



Hochschule für Angewandte Wissenschaften Hamburg
Hamburg University of Applied Sciences

Bachelorarbeit

Sven Bartel

Implementierung von Virtual Reality Funktionalitäten in eine
virtuelle Simulationsumgebung

Sven Bartel

Implementierung von Virtual Reality Funktionalitäten
in eine virtuelle Simulationsumgebung

Bachelorarbeit eingereicht im Rahmen der Bachelorprüfung

im Studiengang Angewandte Informatik
am Department Informatik
der Fakultät Technik und Informatik
der Hochschule für Angewandte Wissenschaften Hamburg

Betreuender Prüfer : Prof. Dr. Philipp Jenke
Zweitgutachter : Prof. Dr. Andreas Meisel

Abgegeben am 14 August 2014

Sven Bartel

Thema der Arbeit

Implementierung von Virtual Reality Funktionalitäten in eine virtuelle Simulationsumgebung

Stichworte

Virtuelle Realität, Probleme, Lösungen, Definition, optische Verfolgungsverfahren, markenbasierte Verfahren, markenlose Verfahren, Navigation in virtuellen Welten, Positionsbestimmung, Rotationsbestimmung, geschlossene Räume, Tiefenkamera, 3D-Video-brille, Raumwahrnehmung

Kurzzusammenfassung

Die Bachelorarbeit untersucht mögliche Erweiterungen für eine bestehende virtuelle Simulationsumgebung hinsichtlich der „Virtuellen Realität“. Hierzu wird eine 3D-Video-brille zum Einsatz kommen, dessen Verwendung Probleme mit sich bringt. Diese Probleme werden vorgestellt und mögliche Lösungsstrategien aufgezeigt. Zusätzlich wird ein System für die Positionsbestimmung in geschlossenen Räumen gesucht. Hierzu werden zwei Produktlösungen vorgestellt, die in einer Marktanalyse ermittelt wurden. Anschließend wird ein Prototyp zur Positionsbestimmung in geschlossenen Räumen geplant, entworfen und implementiert. Der Prototyp wird mit den vorgestellten Produktlösungen verglichen. Zum Schluss wird eine Auswahl eines für die Simulationsumgebung geeigneten Systems getroffen.

Sven Bartel

Title of the paper

Implementation of Virtual Reality capabilities in a virtual simulation environment

Keywords

virtual reality, problems, solutions, definition, optical tracking techniques, markerbased tracking, markerless tracking, navigation in virtual environments, positional tracking, rotational tracking, indoor tracking solutions, depth camera, head mounted display, spatial perception

Abstract

This bachelor thesis analyses possible extensions for an existing virtual simulation environment with regard to Virtual Reality. For this purpose a head-mounted display will be used, which usage raises some problems. These problems will be introduced and some solutions will be shown. Additionally a system for indoor positional tracking has to be found. For this purpose two professional systems will be introduced, which were found on the market. After that a prototype for indoor positional tracking will be planned, designed and implemented. This prototype will be compared with these presented systems. Finally a system for indoor positional tracking will be chosen.

Glossar

A

Avatar

ist ein grafischer Stellvertreter einer echten Person in einer virtuellen Welt.

C

CAD-Modell

ist ein mit Hilfe eines CAD-Systems entworfenes dreidimensionales Modell eines Objektes

CAD-System

CAD-Systeme sind komplexe Expertensysteme für den Entwurf und die Konstruktion technischer Lösungen.
Z.B. CATIA oder AutoCAD

D

Disparität

bezeichnet in der Stereografie den horizontalen Abstand gleicher Bildelemente auf beiden Teilbildern zueinander und ist damit die Abbildung der Parallaxe.

E

egozentrische Kameraperspektive

hierbei erfolgt die Darstellung der virtuellen Welt so, als würde sie, wie in der echten Welt, mit den eigenen Augen wahrgenommen werden. Diese Kameraperspektive ermöglicht einen hohen Grad der Immersion, da scheinbar selbst die Position in der virtuellen Welt eingenommen wird.

EMM-Check

die Software der REKNOW GmbH & Co. KG, die um VR Funktionen erweitert werden soll.

Extrapolation

ist die Bestimmung eines Verhaltens über den gesicherten Bereich hinaus (Vorhersage, Hochrechnung).

F

FAAST

siehe Abschnitt 4.3

Finger-Tracking

analog zu Head-Tracking

Flystick

ist ein Eingabegerät zur Steuerung von VR Anwendungen, das die Funktion eines 3D-Joystick hat, mit dem der Nutzer Positionen und Orientierungen im 3D-Raum angeben kann.

G

geometrischer Sichtbereich

bezeichnet den von der virtuellen Szene dargestellten Bereich, der durch den horizontalen und vertikalen Öffnungswinkel der virtuellen Kamera definiert wird.

GPS

Globales Positionsbestimmungssystem

H

Head-Mounted Display

ist ein auf dem Kopf getragenes visuelles Ausgabegerät. Es präsentiert die Bilder auf einem augennahen Bildschirm. Einige dieser Geräte verfügen noch über Sensoren zur Bewegungserfassung des Kopfes.

Head-Tracking

ist ein Motion-Capture Verfahren (siehe Motion-Capture) zur Erfassung der Position, Lage und Bewegung des Kopfes, um eine blickpunktabhängige Bildgenerierung zu ermöglichen oder eine andere mit dem Kopf erfolgende Steuerung bereitzustellen.

HMD *Siehe* Head-Mounted Display

I

ImmerSight

siehe Abschnitt 3.3.1

Immersion

Definition siehe Abschnitt 2.1.1

immersiv

siehe Immersion

Inside-Out

siehe Abschnitt 2.4.1.6

J

JAR

Java Archive, ein Dateiformat für Java-Programme

L

LED

zu Deutsch Licht-emittierende Diode

M

Motion-Capture

zu Deutsch Bewegungs-Erfassung, ist ein Tracking Verfahren (siehe Tracking), mit dem es möglich wird, jede Art von Bewegungen so zu erfassen, dass sie in ein von Computern lesbares Format umgewandelt werden können. Dadurch können die Bewegungen analysiert, aufgezeichnet und zur Weiterverarbeitung z.B. zur Steuerung von Anwendungen verwendet werden.

O

OpenNI

siehe Abschnitt 4.3

optischer Fluss

ist eine wichtige Repräsentation von Bewegungsinformation in Bildern.

Outside-In

siehe Abschnitt 2.4.1.5

P

Panumbereich

siehe Abschnitt 2.2.2.2

Parallaxe

definiert den Winkel zwischen zwei Geraden, die von verschiedenen Standorten („Basislinie“) auf denselben Punkt (ein Objekt) gerichtet sind.

Präsenz

Definition siehe Abschnitt 2.1.3

Präsenzgefühl

siehe Präsenz

Q

Quaternionen

erlauben eine rechnerisch elegante Beschreibung des dreidimensionalen Raumes, insbesondere im Kontext von Rotationen.

R

REKNOW

REKNOW GmbH & Co. KG, ist das Unternehmen für das die Arbeit angefertigt wird

RGB-Kamera

ist eine Kamera, die eine rechteckige Bildmatrix liefert, bei der jeder Bildpunkt ein Farbwert (Rot, Grün, Blau) zugeordnet ist.

Rigid Body

zu Deutsch starrer Körper, ist ein physikalisches Modell eines nicht verformbaren Körpers.

S

Stereodisplay

ist ein Display der einen plastischen dreidimensionalen Bildeindruck erzeugt.

Stereosehen

siehe Abschnitt 2.2.1.1

stereoskopische Darstellung

siehe Stereosehen

STL-Format

Surface Tessellation Language (deutsch etwa Sprache zur Beschreibung der Oberfläche durch Dreiecke) ist eine Standardschnittstelle vieler CAD-Systeme, welche dazu dient, geometrische Informationen aus dreidimensionalen Datenmodellen bereit zu stellen.

T

Target

siehe Abschnitt 2.4.1.1

Tracking

umfasst alle Arbeitsschritte, die zur Verfolgung von (bewegten) Objekten dienen. Ziel ist das Abbilden der beobachteten tatsächlichen Bewegung zur technischen Verwendung. Das kann z.B. das Zusammenführen eines verfolgten Objektes mit einem nachfolgenden Objekt sein oder auch die Kenntnis des momentane Standort eines verfolgten Objektes.

Tracking-Bereich

ist der Bereich eines Tracking-Systems, indem die Verfolgung eines Nutzers möglich ist.

Tracking-System

ist ein System, das die durch ein Tracking-Verfahren (siehe Tracking) ermittelten Werte, zu einem verfolgten Objekt, Drittanwendungen zur weiteren Verarbeitung bereitstellt.

TÜV

führt technische Sicherheitskontrollen durch, welche durch staatliche Gesetze oder Anordnungen vorgeschrieben sind.

V

Virtual Eye Separation

siehe Abschnitt 2.2.2.1

VR

"Virtual Reality" zu Deutsch "Virtuelle Realität"

VRPN

siehe Abschnitt 4.3

VR-System

ist ein System das den Nutzer in eine Virtuelle Realität versetzt, dazu verwendet es aktuelle Virtual Reality Technologien.

VR-Umgebungen

besteht aus mehreren VR-Systemen (HMD, Head-Tracking, Datenhandschuhen) die durch Zusammenarbeit dem Nutzer ein sehr gutes VR-Erlebnis bieten.

W

Wand (eng.)

ist ein Eingabegerät zur Steuerung von VR Anwendungen, das als Zeigerwerkzeug verwendet werden kann, mit dem entfernte Objekte z.B. selektiert werden können.

Inhaltsverzeichnis

1	Einleitung	13
1.1	Problemstellung, Motivation und Ziele	13
1.2	Vorstellung EMM-Check	14
1.2.1	Gesetzlicher Hintergrund	15
1.2.2	Sichtfeldanalyse am CAD Modell	15
1.2.3	Funktionsumfang von EMM-Check	16
1.2.4	Fahrer-Modus	17
1.2.5	Nutzfahrzeug Inspektion	18
1.3	Struktur der Arbeit	18
2	Grundlagen	20
2.1	Virtual Reality	20
2.1.1	Technologieorientierte Charakterisierungen der VR	20
2.1.2	VR als innovative Form der Mensch-Maschine Interaktion	21
2.1.3	Mentale Aspekte der VR-Erfahrung	22
2.1.4	Wozu VR?	22
2.1.5	Die Perfekte VR	23
2.2	Wahrnehmungsaspekte des Menschen	24
2.2.1	Menschliche Wahrnehmung	25
2.2.1.1.	Visuelle Wahrnehmung	25
2.2.1.2.	Auditive Wahrnehmung	26

2.2.1.3.	Haptische Wahrnehmung.....	26
2.2.2	Phänomene, Probleme, Lösungen.....	27
2.2.2.1.	Abweichende Betrachtungsparameter	27
2.2.2.2.	Doppelbilder	29
2.2.2.3.	Diskrepanzen in der Raumwahrnehmung.....	31
2.2.2.4.	Diskrepanzen in der Bewegungswahrnehmung	32
2.2.2.5.	Cybersickness / Simulatorkrankheit.....	33
2.3	Navigation in virtuellen Welten	33
2.3.1	Bewegungskontrolle	33
2.3.2	Steuerungstechniken zur Bewegungskontrolle	34
2.3.3	Walking als Technik zur Bewegungskontrolle	35
2.4	Allgemeine Tracking Verfahren in geschlossenen Räumen	36
2.4.1	Optisches Tracking	36
2.4.1.1.	Markenbasierte Verfahren	37
2.4.1.2.	Markenlose Verfahren	38
2.4.1.3.	Geometriebasiertes Tracking.....	38
2.4.1.4.	Weitere merkmalsbasierte Tracking-Verfahren	38
2.4.1.5.	Outside-In Verfahren.....	39
2.4.1.6.	Inside-Out Verfahren.....	40
2.4.1.7.	Vergleich der optischen Tracking-Systeme	40
2.4.2	Akustisches Tracking	40
2.4.3	Elektromagnetisches Tracking.....	41
2.4.4	Inertial Tracking	41

3 Analyse von Produktlösungen zur Positionsbestimmung 42

3.1	Untersuchungskriterien.....	42
3.1.1	Anzahl der Freiheitsgrade pro verfolgtem Körper.....	43
3.1.2	Anzahl der gleichzeitig verfolgten Körper	43
3.1.3	Größe des überwachten Bereiches.....	43
3.1.4	Genauigkeit.....	43
3.1.5	Wiederholrate	44
3.1.6	Latenz	44

3.1.7	Empfindlichkeit gegenüber äußeren Rahmenbedingungen.....	44
3.1.8	Kalibrierung	44
3.1.9	Usability.....	44
3.2	Anforderungen des REKNOW Tracking-Systems	45
3.3	Die Produktlösungen	46
3.3.1	ImmerSight	46
3.3.2	WorldViz - Lightweight Stereoglass MotionTracker	48
4	Der Prototyp	50
4.1	Motivation	50
4.2	Systembeschreibung	51
4.3	Hardware und Software	52
4.4	Design	54
4.5	Implementation	56
4.5.1	Client	56
4.5.2	Server	57
4.5.2.1.	DepthCameraHandler	58
4.5.2.2.	TrackingSystemClientHandler.....	59
4.5.2.3.	Transformation	60
4.5.3	Kommunikation zwischen Client und Server	61
4.5.4	Verfahren zur Kalibrierung der Tiefenkameras	61
4.6	TrackingSystemServer Ablauf im Überblick.....	63
4.7	Auswertung hinsichtlich der Untersuchungskriterien	64
4.8	Mögliche Verbesserungen	68
4.9	Auswahl eines geeigneten Tracking-Systems.....	69
5	Schluss	72
5.1	Fazit	72
5.2	Ausblick.....	73

Abbildungsverzeichnis

Abbildung 1 - EMM-Check - Nutzfahrzeug auf dem Prüfkreis zur Sichtfeldanalyse. Die gelben Bereiche stellen die Spiegel- und Kamerasichtfelder dar. Die kleinen Fähnchen markieren mögliche Sichtfeldeinschränkungen auf dem Prüfkreis [2].	16
Abbildung 2 - Fahrer-Modus von EMM-Check, [2]	17
Abbildung 3 - Geometrische Auswirkung der Änderung der Virtual Eye Separation, [23].	28
Abbildung 4 - Zyklopische Skalierung, [23]	30
Abbildung 5 - Darstellung des gleichen virtuellen Raums mit (links) kleinem und (rechts) großem geometrischen Sichtbereichen, [28]	32
Abbildung 6 - merkmalsbasiertes Tracking mit SURF, [70]	38
Abbildung 7 - ImmerSight [51]	46
Abbildung 8 - WorldViz Kalibrierungswerkzeug [54]	48
Abbildung 9 - WorldViz Motion Tracker [53]	48
Abbildung 10 - WorldViz PPT [54]	48
Abbildung 11 - Skizze der Systemidee, Kamerakegel entsprechen nicht den Öffnungswinkeln und Reichweiten der Tiefenkameras	52
Abbildung 12 - FFAST mit aktivem Tracking Fenster, in dem eine Full Body Tracking zusehen ist	53
Abbildung 13 - Verteilungs-Diagramm des Systems mit zwei Tiefenkameras	54
Abbildung 14 - TrackingSystemServer Komponenten Ansicht	55
Abbildung 15 - Klassen und Interfaces des TrackingSystemClient	56
Abbildung 16 - Innenansicht der DepthCameraHandler Komponente	58
Abbildung 17 - Innenansicht der TrackingSystemClientHandler Komponente	59
Abbildung 18 - Innenansicht Transformation Komponente	60
Abbildung 19 - Kalibrierungsphase, eine Aufnahme von zwei verschiedenen Tiefenkameras zur selben Zeit aus der Sicht ihrer lokalen Koordinatensysteme (A).	62
Abbildung 20 - Interaktion zwischen den Komponenten des TrackingSystemServers	63
Abbildung 21 - linkes Bild enthält Störungen, durch überlappende Tiefenkamerakegel im rot markierten Bereich. Gelber Bereich zeigt einen größeren Fehler zwischen linkem und rechtem Bild. Rechtes Bild enthält keine Störungen	66

Tabellenverzeichnis

Tabelle 1 - Systeme und ihre Untersuchungskriterien im vergleichenden Überblick	70
---	----

1 Einleitung

„Virtual Reality ist eine Zukunftstechnologie, die zunehmend an Bedeutung in einer Vielzahl von Anwendungsgebieten in Forschung und Industrie gewinnt. Durch die rasante Entwicklung von Displayhardware, neuen Interaktionsgeräten und Trackingsystemen, werden heute Virtual Reality Anwendungen entwickelt, die vor wenigen Jahren ausschließlich in großen Forschungslaboren realisierbar gewesen wären. Diese Technologierichtung wird in besonderem Maße durch deutsche Forschungs- und Entwicklungseinrichtungen geprägt.“

Zitat aus [1] von Prof. Dr. Oliver Stadt
Universität Rostock
Sprecher der GI-Fachgruppe
Virtual Reality und Augmented Reality

Getrieben von der Virtual Reality als Zukunftstechnologie, die zunehmend an Bedeutung gewinnt, untersucht die vorliegende Bachelorarbeit mögliche Erweiterungen einer bestehenden virtuellen Simulationsumgebung durch den Einsatz von Virtual Reality Funktionalitäten.

1.1 Problemstellung, Motivation und Ziele

Die Firma REKNOW GmbH & Co. KG (kurz: REKNOW) entwickelt und vertreibt die Software EMM-Check zur Sichtfeldanalyse am virtuellen Nutzfahrzeug. In die Software können Konstruktionsdaten des Nutzfahrzeugs aus gängigen CAD-Systemen importiert und das Fahrzeug auf sein Sichtfeld hin untersucht werden. Die Sichtfeldanalysen erfolgen auf Basis international geltender Normen. Diese sollen verhindern, dass Nutzfahrzeuge auf Baustellen und Straßen eingesetzt werden, die auf Grund von eingeschränkten Sichtbedingungen (z. B. tote Winkel) eine Gefahr für im Umfeld der Fahrzeuge arbeitende Menschen sowie andere Verkehrsteilnehmer darstellen. Vor dem Inverkehrbringen eines Neufahrzeugs werden deshalb heute Sichtfeldprüfungen z. B. durch den TÜV oder die Berufsgenossenschaft Bauwirtschaft, durchgeführt. Werden die entsprechenden Normen hierbei nicht erfüllt, müssen gegebenenfalls noch kurz vor Produktionsstart konstruktive Änderungen am Fahrzeug durchgeführt werden, die zusätzliche Kosten für die Entwicklung bedeuten, sowie starke zeitliche Verzögerungen mit sich bringen können. EMM-Check ermöglicht hingegen den Konstrukteuren jederzeit über den gesamten

Produktentwicklungsprozess hinweg die normgerechte Sichtfeldprüfung des Fahrzeugs. Damit wird sichergestellt, dass die Sichtfeldprüfung am physischen Prototyp bestanden wird und teure Nacharbeiten vermieden werden.

EMM-Check bietet darüber hinaus die Möglichkeit, sich virtuell in die Fahrerkabine des Nutzfahrzeuges zu setzen und die Sichtverhältnisse aus einer egozentrischen Kameraperspektive zu überprüfen (Fahrer-Modus). Das Umschauen in der Fahrerkabine erfolgt dabei durch Steuerung mit der Maus. Damit der visuelle Eindruck der Fahrerkabine und dem Fahrzeug sowie den Sichtverhältnissen noch realer erfahrbar ist, soll ein Head-Mounted Display zum Einsatz kommen. Dabei werden die natürlichen Bewegungen des Umschauens mit dem Kopf in der realen Welt auf die virtuelle Kamera in der Fahrerkabine abgebildet. Allerdings bringt der Einsatz von Virtual Reality Funktionalitäten in diesem Kontext auch gewisse Probleme und Einschränkungen mit sich, die diese Arbeit verdeutlichen soll.

Zusätzlich soll sich der Anwender zukünftig virtuell um das Nutzfahrzeug Modell herum bewegen können, um die eingeblendeten Sichtfeldinformationen einzusehen und mögliche Sichtfeldeinschränkungen am Modell aufzuzeigen. Hier soll untersucht werden, welche Möglichkeiten existieren, sich mit Hilfe eines Head-Mounted Displays und der Bewegungs- und Positionserfassung in einem realen Raum zu bewegen und das Modell virtuell auf diese Weise zu erkunden bzw. zu begutachten. Zudem sollen weitere Möglichkeiten hinsichtlich der Steuerung des Programmes mit Gesten in Ausblick gestellt werden.

Hauptziel der Arbeit ist es, eine möglichst kostengünstige Lösung zum Erkunden von virtuellen Nutzfahrzeugmodellen in EMM-Check mit Hilfe von Virtual Reality Funktionalitäten zu finden. Dabei soll durch Verwendung von Bewegungs- und Positionserfassung einer Person im realen Raum in Echtzeit Position und Bewegung dieser Person in die virtuelle Simulationsumgebung abgebildet werden, um dadurch das Fahrzeug virtuell auf natürliche Weise zu erkunden.

1.2 Vorstellung EMM-Check

Im Folgenden wird die Software EMM-Check von REKNOW vorgestellt, in die die Virtual Reality Funktionalitäten integriert werden soll. Dabei wird kurz auf den gesetzlichen Hintergrund eingegangen, darauf folgt eine beispielhafte Sichtfeldanalyse an einem Nutzfahrzeug sowie eine Erklärung zu den Funktionen des Programms, die durch diese Arbeit und mittels Virtual Reality erweitert werden sollen.

1.2.1 Gesetzlicher Hintergrund

Die Unfallursache für schwere und teils tödliche Unfälle beim Einsatz von Baumaschinen, selbstfahrenden Arbeitsmaschinen, Landmaschinen und anderen Nutzfahrzeugen ist, dass im Umfeld dieser Maschinen arbeitende Personen von den Maschinenführern nicht rechtzeitig wahrgenommen und deshalb an- bzw. überfahren werden (vgl. [2]). Deshalb ist beispielsweise die Sicherstellung eines ausreichenden Sichtfeldes und der damit einhergehenden Vermeidung von Unfällen durch die DIN EN 474-1/ISO 5006:2006 vor dem Inverkehrbringen von Erdbewegungsmaschinen seit dem 01.12.2008 verbindlich (vgl. [2]). Darüber hinaus gibt es noch weitere Normen, wie die ISO 5721, ISO 13564, 15830 und StVZO §25b Abs. 2. Die einen Beitrag zur Verbesserung der Sicherheit beim Einsatz von Nutzfahrzeugen leisten.

1.2.2 Sichtfeldanalyse am CAD Modell

Die Verbesserung der Sichtverhältnisse von den Nutzfahrzeugen kann durch konstruktive Maßnahmen, z.B. durch eine veränderte Position der Kabine, durch Drehsitze, erreicht werden (vgl. [2]). Die Möglichkeiten dieser Maßnahmen sind allerdings begrenzt. Die verbleibenden „toten Winkel“ im Sichtfeld des Maschinenführers lassen sich dann nur durch den zusätzlichen Einsatz von Spiegeln, Kamera-Monitor-Systemen usw. vermindern (vgl. [2]). Das sich durch die Gestaltung der Maschinen bzw. der Aufbauten ergebende Sichtfeld ist oft erst zu einem sehr späten Zeitpunkt im Produktentstehungsprozess zu ermitteln. Für die Sichtfeldanalyse am physischen Prototyp sind aufwendige und teure Testaufbauten in einer geeigneten Prüfumgebung erforderlich (vgl. [2]). Notwendige Änderungen und Anpassungen, seien diese konstruktiver Art oder die nachträgliche Berücksichtigung von optischen oder sensorischen Geräten, erzeugen dann hohe Kosten.

Die Lösung kann in einer Produktentwicklung begleitenden Sichtfeldanalyse liegen, die durch EMM-Check geleistet wird. EMM-Check ermöglicht beispielsweise die CAD-neutrale Sichtfeldanalyse von Erdbewegungsmaschinen gemäß DIN EN 474-1/ISO 5006:2006 über den gesamten Produktentstehungsprozess (vgl. [2] , siehe Abbildung 1). Die durch die Norm vorgegebenen Prüfprofile sind in EMM-Check integriert und erlauben die schnelle fahrzeug- und masseabhängige Einstellung der Prüfparameter (vgl. [2]).

Für manche Nutzfahrzeughersteller sind die am Markt verfügbaren Software-Produkte zur Sichtfeldanalyse nicht einsetzbar. Weil sie beispielsweise Funktional zu umfangreich (siehe [3]), nicht ausschließlich auf die Sichtfeldanalyse fokussiert (siehe [3]), hohe Software- und Schulungskosten verursachen und nicht für das eingesetzte CAD-System verwendbar sind. Hierfür wurde EMM-Check entwickelt, damit das Problem der Sichtfeldanalyse schnell und effizient gelöst wird.

1.2.3 Funktionsumfang von EMM-Check

Dieser Abschnitt wurde dem EMM-Check Produktblatt [2] entnommen und enthält zusätzliche Einfügungen, die kursiv gekennzeichnet sind.

„Für die Sichtfeldanalyse in EMM-Check sind nur die 3D-Daten des Fahrzeugs erforderlich, die eine tatsächliche Sichtfeldeinschränkung hervorrufen können. Aus dem CAD-System werden die Konstruktionsdaten des Fahrzeugs im STL-Format exportiert und in EMM-Check eingelesen. Spiegel und Kameras lassen sich aus in EMM-Check integrierten Katalogen auswählen und positionieren.

Anhand von Prüfprofilen berechnet die Software die vom Maschinenführer nicht einsehbaren Bereiche sowie die Spiegel- und Kamerasichtfelder. Die nicht einsehbaren Bereiche werden als Sichtfeldeinschränkungen auf dem Prüfkreis sowie auf der 1m-Sichtlinie um das Fahrzeug dargestellt, ausgewertet und grafisch dokumentiert (*siehe Abbildung 1, die kleinen Fähnchen*). Ebenso lassen sich Spiegel- und Kamerasichtfelder - unter Berücksichtigung der Geometrie des Fahrzeugs - berechnen und auswerten sowie mit den Sichtfeldeinschränkungen in Korrelation bringen (*siehe Abbildung 1, gelbe und schwarze Bereiche*).“ (EMM-Check 3 Produktblatt 2013 [2])

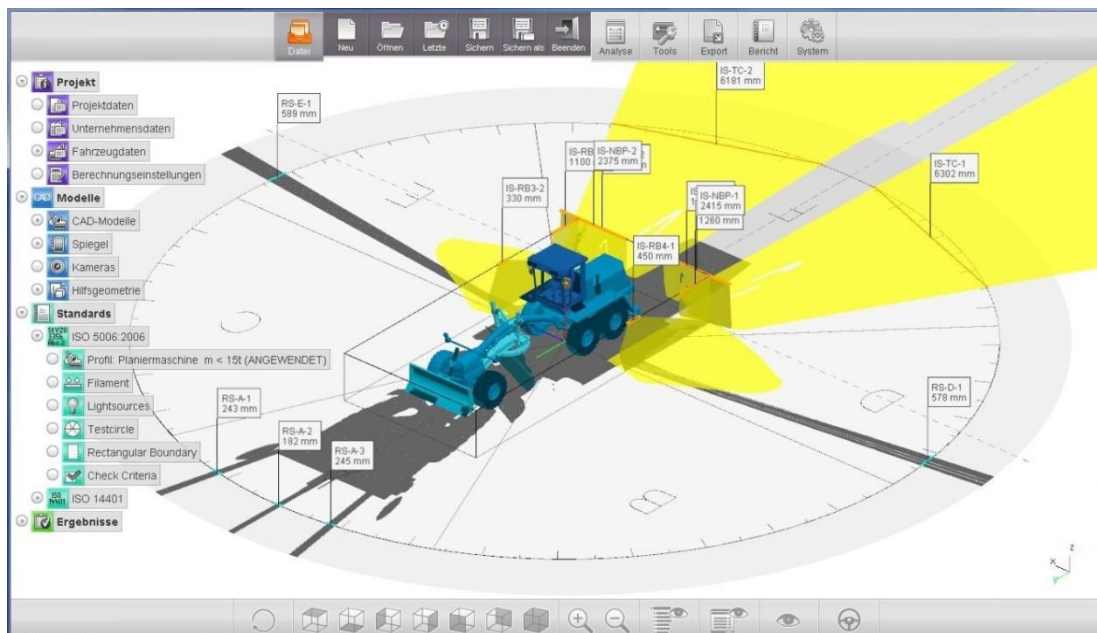


Abbildung 1 - EMM-Check - Nutzfahrzeug auf dem Prüfkreis zur Sichtfeldanalyse. Die gelben Bereiche stellen die Spiegel- und Kamerasichtfelder dar. Die kleinen Fähnchen markieren mögliche Sichtfeldeinschränkungen auf dem Prüfkreis [2].

