Hindawi Publishing Corporation
Mobile Information Systems
Volume 2016, Article ID 3545327, 14 pages
http://dx.doi.org/10.1155/2016/3545327

Research Article

Identification of Partitions in a Homogeneous Activity Group Using Mobile Devices

Na Yu,[1] Yongjian Zhao,[1] Qi Han,[1] Weiping Zhu,[2] and Hejun Wu[3]

[1]*Department of Electrical Engineering and Computer Science, Colorado School of Mines, Golden, CO 80401, USA*
[2]*International School of Software, Wuhan University, Wuhan 430079, China*
[3]*Guangdong Province Key Laboratory of Big Data Analysis and Processing, Sun Yat-Sen University, Guangzhou 510006, China*

Correspondence should be addressed to Qi Han; qhan@mines.edu

Received 5 January 2016; Accepted 31 March 2016

Academic Editor: Peter Brida

Copyright © 2016 Na Yu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

People in public areas often appear in groups. People with homogeneous coarse-grained activities may be further divided into subgroups depending on more fine-grained behavioral differences. Automatically identifying these subgroups can benefit a variety of applications for group members. In this work, we focus on identifying such subgroups in a homogeneous activity group (i.e., a group of people who perform the same coarse-grained activity at the same time). We present a generic framework using sensors built in commodity mobile devices. Specifically, we propose a two-stage process, sensing modality selection given a coarse-grained activity, followed by multimodal clustering to identify subgroups. We develop one early fusion and one late fusion multimodal clustering algorithm. We evaluate our approaches using multiple datasets; two of them are with the same activity while the other has a different activity. The evaluation results show that the proposed multimodal-based approaches outperform existing work that uses only one single sensing modality and they also work in scenarios when manually selecting one sensing modality fails.

1. Introduction

People often appear in groups and participate in various activities in public areas. People with homogeneous coarse-grained activities may be further divided into subgroups based on more fine-grained behavioral differences. For instance, in emergency response situations such as fire evacuation, people have the same coarse-grained activity, that is, walking or running towards emergency exits. However, people may be heading for different exits and with different moving speeds, and people who are moving together can be considered as a subgroup. By monitoring these subgroups, the emergency control center can better guide people by directing each subgroup's route. Therefore, partitioning a group with the same coarse-grained activity into subgroups based on specific activity differences is very important. Similarly, tourists walk around in a park and walking is the same coarse-grained activity. Different walking flocks can be distinguished by the mobility patterns of the tourists; that is, people in the same subgroup should have similar direction and speed. A tour guide can easily manage the tourist group based on the walking flocks and send customized message to different subgroups which are heading to different attractions. Another example is people watching a game. Different subsets of the audience cheer for different teams in a game and the subgroups can be distinguished by the specific actions performed by them; that is, people in support of the same team typically perform certain gesture such as waving hands during the same time period when the team is performing well. Fans of the same team can be easily identified and they can be recommended to be friends to share information for future games. Partitioning groups with the same coarse-grained activity into subgroups based on specific activity differences is exactly the focus of this work.

Lots of work have been done in group detection and activity recognition using mobile devices, but the problem at hand has not been fully addressed by existing work as detailed in Section 2. We have been inspired by the divergence-based affiliation detection (DBAD) approach [1] which provides a framework to identify group affiliation given a sensing

modality to be used for an activity. Different from the group activity recognition problem which typically first recognizes each user's activity and then analyzes their cooperative or collaborative relationship in a group [2], the group affiliation detection problem is about how to identify which users have similar behavior instead of identifying their specific activities. However, one limitation of DBAD is that only one sensing modality can be used at a time to distinguish multiple subgroups, so it cannot accurately partition the groups when behavioral differences can be observed only through multiple sensing modalities. Another limitation of DBAD is that the sensing modality has to be explicitly provided to the framework, which is not practical in many cases since it is not clear which sensing modality works the best. In this work, we focus on building a generic framework that fuses multimodal sensors to identify subgroups in a homogeneous activity group. In other words, the same coarse-grained activity of all the people is provided to the framework as prior knowledge; the framework will divide these people into subgroups based on multiple sensing modalities automatically determined for the given coarse-grained activity. This is also different from the group detection problem studied by some existing work [3–6] as detailed in Section 2 which fuses some manually selected sensor features to group comoving people or devices.

Fine-grained partition of groups raises several interesting challenges.

Sensing Modality Selection. Existing work has shown that sensors on the users' mobile devices produce similar signals when the users have the same fine-grained activity [7]; therefore, group affiliation can be detected by monitoring the sensor signals of the mobile devices. However, with multiple sensing modalities available, it is not clear which sensing modalities can best capture users' activity similarity. It is even harder for a generic approach since it needs to detect group affiliation under any activity. We address this issue in Section 3.

Inconsistent Window Size among Multiple Sensing Modalities. To reduce cost (in particular in terms of energy consumption) of data collection and exchange to measure similarity between users, it is necessary to summarize the sensor data time series into aggregate sensor features. We choose to use probability distribution function (PDF) as the aggregate sensor feature [1]. The length of sensor data time series for summarization significantly impacts similarity measurement, so we need to determine the measurement time window for each sensing modality and deal with the different time window sizes when combining the measurements of multiple sensing modalities. We address this issue in both training phase (Section 3.3) and testing phase (Section 4.1).

Multimodal Clustering. Identifying groups based on the similarity measurements of multiple sensing modalities is nontrivial. Usually, we can apply clustering algorithms on the similarity graph of all users. However, since most sensing modalities are independent of each other, we cannot arbitrarily weigh each sensing modality to combine their similarity measurements into a single value. We address this issue in Sections 4.2 and 4.3.

The main contribution of this paper is that we propose approaches to address these challenges in a generic framework using two phases: phase I is sensing modality selection and phase II is multimodal clustering for group identification. The overall process is presented in Figure 1. We evaluate our approaches using both the dataset provided in DBAD and two datasets we collected. The evaluation results show that our multimodal-based approach outperforms the DBAD approach that uses only one sensing modality by about 10% in group affiliation accuracy. Even though 10% is not a large margin, a distinguishing feature of our approaches is that we can automatically select the right sensing modalities while the best sensing modality has to be explicitly provided to DBAD, which significantly limits its practicality. Further, our approaches work effectively for various activities.

2. Related Work

Group affiliation detection and group identification have been studied using sensor-equipped mobile devices such as smartphones. There exist several ways to identify groups, for instance, based on interactions [8], proximity [9], mobility [3–6], and activity [1, 7]. Most of the existing work relies on mobility for group detection, in which the individuals who have the similar trajectories are considered as in the same group. For example, GruMon [4] determines a group of individuals in a specific location who are traveling together in crowded urban environment. The solution fuses location data of different levels of accuracy using Bluetooth or WiFi with additional data such as semantic labels and smartphone sensor data, and the system shows very promising results based on tests using real-world datasets. In this paper, we focus on the activity-based group detection, in which the individuals who have similar activities are considered in the same group. For example, [7] identifies activity groups based on crowd behavior such as queueing, clogging, and group formation. The solution involves individual activity inference, pairwise activity relatedness, and global behavior inference. Different from the mobility-based group detection, tracking the location data of each individual over time is no longer a requirement. To be more specific, we define a homogeneous activity group as a group of people who perform the same coarse-grained activity at the same time and is one type of activity-based groups (people can have the same coarse-grained activity or different coarse-grained activities). We will use the term "activity" to represent a coarse-grained activity in the rest of the paper.

