

# Evaluation von Headtracking in interaktiven virtuellen Umgebungen auf Basis der Kinect

Dennis Wiebusch\*, Martin Fischbach\*, Alexander Strehler\*, Marc Erich Latoschik\*  
Gerd Bruder †, Frank Steinicke †

\*Human-Computer Interaction

Universität Würzburg

Am Hubland

97074 Würzburg

Tel.: +49 (0)931 / 31 86313

dennis.wiebusch@uni-wuerzburg.de

†Immersive Media Group

Universität Würzburg

Oswald-Külpe-Weg 82

97074 Würzburg

Tel.: +49 (0)931 / 31 85868

gerd.bruder@uni-wuerzburg.de

**Zusammenfassung:** Interaktive Medien mit räumlicher Darstellung virtueller Inhalte finden verstärkt Einzug in verschiedenste Anwendungsfelder. Diese Entwicklung führt zu einer stetig wachsenden Nachfrage nach kostengünstigen Verfahren zur Bestimmung der Kopfposition eines Betrachters, welche zu einer perspektivisch korrekten Darstellung nötig sind. In diesem Beitrag evaluieren wir zwei Headtrackingverfahren in interaktiven virtuellen Umgebungen auf Basis der Microsoft Kinect. Wir vergleichen die beiden Verfahren mit einem professionellen optischen Trackingsystem und zeigen Vor- und Nachteile auf.

**Stichworte:** Headtracking, Virtuelle Umgebungen, Microsoft Kinect

## 1 Einleitung

Ein wichtiger Aspekt immersiver virtueller Umgebungen ist die interaktive Darstellung virtueller Szenen unter Verwendung von perspektivisch korrekten und stereoskopischen visuellen Stimuli. Diese ermöglichen die Exploration virtueller Datensätze vergleichbar mit der Exploration einer Umgebung in der realen Welt. Sie benötigen ein präzises Tracking der Benutzerkopfposition und -ausrichtung mit möglichst geringer Latenz. Verschiedene Technologien, mit denen die Kopfposition eines Betrachters mit hoher Genauigkeit und geringer Latenz erfasst werden kann, sind im Forschungsbereich der virtuellen Realität (VR) vorgestellt worden. Unter Einsatz der erfassten Kopfposition in Trackingkoordinaten sowie einer Abbildung auf eine Kamera in virtuellen Koordinaten ist es damit möglich, einem Betrachter natürliches Feedback zu Bewegungen in der realen Welt zu geben. Die Freiheit, durch natürliche Bewegungen virtuelle Umgebungen zu erkunden, hat großes Potenzial für verschiedene Anwendungen, z.B. für Produktdemonstrationen computer-generierter Datensätze, oder um die Präsenz von Anwendern in virtuellen Welten zu steigern.

Professionelle VR-Installationen verwenden häufig kostenintensive Tracking- und Displaytechnologien aus dem Forschungsumfeld. Dagegen haben sich in den letzten Jahren kostengünstige Technologien im Consumer-Marktsegment etabliert, wie etwa verschiedene

Lösungen zum Headtracking mit der Nintendo Wiimote, der Sony EyeToy oder der Microsoft Kinect. Obwohl die Genauigkeit und Latenzzeiten dieser Trackingtechnologien nicht mit denen professioneller Systeme vergleichbar sind, besteht aufgrund des hohen Kostenunterschieds Interesse, diese für weniger kritische Anwendungen einzusetzen. Dabei stellen insbesondere aus einem oder zwei Bildschirmen bestehende Systeme, wie beispielsweise VR-Workbenches, Multi-Touch-Tables oder Geräte aus dem 3D-Heimkinobereich Anwendungsfälle dar. Bei letzteren handelt es sich z.B. um Projektoren und Großbildschirme unter Einsatz von Shuttertechnologien, deren Verkaufszahlen stetig zunehmen.

In dieser Arbeit beschreiben wir zwei Ansätze zum Headtracking auf Basis der Microsoft Kinect, welche wir mit einem professionellen Trackingsystem (WorldViz PPT X4) vergleichen. Die beiden Kinect-basierten Verfahren unterscheiden sich grundlegend. Der erste Ansatz verwendet eine, durch das OpenNI Framework [Pri12] bereitgestellte, Skelett-basierte Rekonstruktion der Kopfposition, der zweite, im Zuge dieser Arbeit entwickelte Ansatz kombiniert das Skelett-basierte Tracking mit einem marker-basierten Ansatz ohne Hinzunahme weiterer Hardware (abgesehen von einem aktiven Target) [SBWS12].

Drei maßgebliche qualitative Kriterien für die Evaluation eines Trackingsystems sind:

*Präzision:* Streuung der Daten bei wiederholtem Messen des selben Punktes

*Genauigkeit:* Streuung der Daten um den realen Wert

*Latenz:* Zeitdifferenz zwischen Samplezeit und Bereitstellungszeit

Die vorliegende Arbeit untersucht die *Präzision* der drei Verfahren und geht dabei von folgenden Hypothesen aus:

*H1:* Das professionelle System hat insgesamt die höchste Präzision.

*H2:* Das Skelett-basierte Verfahren hat die niedrigste Präzision.

*H3:* Ein hybrides Verfahren erhöht die Präzision des Skelett-basierten Verfahrens.

