

Mehr als oberflächlich: Live Volumetric Streaming für die AR/VR-Kollaboration

Markus Kiesel¹, Daniel Wulf-Misaki²

¹ CMC-Kiesel GmbH, ² oculavis GmbH

Zusammenfassung

Live Volumetric Streaming ist eine von mehreren Technologien, mit denen eine AR/VR-Kollaboration umgesetzt werden kann. Der Beitrag beschreibt Systemarchitektur, technologische Herausforderungen sowie eine prototypische Umsetzung. Die Ergebnisse zeigen, dass Live Volumetric Streaming insbesondere zur besseren Orientierung und Kontexterfassung bei der AR/VR-Kollaboration in der Industrie beitragen kann.

1 Einleitung

Im Forschungsprojekt TeleInteraction-XR wurde ein System zur interaktiven, telepräsenten Zusammenarbeit zwischen Expertinnen oder Experten in einem Backoffice und Servicepersonal im Einsatzfeld entwickelt. Ziel war ein möglichst natürlicher und unterbrechungsfreier Wissensaustausch über räumliche Distanz hinweg. Das Projekt untersuchte hierzu verschiedene technologische Varianten, die sich hinsichtlich Datenquellen, Interaktionsmöglichkeiten und Immersionsgrad unterscheiden [1].

Im Fokus dieses Beitrags steht der Ansatz des Live Volumetric Streamings, der die reale Umgebung des Servicepersonals im Feld mithilfe von Tiefensensoren dreidimensional erfasst, in ein digitales Volumenmodell umwandelt und in Echtzeit an den Experten in der VR-Umgebung überträgt. Im Unterschied zur klassischen Videoübertragungen [2] oder zu statischen 3D-Modellen [3] ermöglicht diese Technologie eine dynamische, echtzeitfähige und perspektivisch freie Sicht in die Einsatzsituation.

Anstatt Informationen mühsam über Sprache, Fotos oder Videos zu übermitteln, nutzt die Person im Feld ein mobiles AR- oder XR-Gerät mit Tiefensensorik. Das damit generierte 3D-Modell wird in Echtzeit an den entfernten Experten bzw. die Expertin gesendet, der oder die die Szene virtuell aus unterschiedlichen Blickwinkeln analysiert, Anomalien erkennt und präzise Handlungsanweisungen erteilen kann – etwa zur Demontage oder Justierung einzelner Komponenten. Der Experte bzw. die Expertin erhält somit ein vollständiges räumliches Verständnis, das über die Möglichkeiten herkömmlicher Videoübertragung hinausgeht. Zudem können digitale Markierungen, Annotationen und digitale Handlungsanweisungen die Kommunikation ergänzen.

2 Grundlagen und Stand der Technik

2.1 Sensorik

Zur Erfassung volumetrischer Daten kommen verschiedene Sensortechnologien zum Einsatz, die geometrische und visuelle Informationen in Echtzeit liefern. Gängige Systeme umfassen RGB-D-Kameras, Time-of-Flight-(ToF)-Sensoren, strukturierte Lichtsysteme und LiDAR-Technologien [4] [5].

Ein etabliertes – und für die Entwicklungen im Projekt TeleInteraction-XR als Vorzugsvariante ausgewähltes – System im Bereich Augmented Reality ist die Microsoft HoloLens, die mittels integriertem Tiefensensor und Spatial-Mapping-Algorithmen ein kontinuierlich aktualisiertes Mesh der Umgebung erzeugt. Dieses dient der Objekterkennung, dem Handling von Verdeckungen (Occlusion) und der präzisen Platzierung virtueller Inhalte. Trotz geringerer Auflösung im Vergleich zu spezialisierten Tiefenkameras bietet die HoloLens durch autarkes Inside-Out-Tracking eine robuste, mobile 3D-Erfassung [5] [6].

