

## 2 Augmented Reality als Querschnittstechnologie für Bildverarbeitung, Geoinformation und Computergrafik

*Sven Wursthorn*

Unter Augmented Reality (engl. für erweiterte Realität, kurz AR) wird die Überlagerung der Sinneseindrücke, die ein Mensch von seiner Umgebung aufnimmt, mit zusätzlichen, vom Computer generierten Informationen in Echtzeit verstanden. Diese überlagerten virtuellen Inhalte fügen sich nahtlos in die vom Nutzer wahrgenommene Umgebung ein und beziehen sich direkt auf diese. Im Unterschied zur AR bewegt sich der Nutzer bei Anwendungen aus dem Bereich der virtuellen Realität in einer vollständig vom Computer generierten künstlichen Welt.

Eine Definition des Begriffs „Augmented Reality“ wurde bereits im Jahr 1992 in einer Veröffentlichung eines Projektes zur Unterstützung der Monteure beim Kabelverlegen in Flugzeugen bei Boeing gegeben (Caudell u. a., 1992). Milgram definiert unter dem Oberbegriff „Mixed Reality“ (Milgram u. a., 1994) den gesamten Bereich zwischen der realen Welt und der rein virtuellen Umgebung, der Virtual Reality (kurz VR).

Ein AR-System (ARS) sollte nach der allgemein anerkannten Definition von Azuma (1997) folgende drei Bedingungen erfüllen:

- Es werden reale und virtuelle, vom Computer generierte Objekte kombiniert.
- Das System ist interaktiv und operiert in Echtzeit.
- Die virtuellen Inhalte überlagern die realen Objekte in der Umgebung des Anwenders.

Durch die Überlagerung in Echtzeit steht eine leicht verständliche Repräsentation von raumbezogenen Datenbeständen zur Verfügung, die sofort mit der direkten Umgebung verglichen und auf Plausibilität geprüft werden können. Die beobachtenden Personen können Verbindungen zwischen der Realität und digitalen Daten herstellen, die ohne den Einsatz eines AR-Systems möglicherweise unerkannt geblieben wären.

Zu einer der ersten Arbeiten im Bereich AR kann man die von Sutherland aus dem Jahr 1968 über Prototypen für Head-Mounted-Displays zählen (Sutherland, 1968).

Dort wird erstmals die Überlagerung eines 3D-Modells eines Raumes mit dem realen Raum in einem experimentellen Hardware-Aufbau beschrieben. Erste Anwendungen außerhalb einer kontrollierten Laborumgebung wurden erst viel später entwickelt. Im „MARS“-Projekt (Mobile Augmented Reality Systems) wurde 1997 ein mobiles Outdoor-AR-System für die Fußgängernavigation entwickelt (Feiner u. a., 1997). Pfeile und ortsbezogene Textmarken mit Informationen zu Gebäuden und Hörsälen sollen den Nutzer bei der Navigation auf einem Campusgelände unterstützen. Derartige mobile Systeme mussten zu dieser Zeit auf einem Rucksack getragen werden. Ein ähnliches System, bei dem alle notwendigen Komponenten auf einem Rucksack montiert waren, hatte etwa 10 kg Gewicht (Wursthorn u. a., 2004). Heute sind spezielle AR-Systeme vollständig in Helme oder Brillen integriert. Normale Smartphones enthalten alle Komponenten, die für den Betrieb einer AR-Anwendung benötigt werden.

### 2.1 Komponenten eines AR-Systems

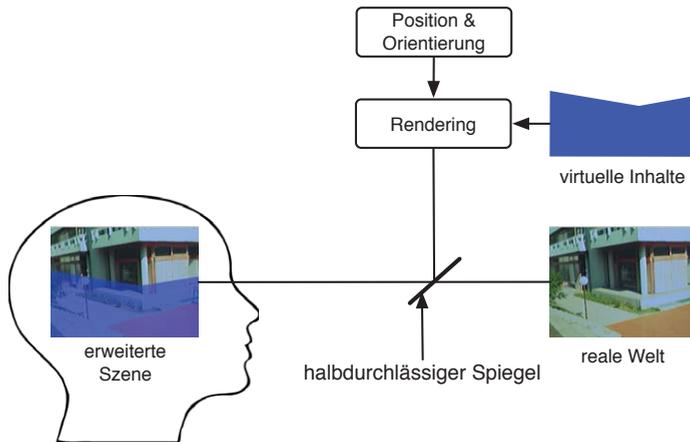
Die notwendigen Komponenten eines AR-Systems können in drei Gruppen aufgeteilt werden:

