A Study on Indoor Pedestrian Flock Detection Using Mobile Devices

（携帯デバイスを用いた屋内における歩行者群検知の研究）

A DISSERTATION

SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL OF

THE UNIVERSITY OF TOKYO

BY

Jose Pablo Alvarez Lacasia

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

MASTER OF INFORMATION AND COMMUNICATION ENGINEERING

Kaoru Sezaki

指導教員　瀬崎　薫　教授

February 2014

Acknowledgements

I would like to thank professor Sezaki for his kind support and guidance for both academic and everyday needs. Also, Iwai sensei, Kobayashi sensei and Ito sensei for their invaluable advice through the whole process of picking my research topic and setting the goals and making them happen. I thank the members of the Sezaki lab for helping with the experiments and helping me bringing me closer to Japanese culture.

I would also like to extend my gratitude to Teemu for being an outstanding and dedicated guide for my research and publications. Without his unparalleled patience, I would not have been able to improve my writing.

It was also due to the extraordinary support of my family and friends, who helped me through the whole process of living in Japan, I can finally bringing this period of my life to a happy conclusion. I will be forever grateful.

Contents

1. Introduction 2
   1.1. Conceptual background 2
   1.2. Previous works in flock detection methods 5
      1.2.1. Outdoor pedestrian flock detection based on GPS data 5
      1.2.2. Indoor pedestrian flock detection based on Wi-Fi signal strength 6
      1.2.3. Indoor pedestrian flock. Multimodal sensor and Wi-Fi approach 7
   1.3. Objectives 8
   1.4. Contribution and Thesis structure 9
2. Problem statement 11
3. Data collection and Ground Truth 13
   3.1. Data collection applications 14
   3.2. Ground truth experiments for Wi-Fi 15
   3.3. Ground truth experiments for sensors 23
4. Features extracted from the ground truth data 28
   4.1. Wi-Fi signal strength similarity 30
   4.2. Sensor Modalities features 31
5. Results: Similarity of the features 34
   5.1. Similarity values for Wi-Fi 34
      5.1.1. Contrasting Wi-Fi with Barometer 41
   5.2. Similarity results for the sensor modalities 42
      5.2.1. Accelerometer 42
      5.2.2. Barometer 45
      5.2.3. Gyroscope 46

5.2.4. Magnetometer 47
5.3. Discussion 49
6. Discussion 51
7. Conclusion 54
7.1. Summary 54
7.2. Suggested Future Work 59

Figure Index

Fig 1. Example of flocks. Spatial and temporal clustering [14]. 3

Fig 2. Concept pedestrian flock detection application architecture [15]. 4

Fig 3. Application architecture for the flock detection method proposed in [9]. 8

Fig 4. Ten Samsung Galaxy sIII smartphones used in the experiments. 13

Fig 5. Graphic user interface of the sensor, Wi-Fi and audio logging Android application 14

Fig 6. Schematic map of the Komaba II campus building. 16

Fig 7. Schematic of the crossing flocks scenario. 18

Fig 8. Schematic of the scenario with flocks moving in the same direction with delay. 18

Fig 9. Schematic of crossing flocks with perpendicular trajectories. 19

Fig 10. Schematic of crossing flocks with an idle person standing in the crossing point. 19

Fig 11. Schematic of perpendicularly crossing flocks with an idle person standing at the crossing point. 20

Fig 12. Schematic of the scenario with flocks moving in the same direction with delay in different floors 21

Fig 13. Schematic of the scenario with crossing flocks in different floors. 22

Fig 14. Schematic of crossing flocks with perpendicular trajectories in different floors. 22

Fig 15. Schematic for experiment 1 and 2 for accelerometer data. 24

Fig 16. Schematic of same turn and different turn scenario for experiment 3 and 4 for sensor data ground collection. Gyroscope and magnetometer comparison. 25

Fig 17. Schematic of experiment 5 for ground truth collection of sensor data. 2 phones carried by an individual and 1 static phone. 26

Fig 18. Schematic of the Wi-Fi similarity calculation process. 31

Fig 19. Accuracy of classifying phones into groups using the average as threshold. 40

Fig 20. Accuracy of classifying phones into groups using the median as threshold........................ 40

Fig 21. Barometer data for scenario 6. 42

Fig 22. Tri-axial magnetometer data for experiment 4 for gathering sensor data. 49

Fig 23. Schematic visualization of the flock detection algorithm using weighted majority voting for three time steps and temporal clustering which identifies two flocks F1 and F2.[13] 52

Fig 24. Schematics of all the experiments for the sensor ground truth data collection experiments. (a) Experiment 1, (b) experiment 2, (c) experiment 3, (d) experiment 4 and (e) experiment 5. 56

Table Index

Table 1. Feature space used in [14] ... 7

Table 2. Wi-Fi Experiment Scenarios: summary. ... 17

Table 3. Summary of ground truth collection experiments for sensor data. ... 23

Table 4. Sample: Ideal similarity value. Average is 0.55 and median is 0.5 ... 34

Table 5.Average threshold scenario 1. Average is 0.63 and standard deviation is 0.26 ... 35

Table 6. Average threshold scenario 2. Average is 0.77 and standard deviation is 0.16 ... 35

Table 7. Average threshold scenario 3. Average is 0.67 and standard deviation is 0.21 ... 35

Table 8.Average threshold scenario 4. Average is 0.66 and standard deviation is 0.22 ... 36

Table 9. Average threshold scenario 5. Average is 0.60 and standard deviation is 0.20 ... 36

Table 10. Average threshold scenario 6. Average is 0.43 and standard deviation is 0.42 ... 36

Table 11. Average threshold scenario 7. Average is 0.41 and standard deviation is 0.40 ... 36

Table 12. Average threshold scenario 8. Average is 0.42 and standard deviation is 0.40 ... 37

Table 13. Median threshold scenario 1. Median is 0.71 ... 37

Table 14. Median threshold scenario 2. Median is 0.77 ... 37

Table 15. Median threshold scenario 3. Median is 0.61 ... 38

Table 16. Median threshold scenario 4. Median is 0.62 ... 38

Table 17. Median threshold scenario 5. Median is 0.63 ... 38

Table 18. Median threshold scenario 6. Median is 0.51 ... 39

Table 19. Median threshold scenario 7. Median is 0.14 ... 39

Table 20. Median threshold scenario 8. Median is 0.54 ... 39

Table 21. Results for accelerometer experiment 1&2. ... 43

Table 22. Results for comparison of the first five coefficients of the fast Fourier transform of the accelerometer data. Experiment 5. 44

Table 23. Results for barometer data in experiment 5 45

Table 24. Results for gyroscope experiments 3 and 4. 46

Table 25. Results for Gyroscope experiment 5. Cross correlation per axis and magnitude 47

Table 26. Results for magnetometer experiment 3 and 4 47

Table 27. Results for the magnetometer experiment 5. Cross correlation per axis and magnitude .. 48

Table 28. Summary of the scenarios for the Wi-Fi ground truth data collection experiment. 55

Table 29. Summary of the sensor ground truth data collection experiments. 56

Table 30. Feature and methods applied to the Wi-Fi and sensor data gathered in the experiments. 58

Abstract

This thesis presents the study of a set of features that are derived from multi-modal sensor data and Wi-Fi signal strength data applied to indoors pedestrian flock detection. Pedestrian flocks are groups of collocated persons walking cohesively for a certain amount of time. Wi-Fi signal strength vectors are commonly used for indoor localization applications, and recently expanded to the application area of detecting pedestrian flocks by comparing the similarities of vectors directly between smartphones. Additionally, this thesis seeks to improve the accuracy of existing methods by using additional modalities of the integrated sensors in smartphones, increasing the feature space needed to determine the similarity in the data vectors to identify the flocks. The utilization of several sensor modalities was studied for this application: accelerometer, magnetometer, gyroscope and barometer. Results show that it is possible to distinguish groups solely from Wi-Fi data under certain circumstances and the data from other modalities can be used to either, complement the results from Wi-Fi or obtain the same results. The results obtained from magnetometer and gyroscope display the highest performance between all the sensor modalities in terms of similarity gap, even for general situations. Therefore, in this thesis, it is argued the possibility of implementing a crowd-sensing application with further sensor modalities for indoor pedestrian flock detection.

1. Introduction

1.1.Conceptual background

Current smartphones are equipped with an increasing amount of sensors, which can be used by researchers to sense and describe the environment surrounding smartphones, therefore indirectly the users [1], and even infer a semantically richer description of a given location [2]. In contrast to these type of applications, this thesis focuses on trying to describe the environment of several users at a time and then, grouping them according to the similarity of their perceived data. This type of application can be categorized as a crowd sensing application [3] as multiple users supply sensing data.

This thesis specifically focus on the detection of pedestrian flocks in indoor environments and argue the possibility of developing a distributed application to solve the problem [15]. A pedestrian flock is a group of people moving cohesively for a given interval of time [4], an example is shown in Figure 1. Developing an application to detect pedestrian flocks can allow researchers to observe how these groups form and dissolve. Applications in a diversity of areas can benefit from the detection of pedestrian clusters. In the area of constructions, the benefit of detecting groups of pedestrians, is the possibility of identifying bottlenecks, or tight passages, where people, in the eventual case of an evacuation, could find difficulties to leave the building as readily and quickly as it is necessary [5]. Current indoor localization applications can be used in conjunction with flock detection to smoothen evacuation processes. People using indoor localization applications in their smartphones seem willing to receive guidance in the case of events, such as emergencies, through their smartphones according to the survey in [6]. Furthermore, detecting how inclined people are to join pedestrian flocks (or avoid joining them) can be used to improve existing social psychology tools [7].

Augmented reality applications can also be devised; they could insert images and sounds according to what a user is currently experiencing, which can be determined by context [22]. The augmented reality applications include navigation, personal information systems, personal awareness systems, activity recognition and moreover enabling collaboration in applications [22]. In these applications, the enabling technologies include handheld displays, tracking sensors and computational power. These are the characteristic features of the smartphones today, which as common appliances are carried along by humans during their everyday lives. Smartphones are computational platforms usable in real-time, interactively and spatially in both indoors and outdoors. As networked devices, smartphones enable collaborative application in ubiquitous computing environments by seamlessly connecting to the services in the local environment.

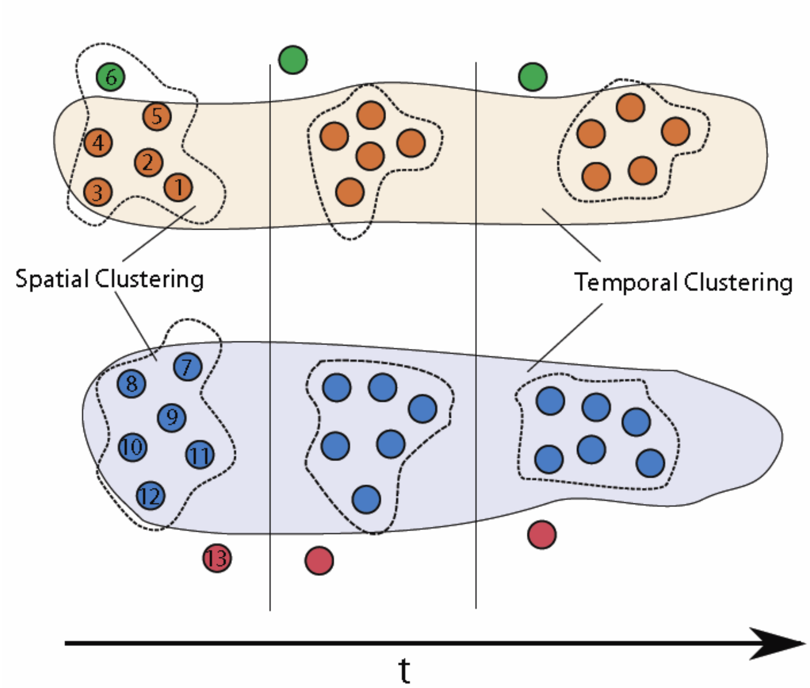

Fig 1. Example of flocks. Spatial and temporal clustering [14].

In earlier work, a conceptual distributed system architecture was described, enabling crowd sensing-based pedestrian flock detection application [15]. The system architecture is shown in Fig. 2. This work was jointly done with Teemu Leppänen, a guest researcher in the Sezaki laboratory during the time this research was conducted. The implementation of the distributed application for pedestrian flock detection will be done by Teemu Leppänen, whereas the features used to derive the flocks by this application are studied in this thesis.

As it can be seen in Figure 2, there are 3 stages that can be identified in the process for detecting flocks that we propose. The application will have to iterate several times, before it can find the optimal set of sensors for detecting flocks, in terms of detection accuracy per sensor and energy consumption. In stage 1 of Figure 2, the phones consult what a "score list", which ranks the sensors according to their flock detection accuracy and energy consumption profile.

Stage 2 is when the phones start logging sensor data and calculate the features which will be used for determining the flocks, using the most fitting sensor modalities indicated by the "score list". The score list has to take into consideration the power consumption profile of each sensor as well as the accuracy. The calculations of the features are done in the phones and the clustering process of phones is done in a distributed manner, sending all the calculated features to a single phone which will calculate the clusters. The flocks are determined depending on how many sensor modalities concur on the same amount of clusters (majority voting as in [9]). In stage 3, the phones perform the clustering calculations and make them available as a web service on the cloud. Following to this step, the determined flocks are evaluated in terms of accuracy and then the score list is updated in case the accuracy is determined to be low (by having several modalities disagreeing on the detected flocks or drops in sensors' accuracy, for example GPS), giving way for the next iteration of the application.

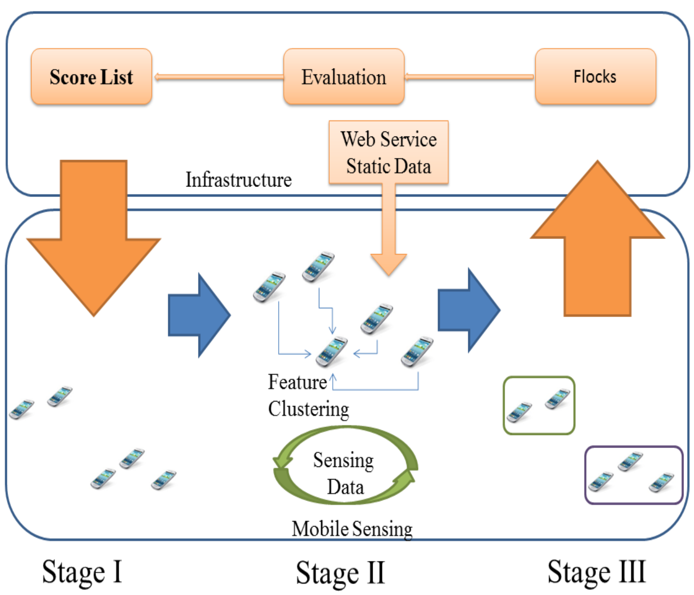

Fig 2. Concept pedestrian flock detection application architecture [15].

Concretely speaking, 4 steps have been defined to calculate the similarity between 2 vectors of Wi-Fi signal strength from different phones in our previous work in [10]: (1) The raw Wi-Fi measurements are handled in the time domain and divided into equally sized groups with incrementing timestamps, (2) the measurements are filtered using an empirically determined threshold [8], as some of the measurements can be seen as noise, (3) next, the average power for each MAC address is calculated over each group of measurements (yielding averages for a time vicinity) and (4) the resultant vector for each phone is compared using the cosine similarity. Finally, all the similarity values for each group are averaged yielding a similarity value for each pair of mobile phones. The work in [10] considered only Wi-Fi data; the results shown in this thesis further this work by adding the analysis of sensor data, such as barometer, accelerometer, magnetometer and gyroscope. For which, the moving average and standard deviation of the data were calculated for each modality.

1.2. Previous works in flock detection methods

The latest work done in pedestrian flock detection, to the best of the knowledge obtained while this thesis was written, is an implementation that uses fusion of multi-modal sensors [9] (Wi-Fi, accelerometer and magnetometer). However, before reaching this solution, there were some prior attempts to solve the pedestrian flock detection problem in a variety of scenarios, which are worth noting.