„Standardprüfprofile für verschiedene Maschinentypen, wie sie in den zugrunde liegenden Normen, z. B. der ISO 5006:2006 für Baumaschinen, spezifiziert sind, sind in EMM-Check integriert und erlauben die schnelle Einstellung der Prüfparameter. Die freie Definition von Prüfkriterien ist ebenso möglich, um die Software auch zur Analyse von nicht der Norm entsprechenden Fahrzeug- bzw. Sichtsituationen nutzen zu können. Die Normkonformität des analysierten Fahrzeugs wird durch EMM-Check auf Basis der durch die gewählte Norm vorgegebenen Parameter beurteilt. Die Prüfergebnisse werden in einem normgerechten Bericht grafisch und tabellarisch dargestellt. Der Re-Import von Sichtfeldeinschränkungen, Spiegel- und Kamerasichtfeldern im CAD-System ermöglicht die unmittelbare Optimierung der sichtfeldbeeinflussenden Geometrie.“ (EMM-Check 3 Produktblatt 2013 [2])

1.2.4 Fahrer-Modus

EMM-Check bietet darüber hinaus die Möglichkeit, sich virtuell in die Fahrerkabine des Nutzfahrzeuges zu setzen und die Sichtverhältnisse aus einer egozentrischen Kameraperspektive zu überprüfen (Fahrer-Modus, siehe Abbildung 2). Das Umschauen in der Fahrerkabine erfolgt in der aktuellen Version der Software durch Steuerung mit der Maus oder die Tastatur. Dieser Modus dient zum einen dazu, einen Eindruck davon zu bekommen, wie die Sichtverhältnisse aus der Fahrzeugkabine sind und zum anderen als Präsentationsmodus für zukünftige Fahrer dieser Nutzfahrzeuge, um sich vorab einmal mit der Fahrerkabine vertraut zu machen.

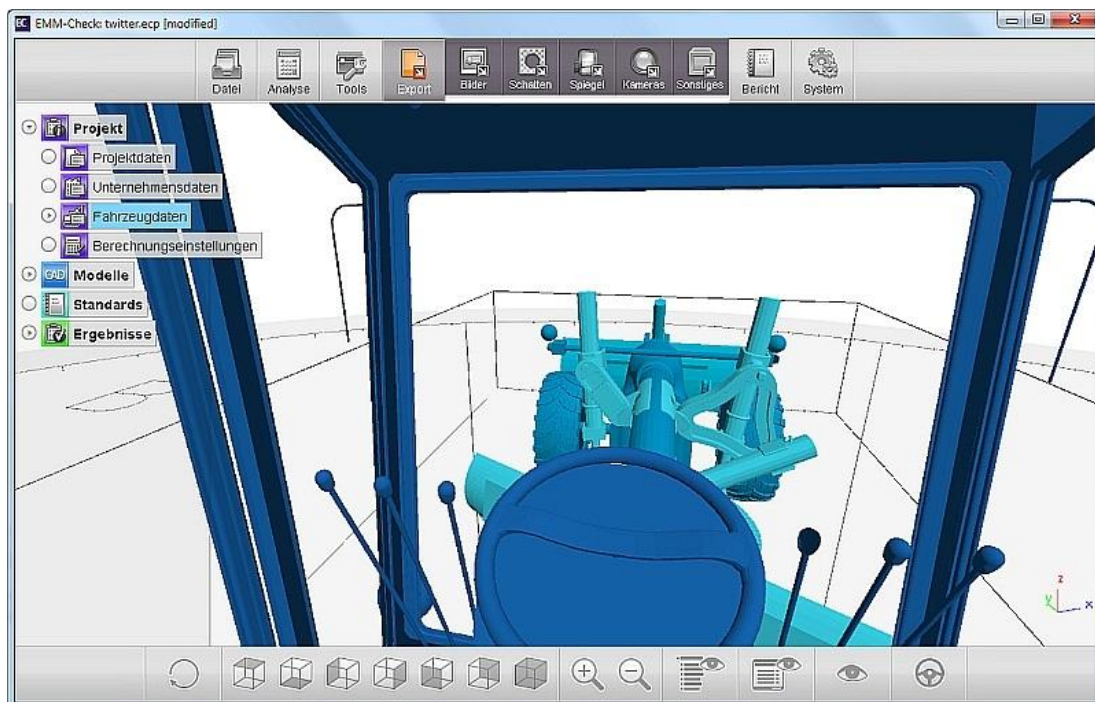


Abbildung 2 - Fahrer-Modus von EMM-Check, [2]

Der Eindruck des Fahrer-Modus und der Fahrerkabine sowie die damit verbundenen Sichtverhältnisse sollen zukünftig noch realer erfahrbar sein. Deshalb wird ein Head-Mounted Display zum Einsatz kommen, das das natürliche Umschauen mit dem Kopf in der realen Welt auf die virtuelle Kamera in der Fahrerkabine der Simulationsumgebung abbildet. Dies hilft dabei die Wahrnehmung und den Eindruck der Fahrerkabine zu verbessern. Allerdings bringt der Einsatz eines Head-Mounted Displays in diesem Kontext auch gewisse Probleme und Einschränkungen mit sich, die diese Arbeit verdeutlichen soll.

1.2.5 Nutzfahrzeug Inspektion

Die Nutzfahrzeug Inspektion von EMM-Check dient dazu, die Sichtfeldeinschränkungen am Nutzfahrzeug zu überprüfen (siehe Abbildung 1). Momentan erfolgt die Steuerung der virtuellen Kamera mit Maus und Tastatur. Dabei kann mit der Kamera um das Nutzfahrzeug rotiert sowie herein und heraus gezoomt werden. In Zukunft soll es möglich sein, die Nutzfahrzeuge virtuell mit einem Head-Mounted Display aus einer egozentrischen Kameraperspektive begutachtet zu können. Der Anwender soll sich innerhalb von EMM-Check frei auf dem Prüfkreis (siehe Abbildung 1) um das Nutzfahrzeug herum bewegen können, indem er sich in der realen Welt in einem ähnlich großen Raum bewegt und seine Bewegungen auf die virtuelle Kamera abgebildet werden. Damit dies möglich wird, wird ein geeignetes System zur Bewegungs- und Positionsbestimmung in geschlossenen Räumen benötigt. In dieser Arbeit werden mögliche Systeme zur Bewegungs- und Positionsbestimmung in geschlossenen Räumen untersucht sowie ein passendes System bestimmt. Außerdem wird ein eigenes System den Anforderungen von REKNOW entsprechend entworfen und entwickelt.

1.3 Struktur der Arbeit

Als erstes wird versucht, eine Charakterisierung von Virtual Reality aus unterschiedlichen Blickwinkeln vorzunehmen. Außerdem wird erklärt, wozu Virtual Reality verwendet werden kann und was die perfekte Virtual Reality auszeichnet.

Dafür behandelt diese Arbeit die Grundlagen der menschlichen Wahrnehmung. Basierend darauf wird auf typische Phänomene und Probleme von Virtual Reality eingegangen. Hier kann das Wissen über die menschliche Wahrnehmung zur Erklärung dieser Probleme und zum Ableiten von Lösungsstrategien genutzt werden.

Darauf folgen einige Konzepte und Techniken zur Navigation in virtuellen Welten. Diese sollen verdeutlichen, welche Möglichkeiten zur Navigation mit Head-Mounted Displays in virtuellen Welten existieren und welche Probleme dabei auftreten können.

Anschließend werden die allgemeinen Verfahren zur Positionsbestimmung in geschlossenen Räumen vorgestellt und auf ihre praktische Anwendbarkeit hin untersucht sowie deren Vor- und Nachteile dargestellt. Hierbei wird speziell auf optische Verfahren eingegangen.

Danach werden bereits vorhandene Systeme zur Positionsbestimmung in geschlossenen Räumen auf die Anforderungen eines für REKNOW in Frage kommenden Systems hin untersucht. Hierzu werden die Untersuchungskriterien sowie die Anforderungen des für REKNOW in Frage kommenden Systems vorgestellt und die vorhandenen System zur Positionsbestimmung in geschlossenen Räumen präsentiert und bewertet.

In Anlehnung daran wird ein eigenes und kostengünstiges System zur Positionsbestimmung in geschlossenen Räumen entworfen, implementiert und auf seine Tauglichkeit hin untersucht. Abschließend wird dieses System mit den vorgestellten Professionellen Systemen verglichen und ein für REKNOW in Frage kommendes System zur Positionsbestimmung in geschlossenen Räumen ausgewählt.

Am Schluss erfolgen das Fazit und ein Ausblick auf die Zukunft von Virtual Reality.

2 Grundlagen

2.1 Virtual Reality

Die Virtuelle Realität oder auch Virtual Reality genannt (im Folgenden als VR bezeichnet) ist ein junges Wissenschaftsgebiet. Die Weiterentwicklung wird unter anderem stark von den rasanten Fortschritten bei der zugrundeliegenden Hardware getrieben. Aus diesem Grund ist es nicht sehr überraschend, dass dieses Wissenschaftsgebiet bisher noch keine einheitliche Definition von „Virtueller Realität“ hervorgebracht hat. (vgl. [4]) Es gibt dennoch weitgehende Übereinkunft zu den wesentlichen bzw. wünschenswerten Merkmalen von VR. Im Folgenden wird versucht, eine Charakterisierung von VR aus unterschiedlichen Blickwinkeln vorzunehmen. Hierzu wird zuerst auf die technologischen Aspekte eingegangen, anschließend auf die Herausstellung der VR als neue Form der Mensch-Computer-Interaktion und als letztes auf die Merkmale der mentalen Erfahrung von VR. Nach dieser Charakterisierung wird geklärt, wozu VR verwendet wird, und im Anschluss wird die perfekte VR erläutert.

2.1.1 Technologieorientierte Charakterisierungen der VR

„The ultimate display would, of course, be a room within which the computer can control the existence of matter. A chair displayed in such a room would be good enough to sit in. Handcuffs displayed in such a room would be confining, and a bullet displayed in such a room would be fatal. With appropriate programming such a display could literally be the Wonderland into which Alice walked.“ (Ivan Sutherland 1965 [5])

In vielen bildlichen Darstellungen von VR-Systemen liegt ein unverkennbares Merkmal auf den besonderen Ein- und Ausgabegeräten, wie am Kopf des Nutzers beispielsweise in Form von Helmen angebrachten Displays (auch Head-Mounted Displays genannt und im Weiteren als HMD bezeichnet) und Datenhandschuhen. Aus diesem Grund kann eine Herausstellung der technologischen Aspekte der VR eine Möglichkeit zur Charakterisierung sein (vgl. [4]). Es besteht jedoch eine gewisse Gefahr bei technologiezentrierten Ansätzen, wenn sich die Definition der VR zu sehr auf Ein- und Ausgabegeräte bezieht (beispielsweise „verkabelte Datenanzüge“). Dies kann durch technologischen Fortschritt schnell veraltet bzw. überholt sein. Deshalb sollten zukunftssichere Definitionen der VR auch mit den

visionären Vorstellungen wie Sutherland's Ultimate Display kompatibel sein. Die beiden folgenden technologieorientierten Charakterisierungen aus den Anfängen der VR treffen auch noch auf heutige VR-Systeme zu (vgl. [4]).

„Virtual Reality (VR) refers to the use of three-dimensional displays and interaction devices to explore real-time computer-generated environments. “(Steve Bryson 1993 [6])

„Virtual Reality refers to immersive, interactive, multi-sensory, viewer-centered, three-dimensional computer generated environments and the combination of technologies required to build these environments. “(Carolina Cruz-Neira 1993, [7])

Der Begriff „immersive“ zu Deutsch Immersion aus Carolina Cruz-Neiras Zitat gründet nach Slater und Wilbur (vgl. [8]) auf vier technischen Eigenschaften von Ausgabegeräten. Erstens sollen die Sinneseindrücke des Menschen möglichst ausschließlich durch den Computer generiert werden, damit der Nutzer soweit wie möglich von der realen Umgebung isoliert wird. Zweitens sollen möglichst viele Sinne angesprochen werden. Drittens sollen die Ausgabegeräte den Nutzer vollständig umgeben. Als viertes und letztes sollen die Ausgabegeräte eine lebendige Darstellung bieten. Diese kann beispielsweise durch eine hohe Auflösung und Qualität der Darstellung erreicht werden.

2.1.2 VR als innovative Form der Mensch-Maschine Interaktion

Die Entwicklung von Mensch-Maschine-Schnittstellen, die im Vergleich zu traditionellen Benutzungsschnittstellen ein besonders natürliches oder intuitives Interagieren mit der virtuellen Welt zulassen, stellt eine weitere Möglichkeit zur Charakterisierung der VR dar (vgl. [4]). Wird statt Maus und Tastatur beispielsweise ein Lenkrad und ein Fußpedal für ein Autorennspiel verwendet, so ist das ein Schritt hin zu einer VR, in der die Bedienung des virtuellen Autos in der virtuellen Welt natürlicher ist (vgl. [4]). Im folgenden Zitat von Robert Stone wird die Zielvorstellung intuitiver Benutzungsschnittstellen im Kontext von VR-Systemen erläutert.

„An intuitive interface between man and machine is one which requires little training and proffers a working style most like that used by the human being to interact with environments and objects in his day-to-day life. In other words, the human interacts with elements of this task by looking, holding, manipulating, speaking, listening, and moving, using as many of his natural skills as are appropriate, or can reasonable be expected to be applied to a task. “(Robert Stone 1993 [9])

Im Vergleich zu anderen innovativen Formen der Mensch-Maschine Interaktion besitzt auch die VR großes Potential zur Realisierung intuitiver Mensch-Maschine-Schnittstellen im Sinne von Robert Stone. In den meisten bisherigen VR-Systemen ist die Zielvorstellung natürlicher Interaktionsformen allerdings noch nicht umgesetzt (vgl. [1]). Dennoch ist die Interaktion

mit den VR-typischen 3D-Ein- und Ausgabegeräten schon um einiges natürlicher als dies bei konventionellen 2D-Schnittstellen der Fall ist (vgl. [4]).

2.1.3 Mentale Aspekte der VR-Erfahrung

„Im Zentrum der VR steht eine Erfahrung - die Erfahrung in einer virtuellen Welt oder an einem fremden Ort zu sein.“ (Rheingold 1992 [10])

Die Konzepte der mentalen Erfahrung von VR folgen der Terminologie von Slater ([11], [12]). Das zentrale Konzept zur Beschreibung der mentalen Aspekte der VR-Erfahrung ist die Präsenz. Es beschreibt im weitesten Sinne das Gefühl, sich innerhalb der virtuellen Welt zu befinden, die durch ein immersives VR-System dargestellt wird. Wenn die Nutzer einer VR die virtuelle Welt so fühlen, als wenn sie in der realen Welt handeln würden, dann ist das ein Hinweis auf das Gefühl der Präsenz. Das Präsenzgefühl setzt sich aus drei verschiedenen Teilaspekten zusammen:

- der Ortsillusion - diese bezieht sich auf das Gefühl, sich an dem vom VR-System dargestellten Ort zu befinden (vgl. [12])
- der Plausibilitätsillusion - diese entsteht, wenn die Ereignisse der simulierten Umgebung so wahrgenommen werden, als ob sie wirklich geschehen würden (vgl. [12])
- der Involviertheit - diese bezieht sich auf den Grad der Aufmerksamkeit bzw. des Interesses des Nutzers an der simulierten Welt (vgl. [12])

2.1.4 Wozu VR?

Es stellt sich nun die Frage, warum eine Virtuelle Realität überhaupt aufgebaut werden sollte, um einen Menschen in diese zu versetzen.

Eine Antwort auf diese Frage ist, dem Menschen das Aufnehmen und Verstehen von Daten zu erleichtern, indem mit Hilfe der VR diese Daten veranschaulicht werden (vgl. [4]). Beispielsweise kann man Architekten dabei helfen, die 2D-Baupläne für die Bauherren mit Hilfe einer VR so zu visualisieren, dass diese einen guten Eindruck von den Gebäuden erhalten und Entscheidungen hinsichtlich Realisierungsalternativen treffen können (vgl. [13]). Ebenfalls können komplexe Ergebnisse von Computersimulationen, wie beispielsweise die an einem neu geplanten Fahrzeug entlang strömende Luft, direkt am virtuellen Fahrzeug visualisiert werden. Ingenieure und Designer können in Zusammenarbeit die Karosserieformen in der virtuellen Welt ästhetischer und ansprechender gestalten (vgl. [14]). Damit lassen sich Luftverwirbelungen vermeiden und der Luftwiderstand des Fahrzeugs wird gesenkt. Heutzutage wird kaum noch ein Auto ohne Nutzung von Methoden aus der VR entwickelt (vgl. [4]). Entwürfe können realitätsnäher visualisiert und digitale Prototypen kostengünstiger erstellt werden, im Vergleich zu

traditionellem Modellbau. Es können beispielsweise auch in einer virtuellen Welt die Roboter einer Fertigungsstraße für Automobile vor Produktionsbeginn auf ein neues Automodell eingestellt und die Simulation den beteiligten Personen in einer VR dargestellt werden (vgl. [14]). So werden Planungsfehler in einer virtuellen Anlage oder Fabrik beseitigt, denn dort ist es deutlich einfacher und verursacht weniger Kosten als in der realen Welt. Schulungen mit Hilfe von Simulatoren, die die Personen in eine virtuelle Welt versetzen und in der es ihnen erlaubt ist, Fehler zu machen, um wiederum daraus zu lernen (vgl. [4]), stellt eine weitere Einsatzmöglichkeit der VR dar. In Flugsimulatoren werden Piloten trainiert, es werden keine realen Flugzeuge verwendet und dadurch spart die Fluggesellschaft Geld ein. Das Training bietet nicht nur finanzielle Vorteile, es entstehen auch weniger CO₂-Emissionen, als bei der Verbrennung von Kerosin bei einem realen Flugzeug und dadurch wird zusätzlich die Umwelt geschont. Auch ist es den Piloten möglich, Extremsituationen gefahrlos üben zu können. Ein weiteres Beispiel ist das Training von Personal für komplexe Anlagen, wie beispielsweise das Bedienen des Leitstandes eines Kraftwerks oder die Wartung von Flugzeugen. Das Personal ist durch die VR bereits zum Zeitpunkt der Inbetriebnahme, durch das Training vor der Fertigstellung, entsprechend ausgebildet (vgl. [4]).

Dieser Abschnitt hat gezeigt, wozu VR verwendet werden kann. Daten können für den Menschen besser veranschaulicht, Kosten eingespart und sogar Extremsituationen trainiert werden. Die vorgestellten Beispiele sind nur ein Ausschnitt von dem, was heutzutage möglich ist (für weitere Beispiele verweise ich hier auf das Buch „Virtual und Augmented Reality (VR/AR)“ [4]) und es wird in Zukunft noch weitere Anwendungsgebiete und Einsatzmöglichkeiten für die VR geben.

2.1.5 Die Perfekte VR

Nach Ralf Dörner [4] sind in einer perfekten VR sämtliche Sinneseindrücke des Nutzers computergeneriert. Diese Sinneseindrücke müssen in gleicher Quantität und Qualität, wie es der Mensch aus der realen Welt gewohnt ist, erzeugt werden. Auch die Handlungen des Menschen hätten in der VR die gleichen Effekte. Virtuelle Objekte müssten genauso auf den Menschen einwirken, wie er es aus der realen Welt gewohnt ist. Nach Ralf Dörner [4] sind heutige VR-Systeme nicht perfekt. Aber die Entwicklung der VR-Technologien unter Einsatz erheblicher Hard- und Software-Ressourcen zielt auf die Erschaffung immer realistischerer erfahrbarer Simulationen ab (vgl. [4]). Damit nun eine perfekte VR zumindest in Ansätzen realisiert werden kann, müssen Reize erzeugt werden, die einen Menschen die Virtuelle Realität wahrnehmen lassen. Die visuellen Reize beispielsweise können mit Hilfe der Computergrafik generiert werden. Auf dem Weg zur perfekten VR ist die Erzeugung der Reize allerdings nur eine Aufgabe. Der Mensch möchte in der virtuellen Welt nicht nur sehen sondern auch handeln können (vgl. [4]). Sieht er beispielsweise einen Fußball, dann möchte er vielleicht auch gegen diesen treten können und ihm vielleicht sogar hinterher laufen. Damit dies möglich wird, muss die virtuelle Welt simuliert werden. Die möglichen

Aktionen des Menschen müssen der Simulation bekannt sein und diese Aktionen müssen die Simulation dementsprechend beeinflussen (vgl. [4]). Die Ergebnisse der Simulation müssen wiederum Auswirkungen auf die Erzeugung der Reize haben. Bewegt sich der Mensch beispielsweise in der virtuellen Welt, muss die neue Position von der Reizerzeugung berücksichtigt werden. Das Verhalten der VR wird durch das Simulationsmodell der Welt festgelegt. Hier müssen zum einen Reaktionen der Welt ausgehend von den Aktionen des Menschen simuliert werden und zum anderen aber auch Änderungen in der virtuellen Welt, die von den Aktionen des Menschen unabhängig sind. Beispielsweise wäre ein Tag-Nacht-Zyklus in einer virtuellen Welt eine solche Aktion.

Ein Computersystem, das die Aktionen des Menschen erkennt und die Simulation der virtuellen Welt sowie die Reizerzeugung für den Menschen zur Aufgabe hat, kann ein hochkomplexes System werden (vgl. [4]). Alleine die Simulation eines einzigen virtuellen Menschen mit realitätsnahen Bildern von Haut und Kleidung, der Sprachsynthese, die Simulation menschlichen Verhaltens und von Emotionen ist heute eine große Herausforderung (vgl. [4]). Die Anforderungen erhöhen sich sogar noch dadurch, dass dieses Computersystem in Echtzeit arbeiten muss. Das setzt voraus, dass alle Berechnungen nicht beliebige Zeit in Anspruch nehmen dürfen, sondern strikte Zeitvorgaben einhalten müssen. Um dem menschlichen Betrachter beispielsweise Bewegungen in der virtuellen Welt als kontinuierlich und natürlich wahrnehmen zu lassen, müssen eine Vielzahl von Bildern pro Sekunde berechnet und erzeugt werden.

Es ist nun möglich VR 's in verschiedenen Ausbaustufen zu erschaffen. Die perfekte VR, wie sie am Anfang dieses Abschnitts beschrieben wurde, ist dabei die extremste Ausbaustufe (vgl. [4]). Es lassen sich aber auch schon heute relativ glaubwürdige virtuelle Welten mit geringem Aufwand realisieren (vgl. [4]).

2.2 Wahrnehmungsaspekte des Menschen

Einem Nutzer die Illusion der Anwesenheit in einer virtuellen Welt zu suggerieren, ist eine der grundlegenden Herausforderungen der VR. Ob und wie gut die Umsetzung jedoch gelingt, ist nicht nur ein technisches Problem, sondern beruht auch auf Prozessen der menschlichen Informationsverarbeitung zur Interpretation der gebotenen Sinnesreize (vgl. [15]). Im Folgenden wird die Grundlage der menschlichen Wahrnehmung behandelt. Basierend darauf wird auf typische Phänomene und Probleme der VR eingegangen. Dabei kann das Wissen über die menschlichen Wahrnehmungsaspekte zur Erklärung dieser Phänomene und zur Ableitung von Lösungsstrategien genutzt werden.

2.2.1 Menschliche Wahrnehmung

Menschen nehmen ihre Umgebung über verschiedene Sinne wahr. Im Rahmen heutiger VR-Technologien sind dabei der visuelle, der akustische und der haptische Sinn die wichtigsten Sinne des Menschen. Sinne, wie der olfaktorische (Riechen) oder gustatorische (Schmecken), werden heute noch nicht virtuell stimuliert. Dadurch werden nahezu sämtliche in der virtuellen Welt dargestellten Informationen durch die Augen, Ohren oder über die Haut wahrgenommen (vgl. [15]).

Die Wahrnehmung in einer virtuellen Welt unterscheidet sich somit auf den ersten Blick nicht von der Wahrnehmung und den damit verbundenen Sinnen und Sinneseindrücken in einer typischen Desktop-Umgebung (vgl. [15]). Mit dem Bildschirm bzw. aus den Lautsprechern erlebt man die visuelle bzw. akustische Stimulation einer virtuellen Welt und durch Maus und Tastatur werden die haptischen Eindrücke vermittelt. Die Möglichkeit eine virtuelle Erfahrung auf immersive Art und Weise zu erleben, macht den Unterschied zwischen VR-Umgebungen und Desktop-Umgebungen aus. VR-Umgebungen nutzen im Gegensatz zu Desktop-Umgebungen nicht nur Maus und Tastatur, sondern werden durch 3D-Eingabegeräte oder sogar durch Bewegungen des Nutzers im realen Raum bedient. Diese werden auf die entsprechenden Bewegungen in der virtuellen Welt abgebildet. Neben diesen Eingaben in die VR-Umgebung existieren weitere Formen der Eingabe, wie z.B. Sprache, Gesten und andere menschliche Ausdrucksformen (siehe [16]).

2.2.1.1. Visuelle Wahrnehmung

Mit Hilfe der visuellen Wahrnehmung kann der Mensch insbesondere Objekte identifizieren. Das Gehirn analysiert das auf die Netzhaut projizierte Bild der virtuellen Welt auf Helligkeit, Kontraste, Farbe und Bewegung. Anschließend wird das Bild durch einen Helligkeitsausgleich oder eine Kontrastverstärkung bearbeitet (vgl. [15]). Das Erkennen von einzelnen Objekten und ihrer Bedeutung erfolgt wahrscheinlich mit Hilfe von Vergleichen mit bereits im Gehirn gespeicherten Erfahrungen, dabei werden die Bilder mit Körpergefühl, Emotionen, Geruch und Geräuschen verknüpft. Die räumliche Lagebeziehung unterstützt das Identifizieren und Unterscheiden von Objekten (vgl. [17]).