- Hardware zur Darstellung der zu überlagernden Inhalte.
- Sensoren und Verfahren zur Positionierung und Orientierung, die eine Überlagerung der raumbezogenen Information mit der Umgebung des Anwenders erlauben. Dazu zählen auch alle Verfahren, die für eine weitere Verschmelzung von virtuellen Objekten mit der realen Umgebung sorgen, wie z. B. die Simulation des Schattenwurfes, die Anpassung an die reale Umgebungsbeleuchtung oder etwa Spiegelungen.
- Raumbezogene Inhalte, die überlagert werden sollen.

Diese drei Gruppen werden im Folgenden näher erläutert.

#### 2.1.1 Displaysysteme

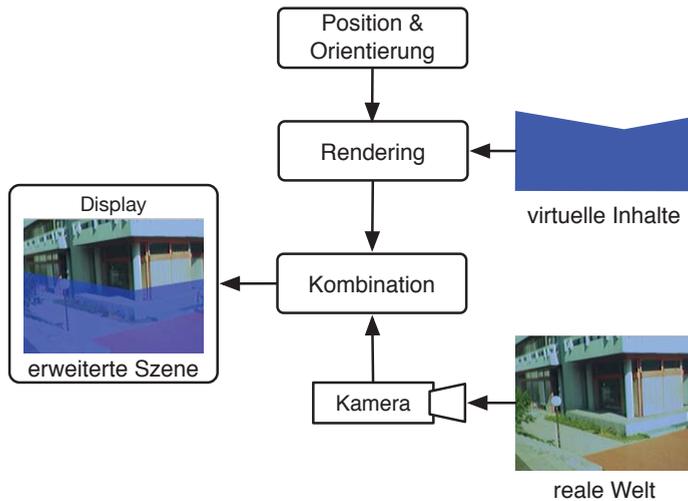
Smartphone- oder Tabletdisplays werden einfach in das Gesichtsfeld gehalten, um die Überlagerungen mit den Live-Bildern der Kamera der Geräterückseite sehen zu können. Diese am weitesten verbreitete Art der AR ist auch als „Magic Lens“ (Bier u. a., 1993, „magische Lupe“) bekannt, da das Smartphone in der AR-Anwendung ähnlich wie eine Lupe genutzt wird. Um bei mobilen Systemen die Hände frei zu haben, können Displays direkt vor dem Auge einer Person getragen werden. Diese Systeme werden allgemein als „Head-Mounted-Display“ (HMD), „Near-Eye-Display“ (NED)



**Abb. 2.1** AR mit See-Through-Displaysystem: Die virtuellen Inhalte werden anhand der Informationen zu Position, Orientierung und Abbildungseigenschaften der Brille entsprechend gerendert und über einen halbdurchlässigen Spiegel in das Gesichtsfeld projiziert. Auf diese Weise wird die Ansicht der realen Umgebung um virtuelles Hochwasser erweitert.

oder schlicht als „Smart Glasses“ bezeichnet. Die Displays sind bei „See-Through“-Systemen transparent. Die Sicht auf die reale Umgebung ist damit auch ohne Betrieb des Displays möglich. Die virtuellen Objekte können die realen Objekte bei den heute am Markt erhältlichen Brillen allerdings nicht verdecken. In Abbildung 2.1 ist der Ablauf für die Überlagerung virtueller Inhalte schematisch dargestellt. Die erweiterte Szene zeigt, wie der virtuelle Inhalt die Ansicht der realen Welt überlagert. In dieser Anwendung ist der virtuelle Inhalt ein 3D-Modell einer Hochwasserfläche, die ein reales Gebäude umfließt. Für das Erzeugen der Überlagerung müssen Position, Orientierung und Abbildungseigenschaften der Brille bekannt sein, damit die Ansicht des 3D-Modells der Hochwasserfläche für die erweiterte Szene entsprechend berechnet werden kann. Diese Berechnung wird in der Computergrafik als Rendering bezeichnet. Diese beispielhafte AR-Anwendung visualisiert somit mögliche Hochwasserszenarien direkt vor Ort, um Hochwasserschutzmaßnahmen entsprechend planen zu können.

