



Bachelorarbeit

im Studiengang Medientechnik

Wahrnehmung räumlicher Tiefe in webbasierten 3D-Darstellungen auf 2D- Monitoren: Eine empirische Nutzerstudie

Eingereicht von: Enis Inep

Matrikelnummer: 7022248

Mail: enis.inep@stud.hs-emden-leer.de

Erstprüfer:

Prof. Dr. rer. nat. Thies Pfeiffer

Zweitprüfer:

Jannik Franssen, M.Eng.

Hamburg, 09 2024

Inhaltsverzeichnis

1 Einleitung	1
1.1 Motivation und Relevanz des Themas	1
1.2 Problemstellung: Tiefenwahrnehmung bei Panoramabildern auf 2D Monitoren	2
1.3 Zielsetzung und Forschungsfragen.....	2
1.4 Aufbau der Arbeit.....	3
2 Theoretische Grundlagen	6
2.1 Wahrnehmungspsychologische Grundlagen (Hergovich, 2022a).....	6
2.1.1 Parallax Effect und Tiefenhinweise (monokular vs. binokular)(Wölfele, 2023).....	6
2.1.2 Grenzen der Tiefenwahrnehmung auf 2D-Displays.....	7
2.2 360°-Panoramen als Darstellungsformat.....	7
2.2.1 Equirectangular Projection und Geometrische Verzerrungen.....	8
2.3 Texturbasierte Verfahren der Tiefendarstellung	9
2.3.1 Texture Baking.....	9
2.3.2 Rolle von Depth Maps.....	10
2.3.3 (Überblick Modifikatoren in Blender)	12
2.4 Überblick bestehender Ansätze	12
2.4.1 Allgemeine Verfahren der Tiefendarstellung in 2D-Panoramen	12
2.4.2 Spezifische Verfahren im Kontext 360°-Panoramen	12
2.4.3 Einschränkungen und Forschungsbedarf	14
2.5 Limitationen	14
3 Methodik	15
3.1.1 Subjektive Tiefenwahrnehmung	15
3.1.2 Verhaltensbasierte Tiefeneinschätzungen	15
3.2 Forschungsdesign: Analyse und Bewertung der Szenarien	16

3.3Definition der Szenarien.....	17
3.3.1Szenario 1: Baseline-3D-Rendering mit gezielten Tiefenherausforderungen.....	18
3.3.2Szenario 2: Depth Map + 360°-Panorama (<i>Babylon.js</i> , Artefakte, Optimierungsstrategien).....	18
3.3.3Szenario 3: Low-Poly-Rekonstruktion aus Panorama	18
3.4Technische Umsetzung	19
3.5Evaluationsdesign.....	19
3.5.1Stimulusmaterial.....	19
3.5.2Auswertungsmethode	19
4Ergebnisse	20
4.1Technische Ergebnisse der Szenario-Umsetzungen.....	20
4.2Beobachtete Optimierungseffekte (z. B. weniger Artefakte, visuelle Glätte).....	20
4.3Ergebnisse der Nutzerbefragung (Darstellung in Tabellen, Diagrammen).....	20
4.4Zusammenfassung der Ergebnisse in Bezug auf Forschungsfrage	20
5Diskussion	21
5.1.1Interpretation der Ergebnisse im Kontext der Forschungsfrage.....	21
5.1.2Vergleich mit Literatur(?)	21
5.1.3Bewertung der Verfahren: Stärken, Grenzen, Anwendungspotenzial	21
6Fazit und Ausblick	22
6.1Beantwortung der Forschungsfrage	22
6.2Beitrag der Arbeit.....	22
6.3Offene Fragen und Ausblick	22
7Anhang	XXIII

1 Einleitung

Im Rahmen meines Praktikums bei einem VR/XR-Start-up in Hamburg habe ich ein ausgeprägtes Interesse an der Tätigkeit als 3D Artist entwickelt. Nach Abschluss des Praktikums habe ich mich am Mixed-Reality-Lab der Hochschule Emden/Leer nach einem geeigneten Thema für eine Bachelorarbeit erkundigt. Durch meinen Betreuer, Herrn Dr. Prof. Theis Pfeiffer, wurde ich darüber informiert, dass das Mixed-Reality-Lab regelmäßig mit 360°-Panoramen im Bereich der virtuellen Realität arbeitet. Diese Bilder enthalten keine Tiefendaten und wirken daher zweidimensional. Mithilfe einer Depth Map ist es möglich, ein Tiefeneindruck zu erzeugen, der dem Panorama mehr Tiefe verleiht. Allerdings weist diese Methode auch einige Schwächen auf. Das Verfahren könnte optimiert werden, um diesen Tiefeneindruck zu erzeugen. Im Rahmen der Analyse werden drei Herangehensweisen erörtert und umgesetzt. Darüber hinaus interessieren mich auch Bilder und Computergrafik im Allgemeinen und daher ist das Thema für mich von großem Interesse.

1.1 Motivation und Relevanz des Themas

360°-Umgebungen und webbasierte 3D-Darstellungen gewinnen in vielen Bereichen an Bedeutung, etwa in Games, Lernumgebungen oder Anwendungen wie Street View. Obwohl im Theorieteil bereits grundlegende Verfahren vorgestellt wurden, besteht weiterhin Forschungsbedarf darin, wie sich räumliche Tiefe auf einem 2D-Monitor überzeugend vermitteln lässt – insbesondere unter den technischen Einschränkungen einer Webumgebung.

Die Frage, wie man flache Szenen durch geeignete Verfahren „räumlicher“ wirken lassen kann, ist dabei sowohl technisch als auch gestalterisch relevant. Da diese Themen in der Medientechnik zunehmend an Bedeutung gewinnen und zudem eine Nische darstellen, die bislang nur begrenzt untersucht wurde, ergibt sich ein interessanter Ansatzpunkt für weiterführende Forschung.

Die Themenwahl entstand auch aus persönlicher Motivation: Das Gebiet wurde mir an der Hochschule durch Thies Pfeiffer erstmals nähergebracht, und die Inhalte decken sich stark mit meinen eigenen Interessen. Ziel dieser Arbeit ist es daher, verschiedene Verfahren praktisch umzusetzen und ihre subjektive Wahrnehmung zu untersuchen. Die Ergebnisse können als Grundlage dienen, um zu beurteilen, welches Verfahren Potenzial für Optimierungen bietet – und welches eher weniger geeignet ist.

1.2 **(Problemstellung: Tiefenwahrnehmung bei Panoramabildern auf 2D Monitoren)**

Kommentiert [EI1]: Kann Denk ich weg?

1.3 Zielsetzung und Forschungsfragen

Das zentrale Ziel dieser Arbeit ist die Entwicklung und Untersuchung von drei unterschiedlichen Verfahren zur Darstellung und Tiefenwahrnehmung auf 2D-Monitoren. Die erstellten Szenarien sollen anschließend im Web laufen und dort auch technisch sinnvoll performen. Eines dieser 3 Szenarien dient hierbei als Basisszenario, aus dem die anderen beiden Varianten hergeleitet werden. Die Szenarien werden individuell bewertet, um die Wirkung von Tiefe zu erfassen und zu bewerten. Spezifischen sollen die Stärken, Schwächen sowie Grenzen jedes Verfahrens untersucht werden.

Ein weiterer Schwerpunkt liegt auf der technischen Performance im Webbrowser, beispielsweise hinsichtlich Ladezeit, Ressourcenverbrauch und flüssiger Darstellung, da diese Faktoren die Nutzbarkeit und die Wahrnehmung der Szenen maßgeblich beeinflussen. Vor diesem Hintergrund adressiert die Untersuchung folgende Fragen:

Wie werden zentrale Aspekte der Tiefenwahrnehmung (z. B. Tiefenstaffelung, Verzerrungen, Realismus) in den drei Szenarien wahrgenommen? Welche Stärken und Grenzen zeigen sich bei jedem Verfahren im Hinblick auf die Erzeugung eines räumlichen Eindrucks auf einem 2D-Display? Wie unterscheiden sich die Szenarien hinsichtlich Performance-Kriterien wie Ladezeit, RAM-Verbrauch und Darstellungsflüssigkeit, und welchen Einfluss hat dies auf die Nutzerwahrnehmung? Darüber hinaus stellt sich die Frage, inwiefern die untersuchten Verfahren Potenzial für zukünftige Optimierungen oder weitere Anwendungsfälle im Bereich webbasierter 3D-Darstellungen bieten.

Auf Grundlage dieser Zielsetzung und Forschungsfragen werden im Folgenden mehrere explorative Hypothesen formuliert, die die anschließende Untersuchung leiten.

H1: Für *Szenario 2* (Depthmap) wird eine geringe und teilweise verzerrte Tiefenwahrnehmung erwartet.

H2: Für *Szenario 1* (High-Poly) wird eine ausgeprägte Tiefenwahrnehmung erwartet.