Das Stereosehen ist ein Beispiel für die Funktionsweise der menschlichen Wahrnehmung und wie sie durch ein VR-System manipuliert werden kann, damit eine Präsenz in der virtuellen Welt erzeugt wird (siehe [15]). Der Mensch hat zwei Augen, nimmt aber keine zwei separaten Bilder von der Realität wahr. Dem visuellen Wahrnehmungssystem des Menschen gelingt es, aus der zweidimensionalen Netzhaut der Augen einen dreidimensionalen Eindruck der Umwelt zu erzeugen (vgl. [15]). Dies geschieht unter anderem durch Triangulierung und wird erst durch zwei Augen möglich. Dadurch kann das visuelle System also Entfernungen von Objekten wahrnehmen. Somit lässt sich in der virtuellen Welt beispielsweise das Stereosehen manipulieren, um einen dreidimensionalen

Eindruck zu bewirken, welcher nur eine zweidimensionale Displayfläche nutzt. Dadurch wird also unter Ausnutzung der menschlichen Wahrnehmung, ein Stereodisplay realisiert, das einen plastischen dreidimensionalen Bildeindruck erzeugt. Dieser kann durch geeignete Wahl der Parallaxe dahin gehend manipuliert werden, dass Objekte vor oder hinter der Bildschirmfläche erscheinen(vgl. [15]).

Die Disparität dient der visuellen Wahrnehmung als Tiefenhinweis für die Raumwahrnehmung (vgl. [15]). Die Entfernung des Betrachters zum jeweiligen Objekt hat einen großen Einfluss auf die Aussagekraft und Zuverlässigkeit dieses Tiefenhinweises. Je weiter ein Punkt vom Betrachter entfernt ist, desto geringer ist die Disparität, die von ihm erzeugt wird. Ein Punkt im Abstand von 2 bis 3 m erzeugt eine sehr geringe Disparität. Ab einem Abstand von 10 m ist die Disparität nahezu nicht mehr wahrnehmbar (vgl. [15]). Somit lohnt sich der Aufwand für den Einsatz eines Stereodisplays bei VR-Systeme, deren virtuelle Welt so gestaltet ist, dass sich bedeutsame Objekte in Armreichweite befinden(vgl. [15]). Wohingegen in virtuellen Welten, bei denen Objekte mehr als 3 m vom Betrachter entfernt sind, der Einsatz eines Stereodisplays nicht viel zur Raumwahrnehmung beiträgt und damit überflüssig sein kann (vgl. [15]). Daraus lässt sich ableiten, dass der Einsatz eines Stereodisplays in Form eines Head-Mounted Displays erheblich zur Verbesserung der Raumwahrnehmung in einem Führerhaus eines Nutzfahrzeuges beitragen kann, weil dort die Bedienelemente in Armreichweite liegen und die Ausmaße eines Führerhauses innerhalb der wahrnehmbaren Disparität liegen.

2.2.1.2. Auditive Wahrnehmung

Mit den Ohren kann der Mensch Luftbewegungen wahrnehmen. Das Ohr nimmt diese in Form von mechanischen Wellen wahr, die durch Luft und Druckschwankungen erzeugt werden. Dabei fängt die Ohrmuschel die Schallwellen auf und leitet diese an das Mittelohr weiter. Dort werden sie in Vibrationen des Trommelfells umgewandelt. Die Haarzellen im Innenohr nehmen die verschiedenen Frequenzen wahr. Hier ist die räumliche Auflösung viel geringer als die des visuellen Sinnes (vgl. [15]). Die Auswertung der Amplituden ist neben den Laufzeitdifferenzen zwischen den Ohren wesentliche Grundlage unseres akustischen Ortungssystems. Die Unterscheidbarkeit hat jedoch deutliche Grenzen, weshalb zwei Geräuschquellen nur dann örtlich unterschieden werden können, wenn sie mehrere Grad auseinander liegen (vgl. [15]).

2.2.1.3. Haptische Wahrnehmung

Die sensorische und motorische Aktivität wird von der Haptik oder haptischen Wahrnehmung beschrieben. Diese ist für das Erfühlen von Objekteigenschaften, wie Größe, Konturen, Oberflächentextur und Gewicht verantwortlich. Über die Haut, durch Muskeln, Gelenke und Sehnen werden die empfunden Sinneseindrücke wahrgenommen (vgl. [18]).

Mit Hilfe des Tastsinns können Berührungen, Wärme und Schmerzen wahrgenommen werden. Dies geschieht durch Rezeptoren in der Haut. Die wichtigsten von ihnen sind die Mechanorezeptoren, die für Druck, Berührungen oder Vibrationen zuständig sind sowie Thermorezeptoren, die Wärme und Kälte wahrnehmen.

2.2.2 Phänomene, Probleme, Lösungen

Beim Einsatz von VR lassen sich überraschende Phänomene beobachten (vgl. [15]): In der einen Sekunde gelingt die Darstellung einer virtuellen Welt in einem Stereodisplay und in der nächsten nicht mehr. Der Betrachter nimmt die Welt nicht mehr plastisch wahr, sondern plötzlich wirkt alles doppelt oder verschwommen für ihn. Nutzer, die sich über einem längeren Zeitraum in einer VR aufhalten, fangen an, sich über Kopfschmerzen zu beklagen oder müssen sich sogar übergeben. Das Raumangebot eines neuen Autos wirkte in der VR sehr großzügig und obwohl der Innenraum dort in Augenschein genommen wurde, wird das Raumangebot in dem realen Auto als enttäuschend eng wahrgenommen, auch wenn die Proportionen zwischen dem realen und virtuellen Auto identisch sind. Diese Phänomene lassen sich anhand des Wissens über die menschliche Wahrnehmung erklären. Somit ist es möglich, Lösungsstrategien zu entwickeln, die die daraus resultierenden Probleme umgehen oder zumindest abmildern können. Dennoch sind wir mit heutiger VR nicht in der Lage, die Realität 1:1 abzubilden, denn es gibt immer Abweichungen (vgl. [15]). Im Folgenden werden VR-typische Phänomene und Problemstellungen thematisiert und auch die heute bekannten Erklärungsversuche und Lösungsansätze dargestellt.

2.2.2.1. Abweichende Betrachtungsparameter

Die Bilder der virtuellen Kamera werden auf einer planaren Bildebene erzeugt. Der Mensch hingegen erzeugt sie auf einer gekrümmten Netzhaut. Der Sichtbereich (field of view) des Betrachters kann vom Öffnungswinkel der virtuellen Kamera abweichen. Diese Abweichung liegt in den unterschiedlichen Variationen der jeweiligen Betrachtungsparameter der virtuellen Kamera und des Menschen (vgl. [15]). Demzufolge treten Vergrößerungen bzw. Verkleinerungen sowie Verzerrungen vom Bildeindruck auf. Dies führt dazu, dass die Entfernungseinschätzung oder auch die Wahrnehmung der Neigung von Objekten beeinflusst wird (siehe [19]). Beim Betrachten von Bildern der virtuellen Welt aus einer anderen Perspektive werden die dadurch entstehenden Verzerrungen als erstaunlich wenig störend wahrgenommen. Hierbei handelt es sich um die Robustheit der linearen Perspektive in der menschlichen Wahrnehmung (vgl. [20]). Dieses Phänomen lässt sich dadurch erklären, dass das visuelle System des Betrachters den verzerrten Bildeindruck aktiv korrigiert (siehe [21]).

Auch wenn abweichende Betrachtungsparameter als nicht sonderlich störend empfunden werden, ist es sinnvoll, eine Minimierung der Abweichung anzustreben. Insbesondere bei Anwendungen, bei denen großer Wert auf eine korrekte Einschätzung von Entfernungen

und Orientierung von Objekten im Raum gelegt wird. Soll die virtuelle Welt nicht nur passiv betrachtet werden, sondern sollen aktive Handlungen aus einer egozentrischen Kameraperspektive, wie beispielsweise das Greifen von Objekten oder Bewegungen, ausgeführt werden, ist diese Minimierung ebenfalls relevant (vgl. [15]). Ein Ansatz zur Minimierung besteht darin, dass die aktuellen Betrachtungsparameter, wie Position und Blickrichtung, z. B. durch Head-Tracking ermittelt werden, um eine blickpunktabhängige Bildgenerierung zu ermöglichen. Bewegt sich der Nutzer hierbei in der realen Welt, so wird die virtuelle Welt automatisch aus dessen neuer Perspektive dargestellt. Ein anderer Ansatz wiederum simuliert größere Brennweiten in der virtuellen Kamera, damit die Verzerrungen verringert werden [22].

Zusätzlich kann eine Abweichung bei Stereodisplays dadurch auftreten, dass die beiden virtuellen Kameras, die das Bild für das linke und rechte Auge erzeugen, einen Abstand (genannt Virtual Eye Separation) haben, der vom Pupillenabstand des Betrachters abweicht. Der durchschnittliche Pupillenabstand beträgt 64 mm, die individuelle Spannweite liegt im Intervall von 45 bis 75 mm. Wie Abbildung 3 zeigt, können kleine Änderungen im Pupillenabstand größere Änderungen in der Tiefenwahrnehmung nach sich ziehen (siehe [23]).

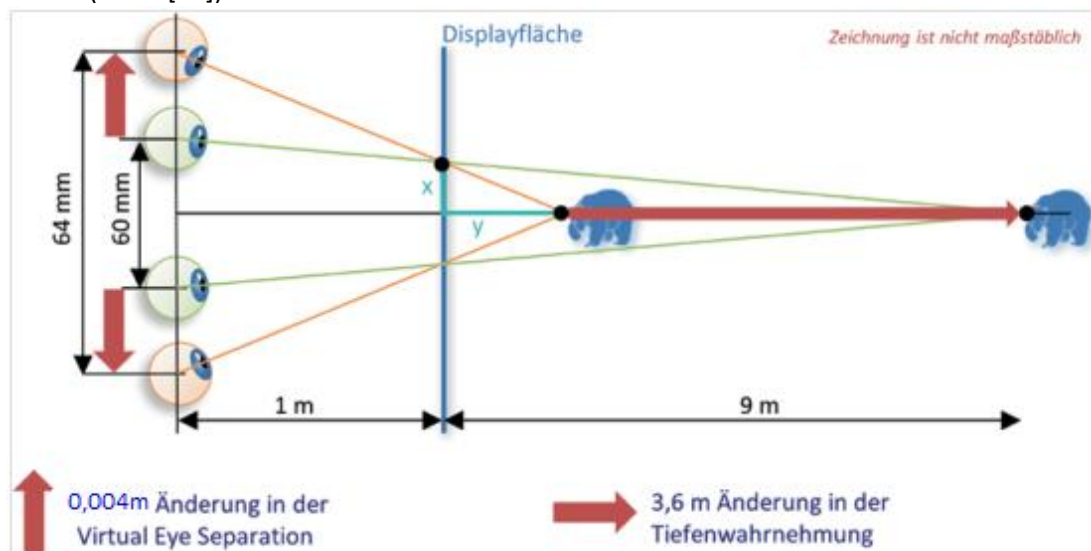


Abbildung 3 - Geometrische Auswirkung der Änderung der Virtual Eye Separation, [23].

Das auf der Projektionsfläche abgebildete Objekt wird bei einem Augpunkt Abstand von 64 mm 9 m hinter der Projektionsfläche lokalisiert. Verringert man nun den Augpunkt Abstand um 4 mm, so folgt durch Anwendung des Strahlensatzes, dass das virtuelle Objekt um 3,6 m nach vorne verschoben wird. Diese Abweichungen im Abstand der Augpunkte werden durch die Anpassungsfähigkeit der Augen so kompensiert, dass sie für den Betrachter ebenfalls als nicht störend wahrgenommen werden. Der Abstand der

virtuellen Kameras kann mehrmals in einer Sekunde geändert werden, ohne dass der Betrachter dies wahrnimmt (siehe [23]). Es ist also in einer VR nicht unbedingt nötig, zunächst den Abstand der beiden virtuellen Kameras entsprechend dem Abstand der beiden Augen des Betrachters anzupassen (vgl. [15]). Dennoch können beim Nutzer Nebenwirkungen wie Übelkeit auftreten, auch wenn ihm die Abweichung im Augenabstand nicht bewusst auffällt.

Das Experiment von Bruder (siehe [23]), bei dem virtuelle Avatare mit unterschiedlichen Betrachtungsparametern dargestellt und die jeweiligen geometrischen Sichtbereiche und virtuellen Augenabständen mit Faktoren relativ zu ihren korrekt kalibrierten Einstellung skaliert wurden, zeigte die empfundenen Effekte von geometrischen Sichtbereichen und dem virtuellen Augenabstand auf die Wahrnehmung von Größen und Distanzen in virtuellen Umgebungen. Die Ergebnisse der Untersuchung verdeutlichen, dass Variationen des geometrischen Sichtbereichs einen großen Effekt auf die Distanzwahrnehmung haben, wohingegen Variationen des virtuellen Augenabstands sich primär auf die Wahrnehmung von Größe auswirken. Als marginal und deutlich geringer als in mathematischen Modellen vorhergesagt (siehe [23]), zeigte sich speziell der Einfluss des geometrischen Sichtbereichs auf die Wahrnehmung von Größe bzw. der Einfluss des virtuellen Augenabstands auf die Distanzwahrnehmung. Die Ergebnisse zeigen ein weiteres Beispiel für die Unterschiede in der Wahrnehmung in der virtuellen und realen Welt.

2.2.2.2. Doppelbilder

Sobald die dem linken und rechten Auge gezeigten beiden unterschiedlichen Bilder nicht vom Betrachter eines Stereodisplays fusioniert werden können, tritt Diplopie (engl. Diplopia) auf. Der Betrachter nimmt Doppelbilder wahr bzw. sieht beide Bilder getrennt. Da dies meist als überaus störend empfunden wird und sich negativ auf das Gefühl der Präsenz in einer VR auswirkt, stellt dies ein schwerwiegendes Problem in einer VR dar (vgl. [15]). Daher ist die Diplopie unbedingt zu vermeiden. Die Ursache für Diplopie ist, dass die korrespondierenden Punkte der beiden Bilder auf der Netzhaut des linken und rechten Auges außerhalb des Panumbereichs liegen und sich deshalb nicht mehr vom Betrachter zu einem Bild fusionieren lassen. Der Panumbereich ist der Bereich, in dem noch eine Fusion der Bildinformationen des linken und rechten Auges möglich ist. Es ist also bei einem Stereodisplay nicht möglich, Objekte beliebig weit vor oder hinter der Displayfläche erscheinen zu lassen (vgl. [15]). Es steht nur ein begrenzter Bereich zur Verfügung, um virtuelle Welten mit Hilfe eines Stereodisplays darzustellen, ohne dass Diplopie auftritt. Die Grenze für den nutzbaren Stereobereich liegt bei -25 bis +60% des Abstandes vom Betrachter zur Displayfläche (siehe [24]). Die Situation wird zusätzlich dadurch verschärft, dass nicht der gesamte Panumbereich ausgeschöpft werden sollte, denn nur in einem Teilbereich ist die Fusion der Bilder über einen längeren Zeitraum hinweg auch ohne Anstrengung für den Betrachter möglich. Dieser Teilbereich umfasst etwa ein Drittel des Panumbereichs und wird als Percival's Zone of Comfort bezeichnet (vgl. [25]).

Bei dem Versuch Diplopie zu umgehen wird eine bestimmte Strategie verwendet. Diese besteht darin, dass der Panumbereich vergrößert wird. Die Größe des Panumbereichs ist abhängig von der Größe und dem Detailreichtum der betrachteten Objekte. Sobald die zu fusionierenden Bilder etwas verschwommen dargestellt werden und der Detailreichtum der betrachteten Objekte reduziert wird, vergrößert sich der Panumbereich (vgl. [15]). Eine andere Strategie besteht darin, Objekte näher an die Displayfläche und damit in den Panumbereich zu holen. Eine Technik, die dazu bereits vorgestellt wurde, ist die Virtual Eye Separation (siehe Abschnitt 2.2.2.1). Damit können hinter der Displayfläche liegende Objekte näher an die Displayfläche heran gebracht werden, wenn der Abstand der virtuellen Kameras verringert wird. Die Änderung der Virtual Eye Separation ist zur Vermeidung von Diplopie sehr hilfreich, weil die menschliche Wahrnehmung robust gegen diese Manipulation ist (vgl. [15]). Darüber hinaus existiert mit der Zyklopischen Skalierung (siehe [26]) eine weitere Strategie. Mit dieser wird die gesamte Szene zwischen den beiden virtuellen Kameras um einen Punkt skaliert (siehe Abbildung 4), um die virtuelle Welt in den Panumbereich zu verschieben.

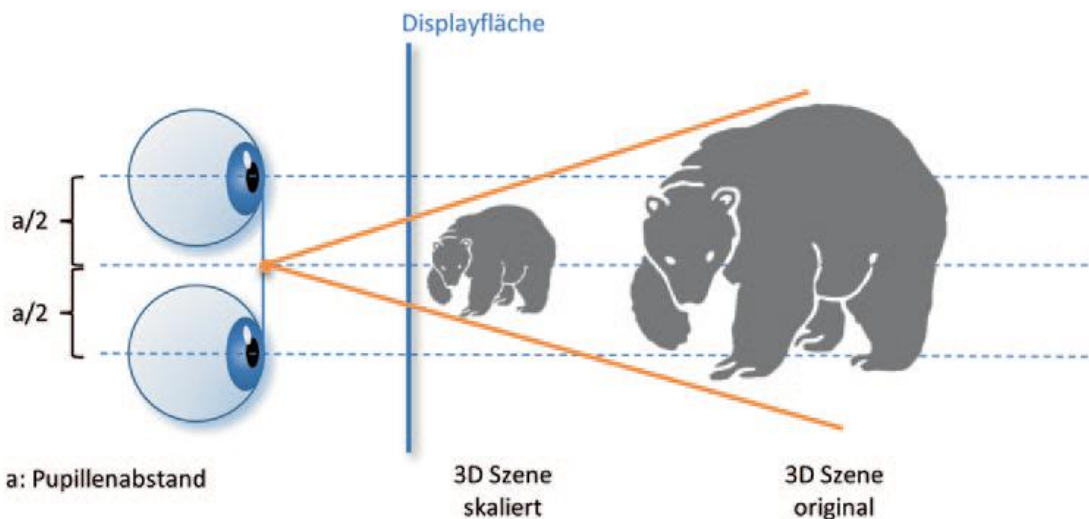


Abbildung 4 - Zyklopische Skalierung, [23]

Die Zyklopische Skalierung lässt sich mit der Manipulation der Virtual Eye Separation kombinieren, dabei sollte darauf geachtet werden, die Skalierung zuerst durchzuführen. Die Skalierung ist nicht nur sinnvoll, um eine räumlich zu ausgedehnte virtuelle Welt in den Panumbereich zu verschieben. Sie kann ebenfalls dazu verwendet werden, eine virtuelle Welt, welche den begrenzten Bereich des Stereodisplays nicht komplett ausnutzt, durch Auseinanderziehen räumlich plastischer darzustellen (vgl. [15]).

2.2.2.3. Diskrepanzen in der Raumwahrnehmung

Anwendungen aus dem Bereich Architektur, CAD, Stadtvisualisierung, Training, Simulation oder Medizin stellen in der Regel dreidimensionale Räume dar, bei denen es essentiell ist, dass der virtuell dargestellte Raum korrekt vom Betrachter wahrgenommen wird (vgl. [15]). Dies ist notwendig, damit die Nutzer Rückschlüsse auf ihre Handlungen und Entscheidungen in der realen Welt ziehen können. Diskrepanzen zwischen der Wahrnehmung von Größe und Distanzen in der virtuellen und der realen Welt sind deshalb in diesen Anwendungen besonders kritisch. Für viele Anwendungsbereiche von VR ist die korrekte Wahrnehmung von Größen und Distanzen also wesentlich (vgl. [15]). Dennoch haben Untersuchungen gezeigt, dass es insbesondere bei der Wahrnehmung des virtuellen Raumes immer wieder zu Diskrepanzen kommen kann. Dabei wurde gezeigt, dass die Nutzer Distanzen in der virtuellen Welt um bis zu 50 % unterschätzen (siehe [27], [28]). In diesen Untersuchungen ist der Einfluss von einigen Faktoren, wie beispielsweise stereoskopische Darstellung, beschränkter Sichtbereich, realistische Beleuchtung oder Schattierungen, auf diese Distanzunterschätzung aufgezeigt worden. Es gibt aber bis heute keine vollständige Erklärung für dieses Phänomen (vgl. [15]). Das Gesetz von Emmert ([29]) besagt, dass es einen klaren Zusammenhang von Größen und Distanzen gibt. Die wahrgenommene Größe verhält sich proportional zum Produkt aus wahrgenommener Distanz und Größe des Bildes auf der Netzhaut. Dadurch lässt sich das Phänomen der Distanzunterschätzung auch anhand einer Überschätzung von Größen beobachten. Andere Untersuchungen haben gezeigt, dass auch die Präsenz einen Einfluss auf die Wahrnehmung von Distanzen hat. Je präsenter sich der Nutzer in einer virtuellen Welt fühlt, desto besser wird seine Einschätzungen von Distanzen (siehe [27], [30]).

Es existieren verschiedene Ansätze um die Einschätzung von Distanzen bzw. Größen in der virtuellen Welt und dem dargestellten Raum oder den darin dargestellten Objekten zu verbessern (vgl. [15]). Beispielsweise könnte die gesamte Geometrie um einen Faktor skaliert werden. Dadurch würde der Nutzer den Raum zwar eher so wahrnehmen, wie er ihn in der realen Welt wahrnehmen würde, dennoch ist das Problem damit nicht gelöst (vgl. [15]). Ein ähnlicher Effekt lässt sich beispielsweise durch eine Vergrößerung des geometrischen Sichtbereichs (engl. geometric field of view) erzielen (vgl. [15]). Der geometrische Sichtbereich bezeichnet den von der virtuellen Szene dargestellten Bereich, der durch den horizontalen und vertikalen Öffnungswinkel der virtuellen Kamera definiert wird. Vergrößert man diesen, sieht der Betrachter einen größeren Bereich der virtuellen Welt. Es wird allerdings immer noch die gleiche Bildschirmfläche verwendet, weshalb dieser vergrößerte Bereich immer noch auf der gleichen Bildschirmfläche dargestellt wird. Das hat zur Folge, dass die Szene verkleinert wird und Objekte weiter entfernt erscheinen [19]. Dies ist in Abbildung 5 dargestellt. Durch Veränderung des Augenabstandes lassen sich ähnliche Effekte erzielen. Nachteil dieser Ansätze ist, dass durch perspektivische Verzerrung normalerweise ein anderer Raum dargestellt wird. Der Nutzer nimmt Größe und Distanz

jetzt zwar besser wahr, allerdings stimmt der dargestellte Raum nicht mehr mit dem realen Raum überein (vgl. [15]).

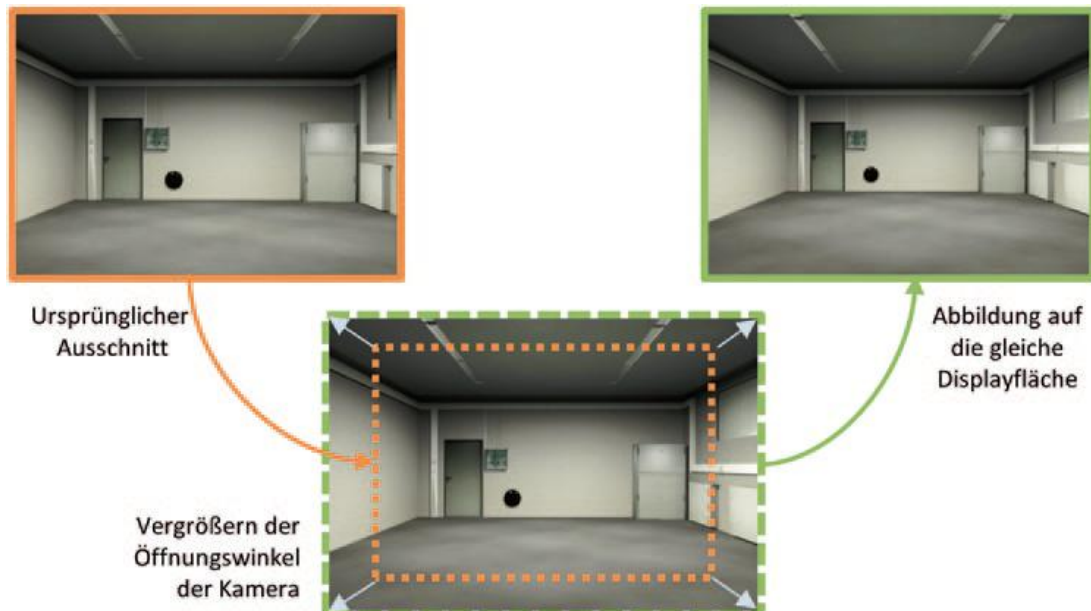


Abbildung 5 - Darstellung des gleichen virtuellen Raums mit (links) kleinem und (rechts) großem geometrischen Sichtbereichen, [28]

2.2.2.4. Diskrepanzen in der Bewegungswahrnehmung

Bei der Bewegungswahrnehmung ist ein ähnlicher Effekt wie bei der Distanzunterschätzung festzustellen. Dabei werden die Geschwindigkeiten der Bewegung oder die zurückgelegten Distanzen über- bzw. unterschätzt (vgl. [15]). Untersuchungen dazu haben ergeben, dass Vorwärtsbewegungen entlang der Blickrichtung unterschätzt werden (siehe [31], [32]). Insbesondere bei Bewegungen die nur visuell dargestellt werden und der Nutzer somit nur den optischen Fluss wahrnimmt. Selbst, wenn der Nutzer sich gleichzeitig bewegt und die Bewegung 1:1 auf die virtuelle Kamera abgebildet wird, unterschätzt der Nutzer die Vorwärtsbewegung entlang der Blickrichtung. Bei virtuellen Rotationsbewegungen hingegen kommt es häufig zu deren Überschätzung (siehe [33]). Diese Diskrepanzen in der Bewegungswahrnehmung lassen sich relativ einfach auflösen. Beispielsweise wird die Vorwärtsbewegungen leicht beschleunigt (ca. 5 bis 15%, [15]), damit sie vom Nutzer als korrekt eingeschätzt wird. Die Rotationsgeschwindigkeit sollte hingegen leicht (ca. 5 bis 10%, [15]) reduziert werden. Durch diese Manipulationen werden die virtuell dargestellten Bewegungen korrekt wahrgenommen, d. h. die visuell wahrgenommen Bewegungen passen zu dem vom Körper wahrgenommenen Bewegungseindruck. Tatsächlich führen die

Nutzer aber nun unterschiedliche Bewegungen in der virtuellen und realen Umgebung durch (vgl. [15]).