1.2.1. Outdoor pedestrian flock detection based on GPS data

In [13], the authors proposed a method to cluster pedestrian in outdoors scenarios based on their GPS data. For this purpose, a 2 stage clustering was devised. The method would cluster samples according to their longitude and latitude. The resulting clusters are then once again clustered according to their time stamp. These spatio-temporally clustered samples are then set as candidate flocks. If flocks maintain their configuration for a time longer than a pre-set threshold, the candidate clusters are considered flocks.

As it was mentioned in [13], GPS samples can be very noisy and even sometimes the visibility from the perspective of the smartphone can be lost. If the samples' accuracy drops below a given threshold, the flock detection application will use a heuristic developed in [13], instead of using interpolation to fill in the sample blanks. The heuristic consists on maintaining the flock configuration for a pre-determined duration, regardless of the actual GPS samples (whose accuracy was determined to be low). This method detects flocks with 91.3% of accuracy according to the authors.

1.2.2. Indoor pedestrian flock detection based on Wi-Fi signal strength

The same authors developed a solution for indoor environments in [14]. As is well known, GPS is not suitable for use in indoor scenarios for positioning devices and consequentially indoor pedestrian flock detection. Therefore, it was necessary to utilize Wi-Fi instead, as it is very commonly used in indoor localization applications [8].

In order to determine how close a pair of devices is, several features had to be calculated. To do so, a server had to be populated with Wi-Fi fingerprints of the premises where the application was meant to be deployed, mapping the fingerprints to a physical location. Fingerprints were calculated by averaging the power level per visible access point from the point of view of the device. Then, to localize a device, the fingerprint calculated at some point of time is compared with the ones stored at the server, yielding the location for the closest match and a feature for clustering. Additionally, the power level of the fingerprints was directly compared. One of the novelties of the method proposed in [14] was to map measurements to a space with a defined distance unit.

The aforementioned features used in this approach are summarized in Table 1. There is a brief explanation in the second column of each type of feature.

As it can be seen from the table 1, a variety of features was derived from exclusively Wi-Fi. Contrasting with the proposed method in this thesis, Wi-Fi fingerprints are directly compared from one another, as opposed to the 3 features derived in [14] and shown in Table 1. This thesis demonstrates that for certain scenarios, only a single feature from Wi-Fi is necessary to differentiate groups of phones. Although, having more features derived from the Wi-Fi data can certainly make the method more robust, the advantage of comparing fingerprints directly, is that there is no need to store the fingerprints in a server prior to the application deployment. The advantages of the method proposed in this thesis will be discussed with more detail in later chapters.

Table 1. Feature space used in [14]

Spatial features	Features calculated from a pre-established coordinate system
Signal-strength features	Features calculated from using directly the data acquired of the signal strengths of Wi-Fi access points
Pseudo-Spatial features	Features calculated which yield an appropriately defined distance

Finally, once the features have been derived, a two stage clustering is performed over the feature data, similarly to the method using GPS for outdoor environments proposed in [13]. The first stage of clustering determines phones spatially collocated. Then the second stage clustering is done according to time. By these means, groups of collocated phones which remain together for a certain interval of time are identified, thus yielding the indoor pedestrian flocks.

1.2.3. Indoor pedestrian flock. Multimodal sensor and Wi-Fi approach

The same authors of [14] furthered their contribution in [9] by adding two more sensor modalities to the feature space spanned from the Wi-Fi signal strength data (magnetometer and accelerometer). In this case however, the first stage clustering consist of clustering according to the feature space spanned by each sensor modality individually and the Wi-Fi data. The additional features added from the accelerometer data were the windowed cross correlation and windowed movement classification comparison. The classification however, was not granular enough to determine whether someone was climbing some stairs or running or walking, thus the authors in [14] settled by determining whether the phone carrier was either or not in movement. There were two features derived from the magnetic compass data as well. The first, similarly to the one derived from the accelerometer, is the windowed cross correlation of the raw data samples. The second feature was the time since the last turn taken by the person carrying the phone.

Then, a majority voting is performed over the spatial clusters derived from the features just mentioned, with an appropriate weighting function to give priority to the feature space of Wi-Fi over the one spanned by the accelerometer and magnetometer as their individual performance was inferior to Wi-Fi's. Finally, the second stage clustering or temporal clustering is performed to determine the final flocks. Figure 3 shows a diagram detailing the process developed in [9] for flock detection.

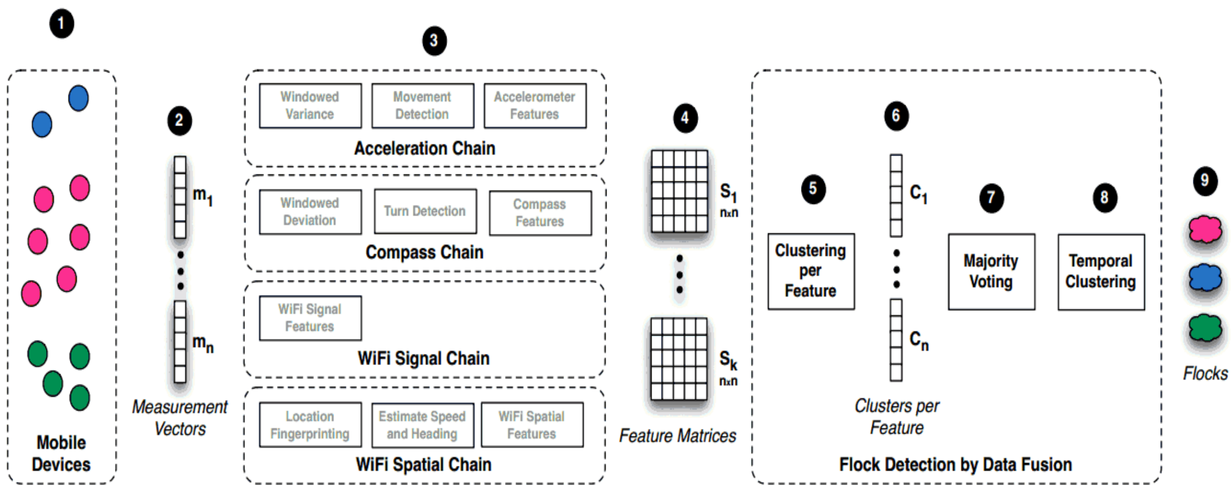

Fig 3. Application architecture for the flock detection method proposed in [9].

1.3. Objectives

This thesis objectives are to determine which features derived from both Wi-Fi signal strength data and sensor data are most suitable for implementing a pedestrian flock detection application. The experiments were conducted in order to gather data in incrementally complex scenarios, trying to simulate real world situations of pedestrian flocks. The features are then compared and depending on whether or not the groups are distinguishable from the results, the feature is determined to be useful or not.

The work written about in this thesis will be used by Teemu Leppänen in order to make a practical implementation of the pedestrian flock detection using his framework for distributed applications for heterogeneous sensor networks demonstrated in [17]. The concept of this application was published in [15]. The power consumption profiling is left to Teemu Leppänen to provide in order to constitute the score list mentioned in figure 2. This thesis is only concerned with the features to be derived by this application, which will be used to determine the pedestrian flocks; according to their accuracy, each feature will be assigned a weight in the score list in the final implementation. The sensor modalities will be fused in the same way proposed in [9], by using majority voting, thus the need for the weights.

Additionally, the change of application for gathering the sensor data for the ground truth experiments was done, because one of the objectives of this work was to test Teemu Leppänen's proposed framework [17] as much as possible.

1.4. Contribution and Thesis structure

The contribution of this thesis is to introduce new features in the feature space for pedestrian flock detection by using cross correlation directly on the raw data and on the average and standard deviation generated vectors in order to evaluate which are the best performing features for determining the similarity between devices according to their Wi-Fi and sensor data. Additionally, a variety of real-world experiments were performed to gather multi-modal sensor data using smartphones carried on by users while moving around in building indoors to apply the features and estimate their performance for pedestrian flock detection.

The introduction shows the survey of the main references and the previous work directly related to this thesis. Additionally, it explicitly states the objectives of this work and concludes by briefly describing how this thesis is structured. Chapter 2 "problem statement" describes the pedestrian flock detection in indoor environments problem alongside the reasons for simplifying the problem and the assumptions kept in mind while working on this thesis. Chapter 3 is about the experiments performed to obtain the data on which the features, explained in Chapter 4, are applied. Chapter 5 "Results", shows the obtained results of the features applied to the experiment data, divided once again in "Wi-Fi" and "Sensors" subsections. Chapter 6 "Discussion" contrasts the method proposed in this thesis with the aforementioned previous works in indoor pedestrian flock detection and propose some methods to further improve the performance of a pedestrian flock detection application. Finally, Chapter 7 "Conclusion" shows the conclusions drawn after conducting the experiments and calculating the features, explicitly stating which sensor modalities are more appropriate for indoor pedestrian flock detection. Additionally there is a subsection of summary, which succinctly explains the work done in this thesis. To finish, open challenges in this thesis work are proposed, which will be continued to be worked on by Teemu Leppänen.

2. Problem statement

Identifying pedestrian flocks in indoor environments is important for all the reasons given in the introduction, however this chapter is about what sort of scenarios are necessary to be studied in order to make meaningful experiments and what are the assumptions made in this thesis in order to simplify the problem to make it more approachable.

This is better explained using an example. Train stations for instance are, depending on the time of the day, very crowded. Identifying groups in these crowds can be a very daunting task.

How can groups be unequivocally identified in situations where GPS is not usable or Wi-Fi fingerprints are possibly not distinctive enough, as the groups of pedestrians are collocated within very close proximity? Through Wi-Fi, it may be possible to distinguish groups which are at a certain distance from each other, but certainly not groups which are very close to each other, as in the case of a train station in rush hour. That is why it is necessary to expand the feature space of the application used to detect pedestrian flocks. Namely, use additional sensors to identify the groups apart.

According to how the groups move across a given place, in this case a train station, the members of a flock are subject to distinctive "fingerprints" of the environments. People who are approaching a commercial area in the train station may be subject to different types of sounds than people who are in a train, for example: music, the sound of cars, etc. as opposed to the quiet environment in a train, with the repetitive noises of the train accelerating and braking.

In this same situation, when comparing the accelerometer readings of someone walking versus someone on a train, it is possible to deduce immediately that the patterns in the readings should be completely different, if correctly processed. Both readings would probably have a strong periodic component, the person walking around the station will probably exhibit accelerometer data which is accentuated as a product of the steps given by the pedestrian. Similarly, in terms of periodicity, the patterns of acceleration and deceleration could be identified from the accelerometer data from someone who is riding a train, however the frequency is probably lower to that of someone who is walking, as implied in [21].

The barometer data, as it will be detailed later in this thesis, can also be used to have a reference of the change in air pressure and thereby in the altitude of individuals and consequentially groups. This can be used for detecting transitions between floors a given building as shown in [18].

The task, as it has been introduced so far, can seem discouraging due to the size of the mass of people in a train station rush hour scenario, however it is possible to simplify the problem and obtain results which are still significant despite the assumptions. Rush hour in a train station, is probably the worst case scenario for a pedestrian flock detection application, as sometimes the crowds can seem to blend into a single mass of people. Besides, rush hours in train stations occur only in small windows of time at certain times and are considered the least common setting for this problem (only few stations have huge influx of commuters at certain hours a day).

The focus of this thesis will be put on identifying the features usable to distinguish groups which are more clearly constituted and apart from each other. Distinguishing groups as granularly as possible at an early stage will organically enable the application to determine when a crowd becomes so cohesive that it constitutes a big group, like in the case of train station in rush hour.

Distinguishing several groups however, is still a difficult task, therefore a bottom-up approach seems to be the most reasonable course of action to solve the problem. Therefore it was decided that the most fundamental problem to solve was to distinguish two groups of data measurements in a variety of scenarios, from the perspective of several sensor modalities.

3. Data collection and Ground Truth

In this chapter, the experimental settings and study of the sensor modalities are presented. In addition to the Wi-Fi signal strength, the sensor modalities observed were barometer and tri-axial sensors such as magnetometer, accelerometer and gyroscope. Two categories were devised for these sensors. Inertial category is for accelerometer and gyroscope sensors, which describe the movement of the person carrying the smartphone. Comparatively, the magnetometer and barometer, describe the environment in which the person carrying the phone is located at any given time. The initial assumption is that the tri-axial magnetometer is considered to describe the environment because its samples can be considered magnetic fingerprints of rooms [12] and the variation on the barometer data can be used to distinguish floors in a given building [18].

Fig 4. Ten Samsung Galaxy sIII smartphones used in the experiments.

In the ground truth experiments from two to ten Samsung Galaxy sIII smartphones were utilized (two models: GT-I9300 and SC-06D), with integrated sensors for the aforementioned modalities, the phones are shown in Figure 4. An Android application was developed to collect the Wi-Fi data, which was deployed in all devices. The experiments for gathering the sensor data have a variety of scenarios with increasing complexity. For Wi-Fi, the emphasis was put into creating scenarios that allowed to observe how fine or coarse Wi-Fi based comparison is. For the sensor modalities, the experiment were designed so it could observed whether the features derived from the data, as posterior analysis shows, were similar enough to identify phones on a same group.

3.1.Data collection applications

For collecting the Wi-Fi data and sensor data a series of experiments were performed. Initially, an Android application was developed to collect sensor data from all the available sensors in the Samsung Galaxy sIII aforementioned, Wi-Fi signal strength per access point and recording sound. The sensor data and visible Wi-Fi networks were displayed on screen and by switching the log button on the sampled data from both Wi-Fi and sensors started to be recorded in text files with the time stamps being affixed to each sample (both the time stamp using the *System.nanoTime()* method and *Calendar.getTimeInMillis()* were taken). The sound was recorded as well, whenever the "log" switch was on. The graphic interface of the application is shown in Figure 5.

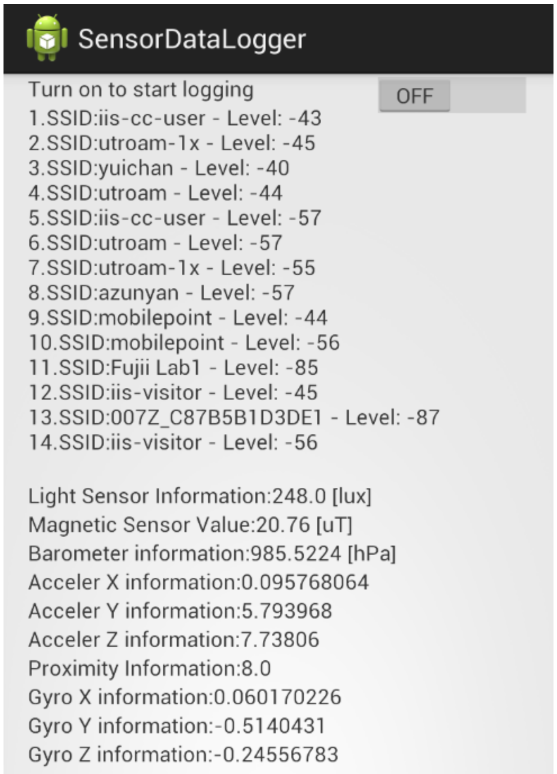

Fig 5. Graphic user interface of the sensor, Wi-Fi and audio logging Android application

The sensor data was sampled at 5 [Hz] and the Wi-Fi data at 0.5[Hz], in average. To the best of the knowledge gained, there is no method to specify the sampling frequency of either sensors or Wi-Fi. For the sensors specifically, it was possible to set the sampling frequency with a limited set of macro, however setting a specific number of samples per second is not possible. Similarly in terms of limitation with Wi-Fi, the scanning of wireless networks is managed by an object in Android which pulls the data whenever available. Unfortunately, due to hardware limitations, the frequency of how often the visible wireless networks can be pulled is low. Therefore, it was decided to execute a thread that would run Wi-Fi scans every 2 seconds. If the thread execution was set to times lower than 2 seconds, there is no improvement in terms of sampling frequency. To calculate the sampling frequency at which the Wi-Fi and sensor data were taken, the difference between the time stamps of 2 consecutive samples was averaged for every sample then calculated the reciprocal for both sensor and Wi-Fi data.