Ein Beispiel für ein modernes See-Through-Display ist Microsofts HoloLens-Brille (Microsoft, 2017). Neben der Optik für die stereoskopische Überlagerung transparenter 3D-Inhalte im Gesichtsfeld des Nutzers sind in der fast 600 g schweren Brille alle Komponenten für den unabhängigen Betrieb integriert. Die Brille ist für den Innenbereich konzipiert. Im Freien, besonders bei Sonnenschein, ist die Lichtstärke des Displays zu schwach. Zudem enthält die HoloLens-Brille Kameras zur Bestimmung der Position und zur Orientierung im Raum, aber keinen GNSS-Sensor (Näheres zur Positionsbestimmung der HoloLens siehe Abschnitt 2.1.2: SLAM). Microsoft bezeichnet die virtuellen 3D-Objekte passenderweise als Hologramme, da die reale



**Abb. 2.2** Videobasiertes AR: Die virtuellen Inhalte werden in das Abbildungssystem der Kamera transformiert und mit der von der Kamera live erfassten Umgebung überlagert. Das Ergebnis kann nur auf einem Display betrachtet werden.

Ansicht hinter den Objekten systembedingt nicht verdeckt werden kann. Die Objekte erscheinen transparent. Es ist auch möglich, dass mehrere HoloLens-Nutzer kollaborativ im gleichen Raum mit Hologrammen arbeiten können. Für die Nutzerinteraktion stehen die Blickrichtung in Form eines fest im Zentrum des Gesichtsfeldes befindlichen Punktes, Handgesten zum Bestätigen von Aktionen, zum Scrollen oder zum Öffnen von Menüs und Sprachbefehle zur Verfügung. Für die bestmögliche Anpassung der HoloLens-Brille an eine bestimmte Person muss eine einfache Kalibrieranwendung zur Erfassung des Pupillenabstands durchlaufen werden.

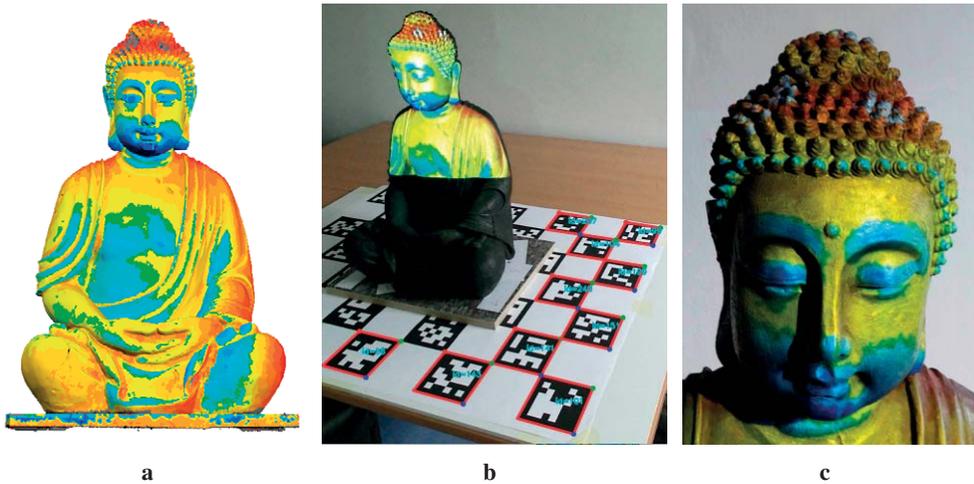
Bei videobasierten Displaysystemen („Video-See-Through“-Displays) wird die Umgebung live mit einer Videokamera erfasst. Der Computer kombiniert die Videobilder mit den Bildern, die er aus den virtuellen Inhalten passend zur Abbildungsvorschrift der Kamera gerendert hat (siehe Abbildung 2.2). Ein solches System kann auch mit einer Virtual-Reality-Brille, welche eine oder zwei nach vorne blickende Kameras besitzt, hergestellt werden. Das Ergebnis der Kombination von realen und virtuellen Objekten wird auf dem eingebauten Display angezeigt. Im Vergleich zu See-Through-Displays kann bei videobasierten Lösungen das Ergebnis der Überlagerung vollständig kontrolliert werden. So können z. B. virtuelle Objekte auch reale Gegenstände verdecken. Bei videobasierten AR-Systemen können bildbasierte Verfahren aus dem Bereich der digitalen Bildverarbeitung und der Computer Vision verwendet werden, um z. B. die Genauigkeit der Überlagerung zu verbessern. Wenn das Videobild sich zu sehr von der natürlichen Wahrnehmung unterscheidet oder es systembedingt zu

