

Media for AR/VR

MOPS WG

IETF 116, March 2023

Cullen Jennings <fluffy@cisco.com>

Photo Realistic Holographic Meetings



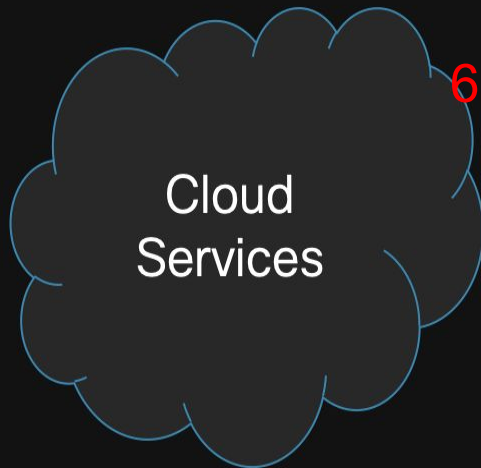
High-Level Overview Today

1. Capture Device



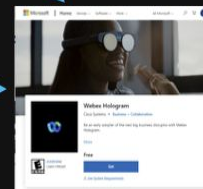
30mbps

2. Cloud Services



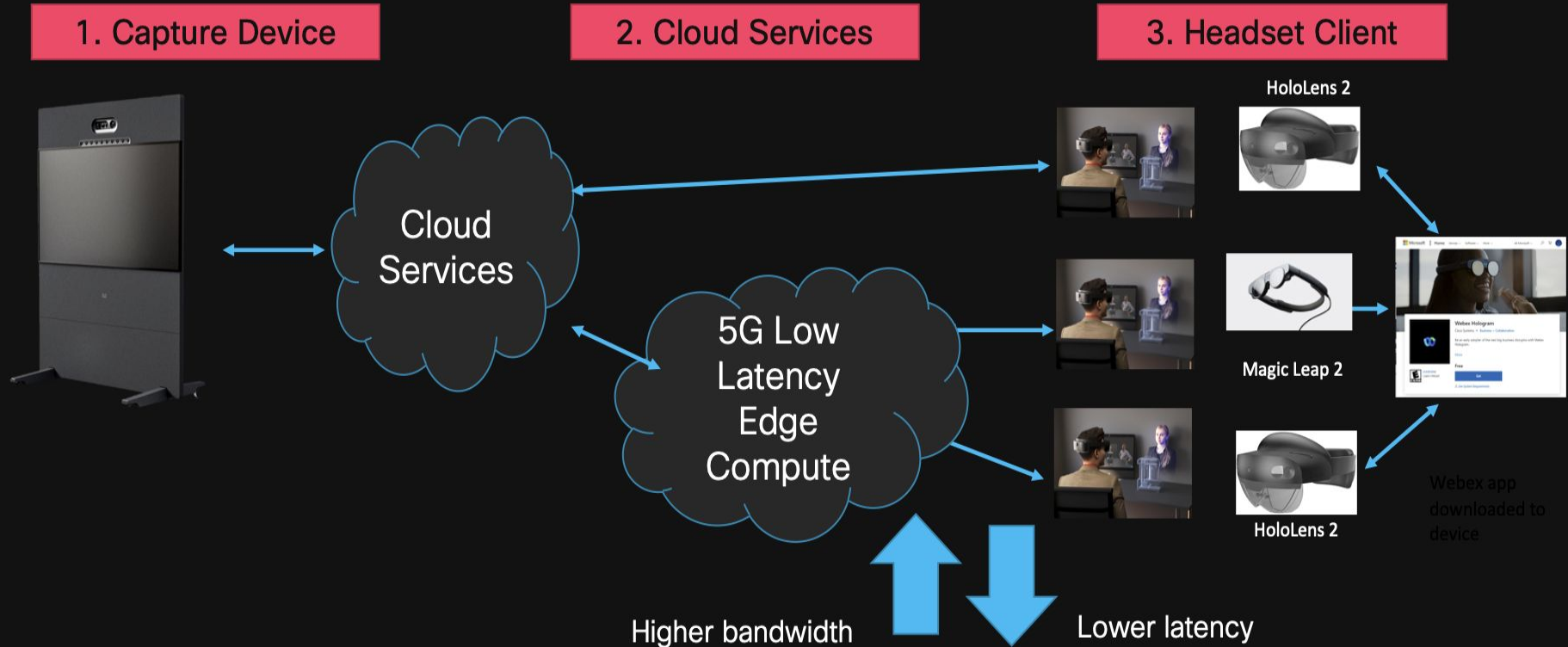
6mbps

3. Headset Client



Webex app
downloaded to
device

High-Level Overview in Future



Latency

Many things contribute to “time to puke” — image lag is one of them.

