Detecting Pedestrian Flocks by Fusion of Multi-Modal Sensors in Mobile Phones

Mikkel Baun Kjærgaard, Martin Wirz, Daniel Roggen and Gerhard Tröster
Wearable Computing Laboratory
ETH Zurich, Switzerland
$mikkelk \| wirzma \| roggen \| troester@ife.ee.ethz.ch$

ABSTRACT
Previous work on the recognition of human movement patterns has mainly focused on movements of individuals. This paper addresses the joint identification of the indoor movement of *multiple* persons forming a cohesive whole - specifically a flock - with clustering approaches operating on features derived from multiple sensor modalities of modern smartphones. Automatic detection of flocks has several important applications, including evacuation management and socially aware computing. The novelty of this paper is, firstly, to use data fusion techniques to combine several sensor modalities (WiFi, accelerometer and compass) to improve recognition accuracy over previous unimodal approaches. Secondly, improve the recognition of flocks using hierarchical clustering. We use a dataset comprising 16 subjects forming one to four flocks walking in a building on single and multiple floors. With the best settings, we achieve a F-score accuracy of up to 87 percent an improvement of up to twelve percent points over existing approaches.

Author Keywords
pattern recognition; crowd behavior sensing; mobile sensing; signal strength-based methods

ACM Classification Keywords
I.5.4 Pattern Recognition: ApplicationsSignal Processing

General Terms
Design, Experimentation, Measurement, Performance.

INTRODUCTION
Current methods for studying human mobility patterns either consider movements of individual persons [11], or the mobility patterns of multiple persons, but often statistically aggregated over long time periods (days, weeks) [17]. However, multiple persons form collective movement patterns in their everyday life. Recognising these patterns is important to understand many real-world situations. One collective movement pattern which can often be observed among pedestrians moving through indoors or outdoors spaces is the *flocking* behavior (in laymen's term a "group" of persons). It captures a number of people walking together as a compact, coherent unit for some time. Such a unit is referred to as a pedestrian *flock* [6].

Today, most people have a location-aware mobile device and many mass gatherings in urban spaces such as festivals or sports events start to offer specially tailored mobile apps to their visitors. The same is seen for venues like airports and shopping malls and even companies are developing apps for their employees as a means to communicate with them. Recent surveys have identified that users of such mobile apps are responsive to getting assistance through apps during an emergency situations [28] making them an ideal platform for emergency management and would allow to provide pedestrians with personalized, targeted information e.g. by receiving direction assistance from rescue personnel which have a better overview of the situation, e.g. thanks to cameras, etc. This helps 'load-balance' an evacuation process by optimally directing people to the most suitable exit (in terms of distance and jamming in front of exits).

We argue that the detection of pedestrian flocks can further smoothen the evacuation process: It has been shown that the majority of pedestrians walk in groups through urban spaces rather than alone [20, 2]. Moussaïd et al. [20] show that up to 70 percent of the observed pedestrians in a commercial street are walking in groups. Further, especially if felt in a dangerous situation, people display affiliative behaviors, as argued for by Mawson [18]. In particular during an evacuation situation, people try to stay in groups of familiar members rather than acting as individuals. Hereby, during such situations, social ties (families and/or friends) remain strong and cooperation is predominating within these groups, as opposed to selfish, uncooperative behavior. Therefore to avoid confusion and to speed up the evacuation process we need to identify socially affiliated groups by detecting their flocking behavior. This enables the sending of personalized messages directed to particular groups rather than uniformly for every individual and thus helps to improve the evacuation process.

Additionally, for research studies of evacuation trials in building complexes flock detection enables better post-hoc analysis tools. For instance, researchers will be able to analyze the size of flocks and how they form, dissolve or are affected by

UbiComp '12, Sep 5-Sep 8, 2012, Pittsburgh, USA.

built environment, such as e.g. being slowed down in narrow passages [24]. Another area is *Socially aware computing* [17] which considers the opportunities that arise when looking at mobile devices as a collaborating collective. Flocking is here an element of monitoring and analyzing social interactions and the aggregated human behavior and social dynamics. For instance, being able to detect flocks would enable the extension of sensing tools for social psychology studies [21] to sense people's mood based on their attitude towards joining flocks.

Previous works on flock detection has mainly focused on the computational efficiency of spatial algorithms on outdoor GPS traces of vehicles or animals [8, 15, 26]. For urban areas a spatial flock detection algorithm to detect outdoor pedestrian flocks from noisy urban GPS positions has been proposed [27]. However, most humans spend a lot of their day indoor and many applications require indoor flock detection. For indoor flock detection individual sensors have been considered, e.g., [23] investigate the potential of using acceleration sensors to detect indoor flocks. While the authors observe a higher correlation in the acceleration patterns among members of the same flock, the lack of spatial information limits the detection accuracy. Previous work has also considered vision-based recognition of crowd movements, but such methods have limited applicability indoors due to the difficulties in placing cameras to get a proper view of the whole crowd [19]. In [13] the authors propose several methods to use WiFi signal strength information to derive spatial information to detect pedestrian flocks in indoor environments using density clustering. While they achieve a reasonable accuracy in most scenarios, the accuracy depends on the concentration of WiFi access points and calibration efforts. Furthermore, their method for clustering is not able to split a crowd into flocks as it will merge them if they are spatially co-located.

These past work among others suffer from the limitation of using a single sensor modality. Depending on a single sensor modality makes recognition prone to sensor failures, e.g., compasses failing due to magnetic disturbances, accelerometers being ill placed to sense motion or WiFi radios providing poor measurements when WiFi access point density is low. In this paper, we propose a multi-modal method to recognize pedestrian flocks in indoor spaces using sensing modalities of common mobile devices: accelerometers, compasses and WiFi radios. Using a multi-modal approach to recognizing pedestrian flocks has several advantages. Firstly, robustness can be improved as the failing of individual sensors can be masked by other sensors. Secondly, detection accuracy might be improved by fusing the measurements of the different modalities. Thirdly, being able to utilize different combinations of sensors opens up for energy savings as methods can select a combination of sensors providing the best results for a given energy budget.