The aforementioned application was used to gather the ground truth Wi-Fi data for the according experiments. However, because this research was conducted cooperatively with Teemu Leppänen, one of the objectives was also to test the framework implemented and shown in the demo in [17], which will also be used to implement the application proposed in [15]; the framework shown in [17] was used for gathering the sensor data for the according ground truth sensor experiments. Unlike the previous application used for data gathering, the framework developed and shown in [17] would store the sensor data in a database located at the phone, and then the data could be accessed remotely, via SQL querying. This application was used exclusively in the experiments for gathering the ground truth experiment data of the scenarios conducted to test the features on sensor data (not Wi-Fi).

The detailed description of the framework designed and developed by Teemu Leppänen escapes the scope of this thesis. It will be however used to implement the application proposed in [15].

3.2. Ground truth experiments for Wi-Fi

Using the application initially developed, Wi-Fi and sensor data were gathered while recording audio in a series of 8 scenarios. Wi-Fi data was the only deeply analyzed in these scenarios, because the first objective in mind was to observe, whether the indoor pedestrian flocks were distinguishable using solely Wi-Fi as it was done in [14].

The scenarios have been summarized in Table 2, later on each one of them will be detailed; and it will also be discussed the reason and necessity for each scenario, and whether or not Wi-Fi was suitable for pedestrian flock detection. The 8 scenarios for this experiment had 10 participants in total, which in most of the scenarios were divided into 2 equally sized groups. Most of the scenarios for both Wi-Fi and sensor data collection only have two groups, as this is the simplest case. It is necessary to determine if it is possible to distinguish between at least 2 groups, with the proposed method, in order to escalate the experiment size and experiment with more than two flocks. The experiment took place at the 6th and 4th floor of the west wing of the E block of the Komaba II campus building, the schematic of the building is shown in Figure 6.

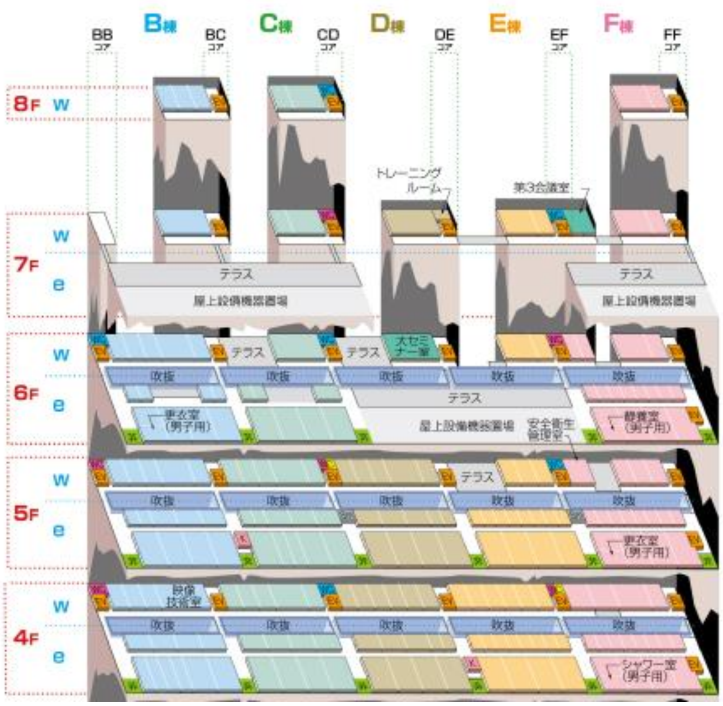

Fig 6. Schematic map of the Komaba II campus building.

Henceforth the indexes from Table 2 will be used as the main way to refer to the experiment scenarios.

Table 2. Wi-Fi Experiment Scenarios: summary.

Index	Floor	Flocks size	Description/Axes involved
1	6th	5,5	Crossing each other
2	6th	5,5	Parallel trajectory
3	6th	5,5	Perpendicular crossing
4	6th	5,4,1	2 flocks crossing. Idle person
5	6th	5,4,1	2 flocks perpendicular crossing.
6	6th and 4th	5,5	Parallel trajectory
7	6th and 4th	5,5	Crossing each other
8	6th and 4th	5,5	Perpendicular crossing

Wi-Fi Scenario 1: Two equally sized groups of 5 persons cross each other in the Komaba campus building. The initial position of each flock is the end position of the other one. The purpose of this experiment is to observe the similarity between pairs of phones, which are grouped into significantly set apart groups, both at the beginning and ending of the experiment. A schematic of the scenario is shown in Figure 7.

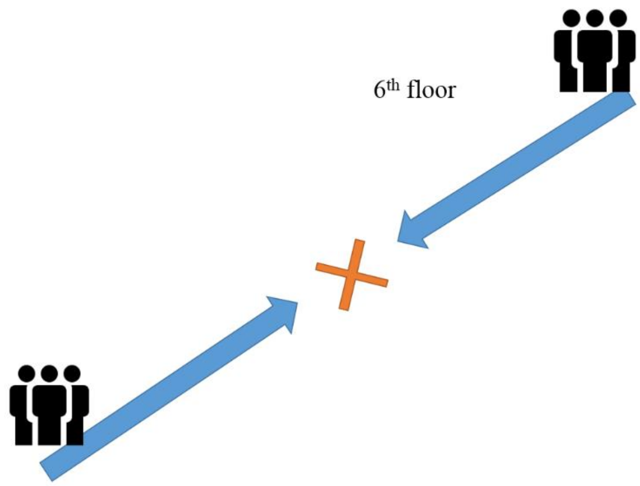

Fig 7. Schematic of the crossing flocks scenario.

Wi-Fi Scenario 2: Two equally sized groups of 5 persons walk in a parallel trajectory in the 6th floor of the Komaba campus building. Both flocks have the same initial position and direction of movement, but one flock starts moving before the second one. The purpose of this scenario is to stress the proposed method by seeing if the two flocks are differentiated despite the identical trajectories and positions. A schematic of the scenario is shown in Figure 8.

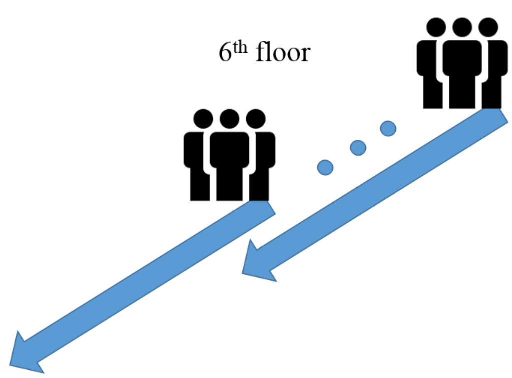

Fig 8. Schematic of the scenario with flocks moving in the same direction with delay.

Wi-Fi Scenario 3: Two equally sized groups of 5 persons walk in a perpendicular trajectory in the 6th floor of the Komaba campus building. The initial position of each flock is closer than the one of scenario 1, but the direction is perpendicular maintaining the same point of crossing than the one in scenario 1. The purpose of this scenario is to set flocks in different yet very proximate initial points. A schematic of the scenario is shown in Figure 9.

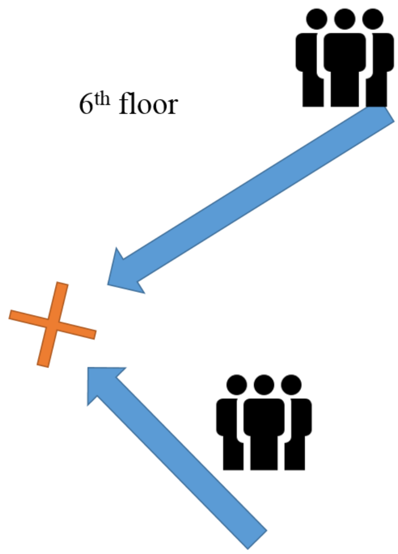

Fig 9. Schematic of crossing flocks with perpendicular trajectories.

Wi-Fi Scenario 4: Two groups, one of size 4 and the other one size 5, walk in opposite directions with an idle person standing in the crossing point of the 2 groups. The initial position of each flock is the end position of the other one. The purpose of this scenario is to observe to which group is the person standing idly most similar. A schematic of the scenario is shown in Figure 10.

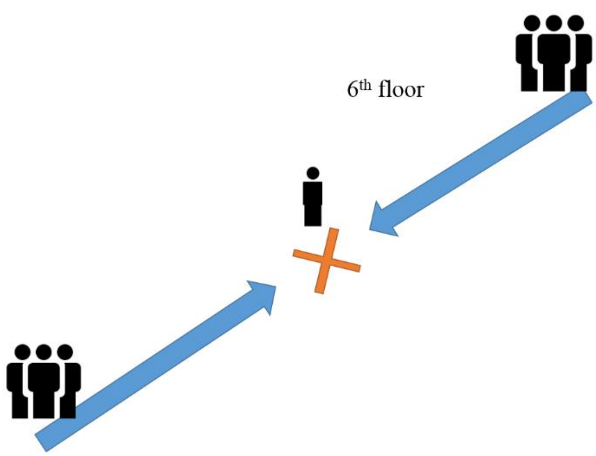

Fig 10. Schematic of crossing flocks with an idle person standing in the crossing point.

Wi-Fi Scenario 5: Two groups, one of size 4 and the other one size 5, walk in perpendicular directions with an idle person standing in the crossing point of the 2 groups. Once again, the initial position of each flock is closer than the one of scenario 4, but the direction is perpendicular maintaining the same point of crossing (where the idle person stands) as in scenario 4. The purpose of this scenario is to observe to which group is the person standing idly most similar and bringing the initial position of the flocks closer to see how the similarity of the measurements is impacted. A schematic of the scenario is shown in Figure 11.

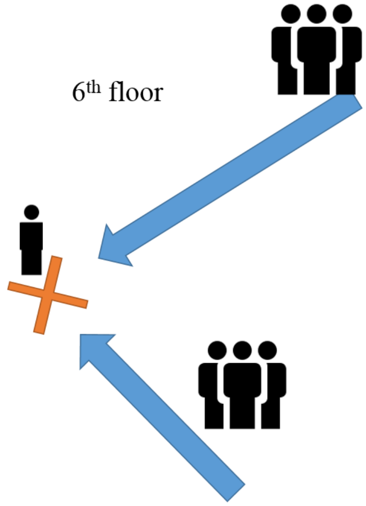

Fig 11. Schematic of perpendicularly crossing flocks with an idle person standing at the crossing point.

At this point of the experiments, it was assumed that only the data gathered in these scenarios was going to be used, therefore it was decided to do these two scenarios with an idle person standing in the crossing point because it was determined that it could be possible to tell the flocks apart from the idle person standing using the accelerometer data in case the Wi-Fi data was not dissimilar enough. This was later discarded, as the experiments for gathering sensor data had completely different scenarios, focusing on generating data that was distinctively different, this will be detailed in the next subsection.

Similarly to the accelerometer just aforementioned, the following three scenarios were done in two different floors of the campus building, in order to visualize the groups by their respective barometer data.

Wi-Fi Scenario 6: Two equally sized groups of 5 persons walk in a parallel trajectory in the 6th and 4th floor of the Komaba campus building. Both flocks have the same initial position but in different floor and same direction of movement, but one flock starts moving before the second one. The purpose of this scenario is to observe whether Wi-Fi data is different enough in the case the flocks are located very close vertically (the distance between floor 6 and 4). Additionally, the results from the data gathered by the barometer sensor were contrasted, as it is later explained. A schematic of the scenario is shown in Figure 12.

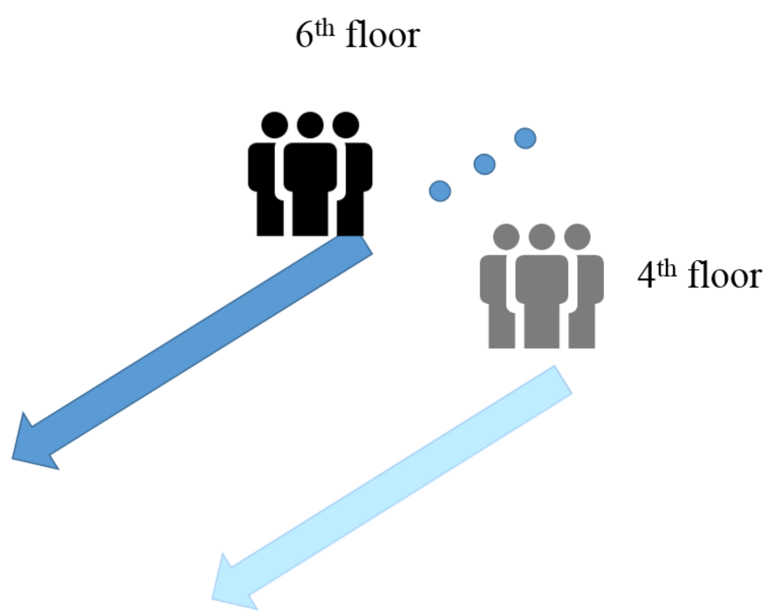

Fig 12. Schematic of the scenario with flocks moving in the same direction with delay in different floors

Wi-Fi Scenario 7: Two equally sized groups of 5 persons cross each other in the 6th and 4th floor of the Komaba campus building. The initial position of each flock is the end position of the other one, but in different floors. The purpose of this scenario is to observe what was considered as the best case scenario in terms of dissimilarity; here the groups of people are as far as possible in their initial and final position. A schematic of the scenario is shown in Figure 13.

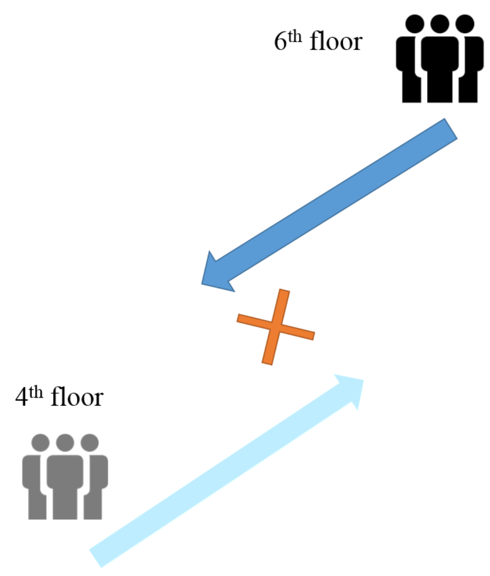

Fig 13. Schematic of the scenario with crossing flocks in different floors.

Wi-Fi Scenario 8: Two equally sized groups of 5 persons cross each other in a perpendicular trajectory in the 6th and 4th floor of the Komaba campus building. Both flock's initial and final position is closer than in scenario 7. This scenario is interesting to observe in order to contrast the results with scenario 3 and visualize the impact in the dissimilarity of the groups in different floors. A schematic of the scenario is shown in figure 14.

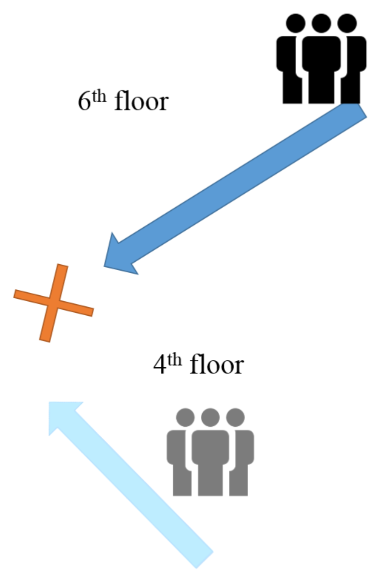

Fig 14. Schematic of crossing flocks with perpendicular trajectories in different floors.

3.3.Ground truth experiments for sensors

From this point onward in the experiments, the application shown in the demo [17] was utilized for gathering sensor data. The purpose of this was to test the application in [17], as this research was done cooperatively with the main author of [17], Teemu Leppänen, and testing this framework was also one of the objectives.

The scenarios for collecting the ground truth data for sensor data were made incrementally complex and are summarized in Table 2. Unlike Wi-Fi, where it was only required to investigate whether or not with the given density of access points in campus, it was possible to determine the groups; the scenarios for sensor data were designed in order to compare phones registering data which was supposed to be surely different from 2 different groups. Unfortunately, it was not possible to gather the same amount of people for these experiments, thereby only 2 to 3 phones were used, from the ones shown in Figure 4. The result are still valid, if the scenarios are designed purposely to register sensor data which it is believed to be distinctive.