2.2.2.5. Cybersickness / Simulatorkrankheit

Ein weiteres Problem ist die Simulatorkrankheit, Bewegungskrankheit oder Cybersickness, welche durch Bewegung in der virtuellen Welt auftreten kann (vgl. [15]). Dabei werden dem visuellen Sinn durch eine virtuelle Welt Bewegungen vorgetäuscht, die denen des eigenen Körpers nicht exakt bzw. gar nicht entsprechen. Diese widersprüchlichen Informationen können zu Übelkeit, erhöhtem Speichelfluss, Benommenheit, Schwindelgefühl und sogar Erbrechen führen (vgl. [15]). Die genannten Symptome können vereinzelt bei Nutzern auftreten, die beispielsweise ein HMD verwenden. Sobald das HMD in Kombination mit einem Head-Tracking eingesetzt wird, kann das angezeigte Bild in Echtzeit zur Kopfbewegung verändert werden. Hierbei können die genannten Symptome auftreten, wenn die Bilddarstellung zu spät oder asynchron angepasst wird (vgl. [15]). Das kann beispielsweise der Fall sein, wenn das System eine zu hohe Latenz von der Kopfbewegung über die Verarbeitung bis zur Darstellung im HMD aufweist (vgl. [15]). Die Symptome lassen sich dadurch reduzieren, dass die Diskrepanzen zwischen den simulierten und tatsächlich empfundenen Bewegungen möglichst gering sind. Damit wird deutlich, dass eine geringe Latenz in der virtuellen Welt sehr wichtig ist. Allgemein sollten die Aufenthalte in der virtuellen Welt nicht zu lange dauern, da die Symptome erst nach einer bestimmten Zeit auftreten (ab ca. 10min, [1]). Es hat sich auch als nützlich erwiesen, den Nutzer langsam mit immer länger andauernden Aufenthalten in der virtuellen Welt an die Symptome zu gewöhnen (vgl. [34]).

2.3 Navigation in virtuellen Welten

Die Navigation ist in der virtuellen Welt eine universelle Interaktionsaufgabe und von zentraler Bedeutung. Damit ein hoher Grad an Immersion und Präsenz erreicht werden kann, muss der Nutzer sich möglichst einfach und natürlich in der virtuellen Welt bewegen können (vgl. [35]). Im Folgenden werden einige Interaktionstechniken zur Bewegungskontrolle in virtuellen Welten vorgestellt. Dabei wird zuerst geklärt, was Bewegungskontrolle ist, anschließend werden die Interaktionstechniken dazu vorgestellt.

2.3.1 Bewegungskontrolle

Bei der Bewegungskontrolle werden nur die grundlegenden Aktionen betrachtet, die benötigt werden, um Position und Orientierung des virtuellen Kameraausschnitts entsprechend der Bewegungen des Nutzers passend zu verändern (vgl. [35]). Dabei gelten Interaktionstechniken für die Bewegungskontrolle als besonders wichtig, da so gut wie jede virtuelle Umgebung es dem Nutzer erlauben muss, sich zumindest in dieser umzuschauen

oder besser noch sich in ihr fortbewegen zu können. Damit andere grundlegende 3D-Interaktionstechniken wie Manipulation oder Systemkontrolle erreicht werden, ist die Bewegungskontrolle ebenfalls notwendig. Ein Konstrukteur kann sich beispielsweise ohne zu einer bestimmten Stelle in einer virtuellen Welt zu gelangen, keinen Überblick über seinen neuen Fahrzeugentwurf verschaffen. Bowman (siehe [36]) definiert drei Aufgaben für die Bewegungskontrolle: Exploration, Suche und Manövrieren. Im Folgenden wird nur die Exploration vorgestellt, für Suche und Manövrieren siehe [36].

Explorationsaufgaben sind besonders in der Architekturvisualisierung, in 3D-Computerspielen und in der Informationsvisualisierung verbreitet, da der Nutzer dort kein konkretes Ziel verfolgt sondern die virtuelle Umgebung erkundend untersucht (vgl. [35]). Diese Aufgabe tritt oft zu Beginn einer Nutzung auf, um eine erste Orientierung zu ermöglichen. Dabei kann schon eine direkte Kontrolle über die virtuelle Kamera hilfreich sein, um interaktiv die Umgebung zu erkunden (vgl. [35]). Im Folgenden werden nun einige Interaktionstechniken zur Bewegungskontrolle in virtuellen Umgebungen betrachtet.

2.3.2 Steuerungstechniken zur Bewegungskontrolle

In vielen VR-Systemen werden virtuelle Steuerungstechniken verwendet, die durch die Angabe eines Richtungsvektors kontrolliert werden. Dabei sind etablierte 3D-Eingabegeräte wie Flystick oder Wand besonders für die handbasierte Steuerung geeignet, denn ihre Position und Orientierung im Raum kann besonders effizient von einem Tracking-System erkannt werden (vgl. [35]). Die Bewegung der virtuellen Kamera startet der Nutzer durch das Eingabegerät, das er in der Hand hält. Diese handbasierten Techniken sind einfach zu realisieren, haben aber den Nachteil, dass für die Bewegungskontrolle eine Hand genutzt werden muss, die dadurch nicht anderweitig verwendet werden kann. Auf der anderen Seite hingegen ist die blickgerichtete Steuerung das Grundprinzip vieler Ego-Shooter und anderer 3D-Computerspiele (vgl. [35]). Dabei rotiert der Spieler seinen virtuellen Avatar aus einer egozentrischen Kameraperspektive mit einem Eingabegerät in eine bestimmte Richtung und bewegt diesen anschließend mit einer bestimmten Geschwindigkeit in diese Richtung vorwärts. Die Bestimmung des Blickvektors kann in einer immersiven Umgebung mit Nutzer-Tracking durch Head-Tracking oder Gesichtserkennung ermittelt werden. Für den Nutzer ist dies einfach und natürlich anzuwenden, bringt aber den Nachteil mit sich, dass man sich nur in die Richtung, in die man schaut, bewegen kann (vgl. [35]). Blickvektor und Bewegungsrichtung können dadurch entkoppelt werden, dass die Hand oder der Körper zur Richtungsbestimmung verwendet wird.

Die Steuerung mit der Hand ist die Grundlage der „Camera-in-Hand“ Technik, welche jedoch den Nachteil hat, dass die Bewegung der Kamera mit der Hand für eine egozentrische Kameraperspektive gewöhnungsbedürftig ist (vgl. [35]). Da der Nutzer sich nicht selbst bewegen muss, werden die Bewegung in der virtuellen Welt von ihm nur visuell

wahrgenommen, weshalb ein Widerspruch zur Wahrnehmung der Eigenbewegung und dem Gleichgewichtssinn entsteht.

2.3.3 Walking als Technik zur Bewegungskontrolle

Das physikalische Laufen ist die am geeignetste Technik für die Bewegungskontrolle (vgl. [35]). Vorteil dieser natürlichen Technik sind die Bewegungshinweise, die das Gleichgewichtsorgan des Menschen bei echter Bewegung liefert. Das Laufen benötigt jedoch einen großen Interaktionsraum und da viele VR-Systeme nicht über diesen notwendigen Raum verfügen, gibt es im folgendem beschriebene alternative Abbildungen der realen Benutzerbewegung auf die virtuelle Kameraposition (vgl. [35]).

Ein Ansatz ist „Walking in Place“, bei dem sich der Nutzer auf der Stelle bewegt und die Geh- und Laufbewegungen von einem geeigneten Tracking-System verfolgt werden. Im Vergleich zu einer rein virtuellen Technik ohne Körperbewegungen erhöht „Walking in Place“ das Präsenzgefühl (vgl. [35]). Die reale Bewegung durch den Raum bietet jedoch eine nochmals höhere Präsenz (siehe [37]), wobei allerdings die Gefahr einer erhöhten Cybersickness (siehe Abschnitt 2.2.2.5) in Kauf genommen werden muss (siehe [38]).

Das „Redirected-Walking“ ist ein Verfahren, mit dem verhindert werden kann, dass sich Nutzer in HMD-basierten virtuellen Umgebungen schnell aus dem Tracking-Bereich bewegen (vgl. [35]). Dies wird dadurch erreicht, dass die Szene visuell langsam und für den Nutzer kaum bemerkbar verändert wird, während er sich physikalisch in eine Richtung bewegt. Der Nutzer passt sich unbewusst so an, dass er zum Beispiel glaubwürdig eine gerade Bewegung in der virtuellen Welt erfährt, obwohl er in der realen Welt im Kreis läuft. Der Nutzer gelangt jedoch bei dieser Technik immer noch zu schnell an den Rand des Tracking-Bereichs oder an Hindernisse, sodass die behutsame Gegenbewegung durch diese Technik nicht stark genug ist, um Probleme beim Tracking oder Kollisionen in der realen Welt zu vermeiden (vgl. [35]).

„Redirected Free Exploration with Distractors“ (siehe [39]) ist eine dreistufige Technik, die verhindern soll, dass sich Nutzer über Tracking-Grenzen hinaus bewegen oder mit realen Hindernissen kollidieren. Diese Technik bestimmt nach jedem Simulationsschritt die voraussichtliche Benutzerrichtung und rotiert die Szene für den Nutzer unbemerkt so, dass dieser im realen Raum in die Mitte des Tracking-Bereichs geleitet wird. Gelangt der Nutzer dennoch durch schnelle Bewegungen zu nah an die Tracking-Grenzen, wird daraufhin eine Ablenkung in der virtuellen Welt generiert. Im Beispiel von [39] fliegt ein Kolibri nah vor dem Nutzer und provoziert die benötigte Kopfbewegung. Reicht diese Ablenkung ebenfalls nicht aus, wird ein Hindernis in Form einer virtuellen Schranke eingeblendet, diese macht deutlich, dass es in dieser Richtung endgültig nicht mehr weitergeht.

Walking unterstützt als natürlichste Art der Bewegung in einer virtuellen Welt konsistent verschiedene physiologische Hinweise wie die Wahrnehmung der Eigenbewegung sowie den Gleichgewichtssinn (vgl. [35]). Die notwendige Größe des Tracking-Bereichs und die Behandlung von Kollisionen bei der Bewegung stellen dabei aber ein Problem dar. Es existieren einige Walking-Techniken, die nur für bestimmte Systeme geeignet sind, z.B. Redirected Walking für HMD's, oder die das Präsenzgefühl nur teilweise unterstützen wie Walking in Place. Mit Hilfe von preiswerten Tiefenkameras wie Microsoft Kinect [40] oder Asus Xtion [41] lassen sich bereits einfache Walking Interfaces kostengünstig realisieren.

2.4 Allgemeine Tracking Verfahren in geschlossenen Räumen

In diesem Abschnitt werden die allgemeinen Verfahren zur Positionsbestimmung in geschlossenen Räumen vorgestellt und auf ihre Anwendbarkeit hin untersucht sowie deren Vor- und Nachteile dargestellt. Dabei wird speziell auf die optischen Verfahren eingegangen.

2.4.1 Optisches Tracking

Optische Trackingverfahren haben sich in den letzten Jahren vermehrt durchgesetzt, weil sie eine hohe Genauigkeit bieten sowie einen flexiblen Einsatz ermöglichen (vgl. [42]). Es werden unterschiedliche Verfahren im Bereich des optischen Tracking verwendet. Allen optischen Verfahren liegt die Idee zugrunde, die relative Positionierung und Orientierung der Objekte zur Kamera (die sogenannten extrinsischen Kameraparameter) mit Hilfe von Objekten zu bestimmen, die im Videostream aufgenommen werden (vgl. [43]).

Diese Verfahren lassen sich in zwei Kategorien einteilen: Die markenbasierten Verfahren verwenden Marken zum Tracking, die anhand von Farbe, Form, Kontrast, Helligkeit oder Reflexionseigenschaften leicht im aufgenommenen Videostream erkennbar sind. Die markenlosen Verfahren die auch ohne Marken funktionieren. Desweiteren können diese Verfahren danach unterschieden werden, ob die Kameras von außen auf das zu überwachende Objekt gerichtet sind (Outside-In) oder ob die Kameras mit dem zu überwachenden Objekt verbunden sind und die Umgebung aufnehmen (Inside-Out) (vgl. [42]).

2.4.1.1. Markenbasierte Verfahren

Um die Berechnungskomplexität zu reduzieren und robuster gegen Fehleranfälligkeit gegenüber unterschiedlichen Beleuchtungssituationen zu sein, werden für optische Tracking-Verfahren oftmals klar spezifizierte Marken verwendet, deren Abbild schnell über Schwellwertfilter im Videostream gefunden werden kann (vgl. [42]). Dabei kann grundsätzlich zwischen aktiven und passiven Marken unterschieden werden, je nachdem, ob die Marken passiv das Licht reflektieren oder ob sie selbst aktiv Licht ausstrahlen.

Beim Einsatz von RGB-Kameras werden hierfür oftmals schwarzweiß Marken mit definierten Größen verwendet. Diese sind sehr zuverlässig, auch unter Verwendung von einfachen Kameras oder bei unterschiedlichen Beleuchtungssituationen, im Bild aufzufinden (vgl. [42]). Zudem sind die Marken sehr günstig herzustellen, meist reicht das Ausdrucken, sodass sie massentauglich sind.

Für den Einsatz im Bereich der VR werden oftmals Infrarotkameras verwendet. Hier kommen entweder passive Reflektoren in Verbindung mit Infrarotleuchten oder aktive Infrarot-LED als Marken zum Einsatz (vgl. [42]). Dadurch sind im Videostream die Marken als sehr helle runde Bereiche zu sehen. Anhand von mehreren Kameraansichten kann so die dreidimensionale Position einer sichtbaren Marke bestimmt werden.

Für den Fall, dass das Tracking nur die Position liefert, ist eine einzelne Marke ausreichend. Demgegenüber ist ein Rigid Body notwendig, bei Tracking-Systemen auch Target genannt, um Position und Orientierung zu berechnen. Demzufolge setzt sich ein Target aus mehreren Einzelmarken zusammen (vgl. [42]). Dem Tracking-System muss mit einem Kalibrierungsschritt der geometrische Aufbau der Targets vermittelt werden. Falls sich alle Targets in ihrem geometrischen Aufbau unterscheiden, kann die Identifikation durch diese Eigenschaft erfolgen, somit entfällt der Kalibrierungsschritt.

Aktive Marken verwenden Infrarot-LED's und müssen zu den Kameras synchronisiert werden. Es erfolgt eine Synchronisation über den IR-Blitz der Kameras, die von den Marken in Richtung der Kameraobjektive reflektiert werden (vgl. [42]). Die reflektierte Strahlung wird von Kameras, die ein bestimmtes Gebiet abtasten, in einem Graustufenbild registriert. Die Bilddaten werden in der Kamera vorverarbeitet und liefern 2D-Markenpositionen mit hoher Genauigkeit unter Verwendung von optimierten Mustererkennungsalgorithmen für die Kreisdetektion (vgl. [42]). Es ist erforderlich, dass mindestens zwei Kameras dasselbe Gebiet gleichzeitig abtasten, damit die Koordinaten eines Targets im Raum bestimmt werden können (vgl. [42]). Dementsprechend sind größere Volumina mit mehreren Kameras auszurüsten.

2.4.1.2. Markenlose Verfahren

Das langfristige Ziel beim Tracking ist es, ein zuverlässiges und genaues Tracking auch ohne Marken zu ermöglichen, um die Verwendung so angenehm wie möglich zu gestalten (vgl. [42]). Eine Technik dafür ist die Extraktion von Eigenschaften (z.B. Kanten, Ecken oder anderen Auffälligkeiten) aus dem Videostream und die Verfolgung über die einzelnen Bilder des Videostreams hinweg (vgl. [42]). Die Verwendung von Farb- und Tiefenkameras in Form von RGBD-Kameras ist eine andere Möglichkeit (vgl. [42]), welche auf Grund des großen Erfolges der ersten Kinect Generation (Eingabegerät für eine Spielekonsole) bekannt wurde (siehe [40]). Für RGBD-Kameras existieren zwei Techniken zur Tiefenerkennung: Zum einen ein Laufzeitverfahren, bei dem die Laufzeit des reflektierten Lichts bestimmt wird. Zum anderen ein Projektionsverfahren, bei dem von der Kamera aus ein Infrarot Muster in die Umgebung projiziert wird.

2.4.1.3. Geometriebasiertes Tracking

Das geometriebasierte Tracking extrahiert aus dem Kamerabild Kanten und/oder Eckpunkte aller Objekte (siehe beispielsweise [44]). Auf Basis einer Extrapolation der aus dem vorangegangenen Kamerabild extrahierten Transformation werden die Abstände zwischen den Kanten und Ecken des errechneten und des aktuellen Bildes als Grundlage für die Veränderung der Transformation verwendet. Dieses Verfahren eignet sich insbesondere im Bereich gleichförmiger geometrischer Formen, speziell wenn der Bereich wenig andere Merkmale zur Extraktion aufweist (vgl. [42]).

2.4.1.4. Weitere merkmalsbasierte Tracking-Verfahren

Im Unterschied zu Ecken und Kanten lassen sich andere visuelle Merkmale für den menschlichen Betrachter häufig nicht ohne weiteres erkennen (vgl. [42]). Sie bieten allerdings den Vorteil, dass sie sich durch Merkmaldetektoren sowohl schnell als auch zuverlässig in einem Kamerabild finden lassen.



Abbildung 6 - merkmalsbasiertes Tracking mit SURF, [70]

Dazu ist aber notwendig, dass sich genügend solcher Merkmale aus dem Kamerabild extrahieren lassen. Um im Anschluss basierend auf ihrer individuellen Beschreibung, dem sogenannten Deskriptor, diese mit vorhandenen Beschreibungen der Merkmale der zu verfolgenden Geometrie vergleichen zu können (vgl. [42]). Verwendete Detektoren für merkmalsbasiertes Tracking sind u. a. *SIFT – Scale Invariant Feature Transform* (siehe [45]) und *SURF – Speeded Up Robust Features* (siehe [46]). Diese Detektoren sind sowohl unempfindlich gegenüber Koordinatentransformationen wie Translation, Rotation und Skalierung als auch robust gegen Beleuchtungsvariation, Bildrauschen und geringere geometrische Deformation höherer Ordnung (wie sie zum Beispiel durch projektive Abbildung eines Objekts von verschiedenen Standpunkten im Raum entstehen). Wie in Abbildung 6 zusehen ist, obwohl der König stark verdeckt ist, hat SURF ihn im Bild anhand der Merkmale (blaue und rote Kreise) identifiziert.

2.4.1.5. Outside-In Verfahren

Bei einem Outside-In Verfahren wird die Szene von außerhalb des Interaktionsbereiches durch eine oder mehrere Kameras aufgenommen. Aus den daraus resultierenden Videostreams werden dann die Tracking-Daten gewonnen (vgl. [42]). Es werden hierzu meist mehrere Kameras in Kombination verwendet, um den Interaktionsbereich zu vergrößern bzw. weniger anfällig gegenüber Verdeckungen zu sein. Mit Hilfe von Testobjekten, deren Form und Größe bekannt sind und die durch den überwachten Raum bewegt werden, erfolgt die Kalibrierung der Kameras. Die Koordinatensysteme der einzelnen Kameras werden durch die so gewonnen Testdaten aufeinander abgeglichen, sodass verfolgte Objekte in einem einheitlichen Koordinatensystem beschrieben werden können (vgl. [42]).

Durch Triangulation der Daten, welche die Kameras an einen zentralen Tracking-Controller senden, werden die dreidimensionale-Position der Marken bzw. die sechsdimensionalen-Daten (Position und Rotation) der Targets berechnet. Damit die Tracking-Software die Triangulation durchführen kann, benötigt sie die exakten Positionen und Orientierungen der Kameras (vgl. [42]). Die Genauigkeitsanforderung in einem typischen VR-System ist hierfür < 1 mm bei der Position und $< 0.1^\circ$ im Winkel. Um die Position und Orientierung mit dieser Präzision zu bestimmen, stellt die Tracking-Software einen einfachen Kalibrierungsschritt zur Verfügung, der aus der Photogrammetrie abgeleitet ist (vgl. [43]) und der das Einmessen in einem kurzen Arbeitsschritt erlaubt. Die Kameras können mit Objektiven unterschiedlicher Brennweiten ausgerüstet werden, um eine Variation in den Öffnungswinkeln zu erreichen. Dadurch ist es möglich, eine ideale Abdeckung des Interaktionsbereichs vom Tracking Volumen zu gewährleisten (vgl. [42]).

Der Nachteil von Outside-In Verfahren ist, dass viele Kameras benötigt werden, um einen großen Interaktionsraum zu überwachen. Dadurch können beim Einsatz von

Spezialkameras die Gesamtkosten schnell steigen. Andererseits muss der Nutzer keine schweren Kameras inklusive deren Auswerteelektronik mit sich führen. Dennoch muss er die Marken in Kombination mit markenbasierten Verfahren am Körper tragen.

2.4.1.6. Inside-Out Verfahren

Bei Inside-Out Verfahren werden die Kameras an diejenigen Objekte angebracht, deren Bewegung erfasst werden soll (vgl. [42]). Beispielsweise am Kopf eines Nutzers zur Umsetzung eines Head-Tracking. Die so aus dem Videostream aufgenommene Umgebung wird dazu verwendet, die Position und Orientierung der Kamera zu einem oder zu mehreren Referenzpunkten in der Umgebung zu bestimmen. Der Nutzer ist damit nicht auf einen bestimmten Interaktionsraum beschränkt und kann sich somit freier bewegen (vgl. [42]). Dadurch muss der Nutzer allerdings auch in Kauf nehmen, dass er Kameras mit sich führen muss.

2.4.1.7. Vergleich der optischen Tracking-Systeme

Ein markenloses Outside-In Verfahren wäre für den Nutzer natürlich wünschenswert, weil damit die Einschränkungen für ihn am geringsten sind (vgl. [42]). Der Nutzer benötigt keine Marken an seinem Körper und muss nichts in den Händen halten. Dadurch kann er sich frei im Interaktionsraum bewegen. In der Praxis hat sich allerdings herausgestellt (vgl. [42]), dass die Präzision bei markenbasierten Systemen höher ist, da markenlose Systeme gegenüber markenbasierten anfälliger gegen Störungen sind (wie z.B. weitere Personen im Raum oder sich abwechselnde Lichtverhältnisse). Hinzukommt, dass der Interaktionsraum von Outside-In Verfahren durch die Kamerapositionen begrenzt ist. Mit Hilfe zusätzlicher Kameras kann dieser zwar vergrößert werden, dennoch bedeutet ein größerer Interaktionsraum entweder Mehrkosten für zusätzliche Kameras oder, auf Grund des größeren Abstandes zu den Kameras, weniger Präzision.

2.4.2 Akustisches Tracking

Tracking-Systeme, die mit Akustik arbeiten, nutzen die Unterschiede in der Laufzeit (Time of Flight Verfahren) oder in der Phase von Schallwellen (vgl. [42]). Hier wird für den Nutzer ein nicht hörbarer Ultraschall verwendet. Mit Hilfe eines Senders und eines Empfängers, wobei einer von beiden mit dem zu überwachenden Objekt verbunden ist, führt die Messung der Laufzeit zwischen ihnen zur Bestimmung des Abstandes. Zur Bestimmung der Position wird ein Aufbau mit einem Sender und drei Empfängern oder entsprechend drei Sendern und

einem Mikrofon benötigt (vgl. [42]). Sollen Position und Orientierung bestimmt werden, werden drei Sender und drei Empfänger benötigt.

Akustische-Tracking-Systeme sind im Vergleich zu anderen 3D-Tracking-Systemen relativ günstig (vgl. [42]). Ein Nachteil ist allerdings, dass das akustische Tracking sehr empfindlich auf Temperatur- oder Luftdruckänderungen reagiert. Jede Änderung kann eine Neukalibrierung des Systems erforderlich machen.

2.4.3 Elektromagnetisches Tracking

Um ein elektromagnetisches Tracking zu ermöglichen, werden mit Hilfe von stromdurchflossenen Spulen Magnetfelder aufgebaut (vgl. [42]). Damit die Position und Orientierung einer Person im Magnetfeld bzw. im Raum bestimmt werden kann, werden ebenfalls Spulen als Sensoren eingesetzt. Diese Spulen messen die induzierten Ströme der Magnetfelder. Deshalb sind sich über die Zeit ändernde Magnetfelder nötig, damit in den Empfängerspulen fortwährend Strom induziert wird. Der Aufbau von drei orthogonal zueinander stehenden Sendern und ebenfalls drei orthogonal stehenden Empfangsspulen erlaubt die Bestimmung von Position und Orientierung im Raum.

Bei elektromagnetischen Tracking-Systemen sind die Empfänger klein, Verdeckungen von Nutzern oder anderen nichtleitenden Objekten sind unproblematisch, und das System ist üblicherweise einfach zu nutzen (vgl. [42]). Demgegenüber dürfen in Räumen, in denen diese Systeme verwendet werden, wenn möglich keine ferromagnetischen Materialien und keine elektromagnetischen Felder existieren, da diese das Magnetfeld stören und somit die Präzision verringern (vgl. [42]).

2.4.4 Inertial Tracking

Beim Inertial-Tracking werden mit Hilfe von Inertialsensoren, auch Trägheits- oder Beschleunigungssensoren genannt, Beschleunigungskräfte gemessen (vgl. [42]). Das Verfahren wird insbesondere zur Bestimmung der Orientierung eingesetzt, beispielsweise bei der Erfassung von Gelenkstellungen eines Nutzers durch Anbringen der entsprechenden Sensoren auf den einzelnen Gliedmaßen. Zur Positionsbestimmung können lineare Inertialsensoren eingesetzt werden. Mit Hilfe der linearen Beschleunigungswerte durch drei senkrecht zueinanderstehende Sensoren kann die aktuelle Geschwindigkeit und damit die Positionsänderung bestimmt werden (vgl. [42]). Durch Messungenauigkeiten kommt es jedoch häufig zu Drifteffekten, welche meist durch eine verhältnismäßig geringe Genauigkeit bei der Umwandlung der analogen Messwerte in digitale Werte entstehen (vgl. [42]).

Bewegt man beispielsweise einen Sensor aus dem Ruhezustand und bremst ihn bis zum Stillstand ab, müssten die Summen der erfassten Beschleunigungswerte sowie auch die errechneten Geschwindigkeitswerte am Ende null ergeben(vgl. [42]). In der Regel ist dies jedoch nicht der Fall, da auch im Ruhezustand eine geringe Restgeschwindigkeit gemessen wird. Diese Restgeschwindigkeit lässt sich auf die Messungenauigkeiten der Sensoren zurückführen. Dadurch kommt es zu einer zunehmenden Abweichung zwischen der tatsächlichen und der gemessenen Position.

3 Analyse von Produktlösungen zur Positionsbestimmung

In diesem Abschnitt werden bereits vorhandene Produktlösungen zur Positionsbestimmung in geschlossenen Räumen auf die Anforderungen eines für REKNOW in Frage kommenden Tracking-Systems untersucht. Zuerst werden die Kriterien, nach denen die Analyse erfolgen soll, vorgestellt. Danach wird auf die Anforderungen an das Tracking-System für REKNOW eingegangen. Zum Schluss werden die vorhandenen Produktlösungen vorgestellt und bzgl. der Kriterien untersucht sowie bewertet.

3.1 Untersuchungskriterien

Für einige Kriterien, wie beispielsweise die Genauigkeit, lassen sich klar quantifizierte Werte bestimmen, für die Usability hingegen ist dies nicht so einfach möglich.

3.1.1 Anzahl der Freiheitsgrade pro verfolgtem Körper

Je nach Tracking-System unterscheidet sich die Anzahl der bestimmaren Freiheitsgrade pro verfolgtem Körper. Im Idealfall ist die Bestimmung aller sechs Freiheitsgrade durch ein Tracking-System möglich. Es werden aber auch Systeme verwendet, die nur die Position bestimmen, was gleichbedeutend mit den drei Freiheitsgraden der Translation ist (vgl. [47]). Systeme die nur die Orientierung bestimmen, sind gleichbedeutend mit den drei Freiheitsgraden der Rotation (vgl. [47]).