To infer pedestrian flocks, we apply hierarchical clustering techniques operating on features computed from measurements of the different sensor modalities. The value of the individual sensor modalities and their features are comparatively assessed. As fusion technique we apply cluster-based weighted majority voting. From acceleration we compute features for movement similarity and correlation in acceleration variance, for compasses turn similarity and correlation in relative heading changes and for WiFi radios the similarity in signal strength, and proximity, speed and heading differences from location-fingerprinting estimated positions. We provide extensive experimental results using a dataset comprising 16 subjects forming one to four flocks walking in a building on single and multiple floors. To evaluate the properties of the multi-modal approach and the use of hierarchical clustering, we consider the improvements in accuracy for both environments with low and high concentrations of WiFi access points and show improvements over both previous work and machine learning-based joint modeling approaches. Furthermore, we study the method's dependency on homogeneity versus heterogeneity in flock sizes.

METHOD FOR FLOCK DETECTION

The notion of a Flock

What people in general regard as a pedestrian flock is hard to capture in a formal definition. Dodge et al. [6] argues that patterns for, e.g., human and animal behavior should be an interpretation of generic movement patterns in a particular application domain. In this work we use the following definition given in [13]. It evaluates the existence of moving clusters with regards to the ground truth location data. That the definition depends on ground truth allows us to apply it to all sensor modalities. Furthermore, it captures that people can join and leave the flock as the moving cluster is allowed over time to change in size as long as it remains larger than ν.

A pedestrian flock F is a moving cluster that exists for the duration $t \geq \tau$ and consists of more than $n \geq \nu$ people where τ and ν are application specific.

Method Overview

Figure 1 shows the proposed flock detection method. We assume that a sensing service will be deployed on persons' mobile devices through existing distribution channels, e.g., as a part of a mobile application for a mall, cultural event or company, or as a system service (1). When requested the n mobile devices will each measure acceleration, compass heading and the signal strength of nearby WiFi access points. The service will as needed stream the measurements to a remote flock detection service as measurement vectors $m_1, \ldots, m_n$ (2). The remote flock detection service might be implemented as a web service running on a cloud platform. The measurements are processed by the feature extraction chains to calculate features from the acceleration, compass and WiFi measurements (3). Afterwards, for each feature, the feature values from each device are pair-wise correlated to obtain a pair-wise similarity measure. This similarity measure indicates how similar the feature values from two devices are with respect to each other. These pair-wise similarity measures will be gathered in a $n \times n$ similarity matrix S_i for each feature space (4).

We divide the clustering process into a two stage process of spatial and temporal clustering as pedestrian flocks are clusters of people where a majority stay together over time. For instance, in comparison a single stage trajectory clustering would fail as it would only detect the people that stay together all the time. We assume that each sensor type spans its own feature space. For each feature space independently, we perform clustering on the similarity estimates to identify clusters of devices with similar feature values in the feature space (5). For the fusion of the clusters identified in each feature space (6), a weighted majority voting is performed that outputs a set of clusters that the majority of features agree on (7). As flocks are clusters of people where a majority stay together over time, temporal clustering is performed to combine highly similar clusters that exists for several successive time steps into flocks (8). The final output consists of devices grouped into flocks and thereby people as well (9). In the following we will discuss the steps in more detail.

Acceleration Features

To help determine if individuals walk together we will compute two features from acceleration measurements: Overlap in Movement Behavior (OMB) and Windowed Cross-Correlation of Acceleration (WCCA).

Overlap in Movement Behavior (OMB)

There has been extensive research in activity recognition on acceleration signals [22, 16]. For flock detection it would be relevant to compare the overlap in activity classifications to identify similarities, e.g., if a person is walking down the stairs potential flock members would mimic this behavior maybe with a small delay. However, the recognition results from only acceleration signals have not been too promising when considering cross device and person performance. E.g., [22] reports a low performance when studying the cross person performance. Lester et al. [16] report improved results but use better and more sensors than what is available in current mobile phones. Therefore we limit ourselves to detecting stationary vs. moving activities but the proposed feature is extensible to more comprehensive activity classification schemes when these can be robustly detected on common mobile phones [17]. Previous work have shown that moving versus still behavior can be robustly classified using a simple threshold based classifier based on the variance of the acceleration magnitude. We compute two lists of moving or stationary classifications M_a and M_b with an entry for each t of the sampling frequency in the time interval $[t_0 - T, \ldots, t_0]$. Each entry of the list, e.g. for M_a is computed as moving if $var(w(A_a, T)) > V$ where V is a variance threshold. Given a function f that returns 1 if $M_a \hat{=} M_b$ and 0 otherwise. With this, the OMB similarity feature computed over a window is given as:

$$s_{a,b} = \frac{\sum_{t=t_0-T}^{t_0} f(M_{a_t}, M_{b_t})}{n}$$

Windowed Cross-Correlation of Acceleration (WCCA)

Related work suggests that windowed cross-correlation can be utilized to compare if two acceleration signals are the result of similar movement behavior [23]. The method considers the variance of the signal magnitude to mask variations in device orientation and small differences in movement trajectories. Furthermore, the maximum cross-correlation for different lags is computed as people part of pedestrian flocks are not constrained to walk in-step. The similarity of a pair of devices a and b will by this feature be computed from measurement streams of acceleration magnitudes A_a and A_b, from each device respectively, up sampled to 50 Hz to mask differences in sampling rates. Given a windowing function $w(x, t_s, t_e)$ where x is the measurement stream and t_s and t_e is the window start and end. We compute two lists of acceleration variances V_a and V_b with an entry for each t of the sampling frequency in the time interval $[t_0 - T, \ldots, t_0]$ where t_0 is the current time and T is a window threshold. Each entry of the list for time t, e.g. for V_a is computed as $var(w(A_a, t - T, t))$. Because behavior changes can be shifted in time between flock members we compute the maximum cross correlation with a lag l between minus one to plus one second. Therefore given a cross correlation function $corr(x, y, l)$ where x and y are value lists and l is the lag the WCCA similarity of a and b is computed as follows:

$$s_{a,b} = max(corr(V_a, V_b, l), l \in [-1, 1])$$

Compass Features

Compass measurements can also help to determine if individuals walk together by measuring the phone's orientation. However, for the compass features we can not depend on the absolute heading for the following two reasons. First, we do not assume that the compass has been recently calibrated to show the correct absolute heading and, secondly, we do not assume that the mobile phone has a specific orientation towards the body so that the system can determine, e.g., if the target is moving backward or forward. Therefore we instead consider relative heading changes which avoids both issues as we only assume the phone to keep the same position with respect to the body. We compute two features based on relative heading changes: the windowed cross correlation in relative heading (WCCH) changes and time since last turn (TSLT).

