Group Activity Recognition Using Wearable Sensing Devices

In Fulfillment of the Requirements for the Degree of

Doctor of Engineering

Faculty of Informatics
Karlsruhe Institute of Technology (KIT)

Approved

Dissertation

by

Dawud Gordon

Born in Istanbul, Turkey

Examination Date: 17 Apr 2014
Primary Examiner: Prof. Dr.-Ing. Michael Beigl
Secondary Examiner: Prof. Dr. Gerhard Tröster

For David Blair,
(1952 – 2013)
who believed in me when no one else did,
including myself.
–AND–
For Dorothea Erhart,
(1953 – 2013)
who accepted me as family
unconditionally.

I would like to thank my father for being my hero, whose strength and warrior mentality in the face of adversity never ceases to amaze and humble me. I thank my mother and my sister Sara for understanding and supporting my decisions, even if it meant not seeing me for 10 years. I would also like to thank my other family, the Erharts, who selflessly welcomed me into their home and lives in that time for absolutely no reason I can think of. My friends who made daily life bearable with this workload, especially Dejan, Tirdad and Markus, I salute you. I thank my colleagues at TecO who put up with my quirks and provided a fantastic platform for creativity and constructive debate. I would also like to thank the myriad of Bachelor, Master and Diploma students who helped me along the way with experiments and implementations. My high school science teachers, John Hoffman and Stephen Edelglass who inspired me to always question and to enjoy finding the answers. Many thanks are due to my secondary adviser Prof. Gerhard Tröster for taking on this role, and for giving me the opportunity to work with him and his group in Zürich. Finally I cannot thank my adviser Prof. Michael Beigl enough for giving me the opportunity to be a part of his research team. Over the years he has always been there with ideas and inspiration, giving me the freedom to follow my own creativity, and the endless pep talks, support and advice I needed to see it through.

Abstract

One goal of ubiquitous and pervasive computing is to enable devices to adapt to the contexts and activities of the individuals interacting with them through a process called context or activity recognition. However, humans are social and spend most of their time in groups, and the behavior of the group is fundamentally different than the behavior of each individual in it, referred to as "emergent behavior" in social psychology. As the number of devices and users increases, using infrastructure and server-side processing for recognition becomes infeasible due to state space explosion and the "curse of dimensionality." As a result, peer-to-peer (P2P) methods for detecting this behavior within the network of sensing devices are needed.

Machine learning and classification are demonstrably effective for recognizing the behavior of individuals based on wearable sensing devices, especially accelerometers. However, systems which recognize contexts and activities of groups of individuals are required in order to support the needs of the group. The mobile devices we carry present an optimal platform, but aggregating data from the multitude of embedded sensors which observe behavior can congest networks. P2P approaches are attractive but challenging, since the behavior of the group cannot be observed at any single location. The challenges are **(1)** to understand the data which is required for recognition, **(2)** to detect different groups who may be in the same environment, and **(3)** to recognize the physical behavior or activities of the group, all in a P2P fashion. Furthermore, all of this must be done while **(4)** respecting the limited resources and primary functions of the sensing devices, e.g. mobile phones.

The contribution of this dissertation consists of the following:

- A formal definition of group activity recognition and differentiation from multi-user activity recognition.

- Methods for using vibration sensors to reduce the power consumption of physical activity sensing.

- Methods for reducing the consumption of the recognition toolchain using prediction for dynamic sensor selection.

- An analysis of different abstraction levels for sensor observations with respect to group activity recognition.

- Algorithms for detection group affiliation in P2P networks of mobile, wearable devices.

- Algorithms for inferring emergent group behavior in P2P networks of mobile, wearable devices.

This dissertation presents novel methods for reducing the power consumption of the machine learning tool-chain for recognition of human behavior in distributed systems. First a vibrational sensor is investigated with respect to its potential for activity recognition. The sensor proves useful for recognizing activities with high-frequency vibrational components such as riding a bicycle, as well as activities with impacts or concussions, such as walking or jogging. The approach consumes 50 times less than an accelerometer and samples high motion frequency information ($3 - 8\,kHz$) which the accelerometer can not. Methods for online sensor selection using the predictability of human behavior are explored, which can greatly reduce the energy footprint of activity sensing without sacrificing the accuracy of recognition. By predicting likely and unlikely activities for the immediate future, sensors required to differentiate unlikely activities can be turned off to conserve energy. Thereby energy savings of around 85% - 90% where achieved in turn for a loss of 1.5 - 3 percentage points of recognition accuracy

To approach the problem of distributed recognition of emergent group activities, the level of abstraction at which to fuse individual behavior information into group information is evaluated. Here power consumption and accuracy of inference is explored for using local features, local behavior classes, as well as unsupervised soft and hard clustering as the basis for inference of group activities. Using single-user activities saves 40% of power consumed by communication, but causes a 47% loss in recognition effectiveness due to technical issues Unsupervised clustering has a high potential, reducing energy consumption by 36%, while only causing 2.8% reduction in recognition rates. The unsupervised clustering abstraction level is therefore used as the abstraction level of choice for this dissertation.

Using unsupervised clustering of behavioral observations, statements can be made about the behavior of a group, based on distributed observations of individual constituents. A method using the Jeffrey's divergence of behavioral clusters and P2P communication called divergence-based affiliation detection (DBAD) is introduced, whereby group affiliations can be detected in multi group environments. When compared to a centralized approach, the DBAD reduces power consumption by up to 43% while maintaining affiliation detection rates and at the same eliminating the need for centralized resources. A method for using distributed probabilistic inference with loopy belief propagation (DPI-LBP) is presented which allows emergent group behavior to be recognized by the distributed network, further reducing energy consumption without the need for infrastructure. Again it is shown that performance is comparable to monolithic approaches, while reducing overall impact on the power consumption of the devices by up to a factor of 40. Local processing and memory usage increase slightly, but are well within tolerable boundaries.

The combined contribution is a methodology for practical recognition of groups and their emergent behavior in a P2P network of mobile phones. Applications are foreseen in adaptive intelligent environments, social networking, crowd monitoring and possibly even crowd management in emergency situations.

Deutsche Zusammenfassung

Eines der Ziele von Pervasive-Computing ist es, anhand mobiler und tragbarer Geräte, die Kontexte und Aktivitäten individueller Nutzer und Träger, zu verstehen. Dieses Ziel wird durch den Prozesses der so genannten Kontext- oder Aktivitätserkennung erreicht. Das Problem hierbei ist, dass das Verhalten eines Menschen in einer Gruppe ein grundlegend anderes ist, als das Verhalten der Gruppe selbst -auch „emergentes" Verhalten" genannt. Mit der stetig wachsenden Anzahl von Geräten und Nutzern steigen auch die zur Erkennung benötigten Ressourcen exponential, verursacht durch die Zustandsexplosion und dem sogenannten „Curse of Dimensionality". Dies führt dazu dass es fast unmöglich sein wird, diese Datenmenge an einem Ort zu verarbeiten. Aus diesem Grund werden Peer-to-Peer (P2P) Ansätze benötigt, da diese das Verhalten innerhalb des Netzwerks von mobilen und tragbaren Sensorgeräten erkennen können.

Maschinelles Lernen und Klassifikation, auf Basis von Sensoren in tragbaren Geräten- insbesondere durch den Einsatz des Accelerometer, haben sich als effektive Werkzeuge zur Verhaltenserkennung einzelner Personen erwiesen. Um die Wünsche und Bedürfnisse einer Gruppe zu erfüllen, müssen die Kontexte und Aktivitäten der Gruppe erkannt werden, was durch die Erweiterung dieser Ansätze möglich wird. Für diesen Zweck stellen mobile und tragbare Geräte eine optimale Plattform dar. Die Aggregation dieser Daten jedoch, kann zu Überlastung der Infrastruktur führen. P2P-Ansätze sind attraktiv aber herausfordernd. Denn es ist nicht möglich, durch die Beobachtung der einzelnen Menschen unabhängig voneinander, auf das Verhalten der Gesamtheit zu schließen. Die Schwierigkeit ist es **(1)** mit einem P2P-System zu verstehen, welche Daten benötigt werden um Gruppenaktivitäten im Allgemeinen zu erkennen, **(2)** verschiedene Gruppen und Gruppenzugehörigkeit zu identifizieren welche sich ggf. in derselben Umgebung aufhalten sowie **(3)** das Verhalten und die Aktivitäten der Gruppen zu verstehen. Dies alles muss unter Beachtung der **(4)** begrenzten Ressourcen und primären Funktionalität der mobilen Geräte, beispielsweise Mobiltelefone, realisiert werden.

Der Beitrag dieser Dissertation besteht aus:

- Einer formalen Definition der Gruppenaktivitätserkennung und einer Unterscheidung zwischen dieser und einer Mehrbenutzeraktivitätserkennung.
- Methoden zur Verwendung von Vibrationssensoren mit dem Ziel, den Stromverbrauch der physikalischen Aktivitätsabtastung zu verringern.
- Methoden zur Verbesserung der dynamischen Sensorselektion durch Aktivitätsprädiktion um den Stromverbrauch der gesamten Aktivitätserkennungswerkzeugkette zu verringern.

- Einer Nützlichkeitsanalyse der verschiedenen Sensordatenabstraktionsebenen zur Gruppenaktivitätserkennung.
- Einem Algorithmus zur Gruppenzugehörigkeitserkennung in P2P-Netzwerken mit mobilen und tragbaren Geräten.
- Einem Algorithmus zur Gruppenaktivitätserkennung in P2P-Netzwerken mit mobilen und tragbaren Geräten.

In dieser Dissertation werden neuartige Methoden vorgestellt, welche den Energieverbrauch der Werkzeugkette zur Erkennung menschlichen Verhaltens reduzieren. Zuerst wird das Potential eines Vibrationssensors zur Aktivitätserkennung untersucht. Dieser Sensor erweist sich als nützlich zur Erkennung von Aktivitäten hochfrequenter Bewegungskomponenten wie z.B. Fahrradfahren oder Aktivitäten mit Aufprall wie Joggen oder Gehen. Dieser Ansatz verbraucht 50-mal weniger Energie als ein Accelerometer und tastet hochfrequente Bewegungsinformation ($3 - 8\,kHz$), welche außerhalb des Messbereichs eines Accelerometers liegen ab. Es werden Methoden zur Sensorselektion in Echtzeit vorgestellt, welche die Vorhersagbarkeit menschlichen Verhaltens zur Reduzierung des Energieverbrauchs nutzen ohne auf Genauigkeit der Erkennung zu verzichten. Dies geschieht indem wahrscheinliche und unwahrscheinliche Aktivitäten in naher Zukunft vorhergesagt werden. Sensoren welche nicht benötigt werden, schalten sich aus, wodurch der Energieverbrauch um 85% - 90% verringert wird, bei einem geringen Genauigkeitsverlust von nur 1,5 - 2 Prozentpunkten.

Als nächstes wird untersucht welche Abstraktionsebene von Sensordaten die von einzelnen Gruppenmitgliedern erzeugt werden, sich am besten zur Erkennung von Gruppenaktivitäten eignet. Hier liegt der Fokus auf der Austauschbeziehung zwischen Energieverbrauch und Genauigkeit der Gruppenaktivitätserkennung, auf Basis von lokalen Signalmerkmalen, Einzelaktivitätsinformationen und nichtüberwachtem Hard- und Softclustering auf verschiedene Abstraktionsebenen. Die Verwendung der Aktivitäten von einzelnen Gruppenmitgliedern spart 40% der Energie durch weniger Datenverkehr ein, reduziert aber die Erkennung um 47% aus technischen und praktischen Gründen. Nichtüberwachtes Clustering weist ein hohes Potential auf, in dem 36% der Energie eingespart wird mit nur 2,8% Erkennungsverlust. Aus diesen Gründen wird in dieser Dissertation das nichtüberwachte Clustering als Abstraktionsebene verwendet.

Durch Clustering der verteilten Verhaltensbeobachtungen der einzelnen Gruppenmitglieder, können Aussagen über das gesamte Gruppenverhalten getroffen werden. Hierzu wird eine Methode vorgestellt, genannt „Divergence-Based Affiliation Detection" (DBAD), welche Verhaltenscluster und P2P-Kommunikation benutzt um Gruppenzugehörigkeit zu identifizieren. Verglichen mit zentralisierten Instanzen, verringert DBAD den Stromverbrauch um bis zu 43% ohne die

Erkennungsraten großartig zu verändern, eliminiert dabei aber die Notwendigkeit eines zentralen Systems. Anschließend wird eine Methode zur Erkennung der Gruppenaktivität der detektierten Gruppen in einem System vorgestellt, anhand verteilter probabilistischer Inferenz mit „Loopy Belief Propagation“ (DPI-LBP). Es wird erneut gezeigt, dass die Erkennungsperformanz vergleichbar ist mit einem zentralisierten monolithischen Ansatz, der den Energieverbrauch auf den Geräten verringert, diesmal um den Faktor bis 40, und es keine Notwendigkeit mehr für zentrale Instanzen und Infrastrukturen gibt. Lokale Berechnungen und Memory-Verbrauch steigen geringfügig, liegen aber für moderne Geräte immer noch im Toleranzbereich.

Der Gesamtbeitrag dieser Dissertation ist ein ganzheitlicher Ansatz um auf praktische Art und Weise die Erkennung von Gruppen und deren emergentes Verhalten in P2P-Netzwerken durchzuführen mit Hilfe von mobilen und tragbaren Geräten. Die Anwendung dieser Methode ist vorgesehen in adaptiv intelligenten Umgebungen, sozialen Netzwerken, Beobachtungen von Menschenmengen und eventuell auch Management von Menschenmengen und Menschenmassen in Notsituationen.

Contents

1 Introduction

The number of devices with which we interact on a daily basis is constantly increasing. So too is the complexity and utility of each device. However, their combined impact is leading to cognitive overload and interruptions [29], rather than the sum of the utility of each device. This overload is partially due to the fact that these devices require explicit input from us in order to behave as we want them to.

Automating this process presents a challenge because what we want from the device, i.e. the correct device behavior, is dependent on our current situation. For example, whether a mobile phone should ring or not is dependent on whether the situation we are in allows it or not, for example when having lunch vs. in a meeting [2]. Devices must therefore be aware of our situations and contexts (context aware or context sensitive) in order to be able to behave correctly without requiring explicit input from the user [17]. To become context aware, devices must perform context recognition, which is the process of discerning different situations from each other, based on available sensor input.

This thesis concerns itself with a subset of the field of context recognition called activity recognition. Although this is a well established field, definitions of what constitutes an activity are inconsistent and conflicting in the literature. Here an activity is defined as the following:

> **Activity:** a human situation (context) with physical motion characteristics.

Activity recognition is then defined as:

> **Activity Recognition:** the process through which a device can discern different activities based on sensor observations.

1.1 GROUP ACTIVITY RECOGNITION

As the amount of instrumentation in our environments increases, devices no longer deal with a single user, but interact with multiple users at once. As such, technology must now be able recognize the activities and contexts of groups of individuals, as well of single subjects. One good example is the field of intelligent environments, where ambient appliances such as projectors, HVAC and lighting are shared by all users, and interactive surfaces such as whiteboards and table tops where multiple users use the same device at once. Multi-user, multi-device interaction calls for devices to be aware of the context and activity of the group, rather than just the individual. A group can simply be defined as the following:

> **Group:** two or more individuals connected to one another by social relationships [10].

70% of all people in public spaces are there as part of a group [22]. Moreover "virtually all the activities of our lives [...] occur in groups rather than isolated from others" [10]. Recognizing these group affiliations would allow applications to provide automated support, such as sharing or tagging. However, the privacy preferences of the group are highly affected by the context and activity of the group [37]. Here group activity recognition can allow applications to assess the group activity and context information, permitting estimation of other conditional parameters.

Since mobile devices, specifically smart phones, are becoming ubiquitous, they represent an intriguing platform for detecting the group activities. These devices accompany many individuals at all times and contain many sensors, as well as communication, processing and memory capabilities. Furthermore they are devices with which we interact often. However, mobile devices serve a different primary purpose, whose functionality must be preserved.

1.2 THE THESIS OF THE DISSERTATION

On one hand, a group is simply a collection of individuals in spatial and temporal proximity, bound together either by physical boundaries as is the case in smart spaces, or by social ties or common goals [10]. The collective behavior of the group is not just the sum of the behavior of each individual, but rather is emergent behavior [3] as a function of the actions, interactions, and internal states of each individual [20]. I define emergent behavior following Kurt Lewin's *field theory* as the following:

> **Emergent Behavior:** the properties of the behavior of the group are different (but not necessarily greater) than the properties of the individual behavior of the group members, or the sum of that behavior [20]

In animals, this is called swarm behavior, where simple rules which homogeneously govern the behavior of each individual can lead to complex emergent behavior of the group [31]. In humans however, individual reactions, interactions and mental states are not homogeneous across all group members, further increasing system complexity.

J. H. Steward states that "systems of the higher level do not consist merely of more numerous and diversified parts," and that it is therefore "methodologically incorrect to treat each part as though it were an independent whole in itself" [30]. The implication is that the distributed system of mobile devices must collaboratively estimate group behavior by incorporating and fusing observations of many participants of the group. This is further complicated by the vast differences between group members. Also, group members may come and go at any time, taking with them both their influence on the group behavior, as well as the computation, sensing and memory of their mobile devices.

The general approach to activity recognition is conducted as follows. Distributed sensors are used to monitor certain parameters of the physical world which are affected by the activity, such as on-body accelerometers, or situated motion sensors [26]. Characteristics called features which are pertinent to these activities are extracted from those sensor using signal processing techniques. Using either classification or clustering algorithms, these features are fused into an indicator of the activity or behavior being conducted [33].

Client-server architectures have been shown to have many advantages for conducting activity recognition using a web-service based approach [2]. The advantages come through being able to offload processing and optimization from mobile, battery-powered devices to remote server locations. The remote location also has access to data from multiple users, sensors or observations, giving it a better vantage point and allowing for cross-user optimization and crowdsourcing [2]. However, there are some situations in which a client-server system architecture for activity recognition is not advantageous.

These situations can be described in terms of their circumstances which fall into three categories [14]. The first is when the bandwidth required to communicate behavior observations to a server is expensive, where the cost associated can be monetary, energy, or simply bitrate saturation. The second is when connectivity between mobile sensing devices and server infrastructure is intermittent, meaning data cannot always be delivered to the recognition apparatus. Thirdly, systems which are designed for offline deployment also cannot rely on remote servers. These conditions will be discussed further in the requirements analysis for distributed GAR in Chap. 4.

The thesis of this dissertation is to **explore the practicality of recognizing emergent group behavior (physical activities) using mobile devices.** The goal is to use the mobile devices as a distributed system for sensing, processing and memory. Each device is capable of monitoring one subject within the group using its local sensors. Collaboratively, the distributed platform must be sentient of the emergent activities of the group of human individuals wearing the devices.

1.3 CHALLENGES FOR GAR

I can now define group activity recognition as a combination of the socio-psychological aspects, human activity recognition and the conditions under which it must occur:

> **Wearable Group Activity Recognition:** an estimation of the emergent group behavior generated by the states and interactions of individual

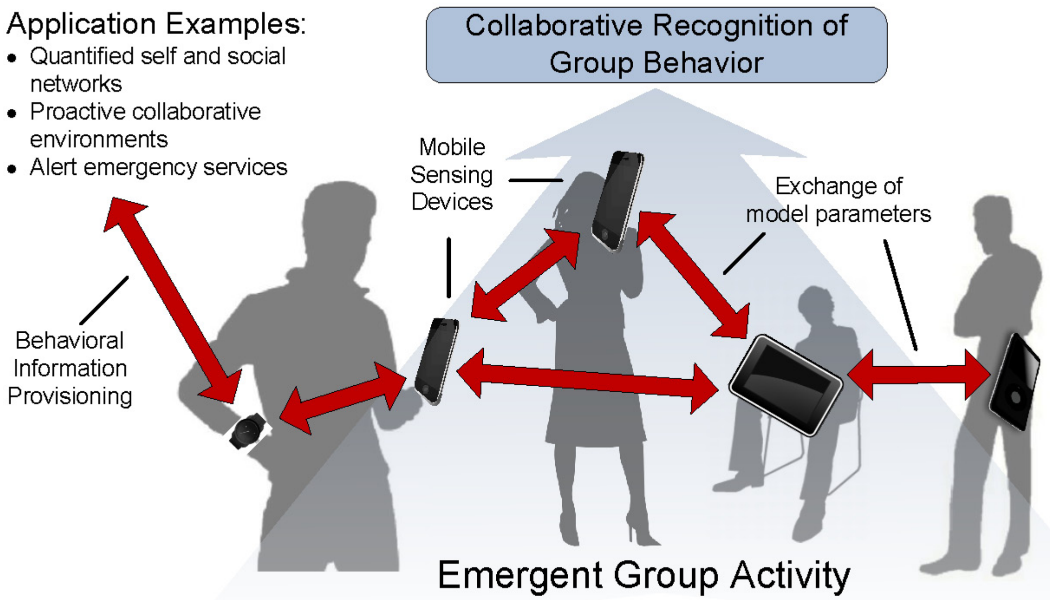


Figure 1.: Group activity recognition using peer-to-peer, ad-hoc, mobile, wearable devices

> members obtained through sparse wearable sensor observations of the distributed states and environments of the individual members.

This definition will be further explored in Chap. 3 with respect to the consequences for algorithmic approaches. The thesis of this dissertation now becomes clearer in terms of how to perform GAR under these conditions, while also observing the restrictions of mobile devices. Based on the conditions under which GAR must be performed and the requirements for operation on mobile phones, a set of challenges emerges. These questions are addressed in this dissertation, although not exhaustively as the field of wearable GAR is introduced here, but in no way concluded.

Challenge 1 - ***Low Power:*** Under normal circumstances, GAR is not the primary function of the user's device [11, 12], although it could become primary in an emergency situation [37]. If the battery is dead the device will be of no use, and if it has negative effects on the usability of the device by drastically reducing battery life it will be rejected by the user and also be of no help. The challenge is to perform GAR, while reducing the incurred power consumption as much as possible.

Challenge 2 - ***Data Abstraction Level:*** Group activity information is extracted from sensor signals at distributed locations, therefore information will have to be communicated and fused across multiple devices [13]. The more processing occurs before communication, the higher the abstraction level of the data [28]. A higher abstraction level means the data has a more symbolic nature and is smaller in size, as opposed to a numerical nature for lower abstraction levels. Processing to a higher level of abstraction reduces the volume of the data and therefore the cost of communicating it, but also discards information which may be of value for group activity recognition [13]. How much abstraction is appropriate before fusion for GAR must be evaluated.

Challenge 3 - *P2P Group Affiliation Detection:* When attempting to recognize group activities, the individual constituents of the group must first be defined [15]. In intelligent environment scenarios this may simply be the set of subjects interacting with, or contained within, the instrumented environment. However if two groups are within the same environment, or if a crowd contains multiple groups, the affiliation of individuals must be identified [25]. While group affiliation detection (GAD) is a different task than GAR, it is also subject to similar constraints and must therefore be addressed.

Challenge 4 - *P2P Group Activity Recognition:* Due to the nature of human groups, devices may fail, users come and go, connectivity may not be reliable, and bandwidth restrictions can prove critical. Connectivity to a server cannot be guaranteed, and neither can the extraction of sensor observations from the users to a remote location in a reasonable period of time [19]. Therefore, the system must be able to process distributed sensor observations to recognize group activities without relying on external resources, nor assume that any task will be performed reliably by a different entity [14]. In other words, the process of GAR must be conducted in an ad-hoc, P2P fashion by the mobile, wearable devices themselves. An analysis of the technical requirements for this challenge will be further explored in Chal. 4.

1.4 THE CONTRIBUTION

The contribution of this dissertation is the **methods, algorithms and documented experience to address these 4 challenges** for achieving GAR using mobile devices. The use of a new type of sensor for observing physical behavior is introduced. Using the sensor for activity recognition requires new data processing methods which are presented and the advantages and limitations of the technology are shown. Each sensor can sense different behavioral parameters with varying costs in terms of energy. Methods for reducing consumption by taking advantage of the inherent predictability in most scenarios are presented. These methods are evaluated with single-user applications, but are also applicable for GAR.

Single-user activities are evaluated as a component of GAR, and the optimal abstraction level for GAR is explored. The energy and accuracy tradeoff for GAR in distributed systems is evaluated using sensor features, single-user activities and unsupervised single-user activity clusters as the data basis. Features are expensive to communicate, but provide optimal recognition accuracy. Using single-user activities recognized with supervised learning techniques can theoretically also produce high recognition for GAR levels with low energy cost. However, selecting the correct activities requires behavioral experts and labeling of both group and single-user activities in parallel, which is a complex and high-effort task. Using unsupervised clustering provides high recognition levels, relatively low power consumption, and eliminates the need for experts and doubly labeling.

Using the unsupervised clustering abstraction level, a method for evaluating group affiliations of multiple subjects in a P2P network is presented. Looking at the divergence of instantaneous distributions over features between two subjects provides an indicator of "social proximity" between group members. To calculate this divergence, only cluster parameters must be exchanged between P2P nodes. A method for filtering this proximity over time is presented which results in an indicator of group affiliation which can then be classified to detect groups. Once group members have been identified, methods for inferring the group activity in a P2P network are presented. By iteratively propagating beliefs through the network, nodes converge to the correct decision over time. Beliefs can be learned at training, either using linear regression over posterior distributions of clusters between nodes, or expectation using the most likely cluster at a given time.

1.5 APPLICATION AREAS

Emergent and swarm behavior has long been studied in the animal kingdom. Many different animals display emergent behavior [31, 24] such as flocking in schools of fish or flocks of birds, or complex emergent intelligence among insect hives [8]. Monitoring of **emergent herd or flocking behavior**, for example migrant bird flocks, can be crucial for understanding these populations, and therefore also for their preservation [9]. However restricting their operation to instrumented areas is not an option, making connectivity a luxury which is often not available, or very limited by the technology [6]. In-network recognition of emergent flocking behavior could therefore alleviate the need for remote connectivity by reducing the amount of information which must be communicated, or allowing this estimation to be carried out autonomously.

Intelligent environments are instrumented spaces in which human beings work and interact, containing "myriad devices that work together to provide users access to information and services" [7]. The goal is to assist the users as much as possible in accomplishing their tasks and goals. Through the fundamental concept of ubiquitous computing, this goal can be best accomplished by allowing the devices themselves to blend into the background of the environment, and allowing users to intuitively and implicitly interact them [34]. Execution of this goal can only be carried out implicitly through context and activity recognition [27].

As people almost always carry out their business in groups [10], and the group behavior and context is fundamentally different from the contexts of the individuals [20], it follows that these environments must also be aware of group contexts and activities [5]. However as the number of devices involved in this process grows as predicted [34], the amount of data to be processed also explodes, a phenomena called "the curse of dimensionality" [23]. In-network processing of context information to a higher, more abstract level can reduce the volume and dimensionality of this data [28], therefore promising the potential to alleviate these problems.

Recently research into life-logging applications has resulted in the "**quantified self**" movement [32]. Applications are now available for individuals to track their own movements, activities and activity levels. Approaches must respect privacy to maintain clientele who will not surrender such personal information without trusting the service provider. But, since most time is spent in groups, observing only the device's user will inevitably not contain the full emergent picture. The contribution here allows social groups to be detected and then using a P2P network of group devices, log the emergent activity as well without behavioral information leaving the group. For individuals interested in sharing information over **social networks**, automated sharing would also be possible. Here a single user can share the emergent group activity without disclosing the identity of other members if wished, and even without knowing or requiring their identity at all.

Groups of people can scale to crowds, where understanding **crowd behavior** can be critical for both safety and security reasons. Smaller groups within the crowd represent social structures [1], and members will maintain proximity to each other during emergency situations [21]. This information is vital when trying to avert human tragedy, as any solution which requires the cooperation of the individuals in the crowd must maintain these structures [36]. Otherwise, instructions will be either partially or entirely ignored, or even disobeyed. Intuitively, understanding the situation and activity of the group is also necessary. Depending on the situation, certain instructions may be more or less likely to be followed [25].

Research has shown that emergency situations coincide with the worst possible conditions for infrastructure of all kinds [4]. Cloud and server-based solutions are inherently dependent on a communication infrastructure and therefore cannot be relied upon for the prevention or mitigation of dangerous situations. However, emergencies are situations in which knowledge of group activity in realtime can indeed be useful and has the potential to avert or ameliorate tragedy [16, 18, 35]. Recognition of emergent crowd and subgroup behaviors within the network of their sensing devices can provide a method for reducing the strain on infrastructure of these processes.

While crowd safety and security is a compelling application area, that research requires special experimental approaches to collect crowd data which are outside the scope of this dissertation. The exploratory research and methods presented here are designed to be scalable, theoretically to perform under crowd conditions, but an explicit evaluation of scalability or crowd activity recognition is not part of this work.

1.6 DISSERTATION STRUCTURE

The composition and structure of the content of this dissertation can be seen in Fig. 2. I begin by giving a formal definition of GAR, and differentiate between multi-user activity recognition (MAR) and single-user activity recognition (SAR) in Chap. 3. In this chapter it is also reasoned why it is not logical or plausible to give a definition of "group activities" as such, but rather only what is necessary to recognize a given behavior (SAR, MAR or GAR).

Chap. 2 gives an overview of the work done in the new field of GAR using wearable sensors. Here the areas of research on which GAR builds are described, such as MAR and SAR techniques. Also research from other disciplines into group behavior and theory as well as other recognition techniques using different sensor modalities are discussed.

Chap. 4 presents a requirements analysis for how to tackle Challenge 4 (P2P Group Activity Recognition), looking at problems that have to be addressed. Here focus is on algorithmic properties, as well as upper and lower bounds for processor load and memory distribution, as well as communication volumes.

Chap. 5 looks at reducing the power consumption footprint of activity sensing. A novel vibration sensor is introduced, as well as methods for integrating that sensor into activity recognition. Energy savings can be achieved from using the new sensor due to its low consumption, and the novel sampling method allows the processing framework to be duty cycled, working towards addressing Challenge 1 (Low Power). That challenge is further addressed in Chap. 6 which researches reducing the energy consumption of local acquisition of activity information. Future activities are predicted, allowing sensors which are not likely to be required to be turned off, saving energy without reducing the accuracy of local single-user recognition. This reduction allows the fusion to global group activity recognition to be conducted at a lower cost by reducing the cost of the information input into the process. While Chaps. 5 and 6 are evaluated for SAR, the results can also be applied to MAR and GAR as well.

Using single-user activities as recognized using SAR is also investigated as input for GAR inference. The trade-off between power consumption and communication volumes is further explored in Chap. 7. There Challenge 2 (Data Abstraction Level) is addressed by exploring GAR accuracy for different degrees of data

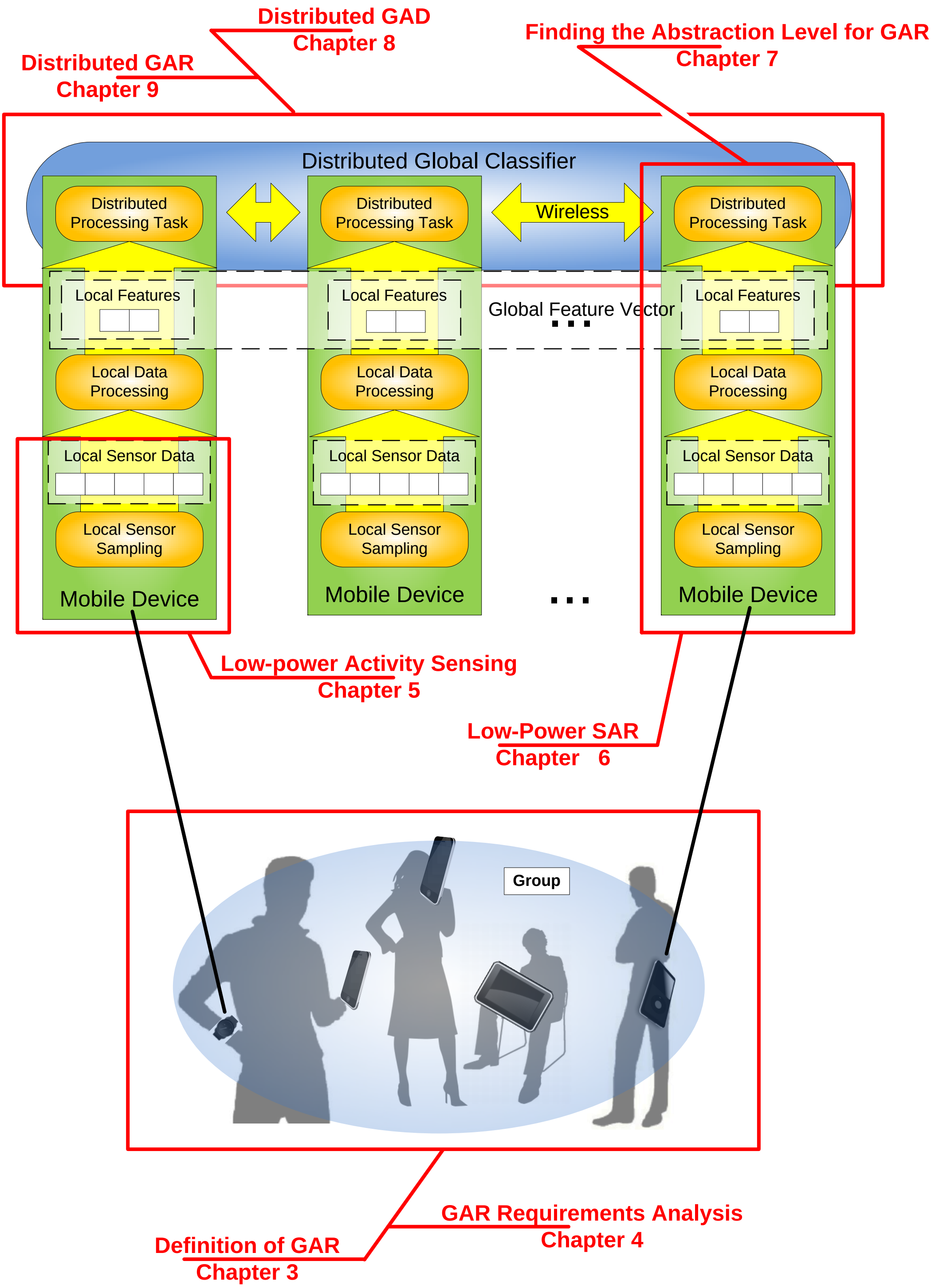


Figure 2.: Structure and composition of this dissertation

abstraction produced by mobile devices. The fact that the abstraction question is researched in an energy-aware fashion also addresses Challenge 1 indirectly as well. This GAR system and experiment is conducted using a distributed, P2P recognition system of mobile devices, addressing Challenge 4 (P2P Group Activity Recognition). However, the system presented in Chap. 7 does not completely fulfill the requirements put forth in Chap. 4and is only a research tool.

Challenge 3 (P2P Group Affiliation Detection) is addressed in Chap. 8, where a novel method is presented for detecting group affiliation using P2P wearable mobile technology. The method models local data from each user, allowing neighboring nodes to compare models with each other and assess similarity using those models. This method also addresses Challenge 1 by reducing the volume of data which must be communicated for affiliation detection, and thereby the power consumption of the detection process as well.

In Chap. 9, Challenge 4 (P2P Group Activity Recognition) is addressed by a method for inferring group behavior using mobile P2P devices. This method uses the optimal abstraction layer obtained in Chap. 7 from each node within the group. These nodes can be selected using the group affiliation information ascertained in Chap. 8. The method uses locally abstracted information to infer group behavior while meeting the requirements set forth in Chap. 4.

The cumulative contribution of this dissertation is the definition and introduction of a new field of research, namely group activity recognition (GAR) for emergent group activities using wearable sensors. The main challenges of the field are defined, and initial research into the requirements and nature of these challenges is presented, indicating that a viable approach must be implemented in a peer-to-peer fashion without using infrastructure or relying on external resources. The challenges are addressed step by step in such a way as to enable the goal of GAR using P2P wearable mobile devices. While the research contribution does not exhaust the solution space to these challenges, nor does it guarantee an optimal result, it defines and introduces the research are of wearable GAR and provides a road-map for continued research towards practical solutions in emergency situations.

Bibliography

[1] Adrian F Aveni. The Not-So-Lonely Crowd: Friendship Groups in Collective Behavior. *Sociometry*, 40(1):96–99, 1977.

[2] Martin Berchtold, Matthias Budde, Dawud Gordon, Hedda Schmidtke, and Michael Beigl. ActiServ: Activity recognition service for mobile phones. In *ISWC'10: Proceedings of the Fourteenth International Symposium on Wearable Computers*, pages 83–90, Seoul, S. Korea, 2010. IEEE Computer Society.

[3] Herbert Blumer. *Collective Behavior*. Irvington Publishers, 1951.

[4] Arjen Boin and Allan McConnell. Preparing for Critical Infrastructure Breakdowns: The Limits of Crisis Management and the Need for Resilience. *Journal of Contingencies and Crisis Management*, 15(1):50–59, March 2007.

[5] Oliver Brdiczka, Jérôme Maisonnasse, Patrick Reignier, and James L. Crowley. Detecting small group activities from multimodal observations. *Applied Intelligence*, 30(1):47–57, July 2007.

[6] Eli S. Bridge, Kasper Thorup, Melissa S. Bowlin, Phillip B. Chilson, Robert H. Diehl, René W. Fléron, Phillip Hartl, Roland Kays, Jeffrey F. Kelly, W. Douglas Robinson, and Martin Wikelski. Technology on the move: Recent and forthcoming innovations for tracking migratory birds. *BioScience*, 61 (9):89–698, 2011.

[7] Barry Brumitt, Brian Meyers, John Krumm, Amanda Kern, and Steven Shafer. Easyliving: Technologies for intelligent environments. In Peter Thomas and Hans-W. Gellersen, editors, *Handheld and Ubiquitous Computing*, volume 1927 of *Lecture Notes in Computer Science*, pages 12–29. Springer Berlin Heidelberg, 2000.

[8] Marco Dorigo, Eric Bonabeau, and Guy Theraulaz. Ant algorithms and stigmergy. *Future Gener. Comput. Syst.*, 16:851–871, June 2000.

[9] John Faaborg, Richard T. Holmes, Angela D. Anders, Keith L. Bildstein, Katie M. Dugger, Sidney A. Gauthreaux Jr., Patricia Heglund, Keith A. Hobson, Alex E. Jahn, Douglas H. Johnson, Steven C. Latta, Douglas J. Levey, Peter P. Marra, Christopher L. Merkord, Erica Nol, Stephen I. Rothstein, Thomas W. Sherry, T. Scott Sillett, Frank R. Thompson III, and Nils Warnock. Conserving migratory land birds in the new world: Do we know enough? *Ecological Applications*, 20:2:398–418, 2010.

[10] D.R. Forsyth. *Group Dynamics*. International student edition. Thomson/Wadsworth, 4th edition, 2006.

[11] D. Gordon, H.R. Schmidtke, M. Beigl, and G. von Zengen. A novel micro-vibration sensor for activity recognition: Potential and limitations. In *Wearable Computers (ISWC), 2010 International Symposium on*, pages 1 –8, oct. 2010.

[12] Dawud Gordon, Jürgen Czerny, and Michael Beigl. Activity recognition for creatures of habit. *Personal and Ubiquitous Computing*, pages 1–17, 2013.

[13] Dawud Gordon, Jan-Hendrik Hanne, Martin Berchtold, Ali Asghar Nazari Shirehjini, and Michael Beigl. Towards collaborative group activity recognition using mobile devices. *Mobile Networks and Applications*, pages 1–15, 2012.

[14] Dawud Gordon, Markus Scholz, Yong Ding, and Michael Beigl. Global peer-to-peer classification in mobile ad-hoc networks: a requirements analysis. In *Proceedings of the 7th conference on Modeling and using context (CONTEXT)*, pages 108–114. Springer, 2011.

[15] Dawud Gordon, Martin Wirz, Daniel Roggen, Michael Beigl, and Gerhard Tröster. Group affiliation detection using model divergence for wearable devices. In *Wearable Computers (ISWC), International Symposium on*, pages 1–10, 2014. (*To Appear).

[16] Dirk Helbing and Pratik Mukerji. Crowd Disasters as Systemic Failures: Analysis of the Love Parade Disaster. *Disasters*, 2012.

[17] Joyce Ho and Stephen S. Intille. Using context-aware computing to reduce the perceived burden of interruptions from mobile devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '05, pages 909–918, New York, NY, USA, 2005. ACM.

[18] inside it.ch. Eine app gegen massenpanik, December 2012.

[19] inside it.ch. Eine app gegen massenpanik - aber ohne masse, February 2013.

[20] K. Lewin. *Field theory in social science: selected theoretical papers*. Social science paperbacks. Harper, New York, 1951.

[21] Anthony R Mawson. Understanding mass panic and other collective responses to threat and disaster. *Psychiatry: Interpersonal and Biological Processes*, 68(2):95–113, 2005.

[22] Mehdi Moussaïd, Niriaska Perozo, Simon Garnier, Dirk Helbing, and Guy Theraulaz. The Walking Behaviour of Pedestrian Social Groups and Its Impact on Crowd Dynamics. *PLoS ONE*, 5(4):7, 2010.

[23] Kevin P Murphy. *Machine learning: a probabilistic perspective*. MIT Press, Cambridge, MA, 2012.

[24] Craig W. Reynolds. Flocks, herds and schools: A distributed behavioral model. *SIGGRAPH Comput. Graph.*, 21:25–34, August 1987.

[25] Daniel Roggen, Martin Wirz, Gerhard Tröster, and Dirk Helbing. Recognition of Crowd Behavior from Mobile Sensors with Pattern Analysis and Graph Clustering Methods. *Networks*, 00(0):1–24, 2011.

[26] Albrecht Schmidt. *Ubiquitous Computing - Computing in Context*. PhD thesis, Lancaster University, November 2002.

[27] Steven A. N. Shafer, Barry Brumitt, and J. J. Cadiz. Interaction issues in context-aware intelligent environments. *Hum.-Comput. Interact.*, 16(2):363–378, December 2001.

[28] S. Sigg, D. Gordon, G. Zengen, M. Beigl, S. Haseloff, and K. David. Investigation of context prediction accuracy for different context abstraction levels. *Mobile Computing, IEEE Transactions on*, PP(99):1, 2012.

[29] Terje Solvoll, Jeremiah Scholl, and Gunnar Hartvigsen. Physicians interrupted by mobile devices in hospitals: Understanding the interaction between devices, roles, and duties. *J Med Internet Res*, 15(3):e56, 2013.

[30] Julian H. Steward. *Area research, theory and practice*. Social Science Research Council, New York, 1950.

[31] D. Sumpter. The principles of collective animal behavior. *Philosophical Trans. of the Royal Society*, 2006.

[32] Melanie Swan. Emerging patient-driven health care models: an examination of health social networks, consumer personalized medicine and quantified self-tracking. *International journal of environmental research and public health*, 6(2):492–525, 2009.

[33] Kristof Van Laerhoven. *Embedded Perception - Concept Recognition by Learning and Combining Sensory Data*. PhD thesis, Lancaster University, Lancaster, 03/2006 2006.

[34] Mark Weiser and John Seely Brown. Beyond calculation. chapter The coming age of calm technolgy, pages 75–85. Copernicus, New York, NY, USA, 1997.

[35] M. Wirz, T. Franke, D. Roggen, E. Mitleton-Kelly, P. Lukowicz, and G. Troster. Inferring crowd conditions from pedestrians' location traces for real-time crowd monitoring during city-scale mass gatherings. In *Enabling Technologies:*

Infrastructure for Collaborative Enterprises (WETICE), 2012 IEEE 21st International Workshop on, pages 367–372, 2012.

[36] M. Wirz, D. Roggen, and G. Troster. Decentralized Detection of Group Formations from Wearable Acceleration Sensors. In *Computational Science and Engineering, 2009. CSE '09. International Conference on*, volume 4, pages 952–959. Ieee, 2009.

[37] M. Wirz, D. Roggen, and G. Troster. User acceptance study of a mobile system for assistance during emergency situations at large-scale events. In *Human-Centric Computing (HumanCom), 2010 3rd International Conference on*, pages 1–6, 2010.

2 Related Work

2.1 HUMAN ACTIVITY RECOGNITION AS A FIELD OF COMPUTER SCIENCE

Early approaches to activity recognition appeared in the 1980's, where early descriptions used the term *plan recognition* [32] or *action recognition* [59] in the fields of knowledge discovery, artificial intelligence, computer vision and robotics. Portable accelerometers had already been developed and were being used in the medical community to quantify energy expenditure [57]. Later on it was also shown to be an effective tool compared to standard approaches for quantifying expenditure [11].

Early publications on recognizing context in the field of mobile and ubiquitous computing sensed user location to recognize context [54, 5]. However it quickly became apparent that more information is required about a user than simply their location to infer context [46]. The usage of low level sensors for context interaction and activity recognition was proposed, and personal digital assistant (PDA) was used to implement context-sensitive interaction [45].

Using machine learning to model and recognize physical activities opened the field of activity recognition in computer science [3]. Wireless sensor nodes were used initially for recognition [47], where extremely limited resources are the main limiting factor and low-power methods gain importance [6]. As smart phones with embedded sensors become ubiquitous, focus shifts towards using these devices as the sensing platform [23]. The issue is to conduct the process of activity recognition accurately without affecting the primary functions of the mobile device [34]. Often, remote server-side resources are used to lighten the resource load for mobile devices by offloading training and optimization [8].

A recent tutorial on human activity recognition describes the problems which are being addressed [13]. From a machine learning standpoint, the common challenges are intra-class variability, inter-class similarity, and dealing with segments which are not one of the classes to be recognized, called the NULL class problem. They state the specific challenges for creating an activity recognition system as defining the diverse activities which should be recognized, handling the problems created when certain activities are performed more often then others, and collecting correctly labeled data for training. A solid description of the activity recognition toolchain, a survey of recognition methods and scenarios, and sensor positioning as well as rates achieved is also presented [13]. Of the surveyed works, recognition rates varied between 68.5% and 99.7%.

A recent publication with the goal of surveying the field of wearable activity recognition has also been presented [35]. Here the challenges for application design issues have been expanded on. Sensor selection is a main challenge, where activities have physical attributes, and selecting sensors which can observe those attributes is crucial for accurate recognition. The survey focuses on environmental sensors, acceleration sensors and location sensors as possible solutions. Other important issues are obtrusiveness, where the system should not require the subject to change their behavior, closely coupled with the data collection methodology and classifier training. Achieving high accuracy is also important as it defines how useful an activity recognition system can be, but the system must also remain flexible to changes in a subject's life, and to applications involving new subjects as well. Finally, the processing paradigm is also a question, either locally on the phone or elsewhere, which has implications for the final challenge, namely the energy consumption of the mobile app.

2.2 GROUP DYNAMICS: THE PSYCHOLOGY AND SOCIOLOGY OF GROUPS

Donelson R. Forsyth states that "virtually all the activities of our lives — working, learning, worshiping, relaxing, playing, and even sleeping — occur in groups rather than isolated from others" his work on Group Dynamics [21] which serves as the basis for this section. For example, 70% of the time we spend in public places is spent inside of a group structure [40]. Both privately and in the workplace, we spend large amounts of time in groups, which help us accomplish tasks, achieve goals, and satiate the social aspects of our human nature. Here we define a group following Forsyth as "two or more individuals who are connected to one another by social relationships," although admittedly there are numerous other but similar definitions [21]. When observing these groups throughout the private and professional lives of individuals, the average size is usually between 2 and 3 individuals [30]. These groups have a natural tendency to gravitate towards smaller sizes, down to the minimum of two individuals [24]. These groups can be social groups, committees, chat partners, or any host of congregations where members identify with each other and share social connections.

Groups have certain characterizations and properties which can be observed. Since the definition of a group is based on social connections, it is clear that there will be a certain level of **social interaction** between group members along these connections. Bales [1, 2] proposes that these interactions fall into two categories. The first is *task interaction* where members interact based on topics of group projects, tasks and goals in order to accomplish these. The second is *socieoemotional interaction*, which is not focused on practical aspects, but rather on strengthening the group's social connections, bonds and norms. This later type of interaction creates an **interdependence** of group members on each other in terms of their emotions, task and goal outcomes, and even survival [52].

Each group also has a certain **structure** which emerges from individual interactions. This structure is made up of roles which are described as "a coherent set of behaviors expected of people who occupy specific positions within a group" [21]. These roles can be characterizations of functional tasks or position behavior such as 'follower' and 'leader', and also social roles such as 'encourager' and 'compromiser' [7]. The behavior expected of each role and member of the group emerges with the structure in the form of emergent norms which define correct and incorrect behavior for group members. The individual behavior of group members conforms to meet these emergent norms, a phenomena which was demonstrated using groups of children [9, 10]. This information proves insightful later on for the purposes of detecting groups and affiliations, as will be shown in Chap. 8.

Groups perform better than individuals at accomplishing goals which is why "much of the world's work is done by groups rather than individuals." Having common **goals** is therefore also a group characteristic. One of the reasons for this are the bonds between individuals of the group. The stronger these pairwise bonds are, the higher the **cohesiveness** of the group becomes. If the bonds are weaker, the group as a system begins to disassociate and approach a natural entropy state of smaller group size or dissolve completely [17]. However, the cohesiveness of the group is stronger than the strength of the individual bonds, due to psychological effects such as feelings of belonging or unity [25]. During times of panic, these bonds are maintained and it has been shown that groups display affiliative behavior under stressed conditions [39].

Tuckman [50, 51] proposes to view the life-cycle of groups as that of an organism and presents 5 discrete stages. The first is **forming** where members of a group come together and begin to interact, followed by **storming** which is a phase of conflict where members vie for power within the group. the third stage is **norming** where conflict subsides and behavior norms and roles emerge and are established. A role is a "coherent set of behaviors expected of people who occupy specific positions within a group" [21]. the definitions of role behavior is also an emergent process, where the roles are generated by the group during the norming process [7]. In social psychology, the role descriptions or labels are very abstract and do not pertain to physical behavior, such "encourager," "compromiser," or "harmonizer."

Next the next stage in the group life-cycle is the **performing** stage where the now structured and normed group sets about reaching goals and completing tasks, after which in the **adjourning** stage the group dissipates when it is no longer required. In the context of this dissertation it would appear that each stage exhibits different behavior and must therefore be modeled separately. However, this thesis focuses on the 'performance' stage to narrow the scope, as here the applications for assistance are easier to identify and assistive systems focus on supporting productivity here.

This view of a group as an organism is validated by the fact that groups display emergent characteristics. Inevitably, observing only the individuals in the group without a holistic view of the group, results in loss of information [21]. Indeed, even human perception of group behavior, or *entitativity*, the basis for all research in Group Dynamics, follows the Gestalt laws, indicating that the entity is indeed not just the sum of its parts. The concept of emergent group behavior was formalized by Kurt Lewin who proposed that the behavior of a group is a function of the individuals and their interactions with each other and their surroundings in his *field theory of group dynamics* [36]. He proposes the principle of *interactionism* which assumes that:

$$B = f(P, E) \tag{1}$$

where group behavior B is a function of the personal characteristics P of the group members and social environmental parameters E. "According to Lewin, whenever a group comes into existence, it becomes a unified system with emergent properties that cannot be fully understood by piecemeal examination" [21].

Lewin states his holistic view of groups along the lines of Gestalt theory "that a dynamic whole has properties which are different from the properties of their parts or from the sum of their parts," also emphasizing that "the whole is not 'more' than the sum of its parts, but it has different properties" [36]. This serves as the basis for the definition of emergent group behavior introduced previously in Chap. 1, where the resulting behavior of the group, produced through the personal characteristics of individuals and their interaction with their social environment, has different characteristics than the 'sum' of the individual behaviors.

J. H. Steward continues further stating that "systems of the higher level do not consist merely of more numerous and diversified parts," and that it is therefore "methodologically incorrect to treat each part as though it were an independent whole in itself" [48]. Lewin also draws a similar conclusion that observation of individuals to obtain information of the group is not the best procedure. However, he surmises that perfectly detailed observations of all aspects of individual behavior of all group members would allow emergent properties of group behavior to be reconstructed [36]. The implications is that the group must be observed as an entity and organism in an of itself, in addition to observations of individual group members. In this thesis I explore creating a distributed discrete estimator of physical group behavior based on incomplete observations of the physical behavior of individual group constituents.

2.3 MODELING EMERGENT BEHAVIOR

Complex behavior of swarms of animals and insects is generated by simple interactions between the agents within the swarms [43, 19, 33]. within computer science, a great deal of research has been conducted on how to generate this emergent behavior using multi-agent systems [19, 33], as well as more recently how to discover the simple social rules based on complex group behavior [22, 31, 37, 20]. These areas of research are focused on observing swarm behavior, either to evaluate correctness or effectiveness of modeling systems [27], or as input for deducing governance rules to model the individual agents in the swarm [22, 31, 37, 20].

Emergent swarm behavior is also present in groups of human beings [49] and is generated as a result of social rules and individual interactions within the group. The difference being, however, the level of complexity governing the social interactions between humans [12], as opposed to swarms of ants or fish [27] for instance. A further difference, is that the humans (or their devices) within the swarm or group may themselves be interested in observing the swarm or group behavior which is being generated through emergent processes.

Here the research direction is to infer the resulting emergent group behavior generated by complex social interactions of the group members. This differs fundamentally from other forms of research into emergent swarm behavior, which is either simulates/generates swarm behavior given rules which govern individuals, or tries to infer individual rules given emergent swarm behavior. The current hypothesis is to infer or estimate discrete classes of emergent swarm behavior based on observations of individual behavior of group or swarm members. The approach which will be followed is to explore HAR techniques with respect to applicability to GAR problems as these have proven successful for recognizing activities [3, 53] and interactions [4] of individuals in the past.

2.4 RECOGNITION OF EMERGENT GROUP BEHAVIOR

Video systems have been shown to perform well for recognizing emergent behavior of groups [38]. Environments are instrumented with video cameras and use communication infrastructure to monitor human subjects. The cameras provide information on both the individuals, their interactions, and the resulting emergent group behavior which is useful for both detecting groups and recognizing their behavior [14]. However, environmental instrumentation as well as the reliance on infrastructure presents a disadvantage under some conditions as will be discussed in Chap. 4.

A similar advantage of sensing both individuals and the emergent group behavior simultaneously can be obtained using microphones [26]. Monitoring individuals' verbal interactions has been shown to pick up small cues in conversations indicative of the type of interaction [16] as well as the role of the actors within the group [18]. Using both role and interaction information, different types group properties can be inferred such as the result of conversations or the amount of inter-departmental communication within an organization [42]. Environmental microphones for audio monitoring can break down as the number of individuals or groups in the same environment increases and the research there is focused on diads and triad-sized groups where the emergent result of binary interactions are easier to model [58].

Wearable motion sensors on the other hand have higher user-fidelity, independent of the size of the group, even up to the crowd level [44]. The first wearable application sensing group behavior was documented in 1992 where a system was used to identify only the presence of groups from office behavior and interactions [54]. For organizational analysis, wearable systems can be used to analyze these structures and extract both individual characteristics such as personality traits [41], finding behavior differences and outliers [28], as well giving insight into the organizational structure as a whole [15]. Wearable systems have been shown to be useful in identifying group characteristics such as group affiliation [55] and crowd conditions such as pressure and density [56]. The aforementioned approaches are focused on centralized infrastructures for extraction of information, which under some circumstances can lead to technical difficulties or even failure [29]. While wearable systems have been used to recognize crowd and group properties, and individual properties as affected by group interactions, appropriate methods for identifying emergent group behavior from observations have not yet been introduced and are presented in this thesis. In Chap. 4 a requirements analysis finds that in order to be applicable to real-world situations, such an approach would have to be P2P in nature. Therefore the research in this dissertation follows that course as well.

Bibliography

[1] R. F. Bales. *Interaction process analysis: A method for the study of small groups*. Addison-Wesley Press, 1951.

[2] R.F. Bales. *Social Interaction Systems: Theory and Measurement*. Transaction Publishers, 2001.

[3] Ling Bao and Stephen S. Intille. Activity recognition from user-annotated acceleration data. In *Pervasive*, pages 1–17, 2004.

[4] Sumit Basu, Tanzeem Choudhury, Brian Clarkson, and Alex (Sandy) Pentland. Learning human interactions with the influence model. Technical report, MIT MEDIA LABORATORY TECHNICAL NOTE, 2001.

[5] Michael Beigl. Memoclip: A location-based remembrance appliance. *Personal Ubiquitous Computing*, 4:230–233, January 2000.

[6] Ari Y. Benbasat and Joseph A. Paradiso. A framework for the automated generation of power-efficient classifiers for embedded sensor nodes. In *SenSys '07: Proceedings of the 5th intl. conf. on Embedded networked sensor systems*, pages 219–232, New York, NY, USA, 2007. ACM.

[7] Kenneth D. Benne and Paul Sheats. Functional roles of group members. *Journal of Social Issues*, 4(2):41–49, 1948.

[8] Martin Berchtold, Matthias Budde, Dawud Gordon, Hedda Schmidtke, and Michael Beigl. ActiServ: Activity recognition service for mobile phones. In *ISWC'10: Proceedings of the Fourteenth International Symposium on Wearable Computers*, pages 83–90, Seoul, S. Korea, 2010. IEEE Computer Society.