3.1.2 Anzahl der gleichzeitig verfolgten Körper

Es existieren Anwendungsfälle, bei denen darauf geachtet werden muss, wie viele Objekte gleichzeitig vom Tracking-System verfolgt werden sollen (siehe [47]). Werden mehrere Körper verfolgt, ist es hilfreich, diese anhand eindeutiger ID's identifizieren zu können. Desweiteren ist es hilfreich, wenn die ID auch dann erhalten bleibt, wenn der Körper vorübergehend außerhalb der Überwachung war.

3.1.3 Größe des überwachten Bereiches

Die Größe des überwachten Bereiches bzw. des überwachten Volumens kann sich je nach Tracking-System unterscheiden. Das ausgewählte Tracking-System sollte einen den Anforderungen entsprechend großen Bereich zur Verfügung stellen. Dieser Bereich kann in Armreichweite liegen oder den Bewegungen eines Kopfes vor dem Monitor entsprechen. Es existieren aber auch Anwendungsfälle, in denen ein Herumlaufen möglich sein soll. Der Grund für Größenbeschränkungen bei kamerabasierten Verfahren liegt in der Auflösung und der Anzahl der Kameras (vgl. [47]).

3.1.4 Genauigkeit

Meistens ist eine hohe Genauigkeit wünschenswert. Diese kann jedoch durch physikalische Einschränkungen der Tracking-Systeme nicht immer erreicht werden (siehe. [47]). Oftmals ist es eine Frage des Kostenaufwands. Ein Kameraaustausch kann beim optischen Tracking die Genauigkeit erhöhen. Das wiederum kann den Preis leicht um den Faktor 100 steigern, wenn statt einer einfachen Webcam eine teure Industriekamera verwendet wird (vgl. [47]). Wie hoch die Genauigkeit sein soll, hängt vom verfügbaren Etat und dem Anwendungsfall ab. Die Spannbreite der Ortsauflösung beginnt bei einer Genauigkeit (von wenigen Millimetern) für beispielsweise Finger-Tracking und geht bis hin zu Ungenauigkeiten von mehreren Metern wie bei der Nutzung von GPS.

3.1.5 Wiederholrate

Die Wiederholrate beschreibt das zeitliche Auflösungsvermögen eines Tracking-Systems (vgl. [47]). Die entsprechenden Freiheitsgrade werden in zeitdiskreten Schritten bestimmt. Die Anzahl dieser Messpunkte pro Sekunde wird als Wiederholrate bezeichnet. Dabei ergibt die Verfolgung der realen kontinuierlichen Bewegung eines Körpers entsprechende Messpunkte.

3.1.6 Latenz

Jedes Tracking-System benötigt eine gewisse Reaktionszeit. Beispielsweise weil es Zeit bis zur nächsten Abtastung abwartet, die durch Laufzeiten von Signalen in Kabeln oder durch die Verarbeitung von Daten durch Algorithmen entsteht. Diese Verzögerung wird Latenz genannt.

3.1.7 Empfindlichkeit gegenüber äußeren Rahmenbedingungen

Die Rahmenbedingungen müssen je nach verwendetem Tracking-System unterschiedlich betrachtet werden. Bei optischen Tracking Verfahren kann die Beleuchtung und die Möblierung des Raums, in dem das VR-System aufgebaut ist, Einfluss auf die Verfolgung haben (vgl. [47]). Gerade bei optischen Tracking Verfahren kann eine gleichmäßige Beleuchtung von großem Vorteil sein, um die Merkmale im Videostream besser extrahieren zu können.

3.1.8 Kalibrierung

Bei einigen Tracking-Systemen ist eine Kalibrierung erforderlich, welche die Messwerte mit einem gegebenen Modell abgleicht (vgl. [47]). Diese Kalibrierung ist notwendig, um die Maße in der virtuellen Welt den verwendeten realen Objekten, die verfolgt werden, anzupassen. Zur Kalibrierung gehört unter anderem auch die Bestimmung von Abbildungsfehlern, welche bei optischen Verfahren durch Linsenverzerrung des Objektivs auftreten.

3.1.9 Usability

Inwieweit der Nutzer durch das Tracking-System eingeschränkt wird, ist entscheidend für eine Anwendung (vgl. [47]). Es kann beispielsweise notwendig sein, Marken, Handschuhe oder Sensoren am Körper anzubringen, damit ein Tracking möglich wird. Ebenfalls ist es für die Verwendung ein Unterschied, ob das Tracking-System kabelgebunden ist oder mittels Funktechnologien arbeitet. Zusätzlich hat die Raumgröße einen Einfluss darauf, ob der

Nutzer gedanklich in die Anwendung eintauchen kann. Der Nutzer sollte nicht dauernd darauf achten müssen, den vorgegebenen Interaktionsbereich innerhalb des Raumes nicht überschreiten zu dürfen. Desweiteren existieren Tracking-Systeme bei denen es notwendig ist, dass der Nutzer ständig zu ihnen hin orientiert ist, um ein Tracking zu ermöglichen.

3.2 Anforderungen des REKNOW Tracking-Systems

Das REKNOW Tracking-System soll in Bezug auf die Anzahl der Freiheitsgrade sowohl Position als auch Orientierung des Nutzers im Raum bestimmen können. Die Position wird dazu verwendet, um sich innerhalb der Simulation fortbewegen zu können. Die Orientierung dient dazu den Blickvektor der virtuellen Kamera und die Bewegungsrichtung des Nutzers zu entkoppeln. Diese Entkopplung sorgt dafür, dass der Nutzer des Tracking-Systems in der Lage ist gerade zu gehen während er sich gleichzeitig dabei umschaun kann. Die Grundlage hierfür wurde in Abschnitt 2.3.2 *Steuerungstechniken zur Bewegungskontrolle* erläutert. Es ist am Anfang erst einmal ausreichend, wenn das System nur einen Nutzer verfolgen kann. Mit Ausblick auf die Zukunft sollte es jedoch das Ziel sein, mehrere Nutzer mit nur einem Tracking-System verfolgen zu können. Diese sollen dann eindeutig über ID's identifiziert werden können, damit sich die Positionsinformationen korrekt zuordnen lassen. Für die Größe des überwachten Tracking-Bereiches gilt: je größer desto besser, denn die CAD-Modelle der Nutzfahrzeuge sollen in der virtuellen Welt nach Möglichkeit im Maßstab 1 : 1 abgebildet werden. Um geringe Bewegungen mit dem REKNOW Tracking-System verfolgen zu können, muss insbesondere die Genauigkeit sehr hoch sein. Jegliche Bewegung, die zu einer Positionsveränderung führt bzw. in der realen Welt durchgeführt wird, soll nach Möglichkeit ebenfalls in der virtuellen Welt abgebildet werden. Die Wiederholrate muss auch hoch sein, damit die Bewegungen in der virtuellen Welt flüssig abgebildet werden können und Änderungen der Bewegungsrichtung sofort erfasst werden. Die Latenz des gesamten Systems muss so gering sein, dass am Ende keine Verzögerungen, die menschliche Wahrnehmung beeinflussen oder andere beeinflussende Phänomene auftreten (siehe 2.2.2.4 Diskrepanzen in der Bewegungswahrnehmung und 2.2.2.5 Cybersickness). Hinzukommt das der Nutzer ein HMD auf hat, bei dem die Darstellung der virtuellen Welt in Echtzeit erfolgt. Er ist dadurch nicht mehr in der Lage seinen eigenen Körper wahrzunehmen sondern nur noch die dynamisch generierte virtuelle Welt. Hier wirkt sich eine hohe Latenz besonders negativ auf das Wohlbefinden des Nutzers aus und es kann zu Schwindelgefühlen kommen (siehe [48]). Die empfohlene Gesamtlatenz für ein Tracking-System mit HMD liegt bei unter 50 ms (siehe [49], [50]). Die Gesamtlatenz setzt sich aus der Tracking Latenz, Transportlatenz, Simulationslatenz, Generierungslatenz und schließlich der Darstellungslatenz zusammen. Deshalb muss die Tracking Latenz nach Möglichkeit zwischen 10-16ms liegen. Um perfekte Rahmenbedingungen für das REKNOW Tracking-System herzustellen, sollte es in geschlossenen Räumen mit gleichmäßiger Beleuchtung verwendet werden. Ebenfalls sollten die Räume keine Möblierung aufweisen, damit es innerhalb des überwachten Tracking-Bereiches zu keinen Kollisionen der Nutzer mit Möbelstücken kommen kann. Nach dem Aufbau sollte das System lediglich einmal

kalibriert werden müssen, auch wenn die Nutzer sich zwischendurch abwechseln. Somit muss nicht für jeden Nutzer ein extra Kalibrierungsschritt durchgeführt werden, damit das System diesen erkennt. Speziell bei der Usability ist es wichtig, dass der Nutzer so wenig wie möglich anlegen oder mit sich führen muss, damit ein reibungsloses Tracking ermöglicht werden kann.

3.3 Die Produktlösungen

Die hier vorgestellten Produktlösungen wurden im Rahmen einer Internet-Recherche ermittelt. Die Analyse erfolgt auf Basis der Informationen der Internetseiten und den Produktbeschreibungen der Hersteller.

3.3.1 ImmerSight

ImmerSight [51] ist ein Tracking-System zur Positionsbestimmung, das mit jedem Head-Mounted Display verwendet werden kann. Das System verwendet eine einfache Webcam



Abbildung 7 - ImmerSight [51]

und ein fünfeckigen Ring, den der Nutzer auf seinem Kopf trägt, wodurch die Translation der Kopfbewegung erfasst und die Immersion der Virtual Reality verbessert wird. Der hierbei verwendete Sensorring wiegt ca. 100g, was dafür sorgt, dass der Nutzer ihn kaum wahrnimmt (vgl. [51]). Das optische Tracking erfolgt mit einer 60Hz Kamera, welche über dem Kopf des Nutzers wie eine Deckenlampe positioniert wird (vgl. [51]). So lassen sich die weißen Marken am Sensorring selbst optimal verfolgen. Die Kamera erkennt die Marken und die Form des fünfeckigen Sensorrings. Dies ermöglicht die Positions- und Orientierungsbestimmung ohne einen Rotationssensor. Der Algorithmus zur Umsetzung dieses Systems wird in [52] beschrieben. Der von der Kamera erfasste Interaktionsraum hat einen Kreisdurchmesser von 1,6 m und könnte in Abhängigkeit von der Deckenhöhe durch Erhöhen der Kameraposition noch vergrößert werden (vgl. [51]). Zusätzlich unterstützt das System einen Controller, der es erlaubt, sich mit der „Camera-in-Hand“ Technik (siehe Abschnitt 2.3.2 Steuerungstechniken zur Bewegungskontrolle) weiter in der virtuellen Umgebung zu bewegen, als der reale Interaktionsraum erlauben würde (vgl. [51]). Dadurch ist es möglich, auch größere virtuelle Welten zu erkunden. Nach der laut Hersteller ca. zehnminütigen Aufbau- und Einstellungsphase ist das System einsatzbereit (vgl. [51]). Das System bietet nur wenig Platz und dieser kann auch nur geringfügig vergrößert werden, folglich kann der Nutzer sich nicht allzu frei bewegen. Dennoch wird es mit der „Camera-in-Hand“ Technik möglich, größere virtuelle Welten zu erkunden. Somit können größere Distanzen zurückgelegt werden und eine Großbereich-Erkundung wird ohne weiteres möglich. Durch die realen Bewegungen in dem vergleichsweise kleinen Tracking-Bereich kann der Nahbereich in der virtuellen Welt erkundet werden. Somit lässt sich das System in Nahbereich- und Großbereich-Erkundung einteilen und ist damit ein guter Kompromiss. Die „Camera-in-Hand“ Technik kann für den Nutzer dennoch unangenehm sein: Der Nutzer ist zum einen in der Lage sich durch reales gehen und zum anderen mit Hilfe eines Kontrollers in der virtuellen Welt fort zu bewegen. Diese Optionen der Fortbewegung können zur Verwirrung der Wahrnehmung führen, wodurch ein erhöhtes Risiko der Cybersickness gegeben ist. Ein weiteres Problem ist, dass der Controller eine Hand bereits komplett blockiert, welche sich dann beispielsweise nicht mehr für Gesten Tracking verwenden lässt. Ebenso kann das System nur einen Nutzer verfolgen, da der Interaktionsraum für mehrere Nutzer zu klein ist. Hinzukommt das die Usability des Systems dadurch reduziert ist, dass der Nutzer einen fünfeckigen Sensor Ring aufsetzen muss, um das Tracking zu ermöglichen. Dieser wiegt nur 100g und stellt deshalb aber keine große Einschränkung dar. Außerdem ist das System sehr mobil und lässt sich schnell aufbauen und kalibrieren. Durch die Verwendung des fünfeckigen Sensorrings, welcher mit Hilfe eines geometrischen Tracking Verfahrens verfolgt wird, ist das System robuster im Hinblick auf äußere Rahmenbedingungen (siehe [52]).

3.3.2 WorldViz - Lightweight Stereoglass MotionTracker

Dieser kleine, leichtgewichtige, kabellose Motion Tracker kann mit allen auf dem Markt verfügbaren 3D-Video Brillen, einschließlich HMDs, verwendet werden (vgl. [53]). Dazu wird er an der 3D-Video Brille oder dem HMD angebracht (siehe Abbildung 9). Er muss in



Abbildung 10 - WorldViz PPT [54]



Abbildung 9 - WorldViz Motion Tracker [53]



Abbildung 8 - WorldViz Kalibrierungswerkzeug [54]

Verbindung mit dem WorldViz PPT (Precision Position Tracking, Abbildung 10) Motion-Capture System verwendet werden und ermöglicht es dem Nutzer, große virtuelle Welten zu erkunden (vgl. [53]). Das WorldViz PPT System bietet ein hoch präzises optisches Bewegungstracking, das ein Interaktionsbereich von mehr als 50m x 50m abdecken kann, und es erfüllt dabei noch die Echtzeit Anforderungen einer immersiven VR-Anwendung (vgl. [54]). Damit dieser Bereich abgedeckt werden kann, werden mehr als 8 Kameras benötigt. Wie viele Kameras genau dafür benötigt werden, ist den Herstellerangaben nicht zu entnehmen. Insgesamt unterstützt das System bis zu 32 Kameras. Die Genauigkeit bei der Positionsbestimmung liegt bei einem Fehler von unter 0,25mm und 0,09° in der Rotationsbestimmung (vgl. [54]). Das System lässt sich schnell und automatisch mit Hilfe eines Kalibrierungswerkzeuges (Abbildung 8, Quadrat mit 4 Led-Marken) kalibrieren (vgl. [54]). Dieses Quadrat sorgt dafür, dass der Tracking-Bereich und das Koordinatensystem präzise bestimmt werden können. Hierzu wird das Quadrat einmal vor der Inbetriebnahme in der Mitte des Tracking-Bereichs ausgelegt, um das System automatisch kalibrieren zu lassen (vgl. [54]). Der Motion Tracker besteht aus zwei aktiven LED Marken, die auf einer Linie mit den Pupillen des Nutzers liegen (vgl. [53]). Der Akku der aktiven LED Marker hat

eine Laufzeit von 6 Stunden(vgl. [53]). Der abgedeckte Interaktionsbereich eines Systemaufbaus mit 2 Kameras beträgt 15m x 15m(vgl. [53]). Die Anzahl der Freiheitsgrade dieses Systems beträgt 6, damit kann es Position und Orientierung des Kopfes im Raum bestimmen. Mit einem HMD-Brillenaufsatz kann nur eine Person verfolgt werden(vgl. [53]). Es lassen sich aber ohne weiteres mehrere dieser Brillenaufsätze parallel durch das Kamerasystem verfolgen. Den hierbei verwendeten aktiven LED-Marken am Brillenaufsatz können andere ID's zugewiesen werden, wodurch ein Tracking von mehreren Personen zur selben Zeit möglich wird(vgl. [53]). Die Größe des überwachten Bereiches liegt mit 4 Kameras bei 15m x 15m und kann bis auf 50m x 50m erhöht werden, indem bis zu 32 Kameras verwendet werden. Dies bringt allerdings enorme finanzielle Kosten mit sich. Die Genauigkeit des Systems ist sehr hoch. Der Fehler in der Positionsbestimmung liegt unterhalb von 0.25mm und in der Rotationsbestimmung unterhalb von 0.09°. Die Wiederholrate liegt bei 180Hz und ist damit ebenfalls sehr gut geeignet, um schnelle und ruckartige Bewegungsabläufe zu verfolgen (vgl. [54]). Die Latenz des Systems liegt bei 20ms und ist damit ebenfalls sehr gut (vgl. [54]). Da das System mit aktiven LED-Marken arbeitet, kann es auch bei wenig bis schlecht beleuchteten Räumen verwendet werden. Auch mit Verdeckung durch Möbel oder Personen kommt das System zurecht, solange sich die verfolgte Person im Sichtfeld zweier Kameras befindet. Damit ist das System beim Einsatz von vielen Kameras extrem unempfindlich gegen äußere Einflüsse. Die Kalibrierung des Systems erfolgt mit Hilfe des vorgestellten Kalibrierungswerkzeuges und ist dadurch sehr einfach. Nachdem das gesamte System aufgebaut wurde, muss das Kalibrierungswerkzeug einmal im Raum ausgelegt werden, um den Ursprung des Koordinatensystems zu kennzeichnen und um die Orientierung der XYZ-Achsen zu bestimmen. Die Usability des Systems ist ebenfalls sehr gut. Der Nutzer muss lediglich das HMD aufsetzen, auf dem der Motion Tracker Aufsatz installiert ist und von da an kann das System diese Person verfolgen. Es müssen also keine speziellen Anzüge oder sonstiges Equipment angelegt werden, um eine Bewegungsverfolgung zu ermöglichen. Die hohe Genauigkeit ist ein besonderes Merkmal dieses Systems. Der Akku der aktiven LED-Marken verbraucht Energie und muss bei einer Akkulaufzeit von ca. 6 Stunden regelmäßig aufgeladen werden.

4 Der Prototyp

In Anlehnung an die in Abschnitt 3 durchgeführte Analyse von Produktlösungen, wird in diesem Abschnitt ein eigenes und kostengünstiges Tracking-System zur Positionsbestimmung in geschlossenen Räumen entworfen, implementiert und auf seine Tauglichkeit hin untersucht. Abschließend wird dieses Tracking-System mit den im vorangegangenen Abschnitt vorgestellten professionellen Tracking-Systemen verglichen. Zum Schluss erfolgt die Auswahl eines für REKNOW geeigneten Tracking-Systems.

4.1 Motivation

Professionelle Tracking-Systeme, die zur Positionsbestimmung in geschlossenen Räumen verwendet werden, benutzen meist markenbasierte optische Infrarot-Verfahren (3.3.2 WorldViz - Lightweight Stereoglass MotionTracker). Diese Systeme können weitläufige Bereiche mit mehreren Benutzern durch eine große Anzahl von Kameras abdecken. Es existieren aber auch Anbieter, die mit Hilfe von Beschleunigungssensoren oder über Funk das Aufzeichnen von Bewegungen ermöglichen (siehe [55], [56]). Die hohen Preise verhindern jedoch den vermehrten Einsatz der Systeme in naher Zukunft. Zusätzlich benötigen diese Lösungen bestimmte Marken oder spezielle Anzüge sowie längere Aufbau- und Kalibrierungsphasen, welche wiederum Einschränkungen in der Usability für die Benutzer darstellen und damit die Einsatzbereitschaft und Akzeptanz solcher Systeme reduzieren.

Hier stellen markenlose optische Outside-In Tracking-Verfahren mit Tiefenkameras eine Alternative dar, welche mit einer Prototypen-Implementation in dieser Arbeit untersucht werden soll. Heutige Tiefenkameras sind mittlerweile zu günstigen Preisen bei dem „privaten Endkunden“ angekommen ([40], [41]). Sie ermöglichen das Aufzeichnen von menschlichen Bewegungen (engl. motion-capture) und die Bestimmung der Körperhaltung (engl. human pose estimation), was auch eine Form der Positionsbestimmung im Raum darstellt (wird im Weiteren nur noch als Full Body Tracking bezeichnet). Dafür benötigen sie jedoch keine Marken oder Anzüge, um den Menschen zu verfolgen. Allein mit Hilfe des Tiefenbildes werden die Bewegungen und die Körperhaltungen in Echtzeit verfolgt (basierend auf dem Verfahren in [57]). Es existieren sofort einsatzbereite Softwaremodule und Middleware-Komponenten, die zum Abrufen der Tiefenkamera-Daten und zur Positionsbestimmung, von einigen Tiefenkameraherstellern vertrieben werden [40], [41].

Die bereitgestellten Schnittstellen machen die einfache Integration in bestehende und neu zu entwickelnde Anwendungen möglich.

Diese Alternative bringt aber auch Einschränkungen mit sich. Zum Beispiel sind die Reichweiten dieser Kameras nicht groß, auch sind die Latenzen insgesamt größer. Werden mehrere Benutzer verfolgt und kommt es zu Überdeckungen bzw. Verdeckungen von Benutzern durch andere Benutzer, können diese nicht mehr eindeutig verfolgt werden. Auch die eindeutige Identifikation der Benutzer ist nicht immer gegeben. Diese Einschränkungen können aber mit Hilfe von mehreren Tiefenkameras, die einen Bereich verfolgen, reduziert werden.

Im Folgenden wird ein Tracking-System beschrieben, das versucht, mit mehreren Tiefenkameras einen vergrößerten Tracking-Bereich zu erschaffen, um damit einen Benutzer in einem Raum so verfolgen zu können, dass die Position und Rotation seines Körpers in eine virtuelle Welt abgebildet werden kann.

4.2 Systembeschreibung

Das System wird für einen leeren quadratischen Raum entworfen. Der Tracking-Bereich innerhalb des Raumes darf die maximale Größe von 8m x 8m nicht überschreiten, damit eine ausreichende Präzision gegeben ist. In den Ecken und entlang der Wände werden mindestens 2 und maximal 8 Tiefenkameras aufgestellt. Die Sichtfelder der Tiefenkameras müssen einen gemeinsamen überlappenden Bereich in der Mitte des Tracking-Bereiches bilden, damit die Tiefenkameras über ein Referenzobjekt, das sich in dem überlappenden Bereich befinden muss, synchronisiert werden können. Dies ist erforderlich, um die lokalen Koordinatensysteme der einzelnen Tiefenkameras in ein gemeinsames globales Koordinatensystem zu transformieren. Die Kalibrierung bzw. Synchronisation der Tiefenkameras nimmt in diesem Fall keine 10 Sekunden in Anspruch. Die Daten der einzelnen Tiefenkameras werden in einem zentralen Server gesammelt. Dort werden sie verarbeitet und ausgewertet, anschließend an einen Client ausgeliefert, der diese Positions- und Rotationsdaten der verfolgten Person einer weiteren Anwendung zur Verfügung stellt. Für den Anfang erlaubt das System, dass nur eine Person zurzeit verfolgt werden kann. Der Hintergrund dazu ist, dass es durch Verdeckung von anderen Personen zu Synchronisationsproblemen innerhalb des zentralen Servers kommt.

Wie in der Abbildung 11 zu erkennen ist, lässt sich der nutzbare Tracking-Bereich durch den Einsatz von mehreren Tiefenkameras erweitern. Der orangene Bereich stellt den überlappenden Bereich dar, in dem das Referenzobjekt für die Synchronisierung der Kameras positioniert werden muss. Die in der Abbildung verwendeten Kamera Kegel (Dreieckssymbole) entsprechen nicht den tatsächlichen Öffnungswinkeln sowie den Reichweiten der Tiefenkameras, sie sollen nur beispielhaft dazu dienen, die Idee zu skizzieren.

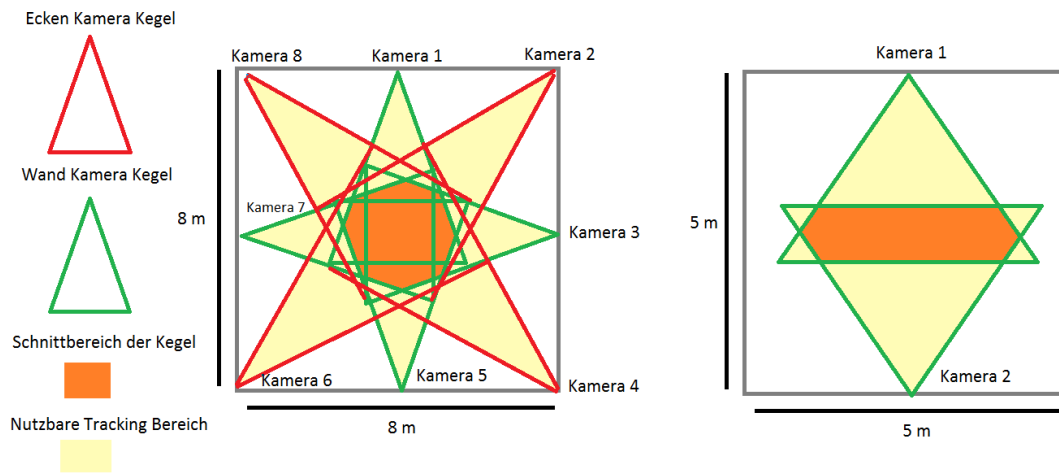


Abbildung 11 - Skizze der Systemidee, Kamerakegel entsprechen nicht den Öffnungswinkeln und Reichweiten der Tiefenkameras

4.3 Hardware und Software

Das System verwendet mehrere Asus Xtion [41] Tiefenkameras zusammen mit dem OpenNI [58] und Flexible Action and Articulated Skeleton Toolkit [59] (im Folgenden als FAAST bezeichnet). OpenNI ist eine Middleware Komponente und bietet auf unterster Ebene eine Schnittstelle für die Kommunikation mit Tiefenkameras. Es unterstützt sowohl Microsofts Kinect [40] als auch Asus Xtion [41]. Es bietet ebenfalls die Möglichkeit des Full Body Tracking anhand von Tiefendaten aus der Tiefenkamera. FAAST ist ein eigenständiges Programm, mit dem vordefinierte Gesten und Bewegungen von Menschen in Tiefenbildern erkannt werden können.

