

# Farbbildverarbeitung unter Echtzeitbedingungen in der Augmented Reality

Ingo Scholz\*, Jochen Schmidt, and Heinrich Niemann

Lehrstuhl für Mustererkennung, FAU Erlangen-Nürnberg  
Martensstr. 3, 91058 Erlangen, Germany  
scholz@informatik.uni-erlangen.de  
<http://www5.informatik.uni-erlangen.de>

**Zusammenfassung** Es wird ein System aus der Augmented Reality (Erweiterte Realität, AR) vorgestellt, das zur Kamerakalibrierung einen farbigen Würfel verwendet, der außerdem als Platzhalter für ein virtuelles Objekt dient. Zur Erkennung des Würfels in einer Szene werden Methoden aus der Farbbildverarbeitung verwendet, wie der Color Structure Code (CSC) und die Klassifikation der entstehenden Regionen nach ihrer Farbe. Zur Beschleunigung der Segmentierung wird ein hierarchisches Verfahren eingesetzt.

**Schlüsselwörter:** Augmented Reality, Farbsegmentierung

## 1 Einleitung

Das Ziel der Augmented Reality (AR), zu deutsch Erweiterte Realität, ist es, die (natürliche) Umgebung, die ein Benutzer wahrnimmt, durch virtuelle Daten zu erweitern. Dabei kann es sich um jede beliebige Art von Information handeln, wobei bei dem hier vorgestellten System ein zusätzliches, virtuelles Objekt in das Sichtfeld des Betrachters eingebracht wird. Dazu existieren bereits eine Reihe von prinzipiellen Lösungsansätzen, die in [2] zusammenfassend dargestellt werden.

Eines der Hauptprobleme bei dieser Art von Anwendung ist die Übereinstimmung der Koordinatensysteme der Umgebung und der physikalischen Aufbauten des AR-Systems, bestehend aus einem Gerät zur Bestimmung der Kopfposition und einem Bildschirm, die normalerweise am Kopf des Benutzers angebracht sind. Die hier gewählte Lösung ist ein videobasiertes System zur Positionsbestimmung – d. h. die Positionsbestimmung erfolgt mittels einer Kamera die ungefähr das gleiche Sichtfeld besitzt wie der Benutzer – kombiniert mit einem „video see-through head-mounted display“ [7]. Letzteres bedeutet, dass der Anwender ausschließlich das von der Kamera aufgenommene und veränderte Bild sieht.

---

\* Diese Arbeit wurde teilweise von der Deutschen Forschungsgesellschaft (DFG) im Rahmen des Sonderforschungsbereiches (SFB) 603 finanziert. Für den Inhalt sind ausschließlich die Autoren verantwortlich.

In vielen anderen AR-Systemen wird das Weltkoordinatensystem durch spezielle Markierungen bestimmt, die an bekannten Punkten in der Szene angebracht sind [4]. In unserem Fall werden diese Markierungen durch einen Metallwürfel bekannter Größe ersetzt, dessen Seiten jeweils mit einer anderen Farbe versehen wurden. Damit können Position und Orientierung des Würfels festgestellt werden, zusammen mit dem Weltkoordinatensystem, dessen Ursprung in den Mittelpunkt des Würfels gelegt wird. Die korrekte Ansicht des virtuellen Objektes wird dann in dem für den Anwender sichtbaren Bild über den Würfel gelegt, sodass dieser verdeckt wird.

Um den Würfel zu identifizieren werden dessen Farbeigenschaften benutzt. Jedes Bild aus einer Kamerasequenz wird farbsegmentiert, wodurch gut voneinander getrennte Regionen mit homogener Färbung entstehen, einschließlich je einer Region für jede sichtbare Seite des Würfels. Durch die anschließende Farbklassifikation werden die Regionen identifiziert, die Teil des Würfels sind. Wurde der Würfel gefunden, so werden seine sichtbaren Ecken detektiert und für die Kalibrierung der Kamera und die Projektion des virtuellen Objektes verwendet. Für die Berechnung der dazu nötigen Projektionsmatrix sind mindestens sechs bekannte Punkte nötig, also die Ecken von mindestens zwei Seiten des Würfels.

Ziel dieser Arbeit war es, ein Augmented Reality System zu entwickeln, bei dem keine künstlichen Markierungen in der Szene angebracht werden müssen, die auch nach der Erweiterung um ein virtuelles Objekt noch sichtbar sind. Durch den stattdessen verwendeten Würfel ist der Anwender sogar in der Lage, das virtuelle Objekt zu berühren. Außerdem sollte gezeigt werden, dass Farbbildverarbeitung in einem Echtzeitsystem anwendbar ist.

