

# Augmented Reality

## Markerbasiertes Tracking für Augmented Reality Applikationen

### Abstract

Tracking ist wohl eines der grössten Probleme in Augmented Reality aktuell. Es wurden viele verschiedene Möglichkeiten dafür vorgeschlagen, aber ich werde mich hier hauptsächlich auf Marker-basiertes Tracking beschränken und da zuerst ein bisschen wie es funktioniert, danach was ein guter Marker haben soll und schliesslich ein Ausblick auf die Möglichkeiten von AR an einem Beispiel das auf Marker-basierten Tracking beruht.

## 1. Marker-basiertes Tracking

### 1.1 Tracking allgemein

Tracking in Augmented Reality beinhaltet die Bestimmung der Position eines Objektes und auch dessen Ausrichtung. Dies sowohl für diejenigen Objekte die man mit virtuellen Objekten erweitern will, als auch für die Position und die Blickrichtung der Person welche das System verwendet.

Tracking wurde schon mit vielen verschiedenen Möglichkeiten vorgeschlagen. Ich möchte hier nun auf einige kurz eingehen und auch deren Vor- und Nachteile ein bisschen aufzeigen.

Zum einen wurden akustische Systeme mit Ultraschall für Tracking vorgeschlagen, dies kann aber vor allem in einer grösseren Umgebung mit vielen verschiedenen zusätzlichen Geräuschquellen zu einigen Störungen führen, was das Tracking dann verunmöglicht.

Eine weitere Möglichkeit sind magnetische Tracking-Varianten die aber in den meisten Anwendungen scheitern, wenn es zu viele metallische Gegenstände im Raum hat, welche das Magnetfeld der jeweiligen Orientierungspunkte verfälschen können. Ein relativ grosses Feld sind die optischen Trackingverfahren. Zu diesem gehören auch die markerbasierten Ansätze, bei welchen das Tracking über speziell angebrachte Marker erfolgt. Auch dieses Verfahren hat verschiedene Nachteile, da zum einen die Marker verdeckt sein können und zum anderen muss auch immer eine gewisse Helligkeit vorhanden sein, da sonst sowohl das Auffinden als auch das Unterscheiden der Marker zu neuen Schwierigkeiten führen kann.

### 1.2 Augmented Reality System von Kato/Billinghurst [1]

#### 1.2.1 System Überblick

Das von Kato und Billinghurst entwickelte AR System soll zur Unterstützung der Zusammenarbeit mehrerer Benutzer verwendet werden. Die Grundidee dabei ist, dass ein Aussendienstmitarbeiter von einem Experten am Desktop-PC unterstützt werden kann. Der Benutzer mit dem AR-Interface trägt eine HMD-Brille (Head-Mounted-Display, see-through) und kann über eine gemeinsame Oberfläche (shared white

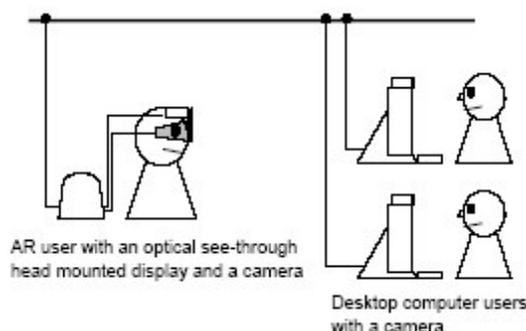


Figure 1: System Aufbau

board) mit seinen Kollegen, welche an den herkömmlichen Computern sitzen kommunizieren. Gleichzeitig besitzt er auch einen Marker je weiteren Mitarbeiter, auf welchem dann das Bild der Kamera angezeigt wird, die auf dem Bildschirm des jeweiligen Users steht. So wird es für die unterschiedlichen Benutzer möglich miteinander zu kommunizieren.