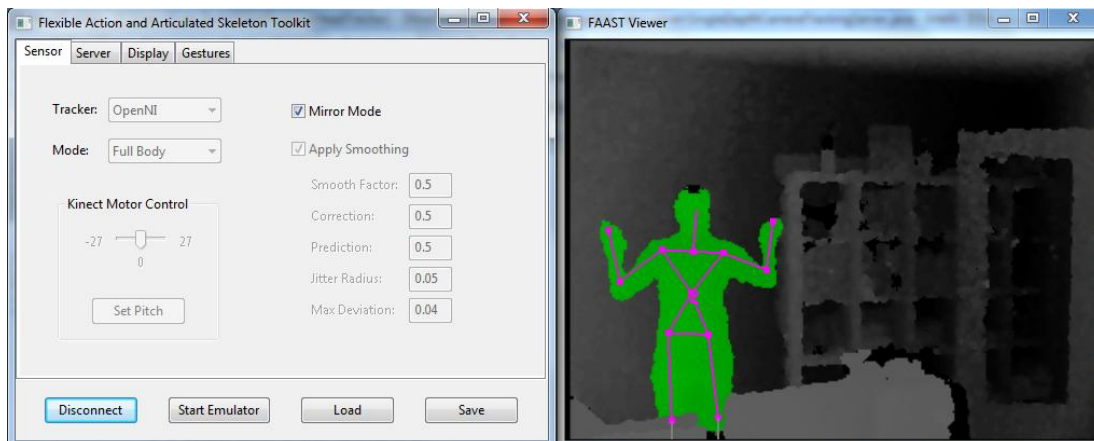


Abbildung 12 - FAAST mit aktivem Tracking Fenster, in dem eine Full Body Tracking zusehen ist

Es verwendet OpenNI für die Durchführung eines Full Body Tracking des menschlichen Körpers, damit im nächsten Arbeitsschritt die Gesten erkannt werden können. Nebenbei stellt es alle Informationen des Full Body Tracking sowie die Gesten durch einen Virtual Reality Peripheral Network Server [60] (im Folgenden als VRPN bezeichnet) im Netzwerk zur Verfügung. Somit werden die Daten für andere Programme nutzbar, um beispielsweise innovative Interaktionsformen zum bedienen des Computers oder anderer Programme zu realisieren. Im Kontext dieses Systems dient FAAST lediglich dazu, die Full Body Tracking-Daten einer Tiefenkamera im Netzwerk bereitzustellen, wodurch der Entwicklungsaufwand des Gesamtsystems reduziert wird. Hierdurch muss kein extra Programm für das System entwickelt werden, das die Kommunikation mit der Tiefenkamera übernimmt und diese Daten im Netzwerk bereitstellt.

Die Tiefenkameras verwenden nur ein Kameraobjektiv zur Verfolgung und funktionieren ohne zusätzliche Marken. Sie projizieren ein Punktmuster mit einem Infrarot Laser Projektor in die Umgebung, diese projizierten Punkte werden von der Kamera verfolgt. Der Projektor und die Kamera sind in einem bekannten Abstand zueinander in der Tiefenkamera verbaut. Durch die aufgezeichneten Punkte im Kamerabild und dem originalen Projektionsmuster kann ein Tiefenbild erstellt werden. Dieses wiederum wird verwendet, um ein menschlichen Körper zu erfassen und seine Körperhaltung zu bestimmen. Die Reichweite der Asus Xtion Tiefenkamera beträgt offiziell 3,5 m [61], aber wie einige meiner Tests gezeigt haben, liefern sie auch bei Reichweiten von bis zu 5 m noch zuverlässige Informationen. Die Kameras arbeiten bei einer Auflösung von 640 x 480 Pixeln mit 30 Bildern pro Sekunde, was einer Abtastrate von 30 Hz entspricht und damit eine Latenz von 33 ms mit sich bringt [61]. Auf Grund dessen, dass mehrere Tiefenkameras zum Einsatz kommen, werden mehrere Laptops mit Netzwerkanschluss benötigt. Hinzu kommt ein 100Mbit Switch und mehrere LAN-Kabel mit Reichweiten von 5 m bis 10 m.

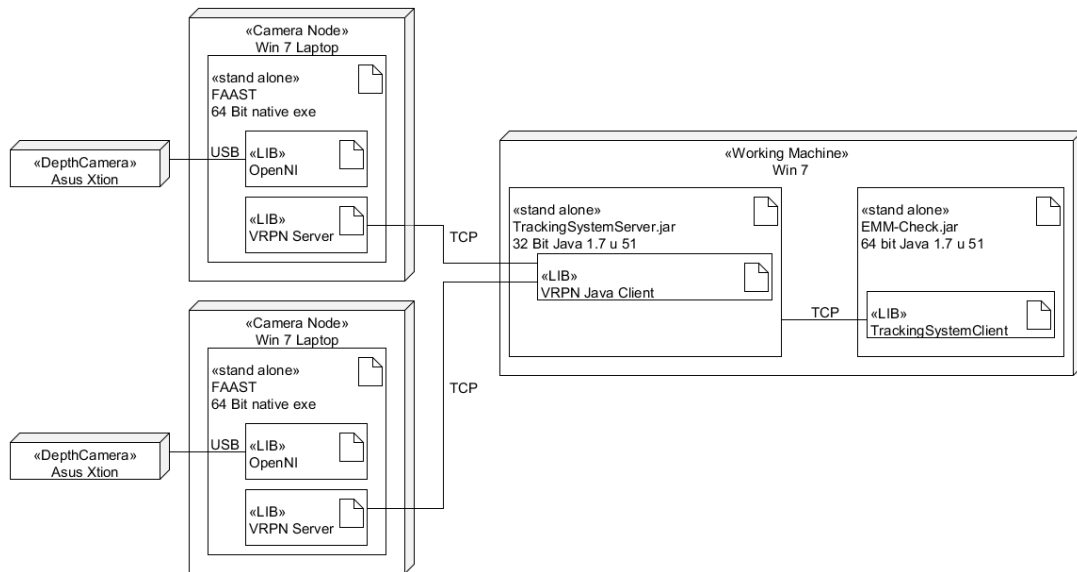


Abbildung 13 - Verteilungs-Diagramm des Systems mit zwei Tiefenkameras

In Abbildung 13 ist ein Aufbau mit 2 Tiefenkameras und die Verteilung der Software auf die einzelnen Computer dargestellt. Der TrackingSystemServer und der TrackingSystemClient auf der „Working Machine“ sind die Komponenten des Gesamtsystems, die entwickelt werden müssen. Dabei verwendet die TrackingSystemServer-Implementation den VRPN Java Client, der die Kommunikation mit dem VRPN-Server übernimmt und so die Full Body Tracking Information dem TrackingSystemServer zur weiteren Verarbeitung bereitstellt. Diese werden in einigen Arbeitsschritten transformiert und an einen oder mehrere TrackingSystemClients verteilt. Der TrackingSystemServer ist ein eigenständiges Java-Programm. Damit die Latenz zwischen TrackingSystemClient und TrackingSystemServer so gering wie möglich gehalten wird, befinden sie sich auf demselben Computer. Der Server kann aber auch auf einem separaten Computer gestartet werden. Der TrackingSystemClient wiederum wird als eine Java-Library zum Integrieren in bereits bestehende Anwendungen entworfen, um den einfachen Einsatz eines Full Body Tracking in Java bereit zu stellen.

4.4 Design

In diesem Abschnitt werden die Komponenten des TrackingSystemServer und TrackingSystemClient erläutert, ohne dabei konkret auf die verwendeten Algorithmen einzugehen.

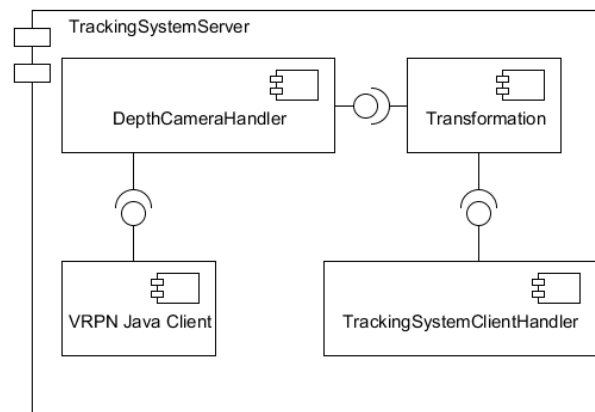


Abbildung 14 - TrackingSystemServer Komponenten Ansicht

Der TrackingSystemServer besteht aus vier Komponenten (siehe Abbildung 14). Der Depth-CameraHandler verwaltet die mit dem Server verbundenen Tiefenkameras, in dem er für jede Tiefenkamera einen VRPN Java Client startet. Außerdem stellt er die Kameradaten der Transformation Komponente zur Verfügung. Die Hauptaufgabe des DepthCameraHandler ist es, immer die Kamera-Informationen bereitzustellen, die die besten Ergebnisse hinsichtlich des Full Body Tracking liefert. Die VRPN Java Client Komponente ist eine Middleware und übernimmt, ohne das Einfluss darauf genommen werden kann, die Kommunikation mit dem VRPN-Server. Sie liefert Positions- und Rotationsdaten der einzelnen Gelenkpunkte des menschlichen Körpers in Form von absoluten dreidimensionalen Koordinaten und Quaternionen. Die Transformation-Komponente stellt die Verbindung zwischen Tiefenkameradaten und dem TrackingSystemClientHandler her. Die Hauptaufgabe dieser Komponente ist es, zu Beginn die Kameras zu kalibrieren, indem sie auf ein gemeinsames Koordinatensystem synchronisiert werden und im weiteren Verlauf die notwendigen Transformationen der Full Body Tracking-Daten zu berechnen und dem TrackingSystemClientHandler zu übergeben. Die TrackingSystemClientHandler-Komponente verwaltet die Sessions mit dem TrackingSystemClient und übernimmt die Kommunikation zwischen Client und Server.

Der TrackingSystemClient ist in keine weiteren Komponenten zerlegt, denn alle notwendigen Berechnungen erfolgen auf dem Server. Der Client dient lediglich dazu, die Positions- und Rotationsdaten des Full Body Tracking als Middleware-Komponente an das Programm auszuliefern, welche den TrackingSystemClient als solche verwenden. Der Client baut eine Verbindung zum Server auf und meldet sich dort an, der Server schickt dann in regelmäßigen Abständen die Daten zum Client.

4.5 Implementation

Das System wird in Java7 implementiert und verwendet FAAST in Version 1.2, um die Daten der Tiefenkameras abzurufen und im Netzwerk über den VRPN-Server bereitzustellen. Der TrackingSystemServer baut zu jeder FAAST-Instanz eine Verbindung zu dessen VRPN-Server auf, um die Full Body Tracking-Daten dort abzurufen. Hierzu müssen die im Netzwerk verteilten FAAST-Instanzen in einem Konfigurationsfile vom TrackingSystemServer mit ihrer IP-Adresse eingetragen sein, damit eine Verbindung aufgebaut werden kann.

4.5.1 Client

Der TrackingSystemClient wird in Java als Middleware für andere Programme entwickelt und als externes JAR zum Einbinden bereitgestellt. Abbildung 15 zeigt die verwendeten Klassen und Interfaces der TrackingSystemClient-Komponente.

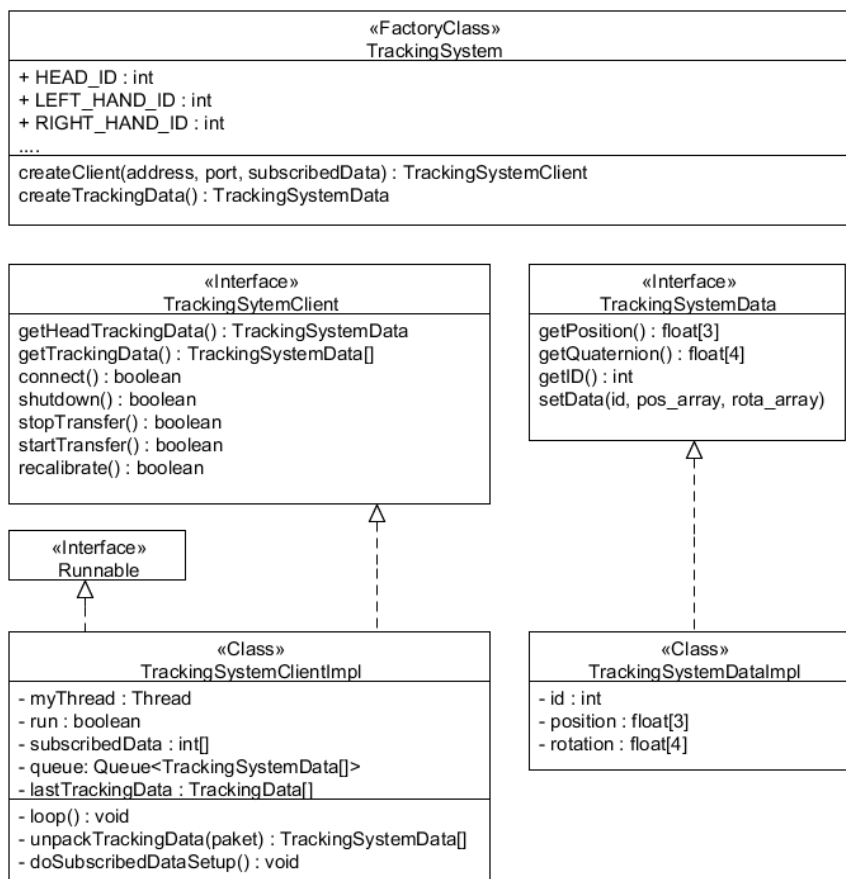


Abbildung 15 - Klassen und Interfaces des TrackingSystemClient

Die TrackingSystem FactoryClass übernimmt das Erzeugen der Client-Objekte und bietet Informationen zu den abzurufenden TrackingData-Elementen an. Außerdem übernimmt es das Pooling von TrackingSystemData-Objekten, um den Speicherverbrauch des Clients konstant zu halten. Zu Beginn wird einmalig eine Anzahl an TrackingSystemData-Objekten erstellt, die im weiteren Programmverlauf immer wieder von TrackingSystemClient an das aufrufende Programm ausgeliefert werden. Das Interface des TrackingSystemClient ist einfach gehalten: Zum einen können die TrackingSystemData des vom Server verfolgten Kopfes abgerufen werden und zum andern auch alle anderen verfügbaren Gelenkpunkte die zu Beginn ausgewählt wurden. Darüber hinaus bietet das Interface einige Steuerungsfunktionen an, mit denen das aufrufende Programm Einfluss auf den Client und den Server nehmen kann. Das TrackingSystemData-Interface stellt einen Datensatz dar, der über im TrackingSystem bekannte ID's identifiziert werden kann. Die Position wird durch einen float array mit 3 Werten repräsentiert, dabei entsprechen die Indices 0 : X, 1 : Y und 2 : Z . Die Rotation wird durch ein float array mit 4 Werten repräsentiert, hier entsprechen die Indices 0 : q0, 1 : q1, 2 : q2 und 3 : q3 der q Komponenten der Quaternionen. Die setData-Methode stellt den Konstruktor für die TrackingSystemData-Objekte dar. Auf Grund des Poolings der TrackingSystemData-Objekte, müssen die neuen Werte über diesen Setter gesetzt werden. Der TrackingSystemClient wird in einem separaten Thread ausgeführt, der ausschließlich zur Kommunikation mit dem TrackingSystemServer dient. Der Thread nimmt die eingehende Pakete in der loop-Methode entgegen und wandelt sie mit der unpackTrackingData-Methode in ein TrackingSystemData-Array um, damit dieser anschließend in der Queue, für die Auslieferung an das aufrufende Programm bereitgestellt werden kann. Für den Fall, dass keine Elemente in der Queue vorhanden sind, wird das zuletzt ausgelieferte TrackingSystemData-Array erneut ausgeliefert. Die Aufrufe auf den TrackingSystemClient sind nicht blockierend und nehmen damit keinen Einfluss auf das aufrufende Programm.

4.5.2 Server

Der TrackingSystemServer wird als eigenständiges Programm in Java entwickelt und kann als JAR vorerst nur auf Windows-Computern ausgeführt werden. Die Abhängigkeit vom VRPN Java Client, der zur Kommunikation mit dem FFAST Tool benötigt wird, führt zu dieser Einschränkung. Auf Grund dessen, dass der VRPN Java Client eine für Windows kompilierte native Library benötigt, die im TrackingSystemServer eingebunden werden muss. Im Folgenden werden die Innenansichten der Komponenten des TrackingSystemServers vorgestellt.

4.5.2.1. DepthCameraHandler

Abbildung 16 zeigt die Innenansicht der DepthCameraHandler-Komponente. Die Komponente verwaltet die verbundenen Tiefenkameras, stellt die Daten aller Tiefenkameras bereit und entscheidet welche Tiefenkamera die besten Daten liefert. Anschließend werden diese Daten an die Transformation-Komponente übergeben und danach dem TrackingSystemClient übersendet.

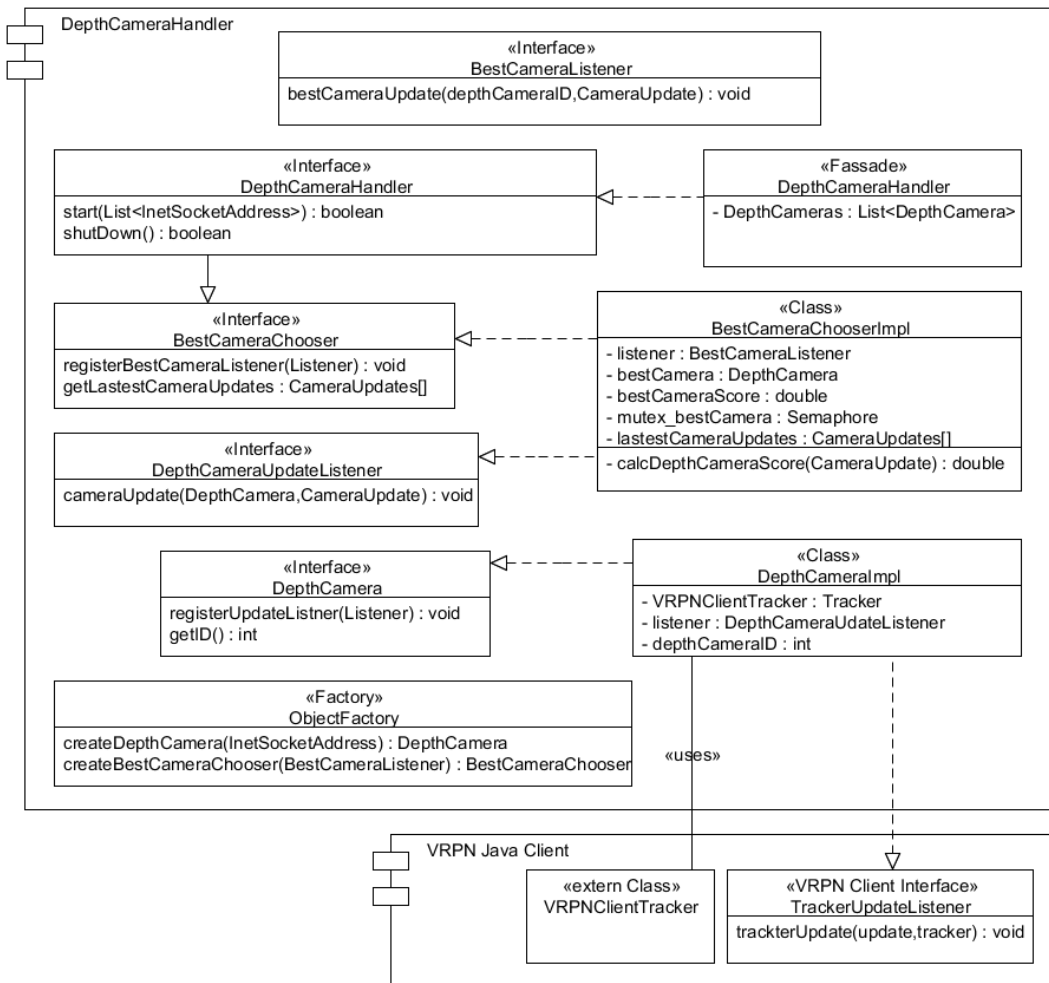


Abbildung 16 - Innenansicht der DepthCameraHandler Komponente

Die DepthCameraHandler-Komponente wird nicht als eigenständiger Thread realisiert, da die DepthCameraImpl, welche den VRPN Java Client kapselt, durch die externe VRPNClientTracker-Klasse bereits als eigenständiger Thread realisiert wird. Hierbei werden über den TrackerUpdateListener die eingehenden Updates vom VRPNClientTracker und

somit der Tiefenkamera an die DepthCameraImpl übergeben. Diese besitzt einen Listener für das DepthCameraUpdateListener Interface, welches die BestCameraChooserImpl-Klasse implementiert. Alle DepthCameraImpl-Instanzen besitzen den BestCameraChooserImpl als diesen Listener und übergeben ihre CameraUpdates an ihn. Der BestCameraChooserImpl ist für die Bestimmung der aktuell besten Tiefenkamera zuständig. Dieser Vorgang erfolgt anhand eines Tiefenkamerawerts, welcher durch die calcDepthCameraScore-Methode berechnet wird. Dieser Wert errechnet sich im Wesentlichen aus der Distanz der Kamera zum verfolgten Objekt. Sobald sich das verfolgte Objekt nah an einer der Tiefenkameras befindet, von denen es verfolgt wird, wird diejenige Tiefenkamera ausgewählt, bei dem der Tiefenkamerawert am niedrigsten ist. Demzufolge repräsentiert ein niedriger Wert eine geringe Distanz zur Tiefenkamera und ein hoher Wert dementsprechend eine große Distanz zur Tiefenkamera. Das ist eine sehr einfache Heuristik, die aber durch die zugrunde liegende Architektur ohne weiteres verbessert werden kann. Nachdem die BestCameraChooserImpl die beste Tiefenkamera bestimmt hat, wird sie an den BestCameraUpdateListener übergeben, welcher von einer Klasse in der Transformation-Komponente des TrackingSystemServer implementiert wird und dort anschließend zur Weiterverarbeitung genutzt werden kann.

4.5.2.2. TrackingSystemClientHandler

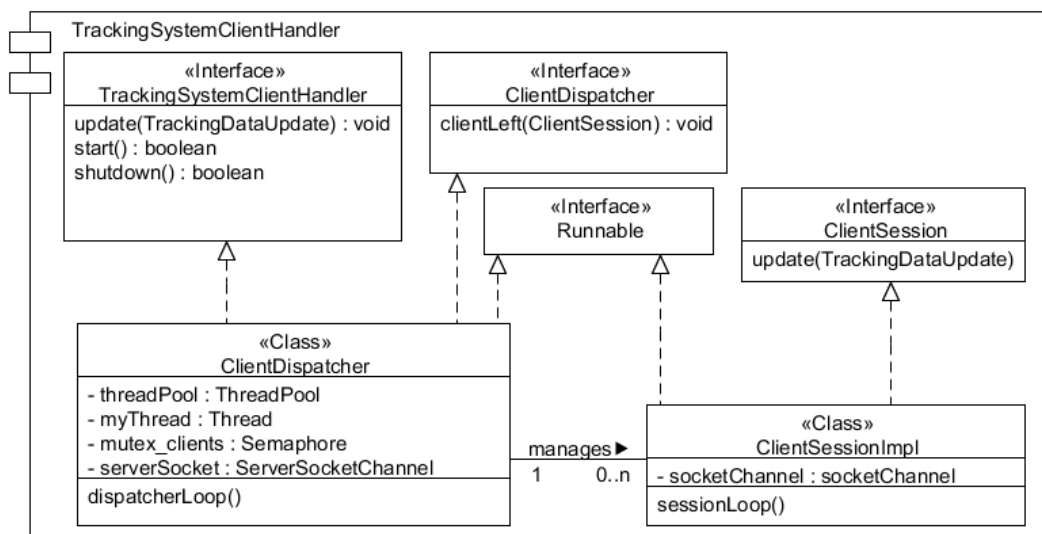


Abbildung 17 - Innenansicht der TrackingSystemClientHandler Komponente

Die Innenansicht der TrackingSystemClientHandler-Komponente wird in Abbildung 17 gezeigt. Sie dient zur Verwaltung und Kommunikation der verbundenen TrackingSystemClients. Der TrackingSystemClientHandler wird durch die ClientDispatcher-Klasse realisiert und ist ein eigenständiger Thread, der in der dispatcherLoop-Methode darauf wartet, eine eingehende Verbindung zu bearbeiten und daraus eine

ClientSessionImpl auf dem ThreadPool zu starten. Dadurch wird es möglich, dass mehrere TrackingSystemClients parallel mit dem TrackingSystemServer kommunizieren können. Die ClientSessionImpl nimmt in der sessionLoop-Methode die Anfragen aus dem in Abschnitt 4.5.3 beschriebenen Protokoll entgegen und bearbeitet sie. Das TrackingSystemClientHandler Interface verfügt über eine update-Methode, damit der TrackingSystemServer und dessen Komponenten Daten an die Clients senden können. Mit Hilfe dieser update-Methode werden die Tracking-Daten an alle verbundenen Clients verteilt, indem über die Liste der aktuellen ClientSessions iteriert wird und auf jede ClientSession die update-Methode aufgerufen wird.

4.5.2.3. Transformation

Die Transformation-Komponente (Abbildung 18) verbindet die DepthCameraHandler- und TrackingSystemClientHandler-Komponente. Sie bekommt die Positions- und Rotationsinformation der besten Tiefenkamera über das BestCameraListener-Interface von der DepthCameraHandler-Komponente bereitgestellt. Anschließend transformiert sie diese mit der transformToGlobalSpace-Methode in das globale Koordinatensystem und übergibt sie der TrackingSystemClientHandler-Komponente, welche diese transformierten Positions- und Rotationsinformation der besten Tiefenkamera an die TrackingSystemClients weiterverteilt.

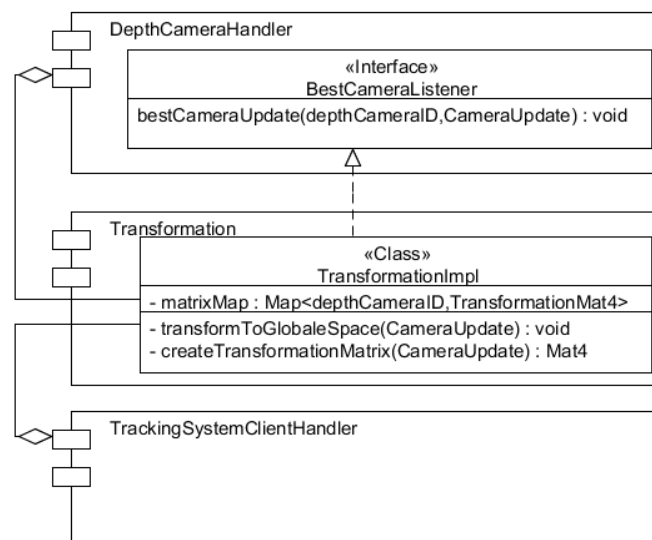


Abbildung 18 - Innenansicht Transformation Komponente

Bei der Transformation von dem lokalen Koordinatensystem der Tiefenkamera in das globale Koordinatensystem des Raumes nutzt sie die in der Kalibrierungsphase erstellten Transformationsmatrizen der jeweiligen Tiefenkamera (siehe 4.5.4 Verfahren zur

Kalibrierung der Tiefenkameras). Die Transformation-Komponente transformiert nur die Koordinaten der besten Tiefenkamera und nicht die aller Tiefenkameras. Dies wird dadurch ermöglicht, dass der DepthCameraHandler die beste Tiefenkamera vorher bestimmt und die bestCameraUpdate-Methode aus dem Listener-Interface nur dann aufgerufen wird.

