Interaktion im öffentlichen Raum: Von der qualitativen Rekonstruktion ihrer multimodalen Gestalt zur automatischen Detektion mit Hilfe von 3-D-Sensoren

Mukhametov, Sergey
s.mukhametov@gmail.com
Institut für Geoinformatik, WWU Münster, Deutschland

Kesselheim, Wolfgang
wolfgang.kesselheim@ds.uzh.ch
Sprache und Raum Laboratorium, Universität Zürich, Schweiz

Brandenbeger, Christina
c.brandenberger@access.uzh.ch
Sprache und Raum Laboratorium, Universität Zürich, Schweiz

Wenn Menschen miteinander interagieren, ist dies für kompetente Gesellschaftsmitglieder selbst aus einem gewissen Abstand heraus leicht zu erkennen: Die Körper von Interaktionsbeteiligten sind auf spezifische Art und Weise aufeinander ausgerichtet. Sie bilden einen "Interaktionsraum". Die Herstellung, Aufrechterhaltung und Auflösung von Interaktionsräumen sind in Soziologie, Psychologie und Linguistik untersucht worden – das in den letzten Jahren vor allem basierend auf Videoaufnahmen authentischer Interaktionsereignisse. Dabei basieren die Erkenntnisse auf feinkörnigen qualitativen Analysen, also auf einer sehr zeitintensiven Auseinandersetzung mit Einzelfällen.

Ziel unseres Vortrags ist es nun, zu zeigen, wie sich die bisherigen Ergebnisse der qualitativen Forschung zu Interaktionsräumen nutzen lassen, um Interaktionsräume automatisch detektieren zu können. Hierfür erheben wir multimodale Datensätze aus 2-D-Videos, Annotationen und 3-D-Sensordaten. Diese machen es mo#glich, pra#zise Informationen zu Bewegungen, Verhalten, und sozialen Interaktionen einer gro#ßeren Anzahl von Probanden automatisiert und beru#hrungsfrei zu erheben und im ra#umlichen Kontext zu analysieren. Während z.B. Computer-Vision-basierte Methoden zur Erkennung von Konversationsgruppen (z.B. Setti et al 2013) die Positionen einzelner Körperteile oft nur unzureichend erkennen, berechnet unsere Methode die präzisen Positionen und Distanzen von Skelett-Daten im 3-D-Raum.

Mit der Darstellung unserer Methode möchten wir exemplarisch aufzeigen, wie ethnografisch oder konversationsanalytisch arbeitende Studien quantitative Ansätze fruchtbar machen können. Unser Beispielfall ist die Untersuchung der körperlich-räumlichen Formationen, die für die soziale Praxis des Museumsbesuchs charakteristisch sind. Mit der quantitativen Auswertung von Interaktionsräumen wird es möglich, im Vergleich zu den bisherigen Studien die Datenbasis beträchtlich zu erweitern, von der aus Generalisierungen zu den typischen Mustern und Merkmalen von Interaktionsräumen formuliert werden. Darüber hinaus lassen sich mit Hilfe der quantitativen Auswertung von körperlich-räumlichen Formationen schnell interessante Fälle der Exponatnutzung oder der Interaktion mit anderen Besuchern identifizieren, die dann qualitativ im Detail untersucht werden können.

Während diese technischen Methoden die Untersuchung von Interaktionsräumen erheblich beschleunigen, bergen sie auch Herausforderungen: die Integration der Daten zeitgleich aufnehmender Tiefenbildkameras, die Integration von geometrischen Körperprofilen in den raum-zeitlichen Ablauf, und die raum-zeitliche Analyse mehrerer Bewegungsprofile zur Identifikation von Gruppen und sozialen Interaktionen, und zur Definition von Interaktionsräumen. Zudem weichen die tatsächlichen Formationen interagierender Personen teilweise auch stark von den in der Literatur beschriebenen ab. Die automatische Erfassung macht es daher notwendig, die bestehenden Definitionen zu verschärfen und zu erweitern (aber liefert eben auch die Grundlagen für diese Arbeit an den Definitionen).