[9] T.J. Berndt. Friends and friends' influence in adolescence. *Current Directions in Psychological Science*, 1:156–159, 1992.

[10] T.J. Berndt. Friendships in adolescence. *A lifetime of relationships*, 1:182–212, 1996.

[11] C.V.C. Bouten, K.T.M. Koekkoek, M. Verduin, R. Kodde, and J.D. Janssen. A triaxial accelerometer and portable data processing unit for the assessment of daily physical activity. *Biomedical Engineering, IEEE Transactions on*, 44(3):136 –147, march 1997.

[12] Rodney A. Brooks. From earwigs to humans. *Robotics and Autonomous Systems*, 20(2-4):291 – 304, 1997.

[13] Andreas Bulling, Ulf Blanke, and Bernt Schiele Schiele. A tutorial on human activity recognition using body-worn inertial sensors. *ACM Computing Surveys*, 2013.

[14] Ming-Ching Chang, Nils Krahnstoever, Sernam Lim, and Ting Yu. Group level activity recognition in crowded environments across multiple cameras. In *Advanced Video and Signal Based Surveillance, IEEE Conference on*, Los Alamitos, CA, USA, 2010. IEEE Computer Society.

[15] Tanzeem Choudhury and Alex Pentland. Sensing and modeling human networks using the sociometer. In *Proceedings of the 7th IEEE International Symposium on Wearable Computers*, ISWC '03, pages 216–, Washington, DC, USA, 2003. IEEE Computer Society.

[16] Jared R Curhan and Alex Pentland. Thin slices of negotiation: predicting outcomes from conversational dynamics within the first 5 minutes. *J Appl Psychol*, 92(3):802–11, 2007.

[17] Kenneth L. Dion. Group cohesion: From 'field of forces' to multidimensional construct. *Group Dynamics: Theory, Research, and Practice*, 4, 2000.

[18] Wen Dong, Bruno Lepri, Alessandro Cappelletti, Alex Sandy Pentland, Fabio Pianesi, and Massimo Zancanaro. Using the influence model to recognize functional roles in meetings. In *Proceedings of the 9th international conference on Multimodal interfaces*, ICMI '07, pages 271–278, New York, NY, USA, 2007. ACM.

[19] Marco Dorigo, Eric Bonabeau, and Guy Theraulaz. Ant algorithms and stigmergy. *Future Gener. Comput. Syst.*, 16:851–871, June 2000.

[20] Anders Eriksson, Martin Nilsson Jacobi, Johan Nyström, and Kolbjørn Tunstrøm. Determining interaction rules in animal swarms. *Behavioral Ecology*, 21(5):1106–1111, 2010.

[21] D.R. Forsyth. *Group Dynamics*. International student edition. Thomson/Wadsworth, 4th edition, 2006.

[22] Daniel Grünbaum, Steven Viscido, and Julia Parrish. Extracting interactive control algorithms from group dynamics of schooling fish. In Vijay Kumar, Naomi Leonard, and A. Morse, editors, *Cooperative Control*, volume 309 of *Lecture Notes in Control and Information Sciences*, pages 447–450. Springer Berlin / Heidelberg, 2005.

[23] Norbert Györbíró, Ákos Fábián, and Gergely Hományi. An activity recognition system for mobile phones. *Mob. Netw. Appl.*, 14(1):82–91, February 2009.

[24] A. P. Hare. *Handbook of small group research*. Free Press, New York, 2nd ed. edition, 1976.

[25] M.A. Hogg. *The social psychology of group cohesiveness: from attraction to social identity*. Harvester Wheatsheaf, 1992.

[26] Jane Yung-Jen Hsu, Chia-Chun Lian, and Wan-Rong Jih. Probabilistic models for concurrent chatting activity recognition. *ACM Trans. Intell. Syst. Technol.*, 2:4:1–4:20, January 2011.

[27] Andreas Huth and Christian Wissel. The simulation of fish schools in comparison with experimental data. *Ecological Modelling*, 75-76(0):135 – 146, 1994. State-of-the-Art in Ecological Modelling proceedings of ISEM's 8th International Conference.

[28] Inseok Hwang, Hyukjae Jang, Taiwoo Park, Aram Choi, Youngki Lee, Chanyou Hwang, Yanggui Choi, Lama Nachman, and Junehwa Song. Leveraging children's behavioral distribution and singularities in new interactive environments: Study in kindergarten field trips. In *Tenth International Conference on Pervasive Computing*, 2012.

[29] inside it.ch. Eine app gegen massenpanik - aber ohne masse, February 2013.

[30] J. James. A preliminary study of thesize determinant in small group nteraction. *American Sociological Review*, 16:474 – 477, 1951.

[31] Yael Katz, Kolbjørn Tunstrøm, Christos C. Ioannou, Cristián Huepe, and Iain D. Couzin. Inferring the structure and dynamics of interactions in schooling fish. *Proceedings of the National Academy of Sciences*, 2011.

[32] H. A. Kautz. *A formal theory of plan recognition*. PhD thesis, Rochester, NY, USA, 1987. UMI Order No. GAX87-18947.

[33] James Kennedy. Swarm intelligence. In Albert Zomaya, editor, *Handbook of Nature-Inspired and Innovative Computing*, pages 187–219. Springer US, 2006.

[34] Jennifer R. Kwapisz, Gary M. Weiss, and Samuel A. Moore. Activity recognition using cell phone accelerometers. *SIGKDD Explor. Newsl.*, 12(2):74–82, March 2011.

[35] O.D. Lara and M.A. Labrador. A survey on human activity recognition using wearable sensors. *Communications Surveys Tutorials, IEEE*, 15(3):1192–1209, 2013.

[36] K. Lewin. *Field theory in social science: selected theoretical papers*. Social science paperbacks. Harper, New York, 1951.

[37] Ryan Lukeman, Yue-Xian Li, and Leah Edelstein-Keshet. Inferring individual rules from collective behavior. *Proceedings of the National Academy of Sciences*, 107(28):12576–12580, 2010.

[38] Yunqian Ma and Petr Cisar. Activity representation in crowd. In Niels Vitoria Lobo, Takis Kasparis, Fabio Roli, JamesT. Kwok, Michael Georgiopoulos, GeorgiosC. Anagnostopoulos, and Marco Loog, editors, *Structural, Syntactic, and Statistical Pattern Recognition*, volume 5342 of *Lecture Notes in Computer Science*, pages 107–116. Springer Berlin Heidelberg, 2008.

[39] Anthony R Mawson. Understanding mass panic and other collective responses to threat and disaster. *Psychiatry: Interpersonal and Biological Processes*, 68(2):95–113, 2005.

[40] Mehdi Moussaïd, Niriaska Perozo, Simon Garnier, Dirk Helbing, and Guy Theraulaz. The Walking Behaviour of Pedestrian Social Groups and Its Impact on Crowd Dynamics. *PLoS ONE*, 5(4):7, 2010.

[41] Daniel Olguın Olguın, Peter A Gloor, and Alex Sandy Pentland. Capturing individual and group behavior with wearable sensors. In *Proceedings of the 2009 aaai spring symposium on human behavior modeling, SSS*, volume 9, 2009.

[42] Alex (Sandy) Pentland. *Honest Signals: How They Shape Our World*. The MIT Press, 2008.

[43] Craig W. Reynolds. Flocks, herds and schools: A distributed behavioral model. *SIGGRAPH Comput. Graph.*, 21:25–34, August 1987.

[44] Daniel Roggen, Martin Wirz, Gerhard Tröster, and Dirk Helbing. Recognition of Crowd Behavior from Mobile Sensors with Pattern Analysis and Graph Clustering Methods. *Networks*, 00(0):1–24, 2011.

[45] Albrecht Schmidt, Kofi Asante Aidoo, and Antti Takaluoma, et al. Advanced interaction in context. In *Proceedings of the 1st international symposium on Handheld and Ubiquitous Computing (HUC'99)*. LNCS, 1999.

[46] Albrecht Schmidt, Michael Beigl, and Hans w. Gellersen. There is more to context than location. *Computers and Graphics*, 23:893–901, 1998.

[47] Mathias Stäger, Paul Lukowicz, and Gerhard Tröster. Power and accuracy trade-offs in sound-based context recognition systems. *Pervasive and Mobile Computing*, 3:300 – 327, 2007.

[48] Julian H. Steward. *Area research, theory and practice*. Social Science Research Council, New York, 1950.

[49] Daniel Thalmann, Soraia Raupp Musse, and Marcelo Kallmann. From individual human agents to crowds. *Memory*, 1(1):6–11, 2000.

[50] Bruce W. Tuckman. Developmental sequences in small groups. *Psychological Bulletin*, 63(6):384–399, 1965.

[51] Bruce W. Tuckman and Mary Ann C. Jensen. Stages of small-group development revisited. *Group & Organization Management*, 2:419–427, 1977.

[52] R Wageman. The meaning of interdependence. *APP SOC RES*, pages 197 – 217, 2001.

[53] Liang Wang, Tao Gu, Xianping Tao, Hanhua Chen, and Jian Lu. Recognizing multi-user activities using wearable sensors in a smart home. *Pervasive Mob. Comput.*, 7(3):287–298, June 2011.

[54] Roy Want, Andy Hopper, Veronica Falcão, and Jonathan Gibbons. The active badge location system. *ACM Trans. Inf. Syst.*, 10(1):91–102, January 1992.

[55] M. Wirz, D. Roggen, and G. Troster. Decentralized Detection of Group Formations from Wearable Acceleration Sensors. In *Computational Science and Engineering, 2009. CSE '09. International Conference on*, volume 4, pages 952–959. Ieee, 2009.

[56] Martin Wirz, Tobias Franke, Daniel Roggen, Eve Mitleton-Kelly, Paul Lukowicz, and Gerhard Troster. Probing crowd density through smartphones in city-scale mass gatherings. *EPJ Data Science*, 2(1):5, 2013.

[57] Thomas C. Wong, John G. Webster, Henry J. Montoye, and Richard Washburn. Portable accelerometer device for measuring human energy expenditure. *Biomedical Engineering, IEEE Transactions on*, BME-28(6):467–471, 1981.

[58] Chenren Xu, Sugang Li, Gang Liu, Yanyong Zhang, Emiliano Miluzzo, Yih-Farn Chen, Jun Li, and Bernhard Firner. Crowd++: Unsupervised speaker count with smartphones. In *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp '13, pages 43–52, New York, NY, USA, 2013. ACM.

[59] J. Yamato, J. Ohya, and K. Ishii. Recognizing human action in time-sequential images using hidden markov model. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'92)*, pages 379 –385, jun 1992.

3 Defining Group Activity Recognition

3.1 ABSTRACT AND CONTEXT

In this chapter the definition of single user, multi-user and group activity recognition is presented and formally defined. The distinction is also made between between cooperative recognition, where remote observations can be used to improve recognition accuracy, and collaborative recognition, where remote observations are required for differentiation. The distinction between collaborative and cooperative recognition presented here is only a necessary one when addressing the distributed recognition or modeling of behavior. Since this is a new new field, it is not defined by related work, but is the contribution of this chapter and is used throughout the dissertation. Recognition of emergent group behavior is shown by definition a collaborative GAR problem, where information from multiple nodes must be fused for accurate recognition. The content of this chapter is based on an invited article in the Journal of Mobile Networks and Applications [7].

3.2 SINGLE-USER, MULTI-USER AND GROUP ACTIVITIES

In this work we differentiate between single-user activity recognition (SAR), multi-user activity recognition (MAR) and group activity recognition (GAR), which are all forms of human activity recognition (HAR). The research here is focused on wearable systems, but results should be generalizable to other approaches such as video or audio recognition systems. These terms, specifically MAR and GAR, have been multiply defined across related work, sometimes synonymously [13] while other times defined in a contradictory manner by different works (compare [8] and [10]). Here a new definition of MAR and GAR is presented with a further classification into collaborative or cooperative recognition.

For all types of HAR, labels from the label space are usually assigned to areas of the activity space in such a way as to make the activities for all labels mutually exclusive. Although this does not have to be the case, overlapping activities create a different kind of recognition problem which must be addressed.

The spaces for activities and for labels are infinite, and the mapping between them is subjective [11]. For example, when observing Fig. 3a, one person may consider it "chopping" and another considers it "cooking", and two people may describe the activity "making coffee" very differently. Therefore defining single-user, multi-user and group activities by making distinctions using labels or activity names is not a valid approach. We can, however, differentiate between these concepts by examining what is necessary in order to infer labels based on the physical characteristics of the behavior. In other words, there is no fundamental difference between single-user, multi-user and group activity *labels* per se, but rather the difference arises only when attempting to distinguish activities from each other in the process of recognition (SAR, MAR and GAR).

3.3 SAR, MAR AND GAR PROBLEMS

Single-user activity recognition (**SAR**) is the problem of recognizing what a user is doing based on sensor measurements taken from that user's body, possessions or environment [8]. This can be seen on the left side of Fig. 3a, where the activities of the single user (subject 1 is "chopping" vegetables) are being monitored. Here, SAR is only concerned with monitoring environmental parameters directly influenced by that subject, e.g. body-worn sensors or utensils which they are using.

Multi-user activity recognition (**MAR**) is the recognition of separate activities of multiple users in parallel, where two or more users are involved [8]. This is demonstrated in Fig. 3b, where the system recognizes several activities, one for each subject.

Group activity recognition (**GAR**) is the process of recognizing activities of multiple users, where a single group activity is a (complex) function of the behavior

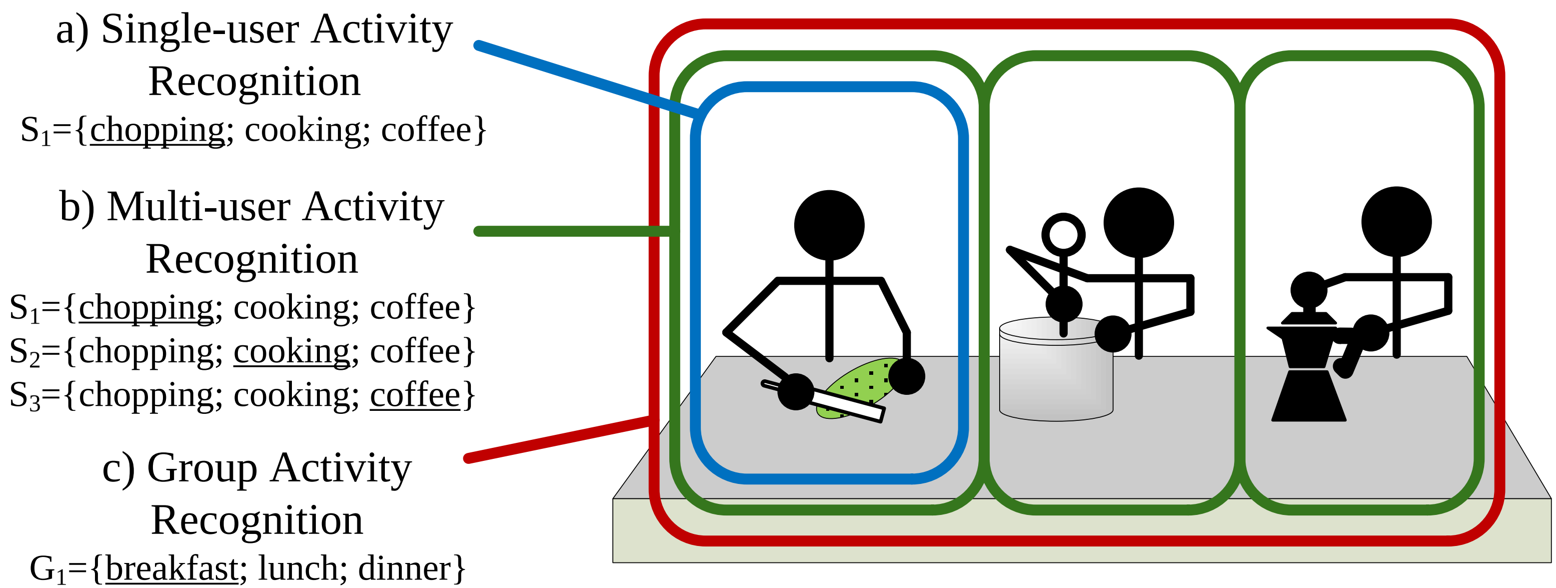


Figure 3.: An Example of single, group and multi-user activity recognition for a group G_1 consisting of 3 subjects $S_1, S_2, S_3 \in G_1$. Individual activities are chopping, cooking, and making coffee, while group activities are the meal which is being prepared

of each and every user within the group [9, 4]. The activity of the group (crowd) can be observed as spontanious emergent behavior, generated from the activities and interactions of the individuals within it [3, 12, 14]. Fig. 3c shows this where without knowledge of all of the users, it is improbable that the system will infer the correct activity, as the activities of each user are ambiguous with respect to the group activities (e.g. "chopping" could be preparation of any meal). Only the 3-equation-problem given by observing all subjects provides enough evidence for accurate inference.

In the same way that multiple sensors placed on the human body, each sensing only "arm", "leg" or "hip" parameters can be used to infer the activity of the entire person (SAR) [1, 2], we propose that sensing of the actors within the group can be used to infer the emergent activity of the group as a whole [5, 6]. This is analogous to treating the group of individuals as an organism in and of itself, rather than the sum of its parts. Although group activities are to the individuals in the group as single-user activities to the limbs of the user, the same methods can not necessarily be applied for distributed approaches. Interactions between humans are far more complex than those between e.g. "knee" and "hip," and therefore their relationship to the behavior generated is far more complex as well. Further research is required to understand what can and can not be used from SAR for GAR.

3.4 cooperative and collaborative recognition

For MAR and GAR problems we define two distinct classes: cooperative recognition problems and collaborative recognition problems. Cooperative and collaborative here do not refer to the type of human interactions occurring between the the users, but rather the type of interaction required between the subject activity models in order to recognize behavior.

Cooperative recognition problems are those in which only sensors local to a single individual are required to infer the activity. Based on these sensors, a model can be constructed in order to recognize behavior based on readings. This can be conducted independently of other individuals in the environment. The activity models, however, may be imperfect which can cause errors in recognition. Using information about neighboring individuals's activity (e.g. if one subject is "chopping" then other subjects may be more likely to be "cooking") can help to correct these errors, therefore the cooperative nature of the recognition approach.

Collaborative recognition problems are those in which recognition of the activity being performed is fundamentally dependent on information from multiple subjects. This class of problems requires that data from multiple individuals be fused in order to infer their activity, while not restricting whether or not it is one activity per user (MAR) or one activity per group (GAR). Since recognition is not possible without fusion, meaning activities cannot be modeled without observing multiple users, this class is referred as "collaborative."

3.4.1 *An Example:*

Imagine an indoor track and field area in which we are monitoring activities. Now imagine a single subject using that track whose activity we are monitoring, and let's say he or she can only perform one of three things: "run," "walk" and take a "break". This is a SAR problem since we are monitoring a single subject (see Fig. 3a) and cooperative problem, since sensor information from that user is used to infer their activity. If they were to be wearing multiple sensing devices, where each device is not capable of inferring activity on its own, we could then refer to this as being collaborative SAR problem.

Now imagine the same situation but one more subject enters the track, where the two do not know each other, are of different skill levels, etc., so that they are not in any way interacting. This is now a MAR problem (see Fig. 3b) and a cooperative problem since we are recognizing individual activities for multiple subjects, where the dependence of each recognition problem is only on the sensors of that subject.

Going one step further, imagine the same situation, except where the two are now acquaintances and take a "break" together at some point, during which we don't know what they do, but only that they do it together. Now the system must recognize "run," "walk" and "break" for both subjects (MAR), but because we cannot be sure if they are taking a break we must now observe both subjects in order to find out, making this a collaborative MAR problem. Notice how two of the activities "run" and "walk" do not change, but depending on what you are trying to distinguish them from, it is either a collaborative or a cooperative problem.

Now imagine the same situation with more users (no effect with respect to MAR/GAR, cooperative/collaborative), where they are all members of a team, meaning all individuals perform one of the activities in unison. Combining all activities together where when every one is walking the single group activity is "walk" converts a MAR problem in a GAR problem due to the fact that we are recognizing a single activity for the group, even though the label has not changed (see Fig. 3c). Similarly, the group activity could be obtained by observing only one subject, since what he or she is doing is also what the group is doing, therefore it is a cooperative GAR problem.

Finally, observe the last example where a team performs the activities "run," "walk" and "break." By changing it such that the team takes breaks together but each individual has a varying skill level such that the activity "walk and run" is possible, we can fundamentally change the nature of the recognition problem. Now, all members of the group must be observed since it cannot be assumed that if one subject is walking then the group is walking and so on. By adding one activity, the

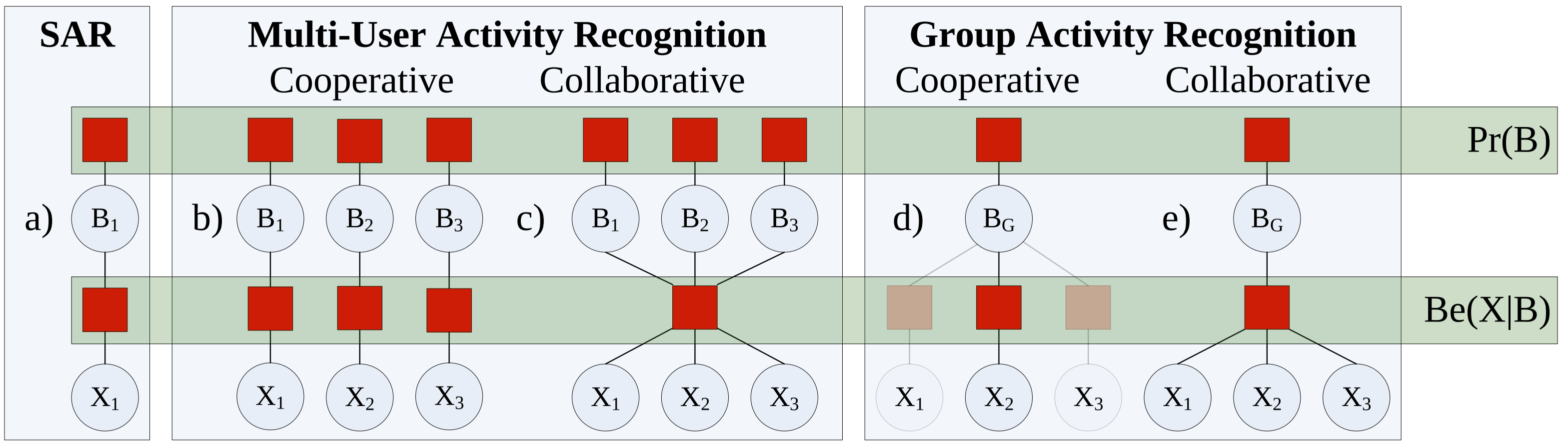


Figure 4.: Factor graphs for belief $Be(X|B)$ and priors $Pr(B)$ over hidden variables (activities) $B_1, B_2, B_3 \in B$ and (group behavior) B_G, as well observable variables (sensors) $X_1, X_2, X_3 \in X$ for **a)** single-user activity recognition (SAR), **b)** cooperative multi-user activity recognition (MAR), **c)** collaborative MAR, **d)** cooperative group activity recognition (GAR), and **e)** collaborative GAR.

problem becomes one of collaborative GAR. Once again notice that the difference between cooperative and collaborative problems is not an intrinsic property of the activities, but rather of the recognition problem.

In reality, natural group behavior is more complex than the constrained example would suggest. The labels of interest are those where the behavior is not constrained, but rather emergent from the personal characteristics of the individuals and the group dynamic [4]. By definition, recognition of emergent group behavior is then a collaborative GAR problem, where only non-emergent behavior can be recognized cooperatively. The intricacies of collaboratively recognizing physical group behavior in a distributed system are discussed further in Chap. 9.

3.5 FORMALIZATION

When observing the definition of MAR vs. GAR and cooperative vs. collaborative from a Bayesian probabilistic viewpoint, we can now formalize it using the prior and belief networks. Fig. 4a) shows the factor graph for SAR, where the hidden variables (activity states) for subject 1 $B_{s_1} \in B$ (shown only as B_1 for legibility) and group behaviors B_G are governed by a prior $Pr(B)$. The symbol B is used to describe the behavior in general, where here the physical properties of that behavior are of interest for activity recognition. The hidden variables are connected to the observable variables (sensors) $X_{s_1} \in X$ (again shown as X_1 for simplicity) by the belief $Be(X|B)$. Here X is the set of all observations for all subjects and $X_s \in X$ are the observations for a single subject. Similarly, Fig. 4b) shows the factor graph for the prior and belief function for collaborative MAR, showing the conditional independence of both belief functions and priors between hidden and observable variables. Fig 4c) shows the belief and priors for collaborative MAR, where subject activities are dependent on information of other subjects.

Fig 4d) shows the factor graph for cooperative GAR. The grayed relationships indicate a "one-of-each" relationship between the hidden variable and the observable variables, meaning that the posterior $p(B|X_s)$ can be evaluated given any X_s. Collaborative GAR is shown in Fig. 4e), where the hidden variables are dependent on all observable variables.

We define $\hat{B}$ as an estimator over the discrete states of B, or $\hat{B}(X)$, which can be thought of as the posterior $p(B|X)$ from a probabilistic standpoint. The error for estimator $\hat{B}$ given a set of observational data X is defined as the following:

$$Err.(\hat{B}(X)) = \frac{1}{N}\sum_{x=1}^{N} b_x \neq \operatorname*{argmax}_{b \in B} \hat{b}(x) \tag{2}$$

Here $b \in B$ indicates the individual activity classes, and b_x represents the label or ground truth activity class for a specific observation x. We can now define $\mathcal{R}$ or accuracy of an estimator over a data set $R(\hat{B}(X))$ as the following:

$$R(\hat{B}(X)) = 1 - Err.(\hat{B}(X)) \tag{3}$$

We can now define a term ϵ which describes the degree to which the behavior recognition rate differs between using local and global evidence: For MAR, this is using individual behaviors B_s where $|S|$ is the cardinality of $s \in S$:

$$\epsilon_{\text{MAR}} = \frac{\sum_{s \in S} \mathcal{R}(\widehat{B}_s(X_s))}{|S|} - \mathcal{R}(\widehat{B}_s(X)) \tag{4}$$

In other words if the difference between average recognition of all users given the observations of that user alone and the recognition of all users' behavior given all observation data. Similarly for GAR:

$$\epsilon_{\text{GAR}} = \frac{\sum_{s \in S} \mathcal{R}(\widehat{B}_G(X_s))}{|S|} - \mathcal{R}(\widehat{B}_G(X)) \tag{5}$$

Using these tools, we can now formalize the mechanism for differentiating collaborative from cooperative classification problems. By setting a threshold θ, we can classify problems with large ϵ values as collaborative and otherwise cooperative.

$$\text{Problem Type} = \begin{cases} \text{Collaborative,} & \text{if } \epsilon > \theta. \\ \text{Cooperative,} & \text{otherwise.} \end{cases} \tag{6}$$

In short, if each node can estimate hidden states "well" based only on local observations (low ϵ) the problem is cooperative. If observations from all individuals are required for estimation, it is a collaborative problem. For defining the difference, some subjective threshold θ must be set by the observer. Recognition of emergent group behavior is therefore by definition a collaborative recognition problem, although not all GAR problems are collaborative. For this reason, ϵ is also referred to as the "degree of emergence" of a specific problem, and is used later on again in Chap. 9. Since a hard categorization into either collaborative or cooperative problems requires a subjective threshold, and there seems to be no literature dictating a hard value for this threshold, using "degree" in a fuzzy sense appears to be a prudent course of action.

3.6 DISCUSSION

This begs the question, why are these distinctions necessary? Usually, they are not of importance to the system used to recognize these activities, as posteriors for hidden variables are inferred using priors, belief functions and observed states of all observable variables. These dependencies can also be modeled explicitly in order to improve system performance [13].

In this work we address approaching this problem from a distributed point of view, where the distributed wireless sensing network is also the platform conducting recognition. Each dependency requires communication between nodes, as they would otherwise not have access to the states of remote variables for inference. Interdependency between users for inference must be explicitly modeled and accounted for, and a distinction between which types of problems require this communication and which types do not must be made.

Since group activities and multi-user activities can be semantically distinct, there is no reason why one system would not be able to recognize a mixture of single-user, multi-user and group activities [13]. Also, there can be certain semantic overlaps between multi-user and group activities, as some activity labels can be attributed to the group as an organism, as well as multiple, and individual users, such as "jogging" for example. Since, however, there are many cases where the emergent group behavior can not be easily deduced from the single and multi-user activities, we argue that GAR is a distinct field from both MAR and SAR [5].

A caveat to this distinction is that the difference between cooperative and collaborative problems is in the amount of recognition power gained through the exchange of information, or lost when information is not exchanged. This is subjective in that a boundary must be set defining which is which, and is also different from application to application, as different applications can suffer different losses in recognition. For now, identifying if a problem is a GAR or MAR problem requires either an expert who can construct a factor graph, or modeling how much recognition is lost when between the cooperative and collaborative models.

3.7 CONCLUSION

In this chapter the fields of single-user (SAR), mutli-user (MAR) and group activity recognition (GAR) were defined and differentiated. The goal of MAR is to obtain the activities of multiple individuals in parallel, where GAR is used to recognize the activity of a group as an organism, often where the activity of the group is not directly evident when observing the activity of the individuals independently. This is given by the definitions of group behavior in social psychology, where the properties of the group behavior cannot be obtained by a piecemeal examination of the constituents: emergent behavior [9, 4]. It was also demonstrated that the differentiation between MAR and GAR cannot be carried out based on the labels used. The problem is that both the semantic meaning of labels, and the assignment of labels to activities are subjective and vary from observer to observer.

Both MAR and GAR can either be cooperative or collaborative, depending on the activities recognized. For cooperative problems, knowledge of other subjects can help to reduce error caused by model error or simplicity. Collaborative issues, however, require exchange of information about other group members in order to model the activities. However the distinction is still somewhat subjective in that a threshold is needed to define if information is crucial for recognition or not. The caveat here is that the difference between collaborative and cooperative is defined either by an expert, or by the application. Later on in Chap. 9 ϵ will be used again to examine the "degree of emergence" of a recognition problem, as emergent group activities are by definition collaborative recognition problems.

Bibliography

[1] Ling Bao and Stephen S. Intille. Activity recognition from user-annotated acceleration data. In *Pervasive*, pages 1–17, 2004.

[2] Martin Berchtold, Matthias Budde, Dawud Gordon, Hedda Schmidtke, and Michael Beigl. ActiServ: Activity recognition service for mobile phones. In *ISWC'10: Proceedings of the Fourteenth International Symposium on Wearable Computers*, pages 83–90, Seoul, S. Korea, 2010. IEEE Computer Society.

[3] Herbert Blumer. *Collective Behavior*. Irvington Publishers, 1951.

[4] D.R. Forsyth. *Group Dynamics*. International student edition. Thomson/Wadsworth, 4th edition, 2006. ISBN 9780534368227.

[5] Dawud Gordon, Jan-Hendrik Hanne, Martin Berchtold, Takashi Miyaki, and Michael Beigl. Recognizing group activities using wearable sensors. In *Mobile and Ubiquitous Systems: Networking Services, MobiQuitous, 2011. MobiQuitous '11. 8th Annual International*, dec 2011.

[6] Dawud Gordon, Markus Scholz, Yong Ding, and Michael Beigl. Global peer-to-peer classification in mobile ad-hoc networks: a requirements analysis. In *Proceedings of the 7th conference on Modeling and using context (CONTEXT)*, pages 108–114. Springer, 2011. ISBN 978-3-642-24278-6.

[7] Dawud Gordon, Jan-Hendrik Hanne, Martin Berchtold, Ali Asghar Nazari Shirehjini, and Michael Beigl. Towards collaborative group activity recognition using mobile devices. *Mobile Networks and Applications*, pages 1–15, 2012. ISSN 1383-469X.

[8] Tao Gu, Zhanqing Wu, Liang Wang, Xianping Tao, and Jian Lu. Mining emerging patterns for recognizing activities of multiple users in pervasive computing. In *The Sixth Annual International Conference on Mobile and Ubiquitous Systems*, pages 1 –10, july 2009.

[9] K. Lewin. *Field theory in social science: selected theoretical papers*. Social science paperbacks. Harper, New York, 1951.

[10] Daniel Roggen, Martin Wirz, Gerhard Tröster, and Dirk Helbing. Recognition of Crowd Behavior from Mobile Sensors with Pattern Analysis and Graph Clustering Methods. *Networks*, 00(0):1–24, 2011.

[11] Albrecht Schmidt. *Ubiquitous Computing - Computing in Context*. PhD thesis, Lancaster University, November 2002.

[12] D. Sumpter. The principles of collective animal behavior. *Philosophical Trans. of the Royal Society*, 2006.

[13] Liang Wang, Tao Gu, Xianping Tao, Hanhua Chen, and Jian Lu. Recognizing multi-user activities using wearable sensors in a smart home. *Pervasive Mob. Comput.*, 7(3):287–298, June 2011. ISSN 1574-1192.

[14] M. Wirz, D. Roggen, and G. Troster. Decentralized Detection of Group Formations from Wearable Acceleration Sensors. In *Computational Science and Engineering, 2009. CSE '09. International Conference on*, volume 4, pages 952–959. Ieee, 2009. ISBN 978-1-4244-5334-4.

4 A Requirements Analysis for Group Activity Recognition

A REQUIREMENTS ANALYSIS FOR GROUP ACTIVITY RECOGNITION

4.1 ABSTRACT AND CONTEXT

In Chap. 1, 4 challenges were presented for achieving P2P GAR using mobile devices. In this chapter I expand on Challenge 4 (P2P Group Activity Recognition) and extract technical requirements for conquering this challenge. First, 3 conditions are described for mobile ad-hoc networking (MANET) scenarios in which P2P GAR would be advantageous. This occurs when **(a)** communication with the outside world is expensive, **(b)** intermittent, or **(c)** nonexistent. In this chapter I use example P2P GAR scenarios under the aforementioned conditions to extract requirements. The emergent nature of group behavior [13] results in the following requirements for P2P GAR. The P2P recognition within the MANET must be able to **(1)** survive node failures, and **(2)** recover from rolling failures as well. The algorithm must also be able to **(3)** approximate the mapping function for estimation of the emergent activity, while at the same time **(4)** preserving the primary function of the device. It is also demonstrated that failure to observe the individuality or role of the group constituents will result in an inability to recognize emergent group behavior. Of these requirements, (1), (3) and (4) are met by the approach presented in Chap. 9, where requirement (2) is left or future work and avenues of investigation are presented in Chap. 10. The content of this chapter is heavily based on a paper published in at CONTEXT 2011 [10].

4.2 INTRODUCTION AND MOTIVATION

One of the advantages of P2P-MANETS [19] over structured client-server network architectures is their ability to adapt to new situations and account for mobility without drastically increasing complexity. The concepts of situational, context and activity recognition have been expanded to include ad-hoc mobile networks, such as wireless sensor nodes and cellular phones. In the ad-hoc network and embedded systems fields, these approaches have been focused on devices which are capable of recognizing their local situations and activities of the user. This information can then be used for local decision making or it can be communicated to a centralized back-end system with various degrees of preprocessing, compression and data fusion.

Local context recognition paradigms while being very useful for many applications [9], stand in contrast to the concept of P2P-MANETs which are used to monitor distributed systems. In the case of group activity recognition, activities and contexts can often be emergent in nature [13], meaning a correct decision cannot be made based on any single measurement within the network. Recognizing such situations is function of all distributed observations and can therefore not be factored to independent local decisions [12]. In distributed sensing systems, transmission of local situational information to a central location allows the system to recognize global situations and reduces the volume of communication when compared to forwarding unprocessed data [11].

For fully distributed ad-hoc wireless systems such as P2P-MANETs however, there is no theoretical, algorithmic or practical support available for global recognition of emergent activities in related work, which builds part of the contribution of this dissertation (see Chap. 9). This chapter will begin by identifying environments and example scenarios for global recognition in P2P-MANETS, extracting a list of requirements based on those scenarios and examining related work for applicable methods.

4.3 APPLICATION SCENARIO

Local activities refer to physical situations occurring in the immediate environment of a network node, or subset of nodes, which can sense parameters of those situations. For example the activity of a single subject is local to the device or devices which they carry on their person. Global activities on the other hand, occur over the domain of the entire MANET and are not directly measurable at any one position, but are rather deducible only when confronted with distributed observations from multiple nodes within the network. Such situations are those which are emergent in nature, where the global activity differs fundamentally from local behavior and observations. Cloud or server systems are optimal for the recognition, as they can aggregate distributed observations of behavior to create an observational representation of the emergent group behavior [12]. However, the necessity to recognize global situations within the P2P network as opposed to using centralized instances arises under the following circumstances:

SPORADIC GLOBAL CONNECTIVITY These circumstances occur when a distributed MANET is used for sensing global behavior but the connectivity state to back-end resources is only intermittently available. A typical application for this kind of setting is an application called "Landmarke" which provides support for firefighter teams using P2P-MANETS [20]. This network is deployed in an environment with unstable communication channel characteristics [22]. However, despite a connection loss to the central uplink, the individual firefighter should still be informed about the situation of the entire team, where the meaning of each activity can only be interpreted in the context of the while.

EXPENSIVE GLOBAL CONNECTIVITY Often the MANET has access to back-end resources but at an exorbitant price in terms of energy consumption, bandwidth, delay, etc. In intelligent environments, understanding a group's behavior can allow the system to proactively support that group in achieving its goals [11]. As interest in wearable devices grows, so too does the number of devices carried per person. The number of sensing modalities on each device is increasing as well, which combined with the sample rate for each sensor works as a further multiplier for the amount of data generated per person. This leads to exponential growth in the amount of data which needs to be processed in realtime known as the "curse of dimensionality" [2], creating a bandwidth bottleneck where high demand and low supply increases the cost. Uploading only emergent situations and activities acquired in the network would greatly reduce the bandwidth consumed [11].

NO GLOBAL CONNECTIVITY These conditions occur when a network without uplink must be aware of global situations and act on that information locally without access to centralized instances. In emergency situations, crowd management systems have the potential to avoid human tragedy and reduce the risk to individuals from both the cause of the emergency, and the resulting behavior of the crowd [24]. In order to conduct management effectively, such a system must be aware of the context of the crowd and subgroups within it [18]. However, infrastructure is often the first casualty of emergencies [3], meaning understanding group contexts must be conducted completely offline. This also has implications for the systems managing the emergency, since they must act autonomously to manage the emergency locally, but require global information for decision making processes. Automated infrastructure-less crowd management systems must still be investigated and are the subject of ongoing research [1].

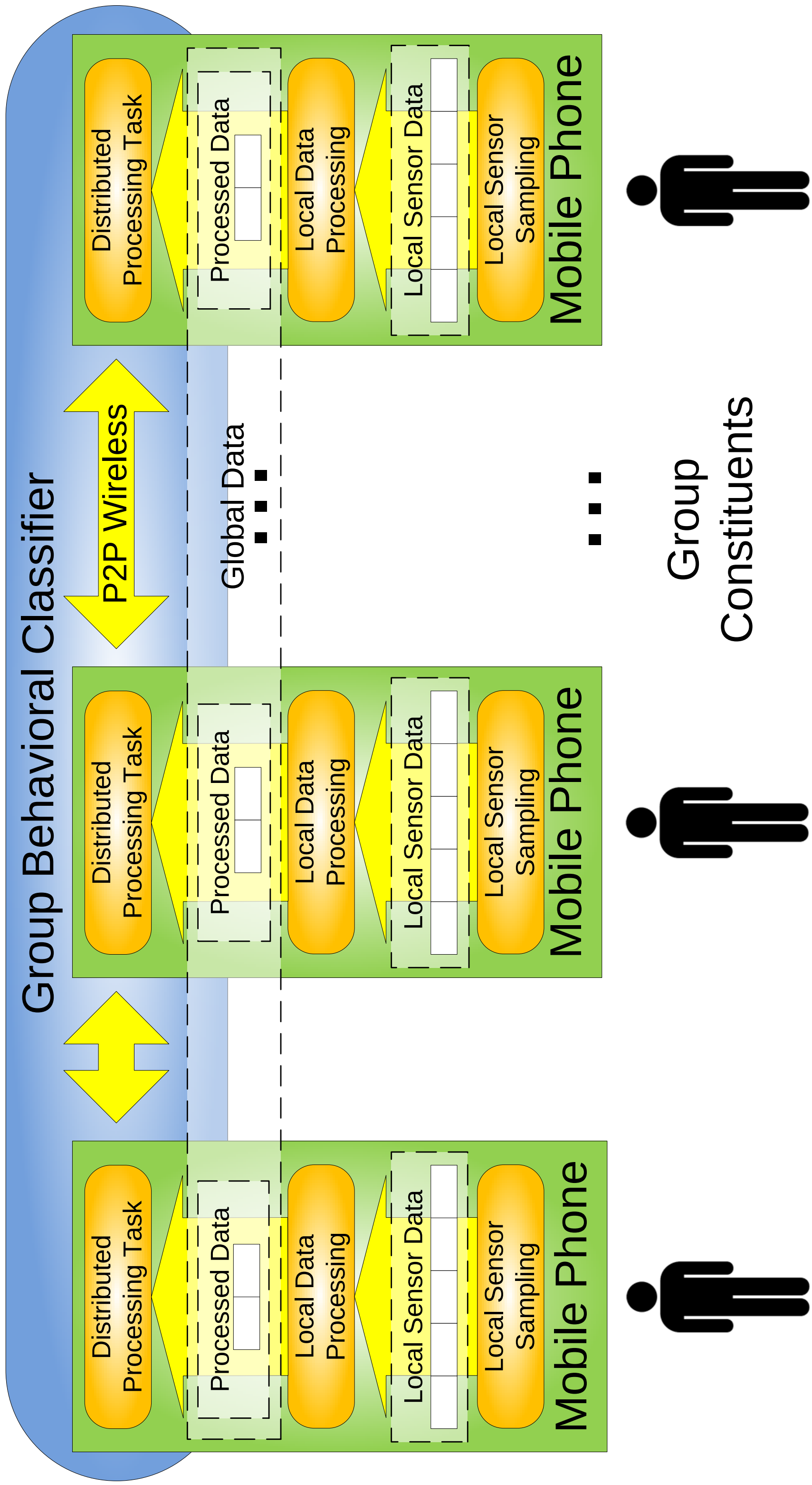


Figure 5.: Distributed classifier architecture in a P2P-MANET for group activity recognition

4.4 REQUIREMENTS ANALYSIS

Based on the application scenarios, it is possible to highlight the requirements which a distributed classification algorithm must fulfill:

Requirement 1 - ***Survival of Node Failures:*** In all three scenarios it is clear that nodes may drop out of the network without warning due to connectivity issues or node failures. This can also occur without a technical cause, for example when individuals leave the group. The global classifier algorithm must be able to continue functioning, even if in an impaired fashion. If there is any single point within the network whose local failure would also cause the global failure of the network, then this requirement has not been met.

Requirement 2 - ***Recovery from Node Failures:*** Not only must the algorithm be able to survive failures, it is also crucial that it can recover from these failures, meaning that successive node replacements do not lead to long term degradation of the algorithm. This can occur when individuals continually leave the group and other individuals join, where the group remains constant but the members change over time. Without this capability the performance of the classification algorithm would slowly degrade over time as each individual which leaves the group causes the irreparable loss of a certain amount of functionality.

Requirement 3 - ***Ability to Approximate the Mapping Function:*** Each of the three scenarios represents a different mapping function from the input signals to the contextual ground truth. Moreover, for different scenarios, the actual contexts which the system should recognize varies. As a result of this, an algorithm which would be able to accomplish these tasks must be able to effectively model the activities (accurate recognition). Due to the emergent nature of group activities, the model is necessarily a function of observations at multiple devices.

Requirement 4 - ***Preservation of the Primary Function of the Device:*** The devices which present themselves for use in global emergent group context or activity recognition often have a primary function other than GAR. Smart-phones are primarily communication and media devices. Therefore any process which conducts global context recognition must not severely impact these functions. This refers not only to the primary concern of power consumption as was already mentioned in Challenge 1, but also processing time, local memory, communication bandwidth, and all other aspects which affect the interaction between the user and their device.

4.5 RELATED WORK

In parallel computing multiple nodes work simultaneously to reduce processing time compared to a sequential approach. Parallel computing system may be categorized regarding coupling of processing elements (shared vs private memory), topology (bus, network, etc.), node type (heterogeneous vs homogeneous, processing elements vs autonomous computing entities). Following a categorization of parallel computing, the proposed distributed MANET classifier is a computing system with loosely coupled, network connected self-contained nodes. A brief review of algorithms from this field showed that these either rely on a central coordinator [5], [21] or are managed by dedicated scheduling instances [16], failing to fulfill Requirement 1 (Survival of Node Failures) or building on different conditions and cost models than MANETs.

Collaborative models and in-network data fusion are one of the most straightforward methods for P2P based classification. Therein each node contributes to a global consensus based on locally recognized situations. However, while approaches [25, 15] from this field employ different strategies to reach a global consensus, they are limited in the complexity of the mapping from local input to the global decision. These approaches observe the state of each of the nodes, and make a decision about the global situation based on these states, but without observing the identity or functionality of each node (voting). The global context algorithm is then only a function quantities of local contexts, in violation of Requirement 3 (Ability to Approximate the Mapping Function) as it can only map a subset of classification functions (see Sec. 4.6 for a discussion). Since group contexts are decidedly emergent in nature, their properties are therefore different than those of each node, or of any abstract sum of node properties [13] (see Def. 1.2). Recently [23] and [14] presented novel methods of processing context data within the nodes of a wireless network. However, there the classification is carried out by a single node, violating Requirement 1 (Survival of Node Failures) and Requirement 2 (Recovery from Node Failures).

In Organic Computing, approaches such as swarm intelligence are distributed paradigms for solving optimization problems inspired by the biological processes of swarming, flocking and herding. Various authors from this field, e.g. [6], [4] present algorithms for the distributed detection and global classification of situations. However, these algorithms conduct this in a collaborative fashion which does not support Requirement 3 (Ability to Approximate the Mapping Function), or use a central unit to perform recognition over a feature map generated in a distributed fashion which is not reconcilable with Requirement 1 (Survival of Node Failures). In short, distributed classification approaches from the area of Organic Computing cannot be directly applied to global situational recognition in P2P-MANETs.

4.6 ANALYSIS AND DISCUSSION

Social Role. As individuals come and go, the population of the group changes over time. Eventually it is possible that a large portion or all of the current members were not present when the classification algorithms where trained. One way to account for this is to simply input the data from the current set of individuals into the existing model for classification. From a practical standpoint, this would involve appending observational data from each member to build a global vector, where a new member simply replaces an old member at that location in the global vector (see Figure 5). However the implicit assumption is that behavior of two individuals is identical for the same global group activity, which we know is not the case. Each member has a different role in personal interactions and in the group dynamic [7] which is dependent on their personal characteristics and those of the group. The result is then degradation over time of the performance of the classification approach, which violates Requirement 2 (Recovery from Node Failures). This leads to the following conclusion:

Lemma 4.6.1 *A system which is individual and role agnostic can be modeled by randomly assigning individual-generated data to classifier inputs at each classification phase.*

A possible way to combat this effect would be to train the classifier using random positions for each partial vector (the data generated by each node) from each object in the total feature vector. This assumes de facto homogeneity among individuals as the information gathered from a certain individual can be input at any location on the feature vector without affecting the output of the classifier. Since the role of a single individual cannot be modeled, the only functions which can be mapped by the classifier at learning time are quantity-based functions (e.g. if the majority is sleeping then the group is sleeping), rather than inferences based on the roles of certain individuals as to the situation of the whole (e.g. inferences based on the dominant roles of certain individuals). This yields the following:

Lemma 4.6.2 *A classifier with randomized individual-to-input assignments can only learn mapping functions over input variables which are symmetric such as sums, products or averages of individual node behaviors.*

Unfortunately, functions over the quantities of objects reduces the system to majority and voting-based collaborative systems [25]. However we know that group behavior is emergent behavior, where Lewin specifically states that its properties cannot be observed through piecemeal observation of the individual members, and that "a dynamic whole has properties which are different from the properties of [its] parts or from the sum of [its] parts" [13]. Standard classifiers implicitly learn object roles in the learning process based on the behavior generated from a persons characteristics and their role in the specific dynamic. A system in which these positions are not constant must therefore explicitly account for these fluctuations. All together, the implication is the following:

Theorem 4.6.3 *A global classifier which does not observe the individuality or role of each of the objects being monitored is only capable of mapping symmetric mapping functions.*

The implication is that ignoring individuality and/or role excludes the possibility of mapping emergent functions, which are a-symmetric according to definition.

4.6.1 *Finding the Boundaries*

In this section we examine two hypothetical approaches. The examination gives insight into some of the limits for resource consumption of different distributed approaches.

Brute Force Method. The simplest solution to the global classifier problem in terms of complexity is the brute force approach. Each node transmits all locally generated data required for global context classification to every other node in the network, and then each node locally classifies the global situation using identical models. Theoretically, if the classifier is identical on each node, and the data vector is also identical, each node should locally classify the identical global situation.

The disadvantages include the amount of memory required by each node to store the entire classifier, the number of transmissions required to transmit all data generated to every other node, as well as energy consumption due to the redundancy. On the other hand, the network is extremely stable as failed nodes do not adversely affect the classification of the rest of the network, as long as the classifier used can accommodate the variable feature vector length (see [8]). Also, new nodes which are added to the network must only receive the parameters for the classifier and be added to the global list of data publishers and subscribers in order to become functioning members of the new system.

A Connectionist Approach Connectionist methods, (e.g. neural network, multi-agent system, spatial reasoning, etc.) which involve processors (neurons) and connections between these processors. This would reduce processor load and memory required when compared to the brute force approach, though it is initially unclear what affect this would have on communication between nodes. Such a method requires time synchronization which is indeed costly in ad-hoc P2P networks, though it would overcome the convergence issues of [17], and increased communication could possibly be combated by P2P self-organization. An example of such a connectionist method for exactly this purpose is presented and evaluated in Chap. 9.

Another approach would be to distribute the data instead of the execution. This could be accomplished by adapting instance-based learning methods such as k-

Table 1.: Resource consumption analysis over the number of nodes (N) for messages passed, total classifier processing (P) and total memory (M) with memory for global (S_G) local (S_L) sensor features

Algorithm	**Messages Passed**	**Processing / Node**	**Memory / Node**
Worst Cast	$N(N-1)$	P	$M+S_G$
Connectionist	$2N$	P/N	$M/N+S_L$
Best Case	N	P/N	$M/N+S_L$

Nearest-Neighbors or Self Organizing Maps to be distributed over multiple nodes along the principle that vectors which are close to each other are also close to each other in terms of hops. Once again, self-organization could be employed to account for varying network structure and mobility, but the amount of communication incurred and the advantages over brute force must be studied.

Resource Consumption Analysis Assuming N peer-to-peer nodes and objects in the network, and a distributable global classification algorithm with memory consumption M and processing load P. The brute force approach incurs the full memory consumption of M and processing load P locally at each node, as the classifier is redundantly stored and executed. The number of messages which have to be passed between nodes is $N-1$, as each node needs to communicate local features to every other node in order to build the global feature vector, or $N(N-1)$ messages in total. The memory consumption is thereby increased to $(M+S_g)$, where S_g is the size (length) of the global feature vector.

For a distributed connectionist reasoning approach, assuming each node is an input, output and hidden processor (e.g. neuron), then each node will have to pass 2 messages. Each processor requires input and generates output, where the input for the input processors is generated locally, and the output processor is output locally. In other words, per classification phase $2N$ messages must be passed by the system. Local memory consumption is now that incurred by 3 of $3N$ processors, where $3N$ processors can be held in M memory, or $\frac{M}{N}$, plus the length of the local feature vector, giving $\frac{M}{N}+S_l$. Each node must execute 3 of $3N$ processors, where the total processing load is P, yielding a load of $\frac{P}{N}$ per node. This indicates that this approach would reduce memory consumption by $\frac{M(N-1)}{N}+(S_g-S_l)$, processing load by $\frac{P(N-1)}{N}$ and the number of messages passed by $N(N-3)$. A comparison of this information is presented in Tab. 1.

Taking this one step further, we can hypothesize about the lower bounds for resource consumption in P2P-MANETs. In an optimal situation, each node sends local information to the exact logical location where it is needed (1 hop), and the system has no redundancy, indicating that each node transmits 1 message per classification phase, for a total of N messages. Also, optimally the system would distribute the memory consumption M and processor load incurred P equally across all nodes, yielding $\frac{M}{N}$ and $\frac{P}{N}$ respectively. This indicates that while being optimal in terms of memory and processor requirements, the connectionist reasoning approach would still be sub-optimal in terms of message passing by a factor of 2.

4.7 CONCLUSION

This work began by identifying the need for peer-to-peer classification of global situations in MANETs such as emergent group activities. The need occurs when either there is no communication with the outside world, that communication is very expensive, or a link is only available from time to time. These situations were elaborated on using group activity recognition in different applications. These scenarios were then analyzed in order to extract requirements for a peer-to-peer classification algorithm in wireless ad-hoc networks. The requirements identified were the ability to survive and recover from node failures, the ability to accurately map the complex classification function, and the need to respect the primary function of the device by reducing resource consumption with special emphasis on conservation of energy. This analysis indicated a further requirement of respecting heterogeneity of the different objects being monitored as well as the limited power supply and primary functions of these devices. Hypothetical upper and lower bounds for processing load, memory usage and communication volumes were elaborated, and a brute force (upper bound) and neural network (close to lower bound) approach were examined.

In this dissertation, Chaps. 5 and 6 address reducing the energy consumption footprint of activity sensing and local processing of individual behavior data, addressing Requirement 4 (Preservation of the Primary Function of the Device). Chap. 7 investigates the mapping function and its ability at different abstraction levels, addressing Requirement 3 (Ability to Approximate the Mapping Function). The distribution of the recognition algorithm in a fully decentralized manner is researched in Chap. 9 addressing Requirement 1 (Survival of Node Failures). However, Requirement 2 (Recovery from Node Failures), or the ability to recover from node failures and the integration of new group members requires further research and is beyond the scope of this dissertation. Possible approaches for accomplishing this are discussed the future work section of Chap. 10.

Bibliography

[1] D. Abusch-Magder, P. Bosch, T. E. Klein, P. A. Polakos, L. G. Samuel, and H. Viswanathan. 911- now: A network on wheels for emergency response and disaster recovery operations. Technical report, Bell Labs Technical Journal,, 2007.

[2] Christopher M Bishop. *Pattern Recognition and Machine Learning*, volume 4 of *Information science and statistics*. Springer, 2006. ISBN 9780387310732.

[3] Arjen Boin and Allan McConnell. Preparing for Critical Infrastructure Breakdowns: The Limits of Crisis Management and the Need for Resilience. *Journal of Contingencies and Crisis Management*, 15(1):50–59, March 2007. ISSN 1468-5973.

[4] Sven A. Brueckner, H. Van, and Dyke Parunak. Swarming distributed pattern detection and classification. In *Lecture Notes in Computer Science*, pages 232–245. Springer Verlag, 2005.

[5] Doina Caragea, Adrian Silvescu, and Vasant Honavar. A framework for learning from distributed data using sufficient statistics and its application to learning decision trees. *Int. J. Hybrid Intell. Syst.*, 1:80–89, April 2004. ISSN 1448-5869.

[6] P. Dasgupta. A multiagent swarming system for distributed automatic target recognition using unmanned aerial vehicles. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, 38(3):549 –563, May 2008. ISSN 1083-4427.

[7] D.R. Forsyth. *Group Dynamics*. International student edition. Thomson/Wadsworth, 4th edition, 2006. ISBN 9780534368227.

[8] D. Gordon, S. Sigg, Y. Ding, and M. Beigl. Using prediction to conserve energy in recognition on mobile devices. In *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2011 IEEE International Conference on*, pages 364 –367, march 2011.

[9] D. Gordon, J. Czerny, T. Miyaki, and M. Beigl. Energy-efficient activity recognition using prediction. In *Wearable Computers (ISWC), 2012 16th International Symposium on*, pages 29 –36, june 2012.

[10] Dawud Gordon, Markus Scholz, Yong Ding, and Michael Beigl. Global peer-to-peer classification in mobile ad-hoc networks: a requirements analysis. In

Proceedings of the 7th conference on Modeling and using context (CONTEXT), pages 108–114. Springer, 2011. ISBN 978-3-642-24278-6.

[11] Dawud Gordon, Jan-Hendrik Hanne, Martin Berchtold, Ali Asghar Nazari Shirehjini, and Michael Beigl. Towards collaborative group activity recognition using mobile devices. *Mobile Networks and Applications*, pages 1–15, 2012. ISSN 1383-469X.

[12] Dawud Gordon, Markus Scholz, and Michael Beigl. Group activity recognition using belief propagation for wearable devices. In *Wearable Computers (ISWC), International Symposium on*, pages 1–10, 2014. (*To Appear).

[13] K. Lewin. *Field theory in social science: selected theoretical papers*. Social science paperbacks. Harper, New York, 1951.

