

# Fernunterstützung und Zusammenarbeit mit 3D Punktwolken

Masterarbeit  
von

Kai Westerkamp

An der Fakultät für Informatik  
Fraunhofer IOSB (IAD)

Erstgutachter:	Prof. Dr.-Ing. Rainer Stiefelhagen
Zweitgutachter:	XXXX
Betreuender Mitarbeiter:	M.Sc. Adrian Hoppe

Bearbeitungszeit: 01.06.2017 – 30.11.2017



# Abstract

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.



# Zusammenfassung

Dies hier ist ein Blindtext zum Testen von Textausgaben. Wer diesen Text liest, ist selbst schuld. Der Text gibt lediglich den Grauwert der Schrift an. Ist das wirklich so? Ist es gleichgültig, ob ich schreibe: „Dies ist ein Blindtext“ oder „Huardest gefburn“? Kjift – mitnichten! Ein Blindtext bietet mir wichtige Informationen. An ihm messe ich die Lesbarkeit einer Schrift, ihre Anmutung, wie harmonisch die Figuren zueinander stehen und prüfe, wie breit oder schmal sie läuft. Ein Blindtext sollte möglichst viele verschiedene Buchstaben enthalten und in der Originalsprache gesetzt sein. Er muss keinen Sinn ergeben, sollte aber lesbar sein. Fremdsprachige Texte wie „Lorem ipsum“ dienen nicht dem eigentlichen Zweck, da sie eine falsche Anmutung vermitteln.



# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>1</b>
1.1	VR . . . . .	1
<b>2</b>	<b>Stand der Technik</b>	<b>3</b>
2.1	Section . . . . .	3
2.1.1	Ungenauigkeiten im Lighthouse Tracking . . . . .	3
<b>3</b>	<b>Punktwolke</b>	<b>5</b>
3.1	Frames aufnehmen und bereinigen . . . . .	6
3.1.1	Aufnahme und Glättung . . . . .	6
3.2	Zusammenfügen von Frames . . . . .	6
3.2.1	Kalibrierung Kinect zu Vive . . . . .	7
3.3	Ergebnisse . . . . .	9
<b>4</b>	<b>3D Tiles un GLTF</b>	<b>11</b>
4.1	3D Tiles . . . . .	11
4.1.1	Tileset und Tiles . . . . .	11
4.2	glTF . . . . .	12
4.2.1	Struktur . . . . .	12
4.2.1.1	Buffers and Accessors . . . . .	13
	<b>Literaturverzeichnis</b>	<b>15</b>





# 1. Einleitung

(TODO)

ToDo

## 1.1 VR



## 2. Stand der Technik

related

### 2.1 Section

paper

#### 2.1.1 Ungenauigkeiten im Lighthouse Tracking

Ein großes Problem sind Ungenauigkeiten im Lighthouse Tracking.

Noise in der Ruhelage 0.3mm [lig]

Im Paper [NLL17] wurden signifikante Fehler nach Tracking Abbrüchen festgestellt. bsi zu 150cm

$2m = 1,98m$

**(bild) (schreiben)**

Tracker trackings

**ToDo**  
**ToDo**



### 3. Punktwolke

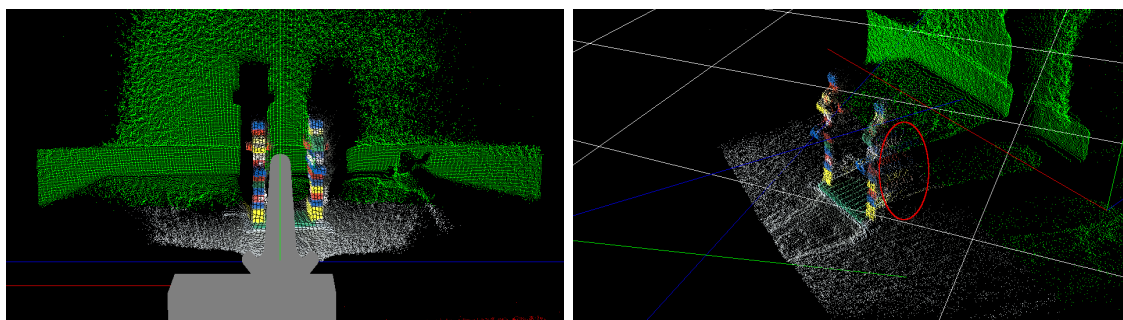
Das Erzeugen der Punktwolke soll einfach und schnell funktionieren und mit einer Kinect erfolgen. Aus einem Frame der Kinect, bestehend aus Farbbilde und Tiefenbild, lässt sich einfach eine Punktwolke relativ zur Tiefenkamera der Kinect errechnen. Eine Aufnahme beinhaltet aber nur alle Informationen die aus den 2D Bilder errechnet werden können. Das heißt man erhält nur eine Seite des Objektes gut aufgelöst und einige Artefakte die sich aus dieser Berechnung ergeben. Für die Darstellung in einer VR Umgebung ist dies nicht ausreichend. Der Betrachter kann sich frei in der virtuellen Welt bewegen und erkennt schnell die nicht vorhandenen Informationen und Fehler.

