

SPIRIT – Videobasierte mobile Augmented Reality Lösung zur interaktiven Informationsvermittlung

HABIBURRAHMAN DASTAGEERI¹, MARTIN STORZ¹ & VOLKER COORS¹

Zusammenfassung: Neben dem Inhalt von Informationen ist die Art und Weise der Informationsvermittlung ausschlaggebend für die User Experience (UX). Eine Option für die Wiedergabe von ortsgebundenen Informationen, auch für weniger technikaffine User, ist der Einsatz von Augmented Reality (AR). Das Projekt SPIRIT setzt eine videobasierte AR Android-Applikation für das Museum Römerkastell Saalburg um. Die Informationsvermittlung findet hierbei in Form von Videos statt, die auf dem Realkamerabild eines Mobile Devices projiziert werden. Voraussetzung hierfür ist ein videobasiertes Tracking, wofür ein Bildverarbeitungsansatz verwendet wird. Dabei musste insbesondere die technische Leistungsfähigkeit der Mobile Devices berücksichtigt werden. Zur Steigerung der Performance wurden die verwendeten Image-Tracking-Verfahren für mobile Endgeräte optimiert. Für den Content wurde der OGC ARML 2.0 Standard erweitert und verwendet.

1 Einleitung

Die mobile und ortsgebundene Informationswiedergabe in Echtzeit stellt Anforderungen an Hardware, Software, Content und User. Die hier vorgestellte Lösung sieht den Schwerpunkt darin, Informationen pädagogisch und visuell ansprechend aufzubereiten, um dann den Aufruf ortsgebunden und kontextspezifisch automatisiert in Form einer Augmented Reality (AR) App wiederzugeben. Im Rahmen des Projektes SPIRIT setzt die Hochschule für Technik Stuttgart (HFT) gemeinsam mit der Hochschule RheinMain (HSRM) einen Prototypen für das archäologische Museum und Römerkastell Saalburg um, das als Projektpartner gewonnen werden konnte.

2 Problemstellung

Als Schauplatz der App dient der archäologische Park und Römerkastell Saalburg, welche aus einer rekonstruierten römischen Militäranlage sowie einer inszenierten Landschaft besteht. Die Bauwerke sind begehbar und dienen zumeist als Ausstellungsräume der Exponate des dazugehörigen Museums. Im umliegenden archäologischen Park können weitere Denkmäler besichtigt werden. Die Saalburg befindet sich auf einer bewaldeten Anhöhe im Taunus nahe Bad Homburg. Damit stellt der Schauplatz die Anforderungen einer Indoor- und Outdoornavigation. Zudem wäre zu berücksichtigen, dass vor Ort keine Internetverbindung besteht, auch das Mobilfunknetz ist nur sehr schwach ausgebaut. Änderungen am Umfeld oder Anbringung von optischen Markern sind nicht möglich, da die Saalburg als Teil des Limes 2005 in die UNESCO-Liste des Welterbes aufgenommen wurde.

Ein weiterer Aspekt ist, dass eine Vielzahl der Besucher Kinder sind. Die Informationen sollen

¹ Hochschule für Technik Stuttgart, Schellingstr. 24, 70174 Stuttgart;
E-Mail: [habiburrahman.dastageeri, martin.storz, volker.coors]@hft-stuttgart.de

pädagogisch wertvoll, die App einfach in der Handhabung und optisch ansprechend sein sowie einen Spielecharakter aufweisen. Bei der Umsetzung des Konzepts muss des Weiteren berücksichtigt werden, dass die Lösung später auf andere Schauplätze übertragbar sein sollte.

3 Herangehensweise

Um den Anforderungen einer unterhaltsamen und vertrauten Informationsvermittlung gerecht zu werden, wurde die Möglichkeit eines virtuellen persönlichen Reiseführers gewählt. In Form eines erweckten Geistes eines ehemaligen Bewohners der Saalburg wird der User angesprochen und mittels einer Geschichte die Saalburg und sein Leben vorgestellt. Es werden mehrere Charaktere konzipiert, die wiederum verschiedene Geschichten erzählen. Damit soll die App auch bei mehrfacher Verwendung langfristig unterhaltsam bleiben. Die Geister werden verkörpert von Museumspädagogen, die auf die historische Korrektheit der Kleidung achten und bei der Erstellung der Geschichten mitwirken. Der Inhalt wird mehrsprachig in Deutsch, Englisch und Latein erstellt. Um den Aspekt der User Experience (UX) zu unterstreichen, sollen die User persönlich angesprochen werden. Dazu soll ein Tablet oder Smartphone als Medium dienen. Zur Gewährleistung einer möglichst einfachen Handhabung wurde die Implementierung einer ortsbezogenen und videobasierten AR App gewählt.