[14] Guang Lu and Wei Xue. Adaptive weighted fusion algorithm for monitoring system of forest fire based on wireless sensor networks. *Computer Modeling and Simulation, International Conference on*, 4:414–417, 2010.

[15] Ping Luo, Hui Xiong, Kevin Lü, and Zhongzhi Shi. Distributed classification in peer-to-peer networks. In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '07, pages 968–976, New York, NY, USA, 2007. ACM. ISBN 978-1-59593-609-7.

[16] Ramesh Natarajan, Radu Sion, and Thomas Phan. A grid-based approach for enterprise-scale data mining. *Future Gener. Comput. Syst.*, 23:48–54, 2007. ISSN 0167-739X.

[17] Mark A. Paskin and Carlos E. Guestrin. Robust probabilistic inference in distributed systems. In *Proceedings of the 20th conference on Uncertainty in artificial intelligence*, UAI '04, pages 436–445, Arlington, Virginia, 2004. AUAI Press. ISBN 0-9749039-0-6.

[18] Daniel Roggen, Martin Wirz, Gerhard Tröster, and Dirk Helbing. Recognition of Crowd Behavior from Mobile Sensors with Pattern Analysis and Graph Clustering Methods. *Networks*, 00(0):1–24, 2011.

[19] R. Schollmeier. A definition of peer-to-peer networking for the classification of peer-to-peer architectures and applications. In *Peer-to-Peer Computing, 2001. Proceedings. First International Conference on*, pages 101 –102, August 2001.

[20] Markus Scholz, Till Riedel, and Christian Decker. A flexible architecture for a robust indoor navigation support device for firefighters. In *Proceedings of the 7th International Conference on Networked Sensing Systems*, 2010.

[21] Markus Scholz, Gesine Flehmig, Hedda R. Schmidtke, and Gerhard H. Scholz. Powering smart home intelligence using existing entertainment systems. In *the 7th International Conference on Intelligent Environments (IE'11)*, 2011.

[22] Erhard Schubert and Markus Scholz. Evaluation of wireless sensor technologies in a firefighting environment. In *Proceedings of the 7th International Conference on Networked Sensing Systems*, 2010.

[23] Stephan Sigg and Michael Beigl. Expectation aware in-network context processing. In *Proceedings of the 4th ACM International Workshop on Context-Awareness for Self-Managing Systems*, CASEMANS '10, New York, NY, 2010. ACM. ISBN 978-1-4503-0213-5.

[24] Dimiter Velev and Plamena Zlateva. A Feasibility Analysis of Emergency Management with Cloud Computing Integration. *International Journal of Innovation Management and Technology*, 3(2):188–193, 2012.

[25] Georg Wittenburg, Norman Dziengel, Christian Wartenburger, and Jochen Schiller. A system for distributed event detection in wireless sensor networks. In *IPSN '10: Proceedings of the 9th ACM/IEEE International Conference on Information Processing in Sensor Networks*, New York, NY, 2010. ACM. ISBN 978-1-60558-988-6.

5 Low-Power Activity Sensing

5.1 ABSTRACT AND CONTEXT

In this chapter the problem of reducing the power consumption footprint of activity sensing in general is addressed in accordance with Challenge 1 (Low Power). This research can be applied for all types of activity recognition, SAR, MAR or GAR, since the sensing aspect of all approaches is identical. In all cases the goal is to sense the inter and intra-activity differences in the physical behavior, such that activities can be differentiated from each either through recognition. At the same time it is important to reduce power consumption in order to address the challenge of respecting the primary function of the device as detailed previously in Chap. 4. Here an analysis of the applicability of a novel sensor for the purpose of activity recognition is presented. The sensor measures vibration and is compared to the accelerometer which is the standard sensor for activity recognition. The results indicate that while the novel sensor consumes vastly less than a standard accelerometer (around 50 times less), it does not sense as much behavior-relevant information. However, the vibration sensor can sense information about concussions and impacts which the accelerometer cannot, improving recognition of activities with these components. These results are generated using SAR experiments but are none the less applicable for MAR and GAR as well [8, 15]. The content of this chapter is based on a publication at ISWC 2010, which was nominated for the Best Paper award that year [4].

5.2 INTRODUCTION AND RELATED WORK

Intelligent devices are increasingly expected to recognize their environment and situations. The most common method of fulfilling these expectations is by using acceleration sensors which are rapidly becoming ubiquitous in modern day technology. They are embedded in devices from cell phones and laptops, to every-day items such as tennis shoes and TV remote controls [3]. Their effects range from smart phones which are capable of adjusting themselves based on their orientation to devices that can recognize individual users and situations [3][11][19].

Several applications have already been developed using multiple acceleration sensors worn at different body locations to recognize different activities [2][9][17][18]. Other examples use one single sensor location but multiple sensor modalities to recognize a variety of activities such as daily routines [19], or a broad spectrum of activities [3][11][20]. The resulting systems can automatically recognize and adjust to certain situations and activities without the user having to explicitly input anything after a training phase. These applications are usually wearable or mobile and must therefore be energy aware in order to avoid maintenance activity such as battery replacement or charging.

In this chapter, a new approach to sensory feature creation for activity recognition in wearable computing is presented. This approach is based on a novel, low-power vibration sensor system which is used to recognize certain activities and situations while consuming significantly less power than an acceleration sensor. Other novel sensors have also been introduced to the activity recognition and wearable community in much the same way [12, 13, 18]. Here a novel ball switch is presented as a tool for context recognition. Along with a method for feature generation and information extraction specifically designed for this type of ball switch, the strengths and weaknesses of the ball switch in the context of wearable activity recognition will also be presented.

The vibration sensor is a miniaturized ball switch (Fig. 6), referred to as a micro-vibrational sensor (MVS) by the manufacturer, available as a commercial, off-the-shelf device (COTS). A conductive sphere rolls between two charged plates, closing the circuit in a certain position. With a diameter of 800 μm, the sphere's physical properties are different than those in traditional ball switches, especially in terms of sensitivity even to extremely low-intensity vibrations, as well as sensitivity in all three dimensions [16].

Initial work done with this sensor indicated that some types of activities generate vibrations on the human body, and that the novel sensor is especially useful for detecting these vibrations [7]. This motivated the hypothesis that the better time resolution of the vibration sensor (over 8000 MVS measurement events per second have been registered) may outweigh the better data resolution of acceleration sensors (3 analog acceleration vs. 1 binary vibration value) in some situations.

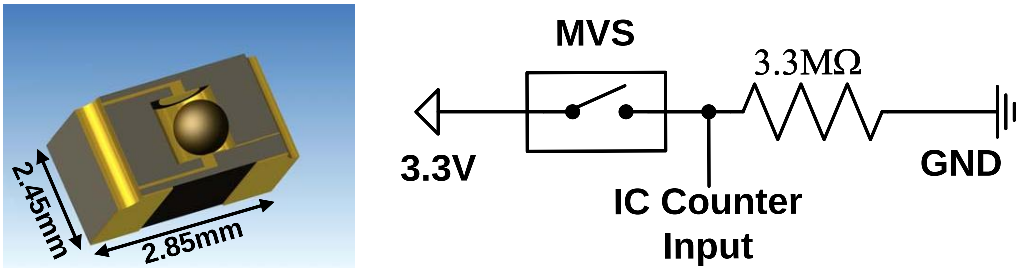


Figure 6.: The MVS [16] and schematic

We also hypothesize that maybe useful for detecting impacts and concussions within some types of activities. The intention of this work is to gain a deeper understanding of the characteristics of the sensor and its suitability for specific types of activity recognition.

Traditional ball or tilt switch sensors have been used before to successfully classify activities [18, 19] based on the evaluation of snapshots from multiple ball switches (tilt switches) to infer limb position and attitude. An approach similar to the one presented here was attempted with multiple switch inputs to a spiking neural network with mixed results [18]. In contrast to the above work, the system presented here uses a single but more sensitive sensor to recognize activity information directly extracted from sensing the *vibrations* on the body of the subject wearing the sensor. The approach in [18] effectively discards information generated by the ball switch between snapshots (samples). The novel methods for feature generation and information extraction presented in this chapter allow us to perform continuous recognition with high resolution but with very low power consumption. In this way the dynamics in vibrations can be taken into account over a period of time with a very fine resolution even at low sampling rates, rather than relying on snapshots of the system state to recognize activities.

5.3 DATA ANALYSIS AND FEATURE GENERATION

In order to evaluate the new activity recognition techniques using the vibration sensor, sampling hardware was used which simultaneously gathered sensory data from the MVS and an accelerometer. The experiment utilizes the Akiba wireless sensor node which conducted measurements using an on-board MVS micro-vibration sensor (MVS) from Sensolute [16] and an external ADXL335 3D accelerometer (referred to as the ADXL) board from Analog Devices [1]. Each axis of the ADXL is directly connected to one of the 10bit-wide A/D ports of the

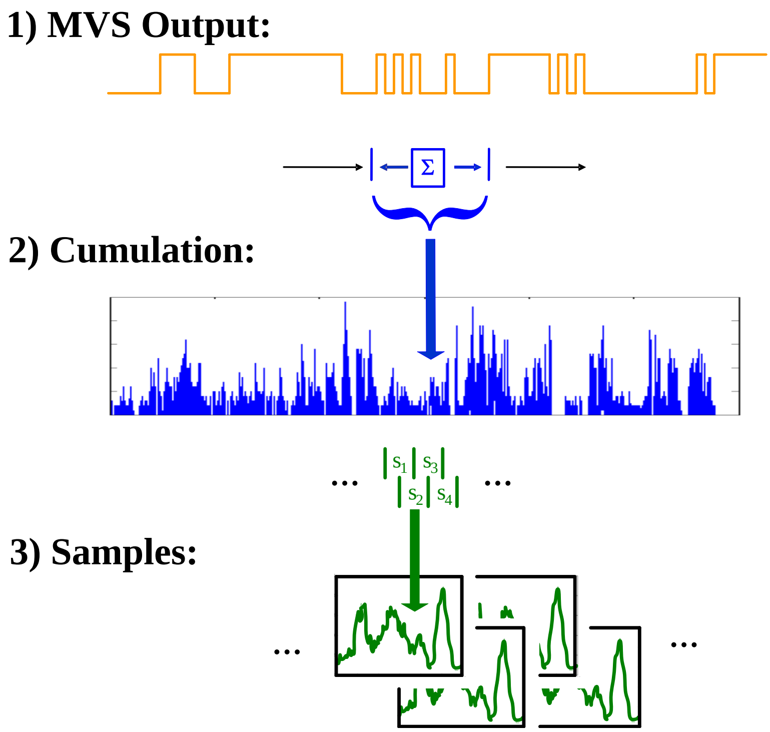


Figure 7.: The MVS preprocessing algorithm

processor (Microchip PIC18F14K22 [10]), and the MVS output is connected to the 16bit timer1 input as seen in Fig. 6.

This constellation allows A/D conversion and counting to run independently of the processing tasks. The sensor node conducted readings from both A/D (ADXL) and timer1 (MVS) registers at a frequency of 60 Hz and outputted the measurements to an external memory management unit which logged the data on a microSD card for further analysis. An in-depth analysis of the sensory device and memory management unit in terms of energy consumption and sampling methods can be found in a previous publication [7].

Unlike the signal produced by the analog acceleration sensor, the output of the MVS is an asynchronous, digital, binary vector as shown in Fig. 7(1). The relevant information in these signals are the unary transitions between the two states of the signal. The vibrational data is a time-series of sequential events whose only important unit is their time stamp, or position on the time line. These events are signaled by a voltage change on the output pin of the vibration sensor, from zero to a logical one or one to zero.

In order to be able to recognize a specific pattern within this system, namely a pattern generated by a certain activity, this signal must be converted into a form which can be analyzed using pattern matching and recognition algorithms. To create such a signal from the time-series, a cumulation method was developed which creates a wave form from the individual events. This function uses a history window to construct a wave based on the number of events in that window. The window is passed over the time line creating a new signal as depicted in Fig. 7(2). This wave, although discrete in nature, can now be treated as a digital representation of an analog signal, namely the vibration levels measured by the MVS. In Fig. 7(3), this wave is cut into separate samples to be classified by the recognition algorithm as to the activity being performed.

5.4 SENSOR HARDWARE COMPARISON

As mentioned in the introduction, the goal is embedded activity and context recognition in an ultra-low power wireless sensor node based on the MVS. For this reason the following power consumption analysis was conducted using the PIC18F14K22 [10] microprocessor from Microchip® based on the circuit in Fig. 6.

5.4.1 *Power consumption*

The ADXL335 3D acceleration sensor was chosen because of its ease of use as well as its typical power consumption signature. In the data sheet the current drawn by the sensor is indicated to be close to $425\,\mu A$ at an operating voltage of $3.3\,V$. At that voltage the rate of consumption of the ADXL is $P_{ADXL} = 1.4\,mW$. The schematic for the integration of the MVS 0608.02 shown in figure 6 implements a $3.3\,M\Omega$ pull-down resistor and therefore pulls a total current of $1\,\mu A$ at $3.3\,V$. This yields a calculated consumption of $P_{MVS-calc} = 3.3\,\mu W$.

The MVS has two states as with any switch: `ON` and `OFF`. In the `ON` state the consumption is $P_{MVS-calc} = 3.3\,\mu W$, but in the `OFF` state the consumption is zero, since no current flows over the sensor. Due to the construction of the MVS, the sensor is in either state at any given time with a probability of 50%, meaning that the actual consumption is only half of the calculated consumption, or $P_{MVS} = P_{MVS-calc}/2 = 1.65\,\mu W$. This is approximately one full order of magnitude less than that of the acceleration sensor.

3 ADC operations are necessary to convert the measured acceleration for each ADXL axis represented in voltage to a digital value, each costing $1.2\,ms$ giving a total of $3.6\,ms$ when the PIC18LF14K22 is in low power mode, e.g. is clocked at 31.25 kHz. Each ADC read requires 2 `MOV` commands to transfer the 10 bit values from the SFR to memory, each costing 1 processor cycle, yields 12 processor cycles. Each processor cycle requires 4 clock cycles yields a total $1.536\,ms$ per ADC read. Together, converting an analog value to a digital one and transferring it to specific

location costs $T_{ADXL} = 1.536\,ms + 3,6\,ms = 4.368\,ms$. Vibration readings and cumulation are directly carried out by a hardware component of the processor, the timer/counter. This is a low power module which operates independently from the rest of the embedded processor [10]. Reading this value, checking and accounting for overflow and subtracting the previously read value incurs on average 64 clock cycles which requires $T_{MVS} = 8,192\,ms$ at $31.25\,kHz$.

As the processor pulls $15.5\,\mu A$, its power consumption is $P_{proc} = 51.15\,\mu W$ at $3.3\,V$. One accelerometer measurement lasts $T_{ADXL} = 4.368\,ms$ with a consumption rate of $P_{proc} + P_{ADXL} = 1.45115\,mW$. For the vibration sensor, one reading uses a total of $P_{proc} + P_{MVS} = 54.45\,\mu W$. This indicates that the energy required to sample the MVS is approximately 14 times less than that necessary to sample the acceleration sensor. The validity of these calculations will be confirmed later in section 5.5.4.

It is important to note that these values will not scale indefinitely for higher clock rates of the processor, as there is a ceiling on minimum A/D conversion time due to capacitor load time, where the MVS wave construction only consists of processor register operations. This implies that for higher clock rates the ratio of power consumption between the two sensors will tip even farther in favor of the MVS, though overall system consumption will increase.

5.4.2 *Size and Cost*

The physical size of both sensors is also comparable; the MVS has a footprint of $2.45\,mm$ x $2.85\,mm$ where the ADXL sensor is significantly larger at $4\,mm$ x $4\,mm$. Both sensors require external circuitry in order to operate properly; the MVS requires one resister where the ADXL requires 4 capacitors, one for each axis and one for power stabilization.

The ADXL335 is one of the more costly acceleration sensors at about 5.50 USD with other comparable models priced as low as 3.00 USD. The MVS on the other hand is a far simpler sensor and is therefore less expensive. The current cost of an MVS (version MVS0608.02) sensor is approximately 1.75 USD, so the sensor is quite competitive, even at the lower end of the acceleration sensor pricing. The costs of the MVS can also be expected to fall as it is a relatively new device and increased production run length and volume would further reduce costs. On a side note, the MVS requires a counter input pin from the processor while the ADXL uses 3 A/D processor inputs.

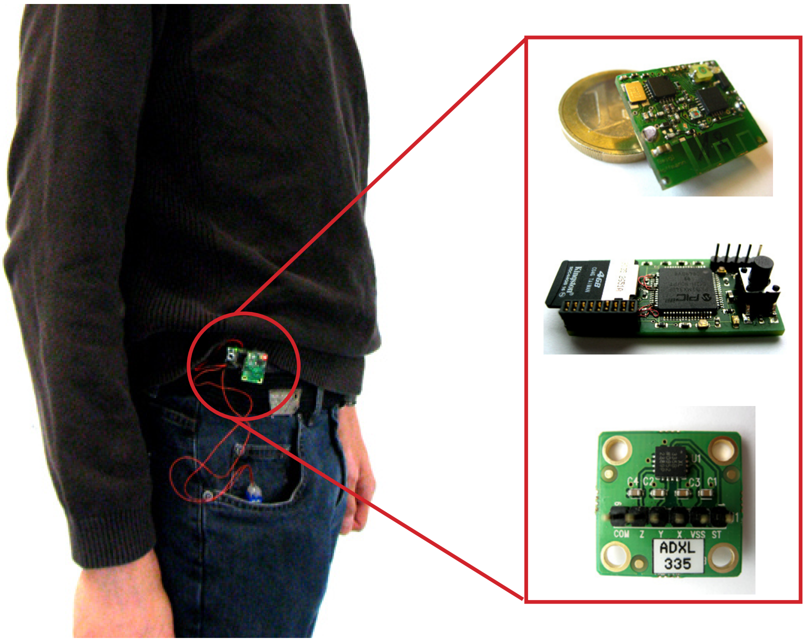


Figure 8.: Subject wearing the Akiba node (top), memory extension and ADXL335 board

5.5 EVALUATION AND RESULTS

In order to evaluate the MVS as an activity recognition tool, a case study was performed involving 8 different everyday activities. The data which was gathered during the course of the case study and was used for this evaluation are available on the Internet [6].

5.5.1 *Experimental Settings and Parameters*

The measurement and logging device described in [7] was used to gather the data for this case study along with an external acceleration sensor. The measurement logging device was powered by a plastic battery pack containing two AAA batteries. The device itself was fixed at the subject's hip between the belt and the subject's pants and the belt was fastened firmly to hold the sensor in place (see Fig. 8).

In total, 5 subjects were used to create a basis for the evaluation. 8 activities were selected consisting of riding the bus, riding a bike, walking, jogging, riding the elevator (lift), typing while seated, climbing the stairs and standing at rest. The subjects performed the selected activities, switching the device on to record and using a button to delimit activities if necessary, creating a method for annotation after the fact. During periods where no relevant activity was being performed the device was turned off, effectively limiting the data the selected activities. Three acceleration axes, the ball switch counter, as well as light and temperature sensors were all sampled synchronously.

The subjects were computer science undergraduate students with technical backgrounds although not extensively in the field of activity recognition. Each user performed all activities sequentially, and data collection was conducted one subject at a time. In total, 142 minutes of data was collected on a university campus from 5 subjects over the course of one week for the evaluative case study.

5.5.2 *Activity recognition*

The WEKA data mining toolkit [21] was selected for activity recognition for its simplification of the pattern-matching algorithms as well as its acceptance in the community [2][3][11][13][14][17]. Specifically, the C4.5 decision tree [14] was used due to its prevalence in the activity recognition literature using acceleration sensors [2][11][17] and its suitability for the intended extremely resource-restricted sensor node platform. Additionally, the IBk k-nearest neighbors and Naive Bayes classification algorithms were also evaluated in order to provide a comparison between standard recognition algorithms [5][12].

Using the samples generated by the algorithm in fig. 7(3), a set of features is generated for each sample which is used to identify the activity. The features used are identical for both the MVS and ADXL, except for the fact that the acceleration data generates 3 sets of features, one per acceleration axis. The other features generated are mean, standard deviation, entropy, area under the curve and FFT-peaks, since these were often cited as being the most decisive [2][9][11][17][20]. The three selected classification algorithms were trained by the WEKA toolkit using the activity feature sets for the vibration data on the one side and the acceleration data on the other. A sample window size of approximately 1 second with 50% overlap was selected for the case study and is constant over all classifications.

In order to evaluate the case study, 3 different classification phases were conducted. In the first phase, the classifiers were trained and tested on the data gathered from all subjects using a 5-fold approach (80% of the data is used for training and 20% for testing). The intention is to analyze how the classifiers performed if data from all subjects was present at training time. In the second phase, data collected from 4 subjects was used to train the classifiers, and the data from the remaining subject was used for testing to provide an indicator of interpersonal variances in the MVS and ADXL output respectively.

In the final phase, the effect of the MVS as a post-hoc addition to a pre-existing activity recognition system was evaluated. To show this, a classifier was trained using the acceleration, light and temperature data of all subjects 5-fold. The C4.5/J48 classifier was selected for this task because of the advantageous property of not

Table 2.: Results of the evaluation of 8 activities in percent

Phase	Type	IBk	J48	Bayes	Average
No. 1	Personalized MVS	46.2	49.2	34.1	43.2
	Personalized ADXL	91.9	96.6	65.6	84.7
No. 2	Generalized MVS	36.1	34.0	21.4	30.5
	Generalized ADXL	23.0	34.1	53.4	36.8
No. 3	ADXL, Light, Temp.		92.8		
	ADXL, Light, Temp., MVS		96.6		

being affected by junk features, meaning that redundant and useless information is automatically discarded at training time [14][21]. Then, the same procedure was conducted again with the addition of the MVS data. The goal of this phase is to assess how much novel information is delivered to a system when the MVS is integrated post-hoc, which would not be otherwise available using conventional sensors.

5.5.3 *Classifier performance*

The results of the three separate classification phases can be seen in Table 2. The acceleration sensor performed far better than the vibration sensor in the personalized classification phase no. 1, with an average classification rate over the 3 algorithms of 84.7% as compared to slightly more than half that value for the vibration sensor. The results of phase no. 2 indicate that the ADXL only slightly outperformed the MVS in this phase with a classification rate of 30.5% on average. In general, the k-nearest neighbors classifier is par with the decision tree, where the Naive Bayes classifier performed poorly compared to the other classifiers. Phase no. 3 indicates a 4% increase in overall system classification rates from 92.8% to 96.6% when the ball switch features were included. An activity per activity comparison between the classification rates of the ADXL and the MVS has been omitted here as the rates for the ADXL were relatively even across all activities and outperformed the MVS.

5.5.4 *Power measurements*

In order to confirm the calculations done in section 5.4, measurements were conducted using a BBC Goerz Metrawatt measurement device in a laboratory setting. These measurements were performed without the data logging unit. Each sensor was connected and sampled individually in an endless loop under heavy agitation to mimic activity, and current flow was measured to quantify power consumption.

Processor activities performed for the ADXL and MVS were conducted as described in sections 5.3 and 5.4. In one cycle (sensor measurement, subsequent processing), an average current flow of $630\,\mu A$ for the ADXL, and $12.8\,\mu A$ for the vibration sensor was measured. At $3.3\,V$ this yields power consumption rates of ca. $2.08\,mW$ for the ADXL ($172.8\,J$/day) and $42.24\,\mu W$ for the MVS ($3.5\,J$/day). The lifetime with a watch-type coin cell (CR1620, $1kJ$) would equate to 6 days using the ADXL and 285 days using the MVS in worst case when assuming 24/7 activity of the user. These results show that the MVS would reduce the total measured consumption of the sensor node system by a factor of almost 50 when compared to the ADXL. The difference between the calculated and measured values (MVS: $2.08\,mW$ vs. $1.45\,mW$ and ADXL: $0.04\,mW$ vs. $0.054\,mW$) is due to the difference between the consumption rates of the processor, A/D and timer unit in the preliminary data sheet and that which was measured. This disparity can either be attributed to measurement device calibration or a documentation error.

5.6 DISCUSSION

5.6.1 *Ramifications for activity recognition*

The hypothesis made is that the MVS sensor is effective for recognizing high-frequency events which occur within activities. This is confirmed by Table 3 which contains the confusion matrix from a personalized classification using the C4.5 (J48) decision tree classifier over the vibration data. The activity jogging contains a series of periodic concussions (footfalls) which stimulate the MVS. For this reason jogging was recognized by the system 79.1% of the time, walking 57.6% of the time and climbing the stairs 47.1%. Another example is the activity of riding a bike, which when conducted outdoors on an uneven surface (as was the case) consists of a series of impacts or free vibrations as the wheels encounter obstacles on the ground, combined with periodic, forced vibrations from peddling. The high frequency vibrations allow bike riding to be classified over the ball switch feature generation, yielding a recognition rate of 49.2%.

The results also indicate that the ball switch is not suitable for tasks such as gesture recognition, which often rely on the relatively low frequencies [13]. This is especially true when these gesture do not involve impulses, impacts or collisions, but are rather rounded motions such as waving or swiping. This is evident in the classification rates for activities which generate low-frequency vibrations such as standing (6.7%), riding the elevator (26%) and riding the bus (27.1%).

An interesting phenomenon is noticeable when observing the activity of typing, where a recognition rate of over 90% was achieved. This would appear to indicate that the sensor is well suited to recognize typing as an activity, when actually this is not the case. Indeed, what occurred is that when subjects were typing, often

Table 3.: Confusion matrix in percent from phase no. 1 for the MVS

a Bus	b Bike	c Walk	d Jog	e Lift	f Type	g Stair	h Stand	
27.1	6.5	3.7	0.4	10.1	40.9	4.5	6.7	**a**
9.1	49.2	12.5	0.9	5.5	2.5	17.2	3.0	**b**
2.1	4.7	57.6	8.4	5.9	0.3	20.8	0.2	**c**
0.6	0.9	9.9	79.1	1.8	0.2	7.3	0.3	**d**
7.2	3.6	11.3	1.5	26.0	35.6	10.8	4.0	**e**
2.6	1.5	0.8	0.4	1.4	90.9	0.8	1.5	**f**
3.5	7.6	21.9	9.8	9.0	0.6	47.1	0.6	**g**
5.8	2.0	1.0	0.5	5.2	77.8	1.1	6.7	**h**

no or very few events were generated by the MVS at all, causing an activity to be classified as typing during periods of no activity. This is evident when examining which other activities were confused with typing: standing (77.8%), riding the bus (40.9%) and riding the elevator (35.6%).

This is due to the fact that these activities generate low-frequency vibrations which often produce little or no activity from the MVS. As typing is an activity which consistently produces almost no output, all of the sample features which do not contain any ball switch events are classified as typing, explaining the high confusion rates. The implication is that an activity recognition system which is based on the MVS would benefit by having a “Zero" class into which all sample windows are classified which do not contain any, or only very few events. This would differentiate between activity samples which have been classified and those which simply did not generate enough vibrations to be classified. A possible method for handling such cases would be to increase the sample and cumulation window lengths, which under certain conditions would reduce the number of samples with 0 events, though at the cost of reduced reaction time.

The results of the three-phase classification study demonstrate that the acceleration sensor is capable of delivering quantitatively more information of relevance for activity recognition when compared to the ball switch. This can be seen clearly when observing phase no. 1 of the case study where personalized classification using the ADXL was significantly more successful than the MVS for the same activities (84.7% compared to 43.2%).

Phase no. 2 on the other hand, indicates that much of this data is largely subject dependent, making it less useful for a generic monolithic approach to context recognition. In this phase the performance of the classifiers dropped for both the vibration and acceleration data, whereby the reduction in recognition rates on average for the vibration data is significantly less than the acceleration data (29.4%

Table 4.: The combined insight of this paper

	Acceleration	Vibration (MVS)
Power consumption[1]	2 mW	42 uW
Resolution	3D 10 bit[2]	1 bit
Size	16 sq. mm	7 sq. mm
Suggested use	low-mid. freq. activities personalized detection	high frequency activities unpersonalized detection

1 at 3.3V
2 for most small microprocessors
3 due to slow A/D in small microprocessors. Sensor max is 1.6kHz

for the MVS versus 56.6% for the ADXL). This would indicate that although the vibration sensor delivers less data than the acceleration sensor, the data is more generic per activity across multiple subjects.

The results from phase no. 3 show an improvement of over 4% in a 3 sensor activity recognition system when the MVS is introduced into the system. This strengthens the assumption that the MVS and the ADXL have complimentary sensitivity ranges in terms of frequency bands and therefore provide activity data which is also complimentary in nature with some overlap.

Lastly, the vibrations which are being measured using the MVS are not usually being generated at that location, but rather these signals must propagate through the human body before arriving at the sensor. This would indicate that sensor location is a crucial aspect when using the MVS for activity recognition as each activity would create a different vibration pattern at a different location, depending on what types of tissue the vibration propagates through. This would suggest that classification rates are only valid for the location where the data was sampled, e.g. are highly location dependent.

5.7 CONCLUSION

This chapter showed the potential of a novel vibration sensor as a tool for continuous, low-power, wearable activity recognition. Table 4 gives an overview of the characteristics of the vibration sensor system and its use in activity recognition, and presents a comparison with activity recognition based on a 3D acceleration sensor.

The MVS is capable of sensing activity data pertinent to standard recognition algorithms. On the one side, the MVS does not deliver as much information as the ADXL acceleration sensor. The MVS can be used well to recognize activities which contain concussions and impacts such as jogging, riding a bike on uneven ground, or presumably tapping on a hard surface. Furthermore, the results indicate that the MVS can generate sensory information which can be better generalized over multiple subjects using a generic monolithic classifier approach.

Finally, the MVS was evaluated as an addition to existing activity recognition systems based on standard sensors including acceleration. The results indicate that the MVS can improve recognition rates while costing one third as much as an ADXL acceleration sensor, taking up one half the size, and consuming 50 times less power, addressing Challenge 1 (Low Power). All of this makes the MVS a resource-effective, simpler alternative to, or extension of, acceleration sensors for low-power, low-cost wearable activity recognition systems for researchers and developers. As the acceleration and MVS based recognition performs significantly better than just acceleration based recognition, there is strong evidence that high-frequency vibrational signals generated by everyday activities is very useful for activity recognition, and that the MVS is capable of sampling that information.

Bibliography

[1] Analog Devices Inc. The ADXL335 3-Dimensional Acceleration Sensor Specification. http://www.analog.com/, 2010.

[2] Ling Bao and Stephen S. Intille. Activity recognition from user-annotated acceleration data. In *Pervasive*, pages 1–17, 2004.

[3] Kenghao Chang, Jeffrey Hightower, and Branislav Kveton. Inferring identity using accelerometers in television remote controls. In *Proceedings of the Seventh International Conference on Pervasive Computing*, Nara, Japan, May 2009.

[4] D. Gordon, H.R. Schmidtke, M. Beigl, and G. von Zengen. A novel micro-vibration sensor for activity recognition: Potential and limitations. In *Wearable Computers (ISWC), 2010 International Symposium on*, pages 1 –8, oct. 2010.

[5] Dawud Gordon, Hedda Schmidtke, and Michael Beigl. Introducing new sensors for activity recognition. In *How To Do Good Research In Activity Recognition: Experimental methodology, performance evaluation and reproducibility. Workshop in conjunction with Pervasive 2010*, Helsinki, Finland, May 2010.

[6] Dawud Gordon, Georg von Zengen, Hedda Rahel Schmidtke, and Michael Beigl. A data set from wearable accelerometer and vibration sensor, May 2010. http://www.teco.edu/~gordon/MVS/.

[7] Dawud Gordon, Florian Witt, Hedda Schmidtke, and Michael Beigl. A long-term sensory logging device for subject monitoring. In *Workshop for Situation Recognition and Medical Data Analysis in Pervasive Health Environments (PervaSense 2010)*, Munich, Germany, March 2010.

[8] Dawud Gordon, Markus Scholz, and Michael Beigl. Group activity recognition using belief propagation for wearable devices. In *Wearable Computers (ISWC), International Symposium on*, pages 1–10, 2014. (*To Appear).

[9] Jonathan Lester, Tanzeem Choudhury, and Gaetano Borriello. A practical approach to recognizing physical activities. In *Pervasive*, pages 1–16, 2006.

[10] Microchip Technology Inc. The PIC18F14K22 Microprocessor Product Specification. http://www.microchip.com/, 2010.

[11] Emiliano Miluzzo, Nicholas D. Lane, Kristóf Fodor, Ronald Peterson, Hong Lu, Mirco Musolesi, Shane B. Eisenman, Xiao Zheng, and Andrew T. Campbell. Sensing meets mobile social networks: the design, implementation and

evaluation of the cenceme application. In *SenSys '08: Proceedings of the 6th ACM conference on Embedded network sensor systems*, pages 337–350, New York, NY, USA, 2008. ACM. ISBN 978-1-59593-990-6.

[12] Georg Ogris, Matthias Kreil, and Paul Lukowicz. Using fsr based muscule activity monitoring to recognize manipulative arm gestures. In *ISWC '07: Proceedings of the 2007 11th IEEE International Symposium on Wearable Computers*, pages 1–4, Washington, DC, USA, 2007. IEEE Computer Society. ISBN 978-1-4244-1452-9.

[13] G. Pirkl, K. Stockinger, K. Kunze, and P. Lukowicz. Adapting magnetic resonant coupling based relative positioning technology for wearable activitiy recogniton. In *Wearable Computers, 2008. ISWC 2008. 12th IEEE International Symposium on*, pages 47–54, 28 2008-Oct. 1 2008.

[14] J. Ross Quinlan. *C4.5: Programs for Machine Learning (Morgan Kaufmann Series in Machine Learning)*. Morgan Kaufmann, 1 edition, January 1993. ISBN 1558602380.

[15] Daniel Roggen, Martin Wirz, Gerhard Tröster, and Dirk Helbing. Recognition of Crowd Behavior from Mobile Sensors with Pattern Analysis and Graph Clustering Methods. *Networks*, 00(0):1–24, 2011.

[16] Sensolute GmbH. MVS 0608.02 Micro-Vibration Sensor Data Sheet. http://www.sensolute.com/, 2010.

[17] Emmanuel Munguia Tapia, Stephen S. Intille, William Haskell, Kent Larson, Julie Wright, Abby King, and Robert Friedman. Real-time recognition of physical activities and their intensities using wireless accelerometers and a heart rate monitor. *Wearable Computers, IEEE International Symposium*, 0:1–4, 2007.

[18] Kristof Van Laerhoven and Hans-Werner Gellersen. Spine versus porcupine: A study in distributed wearable activity recognition. In *ISWC '04: Proceedings of the Eighth International Symposium on Wearable Computers*, pages 142–149, Washington, DC, USA, 2004. IEEE Computer Society. doi: 10.1109/ISWC.2004.4.

[19] Kristof Van Laerhoven, David Kilian, and Bernt Schiele. Using rhythm awareness in long-term activity recognition. In *Proceedings of the 12th Int. Symp. on Wearable Computers (ISWC 2008)*, pages 63–68, Pittsburgh, PA, USA, 2008. IEEE Press. ISBN 978-1-4244-2637-9.

[20] Evan Welbourne, Jonathan Lester, Anthony Lamarca, and Gaetano Borriello. Mobile context inference using low-cost sensors. In *In Proceedings of LoCA*, pages 254–263. Springer-Verlag, 2005.

[21] Ian H. Witten and Eibe Frank. *Data Mining: Practical Machine Learning Tools and Techniques, Second Edition (Morgan Kaufmann Series in Data Management Systems)*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2005. ISBN 0120884070.

6 In-Situ Recognition of Single-User Activities

6.1 ABSTRACT AND CONTEXT

This chapter concerns itself with Challenge 1 (Low Power) by reducing the energy cost of embedded, online classification of local, single-user activities. While the approach itself is monolithic, it is distributed in that it is designed for embedded recognition and does not require a server. The results reveal that leveraging human predictability can yield great energy savings while only sacrificing small amounts of recognition accuracy. On the two data sets evaluated, energy savings of around 85% - 90% where achieved in turn for a loss of 1.5 - 3 percentage points of recognition accuracy.

Chap. 7 evaluates the usefulness of these activities for recognizing group activity compared to features and unsupervised clustering. There clustering is shown to be advantageous for GAR in some cases, as it avoids the complexity of having to label both single-user and group activities in parallel, therefore clustering is used for the GAR approach in Chap. 9. However, the continuing research proposed in Chap. 9 indicates that the potential applications of SAR methods combined with GAR could enable inference of the role of an individual in the group dynamic. It is therefore important to address SAR in the context of GAR to enable further research, and reducing the power consumption of the SAR process addresses the power consumption requirement for GAR in Chap. 4. Since SAR can be a component of GAR, the challenge of respecting the primary function of the device in Chap. 1 is also addressed.

This chapter is based on a publication at ISWC 2012 [13] which received an Honorable Mention and was nominated for the Best Paper award. The content presented here is an extension of that paper which was invited to a special issue in the Journal of Personal and Ubiquitous Computing [16]. An initial research expose was presented in the Work in Progress track at PERCOM 2011 [12].

6.2 introduction

As concepts from pervasive and mobile computing become more mainstream, the community seeks practical approaches for realizing pervasive technology. Situation, context or activity recognition techniques provide a method for machines to recognize human and social situations, allowing them to act proactively without contradicting or offending their users. Modern devices such as smart phones or wireless sensor nodes are now able to support these algorithms [29] as processing power and memory improve over time according to Moore's Law.

Energy storage in such devices is not subject to the same doubling effects and is quickly becoming the limiting factor in pervasive technology. This can be seen clearly when reviewing the battery lifetimes for mobile phones over the past 10 years. The cost of communication in terms of energy consumption is another factor which does not scale according to Moore's Law, indicating that for intelligent wearable applications to be practical, methods for low power situational recognition must be embedded in mobile devices.

Embedded classification for mobile devices is not a new concept and goes as far back as 1997 [8], where Bouten el al. used simple signal processing techniques to measure activity levels of users wearing a mobile device. Several methods for low power embedded context classification have been introduced in the activity and context recognition community. Cacmakci et al. [9] and Stäger et al. [28] introduce straight-forward approaches to low power recognition of contexts and activities in embedded systems using inertial or audio sensors respectively. Krause et al. [18] propose to dynamically reduce sensor sampling rates to conserve energy, thereby greatly increasing battery lifetime. A similar approach was presented by Sun et al. [30] where coarse-grained activity levels were locally recognized to adjust the sensor sample rate. Benbasat et al. [2] introduce a method for conserving energy in a system with redundant sensors which are switched on and off dynamically based on the level of activity currently measured. Roy et al. [26] use sensor configurations which are selected for specific activities based on the minimum requirements of an application. The result of the research mentioned is that there is always a trade off between how well activities can be recognized, and how much energy it costs to do it [29, 4].

We conducted a survey of work done in this field (see Sec. 6.3) revealing that focus is mostly on motion-based sensors, with the accelerometer being the most popular sensing modality. The survey also shows that initially, research was focused on custom hardware and sensor network platforms, but recently it has shifted towards mobile phones. Early methods for low-power recognition began with systems engineering, but more recently dynamic sensor selection and sample rate reduction have been the tools of choice. These dynamic approaches often use the current activity or activity level as an indicator of the optimal current configuration, or wake system states hierarchically to conserve energy.

We present a different, novel method for using context (or activity) prediction to further conserve energy. Many things we do have a certain repetitiveness or periodicity about them [32], and are therefore predictable to a certain extent. This information can be used to improve recognition abilities [23]. Context prediction is the process of using a context history to predict contexts, situations or activities which will occur in the future [21]. This can be done at several different abstraction levels [27], ranging from extrapolating raw sensor data into the future, to predicting abstract concepts such as activities. We propose that it can also be used to reduce the power consumption of the recognizing device as well.

The idea is simple. Traditionally, all sensors are used constantly even though certain sensors may only be necessary to detect specific activities. As a result, energy may be wasted when sensors are enabled which are not necessary to detect the current activities. Given a scenario where activities are performed in a manner which is predictable, probable future activities can be forecast. Sensors which are not needed to decipher probable activities from each other can be turned off (or the sample rate reduced), conserving energy without greatly impacting recognition rates. The risk, is that incorrect predictions cause sub-optimal sensor configurations, further leading to incorrect recognition and prediction.

In this work we propose that by leveraging the predictability of human actions, it is possible to tip the balance of the energy/recognition trade off to conserve energy resources without sacrificing recognition rates. We proposed this concept in a poster [12] and evaluated it initially on a single data set [13]. Since the initial publication, other research in the field has emerged [34], but an exhaustive evaluation of system behavior as presented here has not yet been conducted.

The performance is simulated using two preexisting activity recognition data sets [11, 25], where artificial data sets are generated from these sets in order to evaluate different scenario parameters. We evaluated the algorithms in terms of activity recognition rates, energy savings achieved, and the prediction accuracy with respect to system parameters. The results indicate that the novel approach allows for application and scenario-specific selection of the recognition/energy trade off, producing large energy savings, even for small recognition losses (e.g. recognition losses of 1.5 pp with 84.8% energy savings for first [11], and 2.8 pp and 89.9% for the second data set [25]).

This chapter is structure as follows. A survey of research conducted towards reducing the energy costs of embedded recognition is presented in Sec. 6.3. In Sec. 6.4 the proposed method and algorithm is presented, including the context recognition, prediction and sensor selection processes. The experiment implementation and simulation environment is presented in Sec. 6.5, along with a description of the data sets used and their preparation. Sec. 6.6 contains the results of the simulation with respect to energy consumption, classification and prediction values, the implications of which are discuss in Sec. 6.7. Finally the chapter is concluded in Sec. 6.8.

6.3 a survey of energy-efficient recognition

For embedded and mobile systems, power consumption is one of the most critical attributes [18]. We surveyed system approaches for reducing the power consumption footprint of online, embedded activity recognition in order to generate on overview of this field. To our knowledge, no such survey of applications and attributes has yet been conducted making this a novel contribution of this work.

The survey was conducted based on following parameters which where deemed to be of importance for understanding the breadth of the research. Motivation for these parameters is taken from related work, although the parameters do not cover the entire design space for such systems.

- **Platform:** This parameter describes the hardware platform used for recognition. This information is of importance as it gives the reader an indication of the amount of resources which are available for embedded recognition. For example, embedded recognition on a mobile smart phone [3] probably has an order of magnitude more processing power and memory than an embedded wrist watch platform [18].

- **Sensing Modality:** The sensors used for an application give an indicator of the order of the problem. Sensors have different properties, for example an embedded accelerometer has a far shorter startup time and power consumption [35] compared to a GPS receiver [22].

- **Conservation approach:** There are several different approaches to the problem of energy conservation with different affects on other components of the system. In some of the approaches taken to reducing power consumption, design choices are empirically explored to find systems which consume less energy. Other approaches involve designing dynamic systems which adjust themselves based on the current situation to minimize energy consumption without violating some quality criteria.

- **Control method:** of the aforementioned conservation approaches, several of them dynamically optimize certain parameters, e.g. sensor sample rate, sensor selection, or execution mode. In order to perform these operations, the decision process requires some type of input in order to close the control loop. Here the type of input used to control the conservation approach is surveyed.

Table 5.: Technical details of work surveyed in embedded context and activity recognition

Reference	Year	Sensor Modality	Platform
Bouten et al. [8]	1997	Accelerometer	Sensor node (proprietary)
Cakmakci et al. [9]	2002	Accelerometer	SoundButton sensor sode (proprietary)
Bharatula et al. [5, 6]	2005	Accelerometer, light, microphone	Sensor node (Proprietary)
Krause et al. [18]	2005	Accelerometer, microphone, light, temperature	eWatch wearable platform
Benbasat and Paradiso [2]	2007	Accelerometer, gyro, tilt switch	Gait shoe (proprietary)
Stäger et al. [29]	2007	Microphone, accelerometer, light	Sensor node (proprietary)
Thatte et al. [31]	2010	ECG, accelerometer (wireless)	Smart phone, wireless sensors
Berchtold et al. [3]	2010	Accelerometer	Smart phone
Raffa et al. [24]	2010	Accelerometer, gyroscope	Smart phone, ContextWatch (proprietary)
Lin [19]	2010	GPS, WLAN, Bluetooth, cell	Smart phone
Paek et al. [22]	2010	GPS, WLAN, Bluetooth, cell, accelerometer	Smart phone
Roy et al. [26]	2011	Accelerometer, gyroscope, light, temperature	SunSPOT sensor node
Sun et al. [30]	2011	ECG, accelerometer	Smart phone
Lu et al. [20]	2011	Microphones (internal/external)	Smart phone, sensor node (proprietary)
Gordon et al. [15]	2012	Accelerometer (wireless)	Smart phone, JenniSense sensor node
Wood et al. [34]	2012	Camera, accelerometer	Wearable camera (proprietary)
Au et al. [1]	2012	Accelerometer (wireless)	MicroLeap wearable platform
Yan et al. [35]	2012	Accelerometer	Smart phone
Gao et al. [10]	2012	Accelerometer (wireless)	Smart phone, Shimmer sensor node
Wang et al. [33]	2012	Accelerometer	Smart phone
Gordon et al. [13]	2012	Accelerometer, MVS [11], light, temp	Sensor node (proprietary)

- **Recognition Algorithm:** different algorithms are equipped to specific degrees for certain problems, therefore posing advantages in certain situations. Each algorithm is in itself a trade-off between accuracy and processing power. Certain types can handle missing features and sensors intrinsically such as nearest neighbors approaches and Bayesian inference [13] while others such as neural network approaches and decision trees must be specifically adapted for such issues [3, 30]. The selection of a classifier algorithm therefore provides insight into the energy/recognition trade-off conducted in the work.
- **Application:** which contexts or activities are recognized greatly changes the applicability and general usefulness of the approach. Algorithmic evaluations are also quite specific to the application domain. It is therefore important to note in which domain the research was conducted in order to be able to estimate usefulness in other areas.
- **Reproducibility:** one of the major issues which we see in this field is the reproducibility of results. While methodologies and algorithms may be well defined and formalized, re-implementation is time consuming and effort intensive. A system is considered reproducible if either 1) the hardware platform is available for purchase and the code basis is published, or 2) the data set on which the evaluation was conducted is publicly available.

6.3.1 *Physical Attributes*

In Tab. 5 the surveyed works are listed with respect to their technical details. As indicated there, earlier platforms were often custom built proprietary sensor nodes specially designed for the recognition operation [8, 9, 5, 6, 29], probably due to the unavailability of standardized sensing devices. Although some more recent research projects also incorporated some custom hardware [31, 24, 20, 34], a trend can be seen towards the use of mobile phones [3, 22, 30, 35, 33, 10]. Combinations of devices have also been used, where mobile phones are selected along with customized hardware for the recognition task [31, 24, 20, 12, 10].

When observing the different sensing modalities surveyed, it quickly becomes apparent that the accelerometer is the most popular sensor used. This is not surprising as embedded activity recognition has a large overlap with the wearable sensing community, where sensing motion provides great insight [3]. Often the accelerometer was used alone [8, 9, 3, 35, 33, 1, 12, 10] to recognize physical activities (see Tab. 6). Other times it was combined with other modalities to better capture physical signals [5, 6, 18, 2, 29, 31, 24, 26, 30, 13]. Accelerometers are also used to incorporate physical sensing modalities into other types of recognition systems, for example video [34] or location systems [22]. The sensors used are dependent on the application, i.e. the activities or contexts which were recognized,

Table 6.: Recognition approaches of work surveyed in embedded context and activity recognition

Reference	Recognition Algorithm	Conservation Approach	Control Method	Application	Reproducible
Bouten et al. [8]	Correlation analysis	Low-power design	None, static	Energy expenditure	No
Cakmakci et al. [9]	Bayesian inference	Low-power design	None, static	Basic physical activities	Yes (code)
Bharatula et al. [5, 6]	Decision tree	Low-power design	None, static	Office activities	No
Krause et al. [18]	SVM	Adaptive sample rate	Current activity	Basic physical activities	No
Benbasat and Paradiso [2]	CART decision tree	Adaptive sample rate, hierarchical wake up	Activity level	Wearable gait monitoring, animal monitoring	No
Stäger et al. [29]	C4.5 decision tree	Adaptive sample rate	None, static	Kitchen activities	No
Thatte et al. [31]	SVM, Bayesian inference	Adaptive sample rate	Current activity	Basic physical activities	No
Berchtold et al. [3]	Fuzzy inference	Modular classifier pipeline	Current activity	Basic physical activities	No
Raffa et al. [24]	HMM	Hierarchical pipeline	Activity level	Gesture recognition	No
Lin [19]	Bayesian estimation	Adaptive sample rate	Current location, accuracy requirement	Location	No
Paek et al. [22]	Onset detection	Adaptive sample rate	Location	Location	No
Roy et al. [26]	Classifier independent	Adaptive sample rate, sensor selection	None, static	Basic physical activities	No
Sun et al. [30]	Decision tree	Adaptive sample rate	Activity level	Basic physical activities	No
Lu et al. [20]	Bayesian inference	Hierarchical wake up	Speech (activity) level	Speaker identification	No
Gordon et al. [15]	DT, nB, kNN	Data preprocessing	None, static	Group activities	Yes (data)
Wood et al. [34]	K-Means	Adaptive sample rate	Predicted future activities	Sleep, Office, Cycling	No
Au et al. [1]	HMM	Adaptive sensor selection	Current activity	Basic physical activities	No
Yan et al. [35]	Bayesian inference	Adaptive sample rate	Current activity	Basic physical activities	No

Gao et al. [10]	nB, DT	Sensor selection	Current activity	Basic physical activities	No
Wang et al. [33]	semi-Markov process estimation	Adaptive sample rate	Current activity	Movement detection	Partly (data)
Current Work [13]	HMM, kNN	Adaptive sensor selection	Predicted future activities	Basic physical activities	Yes (Data)

where the effectiveness of the sensing modality is given by the influence of the activity on that sensor.

6.3.2 *Recognition Attributes*

Early systems where exploratory in nature and investigated the performance of recognition under constrained resources. Here, the focus was on system design, where the hardware/software architecture was constructed in such a way as to maintain low consumption using standard algorithms with little or no adaptation at run time [8, 9, 5, 6, 29].

The effects of reducing sensor sample rate to a reduced but constant value throughout operation with respect to energy consumption and recognition loss indicate that for certain applications it can be advantageous [18]. A further improvement can be achieved by adapting the sample rate of the sensors which reduces their consumption. Adaptive sampling, however, requires a control parameter to set the sample rate correctly so as not to cause deterioration in recognition rates. In the surveyed work, this has been done using using some indicator of the current system state, such as an activity level indicator [2, 30], or an actual recognized activity [18, 31, 22, 35, 33] or location [19], or a combination of those [22]. This method is dependent on the sensor modalities (see Tab. 5) with respect to the warm up times and power consumptions for effectiveness.

A second method for reducing energy is to dynamically select sensors to be activated or deactivated during a certain period of time [1]. Here again, a method for selecting sensors for each classification is needed, where Au et al. propose using the last recognized activity to conduct this. Theoretically, once the sensors have been activated, there is no reason why methods of adaptive sampling cannot be applied, although this has not been evaluated in the work surveyed.

Another method for conserving energy is to break the recognition process down into a hierarchical pipeline, where each level activates higher-order processes under certain conditions, thus avoiding superfluous operations if lower levels deem them unnecessary. These methods can range from activating the system based on activity indicators, i.e. activity level for waking up activity recognition [2, 24], modular classification systems [3], movement detection for activation of more expensive location sensors [22], or sensing the presence of voices for speaker identification [20].

In Tab. 5, some instances employ wireless sensors which seems to indicate a distributed approach as apposed to an embedded one [31, 12, 1]. In these works the authors observed the systems as being closed, i.e. the energy consumption of both classification and sensor usage was investigated, and are therefore relevant to this survey. The methods then used for energy consumption optimization must account for the same problems as embedded systems and therefore employ the same

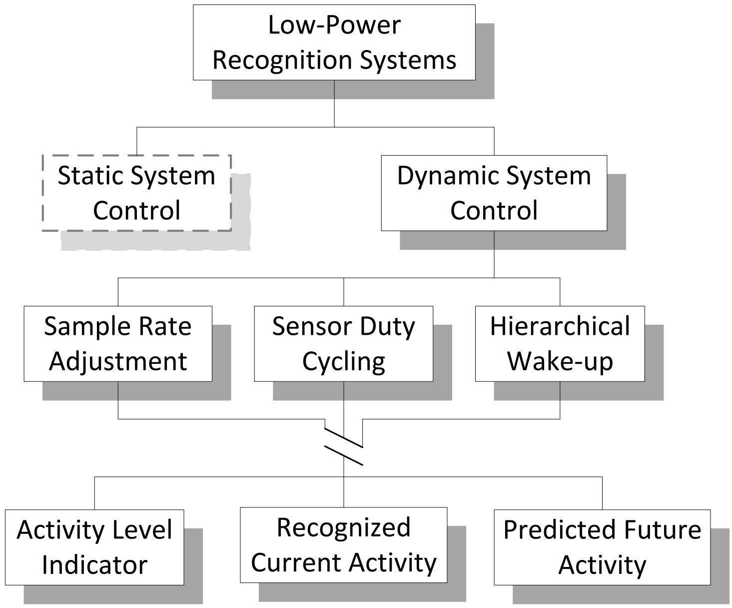


Figure 9.: A taxonomy for low-power, embedded activity recognition

techniques, e.g. adaptive sample rates [31] or sensor selection [1]. One difference is that the volume of sensor data has a great affect on system consumption, which can be addressed using preprocessing [14, 15, 10].

6.3.3 *Summary of the Survey*

To summarize the survey, when designing low-power embedded recognition applications, the system should be designed to reduce overall consumption by using low power components and algorithms. Once this has been accomplished, further savings can be achieved by making the system dynamic in nature to adapt to changing requirements. Here there are 3 methods for energy conservation which have been used. First, the sample rates of the sensors can be dynamically adapted to current requirements, where reducing the sample rate also reduces energy consumption. Second, the designer can opt to turn off sensors entirely for a sample period, further reducing energy consumption but risking deterioration of recognition values if not done correctly.

Third, a method for structuring recognition components can be implemented in a hierarchical way, such that low power components wake those with higher consumption only when they are needed. Often times the third method benefits from dedicated hardware components which conduct low-power listening for activity cues. These three methods can also be combined with each other to further improve consumption rates. Each of these methods has different advantages and disadvantages under different conditions and scenarios. Unfortunately, based on previous publications it is not possible to make comparative statements about the energy consumption and recognition rates for these methods.

Once these design choices have been made, a method for controlling the dynamic conservation approach can be selected. The first method is to use a general indicator of the activity level to control the conservation approach. An activity indicator is especially effective when combined with a hierarchical wake-up pipeline due to the low computational complexity of these indicators [20]. Another method is to use the current activity which has been recognized to adapt the system to the requirements of recognizing that activity. Here the risk is that since an activity must first be detected before adaptation, the system has issues with detecting activity transitions. This design decision structure can be represented as a taxonomy of approaches to embedded recognition which is show in Fig. 9. The static system control mode has been deprecated to indicate that it is not the focus of this survey and is therefore not exhaustive.

In this work, we examine a new method for controlling the energy conserving mechanism. The method is not specific to the energy conservation approach and can be combined with adaptive sampling, sensor selection, or a pipelined activity recognition chain [12, 13, 34], but is evaluated here with a sensor selection approach. Using context prediction, a future probability distribution can be generated which allows the system to be proactive in nature [27], instead of only reacting to the current system state. Using future activities eliminates the lag incurred by having to recognize the current activity state or level [27] before being able to react to it. This can improve the power consumption footprint and the correctness of the recognition during the lag, or activity transition period. However, incorrect predictions may lead to mis-configurations, a research question which is evaluated in the rest of this work.

6.4 PROPOSED ALGORITHMIC APPROACH

The standard process for activity recognition using machine learning algorithms is straightforward. Sensors are sampled in parallel at an arbitrary but constant rate for a period of time. The data is then saved as a discrete multidimensional vector, referred to as a sample window. This window is processed using different algorithms to generate signal features, e.g. standard deviation, average, FFT or cepstral coefficients. Which features are used depends on the application (i.e. which activities we want to recognize), and the type of sensors being used, and are referred to all together as a feature vector. A machine learning algorithm is given the task of recognizing which activity was occurring during the sample window, based on its feature vector.

We propose integrating prediction into the process to improve energy consumption as demonstrated in Fig. 10. First, *activated* sensors are sampled to generate a sample window. The sample window is then processed into a feature vector, and *classified* as to which activity is being performed. Based on the classification history, future activities which are likely to occur are *predicted*. An appropriate sensor configuration is then activated to distinguish only the *likely activities*, and the process repeats itself.

During the course of this research, we identified three parameters which affect the trade off between energy and recognition. The first is the **predictability κ** of the sequence of activities, or the inherent predictability of the scenario itself. A low value for κ indicates that prediction results are little better than random, where a $\kappa = 1$ indicates a 100% prediction accuracy. In real world scenarios, κ simply equates to the prediction rate for a given predictor and scenario. This parameter cannot be influenced by the designer, and can only be quantified by analyzing the scenario and predictor beforehand. The second parameter affecting performance is **ρ, the number of classes which are predicted** at each time step. The more classes which are predicted, the better the chance that the next class is actually among the predicted classes (correct prediction), but the lower the savings will be as the system accounts for more possible activities. Therefore ρ specifies the level of risk, which allows the designer to tip the odds towards recognition or energy as will be seen later. The third parameter is application specific, and is referred to as the **loss parameter λ**, which specifies the amount of recognition which can be sacrificed in order to conserve energy without breaking the application's requirements. A λ value of 0 indicates that optimizations causing any loss at all, however minimal, are not acceptable, and $\lambda = 1$ means energy savings are of the utmost priority, and recognition rates are of no importance.