Table 3. Summary of ground truth collection experiments for sensor data.

Name of the experiment	**Main sensors to be observed and experiment duration**	**Summary**
Experiment 1&2. Still scenario.	Accelerometer (2 and 5 minutes respectively)	2 phones were carried around the Sezaki laboratory. Moving phone was compared to still phone. 2 still phones were compared.
Experiment 1&2. Moving scenario.		
Experiment 3. Same turn.	Gyroscope and Magnetometer (2 minutes)	2 phones were carried while performing turns in same and opposite directions
Experiment 4. Different		

turn.		finishing in different rooms.
Experiment 5.	Acc., Gyro. And Mag.	Real world scenario.

The first experiment only had two scenarios and two phones. The first scenario was to keep one of the phones on the desk and the other one would be carried around by someone. The second scenario was to keep both phones on the desk. This experiment was performed twice, the first time, the experiment lasted 2 minutes and the second time, 5 minutes. Additionally the second time, the gyroscope data was also gathered, and the scenario where one of the phones was moving required the person to go up and down some stairs, as shown in Figure 15, in order to gather data believed to be very dissimilar.

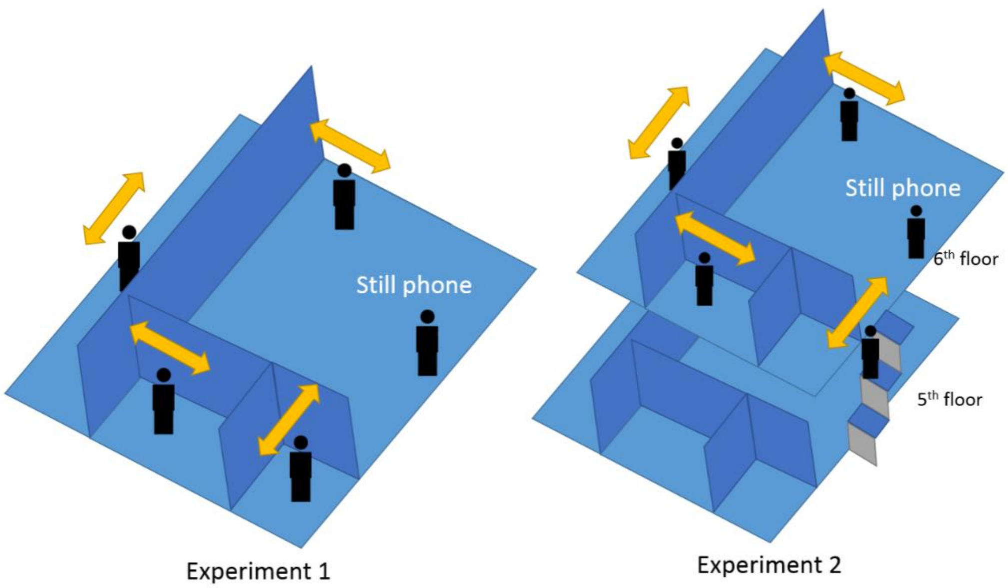

Fig 15. Schematic for experiment 1 and 2 for accelerometer data.

The purpose of these two experiments was to reproduce in a first instance what the authors in [9] did for the accelerometer features. The authors tried to classify the data into activities and then compare the classification in order to cluster the phones according to the classification result. However, according to [9], given the features they calculated from the data, they could only determine whether a phone was being carried by someone moving or not. Therefore, it was decided that the first step should be to reproduce this experiment. According to the authors in [11], it is possible to differentiate the data using the magnitude of the accelerometer data in the time domain, as spectral features are known to be much more effective [11].

The next experiment, similarly to the previous one, has two scenarios. This time however, both phones were being carried inside the laboratory. The first scenario consisted on having two persons walking in the laboratory carrying the smartphones and perform a turn in the same direction. The second scenario was to make these two phones acquire data from persons giving turns in different directions. Additionally, one of the persons finishes its trajectory in a different room. This can be observed schematically in Figure 16.

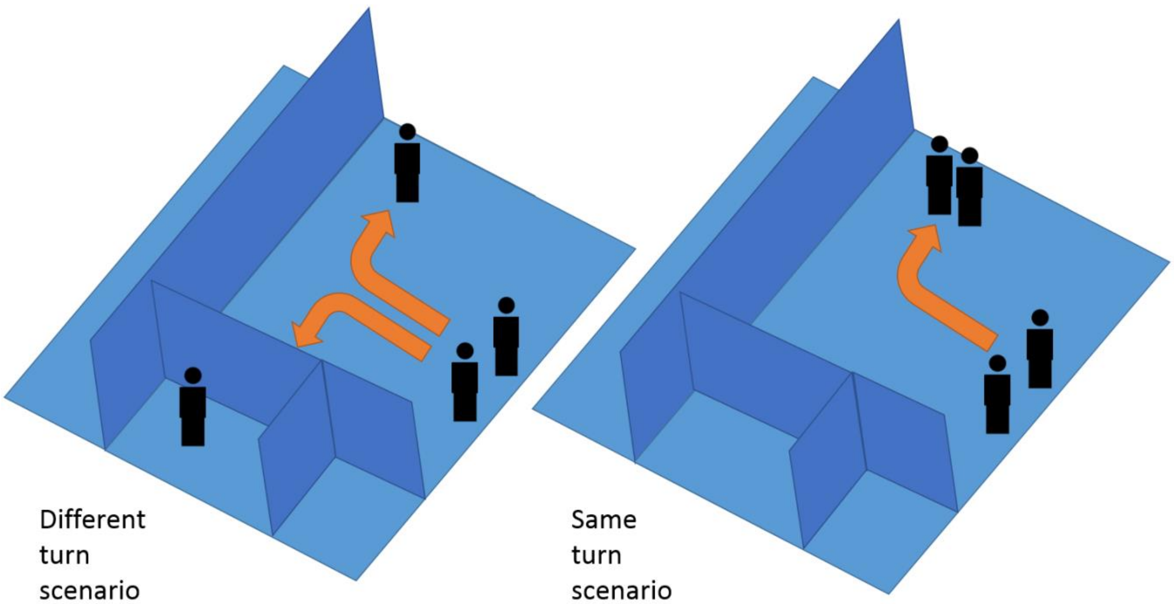

Fig 16. Schematic of same turn and different turn scenario for experiment 3 and 4 for sensor data ground collection. Gyroscope and magnetometer comparison.

As it will be later on detailed in the results chapter, the results for the accelerometer features were not satisfactory, the attention was shifted towards gyroscope and magnetometer in this particular experiment. To achieve so, it was necessary gathering data for an event in which the gyroscope data would be surely different from phones. Therefore comparing people turning in opposite and equal directions was decided to be the simplest scenario and the one which fitted best this purpose.

Additionally, it was decided that the tri-axial magnetometer should be taken into account, as it has been used for fingerprinting in indoor positioning application [12]. For this purpose, both individuals had, for the end of their trajectories, different rooms inside the laboratory.

As this last experiment was very short (2 minutes as well), it was decided to make a longer experiment, similarly to what was done with experiment 1 and 2 for the accelerometer. However for experiment 4 there were some difficulties retrieving the magnetometer data from the application used for gathering the data.

In experiment 5 there were three phones, instead of two like all the previous experiments. Two phones were carried by the same person, simulating a group of 2 persons walking together. One phone remained inside the lab on a desk without particularly moving. The moving phones were carried through a variety of environments in order to generate dissimilar sensor data to the data being gathered by the phone in the lab. The moving phones were carried even to an outdoor environment. The schematic of the experiment can be appreciated in Figure 17.

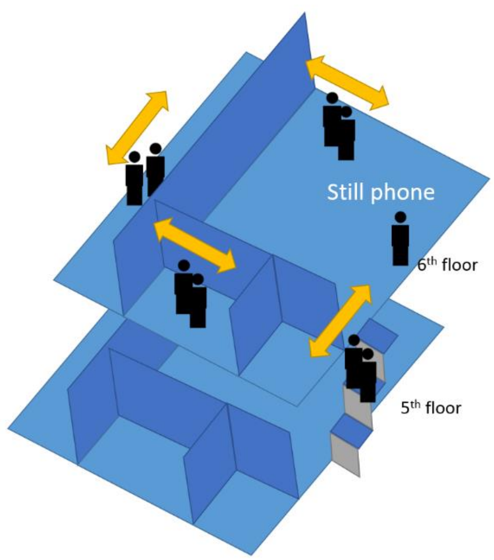

Fig 17. Schematic of experiment 5 for ground truth collection of sensor data. 2 phones carried by an individual and 1 static phone.

The purpose of experiment 5 was to capture gyroscope data in a longer time frame. The results will be further detailed in the results chapter, but because the gyroscope data comparison yielded good enough results for comparing phones, it was deemed necessary to observe if the same results held for a longer time frame.

It was also necessary to compare phones which have been exposed to outdoor environments with phones in indoor environment. From the perspective of the magnetometer, the differences can be determinant, as the metal in the structure of the building can introduce noise [12]. This can in turn be taken as a fingerprint of a given room as well.

Additionally, it was decided to gather more data from different walking activities, namely walking up and down stairs, in order to compare once again the accelerometer data. The features derived this time were spectral features, as their performance according to [11], is high. The results will be detailed later on.

4. Features extracted from the ground truth data

In our earlier work, we presented conceptual distributed system architecture, enabling crowd sensing-based pedestrian flock detection application [15].

For the realization of this architecture, it is necessary to first consider the participatory collection of multimodal sensor data and the features derived from the data, which enables to further the real-world application. Two methods were utilized for comparing both raw samples and features derived from Wi-Fi and sensor data. The first was the cosine similarity, which for the case of Wi-Fi was the only one used. The second one was the normalized cross correlation, which along with the cosine similarity was used on the sensor raw data and features.

Cosine Similarity. Used to calculate the similarity between couples of equally sized processed data vectors. The cosine similarity is equal to the dot product of the vectors divided by the product of their magnitudes.

Cross Correlation. The normalized cross correlation, which along with the cosine similarity, are used on the sensor raw data and features. If normalized the value can be interpreted as percentage of similarity. The values obtained are a function, for which only the maximum value is observed. It represents the maximum similarity of the two functions being compared for a given lag between the functions. The "*numpy.correlate()*" function was used in Python from the "numpy" library of mathematical functions.

Before applying these similarity functions, first is necessary to process the data gathered by the smartphones. The treatment of data varied according to the type of data.

For the Wi-Fi data, each signal strength measurement was grouped according to a predefined group size. This step is crucial because it allows to calculate the feature over a group of measurements which represent a temporal vicinity, instead of calculating it over the entirety of the data for a given scenario. Then the signal strength values were averaged for each MAC address in a group, followed by applying the cosine similarity and finally once again calculating the average of the cosine similarity values yielded between each pair of groups. See Figure 18 for a schematic of the process.

For the sensor data, as the log files did not contain data for an exclusive scenario, it was necessary to trim some of the measurements according to the timestamps of the data. This step was not necessary for Wi-Fi, as the application used for gathering data was different, as it was mentioned. For the sensor data, the highest starting timestamps and the lowest one were taken, and trimmed the data according to these two boundaries. It was assumed that the experiment occurred within those bounds; this was possible because the time in the phones was retrieved from the internet and therefore it was possible to assume there was an absolute source of synchronicity. The methods used for processing the data before comparing were:

Grouping. The data was arranged into groups of 40 measurements. This step is crucial because it allows to calculate the feature over a group of measurements, representing temporal vicinity, instead of calculating it over the entirety of the data, for the given scenario.

Sliding Window. The sliding window function is used whenever there is sudden changes in the data, such as the barometer data from a person going between floors, or cyclic events (such as walking). The windowing function establishes a number of measurements with which the feature is calculated, where overlap of 50% is used whenever the window slides to the next position.

Average and Standard Deviation. Over the grouped or windowed measurements, the average and standard deviation is calculated.

Magnitude. The magnitude of the first Fast Fourier Transform (FFT) coefficients is calculated. This is exclusively used for accelerometer data, because of its outstanding performance, as suggested in [11]. In some cases the magnitude of the tri-axial sensors' data is also calculated, which consists on computing the square root of the summation of squared value of each individual axis.

Over either groups or windows of measurements, the average, standard deviation and the magnitude of the first Fast Fourier Transform (FFT) coefficients were calculated. The FFT was exclusively used for accelerometer data because its performance, as suggested in [11]. In some cases the magnitude of some of the tri-axial sensors data was also calculated.

4.1. Wi-Fi signal strength similarity

The Wi-Fi signal strength measurements were taken at a rate of 0.5 Hz. Initially, all the samples whose power value was below -70 dBm were filtered out, as aforementioned, the threshold was determined empirically, by searching the value that maximized the similarity gap between phones in different groups, [8] ratifies that -70 dBm is appropriate. Either increasing or decreasing the threshold value had a negative impact in the analysis. If the value is increased, few measurements will remain after the filtering process, making the group of measurements not representative of their time vicinity. On the contrary, if the threshold is set lower than -70 dBm, too many low power samples will be taken into account for the analysis, introducing noise to the calculation.

The following step is to divide the full vector of measurements into equally sized groups. Previous methods take into consideration the average of the full length Wi-Fi signal strength vector [9]; however it was deemed essential to the proposed method preserving the time vicinity information, hence the additional partitioning step. Analogously to the determination of the power threshold, it was empirically determined that, by partitioning the total vector into groups of 40 measurements, the similarity gap between phones in different groups in the experiment was maximized (as it will be later discussed in the result chapter, having similarity measurements from different groups as distanced from each other as possible, increases the performance for identifying the groups using Wi-Fi). Each sample contains several measurements (as many as the number of visible access points at the time of the sample). Over each group of measurements, the average power for each MAC address was calculated. Having partition size larger than 40 would mean to lose the time vicinity information, similarly, having less than 40 measurements per group would mean to take into consideration fewer measurements than a sample contains. This is due to the fact that each time the application sampled the visible access points, there were between 10 to 25 access points visible at all times, each constituting a measurement. This was also done taking into consideration that the measuring rate was of 0.5 Hz, meaning that every 2 seconds 10 to 25 of these samples were taken.

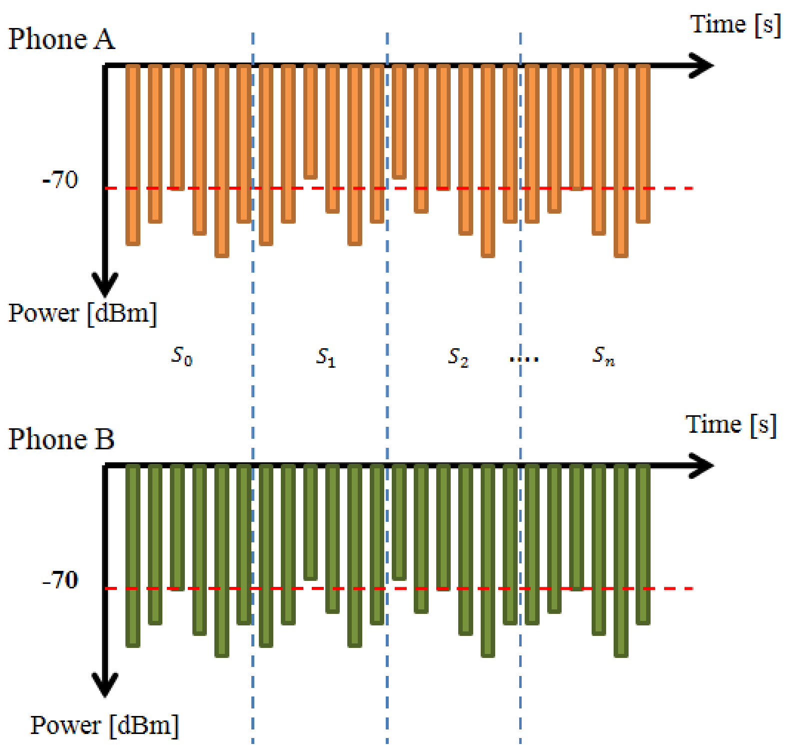

Fig 18. Schematic of the Wi-Fi similarity calculation process.

