

Multimodale Interaktion mit großen Displays

Synergie aus Gesten, Blicken und Berührungen

Raimund Dachselt
Interactive Media Lab
Technische Universität Dresden
raimund.dachselt@tu-dresden.de

Abstract

Among the variety of displays available, large displays are increasingly gaining importance. They can be used for medical imagery, scientific visualization, explorative data analysis, or information displays in public spaces. Traditional input devices are no longer suitable for interacting with them from the distance. Therefore, the research question arises how to develop effective interaction techniques. This paper contributes an overview and classification of possible input modalities and their combination. Three example projects are used for illustrating this growing and exciting field of research along with an outline of central challenges.

Zusammenfassung

Unter den vielfältigen Ausgabegeräten spielen Großdisplays eine zunehmend wichtige Rolle für wissenschaftliche Visualisierungen, explorative Datenanalyse, medizinische Bilder oder Informationsanzeigen im urbanen Raum. Da sich traditionelle Eingabegeräte für deren Bedienung aus der Entfernung nicht eignen, stellt sich die Forschungsfrage nach effektiven Interaktionstechniken. Der Beitrag stellt eine kurze Systematik möglicher Bedienmodalitäten und deren Kombination dar und erläutert anhand eigener Beispielprojekte das vielversprechende Potenzial natürlicher, multimodaler Interaktion. Zentrale Herausforderungen werden abschließend beleuchtet.

Motivation

Während bei klassischen Bürocomputern die Eingabe durch Maus und Tastatur losgelöst von der Ausgabe auf einem Bildschirm erfolgt, erlauben interaktive Oberflächen eine Verschmelzung von Eingabe und Ausgabe. In Form von Multitouch-Smartphones und -Tablets haben sie ihren Siegeszug längst angetreten und sind in unserem Arbeits- und Lebensalltag omnipräsent. Ihre Bedienung erfolgt meistens durch Berührung mit einem oder mehreren Fingern, gelegentlich auch über Stifteingabe. Diese Interaktionsmodalitäten existieren jedoch bisher zumeist alternativ zueinander und werden nicht kombiniert. Forschungsarbeiten, wie z.B. [2], untersuchen hingegen eine effektive *multimodale Interaktion* durch Stift- und Multitouch-Eingabe.

Interaktive Oberflächen kommen jedoch auch zunehmend in deutlich größeren Formfaktoren zum Einsatz. Tabletops, d.h. interaktive Tischdisplays, haben die Forschungslabore längst verlassen und sind kommerziell von verschiedenen Anbietern erhältlich. Gleiches gilt auch für interaktive Wanddisplays, die von großen Fernsehern bis hin zu wandfüllenden, ultra-hochauflösenden Visualisierungssystemen reichen. Auch im öffentlichen Raum ist eine Zunahme an Großdisplays zu beobachten, zumeist zu Informations- oder Werbezwecken.

Viele dieser Großdisplays bieten entweder technisch gar keine direkte Interaktionsmöglichkeit, wie wir sie von Tablets oder Tabletops gewöhnt sind, oder ihr Nutzungskontext gestattet diese nicht. Beispiele dafür sind höher gehängte digitale Anzeigetafeln im öffentlichen Raum oder Monitorhängungen in einem OP-Saal, wobei Ärzte mit den darauf dargestellten medizinischen Bilddaten nur aus der Entfernung interagieren können, weil sie am OP-Tisch stehen. Es ist klar, dass Eingabegeräte wie Maus und Tastatur für die entfernte Interaktion kaum geeignet sind, zumal auch häufig mehrere Nutzer gleichzeitig mit einem derartigen Display interagieren. Somit stellt sich die Forschungsfrage nach effektiven Techniken zur Bedienung entfernter Displays aus einigem Abstand. Dieser Beitrag will zunächst eine kurze Systematik dieses Forschungsgebietes umreißen und dann anhand von eigenen Beispielprojekten neuartige Formen multimodaler Interaktion mit entfernten Displays betrachten.



Abb. 1 Interaktive Visualisierungswand mit 24 Megapixeln am Interactive Media Lab Dresden. Zoom & Pan über Touchbedienung auf einem Smartphone, das gleichzeitig einen Gesamtüberblick bietet.