A useful analogy at each prediction/classification step is that of a wager. Here κ, the predictability of the scenario, can be thought of as the probable outcome of the bet based on previous experience (prior distribution). The number of classes predicted, ρ, allows ρ different outcomes to be bet on at once: the higher ρ is, the better the chances of a correct bet, but the lower the payout in terms of energy saved. In this case, the wager λ is a specified amount of the total recognition rate, and the payout is in energy savings. Losing a bet (meaning a false prediction) is detrimental to classification lowering overall recognition rates.

6.4.1 *Weighting Sensors to Activities*

Here we will present the method for selecting which sensors to activate based on predicted activities. When observing the chain of events in the context classification process, each feature in the set of features used $f \in F$ is implicitly mapped onto a single sensor in the set of sensors $s \in S$. That sensor generates the data for this

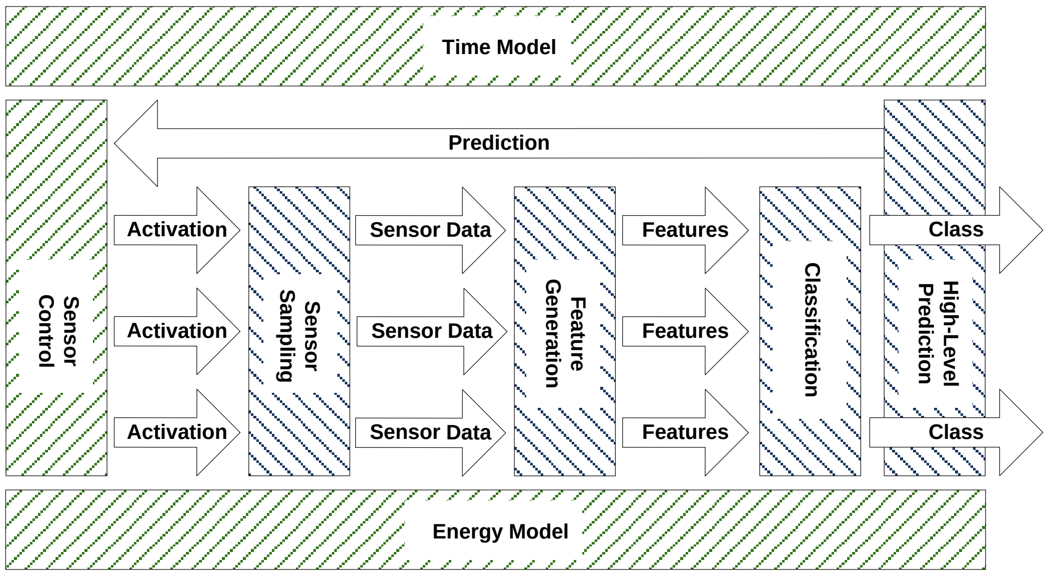


Figure 10.: The novel algorithm (\) and simulation environment (/)

feature, producing the surjective mapping a of features onto sensors $a\colon F \to S$. Reversely, each sensor s_i is then "responsible" for a subset of features $\widetilde{F}_{s_i} \in F$, meaning the features in $\widetilde{F}_{s_i}$ are generated over data stream from sensor s_i.

Mapping activity classes onto the sensors over the features is not as simple. This mapping cannot be carried out independent of the classifying algorithm, as each algorithm has a different method of measuring the distance between two vectors. For example, nearest-neighbors algorithms use a multi-dimensional distance measurement, often euclidean distance, between two vectors to separate them, probability-based models calculate where a vector lies in the probability distribution for a specific class, and decision trees often use entropy as an indicator of distance [7]. An overview of selecting features which best suite an embedded application is presented by Könönen et al. [17], providing a *sensor to application* mapping. While these algorithms potentially improve the quality of classification and reduce the computational load, they do not provide a mapping of features to classes by relevance or importance. A method for generating a *sensor to class (activity)* mapping by relevance or importance was proposed by Roy et al. [26], which they referred to as quality-of-inference (QoINF). As will be discussed later, this method is not effective for the approach and data set presented here.

Turning sensors on and off will result in a dynamic feature vector length, and for this reason we will consider standard classifiers which can natively support this. Specifically, nearest-neighbor classifiers are well suited to this task as omitting a feature represents a dimensional reduction of the labeled training vector space, and

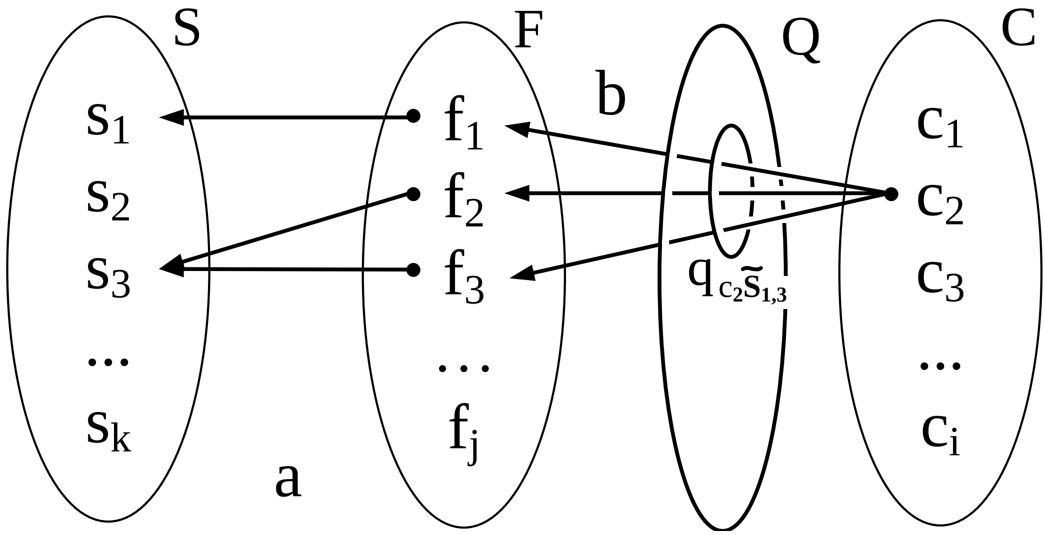


Figure 11.: Features ($\boldsymbol{F}$) from sensors ($\boldsymbol{S}$) $\boldsymbol{s_1}$ and $\boldsymbol{s_3}$ have value ($\boldsymbol{Q}$) $\boldsymbol{q_{c_2\tilde{s}_{1,3}}}$ for class ($\boldsymbol{C}$) $\boldsymbol{c_2}$

the missing features are simply excluded from the distance calculation. Probabilistic models are also well suited as the observational distributions for missing variables can be ignored when calculating the probabilities of the hidden states. Both examples lose only the information that would have been gained from the missing features, but are not negatively affected further [7].

In order to generate the weighted mapping Q (the weight is the dependency of activities on sensors), training data is gathered for each class. Weight calculation was done by testing the trained classifier against all training vectors for each class and simulating different feature combinations. Selected features were turned off and the dependency of each class on those features was evaluated. The degree of dependency is the drop in accuracy compared to the full feature vector: a large drop in recognition indicates a high dependency, a small drop, a low dependency.

Initially the intent was to only evaluate the weight for each feature individually. The cost/dependency weights for a sensor could then be calculated by summation of the weights of its features, assuming $q_{cf_i} + q_{cf_j} \approx q_{cf_{ij}}$ as indicated by Roy et al. [26], and the cost of turning off two sensors, is the cost of the one plus the cost of the other. This however proved to be too inaccurate to be useful due to the conditional dependence of features and sensors, making $q_{cf_i} + q_{cf_j} \leq q_{cf_{ij}}$ [17]. Therefor, Q values were calculated for each class against all possible sensor subsets directly, instead of by summing single feature or sensor values.

In order to correctly estimate the optimal sensor subset $\widetilde{S}$ for a sensing and classification step, the matrix Q must be calculated only once at training time. The resulting mappings can be seen in Fig. 11, showing one mapping of class $c_2 \in C$ onto sensors $\widetilde{S}_{1,3} \in S$ over features generated from $\widetilde{S}_{1,3}$. Each mapping in $b\colon C \xrightarrow{Q} F$ represents one element in the Q matrix, in this case $q_{c_2 \widetilde{S}_{1,3}} \in Q$. The Q matrix is indexed by the power set of S without the empty set, or $\widetilde{S} \in \wp(S) \backslash \{\}$, and the classes $c \in C$, resulting in a $|C| \times (2^{|S|} - 1)$ matrix. The value at each point i, j indexed by c_i and $\widetilde{S}_j$ is the recognition loss when classifying c_i using sensor subset $\widetilde{S}_j$ compared to using all sensors S over a set of evaluation data samples. Now, for each class $c_i \in \widetilde{C}_{t+1}$ where $\widetilde{C}_{t+1}$ is the set of activities predicted to occur at the next time step, a set of sensors $\widetilde{S}_{t+1}$ can be identified which is optimal with respect to λ. This is accomplished by selecting the sensor subset $\widetilde{S}$ for the next period $t+1$ such that it fulfills Eq. (7).

$$\widetilde{S}_{t+1} = \underset{En(\widetilde{S})}{\arg\min}, \forall_{c \in \widetilde{C}_{t+1}} q_{\widetilde{S},c} \leq \lambda \tag{7}$$

Where $En(\widetilde{S})$ is the combined energy consumption of all sensors in $\widetilde{S}$. Simply put, in order to distinguish the classes predicted to occur $\widetilde{C}_{t+1}$ from each other, the sensor configuration $\widetilde{S}$ is selected which saves the most energy $En(\widetilde{S})$ without violating the acceptable loss parameter $q_{\widetilde{S},c} \leq \lambda$ for *any* of the predicted activities $c \in \widetilde{C}_{t+1}$. This selects the sensor configuration with the lowest energy consumption that is still capable of recognizing the predicted classes with acceptable recognition rates. The next section will analyze the use of context prediction to generate a set of classes which are likely to appear in the next sample window ($\widetilde{C}_{t+1}$).

6.4.2 *Context Prediction*

Context prediction is used to estimate a subset of all contexts or activities $\widetilde{C}_{t+1} \in C$ which are most likely to occur at the next time step $t+1$. The cardinality of $|\widetilde{C}_{t+1}| = \rho$ is a parameter which can be adjusted, and allows the designer to select the recognition accuracy risk against the energy reward as will be shown in Sec. 6.6. This approach is independent of the algorithm or abstraction level used for prediction. Important is only the quantification of the predictability parameter κ which is simply an indicator of how well the predictor is able to forecast the given scenario (predictor accuracy). The results presented here should therefore still apply for all scenarios and prediction algorithms.

As indicated by Fig. 10, high-level context information at the activity or context abstraction level is used for prediction. Using low-level, sensory or feature data is also an option, but high-level prediction reduces complexity in terms of training and execution [27]. The algorithm used for prediction is a first-order Markov chain consisting of states $c \in C$. At each time step, the probability $P(c_{i,t+1}|c_t)$ for each $c_i \in C$ is calculated, and the ρ states with the highest probabilities are output as predictions.

6.5 implementation and simulation

This section presents the algorithmic implementation and the simulation environment. Both were programmed using the Python programming language.

6.5.1 *Simulation Environment*

The main concept is to leverage the predictability of human actions in order to conserve a large amount of energy while only sacrificing a small amount of recognition capabilities. The simulation environment was designed to evaluate the method for various degrees of predictability κ. Two published data sets which are publicly available where used in this evaluation.

The data sets where selected because both of them are publicly available sets of numerical data gathered from wearable sensing modalities, making the results presented here easier to reproduce. The MVS data set contains a relatively large number of activities (8), but relatively few sensor modalities. The OPP data set contains relatively few locomotion activities, but with a large number of sensing modalities and locations. The goal was to select data sets which complement each other so as to demonstrate different performance aspects of the proposed approach under different conditions.

micro-vibration sensor data set (mvs): The first data set used for evaluation [11] contains 142 minutes of data from 4 sensors (see Tab. 7), sampled from 5 subjects performing 8 activities (taking a bus, riding a bike, walking, jogging, taking the elevator, typing at a desk, going up/down stairs, and standing).

the opportunity data set (opp): The second data set used for the evaluation [25] contains information from 72 sensors over 12 subjects, yielding 25 hours of data. In order to reduce complexity of the simulation a subset of the sensors was used. The sensors used where the following: acceleration on the hip, right knee, back, right hand and left hand; gyroscope and magnetic field sensors on the back. The data set contains a myriad of labels, including locomotion modes, object interactions, interaction types, and which hand was used. For this evaluation the locomotion mode labels where used, consisting of 4 activities (walk, stand, lie and sit). The system was simulated as a single device connected to all sensors. While the capability of an embedded processor to handle this number of input streams

Table 7.: Energy consumption rates for the simulated hardware components

Element		Dimensions	Energy Cost (mW)		Data Set	
Function	Name		Online	Offline	MVS [11]	OPP [25]
Light	APDS-9003	1	8.25	0.0	√	
Temperature	TC1047	1	0.1155	0.0	√	
Vibration	MVS 0608.02	1	0.0015	0.0	√	
Microprocessor	PIC18LF14K	1	0.0512	0.0	√	√
Acceleration	ADXL335	3	1.4	0.0	√	√
Magnetic Field	HMC 5883L	3	0.33	0.0066		√
Gyroscope	ITG-3200	3	21.45	0.0165		√

or data is questionable, this does not affect the results of the evaluation as the processor is modeled as being always on with constant consumption. The energy consumption values where taken from a standard microprocessor and sensors, identical to that of the MVS data set (see Tab. 7). Actual device specifications and consumption values are not included in the OPP data set.

Both data sets where evaluated using the same preprocessing framework. The system simulates real time through a replay mechanism using the recorded data. The respective data set is cut up into one second windows without overlap, over which features are generated. The resulting feature vectors are then fed to the novel algorithms as if they were being generated in real time. The sensor configurations are simulated, where for a specific sensor configuration $\widetilde{S}$, the features $\widetilde{F}_{\widetilde{S}}$ are present in the feature vector and all others are omitted. Once a sensor configuration $\widetilde{S}$ has been selected, the features $\widetilde{F}_{\widetilde{S}}$ are calculated. Per sensor, the following features are calculated [11]: average, standard deviation, area under the curve, min-max difference, Shannon entropy, and FFT peak.

The energy consumed by the device $En(\widetilde{S})$ is recorded for the time step. The total consumption consists of the consumption of each sensor, as well as the energy consumption of the microprocessor during the course of the sample window, or one second. The energy model is simplistic, ramp-up and ramp-down times/consumptions of the sensors are not modeled, and the processor consumption is modeled as being constant regardless of load. This approximation does not account for the added load of prediction, but the method used here has a computational complexity of only $\mathcal{O}(|C|)$ [12]. The energy consumption rates for each device simulated can be found in Tab. 7. At each time step, a new $\widetilde{S}_t$ is provided by the algorithm, which results in a different feature vector consisting of features $\widetilde{F}_{\widetilde{S}}$, and a different energy cost. The amount of energy consumed can then be compared with the amount consumed for the reference case when $\widetilde{S} = S$, i.e. when all sensors remain on, for comparison.

As with energy consumption, the simulation environment also records classification results, both for the novel algorithm and for the reference case. For each time step, the algorithm classifies $\widetilde{F}_{\widetilde{S}}$ and the result of the classification is recorded, along with the energy consumption. At the same time, the complete feature vector F_S, consisting of the entire feature set F is also classified and the result is stored for comparison with the reference system. In total, the simulator records the energy consumption and classification results for both the novel prediction-based activity recognition algorithm, as well as the reference case when all sensors remain on for both data sets.

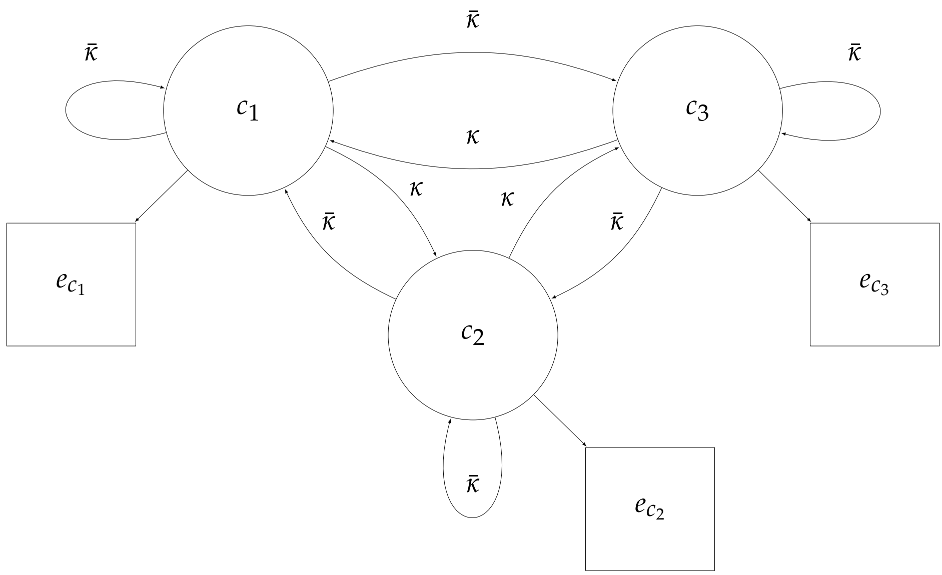


Figure 12.: Generative model for constructing an artificial data set with 3 classes ($c \in \boldsymbol{C}$), emissions ($e \in \boldsymbol{E}$) and predictability $\boldsymbol{\kappa}$

6.5.2 *Artificial Data Set Generation*

In order to evaluate the behavior of the system for different degrees of predictability (κ), artificial data sets are generated using the original data set and a generative probabilistic model shown in Fig. 12. The goal is to generate a data set which is predictable to a specified degree by the predicting algorithm, meaning that it results in a certain prediction accuracy. A Markov chain assumes that the process being modeled holds with the Markov property. It follows that by changing how pronounced the Markov property is in the data, the accuracy of the predictor can be set. The predictability is defined as $\kappa \in [\frac{1}{|C|}, 1]$ where a value of 1 indicates that

$$\forall i \exists j | P(c_{t+1} = c_j | c_t = c_i) = 1$$

and a value of $\frac{1}{|C|}$ indicates that

$$\forall i, j | P(c_{t+1} = c_j | c_t = c_i) = \frac{1}{|C|}$$

or that all transitions are equally likely. Setting $\kappa = \frac{1}{|C|}$ is the lower bound for predictability, as there are $|C|$ transitions leaving each state, and the probabilities of all exit transitions sum to one. Assigning κ a lower value than this means at least one exit transition must have a probability higher than $\frac{1}{|C|}$, increasing predictability.

Using κ, we can generate a HMM (not to be confused with the HMM used for recognition) by ordering states such that each state has 1 and only transition to a different state with probability κ, and only 1 transition from a different state to itself with probability κ. All other transitions have probability $\bar{\kappa} = \frac{1-\kappa}{|C|-1}$. Simply put, as κ approaches 1, the state following the current state becomes more and more certain, and therefore easier to predict. As κ approaches the lower bound of $\kappa = \frac{1}{|C|}$, the next state becomes more random, and harder to predict.

Once this model is created, traversing it generates emissions which are sample windows from the original data set for the given activity. This is demonstrated in Fig. 12 for an example 3-class dataset. Although this artificial data set does not represent a realistic pattern of the human activities in the data set, it does create a data set which is predictable to a specified degree. As will be shown later, the results are only dependent on the prediction accuracy, meaning that for real-world scenarios with identical prediction rates, the results should still hold.

6.5.3 *Experimental Process*

The algorithm presented here is not application specific. It is meant to reduce the cost of embedded activity and context recognition in scenarios with repetitive temporal patterns. Each application is different in terms of the optimal trade off between energy consumption and accuracy [29]. The following evaluation is conducted without a specific cost model, but allows the reader to evaluate the effectiveness for their application scenario at hand.

The classifiers used are the Hidden Markov Model [23] (HMM), and the k-Nearest-Neighbors (kNN) [7] algorithms, as they are both easily adapted to a variable feature vector length. The algorithm requires two separate sets of training data, one to train the classifier and predictor, and a separate one to populate the Q matrix using the trained classifier. A third data set is required for evaluation.

TRAINING PHASE Each artificial data set is partitioned into 3 sections. The data used to train the classifier and predictor $\widetilde{D}_{Train}$ makes up 60% of the original data set D. Another 20% $\widetilde{D}_Q$ is used to calculate the Q matrix, as using $\widetilde{D}_{Train}$ for this purpose results in overfitting, and therefore distorted loss values in Q. Finally, the last 20% $\widetilde{D}_{Eval}$ is used to evaluate the performance of the whole system, and in this experiment contains 3595 sample windows in total.

In the first step, $\widetilde{D}_{Train}$ is used to train the classifier, either HMM or kNN, as well as the Markov chain used for prediction. In this phase sensor selection is not conducted and both of the classifier instances and the predictor are trained on all features $f \in F$. In the second training step, $\widetilde{D}_Q$ is used to populate the Q matrix by evaluating the recognition rate of every class c_i with every permutation of $\widetilde{S}$. Therefore, every combination in $C \times \wp(S) \backslash \{\}$ is evaluated in a separate classification phase, using all vectors for activities of the current subset.

TESTING PHASE In the testing or evaluation phase, the classifier algorithms are run on $\widetilde{D}_{Eval}$ in parallel. At each classification time step, the $\widetilde{S}_t$ resulting from the previous time step is used to generate a new feature vector $\widetilde{F}_t$. This vector is then classified, either by the HMM or kNN classification algorithm. Based on this classification, the algorithm predicts ρ probable classes $\widetilde{C}_{t+1}$ for the next time step. Next, the sensor subset $\widetilde{S}_{t+1}$ is selected such that it fulfills Eq. (7). At the end of each step the simulation environment records the classification result using F, $\widetilde{F}_t$, the ground truth for that sample window, the sensor subset $\widetilde{S}$, the energy consumed $En(\widetilde{S}_t)$ and the predictions $\widetilde{C}_{t+1}$ for the next time step.

6.6 EVALUATION

The evaluation presented here covers several months of simulation time run on a quad-core desktop computer. For each different degree of scenario **predictability** κ (MVS: from 0.125 to 0.875 step 0.125, OPP: from 0.25 to 0.875 step 0.125), a different artificial data set was generated. The **number of predicted states** ρ (MVS: from 1 to 8 step 1, OPP: from 1 to 4 step 1), the **acceptable loss parameter** λ (from 0 to 1 step 0.1) and the classifier (MVS: HMM and kNN, OPP: kNN) were permuted to evaluate the output parameters over each data set. The OPP data set was only evaluated using the one kNN classifier to maintain brevity, and as it suffices to validate the conclusions drawn from the MVS results. These results are multi-dimensional, consisting of dimensions ρ, λ and κ, the classifier, data set, recognition rates and energy consumption. It is impossible to impart this information in its entirety here, therefor we will detail and demonstrate major insights with graphical excerpts.

6.6.1 *Results of the MVS Data Set*

MVS: Recognition Loss

We define recognition loss as the difference in percentage points (pp) between the reference recognition rate (in percent) with all sensors on, and the recognition rate for the novel algorithm for a given set of parameters. For a given loss parameter λ, in can be observed that loss of recognition decreases monotonically (meaning recognition increases) for an increasing ρ (number of states predicted). For the MVS data set, this is demonstrated by Fig. 13 for a κ of 0.125, and again in Fig. 14 for a κ of 0.875 for both the HMM and kNN classifiers. This is again evident in Fig. 15, where for a given λ, increasing ρ either reduces or leaves recognition loss

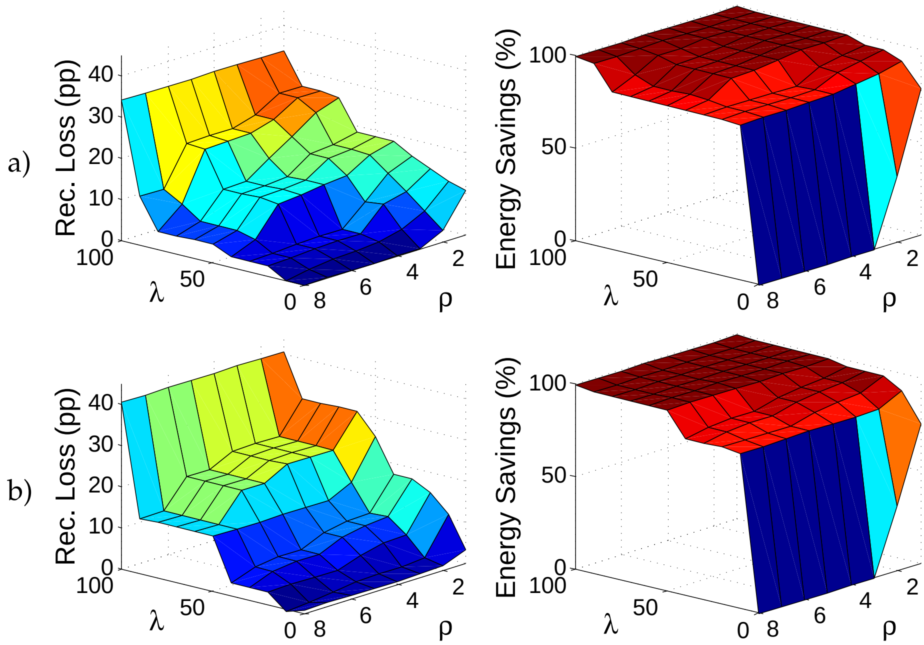


Figure 13.: Recognition loss and energy savings for the MVS data set and the HMM (a) and kNN (b), κ=0.125

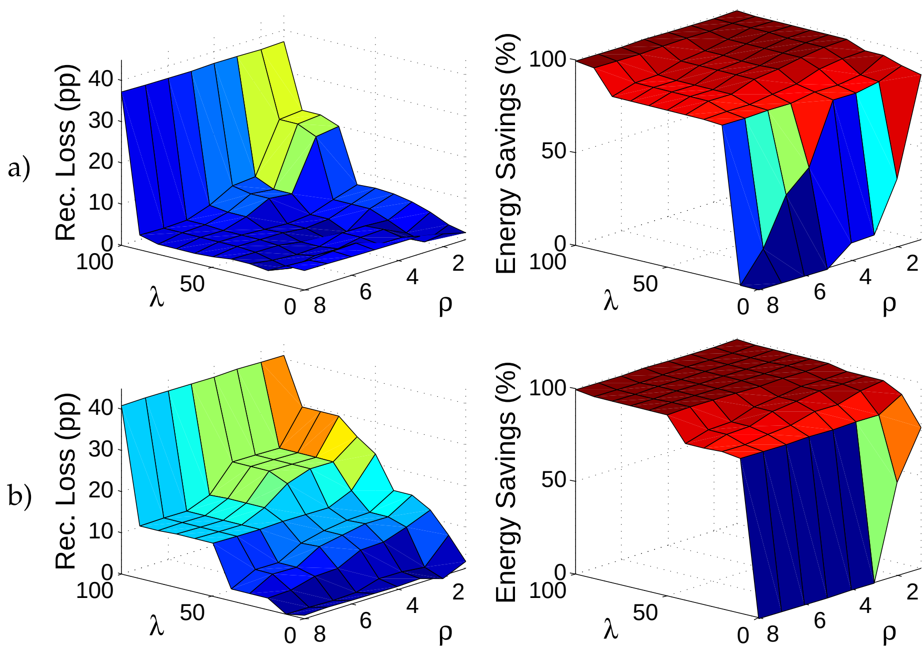


Figure 14.: Recognition loss and energy savings for the MVS data set and the HMM (a) and kNN (b), κ=0.875

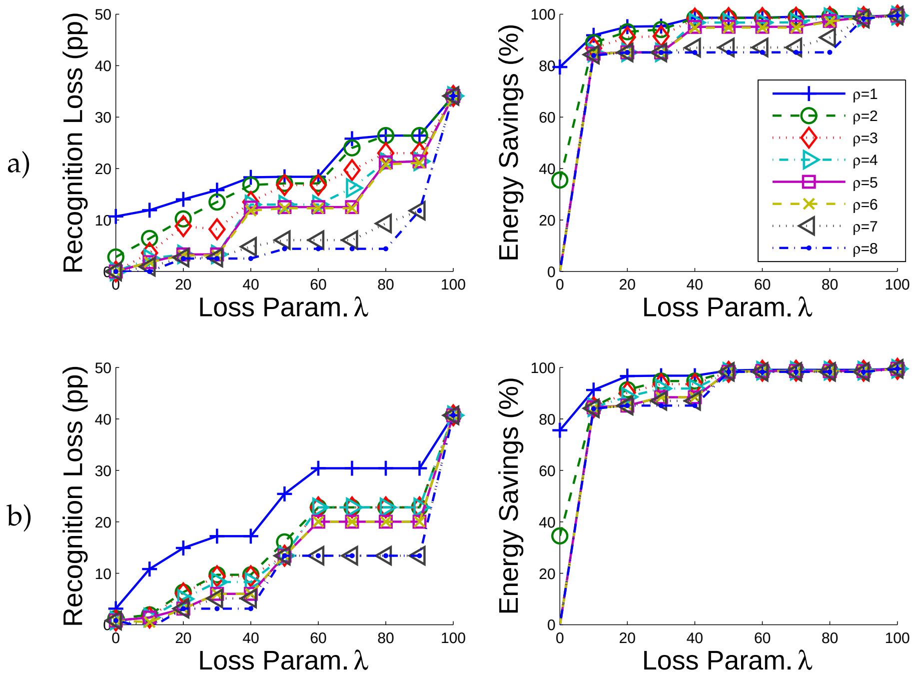


Figure 15.: Recognition loss and energy savings for the MVS data set and the HMM (a) and kNN (b), κ=0.125

unchanged. In other words, for a specific number of classes predicted at each step (ρ), if the parameter λ which identifies how much loss is acceptable for a specific application is increased, the loss in recognition does indeed increase.

The same also applies to the acceptable loss λ, where for a given classifier and ρ, loss in recognition and energy savings increase monotonically with λ. The implication is that the acceptable loss parameter λ does indeed function as an indicator for how much loss can be sacrificed as proposed. The monotonic behavior of recognition loss implies that for a given predictability κ, the lowest recognition loss (best recognition) is obtained by $\rho = |C|$ and $\lambda = 0$, and the highest loss (worst recognition) when $\rho = 1$ and $\lambda = 1$.

Observing accuracy loss over κ for fixed values of λ and ρ is not as clear cut. In Fig. 16, varying κ affects recognition for $\rho = 4$ using the HMM, where the trend in recognition loss is decreasing as κ increases, although not monotonically (compare $\kappa = 0.125$ with $\kappa = 0.875$ for $\lambda = 0.8$). For the kNN classifier, the effects of κ are minimal when compared to the HMM as seen in Figs. 13 and 15.

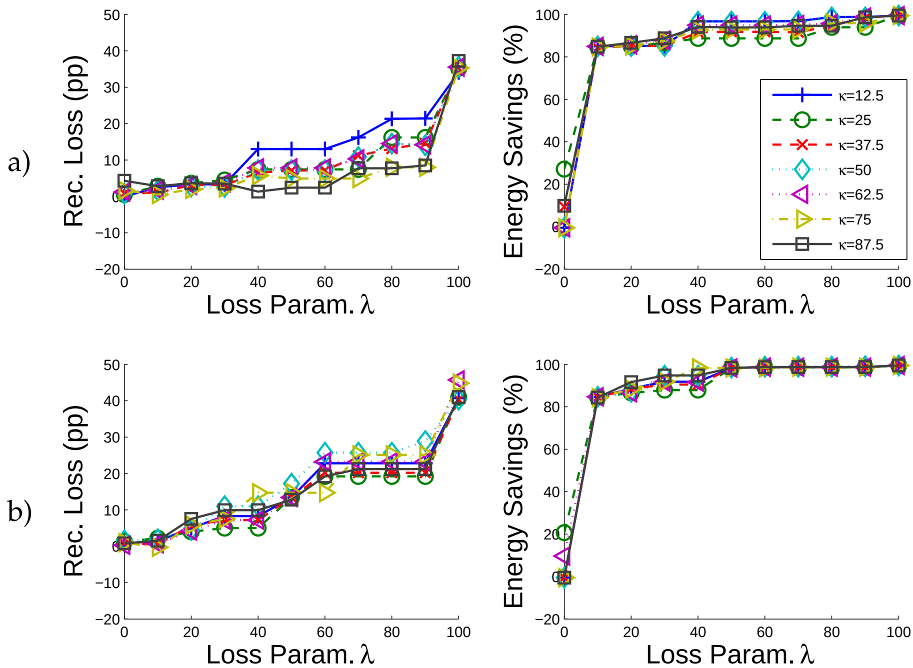


Figure 16.: Recognition loss and energy savings for the MVS data set and the HMM (a) and kNN (b), $\boldsymbol{\rho}$=4

MVS: Energy Consumption Rates

The energy savings is defined as the relative decrease in energy consumed over the evaluation of $\widetilde{D}_{Eval}$ between the reference classifier with all sensors on and the novel algorithm. When observing Fig. 13 and Fig. 14, the acceptable loss parameter λ has a far greater influence on energy savings than either ρ or κ. Fig. 16 and Fig. 15 demonstrate this by showing very little differentiation in energy savings for either κ or ρ respectively. In all cases, a relatively small values of λ (≈ 0.1) suffice for large energy savings (>80%).

All of the images displaying the results clearly show a rapid increase in energy savings for even small acceptable loss values. This increase is caused by the light sensor, which consumes an order of magnitude ($5,500$ times) more energy than the vibration sensor for example (see Tab. 7). The upper bound for energy savings, as well as for recognition loss, is given by using the sensor with the lowest power consumption only, namely the MVS vibration sensor [11]. The light sensor is the first to be shut off, creating the steep climb over low values of λ seen clearly in Figs. 13 and 14. Another slight increase can be seen around $\lambda = 0.5$ corresponding to the acceleration sensor. Shutting off this sensor however, causes large increases in

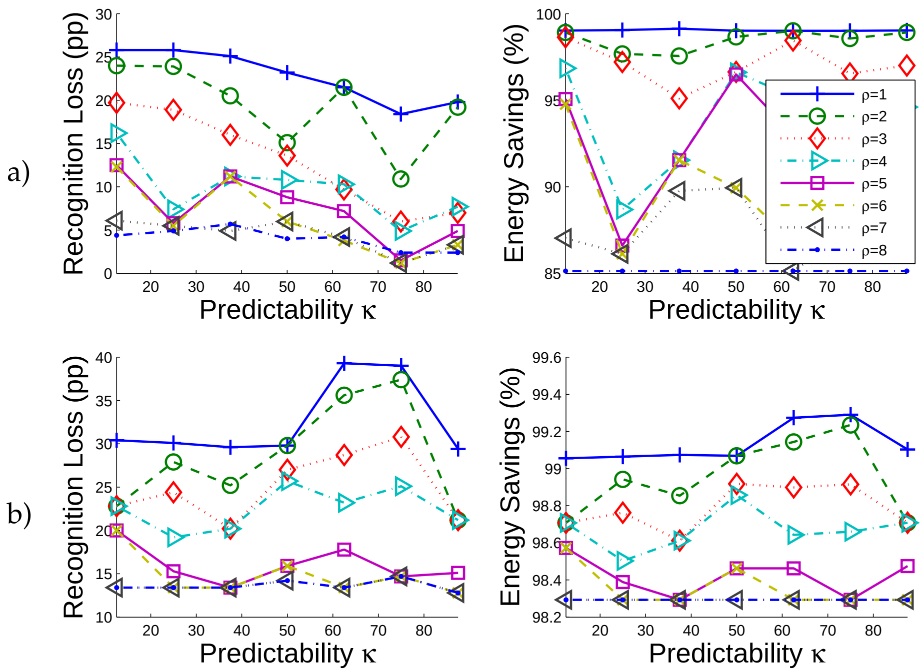


Figure 17.: Recognition loss and energy savings for the MVS data set and the HMM (a) and kNN (b), λ=0.8

recognition loss. In other words, the algorithm filters out those sensors first which contribute little, but cost a lot.

MVS: Classifier Comparison

The kNN classifier performance for the reference case (all sensors on) remained stable across κ with recognition rates between 79.6 - 80.1%. On the other hand, reference recognition rates for the HMM varied in performance from 70.5% for $\kappa = 0.125$ to 81.6% for $\kappa = 0.875$, indicating that the recognition rates of the HMM are quite dependent on the predictability of the scenario. This can be seen again in Fig. 16, where recognition losses vary little for all values of κ for the kNN classifier, but are further spread out for the HMM classifier. However, the recognition loss for the HMM is consistently higher than for the kNN classifier for the same parameters. This can be seen when comparing the top left and bottom left images in Figs. 13 and 14.

On the other hand, the kNN classifier appears to be consistently better at conserving energy than the HMM classifier, as seen in Fig. 15 when comparing energy savings of the HMM and kNN classifiers for $\lambda = 0.1$ or $\lambda = 0.6$ for example. Fig. 14 demonstrates that this is also evident for other values of κ. Both Fig. 13 and

Fig. 14 indicate that the energy consumption of the kNN classifier is also less for higher values of ρ, staying constant where the HMM energy savings fall off. Fig. 17 confirms this (noisily) by indicating higher savings for the kNN classifier compared to the HMM, and less variance over κ for higher values of ρ.

MVS: Prediction Rates

One potential issue which was mentioned earlier is that incorrect predictions can lead to incorrect sensor configurations and incorrect classifications. The apparent problem is that this can then again lead to another incorrect prediction, fueling the cycle. To evaluate the effects of this phenomena, the prediction rates of the system where also evaluated with respect to κ, ρ and λ.

Fig. 18 shows the prediction accuracy with respect to ρ and λ for the a) HMM and b) kNN classifiers with a predictability of κ=0.125. On the left hand side, the prediction accuracy with respect to the ground truth is shown, while on the right the accuracy with respect to system classifications is displayed. The latter indicates the correctness of prediction as seen from the subjective point of view of the algorithm's own classifications. When observing these graphs, the first thing which is clear is that the prediction accuracy is heavily dependent on ρ, or the number of states predicted. The linear relation between ρ and prediction accuracy is to be expected. For $\lambda = 0$, the ratio of correct to incorrect predictions should range from κ for $\rho = 1$, to 1 for $\rho = |C|$, where since all classes are predicted, the prediction is always correct. This is evident in Fig. 18, where the deviance in prediction accuracy with respect to the expected value of κ is due to misclassification. For comparison, Fig. 18 displays the same information for $\kappa = 0.875$, where the linear behavior is still evident but with an increased offset for $\lambda = 0$. Theoretically, this offset should be proportional to κ, or 87.5, where the difference is due to recognition errors.

For lower values of κ, the acceptable loss parameter λ has little effect on the accuracy of the prediction algorithm. This is due to the fact that as κ approaches $\frac{1}{|C|}$, predictions approach random, therefore errors caused by increasing loss in recognition do not affect the randomness of the prediction. As κ increases, the effects of λ also increase, as can be seen when comparing the left column of Fig. 18 with the left column of Fig. 19. Furthermore, these effects are stronger for the kNN classifier as opposed to the HMM classifier, as the later has an internal Markov chain which stabilizes the prediction.

Comparing the left columns of Figs. 18 and 19 with their respective right columns, it is evident that the system's subjective evaluation in terms of its own prediction performance increases with respect to its actual performance. This indicates that because the algorithm is expecting to see certain activities, these activities are recognitized on occasion even if they do not occur. Only for high values of λ is

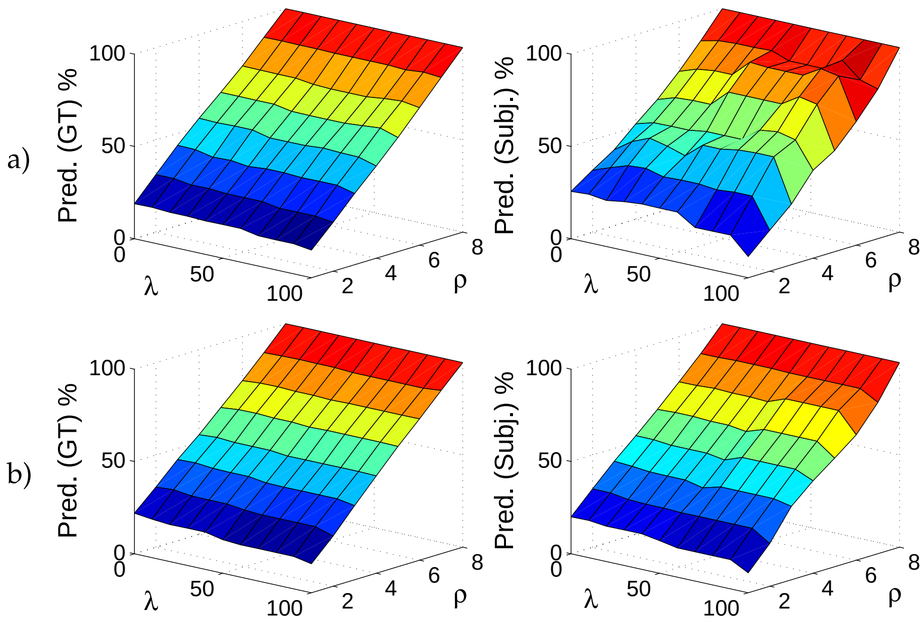


Figure 18.: Prediction against ground truth and classification for the MVS data set and the HMM (a) and kNN (b), κ=0.125

there a noticeable discrepancy, where the discrepancy increases as the predictability κ increases.

6.6.2 *Results of the OPP Data Set*

The activity recognition data set from the OPPORTUNITY project [25] was also evaluated using the prediction based method to confirm initial results from the MVS data set [13]. For this purpose the results of the kNN classifier alone are sufficient, and therefore the HMM results have been omitted for brevity.

OPP: Recognition Loss

The behavior of the recognition loss for the OPP data set with a predictability of $\kappa = 0.25$ is displayed in Fig. 20. Remember, for this data set, this value indicates random ordering as there are only 4 activity classes, as opposed to 8 classes in the MVS data set. Here, a plateau in recognition loss can be clearly seen at 12.2 pp for all values of the loss parameter $\lambda \geq 0.2$. This same plateau can be seen for $\kappa = 0.875$ in Fig. 21 as well.

This plateau behavior is generated by the cheapest sensor in terms of energy cost, in this case the magnetic field sensor (see Tab. 7). As opposed to the MVS data

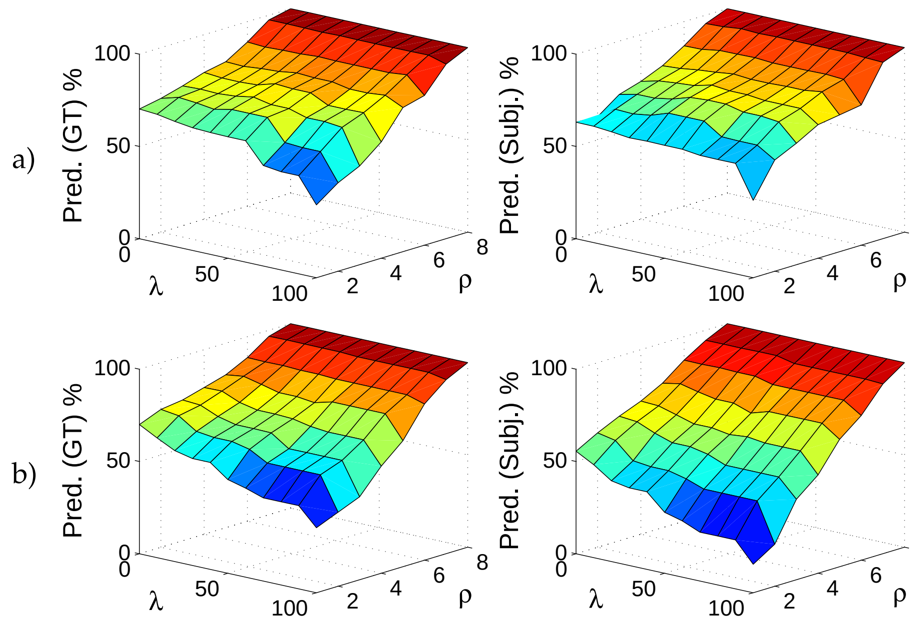


Figure 19.: Prediction against ground truth and classification for the MVS data set and the HMM (a) and kNN (b), κ=0.875

set, this sensor alone provides relatively high accuracy, indicating that increasing λ quickly leads the system to select that sensor alone as the lowest energy sensor configuration for achieving the required recognition.

OPP: Energy Savings

Similar to recognition loss, energy savings also plateaus at 94.8% for values of $\lambda \geq 0.2$, which is also the optimum for energy savings. This can be seen for both $\kappa = 0.25$ in Fig. 20, and for $\kappa = 0.875$ in Fig. 21. Again, this is caused by the loss parameter quickly dropping below the rates achievable using the single cheapest sensor, yielding an optimal configuration of only that sensor.

OPP: Prediction Rates

Fig. 22 displays the prediction rates for the OPP data set with a kNN classifier against ground truth (left) and system classifications (right) of human activities for $\kappa = 0.25$. This is the case where the prior distribution for future activities is completely uniform and random, where all 4 activities are equally likely to occur at the next time step. For $\rho = 1$, prediction is steady at approximately 38%, above the expected 25% given by the generated data set. This is caused by recognition

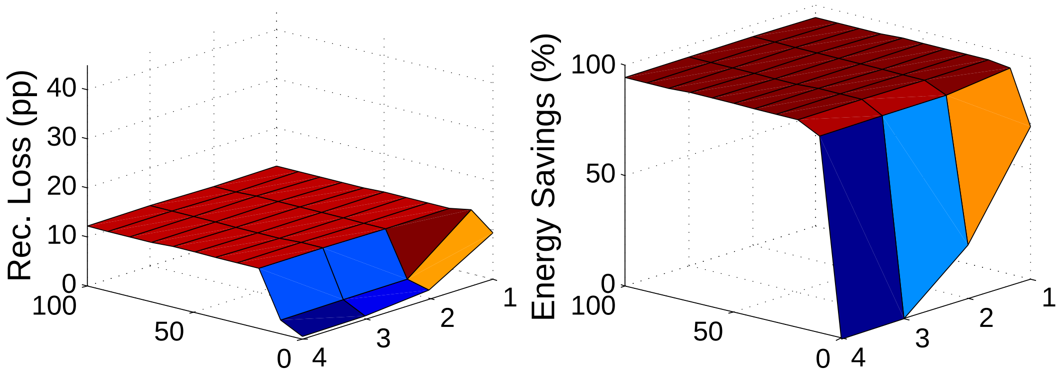


Figure 20.: Recognition loss and energy savings for the OPP data set and kNN with κ=0.025

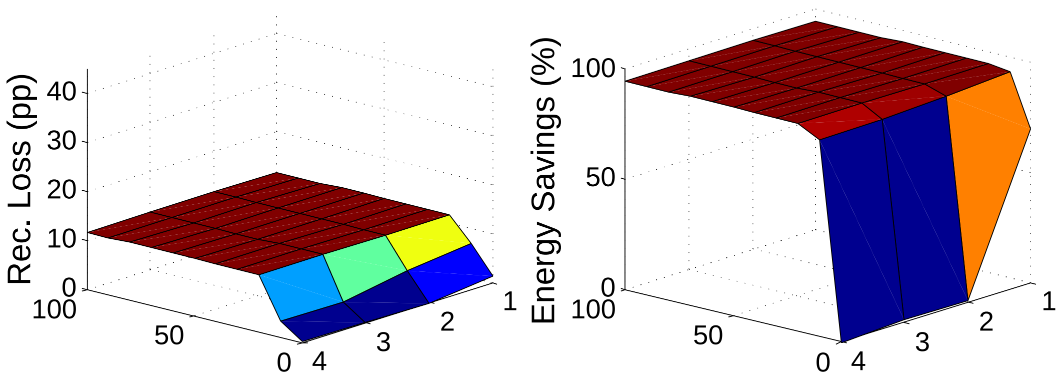


Figure 21.: Recognition loss and energy savings for the OPP data set and kNN with κ=0.875

error which may have a different predictability inside of a single class due to the generation process.

The same information is presented in Fig. 23 for $\kappa = 0.875$. For $\rho = 1$, prediction values begin at around 78% for $\lambda = 0$, but drop off to around 68% once the sensor set is reduced to the gyroscope alone for values of $\lambda \geq 0.2$. Again, the values increase linearly from that point to 100% as ρ increases towards the maximum value of 4.

In both Fig. 22 and Fig. 23 it can be seen that the subjective evaluation of the prediction value for the system is fairly accurate. Here, an increased effect for lower values of λ can be seen, as well as a reduction in prediction values. This reduction is around 10 to 15 pp for $\rho = 1$ and $\kappa = 0.875$, which falls off to 0 for $\rho = |C|$. Intuitively, this can be interpreted as the system correctly predicting the next activity, but judging this as a false positive due to misclassification.

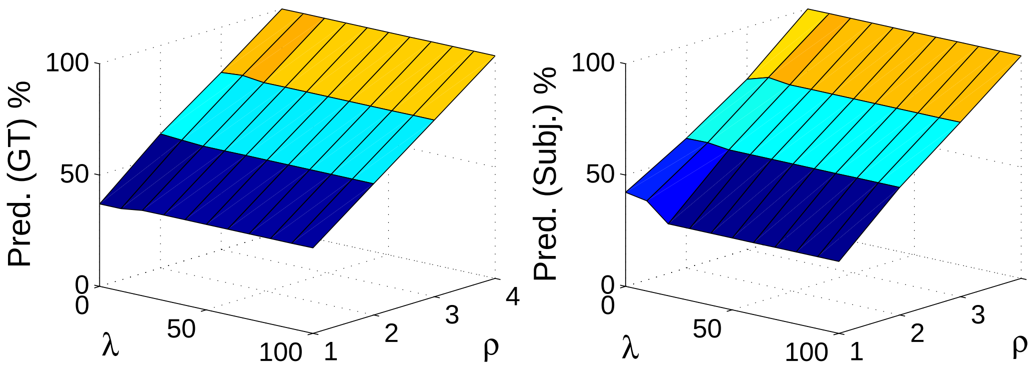


Figure 22.: Prediction against ground truth and Classification for the OPP data set and kNN classifier with κ=0.25

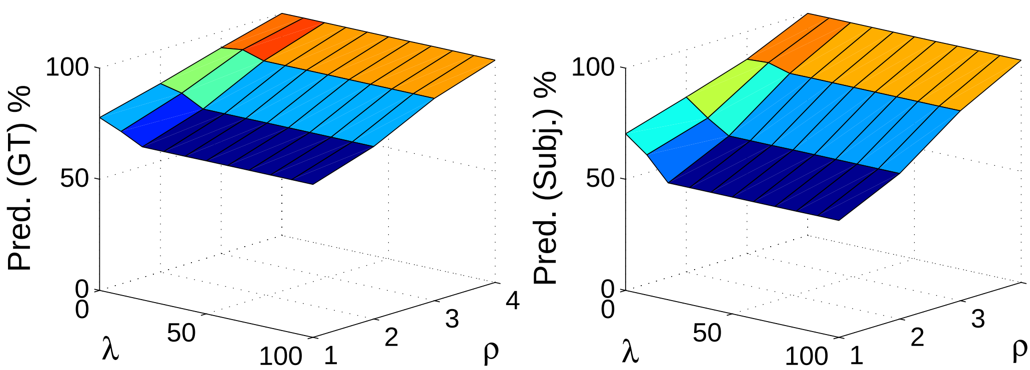


Figure 23.: Prediction against ground truth and Classification for the OPP data set and kNN classifier with κ=0.875

6.7 DISCUSSION AND INSIGHT

In Figs. 13 and 14, non-zero energy savings are present, even when $\lambda = 0$. Intuitively, setting $\lambda = 0$ means that any loss in recognition is unacceptable. For certain classes, some sensors are so insignificant that shutting them off results in an error increase so small that it is approximated to 0. Here the expensive light sensor is useless for most classes, and can be shut off with no loss as long as none of the classes requiring it is predicted. As more states are predicted however, classes requiring that sensor are more frequently predicted, regardless of their occurrence rate, increasing consumption with no effect on recognition.

The situation when $\rho = 1$ is extremely volatile, since only the single most likely future class is predicted. Fig. 13 and Fig. 14 show that $\rho = 1$ has significant negative effects for all non-zero values of λ. False classifications result in false predictions, resulting in false classification again. Introducing a confidence value at this point may allow the system to recognize error occurrence and correct by switching sensors back on when confidence or probabilities for all classes are low. The low recognition rates for $\rho = 1$ indicate that for all real scenarios $\rho = 1$ should probably not be considered. For higher values of κ such as 0.875 in Fig. 14, predicting as few as two states at each step can be sufficient.

The kNN classifier was more resistant to unpredictability within the data set. As κ decreases, Fig. 16 indicates that for a fixed loss parameter λ, the recognition loss expands faster for the HMM than for the kNN. In Fig. 17, the energy consumption for a fixed ρ grows faster and becomes more erratic for the HMM when compared to the kNN. The HMM algorithm models the activities as a Markov process [23], meaning that unpredictable feature vectors not only affect prediction, but classification as well. The kNN algorithm is only affected by κ through incorrect sensor activations which reduce recognition.

Both classification algorithms are influenced by lower values of κ due to sub-optimal sensor activations. The effect can be counteracted by increasing the number of classes predicted ρ, improving recognition accuracy but reducing the gain in energy. The parameter ρ controls the balance between risk and reward. High values of ρ mean less risk but a smaller payoff, and lower values increase the win in energy at the cost of recognition. The predictability of a scenario can be easily obtained for real scenarios by taking the accuracy of the prediction algorithm over the training data. Once κ is known, ρ can be configured to counteract it and select an appropriate risk level using Fig. 17 as a heuristic.

Once the risk and reward trade off between ρ and κ has been found, the loss parameter λ can be assigned to optimize the amount by which recognition may be reduced, and thereby the amount of energy which is conserved. For example, assuming a κ value of 0.5, $\rho = 3$ to counteract and a loss parameter of $\lambda = 0.2$, a

HMM incurs a loss of less than 1.2 pp in recognition but saves up to 84.11% of energy consumed without optimization.

One caveat is that due to the nature of prediction-based optimization, the system may perform badly for recognition of important but rare and unpredictable events or activities. This will arise if those activities require special sensors to distinguish them from other activities. Since the events cannot be easily predicted, the important sensors will not be correctly activated. A possible solution to this could be to exclude the required sensors from the set of sensors which can be deactivated, leaving them on at all times. Also one could use expert knowledge to hard-code the circumstances under which the events occur into the system if possible. If some activities are more important than others in general, investigating the integration of activity importance weights into the the activity-sensor weights in the Q matrix could provide an interesting avenue of research.

Another interesting aspect which has not been addressed here, is that often times the sensors of the mobile device are also used for purposes other than activity recognition alone, such as is the case with mobile phones. Under these circumstances, it is not advisable to switch these sensors off using an algorithm which does not take user preferences into account. In theory, the algorithm could be easily adapted to account for sensors which are in use by a user application $\widetilde{S}_{app}$. At each prediction step, Eq. (7) can be adapted to only search the Q matrix for sensor configurations which are a superset of $\widetilde{S}_{app}$. In this way the optimal sensor subset can be selected given that the subset $\widetilde{S}_{app}$ is activated.

Here we have evaluated the performance of the algorithm for a specified and fixed λ during runtime. Practically speaking, there is no reason why the acceptable loss cannot be changed dynamically during operation. This could have advantages for applications with mobile phones, where requirements on the energy source are also dynamic in nature. For example, when the phone is connected to a power supply, λ can be set to 0 as the power source is effectively unlimited. However as the battery level approaches a critical level, λ can be increased to extend the battery life as long as possible. Alternatively, devices could also try and recognize patterns in the daily lives of users [32], and set λ to appropriately account for the time until the device will probably be recharged.

One issue which has not been addressed is that although λ is proportional to how much recognition will be sacrificed, it does not provide an exact amount. The actual loss is a function of λ, the predictability κ, the number of predicted class ρ, and the number of total classes $|C|$, as well as the reference recognition rate when all sensors are on. The implication is that at training time, actual losses in recognition are unknown, as these are not only dependent on system parameters but also on the reference recognition rate for the given activities. One solution is that when gathering training data a small amount can be set aside for parameter tuning before the system is put online.

6.8 CONCLUSION

We proposed a novel method for saving energy while recognizing human activities using embedded and wearable sensing systems. We conducted a survey of existing techniques which revealed a taxonomy of approaches to this problem. Based on that taxonomy, we introduced the novel method for sensor system control which uses prediction to further conserve energy. Human beings are repetitive and periodic creatures, therefore what we do can be predicted to a certain extent. Sensors which are not needed to decipher probable activities from each other can be turned off, conserving energy without greatly impacting recognition rates. The algorithms are simulated using preexisting data sets [11, 25], which are used to generate artificial scenarios with specific degrees of predictability. The standard classification (Hidden Markov Model and k-Nearest Neighbors) and prediction (Markov Chain) algorithms where used for the evaluation so as to make the effects of the novel methods more pronounced.

