

# SIKOWO

## Gerätefernbedienung durch Zeigehandlungen

Peter Barth  
FORWISS  
Universität Passau  
email: barth@forsip.de

Klaus Donner  
FORWISS  
Universität Passau  
email: donner@forsip.de

**Abstract:** In dieser Arbeit wird die in SIKOWO angefertigte Gerätefernbedienung PoCoDeR (POinting gestures to COntrol DEvices Remotely) vorgestellt, die sich bequem und einfach durch Zeigegesten bedienen lässt. Die Handgesten werden dabei mit zwei HDR-Kameras (high-dynamic-range) aufgenommen. Die rekonstruierte Szene wird durch ein 3D Raummodell dargestellt. Es bietet unterschiedliche Möglichkeiten um neue zu steuernde Geräte hinzuzufügen, wobei diese über den European Installation Bus (EIB) gesteuert werden. Ein kalibriertes Stereokamerasystem dient zur Gewinnung der 3D-Daten. Auf den gelieferten Daten wird dann ein Graph von bewegten Konturen erstellt, der zur Bestimmung der Zeigegesten verwendet wird. Jede erkannte, vordefinierte Geste entspricht dabei einem “Benutzerwunsch” und wird durch das kontextsensitive Steuerungs-/Regelungssystem in eine Aktion zum Schalten eines technischen Geräts umgesetzt.

## 1 Einleitung

Viele alltägliche Regelungsvorgänge werden mit zunehmender Technisierung immer komplexer und irritieren den Benutzer durch Reaktionen, die dem eigentlichen Willen des Interagierenden nicht entsprechen. Zum Beispiel stellen Autofahrer im Hochsommer beim Einstieg in ihr Auto die Klimaanlage sofort auf die minimalste Temperatur. Dies entspricht nicht unbedingt dem Gewünschten, da lediglich versucht wird, die eigentlich gewünschte Temperatur möglichst schnell zu erreichen.

Im Projekt SIKOWO<sup>1</sup> wurde eine typische Sitzungssituation mit Kurzvorträgen betrachtet. Für einen Vortrag sind andere Lichtverhältnisse als während der Diskussion erforderlich. Ein offenes Fenster sollte bei vorbei fahrendem Zug geschlossen werden. Die Heizung ist bedarfsgerecht zu führen. Bei schräg einfallender Sonne auf die Projektionsfläche sind die Jalousien herabzulassen, usw. Die Regelung dieser Umgebungsbedingungen kann zu lästigem Herumhantieren an Regelungskonsolen u.ä. führen. Neben den direkten regelnden Eingriffen durch den Menschen steht im Projekt SIKOWO der geäußerte Gesamtwunsch des Benutzers im Mittelpunkt, der fallweise durch ganz verschiedenartige Reaktionen des Regelungssystems befriedigt werden kann. So kann der Wunsch “vorne etwas heller” sowohl durch Heraufziehen der Jalousien wie durch Zuschalten von Licht erfüllt werden. Dies erfordert die Erfassung des Regelungskontexts, also die Feststellung und Überwachung der Situation im Raum (vgl. [Kel04]).

Die vorliegende Arbeit soll einen Überblick über die im Projekt SIKOWO entwickelte Gerätefernbedienung geben. Dabei werden sowohl die so genannten Zeigehandlungen beschrieben als auch das Wunschsystem, das für die Umsetzung der Wünsche

---

<sup>1</sup>[http://www.forsip.de/?show=projekte\\_sikowo2&page=1](http://www.forsip.de/?show=projekte_sikowo2&page=1)

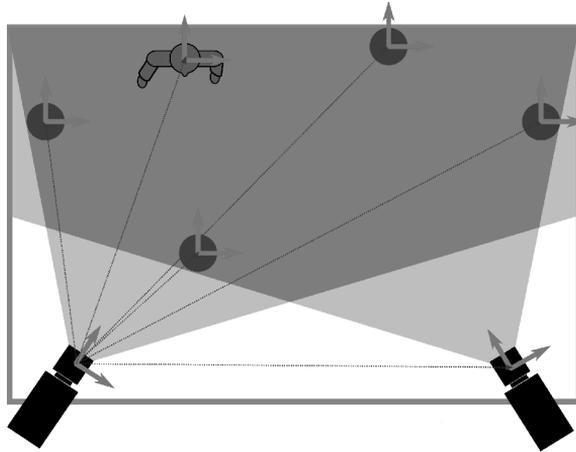


Abbildung 1: Seminarraums mit Stereokamerasystem und einigen technischen Geräten

in konkrete Regelungsaufgaben anhand erweiterter Kontextinformationen zuständig ist.