Ein Fehler entsteht zum Beispiel bei Kanten und Flächen die nicht Senkrecht zur Kamera sind. Die unausreichende Informationen in den Ausgangsdaten werden falsch angenähert und ergeben falsche Flächen. (siehe 3 ). **(Bild 2 ändern / vergrößern?)**

**ToDo**

Das zweite Problem das es zu lösen galt war das zusammenfügen von mehreren Aufnahmen aus unterschiedlichen Perspektiven zu einer großen zusammenhängenden Punktwolke. Um 2 Frames miteinander zu verbinden braucht man die relative Transformation zwischen den beiden Aufnahmen, bzw. absolute Kamerapositionen. Bei bestehenden Algorithmen wird dies zum Beispiel durch Featureerkennung oder zurückrechnen der Kamerabewegung erreicht **(quellen)**. Solche verfahren sind meist rechenaufwändig und zeitintensiv. Für diese Arbeit war es das Ziel die Kinect mit dem Lighthouse Tracking System zu verbinden. Die Trackingdaten aus SteamVR, bzw OpenVR geben uns eine globale Position aller

**ToDo**



(a) Aufnahme aus Sicht der Kinect

(b) Aufnahme von der Seite

Abbildung 3.1: Aufnahmen aus verschiedenen Perspektiven In Bild b) sind falsche Punkte zu sehen die durch die Rekonstruktion aus einem 2D Bild entstehen.

Aufnahmen und vereinfachen das Erzeugen einer großen Punktwolke.

### 3.1 Frames aufnehmen und bereinigen

Als ersten Schritt wird ein Frame mit der Kinect aufgenommen und das Tiefenbild geglättet. Anschließend wird das Tiefen- und Farbbild in 3D Punkte umgewandelt und unerwünschte Punkte verworfen.

#### 3.1.1 Aufnahme und Glättung

Das Aufnehmen einer kleinen Punktwolke wird das Kinect SDK verwendet. Sowohl Tiefenbild auch als auch Farbbild kann man aus der API erhalten. Anschließend wird das Tiefenbild geglättet. Die Rohdaten sind teilweise sehr verrauscht und so erhält man eine Punktwolke mit glatteren Flächen. Hierfür braucht man einen Filter der zwar die Flächen glättet, aber gleichzeitig Objektkanten erhält. Ein Bilateral Filter erzielt den gewünschten Effekt ist aber relativ Rechenaufwändig. Verwendet wurde der Filter der in dem Paper [MCS14] vorgestellt wird. Hierbei wird zunächst das Bild mit einem Gausfilter geglättet und anschließend mit dem Original verglichen um dabei entstehende Artefakte zu entfernen.

Nach der Glättung des Tiefenbildes wird dieses in eine Punktwolke umgewandelt. Hierfür wurde ebenfalls das Microsoft Kinect SDK verwendet das alle benötigten Methoden bereitstellt.

Nach der Umwandlung werden noch weiter Punkte verworfen. Zunächst werden alle Punkte zu denen keine Farbe zugeordnet werden kann verworfen. Auch alle Punkte die zu nah oder zu weit vom Sensor entfernt sind, werden nicht weiter betrachtet. Je weiter das Objekt entfernt, desto ungenauer werden die Aufnahmen. Im folgenden wurde ein Mindestabstand von 30cm und ein Maximalabstand von 90cm verwendet.