This work of identifying subgroups in a homogeneous activity group is inspired by DBAD [1]. The DBAD approach uses probability density functions (PDF) to model sensor data. Each mobile device computes the disparity to its neighbors by computing Jeffrey's divergence between the local PDF and the neighbors' PDF. The DBAD approach has several limitations. First, only one sensing modality is used at a time and this has to be selected manually. In particular, to identify people walking in different groups,

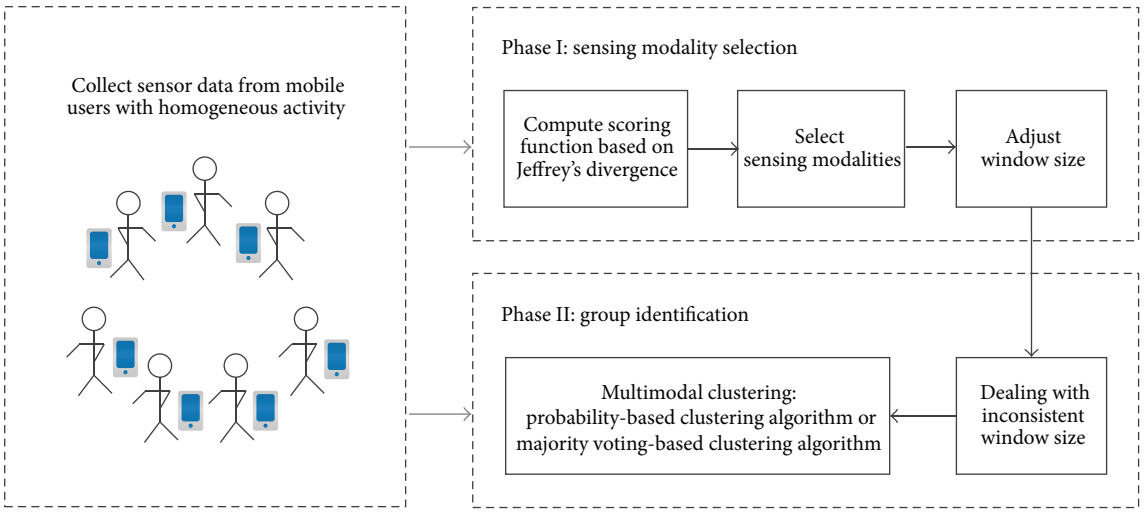


Figure 1: The overall process.

the magnitude of the accelerometer readings is manually selected to identify groups walking with different speeds, and the azimuth sensing modality obtained from the orientation sensor is manually selected to identify groups with different walking directions. However, using only the azimuth will not work when different groups of people walk in the same direction but with different speeds; using only the magnitude can not differentiate groups with different directions. Therefore, multimodal sensing is necessary to distinguish different groups without prior knowledge of the grouping details. Second, in DBAD experiments, wearable mobile devices are attached to the human body with fixed positions to reduce noise in sensor data collected. This is not practical since people may put their phones in pockets or hold them in hand. It is not clear how DBAD performs when noise is present in the collected data.

In activity recognition, the first stage is often sensing modality selection (i.e., feature construction). There are many existing approaches based on mobile devices [10]. In general, either based on some domain knowledge about the physical behavior involved or by making some default assumptions, a fixed set of sensing modalities is manually selected to construct the feature for a specific activity. Further, as discussed in [11], most activity recognition approaches are not generic and they often lead to solutions that are tied to the specific scenarios. Therefore, [11] proposes an algorithm which embeds feature construction into the machine learning process. However, this generic approach only works for the classification and regression problems and cannot be directly applied to the clustering problem we face in this work.

3. Phase I: Sensing Modality Selection

For different activities, different sets of sensing modalities may represent the most distinguishing features. The sensing modality selection process uses a training set for a given activity. The training set consists of one time series for each sensing modality on each mobile device. Each time series may have different sampling rate and may need to be summarized in different time windows. To select the sensing modalities which can provide accurate group affiliation detection results, we first define scoring function as a metric to find the best window size for a sensing modality and then determine whether the sensing modality is qualified for group affiliation detection.

Notations are listed at the end of the paper. The thresholds depend on the activities and sensing modalities. In this work, we determine the practical values of these thresholds using our datasets for various activities. We will determine the thresholds by activity as detailed in Section 6 in our future work.

3.1. Scoring Function. We use a probability-based approach to predict the group affiliation detection accuracy of a sensing modality m_k.

By summarizing m_k on each mobile device over a time window as a PDF, we can compute Jeffrey's divergence [13] (measures the disparity, opposite of similarity) between each device pair. Jeffrey's divergence between two probability distributions PDF_i and PDF_j is given by

$$\mathrm{DJ}\left(\mathrm{PDF}_i \parallel \mathrm{PDF}_j\right) = \int \left(\mathrm{PDF}_i\left(m_k\right) - \mathrm{PDF}_j\left(m_k\right)\right) \cdot \ln\left(\frac{\mathrm{PDF}_i\left(m_k\right)}{\mathrm{PDF}_j\left(m_k\right)}\right) d\left(m_k\right). \tag{1}$$

Scoring function $F(m_k)$ (2) is defined as the conditional probability of any pair of devices in the n devices' training set being in the same group when Jeffrey's divergence between them for sensing modality m_k is no larger than TH_s:

$$F\left(m_k\right) = P\left(G_{i,j} = 1 \mid \mathrm{DJ}\left(\mathrm{PDF}_i \parallel \mathrm{PDF}_j\right) \le \mathrm{TH}_s\right), \quad \forall i, j \in n,\ i \ne j, \tag{2}$$

where $G_{i,j} = 1$ indicates that i and j are affiliated with the same group while $G_{i,j} = -1$ indicates no group affiliation. As

discussed in [1], TH_s highly depends on the sensing modality being used and varies for different activities.

Using Bayes' theorem, (2) is derived as

$$F(m_k) = \frac{P\left(\mathrm{DJ}\left(\mathrm{PDF}_i \parallel \mathrm{PDF}_j\right) \le \mathrm{TH}_s \mid G_{i,j} = 1\right) \times P\left(G_{i,j} = 1\right)}{\sum_{v=\{1,-1\}} P\left(\mathrm{DJ}\left(\mathrm{PDF}_i \parallel \mathrm{PDF}_j\right) \le \mathrm{TH}_s \mid G_{i,j} = v\right) \times P\left(G_{i,j} = v\right)}. \quad (3)$$

The PDF of a sensing modality can be computed using Algorithm 1, assuming the distribution function type is known for the sensing modality. For example, most sensing modalities such as 3D acceleration and 3D rotation rate can be modeled as standard Gaussian distribution, and some sensing modalities such as orientation data have circular features and can be modeled as von Mises distribution [14]. If standard Gaussian is the distribution function type, the parameters are the mean μ and the variance σ^2 of a vector of numerical values in a time series. If von Mises is the distribution function type, the parameters are the circular mean $\mu(\theta)$ and the circular variance $\sigma(\theta)^2$ of a vector of angular values in a time series.

The computational cost of Jeffrey's divergence is related to the number of integration steps when calculating the integration in (1), and the integration steps can be determined based on the time series length l. Therefore, the time complexity of computing Jeffrey's divergence for a time series with length l is about $O(l)$.

3.2. Sensing Modality Selection. The sensing modality selection problem is stated as follows. Given n mobile devices or users in the training set, each has a set of time series S (contains one time series of the time stamped data for each sensing modality under a given activity A), and given the scoring function F to predict the group affiliation detection accuracy (i.e., the ratio of group affiliations that can be determined correctly), find the set of sensing modalities as well as the best window sizes which may result in an accuracy higher than decision threshold TH_d. Since a probability less than 0.5 means that the group affiliation detection has more chance to be incorrectly detected than correctly detected, TH_d should be larger than 0.5. Further, according to different activities, TH_d may vary in order to choose the most significant sensing modalities which have highest scores. The determination of TH_d and the most significant sensing modalities will be discussed in Section 5.

Algorithm 2 depicts how to select the candidate sensing modalities with their corresponding best window sizes which lead to the detection probability higher than TH_d. The time complexity depends on the number of sensing modalities (constant), the number of windows (constant), the number of mobile devices n, and Jeffrey's divergence computation complexity ($O(l)$). Therefore, the overall time complexity of sensing modality selection is $O(nl)$.

3.3. Adjusting Window Size. The sensing modality selection process identifies the best and a few secondary sensing modalities. The window size of each candidate sensing modality is compared against that of the best sensing modality. For any candidate sensing modality, if the new scoring function when using the window size of the best sensing modality is still not smaller than TH_d, the window size of this sensing modality will be modified to the same as that for the best sensing modality; otherwise, it keeps the original window size. The rationale behind this trick is to produce the multimodal fusion results mainly based on the best sensing modality and the results are expected to be improved by considering the secondary sensing modalities. The purpose of this window size matching is to reduce the processing of different window sizes during multimodal clustering in phase II.

Algorithm 3 depicts this process of adjusting window size. Similar to Algorithm 2, the time complexity of adjusting window size is $O(nl)$.

