Detecting Group Formations using iBeacon Technology

Kleomenis Katevas  
Queen Mary University of London, UK  
k.katevas@qmul.ac.uk

Laurissa Tokarchuk  
Queen Mary University of London, UK  
laurissa.tokarchuk@qmul.ac.uk

Hamed Haddadi  
Queen Mary University of London, UK  
hamed.haddadi@qmul.ac.uk

Richard G. Clegg  
Department of Computing  
Imperial College, London, UK  
richard@richardclegg.org

Abstract  
Researchers have examined crowd behavior in the past by employing a variety of methods including ethnographic studies, computer vision techniques and manual annotation based data analysis. However, because of the resources to collect, process and analyze data, it remains difficult to obtain large data sets for study. In an attempt to alleviate this problem, researchers have recently used mobile sensing, however this technique is currently only able to detect either stationary or moving crowds with questionable accuracy. In this work we present a system for detecting stationary interactions inside crowds using the Received Signal Strength Indicator of Bluetooth Smart (BLE) sensor, combined with the Motion Activity of each device. By utilizing Apple's iBeacon™ implementation of Bluetooth Smart, we are able to detect the proximity of users carrying a smartphone in their pocket. We then use an algorithm based on graph theory to predict interactions inside the crowd and verify our findings using video footage as ground truth. Our approach is particularly beneficial to the design and implementation of crowd behavior analytics, design of influence strategies, and algorithms for crowd reconfiguration.

Author Keywords  
Mobile Sensing; Crowd Sensing; BLE; iBeacon; RSSI; Group Formations; Social Interactions; Social Network Analysis
ACM Classification Keywords
Human-centered computing [Ubiquitous and mobile computing]: Empirical studies in ubiquitous and mobile computing

Introduction
Over the years, there have been many attempts to detect social interactions automatically. Most of the initial works are either based on manually annotated videos [8, 5] or use computationally expensive computer vision techniques [16, 3] that rely on external CCTV camera surveillance. With the rapid rise in the variety of available smartphones and their wide range of embedded sensors, researchers have the opportunity to explore social interactions in an automated way that depends entirely on the use of mobile sensing technology [12, 14]. While most early systems using mobile sensing report very accurate results, they are currently only able to detect one-to-one social interactions within crowds. Furthermore, they all rely on pre-trained models that only work with specific smartphone devices.

Unlike previous work in mobile crowd sensing, our approach is able to detect dynamic groups of variety of sizes and is not device dependent. The definition of a social interaction is inspired by the theory of conversation clusters or Kendon’s F-formations where “two or more people cooperate together to maintain a space between them to which they all have direct and exclusive access” [10]. In this preliminary work, we refer to social interaction as stationary groups of variety of sizes that are co-located inside the space. Our approach combines Bluetooth data for estimating the proximity between people and motion activity classification, estimating the stationary vs. moving status of each user by using the motion sensors of the device. Our findings report an accuracy of 89 percent while detecting interactions second-by-second using Bluetooth and Motion Activity sensor data.

Related Work
Hung & Krose proposed a way to identify F-formations based on the proximity and the body orientation of each person [8]. They reported an accuracy of 92 percent. In a similar research, Cristani et al. suggested a system that also detects F-formations with 89 percent accuracy, using information about people’s position and head orientation [5]. Both assume that the information of related proximity or position, as well as the body orientation is known, either using manual annotations or computer vision techniques.

One of the first attempts to identify face-to-face interactions in an automated way was the Sociometer [4], a wearable device that could be placed on each person’s shoulder and identify other people wearing the same device using infrared (IR) sensors. In addition, it is equipped with an accelerometer sensor to capture motion as well as a microphone to capture speech information. During the system evaluation, Sociometer was able to identify social interaction with an accuracy of 63.5 percent overall and 87.5 percent for conversations that lasted for more than one minute. Matic et al. [12] presented a solution based on using the Received Signal Strength Indicator (RSSI) of the Wi-Fi sensor as a way of measuring the distance between people and the embedded magnetometer to detect the body orientation of each participant. Finally, by placing an external accelerometer device into each user’s chest they analyzed the vibrations produced by the user’s vocal chords and detected speech activity. A common drawback with the Sociometer research is that it requires external hardware, making it unrealistic in real-world scenarios.