Als letztes filtern wir alle Flächen die nicht Senkrecht zur Kamera sind (siehe Abb 3 b) Diese Flächen entstehen durch die Umwandlung von einem 2D Tiefenbild in einer 3D Punktwolke. Die benötigten Informationen fehlen an dieser Stelle und Punkte werden auf die Fläche zwischen Oberflächenobjekt und Hintergrund gesetzt. Diese Ebene stimmt nicht mit der wirklichen Oberfläche überein und müssen entfernt werden. Hierfür wird die Oberflächennormale verwendet. Die Normale wird aus dem Tiefenbild geschätzt.

$$\begin{aligned} dzXAxis &= depthAt[x+1, y] - depthAt[x-1, y] \\ dzYAxis &= depthAt[x, y+1] - depthAt[x, y-1] \\ Normale &= Normalize(-dzXAxis, -dzYAxis, 1.0) \end{aligned} \quad (3.1)$$

Mit dem Skalarprodukt lässt sich der Winkel zwischen dem Kameravektor  $(0, 0, 1)$  und Normale ausrechnen. Ein maximaler Winkel von  $65^\circ$  hat in den Tests ein gutes Ergebnis geliefert. **(Quellen auf Kinect und Lighthouse)**

ToDo

### 3.2 Zusammenfügen von Frames

Ein wichtiger Teil beim dem Aufnehmen der Punktwolke ist das zusammenführen von mehreren Frames. Hierfür wurde die Kinect mit dem Lighthouse Tracking System verbunden und verzichtet damit auf aufwändige Berechnungen.

ToDo

**(Foto Halterung)** Im lokalen Koordinatensystem der Kinect, also jedes Frames liegt der Ursprung in dem Tiefensensor. die Transformation *transformControllerToKinect* zwischen dem Koordinatensystem des Controllers und der Kinect wurde bestimmt und



Abbildung 3.2: Effekt einer falschen Kalibrierung  $dx$  auf die endgültige Punktwolke. Aufnahme 1 und 2 sind von lokalen Koordinaten in Welt Koordinaten transformiert

die globale Transformation des Controllers *transformController* ist in der OpenVR API abfragbar. Die Transformation der lokalen Punktwolke in ein globale ist mit diesen beiden Transformationen möglich.

$$globalPosition = transformController * transformControllerToKinect * localPosition \quad (3.2)$$

### 3.2.1 Kalibrierung Kinect zu Vive

Eine wichtige Transformation ist die zwischen dem Koordinatensystem der Kinect und dem des Vive Controllers.

Ist zum Beispiel die Transformation entlang der X Achse der Kinect verschoben so verstärkt sich der Fehler wenn man das Objekt von der Anderen Seite, also um 180° dreht aufnimmt (siehe Abb, 3.2.1). Der Fehler im Lokalen Koordinatensystem wird in das Globale transformiert und ist in dem Fall dann in genau entgegengesetzte Richtungen

Bei der Kinect ist der offiziellen Doku entnehmbar das der Ursprung von Punktwolken in dem Tiefensensor liegt (siehe [Kina]). Aber es gibt keine offizielle Dokumentation wo dieser exakt liegt. Im Bild 3.2.1 aus dem chinesischen Microsoft Forum ist eine von Benutzern vermessene schematische Darstellung der Kinect abgebildet. Der Tiefensensor liegt hinter der kleineren runden Öffnung, aber Fertigungsungenauigkeiten lassen keine exakten Daten finden. Für die Implementation wurde mittig hinter der Öffnung angenommen. Eine digitale Kalibrierung gestaltet sich schwierig, da das Lighthouse Tracking für den Controller zusätzlich einige Ungenauigkeiten mit sich bringt. Der Ursprung des Controllers lässt sich aus den Modellen von SteamVR auslesen. Dieser liegt geschickt für VR Anwendungen ist aber für das Tracking von Objekten ungeschickt. bei der Implementation stand noch keine Vive Tracker zur Verfügung. Für die Arbeit wurde der Controller so nah wie möglich an dem Teifensensor, also direkt darüber angebracht (siehe Abb.3.2.1). **(Bilder)**

**ToDo**





### 3.3 Ergebnisse

Mit dem vorgestellten Verfahren lässt sich einfach und schnell eine Punktwolke erstellen. Jedoch gibt es Ungenauigkeiten in dem Vive Tracking und der Kalibrierung die die Punktwolke unbrauchbar aussehen lassen. **(Bild)** Zwischen 2 Aufnahmen und den daraus resultierenden Punktwolken ist ein Versatz bis zu 2-3 cm sichtbar. In einer 3D Umgebung insbesondere in VR ist das eine zu große Ungenauigkeit. Der Versatz zwischen den Punktwolken ist leider nicht konstant und ändert sich teilweise zwischen Durchläufen. Das Vive Tracking ist hierfür ein Grund. Vergleicht man mit einem 2m Zollstock die reale Distanz zu der in VR gemessenen dann wird daraus 1,98-2m virtuelle Distanz. Die Distanz ist hierbei abhängig von der Orientierung zu den Basistationen und der aktuellen Kalibrierung des Lighthous Tracking Systems. Dieses Problem erschwert es die Kalibrierung zwischen Vive und Kinect zu überprüfen die zusätzlich für einen Versatz der Punktwolke verstärken kann.