When headset users rotate their heads, updates need to be well under 10 ms.

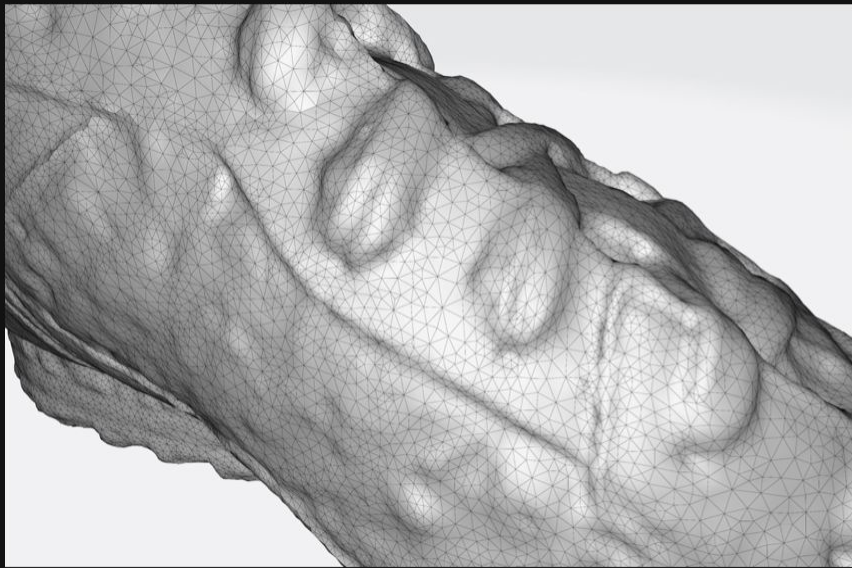
As the headsets resolution in pixels / degree gets better, this update needs to be faster.

When you shift your head, time to update parallax can be much greater than 10 ms.

“Interactive” feel for AR/VR is about the same as audio/video $< 300\text{ms}$

“Live” video is a superset of Interactive and has no lower time bound.

Texture-Mapped Polygons



Point Clouds

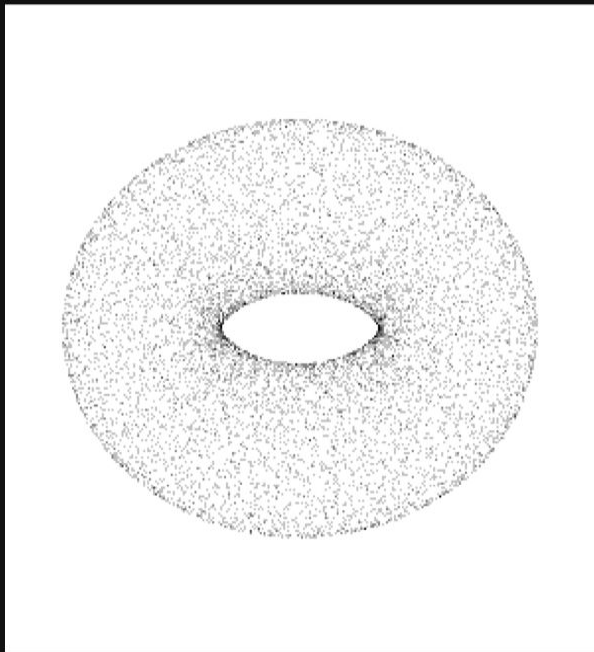
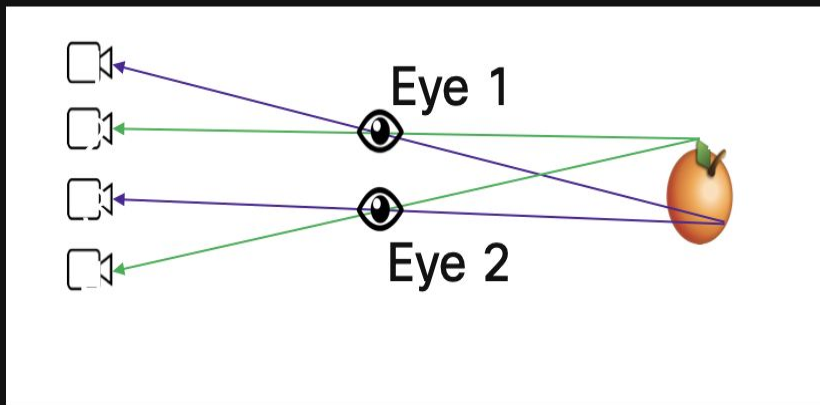