More recently, Palaghias et al. [14] presented a real-time system for recognizing social interactions in real-world sce-
narios. Using the RSSI of Bluetooth radios and a 2-layer machine learning model, they classified the proximity between two devices into three interaction zones, based on the Proxemics theory: a. Public, b. Social and c. Personal. In addition, they used an improved version of the uDirect research [6] that utilizes a combination of acceleromer and magnetometer sensors to estimate the user's facing direction with respect to the earth's coordinates. This work reported results of 81.40 percent accuracy for detecting social interactions, with no previous knowledge of the device's orientation inside the user's pocket. However, this work is only able to detect one-to-one social interactions using a specific device model (HTC One S) and not group interactions of more than two people.

Our approach utilizes the latest specification of Bluetooth, branded as Bluetooth Smart, that provides increased sampling rate, low power consumption compared to Bluetooth Classic radio, and reports more accurate results when used with different types of device models.

Proximity Detection

There have been several ways of estimating the distance between devices using wireless sensors such as Time of Arrival, Time Difference of Arrival, Angle of Arrival and using the Received Signal Strength Indicator (RSSI). At this moment, the only method that is applicable in today's smartphones is using the RSSI of either the Bluetooth or the Wi-Fi sensor.

In the past, researchers have used the RSSI of Bluetooth [11, 7, 14], Wi-Fi [13] or even a combination of them [2] by measuring the RSSI of every wireless sensor available in range and comparing it with a Measured Power constant (also known as txPower) that indicates the signal strength (in dBm) at a known distance (usually 1m). In 2010, the Bluetooth Special Interest Group (SIG) released an updated version of the Bluetooth standard (v4.0) with a Low Energy feature (BLE) that was branded as Bluetooth Smart. Bluetooth Smart is low cost for consumers, has low latency in communications (6 ms) and is power efficient. Moreover, it supports an advertising mode were the device periodically broadcasts specially formatted advertising packets to all devices in range with a sample rate of approximately 3Hz. This packet can contain a unique ID for each device, as well as the measured power constant that was mentioned above. The advantage of using this technology for proximity estimation is that each device can broadcast its own measured power constant, making the proximity estimation more accurate. In addition, devices do not need to be connected in order to measure the RSSI, having a minimum impact on the device's battery life.

Apple developed a closed-source protocol based on Bluetooth Smart, branded as iBeacon™ and supported it as of iOS 7 in June 2013. As of publication, that corresponds to more than 95 percent of all iOS devices available¹. Android recently presented their own open-source protocol based on Bluetooth Smart, titled Eddystone™. Even though scanning for other Eddystone™ beacons is supported in devices with Android Jelly Bean (v4.3) or greater, broadcasting is only fully supported since Android Lollipop (v5.1) and only in the most recent devices (e.g. Nexus 6, Android One).

In this paper, we evaluate the iBeacon™ protocol as a way of proximity estimation between iOS devices. We first conducted a short experiment using an iPhone 5S and an iPhone 6S. Using SensingKit iBeacon™ Proximity sensor in Scan & Broadcast configuration we collected proximity data (RSSI and Accuracy²) for five minutes in 12 distances.

¹As reported by Apple App Store on March 7, 2016
²Apple’s proximity estimation.
from 0.00m to 3.00m, every 0.25m. It is important to mention that each device was both broadcasting and receiving an iBeacon™ signal, thus we ended up with 300 measurements for each device per distance. The experiment took place in the Performance Lab of Queen Mary University of London, an empty room approximately 9x8 meters long that is mainly used for a variety of performance research and recording applications. Figure 1 shows the RSSI patterns over distance, as measured by the two iPhone devices.

![Figure 1: Box plot of Bluetooth Smart RSSI versus distance using two iPhone devices.](image)

The figure demonstrates that the pattern of the RSSI signal is device hardware dependent (iPhone 5S vs. iPhone 6S) and proves the requirement of a measured power constant as a reference point for a pre-known distance as explained above. It also shows the instability and fluctuation of the RSSI signal, mainly because of environmental factors (e.g. reflections, obstacles, background noise) as reported by other similar works [13, 11]. We repeated the same experiment at the Mobile Antenna EMC Screened Anechoic Chamber of Queen Mary University of London, a 6x5x3 meters chamber designed to completely absorb reflections of electromagnetic waves. We used two metallic tripods to hold the two devices. Because of their reflecting material they both contributed to the multi-path components.

Although the pattern of the received RSSI in both environments is similar, the actual captured value in the chamber is lower by approximately 5-7 dBm which highlights the reduction in reflections and multi-path components. In addition, the similarity in the patterns could be contributed to the receiver's own automatic gain control.

