

# Linking physical and virtual worlds with visual markers and handheld devices

**Doctoral Thesis**

**Author(s):**

Rohs, Michael

**Publication date:**

2005

**Permanent link:**

<https://doi.org/10.3929/ethz-a-005067344>

**Rights / license:**

[In Copyright - Non-Commercial Use Permitted](#)

Diss. ETH No. 16193

# Linking Physical and Virtual Worlds with Visual Markers and Handheld Devices

A dissertation submitted to the  
SWISS FEDERAL INSTITUTE OF TECHNOLOGY ZURICH

for the degree of  
Doctor of Sciences

presented by  
MICHAEL ROHS  
Diplom-Informatiker, Darmstadt University of Technology  
born January 13, 1974  
citizen of Germany

accepted on the recommendation of  
Prof. Dr. Friedemann Mattern, examiner  
Dr. Albrecht Schmidt, co-examiner

2005

# Abstract

Linking the physical and the virtual world is a major research theme of ubiquitous and pervasive computing. This dissertation describes concepts and techniques for linking information and services to physical objects as well as for interacting with this information using mobile devices and embodied user interfaces. Such interfaces use gestures on the device body as a means of input. In the recent past, there have been considerable research efforts in linking computation to physical objects. However, these projects were mainly concerned with the physical linking technology per se or with the infrastructure required for identifier resolution. Other work on manipulative and embodied user interfaces focused on improving interaction with a handheld device itself, but did not integrate physical objects of the user's environment. In our work, we combine physical linking and embodied interaction and allow the interaction semantics to be a function of the object and the gestural sequence.

The proposed approach uses camera phones and similar devices as mobile sensors for two-dimensional visual markers. We not only retrieve the value that is encoded in the marker, but also detect the spatial orientation of the device relative to the marker in real time. We use the detected orientation for embodied interaction with the device and augment the live camera image according to the orientation with graphical overlays. By providing a video see-through augmented reality view on the background, the handheld device embodies a “symbolic magnifying glass.” This allows for fine-grained interaction and enhances the currently limited input capabilities of mobile devices. We call this approach *marker-based interaction*. It turns camera phones and similar devices into versatile interfaces to – and mediators for – real-world objects.

In this thesis, we present a system for recognizing two-dimensional visual markers. The markers we developed are called *visual codes*. The recognition system provides a number of parameters for determining the spatial orientation of the device relative to the marker, such as the target point in code coordinates, rotation, tilting, distance, and movement of the device relative to the background. It is specifically designed for the requirements of mobile phones with limited computing capabilities and low resolution cameras. Moreover, the system provides the basis for augmenting objects in the live camera image with precisely aligned graphical overlays. Based on this foundation we have developed several mechanisms and concepts for marker-based interaction, namely: (1) a framework of physical interaction primitives, (2) marker-based interface elements, called *visual code widgets*, (3) interaction techniques for large-scale displays, and (4) handheld augmented reality applications.

Our conceptual framework of physical interaction primitives enables the use of camera-equipped mobile devices as embodied user interfaces, in which users

can specify input through physical manipulations and orientation changes of the device. The framework defines a set of fundamental physical gestures that form a basic vocabulary for describing interaction when using mobile devices capable of reading visual codes. These interaction primitives can be combined to create more complex and expressive interactions. The interaction primitives and their combinations have been evaluated in a usability study.

In comparison to interaction primitives, visual code widgets operate at a higher level of abstraction. Visual code widgets are printable elements of physical user interfaces, comparable to the interactive elements of conventional graphical user interfaces. Each widget type addresses a particular input problem and encapsulates a specific behavior and functionality. Visual code widgets thus define building blocks for applications that incorporate mobile devices as well as resources in the user's environment, such as paper documents, posters, and public electronic displays.

For large-scale displays, we have developed two interaction techniques that rely on visual movement detection and visual code recognition, respectively. The first one enables relative positioning of a cursor and is suited for direct manipulation of objects that are visible on the screen. The second one allows for absolute positioning on the screen and can be used for the selection of displayed objects. Both techniques have been evaluated in a qualitative usability study and are especially useful for displays that are not available for direct touch-based interaction, such as displays in public spaces.

The concepts and techniques that were developed in the scope of this dissertation have been investigated in various application areas. Examples that are detailed in the dissertation are: entry points into a smart campus environment, augmented board games, an interactive photo wall, a collaborative game for large-scale displays, digital annotations of physical objects, and smart product packaging.

# Zusammenfassung

Die Verbindung der physischen mit der virtuellen Welt ist ein zentrales Forschungsthema im Bereich des Ubiquitous und Pervasive Computing. Die vorliegende Dissertation beschreibt Konzepte und Techniken zur Verknüpfung von physischen Objekten mit Information sowie zur Interaktion mit diesen Informationen mittels mobiler Geräte und gestenbasierter Benutzungsschnittstellen. In jüngster Vergangenheit ist die Verknüpfung von physischen Objekten mit digitalen Informationen bereits intensiv erforscht worden. Vorausgegangene Arbeiten haben sich allerdings hauptsächlich mit der Verknüpfungstechnologie oder mit Infrastrukturaspekten beschäftigt. Andere Arbeiten haben sich der Verbesserung der Interaktion mit mobilen Geräten durch physische Manipulationen gewidmet, jedoch keine Objekte in der Umgebung des Benutzers mit einbezogen. In der vorliegenden Arbeit werden physische Verknüpfungen und gestenbasierte Benutzungsschnittstellen kombiniert, wodurch die Interaktionssemantik eine Funktion des Objektes und der physischen Handhabung des mobilen Gerätes wird.

