Thesis for the Degree of Doctor of Philosophy

An Integrative Human Activity Recognition Framework based on Smartphone Multimodal Sensors

Manhyung Han

Department of Computer Engineering
Graduate School
Kyung Hee University
Seoul, Korea

August, 2013

An Integrative Human Activity Recognition Framework based on Smartphone Multimodal Sensors

Manhyung Han

Department of Computer Engineering
Graduate School
Kyung Hee University
Seoul, Korea

August, 2013

An Integrative Human Activity Recognition Framework based on Smartphone Multimodal Sensors

by

Manhyung Han

Advised by

Dr. Sungyoung Lee

Submitted to the Department of Computer Engineering
and the Faculty of the Graduate School of
Kyung Hee University in partial fulfillment
of the requirements for degree of
Doctor of Philosophy

Dissertation Committee:

Prof. Ok-Sam Chae, Ph.D. Ok-Sam Chae
Chairman

Prof. Tae-Choong Chung, Ph.D.

Prof. Young-Koo Lee, Ph.D. Young-Koo Lee

Prof. Tae-Seong Kim, Ph.D.

Prof. Sungyoung Lee, Ph.D.

Abstract

Activity recognition is a highly active research area due to its large number of potential applications such as in healthcare, virtual reality, security, surveillance, and advanced user interface systems. Several years ago, such context aware systems were mostly based on complicated wearable sensors, which are not even commercially available nowadays. However, the recent, rapid development of the smartphone industry has enabled implementation of activity recognition applications using the large number of sensors already integrated within smartphones.

Previous researches investigated an activity recognition of simple user activities using a single type of sensor, such as the accelerometer, GPS or audio. Such an approach is not able to support a comprehensive and realistic recognition device. Motivated by the lack of a comprehensive approach in smartphone-based activity recognition researches, a multimodal activity recognizer utilizing several kinds of sensors in a smartphone is proposed. The proposed system combines and validates the output of the two accelerometer and audio classifier with extra information from the GPS and Wi-Fi functions to produce the final result.

In order to apply proposed comprehensive approach and activity recognition framework to mobile devices such as a smartphone, a lightweight activity modeling and recognizing methodology is required. Existing activity recognition approaches are divided into two phase. Training for modeling with activities samples and recognition using modeled information. Complicated but powerful activity recognition algorithms are impossible to apply to smartphone because they requires bunch of sample data for modeling. Also for collecting personalized data such as life-log, a methodology which trains the subject's specific activity pattern and generates personalized activity model is required. So a lightweight activity modeling and recognition algorithm which enables to the users to build their own activity model on mobile platform is proposed. Also based on the lightweight algorithm, a lightweight activity recognition framework which enables to recognize several activities on smartphone is also proposed in this thesis.

Performance evaluations of the accelerometer and audio data classification schemes showed that the proposed algorithm and system performed better than existing approaches. Proposed system is tested by implementing a smartphone application running on an Android OS. These evaluations also showed that the system works well in real-world environments with the accuracy of 92.43%.

Contents

List of Figures

List of Tables

Chapter 1. Introduction

1.1 Motivation

Activity recognition (AR) is a highly active research area due to its large number of potential applications such as in healthcare, virtual reality, security, surveillance, and advanced user interface systems. As a result, it has caught the attention of researchers from industry, academia, security agencies, consumer agencies, and even the general populace. Several years ago, such context aware systems were mostly based on complicated wearable sensors, which are not even commercially available nowadays. However, the recent, rapid development of the smartphone industry has enabled implementation of AR applications using the large number of sensors already integrated within smartphones [1,2].

Nevertheless, substantial progress has only been made for recognition of simple user activities using a single type of sensor, such as the accelerometer [3], GPS [4], or audio tool [5]. Although some recognition of user activities may be possible with particular sensors, such an approach is not able to support a comprehensive and realistic recognition device. For example, to merely recognize ambulatory activities like walking or jogging, the accelerometer or gyroscope achieves a reasonable accuracy [6,7]. Likewise, to classify acoustic contexts, such as in a bus, subway, or meeting place, the audio data can be utilized [8]. The GPS has also been used as a single source to classify different contexts [4,9,10]. Yet, a comprehensive recognition system should make use of all those sensors in order to be capable of recognizing a higher number of mixed contexts including ambulatory, transportation, and acoustic. Furthermore, the use of multiple sensors can improve the power consumption since some sensors can then be activated only when necessary. For example, a system that recognizes transportation by inferring the user's GPS route [11] can stop collecting GPS data if an accelerometer classifier detects that the user is walking.

Motivated by the lack of a comprehensive approach in smartphone-based AR research, a multimodal activity recognizer utilizing several kinds of sensors in a smartphone is proposed. Also consider that the AR must be performed regardless of what the user is doing with his or her smartphone, such as making a phone call, using applications, playing games, or listening to music. Thus, a position-free recognition system that recognizes a human's activities wherever the smartphone is attached on the body is proposed. It provides high degree of freedom to users, as well as ample practical relevance. Besides the classification aspect, the proposed system pursues the optimal combination of sensors in order to reduce the power consumption, which is a vital issue for any smartphone application [12]. The system utilizes the accelerometer to detect transition points from ambulatory activities to transportation activities and vice versa. The audio classifier is only activated if there is a further need to classify transportation activities, such as riding a bus or subway. By using the above approach, it is possible to reduce power consumption on smartphone devices. Finally, the proposed system combines and validates the output of the two classifiers using extra information from the GPS and Wi-Fi functions to produce the final result. By following this approach, the system is able to classify both ambulatory as well as transportation contexts, while still achieving low power consumption.

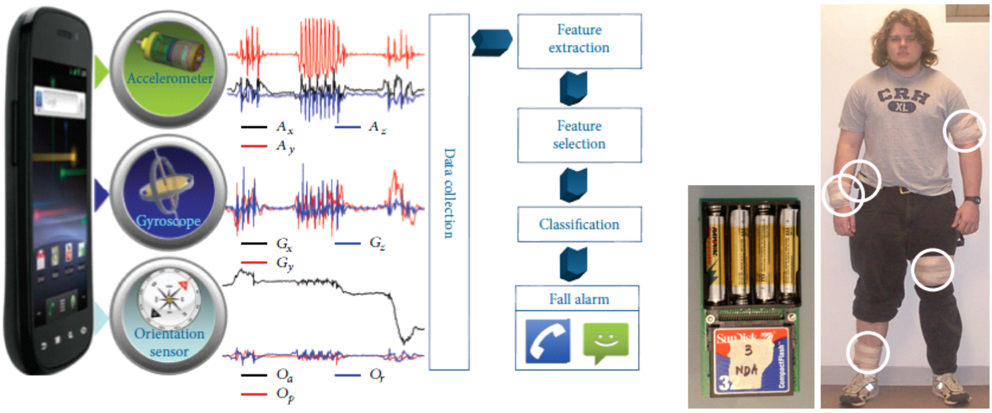
In order to apply proposed comprehensive approach and AR framework to mobile devices such as a smartphone, a lightweight activity modeling and recognizing methodology are required. And life-log data is very private information which required to personalized processing and continuously collected. So the smartphone is a good candidate for the purpose. Existing AR approaches are divided into two phase. Training for modeling with activities samples and recognition using modeled information. Complicated but powerful AR algorithms are impossible to apply to smartphone because they requires a bunch of sample data for modeling. Also for collecting personalized data such as life-log, a methodology which trains the subject's specific activity pattern and generates personalized activity model, not a common activity model applicable to multiple users, is required. So a lightweight activity modeling and recognition algorithm which enables to the users to build their own activity model on mobile platform is proposed. Also based on the lightweight algorithm, a lightweight AR framework which enables to recognize several activities on smartphone is also proposed in this thesis.

1.2 Problem Statement

Most of the studies for human AR used single or simple sensor data among accelerometer, audio, video GPS or light etc. And these researches have several limitations such as position-aware recognition, requiring dedicated devices like wearable sensors or mobile device, and only support off-line processing etc. More detailed problems of current researches are described as below.

- **Considering single or simple sensor:** The high availability of smartphone with built-in sensors is highly advantageous to the research area of context recognition. In [3,6,7], a smartphone accelerometer was used to recognize user's movement contexts such as walking and running. And in [5,8], the authors utilized audio data to classify acoustic environments. The authors of [4,9,11] showed that GPS can be used to recognize transportation routines. However those works merely exploited a particular sensor instead of combining the strength of multiple sensors. To the best of our knowledge, [2] is one of the first works to combine accelerometer and audio classification. The authors demonstrated that the combination of audio helps improve the accuracy of recognizing user activities. But there is no concrete system or framework for smartphone to recognize various types of activities in previous research works.

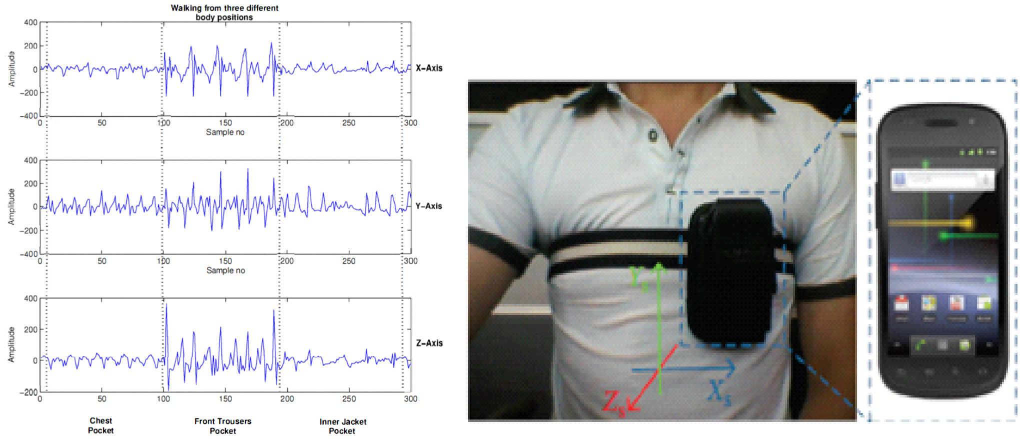
Only algorithms or implementations are proposed and verified in their own environment.



(a) (b)

Figure 1-1. System overview of existing related works. (a) The block diagram of the main data processing scheme which utilizes only one type of sensors – inertial sensors [16]. (b) A biaxial accelerometer sensor for collecting activity data. The research in [3] uses only a single type of sensor and they are attached on 5 different areas of the human body.

- **Position-aware recognition:** In general, the output of any body-worn triaxial accelerometer depends on the position at which it is placed and can vary for the same activity for different positions along the subject's body resulting in high within-class variance. The accelerometer signals for walking, for example, vary at three different positions as shown in Figure 1-2(a). Therefore almost all previous works require accelerometers to be firmly attached to a specific body part such as arm, wrist, chest, thigh etc, making them impractical for long-term activity monitoring during unsupervised free living [17]. Figure 1-2(b) also depicts specific position of the body-worn sensor for recognizing activities of daily living. Therefore, in order to recognize human's ADLs, a position-free data collection or processing techniques are required.



(a) (b)

Figure 1-2. Related works of a position-aware activity recognition. (a) Sample acceleration signals for walking from three different positions(Chest, Front trousers and Inner jacket pocket) [17]. (b) The coordinate system and the placement of the smartphone at the chest of a subject [16].

- **Off-line processing:** Previous works are mainly focused on how accurately recognize activities of human not but considering where the data is processed and how the information is handled throughout a recognition system. General AR system collects activity data from sensor devices and processes the data at back-end system such as server, cloud or any powerful processing device. It may cause privacy problem cause the data is required to sent to an external site [13]. An AR system running purely on a smartphone is presented in [14]. The presented system can be trained on the device and it also does the classification in real-time on the device. The recognition is based on features calculated using geometric template matching and support vector machine (SVM) is used as a classifier. Unfortunately, the article does not include recognition rates: thus, the evaluation of the system is difficult. The system described in [1] can be found from Android Market. It seems to recognize activities with high accuracy, but all the features used are not orientation independent. In addition, personalized mobile AR system for Android phones is presented in [15]. In this application user can select which activities he wants application to recognize but it requires training data collection gathered by the user.

1.3 Contributions

Major problems of previous works are described in section 1.2. To overcome those limitations and apply to resource restricted mobile devices, three major research works are described in this thesis. For recognizing various activities and contexts of human, the methodology and novel framework how to utilize multimodal sensors in both legacy AR system and smartphone are proposed. Proposed framework for a smartphone is able to collect, store and process activity data only on smartphone. Contributions of the thesis are represented as following.

- **Utilizing multimodal sensors:** General AR process is consisted of three steps – Data Collection, Feature Extraction and Classification. Throughout these processes, the most important stage is a feature extraction because the accuracy of total system is highly dependent on the stage. Also an extracting features from different types of raw data is challengeable due to the difficulty of combining and formatting for a classification. So in this research work, a novel hierarchical approach to combine different multimodal sensor data for AR is proposed. In proposed system architecture, a vector typed data such as data from accelerometer, gyroscope, GPS, etc. and a sequential data such as audio data are utilized for decision making. And the recognizable activities and contexts include human's movements, poses, actions, situation and environmental contexts.

- **Novel framework for smartphone:** A framework which utilizes multimodal sensors on smartphone for recognizing personalized activities in real-time is proposed in this thesis. Proposed framework is a conceptual architecture that consisted of several sub modules – sensor data collection, feature extraction, modeling, classification and context fusion etc. - so various classification or extraction algorithms can be selected depending on the target system. In this research work, two types of implementation are proposed and tested. One is for legacy system that a data collection and processing is separated and the other one is for smartphone environment that an activity modeling and classification are run on same platform. Section 3 and 4 describe what kinds of algorithms are used and how the sub components are organized. And the later section represents an implementation of the framework and its evaluation result.

- **On-line processing:** To avoid a privacy problem argued in section 1.2, proposed AR system is able to collect, store and process only on smartphone platform. Due to the resource restriction and relatively-low memory, the lightweight AR framework based on comprehensive AR system is proposed. A lightweight activity modeling and classification algorithm based on the Naïve Bayes, Adaptive Naïve Bayes (A-NB), and the system architecture which based on the A-NB are proposed and tested on smartphone. For an efficient resource management, the lightweight framework utilizes only accelerometer, proximity and GPS sensors except audio.

The proposed common system framework for both legacy system and mobile platform is depicted as Figure 1-3. Data from multimodal sensors on smartphone are collected and processed at accelerometer and audio classifier for acquiring activity labels. Then these labels are fused for making final decision in context fusion module. The activity labels from context fusion module are validated in the Heuristic based Result Validation module for refinement. Result labels from the Comprehensive Activity Recognizer (CAR) are Still, Walking, Jogging, Riding a bus or subway and visit specific location activity contexts.

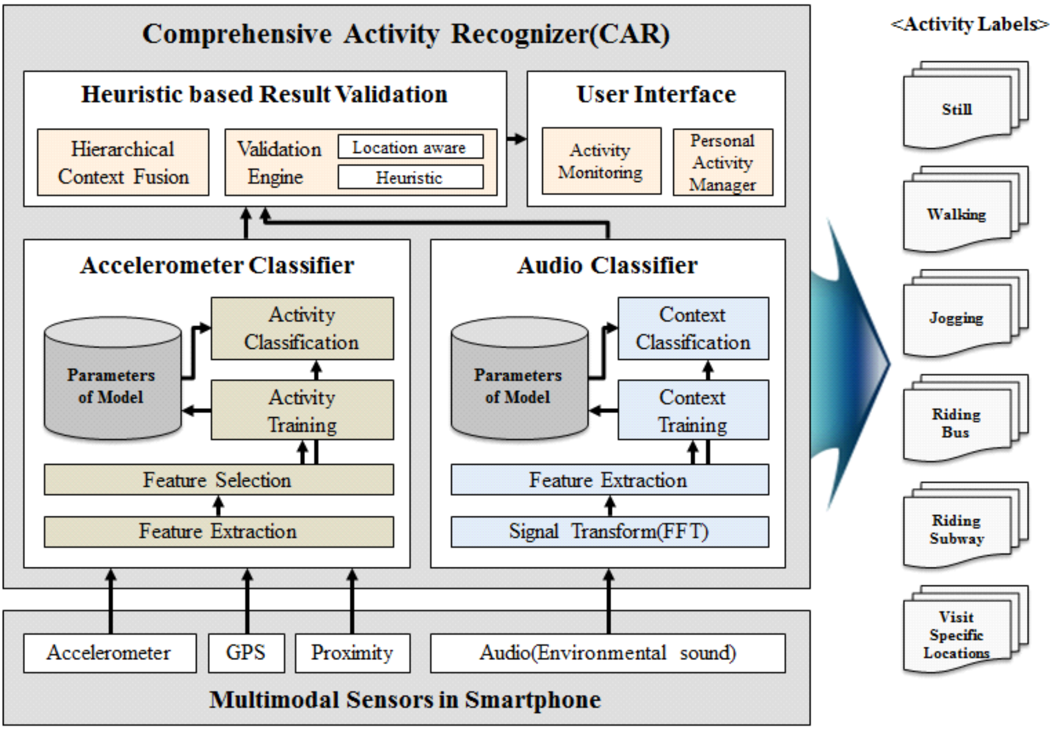


Figure 1-3. Proposed comprehensive AR framework. Totally 5 activity labels are recognized based on smartphone multimodal sensor data.

1.4 Thesis Organization

The thesis is organized in seven chapters as following:

- **Chapter 1 Introduction.** Chapter 1 is the brief introduction of the research works and proposed AR framework. Research challenges, goals and the objective are mainly represented in this chapter.

- **Chapter 2 Activity and Context Recognition.** Chapter 2 provides a definition of activity, situation and context. Also the background of AR, context-aware, user-centric and user-friendly computing are described.

- **Chapter 3 Related Works.** Chapter 3 represents sensors used in existing AR researches and its applications. Detailed recognition techniques, methodologies and context fusion techniques are introduced in this chapter. And the necessity of smartphone for AR is also covered in the chapter.

- **Chapter 4 Comprehensive Activity Recognition Framework.** The proposed AR framework is introduced in this chapter. The components configuring comprehensive activity recognizer system – accelerometer classifier and audio classifier – are discussed.

- **Chapter 5 Lightweight Activity Recognition Framework.** Chapter 5 describes the lightweight activity recognizer for a resource-restricted mobile devices such as smartphone. Detailed algorithms of Adaptive Naïve Bayes and Hierarchical AR Framework are provided.

- **Chapter 6 Implementation and Results.** Detailed performance test results of comprehensive activity recognizer and lightweight activity recognizer with real-world data set are represented. Also the details of heuristic approaches for enhancing an accuracy of AR are described in this chapter.

- **Chapter 7 Conclusion.** This chapter presents the conclusion of this research work and highlights the main contributions.

Chapter 2.
Activity and Context Recognition

2.1 Definition

This chapter gives an overview of activity and context recognition techniques and the important concepts as well as terms related to the design and realization of a user-centric computing system. This overview serves as a basis for the focus and proposal of this thesis. The goal of this work is to investigate suitable approaches that can be applied in the development of a human activity recognition system. Before we introduce the concept of such a system, let us first take a look at the meaning of context and context aware. As an activity itself is a part of the term ‘context’, we define and introduce context, context-aware systems and activity in this chapter.

Context, according to researchers in [18], is defined as location, identities, nearby people and objects, and changes to those objects. This was followed by subsequent work that further investigated the definition of context, context awareness and its potential applications. The other researchers in [19] viewed context as different aspects of the current situation of the user. Researchers at the University of Kent referred to context as the user's location, environment, identity of people around the user, time and temperature [20,21]. Many of these definitions were given either as synonyms for context or by example they were defined as aspects of information needed for their prototypes or applications. A more widely used definition for context was given by authors in [22] as follow:

- **Definition of Context:** Context is any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves.

