

VOLUMETRISCHE VIDEOS ZUR DATENERHEBUNG

Laborbericht einer spielerischen Annäherung im GameLab
Universität Konstanz

Marko Jevtic / Benjamin Schäfer

Abstract: In unserem Laborbericht zur Umsetzung eines Forschungsprojekts im GameLab Universität Konstanz zeichnen wir unsere Abläufe und Erfahrungen im Umgang mit einem volumetrischen Kamerasystem und den daraus resultierenden Aufnahmen nach. In den beiden Hauptabschnitten geht es um die Arbeiten des Aufbaus, Iterierens, Kalibrierens und Abstimmens des Kamerasystems und das anschließende Aufbereiten der Aufnahmen in einer Grafik-Engine. Der Laborbericht zeigt eine mögliche Herangehensweise – die wir nicht als ‚Good Practice‘-Beispiel ansehen – des Einsatzes von volumetrischen Videos als Werkzeug zur Datenerhebung auf. Abschließend geben wir noch einen kurzen Ausblick zur möglichen Analyse volumetrischer Videos, den ethischen Dimensionen und der grundlegenden Anwendbarkeit volumetrischer Videos als Instrument zur Datenerhebung.

Keywords: Laborbericht, volumetrische Videos, Datenerhebung, Postproduktion, Projektarbeit

EINLEITUNG

Die Idee, im GameLab an der Universität Konstanz eine wenig verbreitete Technologie zur Datenerhebung anzusiedeln und ergebnisoffen auszutesten, kam bereits während der ersten Planungsphase zur Sprache. Das grundsätzliche Ziel, Infrastruktur zu schaffen, die bisher in dieser Art und Weise an der Universität Konstanz nicht existiert, war ein zentrales Anliegen seitens der Universität und im Sinne der Profilbildung des GameLab. Volumetrische Videos – so unsere damalige Vermutung – erlauben eine umfangreiche, räumliche Erfassung einer sozialen oder ludischen Interaktion. Dabei können in der Analyse von volumetrischen Aufnahmen nahezu alle Blickwinkel eingenommen werden. Dadurch entfällt das gezielte Platzieren von Kameras, die bestimmte Perspektiven einfangen, die dann eventuell die gewünschten Daten liefern.

Natürlich bringt eine neue Technologie Ungewissheiten, Leerstellen, Fragen und Hürden mit sich. Dementsprechend verlangt der Einsatz volumetrischer Videos neben dem notwendigen technischen Know-how ebenso eine spielerische Offenheit im Erkunden und dem Einsatz der Technologie. Ebendiese spielerische Exploration, die wir in der Vorbereitung und Umsetzung des Forschungsprojekts *Mixed-ReAlity ReflectionS – Design eines Trainingsmoduls*¹ (kurz: MARS-D) von Prof. Dr. Stefanie Findeisen geleistet haben, dokumentieren wir in diesem Laborbericht.

Unser Ziel beim Verfassen dieses Laborberichts besteht im Offenlegen und Festhalten unserer Erfahrungen, die wir im Verlauf des Projekts gesammelt haben. Dabei illustrieren wir die einzelnen Stationen beim Einrichten des Kamerasystems mit aussagekräftigen Bildern und beschreiben, was wir wie und weshalb gemacht haben. Unser Bestreben zielt nicht darauf ab, ein ‚Good Practice‘-Beispiel niederzuschreiben, sondern unsere Herangehensweise als Grundlage für ein mögliches Vorgehen festzuhalten. Hier gibt es verschiedene Dienstleister, die solche Aufnahmesysteme professionell anbieten und wesentlich besser einrichten können als wir es

1 Weitere Informationen zum Projekt: <https://www.wiwi.uni-konstanz.de/jun-prof-dr-stefanie-findeisen/forschung/mixed-reality-selbstreflexion-von-unterrichtshandeln-design-eines-trainingsmoduls-mars-d/>.

mit unserem aktuellen System (Stand: Ende 2024) könnten – das entsprechende Budget vorausgesetzt.

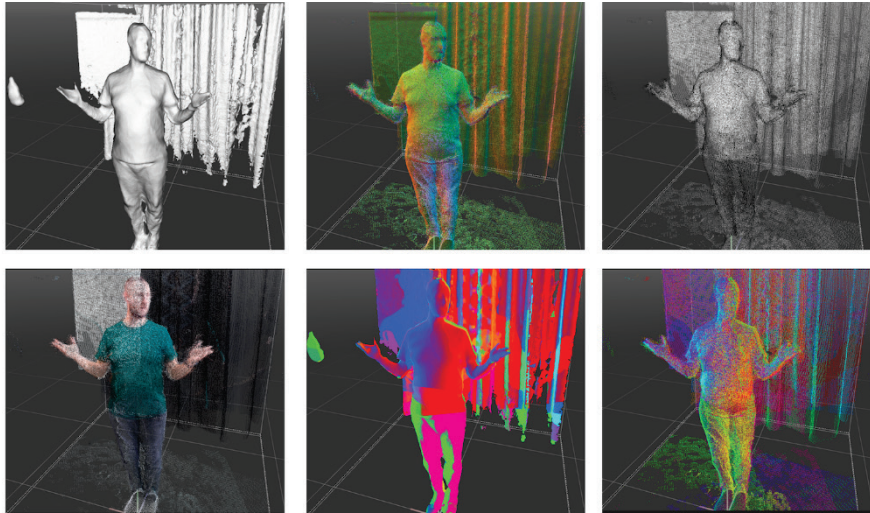


Abb. 1: Collage der verschiedenen Visualisierungen der Kameradaten.

Wir teilen unseren Bericht in vier Abschnitte auf. Im ersten Teil geben wir einen kurzen Einblick in das GameLab Universität Konstanz, eine grobe Einführung in das Forschungsprojekt MARS-D und eine allgemeine Einführung in volumetrische Videos. Dieser grundlegende Teil vermittelt den Rahmen, in den wir unseren Laborbericht einbetten. Im zweiten Teil konzentrieren wir uns auf unsere Erfahrung des Aufbaus, Inbetriebnehmens und Feinjustierens unseres konkreten Kameraaufbaus. Hier geht es um die technische Infrastruktur, den iterativen Aufbau der Kameras, sowie um die Aufnahmesoftware DEPTHKIT (Scatter 2012) und deren Funktionen. Unsere Erfahrungen in der Postproduktion und der Aufbereitung der volumetrischen Aufnahmen bilden den dritten Teil unseres Beitrags. Dabei geht es um das Einbinden der Videos in die Grafik-Engine UNITY (Unity Technologies 2005), das Erstellen einer virtuellen Umgebung, in der die Videos ablaufen, sowie um die Bereitstellung der Aufnahmen als Virtual Reality-Umgebung. Im abschließenden vierten Teil geben wir einen Ausblick auf die Möglichkeiten, die volumetrische Aufnahmen mit sich bringen. Fragen der Analyse, der Forschungsethik und dem generellen Umgang mit den erhobenen, räumlichen Daten bilden in diesem Teil die Dreh- und Angelpunkte.

Wir sehen unseren Laborbericht als einen reinen Erfahrungsbericht an, der unsere Vorgehensweise widerspiegelt und diese weniger in Beziehung zu bereits vorhandener Literatur zum Umgang mit volumetrischen Videos und Aufnahmesystemen setzt. Ebenso sehen wir unseren Text nicht als Methodentext an, der ein replizierbares Vorgehen darstellt. Zudem wurden alle verwendeten Bilder und Screenshots von uns erstellt.

1. DIE CORE FACILITY „GAMELAB UNIVERSITÄT KONSTANZ“

Das *GameLab Universität Konstanz* avancierte 2022 zu einer geisteswissenschaftlichen Core Facility, einer zentralen Einrichtung für die Bereitstellung von Forschungsinfrastruktur. Damit öffnete sich das GameLab in Richtung Forschung. So unterstützt das GameLab nicht nur Lehrveranstaltungen oder Workshops, sondern ebenso Datenerhebungen in Forschungsprojekten und größeren Verbundprojekten wie Sonderforschungsbereiche. Neben Konsolenarbeitsplätzen in der Bibliothek und einem mobilen Gaming-Classroom stellt das GameLab weitere Arbeitsplätze für Forschende im Büro bereit. Dort (Abb. 2) befindet sich ebenfalls unser Forschungssetup für kleinere, einzelne Datenerhebungen. In unserem vielfältig nutzbaren Forschungsraum setzen wir auf knapp 50 Quadratmetern umfangreichere Experimente um, wie die in diesem Bericht beschriebenen volumetrischen Aufnahmen.



