

# **Untersuchung dreier Verfahren zur Darstellung von Panoramabildern mit visuell optimierter Tiefenwahrnehmung und ihren Grenzen**

Bachelorarbeit  
im Studiengang Medientechnik  
des Fachbereichs Technik  
an der Hochschule Emden/Leer

Eingereicht von  
Enis Inep (Matrikelnummer: 7022248)

Erstprüfer:  
Prof. Dr. Thies Pfeiffer

Zweitprüfer:  
M. Eng. Jannik Franssen

Emden, den 30. Dezember 2025

## **Abstract**

Die Vermittlung räumlicher Tiefe in webbasierten 3D-Anwendungen auf 2D-Monitoren stellt eine technische und gestalterische Herausforderung dar. Diese Arbeit untersucht drei Verfahren zur Darstellung von 360°-Panoramabildern mit visuell optimierter Tiefenwahrnehmung. Ein vollständig geometriebasiertes High Poly Szenario, ein bildbasiertes Panorama mit Depth Map sowie eine Low Poly Rekonstruktion mit gebackenen Texturen. Die Szenarien wurden in einer webbasierten Umgebung umgesetzt und in einer explorativen Online-Studie mittels subjektiver Bewertungen sowie relativer Tiefeneinschätzungen verglichen. Die Ergebnisse zeigen, dass High Poly Darstellungen eine konsistente Tiefenwahrnehmung ermöglichen, jedoch für den Webeinsatz ungeeignet sind. Depth Map basierte Verfahren sind sehr effizient, weisen jedoch wahrnehmbare Verzerrungen auf. Low Poly Rekonstruktionen stellen einen ausgewogenen Kompromiss zwischen visueller Qualität und technischer Effizienz dar.

## **Danksagung**

An dieser Stelle möchte ich mich herzlich bei allen bedanken, die mich während der Anfertigung dieser Bachelorarbeit unterstützt haben. Ein besonderer Dank gilt meinem Zweitbetreuer M. Eng. Jannik Franssen, dessen regelmäßiges und konstruktives Feedback eine große Hilfe darstellte und mich während des Arbeitsprozesses fachlich begleitet hat. Zudem möchte ich mich bei meinem Hauptbetreuer Prof. Dr. Thies Pfeiffer für die Betreuung der Arbeit und die fachliche Begleitung bedanken. Darüber hinaus danke ich meinen Kommilitoninnen, Kommilitonen sowie Freunden, die mich während des Studiums und insbesondere in der Phase der Bachelorarbeit unterstützt, motiviert und bei fachlichen Diskussionen begleitet haben. Ein besonderer Dank gilt meiner Familie, die mich stets im Alltag und während des Studiums unterstützt und mir den notwendigen Rückhalt gegeben hat. Abschließend möchte ich mich bei allen bedanken, die mich mit ihrer Zeit, Geduld und Unterstützung bei der Fertigstellung dieser Abschlussarbeit begleitet haben.

# Inhaltsverzeichnis

<b>Abstract</b>	I
<b>Danksagung</b>	II
<b>Abbildungsverzeichnis</b>	V
<b>1 Einleitung</b>	1
1.1 Motivation und Relevanz des Themas . . . . .	1
1.2 Zielsetzung und Forschungsfragen . . . . .	2
1.3 Aufbau der Arbeit . . . . .	3
<b>2 Theoretische Grundlagen</b>	5
2.1 Grundlagen der Tiefenwahrnehmung . . . . .	5
2.1.1 Grenzen der Tiefenwahrnehmung auf 2D-Displays . . . . .	6
2.2 360° Panoramen als Darstellungsformat . . . . .	6
2.2.1 Equirektanguläre Projektion und geometrische Verzerrungen . . . . .	7
2.3 Texturbasierte Verfahren der Tiefendarstellung . . . . .	9
2.3.1 Texture Baking . . . . .	9
2.3.2 Depth Maps . . . . .	10
2.3.3 Low Poly Erstellung und Texturoptimierung für webbasierte 3D-Darstellungen . . . . .	11
2.4 Überblick bestehender Ansätze . . . . .	12
2.4.1 Spezifische Verfahren im Kontext 360° Panoramen . . . . .	12
2.4.2 Einschränkungen und Forschungsbedarf . . . . .	13
2.4.3 Forschung zur Tiefenwahrnehmung auf 2D-Monitoren sowie in VR/AR-Systemen . . . . .	14
2.4.4 Forschung zu KI-generierten Low Poly-Modellen . . . . .	15
<b>3 Methodik</b>	17
3.1 Überblick der Messmethoden . . . . .	17
3.1.1 Subjektive Tiefenwahrnehmung . . . . .	17
3.1.2 Verhaltensbasierte Tiefeneinschätzungen . . . . .	18
3.2 Analyse und Bewertung der Szenarien . . . . .	18
3.3 Stichprobe und Ablauf . . . . .	19
3.4 Datenaufarbeitung und Auswertung . . . . .	21
3.4.1 Grundlegender Auswertungsansatz . . . . .	21

<b>4 Implementierung</b>	<b>24</b>
4.1 Technische Umsetzung . . . . .	24
4.1.1 Szenario 1: Baseline-3D-Rendering mit gezielten Tiefenherausforderungen . . . . .	25
4.1.2 Szenario 2: Depth Map und 360°-Panorama (Babylon.js) . . . . .	26
4.1.3 Szenario 3: Low Poly Rekonstruktion aus Szenario 1 . . . . .	28
<b>5 Ergebnisse</b>	<b>32</b>
<b>6 Diskussion</b>	<b>36</b>
6.1 Einordnung . . . . .	36
6.1.1 Szenario 2: Panorama und Depth map . . . . .	36
6.1.2 Szenario 3: Low Poly Rekonstruktion . . . . .	38
6.1.3 Szenario 1: Baseline High Poly . . . . .	41
<b>7 Fazit</b>	<b>43</b>
<b>Literaturverzeichnis</b>	<b>46</b>
<b>Anhang</b>	<b>51</b>
<b>A Hinweis zur Nutzung von KI-Werkzeugen Und</b>	<b>52</b>
<b>B Hinweis zu begleitenden Materialien und Reproduzierbarkeit</b>	<b>53</b>

# Abbildungsverzeichnis

2.1	Beispiel für eine verzerrte equirektangulare Projektion . . . . .	8
2.2	Schematische Darstellung der Abbildung eines 360° Panoramas auf eine equirektangulare 2:1 Projektion, wobei die Pfeile exemplarisch die Ausdehnung der räumlichen Umgebung sowie die Zuordnung der sphärischen Umgebung zu den horizontalen und vertikalen Bildrichtungen markieren.	8
2.3	Marktentwicklung von VR-Headsets weltweit. Quelle: Statista . . . . .	15
4.1	Showcase A–D . . . . .	25
4.2	Render Ergebnisse von Szenario 2: Panorama Darstellung und zugehörige Tiefeninformation . . . . .	27
4.3	Erster Ansatz für Szenario 3 . . . . .	29
4.4	Beispiel: für Ausgelegtes UV-Layout eines Low Poly Modells . . . . .	30
4.5	Vergleich der Darstellung mit aktiviertem und deaktiviertem Backface Culling in Babylon.js Babylon.js Documentation [2025a] . . . . .	31
5.1	Auswertung relative metrische Schätzungsfragen . . . . .	32
5.2	Auswertung relative Schätzungsfragen . . . . .	33
5.3	Ergebnisse subjektive Einschätzungen, Median . . . . .	34
5.4	Ergebnisse subjektive Einschätzungen, Hauptanteil . . . . .	34
5.5	Ergebnisse subjektive Einschätzungen, Streuung . . . . .	35
6.1	Verzerrung im Spektrum von 0-255 . . . . .	37
6.2	Erster Ansatz für Szenario 3 . . . . .	39
6.3	Darstellung des Meshes ohne Farbgebung . . . . .	39
6.4	Farbige Darstellung derselben Szene mit identischer Geometrie . . . . .	40
6.5	Finales Szenario 2 mit aktivierten Renderingverfahren (Backface Culling und SSAO). . . . .	41
6.6	Finales Basisszenario S1 mit vollständig geometrischer Darstellung . . . . .	42

# 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

Die Verwendung von 360°-Umgebungen und webbasierten 3D-Darstellungen gewinnt in vielen Anwendungsfeldern an Bedeutung, etwa in digitalen Spielen, Lernumgebungen oder kartografischen Anwendungen wie Street View. Trotz der zunehmenden Verbreitung entsprechender Technologien besteht weiterhin Forschungsbedarf hinsichtlich der Frage, wie räumliche Tiefe auf 2D-Monitoren überzeugend vermittelt werden kann, insbesondere unter den technischen Einschränkungen webbasierter Anwendungen.

Die Herausforderung, flache Darstellungen durch geeignete Verfahren räumlicher wirken zu lassen, ist dabei sowohl aus technischer als auch aus gestalterischer Perspektive relevant. Da entsprechende Lösungsansätze bislang nur begrenzt systematisch untersucht wurden, ergibt sich ein geeigneter Ansatzpunkt für weiterführende Forschung im Bereich der Medientechnik.

Die Themenwahl dieser Arbeit ist zudem durch eine persönliche Motivation geprägt. Das Themenfeld wurde mir im Rahmen meines Studiums an der Hochschule Emden/Leer erstmals nähergebracht und deckt sich stark mit meinen eigenen fachlichen Interessen. Ziel dieser Arbeit ist es daher, verschiedene Verfahren zur Darstellung räumlicher Tiefe praktisch umzusetzen und deren subjektive Wahrnehmung systematisch zu untersuchen. Die Ergebnisse sollen dazu beitragen, das Potenzial der einzelnen Ansätze für weiterführende

Optimierungen einschätzen zu können.

## 1.2 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 über das Web in einem Browser laufen und dort auch technisch sinnvoll performen. Eines dieser drei 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. Im Spezifischen sollen die Stärken und 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 (Depth map) 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.

**H3:** Für Szenario 3 (Low Poly) wird eine recht geringe Ladezeit und ein niedriger Ressourcenverbrauch erwartet.

### 1.3 Aufbau der Arbeit

Die drei Herangehensweisen werden auf Basis einer 3D-modellierten Szene umgesetzt. Dabei wird jedes Szenario individuell für seine jeweiligen Anforderungen optimiert.

Das Basisszenario S1 ist ein 3D-Rendering, in dem Objekte bewusst so angeordnet sind, dass Grenzen der Tiefenwahrnehmung ausgereizt werden. Szenario 1 dient hierbei größtenteils als Referenzvorlage und Basis. 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 High Poly Rendering wie in S1 zeigt zwar hohe Detailtreue, bringt aber Herausforderungen hinsichtlich Performance und Dateigröße mit sich, insbesondere in einer Webumgebung.

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

Szenario 3 S3 verfolgt den Ansatz, ein Low Poly 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 nicht sichtbare oder für die Tiefenwahrnehmung irrelevante Flächen gezielt vermieden werden können. Da die konkrete Herangehensweise für S3 noch nicht endgültig festgelegt ist, besteht eine Herausforderung darin, eine geeignete Methode zur Low Poly Erzeugung zu identifizieren. 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 Depth Maps dienen und welche Funktionen sie bieten.

### 2.1 Grundlagen der Tiefenwahrnehmung

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 für die monokulare 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, 2022, S.76]. Dieses Phänomen wird als Bewegungsparallaxe (Parallax Effect) bezeichnet und ist ein klassischer monokularer Tiefenhinweis, da bereits ein einzelnes Auge aus relativen Bewegungen Tiefenunterschiede erschließen kann [Hergovich, 2022, 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., 2023-03, 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.

### 2.1.1 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 [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 erzeugt werden. Auch eine gezielte Nutzung von Licht und Schatten oder eine hierarchische Anordnung der Objekte, bei der weiter entfernte Objekte durch davor platzierte Elemente teilweise verdeckt werden und dadurch Tiefenbeziehungen erkennbar werden [Hergovich, 2022, S.77].

Darüber hinaus kann durch 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 [Miyashita et al., 2022, S.5:2-5:3]. 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 als auch 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 werden 360° Panoramen häufig in VR-Anwendungen oder im Architekturbereich zur Simulation

eingesetzt, da sie in diesen Anwendungsfällen einen besonders hohen Immersionsgrad ermöglichen. Diese Darstellungsform ist besonders hilfreich, um die Umgebung von einem zentralen Punkt aus wahrzunehmen. In der Regel sehen wir unser Sichtfeld um uns herum, ein Panorama bildet dieses Sichtfeld auf einer flachen Ebene ab und erzeugt damit eine Verzerrung [Hoffmann, 2025, S.21-24]. In webbasierten 3D-Anwendungen werden 360° Panoramen in der Regel in equirektangularer Projektion gespeichert. Dabei wird die sphärische Umgebung in eine zweidimensionale, rechteckige Bildfläche mit einem festen Seitenverhältnis von 2:1 transformiert, die anschließend wieder auf eine Kugelgeometrie projiziert werden kann und so eine immersive Rundumsicht ermöglicht Blender Foundation [2025a].

### 2.2.1 Equirektangularare Projektion und geometrische Verzerrungen

Die visuelle Umgebung kann geometrisch als Kugeloberfläche um einen festen Beobachtungspunkt beschrieben werden, da 360° Panoramen den vollständigen Raum in alle Richtungen erfassen. Jeder Punkt innerhalb dieses Raumes lässt sich eindeutig durch Richtungsvektoren oder sphärische Koordinaten beschreiben. Soll diese sphärische Umgebung in Form eines Bildes gespeichert oder dargestellt werden, ist eine Abbildung auf eine zweidimensionale Fläche erforderlich. Bei einer equirektangularen Projektion wird die sphärische Umgebung auf eine rechteckige Bildfläche abgebildet, wobei die Längengrade linear auf die horizontale x-Achse und die Breitengrade linear auf die vertikale y-Achse gemappt werden. Dadurch entsteht ein Bild mit einem Seitenverhältnis von 2:1 Blender Foundation [2025a], Google Developers [2023]. Durch die Übertragung der sphärischen Geometrie auf eine ebene Fläche treten insbesondere im oberen und unteren Bildbereich geometrische Verzerrungen auf [Hoffmann, 2025, S.21–24]. Diese Verzerrungen entstehen, da die lineare Abbildung der Breitengrade die tatsächlichen Flächenverhältnisse der Kugel nicht erhält, wodurch Bildinhalte zu den Polen hin gestaucht und in die Randbereiche des Bildes abgewickelt werden [Snyder and Voxland, 1989, S.24].

