

---

# Erstellung, Suche und Vergleich von Phantombildern in der Augmented Reality

**Alexandra Krien**

Interactive Media Lab  
Technische Universität Dresden  
Dresden, Germany  
alexandra.krien@tu-dresden.de

**Maxime Thebault**

VP, Authoring  
Authorship Holdings, Ltd.  
Awdur SA22 8PP, UK  
Maxime.Thebault@insa-  
rennes.fr

**Heiner Ludwig**

Interactive Media Lab  
Technische Universität Dresden  
Dresden, Germany  
heiner.ludwig@tu-dresden.de

---

Paste the appropriate copyright statement here. ACM now supports three different copyright statements:

- ACM copyright: ACM holds the copyright on the work. This is the historical approach.
- License: The author(s) retain copyright, but ACM receives an exclusive publication license.
- Open Access: The author(s) wish to pay for the work to be open access. The additional fee must be paid to ACM.

This text field is large enough to hold the appropriate release statement assuming it is single spaced in a sans-serif 7 point font.

Every submission will be assigned their own unique DOI string to be included here.

**Abstract**

Phantombilder sind in der Polizeiarbeit ein unverzichtbares Medium bei der Suche nach Verdächtigen. Diese Arbeit beschäftigt sich damit, wie die dabei gängigen Abläufe in die Augmented Reality angehoben werden können um so die Bedienbarkeit und Nutzererfahrung zu verbessern. Fokus liegt dabei auf einer anfragebasierten, unscharfen Suche anhand der Phantombilder auf einer Datenbank von standardisierten Fotografien verschiedener Personen. Der Nutzer hat so die Möglichkeit seine zusammengestellten (Such-)Kriterien anhand reeller Menschen zu vergleichen und möglicherweise seine Anfrage auf der Suche nach dem besten Treffer anzupassen.

**Author Keywords**

Information Retrieval, Phantombilder, Augmented Reality

**Einführung**

Auf der Suche nach bestimmten Personen, beispielsweise im Zuge der Ermittlung einer Straftat, sind Phantombilder ein gängiges Hilfsmittel. Augenzeugen versuchen ihre Erinnerung unter Anleitung eines Beamten zu rekonstruieren. Auch wenn die Erstellung solcher Phantombilder bereits auf eine digitale Ebene angehoben wurde, ist der Prozess dennoch bisher sehr statisch. Die skizzenhaft zusammengestellten Merkmale enthalten möglicherweise große Ungenauigkeiten, Nutzer haben keine Möglichkeit

ihr Ergebnis mit reellen Menschen abzugleichen. Hinzu kommt der große Einfluss eines anwesenden Beamten. Der Zeuge muss seine Erinnerungen sehr genau und eindeutig beschreiben, um auf der gleichen Ebene dessen zu kommunizieren, da beispielsweise die Beschreibung einer "hellen Haut" unterschiedlich aufgefasst werden kann.

Die Idee hinter dieser Arbeit ist es Phantombilder als Anfrage für ein datenbankbasiertes System zu nutzen. Dies bietet die Möglichkeit die skizzenhafte Anfrage mit reellen Menschen zu vergleichen. Personen fallen so möglicherweise eher Fehler oder Ungenauigkeiten auf und haben die Möglichkeit ihre Anfrage iterativ so weit zu verfeinern, dass die Ergebnismenge der verdächtigen Personen eher ihrer Erinnerung entspricht. Diese Menge ist eineindeutig.

Ziel soll ein intuitives System sein, welches nach einer Einführung selbstständig von Personen ohne Vorwissen eingesetzt werden kann. Um dem Nutzer einen persönlichen Bearbeitungsraum zu bieten und an die reelle Begegnung anzuknüpfen, wurde entschieden dieses System in die Augmented Reality anzuheben. Gesten sollen dabei als einfache Interaktionsform dienen.

