

Multimodal data collection for social interaction analysis in-the-wild

Hung, Hayley; Raman, Chirag; Gedik, Ekin; Tan, Stephanie; Quiros, Jose Vargas

DOI

[10.1145/3343031.3351320](https://doi.org/10.1145/3343031.3351320)

Publication date

2019

Document Version

Final published version

Published in

MM 2019 - Proceedings of the 27th ACM International Conference on Multimedia

Citation (APA)

Hung, H., Raman, C., Gedik, E., Tan, S., & Quiros, J. V. (2019). Multimodal data collection for social interaction analysis in-the-wild. In *MM 2019 - Proceedings of the 27th ACM International Conference on Multimedia* (pp. 2714-2715). Association for Computing Machinery (ACM).
<https://doi.org/10.1145/3343031.3351320>

Important note

To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

Multimodal Data Collection for Social Interaction Analysis In-the-Wild

Hayley Hung
h.hung@tudelft.nl
Delft University of Technology

Chirag Raman
C.A.Raman@tudelft.nl
Delft University of Technology

Ekin Gedik
e.gedik@tudelft.nl
Delft University of Technology

Stephanie Tan
S.Tan-1@tudelft.nl
Delft University of Technology

Jose Vargas Quiros
J.D.VargasQuiros@tudelft.nl
Delft University of Technology



Figure 1: Example Snapshots of a mingling event. Taken from the MatchNMingle Dataset [2]

ABSTRACT

The benefits of exploiting multi-modality in the analysis of human-human social behaviour has been demonstrated widely in the community. An important aspect of this problem is the collection of data-sets that provide a rich and realistic representation of how people actually socialize with each other in real life. These subtle coordination patterns are influenced by individual beliefs, goals, and, desires related to what an individual stands to lose or gain in the activities they perform in their every day life. These conditions cannot be easily replicated in a lab setting and require a radical re-thinking of both how and what to collect. This tutorial provides a guide on how to create such multi-modal multi-sensor data sets when holistically considering the entire experimental design and data collection process.

CCS CONCEPTS

• **Hardware** → **Sensor applications and deployments**; *Wireless integrated network sensors*; • **Information systems** → *Social networks*; • **Human-centered computing**

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
MM '19, October 21–25, 2019, Nice, France
© 2019 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-6889-6/19/10.
<https://doi.org/10.1145/3343031.3351320>

→ *Collaborative and social computing devices*; • **Applied computing** → *Psychology*; • **Computing methodologies** → *Camera calibration*.

KEYWORDS

ConfLab, Social Behaviour Analysis, Wearable Sensors, Multimodal Synchronization

ACM Reference Format:

Hayley Hung, Chirag Raman, Ekin Gedik, Stephanie Tan, and Jose Vargas Quiros. 2019. Multimodal Data Collection for Social Interaction Analysis In-the-Wild. In *Proceedings of the 27th ACM International Conference on Multimedia (MM '19), October 21–25, 2019, Nice, France*. ACM, New York, NY, USA, 2 pages. <https://doi.org/10.1145/3343031.3351320>

1 MOTIVATION

Developing automated methods to analyze human social behavior in crowded face to face settings is an important multimedia concern [1, 2, 4]. With the rising importance of mobile and ubiquitous computing, multiple recording modalities are no longer tied to a fixed location in a lab setting. They are distributed in and amongst people as they move around in the world.

Harnessing the ubiquity of these sensing modalities is still an open question. Typically mobile computing applications rely on relatively low resolution data in order to accommodate a reasonable battery life for its users since mobile phones are also used as personal devices. Given the reduction in energy consumption and size of modern day electronics, sensing

devices are also being embedded into everyday objects such as smart ID badges. Such devices allow for a 'grab and go' paradigm to the ubiquitous sensing idea where consent to be recorded does not require a lengthy process of installing a mobile app or having to compete for processing power with other mobile applications. Such smart ID badges allow a tradeoff between battery life and higher sample rates in on-board sensors, such as those measuring body movement.

Exploiting accelerometer readings of body movement have shown promise for detecting social actions such as speaking activity or for detecting social involvement, but the solutions to such problems are still in their infancy [5, 6]. Multimodal attempts have also shown promise [1, 3] At this stage, video and audio data recorded in conjunction with wearable sensor data is vital for us to understand the relationship between automated social behaviour analysis as a multimodal and as a single modality problem.