Windowed Cross Correlation in Relative Heading (WCCH)

For the cross-correlation feature the similarity of a pair of devices a and b is computed from the measurement streams of compass headings C_a and C_b, from each device respectively, up sampled to 50 Hz to mask differences in sampling frequencies. We compute two lists of heading deviations H_a and H_b with an entry for each t of the sampling frequency in the time interval $[t_0 - T, \ldots, t_0]$ where t_0 is the current time and T is a window threshold. We will consider the following two windows $w_1(x) = w(x, t, t - \frac{T}{2})$ and $w_2(x) = w(x, t, t + \frac{T}{2})$. Given a mean function for orientation (handling 360° passes) each entry of the list, e.g. for H_a is computed for t as $mean(w_1(C_a)) - mean(w_2(C_a))$. Again as behavior changes can be shifted in time between flock members we compute the maximum cross correlation with a lag l between minus one to plus one second. Therefore the WCCH similarity of a and b is computed as follows:

$$s_{a,b} = max(corr(H_a, H_b, l), l \in [-1, 1])$$

Time since last Turn (TSLT)

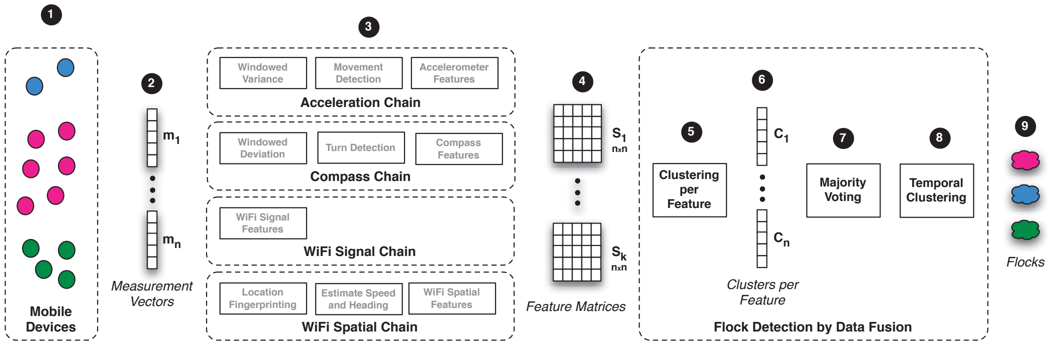

Figure 1. Overview of the method for indoor flock detection

For the turn-based time since last turn feature for each stream of compass orientation measurements, e.g. C_a, we detect turns using the following condition originally proposed in [5] where $y = mean(w_1(C_a)) - mean(w_2(C_a))$, $StdDev(x)$ is the standard deviation of a list and G is a guard factor.

$$y \geq \frac{StdDev(w_1(C_a)) + StdDev(w_2(C_a))}{2} + G$$

For each device we compute a list K which has an entry of zero when a turn is detected and else the previous value plus $1/50 seconds$. From the two lists we compute the TSLT similarity as:

$$s_{a,b} = \sum_{t=t_0-T}^{t_0} ||K_{a_t} - K_{b_t}||$$

WiFi Features

To compute WiFi features we implemented two approaches originally proposed in [13]. The first approach computes a spatial feature (Spatial) where WiFi positioning is applied to map signal strength measurements into a spatial space. The second approach computes signal strength features (Signal) where flocks are detected from features computed directly from the signal strength measurements.

Spatial models the similarity between two mobile devices as the shortest walking distance between their positions estimated via location fingerprinting. To increase robustness we extended this feature with an uncertainty estimation step to add the average uncertainty of the estimated positions to the calculated walking distance. This means that high uncertainty increases the distance between mobile devices thereby making it less likely that in the later clustering the two are clustered together. That we model the uncertainty increases the accuracy of the flock detection in the later evaluation with around one percent point.

Signal features are for two mobile devices the Euclidean distance in signal strength space between their two signal strength vectors.

SpatialSpeed and SpatialHeading are two additional features based on the spatial WiFi positions and captures speed difference (SpatialSpeed) and heading difference (SpatialHeading). Both computed as the minimum sum of differences in speed and heading, respectively within a window threshold T of measurements and allowing for a lag between minus one to plus one second.

Improved Spatial Clustering of Flocks

Previous work have consider the use of DensityJoin-Clustering [29] for spatial clustering to detect flocks. However, this approach has several drawbacks, firstly, this method will often merge nearby flocks due to noise which means that individuals which belongs to different flocks are often wrongly clustered together. Secondly, the method does not provide means to control the geometric stretch of the clusters. To address these issues in this work, we propose to use agglomerative hierarchical clustering in the centroid-link variant [9]. This clustering method has the advantages that it produces more compact cluster and that we can limit the geometric stretch of the clusters, e.g., using study results about the shape and stretch of pedestrian flocks [20]. As we in this work do not assume a priori to know the number of potential flocks we can not easily apply partitional clustering methods such as K-means [9].

Hierarchical clustering structures data into a hierarchical structure according to a feature matrix [9]. The results of hierarchical clustering are usually depicted by a binary tree where the root node represents the whole data set and each leaf node is regarded as a data object. The intermediate nodes describe the extent to which objects are similar to each other. The agglomerative clustering works on this tree starting with clusters including exactly one object and through a series of merge operations creating clusters satisfying a given proximity criteria. In this work we use the centroid-linkage proximity criteria that computes the distance between a new object and the centroid of each cluster. As centroid-linkage is a geometric method it requires that geometric distances are well defined which is only satisfied by the WiFi features: Signal and Spatial. As the acceleration, compass, and WiFi speed and heading features are not geometric we can not use the above method here and therefore we use DensityJoin-Clustering [29]. As illustrated in Figure 2 we run the clustering on the similarity matrix S computed for each feature where the clustering method depends on the type of feature.