## **Verwandte Arbeiten**

### *Erstellung von Phantombildern*

Die klassische Erstellung von Phantombildern findet im Rahmen einer Zeugenaussage auf dem Revier statt. Die Umgebung kann dabei Auswirkungen auf die Genauigkeit der Aussage haben. Erfahrungen zeigten, dass insbesondere anwesende Beamte Einfluss auf die Ergebnisse nahmen. Ihre Aufgabe ist es die Zeugen bei der Rekonstruktion des Gesehenen zu unterstützen. Soziale Faktoren entscheiden dabei darüber wie viele Informationen tatsächlich preisgegeben und ob sie korrekt aufgenommen werden. So können unterschiedliche Vorstellungen und Auffas-

sungen zur Ausprägung von Merkmalen zu Abweichungen führen. Alle beteiligten Personen müssen demnach für ein gutes Ergebnis die gleiche deskriptive Sprache sprechen. [3]

Dieser klassische Prozess wurde bereits in der Vergangenheit versucht auf eine digitale Ebene anzuheben. So existieren eine Vielzahl an einfachen Desktoplösungen, die ein Zusammenstellen verschiedener optischer Merkmale zu einem Phantombild erlauben. Die konkrete Suche aufgrund eines veränderbaren Phantombildes hingegen ist unseren Recherchen zu Folge noch nicht so verbreitet. Die wenigen existierenden Systeme brachten aber bereits Erkenntnisse darüber, wie ein moderner Erstellungsprozess ablaufen könnte und welche Faktoren entscheidend sind, um durch die Anfrage auch signifikante Ergebnisse zu erhalten.

Unterschiede zeigten sich insbesondere durch die Auswahl der kombinierbaren Merkmale für die Suchanfrage. "SpotIt!" beispielsweise ermöglicht die Auswahl mehrerer Ausprägungen für ein Merkmal und interpoliert anschließend zwischen diesen um die optimale Lösung zu finden. [2] Ein anderes Personenidentifikationssystem hingegen zeigte, dass bereits eine Unterscheidung des Gesichts in 3 Bereiche reicht, um in der Suche eine Trefferquote von über 90% zu erreichen. Wichtigster Faktor ist dabei die Ausprägung der Augen. [1]

### *Gesten*

Gestik als Eingabemedium zur Erstellung von Phantombildern ist gänzlich neu, der Aufbau einer geeigneten Gestensprache demnach unumgänglich.

Das Buch Emotionales Interaktionsdesign beschäftigt sich u.a. mit der Semantik von Gestensprachen und zeigt Notwendigkeiten wie Abbruch-Gesten oder die allgemeine Robustheit von Gesten auf [4]. Kaisa Väänänen und Klaus Böhm weisen

auf Schwierigkeiten wie Ermüdung der Gliedmaßen und eingeschränkte Exaktheit während der Gesten-Nutzung hin, betonen jedoch gleichzeitig die vereinfachte Navigation und Manipulation für verschiedenste Nutzergruppen in virtuellen 3D-Räumen [6]. In [8] ist laut einer Nutzerstudie die Metaphern-Nutzung bei Gesten weniger effektiv als die Nutzung von Freihand-Gesten.

Gemäß [7,8] ist allgemein die Nutzung von formalisierten Gesten ("semaphoric gestures") nicht sinnvoll, da sie während der Interaktion unnatürlich wirkt. [8] besagt jedoch, dass beim Großteil der gestenbasierten Eingabesysteme sowohl formalisierte, als auch Freihand-Gesten genutzt werden.

Nicht zuletzt haben Hollywood-Filme wie *Iron Man* [9] oder *Minority Report* [5] zur Inspiration für die User Experience und bei der Erstellung der Nutzungsoberfläche gedient.

Zur Kombination von Virtueller Realität mit Gestenmanipulation haben die Firmen *eyeSight*<sup>1</sup> und *pebbles*<sup>2</sup> unabhängig voneinander eine VR-Brille mit mehreren Kameras kombiniert um die Interaktion mit dem virtuellen Raum für den Nutzer möglich zu machen.