The results indicate that for highly predictable scenarios, significant savings are possible with little loss in recognition. For less predictable scenarios losses are higher, but the predictability can be increased by predicting more than one state per iteration. This however reduces the savings in energy achievable, but limits the loss due to missed predictions. The Hidden Markov Model which takes state transitions into account performed better the kNN which does not, although low predictability has a greater effect on recognition. Finally, energy savings and recognition loss are greatly dependent on the activities being recognized and the sensors being used. The limit is given by the recognition rate using only the cheapest sensor with respect to energy consumption, and the recognition and energy consumption converge to the values given by using only that sensor as the system is granted increasing freedom to optimize at the cost of recognition.

Although this evaluation showed that energy savings are possible for small amounts of recognition loss, there are still some drawbacks to this approach. Applications which attempt to recognize important but unusual events could suffer greatly. This could come about if the important events require special sensors but their occurrence cannot be predicted, where the prediction error would lead to those sensors not being activated in the time of need. Counteracting this effect requires further research where several approaches have been proposed.

The results are also independent of the specific recognition and prediction algorithms used, depending only on the prediction and recognition rates achieved. The condition is that the classifier used can perform unimpared using the variable feature length, i.e. performance with missing features is not less that performance when having been trained without those features entirely. The simulation results can therefore be applied to a new scenario once the predictability (prediction accuracy) has been found. Although the acceptable loss parameter does indeed control how much loss is incurred, the function for this dependence is not straight forward and some training data should be used for parameter fitting. In this way the system designer can ensure that the actual loss in recognition incurred meets application requirements. For a scenario with predictability of 0.5 and 3 classes predicted at each time step, a loss parameter of 0.2 with the kNN classifier would incur a recognition loss of less than 1.5 pp but save 84.8% of energy consumed for the MVS data set and 2.8 pp and 89.9 % for the OPP data set.

These results demonstrate this approach to be an affective method for reducing power consumption of the activity recognition toolchain, thereby addressing Challenge 1 (Low Power). The next chapter evaluates SAR as an integral part of GAR, where using the methods presented in this chapter would greatly reduce the energy footprint of a GAR approach which fuses output from SAR. Furthermore, although not directly evaluated, these methods should be applicable to GAR approaches as well to reduce general GAR power consumption due to sensing. There one would predict likely and unlikely future group activities, and select sensors accordingly. Selecting appropriate sensor sets for different individuals at the same point in time poses and interesting research problem, however how the methods can be distributed in a P2P fashion remains to be seen.

An avenue of further research would be to integrate a reliability measure into the classification and prediction processes. Using such a method could allow the system to identify system configurations which do not allow reliable classification of activities, i.e. when necessary sensors are not active due to incorrect prediction, and take necessary measures. Another open research question is how the system would perform using low-level prediction based on the sensor data streams as opposed to an activity history. This would further decouple the performance of the recognition and prediction processes, and could prevent the negative effects of incorrect classifications on predictions, and then again on the next classifications and predictions, which occurred in some extreme situations. Finally, although the experiment was carefully designed to simulate real conditions, the work presented here would benefit from online experiments, e.g. with a mobile phone, to verify these results in live scenarios.

Bibliography

[1] L.K. Au, A.A.T. Bui, M.A. Batalin, and W.J. Kaiser. Energy-efficient context classification with dynamic sensor control. *Biomedical Circuits and Systems, IEEE Transactions on*, 6(2):167 –178, april 2012. ISSN 1932-4545.

[2] Ari Y. Benbasat and Joseph A. Paradiso. A framework for the automated generation of power-efficient classifiers for embedded sensor nodes. In *SenSys '07: Proceedings of the 5th intl. conf. on Embedded networked sensor systems*, pages 219–232, New York, NY, USA, 2007. ACM. ISBN 978-1-59593-763-6.

[3] Martin Berchtold, Matthias Budde, Dawud Gordon, Hedda Schmidtke, and Michael Beigl. ActiServ: Activity recognition service for mobile phones. In *ISWC'10: Proceedings of the Fourteenth International Symposium on Wearable Computers*, pages 83–90, Seoul, S. Korea, 2010. IEEE Computer Society.

[4] Nagendra B. Bharatula, Mathias Stäger, Paul Lukowicz, and Gerhard Tröster. Empirical Study of Design Choices in Multi-Sensor Context Recognition Systems. In *IFAWC: 2nd International Forum on Applied Wearable Computing*, pages 79–93, March 2005.

[5] Nagendra Bhargava Bharatula, Stijn Ossevoort, Mathias Stäger, and Gerhard Tröster. Towards wearable autonomous microsystems. In *Pervasive*, pages 225–237, 2004.

[6] N.B. Bharatula, M. Stager, P. Lukowicz, and G. Tröster. Power and size optimized multi-sensor context recognition platform. In *Wearable Computers, 2005. Proceedings. Ninth IEEE International Symposium on*, pages 194–195, Oct. 2005.

[7] Christopher M Bishop. *Pattern Recognition and Machine Learning*, volume 4 of *Information science and statistics*. Springer, 2006. ISBN 9780387310732.

[8] C.V.C. Bouten, K.T.M. Koekkoek, M. Verduin, R. Kodde, and J.D. Janssen. A triaxial accelerometer and portable data processing unit for the assessment of daily physical activity. *Biomedical Engineering, IEEE Transactions on*, 44(3):136 –147, march 1997. ISSN 0018-9294.

[9] Ozan Cakmakci, Joelle Coutaz, Kristof Van Laerhoven, and Hans werner Gellersen. Context awareness in systems with limited resources. In *In Proc. of the third workshop on Artificial Intelligence in Mobile Systems (AIMS), ECAI 2002*, pages 21–29, 2002.

[10] Lei Gao, A.K. Bourke, and J. Nelson. Activity recognition using dynamic multiple sensor fusion in body sensor networks. In *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*, pages 1077 –1080, 28 2012-sept. 1 2012.

[11] D. Gordon, H.R. Schmidtke, M. Beigl, and G. von Zengen. A novel micro-vibration sensor for activity recognition: Potential and limitations. In *Wearable Computers (ISWC), 2010 International Symposium on*, pages 1 –8, oct. 2010.

[12] D. Gordon, S. Sigg, Y. Ding, and M. Beigl. Using prediction to conserve energy in recognition on mobile devices. In *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2011 IEEE International Conference on*, pages 364 –367, march 2011.

[13] D. Gordon, J. Czerny, T. Miyaki, and M. Beigl. Energy-efficient activity recognition using prediction. In *Wearable Computers (ISWC), 2012 16th International Symposium on*, pages 29 –36, june 2012.

[14] Dawud Gordon, Jan-Hendrik Hanne, Martin Berchtold, Takashi Miyaki, and Michael Beigl. Recognizing group activities using wearable sensors. In Alessandro Puiatti and Tao Gu, editors, *Mobile and Ubiquitous Systems: Computing, Networking, and Services*, volume 104 of *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*, pages 350–361. Springer Berlin Heidelberg, 2012. ISBN 978-3-642-30972-4.

[15] Dawud Gordon, Jan-Hendrik Hanne, Martin Berchtold, Ali Asghar Nazari Shirehjini, and Michael Beigl. Towards collaborative group activity recognition using mobile devices. *Mobile Networks and Applications*, pages 1–15, 2012. ISSN 1383-469X.

[16] Dawud Gordon, Jürgen Czerny, and Michael Beigl. Activity recognition for creatures of habit. *Personal and Ubiquitous Computing*, pages 1–17, 2013. ISSN 1617-4909.

[17] Ville Könönen, Jani Mäntyjärvi, Heidi Similä, Juha Pärkkä, and Miikka Ermes. Automatic feature selection for context recognition in mobile devices. *Pervasive Mob. Comput.*, 6(2):181–197, April 2010. ISSN 1574-1192.

[18] A. Krause, M. Ihmig, E. Rankin, D. Leong, Smriti Gupta, D. Siewiorek, A. Smailagic, M. Deisher, and U. Sengupta. Trading off prediction accuracy and power consumption for context-aware wearable computing. In *Wearable Computers, 2005. Proceedings. Ninth IEEE International Symposium on*, pages 20–26, Oct. 2005.

[19] Kaisen Lin. Energy-Accuracy Aware Localization for Mobile Devices. In *Proceedings of the 8th international conference on Mobile systems applications and services MobiSys 10*, volume In Mobisys, page 285. ACM Press, 2010. ISBN 9781605589855.

[20] Hong Lu, A J Bernheim Brush, Bodhi Priyantha, Amy K Karlson, and Jie Liu. SpeakerSense: Energy Efficient Unobtrusive Speaker Identification on Mobile Phones. *Pervasive Computing*, 6696:188–205, 2011.

[21] Rene Mayrhofer, Harald Radi, and Alois Ferscha. Recognizing and predicting context by learning from user behavior. In W. Schreiner G. Kotsis, A. Ferscha and K. Ibrahim, editors, *Proc. MoMM 2003: 1st International Conference On Advances in Mobile Multimedia*, volume 171, pages 25–35. Austrian Computer Society (OCG), September 2003. ISBN 3-85403-171-8.

[22] Jeongyeup Paek, Joongheon Kim, and Ramesh Govindan. Energy-efficient rate-adaptive GPS-based positioning for smartphones. In *Proceedings of the 8th international conference on Mobile systems applications and services MobiSys 10*, MobiSys '10, page 299. ACM, ACM Press, 2010. ISBN 9781605589855.

[23] L.R. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257 –286, feb. 1989. ISSN 0018-9219.

[24] Giuseppe Raffa, Jinwon Lee, Lama Nachman, and Junehwa Song. Don't slow me down: Bringing energy efficiency to continuous gesture recognition. In *Wearable Computers ISWC 2010 International Symposium on*, pages 1–8. IEEE, 2010. ISBN 9781424490462.

[25] D. Roggen, A. Calatroni, M. Rossi, T. Holleczek, K. Forster, G. Tröster, P. Lukowicz, D. Bannach, G. Pirkl, A. Ferscha, J. Doppler, C. Holzmann, M. Kurz, G. Holl, R. Chavarriaga, H. Sagha, H. Bayati, M. Creatura, and J. del R Millan. Collecting complex activity datasets in highly rich networked sensor environments. In *Networked Sensing Systems (INSS), 2010 Seventh International Conference on*, pages 233 –240, jun. 2010.

[26] Nirmalya Roy, Archan Misra, Christine Julien, Sajal K Das, and Jit Biswas. An energy-efficient quality adaptive framework for multi-modal sensor context recognition. In *2011 IEEE International Conference on Pervasive Computing and Communications PerCom*, pages 63–73. Institute for Infocomm Research, Singapore, IEEE, 2011. ISBN 9781424495283.

[27] S. Sigg, D. Gordon, G. Zengen, M. Beigl, S. Haseloff, and K. David. Investigation of context prediction accuracy for different context abstraction levels. *Mobile Computing, IEEE Transactions on*, PP(99):1, 2012. ISSN 1536-1233.

[28] Mathias Stäger, Paul Lukowicz, and Gerhard Tröster. Implementation and evaluation of a low-power sound-based user activity recognition system. In *ISWC '04: Proceedings of the Eighth International Symposium on Wearable Computers*, pages 138–141, Washington, DC, USA, 2004. IEEE Computer Society. ISBN 0-7695-2186-X.

[29] Mathias Stäger, Paul Lukowicz, and Gerhard Tröster. Power and accuracy trade-offs in sound-based context recognition systems. *Pervasive and Mobile Computing*, 3:300 – 327, 2007.

[30] Feng-Tso Sun, C. Kuo, and M. Griss. Pear: Power efficiency through activity recognition (for ecg-based sensing). In *Pervasive Computing Technologies for Healthcare (PervasiveHealth), 2011 5th Int. Conf. on*, pages 115 –122, may 2011.

[31] Gautam Thatte, Ming Li, Sangwon Lee, Adar Emken, Shrikanth Narayanan, Urbashi Mitra, Donna Spruijt-Metz, and Murali Annavaram. Knowme: An energy-efficient multimodal body area network for physical activity monitoring. *ACM Trans. Embed. Comput. Syst.*, 11(S2):48:1–48:24, August 2012. ISSN 1539-9087.

[32] Kristof Van Laerhoven, David Kilian, and Bernt Schiele. Using rhythm awareness in long-term activity recognition. In *Proceedings of the 12th Int. Symp. on Wearable Computers (ISWC 2008)*, pages 63–68, Pittsburgh, PA, USA, 2008. IEEE Press. ISBN 978-1-4244-2637-9.

[33] Yi Wang, Bhaskar Krishnamachari, and Murali Annavaram. Semi-Markov State Estimation and Policy Optimization for Energy Efficient Mobile Sensing. In *The 9th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON'12)*, 2012.

[34] Alex L. Wood, Geoff V. Merrett, S.R. Gunn, B.M. Al-Hashimi, N.R. Shadbolt, and W. Hall. Adaptive sampling in context-aware systems: A machine learning approach. In *Wireless Sensor Systems (WSS 2012), IET Conference on*, pages 1–5, 2012.

[35] Zhixian Yan, Vigneshwaran Subbaraju, Dipanjan Chakraborty, Archan Misra, and Karl Aberer. Energy-Efficient Continuous Activity Recognition on Mobile Phones: An Activity-Adaptive Approach. *2012 16th International Symposium on Wearable Computers*, pages 17–24, June 2012.

7 Finding the Optimal Data Abstraction Level

7.1 ABSTRACT AND CONTEXT

This chapter researches the optimal data basis for GAR, looking at signal features, behavioral clusters, or individual activity classes in order to address Challenge 2 (Data Abstraction Level). A system is presented using only distributed, mobile devices, with a P2P client-server architecture used to investigate the abstraction level. In this sense it does not fulfill the requirements in Chap. 4 and is only used for exploratory purposes. The results presented in this chapter indicate that optimal GAR results at 96% accuracy are obtained using low-level activity information such as sensor values or features. Recognizing single-user activities using supervised learning before fusing individual behavior into group activity information can potentially be a good avenue of approach. By doing so, the volume of information which must be communicated drops, reducing power consumption by 40%. However, this experiment demonstrates several technical and practical difficulties which must be addressed for acquiring and working with data at this abstraction level. Specifically, training single-user and group classifiers in parallel requires doubly-labeling those activities, which is effort intensive, and an approach to avoid this problem lead to a drop in accuracy of 47%.

The approach of using cluster information from individuals generated by unsupervised learning is shown to be a good trade-off between recognition accuracy and power consumption and has many advantages GAR. A method using soft probabilistic clustering is shown to reduce power consumption by 36%, while incurring a loss in recognition of only 2.8%. At the same time it circumnavigates several of the practical difficulties incurred by the use of individual activities acquired using supervised learning. Soft clustering is used as the abstraction level of choice in Chap. 8 and in Chap. 9 for fully distributed recognition of group affiliation and emergent group activities. The content of this chapter is based mainly on a portion of a publication in the journal of Mobile Networks and Applications [16], which was preceded by contributions at CONTEXT'11 [13] and MobiQuitous'11 [14].

7.2 INTRODUCTION

Context and activity recognition provide intelligent devices in the environment with the ability to act proactively in the interest of users [25]. Many of us now carry around one or more intelligent devices constantly, and the number of intelligent systems in our environment such as entertainment systems, vending machines and informational displays is steadily increasing [2, 30]. Implicit pro-active interaction based on situational awareness is increasingly more important in order to prevent us from entering a state of permanent distraction and informational overload [12]. This state is a result of constantly having to administrate and respond to the myriad of intelligent devices in our immediate environment [3]. One vision within pervasive and ubiquitous computing sees environments progressing from single-user, private devices to include multi-user devices running private applications for those users who are present. A challenge then becomes not only recognizing the context of the single user interacting with the device as with mobile phones [3], but now attempting to recognize the activity of a group of individuals interacting with the system or the environment [14].

In this work we defined multi-user activity recognition (MAR) as the recognition of distinct activities of multiple users over a period of time. Group activity recognition (GAR) is the recognition of a single activity for an entity consisting of multiple individuals (see Chap. 3). The group activity is not necessarily the same as the sum of the activities of the individuals in it [17], and is often emergent behavior which is a function of the personal characteristics and behavior of all individuals in the group, as well as the group dynamic [21, 10].

Wearable technology has been proven to be effective for human activity recognition (HAR) [3, 1, 17], and is an attractive platform for MAR and GAR as it is already ubiquitous in the form of smart phones, tablets and other accessories. Using a distributed wearable platform for both the sensing and processing aspects of activity recognition is advantageous in that it allows the system to operate independently of existing infrastructure and therefore widens the field of applications [15]. Furthermore, in times of emergency, when GAR may be needed most, the conditions of infrastructure are at their worst [6]. In order to both combat the scalability challenges, and to be robust to infrastructure collapses or partial breakdowns, methods for recognizing group activities using the devices of the individuals within those groups are advantageous. These devices are intrinsically wearable, therefore motivating the field of GAR using mobile devices with wearable sensing modalities.

When using wearable technology such as badges [30], mobile phones [3], coffee cups [2, 14], etc. for group activity or context recognition, it is inherently a hierarchical problem, where data from wearable sensors on multiple users must be aggregated in order to infer the group context [17]. This poses a problem for such systems, as energy storage is a very limiting factor and reducing energy is a main priority [28] (see Chap. 4). Activity recognition approaches must therefore also be acutely aware of this issue and make every effort to reduce their energy consumption footprint on the system as a whole [12]. Preprocessing data locally reduces its volume and therewith the energy required for transmitting that data, but at the same time this process discards information which may be vital for classification [24, 27]. Transmitting unprocessed, raw data guarantees that the maximum amount of information is available for GAR, but the cost of communication is high. The low-power requirements of mobile devices must now be reconciled with the hierarchical nature of GAR, where again a tradeoff between recognition rates and energy consumption is evident. Here this tradeoff is evaluated for GAR in order to find out if an optimal amount of abstraction exists, and where that abstraction level is.

A system for recognizing group activities using only a distributed network of sensor nodes and mobile phones is introduced. A mobile phone is used as a central node for GAR, and wireless sensor nodes are attached to coffee mugs (Smart Mugs) to monitor the activities of the individual subjects. The Smart Mugs can process measured sensor data locally to different abstraction levels before forwarding that data to the mobile phone for GAR. The capability of the system to handle real-world MAR and GAR problems is, however, not evaluated here.

An experiment was designed to create a simple collaborative GAR problem of emergent behavior as it poses issues where nodes must exchange information in order to infer group activities (see Chap. 3). In this chapter the emergent behavior is generated by the construction of the experiment which may or may not be in a natural form. Recognition of naturally emergent behavior is evaluated later on in Chap. 9. The experiment is used to evaluate different levels of abstraction at which the information exchange occurs in terms of its effects on the distributed sensing (energy consumption), information exchange (communication volumes) and recognition (recognition rates) systems. The goal is to identify which abstraction level is optimal for collaborative GAR in terms of the energy savings and loss of recognition values.

Different levels of data processing result in different levels of abstraction [27], from low-level raw sensor data to high-level single-user activity information processed using single-user activity recognition (SAR) techniques. The later approach introduces the problem of having to doubly-label training data in terms of single-user and group activities in order to train both local SAR classifiers on the Smart Mugs and global GAR classifiers on the mobile phone. The term 'local' is used to refer to processes which occur at a single node, while global refers to processes which occur on the mobile phone which has a global view of the network, sensor data, and activities. Two methods for avoiding the doubly-labeling problem are presented and evaluated here: separate training sessions for local and global activities, and using unsupervised clustering techniques. These different modes of operation are evaluated in terms of distributed energy consumption and GAR rates in experiments with multiple subjects.

Furthermore, we present a more advantageous unsupervised learning approach, which solves one of the main problems of GAR using single-user activity information: the doubly-labeling issue. Finally, the data set gathered during the course of these experiments is published (see Section 7.6) to enhance reproducibility and to make this information available for future work within the community.

The results indicate that for the given set of activities, the optimal recognition was achieved using locally extracted features, saving energy without sacrificing recognition. Using locally classified activities presents several issues, and did not perform well under the given circumstances. The use of clustering however shows potential, saving 36 % of energy consumed for communication while only sacrificing 2.8 % of recognition.

7.3 RELATED WORK

The majority of all context and activity recognition work is focused on human subjects and concentrates on single-user activity and context recognition. Traditionally, this is conducted using body-worn acceleration sensors [1, 12] which forward sampled data to a central server for classification. Other approaches range from embedded recognition [28, 12], where emphasis is on the tradeoff between energy expenditure and recognition quality, to server based approaches which optimize classification results using crowd-sourcing [3].

First simple attempts at recognizing the activity of a group as a whole were pioneered with the Active Badge [30] and MediaCup [2] projects, where the status of a user (including meetings or gatherings) was updated based on their location and the location of others. These approaches where not learning-based, but rather static code which recognized activities mostly based on location, proximity, and some sensor measurements.

Other approaches use audio classification to recognize multi-user group activities, such as concurrent chatting activities [19], or for classifying roles of individuals in conversations [8]. These methods have proven effective, but rely heavily on infrastructure for recognition. Theoretically, embedded GAR approaches using audio sensors would be possible [28], but the authors are unaware of research in this direction.

Camera-based systems are well suited to collecting information about multiple individuals within the field of vision. This advantage has been put to use for the purpose of group activity recognition, for example for monitoring activities of groups of individuals in a prison yard [7] or cargo and logistics activities [11, 23]. Another great example of uniquely group-related activities, is recognition

Table 8.: Analysis and comparison of existing multi-user and group activity approaches

Reference	**Application Domain**	**Activ. Type**	**Dependency Type**	**Architecture**	**Sensor Tech.**	**Issues**
Chang et al. [7]	Prisoner activity recognition	GAR	Collaborative	Centralized	Video	Dependent on infrastructure, video requires instrumentation
Want et al. [30]	Office activities	MAR/GAR	Collaborative	Centralized	Multiple	Dependent on infrastructure, static logic, not capable of learning, domain specific
Beigl et al. [2]	Office activities	MAR/GAR	Collaborative	Centralized	Wearable	Dependent on infrastructure, static logic, not capable of learning, domain specific
Hsu et al. [19]	Human conversation	GAR	Collaborative	Centralized	Audio	Dependent on Infrastructure, domain specific
Wirz et al. [31] Roggen et al. [24]	Pedestrian flocking	MAR/GAR	Collaborative	Centralized	Wearable	Domain specific, focused on group affiliation
Hwang et al. [20]	Behavioral singularities	MAR	Collaborative	Centralized	Wearable	Application-specific outlier detection
Gu et al. [17], Wang et al. [29]	ADLs in home	MAR	Collaborative	Centralized	Wearable	Dependent on infrastructure, applicability for GAR unclear
Li et al. [22]	American Football plays	GAR	Collaborative	Centralized	Video	Dependent on infrastructure, video requires instrumentation
Gong et al. [11], Loy et al. [23]	Logistics and public places	MAR	Cooperative	Centralized	Video	Dependent on infrastructure, video requires instrumentation, highly domain specific
Dong et al. [8]	Conversational roles	MAR	Cooperative	Centralized	Audio	Dependent on infrastructure, application for HAR unclear
Present work [13, 14]	Generic (office) activities	GAR	Collaborative	Distributed	Wearable	

of American Football plays based on TV feeds [22]. There, Li et al. track individual trajectories and activities of single users, and then use this information to recognize which play is being orchestrated. The large drawbacks of video-based systems is that they require pre-instrumentation of recognition environments, and commonly require infrastructure to connect sensors and processing architectures.

Research into MAR and GAR using wearable sensors has only recently been introduced to the scientific community, an overview of which can be found in Tab. 8. Gu et al. [17, 29] combine patterns of individual activities to recognize concurrent multi-user activities using probabilistic methods. Here the activities which are recognized range from single-user activities as well as concurrent and conflicting multi-user activities, making this approach collaborative in nature.

Wirz et al. approach recognition of cluster formations and flow patterns in groups of pedestrians [31]. The work presented here expands on that work, as well as its extension done by Roggen et al. [24], where the concept of "behavioral primitives" are introduced as single-user activities. Here, group membership for each subject is monitored (MAR), but also crowd behavior is addressed (GAR), both of which can only be evaluated with knowledge of other group members (collaborative). Similarly Hwang et al. track behavioral singularities in children on field trips, where the behavioral singularity is tracked for each child (MAR), but can only be calculated in comparison with other subjects (collaborative).

Sigg et al. [27] researched the optimal context abstraction level for prediction of future contexts. This was also evaluated for a different application, namely sensor control for embedded SAR using prediction [12]. Since GAR using wearable sensors is inherently a hierarchical problem, these same issues are also present here as well, but with focus on GAR instead of context prediction. A case study on GAR to evaluate the optimal context abstraction level for GAR using sensors from wearable devices was presented in a preliminary poster abstract [13]. A requirements analysis for distributed recognition in peer-to-peer networks of mobile devices was also presented [15]. Preliminary results provided insight into the power-accuracy trade-off for GAR, and uncovered several novel research questions [14].

7.4 SYSTEM DESIGN

The system used here was made up of a wireless sensor network and a mobile phone. Wireless sensor nodes equipped with 3D acceleration sensors were attached to coffee mugs in a university/office setting. The nodes sampled activity and context data at the mugs, processed this data to a specified local abstraction level, and then forwarded it to the smart phone for further classification to the group activity as shown in Fig. 24.

The smart phone was tasked with recognizing group activities based on the data sampled by the wireless nodes on the coffee mugs. These nodes forwarded

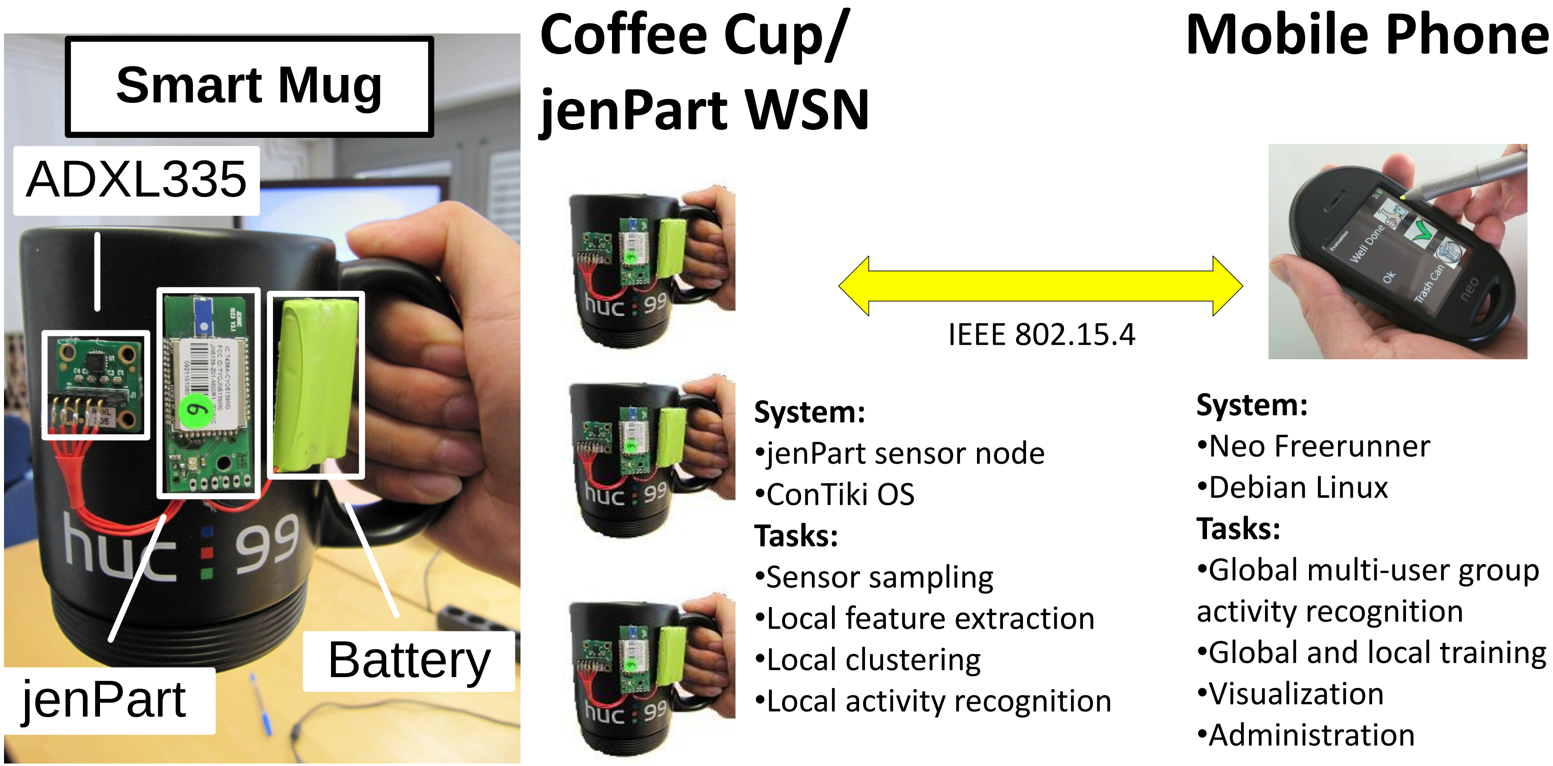


Figure 24.: **Left:** The smart mug with an acceleration sensor, jenPart node and battery **Right:** Network topology of the group activity recognition system and abstraction level experiment

either raw sensor measurements, extracted sensor signal features, local clustering information, or locally recognized single-user activities to the smart phone. The different modes were evaluated in terms of power consumption and recognition accuracy. The classifiers used in this paper are the k-Nearest-Neighbors (kNN) (k=10, Euclidean distance, no feature weighting), Decision Tree (DT) (C4.5), and Naive Bayes (nB) (no kernel estimation, single Gaussian, no covariance modeling) algorithms, selected for their simplicity for embedded purposes.

A hard K-Means clustering algorithm was used which outputs a single cluster candidate (top-1) for each vector, and uses subtractive clustering to identify the number of clusters present [14]. Each node outputs the index of the cluster which is identified given the k-Means clustering algorithm. These values are then fused by the mobile phone into a single group activity using trained classifiers. Since clustering algorithms do not require labels for training, local labels are not required for GAR, making these approaches advantageous.

Here a second method for evaluating the potential of unsupervised clustering as a method of skirting the doubly-labeling issue was investigated. Expectation maximization (EM) for Gaussian mixture models (GMM) [4] was used to cluster the data and a soft clustering approach was used as it has been shown to be advantageous for other approaches [5]. The Gaussian mixture is given by:

$$p(x) = \sum_{k=1}^{K} \pi_k P(x|\mu_k, \Sigma_k), \text{ where } \sum_{k=1}^{K} \pi_k = 1 \tag{8}$$

For each Gaussian component of the GMM, the probability that the vector was generated by that component is calculated. This probability is then normalized and output as the feature indexed by that component. For example, if training of one node yields a 3-component GMM, the output vector is then of length 3, where the k^{th} feature is the posterior for the k^{th} component of the GMM given the feature vector x, or $p(\mu_k, \Sigma_k|x)$.

The wireless sensor nodes used were jenParts [26] from the open-source Jennisense Project[1]. The nodes are based on the JENNIC JN5139 wireless microprocessor, the ConTiki operating system [9], a battery and an analog 3D acceleration sensor[2]. The nodes sample the sensors at a rate of 33 Hz and segment the sample data into windows (1 window = 16 samples $\approx 250ms$ with 50% overlap). Based on the operational mode, the windows are then processed and forwarded to the Neo: either the raw sensor data is forwarded (**Raw Data** mode, low-level data [27]), or the sensor signal features mean and variance are forwarded (**Feature** and **Training** mode), or single-user activity information from a classifier or clusterer is forwarded (**Classification** mode, high-level data [27]).

1 The Jennisense Project: https://github.com/teco-kit/Jennisense/wiki
2 ADXL335 3-Dimensional Acceleration Sensor: http://www.analog.com

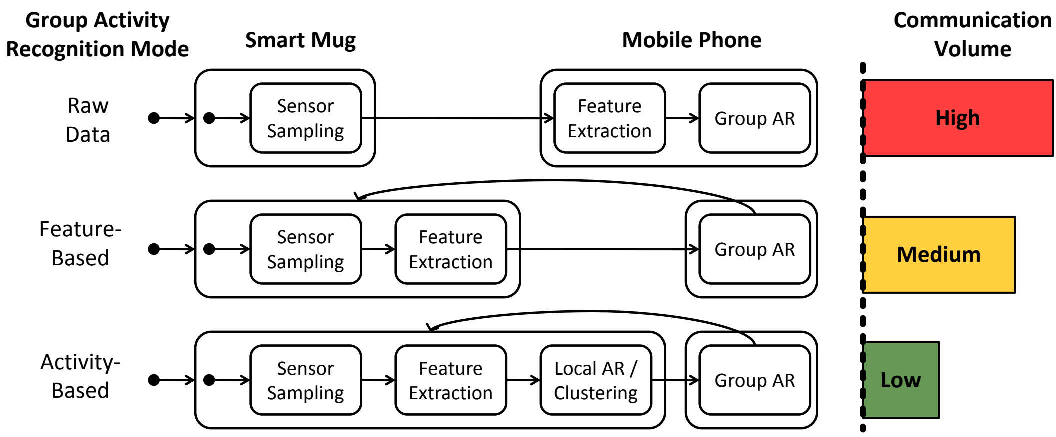


Figure 25.: State charts for the three different system modes for GAR with associated approximate communication volumes

A Neo Freerunner[3] was connected to a jenPart bridge in USB host mode for communication with the Smart Mugs. The Neo serves as a mobile platform for classifying the group activity based on the data aggregated from all nodes in the WSN. This involves a training mode and a classification mode for the global classifier. At training time, a vector consisting of data from the local nodes (either raw, features, clusters or classes) and a global group activity label is input into the global classifier. In classification mode, an unlabeled data vector consisting of the local data from the distributed nodes is input into the classifier, which then outputs the classification, or group activity estimation for that vector.

The Neo also serves as a classifier training platform for the mugs in the WSN. Following the approach presented by Berchtold et al. [3], after being set in training mode by the Neo, each mug gathers data and forwards it to the Neo along with a local annotation indicated by segmenting activities using the button on the jenParts. Once this process is complete, the Neo trains the selected classifier, segments the trained classifier into packet-sized chunks, and sends these chunks sequentially to the nodes in a JSON format. The Mugs are equipped with a JSON interpreter which then reconstructs the classifiers locally and places them in memory so that they can be executed as a module.

For all wireless communication tasks, the data is vital at the receiver. Wireless communication must either be designed in a reliable fashion, or measures must be taken to reconstitute missing data, although these questions are outside the scope of this work. For this experiment the system was designed such that packet loss was not an issue under the controlled conditions of usage, but in a real deployment this must be addressed (see Figs. 24 and 25).

3 http://www.openmoko.org/

7.5 EXPERIMENT

This experiment was designed to evaluate different levels of data abstraction carried out by the mugs in terms of energy consumption, communication volumes and GAR rates. To this end the experiment represents a collaborative GAR problem where single-user activities map to group activities in such a way as to be ambiguous for individual subjects, but explicit when observing all subjects (see Chap. 3) Processing data to the activity abstraction level [27] poses the problem of having to doubly-label the training data in terms of local, single-user activity labels and global, group activity labels. This must either be done using video recordings and offline annotation (time consuming) or multiple annotators in real time, both of which are too elaborate to allow easy deployment in new scenarios.

To counteract this, two methods of skirting the doubly-labeling issue are employed and evaluated. First, local classifiers and global classifiers are trained in two sessions where each session must only be labeled with local or global activities respectively. Second, local activity classifiers are replaced with a hard, top-1, unsupervised k-means clustering, and soft, probabilistic clustering [5], which does not require local activity labels, and can therefore be trained on the same data basis as the group activity classifier. Although the system was implemented on the distributed heterogeneous platform, the classification results were generated offline using the WEKA toolkit [18] for analytical purposes but were cross-checked with online results.

7.5.1 *Activity Recognition Experiment*

During the course of this experiment, 3 subjects performed 7 different activities, 3 of which were group activities and 4 of which were single-user activities involving the Smart Mugs. The activities performed will be detailed in the following subsections. In total, over 45 minutes of data were collected, making over 22,700 sample windows, although some data was discarded at random to ensure that experimental data was independently and identically distributed (i.i.d.). The experiments were conducted in a meeting room in a university setting over the course of a single day. In the first phase, local classifiers were trained and evaluated, followed by the global classifiers in the second.

Phase 1: Local Single-User Classifiers

In the first phase of the evaluation, each user performed a set of activities, each one for a duration from approximately 2 - 15 minutes with the mug in training mode, meaning features and labels were extracted locally and uploaded to the Neo. The activities were local to the mugs, and were not performed as part of group activities, as doubly labeling local and group activities in real time is impractical. The local single-user activities were as follows: the subject has placed the mug on the **table** (or other surface), the subject is holding the mug in their **hand**, the subject is **drinking** from the mug, and the subject is **gesticulating**.

After each activity was performed for the specified period of time, a button press on the node updated the label on the feature vector sent to the Neo and the next activity was performed. The first half of the data generated in this phase was used to train the local classifiers, and the second half was used to evaluate their performance. After all local activities were performed, the local classifiers were trained and communicated to the Smart Mug using JSON packets. The procedure of the process conducted in phase 1 is displayed in the upper portion of the sequence diagram in Fig. 26.

Phase 2: Global Group Training and Evaluation

The evaluation of the global classifier was conducted offline using the data generated in this phase, where again half of the data was used for training and the other for performance evaluation. The subjects conducted the following activities together for 4 - 5 minutes each using the same mugs they trained in the previous phase: **Meeting**, **Presentation** (users 1, 2 and 3) and **Coffee break**. The mappings of group to single-user activities are as follows: meeting consists of all subjects either setting their mugs on the table, holding them in their hand or drinking. In a presentation one subject will be gesticulating or holding their mug while others are either holding, drinking from, or have set the mugs down, and in a coffee break all are either holding, gesticulating with, or drinking from their mugs.

During this period, the nodes transmitted the full locally extracted feature vector, as well as the local classifications of the single-user activities listed previously. The raw sensor data was ignored for reasons which will be explained later. The process flow for phase two is shown in the lower portion of Fig. 26 where feature vectors and local activity classifications are transmitted simultaneously to train global classifiers for each data type respectively.

7.5.2 *Power Measurements*

The power consumption of each device was measured by placing the node in serial with a low error tolerance resistor and measuring the drop in voltage across the resistor. For each of the modes (raw sensor data, extracted feature data and classifier/cluster data) the average rate of consumption and the amount of energy consumed was calculated. The amount of energy consumed over the period of time beginning at t_0 and ending at t_1 is then given by $\int_{t_0}^{t_1} V_{supply} \times I_{supply}\, dt = \int_{t_0}^{t_1} V_{supply} \times \frac{V_{meas}}{R_{meas}}\, dt$ where V_{supply} is the supply voltage, I_{supply} is the current drawn

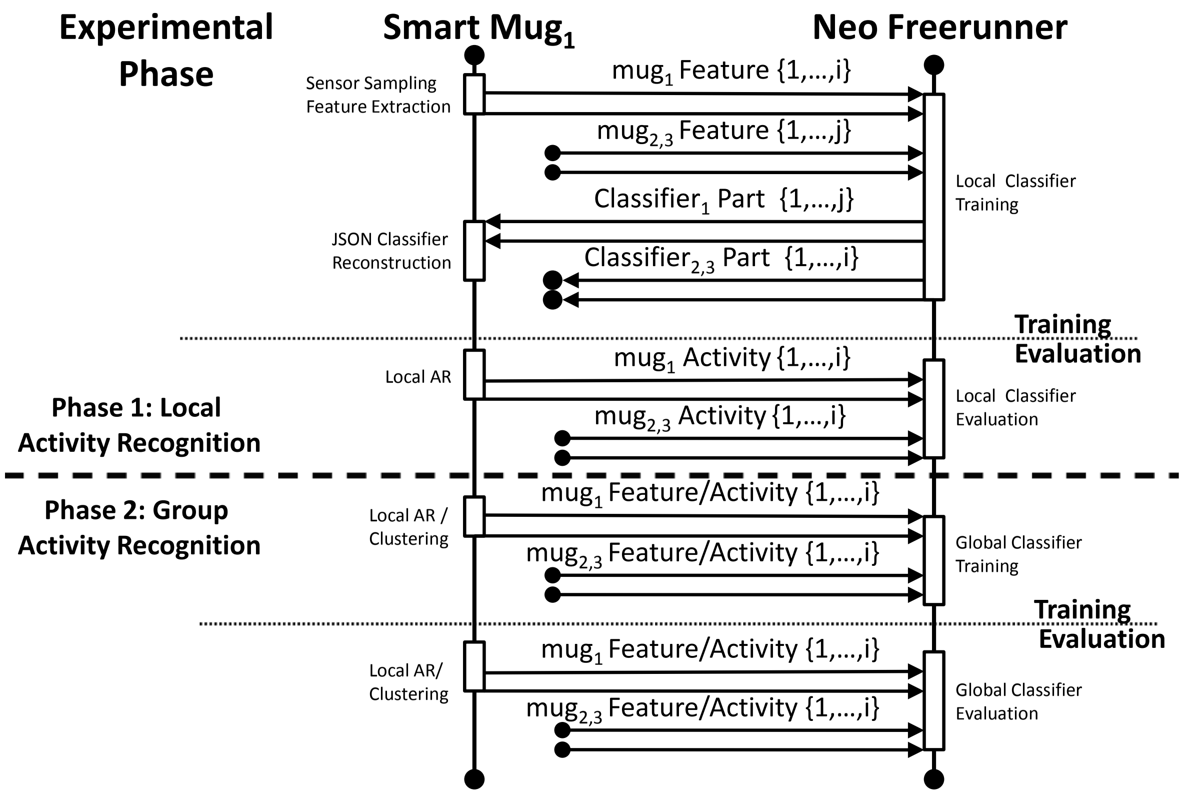


Figure 26.: Sequence diagram for the two-phase group activity recognition experiment.

by the node, which is given by the voltage drop (V_{meas}) over the measurement resistor with resistance R_{meas}.

7.5.3 *Raw Data Issues*

Since the features calculated by the mobile phone and the distributed nodes are identical, the recognition rates for both modes would be identical as well. Theoretically, the Neo is capable of calculating far more complex and extensive feature sets than the sensor nodes, meaning that recognition rates for the raw data phase could be higher than for locally extracted features. That certain features provide better or worse recognition values is however a known fact [12], and the field of feature selection is a different area of research, making this comparison out-of-scope in the context of this work. For this reason, the raw data phase was only used to evaluate data volumes and energy consumption rates, and not to compare classification values.

7.6 THE DATA SET

One of the most difficult and time consuming steps in HAR research is the collection of a data set for evaluating hypotheses and algorithms. Publishing these experimental data sets is a great step towards increasing reproducibility within the research field of activity recognition. Without this, progress in the field is slowed as scientist must redundantly record private data sets independent of each other, and the effort required for reproducing published results is prohibitive. For these reasons, part of the contribution of this publication is the data set used for evaluating this work.

The data set has been made available online [4]. All data has been published in the `.ARFF` file format for compatibility with the WEKA toolkit [18]. The files contain features generated over the sensor data streams for the two phases of the experiment. The raw sensor data was not transmitted in order to reduce the volume of communication on the wireless channel, avoiding collisions (causing possible loss of data) and freeing up bandwidth. The root folder contains two folders `/local/` and `/global/`, which correspond to Phase 1 and Phase 2 respectively.

7.6.1 *Data Set from Phase 1*

In the folder `/local/`, there are two files for each user, one containing training data and one containing testing data. Each file contains a set of vectors of length 7, each containing the average and variance for each axis of the accelerometer, and an

4 `http://www.teco.edu/~gordon/GAR/data_set.zip`

activity label. The activities in these files are the local single-user activity labels from Phase 1.

7.6.2 *Data Set from Phase 2*

Folder `/global/` contains the data from Phase 2 of the experiment. In the folder `/global/single vectors/`, the files `features_train.arff` and `*_test.arff` contain the combined data for all three subjects in one single vector. These vectors have a length 19, with 18 features (6 from each subject) and activity label. The labels here are Phase 2 activities, i.e. group activities.

Files `hard_cluster_train.arff` and `*_test.arff` contain the vectors resulting from the hard clustering experiment. Here, each vector is of length 3 where each feature represents the closest cluster for that subject. The labels in this file are group activities.

Files `soft_cluster_train.arff` and `*_test.arff` contain the vectors resulting from the soft clustering experiment. Here, the length of the vectors is dependent on the number of clusters discovered for each subject. Each feature represents the likelihood of a given cluster. The labels in this file are group activities.

Local single-user classification vectors from Phase 2 are stored in six files in the folder `/multiple vectors/`. The file names begin with the classifier name (`j48, knn, nb`) followed by `_train.arff` or `_test.arff`. Each dimension in the vector represents the activity of the subject indexed by that dimension. These are generated by classifying each vector of the global feature data with the classifiers trained using the data from Phase 1. The value of each position is a Phase 1 single-user activity, where the labels in this file are group activities. Since no doubly-labeling was done, information about the exact correctness of these classifications is not known.

The data set presented here has been normalized using a min-max normalization. The minimum and maximum cannot be taken from the training set. The testing data is used to simulate and evaluate online operation of the system. This data would therefore not be available at the point where the min and max values have to be fixed for the system. In order to correctly model this, the minimum and maximum values for normalization where acquired from the training data only, and used to normalize the entire data set.

7.7 RESULTS

The results are two fold, first the classification rates for local and group activities are presented, followed by the evaluation of the communication load and power consumption of the nodes. The implications of these results and the insights they provide into the field of collaborative GAR will be discussed in Sec. 7.8.

7.7.1 *Classification Results*

The classification results will be presented in two parts. First the local single-user classification rates achieved by the mugs themselves of their local activities are presented, followed by the recognition rates of the global classifier for GAR based on local features and local activities will be presented.

Phase 1: Local Classification

In phase 1 the mugs were trained using the following four classes: subject has set the mug down, subject is holding the mug, subject is drinking and subject is gesticulating. Tab. 9 displays the results of the evaluation of the local classifiers trained in phase 1 of the experiment. The accuracy, precision, recall and F-measure of each mug, as well as the average over the 3 mugs is displayed. All classifiers for local, single-user activities performed at around 95%, where minimal variance across mugs, activities and classifiers was observed.

Phase 2: Global Classification

Similar to phase 1, the global GAR classifier used half of the data generated in phase 2 for training and the other half for classifier evaluation. Tab. 10 displays the results of the evaluation of the global GAR classifiers from phase 2. Each row of the table represents a different data abstraction level of the mugs: either feature transmission, transmission of local activities (the local classifier algorithm is always the same as the global one, e.g. the first column is local single-user DT, with a global GAR DT), or transmission of local clustering results, either soft or hard. In total 12 global GAR classifiers were trained and tested, 3 classifiers (DT, kNN, nB) for each type of local data abstraction.

Tab. 10 indicates that local single-user classification provided poor results with a accuracies of 51% (DT), 49% (nB) and 42% (kNN). Local hard clustering provided better GAR results, with accuracies of 77% (DT, kNN) and 71% (nB). Local soft clustering resulted in a variance across different classifiers, achieving recognition rates of 94% and 91% for the DT and kNN classifiers respectively, but only 57% for the nB classifier (see Sec. 7.8 for a details). The best results were achieved using local features and a DT classifier (96%), where the kNN algorithm achieved relatively high recognition rates (89%), while the nB classifier was only able to achieve GAR with an accuracy of 56% (compare with 33% at random).

Table 9.: Classification rates for local single-user activity recognition

Data	**C4.5 Decision Tree**				**k-Nearest-Neighbors**				**Naive-Bayes**			
Basis	Acc.	Prec.	Rec.	F-meas.	Acc.	Prec.	Rec.	F-meas.	Acc.	Prec.	Rec.	F-meas.
Node 1	0.976	0.976	0.976	0.976	0.971	0.972	0.970	0.971	0.985	0.985	0.985	0.985
Node 2	0.948	0.947	0.947	0.947	0.936	0.941	0.935	0.938	0.906	0.910	0.905	0.908
Node 3	0.951	0.951	0.951	0.951	0.955	0.955	0.954	0.955	0.932	0.940	0.932	0.936
Average	0.958	0.958	0.958	0.958	0.954	0.956	0.953	0.955	0.941	0.945	0.941	0.943

Table 10.: Classification rates for global group activity recognition

Data	**C4.5 Decision Tree**				**k-Nearest-Neighbors**				**Naive-Bayes**			
Basis	Acc.	Prec.	Rec.	F-Sc.	Acc.	Prec.	Rec.	F-meas.	Acc.	Prec.	Rec.	F-meas.
Features	0.962	0.962	0.962	0.962	0.894	0.900	0.896	0.898	0.565	0.612	0.575	0.593
Soft Clust.	0.935	0.919	0.931	0.925	0.914	0.904	0.892	0.898	0.569	0.646	0.753	0.695
Hard Clust.	0.767	0.772	0.656	0.709	0.767	0.764	0.662	0.710	0.716	0.624	0.561	0.591
Activities	0.507	0.531	0.516	0.524	0.424	0.557	0.429	0.485	0.491	0.514	0.495	0.505

Table 11.: Communication volumes and power consumption results

Mode	**Data Volume** (B/s)	**Neo Freerunner** **Avg(P)** (W)	**Wireless Node (Mug)** **Avg(P)** (mW)	**Wireless Node (Mug)** $\mathbf{E_{Tx}}$ (mJ)
Raw data	404.25	1.771	24.574	1.012
Features	107.25	1.723	24.233	0.909
Soft clusters	Variable (avg. 24.75)	Variable ($\approx$ 1.703)	Variable ($\approx$ 23.296)	Variable ($\approx$ 0.648)
Hard clusters	12.375	1.700	23.140	0.605
Local activities	12.375	1.700	23.140	0.605

7.7.2 *Data Transmission and Energy Consumption*

In order to analyze the requirements of the three different system modes in terms of resource consumption the nodes were monitored over different modes of operation. The effects of each mode was analyzed in terms of communication time and volume as well as energy consumption. Tab. 11 displays the amount of time required for communication per second (T_{tx}) and the amount of data communicated per second for each node. The results indicate a drop in data volume of 73.5% between transmitting raw data and features, 88.5% between features and classes/hard clusters, and a 96.9% drop in the amount of data communicated from raw data mode to local context classification mode. Values for soft clustering are approximate as they vary across nodes depending on the number of Gaussian components in the GMM. The values shown in Tab. 11 are achieved using linear approximation based on the average data volume per node.

During the course of these experiments, the energy consumption rates of the different devices were also monitored. Tab. 11 displays the results for the energy measurements for both the mug hardware and the Neo mobile phone as they carried out the necessary operations. The results indicate a decrease in average energy consumption (Avg(P)) at the mugs of 1.4% from raw data to feature modes, a decrease of 4.5% from feature mode to classification mode, and a total drop of 5.8% from raw data to classification mode. For the Neo, a drop of 2.7% in average energy consumption was registered from raw data to features, a drop of 1.33% from features to classes, and a total drop of 4.0% from raw data to classification mode.

Due to the difference in the ratio of operational to transmission power consumption between the 2 device types, the energy consumption due to transmission could only be directly measured accurately at the mugs, but not at the Neo. The right-most column in Tab. 11 indicates the amount of energy consumed by a node for the purpose of wireless communication each second (E_{Tx}). This indicates a 10.17% drop in energy consumed when transmitting features compared to raw data, and a decrease of 33.44% from features to classes, with a total decrease of 40.22% from raw data to classes.

7.8 ANALYSIS AND DISCUSSION

7.8.1 *Using Single-User Activities for GAR:*

One of the most important issues is selecting local activities relevant to discrimination between the global group activities. Here the experiment was designed to avoid this problem by engineering a collaborative GAR problem which can be directly mapped onto the single-user activities in order to evaluate the effects of the abstraction level, rather than the scenario. For real scenarios, either intuitive or experimental knowledge of the relationship between group and individual activities is required for activity selection, otherwise global recognition rates will deteriorate.

In this experiment, global classifiers were trained using the output of the local classifiers in the local classification mode, meaning that local classifier error was present in the training data for global classifiers. Alternatively, doubly-labeling activities would have allowed for training local and global classifiers on the ground truth labels simultaneously. The effects on global rates is unknown; using local labels could allow for the global classifier to account for and correct local errors, though it may also worsen results by distorting global classifier mappings. Furthermore, in this experiment a great deal of the GAR error when using locally classified activities was due to the fact that the data generated in Phase 1 of the experiment differed greatly from the data generated in Phase 2. Although subjects were instructed to conduct local activities as they would in a meeting, they were incapable of reproducing their own behavior under the group activity conditions. This becomes apparent when comparing the averaged maximum feature values for signal average (812 local vs. 1324 global) and variance (6621 local vs. 148271 global) of the two datasets. Eliminating this discrepancy would involve labeling local activities during group activities which would greatly increase labeling effort.

Tab. 11 indicates that the energy consumed by the nodes for the purpose of transmission dropped by 33% when the nodes only transmit a locally classified situation instead of locally generated features. When compared with Tab. 10, it becomes clear that these values come at a high price in terms of the recognition rates for global classification.

In the previous section, the nB classifier performed badly as a global classifier. Both the nB and DT classifiers performed comparably locally, but there is a disparity of up to almost 50% for global group activities based on local features. This indicates that GAR presents problems which are not present for single-user AR, and that not every classifier algorithm used for single-user HAR is appropriate for multi-user GAR. Data analysis indicates that often times group activities create multiple clusters in the multi-dimensional feature (18 dimensions) and activity (3 dimensions) space, for instance group activity "Presentation" consists of 3 clusters, one for the "flavor" of the activity when each different user presents. The nB classifier used here uses a single Gaussian to model each activity without kernel estimation. For GAR, the poor results imply that a probabilistic approach must be combined with clustering and covariance modeling in order to model multiple clusters and dependencies, as the naive Bayes assumption can be detrimental.

7.8.2 *The Energy-Recognition Tradeoff*

The ratio of how much of the total energy consumption is used for communication can be seen in Tab. 11, and is very much system and implementation dependent, where the volume of data falls by 75%, meaning that a large portion of the energy consumed for communication is in overhead.

Also currently, Tab. 11 indicates that the energy used for transmitting data is only around 4% of the total amount of energy consumed by the node, which is due to this fact. The short sample window length (500 ms) means each communication contains only half of one second's worth of data. Increasing the length of this window would increase the amount of data per packet and reduce the packet overhead ratio. These values are heavily system and scenario dependent, where factors such as number of sensors and features, as well as window length and sample rate play a large role.

Changing these parameters could tip the energy and GAR rate trade-off and would require a new evaluation. In this system, only two features are calculated, whereas in the literature, activity recognition systems often implement multiple features in the time and frequency domains [1, 3, 12, 27, 28, 31]. Increasing the number of features calculated would further tip this ratio in favor of local classification and clustering, also increasing the overall effect of the energy consumed for communication on the total energy consumed by the system.

7.8.3 *Using Clustering for GAR*

Although the results of GAR using local hard clustering were significantly lower than using local features (77 % as opposed to 96 %, 20% drop), clustering is quite promising. Clustering does not require a separate phase for local training as local labels are not required (unsupervised learning), and reduces the energy consumption due to transmission by 33%. The 20% drop in GAR rates is prohibitive for most applications.

Soft probabilistic clustering, which showed significant promise for other applications [5], proved to be an effective tool here. The GAR rates in Tab. 10 indicates an accuracy of 94% for soft clustering, compared to the maximum of 96% when using features. This indicates a loss of recognition accuracy of 2.8% from GAR using features to GAR using soft clustering, while maintaining energy savings at approximately 29 %. The implications of these results are two-fold. Depending on the number of clusters identified, a parameter which can be controlled by the system designer, the resulting impact on energy reserves can be varied as well. The amount of data communicated is proportionate to the number of Gaussian components, therefore less clusters means lower consumptions.

Conversely, it stands to reason that increasing the number of clusters increases the quality of the information transferred, thereby increasing recognition rates. In this specific instance, between 1 and 3 clusters were detected per node, with a total of 6 clusters. This configuration generated power consumption values only slightly greater than hard clustering (double the data volume), but already produced high recognition values. Soft clustering allows the application designer to tune the tradeoff between energy consumption and recognition by increasing or decreasing the number of clusters. The range for tuning is given by the recognition rate using raw data (features, in this case 96%) which is the maximum, with the minimum being the values for hard clustering (here 77% accuracy).

Using local activities reduces cost, but also reduces GAR accuracy by an unacceptable 47%. Hard clustering maintained the cost reductions but with an accuracy loss of 20%, the advantage being that unsupervised learning does not require single-user labels. Soft clustering enables high recognition rates without requiring local labels, representing a real and viable solution to the doubly labeling problem. In this instance a loss of only 3% of recongition could be achieved without requireing local labels.

