

# **Ein Navigation Interface basierend auf der Walking-In-Place-Metapher für ein CAVE Automatic Virtual Environment unter Verwendung von Tiefensensor-Kameras**

Sebastian Brandt, Thomas Jung

HTW Berlin,  
Wilhelminenhofstraße 75A , D-12459 Berlin  
eMail: t.jung@htw-berlin.de  
URL: <http://www.htw-berlin.de>

**Zusammenfassung.** Das Cave Automatic Virtual Environment (CAVE) der HTW Berlin wird im Unterschied zu anderen CAVE-Systemen durch ein Natural User Interface (NUI) gesteuert. Tiefensensor-Kameras ermöglichen dabei die Erkennung des Skeletts eines Benutzers. Im Rahmen dieser Arbeit wird ein Navigation Interface präsentiert, das auf der Walking-In-Place (WIP)-Metapher beruht. Dabei muss zwischen realen Laufbewegungen, mit denen sich der Benutzer in der aktuell dargestellten Umgebung bewegt, und Walking-in-Place, mit dessen Hilfe er sich in der virtuellen Welt fortbewegen möchte, unterschieden werden.

## **1 Einleitung**

CAVE-Systeme [1] werden schon seit den 1990er Jahren zur Erzeugung immersiver Virtueller Umgebungen verwendet, die Interaktion erfolgte dabei zunächst über magnetische, später über optische Tracking-Systeme, bei denen Benutzer in der Regel jedoch mit aktiven oder passiven Markern instrumentiert werden. Natural User Interfaces vermeiden weitestgehend sichtbare Bedienelemente, um die Bedienung dadurch natürlicher zu gestalten. Tiefensensor-Kameras, die die Entfernungen zu Bildpunkten bestimmen, ermöglichen dabei die markerlose Erkennung von Bewegungen. Durch die Entwicklung kostengünstiger Geräte (Microsoft Kinect, Asus Wavi Xtion) gewinnt die Entwicklung von NUIs an Bedeutung.

Navigationsinterfaces in virtuellen Umgebungen müssen es dem Benutzer ermöglichen, sich durch eine virtuelle Welt zu bewegen, die viel größer als der real zur Verfügung stehende Interaktionsraum ist. Dieser beträgt in der CAVE der HTW Berlin z. B. nur drei mal drei Meter. Bisher erfolgt die Navigation hier durch ein Steuergerät (Wiimote der Firma Nintendo), wobei vier Tasten für die Drehung nach links und rechts bzw. für die Vorwärts- und Rückwärtsbewegung in Blickrichtung vorgesehen sind.

Während bereits WIP-Navigationsinterfaces für Darstellungen mit Hilfe von Head-Mounted-Displays existieren soll im Folgenden erstmals untersucht werden, wie mit

Hilfe von kostengünstigen Tiefenkameras in einer CAVE ein WIP-Navigationsinterface realisiert werden kann.