4. Phase II: Group Identification Using Multimodal Clustering

Once we have determined a set of candidate sensing modalities along with their window sizes, the next process is to use the test set to identify subgroups whose members have high similarity in these sensing modalities within a homogeneous activity group. Unlike the precollected training set, the test set can be recorded in real time and the sensor data distributions of all mobile devices can be periodically (i.e., according to the window sizes of the sensing modalities) sent to a central server in an infrastructure-based environment or collected by a sink node via data collection protocols in mobile ad hoc networks. Therefore, the group identification can also be done in real time in addition to using a precollected test set.

The multimodal sensor fusion-based group identification problem is actually the multimodal clustering problem, which has commonly been treated using early fusion or late fusion [15]. Early fusion combines the sensing modalities in a specific representation before the clustering process, while late fusion first applies the clustering process to each sensing modality separately and then combines the results from each sensing modality. According to the comparison in [16], the advantage of early fusion is that it requires one learning phase only, while the disadvantage is the difficulty to combine multiple sensing modalities in a common representation. Although late fusion avoids this issue, it has other drawbacks such as the expensiveness in learning since every sensing modality requires a separate learning phase and potential loss of correlation in multidimensional space. We believe that early fusion may outperform late fusion in certain scenarios, but not in others. Therefore, we investigate and compare two

Input: time series s, time series length l, window size w, distribution function type f
Output: series of mixture model parameters p
(1) **for** $i \in [0, l/w]$ **do**
(2) Use expectation maximization [12] to calculate the parameters of f for values $s[i \times w]$ to $s[(i+1) \times w - 1]$ in the vector of time series s;
(3) $p[i] \leftarrow \{parameters\}$;
(4) **end for**

Algorithm 1: Compute PDF.

Input: training set of time series $S_1, \ldots, S_n$ from n mobile devices under activity A, x sensing modalities in each set of time series, window size range $w_{\min}$ and $w_{\max}$ according to the sampling rate in the training set, scoring function F, decision threshold TH_d
Output: set of candidate sensing modalities C
(1) $C \leftarrow \emptyset$;
(2) $\mathrm{score}_{\mathrm{bestmodality}} \leftarrow 0$;
(3) $w_{\mathrm{bestmodality}} \leftarrow 0$;
(4) **for** $k \in [1, x]$ **do**
(5) $m_k.\mathrm{index} \leftarrow k$;
(6) $m_k.\mathrm{score}_{\mathrm{best}} \leftarrow 0$;
(7) $m_k.w_{\mathrm{best}} \leftarrow w_{\min}$;
(8) **for** $w \in [w_{\min}, w_{\max}]$ **do**
(9) **for** $i \in [0, n)$ **do**
(10) $\mathrm{PDF}[i] \leftarrow \mathrm{ComputePDF}(s_i \leftarrow S_i[k], w)$;
(11) **end for**
(12) **if** $F(m_k) \geq m_k.\mathrm{score}_{\mathrm{best}}$ **then**
(13) $m_k.\mathrm{score}_{\mathrm{best}} \leftarrow F(m_k)$;
(14) $m_k.w_{\mathrm{best}} \leftarrow w$;
(15) **end if**
(16) **end for**
(17) **if** $m_k.\mathrm{score}_{\mathrm{best}} \geq \mathrm{TH}_d$ **then**
(18) $C \leftarrow C \cup \{m_k\}$;
(19) **if** $m_k.\mathrm{score}_{\mathrm{best}} \geq \mathrm{score}_{\mathrm{bestmodality}}$ **then**
(20) $\mathrm{score}_{\mathrm{bestmodality}} \leftarrow m_k.\mathrm{score}_{\mathrm{best}}$;
(21) $w_{\mathrm{bestmodality}} \leftarrow m_k.w_{\mathrm{best}}$;
(22) **end if**
(23) **end if**
(24) **end for**

Algorithm 2: Select sensing modalities.

clustering approaches, probability-based clustering for early fusion and majority voting-based clustering for late fusion.

Before we discuss the two clustering algorithms, we need to explain how to deal with different window sizes among different sensing modalities selected.

4.1. Dealing with Inconsistent Window Size. We use the window size of the best sensing modality for group identification, so the best sensing modality delivers one pairwise group affiliation result in each time window of group identification, and the secondary sensing modalities deliver multiple or no results in such a time window. Figure 2 shows an example with time series of three candidate sensing modalities provided by a mobile device, where s_1 is for the best sensing modality m_1 and the window size w_1 of m_1 is used as the group identification time window. The window size of each sensing modality is the same on all mobile devices. Therefore, by collecting the information of all sensing modalities on all mobile devices, m_1 delivers one pairwise group affiliation result in each of the w_1 windows, m_2 (corresponding to s_2) delivers one or no result, and m_3 (corresponding to s_3) delivers one or multiple results.

To determine pairwise group affiliation between a pair of mobile devices i and j, Jeffrey's divergence is compared against threshold TH_s: if $\mathrm{DJ}(\mathrm{PDF}_i \parallel \mathrm{PDF}_j) \leq \mathrm{TH}_s$, then use the temporary result $v = 1$ to indicate positive group affiliation; otherwise, use $v = -1$ to indicate no group affiliation. Moreover, since the sensing modality m_k may deliver multiple results or no result in the group identification time window w_1, we define the aggregated result delivered by m_k in each w_1 window as $r_{m_k} \in \{1, 0, -1\}$, indicating whether the sum of v during the window is positive, zero, or negative.

Input: training set of time series $S_1, \ldots, S_n$ from n mobile devices under activity A, scoring function F, decision threshold TH_d, set of candidate sensing modalities C
Output: C with adjusted window sizes
(1) **for** $m_c \in C$ **do**
(2) **if** $m_c.\mathrm{score}_{\mathrm{best}} < \mathrm{score}_{\mathrm{bestmodality}}$ **then**
(3) **for** $i \in [0, n)$ **do**
(4) $\mathrm{PDF}_i \leftarrow$
(5) $\mathrm{ComputePDF}(S_i[m_c.\mathrm{index}], w_{\mathrm{bestmodality}})$;
(6) **end for**
(7) **if** $F(m_c) \geq \mathrm{TH}_d$ **then**
(8) $m_c.\mathrm{score}_{\mathrm{best}} \leftarrow F(m_c)$;
(9) $m_c.w_{\mathrm{best}} \leftarrow w_{\mathrm{bestmodality}}$;
(10) **end if**
(11) **end if**
(12) **end for**

Algorithm 3: Adjust window size.

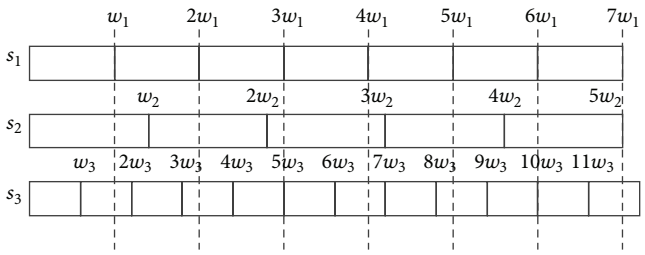


Figure 2: Example time series with different window sizes.

This is because positive summation implies that most of the time positive group affiliation is suggested and vice versa. The aggregated result 0 may be caused by no result delivered in this time window or multiple results canceling out each other. In this case, the impact of m_k on group identification does not need to be considered. Therefore, sensing modality m_k is taken into account in a group identification time window only when it provides an aggregated result 1 or −1.

4.2. Early Fusion: Probability-Based Clustering. We present an early fusion multimodal clustering approach which combines the pairwise group affiliation results delivered by all sensing modalities in each group identification time window into a single result. A common approach for early fusion is to assign weights to each sensing modality. However, it is difficult to determine the appropriate weights, either manually or using a search procedure. Moreover, we have sensing modalities which deliver the pairwise group affiliation results with different accuracies. Intuitively, the best sensing modality should be given the highest weight in the early fusion process. If we assign a percentage as the weight to each of the sensing modalities and then sum them up, the fusion function has no physical meaning and it is even more confusing than using only the best sensing modality. On the other hand, as discussed in Section 2, using a single sensing modality without prior knowledge of grouping details is insufficient for many scenarios such as different groups of people walking in the same direction but with different speeds. Therefore, instead of using a single sensing modality or arbitrarily providing weights to different sensing modalities, we use the joint probability of correct pairwise group affiliation detection as a fusion method to combine the pairwise group affiliation results delivered by all the selected sensing modalities.