After averaging the signal strength value over for each repeated MAC address in a group of measurements, the cosine similarity between the segments of the vector of Wi-Fi measurements was finally calculated. Then the segment-wise similarity values were averaged obtaining a similarity value for each pair of smartphones, or simply put, the average value of the n values of S shown in Figure 18, which describes the complete process for calculating the cosine similarity, schematically for any pair of phones.

4.2. Sensor Modalities features

In addition to the Wi-Fi signal strength, the sensor modalities we observed are barometer and tri-axial sensors such as accelerometer, gyroscope and magnetometer. After applying the trimming function according to the timestamps (using the highest starting timestamp as the lower bound and the smallest finishing timestamps as the upper bound), the features were calculated for each sensor.

The size of the window was 2 and 10 seconds long [11]; (because the sampling rate was of 5 Hz in average, it is equivalent to say that the window covered 10 or 50 samples at any given time of the calculation) with 50% overlap [11], as it is more likely to capture a cyclic event, such as a person's steps, in that interval. The overlap was necessary because the start of events e.g. steps in case of the accelerometer or turns in case of the gyroscope, is unknown.

Accelerometer. The sliding window function was applied to the accelerometer data. This was required because of the cyclic nature of walking and the uncertainty of when the steps are given. The first features derived were calculating the average and standard deviation after applying the grouping function. However in [9], it was already determined that the data from the accelerometer was not descriptive enough to allow coarse movement classification. For the accelerometer the features were both spectral and from the time domain. For the spectral features, the similarity of the magnitude of the first five coefficients of the FFT was calculated. To compare these magnitudes, the maximum value of the normalized cross correlation function was observed.

The time domain features consisted of the average and standard deviation per axis of the accelerometer data. The method for comparison was both the cosine similarity and the maximum value of the cross correlation function, however due to these values being close, it was decided to only show the value for the cosine similarity.

Barometer. The barometer was already studied in [18]. Barometer data was simultaneously gathered during the Wi-Fi experiments, to see how effectively it was possible to reach the same detected groups using different modalities. The purpose of the windowing function for the barometer data, despite it being regular, as shown in Figure 21, was to mask big variations on data. When the person carrying the phone goes to a different floor; the values of the barometer signal change abruptly and are maintained relatively constant as long as the person stays on the same floor, therefore keeping the sharp step function-like shape of the barometer signal is desirable. The maximum value of the normalized cross correlation was calculated for the raw data and the windowed average and standard deviations values.

Gyroscope. No windowing function was used as turning events lack periodicity. The maximum value of the normalized cross correlation function was then calculated over the raw data and magnitude.

Magnetometer. No windowing function was applied as these features are used to capture turning events and magnetic fingerprints of the building, respectively. Unlike walking and turning events, this data lacks periodicity. The maximum value of the normalized cross correlation function was therefore calculated over the raw data per axis and the magnitudes for both the experiment 3 and 4 from table 3. It was also necessary to compare phones, which had been exposed to outdoor environments with phones in indoor environment. From the perspective of the magnetometer, the differences can be determinant, as the metals and electrical circuits in the structure of the building can alter the magnetometer readings which in outdoor environments are completely clear. This can be used as a form of fingerprinting according to [12].

5. Results: Similarity of the features

In this section the obtained results from comparing the ground truth to the features applied to the data for Wi-Fi and the sensors gathered during the ground truth experiments are shown. For the case of the experiments to gather the Wi-Fi data, the barometer data was also gather, in order to be able to contrast the results obtained for barometer and Wi-Fi, reaching the same conclusion in terms of group detection. Then the results for the analysis of the ground truth data for the sensor are shown. The section finishes with some comments and conclusions drawn in the "discussion" subsection.

5.1. Similarity values for Wi-Fi

Concerning Wi-Fi signal strength measurements, extended from the work in [10], the smartphone pair-wise cosine similarity values for all the scenarios in Table 2 are shown from Table 5 to 12. These scenarios represent situations in which the two groups of indoor pedestrians were walking on same and different floors. It is of great interest to observe these two types of scenarios, as the same conclusion can be drawn from Wi-Fi and the barometer data.

The colors in the tables of this section represent whether the value surpasses or not the average or median similarity value of each scenario. Table 4 shows the ideal case, where the two groups are clearly distinguishable. The red values would represent the similarity value between smartphones in different groups equal to 0 ideally and the green values represent the phones on the same group with similarity value ideally equal to 1. As the tables are symmetrical respect the diagonal, the values above the diagonal have been omitted.

Row and column labels from one to five correspond to phones in one group and the ones from six to ten are the other group. As it can be appreciated from the Tables 5 to 12, the similarity values for phones on a same group are higher than the ones for phones in different groups. The difference becomes even more accentuated for scenarios 6, 7 and 8, where the groups where in different floors.

Table 4. Sample: Ideal similarity value. Average is 0.55 and median is 0.5

	1	2	3	4	5	6	7	8	9	10
1	1.00									
2	1.00	1.00								

3	1.00	1.00	1.00							
4	1.00	1.00	1.00	1.00						
5	1.00	1.00	1.00	1.00	1.00					
6	0.00	0.00	0.00	0.00	0.00	1.00				
7	0.00	0.00	0.00	0.00	0.00	1.00	1.00			
8	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00		
9	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	1.00	
10	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	1.00	1.00

Table 5.Average threshold scenario 1. Average is 0.63 and standard deviation is 0.26

	1	2	3	4	5	6	7	8	9	10
1	1.00									
2	0.87	1.00								
3	0.81	0.85	1.00							
4	0.91	0.78	0.82	1.00						
5	0.76	0.74	0.80	0.75	1.00					
6	0.32	0.32	0.33	0.35	0.36	1.00				
7	0.30	0.34	0.48	0.27	0.41	0.74	1.00			
8	0.37	0.35	0.51	0.37	0.50	0.71	0.77	1.00		
9	0.35	0.31	0.51	0.27	0.38	0.71	0.74	0.78	1.00	
10	0.40	0.27	0.42	0.31	0.39	0.70	0.65	0.78	0.76	1.00

Table 6. Average threshold scenario 2. Average is 0.77 and standard deviation is 0.16

	1	2	3	4	5	6	7	8	9	10
1	1.00									
2	0.79	1.00								
3	0.74	0.68	1.00							
4	0.68	0.80	0.60	1.00						
5	0.77	0.65	0.71	0.60	1.00					
6	0.54	0.72	0.67	0.78	0.49	1.00				
7	0.55	0.63	0.60	0.69	0.44	0.74	1.00			
8	0.58	0.78	0.62	0.73	0.47	0.80	0.83	1.00		
9	0.80	0.87	0.92	0.87	0.55	0.96	0.92	0.91	1.00	
10	0.60	0.81	0.68	0.77	0.50	0.81	0.72	0.86	0.92	1.00

Table 7. Average threshold scenario 3. Average is 0.67 and standard deviation is 0.21

	1	2	3	4	5	6	7	8	9	10
1	1.00									
2	0.82	1.00								
3	0.88	0.81	1.00							
4	0.76	0.83	0.78	1.00						
5	0.80	0.84	0.84	0.92	1.00					
6	0.49	0.39	0.57	0.56	0.54	1.00				
7	0.52	0.38	0.59	0.46	0.51	0.58	1.00			
8	0.53	0.55	0.60	0.63	0.63	0.62	0.50	1.00		
9	0.54	0.68	0.48	0.61	0.61	0.43	0.33	0.59	1.00	
10	0.59	0.40	0.66	0.46	0.51	0.67	0.55	0.48	0.27	1.00

Table 8.Average threshold scenario 4. Average is 0.66 and standard deviation is 0.22

	1	2	3	4	5	6	7	8	9	10
1	1.00									
2	0.69	1.00								
3	0.84	0.62	1.00							
4	0.71	0.84	0.59	1.00						
5	0.53	0.68	0.50	0.62	1.00					
6	0.36	0.55	0.29	0.61	0.62	1.00				
7	0.61	0.40	0.33	0.33	0.35	0.64	1.00			
8	0.47	0.30	0.24	0.36	0.43	0.59	0.81	1.00		
9	0.54	0.48	0.47	0.60	0.59	0.63	0.67	0.67	1.00	
10	0.62	0.54	0.42	0.52	0.58	0.75	0.67	0.69	0.83	1.00

Table 9. Average threshold scenario 5. Average is 0.60 and standard deviation is 0.20

	1	2	3	4	5	6	7	8	9	10
1	1.00									
2	0.73	1.00								
3	0.88	0.55	1.00							
4	0.78	0.86	0.51	1.00						
5	0.80	0.79	0.63	0.83	1.00					
6	0.53	0.49	0.51	0.50	0.51	1.00				
7	0.59	0.42	0.42	0.47	0.54	0.88	1.00			
8	0.66	0.43	0.42	0.48	0.48	0.81	0.53	1.00		
9	0.59	0.51	0.42	0.54	0.62	0.91	0.77	0.60	1.00	
10	0.62	0.59	0.42	0.62	0.67	0.77	0.70	0.66	0.75	1.00

Table 10. Average threshold scenario 6. Average is 0.43 and standard deviation is 0.42

	1	2	3	4	5	6	7	8	9	10
1	1.00									
2	0.63	1.00								
3	0.81	0.51	1.00							
4	0.57	0.53	0.63	1.00						
5	0.65	0.34	0.79	0.50	1.00					
6	0.00	0.00	0.00	0.00	0.00	1.00				
7	0.00	0.00	0.00	0.00	0.00	0.65	1.00			
8	0.00	0.04	0.00	0.00	0.00	0.71	0.72	1.00		
9	0.00	0.00	0.00	0.00	0.00	0.82	0.83	0.83	1.00	
10	0.00	0.00	0.00	0.00	0.00	0.84	0.79	0.82	0.91	1.00

Table 11. Average threshold scenario 7. Average is 0.41 and standard deviation is 0.40

	1	2	3	4	5	6	7	8	9	10
1	1.00									
2	0.46	1.00								
3	0.62	0.60	1.00							
4	0.54	0.72	0.78	1.00						

5	0.76	0.59	0.78	0.74	1.00					
6	0.00	0.02	0.02	0.02	0.00	1.00				
7	0.00	0.11	0.13	0.13	0.14	0.12	1.00			
8	0.00	0.01	0.00	0.00	0.00	0.68	0.12	1.00		
9	0.00	0.00	0.00	0.00	0.00	0.84	0.00	0.77	1.00	
10	0.00	0.03	0.05	0.05	0.04	0.77	0.24	0.72	0.83	1.00

Table 12. Average threshold scenario 8. Average is 0.42 and standard deviation is 0.40

	1	2	3	4	5	6	7	8	9	10
1	1.00									
2	0.89	1.00								
3	0.54	0.54	1.00							
4	0.76	0.77	0.65	1.00						
5	0.76	0.77	0.64	0.76	1.00					
6	0.00	0.00	0.04	0.00	0.00	1.00				
7	0.00	0.00	0.07	0.00	0.00	0.58	1.00			
8	0.00	0.00	0.05	0.00	0.00	0.62	0.60	1.00		
9	0.00	0.00	0.00	0.00	0.00	0.49	0.62	0.62	1.00	
10	0.00	0.00	0.08	0.00	0.00	0.67	0.65	0.64	0.54	1.00

The following tables, from Table 13 to Table 20, have been colored according to whether the values in the table surpassed the median instead of the average. The difference is minimal but it is necessary to try different criteria for the group determination. Once again, the indexes from 1 to 5 of every table represent one group, and the indexes from 6 to 10 are the other group.

Table 13. Median threshold scenario 1. Median is 0.71

	1	2	3	4	5	6	7	8	9	10
1	1.00									
2	0.87	1.00								
3	0.81	0.85	1.00							
4	0.91	0.78	0.82	1.00						
5	0.76	0.74	0.80	0.75	1.00					
6	0.32	0.32	0.33	0.35	0.36	1.00				
7	0.30	0.34	0.48	0.27	0.41	0.74	1.00			
8	0.37	0.35	0.51	0.37	0.50	0.71	0.77	1.00		
9	0.35	0.31	0.51	0.27	0.38	0.71	0.74	0.78	1.00	
10	0.40	0.27	0.42	0.31	0.39	0.70	0.65	0.78	0.76	1.00

Table 14. Median threshold scenario 2. Median is 0.77

	1	2	3	4	5	6	7	8	9	10
1	1.00									
2	0.79	1.00								
3	0.74	0.68	1.00							

4	0.68	0.80	0.60	1.00						
5	0.77	0.65	0.71	0.60	1.00					
6	0.54	0.72	0.67	0.78	0.49	1.00				
7	0.55	0.63	0.60	0.69	0.44	0.74	1.00			
8	0.58	0.78	0.62	0.73	0.47	0.80	0.83	1.00		
9	0.80	0.87	0.92	0.87	0.55	0.96	0.92	0.91	1.00	
10	0.60	0.81	0.68	0.77	0.50	0.81	0.72	0.86	0.92	1.00

Table 15. Median threshold scenario 3. Median is 0.61

	1	2	3	4	5	6	7	8	9	10
1	1.00									
2	0.82	1.00								
3	0.88	0.81	1.00							
4	0.76	0.83	0.78	1.00						
5	0.80	0.84	0.84	0.92	1.00					
6	0.49	0.39	0.57	0.56	0.54	1.00				
7	0.52	0.38	0.59	0.46	0.51	0.58	1.00			
8	0.53	0.55	0.60	0.63	0.63	0.62	0.50	1.00		
9	0.54	0.68	0.48	0.61	0.61	0.43	0.33	0.59	1.00	
10	0.59	0.40	0.66	0.46	0.51	0.67	0.55	0.48	0.27	1.00

Table 16. Median threshold scenario 4. Median is 0.62

	1	2	3	4	5	6	7	8	9	10
1	1.00									
2	0.69	1.00								
3	0.84	0.62	1.00							
4	0.71	0.84	0.59	1.00						
5	0.53	0.68	0.50	0.62	1.00					
6	0.36	0.55	0.29	0.61	0.62	1.00				
7	0.61	0.40	0.33	0.33	0.35	0.64	1.00			
8	0.47	0.30	0.24	0.36	0.43	0.59	0.81	1.00		
9	0.54	0.48	0.47	0.60	0.59	0.63	0.67	0.67	1.00	
10	0.62	0.54	0.42	0.52	0.58	0.75	0.67	0.69	0.83	1.00

Table 17. Median threshold scenario 5. Median is 0.63

	1	2	3	4	5	6	7	8	9	10
1	1.00									
2	0.73	1.00								
3	0.88	0.55	1.00							
4	0.78	0.86	0.51	1.00						
5	0.80	0.79	0.63	0.83	1.00					
6	0.53	0.49	0.51	0.50	0.51	1.00				
7	0.59	0.42	0.42	0.47	0.54	0.88	1.00			
8	0.66	0.43	0.42	0.48	0.48	0.81	0.53	1.00		
9	0.59	0.51	0.42	0.54	0.62	0.91	0.77	0.60	1.00	
10	0.62	0.59	0.42	0.62	0.67	0.77	0.70	0.66	0.75	1.00

Table 18. Median threshold scenario 6. Median is 0.51

	1	2	3	4	5	6	7	8	9	10
1	1.00									
2	0.63	1.00								
3	0.81	0.51	1.00							
4	0.57	0.53	0.63	1.00						
5	0.65	0.34	0.79	0.50	1.00					
6	0.00	0.00	0.00	0.00	0.00	1.00				
7	0.00	0.00	0.00	0.00	0.00	0.65	1.00			
8	0.00	0.04	0.00	0.00	0.00	0.71	0.72	1.00		
9	0.00	0.00	0.00	0.00	0.00	0.82	0.83	0.83	1.00	
10	0.00	0.00	0.00	0.00	0.00	0.84	0.79	0.82	0.91	1.00

Table 19. Median threshold scenario 7. Median is 0.14

	1	2	3	4	5	6	7	8	9	10
1	1.00									
2	0.46	1.00								
3	0.62	0.60	1.00							
4	0.54	0.72	0.78	1.00						
5	0.76	0.59	0.78	0.74	1.00					
6	0.00	0.02	0.02	0.02	0.00	1.00				
7	0.00	0.11	0.13	0.13	0.14	0.12	1.00			
8	0.00	0.01	0.00	0.00	0.00	0.68	0.12	1.00		
9	0.00	0.00	0.00	0.00	0.00	0.84	0.00	0.77	1.00	
10	0.00	0.03	0.05	0.05	0.04	0.77	0.24	0.72	0.83	1.00

Table 20. Median threshold scenario 8. Median is 0.54

	1	2	3	4	5	6	7	8	9	10
1	1.00									
2	0.89	1.00								
3	0.54	0.54	1.00							
4	0.76	0.77	0.65	1.00						
5	0.76	0.77	0.64	0.76	1.00					
6	0.00	0.00	0.04	0.00	0.00	1.00				
7	0.00	0.00	0.07	0.00	0.00	0.58	1.00			
8	0.00	0.00	0.05	0.00	0.00	0.62	0.60	1.00		
9	0.00	0.00	0.00	0.00	0.00	0.49	0.62	0.62	1.00	
10	0.00	0.00	0.08	0.00	0.00	0.67	0.65	0.64	0.54	1.00

Additionally, figure 19 and 20 show the graphs of the percentage of correctly classified phones for both average and median thresholds. Tables from 5 to 20 are compared to Table 4 (the ideal scenario, where all classifications are correct). It can be shown that the classification accuracy increases when the standard deviation is higher, which can be interpreted as the similarity values being farther apart from each other, which is desirable in order to correctly distinguish the groups.

