BACHELORARBEIT

Head-Motion Parallax in gerahmtem, stereoskopischem 3DoF+ Video

Experimentelle Bestimmung des Einflusses von Rahmung und Skalierung einer 3D-Szene auf die Wahrnehmung selbstinduzierter, lateraler Head-Motion Parallax in VR

vorgelegt am 9. Februar 2024

Kai Nüske

Erstprüfer: Prof. Dr. Marco Grimm
Zweitprüfer: Prof. Dr. Eike Langbehn

HOCHSCHULE FÜR ANGEWANDTE WISSENSCHAFTEN HAMBURG
Department Medientechnik
Finkenau 35
22081 Hamburg

Zusammenfassung

Im VR-Kontext ist eine natürliche Übersetzung der realen in die virtuelle Bewegung essenziell für eine angenehme Erfahrung. Vom durch 3DoF+/6DoF[1] gewonnenen Tiefenhinweis Motion Parallax können nicht nur 360°-Medien, sondern auch gerahmte, virtuelle Szenen profitieren: Stereoskopisches Video, bei dem der Kamerablickwinkel passend zur Translation des Betrachters reagiert. Dies ist die erste Arbeit, die die Wahrnehmung natürlicher, lateraler Translationsübersetzung in gerahmten Szenenabbildern zum Gegenstand hat: Der Einfluss dieser Rahmung, sowie der Skalierung des Rahmeninhalts (Brennweite) auf die natürlich empfundene Motion Parallax wird mithilfe eines psychophysischen VR-Experiments untersucht. Rahmung alleine bewirkte keine Veränderung der natürlichen Übersetzung, allerdings der Empfindlichkeit für Abweichungen: Diese fiel deutlich höher aus, da die unverändert gezeigte Umgebung als Referenz dienen kann – die Teilnehmer passten ihre reale Translationsgeschwindigkeit deutlich weniger an die virtuelle an. Eine 2x-Vergrößerung wurde mit einer um den Faktor $g_t = 0,75$ gestauchten, eine 0,5x-Vergrößerung um $g_t = 1,51$ gestreckten Übersetzung als realistisch empfunden. Die Sensibilität für Abweichungen zeigte sich deutlich niedriger für die verkleinerte Projektion und leicht erhöht für die vergrößerte.

Abstract

In the VR context, a natural translation of real movement into virtual movement is essential for a pleasant experience. Not only 360° media, but also windowed virtual scenes can benefit from the depth cue motion parallax gained through 3DoF+/6DoF[2]: Stereoscopic video, where the camera viewing angle reacts to match the viewer's translation. This is the first work to address the perception of natural lateral translation in windowed scene images: The influence of this windowing, as well as the scaling of the windowed content (focal length) on naturally perceived motion parallax is investigated using a psychophysical VR experiment. Windowing alone did not change the natural translation, but the sensitivity to deviations did: This turned out to be significantly higher, as the unchanged surroundings can serve as a reference - the participants adapted their real translation speed significantly less to the virtual one. A 2x magnification was perceived as realistic with a translation compressed by the factor $g_t = 0.75$, a 0.5x magnification stretched by $g_t = 1.51$. The sensitivity to deviations was significantly lower for the downscaled projection and slightly higher for the enlarged projection.

[1] Bewegungs- und Rotationsfreiheit entlang aller Raumachsen

[2] freedom of movement and rotation along all spatial axes

Inhaltsverzeichnis

Abbildungsverzeichnis

Tabellenverzeichnis

Abkürzungsverzeichnis

Abkürzung	Bedeutung
2AFC-Methode	Methode der erzwungenen Wahl zwischen zwei Alternativen (engl. two-alternative forced choice)
3DoF	drei Freiheitsgrade der Bewegung: Rotation um alle Raumachsen
3DoF+	s. 3DoF; erweitert um räumlich begrenzte Translation entlang aller Raumachsen
6DoF	s. 3DoF; erweitert um räumlich unbegrenzte Translation entlang aller Raumachsen
CI	Konfidenzintervall (engl. confidence interval)
DoF	Freiheitsgrade der Bewegung (engl. degrees of freedom)
DT	Erkennungsschwellwert (engl. detection threshold)
HVS	menschliches, visuelles System (engl. human visual system)
IPD	Interpupillardistanz
M/PR	Motion/Pursuit Ratio nach Nawrot und Stroyan (2009)
MLE	Maximum Likelihood Estimation
PSE	Punkt subjektiver Gleichheit (engl. point of subjective equality)
RU	Realumgebung
SEM	Standardfehler des Mittelwerts (engl. standard error of the mean)
VU	virtuelle Umgebung

1 Einführung

Immersive Bewegtbildtechnologien erweitern die Betrachtungserfahrungen gegenüber herkömmlichem Bewegtbild, um durch visuellen Realismus und Interaktivität die Empfindung von Telepräsenz (sich virtuell am Ort der gezeigten Szene zu befinden) zu ermöglichen und in das Geschehen „einzutauchen“ (engl. to immerse). (Alain et al., 2023, S. 3) Durch stetig günstiger und technisch ausgereifter werdende Virtual-Reality-Headsets sinkt die Hürde der Verbraucher für den Einstieg in immersive Bewegtbilderfahrungen. Dass erst Meta (Meta, 2024) und nun auch Apple (Apple, 2024a) in die Entwicklung und Vermarktung der Idee Virtual-/Mixed-Reality investieren, gibt Einblick in die Zukunftsvision der Konzerne von Mensch-Computer-Interaktion und damit einhergehendem Bewegtbildkonsum.

Stereoskopisches Video (3D-Video) erweiterte den visuellen Realismus von herkömmlichem Video (2D-Video) durch die Darstellung von je einer Szenenperspektive pro Auge statt nur einer Perspektive für beide Augen (mehr dazu in Abschnitt 2.3). Die Technologie wurde umfangreich erforscht (Tam et al., 1998, Stelmach et al., 1999, Devernay und Beardsley, 2010), blieb aber bisher primär im Anwendungsfeld Kino relevant. Heimanwender- und Verbrauchergeräte erreichten langfristig nur geringen Absatz.

1.1 Identifikation der Forschungslücke

VR-Headsets bieten ebenfalls ein Display pro Auge, weshalb 3D-Videoinhalte auch außerhalb des Kinos wieder an Relevanz gewinnen könnten.

Nach einer Umfrage aus 2012 vom NPD war für vier von fünf Befragten der Grund gegen die Anschaffung eines 3D-Fernsehers, dass spezielle Brillen zum Anschauen benötigt wurden (Follows, 2017) – Wenn nun die VR-Headsets („VR-Brillen“) den Weg in den Alltag der Verbraucher schaffen, könnte diese Art Brille wiederum eine neue Welle verbraucherseitiger Nachfrage an 3D-Video bewirken.

Auch autostereoskopische Displays finden sich mittlerweile vermehrt in Produkten wie Laptops und Tablets (ASUS, 2024, Acer, 2024). Mit den darin auf dem Display verbauten Mikrolinsen-Arrays ist stereoskopische Szenendarstellung nun auch ohne 3D-Brille möglich und das nicht mehr rein statisch, sondern auch abhängig vom Betrachtungswinkel. Realisiert

wird dies durch eingebaute Head- und Eye-Tracking-Technologie: Anhand der gewonnenen Positionsdaten der Augen des Betrachters, kann das Bild je Auge der Perspektive entsprechend angepasst und ausgegeben werden. VR-Headsets erfassen ebenso die Kopf- und teilweise Augenposition, mit gleichem Resultat einer dynamischen Perspektivanpassung.

1.1.1 3DoF+/6DoF Video ermöglicht Motion Parallax

Diese Dimension der Bildbetrachtung wird in 3D-Video nicht berücksichtigt – Der Betrachter hat keinen Freiheitsgrad in der Wahl der eigenen Betrachtungsperspektive, genannt 0DoF (DoF = degrees of freedom z. Dt. Freiheitsgrade). Die Bezeichnungen verschiedener Abstufungen von Freiheitsgraden wird zumeist nach dem Entwurf der MPEG-I (Champel et al., 2018) vorgenommen: In 360°-Videos kann die Blickrichtung frei gewählt werden, also die *Rotation* der Perspektive um alle drei Raumachsen (x, y und z), wodurch die Bezeichnung *3DoF* geprägt wird (Champel et al., 2018, S. 3, Alain et al., 2023, S. 4, Van Der Hooft et al., 2023, S. 1336). Die Translation (Positionsänderungen im Raum) wird allerdings auch hier nicht übersetzt. Wenn die Blickwinkelwahl durch *Translation und Rotation* bezüglich aller drei Raumachsen uneingeschränkt erfolgen kann, spricht man von *6DoF* (Champel et al., 2018, S. 3, Boyce et al., 2021, S. 1522, Alain et al., 2023, S. 4). *3DoF+* ist eine eingeschränkte Form von *6DoF*, wo die Translation auf ein Volumen (engl. viewing baseline) beschränkt ist, in dem Perspektiven für den Betrachter zur Verfügung stehen – zum Beispiel sitzend (Champel et al., 2018, S. 3, Jeong et al., 2019, S. 136399, Alain et al., 2023, S. 4). In der Literatur wird 3DoF+ teils auch 6DoF genannt (Thatte et al., 2017, S. 1, Broxton et al., 2020, S. 1). Da Unterscheidung von 3DoF+ und 6DoF eine präzisere Beschreibung ermöglicht, werden diese Schreibweisen im Folgenden verwendet. Der MPEG-I Entwurf führt zudem gerahmtes (engl. „windowed“) 6DoF-Video: Dies ist in der Translationsfreiheit in der Vorwärts-Achse eingeschränkt und durch Fensterung/Rahmung einer virtuellen Szene nicht mehr 360° um den Betrachter, sondern vor ihm als Ausschnitt – die anderen Bewegungsachsen sind uneingeschränkt (Champel et al., 2018, S. 3). Diese Arbeit beschäftigt sich mit einer Mischform der aufgeführten Formate: **gerahmtes, stereoskopisches 3DoF+ Video**. Es übernimmt den Rahmungsaspekt, ist jedoch zur Betrachtung innerhalb eines in allen Achsen begrenzten Volumens angedacht (z. B. sitzend; Dies ist aufnahmeseitig begründet, mehr dazu in Abschnitt 1.3.2.). Die Zusammenhänge der MPEG-I Definitionen werden in Abbildung 1.1 veranschaulicht (Champel et al., 2018, S. 4–5).

Durch Translation bei Betrachtung einer statischen Szene bewegen sich Szenenelemente im Bild in unterschiedlichen Geschwindigkeiten und Richtungen, abhängig von ihrer relativen Tiefe – Dieser Zusammenhang wird in Modellen zur Erklärung der menschlichen Tiefenwahrnehmung genutzt und als Motion Parallax bezeichnet (Rogers und Graham, 1979, S. 125, Nawrot und Stroyan, 2009, S. 1, Yoonessi und Baker, 2013, S. 1, Nawrot et al., 2014, S. 1;

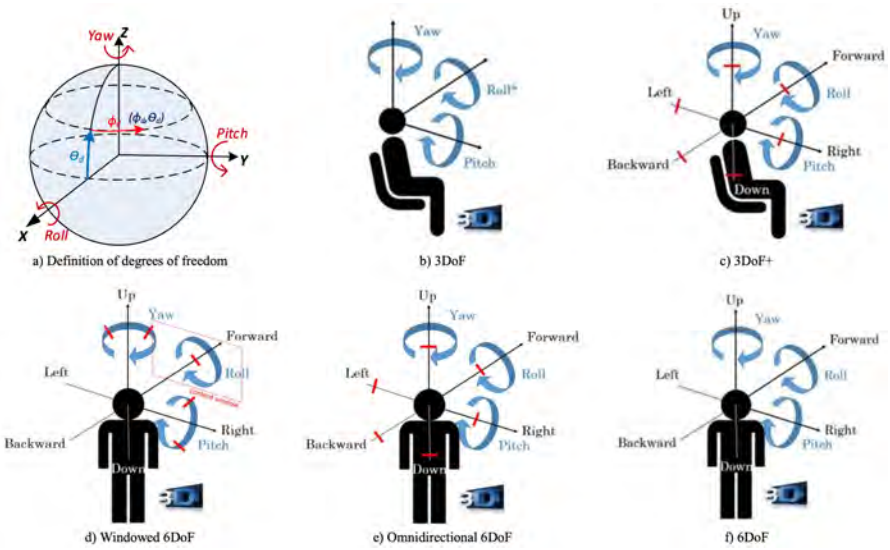


Abbildung 1.1: Illustration der Freiheitsgrade nach Definition aus MPEG-I Entwurf – Champel et al., 2018 S. 4–5

mehr dazu in Abschnitt 2.1.1). Mehr als drei Freiheitsgrade (3DoF+, 6DoF) machen somit die Wahrnehmung selbstinduzierter Motion Parallax erst möglich.

In der herkömmlichen Filmproduktion wird Motion Parallax als Gestaltungsmittel zur Verstärkung der Tiefenwahrnehmung mittels Kameraführung eingesetzt; Beispiele hierfür sind Kamerafahrten um einen Fixationspunkt herum oder parallel einem bewegten Fixationspunkt folgende Dolly-Fahrten. Diese Translationen sind somit jedoch fremdinduziert und nicht selbstinduziert vom Betrachter (Nielsen, 2007, S. 53–56) – folglich nicht interaktiv. Die selbstinduzierte Motion Parallaxe stellt somit das Novum von 3DoF+/6DoF Video gegenüber herkömmlichem 3D-Video dar.

1.1.2 Translationsverstärkung und Rahmung

Die technische Realisierung dieser Freiheitsgrade stellt die gesamte Produktionskette vor neue Herausforderungen: Es werden für jeden Blickwinkel innerhalb des erlaubten Betrachtungsvolumens Bilddaten benötigt, um flüssige Übergänge im Szeneninhalt während jeder Bewegung des Betrachters zu zeigen. Während dies im Fall von computergenerierten Bildern über Simulationen aus expliziter Geometrie, Texturen und Lichtberechnungen möglich ist, ist das Erfassen und Verarbeiten realer Szenendaten weitaus komplexer (mehr dazu in Abschnitt 1.3.2). Für beide Szenarien stellt sich dennoch die wichtige Frage, wie die in der Realumgebung (ab hier RU) gemessene Kopftranslation T_r in die Kopftranslation T_v in der virtuellen Umgebung (ab hier VU) übersetzt werden sollte, sodass die Wahrnehmung der Motion Parallax vom Betrachter als realitätsnah empfunden wird. Diese Übersetzung wird in

der Literatur Translationsverstärkung g_t genannt (z. B. Steinicke et al., 2010, S. 20, Langbehn et al., 2018, S. 3, Serrano et al., 2020a, S. 3, mehr dazu in Abschnitt 2.4 und 3) und definiert durch:

$$g_t = \frac{T_v}{T_r} \tag{1.1}$$

Gilt die intuitive Antwort der 1:1-Übersetzung ($g_t = 1$) zwischen RU und VU in jedem Fall? In welchem Wertebereich von g_t (bei welcher Stauchung oder Streckung der virtuellen Translation T_v) bleiben Veränderungen in der Übersetzung unbemerkt? Im Forschungsfeld „VR Locomotion“ (dt. VR-Fortbewegung)/„Redirected Walking“ (dt. umgeleitetes Laufen) konnten bereits viele Erkenntnisse zur Translationsverstärkung g_t bei gehender Fortbewegung in VR-Anwendungen gesammelt werden (dazu mehr in Abschnitt 3.1). Zur Wahrnehmung der Translationsverstärkung g_t bei reiner Kopf-/Oberkörperbewegung im Sitzen (engl. „seated VR“) gibt es deutlich weniger Daten. Die dafür vorhandenen Ergebnisse beschränken sich bisher überdies auf vollumfängliche VU, seltener sind auch stereoskopische 3DoF+ 360° Videos Untersuchungsgegenstand (Teng et al., 2023, S. 404, Serrano et al., 2020a, S. 9). Daten zur Wahrnehmung beim Blick durch ein virtuelles Fenster/einen virtuellen Rahmen, in dem der Szenenbildausschnitt der Kopftranslation des Betrachters folgt, fehlen zum Zeitpunkt dieser Arbeit noch in Gänze.

1.2 Forschungsfragen und Zielsetzung

Ziel dieser Bachelorarbeit ist, Daten in der beschriebenen Forschungslücke zu erheben, die einen ersten Eindruck der Wahrnehmung realitätsgetreuer Motion Parallax in gerahmtem, stereoskopischen 3DoF+ Video ermöglichen. Die ersten beiden Forschungsfrage zielen daher auf den Einfluss der Rahmung auf jene Wahrnehmung ab:

FF1: Inwiefern beeinflusst die Rahmung einer Szene die als natürlich wahrgenommene Translationsverstärkung g_t der im Rahmen sichtbaren, durch laterale Kopfbewegungen selbstinduzierten Motion Parallax in VR?

FF2: Um welchen Faktor lässt sich die virtuelle Kopftranslation T_v (für den Blick durch den Rahmen) gegenüber der realen T_r stauchen oder strecken, ohne dass die Veränderung eindeutig vom Betrachter wahrgenommen wird?

In einer 360° Umgebung sind die Größen- und Tiefenverhältnisse stets perspektivisch unverzerrt und getreu der Tiefenstaffelung der Szene zu zeigen. Nur so entsteht der Eindruck einer vollumfänglich realistischen VU und die Blickwinkeländerung durch Rotation verhält sich wie erwartet. Einzig die spezifisch vom virtuellen Betrachtungssystem hervorgerufenen Verzerrungen der Szenenwahrnehmung sind zu entzerren (Thatte et al., 2017, S. 2–3,

Broxton et al., 2020, S. 5). Durch die Rahmung des gezeigten Szeneninhalts wird es erst möglich, den Szeneninhalt *im Rahmen* in Skalierung und Tiefenstaffelung anzupassen. Wie bei herkömmlichem 3D-Video können Szenen dann überlebensgroß oder verkleinert gezeigt werden – ein essenzielles Gestaltungsmittel des Mediums Bewegtbild. Diese Möglichkeit der Perspektivverzerrung wirft zwei weitere Forschungsfrage auf:

FF3: Inwiefern beeinflusst eine Skalierung des gerahmten Szeneninhalts (Brennweite f der Kamera) die als natürlich wahrgenommene Translationsverstärkung g_t der im Rahmen sichtbaren, durch laterale Kopfbewegungen selbstinduzierten Motion Parallax in VR?

FF4: Wie wird die Sensibilität der Betrachter für Abweichungen von der als natürlich wahrgenommenen Translationsverstärkung g_t durch die Veränderung der Skalierung des gerahmten Szeneninhalts (Brennweite f der Kamera) beeinflusst?

Eingangs wird die aktuelle Relevanz des Formats „gerahmtes, stereoskopisches 3DoF+ Video“ aufgezeigt, indem sein Mehrwert, sowie aktuelle technische Entwicklungen und Hürden der Herstellung und Verbreitung dargelegt werden, zu deren Minderung die Ergebnisse dieser Arbeit beitragen könnten. Um eine Methode zur Beantwortung der Forschungsfragen zu entwerfen, wird in die Theorie der Tiefenwahrnehmung eingeführt, um dann die Rolle der untersuchten Motion Parallax herauszuarbeiten. Das Kapitel 2 ist wie eine Übersichtslektüre zu wahrnehmungsseitigen Parametern der Projektion in dem Videoformat zu verstehen: Der Forschungsstand und wichtige Modelle zu relevanten Tiefenhinweisen werden zusammengefasst und Wechselwirkungen aufgezeigt. Die geometrischen Zusammenhänge der Stereoskopie, Projektion (inkl. Skalierung/Brennweite f) und Translationsverstärkung g_t werden mithilfe eines entworfenen Geometrie-Tools[1] veranschaulicht. Dadurch lassen sich in Kapitel 3 die aktuellen Forschungsergebnisse zur Translationsverstärkung g_t in VR einordnen und die Wahl der Methode theoretisch fundiert treffen – relevante Daten aus dem verwandten Forschungsfeld VR-Locomotion werden hinsichtlich Implikationen für das eigene Experiment zusammengeführt. Aus den bisher wenigen Arbeiten zu lateraler Translationsverstärkung wird eine Abwandlung bezüglich des untersuchten Rahmungs- und Projektionseffekts entworfen: Das resultierende, psychophysische VR-Experiment wird in Kapitel 4 beschrieben. Die Forschungsfragen lassen sich dann anhand der statistisch ausgewerteten Ergebnisse in Kapitel 5 beantworten, die in Kapitel 6 eingeordnet werden. Abschließend werden Thesen zu Ursachen und Implikationen der Messwerte diskutiert.

[1] s. Anhang GeoGebra Online-Tool, S. 70

Abbildung 1.2: Beispielbild: gerahmtes, stereoskopisches 3DoF+ Video (Apple Spatial Video) – Apple, 2024a

1.3 Relevanz der Arbeit

Im Vorstellungsvideo der neu erschienen Apple Vision Pro wurde „Spatial Video“ (dt. räumliches Video) als neues Feature vorgestellt (Apple, 2024b, Minute 3:04 – 3:36): ein gerahmtes, stereoskopisches Videoformat mit der Möglichkeit, sich in der Szene umzuschauen – im Prinzip gerahmtes, stereoskopisches 3DoF+ Video (s. Abb. 1.2). Auch die Forschung beschäftigt sich mit dem Mehrwert, den Motion Parallax bei stereoskopischen Bildinhalten bieten kann:

1.3.1 Mehrwert von Motion Parallax

So folgerten Kongsilp und Dailey (2017) aus ihrer Studie, dass durch Motion Parallax das Empfinden von Präsenz erhöht und visuelle Ermüdung reduziert werden könne (S. 76–77). Thatte und Girod (2018) schlossen aus ihrem erfassten, subjektiven Meinungsbild aus randomisierten Doppelvergleichen, dass in 360°-Videos Motion Parallax einen größeren, positiven Einfluss auf den Qualitätseindruck habe als die stereoskopische Darstellung (S. 2). Auch die Befragung von Serrano et al. (2019) zeigte eine klare Präferenz von 360° Video mit Motion Parallax verglichen mit herkömmlichem 360° Video, sowie weniger Unwohlsein der Betrachter (S. 7–8). Miyashita et al., 2022 untersuchten gerahmtes, stereoskopisches 3DoF+ Video: Stereoskopische Darstellung und Motion Parallax verstärkten den Eindruck der ausgewerteten Parameter „Power“ und „Space“ (dt. Ausdrucksstärke und Räumlichkeit), wobei sich die stereoskopische Darstellung insgesamt negativ auf „Visibility“ (dt. Sichtbarkeit) auswirkte – vermutlich der technischen Umsetzung der Stereoskopie geschuldet [2] (S. 7–10). Fulvio et al. (2021) konnten zeigen, dass selbst kleinste translative Bewegungen bei eigentlich statischer Betrachtung bereits die Tiefenwahrnehmung verbessern (S. 8).

[2] siehe Diskussionsteil Miyashita et al., 2022 S. 11–12

Die Berichte hochwertigerer und angenehmerer Erfahrungen könnte auf die sensorischen Konflikte zurückzuführen sein, die durch die statisch vorgegebene Perspektive bei herkömmlichen 3D-Videos entstehen (LaViola, 2000, S. 50–52). Die Betrachtung herkömmlicher 3D-Videos in VR schnitt verglichen mit Betrachtung auf einem 3D-Bildschirm oder einer panoramischen 3D-Projektion bei Choy et al. (2021) hinsichtlich der Betrachtungserfahrung deutlich schlechter ab, mit mehr Motion Sickness und visuellen Ermüdungserscheinungen (S. 9591) – hierbei könnte die Ergänzung von Motion Parallax demnach möglicherweise für eine signifikante Verbesserung sorgen. Sie hilft laut IJsselsteijn et al. (2008) ferner dabei, den Eindruck eines realitätsnahen Fensters zu vermitteln (S. 278).

Die freie Wahl der Betrachtungsperspektive verspricht somit viele Vorteile – die richtige technische Umsetzung dieser ist allerdings unbedingt sicherzustellen: Selzer et al. (2022) zeigten, dass zu hohe g_t zu Cybersickness führen kann (S. 1466). Die Forschungsfragen dieser Arbeit sind somit auch hinsichtlich des Betrachtungskomforts wichtig zu beantworten.

1.3.2 4.5D/5D Lichtfelderfassung

Die Implementierung von Motion Parallax in die stereoskopische Bewegtbildwiedergabe könnte somit ein wichtiger nächster Schritt für noch immersivere und angenehmere Betrachtungserfahrungen sein. Die technischen Herausforderungen der Implementierung sind tagesaktueller Gegenstand in der Forschung:

Einordnung des untersuchten Videoformats Beginnend bei der Aufzeichnung von gerahmtem, stereoskopischem 3DoF+ Video, ist das Forschungsfeld um Lichtfelderfassung und -repräsentation zu erwähnen: Gerahmtes, stereoskopisches 3DoF+ Video ließe sich auch als Lichtfeldvideo (teils genannt 4.5D Lichtfeld Herfet et al., 2023, S. 180) oder 5D Lichtfeld beschreiben. Diese zeichnen sich dadurch aus, die aus einem beliebigen Punkt im Raum (x, y, z) sichtbare Lichtintensität aus beliebiger Richtung (θ, ϕ) über die Dimension der Zeit (t) aufzunehmen (Chelli et al., 2020, S. 1). Die vollständige siebendimensionale plenoptische Funktion zur Beschreibung eines jeden Lichtstrahls im Raum über die Zeit, würde ergänzend die Wellenlänge λ enthalten (J. Liu et al., 2023, S. 469):

$$L_{7D} = P(x, y, z, \theta, \phi, t, \lambda) \tag{1.2}$$

Da die Radianz entlang eines Lichtstrahls zwischen zwei Punkten im Raum zum selben Zeitpunkt als konstant angenommen wird, ist eine der Dimensionen redundant; Es bleibt eine fünfdimensionale Funktion, weshalb von 5D Lichtfeldern gesprochen wird (Van Der Hooft et al., 2023 S. 1359, Bae et al., 2019, S. 1, Chelli et al., 2020, S. 1). Da die Lichtstrahlen durch ihre Position und Richtung eindeutig beschreibbar sind, können sie als Schnittpunkte zwischen

zwei Ebenen in Koordinatenschreibweise ($a,\ b$) und ($x,\ y$) mit beschrieben werden – Es ergibt sich (Chelli et al., 2020, S. 1, Bae et al., 2019, S. 3, Alain et al., 2023, S. 12):

$$L(a, b, x, y, t) \tag{1.3}$$

Video bestehend aus sequentiell mit synchroner Bildrate aufgezeichneten 4D Lichtfeldern (Standbilder ohne Zeitdimension t), wird 4.5D Lichtfeld genannt, da die Zeitdimension t nicht kontinuierlich, sondern nur in den diskret aufgenommenen Zeitschritten dargestellt wird (Herfet et al., 2023, S. 180). Werden die Zeitschritte dazwischen interpoliert, ist die Bezeichnung 5D Lichtfeld geläufig. Die Aufzeichnung der Lichtintensität aus jedem beliebigen Blickpunkt mit beliebiger Ausrichtung ermöglicht eine freie Wahl des Blickwinkels in das aufgezeichnete View Frustum, dem Volumen der aufgezeichneten Lichtstrahlen (Flynn et al., 2019, S. 2367, Broxton et al., 2020, S. 2).

Die Beschreibung „gerahmtes, stereoskopisches 3DoF+ Video“ wurde für diese Arbeit gewählt, da sie die Betrachtungsmodalitäten in den Vordergrund stellt und nicht die Art der Informationsbeschreibung/Gewinnung, wie es die Lichtfeld-Begrifflichkeit tut. Auch gänzlich computergenerierte Bildinhalte können in dem Videoformat wiedergegeben werden; Bei diesen handelt es sich jedoch nicht um aufgezeichnete Lichtfelder, weshalb die in dieser Arbeit gewählte Bezeichnung als Ausspielformat zu verstehen und den 4.5D/5D Lichtfeldern überzuordnen ist.

Relevanz der Arbeit für 4.5D/5D Lichtfelderfassung Die untersuchten Fragestellungen sind im Besonderen für die Lichtfelderfassung relevant: Gängige Lichtfeldaufnahmesysteme setzen auf Kamera-Arrays zur Erfassung vieler Blickwinkel auf eine Szene (Wilburn et al., 2001, S. 33–34, Smith et al., 2009, S. 1, Anderson et al., 2016, S. 4–5 Song et al., 2017, Milliron et al., 2017, S. 1, Wu et al., 2017, S. 930, Flynn et al., 2019, S. 2367, J. Zhang et al., 2020, S. 3, Chelli et al., 2020, S. 1–2, Broxton et al., 2020, S. 3–4, Thatte und Girod, 2021, S. 31). Für den Fall gerahmter Szenen (nicht 360°) sind planare Arrays üblich: Die Kameras sind in einer Ebene angeordnet und gleichmäßig verteilt. Je nach Kamera-Array Konfiguration entsteht ein unterschiedlich großes Volumen, in dem der Betrachter die Blickwinkel frei wählen kann (genannt viewing baseline, entsprechend der 3DoF+ Beschreibung). Grundlegend gilt vereinfacht, je größer und dichter das Kamera-Array, desto größer die viewing baseline (Keinert et al., 2023, S. 235–237, Thatte et al., 2017, S. 2, Flynn et al., 2019, S. 2367). Gängig sind hierbei Abmessungen in der Größenordnung $\pm 20 \ldots 40$ *cm* Translation vom Ursprungspunkt (Thatte et al., 2017, S. 2, Broxton et al., 2020, S. 2, Bertel et al., 2020, S. 127, Keinert et al., 2023, S. 237). Thatte und Girod (2018) fanden bei sitzender Betrachtung von stereoskopischen 360° 3DoF+ Videos, dass über 90 % der horizontalen, translativen Auslenkung unterhalb von 30 *cm* fallen, selbes gilt für vertikale, translative Auslenkung unterhalb 8, 5 *cm* (S. 3). Bei freier Betrachtung gerahmter, stereoskopischer 3DoF+ Videoinhalte zeigte das Teilnehmerverhalten

bei Miyashita et al. (2022), dass 95 % der gemessenen Kopftranslationen in einen Wertebereich von 31 … 47 *cm* lateraler Auslenkung fielen (S. 10).