4 Umsetzung

Im ersten Schritt wurde ein Teaser als Proof of Concept erstellt. Gemeinsam mit einem Museumspädagogen wurde ein Einführungsvideo konzipiert und im Studio an der HSRM aufgenommen und entsprechend bearbeitet.

4.1 Generierung des Videoinhaltes

Die dargestellte Figur soll später in der AR-App als einzelne Figur über das aktuelle reale Kamerabild der Umgebung eingeblendet werden. Hierzu wurden die Aufnahmen im BlueScreen- bzw. GreenScreen-Verfahren mit einer Panasonic PHX 301 gefilmt. Aus dem aufgenommenen Rohmaterial wurden dann Videoschnipsel erzeugt und in sekundengenaue Sinnabschnitte unterteilt. Daraufhin wurde mittels des Chromakey Effekts die Hintergrundfarbe blau bzw. grün aus dem Bild gekeyed bzw. entfernt. Die Figur konnte so vom Hintergrund freigestellt werden. Weitere Schritte der Postproduktion waren die Bearbeitung der Aufnahmen mit bestimmten Videoeffekten. Nach Überprüfung diverse Effekte wurden die folgenden Einstellungen und Effekte vorgenommen, um die Figur einer Geistererscheinung möglichst nachzuempfinden. Abbildung 1 stellt die einzelnen Schritte dar. Der erste Bildabschnitt stellt das Rohmaterial im BlueScreen-Verfahren dar. Im zweiten Bildabschnitt wird die Figur mittels des Chromakey Effekts freigestellt. Zur Abschwächung der Konturen wird daraufhin ein Störeffekt angewandt und dann die Figur in einen Lilafarbtönen eingefärbt. Im fünften Schritt wird die Helligkeit erhöht. Um die Figur freischwebend einzublenden wird im sechsten Schritt ein Transparenzverlauf eingearbeitet. Schließlich soll die Assoziation von einem Geist durch einen Scheineffekt verstärkt werden. Das bearbeitete Video wird im Quicktime-Format mit dem Apple ProRes4444 Codec exportiert. Dieser Codec unterstützt die Darstellung transparenter und halbtransparenter

Flächen innerhalb eines Videos. Als letzter Bearbeitungsschritt wird das Video vom mov.-Format zwecks Plattformunabhängigkeit und Komprimierung in das mp4.-Format umgewandelt und kann nun als Video mit einem Alphakanal in der App verwendet werden.

4.2 Implementierung der AR

Um das Erlebnis einer AR-Anwendung gerecht zu werden, muss die Latenzzeit geringstmöglich sein. Sollte die App nicht verzögerungsarm auf den User reagieren können, würde die App nicht als AR-App wahrgenommen werden. AR-Apps charakterisieren sich nach AZUMA (1997) anhand von drei Eigenschaften:

- Kombination der Realität und Virtualität
- Verzugsarme Interaktivität
- Erfassung in 3D



Abb. 1 Verarbeitungsschritte des Videos

Bei der Kombination der Realität und Virtualität sind nach ALAN (2013) wiederum vier Schlüsselaspekte zu berücksichtigen:

1. Die physische Welt wird angereichert durch digitale Informationen indem über eine Ansicht der physischen Welt digitale Informationen eingeblendet werden.
2. Das AR-System erfasst autonom die Position in der realen Welt, in der Informationen eingeblendet werden sollen.
3. Die Informationen werden in Abhängigkeit der lokalen Position und Ausrichtung des Users in der physischen Welt dargestellt.
4. AR ist ein interaktives Erlebnis, d.h. eine Person kann Informationen wahrnehmen und diese wiederum in die jeweils gewünschten Informationen abändern. Der Grad der Interaktivität kann variieren zwischen einer einfachen Änderung der Ausrichtung bis hin zur Beeinflussung oder gar Erstellung neuer Informationen.

4.2.1 Spezifikation der Hard- und Software

Um den Anforderungen der AR gerecht zu werden, bedarf es eines Mobile Devices ausgestattet mit Kamera, Kompass, Lagesensoren, Bluetooth und GPS, Unterstützung von OpenGL ES 2.0 sowie den notwendigen Codecs für die Wiedergabe der Audio- und Videoformate. Für die Darstellung videobasierter AR ist die Unterstützung von Videos auf Texturen notwendig. Android bietet dies ab Version 4.0.3 (Min. Api Level 15) an.