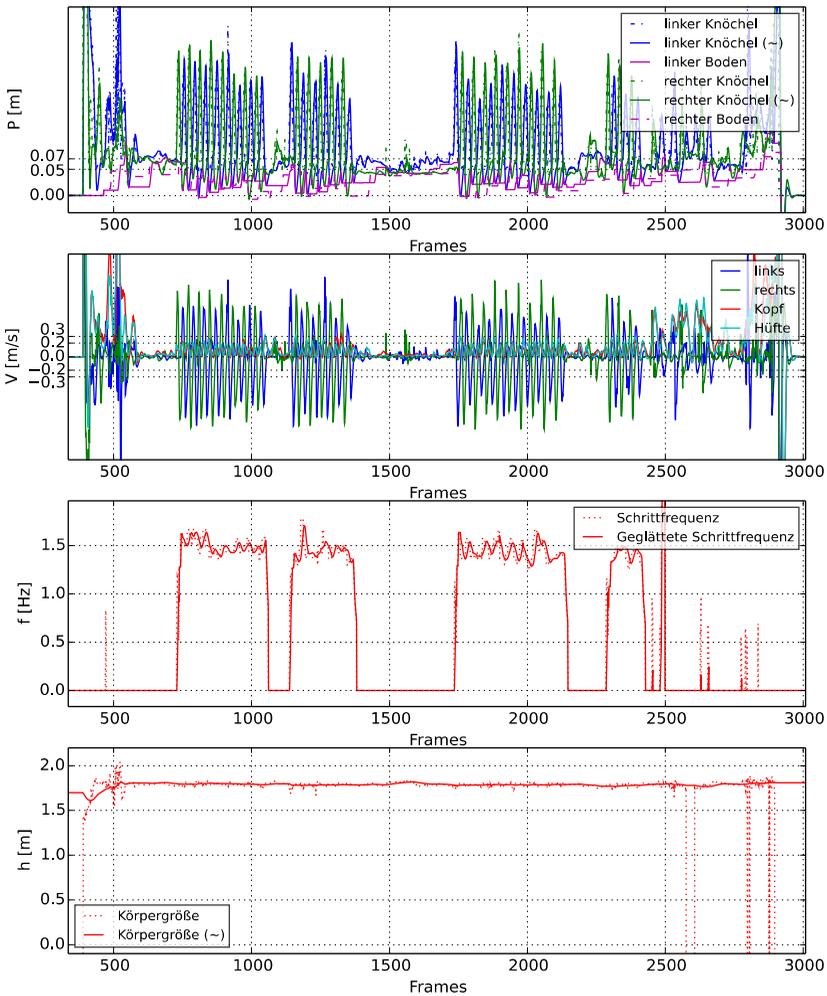
## **2 Navigation unter Verwendung der Walking-In-Place-Metapher**

Slater, Usoh und Steed haben als erste ein Navigation Interface für die Fortbewegung in virtuellen Welten basierend auf der WIP-Metapher präsentiert. Sie konnten zeigen, dass der Grad der Präsenz gegenüber der Navigation durch ein Steuergerät deutlich erhöht werden kann. Sie verwendeten ein Head-Mounted-Display (HMD) mit Headtracking-System, wobei sie die Positionsveränderung des Kopfes zwischen zwei aufeinanderfolgenden Frames bestimmten und ein neuronales Netz auf die Erkennung der Schrittbewegungen trainierten [2]. Die Erkennung der Laufbewegung erfolgte mit einer Erfolgsrate von ca. 90%, unterschiedliche Laufgeschwindigkeiten konnten dabei nicht bestimmt werden.

Templeman, Denbrook und Sibert verwendeten 1999 Sensoren an Knien mit sechs Freiheitsgraden sowie Kraftmesser in Schuhen, wodurch sogar weitere Gesten wie Seitwärtsschritte möglich waren. Problematisch waren hier u. a. die hohe Anfangslatenz und die hohe Anfälligkeit für zeit- und räumlich nicht exakte Messwerte [3]. Yan und Rushton bestimmten 2004 anhand von Nulldurchgängen und Extrema der vertikalen Geschwindigkeiten der Beine bereits Fortbewegungsgeschwindigkeiten. Dabei wurde ein relativ präzises akustisches Trackingsystem verwendet [4]. Feasel, Whitton und Wendt verwendeten 2008 die Summe der vertikalen Geschwindigkeiten beider Fersen, wodurch sie die Latenz deutlich verringern konnten. Die resultierenden Bewegungen waren jedoch nicht sehr natürlich, sondern eher ruckartig [5]. Wendt, Whitton und Brooks entwickelten 2010 schließlich ein System, das auf der Unterscheidung von drei Bewegungsphasen während der menschlichen Laufbewegung beruht. Basierend auf biomechanischen Untersuchungen und unter Verwendung eines Zustandsautomaten können Sie in jeder Phase anhand mehrerer Parameter die Laufgeschwindigkeit bestimmen. Zur Fortbewegung reicht es, bereits eine einzelne Phase zu erkennen, wodurch die Latenz insgesamt sehr gering gehalten werden kann. Die Darstellung der virtuellen Umgebung erfolgte über ein HMD, beide Schienbeine des Benutzers wurden über ein optisches Tracking-System getrackt [6].