The definition was not intended to specify how context should be modeled in an implementation. It gives a generalized view on the idea of context. This work views this abstract but useful definition as a general basis for all potential context aware applications. The evolution of context in the work described in [22] has given a clearer picture how context can be applied in an implementation. Authors in [23] pointed out that the researches in [22] did not clarify what was meant by 'the situation of an entity'. The situation can be seen as a complex but definable concept. The Oxford online dictionary defines situation1 as follows:

- **Definition of Situation:**
  1. a set of circumstances in which one finds oneself; a state of affairs
  2. the location and surroundings of a place
  3. formal a position of employment; a job

The first two definitions of the word situation can be applied to context awareness. For a person, his situation can refer to what he is doing, where he is, and his condition at that particular moment. The situation information of a system, both software and hardware, can represent its current state, executable functions or even system related information. In other words, context provides a mean for human users and systems to understand relevant information around them.

In some of the earlier research, contexts were mainly used to reveal further information for desired automatic change of system behaviour. For example, the Active Badge system [24] allows the receptionists to see a user's possible last detected location. In [18], authors demonstrated in their Palo Alto Research Center

Incorporated Tab (PARCTAB) system, a context aware application that is able to present information to users based on proximity to services. Devices can be turned on or reconfigured according to the location of users and selected services can be executed automatically. In these examples, users can see which contexts have been recognized by the application (e.g. location of a user). At the same time, devices and services were initialized and executed based on the obtained contexts. Such automatic execution of devices and services is known as service adaptation. Since the adaptation takes place due to the obtained contexts or their changes, a context aware system can be defined as follows:

- **Definition of Context-aware system:** A context-aware system is a system that delivers and understands the available contexts perceived from the users and the surrounding through the use of sensor information. It also performs the appropriate service adaptations based on these contexts and their changes.

If the definition of context is adopted as abstract as it is, the information measured and provided by a sensor is considered as a context. The interpreted information based on this sensor measurement is also a context. The abstraction levels can be seen as a way to view context from a computational point of view [25]. The first hand information obtained from a sensor device can be referred to as raw data. The processing of this raw data produces low level context. The low level context can be further processed to obtain high-level context. The process is illustrated in Figure 2-1.

A low-level context is regarded as a direct interpretation of the information obtained from a sensor source. It gives a semantic meaning to the obtained value in order to allow further usage of this context. For example, the temperature sensor gives a reading of the voltage potential differences that represent the current temperature of the object or environment it measures. Given the corresponding calculation to this value, one can then obtain the current temperature with a desired unit, such as 23C. The third abstraction level of context is called the high-level context. It expresses the information that is usually interpreted from one or more low-level contexts. For example, a person may regard the surrounding temperature 23C as warm. One can also conclude that a person is busy, when he is located in his office while his computer is turned on and is currently not idle. Usually, high-level contexts are semantically understandable and are implicitly perceived by human automatically. In context awareness, the computers are expected to be able to produce and use these contexts. In other words, the computers can potentially understand what and how a human thinks and perceives.

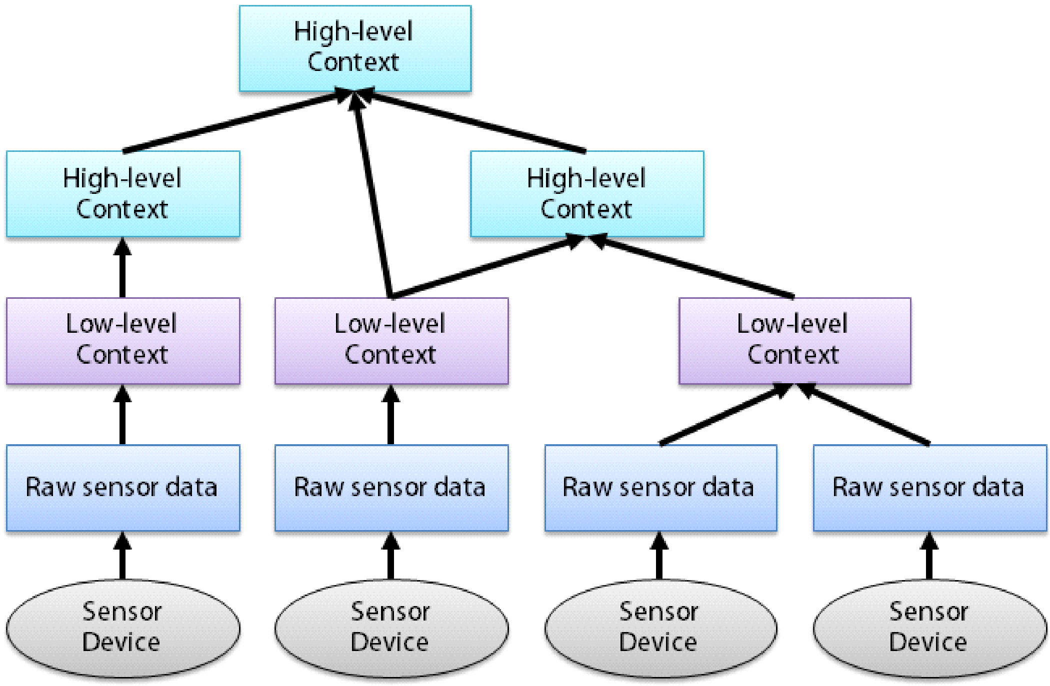


Figure 2-1. Three levels of abstraction for contexts.

To the best of our knowledge, the distinctions and definitions for raw data, low-level contexts and high-level contexts are not often discussed in the literature. Nevertheless, these distinctions can provide designers and users of context aware systems a better understanding of the required contexts in a desired implementation and usage. The abstraction levels are useful to allow human users to understand what a system has processed and interpreted. For a computing device, the different abstractions are basically still strings, where these strings are results from the processing of the sensor data [25].

Besides the three abstraction level of contexts, the concept of contexts can also be understood by the categorization of the aspects of context. For example, in [18], the authors mentioned that three important aspects of contexts are ‘where you are, who you are with, and what resources are nearby.’ In [26] researchers adopted and extended the definition of [22] by categorizing contexts into different aspects such as geographical, physical, organizational, social, user, task, action, technological and time. Authors in [25] presented a more complete picture on the aspects of contexts by naming 14 different aspects of contexts. They grouped these 14 aspects into 5 main aspects, which are time, location, constitution, environment and identity. The aspects of context help us to have a clearer picture how contexts are defined and applied in a context aware system. At the same time, it also clarifies the ambiguity that may occur when the definition of [22] is adopted. The process to obtain low and high-level contexts from sensor data is commonly known as context modeling. A context model defines contexts that are understandable by machines and users. Authors in [27] presented a summary on most relevant context modeling approaches. These approaches were key-value, markup scheme, graphical, object oriented, logic based and ontology based models. There are cases where an ontology is required for its expressiveness, but there are also cases where simple key-value implementations will suffice. To achieve the balance between user control and automation, the selection of a user accessible and comprehensible context model are factors to be considered.

In this thesis, the activity of a user, a specific aspect of context, is mainly focused. The expected service adaptation in a context aware system is usually dependent on what a user is doing and his situation at a given time. Besides the context of time and location, this aspect of context is seen as equally important in the context aware environment because a user is constantly performing a certain activity. The word ‘activity’, according to the online Oxford dictionary, is defined as follows:

- **Definition of Activity:**
  1. a condition in which things are happening or being done.
  2. busy or vigorous action or movement.
  3. an action taken in pursuit of an objective.
  4. a recreational pursuit.
  5. the degree to which something displays its characteristic property or behaviour.

For the use of activity recognition in context awareness, the first three definitions can be seen as appropriate. The concept of activity describes things that are happening around a person. By including the second and third definitions, the word activity extends the concept to additional details of a given activity. These details include the involved actions or movements and possible description of the objective for the occurrence of a given activity.

2.2 Activity Recognition

As mentioned in the Section 2.1, this thesis focuses on the activity context of a user. Activity context can be seen as a subset of situation information, because it explains what a user is doing at a certain time. The activity context can be further broken down into more detailed categories. An activity is usually time and location dependent. A user's action or movement can also be grouped as his activity. Since activities are not always directly measurable, activity recognition techniques are applied to enable the acquisition of users' activity contexts. The word recognition is defined as ‘The action or process of recognizing or being recognized, in particular the identification of a thing or person from previous encounters or knowledge.’

An activity recognition system is able to detect a person's current action or movement based on the available information. This information is usually obtained from the available sensors that are placed either on a user and/or around him in the environment where he is situated. The selected activity recognition is responsible to find correlations and relationships in these sensors data in order to discover the corresponding activity contexts. Examples of sensors used in past investigations for the area of activity recognition are accelerometer, microphone, camera, heart rate belt or Radio-frequency identification (RFID) tags. The types of desired activities related contexts include movements, tasks, locations and presence information (such as busy, available or away). These sensors can be categorized as their designated deployments.

- **Wearable sensor devices:** In the first category, a user is usually equipped with one or more devices, placed at different parts of the body. Each device may have more than one sensor built-in, together with the necessary processing and communication modules. The processing modules enable simple or even intensive computation and manipulation of sensor data. The communication modules are responsible to transmit the obtained and processed sensor data to a remote device for further processing and storage tasks. The wearable sensors are responsible to record sensor values and their changes, which should generally correlate to the activities undertaken by the wearer. By applying appropriate algorithms, one can find useful information from these sensor values that can eventually be used to detect the corresponding activities.

- **Multimodal sensors on smartphone:** It is a kind of wearable sensors but various sensor devices are embedded on a single platform. An activity data is collected by multiple devices simultaneously. Not only data collection but also data storage and processing are performed in same device. Above all, sensing from smartphone sensors does not required any dedicated sensor devices and the user always carrying it.

- **Environmental sensor devices:** It is also possible to recognize activities in a given place such as a room. Sensors such as camera, microphone and RFID tags are deployed to provide implicit information that can be related to the activities carried out by the occupants at a given place. Examples of the activity information range from location of the occupants to the tasks a person is currently doing at his desktop. Such examples are found among the typical scenarios mentioned in the related investigations.

- **Combination of both wearable and environmental sensor devices:** There exists investigations that combine the above categories for designated activity recognition. Sensor information from various type of sensors is aggregated and analyzed to produce activity and other useful contexts for potential further usages and adaptations.

Similar to the abstraction levels for contexts, activities can be grouped and defined into three different categories as shown in Figure 2-2. The first category of activities consists of gestures and movements. It can be seen as the simplest form of activities that can be useful for potential context aware adaptation. Gestures and movements are usually short and possibly repetitive. The recognition of gestures and movements can be applied in areas such as recognizing sign languages [28] and intuitive human-computer interface input [29,30]. It can also be the input information for the second category of activity - basic activities.

Basic activities can be seen as activities that involve different combinations of gestures and movements. For instance, a sequence of repeating steps can indicate that a person is walking or running, depending of the speed of the steps. This type of activities last longer than gestures and movements, ranging from seconds to minutes or hours. These basic activities are considered as important information, because people perform them in daily lives. Once basic activities are recognized, they can be used to derive specific activities.

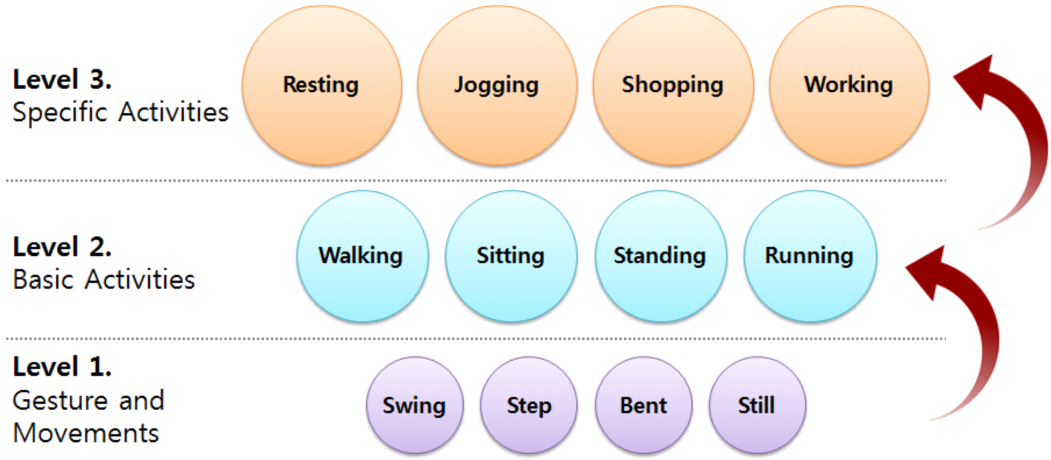


Figure 2-2. Three different levels of human activities.

The specific activities are defined as high level descriptions of task and event one performs at a given time and situation, i.e. working, jogging, shopping, sightseeing or relaxing. As an example, if an activity recognition system recognizes that a person is walking, it may be possible to use additional information to derive whether the person is currently sightseeing or shopping. For the former, the system may need to obtain his calendar information to know he is on holiday and he is walking around a tourist location away from his home. For the latter, the person should be located in a place where people usually shop, such as a shopping center.

The first two categories of activity are normally derivable via sensor data. Once distinct body motions are detectable, a system may be able to tell some gestures or basic activities apart. Depending on the available sensors and recognition approaches, basic activities can be recognized based on gestures and their transitions, or they can also be directly modeled and detected. For the last category one needs to model the possible relationships between the available context information that may help to reveal a user's specific activity at a given time.

2.3 Activity and Context-aware Framework

There had been a lot of important investigations and efforts in bringing the vision into working prototypes and applications in activity and context-aware system. The Active Badge system [24] and the PARCTAB work [18] are two of the earliest context aware applications. A similarity among these earliest applications is the focus on location-aware adaptations and functions. In other words, the location information is the main contexts considered and applied. These systems are commonly known as location-based services today. Another common demonstration of location-based context aware application is a tour guide system. As tourists move around different attractions in a town, the tour guide can make use of the location context to present useful tourist information of different nearby highlights. Some examples are the work of [31,32]. Currently, there are already a number of emerging location-based services such as FourSquare and Google Latitude that are made available as products and services for the masses.

As mentioned by the authors in [33], there are more contexts than location. This can be observed in later work after the end of the 1990s. For example, the CybreMinder system [22] is a context aware application that sends out reminders based on time, location and situational contexts. Several other domains that have adopted the context aware approaches are smart homes [34], personalization [35] and health care [36,37]. Nevertheless, regardless of how these investigations have selected the relevant contexts for the respective application domains, the choices and types of the selected contexts are consistent with the definitions of context and context aware system mentioned earlier on section 2.1.

There are also different context-aware frameworks proposed in the past years. These frameworks include the Context Toolkit [38], the Context Management Framework (CMF) [39], the Context Broker Architecture (CoBrA) [40], the Context-aware Sub-Structure (CASS) middleware [41] and the Service-Oriented Context Aware Middleware (SOCAM) [42]. There are a number of similarities among them. These systems may use different names or categorizations of their functions, but it is possible to summarise them in a generalized procedures for understanding purposes. This procedures, as illustrated in Figure 2-3, gives an overview of the core functions of the different roles and components found in a context-aware system.

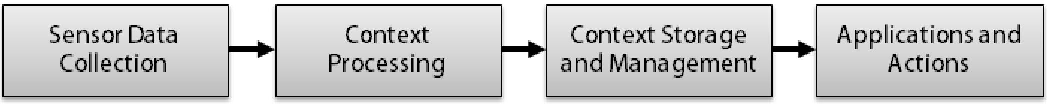


Figure 2-3. A general procedure of previous context-aware system.

The first step is sensor data collection. The sensors are responsible for the sensing of information that can be used to derive usable contexts. Commonly used sensors are the physical sensors, such as temperature, barometer or accelerometer. This type of sensor information usually provides raw data or low level contexts. Besides that, it is also possible to extract information from software and hardware and offer them via virtual sensors. Examples of virtual sensor information are personal calendar entries and smartphone call status. The contexts provided by virtual sensors are generally low level or high level contexts.

The processes that derive contexts are found in the second step. Context processing includes a number of functions, such as filtering, context learning, context interpretation, context reasoning and prediction. Many context awareness related research projects focus in this step. The next step of context processing is the context storage and management. It deals with the issues on how the derived and produced contexts are stored and managed. Components, such as context repository and context broker, are found here. The final step utilizes components from the second and third steps for potential applications and actions. The available contexts for a specific application domain may be used by respective applications for display and adaptation purposes. It is also possible that a context aware application automatically invokes desired actions that are triggered by context-changes.

Currently, a conventional system usually requires human user input in order to provide appropriate services that satisfy the user's needs. In cases where service wishes are repetitive, there are systems that remembers the wishes, so that at a recurrence of these wishes, the system can provide the functions in an automatic manner. Current conventional systems usually use two direct methods to automate certain processes. Firstly, they remember the necessary settings at the point where the users last used the system. Typical examples are like radio or entertainment systems that remember last played station or song. A basic assumption made here is that the user wishes to continue where he was in a previous usage. Secondly, the systems support the utilization of more complex settings by allowing users to program or customize their service behaviors. By stating and defining how, what and when the system should perform the desired functions, a system can execute these service wishes as defined. This ability that enables a system to act and respond according to pre-defined instructions is commonly known as an adaptivity.

The vision of context aware computing intends to bring such adaptation and automation to the next level. Instead of providing just passive adaptivity, a context aware system aims to provide active adaptivity. In other words, the service wishes and behaviors should be recognized by the system in order to learn and recognize them correctly. Once this is done, the system can then provide appropriate adaptations that correspond to the user's needs. There should be a minimum amount of direct input to enable this ability.

2.4 User-centric and User-friendly Computing

Many of the past investigations were made strongly based on the design and thoughts of the system developers. The designers and developers were responsible for the assumptions and decisions set on behalf of the potential users. The desired prototypes and systems were then built based on these assumptions and decisions. While this is perhaps unavoidable for most of the cases, if the issue is addressed and considered from the beginning of design and development of an activity recognition system, the end product may be more likely to be well received by the designated users [43].

An activity recognition is one of the most important element in a user-centric environment. The complexity and technicality should be appropriately reduced and hidden, if not removed, so that users can have the necessary control of the system without being overwhelmed. This vision has resulted as the idea of a user-centric activity recognition system. In this thesis, a user-centric computing is defined as a system that is designed and developed with the users and their needs placed at the center. Based on this definition, some design requirements for a user-centric activity recognition system are listed below:

- Users and their needs should be taken into consideration during the design of a context aware system.
- Users should be empowered to be able to use and to have control over the system and its behaviors.
- Users should be comfortable with the deployment and usage of the system.
- The system should deliver reliable performance and results, so that users do not relinquish it just because it fails to deliver.

The first requirement is usually considered in many past investigations. Most of the investigations have proposed solutions that automate tasks and functions on behalf of the users upon context changes. Nevertheless, there are also approaches where systems automatically provide users with information and service adaptations that are relevant only to the opinion of the developers [44]. The second requirement is however not always found [45]. In investigations such as [18,32], the proposed solution and applied methods for context models are static in nature. There are also approaches that react fully automatic on behalf of the users. The users are not provided with the means to alter the provided adaptation possibilities. If there is an appropriate method that can at least give users a better understanding and overview of what the context aware system is going to perform, it may reduce the chances where users feel a lack of control.

The third requirement takes users' acceptance and comfort for an activity recognition system into consideration. An activity recognition system can be designed in such a way where multiple sensors and devices work seamlessly to acquire useful contexts. If the system utilizes body-worn sensors, the user is required to wear and carry these sensors throughout the whole day. He may likely to reject the system if the setup of the sensors and devices is troublesome or makes him feel uncomfortable by wearing them, or if both sensors and devices need frequent replacement because they run out of battery. Similarly, the fourth requirement also emphasizes on user experience from the perspective of performance. An activity recognition system should deliver reliable performance and results that fulfill the promised features it claims to deliver. If it creates mistakes and frustration frequently, the users will give up using the system. Summarizing the third and fourth requirements, one important factor to be highlighted is the possibility for an activity recognition system being obtrusive. The word obtrusive is defined as noticeable or prominent in an unwelcome or intrusive way.