Fusion of Clusters for Improved Robustness

To improve the robustness of our method we propose to fuse the clusters identified for the individual features. Previous works on clustering have shown that this technique is able

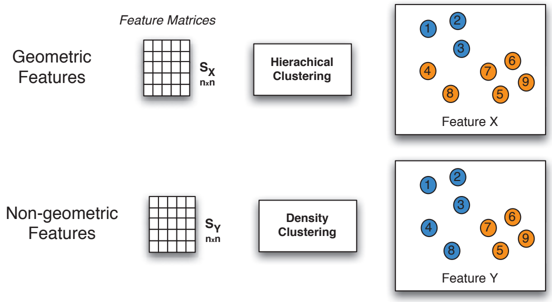

Figure 2. Clustering shown with coloring for a geometric feature (WiFi) and for a non-geometric feature (Accelerometer, Compass)

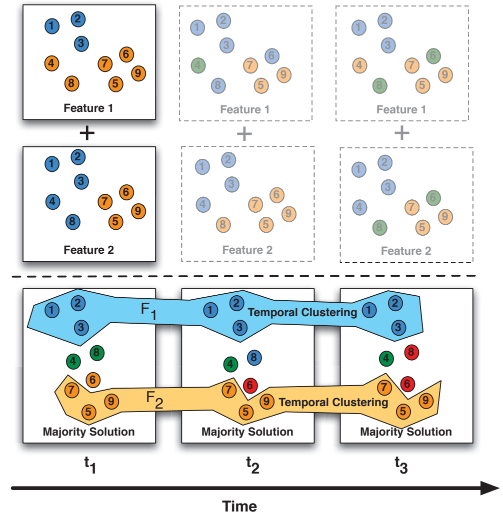

Figure 3. Schematical visualization of the flock detection algorithm using weighted majority voting for three time steps and temporal clustering which identifies two flocks F_1 and F_2.

to improve the overall clustering quality if the quality of the individual clustering is not very bad [25]. In this paper we propose to apply weighted majority voting to combine clusters identified in the different feature spaces.

The fusion by majority voting works on the sets of clusters outputted by the feature clustering. We assume that we select weights for each feature space that capture the quality of the feature. The voting proceeds as follows: for each device pair, calculate the sum of the weight of the feature spaces where the devices appear in the same cluster. If the summed weight is larger than fifty percent, the two devices are added as a cluster in the majority solution. If one of the devices is already part of a solution cluster, instead of adding a new cluster, both devices joins the existing cluster.

As an example Figure 3 shows the majority voting for three time steps assuming similar weights for the two features. For time step t_1 we have two different sets of clusters outputted by the clustering of Feature 1 and 2. Then a majority voting is performed. In our example, as device 4 and 8 are in different clusters for Feature 1 and 2 they will end up in their own cluster after the majority voting. Therefore the majority voting, e.g., can help address situations where features agree that two devices correctly are in the same flock but differ in their groupings with respect to other devices.

The fused clusters in the majority solution represent groups of devices with similar features. Flocks, however, are groups of people that stay together most of the time. Hence, lastly, the algorithm then performs a temporal clustering based on the method proposed by Kang et al. [10]. Clusters that exist for several successive time steps are combined into flocks by respecting a set of rules. The rules are parameterized by three thresholds: two thresholds for promoting and dissolving flocks that captures the temporal timespan of a cluster before it is considered a flock and not considered a flock anymore, and a similarity measure for the temporal clustering between zero and one to capture the needed overlap in flock members. As a result, in the example presented in Figure 3 two flocks are detected with three devices in each. The remaining devices 4, 6, and 8 are not part of a flock as they are not members of a similar cluster over several time steps with at least two members.

EVALUATION

We evaluate the flock detection accuracy of our data set for the individual features, the improvements using data fusion and hierarchical clustering, the influence of the different parameter settings, the effect of the availability of WiFi access points, the method's dependency on homogeneity versus heterogeneity in flock sizes.

Data Set

We conducted an experiment in an indoor environment to obtain a data set. A subset of this data was considered by our earlier work [13]. The experiment was conducted according to a predefined script and monitored with video cameras and manual annotation in order to obtain a ground truth against which the method's performance can be evaluated. For the ground truth we analysed the video and applied Definition with τ high enough (15 seconds) so that we do not regard a moving flock passing a stationary flock as a merge and ν as two individuals.

The experiment took place on two floors of an office building as shown in Figure 4 with 19 participants. We use data from 16 participants as a few of the phones went into sleep mode during the experiment. All of them were equipped with different kinds of smart phones, here we consider data from the following combination of phones Google Nexus One (2), HTC Desire (8), Nokia N8 (3), Nokia C7 (2) and Nokia N97 (1) reflecting the heterogeneous hardware platforms experienced "in the wild" deployments. All the phones ran dedicated sensor logging software to record WiFi signal strength, acceleration and compass heading with a sampling rate of 0.5-1 Hz, 20-40 Hz and 20-40 Hz, respectively on the Android phones and 0.1 Hz, 38 Hz and 10 Hz, respectively on the Nokia phones.

For the WiFi positioning we collected fingerprints throughout the building in a 1.5 meter grid with Google Nexus One phones standing at each grid point for 30 seconds - in all the fingerprinting process took one hour for two persons.

During the experiment, the participants formed flocks of sizes

Table 1. Scenes of the experiment

Name	Flocks	Sizes	Description
Horiz. A	$1-3$	3,5,8	Flocks both stationary and moving on the same floor with changed flock members (4:08 min)
Horiz. B	$1-3$	2,6,8	Repetition with changed flock members and sizes (5:42 min)
Vertical A	$1-4$	3,3,5,5	flocks both stationary and moving on both floors (4:27 min).
Vertical B	$1-4$	3,3,5,5	moving flocks on both floors (10:25 min).
Vertical C	$1-4$	1,2,5,8	flocks both stationary and moving on both floors (6:39 min).
Vertical D	$1-4$	1,2,5,8	moving flocks on both floors (7:23 min).
Training A	$1-3$	3,5,8	flocks both stationary and moving on the same floor (4:30 min).
Training B	1	16	stationary flock on the same floor (1 min).

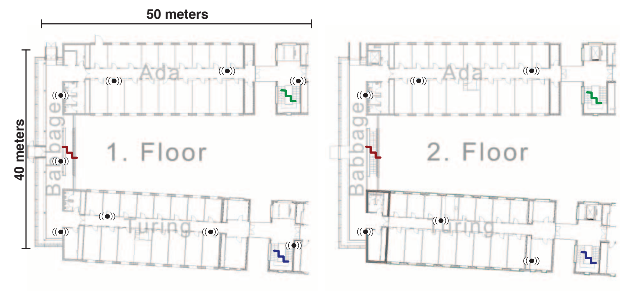

Figure 4. Building layout with icons for the 3 stair cases and 15 WiFi access points (Additional 36 WiFi access points were visible from other buildings).

spanning from 1 to 16 members. The experiment was organised into scenes where each scene was designed to test different elements of our method as listed in Table 1. During each scene the flocks would meet, merge and split several times. In order to obtain recordings of natural behavior, we did not instruct the subjects on how to move or behave and only told them to put the mobile phones into their trouser pockets and to always follow and stay close together (where 'close' was not further specified but was observed to be around $1.5m$).