Schon in den 1970er Jahren haben sich Forscherinnen und Forscher aus Soziologie und Psychologie für die räumlichen Aspekte der sozialen Interaktion interessiert. So geht Goffman (1963, 1971) aus alltagssoziologischer Perspektive der Frage nach, unter welchen Bedingungen Interaktion in Gang kommen kann, wenn Menschen sich im öffentlichen Raum begegnen. Inspiriert durch Goffman, untersucht der Psychologe Adam Kendon (z.B. 1990 [1973, Ciolek / Kendon 1980) die räumlichen Bedingungen des Entstehens von Interaktion. Anders als Goffman, dessen Analysen auf direkter visueller Beobachtung und Feldnotizen beruhen, arbeitet Kendon erstmals mit Kameras und der detaillierten multimodalen Transkription und Annotation der aufgenommenen Daten. Dabei interessiert sich Kendon nicht nur für den Beginn der Interaktion, sondern zunehmend auch für deren Aufrechterhaltung durch die dynamische Abstimmung der Körperpositionen der Interaktionsteilnehmer im Raum. Zentral sind hier die Konzepte der "F-Formation" und des "o-space", die bis in die aktuelle Forschung hinein folgenreich geblieben sind. Mit "F-Formation" ist eine dynamische Konfiguration der Interaktionsteilnehmer im Raum gemeint, die die räumlichen Voraussetzungen für das gemeinsame Interagieren zur Verfügung stellt. In einer F-Formation bringen die Interaktionspartner ihre "transactional segments" zur Übereinstimmung, also die Raumbereiche, in denen ihr Körper mit

der Umwelt in Kontakt treten kann. Durch eine F-Formation entsteht der "o-space": der Raum, der von den Interaktionsteilnehmern umständen wird, in den hinein ihr Sprechen und Gestikulieren gerichtet ist und dem deshalb ihre Aufmerksamkeit gewidmet ist. Alle Teilnehmer haben gleichberechtigten Zugang zu diesem Raum und grenzen ihn gegen die räumliche Umwelt aktiv ab.

Die Linguistik hat erst in den letzten Jahren begonnen, sich für das Zusammenspiel von Raum und Interaktion zu interessieren, parallel in der textlinguistischen Multimodalitätsforschung und in der linguistischen Gesprächsanalyse. Dabei baut die linguistische Gesprächsanalyse auf einer gut etablierten Forschung in der angelsächsischen Soziologie und Anthropologie auf (etwa Goodwin 1986 oder Heath 1986), in deren videobasierter Praxis die Relevanz des Raums schon seit den 1980er Jahren erkennbar geworden ist.

An dieser Tradition knüpft die linguistische Forschung zum "Interaktionsraum" an (s. etwa Mondada 2009) und präzisiert dabei Kendons Untersuchungen zu "F-Formations" und dem "o-space" in wichtigen Punkten. Zum einen differenziert sie den gemeinsamen Interaktionsraum in drei separate Räume, den "Wahrnehmungs-", den "Bewegungs-" und den "Handlungsraum" (Hausendorf 2010). Diese Räume entstehen durch jeweils eigene Koordinationsleistungen: die Koordination der Körperbewegungen im Raum, die der visuellen Wahrnehmung und das aufeinander abgestimmte soziale Handeln. Gleichzeitig arbeitet die linguistische Forschung heraus, wie die konkrete Gestalt des Interaktionsraums auf die Besonderheiten des gebauten Raums reagiert (Hausendorf / Kesselheim 2016) oder auf die spezifischen Notwendigkeiten der gemeinsam ausgeführten Aktivität (etwa die Nutzung bestimmter Objekte, vgl. Nevile et al. 2014). Unsere Methode der automatischen Erkennung von Interaktionsräumen setzt diese Präzisierungen um, indem sie zum einen separat Wahrnehmungs- und Bewegungsräume identifiziert, und zum anderen, indem sie die Rolle von für die Interaktion relevanten Objekten im Raum mitberücksichtigt. Beides ist, wie wir zeigen werden, für das Verständnis des Interaktionsgeschehens in unserem Museumssetting essenziell.