Natürliche Interaktion mit großen Displays

Großdisplays, wie z.B. in Abb. 1 dargestellt, eignen sich besonders für Anwendungsdomänen, wo sehr große Informationsräume, wissenschaftliche Datensätze oder Visualisierungen angezeigt werden sollen. Typische Beispiele sind Anwendungen in Biologie, Medizin, Stadtplanung, Architektur, Automobildesign, Fabrikplanung, Rohstoffförderung oder Energieversorgung.

Für die Bedienung derartiger Anwendungen lassen sich folgende typische Interaktionsaufgaben identifizieren: die *Navigation* durch große Informationsbestände, z.B. durch Zooming & Panning [4], das Zeigen auf Objekte (*Pointing*) und ihre *Selektion* bzw. Gruppierung, ihre *Manipulation* (z.B. Verschieben, Kopieren, Löschen, affine Transformationen) sowie das Auslösen von Steuerbefehlen zur *Systemkontrolle*, z.B. über Menüs.

Während bei Desktopsystemen Maus und Tastatur für all diese Aufgaben verwendet werden, sind in den letzten Jahren vielfältige Interaktionsmodalitäten erforscht worden, die natürlichere und an unseren Erfahrungen aus dem realen Leben orientierte Bedienhandlungen erlauben [3]. Während sich Multitouch-Eingabe, Stiftinteraktion oder auch die Benutzung von physischen Objekten (z.B. ein Schwamm) für die direkte Interaktion *auf* der Oberfläche eignen, bieten sich für eine *entfernte Interaktion* mit Großdisplays u.a. folgende Modalitäten an (von denen einige z.B. in der vergleichenden Studie [4] getestet wurden):

- Smartphones und Tablets, auf denen Multitouch-Eingaben zur Steuerung genutzt werden, oder auch Bewegungen bzw. Gesten *mit* dem Gerät;
- In der Hand gehaltene Eingabegeräte mit 6 Freiheitsgraden, wie sie vor allem in Virtual Reality Anwendungen verwendet werden, z.B. Flysticks;
- Freihandgesten oder Körperbewegungen, die z.B. mithilfe einer Tiefenkamera erfasst werden;
- Augensteuerung oder Nutzung der Kopforientierung zur Blickapproximation;
- Spracheingabe, jedoch weniger für kontinuierliche Eingabe und kollaboratives Arbeiten geeignet;
- Spezialgeräte, z.B. physische Tangibles – wie kleine Automodelle, die man auf einem Tisch zur Ansichtsteuerung verschieben kann;
- Digitale Stifte und Papier, auf dem Eingaben gezeichnet oder gedruckte Interfacelemente berührt werden können.

Die Bedienmodalitäten unterscheiden sich zunächst in ihren Freiheitsgraden bzw. der Ausdrucksmächtigkeit, in der Präzision der Eingabe, im technischen Aufwand für das Erfassen der Interaktion, in der notwendigen Instrumentierung des Raums mit Zusatzsensorik, in der einfachen Verfügbarkeit der Geräte, den Kosten und in der Tauglichkeit für mehrere Benutzer. Manche erlauben eine Identifikation des jeweiligen Nutzers für kollaboratives Arbeiten, bei anderen ist dies nur mit Zusatzsensorik möglich. Für Gelegenheitsnutzer, z.B. im öffentlichen Raum, kommen natürlich nur sehr einfache Bedienformen oder die Nutzung von mobilen Endgeräten in Betracht, während eine Expertenanwendung (VR-CAVE, OP-Saal oder hochauflösende Visualisierungswand) speziellere und technisch aufwändigere Lösungen gestattet oder sogar erfordert. Schließlich macht es auch einen Unterschied, ob nur der eigene Körper (Stimme, Gesten, Bewegungen) zur Interaktion verwendet wird, ein reines Eingabegerät genutzt wird (z.B. Flystick, Air Mouse, Tangibles) oder aber Eingabe und Ausgabe in einem mobilen Gerät kombiniert sind (z.B. Handy), wodurch auch Zusatzinformationen oder ein Überblick (wie in Abb. 1) dargestellt werden können.

