

Binokulare Blickverfolgung als Eingabe für stereoskopische VR/MR-Systeme

Michael Kleiber, Carsten Winkelholz

FGAN-Forschungsinstitut für Kommunikation,
Informationsverarbeitung und Ergonomie
Abteilung Mensch-Maschine-Systemtechnik
Neuenahrer Str. 20, 53343 Wachtberg
{kleiber, winkelholz}@fgan.de

Abstract: Am Beispiel des Geräts Tobii X-120 wird die Genauigkeit von binokularen Gaze-Tracking-Systemen analysiert. Der Fehler des 3D-Blickpunkts wird mittels Wahrscheinlichkeitsfunktionen modelliert. Eine Abschätzung bezüglich der Genauigkeit der 3D-Blickpunkterfassung ist für die Datenanalyse sowie bei der Konzeption von Interaktionstechniken wichtig. Im Beitrag wird eine Anwendung von binokularem Gaze-Tracking für ein Desktop-VR-System vorgestellt.

1 Einleitung

Blickverfolgung, auch Gaze- oder Eye-Tracking genannt, und die daraus resultierenden Blickbewegungsdaten stellen ein direktes Eingabemedium dar [Hin03]. Blickbewegungen erlauben eine direkte Interaktion zwischen Mensch und Maschine mittels minimaler motorischer Arbeit [SJ00]. Allen praktischen Anwendungen von Blickbewegungssystemen gemein ist, dass die eingesetzten Interaktionstechniken der Genauigkeit der Blickpunkterfassung angepasst sind und eine implizite Steuerung erlauben [Jac90].

Der Einsatz von Gaze-Tracking als Interaktionsform in VR-/MR-Systemen führt zu einem höheren Grad der Immersion, so dass der Nutzer das Gefühl hat die VR-Umgebung durch Gedanken zu steuern [JGL08]. Zur Nutzung von Eye-Tracking bei vollimmersiven VR-, MR- oder AR-Systemen muss entweder ein optisches Blickbewegungsmesssystem eingesetzt werden, welches direkt in das Head-Mounted-Display (HMD) integriert ist, oder aber die Blickbewegungen müssen über ein Elektrookulogramm erfasst werden. Besser als vollimmersive VR-Systeme eignen sich sogenannte Fish-Tank VR-Systeme zur Kombination mit kontaktlosen Eye-Trackern [WAB93]. Werden die VR-Systeme mit autostereoskopischen Monitoren betrieben, so gibt es hinsichtlich der Wahl des Eye-Trackers keine Einschränkungen. Die nahtlose Integration in den bestehenden 2D-Arbeitsplatz erlaubt außerdem einen schnellen Wechsel zwischen VR-System und 2D-Arbeitsplatz.

Moderne Eye-Tracking-Systeme erlauben das Erfassen von Blickpunkten für linkes und rechtes Auge. Bei stereoskopischen Anzeigen, welche ebenfalls eine 2D-Anzeigefläche besitzen, wird dem Betrachter dennoch ein räumlicher Eindruck durch zwei unterschied-

liche Bilder für das linke und rechte Auge vermittelt. Unterschiede zwischen rechtem und linkem Bild (Disparität) aufgrund der perspektivischen Projektion werden vom Betrachter als stereoskopischer Tiefenreiz wahrgenommen. Um ein virtuelles Objekt scharf zu sehen, muss der Betrachter linkes und rechtes Auge jeweils auf den Bereich der Anzeige richten, an dem das Objekt dargestellt ist. Aus den unterschiedlichen Blickpositionen kann der 3D-Blickpunkt berechnet werden. Dieser hat mehr Aussagekraft bezüglich der Intention des Nutzers, als der 2D-Blickpunkt.

2 Blickverfolgung

Auch wenn Blickverfolgungssysteme im Allgemeinen immer besser werden, so ist die Blickpunktbestimmung ungenau. Bei der Nutzung von Blickbewegungsdaten zur Analyse oder für die Interaktion ist es jedoch wichtig die Genauigkeit der Blickpunkterfassung zu kennen: Wird beispielsweise der Blickpunkt zur interaktiven Selektion genutzt, so kann über die Abschätzung des wahrscheinlichen Fehlerbereichs der kleinstmögliche Abstand zwischen Selektionsobjekten bestimmt werden.

Die Genauigkeit eines Eye-Tracking-Systems kann unter anderem durch folgende Parameter beschrieben werden: Auflösung, Drift und Bewegungsfehler. Aus der Summe der möglichen Fehler ergibt sich die systemspezifische Genauigkeit – beim Tobii X-120 wird diese mittels der mittleren absoluten Abweichung mit $E_{e_\alpha} = 0,5^\circ$ angegeben [AB09].