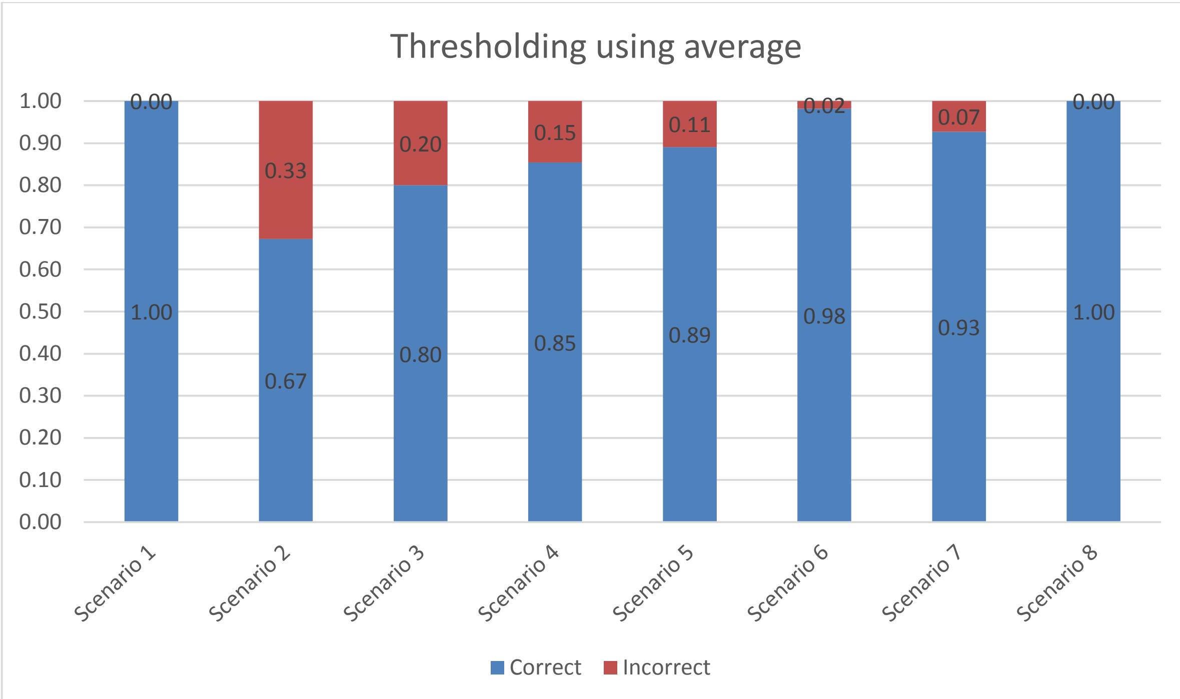

Fig 19. Accuracy of classifying phones into groups using the average as threshold.

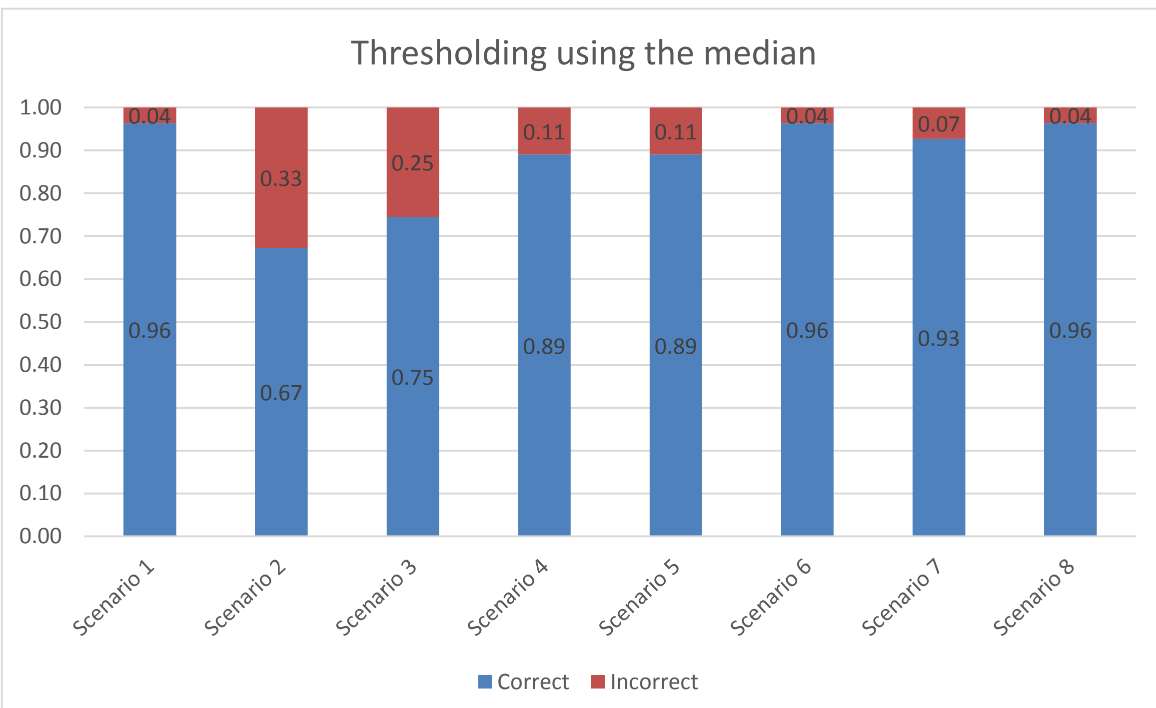

Fig 20. Accuracy of classifying phones into groups using the median as threshold.

As it can be observed from the values of tables from 5 to 12, scenarios in which the standard deviation is high, the accuracy of grouping by either average or median threshold tends to rise. This is due to the data being more dispersed, therefore it is easier to distinguish either group. However, in scenarios 4 and 5, for which phone number 10 was the idle person standing in the crossing points of the flocks, was not possible to distinguish the phone from either group. It was concluded that Wi-Fi is not coarse enough to do so, at least by only calculating the feature proposed in this work.

In the other hand, for phones in groups where the initial and final position distances of their trajectories was far apart, more specifically, scenario 1 and scenarios from 6 to 8, seem to be easily classified into groups by the sole use of threshold with the average. It is possible to see from figure 19 and 20 that the average is a better threshold than the median, as the median, in case of very polarized data (similarity values which are either very close to 1 or 0) can set the threshold too high or too low, leaving out measurements which in the case of the average threshold would be considered as in either the same group or the opposite case.

5.1.1. Contrasting Wi-Fi with Barometer

According to the barometer data shown in figure 21, in order to correctly detect groups of phones by analyzing exclusively the raw data, calibration of the mobile phones is required. Despite having exclusively Samsung Galaxy sIII phones for the Wi-Fi and barometer, the difference in the phone models was enough to log different barometer readings. Requiring calibration however is relative, as it could be seen in a later experiment performed, by using the standard deviation was ir could be possible to observe only the changes on the sensor value rather than the absolute value, rendering calibration unnecessary for this particular application.

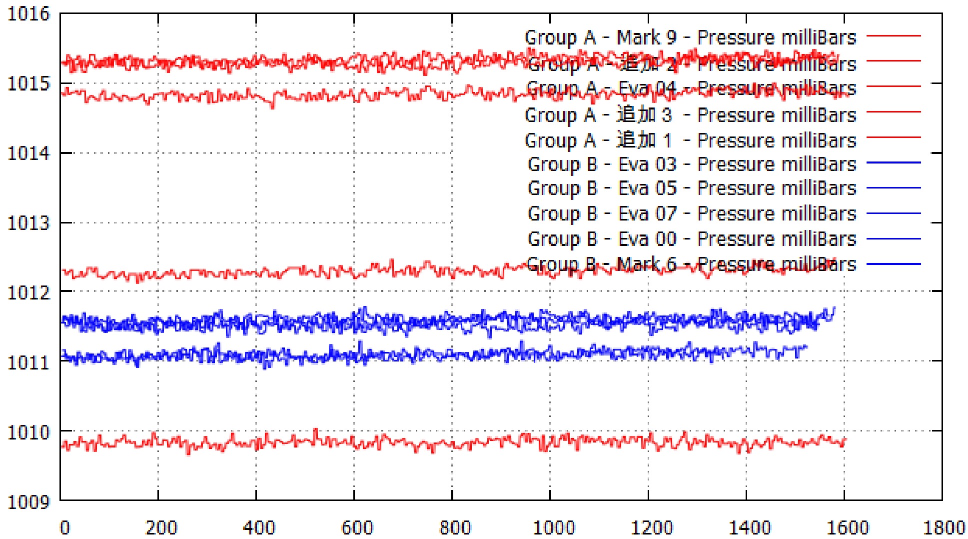

Fig 21. Barometer data for scenario 6.

5.2.Similarity results for the sensor modalities

The cosine similarity and maximum cross correlation values are shown in this section for the features proposed in subsection 4.2 “Sensor modalities features”. This chapter is divided into each sensor modality used in the ground truth collection experiment, so the reader can direct its attention to whichever sensor modality is required.

5.2.1. Accelerometer

Sliding window functions were applied to the accelerometer. The accelerometer data, as it was mentioned earlier, required the use of a windowing function because of the cyclic nature of walking. The overlap is necessary as there is no certainty of when the steps given by the pedestrian will occur. The first features that was derived was the average and standard deviation after applying the trimming and grouping functions, with group sizes of 10 and 50 measurements.

There were two experiments done to observe the accelerometer data. In the first experiment, there were two scenarios. The “moving scenario” in which one person carrying a smartphone moved around the Sezaki laboratory in Komaba II campus building and the other phone was left unmoved on a desk. Then the “still scenario” was done to compare two phones left on a desk unmoved.

The second experiment was design as closely as possible to a real world scenario. Two smartphones were carried around the Komaba II campus building by one person, two simulate a pedestrian flock of two persons; while one smartphone was left in the Sezaki laboratory unmoved. The person moving was meant to walk around the campus building, going up and down stairs to gather distinctive activities from the perspective of the accelerometer.

Both experiments correspond to experiments 1, 2 and 5 from Table 3 in the section “2,3 Sensor ground truth data collection”. The results of the data comparison using the cosine similarity and the maximum value of the cross correlation function for experiment 1&2 and 5 are shown in Table 21 and Table 22, respectively.

Table 21. Results for accelerometer experiment 1&2.

Moving Sce.	**Cosine Similarity**			
	Average for partition size of		***Standard Deviation for partition size of***	
	2 seconds	***10 seconds***	***2 seconds***	***10 seconds***
X Axis	-0.1359	-0.1963	0.5110	0.5906
Y Axis	0.7411	0.7540	0.5783	0.7060
Z Axis	0.7416	0.7546	0.5477	0.7146
Still Sce.				
X Axis	-0.7778	-0.8185	0.5364	0.6283
Y Axis	0.7063	0.7176	0.6135	0.6730
Z Axis	0.7036	0.7138	0.5888	0.6631

Table 22. Results for comparison of the first five coefficients of the fast Fourier transform of the accelerometer data. Experiment 5.

Windows size 10	**Max. Normalized Cross Correlation of 5 coefficients of the Fast Fourier Transform**		
	Mov1 vs. Mov2	***Mov1 vs Still***	***Mov2 and Still***
X Axis	0.9038	0.9490	0.8813
Y Axis	0.9445	0.9582	0.9555
Z Axis	0.9985	0.9984	0.9990
W. Size 50			
X Axis	0.9112	0.9417	0.8843
Y Axis	0.9766	0.9676	0.9901
Z Axis	0.9988	0.9985	0.9995

Also, according to [11], the best performing feature for comparison between accelerometer signals is comparing the magnitude of the first 5 coefficients of the Fourier transform. This feature was derived for a window size of 10 and 50 measurements (or equivalently 2 and 10 seconds, respectively) with a 50% overlap. The results are shown in Table 22. As it can be observed, even the best performing feature is not usable to compare the accelerometer raw data between smartphones.

The analysis of the accelerometer data was originally meant by the authors in [9] to be used to classify the data and then compare the classifications, however they had the same difficulties that can be observed from the results in Table 22, in terms of the accelerometer data not being representative enough of the highly distinctive activities it is meant to describe. the location of the phone respect the body or any other well-known techniques to improve the clarity of the data was not taken into consideration in this thesis, thus arriving to the same conclusion stated in [9], which were that by utilizing the raw data to calculate the features directly is not descriptive enough for classification or, furthermore, direct comparison.

5.2.2. Barometer

The purpose of the windowing function for the barometer data, despite it being quite regular, as shown in Figure 21, was to mask big variations on the data. When the person carrying the phone goes to a different floor; the values of the barometer signal change abruptly and are maintained relatively constant as long as the person stays on the same floor, therefore keeping the sharp step function-like shape of the barometer signal is desirable. The maximum value of the normalized cross correlation was calculated for the raw data and the windowed average and standard deviations values. The barometer data was gathered in the sensor experiment 5.

It is possible to distinguish from the data from the barometer the phones that were on different floors. These results were expected as it was shown in [18]. This result is relevant because it allows the pedestrian flock detection application switching from using Wi-Fi to barometer decreasing the energy consumption. This would be the case for scenarios such as scenario 6 from the Wi-Fi ground truth data collection experiments. This data has already been reported; the barometer data plot is shown in Figure 21, for scenario 6 of the Wi-Fi experiments. The results for the sensor experiment 5 are shown in Table 23.

Table 23. Results for barometer data in experiment 5

Windows size = 10	**Max. Normalized Cross Correlation**		
	Mov1 vs. Mov2	***Mov1 vs Still***	***Mov2 and Still***
Stand dev	0.9210	0.4932	0.5052
Average	0.9918	0.5495	0.5450
W. Size = 50			
Stand dev	0.9756	0.5112	0.5242
Average	1.0000	0.5362	0.5362
Raw data			
	0.5923	0.3032	0.0769

5.2.3. Gyroscope

No windowing function was applied as these features are used to capture turning events. Unlike walking, turning events lack periodicity. The maximum value of the normalized cross correlation function and cosine similarity was therefore calculated over the raw data and the magnitudes.

Similarly to accelerometer, three experiments were conducted to observe the gyroscope performance for detecting groups. The first two experiments, experiments 3 and 4 from Table 3, were design to gather data of pedestrian turning to the same direction and opposite direction respectively. These experiments were meant to observe the gyroscope data in absolutely opposite situation. The result for these experiments are shown in Table 24.

Experiment 5 is the experiment closest to a real world scenario, and has data from two moving phones and one still phone. The results for this experiment are shown in Table 25.

The results from Table 24 might seem inconclusive, as the similarities have very low and irregular values, however, for experiment 5, the similarity of the magnitude of the gyroscope data exhibits a 20% gap between phones in different groups, as shown in the results in Table 25. This result is good because by just using the median or the average of the data as a threshold it is possible to easily distinguish the phones in different groups.

