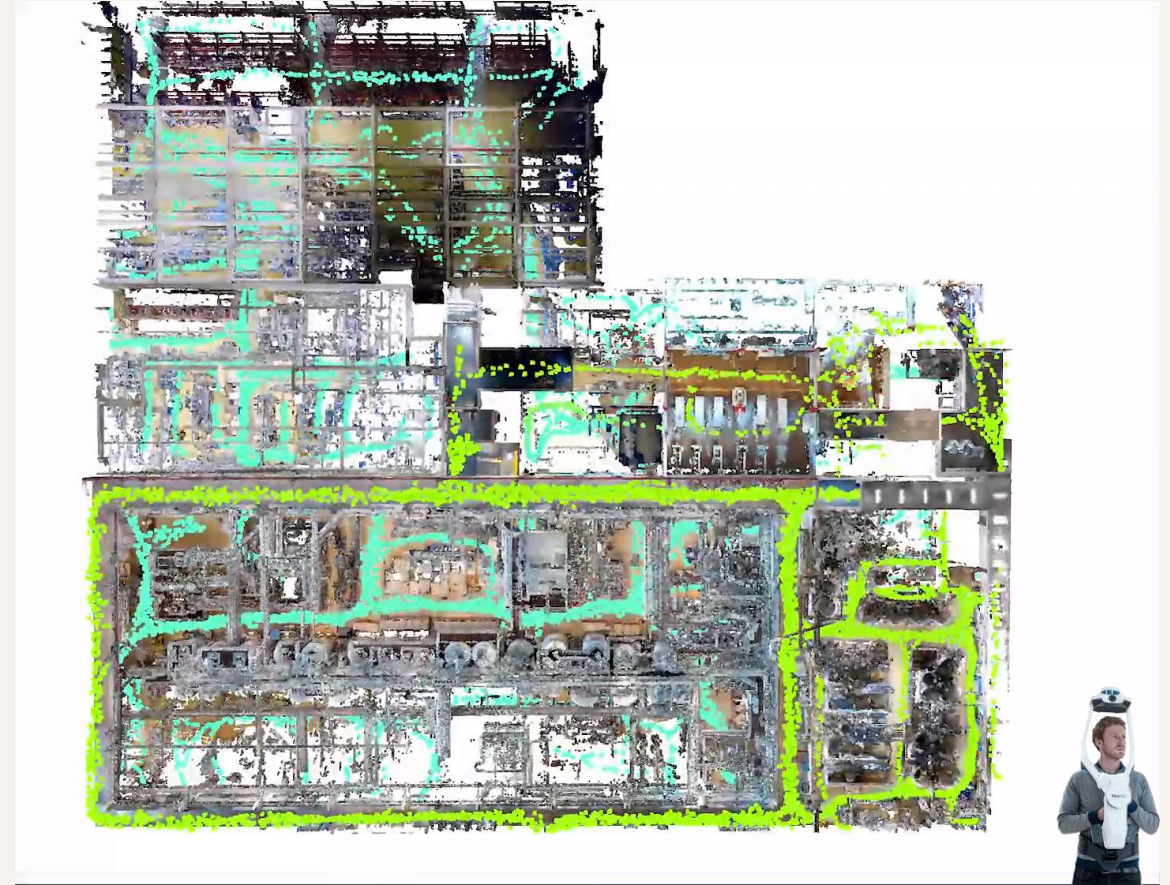


Cloud



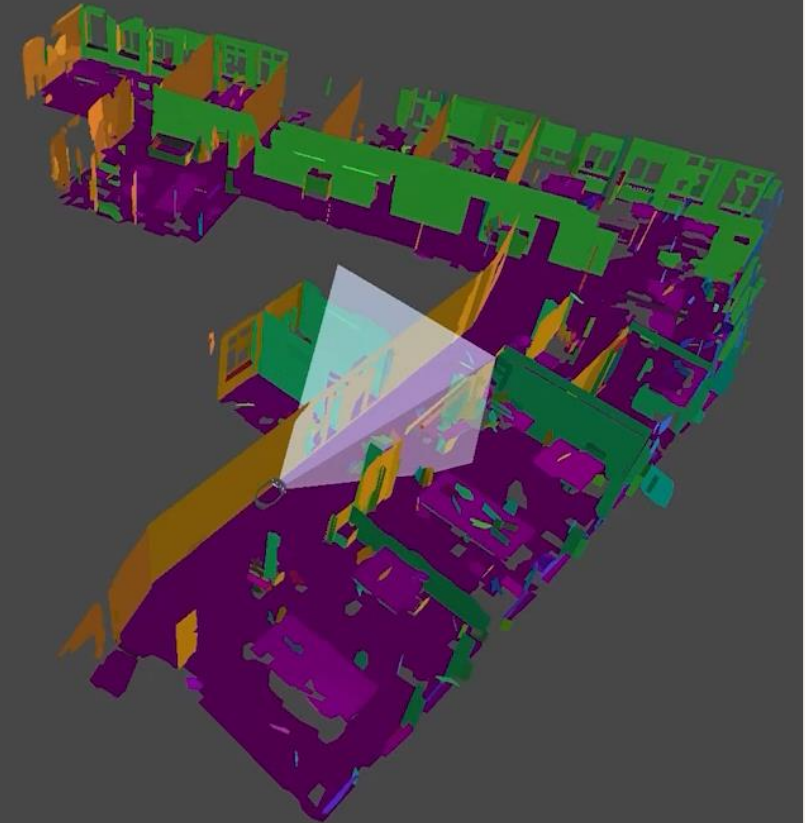
