

# Vorlesungsnachverfolgung zur Unterstützung von Hybrid-Vorlesungen

Prof. Dr.-Ing. Hartwig Grabowski, Philipp Schweizer B.Sc.

Es wird ein System zur Vorlesungsnachverfolgung (engl. Lecture Tracking) vorgestellt, das eine Kamera automatisch in Richtung des Vortragenden ausrichtet. Die kontinuierliche Positionsbestimmung der Kamera als auch die des Vortragenden erfolgt dabei durch Smartphones. Die Kamera und ein Smartphone sind an einen Roboter montiert, der als Schwenkeinheit dient. Das andere Smartphone trägt der Vortragende. Beide Smartphones können ihre Position im Raum bestimmen, sodass der erforderliche Drehwinkel berechnet und an den Roboter gesendet werden kann. Dieser führt dann die entsprechende Rotation durch, sodass der Vortragende immer in der Bildmitte zu sehen ist.

*A real-time lecture tracking system is presented, which uses smartphones for continues location detection. Together with a camera, one smartphone is mounted on a robot, which serves as a pan-tilt-unit. The other smartphone must be worn by the speaker. Both smartphones detect their location in the room using spatial anchors and SLAM tracking. They exchange their information through a common server in the cloud. Based on this information, the required angle of rotation for the camera can be calculated and the according control signals can be sent to the robot.*

## Motivation

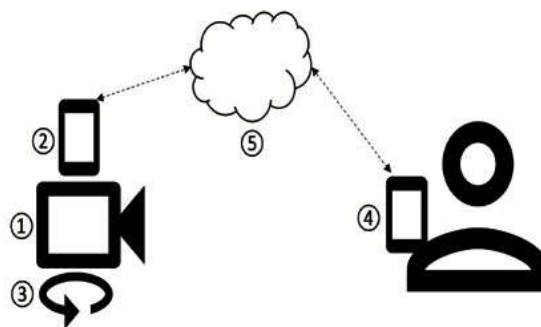
Wird bei Vorlesungen der Vortragende und dessen Tafelaufschrieb per Video übertragen, muss die Kamera dem Vortragenden folgen, damit dieser nicht aus dem Bildschirm verschwindet. Weiterhin muss sichergestellt werden, dass das Mikrofon immer auf den Vortragenden ausgerichtet ist, damit er klar und deutlich zu hören ist. Dies erfordert entweder einen zusätzlichen Kameramann oder zusätzliches teures Equipment, wie beispielsweise Infrarot-Kameras oder Weitwinkel-Kameras [1, 2, 3, 4]. Wir präsentieren ein auf Smartphones basierendes Trackingsystem, welches eine PTZ-Kamera oder ein entsprechender Roboter steuern kann.

## Systemarchitektur

Das hier vorgestellte System besteht aus folgenden Komponenten:

1. Die Kamera, die den Vortragenden aufnimmt und per Stream den Zuhörern bereitstellt. Wir verwenden hierfür ein gewöhnliches Smartphone, welches das Videobild live via Zoom-App den Zuhörern bereitstellt.
2. Ein weiteres Smartphone, welches starr mit der Kamera verbunden und entsprechend der Kamera ausgerichtet ist. Dieses bestimmt neben der Position der Kamera im Raum auch die Ausrichtung der Kamera.

3. Die Schwenkeinheit, auf der sowohl Kamera als auch Smartphone montiert sind. Hierfür verwenden wir den Double-Robotics Telepräsenz-Roboter [5], der durch das montierte Smartphone (2) gesteuert wird.
4. Ein weiteres Smartphone, welches der Vortragende trägt. Dieses bestimmt die genaue Position des Vortragenden im Raum.
5. Einen zentralen Rechner in der Cloud, über den die beiden Smartphones (2) und (4) ihre Positionsinformationen austauschen.



**Abb. 1:** Kameraseitig werden die drei Komponenten Kamera (1), Smartphone (2) und Schwenkeinheit (3) benötigt. Aufseiten des Vortragenden reicht ein Smartphone (4) zur Positionsbestimmung. Die beiden Smartphones tauschen ihre Positionsinformation über einen Server in der Cloud (5) aus.