H3: Für *Szenario 1* (High-Poly) wird eine hohe Ladezeit und ein hoher Ressourcenverbrauch erwartet.

H4: Für *Szenario 3* (Low-Poly) wird eine klare und stabile Tiefenwahrnehmung erwartet.

1.4 Aufbau der Arbeit

Die drei Herangehensweisen werden in einer 3D-Umgebung erstellt und jedes Szenario wird individuell optimiert.

Szenario 1: S1

Das Basisszenario S1 ist ein 3D-Rendering, in dem Objekte bewusst so angeordnet sind, dass Grenzen der Tiefenwahrnehmung ausgereizt werden. Die Herausforderung besteht darin, dass Nutzer:innen bei der Betrachtung Schwierigkeiten haben können, Tiefe, Abstände und Relationen korrekt einzuschätzen. Damit dient S1 als Testszene, um systematisch zu prüfen, welche Probleme bei der Wahrnehmung von Tiefe auftreten.

Aufbauend auf S1 werden weitere Szenarien erzeugt. Diese stellen verschiedene technische Lösungsansätze dar, um die Wahrnehmung von Tiefe zu verbessern. Ein Highpoly-Rendering wie in S1 zeigt zwar hohe Detailtreue, bringt aber Herausforderungen hinsichtlich Performance und Dateigröße mit sich, insbesondere in einer Web-Umgebung.

Szenario 2: S2

Szenario 2 S2 erweitert die Basisszene S1 durch die Fusion einer Depth Map mit dem 360°-Panorama. Mithilfe von Babylon.js wird auf einer unterteilten Sphere ein Tiefeneindruck erzeugt. Jedes Face erhält den Tiefenwert seines Mittelpunktes, wodurch die Szene räumlich dargestellt wird.

Bei der Analyse von S2 treten jedoch neue Herausforderungen auf. Insbesondere an Spitzen, Ecken und Außenkanten entstehen Verzerrungen, wenn sich die Nutzer:innen im Raum bewegen oder drehen. Dies liegt daran, dass die Kanten der Faces nach hinten abknicken und in der Mitte nadelartige Spitzen entstehen, die den Gesamteindruck der Szene beeinträchtigen.

Szenario 3 S3 verfolgt den Ansatz, ein Lowpoly-Mesh aus der Basisszene S1 zu erzeugen, das mit gebackenen Texturen versehen werden kann. Ziel ist es, ein visuell zufriedenstellendes Ergebnis zu erreichen, ohne ein neues Mesh manuell zu modellieren. Im Unterschied zu S2 soll das Mesh hier nicht zwingend als durchgehendes, geschlossenes Modell umgesetzt werden.

Vielmehr bietet sich die Möglichkeit, Objekte im Raum vom Hintergrund zu trennen, sodass unnötige Flächen insbesondere hinter dem eigentlichen Mesh eingespart werden können.

Da die konkrete Herangehensweise für S3 noch nicht endgültig festgelegt ist, besteht eine Herausforderung darin, eine geeignete Methode zur Lowpoly-Erzeugung zu identifizieren. Dabei müssen einerseits visuelle Qualität und Tiefeneindruck gewahrt bleiben, andererseits dürfen Performance und Dateigröße nicht aus dem Ruder laufen. Die Trennung von Objekten und Hintergrund erfordert zudem eine saubere Segmentierung, um Darstellungsfehler zu vermeiden. Für eine fundierte Bewertung wird es sinnvoll sein, die Ergebnisse online oder in Form von Videodemonstrationen zu präsentieren, damit der Parallax-Effekt während der Beurteilung berücksichtigt werden kann.

2 Theoretische Grundlagen

Das Thema der Untersuchung lautet: "Untersuchung dreier Verfahren zur Darstellung von Panoramabildern mit visuell optimierter Tiefenwahrnehmung und ihren Grenzen". Für die Untersuchung ist ein gewisses Maß an Verständnis der Wahrnehmungspsychologie erforderlich. Es stellt sich die Frage, wie Tiefe auf einem 2D-Monitor wahrgenommen wird, wann genau Tiefe erkannt wird, was der Parallaxe-Effekt ist und wozu er wichtig ist. Zudem wird erörtert, wozu DepthMaps dienen und welche Funktionen sie bieten.

Im Folgenden erfolgt eine Definition des Begriffs *360°-Panorama*. Ein 360°-Panorama liegt in equirektangularer Projektion vor. Dabei wird die sphärische Umgebung in eine zweidimensionale, rechteckige Fläche mit einem Seitenverhältnis von 2:1 transformiert. Dieses Format kann in 3D-Anwendungen wieder auf eine Kugelgeometrie projiziert werden, wodurch eine immersive Rundumsicht entsteht. In der vorliegenden Arbeit wird ein direkt anschließendes, umlaufendes Bild als 360°-Panorama bezeichnet.

Kommentiert [Inep, Eni2]: Passt das hier? Und beleg fehlt!

2.1 Wahrnehmungspsychologische Grundlagen (Hergovich, 2022a)

Unnötiges Kapitel

Kommentiert [Inep, Eni3]: Kann weg? Kann Weg!

2.1.1 Parallax Effect und Tiefenhinweise (monokular vs. binokular)(Wölfel, 2023)

Wie bereits in den „Theoretischen Grundlagen“ angedeutet, setzt dieses Kapitel grundlegende Kenntnisse der Wahrnehmungspsychologie voraus. Im Folgenden wird daher die Tiefenwahrnehmung in ihren monokularen und binokularen Formen näher erläutert. Ein anschauliches Beispiel hierfür ist die Wahrnehmung von Objekten entlang einer Bahnstrecke. „Schauen wir während der Fahrt aus dem Zug, rauschen nahe Objekte schnell an uns vorbei, ohne dass wir sie scharf wahrnehmen können, während Objekte am Horizont nahezu unverändert bleiben“ (Hergovich, 2022b, S. 76). Dieses Phänomen wird als Bewegungsparallaxe (*Parallax Effect*) bezeichnet und gilt als monokularer Tiefenhinweis, da bereits ein Auge genügt, um relative Tiefenunterschiede wahrzunehmen (Hergovich, 2022b, S. 76).

Im Rahmen der binokularen Tiefenwahrnehmung kommt es aufgrund der leichten Versetzung der Augen zum Auftreten von zwei verschiedenen Bildern, die in der Fachliteratur als Querdisparation bezeichnet werden. Das Gehirn gleicht diese beiden Bilder ab und unterscheidet, inwiefern sich die Positionen der Objekte voneinander versetzt sind (Schönhammer, 2013, S. 196–197). Die daraus resultierende Tiefenwahrnehmung übertrifft die

eines einzelnen Auges (Hibbard et al., 2023, S. 7–8). Um die Thematik zu veranschaulichen, kann der Versuch unternommen werden, einen Faden durch ein Nadelöhr zu fädeln, zunächst mit einem geschlossenen Auge und anschließend mit beiden Augen.

2.1.2 Grenzen der Tiefenwahrnehmung auf 2D-Displays

Der Versuch, Tiefe auf einem 2D-Monitor darzustellen, auf den beide Augen fokussiert sind, führt nicht zur Querdisparation, da das Bild lediglich auf einer flachen Oberfläche betrachtet wird (vgl. Lehn et al., 2022, S. 422). Die Tiefenwahrnehmung erfolgt daher ausschließlich monokular, auch wenn beide Augen genutzt werden. Um dennoch einen Eindruck von Tiefe zu erzeugen, können verschiedene Gestaltungsmittel eingesetzt werden. So lassen sich durch perspektivische Fluchtpunkte oder Größenunterschiede von Objekten räumliche Effekte hervorrufen. Auch eine gezielte Nutzung von Licht und Schatten oder Hierarchie der Objekte, wenn im Vordergrund stehende Objekte, die im Hintergrund verdecken, können Tiefe im Bild schaffen. (vgl. Hergovich, 2022b, S. 77).

Darüber hinaus kann Bewegung eine Bewegungsparallaxe erzeugen und die Tiefenwahrnehmung verstärken. Auch externe Faktoren beeinflussen die Wahrnehmung: Untersuchungen zeigen, dass sowohl der Betrachtungsabstand als auch die Displaygröße eine Rolle spielen. Je näher der Betrachter am Monitor sitzt und je größer das Display ist, desto stärker wirken Fluchtpunkte und Größenunterschiede, wodurch die Tiefe intensiver wahrgenommen wird (vgl. Miyashita et al., 2022, S. 5:2–5:5). Es ist wichtig auf diese Methoden zuzugreifen, um mehr Tiefe auf 2D-Monitoren zu erzeugen, denn auf Standbildern bleibt der Tiefeneindruck eher flach (Miyashita et al., 2022, S. 5:10-5:11).