Weitere relevante Systeme zur mobilen volumetrischen Erfassung sind:

- **Azure Kinect:** Hochauflösende RGB-D-Kamera mit Weitwinkelobjektiv, besonders für präzise Nahbereichserfassung geeignet.
- **Intel RealSense:** Portable Kameraserie mit Varianten wie D435i und L515, geeignet für flexible Anwendungsszenarien.
- **LiDAR-Sensoren in Consumer-Geräten** (z. B. iPad Pro, iPhone Pro): Erlauben eine mobile Datenerfassung ohne Zusatzhardware.
- **Stereoskopische Verfahren:** Nutzung von zwei oder mehr RGB-Kameras zur Tiefenerkennung, etwa durch Googles ARCore Depth API.

Neben der Hardware ist auch die Software-Pipeline entscheidend. Während Geräte wie die HoloLens die komplette Rekonstruktion direkt auf dem Gerät ausführen (On-Device), übertragen andere Systeme Rohdaten oder Punktwolken zur externen Weiterverarbeitung. Abhängig vom Einsatzszenario werden die Daten entweder als Punktwolken oder als weiterverarbeitete Oberflächen-Meshs genutzt.

Die Auswahl geeigneter Sensortechnologie richtet sich nach Anwendungskontext, erforderlicher Mobilität, Datenqualität, Energieverbrauch und Echtzeitanforderungen. Für industrielle Szenarien sind Systeme notwendig, die kontinuierlich aktualisierte volumetrische Daten bereitstellen und nahtlos mit AR-/VR-Umgebungen zu synchronisieren sind.

2.2 Mesh-Erzeugung und Datenverarbeitung

Die Echtzeit-Erzeugung dynamischer 3D-Meshes aus Sensordaten stellt hohe Anforderungen an Verarbeitungsgeschwindigkeit und Stabilität. Der Prozess erfolgt in mehreren Schritten:

1. **Erfassung von Punktwolken:** Aus den Eingangsdaten werden kontinuierlich Punktwolken generiert.
2. **Vorverarbeitung:** Rauschreduktion, Glättung und Normalenschätzung verbessern die Qualität und Stabilität der späteren Geometrie.
3. **Mesh-Erzeugung:** Algorithmen wie Marching Cubes, Poisson Surface Reconstruction oder Delaunay-Triangulierung erzeugen daraus stabile 3D-Modelle.

Die Microsoft HoloLens 2 zeigt eine effiziente Verarbeitung: Ihr Spatial Mapping rekonstruiert die Umgebung laufend auf dem Gerät. Neue Bereiche werden automatisch integriert, bestehende aktualisiert. Das System basiert auf voxelbasierten Partitionierungen und sogenannten Spatial Surface Observers, was gezielte und ressourcenschonende Aktualisierungen ermöglicht [4] [5] [6].

3 Systemarchitektur und technische Umsetzung

3.1 Gesamtarchitektur

Die im Projekt TeleInteraction-XR entwickelte Systemarchitektur besteht aus mehreren miteinander verknüpften Kernkomponenten:

- **Kollaborations-Backend-Server**
Das zentrale Backend empfängt die vom AR-Nutzer erfassten volumetrischen Daten, verarbeitet sie in Echtzeit und stellt sie dem VR-Client zur Verfügung. Neben dem Mesh-Routing synchronisiert der Server zusätzliche Datenquellen, darunter Transformationsdaten aus 3D-Geometrien und Avatare. Dadurch entsteht eine konsistente, synchrone Datenbasis für alle Beteiligten.
- **CMC ViewR als VR-Client**
Zur Darstellung und Interaktion dient die VR-Plattform CMC ViewR. Sie erlaubt die dreidimensionale Visualisierung und Transformation (Skalierung, Rotation, Translation) empfangener oder importierter 3D-Geometrien. Die Navigation erfolgt durch Raumbewegung (Roomscale), Teleportation oder einen Flugmodus – letzterer eignet sich besonders für großflächige oder komplexe Szenarien.
- **Oculavis Share auf der AR-Seite**
Die AR-Komponente basiert auf der Plattform oculavis Share und wird in Form der MR Share-App auf der Microsoft HoloLens 2 eingesetzt. Diese dient sowohl als Sensorplattform zur volumetrischen Datenerfassung als auch zur AR-Ausgabe von Informationen und Annotationen. Die erfassten Mesh-Daten werden kontinuierlich in Echtzeit an den Backend-Server und von dort an die VR-Anwendung weitergeleitet.