Initial Data Analysis

As stated before, compared to earlier work we relax for our dataset some of the assumptions of prior work. For instance, that we utilize acceleration and compass measurements from mobile phones loosely worn in the pocket compared to dedicated sensors firmly attached at a stationary position. To visualize the quality of the data Figure 5 to 10 plots raw acceleration and compass heading for four devices two of the same flock (1) and two others from different flocks (2) and (3), where (2) is a stationary flock and (3) is moving. To configure the methods we used the best parameters identified on our training data set. For *OMB*, *WCCA*, *WCCH* and *TSLT* we use a window threshold of 15 seconds, for *OMB* a variance threshold of 0.25 and for *TSLT* a turn detection guard value of 20°.

The acceleration data shows differences among the flocks (Figure 5) which is also present in the windowed variance (Figure 7). However, there are also differences, e.g. in signal magnitude for devices within the same flock. The movement or still state for the devices can be stablely detected

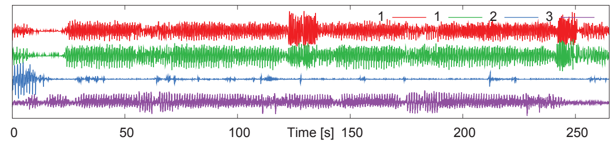

Figure 5. Acceleration

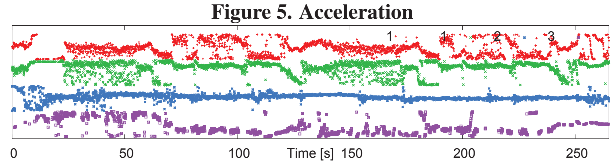

Figure 6. Compass

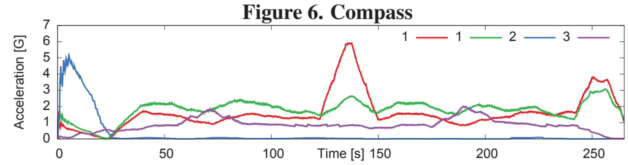

Figure 7. Variance of Acceleration for WCCA

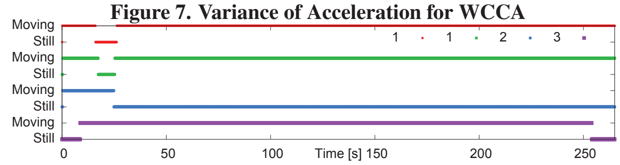

Figure 8. Movement versus Still Detection for OMB

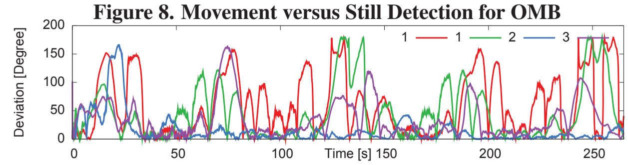

Figure 9. Windowed Deviation of Compass Headings for WCCH

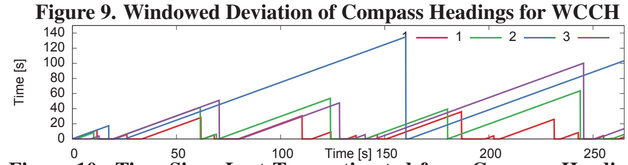

Figure 10. Time Since Last Turn estimated from Compass Headings for TSLT

from the data (Figure 8). The compass data also shows differences among the flock (Figure 6) but in some periods a high level of noise is also visible. The computed windowed deviations has overlap within the flock but the high level of noise create a number of noise peaks (Figure 9). The noise makes the difference to the other moving flock (3) less apparent. The value estimating the time since last turn shows a more similar pattern within the flock and dissimilar to other flocks (Figure 10). However, the symmetry in the building layout and the partly synchronized flocking behavior in the experiment creates additional similarity in the turning behavior between the flocks. In summary these observations suggest that there is some information to gain for flock detection from acceleration and compass measurements but the value is constrained due to noise in the data and the information content.

Flock Detection Accuracy

By evaluating the flock detection accuracy, we are interested in how well the individual subjects can be assigned to the correct flock. For this, we obtain a meaningful evidence by calculating the F-measure as applied in the context of clustering [14]. The F-measure is calculated as the average precision and recall when comparing each predicted flock to the most similar ground truth flock. To aggregate the results over

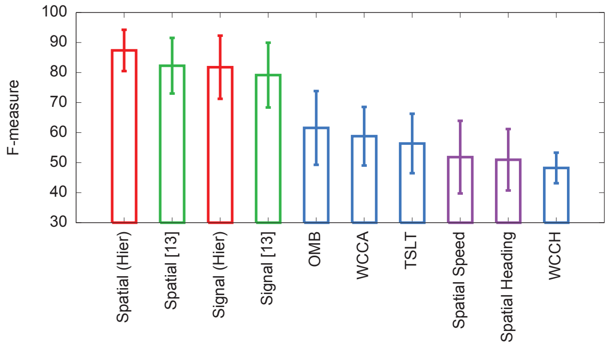

Figure 11. Accuracy results by the F-Measure for the ten individual features.

time we calculate the average F-measure for all time steps in a scene.

In our evaluation we consider eight feature setups: *Spatial (Hierarchical)*, *SpatialHeading*, *SpatialSpeed*, *Signal (Hierarchical)*, *OMB*, *WCCA*, *WCCH* and *TSLT*. We use the best parameters, e.g. for clustering thresholds, identified using our training data set. Furthermore, we compare our method to previous work: *Spatial [13]* and *Signal [13]* were we use the parameters given in [13]. The individual results are shown in Figure 11 as the average F-measure and the standard deviation for the results of the individual scenes.

The results highlight that there is a large difference in the accuracy provided by the different features and for the individual scenes as shown by the standard deviation. The generally best performing method is *Spatial (Hier.)* improving with five percent points over prior work *Spatial [13]* and at the same time reducing the standard deviation with several percent points. For the accelerometer based features *OMB* provides the best accuracy and for the compass *TSLT* provides the highest accuracy. In comparison both *SpatialHeading* and *SpatialSpeed* has the lowest performance together with *WCCH*.

