

Dreidimensionale Erfassung visueller Aufmerksamkeit für Usability-Bewertungen an virtuellen Prototypen

Dipl.-Inform. Thies Pfeiffer, Prof. Dr. Ipke Wachsmuth

AG Wissensbasierte Systeme, Technische Fakultät, Universität Bielefeld

Universitätsstraße 25, 33615 Bielefeld

Tel. 0521/106-6999

E-Mail: [tpfeiffe,ipke}@techfak.uni-bielefeld.de](mailto:{tpfeiffe,ipke}@techfak.uni-bielefeld.de)

Zusammenfassung

Die Messung visueller Aufmerksamkeit mittels Eye-Tracking ist eine etablierte Methode in der Bewertung von Ergonomie und Usability. Ihr Gegenstandsbereich beschränkt sich jedoch primär auf 2D-Inhalte wie Webseiten, Produktfotos oder –videos. Bewegte Interaktion im dreidimensionalen Raum wird selten erfasst, weder am realen Objekt, noch am virtuellen Prototyp. Mit einer Aufmerksamkeitsmessung im Raum könnte der Gegenstandsbereich um diese Fälle deutlich erweitert werden. Der vorliegende Artikel arbeitet den aktuellen Stand der Forschung zur Messung visueller Aufmerksamkeit im Raum auf. Dabei werden insbesondere die zu bewältigenden Schwierigkeiten herausgearbeitet und Lösungsansätze aufgezeigt. Als Schwerpunkt werden drei Themen an eigenen Arbeiten diskutiert: Aufbau und Kalibrierung der Systeme, Bestimmung des betrachteten Volumens und Visualisierung der Aufmerksamkeit im Raum.

Schlüsselwörter

Mensch-Maschine-Schnittstelle, Eye-Tracking, Virtuelle Realität, Digital Design Review, Ergonomieuntersuchungen

1 Einleitung

Anwendungsgebiete für eine Aufmerksamkeitsmessung am dreidimensionalen virtuellen Prototyp sind unter anderem die Ausgestaltung von Cockpits, wie z.B. bei Bahn, Auto oder Flugzeug, sowie die Gestaltung von Schalttafeln und Information Dashboards. Interessante Fragestellungen betreffen dabei unter anderem die gezielte Aufmerksamkeitsführung, die Verringerung des Informationsdurcheinanders und insbesondere die Wahrnehmung von Warnsignalen. Durch die Aufmerksamkeitsmessung am virtuellen Prototypen können frühzeitig Designentscheidungen in der dreidimensionalen Interaktion evaluiert und abgesichert werden, ein Prozess, der normalerweise erst später im Entwicklungszyklus stattfinden kann.

Im vorliegenden Beitrag wird neben einer Darstellung der wesentlichen Grundlagen zur visuellen Wahrnehmung der aktuelle Stand der Technik vorgestellt. Dabei wird sowohl auf Arbeiten zur Detektion von Blickbewegungen im Raum, als auch auf aktuelle eigene Arbeiten zu Aufnahme- und Auswertungstechniken eingegangen.

1.1 Augenbewegungen als Indikator für visuelle Wahrnehmung

Die Augen sehen nur in einem kleinen Bereich mit einem Durchmesser von etwa $1,3^\circ$ (Foveola) bis 5° (Fovea) um die Sehachse scharf – dort wo im Auge die Sehgrube ist, eine kleine Einbuchtung mit einer hohen (Fovea) bis sehr hohen (Foveola) Dichte an lichtempfindlichen Zellen. Um feine Details visuell erfassen zu können, muss das Auge daher aktiv ausgerichtet werden. Diese Augenbewegungen, *Sakkaden* genannt, lassen sich messen, genauso wie die Ruhepunkte, oder *Fixationen*, in denen die visuelle Information aufgenommen wird. Der Zielpunkt auf dem das Auge ruht wird auch als *Point-of-Regard* bezeichnet.

Yarbus präsentierte 1967 erstmals Untersuchungsergebnisse über die Analyse von Blickbewegungen auf komplexen Szenen unter variierendem Kontext in Form von Blicktrajektorien [Yar67]. Mit seiner Visualisierungstechnik konnte er die sequentielle Abfolge der Fixationen darstellen (später als *Scanpath* bezeichnet [NS71]) und signifikante Unterschiede im Verlauf nachweisen, der je nach Aufgabenstellung variierte. Diese Ergebnisse stellten den Ausgangspunkt für die Interpretation von Blickbewegungen in Bezug auf Entscheidungsprozesse dar, welche für die Bewertung von Ergonomie und Usability relevant ist.

1.2 Usability-Untersuchungen mittels Blickbewegungsmessung

Mittels Scanpath wird heute z.B. der Aufmerksamkeitsverlauf einer einzelnen Person beim Betrachten einer Webseite analysiert, um Rückschlüsse auf Verbesserungsmöglichkeiten im Layout zu gewinnen (Bild 1, links). Die einzelnen Kreise markieren dabei die Fixationen, ihr Durchmesser die Fixationsdauer und die Verbindungslinien stellen die Sakkaden dar. An Hand von Scanpaths lässt sich die Aufmerksamkeitsführung einer Webseite optimieren oder die Interaktion zwischen inhaltlich unterschiedlichen oder sich ergänzenden Bereichen ermitteln.