4.2.2 Kombination von Realität und Virtualität

Entsprechend dem 1. Schlüsselaspekt nach A. B. ALAN werden die vorproduzierten Videoaufnahmen auf das Realkamerabild überlagert, allerdings sind Anpassungen im Bereich der Videodarstellung notwendig. Android bietet keine Unterstützung für den Alphakanal an. Die Transparenz wurde mittels des SurfaceTexture von OpenGL manuell implementiert. So konnte anhand des Fragment Shaders ein Alphakanal realisiert werden. Zur Steigerung der Performance konnten aufgrund der reinen Verwendung von Shadern die Berechnungen direkt auf der Grafikkarte ausgeführt werden. Diese Methode erwies sich als sehr effizient und ermöglichte ein flüssiges Abspielen von HD-Videos mit mindestens 24 Frames pro Sekunde.

4.2.3 Voraussetzung für eine verzugsarme Interaktivität

Bei einer ortsgebundenen AR-Anwendung ist eine schnelle und genaue Ortung ausschlaggebend. Eine reine GPS-Lokalisierung für den Outdoor-Bereich der Saalburg war systembedingt zu langsam und zu ungenau. Eine alternative schnellere Lösung wurde gesucht. Als erfolgreiche Methode stellte sich eine Kombination aus GPS und Image-Tracking heraus.

4.2.4 Ortungsverfahren

Im ersten Schritt werden Referenzbilder der relevanten Ort erstellt, die ein Ereignis triggern sollen. Hierbei ist zu beachten, dass im Außenbereich das Wetter berücksichtigt werden muss. Vom selben Ort sollten Bilder in den verschiedenen Jahreszeiten vorliegen. Besonders muss hier auf Schnee und unterschiedliche Belaubung Rücksicht genommen werden. Diese Bilder werden dann den Orten zugeordnet. Per GPS kann für den folgenden Bildvergleich so aus der Gesamtbilderdatenbank die in Frage kommenden Bilder selektiert werden.

Im zweiten Schritt werden die Referenzbilder mit dem Realkamerabild mittels des Image-Tracking-Verfahrens verglichen. Bei erfolgreichem Bildabgleich kann so die genaue Position und Ausrichtung des Tablets festgelegt werden.

Im dritten und letzten Schritt werden die Positionsdaten ausgewertet und entsprechende Ereignisse berührungsfrei automatisch getriggert. Je nach Reaktion des Users reagiert die App und leitet den User weiter an.

4.2.5 Image-Tracking-Verfahren

Beim Image-Tracking-Verfahren werden Gemeinsamkeiten zwischen dem Referenzbild und dem Realkamerabild gesucht. Hierzu werden markante Merkmale, die Keypoints, kontinuierlich aus dem Kamerabild extrahiert und mit Referenzbildern abgeglichen. Je größer die Übereinstimmung, umso kleiner die Differenz. Die Methode verwendet den ORB-Algorithmus von RUBLEE et al (2011) der Open-Source Bibliothek Open Computer Vision (OpenCV). Vier Werte sind für den Vergleich relevant; die Anzahl der Vergleichspunkte, die minimale Differenz (Min_dist), maximale Differenz (Max_dist) und die durchschnittliche Differenzen (Avg_dist). Zur Qualitätssteigerung der Ergebnisse wurden drei Filtermethoden implementiert. Der erste Filter führt einen symmetrischen Abgleich aus. Hierbei werden Vergleichsbilder von Bild 1 in Bild 2 gesucht und darauf folgend die gleichen Vergleichspunkte in entgegengesetzter Reihenfolge von Bild 2 in Bild 1 gesucht. Alle Vergleichspunkte, die diese Eigenschaft nicht besitzen, werden gelöscht. Der zweite Filter schließt alle Vergleichspunkte mit schlechten

Bewertungen aus. Hierbei handelt es sich um Vergleichspunkte, deren Max_dist den doppelten Wert von Min_dist übersteigen. Beim dritten und letzten Filter handelt es sich um den RANSAC Filter nach FISCHLER & BOLLES (1981). Hierbei werden entfernte Punkte gefunden, als Ausreißer erkannt und anschließend gelöscht.

Zur Minimierung der Latenzzeit wurden Berechnungen zur Definierung der Anzahl der Keypoints durchgeführt. Das Referenzbild hatte dabei eine Auflösung von 175x395 Pixel, das Realkamerabild 400x300 Pixel. Tabelle 1 stellt die Messwerte bei variierten Keypoints dar. Erfolgreiche Zuweisungen konnten ab 2500 Keypoints durchgeführt werden. Die Messungen fanden auf einem Samsung S3 mit einem 1,4 GHz Quad-Core ARM Cortex-A9 und 1GB RAM statt.