Das „Shared White Board“ kann von den verschiedenen Benutzertypen unterschiedlich bedient werden. Während die Benutzer an den PCs herkömmlich mit Maus und Tastatur arbeiten, verwendet



Figure 2: Benutzung des AR-Interfaces

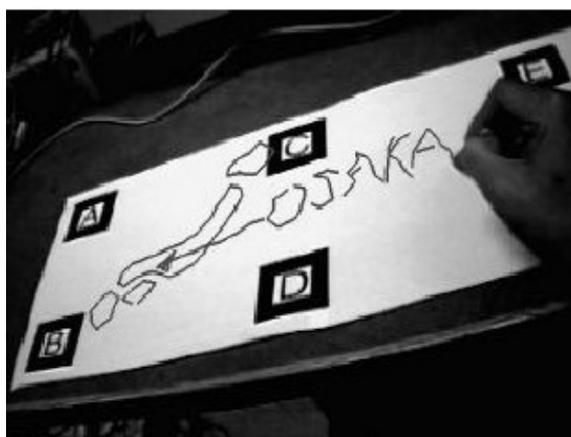


Figure 3: Shared White Board



Figure 4: Viedokonferenz

derjenige mit dem HMD einen speziellen Stift, mit dem er virtuelle Objekte auf die Oberfläche „zeichnen“ kann. Diese Objekte können dann von allen Benutzern eingesehen und bearbeitet/gelöscht werden.

### 1.2.2 Positionsbestimmung von Markern

Zur Feststellung der Position werden Marker mit bekannter Größe verwendet. Die Transformationsmatrix (Fig 5) wird mit Bildanalyse abgeschätzt. Dies geschieht indem die Koordinaten der Ecken des Markers, sowohl im Realen als auch auf der Bildschirmoberfläche verwendet werden, um alle Unbekannten zu bestimmen.

Das innere Bild wird dann ebenfalls normalisiert und dann mit den in der Datenbank vorhanden abgeglichen.

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} = \begin{bmatrix} V_{11} & V_{12} & V_{13} & W_x \\ V_{21} & V_{22} & V_{23} & W_y \\ V_{31} & V_{32} & V_{33} & W_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_m \\ Y_m \\ Z_m \\ 1 \end{bmatrix} = \begin{bmatrix} V_{3 \times 3} & W_{3 \times 1} \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_m \\ Y_m \\ Z_m \\ 1 \end{bmatrix} = T_{cm} \begin{bmatrix} X_m \\ Y_m \\ Z_m \\ 1 \end{bmatrix} \quad (\text{eq. 1})$$

Figure 5: Beziehung zwischen Markerkoordinaten und Kamerakoordinaten.

### 1.2.3 Kalibrierung

In einem System mit einem „see-through“ HMD müssen die Koordinaten der Augen mit denjenigen der Kamera und dem Display abgeglichen werden.

Dazu wird vor der eigentlichen Nutzung des Systems eine Kalibrierungsphase gemacht, in welcher die benötigten Transformationsmatrizen bestimmt werden können.

Zur Kamerakalibrierung wird ein spezielles Stück Karton mit einem Gitternetz drauf vor die Kamera gehängt. Da die Koordinaten der Gitterpunkte bekannt sind, kann man mit Hilfe vieler solcher Punkte die Transformationsmatrix zwischen Kamera-Koordinaten und Kamera-Schirm-Koordinaten ermitteln.

Die Kalibrierung des HMD wird für jedes Auge einzeln gemacht. Dazu verwendet man Marker und die vorher genannte Tracking Methode. Der Benutzer hat dazu einen Marker in der Hand und versucht dessen Zentrum mit einem auf dem HMD eingeblendeten Fadenkreuz in Deckung zu bringen. Nach einigen Wiederholungen dieses Zentriervorganges können dann die Matrizen  $Q_{se}$  und  $T_{ec}$  bestimmt werden. Diese beiden Matrizen werden dann für den gesamten Arbeitsvorgang verwendet und nur die Transformationsmatrix  $T_{cm}$  muss bei jeder Bewegung des Kopfes neu berechnet werden.

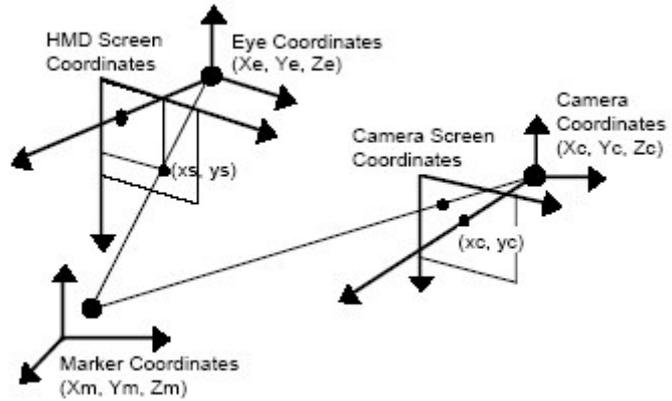


Figure 6: Koordinatensystem im Kalibrierungsprozess

$$\begin{bmatrix} ix_s \\ iy_s \\ i \\ 1 \end{bmatrix} = Q_{se} \begin{bmatrix} X_e \\ Y_e \\ Z_e \\ 1 \end{bmatrix} = Q_{se} T_{ec} \begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} = Q_{se} T_{ec} T_{cm} \begin{bmatrix} X_m \\ Y_m \\ Z_m \\ 1 \end{bmatrix}$$