Um die Aufmerksamkeitsverteilung über mehrere Personen und über einen längeren Zeitraum darzustellen, wurden *Aufmerksamkeitslandschaften* – heute oft als *Heatmaps* bezeichnet – entwickelt (Bild 1, rechts) [PRV96]. Diese heben die Aufmerksamkeitsverteilung auf dem Originalbild optisch hervor, entweder durch Aufhellung der häufig und Abschattung der weniger betrachteten Bereiche, oder wie eine Hitzekarte durch unterschiedliche Farbtöne. So wird auf einen Blick die Relevanz individueller Bereiche für die visuelle Wahrnehmung reflektiert. Bei Webseiten wird damit die Platzierung von Marken geprüft oder eine Bewertung von Werbeplätzen vorgenommen.

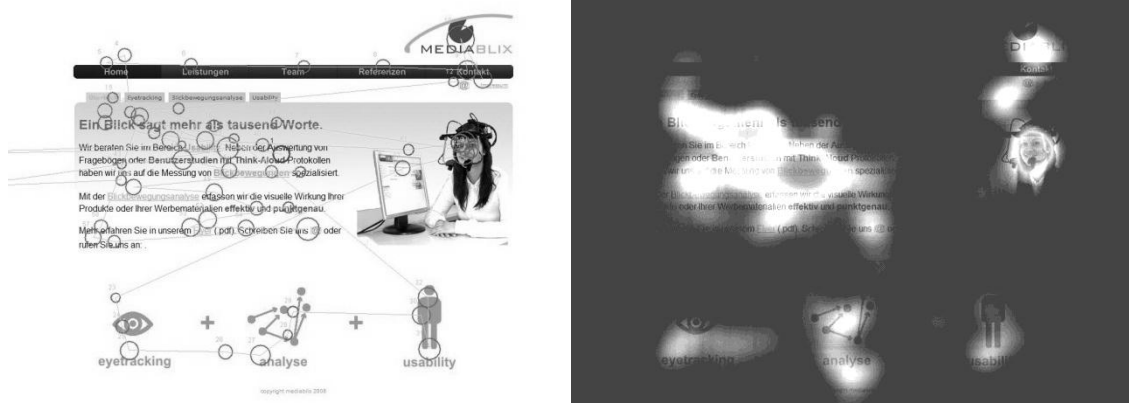


Bild 1: Zwei Darstellungen der gleichen Homepage. Ein Scanpath (links) zeigt den Aufmerksamkeitsverlauf einer einzelnen Person. Die Aufmerksamkeitslandschaft (rechts) ermittelt die Relevanz von Bereichen für mehrere Personen.

Die beschriebenen Methoden haben gemeinsam, dass sie sich zunächst nur mit Blicken auf einer zweidimensionalen Ebene auseinandersetzen. Im Folgenden wird genauer auf die Herausforderung und existierende Lösungsansätze für die Erfassung und Verwertung von Blickinformationen im Raum eingegangen, bevor dann im Anschluss in den Abschnitten 3 und 4 eigene Arbeiten vorgestellt und in Abschnitt 5 diskutiert werden.

2 Stand der Forschung: Aufmerksamkeit im Raum

Einen frühen halbautomatischen Ansatz zum Erfassen der Aufmerksamkeit auf realen Objekten im Raum realisierten Rötting, Göbel und Springer [RGS99]. Sie kombinierten einen Eye-Tracker mit Szenenkamera und einen 6DOF-Tracker für Kopfposition und Orientierung. Die Zuordnung von Blicken zu Objekten im Raum wurde in einem zweistufigen Verfahren im Anschluss an die Aufzeichnung ermittelt. Zuerst wurden die Objektkonturen auf der Bildebene der Szenenkamera aus zwei Perspektiven markiert. Damit rekonstruieren die Autoren ein geometrisches Näherungsmodell der Objekte im Raum. Dann wurden die Fixationen, die als 2D-Punkte auf dem Bild der Szenenkamera vorlagen, klassifiziert und zu jedem Zeitpunkt das betrachtete Objekt (*Model-of-Interest*) ermittelt. Mit diesem einfachen Verfahren konnte zumindest eine Verschiebung auf einer Achse automatisch ausgeglichen werden. Das System wurde nach eigenen Angaben für die Entwicklung eines neuen 3D-Radar-Displays verwendet.

2.1 Objektbezogene Aufmerksamkeitserfassung

Der oben beschriebene Ansatz nutzt zur Auswertung der Blickinformationen geometrische Näherungsmodelle der Zielobjekte und liefert objektbezogene Informationen über den Aufmerksamkeitsverlauf. Die Zuordnung der Fixationen erfolgt nach einer Projektion der Modelle auf die Bildebene. Auf die explizite Berechnung einer Blickrichtung wurde dabei verzichtet. Für die realen Objekte wurden diese Modelle manuell erstellt.