Die Architektur ist offen konzipiert, um verschiedene VR-Systeme zu integrieren. Im Projekt kamen beispielsweise die HTC Vive Pro und die Meta Quest 3 zum Einsatz, je nach spezifischem Nutzungsszenario – von stationärer industrieller Anwendung bis hin zu mobilen Einsatzkontexten.

Ein entscheidender Vorteil des Systems besteht darin, dass beide Nutzer – AR- und VR-Seite – dieselbe volumetrische Datenbasis verwenden. Dies erlaubt exakte, positionsgenaue Annotationen und eine unmissverständliche Kommunikation.

3.2 Erfassungssystem

Die Auswahl eines geeigneten Erfassungssystems fokussierte sich frühzeitig auf AR/XR-Headsets, da alternative Technologien aufgrund gravierender Nachteile für den Einsatz zur Fernunterstützung in der Industrie ausschieden:

- **LiDAR-Sensoren**, wie sie zum Beispiel im iPad Pro verbaut sind, liefern sehr präzise Tiefendaten, benötigen aber externe Verarbeitungseinheiten für die Mesh-Erzeugung, was ihre Mobilität einschränkt.
- **RGB-D-Kameras** (z. B. Azure Kinect) bieten eine höhere Auflösung, sind aber ebenfalls auf externe Rechner angewiesen und somit nicht autark.
- **Stereoskopische Systeme** wie die ARCore Depth API sind preiswert, aber haben Probleme bei schwierigen Lichtverhältnissen und reflektierenden Oberflächen.

Die Eignung gängiger AR/XR-Headsets wurde anhand von AR-Fähigkeit, Kamera-Zugriffsrechten, Leistungsfähigkeit und Marktverfügbarkeit analysiert (Tabelle 1). Obwohl Geräte wie Oculus Quest 3 und Apple Vision Pro technisch leistungsfähig sind, zeigten sie im Projektkontext erhebliche Einschränkungen. Meta- und Apple-Geräten schieden aufgrund der zum Zeitpunkt der Analyse restriktiven Zugriffsrechte auf Kamera- und Tiefensensor-Daten aus.

Gerät	AR-fähig	Verfügbarkeit	Kamera-Stream	Leistungsfähigkeit
HoloLens 2	✓	✓	✓	+
Magic Leap 2	✓	(✓)	✓	+
Oculus Quest 3	(✓)	✓	(X)	+++
Pico 4	(✓)	✓	X	+++
Vision Pro	(✓)	✓	✓	+++
Pico 4 Ultra Enterprise	(✓)	✓	✓	+++

Tabelle 1: Gerätevergleich

Im Projekt TeleInteraction-XR wurde daher für die volumetrische Datenerfassung und Darstellung auf die Microsoft HoloLens 2 gesetzt. Der Fokus lag auf

einer mobilen, autarken Lösung, die ohne externe Hardware auskommt und sich direkt in industriellen Anwendungsszenarien einsetzen lässt. Die HoloLens 2 erfüllt als einziges getestetes Gerät alle relevanten Anforderungen für den produktiven Einsatz: Sie bietet integrierte Tiefensensorik, On-Device-Verarbeitung, eine robuste Spatial-Mapping-Funktionalität und direkten Zugriff auf die Kameras.

Die HoloLens 2 nutzt ein integriertes Time-of-Flight-(ToF)-System, das mittels Infrarotimpulsen Tiefeninformationen erfasst. Zusammen mit RGB-Kameras entsteht so ein adaptives Mesh, dessen Polygonauflösung automatisch an die Komplexität der Umgebung angepasst wird. Die Mesh-Generierung, Verarbeitung und Aktualisierung erfolgen vollständig Onboard – externe Recheneinheiten sind überflüssig. Änderungen in der Szene werden kontinuierlich erkannt und in das Modell integriert [4] [5] [6].