2.2 360°-Panoramen als Darstellungsformat

Monitore und Bildformate besitzen feste Grenzen und beschränken somit den sichtbaren Bildausschnitt (Hoffmann, 2025, S. 10–11). Sie unterliegen normierten und standardisierten Größenverhältnissen, die sich historisch aus technischen Entwicklungen ergeben haben (Hoffmann, 2025, S. 19–20). Um den wahrnehmbaren Darstellungsbereich über diese Festgelegten Grenzen zu erweitern, werden Panoramen eingesetzt. Diese ermöglichen eine erweiterte bildliche Wahrnehmung sowohl auf zweidimensionalen Bildschirmen oder im VR-Bereich. (Hoffmann, 2025, S. 11–12).

Dabei kann man unter verschiedenen Arten von Panoramabildern unterscheiden, am bekanntesten sind 180°-Darstellungen. Für diese Bachelorarbeit werden jedoch 360°-

Panoramen betrachtet, da diese im Kontext von 3D-Visualisierung und Tiefenwahrnehmung relevanter sind. Diese Panoramen decken das gesamte Sichtfeld ab, das sich über 360° horizontal und 180° vertikal erstreckt. (Hoffmann, 2025, S. 23–25). Praktisch nutzt man 360°-Panoramen oft in VR-Anwendungen oder im Architekturbereich zur Simulation, in den anwendungsfällen wirken sie besonders Immersiv.

Diese Darstellungsform ist besonders hilfreich, um seine Umgebung von einem Zentralen Punkt aus wahrzunehmen. In der Regel sehen wir unser Sichtfeld um uns herum, ein Panorama bildet dieses Sichtfeld auf eine Fläche ebene ab und erzeugt damit eine Verzerrung (Hoffmann, 2025, S. 21–24).

Kommentiert [Inep, Eni4]: Eigentlich gute Überleitung für nächstes Thema lol

2.2.1 Equirectangular Projection und Geometrische Verzerrungen

Wir betrachten die Welt in einer sphärischen Geometrie unser gesamtes Sichtfeld bildet eine 360° Kugeloberfläche. Jeder Punkt in diesen Raum lässt sich durch Richtungsvektoren beschreiben oder Sphärischen Koordinaten beschreiben. Sobald man diese Umgebung auf eine 2D Fläche überträgt, in Form eines Bilds oder ähnliches wird im Oberen und unterem Bereich das Bild verzerrt. Das 360°-Panorama wird an der Oberen und Unteren Bildfläche gestaucht und in die Ecken des Equirectangulars „abgewickelt“, dass Breitengrade linear auf die y-Achse und Längengrade linear auf die x-Achse gemappt werden. Mathematisch Beschrieben:

$$x = \lambda, y = \phi$$

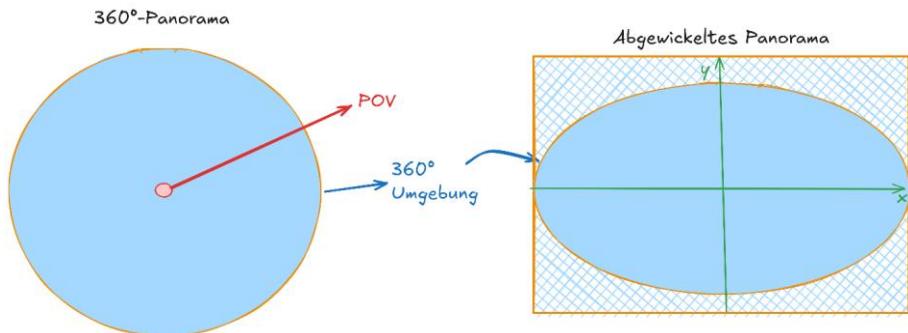
Längengrad $\lambda \in [-180, 180]$ und Breitengrad $\phi \in [-90, 90]$.

Der Maßstab M Wächst mit der breite:

$$M = \frac{1}{\cos(\phi)}$$

Am Äquator ($\phi = 0^\circ$)ist $M = 1 \rightarrow$ keine Verzerrung

An den Polen ($\phi = \pm 90^\circ$)geht $M \rightarrow \infty \rightarrow$ maximale Verzerrung



2.3 Texturbasierte Verfahren der Tiefendarstellung

2.3.1 Texture Baking

Texture Baking ist ein Verfahren, bei dem Rendering-Ergebnisse eines High-Poly Modells auf eine zweidimensionale Bitmap Textur übertragen werden, um diese anschließend auf ein Low-Poly Modell anzuwenden (Zhang et al., 2023, S. 647–648). Texture Baking wird oft für Spiele, Web-basierte oder VR/AR Anwendungsfälle verwendet, da es Speicher und Rechenressourcen sparen kann. Andernfalls müssen beispielsweise für ein 3D-Modell einer Schreibtischlampe mehrere Texturen in hoher Auflösung gerendert werden. (Adobe, 2025)

Der Schirm der Lampe ist beispielsweise grünlich transparent und matt, der Lampenfuß besteht aus braunem Holz, und die Beine sind aus glänzendem Metall. Das wären 3 aufwändige

Texturen für ein 3D Modell, die viel Speicherplatz benötigen. Mit Texture Baking könnte man die 3 verschiedenen Texturen zu einer kombinieren (Knott et al., 2023). So werden beispielsweise Farben (Base Color), Roughness map, Normal Map und Metallizität (Metallic) in einem „Atlas“ gespeichert (Blender Foundation, 2025a). Das reduziert den Speicherbedarf und auch den Rechenaufwand bei der Laufzeit. Um Texture Baking in Blender anzuwenden, braucht das 3D-Modell eine UV-Map und eine „Image Texture Node“ um die Ergebnisse zu speichern (Blender Foundation, 2025a). Der Prozess ist ähnlich wie ein Rendering verfahren, nur mit dem Unterschied, dass das Ergebnis nicht mit einem Framebuffer sondern auf eine Texture gespeichert wird.

In Blender kann man zwischen zwei grundlegenden Bake-Verfahren unterscheiden. Im ersten Verfahren wird von einem Quellobjekt, gängig wird hier ein High Poly verwendet, auf ein Lowpoly oder das gewünschte Objekt gebacken. Dieses Vorgehen ist die gängigste Methode, da sich damit beliebige Texture Maps wie Normal, Ambient Occlusion oder Roughness vom detaillierten Modell auf ein vereinfachtes übertragen lassen. (Blender Foundation, 2025b)

Im zweiten Verfahren wird direkt auf dasselbe Mesh gebacken. Dieses Verfahren wird verwendet, um Material- oder Beleuchtungsinformationen innerhalb eines einzelnen Modells zu konsolidieren, etwa wenn mehrere Shader oder Materialien zu einer einheitlichen Textur zusammengeführt werden sollen. Dadurch kann man komplexe Node-Strukturen reduzieren oder statische Lichtinformationen, beispielsweise Lightmaps, direkt in das Objekt integrieren. (Blender Foundation, 2025b)

Das Verfahren erfordert eine saubere Topologie und korrekte UV-Koordinaten, da jede geometrische Inkonsistenz (z. B. überlappende Faces, inverted Normals oder non-manifold Edges) zu Fehlberechnungen führt. Wichtig ist hierbei die Formgebenden Edges im Low-Poly beibehalten werden, um so genannte Fehler zu auszuschließen.

Kommentiert [Inep, Eni5]: Quellen Aktualisieren und Prüfen.

2.3.2 Rolle von Depth Maps

Eine Depth Map ist ein Bild, das Tiefeninformationen einer Szene speichert (Jähne, 2024, S. 271–272). Sie gehört wie eine Normal Map oder eine Bump Map zur Kategorie der Textur Maps. Gespeichert wird dabei der Abstand zwischen der Kamera und dem jeweils nächstgelegenen 3D Punkt der Szene (Jähne, 2024, S. 283–289). Die Darstellung erfolgt in der Regel über einen Farbverlauf, meist zwischen Schwarz und Weiß. Helle Bereiche stehen für nahe Objekte, während dunkle Bereiche größere Entfernung kennzeichnen. Grundsätzlich sind auch andere Farbschemata möglich, die Monokular-Depth Graustufenabbildung gilt

Kommentiert [Inep, Eni6]: Dopplung

jedoch als Standard (Divyanth et al., 2023, S. 2–3). Wir wissen, dass Farben in Bildern, durch die Kombination von drei Farbkanälen (RGB) entsteht. Jedes dieser Kanäle enthält einen Helligkeitswert der Farbe, in einem Bereich von 0 bis 255, das Mischen der 3 Farbkanäle ermöglicht das additive Farbmischen. Eine Depth Map besteht nur aus einem 8-Bit Graustufenbild, mit einem Graukanal ebenfalls von 0 bis 255. (*What Is Depth Anything V2*, 2025)

Kommentiert [El7]: Ich glaub hier fehlt noch ne quelle?