Tab 1 Auszug aus den Testergebnissen der zeitlichen Messungen

| Messungen | Keypoints | Minimale Zeit | Maximale Zeit | Durchschnitt |
|-----------|-----------|---------------|---------------|--------------|
| 100 | 500 | 439 ms | 1246 ms | 468 ms |
| 100 | 750 | 527 ms | 701 ms | 571 ms |
| 100 | 2500 | 1120 ms | 1437 ms | 1194 ms |
| 100 | 7500 | 1604 ms | 1941 ms | 1699 ms |

4.3 Video als AR-Element

Sobald das Video eines Geistes getriggert werden soll, wird ein Referenzfoto der realen Umgebung automatisch erstellt. Wenn das Realkamerabild mit dem Referenzfoto übereinstimmt, wird der Geist auf dem Realkamerabild platziert und das Video beginnt. Hierbei wird die Größe und Position des Geistes auch bei Änderungen des Smartphones beibehalten. Verwendet wurde das "User Defined Targets"-Verfahren mit extended tracking von QUALCOMM VUFORIA (2015).

Beim Darstellen des Videos musste berücksichtigt werden, dass derzeit der Android MediaPlayer kein Videoformat mit Alpha Channel unterstützt. Eine eigene Lösung musste implementiert werden. WIKITUDE (2015) hat das Problem gelöst, indem es die Videohöhe verdoppelt. Im oberen Bild ist das Video als RGB-Video gespeichert, im unteren Bereich wird der Alpha-Wert durch Grautöne dargestellt, wobei hier gilt: Schwarz = unsichtbar, Weiß = voll sichtbar.

Dieser Lösungsansatz wird im Projekt SPIRIT durch eine eigene Implementierung genutzt. Abbildung 2 stellt ein so für SPIRIT aufbereitetes Video dar.

4.4 ARML 2.0

Zukünftig soll die SPIRIT-App auch für weitere Museen und touristisch attraktiven Einrichtungen einsetzbar sein. Eine einfache standardisierte Möglichkeit zur Erstellung und Veränderung der Abläufe der Geschichten des Contents ist eine wesentliche Anforderung. Eine Möglichkeit wäre die Verwendung der Augmented Reality Markup Language (ARML) entsprechend der Version (OGC) ARML 2.0 SWG. SPIRIT wurde dahingehend erweitert und verfügt über ein Interface für ARML 2.0 Dokumente mit den folgenden Parametern: Points of Interest (POI) Name, Longitude, Latitude, Video Name (im mp4.-Format) und Referenzbild Name. Die Anordnung und die Anzahl der POIs sowie deren Parameter (Position, Referenzbild und das jeweils passende Video) können nun unabhängig von der Android-App verändert und angepasst werden.