Mathematisch beschrieben:

$$x = \lambda, \quad y = \varphi$$

Dabei gilt für den Längengrad:  $\lambda \in [-180^\circ, 180^\circ]$

und für den Breitengrad:  $\varphi \in [-90^\circ, 90^\circ]$ .

Der Maßstab  $M$  wächst mit der Breite:

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

Am Äquator ( $\varphi = 0^\circ$ ) ergibt sich ein Maßstab von  $M = 1$ , sodass keine Verzerrung auftritt. In Richtung der Pole ( $\varphi = \pm 90^\circ$ ) wächst der Maßstab gegen unendlich, wodurch dort die stärksten Verzerrungen entstehen.



Abbildung 2.1: Beispiel für eine verzerrte equirektanguläre Projektion

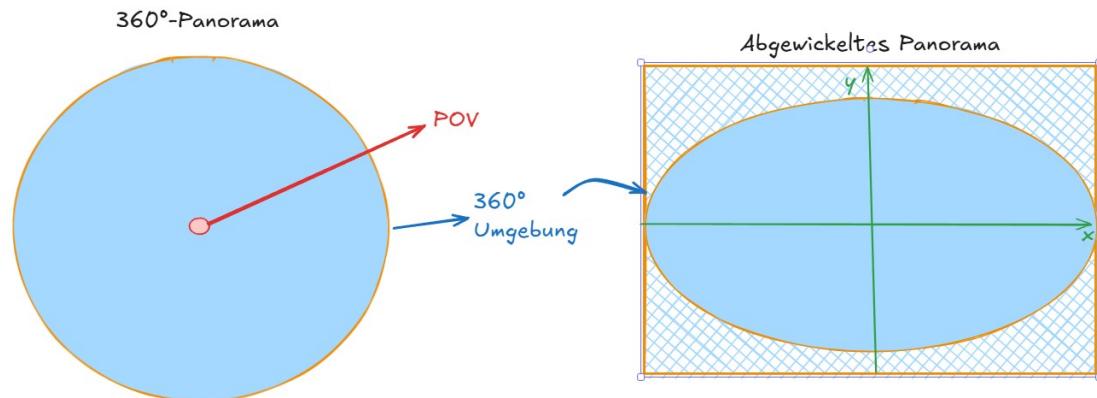


Abbildung 2.2: Schematische Darstellung der Abbildung eines  $360^\circ$  Panoramas auf eine equirektanguläre 2:1 Projektion, wobei die Pfeile exemplarisch die Ausdehnung der räumlichen Umgebung sowie die Zuordnung der sphärischen Umgebung zu den horizontalen und vertikalen Bildrichtungen markieren.

## 2.3 Texturbasierte Verfahren der Tiefendarstellung

### 2.3.1 Texture Baking

Texture Baking ist ein Verfahren, bei dem Rendering-Ergebnisse eines 3D-Modells auf eine zweidimensionale Bitmap-Textur übertragen werden, um rechenintensive Berechnungen zur Laufzeit zu vermeiden [Zhang et al., 2023, S.647–648]. Texture Baking wird oft für Videospiele, Web-basierte oder VR/AR Anwendungsfälle verwendet, da es Speicher- und Rechenressourcen sparen kann. Andernfalls müssten 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 drei verschiedenen Texturen zu einer kombinieren [Knodt et al., 2023]. So werden beispielsweise für unterschiedliche Materialien gemeinsame Texturatlanten pro Map-Typ (z. B. Base Color, Normal, Roughness, Metallic) erzeugt und gespeichert [Blender Foundation, 2025b]. Das reduziert den Speicherbedarf und auch den Rechenaufwand bei der Laufzeit. Um Texture Baking im 3D-Modellierungsprogramm Blender anzuwenden, braucht das 3D-Modell eine UV-Map und eine „Image Texture Node“, um die Ergebnisse zu speichern [Blender Foundation, 2025b]. Der Prozess läuft ähnlich wie ein Renderingverfahren ab, nur mit dem Unterschied, dass das Ergebnis nicht in einem Framebuffer, sondern auf eine Textur gespeichert wird [Blender Foundation, 2025b].

In Blender lassen sich zwei grundlegende Baking Verfahren unterscheiden. Im ersten Verfahren wird von einem Quellobjekt auf ein Zielobjekt gebacken. Gängig sind hierbei Projektionen von einem High Poly Objekt auf ein entsprechendes Low Poly Objekt. Dieses Vorgehen ist die am häufigsten verwendete Methode, da sich damit verschiedene Texture Maps wie Normal Maps, 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 können komplexe Node Strukturen reduziert werden oder statische Lichtinformationen, beispielsweise Lightmaps, direkt in das Objekt integriert werden [Blender Foundation, 2025b].

Das Verfahren erfordert eine saubere Topologie und korrekte UV-Koordinaten, da jede geometrische Inkonsistenz (z. B. überlappende Flächen, invertierte Normalen oder nicht-manifold Geometrie) zu Fehlberechnungen führen kann. Wichtig ist hierbei, dass die formgebenden Kanten im Low Poly Modell erhalten bleiben, um die genannten Probleme zu vermeiden.

### 2.3.2 Depth Maps

Eine Depth Map ist ein Bild, das Tiefeninformationen einer Szene speichert [Jähne, 2024, S. 271–272]. Sie gehört wie Normal Maps oder Bump Maps zur Kategorie der Texture Maps. Gespeichert wird dabei für jedes Pixel der Abstand von der Kamera zur jeweils ersten sichtbaren Oberfläche entlang der entsprechenden Blickrichtung [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 jedoch als Standard [Divyanth et al., 2023, S.2-3]. Farbbilder werden üblicherweise über drei Farbkanäle (RGB) dargestellt, wobei jeder Kanal einen Intensitätswert speichert. Eine Depth Map selbst enthält keine Farbinformationen, sondern speichert pro Pixel einen einzelnen Tiefenwert. Dieser wird häufig als Graustufenbild kodiert, wobei je nach Anwendung unterschiedliche Bittiefen verwendet werden (z. B. 8-Bit, 16-Bit oder 32-Bit), um den Abstand zur Kamera mit entsprechender Genauigkeit abzubilden Roboflow [2025].

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]. In 3D-Grafiksystemen können Depth Maps direkt während des Renderprozesses erzeugt werden, indem für jedes Pixel der Abstand zur Kamera entlang der Sichtachse bestimmt wird. Diese Tiefenwerte werden typischerweise im sogenannten Z-Buffer gespeichert und ergeben in ihrer Gesamtheit eine Depth Map Blender Foundation [2025b].

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 and Jiang, 2011, 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].

Schwierigkeiten bei der Tiefeneinschätzung treten insbesondere bei spiegelnden oder transparenten Oberflächen auf. Dieses Problem ist jedoch nicht auf Depth Maps beschränkt, sondern betrifft allgemein Verfahren zur optischen Tiefenerfassung in realen Umgebungen. Depth Maps erfassen zudem ausschließlich die jeweils vordersten sichtbaren Oberflächen entlang der Blickrichtung und liefern dadurch nur auf den aktuellen Beobachtungspunkt beschränkte Tiefeninformationen. Bei KI-basierten Verfahren kommen zusätzlich Rauschen und Schätzungsfehler hinzu, insbesondere bei komplexen Materialien oder Geometrien [Divyanth et al., 2023, S. 6–7]. Dennoch eröffnen sich durch den Gebrauch von Depth Maps vielseitige Einsatzmöglichkeiten. Kombinierte Depth Maps aus verschiedenen Blickwinkeln können zu einer dichten 3D-Rekonstruktion verhelfen [Izadi et al., 2011, S.1]. Neuerdings nutzen Open-Source-Frameworks wie NeRF (Neural Radiance Fields) Depth Maps zur Rekonstruktion realer Umgebungen, indem aus bildbasierten Aufnahmen konsistente, fotorealistische 3D-Szenen erzeugt werden, die insbesondere für das Scannen und die virtuelle Präsentation real existierender Räume eingesetzt werden [Ito et al., 2024-03-08, 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.

### **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].

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].

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

Im Rahmen der Literaturrecherche zeigte sich, dass bereits verschiedene Verfahren existieren, um Tiefe in zweidimensionalen Panoramen darzustellen. Eine Studie von 2007 zeigt, wie Layers verwendet werden können, um Tiefe in Panoramabildern 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 Layers befindet sich keine dreidimensionale Tiefe, der parallaxe Effekt erzeugt dennoch ein Gefühl von Tiefe. Falls sich aber den Layers genähert wird, wird deutlich, dass diese flach sind.

### 2.4.1 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., 2022-06, S.5173-5174] [Wang et al., 2023, S.1-2] [Silveira 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., 2022-06, S.5173–5174] [Rey-Area et al., 2022, S.1-2]. Durch dieses Training lernt das neuronale Netz, die Tiefenstruktur einer Szene aus einem oder 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 ressourcenschonende Tiefendarstellung abzielen, etwa zur visuellen Unterstützung in webbasierten, interaktiven Interfaces, sind diese Verfahren daher nur eingeschränkt geeignet. Die beschriebenen Deep-Learning Ansätze erfordern in der Regel hohen Rechenaufwand, umfangreiche Trainingsdaten und erzeugen häufig komplexe geometrische Repräsentationen mit entsprechend großem Speicherbedarf. Zudem sind viele dieser Verfahren primär auf Innenraumszenarien ausgelegt oder auf spezifische Annahmen zur Szenengeometrie beschränkt. Im Kontext dieser Arbeit wird daher ein stärker vereinfachter Ansatz verfolgt, bei dem nicht die vollständige geometrische Rekonstruktion einer Szene im Vordergrund steht, sondern eine visuell nachvollziehbare Tiefenwirkung bei gleichzeitig geringem Rechen- und Speicheraufwand.

#### 2.4.2 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. Die Qualität der Tiefenschätzung lässt sich mithilfe quantitativer Fehlermetriken bewerten. Eine gängige Kennzahl ist dabei der Root Mean Squared Error (RMSE), der angibt, wie stark die vorhergesagten Tiefenwerte von den Referenzwerten (Ground Truth) abweichen. 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 auch gemessen werden, wie viele Pixel ungefähr richtig sind. Das führt dazu, dass Tiefe eine Eindeutige, metrisch Bewertbare Zielgröße wird. Es könnte also gesagt werden, dass Depth Maps standardisierte vergleichbare Größen sind, die sich optimieren lassen [Silveira et al., 2023, 7-8]. Da es sich bei Depth Maps um zweidimensionale Bilder handelt, ist für die vollständige Rekonstruktion eines 3D-Modells eine weiterführende Verarbeitung erforderlich. Silveira et al. betonen hierzu ausdrücklich, dass „per-face depth estimation tends to produce discontinuities along face edges [...] and post-processing is required to obtain the full 3D model of the scene“ [Silveira et al., 2023, S. 20]. Die Ableitung eines expliziten 3D-Modells aus einer Depth Map ist grundsätzlich

möglich, erfordert jedoch zusätzliche Rekonstruktions und Optimierungsschritte. Diese führen häufig zu hochauflösten Punktwolken oder Meshes, die für den Einsatz in webbasierten Anwendungen hinsichtlich Speicherbedarf und Renderperformance nur eingeschränkt geeignet sind. Insbesondere bei 360° Panoramen steigt die Datenmenge durch die vollständige Rundumabbildung deutlich an, was eine effiziente Low Poly Rekonstruktion zusätzlich erschwert. Aus diesem Grund konzentriert sich die Forschung häufig auf die Erzeugung und Bewertung von Depth Maps als Zwischendarstellung.

#### **2.4.3 Forschung zur Tiefenwahrnehmung auf 2D-Monitoren sowie in VR/AR-Systemen**

In diesem Abschnitt wird der Forschungsstand zur Tiefenwahrnehmung auf 2D-Displays sowie in VR/AR-Systemen zusammengefasst. Die Tiefenwahrnehmung auf 2D-Monitoren ist seit mehreren Jahrzehnten Gegenstand wahrnehmungspsychologischer Forschung und gilt heute als gut untersucht. Monitore sind weit verbreitet, technisch standardisiert und ermöglichen kontrollierte experimentelle Bedingungen, wodurch sich ein umfangreicher und konsolidierter Forschungsstand zur Tiefenwahrnehmung auf zweidimensionalen Darstellungsmedien etabliert hat [Goldstein, 2013, S.234–235]. Im Vergleich zu 2D-Monitoren ist die Forschung zur Tiefenwahrnehmung in VR- und AR-Systemen noch vergleichsweise jung und weist weiterhin offene Fragestellungen auf [Hoffmann, 2025, S.92–96]. Dies steht im Zusammenhang mit der fortschreitenden technischen Entwicklung sowie der zunehmenden Verfügbarkeit von VR- und AR-Systemen Statista [2025], Coban et al. [2022].

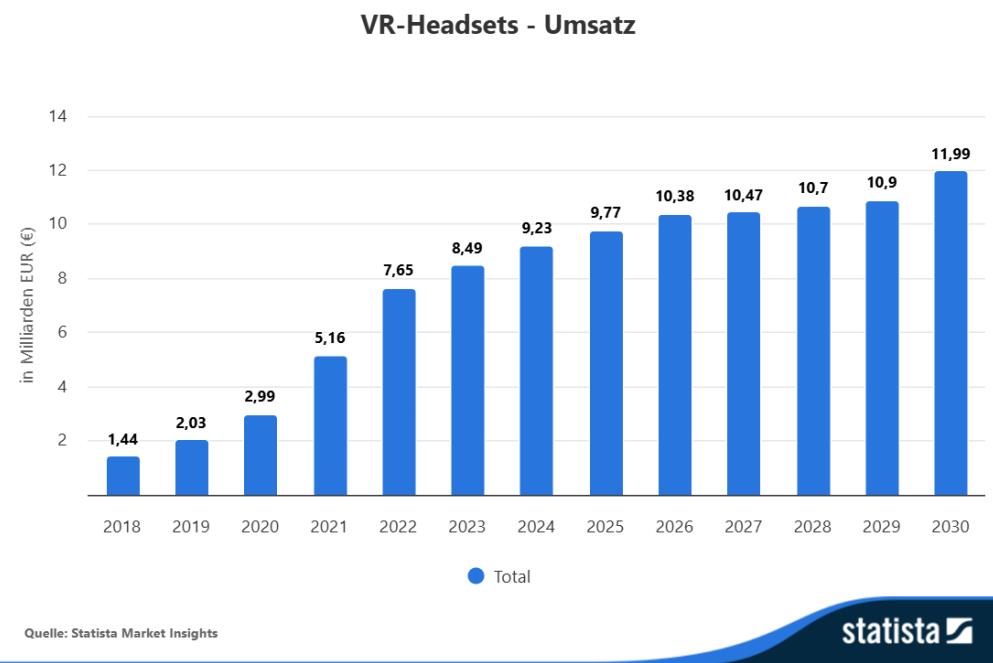


Abbildung 2.3: Marktentwicklung von VR-Headsets weltweit. Quelle: Statista

<https://de.statista.com/outlook/cmo/consumer-electronics/gaming-equipment/vr-headsets-weltweit>

Viele Studien zur Tiefenwahrnehmung stützen sich auf grundlegende Erkenntnisse der Wahrnehmungspsychologie, wie sie in etablierten Lehrbüchern zusammengefasst sind. Ein häufig zitierter Überblick findet sich bei Goldstein (2013), der zentrale Mechanismen und Tiefenhinweise der visuellen Wahrnehmung systematisch darstellt Goldstein [2013]. Diese Grundlagenforschung bildet bis heute die Basis für zahlreiche Anwendungsfelder, von der Bildverarbeitung über grafische Benutzeroberflächen bis hin zur computergestützten Visualisierung. Aktuelle Studien konzentrieren sich zunehmend auf grundlegende Effekte von Immersion, Präsenz und räumlicher Wahrnehmung Li et al. [2025].