Datengrundlage unseres Vortrags ist ein Teilkorpus von 42 Stunden (1 Woche) Aufnahmen in einem Museum im Norden Deutschlands (Gesamtkorpus: 100 Wochen). Mithilfe von mehreren modifizierten Kinect-v2-Geräten wurden gleichzeitig von mehreren Standpunkten aus 2-D-Videoaufnahmen und 3-D-Tiefenbild-Sensordaten eines Teiles der Dauerausstellung aufgezeichnet. Um eine optimale Erfassung der Besucheraktivitäten zu ermöglichen, kalkulieren wir die Anordnung der Sensoren mit einem speziellen Verfahren, um Okklusion zu vermeiden. Die Nutzung von ToF-Kameras erlaubt es, mit einer zeitlichen Auflösung von 20 ms die 3-D-Koordinaten der Körperteile von Besuchern zu ermitteln, die während der Beobachtungszeit im Blickfeld der Sensoren erschienen sind.

Unsere Methode besteht in dem Tracking möglichst aller menschlicher Aktivitäten in einem mit Sensoren erfassten Innenraum und der nachfolgenden Analyse der entstehenden Tracking-Datensätze. Zunächst werden die Daten mehrerer Sensoren kombiniert. Aus den kombinierten und gefilterten Datensätzen werden die Trajektorien einzelner Besucher zusammengestellt, alle Stellen des Hovering-Verhaltens und Stopp-Positionen detektiert und diese mit den Positionen naheliegender Ausstellungsobjekte in Beziehung gesetzt. Auf der Grundlage der Positionen von Schultern, Becken und Köpfen der betreffenden Personen werden alle “transactional segments” (Kendon, s.o.) berechnet, um Überschneidungen zu finden und diese bestimmten Mustern zuzuordnen. Diese Muster werden dann weiter analysiert, um gemeinsame Interaktionsräume zu berechnen.

So lassen sich sowohl Gruppen von Besuchern definieren, die sich miteinander in Interaktion befinden, als auch körperlich-räumliche Bezugnahmen der Interagierenden auf Exponate im Raum. Dies erlaubt z.B. zu bestimmen, mit welchen Exponaten sich die Besucher auseinandersetzen und ob sie dies alleine oder gemeinsam tun. Aufgrund von einer Sammlung akkumulierter Interaktionsräume werden schließlich die Konturen von typischen Interaktionsräumen eines Exponats berechnet (typische Distanz zum Exponat, Blickwinkel, Dauer und Fragmentierung der Betrachtung). Daraus ergeben sich wichtige Hinweise für die Ausstellungsgestaltung und Wissensvermittlung als auch neue Möglichkeiten zur Auswertung der Nutzung von Ausstellungen durch die Besucher.

Ein Vorteil der Methode besteht schließlich auch in Bezug auf ethische und rechtliche Fragen, die sich im Zusammenhang mit videobasierten Besucherstudien stellen. Die Analyse der Besucherinteraktion auf Grundlage der 3-D-Skelette und schematischen Visualisierungen der Interaktionsräume, die aus den Sensordaten berechnet worden sind, ist vom Gesichtspunkt des Persönlichkeitsschutzes deutlich weniger heikel als die Arbeit mit Video-Daten, die nur mit hohem technischen Aufwand anonymisiert werden können.

Zur Gliederung unseres Vortrags.

Zunächst werden wir die von der Interaktionsraum-Forschung herausgearbeiteten Merkmale beschreiben, die für die räumlichen Formationen von Interaktionsbeteiligten charakteristisch sind. Dann erläutern wir, wie wir diese Einsichten genutzt haben, um in unserem Korpus Interaktionsräume automatisch zu identifizieren. Dabei arbeiten wir heraus, worin unsere Methode den im Rahmen der Computer Vision entwickelten Methoden zur Interaktionsraumerkennung überlegen ist, sowohl in technischer Hinsicht als auch im Hinblick auf die Differenziertheit der Abbildung qualitativer Forschungsergebnisse in der mathematischen Modellierung der Interaktionsräume.

Unser Beitrag zur Forschung.