Werden Inhalte ohnehin bereits geometrisch beschrieben und über die Zeit verfolgt, wie in der Erweiterten oder Virtuellen Realität, so vereinfacht sich dieser Prozess.

In ersten Arbeiten zur Aufmerksamkeitsmessung in der Virtuellen Realität, wurde auf Head-Mounted Displays (HMDs) zurückgegriffen [TJ00][DMG+01]. Auch hier wurde objektbezogen gearbeitet. Als „Szenenkamerabild“ kann dabei die für das jeweilige Auge berechnete Projektion verwendet werden, was den Prozess weiter vereinfacht. Ob eine Fixation auf einem Objekt liegt oder nicht, kann so direkt auf der Bildebene festgestellt werden, wie in klassischen Picking-Verfahren für die Interaktion mit 3D-Objekten. Alternativ, wie auch bei [DMG+01], wird die Blickrichtung aus der bekannten Position des Auges, der Lage des Displays im HMD und der vom Eye-Tracker gemessenen 2D-Fixation auf dem Display berechnet. Die vorliegende Objektgeometrie kann dann mit der Blickrichtung geschnitten werden, um den *3D-Point-of-Regard* (3D-POR) und damit das Model-of-Interest zu ermitteln.

Heute werden leichte, kleine Eye-Tracker auch vor PowerWalls oder in CAVEs eingesetzt. Diese Systeme lassen sich mit der für Projektion und Tracking notwendigen Technik kombinieren, wie im Bild 2 am Eigenbau einer Kombination von Eye-Tracker, Polarisationsfilter und optischem Marker zu sehen ist.

Der größte Nachteil der Bestimmung des 3D-Point-of-Regard über Geometriemodelle ist, dass aus der einzelnen Blickrichtung noch keine Aussage über die Tiefe getroffen werden kann. Liegen mehrere Objekte hintereinander in Blickrichtung wird dann fälschlich ein 3D-POR auf dem nächstliegenden Objekt bestimmt. Der umgekehrte Fall kann bei der Betrachtung kleiner Objektprojektionen eintreten: geht der gemessene Blick am Objekt vorbei, wird fälschlich ein 3D-POR auf einem weiter weg gelegenen Objekt bestimmt. Diese Probleme werden noch verstärkt durch Materialeigenschaften wie Transparenz (Blick durch ein Fenster) oder Spiegelung, sowie durch die Verwendung von Geometrie-Shadern, da dann die Objektgeometrie nicht direkt verfügbar ist.

2.2 Holistische Aufmerksamkeitserfassung

Bei der holistischen Aufmerksamkeitserfassung wird auf Geometriemodelle verzichtet. Stattdessen werden mehrere gemessene Informationen integriert, um im einfachsten Fall an Hand einer Triangulierung den 3D-POR direkt zu ermitteln. Notwendig sind dazu Blickrichtungsinformationen aus mindestens zwei Perspektiven. Diese können entweder über einen zeitlichen Abstand von Blickrichtungsmessungen eines einzelnen Auges eines bewegten Betrachters [KJK+06] oder durch simultanes Messen der Blickrichtungen beider Augen gewonnen werden [EPR06].

Da bei der holistische Aufmerksamkeitserfassung explizit auch die Tiefe des 3D-POR bestimmt wird, werden die in 2.1 beschriebenen Nachteile der geometriebasierten Verfahren vermieden. Dieser Ansatz wird daher auch in den eigenen Arbeiten verfolgt.

2.3 Aufmerksamkeit im Raum visualisieren

Wurden die 3D-POR im Raum ermittelt, entweder durch Schnitte mit Geometriemodellen oder holistisch, ist eine Darstellung als *3D-Scanpath* einfach. Dazu werden Fixationen als Kugeln modelliert, deren Durchmesser wie bei den Fixationskreisen als Funktion der Fixationsdauer festgelegt wird. Die Sakkaden können dann z.B. als Zylinder zwischen den Kugeln realisiert werden (Bild 3, links).

Im Bereich der objektbezogenen Darstellungen gibt es mehrere Alternativen: die einfache objektzentrierte Einfärbung des gesamten Objektes (Model-of-Interest) oder eine oberflächenzentrierte Darstellungen, bei der ähnlich zu den 2D-Heatmaps eingefärbte Texturen generiert und überlagert werden [SND10]. Die objektbezogenen Darstellungen basieren dabei stets auf modellbasierten Fixationsberechnungen. Dies ist jedoch nicht zwingend, die Färbung könnte auch durch die Projektion holistisch ermittelter Fixationen bestimmt werden. Ein direktes Pendant für 2D-Heatmaps zur Darstellung holistisch ermittelter Fixationen existierte bisher nicht, wird aber in Abschnitt 4 eingeführt.

2.4 Aufmerksamkeit im Raum nutzen

Die Analyse der Aufmerksamkeitsverteilung im Raum könnte für die Bewertung von Usability und Ergonomie am virtuellen Prototyp einen ähnlichen Nutzen haben, wie die 2D-Verfahren, die heute schon zum Alltag von Mediengestaltung und Produktevaluierung von der Webseite bis zum Point-of-Sale gehören.