Influence of Thresholds

To study the impact of different thresholds on the method's performance Figure 12 to 14 plot results with a range of values for each threshold. We focus on the two WiFi methods as these have the best performance. The results on each figure is plotted with the best parameter setting identified for the two other parameters for each feature. The impact of the distance threshold used for the centroid-linkage is plotted in Figure 12. The threshold is given in multiples of standard deviations from the mean where the mean and the standard deviation is the intra-flock subject feature distance recorded during the training scene. The best performance is for *Spatial (Hier.)* and *Signal (Hier.)* achieved at 5 and 6 times the standard deviation, respectively. The impact of this threshold is stronger for *Signal (Hier.)* than for *Spatial (Hier.)*. Figure 13 shows results for the similarity threshold of temporal clustering which has a strong impact on the accuracy. The best performance is for *Spatial (Hier.)* and *Signal (Hier.)* achieved with a similarity threshold of 0.3 and 0.4, respectively. This means that the best performance is achieved allowing a significant change in flock members between different time steps. The final parameter is the time thresholds for promoting and dissolving flocks. From the graph we can observe that this parameter is the least significant of the three as it only have a minor impact on the accuracy.

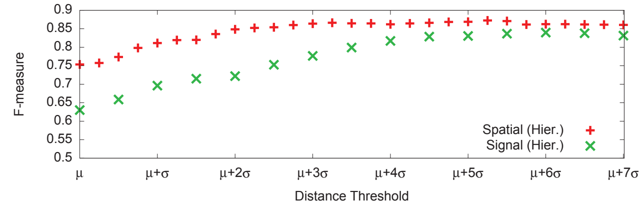

Figure 12. Accuracy results by the F-Measure for clustering thresholds

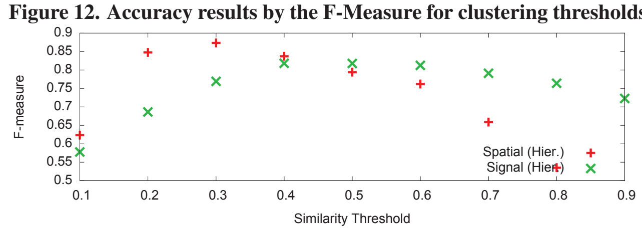

Figure 13. Accuracy results by the F-Measure for the similarity threshold for temporal clustering

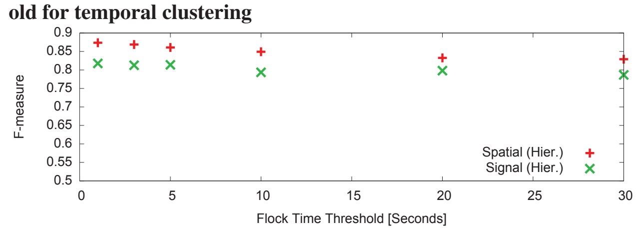

Figure 14. Accuracy results by the F-Measure for the time threshold for promoting and dissolving flocks.

Improvements by Data Fusion

To quantify the improvements made possible using data fusion by majority voting we have evaluated different combinations of features. As *Spatial (Hier.)* and *Signal (Hier.)* individually provide the best results we focus on combining them with one or more of the accelerometer and compass features to learn what combinations give the best accuracy. In the weighted majority voting we also assigns a weight of 50% to these features and then divide the remaining 50% between the accelerometer and compass features. It is left as future work to consider automatic training of these weights. In comparison, the best combination of the accelerometer and compass features is *OMB* and *TSLT* when combined achieve an accuracy of 62.9%.

As the accuracy of the WiFi based methods depend on the density of WiFi access points we consider in the following three scenarios: One were 98% of the access points are present (51 in total), one where only 50% of them are available and a final one with only 20%. The scenarios were simulated as a decimation of the number of access points and repeated ten times to account for the random selection of access points to decimate. We therefore simulate different deployments as they would be found in different buildings. The best results for combinations with Spatial are shown in Figure 15 and for Signal in Figure 16.

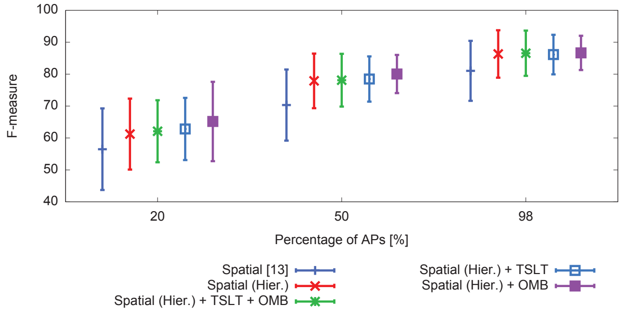

Figure 15. Accuracy results by the F-Measure for combinations with *Spatial (Hier.)* (Results offset for visibility).

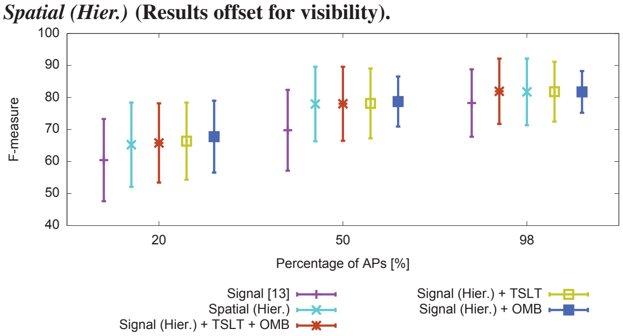

Figure 16. Accuracy results by the F-Measure for combinations with *Signal (Hier.)* (Results offset for visibility).

Table 2. Flock membership assignment accuracy by the F-measure for different group sizes for *Spatial (Hier.)*

	3,3,5,5	1,2,5,8
Vertical A / C	82.1	78.3
Vertical B / D	90.0	89.7

For *Spatial (Hier.)* the best performance is achieved when combined with OMB alone of 86.7%, 80.0% and 65.2% in the 98%, 50% and 20% cases, respectively. When the feature is combined with more features the accuracy decreases. *Signal (Hier.)* achieves with 98% WiFi access points the best performance of 82.0% when combined with the two features *OMB* and *TSLT*. In the cases of 50% and 20% access points the combination with *OMB* outperforms the other options. Again combining with more features does not improve the accuracy. Again we see a clear improvement over *Signal [13]* and *Spatial [13]* both in terms of improved accuracy of up to twelve percent points and lower standard deviation.

Impact of Flock Sizes

In this section we compare the results of the four vertical scenes as we in these have flocks with different sizes. Furthermore, as an additional challenge Vertical C and D also include a single person walking alone that should not be included in the other flocks. The results are listed in Table 2 for *Spatial (Hier.)*. One can observe that the difference in scene setup between (A/C) versus (B/D) has the largest impact on the individual results. For similar scenes with variations in flock sizes there is only a minor difference. The largest difference is for A / C where it is 4% and therefore the results indicate that differences in flock sizes have a small impact on the method's accuracy.