Despite interest community interest in this topic, research works still remain relatively few. This could be strongly related to the difficulty of collecting relevant and sufficiently large data sets. Practical advice is needed to help researchers in the Multimedia community to design the data collection process appropriately. It calls upon an interdisciplinary approach to the data collection and experimentation process which is often not part of the standard educational curriculum of many multimedia researchers.

2 COURSE DESCRIPTION

This tutorial is tied closely to a social experiment carried out during ACM Multimedia 2019 called *ConfLab*. ConfLab aims to create a community level data gathering event. It involves recording conference attendees while they socialize as a means of helping attendees and organizers to network. To our knowledge, ConfLab is the first of its kind to turn the data collection process as a tool for community level introspection. The aim of ConfLab is to also make the data collected to be shared under appropriate levels of privacy restriction with the wider research community as a stimulus for grand challenge innovation.

The tutorial also acts as a debriefing and a moment for community reflection where participants or non-participants of ConfLab can share their thoughts on the initiative with others. We particularly encourage both newcomers and experienced members of the community to join. All perspectives of the conference from the attendees are vital for enabling a balanced and diverse learning experience.

This half day tutorial covers many of the practical considerations of collecting data which are often not documented in research papers, being not considered of academic interest for research on the automated human behaviour analysis in the wild. However, the issues highlighted in this tutorial are crucial to consider when collecting data in semi-public spaces. The tutorial is partially based on the book chapter by Hung *et al.*[7] providing more detailed practical advice at all levels of the collection process. In addition, themes more specific the data collection process of ConfLab will also be discussed.

The structure of the tutorial is divided into lectures and guided group discussion. The group discussions will primarily be focused on the debriefing of the ConfLab experiment and discussions about data sharing, how to incentivize participation, and how to include a participatory design approach into community data collection through thinking about ethical AI practices. Concretely, the following themes will be covered:

- Theory
 - Defining In-The-Wild vs Ecological Validity
 - Thin slice approach to Behavior Analysis
 - The inductive vs deductive approach when collecting data
- Practical issues (lectures)
 - Wearable sensor design
 - Multicamera video
 - Audio and privacy concerns
 - Multimodal data synchronization
 - Pilot Tests
- Ethical issues (lectures)
 - Applying for ethical approval
 - Informed consent
 - Data sharing (with particular focus on recent European Union laws on the General Data Protection Regulations (GDPR) introduced in 2018.
- Human concerns (group discussion)
 - Debriefing on ConfLab
 - How to incentivize participation: identifying stakeholders and their needs
 - Ethical data sharing practice

ACKNOWLEDGMENTS

This tutorial was partially funded by the Netherlands Organization for Scientific Research (NWO) under the MINGLE project number 639.022.606 and 015.012.018.

REFERENCES

- [1] Xavier Alameda-Pineda, Yan Yan, Elisa Ricci, Oswald Lanz, and Nicu Sebe. 2015. Analyzing free-standing conversational groups: a multimodal approach. In *Proceedings of the 23rd ACM international conference on Multimedia*. ACM, 5–14.
- [2] Laura Cabrera-Quiros, Andrew Demetriou, Ekin Gedik, Leander van der Meij, and Hayley Hung. 2018. The MatchNMI dataset: a novel multi-sensor resource for the analysis of social interactions and group dynamics in-the-wild during free-standing conversations and speed dates. *IEEE Transactions on Affective Computing* (2018).
- [3] L. Cabrera-Quiros, D. M. J. Tax, and H. Hung. 2019. Gestures in-the-wild: detecting conversational hand gestures in crowded scenes using a multimodal fusion of bags of video trajectories and body worn acceleration. *IEEE Transactions on Multimedia* (2019).
- [4] Tian Gan, Yongkang Wong, Daqing Zhang, and Mohan S Kankanhalli. 2013. Temporal encoded F-formation system for social interaction detection. In *Proceedings of ACM Multimedia*. ACM.
- [5] Ekin Gedik and Hayley Hung. 2017. Personalised models for speech detection from body movements using transductive parameter transfer. *Personal and Ubiquitous Computing* 21, 4 (2017), 723–737.
- [6] Hayley Hung, Gwenn Englebienne, and Laura Cabrera Quiros. 2014. Detecting conversing groups with a single worn accelerometer. In *Proceedings of the 16th international conference on multimodal interaction*. ACM, 84–91.
- [7] Hayley Hung, Ekin Gedik, and Laura Cabrera Quiros. 2019. Complex conversational scene analysis using wearable sensors. In *Multimodal Behavior Analysis in the Wild*. Elsevier, 225–245.