#### 2.4.4 Forschung zu KI-generierten Low Poly-Modellen

Automatisierte Verfahren zur Erstellung KI-generierter 3D-Modelle, etwa auf Basis von Bildern oder Scans, sind Gegenstand aktueller Forschung und inzwischen auch praktisch verfügbar. Insbesondere bildbasierte Verfahren zur Rekonstruktion von 3D-Geometrie aus realen Objekten haben in den letzten Jahren an Bedeutung gewonnen. Die dabei

resultierenden Modelle weisen häufig eine hohe geometrische Detaildichte sowie eine entsprechend hohe Polygonanzahl auf. In ihrer Rohform sind solche Modelle daher meist nur eingeschränkt für webbasierte Echtzeitanwendungen geeignet, bei denen, wie in Kapitel 2.3.3 beschrieben, eine reduzierte Geometrie und optimierte Texturierung eine zentrale Rolle spielen. [Tewari et al., 2023, S.6-7] [Huang et al., 2024, S.5-7] [Lee et al., 2024, S.1-2]

Übersichtsarbeiten zur bildbasierten 3D-Rekonstruktion zeigen, dass KI-gestützte Verfahren primär auf die möglichst detailgetreue Erfassung von Oberflächen abzielen und weniger auf eine kontrollierte Low Poly-Topologie [Han et al., 2021, S.1]. Die zunehmende praktische Relevanz dieser Ansätze wird zudem durch kommerzielle Werkzeuge wie Meshy-AI verdeutlicht, die automatisierte Bild-zu-3D-Workflows anbieten Meshy AI [2025]. Eine nachträgliche Überführung solcher hochauflösten Modelle in performante Low Poly-Varianten erfordert in vielen Fällen zusätzliche Verarbeitungsschritte wie Retopologie, Polygonreduktion und Texture Baking, die nur begrenzt automatisierbar sind und häufig manuelle Nachbearbeitung erfordern.

## 3 Methodik

Es gibt verschiedene Möglichkeiten, die Wahrnehmung von Tiefe in 3D-Szenarien zu untersuchen. In dieser Arbeit wurde eine Nutzerbefragung gewählt, weil diese 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 schnell stark variieren. Deshalb wurde dieser Ansatz nicht weiterverfolgt. Es existieren auch technisch-physikalische Messmethoden sowie Eye-Tracking Verfahren zur Erhebung der Wahrnehmung und des Blickverhaltens von Testpersonen. Diese kamen im Rahmen dieser Arbeit jedoch nicht zum Einsatz, da sie spezielle Hardware voraussetzen und die Studie explizit für eine webbasierte Nutzung auf 2D-Monitoren konzipiert ist.

### 3.1 Überblick der Messmethoden

#### 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. [2024] nutzt das Subjektive Einschätzungsverfahren über Ratingskalen als ein etabliertes Verfahren zur Erfassung wahrgenommener räumlicher Tiefe [Lesar et al., 2024, s. 1-2].

### 3.1.2 Verhaltensbasierte Tiefeneinschätzungen

Ansätze, die stärker auf die objektive Wahrnehmungsleistung abzielen, basieren auf Aufgaben, bei denen Teilnehmende räumliche Tiefe nicht nur subjektiv bewerten, sondern aktiv einschätzen. Zu den etablierten Verfahren zählen Relative Depth Ordering, relative Distanzeinschätzungen sowie Pairwise- und Forced-Choice Methoden, bei denen jeweils vergleichende Entscheidungen zwischen Objekten getroffen werden. Bei der Relative Depth Ordering Methode beurteilen Teilnehmende, welches Objekt näher oder weiter entfernt ist. Die Methode kommt ohne metrische Angaben aus und eignet sich daher insbesondere für Entfernungsschätzungen, bei denen relative Tiefenrelationen im Vordergrund stehen [Zabih and Woodfill, 1994, s. 1–2]. Auch in aktuellen Arbeiten wird dieses Prinzip weiterhin eingesetzt. So nutzen Singh et al. [2025] ein vergleichbares relatives Verfahren, bei dem Tiefenpositionen visuell eingeschätzt und im Rahmen eines Matching-Tasks (visueller Abgleich zweier Tiefenpositionen) beurteilt werden [Singh et al., 2025, s. 5].

Eine weitere Methode ist die relative Distanzschätzung, bei der Teilnehmende eine Entfernung innerhalb eines vorgegebenen Distanzbereichs einordnen (z. B. 5–10 m, 10–15 m). Auch dieses Verfahren kommt ohne exakte metrische Angaben aus, erlaubt jedoch nur eine grobe Klassifizierung von Entfernungen und keine präzisen Distanzwerte [Committeri et al., 2020, s. 1–2] [Lesar et al., 2024, s. 4]. Zuletzt gibt es noch die Pairwise- bzw. Forced-Choice Methode. Dabei werden zwei Darstellungen direkt miteinander verglichen, indem sie nacheinander präsentiert werden und anschließend beurteilt wird, welche Darstellung im Hinblick auf die Tiefenwahrnehmung realistischer wirkt. Dieses Verfahren eignet sich insbesondere für den direkten Vergleich mehrerer Szenarien [Ashtiani et al., 2023, s. 4].

Auf Eye-Tracking wird in dieser Arbeit verzichtet, da die notwendige Hardware nicht zur Verfügung stand. Eye-Tracking eignet sich grundsätzlich zur detaillierten Erfassung von Blickverläufen und Aufmerksamkeitsmustern. Die Methode liefert dabei indirekte Hinweise auf Wahrnehmungsprozesse, etwa durch typische Fixations- und Blickbewegungsmuster, die mit der Einschätzung räumlicher Tiefe in Zusammenhang stehen können [Choi and Nam, 2025, s. 4].

## 3.2 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 Arten 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 Webbrowsert

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 unter eingeschränkten Bewegungsmöglichkeiten untersucht. Die Teilnehmenden können sich nur begrenzt innerhalb der Szene bewegen, sodass einerseits Bewegungsparallaxe als Tiefenhinweis möglich bleibt, andererseits jedoch keine beliebigen Perspektivwechsel möglich sind. Ziel dieser Einschränkung ist es, zu verhindern, dass Entfernung durch gezielte Positionierung entlang einer gemeinsamen Blicklinie oder durch andere geometrische Hilfskonstruktionen abgeschätzt werden. Die Tiefenwahrnehmung soll dabei hauptsächlich aus den Veränderungen der Objektpositionen relativ zum Beobachtungspunkt entstehen und nicht aus frei wählbaren Standpunkten oder direkten Vergleichsstrategien. 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, durch welches die schwächste Tiefenwirkung erwartet wird. Anschließend folgt Szenario 3 (Low Poly Rekonstruktion) und zum Schluss Szenario 1 als High Poly Basisszene 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.

## 3.4 Datenaufarbeitung und Auswertung

Die im Rahmen der Online-Studie erhobenen Daten wurden mithilfe des Umfragetools LimeSurvey erfasst und anschließend in tabellarischer Form ausgewertet. Die Daten liegen dabei in einer strukturierten Matrix vor, in der jede teilnehmende Person durch eine Zeile und jede Frage durch eine Spalte repräsentiert wird. Zur Beschreibung der zentralen Tendenz und der Streuung wurden robuste Kennwerte herangezogen, die den ordinalen Charakter der Skalen berücksichtigen. Die Ergebnisse werden in tabellarischer Form dargestellt.

Für die metrischen Distanzschätzungen aus Phase 1 wurden reale Referenzwerte herangezogen, die als Ground Truth Distanzen definiert wurden. Die Einschätzungen der Teilnehmenden wurden diesen Referenzwerten gegenübergestellt, um Abweichungen zwischen wahrgenommener und tatsächlicher Entfernung zu erfassen. Die Ergebnisse werden dabei nicht als exakte Genauigkeitsmessung interpretiert, sondern als relative Nähe zu den jeweiligen Referenzwerten. Nicht-metrische relative Tiefenurteile, etwa zu Nähe- und Distanzrelationen, wurden über Häufigkeitsverteilungen ausgewertet. Ein hoher Anteil konsistenter Antworten wird dabei als Hinweis auf eine stabile Wahrnehmung relativer Tiefenrelationen interpretiert, während häufige unsichere Angaben auf Wahrnehmungsunsicherheiten hindeuten. Freitextantworten und Begründungen wurden ergänzend qualitativ gesichtet und thematisch zusammengefasst. Wiederkehrende Beobachtungen, etwa zu wahrgenommenen Verzerrungen oder auffälligen Tiefeneffekten, werden in den Ergebnissen exemplarisch aufgegriffen und zur Einordnung der quantitativen Resultate herangezogen.

### 3.4.1 Grundlegender Auswertungsansatz

Aufgrund der explorativen Zielsetzung und der begrenzten Stichprobengröße wurden keine statistischen Signifikanztests durchgeführt. Der Fokus der Auswertung liegt auf der Beschreibung von Wahrnehmungstendenzen innerhalb der untersuchten Szenarien, nicht auf einer statistischen Generalisierung der Ergebnisse auf eine Grundgesamtheit. Zur Analyse wurden Häufigkeiten, prozentuale Anteile sowie Lage und Streuungsmaße herangezogen. Diese Kennwerte ermöglichen es, Unterschiede zwischen den Szenarien sichtbar zu machen und zugleich Unsicherheiten oder Streuungen in den Antworten der Teilnehmenden abzubilden.

### **Metrische Fragen und Ground Truth Kategorien**

Die metrischen Fragen zielten auf die Einschätzung von Distanzen innerhalb vorgegebener Entfernungsklassen ab. Für jede dieser Fragen wurden Referenzwerte definiert, die als Ground Truth Distanzen (GT) dienen. Die Auswertung erfolgte nicht auf Basis exakter numerischer Abweichungen, sondern durch die Zuordnung der Antworten zu diskreten Distanzklassen. Als korrekt (GT-Klasse) wurde eine Antwort gewertet, wenn sie derjenigen Distanzklasse entsprach, in der der Referenzwert lag. Zusätzlich wurde eine erweiterte Kategorie  $GT \pm 1$  eingeführt, die auch Antworten berücksichtigt, die maximal eine Distanzklasse ober oder unterhalb der Ground Truth Klasse liegen. Diese Kategorie dient als Toleranzmaß und erlaubt eine differenziertere Betrachtung der Nähe zur tatsächlichen Entfernung. Der Anteil der Antworten in der GT-Klasse sowie der Anteil der Antworten in der Kategorie  $GT \pm 1$  wurden jeweils als prozentuale Kennwerte berechnet. Die Kategorie  $GT \pm 1$  stellt dabei keinen Mittelwert dar.

### **Auswertung relativer Vergleichsfragen**

Die relativen Vergleichsfragen zielten darauf ab, Tiefenrelationen zwischen Objekten zu erfassen, etwa durch Einschätzungen wie „Objekt A ist näher“, „Objekt B ist näher“ oder „beide Objekte sind gleich weit entfernt“. Zusätzlich stand den Teilnehmenden eine Unsicherheitsoption zur Verfügung. Antworten wurden als korrekt gewertet, wenn sie mit der tatsächlichen Tiefenrelation übereinstimmten, wobei die Option „gleich weit entfernt“ dann als korrekt galt, wenn keine Tiefendifferenz zwischen den Objekten vorlag. Eindeutig widersprechende Antworten wurden als falsch kategorisiert. Die Option „unsicher“ wurde unabhängig davon als eigenständige Kategorie erfasst und dient als Indikator für Wahrnehmungsunsicherheiten. Für jedes Szenario wurden die prozentualen Anteile korrekter, falscher und unsicherer Antworten ermittelt. Der Anteil unsicherer Antworten ergibt sich dabei als Differenz zwischen der Gesamtzahl der Antworten und der Summe aus korrekten und falschen Einschätzungen. Diese Differenzierung ermöglicht es, nicht nur die Treffergenauigkeit, sondern auch die Sicherheit der Wahrnehmung abzubilden.

### **Auswertung subjektiver Likert-Skalen**

Die subjektiven Einschätzungen der Teilnehmenden wurden mithilfe von siebenstufigen Likert-Skalen erhoben. Da es sich hierbei um ordinalskalierte Daten handelt, wurden zur Auswertung robuste Kennwerte herangezogen, die diesem Skalenniveau Rechnung

tragen. Als Maß der zentralen Tendenz wurde der Median verwendet, da er gegenüber Ausreißern unempfindlich ist und eine geeignete Beschreibung ordinaler Daten ermöglicht. Ergänzend wurde der Modus bestimmt, um den am häufigsten gewählten Skalenwert als Hauptanteil der Antworten abzubilden. Zur Beschreibung der Streuung wurde die Standardabweichung herangezogen, um das Ausmaß der Uneinigkeit oder Polarisierung innerhalb der Bewertungen sichtbar zu machen. Die Kombination aus Median, Modus und Streuungsmaß erlaubt eine differenzierte Betrachtung der subjektiven Wahrnehmung, ohne Annahmen über Normalverteilungen oder intervallskalierte Daten zu treffen.