## **3 Walking-In-Place-Navigation in der CAVE der HTW Berlin**

Alle bisher beschriebenen WIP-Interfaces basieren auf relativ präzisen Trackingsystemen, die jedoch nur wenige Freiheitsgrade liefern. Die Besonderheit beim Ansatz der CAVE der HTW besteht darin, dass die Kinect-Kameras ein komplettes Skelett des Benutzers liefern, die Daten jedoch rauschbehaftet oder durch Glättung relativ ungenau sind. Im Folgenden wird untersucht, wie unter diesen Bedingungen ein WIP-Interface realisiert werden kann.



**Abb. 1:** Erkennung des Walking-In-Place mit Höhen der Fußknöchel, abgeleiteten vertikalen Geschwindigkeiten, erkannter Schrittfrequenz und erkannter Körpergröße. Die Schrittgeschwindigkeit leitet sich direkt aus letzteren beiden ab.

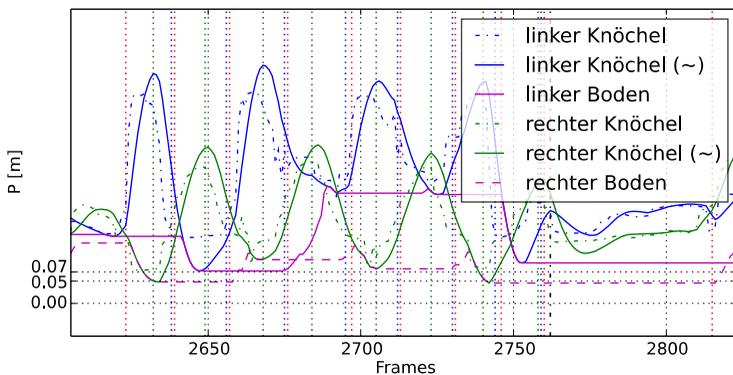
Das hier beschriebene System basiert grundsätzlich auf dem Ansatz von Wendt, Whitton und Brooks [8]. Die Ausgabe des Algorithmus besteht aus der zum Zeitpunkt des Frames momentanen positiven Schrittgeschwindigkeit des Benutzers. Von der Erkennung des Rückwärtslaufens wird abgesehen, da dies einerseits eine eher aus Computerspielen bekannte Geste ist und in der realen Bewegung nur selten zur Navigation verwendet wird, andererseits stellen sich die durchaus komplexen Fragen, wie diese Geste *natürlich* signalisiert werden soll und welchen Einfluss das Rückwärtsgehen auf die Entstehung von Motion Sickness haben würde.

Während Wendt, Whitton und Brooks das Trackingsystem am Schienbein des Benutzers befestigen, kommen bei der Verwendung der Kinect-Kamera mehrere Gelenkpunkte der Beine in Frage. Die Analyse der Skelettdaten zeigt, dass mit dem im Kinect-SDK implementierten Algorithmus die Fußspitzen nur relativ ungenau getrackt werden. Die Genauigkeit zwischen Knie und Fußgelenk scheint ähnlich zu sein, die Bewegung des Fußgelenks weist für die menschliche Laufbewegung jedoch eine größere Signifikanz auf.

Die Skelettdaten des Fußgelenks werden anschließend mit Holt-Winter Double Exponential Smoothing ( $\alpha=0,2$ ,  $\beta=0,2$ ) geglättet, die Höhe wird dann für die Schritterkennung verwendet. In Abbildung 1 ist zu erkennen, dass bei gleichmäßigem Laufen am Platz eine relativ gleichmäßige Schrittfrequenz erkannt wird.

Eine wichtige Eigenschaft eines WIP-Interfaces ist die Unterscheidung zwischen WIP und realem Laufen. Dem Benutzer muss ermöglicht werden, sich innerhalb des Interaktionsraums der CAVE zu bewegen, ohne das WIP auszulösen.

Der WIP-Modus wird anhand mehrerer Merkmale erkannt. WIP-Zyklen erfordern eine minimale Knöchelhöhe und vertikale Geschwindigkeit, so dass man mit „flachen“ Laufbewegungen den WIP-Modus vermeiden kann. WIP wird weiterhin abgebrochen oder nicht begonnen, wenn die horizontale Geschwindigkeit des Beckens während des ersten Schrittes einen Schwellwert von 0,4m/s überschreitet.