**ToDo**

Die Rotation der einzelnen Aufnahmen war kein Problem und hat keine sichtbaren Probleme produziert.



## 4. 3D Tiles un GLTF

In diesem Kapitel wird ein grober Überblick über die Struktur und die Komponenten des GL Transmission Formats und der 3D Tiles gegeben. Diese wurden verwendet um und ie Punktwolken zu speichern.

3D Tiles und gltf 3D Tiles are an open specification for streaming massive heterogeneous 3D geospatial datasets Tile Struktur (Tielset) different Tiles (batched, instanced, points ...) Bounding Volumes und LOD Dynamic loading GLTF Struktur Optimized for OpenGL and streaming scenes, nodes, meshes materials animations ignored

### 4.1 3D Tiles

3D Tiles [3DT] ist eine neue offene Spezifikation für das streamen von massiven, heterogenen, geospatialen 3D Datensätzen. Die 3D Tiles können genutzt werden um Gelände , Gebäude, Bäume und Punktwolken zu streamen und beiden Features wie Level of Detail (LOD). Für die Arbeit wurde erwartet das insbesondere LOD notwendig werden könnte, es wurde aber nicht verwendet.

#### 4.1.1 Tileset und Tiles

Als Basis der 3D Teils wird JSON formatiertes Tileset verwendet das auf die eigentlichen Daten in Tiles verweist. Das Tileset hat eine baumartige Struktur aus Tiles und deren Metadaten. Jedes Tile hat hierbei ein 3D Volumen der den geografischen Bereich beschreibt, einen geometrischen Fehler zur Echtwelt. Außerdem können Kinder und deren Transformationen zu dem Elternteil angegeben werden. Alle Kinder liegen hierbei in dem Volumen des Elternknotens und können mit verschiedenen Datenstrukturen, wie K-D Bäumen Quadrees oder ähnlichem die Region genauer spezifizieren (siehe Bild 4.1.1. Hierbei können die Kinder das Elterntile ersetzen (replace, z.B. genaueres Mesh) oder das bestehende Tile ergänzen (refine, zusätzliche Gebäude oder Details). Die eigentlichen Daten der Tiles sind durch eine URL verlinkt und können dynamisch nachgeladen werden.

Tiles können in unterschiedlichen Formaten sein zum Beispiel:

**Batched3D Model** 3D Daten die als glTF übertragen werden. Zusätzlich können pro Modell Metadaten für das Visualisieren enthalten sein.

**Instanced3D Model** Tileformat für Instancing. Die Geometrie wird als glTF übertragen und zusätzlich eine Liste aus Positionen an denen die Objekte Instanziiert werden sollen. Kann zum Beispiel für Bäume genutzt werden.