Neben der Nutzung der Aufmerksamkeitsverteilung für die Weiterentwicklung des Designs kann auch im Digitalen Design Review das Wissen über die aktuelle Aufmerksamkeit eines Kooperationspartners online verwendet werden – zum Beispiel dann, wenn sich die Partner in einer verteilten VR-Umgebung befinden [DCC+04].

3 Aufmerksamkeit im Raum messen

Im Folgenden wird der eigene Ansatz beschrieben, mit dem visuelle Aufmerksamkeit im Raum gemessen wird. In unseren aktuellen Forschungsprojekten werden die so gewonnenen Informationen nicht nur in Studien eingesetzt, sondern auch um virtuelle Objekte über Blicke zu selektieren, um im Fokusbereich Informationen lokal zu präsentieren und um mit virtuellen Agenten zu kommunizieren.

Als wesentliche Herausforderungen kristallisieren sich die Integration der Hardware, die Kalibrierung der Systeme und die robuste Bestimmung des 3D-Point-of-Regard heraus.

3.1 Hardware-Auswahl und Integration

Notwendige Grundvoraussetzung ist eine geeignete Hardware-Ausstattung, die kombiniert und aufeinander abgestimmt werden muss. Dies umfasst beim Einsatz in der virtu-

ellen Realität mindestens drei Komponenten: ein Projektionssystem, ein Trackingsystem für Kopf und Körper sowie einen geeigneten Eye-Tracker.

Die meisten der Komponenten werden in existierenden Systemen bereits vorhanden sein. Im Bielefelder Labor für Künstliche Intelligenz und Virtuelle Realität wird die Stereo-Projektion mittels Polarisationsfiltern erreicht. Möchte man diese mit einem Eye-Tracker kombinieren, so ist darauf zu achten, dass die Polarisationsfilter das Infrarotlicht des Eye-Trackers passieren lassen. Zwar kann die Blickerfassung auch von unterhalb des Rahmens der VR-Brille erfolgen, dies führt jedoch zu einem steilen Winkel auf das Auge, was die Kalibrierung erschwert und die Genauigkeit verringert.

Die Verfolgung der Kopfposition und -orientierung, sowie der gestischen Interaktion erfolgt in unserem Labor mit einem System der A.R.T. GmbH (DTrack). Dieses optische Tracking System arbeitet sowohl mit passiven, als auch mit aktiven Markern. Für beide Varianten werden Infrarot-Blitze eingesetzt, die im Fall der passiven Marker für eine gezielte Reflektion, im Fall der aktiven Marker zur Synchronisation der Schaltung der Marker-LEDs dienen. In der Praxis wurde dabei keine Interferenz mit dem Eye-Tracking-System festgestellt.

Nach Evaluation eines SMI EyeLink I und eines Arrington Research ViewPoint PC-60 BS007 [PLW09], wurde für den interaktiven Einsatz das Modell von Arrington gewählt. Es ist günstiger in der Anschaffung, leicht zu kalibrieren und sehr flexibel. Wie in Bild 2 zu sehen, konnte das Gerät auf einfache Weise mit den Polarisationsfiltern (Mitte) und den Markern des Tracking-Systems (rechts) kombiniert werden. Im Bild sieht man jeweils eine Kamera für jedes Auge, sowie getrennt davon einstellbar jeweils eine Infrarot-LED zur besseren Ausleuchtung der Augen unter den Bedingungen in der CAVE. Wünschenswert wäre eine kabellose Konstruktion, die aber derzeit von keinem Hersteller angeboten wird.



Bild 2: Der Eigenbau eines für den Einsatz in der Virtuellen Realität angepassten Systems zur Aufmerksamkeitsmessung kombiniert einen Eye-Tracker von Arrington Research (Mitte, unten) mit Polarisationsfiltern (Mitte) und Markern für ein optisches Tracking System (Oben, rechts).

3.2 Integration von Eye- und Head-Tracking

Die gewählte Hardware-Kombination von Eye-Tracker, optischem Marker und Polarisationsfiltern vereinfacht die Integration der beiden Verfahren. Im Szenengraph der VR-Anwendung wird der Kopf des Benutzers als Transformationsknoten repräsentiert, dessen Transformationsmatrix direkt aus den Tracking Daten des A.R.T.-Systems gespeist wird. Durch die seitliche Anbringung des optischen Markers an der Brille muss noch eine Korrekturmatrix angewendet werden, um die Standardreferenz für die Kopfposition als Punkt zwischen beiden Augen mit einer Orientierung entlang negativer Z-Achse zu erhalten. Unterhalb dieses Transformationsknotens repräsentieren zwei weitere Knoten die Position des linken und des rechten Auges, mit einer entsprechenden Translation, die benutzerspezifisch ermittelt wird. Die Welttransformation der Knoten für die beiden Augen kann jeweils für die Berechnung der zugehörigen Projektionsmatrix für die Visualisierung im VR-System verwendet werden.