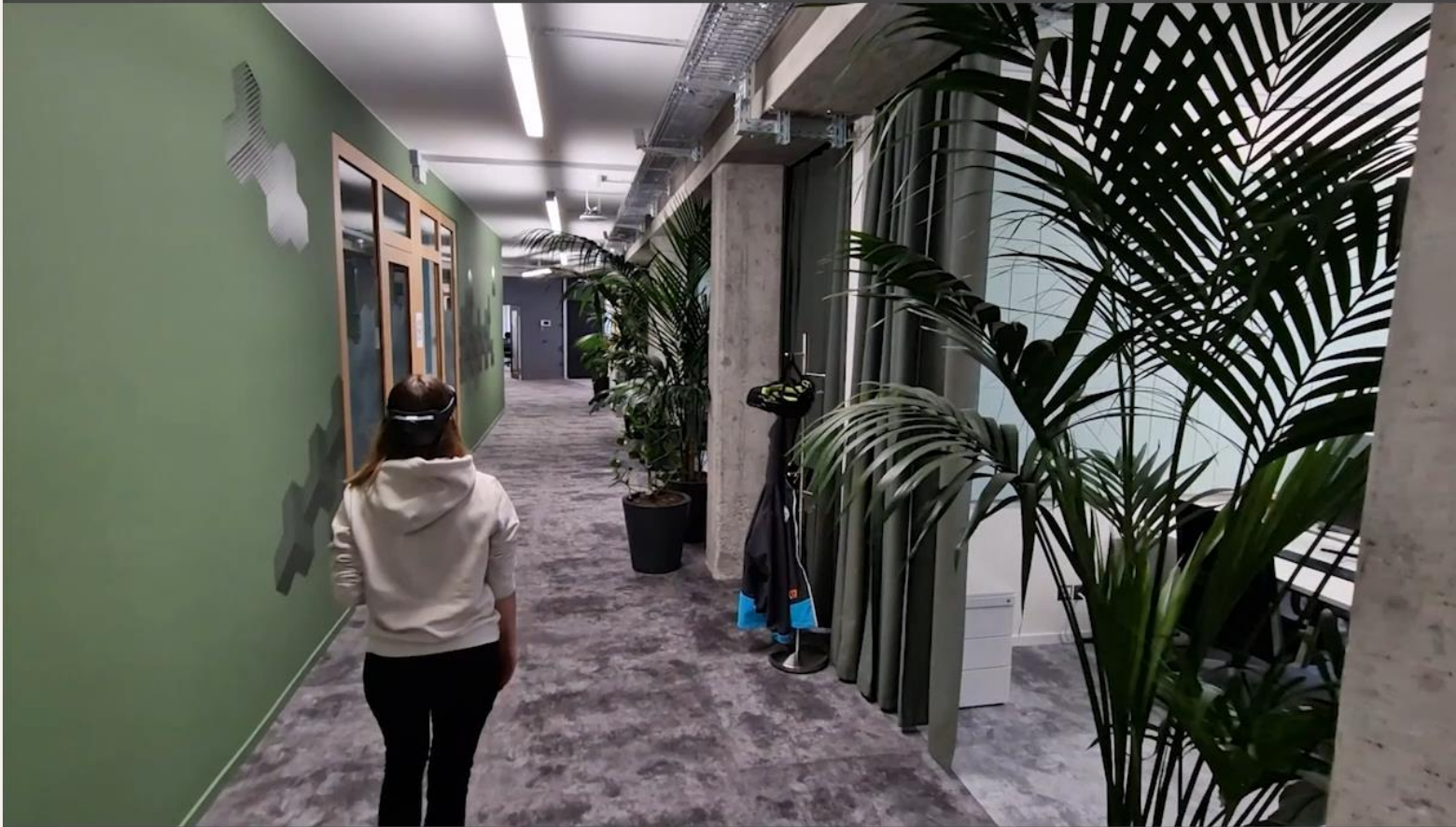
combine multiple maps in the cloud