Im praktischen Einsatz zeigten sich jedoch auch folgende Einschränkungen:

- Die Auflösung des generierten Meshs ist durch die Hardware begrenzt.
- Das Sichtfeld von ca. 52° erschwert die vollständige Erfassung großer Maschinen oder komplexer Szenen.
- Die Akkulaufzeit liegt bei etwa 2 bis 3 Stunden und limitiert die Einsatzdauer.

Speziell bei der Datenerfassung traten folgende Herausforderungen auf:

- Mesh-Lücken bei schlechten Lichtverhältnissen oder stark reflektierenden Oberflächen,
- Latenzen bei hoher Rechenlast auf dem Gerät und
- die Notwendigkeit manueller Kalibrierungsschritte, um die Genauigkeit über die standardmäßig integrierten Routinen hinaus zu verbessern.

Dennoch hat sich die HoloLens 2 als derzeit einzig praxistaugliche Lösung für die mobile volumetrische Echtzeit-Erfassung bewährt – insbesondere in Verbindung mit gängigen Frameworks wie Unity AR Foundation.

3.3 Datenverarbeitung und technische Umsetzung des Volumetric Streaming

Die von dem AR/XR-Headset generierten Meshs werden über eine eigens entwickelte WebSocket-Schnittstelle an den Backend-Server übertragen. Die WebSockets erlauben eine latenzarme, stabile und plattformunabhängige

Übertragung in Echtzeit. Der Server verarbeitet die eingehenden Daten direkt weiter und verteilt sie über eine optimierte gRPC-Schnittstelle an alle aktiven VR-Teilnehmenden einer Kollaborationssession. Die gRPC-Technologie gewährleistet minimale Kommunikationslatenz und geringen Overhead. Das System unterstützt mehrere parallele Sessions. Alle jeweils registrierten Nutzenden können – unabhängig vom Standort – auf die identischen volumetrischen Daten zugreifen. Eine Übersicht ist in Abbildung 1 dargestellt.