One potential hindrance for a user-centric and user-friendly computing is the obtrusiveness of the system. A system can be obtrusive in the usage and control of its functions. Complicated interfaces or control mechanisms may cause the users to feel frustrated. Devices can also be obtrusive in different ways, such as the wearing and placement of multiple sensor devices, inconveniences caused by slow response time or short battery life of a device, as well as the control and management of the devices that require constant user attention and intervention. In other words, the obtrusiveness factor in a context aware system brings inconvenience to the users or requires them to make changes to their normal routines and habits.

The proposed vision of Mark Weiser had already emphasized on this factor. In his paper [46] he used the expressions ‘invisible’ and fade into the background to describe how unobtrusive technologies are necessary in a future computing environment. On the one hand, the system should be invisible or unconscious to the users so that no unnecessary interaction or direct user input are required. On the other hand, users should not feel lost among these invisible technologies. This vision motivates us to formulate a user-centric activity recognition system that emphasizes on the use of techniques and approaches that are user-centric and user-friendly.

This chapter reviewed and refined the definitions and concepts that were relevant to the thesis. Based on these definitions and concepts, the proposal of a human activity recognition system was presented. As users' needs and acceptance were set as requirements, the need use multimodal sensors on smartphone and to include unobtrusive approaches in the design and development are described in this chapter.

Chapter 3. Related Works

In this chapter, the ideas of activity recognition using various sensor devices, its applications and recognition algorithms are introduced. As mentioned in the previous chapter, obtrusiveness of a system and its devices can be a hindrance for the system to be accepted and used by the designated users. For such cases, we need to identify approaches and devices that are less or unobtrusive and at the same time deliver equivalent performance like the obtrusive approaches. This chapter first presents related works in activity recognition and then review previous activity recognition system architectures using multimodal sensors on smartphone.

3.1 Sensors used in Activity Recognition

There are numerous techniques to recognize user activities. Popular methods make use of different types of sensors such as microphone and camera, body-worn sensors as well as wireless sensor nodes in the environment. The different types of sensors require a different kind of techniques for the desired recognition. The following reviews elaborate the selected work that have investigated the use of different sensors and techniques in this area.

The earlier work investigated the possibilities of using body-worn sensors for activity recognition. In the late 90s, researchers in Massachusetts Institute of Technology (MIT) [47] explored gestures and activity recognition techniques based on video [48] and audio data [49]. For example, in [48] one can use a small camera, installed on a cap, to track hand movements to interpret the performed sign language signals. With the help of a microphone and a Personal Digital Assistant (PDA), the authors in [49] processed audio streams to detect speeches or even proximity of several users. Video-based approaches are also found in premises monitoring. In the work of [50], video was used to detect Activities of daily living (ADL) for the care of elderly people. The use of audio and video as sensor data has a problem, where installation and monitoring using microphones and video cameras often raise the issue of trust and privacy.

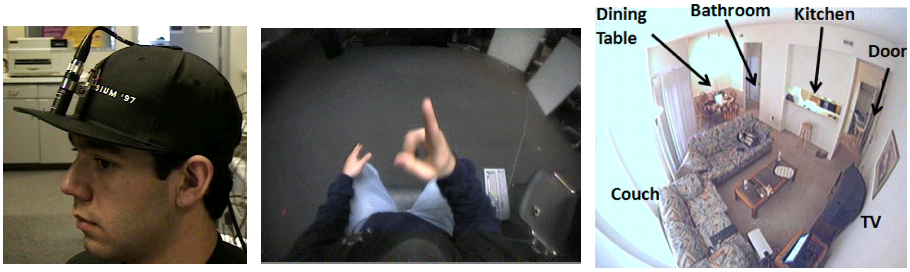


Figure 3-1. A small camera sensor installed on a cap for tracking hand movements [48]. Video sensor for recognizing ADLs for the care of elderly people [50].

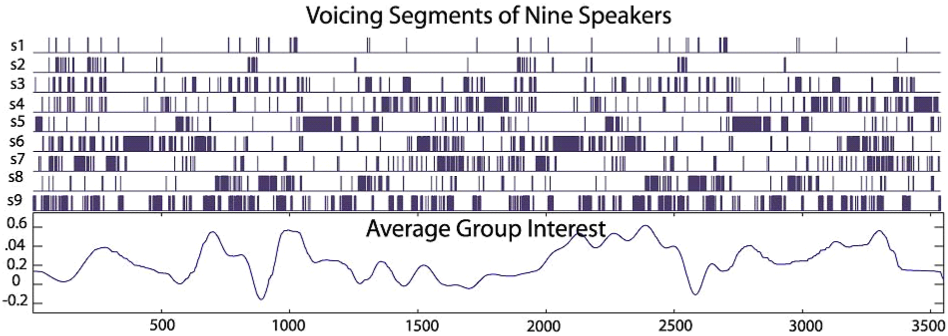
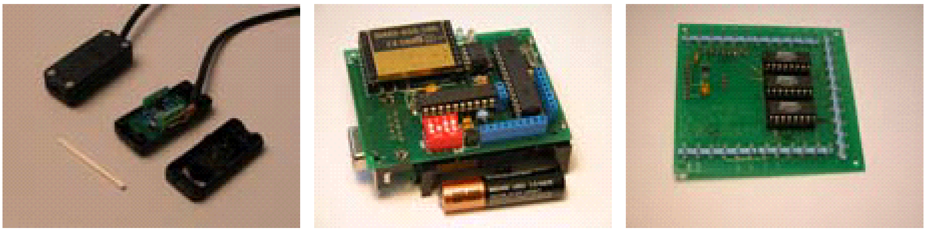


Figure 3-2. Audio streams to detect speeches of proximity of several users [49].

Besides audio and video, other sensors have been tested for potential activity recognition. Body-worn accelerometers are seen as a popular choice as the main sensor for activity recognition. Works such as [3,7,52,54,55] demonstrated that the use of dedicated accelerometers can provide good results in the area of activity recognition. For example, in [3], the recognition accuracy was 84.26% using five biaxial accelerometers on different body parts [56]. Similarly, researchers in [52] achieved accuracy up to 84% using 12 triaxial accelerometers and a Naïve Bayes classifier. For example, authors in [7] obtained accuracy up to 99.82% using a dedicated triaxial accelerometer mounted on a harder board placed near the pelvic region of a test person and meta-level classifiers.



(a) 3D acceleration sensor nodes and their robust casing

(b) Lancaster Smart-It

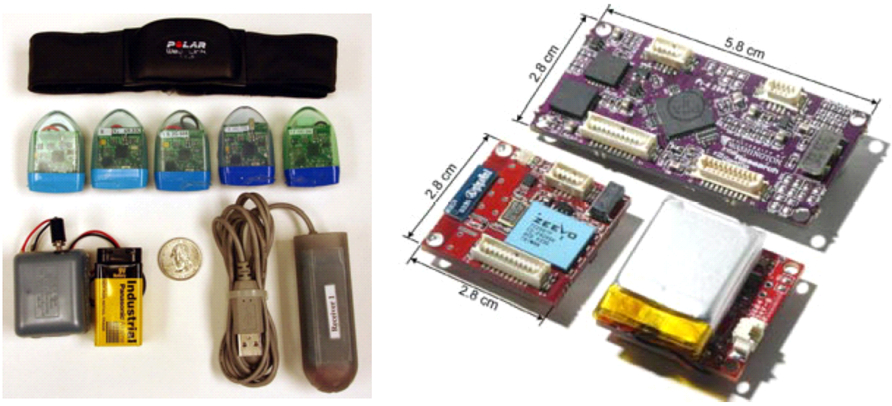
(c) Multiplexer board to connect six triaxial acceleration sensors

Figure 3-3. The different parts of the acceleration recording platform [52].

The different investigations have one thing in common - a single accelerometer or multiple accelerometers were placed on different parts of the body, either wired or wireless, and users were required to perform designated movements. The recorded accelerometer values were processed and the resulted features were evaluated using classification algorithms for potential recognition. The accelerometer can be seen as an appropriate unobtrusive sensor device. It is small and uses relatively small amount of energy. However, these previous investigations focused more on off-line recognitions and did not propose how suitable is the usage of accelerometer together with a mobile computing device in a real-time recognition system. The usage of five up to twelve accelerometers was also seen as a rather obtrusive approach, since users need to wear all the accelerometers on the designated positions in order to make use of the proposed systems.

The ideas were expanded with the inclusion of additional sensor information. For example, in [57] a heart rate monitor was coupled with data, taken from five accelerometers, to detect physical activities. Researchers at the Intel Research in

Seattle and the University of Washington used the Multi-modal sensor board (MSB) that had accelerometer, audio, temperature, Infrared(IR)/visible/high-frequency light, humidity, barometric pressure and digital compass [58]. They investigated activity recognition classification of physical activities with multiple MSBs. The group in [59] used a triaxial accelerometer together with a wearable camera to recognize human activity. The combination of these two sensors was used to recognize whether a user was walking forward or backward, standing, sitting, turning or taking the elevator. The inclusion of additional sensors can help improve the recognition of specific activities, particularly when the information from an accelerometer sensor was insufficient to correctly recognize them.

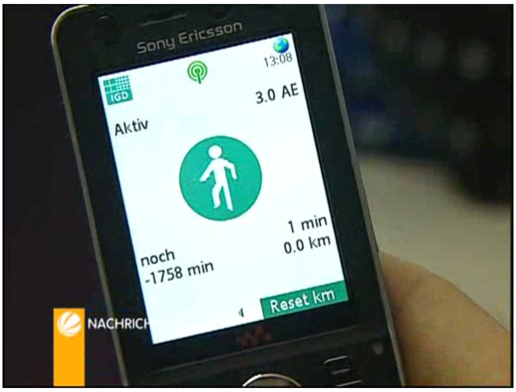


(a) (b)

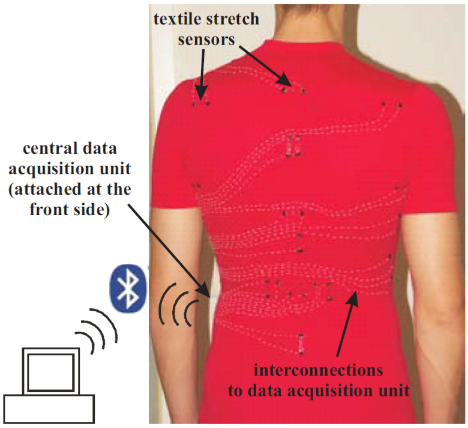
Figure 3-4. Sensors used for recognizing user's physical movements. (a) Five 3-axis wireless accelerometers, a heart rate monitor and USB wireless receiver[57], (b) The multimodal sensor board (top), a Bluetooth iMote (lower left), and USB rechargeable battery board (lower right) [58].

Recently, the 3D accelerometer integrated in smartphones was also investigated as a potential sensor for movement recognition. In [60], the accelerometer of a Nokia N95 was used as a step counter. The results showed that such smartphones can provide accurate step-counts, comparable to some of the commercial and dedicated step counter products, provided the phone is firmly attached to the body. The DiaTrace project [61] uses a mobile phone with accelerometers for physical activity monitoring. The prototype obtained accuracy of over 95% for activity types of resting, walking, running, cycling and car driving. Researcher in [62] also used the acceleration data collected with a Nokia N95 with K-nearest neighbor algorithm to detect common movements [62]. This approach is regarded as suitable, since a smartphone is getting more and more common and can be used as an unobtrusive device for the purpose of activity recognition. However, the above investigations did not compare the respective applied methods with other classification algorithms, used in related investigations with one or multiple dedicated accelerometers. There was also no investigation to compare performance related issues, such as recognition speed and influence on the battery life of the smartphones.

There are also investigations in integrating sensors on garments for various activity recognition tasks. A researcher from the ETH Zurich used tight-fitting clothing and strain sensors to measure body posture [63]. The SMArt SHirt (SMASH) [64] uses accelerometers integrated in the garment for potential rehabilitation applications, such as movement and posture rehabilitation for children. The Konnex unit has a TI's Mixed Signal Processor 430 (MSP430) microprocessor and performs the designated pattern recognition tasks. The SMASH Gateways is designed to acquire sensor data from the accelerometers and to perform feature extraction on these data before sending them to the Konnex unit. University of Passau used conductive textile based electrodes that are integrated in a garment to detect specific activities such as chewing, swallowing, speaking or sighing [65]. This capacitive-based sensing approach measures capacitance changes inside the human body that can give correlated information related to movements and shape changes of muscle, skin, and other tissues [65]. These investigations using smart garments are also seen as unobtrusive approaches. However, they are still prototypes and proof of concepts garments. It is not possible to obtain one in the commercial market for immediate usage and integration for everyday activities.



(a)



(b)

Figure 3-5. Prototypes in integrating sensors on garments. (a) Phone with integrated sensor showing actual activity [61], (b) Architecture overview of a prototype recognizing upper body postures using strain sensors [63].

Apart from body-worn sensors, there are also approaches that utilize small sensors embedded in mobile devices and artifacts. The Technology Enabling Awareness (TEA) project investigated how multiple sensor context awareness can be realized in a self-contained device [66]. A TEA device, consists of photo diodes, microphones, accelerometer, digital temperature sensor and touch sensor, as well as a micro controller, can be attached to mobile devices for the use of activity recognition [66,67]. Similar embedded sensor devices can also be integrated in normal daily life objects. For example, Active Badge [68] uses RFID for the recognition of the user's location. The MediaCup [69] is another example of a recognition technology augmented non-computational artifact. The system can detect context changes such as temperature of the cup, keyboard activities (via keyboard clicking sound) and cup movements. The researchers in University of Linz developed a cube with sensors such as accelerometer and gyroscope to be used as a remote control device with tangible user interface for set-top boxes [70]. These approaches are also considered as unobtrusive, as long as the sensor devices can be kept small and can be accepted by their potential users without requiring them to make big adjustments in their everyday lives.

3.2 Activity Recognition Applications

A major area where activity recognition plays an important role is the health care. With the possibility to detect activities automatically and reliably, advanced services such as remote patient monitoring or therapy can be offered. For instance, authors in [71] investigated detection of human motion states using the SenseWear Pro2 sensor armband. They aim to use activity recognition for diagnosis of sleep disorder, screening of treatment effort as well as monitoring of motions and exercise prescription for patients with chronic disease. University of Karlsruhe [72] used a dedicated acceleration sensor to monitor physical activity of a patient. By combining ECG signal, blood-pressure and physical activity, the developed system provides continuous monitoring of patients for the prevention or assisted therapy of cardiovascular diseases in their daily lives. In the investigation of [73], the pedometer and accelerometers were used to assist the quantification of the 6-minute walk test (6MWT) performance for patients with Chronic Heart Failure (CHF).

Figure 3-6. Body-worn sensor for chronic disease patients. SenseWear Pro2 armband sensor for diagnosis of sleep disorder, screening of treatment and exercise prescription [71].

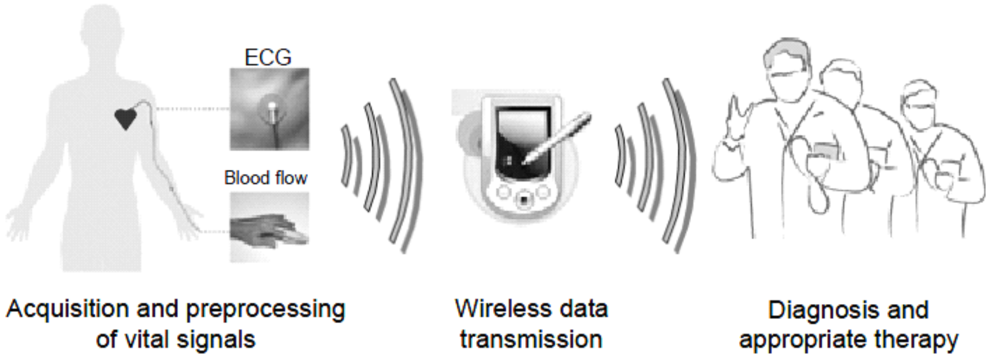


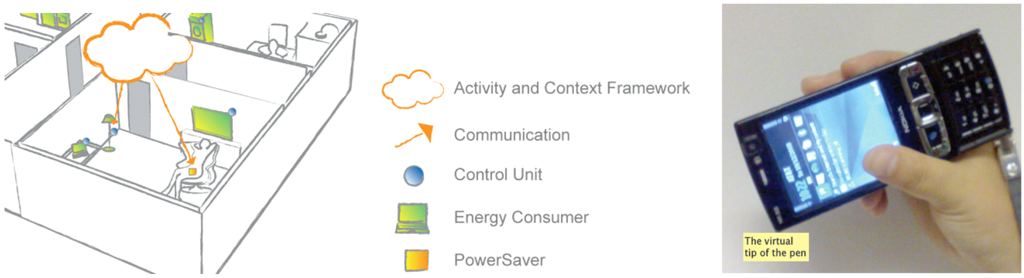
Figure 3-7. Overview of the CALM project which provides continuous monitoring of patients for the prevention or assisted therapy of cardiovascular diseases [72].

The availability of activity contexts is also investigated and used in smart environment and assisted living areas. This specific area of interest intends to use the recognition to provide appropriate services and adaptations. The Gator Tech Smart House [74], located in Gainesville, Florida, is a project that aims to create assisting environments using sensors, actuators and services running on a middle ware. Its goal is to realize homes with the ability to sense context information on the buildings and their residents, such as arrival of new mails in the mailbox, sleep pattern monitoring, contents in the refrigerator and floor that tracks location of the occupants and fall detection. Another smart home project is the Managing An Intelligent Versatile Home (MavHome) from the University of Texas of Arlington [75]. With the use of sensor information, the MavHome intelligent agent predicts next action of the inhabitant to automate the repetitive tasks for them. Activities were recorded and analyzed to discover available patterns for the designated prediction tasks. The iDorm1 and iDorm2 projects from the University of Essex [76] used embedded agents to sense and recognize abnormal activities that take place in an environment such as a flat.

Other applications of activity recognition for smart environment and assisted living include energy management, safety and elderly care. For example, authors in

[77] used wireless camera sensor nodes in a building to predict occupancy for the rooms. With this occupancy information, the system automatically makes adjustments and controls the heating, ventilating and air conditioning (HVAC) systems installed in the rooms. Researchers in [78] proposed a system using a small sensor box with a built-in accelerometer to recognize three simple movements sitting, standing and walking. The different electrical appliances in the room are controlled via a management system by receiving different context information from the users and the environment. The researchers at University of Washington and Microsoft proposed a video surveillance system that analyzes video images to detect new and unusual activities [79].

Another emerging application of activity recognition techniques focuses on how users interact with devices. The utilization of wireless controller that recognizes human gestures as game controls have been popular since the last few years. Nintendo Wii game console as well as smartphones such as Apple iPhones, Nokia N-series and Android-based phones are some examples that took advantage of this possibility. The gesture-based game control has provided new experiences and new ways to interact with the supported games. Besides gaming, these applications can also provide more intuitive interaction between users and devices in a similar manner. As mentioned earlier, the research area of tangible user interfaces investigates how gestures and movements can be used for different purposes. For example, the uWave gesture recognition system supports devices such as Wii remote controller, custom-built micro-controller and smartphone. Applications tested with the uWave system include gesture-based authentication and 3D mobile user interface control. Duke University Durham presented the PhonePoint Pen system that turns a Nokia N95 smartphone into a hand-writing and drawing tool [80]. These investigations have demonstrated possibilities for users to interact with devices in a more natural and unobtrusive manner.