In a group identification time window, given a set of sensing modalities $\{m_1, \ldots, m_z\}$, each delivers a pairwise group affiliation result $r_{m_y} \in \{1, -1\}$, where $y \in \{1, \ldots, z\}$. The probability of correct pairwise group affiliation detection (i.e., the fusion function) is calculated as shown in what follows using Bayes' theorem:

$$P\left(G_{i,j} = 1 \mid r_{m_1}, \ldots, r_{m_z}\right) = \frac{P\left(r_{m_1}, \ldots, r_{m_z} \mid G_{i,j} = 1\right) \times P\left(G_{i,j} = 1\right)}{\sum_{v=\{1,-1\}} P\left(r_{m_1}, \ldots, r_{m_z} \mid G_{i,j} = v\right) \times P\left(G_{i,j} = v\right)}. \quad (4)$$

Further, we assume that each sensing modality can deliver a pairwise group affiliation result independently, so we can rewrite (4) as

$$P\left(G_{i,j} = 1 \mid r_{m_1}, \ldots, r_{m_z}\right) = \frac{\left(\prod_{y=1}^{x} P\left(r_{m_y} \mid G_{i,j} = 1\right)\right) \times P\left(G_{i,j} = 1\right)}{\sum_{v=\{1,-1\}} \left(\prod_{y=1}^{z} P\left(r_{m_y} \mid G_{i,j} = v\right)\right) \times P\left(G_{i,j} = v\right)}, \quad (5)$$

where the probabilities $P(r_{m_y} \mid G_{i,j} = v)$ and $P(G_{i,j} = v)$ are computed in the same way as the calculations in Section 3.1 using the training set. These precomputed probability values can be directly applied to the clustering algorithm in which the test set is being used for group identification.

Using the test set, we can compute the pairwise group affiliation probabilities $P(G_{i,j} = 1 \mid r_{m_1}, \ldots, r_{m_z})$ in each group identification time window. We use a probability threshold TH_p to convert the pairwise group affiliation probabilities into a binary matrix $\mathbf{V}$ of the fused pairwise

Input: test set of time series $S_1, \ldots, S_n$ on n mobile devices under activity A, x selected sensing modalities in each set of time series, probability threshold TH_p
Output: device groups in each group identification time window
(1) Each mobile device uses its local time series to compute the PDFs for each selected sensing modality according to its window size;
(2) The server or sink node collects the PDFs from all the n mobile devices once in each group identification time window and run the following process:
(3) Initialize group affiliation matrix $\mathbf{V}$;
(4) **for** each device pair (i, j) **do**
(5) $M \leftarrow \emptyset$;
(6) **for** $k \in \{1, \ldots, x\}$ **do**
(7) Compute r_{m_k};
(8) **if** $r_{m_k} \neq 0$ **then**
(9) $M \leftarrow M \cup \{(k, r_{m_k})\}$;
(10) **end if**
(11) **end for**
(12) Compute $p \leftarrow P(G_{i,j} = 1 \mid \forall r_{m_k} \in M)$;
(13) **if** $p \geq \mathrm{TH}_p$ **then**
(14) $V_{i,j} \leftarrow 1$;
(15) **else**
(16) $V_{i,j} \leftarrow -1$;
(17) **end if**
(18) **end for**
(19) Apply DJ-Cluster algorithm on matrix $\mathbf{V}$;

Algorithm 4: Probability-based clustering algorithm.

group affiliation results. The value corresponding to the mobile devices i and j in the matrix $\mathbf{V}$ is denoted as $V_{i,j} \in \{1, -1\}$. If $P(G_{i,j} = 1 \mid r_{m_1}, \ldots, r_{m_z}) \geq \mathrm{TH}_p$, then $V_{i,j} = 1$; otherwise $V_{i,j} = -1$. TH_p may also vary for different activities, and its determination will be discussed in Section 5.

Based on the group affiliation matrix, we can use existing clustering algorithms in one-dimensional space. We apply the density joint clustering algorithm (DJ-Cluster) [17] which is used by existing work of pedestrian flocks detection [3] to cluster the mobile devices into different groups.

The process of the probability-based clustering approach is given in Algorithm 4. Note that a sensing modality m_k is taken into account in computing the fused pairwise group affiliation result only when it provides the result $r_{m_k} \neq 0$. The time complexity depends on the number of device pairs (n^2), the number of selected sensing modalities (constant), computation of r_{m_k} (the complexity is the same as computing Jeffrey's divergence, i.e., $O(l)$), and DJ-Cluster algorithm ($O(n^2)$). Therefore, the overall time complexity of the probability-based clustering algorithm is $O(n^2 l)$.

4.3. Late Fusion: Majority Voting-Based Clustering. We present a late fusion multimodal clustering approach which combines the clusters generated by each sensing modality in each group identification time window. We first use the DJ-Cluster algorithm to generate the clusters for each sensing modality separately. Similar to Algorithm 4, a sensing modality m_k is taken into account in the final cluster determination for two mobile devices only when it provides the result $r_{m_k} \neq 0$. We modify the majority voting approach used in [3], where the fusion is calculating the summed weight of the sensing modalities where a pair of mobile devices are clustered into the same group. The two mobile devices are added as a cluster in the majority solution if the summed weight is larger than 50%. If one of the them is already inside a solution cluster, the other one joins the same cluster instead of adding a new cluster. However, in [3], it simply assigns a weight of 50% to the features which may give the best accuracy and then divide the remaining 50% among the other features. It does not search for the best weights assignment or automatic training of these weights. Therefore, the weight assignment is still a problem in this late fusion multimodal clustering approach. Since we already have a sensing modality selection process before the clustering process, as long as the sensing modalities are well selected, all the selected sensing modalities should play important roles in the group identification. Therefore, we apply the same weight on all selected sensing modalities.

Algorithm 5 gives the process of the majority voting-based clustering approach. Similar to Algorithm 4, the time complexity of separate clustering for all the selected sensing modalities is $O(n^2 l)$. Further, the time complexity of applying majority voting on all device pairs is $O(n^2)$. Therefore, the overall time complexity of the majority voting-based clustering algorithm is $O(n^2 l)$, which is the same as the probability-based clustering algorithm.

Complexity Comparison with DBAD. The DBAD approach computes pairwise group affiliations on each device. The complexity of computing a pairwise group affiliation is basically Jeffrey's divergence computation ($O(l)$). Each device

Input: test set of time series $S_1, \ldots, S_n$ on n mobile devices under activity A, x selected sensing modalities in each set of time series
Output: device groups in each group identification time window
(1) **for** each device pair (i, j) **do**
(2) $M_{i,j} \leftarrow \emptyset$;
(3) **end for**
(4) **for** $k \in \{1, \ldots, x\}$ **do**
(5) Initialize group affiliation matrix **V**;
(6) **for** each device pair (i, j) **do**
(7) Compute r_{m_k};
(8) **if** $r_{m_k} \neq 0$ **then**
(9) $M_{i,j} \leftarrow M_{i,j} \cup \{k\}$;
(10) $V_{i,j} \leftarrow r_{m_k}$;
(11) **else**
(12) $V_{i,j} \leftarrow -1$;
(13) **end if**
(14) **end for**
(15) Apply DJ-Cluster algorithm on matrix **V**;
(16) **end for**
(17) **for** each device pair (i, j) **do**
(18) Apply majority voting to the clusters generated by the sensing modalities in $M_{i,j}$;
(19) **end for**

Algorithm 5: Majority voting-based clustering algorithm.

needs to compute $n - 1$ pairwise group affiliations against other devices. Therefore, the overall time complexity of DBAD is $O(nl)$. In our approach, we not only compute the pairwise group affiliations, but also identify the group partitions. Therefore, our approach needs to compute Jeffrey's divergence ($O(l)$) over all pairs of devices (n^2), leading to the overall time complexity ($O(n^2 l)$). The complexity added in our approach is necessary to solve the group partition problem.

