Detecting Social Interactions using Multi-Modal Mobile Sensing

Kleomenis Katevas†⋄, Katrin Hänsel⋄, Richard Clegg⋄, Ilias Leontiadis♮,
Hamed Haddadi†, Laurissa Tokarchuk⋄
† Imperial College London, ⋄ Queen Mary University of London, ♮ Telefónica Research

Over the years, there have been many attempts for detecting social interactions automatically. Most of the initial works are either based on manual annotated videos [1] or use computationally expensive computer vision techniques that rely on external CCTV camera surveillance. Other approaches with custom made wearable hardware [4] that use advanced sensors (*e.g.* infrared light) report increased accuracy, but at the same time are expensive and not easy to scale in larger environments. With the rapid rise in the variety of available smartphones and their wide range of embeded sensors, researchers have the opportunity to explore social interactions in an automated way that depends entirely on the use of mobile sensing technology [3], without the need of additional wearable equipment or external computer vision systems that consequently have implications with the user's privacy. While earlier works that depend on mobile sensing report accurate results, they have only been focused on and evaluated in detecting one-to-one social interactions, a situation that only covers a subset of the formations that occur in natural environments. Furthermore, they rely on pre-trained models that only work with specific mobile devices.

In this work, we present a real-time system based on gradient-boosted trees capable of detecting stationary social interactions inside crowds, depending on multi-modal mobile sensing data such as Bluetooth Smart (BLE), accelerometer and gyroscope. To overcome the restrictions applied from the mobile operating system (*i.e.* Apple iOS), wearable BLE beacons were used, configured to broadcast a beacon signal with a custom advertising packet per participant. This also allowed us to customize the broadcasting power of each beacon, achieving better accuracy in estimating the social space of each participant. In addition, the use of external BLE beacons installed in the ceiling of the room were used, providing localization based information to the classifier.

We evaluated our approach in a speed-networking study with 24 participants (9 male and 15 female). Each participant was equipped with a wearable beacon and an iPhone device enabled to collect sensor data using SensingKit[1] continuous sensing framework [2]. Two HD video cameras recorded the event from different angles. These videos were manually annotated by three independent researchers to provide an accurate ground truth. Participants were instructed to socially network with each-other for 45 minutes. In total, 99 one-to-one interactions were observed with a mean duration of $254.9s$ (± 161.7) and 22 group interactions (*i.e.* interactions that include more than two participants) with a mean duration of $117.2s$ (± 139.4).

A series of common features were computed for all $C(24, 2)$ combinations of the participant pairs. Features reflecting the current moment were initially computed, in a static window of 1 second, following with features reflecting past information. A set of features that are commonly included in mobile sensing problems were used, such as features extracted from motion and orientation sensors. Additional features were explored that provide more precise information for detecting the social interactions (*i.e.* interpersonal space, device position and indoor positioning). Our results report a performance of 86.4% Average Precision when detecting in real-time, the social interactions of each pair (*i.e.* every combination of the participants) per second. That is a performance increase of 26.7% over a proximity-based simplistic approach that was used as a baseline.

To continue this work, we will explore the use of orientation related features to improve the performance of the model. Moreover, we plan to extend this approach in detecting other types of interactions such as the pedestrian flocking behavior. We believe that this work can be particularly beneficial to the design and implementation of in-the-wild crowd behavioral analysis, design of influence strategies, and algorithms for crowd reconfiguration.

[1] https://www.sensingkit.org

REFERENCES

[1] H. Hung and B. Kröse. Detecting f-formations as dominant sets. In *Proceedings of the 13th International Conference on Multimodal Interfaces*, ICMI '11, pages 231–238, New York, NY, USA, 2011. ACM.

[2] K. Katevas, H. Haddadi, and L. Tokarchuk. SensingKit: Evaluating the sensor power consumption in ios devices. In *2016 12th International Conference on Intelligent Environments (IE)*, pages 222–225, Sept 2016.

[3] A. Matic, V. Osmani, A. Maxhuni, and O. Mayora. Multi-modal mobile sensing of social interactions. In *Pervasive Computing Technologies for Healthcare (PervasiveHealth), 2012 6th International Conference on*, pages 105–114, May 2012.

[4] A. Montanari, Z. Tian, E. Francu, B. Lucas, B. Jones, X. Zhou, and C. Mascolo. Measuring interaction proxemics with wearable light tags. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2(1):25, 2018.