Die untersuchten Forschungsfragen FF2 und FF4 sind daher relevant für die Lichtfeldaufnahme, da sie zeigen können, wie weit sich die virtuelle Translation T_v unbemerkt stauchen lässt. Dadurch wird eine weniger große, reale viewing baseline benötigt. Die enormen Datenmengen der vielen Einzelkameras des Arrays sind aktuell eine Hürde in der Verarbeitung von 3DoF+ Inhalten, die so reduziert werden könnte. Das aufzuzeichnende Sichtvolumen der Szenengeometrie kann zudem reduziert werden, was auch bei der Verbreitung der Ergebnisse Datenmenge einspart. Ferner lässt sich aus der Beantwortung von Forschungsfragen FF1 und FF2 ableiten, wie die Größenverhältnisse des Kamera-Arrays je Brennweite f zu wählen sind.

Alle Blickwinkel innerhalb der viewing baseline, die nicht exakt von den Einzelkameras erfasst wurden, müssen durch Blickwinkelsynthese (engl. view synthesis) und/oder geometrischer Rekonstruktion interpoliert werden. Die Forschung ist in diesem Bereich in den letzten Jahren sehr aktiv, mit verschiedenen Ansätzen (siehe Keinert et al., 2023 für eine umfassende Übersicht), dennoch sind weiterhin Artefakte in z. B. verdeckten, hochkomplexen, fein aufgelösten oder reflektierenden Szenenbereichen in den Interpolationen enthalten (Broxton et al., 2020, S. 7 Fig. 6, Richardt et al., 2020, S. 16 Fig. 6, Bonatto et al., 2021, S. 146876 Fig. 10, Attal et al., 2021, S. 5–7, Yan et al., 2022, S. 3897, Zou et al., 2022, S. 645, Jin et al., 2022, S. 594, Keinert et al., 2023, S. 242). Durch eine gestauchte Translationsübersetzung könnte bei selber Kameraanzahl und somit selber Rohdatenmenge die Dichte erhöht und weniger Interpolation benötigt werden. Erste Ergebnisse von Serrano et al., 2020a zeigten, dass durch dynamische Stauchung der Translationsgeschwindigkeit g_t die Sichtbarkeit von Interpolationsartefakten in 360° 3DoF+ Videos reduziert und die wahrgenommene Qualität somit erhöht werden konnte (S. 9–11; mehr dazu in Abschnitt 3.3).

Thematische Abgrenzung Das Zielbild von Lichtfelderfassung stellt eine möglichst akkurate, physikalische Repräsentation des aufgenommenen Sichtvolumens dar. Das hier untersuchte Videoformat erfordert diese Realitätsentsprechung nicht, weshalb Skalierung (Brennweite f) und Projektion (Abstand und Maße des Rahmens) des Volumens im Rahmen nicht nur möglich, sondern explizit als Gestaltungsmittel erwünscht sind. Der dem untersuchten Videoformat verwandte Begriff „Fishtank VR“ (dt. Aquarium VR) wird in dieser Arbeit nicht verwendet, da die praktischen Umsetzungen davon sehr unterschiedlich (von kugelförmigen, über kubisch transparenten bis vergleichbar gerahmten Volumina; Kongsilp und Dailey, 2017, S. 73) ausfallen und der Begriff kontextuell eher bei VR-Anwendungen als bei der Bewegtbildwiedergabe verortet wird.

2 Theoretischer Hintergrund

Um fundiert eine Methode entwerfen zu können, die zur Beantwortung der Forschungsfragen aus Abschnitt 1.2 beitragen kann, gilt es, die technischen und wahrnehmungsseitigen Eigenschaften des Videoformats „gerahmtes, stereoskopisches 3DoF+ Video“ zu beleuchten. Der Fokus dieser Arbeit liegt auf der Untersuchung der Wahrnehmung von Motion Parallax. Motion Parallax tritt weder in realen Betrachtungsumgebungen, noch im Zielbild des Videoformats völlig isoliert auf, sondern bildet stets einen Teil des gesamtheitlichen Szeneneindrucks des HVS. Somit wird im Folgenden eingangs die Unterteilung der Wahrnehmung in sogenannte „Tiefenhinweise“ – wie Motion Parallax – diskutiert und der Begriff Motion Parallax, sowie seine Funktion als Tiefenhinweis eingeordnet. Anschließend werden Modelle beschrieben, die Erklärungsversuche dafür anbieten, wie das HVS aus Sinnesreizen und projektiver Geometrie (also dem retinalen Abbild der Szene) auf die reale Geometrie der Szene schließen könnte. Außerdem werden Wechselwirkungen und Ähnlichkeiten mit anderen im untersuchten Videoformat relevanten Tiefenhinweisen aufgezeigt. Dadurch wird die theoretische Grundlage der Motion Parallax Wahrnehmung für den Methodenentwurf und die Ergebnisdiskussion geschaffen.

Um darüber hinaus die Forschungsfragen FF3 und FF4 fundiert beantworten zu können, ist der Einfluss der Skalierung des gerahmten Szenenabbilds auf die projektive Geometrie in Abschnitt 2.4 erörtert. Basierend auf den zuvor gewonnen Erkenntnissen zu Modellen der Funktionsweise von Motion Parallax, lassen sich so Vermutungen hinsichtlich des Einflusses der Skalierung aufstellen.

Da das Videoformat stereoskopischer Natur ist und sich in seiner technischen Umsetzung auf das Errechnen virtueller Kameraperspektiven stützt (s. Abschnitt 1.3.2), werden abschließend für den Methodenentwurf relevante, geometrischen Grundgrößen und -begriffe der Stereoskopie, sowie ihre Auswirkung auf die Tiefenwahrnehmung umrissen.

2.1 Theorie der Tiefenhinweise

In der Fachliteratur zur menschlichen Tiefenwahrnehmung ist das Konzept einzelner „Tiefenhinweise“ (engl. „depth cues“) sehr verbreitet (James J. Gibson., 1950 S. 19–22, Cutting und

Tabelle 2.1: Kategorische Einordnung von Tiefenhinweisen

	Monokular	**Binokular**
Visuell	Motion Parallax, Schattierung, Okklusion, Höhe im Sichtfeld Texturgradienten lineare Perspektive atmosphärische Perspektive relative Größe …	Binokulare Disparität
Okulomotorisch	Akkomodation	Vergenz

Vishton, 1995, S. 78–79, J. Liu et al., 2023, S. 473). Diese Unterteilung der gesamtheitlichen Tiefenwahrnehmung wird auf Helmholtz „Handbuch der Physiologischen Optik“ zurückgeführt, der hier vom Begriff „Zeichen“ (von Helmholtz und Nagel, 1910 nach Rogers, 2022, S. 295) Gebrauch machte. Die einzelnen Hinweise sind heutzutage in der Literatur meist in vier Kategorien nach zwei Eigenschaften eingeordnet: Einerseits wird zwischen monokularen (mit einem Auge wahrnehmbaren) und binokularen (nur mit beiden Augen im Verbund wahrnehmbaren) Hinweisen unterschieden. Andererseits werden visuelle (aus dem retinalen Szenenbild wahrnehmbar) und okulomotorische (aus kinästhetischen Empfindungen von Muskeln um den optischen Apparat wahrnehmbar) Hinweise getrennt (J. Liu et al., 2023, S. 473). Die kategorische Einteilung gebräuchlicher Tiefenhinweise ist in Tabelle 2.1 aufgeführt:

In der Klassifizierung immersiver Bewegtbildformate ist diese Benennung einzelner Tiefenhinweise unerlässlich, um die Formate anhand ihrer Charakteristika zu unterscheiden: Monoskopische 2D-Bildwiedergabesysteme stellen ausschließlich monokulare, visuelle Tiefenhinweise dar, während stereoskopische 3D-Videos die Tiefenwahrnehmung aus binokularen Tiefenhinweisen ermöglichen (J. Liu et al., 2023, S.472). Das untersuchte Format „gerahmtes, stereoskopisches 3DoF+ Video“ übernimmt diese stereoskopischen Tiefenhinweise des 3D-Video und erweitert sie um den Untersuchungsgegenstand dieser Arbeit: die vom Betrachter selbstinduzierte Motion Parallax.

2.1.1 Definition und Klassifizierung von Motion Parallax

Motion Parallax wird als monokular und visuell klassifiziert. (Hartle und Wilcox, 2021 S. 1, J. Liu et al., 2023, S. 473) – wenn auch die Bewegung der Augen essenzieller Bestandteil in

Modellen der Gewinnung von Tiefeninformation aus Motion Parallax ist, wie im weiteren Verlauf deutlich wird (s. Abschnitt 2.2.2).

Rogers und Graham, 1979 beschrieben Motion Parallax als „the relative movement of images across the retina resulting from movement of the observer or the translation of objects across his field of view“ ([Die relative Bewegung von Bildern über die Retina, resultierend aus Bewegung des Betrachters oder der Translation von Objekten über sein Gesichtsfeld] (Übers. d. Verf.), S. 125). Bewegt sich ein Betrachter lateral in einer statischen Szene und fixiert dabei einen Punkt im Raum, kann er aus den wahrgenommenen, relativen Geschwindigkeiten der Objekte im Sichtfeld auf die Tiefenstaffelung der Szene schließen. Je nach Tiefe der Szenenobjekte im Raum geschieht dies in unterschiedliche Richtungen, mit unterschiedlichen Geschwindigkeiten (Nawrot und Stroyan, 2009, S. 1969, Nawrot et al., 2014, S. 1): Objekte, die vor dem fixierten Objekt liegen, bewegen sich schneller als dies und entgegen der Bewegungsrichtung des Betrachters. Objekte hinter dem fixierten Objekt bewegen sich langsamer und mit der Bewegungsrichtung (Nawrot et al., 2014, S. 1). Neben der Information aus den Veränderungen des retinalen Szenenabbilds, stehen dem bewegten Betrachter ergänzende nicht-visuelle Signale als mögliche Informationsquellen zur Tiefenwahrnehmung aus Motion Parallax zur Verfügung: Rogers und Graham, 1979 führten hierfür kinästhetische (Kinästhetik = Lehre von der Bewegungsempfindung) und vestibuläre (Gleichgewichtssystem im Innenohr) Informationen auf (S. 127), Nawrot und Joyce, 2006 fassten als „extra-retinal signal“ (S. 4710) zusammen. Da Motion Parallax auch im Falle eines statischen Betrachters wirkt, wird im aktuellen Forschungsstand besonders der Verfolgungsbewegung der Augen (um Fixation auf die Objekte zu bewahren) eine Schlüsselrolle für eine eindeutige Tiefenwahrnehmung zugeschrieben (Holmin und Nawrot, 2015, S. 1, mehr dazu in Abschnitt 2.2.2). Diese ist dem statischen und bewegten Betrachter gemein.

2.1.2 Kritik am Konzept der Tiefenhinweise

Im Forschungsfeld der menschlichen Tiefenwahrnehmung wurde häufig die Isolation einzelner Tiefenhinweise (Nawrot und Stroyan, 2009, S. 1974–1977,Buckthought et al., 2017, S. 2 Teng et al., 2023, S. 401–406) oder derer vermuteten Bestandteile (Rogers und Graham, 1979, S. 127, Yoonessi und Baker, 2011, S. 2, Yoonessi und Baker, 2013, S. 2) in experimentellen Aufbauten mit speziell dafür entworfenen Stimuli versucht, um Einsicht in die einzelnen Wahrnehmungszusammenhänge zu erlangen; so auch für den hier untersuchten Tiefenhinweis Motion Parallax: Rogers und Graham, 1979 deklarierten, dass Motion Parallax nach ihren Ergebnissen einen „zuverlässigen, konsistenten und eindeutigen Eindruck von relativer Tiefe *in Abwesenheit aller anderen Hinweise auf Tiefe und Distanz* [produziert]“ ([Übers. d. Verf.], [Hervorh. d. Verf.] Rogers und Graham, 1979, S. 1). Somit zeigten sie, dass der Mensch aus

Motion Parallax als isoliertem Hinweis, Informationen über die Szenengeometrie erlangen kann.

Nun ist es mitunter einer der Verfasser selbst, der aktuell Kritik am Konzept der wahrnehmungsseitig isolierten Kausalität der Gewinnung von Tiefeninformation aus einzelnen Tiefenhinweisen übt (Rogers, 2022): Während er die okulomotorischen Tiefenhinweise Vergenz und Akkomodation (für Erläuterungen zu diesen, siehe Hoffman et al., 2008, S. 2–3) hierbei ausklammert, merkt er an, dass alle visuellen Tiefenhinweise grundlegend einer gemeinsamen Informationsquelle – dem retinalen Szenenabbild – entspringen. Denn alle visuellen Tiefenhinweise sind prinzipiell auf den Zusammenhang projektiver Geometrie zurückzuführen (Rogers, 2022, S. 295–296): Das von Szenenelementen reflektierte Licht fällt auf die Retina. Die visuellen Tiefenhinweise sind damit insgesamt als eine Projektion der dreidimensionalen Szenenumgebung auf die zweidimensionale Retina zusammenfassbar. Die Tiefenhinweise folgen dann aus der Unterscheidung einzelner Erscheinungen *innerhalb* der projektiven Geometrie des optischen Apparats. Rogers kritisiert, dass die konzeptionelle Trennung der visuellen Tiefenhinweise andeute, dass die Tiefenwahrnehmung „höhere, kognitive Prozesse“ aus der Abwägung einzelner Hinweise impliziere und die Hinweise ein Produkt der „Erkenntnis über die Wahrnehmung“ ([Übers. d. Verf.] Rogers, 2022, S. 295) beschreiben, nicht die Wahrnehmung selbst. Er schließt sich damit der Kritik von Pagel, 2019 an, von dem die beiden Ausdrücke übernommen sind. Auch Maniatis, 2021 veröffentlichte eine Reihe an Preprints mit dem Titel „The myth of visual depth cues“ mit einer ähnlichen Einschätzung.

Entscheidend festzuhalten ist im Kontext dieser Arbeit die Erkenntnis, dass unumstritten projektive Geometrie, also perspektivische Projektion, informationsseitig der Grundstein visueller Tiefenwahrnehmung ist. Der Tiefenhinweis binokulare Disparität ließe sich laut Rogers auch als „binocular perspective“ (Rogers, 2022, S. 296) und Motion Parallax als „motion perspective“ (Rogers, 2022, S. 296) ausdrücken. Motion Parallax setzt somit die Informationen aus der Perspektive des Auges auf die Szene in einen *zeitlichen Zusammenhang* (Veränderungen der Perspektive durch Bewegung) und binokulare Disparität die Perspektiven beider Augen in einen *örtlichen Zusammenhang* (Unterschied der Perspektiven zum selben Zeitpunkt).

Die Erkenntnisse im Bereich Vision Science zu diesen beiden Tiefenhinweisen, auf denen der Fokus dieser Arbeit liegt, sind dennoch äußerst relevant zur Beantwortung der Forschungsfragen: Die Beobachtungen und entworfenen Modelle für sich liefern in ihren parametrischen Annahmen akkurate und reproduzierbare Ergebnisse. Auch wenn die einzelnen Tiefenhinweise möglicherweise nicht ursächlich für die Tiefenwahrnehmung sind, wie die Kritik verlauten lässt (Maniatis, 2021, S. 2, Rogers, 2022, S. 297), können die Modelle dazu dienen, valide Prognosen für neue visuelle Reize und ihre Wahrnehmung aufzustellen. Ferner ist die tatsächlich akkurate Wahrnehmung von Tiefe oder die zutreffendste Beschreibung der Zusammenhänge im HVS nicht direkt Untersuchungsgegenstand.

2.2 Tiefenwahrnehmung aus Motion Parallax

Wie in Abschnitt 2.1.1 beschrieben, stehen dem HVS zur Tiefenwahrnehmung aus Motion Parallax retinale (projektive Geometrie) und extra-retinale (Kinästhetik, Okulomotorik) Informationen zur Verfügung. Untersuchungen hinsichtlich der projektiven Geometrie versuchen, Erscheinungen *innerhalb* der projektiven Geometrie zu isolieren und den Einfluss und die Wirksamkeit dieser im Vergleich miteinander herauszuarbeiten (Yoonessi und Baker, 2011, Yoonessi und Baker, 2013, Nawrot und Stroyan, 2009).

2.2.1 Visuelle Wirkung von Motion Parallax

Da für verschiedene Punkte im Raum, abhängig von ihrer Tiefe und dem Betrachterverhalten (Translation und Fixation), unterschiedliche Bewegungen ihrer retinalen Abbilder entstehen, werden besonders die Kanten/Übergänge von Objekten analysiert. Das Verhalten an parallel zur Translationsrichtung verlaufenden Kanten wird von dem an orthogonal zur Translationsrichtung verlaufenden Kanten getrennt:

Motion Parallax an Kanten parallel zur Translation Kanten, die parallel zur Translationsrichtung verlaufen, zeigen eine Scherbewegung (engl. „shearing motion“). Ein weiter entfernt liegender Punkt bewegt sich langsamer als ein näher liegender, wodurch an Kanten mit abrupten Tiefenübergängen eine scherenartige Bewegung im retinalen Abbild beobachtet werden kann (Yoonessi und Baker, 2011, S. 1–2). Diese Bewegungsmuster von Punkten auf der Retina werden auch „Optic Flow“ Komponenten/Muster genannt. Fixiert der Betrachter einen Punkt zwischen zwei Objekten bei der Translation, sind die Optic Flow Muster bidirektional: die Punkte bewegen sich in unterschiedliche Richtungen. (s. Abb. 2.1(a) und Abb.2.1(b)) Die Informationen aus den Scherbewegungen sind nach den Ergebnissen von Yoonessi und Baker, 2011 zum einen bei der Erkennung von Objektkanten mit abrupten Tiefenübergängen hilfreich. Bei einer Tiefenordnungsaufgabe hingegen, wo nicht die Kantenerkennung, sondern die tatsächliche Tiefenreihenfolge untersucht wurde, schnitten die Teilnehmer bei feineren Abstufungen (also graduellen Tiefenunterschieden) besser ab als bei abrupten (Yoonessi und Baker, 2011, S. 9–11). Schlussfolgern ließe sich daraus, dass die Scherbewegungen zur Segmentierung von Objekten und nicht so sehr ihrer Tiefenrelation beiträgt, dafür aber einen Beitrag zu Wahrnehmung der Tiefenrelation und plastischen Tiefe eines Objekts selbst leisten könnte (Yoonessi und Baker, 2011, S. 18).

Motion Parallax an Kanten orthogonal zur Translation An Kanten, die orthogonal zur Translationsrichtung verlaufen, wird zwischen zwei Phänomenen unterschieden:

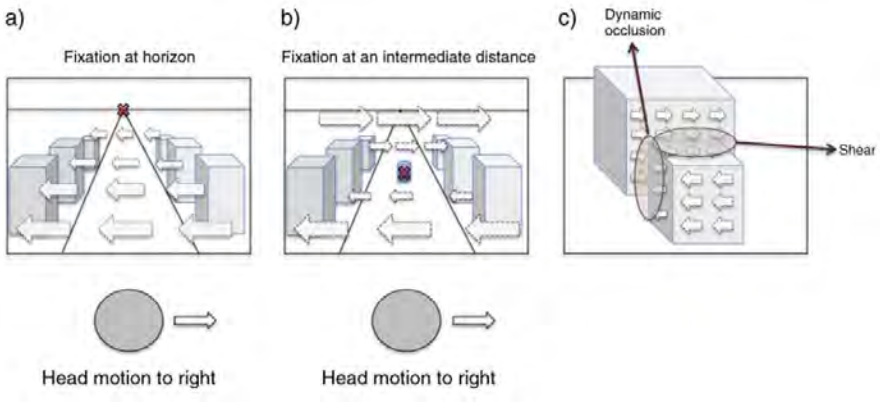


Abbildung 2.1: Muster retinaler Bildbewegung „Optic Flow Muster“: (a) Fixation auf den Horizont (b) Fixation mittlerer Distanz (c) Verhalten an Kanten orthogonal (Dynamische Okklusion) und parallel (Scherbewegung) zur Translationsrichtung – Yoonessi und Baker, 2011 Fig. 1

Expansion/Kompression (engl. „expansion“/„compression“) sowie Zu-/Aufdecken (engl. „deletion“/„accretion“). Zusammengefasst werden sie als dynamische Okklusion (engl. „dynamic occlusion“) bezeichnet (s. Abb. 2.1(c);Yoonessi und Baker, 2013, S. 1). Expansion/Kompression ist dabei ähnlich der Scherbewegung eine „Optic Flow“-Komponente, die die gemeine Bewegungsrichtung und -geschwindigkeit von Punkten innerhalb einer Tiefenebene beschreibt. Sie ist als Tiefenhinweis ohne Translation des Betrachters uneindeutig: Unterschiedliche Geschwindigkeiten und Bewegungsrichtungen von z. B. Punktwolken bieten für sich keinen Aufschluss über die Tiefenstaffelung, nur über die Segmentierung (Yoonessi und Baker, 2013, S. 2). Zu-/Aufdecken beschreibt die Beobachtung, dass näher liegende Punkte weiter entfernte abhängig von der Betrachterposition verdecken/aufdecken. Auch ohne Translation des Betrachters liefert Zu-/Aufdecken eindeutige Tiefeninformationen; die verdeckenden Elemente liegen eindeutig vor den von ihnen zu-/aufgedeckten Bildbereichen (Yoonessi und Baker, 2013, S. 2).

In den Experimenten von Yoonessi und Baker (2013) wurden die beiden Phänomene Expansion/Kompression und Zu-/Aufdecken mit Tiefenordnungsaufgaben untersucht: Wurden die Phänomene widersprüchlich dargestellt, antworteten die Teilnehmer gemäß der Expansion/Kompression bei geringer relativer Tiefe und gemäß Zu-/Aufdecken bei großen relativen Tiefenunterschieden (Yoonessi und Baker, 2013, S. 7–8): Denn wurde Expansion/Kompression isoliert, konnten Teilnehmer geringe relative Tiefen gut unterscheiden, mit abnehmender Akkuratesse zu größeren relativen Tiefenunterschieden (Yoonessi und Baker, 2013, S. 8–9). Ohne Bewegung des Betrachters bei konsistenter Wiedergabe beider Phänomene (bewegte Stimuli, statischer Betrachter) verlief die Akkuratesse gegenläufig dazu (Yoonessi und Baker, 2013, S. 6–7) – hier kann Expansion/Kompression keine eindeutigen Tiefeninformationen liefern und der Einfluss von Zu-/Aufdecken wird deutlich: Bei größeren relativen Tiefenun-

terschieden bietet Zu-/Aufdecken mit dem Kontext der Expansion/Kompression eindeutige Tiefenwahrnehmung. Wird Zu-/Aufdecken jedoch isoliert (also nur Verdeckung und keine relativen Geschwindigkeiten der Texturen gezeigt), konnten die Teilnehmer keine Tiefe wahrnehmen. Zu-/Aufdecken ist ohne den Kontext der Expansion/Kompression unzureichend und somit kein alleinstehender Tiefenhinweis (Yoonessi und Baker, 2013, S. 9).

Die Einblicke in die Wirkungszusammenhänge und -bereiche der visuellen Phänomene (Schereffekt, Expansion/Kompression und Zu-/Aufdecken) innerhalb der geometrischen Projektion helfen zur Beantwortung der Forschungsfragen dabei, die Sachzusammenhänge im Experiment benennen zu können. Insbesondere im Kontext der untersuchten Translationsverstärkungen g_t (engl. translation gain) sind sie zielführend, da die Tiefenrelationen in den Experimenten mitunter über sogenannte Synchronisationsverstärkungen (engl. syncing gain) umgesetzt wurden, die grundlegend vergleichbar funktionieren: Die Geschwindigkeiten der virtuellen Veränderungen im Bild (Geschwindigkeiten der Punktwolken bzw. Verdeckungen) wurden gemäß verschiedener Skalierungen der realen Kopftranslation T_r angepasst, um den Eindruck relativer Tiefe zu erzeugen (Yoonessi und Baker, 2011, S. 3, Yoonessi und Baker, 2013, S. 3).

2.2.2 Motion/Pursuit Ratio

Das bisher vollständigste Modell zur Tiefenwahrnehmung aus Motion Parallax wurde von Nawrot und Stroyan (2009) durch das „Motion/Pursuit Ratio“ (dt. Bewegungs-/ Verfolgungsverhältnis, ab hier: M/PR) beschrieben. Das M/PR ermöglicht im Gegensatz zu den Untersuchungen der retinalen Phänomene eine geometrische Erklärung für die Tiefenwahrnehmung aus der Kombination von retinalen und extra-retinalen Informationen: Das HVS kombiniert die Bewegung der Bildinformation auf der Retina (engl. „retinal image *motion*“) mit der zum Beibehalten einer Fixation benötigten Drehung der Augen (engl. „*pursuit* eye movement“) (Nawrot und Joyce, 2006 S. 4710, Nawrot und Stroyan, 2009, S. 1969).

Das M/PR ist dann die entsprechende mathematische Formel, die andeutet, wie das HVS die Tiefeninformationen aus der Bildinformation auf der Retina und der Augendrehung geometrisch schlussfolgern könnte (Nawrot und Stroyan, 2009, S. 1969–1970). Die Formel basiert auf den in Abb. 2.2 gezeigten Größen.

Führt der Betrachter eine laterale Kopfbewegung aus und behält dabei den Fixpunkt F (in der Distanz f) durch die Verdrehung der Augen α entgegen der Kopfbewegung in der Mitte der Retina, so ändert sich die Position des retinalen Abbilds des Distraktors D (in der Distanz d) um den Winkel θ. Diese distanzabhängigen Winkel umfassen geometrisch alle im vorangegangenen Abschnitt 2.2.1 beschriebenen Phänomene in einem Winkelzusammenhang. Wird statt des Verfolgungswinkels α die Verfolgungsgeschwindigkeit $d\alpha/dt$ angegeben, ergibt sich

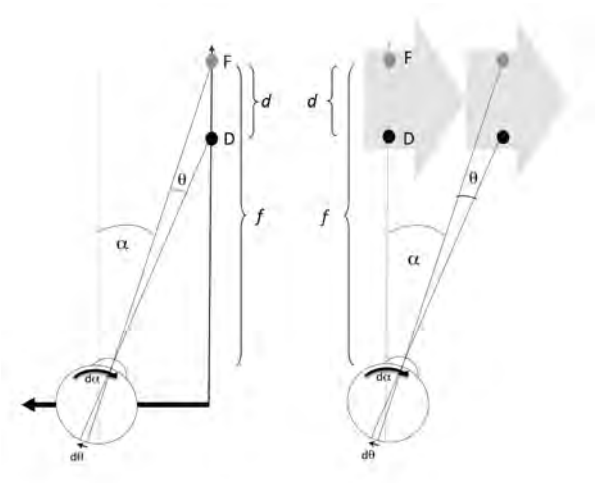


Abbildung 2.2: Geometrie des Motion/Pursuit Ratio – Nawrot et al., 2014, S. 2 Fig. 1

für den Distraktor die sogenannte Retinalgeschwindigkeit $d\theta/dt$ (häufig nur $d\alpha$ und $d\theta$ als Angabe in kleinen Zeitschritten). Aus diesen Kenngrößen lässt sich dann geometrisch auf die relative Tiefe von Fixationspunkt F und Distraktor D nach dem M/PR schließen:

$$\frac{d}{f} = \frac{d\theta}{d\alpha} \cdot \frac{1}{1 - \frac{d\theta}{d\alpha}} \tag{2.1}$$

Für den Fall $d\theta \ll d\alpha$ gilt näherungsweise:

$$\frac{d}{f} \approx \frac{d\theta}{d\alpha} \tag{2.2}$$

Das M/P Ratio beschreibt den geometrischen Zusammenhang akkurat, die menschliche Wahrnehmung ist jedoch fehlerbehaftet: Tiefenverhältnisse aus Motion Parallax werden in der Realität verkürzt wahrgenommen – Die Objekte erscheinen näher am Fixationspunkt als das Verhältnis der Retinalgeschwindigkeit $d\theta/dt$ zur Verfolgungsgeschwindigkeit $d\alpha/dt$ geometrisch schlussfolgern ließen (Durgin et al., 1995, S. 685 Fig. 3, Nawrot et al., 2014, S. 10). Objekte hinter dem Fixationspunkt werden also zu nah, Objekte vor dem Fixationspunkt zu weit entfernt wahrgenommen. Empirisch bestimmt wurde das M/PR von Nawrot et al. (2014) als (S. 11):

$$\frac{d}{f} = \frac{d\theta^{0,416}}{d\alpha^{0,192}} \cdot 0,0313 \tag{2.3}$$

Das Verhältnis der Winkelgeschwindigkeiten wurde somit nicht-linear gestaucht gegenüber dem Verhältnis der Abstände, entsprechend der verkürzten Wahrnehmung. Das M/PR schafft hinsichtlich der Forschungsfragen eine wichtige Diskussionsgrundlage zur Wahrnehmung der Motion Parallax: Die Translationsverstärkung g_t nimmt durch die Veränderung der

Betrachtungsposition direkt Einfluss auf die Winkelgeschwindigkeiten $d\theta/dt$ und $d\alpha/dt$. Die Skalierung der Szene im Rahmen verändert darüber hinaus die Retinalgeschwindigkeit $d\theta/dt$: Eine doppelte Vergrößerung sorgt nicht nur dafür, dass ein Szenenobjekt die doppelten virtuellen Abmessungen erhält, es resultiert auch eine Verdopplung der Größe jeder Bewegung im Rahmen und somit auch näherungsweise auf der Retina[1].