Verzögerungen bei der Anzeige kommt, dann kann dies zu Übelkeit bei Anwendern führen („operator sickness“). Eine zu geringe Auflösung des Kameradisplaysystems kann sich außerdem störend auf die Wahrnehmung auswirken, wenn z. B. reale Texte im Blickfeld nicht mehr gelesen werden können. Ein Systemausfall würde die Sicht komplett versperren und erhöht so die Verletzungsgefahr im mobilen Betrieb. Spezielle Vor- und Nachteile der AR-Systemarten HMD und Magic Lens werden von Bach u. a. (2018) diskutiert.

Außer mittels Smartphones und HMDs können Szenen und Objekte auch mithilfe eines oder mehrerer Projektoren mit projizierten Texturen erweitert werden. Dieses Prinzip wird von Bandyopadhyay u. a. (2001) auch als „shader lamp“ bezeichnet. Auf diese Weise können eine oder mehrere Personen gleichzeitig ohne Vorbereitung oder spezielle Hardware gemeinsam eine AR-Anwendung nutzen. Projektorbasierte AR wird im Buch „Spatial Augmented Reality“ von O. Bimber ausführlich behandelt (Bimber u. a., 2005).

Abbildung 2.3 zeigt ein Beispiel für eine projektorbasierte AR-Anwendung. Ein 3D-Modell einer Buddhastatue mit einer Textur, die farbcodierte Abweichungen zweier 3D-Messergebnisse dieser Statue zeigt (Abbildung 2.3a), wird direkt auf die reale Statue projiziert. Auf diese Weise können mehrere Personen gleichzeitig auf Anhieb sehen, an welchen Stellen am Objekt die größten Abweichungen auftreten. Die Registrierung von Objekt und Projektion wird mithilfe einer am Projektor befestigten Kamera und einer Platte mit codierten Zielmarken (sog. „Marker“, s. u.) durchgeführt. Position und Orientierung des Objekts zur Platte müssen bekannt sein. Das Projektor-Kamera-System befindet sich auf einem beweglichen Stativ, sodass die Überlagerung an beliebigen Stellen am Objekt gezeigt werden kann, solange sich eine ausreichende Anzahl von Zielmarken im Gesichtsfeld der Kamera befindet (siehe Abbildung 2.3b). Die Detailansicht des Kopfes der Statue in Abbildung 2.3c zeigt, wie genau die Überlagerung auf diese Weise sein kann (siehe Augenbereich und Punkt auf der Stirn).