## 4 Implementierung

### 4.1 Technische Umsetzung

Zu Beginn wurde ein einheitliches Basisszenario definiert, das als gemeinsame Grundlage für alle weiteren Szenarien dient. Sämtliche Szenarien wurden mithilfe der 3D-Software Blender erstellt. Ein Teil der verwendeten Modelle wurde eigens modelliert und texturiert, darunter beispielsweise das Holzfällerhaus. Weitere Objekte stammen aus der Objektbibliothek des Online-Asset-Anbieters *Poly Haven Asset Browser* [Poly Haven, 2025]. Die Szene besteht aus einer zentralen Bodenfläche sowie vier thematisch unterschiedlichen Showcases, die in vier Blickrichtungen um den Ausgangspunkt angeordnet sind. Diese Showcases umfassen eine Autogarage, ein mittelalterlich dekoriertes Zimmer, einen Pavillon mit Sitzplätzen sowie ein Holzfällerhaus. Die Anordnung wurde so gewählt, dass sich aus der zentralen Betrachtungsposition unterschiedliche Tiefenstaffelungen ergeben und die räumlichen Relationen zwischen Vorder- und Hintergrundelementen gut wahrnehmbar sind. Die gesamte Szenerie ist in eine HDRI-basierte Umgebungsbeleuchtung eingebettet, die eine gleichmäßige Grundausleuchtung sicherstellt. Ergänzend wurden gezielt Lichtquellen eingesetzt, um relevante Objekte hervorzuheben und die räumliche Struktur der Szene zu unterstützen. Dieses Szenario bildet das Basisszenario (S1), aus dem die beiden weiteren Szenarien (S2 und S3) technisch abgeleitet wurden.

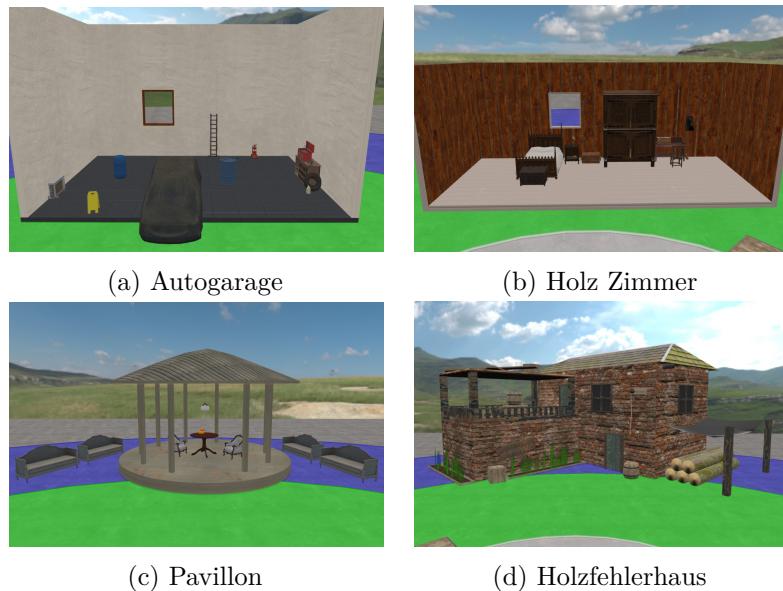


Abbildung 4.1: Showcase A–D

#### 4.1.1 Szenario 1: Baseline-3D-Rendering mit gezielten Tiefenherausforderungen

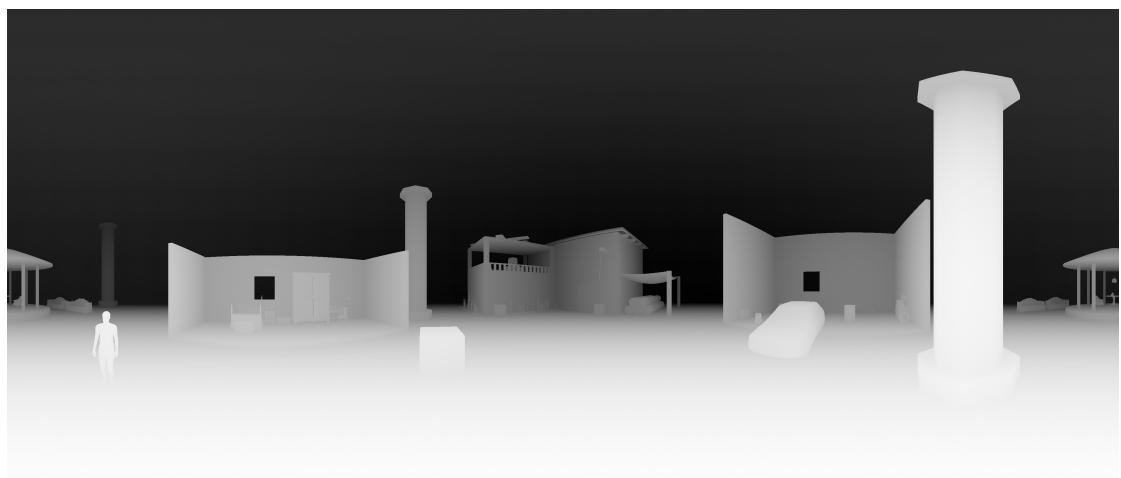
Szenario 1 dient als Basisszenario und stellt ein klassisches 3D-Rendering dar, das webbasiert mithilfe von Babylon.js dargestellt wird. Die Szene wurde vollständig in Blender erstellt und besteht aus einer Kombination aus verschiedenen 3D modellen. Sämtliche Elemente wurden zu einer gemeinsamen Szene zusammengeführt und anschließend als .glb Datei exportiert, da dieses Format für die effiziente Darstellung komplexer 3D-Szenen im Web geeignet ist The Khronos Group [2020]. Die Szene ist bewusst so gestaltet, dass typische Herausforderungen der Tiefenwahrnehmung auf zweidimensionalen Displays auftreten. Dazu zählen unter anderem Überlagerungen von Objekten, unterschiedliche Entfernungsstaffelungen sowie varierende Objektgrößen ohne explizite Tiefenhinweise. Die Kameraposition ist fixiert und entspricht in allen Szenarien derselben Startperspektive, um eine vergleichbare Wahrnehmungsgrundlage für die nachfolgenden Untersuchungen sicherzustellen. Szenario 1 fungiert als Referenzszene, auf deren geometrischer und räumlicher Struktur die beiden weiteren Szenarien aufbauen. Veränderungen an Geometrie, Darstellungsform oder Tiefeninformation erfolgen in den nachfolgenden Szenarien ausschließlich auf Basis dieser Ausgangsszene.

#### 4.1.2 Szenario 2: Depth Map und 360°-Panorama (Babylon.js)

Aufbauend auf dem Basisszenario (Szenario 1) wurde für Szenario 2 ein 360°-Panorama der bestehenden 3D-Szene erzeugt. Die vollständige Szenerie wurde hierzu in Blender gerendert, wobei die Kamera auf eine sphärische Panorama Projektion eingestellt wurde. Das resultierende Panorama wurde in äquirektangulärer Projektion mit einem Seitenverhältnis von 2:1 ausgegeben, um eine korrekte Abbildung auf eine Kugelgeometrie in der webbasierten Darstellung zu ermöglichen. Parallel zum Farbrendering wurde eine Depth Map der Szene erzeugt. Hierzu wurde der Z-Depth Pass im View Layer aktiviert und im Compositing Modus aus dem Render Layer extrahiert Blender Manual [2025]. Die Tiefenwerte wurden normalisiert und invertiert, sodass aus der Renderperspektive nahe Bildbereiche hell und weiter entfernte Bereiche dunkel kodiert sind. Das resultierende Tiefenbild wurde als Graustufenbild exportiert und gemeinsam mit dem Panorama gespeichert. Abbildung 4.2 zeigt die resultierenden Renderergebnisse des Panoramas sowie die zugehörige Depth Map von Szenario 2.



(a) Farbpanorama der Szene (Szenario 2)



(b) Zugehörige Depth Map des Panoramas

Abbildung 4.2: Render Ergebnisse von Szenario 2: Panorama Darstellung und zugehörige Tiefeninformation

Das 360° Panorama sowie die zugehörige Depth Map wurden anschließend über feste URLs

bereitgestellt und in Babylon.js geladen. Beide Texturen wurden auf eine segmentierte Kugelgeometrie projiziert, wobei die Depth Map zur positionsabhängigen Verformung der Oberfläche verwendet wurde, um einen pseudo-räumlichen Tiefeneindruck zu erzeugen. Innerhalb der Webszene wurden grundlegende Kamerasteuerungen sowie Parameter zur Steuerung der Stärke der Tiefenverformung implementiert. Das Szenario ist dadurch interaktiv betrachtbar und stellt ein texturbasiertes 2.5D-Verfahren dar, bei dem insbesondere bei Kamerabewegungen charakteristische geometrische Verzerrungen auftreten.

#### **4.1.3 Szenario 3: Low Poly Rekonstruktion aus Szenario 1**

Szenario 3 basiert auf einer geometrisch vereinfachten Rekonstruktion des Basisszenarios (Szenario 1). Ziel war es, die vollständige dreidimensionale Tiefeninformation der Szene beizubehalten und zugleich die Dateigröße sowie die Laufzeitperformance 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 einer High Poly Geometrie erweitert. Für Szenario 3 wurden diese Modifier gezielt deaktiviert, um auf die ursprüngliche Low Poly Struktur zurückzugreifen. Zur konsistenten Verwaltung mehrerer Objekte wurde hierfür ein skriptbasierter Workflow eingesetzt. Hochauflöste Assets aus externen Quellen wurden mithilfe des Decimate Modifiers geometrisch vereinfacht. Dies ist in Abbildung 4.3 zu sehen.

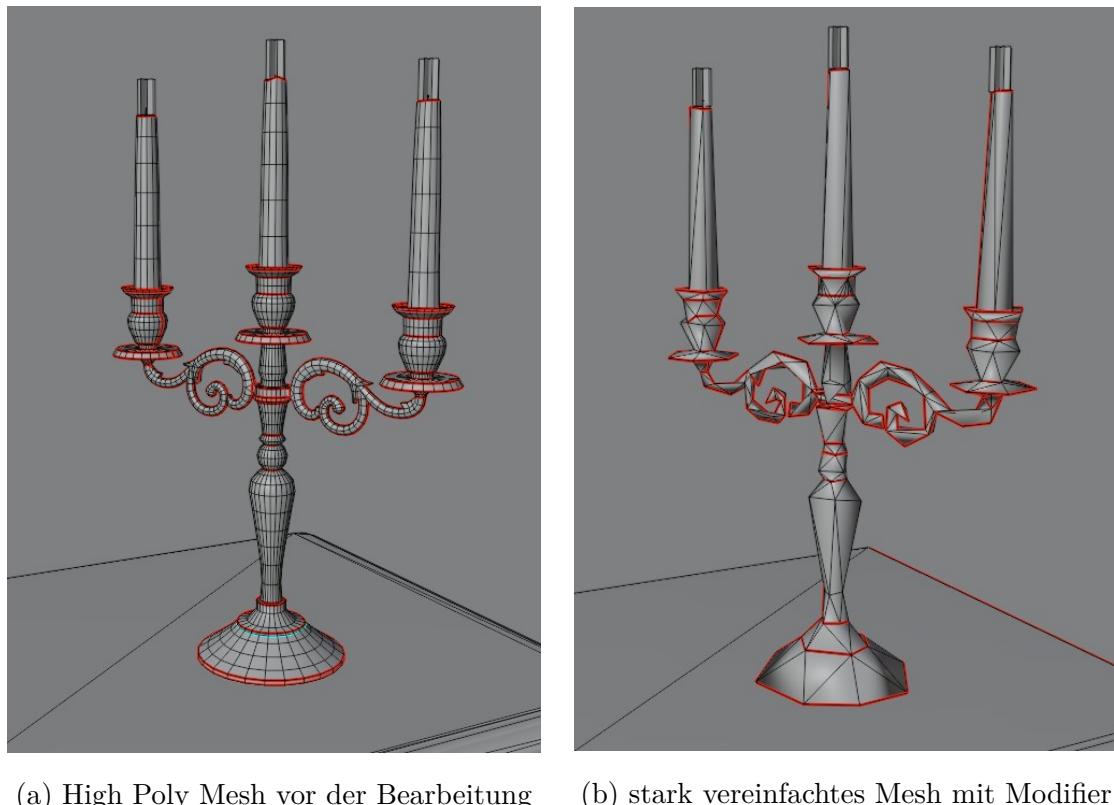


Abbildung 4.3: Erster Ansatz für Szenario 3

Dieses Vorgehen stellte einen praktikablen Kompromiss zwischen visueller Qualität und Arbeitsaufwand dar, da alternative Verfahren wie z.B. eine manuelle Retopologie den zeitlichen Rahmen der Arbeit überschritten hätten. Zur Vorbereitung des Texture Baking Prozesses wurden für alle relevanten Objekte konsistente UV-Layouts erstellt. Hierfür wurden die Meshes gezielt mit Seam-Markierungen versehen, insbesondere entlang von Kanten mit Winkeln nahe oder über  $90^\circ$ , um Überlappungen und Verzerrungen bei der UV-Abwicklung zu vermeiden. Die Platzierung der Seams erfolgte teilautomatisiert mithilfe eines unterstützenden Skripts, das entsprechende Kanten anhand ihres Normalenwinkels identifizierte.

Dieser Schritt war insbesondere bei Meshes erforderlich, deren Geometrie durch den Einsatz des Decimate Modifiers stark reduziert wurde, da hierbei die ursprüngliche Topologie und Kantenführung teilweise verloren ging, was eine saubere UV-Abwicklung erschwerte. Durch manuelle Nachbearbeitung der UVs konnte dennoch ein konsistentes Baking Ergebnis erzielt und die visuelle Qualität der Low Poly Modelle verbessert werden.

Neben der Reduktion der Polygonanzahl wurde auch die Texturauflösung optimiert. Ursprünglich hochauflösende Texturen wurden mithilfe von Substance Painter neu eingebunden und in komprimierter Form mit reduzierter Auflösung exportiert. Dadurch konnte die Gesamtdateigröße der Szene erheblich verringert werden, ohne den grundlegenden visuellen Eindruck maßgeblich zu verändern Blender Manual Decimate [2025].