Im Folgenden werden in Abschnitt 2 und 3 die Segmentierung des Eingabebildes und die Klassifikation der resultierenden Regionen beschrieben. Abschnitt 4 stellt einige experimentelle Ergebnisse dar, gefolgt von einer abschließenden Zusammenfassung.

## 2 Farbsegmentierung

Der erste Schritt in der Erkennung des Würfels, der als Platzhalter für das virtuelle Objekt fungiert, ist die Farbsegmentierung jedes Bildes aus der Eingabesequenz. Abbildung 2 zeigt ein typisches Bild des Würfels.

Als Algorithmus zur Farbsegmentierung wurde der Color Structure Code (CSC) verwendet, der in [5] beschrieben wird. Der CSC basiert auf dem Hierarchical Structure Code (HSC) [3], dessen grundlegendes Prinzip die Verwendung von hexagonalen Regionen, sogenannten Inseln, in einer hierarchischen Struktur ist. Während der HSC zur Codierung von Konturen und Regionen in Grauwertbildern verwendet wird, wurde der CSC für die Segmentierung von Regionen in Farbbildern erweitert.

Der CSC arbeitet auf anwachsenden Regionen und besteht, wie andere Algorithmen dieser Art, aus zwei Phasen. In der ersten Phase, dem Zusammenfügen, werden in jeweils einer Hierarchieebene Regionen aus der nächstniedrigeren Ebene zu neuen Regionen zusammengefaßt. Abwechselnd mit dem Zusammenfügen

in jeder Ebene werden in der zweiten Phase solche Regionen geteilt, die aufgrund ihrer Größe eine inhomogene Farbe haben. Dadurch, daß auf Regionen niedrigerer Ebenen nur Zeiger gesetzt werden, wird die hohe Geschwindigkeit des Algorithmus erreicht. Das Ergebnis der Segmentierung ist in Abbildung 3 zu sehen.

Um entscheiden zu können, ob zwei Regionen zusammengefasst werden sollen, ist ein Farbähnlichkeitsmaß nötig. Das in [5] beschriebene Maß basiert auf dem HSV-Raum. Das Problem mit diesem Farbraum ist, dass der Farbwert (H) instabil wird, wenn die Werte für Sättigung (S) und Helligkeit (V) zu gering werden. Aus diesem Grund werden in [5] die Schwellwerte für die Ähnlichkeit zweier Farben in Abhängigkeit von S und V gewählt.

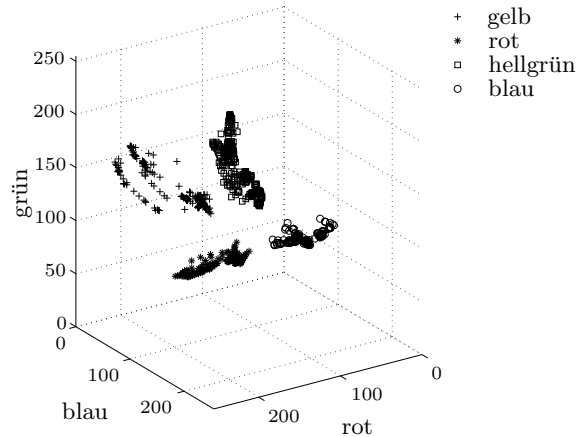
Im Gegensatz dazu wurde für das vorliegende System ein einfacheres, euklidisches Maß verwendet, das auf dem RGB-Raum arbeitet. Für die Trennung der hellen und satten Farben der Würfelseiten vom Hintergrund des Bildes hat sich dieses als ausreichend erwiesen.

Obwohl der CSC ein vergleichsweise schneller Vertreter der Segmentierungsalgorithmen ist, stellt er doch den größten Flaschenhals des Systems dar. Um die Anzahl der Verarbeitungsschritte möglichst gering zu halten, wurde zur Aufnahme der Bildsequenzen eine digitale Kamera verwendet, die mit drei CCD-Chips arbeitet und außerdem über Progressive Scan verfügt, sodass ein Bild nicht aus zwei Halbbildern zusammengesetzt wird. Durch die so erreichte hohe Qualität der Bilder konnte die Vorverarbeitung auf einen einzigen Filter zur Rauschreduktion beschränkt werden, den in [5] ebenfalls verwendeten Symmetric Nearest Neighbour (SNN) Filter.