Multimodale Interaktion: Anwendungsbeispiele

Die genannten Bedienmodalitäten können teilweise auf sehr effektive Weise miteinander kombiniert werden, um ihre Nachteile in der alleinigen Nutzung auszugleichen, parallele Aktionen ausführen zu können oder nahtlose Übergänge zwischen verschiedenen Interaktionsformen zu ermöglichen. Anhand von Forschungsbeispielen der eigenen Arbeitsgruppe, dem Interactive Media Lab Dresden an der Informatikfakultät der Technischen Universität Dresden, soll das Potenzial multimodaler Interaktion für große Displays illustriert werden.

Physische Gesten und Berührungen

In unserer Forschungsanwendung *Throw & Tilt* [1] wird eine sehr einfache und natürliche Form der Übertragung von Medienobjekten zwischen persönlichem Handy und Wanddisplay ermöglicht, und das mobile Gerät dient gleichzeitig als multimodales Steuerzentrum für die auf der Wand angezeigte Medienanwendung. Nach lokaler Auswahl von Fotos auf dem Handy werden diese mithilfe einer Wurfgeste an die Wand „geschleudert“, wo sie nach einer kurzen WLAN-Übertragungszeit erscheinen. Das Mobiltelefon wird nun zum reinen Eingabegerät; durch Neigegesten in verschiedene Richtungen kann man zwischen den Bildern blättern (Navigation) und diese einzeln zoomen (Manipulation). Ein erweiterter Prototyp, der auch auf der ACM CHI 2009 demonstriert wurde, erlaubt durch die Verwendung eines iPhones noch zusätzliche *Toucheingabe*. Damit können Moduswechsel vorgenommen oder Bilder bzw. eine GoogleEarth-Ansicht verschoben werden – und das bei sequenziellem oder sogar gleichzeitigem Zoomen durch *Neigen* des Handys.

Blickgestützte Multimodale Interaktion

Auch wenn die Augen nicht für Manipulation und Interaktion (wie z.B. die Hände) geschaffen sind, so haben sie doch ein faszinierendes Potenzial, da der Blick typischerweise den Handlungen vorausgeht und Augenbewegungen implizite Positionsinformationen liefern. Gleichzeitig sind Nachteile – wie physiologische Beschränkungen und Tracking-Ungenauigkeiten – auszugleichen und das Midas-Touch-Problem zu adressieren. Wenn z.B. eine Selektion durch längeres Ansehen eines Objektes vorgenommen werden soll (*Dwell-time Selection*), dann kann dies auch unbeabsichtigt passieren, nur weil der Blick zu lange auf einer Stelle verweilt hat.

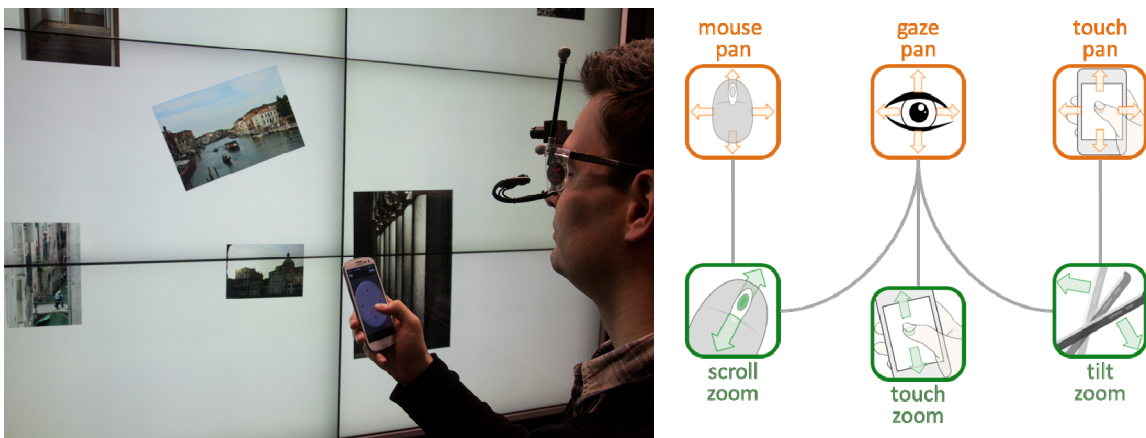


Abb. 2 Blickgestützte multimodale Interaktion. Das auf der Wand betrachtete Bild kann mit einfachen Touchgesten selektiert und manipuliert werden (li.) [8]. Verschiedene Kombinationen von Augensteuerung mit ergänzenden Modalitäten für Navigationsaufgaben (re.) [7].