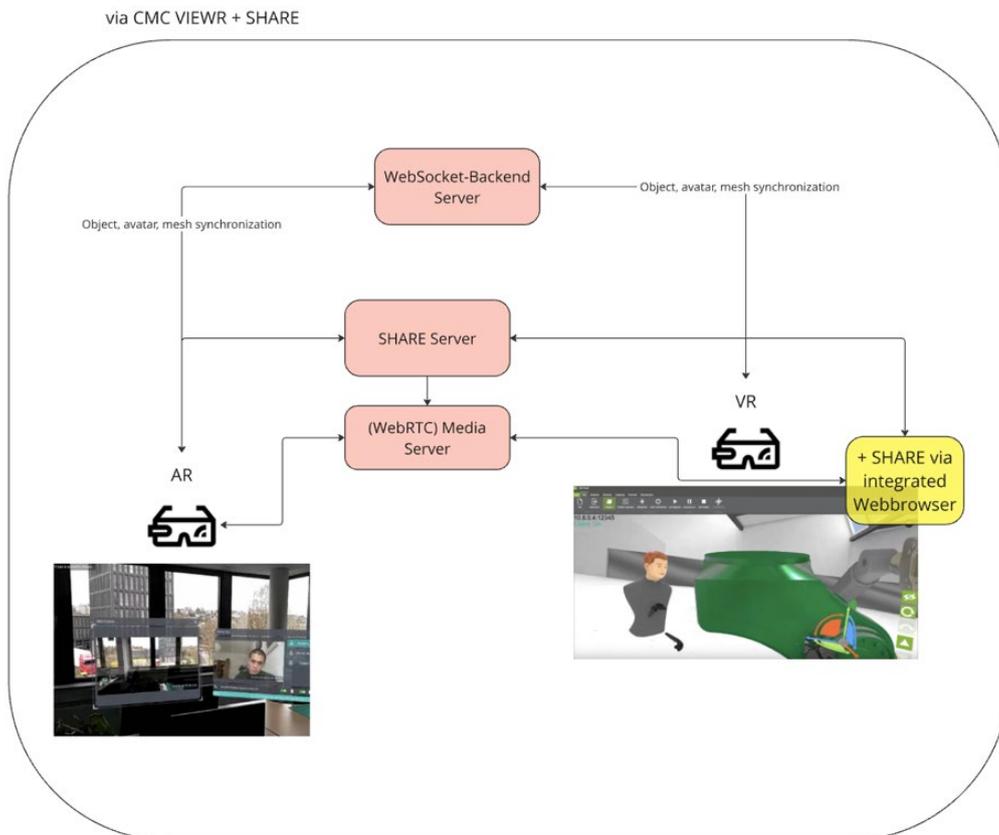


Abbildung 1: Übersicht der Systeme

Optimierungen zur Reduktion von Latenz und Ressourcenverbrauch: Um eine möglichst stabile und angenehme Erfahrung der Nutzenden zu gewährleisten – auch in Hinblick auf VR-typische Herausforderungen wie Motion Sickness – wurden zwei zentrale Optimierungsmaßnahmen implementiert:

- **On-Demand-Refresh:**
Mesh-Updates werden nicht in festen Intervallen, sondern ausschließlich bei signifikanten Veränderungen im Umgebungsmodell gesendet. Zusätzlich wird die Anzahl der Aktualisierungen pro Minute begrenzt. Dadurch sinken sowohl Datenvolumen als auch Server- und Clientlast erheblich.
- **Asynchrones Mesh-Loading und segmentierte Verarbeitung in ViewR:**
Innerhalb der ViewR-VR-Plattform erfolgt das Laden und Konvertieren der Mesh-Daten asynchron in separaten Verarbeitungsthreads. Dies führt zu einer stabileren Bildrate und reduziert wahrgenommene Latenzen deutlich.
Die Meshs sind in **Segmenten** organisiert, was eine inkrementelle Aktualisierung erlaubt: Neue Raumbereiche werden ergänzt, ohne bereits verarbeitete Mesh-Bereiche erneut zu laden. Das spart Rechenleistung und verbessert die Skalierbarkeit des Systems.

Latenz und visuelle Konsistenz: Trotz aller Optimierungen lässt sich eine geringe Verzögerung zwischen realer Umgebungsveränderung und VR-Darstellung technisch nicht vollständig vermeiden. Diese Latenz bleibt jedoch in einem akzeptablen Rahmen und beeinträchtigt die Anwendung nicht spürbar, da das virtuelle Modell persistent erhalten bleibt und kontinuierlich erweitert wird. Das Ergebnis ist ein stabiles, aktuelles 3D-Abbild der realen Arbeitsumgebung, das eine präzise Entscheidungsunterstützung, Annotation und Interaktion im Kontext der Fernkollaboration ermöglicht.

3.4 Anwendungsszenarien

Im Projekt wurden zwei reale Szenarien mit Industriepartnern umgesetzt welche im Folgenden kurz beschrieben werden. Eine vollständige Beschreibung der Anwendungsfälle findet sich in [7].

- **Remote-Fehlerdiagnose an einer industriellen Filteranlage**
Der Experte unterstützte den Techniker vor Ort bei der Analyse und Behebung einer Störung. Obwohl die Mesh-Auflösung der HoloLens nicht ausreichte, um die Ursache visuell eindeutig zu identifizieren, bot die volumetrische Übertragung wertvolle Kontextinformationen – insbesondere zur räumlichen Orientierung im VR-System (Abbildung 2).
- **Wartungseinsatz an einem XR-CAVE-System**
Auch hier wurde die Umgebung volumetrisch erfasst und in Echtzeit übertragen. Die Mesh-Qualität genügte nicht für eine vollständige Fehleranalyse, erleichterte jedoch das räumliche Verständnis und die Kommunikation zwischen den Beteiligten.

In beiden Fällen wurde deutlich, dass volumetrische Daten derzeit nicht als alleinige Informationsquelle ausreichen, jedoch eine sinnvolle Ergänzung klassischer Remote-Werkzeuge darstellen.