Der Eye-Tracker liefert nach Vorverarbeitung und Kalibrierung (im nächsten Abschnitt genauer erläutert) die Orientierung der beiden Knoten für die Augen. Damit spiegeln diese Knoten im Szenengraph nicht nur die aktuelle Position der Augen mit einberechneter Kopforientierung wieder, sondern lassen an der Ausrichtung der negativen Z-Achse auch die aktuelle Blickrichtung erkennen.

3.3 Kalibrierung der Systeme

Das optische Tracking-System detektiert in mehreren 2D-Kamerabildern die Positionen der Marker. Durch die Kalibrierung des Systems, für die Verfahren vom Hersteller bereitgestellt werden, wird eine Abbildung der ermittelten 2D-Positionen auf 3D-Position und Orientierung (bei Verwendung von komplexen Targets) im Raum erreicht.

Die von den Herstellern der Eye-Tracker entwickelte Software leistet in der Regel die Abbildung der detektierten 2D-Pupillenposition im Kamerabild des Eye-Trackers auf eine 2D-Position auf einer Ebene. Zur Kalibrierung werden mehrere Referenzpunkte auf dieser Ebene präsentiert und jeweils die dazu von den Augen eingenommenen Orientierungen ermittelt. Bei stationären Systemen entspricht diese Ebene in der Regel einer Bildschirmenebene, auf der dann Stimuli präsentiert werden. Bei mobilen Geräten wird häufig mit einer Szenenkamera gearbeitet, die ein Live-Bild aus der Nutzerperspektive aufzeichnet. In diesem Fall wird auf eine bestimmte Ebene im Aufnahmebereich der Szenenkamera kalibriert. Dies kann z.B. eine weiße Wand sein, auf welcher die Referenzpunkte von einer zweiten Person manuell per Laserpointer angezeigt werden. Für beide Formen der Kalibrierung ist es wesentlich, dass die Kalibrierungsebene senkrecht zur entspannten Blickrichtung steht und dass der Kalibrierungsbereich möglichst das alleine durch Augenbewegungen erfassbare Blickfeld einnimmt.

Für die Integration der beiden Tracking-Systeme kann die vom Hersteller angebotene 2D-Kalibrierung des Eye-Trackers auf einer Bildschirm- oder Projektionsebene ver-

wendet werden. Dann wird aus der bekannten Größe des Bildschirms, den vom Eye-Tracker gelieferten 2D-Koordinaten auf dieser Ebene und dem Abstand des Nutzers die Orientierung der Augen ermittelt und als Rotationsmatrix wie oben beschrieben in den Szenengraph eingespeist.

Dieses Vorgehen hat in der Praxis jedoch mehrere Nachteile. Zum einen muss zwischen der Darstellung der 2D-Kalibrierung und der virtuellen Realität umgeschaltet werden. Da eine Kalibrierung auch während einer längeren Anwendungssitzung eventuell erneuert werden muss, kann sich dieser Faktor mehrfach störend auswirken. Zum anderen muss der Nutzer während der Kalibrierung den Kopf möglichst ruhig und in konstantem Abstand von der Kalibrierungsebene halten. Zudem muss die Anordnung von Kalibrierungsebene und Kopf des Nutzers dem System möglichst exakt bekannt sein. Dies wird häufig durch den (unkomfortablen) Einsatz einer Kinnstütze erreicht.

Speziell für den Einsatz in der virtuellen Realität wurde deswegen in unserem Labor ein neues Verfahren entwickelt. Ausgangspunkt ist dabei die Erkenntnis, dass für die Kalibrierung des Eye-Tracking-Systems im Wesentlichen eine Ebene senkrecht zur entspannten Blickrichtung mit konstantem Abstand zum Nutzer benötigt wird. Eine solche Ebene lässt sich in der virtuellen Realität leicht konstruieren. Wie oben beschrieben ist die Weltposition jedes Auges über das optische Trackingsystem bekannt und im Szenengraph bereits als Transformationsknoten repräsentiert. An diesen Knoten wird nun eine entsprechende Ebene mit konstanter Translation geknüpft, auf der sich anwendungssteuert virtuelle Referenzpunkte präsentieren lassen. Da sich die Ebene stets mit dem Kopf des Nutzers mitbewegt, kann sich dieser auch während der Kalibrierung frei bewegen, ohne das Ergebnis zu gefährden. Durch die Einbettung der Kalibrierungsebene in die VR-Anwendung kann die Kalibrierung direkt im Kontext geschehen. Damit wird der Immersionsbruch reduziert.

3.4 Bestimmung der Blickrichtung

Nach der Kalibrierung liefert der Eye-Tracker 2D-Fixationen als Koordinaten auf der virtuellen Ebene. Nachdem wie in 3.2 und 3.3 beschrieben die Positionen des Auges und die relative Lage der Ebene bekannt sind, lässt sich die Blickrichtung als Vektor vom Zentrum des Auges zum 2D-Fixationspunkt auf der Ebene bestimmen. Damit lässt sich die Rotation des Auges berechnen und als Matrix im Szenengraph repräsentieren. Die so bestimmte Blickrichtung kann für die Interaktion (z.B. mit Standard-Picking-Verfahren) und die Aufzeichnung von Messdaten verwendet werden [Pfe08].