Dieser Beitrag ist wie folgt strukturiert: Abschnitt 2 gibt einen Überblick über verwandte Arbeiten. In Abschnitt 3 stellen wir die Verfahren zum Headtracking auf Basis der Kinect vor. In Abschnitt 4 beschreiben wir die durchgeführte Evaluation der Verfahren. Abschnitt 5 fasst den Beitrag zusammen und gibt einen Überblick über zukünftige Forschungen.

## 2 Verwandte Arbeiten

Interaktive virtuelle Umgebungen ermöglichen Betrachtern einen räumlichen Eindruck dreidimensionaler virtueller Datensätze zu gewinnen. Dies wird insbesondere durch stereoskopische Darstellung virtueller Szenen ermöglicht, wodurch Distanzeinschätzungen vereinfacht und Mehrdeutigkeiten in räumlichen Verhältnissen reduziert werden, sowie der Aufbau einer mentalen Repräsentation dargestellter Szenen unterstützt wird [TFCRS11]. Darüber

hinaus ermöglicht Bewegungsparallaxe bei getrackten Eigenbewegungen einem Betrachter zusätzliche Tiefenhinweise zu gewinnen, welche von traditionellen Darstellungsmedien nicht unterstützt werden [RG79, NJ06]. Speziell für diese beiden Tiefenhinweise ist es essenziell, dass die Position der Augen des Betrachters korrekt und in Echtzeit erfasst wird.

Verschiedene Technologien existieren, um dieses Problem zu lösen [MG01, MCT09]. Dazu zählen marker-basierte *outside-in* Ansätze auf Basis von Triangulation mit mehreren optischen Sensoren (z.B. Zeilenkameras [Atr12] oder Kameras mit 2D-Pixelrastern [Wor12]), Tiefen-, Form- oder Featurepunkt-basierte Trackingverfahren auf Basis einer oder mehrerer optischer Sensoren [See12], Ansätze zum optischen *inside-out* Tracking [WBV<sup>+</sup>99] sowie diverse Verfahren zum magnetischen, mechanischen oder inertialen Tracking [WF02]. Mehrere dieser Ansätze sind zu professionellen Trackingsystemen entwickelt worden und finden Einsatz in verschiedenen VR Installationen.

Während diese Trackingsysteme weltweit häufig nur in ein-, zwei- oder dreistelligen Mengen abgesetzt werden, was einen direkten Einfluss auf den Preis dieser Systeme hat, sind in den letzten Jahren kostengünstigere Lösungen (z.B. von Nintendo, Sony und Microsoft) verfügbar geworden, deren Preis sich aufgrund von Massenproduktion deutlich von professionellen Systemen unterscheidet. Insbesondere zu nennen ist der Nintendo Wii Remote Controller, welcher eine einfache Kamera mit Infrarotfilter beinhaltet, in deren Kamerabild die Position von IR-Punkten detektiert werden kann. Durch Triangulation einer IR-LED von mehreren Wiimotes, bzw. Erkennung mehrerer IR-LEDs durch eine Wiimote haben verschiedene Forscher gezeigt, dass sich praktisch einsetzbare Trackingsysteme für wenige hundert Euro entwickeln lassen, auch wenn die Genauigkeit und Präzision i.A. nicht das Niveau professioneller Systeme erreicht [Lee08]. Ähnlich kostengünstige Trackingsysteme sind auf Basis des natürlichen Farbraums vorgestellt worden, z.B. für die Sony EyeToy oder unter Verwendung einer oder mehrerer Webcams [See12, PDS<sup>+</sup>09, FWGVG11, FMO12].

Obwohl diverse anwendbare Verfahren zum Benutzertracking bereits existierten, wird die Microsoft Kinect seit ihrer Vorstellung in 2010 für verschiedenste Anwendungen bevorzugt als Trackingsystem eingesetzt. Gründe hierfür sind insbesondere die Fähigkeit, den gesamten Körper eines oder mehrerer Benutzer markerlos zu tracken, sowie die bereitgestellten Computer Vision Frameworks (z.B. Primesenses OpenNI SDK und Microsofts Kinect SDK) und mehrere freie Open Source Entwicklungen und Treiber [Pri12, Mic12, SLR<sup>+</sup>11, GP11].

Im Folgenden zeigen wir, dass sich die vom OpenNI-Framework bestimmte Kopfposition aufgrund unpräziser Trackingdaten nur bedingt für Headtracking einsetzen lässt. Dem gegenüber zeigen wir ein neues Verfahren als Erweiterung auf, mit dem die Vorteile der Kinect gegenüber anderen Trackingansätzen erhalten bleiben (insbesondere das Vollkörpertracking), jedoch die Erkennung der Kopfposition auf einfache Weise verbessert werden kann.

### 3 Kinect-basierte Headtrackingverfahren

In diesem Abschnitt beschreiben wir zwei Verfahren zum Headtracking in interaktiven virtuellen Umgebungen auf Basis der Kinect: (1) die Approximation der Augenpositionen unter