# Emmi plant



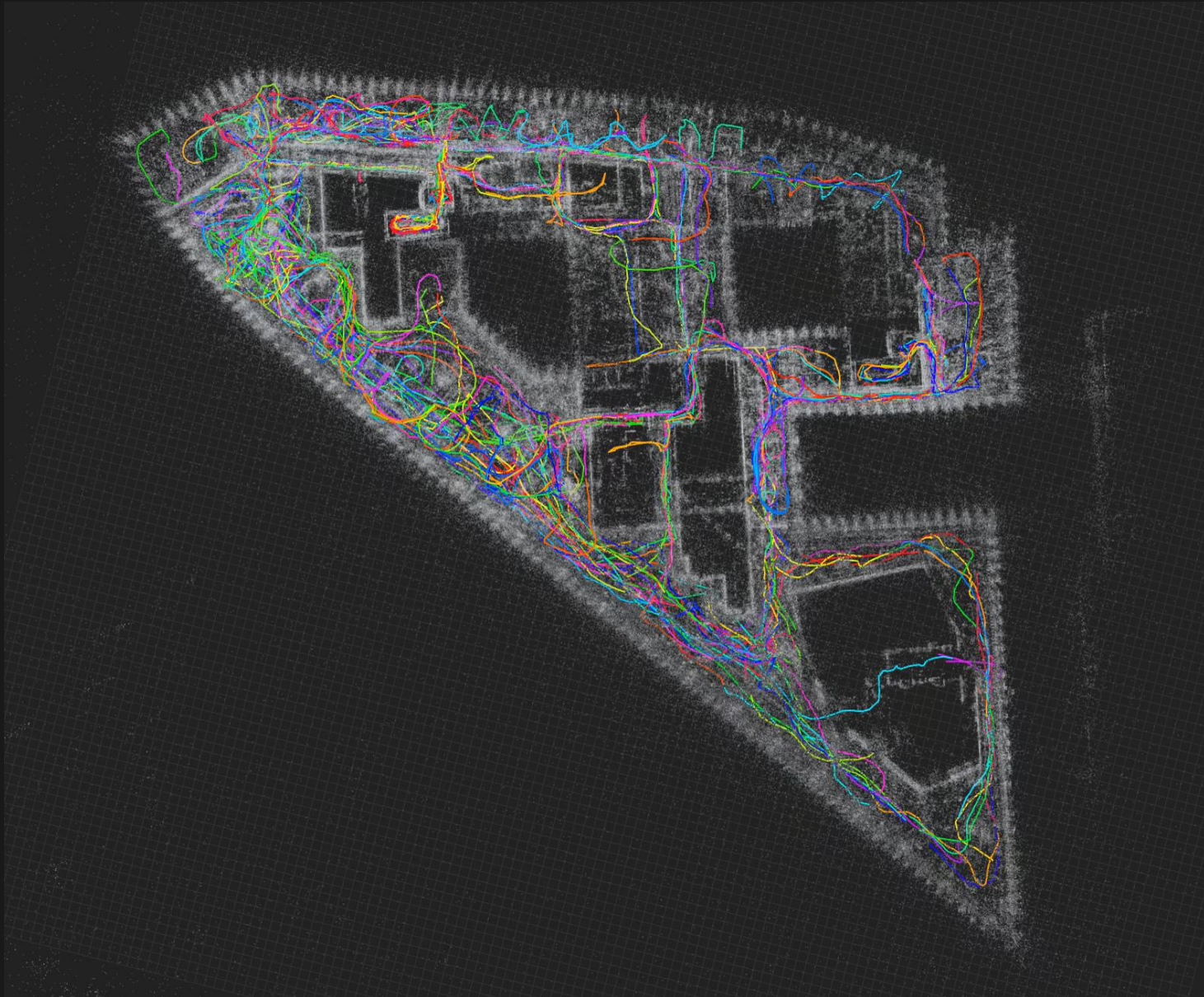


# Spatial data access





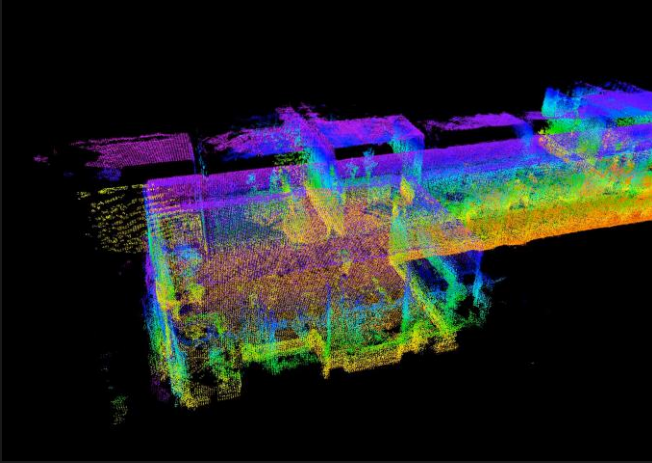
# The Circle (Zurich Airport)