Figure 6: Beziehung zwischen Augen- und Kamerakoordinaten

### 1.2.4 Genauigkeit des Systems

Die Abschätzung der Genauigkeit wurde mit quadratischen Markern mit 80mm Seitenlänge durchgeführt, indem man sie in verschiedenen Abständen und verschiedenen Winkeln zur Kamera gehalten hat. Wie unschwer festzustellen ist, wird die Genauigkeit schlechter, je weiter sich der Marker von der Kamera entfernt. Es zeigt sich auch, dass das System wirklich nur für Benutzung im Nahbereich geeignet ist und die Anwendung sollte nicht extrem empfindlich sein auf kleine Fehler.

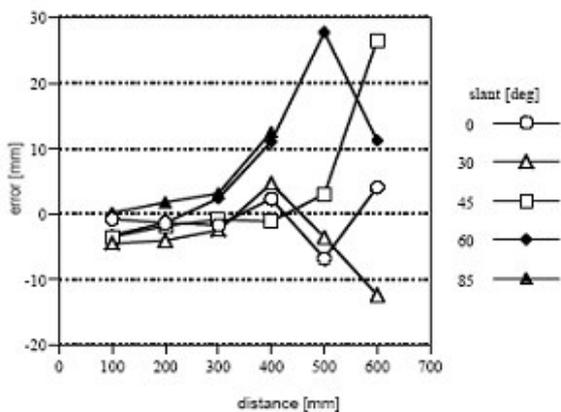


Figure 7: Positionsabweichung

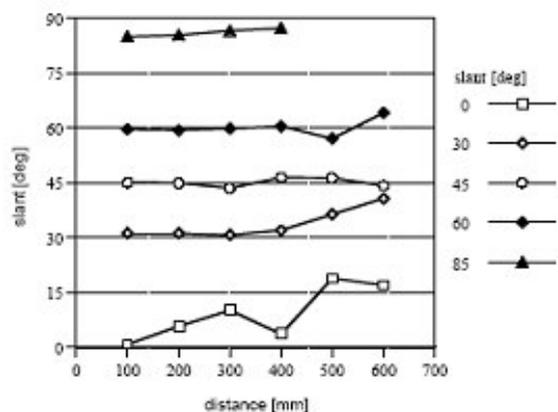


Figure 8: Entdeckter Winkelfehler

### 1.2.5 Zusammenfassung

Grundsätzlich hat sich das System für diese Anwendung als recht gut erwiesen. Die Marker-Erkennung auf nahe Distanzen ist nicht schlecht, aber nimmt auf grössere Distanzen relativ schnell ab. Recht vorteilhaft ist auch das Kalibrierungssystem, welches relativ einfach funktioniert und bei dem der Kopf nicht die ganze Zeit ruhig gehalten werden muss. Bei einigen Tests damit wurde festgestellt, dass auch ein weniger geübter Benutzer relativ einfach das System kalibrieren kann.

## 2. Best fiducials [2]

### 2.1 Anforderungen an Marker

Bei optischen Tracking-Systemen im Allgemein muss man sich immer überlegen, wie man ein Objekt am besten identifizieren kann. Deshalb drängt sich auch bei den Markern die Überlegung auf, was den optimal ist.

Hier nun ein paar Überlegungen was ein Marker alles für Eigenschaften braucht. Zuerst einmal sollte ein Marker aus allen Richtungen gleich gut erkennbar sein, er sollte also egal wie er liegt erkannt werden. Dazu ist es auch wichtig, dass man einfach erkennen kann wo er ist und wie er da liegt. Das heisst man muss ihn einfach lokalisieren können im gesamten Bild, der Kontrast zwischen realer Welt und Marker sollte möglichst gross sein. Wenn er lokalisiert ist, braucht es einen guten Algorithmus um den Marker auch eindeutig identifizieren zu können. Und zu guter letzt nützt ein einzelner Marker nicht viel, deshalb sollte er einer „Familie“ von Markern angehören, damit man auch wirklich sinnvoll damit arbeiten kann. Schliesslich reicht es nicht einen einzigen Supermarker zu haben, da jede Anwendung auf einer Menge operiert und für einen grösseren Raum können das ein paar Hundert Marker sein. Nach all diesen Überlegungen sind dann folgende Fragen aufgetaucht:

- Was ist eine geeignete Form?
- Welche Farben sollen benutzt werden? Schwarz-weiss, farbig oder in Graustufen?
- Wie identifiziert man den Marker?
- Wie findet man den Marker im Bild?