Image from Lucas Vieira

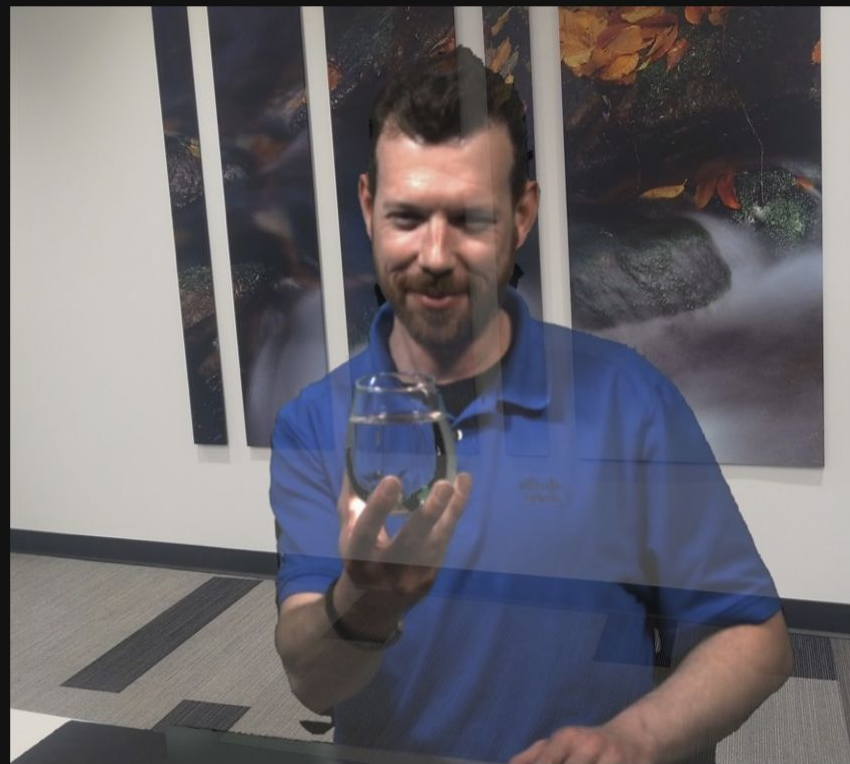


Model from Michael Nicolayeff

How Light Fields Work



Light Field Holograms



Comparisons

	Captured Polygons	Point Clouds	Light Fields
Capture	One depth camera	Several depth camera	Many cameras
Data size	Small	Large	Huge
Fidelity	Meh	OK	Great
Impact of Artifacts	Horrible	Bad	Subtle
Interaction and Intersection	Easy	Hard	Challenging
Translucent Objects	No	No	Yes
Reflective Objects	Bad	No	Awesome
Real-Time Capture	Not great	OK	Super

Neural Radiance Fields (NeRF) are coming



See: NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis by Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, Ren Ng in CVPR 2020

Media for object manipulation & hand gestures



Encode the location of joints in your hand

Encode the location of objects in the virtual scene

<https://datatracker.ietf.org/doc/html/draft-jennings-dispatch-game-state-over-rtp>

Key things for MOPs WG

Photorealistic AR/VR hologram and high-quality avatars are possible on today's networks.

We have specific information about latency and implications from Webex Hologram that partially apply to other future 3D collaboration systems.

We should be thinking about implications of 3D non-video media types.

Better definitions of Interactive and Live media.

The MoQ protocol looks like it may be better than RTP for this.