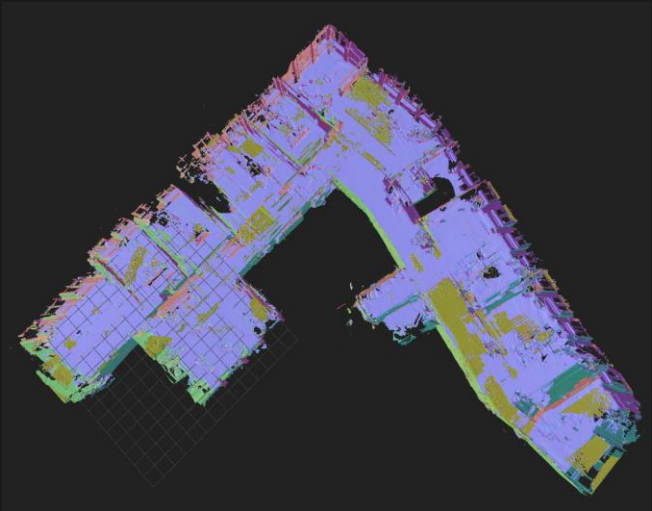


# Looking beyond visual features

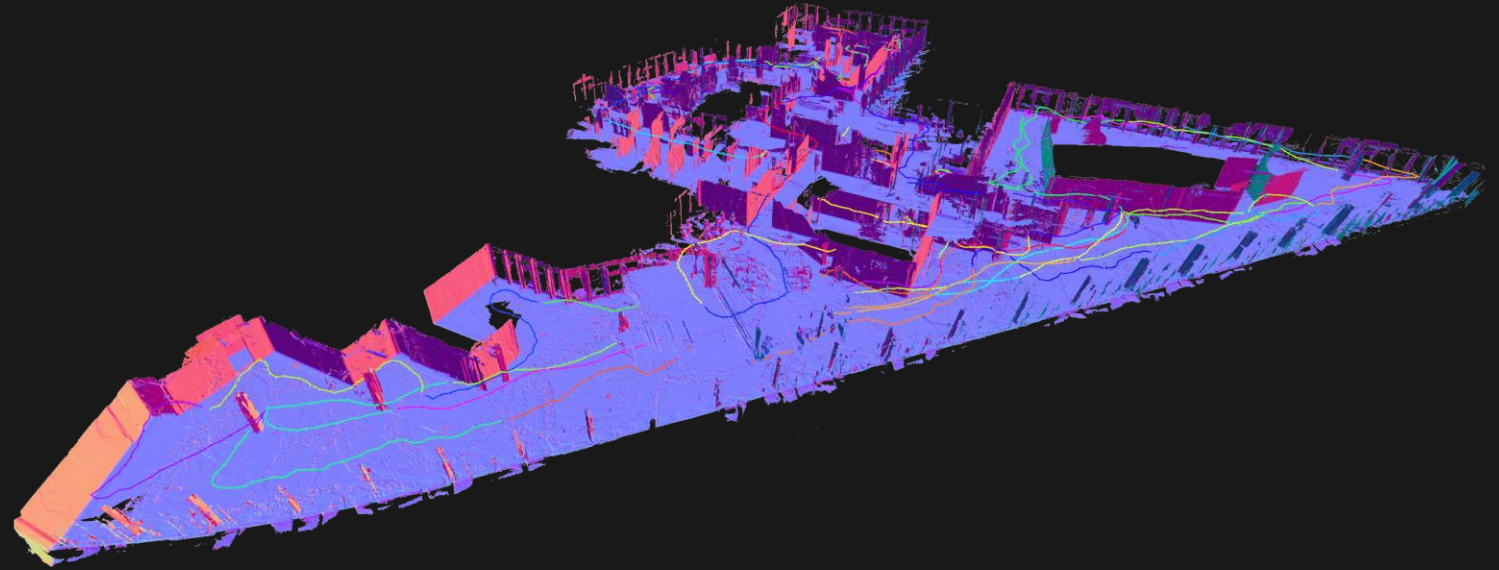
depth, 3p ingestion (digital twin, navvis), lidar, robots



Raw ouster lidar (mounted on spot) point cloud using map poses



Dense single-session meshing result using HL2 data



Dense multi-session meshing result using HL2







Emergency Stop