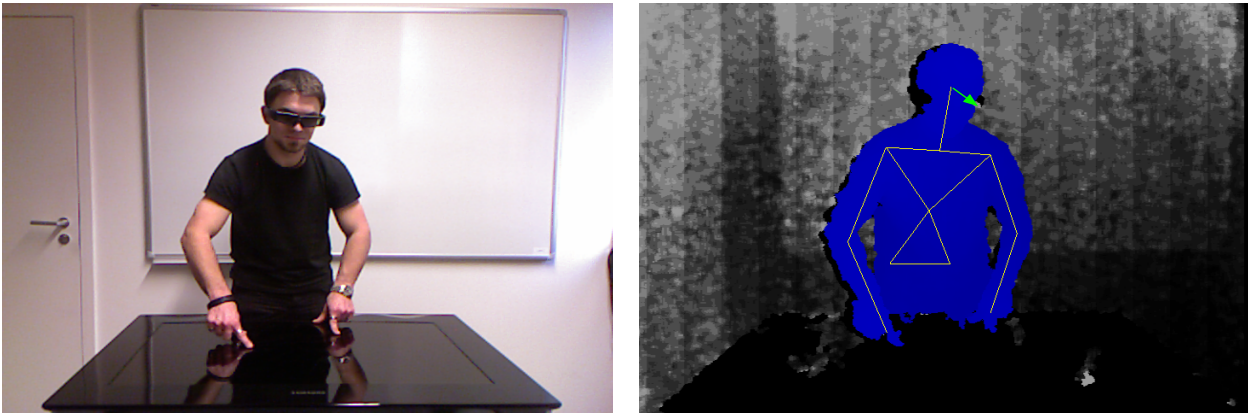


Abbildung 1: Illustration von Skelett-basiertem Headtracking mit OpenNI: (links) RGB-Bild der Kinect und (rechts) intern berechnetes Tiefenbild, sowie resultierendes Skelettmodell mit Position und approximierter Orientierung des Kopfes.

Verwendung des markerlosen, form-basierten Skeletttrackings des OpenNI-Frameworks. (2) ein hybrides Verfahren, welches sowohl die OpenNI-Skelettdaten als auch die nativ verfügbaren Daten der Kinect verwendet, um die Position eines aktiven IR-Markers zu bestimmen.

### 3.1 Skelett-basiertes Headtracking

Die Microsoft Kinect ist ein Multisensorsystem, welches unter anderem eine RGB- und IR-Kamera mit einer Auflösung von  $640 \times 480 @ 30\text{Hz}$  (bzw.  $1280 \times 1024 @ 15\text{Hz}$ ) sowie einen IR-Projektor beinhaltet, der *strukturiertes Licht* [SS03] emittiert. Die bekannte Anordnung dieser Elemente im Gehäuse der Kinect ermöglicht es, für kurze Distanzen bis ca. 3–4m, eine RGB-D Tiefenkarte für die Pixel der RGB-Kamera zu bestimmen [SJP11, KE12]. Anhand der Tiefeninformationen lassen sich verschiedene Benutzerkörper unterscheiden sowie jeweils ein vereinfachtes Skelettmodell an die Tiefenwerte annähern [Pri12, Mic12]. Obwohl sich die Skelettmodelle in den verbreitetsten SDKs von PrimeSense und Microsoft unterscheiden, haben sie gemein, dass jeweils eine Kopfposition und ihre Orientierung angenähert wird.

Im Rahmen dieser Arbeit verwenden wir das OpenNI-Framework für das Skeletttracking eines Benutzerkörpers. Dieses bietet die Möglichkeit auf alle Gelenkpositionen eines aus den Tiefendaten berechneten Ganzkörperskeletts zuzugreifen (siehe Abb. 1). Die durch das OpenNI-Framework bereitgestellten Positions- und Orientierungsdaten des Kopfes – letztere werden auf Grundlage der Orientierung des Torso approximiert – kann zum Headtracking eingesetzt werden. Um die Position der Augen zu bestimmen, wird unter Verwendung dieser Orientierung ein experimentell ermittelter, konstanter Versatz auf die Kopfposition addiert.

Die Aktualisierung der Kopfposition und -orientierung ist, aufgrund der Spezifikationen des Kinect Sensors, auf 30 Hz beschränkt. Statt direkt auf das Skelettmodell von OpenNI zuzugreifen wird in interaktiven virtuellen Umgebungen häufig ein Proxy eingesetzt, wie z.B. das *Flexible Action and Articulated Skeleton Toolkit* (FAAST) [SLR<sup>+</sup>11], welches die berechneten Gelenkdaten von OpenNI via VRPN [THS<sup>+</sup>01] im Netzwerk bereitstellt.



Abbildung 2: Illustration von Marker-basiertem Headtracking mit der Kinect: RGB-Bild (links), intern berechnetes Tiefenbild (Mitte) sowie IR-Bild der Kinect (rechts). Der IR-Marker erscheint als hellster Punkt im IR-Bild und führt zu einer lokalen Störung im berechneten Tiefenbild (weiße Bereiche stellen undefinierte Werte dar).

### 3.2 Hybrides Headtracking

Basierend auf den nativ von der Kinect zur Verfügung gestellten RGB-, IR- und Tiefenbilddaten stellen wir im Folgenden einen hybriden markerbasierten Ansatz für das Headtracking mit der Kinect vor. Dieser hat das Potenzial die Qualität der detektierten Kopfposition gegenüber der Verwendung der unveränderten Kopfposition der Skelettrekonstruktion in Abschnitt 3.1 zu verbessern. Erreicht wird dies durch die Anbringung eines aktiven kugelförmigen IR-Markers am Kopf des Benutzers (z.B. an einer Shutterbrille, siehe Abb. 2). Der Vorteil der Verwendung eines solchen Markers ist die Unabhängigkeit von der Umgebungsbeleuchtung. Deren Ausmaß muss beispielsweise beim Einsatz von Multi-Touch-Tables gering gehalten werden, um ihre Funktionalität zu gewährleisten.