7.8.4 *Generalization of Results*

As a strong caveat, the absolute values for GAR rates and energy consumptions cannot be assumed for other collaborative GAR problems. The activities here were designed to present a collaborative GAR problem for experimentation, and to be fairly straight-forward to recognize. Results for different scenarios, as is also the case with SAR and MAR, would depend on a multitude of factors such as sensing modalities, type and number of activities, amount of training data, etc.. What can be generalized, however, are the relative rates for energy consumption, communication volumes and recognition rates with respect to the abstraction levels under similar conditions.

7.9 CONCLUSION

This paper introduced a system for group activity recognition using only wearable and mobile devices for both sensing and recognition purposes. The problems of multi-user (MAR) and group activity recognition (GAR) where defined and further classified into cooperative and collaborative problems, where information exchange between nodes was analyzed. An experiment was designed to investigate the effects of the abstraction level for information exchange on energy consumption and recognition rates.

The experiment was conducted in an office scenario where nodes attached to mugs were used to monitor user's activities and perform collaborative group activity recognition on a mobile phone. Different levels of context preprocessing at the mugs were examined and evaluated in terms of power consumption and activity recognition rates. Specifically, using raw data, signal features, locally classified single-user activities and local clustering were examined as the basis for GAR and evaluated in terms of the cost of transmission incurred as well as GAR rates. The dataset was presented as part of the scientific contribution of this work.

Results indicate that for the given set of activities, the optimal recognition was achieved using locally extracted features, with GAR accuracy of 96 % and a 10 % drop in the amount of energy consumed for the purpose of wireless communication. Locally classifying activities and using these to classify the global group activity reduced power consumption by a further 33 % to 40 % total, but incurred a 47 % drop in global multi-user GAR rates due to subjects' inability to recreate their own behavior under different conditions. Using local hard clustering showed potential by maintaining the reduced power consumption at 40 %, but still incurred a recognition drop of 20 %.

The investigations presented here into soft clustering for GAR showed two major insights. First, probabilistic soft clustering using Gaussian mixtures can be used to tweak the tradeoff between accuracy and power consumption of the GAR application. Second, by reducing power consumption (here only twice the communication volume compared to local activities, saving approximately 36 %) and maintaining high GAR accuracy (only a loss of 2.8 %), probabilistic soft clustering represents a method for tackling the doubly-labeling issue which is intrinsic in GAR.

Bibliography

[1] Ling Bao and Stephen S. Intille. Activity recognition from user-annotated acceleration data. In *Pervasive*, pages 1–17, 2004.

[2] Michael Beigl, Hans-W. Gellersen, and Albrecht Schmidt. Mediacups: experience with design and use of computer-augmented everyday artefacts. *Comput. Netw.*, 35:401–409, March 2001. ISSN 1389-1286.

[3] Martin Berchtold, Matthias Budde, Dawud Gordon, Hedda Schmidtke, and Michael Beigl. ActiServ: Activity recognition service for mobile phones. In *ISWC'10: Proceedings of the Fourteenth International Symposium on Wearable Computers*, pages 83–90, Seoul, S. Korea, 2010. IEEE Computer Society.

[4] Christopher M Bishop. *Pattern Recognition and Machine Learning*, volume 4 of *Information science and statistics*. Springer, 2006. ISBN 9780387310732.

[5] Ulf Blanke and Bernt Schiele. Daily routine recognition through activity spotting. In *Proceedings of the 4th International Symposium on Location and Context Awareness*, LoCA '09, pages 192–206, Berlin, Heidelberg, 2009. Springer-Verlag. ISBN 978-3-642-01720-9.

[6] Arjen Boin and Allan McConnell. Preparing for Critical Infrastructure Breakdowns: The Limits of Crisis Management and the Need for Resilience. *Journal of Contingencies and Crisis Management*, 15(1):50–59, March 2007. ISSN 1468-5973.

[7] Ming-Ching Chang, Nils Krahnstoever, Sernam Lim, and Ting Yu. Group level activity recognition in crowded environments across multiple cameras. In *Advanced Video and Signal Based Surveillance, IEEE Conference on*, Los Alamitos, CA, USA, 2010. IEEE Computer Society. ISBN 978-0-7695-4264-5.

[8] Wen Dong, Bruno Lepri, Alessandro Cappelletti, Alex Sandy Pentland, Fabio Pianesi, and Massimo Zancanaro. Using the influence model to recognize functional roles in meetings. In *Proceedings of the 9th international conference on Multimodal interfaces*, ICMI '07, pages 271–278, New York, NY, USA, 2007. ACM. ISBN 978-1-59593-817-6.

[9] Adam Dunkels, Björn Grönvall, and Thiemo Voigt. Contiki - a lightweight and flexible operating system for tiny networked sensors. In *Proceedings of the First IEEE Workshop on Embedded Networked Sensors (Emnets-I)*, Tampa, Florida, USA, November 2004.

[10] D.R. Forsyth. *Group Dynamics*. International student edition. Thomson/Wadsworth, 4th edition, 2006. ISBN 9780534368227.

[11] Shaogang Gong and Tao Xiang. Recognition of group activities using dynamic probabilistic networks. In *Proceedings of the Ninth IEEE International Conference on Computer Vision - Volume 2*, ICCV '03, pages 742–, Washington, DC, USA, 2003. IEEE Computer Society. ISBN 0-7695-1950-4.

[12] D. Gordon, J. Czerny, T. Miyaki, and M. Beigl. Energy-efficient activity recognition using prediction. In *Wearable Computers (ISWC), 2012 16th International Symposium on*, pages 29 –36, june 2012.

[13] Dawud Gordon, Jan-Hendrik Hanne, Martin Berchtold, Takashi Miyaki, and Michael Beigl. An experiment in hierarchical recognition of group activities using wearable sensors. In *Proceedings of the 7th conference on Modeling and using context (CONTEXT)*, pages 108–114. Springer, 2011.

[14] Dawud Gordon, Jan-Hendrik Hanne, Martin Berchtold, Takashi Miyaki, and Michael Beigl. Recognizing group activities using wearable sensors. In *Mobile and Ubiquitous Systems: Networking Services, MobiQuitous, 2011. MobiQuitous '11. 8th Annual International*, dec 2011.

[15] Dawud Gordon, Markus Scholz, Yong Ding, and Michael Beigl. Global peer-to-peer classification in mobile ad-hoc networks: a requirements analysis. In *Proceedings of the 7th conference on Modeling and using context (CONTEXT)*, pages 108–114. Springer, 2011. ISBN 978-3-642-24278-6.

[16] Dawud Gordon, Jan-Hendrik Hanne, Martin Berchtold, Ali Asghar Nazari Shirehjini, and Michael Beigl. Towards collaborative group activity recognition using mobile devices. *Mobile Networks and Applications*, pages 1–15, 2012. ISSN 1383-469X.

[17] Tao Gu, Zhanqing Wu, Liang Wang, Xianping Tao, and Jian Lu. Mining emerging patterns for recognizing activities of multiple users in pervasive computing. In *The Sixth Annual International Conference on Mobile and Ubiquitous Systems*, pages 1 –10, july 2009.

[18] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H. Witten. The weka data mining software: an update. *SIGKDD Explor. Newsl.*, 11:10–18, November 2009. ISSN 1931-0145.

[19] Jane Yung-Jen Hsu, Chia-Chun Lian, and Wan-Rong Jih. Probabilistic models for concurrent chatting activity recognition. *ACM Trans. Intell. Syst. Technol.*, 2: 4:1–4:20, January 2011. ISSN 2157-6904.

[20] Inseok Hwang, Hyukjae Jang, Taiwoo Park, Aram Choi, Youngki Lee, Chanyou Hwang, Yanggui Choi, Lama Nachman, and Junehwa Song. Leveraging children's behavioral distribution and singularities in new interactive environments: Study in kindergarten field trips. In *Tenth International Conference on Pervasive Computing*, 2012.

[21] K. Lewin. *Field theory in social science: selected theoretical papers*. Social science paperbacks. Harper, New York, 1951.

[22] Ruonan Li, R. Chellappa, and S.K. Zhou. Learning multi-modal densities on discriminative temporal interaction manifold for group activity recognition. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 2450 –2457, june 2009.

[23] Chen Change Loy, Tao Xiang, and Shaogang Gong. Time-delayed correlation analysis for multi-camera activity understanding. *Int. J. Comput. Vision*, 90(1): 106–129, October 2010. ISSN 0920-5691.

[24] Daniel Roggen, Martin Wirz, Gerhard Tröster, and Dirk Helbing. Recognition of Crowd Behavior from Mobile Sensors with Pattern Analysis and Graph Clustering Methods. *Networks*, 00(0):1–24, 2011.

[25] Albrecht Schmidt. *Ubiquitous Computing - Computing in Context*. PhD thesis, Lancaster University, November 2002.

[26] Philipp M. Scholl, Kristof Van Laerhoven, Dawud Gordon, Markus Scholz, and Matthias Berning. jnode: A sensor network platform that supports distributed inertial kinematic monitoring. In *Ninth International Conference on Networked Sensing Systems (INSS)*, june 2012.

[27] S. Sigg, D. Gordon, G. Zengen, M. Beigl, S. Haseloff, and K. David. Investigation of context prediction accuracy for different context abstraction levels. *Mobile Computing, IEEE Transactions on*, PP(99):1, 2012. ISSN 1536-1233.

[28] Mathias Stäger, Paul Lukowicz, and Gerhard Tröster. Power and accuracy trade-offs in sound-based context recognition systems. *Pervasive and Mobile Computing*, 3:300 – 327, 2007.

[29] Liang Wang, Tao Gu, Xianping Tao, Hanhua Chen, and Jian Lu. Recognizing multi-user activities using wearable sensors in a smart home. *Pervasive Mob. Comput.*, 7(3):287–298, June 2011. ISSN 1574-1192.

[30] Roy Want, Andy Hopper, Veronica Falcão, and Jonathan Gibbons. The active badge location system. *ACM Trans. Inf. Syst.*, 10(1):91–102, January 1992. ISSN 1046-8188.

[31] M. Wirz, D. Roggen, and G. Troster. Decentralized Detection of Group Formations from Wearable Acceleration Sensors. In *Computational Science and Engineering, 2009. CSE '09. International Conference on*, volume 4, pages 952–959. Ieee, 2009. ISBN 978-1-4244-5334-4.

8 Distributed Group Affiliation Detection

8.1 ABSTRACT AND CONTEXT

Chap. 7 demonstrated the potential of using unsupervised learning to characterize individual behavior within the group. In this chapter the use of similar methods for P2P detection of affinity and affiliation between individuals is investigated to address Challenge 3 (P2P Group Affiliation Detection). By modeling individuals' behavior independently using the unsupervised model fitting approach presented in Chap. 7, similarity between subjects can be judged based on the similarity of the models. The results presented here demonstrate that the novel methods work optimally over shorter windows of time (around 5 seconds), but that this optimum is fairly poor compared to a standard centralized approach. Taking affiliation values generated from these short windows and filtering them over longer periods of time (1 - 2 minutes) can however generate comparable results to algorithms from centralized related work. This chapter thereby addresses the Challenge 3 of recognizing group affiliations in a P2P network of sensing devices. Once affiliations between subjects has been detected, a P2P recognition of the emergent group activity can be carried out over members of the same group, where a method for this purpose is presented in Chap. 9. The work in this chapter has been accepted for publication at ISWC 2014 [12]. The work is the result of a 3 month stay as a Visiting Researcher at the ETH Zürich in the Wearable Computing Lab[1] under the supervision of Prof. Gerhard Tröster.

1 http://www.wearable.ethz.ch/

8.2 introduction and motivation

Around 70% of the time we spend in public areas is done together with other people [17]. In general we are social creatures and spend a great deal of our time in groups of one form or another [9]. This social behavior is also useful, as it has been shown that groups are better than individuals at accomplishing tasks, which is often why they are formed in the first place [9]. Understanding group behavior and context is then crucial for systems which are trying to assist these groups in some fashion. Before an understanding of the group's context can be reached, group and individual affiliations must be identified through the precess of group affiliation detection (**GAD**). Often times several groups can occupy the same space at once [17], making it important to detect non-affiliation as well as affiliation.

A group are two or more individuals who are connected to each other by social relationships [9]. Humans have an innate ability to visually recognize these groups quickly [17], using unconscious processes which can be described using the Gestalt Laws [9]. Our minds automatically observe and group objects together based on proximity, similarity and interaction: objects which are similar or near each other belong together. It is this perception process of detecting groups and affiliations which GAD proposes to emulate[18]. Since human-like perception is the goal, we are therefore bound to that perception as it defines correct and incorrect affiliation decisions. The problem is then to differentiate inter-group similarity from intra-group similarity.

Individuals who are within the same group at the same location will often behave similarly to each other, e.g. walking in the same direction, at the same speed, or conducting activities at the same time [18]. The behavior of group members is similar as members adapt their behavior to the group in a process called norming [9], as well as the "Chameleon Effect" where individuals mimic each other's physical behavior [7]. The principles of homophily and similarity also state that individuals who are already similar will also tend to to form groups with each other [9]. However the type of similarity differs from group to group based on the type social relationship which defines that group. A sensor which can measure and provide an indicator of "social proximity" between individuals for a given group can therefore be used to detect group affiliation. For example in groups performing similar physical activities, wearable sensor signals from group members can be used to estimate group affiliation [18], and recognize affiliative behavior.

Such methods aggregate data from distributed sensors and sensing modalities [15] and then analyze the emergent result [18]. However, there are conditions where centralized aggregation cannot be achieved [11]. This can occur in non-instrumented environments where a centralized resources is not reachable, or when the bandwidth is too expensive either in terms cost to the user device or to the infrastructure provider. Finally, under some circumstances access is simply not available, such as during emergencies where infrastructure is usually the first casualty [4]. For these conditions, new methods for evaluating group affiliation using P2P analysis systems must be explored.

By changing the angle of approach from centralized to P2P, the definition of the problem also changes. The point of view which we wish to imitate changes from that of the observer of the emergent behavior to the point of view of the individual in the P2P network. The problem now becomes assessing individual-to-individual affiliation across neighboring individuals and nodes. Complexity moves from the method for clustering groups, to the method for evaluating social proximity in a distributed fashion.

We present a method for assessing P2P affinity by modeling the data as a distribution and then calculating the disparity (or similarity) as the Jeffrey's divergence between models from different individuals. We call this method divergence-based affiliation detection (**DBAD**). We compare DBAD with centralized and distributed approaches using signal correlation which is the basis for previous approaches [18]. Such approaches require sensor data exchanges between nodes in order to perform time-series analysis (TSA). The DBAD approach is sensor-independent, requiring only a sensor which measures personal characteristics which are in some way indicative of inter and intra-group social proximity and can be modeled as distributions, e.g. BlueTooth fingerprints, GPS locations, physical proximity sensors, etc. Furthermore DBAD has the potential to use multiple modalities for a single pair-wise affinity analysis which would solve several existing issues [15], although this is beyond the scope here. The contribution of this work is the following:

- A method for detecting P2P social proximity by exchanging only model parameters of mobile sensor data instead of exchanging sensor data itself for TSA.
- A method for filtering social proximity over time to indicate group affiliation between individuals.

We present 2 methods for accomplishing GAD, one where nodes exchange Gaussian probability density functions (DBAD-P) of sensor data, and another where they exchange histograms of observations (DBAD-H). We evaluate these methods with an experiment involving 10 individuals with varying group numbers, sizes and affiliations conducting a homogenous activity: a scenario with high difficulty. The resulting data set is also published as part of the contribution of this work (see Sec. 8.5). We evaluate the methods using two different types of sensor data each with different types of distributions; accelerometers and magenetometers, modeled as normal and Von Mises distributions respectively. The DBAD methods perform significantly worse in terms of identifying inter and intra-group similarities at any given instant with a maximum of 63% compared to a 74% for centralized correlation. However, filtering similarity values over time improves recognition to 93%. The reduced P2P communication range limits the number of inter-group neighbors increasing accuracy to 80% even without filtering.

For a single classification, a centralized correlation approach requires only under 2kB and 13.6 mJ, and the total response time for each node is low as processing is offloaded. Distributed applications of correlation algorithms are however are not viable due to the time and energy required for communicating sensor data. The DBAD-H requires around 2.1kB of memory, and decreases response time and totally energy consumed by 8% and 24% respectively compared to the centralized approach due to reduced communication. DBAD-P has a higher response time due to local processing and requires double the memory, but the total energy expenditure is less than both centralized correlation and DBAD-H by 43% and 24% respectively due to reduced communication.

8.3 RELATED WORK

The behavior of groups or crowds is emergent behavior resulting from individual members' actions, their interactions with each other and the environment, as well as their initial states and predispositions [9]. In animals, individual behavior, interactions and states can be quite simple yet still generate complex group behavior [19], allowing straightforward modeling. For humans, modeling such systems is a very difficult problem due to the complexity and cardinality of variables.

GAD differs from behavioral modeling in that we are not interested in understanding the behavior, but rather in assessing if the behaviors are affiliated, regardless of the form of the behaviors themselves. Marin-Perianu et al. [16] proposed detecting groups of smart goods in the supply chain using the degree of correlation between the sensor signals. This approach was later applied to human beings, where the correlation of acceleration signal variance was used to identify group affiliation [18]. This was conducted with heterogeneous behavior over groups (e.g. walking, climbing stairs, etc.) and homogeneous behavior between individuals within the group.

First behavior-relevant information is extracted from sensor signals (signal features) and a correlation analysis is conducted to generate a pair-wise disparity (or similarity) matrix. The relational graph represented by the disparity matrix can then be clustered in order to obtain a fairly accurate group affiliation label for all individuals [18]. For multi-modal sensing systems, the clusters generated from each sensing modality can be fused in order to combine information from both multiple modalities [15]. Here the focus is now on trying to create methods which achieve the same goals, but without requiring centralized resources.

Brdiczka et al. [5] recognized changes in group configurations by calculating the Jeffrey's divergence over histograms of multi-modal sensor data. There, divergences have been shown to indicate changes in the group dynamic based on the emergent image of sensor data. Here we we investigate if these methods can also be an indicator of one to one group affinity. Since probability density functions over human trajectories characterize them well [6], it follows that these models could be useful for detecting similarities in that behavior.

BlueTooth has also been used as a sensing modality to recognize device proximity [8], as have microphone sensors [20]. In both cases fingerprinting methods were used to compare individuals to each other. However the principle is the same. At any given time, similarity metrics can give us an indication of "proximity" between individuals, often corresponding to physical proximity. However it is the similarity in these proximities over time which indicate group affiliation, requiring exchanging time-lines of measurements or features. We present a method to avoid exchanging observations, using model parameters instead, and a method for filtering these similarities over time to create an indicator of group affiliation.

8.4 DIVERGENCE-BASED AFFILIATION DETECTION

The previous work on centralized approaches [18] describes GAD as following. Sensor data streams from devices monitoring potential group members are analyzed and behavior-relevant information is extracted, e.g. acceleration variance, as indicators of individual activity cues [18]. A cross-correlation ρ analysis of a given time window of these extracted signals is conducted in a pair-wise fashion, resulting in a disparity matrix $\overline{\mathcal{M}}$ in which index i, j indicates the strength of the correlation between the observational data X of subject i and subject j over a period of time t.

$$\overline{\mathcal{M}}_{ij}^{t} = \rho(X_i^t, X_j^t) = \frac{\gamma(X_i^t, X_j^t)}{\sigma(X_i^t)\sigma(X_j^t)} \tag{9}$$

where X_i represents observational sensor data from subject i over the time period t, γ is the covariance and σ the variance over the windows. The multi-dimensional similarity graph represented by $\overline{\mathcal{M}}$ can then be clustered using semi-supervised clustering algorithms, resulting in an assignment of group affiliation.

The problem with this approach is that in order to evaluate the similarity between two subjects the data streams from both subjects are required. This is due to the cross-correlation algorithm γ in the numerator of Eq. (9), which requires calculation of a function of the point-wise multiplication of both signals. In order to avoid communicating raw sensor data, new methods of analysis which do not rely on time-based signal analysis are required.

We present a model-based approach to this problem called divergence-based affiliation detection (DBAD). The approach works for any sensing modality which delivers similar values for similar inter-individual behavior and can be expressed

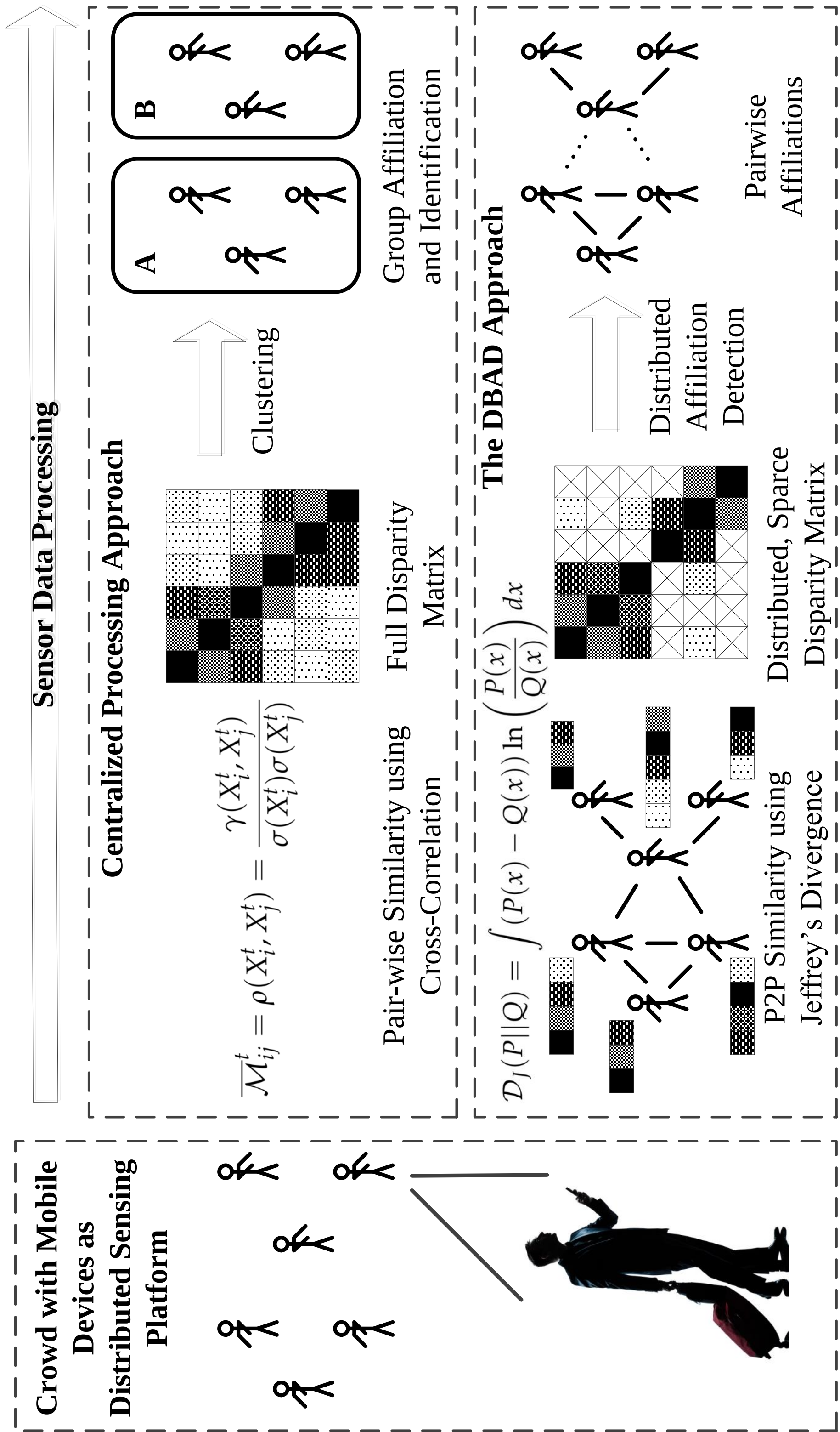


Figure 27.: The centralized and the novel distributed approach to group affiliation detection

as a histogram over a window. Theoretically it can also be used to combine several modalities into one similarity measurement although we have not yet evaluated this aspect. Each device computes a model of local data based on the sensor signals it has collected over a specified time window. Here the specific modeling approach taken is to use probability density functions (PDF) for modeling windows of local data. Once these models have been fitted, devices exchange parameters of these models with their neighbors. Each device calculates similarity to its neighbors using its own data and the models from its neighbors by calculating the divergence of the PDFs using an extension of the Kullback-Leibler divergence. Based on this information, neighboring devices then collaboratively decide if they are affiliated with each other or not. Depending on the model, this should reduce communication volume and hopefully maintain affiliation detection accuracy with respect to related work. As we will show in Sec. 8.6, this is indeed the case with a few caveats.

The volume of raw data communication is dependent on the sensor sample and re-sample rate and is therefore not fair to directly compare these rates. One approach to reducing this communication load without sacrificing recognition would be to compress the data losslessly before transmission, thereby reducing the amount of information which must be transmitted. We therefore evaluated how lossless compression would affect communication volumes by compressing the data using a two-step differential encoding followed by the DEFLATE[2] algorithm. Initially lossless audio compression algorithms where tried as these outperform DEFLATE for audio data which is of a somewhat similar nature to the sensor data used here, but these performed poorly.

8.4.1 *Distributed Modeling*

DBAD is then as follows. For each sample window, nodes extract relevant activity cues from the sensors. In this work, we monitored the acceleration and orientation of the subjects, using accelerometers and magnetometers respectively. For the acceleration we used the variance of signal magnitude, indicative of walking speed [18]. For orientation the circular mean of the azimuth, indicative of the direction of walking heading. The circular mean of a vector of angles $\bar{\theta}$ consisting of N angles θ is given by [2]:

$$\mu(\bar{\theta}) = \text{atan2}\left(\frac{\text{imag}(\bar{r})}{\text{real}(\bar{r})}\right), \text{ where } \bar{r} = \frac{1}{N}\sum_{j}^{N} e^{i\theta_j} \qquad (10)$$

These signals build the basis for the comparative analysis of individuals. Based on these signals each node fits a mixture model of distributions to the window,

2 https://tools.ietf.org/html/rfc1951

where the type of distribution used is based on the type of sensor used. For the acceleration signal, this is then modeled using a mixture of Gaussians. For the orientation sensor, the data is modeled using a mixture of von Mises distributions [3] due to the circular nature of the data [6], given by:

$$\text{vonMises}(\theta|\mu, m) = \frac{1}{2\pi I_0(m)} e^{m\cos(\theta-\mu)} \tag{11}$$

where the circular variance σ is given by $\sigma(\bar{\theta}) = 1 - \bar{r}$ and $I_0(m)$ is a normalization coefficient, given the zeroth-order modified Bessel function of the first kind:

$$I_0(m) = \frac{1}{2\pi} \int_0^{2\pi} e^{m\cos(\theta)} d\theta \tag{12}$$

For both models, the number of components is identified using subtractive clustering, with expectation maximization for parameter fitting [3, 6]. The results is a mixture model consisting of K Gaussian components:

$$P(x) = \sum_{k=1}^{K} \pi_k Distr_k(X) \tag{13}$$

where the type of distribution $Distr_k(X)$ used depends on the data being modeled, using standard Gaussians $\mathcal{N}(x|\mu_k, \sigma_k)$ for acceleration, or $\text{vonMises}(\theta|\mu_k, m_k)$ for orientation data.

8.4.2 *Distributed Affinity Analysis*

Once these mixture models haven been built, nodes (belonging to individuals) exchange these models with their single-hop neighbors. Each node n_i in the set of all nodes with dimension N can now calculate their disparity to neighboring nodes based on the Jeffrey's divergence of these two nodes. The Jeffrey's divergence is an extension of the Kullback-Leibler divergence, selected because it is numerically stable and symmetric. The Jeffrey's divergence D_j between two distributions P and Q is given by:

$$\mathcal{D}_J(P||Q) = \int (P(x) - Q(x)) \ln\left(\frac{P(x)}{Q(x)}\right) dx \tag{14}$$

Each node calculates its pairwise disparity to all other nodes within its single-hop communication neighborhood $\mathcal{V}^t$ at time t. Which nodes are in this neighborhood is dependent on the range of communication ψ (complexities in wireless communication are ignored here), and the physical Euclidean distance between two nodes at the time:

$$\mathcal{V}^t = \{[n_i, n_j]\}|\text{dist}^t(n_i, n_j) \le \psi \tag{15}$$

The behavioral distance between neighboring nodes can then be acquired as the value of the Jefferey's divergence between distributions of the sensor data of the two nodes.

$$\forall_{[n_i,n_j]\in\mathcal{V}^t} | \overline{\mathcal{M}}_{ij}^t = \mathcal{D}_J(Dist(X_i^t)||Dist(X_j^t)) \tag{16}$$

In this way the $\mathcal{D}_J$ is commutative and both nodes will conclude the same similarity based on the same models. In the centralized approach, the results of the complete pairwise metrics are centrally calculated, yielding a complete similarity matrix for all nodes as shown Fig. 27. Clustering this matrix to find affinity is a relatively straight-forward task, requiring only parameter fitting for clustering thresholds [18]. In a distributed approach this is not the case.

Each mobile device can only communicate with other nodes within reach of local p2p communication, which has 2 important repercussions. First, the similarity matrix is distributed across the complete set of user devices and is not available to any single device. Since the assumption is that global communication is either unavailable, intermittently unavailable, or cannot be used for cost reasons (i.e. bandwidth), it also implies that this distributed data entity cannot be fully queried by any single device. Second, its distributed nature also means that the disparity matrix is incomplete or sparse, as disparity is not measured between devices which are not within communication range. This presents a challenge of evaluating a distributed, sparse disparity matrix across multiple devices. The result is a distributed, sparse disparity matrix as shown in the bottom portion of Fig. 27, where each row of the disparity matrix is located on a different device, and several positions contain no data (when $[n_i, n_j] \notin \mathcal{V}$). Since individuals are mobile over time, the vacancy of a position in the disparity matrix $\mathcal{M}_{ij}^t$ at time t also varies over time as well.

The output of the distributed similarity analysis may fluctuate from window to window, therefore a moving average of the disparity matrices is used as a low-pass filter to smooth the output. A fifo buffer of length b is taken, where the length of the buffer represents how many disparity matrices where used in the average process. This buffer forms the basis for the low-pass moving average filter, where at any point in time t the smoothed disparity matrix $\widetilde{\mathcal{M}}^t$ is given by point-wise average of the disparity matrices in the buffer:

$$\widetilde{\mathcal{M}}_{ij}^t = \frac{1}{b}\sum_{\tau=0}^{b-1} \overline{\mathcal{M}}_{ij}^{t-\tau} \tag{17}$$

For example, a buffer length of one is the same as using no buffer, where a buffer of length 3 means that two previous matrices as well as the current one are averaged together before clustering.

Distributed Threshold-based Clustering

Once distributed disparity has been assessed, nodes must then convert this into affiliation information. A threshold-based approach was followed where each device makes a decision based on locally observed disparity values and a predefined threshold ϕ. For each node n_i, the clustering is conducted using only the information in $\overline{\mathcal{M}}_{ij}^{t}$ where $[n_i, n_j] \in \mathcal{V}^t$, or the information local to the node at time t. The result is a subset $\mathcal{V}^t_{\text{affil}_i} \in \mathcal{V}$ of nodes which are affiliated withe node n_i at time t, based on their disparity:

$$\mathcal{V}^t_{\text{affil}_i} := [n_i, n_j] \in \mathcal{V}^t | (\overline{\mathcal{M}}_{ij}^{t} \leq \phi) \tag{18}$$

Where the converse is true for local non-affiliation decision:

$$\mathcal{V}^t_{\text{non-affil}_i} := [n_i, n_j] \in \mathcal{V}^t | (\overline{\mathcal{M}}_{ij}^{t} > \phi) \tag{19}$$

From a global point of view one can ignore the pairwise neighborhood memberships and cluster the full disparity matrix using the threshold value ϕ, ignoring $\mathcal{V}^t$. However, in real situations, this information is not available, therefore making local decisions based on local information necessary. The optimal value used for ϕ is dependent on the physical activity of the subject, as well as the sensors used to monitor that behavior. For practical purposes, the threshold can be experimentally obtained by maximizing the accuracy.

8.5 GROUP BEHAVIOR EXPERIMENT

Previous experiments with centralized group behavior detection [18] were conducted with groups performing various heterogeneous activities and acceleration sensors. This configuration presents a large difference between inter and intra-group similarities. However inter-group behavioral differences may not occur in this fashion. It is quite possible that the activity performed by all participants is homogeneous, e.g. walking, queuing in crowds [14]. To evaluate performance under these more difficult conditions, an experiment and data set was created using homogeneous activity behavior, namely walking, of several individuals in different group configurations.

The experiment was conducted in a large open room in a university setting. 12 subjects walked through the room in various group configurations while being monitored by wearable android mobile sensing devices attached to the hip of each subject as shown in Fig. 29. The devices monitored a single subject each using 3D accelerometers and magnetic field sensors, as well as ambient audio. For each subject, the data set contains 51 minutes of data, although 2 devices contained faulty motion sensors, leaving 10 usable subjects.

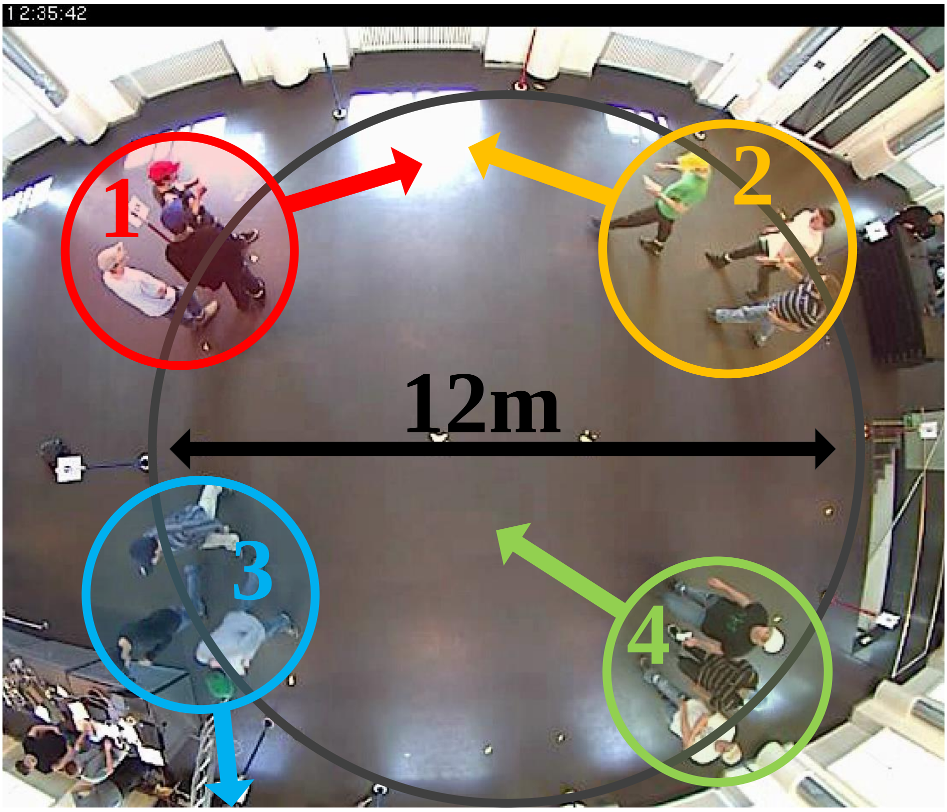


Figure 28.: A still image from the experiment video showing a four group configuration, with annotated group affiliating and heading

The experiment was recorded using a wide-angle lens on the ceiling of the room, and each subject was given head-gear of a different color to enable offline individual identification as shown in Fig. 28. 12 labeled posts where set up in a circle with a diameter of 12 meters inside of a large room, where each post displayed a unique number clearly on a sign in clock-wise order. A single member from each group was given a list of numbers, and each group then followed that member from post to post in the randomly assigned order on the list. Between experiments, group affiliations where reassigned and the experiment was repeated in the following configurations: one group (all together), 2 groups, 3 groups, 4 groups, no groups (each subject was given a separate list). Before each group experiment, subjects hopped in unison 3 times which was used to synchronize data by aligning the periods of free-fall (zero acceleration) across subjects.

Location data for each subject was annotated after the fact using a mixture of manual and automated color tracking software. For this purpose the video of the experiment was taken and the pixel coordinates of the subject's hat was tracked

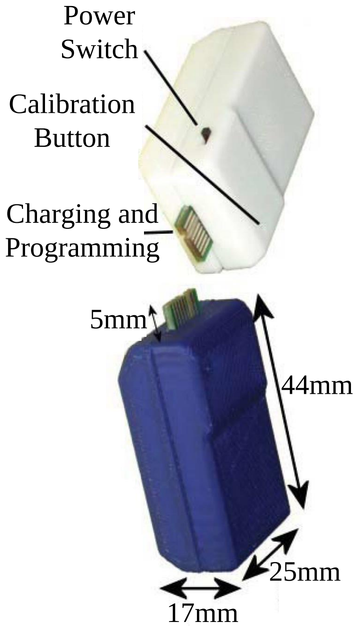


Figure 29.: The Android devices used for subject monitoring (left) and the on-body position of the devices (right)

throughout the experiments. The location is given in pixel coordinates from the top left of the video. We converted these coordinates into meters using the diameter of the circle (12 meters = 430 pixels) as a reference. These coordinates contain the elliptical distortion of the wide-angle lens, but can theoretically be transposed into spacial coordinates using the known dimensions of the room and the location of the camera. We argue that for the purpose of this research, this approximation suffices.

The performance of both centralized and the DBAD algorithms was implemented in MATLAB and then simulated using this data set. The simulation was performed on both the accelerometer and orientation data respectively to evaluate new and previous methods for the emergency situation scenario. For this purpose, the data from the experiments was cut up into windows whose length was varied. The variance of the acceleration data was calculated over a 15 second moving window, as this was shown to be effective for centralized forms of group affiliation detection in other scenarios [18]. For the orientation data, the azimuth was taken around the vertical axis of the subject, and a moving average of one second was used as an indicator of walking direction.

GAD was then performed using the centralized approach based on the signal cross-correlation, as well as the DBAD algorithms. Numerical integration of a PDF is carried out by estimating a histogram of the PDF. In order to evaluate the effect of modeling error on performance, the same process was also conducting using histograms of the individual sample windows constructed using the data windows directly as well. The resulting sparse, distributed similarity matrices (see Fig. 27 where then clustered for both the PDF-based and histogram-based data, and the results where evaluated in terms of correct and incorrect pairwise affiliation detections.

Pairwise affiliations are binary in nature, either indicating affiliation or non-affiliation of two subjects. However for a given group configuration, the distribution of affiliation and non-affiliation is not independent and identical. Given N subjects divided into M groups, with size(m_i) subjects in each group $m_i \in M$, the total number of pairwise comparisons is:

$$|S \times S|_{\text{total}} = \frac{(N)(N-1)}{2} \tag{20}$$

The number of subject affiliations is given by:

$$|S \times S|_{\text{affil}} = \sum_{i}^{M} \frac{\text{size}(m_i)(\text{size}(m_i) - 1)}{2} \tag{21}$$

The number of non-affiliations is given by:

$$|S \times S|_{\text{non-affil}} = \prod_{i}^{M} \text{size}(m_i) = X_{\text{total}} - X_{\text{affil}} \tag{22}$$

The accuracy for recognition is then defined as the fraction of pairwise affiliations correctly estimated by the system. A true positive is an affiliation which is judged as an affiliation, as true negative is for non-affiliations. False positive is a non-affiliation judged to be an affiliation, and false negative when affiliated subjects are judged non affiliated. In experiments where there is either only one group, or experiments where everyone acts independently, a row of the confusion matrix is then empty or zero. It is important to note that standard evaluation metrics such as precision, recall and f-measure are then undefined in these cases due to division by zero.

8.6 EVALUATION

Before evaluating communication range, all algorithms where evaluated using a sliding window whose length was varied between 1 to 60s. The results of this simulation are shown in Fig. 30. For all algorithms, results using the acceleration sensor are shown to be independent of window length and remain constant at around 63.5%. This value represents the noise level, defined as the performance of a classifier guessing at random, meaning that for these groups acceleration

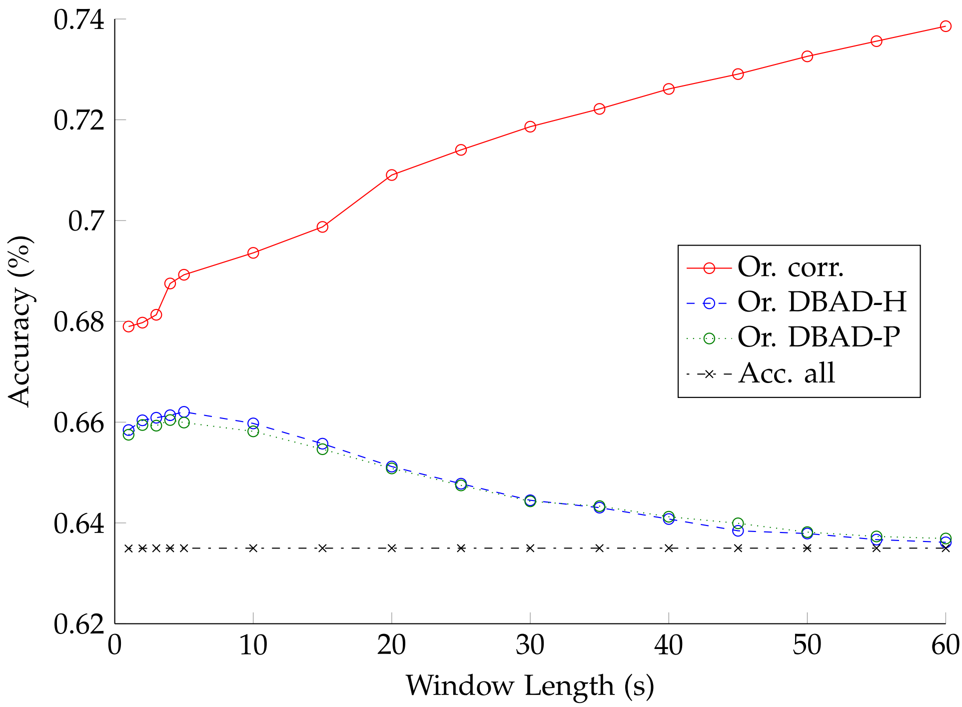


Figure 30.: Performance for cross-correlation (corr.) and DBAD over orientation (or.) and acceleration (acc.) signals

does not represent proximity. The reason the value is above 50% is the imbalance between the number of affiliative and non-affiliative links as was mention in Sec. 8.5. The centralized approach performed best of the three algorithms and improves monotonically with the length of the window, achieving just under 74% for a window length of 60 seconds. Using DBAD-H on the histograms and clustering the resulting complete similarity matrices yields an optimum of around 66.2% at 5 seconds, indicating weak representation of social proximity. Further increasing the sample window reduces the accuracy of the algorithm, as it asymptotically approaches the noise level at 60 seconds. DBAD-P performs only slightly worse than using a histogram, behaving similarly with an optimum of 66.0% at a window length of 5 seconds which then drops off into noise.

While not necessarily a negative result, the recognition rates achieved are not high enough to be useful in such situations. On further inspection of performance, we identified that individual affiliation values extracted from sample windows where noisy from frame to frame. A simulation was conducted using a low pass filter, in this case a sliding window moving average, in order to assess the informational content provided by the distributed methods. For this purpose the

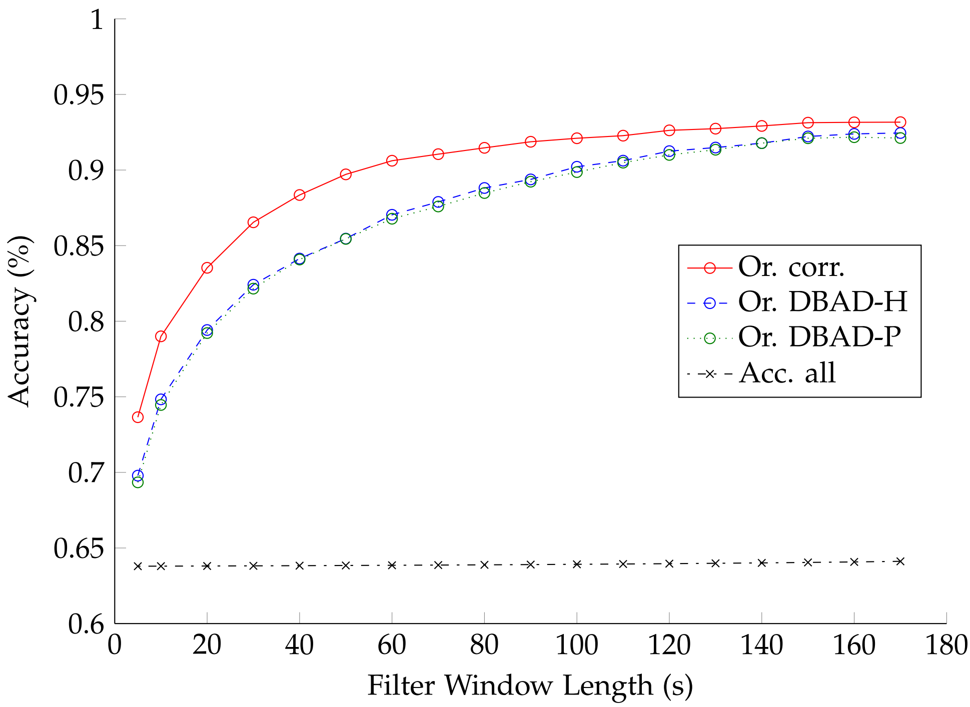


Figure 31.: Performance for cross-correlation (corr.) and DBAD over orientation (or.) and acceleration (acc.) signals filtering over a 5s window

optimum window length for the distributed algorithms of 5 seconds was used, as well as the filter in Eq. (17).

The results of this evaluation are shown in Fig. 31. Here the accuracy using the acceleration sensor remains almost constant showing only slight increases with filtering, indicating that noisy data is not causing the low values. Using orientation data however, it can be clearly seen that the centralized approach as well as both distributed approaches benefit from filtering, eventually all converging at values of around 93.3%. Optimum values are reached after about 250 seconds of monitoring, or a filter of length 50 classifications over 5 second windows.

One of the goals of the proposed methods is to reduce communication volumes, thereby alleviating stress on the network and reducing battery life of the individual devices. We monitored the rate of exchange of data during the course of the simulations for the different algorithms, the results of which can be seen in Fig. 32. The centralized approach requires each node to exchange the entire sample window's worth of sensor data, in this case sampled at 50 Hz. Regardless of window length, 50 Hz of sensor data must be transferred per second, requiring 4 bytes of data per measurement, or 200 B/s.

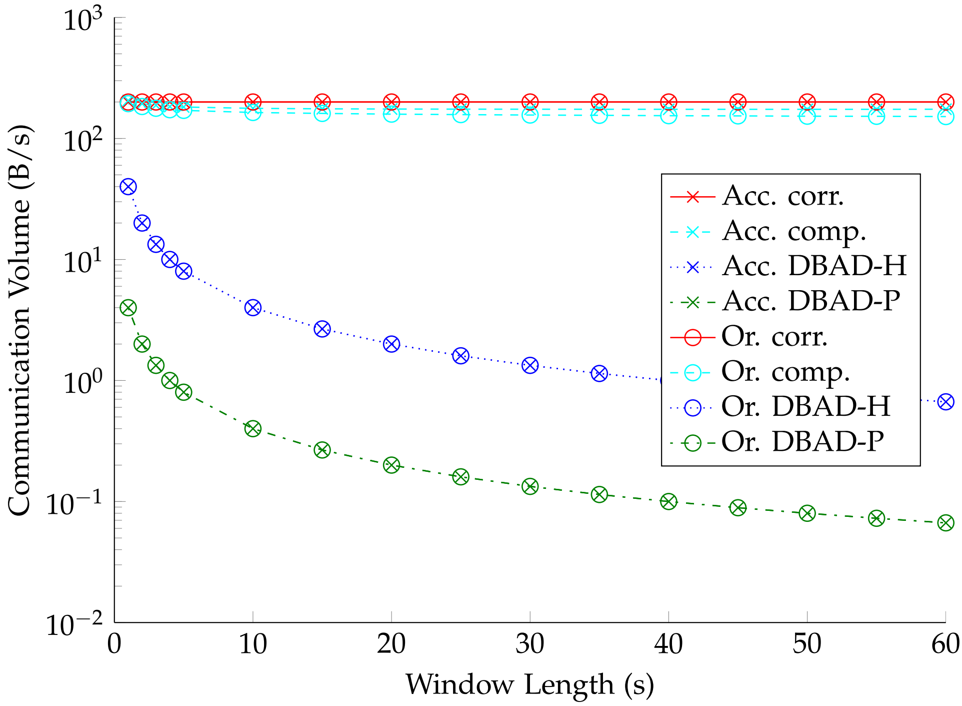


Figure 32.: Communication volumes for cross-correlation (corr.), cross-correlation using compressed values (comp.) and DBAD over orientation (or.) and acceleration (acc.) signals

For smaller window sizes, the compression overhead reduces the advantages of compression (orientation) or even makes it counter-productive (acceleration), where as window size increases the savings become more pronounced, at about 175 B/s for acceleration and 150 B/s for orientation, being able to save around 12.5% and 25% respectively. The distributed algorithms however greatly outperform the centralized approaches. At their optimal window length of 5 seconds, communicating histograms between nodes (in this case 20 buckets) requires only 8 Bytes/second of communication, and communicating models a factor of 10 less. Concretely these are either π, μ and σ values for acceleration data, π, θ and m values for orientation data respectively), as shown in Eq. (13). This is 94.7% and 99.5% reduction when compared even to the centralized approach with lossless data compression for the histogram and model-based distributed methods respectively.

One major difference between the distributed approaches and the centralized approach is the use of P2P communication which has a limited communication range. We evaluated the effect of this by varying the effective communication range of individual nodes using the location information annotated from the video. For

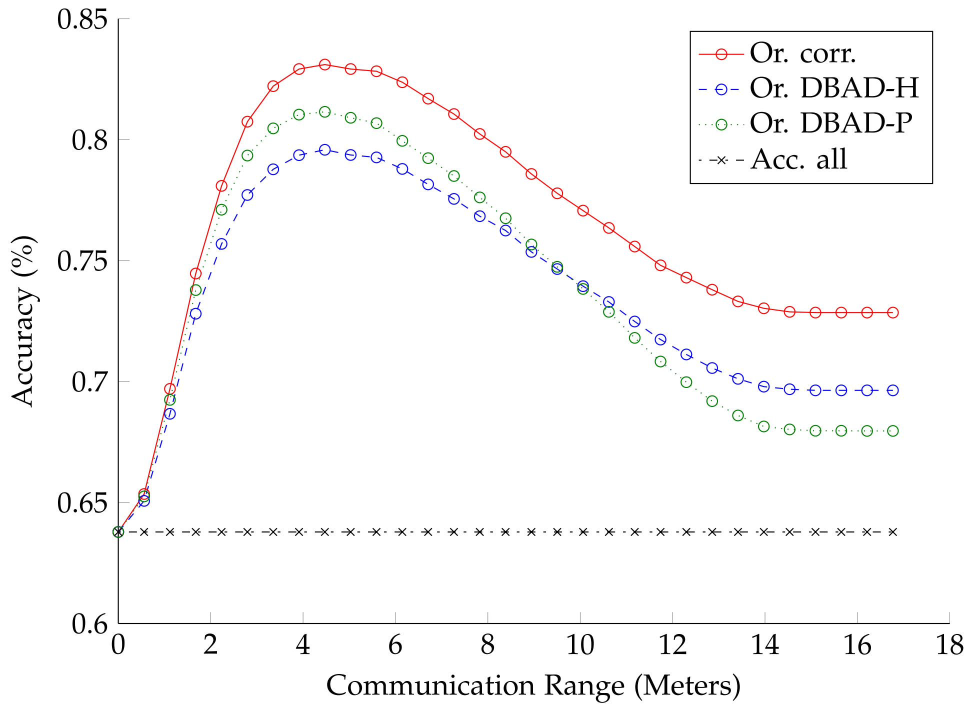


Figure 33.: Performance over communication range for cross-correlation (corr.) and DBAD over orientation (or.) and acceleration (acc.) signals for a 5s window

a given range, nodes are able to only communicate with other nodes which are within a circle with radius equal to the range.

Fig. 33 shows the accuracy results when the communication range of the devices is limited in simulation. At maximum range all nodes can communicate with each other across all experiments. As the range is decreased, the accuracy of the all methods increases to an optimum at 4.5m of 83.1% for the centralized approach, 79.6% for the histogram-based approach, and 81.2% for the approach using model divergence. Decreasing the communication range further incurs a sharp drop, with accuracy eventually dropping off to noise as the distance approaches 0. The optimum of 4.5m is there length where affiliated links are maximized and non affiliated links are minimized within the neighborhood of each node.

The results are demonstrated in Fig. 34, where similarity matrices are displayed instead of disparity for visibility reasons. Each row and column are subjects from 1 to 10, and index i, j is the similarity between subject i and j. In Fig. 34a) a typical clustering of a 5 second window by DBAD-P algorithm is shown for two groups. The difference in the similarity between subjects can be seen, but two groups can be identified, one in the upper left and one on the lower right. This also leads to noise in the identification of group affiliation in the same column of Tab. 12. In Fig. 34b) both are in different locations but the heading is similar, as is the case with groups 2 and 4 in Fig. 28. This leads to a drop in precision in Tab. 12 for that window. In Fig. 34c), a communication range of 5m greatly increases precision as most inter-group links are removed, but recall lags, as intra-group similarity fails to correlate group affinity. Filtering over the entire experiment Fig. 34d) improves all values, but errors are still caused by intra-group similarity values. The problem with intra-group similarities is demonstrated by Fig. 28(1), where the heading of the individuals in the group differs dramatically. Note that here we use precision and recall for demonstration purposes, but for experiments where with 1 or no groups, F-score and either one or the other of these metrics is undefined.

Finally, we ran a simulation to compare the local resource footprints of the various approaches. The values presented in Tab.13 are simplified approximations, calculated from the bitrate and power consumption of different communication technologies [1][3], and processing times and consumptions. This is modeled on an Android Nexus 4 device where processing occurs on a single core which has a consumption of 0.5W. Detecting affiliation using distributed cross-correlation is impractical due to the high response time and total energy cost of classification. The costs are due to the high communication volumes and consumptions caused by communicating raw sensor data over P2P channels. The centralized approach however has an expensive communicator, but the high bandwidth means low communication times. Processing is also offloaded, therefore processing time is low, and total energy is low as well.

DBAD-H has low processing time because model fitting is avoided, and P2P communication reduces the cost of communication even with the reduced bitrate. The total cost of energy of DBAD-H is therefore 24% lower than for centralized cross-correlation. DBAD-P has a more processing for model fitting and analysis than DBAD-P, and therefore increased response time as well, but the total energy required drops due to reduced communication. Nonetheless, DBAD-P reduces total energy consumption with respect to DBAD-H by a further 24% or by 43% compared to centralized cross-correlation.

8.7 DISCUSSION

Due to the nature of the problem, subjects who are in the same group generate similar sensor patterns for reasons discussed in Sec. 8.2. However, subjects in different groups may appear to be similar for periods of time, e.g. when both groups walk in the same direction, as is the case with groups 2 and 4 in Fig. 28. By observing subjects for a long enough period (extending window

3 http://www.csr.com/sites/default/files/white-papers/comparisons_between_low_power_wireless_technologies.pdf

Table 12.: Confusion matrices of affiliation (AF) and non-affiliation (NAF) with ground truth (GT) and classification (CL) and resulting metrics accuracy (acc.), precision (prec.), recall (rec.) and F-score, corresponding to the disparity matrices in Fig 34.

↓ GT, CL →	AF	NAF
AF	38	12
NAF	6	44

Acc.	Prec.	Rec.	F-score
0.82	0.86	0.76	0.81

↓ GT, CL →	AF	NAF
AF	40	10
NAF	44	6

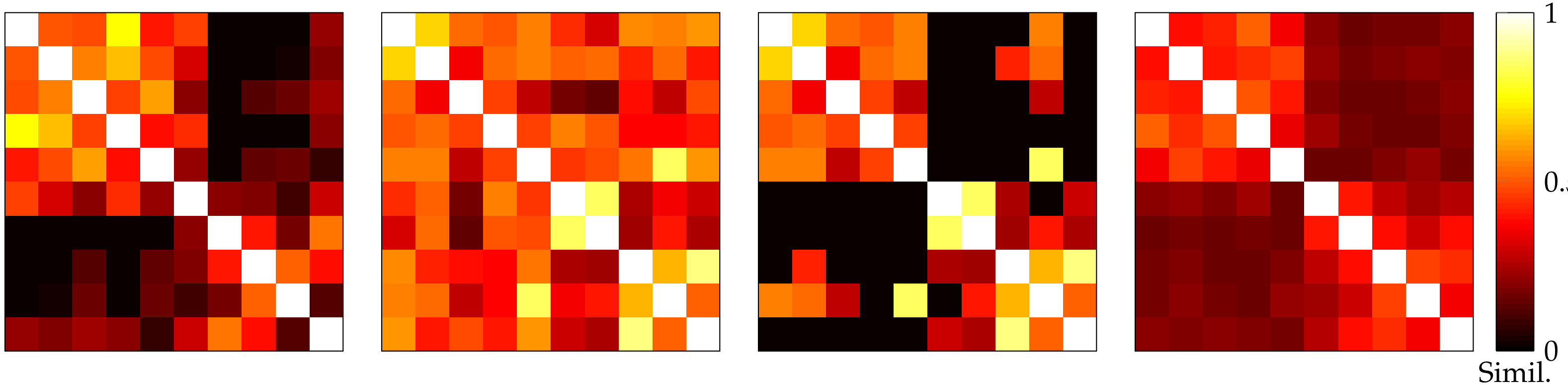
Acc.	Prec.	Rec.	F-score
0.46	0.48	0.8	0.6

↓ GT, CL →	AF	NAF
AF	38	12
NAF	8	42

Acc.	Prec.	Rec.	F-score
0.8	0.83	0.76	0.79

↓ GT, CL →	AF	NAF
AF	42	8
NAF	0	50

Acc.	Prec.	Rec.	F-score
0.92	1	0.84	0.91



a) b) c) d)

Figure 34.: Similarity between subjects in a two-group experiment for a window size of 5s using the DBAD-P method and orientation. **a)** under normal conditions, **b)** when both groups have similar headings, **c)** when the communication range is 5m and **d)** when averaged over the whole experiment.