Abb. 2: Das Büro des GameLab mit Arbeitsplätzen und dem Forschungssetup.

Komplementär unterstützen wir als GameLab Forschende bei der Projektplanung und Antragsstellung, wenn es um die methodische Einbindung von Spielen geht oder spezielle Spiele gesucht werden, um einen Aspekt der Forschung betrachten zu können. Zusammenfassend sehen wir das GameLab und dessen Arbeit als wissenschaftsunterstützend an, allerdings aus dem wissenschaftlichen Bereich heraus. Unabhängig davon, ob technische Infrastruktur benötigt wird, Beratung im Vordergrund steht, wir Forschende in der Lehre unterstützen oder innerhalb der Universität an andere Stellen verweisen, versuchen wir im GameLab Forschungsprojekte zu ermöglichen und bestmöglich umzusetzen.

2. DAS MARS-D PROJEKT

Eines dieser durch das GameLab ermöglichte Projekte heißt *Mixed-ReAlity ReflectionS – Design eines Trainingsmoduls* und wird von Tenure-Track-Professorin Dr. Stefanie Findeisen durchgeführt. Stefanie Findeisen erforscht an ihrem Lehrstuhl für Wirtschaftspädagogik Auswirkungen von Lernprozessen in der Berufsbildung und deren Gestaltung. Ein Schwerpunkt findet sich daher in der Lehrer*innenausbildung, um die es sich im MARS-D-Projekt dreht.

Unterrichtssimulationen stellen ein anerkanntes Mittel zur Reflexion für Lehramtsstudierende dar. Auf Videoaufnahmen sehen Studierende, wie sie sich in ihrer selbst vorbereiteten Lerneinheit verhalten haben: wie war der Blickkontakt, wie die Gestik, Mimik und Ansprache usw. In einer solchen Unterrichtssimulation mimen andere Studierende die Schüler*innen. Um typische (Konflikt-)Situationen aus dem Klassenzimmer in die Lerneinheit zu bringen, bekommen die simulierten Schüler*innen mitunter Aufgaben zugeteilt. Darüber üben die Studierenden in der Rolle der Lehrkraft den Umgang mit Störungen sowie dem generellen Unterricht und können ihre Reaktion im Video reflektieren.

MARS-D nimmt diese Situation und fragt, ob das Erstellen von volumetrischen Videoaufnahmen der Lehrperson eine bessere oder schlechtere Reflexion erlaubt. Dabei sieht das Forschungsdesign neben der Kooperation mit dem GameLab zum Erstellen der volumetrischen Videos

eine umfangreiche Inputphase zur Unterrichtsinteraktion vor sowie eine Kontrollgruppe, die ihre Unterrichtssimulationen weiterhin per Videoaufnahme reflektiert. Zusätzlich kommen in diesem Projekt Studierende der Universität Konstanz mit Studierenden der Universität Göttingen zusammen. Dazu arbeitet Stefanie Findeisen mit Professorin Dr. Viola Deutscher zusammen, die sich in ihrer Professur für Wirtschaftspädagogik und digitales berufliches Lernen ebenfalls mit der Lehrer*innenausbildung und der Qualität der Ausbildung beschäftigt.

Das von der Stiftung Innovation in der Hochschullehre geförderte Projekt läuft bereits seit April 2024 mit einer Förderdauer bis März 2026. Die Projektvorbereitungen seitens des GameLab fanden im März und August 2024, die ersten Aufnahmen mit Studierenden im November 2024 statt. Weitere Aufnahmen sind für das Sommersemester 2025 geplant.

3. VOLUMETRISCHE VIDEOS

Volumetrische Videos nehmen in Erweiterung zu gewöhnlichen Videoaufnahmen Tiefendaten der gefilmten Szene auf. Bei Verwendung mehrerer Kameras entsteht so eine räumliche Repräsentation der Szene bzw. des Volumens. Dazu werden die Bilder und Tiefendaten aller Kameras per so genanntem Stitching zusammengebracht: Die Daten werden an ihren sich überlagernden Kanten übereinandergelegt, woraus ein dreidimensionales Video des gefilmten Volumens entsteht. Dieses kann im Nachhinein über eine Grafik-Engine aufbereitet werden. Damit werden die Aufnahmen in dem Sinne begehbar, dass diese entweder am Computer per Controller oder in VR durchschritten werden können. Dadurch entsteht die Möglichkeit, unabhängig von den Kamerapositionen verschiedene Perspektiven auf die gefilmte Szene einzunehmen.

4. BERICHT AUS DEM LABOR

In den folgenden Abschnitten gehen wir auf die Aspekte ein, die unsere direkte Arbeit im Forschungsraum des GameLab betreffen. Dabei fokussieren wir den iterativen und spielerischen Ablauf unserer Herangehensweise, legen darüber hinaus ebenso Wert darauf, die groben Schritte, die

beim Aufbau unseres Kamerasystems zum Tragen kommen, nachvollziehbar zu machen. Wir arbeiten in unserem Laborbericht hauptsächlich mit Fotografien aus unserem Forschungsraum und Screenshots aus den verschiedenen Softwarelösungen. In unseren Texten ordnen wir die Abbildungen in den Ablauf ein und geben ihnen Kontext.

4.1. UNSERE TECHNISCHE INFRASTRUKTUR

In unserem Aufbau haben wir zehn Kameras an einen Computer (siehe Abb. 3) angeschlossen, die zusammen das Volumen aufnehmen. Grundsätzlich existieren verschiedene Abstufungen von volumetrischen Aufnahmesystemen mit ein oder zwei Kameras bis hin zu 100 oder mehr Kameras. Dabei ändert sich die benötigte Infrastruktur, sobald mehr als zehn Kameras in einem Verbund betrieben werden. Aufgrund der explorativen Ausrichtung haben wir das volumetrische Kamerasystem im GameLab auf zehn Kameras beschränkt, womit wir gleichzeitig die Kosten und den administrativen Aufwand gering halten konnten.



Abb. 3: Umgebauter Aktenschrank mit Aufnahmecomputer, zusätzlicher Kühlung und Anschlusskabeln der volumetrischen Kameras.

Alle zehn Kameras sind über spezielle USB-Erweiterungskarten an den Computer angeschlossen und untereinander über Klinkekabel zur Synchronisation verbunden. In unserem Setup kommen die in Abb. 4 zu sehenden *Microsoft Azure Kinect DK* Kameras zum Einsatz, die, wie der Name bereits vermuten lässt, eine Weiterentwicklung der Kinect-Kameras der Spielkonsole *Xbox 360* darstellen. Die Kameras werden per USB-Kabel an den Computer angeschlossen und liefern daher kein Bildsignal, sondern ein Datensignal, das die Daten mehrerer Kamerabilder, der Lage der Kamera im Raum und (bei Bedarf) einen Audiostream beinhaltet. Insgesamt werden so zwei Aufnahmen gleichzeitig produziert: zum einen nehmen die Kameras ein Farbbild auf, zum anderen ein Tiefenbild (vgl. Abb. 1 mit unterschiedlichen Visualisierungen der Kameradaten). Während ersteres bei der Aufnahme zu einer Textur umgewandelt wird, liefert das Tiefenbild Entfernungsdaten, in dem es Infrarotlicht aussendet und mittels eines Time-of-Flight-Sensors die Zeit misst, die das Infrarotlicht vom Aussenden über die Reflexion bis zum Einfangen braucht, um so die Entfernung der gefilmten Objekte zu ermitteln.



Abb. 4: *Microsoft Azure Kinect DK* Kameras, die zur Aufnahme der volumetrischen Videos verwendet werden. Links: Vorderseite. Rechts: Rückseite mit Synchronisations-Kabeln (oberes und unteres Kabel), USB-C-Kabel und Stromkabel (schwarz).

Um die Kameras dementsprechend flexibel aufhängen zu können, haben wir zusammen mit den wissenschaftlichen Werkstätten – einer weiteren Core Facility im wissenschaftsunterstützenden Bereich der Universität Konstanz – eine Deckentraverse planen und anbringen lassen, die verschiedene Aufgaben übernimmt. Ganz grundlegend verlegt die Traverse alle Kabel an die Decke, sodass der gesamte Boden des Forschungsraumes ohne Stolperfallen nutzbar bleibt. Zum Aufhängen der Kameras verwenden wir Doppelscheren aus der Studiotchnik (Abb. 5), die von den wissenschaftlichen Werkstätten angepasst wurden. Die Doppelscheren erlauben ein Abhängen der Kameras bis fast auf Bodenhöhe. Das ermöglicht es uns, die Tiefenkameras konisch in verschiedenen Höhen, um das aufzunehmende Volumen herum aufzuhängen.