Depth Maps können auf unterschiedliche Weise erzeugt werden. Zu den gängigen Methoden zählen Stereovision, Time-of-Flight, Structured Light oder monokulare Schätzungen mithilfe von KI (Jähne, 2024, S. 283–289). Im hierfür wurde Blender verwendet, wobei über die Compositing-Nodes ein sogenannter Z-Depth-Pass aus dem Render extrahiert wurde. Blender kann die Tiefeninformationen direkt aus der 3D-Szene berechnen, indem für jedes Pixel der Abstand zur Kamera entlang der Z-Achse bestimmt wird. Diese Werte werden im sogenannten Z-Buffer gespeichert und ergeben in ihrer Gesamtheit die Depth Map (Blender Foundation, 2025a).

Gängige Einsatzfelder von Depth Maps sind beispielsweise die Hinderniserkennung in virtuellen Umgebungen oder die Modellierung von Szenen. Darüber hinaus werden sie im sogenannten Depth-Image-Based Rendering (DIBR) genutzt, um neue Kameraperspektiven aus vorhandenen Tiefeninformationen zu berechnen (Schmeing & Jiang, o. J., S. 279–280) (Lipski et al., 2014, S. 943–944). Auch in modernen Anwendungen, etwa im autonomen Fahren oder in der Robotik, dienen Depth Maps als Grundlage für die Erstellung topologischer Karten zur Erfassung von Hindernissen und Entfernung (Saha et al., 2022, S. 1).

Kommentiert [Inep, Eni8]: Keine Wissenschaftliche Quelle

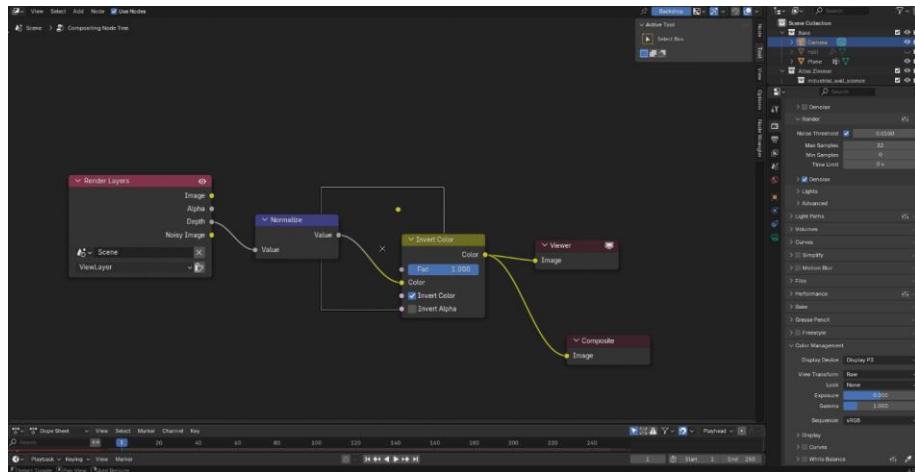
Schwierigkeiten zur Tiefeneinschätzung hat eine Depth Map mit spiegelnden Oberflächen, zudem fassen sie nur die Vorderfläche von einer Szene auf und liefern dadurch eine eingeschränkte Sicht. Außerdem sind sie häufig verrauscht und enthalten oft Schätzungsfehler, wenn man sie mithilfe einer KI erstellt (Divyanth et al., 2023, S. 6–7).

Kommentiert [Inep, Eni9]: Jahr

Mit all diesen Möglichkeiten kann man weiterführend viel mit Depth Maps machen, kombiniert Depth Maps aus verschiedenen Blickwinkeln können zu einer dichten 3D-Rekonstruktion verhelfen (Izadi et al., 2011, S. 1). Neuerdings kann man mit Open-Source Frameworks, wie zum Beispiel NeRF (Neural Radiance Fields), realistische 3D-Szenen mit ihnen erstellen (Ito et al., 2024, S. 5).

In dieser Arbeit liegt der Fokus nicht auf der algorithmischen Erzeugung von Depth Maps, sondern auf deren Anwendung zur visuellen Tiefendarstellung. Entsprechend werden die

verschiedenen Verfahren hier nur überblicksartig erwähnt. (In dieser wird das Szenario 3 nach dem Prinzip erstellt und ist daher für die Theorie wichtig.)



2.3.3 Low-Poly Erstellung und Texturoptimierung für Webbasierte 3D-Darstellungen

Die Erstellung performanter Low-Poly-Modelle ist ein etabliertes Verfahren in der Echtzeit- und Webgrafik. Ziel ist es, den Rechenaufwand beim Rendern und Laden von 3D-Modellen zu reduzieren, ohne den visuellen Gesamteindruck wesentlich zu beeinträchtigen. Eine Verringerung der Polygonanzahl führt zu einer Reduktion der geometrischen Komplexität und kann dadurch zu kürzeren Renderzeiten und geringerem Rechenaufwand in webbasierten 3D-Anwendungen beitragen. (Grande et al., 2024, S. 1)

Neben der Geometrie stellt jedoch insbesondere die Größe der verwendeten Texturen einen zentralen Einflussfaktor auf Speicherbedarf und Ladezeiten dar. Hochauflösende Texturen, wie sie häufig bei High-Poly-Assets eingesetzt werden, können die Gesamtgröße eines Modells maßgeblich erhöhen und damit die Performance webbasierten Renderings negativ beeinflussen. Studien zur 3D-Asset-Optimierung zeigen, dass sowohl die Reduktion der Polygonanzahl als auch die Verkleinerung und Komprimierung von Texturen integrale Schritte sind, um Modelle performancefreundlich zu gestalten, ohne dabei wesentliche visuelle Details zu verlieren. (Grande et al., 2024, S. 28–29)

Wie in Abschnitt 2.3.1 beschrieben, wird zur Erhaltung feiner Oberflächendetails häufig das Texture-Baking-Verfahren eingesetzt. Durch das Übertragen hochfrequenter geometrischer

Details in Texturen kann die Polygonanzahl deutlich reduziert werden. In Kombination mit einer Reduktion der Texturauflösung auf etwa 1K–2K lassen sich Speicherbedarf und Ladezeiten weiter verringern. Im Kontext dieser Arbeit umfasst der Begriff *Low-Poly-Erstellung* daher sowohl die Vereinfachung der Geometrie als auch eine gezielte Optimierung der verwendeten Texturen.

2.4 Überblick bestehender Ansätze

2.4.1 Allgemeine Verfahren der Tiefendarstellung in 2D-Panoramen

Im Rahmen der Literaturrecherche zeigte sich, dass bereits verschiedene Verfahren existieren, um Tiefe in zweidimensionalen Panoramen darzustellen. Eine Studie von 2007 zeigt wie man Layer verwenden kann, um Tiefe in Panorama Bildern zu erstellen, dafür wurden mehrere Bilder von einer Szene aufgenommen, um gezielt den Parallaxeneffekt zu erzeugen. Das Konzept heißt Layered Depth Panorama (LDP) und ist eine Erweiterung des klassischen 2D-Panoramas. Während herkömmliche Panoramen nur ein einziges Tiefenniveau pro Pixel abbilden, speichert das LDP mehrere Tiefenschichten (Layer) pro Blickrichtung ähnlich wie ein „Layered Depth Image“ (LDI), aber in zylindrischer Projektion. (Zheng et al., 2007)

Hier wurde aber nur ein 180-220° Panorama verwendet und hinter den Layern befindet sich keine 3 dimensionale tiefe, der Parallaxe Effekt erzeugt dennoch ein Gefühl von tiefe, aber falls man sich den Layern nährt, bemerkt man das diese Flach sind.

2.4.2 Spezifische Verfahren im Kontext 360°-Panoramen

Eine bereits weit erforschte Variante im Indoor-Bereich ist die Rekonstruktion von 3D-Umgebungen aus einem oder mehreren 360°-Panoramen unter Verwendung von überwachtem Deep Learning (He et al., 2022, S. 5173–5174; Silveira et al., 2023, S. 1–2; Wang et al., 2023, S. 1–2). Dabei werden die Panoramabilder in ein neuronales Netz eingespeist, das zuvor mit zahlreichen Panoramen und deren zugehörigen Tiefeninformationen trainiert wurde (He et al., 2022, S. 5174). Diese sogenannten Ground-Truth-Informationen unterscheiden sich je nach Studie und können beispielsweise Depth Maps, Normalenvektoren oder die Kamerahöhe umfassen (He et al., 2022, S. 5173–5174; Rey-Area et al., 2022, S. 1–2). Durch dieses Training lernt das neuronale Netz, die Tiefenstruktur einer Szene aus einem in manchen Fällen auch aus mehreren Panoramabildern zu rekonstruieren (Silveira et al., 2023, S. 11)

Die Ergebnisse werden typischerweise in Form einer Depth Map, einer Point Cloud oder eines 3D-Meshes ausgegeben, aus denen sich die räumliche Struktur ableiten lässt (Silveira et al.,

2023, S. 6–8). Solche Verfahren finden bislang vor allem Anwendung in Bereichen wie Innenarchitektur, Schulungsumgebungen oder Visualisierungsanwendungen, da sie zwar sehr detaillierte Geometrien erzeugen, diese jedoch häufig große Datenmengen verursachen oder sich primär auf Innenräume beschränken (Zeng et al., 2020, S. 1–2; Silveira et al., 2023, S. 9–10). Für Anwendungen, die auf eine kompakte und effiziente Tiefendarstellung abzielen etwa zur visuellen Unterstützung in interaktiven Interfaces oder mobilen Systemen sind diese Verfahren daher nur bedingt geeignet. Gezielt wird eine Ressource sparsame Variante gesucht, die Webend gut aussieht (Bro ich bin tod was heißt gut aussieht?) und Performen kann. Die meisten Verfahren sind für innenraum szenarien gedacht oder erzeugen nur ein

Kommentiert [Inep, Eni10]: Kann weg oder in die nächste Spalte?