Table 13.: Resource consumption analysis for a centralized correlation approach, distributed correlation approaches and novel DBAD-based methods for 1 Classification of 5 Seconds

Approach	Memory Used (B)	Comm. Tech.	Comm. Per Classification (B)	Comm. Time (ms)	Comm. Energy (mJ)	Proc. Time (ms)	Proc. Energy (mJ)	Total Time (ms)	Total Energy (mJ)
Cent. Corr.	2000	3G	1500	29.8	13.53	61.95	0.03	91.75	13.56
Acc. Corr. Comp.	2000	BT 4.0	18000	417.66	191.85	110.4	0.06	528.05	191.91
Or. Corr. Comp.	2000	BT 4.0	21000	487.27	223.83	110.4	0.06	597.66	223.89
DBAD-Hist	2160	BT 4.0	960	22.28	10.23	62.98	0.03	85.26	10.26
DBAD-PDF	4280	BT 4.0	720	16.71	7.67	168.47	0.08	185.17	7.76

size), the centralized approach can make these temporary phenomena irrelevant as demonstrated in Sec. 8.6. For the distribution-based approaches however, extending the window size reduces effectiveness as the characteristics of the signal disappear into a flat distribution after enough directional changes [13]. This effect is also compounded by a weakness in the distributed methods themselves, as PDFs and histograms both ignore the time component of the signals.

Take for example two individuals who walk in opposite directions for a period of time, then turn around 180 degrees and walk back the way they came for the same period. In this scenario, although the individuals exhibited very different behavior, the distribution over orientation for that period would appear identical. For this same reason, intra-group affiliations are difficult to correctly recognize, as heading varies over members depending on their location. This is a sensor issue, which indicates that the heading feature is not a perfect fit for intra-group affiliation. However, correlation does not use the absolute value of the signal but rather analyzes covariance over time. The distributed method is therefore slightly worse, even with filtering, compared to the correlation approach which is more robust in this respect. The indication is that the P2P DBAD methods are weak against variance from sensors with respect to intra-group affiliation.

The fact that the distribution-based approaches bring with them this inherent weakness also explains why the low-pass filter is so effective. The filter allows the p2p methods to deal with short-term similarity between non-affiliated subjects by extending the observation range for any given affiliation decision. Reducing communication range however can remove these ambiguities entirely, as the members of different two groups are often not compared with each other if they are outside the communication range ψ (again observe groups 2 and 4 in Fig. 28).

One application is for support of social network applications by allowing automated sharing or tag recommendation based on user affiliations. Other applications include life-logging systems which could document who we spent time with. The DBAD approach can also be used to support P2P group activity recognition [10] by allowing group constituents to be identified. The novel algorithms presented here have not been evaluated in large groups or crowds, however the evaluation gives some insight into the uses there. The P2P methods only use neighboring nodes, meaning that the effort required by each device is dependent on the density of the crowd and not the size as a whole. The complexity of DBAD-P for each node at each iteration therefore scales with $\mathcal{O}(K|\mathcal{V}|)$, where K is the number of modalities in the behavior of the individual and $|\mathcal{V}|$ is the number of immediate neighbors, where DBAD-H scales with $\mathcal{O}(|\mathcal{V}|)$. In emergency situations, prevention or management systems must be aware of group affiliations in order to manage groups of individuals as whole. Contradictory instructions to different individuals of the same social group will cause confusion and may be partially or fully ignored or disobeyed. Using these methods, management systems could disseminate messages to different individuals, and then allow these messages to disseminate along P2P links classified as intra-group affiliation. Furthermore a combination of in-network similarity assessment and server-side clustering approaches would alleviate bandwidth consumption caused by GAD in crowds while enabling a full emergent group analysis.

8.8 CONCLUSION

Humans often build groups for social reasons, and because groups can be better at reaching goals than the individuals separately [9]. However, often several different groups have different goals and occupy the same space, and must therefore be differentiated. Current differentiation methods consist of centrally aggregating sensor information and then clustering the emergent sensor image. However this approach is not feasible when network communication is too expensive, either due to the scale or the environment.

In response to the Challenge 3 (P2P Group Affiliation Detection) in Chap. 1, a method for distributed, P2P recognition of group affiliations was presented in this chapter, using the divergence of sensor data distributions as an indicator of similarity (DBAD). When addressing the problem from a P2P standpoint, the challenge changes slightly from recognizing group boundaries from the observers point of view, to recognizing subjective affiliations to local neighbors from the point of view of each group member. Divergences can either be calculated using models of individual behavior (DBAD-P) or using histograms of sensor data (DBAD-H). The requirement is that the sensor used is an indicator of social proximity for the kinds of social connections which define the particular group. The results show that the output of the proposed method fluctuates with instantaneous recognition rates only slightly over random. However group affiliations can still be detected 93% of the time by applying a low-pass filter to that output signal.

We show that only having a limited range of communication actually improves system performance, by allowing the devices to implicitly use location information without requiring a further sensor. Analysis of resource consumption indicates that time-series analysis approaches in the network are infeasible due to time and energy required for communication. DBAD-H and DBAD-P reduce energy consumption by 24% and 43% respectively, where DBAD-H reduces response time by 7%, but DBAD-P doubles it, indicating there is a trade-off between energy consumption and response time. Both distributed methods increase the amount of memory used well, although usage remains under 4.5 kB. None the less, both methods are independent of centralized resources and can be applied in distributed P2P systems. DBAD therefore addresses the challenge of P2P GAR while respecting the challenge of preserving the primary function of the devices.

Bibliography

[1] Niranjan Balasubramanian, Aruna Balasubramanian, and Arun Venkataramani. Energy consumption in mobile phones: a measurement study and implications for network applications. In *Proceedings of the 9th ACM SIGCOMM conference on Internet measurement conference*, IMC '09, pages 280–293, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-771-4.

[2] Philipp Berens. Circstat: A matlab toolbox for circular statistics. *Journal of Statistical Software*, 31(10):1–21, 9 2009. ISSN 1548-7660.

[3] Christopher M Bishop. *Pattern Recognition and Machine Learning*, volume 4 of *Information science and statistics*. Springer, 2006. ISBN 9780387310732.

[4] Arjen Boin and Allan McConnell. Preparing for Critical Infrastructure Breakdowns: The Limits of Crisis Management and the Need for Resilience. *Journal of Contingencies and Crisis Management*, 15(1):50–59, March 2007. ISSN 1468-5973.

[5] Oliver Brdiczka, Jérôme Maisonnasse, Patrick Reignier, and James L. Crowley. Detecting small group activities from multimodal observations. *Applied Intelligence*, 30(1):47–57, July 2007. ISSN 0924-669X.

[6] S. Calderara, A. Prati, and R. Cucchiara. Mixtures of von mises distributions for people trajectory shape analysis. *Circuits and Systems for Video Technology, IEEE Transactions on*, 21(4):457–471, 2011. ISSN 1051-8215.

[7] T L Chartrand and J a Bargh. The chameleon effect: the perception-behavior link and social interaction. *Journal of personality and social psychology*, 76(6): 893–910, June 1999. ISSN 0022-3514.

[8] Trinh Minh Tri Do and D. Gatica-Perez. Groupus: Smartphone proximity data and human interaction type mining. In *Wearable Computers (ISWC), 2011 15th Annual International Symposium on*, pages 21–28, 2011.

[9] D.R. Forsyth. *Group Dynamics*. International student edition. Thomson/Wadsworth, 4th edition, 2006. ISBN 9780534368227.

[10] D. Gordon, M. Berning, R.E. Masri, M. Beigl, J. Blanckenstein, and J. Klaue. Wor-mac: Combining wake-on-radio with quality-of-service for intelligent environments. In *Networked Sensing Systems (INSS), 2012 Ninth International Conference on*, pages 1–8, 2012.

[11] Dawud Gordon, Markus Scholz, Yong Ding, and Michael Beigl. Global peer-to-peer classification in mobile ad-hoc networks: a requirements analysis. In *Proceedings of the 7th conference on Modeling and using context (CONTEXT)*, pages 108–114. Springer, 2011. ISBN 978-3-642-24278-6.

[12] Dawud Gordon, Martin Wirz, Daniel Roggen, Michael Beigl, and Gerhard Tröster. Group affiliation detection using model divergence for wearable devices. In *Wearable Computers (ISWC), International Symposium on*, pages 1–10, 2014. (*To Appear).

[13] Gilbert R. Gordon and Julius Surkis. A general simulation methodology for optical bar code evaluation. *Journal of Statistical Computation and Simulation*, 8 (2):85–101, 1978.

[14] Dirk Helbing and Pratik Mukerji. Crowd Disasters as Systemic Failures: Analysis of the Love Parade Disaster. *Disasters*, 2012.

[15] Mikkel Baun Kjærgaard, Martin Wirz, Daniel Roggen, and Gerhard Tröster. Detecting pedestrian flocks by fusion of multi-modal sensors in mobile phones. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, UbiComp '12, pages 240–249, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1224-0.

[16] Raluca Marin-Perianu, Mihai Marin-Perianu, Paul Havinga, and Hans Scholten. Movement-based group awareness with wireless sensor networks. In *Proceedings of the 5th international conference on Pervasive computing*, PERVASIVE'07, pages 298–315, Berlin, Heidelberg, 2007. Springer-Verlag. ISBN 978-3-540-72036-2.

[17] Mehdi Moussaïd, Niriaska Perozo, Simon Garnier, Dirk Helbing, and Guy Theraulaz. The Walking Behaviour of Pedestrian Social Groups and Its Impact on Crowd Dynamics. *PLoS ONE*, 5(4):7, 2010.

[18] Daniel Roggen, Martin Wirz, Gerhard Tröster, and Dirk Helbing. Recognition of Crowd Behavior from Mobile Sensors with Pattern Analysis and Graph Clustering Methods. *Networks*, 00(0):1–24, 2011.

[19] D. Sumpter. The principles of collective animal behavior. *Philosophical Trans. of the Royal Society*, 2006.

[20] M. Wirz, D. Roggen, and G. Troster. A wearable, ambient sound-based approach for infrastructureless fuzzy proximity estimation. In *Wearable Computers (ISWC), 2010 International Symposium on*, pages 1–4, 2010.

9 Distributed Recognition of Group Activities

9.1 ABSTRACT AND CONTEXT

In Chap. 4 requirements were presented which distributed GAR approaches should fulfill:

- Requirement 1 (Survival of Node Failures)
- Requirement 2 (Recovery from Node Failures)
- Requirement 3 (Ability to Approximate the Mapping Function)
- Requirement 4 (Preservation of the Primary Function of the Device)

Chap. 7 demonstrated the potential of unsupervised clustering to characterize individual behavior within a group. Chap. 8 used the clustering abstraction level to identify affiliation between individuals. In this chapter I present a method for conducting GAR in a fully distributed manner using the abstraction level of unsupervised individual behavior clustering, combined with a supervised learning approach. The novel method is evaluated as to how well this approach fulfills the requirements 1, 3 and 4 set forth in Chap 4. Requirement 2, namely that the method be able to recover from failures by incorporating new members into the group, is left for future work (see Sec. 10.4). Some approaches and a plan for continuing research in this direction is presented in Chap. 10. In total, this chapter here addresses Challenge 4 (P2P Group Activity Recognition), namely the ability to infer emergent group behavior in a P2P network. The content of this chapter has been accepted for publication at ISWC 2014 [10].

9.2 INTRODUCTION

Human beings are social creatures, and as such we spend most of our time in groups [18]. It has been shown that groups are better than individuals at accomplishing tasks, which is often why they are formed in the first place [7]. Understanding group behavior and context is then crucial for intelligent environments. The process of understanding what a group is doing, or the physical attributes of the group behavior, is called group activity recognition (GAR) [9].

The behavior of the group is emergent behavior, emerging from the personal characteristics of the individual members and the group dynamic [15, 7]. Human perception of group behavior can be explained by Gestalt Theory, where only when observing the complete whole can its properties be described (see Fig. 35) [15, 7]. Recognition of that behavior is irrefutably bound to human perception, as it is the human who labels a group activity based on his/her perception. Kurt Lewin, a pioneer of modern social psychology, uses the term "emergence" to signify that the properties of the behavior of the group are fundamentally different than the properties of the behavior of the individuals, or of the "sum" of those behaviors [15], a definition which we follow. This is a generalization of many definitions of the term emergence [5], where all agree that emergence is a difference between (human) observations of micro and macro properties.

Mobile devices such as Smart Phones present an attractive platform both for human activity recognition (HAR) and the recognition of emergent group activities. Sensor information from these devices is used by a recognition algorithm to learn the ability to make the same observations as a human would. This paper shows that a global observer - a centralized detection algorithm - having the complete picture can perform detection of emergent group actives. It then analyzes if a local observer - a decentralized algorithm running on individual devices - having limited peer-to-peer communication with other peers can also deliver such observations and studies how well such local detection performs in comparison with a centralized approach. It also studies the communication range required to detect the emergent behavior with respect to the spatial size of the group, and if sparse communication can still reach acceptable detection rates compared to a global observer. We also study how much energy can be saved using the decentralized approach and how much energy needs to be invested for local processing instead.

We present novel methods for distributed GAR using distributed probabilistic inference (DPI) combined with loopy belief propagation (LBP) [19]. For each group activity, the behavior is broken down into individual clusters using unsupervised methods. Each node then estimates its belief over its local clusters for all group activities given current sensor observations, and then communicates this information to its neighbors. All nodes then iteratively update and re-communicate their beliefs based on their local sensory evidence, the belief estimates received, and a

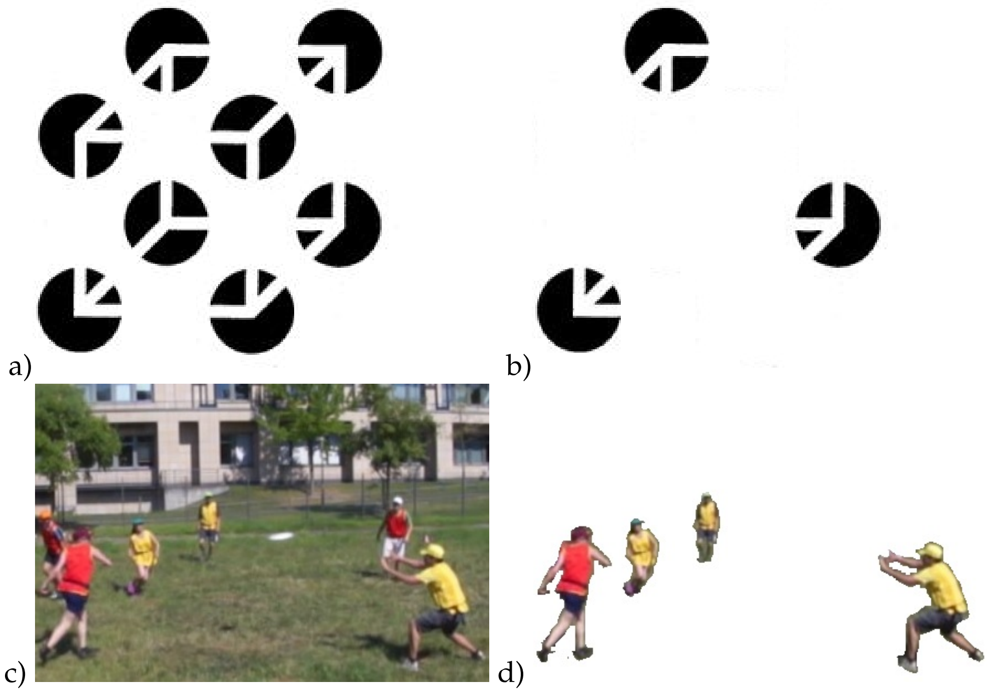


Figure 35.: Following Gestalt theory, an image of a cube emerges from distinct objects (**a**). An incomplete set obscures the emergent properties (**b**). The same is valid for group activities where from a complete image a sport can be identified from context (**c**). A partial view of player's physical behavior without context makes identification difficult (**d**).

model of individual-to-individual group dynamics. The network then iterates and converges towards a response prediction. We present two methods for LBP, one linear regression over soft posterior probabilities over user behaviors (SLBP) and one using expectations based on hard classifications (HLBP).

The novel algorithms are evaluated using an experiment in team sports. 10 subjects play 6 different sports and are monitored using Android phones as wearable sensors. The experiment naturally creates emergent group behavior where the algorithms are then evaluated in terms of their effectiveness at recognizing that behavior. The evaluation is in terms of performance with respect to the number of iterations required for convergence. The effects of P2P communication range are also evaluated by simulating local links using the devices' GPS locations. The results are then compared with centralized inference of group behavior, where the complete set of sensor data provides the complete picture of the emergent behavior.

The results show that centralized inference of emergent group behavior when presented with the complete set of group sensor data is relatively straight-forward, approaching an F-score of 0.81 for a window length of 2 seconds and 0.96 for 10 seconds. However inference using solely the data of each subject individually is poor at around 0.55 for the same window. The novel DPI-SLBP approach begins at iteration 0 at the same value as with individual subject inference, but then rapidly improves with each iteration, surpassing the centralized naive Bayes approach after three iterations and converging to an F-score of 0.84 after about 10 iterations. This method incurs an increase in the amount of local memory consumed and processing required, but reduces the amount of energy required overall for classification factor of almost 7.

The simplified DPI-HLBP algorithm performs similarly but converges to a lower value of 0.81, just under the centralized approach. However compared to DPI-SLBP, memory consumption energy required for classification drop by a factor of more than 6, which is 40 times less than that required by the centralized approach.

9.3 RELATED WORK

Recognizing emergent behavior has historically been a topic in HAR for quite some time, although it has not been named as such directly. Human perception of other human activities is also governed by Gestalt Theory in that we observe a single individual instead of a collection of limbs, therefore inferring behavior of a single individual from the distributed behavior of their body is emergent [2]. The problem is however simpler, as limbs don't change roles, and there interactions with each other are mechanical in nature.

Multi-user activity recognition (MAR) is the process of recognizing the activities of multiple individuals in parallel [9]. Wearable sensing approaches leverage centralized inference structures to infer multiple activities in parallel. Subjects may be interacting with each other or may even be in the same group, but the problem presented is of a different nature, recognizing distinct activities for different subjects [12], as opposed to emergent group activities. However often these approaches gray the boundaries between MAR and GAR, where some of the activities recognized are group labels of emergent activities, where behavior of multiple subjects is necessary to infer certain activities, and others are single-user activities [23].

Approaches have also been presented to distributing the recognition process for contexts and activities across the network of sensing nodes [25]. Distributed methods leverage knowledge about the conditions which govern distributed sensing to fuse information into recognition, e.g. someone climbing a fence will create similar disturbances at multiple measurement locations [25]. However these approaches are not focused on emergent behavior. One approach which was inspirational for our research here is a method of distributed probabilistic inference for sensor calibration [21]. The approach uses the assumption that the distance between the measurement locations of nodes will provide temperature measurements which are correlated with each other, over which a potential function can be built. The approach is however fundamentally different from emergent GAR, as it does not address the human factor, where this factor is the main cause of complexity in HAR as a field in general. It is more akin to MAR, where each node must estimate its own bias under the assumption that measurements are correlated. For that application, loopy belief propagation does not converge, requiring a complex networking architecture for clique structuring and belief propagation [20].

Other sensing modalities have also been used for recognizing group activities. Video systems present an advantage as they are able to view local individual behavior and the resulting emergent group behavior simultaneously [4], and are also able to scale to larger groups. They are also able to measure certain properties of individual roles, for example a player's position in an American Football team [16]. However such approaches are accompanied with infrastructure requirements for communicating and processing the constant flow of video data, and therefore can only be applied in instrumented environments. Many human interactions are verbal, and monitoring these conversations using microphones also provides insight into the group activity [13]. An understanding of the audio situation can even allow extraction of certain types of role information present in the group behavior [6]. However for activity recognition, microphones are an orthogonal sensing approach as they do not sense the physical parameters of the behavior directly, and extracting this information from audio data is a different branch of research with its own set of challenges.

Monitoring location has also been shown to give insight into emergent properties of larger groups or crowds [24]. Here emergent spacial properties can be computed as a function of the location of multiple individuals and the properties of the space in which they are located. Adding motion sensors also allows properties such as affiliation of users to each other and to groups, building subgroups within a larger group or crowd [22]. Emergent behavior has also been addressed in the separate but related field of swarm intelligence, usually addressing this behavior in animals and insects [11]. Here the problems addressed usually have one of three different goals, either looking to simulate the emergent group behavior based on models of individuals (generation) [14], discover the rules governing individuals based on the emergent behavior produced (discovery) [17], or evaluate the correctness of assumptions about the relationship between local agents and emergent group behavior (evaluation) [14]. Our approach here differs from this field because we wish to **predict the emergent group behavior** based on observations of agents (humans) who are admittedly far too complex to model using expert knowledge. We therefore approach the problem from a machine learning standpoint in order to discover and model pertinent characteristics of agents in an automatic fashion, using only the sensing devices.

9.4 CONCEPTS AND APPROACH

In this section we present the concepts and theories which motivate the design decisions made. We begin with the fundamental principles which govern group behavior from the field of group dynamics and social psychology. Inspired by these abstract models and theories, we construct concrete models and methods for modeling and classifying group behavior in a probabilistic fashion. The goal of this section is to create models for centralized and distributed recognition of emergent group behavior, methods for evaluating them independently, and a metric for judging the degree of emergence of a recognition problem given specific models.

9.4.1 *From Field Theory to Probability Theory*

Kurt Lewin's "Field Theory" [15] states that the individual behavior $B^{\text{ind.}}$ of members of a group is a function of their individual attributes and characteristics c and the social environment of the group E. He quantified this as "interactionism" in Eq. (23).

$$B^{\text{ind.}} = f(c, E) \tag{23}$$

He stated that the resulting group behavior is "a dynamic whole [that] has properties which are different from the properties of [its] parts or from the sum of [its] parts" [15]. "According to Lewin, whenever a group comes into existence, it becomes a unified system with emergent properties that cannot be fully understood by piecemeal examination" [7]. However, the behavior of an individual is not only governed by their individual attributes, but also their role in the group dynamic [3]. These roles, as with group behavior, are generated as emergent norms when the group is formed, and members adapt their behavior to fit the norms for different roles [7]. As a result we can update Lewin's equation to account for emergent roles $\rho \in R$: $B^{\text{ind.}} = f(c, \rho, E)$.

From a probabilistic standpoint, we can model the probability p of all group behaviors $p(B)$ as the joint probability of all individuals. We know that this is the joint distribution of C, R and E which symbolizes the social dynamic:

$$p(B) = p(B^{\text{ind.}}_{s_1}, B^{\text{ind.}}_{s_2}, \ldots, B^{\text{ind.}}_{s_n}) = p(C, R, E) \tag{24}$$

where C is the set of characteristics of all group members $c \in C$. When Lewin used this term, he was referring to all possible relevant characteristics of the individual, psychological, sociological, physiological, metaphysical, etc., meaning the state space of C approaches infinite. However, for activity recognition we focus on the physical characteristics of contexts, activities and behaviors. These physical properties can be observed and differentiated using sensors (the premise for HAR), therefore we make the assumption that we can replace the infinite state-space of C

with our observations of the physical properties of C, referred to as X. Here we use the notation x_s^τ to indicate a single observation, or observations over a window, for subject s at time τ. X_s refers to all observations for subject s, X^b refers to the evidence of all subjects for a single group activity, and X is the complete set of observations for all subjects and activities. We now have the following equation for the joint probabilities of group activities: $p(B) = p(B_{s_1}^{\text{ind.}}, B_{s_2}^{\text{ind.}}, \dots, B_{s_n}^{\text{ind.}}) = p(X, R, E)$. We can break down the right hand side to approach the problem of differentiating $b \in B$ given observations and models as the following:

$$p(B|X, R, E) = \prod_{s_i \in G} p(B|X_{s_i}, \rho_{s_i}, E) \tag{25}$$

However, we still have the role of each user in the equation. Identification and annotation of roles and individual-to-role affiliation requires behavioral experts, meaning this approach lacks versatility and requires a great deal of preparation. Also, the double annotation of group activity, and role greatly increases the effort required for training. To circumnavigate this issue, we make a key assumption. The evidence X is conditionally dependent on both the individuals characteristics, and the role of the individual in the group [7]. We can therefore use this conditional dependence to gain the pertinent information about the observations and roles. This is done by combining the evidence in its conditionally dependent form using a transformation into a different space:

$$K = f(X \otimes R) = \forall_{b \in B} \forall_{s \in G} \text{Clust.}(X_s|b) \tag{26}$$

We cluster the evidence into clusters $\kappa \in K$, where κ_s^b is a cluster from subject s generated by group behavior b and their role ρ_s^b in that behavior. To be clear, we are not making the assumption that these clusters equate semantically to the role of the individual in the activity. Our assumption is that the clusters contain a factorization of the conditional dependencies between the evidence, the roles and the group behavior, or $p(K|B) = p(X, R|B)$. For example, assuming experts in the sport soccer inform us that one of the roles is goal-tender, no single cluster would equate to this role for a specific subject. The assumption is that the role goal-tender for a specific subject will however generate one or several clusters in which the different modalities in which this user behaves in this role are quantified. It is also possible that a similar behavior from the same or different subject in the same or different group activity could generate a cluster of the same dimensions.

9.4.2 *Modeling and Classifying Group Activities*

The clustering approach used is a probabilistic clustering using Expectation Maximization. For each group activity and subject, X is separated into X_s^b and then clustered, yielding clusters K_s^b. The probability density function (PDF, or P) of the clusters for a subject and group activity is given by a Gaussian mixture model (GMM) [7]:

$$P(X_s^b|K_s) = \sum_{\kappa_s^b \in K_s^b} \pi_{\kappa_s^b} \mathcal{N}(X_s|\mu_{\kappa_s^b}, \Sigma_{\kappa_s^b}) \tag{27}$$

Each node s has clusters K_s where each cluster κ_s is generated by a certain group behavior b, giving a subset of clusters for each group activity $\kappa_s^b \in K_s^b$. These clusters now build the evidence function for inference of group activities. The posterior probability distribution $p(K|X)$ can be obtained using Bayesian inference, where each posterior is normalized using the following equation:

$$p(\kappa_s^b|x_s^\tau) = \underbrace{\text{Post.}(\kappa_s^b|x_s^\tau)}_{\text{GMM posterior}} \underbrace{\frac{\text{Like.}(K_s^b|x_s^\tau)}{\sum_{b' \in B} \text{Like.}(K_s^{b'}|x_s^\tau)}}_{\text{GMM likelihoods normalization}} \tag{28}$$

Here posteriors are generated over the Gaussian mixtures for each class K_s^b given an observation x_s^τ, after which the posterior distribution is normalized by the likelihood of all activity cluster models for that subject. Both the likelihood of a GMM and the posterior of a cluster given an observation are obtained by applying Bayesian inference and the Law of Total Probability [19]. It is important to note that due to the normalization in Eq. (28), the resulting probability distribution over all clusters for all activities for each subject (K_s) sums to 1. As will be explained later on, this step is necessary in order for nodes to be able to learn relative probability distributions of neighboring nodes based on histories of these distributions generated by observations. Classification of the current group activity at any point in time for a single subject is achieved by Eq. (29).

$$p(B = b|K_{s_i}, x_{s_i}^\tau) = \sum_{\kappa_{s_i}^b \in K_{s_i}^b} p(\kappa_{s_i}^b|x_{s_i}^\tau) \tag{29}$$

The classification approach of evaluating local posteriors using local evidence (Eq. (28)) can be used to evaluate the ability of a single node to infer the group activity based on local observations alone, which we call the **independent local inference (ILI)** method.

Returning to the original problem of inferring group behavior, we have now combined the user's role with the evidence in clusters. We now have the following equation: $p(B) = p(K, E)$. The term E – the social environment – is problematic, since it is difficult to quantify. We know, however, that group behavior can be observed by applying Gestalt Theory, meaning that observation of the whole allows it to appear in its emergent form, rather than as a sum of unrelated aspects. The indication is that for complete set of K, the effects of E are already present with respect to the interpretation of the group behavior B. The same concept can be seen in Fig.35, where presented presented with the complete image, a cube appears through the Gestalt principles in 35a. This is the emergent whole with properties different than the sum of the individual circles, which are actually not circles, but appear as such on the right in in 35b. Therefore observing all distributed observations in a single location should allow a complete view of the emergent group behavior. To examine this hypothesis, we used two methods of central inference.

The first is Bayesian inference using the complete probability distributions of K. For this purpose, $\overline{\xi}$ is constructed such that:

$$\overline{\xi}^{\tau} := \underset{K_s \in K}{\forall} \text{append}(p(K_s|x_s^{\tau})) \tag{30}$$

For each time-step τ, $\overline{\xi}^{\tau}$ is then a vector of the complete normalized posters across K. Using this set as observations, a naive Bayesian classifier is constructed to model $P(\overline{\xi}|B)$ and then to infer $p(B|\overline{\xi}^{\tau})$ for each time-step τ. This method is referred to as **centralized cluster-based inference (CCI)**. A more concrete description of the training process will be explained in Sec. 9.5. The second centralized method a more traditional naive Bayesian inference method using the observations directly. Here $P(X|B)$ is modeled as a GMM using the Expectation Maximization (EM) algorithm [7], and $p(B|X)$ can then be inferred, referred to hereon as **centralized naive Bayes (CnB)**.

Our goal is to recognize emergent group behavior using distributed mobile phones, where E can no longer be ignored. We propose to approach this problem using DPI with LBP. The missing information sampled by other nodes which is necessary in order to infer the emergent behavior is propagated through the network in the form of beliefs from other nodes. The equation for exact inference is shown in Eq. (31) where each node calculates its own belief based on its evidence, as well as its belief of other nodes states based on its own local evidence. Evidence is propagated through the network in the form of posteriors known as beliefs

$$p(K|X) = \prod_{s_i \in G} \underbrace{p(K_{s_i}|X_{s_i})}_{\text{local evidence}} \prod_{s_{j \neq i} \in G} \underbrace{p(K_{s_i}|X_{s_j})}_{\text{distributed belief}} \tag{31}$$

This method has the advantage of being exact, meaning the accuracy achieve is equal to that of a centralized system [19]. However, the state space of all random variables must be modeled redundantly at every node at process at each iteration step. More attractive are methods of approximate inference where each node propagates beliefs for other nodes based on its internal beliefs and a model for relations between its random variables and those of other nodes [19].

For standard DPI problems, clique graphs can be built to factor priors using some form of expert knowledge or assumptions about conditional independence between nodes [21]. These clique graphs are structured as directed a-cyclical graphs (DAG) and then traversed for belief propagation, guaranteeing that loops do not occur.

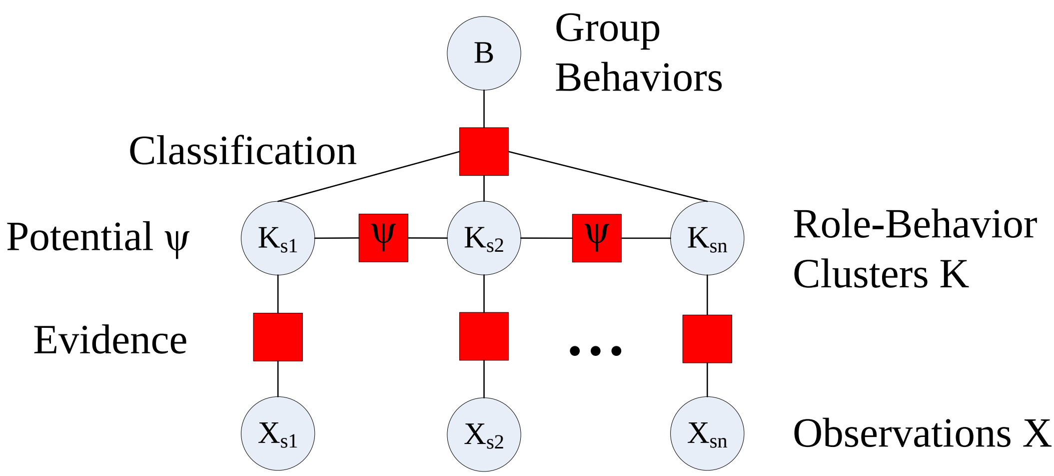


Figure 36.: Factor graph for DPI-LBP with evidence $p(K, X)$, potential $\psi(K_{s_i}, K_{s_j})$ (for All i, j, some are omitted), and classification $p(B, K)$

Constructing a recognition system in this manner is guaranteed to converge to the optimal solution with respect to a centralized system with a full sensory image of the emergent behavior. For group activity recognition this is not the case, as all variables within are influenced by the group dynamic E, making the entire group a single clique graph.

One approach which may or may not work in such situations is loopy belief propagation (LBP) where cyclical belief propagation paths are allowed. However several problems may occur depending on the inference problem. Several types of inference problems do not converge to single solution, and it is unclear which types of problems do and do not converge [19]. Also, the convergence rate, meaning how many iterations of belief propagation are required for convergence, are unknown. Luckily for emergent GAR, the system does converge in relatively few iterations with a resulting high accuracy, as we will see in Sec. 9.6. The equation for loopy and non-loopy belief propagation is given in Eq. (32).

$$p(K|X) = \prod_{s_i \in G} \underbrace{p(K_{s_i}|X_{s_i})}_{\text{local evidence}} \prod_{s_{j \neq i} \in G} \underbrace{\psi_{i,j}(K_{s_i}, K_{s_j})}_{\text{potential function}} \tag{32}$$

The potential function ψ can be any positive function which defines the relationship between the variables at subject s_i and s_j [7]. For this function we used linear regression [19] to model the relationship between the variables of each pair of subjects, or K_{s_i} and K_{s_j}. As stated before, the evidence function is trained using EM for unsupervised clustering of each subjects data for each group activity. Each potential function is trained using linear regression from the variables K_{s_j} of other subjects to each cluster κ_{s_i} separately. The resulting linear mapping takes the form:

$$\psi_{i,j} = \underset{\kappa_{s_i} \in K_{s_i}}{\forall} \underset{s_{j \neq i} \in G}{\forall} \alpha + [\beta_1, \beta_2, \ldots, \beta_n] \times [p(K_{s_j})] \tag{33}$$

Where $[p(K_{s_j})]$ is a column vector of all cluster posteriors $\kappa_{s_j} \in K_{s_j}$. This method we call **DPI with soft LBP (DPI-SLBP)** due to the "soft" posterior probability distributions which are mapped.

Each iteration consists of a local inference step followed by several update and classification steps. In the inference step, each node s_i generates a posterior distribution over its clusters using its local evidence function from Eq. (32), creating an initial estimate of the group activity based only on local estimates. In the first update step, this information is propagated to all neighboring nodes s_j, i.e. all nodes within range of one-hop communication. These nodes then convert this estimation of the posterior probability distribution over K_{s_i} to a belief over K_{s_j} using the mappings generated from Eq. 33. These beliefs are then combined with the current beliefs of node s_j over K_{s_j} and the resulting classification of the group behavior is reevaluated using Eq. (29) in the classification step. The update and classification steps then repeated until the network is satisfied that convergence has been reached, where we will empirically evaluate how man update steps are required in Sec. 9.6.

We also present a simplified version of the aforementioned DPI with LBP approach. That method requires each node to broadcast its posterior $p(K_s|X_s)$ to all neighboring nodes. Probabilistic classification works on the assumption that the most likely model given specific evidence is the correct model for a given instance. Based on this assumption, the most valuable information $p(K_s|X_s)$ is the most likely cluster in the most likely activity, namely $\text{argmax}_{k^b_{s_j}}\, p(k^b_{s_j})$. We present a simplified method where beliefs are calculated using only this information, instead of the full cluster posteriors $p(K_s|X_s)$. This simplified method takes the same form as Eq. (32) with a modified potential function presented in Eq. (34).

$$\psi_{ij}^{\text{simp.}} = p(K_{s_i}| \underset{k^b_{s_j}}{\text{argmax}}\, p(k^b_{s_j})) \tag{34}$$

Training for the simplified potential model is done by calculating the expectation $\mathbb{E}$ instead of the method using regression previously introduced. Training $\psi_{ij}^{\text{simp.}}$ for node s_i to s_j for $\text{argmax}_{k^b_{s_j}} \sum_{k^b_{s_j} \in K^b_{s_j}} p(k^b_{s_j}) = \kappa^b_{s_j}$ is done by creating a posterior probability for $p(K_{s_j})$ given the posteriors of instances of training data data where the most like behavior for node s_i is $\kappa^b_{s_j}$. Intuitively, we model a belief for the behavior of node j at times when node i is behaving in a specific manner. For example, if node i is behaving as a goal keeper in a soccer game, the belief that node j is playing soccer as a midfielder would (assumedly) be higher than than the belief that node j is serving a volleyball. The equation for computation of $\psi_{ij}^{\text{simp.}}$ is the following. First we define $\overline{\kappa}_{s_i}$ to be the most probable role-behavior cluster for s_i:

$$\overline{\kappa}_{s_i} = \operatorname*{argmax}_{k^b_{s_i} \in K^b_{s_i}} p(k^b_{s_i} | x^t_{s_i}) \tag{35}$$

Then, for each cluster κ_{s_i} the expectation is calculated given K_{s_i} and $x^\tau_{s_i}$:

$$\psi_{ij}^{\text{simp.}}(\kappa_{s_i}) = \forall_{\tau | \overline{\kappa}_{s_j} = \kappa^b_{s_j}} \mathbb{E}(\kappa_{s_j} | K_{s_i}, x^\tau_{s_i}) \tag{36}$$

where the probability of $K_{s_i}, x^\tau_{s_i}$ is given by Eq. (28). We refer to this method as **DPI with hard LBP (DPI-HLBP)** due to the hard role-behavior classification in the potential function. Lewin's definition of emergence in group behavior as the whole having properties different than the parts or the "sum" of those parts [15], and emergence is a function of observational difference between the micro and the macro [5]. We define a metric for evaluating this disparity. For a physical activity recognition system, trained to recognize a set of group activities identified by human observations, we define the **"degree of emergence"** ϵ as the proportional information gain, quantified using the F-score, of activity recognition with the complete picture, to the mean of activity recognition of all nodes using their local observations.

$$\epsilon(B|X) = \frac{\text{F-score}(p(B|X)) - \frac{\sum_{s \in G} \text{F-score}(p(B_s|X_s))}{|G|}}{\text{F-score}(p(B|X))} \tag{37}$$

This measure is dependent on and specific to the models used, the subjective observations (labels), and only for the behavior recognition problem, and does not necessarily be generalized over these parameters, other definitions of emergence, or other recognition problems.

9.5 EXPERIMENT AND PROCEDURE

To evaluate the approach detailed in the preceding section we constructed an experiment with emergent group activities. The activities performed were team sports, where the emergent behavior is the sport being played itself, based on the observations of the physical behavior of the individuals.

The devices used LG Nexus 4 Android devices with a custom application. The software sampled the accelerometer, magnetometer and gyroscope, each a 3 axis vector value, with the maximum sampling rate. The accelerometer measures on-body acceleration, the magnetometer delivers orientation and heading information relative to the local ambient magnetic field, and the gyroscope samples rotation information. Effectively a sample rate of about 50 Hz was delivered for

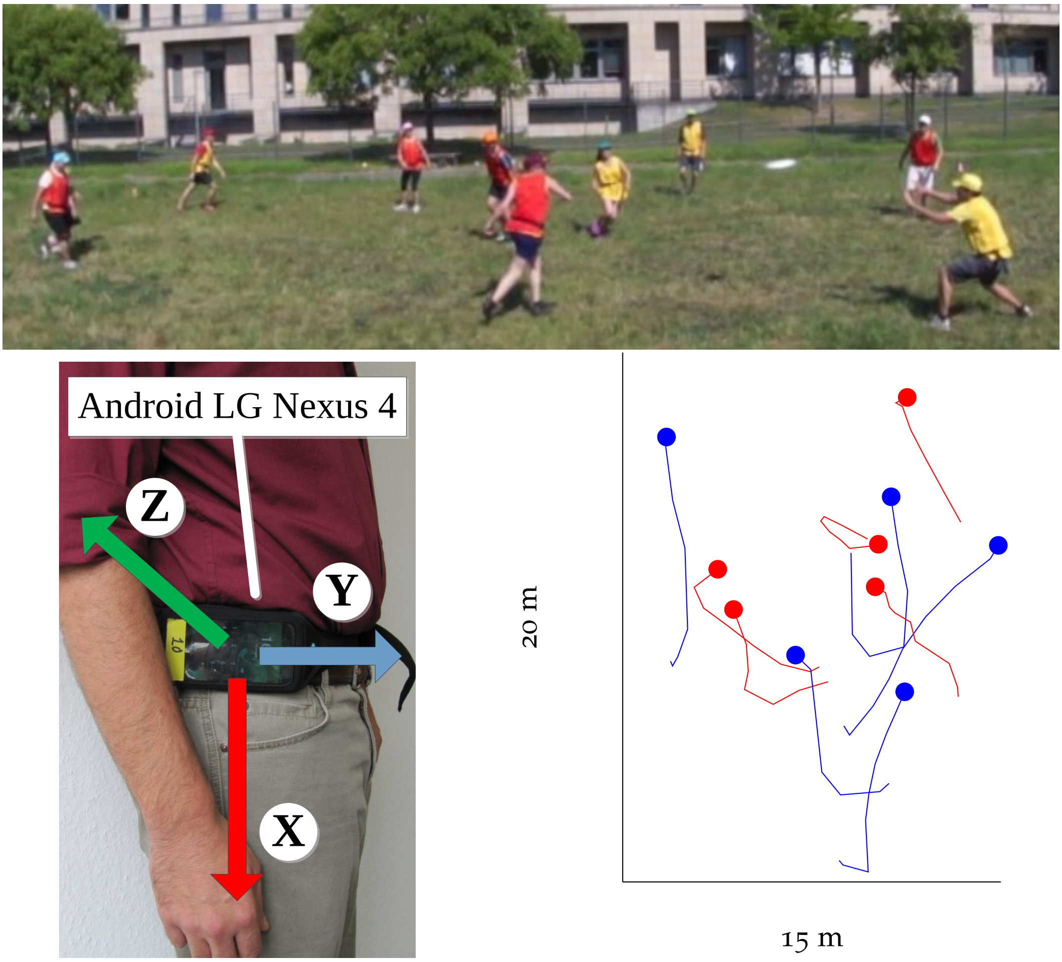


Figure 37.: The team sports group activity experiment (top) with on-body device placement (bottom left) and relative subject locations on the field (bottom right)

the accelerometer and gyroscope, while the orientation sensor only delivered an approximate sample rate of 20 Hz. In addition to the behavioral sensing, the devices sampled their absolute location using the GPS sensor. The location information was not used for group activity recognition, but was used for simulation of performance of a the P2P recognition system.

The devices where attached at the right side of the hip, as the hip has been shown to be the most beneficial single location for activity sensing [2]. This was done using an elastic sports belt for the device, where the phones where inserted into the belt with the face outwards and the top of the phone forward as shown in Fig. 37. 6 different team sports where performed by all subjects: **volleyball, badminton,**

football (soccer), ultimate Frisbee, touch rugby, and flunky-ball. Each sport was performed for 10 minutes, with a break between each type of sport. The experiment was conducted outdoors in an open field with a natural turf of dimensions 15m by 20m, and a video recording was made from an elevated standpoint of the experiment. The day was sunny with high temperatures around 29°C, making breaks between activities necessary. The subjects were made up of 7 males and 3 females. On a scale of 0 (no experience in any of the sport) to 10 (very experienced in all of the sports) over all sport types, the average experience was 4.5 with a variance of 3.5.

The data recorded was synchronized and input into an offline sensor replay mechanism in a MATLAB simulation environment, where the algorithms are implemented. 50% of the data is used to train the algorithms, and the other 50% for evaluating algorithmic performance. All sensor measurements where then hold-resampled to 50 Hz to provide equidistant measurements for feature calculation. GPS location annotations where also resampled and smoothed to account for asynchronous updates. This sensor data was cut into windows of lengths from 1 to 10 seconds, where the window is advanced by 0.5 seconds each iteration over which features where calculated. The features used were the mean and variance of the total acceleration signal, the mean and variance of the azimuth orientation with respect to the subject's body, and the mean and variance of the rotation around the X and Z axes (see Fig. 37 for orientation). These features calculated for subject s then represent the observations X_s of the subject, where τ is the last timestamp of a sensor data window. For each window length, all models are retrained and reevaluated using the features generated over the windows.

Based on these locations we simulated performance under different communication capabilities. We then simulated performance for a communication range ϕ of 5m, 10m, 15m, and 20m sequentially, compared to the diagonal of the field of 25m which is also a good approximation of the radius of the group. We used the relative Euclidean distance between two subjects $\text{dist}(s_i, s_j)$ based on their GPS coordinates, and judged them to be able to communicate if $\text{dist}(s_i, s_j) \leq \phi$. The timestamp used to evaluate $\text{dist}(s_i, s_j)$ is the final timestamp of the window τ, as this is the point where the network is able to evaluate the distributed evidence functions and communicate beliefs. we simulated performance with full inter-connectivity of all nodes in the network, meaning the range local P2P communication was greater than the maximum distance between any two subjects during the course of the experiment, or $\phi = \infty$. No multi-hop communication is implemented, simulated or required for the methods presented here. The results are generated using only loopy belief propagation and the models previously trained for this window length.

During the course of the simulation we evaluate the F-score of the described algorithms by constructing confusion matrices over the output of the algorithm for each node. How the output is determined based on a node's belief is described in

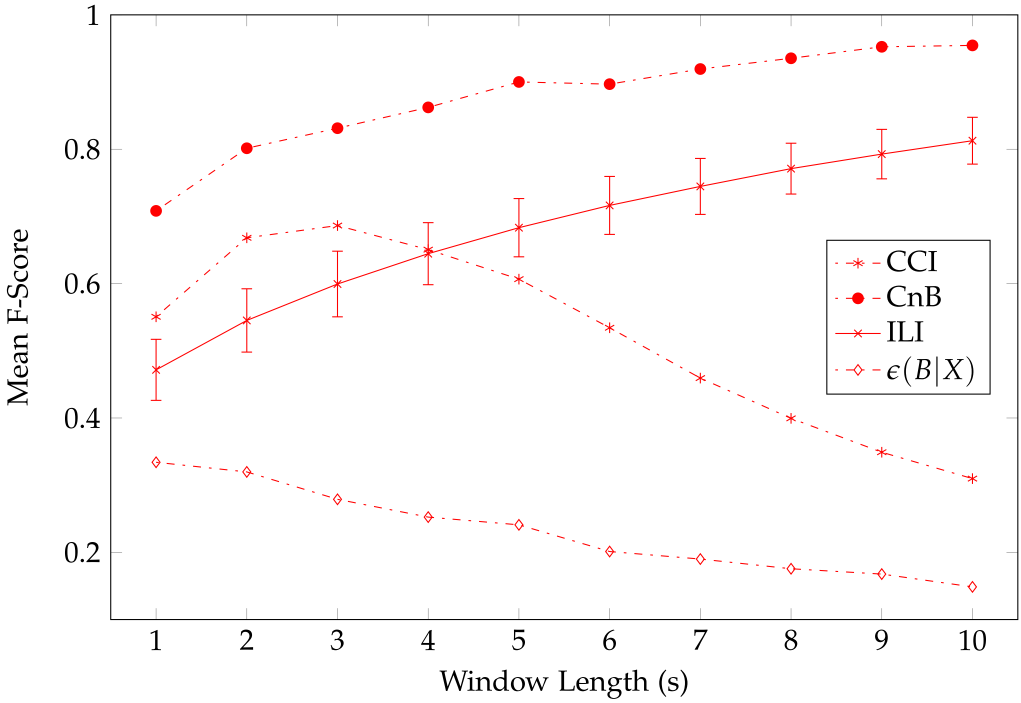


Figure 38.: Performance for centralized algorithms, distributed independent classification and the resulting degree of emergence

Eq. (29). This is monitored for each node, at each iteration of the belief propagation algorithm. We also monitored the processor time required for each operation and iteration, as well as the memory required to store and process information.

9.6 EVALUATION

The goal of the algorithms presented in Sec. 9.4 is to allow distributed mobile devices sensing the physical activities of individuals to be able to recognize the activity of the group. Since group behavior is emergent, the correct response is not dependent on any single node, but the combine implications of all distributed measurements. The presented DPI with LBP methods may be a solution, but there are open questions in the literature about their performance under GAR circumstances. For one, it is unclear if the algorithms will converge to response. If they do, it is unclear what the accuracy of that response will be, or how long it will take to converge.

9.6.1 *Centralized Recognition of Emergent Group Behavior*

To analyze performance of centralized inference of emergent group behavior using the complete picture of sampled sensor data, we looked at 3 different approaches which where explained in Sec. 9.4, namely CnB, CCI, ILI and the degree of emergence ϵ of this specific problem.

The results of the centralized analysis are presented in Fig. 38 where performance is shown in the form of the F-measure over varying lengths of the feature analysis window. The CnB algorithm performs the best, with F-measures of 0.71 for a window of 1 second, increasing up to a recognition rate of 0.96 for a window of 10 seconds. The implications are that for the given scenario and set of conditions, the emergent group behavior can be recognized using relatively straight forward methods, *if* observations of all members of the group are present. Admittedly, there are many other issues in GAR which are not present in this experiment, such as variance of group members and the number of group members over time, device location, etc. [8], however these problems are outside the scope of this work.

The CCI approach yields an F-score of 0.52 for an observation window length of 1 second, with an optimum of 0.70 for a window length of 3 seconds, after which it subsides towards random classifications with an F-score of 0.31 at 10 seconds. This would appear to indicate that posteriors over role-behavior clusters do not contain the pertinent information required to infer group behavior. However, as we will see later, the is not the case. The implication is therefore only that naive Bayesian inference is not the correct method for inference using these posteriors. This is due to the fact that Bayesian inference using GMMs separates the data probabilistically using EM for clustering, but the posteriors themselves do not separate well into such clusters.

The evaluation of the accuracy of the ILI method provides insight into the nature of the experiment. For a window size of 1 second, the mean F-measure of all nodes across all experiments was 0.48, with a variance of 0.05. For 10 seconds, the mean increases to 0.82 and variance drops slightly to 0.03. The longer the time-line of data used to classify the group activity, the better the group activity can be recognized, both for the centralized as distributed evidence functions. Also the quantified emergence of the group activity shrinks with the size of the window from 0.34 for 1 second to 0.16 for 10 seconds.

Sports activities in general are very dynamic in nature, where players change roles rapidly. For a longer observation time, a single player may change roles enough, allowing a classifier to observe the majority of role-behaviors from a single subject in that time, and therefore improve classification of the emergent behavior. This effect cannot be generalized to other forms of group activities such as social gatherings or meetings and is specific to the experiment conducted here. For the remaining evaluation of the novel distributed algorithms, a window size of 2

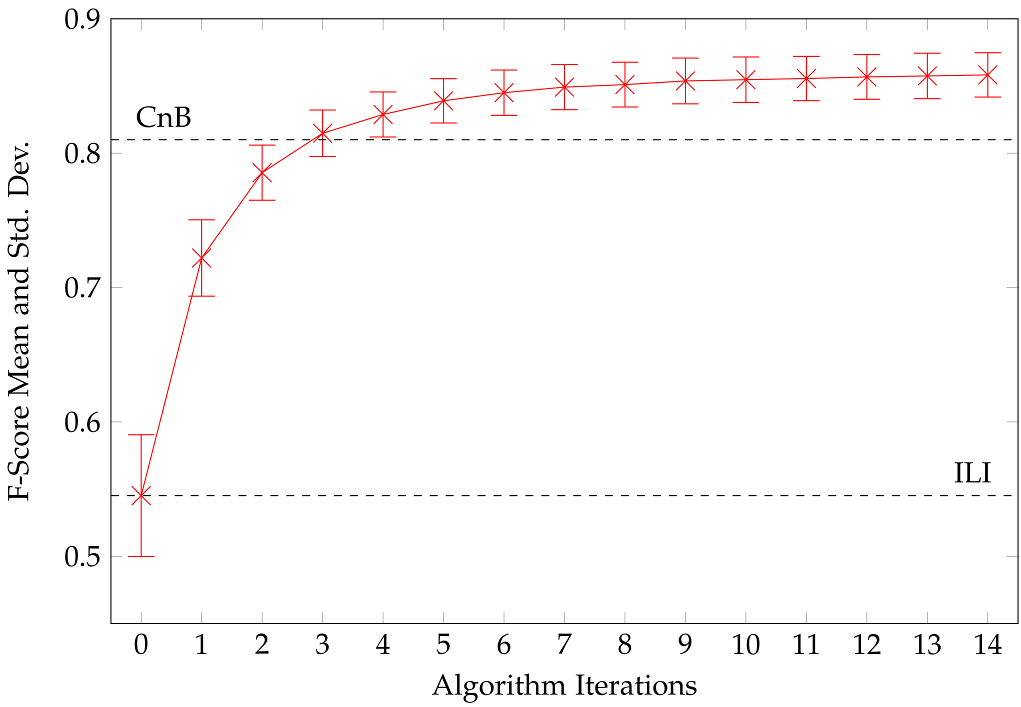


Figure 39.: DPI-SLBP for a window of 2 seconds and full inter-connectivity ($\phi = \infty$)

seconds has been selected, as it represents a a good level of emergence, and none of the algorithms in Fig. 38 have saturated or reached their peak results, allowing us to compare relative values.

9.6.2 *DPI with LBP*

The results of DPI with LBP for a window size of 2 seconds and a communication range of $\phi = \infty$ are displayed in Fig. 39. The shape of the curve presented demonstrates clearly that the distributed algorithm does indeed converge to a solution. This solution is reached after 15 iterations at an F-measure of 0.86. At iteration 0, the lower bound is given by the evaluation of the local evidence functions of each node separately, and corresponds to the value for a window size of 2s in Fig. 38. This value even exceeds the centralized approach at 0.81 after 3 iterations where 95% of convergence, a value of 0.84 is already reached after 6 iterations. It must be noted here that the indication is not distributed inference performs better, but that the potential performance using posteriors over K is higher than the performance of a nB classifier over X. The standard deviation across nodes is 0.045 for iteration 0, but drops to 0.027 already after one iteration and then converges to a value of 0.021.

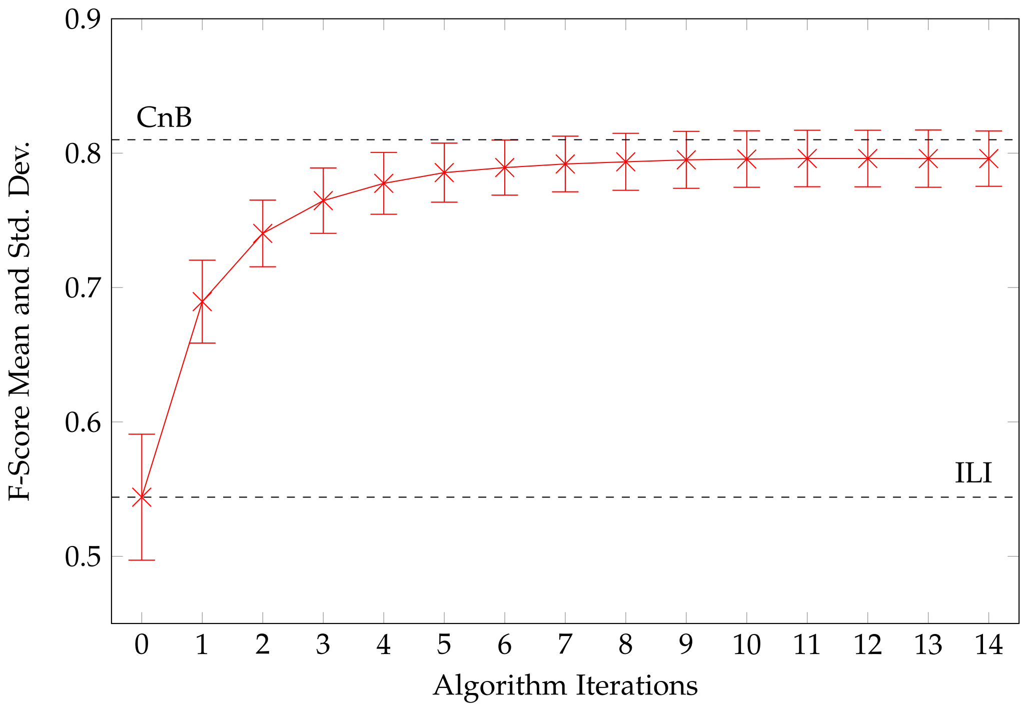


Figure 40.: DPI-HLBP for a window of 2 seconds and full inter-connectivity ($\phi = \infty$)

The results of DPI with LBP with the same parameters ($\phi = \infty$, ws = 2) but with the simplified potential function $\psi^{\text{simp.}}$ is shown in Fig. 40. Iteration 0 also begins at the same lower bound as in Fig. 39. A similar convergence is also clearly visible, but convergence occurs at 0.80, as compared to a value of 0.86 for the full potential method. The standard deviation also drops dramatically after one iteration from 0.045 to 0.037, and then iteratively converges to 0.031. This value is however greater than the standard deviation of 0.021 for the regression-based potential function. Here again, 95% of convergence is reached fairly quickly after 5 iterations.