Abb. 5: Deckentraverse und Doppelscheren an denen die Kameras von der Decke abgehängt werden.

4.2. VERSUCHSAUFBAU

Unser erster Aufbau der Kameras orientierte sich an den Vorschlägen aus der Dokumentation² der Aufnahmesoftware DEPTHKIT. Hier verwendeten wir neun von zehn Kameras, die wir in drei Dreiecke auf drei unterschiedliche Höhen aufteilten und entsprechend an der Traverse befestigten (siehe Abb. 6). Neben dem Aufhängen der Doppelscheren und der Kameras galt es, alle Kabel ordentlich über die Traverse zum Computer bzw. zwischen den Kameras zu verlegen. Insgesamt verlegten wir 30 Kabel mit einer Gesamtlänge von 360 Metern, worunter zehn Glasfaserkabel fallen, die mit besonderer Sorgfalt verlegt werden müssen, damit sie aufgrund eines zu engen Biegeradius nicht (an-)brechen. Um nichts falsch zu verlegen, führten wir auf einem digitalen Whiteboard Protokoll (siehe Abb. 7) und beschrifteten die Klinkenkabel zur Synchronisation gemäß ihrer Anschlussrichtung.

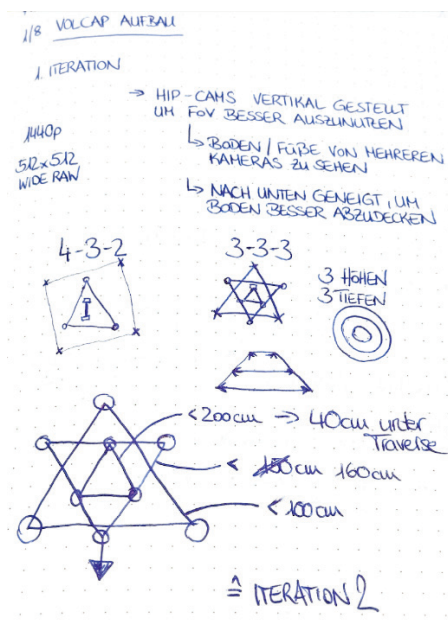


Abb. 6: Auszug aus unseren Notizen bei der Anordnung der Kameras.

2 Siehe <https://docs.depthkit.tv/docs/sensor-configurations> (Stand: 23. Januar 2025).

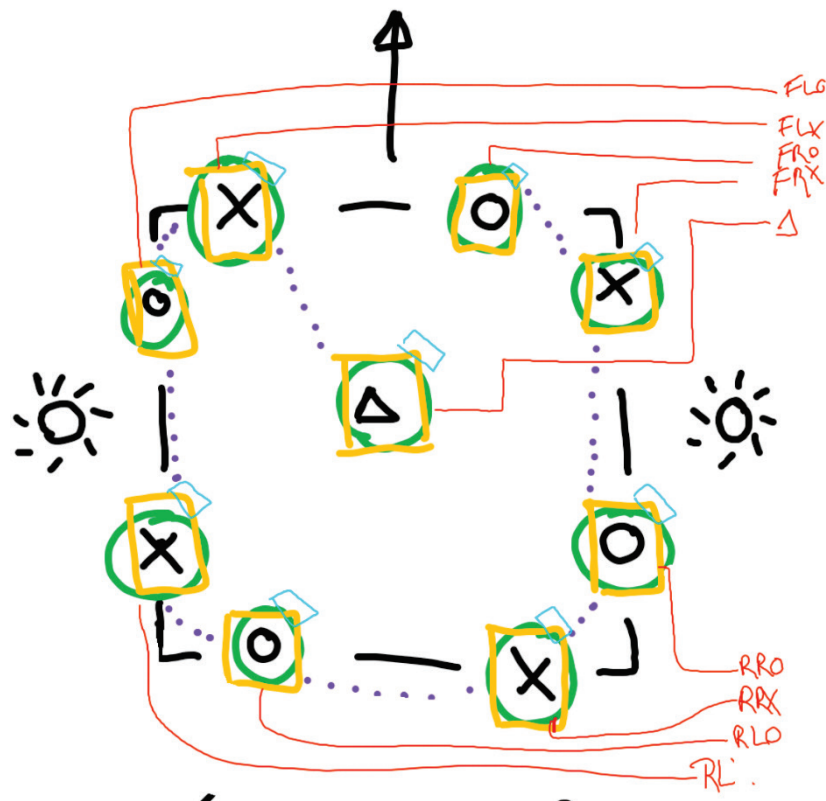


Abb. 7: Screenshot zur Verkabelung. ○, X und △ stehen für Kameras, die auf unterschiedlichen Höhen angebracht sind. Grüne und gelbe Markierungen zeigen, dass diese Kameras bereits mit USB-C und Stromkabel verbunden wurden. Die violett gepunkteten Linien stehen für die Synchronisationskabel.

Bis zu unserer finalen Iteration durchschritten wir drei verschiedenen Hängen und Einstellungen der Kameras (siehe Abb. 8), bis wir ein sehr gutes Ergebnis mit einem Halbkreis aus sieben Kameras auf Kopfhöhe und einem Dreieck aus drei Kameras auf Hüfthöhe erreichten. Jede Iteration wurde von einer erneuten Kalibrierung und dem Einrichten des Studilichtes begleitet.

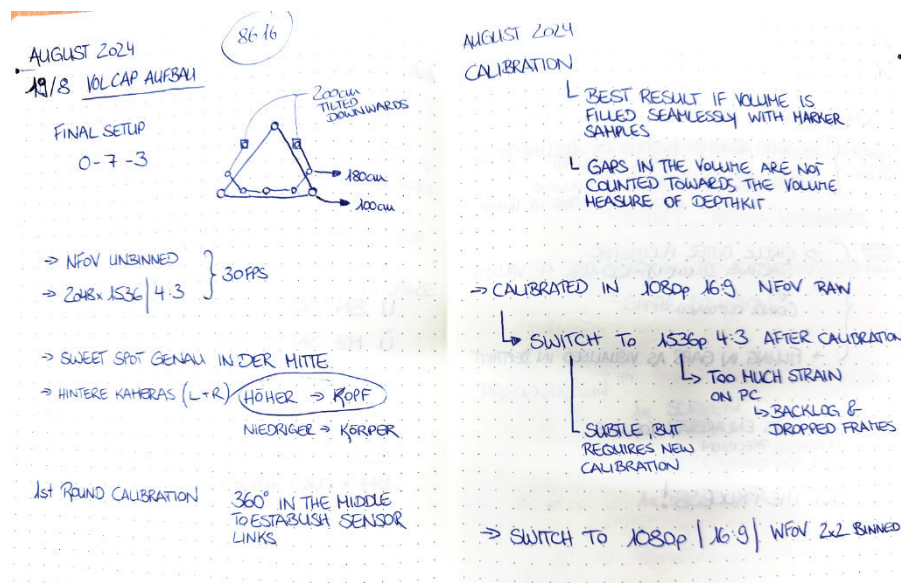


Abb. 8: Auszug aus unseren Notizen zur finalen Anordnung der Kameras (links). Rechts stehen Notizen zu verschiedenen Einstellungen und deren Auswirkungen auf die Aufnahmen.

Dabei kommen beim Platzieren der Kameras im Raum mehrere Faktoren zum Tragen: Allen voran spielt die Entfernung zum Mittelpunkt des gefilmten Volumens eine entscheidende Rolle. Der Idealbereich der Kameras liegt zwischen einem halben Meter und – je nach Sichtfeldeinstellung – circa zwei Metern. Dadurch ergibt sich, dass die Kameras idealerweise kreisförmig um den Mittelpunkt des Volumens angeordnet werden. Gleichzeitig gilt es, darauf zu achten, dass sich die Sichtfelder der Kameras überlappen. Die Sichtfelder bzw. -winkel der Kameras wiederum hängen – je nach eingestelltem Sichtfeldmodus – davon ab, ob sie horizontal oder vertikal aufgehängt wurden. Dieses Überlappen der Kamerabilder wirkt sich auf die Qualität des Stitching aus. Durch unsere spielerischen Herangehensweise haben wir durch Ausprobieren immer besser verstehen können, wie die Kameras und die Software gemeinsam die Bilddaten verarbeiten und wie wir durch die Aufhängung der Kameras das Ergebnis beeinflussen konnten.