3.5 Bestimmung der Blickposition

Zur holistischen Bestimmung der Blickposition eines bewegten Nutzers wurden die binokulare Triangulierung und ein maschinelles Lernverfahren [EPR06] erweitert und gegeneinander verglichen [PLW09]: Die mit dem maschinellen Lernverfahren erzielten

Ergebnisse waren signifikant genauer. Die ermittelten Fehler der Tiefenschätzung lagen bei -2 cm mit einer Standardabweichung von 10 cm. Als Ziel wurden Baufix-Teile in einer Konstruktion mit einer Kantenlänge von nur 30 cm verwendet (Bild 3, links, Blickrichtung von links unten nach rechts oben). Die einzelnen Objekte hatten eine Ausdehnung von 1 – 3 cm. Damit ist die Zielkonstruktion mit fünf innerhalb von 30 cm zu unterscheidenden Tiefenebenen komplexer als z.B. typische Cockpit-Systeme.

Bild 3, links, zeigt die so ermittelten 3D-PORs einer Testperson als 3D-Scanpath. Die Abweichungen in Breite und Höhe sind minimal, in der Tiefe sind kleine Abweichungen erkennbar, die jedoch systematisch sind und eine Zuordnung zu den fünf Tiefenebenen zulassen. Die Struktur der Zielkonstruktion ist bis auf eine Streckung in der Tiefe ebenfalls deutlich zu erkennen.

4 Aufmerksamkeit visualisieren: Aufmerksamkeitsvolumina

Für die holistische Visualisierung der Aufmerksamkeit werden die Bereiche hoher Auflösung um die 3D-POR als *Aufmerksamkeitsvolumina* (*Attention Volumes*) modelliert. Analog zum Punkt der Aufmerksamkeit (Point-of-Regard, $\overrightarrow{p_{POR}}$) oder der *Region von Interesse* (*Region-of-Interest, ROI*) in den 2D-Modellierungen, wird hier also das *Volumen von Interesse* (*Volume-of-Interest, VOI*) modelliert. In erster Näherung kann dies durch eine Kugel geschehen, die leicht verzerrt ist, da der Radius bei gegebenem Winkel scharfen Sehens α ($2,5^\circ$ für den Bereich der Fovea) mit der Entfernung vom Auge des Betrachters ($\overrightarrow{p_{eye}}$) wächst:

$$\begin{aligned} \text{Aufmerksamkeitsvolumen: } (\vec{x} - \overrightarrow{p_{POR}})^2 &\leq r(\vec{x})^2 \\ \text{Radius: } r(\vec{x}) &= |\vec{x} - \overrightarrow{p_{eye}}| \tan \alpha \end{aligned}$$

Durch eine 3D-Gaussverteilung kann die Modellierung noch feiner gestaltet werden, in dem jedem Punkt \vec{x} im Raum noch ein Gewicht $w(\vec{x})$ zugeordnet wird. Dieses Gewicht berücksichtigt den Abstand jedes Punktes zum Punkt der Aufmerksamkeit (3D-POR) sowie mit dem Gewichtungsfaktor $d(t)$ die Dauer der Fixation:

$$\text{Gewicht: } w(\vec{x}) = d(t) e^{-\frac{|\vec{x} - \overrightarrow{p_{POR}}|^2}{r(\vec{x})}}$$

Die Aufmerksamkeitsvolumina können dann mit Verfahren des Volume Renderings im 3D-Raum dargestellt und an Hand der Gewichte eingefärbt werden (Bild 3, rechts). Die Visualisierung ist dabei interaktiv und kann direkt in der VR-Anwendung über dem Prototypen dargestellt und von allen Seiten inspiziert werden.

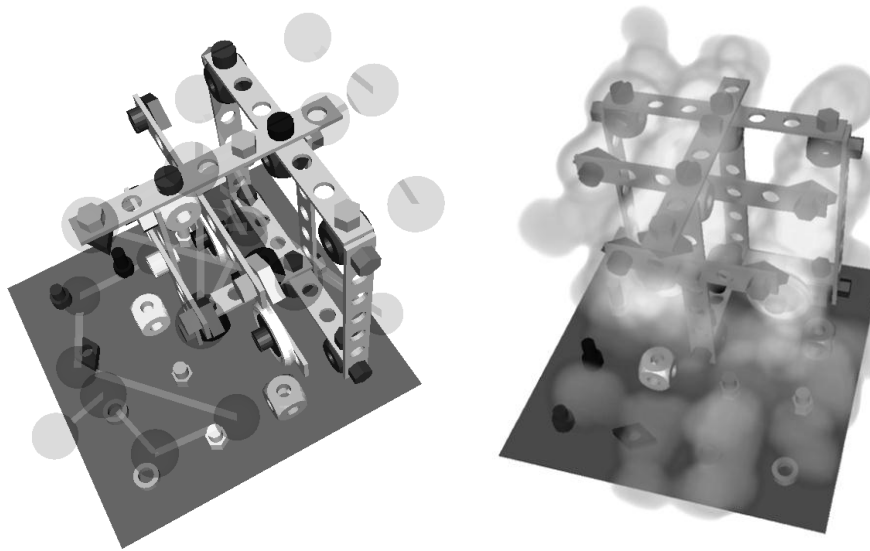


Bild 3: Links der 3D-Scanpath einer einzelnen Person, rechts ein Volume Rendering des Aufmerksamkeitsvolumens für zehn Personen über einer Beispielkonstruktion mit kleinen Bauteilen. Die Kantenlänge der Konstruktion beträgt 30 cm.