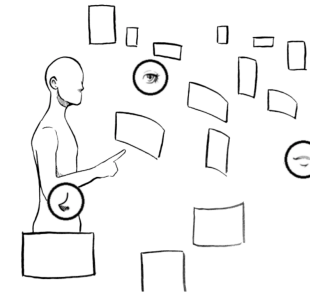
## Konzept

Das Konzept zur digitalen Erstellung von Phantombildern ist ein digitales System, welches den Nutzer in eine virtuelle Realität versetzt, die mittels Gesten manipulierbar ist. Durch die Interaktion mit suchanfragen- und ergebnisrepräsentierenden Objekten steuert der Nutzer die Erstellung des Phantombildes, indem er diese Objekte auswählt, verschiebt, löscht oder kombiniert.

Der Nutzer navigiert das System von einem Punkt aus. Um einzelne Objekte genauer zu betrachten, bewegt er sich

<sup>1</sup>[eyesight-tech.com/vr-ar.html](http://eyesight-tech.com/vr-ar.html)

<sup>2</sup>[www3.oculus.com/en-us/blog/pebbles-interfaces-joins-oculus/](http://www3.oculus.com/en-us/blog/pebbles-interfaces-joins-oculus/)



**Figure 1:** Virtuelle Objekte im System

also nicht körperlich, sondern verändert die Position des virtuellen Raums mittels Translations- und Skalierungsgesten.

Die Objekte sind im 180-Winkel "schwebend" um den Nutzer angeordnet (Figure 1). Durch sowohl freie, als auch objektorientierte Gesten, kann er, mittels ein- oder zweihändiger Gestikulation, Suchanfragen, Ergebnisse oder Ergebnisvorschläge speichern, löschen, zum Ergebnis-pool hinzufügen, neu kombinieren oder rückgängig machen. Das System wird immer nur von einem Nutzer gleichzeitig benutzt, weitere Personen können während des gesamten Findungsprozesses ausschließlich als Zuschauer involviert werden. So ist eine Moderatorenrolle eines Experten denkbar, der Hilfestellung gibt.

Aus der Sicht der Nutzbarkeit bietet das System folgende Vorteile:

- Effizientes und selbstständiges Arbeiten durch intuitive Gesten und Hilfestellung durch das System.
- Sehr direkte Interaktion vom Nutzer mit der Benutzungsoberfläche.

- Effektive Ergebnisfindung durch ständiges Feedback und Anpassung vom System.
- Nutzerbefriedigung durch immer sichtbaren Fortschritt.
- Einsatz ist ortsunabhängig und verlangt nur kleine Räumlichkeiten

## Systemnutzung

### *Systemstart und erste Suche*

Nachdem der Nutzer eine Virtual-Reality-Brille erhalten hat, kann er mit Hilfe der *smile*-Geste eine Übersicht öffnen, die in alle Gesichtsbereiche untergliedert ist. Um eine Suchanfrage zu starten, wählt der Nutzer ein Gesichtsmerkmal aus, indem er das entsprechende virtuelle Objekt mit einer Hand anfasst und mittig vor sich zieht. Sobald der Nutzer das Objekt loslässt, werden verschiedene Variationen des gewählten Gesichtsteils, wieder in Form von Objekten, angezeigt. Der Nutzer kann die passendste Ausprägung auswählen und das Merkmal so seiner Suchanfrage hinzufügen.

Das Anfragemenü kann anschließend mittels der umgekehrten Geste wieder geschlossen werden. Die Suche wird automatisch gestartet. Die einzelnen Anfrageobjekte platzieren sich gleichmäßig im Raum, die Ergebnisse befinden sich zwischen diesen. Dabei ist die räumliche Entfernung zwischen einzelnen Suchergebnissen direkt proportional zu deren Ähnlichkeit. Ergebnisse die sich besonders nah an einem Anfrageobjekt bewegen erfüllen dieses Kriterium besonders gut, Ergebnisse die allen Kriterien entsprechen, befinden sich entsprechend mittig.