Eine elegantere Lösung, die dieses Problem vermeidet und auch Ungenauigkeiten kompensiert, ist die multimodale blickgestützte Interaktion (*gaze-supported interaction*) [7,8]. Dabei wird der Blick grundsätzlich zur schnellen, teilweise initialen Positionierung genutzt, während ergänzende Interaktionsmodalitäten eine Verfeinerung erlauben, eine Auswahl bestätigen oder eine andere Aktion auslösen. So haben wir in [8] z.B. ein touchfähiges Handy in Ergänzung zum Blick genutzt, der mit einem mobilen Eyetracker verfolgt wird. Nach Ansehen der gewünschten Zielposition auf dem entfernten Display kann durch Fingerbewegung auf dem Touchdisplay dann eine Feinpositionierung des Cursors vorgenommen werden (s. Abb. 2). Neben dieser Technik wurde auch ein komplettes Set von Interaktionstechniken entwickelt, das neben der Selektion auch die Positionierung und Manipulation von entfernten Objekten erlaubt [8].

Blick kann auch mit anderen Modalitäten kombiniert werden. In [7] wurde in einer systematischen Studie untersucht, wie Augensteuerung mit einem Musrad, mit Touchinteraktion auf einem Handy und mit Handy-Neigegesten für Navigationsaufgaben, d.h. Zoom & Pan, genutzt werden kann (vgl. Abb. 2, rechts). So hat sich

beispielsweise das *Gaze-Pivot-Zooming*, d.h. das Anschauen der Zielposition in Kombination mit einer Zoommodalität, wie Neigen oder Touchgesten auf einem Smartphone, als intuitiv und schnell herausgestellt.

Tangible Magic Lenses & Tangible Windows

Wie Abb. 3 zeigt, kann auch der Raum vor einem Display oder über einem Tabletop genutzt werden, um Informationsräume darin zu explorieren [6]. Durch Bewegung einer in der Hand gehaltenen, lagebewussten magischen Linse kann z.B. an der entsprechenden Stelle des Mikroskopbildes hineingezoomt werden. Touch- und Stiftinteraktion erlauben zusätzlich Annotationen und Moduswechsel. In [5] haben wir beschrieben, wie neben der Lage der *Tangible Magic Lens* auch die Kopfposition und -orientierung des Nutzers erfasst wird. Damit und durch entsprechende perspektivische Korrekturen lassen sich auch dreidimensionale Inhalte glaubwürdig durch oder auf der Linse betrachten, die ein Fenster in die Virtualität darstellt. Auch hier steht Touch zur Verfügung.

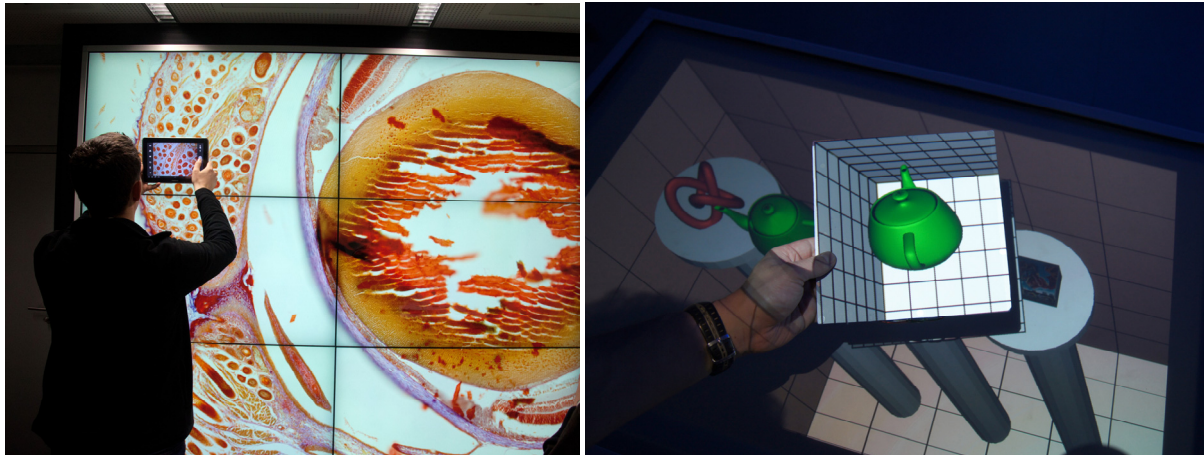


Abb. 3 Eine lagebewusste Linse als Fenster in einen virtuellen Informationsraum vor dem Display. Zoomen durch Bewegung des Displays an der jeweiligen Stelle (li.) und 3D-Objektinspektion durch zusätzliches Kopftracking (re.) [5].