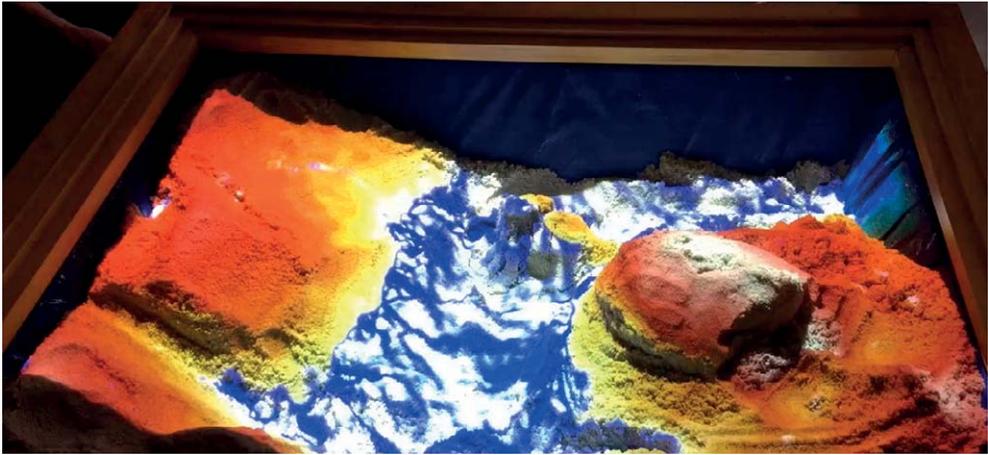
Ein weiteres Beispiel für ein projektorbasiertes Augmented-Reality-System ist die AR-Sandbox (auch „SARndbox“). Dabei wird die Oberfläche eines Sandkastens von oben mit Bildern eines Projektors überlagert. Änderungen an der Sandoberfläche werden mit einer Tiefenkamera (oder auch Entfernungskamera) in Echtzeit registriert. Die Pixelwerte des Bildes der Tiefenkamera enthalten jeweils den Abstand der Kamera zur Sandoberfläche. Bewegte Objekte und Rauschen werden aus den Tiefenwerten herausgefiltert und Bereiche ohne Tiefenwerte (Löcher) werden mit Methoden der digitalen Bildverarbeitung gefüllt. Mithilfe des Projektors wird die Sandoberfläche entsprechend eingefärbt (siehe Abbildung 2.4). Das Rendering der Textur erfolgt mit



**Abb. 2.3** Beispiel für projektorbasiertes AR: Die Buddhastatue wird mit einem virtuellen Objekt (a) überlagert. In (b) sind die Marker rot hervorgehoben, die zur Registrierung genutzt wurden. Die Detailansicht in (c) zeigt die Genauigkeit der Überlagerung (Kern u. a., 2017).

GLSL-Shadern. Tiefenkamera, Projektor und Sandkasten sind fest verbunden. Die Position und Orientierung zueinander sowie die Abbildungseigenschaften von Tiefenkamera und Projektor müssen über eine Kalibrierung vorab bestimmt werden. Die Wasseroberfläche wird mithilfe von Saint-Venant-Gleichungen für Flachwasseroberflächen simuliert und ebenfalls mit GLSL-Shadern visualisiert. Für die Wasserströmung wird ein Modell basierend auf Navier-Stokes-Gleichungen genutzt. Grundlage für die Berechnungen ist die gemessene Sandoberfläche. Diese Berechnungen laufen in Echtzeit auf der GPU der Grafikkarte (Pantuwong u. a., 2016). Auf diese Weise können Nutzer im echten Sand virtuelles Wasser in Mulden fließen lassen und Staudämme bauen. Das System wurde zu Lehrzwecken für kleine und große Kinder entwickelt und da die Software frei verfügbar ist, gibt es viele Sandbox-Installationen weltweit (Kreylos, 2018). Die AR-Sandbox ist ein perfekter AR-Demonstrator, nicht nur für projektorbasierte Systeme: Sie ist durch das Zusammenspiel von echtem Sand und virtuell fließendem Wasser interaktiv und ohne spezielle Einarbeitung intuitiv nutzbar, da keine ungewohnten Gesten gelernt werden müssen.

Systembedingt sind projektorbasierte Anwendungen besser für den Innenbereich geeignet, oder der Betrieb wird für den Außenbereich auf die Abend- und Nachtstunden verlegt. Mit Laserprojektoren kann aber im Außenbereich auch bei Tageslicht gearbeitet werden. Solche Systeme werden bei industriellen Anwendungen eingesetzt (Extend3d, 2016). Projektorbasierte Anwendungsszenarien gibt es im Bereich Medizin (Lahanas u. a., 2014; Nicolau u. a., 2011; Kilgus u. a., 2014), in der industriellen



**Abb. 2.4** Beispiel für projektorbasierte AR: AR-Sandbox. Das Oberflächenmodell des Sandes wird von einer Tiefenkamera in Echtzeit erfasst und die Projektion auf die Sandoberfläche entsprechend angepasst. Das so projizierte virtuelle Wasser reagiert auf Änderungen und fließt in neu geschaffene Senken oder wird an Dämmen gestaut.