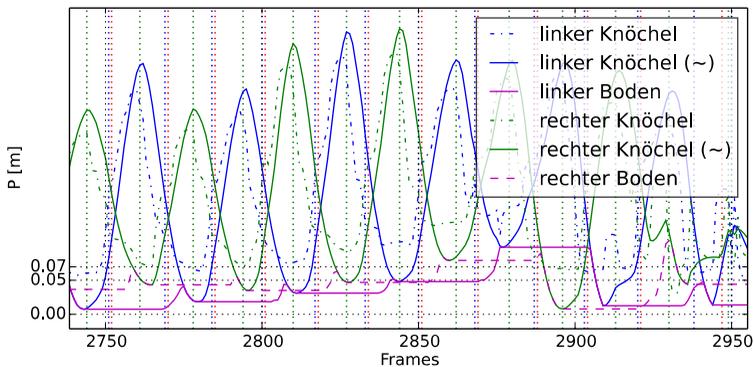


**Abb. 2:** Verdeckung des linken Fußes durch den rechten

Beobachtungen zeigen, dass selbst das kurze Auslösen von WIP durch eine zu geringe Hüftgeschwindigkeit am Anfang des Schrittes vom Benutzer nicht als störend wahrgenommen wird. Vermutlich ist dies auf „Schwächen“ im menschlichen Wahrnehmungsapparat zurückzuführen, welche Razzaque bereits für die Kaschierung seiner Drehung der virtuellen Welt verwendete.

Während sich das Positionsrauschen des Skelettes mit dem verwendeten Algorithmus relativ gut heraus rechnen lässt, stellen falsch-positiv erkannte Positionen ein größeres Problem dar. So ist es möglich, dass ein teilweise verdeckter Fuß beim Eintauchen in den "Schatten" des verdeckenden Fußes als schwebend erkannt wird (siehe Abbildung 2) wodurch die Lauferkennung unnötig abbricht. Um toleranter gegenüber solchen und anderen Fehlinterpretationen zu sein, wird ständig der Status des jeweils anderen Fußes überwacht. So kann ein "eindeutiges Anheben des anderen Fußes" ( $1,5 \cdot$  Anhebe-Schwellwert) einen Schritt auch ohne erkanntes Aufsetzen beenden, womit die Kontinuität in der Lauferkennung gewahrt bleibt. Als positiver Nebeneffekt werden weitere Situationen mit Fehlerkennungen durch aus anderen Gründen deformierten Kurven verbessert.

Ein weiteres Problem ergibt sich aus von Schritt zu Schritt an unterschiedlichen Positionen erkannten Gelenkpunkten. So kann es sein, dass sich die Höhe des Fußknöchels beim Laufen auf exakt derselben Stelle ständig verändert (siehe Abbildung 3). Wichtig für die Lauferkennung sind jedoch nur die relative Höhe über dem Boden bzw. tiefstem Punkt sowie die vertikale Geschwindigkeit, sodass die Berechnung eines relativen Bodenniveaus ausreicht. Dieses wird durch den tiefsten Punkt des letzten Schrittes abhängig von der erkannten Schrittfrequenz festgelegt.



**Abb. 3:** Veränderliches Bodenniveau

Des Weiteren wurde untersucht, ob die Verwendung des vom Kinect SDK gelieferten Status eines jeden Gelenks, welcher *erkannt*, *vermutet* und *nicht erkannt* sein kann, zur weiteren kontextsensitiven Filterung von Fehlern hinzuziehbar ist. Es hat sich jedoch herausgestellt, dass diese Angaben hierfür zu ungenau sind und teilweise ein zu häufiges Fehlverhalten zur Folge hätten. Einzig der erkannte Status des Körpers wird genutzt, um das Eintreten des Benutzers in das Sichtfeld der Kinect zu

erkennen und unsinnig erkannte Geschwindigkeiten während dieser kurzen Phase zu verhindern.

