MIXED REALITY PASSTHROUGH

Meta Quest / Oculus and Apple Vision Pro

Laboratorio Realtà Virtuale 2025/2026

eleonora.chitti@unimi.it

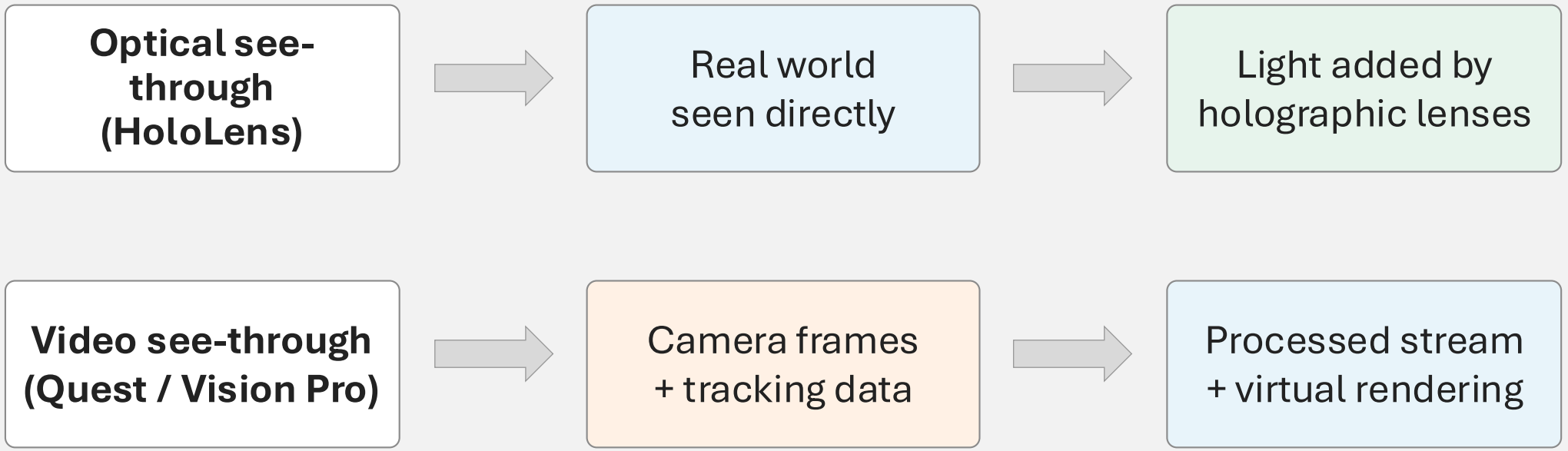
Mixed Reality

Mixed Reality integrates virtual objects with the real world so that digital content is spatially anchored, responsive, and partially occluded by physical geometry.

- In optical AR/MR, the user sees the real world directly through transparent optics (Holograms).
- In video see-through MR, the user sees a real-time camera reconstruction of the world on displays (Passtrhough).
- Meta Quest and Apple Vision Pro use the second approach: the headset modifies a camera stream and composites virtual content into it.

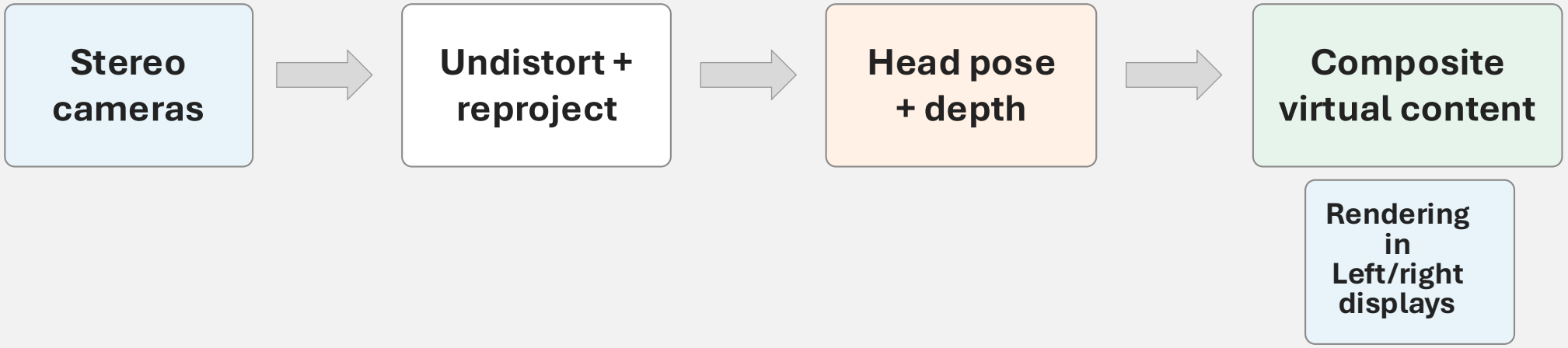
From Holograms to Passthrough

HoloLens creates “holograms” by adding light to what the user directly sees through waveguides. Passthrough headsets rebuild the real world as video first, then draw virtual content on it.