2.4.3 Einschränkungen und Forschungsbedarf

Im vorherigen Kapitel wurden Verfahren vorgestellt, die sich mit der Tiefenschätzung aus 360°-Panoramen befassen. Dabei zeigte sich, dass nahezu alle aktuellen Arbeiten ihren Fokus auf die Erzeugung von Depth Maps legen. Hier soll nun erläutert werden, wieso das so ist und warum die Forschung den Schritt umgeht, mithilfe der erstellten Depth Map ein 3D-Modell zu erstellen. Diese Größe lässt sich mithilfe von einem Vergleichswert messen, zum Beispiel misst Root Mean Squared Error (RMSE), wie weit die vorhergesagte Tiefe vom echten (Ground-Truth) Wert entfernt ist. Zudem kann auch der Durchschnittlicher absoluter Fehler Mean Absolute Error (MAE) gemessen werden, dieser misst den Durchschnitt aller Fehler, egal ob zu groß oder zu klein. Zuletzt kann man auch messen, wie viele Pixel ungefähr richtig sind. Das führt dazu, dass Tiefe eine Eindeutige, metrisch Bewertbare Zielgröße ist. Man könnte also sagen, dass Depth Maps standardisierte vergleichbare Größen sind, die sich Optimieren lassen. (Silveira et al., 2023, S. 7–8).

Da es sich bei Depth Maps um zweidimensionale Bilder handelt, ist für die vollständige Erstellung eines 3D-Modells eine weitere Bearbeitung erforderlich. Silveira et al. betonen hierzu ausdrücklich: „Per-face depth estimation tends to produce discontinuities along face edges [87, 119, 32], and post-processing is required to obtain the full 3D model of the scene.“

(Silveira et al., 2023, S. 20)

Diese Nacharbeitungen kann je nach Verfahren, zu einem hochauflösenden High-Poly Mesh oder Point Cloud, die für Performance Zwecke im Web problematisch werden könnten. Ich möchte hier darauf hinweisen, dass es prinzipiell nicht unmöglich ist, aus einer Depth Map eine

Kommentiert [Inep, Eni11]: Quelle nach 2022

visuell ansehnliches Low Poly zu erzeugen. Diese Nachbearbeitung könnte aufwendig werden, in diesem Fall arbeiten wir mit einem 360°-Panorama. Daher schätze ich das speziell der Aufwand Groß ist ein Optimierte LP zu erstellen, da du hier viel mehr Bildinformationen hast als sonst mit einem normalen Bild.

2.5 Limitationen

Ja es gibt viel. 😊

Kommentiert [Inep, Eni12]: Ich bin mir nicht Sicher ob ich hier gerade eine These Geschrieben habe, die ich nicht belegen kann.

3 Methodik

Es gibt verschiedene Möglichkeiten, die Wahrnehmung von Tiefe in 3D-Szenarien zu untersuchen. In dieser Arbeit wurde eine Nutzerbefragung gewählt, weil sie für experimentelle Szenarien gut geeignet ist und sich ohne spezielle Hardware umsetzen lässt. Die Befragung konzentriert sich darauf, wie Personen die Tiefe in den Szenen einschätzen und beschreiben.

Grundsätzlich lassen sich Methoden zur Erfassung von Tiefe grob in unterschiedliche Bereiche einteilen. Dazu gehören subjektive Verfahren, relative Einschätzungen und objektivere Ansätze. Absolute Messungen, also konkrete Werte wie „das Objekt ist genau X Meter entfernt“ sind in einer Online-Studie jedoch kaum sinnvoll durchführbar. Menschen nehmen Entfernung sehr unterschiedlich wahr, und ohne Hilfsmittel oder Referenzpunkte können solche Angaben leicht stark variieren. Deshalb wurde dieser Ansatz nicht weiterverfolgt.

Es gibt auch physikalisch-technische Messmethoden oder Eye-Tracking Verfahren, die für Tiefe genutzt werden können. Diese kamen hier jedoch nicht infrage, weil sie spezielle Hardware voraussetzen, außerdem ist die Entwicklungsumgebung für 2D-Monitore vorgesehen.

3.1.1 Subjektive Tiefenwahrnehmung

Der wichtigste Teil der Untersuchung basiert auf der subjektiven Wahrnehmung. Die Teilnehmenden geben an, ob sie in der Szene Tiefe erkennen können und wie realistisch diese wirkt. Hier geht es bewusst um persönliche Einschätzungen, also darum, wie „stimmig“ oder glaubwürdig die räumliche Darstellung wirkt. Diese Methode eignet sich gut für eine erste Orientierung und ist leicht über Fragebögen abbildbar. Eine Studie von Lesar et al. nutzt das Subjektive einschätzungsverfahren über Ratingskalen ein etabliertes Verfahren zur Erfassung wahrgenommener räumlicher Tiefe (Lesar et al., 2024, S. 1–3).

3.1.2 Verhaltensbasierte Tiefeneinschätzungen

Ansätze, die stärker in Richtung objektive Wahrnehmungsleistung gehen, beziehen sich auf Aufgaben, bei denen Teilnehmende die räumliche Tiefe nicht nur subjektiv bewerten, sondern aktiv einschätzen.

Zu den bekannten Verfahren gehören, Relative Depth Ordering, relative Distanzeinschätzung und Pairwise oder Forced-Choice Methoden. Bei der Relative Depth Ordering Methode,

urteilen Teilnehmer*innen, welches Objekt näher oder weiter entfernt ist. Die Methode kommt ohne metrische Werte aus und eignet sich daher gut für Entfernungsschätzungen ohne genaue Wertbestimmung (Zabih & Woodfill, 1994, S. 1–3). Auch in aktuellen Studien wird die Methode weiterhin verwendet. Singh et al. (2025) nutzen in ihrer Untersuchung zur Tiefenwahrnehmung ein vergleichbares relatives Verfahren, bei dem Teilnehmende Tiefenpositionen visuell einschätzen und anhand eines Matching-Tasks beurteilen (Singh et al., 2025, S. 5).

Eine weitere Methode ist die relative Distanzschätzung, bei der Teilnehmende eine Entfernung innerhalb von einem bestimmten Distanz Spektrum schätzen (z.B. 5-10m, 10-15m...). Auch diese Methode benötigt keine Absoluten Werte, reduziert jedoch die Genauigkeit freier Schätzwerte (Committeri et al., 2020, S. 1–2; Lesar et al., 2024, S. 4). Zuletzt gäbe es noch die Pairwise- bzw. Forced-Choice Methode. Dabei vergleichen Teilnehmende zwei Darstellungen direkt miteinander und wertet dann welche Darstellung den besseren Tiefeneindruck übermittelt. Allgemein ist das Verfahren gut dafür geeignet und zum Beispiel die Szenarien gegeneinander zu bewerten, da müsste man beide hintereinander zeigen und man fragt dann welche Darstellung realistischer wirkt (Ashtiani et al., 2023, S. 4).

Auf Eye-Tracking wird in dieser Arbeit aus Gründen der fehlenden Hardware verzichtet, obwohl die Methode grundsätzlich gut geeignet wäre, Blickverläufe und Tiefenwahrnehmung detailliert zu erfassen(Choi & Nam, 2025, S. 1–2).

(Theoretisch fehlt hier noch Physikalisch, aber das will ich nur ausbauen, wenn noch Zeit ist.)

3.2 Forschungsdesign: Analyse und Bewertung der Szenarien

Um die drei Szenarien bewerten zu können, wurde ein experimenteller Vergleich als Forschungsdesign gewählt. Die Szenen werden mithilfe von Babylon.js webbasiert dargestellt, sodass sie plattformunabhängig und ohne zusätzliche Software genutzt werden können. Die Evaluation erfolgt über einen Online-Fragebogen, den die Teilnehmenden am eigenen Computer ausfüllen. Szenario 1, 2 und 3 werden dabei **jeweils einzeln** betrachtet und nicht direkt gegeneinander gestellt.