(a) (b)

Figure 3-8. Utilizing smartphone for a gesture recognition. (a) Activity recognition using accelerometer for the PowerSaver solutions [78], (b) Pretending the phone's corner to be the pen-tip [80].

3.3 Recognition Methodology

3.3.1 Learning Techniques

The types of recognition techniques are usually dependent on the types of sensors and available data used in the respective investigations. Generally, the methods used can be categorized into two main groups. The first group is the supervised learning approach to train the given algorithm with labeled data to generate models that can be used for the designated recognition. The second group is known as unsupervised learning approach. It attempts to construct usable models without the need of having the data labeled. In other words, the unsupervised approach aims to discover available information within a set of unlabeled data that allows future recognition. Typically, both methods do not work on the raw data directly. The available measured data are first pre-processed to be transformed into features. Selected learning algorithms are then applied to analyze these features to draw useful observation and patterns from the transformed data.

The supervised learning methods are considered as the most predominantly applied approach in the field of activity recognition [81,82]. For example, simple but efficient base-level classification algorithms such as the k-nearest neighbor, decision tree, Naïve Bayes and Bayesian Network were used in various investigations [83,84,85,87]. There are also investigations that used other base-level classification algorithms such as Support Vector Machines (SVM) [86,88,89], Hidden Markov Model (HMM) [90,91] and neural network based classifiers [92,93].

It is also possible to combine more than one base-level classier or to reuse the same base-level classier in multiple iteration in order to improve the recognition accuracy. This approach, also known as meta-level classification, is also used in different previous activity recognition investigations. For example, the meta-level classifiers are used in investigations such as [94,95]. It is shown that the meta-classifiers generally improve the recognition accuracies as compared to the respective single base-level classifiers. Researchers in [58] has combined static classifiers with HMM to improve the intended recognition accuracy.

The unsupervised learning methods are not as common as the supervised approaches. A popular approach is to utilize clustering techniques to group similar patterns found in the data as possible activities. In [96], authors have used hierarchies of HMMs from audio and video data to perform unsupervised time series clustering on activities. They also have applied the use of multiple Eigenspaces to enable unsupervised learning of basic activities based on the measured triaxial accelerometer data. Authors in [97] have shown that motifs (recurring patterns) found in data can be detected using unsupervised methods such as Symbolic Aggregate Approximation (SAX) technique developed by [98]. They have investigated efficient approaches for multi variate motif discovery applicable for different domains such as activity discovery using body-worn accelerometer and gyroscope data. In the investigation of [82], an unsupervised fingerprint-based algorithm has been proposed to recognize activities in a smart home environment using a wearable RFID system.

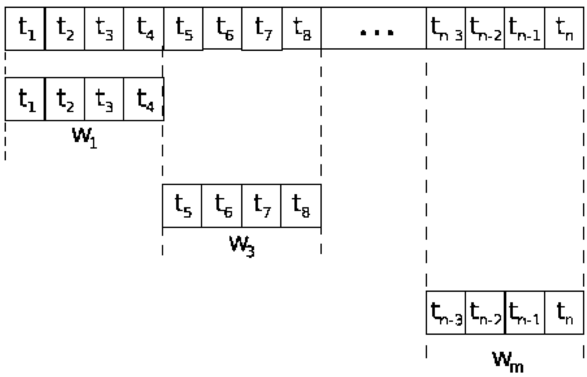
3.3.2 Feature Extraction Techniques

Feature extraction is a technique that enables the reduction of dimensionality of the data and the discovery of useful patterns. A feature is defined as a new attribute generated from the original raw data. Example of this technique can be seen in various fields such as image processing and signal pattern matching. Feature extraction is useful especially when the original data are not directly usable for potential processing using algorithms such as classification or clustering. One specific type of such data is a time series data. In [99], researchers stated that a time series is not suitable to be directly analyzed and processed by classification algorithms. Similarly, the accelerometer data, which are also time series, need to be transformed to obtain suitable features that can be applicable in the activity recognition processes.

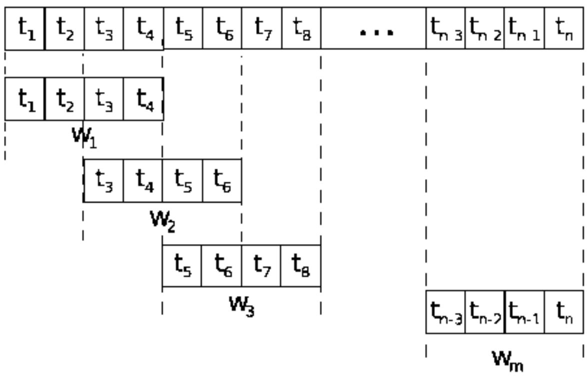
3.3.2.1 Sliding Window

The first step in feature extraction process is to split the given accelerometer data into data segments with a fixed interval. This technique is usually known as the sliding window algorithm [100]. The algorithm is useful if one wishes to compare the segments to discover recurring patterns. A data segment is grown until it exceeds the given interval to form a so-called window. If no overlapping is desired, the following window starts from where the previous window stops at. This can be illustrated as in the Figure 3-9(a). In this figure, a time series sample is split into m windows. Each window has a window length of four samples. An overlapping sliding window has its following segments starting from a certain position (depending on the percentage of the overlap) in the newly created segment. An example is shown in Figure 3-9(b) where new data segments are generated using a sliding window of four samples with 50% overlap.

The sliding window technique is commonly used in previous investigations. The accelerometer data are processed to produce three sets of data segments before the feature extraction step is performed. Most of the investigations using classification techniques selected 50% as the overlap percentage and the research described in [101] used 75% overlapping. The use of overlapping windows has the advantage of retaining the similarity of data segments. The repeated samples in two subsequent windows may increase the similarity between them. It is also useful when the sample size of the training data is relative small [102]. A set of training data produces more instances with a higher percentage of overlapping than the same training data with a lower percentage of overlapping.



(a) Sliding window with no overlapping



(b) Sliding window with 50 % overlapping

Figure 3-9. Data segmentation using sliding window technique [100,101].

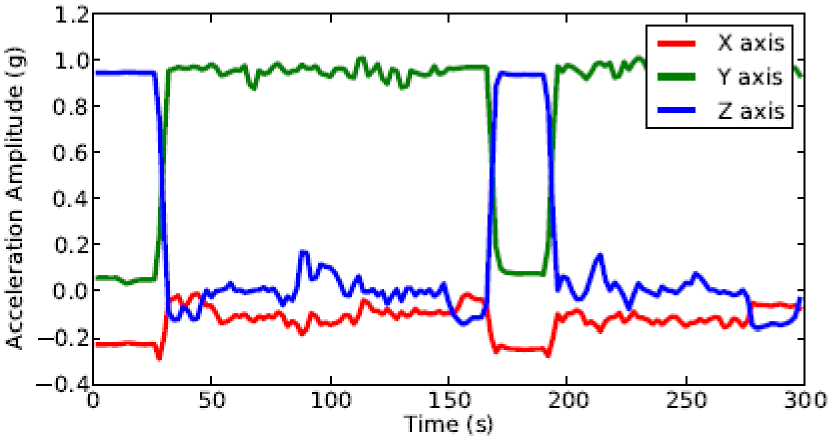
3.3.2.2 Computation and Transformation

For the classification-based activity recognition, feature extraction is an essential process. As a comparison, the selected features in the previous investigations are listed in Table 3-1. Among them, most frequently selected features are mean, standard deviation and the Fast Fourier Transform (FFT) energy. The choice of simple statistic features is due to the simplicity and low computational cost.

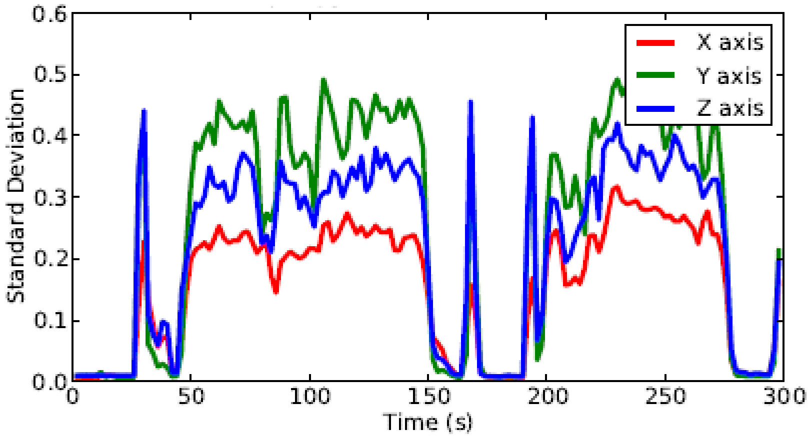
Table 3-1. Features selected and evaluated in previous investigations.

Research works	No. of Features	Features selected and evaluated
Mäntyjärvi et al. [101]	2	Principal Component Analysis (PCA), Independent Component Analysis (ICA)
Huynh and Schiele [102]	2	mean, variance
Van Laerhoven and Gellersen [103]	3	average, variance, peak set descriptors
Bao and Intille [3]	10	mean, variance or standard deviation, frequency domain energy, frequency domain entropy, correlation between axes
Ravi et al. [7]	4	mean, standard deviation, frequency domain energy, correlation between axes
Kern et al. [52]	2	mean, variance
Pärkkä et al. [92]	5	peak frequency, median, peak power, variance, sum of variances
Laerhoven and Cakmaci [54]	3	mean of the sum of maximum and standard deviation, zero-crossings, mean of the standard deviation

The FFT features are suitable to identify movements with distinguishable frequencies. Feature extraction involves the computation and transformation of the sliding windows into newer values. A simple example is to compute the mean value for every window. The result of this computation is a new series of values that represent mean of the original time series at a given interval. An example illustration is shown in Figure 3-10, where two features have been computed using mean and standard deviation for the given accelerometer data.



(a) Mean extracted from windows of 4 seconds



(b) Standard deviation extracted from windows of 4 seconds

Figure 3-10. Comparison between extracted mean values and extracted standard deviation values of same accelerometer raw data.

The corresponding features shown in the Figure 3-10 reflect the acceleration changes taken place in the top graph. For instance, the standard deviation values showed an increase and changes when the test user is in motion or when a transition takes place in his movements. The mean values have less significant differences between acceleration changes, but there are still noticeable patterns that can be used to recognize the movements. Transformations such as mean and standard deviation computations are relatively simpler features obtainable in the time domain. There are also transformations that attempt to extract information and patterns in another domain space such as frequency domain. This can be achieved by using the Fourier transform method to convert the time series to a representation in the frequency domain. Regardless of the transformation choices, the goal of feature extraction is to identify possibly enough different features that help the recognition system to better differentiate and detect the correct movements.

There are five features have been selected to be evaluated. Mean, variance and standard deviation are the selected simple statistic features. Besides that, two frequency domain FFT-based features, energy and information entropy have also been selected. The formulas for the selected features are listed in Table 3-2. The mean of the acceleration value for each axis represents the DC component of the acceleration data. The variance and standard deviation values are used to represent the range of acceleration differences from the mean, which may be representative if different movements demonstrate distinguishable ranges.

The FFT-based features (energy and information entropy) are chosen because the frequency domain characteristics in the acceleration may be usable to discriminate movements with different frequencies. The energy is calculated using the sum of the squared FFT component magnitudes of the acceleration of each axis and divided by the number of samples for normalization. Similarly, the information entropy of the discrete FFT component magnitudes of the acceleration values is also normalized. The information entropy may be used to support discrimination of movements with similar energy values [54].

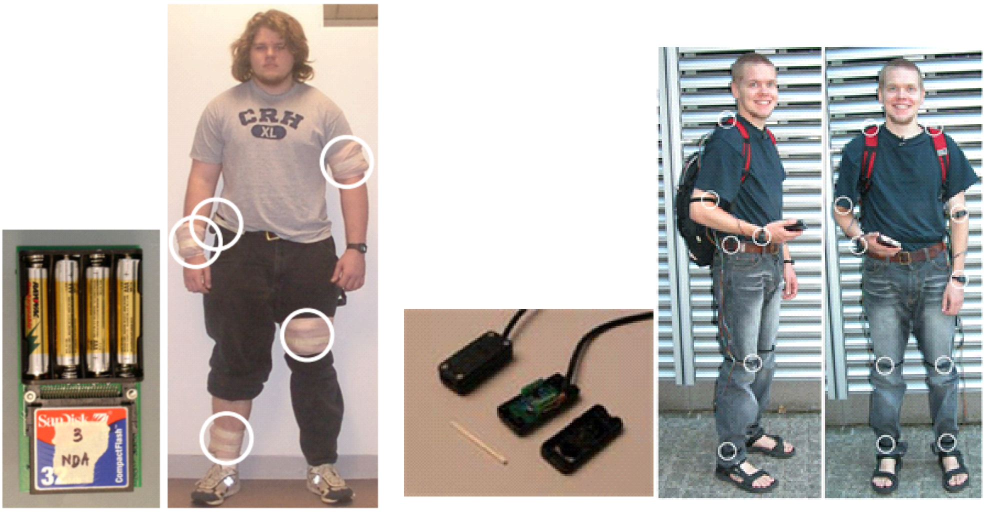
Table 3-2. Formulas for the computation of the selected features.

Features	Formula
mean	$\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i$
variance	$\sigma^2=\frac{1}{n}\sum_{i=1}^{n}x_i-\bar{x}$
standard deviation	$\sigma=\sqrt{\frac{1}{n}\sum_{i=1}^{n}x_i-\bar{x}}$
energy	$E=\frac{1}{n}\sum_{i=0}^{n}\lvert FFT_i\rvert^2$
information entropy	$I=-\frac{1}{\ln n}\sum_{i=0}^{n}\lvert FFT_i\rvert \bullet \ln(\lvert FFT_i\rvert)$

The total number of samples for each feature is relatively smaller than the original accelerometer data. For example, if the original data has 1000 samples and a window interval of 10 is selected, the total number of samples for each feature is 100 (with no overlapping) and 199 (with 50 % overlapping). Therefore, if the feature extraction process is not computational complex, this reduction of the total number of samples may help improve recognition speed, provided the number of selected features are also kept relatively low, and the features are suitable and applicable for the designated recognition.

3.4 Smartphone-based Activity Recognition

As the related work elaborated in the previous section 3.1 and 3.2, some of the popular activity recognition approaches have proposed the use of body-worn sensors. Different sensors are commonly placed at different parts of the body. The two following examples have selected to show how the placement of sensors and corresponding devices was proposed. Researchers in [3] used 5 sensor boards on different parts of the body such as arm, wrist, knee, ankle and waist. Each sensor board consists of a biaxial accelerometer, four AAA batteries and a memory card for storage. As shown in Figure 3-11, authors in [52] investigated activity recognition using 12 body-worn triaxial accelerometers. Both investigations have shown that accelerometer-based activity recognition can give up to around 90% accuracy. However, in order to enable the recognition of basic activities, the approaches suggest the use of a few sensors placed at fixed strategic positions depending on the targeted activities.



(a) (b)

Figure 3-11. Sensor replacements in two activity recognition researches using body-worn sensors [3], [52].

In [87], authors claim that such approaches are obtrusive for a person. In the investigation of [3], the authors mentioned that some of the experiment participants have reported that they felt self conscious in public spaces. This was because the sensor devices used were visually noticeable. Researchers in [7] stated that the placement of sensors in multiple predefined locations can be quite obtrusive. They contended this as a limitation for common activity recognition approaches using body-worn sensors [105]. There are also other people who have made similar observations on the same issue. Authors in [58] suggested the use of a single sensor placement as a less obtrusive alternative. Authors in [78] discussed on this issue by saying that wearable multi-sensor solutions are very obtrusive since wired techniques are used and needed most of the time and users have to strap sensors with Velcro strips or even wear special suits for the intended recognitions. Instead of placing different sensors on the person's body for continuous monitoring, we propose to use unobtrusive and minimum number of devices. The person should not consciously feel intruded or disturb with the number of sensors and the placement of the sensors.

A smartphone can be seen as a potential unobtrusive sensor device. Currently, most smartphones in the market have multiple built-in sensors, such as accelerometer, microphone, proximity, GPS, light and so on. Particularly in the case where the person already owns a smartphone, it is not necessary for him to use additional device for sensor data collection. If there is such a demand in the near future, additional external sensor devices can still be connected wirelessly to the smartphone for data collection, processing, transfer and even evaluation. The smartphone can be used as an alternative to current body-worn sensor devices based on the following factor:

① **The available sensors are built-in.** As long as the desired context can be derived and recognized from the data of the built-in sensors, the users are not required to use external sensors in order to collect needed information. In cases where the need occurs, additional sensors and devices can be interfaced to smartphones to extend necessary sensors other than the ones built-in. The flexibility and the readiness of the smartphone as a sensor device are seen as advantages.

② **The smartphones already have many properties that enable activity recognition related implementations**. Most smartphones have relatively high processing power and sufficient memory for data processing tasks. They also contain more than adequate storage space for the storage of raw and computed data. The smartphones also provide communication possibilities that allow information exchange between user and external services. The smartphone itself can be seen as a small computing device with common connectivity integrated.

③ **A smartphone is likely to be with a user during daily activities.** It can be seen as a natural choice of an unobtrusive device. The chances of users feeling awkward or uncomfortable will be much lower as compared to approaches that affect the usage habits of the users.

④ **Most smartphones have also relatively long operation durations.** For an average user, under normal usage patterns (some daily phone conversations and text messages), a smartphone should have at least a day's operation time before a recharge is required. With proper management for sensor data polling, whole day sensor data collection and processing may be achievable.

A smartphone plays five main roles for the tasks of activity recognition. It is a device that supports sensor, computing, storage, communication and user interaction. The available built-in physical and virtual sensors can be used to acquire information about the users and their environment. The acquired sensor data are stored for storage and further context processing steps. If necessary, the sensor information or the processed contexts can be sent to a remote server for further storage and processing. The communication channels offered by a smartphone range from short-range communication such as Bluetooth and Wireless Local Area Network (WLAN) to mobile broadband such as Universal Mobile Telecommunications System (UMTS).

The accelerometer of a smartphone is seen as a good candidate as sensor device for the desired activity recognition. It measures the acceleration of a person, which correlates to his movements. Many users usually carry a smartphone with them most of the time throughout a day [106]. If the measurable sensor information is able to be used to derive implicit information that reveal their contexts, such as their current activities and situations, the smartphone is an ideal all-in one device that complements the existing approaches.

3.5 Context Fusion

3.5.1 Fusion Techniques

Many well-developed algorithms can be applied to the competitive type of context fusion. The commonly used sensor fusion techniques for activity and context recognition are classical and Bayesian inference, voting, and fuzzy logic [115]. This section examines these commonly used sensor fusion methods in order to choose one as a module for fusing activity labels from both accelerometer and audio classifier.