Table 3. Best performing method for different conditions.

	WiFi Access Point Availability: No	Low	High
In the Wild	*OMB + TSLT*	*Signal (Hier.) + OMB*	*Signal (Hier.) + OMB*
Controlled	*OMB + TSLT*	*Signal (Hier.) + OMB*	*Spatial (Hier.) + OMB*

DISCUSSION

We have presented a variety of results to show how our method perform using various combinations of the mentioned sensors, and by emulating with different amounts of visible access points. Table 3 shows a summary of these results. From this table, we can draw a number of conclusions about when each combination is the most appropriate.

Accelerometer and Compass Combinations: The advantage of using only accelerometer and compass features is that no infrastructure is required. The main problem with this approach is that the accuracy is low. Therefore this approach should only be considered an option when no WiFi infrastructure is available where OMB + TSLT is the best combination.

Combinations with Signal (Hier.): The advantages of using Signal (Hier.) are that coverage of the resulting method only depend on whether WiFi access points are present or not. The drawback is that the method's parameters depend slightly on the number of access points and the distance to them. When combined with OMB the accuracy is further improved. The advantages make this approach relevant for "in the wild" applications because no a-priori fingerprinting is needed. For controlled environments the method is also relevant when there is a low density of WiFi access points as Signal (Hier.) + OMB has a better performance than combinations with Spatial (Hier.) in this case.

Combinations with Spatial (Hier.): For Spatial (Hier.) we utilize fingerprinting-based positioning, which is a positioning method that is fairly accurate, requires no extra hardware, and which coverage is easy to increase by dedicated or user-based fingerprinting [7]. The drawbacks of the positioning method are the time needed to collect fingerprints and update them over time and for our case that the positioning accuracy is negatively impacted when people cluster [4]. These drawbacks of the positioning method might make it unfeasible to use Spatial (Hier.) for "in the wild" applications. However, the work needed for preparing a few buildings to support evacuation procedures or a study is still limited to a couple of hours. Spatial (Hier.) provides generally a high detection accuracy which is further improved when combined with OMB. Additionally, Spatial (Hier.) has the added benefit that the spatial position of the flocks are available.

Other Sensor Modalities

For indoor positioning and proximity sensing many sensor modalities have been explored that might also be relevant to consider in future work on flock detection. For instance, other radio technologies have been used for proximity sensing, e.g. Bluetooth [3] or Zigbee [4]. However, Bluetooth has drawbacks in crowded areas due to signal attenuation and interference, and Zigbee is a technology not widely deployed as of today in mobile devices. WiFi signal strength measurements measured at the base stations could also be used and integrates directly into the proposed methods. Furthermore, for activity classification sensors, such as, gyros and barometers are now also entering modern smartphones enabling more extensive classification schemes [16, 17]. For all of such other sensor types the proposed approach is general enough to fuse such additional information to further improve recognition accuracy.

Machine-learning based Joint Modeling

In this work we fuse the clustering results for each individual feature. Therefore, one might ask if joint modeling of features would improve the performance. To evaluate this we extended our processing chain with a step that given the individual features for each pair of devices uses a machine-learning algorithm to predict if they are flocking together or not. These pair-wise results are then feed into the clustering part of our method. Given that this data is none geometric we are limited to using DensityJoin-Clustering. We evaluated this approach using both a Naive Bayes and a Support Vector Machine (SVM) classifier trained on our training data. However, the results were not encouraging as the best average F-score was with Naive Bayes of 76.6% for WiFi Spatial and all four accelerometer and compass features. This result is eleven percent points lower than our method based on fusion of independent feature clustering.

One of the main reasons for this failure is the existing labeling of training data into a binary decision for each pair: flocking together or not. Firstly, as binary classifications are non-geometric data we can not apply hierarchical clustering. Secondly, given the binary classification a single wrong classification might cause the clustering to wrongly join two flocks. Thirdly, for the category not flocking the classifier is trained with data covering very different situations. E.g. two flocks passing close by versus two flocks faraway from each other. A more detailed labeling might make it possible for a classifier to better detect these different situations and thereby make the overall classification more robust.

CONCLUSION

In this paper we addressed the joint identification of the indoor movement of *multiple* persons forming a cohesive whole - specifically flocks - from smartphone sensor data. The novelty of this work is, firstly, to use data fusion techniques to combine several sensor modalities (WiFi, accelerometer and compass) to improve recognition accuracy over previous unimodal approaches. Secondly, we improved the recognition of flocks using hierarchical clustering. Thirdly, we studied the method's dependency on homogeneity versus heterogeneity in flock sizes. The conclusion was that a method based on Spatial (Hier.) + OMB provides the best accuracy with an average F-score accuracy of 87%. We analysed the advantages and drawbacks of the feature combinations considering their suitability for use "in the wild" or in controlled environments with no, low or high density of WiFi access points. The conclusion is that when there is no WiFi infrastructure a combination of OMB and TSLT is the most appropriate. For "in the wild" a method combining Signal (Hier.) and OMB is the best option. For controlled environments if there is a low access point density Signal (Hier.) and OMB is the best option whereas when there is a medium to high density Spatial (Hier.) and OMB is the best option.

For the mentioned application domains our results indicate that the flock detection methods can be used as a tool for sensing flocking patterns. Given the accuracy they can provide valuable input to new mobile-based evacuation or social-aware systems. How these systems deal with the uncertainty in the recognition is an interesting question which is a general challenge for ubicomp systems. The next step is to deploy our methods in the mentioned domains to realize the potentials and to study the performance of the method for flocking patterns as they appear in the different application domains.

In our ongoing work we pursue the following goals: Firstly, we would like to further improve the method and features, e.g., considering other types of sensors, clustering methods, e.g., spectral clustering and extend OMB with other types of behaviors such as walking on stairs [16]. Secondly, we would like to study system issues, such as, energy consumption [12] and privacy [1]. Thirdly, we would like to deploy the methods for flock detection within evacuation management and social-aware computing.

ACKNOWLEDGMENTS

This work is supported under the FP7 ICT Future Enabling Technologies Programme under grant agreement No 231288 (SOCIONICAL) and by a grant from the Carlsberg Foundation. We thank Michael Lind Mortensen and Henrik Blunck for helping with the data collection.