Im Fragebogen kommen vor allem zwei Gruppen von Verfahren zum Einsatz: subjektive Ratings und relative Tiefeneinschätzungen. Die subjektive Tiefenwahrnehmung wird über Likert-Skalen erfasst, zum Beispiel mit Aussagen wie „Eine eindeutige und deutliche Tiefenwahrnehmung zwischen den Bereichen Grau, Grün und Blau ist klar erkennbar“.

Bewertet wird jeweils auf einer Skala von 1 bis 7, wobei 1 eine schlechte bzw. kaum vorhandene Tiefenwahrnehmung und 7 eine besonders klare Tiefenwahrnehmung bezeichnet. Auf diese Weise lässt sich zunächst allgemein erfassen, wie überzeugend die Tiefe in einem Szenario wirkt.

Ergänzend dazu werden relative Tiefenurteile abgefragt, etwa durch Fragen wie „Wie weit entfernt ist die Säule im grünen Bereich von Ihrer aktuellen Position?“ oder „Welches Objekt ist näher an Ihrer Startposition?“. Hier interessieren vor allem Abstände in Kategorien bzw. Relationen (z.B. näher/weiter, Distanzbereiche), nicht exakte metrische Werte. Auf die Erhebung absoluter Distanzen in Metern wird bewusst verzichtet, da solche Angaben in einer Online-Studie ohne Referenz und Hilfsmittel nur begrenzt zuverlässig wären und für die Beantwortung der Forschungsfragen nicht zwingend notwendig sind. Verfahren wie Pairwise- oder Forced-Choice-Vergleiche, bei denen Darstellungen direkt gegeneinander bewertet werden, wurden ebenfalls nicht verwendet, da der Fokus dieser Arbeit auf der individuellen Beurteilung jedes einzelnen Szenarios liegt.

Insgesamt eignet sich diese Kombination aus subjektiven Ratings und relativen Tiefeneinschätzungen gut für ein webbasiertes Setting: Sie ist für Teilnehmende ohne Spezialhardware durchführbar, bildet sowohl den allgemeinen Eindruck der Tiefenwahrnehmung als auch einfache Tiefenrelationen ab und erlaubt eine quantitative Auswertung hinsichtlich visuellem Eindruck, Tiefenwirkung, Verzerrungen und allgemeiner Darstellungsqualität.

3.3 Stichprobe und Ablauf

Im Rahmen dieser Bachelorarbeit wurde eine Stichprobengröße von etwa 20 Personen angestrebt, was unter den gegebenen Bedingungen als realistisch angesehen wurde. Rekrutiert wurde über den Bekannten und Freundeskreis sowie über den Mailverteiler des Studiengangs Medientechnik. Die Online-Befragung wurde mit dem webbasierten UmfrageTool *LimeSurvey* durchgeführt, das die strukturierte Erhebung der Antwortdaten ermöglichte.

Zur Teilnahme galten folgende Ein- bzw. Ausschlusskriterien:

- Mindestalter 18 Jahre
- Nutzung eines internetfähigen Computers oder Laptops (keine Teilnahme über Smartphone oder Tablet), da die Tiefenwahrnehmung auf kleinen Displays eingeschränkt sein kann

- Grundlegende Erfahrung im Umgang mit Computer und Webbrowswer

Damit sollte sichergestellt werden, dass die Szenen technisch korrekt dargestellt werden und die Teilnehmenden die Bedienung ohne zusätzliche Anleitung bewältigen können.

Der Ablauf der Online-Studie gliedert sich in zwei Phasen, in denen die drei Szenarien jeweils nacheinander durchlaufen werden. In Phase 1 werden die Szenarien mit eingeschränkter Bewegungsmöglichkeit betrachtet. Die Teilnehmenden sollen sich nur begrenzt in der Szene bewegen, damit einerseits Bewegungsparallaxe als Tiefenhinweis entsteht, andererseits aber keine beliebigen Perspektivwechsel genutzt werden können, um Entfernung „trickreich“ abzuschätzen. Zudem soll so verhindert werden, dass sich Personen konkrete Abstände merken und diese Schätzungen anschließend einfach auf andere Szenarien übertragen, ohne die dort tatsächlich wahrgenommene Tiefe neu zu beurteilen.

Die Reihenfolge der Szenarien ist in beiden Phasen identisch: Zunächst wird Szenario 2 präsentiert, das erwartungsgemäß die schwächste Tiefenwirkung aufweist. Anschließend folgt Szenario 3 (Low-Poly-Rekonstruktion) und zum Schluss Szenario 1 als High-Poly-Basiszene mit der voraussichtlich überzeugendsten Tiefenwahrnehmung. Da insbesondere Szenario 1 längere Ladezeiten und einen höheren Ressourcenverbrauch aufweisen kann, wird für jedes Szenario zusätzlich ein Video auf YouTube bereitgestellt, das bei technischen Problemen als Alternative zur interaktiven Darstellung genutzt werden kann.

In Phase 1 werden ausschließlich relative Fragen gestellt (z. B. zu Entfernung oder Tiefenrelationen), die jeweils um eine kurze Begründung der Antwort ergänzt werden. Auf explizit subjektive Gesamturteile zur Szene wird in dieser Phase bewusst verzichtet. Diese folgen erst in Phase 2, in der sich die Teilnehmenden frei durch die Szenen bewegen können. In dieser zweiten Phase sollen die Szenarien detaillierter betrachtet werden, um anschließend die subjektive Tiefenwahrnehmung, den Realismus und eventuelle Verzerrungen zu bewerten. Die freie Bewegung soll dabei helfen, die Einschätzungen zu erleichtern und zu präzisieren, weil Entfernung und räumliche Staffelung gezielt aus unterschiedlichen Blickwinkeln überprüft werden können.

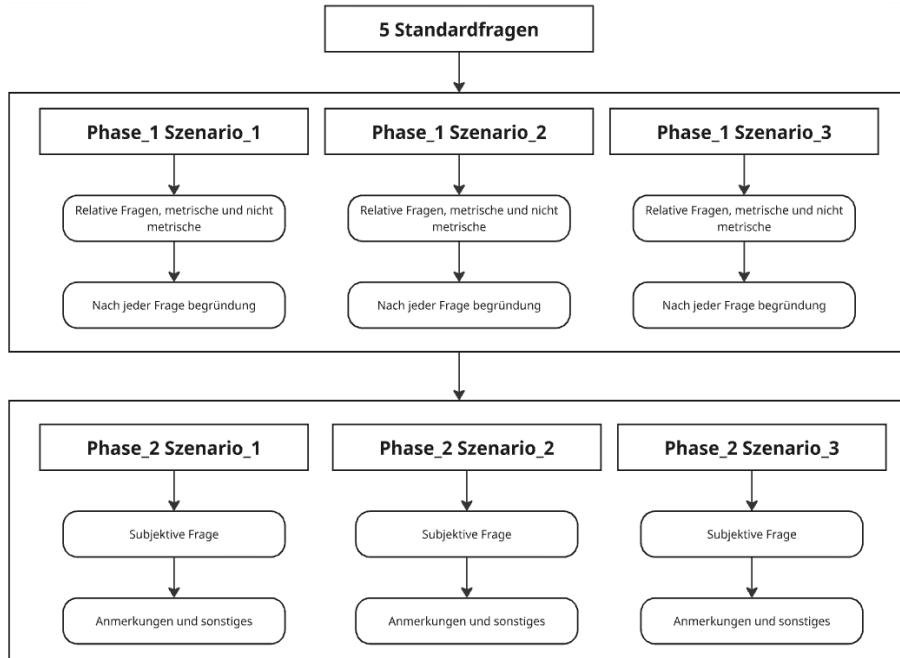


Abbildung: hallo ich bin schwul

3.4 Datenaufarbeitung und Auswertung

Die im Rahmen der Online-Studie erhobenen Daten wurden mit dem Umfragetool *LimeSurvey* erfasst und anschließend in tabellarischer Form exportiert. LimeSurvey speichert alle Antworten strukturiert, sodass jede teilnehmende Person einer Datenzeile und jede Frage einer Datenspalte entspricht.

Die numerischen Antworten der Likert-Skalen wurden deskriptiv ausgewertet. Für jede Frage und jedes Szenario wurden Mittelwerte sowie Streuungsmaße berechnet und in Tabellen und Balkendiagrammen dargestellt. Ziel dieser Auswertung ist die Beschreibung von Bewertungstendenzen innerhalb der einzelnen Szenarien, nicht die Durchführung inferenzstatistischer Tests.