4.5.3 Kommunikation zwischen Client und Server

Die Kommunikation zwischen dem TrackingSystemServer und dem TrackingSystemClient erfolgt mit TCP und ist mit Java New IO API [62] implementiert. TCP wird verwendet, um die Reihenfolge, in der die Pakete versandt werden, zu erhalten und keine Pakete verloren gehen. Aus diesem Grund wird eine geringfügig höhere Latenz gegenüber UDP zu erwarten sein, dennoch ist die zuverlässige Reihenfolge erhaltene Kommunikation das ausschlaggebende Kriterium bei der Wahl zwischen TCP und UDP. Sollte die Latenz letztendlich doch zu hoch sein, würde sich an dieser Stelle noch weiteres Optimierungspotenzial anbieten, welches der modularen Struktur des Systems zu verdanken ist. Dazu müsste die darunterliegende Kommunikationsstruktur von Client und Server in einer weiteren Implementation anhand des Interfaces neu implementiert werden. Dadurch könnte der Nutzer später entscheiden, welche Kommunikationsform für ihn die geeignetere ist.

Die Kommunikation zwischen TrackingSystemServer und TrackingSystemClient findet mit einem speziell dafür entworfenen Protokoll statt. Das Protokoll sorgt dafür, dass sich Clients am Server anmelden können. Zudem erlaubt es dem Client, dem Server mitzuteilen, welche Full Body Tracking-Daten dieser vom Server zugesandt bekommen soll. Der Server stellt insgesamt 19 verschiedene Gelenkpunkte des menschlichen Körper bereit, von denen die Position und Rotation im Raum bekannt sind. Gibt der Client die Anweisung zum Übertragen der Daten vom Server an den Client, so sendet der Server daraufhin alle 33ms die Daten an den Client. Hierbei hat der Client wiederum die Möglichkeit, die Übertragung anzuhalten oder ganz zu beenden. Der Client ist ebenfalls dazu in der Lage eine neue Kalibrierung der Tiefenkameras zu veranlassen, um den Ursprung des globalen Koordinatensystems neu auszurichten.

4.5.4 Verfahren zur Kalibrierung der Tiefenkameras

Nachdem alle Tiefenkameras im Raum positioniert wurden, muss eine einmalige Kalibrierung des Systems erfolgen. Danach kann das System beliebig oft verwendet werden. Sobald aber die Position oder Orientierung einer Tiefenkamera verändert wurde, muss eine Neukalibrierung durchgeführt werden. Die Kalibrierung bzw. Synchronisation der Kameras erfolgt mit Hilfe eines Referenzobjektes, welches sich in dem überlappenden Bereich aller Tiefenkamerakegel befinden muss und dauert keine 10 Sekunden. Das Referenzobjekt ist

ein menschlicher Körper, zum Beispiel die Person die verfolgt werden soll. Diese Person positioniert sich zu Beginn einmal in der Mitte des Raumes und bleibt dort, ohne sich zu bewegen, ruhig stehen. Der TrackingSystemServer erhält jetzt von allen verbundenen Tiefenkameras die Positions- und Rotationsdaten dieser Person in den lokalen Koordinatensystemen der Tiefenkameras. Die Person repräsentiert damit die Position und die Achsenorientierung des Ursprungs vom globalen Koordinatensystem (B) in den lokalen Koordinatensystemen (A) der einzelnen Tiefenkameras.

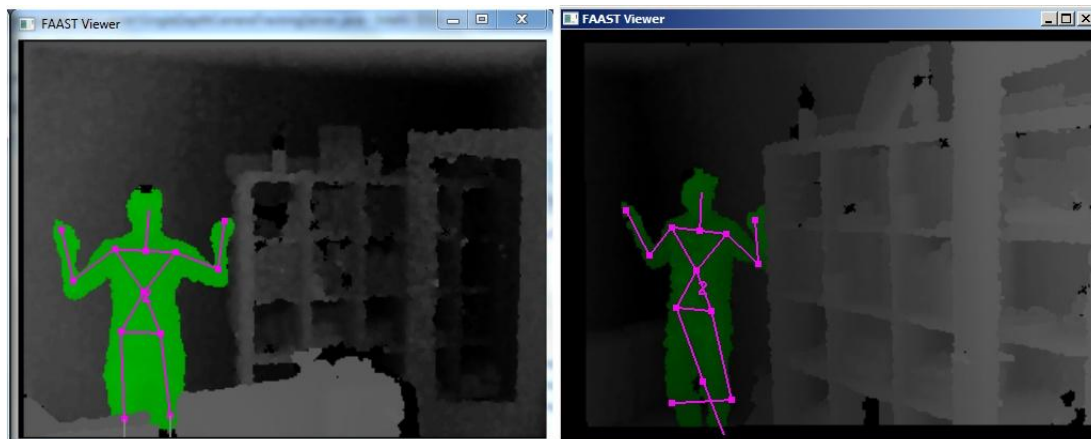


Abbildung 19 - Kalibrierungsphase, eine Aufnahme von zwei verschiedenen Tiefenkameras zur selben Zeit aus der Sicht ihrer lokalen Koordinatensysteme (A).

Mit Hilfe einer Koordinatensystem-Transformation, die den Ursprung eines Koordinatensystems und dessen Achsenorientierung benötigt, kann eine Transformationsmatrix ($B \rightarrow A$) gebildet werden. Mit dieser Matrix kann vom globalen Koordinatensystem (B) in das jeweilige lokale Koordinatensystem (A) der Tiefenkamera transformiert werden. Wird von dieser Matrix die inverse Matrix $((B \rightarrow A)^{-1})$ gebildet, kann vom lokalen Koordinatensystem (A) der Tiefenkamera in das globale Koordinatensystem (B) transformiert werden.

Der TrackingSystemServer erstellt in seiner Kalibrierungsroutine für jede Tiefenkamera eine inverse Transformationsmatrix $((B \rightarrow A)^{-1})$, um eine Transformation vom lokalen Koordinatensystem (A) der Tiefenkamera in das globale Koordinatensystem (B) des Raumes zu ermöglichen. Das Erstellen der Transformationsmatrix erfolgt, indem aus den Rotationsdaten des Referenzobjektes eine Rotationsmatrix erzeugt wird. Hier wird die Position des Referenzobjektes als Translationskomponente gesetzt und die Matrix invertiert. Nachdem für alle Tiefenkameras die Transformationsmatrizen $((B \rightarrow A)^{-1}_{Cx})$ erstellt wurden, ist die Kalibrierungsphase abgeschlossen. Im Weiteren werden alle Tiefenkamera Koordinaten (P_{Acx}) immer mit den dazu gehörigen Transformationsmatrizen in das globale Koordinatensystem überführt ($P_{Bcx} = P_{Acx} * (B \rightarrow A)^{-1}_{Cx}$).

Die gesamte Kalibrierung wird dadurch vereinfacht, dass OpenNI zum einen die Abbildungsfehler durch die Linsenverzerrung der Kamera korrigiert und zum andern die Möglichkeit zur Bestimmung der Körperhaltung (engl. human pose estimation, siehe [63]) in Tiefenbildern bietet. OpenNI erkennt, ob eine Person seitlich, mit der Brust, mit dem Rücken oder in einem anderen Winkel zur Tiefenkamera steht. Hierbei bestimmt es die Rotation des Körpers, welche anschließend als Grundlage zur Erstellung der Transformationsmatrizen verwendet wird. Die Idee hinter dem Kalibrierungsverfahren des TrackingSystemServers ist die Annahme, dass die Person, die das Referenzobjekt der Kalibrierung repräsentiert, als Koordinatensystemursprung mit seiner Position und Rotation im Raum angesehen wird.

4.6 TrackingSystemServer Ablauf im Überblick

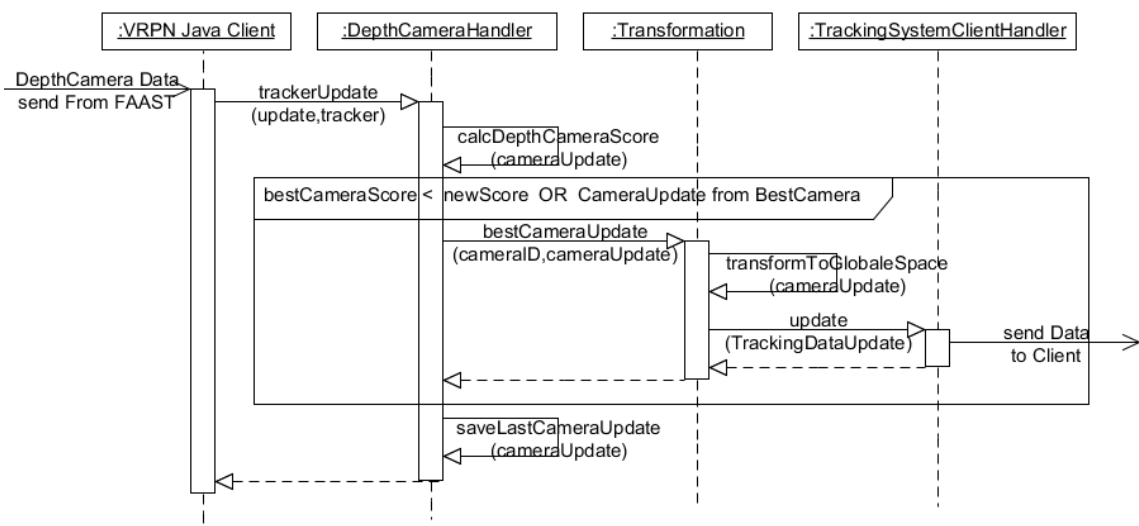


Abbildung 20 - Interaktion zwischen den Komponenten des TrackingSystemServers

Abbildung 20 zeigt die Interaktion zwischen den Komponenten des TrackingSystemServers, nachdem eine Kalibrierung des Systems erfolgt ist. Der VRPN Java Client erhält von FFAST die Positions- und Rotationsdaten der Tiefenkamera. Ein Systemaufbau mit vier Tiefenkameras verfügt über vier VRPN Java Client-Instanzen, welche von einer DepthCameraHandler-Instanz verwaltet werden. Die VRPN Java Client-Instanzen rufen jeweils alle 33ms die trackerUpdate-Methode auf die DepthCameraHandler-Instanz auf. Die BestCameraChooserImpl entscheidet innerhalb der DepthCameraHandler-Komponente durch die calcDepthCameraScore-Methode, ob das cameraUpdate besser als das aktuell beste Update ist. Daraufhin wird der bestCameraUpdateListener aufgerufen, der in der Transformation-Komponente implementiert ist. Dort wird das CameraUpdate mit der für die jeweilige Tiefenkamera geltenden Transformationsmatrix in das globale Koordinatensystem des Raumes transformiert. Anschließend werden die Daten an den

TrackingSystemClientHandler zum Versand übergeben. Dieser erstellt ein Paket des TrackingSystem Protokolls und versendet es an alle verbunden TrackingSystemClients. Abschließend wird das CameraUpdate noch in der DepthCameraHandler-Komponente als letztes Update von der Tiefenkamera gespeichert.

4.7 Auswertung hinsichtlich der Untersuchungskriterien

Das hier vorgestellte Tracking-System unterstützt 6 Freiheitsgrade. Es ist damit also in der Lage, Position und Rotation einer verfolgten Person zu bestimmen. Dies gilt allerdings nur für den Oberkörper ohne den Kopf der Person. Dreht die Person beispielsweise den Kopf nach rechts, dann kann das System nur die Rotation des Oberkörpers und nicht die exakte Rotation des Kopfes bestimmen. Wird also nur der Kopf einer Person verfolgt, unterstützt das System nur 3 Freiheitsgrade. Desweiteren ist zu beachten, dass das System nur eine Person zurzeit verfolgen kann. Durch mehr Entwicklungsaufwand kann diese Annahme aus dem Systementwurf jedoch aufgehoben werden. Eine Tiefenkamera alleine kann 4 Personen gleichzeitig verfolgen. Der TrackingSystemServer müsste daher so modifiziert werden, dass die Personen über alle verbundenen Tiefenkameras hinweg eindeutig identifiziert werden können. So wäre das System in der Lage, 4 Personen gleichzeitig verfolgen zu können. Die Größe des überwachten Bereiches liegt, wie im Systementwurf angegeben, bei maximal 8m x 8m und bietet damit ausreichend Platz, um ein freies Erkunden einer virtuellen Welt zu ermöglichen. Diese Größenbeschränkung hat ihren Ursprung in der vom Tiefenkamerahersteller vorgegebenen maximalen Reichweite von 3,5 m [61]. Einige meiner Versuche haben jedoch gezeigt, dass die Asus Xtion auch bei Reichweiten von bis zu 5 m noch zuverlässige Informationen liefert. Die Genauigkeit der Positionsbestimmung des Tracking-Systems liegt in einem Intervall von -10 cm bis +10 cm entlang der Z-Achse und bei -3cm bis +3cm entlang der X- und Y-Achse. Die Genauigkeit ist stark abhängig davon, wie gut die Kalibrierung des Systems erfolgt ist. Eine geringe Abweichung kann, bei der Positions- und Rotationsbestimmung des Referenzobjektes über die verschiedenen Tiefenkameras hinweg, große Fehler bei der Umwandlung vom Koordinatensystem der Tiefenkamera in das globale Koordinatensystem des Raumes mit sich bringen. Der folgende Ausschnitt aus einem Logfile soll die angegebene Genauigkeit verdeutlichen. Zwei Tiefenkameras haben hierbei eine Person über 60 Sekunden verfolgt, welche ohne sich zu bewegen im Raum stand. Dabei ist Node0 und Node1 jeweils eine Tiefenkamera. Die angegebenen Koordinaten (x,y,z,w) sind in Metern und entsprechen den Koordinaten des globalen Koordinatensystems im Raum. Es wurde alle 5 Sekunden eine Positionsbestimmung durchgeführt.

```
Node0 Camera space -> World space : (0.39423123, -0.33948094, -0.16038918, 1.0)
Node1 Camera space -> World space : (0.37855852, -0.29620633, -0.17196691, 1.0)
-----5 seconds later-----
```



```

Node0 Camera space -> World space : (0.37820038, -0.3704291, -0.18628144, 1.0)
Node1 Camera space -> World space : (0.3689654, -0.32200584, -0.21440697, 1.0)
-----5 seconds later-----
Node0 Camera space -> World space : (0.39515412, -0.3826896, -0.17125487, 1.0)
Node1 Camera space -> World space : (0.37601388, -0.32269678, -0.18772769, 1.0)
-----5 seconds later-----
Node0 Camera space -> World space : (0.39272332, -0.3796636, -0.16175914, 1.0)
Node1 Camera space -> World space : (0.37690163, -0.3242941, -0.19769454, 1.0)
-----5 seconds later-----
Node0 Camera space -> World space : (0.38131428, -0.33681422, -0.1764574, 1.0)
Node1 Camera space -> World space : (0.38355434, -0.3020517, -0.20749927, 1.0)
-----5 seconds later-----
Node0 Camera space -> World space : (0.3683861, -0.3304739, -0.17932749, 1.0)
Node1 Camera space -> World space : (0.37384558, -0.30161244, -0.19464457, 1.0)
-----5 seconds later-----
Node0 Camera space -> World space : (0.38220286, -0.36084473, -0.17979217, 1.0)
Node1 Camera space -> World space : (0.3695531, -0.32031775, -0.20801806, 1.0)
-----5 seconds later-----
Node0 Camera space -> World space : (0.36847988, -0.35638118, -0.2163806, 1.0)
Node1 Camera space -> World space : (0.3647927, -0.31802607, -0.20809114, 1.0)
-----5 seconds later-----
Node0 Camera space -> World space : (0.367909, -0.36241478, -0.2101233, 1.0)
Node1 Camera space -> World space : (0.3616109, -0.3220145, -0.21871114, 1.0)

```

Im folgenden Logfileausschnitt wurde vom Koordinatensystemursprung des Raumes ein Schritt nach rechts gemacht.

```

Node0 Camera space -> World space : (-0.0013974309, -0.002557218, 0.0062651634, 1.0)
Node1 Camera space -> World space : (-0.013339043, 3.033504E-4, 0.011736512, 1.0)
-----schritt nach rechts-----
Node0 Camera space -> World space : (0.7268992, 0.0044, -0.052053213, 1.0)
Node1 Camera space -> World space : (0.7935864, -0.005929902, -0.11342478, 1.0)

```

Im folgenden Logfileausschnitt wurde vom Koordinatensystemursprung des Raumes ein Schritt zurück gemacht.

```

Node0 Camera space -> World space : (-0.0014801025, -0.0025572926, 8.392334E-4, 1.0)
Node1 Camera space -> World space : (-0.0072431564, 0.008981854, -0.0018662214, 1.0)
-----schritt nach hinten-----
Node0 Camera space -> World space : (0.01479885, 0.06407204, 0.7502761, 1.0)
Node1 Camera space -> World space : (-0.02644062, 0.06828159, 0.7710422, 1.0)

```

Im folgenden Logfileausschnitt wurde sich vom Koordinatensystemursprung des Raumes geduckt.

Node0 Camera space -> World space : (-8.41707E-4, 0.0078006983, 0.0013039112, 1.0)

Node1 Camera space -> World space : (-2.4795532E-4, 0.0025345385, 4.5585632E-4, 1.0)

-----ducken-----

Node0 Camera space -> World space : (0.0667322, -0.492973, 0.09475732, 1.0)

Node1 Camera space -> World space : (0.04872179, -0.54737973, 0.100128055, 1.0)

Die Genauigkeit wurde ebenfalls dadurch reduziert, dass sich die Tiefenkameras untereinander mit ihren Infrarot-Projektoren stören. In Abbildung 21 im linken Bild wird das Regal und die Tür im Hintergrund von dieser und einer weiteren Tiefenkamera abgetastet. Die Konturen der Tür und des Regals sind schlechter zu erkennen als im rechten Bild (rot markierter Bereich), dieses wurde mit nur einer Tiefenkamera abgetastet. Im Vergleich zum rechten Bild, lassen sich im linken Bild viel mehr der kleinen grauen runden Stellen erkennen, welche als Störungen angesehen werden können. Der gelbe Bereich zeigt einen größeren Fehler zwischen linkem und rechtem Bild.

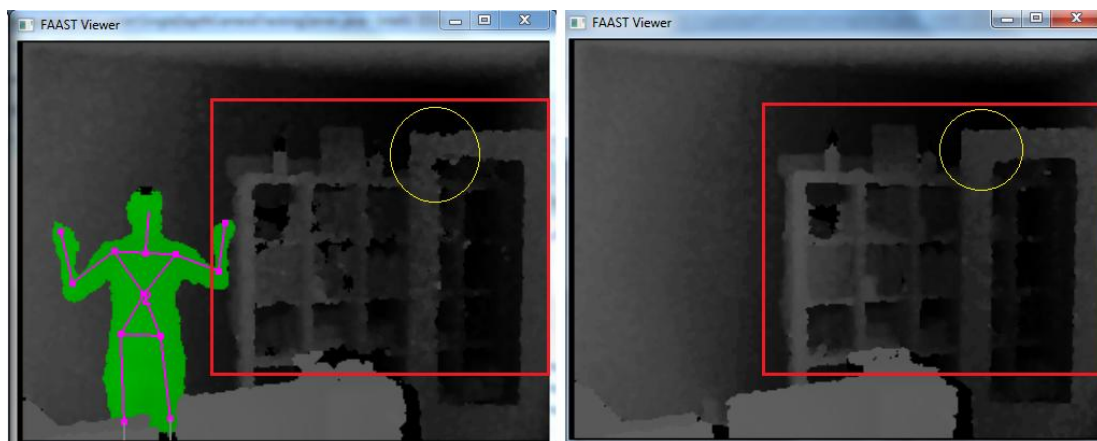


Abbildung 21 - linkes Bild enthält Störungen, durch überlappende Tiefenkamerakegel im rot markierten Bereich. Gelber Bereich zeigt einen größeren Fehler zwischen linkem und rechtem Bild. Rechtes Bild enthält keine Störungen.

Für den Kalibrierungsprozess von mehr als 3 Tiefenkameras könnte dies ein größeres Problem darstellen, da die überlappenden Kamerakegel, in denen sich die Person zur Kalibrierung positionieren muss, einen größeren Fehler erzeugen könnten. So könnten die Positions- und Rotationsdaten, die zur Bestimmung des globalen Koordinatensystems verwendet werden, verfälscht werden. Um an dieser Stelle größere Fehler bzw. Abweichungen bei der Umrechnung vom lokalen in das globale Koordinatensystem zu vermeiden und somit eine ausreichende Genauigkeit zu gewährleisten, könnte die Kalibrierungsphase verlängert werden. Es bietet sich zum einen eine Mittelung der Positions- und Rotationsdaten über einen längeren Zeitraum an, welche das Rauschen bzw.

die Fehler reduziert. Zum anderen könnten die Kameras nacheinander kalibriert werden, sodass nur eine Kamera zurzeit aktiv ist und den Bereich des Referenzobjekts abtastet. Damit würde sich die Kalibrierungsphase zwar auf ein oder zwei Minuten verlängern, jedoch wäre der Fehler und somit die Abweichung stark reduziert.

Die Wiederholrate des Systems beträgt 30Hz und wird durch die verwendete Tiefenkamera limitiert. Diese Wiederholrate ist ausreichend, um verfolgte Bewegungen einer Person gut wiederzugeben. Führt die verfolgte Person jedoch ruckartige oder blitzartige Bewegungen durch, kann es zu einer schlechteren Wiedergabe der Bewegungen kommen. Die Wiederholrate hat ebenfalls Einfluss auf die Latenz. 30Hz entsprechen einer Latenz von 33ms. Hinzu kommt die Transportlatenz des aufgebauten Netzwerks, die Verarbeitungslatenz am TrackingSystemServer, die Auslieferung an den Client und die Darstellungslatenz. Es wird angenommen, dass dies bis zu 30ms in Anspruch nimmt. Damit liegt die endgültige Latenz des Systems in einem Intervall von 48ms bis 63ms und bewegt sich damit an der empfohlenen Grenze von 50ms (siehe [49], [50]).

Wie bereits in der Motivation zur Entwicklung des Systems geschrieben wurde, bringen Tiefenkameras auch einige Einschränkungen mit sich. Diese schlagen sich bei der Empfindlichkeit gegenüber den äußeren Rahmenbedingungen nieder. Die Tiefenkameras stören sich untereinander bei der Erstellung des Tiefenbilds, da sie alle den gleichen Infrarotprojektor verbaut haben und somit alle dasselbe Projektionsmuster verwenden. Das hat zur Folge, dass ein geringes Rauschen in den Tiefenbildern entsteht, das Auswirkungen auf die Genauigkeit hat. Zum anderen kommen die Tiefenkameras schlecht mit Verdeckungen, z.B. durch Möblierung des Raumes oder durch andere Personen, zurecht. Deshalb wurde unter anderem zu Beginn die Annahme eines leeren Raumes getroffen, indem das System aufgebaut werden muss. Das in Abschnitt 4.5.4 beschriebene Kalibrierungsverfahren ist ein sehr einfaches Verfahren. Es hat allerdings den Nachteil, dass die Genauigkeit dieses Systems sich nicht im Millimeter Bereich bewegt, sondern, wie anfangs beschrieben, im Zentimeter-Bereich.

Die Usability des Systems ist ausreichend, denn der Nutzer braucht keinerlei Marken oder Anzüge am Körper zutragen, um vom System verfolgt zu werden. Desweiteren ist die Kalibrierungsphase schnell und ohne großen Aufwand erledigt. Der Aufbau des gesamten Systems ist allerdings mit allen verwendeten Tiefenkameras, den dazugehörigen Laptops und den Netzwerkkomponenten äußerst zeitaufwendig und umständlich. Ist das System jedoch einmal aufgebaut, benötigt es nur eine einmalige Kalibrierung. Auf Grund der längeren Aufbauphase und der Tatsache, dass der gesamte Aufbau als Festinstallation im Raum dienen soll, ist es als mobiles System für schnelle Standortwechsel nicht geeignet. Die Startphase des Systems erfordert in diesem Stadium der Implementation einen manuellen Start der FFAST-Anwendung und zwar auf allen verwendeten Laptops, an denen Tiefenkameras angeschlossen sind. FFAST bietet zudem keine Möglichkeit zur Auswahl der angeschlossenen Tiefenkamera (dies ist ebenfalls unvorteilhaft). Die Konsequenz daraus ist,

dass pro Tiefenkamera ein Computer benötigt wird. Diese Einschränkung lässt sich jedoch durch eine Eigenentwicklung eines Clients, der die Tiefenkameradaten bereitstellt und damit FFAST ersetzt, sofort beheben.

Besondere Merkmale des Systems sind, dass es mit einem markenlosen optischen Tracking-Verfahren arbeitet und somit nicht auf Marken oder Anzüge angewiesen ist. Mit einer Erweiterung der Implementation lassen sich Gesten zur Steuerung einer Benutzeroberfläche definieren.

4.8 Mögliche Verbesserungen

Das Trackingsystem verfügt dennoch über ein großes Erweiterungspotenzial. Der überwachte Bereich von 8m x 8m könnte beispielsweise vergrößert werden. Dazu müssten mehrere TrackingSystem-Systeme und deren Server, die einen Bereich mit mehreren Kameras abdecken, in einem übergeordneten zentralen Server gesammelt werden. Nun kann unter Zusammenschluss der aneinander liegenden TrackingSystem-Systeme ein noch größerer Gesamtbereich gebildet werden. Dieser Zusammenschluss von mehreren TrackingSystem-Systemen könnte allerdings zu noch größeren Latenzen führen und damit das VR-Erlebnis sowie die Präsenz reduzieren. Sollte es aber gelingen, dann wäre es möglich noch größere nutzbare überwachte Bereiche zu schaffen. Denkbar wäre z.B. bei dem Zusammenschluss von 2 TrackingSystem-Systemen, mit 8m x 8m, ein nutzbarer Gesamtbereich von 16m x 8m.

Innerhalb von FFAST Gesten zu definieren, wäre eine weitere mögliche Erweiterung. Sie ließe sich dazu nutzen, mit der Umgebung der virtuellen Welt oder sogar der Anwendung selbst zu interagieren, Manipulationen vorzunehmen oder die Steuerung durchzuführen. Der VRPN-Server von FFAST bietet eine einfache Möglichkeit, diese Gesten abzurufen bzw. im TrackingSystemServer auslösen zu lassen. Der TrackingSystemClient würde informiert werden, damit die Anwendung dann auf dieses Ereignis entsprechend reagieren kann.

Um den Entwicklungsaufwand gering zu halten und schnell einen einfachen Prototypen entwickeln zu können, wurde bei der Implementation des TrackingSystem zur Unterstützung FFAST verwendet. FFAST könnte nun durch eine Eigenimplementation ersetzt werden, damit das TrackingSystem als ein geschlossenes Gesamtsystem vertrieben werden kann. Diese Änderung ist ohne weiteres möglich, da OpenNI auch in Java zur Verfügung steht. Hierdurch könnte die Start- und Kalibrierungsphase des Systems für den Nutzer angenehmer gestaltet werden.

Eine weitere mögliche Erweiterung könnte sein, die darunterliegende Kommunikationsstruktur des TrackingSystemServer von TCP auf UDP umzustellen. Diese Änderung führt dazu, dass geringere Latenzen durch die Kommunikation zwischen Client und Server zu erwarten sind. Allerdings ist die Reihenfolgeerhaltende Kommunikation von TCP somit nicht

mehr gegeben und müsste durch ein eigenes Verfahren, zum Beispiel Zeitstempel an den UDP Paketen, implementiert werden. Wie stark sich jedoch die Änderung von TCP auf UDP auf die Latenz auswirkt muss untersucht werden.