3.5.1.1 Classical Sensor Fusion

The classical inference method and Bayesian inference network method are often referred as the classical or canonical sensor fusion methods because not only they are the most widely used, but also they are the bases of, or the starting points for, many new methods. Classical inference methods seek to judge the validity of a proposed hypothesis based on empirical probabilities. Given an assumed hypothesis H_i (a contextual fact is true or an event has happened), the joint probability P that an observation E_k would be reported by the sensors is:

$$P(E_k\, would\, be\, observed \mid H_k\, is\, true) \tag{1}$$

Many decision rules can be used to form the judgment in the classical inference method [116]. For example, the likelihood comparison rule suggests accepting the hypothesis H_i if the probability relationship satisfies Equation (2), otherwise, the system should believe that the contextual fact or event is not true or has not happened.

$$P(E_k|H_i)P(H_i) > P(E_k|\neg H_i)P(\neg H_i) \tag{2}$$

Another example of the decision rules is to use statistical significant test techniques. In the case where there are several alternative hypotheses, then the joint probability for each hypothesis needs to be computed and the results compared. The classical inference method quantitatively compares the probability that an observation can be attributed to a given assumed hypothesis. But it has the following major disadvantages [117]. (i) difficulty in obtaining the density functions that describe the observables used to classify the object, (ii) complexities that arise when multi variate data are encountered, (iii) its capability to assess only two hypotheses at a time, and (iv) its inability to take direct advantage of a priori likelihood probabilities. Bayesian inference overcomes some of these limitations by updating the likelihood of a hypothesis given a previous likelihood estimate and additional new observations. It is applicable when two or more hypotheses are to be assessed. Given the observed phenomena or evidence E, Bayesian inference calculates the likelihood $P(H_i|E)$ that the contextual fact or event H_i should be true or should have occurred in the form of [117]:

$$P(H_i|E) = \frac{P(E|H_i)P(H_i)}{\sum_j P(E|H_j)P(H_j)} \tag{3}$$

where, $P(H_i)$ is the a priori probability that the contextual fact or event H_i has occurred; $P(E|H_i)$ is the likelihood that the phenomenon or evidence E can be observed given the contextual fact or event H_i has occurred.

3.5.1.2 Voting Fusion

Voting sensor fusion imitates voting as a means for human decision-making. It combines detection and classification declarations from multiple sensors by treating each sensor's declaration as a vote, and the voting process may use majority or decision-tree rules. The most commonly used voting architecture is a Boolean combination of outputs from multiple sensors, where additional discrimination can be introduced via weighting each sensor's specific declaration [116,117].

The principle of the underlying mechanism of voting fusion is estimation of the joint detection probability based on the participating sensors detection confidence levels, which are in turn based on predetermined detection probabilities for an object or an event. Given that all sensors□observations are independent and non-nested, the probability that a hypothesis is true can be estimated as illustrated by the following example. For the proposition the context fact H_k is true or event H_k occurs, the inputs of voting fusion are the sensor s_i and s_j's detection probabilities $P_i(H_k)$ and $P_j(H_k)$, and their false alarm probabilities $P_{fai}(H_k)$ and $P_{faj}(H_k)$. The outputs of the voting algorithms are the detection probability $P(H_k)$ and the false alarm probability $P_{fa}(H_k)$, as shown in Equations (4) and (5).

$$P(H_k) = P_i(H_k) + P_j(H_k) - P_{i\cap j}(H_k) \tag{4}$$

$$P_{fa}(H_k) = P_{fai}(H_k) + P_{faj}(H_k) - P_{fai\cap j}(H_k) \tag{5}$$

The voting method greatly simplifies the sensor fusion process, and it can provide a prediction of object detection probability as well as false alarm probability. However, voting fusion is more suitable with 'yes/no' problems like the classical inference method. This granularity of reasoning, generally speaking, is not good enough for multiple status context discrimination, which is often required in context-aware computing applications. For a multiple status problem to be solved using the voting method, it has to be converted into multiple 'yes/no' problems first. Further, the more serious disadvantage inherent in the voting fusion method is that it treats each 'yes/no' problem separately rather than taking them as a whole package.

3.5.1.3 Fuzzy Logic Method

The fuzzy logic method accommodates imprecise states or variables. It provides tools to deal with context information that is not easily separated into discrete segments and is difficult to model with conventional mathematical or rule-based schemes. One example of such information kind is the room temperature: though it is commonly referred to with some descriptive words like 'cold', 'warm' or 'hot', it does not have hard boundaries between these states.

There are three primary elements in a fuzzy logic system, namely, fuzzy sets, membership functions, and production rules. Fuzzy sets consist of the imprecisely labeled groups of the input and output variables that characterize the fuzzy system, like the 'cold', 'warm' and 'hot' status in the above example of room temperature. Each fuzzy set has an associated membership function to provide a representation of its scope and boundaries. A variable of a fuzzy set takes on a membership value between the limits of 0 and 1, with 0 indicating the variable is not in that state and 1 indicating it is completely in that state. An intermediate membership value means a 'fuzzy' state, somewhat between the 'crisp' limits. A variable may belong to more than one fuzzy set. For example, a room temperature of 90°F may be regarded simultaneously as 0.25 'warm' and 0.65 'hot'.

Production rules specify logic inference in the form of IF-THEN statements, which are also often referred to as fuzzy associative memory. The basic algorithm is that the 'AND' operation returns the minimum value of its two arguments, and the 'OR' operation returns the maximum value of its two arguments. The output fuzzy set is defuzzified to convert the fuzzy values, represented by the logical products and consequent membership functions, into a fixed and discrete output that can be used by target applications.

Regarding human-users contextual information, there is a broad range of 'fuzzy' situations, where the boundaries between sets of values are not sharply defined, events occur only partially, or the specific mathematical equations that govern a process are not known. With its capability of dealing with this kind of information, and with its cheap computation to solve very complicated problems, the fuzzy logic method is expected to develop extensively in some context-aware computing applications.

The fuzzy logic sensor fusion method provides an effective tool to handle requirements of human daily-life, where imprecision is an inherent property in nature. However, the fuzzy logic sensor fusion method cannot be the main sensor fusion method in a generalizable architectural solution in building a context-aware computing system for two reasons. First, it is not applicable to situations where the objects inherently have clear-cut boundaries (e.g., it does not make sense to say, this is 0.6 person-A and 0.4 person-B). Second, the fuzzy set, membership function assignment, and production rules are usually extremely domain- and problem-specific, making it difficult to implement the method as a general approach.

3.5.1.4 Neural Network Method

Neural networks open a new door for fusing outputs from multiple sensors. A neural network can be thought of as a trainable non-linear black box suitable for solving problems that are generally ill defined and that otherwise require large amounts of computation power to solve. A neural network consists of an array of input nodes to accept sensors' output data, one or a few output nodes to show sensor fusion results, and sandwiched in between the input and output nodes is a network of interconnecting data paths. The weights along these data paths decide the input-output mapping behavior, and they can be adjusted to achieve desired behavior. This weight-adjusting process is called training, which is realized by using a large number of input-output pairs as examples.

The neural network training process can be simplified as follows. From the input nodes to output nodes, the data-path network provides many ways to combine inputs: those that lead to the desired output nodes are strengthened, whereas those that lead to undesired output nodes are weakened. Thus, after using the large number of input-output pair as training examples to adjust weights, the input data are more easily transferred to desired output nodes through the strengthened paths. The neural networks can work in a high-dimensional problem space and generate high-order nonlinear mapping. Many successful applications have been reported. However, it has some well-known drawbacks too. The three major problems are (i) it is difficult to select a network architecture that reflects the underlying physical nature of the particular applications; (ii) training a network is typically tedious and slow, and (iii) training can easily end up with local minima, as there is no indication whether the global minimum has been found [118].

The neural network method is not suitable for the main sensor fusion method mainly because of the drawbacks. First, the mapping mechanism is not well understood even if the network can provide the desired behavior - only in the simplest toy-like problems does examination of the weights in the trained network give any clue as to the underlying analytical connection between the inputs and outputs. Thus, such a solution cannot be easily generalized. Second, the neural network method is, generally speaking, not suitable to work in a dynamic sensor configuration environment, because each sensor needs a unique input node and each possible sensor-set configuration needs to be specifically trained. Third, the neural network sensor fusion method has the 'local minimum problem' during its training process, which cannot be easily overcome.

3.5.2 Fusion-based Classification Model

Local model assumes that each sensor node performs classification individually without communicating and cooperating with others. Figure 3-12 illustrates processing model of local classification, which consists of (i) a number of sensors providing input to the classifier, (ii) the classifier, which is responsible for activity recognition and determining the belogness of each instance to an activity class, and (iii) classification output, which is called activity. One should note that not all sensor nodes need to have the same classifiers.

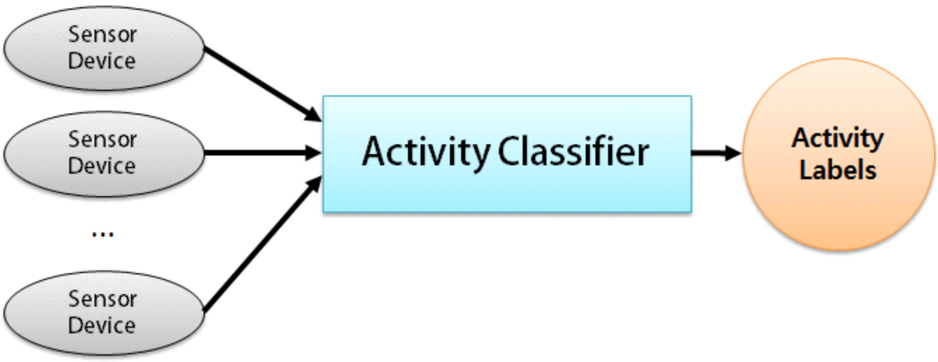


Figure 3-12. Local classification model.

The local approach is simple and works fine in situations, in which the sensor nodes are highly accurate and not prone to noises. However, generally speaking sensors, sensor nodes, and communication links are not always reliable and their failure is a common practice. Fusion-based classification model tolerates individual sensor and sensor node failures and involves more than one sensor node in the classification process. By doing so, it ensures that there are always some sensor nodes contributing to the classification process and compensating for the errors. The fusion-based approach uses the basic notions of the local approach and lets individual sensor nodes first classify and detect activities on their own.

Then, the classification results are all sent to a fusing / voter node (e.g., a cluster head) to reach a consensus. Figure 3-13 illustrates processing model of fusion-based classification. Similar to the local model, not all sensor nodes (including the fusing node) need to have the same classifiers. In this thesis we use fusion-based classification model for activity recognition.

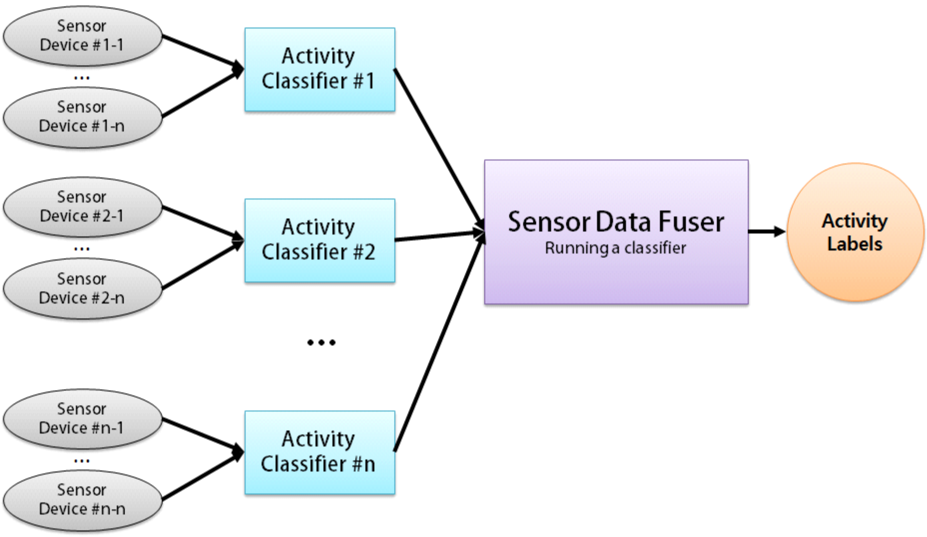


Figure 3-13. Fusion-based classification model.

Chapter 4. Comprehensive Activity Recognition Framework

In this thesis, the major contribution is to propose novel framework for smartphone which enable to utilize multimodal sensors on smartphone and support on-line processing as described in section 1.3. A comprehensive activity recognition framework is consisted of accelerometer classifier, audio classifier, heuristic-based result validation and user interface modules. Because of a proposed framework is conceptual architecture, specific algorithms for each steps in sub-modules are represented in this chapter.

For an accelerometer data processing, mixture model which is suitable for representing multiple distributions of collected data is chosen because of using multiple dimensions of features. Before modeling and classifying acceleration data, a prior processing including feature extraction and selection generates bunch of features to be used for a classification. In proposed framework, Gaussian Mixture Model (GMM) is used for the acceleration data classification. Also it fits to process mean and variance value [113]. Other classification techniques such as Gaussian Process is more appropriate for considering small number of variables or features. For the audio classification, Hidden Markov Model (HMM) algorithm is used for training and testing audio data because the module needs to be classify only two activities－bus and subway－and requires running on a smartphone in real-time. There are other audio classification algorithms such as Conditional Random Field (CRF) and Support Vector Machine (SVM), but proposed approach using HMM is lighter than other algorithms and it fits in classifying similar audio data both collected from bus and subway [114].

Overall architecture of the proposed Comprehensive Activity Recognizer (CAR) is described in Figure 4-1. CAR classifies several activities including ambulatory activities such as walking, jogging or still and transportation activities – bus and subway. Later section describes specific techniques for accelerometer classifier, audio classifier and heuristic based result validation. Also provides how two different classification results from accelerometer and audio classifier are combined for recognizing activities.

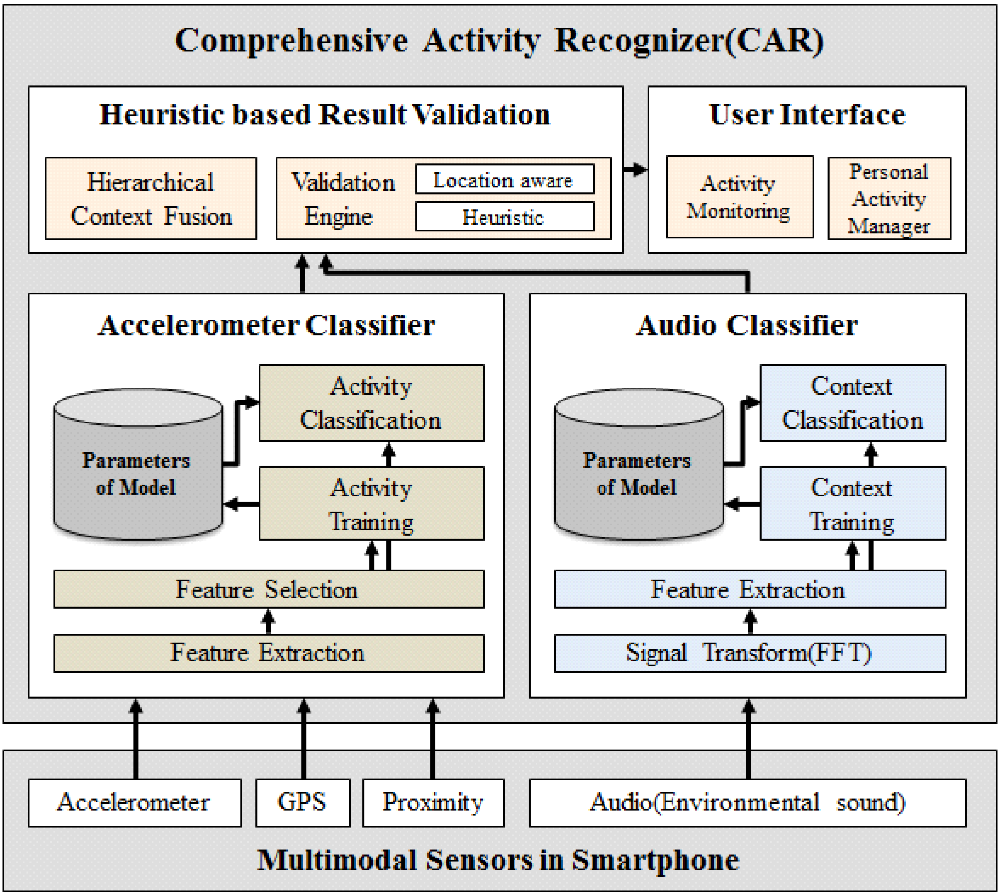


Figure 4-1. Overall architecture of the comprehensive activity recognition framework.

The system starts by recording three seconds of accelerometer data and then classifying those data into two categories:

- *Ambulatory activities－Walking, Jogging or Still*
- *Transportation*

For classifying ambulatory activities and transportation, acceleration data from accelerometer is utilized. If collected data has regular pattern such as walking, jogging and still, the system classify it as an ambulatory activity, but if it shows an irregular pattern, the output is regarded as transportation. If the output is ‘ambulatory activities’, the system determines whether the user is walking or jogging at a reasonable speed based on the speed information from the GPS interface. If the speed is reasonable or if a GPS signal is not available, the system outputs the final recognized context. Occasionally, a running bus may be mis-recognized as ‘walking’ or ‘jogging’ and in such a case, the speed validation module will redirect the next processing step to the ‘transportation’ branch. In the ‘transportation’ branch, the system first determines whether a transition point occurred (i.e., the previous recognized context was not ‘transportation’). Then, if a transition point did occur, the audio recorder will activate to record another three seconds of audio data. The system will then classify these three seconds of sound into three categories:

- *Bus*
- *Subway*
- *Others (all other sound that is not a bus or a subway)*

The result of the audio classifier can be further validated using a Wi-Fi pattern. More specifically, subway systems possess only a small number of well-known Wi-Fi services, and private wireless networks are nearly non-existent inside subways. In contrast, buses run on streets where private wireless networks from the passing buildings are abundant and often appear in the user range only to disappear a short time later. Consequently, these different Wi-Fi patterns can be used to validate the result of the audio classifier and avoid ambiguity in recognizing a bus and a subway.

Further validation can be done through the use of GPS readings, if available. For example, manually prerecorded the locations of all the subway stations in Seoul, which totaled around 100 stations. Hence, if a user approaches a subway, his latest location should be near a station (i.e., within a radius of 200 m). In short, the proposed system makes use of several sensors, including the accelerometer, audio tool, GPS, and Wi-Fi, and is able to recognize at least five different contexts:

- *User is walking*
- *User is jogging*
- *User is riding a bus*
- *User is riding a subway*
- *Other contexts (the context that is not one of the above four target contexts)*

The system mainly employs the accelerometer and audio recordings to classify the contexts. It uses extra information from the GPS and Wi-Fi systems to validate the results of the classification modules.

4.1 Accelerometer Classification

4.1.1 Feature Extraction and Selection

In proposed system, instead of using a single method, several kinds of well-known feature extraction techniques are utilized to construct a high number of features; then select the best features using own feature selection algorithms. The following features are considered:

- *Time domain features: standard deviation, mean crossing rate, Pearson correlation coefficients*
- *Frequency domain features* [3]
- *Linear Predictive Coding (LPC) features* [51]

Since there are a large number of features, using all of them may not increase the accuracy due to the problem known as 'the curse of dimensionality.' Consequently, it is necessary to select the best features from the extracted ones in order to construct a good feature set. The proposed method [53] measures the quality of a feature based on two criteria: the relevancy of the feature (or the classification power) and the redundancy of the feature (or the similarity between two selected features). These two criteria are computed from the mutual information of the feature as described in Equations (6) and (8):

$$Rel(X) = \frac{I(C;X)}{\log_2(|\Omega_C|)} \tag{6}$$

where X is a feature variable, C is a class variable, and Ω_C is the state space of C. Note that $I(C;X)$ is the mutual information between C and X, which can be calculated by:

$$I(C;X) = \sum_{C \in \Omega_C} \sum_{X \in \Omega_X} p(c,x) \log_2\left(\frac{p(c,x)}{p(c)p(x)}\right) \tag{7}$$

where Ω_X is the state space of the variable X; $p(c,x)$, $p(c)$, and $p(x)$ are, respectively, the joint and marginal probability distributions:

$$Red(X,Y) = \frac{I(X;Y)}{\log_2(|\Omega_X|)} \tag{8}$$

Algorithm 1. Feature Quantization.

```
Input:   M – Total number of features
         X(1..M) – Training data
         Δ – The quantization error
Output:  N – Number of quantization levels
         Y(1..M) – Quantized data
Quantization
    N = 2;
    while 1 do
        MaxError = −1e+16;
        for m = 1 to M do
            Upper = max(X(m));
            Lower = min(X(m));
            Step = (Upper – Lower) / N;
            Partition = [Lower : Step : Upper];
            CodeBook = [Lower – Step, Lower : Step : Upper];
            [Y(m), QError] = Quantiz(X(m), Partition, CodeBook);
            if QError > MaxError then
                MaxError = QError;
            end if
        end for
        if MaxError < Δ then
            break;
        end if
        N = N + 1;
    end while
end Quantization
```

Algorithm 2. Greedy Forward Searching for Feature Selection.

```
Input:   M – Total number of features
         N – Total number of data samples
         K – Number of features to be selected
         X – Training data matrix (M × N)
         C – Class labels (1 × N)
Output:  S – The index vector of the selected features (1 × K)
Forward
    S = Φ;
    for m = 1 to M do
        X_m = X_m − μ(X_m);
        X_m = X_m / σ(X_m);
    end for
    X = Quantiz(X);
    for k = 1 to K do
        BestScore = −1e+16;
        BestIndex = 0;
        for i = 1 to M do
            if X_i not in S then
                f = 0; c = 0;
                for X_j in S do
                    c = c + 1; f = f + Red(X_i, X_j);
                end for
                f = Rel(X_i) – f/c;
                if (f > BestScore) then
                    BestScore = f;
                    BestIndex = i;
                end if
            end if
        end for
        S = {S, BestIndex};
    end for
end Forward
```

In the above Equations (6) and (8), the mutual information can be computed by summing over the state space of the variable; therefore, the variables should be discretized before such a calculation can be performed. The discretization algorithm is illustrated in Algorithm 1. Once the relevance and the redundancy have been computed, well-known searching mechanism called 'greedy forwarding' is used to gradually extend the selection of features. The whole selection process is illustrated in Algorithm 2.