Desweiteren wurde zur Vermeidung von unnötigen Berechnungen eine zwei-stufige AuflösungsPyramide verwendet, um den CSC nicht auf das gesamte Bild anwenden zu müssen. In einem ersten Schritt wird das Eingabebild auf ein Fünftel seiner Größe verringert und auf dieses die Farbsegmentierung angewendet (siehe Abbildung 4). Mit Hilfe der Farbklassifikation aus Abschnitt 3 und dem Wissen, dass der Würfel durch benachbarte Regionen bestimmter Farben dargestellt wird, können Bereiche des Bildes identifiziert werden, in denen sich der Würfel befinden könnte. Dabei wird angenommen, dass immer mindestens zwei Seiten des Würfels sichtbar sind. Wäre nur eine Seite sichtbar, könnte auch die Projektionsmatrix aus den verbleibenden vier sichtbaren Eckpunkten nicht berechnet werden. Im zweiten Schritt wird wieder das Originalbild betrachtet und ein Fenster um den Würfelkandidaten herum ausgeschnitten. Auf dieses wird die endgültige Farbsegmentierung angewendet.

### 3 Farbklassifikation

Der erste Schritt zur Identifikation des Würfels ist die Klassifikation der Regionen aus der Farbsegmentierung. Die Farben der Seiten des Würfels bilden jeweils eine Region im Merkmalsraum, wie in Abbildung 1 für den RGB Farbraum dargestellt. Die Streuung der Farbwerte entsteht durch verschiedene Beleuchtungen der Szene. Die Regionen im Bild werden nun in sieben Klassen eingeteilt, jeweils



**Abbildung 1.** Vier von sechs Farbklassen im RGB Farbraum.

eine für jede Seite des Würfels, und eine Rückweisungsklasse. Dazu wird ein numerischer Klassifikator verwendet, als Abstandsmaß dient der Mahalanobis-Abstand.

Zur Klassifikation werden zwei verschiedene Merkmalsvektoren verwendet. Zum einen die RGB-Werte der Durchschnittsfarbe einer Region, und zum anderen Sättigung und Farbwert, H und S, aus dem HSV-Raum. Die Helligkeit (V) einer Farbe dagegen stellt kein geeignetes Merkmal dar, da sie hauptsächlich von der Beleuchtung der Szene abhängig ist. Die Farben des Würfels wurden so ausgewählt, dass sie möglichst kompakte Regionen mit großem Interklassenabstand in diesen Merkmalsräumen bilden. Der Farbklassifikator wird für beide Merkmalsvektoren getrennt angewendet. Eine Region wird schließlich als Würfelseite identifiziert wenn die Ergebnisse beider Merkmalsvektoren über einem bestimmten Schwellwert liegen.

Für das Training der Parameter des Mahalanobis-Abstandsmaßes, d. h. die durchschnittlichen Farbvektoren und die Kovarianzmatrix, wurden Bilder des Würfels mit unterschiedlicher Beleuchtung und verschiedenen Winkeln zur Lichtquelle verwendet. Die Bilder wurden farbsegmentiert bevor die Farbdurchschnittswerte der Regionen zur Trainingsstichprobe hinzugefügt wurden.

Die folgenden Schritte zur Identifikation des Würfels aus den klassifizierten Regionen benutzen das Wissen über seine bekannten geometrischen Eigenschaften. Wie schon in Abschnitt 2 erwähnt, wird davon ausgegangen dass mindestens zwei Seiten des Würfels sichtbar sind. Daher werden alle im Bild vorhandenen Paare und Tripel solcher Regionen, die aneinander angrenzen, als Würfelkandidat betrachtet. Dabei müssen die Regionen eine mögliche Kombination von Farben besitzen, d. h. es dürfen keine zwei dieselbe Farbe besitzen oder die Farben von einander gegenüberliegenden Seiten aneinander grenzen.

Weitere Kriterien sind das Größenverhältnis und die Länge der gemeinsamen Grenzen zweier Regionen. Liegt ein Kandidat hier in einem sinnvollen Bereich, so wird mittels der Hough-Transformation die Kontur jeder seiner Regionen durch vier Linien approximiert. Aus der Güte der Approximation und der Existenz von parallelen Linien – es wird angenommen dass die Seite eines Würfels näherungsweise ein Parallelogramm ist – wird ein Gütemaß für jeden Kandidaten abgeleitet.

Nach der Anwendung dieser Restriktionen bleibt nur eine kleine Anzahl (im besten Fall nur einer) an Würfelkandidaten übrig. Die Ecken des wahrscheinlichsten Kandidaten werden nun aus den Schnittpunkten der vorher berechneten Kantenapproximationen gewonnen. Diese (mindestens sechs) Punkte werden dazu verwendet, die Projektionsmatrix zu berechnen die das Modell eines virtuellen Objekts auf die Position des Würfels im Bild abbildet. Das dabei angewandte Verfahren wird in [6] genauer beschrieben.