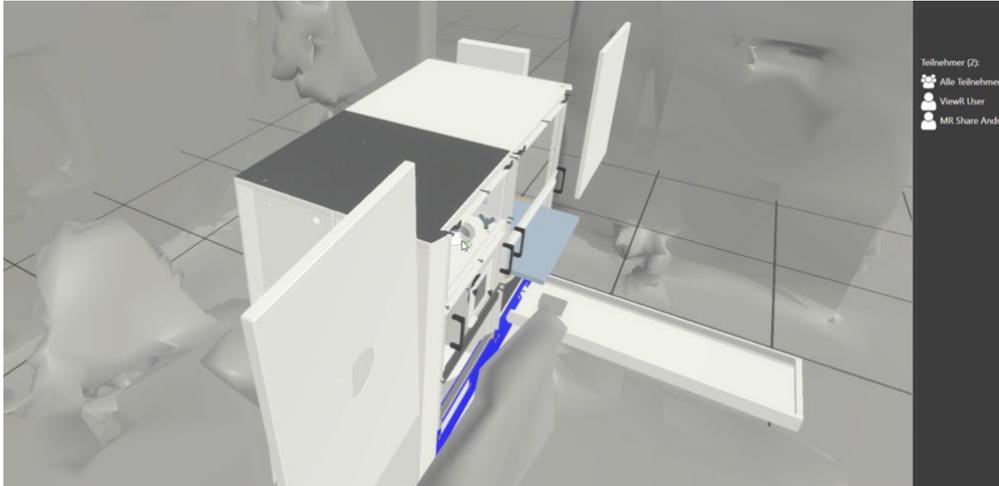


Abbildung 2: 3D-Mesh innerhalb eines volumetrischen Meshs im Anwendungsszenario Filteranlage

4 Diskussion und Ausblick

Die Ergebnisse zeigen, dass volumetrische Echtzeitdaten einen Beitrag insbesondere bei der Bereitstellung räumlicher Kontextinformationen leisten können, die über klassische 2D-Videostreams oder statische 3D-Modelle hinausgehen. Dies ist insbesondere für komplexe Wartungs- und Schulungssituationen von hohem Nutzen.

Technologische Grenzen: Die betrachteten AR-Systeme – einschließlich der Microsoft HoloLens 2 – weisen erhebliche Einschränkungen auf hinsichtlich Rechenleistung, Übertragungskapazität, Mesh-Qualität und allgemeiner Robustheit im industriellen Einsatz.

Die HoloLens 2 erwies sich als das derzeit bestgeeignete Gerät, das ohne externe Komponenten auskommt und eine zuverlässige On-Device-Mesherstellung bietet. Dennoch bilden die begrenzte räumliche Auflösung und Detailtreue der erzeugten Meshs eine erhebliche Hürde für den Industrieeinsatz.

Herstellerseitige Einschränkungen: Ein weiteres Hindernis stellen restriktive Sicherheits- und Zugriffsvorgaben führender Hersteller dar. Insbesondere bei Geräten von Meta und Apple war zum Zeitpunkt des Forschungsprojektes der direkte Zugriff auf Rohdaten wie Tiefensensorwerte oder Kamera-Streams nicht möglich.

Praktische Erkenntnisse aus Anwendungsszenarien: In beiden untersuchten Szenarien – Fehlerdiagnose an einer Filteranlage und Serviceeinsatz in einem XR-CAVE-System – wurde deutlich, dass volumetrische Daten allein derzeit nicht ausreichen, um komplexe Probleme zu lösen. Dennoch ermöglichten sie eine verbesserte räumliche Orientierung für den VR-Experten. Die zusätzliche Kontextualisierung trug dazu bei, die Kommunikation zu erleichtern und die Interaktion zu strukturieren.