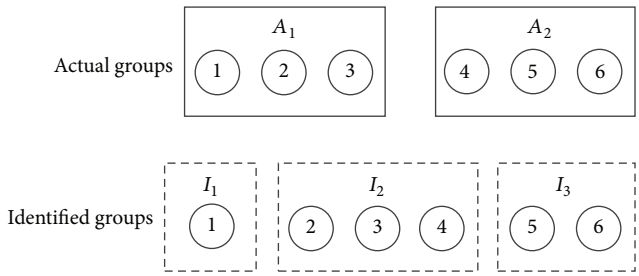


Figure 3: Sample group identification result.

5. Performance Evaluation

5.1. Performance Metrics. Since the DBAD approach only detects pairwise group affiliation, its evaluation only considers the accuracy of pairwise group affiliation detection results. In contrast, our final results are the identified groups; therefore we use the performance metrics pairwise group affiliation accuracy and group membership similarity to evaluate the intermediate and the final results, respectively. For group identification, since the groups are preconfigured and unchanged during an experiment, we determine the final groups when the grouping results are stable; that is, groups remain for at least five group identification time windows. The group membership similarity is calculated as the average Jaccard similarity [18] between an identified group and the corresponding actual group. The pairwise group affiliation accuracy is calculated as ratio of the correctly determined group relationships over the total number of pairwise group relationships when the final groups are identified.

Figure 3 shows a sample result of group identification comparing to the actual groups. We first match each identified group to an actual group which has the most common members, so I_1 is matched to A_1, I_2 to A_1, and I_3 to A_2. Then, the Jaccard similarity is 1/3 between I_1 and A_1, 2/4 between I_2 and A_1, and 2/3 between I_3 and A_2. Therefore, the group membership similarity is 0.5 (the average Jaccard similarity). In the meantime, there are $C_6^2 = 15$ pairwise group relationships in total, but only 9 pairs (with or without group affiliation) are determined correctly, that is, $(1, 4)$, $(1, 5)$, $(1, 6)$, $(2, 3)$, $(2, 5)$, $(2, 6)$, $(3, 5)$, $(3, 6)$, and $(5, 6)$. Therefore, the pairwise group affiliation accuracy is $9/15 = 0.6$.

5.2. Datasets. In performance evaluation, we first use the dataset provided in DBAD [1] where the activity is people walking together. The DBAD dataset contains the sensor data obtained from 10 homogeneous Android devices which are attached to the hip of each person. The experiments are conducted with different group configurations (from 1 to 10 groups), and each experiment lasts 51 minutes. The sampling rate is about 25 Hz for each sensor. To compute the activity similarity for people walking together, we consider the following sensing modalities available in the dataset: x-acceleration, y-acceleration, z-acceleration, and magnitude

(obtained from the 3D accelerometer); azimuth, pitch, and roll (obtained from the orientation sensor). The magnitude is the square root of the square sum of the 3D accelerations, and the DBAD evaluation uses it instead of the 3D acceleration measurements. There are two limitations of the DBAD dataset as discussed in Section 2: one is that wearable mobile devices are attached to the human body with fixed positions in order to reduce noise in the collected sensor data; the other is that there is only one activity (i.e., people walking together) involved. Therefore, we also collect our own datasets—one for the park scenario and one for the game scenario as discussed in Section 1.

The park scenario has the same activity with the DBAD dataset and uses the same sampling rate, but with less controlled phone positions to allow for more noisy data and with more sensing modalities to allow for consideration of multiple modalities. Since the DBAD dataset only contains accelerometer and orientation sensor, we collect our own dataset with more motion sensors on smartphones for the same activity in which people walk together. It contains the sensor data obtained from 8 heterogeneous smartphones (e.g., Nexus and Samsung Galaxy phones) held in hands by people walking in 3 groups for about 10 minutes. These groups have different walking directions and are slightly different in walking speed. The sensors recorded are 3D accelerometer, 3D gyroscope, and orientation sensor. We consider the following sensing modalities: x-acceleration, y-acceleration, and z-acceleration (obtained from the 3D accelerometer); x-rotation, y-rotation, and z-rotation (obtained from the 3D gyroscope); azimuth, pitch, and roll (obtained from the orientation sensor).

The game scenario has a different activity (i.e., audience wave hands for different teams) from the DBAD dataset and it is used to demonstrate that our approaches are general and can handle different activities. The sampling rate is also the same. This dataset contains the sensor data obtained from 8 heterogeneous smartphones for about 10 minutes. Each group waves their smartphones in different time periods, mimicking the activity that audience cheer for the two competitor teams in a game. The sensors recorded are the same as in the park scenario dataset.

For each dataset, we divide it into two parts—the first half as the training set for sensing modality selection and the second half as the test set for identification of subgroups within a homogeneous activity group. We implement our algorithms in Python and run Algorithms 2 and 3 on the training set and Algorithms 4 and 5 on the test set.

5.3. Experimental Results

5.3.1. Results Using the DBAD Dataset. In the training set, we set the minimum and maximum window sizes as 5 seconds and 50 seconds, respectively. The minimum window size is set according to the sampling rate 25 Hz, so we can have more than 100 samples within each window to compute the PDF. The maximum window size cannot be too large (within a minute); otherwise it takes too long to make the grouping decision. Table 1 shows the results for each sensing modality,

Table 1: Sensing modality selection using DBAD dataset.

Sensing modality	Best window size	Best score	New score
x-acceleration	15 s	0.65	0.5
y-acceleration	15 s	0.64	0.5
z-acceleration	15 s	0.55	0.49
Magnitude	15 s	0.58	0.5
Azimuth	**5 s**	**0.75**	**0.75**
Pitch	5 s	0.45	0.45
Roll	5 s	0.48	0.48

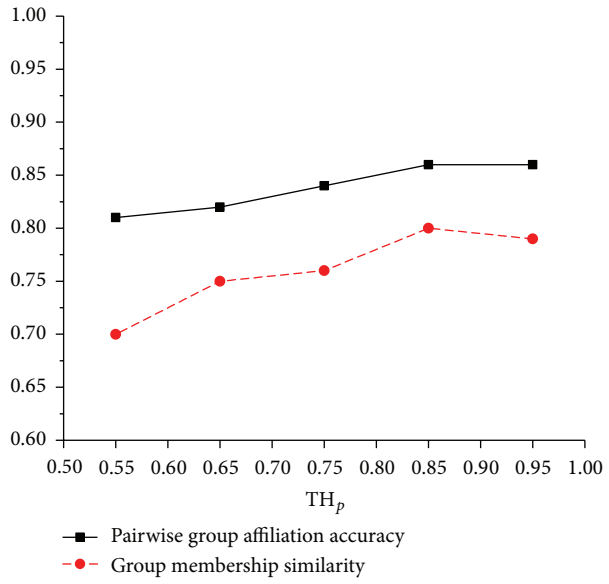
where the best score is the scoring function with the best window size for that sensing modality and the new score is the recalculated scoring function using the best sensing modality's best window size.

As discussed in Section 3.2, the decision threshold TH_d should be larger than 0.5. Here we set $\mathrm{TH}_d = 0.55$; then the azimuth (window size 5 s), x-acceleration (window size 15 s), y-acceleration (window size 15 s), z-acceleration (window size 15 s), and magnitude (window size 15 s) are selected. Since magnitude is a redundant sensing modality to the 3D acceleration and it yields very similar score as the 3D acceleration, we use the 3D acceleration sensing modalities in Algorithms 4 and 5 instead of magnitude. We next use the test set to evaluate Algorithms 4 and 5.