Ergänzend untersuchten Holmin und Nawrot (2015) die Schwellwerte dafür, wie viel Verfolgungsgeschwindigkeit $d\alpha/dt$ nötig ist, um aus Motion Parallax wirksam auf Tiefenstaffelung schließen zu können: Innerhalb eines Wertebereichs von $5\,^\circ/s \leq d\alpha/dt \leq 20\,^\circ/s$ war dies der Fall (Holmin und Nawrot, 2015, S. 45). Es gibt somit eine Unter- wie Obergrenze für die Verfolgungsgeschwindigkeit zur effektiven Tiefenwahrnehmung aus Motion Parallax. In einer Veröffentlichung von Hartle und Wilcox (2021) zur Untersuchung der Wechselwirkung zwischen Motion Parallax und Disparität wurde so z. B. eine Verfolgungsgeschwindigkeit von $d\alpha/dt \approx 9\,^\circ/s$ eingesetzt (S. 53). Ergebnisse von Kellnhofer et al. (2016) implizierten, dass mindestens eine Retinalgeschwindigkeit $d\theta/dt > 3\ arcmin/s$ benötigt wird, um Tiefeninformationen aus Motion Parallax zu gewinnen. Die beschriebenen Wertebereiche sind für das zu entwerfende Experiment dieser Arbeit anzustreben.

2.2.3 Motion Parallax im Zusammenwirken mit binokularer Disparität

Ziel von Kellnhofer et al. (2016) war es, das M/PR zu erweitern: Während Nawrot et al., 2014 das Ziel hatte, die Tiefenwahrnehmung aus isolierter Motion Parallax empirisch zu modellieren (S. 3), untersuchten Kellnhofer et al. (2016) nun das Zusammenwirken von Motion Parallax mit binokularer Disparität. Allerdings wurde keine Bewegung des Betrachters berücksichtigt, alleine die Stimuli waren dynamisch (fremdinduzierte Motion Parallax). Da das untersuchte Videoformat „gerahmtes, stereskopisches 3DoF+ Video“ binokulare Disparität mit Motion Parallax vereint, sind die Ergebnisse dennoch relevant. Auch Hartle und Wilcox (2021) untersuchten die Wechselwirkungen und Zuverlässigkeit von binokularer Disparität und Motion Parallax – hier war die Motion Parallax selbstinduziert durch Bewegung auf einer Kinnablage (S. 52–54). Kongsilp und Dailey (2018) führten ihre Experimente unter gänzlich freier Betrachterbewegung aus, wobei die beiden Tiefenhinweise einzeln und im Verbund bei Tiefenordnungsaufgaben gezeigt wurden (S. 209–211).

Alle drei kamen zu dem Ergebnis, dass Motion Parallax ein unzuverlässigerer Tiefenhinweis ist als binokulare Disparität, da Tiefe – analog zur Veröffentlichung von Nawrot et al., 2014 (S. 10) – verkürzt wahrgenommen wird (Kellnhofer et al., 2016, S. 5 Fig. 4a, Hartle und Wilcox, 2021, S. 56 Fig. 5, Kongsilp und Dailey, 2018, S. 217–218 Fig. 7 Fig. 9). Hartle und Wilcox (2021)

[1] Bei gleicher Translationsverstärkung g_t und gleicher Translationsbewegung, Details siehe Abschnitt 2.4

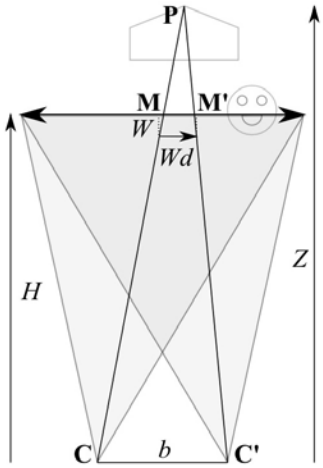
konnten darüber hinaus eine Veto-Funktion des zuverlässigeren Tiefenhinweises binokulare Disparität gegenüber der unzuverlässigeren Motion Parallax feststellen (S. 61). Auch French und DeAngelis, 2022 zeigten, dass die Akkuratesse der Tiefenwahrnehmung, wenn beide Tiefenhinweise präsent sind, deutlich verbessert ausfällt, verglichen mit Motion Parallax in rein monokularen Stimuli (S. 9). Das HVS verlässt sich somit eher auf binokulare Disparität, wenn beide präsent sind. Daher ist es für das in dieser Arbeit untersuchte Videoformat insbesondere wichtig, Stimuli mit akkurater binokularer Disparität zu zeigen (Die technischen Details zur Umsetzung eines diesen Anforderungen entsprechenden, virtuellen Stereo-Kamerasystems sind im nachfolgenden Abschnitt 2.3 und 4.2 aufgeführt).

Ferner ist für die Untersuchungen in dieser Arbeit relevant, dass in den Experimenten von Hartle und Wilcox (2021) Stimuli aus der RU im Vergleich mit Stimuli aus der VU (über ein VR-Headset) hinsichtlich akkurater Tiefenwahrnehmung verglichen wurden. Insofern die Stimuli Tiefenhinweise oberhalb des Erkennungsschwellwerts zeigten, konnten keine signifikanten Unterschiede zwischen RU und VU festgestellt werden (S. 60). Der immanente Konflikt stereoskopischer Bildwiedergabe zwischen Vergenz (Augenrotation zur Fixation auf Punkt in der Tiefe; muss nicht der Bildebene entsprechen) und Akkommodation (Adaption der Linsen auf die Fokaldistanz; entspricht stets der Bildebene) – genannt VAC (engl. vergence accomodation conflict, Hoffman et al., 2008, S. 2–3) – scheint keinen Einfluss auf die Kombination von Motion Parallax und binokulare Disparität zu haben.

2.3 Stereoskopie

Die Anforderungen einer akkuraten Darstellung stereoskopischer Tiefe bei der Untersuchung von Motion Parallax aufgrund der Veto-Funktion der binokularen Disparität wurde im letzten Abschnitt aufgezeigt. Gerahmtes, stereoskopisches 3DoF+ Video wird technisch dadurch realisiert, dass für jedes Auge ein virtueller Blickwinkel errechnet und ausgegeben wird (siehe Abschnitt 1.3.2) – im Prinzip liegt also das Äquivalent eines virtuellen Stereo-Kamerasystems vor. Wie die geometrischen Zusammenhänge der Stereoskopie umzusetzen sind, damit der Betrachter die gewünschte Wahrnehmung der Szenentiefe erhält, wird in diesem Abschnitt erläutert:

Monoskopische Bildinhalte (2D) werden mit einer einzelnen Kamera ausgezeichnet, liefern entsprechend nur einen Blickwinkel auf die Szenenumgebung. Das Licht der Szene wird durch das Objektiv gebündelt, fällt auf den Kamerasensor und wird als zweidimensionale Repräsentation der auftreffenden Lichtintensitäten ausgewertet und gespeichert. Dadurch wird ein Abbild des 3D-Raums innerhalb des sogenannten „View Frustum“ des Kamerasystems – ein Kegelstumpf, beschrieben durch die Sichtwinkel (horizontal α_h, vertikal α_v) – auf eine 2D-Ebene projiziert (mehr zu Projektion in Abschnitt 2.4). 2D-Bildwiedergabe ermöglicht



Symbol	Camera	Display
C, **C**′	camera optical center	eye optical center
P	physical scene point	perceived 3-D point
M, **M**′	image points of **P**	screen points
b	camera interocular	eye interocular
H	convergence distance	screen distance
W	width of convergence plane	screen size
Z	real depth	perceived depth
d	disparity (as a fraction of W)	

Abbildung 2.3: Aufnahme- und Betrachtungsgeometrie mit Notationen nach Devernay und Beardsley, 2010, S.20 Fig. 8

nur monokulare („einäugige“), visuelle Tiefenhinweise, wie Okklusion (Verdeckung), relative Größe, Texturgradienten, Schattierung, Höhe im Sichtfeld (Devernay und Beardsley, 2010 S. 7–8, Brenner und Smeets, 2018, S. 4–11, J. Liu et al., 2023, S. 473).

In stereoskopischen Bildaufnahmesystemen (3D) kommen zwei Kameras zu Einsatz, die horizontal in einem gewählten Abstand (genannt: Interokulardistanz b) zueinander angeordnet sind. Stereoskopische Bildwiedergabesysteme, die jedem Auge des Betrachters eines der Kamerabilder zuspielen können, ermöglichen durch den relativen Perspektivversatz die Wahrnehmung von Tiefe. Dabei werden der binokularen („zweiäugigen“) Tiefenhinweise Disparität (visueller Tiefenhinweis), sowie Vergenz (okulomotorischer Tiefenhinweis) genutzt. (Devernay und Beardsley, 2010 S. 8–9, ITU, 2023, S. 94, J. Liu et al., 2023, S. 472).

Geometrische Zusammenhänge Dafür, wie die Tiefe in stereoskopischer Bildwiedergabe wahrgenommen wird, sind die geometrischen Zusammenhänge des Stereo-Kamerasystems, sowie die Betrachtungsbedingungen maßgeblich. In dieser Arbeit werden die Notationen von Devernay und Beardsley, 2010 für die zugehörigen Parameter verwendet (s. Abb. 2.3). Betrachtungsseitige Variablen sind hierbei wie ihr aufnahmeseitiges Pendant benannt und durch einen Apostroph gekennzeichnet. Z' beschreibt die wahrgenommene Distanz jedes beliebigen Szenenelements, welches sich in der Distanz Z zu den Kameras befindet. Im Kontext der Betrachtung sind für die Tiefenwahrnehmung drei Parameter verantwortlich: die Interokulardistanz (Augenabstand) des Betrachters b' , die Breite des gezeigten Bilds (Display/-Leinwand) W', sowie die Distanz zwischen Betrachter und Bildebene H'. Aufnahmeseitig entsprechen diese der Interokulardistanz der Kameras b, der Breite der Konvergenzebene W und der Distanz der Kameras zur Konvergenzebene (genannt Konvergenzdistanz) H.

Die Konvergenzebene beschriebt die Ebene, in der das rechte und linke Kamerabild keine Disparität $d = 0$ zeigen. Das bedeutet, dass sie keinen relativen Versatz aufweisen, also dem rechten, wie linken Auge des Betrachters ein identisches Bild liefern. Diese Ebene wird vom Betrachter als in der Tiefe der Bildebene (Display/Leinwand) wahrgenommen ($Z' = H'$). Die Konvergenzdistanz H der beiden Kameras liegt im Schnittpunkt ihrer optischen Achsen. Im Falle zweier paralleler Kameras, gibt es keinen Schnittpunkt und die Konvergenzdistanz H liegt im Unendlichen. Alle Szenenelemente werden vor der Bildebene wahrgenommen. Nähern sich die optischen Achsen der Kameras an, wird von einer „toed-in“ (Devernay und Beardsley, 2010, S. 19) Kamera-Konfiguration gesprochen. Wie ein Paar Augen richten sie sich jeweils auf das fixierte Szenenobjekt. In der zugehörigen Distanz H des Objekts liegt dann die Konvergenzebene. Szenenelemente hinter der Konvergenzebene zeigen dann positive Disparität $d > 0$, die Augen des Betrachters müssen weniger konvergieren als beim Blick auf Elemente in der Bildebene. Umgekehrt ist die Disparität negativ $d < 0$ für Elemente vor der Konvergenzebene, die optischen Achsen treffen sich in einem steileren Winkel.

Homothetische Stereo-Kamera Der geometrisch einfachste Fall liegt bei einer homothetischen Kamerakonfiguration vor. Es gilt:

$$\frac{W'}{W} = \frac{H'}{H} = \frac{b'}{b} \tag{2.4}$$

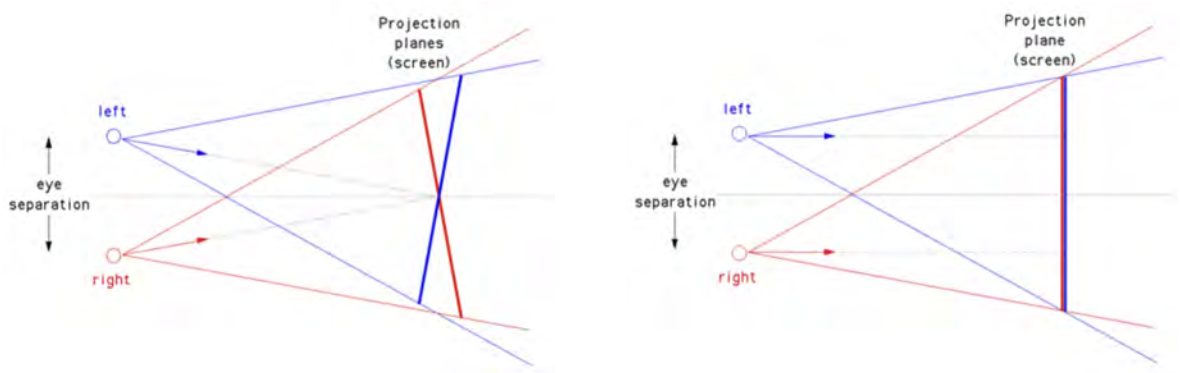
(Devernay und Beardsley, 2010, S. 24 Formel (9))

Zwischen der Aufnahme- und Betrachtungsgeometrie besteht ein linearer Zusammenhang (Skalierung). Die Szenentiefe wird unverzerrt (1:1) dargestellt. (Devernay und Beardsley, 2010, S.23–24)

Kanonische Stereo-Kamera Wenn sich nun das Verhältnis der Konvergenzdistanz zur Betrachtungsdistanz $\frac{H'}{H}$ ändert und die anderen Parameter gleich bleiben (s. Formel (2.5); z. B. durch eine kürzere Betrachtungsdistanz H'), wird die Konfiguration kanonisch (engl. „canonical“ Devernay und Beardsley, 2010, S. 22) genannt:

$$\frac{W'}{W} = \frac{b'}{b} \tag{2.5}$$

Die wahrgenommene Szenentiefe ist nicht mehr unverzerrt (1:1), bleibt jedoch linear skaliert um den Faktor $\frac{H'}{H}$. Dadurch wirkt die Szene in der Tiefe gestreckt ($H' > H$), beziehungsweise gestaucht ($H' < H$). Für den Sonderfall $H' = H$ bleibt der Zusammenhang unverzerrt. Während $\frac{H'}{H}$ die Skalierung der Tiefe (in Schreibweise als Kamera-Koordinaten: z-Achse) festlegt, gibt das Verhältnis der Bildbreite zu Breite der Konvergenzebene $\frac{W'}{W}$ die Skalierung in der xy-Ebene vor.



(a) tied-in (b) schiefachsig

Abbildung 2.4: Stereo-Kamera: Vergleich toed-in (a) und schiefachsiger (b) Konfiguration – Bourke, 1999

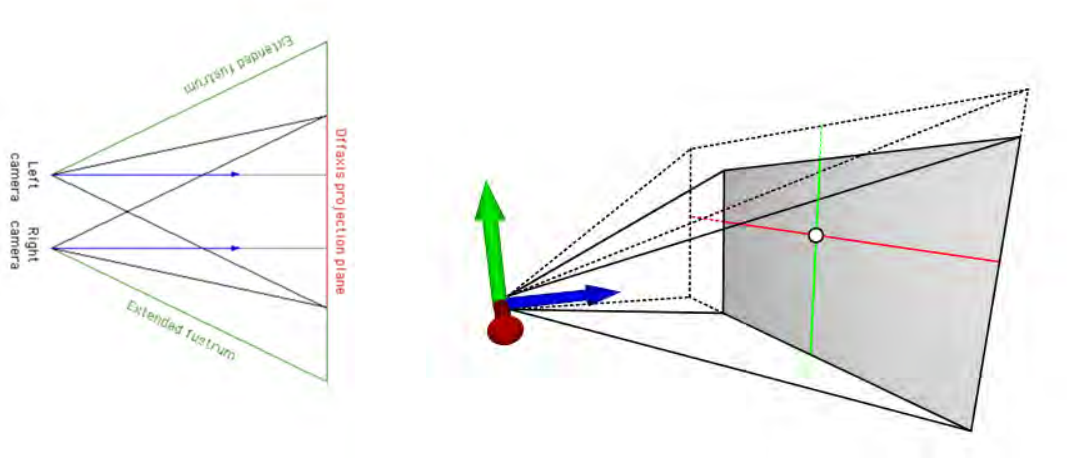
Schiefachsige Projektion Da in der Regel Tiefe jenseits der Bildebene ($Z' > H'$) gewünscht ist, damit der Rahmen des Bilds dem Eindruck eines Fensters gleicht, wird in szenischen Anwendungen meist mit „toed-in“ Kameras gearbeitet. Durch das Kreuzen der optischen Achsen sind die Bildebenen dann allerdings zueinander verdreht (s. Abb. 2.4(a)). Es entsteht vertikale Disparität (Ukai und Howarth, 2008, S. 108), die als unangenehm empfunden und daher üblicherweise in der Nachbearbeitung korrigiert wird.

In der Herstellung computergenerierter Stereoskopie wird daher mit asymmeterischen View Frusta gearbeitet (Zellmann und Amstutz, 2023, S. 1–2). Die optischen Achsen der Kameras werden dafür parallel ausgerichtet. Ihr horizontaler Sichtwinkel α_h zentriert sich nun allerdings um diejenigen optischen Achsen, die für die gewünschte Konvergenzdistanz in der „toed-in“ Konfiguration benötigt würden (s. Abb. 2.4(b)). Dadurch wird die Projektion schiefachsig.

Dies entspricht bei der Verwendung von realen Kameras einer Verschiebung der Bildsensoren (linker Sensor nach links, rechter Sensor nach rechts). Ohne Verschiebung der Bildsensoren lässt sich der Effekt konstruieren, indem der horizontale Sichtwinkel α_h der Kameras so erweitert (herausgezoomt), hochskaliert und zugeschnitten wird (Zellmann und Amstutz, 2023, S. 5, s. Abb. 2.5). Dadurch wird Auflösung eingebüßt. Der benötigte Sichtwinkel α_h^* entspricht dem verdoppelten Winkel zwischen optischer Achse und dem jeweils entfernteren Rand der gewünschten Konvergenzebene α_{off}:

$$\alpha_h^* = 2 \cdot \alpha_{off} \tag{2.6}$$

Für die Untersuchung der Forschungsfragen ist eine schiefachsige, kanonische Projektion mit $H' = H$ für alle Skalierungen (Brennweiten f) zu wählen, um die Tiefe der Szene



(a) Aufsicht (b) Perspektivisch

Abbildung 2.5: Stereo-Kamera: Konstruktion einer schiefachsigen Projektion durch Vergrößerung des horizontalen Sichtwinkels α_h: (a) Aufsicht – Bourke, 1999 (b) Perspektivische Darstellung aus einer Kamera – Zellmann und Amstutz, 2023, S. 6 Fig. 2

durchweg verzerrungsfrei darzustellen. Nur so sind Stimuli zwischen den Brennweiten f vergleichbar hinsichtlich der Wahrnehmung natürlicher Translationsverstärkungen g_t und der resultierenden Motion Parallax (mehr dazu in Abschnitt 4.2).

2.4 Brennweite, Translationsverstärkung und Projektion

Wie in Abschnitt 2.1 diskutiert, steckt als Informationsquelle hinter visuellen Tiefenhinweisen projektive Geometrie. Die Zusammenhänge der Projektion im untersuchten Videoformat werden nachfolgend anhand einer Veranschaulichung erörtert:

Aus der effektiven Brennweite f eines Objektivs lassen sich abhängig von der verwendeten Sensorgröße ($W_{sensor} \times H_{sensor}$) die Sichtwinkel ($\alpha_h$, α_v) bestimmen:

$$\alpha_h = 2 \cdot \arctan(\frac{W_{sensor}}{2f}), \quad \alpha_v = 2 \cdot \arctan(\frac{H_{sensor}}{2f}) \tag{2.7}$$

Die Position der Projektion eines beliebigen Punkts P_1 im Szenenraum ergibt sich daraus, wie groß die Winkel zur optischen Achse der Kamera ausfallen (horizontal und vertikal). Sind diese kleiner als die Sichtwinkel α_h und α_v, ist der Punkt P_1 im Bild sichtbar. Seine Projektion $P_{1,proj}$ liegt dann auf dem Sensor in der relativen Breite und Höhe, die P_1 in der vom View Frustum begrenzten Ebene in seiner Distanz (Objektebene) aufweist – solange er nicht von

anderen Szenenelementen verdeckt wird. Die Breite und Höhe der Objektebene W_{d_P} und H_{d_P} in der Distanz Z_P betragen:

$$W_{d_P} = 2 \cdot Z_P \cdot \tan(\frac{\alpha_h}{2}), \quad H_{d_P} = 2 \cdot Z_P \cdot \tan(\frac{\alpha_v}{2}) \tag{2.8}$$

Die Größe der Projektion eines Objekts im Bild entspricht seinen Abmessungen in der Objektebene in Relation zu den Abmessungen der Ebene selbst. Die nachfolgenden Rechenbeispiele zeigen die Zusammenhänge in einer simulierten Aufsicht (es wird somit nur a_h und H_{d_P} betrachtet), da die laterale Translation Gegenstand der Arbeit ist:

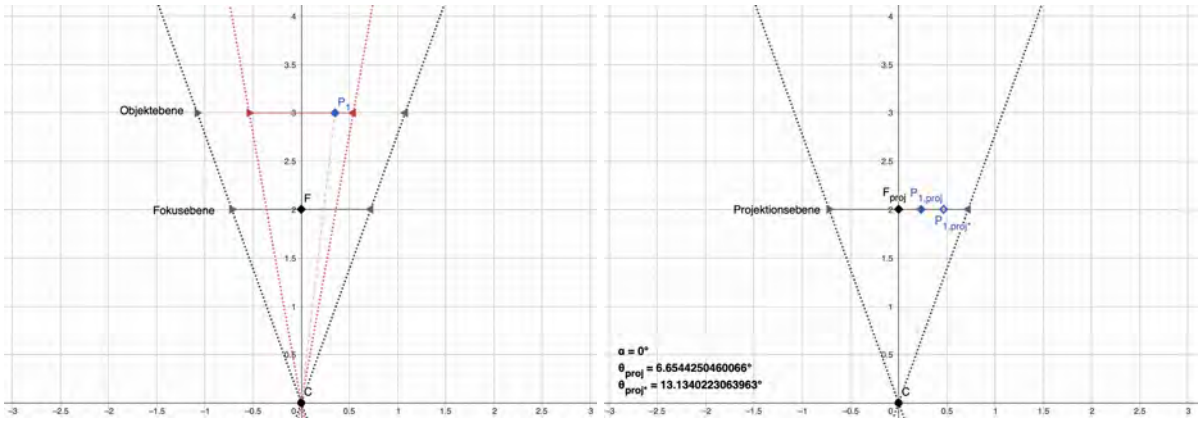
Beschreibt P_1 den äußersten Punkt eines Objekts in horizontaler Richtung, wird P_{1_x} als dessen Abstand zur optischen Achse der Kamera festgelegt (x-Koordinate in Abb. 2.6). S_x ergibt dann den Anteil, den dieser Abstand in der Objektebene H_{d_P} ausgehend vom Bildmittelpunkt einnimmt ($S_x = 0$ ist die Bildmitte, $S_x = 1$ ist der Bildrand, daher ist H_{d_P} halbiert).

$$S_x(Z_P) = \frac{P_{1_x}}{0,5 \cdot H_{d_P}} = \frac{P_{1_x}}{Z_P \cdot \tan(\frac{\alpha_h}{2})} \tag{2.9}$$

Einfluss der Brennweite auf die Retinalgeschwindigkeit Wird von einer homothetischen Konfiguration ausgehend (s. Abschnitt 2.4) die Brennweite f_1 verdoppelt, halbieren sich die Abmessungen der Objektebene und das Objekt wird in Relation doppelt vergrößert im virtuellen Rahmen dargestellt. Betrachtet man die Projektionen als Koordinaten bezüglich der Bildmitte als Koordinatenursprung, bewirkt eine Verdopplung der Brennweite $f_2 = 2f_1$ somit eine Skalierung der Koordinaten jedes Punkts um den Faktor zwei; das Bild wird von der Bildmitte ausgehend „aufgezogen“. Abbildung 2.6 veranschaulicht die Zusammenhänge: Die homothetische Konfiguration ist in Grau angedeutet, die Projektion $P_{1,proj}$ des Punkts P_1 liegt exakt in der realitätsgetreuen Sichtachse. Die Projektion $P_{1,proj*}$ aus der doppelten Brennweite f_2 (rotes View Frustum) liegt auf Grund der halbierten Objektebene doppelt so weit von der optischen Achse entfernt.

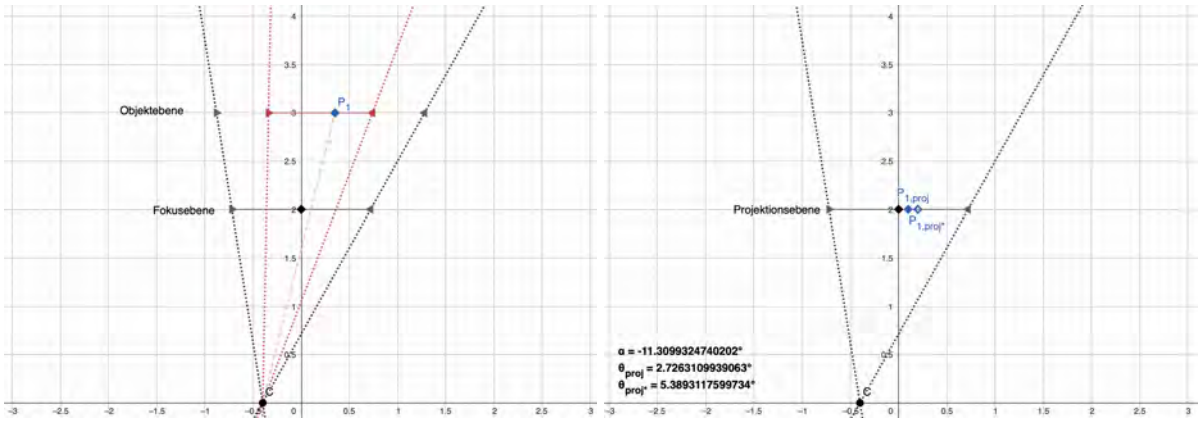
Bewegt sich der Betrachter nun lateral (Beispiel 40 *cm* – zur Veranschaulichung sehr groß gewählt – entlang der x-Achse) und wird seine reale Translation T_r exakt in die virtuelle Kameratranslation $T_{v_{cam}}$ übersetzt ($g_t = 1$; s. Formel (2.10)) verändert sich die Perspektive bei gleichbleibender Beschränkung des Sichtfelds durch den Rahmen (asymmetrisches View Frustum, s. Abschnitt 2.3). Durch die zweifache Vergrößerung von der Bildmitte ausgehend bei f_2 verdoppelt sich auch die Strecke, die die Projektion $P_{1,proj*}$ auf der Projektionsebene verglichen mit $P_{1,proj}$ zurückgelegt hat (s. Abb. 2.7). Somit fällt auch die Änderung des Retinalwinkels θ der Abbildung (bei Fixation von F_{proj} in der Fokusebene) näherungsweise doppelt (genau: Faktor[2] 1,97) so schnell aus (siehe Werte für θ und θ_{proj*} in Abb. 2.6 (b)

[2]Die gleiche Auslenkung in entgegengesetzte Richtung gibt Faktor 2,06.



(a) Aufzeichnung (b) Betrachtung

Abbildung 2.6: Veranschaulichung der Projektion von Aufzeichnung (a) zur Betrachtung (b) in Ursprungsposition mit doppelter Brennweite f_2



(a) Aufzeichnung (b) Betrachtung

Abbildung 2.7: Veranschaulichung der Projektion von Aufzeichnung (a) zur Betrachtung (b) mit lateraler Translation und doppelter Brennweite f_2

und 2.7(b), was in einer Verdopplung der Retinalgeschwindigkeit $d\theta/dt$ für den Betrachter resultiert.

Es entsteht eine leichte Abweichung durch die perspektivische Verzerrung aus dem lateral versetzten Blickpunkt auf die planare Projektionsebene (Objekte an der näher gelegenen Rahmenkante erscheinen minimal größer als an der weiter entfernten). Neben dem exakten Faktor für das spezifische, abgebildete Rechenbeispiel wird in den Fußnoten der Faktor für die laterale Gegenbewegung als Vergleichswert angeführt.[3]. Alle Berechnungen fußen auf

[3]Der Effekt lässt sich mit dem beiliegenden GeoGebra Online-Tool, aus dem die Grafiken entstammen, veranschaulichen (s. Anhang GeoGebra Online-Tool, S. 70

trigonometrischen und vektoriellen Zusammenhängen, die hier aus Gründen der Übersicht nicht detailliert aufgeführt werden.)

Eine Verdopplung der Brennweite $f_2 = 2f_1$ hat somit näherungsweise eine Verdopplung der Retinalgeschwindigkeit $d\theta/dt$ bei lateraler Translation zufolge (eine Halbierung $f_{0.5} = 0,5f_1$ eine Halbierung von $d\theta/dt$).