Zur Fehlerabwägung können Wahrscheinlichkeitsfunktionen benutzt werden. Das Datenblatt des Tobii X-120 enthält allerdings keine Angabe über die Verteilung der Fehler, so dass im Folgenden von einer Normalverteilung ausgegangen wird. Die radiale Abweichung (e_α) des Blickpunkts kann damit über die Verteilungsfunktion beschrieben werden. Die Standardabweichung lässt sich über die mittlere absolute Abweichung bestimmen: $\sigma_\alpha = E_{e_\alpha} \left(\frac{2}{\pi}\right)^{-\frac{1}{2}}$.

Zur Erhöhung der Genauigkeit kann das gleitende Mittel gebildet werden. Bei Mittelung von n Messwerten verringert sich die Standardabweichung der Messung um den Faktor $n^{-\frac{1}{2}}$. Eine Erhöhung der Anzahl der Blickpunkte führt zu einer Verbesserung der Genauigkeit. Dies ist jedoch nicht beliebig fortführbar. Der limitierende Faktor ist dabei der Mensch, welcher seine Augen fast kontinuierlich bewegt¹. Neben den willentlichen Augenbewegungen gibt es noch die unwillentlichen, welche sehr viel kleiner sind und daher auch als Mikrobewegungen bezeichnet werden. Dazu zählen Tremore und Drift, welche binokular unkorreliert stattfinden, so dass es zu gegensätzlichen und gleichgerichteten Bewegungen kommt. Diese Mikrobewegungen sind daher eine Quelle der Ungenauigkeiten des Blickbewegungsmesssystems.

Der limitierende Faktor bei der Mittelung von n Blickpunkten ist demnach die Frequenz der Sakkadensprünge (ca. 5 Hz). Für die Bestimmung des gleitenden Mittels bedeutet dies bezüglich des eingesetzten Eye-Trackers mit einer Samplingrate von 120 Hz, dass maximal 24 Werte genutzt werden können.

¹Die durchschnittliche maximale Sakkadenfrequenz beträgt ca. 5 Hz.

Bisher wurde die Genauigkeit der Blickpunktbestimmung für ein einzelnes Auge beschrieben. Bei monoskopischen Anwendungen kann das Mittel aus linkem und rechtem Blickpunkt bestimmt werden, so dass die Genauigkeit noch einmal gesteigert wird. Soll aber der 3D-Blickpunkt bestimmt werden, so ist eine Mittelung nicht möglich, da beide Freiheitsgrade dazu benötigt werden Distanz und Richtung zu bestimmen. Vor allem die Streuung in der Tiefe ist von Interesse. Diese jedoch ist abhängig von der Interokularstanz (pd) des Betrachters und von der mittleren Blickentfernung, welche sich über die Vergenzwinkel ($\alpha_{l_0}, \alpha_{r_0}$) der Augen ausdrücken lässt²:

$$z = \frac{pd}{\tan \alpha_r - \tan \alpha_l} \approx \frac{pd}{\alpha_r - \alpha_l} = \frac{pd}{\Delta\alpha} \text{ für } |\Delta\alpha| < 10^\circ$$

$$f(\Delta\alpha, \mu_{\Delta\alpha}, \sigma_{\Delta\alpha}) = \text{norm}(\Delta\alpha, \mu_{\Delta\alpha}, \sqrt{2}\sigma_\alpha) \quad (1)$$

Durch die Näherung³ $\tan \alpha \approx \alpha$ ist es möglich eine Aussage über die Standardabweichung des Konvergenzwinkels ($\Delta\alpha$) und damit eine Abschätzung über den Distanzfehler zu treffen. Folgende Tabelle zeigt die Näherung des Fehlers im Vergleich zu einer Monte-Carlo-Simulation (100 000 Berechnungen) für $\alpha_{r_0} = -\alpha_{l_0}$:

Distanz	$\mu_z; \sigma_z$ (M.-C.)	$z(\Delta\alpha); z(\Delta\alpha \pm \sigma_{\Delta\alpha})$	$\sigma_z^{n=12}; \sigma_z^{n=24}$ (M.-C.)
40 cm, $\alpha_{l_0} \approx 4,7^\circ$	-40,4; 4,0	-40,1; -36,6; -44,3	1,16; 0,82
60 cm, $\alpha_{l_0} \approx 3,1^\circ$	-61,2; 9,5	-60,0; -52,5; -70,1	2,75; 1,94
80 cm, $\alpha_{l_0} \approx 2,3^\circ$	-83,3; 18,8	-80,0; -67,2; -99,0	5,33; 3,83

Die Ergebnisse der Näherung liegen sehr nah bei den Ergebnissen der Monte-Carlo-Simulation – die Näherung ist demnach aussagekräftig. Die große Streuung macht aber eine direkte Nutzung der Distanz als Eingabe für die Interaktion unmöglich. Erst bei Mittelung von mehreren Blickpunkten erreicht die Streuung einen Bereich der praktisch nutzbar ist. So liegt beispielsweise, bei einer wirklichen Blickdistanz von 60 cm, die ermittelte Blickdistanz mit 95% Wahrscheinlichkeit innerhalb von $\pm 3,88$ cm, wenn 24 Blickpunkte gemittelt werden.