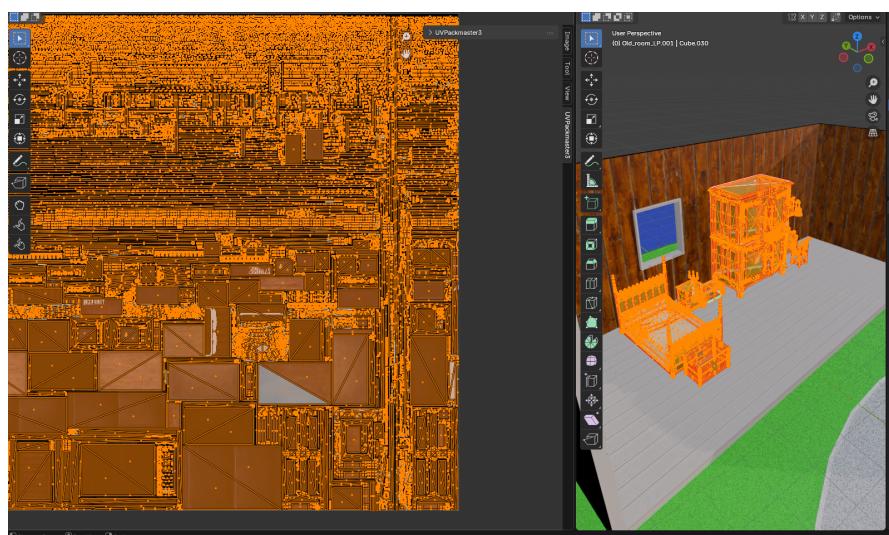


Abbildung 4.4: Beispiel für Ausgelegtes UV-Layout eines Low Poly Modells

Das final optimierte Low Poly Modell wurde anschließend erneut in Blender zusammengeführt und als .glb Datei exportiert. Die Szene wurde über einen 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 einerseits die visuelle Tiefenwirkung zu unterstützen und andererseits die Laufzeitperformance zu verbessern. Babylon.js Documentation [2025b] Babylon.js Documentation [2025a]

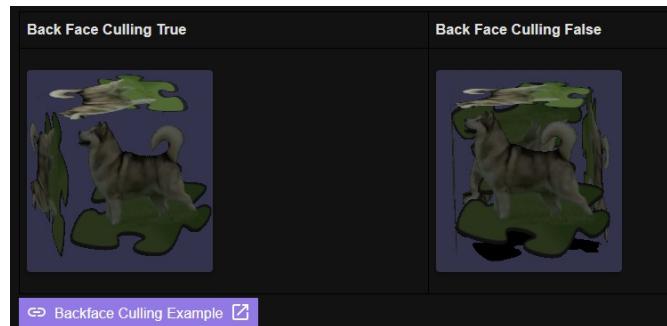


Abbildung 4.5: Vergleich der Darstellung mit aktiviertem und deaktiviertem Backface Culling in Babylon.js Babylon.js Documentation [2025a]

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.

## 5 Ergebnisse

In den folgenden Abbildungen werden die Ergebnisse der Befragung dargestellt. Die metrischen Distanzschätzungen werden anhand der Kategorien **GT-Klasse** und **GT  $\pm$  1** ausgewertet, wobei die **GT-Klasse** Antworten umfasst, die der tatsächlichen Distanzstufe entsprechen, und **GT  $\pm$  1** Antworten, die maximal eine Distanzstufe vom Referenzwert abweichen. Die **Streuung** beschreibt die Verteilung der gegebenen Antworten innerhalb derselben Frage und dient als Maß für die Uneinheitlichkeit der Einschätzungen.

Die subjektiven Bewertungen basieren auf siebenstufigen Likert-Skalen und werden durch **Median**, **Hauptanteil (Modus)** und **Streuung** beschrieben. Die relativen Tiefenschätzungen werden in die Kategorien **korrekte Ordnung**, **falsche Ordnung** und **unsicher** unterteilt, um die Einschätzung der relativen Tiefenrelationen zwischen Objekten darzustellen.

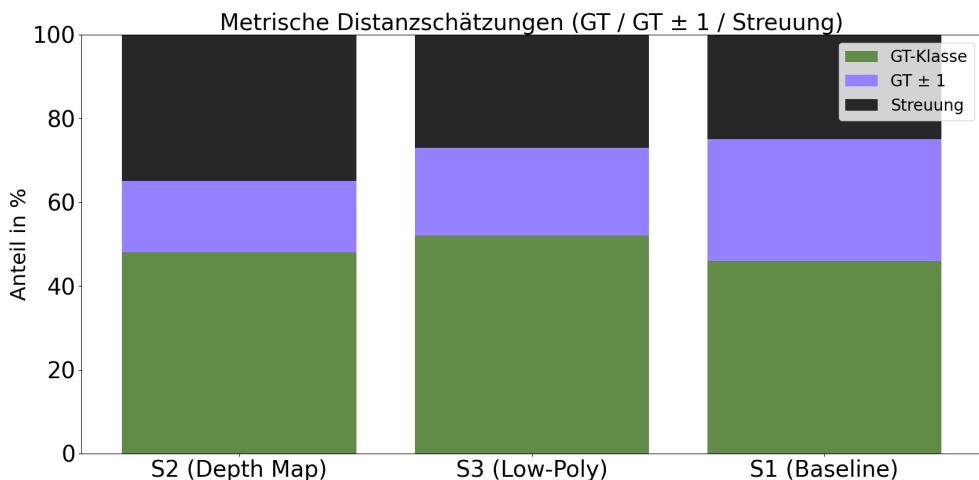


Abbildung 5.1: Auswertung relative metrische Schätzungsfragen

Abbildung 5.1 zeigt die Ergebnisse der metrischen Distanzschätzungen in Phase P1. In allen Szenarien weist die **GT-Klasse** den größten Antwortanteil auf. Unterschiede zwischen den Szenarien zeigen sich in der **Streuung** sowie im Anteil der Toleranzkategorie (**GT  $\pm$  1**), wobei das Depth-Map-Szenario die höchste Streuung, das Low-Poly-Szenario

den höchsten Anteil exakt korrekter Schätzungen und das Basisszenario den höchsten **GT ± 1**-Anteil aufweist.

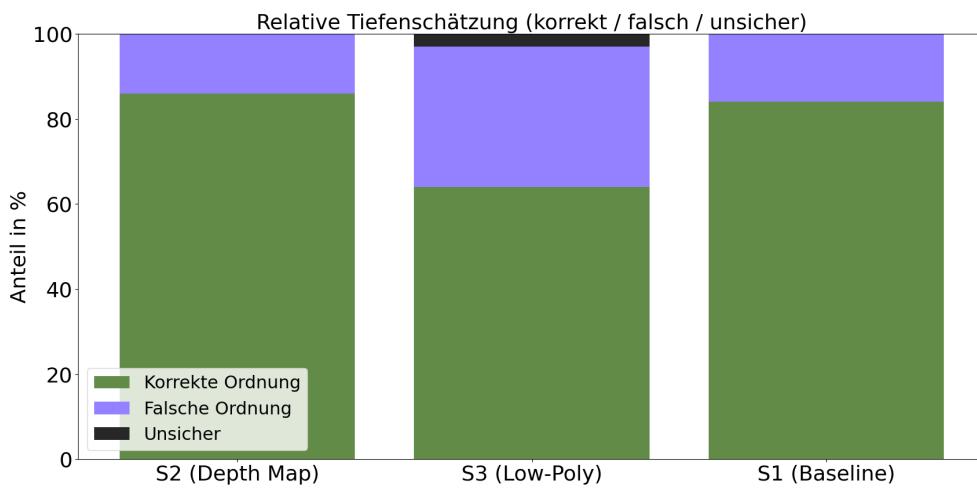


Abbildung 5.2: Auswertung relative Schätzungsfragen

Abbildung 5.2 zeigt die Ergebnisse der relativen Tiefenschätzungen in Phase P1. In den Szenarien S1 (Baseline) und S2 (Depth Map) liegt der Anteil **korrekt geordneter** Antworten auf einem vergleichbaren Niveau von jeweils über 80%, wobei in beiden Fällen keine **unsicheren** Einschätzungen auftreten. Entsprechend fällt auch der Anteil **falscher Ordnungen** in beiden Szenarien ähnlich gering aus. Im Szenario S3 (Low Poly) liegt der Anteil **korrekter Ordnungen** mit 64% deutlich geringer, während zugleich ein kleiner Anteil **unsicherer** Einschätzungen von 3% auftritt; der verbleibende Anteil entfällt auf **falsch** eingeordnete Tiefenrelationen.

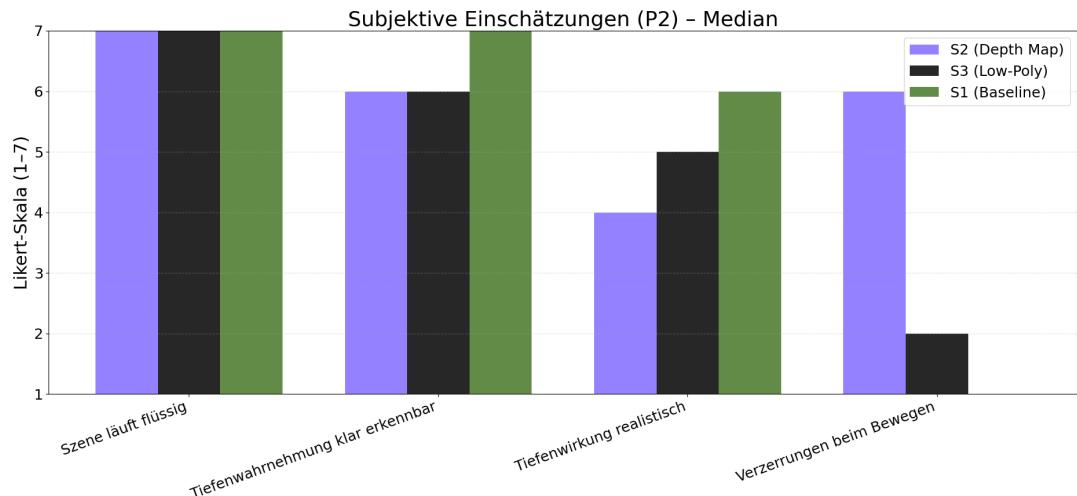


Abbildung 5.3: Ergebnisse subjektive Einschätzungen, Median

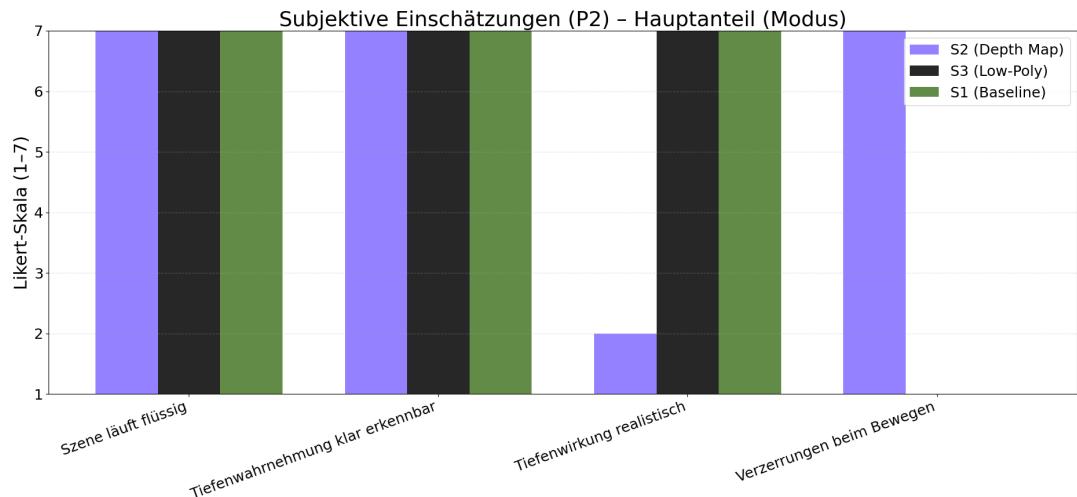


Abbildung 5.4: Ergebnisse subjektive Einschätzungen, Hauptanteil

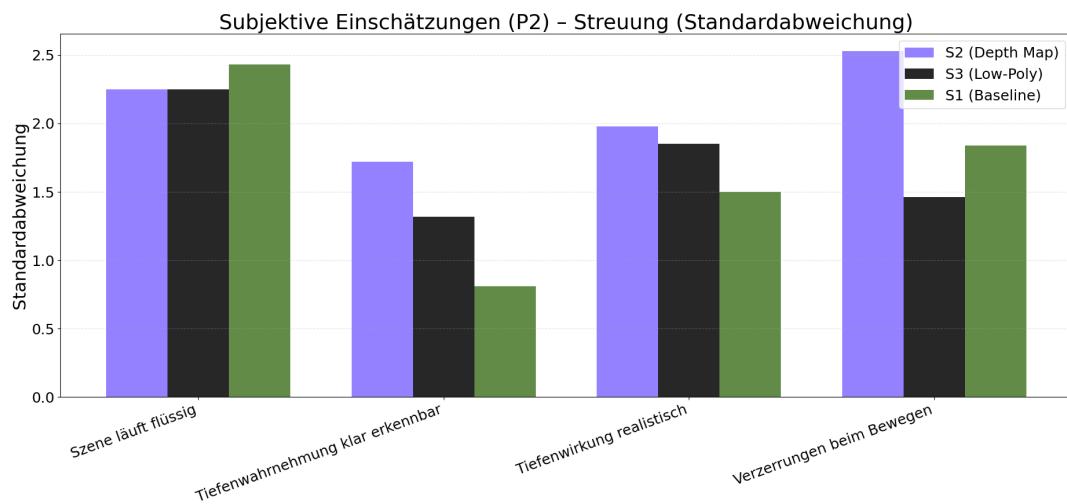


Abbildung 5.5: Ergebnisse subjektive Einschätzungen, Streuung

Die Abbildungen 5.3 bis 5.5 zeigen die Ergebnisse der subjektiven Einschätzungen in Phase P2, dargestellt anhand von **Median**, **Hauptanteil (Modus)** und **Streuung**. Über alle Szenarien hinweg weisen die Mediane überwiegend hohe Werte für die Klarheit der Tiefenwahrnehmung sowie die wahrgenommene Performance auf, während die Einschätzung der Realitätsnähe und der Verzerrungen stärkere Unterschiede zwischen den Szenarien aufzeigt. Die Darstellung des Hauptanteils bestätigt diese Tendenzen durch wiederkehrende Häufungen an den oberen Skalenwerten. Die Streuung verdeutlicht zudem, dass einzelne Fragen eine stärkere Varianz in den Antworten aufweisen. Hier sticht insbesondere die Wahrnehmung der Verzerrungen während der Bewegung hervor.