Table 24. Results for gyroscope experiments 3 and 4.

Same turn	**Cosine Similarity**			
	Average for partition size of		***Standard Deviation for partition size of***	
	2 seconds	***10 seconds***	***2 seconds***	***10 seconds***
X Axis	0.0062	0.0822	0.0686	0.1006
Y Axis	-0.0121	0.0783	0.0005	0.1253
Z Axis	0.0016	0.0685	0.0194	0.0943
Opp. turn				
X Axis	-0.0173	0.1173	-0.1120	0.2724
Y Axis	-0.0234	0.1699	-0.0437	0.2592
Z Axis	-0.0213	0.1077	-0.0157	0.1069

Table 25. Results for Gyroscope experiment 5. Cross correlation per axis and magnitude

	Max. Normalized Cross Correlation		
	Mov1 vs. Mov2	***Mov1 vs Still***	***Mov2 and Still***
X Axis	0.1806	0.1290	0.3139
Y Axis	0.2889	0.1877	0.3151
Z Axis	0.4382	0.1868	0.5162
Magnitude	0.4201	0.2675	0.2190

5.2.4. Magnetometer

No windowing function was applied as this sensor is used to capture turning events and magnetic fingerprints of the building. Unlike walking, this data lacks periodicity. The maximum value of the normalized cross correlation function was therefore calculated over the raw data per axis and the magnitudes for both the experiment 3 and 4 from Table 3. Results for both experiments are shown in Tables 26 and 27, respectively.

The magnetometer data based features display a good performance when differentiating groups. Some problems were encountered while retrieving the samples however; some of the samples from previous experiments were mixed in, despite this the conclusion is the same, the features derived from the magnetometer are robust because it is possible to distinguish the groups despite the difficulties retrieving the data.

Table 26. Results for magnetometer experiment 3 and 4

Max. Normalized Cross Correlation		
	Same turn	***Opposite turn***
X Axis	0.434770537	0.074337106
Y Axis	0.428146936	0.084175992
Z Axis	0.586855569	0.033560895
Magnitude	0.270406646	0.082014476

Table 27. Results for the magnetometer experiment 5. Cross correlation per axis and magnitude

	Max. Normalized Cross Correlation		
	Mov1 vs. Mov2	***Mov1 vs Still***	***Mov2 and Still***
X Axis	0.5716	0.4415	0.2071
Y Axis	0.8058	0.8592	0.4152
Z Axis	0.3216	0.8019	0.2289
Magnitude	0.7603	0.5678	0.5064

Both the raw data and the magnitude of the tri-axial magnetometer are usable. The difference between groups is clearer when comparing axis by axis instead of the magnitude.

Figure 22 has the magnetometer raw data plotted to support the argument in favor of using this data as room fingerprints. The saturated part of the graph corresponds to the magnetic fingerprint of the Sezaki laboratory. The initial data that can be seen, which is quite unstable is due to readings from previous experiments, as it was just mentioned, there were problems retrieving the data from the phones. The person who was to remain in the laboratory moved around inside the laboratory but then settled in a desk, showing that even inside the laboratory there are different fingerprints. Further work should be done in this direction in order to determine how granular magnetometer data can be.

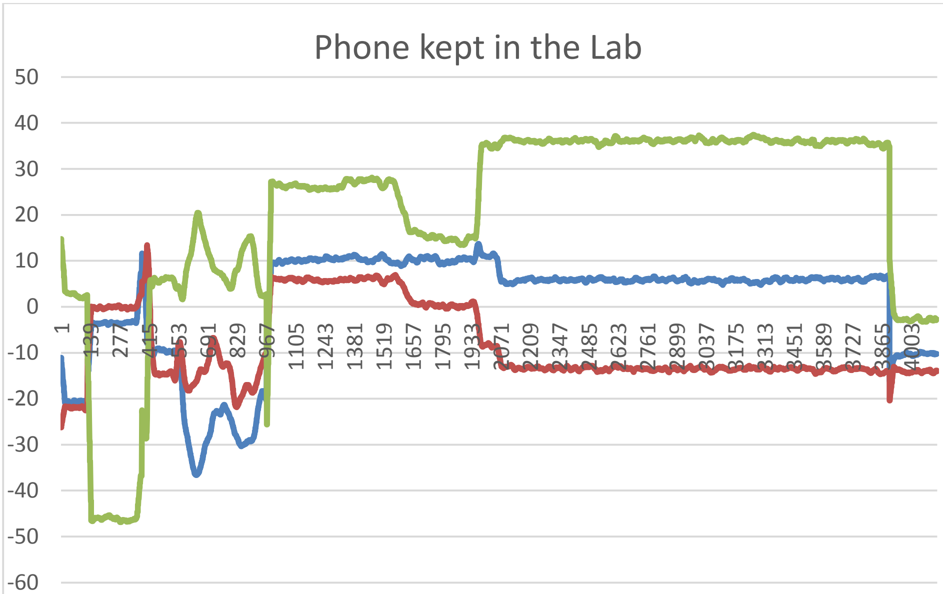

Fig 22. Tri-axial magnetometer data for experiment 4 for gathering sensor data.

5.3. Discussion

The Wi-Fi data has a good performance as long as the initial position of both groups is distant from each other. It was also possible to see that for the case of the flocks being located on different floors, the use of barometer could replace the use of Wi-Fi (for a two group scenario, however if there were 3 groups and two of them were on the same floor, they would be undistinguishable from solely the barometer data perspective). Switching from a power hungry sensor such as Wi-Fi over to barometer is an attractive option. The concept application designed in [15] could switch from Wi-Fi to barometer is the same amount of flocks is detected with both methods. There is no real implementation of this, at this point is only possible to theorize about it.

It is important to remark that the performance for distinguishing groups dramatically increased as the standard deviation of the samples was higher. This can be interpreted as the data being more dispersed and therefore more distinct. Having the samples significantly different to the average helps classification.

It is possible to conclude that the usage of the accelerometer data can be discarded for this application. In this experiment, the participants held the devices in front of their faces, gathering very similar measurements despite one of the participants performed different types of movement compared to the other (going up and down stairs). Without considering the location of the phone respect the pedestrian's body, the information that can be obtained, using the best performing feature, namely the magnitude of the first five coefficients of the Fast Fourier Transform, is useless for using classification of activities for pedestrian flock detection.

However, observing the inertial sensors is still necessary, as the gyroscope has a 20% similarity gap between phones in different groups. This gap was negatively impacted by the problems experienced when recovering the samples from the application used for gathering the sensor ground truth data. These difficulties recovering the data were tied to a single specific smartphones, as the same was experienced with the sensor data of the same smartphone.

In the sensor experiment 5, the phone kept still in the laboratory had remarkably less samples than the other two which were in motion, despite this the cross correlation of the magnitude is still able to distinguish correctly between groups of phones. Therefore the best performing features are the magnetometer raw data and magnitude and the gyroscope magnitude and standard deviation of the raw data.

6. Discussion

This thesis seeks to further the addition of modalities to the feature space in the application of pedestrian flock detection in a crowd-sensing application. Previous approaches to solve the pedestrian flock detection problem have shown advances both in outdoors and indoors scenarios as shown in [9], [13] and [14].

In the case of the outdoor scenarios only GPS (Global Positioning System) was used in [13]. As it was aforementioned, the GPS samples were subject to a 2 stage clustering process; spatial clustering prior to temporal clustering.

For indoor scenarios, features such as position and direction of movement derived from Wi-Fi as well as direct comparison of power measurements, were utilized in [14]. The amount of features derived from Wi-Fi only data was possible due to the prior fingerprinting of the area of the application deployment, requiring a database visible to all the smartphones.

The authors extended their work in [14] by adding accelerometer and magnetometer features to the already existing Wi-Fi approach in [9]. The accelerometer features were calculated by using a window function and compared using cross correlation. Additionally, the authors used the variance of the magnitude to determine whether the person carrying the phone was moving (originally wanting to classify their movements and compare the classification, but they concluded that the accelerometer data from a smartphone is not distinctive enough to classify movement using solely the technique they proposed).

The features they derived from the compass was the windowed maximum cross correlation and the time since the last detected turn. The magnetic compass, however, is subject to noise according to the metal present in the environment where the data is being taken.

Furthermore, results are sensitive to changes in the infrastructural Wi-Fi network of the place of the experiment, as the Wi-Fi derived features are calculated using fingerprints previously stored in a server, instead of comparing them directly to each other as it was shown in the method presented in this thesis. Not having to compare the Wi-Fi fingerprints with previously stored fingerprints in a server decentralizes the method making it distributable. This has the advantage of reducing the maintenance costs, by not having to update the database if there is the addition or removal of an access point, and avoiding any delays incurred by the network latency due to the communications between the smartphones and the database.

Once the features per sensor are derived, the clustering process starts by each feature space, when the flock candidates are derived, then according to weights defined by the authors in [13], a majority voting is performed over the flock candidates yielding the groups which will then be temporally clustered, similarly to the work done in [13]. A schematic view of the majority voting can be seen in Figure 23.

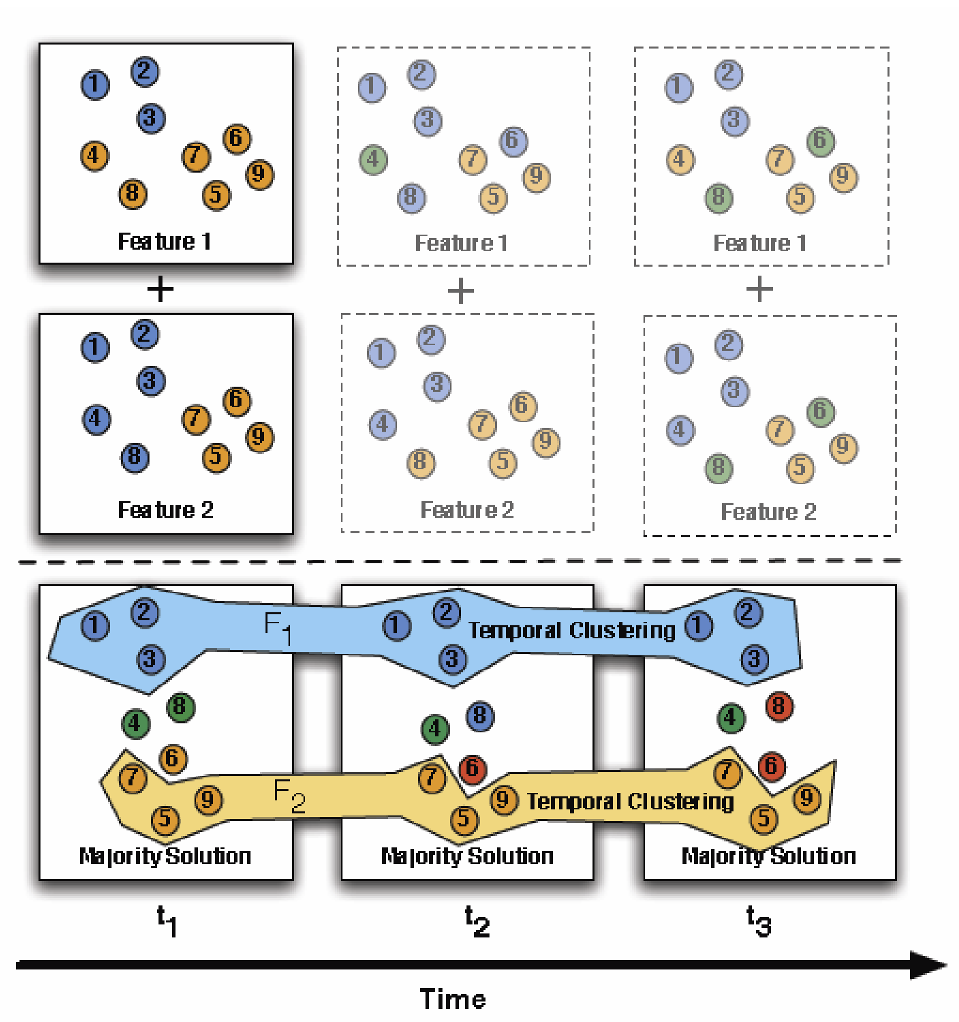

Fig 23. Schematic visualization of the flock detection algorithm using weighted majority voting for three time steps and temporal clustering which identifies two flocks F1 and F2.[13]

This thesis proposal is that by increasing the variety of sensor modalities in the feature space of the pedestrian flock application, it is possible to increase the accuracy of the application, as some sensors are more appropriately used in certain circumstances than others. For instance, the case of a scenario where the flocks are located in different floors; as it was demonstrated, this situation can be identified both with the use of Wi-Fi and barometer. The amount of energy consumed by each modality must be taken into consideration, however it is of great interest to, at least at first, expand the feature space as much as possible. Therefore, advances were investigated in grouping pedestrian flocks using solely sound [16], which broadens the feature space improving the accuracy and a variety of scenarios of the application. An alternative approach to using sound was proposed in [20] where the silences in sounds samples were compared to determine whether a couple of devices were collocated or not. Despite all the available options to broaden the feature space it is also necessary to determine more energy efficient solutions as smartphones are very constrained in terms of resources, as it is also discussed in [19].

The other possible approach to improve the accuracy detection of flocks is to, instead of expanding the feature space, improve its quality. Techniques used for identifying the modes of transportation [21], using solely the accelerometer data. As it was commented in chapter 2, in the case where the data gathered by people riding a train and people are just walking around the commercial area of a train station, the benefit of distinguishing these activities would be a great advantage. It would even render the conclusion, about the accelerometer not being usable drawn in this thesis and in [9], refuted. As it was suggested in [9], these classification of modes of transport could be directly compared.

7. Conclusion

Sparse values respect to the average for the Wi-Fi data were observed in scenarios were the classification rate was high. This suggests that for successfully detecting the groups, the cosine similarity over the grouped average was distinctive enough. Additionally, for certain scenario in which there are two groups in different floors, it is possible to switch to barometer usage instead, saving battery life. The density of Wi-Fi access points is also an important factor to consider while using this sort of methods.

For the sensor data, the accelerometer specifically was once again proven to be not suitable for pedestrian flock detection, unless there is a prior step of additional processing. In the case of the barometer and magnetometer data, the comparison of the raw data directly is enough to differentiate the groups from the experiments. The case of the gyroscope, it was necessary to calculate the magnitude of the data prior to comparison. Sensors are by far less power hungry than Wi-Fi, according to the phone manufacturer, so finding a subset of sensors from the ones studied is desirable.

As it was shown with the case of the barometer and Wi-Fi data, the usage of either sensors, for that experimental setting, yields seamless results. Therefore, it can be concluded that there might be a best suitable subset of sensors for each situation. The application to be developed by Teemu Leppänen will have to propose a solution on how to identify the situation and then utilize the most appropriate subset of sensors accordingly.

The experiments settings are limited however, situations which have only 2 groups are far from reality. This is only a first step to evaluate the appropriateness of the features proposed in this thesis. Larger experiment settings will have to be considered to develop an “all situation fitting” solution.

7.1. Summary

There were two set of experiments to gather the data from the ground truth. In almost all the experiments, the total amount of persons were divided into two equally sized groups. This is because distinguishing two groups of people is the fundamental case. The purpose of the first set of experiments was to gather the ground truth data for the Wi-Fi and then compare. This experiment had 8 scenarios, each one was designed posing a different challenge for pedestrian flock detection. A summary of the purpose and description of each scenario for the Wi-Fi ground truth data collection is presented in the following table.

Table 28. Summary of the scenarios for the Wi-Fi ground truth data collection experiment.