Durch das vom Marker emittierte Infrarotlicht erscheint dieser als hellster Punkt im Bild der IR-Kamera der Kinect und kann somit leicht (z.B. mit Hilfe von OpenCV [Bra00]) detektiert werden. Um die Stabilität dieser Bestimmung zu erhöhen, wird das IR-Bild in einem ersten Schritt mit morphologischen Operatoren gefiltert. Dadurch werden einzelne helle Bildpunkte abgedunkelt, welche z.B. durch das strukturierte Licht der Kinect entstehen. Anschließend wird das Zentrum des Markers im Bild bestimmt, indem der Mittelpunkt aller in der lokalen Umgebung des hellsten Pixels vorkommenden, ausreichend hellen Pixel ermittelt wird. Die in diesem Zusammenhang verwendete Mindesthelligkeit wird dynamisch, in Abhängigkeit von der Intensität des hellsten Pixels in der Suchregion bestimmt. Je nach Größe des Markers im Bild kann dessen Position so mit Subpixelgenauigkeit bestimmt werden.

Dieses Verfahren ist zunächst anfällig für helle Reflexionen im IR-Bild der Kinect, welche z.B. durch reflektierende Oberflächen hervorgerufen werden können. Mittels eines hybriden Ansatzes kann dieses Problem jedoch behoben werden: Um zusätzliche Stabilität der Detektion des Markers zu erreichen, wird dieser nicht im gesamten Kamerabild, sondern in einer rechteckigen Umgebung um die von der OpenNI-Skelettrekonstruktion ermittelten Kopfposition gesucht. Die Größe der Suchregion wird aus der Bewegungsrichtung und Geschwindigkeit des Markers zwischen den letzten beiden Bildern berechnet: Auf eine festgelegte Mindest-

länge und -breite wird die letzte bekannte Geschwindigkeit des Markers in horizontaler und vertikaler Richtung aufaddiert, so dass sich die Suchregion bei Bewegungen vergrößert.

Wie in Abbildung 2 (Mitte) zu erkennen ist, sind die Werte der Tiefenkarte im Bereich des Markers undefiniert. Dies ist der Tatsache geschuldet, dass der Marker eine Störung des strukturierten Lichts verursacht. Um die Position des Markers im Raum zu bestimmen, wird zunächst das arithmetische Mittel der Werte gebildet, welche sich in der Tiefenkarte in einer  $11 \times 11$  Umgebung des zuvor ermittelten Pixels befinden. Die endgültige Position wird dann aus dem so errechneten Tiefenwert und den subpixelgenauen Bildkoordinaten berechnet.

## 4 Evaluation

Zur Evaluation der verschiedenen Verfahren wurden sowohl Positionsdaten eines still stehenden als auch eines sich bewegenden Benutzers aufgenommen. Erstere dienen dazu, die Präzision der verschiedenen Verfahren zu ermitteln, wohingegen die Aufnahmen unter Bewegung die Leistungsfähigkeit im Anwendungsfall zeigen sollen. Die Ergebnisse finden sich in den Abbildungen 4 und 5.

### 4.1 Aufbau und Methoden

Um eine möglichst stabile Position des Markers sowie des Kopfes für die statische Kondition zu erhalten, wurde eine Schaufensterpuppe verwendet, an deren Stirn ein Marker des World-Viz PPT Trackingsystems angebracht wurde (siehe Abb. 4d). Für die Aufnahme dynamischer Bewegungsdaten wurde eine Versuchsperson auf die gleiche Weise ausgestattet.

Die statischen Daten wurden an drei Positionen im Raum aufgenommen, jede der Aufnahmen fand über einen Zeitraum von 30 Sekunden statt. Des Weiteren wurden Daten für fünf Bewegungsmuster aufgenommen: Bewegungen entlang der 3 Raumachsen sowie eine kreisende Bewegung und eine freie Bewegung im Raum. Pro Verfahren und Aufnahme wurden die ermittelten Positionswerte sowie der Zeitstempel des Aufnahmezeitpunkts und des Speicherzeitpunkts festgehalten.

Der Versuchsaufbau ist in Abbildung 3 skizziert: Die vier Kameras des Trackingsystems wurden in einem Halbkreis vor dem Benutzer aufgestellt, um den Marker auf dessen Stirn aufnehmen zu können. Die Kinect wurde so positioniert, dass der Benutzer während der Datenakquise in Frontalansicht aufgenommen wurde. Während der Bewegung wurde darauf geachtet, dass der Kopf in Richtung der Kinect ausgerichtet blieb, wie es bei der Verwendung von 3D-Displays, Multi-Touch-Tables und ähnlichen, im Consumer Bereich eingesetzten Anzeigegeräten der Fall ist.

Da die Koordinatensysteme der Kinect und des professionellen Systems nicht übereinstimmen, und die entsprechende Koordinatentransformation aufgrund der Umfassung der Kinect nicht ausreichend genau bestimmt werden konnte, wurde diese mit Hilfe der aufgenommenen Daten berechnet. Zu diesem Zweck wurden die zeitgleich aufgenommenen Punkte ermittelt und die entsprechende Transformation errechnet (cf. [BM92]).