### *Ergebnisexploration und -speicherung*

Durch die *navigation*-Geste kann der Nutzer den „Ergebnis-Raum“ translatieren, skalieren und rotieren. Ist bspw. ein Suchergebnis passend, kann die Menge ähnlicher Suchergebnisse vergrößert und die Granularität verfeinert werden wer-

den, indem sich der Nutzer virtuell zum Ergebnis-Objekt bewegt und eine Skalierung vornimmt, um sich die Ergebnisse diesen Typs genauer anzuschauen. Um Ergebnisse zu speichern werden diese per Zeige-Geste markiert und mit einer *pull*-Geste zum Nutzer gezogen. Sind Ergebnisse auszuschließen, können diese markiert und anschließend mit der *remove*-Geste entfernt werden. Soll die nächste Suchanfrage durch bereits gefundene Gesichtsmkmale konkretisiert werden, wird das Ergebnis-Objekt mit beiden Händen gefasst und mittig vor den Nutzer gezogen und so mit dem Suchanfrage-Objekt der vorherigen Suche verknüpft.

### *Weitere Suchanfragen*

Nach jeder weiteren Suche, kann das System anhand der vom Nutzer ausgewählten und gelöschten Ergebnis-Objekte zunehmend genauere Vorschläge für die folgenden Suchanfragen aufstellen. So ist bspw. die Wahrscheinlichkeit, dass eine blauäugiger Mensch eher einen (nord-)europäischen Hauttyp besitzt höher, als bei braunäugigen Menschen [7].

### *Rückgängig machen und Verlauf hervorheben*

Um eine Suchanfrage oder Objektmanipulation rückgängig zu machen, kommt die *undo*-Geste zum Einsatz.

Sollen multiple Arbeitsschritte widerrufen werden, wird die *erweiterte undo*-Geste genutzt: Während die eine Hand gehalten wird, erscheint vor dem Nutzer eine virtuelle Zeitleiste, anhand der im Arbeitsverlauf vor- und zurückgesprungen werden kann. Dazu wird die Hand horizontal nach links bzw. rechts bewegt.

### *Speichern und Ergebnis*

Wie bereits unter Abschnitt *Ergebnisexploration und -speicherung* erwähnt können einzelne Suchergebnisse jederzeit über die *pull*-Geste gespeichert werden. Mittels der *down*-Geste

öffnet sich eine Übersicht aller bereits gespeicherten Ergebnisse. Hier befindliche Objekte können jederzeit wieder entfernt werden oder auch nachträglich als Anfrageobjekte eingesetzt werden.

Beendet der Nutzer das System, werden die zu diesem Zeitpunkt gespeicherten Ergebnisse als finales Ergebnis übernommen. Dieses ist auch nach Beenden der Session vom Beamten einsehbar und kann für die weiteren Ermittlungen eingesetzt werden. Wurde das System von mehreren Zeugen genutzt hat der Beamte so die Möglichkeit die Ergebnisse zu vergleichen und auf Überschneidungen der Ergebnismengen zu überprüfen. Die Ergebnisse können ebenfalls einer späteren Session hinzugefügt werden. Der Beamte kann die Auswertung so direkt im System vornehmen.