Mittels der in diesem Abschnitt vorgestellten Näherung kann sehr einfach eine Abschätzung des zu erwartenden Fehlers bei der Ermittlung des 3D-Blickpunkts durchgeführt werden. Die Näherung ist dabei unabhängig vom eingesetzten Blickerfassungssystem und lässt sich über die Art der Verteilung und die Streuung parametrisieren. Voraussetzung für die Anwendbarkeit ist jedoch die zufällige Verteilung der Fehler: Systematische Fehler werden nicht erfasst.

²Für zwei normalverteilte Zufallsgrößen gilt $\sigma_{a+b}^2 = \sigma_a^2 + \sigma_b^2$, so dass gilt $\sigma_{\Delta\alpha} = \sqrt{2}\sigma_\alpha$.

³Die Abweichung der Näherung beträgt weniger als 1% für $\alpha < 10^\circ$.

3 Anwendung

Die Navigation bei Desktop-VR-Systemen findet meist mittels herkömmlicher Eingabegeräte statt: Die Tastatur wird zur Anpassung der Position, die Maus zur Änderung der Orientierung genutzt. Praktisch bedeutet diese Navigationstechnik, dass die Maus für die Interaktion nicht mehr zur Verfügung steht. Als Alternative bietet sich eine Steuerung der Navigation über die Änderung des 3D-Blickpunkts an [EPR06] – die Genauigkeit des 3D-Blickpunkts ist dabei vernachlässigbar. Andererseits kann aber die Navigation auch weiter über die herkömmlichen Eingabegeräte erfolgen, wenn die Interaktion mit dem VR-System mittels 3D-Blicksteuerung erfolgt [KJK⁺06]. Dabei müssen die Interaktionselemente entsprechend der Genauigkeit der Blickerfassung gewählt werden.

Die Fehlerabschätzung hat gezeigt, dass eine Genauigkeit von $\pm 3,88$ cm möglich ist. Für die praktische Nutzung bei einem Desktop-VR-System (60 cm Betrachterdistanz, 35 cm Tiefe des Sichtvolumens) bedeutet dies, dass ca. 5 Bereiche in der Tiefe unterschieden werden können. Eine erste empirische Überprüfung hat dies bestätigt. Binokulare 3D-Blickpunkterfassung hat damit das Potential in VR-/MR- und AR-Anwendungen zur Analyse aber auch als direkte Eingabe eingesetzt zu werden.

Literatur

- [AB09] Tobii Technology AB. Tobii T/X Series Eye Trackers (Product Description) Rev. 2.0, 2009.
- [EPR06] Kai Essig, Marc Pomplun und Helge Ritter. A neural network for 3D gaze recording with binocular eye trackers. *International Journal of Parallel, Emergent and Distributed Systems*, 21(2):79, 2006.
- [Hin03] Ken Hinckley. Input technologies and techniques. In *The human-computer interaction handbook: fundamentals, evolving technologies and emerging applications*, Seiten 151–168. L. Erlbaum Associates Inc., 2003.
- [Jac90] Robert J. K. Jacob. What you look at is what you get: eye movement-based interaction techniques. In *Proceedings of the SIGCHI conference on Human factors in computing systems: Empowering people*, Seiten 11–18, Seattle, Washington, USA, 1990. ACM.
- [JGL08] J. Jimenez, D. Gutierrez und P. Latorre. Gaze-based Interaction for Virtual Environments. *Journal of Universal Computer Science*, 14(19):3085–3098, 2008.
- [KJK⁺06] Yong-Moo Kwon, Kyeong-Won Jeon, Jeongseok Ki, Qonita M. Shahab, Sangwoo Jo und Sung-Kyu Kim. 3D Gaze Estimation and Interaction to Stereo Display. *The International Journal of Virtual Reality*, 5(3):41–45, 2006.
- [SJ00] Linda E. Sibert und Robert J. K. Jacob. Evaluation of eye gaze interaction. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, Seiten 281–288, The Hague, The Netherlands, 2000. ACM.
- [WAB93] Colin Ware, Kevin Arthur und Kellogg S. Booth. Fish tank virtual reality. In *Proceedings of the INTERACT '93 and CHI '93 conference on Human factors in computing systems*, Seiten 37–42, Amsterdam, The Netherlands, 1993. ACM.