Das Endergebnis der oben beschriebenen Verfahren, die Überlagerung des Würfels, ist in Abbildung 5 zu sehen. Als virtuelles Objekt wurde hier die bekannte Teekanne verwendet.

## 4 Experimentelle Ergebnisse

Die Bilder für die folgenden Experimente wurden mit einer „Digital Video“-Kamera aufgenommen und ihre Größe auf  $360 \times 288$  Pixel halbiert. Die Größe der Trainingsstichprobe für den Farbklassifikator lag zwischen 577 und 885 Merkmalsvektoren pro Farbe des Würfels. In einer Testsequenz mit ähnlicher Beleuchtung wie in den Bildern, die für das Farbtraining verwendet wurden, wurde damit der Würfel in 93% der Bilder korrekt lokalisiert. Dieses Ergebnis verschlechtert sich natürlich für schlechtere, d. h. vor allem dunklere, Beleuchtung oder für teilweise Verdeckung des Würfels. Dass das System aber trotz Verdeckungen noch korrekt arbeiten kann zeigt Abbildung 6.

Eines der Ziele dieser Arbeit war es, die Anwendbarkeit von Algorithmen aus der Farbbildverarbeitung in Echtzeit-Systemen für Augmented Reality zu untersuchen. Bisher konnte das Ziel einer Echtzeit-Anwendung allerdings nicht erreicht werden. Auf einem 800 MHz Pentium III dauerte die Verarbeitung eines Bildes zwischen 460 und 700 ms, exklusive der Bilderfassung und Übertragung von der Kamera und dem Rendern des Ergebnisbildes. Da eine sinnvolle Echtzeitanwendung Bildwiederholraten von mindestens 10 Bildern pro Sekunde benötigt, sind die erreichten Werte noch um mindestens eine Größenordnung zu hoch. Trotzdem sind die Ergebnisse ermutigend: der CSC stellt immer noch den hauptsächlichen Flaschenhals des Systems dar, da er zwischen 250 und 350 ms für die in Abschnitt 2 beschriebenen zwei Arbeitsschritte benötigt. Im Vergleich dazu benötigte eine Testversion der Autoren von [5] nur etwa 40 bis 130 ms für dieselben Bilder. Dies zeigt, dass durch weitere Optimierung noch eine nennenswerte Beschleunigung möglich ist.

## 5 Zusammenfassung

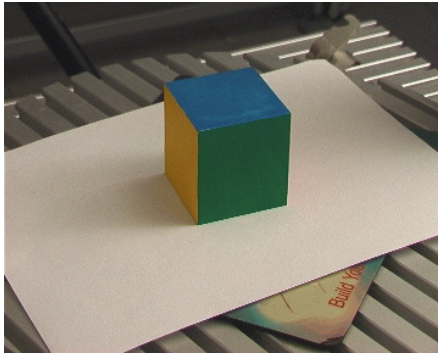
In den vorangegangenen Abschnitten wurde ein Überblick über die Funktionsweise eines Systems der Augmented Reality gegeben, das Methoden aus der Farbbildverarbeitung verwendet. Dabei wird ein farbiger Würfel als Platzhalter für ein virtuelles Objekt verwendet. Zur Erkennung des Würfels wird das Bild mittels des CSC farbsegmentiert und die entstehenden Regionen nach ihrer Farbe als dem Würfel möglicherweise zugehörig oder nicht klassifiziert. Aus der entstehenden Menge an Kandidaten wird der tatsächliche Würfel extrahiert und seine Eckpunkte bestimmt. Aus diesen schließlich wird eine Projektionsmatrix berechnet, die ein virtuelles Objekt in die Szene an der Stelle des Würfels abbilden kann.

Das derzeitige Ergebnis unseres Projektes zeigt, dass Farbbildverarbeitung auch in Anwendungen mit Echtzeit-Anforderung, wie Augmented Reality, anwendbar ist. Durch die Verwendung neuartiger Algorithmen und besserer Prozessoren kann die zusätzliche Information, die in Farbbildern gegenüber Grauwertbildern enthalten ist, genutzt werden, um die Robustheit von Segmentierung und Klassifikation zu erhöhen. Dadurch wird das hier entwickelte System unabhängig von Informationen von zusätzliche Sensoren, und arbeitet auf denselben Daten die auch dem Anwender zur Verfügung stehen.