4.3. DEPTHKIT

Als Aufnahmesoftware verwenden wir DEPTHKIT des Entwicklers Scatter. Im Vorfeld und während der Planung des Aufnahmesystems probierten wir verschiedene Softwarelösungen aus, darunter Open-Source-Programme sowie kostenfreie und kostenpflichtige Lösungen. Da die Aufnahme von volumetrischen Videos lediglich eine kleine Nische darstellt, gab und gibt es kein breites Angebot und Open-Source-Lösungen besitzen keine Support-Community, die sich der (Weiter-)Entwicklung widmet. Scatter stellt hier mit DEPTHKIT eine Ausnahme dar, da sich Scatter im Bereich der Filmschaffenden etablieren konnte. Aktuell nehmen wir den hohen Lizenzpreis in Kauf, da DEPTHKIT für uns eine funktionale, einfach zu bedienende und zuverlässige Softwarelösung darstellt. Dennoch behalten wir im Blick, wie sich andere Softwarelösungen entwickeln und werden zu Beginn des Jahres 2025 das umstrukturierte Angebot des nun chinesischen Herstellers Djinn Technologies Limited SPATIALSCAN3D (2022) ausprobieren.

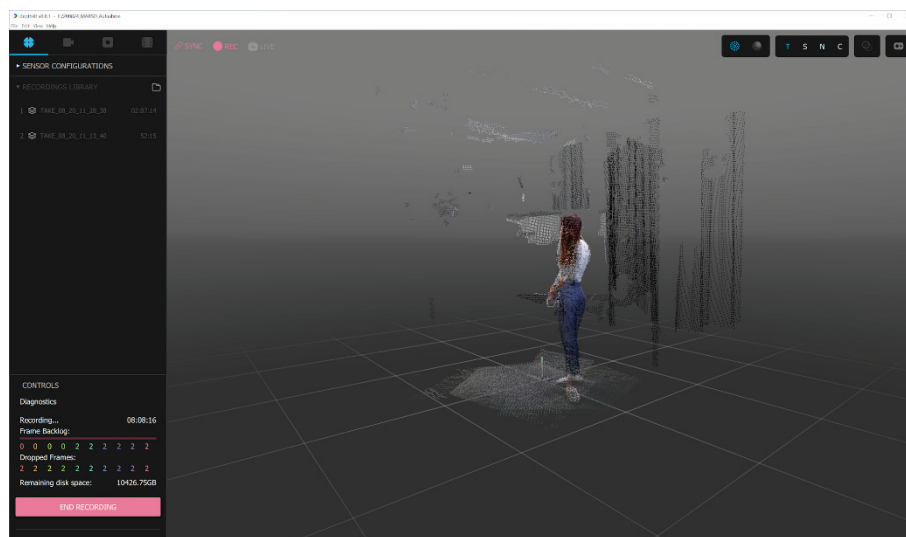


Abb. 9: Aufnahmesituation in DEPTHKIT während einer Aufnahme. Rechts kann in Echtzeit die Aufnahme direkt in 3D verfolgt werden. Links unten wird der Aufnahmeprozess visualisiert.

4.4. KALIBRIERUNG

Ein entscheidender Prozess im Aufbau eines volumetrischen Aufnahmesystems liegt in der Kalibrierung der Kameras. Hierbei lernt das Programm, wo sich die Kameras im Raum befinden und wie sie zueinander ausgerichtet sind. Für den Kalibrierungsprozess benötigt die Software QR-Code-basierte Marker, die von den Kameras erkannt werden und als Ankerpunkte im Raum dienen, um das Volumen aufzubauen. Dazu haben wir die Marker an einem rollbaren Stativ befestigt, um sie einfach durch das zu erfassende Volumen bewegen zu können (Abb. 10). Dabei erfasst die Software jede Position der Marker im Raum und errechnet daraus die Positionen der Kameras zueinander, die Größe des Volumens, sowie die erwartete Qualität und räumliche Präzision der Aufnahme.

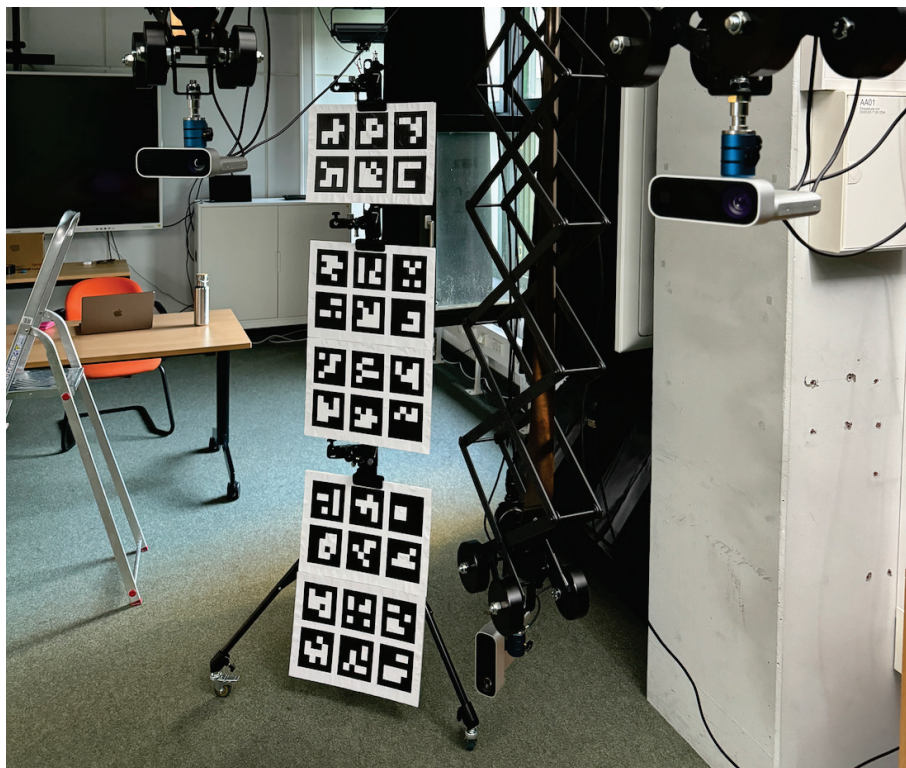


Abb. 10: Unser rollbares Stativ mit mehreren Markern zur Kalibrierung.

Die von uns verwendete Software DEPTHKIT übernimmt den Kalibrierungsprozess nativ und ermöglicht es uns, einen Aufbau mit zehn Kameras in

circa 15 Minuten zu kalibrieren. Im Zuge unserer Testläufe probierten wir verschiedene Kalibrierungsabläufe aus. Unseren Erfahrungen nach liefern konzentrische, größer werdende Kreise um den geschätzten Mittelpunkt des Volumens das beste Ergebnis (Abb. 11). Dabei hilft es der Software keine Lücken zwischen den Positionen der Marker zu lassen, da so das Volumen besser eingerichtet und aufgenommen werden kann.

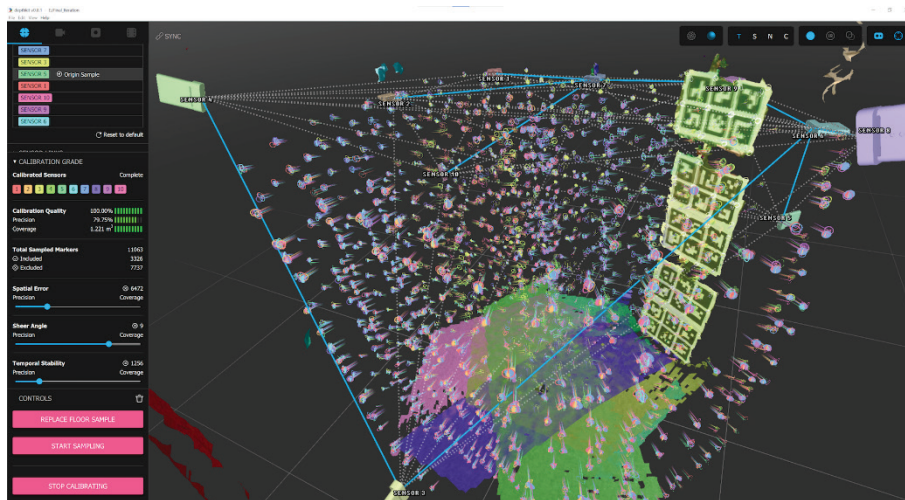


Abb. 11: Screenshot aus der Aufnahmesoftware DEPTHKIT während einer Kalibrierung. Die Farbe der Punkte gibt Aufschluss darüber, von welcher Kamera der Marker gesehen wurde und aus welcher Richtung.