Für die metrischen Distanzschätzungen aus Phase 1 wurden ergänzend reale Referenzwerte herangezogen. Diese absoluten Distanzen wurden direkt in der 3D-Szene in Blender bestimmt

und als Ground-Truth-Werte dokumentiert. Die mittleren Schätzwerte der Teilnehmenden wurden anschließend mit diesen realen Distanzen verglichen, um die Abweichung zwischen wahrgenommener und tatsächlicher Entfernung zu beurteilen. Die Ergebnisse werden dabei als relative Nähe zum Realwert (z. B. geringe oder größere Abweichung) interpretiert, nicht als exakte Genauigkeitsmessung.

Nicht-metrische relative Tiefenurteile (z. B. näher/weiter/zu ungenau) wurden über Häufigkeitsverteilungen ausgewertet. Ein hoher Anteil konsistenter Antworten wird als Hinweis auf eine stabile Tiefenrelation gewertet, während häufige „zu ungenau“-Angaben auf Unsicherheiten in der Tiefenwahrnehmung hindeuten.

Freitextantworten und Begründungen wurden qualitativ gesichtet und thematisch zusammengefasst. Wiederkehrende Beobachtungen, etwa zu Verzerrungen oder auffälligen Wahrnehmungseffekten, werden in den Ergebnissen exemplarisch aufgegriffen und zur Einordnung der quantitativen Resultate herangezogen.

3.5 Technische Umsetzung

3.5.1 Definition der Szenarien

Zuerst sollte eine Szenario definiert werden, um ein klares Verständnis zu schaffen. Alle Szenarien wurden mithilfe von Blender erstellt. Einige Modelle wurden selbst modelliert texturiert, wie zum Beispiel das Holzfällerhaus. Andere stammen aus dem Poly Haven Add-ons importiert welches eine Objekt Bibliothek bereitstellt (*Poly Haven Asset Browser*, 2025). Die Szene besteht aus einer Bodenfläche und vier thematisch unterschiedlichen Showcases, die in vier Blickrichtungen angeordnet sind. Diese Showcases umfassen eine Autogarage, ein mittelalterlich dekoriertes Zimmer, einen Pavillon mit Sitzplätzen und ein altes Holzfällerhaus. Die Anordnung wurde so gewählt, dass sich Betrachter:innen frei umsehen können und dadurch eine möglichst gute Einschätzung der räumlichen Tiefe ermöglicht wird. Die gesamte Szenerie befindet sich innerhalb einer HDRI-Kuppel, Licht wurde auch noch hinzugefügt. Diese Umgebung bildet das erste Szenario (Baseline), aus dem die Szenarien 2 und 3 abgeleitet werden.

3.5.2 Szenario 1: Baseline-3D-Rendering mit gezielten Tiefenherausforderungen

Szenario 1 dient als Basisline und ist einfach nur ein 3D Rendering das auf Babylon.js hochgeladen wird und gerändert wird. Zur Erstellung des Baseline-Szenarios wurde in Blender ein 3D-Modell entwickelt. Dabei kam das Add-on *Poly Haven Asset Browser* (*Poly Haven*

Asset Browser, 2025) zum Einsatz, das eine integrierte Bibliothek hochauflösender, texturierter 3D-Modelle bereitstellt, die per Drag-and-drop in die Szene eingefügt werden können. Anschließend wurden alle Modelle zusammengeführt und als .glb-Datei exportiert, das Format eignet sich besonders für 3d Modelle, die man im Web-Visualisieren möchte (glTF - Runtime 3D Asset Delivery, 2020). Die Datei wurde anschließend auf Babylon.js eingebettet und dort dargestellt. Hierbei wurde darauf wertgelegt das die Szene den Bewertungsansprüchen entspricht.

Kommentiert [El13]: Überflüssig gehört nicht in Methode.

Kommentiert [Inep, Eni14]: Absichtlich noch schwammig, da ich noch nicht konkret weiß wie bewertet wird.

3.5.3 Szenario 2: Depth Map + 360°-Panorama (Babylon.js)

Aufbauend auf dem Basisszenario (Szenario 1) wurde für Szenario 2 ein 360°-Panorama der bestehenden 3D-Szene erstellt. Hierzu wurde die vollständige Szenerie in Blender gerendert, wobei die Kamera auf eine sphärische Panorama-Projektion eingestellt wurde. Das Panorama wurde in equirektangularer Projektion mit einem Seitenverhältnis von 2:1 erzeugt, um eine korrekte Abbildung auf eine Kugelgeometrie in der Webdarstellung zu ermöglichen.

Parallel zum Farbrendering wurde eine Depth Map der Szene erzeugt. Dazu wurde im View-Layer der Z-Depth-Pass aktiviert und im Compositing-Modus von Blender aus dem Render-Layer extrahiert. Die Tiefenwerte wurden anschließend normalisiert und invertiert, sodass nahe Bereiche hell und entfernte Bereiche dunkel dargestellt werden. Das resultierende Tiefenbild wurde als Graustufenbild exportiert und gemeinsam mit dem Panorama gespeichert.

Sowohl das 360°-Panorama als auch die zugehörige Depth Map wurden anschließend auf einem externen Server abgelegt, um einen direkten Zugriff über feste URLs zu ermöglichen. Diese Bildressourcen wurden in Babylon.js geladen und auf eine segmentierte Kugelgeometrie projiziert. Die Depth Map wurde dabei genutzt, um eine pseudo-räumliche Verformung der Oberfläche zu erzeugen, wodurch ein Tiefeneindruck entsteht.

Innerhalb der Webszene wurden zusätzlich Kamerasteuerung, Beleuchtung sowie Parameter zur Stärke der Tiefenverformung implementiert. Das Szenario konnte dadurch interaktiv betrachtet werden und diente als Beispiel für ein texturbasiertes 2.5D-Verfahren mit sichtbaren Verzerrungen bei Bewegung.

3.5.4 Szenario 3: Low-Poly-Rekonstruktion aus Szenario 1

Szenario 3 basiert auf einer vereinfachten geometrischen Rekonstruktion des Basisszenarios (Szenario 1). Ziel war es, die vollständige dreidimensionale Tiefeninformation der Szene

beizubehalten, gleichzeitig jedoch die Performance und Dateigröße für die webbasierte Darstellung deutlich zu reduzieren.

Selbst modellierte Objekte lagen bereits in einer Low-Poly-Grundform vor und wurden in Szenario 1 lediglich durch Modifier zu High-Poly-Geometrie erweitert. Für Szenario 3 wurden diese Modifier gezielt deaktiviert, um wieder auf die ursprüngliche Low-Poly-Struktur zurückzugreifen. Zur effizienten Verwaltung der Modifier über mehrere Objekte hinweg wurde ein skriptbasierter Workflow eingesetzt.

Hochauflöste Assets aus externen Quellen wurden mithilfe des *Decimate*-Modifiers geometrisch vereinfacht. Dieses Verfahren stellte einen praktikablen Kompromiss zwischen visueller Qualität und Arbeitsaufwand dar, da alternative Methoden zur manuellen Retopologie den zeitlichen Rahmen der Arbeit überschritten hätten. Die resultierenden Low-Poly-Modelle wiesen eine ausreichende visuelle Qualität für die geplante Evaluation auf.

Neben der Reduktion der Polygonanzahl wurde auch die Texturauflösung optimiert. Die ursprünglichen hochauflösenden Texturen wurden in einem externen Textur-Workflow neu exportiert und in komprimierter Form mit reduzierter Auflösung eingebunden. Dadurch konnte die Gesamtdateigröße der Szene erheblich verringert werden, ohne den grundlegenden visuellen Eindruck maßgeblich zu beeinträchtigen.

Das final optimierte Low-Poly-Modell wurde anschließend erneut in Blender zusammengeführt und als GLB-Datei exportiert. Diese wurde auf einem externen Server bereitgestellt und in Babylon.js eingebettet. In der Webdarstellung kamen zusätzliche Rendering-Optimierungen wie Screen-Space Ambient Occlusion (SSAO) und Backface Culling zum Einsatz, um die visuelle Tiefenwirkung zu unterstützen und gleichzeitig die Laufzeitperformance zu verbessern.

Szenario 3 dient damit als Referenz für ein performantes, vollständig geometriebasiertes Verfahren zur Tiefendarstellung im Web, das im Gegensatz zu Szenario 2 keine rein bildbasierten Tiefeninformationen nutzt.