- Main consequence: the operating system owns the camera-to-display pipeline, because latency and user safety are critical.

Video See-through MR Pipeline (Passthrough)



- The user does not receive “raw reality”; they receive a stereo reconstruction optimized for comfort and latency.
- The app usually renders transparent virtual layers; the system compositor blends them with passthrough.
- Depth and scene meshes allow occlusion: real tables, walls, and hands can hide virtual objects.

Oculus / Meta Quest

Meta Quest 3 and Quest 3S are standalone VR/MR headsets. For mixed reality they use full-color passthrough: front cameras capture the room, the headset reconstructs it, and apps render content into that view.

- Standalone device: no PC is required for native apps.
- Passthrough is a video feed, not transparent optics.
- Supports controllers, direct hand tracking, spatial anchors, scene data, and depth-based occlusion.

Source: meta.com/quest/quest-3

Mixed reality goal:
make digital content behave as if it belongs in the real room.

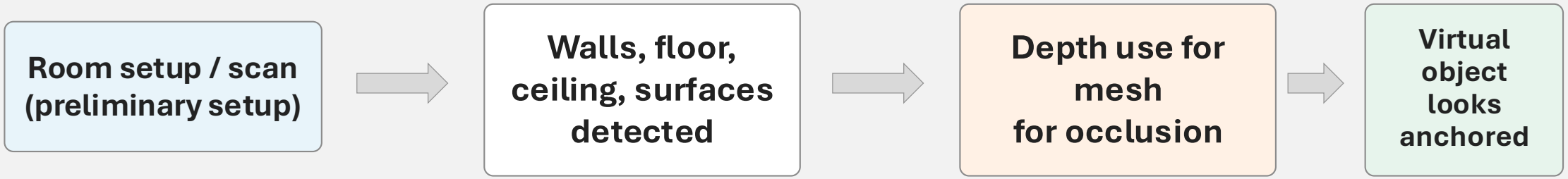
Meta Quest: Sensors and Data

- RGB cameras: full-color passthrough for the physical environment.
- Tracking cameras and IMU: inside-out tracking of head pose and controllers.
- Depth understanding: estimates room geometry for occlusion and physics.
- Microphones and spatial audio: voice input and spatialized output.

Key idea for developers:
render in the same spatial coordinate system used by the tracked room.

Meta Quest: Scene Understanding

For MR, the headset needs an approximate model of the room so virtual objects can be *placed*, *hidden*, and *interacted with* consistently.



- Spatial anchors keep content at stable real-world locations.
- Scene API / MR Utility Kit (Meta XR API) expose room data to applications.
- Depth API (Meta XR API) enables realistic occlusion: a real desk can hide a virtual ball.

Meta Quest Interaction Model

Quest interaction is centered on controllers and hands, with passthrough making the physical room visible while still running a VR-style application.

- Touch Plus controllers: buttons, triggers, thumbsticks, haptics, 6-DoF tracking.
- Hand tracking: direct touch, pinch, grab, poke, and ray-based UI selection.
- Head and hand pose are used together: the user can reach for near objects or point at distant UI.
- Voice input is available at system or app level depending on the software stack.

Apple Vision Pro

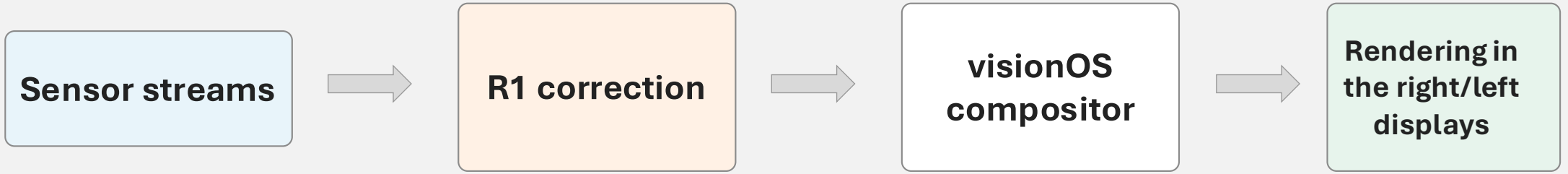
Apple Vision Pro also uses video see-through: cameras capture the environment, the R1 chip processes sensor streams with low latency, and visionOS blends digital content with physical space.