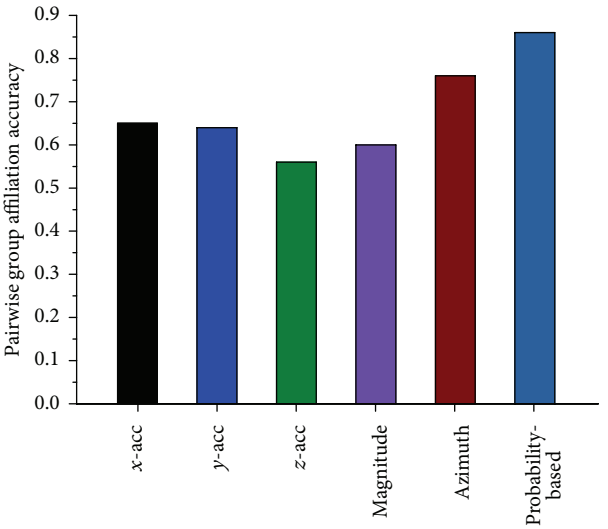
First, we consider the probability threshold TH_p in Algorithm 4. Similar to the decision threshold TH_d, it should also be larger than 0.5. Therefore, we vary it from 0.55 to 0.95. Figure 4(a) shows that the group membership similarity is slightly smaller than the pairwise group affiliation accuracy. This is because there exist some critical links in the graph-based clustering algorithms. If a critical link is determined with incorrect group affiliation result, it will significantly impact the group identification results. In general, the pairwise group affiliation accuracy increases when TH_p increases. Using the DBAD dataset, $\mathrm{TH}_p = 0.85$ leads to both the highest pairwise group affiliation accuracy and the highest group membership similarity. Next, we will compare the results of the probability-based clustering algorithm using $\mathrm{TH}_p = 0.85$ with the results of using the DJ-Cluster algorithm on each single sensing modality as well as using the majority voting-based clustering algorithm on all sensing modalities.

Figure 4(b) shows the pairwise group affiliation accuracy and Figure 4(c) shows the group membership similarity. We put the results of different sensing modalities together with the results of different approaches in order to compare not only the approaches but also multimodal against each individual sensing modality. Also note that, since the majority voting-based clustering algorithm outputs the final clusters based on the clusters computed from each sensing modality, it does not output the combined pairwise group affiliation results of all sensing modalities; we only compare the probability-based approach with each single sensing modality for the pairwise group affiliation accuracy.

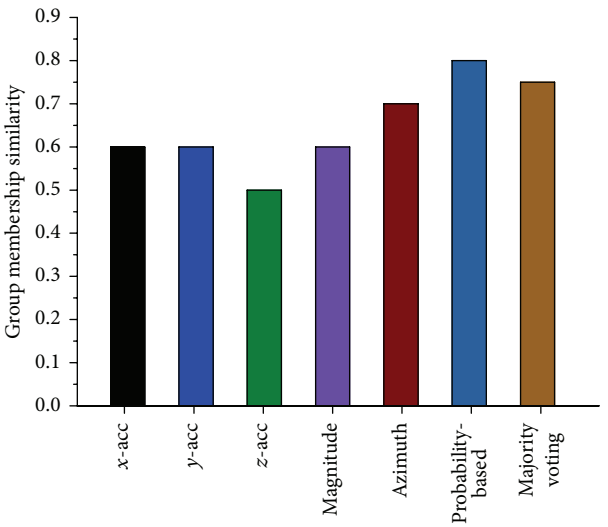
In Figure 4(b), the 3D acceleration sensing modalities lead to an accuracy around 0.6 while the azimuth related to the orientation sensor leads to an accuracy about 0.76.



(a) Impact of TH$_p$



(b) Pairwise group affiliation accuracy



(c) Group membership similarity

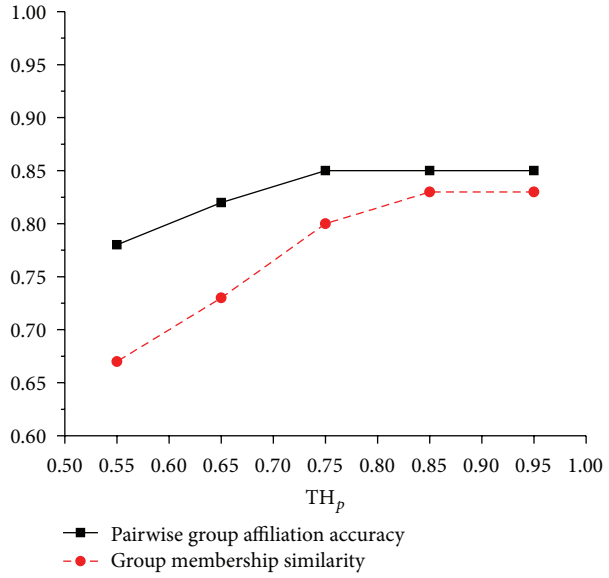
Figure 4: Results using DBAD dataset.

These results are consistent with the findings in the DBAD approach, where the azimuth delivers the best pairwise group affiliation accuracy. Beyond their findings, our sensing modality selection approach automatically selects the azimuth as the most significant sensing modality. Further, the probability-based approach leads to an accuracy about 0.86, which shows that the multimodal-based approach outperforms the original DBAD approach which uses a single sensing modality.

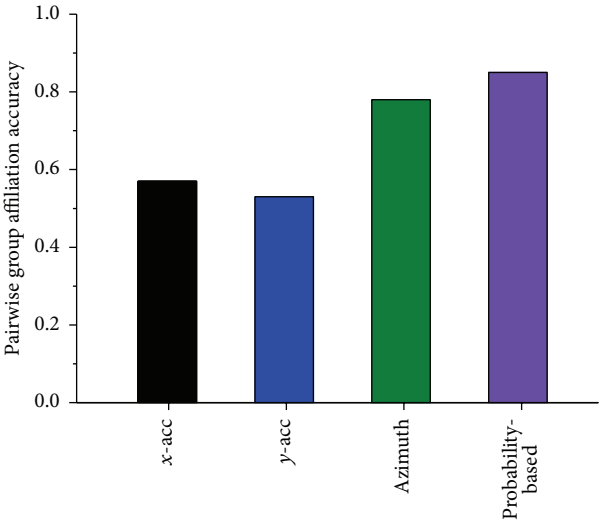
In Figure 4(c), the comparisons are similar to Figure 4(b). In addition, the probability-based approach outperforms the majority voting-based approach using the DBAD dataset. This is because the sensing modalities other than azimuth do not have high scores, so their contributions in the majority voting-based approach are not significant. However, the majority voting-based approach still provides a higher group membership similarity than using the 3D acceleration or the azimuth separately.

5.3.2. Results Using the Park Scenario Dataset. We use the same minimum/maximum window sizes as in the DBAD training set. Table 2 shows the results, where the azimuth also leads to the best score as in Table 1.

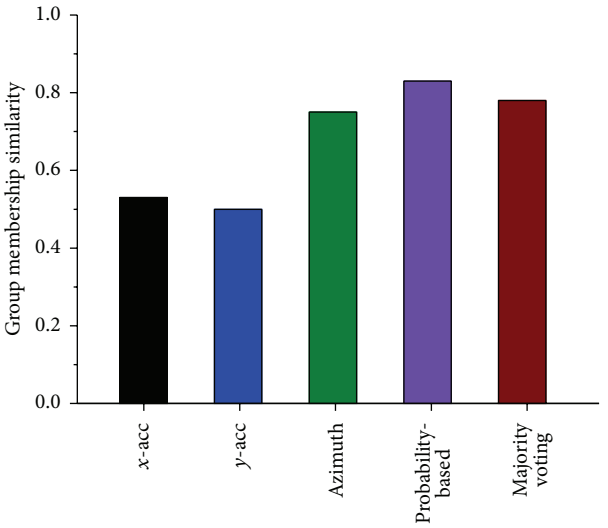
We also choose the decision threshold $\text{TH}_d = 0.55$, so the azimuth (window size 5 s), x-acceleration (window size 15 s), and y-acceleration (window size 15 s) are the selected sensing modalities. Although z-acceleration is not selected here, it does not contribute significant results for DBAD dataset either. Figure 5(a) shows the results of the probability-based approach when we vary the probability threshold TH$_p$



(a) Impact of TH$_p$



(b) Pairwise group affiliation accuracy



(c) Group membership similarity

Figure 5: Results using park scenario dataset.

Table 2: Sensing modality selection using park scenario dataset.

Sensing modality	Best window size	Best score	New score
x-acceleration	15 s	0.58	0.45
y-acceleration	15 s	0.55	0.45
z-acceleration	15 s	0.51	0.4
x-rotation	15 s	0.42	0.4
y-rotation	15 s	0.35	0.33
z-rotation	15 s	0.35	0.33
Azimuth	**5 s**	**0.78**	**0.78**
Pitch	5 s	0.4	0.4
Roll	5 s	0.4	0.4

from 0.55 to 0.95. Similar to the findings in the DBAD test set, the group membership similarity is slightly lower than the pairwise group affiliation accuracy, and the pairwise group affiliation accuracy increases when TH$_p$ increases. We choose TH$_p$ = 0.85 for the probability-based approach in the following comparisons using the test set.