Die automatisierte Detektion von Interaktionsräumen ermöglicht es, die Analysebehauptungen der bisherigen

Forschung zu Interaktionsräumen auf breiter Datenbasis zu überprüfen und zu konsolidieren. So zeigen unsere Analysen etwa, dass die Kendon'sche Beschreibung von F-Formations zu schematisch ist. Tatsächlich lassen sich in unseren Daten unterschiedliche Ausprägungen von Interaktionsräumen beobachten: beispielsweise solche, zu denen nicht alle Teilnehmer gleichberechtigten Zugang haben, oder solche, in denen Objekte die Position von 'Beteiligten' zugewiesen bekommen. Darüber hinaus kann die Liste der in einem Korpus detektierten Interaktionsräume von qualitativ Forschenden genutzt werden, um in ihrem Material schnell Fälle von 'unproblematischen' Interaktionsräumen zu identifizieren und ausgehend hiervon die Besonderheiten von weniger eindeutigen oder von den Erwartungen abweichenden Interaktionsräumen im Kontrast profilieren zu können.

Bibliographie

Ciolek, T. Matthew / Kendon, Adam (1980): *Environment and the Spatial Arrangement of Conversational Encounters*, in: Sociological Inquiry 50 (3-4), 237–271.

Ge, Weina / Collins, Robert T. / Ruback, R. Barry (2012): *Vision-based analysis of small groups in pedestrian crowds*, in: IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI). 34(5)1003–1016. PMID: 21844622 https://ieeexplore.ieee.org/document/5989835 [letzter Zugriff 02. Januar 2019].

Goffman, Erving (1963): *Behavior in public places. Notes on the social organization of gatherings*. New York, NY: Free Press.

Goffman, Erving (1971): *Relations in public. Microstudies of the public order*. New York: Basic Books.

Goodwin, Charles (1986): *Gesture as a Resource for the Organization of Mutual Orientation*, in: Semiotica 62 (1-2) 29–49.

Hausendorf, Heiko (2010): *Interaktion im Raum. Interaktionstheoretische Bemerkungen zu einem vernachlässigten Aspekt von Anwesenheit*, in: **Deppermann, Arnulf / Linke, Angelika (eds.):** *Sprache intermedial. Stimme und Schrift, Bild und Ton*. Jahrbuch 2009 des Instituts für deutsche Sprache. Berlin: de Gruyter (Jahrbuch des Instituts für deutsche Sprache, 2009) 163–197.

Hausendorf, Heiko / Kesselheim, Wolfgang (2016): *Die Lesbarkeit des Textes und die Benutzbarkeit der Architektur. Text- und interaktionslinguistische Überlegungen zur Raumanalyse*, in: **Hausendorf, Heiko / Schmitt, Reinhold, Kesselheim, Wolfgang (eds.):** *Interaktionsarchitektur, Sozialtopographie und Interaktionsraum*. Tübingen: Narr Francke Attempto (Studien zur deutschen Sprache 72) 55–85.

Heath, Christian (1986): *Body movement and speech in medical interaction*. Cambridge: Cambridge University Press.

Hung, Hayley / Kro#se, Ben (2011): *Detecting F-formations as Dominant Sets*, in: International Conference on Multimodal Interfaces (ICMI) 231–238. homepage.tudelft.nl/3e2t5/HungKrose_ICMI2011.pdf [letzter Zugriff 02. Januar 2019].

Mondada, Lorenza (2009): *Emergent focused interactions in public places. A systematic analysis of the multimodal achievement of a common interactional space*, in: Journal of Pragmatics (41) 1977–1997. DOI: 10.1016/j.pragma.2008.09.019.

Qin, Zhen / Shelton, Christian R. (2012): *Improving Multi-target Tracking via Social Grouping*, in: IEEE International Conference on Computer Vision and Pattern Recognition (CVPR) 1972–1978. https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=6247899 [letzter Zugriff 02. Januar 2019].

Nevile, Maurice / Haddington, Pentti / Heinemann, Trine / Rauniomaa, Mirka (eds.) (2014): *Interacting with objects. Language, materiality, and social activity*. Amsterdam: Benjamins.

Setti, Francesco / Hung, Hayley / Cristani, Marco (2013): *Group detection in still images by F-formation modeling: A comparative study*, in: International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS) 1–4. https://ieeexplore.ieee.org/document/6616147 [letzter Zugriff 02. Januar 2019].