Durch die perspektivenunabhängige Modellierung kann das Aufmerksamkeitsvolumen auf verschiedene Perspektiven projiziert werden, so dass z.B. auch Videos mit Kamerafahrten möglich sind. Die Modellierung lässt sich auch auf modellfrei ermittelte 3D-POR anwenden. So kann die Projektion z.B. auch auf das Bild einer Szenenkamera erfolgen, um Aufmerksamkeitsvolumina auf realen Objekten darzustellen. Die visuelle Qualität der Darstellung kann durch Informationen über die Geometriemodelle noch verbessert werden, um Verdeckungen in der Darstellung berücksichtigen zu können.

5 Resümee und Ausblick

Ausgangspunkt für die Betrachtung war, dass eine Bewertung von Ergonomie und Usability am virtuellen dreidimensionalen Prototypen Vorteile bringen kann, die Methode der visuellen Aufmerksamkeitsmessung jedoch derzeit in der Regel auf 2D-Stimuli angewiesen ist. Für eine Analyse der visuellen Aufmerksamkeit im Raum wurden daher Verfahren zur Bestimmung des 3D-Point-of-Regards, zur immersiven Kalibrierung und zur 3D-Visualisierung entwickelt, die den notwendigen Prozess vollständig abbilden.

Objektbasierte Verfahren zur Berechnung der Aufmerksamkeit im Raum eignen sich gut für den Einsatz in virtueller Realität, da dort die benötigten Geometrieinformationen zwangsläufig vorliegen und direkt genutzt werden können – mit den in 2.1 beschriebenen Einschränkungen. Dann kann eventuell auf binokulares Tracking verzichtet werden, um Kosten zu reduzieren. Es sollte jedoch das dominante Auge des Benutzers gemessen werden, da sich sonst größere Messfehler ergeben können. In bestimmten Kontexten, wie z.B. beim Zeigen mit der Hand, wurde aber auch ein dynamischer Wechsel in der Augendominanz nachgewiesen, so dass genaue Ergebnisse nur mit binokularem Tra-

cking erzielt werden können, auch wenn nur die Blickrichtung bestimmt werden soll. Das entwickelte holistische Verfahren kommt ohne diese Nachteile aus und kann schneller und flexibler eingesetzt werden. Durch Verzicht auf Geometriemodelle eignet es sich neben dem Einsatz in der virtuellen Realität auch für die reale Welt.

Die neu entwickelten Attention Volumes erlauben eine immersive, interaktive Darstellung der Ergebnisse am virtuellen Prototyp. Zusammen mit 3D-Scanpaths und objekt- oder oberflächenzentrierten Aufmerksamkeitslandschaften, liegen damit nun Pendants der 2D-Visualisierungen für die Analyse im Raum vor. Somit erschließt sich der Bereich des Virtual Prototyping der Analyse durch Usabilitybetrachtungen.

Wesentliche Fortschritte lassen sich in einer genaueren Modellierung des Aufmerksamkeitsvolumens um einen 3D-Point-of-Regard erwarten. Die vorgestellte Modellierung liefert eine erste Näherung, vergleichbar mit der Genauigkeit der Modellierung für 2D-Inhalte. Eine genauere Modellierung sollte die Form der Ausdehnung in der Tiefe und die Interaktion zwischen beiden Augen bei binokularem Sehen stärker berücksichtigen.

In der praktischen Anwendung spielen methodische Aspekte eine wesentliche Rolle. Das System muss schnell und möglichst unauffällig kalibriert werden können. Weiterhin sollte eine genaue Messung über einen langen Zeitraum mit wenigen Kalibrierungen aufrechterhalten werden können. Aktuelle Arbeiten beschäftigen sich daher damit, die Handhabung des Systems zu vereinfachen, indem Verfahren zur kontinuierlichen Beobachtung der Messgenauigkeit und zur unauffälligen Kalibrierung entwickelt werden.

Neben der Erfassung der Aufmerksamkeit von Testpersonen für die Bewertung der Usability kann diese Technik auch in der visuellen Analyse vom Experten genutzt werden. Erste Visionen zur Interaktion mit Blicken finden sich dabei bereits bei Bolt [Bol81]. Durch die Einblendung von Informationen im Bereich der aktuellen Aufmerksamkeit lässt sich z.B. eine visuelle Überfrachtung der Darstellung verringern [PM09].