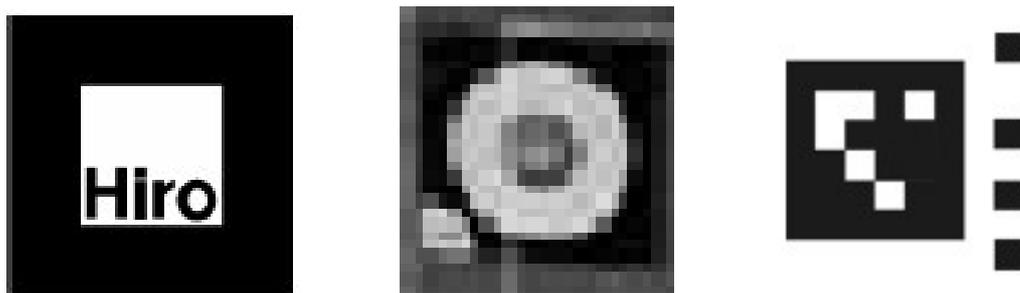


Figure 9: Beispiele einiger Marker die von Verschiedenen Systemen verwendet wurden.

### 2.2 Der „optimale“ Marker

Aufbauend auf den vorher genannten Kriterien wurde von Owen folgende Punkte für einen guten Marker vorgeschlagen.

#### *Form:*

Die Form des Markers sollte möglichst gleichmässig sein, damit man ihn aus allen

Richtungen gut erkennen kann. Dazu sind mindestens 4 unabhängige, eindeutig identifizierbare Punkte nötig, damit der Marker beliebig angeordnet sein kann und in jede mögliche Richtung im Raum ausgerichtet sein kann. Sie schlagen deshalb ein Quadrat als optimale Form vor, da es alle obengenannten Kriterien erfüllt und dazu eine einfach Figur darstellt.

#### *Farbe:*

Es wurde vorgeschlagen nur mit Graustufen zu arbeiten, da diese deutlich schneller verarbeitet werden können, als farbige Bilder. Dazu kommt bei Farben, dass sie je nach Kamera und Lichtverhältnissen nicht immer genau gleich erkannt werden, was natürlich zu grösseren Problemen führen könnte. Schwarz-weiss Bilder bieten im Normalfall einen sehr guten Kontrast zur meist farbigen Umgebung, welcher bei der Identifikation der Marker von Vorteil ist.

#### *Markerlokalisierung:*

Um den Marker in einer realen Umgebung gut zu finden, wurde ein dicker, schwarzer, quadratischer Rand vorgeschlagen und dazu das ganze auf weissem Hintergrund. So setzt sich der Marker wohl in jeder Umgebung ziemlich eindeutig vom Rest ab. Sie haben sich hier dazu entschlossen den Rand 15% des gesamten Bildes zu machen, damit sowohl die Identifikation gut ist, als auch ein möglichst grosser Bereich für das innere Bild vorhanden ist

#### *Bild Identifikation:*

Zur eigentlichen Identifikation des Markers wird ein soeben entdeckter Marker (I) mit möglichen Bildern (P) aus der Datenbank verglichen. Dies wird durch Bestimmung eines Korrelationskoeffizienten gemacht. Dieser Koeffizient wird mit folgender Formel berechnet:

$$\rho = \frac{\sum_x \sum_y (I(x, y) - \mu_I)(P(x, y) - \mu_P)}{\sigma_I \sigma_P}$$

$$\sigma_I = \sqrt{\left( \sum_x \sum_y (I(x, y) - \mu_I)^2 \right)} \quad \mu_I = \frac{1}{xy} \sum_x \sum_y I(x, y)$$

Die Werte für die Bilder aus der Datenbank können natürlich schon vorher berechnet werden, so dass beim Vergleich weniger gemacht werden muss. Bei der Auswahl welches Bild denn gefunden wurde, wählt man einfach dasjenige mit dem höchsten  $\rho$  und dieser Wert muss auch noch grösser als ein vorher abgemachter Grenzwert sein (0.5 im Normalfall). Es ist klar dass die verschiedenen Bilder der gewählten Marker-Familie möglichst ein kleines  $\rho$  untereinander haben sollten. Im Idealfall sollte es sogar 0 sein.