4.5. AUFNAHME

Die Aufnahmesituation gestaltet sich unkompliziert. Im Projekt MARS-D nehmen wir in simulierten Unterrichtssituationen angehende Lehrpersonen als volumetrische Videos auf. Gleichzeitig filmen wir mit Videokameras die von Studierenden dargestellten Schüler*innen, die auf Fragen der Lehrperson reagieren (siehe Abb. 12). Jede Unterrichtssituation wird von einer Präsentation begleitet, deren Ablauf per OPEN BROADCASTER SOFTWARE STUDIO (Bailey 2012) via Bildschirmaufnahme auf einem separaten Computer aufgenommen wird.



Abb. 12: Foto einer Unterrichtssituation während einer Aufnahme.

Um ein gutes Ergebnis zu erhalten, erklären wir den Studierenden, dass sie darauf achten, sich nicht zu weit von der auf dem Boden markierten Mitte des Volumens zu entfernen. Ebenso achten wir genau darauf, dass die Kameras nicht berührt werden, da wir sonst erneut kalibrieren müssen. Den Start der eigentlichen Aufnahmen markieren wir durch einen lauten Ruf zur Synchronisation aller Kameraaufnahmen. Der Einsatz einer Filmklappe oder ähnlicher Hilfsmittel führt zu einem längeren volumetrischen Video und das Hilfsmittel kann aufgrund seiner Größe im volumetrischen Video nicht erkannt werden. Ein Problem, das wir im Ausblick noch einmal ansprechen.

Während einer Aufnahme halten wir uns idealerweise komplett zurück und versuchen nur in Ausnahmefällen zu interferieren. Sobald die Unterrichtssimulation zu ihrem Ende gekommen ist, beenden wir alle Aufnahmen, benennen diese entsprechend ihrem Inhalt und tragen nach und nach alles an einem zentralen Speicherplatz zusammen – in unserem Fall dem zentralen Storage Server des GameLab.

4.6. EXPORT

Bevor wir die volumetrischen Videos für die Postproduktion und Analyse verwenden können, müssen die Rohdaten aus DEPTHKIT exportiert werden. Je nach beabsichtigter Verwendung bietet DEPTHKIT verschiedene Exportvarianten an. Für unsere Zwecke kommt die Variante *Combined Per Pixel Video* in Frage, die die Rohdaten so verpackt, dass wir die Videos direkt in die Game-Engine UNITY einbinden können. In dieser Variante bekommen wir einerseits die Möglichkeit die Bitrate, den Codec und Farbraum zu bestimmen oder andererseits die für uns sinnvollere Option des Presets für das autarke ‚Meta Quest 2‘-VR-Headset zu wählen, das im Projekt MARS-D von den Studierenden zur Rezeption verwendet wird.

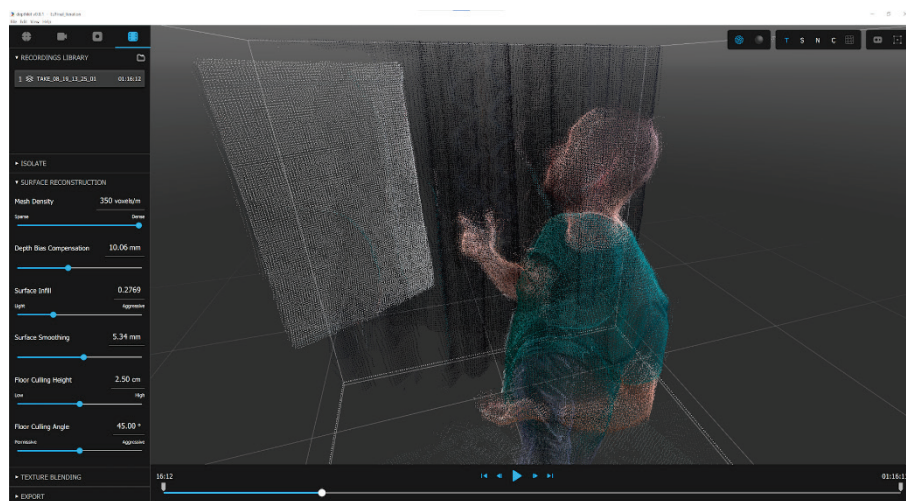


Abb. 13: Exportmodus in DEPTHKIT. Links können die Parameter für den Export angepasst werden. Bei den angezeigten Einstellungen geht es um die Komplexität der Oberflächen.

Zwischen der Aufnahme und dem Playback des aufgenommenen volumetrischen Videos isoliert DEPTHKIT automatisch den Boden sowie weiter entfernte Objekte, die sich außerhalb des eigentlichen Volumens befinden. Diese Grenzen bestimmen gleichzeitig, welche Teile der Aufnahme Teil des exportierten Videos werden. Dementsprechend passen wir diese räumlichen Begrenzungen unseren Begebenheiten an, um nur den Raum,

den die Studierenden ausgenutzt haben, in den exportierten Videos abzubilden. Dabei achten wir darauf, dass gerade bei ausladenden Gesten etwa die Arme den festgelegten Bereich nicht verlassen und somit abgeschnitten werden würden.

5. AUFBEREITUNG DER AUFNAHMEN UND POSTPRODUKTION

5.1. UNITY

Das von DEPTHKIT nativ unterstützte Programm zum Import und der Aufbereitung der volumetrischen Daten ist die Game Engine UNITY, die sich sowohl in der Entwicklungsphase als auch als nutzbares Endprodukt als einsteigerfreundlich und universell zugänglichste Lösung offenbarte. UNITY ist kostenfrei nutzbar und hat weiterhin eine Vielzahl von Installationspaketen sowohl von UNITY selbst als auch großen Firmen wie VR-Entwicklerfirmen wie *Meta*, Endnutzern oder eben DEPTHKIT. Dazu kommt, dass dank einer aktiven Community sich Hilfe bei fast jedem Problem finden lässt, sei es in Textform oder als YOUTUBE-Video. Zudem unterstützten uns kostenfreie KI-Programme wie CHATGPT beim Schreiben von komplexen, individuellen UNITY-Skripts.

So ist es auch für Laien ohne einschlägige Erfahrung mit Scripting, Game Design, oder dem Erstellen von Benutzeroberflächen möglich, mit UNITY zu arbeiten. Für unser Projekt wurde schnell klar, dass nicht nur fundamentales Game Design benötigt wird, etwa zum Erstellen der virtuellen Welten, sondern auch Video- und Audioediting, sowie stellenweise Erfahrung mit Bildbearbeitungsprogrammen. Besonders das Zurechtschneiden von zusätzlichen, lehrunterstützenden Videos – bspw. Videos von der Mitarbeit und den Reaktionen der Schüler*innen – und der Synchronisierung von diesen mit den volumetrischen Aufnahmen und den jeweiligen Tonspuren offenbarte sich als größerer Teil der Arbeit als ursprünglich gedacht. Diese Dateien in UNITY einzuspeisen, ist problemlos möglich, doch zur Bearbeitung von Audio-, Bild- oder Videodateien sind die Möglichkeiten der

Game Engine zu rudimentär. Hierfür wurden die einschlägigen Adobe-Programme verwendet, insbesondere PREMIERE PRO und PHOTOSHOP (siehe Abb. 14).



Abb. 14: Die gleichzeitige Bearbeitung in ADOBE PREMIERE PRO von Video- und Audiodateien für die synchrone Einarbeitung mehrerer Objekte in UNITY.