4.1.2 Gaussian Mixture Classifier

After extracting and selecting features, let us assume that X^C is a training data matrix *(N × K)* for class C, where each row is a training sample, and each column is a feature value. Gaussian mixture model (GMM) is used to determine the parametric probability density function of each class, denoted by $p(X^C|\lambda^C)$, where λ^C is the parameter set that includes the mixing weights and individual Gaussian mean vectors and covariance matrices:

$$p(X^C|\lambda^X) = \sum_{i=1}^{M} \omega_i N(X^C|\mu_i, \Sigma_i) \tag{9}$$

where N is a Gaussian distribution and is given by:

$$N(x|\mu_i, \Sigma_i) = \frac{1}{(2\pi)^{\frac{D}{2}} |\Sigma_i|^{\frac{1}{2}}} exp\left\{-\frac{1}{2}(x-\mu_i)' \Sigma_i^{-1}(x-\mu_i)\right\} \tag{10}$$

The mixing weights must satisfy the following condition:

$$\sum_{i=1}^{M} \omega_i = 1 \tag{11}$$

During the training phase, the parameters $\lambda^C = \{\omega, \mu, \Sigma\}$ are determined to maximize the training data likelihood $p(X^C|\lambda^C)$. In the inference phase, given all the class parameter sets λC1, λC2, ⋯ , λCm and an input vector x, the class label is determined by:

$$C = \mathrm{arg}max_C(p(x|\lambda^C)) \tag{12}$$

Figure 4-2 shows that the two different activities – walking and jogging – have different distribution of extracted features. In this example, among twelve features the classifier used, *standard deviation* and *mean crossing rate* of X-axis sensor data are used. The other ambulatory activities also classified by this methodology with twelve different features from raw sensor data – Mean, Standard deviation, Mean Crossing Rate of X, Y and Z axis sensor data, and even XY, YZ and ZX correlation data are employed.

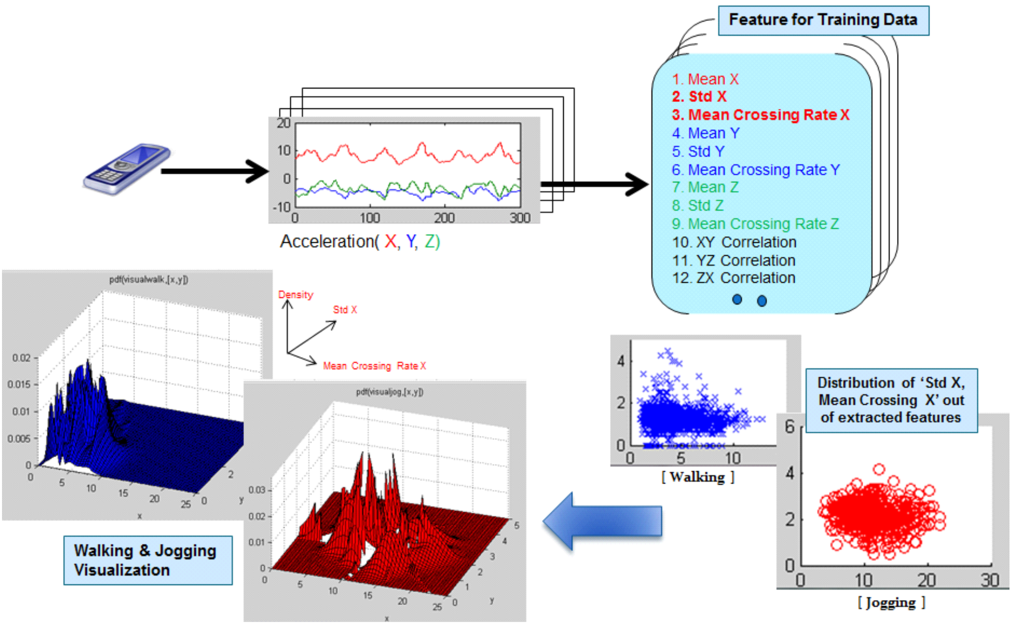


Figure 4-2. Classification visualization example of walking and jogging activities.

4.2 Audio Classification

For the audio classification module, collected audio signal is processed throughout several procedures. At first, analog signal formatted audio data is transformed to discrete data using Fast Fourier Transform (FFT). Then features are extracted from transformed data. MFCCs [5] features are utilized in proposed framework and the conventional classification method using the Hidden Markov Model (HMM) trains and classifies human activities such as riding a bus or subway. Figure 4-3 illustrates the audio classification module.

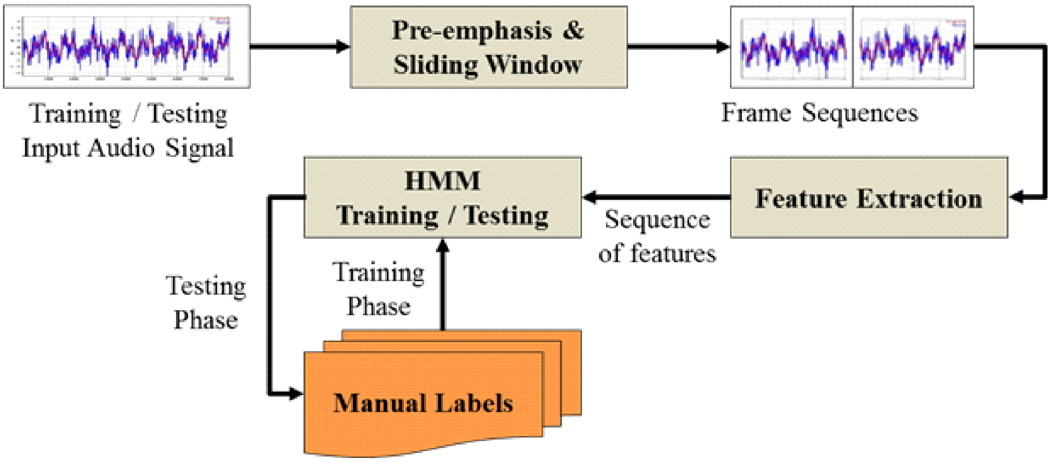


Figure 4-3. General sequence of the audio classification module.

4.2.1 MFCC Feature Extraction

Before the computation of MFCCs, a pre-emphasis filter is applied to the input audio signal *x(n)*, which eliminates the high frequencies:

$$x(n) = x(n) - 0.9x(n-1) \quad (13)$$

Next, the filtered signal is divided into shorter frames and multiplied with a

Hamming window function such that:

$$w(n) = 0.54 - 0.46\cos\left(\frac{2n\pi}{N-1}\right) \tag{14}$$

$$y(n) = w(n)x(n) \tag{15}$$

where N is the length of a window.

The feature extraction component then transforms the signal frames into the frequency domain using a discrete Fourier transform (DFT):

$$S(n) = DFT[y(n)] = R(n) + jI(n) \tag{16}$$

$$P(n) = |S(n)| = \sqrt{R^2(n) + I^2(n)} \tag{17}$$

where R and I are the real and imaginary parts of the Fourier transform respectively. The magnitude spectrum, *P(n)*, is then multiplied with Mel filter bands as follows:

$$P_{Mel}(m) = \sum_{n=0}^{\frac{N}{2}-1} H_m(n)P(n) \tag{18}$$

$$H_m(n) = \begin{Bmatrix} 0, f(n) < f_c(m-1) \\ \dfrac{f(n)-f_c(m-1)}{f_c(m)-f_c(m-1)}, f_c(m-1) \leq f(n) \leq f_c(m) \\ \dfrac{f(n)-f_c(m+1)}{f_c(m)-f_c(m+1)}, f_c(m) \leq f(n) \leq f_c(m+1) \\ 0, f_c(m+1) \leq f(n) \end{Bmatrix} \tag{19}$$

$$f_c(m) = 700(10^{\frac{\epsilon(m)}{2595}} - 1) \tag{20}$$

$$\epsilon = 2595\log_{10}(\frac{f}{700}+1) \quad (21)$$

The MFCCs are finally extracted by applying a discrete cosine transform to $P_{Mel}(m)$:

$$MFCC(k) = \sum_{m=0}^{M-1} P_{mel}(m)\cos\left(\frac{(m+0.5)k\pi}{M}\right) \quad (22)$$

where *M* is the number of Mel filters and *MFCC(k)* is the k^{th} coefficient.

4.2.2 Hidden Markov Model

A hidden Markov model (HMM) is a parametric model that determines the characteristics of data sequences. A HMM parameter set is defined as follows:

$$\Lambda = \{\pi, A, B\} \quad (23)$$

where π is a *1* × *N* vector containing the prior probability distribution of *N* states, *A* is a *N* × *N* transition probability matrix, and *B* is a set of *N* observation density functions. In this case, the continuous input where *B* was defined as:

$$B(i,x) = \sum_{m=1}^{M} \omega_m G(x, \mu_m, \Sigma_m) \quad (24)$$

where $i = 1,2,\cdots,N$ indicates the state index, *M* is the number of Gaussian components, ω_m is the mixing weight of the m^{th} Gaussian component, and $G(x,\mu_m,\Sigma_m)$ is a Gaussian density function with mean μ_m and covariance matrix Σ_m.

In the training phase of the HMM, given the input sequence $X = x_1, x_2, \dots, x_T$, the model parameters are updated to maximize the training likelihood $P(X|\Lambda)$. More details about the training algorithm can be found in [107]. After the training phase, each audio class has a corresponding HMM defined by the parameter sets $\Lambda^{Subway}, \Lambda^{Bus}, \Lambda^{Other}$. In the inference phase, given an input sequence $X = x_1, x_2, \ldots, x_T$, the likelihood of X given a HMM can be computed by:

$$P(X|\Lambda^C) = \sum_{h_1, h_1, \ldots, h_T} \pi(h_1) B(h_1, x_1) \prod_{t=2}^{T} A(h_{t-1}, h_t) B(h_t, x_t) \tag{25}$$

where $h_t (t = 1, 2, \ldots, T)$ is a hidden state value at time t and $h_t \in [1, 2, \ldots, N]$. The likelihood is calculated by using a forward or backward algorithm as described in [107]. Ultimately, the final class label is decided by:

$$Audio\ Class = Argmax_{C \in \{Bus, Subway, Other\}} P(X|\Lambda^C) \tag{26}$$

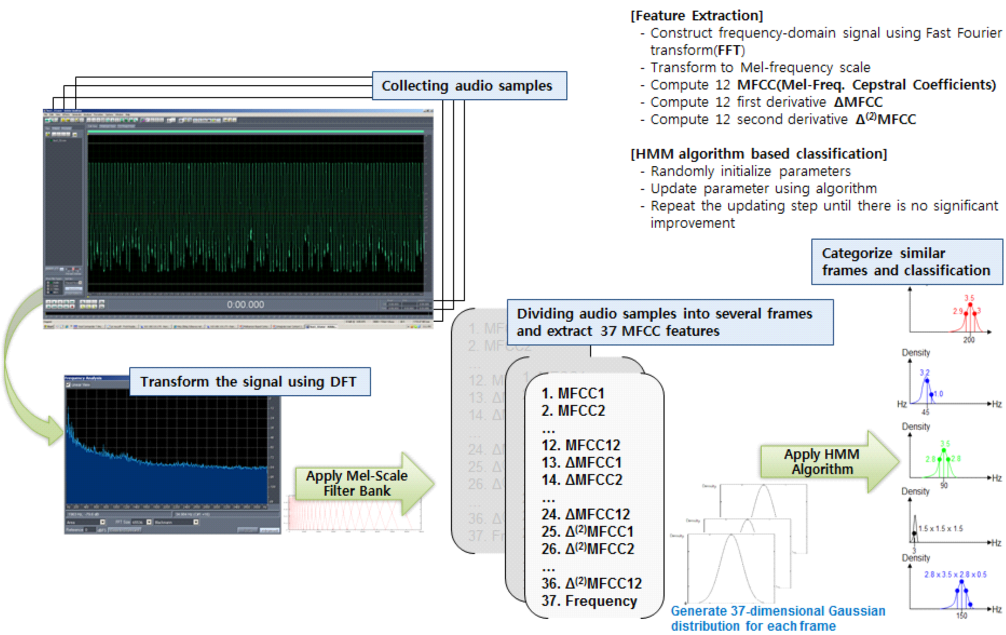


Figure 4-4. Overall process of audio classification.

Figure 4-4 shows the whole process of audio classification. Once the audio samples are collected, it is transformed from time domain data to frequency domain using the Fast Fourier Transform. Then 37 MFCC features are extracted from each audio samples for training and classification. Proposed audio classifier utilizes HMM algorithm for classification. The reason why HMM algorithm is chosen for training and testing audio data is that the module needs to be classify only two activities－bus and subway－and requires running on a smartphone in real-time. Also HMM is relatively lighter than other audio classification algorithms such as CRF and SVM.

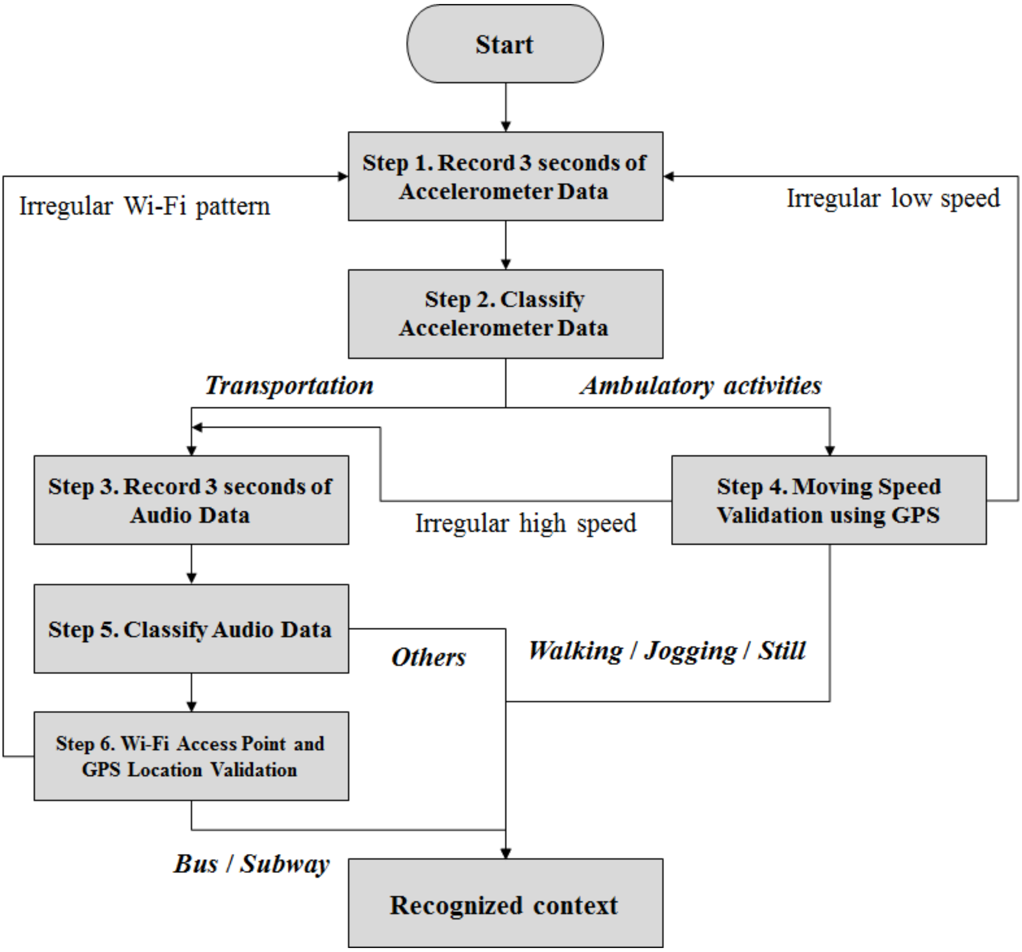


Figure 4-5. Overall process of the context fusion.

4.3 Hierarchical Context Fusion

In both accelerometer and audio classifier, activity labels of human's behavior and context are generated using GMM and HMM algorithm-based classification processes. Once these information are generated, they are delivered to the heuristic-based result validation sub-module. Main role of the module is to combine both activity labels from accelerometer and audio classifier for making final decision. Figure 4-5 represents the overall process to recognize human's activity information using hierarchical approach. For combining each outputs of accelerometer and audio classifier, we have applied several rules for fusing contexts. In this chapter, the steps and rules for fusing outputs of two classifier are introduced.

Step 1. Record 3 seconds of Accelerometer Data - When the recognition process begins, the system records 3-axis accelerometer data for 3 seconds. Collected raw sensor data will be delivered to the accelerometer classifier for recognizing ambulatory activities.

Step 2. Classify Accelerometer Data – As described in section 4.1, some features for classification are extracted from collected raw sensor data. Then the module decides that the activity is ambulatory or not. If the result of the accelerometer classification is a 'Transportation' activity, the process goes to the step 3. But if the result is an 'Ambulatory activities', it goes to the step 4 by applying Rule 1.

```
Rule 1.
if result label == 'transportation', then go to step 3;
else if result label ∈ ambulatory activities(Walking, Jogging,
Still), then go to step 4;
```

Step 3. Record 3 seconds of Audio Data - When the result of accelerometer classification is transportation, the system records environmental sound for 3 seconds. Collected raw audio data will be delivered to the audio classifier for recognizing transportation activities.

Step 4. Moving Speed Validation using GPS – If the classification result is one of the ambulatory activities, the systems validates the result with speed information using GPS. This step checks the current moving speed and confirm current activity label among 3 activities(Walking, Jogging and Still). If moving speed is irregularly high, the process goes to the step 3, but if it is too low, it goes to the step 1 by applying Rule 2.

```
Rule 2.
if current speed >= 3 km/h and current speed <= 15 km/h, then
confirm current activities;
else if current speed < 3 km/h, then go to step 3;
else if current speed > 15 km/h, then go to step 1;
```

Step 5. Classify Audio Data – As described in section 4.2, if the current activity is ‘transportation’ the systems tries to recognize what transportation method the use is currently riding by using environmental audio data. If the result of the audio classification is an ‘Others’ activity, it confirms current activity label. But if the result is one of the ‘Transportation’ activities, it goes to the step 6 by applying Rule 3.

```
Rule 3.
if result label == ‘others’, then confirm current activity;
else if result label ∈ transportation activities(Bus, Subway),
then go to step 6;
```

Step 6. Wi-Fi Access Point and GPS Location Validations – If the classification result is one of the transportation activities, the systems validates the result with registered Wi-Fi Access Point information and location information using GPS. This step checks the current location is in the registered locations. If Wi-Fi pattern is not matched to the registered patterns, the process goes back to the step 1, but if not it confirms current activity label.