We evaluated different methods of estimating distance using the Bluetooth Smart RSSI and the measured power as recorded by the iBeacon™ Proximity sensor (-57 for iPhone 5S and -56 for iPhone 6S). We tested the Path Loss Model (PLM) [13], based on the theory of signal loss as it propagates through space:

$$ P(d) = P(d_0) - 10 \times n \times \log_{10}(d) - X $$

where $P(d_0)$ is the measured power (in dBm) at 1 meter distance, $n$ the path loss exponent, $d$ the distance in which the RSSI is estimated and $X$ a component that describes the path loss by possible obstacles between the transmitter and the receiver. We chose $n = 1.5$ and $X = 0$ as the environment was indoors and no obstacles between the two devices existed [12]. The same model can easily be transformed in order to estimate the distance ($d$) from the RSSI ($P(d)$), as shown in the following formula:

$$ d = 10 \frac{P(d_0) - P(d) - X}{10 \times n} $$

We also tested a model given by Radius Networks [15], a
company that produces beacon technology solutions compatible with both iBeacon™ and Eddystone™:

\[ r = \frac{P(d)}{P(d_0)} \]  

\[ d = \begin{cases} r^{10}, & \text{if } r < 1.0 \\ 0.89976 \times r^{7.7095} + 0.111, & \text{otherwise} \end{cases} \]  

According to Radius Networks, “The three constants in the formula (0.89976, 7.7095 and 0.111) are a best fit curve based on a number of measured signal strengths at various known distances from a Nexus 4”.

Finally, we compared the two methods with iBeacon’s distance estimation (reported as Accuracy by Apple). As iBeacon™ technology is closed-source, the details of this method are unknown. We observed a delay of 20 seconds each time we change the distance, indicating that Apple’s method uses a buffer for estimating the distance, making it less responsive and not suitable for the aim of this work.

Figure 2 shows the accuracy of the distance estimated by the three methods described above. Since two RSSI measurements exist, one reported by each device, we applied each model twice and used the average of the two distances. The results indicate a low accuracy in all three models, especially at the distance of 3 meters that the overall error is more than 1.5 meters. Apple and PLM models reported the worst results, with a Root Mean Squared Error (RMSE) of 0.78 and 0.95 accordingly. The Radius model showed slightly better results with a RMSE of 0.68 overall.

**Motion Activity Detection**

Modern mobile devices are equipped with motion co-processors that analyze data from the embedded accelerometer, gyroscope and magnetometer sensors and classify the user’s motion activity into Stationary, Walking, Running, Driving and Cycling. We tested all motion co-processors available in iPhone devices today (M7 to M9 models). When an activity changed from stationary to walking, we observed a delay of 3-5 seconds until this was reported. As responsive motion activity recognition is a key feature for identifying stationary interactions, we analyzed the motion sensor data from the device.

We chose the linear acceleration sensor that measures the coordinate acceleration of the device by removing the gravity from the raw accelerometer data. We averaged the data on every 1 second and empirically classified a device as Stationary if the magnitude of the acceleration is less than 0.15 g and as In Motion otherwise.
Detecting Group Formations
This study aims to automatically detect group formations taking place in social events with the use of Mobile Sensing technology. In contrast to all other related work discussed in Sec. Related Work, our approach is capable of detecting group formations with more than two people involved.

This section starts with a detailed description of the experiment, including the participants and the materials used, a description of the space that the experiment took place and the procedure followed.

Participants
The experiment involved the recruitment of six participants, all research students from Queen Mary University of London that had no previous experience with the current study. Table 1 shows the demographic information of the recruited participants as well as the iPhone device and iOS version they used in the experiment.

<table>
<thead>
<tr>
<th>Gender</th>
<th>Height</th>
<th>Weight</th>
<th>Device</th>
<th>iOS</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1 Male</td>
<td>1.80m</td>
<td>74kg</td>
<td>iPhone 6S</td>
<td>9.1</td>
</tr>
<tr>
<td>P2 Male</td>
<td>1.88m</td>
<td>63kg</td>
<td>iPhone 6</td>
<td>8.3</td>
</tr>
<tr>
<td>P3 Male</td>
<td>1.85m</td>
<td>72kg</td>
<td>iPhone 6</td>
<td>9.1</td>
</tr>
<tr>
<td>P4 Female</td>
<td>1.72m</td>
<td>97kg</td>
<td>iPhone 6</td>
<td>9.1</td>
</tr>
<tr>
<td>P5 Female</td>
<td>1.64m</td>
<td>59kg</td>
<td>iPhone 5S</td>
<td>9.1</td>
</tr>
<tr>
<td>P6 Male</td>
<td>1.90m</td>
<td>95kg</td>
<td>iPhone 5S</td>
<td>9.0.2</td>
</tr>
</tbody>
</table>