Figure 5(b) compares the pairwise group affiliation accuracy results. Similar to Figure 4(b), the azimuth leads to a higher accuracy than the 3D acceleration, and the probability-based approach leads to an even higher accuracy than the azimuth. Figure 5(c) compares the group membership similarity results. The comparison is consistent with that of the pairwise group affiliation accuracy. In addition, the majority voting-based approach leads to a lower group membership similarity than the probability-based approach, but the similarity is still higher than using the x-acceleration, y-acceleration, or azimuth individually. All these results

again verify that the multimodal-based approaches outperform the original DBAD approach that works with a single sensing modality. Further, unlike the controlled experiments with homogeneous phones and fixed phone positions in DBAD, our experiments are less controlled and have more uncertainty in the collected sensor data. Despite all these, the results using our dataset are still promising (e.g., the group membership similarity for the probability-based approach is still above 0.8), indicating that our approaches can inherently deal with sensor data noises. This is because sensing modalities are selected in the presence of data noises.

Moreover, the results using the park scenario dataset are consistent with those using the DBAD dataset because of the same activity involved. This indicates that the same training set for the same activity may be used to test both the datasets if the training set is well collected and the parameters involved in the algorithms are well studied.

5.3.3. Results Using the Game Scenario Dataset. Table 3 shows the results of sensing modality selection. Different from Tables 1 and 2, the 3D rotations lead to the highest scores. The 3D accelerations may still work, but the azimuth does not make much sense in this activity. This implies that the DBAD approach of manually selecting one single sensing modality will not work in such a scenario.

We can still choose the decision threshold $TH_d = 0.55$, so the x-acceleration, y-acceleration, z-acceleration, x-rotation, y-rotation, and z-rotation are selected. Figure 6(a) shows the results of the probability-based approach. Similar to the findings in both the DBAD test set and the park scenario test set, we can choose $TH_p = 0.95$ for the probability-based approach to compare with using each single sensing modality as well as the majority voting-based approach.

Figure 6(b) shows that the y-rotation leads to a higher accuracy than each other sensing modality, and the probability-based approach leads to even higher accuracy than using only y-rotation. Figure 6(c) shows a consistent trend as in Figure 6(b). However, different from both Figures 4(c) and 5(c), the majority voting-based approach leads to a slightly higher group membership similarity than the probability-based approach. This is because there are several significant sensing modalities (i.e., x-rotation, y-rotation, and z-rotation) which contribute accurate results in this activity. Unlike the activity that people walk together, only the azimuth makes significant contribution in the final results of the multimodal-based approaches; here all the 3D rotations make significant contributions; therefore the majority voting is more significant.

In summary, the activity significantly impacts the sensing modality selection as well as the group identification results. This verifies our hypothesis in Section 3 that a selection process is needed to automatically select sensing modalities for different activities. In addition, the comparison of the probability-based approach and the majority voting-based approach verifies our hypothesis in Section 4 that early fusion multimodal clustering may outperform late fusion in some activities, but not always. All things considered that all the approaches proposed in this work (i.e., Algorithms 2, 3, 4, and 5) are effective for various activities.

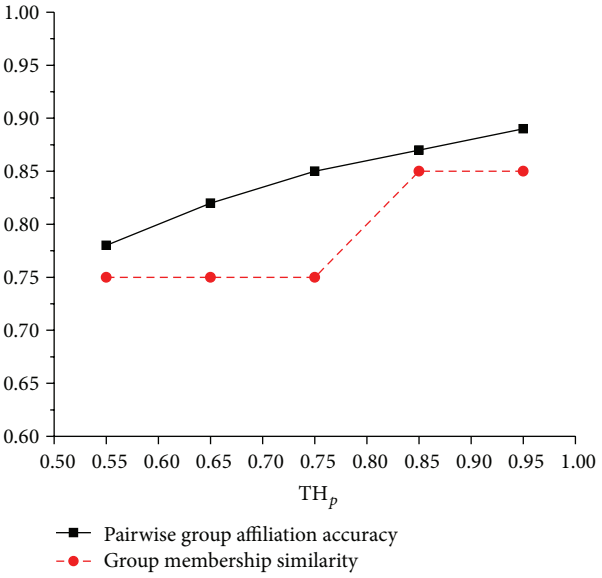
Table 3: Sensing modality selection using game scenario dataset.

Sensing modality	Best window size	Best score	New score
x-acceleration	15 s	0.66	0.66
y-acceleration	15 s	0.65	0.65
z-acceleration	15 s	0.58	0.58
x-rotation	15 s	0.75	0.75
y-rotation	**15 s**	**0.8**	**0.8**
z-rotation	15 s	0.72	0.72
Azimuth	5 s	0.54	0.51
Pitch	5 s	0.52	0.5
Roll	5 s	0.46	0.45

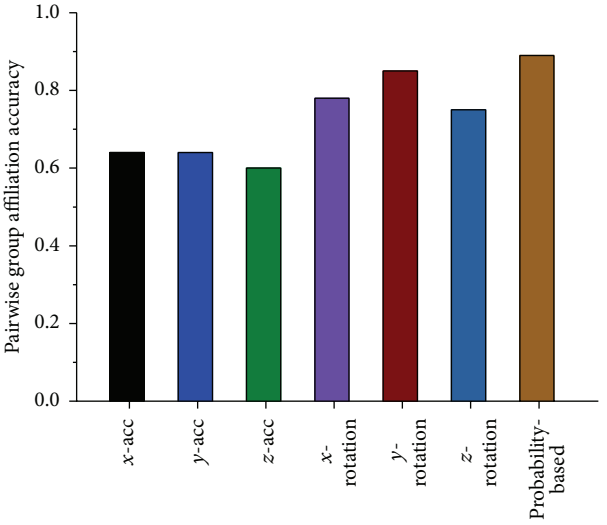
6. Conclusion

In this paper, we have presented a generic framework to identify subgroups in a homogeneous activity group using sensor-equipped mobile devices. We have first proposed a sensing modality selection approach given a coarse-grained activity. We have then provided an approach to deal with multiple window sizes among all the selected sensing modalities. By setting the group identification window size the same as that of the best sensing modality, we have further developed two multimodal clustering approaches—probability-based approach for early fusion and majority voting-based approach for late fusion. Finally, we have evaluated our approaches using a publicly available dataset and also two others collected by ourselves. The evaluation results have shown that our framework of multimodal approaches outperforms the original DBAD approach which works on a single sensing modality, and the framework is effective for various activities.

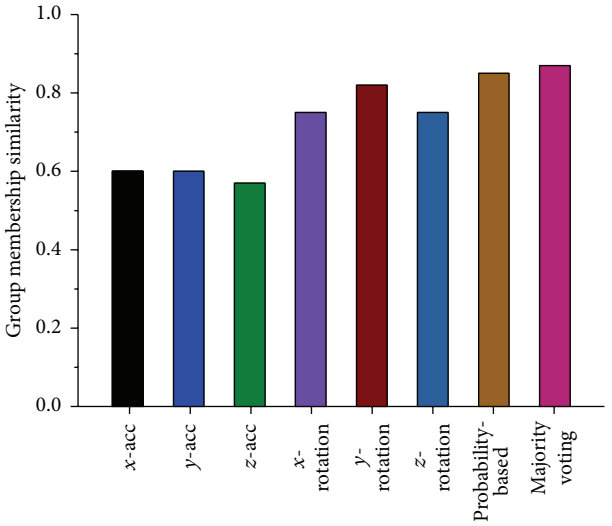
Several improvements are considered for future work. First, in this framework, activity is considered as an input to the algorithms. Although we have not yet studied the sensing modality selection training per activity, our evaluation results of different datasets but with the same activity tend to be very similar, indicating that using the same training set for an activity and test on different datasets regarding this activity is possible. Second, in this work, we assume that the sensor data distributions of all mobile devices are periodically sent to a central server in an infrastructure-based environment or collected by a sink node via data collection protocols in mobile ad hoc networks. Therefore, the central server or the sink node has the complete information in the network to calculate pairwise similarities and apply clustering algorithms on the group affiliation matrix based on the pairwise similarities. In our future work, we will further consider a pure peer-to-peer environment where neighboring mobile devices exchange their sensor data distributions. Since some pairwise similarities between multihop neighbors may not be computed due to limited hops of data exchange, the clustering algorithms need to be revised accordingly to work with a local partial group affiliation matrix on each mobile device. Last, we will apply Jeffrey's divergence directly to multiple sensing modalities when a practical mathematical method is available.