> **Rule 4.**
> if *current AP name* ∈ *registered_Wi-Fi_AP_list*, or *current location* ∈ *registered_GPS_location_list*, then confirm current activities;
> else if *current AP name* ∉ *registered_Wi-Fi_AP_list*, then go to step 1;

Chapter 5. Lightweight Activity Recognition Framework

As described in Chapter 4, the proposed comprehensive activity recognizer collects activity data from multimodal sensors and classifies five activities. But the proposed framework is a conceptual architecture so various classification or extraction algorithms can be selected depending on the target system. In this chapter, lightweight activity recognition framework which activity modeling and classification are run on same platform for mobile devices such as smartphone is proposed.

In section 3.4, several advantages of smartphone-based activity recognition are described. The first one is an availability of various sensor devices embedded in. So the user are not required to use external sensors in order to collect needed information. The flexibility and readiness of the smartphone as a sensor device are also seen as advantages. The second one is that the smartphones already have many properties that enable activity recognition related implementations. it provides processing capability, adequate storage space and communication ability etc. The third one is that a smartphone is likely to be with a user during daily activities. So the system easily acquires activity and context data from users in unobtrusive way. The last one is that the most smartphones have also relatively long operation durations. In order to recognize whole day activities of users, a smartphone should have at least a day's operation time before a recharge is required. Current smartphone has already sufficient reliability of operating system and battery power.

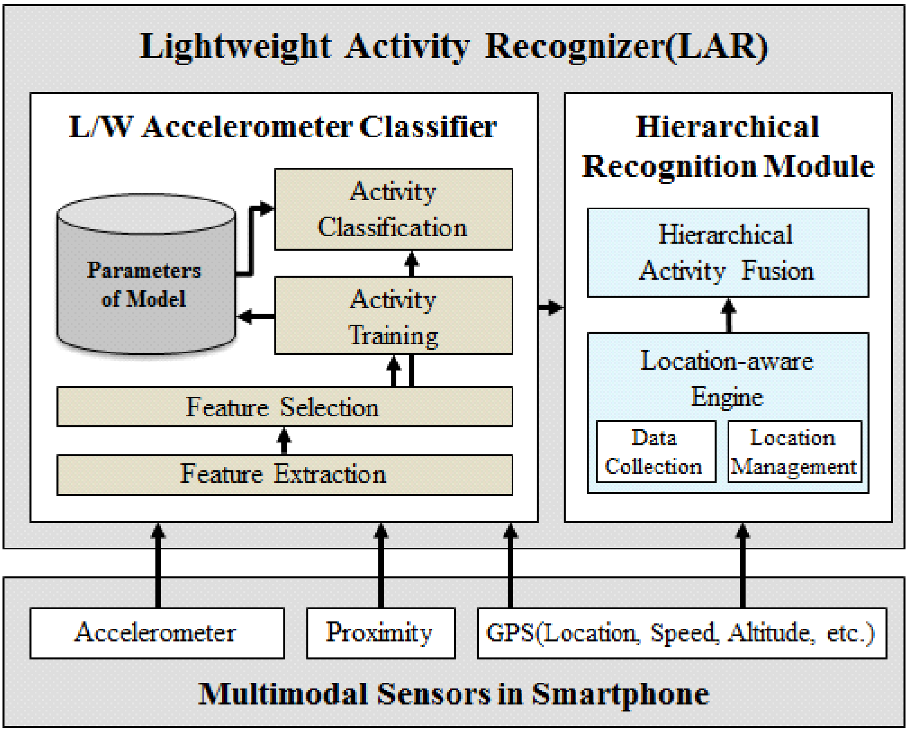


Figure 5-1. Overall architecture of proposed lightweight activity recognition framework.

For the lightweight activity classification framework, Naïve Bayes algorithm is used as a basic algorithm for recognizing human's activities. If the activity information of users is matched to the candidate which has highest possibilities among pre-constructed activities models, the one is chosen by the algorithm. And the algorithm must be lightweight and fast especially if the target platform is a mobile devices such as smartphone. There is a comparison of several classification methodologies in [108]. It shows that Naïve Bayes classifier achieves the fastest modeling time than other machine learning algorithms.

Although the Naïve Bayes classifier generates activity model fast, it has several limitation such as relatively low processing speed, hard to apply into mobile environment which has less resources. First, in modeling phase - extracting features from collected data – memory error can be happened because of insufficient memory allocation in mobile device. Second, one of the inherent characteristics of Naïve Bayes classifier, every attribute have same priority, causes lower accuracy of posterior probability. So in order to resolve above problems, Adaptive Naïve Bayes (A-NB) is proposed.

5.1 Naïve Bayes Classifier

Naïve Bayes based on Bayes theorem is a statistical classification method which can estimate the possibility of a given sample. Compare to existing machine learning algorithm − Decision Tree or NN − it results higher accuracy and speed on large-scale database. Also it processes fast and requires relatively low resources than HMM, GMM and SVM [109]. Naïve Bayes probabilistic model assumes that sample data F_1 to F_n are included to independent class C. The probability of C after the sample data $F_1 \ldots F_n$ are collected is $p(C|F_1, \ldots, F_n)$ and it called a posteriori probability. In order to calculate $p(C|F_1, \ldots, F_n)$, $p(F_1, \ldots, F_n)$ and $p(C)$ are required. These can be estimated by trained data and it called boundary probability. By using Bayes' theorem a posteriori is defined as below:

$$p(C|F_1, \ldots, F_n) = \frac{p(C)\,p(F_1, \ldots, F_n|C)}{p(F_1, \ldots, F_n)} \tag{27}$$

Only considering a maximization of $p(C)p(F_1, \ldots, F_n|C)$ because $p(F_1, \ldots, F_n)$ has equal values to every class. If it is not able to know the boundary probability of the class, only $p(F_1, \ldots, F_n|C)$ might be considered. $p(F_1, \ldots, F_n|C)$ is calculated by independent assumption of Naïve Bayes. As a result, $F_1 \ldots F_n$ can be classified as the class which has the biggest posteriori probability. So the posteriori probability is defined as below:

$$V_{\max} = p(C_j|F_1, \ldots, F_n) = p(C)\prod_{i=1}^{n} p(F_i|C) \tag{28}$$

In above equation, C is a probability of given class is class i, F_i is a probability of *ith* attribute, $p(C)$ is a probability of class C among whole class and $p(F_i|C)$ is a probability of specific attribute of class C. So the maximum posteriori probability $V_{\max}$ is selected after calculating each classes.

By the way, if a sample data F_i is a classification attribute contains one value out of several limited values, a calculation of $p(F_i|C)$ is easy by traditional probability equation. But the training data for activity recognition is mostly continuos value. In this case, distribution of probability is utilized for calculating conditional probability. If the mean value of F_i in class C is μ_c and distribution is σ_c^2, gaussian distribution is utilized for representing a distribution of F_i.

$$P(F_i = v|C) = \frac{1}{\sqrt{2\pi\sigma_C^2}} e^{-\frac{(v-\mu_c)^2}{2\sigma_C^2}} \tag{29}$$

5.2 Adaptive Naïve Bayes Algorithm

A lightweight activity recognition framework utilized multiple sensor data which embedded in a smartphone. So the lightweight modeling and recognition algorithm are required because of the limitation of resources. Also the Naïve Bayes performs relatively low accuracy because of its inherent characteristics. So Naïve Bayes based lightweight classification algorithm A-NB which enables activity modeling and recognition in a Smartphone is proposed.

When building activity model using a Naïve Bayes, complexity of calculation is dependant on a number of sample data i. If considering factors are increased, data overhead has happened while calculating mean value μ_c and distribution σ_c^2 of data F_i. Upon the processing environment, memory overflow could be happened in mobile environment. It causes not only activity recognition performance but also total systems delay.

In order to resolve the problem happened in mobile environment, A-NB algorithm is proposed. For overcoming memory overflow which can happened during real-time activity training, A-NB calculate the mean and distribution values of data F_i periodically. While total training time, if the number of calculated mean and distribution is j, F_i has the matrix below:

$$\left\{(\mu_1, \sigma_1^2), (\mu_2, \sigma_2^2), (\mu_3, \sigma_3^2), \dots, (\mu_j, \sigma_j^2)\right\} \tag{30}$$

where j is a number of time slices. Proposed A-NB calculate total mean value by combining μ_1 to μ_j, total distribution value by mean value of σ_1^2 to σ_j^2.

$$V'_{\max} = p\left(C_j | F_1, \dots, F_n\right) = p(C) \prod_{i=1}^{n} p'(F_i | C) \tag{31}$$

In order to calculate $p'(F_i|C)$ value, equation (29) is transformed to equation (32)

$$P'(F_i = v | C) = \frac{1}{\sqrt{2\pi\mu_v}} e^{-\frac{(v-\mu_m)^2}{2\mu_v}} \tag{32}$$

μ_m is a mean value of $\mu_1 \cdots \mu_j$, μ_v is also mean value of distribution $\sigma_1^2 \cdots \sigma_j^2$. So the mean and distribution are calculated by data sample F_i, and by using these values posteriori probability $P'(F_i = v|C)$ and $V'_{\max}$ are able to calculated.

5.3 Hierarchical Activity Recognition Framework

Although the activity recognition using multimodal sensors can increase recognizable activities and enable to recognize various situations, it lowers the accuracy of recognition result because of the classifier is required to consider more factors from input data. In order to overcome above weakness, HARF which recognize activities in hierarchical approach has been proposed. Life log includes not only recognition of a simple act, but also visit a specific location and boarding of the vehicle etc. Also considering that there may have different meanings depending on the location, activities can be classified as 3 types. Table 5-1 shows each activities in different categories and sensors required to recognize them.

Table 5-1. Activity categorization for hierarchical activity recognition.

Type	Area	Activity	Sensors
Location & Multimodal sensor based activity recognition	Home	Walking	Accelerometer, Gyroscope, Proximity and GPS
		Sitting	
		Standing	
	Office	Walking	
		Sitting	
		Standing	
	Outdoor	Walking	
		Sitting	
		Standing	
		Jogging	
Location based activity recognition	Outdoor	Waiting bus at bus stop	GPS
		Having meal at cafeteria	
		Exercising at gym	
		Visiting a park	
Heuristic based activity recognition	Outdoor	Riding a car	Accelerometer, Gyroscope, Proximity, GPS and Heuristic Rule

5.4 Hierarchical Activity Fusion

In both L/W accelerometer classifier and location-aware engine, activity labels of human's behavior and current location are generated. Once these information are delivered to the activity fusion sub-module, it tries to combine both information for making final decision.

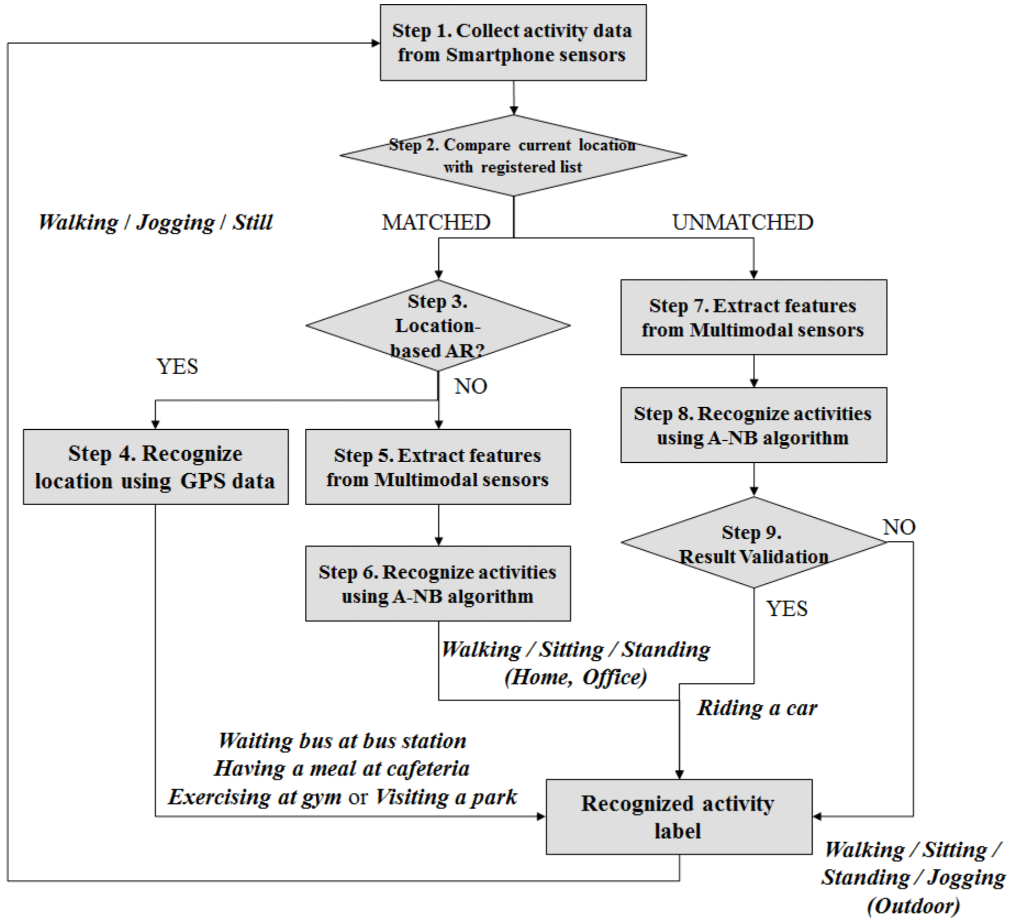


Figure 5-2. Overall process of the activity fusion.

Figure 5-2 represents the overall process to recognize human's activity information using hierarchical approach. It shows the real-time activity recognition process based on tha A-NB algorithm. If the A-NB is applied to activity recognition, classification is performed using location information first and heuristic approach is applied as described in Table 5-1. Once recognition is performing, system recognize the location first for differentiating indoor(Home and Office) and outdoor. But the user is at unregistered location, the system only uses physical sensor data. In this chapter, the steps and rules for fusing outputs of two classifier – L/W accelerometer classifier and location aware engine - are introduced.

Step 1. Collect activity data from smartphone sensors - When the recognition process begins, the system collects activity data from 3-axis accelerometer, 3-axis gyroscope, GPS and proximity sensor. Collected raw sensor data will be delivered to the accelerometer classifier and location-aware engine for recognizing current activities.

Step 2. Compare current location with registered list – As described in section 5.3, the system compares the current location with registered location list(Home, Office) first. If the current location is in registered location lists, the process goes to the step 3. But if the current location is not in the list, it goes to the step 8 by applying Rule 5.

```
Rule 5.
if current location ∈ registered_list, then go to step 3;
else if current location ∉ registered_list, then go to step 8;
```

Step 3. Location-based AR – If the current location is in registered locationion list, the system differentiate location-based in this step. If the current location is in the registered list and location-aware AR, the process goes to the step 4. But if the current location is one of the ‘Home’ or ‘Office’, it goes to the Step 5 by applying Rule 6.

```
Rule 6.
if current location ∈ Location-based AR, then go to step 4;
else if current location == 'Home' or current location == 'Office', then go to step 5;
```

Step 4. Recognize location using GPS data – If the current location is one od the location-based AR, the system confirms current activity among registered outdoor activity lists - Waiting bus at bus station, Having a meal at cafeteria, Exercising at gym or Visiting a park – by applying Rule 7.

```
Rule 7.
if current location ∈ outdoor_registered_list, then confirm current activities(Waiting bus at bus station, Having a meal at cafeteria, Exercising at gym or Visiting a park);
```

Step 5. Extract features from Multimodal Sensors – In this step, the system extracts some features for a recognition from collected raw sensor data. Mean, Standard Deviation feature are utilized for A-NB algorithm.

Step 6. Recognize activities using A-NB algorithm – As described in section 5-2, Naïve Bayes based recognition algorithm is used for recognizing activities in real-time manner. If the classification result is one of the physical movement activities in both 'Home' or 'Office', the systems confirms activity label among 'Walkinh', 'Sitting' or 'Standing' by applying Rule 8.

```
Rule 8.
if result label ∈ indoor activities, then confirm current activities(Walking, Jogging or Standing);
```

Step 7. Extract features from Multimodal Sensors – In this step, the system extracts some features for a recognition from collected raw sensor data. Mean, Standard Deviation feature are utilized for A-NB algorithm.

Step 8. Recognize activities using A-NB algorithm – As described in section 5-2, Naïve Bayes based recognition algorithm is used for recognizing activities in real-time manner. If the classification result is one of the physical movement activities in 'Outdoor', the process goes to the step 9 for validation.

Step 9. Result Validation – If the classification result is one of the outdoor activities, the system validates whether the current activity is 'Riding a car' activity of not using moving speed from GPS sensor. So if the current moving speed is over 25 km/h, the system confirms the activity label. But if not, it confirms the activity label among 4 outdoor activities – Walking, Sitting, Standing and Jogging – by applying Rule 9.

```
Rule 9.
if moving speed >= 25 km/h, then confirm current activity as
'Riding a car';
else if moving speed < 25 km/h, then confirm current
activities(Walking, Sitting, Standing or Jogging);
```

5.5 Implementation of the A-NB based HARF

For an evaluating the proposed algorithm in smartphone environment, a real-time activity training and recognition system called pARNL (Personalized Activity Recognizer and Logger) is proposed. Implementation using the pARNL enables personalized activity recognition and life-logging and users can add or monitoring their own activities of contexts. pARNL is described as Figure 5-3.

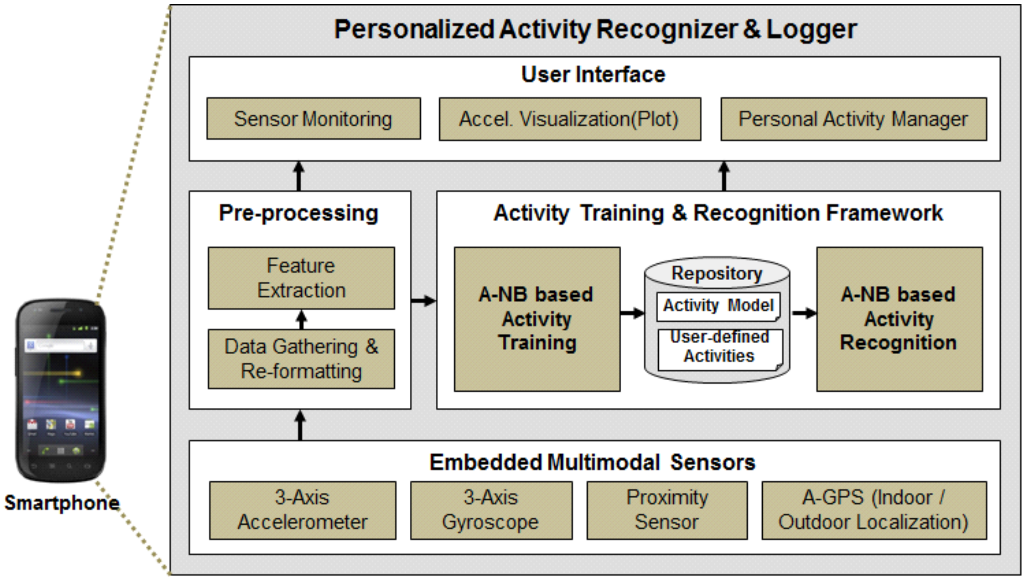


Figure 5-3. HARF based real-time activity recognition system.