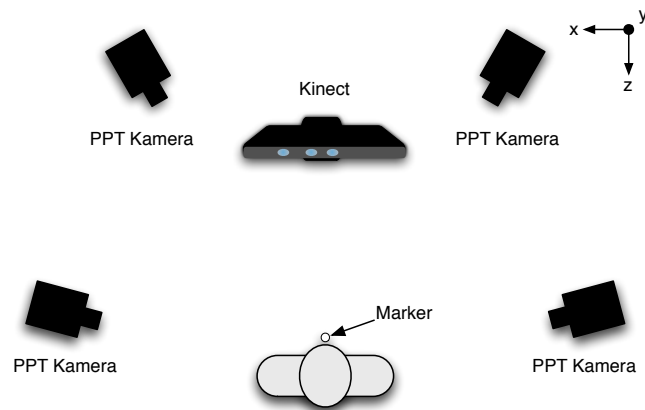


Abbildung 3: Schematische Darstellung des Versuchsaufbaus mit WorldViz PPT X4, Microsoft Kinect und IR-Marker an der Benutzerposition.

## 4.2 Ergebnisse

In der statischen Kondition haben wir einen signifikanten Effekt für die drei untersuchten Methoden auf die euklidische Distanz zwischen zwei zeitlich aufeinanderfolgenden Positionswerten (als Maß für den Jitter) gefunden (ANOVA,  $F(2,8346)=2322.52$ ,  $p<.001$ ). Post-hoc Vergleiche mit dem Tukey HSD Test zeigen für ein  $p<.05$  Niveau, dass sich die mittlere Distanz (in Metern) sowohl für PPT ( $M=0.0002$ ,  $SD=0.0001$ ), als auch für das Skelett-Tracking ( $M=0.0034$ ,  $SD=0.0032$ ) sowie für das hybride Tracking ( $M=0.0008$ ,  $SD=0.0007$ ) jeweils signifikant von den jeweils anderen Methoden unterscheidet.

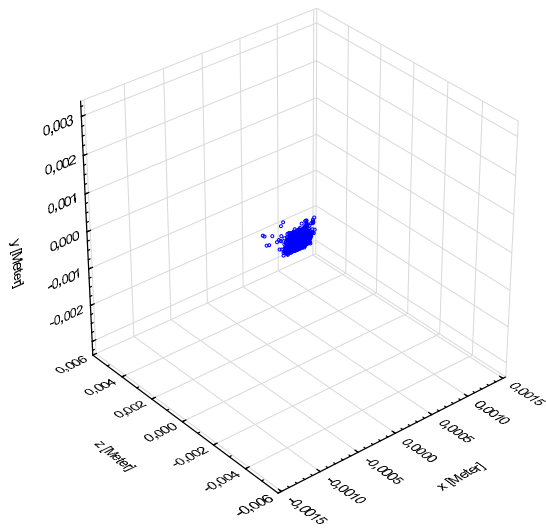
In der statischen Kondition haben wir ebenfalls einen signifikanten Unterschied der drei Methoden hinsichtlich der Streuung der Positionswerte um ihren Mittelpunkt (als Maß für die Präzision) gefunden (ANOVA,  $F(2,11144)=5952.09$ ,  $p<.001$ ). Post-hoc Vergleiche mit dem Tukey HSD Test zeigen für ein  $p<.05$  Niveau, dass sich die mittlere Distanz der Streuung (in Metern) sowohl für PPT ( $M=0.0003$ ,  $SD=0.0003$ ), als auch für das Skelett-Tracking ( $M=0.0033$ ,  $SD=0.0022$ ) sowie für das hybride Tracking ( $M=0.0011$ ,  $SD=0.0008$ ) jeweils signifikant von den übrigen Methoden unterscheidet.

Diese Ergebnisse sowie die Daten der dynamischen Testreihen (siehe Abb. 5) werden im Folgenden in Bezug auf die anfänglich aufgestellten Hypothesen diskutiert.

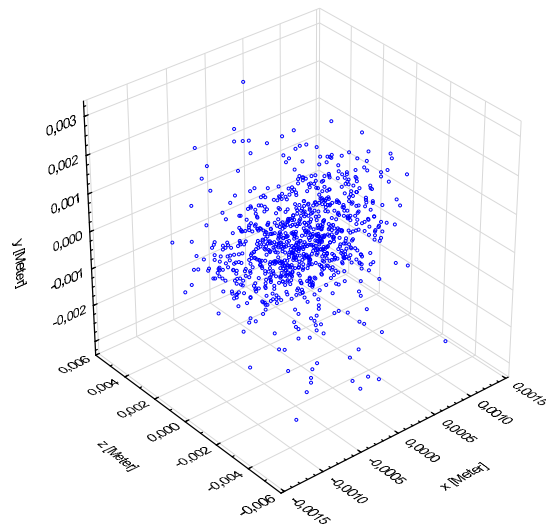
### 4.2.1 H1: Das PPT-System hat insgesamt die höchste Präzision

Im statischen Fall (siehe Abb. 4) zeigt das PPT-System im Vergleich zur Skelett-basierten Variante eine signifikant (im Mittel ca. um den Faktor 10) geringere Streuung, womit H1 bestätigt wird. Gegenüber dem hybriden Verfahren fällt der Unterschied weniger stark aus (im Mittel ungefähr um den Faktor 3). Das hybride Verfahren streut verstärkt in der Tiefe und bildet Liniencluster fast senkrecht zur XY-Ebene (siehe Abschnitt 4.2.3). Für die beiden dynamischen Fälle zeigt Abbildung 5 präzise Trajektorien des PPT-Systems im Vergleich zu den anderen Trackingverfahren.