Navigation toolbar containing a mouse cursor icon, three location pin icons (one with a plus sign, one with an 'x', and one with a plus sign), and the text 'DS'.

Start Mission

Next Mission: 08:00



**WAYPOINTS**

Waypoint management panel with a large empty area and navigation controls (up, down, and close icons) at the bottom.

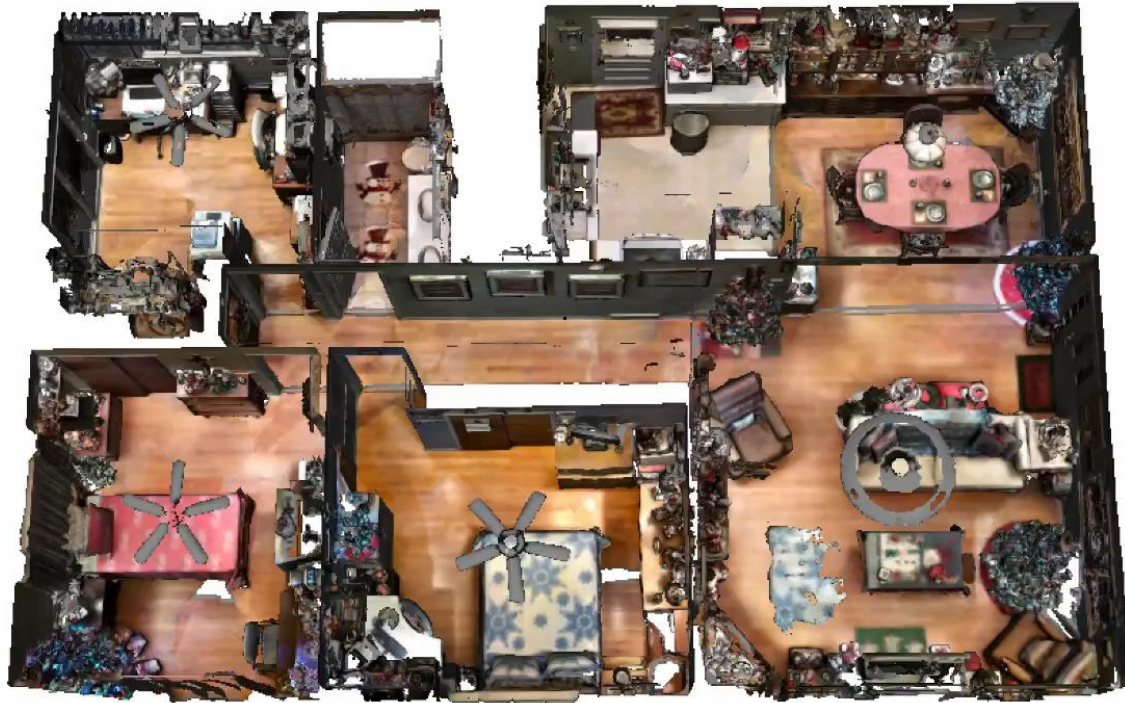
Import...