- Input: eyes, hands, and voice.
- Output: high-resolution micro-OLED displays and spatial audio.
- System model: immersive spaces in 3D.

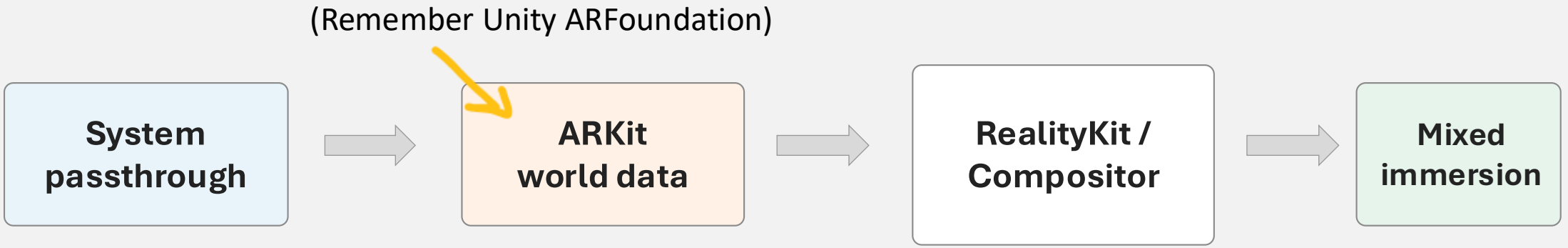
Apple avoids exposing passthrough as a normal app-owned camera stream for privacy and safety.

Vision Pro: Sensors and Compute

- Two high-resolution cameras create the stereoscopic view of the real world.
- Six world-facing tracking cameras, LiDAR, TrueDepth, IMUs, flicker, and ambient-light sensors support spatial tracking.
- Four eye-tracking cameras enable gaze-based targeting and Optic ID.
- M-series chip runs apps; the R1 chip low-latency camera-to-display processing.



visionOS: Modifying the Camera Stream



- Use Mixed Immersion to blend app content with passthrough in a Full Space.
- Use anchors, plane detection, scene geometry, hand tracking, and lighting estimates for spatial consistency (remember ARFoundation exercise).
- Direct camera access is restricted; enhanced camera access exists for enterprise use cases with specific entitlements.

Vision Pro Interaction Model

Vision Pro interaction is designed to be "*look, then pinch*": gaze targets the element and a finger gesture confirms the action.

Hand tracking also supports direct interactions with nearby content.

- Eyes: select the target by looking at it.
- Hands: tap fingers, pinch, drag, or directly touch nearby spatial UI.
- Voice: dictate text, invoke system commands, or control app features.
- Upper limbs (arms+hands - extended FoV) can be shown in front of or behind rendered content to reduce visual conflicts with passthrough.

Design principle:
keep the real environment understandable because users rely on passthrough for safety and orientation.

HoloLens vs Passthrough Headsets

Aspect	HoloLens	Meta Quest	Vision Pro
Reality	Direct optical view	Camera passthrough	Camera passthrough
Virtual layer	Light added by waveguides	System passthrough layer + app render	visionOS compositor + app render
Occlusion	Spatial mesh blocks holograms	Depth / scene mesh	ARKit scene geometry + depth
Input	Gaze, gestures, voice	Controllers, hands, voice	Eyes, hands, voice
Trade-off	Real world is naturally visible but FOV is narrow	Wide VR content, but camera artifacts/latency matter	High quality blend, but stricter access to sensor data

Passthrough MR Pros and Cons

Pros:

- Combines full VR rendering with awareness of the physical room.
- Can switch continuously from VR to MR by changing opacity and immersion.
- Good for labs: safe demos, spatial UI, room-scale interaction, and CV/ML experiments.

Cons:

- The real world is mediated by cameras: exposure, blur, warping, and latency can affect comfort.
- Occlusion and physics depend on imperfect depth and scene reconstruction.
- Raw camera access is controlled by the platform, especially on Vision Pro.