(a) Impact of TH$_p$



(b) Pairwise group affiliation accuracy



(c) Group membership similarity

Figure 6: Results using game scenario dataset.

Notations

m: Sensing modality
w: Window size
F: Scoring function
TH$_s$: Jeffrey's divergence threshold (varies by modality and activity)
TH$_d$: Sensing modality decision threshold (varies by activity)
TH$_p$: Group probability threshold with multiple sensing modalities.

Competing Interests

The authors declare that they have no competing interests.

Acknowledgments

This project is supported in part by NSF Grant CNS-0915574 and National Natural Science Foundation of China-Guangdong Government Joint Funding (2nd) for Super Computer Application Research.

References

[1] D. Gordon, M. Wirz, D. Roggen, G. Tröster, and M. Beigl, "Group affiliation detection using model divergence for wearable devices," in *Proceedings of the ACM International Symposium (ISWC '14)*, pp. 19–26, Seattle, Wash, USA, September 2014.

[2] D. Gordon, J.-H. Hanne, M. Berchtold, A. A. N. Shirehjini, and M. Beigl, "Towards collaborative group activity recognition

using mobile devices," *Mobile Networks and Applications*, vol. 18, no. 3, pp. 326–340, 2013.

[3] M. B. Kjærgaard, M. Wirz, D. Roggen, and G. Tröster, "Detecting pedestrian flocks by fusion of multi-modal sensors in mobile phones," in *Proceedings of the 14th International Conference on Ubiquitous Computing (UbiComp '12)*, pp. 240–249, Pittsburgh, Pa, USA, September 2012.

[4] R. Sen, Y. Lee, K. Jayarajah, A. Misra, and R. K. Balan, "GruMon: fast and accurate group monitoring for heterogeneous urban spaces," in *Proceedings of the 12th ACM Conference on Embedded Networked Sensor Systems (SenSys '14)*, pp. 46–60, Memphis, Tenn, USA, November 2014.

[5] A. Srivastava, J. Gummeson, M. Baker, and K. Kim, "Step-by-step detection of personally collocated mobile devices," in *Proceedings of the 16th International Workshop on Mobile Computing Systems and Applications (HotMobile '15)*, pp. 93–98, Santa Fe, NM, USA, February 2015.

[6] M. B. Kjaergaard, M. Wirz, D. Roggen, and G. Troster, "Mobile sensing of pedestrian flocks in indoor environments using WiFi signals," in *Proceedings of the 10th IEEE International Conference on Pervasive Computing and Communications (PerCom '12)*, pp. 95–102, Lugano, Switzerland, March 2012.

[7] D. Roggen, M. Wirz, G. Tröster, and D. Helbing, "Recognition of crowd behavior from mobile sensors with pattern analysis and graph clustering methods," *Networks and Heterogeneous Media*, vol. 6, no. 3, pp. 521–544, 2011.

[8] B. Guo, H. He, Z. Yu, D. Zhang, and X. Zhou, "GroupMe: supporting group formation with mobile sensing and social graph mining," in *Mobile and Ubiquitous Systems: Computing, Networking, and Services*, vol. 120, pp. 200–211, Springer, Berlin, Germany, 2013.

[9] N. Yu and Q. Han, "Grace: recognition of proximity-based intentional groups using collaborative mobile devices," in *Proceedings of the 11th IEEE International Conference on Mobile Ad Hoc and Sensor Systems (MASS '14)*, pp. 10–18, Philadelphia, Pa, USA, October 2014.

[10] Ó. D. Lara and M. A. Labrador, "A survey on human activity recognition using wearable sensors," *IEEE Communications Surveys and Tutorials*, vol. 15, no. 3, pp. 1192–1209, 2013.

[11] R. Cachucho, M. Meeng, U. Vespier, S. Nijssen, and A. Knobbe, "Mining multivariate time series with mixed sampling rates," in *Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '14)*, pp. 413–423, Seattle, Wash, USA, September 2014.

[12] G. McLachlan and T. Krishnan, *The EM Algorithm and Extensions*, John Wiley & Sons, New York, NY, USA, 2nd edition, 2008.

[13] M. Budka, B. Gabrys, and K. Musial, "On accuracy of PDF divergence estimators and their applicability to representative data sampling," *Entropy*, vol. 13, no. 7, pp. 1229–1266, 2011.

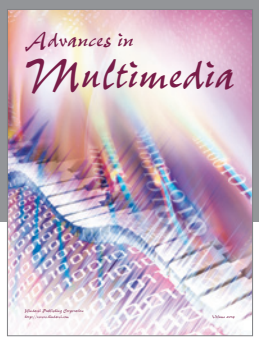
[14] S. Calderara, A. Prati, and R. Cucchiara, "Mixtures of von Mises distributions for people trajectory shape analysis ," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 21, no. 4, pp. 457–471, 2011.

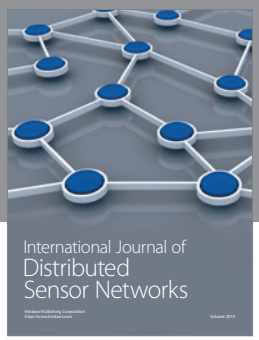
[15] G. Petkos, S. Papadopoulos, E. Schinas, and Y. Kompatsiaris, "Graph-based multimodal clustering for social event detection in large collections of images," in *MultiMedia Modeling*, C. Gurrin, F. Hopfgartner, W. Hurst, H. Johansen, H. Lee, and N. O'Connor, Eds., vol. 8325 of *Lecture Notes in Computer Science*, pp. 146–158, 2014.

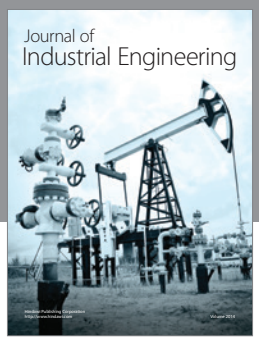
[16] C. G. M. Snoek, M. Worring, and A. W. M. Smeulders, "Early versus late fusion in semantic video analysis," in *Proceedings of the 13th ACM International Conference on Multimedia (MM '05)*, pp. 399–402, November 2005.

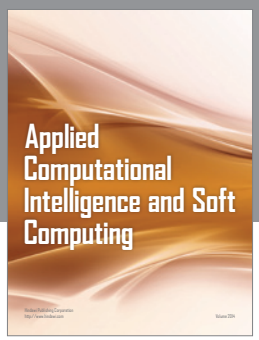
[17] C. Zhou, D. Frankowski, P. Ludford, S. Shekhar, and L. Terveen, "Discovering personal gazetteers: an interactive clustering approach," in *Proceedings of the 12th Annual ACM International Workshop on Geographic Information Systems (GIS '04)*, pp. 266–273, Washington, DC, USA, November 2004.

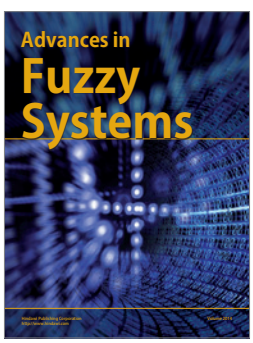
[18] P. Tan, M. Steinbach, and V. Kumar, *Introduction to Data Mining*, Addison Wesley, 2006.

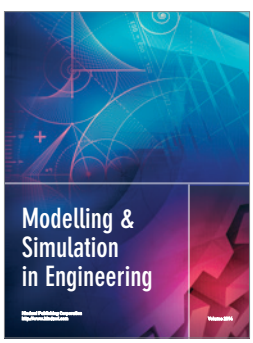


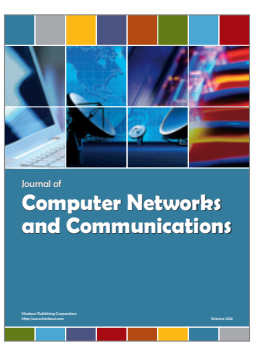












Hindawi
Submit your manuscripts at
http://www.hindawi.com