## 2 Problemstellung und Use-Cases

Bereits am Beispiel des Sitzungsraumszenarios sieht man schnell, dass allein die Anzahl der verschiedenen technischen Geräte einem typischen Benutzer schnell den Überblick verlieren lässt. So sind z.B. im verwendeten Sitzungsraum mehr als 6 Schalter nur zur Steuerung des Lichts vorhanden, in Vorlesungsräumen sogar noch bedeutend mehr.

Ein Sitzungsraum stellt das Testszenario für das in SIKOWO entwickelte Assistenzsystem dar. Es wird ermöglicht, über einfache Gesten mit dem System zu interagieren. Dabei ist es wichtig, die Steuerung über das kontextsensitive Wunschsystem abzuwickeln, so dass das System abhängig von Tageszeit, Raumhelligkeit, usw. selbst entscheiden kann, wie einem Wunsch idealerweise nachgekommen werden soll. Um durch das neue System nicht eine erneute Komplexität zu generieren soll das System ausschließlich durch Gesten bedient werden. Es sind somit keine weiteren Eingabegeräte zur Steuerung bzw. Interaktion mit dem System notwendig.

## 3 Lösungskonzept

### 3.1 Systemaufbau

Der Hauptsensor der Gerätefernbedienung stellt ein Stereokamerasystem dar. Es wird mit einem Abstand von ca. 3 Metern so aufgebaut, dass der vordere Bereich des Raumes vollständig im Blickfeld beider Kameras ist. Durch eine Stereokalibrierung der beiden Kameras lässt sich für ein in beiden Kamerabildern erkanntes Objekt die exakte 3D-Position im Raum bestimmen. Ausßerdem ist der komplette Sitzungsraum mit EIB-fähigen Geräten ausgestattet. So lassen sich alle vorhandenen Lampen einzeln schalten und dimmen, aber auch die Jalousie und Sonnenblende sind regelbar. Damit

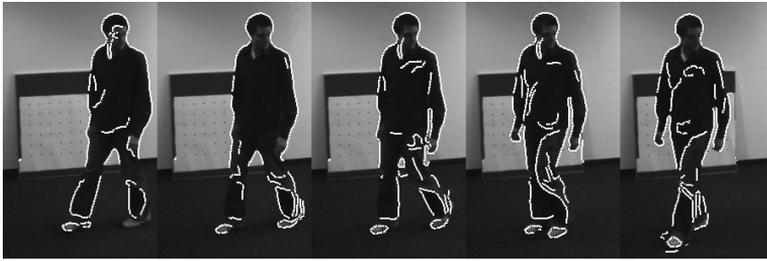


Abbildung 2: Videosequenz der bewegten Konturen einer Person

diese Geräte auch durch den Computer steuerbar sind, wurde neben dem Stereokamerasystem auch der EIB-Bus über eine USB-Schnittstelle mit dem Rechner verbunden.

Zur Systemkalibrierung sind verschiedene Initialisierungsarbeiten zu erledigen. Ein wichtiger Schritt besteht dabei in der Einzelkalibrierung und der anschließenden Stereokalibrierung der beiden Kameras (vgl. [HGP04] und [Han06]). Sind Position und Lage der Kameras bekannt, kann eine Initialisierung der Raumgeometrie stattfinden. Von Bedeutung sind hier die Wand, Decke und Boden des Raumes, die es im Systembetrieb erlauben, die Position einer Person zu erfassen.

### 3.2 Körperteilklassifikation

Nach der Systeminitialisierung kann die Gerätefernbedienung in Betrieb genommen werden. Das Stereokamerasystem nimmt dann kontinuierlich den Sitzungsraum auf und versucht, Personen zu lokalisieren und zu verfolgen. Dabei werden nur sich verändernde Regionen im Kamerabild ausgewertet, um die benötigte Rechenleistung möglichst gering zu halten (vgl. [KH04]). In den bewegten Bildregionen findet eine Personenerkennung statt. Dazu wird in einem ersten Schritt über eine Klassifikation der Kopf einer Person detektiert und mit der Position im vorhergehenden Bild abgeglichen, um eine zeitliche Verfolgung der Person durchzuführen. Ausgehend vom Kopf der erkannten Person werden dann Oberkörper, Beine und vor allem Arme und ihre räumliche Lage detektiert. Durch die Körperteilklassifikation lassen sich die verschiedenen Benutzerwünsche erkennen und schließlich in konkrete Aktionen umsetzen. Für die Wunschäußerung des Benutzers wurden Armgesten verwendet. Ein einfaches Zeigen auf ein technisches Gerät oder auf eine zuvor markierte Positionen im Raum wird durch das System als Wunsch interpretiert. Als Beispiel kann das Zeigen auf die Deckenbeleuchtung als Wunsch nach mehr Helligkeit interpretiert werden.