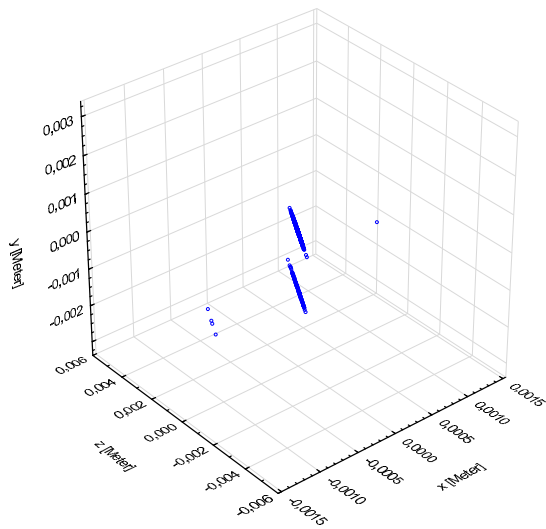




(a) PPT X4



(b) Skelett-basiert



(c) Hybrid



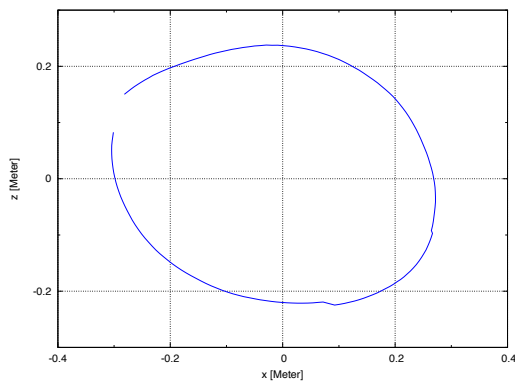
(d) Testobjekt

Abbildung 4: Um ihren Mittelpunkt zentrierte Messwerte von (a) PPT X4, (b) Skelett-basiertem Tracking und (c) dem hybriden Verfahren für eine statische Kopfposition sowie (d) die zur Erhebung der Daten eingesetzte Schaufensterpuppe.

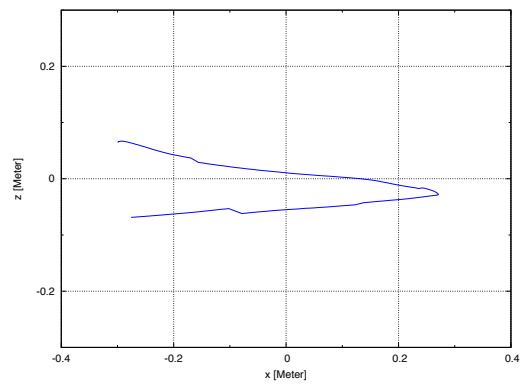
#### 4.2.2 H2: Das Skelett-basierte Verfahren hat die niedrigste Präzision

Im statischen wie im dynamischen Fall zeigen Abbildungen 4 und 5 die deutliche Streuung der Daten des Skelett-basierten Ansatzes. Insgesamt bestätigen die Resultate die im Vergleich geringste Präzision dieses Trackingverfahrens ( $M=0.0033$ ,  $SD=0.0022$ ). Als Erklärung für die starke Streuung kommt in erster Linie das verwendete Skelett-Fitting in Frage. Diese Funktion ist nicht trivial und kann an mehreren Stellen zu Mehrdeutigkeiten und damit zu Sprüngen ( $M=0.0034$ ,  $SD=0.0032$ ) in konsekutiven Messungen führen.