Wenn man nun diese Bilder spontan auswählt besteht natürlich die Möglichkeit, dass eine grosse Korrelation zwischen den einzelnen Bildern besteht.

Grundsätzlich erscheint der Gedanke Buchstaben als Marker zu nehmen relativ in-

interessant und wir können diese von Auge auch sehr gut unterscheiden. Als Beispiel habe ich hier 4 Marker vom vorhergehenden Beispiel. Mit der vorhin genannten Formel kommen wir auf einen Korrelationswert von 0.498 im für das Kanji-Symbol und das A, was also schon beinahe für eine positive Übereinstimmung reichen würde. Noch extremer ist es bei C und G die sogar eine Korrelation von 0.820 erreichen. Dieser Wert liegt damit bei weitem über dem Grenzwert. Vor allem wenn man bedenkt, dass auch einmal ein Teil des Markers verdeckt sein kann oder dass die Lichtverhältnisse nicht immer so gut sind.

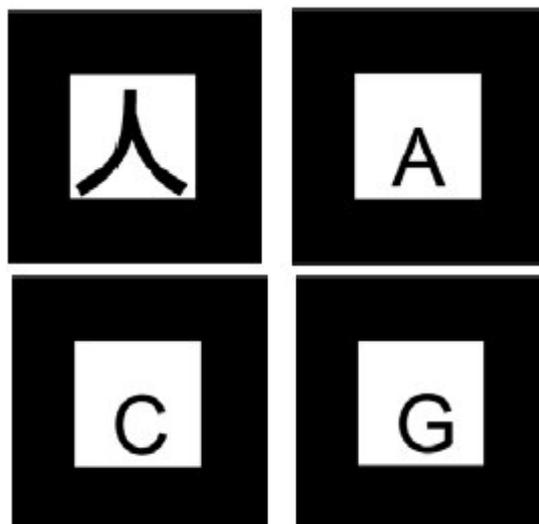


Figure 10: Beispiele von ähnlichen Markern

Um also ein besseres Bild zu erhalten haben sie sich überlegt mit was für einer Funktion  $\rho=0$  entstehen würde. Sie haben sich dann für einen Marker basierend auf folgender Funktion entschlossen, da es so möglich ist eine grössere Gruppe von verschiedenen Bildern zu erstellen und auch eine relativ einfach Möglichkeit sie zu vergleichen.

$$I_{u,v}(x,y) = \frac{B_{u,v}(x,y) + B_{1,0}(x,y) + 2}{4}$$

$$B_{u,v}(x,y) = \cos\left(\frac{(2x+1)u\pi}{2N}\right) \cos\left(\frac{(2y+1)v\pi}{2N}\right)$$

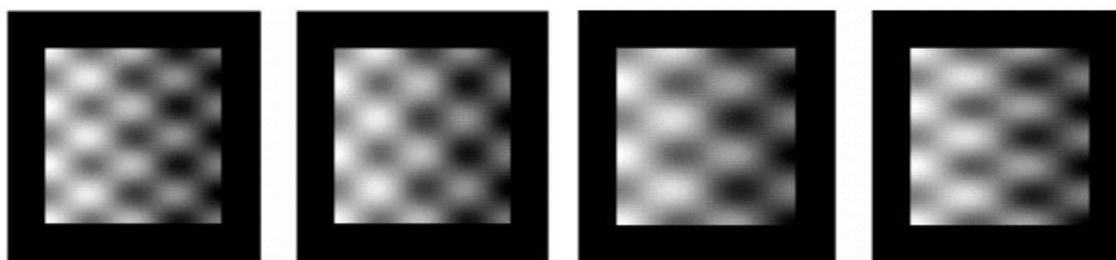


Figure 11: Vier verschiedene Marker aus der von ihnen vorgeschlagenen Marker Familie

Diess Ergebnis erscheint für den Betrachter etwas unerwartet, da diese Bilder von Auge lange nicht so gut unterschieden werden können als die oben gezeigten.