# 6 Diskussion

## 6.1 Einordnung

### 6.1.1 Szenario 2: Panorama und Depth map

Die drei untersuchten Szenarien unterscheiden sich primär in ihrer technischen Umsetzung der Tiefendarstellung, während Aufbau, Objektanordnung und Blickrichtungen bewusst konstant gehalten wurden. Dadurch lassen sich Unterschiede in der Wahrnehmung räumlicher Tiefe unmittelbar auf die jeweilige Darstellungsform zurückführen und nicht auf inhaltliche oder kompositorische Abweichungen innerhalb der Szene. Beim Vergleich der Szenarien fällt insbesondere Szenario 2 (Depth Map basiertes Panorama) aus dem Gesamtbild heraus. Während Szenario 1 (Baseline 3D-Rendering) und Szenario 3 (Low Poly Geometrie) in den objektiven und subjektiven Ergebnissen überwiegend konsistente Werte mit geringer Streuung aufweisen, zeigt Szenario 2 Abweichungen. Dies betrifft sowohl die metrischen und relativen Schätzungen als auch die subjektive Einschätzung der Realitätsnähe und der Bewegungsstabilität.

Die Ergebnisse der subjektiven Bewertung legen nahe, dass die bildbasierte Tiefenrekonstruktion mittels Depth Map in diesem Szenario mit wahrnehmbaren Verzerrungen einhergehen könnte. Insbesondere bei Kamerabewegungen scheinen Verschiebungen und Verformungen der Szene aufzutreten, die den Eindruck einer konsistenten räumlichen Struktur beeinträchtigen könnten. Diese Effekte könnten zu einer reduzierten Immersion beitragen und das Szenario im Vergleich weniger realistisch erscheinen lassen.

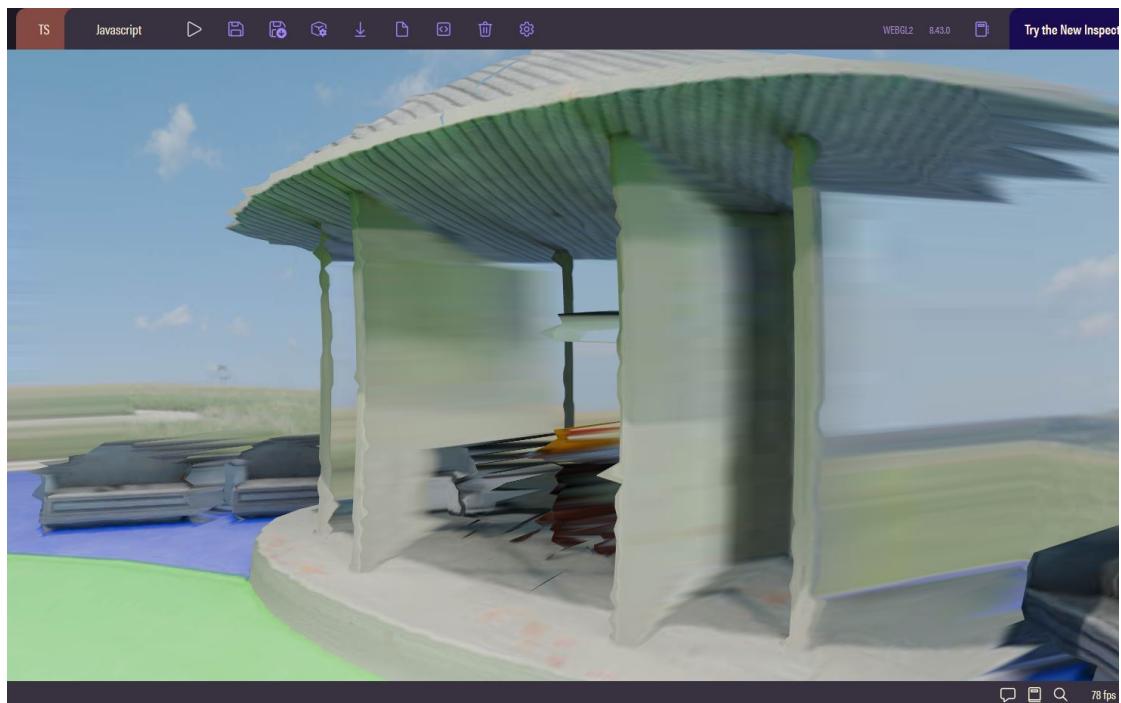


Abbildung 6.1: Verzerrung im Spektrum von 0-255

Die Tiefenverformung erfolgt über die Funktion `applyDisplacementMap`, bei der Graustufenwerte einer Depth Map auf eine sphärische Geometrie abgebildet und entlang der Normalen verschoben werden. Der zugrunde liegende Graustufenbereich von 0 - 255 kodiert dabei ausschließlich relative Tiefenunterschiede Babylon.js Contributors [2025]. Die Parameter für minimale und maximale Verschiebung definieren lediglich den Wirkungsbereich der Verformung, ohne zusätzliche Tiefeninformation zu erzeugen. Eine Erhöhung des maximalen Displacement Werts über den effektiven Wertebereich der Depth Map hinaus führt daher nicht zu einer genaueren Tiefenabbildung, sondern verstärkt geometrische Artefakte bis hin zur Selbstintersektion der Geometrie.

Umgekehrt resultiert eine starke Einschränkung des Displacement Bereichs in einer deutlichen Abflachung der Szene, da nahe Tiefenbereiche komprimiert und feine Tiefenabstufungen unterdrückt werden. Auch eine Erhöhung der geometrischen Auflösung der sphärischen Projektion konnte diese Effekte nicht grundlegend beheben, da die Limitationen primär aus der bildbasierten Tiefenrepräsentation resultieren. Ein wesentlicher Faktor für das vergleichsweise schwache Abschneiden von Szenario 2 in den metrischen Auswertungen liegt im Vergleich mit den Ground Truth Werten. Eine durchgehend fehlerfreie Darstellung der Distanzen anhand der Ground Truth Distanzen ist daher kaum möglich und stellt eine erhebliche Limitation dieses Ansatzes dar und ist nicht auf eine

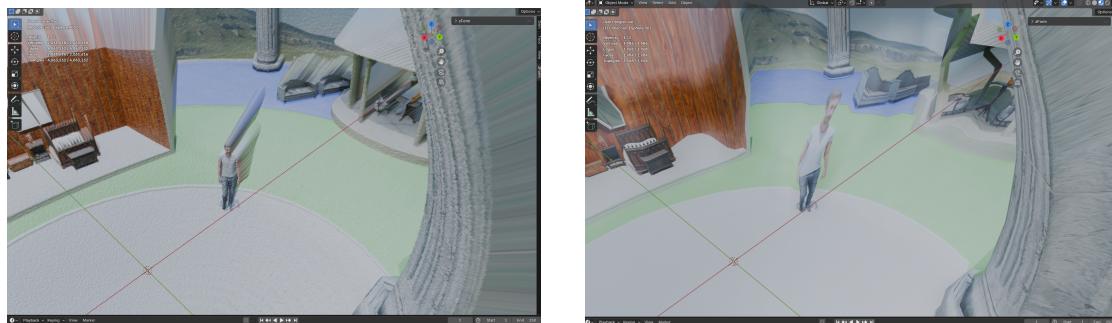
fehlerhafte Implementation zurückzuführen. Der direkte Vergleich mit Ground Truth Distanzen macht diese grundsätzliche Limitation des Ansatzes besonders deutlich, ist jedoch nicht als Hinweis auf eine fehlerhafte Implementierung zu verstehen.

Die Unterschiede in der wahrgenommenen initialen Ladezeit lassen sich auch durch die stark variierenden Datenmengen der Szenarien erklären. Szenario 1 erfordert den Transfer einer hochauflösenden 3D-Szene mit einer Gesamtgröße von 428 MB, während Szenario 3 mit rund 88 MB deutlich kompakter ausfällt. Szenario 2 basiert hingegen ausschließlich auf einem Panorama (ca. 1,9 MB) und einer zugehörigen Depth Map (ca. 262 KB). Diese Größenunterschiede könnten eine technische Erklärung dafür liefern, warum Szenario 2 von den Teilnehmenden als schneller verfügbar wahrgenommen wurde. Da jedoch keine instrumentierte Messung der Ladezeiten durchgeführt wurde, lassen sich daraus keine quantitativen Aussagen zur Performance ableiten. Die Beobachtungen stellen daher ausschließlich eine qualitative Einordnung dar.

### 6.1.2 Szenario 3: Low Poly Rekonstruktion

Szenario 3 verfolgte das Ziel, die in Szenario 2 beobachteten Verzerrungsartefakte durch eine vollständig geometriebasierte Tiefendarstellung zu vermeiden. Der ursprüngliche Ansatz sah vor, ein Low Poly Modell zu erzeugen, das ausschließlich die für die Tiefeinwahrnehmung relevanten Vordergrundobjekte enthält, während der Hintergrund als vereinfachte und speichereffiziente Umgebung realisiert wird. Auf diese Weise sollte ein 360°-Szenario entstehen, das im Gegensatz zum durchgängigen sphärischen Mesh von Szenario 2 keine global zusammenhängende Geometrie aufweist und dadurch Verzerrungen bei Bewegung reduziert.

Ein naheliegender Ansatz bestand darin, das Low Poly Modell direkt aus der Depth Map des Panoramas abzuleiten, indem die Tiefeninformation mittels Displacement Verformung in Geometrie überführt wird. In Blender erfordert dieser Ansatz eine sehr hohe Meshdichte, da das Displacement ausschließlich auf vorhandene Vertices wirkt. Um feine Tiefenabstufungen der Depth Map sichtbar zu machen, musste die zugrunde liegende Geometrie stark unterteilt werden, typischerweise durch einen hohen Grad an Subdivision. Diese notwendige Erhöhung der geometrischen Auflösung führte unmittelbar zu einer sehr hohen Polygonanzahl, sodass das resultierende Modell faktisch einem High Poly Mesh entsprach. Eine Reduktion der Subdivision Stufen führte hingegen zu einer starken Abflachung der Szene, da die Tiefe nur noch grob oder gar nicht mehr abgebildet wurde. Für webbasierte Anwendungen erwies sich dieser Ansatz daher als ungeeignet (vgl. Abbildung 6.2).



(a) resultierendes High Poly Mesh bei hoher Subdivision zur Darstellung feiner Tiefenabstufungen

(b) stark vereinfachtes Mesh bei reduzierter Subdivision mit deutlich abgeflachter Tiefenstruktur.

Abbildung 6.2: Erster Ansatz für Szenario 3

Alternative Vereinfachungsstrategien führten ebenfalls nicht zum gewünschten Ergebnis. Der Einsatz eines Decimate Modifiers auf ein zuvor hochauflöste Mesh bewahrte zwar global die grobe geometrische Form, führte jedoch zu lokalen Verzerrungen, stark ungleichmäßigen Dreiecksstrukturen und ausgeprägten visuellen Artefakten. Dies ist vermutlich darauf zurückzuführen, dass das Mesh vor der Anwendung des Decimate Modifiers keine semantisch stabile Flächenstruktur aufweist. Diese Struktur wird durch die polygonale Reduktion nicht konsistent erhalten. Beide Verfahren erwiesen sich somit als nicht praktikabel.

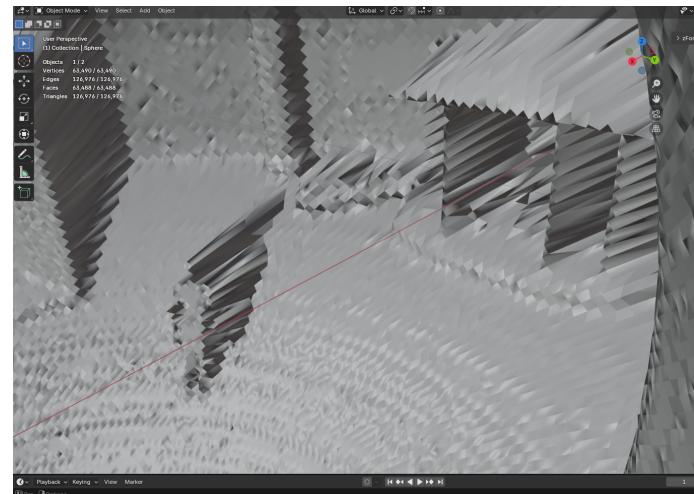


Abbildung 6.3: Darstellung des Meshes ohne Farbgebung

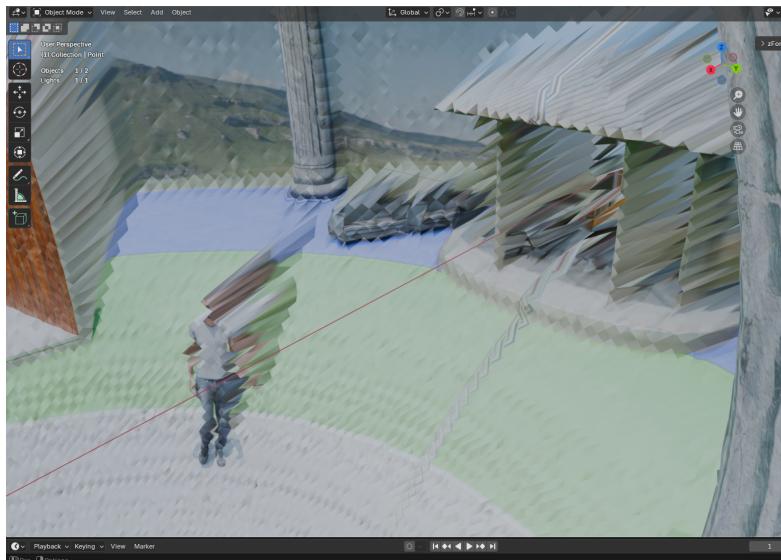


Abbildung 6.4: Farbige Darstellung derselben Szene mit identischer Geometrie

Vor diesem Hintergrund wurde für Szenario 3 ein alternativer Ansatz gewählt, der sich an Szenario 1 orientierte. Anstatt die Depth Map direkt in zusätzliche Geometrie zu überführen, wurde im finalen Szenario auf bildraumbasierte Renderingverfahren zurückgegriffen, die die Wahrnehmung vorhandener Geometrie gezielt beeinflussen. Durch aktiviertes Backface Culling wurden Rückseiten von Polygonen konsequent von der Berechnung ausgeschlossen, wodurch unnötiger Overdraw vermieden und die Laufzeitperformance stabilisiert werden konnte. Ergänzend kam Screen Space Ambient Occlusion (SSAO) als Post-Processing Verfahren zum Einsatz. SSAO berechnet lokale Kontaktverschattungen auf Basis von Tiefen- und Normaleninformationen im Bildraum und verstärkt insbesondere an Objektkanten, Übergängen und Vertiefungen die visuelle Trennung räumlicher Strukturen. Dadurch wird keine neue Tiefeninformation erzeugt, sondern die Wahrnehmbarkeit bestehender geometrischer Tiefe erhöht. Um durch SSAO entstehendes Bildrauschen bei Kamerabewegungen zu reduzieren, wurde zusätzlich temporales Anti Aliasing eingesetzt. Diese Kombination führte zu einem visuell stabileren und realistischeren Gesamteindruck, ohne die geometrische Komplexität der Szene weiter zu erhöhen. Szenario 3 stellt damit einen technisch begründeten Kompromiss dar, der geometriebasierte Tiefenkonsistenz mit vertretbarem Rechenaufwand verbindet.