Materials
As a sensor data collection tool, we used the open-source CrowdSense App. CrowdSense is an iOS application based on the continuous sensing library Sensing Kit [9], an open-source library for iOS and Android platforms. It provides access to all sensors supported by SensingKit, when these are available in the device. The application collects sensor data and saves them into the device's memory in CSV format. It can be run in all iOS devices with iOS 8 or greater and is available for free in Apple’s App Store.

As shown in Table 1, the experiment used one iPhone 6S, three iPhone 6 and two iPhone 5S devices, equipped with CrowdSense App and configured to collect Accelerometer as well as iBeacon™ Proximity data in Scan & Broadcast mode. We decided to avoid iBeacon™ re-calibration and chose the default measured power value in iBeacon™ Broadcast, as defined by the manufacturer. The reason is that we wanted to make sure that this experiment can easily be replicated in large scale with unknown type of devices. We set the iBeacon™ identifier (Minor value in iBeacon™ protocol) with a unique ID per device. A short audio sample of 10 seconds was also recorded and used to synchronize the sensor data between the six devices.

Two full-HD video cameras were used to record the experiment from two different angles. This video recording was used as the ground truth for evaluating the accuracy of our detection of the interactions happening inside the room.

Procedure
The experiment took place at the Performance Lab of Queen Mary University of London. The Performance Lab is a large space (approximately 9x8 meters), suitable for live performance experiments. A DMX lighting rig is also installed in the ceiling of the space, useful for installing Lights, Cameras or other related equipment. We installed the two HD video cameras in the lighting rig.

All six participants were informed that the experiment explores the ways people interact in social events and received a short demonstration of CrowdSense application. They signed the required consent forms and completed a short questionnaire with demographic questions (Gen-
Table 2: Scenes with group formations tested during the experiment.

<table>
<thead>
<tr>
<th>Name</th>
<th>Groups</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>3</td>
<td>A group of three people walk in and form a stationary F-formation.</td>
</tr>
<tr>
<td>S2</td>
<td>3, 3</td>
<td>Another group of three people walk in and form another stationary F-formation.</td>
</tr>
<tr>
<td>S3</td>
<td>2, 4</td>
<td>One participant leaves group A and joins group B.</td>
</tr>
<tr>
<td>S4</td>
<td>3, 3</td>
<td>Another participant leaves group B and joins group A.</td>
</tr>
<tr>
<td>S5</td>
<td>5, 1</td>
<td>Two participants leave group A and joins group B. The remaining person walks on the opposite direction.</td>
</tr>
<tr>
<td>S6</td>
<td>6</td>
<td>All participants join group B.</td>
</tr>
<tr>
<td>S7</td>
<td>2, 4</td>
<td>Two people from group B leave and form a separate formation (group A).</td>
</tr>
<tr>
<td>S8</td>
<td>4, 1, 1</td>
<td>Group A is now split and one of the participant walks close to group B twice, but do not join their formation.</td>
</tr>
<tr>
<td>S9</td>
<td>5, 1</td>
<td>The same participant now joins group B.</td>
</tr>
<tr>
<td>S10</td>
<td></td>
<td>Group split and participants leave the room.</td>
</tr>
</tbody>
</table>

All data collection and analysis was made with informed consent and approved by the Queen Mary University of London research ethics committee (Reference: QMREC1543a).

Results and Discussion
Before starting the data analysis, we synchronized the sensor data (iBeacon™ Proximity and Linear Acceleration) using the audio sample described previously, having an accuracy of ±50ms. We also filtered the proximity data by removing values reported as Unknown with -1 as RSSI. This usually occurs at the beginning of the iBeacon™ ranging process due to insufficient measurements to determine the state of the other device [1]. We also applied Radius proximity model for estimating the proximity between devices and classified the motion activity using the Linear Acceleration sensor data.