Produktion (Otto u. a., 2014; Sand u. a., 2016) und für Ausstellungen (Ridel u. a., 2014). Mit kleinen tragbaren Projektoren ist auch ein mobiler Einsatz realisierbar.

Tabelle 2.1 stellt einige Merkmale der hier vorgestellten Geräteklassen für AR-Systeme gegenüber. Die Angaben in der Zeile „Anzahl Personen“ bezieht sich auf die Menge, die mit jeweils nur einem Gerät eine Erweiterung betrachten kann. Für kollaborative Anwendungen bieten Plattformen wie Apples ARKit auf Smartphones, oder Microsoft für die HoloLens-Brille, entsprechende Funktionalitäten an, die es mehreren Personen mit jeweils eigenen Geräten erlauben, gemeinsam mit virtuellen Objekten zu interagieren und diese aus dem jeweils eigenen Blickwinkel zu betrachten.

Bei der Kalibrierung der Systeme müssen nur bei See-Through-Geräten die individuellen Unterschiede der Brillenträger durch jeweils eine eigene Kalibrierung berücksichtigt werden. Alle anderen Systeme können einmalig kalibriert werden.

### 2.1.2 Positionierung und Orientierung

Für die in der Einleitung geforderte Überlagerung in Echtzeit müssen Position und Orientierung des AR-Systems zu jedem Zeitpunkt bekannt sein. Im Fall einer videobasierten AR-Lösung müssen die 3D-Punkte  $X$  eines virtuellen Objekts auf das

	Videobasiert	See-through	Projektorbasiert	
<b>Geräteart</b>	Smartphone, Tablet	VR-Brille + Kamera(s)	HMD, Smart Glasses	DLP, LCD, LASER
<b>Sicht der Umgebung</b>	muss in Gesichtsfeld gehalten werden	nur durch Kameras; ungewohnt	fast ungestört; natürlich	normal
<b>Anzahl Personen</b>	wenige	eine	eine	viele
<b>Outdoor-Nutzung</b>	Abhängig von Displayhelligkeit	Abhängig von Kamerahelligkeit	schlechte Helligkeit bei Sonnenlicht	nur im Dunkeln oder mit LASER
<b>Kalibrierung</b>	einmalig	einmalig	vor jeder Nutzung bzw. für jeden Nutzer	einmalig
<b>Sonstiges</b>	allgemein verfügbar	spezielle Hardware; unbequem; Gewicht		

**Tab. 2.1** Vergleich der Darstellungsmöglichkeiten für AR-Systeme.

Kamerabild als 2D-Bildpunkte  $\mathbf{x}'$  mit der Projektionsmatrix  $\mathbf{P}$  abgebildet werden, sodass

$$\mathbf{x}' = \mathbf{P} \cdot \mathbf{X}$$

gilt. Die  $3 \times 4$ -Matrix für die projektive Abbildung  $\mathbf{P}$  mit

$$\mathbf{P} = \mathbf{K}[\mathbf{R}|\mathbf{t}]$$

ist hier aus Kameramatrix  $\mathbf{K}$ , Rotationsmatrix  $\mathbf{R}$  und Translationsvektor  $\mathbf{t}$  aufgebaut. Die Kameramatrix  $\mathbf{K}$  kann mittels Kalibrierung bestimmt werden. Die grundlegenden Zusammenhänge der projektiven Abbildung sind im Detail in den umfassenden Werken von Hartley u. a. (2004) und Förstner u. a. (2016) beschrieben. Die Pose  $[\mathbf{R}|\mathbf{t}]$  enthält Position und Orientierung und muss in Echtzeit ermittelt werden. Sie hat sechs Freiheitsgrade: drei für die Position und drei für die Orientierung im Raum.