Abbildung 6.5: Finales Szenario 2 mit aktivierte Renderingverfahren (Backface Culling und SSAO).

Die Ergebnisse zeigen, dass Szenario 3 in den metrischen und relativen Schätzungen stabile Werte liefert und in der subjektiven Wahrnehmung ähnlich beurteilt wird wie das Baseline Szenario. Der erhöhte Anteil der Kategorie  $GT \pm 1$  weist darauf hin, dass Distanzen überwiegend innerhalb des vorgesehenen Toleranzbereichs eingeordnet wurden. Die Tiefenrelationen wurden damit grundsätzlich korrekt wahrgenommen, auch wenn die Einordnung nicht in allen Fällen der exakten Referenzklasse entsprach. Dies ist plausibel, da die Low Poly Rekonstruktion zwar die räumliche Struktur beibehält, jedoch nicht in allen Bereichen exakt mit der hochauflösenden Geometrie von Szenario 1 übereinstimmt. Insgesamt zeigt Szenario 3, dass eine manuell erstellte, geometriebasierte Low Poly Darstellung einen praktikablen Kompromiss zwischen visueller Qualität, Tiefenkonsistenz und Speicherbedarf darstellt. Im Vergleich zu bildbasierten Verfahren vermeidet dieser Ansatz strukturelle Verzerrungen, erfordert jedoch einen deutlich höheren Modellierungs- und Vorbereitungsaufwand.

### 6.1.3 Szenario 1: Baseline High Poly

Szenario 1 diente als bewusst hochauflöste Basisszenario und stellte den Referenzpunkt für die beiden alternativen Darstellungsverfahren dar. Ziel war es, eine möglichst detaillierte und geometrisch konsistente 3D-Szene zu erzeugen, die als visuelle und strukturelle

Vergleichsgrundlage für Szenario 2 und Szenario 3 fungiert. In diesem Sinne übernimmt Szenario 1 die Rolle einer Blackbox-Referenz, an der sich sowohl bildbasierte als auch vereinfachte geometrische Verfahren messen lassen. Das Szenario repräsentiert einen idealisierten Fall der virtuellen Rekonstruktion realer Umgebungen, wie er beispielsweise durch aufwendige 3D-Scans, Pointcloud-Rekonstruktionen oder RGB-D-Aufnahmen entstehen könnte. Die vollständige geometrische Abbildung der Szene erlaubt es, Tiefenrelationen konsistent und ohne rekonstruktionsbedingte Artefakte darzustellen. Entsprechend dient Szenario 1 zugleich als Grundlage für die Definition der Ground-Truth-Werte, die in den metrischen Distanzschätzungen als Referenz herangezogen wurden und prinzipiell auch reale Messwerte einer physischen Umgebung repräsentieren könnten.



Abbildung 6.6: Finales Basisszenario S1 mit vollständig geometrischer Darstellung

In den Ergebnissen zeigt Szenario 1 durchgehend stabile und hohe Werte in den metrischen, relativen und subjektiven Bewertungen der Tiefenwahrnehmung. Auffällige Abweichungen oder unerwartete Effekte treten nicht auf, was die Rolle des Szenarios als konsistente Referenz unterstreicht. Gleichzeitig zeigen die Ergebnisse, dass Szenario 3 in vielen Aspekten ein vergleichbares Wahrnehmungsniveau erreicht, obwohl dort eine deutlich reduzierte geometrische Komplexität vorliegt. Unabhängig von der guten Wahrnehmungsqualität ist Szenario 1 jedoch nur eingeschränkt für webbasierte Anwendungen geeignet. Wie in den vorangegangenen Kapiteln bereits dargestellt, stellt insbesondere die hohe Datenmenge eine wesentliche Einschränkung für den Einsatz im Webkontext dar.

## 7 Fazit

Ziel dieser Arbeit war es, drei unterschiedliche Verfahren zur Darstellung von Panoramabildern mit visuell optimierter Tiefenwahrnehmung auf 2D-Monitoren zu untersuchen und deren jeweilige Stärken, Grenzen sowie potenzielle Einsatzbereiche zu analysieren. Im Fokus stand dabei nicht die Entwicklung eines optimalen oder universell einsetzbaren Verfahrens, sondern eine explorative Gegenüberstellung unterschiedlicher technischer Ansätze unter realistischen Rahmenbedingungen einer webbasierten Darstellung.

Hierzu wurden drei Szenarien implementiert, technisch beschrieben, iterativ optimiert und anschließend mithilfe einer Onlinestudie hinsichtlich ihrer wahrgenommenen Tiefenwirkung, Stabilität und Nutzbarkeit bewertet. Die Erhebung kombinierte subjektive Einschätzungen mit relativen und kategorialen Tiefenurteilen, um sowohl qualitative Eindrücke als auch vergleichbare Wahrnehmungstendenzen zu erfassen. Die explorative Ausrichtung der Studie ergibt sich aus der begrenzten Stichprobengröße sowie dem ordinalen Skalenniveau der erhobenen Daten. Ziel war keine statistische Hypothesenprüfung im inferenzstatistischen Sinne, sondern die Identifikation robuster Wahrnehmungstendenzen.

Auf Grundlage der theoretischen Vorüberlegungen wurden zu Beginn der Arbeit mehrere Hypothesen formuliert, die im Folgenden eingeordnet werden können.

Für Szenario 1 (High Poly Baseline) wurde eine ausgeprägte Tiefenwahrnehmung sowie ein hoher Ressourcenbedarf erwartet (H2, H3). Die Ergebnisse der Befragung stützen diese Annahmen: Szenario 1 erreichte durchgehend hohe Werte in den subjektiven Bewertungen der Tiefenwahrnehmung und zeigte stabile Ergebnisse in den relativen Einschätzungen. Gleichzeitig wurde das Szenario von vielen Teilnehmenden als vergleichsweise langsam ladend wahrgenommen. Diese Beobachtung steht im Einklang mit der sehr hohen Datensumme der Szene, wenngleich keine instrumentierten Ladezeitmessungen durchgeführt wurden. Szenario 1 kann somit als visuell überzeugende, geometrisch konsistente Referenz betrachtet werden, ist jedoch aufgrund seiner Größe und Komplexität für webbasierte Anwendungen nur eingeschränkt geeignet.

Für Szenario 2 (Panorama mit Depth Map) wurde eine geringe und potenziell verzerrte Tiefenwahrnehmung angenommen (H1). Die Ergebnisse deuten darauf hin, dass dieser Ansatz zwar eine sehr kompakte und ressourcenschonende Darstellung ermöglicht, die Tiefenwirkung jedoch häufig als instabil oder wenig realistisch wahrgenommen wurde. Insbesondere bei Kamerabewegungen traten Verzerrungen auf, die den Eindruck einer

konsistenten räumlichen Struktur beeinträchtigten. Gleichzeitig wurde Szenario 2 von vielen Teilnehmenden als schnell verfügbar wahrgenommen, was sich plausibel, jedoch nicht quantitativ abgesichert, durch die sehr geringe Datenmenge erklären lässt. Insgesamt zeigt Szenario 2 damit ein hohes Potenzial im Hinblick auf Effizienz und Einfachheit, jedoch deutliche Einschränkungen hinsichtlich der wahrgenommenen Tiefenqualität.

Für Szenario 3 (Low Poly Rekonstruktion) wurde eine klare und stabile Tiefenwahrnehmung bei reduziertem Ressourcenbedarf erwartet (H4). Die Ergebnisse legen nahe, dass dieses Szenario in vielen Aspekten ein Wahrnehmungsniveau erreicht, das dem High Poly Basisszenario nahekommt, obwohl die geometrische Komplexität deutlich reduziert ist. Die Kombination aus vereinfachter Geometrie, gebackenen Texturen und gezielten Rendering Optimierungen führte zu einer insgesamt stabilen und plausiblen Tiefenwirkung. Szenario 3 kann daher als technisch begründeter Kompromiss zwischen visueller Qualität und Performance eingeordnet werden.

**H1 (Depth Map):** wurde bestätigt – verzerrte und instabile Tiefenwahrnehmung

**H2 (High Poly):** wurde bestätigt – höchste Tiefenqualität

**H3 (High Poly Performance):** wurde tendenziell bestätigt, jedoch nur qualitativ

**H4 (Low Poly):** wurde weitgehend bestätigt – stabile Tiefenrelationen bei reduzierter Komplexität

Zusammenfassend zeigt die Untersuchung, dass sich die drei betrachteten Verfahren deutlich hinsichtlich ihrer Eignung für webbasierte Anwendungen unterscheiden. Während vollständig geometriebasierte High Poly Szenen eine hohe Tiefenqualität ermöglichen, sind sie mit erheblichen Nachteilen hinsichtlich Datenmenge und Ladezeit verbunden. Rein bildbasierte Ansätze auf Basis von Depth Maps sind äußerst effizient, weisen jedoch wahrnehmbare geometrische Verzerrungen auf, die ihre Einsatzmöglichkeiten einschränken. Low Poly Rekonstruktionen mit gebackenen Texturen erscheinen demgegenüber als besonders ausgewogener Ansatz, der sowohl visuell als auch technisch praktikabel ist.

Die Arbeit ist bewusst explorativ angelegt und unterliegt mehreren Einschränkungen. Die Stichprobengröße ist begrenzt, es wurden keine statistischen Signifikanztests durchgeführt, und Performance Aspekte wurden überwiegend auf Basis subjektiver Wahrnehmung bewertet. Objektive Performance-Messungen hätten eine weiterführende Quantifizierung ermöglicht, lagen jedoch außerhalb des Fokus dieser wahrnehmungsorientierten Untersuchung. Entsprechend lassen sich aus den Ergebnissen keine generalisierbaren Aussagen ableiten, sondern vielmehr Tendenzen und qualitative Einordnungen.

Für zukünftige Arbeiten könnten insbesondere hybride Ansätze weiter untersucht werden, etwa die Kombination mehrerer Tiefenschichten oder segmentierter Depth Maps, um Verzerrungen bildbasierter Verfahren zu reduzieren. Ebenso wäre eine instrumentierte Messung von Ladezeiten und Ressourcenverbrauch sinnvoll, um subjektive Wahrnehmungen mit objektiven Kennwerten zu verknüpfen. Trotz der genannten Einschränkungen liefert die vorliegende Arbeit einen strukturierten Vergleich dreier grundlegender Ansätze und zeigt auf, unter welchen Bedingungen welche Verfahren für webbasierte 3D-Darstellungen mit Tiefenwirkung sinnvoll eingesetzt werden könnten. (Die genannten Einschränkungen beeinflussen vor allem die Generalisierbarkeit der Ergebnisse, nicht jedoch deren interne Vergleichbarkeit innerhalb der drei Szenarien.)

## Literaturverzeichnis

Adobe. What is Baking ? | Substance 3D bakers, 2025. URL <https://helpx.adobe.com/content/help/en/substance-3d-bake/getting-started/what-is-baking.html>. [9]

Omeed Ashtiani, Hung-Jui Guo, and Balakrishnan Prabhakaran. Impact of motion cues, color, and luminance on depth perception in optical see-through AR displays. 4:1243956, 2023. ISSN 2673-4192. 10.3389/frvir.2023.1243956. URL <https://www.frontiersin.org/articles/10.3389/frvir.2023.1243956/full>. [18]

Babylon.js Contributors. Babylon.mesh.applydisplacementmap, 2025. URL <https://doc.babylonjs.com/typedoc/classes/BABYLON.Mesh#applydisplacementmap>. Zugriff am 25.12.2025. [37]

Babylon.js Documentation. Back face culling, 2025a. URL [https://doc.babylonjs.com/features/featuresDeepDive/materials/using/materials\\_introduction/#back-face-culling](https://doc.babylonjs.com/features/featuresDeepDive/materials/using/materials_introduction/#back-face-culling). Zugriff am 29. Dez. 2025. [V, 30, 31]

Babylon.js Documentation. Ssao rendering pipeline, 2025b. URL <https://doc.babylonjs.com/features/featuresDeepDive/postProcesses/SSAORenderPipeline>. Zugriff am 29. Dez. 2025. [30]

Blender Foundation. Cameras - Blender 5.0 Manual, 2025a. URL [https://docs.blender.org/manual/en/latest/render/cycles/object\\_settings/cameras.html](https://docs.blender.org/manual/en/latest/render/cycles/object_settings/cameras.html). [7]

Blender Foundation. Render Baking - Blender 4.5 LTS Manual, 2025b. URL <https://docs.blender.org/manual/en/latest/render/cycles/baking.html>. [9, 10]

Blender Manual. Passes, 2025. URL <https://docs.blender.org/manual/en/latest/render/layers/passes.html>. Zugriff am 29. Dez. 2025. [26]

Blender Manual Decimate. Decimate modifier, 2025. URL <https://docs.blender.org/manual/en/latest/modeling/modifiers/generate/decimate.html>. Zugriff am 29. Dez. 2025. [30]

Haram Choi and Sanghun Nam. Effect of eye-tracking-based depth perception visualization interface on virtual reality user experience. 30(1):5, 2025. ISSN 1434-

9957. 10.1007/s10055-025-01260-8. URL <https://link.springer.com/10.1007/s10055-025-01260-8>. [18]
- Murat Coban, Yusuf Islam Bolat, and Idris Goksu. The potential of immersive virtual reality to enhance learning: A meta-analysis. *Educational Research Review*, 36:100452, 2022. 10.1016/j.edurev.2022.100452. URL <https://www.sciencedirect.com/science/article/pii/S1747938X22000215>. [14]
- Giorgia Committeri, Valentina Sebastiani, Francesco De Pasquale, Massimiliano Stocchi, and Chiara Fini. Functional Autonomy Affects Elderly Spatial Perception in Body-Centered Coordinates. 2020:1–8, 2020. ISSN 2090-2204, 2090-2212. 10.1155/2020/5694790. URL <https://www.hindawi.com/journals/jar/2020/5694790/>. [18]
- L. G. Divyanth, Divya Rathore, Piranav Senthilkumar, Prakhar Patidar, Xin Zhang, Manoj Karkee, Rajendra Machavaram, and Peeyush Soni. Estimating depth from rgb images using deep-learning for robotic applications in apple orchards. *Smart Agricultural Technology*, 6:100345, 2023. 10.1016/j.atech.2023.100345. URL <https://linkinghub.elsevier.com/retrieve/pii/S2772375523001740>. [10, 11]
- E. Bruce Goldstein. *Sensation and Perception*. Cengage Learning, Belmont, CA, 8 edition, 2013. ISBN 9781133958499. [14, 15]
- Google Developers. Photo Sphere XMP Metadata | Street View, 2023. URL <https://developers.google.com/streetview/spherical-metadata>. [7]
- Rubén Grande, Javier Albusac, David Vallejo, Carlos Glez-Morillo, and José Jesús Castro-Schez. Performance Evaluation and Optimization of 3D Models from Low-Cost 3D Scanning Technologies for Virtual Reality and Metaverse E-Commerce. 14(14):6037, 2024. ISSN 2076-3417. 10.3390/app14146037. URL <https://www.mdpi.com/2076-3417/14/14/6037>. [11]
- Xian-Feng Han, Hamid Laga, and Mohammed Bennamoun. Image-based 3D Object Reconstruction: State-of-the-Art and Trends in the Deep Learning Era. 43(5):1578–1604, 2021. ISSN 0162-8828, 2160-9292, 1939-3539. 10.1109/TPAMI.2019.2954885. URL <http://arxiv.org/abs/1906.06543>. [16]
- Lu He, Bing Jian, Yangming Wen, Haichao Zhu, Kelin Liu, Weiwei Feng, and Shan Liu. Rethinking Supervised Depth Estimation for 360° Panoramic Imagery. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 5169–5177. IEEE, 2022-06. ISBN 978-1-6654-8739-9.