Inspired by the work of [8], our concept for detecting Social Interactions is based on Graph Theory. Each moment (in seconds) is represented as an undirected weighed graph $G = (V, E, w)$, with a set of vertices $V$ and weighed edges $E(w)$. Each vertex corresponds to a participant, and each weighted vertex correspond to the distance as estimated by the iBeacon™ Proximity sensors, as discussed in Sec. Proximity Detection. We use an array of weighed adjacency matrices to represent the estimated distance of each participant for each second (Figure 3a). Since this work only focuses on stationary interactions, when a participant is classified as non-stationary (based on the method discussed in Sec. Motion Activity Detection), he is equally classified as non-interacting. We convert the weighed adjacency matrices into binary adjacency matrices (Figure 3b) based on Algorithm 1. The algorithm is based on two variables: A range (from - to) of estimated distance that two participants should have, and $x$ that represents the time (in
seconds) that this distance should be maintained. When these conditions are met, an edge between the two nodes is created and represented by 1 in the matrix. An interaction is identified when a connected component exists in the graph, as shown in Figure 3c (Participants 3, 4, 5 and 6).

**Algorithm 1 Vertice creation of binary adjacency matrix**

if activity is not stationary then
  w ← 0
else
  if from ≤ distance ≤ to and time ≥ x then
    w ← 1
  else
    w ← 0
end if
end if

We evaluated the performance of the algorithm by using the video recording as ground truth. We asked an external researcher to annotate the group formations using ELAN multimedia annotator software [17]. The exact instructions, based on Kendon’s F-formation [10] explained before, were: “An interaction begins at the moment two or more people are stationary and cooperate together to maintain a space between them to which they all have direct and exclusive access.”.

We evaluated the algorithm with different values for time and distance variables and discovered that it performs best with an estimated distance between 0 and 5 meters for at least 3 seconds (Figure 4). The maximum accuracy that we achieved is 89 percent with precision 0.81 percent. Table 3 summarizes the results of our experiment using a confusion matrix.

<table>
<thead>
<tr>
<th></th>
<th>Positive</th>
<th>Negative</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>1052</td>
<td>248</td>
<td>1139</td>
</tr>
<tr>
<td>Negative</td>
<td>87</td>
<td>1598</td>
<td>1846</td>
</tr>
<tr>
<td>Total</td>
<td>1155</td>
<td>1830</td>
<td>2985</td>
</tr>
</tbody>
</table>

Precision = 0.809, Accuracy = 0.888

Even though the method depends on using anonymous IDs when broadcasting iBeacon™ data and no other personal information is broadcast, there is always the danger that this anonymity can be compromised by tracking the openly available ID of a user. This is the reason that in iOS devices, a device is only allowed to broadcast as an iBeacon™ while the app is actively running in the foreground. A solution to this can be the use of the latest Eddystone-EID frame type, released in April 2016 by Google, that encrypts the broadcast data so that only authorized people or applications can access it.

**Conclusions and Future Work**

In this work, we introduced a system that automatically detects formations of groups with a variety of sizes inside crowds. The detection depends entirely on the embedded sensors of smartphone devices. By utilizing the Bluetooth Smart and motion sensors of each device, we were able to estimate the distance between people as well as to classify the user’s activity between stationary and in-motion. In addition, by using an algorithm that is based on graph theory models, we were able to predict group formations inside a small group of seven people, reaching an accuracy of 89 percent.

To continue this work we will explore the use of the gyroscope sensor as a way of estimating the orientation of each
user related to the environment and the effect it will have when including this feature in our group formation detection method. We also plan to improve the proximity estimation using trained regression models. One of the limitation of our approach is that we do not make use of the orientation of each user, hence identifying an interaction based on proximity and motion activity. This approach will have an impact when applying the algorithm in crowded places were people are standing close to each-other. The less accurate distance estimation based on RSSI is another challenge of this research. As previously mentioned, the only option available in today’s smartphone devices is based on the RSSI. If other technologies ever become available such as the ones mentioned in Sec. Related Work, the accuracy of this method will greatly benefit. In future works with larger crowds we are planning to improve this aspect.

We plan to repeat the experiment in a real-world social event, evaluate its accuracy and finally present analytics.
about the ways in which people are interacting. We believe that this work will lead to more complex techniques capable of detecting actual social interactions in variety of crowd sizes.

Acknowledgments
The authors wish to thank the participants in the trial. They also wish to thank Alessio Xompero and Riccardo Mazzon from Queen Mary University of London for sharing their computer vision expertise as well as Akram Alomainy for his advise during the proximity detection experiment. This work is supported by funding from the UK Defense Science and Technology Laboratory.

REFERENCES