Herausforderungen und Ausblick

Aus Sicht der Informatik ergeben sich mehrere zentrale Herausforderungen für die Interaktion mit großen Displays. Das qualitativ hochwertige und verzögerungsfreie/-arme Rendering komplexer Informationsräume im Megapixel-Bereich benötigt häufig mehrere Grafikkarten, Rechnercluster und dedizierte verteilte Architekturen. Die Interaktion mit großen Datenmengen, z.B. schnelle Ansichtsänderungen durch Zoomvorgänge oder die interaktive Parameteränderung in komplexen Simulationen, stellt zusätzlich hohe Ansprüche an ausreichende Bildwiederholraten und geringe Latenz. Ebenfalls zentral für die meisten der oben genannten Interaktionsmodalitäten ist die Herausforderung des Trackings. Im einfachsten Fall müssen nur Multitouch-Gesten erkannt werden; das gleichzeitige Erfassen von Augen-, Hand- oder Körperbewegungen für mehrere Nutzer ist erheblich aufwändiger. Steuergesten müssen zuverlässig erkannt werden, wobei z.B. Verfahren des Machine Learnings zum Einsatz kommen. Auch aus Sicht der Usability ergeben sich wesentliche Herausforderungen, z.B. bezüglich der Eignung und Erlernbarkeit von Gesten, ihrer Erinnerung, der Ermüdungsfreiheit, der kollaborativen Nutzung oder der effektiven Kombination von Modalitäten. Informelle Benutzerstudien und kontrollierte Experimente mit gründlicher statistischer Auswertung dienen u.a. der methodischen Überprüfung von Usability und User Experience.

Der Beitrag hat Motivation, Formen, Anwendungen und Herausforderungen der multimodalen Interaktion mit großen Displays beleuchtet. Es ist deutlich geworden, dass es sich um ein komplexes Forschungsgebiet mit vielen spannenden Fragestellungen und großem Anwendungspotenzial handelt.

Referenzen

1. Dachsel, R. Buchholz, R.: Throw and Tilt - Seamless Interaction across Devices Using Mobile Phone Gestures. GI Jahrestagung (1), 2008, 272-278
2. Frisch, M., Dachsel, R.: Kombinierte Multi-Touch und Stift-Interaktion: Ein Gesten-Set zum Editieren von Diagrammen. In: Schlegel, T. (Hrsg.): Multi-Touch: Interaktion durch Berührung, S. 89-116, Berlin, Heidelberg: Springer 2013

3. Jacob, R.J.K. et al.: Reality-based interaction: a framework for post-WIMP interfaces. In Proc. CHI '08, ACM (2008), 201-210
4. Nancel, M., Wagner, J., Pietriga, E., Chapuis, O., Mackay, W.: Mid-air pan-and-zoom on wall-sized displays. In Proc. CHI '11, ACM (2011), 177-186
5. Spindler, M., Büschel, W., Dachzelt, R.: Use your head: tangible windows for 3D information spaces in a tabletop environment. In Proc. ITS '12, ACM (2012), 245-254
6. Spindler, M., Büschel, W., Winkler, C., Dachzelt, R.: Tangible Displays for the Masses: Spatial Interaction with Handheld Displays by Using Consumer Depth Cameras. *Personal and Ubiquitous Computing*. Springer 2013.
7. Stellmach, S., Dachzelt, R.: Investigating gaze-supported multimodal pan and zoom. In Proc. ETRA '12, ACM (2012), 357-360
8. Stellmach, S., Dachzelt, R.: Still looking: investigating seamless gaze-supported selection, positioning, and manipulation of distant targets. In Proc. CHI '13, ACM (2013), 285-294