Index	Floor	size	Description/Axes involved	Description summary
1	6th	5,5	Crossing each other	The initial position of each flock is the end position of the other one. The purpose of this experiment is to observe the similarity between pairs of phones, which are grouped into significantly set apart groups, both at the beginning and ending of the experiment
2	6th	5,5	Parallel trajectory	Both flocks have the same initial position and direction of movement, but one flock starts moving before the second one. The purpose of this scenario is to stress the proposed method by seeing if the two flocks are differentiated despite the identical trajectories and positions.
3	6th	5,5	Perpendicular crossing	Perpendicular trajectory in the 6th floor of the Komaba II campus building. The initial position of each flock is closer than the one of scenario 1; maintaining the same point of crossing than the one in scenario 1. The purpose of this scenario is to set flocks in different yet close initial points.
4	6th	5,4,1	2 flocks crossing. Idle person	Groups with opposite directions with an idle person standing in the crossing point of the 2 groups. The initial position of each flock is the end position of the other one. The purpose of this scenario is to observe to which group is the person standing idly most similar.
5	6th	5,4,1	2 flocks perpendicular crossing.	Groups walk with perpendicular directions with an idle person standing in the crossing point of the 2 groups. Similarly to the previous scenarios, the purpose is to set the flocks to closer initial points and observer the idle person.
6	6th & 4th	5,5	Parallel trajectory	Groups with parallel trajectories in the 6th and 4th floor of the Komaba II campus building. Both flocks have the same initial position but in different floor and same direction of movement, but one flock starts moving before the second one. The purpose of this scenario is to observe whether Wi-Fi data is different enough in the case the flocks are located very close vertically
7	6th & 4th	5,5	Crossing each other	Groups cross each other in the 6th and 4th floor of the Komaba II campus building. The initial position of each flock is the end position of the other one, but in different floors. The purpose of this scenario is to observe what was considered as the best case scenario in terms of dissimilarity; here the groups of people are as far as possible in their initial and final position.
8	6th & 4th	5,5	Perpendicular crossing	Groups cross each other in a perpendicular trajectory in the 6th and 4th floor of the Komaba campus building. Both flock's initial and final position is closer than in scenario 7. This scenario was believed to be meaningful because it would allow to contrast the results with the ones obtained in scenario 3 and visualize the impact in the dissimilarity of the groups in different floors.

The second set of experiments was the ground truth collection for the sensor data; which had several experiments because single sensors were tested in specific scenarios. Similarly to the case of scenarios 3 and 4 from the sensor experiments shown in Table 29, special emphasis was put on gathering particularly distinctive sensor data.

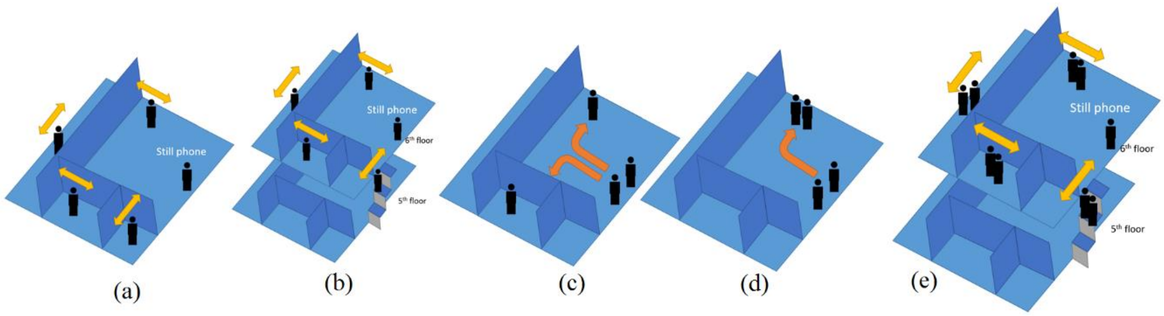

Fig 24. Schematics of all the experiments for the sensor ground truth data collection experiments. (a) Experiment 1, (b) experiment 2, (c) experiment 3, (d) experiment 4 and (e) experiment 5.

As Table 29 shows, the experiment 1 and 2 shown in the schematic in Figure 24,a and 24,b had the purpose of gathering data to compare accelerometer data. The following two experiments, 3 and 4 (figure 24,c and 24,d) had for objective, as it is reiterated in Table 29, to gather magnetometer and gyroscope data which is distinctive enough to separate the groups. Finally experiment 5 presents a real world scenario, also shown in figure 24,e.

Table 29. Summary of the sensor ground truth data collection experiments.

Name of the Experiment	**Main Sensors observed**	**Time span**	**Description summary**
Experiment 1.	Accelerometer	2 minutes	Two phones were kept on a desk in order to visualize the similarity of two static phones and then one was moved around the laboratory. Figure 24,a
Experiment 2.	Accelerometer	5 minutes	Two phones were used. One phone was kept on a desk and the other was taken around the Komaba II campus building. The purpose was to contrast data from movement. Figure 24,b
Experiment 3.	Gyroscope	2	Two persons walked and gave a single turn in opposite directions

Different turn	and Magnetometer	minutes	having as end point of their trajectory different rooms. See figure 24,c
Experiment 4. Same turn			Two persons walked and gave a single turn in the same direction having as end point of their trajectory the same room. See figure 24,d
Experiment 5.	Acc., Gyro., Mag. And Bar.	5 minutes	3 phones were used. Two phones were carried together to simulate a group and one was left on a desk. See figure 24,e

Over the data gathered a series of functions were utilized to extract features from the data. For the Wi-Fi data and sensor data a sliding window function and a grouping function (which, in other words, is a sliding window function with 0% overlap between consecutive windows) were utilized. For the Wi-Fi data, the power measurements were analyzed in groups of 40 measurements and the average was calculated for power measures with repeated MAC addresses (after filtering measurements whose power was below -70 dBm). Then the cosine similarity was applied to the averaged groups and finally all the groups' cosine similarity values were averaged. This yields a similarity measurement for each pair of phones.

For the sensor data, as it was aforesaid, the grouping and sliding window functions were used. The features extracted from the data were the average, standard deviation and magnitude in the case of tri-axial sensors. For the accelerometer, the similarity between the magnitudes of the first 5 coefficients of the Fast Fourier Transform was additionally calculated, as its performance is superior to features derived in the time domain, according to [11].

The two methods for sensor and Wi-Fi data derived features comparison were the cosine similarity and the cross correlation function. The cosine similarity consists of the dot product of two vectors of measurements divided by the product of the magnitude of said vectors (in the case of Wi-Fi the vectors were constituted by the power measurements, each one identified by its associated MAC address). The cross correlation function was normalized and only the maximal value was taken into account, which when the function is normalized, represents the maximum similarity for a given lag between the functions or vectors being compared. The details on which specific means of comparison and which features were derived from which sensor has been put in Table 30.

Two categories of sensors have been devised: inertial sensors, such as the accelerometer and gyroscope, and environmental sensors, such as the barometer and magnetometer. The former type of sensor describes the movement of the person carrying the smartphones, whereas the latter describes the surroundings.

Table 30. Feature and methods applied to the Wi-Fi and sensor data gathered in the experiments.

Feature and methods	**Sensors**
Cosine similarity	Wi-Fi, Gyroscope
Cross Correlation	Accelerometer, Barometer, Gyro, Magnetometer
Grouping	Wi-Fi, Gyroscope
Sliding Window	Accelerometer, Barometer
Average and SD	Accelerometer, Barometer, Magnetometer
Magnitude	Accelerometer, Gyroscope, Magnetometer
FFT first 5 coefficient magnitude	Accelerometer

The qualitative results are that by comparing the similarity per axis for the gyroscope data, the two groups in this experiment setting can be distinguished. The results are also good for the magnetometer data; for which the similarity was calculated for each axis and the magnitude. The barometer, as it was demonstrated in [18], can be used only in scenarios in which the flocks are located in different floors. Finally, the accelerometer was disproved to be useful for experiment setting. Despite the features calculated for the data, the similarity values remained almost the same for phones in different groups.

The Wi-Fi data demonstrated to be useable in most situations, except in those where the flocks had very close initial positions and walked very close to each other (about 5 meters). This can be due to the access points in the Komaba campus being too scarce for this purpose, or the sampling rate of the smartphones being too slow.

The details of the results can be found in the “Results” section, where a quantitative analysis of the similarity values obtained for each experiment has been given.

7.2. Suggested Future Work

The results in this thesis are directly usable in the application of pedestrian flock detection. Furthermore, this work enables the future collaborative and distributed application for assisting human movements in urban areas [13], using smartphones. The benefits of this method include adding other sensor modalities, such as magnetometer, enabling flock detection in areas where other modalities, such as Wi-Fi, may not be available. Also, this work makes possible the real-time comparison of Wi-Fi signal strength and other sensor data vectors without any previously stored information. However, without this information, the capability of localizing the detected flocks is lost.

The barometer data was simultaneously gathered to see how effectively it was possible to reach to the same detected groups using different sensor modalities. The energy consumed by the barometer sensor is considerably less than the one used by Wi-Fi, based on readings provided by the smartphone manufacturer, therefore switching to barometer without negatively impacting the detection performance, is desirable. Of course, this is smartphone model-dependent. To further this work, it is also necessary to rank the sensor modalities in terms of energy-consumption and environmental conditions, as described in [15].

Future work will have to consider larger real-world scenarios, both in terms of the amount of persons and groups observed. Only the fundamental scenarios were dealt with in this thesis, in which distinct features from the sensor data were considered. The addition of larger groups poses yet another challenge, as the quality of available sensor data is unknown for accurate detection. This poses yet another difficulty, as sensor data from different phone manufacturers is bound to be different. Future work will have to take this into consideration and then develop a way of auto-calibration for a diversity of smartphone models.

Publication List

[1] Alvarez, J.; Teemu, L.; Masayuki, I.; Kobayashi, H.; Sezaki, K.,"A method for grouping smartphones based on Wi-Fi signal strength," Forum of Information Technologies 2013. FIT 2013, vol. 3, pp. 449-452, September 2013.

[2] Leppänen, T.; Alvarez , J.; Iwai, M.; Sezaki, K.; Tobe, Y.; Riekki, J., "A distributed system architecture for pedestrian flock detection with participatory sensing,"In Proceedings of the 37th IPSJ UBI Reasearch Conference, vol. 2013-UBI-37, no. 40, pp. 1-7, January 2013.

[3] Leppänen, T.; Alvarez , J.; Ramalingam, A.; Lui, M.; Harjula, E.; Närhi, P.; Ylioja, J.; Riekki, J.; Sezaki, K.; Tobe, Y., "Interoperable mobile agents in heterogeneous wireless sensor networks", In Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems, pp. 64, 2013

[4] Alvarez, J.; Leppänen, T.; Sezaki, K., "A Study on Features Derived from Smartphones' Wi-Fi and Sensor Data for Pedestrian Flocks Detection", 5th ACM Augmented Human Conference, March 2014 (Under review)

References

[1] Kang, J. H.; Welbourne, W.; Stewart, B.; Borriello, “Extracting places from traces of locations,” SIGMOBILE Mob. Comput.Commun. Rev. 9, pp. 58-68, March 2005.

[2] C. Zhou, D. Frankowski, P. Ludford, S. Shekhar, and L. Terveen. “Discovering personal gazetteers: an interactive clustering approach,” in proceedings of the 12th annual ACM international workshop on Geographic information systems 2004. GIS 2004, pp. 266-273, November 2004.

[3] Lukowicz, P.; Pentland, A.; Ferscha, A., "From Context Awareness to Socially Aware Computing," Pervasive Computing, IEEE , vol.11, no.1, pp.32-41, January-March 2012.

[4] Dodge, S.; Weibel, R.; Lautenschutz, A.-K., “Towards a taxonomy of movement patterns,” Information Visualization, vol. 7, no. 3-4, pp.240–252, September 2008.

[5] Schadschneider, A.; Klingsch, W.; Klupfel, H.; Kretz, T.; Rogsch, C.; Seyfried, A., Evacuation Dynamics: Empirical Results, Modeling and Applications, in: Meyers, R. A., Extreme Environmental Events.Springer New York, 2009, pp 517-550.

[6] Wirz, M.; Roggen, D.; Troster, G., "User Acceptance Study of a Mobile System for Assistance during Emergency Situations at Large-Scale Events," Human-Centric Computing (HumanCom), 2010 3rd International Conference on , vol., no., pp.1-6, 11-13 Aug. 2010.

[7] Rachuri, K. K.; Musolesi, M.; Mascolo, C.; Rentfrow, P. J.; Longworth, C.; Aucinas, A., “Emotionsense: a mobile phone based adaptive platform for experimental social psychology research,” In proceedings of the 12th ACM International Conference on Ubiquitous Computing 2010. Ubicomp 2010, pp. 281-290, September 2010

[8] Kaemarungsi, K.; Krishnamurthy, P., "Properties of indoor received signal strength for WLAN location fingerprinting," Mobile and Ubiquitous Systems: Networking and Services, 2004. MOBIQUITOUS 2004, pp.14,23, 22-26 August 2004.

[9] Kjaergaard, M. B.; Wirz, M.; Roggen, D.; Tröster, G., “Detecting pedestriand flocks by fusion of multi-modal sensors in mobile phones,” Proceedings of the 2012 ACM Conference on Ubiquitous Computing . UbiComp 2012, pp. 240-249, September 2012.

[10] Alvarez, J.; Teemu, L.; Masayuki, I.; Kobayashi, H.; Sezaki, K.,“A method for grouping smartphones based on Wi-Fi signal strength,” Forum of Information Technologies 2013. FIT 2013, vol. 3, pp. 449-452, September 2013.

[11] Wu, W.; Dasgupta, S.; Ramirez, E.; Peterson, C.; Norman, G. J.,"Classification accuracies of physical activities using smartphone motion sensors," J Med Internet Res, vol. 14, no. 5, pp. e130, October 2012.

[12] Wang, H.; Sen, S.; Elgohary, A.; Farid, M.; Youssef, M.; Choudhury, R. R., "No need to war-drive: unsupervised indoor localization," In Proceedings of the 10th ACM international conference on Mobile systems, applications, and services 2012. MobiSys 2012, pp. 197-210, June 2012

[13] Wirz, M.; Kjaergaard, M. B.; Feese, S.; Schlapfer, P.; Roggen, D.; Tröster, G., "Towards an online detection of pedestrian flocks in urban canyons by smoothed spation-temporal clustering of GPS trajectories," In Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Location-Based Social Networks, pp 17-24, November 2011.

[14] Kjaergaard, M.B.; Wirz, M.; Roggen, D.; Troster, G., "Mobile sensing of pedestrian flocks in indoor environments using Wi-Fi signals," Pervasive Computing and Communications (PerCom), 2012 IEEE International Conference on , vol., no., pp.95-102, 19-23 March 2012.

[15] Leppänen, T.; Alvarez , J.; Iwai, M.; Sezaki, K.; Tobe, Y.; Riekki, J., "A distributed system architecture for pedestrian flock detection with participatory sensing,"In Proceedings of the 37th IPSJ UBI Reasearch Conference, vol. 2013-UBI-37, no. 40, pp. 1-7, January 2013.

[16] Satoh, H.; Suzuki, M.; Yamashita, Y.; Lee, D.; Morikawa, H., "Ambient sound-based proximity detection with smartphone," IEICE Society Conference, pp. 439, September 2013. (In Japanese)

[17] Leppänen, T.; Alvarez , J.; Ramalingam, A.; Lui, M.; Harjula, E.; Närhi, P.; Ylioja, J.; Riekki, J.; Sezaki, K.; Tobe, Y., "Interoperable mobile agents in heterogeneous wireless sensor networks", In Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems, pp. 64, 2013.

[18] Stiefmeier, T.; Tröster, G., "Detection of Walking Activities for Localization using Acceleration and Air Pressure Sensors", 2nd International Forum on Applied Wearable Computing (IFAWC), Zurich, Switzerland, March 17-18, 2005.

[19] Harle, R., "A Survey of Indoor Inertial Positioning Systems for Pedestrians," *Communications Surveys & Tutorials, IEEE* , vol.15, no.3, pp.1281,1293, Third Quarter 2013

[20] Tan, W-T.; Baker, M.; Lee, B.; Samadani, R., "The Sound of Silence", Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems, SenSys'13, Article No. 19, November 11-15, 2013

[21] Hemminki, S.; Nurmi, P.; Tarkoma, S., "Accelerometer-Based Transportation Mode Detection on Smartphones", Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems, SenSys'13, Article No. 19, November 11-15, 2013

[22] Van Krevelen, D.W.F.; Poelman, R., "A Survey of Augmented Reality Techologies, Applications and Limitations," The International Journal of Virtual Reality, 9(2):1-20, 2010.