Der vorgeschlagene Ansatz nutzt Mobiltelefone und ähnliche Geräte mit integrierter Kamera als mobile Sensoren für zweidimensionale visuelle Marker. Es werden nicht nur die im Marker gespeicherten Daten, sondern auch die räumliche Orientierung des Gerätes relativ zum Marker in Echtzeit ermittelt. Die erkannte Orientierung wird zur physischen Interaktion mit dem Gerät genutzt und das Kamerabild entsprechend der Orientierung mit grafischen Ausgaben überlagert. Das Gerät liefert eine um Informationen erweiterte Sicht auf den mit der Kamera fokussierten Bereich und implementiert damit die Metapher einer „symbolischen Lupe“. Dies ermöglicht eine fein-granulare Interaktion und erweitert die bisher eingeschränkten Eingabemöglichkeiten von mobilen Geräten. Wir bezeichnen diesen Ansatz als *Marker-basierte Interaktion*. Er verwandelt mobile Geräte in vielseitig verwendbare Schnittstellen und Mediatoren für Objekte der realen Welt.

In der vorliegenden Arbeit wird ein System zur Erkennung von zweidimensionalen visuellen Markern vorgestellt. Die von uns entwickelten Marker bezeichnen wir als *Visual Codes*. Das Erkennungssystem liefert verschiedene Parameter zur Ermittlung der räumlichen Orientierung der Kamera relativ zum Marker, wie den anvisierten Punkt in der Ebene des Markers, die Rotation, die Neigung, die Distanz und die Bewegung des Gerätes relativ zum Hintergrund. Das System wurde speziell für die Anforderungen von Mobiltelefonen mit eingeschränkter Rechenkapazität und gering auflösender integrierter Kamera entworfen. Es dient darüber hinaus als Basis für die Erzeugung grafischer Einblendungen im Kamerabild, die einzelne Bildelemente geometrisch präzise überlagern. Aufbauend auf dieser Basis wurden verschiedene Mechanismen und Konzepte für Marker-basierte Interaktion entwickelt: (1) ein System kombinierbarer physischer Interaktionsprimitive, (2) Marker-basierte Bausteine für physische Benutzungsschnittstellen, sogenannte

*Visual-Code-Widgets*, (3) Interaktionstechniken für Grossbildschirme und (4) mobile Augmented-Reality-Anwendungen.

Das System kombinierbarer physischer Interaktionsprimitive erlaubt die Nutzung mobiler Geräte mit integrierter Kamera als *embodied user interfaces*, bei denen der Benutzer Eingaben mit Hilfe von physischen Manipulationen und Orientierungsänderungen des Gerätes vornehmen kann. Die Interaktionsprimitive definieren dabei eine Menge von grundlegenden physischen Gesten, die ein Basis-Vokabular zur Interaktion mit visuellen Markern mittels Kamera-Mobiltelefonen und ähnlichen Geräten darstellen. Die Interaktionsprimitive können kombiniert werden, um komplexere und ausdrucksstärkere Interaktionen zu ermöglichen. Die Eigenschaften der verschiedenen Interaktionsprimitive und ihrer Kombinationen wurden anhand einer Nutzerstudie evaluiert.

Im Vergleich zu den Interaktionsprimitiven befinden sich die *Visual-Code-Widgets* auf einer höheren Abstraktionsstufe. *Visual-Code-Widgets* sind auf Papier druckbare Elemente physischer Benutzungsschnittstellen, vergleichbar den interaktiven Elementen herkömmlicher grafischer Benutzungsschnittstellen. Jeder Widget-Typ definiert ein bestimmtes Eingabeparadigma, kapselt eine bestimmte Funktionalität und zeigt gegenüber dem Benutzer ein charakteristisches Verhalten. *Visual-Code-Widgets* stellen damit Grundbausteine für Anwendungen zur Verfügung, die sowohl mobile Geräte, als auch Ressourcen in der Umgebung des Benutzers, wie Papierdokumente, Poster und Grossbildschirme, mit einbeziehen.

Für Grossbildschirme wurden zwei Interaktionstechniken entwickelt, die auf optischer Bewegungserkennung und der Erkennung von visuellen Markern basieren. Die erste Technik erlaubt die relative Positionierung eines Cursors auf dem Grossbildschirm und ist für die direkte Manipulation von angezeigten Objekten geeignet. Die zweite Technik ermöglicht die absolute Positionierung auf dem Bildschirm und kann zur Selektierung von Bildschirm-Elementen verwendet werden. Die entwickelten Techniken wurden im Rahmen einer qualitativen Nutzerstudie untersucht und sind besonders für die Anwendung im öffentlichen Raum geeignet, wie z.B. in Bahnhöfen, Flughäfen oder Museen.

Die im Rahmen dieser Dissertation behandelten Konzepte und Techniken wurden in verschiedenen Anwendungsgebieten prototypisch untersucht. Beispiele, die in dieser Arbeit vorgestellt werden, sind: Einstiegspunkte in einen virtuellen Campus, virtuell erweiterte Brettspiele, eine interaktive Foto-Wand, ein kollaboratives Spiel für Grossbildschirme, die digitale Annotation physischer Objekte und smarte Produktverpackungen.