5.2. AUFBEREITUNG DER DATEN IN UNITY

Die volumetrischen Videos in UNITY zu importieren, gestaltet sich weitaus nutzerfreundlicher, als es die Komplexität der Daten und ihrer Erstellung vermuten lassen. Nach einer manuellen Installation der von DEPTHKIT zur Verfügung gestellten Installationspakete müssen die richtigen 'Prefabs' – also vorgefertigte Materialien oder Skripte für die Game Engine – per simplem Drag-and-Drop in Unity eingefügt werden. Die Daten aus dem Export der volumetrischen Aufnahmen teilen sich in drei Formaten auf: ein Referenz-Bild, die Videodatei mit Aufnahmen aus allen Kameraperspektiven sowie eine Textdatei bestehend aus kodifizierten Metadaten über die sich die zehn separaten Perspektiven in ein dreidimensionales Hologramm bzw. Video 'übersetzen' lassen. Diese drei Dateien werden ebenfalls via Drag-and-Drop in den DEPTHKIT-Prefab eingefügt. Da es sich hier um drei verschiedene Dateiformen handelt (Bild, Video und Text) und jedes Feld nur die passende Form akzeptiert, ist der einzig mögliche Fehler, dass man die Dateien aus verschiedenen Aufnahmen untereinander vermischt und

der volumetrische Clip somit beim Abspielen falsch zusammengesetzt wird.

Die Möglichkeiten der Kontrolle, Nachbearbeitung und Feinjustierung der volumetrischen Daten sind nach dem Export immens. Innerhalb von UNITY kann beispielsweise jede Kameraperspektive neu eingesehen werden und für jede der zehn volumetrischen Perspektiven der Beitrag zur Farb- und Geometrieerstellung angepasst, abgeschwächt oder sogar komplett ausgeschaltet werden (siehe Abb. 15). Wenn sich bei der Einarbeitung der volumetrischen Daten offenbart, dass die vortragende Person sich häufiger in eine bestimmte Richtung dreht (etwa Richtung Tafel), oder dass eine Kameraperspektive wenige Millisekunden langsamer aufgenommen hat und damit einen asynchronen Effekt erzeugt, können solche Probleme direkt in UNITY abgemildert oder gelöst werden. Vereinfacht wird dieser Prozess durch eine optional darstellbare *Heatmap*, die wahlweise die Dichte oder die Quelle der Daten als Textur auf die volumetrische Aufnahme projiziert.

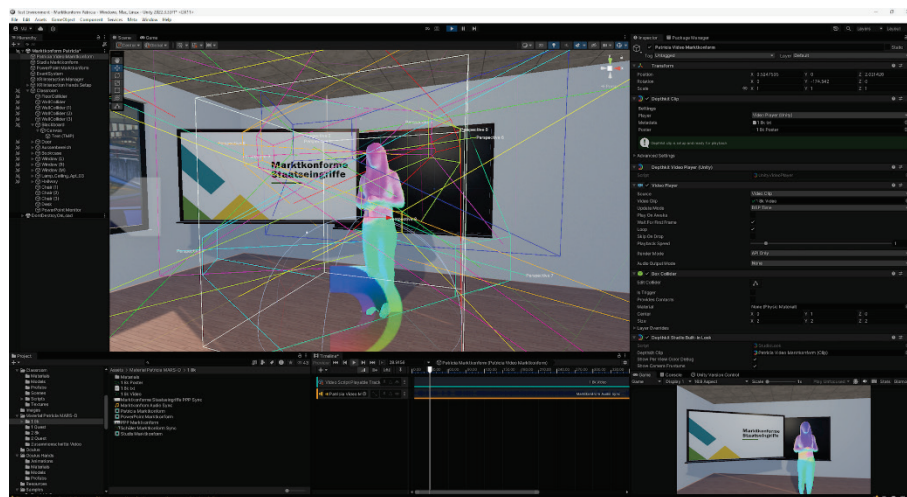


Abb. 15: Jede Kameraperspektive wird auf die volumetrische Aufnahme projiziert mit nachvollziehbarem Ursprungsort sowie dem jeweiligen Anteil am Gesamtbild.

Die Problemstellen der Aufnahmen, wie auch die Ergebnisse der Nachbearbeitung, sind also unmittelbar visuell dargestellt und können somit in der virtuellen Umgebung im buchstäblichen Vorbeigehen beobachtet und

nachjustiert werden. Denn insbesondere, wenn sich die entwickelnde Person per Headset selbst in die eigens erstellte virtuelle Umgebung begibt und der volumetrischen Aufnahme zur Problemfindung und -lösung *auf Augenhöhe* begegnet, ist die gleichzeitige Beobachtung und Korrektur der Daten am produktivsten und einfachsten zu bewerkstelligen.

In diesem Prozess wird deutlich, dass eine Game Engine nicht nur im fertigen Produkt Vorteile durch die freie, multiperspektivische Beobachtbarkeit bietet, sondern ebenso in der Erstellung und Komposition der Aufnahmen selbst. Die Vorgänge der Einarbeitung, Feinjustierung und Beobachtung der volumetrischen Daten finden hier simultan statt und sind schon aus praktikablen Gründen oft auch dreidimensionale, immersive und quasi-ludische Abläufe.

5.3. ERSTELLEN EINER VIRTUELLEN UMGEBUNG

Bei der Erstellung der virtuellen Umgebung liegt unser Ziel darin, ein immersives und glaubhaftes Ambiente zu erzeugen, das nicht ablenkt oder die Performanz des fertigen Produktes beeinträchtigt. Wir haben uns entschieden, ein generisches Klassenzimmer zu konstruieren, das mehr die allgemeine Idee eines Unterrichtsraums kommuniziert statt eines konkreten Klassenzimmers aus einer bestimmten Schule.

Zunächst wurde diese Entscheidung aus reiner Zweckmäßigkeit getroffen: Das Einfügen der volumetrischen Videos in einem komplett leeren virtuellen Raum macht es unmöglich, etwa die Größe des eingefügten Videos richtig einzuschätzen, weshalb schnell klar wurde, dass Wände, Decken und ein Boden benötigt werden. Dann wurde bei der Bearbeitung in der Engine deutlich, dass das digitale Simulakrum einer Person in einer großen, weißen, leeren ‚Zelle‘ isoliert wirken kann. Es wurde schon im frühesten Stadium dieses Entwicklungsprozesses deutlich, wie diese hinzugekommene Dreidimensionalität und Nahbarkeit immersive und sogar affektive Qualitäten kreieren können. Der leere Raum musste also zunächst mit Fenstern und Türen erweitert werden, um eine gewisse Freiheit der Räumlichkeiten und Freiwilligkeit der in ihnen stehenden Person zu kommunizieren; das *Uncanny Valley* der sterilen Zelle war also imminently spürbar,

mit einer gewissen narrativen Komponente von angedeuteter Lebendigkeit des Raums als intuitiv korrekte und naheliegende Lösung um dieses einengende Gefühl der Isolation zu umgehen. Ausgehend von dieser Basis war der Schritt zum Erstellen eines kompletten virtuellen Klassenzimmers ein vergleichsweise kleiner.

Da uns die Erfahrung mit Texturenerstellung oder der Arbeit mit geometrischen Objekten in UNITY fehlt, haben wir uns entschieden, passende Objekte im *Unity Asset Store* zu besorgen. Unsere Wahl fiel hierbei auf die kostenpflichtige Kollektion *Simple Classroom* des User NeutronCat³, die 22 Modelle und 23 Prefabs enthält, darunter etwa Stühle, Schreibtische oder eine Tafel mit passender Kreide. Unter allen verfügbaren Alternativen war das die mit Abstand günstigste und sollte zunächst nur als Probelauf dienen, stellte sich aber in der Praxis schnell als qualitativ und quantitativ ausreichend heraus, weshalb wir es weiterhin nutzen.

Wie auch die DEPTHKIT-Prefabs war die Installation und Einarbeitung der Objekte sehr nutzerfreundlich. Mit wenigen Klicks und Drag-and-Drop waren Tafeln, Türen und Fenster an den Wänden, sowie Stühle, Tische und dekorative Objekte wie ein Bücherregal auf dem Boden platziert (siehe Abb. 16). Eine Deckenlampe mit bereits voreingestellten Lichteffekten konnte angebracht werden, sodass alle Möbel und sogar die Person im volumetrischen Video beleuchtet werden und korrekte Schatten werfen. Außerhalb des Klassenzimmers platzierten wir dann noch ein simples Quadrat vor den Fenstern, das als Textur ein Foto der bunten Dächer des Foyers der Universität Konstanz verwendet – ein bekanntes Bild für Konstanzer Studierende, das jedoch problemlos und je nach Wunsch und Verwendungszweck frei ausgetauscht werden kann. Zur weiteren atmosphärischen Unterstützung wurden hinter dem Fenster noch eine weitere Lichtquelle in Unity hinzugefügt, wie auch eine Audiodatei, die in regelmäßigen Abständen Vogelzwitschern abspielt. Dies erzielt den Effekt, dass von außen kommende Sonnenstrahlen simuliert werden, und einer potenziell unwohlens Sterilität des virtuellen Raums mit passiv wahrgenommenen Naturgeräuschen entgegenzuwirken.