The effects of the simplified potential function are clear. Convergence occurs slightly faster (1 iteration less for 95%), but converges to an optimum 7% less than when using a full regression-based potential function, and the standard deviation across nodes also increases by 68%. As we will see later, the reduced F-measure and increased standard deviation come with reductions in resource consumption, where the performance trade-off can be advantageous for certain applications.

9.6.3 *Effects of P2P Communication*

The two novel distributed methods where also simulated for various communication ranges. The range ϕ was simulated for 5m, 10m, 15m, 20m, and ∞, or full

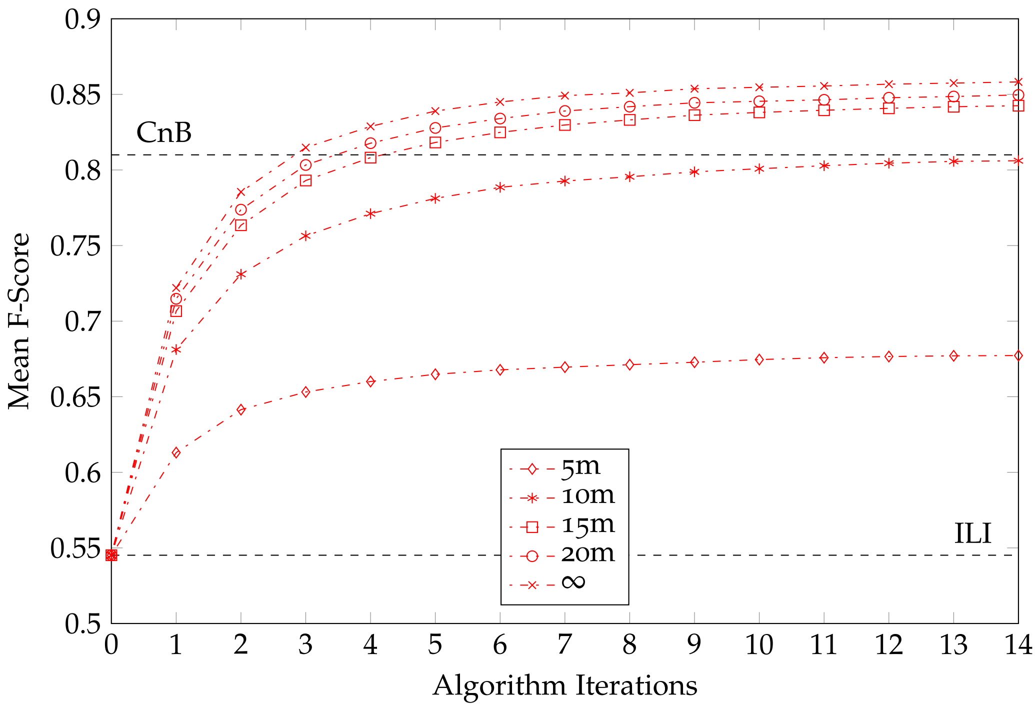


Figure 41.: Convergence curves for DPI-SLBP for varying ranges ϕ

connectivity. The mean F-measure results for regression-based potential function are displayed in Fig. 41. There the man value for $\phi = \infty$ corresponds to the same mean in Fig. 40. Mean values for 20m and 15m perform similarly to full connectivity, converging at an almost identical rate to a value of 0.85 and 0.84 respectively, compared to 0.86 for full connectivity. Reducing communication to 10m however incurred larger losses, converging to a value of 0.80, although with an identical rate of convergence as well. At 5m, convergence only achieved and F-measure of 0.68, although the rate of convergence remained constant.

Similar behavior was also observed for performance using the simplified potential function for the same simulated communication distances in Fig. 42. Communication ranges of 20m and 15m iteratively incur a loss of less than one F-measure point, although 95% of convergence requires one further iteration, namely 6 iterations. At 10m, convergence occurred at an F-measure of 0.76 with 95% reached after 8 iterations. Reducing communication further to 5m also required 8 iterations and converged to an F-measure of 0.65.

A survey of convergence values for both algorithms after 5 iterations can be seen in Tab. 14, where the coverage is simply the ratio of the of ϕ to the diameter of the group, assumed to be the diagonal of the field 25m. From full connectivity to 15m range there is little effect on the convergence times, although the using the

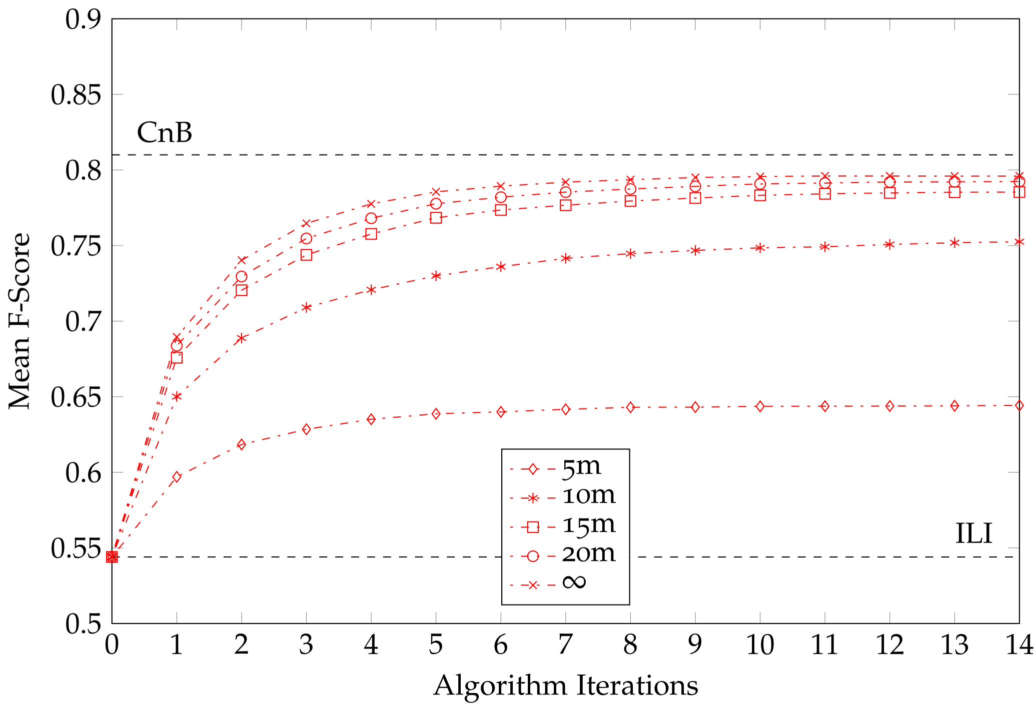


Figure 42.: Convergence curves for DPI-HLBP for varying ranges ϕ

simplified potential function incurred a greater reduction of 4.9 percentage points (pp) as apposed to 2.4 pp for regression-based potentials. This effect is due to the speed of belief propagation for the two algorithms. For $\psi^{\text{simpl.}}$ the propagation takes more "effort" as a node must receive enough belief contrary to its current state before its internal belief about its must probable cluster changes. For the regression-based approach, this occurs more quickly as beliefs are integrated and propagated in a continuous manner. For these communication ranges, the large majority of nodes are in the same network with occasional disconnection of individuals as they leave the group, e.g. to collect the ball. Hence, only the small changes in recognition rates over these ranges as belief propagates over intermediary nodes throughout the network.

For a communication distance of 10m, both algorithms propagate information at the same speed as before, but the network breaks apart into disjoint sub-networks as groups of nodes and individuals are out of range of each other. This is also the cause of the reduced recognition rates in Figs. 41 and 42 for a range of 10m, where necessary information cannot propagate to all nodes due to the lack of a link between nodes in different sub-networks. For a range of 5m the problem is exacerbated as the network breaks up into many different subgroups, and nodes only have one or two other nodes in their neighborhoods, many disjunct

Table 14.: Convergence in % after 5 iterations

Range	Coverage (%)	Convergence SLBP (%)	Convergence HLBP (%)
∞	100	91.2	94.6
20m	80	89.6	91.5
15m	60	88.8	89.7
10m	40	86.5	86.5
5m	20	91.1	91.7

neighborhoods appear. The results can be seen clearly in the low convergence rates in Figs. 41 and 42. However, convergence occurs quickly, as beliefs are only propagated to small subgroups of G.

9.6.4 *Resource Consumption Analysis*

The resource consumption is only for recognition, where training would incur higher costs and is more efficient when conducted in a centralized manner. The values presented here are simplified approximations, calculated from the bitrate and power consumption of different communication technologies [1][1]. The model assumes an Android Nexus 4 Device with processing on a single core with a consumption of 0.5W for that core. The results of the embedded resource consumption analysis for the different approaches are presented inn Tab. 15. For the CCI and CnB algorithms, we simulated communication of local information to a centralized instance using 3G networks. For the DPI algorithms, 10 iterations are assumed which is well over the amount required for 95% convergence presented in Tab. 14. Here DPI-SLBP reduced power consumption due to communication by 84% compared to CnB, and DPI-HLBP presents a reduction of 97.5%. In terms of time required for a classification, DPI-SLBP increases response time by a factor of 2.5, although server-side calculations for CnB are not taken into account [9]. DPI-HLBP however reduces the reaction time of the system by 51% with respect to CnB, which is around 5.5 times less then the reaction time of DPI-SLBP. It is important to note than the necessity to communicate with a server or centralized instance is removed for DPI-LBP algorithms.

The memory required to perform CnB is only the amount of memory required to store 1 window of sensory data. For DPI-SLBP, around 30 times more storage is required or almost 100 kB. DPI-HLBP only requires around 5 times more memory than CnB, representing a reduction of over 83% compared to DPI-SLBP due to

1 http://www.csr.com/sites/default/files/white-papers/comparisons_between_low_power_wireless_technologies.pdf

Table 15.: Resource consumption analysis for all algorithms for 1 classification after 10 iterations

Approach	Communication Technology	Memory Used (kB)	Comm. Per Classification (B)	Comm. Time (ms)	Comm. Energy (mJ)	Processing Time (ms)	Processing Energy (mJ)	Total Time (ms)	Total Energy (mJ)
CnB	(3G)	3.6	3600	71.53	32.47	2.13	0	73.65	32.47
CCI	(3G)	14.48	51.5	1.02	0.46	7.16	0	8.18	0.47
ILI	n/a	14.48	0	0	0	2.56	0	2.56	0
DPI-SLBP	(BT)	99.35	5150	119.5	4.57	76.31	0.38	195.81	4.96
DPI-HLBP	(BT)	16.6	800	18.56	0.71	17.7	0.09	36.26	0.8

the reduced size of the expectation look-up table compared to linear regression mappings.

9.7 DISCUSSION

The large reductions in resource consumption and low convergence time make DPI-HLBP an attractive approach. However for many applications there are some drawbacks. The effect of reducing simulated communication range was more pronounced than with DPI-SLBP. For both algorithms, conversion time increases as the group grows proportional to the communication range (see Tab. 14), but it grows slower for DPI-SLBP then for DPI-HLBP. For applications where the surface area of the group is large proportional to the communication range of the group, e.g. groups or crowds in public areas, propagation rates for DPI-SLBP could be greatly affected. For such applications the indications are that DPI-SLBT is the best approach to take, although performance and scalability to large groups was not evaluated here. However through the use of LBP, each node is only dependent on neighboring nodes, meaning the approach is very scalable, where the limiting factor is the time required for information to propagate over the group. For small groups such as the one analyzed here, this time is negligible. However if the required response time of the system drops below the processing time required, the number of iterations possible becomes limited and may not suffice for convergence.

For both algorithms however, it is important that the communication range be proportional to the surface area of the group such that the vast majority of group members are connected to at least one other member by one link, and to all members by at least one multi-hop path so that belief may propagate. In the case of sport activities, this requirement is fulfilled by a range of around 12.5m-15m, or 50% of the surface area of the group. For each iteration, the computational complexity at each node scales with the number of neighboring nodes and the number internal behavioral modalities. For DPI-SLBP, complexity scales with the number of neighboring nodes and the number of role-behavior clusters of the local node and each neighbor, or $\mathcal{O}(NK^2)$, where N is the size of the 1-hop neighborhood, and K^2 is due to the $K_i \times K_j$ multivariate regression. For DPI-HLBP, this is reduced to $\mathcal{O}(NK)$ since only the the expectation given the most like $k \in K_i$ is computed. Although complexity scales with the number of neighbors only (group density, and not group size), the number of iterations required is still a function of group size and the communication range.

For the presented experiment, it is conceivable that distributed majority voting techniques could achieve high GAR rates. However, in general for GAR, this will not be applicable. For problems with a higher degree of divergence ϵ, majority voting will inevitably degrade into noise by the definition of majority voting and

ϵ. For this reason an analysis and comparison of such methods has been omitted here.

In the field of group activity recognition there are other aspects which are not addressed here [8]. Group members can come and go over time, leading to changing group sizes and changes in individual and group behavior characteristics. These aspects are outside the scope of this work and must still be researched, for GAR in general and for GAR using DPI-LBP. Integration of explicit roles into the approaches presented here, along with generalized models for each role and automatic role detection is a path for future research which we will follow, and which could potentially further address these open issues in the field of GAR.

9.8 conclusion

Group activities are emergent from the individual characteristics of group members, their roles in the group, and the group dynamic [7]. The group behavior therefore has properties which are different from the properties of the behavior of the individuals, as well as the "sum" of those individual properties [15]. Recognition of these activities is the process of inferring the properties of the whole, based on the properties of the individual behaviors.

We have shown that the emergent behavior of the group can be inferred using centralized inference methods where the distributed observations of all members are present with F-scores upwards of 95% possible. We use clustering to address the problem of inference without explicitly requiring role. We presented two methods of inferring emergent behavior in a distributed fashion, based local estimations (distributed probabilistic inference DPI) and exchange of belief estimates (loopy belief propagation LPB). The first (DPI-SLBP) propagates beliefs based on linear potentials over posteriors from subject to subject. The second (DPI-HLBP) propagates beliefs as expectation based on the most likely behavior of an individual.

DPI-SLBP and DPI-HLBP converged to relatively high rates of recognition, with F-scores of 0.84 and 0.80 respectively compared to a centralized inference of 0.81 for the same parameters. The comparatively high recognition rates for the DPI approaches also demonstrates fulfillment of the requirement to be able to accurately model behavior with a *mapping function* from Chap. 4. Reducing the the communication range to 50% of the diameter of the group only marginally affected the value which the distributed algorithms converged to, as long as the range did not create disjunct networks out of the single group. However it did affect convergence time, where the effect on DPI-HLBP was greater, increasing the number of iterations needed. Further reductions of the communication range with respect to the group area incurred loss, but the reduced rates scale with the number of nodes lost, fulfilling the requirement for *surviving node failures*.

For larger groups such as crowds where local communication range is small in proportion to the surface area of the group, DPI-SLBP is then preferable. However, DPI-HLBP greatly reduces local resource consumption compared to DPI-SLBP, making it attractive for small group applications. In total, the distributed approaches allow inference of emergent group behavior using only local observations and classification from the mobile devices themselves, without the need for a centralized instance or infrastructure. They also reduce local energy consumption of the nodes themselves by a factor of 7 to 40, although for both algorithms the memory required locally increases, although still remaining under 100 kB. Response time also increases slightly for DPI-SLBP, although DPI-HLBP reduces response time against a cloud or server based centralized system. The relatively low energy, memory and computational footprint of the approaches when compared to the resources of modern smart phones [1] fulfills the requirement for a from Chap.4.

In Chap. 1, 4 challenges where presented. By fulfilling the technical requirements for P2P GAR in Chap. 4, the DPI-LBP also addressed the challenge of P2P GAR. By reducing the resource consumption footprint, this work also addressed the challenge of respecting the primary function of the devices. One requirement for P2P GAR is however not addressed here, namely the ability to recover from node failures by incorporating new group members and devices into the recognition approach. This is left for future work, and a road-map for future research in this direction is presented in Chap. 10.

Bibliography

[1] Niranjan Balasubramanian, Aruna Balasubramanian, and Arun Venkataramani. Energy consumption in mobile phones: a measurement study and implications for network applications. In *Proceedings of the 9th ACM SIGCOMM conference on Internet measurement conference*, IMC '09, pages 280–293, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-771-4.

[2] Ling Bao and Stephen S. Intille. Activity recognition from user-annotated acceleration data. In *Pervasive*, pages 1–17, 2004.

[3] Kenneth D. Benne and Paul Sheats. Functional roles of group members. *Journal of Social Issues*, 4(2):41–49, 1948. ISSN 1540-4560.

[4] Ming-Ching Chang, Nils Krahnstoever, Sernam Lim, and Ting Yu. Group level activity recognition in crowded environments across multiple cameras. In *Advanced Video and Signal Based Surveillance, IEEE Conference on*, Los Alamitos, CA, USA, 2010. IEEE Computer Society. ISBN 978-0-7695-4264-5.

[5] Joris Deguet, Laurent Magnin, and Yves Demazeau. Elements about the emergence issue: A survey of emergence definitions. *ComPlexUs*, 3:24–31, 2006.

[6] Wen Dong, Bruno Lepri, Alessandro Cappelletti, Alex Sandy Pentland, Fabio Pianesi, and Massimo Zancanaro. Using the influence model to recognize functional roles in meetings. In *Proceedings of the 9th international conference on Multimodal interfaces*, ICMI '07, pages 271–278, New York, NY, USA, 2007. ACM. ISBN 978-1-59593-817-6.

[7] D.R. Forsyth. *Group Dynamics*. International student edition. Thomson/Wadsworth, 4th edition, 2006. ISBN 9780534368227.

[8] Dawud Gordon, Markus Scholz, Yong Ding, and Michael Beigl. Global peer-to-peer classification in mobile ad-hoc networks: a requirements analysis. In *Proceedings of the 7th conference on Modeling and using context (CONTEXT)*, pages 108–114. Springer, 2011. ISBN 978-3-642-24278-6.

[9] Dawud Gordon, Jan-Hendrik Hanne, Martin Berchtold, Ali Asghar Nazari Shirehjini, and Michael Beigl. Towards collaborative group activity recognition using mobile devices. *Mobile Networks and Applications*, pages 1–15, 2012. ISSN 1383-469X.

[10] Dawud Gordon, Markus Scholz, and Michael Beigl. Group activity recognition using belief propagation for wearable devices. In *Wearable Computers (ISWC), International Symposium on*, pages 1–10, 2014. (*To Appear).

[11] Daniel Grünbaum, Steven Viscido, and Julia Parrish. Extracting interactive control algorithms from group dynamics of schooling fish. In Vijay Kumar, Naomi Leonard, and A. Morse, editors, *Cooperative Control*, volume 309 of *Lecture Notes in Control and Information Sciences*, pages 447–450. Springer Berlin / Heidelberg, 2005. ISBN 978-3-540-22861-5.

[12] Tao Gu, Zhanqing Wu, Liang Wang, Xianping Tao, and Jian Lu. Mining emerging patterns for recognizing activities of multiple users in pervasive computing. In *The Sixth Annual International Conference on Mobile and Ubiquitous Systems*, pages 1 –10, july 2009.

[13] Jane Yung-Jen Hsu, Chia-Chun Lian, and Wan-Rong Jih. Probabilistic models for concurrent chatting activity recognition. *ACM Trans. Intell. Syst. Technol.*, 2: 4:1–4:20, January 2011. ISSN 2157-6904.

[14] Andreas Huth and Christian Wissel. The simulation of fish schools in comparison with experimental data. *Ecological Modelling*, 75-76(0):135 – 146, 1994. ISSN 0304-3800. State-of-the-Art in Ecological Modelling proceedings of ISEM's 8th International Conference.

[15] K. Lewin. *Field theory in social science: selected theoretical papers*. Social science paperbacks. Harper, New York, 1951.

[16] Ruonan Li, R. Chellappa, and S.K. Zhou. Learning multi-modal densities on discriminative temporal interaction manifold for group activity recognition. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 2450 –2457, june 2009.

[17] Ryan Lukeman, Yue-Xian Li, and Leah Edelstein-Keshet. Inferring individual rules from collective behavior. *Proceedings of the National Academy of Sciences*, 107(28):12576–12580, 2010.

[18] Mehdi Moussaïd, Niriaska Perozo, Simon Garnier, Dirk Helbing, and Guy Theraulaz. The Walking Behaviour of Pedestrian Social Groups and Its Impact on Crowd Dynamics. *PLoS ONE*, 5(4):7, 2010.

[19] Kevin P Murphy. *Machine learning: a probabilistic perspective*. MIT Press, Cambridge, MA, 2012.

[20] Mark Paskin, Carlos Guestrin, and Jim McFadden. A robust architecture for distributed inference in sensor networks. In *Proceedings of the 4th international*

symposium on Information processing in sensor networks, IPSN '05, Piscataway, NJ, USA, 2005. IEEE Press. ISBN 0-7803-9202-7.

[21] Mark A. Paskin and Carlos E. Guestrin. Robust probabilistic inference in distributed systems. In *Proceedings of the 20th conference on Uncertainty in artificial intelligence*, UAI '04, pages 436–445, Arlington, Virginia, 2004. AUAI Press. ISBN 0-9749039-0-6.

[22] Daniel Roggen, Martin Wirz, Gerhard Tröster, and Dirk Helbing. Recognition of Crowd Behavior from Mobile Sensors with Pattern Analysis and Graph Clustering Methods. *Networks*, 00(0):1–24, 2011.

[23] Liang Wang, Tao Gu, Xianping Tao, Hanhua Chen, and Jian Lu. Recognizing multi-user activities using wearable sensors in a smart home. *Pervasive Mob. Comput.*, 7(3):287–298, June 2011. ISSN 1574-1192.

[24] Martin Wirz, Tobias Franke, Eve Mitleton-Kelly, Daniel Roggen, Paul Lukowicz, and Gerhard Tröster. Coenosense: A framework for real-time detection and visualization of collective behaviors in human crowds by tracking mobile devices. In *Proceedings of European Conference on Complex Systems*. Springer, 2012.

[25] Georg Wittenburg, Norman Dziengel, Christian Wartenburger, and Jochen Schiller. A system for distributed event detection in wireless sensor networks. In *IPSN '10: Proceedings of the 9th ACM/IEEE International Conference on Information Processing in Sensor Networks*, New York, NY, 2010. ACM. ISBN 978-1-60558-988-6.

10 Conclusion and Outlook

Human activity recognition (HAR) is the process of recognizing the physical behavior of individuals by sensor-generated observations. Wearable sensing has great potential for HAR because wearable sensors are embedded in ubiquitous smart phones. Machine learning is used to interpret these signals to extract the activity being performed. However, wearable sensors can only sense a single individual with high fidelity. Recognizing the complex behavior of groups using this technology requires fusing observations of group members into group information. The algorithms and experience to practically achieve this is the contribution of this dissertation.

10.1 IDENTIFYING THE PROBLEM

In Chap. 1 an activity is defined as a human context with a physical motion characteristic. Activity recognition is the process through which a device can understand the activity of the person using that device. We spend most of our time in groups, where a group is simply a set of individuals connected by social interactions. By taking a look at social psychology it quickly becomes clear that the relationship between the behavior of the grou and of the individual group members is not straightforward. Group behavior is emergent from the characteristics of the individuals within the group and the group dynamic. Emergence signifies that the group behavior is fundamentally different than the behavior of the individuals within it, and cannot be observed by a piecemeal examination of those individuals. It is therefore necessary to understand group activities as well as activities of single users in order to fully comprehend the human situation.

One way to recognize emergent behavior is to aggregate distributed observations (sensor measurements) together to be able to identify the group behavior based on a sensory picture of the emergent phenomena. From an algorithmic standpoint, observing all the combined behavioral measurements from all members together allows recognition of the emergent group behavior. However there are situations in which aggregation in a single location is not possible, such as when infrastructure and connectivity is sporadic, expensive or missing completely (see Chap. 4). In these cases it can still be advantageous to be able to recognize the emergent behavior, even without global network connectivity and access to remote resources: peer-to-peer (P2P) approaches. The challenges which must be overcome in order to accomplish this were stated in Chap. 1:

- **Challenge 1** – the amount power (and other resources) consumed for recognizing the group activity must not have a significant negative affect the primary function of the device.
- **Challenge 2** – the trade-off between power consumption and recognition accuracy must be evaluated in order to find the optimal point for GAR.

- **Challenge 3** – individuals who are affiliated with each other must be recognized as such based on sensory information in a P2P fashion to establish group membership.

- **Challenge 4** – given a group, the emergent activity must be recognized in a distributed fashion based on distributed observations of the group members.

An exploration of related work in Chap. 2 reveals research in different fields with similar goals. HAR for a single person using distributed measurements of the same body is also an emergent problem. However approaches use centralized resources which aggregate distributed measurements to recognize that behavior. The field of social psychology has a very detailed understanding of the intricacies of group and individual behavior, but is not focused on practical solutions to recognize it. There are however researchers who focus on modeling emergent behavior. Approaches are either bottom-up simulation, where rules governing individuals are used to simulate generation of the emergent properties, or top-down exploration of the individual rules based on observations of group or swarm properties. There are no approaches found which look to estimate or approximate emergent group behavior based on observations of individuals.

There is however some research into recognition of group activities available. In Chap. 3 the definition of single-user activity recognition (SAR), multi-user activity recognition (MAR) and GAR is created and explored. I defined MAR as the recognition of multiple activities for multiple users, and differentiated it from GAR which recognized a single activity for an "organism" consisting of group of individuals. It is also shown that the distinction between GAR and MAR cannot be made based on the labels used as these are subjective interpretation of the underlying phenomena.

In Chap. 4, the hurdles for achieving Challenge 4 (P2P Group Activity Recognition) are explored to create formal requirements for distributed GAR. There 4 requirements emerged. First, the approach must be able to survive node failures (such as hardware/software failures or individuals leaving) without suffering catastrophic loss of recognition capabilities. Second, the algorithm must be able to approximate the mapping function, evaluated in the form of recognition rates for GAR. Third, the algorithm must preserve the primary function of the mobile devices with respect to resource consumption. The fourth requirement is that the approach be able to recover from node failures, meaning integrate new nodes or even individuals into the recognition approach at training. This requirement is not addressed in this dissertation and is left for future work. In order to fulfill this final requirement, significant investigation is required, a road-map for which is given in future work in Chap. 10.4. Although not a requirement, it was also stated that a recognition algorithm which does not take the individuality or role of group members into account will not be able to recognize the emergent behavior of the group.

10.2 ADDRESSING THE CHALLENGES

The first challenge I address is Challenge 1 (Low Power). In Chap 5 the use of a highly sensitive vibration sensor is introduced for activity recognition called the micro-vibration sensor (MVS). The sensor can be used for sensing and discriminating activities which have a high-frequency component. The optimal vibrational frequency range for the sensor is between 3 kHz and 8 kHz, a range where most mobile accelerometers can not sense. The sensor is useful for recognizing activities with impacts such as walking, jogging or riding a bike, while slower movements such as gestures can not be easily discriminated. While it cannot replace a more energy-hungry accelerometer for many activities, it can sense useful information which the accelerometer can't. The 50x reduction of power consumption creates new opportunities for low-power recognition such low-power activity listening and machine monitoring.

Challenge 1 is further addressed in Chap. 6 where a method for reducing power consumption for embedded SAR is introduced. The predictability of human subjects is used to turn off unnecessary sensors when they are not needed. The decision is made based on predicted future activities and the dependence of the recognition rate for each activity on each sensor. The method is evaluated using two data sets, on of which is uses the MVS sensor from the previous section. The result is that for a small loss in recognition accuracy of 1.5 pp - 2.8 pp 84 % to 89 % of energy consumption can be saved. The implications for GAR are 1) a low-power approach for acquiring single-user activity information for fusion, and 2) a method for performing sensor selection based on prediction for GAR as well.

The next challenge which is addressed is Challenge 2 (Data Abstraction Level) for performing GAR in Chap. 7. The abstraction levels from the previous two chapters, namely sensor data, features, and single-user activities are compared, as well as using unsupervised clustering of single-user sensor data. Each of these represents a different level of abstraction, either low abstraction for features, mid for clusters, or high for single-user activities. Each has a different computational effort, communication volume, and GAR accuracies associated with it, where the tradeoff is investigated. The different abstraction levels are tested on 3 subjects in 3 different group activities, representing a simple scenario designed to naturally generate emergent group behavior.

Sensor data or features from all individuals allows group activities to be recognized 0.96% of the time for the scenario, but are expensive at 0.91 mJ per classification due to the cost of transmitting data. For single-user activities, the cost drops significantly to 0.61 mJ per classification due to the low data volume (high abstraction level) but recognition rates are inconclusive. Only 0.63 percent of activities could be recognized. However, this was due to behavioral differences in individuals during SAR classifier training and group activities. This result highlights the issue of having to doubly label group and single-user activities in parallel to train GAR algorithms. The cluster abstraction level avoids this issue by allow unsupervised single-user clusters and supervised GAR classifiers to be trained in parallel, costing approximately the same as SAR at 0.61 mJ but achieving recognition rates almost equal to features at 0.93%. The potential of clusters is therefore shown for GAR and this abstraction level is used in the following chapters for both recognizing affiliations and activities in the group.

Chap. 8 addresses Challenge 3 (P2P Group Affiliation Detection). The problem is to detect similarity – of any kind – between different members of a group without exchanging the sensor data between subjects as would be the case for contemporary centralized time-series analysis approaches. The abstraction level of clustering from the previous chapter is selected, where clusters from different individuals are compared with each other over a given window. For this purpose only the parameters of the clusters must be transmitted, and the Jeffrey's divergence over these parameters is used as an indicator of social proximity. I call this method divergence-based affiliation detection (DBAD). DBAD can be conducted using probability density functions (DBAD-P) or histograms of sensor data (DBAD-H). Using the resulting divergence as an indicator of social proximity, a method for filtering this proximity into an indicator of affiliation is presented.

Both methods with filtering have the potential to reach a GAD rates of 93% which is comparable to time-series analysis methods, however the resource consumption differs. DBAD-H and DBAD-P reduce energy consumption by 24% and 43% respectively, where DBAD-H reduces response time by 7%, but DBAD-P doubles it, indicating there is a trade-off between energy consumption and response time. Both distributed methods increase the amount of memory used, although usage remains under 4.5 kB. The analysis shows that distributing contemporary time-series analysis approaches by communicating data is prohibitively expensive in terms of time and energy required. However, both DBAD methods are independent of centralized resources and can be applied in a practical way with respect to device resources in distributed, P2P systems.

The final challenge addressed in this dissertation is Challenge 4 (P2P Group Activity Recognition). In Chap. 9, I introduce a method for recognizing group activities using distributed probabilistic inference with loopy belief propagation (DPI-LBP). The clustering abstraction level is again used from Chap. 7, where DPI-LBP is used to infer group activities on top of individual behavioral clusters in a distributed fashion. Individual devices exchange probabilities with each other in order to converge together to a decision on the most probable emergent group behavior given all sensor measurements in the network.

Two methods for propagating beliefs through the network were presented and evaluated, one using hard classifications (HLBP) and one using soft probability distributions (SLBP). HLBP performed slightly worse than SLBP, converging to F-scores of 0.84 and 0.80 respectively compared to centralized inference of 0.81. These values remain relatively stable as long as group members are within communication distance. HLBP converges faster than SLBP, but the negative affects of multiple hops for belief are greater for HLBP, indicating that SLBP is perhaps the better choice for groups which are large with respect to the communication range of each device (crowds). SLBP and HLBP both reduce energy consumption greatly by a factor of 7 or 40 respectively, although memory consumption increases to up to 100 kB and 17 kB respectively. It is also demonstrated that this approach meets 3 of 4 requirements put forth in Chap. 4, while the ability to incorporate new individuals into existing groups is not explored.

10.3 SUMMARY AND APPLICATIONS

GAR in P2P mobile devices is a challenging problem due to the conceptual disparity between the emergent nature of group behavior, and the local scope of observations of mobile devices. At the beginning of this dissertation I identified the challenges which must be addressed in order to accomplish these tasks. Each challenge was investigated and approaches, algorithms and methods for addressing them where introduced and evaluated. The evaluations of the proposed approaches demonstrated positive results, indicating a set of solutions to the individual challenges. The combined contribution however is an understanding of GAR in P2P mobile devices, and a methodology for achieving that goal.

This technological advance promises improvement in current applications as well as opening new application areas. Detecting emergent group behavior can allow intelligent environments to improve their understanding of the groups working within them, without having to understand the individuals themselves. The mobile devices of the individuals collaboratively work to understand the behavior of the group, enabling proactive environmental support without incurring load on the network infrastructure. Furthermore, because the algorithms are P2Pin nature and work on an abstract but unlabeled data basis, they also reduce the amount of privacy which must be sacrificed for GAR. Similar approaches could be used to monitor social animals in the wild with small P2P devices for monitoring individuals, providing a time-line of emergent behavior. For example, the social behavior of livestock, migratory birds and endangered species with social aspects (wolves, primates, whales, etc.) could be monitored outside of the confines of instrumented areas.

For individuals looking to use machine learning approaches to better understand and document themselves, the contribution of this dissertation allows them to improve the view of the device. I have shown that the behavior of the group can be recognized collaboratively by devices, but observing the behavior of the individual alone does not provide the full picture. These methods would allow devices to log the emergent group behavior as well, putting the individual behavior in context, without sacrificing the privacy of each individual. This emergent picture of behavior can then be used quantify the life of a user for their own purposes, or to allow them to share the full picture with friends across social networks without sacrificing the privacy of other group constituents.

As small groups coalesce into larger crowds, these P2P approaches can be used to recognize subgroups within the crowd and infer their behavior without requiring or overloading infrastructure. The scalable design of the novel methods may even allow inference of emergent crowd behavior as a whole, thereby having applications for crowd management and possibly even emergency and catastrophe management systems.

10.4 further work

One requirement for P2P GAR which was not evaluated in Chap. 9 was the requirement from Chap. 4 to be able to incorporate new members into the group who either join or replace existing members over time. I believe however that the tools required to accomplish this have already been presented here, but not yet evaluated.

My avenue of approach would be to combine methods from DBAD and DPI-LBP. When a new member arrives, an affinity analysis could indicate similarities in individual behavior. I would like to investigate how by selecting the individual with the greatest similarity using DBAD, and then adopting their models to the new user would perform. I would also look at adapting the models from the most similar member based on the disparity given by DBAD could improve model fit. Alternatively, one could look at transfer learning approaches, where a new subject learns from existing members at classification time. The classified emergent group activity would then be used as a label to train the device of the new member to participate in future iterations. It would also be interesting to evaluate this approach with a combined DBAD-DPI method.

Another point of interest would be to explicitly include the role of users into the classification process. In this dissertation the cluster-based approach from Chap. 7 was used because it eliminates the need for specific subject activity, subject role, and group activity labels in parallel. For accurate role descriptions, behavior experts and social psychologists are required to correct labels. With this information however, DPI-LBP could be adapted to better infer group behavior given the roles of the constituents, or to infer the role given a known behavior. If the role of members is known and the group behavior is classified, it may even be possible to explore inference of more subtle psychological attributes of members, such as personality type or mood, moving further towards computational social psychology.

Finally, these approaches have been designed to be P2P and hopefully inherently scalable due to dependency only on neighboring nodes. The implication is that the computational complexity is not dependent on the size of the group, but on the number of neighbors. Here scalability was not explicitly evaluated, but for crowds this implication means that complexity is dependent on crowd *density* rather than size, although for DPI-LBP propagation of beliefs through the crowd could create issues. Experiments using large groups or crowds would provide insight into the applicability of these approaches for crowd emergencies, or indicate if further research is required.

A

About the Author

DAWUD GORDON

ABOUT THE AUTHOR

PERSONAL INFORMATION

Born in Istanbul, Turkey, 16 February 1982

Nationality United States of America

email gordon@teco.edu

website http://www.teco.kit.edu/people/gordon

GOAL

My current interests are towards empowering users with respect to their own personal data. My research has been focused on activity recognition, in particular embedded and distributed recognition using mobile devices with respect to emergent group activities. In the future my goal is to understand how human and group activity recognition can be applied in the field of computational social sciences to better understand human behavior and interaction in the wild.

WORK EXPERIENCE

TwoSense Labs

2013 – CEO and Co-Founder, TwoSense

Empowering users with respect to their personal data by creating an information marketplace.

Karlsruhe Institute of Technology

2010 – 2013 Researcher, Pervasive Computing Systems / TecO — Karlsruhe, Germany

Research focus: group activity recognition using mobile, P2P devices.
Under the advisement of Prof. Michael Beigl

ETH Zurich

Sep-Nov 2012 Visiting Researcher, Wearable Computing Lab — Zurich, Switzerland

Research focus: group affiliation detection using mobile, P2P Devices
Under the advisement of Prof. Gerhard Tröster

TU Braunschweig

2009 - 2010 Researcher, Distributed and Ubiquitous Systems — Brunswick, Germany

Research focus: embedded systems and wireless sensor networks for Ubicomp.
Under the advisement of Prof. Michael Beigl

TU Braunschweig

2008 - 2009 Student Researcher, Distributed and Ubiqitous Systems — Brunswick, Germany

Research focus: embedded systems and wireless sensor networks for Ubicomp.
Under the advisement of Prof. Michael Beigl

TU Braunschweig

2006 - 2007 Student Researcher, Institute of Control Engineering — Brunswick, Germany

Focus: CAN bus interface for automotive error diagnostics and emulation.
Under the advisement of Dr. Tobias Müller

Miltenyi Biotec

2005 - 2006 Internship, Miltenyi Biotec — Cologne, Germany

Focus: intranet application development in C#.NET

LeCroy Corporation

2002 - 2004 Internships, LeCroy Corp. — New York, U.S.A.

Focus: IT and quality assurance through invariance testing

EDUCATION

Ph.D. Candidate in Computer Engineering

2010-2013 KIT — Karlsruhe, Germany

Thesis: *Recognition of Group Behavior using Mobile Sensing Devices*, expected date of completion 17 Jan 2014
Description: This thesis explores the ability of mobile devices, each sensing only the behavior of a single person, to collaboratively estimate the emergent behavior of a group.
Adviser: Prof. Michael Beigl

Study Abroad: Mexico

2007 - 2008 Universidad de Guadalajara — Jalisco, Mexico

Description: This exchange focused on machine learning and Spanish language skills.

M.Sc. in Informatics

2006-2009 TU Braunschweig — Brunswick, Germany

Thesis: *Akiba-Node: System Design, Development, Implementation and Evaluation of Components for a Wireless Sensor System Considering Particular Applications*
Description: This thesis explored the design space for a low-power, low-cost sensor node for rapid prototyping and retail applications.
Adviser: Prof. Michael Beigl

Study Abroad: Germany

2004 - 2005 Universität Würzburg — Würzburg, Germany

Description: This exchange focused on general education requirements and German language skills.

B.Sc. in Computer Science Combined

2000 - 2005 State University of New York at Albany

Description: This degree focused on a general computer science background with a specialization in physics.

OTHER INFORMATION

Awards and Citations

2013 · Best Paper Award, MobiCASE'13, Reconciling Cloud and Mobile Computing using Activity-Based Predictive Caching

2012 · Honorable Mention and Best Paper Nominee: ISWC'12, Energy-Efficient Activity Recognition using Prediction, Acceptance Rate 18%

2011 · Best Demo Nominee: Pervasive'11, Program Your Reality with dinam-mite

2010 · Best Paper Nominee: ISWC'10, A Novel Micro-Vibration Sensor for Activity Recognition: Potential and Limitations, Acceptance Rate 21%

2010 · Best Paper Nominee: ISWC'10, ActiServ: Activity Recognition Service for Mobile Phones, Acceptance Rate 21%

2004 · Deans List, SUNY Albany,

Languages

English · Mother-tongue

German · Fluent (spoken and written)

Spanish · Fluent (spoken and written)

Interests

Generative art · Billiards · Travel · Languages · Rugby Union

February 17, 2014

A.2 LIST OF PUBLICATIONS

This work is based on and composed of the content of several peer-reviewed journal articles, conference proceedings and workshop contributions which I have authored along with my colleagues and students. A proposal abstract of this dissertation was presented at the Pervasive doctoral consortium in 2012 [10]. I contributed the majority of the content and work to the following publications.

A.3 PUBLICATIONS USED IN THIS DISSERTATION

Using a novel micro-vibration sensor for activity recognition and a method for processing the sensor output were published at ISWC 2010 [6] which was nominated for the Best Paper Award. This was preceded by an exploratory study on the use of this sensor at the workshop PervaSense 2010 [15]. A general methodology for introducing and evaluating novel sensors in terms of their usefulness for activity recognition was published at the Workshop on How To Do Good Research in Activity Recognition [14] which was held in conjunction Pervasive 2010.

The concept of predicting human behavior to allow for better sensor selection was first published in the Work in Progress track at PERCOM 2011 [7]. This contribution was later expended on at ISWC 2012 [9] where it received an Honorable Mention Award and a nomination for the Best Paper Award. An expanded evaluation was invited to the Journal of Personal and Ubiquitous Computing (PUC) in 2013 [21].

The requirements analysis for performing GAR using only distributed wearable sensing devices was published in the proceedings CONTEXT 2011 [19]. The experiment concept for assessing the correct abstraction level for GAR was first proposed in a short paper and poster contribution in the proceedings of CONTEXT 2011 as well [16]. The results of this experiment where published at MobiQuitous 2011 [17]. An extended version of these proceedings was published in the Journal of Mobile Networks and Applications (MONET) as an invited submission in 2012 [20]. The work on group affiliation detection [24], and the work on distributed probabilistic inference for GAR [23] have both been accepted in parallel to ISWC 2014. The dependencies of the individual chapters of this dissertation on specific publications is listed in the introduction section of each chapter.

A.4 CONTRIBUTIONS TO SUPPORTING PUBLICATIONS

I have also made smaller contributions to works which are not directly used in this dissertation but which have a supporting role. Sigg et al. [27] researched the effects of different extraction layers on context prediction, a work which inspired similar questions for the task of GAR researched here Chap. 7. Berchtold et al. [2]

researched the use of an activity recognition service for offloading of processor load for both training and activity recognition phases, as well as for cross-individual optimization and crowd-sourcing. I was also involved with other researchers in developing the JNode wireless sensor network (WSN) platform for conducting activity recognition [25].

A.5 publications and contributions not used in this dissertation

In the past I have researched various other aspects of wireless sensor networks. Focus here was on the development of novel WSN technology to improve usability and expand the application and user space [11, 5, 28, 25, 26]. Here I developed a new method to approach WSN application-building problems by embedding the IDE and all required resources into the nodes themselves [12, 13, 4]. A demonstration of this technology was nominated for Best Demo at Pervasive 2011 [18]. I also contributed to work on collaborative transmission and reception in wireless sensor networks [1].

One contribution in this area is WoR-MAC, a media access control (MAC) protocol for low-power wireless communication in sensor networks [8]. This protocol utilizes low-power listening, either native or with dedicated hardware, to reduce consumption for intermittent wireless communication. The reduction is achieved by allowing nodes to duty-cycle transceivers, thereby reducing overall consumption of communication. WoR-MAC was designed with ad-hoc, p2p recognition of emergent group behavior as a target application and would theoretically reduce the energy footprint caused by increased communication [8]. However, I view this work to be slightly outside the scope of this dissertation and have opted not to include it for brevity. This work inspired the research on the MVS sensor in Chap. 5, as a form of low-power listening for activities where the channel is human behavior.

Within the field of activity recognition I have also looked at uses for various purposes. Recently, I worked on using activity recognition to create caching recommendations for mobile applications by predicting disconnection events [22] which won the Best Paper award at MobiCASE'13. I also contributed to work investigating activity recognition algorithms as a tool for transferring skills between individuals [3].

Bibliography

[1] Behnam Banitalebi, Dawud Gordon, Stephan Sigg, Takashi Miyaki, and Michael Beigl. Collaborative channel equalization: Analysis and performance evaluation of distributed aggregation algorithms in wsns. In *Mobile Adhoc and Sensor Systems (MASS), 2011 IEEE 8th International Conference on*, pages 450–459. IEEE, 2011.

[2] Martin Berchtold, Matthias Budde, Dawud Gordon, Hedda Schmidtke, and Michael Beigl. ActiServ: Activity recognition service for mobile phones. In *ISWC'10: Proceedings of the Fourteenth International Symposium on Wearable Computers*, pages 83–90, Seoul, S. Korea, 2010. IEEE Computer Society.

[3] Yong Ding, Matthias Budde, Dawud Gordon, Dimana Shishkova, Nadezda Sackmann, Hedda R. Schmidtke, and Michael Beigl. Ubiquitous know-how transfer based on a mobile learning and classification system. In *Adjunct Proceedings of the Seventh International and Interdisciplinary Conference on Modeling and Using Context (CONTEXT'11)*, 2011.

[4] Yong Ding, M.A. Neumann, D. Gordon, T. Riedel, T. Miyaki, M. Beigl, Wenzhu Zhang, and Lin Zhang. A platform-as-a-service for in-situ development of wireless sensor network applications. In *Networked Sensing Systems (INSS), 2012 Ninth International Conference on*, pages 1–8, 2012.

[5] D. Gordon and M. Beigl. D-bridge: A platform for developing low-cost wsn product solutions. In *Networked Sensing Systems (INSS), 2009 Sixth International Conference on*, pages 1–4, 2009.

[6] D. Gordon, H.R. Schmidtke, M. Beigl, and G. von Zengen. A novel micro-vibration sensor for activity recognition: Potential and limitations. In *Wearable Computers (ISWC), 2010 International Symposium on*, pages 1 –8, oct. 2010.

[7] D. Gordon, S. Sigg, Y. Ding, and M. Beigl. Using prediction to conserve energy in recognition on mobile devices. In *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2011 IEEE International Conference on*, pages 364 –367, march 2011.

[8] D. Gordon, M. Berning, R.E. Masri, M. Beigl, J. Blanckenstein, and J. Klaue. Wor-mac: Combining wake-on-radio with quality-of-service for intelligent environments. In *Networked Sensing Systems (INSS), 2012 Ninth International Conference on*, pages 1–8, 2012.

[9] D. Gordon, J. Czerny, T. Miyaki, and M. Beigl. Energy-efficient activity recognition using prediction. In *Wearable Computers (ISWC), 2012 16th International Symposium on*, pages 29 –36, june 2012.

[10] Dawud Gordon. Group activity recognition using mobile devices. In *Doctoral Consortium at the Tenth International Conference on Pervasive Computing*, 2012.

[11] Dawud Gordon, Michael Beigl, and Masayuki Iwai. A study on the use of wireless sensor networks in a retail store. In *Workshop in Conjunction with Pervasive*, 2009.

[12] Dawud Gordon, Michael Beigl, and Martin Alexander Neumann. dinam: A wireless sensor network concept and platform for rapid development. In *Proceedings of the Seventh International Conference on Networked Sensing Systems (INSS10)*, Kassel, Germany, June 2010.

[13] Dawud Gordon, Matthias Berning, Hedda R. Schmidtke, and Michael Beigl. Rapid prototyping for urban sensing. In *UBI Challenge Workshop in conjunction with Ubicomp*, 2010.

[14] Dawud Gordon, Hedda Schmidtke, and Michael Beigl. Introducing new sensors for activity recognition. In *How To Do Good Research In Activity Recognition: Experimental methodology, performance evaluation and reproducibility. Workshop in conjunction with Pervasive 2010*, Helsinki, Finland, May 2010.

[15] Dawud Gordon, Florian Witt, Hedda Schmidtke, and Michael Beigl. A long-term sensory logging device for subject monitoring. In *Workshop for Situation Recognition and Medical Data Analysis in Pervasive Health Environments (PervaSense 2010)*, Munich, Germany, March 2010.

[16] Dawud Gordon, Jan-Hendrik Hanne, Martin Berchtold, Takashi Miyaki, and Michael Beigl. An experiment in hierarchical recognition of group activities using wearable sensors. In *Proceedings of the 7th conference on Modeling and using context (CONTEXT)*, pages 108–114. Springer, 2011.

[17] Dawud Gordon, Jan-Hendrik Hanne, Martin Berchtold, Takashi Miyaki, and Michael Beigl. Recognizing group activities using wearable sensors. In *Mobile and Ubiquitous Systems: Networking Services, MobiQuitous, 2011. MobiQuitous '11. 8th Annual International*, dec 2011.

[18] Dawud Gordon, Martin Alexander Neumann, and Michael Beigl. Program your reality with dinam-mite. In *Demonstration at the Ninth International Conference on Pervasive Computing*, 2011.

[19] Dawud Gordon, Markus Scholz, Yong Ding, and Michael Beigl. Global peer-to-peer classification in mobile ad-hoc networks: a requirements analysis. In *Proceedings of the 7th conference on Modeling and using context (CONTEXT)*, pages 108–114. Springer, 2011. ISBN 978-3-642-24278-6.

[20] Dawud Gordon, Jan-Hendrik Hanne, Martin Berchtold, Ali Asghar Nazari Shirehjini, and Michael Beigl. Towards collaborative group activity recognition using mobile devices. *Mobile Networks and Applications*, pages 1–15, 2012. ISSN 1383-469X.

[21] Dawud Gordon, Jürgen Czerny, and Michael Beigl. Activity recognition for creatures of habit. *Personal and Ubiquitous Computing*, pages 1–17, 2013. ISSN 1617-4909.

[22] Dawud Gordon, Sven Frauen, and Michael Beigl. Reconciling cloud and mobile computing using activity-based predictive caching. In *Fifth International Conference on Mobile Computing, Applications and Services (MobiCASE)*. Springer Verlag, 2013.

[23] Dawud Gordon, Markus Scholz, and Michael Beigl. Group activity recognition using belief propagation for wearable devices. In *Wearable Computers (ISWC), International Symposium on*, pages 1–10, 2014. (*To Appear).

[24] Dawud Gordon, Martin Wirz, Daniel Roggen, Michael Beigl, and Gerhard Tröster. Group affiliation detection using model divergence for wearable devices. In *Wearable Computers (ISWC), International Symposium on*, pages 1–10, 2014. (*To Appear).

[25] Philipp M. Scholl, Kristof Van Laerhoven, Dawud Gordon, Markus Scholz, and Matthias Berning. jnode: A sensor network platform that supports distributed inertial kinematic monitoring. In *Ninth International Conference on Networked Sensing Systems (INSS)*, june 2012.

[26] Markus Scholz, Dawud Gordon, Leonardo Ramirez, Stephan Sigg, Tobias Dyrks, and Michael Beigl. A concept for support of firefighter frontline communication. *Future Internet*, 5(2):113–127, 2013. ISSN 1999-5903.

[27] S. Sigg, D. Gordon, G. Zengen, M. Beigl, S. Haseloff, and K. David. Investigation of context prediction accuracy for different context abstraction levels. *Mobile Computing, IEEE Transactions on*, PP(99):1, 2012. ISSN 1536-1233.

[28] Wenzhu Zhang, Lin Zhang, Yong Ding, Takashi Miyaki, Dawud Gordon, and Michael Beigl. Mobile sensing in metropolitan area: Case study in beijing. In *Workshop on Mobile Sensing: Challenges, Opportunities and Future Directions at Ubicomp'11*, 2011.

List of Symbols and Abbreviations

A

A	Ampere, also a mapping from features onto the sensors used to generate them.
Acc.	Acceleration, also accuracy.
A/D	Analog to digital converter.
ADXL	ADXL 335 3D accelerometer from Analog Devices.
AF	Affinity.
AR	Activity Recognition.

B

B	Behavior of an individual or group.
b	A weighted mapping from classes onto features.
Be.	Belief function.

C

C	The set of all activity, context or behavioral classes $c \in C$, also personal characteristics $c \in C$.
C4.5	C4.5 decision tree datamining algorithm.
CCI	Centralized cluster-based inference.
Cent.	Centralized.
Chal.	Challenge.
Chap.	Chapter.
CL	"Classified as".
Clust.	Clustering.

CnB Centralized naïve Bayes.

Comm. Communication.

Comp. Compressed.

Corr. Correlation.

COTS Commercial, off of the shelf.

D

D A set of data used for a experiment.

D Divergence.

DBAD Divergence-based affiliation detection.

DBAD-H Divergence-based affiliation detection using histograms.

DBAD-P Divergence-based affiliation detection using probability density functions.

dist A distance metric, usually Euclidean.

Distr A distribution of any kind.

D$_J$ Jeffrey's divergence.

DPI Distributed probabilistic inference.

DPI-HLBP Distributed probabilistic inference with hard loopy belief propagation.

DPI-SLBP Distributed probabilistic inference with soft loopy belief propagation.

DT Decision tree.

E

E Social environment or "field", also: emissions of a generative probabilistic model $e \in E$, and energy.

ECG Electrocardiography.

EM Expectation maximization.

En Energy.

F

F A set of sensor signal features $f \in F$.

FFT Fast Fourier transformation.

G

G A group of individuals.

GAD Group affiliation detection.

GAR Group activity recognition.

GMM Gaussian mixture model.

GND Ground.

GPS Global positioning system.

GT Ground truth.

Gyro Gyroscope.

H

HAR Human activity recognition.

HMM Hidden Markov model, a datamining algorithm.

HVAC Heating, ventilation, and air conditioning.

Hz Herz.

I

I Current.

IBK A k-nearest neighbors datamining algorithm implementation.

IC Integrated circuit.

i.i.d. Independently and identically distributed.

ILI Independen local inference.

Ind. Individual.

J

J Joule.

J48 A java implementation of the C4.5 algorithm.

K

K Set of all clusters $\kappa in K$.

kNN k-nearest neighbors machine learning algorithm.

L

LBP Loopy belief propagation.

Like. Likelihood.

M

M Memory consumption.

MANET Mobile, ad-hoc network.

MAR Multi-user activity recognition.

MVS Micro-vibration sensor.

N

N The number of users, subjects or devices.

NAF Non-affinity.

nB Naïve Bayes datamining algorithm.

O

OPP The OPPORTUNITY project and data set.

Or. Orientation.

OS Operating system.

P

P	Personal characteristics of an individual, also power consumption.
p	Probability, also processing load.
P2P	Peer-to-peer.
PDA	Personal digital assistant.
PDF	Probability density function.
PIC	PIC microcontroller from Microchip.
Post.	Posterior probability.
pp	Percentage points.
Pr.	Prior distribution.
Prec.	Precision.
Proc.	Processor, processing.

Q

Q	A set of weights for the mapping b which indicate the importance of features for recognizing classes $q \in Q$.

R

R	Accuracy or recognition rate, also resistance and the set of all individual roles in the group $\rho \in R$.
Rec.	Recall.
Req.	Requirement.

S

S	Set of all subjects, also used for sensors.
s	One single subject, also used for one single sensor.
SAR	Single-user activity recognition.
Sec.	Section.

T

t Current time or timestamp.

T Measurement of time.

Tab. Table.

TSA Time series analysis.

U

USB Universal serial bus.

USD U.S. dollars.

V

V Voltage.

vonMises A circular von Mises distribution.

W

W Watt.

WEKA An open-source datamining toolkit.

WLAN Wireless local area network.

WSN Wireless sensor network.

X

X Set of observations.

x One single observation.

SYMBOLS

β Linear coefficient.

Δ Difference.

$\mathcal{D}$	Divergence.
ϵ	Degree of emergence of a recognition problem.
$\mathbb{E}$	Expectation.
κ	The predictability of a given scenario, also, a behavior cluster.
λ	A parameter which specifies the acceptable amount of recognition loss for an application.
μ	Mean.
$\mathcal{M}$	A disparity matrix.
$\mathcal{N}$	Gaussian distribution.
$\mathcal{O}$	Computational complexity.
π	Mixing coefficient.
ψ	Potential function.
ϕ	Decision threshold for affinity, also communication range of mobile devices.
ρ	The number of classes predicted at a specific point in time, also the role of the individual in a group dynamic $\rho \in R$.
σ	Variance.
Σ	Covariance matrix.
θ	An angle.
$\mathcal{V}$	Communication neighborhood.
$\wp$	Power set.
ξ	A vector of posterior probabilities.

MODIFIERS

$\vert\ \vert$	Cardinality.
$\widehat{\ }$	Behavioral estimator.
$\widetilde{\ }$	A subset of a set.