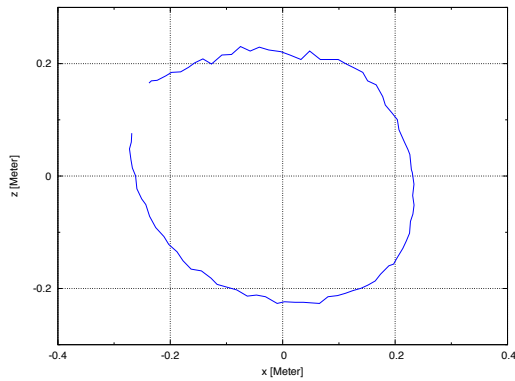




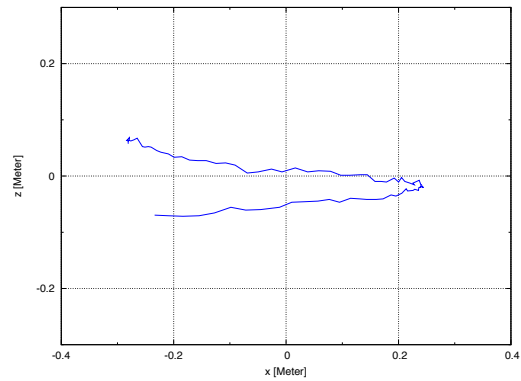
(a) PPT X4 - Kreis



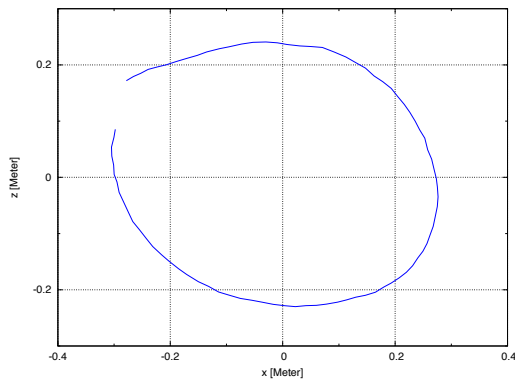
(b) PPT X4 - Seitwärts



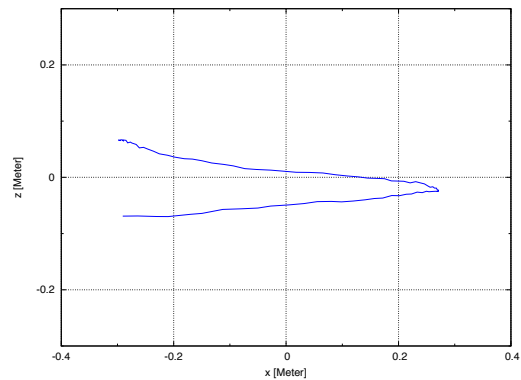
(c) Skelett-basiert - Kreis



(d) Skelett-basiert - Seitwärts



(e) Hybrid - Kreis



(f) Hybrid - Seitwärts

Abbildung 5: Um ihren Mittelpunkt zentrierte Messwerte für zwei unterschiedliche Bewegungen eines Benutzers. Eine Kreisbewegung getrackt mit (a) PPT X4, (c) Skelett-basiertem Tracking und (e) dem hybriden Verfahren, sowie eine seitliche Schwingung getrackt mit (b) PPT X4, (d) Skelett-basiertem Tracking und (f) dem hybriden Verfahren.

### 4.2.3 H3: Ein hybrides Verfahren erhöht die Präzision des Skelett-basierten Verfahrens

Für den statischen Fall weist das hybride Verfahren eine signifikant geringere Streuung als das Skelett-basierte Verfahren auf (im Mittel ca. um den Faktor 3), womit H3 bestätigt wird. Abbildung 4c zeigt Liniencluster fast senkrecht zur XY-Ebene. Dieses Resultat ist mit

hoher Wahrscheinlichkeit auf das eingesetzte Region-basierte Suchverfahren des Markermitelpunktes im Pixelraster des IR-Bildes der Kinect zurückzuführen, welches nur zwischen einer endlichen Anzahl von Positionen diskretisieren kann und durch die Auflösung der IR-Kamera limitiert ist.

Für den dynamischen Fall zeigt das hybride Verfahren dem PPT-System sehr ähnliche Ergebnisse (siehe Abb. 4). Die Ergebnisse sind ein starker Indikator dafür, dass die Markerbasierte Erweiterung zu einer erhöhten Präzision in den Trackingdaten führt und somit die Anwendbarkeit der Kinect für Headtracking verbessern kann.

## 5 Zusammenfassung

In der vorgestellten Arbeit wurden zwei Verfahren (markerlos und markerbasiert) zur Bestimmung der Kopfposition eines Anwenders auf Basis der Microsoft Kinect dargestellt und mit einem kommerziellen optischen, markerbasierten Trackingsystem verglichen.

Zum einen bietet die Verwendung des Skelett-basierten Trackingansatzes der Kinect zwar eine geringe Präzision und resultiert in vergleichbar starken Sprüngen in den Daten, stellt aber nativ keine Anforderungen an die Instrumentierung des Benutzers. Durch den geringen Preis der Kinect ist das System ubiquitär einsetzbar. Zum anderen konnten wir zeigen, dass sich die Vorteile eines markerbasierten optischen Trackingsystems und des markerlosen Trackingansatzes der Kinect kombinieren lassen, indem ein zusätzlicher IR-Marker für das Headtracking eingesetzt wird. Dieser Marker lässt sich an Shutterbrillen oder Head-Mounted Displays anfügen, so dass keine zusätzliche Instrumentierung des Benutzers notwendig wird. Die notwendige Größe und Leuchtintensität des Markers sind abhängig davon, welchen maximalen Abstand zwischen Benutzer und Kinect die jeweilige Anwendung erfordert. Prinzipiell lässt sich hierfür jede Infrarotlichtquelle verwenden. In Szenarien in denen der Abstand zur Kinect gering ist könnte auch auf einen retroreflektiven Marker zurückgegriffen werden.

Wie erwartet zeigen die Ergebnisse die höchste Präzision in den Daten des professionellen Trackingsystems. Im Vergleich der Ergebnisse des Skelett-basierten Ansatzes und unserer markerbasierten Erweiterung stellt sich heraus, dass letztere diskreten Jitter in der Bildebene aufweist. Dieser Effekt ist dem Verfahren inhärent, da es auf den Pixeln der Bildebene operiert. Unabhängig von der Größe des detektierten Markers lässt sich dies somit auch bei einer Subpixel-genauen Ermittlung der Markerposition nicht vollständig vermeiden. Ungeachtet dessen ist die gemessene Präzision des Verfahrens höher als die des Skelett-basierten Verfahrens. Somit bietet es eine sinnvolle Erweiterung zur Bestimmung der Kopfposition.

Das Verfahren lässt sich für beliebige Objekte, welche die Größe des verwendeten Markers überschreiten, anwenden, so dass es sich nicht nur für kostengünstiges Headtracking einsetzen lässt, sondern auch Perspektiven für verschiedene Interaktionstechniken bietet. Durch die Verwendung mehrerer Marker für das gleiche Objekt ließe sich darüber hinaus die Genauigkeit des Verfahrens voraussichtlich erhöhen, bzw. ebenfalls die Orientierung beim Headtracking bestimmen. Nachdem hier die Präzision der Verfahren untersucht wurde, bleibt der Vergleich von Latenz und Genauigkeit ein Thema für weitere Arbeiten.