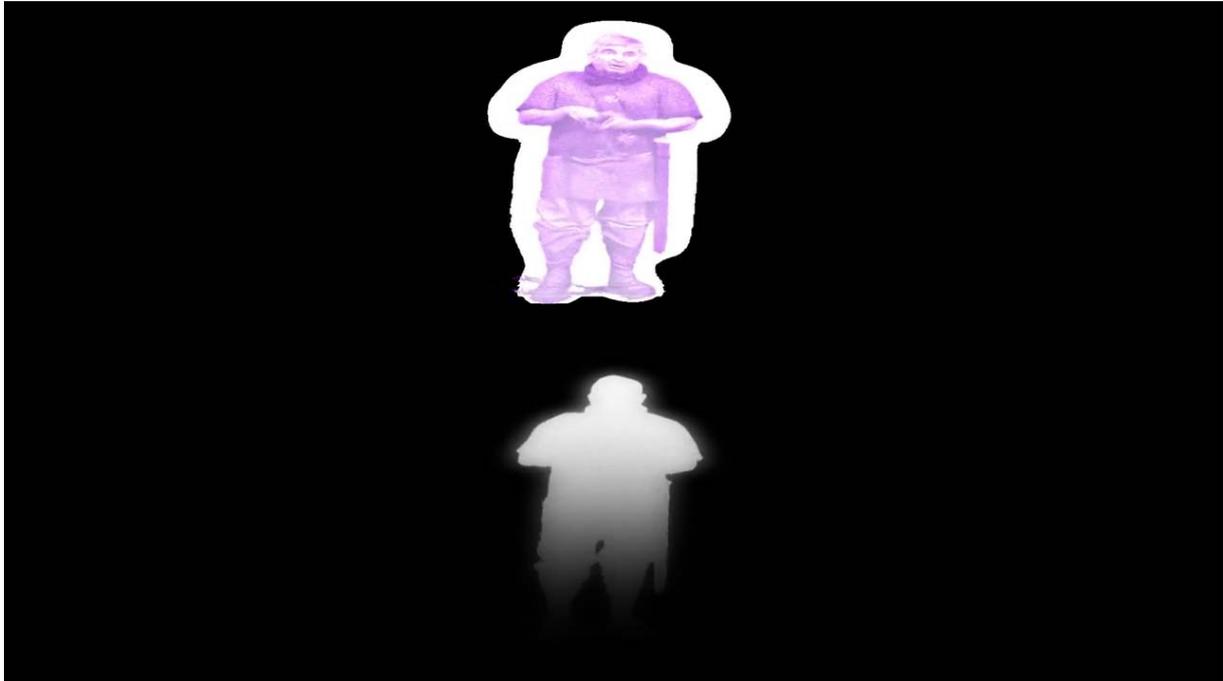


Abb. 2 Transparenz des Videos

5 Fazit und Ausblick

Eine geringe Latenzzeit ist ausschlaggebend für den Gebrauch einer AR-App. Wenn eine erhoffte Reaktion ausbleibt, erwartet der User eine prompte Rückmeldung zum weiteren Vorgehen. Eine Hilfefunktion könnte beispielsweise Hinweise zur Änderung der Position oder Ausrichtung des Smartphones geben. Besonders in Bereichen mit einer notwendig hohen Ortungsgenauigkeit sind schnelle Rückmeldungen entscheidend. Daher soll zukünftig auch im Innenbereich eine Selektierung der Referenzbilder stattfinden. Nach erfolgreicher Indoor-Navigation werden dann nur die in Frage kommenden Referenzbilder zum Bildabgleich verwendet. Gegenwärtig wurde eine Ortung mittels Beacons prototypisch umgesetzt. Diese Funktion soll zukünftig verfeinert und einbezogen werden.

Unter Beachtung des schnelllebigen Marktes für Mobile Devices sollten weitere Plattformen berücksichtigt werden. Bei der Implementierung werden gegenwärtig Android-spezifische Features näher betrachtet und nach möglichen Lösungen für eine iOS-Unterstützung gesucht, um eine zukünftige und aufwendungsarme iOS-Implementierung zu ermöglichen.

6 Danksagung

Das Projekt SPIRIT wird unter dem Förderkennzeichen 03FH035PA3/B3 vom BMBF gefördert. Dank gilt K. Stöbener und A. Kampa von der Hochschule RheinMain und dem Saalburgmuseum für die Unterstützung bei der Inhaltsgestaltung.

7 Literaturverzeichnis

- AZUMA, R. T., 1997: A Survey of Augmented Reality. *Presence: Teleoperators and Virtual Environment*, **6** (4), S. 355-385
- ALAN, A. B., 2013: *Understanding Augmented Reality, Concepts and Application*. Elsevier Inc., S. 16
- KAMPA, A., DASTAGEERI, H, STORZ, M, COORS, V. & SPIERLING, U., 2014: SPIRIT – Ereignisgesteuerte Informationsvermittlung, Inspiration und Unterhaltung im urbanen Umfeld auf Basis mobile Augmented Reality Technologien. 11. Workshop der GI-Fachgruppe VR/AR, Tagungsband *Virtuelle und Erweiterte Realität*, Gabriel Zachmann, René Weller, André Hinkenjann (Hrsg.), Shaker Verlag, ISBN 978-3-8440-3054-9
- RUBLEE, E., RABAUD, V., KONOLIGE, K. & BRADSKI, G. R., 2011: ORB: An efficient alternative to SIFT or SURF. *Proceedings of IEEE Computer Vision (ICCV 2011)*, S. 2564-2571.
- FISCHLER, M. A. & BOLLES, R. C., 1981: Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Proceedings of Communications of the ACM* **24**(6), S. 381-395.
- QUALCOMM VUFORIA, 2015: Vuforia Features – User Defined Targets.
<https://developer.vuforia.com/resources/sample-apps/user-defined-target-sample-app>,
aufgerufen am 19.01.2015
- WIKITUDE, 2015: Bonus: Transparent Video (4/4). Wikitude SDK Web Documentation version 4.0.
<http://www.wikitude.com/external/doc/documentation/latest/web/video.html#transparentvideo>,
aufgerufen am 19.01.2015.
- OPEN GEOSPATIAL CONSORTIUM, 2015: AMRL 2.0 Standard Working Group.
<http://www.opengeospatial.org/projects/groups/arm2.0swg>, aufgerufen am 19.01.2015.