4 Ergebnisse

- 4.1 Technische Ergebnisse der Szenario-Umsetzungen**
- 4.2 Beobachtete Optimierungseffekte (z. B. weniger Artefakte, visuelle Glätte)**
- 4.3 Ergebnisse der Nutzerbefragung (Darstellung in Tabellen, Diagrammen)**
- 4.4 Zusammenfassung der Ergebnisse in Bezug auf Forschungsfrage**

5 Diskussion

5.1.1 Interpretation der Ergebnisse im Kontext der Forschungsfrage

5.1.2 Vergleich mit Literatur(?)

5.1.3 Bewertung der Verfahren: Stärken, Grenzen, Anwendungspotenzial

6 Fazit und Ausblick

6.1 Beantwortung der Forschungsfrage

6.2 Beitrag der Arbeit

6.3 Offene Fragen und Ausblick

7 Anhang

Quellen:

Adobe. (2025). *What is Baking? | Substance 3D bakers.*

<https://helpx.adobe.com/content/help/en/substance-3d-bake/getting-started/what-is-baking.html>

Ashtiani, O., Guo, H.-J., & Prabhakaran, B. (2023). Impact of motion cues, color, and luminance on depth perception in optical see-through AR displays. *Frontiers in Virtual Reality*, 4, 1243956. <https://doi.org/10.3389/frvir.2023.1243956>

Blender Foundation. (2025a, Oktober 5). *Glossary—Blender 4.5 LTS Manual.*
<https://docs.blender.org/manual/en/latest/glossary/index.html#term-Z-buffer>

Blender Foundation. (2025b, Oktober 19). *Render Baking—Blender 4.5 LTS Manual.*
<https://docs.blender.org/manual/en/latest/render/cycles/baking.html>

Choi, H., & Nam, S. (2025). Effect of eye-tracking-based depth perception visualization interface on virtual reality user experience. *Virtual Reality*, 30(1), 5.
<https://doi.org/10.1007/s10055-025-01260-8>

Committeri, G., Sebastiani, V., De Pasquale, F., Stocchi, M., & Fini, C. (2020). Functional Autonomy Affects Elderly Spatial Perception in Body-Centered Coordinates. *Journal of Aging Research*, 2020, 1–8. <https://doi.org/10.1155/2020/5694790>

Divyanth, L. G., Rathore, D., Senthilkumar, P., Patidar, P., Zhang, X., Karkee, M., Machavaram, R., & Soni, P. (2023). Estimating depth from RGB images using deep-learning for robotic applications in apple orchards. *Smart Agricultural Technology*, 6, 100345. <https://doi.org/10.1016/j.atech.2023.100345>

gltf - Runtime 3D Asset Delivery. (2020, Dezember 3). The Khronos Group.

<https://www.khronos.org/gltf/>

He, L., Jian, B., Wen, Y., Zhu, H., Liu, K., Feng, W., & Liu, S. (2022). Rethinking Supervised

Depth Estimation for 360° Panoramic Imagery. *2022 IEEE/CVF Conference on*

Computer Vision and Pattern Recognition Workshops (CVPRW), 5169–5177.

<https://doi.org/10.1109/CVPRW56347.2022.00565>

Hergovich, A. (2022a). *Allgemeine Psychologie: Wahrnehmung und Emotion* (3. Aufl.). utb

GmbH. <https://doi.org/10.36198/9783838554341>

Hergovich, A. (2022b). *Allgemeine Psychologie: Wahrnehmung und Emotion* (3. Aufl.). utb

GmbH. <https://doi.org/10.36198/9783838554341>

Hibbard, P. B., Hornsey, R. L., & Asher, J. M. (2023). Binocular Information Improves the

Reliability and Consistency of Pictorial Relief. *Vision*, 7(1), 1.

<https://doi.org/10.3390/vision7010001>

Hoffmann, P. (2025). *Beyond (Multi-) Media: Multimediaformen erklärt: Von*

Panoramabildern über 3D bis zu den immersiven Welten des Metaversums. Springer

Fachmedien Wiesbaden. <https://doi.org/10.1007/978-3-658-48567-2>

Ito, S., Miura, K., Ito, K., & Aoki, T. (2024). Neural Radiance Field-Inspired Depth Map

Refinement for Accurate Multi-View Stereo. *Journal of Imaging*, 10(3), 68.

<https://doi.org/10.3390/jimaging10030068>

Izadi, S., Kim, D., Hilliges, O., Molyneaux, D., Newcombe, R., Kohli, P., Shotton, J., Hodges,

S., Freeman, D., Davison, A., & Fitzgibbon, A. (2011). KinectFusion: Real-time 3D

reconstruction and interaction using a moving depth camera. *Proceedings of the 24th*

Annual ACM Symposium on User Interface Software and Technology, 559–568.

<https://doi.org/10.1145/2047196.2047270>

Jähne, B. (2024). *Digitale Bildverarbeitung: Und Bildgewinnung*. Springer Berlin Heidelberg.

<https://doi.org/10.1007/978-3-662-59510-7>

Jan, A., Khan, S., & Suyoung, S. (2023). Deep Learning-based Depth Map Estimation: A Review. *Korean Journal of Remote Sensing*, 39(1), 1–21.

<https://doi.org/10.7780/kjrs.2023.39.1.1>

Lesar, Ž., Bohak, C., & Marolt, M. (2024). Evaluation of depth perception in crowded volumes.

Computers & Graphics, 120, 103918. <https://doi.org/10.1016/j.cag.2024.103918>

Lipski, C., Klose, F., & Magnor, M. (2014). Correspondence and Depth-Image Based Rendering a Hybrid Approach for Free-Viewpoint Video. *IEEE Transactions on Circuits and Systems for Video Technology*, 24(6), 942–951.

<https://doi.org/10.1109/TCSVT.2014.2302379>

Miyashita, Y., Sawahata, Y., Sakai, A., Harasawa, M., Hara, K., Morita, T., & Komine, K. (2022). Display-Size Dependent Effects of 3D Viewing on Subjective Impressions.

ACM Transactions on Applied Perception, 19(2), 1–15.

<https://doi.org/10.1145/3510461>

Poly Haven Asset Browser. (2025, September). Superhive (Formerly Blender Market).

<https://superhivemarket.com/products/poly-haven-asset-browser?ref=3841>

Rey-Area, M., Yuan, M., & Richardt, C. (2022). *360MonoDepth: High-Resolution 360° Monocular Depth Estimation* (No. arXiv:2111.15669). arXiv.

<https://doi.org/10.48550/arXiv.2111.15669>

Saha, A., Dhara, B. C., Umer, S., Yurii, K., Alanazi, J. M., & AlZubi, A. A. (2022). Efficient Obstacle Detection and Tracking Using RGB-D Sensor Data in Dynamic Environments

for Robotic Applications. *Sensors (Basel, Switzerland)*, 22(17), 6537.

<https://doi.org/10.3390/s22176537>

Schmeing, M., & Jiang, X. (o. J.). *Depth Image Based Rendering*. https://doi.org/10.1007/978-3-642-22407-2_12

Schönhammer, R. (2013). *Einführung in die Wahrnehmungspychologie: Sinne, Körper, Bewegung* (2., überarb., aktualisierte u. erw. Aufl). Facultas.wuv.

Silveira, T. L. T. da, Pinto, P. G. L., Llerena, J. E. M., & Jung, C. R. (2023). 3D Scene Geometry Estimation from 360° Imagery: A Survey. *ACM Computing Surveys*, 55(4), 1–39. <https://doi.org/10.1145/3519021>

Singh, S., Sharma, Y., Liaqat, A., & Kalawsky, R. S. (2025). Evaluation of XR device's real-world tracking accuracy and depth perception from an industrial point of view. *Virtual Reality*, 29(3), 118. <https://doi.org/10.1007/s10055-025-01192-3>

Wang, G., Wang, P., Chen, Z., Wang, W., Loy, C. C., & Liu, Z. (2023). *PERF: Panoramic Neural Radiance Field from a Single Panorama* (No. arXiv:2310.16831). arXiv. <https://doi.org/10.48550/arXiv.2310.16831>

What Is Depth Anything V2: Depth Estimation Tutorials. (2025, Juli 24). Roboflow Blog. <https://blog.roboflow.com/depth-anything/>

Wölfel, M. (2023). *Immersive Virtuelle Realität: Grundlagen, Technologien, Anwendungen*. Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-662-66908-2>

Zabih, R., & Woodfill, J. (1994). Non-parametric local transforms for computing visual correspondence. In J.-O. Eklundh (Hrsg.), *Computer Vision—ECCV '94* (Bd. 801, S. 151–158). Springer Berlin Heidelberg. <https://doi.org/10.1007/BFb0028345>

Zeng, W., Karaoglu, S., & Gevers, T. (2020). Joint 3D Layout and Depth Prediction from a Single Indoor Panorama Image. In A. Vedaldi, H. Bischof, T. Brox, & J.-M. Frahm

(Hrsg.), *Computer Vision – ECCV 2020* (Bd. 12361, S. 666–682). Springer International Publishing. https://doi.org/10.1007/978-3-030-58517-4_39

Zhang, C., He, B., Guo, R., & Ma, D. (2023). When a tree model meets texture baking: An approach for quality-preserving lightweight visualization in virtual 3D scene construction. *International Journal of Digital Earth*, 16(1), 645–670. <https://doi.org/10.1080/17538947.2023.2177758>