Export...

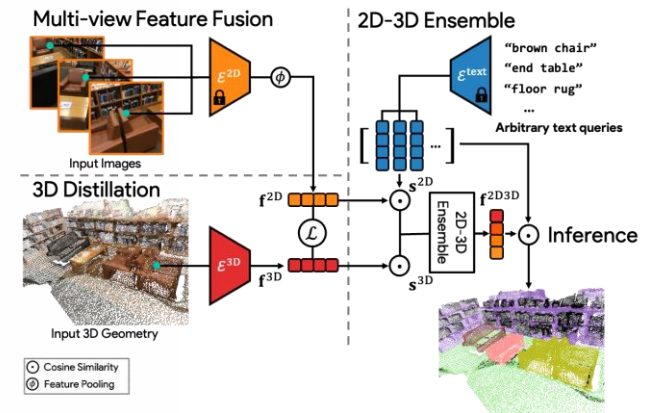
Export To Robot...

This Unity application shows how we can design autonomous robot missions

Text queries:



Extends OpenAI CLIP joint text-image embedding towards 3D geometry



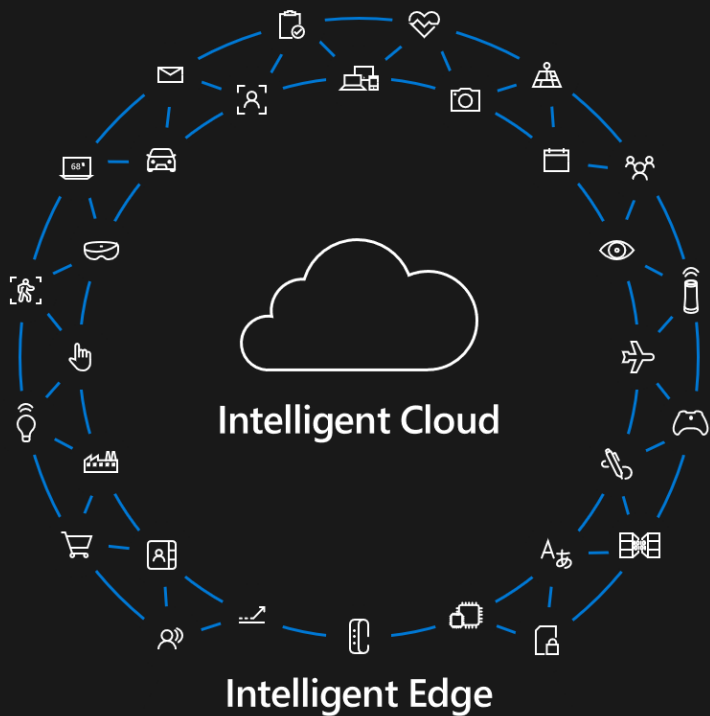
# Mixed Reality & AI



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



# Opportunities & open issues



- Combine edge and cloud computing
- Tele-presence
  - Project remote presence of user
  - Allow user to access remote places
- Synergies between HMDs, robotics, ambient sensing and IoT
- Privacy

**Thank you!**  
**Questions?**