### **Gesten**

Gesten sollen logisch, intuitiv einfach zu erlernen und ergonomisch auszuführen sein [8]. Da das System hauptsächlich bei Laien zum Einsatz kommt, ist es von hoher Wichtigkeit, eingängige Gesten zu finden, um eine gute Nutzbarkeit des Systems zu garantieren. Um die eindeutige Interpretation auf technischer Seite und die Unverwechselbarkeit der Gesten Nutzer-seitig zu gewährleisten, sollen Gesten robust sein, sich also klar voneinander unterscheiden [8, 4]. Des Weiteren benötigt jede Geste eine Abbruch-Möglichkeit [4]. Im folgenden Teil werden alle System-Gesten aufgeführt: (Abb. 2)

smile-Geste: Öffnen bzw. Schließen des Such-Menüs

move-Geste: Navigation durch den virtuellen Raum

select-Geste: Auswählen von Such- oder Ergebnis-Objekten

add-Geste: Hinzufügen eines Ergebnis-Objektes zur Suche

remove-Geste: Löschen eines ausgewählten Objektes

undo-Geste: Rückgängig machen des letzten Schritts

erweiterte undo-Geste: Aufzeigen und Steuern des Arbeitsverlaufs

redo-Geste: Wiederholen des letzten Arbeitsschritt

pull-Geste: Speichern der Ergebnisse

down-Geste: Öffnen der gespeicherten Ergebnisse

### **Diskussion**

Sowohl die computergestützte Phantombild-Erstellung, als auch die Gestik als Eingabemedium wurden separat vielfach wissenschaftlich untersucht. Die Kombination beider Themengebiete ist neu und bietet interessante Möglichkeiten zum Aufbau einer innovativen Nutzungsoberfläche.

Der stark reglementierte Prozess zum stufenweisen Aufbau eines Phantombildes ist für die Gesten-Eingabe gut geeignet, da keine Notwendigkeit für eine „freie Suche“ besteht und somit bspw. auf Text-Eingaben durchweg verzichtet werden kann. Das erlaubt es, das System sehr einfach zu halten.

Die Gesten bestehen nur aus wenigen Bewegungen, zumeist nur mit einer Hand. Dadurch wird der Ermüdung der Gliedmaßen des Nutzers entgegengewirkt, die simplen Abläufe sind schnell zu erlernen [6]. Da Suchanfragen und Ergebnisse prinzipiell als einfache Objekte dargestellt sind, ergibt sich die Möglichkeit des Einsatzes vieler Freihand-Gesten [8].

Der Bezug zur Realität (“Objekt nehmen und verschieben”) erleichtert die Verständlichkeit für den Nutzer. Zusätzlich bieten Objekte eine Große Interaktions-Oberfläche, wodurch

die fehlende Exaktheit von Gesten kompensiert wird [6].

Prinzipiell sollte sich die Frage gestellt werden, ob das Verstehen einer neuen Nutzungsoberfläche, oder das Erlernen von Gesten aufwändiger ist. Im Falle dieser Untersuchung konnte die Notwendigkeit von komplexen Gesten nur durch den unkonventionellen Aufbau der virtuellen Oberfläche beschränkt werden. So wird das WIMP-Paradigma aufgeweicht, welches möglicherweise zur Verunsicherung bei (routinierten) Computer-Nutzern führt. Bspw. wurde auf die gewöhnliche Fensteranordnung oder einen Pointer verzichtet, jedoch spielt die räumlich Entfernung von Objekten eine wichtige Rolle.

Zusammenfassend profitiert der Nutzer also nur dann von den einfachen Gesten, wenn er das Prinzip des Systems verinnerlicht hat.

### Schlussfolgerung

Die Phantombilderstellung scheint eine bislang nicht ausgereizte Domäne zu sein. Zwar widmen sich viele wissenschaftliche Untersuchungen der technischen Umsetzung zur Gesichtserkennung, im Bereich Nutzungsoberflächen gibt es jedoch Nachholbedarf.

Mit dem gestenbasierten System zum stufenweisen Aufbau von Phantombildern wird ein Konzept vorgestellt, welches die Möglichkeiten von Gestennutzung in einer innovativen Augmented-Reality-Oberfläche aufzeigt.