Einfluss der Translationsverstärkung auf die Retinalgeschwindigkeit Wird statt der doppelten Brennweite f_2 wiederum f_1 verwendet, allerdings mit einer Translationsverstärkung von $g_t = 2$, findet keine Skalierung der Projektion statt, jedoch ein aufzeichnungsseitiger Perspektivversatz ($C_* \neq C$) zur Sichtlinie der homothetischen Konfiguration: Die betrachtungsseitige virtuelle Translation des Betrachters T_v folgt stets akkurat seiner realen Translation T_r – der Blickwinkel auf den Rahmen bleibt somit unbeeinflusst (s. Betrachtungspunkt C' in Abb. 2.8(b)). Der Inhalt des Rahmens zeigt sich verändert, da die Translationsverstärkung g_t ausschließlich die Auslenkung der Kamera $T_{v_{cam}}$ betrifft (s. Abb. 2.8(a)). Untersuchungsgegenstand ist schließlich die als natürlich wahrgenommene Translationsverstärkung *innerhalb des Rahmens*. Die Translationsverstärkung g_t im Zusammenhang mit gerahmtem, stereoskopischem 3DoF+ Video meint daher:

$$g_t = \frac{T_{v_{cam}}}{T_r} \tag{2.10}$$

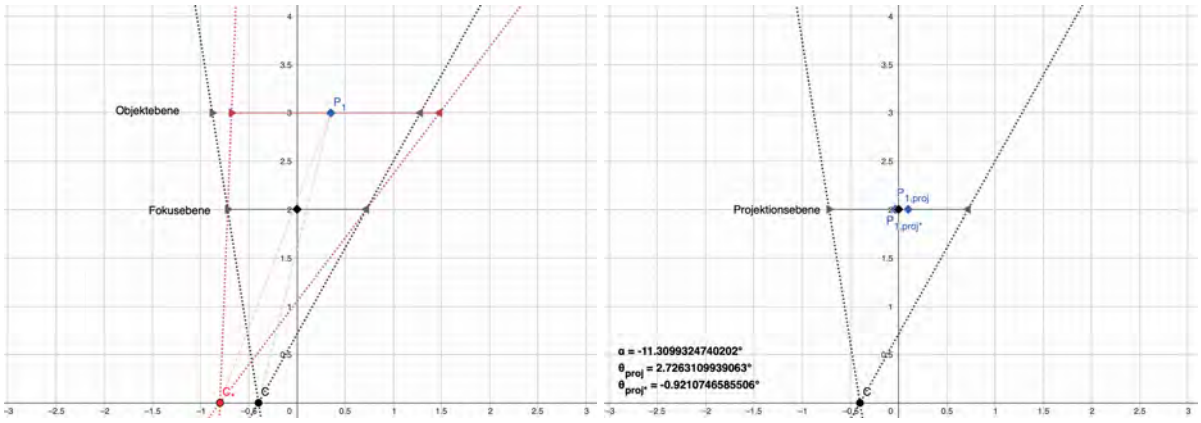
Vergleicht man die Veränderung der Retinalwinkel θ zwischen Abb. 2.6(a) und Abb. 2.8(a), fällt auf, dass sich diese näherungsweise gleich zu der Retinalwinkelveränderung bei doppelter Brennweite f_1 und $g_t = 1$ verhält. Eine Verdopplung der Translationsverstärkung g_t bei gleichbleibender Brennweite f bewirkt somit auch näherungsweise eine Verdopplung (genau: Faktor[4] 1, 93) der Retinalgeschwindigkeit $d\theta/dt$ (eine Halbierung von g_t eine Halbierung der Retinalgeschwindigkeit $d\theta/dt$).

Zusammenwirken von Brennweite und Translationsverstärkung Werden beide Anpassungen (f_2 mit $g_t = 2$) kombiniert, lässt sich eine Multiplikation ihrer Wirkung auf die Retinalgeschwindigkeit $d\theta/dt$ beobachten – so fällt diese im Rechenbeispiel näherungsweise um den Faktor vier (genau: Faktor[5] 3, 81) schneller aus (s. Abb.2.8(b) und 2.9(b)).

Mithilfe dieser Erkenntnisse lässt sich der Einfluss der Brennweite f und Translationsverstärkung g_t auf die Wahrnehmung der Tiefe aus Motion Parallax mittels des M/PR (s. Abschnitt 2.2.2) abschätzen. Hypothesen dafür, welche Translationsverstärkung g_t für eine Brennweite $f_X = X \cdot f_1$ als natürlich wahrgenommen wird, lassen sich nur bedingt aufstellen, da die Wahrnehmung der Perspektivveränderung nicht isoliert von der Tiefenwahrnehmung aus Motion

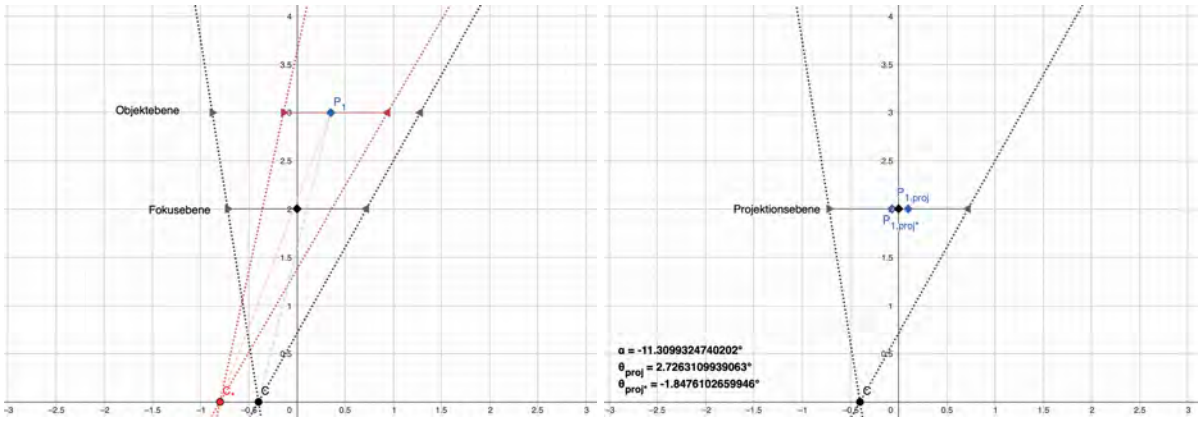
[4]Die gleiche Auslenkung in entgegengesetzte Richtung gibt Faktor 2, 03.
[5]Die gleiche Auslenkung in entgegengesetzte Richtung gibt Faktor 4, 02.

Parallax abhängt – die Szene wird schließlich sowohl hinsichtlich der Retinalgeschwindigkeit $d\theta/dt$, als auch in der Skalierung verändert.



(a) Aufzeichnung (b) Betrachtung

Abbildung 2.8: Veranschaulichung der Projektion von Aufzeichnung (a) zur Betrachtung (b) mit lateraler, verstärkter Translation und homothetischer Brennweite f_1



(a) Aufzeichnung (b) Betrachtung

Abbildung 2.9: Veranschaulichung der Projektion von Aufzeichnung (a) zur Betrachtung (b) in Ursprungsposition mit doppelter Brennweite f_2

3 Forschungsstand: Translationsverstärkung & Motion Parallax

In Abschnitt 2.2.1 wurde bereits gezeigt, dass zur Simulation visueller Phänomene und Bestimmung deren Einflusses auf die Wahrnehmung von Motion Parallax, von einer Synchronisationsverstärkung bezüglich der realen Kopftranslation T_r Gebrauch gemacht wurde. Experimente zum Einfluss veränderter Translationsverstärkungen g_t aus der RU in die VU sind aktuell hingegen insbesondere im Forschungsfeld „VR Locomotion“ (dt. VR-Fortbewegung)/„Redirected Walking“ (dt. umgeleitetes Laufen) geläufig, meist mit dem Ziel räumliche Begrenzungen mittels unbemerkter Veränderungen im virtuellen Sichtfeld des Anwenders zu umgehen:

3.1 Psychophysische Experimente zur Schwellwerterkennung

Dafür kommt in der Regel das psychophysische Verfahren von konstanten Stimuli mit einer erzwungenen Wahl zwischen zwei Alternativen zum Einsatz (engl. two-alternative forced-choice = **2AFC**; J. Zhang et al., 2018 S. 1675, Langbehn et al., 2018 S. 5, Kim et al., 2021, S. 656, Kim et al., 2022, S. 383). Die Teilnehmer werden einem Stimulus einzeln mit einer der untersuchten Translationsverstärkungen g_t ausgesetzt, wonach ihnen eine Frage gestellt wird, auf die sie mit einer von zwei Möglichkeiten antworten müssen. Im Kontext der Translationsverstärkung g_t lautet die Frage meist sinngemäß:

„*War die in der virtuellen Umgebung gezeigte Bewegung größer oder kleiner als deine reale Bewegung?*“

(Steinicke et al., 2010, S.21, J. Zhang et al., 2018, S. 1675,Kim et al., 2021, S. 656, Kim et al., 2022, S. 383)

Dieser Vorgang wird dann für alle zu untersuchenden Stimuli wiederholt. Es muss mit „*größer*“ oder „*kleiner*“ geantwortet werden: So ergibt sich für den Stimulus, bei dem die Teilnehmer

zu gleichen Teilen (Wahrscheinlichkeit „*größer*“ zu antworten $P_{larger} = 0,5$) „*größer*“ oder „*kleiner*“ geantwortet haben, die Translationsverstärkung g_t, für die die virtuelle Bewegung der reellen entsprechend empfunden wurde. Dieser Punkt wird PSE (engl. point of subjective equality) genannt. Die Punkte, für die $P_{larger} = 0,25$, respektive $P_{larger} = 0,75$ gelten, werden als Erkennungsschwellwerte $DT_{25\%}$ und $DT_{75\%}$ bezeichnet: Der Bereich zwischen den Erkennungsschwellwerten wird in der Literatur als derjenige beschrieben, in dem die untersuchte Translationsanpassung g_t von Teilnehmern nicht zuverlässig als unterschiedlich zum *PSE* identifiziert werden kann. Um die Werte zu erhalten, wird eine psychometrische Funktion aus den Antworten angenähert, die die Wahrscheinlichkeit „*größer*“ zu antworten P_{larger} in Abhängigkeit der Translationsverstärkung g_t aufträgt. (Steinicke et al., 2010, S. 21, Langbehn et al., 2018, S. 8, Serrano et al., 2020a, S. 6, Y. Zhang et al., 2022, S. 831, Teng et al., 2023, S. 404)

Diese 2AFC-Methode mit konstanten Stimuli ermöglicht somit die Erhebung von Messwerten, die zur Beantwortung der Forschungsfragen dieser Arbeit beitragen: Die Bestimmung von *PSE*-Werten gibt Aufschluss über Forschungsfragen FF1 und FF3, während die Erkennungsschwellwerte $DT_{25\%}$ und $DT_{75\%}$ für die Forschungsfragen FF2 und FF4 zielführend sind. Eine entsprechende Methode wurde daher im Experiment dieser Bachelorarbeit eingesetzt (s. nachfolgender Abschnitt 4).

3.2 Abstandsabhängige Wahrnehmung der Translation beim Vorwärtsgehen

Die Begrenzung des Videoformats auf 3DoF+ bedeutet prinzipiell, dass dem Anwender sitzend (selten auch stehend) ein diesem Umstand entsprechender Bewegungsfreiraum ohne örtliche Fortbewegung (wie Gehen) zur Verfügung steht. Dennoch bieten einige Forschungsergebnisse aus dem Bereich „VR-Locomotion“ interessante und hinsichtlich der Forschungsfragen relevante Einblicke in die Wahrnehmung von Translationsverstärkungen g_t. Diese sind hier kurz zusammengefasst:

Relevant erscheint zum einen die Erkenntnis, dass die als natürlich wahrgenommene Translationsverstärkung g_t bei gehender Fortbewegung geradeaus abhängig von den Abmessungen der VU ist: Je kleiner die VU, desto geringer fielen die *PSE*-Werte aus, zu größeren VU nahmen sie zu (Kim et al., 2021, S. 658, Y. Zhang et al., 2022 S. 831, Kim et al., 2022, S. 384). Dies könnte darauf zurückzuführen sein, dass die Motion Parallax bei geringen Distanzen und entsprechend dazu in Relation großen Distanzunterschieden besonders deutlich visuelle Hinweise auf die eigene Fortbewegung gibt (Cutting und Vishton, 1995, S. 80) – bei größeren Geschwindigkeiten wird der Tiefenhinweis auch bei weiteren Distanzen wirksam (Cutting und Vishton, 1995, S. 84), was die erhöhten *PSE*-Werte erklären könnte.

Die Untersuchung von Y. Zhang et al. (2022) zeigte darüber hinaus, dass in möblierten VU die Wahrnehmung der eigenen Translation zwischen verschiedenen Raumgrößen konsistenter und insgesamt akkurater wurde gegenüber leerer VU (S. 831). Dies deckt sich mit den Beobachtungen, dass komplexere Umgebungen zu akkuraterer Tiefenwahrnehmung führen (Murgia und Sharkey, 2009, S. 72, Vienne et al., 2020, S. 29103).

Auch die prozentualen Abweichungen der Erkennungsschwellwerte $DT_{25\%}$ und $DT_{75\%}$ von ihren gemessenen *PSE* sind interessant als Referenz für die eigene Messung: Sie liegen für die Experimente zur Translationsverstärkung g_t des Vorwärtsgehens in Wertebereichen von rund $\pm 7,5 \ \ldots 25$ % und fallen tendenziell bei kleineren Abmessungen der VU geringer aus: Hierfür sind in den Arbeiten von Y. Zhang et al., 2022 (S. 831, Szenenbreite 3 $m \Longrightarrow \pm 9,2$ %, 100 $m \Longrightarrow \pm 13,1$ %) und Kim et al., 2022 (S. 384, bei möblierten Szenen mit Breite $5,2\ m \Longrightarrow \pm 12,7$ %, 17 $m \Longrightarrow \pm 21,8$ %) direkte Vergleichswerte vorhanden. Weitere Einzelmessungen von J. Zhang et al., 2018 (S. 1676) mit ca. 2 m Szenenbreite und $\pm 7,6$ %, sowie Steinicke et al., 2010 (S. 23) mit mindestens 10 m Szenenbreite und ± 18 % stützen diese These. Die erhöhte Empfindlichkeit bezüglich der Translationsverstärkung g_t bei kleineren VU könnten ebenfalls in Teilen durch die stärkere Wirkung der Motion Parallax bei geringeren Distanzen begründet sein (s. oben).

3.3 Forschungsstand zur Wahrnehmung lateraler Kopftranslation

Die vorliegende Arbeit hat allerdings laterale Kopftranslationen im Sitzen zum Gegenstand. Vertikale Auslenkung (Strecken/Ducken) wird im Sitzen kaum genutzt (Thatte und Girod, 2018, S. 3) und hinsichtlich der Tiefenwahrnehmung ist laterale Auslenkung weitaus zuverlässiger als Vor-/Zurücklehnen (S. Liu et al., 2023, S. 6–7). Zu lateraler Translationsverstärkung g_t ist die Studienlage verglichen mit dem Vorwärtsgehen noch deutlich weniger fortgeschritten: Serrano et al., 2020a waren die ersten, die Erkennungsschwellwerte für Anpassungen der Übersetzung realer, lateraler Kopftranslationen T_r in virtuelle T_v untersuchten (S. 3) – seither sind nur Teng et al. (2023) gefolgt.

Experiment von Teng et al. (2023) Letztere untersuchte eine abstrakte, texturierte Winkelform als Stimulus, die ober- und unterhalb von schwarzen Balken umrahmt vor einer Landschaftstextur in unendlicher Tiefe gezeigt wurde (Teng et al., 2023, S. 399–400 Fig. 2a). Die Teilnehmer standen fest im Abstand von $1,5\ m$ zu den Stimuli. Sie werteten Translationsverstärkungen g_t im Wertebereich $0,5 \ \leq g_t \leq 2,0$ mittels der oben beschriebenen Methode zur Schwellwerterkennung aus. Vor-/Zurücklehnen wurde technisch dadurch unterbunden, dass bei einer Auslenkung von $\pm 3,5\ cm$ in dieser Achse das Bild gedimmt wurde – so konnte

nur laterale Auslenkung effektiv genutzt werden (Teng et al., 2023, S. 400). Das Ergebnis zeigte für die Kombination binokularer Disparität mit Motion Parallax einen *PSE* von $g_t = 1,14$ mit den Erkennungsschwellwerten $DT_{25\%}$: $g_t = 1,02$ und $DT_{75\%}$: $g_t = 1,26$ (entspricht $\pm 10,5$ % Abweichung in Verhältnis zum *PSE*; Teng et al., 2023, S. 404).

Experiment von Serrano et al. (2020a) Serrano et al. (2020a) hingegen führten ein solches psychophysisches Experiment unter völlig freien Betrachtungsbedingungen (inklusive Rotation) in einer realistischen Szenenumgebung durch. In einem möblierten Raum der Größe $12 \times 12\ m$ erschienen an drei von vier Orten um die zentral positionierten Teilnehmer graue Ebenen. Einen Meter hinter einer dieser Ebenen verbarg sich eine Münze, die es durch laterale Kopftranslation zu finden galt. Die Suchaufgabe sollte die laterale Translation auf natürliche Weise motivieren – die Teilnehmer führten das Experiment sitzend auf einem Drehhocker aus. Die drei Ebenen erschienen je Durchlauf in einer der untersuchten Distanzen $\{1\ m;\ 1,75\ m;\ 2,5\ m;\ 3,25\ m;\ 4\ m\}$ und dienten als Fixationspunkte bei der Translation (s. MP/R in Abschnitt 2.2.2). Nach Erfüllen der Suchaufgabe wurde die Frage nach der Wahrnehmung der Translationsübersetzung gestellt, wodurch sich für jede Distanz eine psychometrische Funktion ergab. Die *PSE* erwiesen sich als distanzabhängig, je größer die Distanz, desto höher der *PSE* (vergleichbar mit der These zur Wahrnehmung der Translationsverstärkung beim Vorwärtsgehen, s. Abschnitt 3.2): Die *PSE*-Werte ergaben $g_t = \{0,87;\ 0,92;\ 1,1;\ 1,29;\ 1,4\}$ (Reihenfolge wie oben) mit weitaus größeren, prozentualen Abweichungen von $DT_{25\%}$ und $DT_{75\%}$ der Werte $\{51,7\ \%;\ 39,1\ \%;\ 39,1\ \%;\ 42,6\ \%;\ 40,0\ \%\}$ (Reihenfolge wie oben; Serrano et al., 2020a, S. 6 Fig. 4).

Die untersuchten relativen Distanzen wurden in zugehörige Retinalgeschwindigkeiten $d\theta/dt$ umgerechnet, um für komplexe 360° 3DoF+ Videos modellieren zu können, wie – je nach Fixation der Betrachter und der dortigen Distanzrelationen – die Translationsgeschwindigkeit g_t dynamisch gestaucht werden könnte (zur Artefaktminderung, s. Abschnitt 1.3.2; Serrano et al., 2020a, S. 9–11). Die sogenannten „translation gain maps“ der Videos (dt. Translationsverstärkungs-Kartierungen) sind daher relevant, da die *PSE*-Messwerte eindeutig zeigten, dass die natürliche Translationsübersetzung für VU abstandsabhängig ist (Serrano et al., 2020a, S. 6): Dies deckt sich mit den zuvor aufgeführten VR-Locomotion-Studien. Ob sich diese Beobachtung auch auf gerahmte und durch Projektion veränderte Videos übertragen lässt, bleibt zu untersuchen (mehr dazu in der Diskussion: Abschnitt 6).

Entwurf des eigenen Experiments Die Messwerte der Studien (Serrano et al., 2020a, Teng et al., 2023) können für das im folgenden Abschnitt entworfene, psychophysische Experiment als Vergleichswerte in der Auswertung dienen. Um die Forschungsfragen FF1 und FF2 beantworten zu können, ist ein Vergleich der Wahrnehmung der lateralen Translationsverstärkung g_t bei Betrachtung einer *gerahmten* Szene mit Betrachtung einer *ungerahmten*

Szene erforderlich – Nur so lässt sich der Einfluss der Rahmung auf *PSE*- und *DT*-Werte isoliert betrachten. Da das psychophysische Experiment in dieser Arbeit auch Einblicke in den Einfluss der Brennweite f im Kontext der Rahmung geben soll (Forschungsfragen FF3 und FF4), wurde nur *ein* fester Betrachtungsabstand für die gerahmten Stimuli gewählt[1]. Dem zeitlich begrenzten Umfang der Bachelorarbeit geschuldet, wäre eine eigene Vergleichsstudie zwischen gerahmten und ungerahmten Szenen zusätzlich zum Brennweiteneinfluss nicht umsetzbar. Somit wird im nachfolgenden Kapitel eine Methode entworfen, die auf der von Serrano et al. (2020a) verwendeten aufbaut. Die Ergebnisse von Serrano et al. (2020a) bieten dann die benötigten Vergleichswerte. Entscheidend dafür ist dann, einen möglichst vergleichbaren Versuchsaufbau mit der Suchaufgabe innerhalb von Rahmen umzusetzen. So muss zum Beispiel die Tiefenwiedergabe innerhalb des Rahmens stets einer realistischen VU entsprechen und möglichst unverzerrt sein (mehr dazu in Abschnitt 4.2).

Eine Abwandlung der Methode von Teng et al. (2023) kam nicht infrage, da diese bereits durch die schwarzen Balken gewissermaßen gerahmt ist und keine realistische Szenenumgebung zeigt (somit weniger Aussagekraft für das Zielbild eines Videoformats hat).

[1] Erläuterung der Wahl von H' im nachfolgenden Kapitel 4.3.

4 Psychophysisches Experiment

4.1 Teilnehmer

19 Teilnehmer (10 männlich und 9 weiblich, Alter 24–55, $M = 30,1$) haben an dem Experiment teilgenommen. Alle gaben an, weniger als einmal im Monat stereoskopische Bildinhalte zu konsumieren. Drei Teilnehmer haben noch nie ein VR-Headset getragen, 14 bis zu 5-mal und zwei Teilnehmer häufiger als 5-mal. 13 Teilnehmer wählten die Selbsteinschätzung, seltener als einmal im Monat Video-Spiele auf PC oder Konsole zu spielen, zwei wählten ein- bis viermal im Monat und vier wählten häufiger als einmal in der Woche.

Die meisten der Teilnehmer waren Angestellte eines großen deutschen Medienunternehmens. Die Selbsteinschätzungen der Vorerfahrung in den Bereichen 3D-Grafik und Videoproduktion (vorgenommen in vier Stufen verrechnet als {0; 0,33; 0,67; 1}, s. Anhang Fragebogen, S. 73), Anhang Datenschutz-Formular, S. 72) spiegeln dies wider: $M = 0,31$ für den Bereich 3D-Grafik und $M = 0,72$ für Videoproduktion. Die Teilnehmer gaben an, dass sie mindestens einmal pro Monat Filme und Fernsehsendungen über den Computer (16), den Fernseher (12), das Smartphone (11) oder das Kino (8) anschauen. Kein Teilnehmer wählte hier die Option VR-Headset (0).

Test des Sehvermögens Vor der Durchführung des Experiments wurden alle Teilnehmer – gemäß der Empfehlung ITU-R BT.500-15 zur subjektiven Beurteilung von stereoskopischen 3DTV Systemen (ITU, 2023 S. 102)– auf ihr Sehvermögen (Scharfsehen, Farbsehen und stereoskopisches Sehen) überprüft: Zwei Teilnehmer trugen eine Brille (einmal zur Korrektur von Kurzsichtigkeit, einmal aufgrund von Astigmatismus), ein Teilnehmer trug Kontaktlinsen (zur Korrektur von Kurzsichtigkeit), alle übrigen nahmen ohne Sehhilfe teil. Zwei Teilnehmer gaben eine Rot-Grün-Schwäche an. Technisch wurden alle Sehtests in einem separaten, abgedunkelten Raum mit einem Binoptometer® 4P durchgeführt. Abbildung 4.1 zeigt den Testaufbau.

Bei allen Teilnehmer wurde in der binokularen Fernsicht ein Visus von 1,0 oder besser gemessen, bis auf eine Ausnahme mit einem Visus von 0,8. In der binokularen Nahsicht (Distanz von 0,4 m) ergab ebenfalls nur eine Messung einen Wert unter 1,0 mit einem Visus von 0,4.

Abbildung 4.1: Aufbau des Sehtests

Die Fernsicht ist für VR-Anwendungen die wichtigere Metrik, da angenommen wird, dass das verwendete VR-Headset (Meta Quest 3) das virtuelle Bild in einem Abstand der Größenordnung 1–2 m darstellt; Hierzu fehlen belastbare Daten, jedoch hatte sich der ehemalige CTO von Oculus VR via Twitter zu den beiden Vorgängermodellen geäußert: „The design focal distance for the Quest/Quest 2 optics is 1.3 meters." (John Carmack [@ID_AA_Carmack], 2021). Die Farb- und Stereosehtest wurden daher mit einer Distanzeinstellung von 1,5 m vorgenommen.

Die Farbsehtests (5 Ishihara Farbtafeln) bestätigten die zwei Angaben von Rot-Grün-Schwäche, die übrigen Teilnehmer konnten alle Farbtafeln sicher identifizieren. Bei den Stereosehtests erreichten alle Teilnehmer ein Auflösungsvermögen von 15", mit Ausnahme von zwei Teilnehmern mit 100"[1] und einem Teilnehmer, der kein stereoskopisches Sehvermögen (>600") aufwies[2].

Alle Teilnehmer wurden trotz Farb- oder Stereosehschwäche im Experiment berücksichtigt, da der untersuchte Tiefenhinweis Motion Parallax monokular (s. Abschnitt 2.1.1) und nicht farbabhängig ist.

[1] Eine naheliegende Erklärung liefert die Messung der monokularen Fernsicht (hier: bei beiden Teilnehmern je mindestens ein Auge mit Visus ≥ 0,63). Bei der Distanz von 1,5 m könnte somit das Scharfsehen nicht ausreichen, um Informationen aus dem Tiefenhinweis Disparität zu erschließen.

[2] Auch hier: Ein Auge zeigte in Fern- wie Nahsicht nur einen Visus von 0,2.

4.2 Stimuli

Die Stimuli bilden in einem Abstand von $H' = 2\ m$ vom Betrachter[3] platzierte virtuelle Rahmen. Die Rahmen messen in der Breite $W' = 1,44\ m$ mit einem Seitenverhältnis von 16:9 (Horizontaler Bildwinkel von $\alpha_{FOV_H} \approx 39,6°$ entsprechend der homothetischen Brennweite $f_1 = 50\ mm$, mehr dazu in Abschnitt 2.4). In den Rahmen wird das von einem Stereo-Kamerasystem bereitgestellte Videobild einer statischen Szene gezeigt. Die Translationsverstärkung g_t betrifft hier – wie in Abschnitt 2.4 beschrieben – die Translation des Stereo-Kamerasystems $T_{v_{cam}}$ (s. Formel (2.10)). Der Rahmen selbst bleibt in seiner Position im VU fixiert, die sich bis auf den Rahmen gemäß einer 1:1 Übersetzung realer Kopftranslation T_r in virtuelle T_v verhält.

Stereo-Kamerasystem Das Stereo-Kamerasystem besteht aus zwei Kameras, die orthogonal zur optischen Achse um die Hälfte der gewählten Interokulardistanz b in entgegengesetzte Richtungen verschoben sind (die rechte Kamera also um $b/2$ nach rechts, die linke Kamera um $b/2$ nach links). Jedes Auge des Teilnehmers erhält dabei das zugehörige Bild. Die Projektionsmatrix der Kameras ist entsprechend dem Entwurf aus Abschnitt 2.4 schiefachsig: Die optischen Achsen der beiden Kameras (R, L) bleiben stets parallel zueinander. Ihr View Frustum ist so verschoben, dass die Konvergenzebene beider Kameras stets in der Ebene des zugehörigen Rahmens liegt (somit asymmetrisch zur optischen Achse). Szeneninhalte in der Konvergenzdistanz H zum Stereo-Kamerasystem zeigen keine Disparität und werden vom Teilnehmer als in der Ebene des Rahmens wahrgenommen (Details und Abbildungen s. Abschnitt 2.3).

Dementsprechend ist die Konvergenzdistanz H so gewählt worden, dass die dem Betrachtungsabstand $H' = 2\ m$ entspricht. Darüber hinaus ist das Stereo-Kamerasystem als kanonische Konfiguration (s. Abschnitt 2.3) realisiert. Aus diesen Rahmenbedingungen ergibt sich eine lineare Tiefenrelation aus dem Tiefenhinweis Disparität, bei dem der gesamte Tiefenbereich der Szene unverzerrt im Rahmen wiedergegeben und Divergenz im Unendlichen verhindert wird (Devernay und Beardsley, 2010, S. 22). Die Sicherstellung einer unverzerrten Tiefenrelation hinsichtlich der binokularen Disparität ist entscheidend, aufgrund der in Abschnitt 2.2.3 beschriebenen Veto-Funktion der binokularen Disparität über Motion Parallax (Hartle und Wilcox, 2021, S. 61). Nur so sind die untersuchten Variablen, die Einfluss auf die Szenendarstellung nehmen, valide miteinander vergleichbar.

Szene im Rahmen Die vom Stereo-Kamerasystem erfasste, statische Szene (s. Abb. 4.2(b) zeigt im Vordergrund ein graues Quadrat mit den Abmessungen $0,2 \times 0,2\ m$, welches in der

[3] Details s. Abschnitt 4.3

Konvergenzebene, somit in der Rahmenebene $Z = H$ liegt. Dieses dient als Fixationspunkt für die Teilnehmer während der lateralen Translationsbewegungen. Alle weiteren Szenenelemente befinden sich hinter der Rahmenebene ($Z > H$), um „window violations“[4] zu verhindern (Devernay und Beardsley, 2010, S. 27, Gardner, 2011, S. 3, Miyashita et al., 2022, S. 5). An den Kanten des grauen Quadrats ist der Tiefenübergang abrupt, wodurch dort der Unterschied der Retinalgeschwindigkeiten $d\theta/dt$, sowie die Scherbewegung, Expansion/Kompression und Zu-/Aufdecken bei Translationsbewegungen besonders deutlich werden (siehe Abschnitt 2.2.1).

Da komplexe Hintergründe für eine akkuratere, linearere Tiefenwahrnehmung sorgen (Murgia und Sharkey, 2009, S. 72, Vienne et al., 2020, S. 29103) und bei Motion Parallax relativen Retinalgeschwindigkeiten $d\theta/dt$ verschiedener Objekte in Abhängigkeit ihrer Tiefe eine wichtige Rolle zugeschrieben wird (s. Abschnitt 2.2.2), zeigt die Szene in der Tiefe gestaffelte Szenenelemente. Diese sind so entworfen, dass möglichst viele monokulare, visuelle Tiefenhinweise (s. Abschnitt 2.3) akkurat wiedergegeben werden. Der Stimulus ist in Abbildung 4.2(b) dargestellt.