Literatur

- [Bol81] BOLT, R.: Gaze-orchestrated dynamic windows. Proc. of the 8th Annual Conference on Computer Graphics and Interactive Techniques, pp. 109–119. ACM Press New York, 1981
- [Dod07] DODGE, R.: Studies from the psychological laboratory of Wesleyan University: An experimental study of visual fixation. Psychological Monographs. APA, 1907
- [DMG+01] DUCHOWSKI, A. T.; MEDLIN, E.; GRAMOPADHYE, A.; MELLOY, B.; NAIR, S.: Binocular Eye Tracking in VR for Visual Inspection Training. Virtual Reality Software and Technology ACM: Symposium on VR Software and Technology. ACM Press, 2001
- [DCC+04] DUCHOWSKI, A. T.; COURNIA, N.; CUMMING, B.; MCCALLUM, D.; GRAMOPADHYE, A.; GREENSTEIN, J.; SADASIVAN, S.; TYRRELL, R. A.: Visual Deictic Reference in a Collaborative Virtual Environment, ETRA Symposium 2004, pp. 35–40. ACM Press, 2004
- [EPR06] ESSIG, K.; POMPLUN, M.; RITTER, H.: A neural network for 3D gaze recording with binocular eye trackers. Journal of Parallel, Emergent and Distributed Systems (21), pp. 79 – 95, 2006

- [KJK+06] KWON, Y.-M.; JEON, K.-W.; KI, J.; SHAHAB, Q. M.; JO, S.; KIM, S.-K.: 3D Gaze Estimation and Interaction to Stereo Display. *Journal of Virtual Reality*, 5, pp. 41–45, 2006
- [Pfe08] PFEIFFER, T.: Towards Gaze Interaction in Immersive Virtual Reality: Evaluation of a Monocular Eye Tracking Set-Up. In Schumann, M.; Kuhlen, T.: *Virtuelle und Erweiterte Realität – 5. Workshop der GI-Fachgruppe VR/AR*, pp. 81–92. Shaker Verlag GmbH, 2008
- [PLW09] PFEIFFER, T.; LATOSCHIK, M. E.; WACHSMUTH, I.: Evaluation of Binocular Eye Trackers and Algorithms for 3D Gaze Interaction in Virtual Reality Environments. In Herder, J.: *Journal of Virtual Reality and Broadcasting*, 2009
- [PM09] PFEIFFER, T.; MATTAR, N.: Benefits of Locating Overt Visual Attention in Space Using Binocular Eye Tracking for Mixed Reality Applications. In Kain, S.; Struve, D.; Wandke, H.: *Workshop-Proceedings der Tagung Mensch & Computer 2009: Grenzenlos frei!?*, pp. 272–274. Logos, Berlin, 2009
- [PRV96] POMPLUN, M.; RITTER, H.; VELICHKOVSKY, B.: Disambiguating complex visual information: Towards communication of personal views of a scene. *Perception* 25(8), pp. 931–948, 1996
- [RGS99] RÖTTING, M.; GÖBEL, M.; SPRINGER, J.: Automatic object identification and analysis of eye movement recordings. *MMI-Interaktiv*, 1999
- [SND10] STELLMACH, S.; NACKE, L.; DACHSELT, R.: 3D Attentional Maps - Aggregated Gaze Visualizations in Three-Dimensional Virtual Environments. *Proc. of the Int. Working Conference on Advanced Visual Interfaces (AVI 2010)*, pp. 345–348. ACM Press, 2010
- [TJ00] TANRIVERDI, V.; JACOB, R. J. K.: Interacting with eye movements in virtual environments. *Proc. of the CHI 2000*, pp. 265–272. ACM Press, 2000
- [Ver03] VERTEGAAL, R.: Attentive User Interfaces. *Communications of the ACM* 46 (3), 2003
- [Yar67] YARBUS, A.L.: Eye movements during perception of complex objects. *Eye movements and vision* (7), pp. 171–196, 1967

Autoren

Thies Pfeiffer schloss sein Diplom in Informatik 2003 an der Universität Bielefeld ab, arbeitete danach in Vorhaben des SFB 360 „Situierete Künstliche Kommunikatoren“ und forscht seit 2006 in der Gruppe von Prof. Dr. Ipke Wachsmuth zum Thema Mensch-Maschine-Interaktion. Er promovierte 2010 über multimodale deiktische Ausdrücke mit Schwerpunkt auf Blicken und Zeigegesten. Gefördert vom BMBF gründete Herr Pfeiffer 2006 den Usability-Dienstleister Mediablix und ist dort als Berater aktiv.

Prof. Dr. Ipke Wachsmuth studierte Mathematik und Informatik an der TU Hannover, wo er 1980 promovierte. Nach Tätigkeiten an der Universität Osnabrück, der Northern Illinois University und bei IBM Deutschland habilitierte er sich 1989 an der Universität Osnabrück; im gleichen Jahr wurde er auf eine Professur für Wissensbasierte Systeme an der Universität Bielefeld berufen. Er war dort Gründungsmitglied der Technischen Fakultät (1990), Mitinitiator zweier Sonderforschungsbereiche und von 2002 bis 2009 geschäftsführender Direktor des Zentrums für interdisziplinäre Forschung (ZiF). Er ist derzeitiger Sprecher des SFB 673 „Alignment in Communication“.