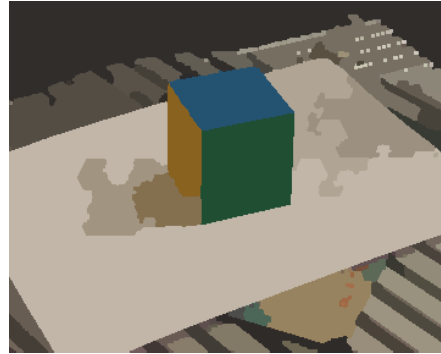
Für weitere Informationen, Beispielbilder und -sequenzen siehe [1].

## Literatur

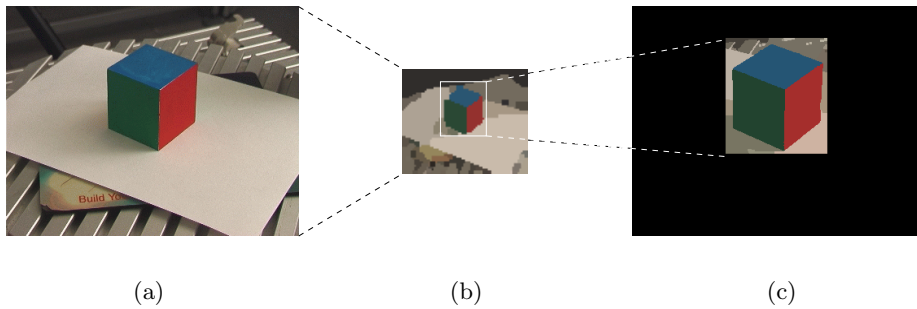
1. <http://www5.informatik.uni-erlangen.de/~ar>.
2. Ronald T. Azuma. A survey of augmented reality. *Presence: Teleoperators and Virtual Environments*, 6(4):355–385, August 1997.
3. G. Hartmann. Recognition of Hierarchically Encoded Images by Technical and Biological Systems. *Biological Cybernetics*, 57:73–84, 1987.
4. Y. Ohta and H. Tamura, editors. *Mixed Reality – Merging Real and Virtual Worlds*. Springer, Berlin, 1999.
5. V. Rehrmann and L. Priese. Fast and Robust Segmentation of Natural Color Scenes. In *Proceedings of the 3<sup>rd</sup> Asian Conference on Computer Vision*, volume 1, pages 598–606, HongKong, January 1998.
6. J. Schmidt, I. Scholz, and H. Niemann. Placing Arbitrary Objects in a Real Scene Using a Color Cube for Pose Estimation. In *DAGM 2001, 23. Symposium für Mustererkennung*, Munich, September 12-14 2001. to appear.
7. I. Scholz. Augmented Reality: A System for the Visualization of Virtual Objects Using a Head-mounted Display by Localization of a Real Object of Known Geometry and Color. Diplomarbeit, Lehrstuhl für Mustererkennung, Universität Erlangen-Nürnberg, 2000.



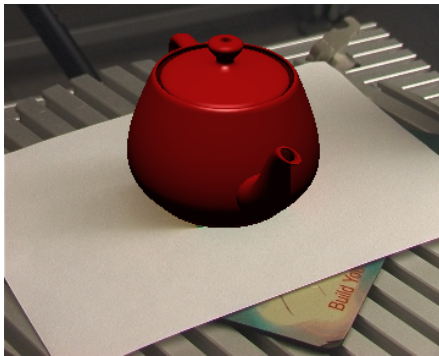
**Abbildung 2.** Originalbild aus der Eingabesequenz.



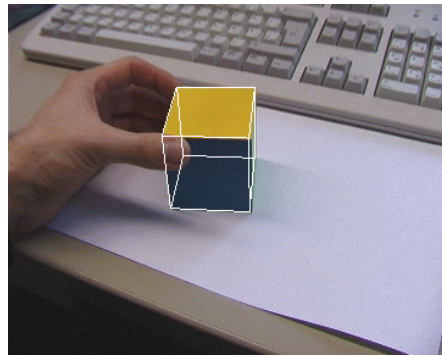
**Abbildung 3.** Ergebnis der Anwendung des Color Structure Code (CSC).



**Abbildung 4.** Auffinden einer möglichen Position des Würfels in einem Bild mit verringerter Auflösung. (a) Originalbild, (b) größenreduziertes Bild, (c) mögliche Position des Würfels.



**Abbildung 5.** Durch eine virtuelle Teekanne erweiterte Szene.



**Abbildung 6.** Teilweise verdeckter Würfel.