3 <https://assetstore.unity.com/packages/3d/environments/simple-classroom-97585>.



Abb. 16: Das digitale Klassenzimmer in fertiger Fassung für unser erstes Projekt mit volumetrischen Aufnahmen.

Das Endprodukt des virtuellen Klassenzimmers besteht somit aus einem Schreibtisch mit drei Stühlen (ein Stuhl pro Schüler*in in der Aufnahme, die als ‚flaches‘ Video so projiziert werden, dass es wirkt, als würden sie am Tisch sitzen), einem Monitor, der ein Video der benutzten PowerPoint-Präsentation abspielt, eine Tafel mit dem Thema der digitalen Stunde in simulierter Kreideaufschrift und dekorative Elemente in Form von Türen, Fenstern mit Außenansicht, einer Wanduhr, einer Deckenleuchte mit simuliertem Licht sowie einem gefüllten Bücherschrank.

5.4. EXPORT IN VR

Für dieses Projekt wurden zwei Unterrichtseinheiten aufgenommen, jeweils eine zu marktkonformen beziehungsweise marktkonträren Staats- eingriffen. Beide Aufnahmen wurden separat als einzelne ‚Level‘ in der Game Engine erstellt, damit sie als eine App auf den VR-Headsets zur Besichtigung installiert werden können. Um zwischen den Sitzungen auswählen zu können, wurde ein Hauptmenü erstellt, das den PowerPoint-Monitor als ‚Levelauswahl-Menü‘ zweckentfremdet, die mit einem Klick über den Controller die gewünschte Unterrichtseinheit lädt. Da zweidimensionale Menüs in VR schnell wie ein Fremdkörper wirken, wurde das

Hauptmenü als das virtuelle Klassenzimmer aus den Sitzungen gestaltet, nur dass die Lehrperson und Schüler*innen fehlen. Stattdessen haben wir Audio-Aufnahmen eingefügt, die das Schreiben auf Papier und Schritte im Flur simulieren sollen, damit die Illusion von einer Pause zwischen zwei Lehrstunden einen hörbaren Kontrast zu den tatsächlichen Unterrichtseinheiten herstellt und damit in der virtuellen Umgebung klar und unmittelbar die Situation vermittelt.

Um die Immersion und das Gefühl der körperlichen Anwesenheit im virtuellen Raum weiter zu steigern, wurden zwei weitere Elemente hinzugefügt: digitale Hände (eingefügt als Prefab von UNITY) und ein schlichtes, zweidimensionales Menü mit einem sehr simplen und rudimentären Aufbau, die das Pausieren und Wiederabspielen der volumetrischen Clips und eine Rückkehr in das Hauptmenü ermöglichen. Dieses Menü wird geöffnet, sobald die spielende Person auf die linke Handfläche schaut, sobald diese offen nach oben zeigt (siehe Abb. 17). Hier hilft ein Prefab, das von UNITY selbst im Basic-Kit für VR-Entwicklung bereitgestellt wird. Die Hände der Person in VR werden dabei automatisch getrackt und im Raum korrekt dargestellt, die Interaktion erfolgt somit durch ein Tippen mit dem digital simulierten Finger auf die virtuellen Knöpfe. Bei der aktuellen Version der App handelt es sich bei alledem um simple, aber funktionale Platzhalter, die für die nächsten derartigen Projekte optisch und durch ein besseres Interaktionsskript verbessert werden sollen.

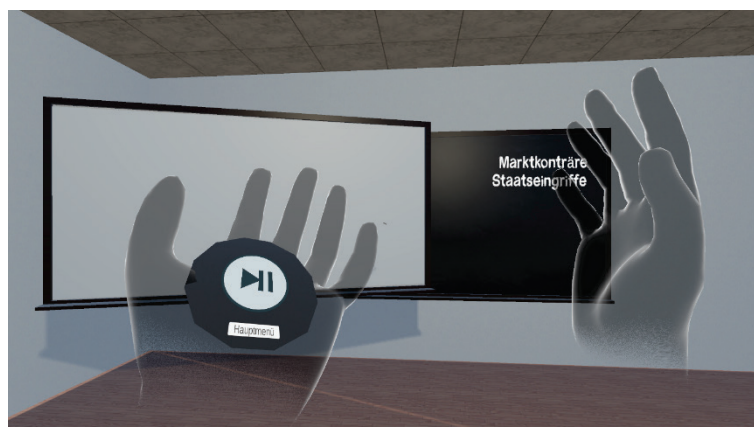


Abb. 17: Das Interaktionsmenü an der linken Handoberfläche, wie es in VR sichtbar ist.

Der fertige Build dieses VR-Projekts besteht also aus drei anwählbaren Szenen – dem Hauptmenü und zwei Unterrichtseinheiten –, wobei die App im Hauptmenü startet. In der jetzigen Version wird diese App als eine Datei für Android-Betriebssysteme gespeichert und im Anschluss direkt auf die Quest-Headsets im Entwicklermodus installiert. Die App kann somit auch ohne Verbindung mit einem Computer und dank Hand-Tracking wahlweise auch ohne zusätzliche Controller gestartet und genutzt werden.

6. AUSBLICK

Ausgehend von unserem Wissensstand, den wir Ende 2024 erreicht haben, eröffnen sich drei größere Themenfelder, die im Zuge der Evaluierung und der Weiterentwicklung einer methodischen Einbettung von volumetrischen Videos unserer Meinung nach angegangen werden müssen. Die Arbeit im MARS-D-Projekt erlaubt uns abseits des Auslotens der Grenzen und Hürden unseres volumetrischen Aufnahmesystems, zu verstehen, wie ein solches System in Forschungsdesigns eingebettet werden kann, welche Bedingungen beachtet werden müssen und welche Spezifika die Daten mit sich bringen, die wir generieren.

Daraus ergibt sich (1) die Notwendigkeit, zu eruieren, wie volumetrische Daten analysiert werden können, ohne sie auf ein anderes Medium herunterzubrechen. Wie in unserem Bericht beschrieben, haben wir bis jetzt eine reine Betrachtungsmöglichkeit für die erhobenen Daten entwickelt. Etwasige Annotationen müssen hier noch außerhalb der virtuellen Umgebung gemacht werden. Der Vorteil und damit die Hürde liegt darin, direkt in der virtuellen Umgebung und somit in Raum und Zeit des Videos zu annotieren. Analog zu einer Annotation eines Bildes oder Videos, in denen der Ort im Bild und entlang der Zeitachse für eine nachvollziehbare Anmerkung vonnöten ist, bedarf es in volumetrischen Videos noch des Blickwinkels bzw. der Position des Betrachtenden im Raum in Verbindung mit der Zeit im Video und der betrachteten Position im entsprechenden Bildausschnitt, um eine ähnliche Nachvollziehbarkeit zu erreichen. Um eine solche Anno-

tation zu ermöglichen, planen wir eine Zusammenarbeit mit der Forschungsgruppe Human-Computer Interaction an der Universität Konstanz.

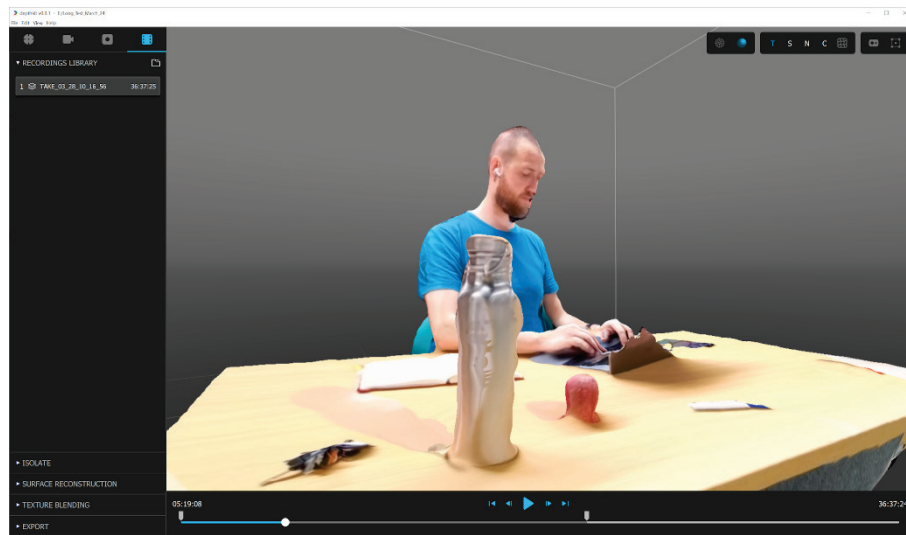


Abb. 18: Die Flasche und der Apfel werden nicht in ihrer eigentlichen Geometrie dargestellt. Die Schlüssel (vorne links) und der Stift (rechts neben Apfel) werden nicht als Objekte erfasst, sondern nur als Textur angezeigt.