- 10.1109/CVPRW56347.2022.00565. URL <https://ieeexplore.ieee.org/document/9857323/>. [12]
- Andreas Hergovich. *Allgemeine Psychologie: Wahrnehmung und Emotion*. utb GmbH, 3 edition, 2022. ISBN 978-3-8385-5434-1. 10.36198/9783838554341. URL <https://elibrary.utb.de/doi/book/10.36198/9783838554341>. [5, 6]
- Paul B. Hibbard, Rebecca L. Hornsey, and Jordi M. Asher. Binocular Information Improves the Reliability and Consistency of Pictorial Relief. 7(1):1, 2023-03. ISSN 2411-5150. 10.3390/vision7010001. URL <https://www.mdpi.com/2411-5150/7/1/1>. [5]
- Peter Hoffmann. *Beyond (Multi-) Media: Multimediaformen erklärt: Von Panoramabildern über 3D bis zu den immersiven Welten des Metaversums*. X.media.press. Springer Fachmedien Wiesbaden, 2025. ISBN 978-3-658-48566-5 978-3-658-48567-2. 10.1007/978-3-658-48567-2. URL <https://link.springer.com/10.1007/978-3-658-48567-2>. [6, 7, 14]
- Dongjin Huang, Nan Wang, Xinghan Huang, Jiantao Qu, and Shiyu Zhang. Mesh-controllable multi-level-of-detail text-to-3d generation. *Computers & Graphics*, 123: 104039, 2024. 10.1016/j.cag.2024.104039. URL <https://www.sciencedirect.com/science/article/pii/S0097849324001742>. [16]
- Shintaro Ito, Kanta Miura, Koichi Ito, and Takafumi Aoki. Neural Radiance Field-Inspired Depth Map Refinement for Accurate Multi-View Stereo. 10(3):68, 2024-03-08. ISSN 2313-433X. 10.3390/jimaging10030068. URL <https://pmc.ncbi.nlm.nih.gov/articles/PMC10971679/>. [11]
- Shahram Izadi, David Kim, Otmar Hilliges, David Molyneaux, Richard Newcombe, Pushmeet Kohli, Jamie Shotton, Steve Hodges, Dustin Freeman, Andrew Davison, and Andrew Fitzgibbon. KinectFusion: Real-time 3D reconstruction and interaction using a moving depth camera. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology*, pages 559–568. ACM, 2011. ISBN 978-1-4503-0716-1. 10.1145/2047196.2047270. URL <https://dl.acm.org/doi/10.1145/2047196.2047270>. [11]
- Bernd Jähne. *Digitale Bildverarbeitung: und Bildgewinnung*. Springer Berlin Heidelberg, 2024. ISBN 978-3-662-59509-1 978-3-662-59510-7. 10.1007/978-3-662-59510-7. URL <https://link.springer.com/10.1007/978-3-662-59510-7>. [10]
- Julian Knodt, Zherong Pan, Kui Wu, and Xifeng Gao. Joint UV Optimization and

- Texture Baking. 43(1):2:1–2:20, 2023. ISSN 0730-0301. 10.1145/3617683. URL <https://dl.acm.org/doi/10.1145/3617683>. [9]
- H. Lee, Manolis Savva, and Angel X. Chang. Text-to-3d shape generation. *Computer Graphics Forum*, 43(2):e15061, 2024. 10.1111/cgf.15061. URL <https://onlinelibrary.wiley.com/doi/10.1111/cgf.15061>. [16]
- Karsten Lehn, Merijam Gotzes, and Frank Klawonn. *Grundlagen der Computergrafik: Eine Einführung mit OpenGL und Java*. Springer Fachmedien Wiesbaden, 2022. ISBN 978-3-658-36074-0 978-3-658-36075-7. 10.1007/978-3-658-36075-7. URL <https://link.springer.com/10.1007/978-3-658-36075-7>. [6]
- Žiga Lesar, Ciril Bohak, and Matija Marolt. Evaluation of depth perception in crowded volumes. 120:103918, 2024. ISSN 00978493. 10.1016/j.cag.2024.103918. URL <https://linkinghub.elsevier.com/retrieve/pii/S0097849324000530>. [17, 18]
- Shuai Li, Shufang He, Yuanrui Dong, Caihong Dai, Jinyuan Liu, Yanfei Wang, and Hiroaki Shigemasu. Depth Perception Based on the Interaction of Binocular Disparity and Motion Parallax Cues in Three-Dimensional Space. 25(10):3171, 2025. ISSN 1424-8220. 10.3390/s25103171. URL <https://www.mdpi.com/1424-8220/25/10/3171>. [15]
- Christian Lipski, Felix Klose, and Marcus Magnor. Correspondence and Depth-Image Based Rendering a Hybrid Approach for Free-Viewpoint Video. 24(6):942–951, 2014. ISSN 1558-2205. 10.1109/TCSVT.2014.2302379. URL <https://ieeexplore.ieee.org/document/6727475/>. [10]
- Meshy AI. Meshy ai: Ki-basierte bild-zu-3d-modellerstellung, 2025. URL <https://www.meshy.ai/de>. [16]
- Yamato Miyashita, Yasuhito Sawahata, Akihiro Sakai, Masamitsu Harasawa, Kazuhiro Hara, Toshiya Morita, and Kazuteru Komine. Display-Size Dependent Effects of 3D Viewing on Subjective Impressions. 19(2):1–15, 2022. ISSN 1544-3558, 1544-3965. 10.1145/3510461. URL <https://dl.acm.org/doi/10.1145/3510461>. [6]
- Poly Haven. Poly haven asset browser, 2025. URL <https://superhivemarket.com/products/poly-haven-asset-browser>. [24]
- Manuel Rey-Area, Mingze Yuan, and Christian Richardt. 360monodepth: High-resolution 360° monocular depth estimation. *arXiv preprint*, arXiv:2111.15669, 2022. 10.48550/arXiv.2111.15669. URL <https://arxiv.org/abs/2111.15669>. [12]

Roboflow. What is depth anything v2: Depth estimation tutorials. <https://blog.roboflow.com/depth-anything/>, 2025. Accessed: 2025-10-23. [10]

Arindam Saha, Bibhas Chandra Dhara, Saiyed Umer, Kulakov Yurii, Jazem Mutared Alanazi, and Ahmad Ali AlZubi. Efficient Obstacle Detection and Tracking Using RGB-D Sensor Data in Dynamic Environments for Robotic Applications. 22(17):6537, 2022. ISSN 1424-8220. 10.3390/s22176537. URL <https://pmc.ncbi.nlm.nih.gov/articles/PMC9460380/>. [10]

Michael Schmeing and Xiaoyi Jiang. Depth image based rendering. In Patrick S. P. Wang, editor, *Pattern Recognition, Machine Intelligence and Biometrics*, pages 279–310. Springer, Berlin, Heidelberg, 2011. 10.1007/978-3-642-22407-2\_12. URL [https://link.springer.com/chapter/10.1007/978-3-642-22407-2\\_12](https://link.springer.com/chapter/10.1007/978-3-642-22407-2_12). [10]

Rainer Schönhammer. *Einführung in die Wahrnehmungspychologie: Sinne, Körper, Bewegung*. Number 3142 in UTB. Facultas.wuv, 2., überarb., aktualisierte u. erw. aufl edition, 2013. ISBN 978-3-8252-4076-9. [5]

Thiago Lopes Trugillo da Silveira, Paulo Gamarra Lessa Pinto, Jeffri Erwin Murrugarra Llerena, and Claudio Rosito Jung. 3d scene geometry estimation from 360° imagery: A survey. *ACM Computing Surveys*, 55(4):1–39, 2023. 10.1145/3519021. URL <https://dl.acm.org/doi/10.1145/3519021>. [12, 13]

Shubham Singh, Yash Sharma, Amer Liaqat, and Roy S. Kalawsky. Evaluation of XR device's real-world tracking accuracy and depth perception from an industrial point of view. 29(3):118, 2025. ISSN 1434-9957. 10.1007/s10055-025-01192-3. URL <https://link.springer.com/10.1007/s10055-025-01192-3>. [18]

John P. Snyder and Philip M. Voxland. An album of map projections. Technical Report 1453, U.S. Geological Survey, 1989. URL <https://pubs.usgs.gov/pp/1453/report.pdf>. [7]

Statista. Vr-headsets – weltweit: Marktprognose und umsatzdaten, 2025. URL <https://de.statista.com/outlook/cmo/consumer-electronics/gaming-equipment/vr-headsets/weltweit>. Abgerufen am 18. Dezember 2025. [14]

Ayush Tewari, Justus Thies, Ben Mildenhall, Ricardo Martin-Brualla, Stephen Lombardi, Wang Yifan, Jonathan T. Barron, Paul Debevec, Deva Ramanan, Steven M. Seitz, and Maneesh Agrawala. Neural rendering and reconstruction. *ACM SIGGRAPH Courses*, 2023. 10.1145/3587421. URL <https://arxiv.org/abs/2206.07631>. [16]

The Khronos Group. gltf — runtime 3d asset delivery, 2020. URL <https://www.khronos.org/gltf/>. [25]

Guangcong Wang, Peng Wang, Zhaoxi Chen, Wenping Wang, Chen Change Loy, and Ziwei Liu. PERF: Panoramic Neural Radiance Field from a Single Panorama, 2023. URL <http://arxiv.org/abs/2310.16831>. [12]

Ramin Zabih and John Woodfill. Non-parametric local transforms for computing visual correspondence. In Jan-Olof Eklundh, editor, *Computer Vision — ECCV '94*, volume 801, pages 151–158. Springer Berlin Heidelberg, 1994. ISBN 978-3-540-57957-1 978-3-540-48400-4. 10.1007/BFb0028345. URL <http://link.springer.com/10.1007/BFb0028345>. [18]

Wei Zeng, Sezer Karaoglu, and Theo Gevers. Joint 3D Layout and Depth Prediction from a Single Indoor Panorama Image. In Andrea Vedaldi, Horst Bischof, Thomas Brox, and Jan-Michael Frahm, editors, *Computer Vision – ECCV 2020*, volume 12361, pages 666–682. Springer International Publishing, 2020. ISBN 978-3-030-58516-7 978-3-030-58517-4. 10.1007/978-3-030-58517-4\_39. URL [https://link.springer.com/10.1007/978-3-030-58517-4\\_39](https://link.springer.com/10.1007/978-3-030-58517-4_39). [13]

Chen Zhang, Biao He, Renzhong Guo, and Ding Ma. 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, 2023. 10.1080/17538947.2023.2177758. URL <https://www.tandfonline.com/doi/full/10.1080/17538947.2023.2177758>. [9]

Ke Colin Zheng, Sing Bing Kang, Michael F. Cohen, and Richard Szeliski. Layered Depth Panoramas. In *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE, 2007. ISBN 978-1-4244-1179-5 978-1-4244-1180-1. 10.1109/CVPR.2007.383295. URL <http://ieeexplore.ieee.org/document/4270320/>. [12]

## **Anhang A**

### **Hinweis zur Nutzung von KI-Werkzeugen Und**

Zur sprachlichen und stilistischen Überarbeitung dieser Arbeit wurden unterstützend die Werkzeuge ChatGPT (OpenAI) sowie DeepL Write verwendet.

Die Nutzung beschränkte sich auf Formulierungs-, Grammatik- und Stilkorrekturen. Die inhaltliche Konzeption, Analyse, Auswertung sowie die Interpretation der Ergebnisse erfolgten eigenständig durch den Autor.

## Anhang B

# Hinweis zu begleitenden Materialien und Reproduzierbarkeit

Alle im Rahmen dieser Arbeit entwickelten Szenarien, Quellcodes, 3D-Modelle, Texturen, Depth Maps, Panoramen sowie Auswertungsdateien und Arbeitsmaterialien wurden in begleitenden Online-Repositorien bereitgestellt.

Das GitHub-Repository enthält insbesondere:

- den vollständigen Quellcode der Babylon.js-Szenarien,
- die verwendeten 3D-Modelle im GLB-Format,
- Panorama- und Depth-Map-Daten,
- Blender-Arbeitsdateien,
- sowie direkte Verweise zu den lauffähigen Szenarien.

Zusätzlich wurden umfangreiche Arbeits- und Auswertungsdateien, darunter Tabellen, Rohdaten und weitere projektbezogene Materialien, in einem separaten Cloud-Speicher abgelegt. Die bereitgestellten Materialien dienen der Transparenz, Nachvollziehbarkeit und Reproduzierbarkeit der Arbeit und wurden aus Umfangsgründen nicht vollständig in den Anhang der PDF-Version integriert.

Die begleitenden Materialien sind unter folgenden Links abrufbar:

- GitHub-Repository: [https://github.com/NisVison/Enis\\_Bachelorarbeit\\_Abgabe25\\_Archiv.git](https://github.com/NisVison/Enis_Bachelorarbeit_Abgabe25_Archiv.git)
- Cloud-Speicher (Auswertungen und Arbeitsmaterialien): <https://sync.academiccloud.de/index.php/s/XoxvDBu8LWw8j31>