## Literatur

- [Atr12] ATRACSYS ADVANCED TRACKING SYSTEMS: *accuTrack*. [http://atracsys.com/\\_products/accuTrack.php](http://atracsys.com/_products/accuTrack.php), 2012. [Online; last accessed 14-June-2012].
- [BM92] BESL, P.J. und N.D. MCKAY: *A Method for Registration of 3-D Shapes*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 14:239–256, 1992.
- [Bra00] BRADSKI, G.: *The OpenCV Library*. Dr. Dobb’s Journal of Software Tools, 2000.
- [FMO12] F. MORA, K. ALBERTO und J.-M. ODOBEZ: *Gaze Estimation from Multimodal Kinect Data*. In: *CVPR 2012 Workshop on Gesture Recognition*, 2012.
- [FWGVG11] FANELLI, G., T. WEISE, J. GALL und L. VAN GOOL: *Real time head pose estimation from consumer depth cameras*. Pattern Recognition, Seiten 101–110, 2011.
- [GP11] GARSTKA, J. und G. PETERS: *View-dependent 3D Projection using Depth-Image-based Head Tracking*. In: *8th IEEE International Workshop on Projector-Camera Systems PROCAMS*, Seiten 52–57, 2011.
- [KE12] KHOSHELHAM, K. und S. O. ELBERINK: *Accuracy and Resolution of Kinect Depth Data for Indoor Mapping Applications*. Sensors, 12:1437–1454, 2012.
- [Lee08] LEE, J.C.: *Hacking the nintendo wii remote*. Pervasive Computing, 7(3):39–45, 2008.
- [MCT09] MURPHY-CHUTORIAN, E. und M.M. TRIVEDI: *Head pose estimation in computer vision: A survey*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 31(4):607–626, 2009.
- [MG01] MOESLUND, T.B. und E. GRANUM: *A survey of computer vision-based human motion capture*. Computer Vision and Image Understanding, 81(3):231–268, 2001.
- [Mic12] MICROSOFT: *Kinect for Windows*. <http://www.microsoft.com/en-us/kinectforwindows/>, 2012. [Online; last accessed 14-June-2012].
- [NJ06] NAWROT, M. und L. JOYCE: *The pursuit theory of motion parallax*. Vision research, 46(28):4709–4725, 2006.
- [PDS<sup>+</sup>09] PERREIRA DA SILVA, M. et al.: *Fast, low resource, head detection and tracking for interactive applications*. PsychNology Journal, 7(3):243–264, 2009.
- [Pri12] PRIMESENSE: *OpenNI*. <http://www.openni.org>, 2012. [Online; last accessed 14-June-2012].

- [RG79] ROGERS, B. und M. GRAHAM: *Motion parallax as an independent cue for depth perception*. Perception, 8(2):125–134, 1979.
- [SBWS12] SATYAVOLU, S., G. BRUDER, P. WILLEMSSEN und F. STEINICKE: *Analysis of IR-based virtual reality tracking using multiple Kinects*. In: *Proceedings of IEEE Virtual Reality*, Seiten 149–150, 2012.
- [See12] SEEING MACHINES: *faceAPI - the real-time face tracking toolkit for developers and oems*. <http://www.seeingmachines.com/pdfs/brochures/faceAPI-techspecs.pdf>, 2012. [Online; last accessed 14-June-2012].
- [SJP11] SMISEK, J., M. JANCOSEK und T. PAJDLA: *3D with Kinect*. Technischer Bericht, Czech Technical University Prague, 2011.
- [SLR<sup>+</sup>11] SUMA, E.A., B. LANGE, A. RIZZO, D.M. KRUM und M. BOLAS: *Faast: The flexible action and articulated skeleton toolkit*. In: *IEEE Virtual Reality Conference*, Seiten 247–248, 2011.
- [SS03] SCHARSTEIN, D. und R. SZELISKI: *High-Accuracy Stereo Depth Maps Using Structured Light*. In: *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR)*, Seiten 195–202, 2003.
- [TFCRS11] THOMPSON, W., R. FLEMING, S. CREEM-REGEHR und J. K. STEFANUCCI: *Visual Perception from a Computer Graphics Perspective*. A. K. Peters / CRC Press, 2011.
- [THS<sup>+</sup>01] TAYLOR, R. M., T. C. HUDSON, A. SEEGER, H. WEBER, J. JULIANO und A. T. HELSER: *VRPN: A Device-Independent, Network-Transparent VR Peripheral System*. In: *Proceedings of the ACM Symposium on Virtual Reality Software and Technology*, Seiten 55–61, 2001.
- [WBV<sup>+</sup>99] WELCH, G., G. BISHOP, R. VICCI, S. BRUMBACK und K. KELLER: *The HiBall Tracker: High-Performance Wide-Area Tracking for Virtual and Augmented Environments*. In: *Proceedings of the ACM Symposium on Virtual Reality Software and Technology*, Seiten 1–10, 1999.
- [WF02] WELCH, G. und E. FOXLIN: *Motion tracking: No silver bullet, but a respectable arsenal*. Computer Graphics and Applications, IEEE, 22(6):24–38, 2002.
- [Wor12] WORLDVIZ: *PPT Precision Position Tracker*. <http://www.worldviz.com/products/ppt/>, 2012. [Online; last accessed 14-June-2012].