### 3.3 Kontextsensitive Regelung

Ein entscheidender Faktor im Assistenzverhalten der Gerätefernbedienung stellt das kontextsensitive Regelungssystem dar. Es handelt sich um einen endlichen Automaten für den Zustandswechsel der zu steuernden Geräte unter Einbeziehung von Kontextwissen. Zur Erfassung des Kontextes gibt es verschiedene Sensoren, die sich vorwiegend auf Bildverarbeitung stützen. Es kann z.B. der Zustand eines Fensters oder einer Tür beobachtet und sehr genau der Öffnungsgrad bestimmt werden (vgl. [Kel04]). Darüber hinaus wurden auch Sensoren zur Erfassung der Raumhelligkeit oder auch des Sonnenstands entwickelt.

Bei Auftreten eines Wunsches entscheidet das System nun, wie der Wunsch um-

gesetzt werden soll. Dazu unterscheidet man verschiedene Arten von Wünschen, die durch den Benutzer geäußert werden können. Zum einen wird zwischen abstrakten und konkreten Wünschen unterschieden. Der Wunsch nach “mehr Helligkeit im Raum” würde dabei als abstrakt eingestuft, wohingegen der Wunsch nach “mehr Helligkeit durch Einschalten der Deckenlampe” einen konkreten Wunsch darstellt. Zum anderen wird zwischen absoluten und relativen Wünschen unterschieden. Wird der Wunsch nach “mehr Helligkeit im Raum” gestellt, so ist dies als relativer Wunsch zu werten, da das System selbst bestimmen muss um wieviel heller der Raum werden soll. Hingegen ist der Wunsch “mach das Licht an” ein absoluter Wunsch und kann durch das System direkt umgesetzt werden.

Die Umsetzung der Wünsche geschieht im Wesentlichen in 2 Stufen. Jeder abstrakte Wunsch wird in einem ersten Schritt in einen konkreten Wunsch umgesetzt, wobei die Umsetzung vom aktuellen Kontext abhängt. Es ist z.B. nicht möglich, Nachts durch Öffnen der Jalousie eine Erhellung eines Raumes zu bewirken. Relative Wünsche wie z.B. “mach das Licht heller” haben einen größeren Spielraum zur Umsetzung. Das System muss erneut basierend auf den Kontextinformationen entscheiden, um wieviel Prozent die Lichtstärke erhöht wird. Es werden somit alle Wünsche in konkrete und absolute Wünsche überführt, die schließlich durch das System ausgeführt wird.

## 4 Ergebnisse

Die im Sitzungsraum aufgebaute und getestete Gerätefernbedienung erzielte eine hohe Erkennungsrate (vgl. [BKH07]), wurde jedoch teilweise durch stark wechselnde Hintergründe erschwert. Das Wunschsysteem unter Berücksichtigung von Kontextinformationen ist auch für Benutzer, die das System nicht kennen, gut geeignet. In zukünftigen Arbeiten sollte jedoch die Sensorik zur Erfassung des Kontexts noch weiter ausgebaut werden und auch Sensorsysteme Verwendung finden, die nicht auf Bildverarbeitung basieren.

## Literatur

- [BKH07] P. Barth, M. Kellner und T. Hanning. PoCoDeR: Device control by pointing gesture. In *International Conference on Artificial Intelligence and Applications (AIA)*, 2007.
- [Han06] T. Hanning. Re-Projective Calibration of a Stereo Camera System. In *International Conference on Visualization, Imaging and Image Processing (VIIP)*, Seiten 7 – 12, Palma de Mallorca, Spain, 2006.
- [HGP04] T. Hanning, S. Graf und G. Pisinger. Extrinsic Calibration of a stereo camera system fulfilling generalized epipolar constraints. In *International Conference on Visualization, Imaging and Image Processing Conference (VIIP)*, Seiten 1 – 5, Marbella, Spain, 2004.
- [Kel04] Michael Kellner. Image Processing Modules in Smart Homes. *KONNEX Scientific Conference*, 2004.
- [KH04] M. Kellner und T. Hanning. Motion Detection based on Contour Strings. In *International Conference of Image Processing (ICIP)*, Seiten 2599 – 2603, Singapore, October 2004. IEEE.