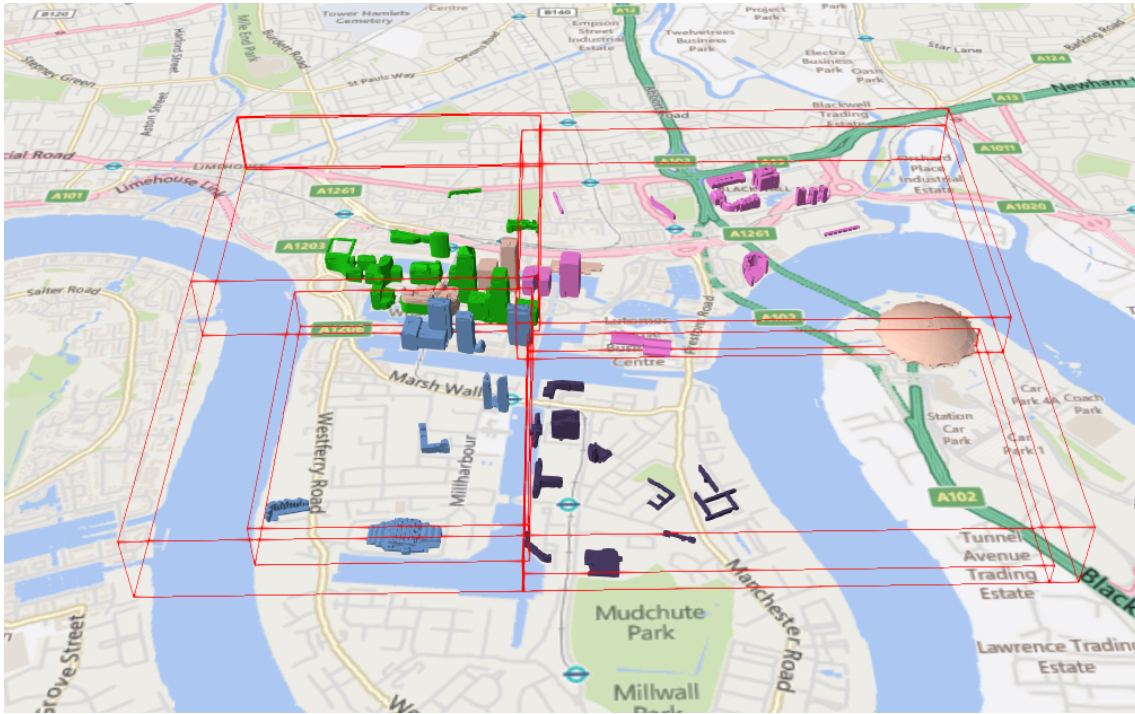


Abbildung 4.1: Ein Tile mit 4 Kindern. Die 4 Kinder fügen die Gebäude hinzu und liegen im Volumen des Elterntiles. Als Datenstruktur liegt ein nicht uniformer Quadtree vor.

**Point Cloud** Format um Punktwolken zu übertragen. Das Teileformat enthält einen kleinen Header mit Metadaten und der Anzahl an Punkten. Außerdem ist enthalten welche und wie die Daten wie Position und Farbe vorliegen. Die eigentlichen Daten werden als Binärdaten übertragen und können so ohne Parsen direkt in den Speicher geladen werden.

**Composite** Tileformat zum gleichzeitigen Übertragen mehrerer einzelner Tileformate in einem. Es lässt sich zum Beispiel ein Batched3D Modell für Gebäude mit Instanced3D Modell für Bäume verbinden und als ein Tile übertragen.

## 4.2 glTF

Das GL Transmission Format (glTF [GLT]) ist ein Format zum effizienten Übertragen von 3D Szenen für GL Api's wie WebGL, OpenGL ES und OpenGL. glTF dient als effizientes, einheitliches und erweiterbares Format zur Übertragung und Laden von 3D Daten. Im Vergleich zu aktuellen Standards wie COLADA ist glTF optimiert, schnell übertragen und kann schnell in eine Applikation geladen werden. In einer JSON formatierten Datei (.gltf) wird eine komplette Szene samt Szenegraf, Materialien und deren zugehörigen Shadern, Kamerapositionen, Animationen und Skinning Informationen übertragen. Dabei kann auf externe Dateien verwiesen werden. Diese sind zum Beispiel Binärdaten oder Bildern die für das einfache und effiziente Übertragen von Geometrie, Texturen oder den nötigen GLSL Shadern genutzt werden

### 4.2.1 Struktur

Die .gltf Datei bildet den Kern jedes Modells und ist eine JSON formatierte Datei. In ihr werden alle grundlegenden Informationen wie zum Beispiel die Baumstruktur und die Ma-