Variablen Die unabhängigen Variablen des Experiments bilden Brennweite f und Translationsverstärkung g_t. Die abhängigen Variablen sind die zu erhebenen Kennwerte der psychometrischen Funktionen der Translationswahrnehmung (s. Abschnitt 3.1). Es werden als Translationsverstärkungen analog zum Experiment von Serrano et al., 2020a (S. 4) die Werte $g_t = \{0,4; 0,6; 0,8; 1,0; 1,25; 1,67; 2,5\}$ gewählt, um die Ergebnisse im Vergleich einordnen zu können. Ausschließlich laterale Kopftranslationen T_r werden in virtuelle Translationen des Stereo-Kamerasystems $T_{v_{cam}}$ übersetzt.[5]

Hinsichtlich der Brennweite f ist die Referenz mit $f_1 = 50\,mm$ auf einem Vollformat-Bildsensor ($W_{sensor} = 36\,mm$, Seitenverhältnis: 16:9) gewählt worden. Die vier Ecken des resultierenden View Frustum passen dann genau durch die vier Ecken des Rahmens, was gepaart mit der unverzerrten Tiefendarstellung (s. oben) dem Eindruck eines Fensters in der RU entsprechen soll (analog zu Experimenten wie Miyashita et al., 2022 S. 4). Wichtig zu ergänzen ist, dass die Brennweite f nur für die Ausgangslage des Betrachters gilt, da diese die Sichtwinkel (α_h, α_v) und somit die Ausgangsposition des virtuellen Rahmens und seine Abmessungen (z. B. Breite der Konvergenzebene W) im Szenenraum festlegt. Bei $T_v \neq 0$ verändern sich die Sichtwinkel stets so, dass die vier Ecken des virtuellen Rahmens stets als Begrenzung des View Frustum fungieren (s. Abschnitt 2.4). Neben f_0 sind die Brennweiten $f_{0,5} = 25\,mm$ und $f_2 = 100\,mm$ untersucht worden. Die Breite der Konvergenzebene W fällt durch den

[4] Als „window violation“ (dt. Fensterverletzung) wird die Überschneidung eines Szenenelements, das vor der Konvergenzebene liegt, mit den Rändern des View Frustum bezeichnet. In der Realität würde es aus dem Bildrahmen hinausragen, durch die Begrenzung der Kamerablickwinkel wird es abgeschnitten.

[5] Begründung s. Abschnitt 4.4

(a) Brennweite $f_{0.5}$ (b) Brennweite f_1 (c) Brennweite f_2

Abbildung 4.2: Gerahmte Stimuli in Ursprungsposition für (a) $f_{0.5}$ (b) f_1 und (c) f_2.

veränderten Sichtwinkel α_h dann doppelt ($f_{0,5}$), beziehungsweise halb (f_2) so groß aus. Da das Bild des Stereo-Kamerasystems den Rahmen stets ausfüllt und dessen Abmessungen und Abstand konstant bleiben, entstehen Vergrößerungen des Szenenbilds im Rahmen um die Faktoren 0,5 ($f_{0,5}$) und 2 (f_2).

Die Größe des grauen Quadrats in der Rahmenebene ist somit zwischen den Brennweiten f nicht identisch. Dieser Umstand ist notwendig, da sonst das Zu-/Aufdeckverhalten des Szenenhintergrunds mit der Brennweite f verändert würde. Dadurch wäre das Verhältnis der Retinalgeschwindigkeiten $d\theta/dt$ zu den Verfolgungsgeschwindigkeiten $d\alpha/dt$ an den Tiefenübergängen zwischen Quadrat und Hintergrund je Brennweite f unterschiedlich und die Vergleichbarkeit der brennweitenabhängigen Ergebnisse untereinander könnte beeinträchtigt werden.

4.3 Freie Betrachtungsumgebung

Da die Daten von Serrano et al., 2020a als Vergleichswerte für laterale Translationswahrnehmung in VU ohne Rahmung dienen sollten, wurde das Experiment ebenfalls unter freien Betrachtungsbedingungen (engl. „free viewing conditions“ Serrano et al., 2020a, S. 4) in einem in der VU simulierten Raum mit den Abmessungen $12 \times 12\ m$ durchgeführt. Die Teilnehmer saßen in der Mitte des virtuellen Raums auf einem Drehhocker (s. Abb. 4.3).

Suchaufgabe zur Motivation lateraler Translation Zur Motivation der lateralen Translationsbewegung ist bei Serrano et al., 2020a eine Suchaufgabe für die Teilnehmer verwendet worden: Um die Teilnehmer herum sind zufällig in drei der vier Himmelsrichtungen graue Ebenen im gleichen Abstand erschienen, wobei sich einen Meter hinter einer der Ebenen eine Münze verborgen hat. Die Aufgabe war, sich mittels des Drehhockers zu den Ebenen auszurichten und die Münze dann durch laterale Kopftranslation zu finden (Serrano et al., 2020a, S. 4). Die Ebenen haben als Fixationspunkte während der Kopftranslation gedient –

Abbildung 4.3: Versuchsaufbau

Sie wurden für diese Arbeit als graue Quadrate abgewandelt, um in dem begrenzten Rahmen der Stimuli auch ober- und unterhalb Scherbewegungen und Retinalgeschwindigkeiten $d\theta/dt$ als Tiefenhinweise zu ermöglichen (s. Abschnitt 4.2). Mit den gerahmten Stimuli wurde eine entsprechende Suchaufgabe erstellt: Drei der gerahmten Stimuli erschienen zufällig in einer von zwei Konstellationen[6] um die Teilnehmer (s. Abb. 4.5). Alle Stimuli waren je Durchlauf identisch hinsichtlich der Variablen f und g_t konfiguriert und zeigten die gleiche Szene. In einem der drei Rahmen befand sich hinter dem grauen Quadrate ein blaues[7] Rechteck, welches es zu finden galt (s. Abb. 4.4). Dies war bei jeder Brennweite im Fall $g_t = 1$ bei einer Translationsauslenkung von $T_{v_{cam}} = \pm 20\ cm$ für das erste Auge sichtbar:

Die Brennweite f verändert die Breite der Konvergenzebene W bei konstanter Konvergenzdistanz H (s. Abschnitt 2.4). In einer kanonischen Konfiguration (s. Abschnitt 2.3) erfordert dies die Anpassung der Kamera-Interokulardistanz b. Durch den veränderten Versatz der Kameras vom Mittelpunkt des Systems wäre das blaue Rechteck bei $f_{0.5}$ mit etwas weniger und bei f_2 mit etwas mehr virtueller Auslenkung hinter dem grauen Quadrat zu erkennen. Die Breite des blauen Rechtecks wurde daher skaliert, sodass die oben beschriebene Vorgabe für alle f gilt.

[6] Da die Rahmen deutlich größer ausfallen als die Ebenen bei Serrano et al., 2020a, musste mit mehr Abstand zwischen den Stimuli gearbeitet werden, um sicherzustellen, dass sich stets nur ein Stimulus im Sichtfeld der Teilnehmer befindet.

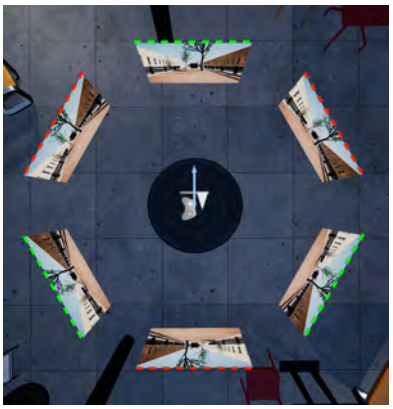
[7] In der Farbe Blau, um auch bei Rot-Grün-Schwäche gut erkennbar zu sein.

(a) Ursprungsposition

(b) Translation nach links

(c) Stimulus-Erfassung

Abbildung 4.4: Gerahmter Stimulus für f_1 (a) in Ursprungsposition (b) mit lateraler Translation $T_{v_{cam}}$ nach links, sodass blaues Zielrechteck sichtbar wird (nur in einem der drei Rahmen). (c) zeigt die Stimulus-Erfassung mit dem virtuellem Stereo-Kamerasystem.



(a) Aufsicht Rahmenkonstellationen

(b) Teilnehmersicht (simuliert)

Abbildung 4.5: Ansichten der Betrachtungsumgebung: (a) Aufsicht zeigt Rahmenanordnung 1 in Grün und Anordnung 2 in Rot), (b) zeigt Sicht der Teilnehmer auf einen gerahmten Stimulus mit f_1

Wahl der Umgebung und des Betrachtungsabstands Der Raum ist mit Einrichtungsgegenständen und Säulen möbliert, Boden wie Decke sind texturiert, um – wie auch beim Stimulus selbst (s. Abschnitt 4.2) – möglichst viele monokulare, visuelle Tiefenhinweise, sowie Anhaltspunkte für akkurate Stereopsis zu schaffen (s. Abb. 4.5(b); Murgia und Sharkey, 2009, S. 72, Devernay und Beardsley, 2010, S. 7–8, Vienne et al., 2020, S. 29103). So kann die Umgebung um die Stimuli als Referenz für die reale Kopftranslation verwendet werden. Diese Referenzfunktion hat außerdem die Wahl des festen Betrachtungsabstands von $H' = 2\,m$ motiviert:

Als Betrachtungsabstand wurde $H' = 2\,m$ gewählt, da die Ergebnisse von Serrano et al., 2020a andeuten, dass in VU bei diesem Abstand die laterale Translationsverstärkung $g_t = 1$ den PSE bildet. Das bedeutet, dass dort die Translation in der VU identisch zu der ausgeführten Translation in der RU wahrgenommen wird. Da die Rahmen selbst Verdeckungen der dahinterliegenden Wände bewirken (also sich bei lateralen Kopftranslationen ähnlich der grauen Fixations-Ebenen bei Serrano et al., 2020a (S. 4) verhalten), ließe sich so der Einfluss der Umgebung, sowie der durch das VR-Headset veränderten Tiefenwahrnehmung theoretisch minimieren. So könnten die Ergebnisse potenziell auf Mixed Reality Anwendungen (z. B. mit halbdurchsichtigen Displays) übertragen werden. Der Sichtwinkel auf dem Rahmen entspricht überdies optimalen Betrachtungsbedingungen in einem Kino-Saal mit einem Abstand von $H' \approx 1,39 \cdot W'$ und $\alpha'_h = 39,6\,^\circ$ (Szita et al., 2024 S. 5, Gjestland, 2022 S. 67)

4.4 Apparatus

Die Teilnehmer wurden instruiert, ein Meta Quest 3 VR-Headset aufzusetzen, welches eine Auflösung 2.064 × 2.208 Pixeln pro Auge bietet bei einem Sichtfeld von $\alpha_{FOV_H} = 110^\circ$ horizontal und $\alpha_{FOV_V} = 96^\circ$ vertikal. Für das Positions- und Rotationstracking wurde das Tracking der Sensoren im HMD selbst genutzt. Der rechte Meta Quest Touch Plus-Controllers diente den Teilnehmern als Eingabegerät. Die VU wurde mit Unreal Engine 5.3 simuliert, ausgeführt auf einem HP OMEN Desktop Computer[8] unter Windows 11. Die Bildwiederholrate des HMD wurde auf die für das System empfohlenen 72 Hz bei der vollen Auflösung (4.128 × 2.208 Pixel insgesamt; entspricht 2.064 × 2.208 Pixel pro Auge) eingestellt. Unreal Engine 5.3 lieferte das Bild über die Oculus OpenXR-Laufzeitumgebung via Link-Kabel (Länge 5 m) an das Headset. Die Teilnehmer saßen auf einem Drehhocker, das Link-Kabel war an einer Zugentlastung über ihnen aufgehängt, sodass sie sich sitzend frei bewegen und drehen konnten. (s. Grafik 4.3)

Da Unreal Engine 5.3 im Gegensatz zu Unity keine Einstellmöglichkeit eines asymmetrischen View Frustum für virtuelle Kameras bereitstellt, wurde die entsprechende Funktionalität ei-

[8] Spezifikationen: AMD Ryzen 9 5900X 12-Core Prozessor, 32 GB RAM, NVIDIA GeForce RTX 3080 10 GB

genhändig im Unreal Engine Blueprint des Stereo-Kamerasystems realisiert. Dafür wurde die in Abschnitt 2.3 beschriebene Methode verwendet. Dem zeitlichen Rahmen der Bachelorarbeit geschuldet, wurde von der Implementierung der Translation in den Achsen „hoch-runter“ und „vor-zurück“ abgesehen, da diese nicht Untersuchungsgegenstand sind. Das virtuelle Stereo-Kamerasystem folgt somit ausschließlich dem lateralen Translationsanteil der realen Kopftranslation T_r auf einer Geraden parallel zur Konvergenzebene. Die Teilnehmer wurden für ihren visuellen Komfort angehalten, ihren Kopf nur lateral parallel zur Ebene des betrachteten Rahmens zu bewegen und ihn bei den Bewegungen aufrecht halten. Letzteres ist ferner empfehlenswert, da – wie bei herkömmlichem 3D – mögliche Kopfneigungen in der Roll- oder Gierachse (engl. roll/yaw axis) und der somit entstehende, leichte Versatz der Augen zu dieser Geraden programmseitig nicht berücksichtigt worden sind. Diese beiden Einschränkungen hinsichtlich des Betrachtungskomforts lassen sich in zukünftigen Experimenten ergänzen.

4.5 Verfahren

Um die Wertebereiche der Translationsverstärkung g_t, die als realitätsnah empfunden werden, für die drei Brennweiten f zu bestimmen, wurde das in der Literatur geläufige psychophysische Verfahren von konstanten Stimuli mit einer erzwungenen Wahl zwischen zwei Alternativen verwendet (2AFC, s. Abschnitt 3.1). Den Teilnehmern wurde nach Erfüllung der Suchaufgabe (s. Abschnitt 4.3) die Frage gestellt:

„War die in den virtuellen Fenstern gezeigte Bewegung schneller oder langsamer als deine reale Kopfbewegung?“

Sie mussten mit *„schneller“* oder *„langsamer“* antworten. Die Bezeichnung „Fenster“ wurde statt „Rahmen“ verwendet, um die bildliche Referenz aus der RU in der Frage zu benennen. Die Teilnehmer saßen während des Briefings auf einem Stuhl im Abstand von 2 m zur Fensterfront des Raums, sodass sie sich vor dem Experiment die Referenz der Geschwindigkeitsverhältnisse im Fenster einprägen konnten (s. Abb. 4.3). Die in ähnlichen Arbeiten für die 2AFC-Methode verwendete Frage, ob die virtuelle Bewegung *„größer“* oder *„kleiner“* als die reale Bewegung war (s. Abschnitt 3.1), wurde hier in *„schneller“* oder *„langsamer“* abgewandelt, da es sonst durch die Vergrößerung/Verkleinerung aufgrund der Brennweiten f zu Missverständnissen kommen könnte. Der Zusammenhang in Formel (4.1) zeigt, dass die Frage nach Auslenkung oder Geschwindigkeit im Kontext dieser Untersuchung synonym sind.

$$\frac{v_{v_{cam}}}{v_r} = \frac{T_{v_{cam}}/t}{T_r/t} = \frac{T_{v_{cam}}}{T_r} = g_t \tag{4.1}$$

Die Daten wurden in einem mehrfaktoriellen, experimentellen Design mit Messwiederholung (engl. full-factorial within subjects design) erhoben: Alle unabhängigen Variablen (f, g_t)

wurden je Testsubjekt variiert. Durch die drei Brennweiten ($f_{0,5} = 25\ mm$, $f_1 = 50\ mm$ und $f_2 = 100\ mm$), sowie sieben Translationsverstärkungen $g_t = \{0,4; 0,6; 0,8; 1,0; 1,25; 1,67; 2,5\}$ kamen insgesamt 21 verschiedene Stimuli zustande. Diese wurden in drei direkt aneinandergereihten Durchläufen in zufälliger Reihenfolge präsentiert ($3 \times 7 \times 3 = 63$ Versuche).

Ablauf Vor dem Experiment erteilten die Teilnehmer ihr informiertes Einverständnis zur Erfassung der Daten[9] und füllten einen Fragebogen zu ihrem Nutzungsverhalten und ihrer Vorerfahrungen hinsichtlich digitaler Bewegtbildmedien[10] aus. Anschließend wurde ihre Interokulardistanz b' (engl. interpupillary distance = IPD) gemessen und im VR-Headset, wie auch Unreal Engine eingestellt, um die kanonische Konfiguration des Stereo-Kamerasystems als (s. Abschnitt 4.2) für jeden Teilnehmer individuell sicherzustellen. Das Sehvermögen wurde gemessen (s. Abschnitt 4.1), worauf das Briefing der Teilnehmer in Form einer Präsentation über einen Beamer folgte.

Die Aufgabe der Teilnehmer bestand darin, sitzend durch laterale Kopftranslationen das blaue Rechteck hinter dem grauen Quadrat in einem der drei zufällig um sie erscheinenden Rahmen zu finden (s. Abschnitt 4.3). Sie wurden instruiert, sich mit dem Drehhocker zu den Rahmen auszurichten und, durch seitliche Verschiebung der oberen Körperhälfte, hinter die grauen Quadrate schauen. Während der Bewegung war darauf zu achten, wie sich die Geschwindigkeit innerhalb des Rahmens, verglichen mit der Geschwindigkeit, die sie in einem realen Fenster erwarten würden, verhält. Sie wurden im Vorfeld ermutigt, beim Blick durch die Rahmen in Bewegung zu bleiben (um oberhalb des Schwellwerts der wahrnehmbaren Motion Parallax zu sein; s. Abschnitt 2.2.2). Sollten sie das blaue Rechteck im Rahmen vor sich nicht gefunden haben, wurden sie angewiesen, sich in die aufrechte Ausgangsposition zurückzubegeben, neu ausrichten und in einem der weiteren Rahmen suchen. Ihnen wurde mitgeteilt, dass alle Rahmen technisch völlig identisch konfiguriert sind – Somit galten alle ihre Beobachtungen aus einem Rahmen auch für die übrigen beiden. Sobald sie das blaue Rechteck gefunden hatten, sollten sie die Trigger-Taste ihres rechten Controllers betätigen. Die Rahmen verschwanden und ihnen wurde mündlich die Frage gestellt: *„War die in den virtuellen Fenstern gezeigte Bewegung schneller oder langsamer als deine reale Kopfbewegung?“* (s. oben), worauf sie ebenfalls mündlich mit *„schneller“* oder *„langsamer“* antworten mussten. Es gab keine zeitliche Beschränkung – weder für das Erfüllen der Suchaufgabe, noch für die Antwort. Ihnen wurde mitgeteilt, dass sie das Experiment jederzeit abbrechen durften. Vor dem ersten Versuch war es ihnen erlaubt, Verständnisfragen zu stellen.

Nach dem Briefing wurden die Teilnehmer instruiert, sich das VR-Headset aufzusetzen und für sich einzustellen. Daraufhin startete die Simulation der VU (s. Abschnitt 4.3) ohne Stimuli (Rahmen), in der sich die Teilnehmer mit der Umgebung vertraut machen und umschauen

[9] Formular s. Anhang Datenschutz-Formular, S. 72
[10] Formular s. Anhang Fragebogen, S. 73

durften. Sobald sie bereit waren, wurden alle Versuche wie beschrieben nacheinander durchgeführt. Die mündlich mitgeteilten Antworten wurden in Unreal Engine eingegeben und mit den übrigen erfassten Daten je Versuch tabellarisch gespeichert.

Das gesamte Verfahren dauerte ca. 45 Minuten, wovon das Experiment 20 Minuten einnahm (Sehtest 15 Minuten, Briefing und Fragebogen je 5 Minuten). Kein Teilnehmer berichtete nach dem Experiment von visueller Ermüdung oder Unwohlsein.

4.6 Überprüfung des Teilnehmerverhaltens

Durch die Möglichkeit der freien Betrachtung in der VU sind die visuellen Stimuli für jeden Teilnehmer unterschiedlich. Somit war sicherzustellen, dass die Teilnehmer sich wie erwartet verhalten, wofür zwei Bedingungen formuliert wurden (Vgl. Serrano et al., 2020a, S. 3)

1. Signifikante Fixationszeit auf den Stimuli
2. Plausible Erfüllung der Suchaufgabe

Signifikante Fixationszeit auf Stimuli Da die Meta Quest 3 kein Eye-Tracking unterstützt, um sicherzustellen, dass die Teilnehmer beim Erfüllen der Suchaufgabe lange genug auf die grauen Fixations-Quadrate schauen, wurde eine statistisch fundierte Annäherung von Eye-Tracking implementiert: Wenn die optische Achse der Teilnehmer bekannt ist, lässt sich nach aktuellen Studienergebnissen eine Aussage darüber treffen, in welchem Blickwinkel um die optische Achse ihre Fixation in der VU am wahrscheinlichsten liegt: Kollenberg et al., 2010 (S. 3) ermittelten eine Augenexzentrizität (Abweichung der Augen- von der Kopfausrichtung, angegeben als Sichtwinkel) von $M = 11,46°$; Sitzmann et al., 2018 (S. 1637) zeigten, dass diese während Fixationen im Mittel $M = 11,67°$ beträgt. Die Studie von David et al., 2022 (S. 15) fand 87,6 % der Augenbewegungen unterhalb von 15° Amplitude. Modelle zur Vorhersage sakkadischer Augenbewegungen[11] unter freien Beobachtungsbedingungen, wählten z. B. Wertebereiche der Größenordnung 12,2° (Rai et al., 2016, S. 4 Fig. 4) bis 20° (Le Meur und Liu, 2015, S. 155) für Modellierung der Wahrscheinlichkeitsverteilungen.

In Unreal Engine wurde daher die optische Achse aus den erfassten Positions- und Rotationsdaten des VR-Headsets bestimmt und der Sichtwinkel berechnet, der für die Fixation des Stimulus-Mittelpunkts benötigt würde. Wenn dieser den Wert von 15° Sichtwinkel unterschritt,[12] wurde davon ausgegangen, dass der Teilnehmer den Stimulus fixiert hatte. Die

[11] = schnelle Bewegung der Augen zum Wechsel von einer Fixation zur nächsten

[12] Der Wert wurde konservativ gewählt, da der Sichtwinkel zum exakten Mittelpunkt des Stimulus bestimmt worden ist (die Mitte des grauen Quadrats im Rahmen, dem der Teilnehmer zugewandt ist). Das graue Quadrat selbst ist schließlich größer als dieser Punkt.

Fixationszeit wurde erfasst und konnte dann mit der Gesamtzeit des Versuchs in Relation gesetzt werden.

Plausible Erfüllung der Suchaufgabe Um zu prüfen, ob die Teilnehmer die Suchaufgabe tatsächlich erfüllt haben konnten, wurde die maximale laterale Auslenkung vom Ursprungspunkt aus den Tracking-Daten der Meta Quest 3 erfasst. Ergänzend dazu zeigte ein externer Bildschirm durchgängig das virtuelle Sichtfeld der Teilnehmer über das Unreal Engine VR Preview Fenster zur Überwachung des Verhaltens (s. Abb. 4.3)

5 Ergebnisse

In diesem Kapitel wird eingangs ausgewertet, ob das Teilnehmerverhalten im Kontext freier Betrachtung der formulierten Erwartung (s. Abschnitt 4.6) entspricht. Um robuste Ergebnisse zu erhalten, wird der Datensatz von Teilnehmern, die inkonsistente Extremwerterkennung zeigten, bereinigt und anschließend eine psychometrische Funktion für jede Brennweite f zur Schwellwerterkennung der Translationsverstärkung g_t angenähert. Die Signifikanz der Unterschiede zwischen den Resultaten wird beleuchtet. Um die Forschungsfragen zu beantworten, wird im Fall der Forschungsfragen FF1 und FF2 der Vergleich der psychometrischen Funktion für f_1 mit den Messergebnissen von Serrano et al. (2020a), wie in Abschnitt 3.3 erläutert, gezogen. Forschungsfragen FF3 und FF4 werden dann im Vergleich der Ergebnisse unter den Brennweiten $f_{0.5}$, f_1 und f_2 beantworten.

Abschließend geben Auswertungen der durchschnittlichen Translationsgeschwindigkeit v_{avg} Einsicht darein, ob die Brennweiten f und Translationsverstärkungen g_t mit unterschiedlichem Teilnehmerverhalten in Verbindung gebracht werden können. Das Kapitel dient der Vorstellung der Ergebnisse und Beantwortung der Forschungsfragen. Die Einordnung, sowie Diskussion folgen in Kapitel 6.

5.1 Teilnehmerverhalten und Datenbereinigung

Teilnehmer brauchten im Durchschnitt $11,61\ s$ ($SD = 2,71\ s$[1]) zum Erfüllen der Suchaufgaben. Die Daten des angenäherten Eye-Trackings lieferten eine mittleren, prozentualen Fixationsanteil der Gesamtdauer von $52,94\ \%$ ($SD = 12,51\ \%$). Die Teilnehmer verbrachten somit einen signifikanten Anteil der Gesamtzeit damit, die Stimuli zu fixieren. Durch einen Programmfehler waren die Daten zur maximalen, lateralen Auslenkung unzuverlässig und konnten nur in acht Fällen korrekt erhoben werden. Über das VR-Preview Fenster von Unreal Engine auf dem externen Bildschirm wurde jedoch für alle Teilnehmer und Versuche bestätigt, dass das blaue Zielrechteck im Sichtfeld des Teilnehmers war – somit ausreichend virtuelle Auslenkung stattgefunden hatte, um die Suchaufgaben zu erfüllen.

[1] Die *SD*-Angaben beziehen sich in diesem Abschnitt auf die Abweichung der arithmetischen Mittel je Teilnehmer zum Mittel aller Teilnehmer.

Als Kriterium für die Berücksichtigung der Antwortdaten im untersuchten Datensatz *DS* wurde die konsequente Erkennung der beiden Extremfälle herangezogen ($g_t = 0,4$; $g_t = 2,5$). Da die Extremfälle von den meisten Teilnehmern sehr akkurat erkannt worden sind, wurde nur maximal eine Falschantwort pro Extremfall[2] aus den 9 Stimuli je Extremfall (3 f × 3 Durchläufe) zugelassen. Messungen der Teilnehmer, die einen Extremfall zweimal falsch einstuften, wurden im Datensatz *DS* nicht berücksichtigt. Dadurch sind vier Teilnehmer eliminiert worden (15 verbleibend). Im Folgenden wird vornehmlich der Datensatz *DS* untersucht, zum Vergleich werden allerdings auch die Ergebnisse aus den unbereinigten Daten DS_{all} angeführt.

Die ausgeschlossenen Teilnehmer waren alle weiblich, in einer Altersspanne von 26 bis 38, hatten sehr gutes Sehvermögen (Visus 1.0 nah und fern, unbeeinträchtigtes Farb- und Stereosehen). Sie gaben an, weniger als einmal im Monat Videospiele zu spielen und „*weniger*“ bis „*gar nicht*“ erfahren in der Herstellung von 3D-Grafik zu sein. Hinsichtlich VR-Vorerfahrungen waren Angaben von „*> 20-mal*“ ein VR-Headset getragen bis „*noch nie*“ enthalten, hinsichtlich Erfahrungen mit der Produktion von Bewegtbildinhalten reichten die Antworten von „*gar nicht*“ bis „*voll und ganz*“.

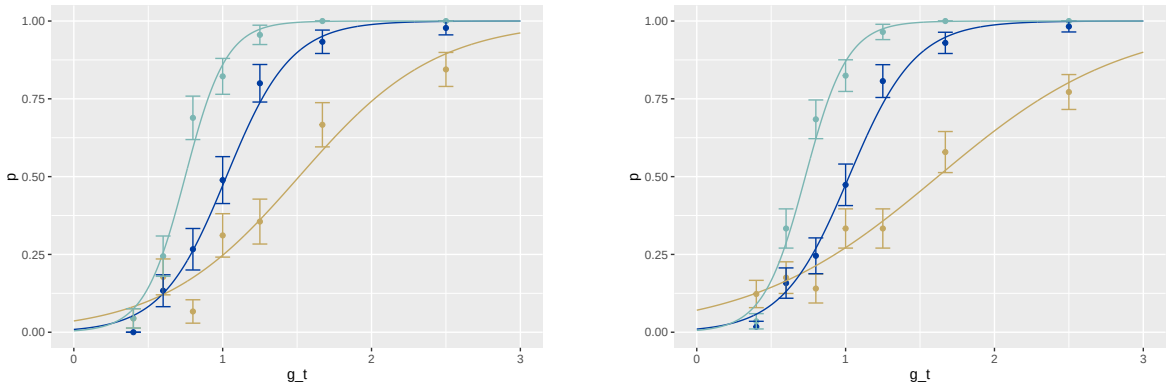
5.2 Psychometrische Funktionen je Brennweite f

Die erhobenen Daten *DS* wurden je Teilnehmer für die 21 Variablenkombinationen (3 f × 7 g_t) über die drei Durchläufe arithmetisch gemittelt. Die Antwort „*schneller*“ wurden zuvor als Wert ‚1‘ gespeichert, „*langsamer*“ als Wert ‚0‘. So ergibt sich für jeden Teilnehmer die Wahrscheinlichkeit, bei einer Variablenkombination mit „*schneller*“ geantwortet zu haben (genannt P_{faster}), direkt aus diesem Wert. Für jede Brennweite f wurde dann über ein generalisiertes, lineares Modell (GLM) eine logistische, psychometrische Funktion über die Maximum-Likelihood-Methode (engl. Maximum Likelihood Estimation = MLE) angenähert. Diese Kurve zeigt einen sigmoidalen Verlauf (S-Kurve) und wird beschrieben durch:

$$f(x) = \frac{1}{1 + e^{-(a+bx)}} \tag{5.1}$$

mit $a, e \in \mathbb{R}$. Diese Art Funktion ist in der Literatur zur psychophysischen Bestimmung von Erkennungsschwellwerten bei Anpassung von Translationsverstärkungen üblich (J. Zhang et al., 2018, S. 1676, Serrano et al., 2020a, S. 6., Langbehn et al., 2018, S. 7, Hartle und Wilcox, 2021, S. 56, Teng et al., 2023 S.404). Die Abbildung 5.1 zeigt die psychometrischen Funktionen für *DS* (a) und DS_{all} (b) im Vergleich: $f_{0.5}$ (gelb), f_1 (blau) und f_2 (grün). Die arithmetischen Mittelwerte sind als Punkte mit dem Streuungsmaß der zugehörigen Standardfehler (engl. standard error of the mean = SEM) angegeben.