Vor die Analyseproblematik stellt sich dabei grundlegend die Frage nach dem (2) Nutzen volumetrischer Videos in der Forschung. Diesen Punkt können wir nur schlaglichtartig ausführen, da wir nur ein volumetrisches Kamerasystem haben und dieses vergleichsweise einfach aufgebaut ist. In den Aufnahmen während unserer Testläufe merkten wir, dass Gegenstände wie Trinkflaschen, Äpfel oder Laptops unvollständig oder in einer anderen Geometrie erfasst werden (siehe Abb. 18). Notizbücher wurden zur Tischtextur bzw. verschwanden, wenn sie in die Hand genommen wurden. Bildschirme werden von unseren Kameras aufgrund ihrer Oberfläche nicht erkannt. In einem Fall fehlten uns die Beine einer Person, da das Material der Hose das Infrarotlicht nicht reflektierte und somit keine Tiefendaten lieferte. Dadurch eignet sich unser System nur bedingt, sobald zu kleine oder zu dünne Objekte Teil der Aufnahme sind. Somit entsteht ein Nachteil in der Nutzung volumetrischer Aufnahmen im Vergleich zu einer Videoaufnahme, wenn es bspw. um eine Brettspielsituation geht. Ebenfalls

verringert sich der Nutzen, sobald sich z.B. mehrere Menschen im aufgenommenen Volumen befinden, da sie sich dabei gegenseitig verdecken und somit Okklusion entsteht (vgl. Abb. 19). Aktuell ziehen wir den größten Nutzen aus unserem System, wenn wir einzelne Personen aufnehmen, die entweder allein im Raum oder mit weiteren Personen außerhalb des Volumens interagieren.

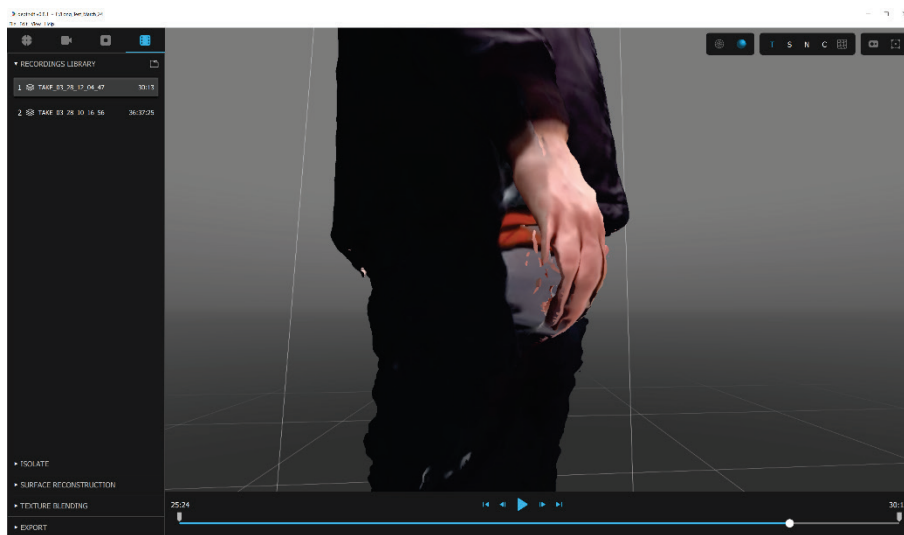


Abb. 19: Die Hand verdeckt Teile des Beines, dadurch entstehen für den Zwischenraum von Hand und Bein keine Tiefendaten auf die eine Textur gelegt werden kann.

Letztendlich reflektieren wir ebenso (3) die ethischen Dimensionen der erhobenen Daten eines volumetrischen Videos. Hier geht es nicht um personenbezogene Daten im datenschutzrechtlichen Sinne, als vielmehr um die möglichen Risiken aus der Zusammenführung und Synthetisierung basierend auf den erhobenen volumetrischen Daten. Mit Hilfe unterschiedlicher künstlicher Intelligenzen besteht die Möglichkeit, aus den Daten volumetrischer Videos animierbare 3D-Modelle der Personen zu erstellen und die Videos als Trainingsdaten zu verwenden, um Gestik und Bewegungsabläufe zu erhalten. Ebenso kann aus den Tondaten ein synthetisiertes Stimmprofil erstellt werden. In der Bündelung der genannten Verfahren liefern volumetrische Videos die Grundlage, um (rudimentär) einen

digitalen Zwilling einer Person zu erstellen. Zwar braucht es dazu umfangreiches Wissen im Umgang mit diesen Möglichkeiten, dennoch sind wir der Meinung, dass wir uns diesem Risiko der Nutzung bewusst sein sollten. Es geht hierbei (aktuell) nur indirekt um datenschutzrechtliche Belange, sondern vielmehr um die ethische Dimension in der weiteren Nutzung, gerade im Hinblick auf das Forschungsdatenmanagement, die mögliche Anonymisierung und das Teilen der erhobenen Daten im Sinne der Open Science-Bewegung.

Abschließend hoffen wir, dass wir mit unseren Ausführungen zumindest einen kurzen Einblick in unsere Laborarbeit zum Erstellen volumetrischer Videos und einem Teil der Arbeit des GameLab geben konnten. Wir haben in den Vorbereitungen und der Durchführung der Aufnahmen unser Verständnis und Wissen darüber, wie ein solches System funktioniert, welche Reibungspunkte entstehen, wie wir Proband*innen in der Aufnahme begleiten und wie wir die Daten bestmöglich aufbereiten, ausgebaut. In Zukunft gilt es nun, die Anwendungsszenarien für Spielsituationen auszutesten, die Postproduktion zu verfeinern und die Bereitstellung und Nachnutzung der Daten auszuloten.

SOFTWARE

CHATGPT (2022), OpenAI.

DEPTHKIT (2012), Scatter.

OPEN BROADCASTER SOFTWARE STUDIO (2012), Lain Bailey.

PHOTOSHOP (1990), Adobe.

PREMIERE PRO (2003), Adobe.

SPATIALSCAN3D (2022), Djinn Technologies Limited.

UNITY (2005), Unity Technologies.

YOUTUBE (2005), Google LLC.

ÜBER DIE AUTOR*INNEN

Marko Jevtic, M.A. ist ein Promotionsstudent an der Universität Konstanz. Seinen Masterabschluss machte er in der Amerikanistik, wechselte aber zum Promotionsstudium in die Medienwissenschaft, um zu Videospielen zu forschen. Dabei schaut er aus einer primär literaturwissenschaftlichen Perspektive und mit narratologischer Methodik auf Rassismus in Games, sowie das Zusammenspiel zwischen Interaktivität und Aktivismus. Neben seiner Promotion arbeitete Marko Jevtic als freiberuflicher Journalist bei Mein-MMO.de und unterrichtet Seminare über Hip-Hop-Storytelling, Narrativen der Radikalisierung in Film und Serien, oder Verhandlungen von Genderidentitäten in Hard-Boiled Crime Fiction und dem Film Noir. Außerdem entwickelt er gemeinsam mit Studierenden eine Spieladaption von Road Trip-Novels und hilft bei der Entwicklung von Accessibility-Software für Maus-Inputs am PC.

Benjamin Schäfer, M.Sc. promoviert an der Universität Konstanz und leitet dort das GameLab als Academic Manager. In dieser Rolle steht er einerseits als Ansprechpartner für Forschende zur Verfügung und berät diese von der Antragsstellung bis zur Umsetzung eines Forschungsprojekts, andererseits betreut, pflegt und erweitert er die gesamte technische Ausstattung. In seinem Promotionsprojekt interviewte er Entwickler*innen von Virtual Reality Hardware und Software im Silicon Valley, in Seattle und Stockholm, um herauszufinden, wie sie Realität in virtuelle Realitäten übersetzen. Grundsätzlich liegen seine Forschungsinteressen in den Funktionsweisen von Computerspielen und den Produktionsprozessen, die beim Entwickeln von Computerspielen zum Tragen kommen.