Ein Problem stellt die Wiederholrate von 30Hz dar, welche die Tiefenkamera alle 33ms eine Tiefenmessung durchführen lässt. Bei sehr schnellen Bewegungen reichen 30Hz jedoch nicht mehr aus und so kommt es zum Verlust von Messpunkten, die dazu nötig wären, schnelle Bewegungen in der virtuellen Welt auf dem Bildschirm flüssig darzustellen. Ein Ansatz wäre der Versuch, die Bewegungen vorherzusagen, um dadurch eine Interpolation von 30Hz auf 60Hz zu ermöglichen. Dies kann durch Bildung einer Differenz zwischen der aktuellen und der letzten Position/Rotation geschehen. So lässt sich der nächste Schritt vorhersagen und durchführen. Die Asus Xtion bietet darüber hinaus die Möglichkeit, die Auflösung von 640 x 480 auf 320 x 240 zu reduzieren. Die Reduzierung der Auflösung würde zu einer erhöhten Wiederholrate von 60Hz statt 30Hz führen und zu einer niedrigeren Genauigkeit. Es könnten so schnelle Bewegungsabläufe mit weniger Latenz auf den Bildschirm dargestellt werden, welche dann flüssiger wahrgenommen werden. Dies wäre für den Nutzer deutlich angenehmer, denn wenn die Latenz zwischen den Körperlich wahrgenommenen Bewegungen zu groß ist, kann das Gefühl der Präsenz beeinflusst werden und sogar Cybersickness hervorrufen. An dieser Stelle müsste untersucht werden, wie stark der Einfluss durch die Reduzierung der Auflösung auf die Genauigkeit ist. So könnten bessere Abwägungen zwischen Genauigkeit und Latenz erfolgen.

Festzuhalten ist, dass es einige Erweiterungen gibt, welche das Trackingsystem verbessern können und den Einsatz des Systems angenehmer machen. Diese Erweiterungen können mit dem notwendigen Wissen und Zeitaufwand leicht umgesetzt werden.

4.9 Auswahl eines geeigneten Tracking-Systems

Durch Entwicklung des Prototypen und Tests der Anwendbarkeit hat sich herausgestellt, dass es mit gewissen Einschränkungen möglich ist, ein einfaches Tracking-System für die Positionsbestimmung in geschlossenen Räumen mit Hilfe von mehreren kostengünstigen Tiefenkameras zu realisieren. Das System benötigt keinerlei Marken oder Anzüge zur Verfolgung von Personen. Zudem arbeitet es rein mit optischen Verfahren zur Positionsbestimmung und ist somit sehr angenehm in der Verwendung für den Nutzer. Darüber hinaus besteht noch die Möglichkeit Gesten zur Steuerung zu verwenden. Nun erfolgt die Auswahl eines geeigneten Systems für die Nutzfahrzeug Inspektion innerhalb von EMM-Check.

Tabelle 1 zeigt die in der Arbeit vorgestellten Tracking-Systeme und die Eigenimplementierung im Vergleich. Alle 3 Systeme unterstützen 6 Freiheitsgrade. Hierbei kann sich das TrackingSystem hervorheben, denn es verfolgt den gesamten Körper und

nicht nur den Kopf, wie es bei den in Abschnitt 3 vorgestellten Systemen der Fall ist. Das TrackingSystem bietet daher beispielsweise auch die Möglichkeit die Arme mit den Händen in der virtuellen Welt abzubilden. Somit können in der virtuellen Welt Interaktionen wie zum Beispiel das Betätigen von Hebeln oder Öffnen von Türen durchgeführt werden. Die anderen beiden Systeme bieten dort eine Controllerlösung, mit der in der virtuellen Welt wie mit einer Maus interagiert werden kann.

Tabelle 1 - Systeme und ihre Untersuchungskriterien im vergleichenden Überblick

	ImmerSight	WorldViz	TrackingSystem
Anzahl Freiheitsgrade	6, nur Kopf	6, nur Kopf	6, gesamter Körper
Anzahl Gleichzeitig verfolgter Körper	1	1	1
Größe des überwachten Bereiches	1.5m x 1.5m bis 2m x 2m	15m x 15m bis 50m x 50m	5m x 5m bis 8m x 8m
Genauigkeit	gut	sehr gut	ausreichend
Wiederholrate	60 Hz	180 Hz	30 Hz
Latenz	maximal 30ms	maximal 20ms	maximal 63ms
Empfindlichkeit gegenüber äußere Rahmenbedingungen	anfälliger bei schlechter Beleuchtung	unempfindlich	anfällig gegen Verdeckung, Tiefenkameras können sich gegenseitig stören
Kalibrierung	einfach	einfach	einfach
Usability	befriedigend	gut	gut
Kosten	niedrig	sehr hoch	niedrig
Besondere Merkmale	extra Controller für Camera in Hand Technik, sehr mobil, einfach aufzubauen	akku der aktiven LED Marker hält 6 Stunden, sehr hohe Genauigkeit	kann Gesten zur Verfügung stellen, rein optisches Verfahren

Alle Systeme können zurzeit nur eine Person verfolgen. Dies liegt bei dem ImmerSight System zum einen daran, dass es nur wenig Platz bietet und zum anderen dass nur die Person verfolgt werden kann, die den Sensoring auf dem Kopf hat (siehe Abschnitt 3.3.1). WorldViz unterstützt mehrere Personen gleichzeitig, sofern mehrere MotionTracker Brillenaufsätze verwendet werden. Das TrackingSystem kann durch mehr Entwicklungsaufwand auch 4 Personen unterstützen. Diese Grenze ist durch Einschränkungen von OpenNI und FFAST endgültig. Hinsichtlich der Größe des überwachten Bereiches ist das WorldViz System die beste Wahl, denn es kann große Bereiche mit Hilfe einer großen Anzahl an Kameras abdecken. Allerdings bringt das hohe Kosten mit sich. Hier bietet das TrackingSystem eine Alternative, da es ebenfalls einen ausreichenden Bereich abdecken

kann und um ein Vielfaches günstiger ist. Hinsichtlich der Genauigkeit bietet das WorldViz System die besten Ergebnisse. Das Trackingsystem ist jedoch auf Grund von Störungen, dem einfachen Kalibrierungsverfahren und der nur ausreichende Genauigkeit bedingt konkurrenzfähig. Die Wiederholrate aller Systeme ist gut bis sehr gut. Das TrackingSystem bietet eine nur ausreichende Wiederholrate mit 30Hz, das, wie bereits angesprochen, zu nicht flüssigen Bewegungsabläufen bei der Wiedergabe führen kann. Die Latenz ist ebenfalls, bis auf die des TrackingSystems, gut bis sehr gut.

Die Kalibrierung aller Systeme ist einfach und ist bei allen Systemen nur einmal nach dem Aufbau durchzuführen. Im Hinblick auf die Usability ist das ImmerSight Systems nur befriedigend, da zur Erkundung größerer virtueller Welten ein zusätzlicher Controller notwendig ist. Der Controller dient dazu, sich alternativ zum normalen gehen, mit der „Camera-in-Hand“ Technik durch die virtuelle Welt zu bewegen. Die anderen Systeme bieten dort einen größeren überwachten Bereich, sodass diese Systeme das natürliche Gehen in der realen Welt besser unterstützen. Im Vergleich zum professionellen System WorldViz mit hohen Kosten, stellt das ImmerSight und TrackingSystem mit jeweils niedrigem Preis eine kostengünstige Alternativen dar.

Schlussendlich fällt die Wahl auf das ImmerSight-System, welches ein mobiles System ist, wenig Platz benötigt und dennoch die Möglichkeit bietet mit Hilfe der „Camera-in-Hand“ Technik größere virtuelle Welten zu erkunden. Es müssen jedoch Abstriche bei dem Präsenzgefühl sowie bei der Bewegungswahrnehmung in Kauf genommen werden, doch dies wird durch den niedrigen Preis und geringe Latenz wieder wettgemacht. EMM-Check ist eine produktentwicklungsbegleitende Software, welche primär den Konstrukteuren dabei hilft, die Sichtfeldeinschränkungen am entworfenen CAD-Modell aufzuzeigen. Die Konstrukteure arbeiten dabei überwiegend in Büros. Hier ist ein schnell aufzubauendes und wenig Platz benötigendes System genau am richtigen Einsatzort. Es können die Sichtverhältnisse und das entworfene CAD-Modell in einer virtuellen Welt betrachtet werden. Letztendlich kann dies zwar nur teilweise maßstabsgetreu erfolgen, dennoch stellt es eine innovative Lösung für den Konstrukteur dar.

5 Schluss

5.1 Fazit

Die vorliegende Arbeit hat gezeigt, welche Eigenschaften die perfekte Virtual Reality auszeichnen und das sie heutzutage nur in Ansätzen realisiert werden kann. Dies ist dennoch sehr hilfreich, um Informationen für den Menschen zu visualisieren, damit diese von ihm besser aufgenommen und bewertet werden können. Auch die Wahrnehmungsaspekte zur Lösung bestimmter Probleme von HMD's, wie Cybersickness oder die korrekte Raumwahrnehmung, wurden analysiert und erläutert. Mit den in dieser Arbeit vorgestellten Lösungsstrategien kann den vorgestellten Problemen entgegengewirkt werden, um die VR-Erfahrung so angenehm und intensiv wie möglich zu gestalten. Ebenfalls wurden die allgemeinen Verfahren zur Positionsbestimmung in geschlossenen Räumen vorgestellt und miteinander verglichen. Wünschenswert für den Nutzer wäre dabei ein markenloses Outside-In Verfahren, wodurch die Einschränkungen für den Nutzer am geringsten wären. Es hat sich in der Praxis allerdings gezeigt, dass die Präzision bei markenbasierten Systemen höher ist, da markenlose Systeme gegenüber markenbasierten anfälliger gegen Störungen sind. Die darauf folgende Analyse der bereits vorhandenen Systeme zur Positionsbestimmung in geschlossenen Räumen hat mögliche Lösungen für das Problem der Positionsbestimmung in geschlossenen Räumen vorgestellt, anhand von einigen Untersuchungskriterien wurden die Systeme durch eine Internetrecherche analysiert und bewertet. Die Analyse hat ergeben, dass viele professionelle Systeme existieren, die zum Großteil mit optischen Verfahren arbeiten und hohe Anschaffungskosten haben. Anschließend wurde ein Prototyp entworfen, der mit Hilfe eines markenlosen Outside-In Verfahren und mehreren Tiefenkameras in der Lage ist, eine Positionsbestimmung des gesamten Körpers im geschlossenen Raum durchzuführen und diese Daten einer anderen Anwendung zur weiteren Nutzung zur Verfügung stellen kann. Dieser Prototyp wurde daraufhin ebenfalls hinsichtlich der Untersuchungskriterien der vorangegangenen Analyse untersucht und bewertet. Es hat sich herausgestellt, dass das System Potenzial zur Positionsbestimmung in geschlossenen Räumen hat und mit zusätzlichem Entwicklungsaufwand eine durchaus gute Problemlösung sein kann, vorausgesetzt die vorgeschlagenen Verbesserungen werden implementiert. Daraufhin erfolgte die Auswahl eines geeigneten Systems für den Einsatz in der vorgestellten Software EMM-Check der REKNOW GmbH und Co. KG.

5.2 Ausblick

Durch den Erwerb von OculusVR Inc. durch Facebook Inc. (siehe [64]) und die für das Jahr 2015 angekündigten HMD's (siehe [65] [66]) ist zu erwarten, dass der Bereich Virtual Reality einen großen Schub erleben wird. Nicht nur die Spieleindustrie wird davon profitieren, sondern auch viele weitere Branchen, in denen der Einsatz von VR zur Verbesserung der Anwendungen beitragen wird. Dadurch, dass diese HMD's dann gegebenenfalls zu günstigeren Preisen verfügbar sein werden, wird in naher Zukunft auch der Bedarf an Systemen zur Positionsbestimmung in geschlossenen Räumen zunehmen. Ein HMD kann alleine zwar in Ansätzen ein gutes VR-Erlebnis bieten, um jedoch das Präsenzgefühl noch zu steigern und dem Nutzer wirklich das Gefühl geben zu können, dabei zu sein, muss er sich auch in der realen Welt bewegen und Interaktionen ausführen können, welche die virtuelle Welt beeinflussen. An diesem Punkt können dann Prototypen, wie der in dieser Arbeit vorgestellte, mögliche Lösungsansätze für den Nutzer darstellen.

Literaturverzeichnis

- [1] R. Dörner, B. Jung, P. Grimm und W. Broll, *Virtual und Augmented Reality (VR/AR) - Grundlagen und Methoden der Virtuellen und Augmentierten Realität*, Heidelberg: Springer Vieweg, 2013.
- [2] REKNOW GmbH & Co. KG, „EMM-Check 3 Produktblatt,“ 2013. [Online]. Available: <http://reknow.de/includes/download.php?id=31>. [Zugriff am 04 05 2014].
- [3] HUMAN SOLUTIONS GmbH, „HS Group - Produkte Mobility - RAMSIS 2008/2/EG - EU-Richtlinie schnell und rentabel erfüllen,“ 2014. [Online]. Available: http://www.human-solutions.com/mobility/front_content.php?idcat=326&lang=3. [Zugriff am 05 05 2014].
- [4] R. Dörner, B. Jung, P. Grimm und W. Broll, „Worum geht es bei VR/AR? , Was ist VR?, VR/AR Systeme,“ in *s Virtual und Augmented Reality (VR/AR)*, Berlin Heidelberg, Springer-Verlag, 2013, pp. 1 - 24.
- [5] I. E. Sutherland, „The ultimate display,“ in *s Proc IFIP Congress*, 1965.
- [6] S. Bryson, „Call for Participation,“ in *s IEEE Symposium on Research Frontiers in Virtual Reality*, 1993.
- [7] C. Cruz-Neira, „SIGGRAPH '93 Course Notes „Virtual Reality Overview“,“ in *s SIGGRAPH '93*, 1993.
- [8] M. Slater und S. Wilbur, „A Framework for Immersive Virtual Environments (FIVE): speculations on the role of presence in virtual environments,“ *Teleoperators and Virtual Environments*, Bd. 6, Nr. 6, pp. 603-616, 1997.
- [9] R. J. Stone, Earnshaw RA, Gigante MA, Jones H (eds) *Virtual Reality Systems*, London: Academic Press, 1993.
- [10] H. Rheingold, *Virtuelle Welten – Reisen im Cyberspace*, Berlin: Rowohlt Verlag, 1992.
- [11] M. Slater, *A note on presence terminology*, 2003.
- [12] M. Slater, *Place illusion and plausibility can lead to realistic behaviour in immersive virtual*, Royal Society, 2009.
- [13] A. Virtual, „Architectural Visualization,“ *Arch Virtual*, [Online]. Available: <http://archvirtual.com/architectural-visualization/>. [Zugriff am 30 04 2014].
- [14] H. Payne, „Q&Auto: Ford's virtual reality pioneer,“ *detroitnews.com*, 12 04 2014. [Online]. Available: <http://www.detroitnews.com/article/20140412/AUTO0102/304120021>. [Zugriff am 30 04 2014].
- [15] F. Steinicke und R. Dörner, „Wahrnehmungsaspekte von VR,“ in *s Virtual und Augmented Reality (VR AR)*, Berlin Heidelberg, Springer-Verlag , 2013, pp. 33 - 61.
- [16] M. Dahm, *Grundlagen der Mensch-Computer-Interaktion*, New York: Pearson Studium, 2006.

- [17] D. Marr, *Vision: a computational investigation into the human representation and processing of visual information.*, Cambridge: MIT, 1982.
- [18] V. Hayward, O. R. Astley, M. Cruz-Hernandez und D. Grant, „Haptic interfaces and devices,“ *Sensor Review*, Bd. 1, Nr. 24, pp. 16 - 29, 2004.
- [19] S. A. Kuhl, W. B. Thompsen und S. H. Creem-Regehr, „Minification influences spatial judgments in virtual environments,“ in *s APGV '06 Proceedings of the 3rd symposium on Applied perception in graphics and visualization* , New York, 2006.
- [20] M. Kubovy, *The psychology of linear perspective and renaissance art*, Cambridge: Cambridge University, 1986.
- [21] D. Vishwanath, A. R. Girshick und M. S. Banks, „Why pictures look right when viewed from the wrong place,“ *Nature neuroscience*, Bd. 8, Nr. 10, pp. 1401-1410, 2005.
- [22] M. A. Hagen und H. B. Elliot., „An investigation of the relationship between viewing condition and preference for true and modified linear perspective with adults.,“ *Journal of experimental psychology: human perception and performance*, Bd. 4, Nr. 2, p. 479, 1976.
- [23] G. Bruder, A. Pusch und F. Steinicke, „Analyzing effects of geometric rendering parameters on size and distance estimation in on-axis stereographics.,“ in *s Proceedings of the ACM Symposium on Applied Perception. ACM*, 2012.
- [24] S. P. Williams und R. V. Parrish, „New computational control techniques and increased understanding for stereo 3-D displays.,“ *SC-DL tentative. International Society for Optics and Photonics*, pp. 73 - 82, 1990.
- [25] D. M. Hoffman, „Vergence–accommodation conflicts hinder visual performance and cause visual fatigue.,“ *Journal of vision*, Nr. 8, p. 33, 2008.
- [26] C. Ware, C. Gobrecht und M. A. Paton, „Dynamic adjustment of stereo display parameters.,“ *Systems, Man and Cybernetics, Part A: Systems and Humans. IEEE Transactions* , pp. 56-65, 1998.
- [27] V. Interrante, B. Ries und L. Anderson, „Distance perception in immersive virtual environments, revisited.,“ in *s Virtual Reality Conference*, 2006.
- [28] F. Steinicke, „Gradual transitions and their effects on presence and distance estimation.,“ *Computers & Graphics*, Bd. 34, pp. 26 - 33, 2010.
- [29] E. Emmert, „Größenverhältnisse der Nachbilder,“ *Klinische Monatsblätter für Augenheilkunde*, Nr. 19, pp. 443-450, 1881.
- [30] F. Steinicke, „Judgment of natural perspective projections in head-mounted display environments.,“ in *s Proceedings of the 16th ACM Symposium on Virtual Reality Software and Technology*, 2009.
- [31] M. Lappe, M. Jenkin und L. R. Harris, „Travel distance estimation from visual motion by leaky path integration.,“ *Experimental Brain Research*, Nr. 180, pp. 35 - 48, 2007.
- [32] J. M. Loomis und J. M. Knapp, „Visual perception of egocentric distance in real and virtual environments.,“ *Virtual and adaptive environments*, Nr. 11, pp. 21 - 46, 2003.

- [33] F. Steinicke, „Estimation of detection thresholds for redirected walking techniques.,“ *Visualization and Computer Graphics, IEEE Transactions on*, Nr. 16, pp. 17 - 27, 2010.
- [34] M. E. McCauley und T. J. Sharkey, „Cybersickness: perception of self-motion in virtual environments,“ *Presence: Teleoperators and Virtual Environments*, Bd. 1, Nr. 3, pp. 311 - 318, 1992.
- [35] L. Oppermann, C. Geiger, R. Dörner und V. Paelke, „Interaktionen in Virtuellen Welten - Navigation,“ in *s Virtual und Augmented Reality (VR/AR)*, Berlin, Springer-Verlag, 2013, pp. 168 - 177.
- [36] D. A. Bowman und L. F. Hodges, „Formalizing the design, evaluation, and application of interaction,“ *J Vis Lang & Comp*, Bd. 10, pp. 37 - 53, 1999.
- [37] M. Usoh und K. Arthur, „Walking > walking-in-place > flying, in virtual environments,“ in *s Proceedings of the 26th annual conference on Computer graphics and interactive techniques*, 1999.
- [38] E. A. Suma, S. L. Finkelstein, M. Reid, A. Ulinski und L. F. Hodges, „Real walking increases simulator sickness in navigationally complex virtual environments.,“ in *s Virtual Reality Conference, 2009.*
- [39] T. C. Peck, H. Fuchs und M. C. Whitton, „An evaluation of navigational ability comparing Redirected Free Exploration with Distractors to Walking-in-Place and joystick locomotion interfaces.,“ in *s Virtual Reality Conference (VR)*, Singapore, 19-23 March 2011.
- [40] Microsoft, „Kinect for Windows,“ [Online]. Available: <http://www.microsoft.com/en-us/kinectforwindows/>. [Zugriff am 04 06 2014].
- [41] A. C. Inc, „Use Xtion PRO developer solution to make motion-sensing applications and games,“ [Online]. Available: http://www.asus.com/Multimedia/Xtion_PRO_LIVE/. [Zugriff am 04 06 2014].
- [42] P. Grimm, R. Herold und J. Hummel, „Optisches Tracking, Weitere Eingabegeräte,“ in *s Virtual und Augmented Reality (VR/AR)*, Berlin Heidelberg, Springer-Verlag, 2013, pp. 104 - 115.
- [43] R. Hartley und A. Zisserman, *Multiple view geometry in computer vision*, Cambridge university press, 2003.
- [44] G. Reitmayr und T. Drummond, „Going out: robust model-based tracking for outdoor augmented reality,“ in *s Fifth IEEE/ACM International Symposium on Mixed and Augmented, ISMAR 2006*, Santa Barbara, CA, USA, 2006.
- [45] D. G. Lowe, „Distinctive Image Features from Scale-Invariant Keypoints,“ *International Journal of Computer Vision.*, Bd. 60, Nr. 2, pp. 91-110, 2004.
- [46] H. Bay, T. Tuytelaars und L. Van Gool, „SURF: Speeded Up Robust Features,“ in *s Proceedings of the 9th European Conference on Computer Vision, Springer Verlag*, 2006.
- [47] P. Grimm, R. Herold und J. Hummel, „VR-Eingabegeräte - Grundlagen,“ in *s Virtual und*

- Augmented Reality (VR/AR)*, Berlin Heidelberg, Springer-Verlag, 2013, pp. 98 - 104.
- [48] M. Meehan, S. Razzaque, M. C. Whitton und F. P. Brooks, „Effect of Latency on Presence in Stressful Virtual Environments,“ in *s VR '03 Proceedings of the IEEE Virtual Reality 2003*, IEEE Computer Society Washington, DC, USA.
- [49] F. P. Brooks, „What's real about virtual reality?.,“ *Computer Graphics and Applications*, Bd. 6, Nr. 19, pp. 16 - 27, 1999.
- [50] S. R. Ellis und e. al., „Factors influencing operator interaction with virtual objects viewed via head-mounted see-through displays: viewing conditions and rendering latency,“ in *s irtual Reality Annual International Symposium, 1997., IEEE 1997.*, 1997.
- [51] ImmerSight GmbH, „ImmerSight,“ ImmerSight GmbH, 2014. [Online]. Available: <http://www.immersight.de/>. [Zugriff am 26 04 2014].
- [52] S. Hoermann, „Robot localization using 3D-models and an off-board monocular camera,“ in *s Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on*, Barcelona, 2012.
- [53] World Viz LLC, „Wand & Eyes,“ World Viz, 2014. [Online]. Available: <http://www.worldviz.com/products/ppt/wand-and-eyes#Wand>. [Zugriff am 28 04 2014].
- [54] World Viz LLC, „PPT E Highest Performance Wide-area Optical Motion Tracking,“ World Viz, 2014. [Online]. Available: <http://www.worldviz.com/products/ppt/ppt-e>. [Zugriff am 28 04 2014].
- [55] Xsens Technologies B.V., „Xsens MVN is a full-body, camera-less inertial motion capture (MoCap) solution. It is a flexible system that can be used indoors or outdoors (on-set).,“ Xsens, 2014. [Online]. Available: <http://www.xsens.com/products/xsens-mvn/>. [Zugriff am 27 06 2014].
- [56] Xsens Technologies B.V., „ MVN MotionGrid is a position aiding system for Xsens MVN. MotionGrid enables real-time driftless multi-person recording.,“ Xsens, 2014. [Online]. Available: <http://www.xsens.com/products/mvn-motiongrid/>. [Zugriff am 27 06 2014].
- [57] J. S. Shotton, A. W. Fitzgibbon und T. Sharp, „Real-time human pose recognition in parts from single depth images,“ in *s 2011 IEEE Conference on Computer Vision and Pattern Recognition CVPR'11*.
- [58] Primesense, „OpenNI 2 SDK Binaries & Docs,“ [Online]. Available: <http://structure.io/openni>. [Zugriff am 04 06 2014].
- [59] E. A. Suma, B. Lange, S. Rizzo, D. Krum und M. Bolas, „University of Southern California - Flexible Action and Articulated Skeleton Toolkit (FAAST),“ [Online]. Available: <http://projects.ict.usc.edu/mxr/faast/>. [Zugriff am 04 06 2014].
- [60] D. o. C. S. University of North Carolina, „VRPN 07.31 Virtual Reality Peripheral Network,“ [Online]. Available: <http://www.cs.unc.edu/Research/vrpn/>. [Zugriff am 04 06 2014].
- [61] A. C. Inc, „Xtion PRO LIVE specifications,“ [Online]. Available:

- http://www.asus.com/Multimedia/Xtion_PRO_LIVE/specifications/. [Zugriff am 04 06 2014].
- [62] Oracle, „New I/O APIs,“ [Online]. Available: <http://docs.oracle.com/javase/1.5.0/docs/guide/nio/index.html>. [Zugriff am 29 05 2014].
- [63] J. Shotton, „Efficient human pose estimation from single depth images,“ *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, Bd. 12, Nr. 35, pp. 2821-2840, 2013.
- [64] A. Donath, „Übernahme: Facebook kauft Oculus VR für 2 Milliarden US-Dollar,“ *Golem.de*, 26 03 2014. [Online]. Available: <http://www.golem.de/news/uebernahme-facebook-kauft-oculus-vr-fuer-2-milliarden-us-dollar-1403-105375.html>. [Zugriff am 11 06 2014].
- [65] OculusVR, „Oculus Rift: Next-Gen Virtual Reality,“ OculusVR, 2014. [Online]. Available: <http://www.oculusvr.com/rift/>. [Zugriff am 11 06 2014].
- [66] Sony Computer Entertainment Inc., „Sony Computer Entertainment Announces "Project Morpheus" - A Virtual Reality System That Expands The World Of Playstation®4,“ 18 03 2014. [Online]. Available: <http://www.sony.com/SCA/company-news/press-releases/sony-computer-entertainment-america-inc/2014/sony-computer-entertainment-announces-project-morp.shtml>. [Zugriff am 11 06 2014].
- [67] M. A. Fischler und R. C. Bolles, „Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography,“ *Commun. ACM*, pp. 381-395, June 1981.
- [68] B. G. Witmer und M. J. Singer, „Measuring presence in virtual environments: a presence questionnaire,“ *Teleoperators and Virtual Environments*, pp. 225 - 240, 1998.
- [69] M. Leach, „On Making Joseph Wright AR - Location-Based AR,“ 01 2012 . [Online]. Available: http://www.johngoto.org.uk/Joseph_Wright/pages/Image_Tracking_AR.htm. [Zugriff am 13 06 2014].
- [70] E. Zatepyakin, „AS SURF,“ 2012. [Online]. Available: <https://code.google.com/p/in-spirit/wiki/ASSURF>. [Zugriff am 21 06 2014].

Versicherung über Selbstständigkeit

Hiermit versichere ich, dass ich die vorliegende Arbeit ohne fremde Hilfe selbstständig verfasst und nur die angegebenen Hilfsmittel benutzt habe.

Hamburg, den _____