[2] = maximal 1 × „*schneller*“ geantwortet bei $g_t = 0,4$ UND maximal 1 × „*langsamer*“ geantwortet bei $g_t = 2,5$



(a) *DS* (b) DS_{all}

Abbildung 5.1: Angenäherte, psychometrische Funktion je Brennweite $f_{0.5}$ (gelb), f_1 (blau) und f_2 (grün). Die x-Achse zeigt die Translationsverstärkung g_t, die y-Achse die Wahrscheinlichkeit P_{faster} dass „*schneller*“ geantwortet wurde. (a) Datensatz bereinigt *DS* (b), Datensatz aller Teilnehmer D_{all}.

Tabelle 5.1: *PSE*, $DT_{25\%}$ und $DT_{75\%}$ aus den psychometrischen Funktionen (s. Abb. 5.1) je Brennweite f: (a) Datensatz bereinigt *DS* (b) Datensatz aller Teilnehmer DS_{all}

(a) *DS*

	$f_{0.5}$	f_1	f_2
$DT_{25\%}$	1,01	0,79	0,60
PSE	1,51	1,02	0,75
$DT_{75\%}$	2,02	1,26	0,90

(b) DS_{all}

	$f_{0.5}$	f_1	f_2
$DT_{25\%}$	0,93	0,78	0,57
PSE	1,62	1,02	0,73
$DT_{75\%}$	2,31	1,27	0,89

Aus den genäherten Funktionen wurde der *PSE* (engl. point of subjective equality) und die Erkennungsschwellwerte $DT_{25\%}$ und $DT_{75\%}$ (engl. detection threshold = DT) bestimmt: Der PSE beschreibt hierbei diejenige Translationsverstärkung g_t, bei der $P_{faster} = 0,5$ gilt, also die Teilnehmer zu gleichen Teilen mit „*schneller*“ oder „*langsamer*“ geantwortet haben. Daraus wird geschlussfolgert, dass dieser Wert für g_t als der erwarteten Geschwindigkeit entsprechend und somit als am natürlichsten wahrgenommen wird. $DT_{25\%}$ und $DT_{75\%}$ bilden die Werte von g_t, für die $P_{faster} = 0,25$, respektive $P_{faster} = 0,75$ gilt. Durch den sigmoidalen Verlauf der angenäherten Funktion liegen ihre Werte stets symmetrisch um den *PSE*. Tabelle 5.1 zeigt die erhobenen Ergebnisse für alle Brennweiten f.

5.2.1 Beantwortung der Forschungsfragen FF1 und FF2

Die Messwerte der *PSE* bei Serrano et al. (2020a) deuteten an, dass eine Distanz von ca. $H' = 2\,m$ in der Wahrnehmung von $g_t = 1$ als natürlich resultieren würde ($H' = 1,75\,m \implies PSE = 0,92$; $H' = 2,5\,m \implies PSE = 1,10$). Der in dieser Arbeit erhobene Messwert von $PSE = 1,02$ (Konfidenzintervall $CI_{95\%}$ $[0,97; 1,09]$) fällt ziemlich exakt auf diesen Wert. Somit lässt sich die **Forschungsfrage FF1** beantworten: Es ist kein signifikanter Einfluss der Rahmung einer Szene auf die als natürlich wahrgenommene Translationsverstärkung g_t der im Rahmen sichtbaren, durch laterale Kopfbewegungen selbstinduzierten Motion Parallax feststellbar.

Forschungsfrage FF2 ist über die Erkennungsschwellwerte $DT_{25\%}$ und $DT_{75\%}$ bei Brennweite f_1 aufgelöst: Eine prozentuale Abweichung der Translationsverstärkung g_t im Bereich $\pm 23,5$ % vom *PSE* wird nicht eindeutig als verändert wahrgenommen. Um diesen Faktor lässt sich die Translation des Stereo-Kamerasystems $T_{v_{cam}}$ relativ zum *PSE* unbemerkt stauchen oder strecken. Die Messwerte von Serrano et al. (2020a) zeigen hier eine deutlich niedrigere Sensibilität der Teilnehmer für die veränderte Translationsübersetzung ($\pm 39,1$ % vom *PSE*). Daraus lässt sich schlussfolgern, dass die Sensibilität für veränderte Translationsübersetzungen erhöht ausfällt, wenn von der Translationsverstärkung g_t nur der Blickwinkel durch einen Rahmen in einer sonst natürlich repräsentierten VU betroffen ist.

5.2.2 Beantwortung der Forschungsfragen FF3 und FF4

Zur Beantwortung der Forschungsfragen FF3 und FF4 wird der Vergleich der gemessenen psychometrischen Funktionen je Brennweite f herangezogen und mittels Signifikanztests ermittelt, ob die Unterschiede statistisch signifikant ausfallen:

Vergleich der psychometrischen Funktionen Die psychometrischen Funktionen der Brennweiten f zeigen sichtbar unterschiedliche Verläufe: Während der Wert für f_1 mit $PSE = 1,02$ ($CI_{95\%}$ $[0,95; 1,10]$) nah an $g_t = 1$ liegt, fällt er bei doppelter Vergrößerung f_2 um rund ein Viertel kleiner ($PSE = 0,75$ mit $CI_{95\%}$ $[0,69; 0,80]$) und bei halber Vergrößerung $f_{0.5}$ um rund die Hälfte größer ($PSE = 1,51$ mit $CI_{95\%}$ $[1,38; 1,65]$) aus.

Auch hinsichtlich der Erkennungsschwellwerte $DT_{25\%}$ und $DT_{75\%}$ ergeben sich Unterschiede durch die variierende Steilheit der Funktionen: Bei f_2 hat der annähernd lineare Anteil um den *PSE* die größte Steigung, bei $f_{0.5}$ wurde die geringste Steigung verzeichnet. Durch die unterschiedlichen *PSE* ist der absolute Wertebereich (die absolute Steigung der Funktion um den *PSE*) im Vergleich weniger aussagekräftig, als der relative Wertebereich abhängig vom

PSE: Bei f_2 liegen die Schwellwerte im Bereich ±20 % vom *PSE*, bei f_1 im Bereich ±23, 5 % und bei $f_{0.5}$ im Bereich ±33, 8 %.

Signifikanztests von *PSE* und *DT* Um zu bestimmen, ob zwischen den erhobenen Kennzahlen (*PSE*, $DT_{25\%}$ und $DT_{75\%}$) der Brennweiten f statistisch signifikante Unterschiede vorliegen, ist es entscheidend, den für die Daten passenden Signifikanztest durchzuführen: Dazu sind für jeden Teilnehmer einzeln psychometrische Funktionen angenähert und die Werte für $DT_{25\%}$, $DT_{75\%}$ und *PSE* gesondert in Data Frames untersucht worden. Nur so können die gemittelten Messwerte für die Gesamtheit der Teilnehmer mit der Stichprobenzahl (Anzahl der Teilnehmer) und ihren Einzelergebnissen in Verbindung gebracht werden. Die Data Frames sind nicht normalverteilt ($p < 0,05$ nach Shapiro-Wilk Test), weshalb die nichtparametrische Version der ANOVA mit Messwiederholung – der Friedman Test – verwendet wird (Marino, 2018, S. 134). Dieser ergab für alle Data Frames Werte von $p < 0,5$. Die Nullhypothese des Friedman-Tests, dass die Unterschiede der Medianwerte der Kennzahlen zwischen den Brennweiten f nur dem Zufall zuzuschreiben sind, kann somit abgelehnt werden (Marino, 2018, S. 134): Es gibt folglich signifikant unterschiedliche Ergebnisse zwischen den Brennweiten f für *PSE*, $DT_{25\%}$ und $DT_{75\%}$.

Zur Feststellung, welche Werte sich signifikant unterscheiden, ist die Auswertung mittels Dunn-Bonferroni Post-hoc und Bonferroni Korrektur für multiple Tests erfolgt (Marino, 2018, S. 134). Tabelle 5.2 zeigt die paarweisen Vergleiche und die zugehörigen Wahrscheinlichkeiten p[3]: Die *PSE*- und $DT_{75\%}$-Werte sind zwischen allen Brennweiten f signifikant unterschiedlich. Selbes gilt für $DT_{25\%}$ zwischen f_1 und f_2, sowie zwischen $f_{0.5}$ und f_2. Einzig der Unterschied der Mediane von $DT_{25\%}$ zwischen $f_{0.5}$ und f_1 ist nicht signifikant eingestuft. Dies könnte auf die schwankenden Antworten der Teilnehmer bei $f_{0.5}$ und niedrigen g_t-Werten (im Bereich $0,4 \leq g_t \leq 1,25$) zurückzuführen sein. Diese Schlussfolgerung wird von den Signifikanztests für den Datensatz aller Teilnehmer DS_{all} bestärkt (s. Abschnitt 5.4).

Bezüglich **Forschungsfrage FF3** ist somit ein signifikanter Einfluss der Skalierung (Brennweite f) auf die als natürlich wahrgenommene Translationsübersetzung g_t der im Rahmen sichtbaren, durch laterale Kopfbewegungen selbstinduzierten Motion Parallax vorhanden: Bei doppelter Vergrößerung der Szenenprojektion (f_2) gegenüber der homothetischen Konfiguration (f_1) wird eine 26, 5 % langsamere virtuelle Translation $T_{v_{cam}}$ als natürlich empfunden, während dies für eine 48, 0 % schnellere virtuelle Translation bei halber Vergrößerung ($f_{0.5}$) zutrifft. Der Versatz der psychometrischen Kurven zeigte sich auch in den (mit nur einer Ausnahme) signifikant unterschiedlichen *DT*-Werten zwischen den Brennweiten f. Es lässt sich somit feststellen: Wird eine Szene gegenüber der homothetischen Konfiguration (bei gleichem Betrachtungsabstand H' und unverzerrter Tiefendarstellung) vergrößert gezeigt,

[3]Bedeutung wie bei Friedman Test, nun im paarweisen Vergleich

Tabelle 5.2: Ergebnisse der Signifikanztest (Dunn-Bonferroni Post-hoc mit Bonferroni Korrektur) für Unterschiede von *PSE*, $DT_{25\%}$ und $DT_{75\%}$ zwischen den Brennweiten f: Bei $p < 0,5$ (mit ‚*' markiert) liegt ein signifikanter Unterschied vor: (a) Datensatz bereinigt *DS* (b) Datensatz aller Teilnehmer DS_{all}

(a) *DS*

	$f_{0.5}$- f_1	f_1- f_2	$f_{0.5}$- f_2
p für $DT_{25\%}$	0,4806	0,0302*	0,0002*
p für *PSE*	0,0356*	0,0088*	0,0000*
p für $DT_{75\%}$	0,0104*	0,0290*	0,0000*

(b) DS_{all}

	$f_{0.5}$- f_1	f_1- f_2	$f_{0.5}$- f_2
p für $DT_{25\%}$	1,0000	0,0150*	0,0022*
p für *PSE*	0,1322	0,0052*	0,0000*
p für $DT_{75\%}$	0,0370*	0,0098*	0,0000*

ist eine langsamere Blickpunktveränderung zu wählen, bei Verkleinerung eine beschleunigte – insofern eine als natürlich wahrgenommene Motion Parallax gewünscht ist.

Die Betrachtung der von $DT_{25\%}$ und $DT_{75\%}$ begrenzten, relativen Wertebereiche um den *PSE*-Wert deuten eine Tendenz hinsichtlich **Forschungsfrage FF4** an: So werden relative Veränderungen von g_t um den natürlich empfundenen Wert *PSE* bei doppelter Vergrößerung f_2 schon mit 3, 5 % weniger Veränderung erkannt, verglichen mit f_1. Die halbe Vergrößerung $f_{0.5}$ zeigt im Vergleich mit f_1 eine Desensibilisierung für die g_t-Veränderung um 10, 3 %. Zwischen halber und doppelter Vergrößerung liegen somit 13, 8 %, was einen signifikanten Einfluss der Brennweite f auf die Erkennung vom *PSE* abweichender Translationsverstärkungen g_t vermuten lässt. Eine Bestimmung der statistischen Signifikanz der Unterschiede dieser prozentualen Abweichungen wurde dem zeitlichen Rahmen der Bachelorarbeit geschuldet nicht vorgenommen – die Ergebnisse bieten somit erst einmal einen nicht statistisch validierten Eindruck der Wirkung.

5.3 Einfluss der Variablen auf das Teilnehmerverhalten

Der Boxplot in Abbildung 5.2 zeigt die statistische Verteilung der Durchschnittsgeschwindigkeiten v_{avg} der realen, lateralen Translationsbewegungen für alle Teilnehmer aus DS_{all} (die ID-Nummer ist auf der x-Achse aufgetragen, v_{avg} auf der y-Achse): Die Mediane reichen von

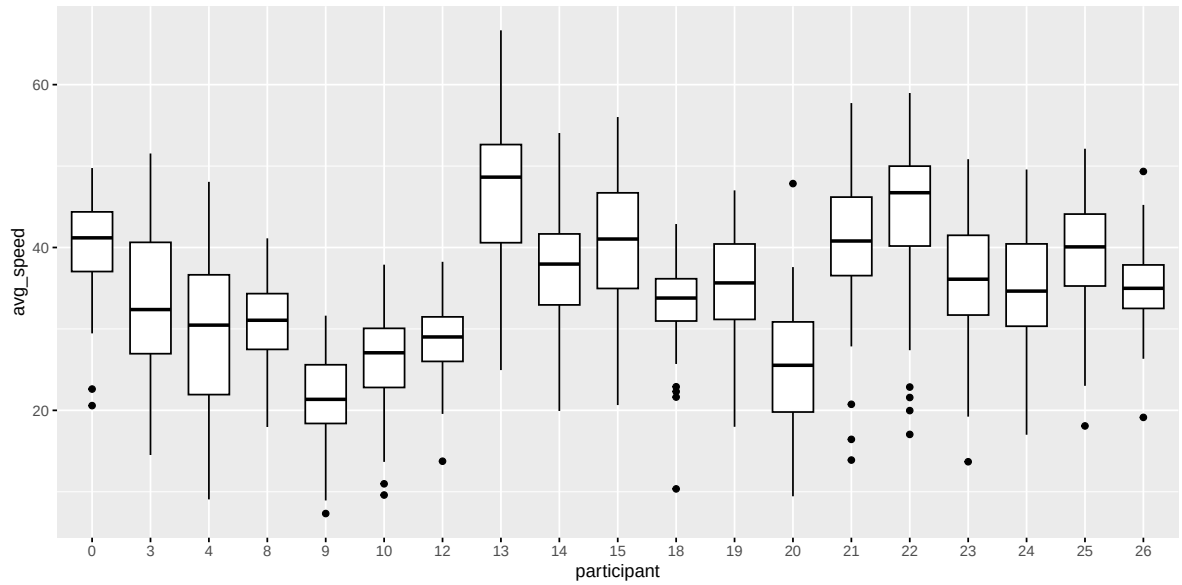


Abbildung 5.2: Boxplots der durchschnittlichen realen, lateralen Translationsgeschwindigkeit v_{avg} abhängig vom Teilnehmer (ID-Nummer)

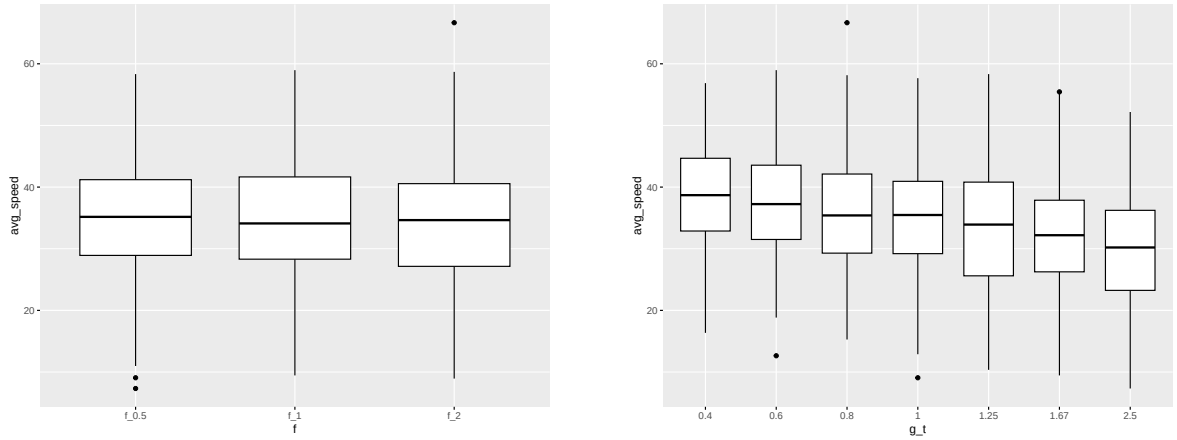
$0.22\,m/s$ bis $0,48\,m/s$, was bei einer Fixationsdistanz von $H' = 2\,m$ Winkelgeschwindigkeiten für die Verfolgung $d\alpha/dt$ von $6,28\,°/s$ bis $13,50\,°/s$ nach der Formel

$$d\alpha/dt = \arctan(\frac{v_{avg}}{H'}) \tag{5.2}$$

ergibt. Somit lag das Teilnehmerverhalten innerhalb des zur akkuraten Wahrnehmung von Motion Parallax benötigten Wertebereichs von $5\,°/s \leq d\alpha/dt \leq 20\,°/s$ (Holmin und Nawrot, 2015, S. 45; s. Abschnitt 2.2.2).

Abbildung 5.3(a) trägt v_{avg} abhängig der Brennweite f auf. Hier ist kein nennenswerter Unterschied zwischen $f_{0.5}$, f_1 und f_2 erkennbar: Somit passten die Teilnehmer ihre Translationsgeschwindigkeit nicht an die durch die Skalierung nachweislich veränderte Wahrnehmung ihrer Translation an.

Abbildung 5.3(b) lässt hingegen einen Einfluss der Variable g_t auf v_{avg} vermuten: So liegt der Median für $g_t = 0,4$ bei $v_{avg} = 0,38\,m/s$ und fällt zum größten $g_t = 2,5$ kontinuierlich auf $v_{avg} = 0,30\,m/s$ ab. Auf einen Signifikanztest der Unterschiede wurde, auch hier dem Zeitrahmen der Bachelorarbeit geschuldet, verzichtet. Die Tendenz legt dennoch nahe, dass die Teilnehmer ihre reale Translation T_r an die Geschwindigkeit der virtuellen Translation $T_{v_{cam}}$ anpassten.



(a) v_{avg} abhängig von f

(b) v_{avg} abhängig von g_t

Abbildung 5.3: Boxplots der durchschnittlichen realen, lateralen Translationsgeschwindigkeit v_{avg} abhängig von f (a) und g_t (b)

5.4 Kurzüberblick: Unbereinigter Datensatz DS_{all} aller Teilnehmer

Für den Datensatz DS_{all} wurden ebenfalls psychometrische Funktionen angenähert und Signifikanztests zum Vergleich durchgeführt: Die beschriebenen Erkenntnisse für DS gelten insgesamt auch für den Datensatz DS_{all}. Die psychometrischen Funktionen und resultierenden Werte (PSE, $DT_{25\%}$ und $DT_{75\%}$) für f_1 und f_2 sind nahezu identisch, wenn auch minimal konservativer. Die Werte für $f_{0.5}$ zeigen einen deutlich größeren $PSE = 1,62$, sowie einen größeren relativen Wertebereich von $\pm 42,6$ % der Schwellwerte $DT_{25\%}$ und $DT_{75\%}$ um den PSE. Dies liegt darin begründet, dass die ausgeschlossenen Teilnehmer hier die meisten falschen Antworten der Extremfälle verzeichneten (Ausschlusskriterium, s. Abschnitt 5.1). Sie antworteten im Mittel zu 41,75 % für $g_t = 0,4$ fälschlicherweise mit „*schneller*“ und zu 50 % fälschlicherweise „*langsamer*“ für $g_t = 2,5$. Die 15 in DS enthaltenen Teilnehmer zeigen hier Werte von 4,44 % und 15,56 %. In DS liegen die Werte durch Berücksichtigung der vier ausgeschlossenen Teilnehmer dann bei 12,28 % und 22,81 %, was durch die weniger eindeutige Erkennung zu dem flacheren Kurvenverlauf und dem Versatz des PSE führt.

Die Ergebnisse der Signifikanztests für die paarweisen Wertvergleiche fallen für DS_{all} identisch zu DS aus, mit der Einschränkung, dass neben dem unteren Schwellwert $DT_{25\%}$ auch der PSE zwischen $f_{0.5}$ und f_1 nicht signifikant unterschiedlich ausfällt. Die Unsicherheit in den Antworten bei halber Vergrößerung $f_{0,5}$ (s. Abschnitt 5.2) streut die Ergebnisse hier noch mehr als schon in DS (s. Fehlerbalken in Abb. 5.1(b)), wodurch im Fall von DS_{all} auch der Bereich des PSE von f_1 betroffen ist. Dies bekräftigt die These, dass in DS durch die

schwankenden Antworten in niedrigen Wertebereichen von g_t kein signifikanter Unterschied zwischen $DT_{25\%}$ von $f_{0.5}$ und f_1 festgestellt werden konnte.

Ferner ist zu erkennen, dass die beinahe identischen *PSE*- und *DT*-Werte für f_1 und f_2 zwischen *DS* und DS_{all} (s. Tabelle 5.1) in den paarweisen Vergleichen an Signifikanz eingebüßt haben; Durch die geringere Teilnehmeranzahl fällt bei gleichen arithmetischen Mittelwerten der *SEM* größer aus (s. Fehlerbalken in Abb. 5.1) und die Wahrscheinlichkeit, dass die Unterschiede auf Zufall zurückzuführen sind, steigen.

6 Diskussion

In diesem Kapitel werden die Ergebnisse des psychophysischen Experiments eingeordnet und mögliche Erklärungsansätze für Abweichungen von anderen Messungen oder theoretischen Modellen erörtert. Wahrnehmungsseitige Wechselwirkungen der Tiefenhinweise in den Stimuli werden abgewogen und diskutiert, ob und inwiefern die bisherigen Modelle im Fall projizierter Geometrien gelten. Offene Fragen und Einschränkungen des Experiments werden beleuchtet, um weitere Forschung im Bereich gerahmter, stereoskopischer Bildmedien zu motivieren.

6.1 Einfluss der Rahmung

Die Rahmung zeigte in den Ergebnissen (im Vergleich mit den Messwerten von Serrano et al., 2020a, S. 6) keinen messbaren Einfluss auf die Wahrnehmung natürlicher Translationsverstärkung g_t. Die Messwerte der gerahmten Stimuli glichen den der äquivalenten ungerahmten. Dies ist zu erwarten, da sich der Stimulus in homothetischer Konfiguration prinzipiell nur wie ein schwebendes Fenster in einen anderen Raum/in die Außenwelt verhalten sollte. Die Ergebnisse bestätigen somit zum einen die Werte von Serrano et al. (2020a) und legen nahe, dass, trotz der vielen projektiv zu berücksichtigen Parameter (s. Abschnitt 2.3, 2.4), ein vergleichbarer Aufbau realisiert werden konnte.

Die erhöhte Sensibilität bei gerahmten Stimuli für abweichende Translationsverstärkungen g_t vom *PSE* (Bereich zwischen $DT_{25\%}$ und $DT_{75\%}$) verglichen mit den ungerahmten Stimuli bei Serrano et al. (2020a) ließe sich durch die Betrachtungsumgebung erklären: Von g_t ist im Experiment dieser Arbeit nur die Translation des Blickwinkels im Rahmen $T_{v_{cam}}$ betroffen, die eigentliche Bewegung des Betrachters in der VU T_v bleibt stets realitätsgetreu. Somit bieten die VU um die gerahmten Stimuli – und die fixierte Position der Rahmen selbst – Anhaltspunkte für den Betrachter hinsichtlich seiner unverzerrten, virtuellen Translation: Folglich werden Veränderungen vom *PSE* schneller erkannt als ohne Referenz.

Diese These wird dadurch unterstützt, dass die Teilnehmer ihre durchschnittliche Translationsgeschwindigkeit v_{avg} deutlich weniger an die veränderte Translationsverstärkung g_t im Rahmen anpassten: Der Median von v_{avg} fiel zwar kontinuierlich im Wertebereich von

$g_t = \{0,4; \ldots; 2,5\}$ ab $\tilde{v}_{avg} = \{0,38\ m/s; \ldots; 0,30\ m/s\}$, jedoch weitaus weniger als beim Vergleichsexperiment mit $\tilde{v}_{avg} \approx \{0,55\ m/s; \ldots; 0,08\ m/s\}$ über denselben Wertebereich (abgelesene Werte aus Serrano et al., 2020b, S. 4 Fig. 3). Die Teilnehmer kompensierten die Translationsanpassung somit nicht vollständig, um konstant die gleiche virtuelle Translationsgeschwindigkeit zu erhalten (bei Serrano et al., 2020a, S. 5 im Mittel $0,2\ m/s$). Sie bemerkten die Translationsgeschwindigkeit in der VU um die gerahmten Stimuli. Folgeuntersuchungen dazu, wie Rahmengröße (Größe im Sichtfeld aus α'_h und α'_v) und Rahmenabstand H' die Schwellwertwahrnehmung beeinflussen, wären interessant – Die naheliegende Hypothese, dass mit steigender Größe im Sichtfeld (und somit weniger Referenz) die Sensibilität für Veränderungen sinkt, wäre zu prüfen.

6.2 Einfluss der Skalierung

Die Begründung des gemessenen Einflusses der Skalierung (Brennweite f) der Szene im Rahmen auf die Wahrnehmung der natürlichen Translationsverstärkung g_t ist hingegen deutlich unübersichtlicher: Ziel des Experiments war, einen ersten Eindruck über die gesamtheitliche Wirkung der Brennweitenveränderung zu erlangen. Aufschluss über die resultierenden Wechselwirkungen der einzelnen Projektionsbestandteile und wie stark diese jeweils Einfluss nehmen, sind aus den Daten nicht direkt erkennbar. Da dies die erste Arbeit zu Translationsverstärkungen g_t im Kontext gerahmter VU – und durch die Untersuchung der Brennweiten auch die erste im Kontext gerahmter Projektionen von VU – ist, sind die Messwerte schwer in den Forschungsstand einzuordnen. Die Skalierung der Szene im Rahmen stellt kein in der RU mögliches Seherlebnis dar und widersprüchliche Tiefenhinweise sind somit immanent. Dennoch werden die erhobenen Werte im Folgenden beleuchtet und Einflussfaktoren, die zu den Ergebnissen beigetragen haben könnten, aufgezeigt:

6.2.1 Konsistenz zwischen Retinalgeschwindigkeit und binokularer Tiefe

In Abschnitt 2.4 wurde dargelegt, dass die doppelte Brennweite f_2 der homothetischen Brennweite f_1 bei lateraler Kopftranslation näherungsweise eine Verdopplung der Retinalgeschwindigkeit $d\theta/dt$ nach sich zieht. Selbes gilt für eine Verdopplung von g_t bei gleichbleibender Brennweite f_1. Da $d\alpha/dt$ von der Translationsverstärkung g_t und Brennweite f unabhängig ist, ergibt sich nach dem geometrischen Modell des M/PR (s. Abschnitt 2.2.2; Nawrot und Stroyan, 2009, S. 1970) für eine Verdopplung der Retinalgeschwindigkeit $d\theta/dt$ auch eine Verdopplung des Distanzverhältnisses $\frac{d}{f}$. Nach dem empirischen Modell (Nawrot et al., 2014,

S. 11) hingegen ist eine Veränderung der tatsächlich wahrgenommenen relativen Distanz um den Faktor $1,33$ zu erwarten:

$$\frac{\frac{d^*}{f}}{\frac{d}{f}} = \frac{\frac{(2d\theta)^{0,416}}{d\alpha^{0,192}} \cdot 0,0313}{\frac{d\theta^{0,416}}{d\alpha^{0,192}} \cdot 0,0313} \Rightarrow \frac{d^*}{d} = 2^{0,416} \approx 1,33 \tag{6.1}$$

Eine Halbierung der Brennweite f_1 zu $f_{0.5}$ würde analog einen Faktor von $\frac{d^*}{d} = 0,5^{0,416} \approx 0,75$ nach sich ziehen. Da die binokulare Disparität die Szene in der Tiefe unverzerrt zeigt, wäre für die Tiefenstaffelung jedoch ein Verhältnis von $\frac{d^*}{d} = 1$ zu erwarten. Somit wäre für die Retinalgeschwindigkeit der ausgleichende Faktor $g_t = 0,5$, respektive $g_t = 2$ vorzunehmen (s. Abschnitt 2.4).