Perspektiven für Forschung und Entwicklung: Zukünftige Entwicklungen sollten sich auf die AR-Hardware sowie Software- und Streaming-Architekturen konzentrieren. Besonders die Auseinandersetzung mit folgenden Aspekten erscheint vielversprechend:

- Leistungsfähigere AR-Geräte mit besserer Mesh-Auflösung,
- offene Schnittstellen für direkten Sensorzugriff,
- effizientere inkrementelle Datenübertragung,
- KI-gestützte Mesh-Optimierung und Objektsegmentierung zur Reduktion von Latenzen und Rechenlasten.

5 Fazit

Live Volumetric Streaming bietet ein hohes Innovationspotenzial für die AR/VR-Kollaboration. Der aktuelle Stand der Technik erlaubt bereits sinnvolle Einsätze zur räumlichen Kontextunterstützung. Für eine vollumfängliche Nutzung in der industriellen Praxis sind jedoch gezielte Weiterentwicklungen auf Hardware-, Software- und Systemebene erforderlich. Mit zunehmender Reife der Technologie wird es möglich sein, Live Volumetric Streaming als integralen Bestandteil digitaler Fernunterstützungslösungen zu etablieren.

Die Forschungs- und Entwicklungsarbeit wurde vom Bundesministerium für Bildung und Forschung im Fachprogramm „Zukunft der Wertschöpfung“ und der Fördermaßnahme „Innovative Arbeitswelten im Mittelstand“ im Projekt TeleInteraction-XR (Förderkennzeichen 02L21B550-4) gefördert und vom Projektträger Karlsruhe (PTKA) betreut. Die Verantwortung für den Inhalt der Veröffentlichung liegt beim Autor.

6 Literaturverzeichnis

- [1] Kögel, A. (2025). Wer und was ist wo? - Synchronisation von Virtueller und Erweiterter Realität. In: Bullinger-Hoffmann, A.C. (2025). Telepräsenste Fernunterstützung in der Industrie. aw&I – Wissenschaft und Praxis. DOI
- [2] Wulf-Misaki, D. (2025). Kamera, Licht, Action! - Videostream-basierte AR/VR Kollaboration. In: Bullinger-Hoffmann, A.C. (2025). Telepräsenste Fernunterstützung in der Industrie. aw&I – Wissenschaft und Praxis. DOI
- [3] Kiesel, M., Kögel, A. (2025). Ich habe etwas vorbereitet – 3D-Modell-basierte AR-/VR-Kopplung. In: Bullinger-Hoffmann, A.C. (2025). Telepräsenste Fernunterstützung in der Industrie. aw&I – Wissenschaft und Praxis. DOI
- [4] Microsoft (2023). HoloLens 2 hardware specifications. Online verfügbar unter: <https://learn.microsoft.com/de-de/hololens/hololens2-hardware> (letzter Zugriff: 31.03.2025).
- [5] Hübner, P., Clintworth, K., Liu, Q., Weinmann, M., Wursthorn, S. (2020). Evaluation of HoloLens Tracking & Depth Sensing for Indoor Mapping Applications. Sensors, 20(4), 1021.
- [6] Microsoft Corporation (2022). Spatial Mapping overview. <https://learn.microsoft.com/de-de/windows/mixed-reality/design/spatial-mapping>
- [7] Berger, M., Holz, N., Ruffert, D., Kögel, A., (2025). TeleInterACTION! – VR/AR-Kopplung im Praxistest. In: Bullinger-Hoffmann, A.C. (2025). Telepräsenste Fernunterstützung in der Industrie. aw&I – Wissenschaft und Praxis. DOI

Autoren



Kiesel, Markus

Markus Kiesel ist studierter Maschinenbauingenieur (M.Eng.) Seit 2016 ist er als Unternehmer im Bereich der industriellen grafischen Simulation tätig. Mit seinem Team entwickelt er Anwendungen welche weltweite Verbreitung finden.



Wulf-Misaki, Daniel

Daniel Wulf-Misaki ist studierter Volkswirt (B.Sc.) und hat im Anschluss Mensch-Computer-Interaktion (M.Sc.) studiert. Seit 2022 ist er als Produkt Manager für die oculavis GmbH tätig und ist u. a. für die operative Leitung und Gestaltung von Forschungsprojekten zuständig.