### 2.3 Ergebnisse

Grundsätzlich lassen sich die hier gezeigten Marker sehr gut und auch schnell

durch optische Methoden identifizieren. Einzig bei schlechter Beleuchtung lässt die Erkennungsrate nach, was natürlich die Frage aufwirft ob diese Marker den wirklich für den Einsatz in der Praxis geeignet sind, da die meisten Räume alles andere als perfekt ausgeleuchtet sind.

### 3. AR mit Handeld Devices [3]

#### 3.1 HMD vs. PDA

In den meisten bisher entwickelten Systemen wurde ein HMD verwendet. Doch diese Systeme haben gewisse Nachteile. Dadurch dass sie aus verschiedenen Komponenten (Notebook, Display, Kamera ...) bestehen, sind sie sehr wartungsintensiv und anfällig für Hardware-Störungen aller Art. Dazu werden sie im aktuellen Stadium wohl kaum von den durchschnittlichen Benutzern akzeptiert, da schliesslich die wenigsten mit so einem HMD durch die Gegend laufen wollen. Ein PDA dagegen ist für einige Anwendungen aktuell sicherlich besser geeignet, da er deutlich leichter ist und auch einiges weniger Energie verbraucht. Dadurch dass alles an einem Stück ist, wird es auch zu weniger Problemen mit der Hardware führen und schlussendlich wird er beim „normalen“ User sicherlich besser akzeptiert als die herkömmlichen Rucksack-Systeme. Natürlich hat der PDA nicht nur Vorteile. Insbesondere die Tatsache, dass man ihn immer in der Hand halten muss, macht diese Lösung für einige Einsatzgebiete unbenutzbar. Dazu kommt dass nur ein sehr kleiner Bereich des Blickfeldes auch wirklich mit virtuellen Objekten ergänzt werden kann.



Figure 12: Gegenüberstellung der beiden Systeme.

#### 3.2 The Invisible Train

Als Beispielanwendung für AR mit PDA wurde ein kleines Multiplayer-Spiel entwickelt, bei dem der Spieler einen virtuellen Zug über reale Schienen fahren lässt. Dabei hat er die Möglichkeit über den Touchscreen seines PDA die Geschwindigkeit seines Zuges und auch die Weichen auf der Strecke zu steuern.

Die Berechnungen laufen alle auf den PDA's ab und die Kommunikation zwischen den einzelnen Benutzern erfolgt über Wireless LAN. Das Tracking wird in diesem Fall über Marker gemacht und ist sehr ähnlich wie im vorhergehenden Beispiel erklärt.



Figure 13: The Invisible Train

Das Spiel wurde dann zur Erprobung von circa 5000 verschiedenen Benutzern gespielt. Waren diese Leute anfangs aus der AR Forschung so wurde es am Schluss der Testphase auch von Personen gespielt, welche vorher noch nie mit AR in Kontakt gekommen waren.

### 3.3 Ergebnisse

Es hat sich gezeigt, dass dieses System mit dem PDA und der Kamera recht robust ist und auch von den Benutzern ohne Vorbehalte akzeptiert wurde. Dazu zeigte sich auch, dass es wirklich einfach zu bedienen ist und auch für Leute ohne Erfahrung mit AR oder langer Einarbeitungszeit problemlos zu benutzen war.

Die einzigen Probleme mit der Hardware traten auf, wenn die Kamera während dem Betrieb rausgenommen wurde. Als weiteres Problem erwies sich noch die relativ kurze Akkulaufzeit, welche mit circa 2 Stunden noch nicht das bringt was man von einem System erwarten würde, dass den ganzen Tag laufen sollte.

### References

- [1]H. Kato, Mark Billinghurst, **Marker Tracking and HMD Calibration for a Video-based Augmented Reality Conferencing System.**  
Proceedings of the 2nd International Workshop on Augmented Reality (IWAR 99), San Francisco, USA, October 1999
- [2]Charles B. Owen, Fan Xiao, Paul Middlin, **What is the best fiducial?**  
The First IEEE International Augmented Reality Toolkit Workshop, pp. 98-105, Darmstadt, Germany, September 2002
- [3]Daniel Wagner, Thomas Pintaric, Florian Ledermann, Dieter Schmalstieg, **Towards Massively Multi-User Augmented Reality on Handheld Devices.**  
Third International Conference on Pervasive Computing (Pervasive 2005), Munich, Germany, May 2005