Die *PSE* Messwerte für f_2 mit $g_t = 0,75$ und für $f_{0.5}$ mit $g_t = 1,51$ spiegeln dies jedoch nicht wider. Die Skalierung der Szene beeinflusst die Wahrnehmung natürlicher Motion Parallax somit hin zu moderateren Translationsanpassungen, als aus den MP/R Modellen anzunehmen wären. Dadurch lässt sich festhalten, dass die Retinalgeschwindigkeit $d\theta/dt$ **nicht** passend zur suggerierten Tiefe aus binokularer Disparität als natürlich empfunden wird. Dies steht im Widerspruch zur Veto-Funktion der binokularen Disparität gegenüber der Motion Parallax (Hartle und Wilcox, 2021, S. 61; s. Abschnitt 2.2.3).Die beiden Tiefenhinweise treten zwar nicht völlig isoliert auf, da auch alle anderen monokularen Tiefenhinweise durch den veränderten Bildwinkel verändert werden. Allerdings ist die Veto-Funktion bei Szenen ohne Projektion so gut belegt (s. Abschnitt 2.2.3), dass die Erkenntnis ihres Außerkraftsetzens die Zweifel an dem Konzept einzelner wirklich separierter Tiefenhinweise (s. Abschnitt 2.1.2) bestärken könnte.

Es lässt sich definitiv festhalten, dass in der Umsetzung des Videoformats die als natürlich wahrgenommenen Translationsgeschwindigkeiten g_t nicht der binokularen Tiefendarstellung (bzw. gewählter Tiefenstaffelung aus Interokulardistanz b der virtuellen Kameras) und so resultierenden Retinalgeschwindigkeiten $d\theta/dt$ entsprechend gewählt werden sollten. Die Umsetzung von translation gain maps (s. Abschnitt 3.3) aus den Retinalgeschwindigkeiten $d\theta/dt$ der Tiefenstaffelung ist somit zumindest nicht trivial möglich. Es werden mehr Daten verschiedener Tiefenstaffelungen in unterschiedlichen Projektionen und Kamera-Interokulardistanzen b benötigt, um ein umfassenderes Bild des Einflusses aller Parameter zu ermöglichen.

Die Wahrnehmung scheint von den durch die Projektion veränderten, monokularen, visuellen Tiefenhinweisen (s. Tabelle 2.1) beeinflusst: Eine zur homothetischen Konfiguration identische Retinalgeschwindigkeit $d\theta/dt$ für die fundamental durch f_2 und $f_{0.5}$ veränderten Projektionen (s. Abschnitt 2.4), wäre möglicherweise auch verwunderlich. So ist davon auszugehen, dass die drei Skalierungen im direkten Vergleich von den Teilnehmern klar erkannt wurden und sie daher auch ein anderes Verhalten der Retinalgeschwindigkeiten $d\theta/dt$ erwarten würden.

6.2.2 Theorie der veränderten Distanzwahrnehmung

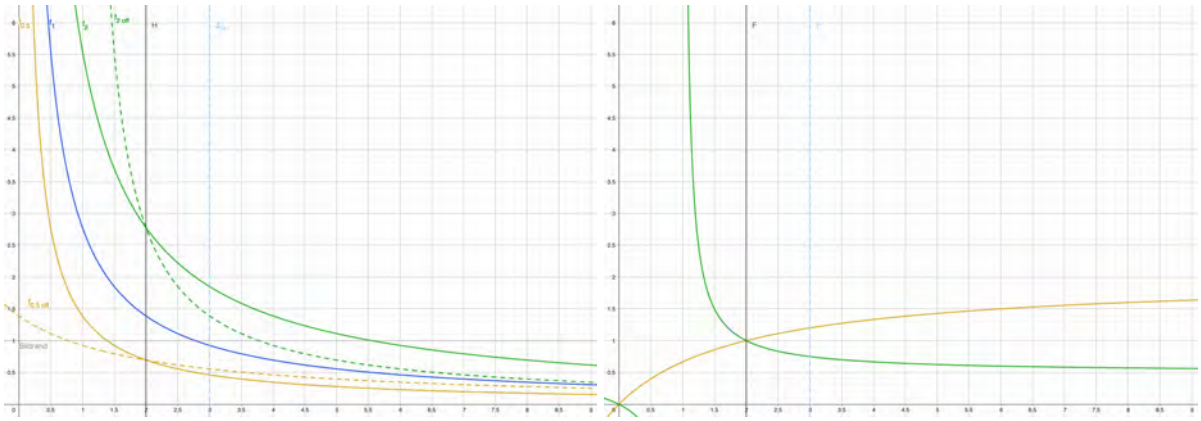
Ein intuitiver Ansatz wäre der Eindruck der Teilnehmer, dass statt einer Skalierung der Szenengeometrie entlang aller Achsen[1] im Rahmen, ein Blickpunkt mit einer anderen Distanz zur Fokusebene gezeigt würde. Eine doppelte Vergrößerung f_2 der Objekte in der Fokusebene wäre dann einer Halbierung des Abstands $H'^* = 0,5 \cdot H'$ gleichzusetzen. Einige Teilnehmer äußerten sich bei ihren ersten Durchläufen mit vergrößerten Stimuli, dass sie die Szene als „näher" empfanden (umgekehrt bei verkleinerten Stimuli als „weiter weg"). Dem HVS steht eine Veränderung seines Sichtwinkels nicht zur Verfügung. Der einzige Weg, ein Objekt doppelt so groß anzusehen, ist es, näher heranzutreten (Tiefenhinweis relativer Größe, s. Tabelle 2.1). Hierfür wäre es für zukünftige Experimente interessant, das Fixationsquadrat in konstanter Größe trotz Vergrößerung mit der hier untersuchten realitätsgetreuen Vergrößerung gegenüberzustellen, um so diesen Zusammenhang näher zu betrachten. Die perspektivischen Implikationen der Distanzveränderung werden geometrisch aufgezeigt:

Dafür lässt sich auf die Projektionszusammenhänge aus Abschnitt 2.4 zurückgreifen. Eine repräsentative Skalierungsfunktionen $S_x(Z_p)$ nach Formel (2.9) wird für einen Punkt P_1 im Abstand von $P_{1_x} = 1\ m$ zur optischen Achse[2] für die Brennweiten $f_{0.5}$ (gelb) , f_1 (blau) und f_2 (grün) aufgetragen (s. Abb. 6.1(a)). Ab einem Wert von $S_x \leq 1$ ist die Projektion $P_{1,proj*}$ im Bild zu sehen (in Abb. 6.1(a) als „Bildrand" eingezeichnet). Für f_2 und $f_{0.5}$ sind gestrichelt die Kurven $f_{2,off}$(grün gestrichelt)und $f_{0.5,off}$(gelb gestrichelt)mit den vermuteten Distanzveränderungen (f_2: Halbierung, $f_{0.5}$: Verdopplung), die ihre Skalierungsveränderung des grauen Fixations-Quadrats ersetzen würden, eingezeichnet. Für sie wurde folglich die homothetische Brennweite f_1 eingestellt. Die Schnittpunkte der gleichfarbigen Kurven in der Distanz $Z_P = 2\ m = H = H'$ zeigen, dass die Skalierung in der Projektionsebene somit identisch ausfällt.

Aus den Kurven in Abb. 6.1(b) lässt sich erkennen, dass eigentlich eine deutlich sichtbare Perspektivverzerrung der Objekte hinter dem Fixations-Quadrat stattfinden würde durch die Abstandsveränderungen mit homothetischer Brennweite f_1. In der Szenenprojektion wäre der näher gelegene Baum hinter dem Quadrat ($Z = 8,3\ m$) mit Konfiguration f_{2off} um 43% verkleinert, mit $f_{0.5off}$ um 61% vergrößert zu sehen, verglichen mit f_2, respektive $f_{0.5}$. Die Skalierung des Fixations-Quadrats ist somit zwar wie erwartet, was die Distanzänderung plausibel macht, der Szenenhintergrund würde sich dennoch sichtlich anders zeigen. Ob

[1] Ausgenommen der Tiefenachse, sollte trotz der ungewohnten projektiven Geometrie weiterhin von der Veto-Funktion binokularer Disparität ausgegangen werden.

[2] Die Verläufe bleiben unbeeinflusst von dieser Wahl, da sie alle gleichermaßen linear mit ihm skalieren. Der Wert ist groß gewählt für eine leserliche Skalierung der Achsen, ein Objekt dieser Größe (2 m breit, wenn zentriert um optische Achse) ist erst bei großen Distanzen im Bild sichtbar, wie der eingezeichnete „Bildrand" in Abb. 6.1(a) zeigt.



(a) Vergleich: Distanz- statt Brennweitenänderung (b) Resultierende relative Skalierungsänderung

Abbildung 6.1: Vergleich perspektivischer Skalierungsfunktionen: Die x-Achse zeigt die Objektdistanz Z_P in Metern, Fokusebene ($H = 2\ m$) und Distanz des blauen Zielrechtecks ($Z_R = 3\ m$) sind markiert. Die y-Achse zeigt: (a) die Größe der Projektion $S_x(Z_P)$ für: $f_{0.5}$ (gelb) | $f_{0.5off}$ (gelb gestrichelt) = f_1 mit doppelter Distanz | f_1 (blau) | f_2 (grün) | f_{2off} (grün gestrichelt) = f_1 mit halber Distanz. (b) das Verhältnis der Skalierungsfunktionen: $\frac{f_{0.5off}}{f_{0.5}}$ (gelb) | $\frac{f_{2off}}{f_2}$ (grün).

diese Transferleistung des tatsächlich realistischen Skalierungsabfalls von den Teilnehmenden intuitiv wahrgenommen werden könnte, ist jedoch fraglich. Vielmehr ist die dann zu erwartende Retinalgeschwindigkeit $d\theta/dt$ interessant:

Da sich durch die Änderung der Fixationsdistanz f alle relativen Distanzen $\frac{d}{f}$ ändern würden, lässt sich kein linearer Skalierungsfaktor als Einfluss der Distanzänderung errechnen. Mithilfe des GeoGebra-Online Tools aus Abschnitt 2.4 (s. Anhang GeoGebra Online-Tool, S. 70) wurde das Verhalten einiger Bildpunkte im Rahmen so simuliert: Das blaue Zielrechteck ($Z = 3m;\ P_x = 0m$), der nächstgelegene Baum ($Z = 8,3m;\ P_x = 1,1m$), die erste Bank auf der linken Bildseite ($Z = 15m;\ P_x = 2m$) und der Fluchtpunkt ($Z = \infty;\ P_x = 0m$): Die Retinalgeschwindigkeiten $d\theta/dt$ multiplizierten sich gegenüber der homothetischen Konfiguration um die Faktoren $\{1,51; 1,14; 1,07; 1\}$ für f_2 und um $\{0,60; 0,82; 0,89; 1\}$ für $f_{0.5}$. Somit wäre für Objekte nah der Konvergenzebene eine höhere/niedrigere Retinalgeschwindigkeit erwartet – Richtung Fluchtpunkt nähert sich der Verlauf der homothetischen Erwartung. Um die durch die Skalierung verdoppelte/halbierte Retinalgeschwindigkeit $d\theta/dt$ auf diese Erwartungswerte zu korrigieren (sodass der Faktor ‚2' in Formel (6.1) statt ‚1' diesen Wert annimmt), wären die Faktoren $\{0,76; 0,57; 0,54; 0,5\}$ für f_2 und $\{1,20; 1,64; 1,78; 2,0\}$ für $f_{0.5}$ nötig. Achtet der Betrachter also vor allem auf die Elemente näher am Fixationspunkt für seine Einschätzung, zeigt die Hypothese der wahrgenommenen Distanzänderung eine bessere Übereinstimmung mit den Messwerten. Die Distanzänderung als Einflussfaktor für die hier

erhobenen *PSE*-Werte je Vergrößerung hin zu konservativeren Translationsanpassungen wäre somit grundlegend plausibel. Hier könnte weiterführende Forschung ansetzen und tatsächliche Distanzveränderungen mit Skalierungsveränderungen (Projektionsveränderungen) in ihrer Wirkung vergleichen.

6.2.3 Translationsverstärkung und optische Achse

Interessant an den Messwerten ist die geringere Empfindlichkeit für Abweichungen der Translationsverstärkung g_t vom *PSE* bei halber Vergrößerung $f_{0.5}$ – respektive der leichte Anstieg bei f_2 – gegenüber f_1 (f_2: ±20 %; f_1: ±23, 5 %; $f_{0.5}$: ±33, 8 %). Die oben beschriebene Referenz in Form des umliegenden Raums war für alle Stimuli gleich. In den Forschungsergebnissen zu VR-Locomotion wird eine höhere Empfindlichkeit für Translationsveränderungen mit engeren Räumen in Verbindung gebracht, während Teilnehmer in VU mit mehr Szenenweite geringere Empfindlichkeiten zeigten (s. Abschnitt 3.2). Zudem werden bei mehr Szenenweite höhere *PSE*-Werte für die Translationsgeschwindigkeit g_t gemessen (s. Abschnitt 3.2). Dies deckt sich mit der Verengung/Öffnung der Sichtwinkels (α_v, α_h) in den veränderten Projektionen dieser Arbeit und den unterschiedlichen *PSE* je Brennweite f.

Wichtig anzumerken ist, dass durch eine Anpassung der Translationsverstärkung g_t die dynamische Okklusion (s. Abschnitt 2.2.1) verändert wird. Expansion und Kompression in der Konvergenzebene bleiben unverändert (die Fixationsquadrate bleiben starr an ihrer Position), während Zu-/Aufdecken der variierende Faktor ist. Nur bei Beibehaltung der Translationsverstärkung $g_t = 1$ bleibt die optische Achse identisch und die gleichen Szenenelemente werden mit der gleichen Translation $T_{v_{cam}}$ zu- oder aufgedeckt, unabhängig von der Brennweite f (wenn auch auf der Retina mit doppelter/halber Geschwindigkeit). Das räumliche Verständnis der Teilnehmer und das unterbewusste – vielleicht gar bewusste – Verorten der Objekte in der Szene, könnte Einfluss nehmen: Gerade durch den direkten Vergleich mit der homothetischen Konfiguration, könnten die Teilnehmer die zu erwartenden Verdeckungen durch das Szenenlayout berücksichtigen und so beeinflusst sein. Durchführen zweier Versuche, einmal nur mit veränderter Projektion und einmal mit beiden im Vergleich könnten diesen Effekt näher beleuchten.

Die erhobenen *PSE* Messwerte für f_2 mit $g_t = 0,75$ und für $f_{0.5}$ mit $g_t = 1,51$ ordnen sich zwischen den Extremfällen der Beibehaltung der optischen Achse ($g_t = 1$) und der Beibehaltung der erwarteten Retinalgeschwindigkeit ($g_t = 0,5$, respektive $g_t = 2$) aus der gezeigten Szenentiefe ein. Aus den erhobenen Daten lassen sich keine fundierte Aussagen darüber treffen, wie die Teilnehmer die Stimuli in Wirklichkeit wahrgenommen haben, an welchen Tiefenhinweisen sie sich orientierten, oder wo im Raum sie sich oder die virtuelle Kamera gedanklich positionierten. Die aufgeworfenen Hypothesen und vermuteten Zusammenhänge gilt es in zukünftiger Recherche zu prüfen.

7 Fazit und Ausblick

In dieser Arbeit wurde ein psychophyisches VR-Experiment zur Bestimmung der Wahrnehmung von Motion Parallax in gerahmtem, stereoskopischem 3DoF+ Video theoretisch fundiert entworfen, technisch realisiert und die Messergebnisse statistisch ausgewertet und eingeordnet. Gerahmte Stimuli, die die Projektion des Blickwinkels eines virtuellen Stereo-Kamerasystems auf eine realistische Szene zeigten, wurden den Teilnehmern als schwebende Fenster in einem virtuellen Raum gezeigt. Dabei sind drei Skalierungsstufen der Projektion im Rahmen, der in seiner Größe und Abstand konstant blieb, untersucht worden (halbe Vergrößerung $f_{0.5}$, homothetische Konfiguration f_1 und doppelte Vergrößerung f_2). Die laterale Translation der virtuellen Stereo-Kamera $T_{v_{cam}}$ war an die Translation des Betrachters T_r über die Translationsverstärkung g_t gekoppelt. Motiviert durch eine Suchaufgabe, induzierten die Teilnehmer Motion Parallax, um dann mit einer 2AFC-Methode die *PSE* und $DT_{25\%}/DT_{75\%}$ Wertebereiche für die als natürlich wahrgenommene Translationsübersetzung g_t je Skalierung zu bestimmen.

Die Ergebnisse zeigen, dass die Rahmung einer virtuellen Szene keinen Einfluss auf die darin als natürlich wahrgenommene Translationsverstärkung g_t hat: Der methodische Aufbau des Experiments ist eine Abwandlung der Studie von Serrano et al. (2020a) , der jene primär um die Rahmung der Stimuli erweitert; So konnte bestätigt werden, dass die *PSE*-Werte der gerahmten Stimuli mit den ungerahmten übereinstimmen. Die Schwellwerte $DT_{25\%}/DT_{75\%}$ hingegen sind durch die Rahmung beeinflusst: Es wird eine deutlich erhöhte Sensibilität für Translationsveränderungen festgestellt: So werden ±23 % Abweichung vom *PSE* nicht eindeutig unterschiedlich erkannt, verglichen mit ±39, 1% (Serrano et al., 2020a, S. 6). Das Teilnehmerverhalten änderte sich dadurch, dass die Translationsverstärkung g_t nur die im gerahmten Stimulus sichtbare Translation $T_{v_{cam}}$ betraf – Die Teilnehmer passten die Geschwindigkeit ihrer realen Bewegungen weitaus weniger an die veränderte Translationsübersetzung an.

Eine Skalierung der Projektion bewirkt, dass eine gestauchte/gestreckte Übersetzung als realitätsnah empfunden wird. Bei zweifacher Vergrößerung gegenüber homothetischer Konfiguration wurde eine Stauchung um 26 %, bei halber Vergrößerung eine Streckung um 48 % gemessen. Die Empfindlichkeit für Abweichungen von diesen als natürlich empfundenen Übersetzungen fiel für die vergrößerte Projektion leicht höher, für die verkleinerte Projektion deutlich niedriger aus. Die Messwerte stehen im Widerspruch zur von Hartle

und Wilcox, 2021 beobachteten Veto-Funktion der binokularen Disparität gegenüber Motion Parallax als Tiefenhinweis (S. 61). Die *PSE*-Werte der skalierten Projektionen folgen nicht der aus binokularer Disparität zu schlussfolgernden Retinalgeschwindigkeit $d\theta/dt$. Einen möglichen Erklärungsansatz liefert die Theorie, dass die Teilnehmer statt einer Skalierung eine Distanzveränderung vermuteten.

Die gewonnenen Erkenntnisse leisten einen wichtigen Beitrag zur Erschließung gerahmter Videoformate mit mehr als drei Freiheitsgraden und bringen Implikationen für die Herstellung (Kamera-Array-Geometrie) und Wiedergabe (Größe der viewing baseline) mit sich. Es ist die erste Arbeit zur Wahrnehmung lateraler Übersetzung von Motion Parallax beim Blick in gerahmte und überdies skalierte Szenenprojektionen.

In zukünftiger Forschung zu untersuchen ist im Besonderen der Zusammenhang zwischen Skalierung/Projektion und Motion Parallax Wahrnehmung. Wie wirken andere Betrachtungsabstände und Rahmengrößen? Wird eine Vergrößerung des Bildinhalts als Veränderung der Betrachtungsposition wahrgenommen? Inwiefern gelten bewährte Modelle zur Tiefenwahrnehmung in geometrischen Projektionen, die menschliches Sehvermögen übersteigen? Und welche Wechselwirkungen treten auf, wenn bewegte Szenenelemente gezeigt, oder Kameraführung induziert würde?

Die Relevanz des Forschungsgegenstands wird unter anderem durch die Produkteinführung der Apple Vision Pro (Apple, 2024a) unterstrichen: Wenn die Zukunft der Computer in virtuell projizierten Fenstern liegt, sollten die darin sichtbaren Szenen – auch wenn sie mal nicht dem realitätsgetreuen Blickwinkel der Echtwelt entsprechen – auf natürliche Weise interaktive Darstellungen zeigen.

Literatur

Acer. (2024). *SpatialLabs™ Stereoscopic 3D-Lösungen*. Acer Deutschland. Verfügbar 11. Februar 2024 unter https://www.acer.com/de-de/spatiallabs

Alain, M., Zerman, E., Ozcinar, C., & Valenzise, G. (2023, 1. Januar). Chapter 1 - Introduction to Immersive Video Technologies. In G. Valenzise, M. Alain, E. Zerman & C. Ozcinar (Hrsg.), *Immersive Video Technologies* (S. 3–24). Academic Press. https://doi.org/10.1016/B978-0-32-391755-1.00007-9

Anderson, R., Gallup, D., Barron, J. T., Kontkanen, J., Snavely, N., Hernandez, C., Agarwal, S., & Seitz, S. M. (2016). Jump: Virtual Reality Video. *35*(6), 198. https://doi.org/10.1145/2980179.2980257
MAG ID: 2531849700.

Apple. (2024a). *Apple Vision Pro — Apples erster räumlicher Computer*. Apple Newsroom (Deutschland). Verfügbar 9. Februar 2024 unter https://www.apple.com/de/newsroom/2023/06/introducing-apple-vision-pro/

Apple. (2024b, 19. Januar). *A Guided Tour of Apple Vision Pro*. Verfügbar 7. Februar 2024 unter https://youtu.be/Vb0dG-2huJE?t=184&feature=shared

ASUS. (2024). *ASUS Spatial Vision – Let your creativity shine in 3D*. ASUS Global. Verfügbar 11. Februar 2024 unter https://www.asus.com/content/asus-spatial-vision-technology/

Attal, B., Laidlaw, E., Gokaslan, A., Kim, C., Richardt, C., Tompkin, J., & O'Toole, M. (2021). TöRF: Time-of-Flight Radiance Fields for Dynamic Scene View Synthesis: NeurIPS 2021: Conference on Neural Information Processing Systems. *Advances in Neural Information Processing Systems*, *2021*. Verfügbar 11. Februar 2024 unter https://imaging.cs.cmu.edu/torf/

Bae, K., Ivan, A., Nagahara, H., & Park, I. K. (2019, 23. Dezember). *5D Light Field Synthesis from a Monocular Video*. arXiv: 1912.10687 [cs]. Verfügbar 7. Februar 2024 unter http://arxiv.org/abs/1912.10687

Bertel, T., Xu, F., & Richardt, C. (2020). Image-Based Scene Representations for Head-Motion Parallax in 360° Panoramas. In M. Magnor & A. Sorkine-Hornung (Hrsg.), *Real VR – Immersive Digital Reality: How to Import the Real World into Head-Mounted Immersive Displays* (S. 109–131). Springer International Publishing. https://doi.org/10.1007/978-3-030-41816-8_5

Bonatto, D., Fachada, S., Rogge, S., Munteanu, A., & Lafruit, G. (2021). Real-Time Depth Video-Based Rendering for 6-DoF HMD Navigation and Light Field Displays. *IEEE Access*, *9*, 146868–146887. https://doi.org/10.1109/ACCESS.2021.3123529

Bourke, P. (1999). *Calculating Stereo Pairs*. Calculating Stereo Pairs. Verfügbar 20. Dezember 2023 unter https://paulbourke.net/stereographics/stereorender/

Boyce, J. M., Dore, R., Dziembowski, A., Fleureau, J., Jung, J., Kroon, B., Salahieh, B., Vadakital, V. K. M., & Yu, L. (2021). MPEG Immersive Video Coding Standard. *Proceedings of the IEEE*, *109*(9), 1521–1536. https://doi.org/10.1109/JPROC.2021.3062590

Brenner, E., & Smeets, J. B. J. (2018, 23. März). Depth Perception. In J. T. Wixted (Hrsg.), *Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience* (1. Aufl., S. 1–30). Wiley. https://doi.org/10.1002/9781119170174.epcn209

Broxton, M., Flynn, J., Overbeck, R., Erickson, D., Hedman, P., Duvall, M., Dourgarian, J., Busch, J., Whalen, M., & Debevec, P. (2020). Immersive Light Field Video with a Layered Mesh Representation. *ACM Transactions on Graphics*, *39*(4), 86:86:1–86:86:15. https://doi.org/10.1145/3386569.3392485

Buckthought, A., Yoonessi, A., & Baker, C. L. (2017). Dynamic perspective cues enhance depth perception from motion parallax. *Journal of Vision*, *17*(1), 10. https://doi.org/10.1167/17.1.10

Champel, M.-L., Koenen, R., Lafruit, G., & Budagavi, M. (2018, April). *Proposed Draft 1.0 of TR: Technical Report on Architectures for Immersive Media* (Doc. Nr. N17685). ISO/IEC JTC1/SC29/WG11 MPEG. San Diego. Verfügbar 9. Februar 2024 unter https://mpeg.chiariglione.org/standards/mpeg-i/technical-report-immersive-media/text-pdtr-isoiec-23090-1-immersive-media

Chelli, K., Lange, T., Herfet, T., de, Solony, M., & Smrz, P. (2020). A Versatile 5D Light Field Capturing Array. Verfügbar 7. Februar 2024 unter https://api.semanticscholar.org/CorpusID:246821950

Choy, S.-M., Cheng, E., Wilkinson, R. H., Burnett, I., & Austin, M. W. (2021). Quality of Experience Comparison of Stereoscopic 3D Videos in Different Projection Devices: Flat Screen, Panoramic Screen and Virtual Reality Headset. *IEEE Access*, *9*, 9584–9594. https://doi.org/10.1109/ACCESS.2021.3049798

Cutting, J. E., & Vishton, P. M. (1995, 1. Januar). Chapter 3 - Perceiving Layout and Knowing Distances: The Integration, Relative Potency, and Contextual Use of Different Information about Depth*. In W. Epstein & S. Rogers (Hrsg.), *Perception of Space and Motion* (S. 69–117). Academic Press. https://doi.org/10.1016/B978-012240530-3/50005-5

David, E. J., Lebranchu, P., Perreira Da Silva, M., & Le Callet, P. (2022). What are the visuomotor tendencies of omnidirectional scene free-viewing in virtual reality? *Journal of Vision*, *22*(4), 12. https://doi.org/10.1167/jov.22.4.12

Devernay, F., & Beardsley, P. (2010, 30. Mai). Stereoscopic Cinema. In *Image and Geometry Processing for 3-D Cinematography*. https://doi.org/10.1007/978-3-642-12392-4_2

Durgin, F. H., Proffitt, D. R., Olson, T. J., & Reinke, K. S. (1995). Comparing depth from motion with depth from binocular disparity. *Journal of Experimental Psychology: Human Perception and Performance*, *21*(3), 679–699. https://doi.org/10.1037/0096-1523.21.3.679

Flynn, J., Michael Broxton, Broxton, M., Debevec, P., Paul Debevec, DuVall, M., Fyffe, G., Overbeck, R., Snavely, N., & Tucker, R. (2019). DeepView: View Synthesis With Learned Gradient Descent, 2367–2376. https://doi.org/10.1109/cvpr.2019.00247
MAG ID: 2949657144.

Follows, S. (2017, 20. November). *Are audiences tiring of 3D movies?* Stephen Follows. Verfügbar 11. Februar 2024 unter https://stephenfollows.com/audiences-tiring-of-3d-movies/

French, R. L., & DeAngelis, G. C. (2022). Scene-relative object motion biases depth percepts. *Scientific Reports*, *12*(1), 18480. https://doi.org/10.1038/s41598-022-23219-4

Fulvio, J. M., Miao, H., & Rokers, B. (2021). Head Jitter Enhances Three-Dimensional Motion Perception. *Journal of Vision*, *21*(3), 12. https://doi.org/10.1167/jov.21.3.12

Gardner, B. R. (2011, 10. Februar). The Dynamic Floating Window: A new creative tool for 3D movies. In A. J. Woods, N. S. Holliman & N. A. Dodgson (Hrsg.). https://doi.org/10.1117/12.872608

Gjestland, R. (2022, Oktober). *How to design a cinema auditorium*. Union Internationale des Cinémas. Verfügbar 11. Februar 2024 unter https://www.unic-cinemas.org/fileadmin/user_upload/Publications/2022/UNIC_handbook_online_Okt22_.pdf

Hartle, B., & Wilcox, L. M. (2021). Cue vetoing in depth estimation: Physical and virtual stimuli. *Vision Research*, *188*, 51–64. https://doi.org/10.1016/j.visres.2021.07.003

Herfet, T., Chelli, K., & Le Pendu, M. (2023, 1. Januar). Chapter 7 - Light Field Representation: The Dimensions in Light Fields. In G. Valenzise, M. Alain, E. Zerman & C. Ozcinar (Hrsg.), *Immersive Video Technologies* (S. 173–199). Academic Press. https://doi.org/10.1016/B978-0-32-391755-1.00013-4

Hoffman, D. M., Girshick, A. R., Akeley, K., & Banks, M. S. (2008). Vergence-Accommodation Conflicts Hinder Visual Performance and Cause Visual Fatigue. *Journal of Vision*, *8*(3), 33–33. https://doi.org/10.1167/8.3.33
MAG ID: 1991012411.