REFERENCES

1. M. S. Andersen and M. B. Kjærgaard. Towards a new classificaion of location privacy methods in pervasive computing. In *Proc. of the 8th Int. ICST Conf. on Mobile and Ubiquitous Systems (Mobiquitous)*. Springer, 2011.
2. A. F. Aveni. The not-so-lonely crowd: Friendship groups in collective behavior. *Sociometry*, 40(1):pp. 96–99, 1977.
3. N. Banerjee, S. Agarwal, P. Bahl, R. Chandra, A. Wolman, and M. D. Corner. Virtual compass: Relative positioning to sense mobile social interactions. In *Proceedings of the 8th International Conference on Pervasive Computing*, pages 1–21, 2010.
4. L.-W. Chan, J.-R. Chiang, Y.-C. Chen, C. nan Ke, J. Y. jen Hsu, and H.-H. Chu. Collaborative localization: Enhancing wifi-based position estimation with neighborhood links in clusters. In *Proceedings of the 4th International Conference on Pervasive Computing*, pages 50–66, 2006.

5. I. Constandache, X. Bao, M. Azizyan, and R. R. Choudhury. Did you see bob?: human localization using mobile phones. In *Proceedings of the 16th Annual International Conference on Mobile Computing and Networking*, pages 149–160, 2010.

6. S. Dodge, R. Weibel, and A.-K. Lautenschütz. Towards a taxonomy of movement patterns. *Information Visualization*, 7(3-4):240–252, 2008.

7. J. geun Park, B. Charrow, D. Curtis, J. Battat, E. Minkov, J. Hicks, S. J. Teller, and J. Ledlie. Growing an organic indoor location system. In *Proceedings of the 8th International Conference on Mobile Systems, Applications, and Services (MobiSys 2010)*, pages 271–284, 2010.

8. J. Gudmundsson and M. van Kreveld. Computing longest duration flocks in trajectory data. In *Proceedings of the 14th annual ACM international symposium on Advances in geographic information systems*, pages 35–42. ACM, 2006.

9. A. K. Jain, M. N. Murty, and P. J. Flynn. Data clustering: a review. *ACM Computing Surveys*, 31:264–323, September 1999.

10. J. H. Kang, W. Welbourne, B. Stewart, and G. Borriello. Extracting places from traces of locations. *ACM SIGMOBILE Mobile Computing and Communications Review*, 9(3):58–68, 2005.

11. D. H. Kim, Y. Kim, D. Estrin, and M. B. Srivastava. Sensloc: sensing everyday places and paths using less energy. In *Proceedings of the 8th International Conference on Embedded Networked Sensor Systems*, pages 43–56, 2010.

12. M. B. Kjærgaard. Location-based services on mobile phones: Minimizing power consumption. *IEEE Pervasive Computing*, 11(1):67–73, 2012.

13. M. B. Kjærgaard, M. Wirz, D. Roggen, and G. Tröster. Mobile Sensing of Pedestrian Flocks in Indoor Environments using WiFi Signals. In *Proceedings of the 10th IEEE International Conference on Pervasive Computing and Communications*, 2012.

14. B. Larsen and C. Aone. Fast and effective text mining using linear-time document clustering. In *Proc. of the 5th ACM SIGKDD int. conf. on Knowledge discovery and data mining*, pages 16–22. ACM, 1999.

15. P. Laube, M. Kreveld, and S. Imfeld. Finding REMO: detecting relative motion patterns in geospatial lifelines. *Developments in Spatial Data Handling*, 2005.

16. J. Lester, T. Choudhury, and G. Borriello. A practical approach to recognizing physical activities. In *Proceedings of the 4th International Conference on Pervasive Computing*, pages 1–16, 2006.

17. P. Lukowicz, S. Pentland, and A. Ferscha. From context awareness to socially aware computing. *Pervasive Computing, IEEE*, 11(1):32 –41, jan. 2012.

18. A. Mawson. Understanding mass panic and other collective responses to threat and disaster. *Psychiatry*, 68(2):95–113, 2005.

19. B. E. Moore, S. Ali, R. Mehran, and M. Shah. Visual crowd surveillance through a hydrodynamics lens. *Commun. ACM*, 54(12):64–73, 2011.

20. M. Moussaid, N. Perozo, S. Garnier, D. Helbing, and G. Theraulaz. The walking behaviour of pedestrian social groups and its impact on crowd dynamics. *PLoS ONE*, 5(4), 04 2010.

21. K. K. Rachuri, M. Musolesi, C. Mascolo, P. J. Rentfrow, C. Longworth, and A. Aucinas. Emotionsense: a mobile phones based adaptive platform for experimental social psychology research. In *Proc. of the 12th Int. Conf. on Ubiquitous Computing*, pages 281–290, 2010.

22. N. Ravi, N. Dandekar, P. Mysore, and M. L. Littman. Activity recognition from accelerometer data. In *Proceedings of The Twentieth National Conference on Artificial Intelligence*, pages 1541–1546, 2005.

23. D. Roggen, M. Wirz, D. Helbing, and G. Tröster. Recognition of crowd behavior from mobile sensors with pattern analysis and graph clustering methods. *Networks and Heterogeneous Media*, 2011.

24. A. Schadschneider, W. Klingsch, H. Klüpfel, T. Kretz, C. Rogsch, and A. Seyfried. Evacuation dynamics: Empirical results, modeling and applications. In *Encyclopedia of Complexity and Systems Science*, pages 3142–3176. Springer New York, 2009.

25. A. P. Topchy, A. K. Jain, and W. F. Punch. Combining multiple weak clusterings. In *Proceedings of the 3rd IEEE International Conference on Data Mining (ICDM 2003)*, pages 331–338, 2003.

26. M. Vieira, P. Bakalov, and V. Tsotras. On-line discovery of flock patterns in spatio-temporal data. In *Int. Conference on Advances in Geographic Information Systems*. ACM, 2009.

27. M. Wirz, M. B. Kjærgaard, S. Feese, P. Schläpfer, D. Roggen, and G. Tröster. Towards an online detection of pedestrian flocks in urban canyons by smoothed spatio-temporal clustering of gps trajectories. In *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Location-Based Social Networks*, 2011.

28. M. Wirz, D. Roggen, and G. Tröster. User acceptance study of a mobile system for assistance during emergency situations at large-scale events. In *Proceedings of The 3rd International Conference on Human-centric Computing*, 2010.

29. C. Zhou, D. Frankowski, P. Ludford, S. Shekhar, and L. Terveen. Discovering personal gazetteers: an interactive clustering approach. In *Int. workshop on Geographic information systems*. ACM, 2004.