Wendts Algorithmus ermöglichte das Einbeziehen der Körpergröße des Benutzers, um eine natürlichere Laufgeschwindigkeit zu berechnen. Die Berechnung dieses Maßes ist mit den Skelettdaten der Kinect grundsätzlich möglich, erfordert jedoch einige Anpassungen. Die Länge der Vektorpfade vom Kopf über das Zentrum der Schultern bis zum Zentrum der Hüfte, sowie die Länge der Beine von den Seiten der Hüfte über die Knie bis zu den Knöcheln können einfach verwendet werden. „Fehlende“ Höhen, das heißt vom Kopfmittelpunkt bis zum höchsten Punkt, von der Hüftmitte bis zu den äußeren Hüftpunkten und vom Knöchel bis zum Boden können durch Verdoppeln der Vektorbeträge Kopf-Schulter und Hüfte-seitliche Hüfte, sowie dem Addieren von 5cm als geschätzte Fußknöchelhöhe ausgeglichen werden. Da dieser Wert mit einem hohen Rauschen versehen ist, sollte er mit einem einfachen trägen Exponentialfilter ( $\alpha=0,01$ ) geglättet werden.

#### **4 Steuerung der Bewegungsrichtung**

Die bisher bekannten Ansätze verwendeten jeweils ein besonderes Tracking-System, das die Instrumentierung des Benutzers erfordert. Durch die Verwendung eines HMDs konnte sich der Benutzer jeweils frei in der virtuellen Welt drehen. Die CAVE der HTW verfügt jedoch nur über drei Seitenwände, so dass die realen Blickrichtungen des Benutzers begrenzt werden müssen. Einen grundsätzlichen Lösungsansatz stellt in diesem Zusammenhang Redirected WIP [9] dar. Razaque und Kollegen testeten ein solches Interface in Ihrer CAVE unter Verwendung eines Headtrackers und eines Trackers am Körper des Benutzers, um dessen horizontale Ausrichtung zu bestimmen. Die Bewegungsrichtung des Benutzers in der virtuellen Welt wurde dabei während der Bewegung leicht gedreht, um ihn davon abzuhalten, auf die nicht existente Rückwand zu schauen. Razaque und Kollegen räumen ein, dass die Anzahl der Situationen, in denen der Benutzer nach hinten schaut, entgegen Ihrer Erwartung in Ihrem System leider nicht signifikant verringert werden konnte. Sie stellten darüber hinaus fest, dass der Grad der Präsenz durch das bewusste Wahrnehmen der Rotation gemindert wurde [7].

Im System von Razaque und Kollegen wurde die Ausrichtung des Benutzers in der virtuellen Welt proportional zum Sinus des doppelten Winkels zwischen Torsoorientierung des Benutzers und Frontseite der CAVE verändert. Durch die relativ große Steigung der Sinusfunktion im Nulldurchgang ist es für den Benutzer kaum möglich, seine derzeitige Ausrichtung beizubehalten. Vermutlich ist dies eine Hauptursache des relativ geringen Erfolgs des beschriebenen Ansatzes.

Hier wird deshalb als Funktion  $(2x)^3$  verwendet, so dass dem Benutzer aufgrund der geringen Steigung im Nulldurchgang auch bei leichter Änderung der Torsoorientierung während der Interaktion mit einer Anwendung eine bessere Kontrolle seiner Ausrichtung in der virtuellen Welt ermöglicht wird. Des Weiteren wird der maximale Winkel der Torsoorientierung durch die verwendeten Kinect-Sensoren im günstigsten Fall auf ca.  $\pm 45^\circ$  beschränkt. Die höhere Steigung unserer Funktion bei größeren Argumenten erlaubt es dem Benutzer, die virtuelle Welt trotz dieser Einschränkung in allen Situationen ausreichend schnell zu drehen. Außerdem verhindert sie

durch die zunehmende Drehgeschwindigkeit wirksam, dass die hintere Wand der CAVE ins bewusste Sichtfeld des Benutzers rückt.

## 5 Benutzertests

Zum Abschluss des Projekts wurde ein Hallway-Test [8] durchgeführt. An diesem haben 15 Personen teilgenommen, davon drei weiblich und zwölf männlich. Der Altersdurchschnitt lag bei ungefähr 30 Jahren, bis auf eine Person hatte niemand Erfahrung mit einer CAVE. Die Tests wurden einzeln durchgeführt und beinhalteten eine beliebig lange Einführungsphase für die Person in die CAVE. Der eigentliche Test beinhaltete das Ausführen derselben Aufgabe einmal mit der bestehenden Steuerung per Wiimote und anschließend mit WIP. Als Szene diente hier ein Nachbau der Räumlichkeiten der HTW Berlin und als Aufgabe das Navigieren durch drei einfache Räume mit zwei definierten Wegpunkten.