Holmin, J., & Nawrot, M. (2015). Motion parallax thresholds for unambiguous depth perception. *Vision Research*, *115*, 40–47. https://doi.org/10.1016/j.visres.2015.07.002

IJsselsteijn, W. A., Oosting, W., Vogels, I. M. L. C., De Kort, Y. A. W., & Van Loenen, E. (2008). A Room with a Cue: The Efficacy of Movement Parallax, Occlusion, and Blur in Creating a Virtual Window. *Presence: Teleoperators and Virtual Environments*, *17*(3), 269–282. https://doi.org/10.1162/pres.17.3.269

ITU. (2023, Mai). *Methodologies for the subjective assessment of the quality of television images*. Verfügbar 11. Februar 2024 unter https://www.itu.int/rec/R-REC-BT.500-15-202305-I/en

James J. Gibson. (1950). *The perception of the visual world*. Houghton Mifflin. Verfügbar 2. Februar 2024 unter http://archive.org/details/perceptionofvisu00jame

Jeong, J.-B., Lee, S., Jang, D., & Ryu, E.-S. (2019). Towards 3DoF+ 360 Video Streaming System for Immersive Media. *IEEE Access*, *7*, 136399–136408. https://doi.org/10.1109/ACCESS.2019.2942771

Jin, C., Peng, Z., Chen, F., & Jiang, G. (2022). Subjective and Objective Video Quality Assessment for Windowed-6DoF Synthesized Videos. *IEEE Transactions on Broadcasting*, *68*(3), 594–608. https://doi.org/10.1109/TBC.2022.3165473

John Carmack [@ID_AA_Carmack]. (2021, 15. März). *The design focal distance for the Quest/Quest 2 optics is 1.3 meters. Some older headsets were 2.0 meters, and I have been saying that incorrectly for a while.* Twitter. Verfügbar 21. Januar 2024 unter https://twitter.com/ID_AA_Carmack/status/1371485209603022853

Keinert, J., Fink, L., Goldmann, F., Gul, M. S. K., Jaschke, T., Prappacher, N., Ziegler, M., Bätz, M., & Fößel, S. (2023, 1. Januar). Chapter 9 - Light Field Processing for Media Applications. In G. Valenzise, M. Alain, E. Zerman & C. Ozcinar (Hrsg.), *Immersive Video Technologies* (S. 227–264). Academic Press. https://doi.org/10.1016/B978-0-32-391755-1.00015-8

Kellnhofer, P., Didyk, P., Ritschel, T., Masia, B., Myszkowski, K., & Seidel, H.-P. (2016). Motion Parallax in Stereo 3D: Model and Applications. *35*(6), 176. https://doi.org/10.1145/2980179.2980230
MAG ID: 2557137016.

Kim, D., Kim, J., Shin, J.-E., Yoon, B., Lee, J., & Woo, W. (2022). Effects of Virtual Room Size and Objects on Relative Translation Gain Thresholds in Redirected Walking. *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 379–388. https://doi.org/10.1109/VR51125.2022.00057

Kim, D., Shin, J.-e., Lee, J., & Woo, W. (2021). Adjusting Relative Translation Gains According to Space Size in Redirected Walking for Mixed Reality Mutual Space Generation. *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*, 653–660. https://doi.org/10.1109/VR50410.2021.00091

Kollenberg, T., Neumann, A., Schneider, D., Tews, T.-K., Hermann, T., Ritter, H., Dierker, A., & Koesling, H. (2010). Visual Search in the (Un)Real World: How Head-Mounted Displays Affect Eye Movements, Head Movements and Target Detection. In C.H. Morimoto & I. Howell (Hrsg.), *Proceedings of the Eye Tracking Research & Applications Symposium* (S. 121–124). ACM. https://doi.org/10.1145/1743666.1743696

Kongsilp, S., & Dailey, M. N. (2017). Motion parallax from head movement enhances stereoscopic displays by improving presence and decreasing visual fatigue. *Displays*, *49*, 72–79. https://doi.org/10.1016/j.displa.2017.07.001

Kongsilp, S., & Dailey, M. N. (2018). User Behavior and the Importance of Stereo for Depth Perception in Fish Tank Virtual Reality. *Presence: Teleoperators and Virtual Environments*, *27*(2), 206–225. https://doi.org/10.1162/pres_a_00327

Langbehn, E., Steinicke, F., Lappe, M., Welch, G., Welch, G. F., & Bruder, G. (2018). In the Blink of an Eye: Leveraging Blink-Induced Suppression for Imperceptible Position and Orientation Redirection in Virtual Reality. *37*(4), 66. https://doi.org/10.1145/

3197517.3201335
MAG ID: 2810558896.

LaViola, J. J. (2000). A discussion of cybersickness in virtual environments. *ACM SIGCHI Bulletin*, *32*(1), 47–56. https://doi.org/10.1145/333329.333344

Le Meur, O., & Liu, Z. (2015). Saccadic model of eye movements for free-viewing condition. *Vision Research*, *116*, 152–164. https://doi.org/10.1016/j.visres.2014.12.026

Liu, J., Zhong, F., Mantel, C., Forchhammer, S., & Mantiuk, R. K. (2023, 1. Januar). Chapter 17 - Computational 3D Displays. In G. Valenzise, M. Alain, E. Zerman & C. Ozcinar (Hrsg.), *Immersive Video Technologies* (S. 469–500). Academic Press. https://doi.org/10.1016/B978-0-32-391755-1.00023-7

Liu, S., Kersten, D. J., & Legge, G. E. (2023). Effect of expansive optic flow and lateral motion parallax on depth estimation with normal and artificially reduced acuity. *Journal of Vision*, *23*(12), 3. https://doi.org/10.1167/jov.23.12.3

Maniatis, L. (2021, 30. März). *The myth of visual depth cues V: Motion parallax*. https://doi.org/10.31234/osf.io/dwf5t

Marino, M. J. (2018). Chapter 3 - Statistical Analysis in Preclinical Biomedical Research. In M. Williams, M. J. Curtis & K. Mullane (Hrsg.), *Research in the Biomedical Sciences* (S. 107–144). Academic Press. https://doi.org/10.1016/B978-0-12-804725-5.00003-3

Meta. (2024). *Meta Quest 3: New mixed reality VR headset – Shop now | Meta Store*. Verfügbar 9. Februar 2024 unter https://www.meta.com/de/en/quest/quest-3/

Milliron, T., Szczupak, C., & Green, O. (2017). Hallelujah: The world's first lytro VR experience. *ACM SIGGRAPH 2017 VR Village*, 1–2. https://doi.org/10.1145/3089269.3089283

Miyashita, Y., Sawahata, Y., Sakai, A., Harasawa, M., Hara, K., Morita, T., & Komine, K. (2022). Display-Size Dependent Effects of 3D Viewing on Subjective Impressions. *ACM Transactions on Applied Perception*, *19*(2), 1–15. https://doi.org/10.1145/3510461

Murgia, A., & Sharkey, P. M. (2009). Estimation of Distances in Virtual Environments Using Size Constancy. *International Journal of Virtual Reality*, *8*(1), 67–74. https://doi.org/10.20870/IJVR.2009.8.1.2714

Nawrot, M., & Joyce, L. (2006). The pursuit theory of motion parallax. *Vision Research*, *46*(28), 4709–4725. https://doi.org/10.1016/j.visres.2006.07.006

Nawrot, M., Ratzlaff, M., Leonard, Z., Stroyan, K. D., & Stroyan, K. (2014). Modeling Depth from Motion Parallax with the Motion/Pursuit Ratio. *Frontiers in Psychology*, *5*, 1103–1103. https://doi.org/10.3389/fpsyg.2014.01103
MAG ID: 1981619594.

Nawrot, M., & Stroyan, K. (2009). The motion/pursuit law for visual depth perception from motion parallax. *Vision Research*, *49*(15), 1969–1978. https://doi.org/10.1016/j.visres.2009.05.008

Nielsen, J. I. (2007). *Camera Movement in Narrative Cinema: Towards a Taxonomy of Functions*. Department of Inf. & Media Studies, University of Aarhus. https://pure.au.dk/ws/files/52113417/Camera_Movement_0910.pdf

Pagel, R. (2019). The concept of (depth) cues: An exemplification of homuncular language in vision science. *Theory & Psychology*, *29*(1), 66–86. https://doi.org/10.1177/0959354318810184

Rai, Y., Le Callet, P., & Cheung, G. (2016). Quantifying the relation between perceived interest and visual salience during free viewing using trellis based optimization. *2016 IEEE 12th Image, Video, and Multidimensional Signal Processing Workshop (IVMSP)*, 1–5. https://doi.org/10.1109/IVMSPW.2016.7528228

Richardt, C., Christian Richardt, Tompkin, J., & Wetzstein, G. (2020). Capture, Reconstruction, and Representation of the Visual Real World for Virtual Reality, 3–32. https://doi.org/10.1007/978-3-030-41816-8_1
MAG ID: 3011969621.

Rogers, B. (2022). Cues, clues and the cognitivisation of perception: Do words matter? *Perception*, *51*(5), 295–299. https://doi.org/10.1177/03010066221080617

Rogers, B., & Graham, M. (1979). Motion Parallax as an Independent Cue for Depth Perception. *Perception*, *8*(2), 125–134. https://doi.org/10.1068/p080125

Selzer, M. N., Larrea, M. L., & Castro, S. M. (2022). Analysis of translation gains in virtual reality: The limits of space manipulation. *Virtual Reality*, *26*(4), 1459–1469. https://doi.org/10.1007/s10055-022-00640-8

Serrano, A., Kim, I., Chen, Z., DiVerdi, S., Gutierrez, D., Hertzmann, A., & Masia, B. (2019). Motion parallax for 360° RGBD video. *IEEE Transactions on Visualization and Computer Graphics*, *25*(5), 1817–1827. https://doi.org/10.1109/TVCG.2019.2898757

Serrano, A., Martin, D., Gutierrez, D., Myszkowski, K., & Masia, B. (2020a). Imperceptible Manipulation of Lateral Camera Motion for Improved Virtual Reality Applications. *ACM Transactions on Graphics*, *39*(6), 267:1–267:14. https://doi.org/10.1145/3414685.3417773

Serrano, A., Martin, D., Gutierrez, D., Myszkowski, K., & Masia, B. (2020b, Dezember). Supplementary material: Imperceptible manipulation of lateral camera motion for improved virtual reality applications. Verfügbar 21. Dezember 2023 unter https://ana-serrano.github.io/projects/VR-LateralMotion

Sitzmann, V., Serrano, A., Pavel, A., Agrawala, M., Gutierrez, D., Masia, B., & Wetzstein, G. (2018). Saliency in VR: How Do People Explore Virtual Environments? *IEEE Transactions on Visualization and Computer Graphics*, *24*(4), 1633–1642. https://doi.org/10.1109/TVCG.2018.2793599

Smith, B. M., Zhang, L., Jin, H., & Agarwala, A. (2009). Light Field Video Stabilization, 341–348. https://doi.org/10.1109/iccv.2009.5459270
MAG ID: 2112673789.

Song, A., Frank, J., Hernandez Zaragoza, J. C., Green, O., Cooper, S., Braunstein, A., Milliron, T., Pitts, C., Yasui, Y., Shahhosseini, S., & Zhang, B. (2017). *Wedge-based light-field video capture* (US-Pat. Nr. 10,275,898). Verfügbar 20. November 2023 unter https://uspto.report/patent/grant/10,275,898

Steinicke, F., Bruder, G., Jerald, J., Frenz, H., & Lappe, M. (2010). Estimation of Detection Thresholds for Redirected Walking Techniques. *IEEE Transactions on Visualization and Computer Graphics*, *16*(1), 17–27. https://doi.org/10.1109/TVCG.2009.62

Stelmach, L. B., Tam, W. J., & Meegan, D. V. (1999, 24. Mai). Perceptual basis of stereoscopic video. In J. O. Merritt, M. T. Bolas & S. S. Fisher (Hrsg.). https://doi.org/10.1117/12.349387

Szita, K., Moss-Wellington, W., Sun, X., & Ch'ng, E. (2024). Going to the movies in VR: Virtual reality cinemas as alternatives to in-person co-viewing. *International Journal of Human-Computer Studies*, *181*, 103150. https://doi.org/10.1016/j.ijhcs.2023.103150

Tam, W. J., Stelmach, L. B., & Corriveau, P. J. (1998, 30. April). Psychovisual aspects of viewing stereoscopic video sequences. In M. T. Bolas, S. S. Fisher & J. O. Merritt (Hrsg.). https://doi.org/10.1117/12.307169

Teng, X., Allison, R. S., & Wilcox, L. M. (2023). Manipulation of Motion Parallax Gain Distorts Perceived Distance and Object Depth in Virtual Reality. *2023 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*, 398–408. https://doi.org/10.1109/VR55154.2023.00055

Thatte, J., & Girod, B. (2018). Towards Perceptual Evaluation of Six Degrees of Freedom Virtual Reality Rendering from Stacked OmniStereo Representation. *Electronic Imaging*, *30*(5), 352-1-352–6. https://doi.org/10.2352/ISSN.2470-1173.2018.05.PMII-352

Thatte, J., & Girod, B. (2021). Real-World Virtual Reality With Head-Motion Parallax. *IEEE Computer Graphics and Applications*, *41*(4), 29–39. https://doi.org/10.1109/mcg.2021.3082041

MAG ID: 3161599457.

Thatte, J., Lian, T., Wandell, B., & Girod, B. (2017). Stacked Omnistereo for virtual reality with six degrees of freedom. *2017 IEEE Visual Communications and Image Processing (VCIP)*, 1–4. https://doi.org/10.1109/VCIP.2017.8305085

Ukai, K., & Howarth, P. A. (2008). Visual fatigue caused by viewing stereoscopic motion images: Background, theories, and observations. *Displays*, *29*(2), 106–116. https://doi.org/10.1016/j.displa.2007.09.004

Van Der Hooft, J., Amirpour, H., Vega, M. T., Sanchez, Y., Schatz, R., Schierl, T., & Timmerer, C. (2023, Sommer). A Tutorial on Immersive Video Delivery: From Omnidirectional Video to Holography. *IEEE Communications Surveys & Tutorials*, *25*(2), 1336–1375. https://doi.org/10.1109/COMST.2023.3263252

Vienne, C., Masfrand, S., Bourdin, C., & Vercher, J.-L. (2020). Depth Perception in Virtual Reality Systems: Effect of Screen Distance, Environment Richness and Display Factors. *IEEE Access*, *8*, 29099–29110. https://doi.org/10.1109/ACCESS.2020.2972122

von Helmholtz, H., & Nagel, W. (1910). *Handbuch Der Physiologischen Optik: Die Lehre von Den Gesichtswahrehmungen, Hrsg. von J. von Kries*. L. Voss. https://books.google.de/books?id=Sc45AQAAMAAJ

Wilburn, B., Smulski, M., Lee, H.-H. K., & Horowitz, M. (2001). Light Field Video Camera. *electronic imaging*, *4674*, 29–36. https://doi.org/10.1117/12.451074
MAG ID: 2091663372.

Wu, G., Masia, B., Jarabo, A., Yuchen Zhang, Zhang, Y., Wang, L., Dai, Q., Chai, T., Chai, T., & Liu, Y. (2017). Light Field Image Processing: An Overview. *IEEE Journal of Selected Topics in Signal Processing*, *11*(7), 926–954. https://doi.org/10.1109/jstsp.2017.2747126
MAG ID: 2753630964 S2ID: ed7038b052af85886bf96c3511910ad654e1811a.

Yan, J., Li, J., Fang, Y., Che, Z., Xia, X., & Liu, Y. (2022). Subjective and Objective Quality of Experience of Free Viewpoint Videos. *IEEE Transactions on Image Processing*, *31*, 3896–3907. https://doi.org/10.1109/TIP.2022.3177127

Yoonessi, A., & Baker, C. L. (2011). Contribution of motion parallax to segmentation and depth perception. *Journal of Vision*, *11*(9), 13–13. https://doi.org/10.1167/11.9.13

Yoonessi, A., & Baker, C. L. (2013). Depth perception from dynamic occlusion in motion parallax: Roles of expansion-compression versus accretion-deletion. *Journal of Vision*, *13*(12), 10–10. https://doi.org/10.1167/13.12.10

Zellmann, S., & Amstutz, J. (2023, 18. November). *A Practical Guide to Implementing Off-Axis Stereo Projection Using Existing Ray Tracing Libraries*. arXiv: 2311.05887 [cs]. Verfügbar 9. Februar 2024 unter http://arxiv.org/abs/2311.05887

Zhang, J., Tianyi, Z., Zhang, A., Xiaoyun Yuan, Yuan, X., Wang, Z., Wang, Z., Beetschen, S., Xu, L., Lin, X., Dai, Q., & Fang, L. (2020). Multiscale-VR: Multiscale Gigapixel 3D Panoramic Videography for Virtual Reality, 1–12. https://doi.org/10.1109/iccp48838.2020.9105244
MAG ID: 3033011849.

Zhang, J., Langbehn, E., Krupke, D., Katzakis, N., & Steinicke, F. (2018). Detection Thresholds for Rotation and Translation Gains in 360° Video-Based Telepresence Systems. *IEEE Transactions on Visualization and Computer Graphics*, *24*(4), 1671–1680. https://doi.org/10.1109/TVCG.2018.2793679

Zhang, Y., Liu, Q., & Wang, Y. (2022). Redirected Walking in 360° Video: Effect of Environment Size on Detection Thresholds for Translation and Rotation Gains. *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, 830–831. https://doi.org/10.1109/VRW55335.2022.00266

Zou, W., Peng, Z., Ma, X., Chen, F., & Jiang, G. (2022). Subjective quality assessment of synthesized videos for windowed six degrees of freedom video system. *Electronics Letters*, *58*(17), 645–647. https://doi.org/10.1049/ell2.12554

Anhang

GeoGebra Online-Tool

Das erstellte Online-Tool zur Simulation des Projektionsverhaltens liegt als Datei „geogebra-tool.ggb“ auf dem USB-Stick. Es kann in den GeoGebra Grafikrechner geladen und verwendet werden, um eigene Rechenbeispiele zu bemessen. In Abschnitt 2.4 wurden Aufzeichnung und Betrachtung der Übersicht halber getrennt dargestellt. Da im Fall des Experiments stets $H' = H$ galt, werden beide überlagert gezeigt (in der homothetischen Konfiguration sind sie schließlich äquivalent).

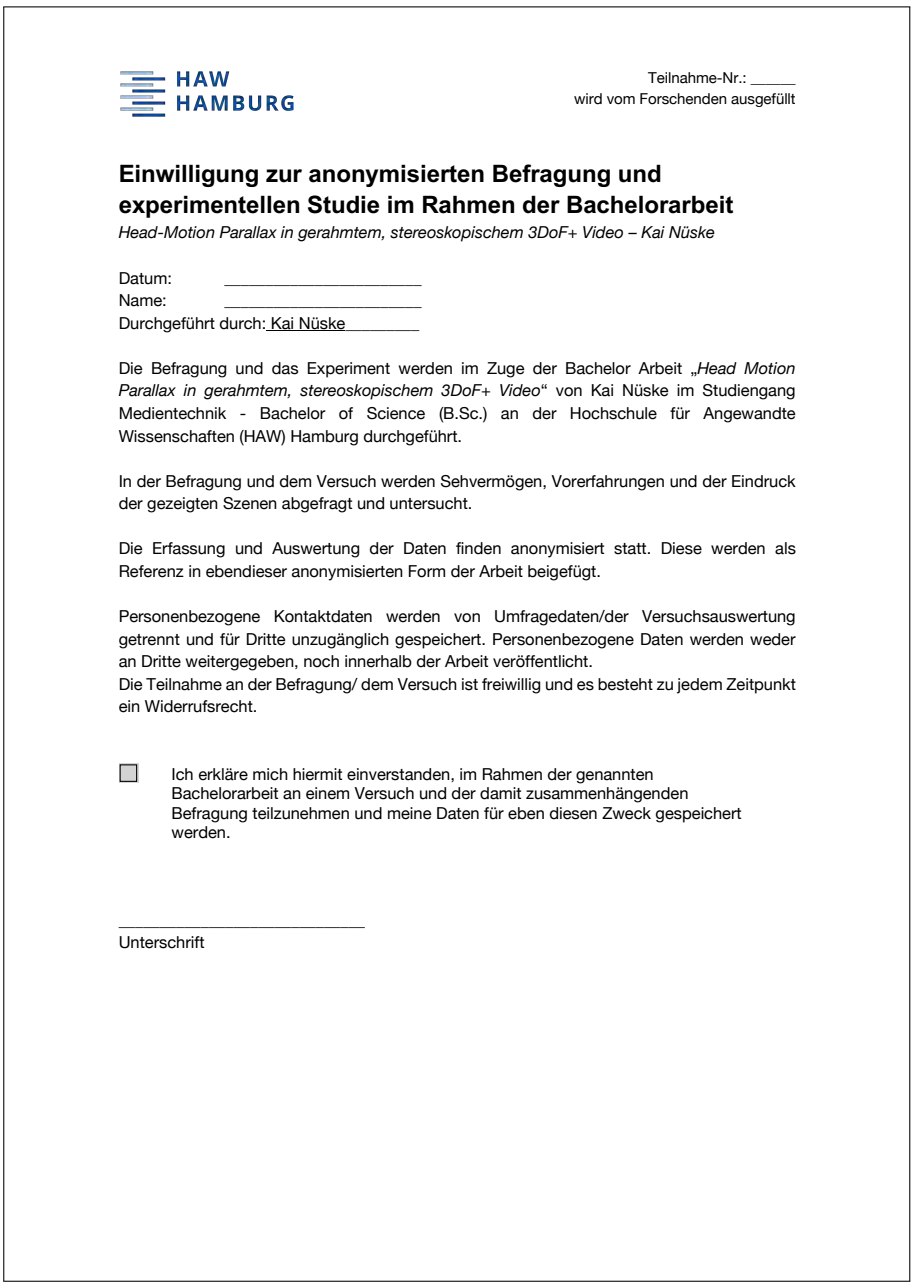
Eingabeparameter

- f_1: Einstellung der homothetischen Brennweite in Ursprungsposition in Millimetern
- T_r: Einstellung der realen, lateralen Auslenkung des Betrachters in Metern
- H: Einstellung des Abstands der Konvergenzebene. Verändert stets Betrachtungsdistanz H' zum gleichen Wert (Bedingung des Experiments).
- g_t: Einstellung der Translationsverstärkung
- s_f: Einstellung des Skalierungsfaktors ($f_{s_f} = s_f \cdot f_1$)
- H_{C*}: Einstellung eines Versatzes der Kamera vom Ursprungspunkt aus (vor/zurück)
- P_1: Koordinaten des hinsichtlich seiner Projektion zu untersuchenden Punkts P_1 (sind im Seitenmenü vorzunehmen; Koordinaten in Metern)

Ausgabeparameter

- α: Verfolgungswinkel
- θ_{proj}: Retinalwinkel zur Projektion $P_{1,proj}$ des Punkts P_1 aus der homothetischen Konfiguration $P_{1,proj*}$
- θ_{proj*}: Retinalwinkel zur Projektion $P_{1,proj*}$ des Punkts P_1 aus der mit {} eingestellten Konfiguration $P_{1,proj*}$

Datenschutz-Formular



HAW HAMBURG

Teilnahme-Nr.: _____
wird vom Forschenden ausgefüllt

Einwilligung zur anonymisierten Befragung und experimentellen Studie im Rahmen der Bachelorarbeit

Head-Motion Parallax in gerahmtem, stereoskopischem 3DoF+ Video – Kai Nüske

Datum: _____
Name: _____
Durchgeführt durch: Kai Nüske

Die Befragung und das Experiment werden im Zuge der Bachelor Arbeit „*Head Motion Parallax in gerahmtem, stereoskopischem 3DoF+ Video*“ von Kai Nüske im Studiengang Medientechnik - Bachelor of Science (B.Sc.) an der Hochschule für Angewandte Wissenschaften (HAW) Hamburg durchgeführt.

In der Befragung und dem Versuch werden Sehvermögen, Vorerfahrungen und der Eindruck der gezeigten Szenen abgefragt und untersucht.

Die Erfassung und Auswertung der Daten finden anonymisiert statt. Diese werden als Referenz in ebendieser anonymisierten Form der Arbeit beigefügt.

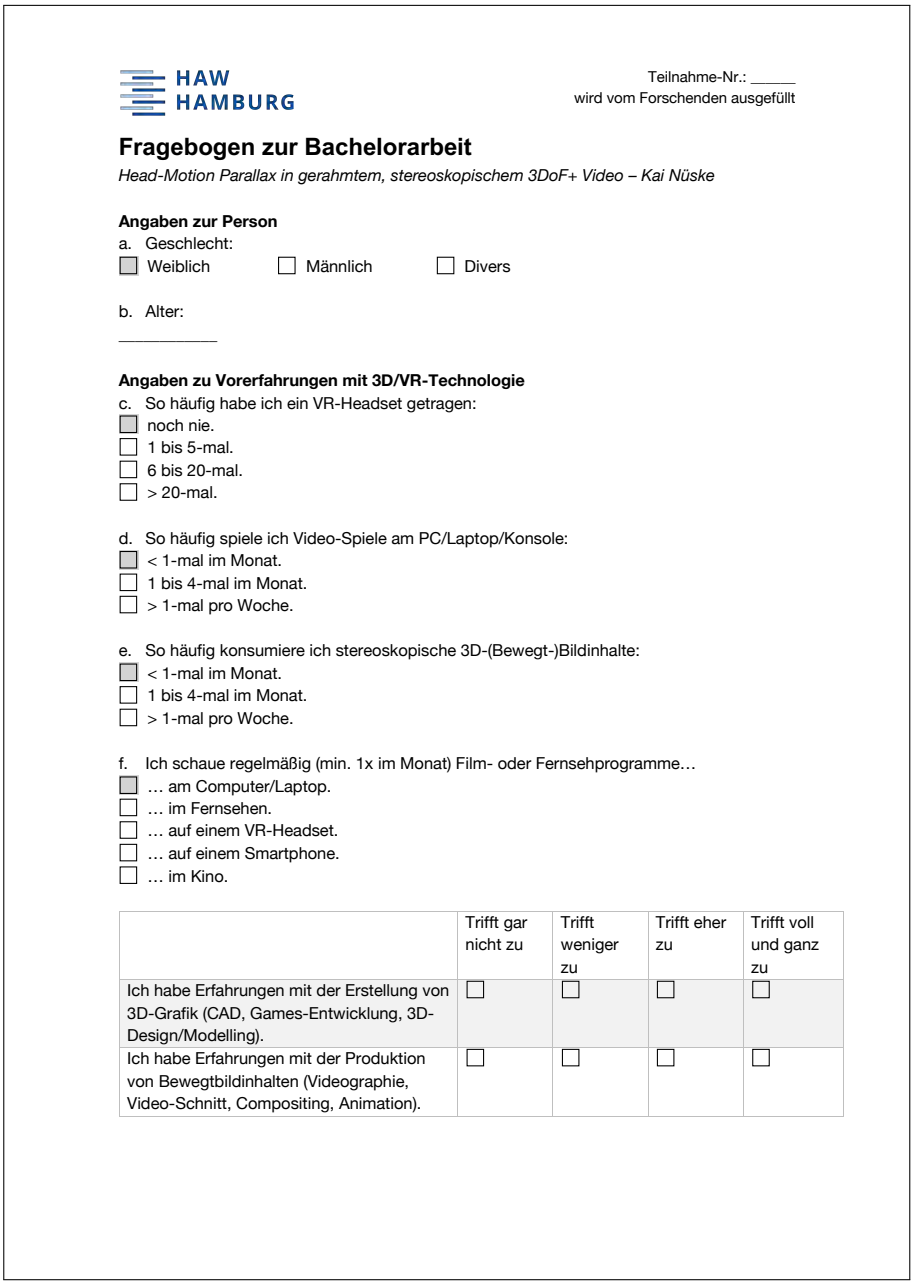
Personenbezogene Kontaktdaten werden von Umfragedaten/der Versuchsauswertung getrennt und für Dritte unzugänglich gespeichert. Personenbezogene Daten werden weder an Dritte weitergegeben, noch innerhalb der Arbeit veröffentlicht.
Die Teilnahme an der Befragung/ dem Versuch ist freiwillig und es besteht zu jedem Zeitpunkt ein Widerrufsrecht.

☐ Ich erkläre mich hiermit einverstanden, im Rahmen der genannten Bachelorarbeit an einem Versuch und der damit zusammenhängenden Befragung teilzunehmen und meine Daten für eben diesen Zweck gespeichert werden.

_____
Unterschrift

Abbildung 1: Datenschutzformular, das vor dem Experiment ausgefüllt wurde.

Fragebogen



HAW HAMBURG

Teilnahme-Nr.: _____
wird vom Forschenden ausgefüllt

Fragebogen zur Bachelorarbeit

Head-Motion Parallax in gerahmtem, stereoskopischem 3DoF+ Video – Kai Nüske

Angaben zur Person

a. Geschlecht:
☐ Weiblich ☐ Männlich ☐ Divers

b. Alter:
_____

Angaben zu Vorerfahrungen mit 3D/VR-Technologie

c. So häufig habe ich ein VR-Headset getragen:
☐ noch nie.
☐ 1 bis 5-mal.
☐ 6 bis 20-mal.
☐ > 20-mal.

d. So häufig spiele ich Video-Spiele am PC/Laptop/Konsole:
☐ < 1-mal im Monat.
☐ 1 bis 4-mal im Monat.
☐ > 1-mal pro Woche.

e. So häufig konsumiere ich stereoskopische 3D-(Bewegt-)Bildinhalte:
☐ < 1-mal im Monat.
☐ 1 bis 4-mal im Monat.
☐ > 1-mal pro Woche.

f. Ich schaue regelmäßig (min. 1x im Monat) Film- oder Fernsehprogramme…
☐ … am Computer/Laptop.
☐ … im Fernsehen.
☐ … auf einem VR-Headset.
☐ … auf einem Smartphone.
☐ … im Kino.

	Trifft gar nicht zu	Trifft weniger zu	Trifft eher zu	Trifft voll und ganz zu
Ich habe Erfahrungen mit der Erstellung von 3D-Grafik (CAD, Games-Entwicklung, 3D-Design/Modelling).	☐	☐	☐	☐
Ich habe Erfahrungen mit der Produktion von Bewegtbildinhalten (Videographie, Video-Schnitt, Compositing, Animation).	☐	☐	☐	☐

Abbildung 2: Fragebogen, der vor dem Experiment ausgefüllt wurde. Die grauen Kästchen sind Resultat einer fehlerhaften PDF-Ausgabe.

Eigenständigkeitserklärung

Hiermit versichere ich, dass ich die vorliegende Bachelorarbeit mit dem Titel

Head-Motion Parallax in gerahmtem, stereoskopischem 3DoF+ Video

selbstständig und nur mit den angegebenen Hilfsmitteln verfasst habe. Alle Passagen, die ich wörtlich aus der Literatur oder aus anderen Quellen wie z. B. Internetseiten übernommen habe, habe ich deutlich als Zitat mit Angabe der Quelle kenntlich gemacht.

Hamburg, 09. Februar 2024