Durch den iterativen Suchprozess, wird die Suche automatisch immer weiter verfeinert. Indem verschiedene Arten von Gesichtsteilen vom Nutzer ausgewählt und u.U. zu neuen Suchanfragen zusammengefügt werden, ergibt sich ein Phantombild, welches mit jeder Ergebnisauswahl konkreter wird. Der virtuelle Arbeitsraum in Kombination mit der Gestennutzung soll hierbei den Nutzer unterstützen,

indem es die Exploration großer Datenmengen mit einer neuartigen Oberfläche und einfacher Bedienung kombiniert.

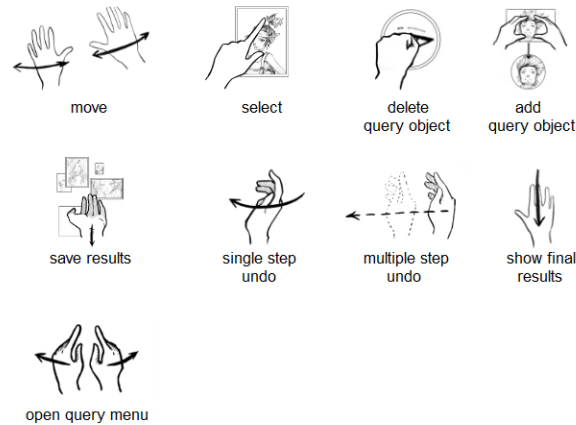
Im Idealfall ist das System so direkt zu bedienen, dass der Nutzer sich ausschließlich auf die Inhalte konzentriert, nicht auf die Bedienung. Da im Szenario Menschen im Vordergrund stehen, die das System sehr selten nutzen, wurde großer Wert auf die Verständlichkeit des Systems und die allgemeine Erlernbarkeit der Gesten gelegt.

Allgemein soll das Konzept zeigen, dass komplexe Computersysteme mit Hilfe neuer Bedienungs-Technologien neue Ansätze hinsichtlich der Nutzbarkeit bieten. Es gilt nun zu evaluieren, ob es grundsätzlich funktioniert und welche Verständnisprobleme auftreten.

Eine künftige Ergänzung könnte der Aufbau eines dreidimensionalen Phantombildes darstellen, das die plastische Form des Gesichts verdeutlichen würde. Dafür müssten die entsprechenden Datenbanken im Back-End jedoch 3D-Scans der Gesichter aller Personen bereit stellen, um einen Abgleich starten zu können.

### REFERENCES

1. Janusz Bobulski and Mariusz Kubanek. 2012. Person identification system using an identikit picture of the suspect. *Optica Applicata* 42, 4 (2012), 865–873.
2. Roberto Brunelli and Ornella Mich. 1996. SpotIt! an interactive identikit system. *Graphical Models and Image Processing* 58, 5 (1996), 399–404.
3. Helmut Buchholz. 2012. Wie die Polizei Phantombilder erstellt. (2012). <http://www.stimme.de/heilbronn/nachrichten/region/sonstige-Wie-die-Polizei-Phantombilder-erstellt;art16305,2541477>



**Figure 2:** Gesten

4. Rainer Dorau. 2011. *Emotionales Interaktionsdesign*. Springer Vieweg, Heidelberg. DOI : <http://dx.doi.org/10.1007/978-3-642-03101-4>

5. 20th Century Fox Dream Works Pictures. 2002. *Minority Report*. cinematic movie. (3 October 2002).
6. R. A. Earnshaw. 1993. *Virtual Reality Systems*. Academic Press.
7. Peter Frost. 2006. European hair and eye color A case of frequency-dependent sexual selection? *Evolution and Human Behavior* 27 (2006). <http://www.sciencedirect.com/science/article/pii/S1090513805000590>
8. Joao Rodrigues. 2016. *Handbook of Research on Human-Computer Interfaces, Developments, and Applications (Advances in Human and Social Aspects of Technology)*. IGI Global. <http://www.igi-global.com/book/handbook-research-human-computer-interfaces/146921>
9. Marvel Studios. 2010. *Iron Man 2*. cinematic movie. (6 May 2010).