Die Testpersonen füllten nach jedem Teilttest drei Fragebögen aus, deren Ergebnisse des Vor- und Nachtests verglichen wurden: AttrakDiff (<http://attrakdiff.de/>), der Nasa Task Load Index (TLX) [9] und ein Präsenzfragebogen [10]. In letzterem verbesserte sich der Durchschnittswert auf einer Skala von 1 bis 7 um 0,8 Punkte oder 18% bei einer Standardabweichung von 0,7. Der TLX nahm um 21% von 2,8 auf 3,4 im selben Skalenbereich zu, wobei hier die einzelnen Belastungsarten genauer betrachtet werden sollten. Die körperliche Anforderung stieg beträchtlich von 2,0 auf 3,8, während die allgemeine Zufriedenheit mit der Ausführung der Aufgabe jedoch stieg und somit eine Verringerung der Belastung um 0,4 Punkte bzw. 12% in dieser Kategorie zu beobachten war. AttrakDiff hat letztendlich gezeigt, dass dieselbe Anwendung mit unterschiedlichen Navigationsmetaphern mit gleichbleibender pragmatischer, jedoch im Falle des WIP mit einer deutlich höheren hedonischen Qualität und somit mit einer höheren Gesamtattraktivität der Applikation bewertet wurde. Die größten Verbesserungen zeigen die Bewertungen der Menschlichkeit, Originalität und Neuartigkeit.

Weiterhin konnte beobachtet werden, dass die Benutzer sich schnell an die höheren Drehgeschwindigkeiten der virtuellen Umgebung gewöhnt haben und die fehlende Rückwand der CAVE nicht mehr störend ins Blickfeld rückte.

## Literatur

- [1] Cruz-Neira C., D. J. Sandin and T. A. DeFanti, "Surround-Screen Projection-Based Virtual Reality: The Design and Implementation of the CAVE", Computer Graphics, SIGGRAPH Annual Conference Proceedings, 1993.
- [2] Slater, M., Usoh M., Steed A., 1995: "Taking Steps: The Influence of a Walking Technique on Presence in Virtual Reality," ACM Trans. on CHI, Special Issue on Virtual Reality Software and Technology, 2, 3: 201-219, September.
- [3] Templeman, J. N. ; Denbrook, P. S. ; Sibert, L. E.: "Virtual locomotion: Walking in place through virtual environments". Proc. Presence 8 (1999),Nr. 6, pp. 598–617
- [4] Yan, L ; Allison, R.S. ; Rushton, S.K. "New simple virtual walking methodwalking on the spot". Proceedings of the IPT Symposium 2004
- [5] Feasel J. ; Whitton, M. C.; Wendt, J.D.: "LLCM-WIP: Low-latency, continuous-motion walking-in-place". Proc. IEEE 3D User Interfaces 2008 pp. 97–104
- [6] Wendt, J. D. ; Whitton, M. C. ; Brooks, F. P.: GUD WIP: Gait-understanding-driven walking-in-place. Proc. IEEE Virtual Reality Conference 2010, pp. 51–58
- [7] Razaque, S.; Swapp, D. Slater, M.; Whitton, M.; Steed, A. : "Redirected walking in place",Proc. of the workshop on Virtual environments 2002, pp 123-130, Aire-la-Ville, Switzerland
- [8] Nielsen, J.: „Why You Only Need to Test with 5 Users“, <http://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/>, Abrufdatum: November 2013
- [9] Human Performance Research Group: NASA Task Load Index (NASA-TLX). Version 1.0. NASA Ames Research Center, Moffett Field. California: NASA, 1986
- [10] Slater, Mel, John McCarthy, and Francesco Maringelli. "The influence of body movement on subjective presence in virtual environments." Human Factors: The Journal of the Human Factors and Ergonomics Society 40.3 (1998): 469-477.