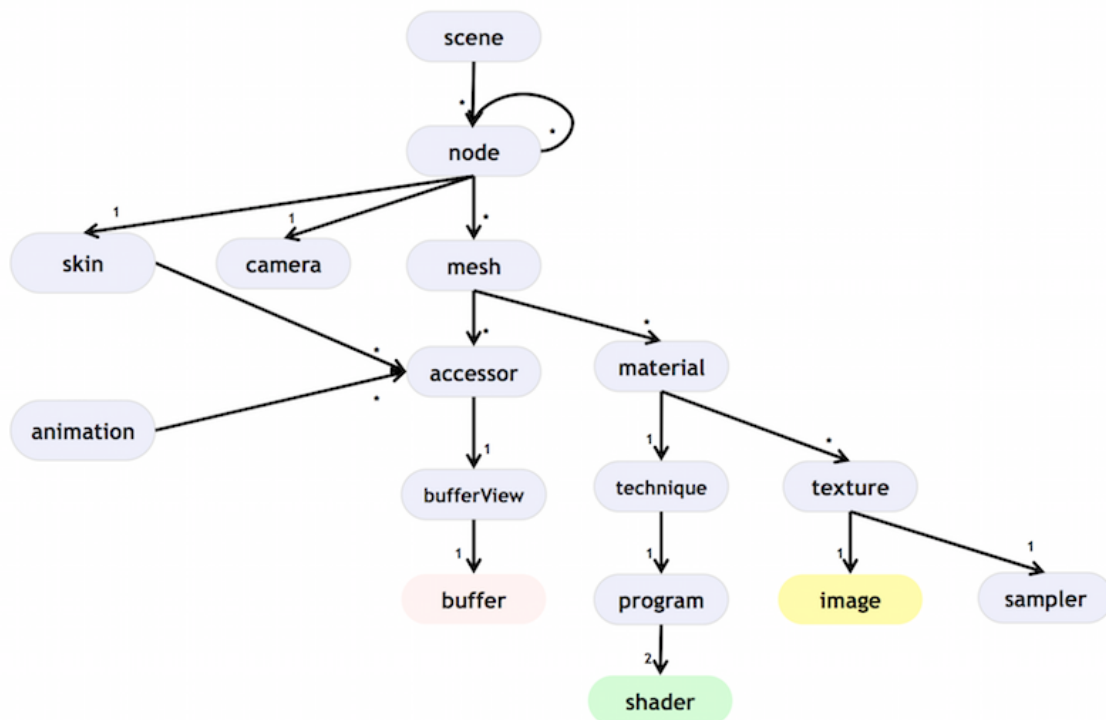


Abbildung 4.2: Struktur einer glTF Datei

terialien gespeichert (siehe Abb. 4.2.1). Eine Szene bildet hierbei den Startpunkt für die zu rendernde Geometrie. Szenen bestehen aus Knoten (Nodes) im Szenengraf, die wiederum Knoten als Kinder haben können. Jeder Knoten kann eine Transformation im lokalen Raum definieren, bestehend aus einer Translation, einer Rotation und einer Skalierung. Jeder Knoten kann eine Mesh und damit die eigentliche Geometrie referenzieren.

#### 4.2.1.1 Buffers and Accessors

Buffer sind die eigentlichen Daten in einem Binären Block. Diese können entweder als externe Datei (.bin) oder als BASE64 encodierter String in der JSON Datei angefügt werden. Die Hauptaufgabe der Buffer ist es große mengen an Daten wie die Geometrie effizient zu übertragen.

glf right Handed y Axis up in Meters and radians



# Literaturverzeichnis

- [3DT] *3d tiles spezifikation.* <https://github.com/AnalyticalGraphicsInc/3d-tiles>. Accessed: 2017-09-13.
- [GLT] *Gltf spezifikation.* <https://github.com/KhronosGroup/glTF>. Accessed: 2017-09-13.
- [Kina] *Kinect dokumentation koordinatensysteme.* <https://msdn.microsoft.com/de-de/library/dn785530.aspx>. Accessed: 2017-11-02.
- [Kinb] *Kinect tiefensensor position.* <https://social.msdn.microsoft.com/Forums/sqlserver/ja-JP/05a6d2b3-9096-4236-b77a-691c5f047066/kinect-for-windows-v2-?forum=windowsgeneraldevelopmentissuesja>. Accessed: 2017-11-02.
- [lig] *Lighthouse tracking examined.* <http://doc-ok.org/?p=1478>. Accessed: 2017-11-02.
- [MCS14] Manuel Martin, Florian van de Camp, and Rainer Stiefelhagen: *Real time head model creation and head pose estimation on consumer depth cameras.* In *Proceedings of the 2014 2Nd International Conference on 3D Vision - Volume 01*, 3DV '14, pages 641–648, Washington, DC, USA, 2014. IEEE Computer Society, ISBN 978-1-4799-7000-1. <http://dx.doi.org/10.1109/3DV.2014.54>.
- [NLL17] Diederick Christian Niehorster, Li Li, and Markus Lappe: *The accuracy and precision of position and orientation tracking in the htc vive virtual reality system for scientific research.* In *i-Perception*, 2017.





---

Ich versichere wahrheitsgemäß, die Arbeit selbstständig angefertigt, alle benutzten Hilfsmittel vollständig und genau angegeben und alles kenntlich gemacht zu haben, was aus Arbeiten anderer unverändert oder mit Abänderungen entnommen wurde.

**Karlsruhe, xx.xx.xx**

.....  
(Max Mustermann)