## Referenzen/References:

- [1] Yoshitaka, A., Nomi, K., Kawano, A., Hirashima, T. "Speaker Tracking for Automated Lecture Archiving Using Tagged Microphone", Tenth IEEE International Symposium on Multimedia, ISM 2008, pp. 45-52
- [2] Vitaglione, et al., "Lecture Object: An Architecture For Archiving Lectures On The Web," CERIN-OPEN-2001-070 (2001)
- [3] Rui, Y., Gupta, A., Grudin, J., and He, L. "Automating lecture capture and broadcast: Technology and videography," ACM Multimedia. Systems Journal, 10, 1, 2004, pp. 3-15
- [4] Tan, T. H., Kuo, T. Y., & Liu, H. (2019). „Intelligent Lecturer Tracking and Capturing System Based on Face Detection and Wireless Sensing Technology". Sensors (Basel, Switzerland), 19(19), 4193. <https://doi.org/10.3390/s19194193>
- [5] Double Robotics, Inc. (2022), „Double 2 – Features", <https://www.doublerobotics.com/double2.html>. Accessed 1.3.2022
- [6] Microsoft (2020), „Azure Spatial Anchors overview", <https://docs.microsoft.com/de-de/azure/spatial-anchors/overview>. Accessed 1.3.2021
- [7] Google Inc. (2018a), „ARCore, Fundamental Concepts", <https://developers.google.com/ar/discover/concepts>. Accessed 1.3.2021
- [8] Grabowski, H. „Physical Distancing mit Spatial Anchors", forschung im fokus (ffif), 07/2021, pp. 37-40
- [9] Grabowski, H. „Lecture Tracking", <https://www.youtube.com/watch?v=Ygek1cthrKo>. Accessed 1.2.2022

## Positionsbestimmung

Die Positionsbestimmung der Smartphones erfolgt initial durch einen virtuellen Referenzpunkt (Spatial Anchors), der zuvor im Raum gesetzt wurde. Die Platzierung und Erfassung dieses virtuellen Referenzpunkts erfolgt durch den von Microsoft bereitgestellten Azure Spatial Anchors Dienst [6]. Dieser berechnet aus den Kamerabildern des Smartphones eine charakteristische Punktwolke, mit deren Hilfe eine Position im Raum eindeutig bestimmt werden kann. Nach dieser initialen Positionsbestimmung erfolgt nun auf jedem der beiden Smartphones (Smartphone an der Kamera und Smartphone des Vortragenden) ein SLAM-basiertes Tracking auf Basis von Google AR Core [7]. Dieses zur Positionsbestimmung eingesetzte Verfahren wurde von uns bereits in [8]. beschrieben. Neben der Bestimmung der genauen Position der beiden Smartphones im Raum wird jetzt jedoch zusätzlich die Ausrichtung der Smartphones ermittelt, sodass in dem an der Kamera montierten Smartphone der Drehwinkel berechnet werden kann, um den die Kamera gedreht werden muss, damit sie auf den Vortragenden ausgerichtet ist. Dieser Drehwinkel wird an die Schwenkeinheit - bei uns der Double-Robotics Roboter - gesendet, die dann die entsprechende Rotation durchführt.

### Abb. 2:

Das Roboter-System im Einsatz (links) und die an den Roboter montierten Smartphones (rechts)



## Fazit

Bei Lecture-Tracking-Systemen, die auf Basis von Gesichtserkennung oder IR-Kameras basieren, kann das Tracking abbrechen, wenn sich der Vortragende zu schnell aus dem Bild entfernt oder hinter einem Pult verschwindet. Das erneute Erfassen des Vortragenden kann dann problematisch werden, weil der Vortragende möglicherweise wieder außerhalb des Sichtbereichs des Kamerasystems erscheint. Dies kann bei dem hier vorgestellten System nicht passieren. Selbst wenn der Vortragende den Raum verlässt und kein Sichtkontakt zwischen Kamera und Vortragendem besteht, funktioniert die Ausrichtung korrekt.

Die Positionsbestimmung durch den virtuellen Referenzpunkt und das anschließende Tracking haben sich als hinreichend stabil herausgestellt, sodass auch große Bewegungen des Vortragenden über längere Zeit hinaus hinreichend genau erfasst werden [9].

Statt dem Double-Robotics-Roboter kann auch eine kostengünstige Pan-Tilt-Kamera-Plattform verwendet werden. Weil der Abstand des Vortragenden zur Kamera berechnet wird, kann auch der Kamera-Zoom entsprechend gesteuert werden, sofern dies die Kamera zulässt. Da das hier vorgestellte Gesamtsystem aus mehreren Komponenten besteht, ist die initiale Einrichtung recht aufwendig, sodass für einen routinemäßigen Einsatz dauerhaft vorkonfigurierte Komponenten zu verwenden wären.

### AUTOREN



Prof. Dr.-Ing. Hartwig Grabowski  
Institute for Machine Learning and Analytics (IMLA), Mobile Computing, Enterprise Applications  
[hartwig.grabowski@hs-offenburg.de](mailto:hartwig.grabowski@hs-offenburg.de)



Phillip Schweizer B.Sc.  
[philipp.schweizer@hs-offenburg.de](mailto:philipp.schweizer@hs-offenburg.de)