- Embedded Multimodal Sensors: Consisted of 4 sensors (3-axis accelerometer, 3-axis gyroscope, proximity sensor and GPS) embedded in smartphone for recognizing activities.

- Pre-processing: Collecting sensing data from multi-modal sensors periodically (50Hz) and unifying data format for efficient data processing sharing. Also extracting features from collected sensor data for activity modeling and recognition.

- Activity Training & Recognition: Composed of proposed A-NB algorithm based activity training and recognition modules. Activity models and user-defined activities are stored at the Repository

- User Interface: Let the users to monitor collecting data from multimodal sensors and visualize 3-axis accelerometer data for the users. Also provide an interface for users to add or train their own activities.

Chapter 6.

Implementation and Results

6.1 Comprehensive Activity Recognizer

To evaluate proposed comprehensive activity recognizer, experiments with the accelerometer and audio classification are inducted independently. As described in chapter 4, the proposed system classifies activities into four contexts first, and then if the system identifies a 'transportation' mode, it starts to collect audio data to determine whether this transportation is via bus or subway. Next, an integrated system that combined the accelerometer and audio classifiers is evaluated. For the evaluation and testing, over 10,000 data samples are collected from 10 volunteer graduate students during a month-long period at various locations.

Also for achieving position-free approach, volunteers are allowed to hold a smartphone at anywhere on their body such as attach it on waist, put it in trousers' pocket or just hold it by hands. After collecting sensor data from all volunteers, categorized them into each activity types based on activity label. Then activity model of each activity labels—walking, jogging, still, bus (run, jam, stop) and subway (run, stop)—by GMM-based modeling and classification module in the accelerometer classifier are constructed. As noted previously, the proposed system utilizes sensor data which is collected previous 3 seconds for real-time processing. It means the system does not use previous contexts for recognition processing. The approaches described above enable position-free recognition. The smartphones used for evaluation are Android HTC Desire smartphones, Samsung Galaxy S smartphones, and Samsung Galaxy S II smartphones.

Table 6-1. Types and number of Collected sensor data for evaluation.

Activity Type	Sensors used	Data format	No. of samples
Walking	Accelerometer, GPS, Wi-Fi	Text	1244
	Audio	Raw	1244
Jogging	Accelerometer, GPS, Wi-Fi	Text	591
	Audio	Raw	591
Bus	Accelerometer, GPS, Wi-Fi	Text	4645
	Audio	Raw	13023
Subway	Accelerometer, GPS, Wi-Fi	Text	3864
	Audio	Raw	3387
Car	Accelerometer, GPS, Wi-Fi	Text	955
	Audio	Raw	2829
Others	Accelerometer, GPS, Wi-Fi	Text	3106
	Audio	Raw	5472
Total number of samples (Accelerometer etc.)			**14405**
Total number of samples (Audio)			**26546**

6.1.1 Accelerometer Classification

In order to validate the accelerometer classification module, acceleration data is collected in four contexts: walking, jogging, transportation (bus and subway), and still. An assortment of features is investigated, including frequency, time, and LPC features. To combine the strength of different feature extraction methods, proposed feature selection algorithm is used to select the best candidate from a large set of features extracted by the existing method. Table 6-2 and Figure 6-1 show the 10-fold cross validation test results for different features. Table 6-3 shows which features were selected from the features generated by the existing feature extraction methods using proposed feature selection algorithm.

Table 6-2. Accelerometer classification accuracy with different features.

	Frequency Features	Time Features	LPC Features	Selected Features
Fold1	87.16	87.07	89.75	**90.71**
Fold2	87.23	86.41	89.89	**90.13**
Fold3	86.55	88.26	89.24	**91.01**
Fold4	86.49	87.89	89.30	**89.82**
Fold5	86.93	87.97	88.03	**89.97**
Fold6	88.18	88.70	87.98	**90.78**
Fold7	87.08	88.78	88.27	**89.68**
Fold8	86.19	89.09	90.04	**90.41**
Fold9	86.41	86.48	89.00	**89.90**
Fold10	86.47	89.24	89.83	**90.34**
Average	86.87	87.99	89.13	**90.27**

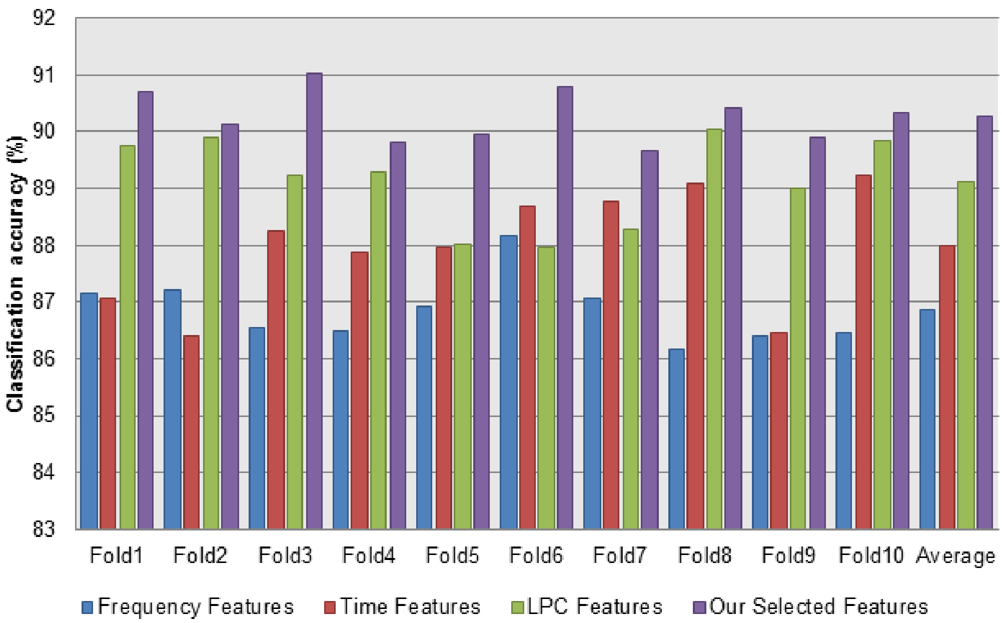


Figure 6-1. Accelerometer classification accuracy comparison based on Table 6-2.

Table 6-3. Selected features from extracted by existing feature extraction methods.

	Features	Selected (X = yes, O = no)
Frequency Features	Over spectral energy	X
	Spectral sub-band 1 energy	X
	Spectral sub-band 2 energy	X
	Spectral sub-band 3 energy	O
	Spectral sub-band 4 energy	O
	Spectral sub-band 5 energy	O
	Spectral sub-band 6 energy	O
	Spectral sub-band 7 energy	O
	Spectral sub-band 8 energy	O
Linear Predictive Coding(LPC) Features	LPC coefficient 1	X
	LPC coefficient 2	X
	LPC coefficient 3	O
	LPC coefficient 4	O
	LPC coefficient 5	O
	LPC coefficient 6	X
	LPC estimation error	X
Time Domain Features	Mean value	O
	Standard deviation value	X
	Mean crossing rate	X
	XY correlation	X
	YZ correlation	O
	ZX correlation	O

6.1.2 Audio Classification

The dataset used to evaluate the audio classification was collected and provided by the School of Computing Sciences, University of East Anglia, UK, and is available in [110]. This dataset contained WAV formed audio files (sampling rates: 8 kHz, 8 bit, mono) taken using a Samsung YP55H MP3 recorder in 2004. It had twelve different audio files, but seven different contexts were used: Building Site, Bus, Car (city), Supermarket, Office, Presentation and Street (traffic). Table 6-4 shows the confusion matrix of the classification measured using a k-fold (k = 10) cross-validation rule.

Table 6-4. Accuracy table of audio classification confusion matrix (Ma, L. [112] Dataset).

	Building Site	**Bus**	**Car (City)**	**Supermarket**	**Office**	**Presentation**	**Street Traffic**	**Total**
Building Site	**100%**	-	-	-	-	-	-	100%
Bus	-	**100%**	-	-	-	-	-	100%
Car	-	4%	**95%**	1%	-	-	-	100%
Supermarket	-	-	-	**100%**	-	-	-	100%
Office	-	-	-	-	**100%**	-	-	100%
Presentation	-	-	-	-	-	**99%**	1%	100%
Street	-	-	-	1%	1%	10%	**88%**	**100%**

The average accuracy of proposed audio classification system was about 97.43%. In addition, audio dataset were manually collected for three contexts－bus, subway, and other (anything except bus and subway)－using various Android smartphones. Using a k-fold (k = 10) cross-validation rule, the accuracy is acquired shown in

Table 6-5.

Table 6-5. Accuracy table of audio classification using own dataset.

	Bus	Subway	Other
Bus	89.34%	5.60%	10.66%
Subway	4.25%	91.20%	4.55%
Other	4%	4%	92%

These results present a reasonably high accuracy level, suggesting that audio is an important data source for proposed context-aware system.

6.1.3 Performance Evaluation of the Integrated System

After validating the individual classification module, accelerometer and audio classifiers were combined into one integrated system, with extra information acquired from the GPS and Wi-Fi schemes. The integrated system was tested on the field with realistic and real-time sensory data. More specifically, a user launched the system via a smartphone, and as this user moved－e.g., riding a bus or subway－an observer recorded all of the ground truth labels by hand while the system wrote the recognized labels to a log file. After the test, the recognized labels were compared with the hand-recorded truth tables.

As described in Table 6-6, eight different recognizable activities are collected and tested. Three of them are ambulatory activities and the rest of them are transportation activities. Especially riding a bus has another situation ‘Jam’ which might be occurred when a bus is stopped by traffic signal or bad traffic condition. Table 6-6 shows a confusion matrix of different contexts. Over a thousand activities for each context were collected and used. Figure 6-2 is a comparison graph of the true positive with the false negative of each activity, which highlights the accuracy of the recognized labels.

The results of the audio classification shows that, by selecting the good features from different feature sets, it significantly improve the classification accuracy. To validate the significance of the difference between the achievements (when comparing the recognition results of selected feature set with those of the other feature sets), the paired t-test is used to calculate the p-values, which were always smaller than 0.05 (note that a p-value < 0.05 indicates that the achievements are significantly different from a statistical point of view).

Table 6-6. Evaluation of the integrated system with realistic and real-time data.

	Ambulatory Activities			Bus			Subway		Total Samples
	Walk	Jogging	Still	Run	Jam	Stop	Run	Stop	
Walk	**1109**	36	48	-	-	-	-	-	**1193**
Jogging	25	**767**	42	-	-	-	-	-	**834**
Still	-	-	**1915**	-	-	-	20	60	**1995**
Bus(run)	65	86	-	**2000**	-	-	-	-	**2151**
Bus(jam)	-	-	52	-	**782**	-	-	35	**869**
Bus(stop)	-	-	16	-	-	**279**	-	19	**314**
Subway (run)	-	-	24	-	49	-	**2341**	-	**2414**
Subway (stop)	-	-	18	-	11	7	-	**314**	**350**

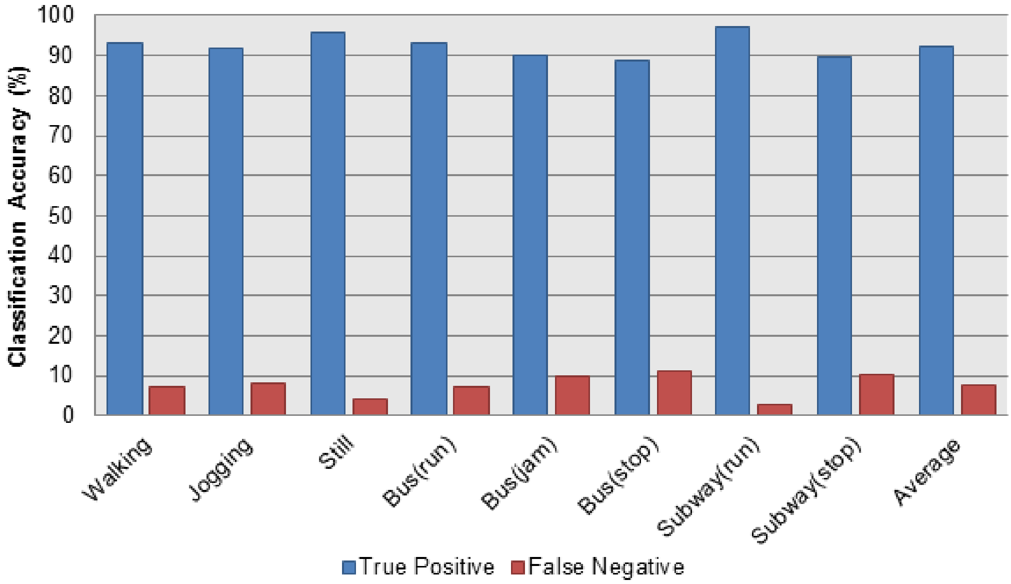


Figure 6-2. Classification accuracy of the integrated system based on Table 6-6.

The experiments clearly show that each individual classifier performed reasonably well, with an average accuracy around 90%. Furthermore, using proposed feature selection method with the accelerometer classifier was more accurate than using some specific kind of features (p-value < 0.05). By combining the two classifiers with other sensor information, integrated system successfully recognized different contexts, including not only ambulatory contexts like walking and jogging, but also transportation contexts like the bus and subway. Although the category is still limited by a small number of contexts, proposed multimodal sensor approach has the potential to recognize different kind of contexts. The proposed algorithm for context recognition is mainly focus on how to acquire better classification result by combining accelerometer and audio sensor data. Therefore the accuracy of proposed classification algorithm is presented in Figure 6-2.

In order to test and evaluate the proposed system in the real-world environment, the system is implemented on an Android smartphone as an application. In Figure 6-3, (a) indicates the initial state of the context recognizer—*i.e.,* 'Still'—with a red line, (b) shows that the user is walking with the smartphone in his hand, and (c) shows that the application recognized his activity as ‘Walking’ with a green line. When the user started jogging with the smartphone in his pocket, as denoted by (d), the proposed system detected his activity as ‘jogging’ and displayed the movement with a blue line, shown by (e). Subsequently, (f) and (g) show that the user is riding a bus, which is recognized and displayed by the system with a yellow line. The user is riding a subway in (h), which can be recognized even the subway is stopped in (i). A full version of the demonstration movie recorded in real world setting is available on YouTube in [111].

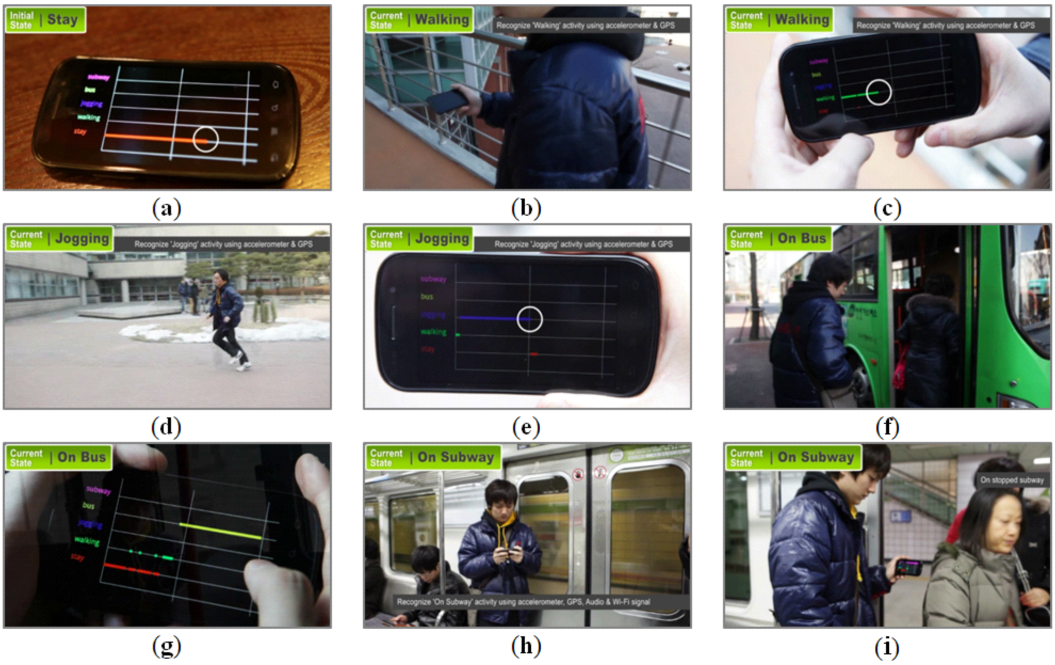


Figure 6-3. Demonstration of the integrated system via testing in a real-world environment.

6.2 Heuristic Approach

In real world environment, the accuracy of the comprehensive activity recognizer is awfully low because of unexpected situations. To enhance the recognition accuracy, several rules from our experiments are applied to the results from comprehensive activity recognizer. In Figure 6-4, total procedure of heuristic-based revision is described. In order to get a Revised Result (RR), 4 rules are applied to Classification Result (CR).

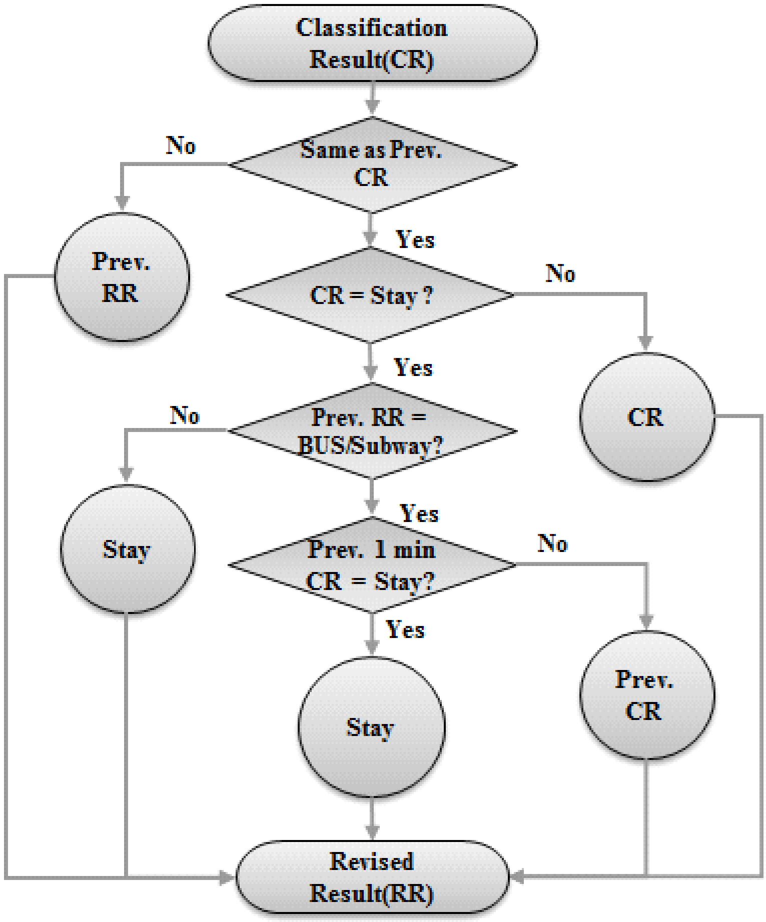


Figure 6-4. Heuristic-based enhanced decision making and revision process.

Table 6-7 and 6-8 are comparisons of classification results and revised results. According to the rules in Figure 6-4 classification results are changed. By applying presented heuristic approach, higher accuracy of whole activity recognition system is expected.

Table 6-7. Heuristic-based result revision, example case 1.

Case 1	CR	RR
1	Stay	- (Initializing)
2	Stay	- (Initializing)
3	Stay	- (Initializing)
4	Walking	- (Initializing)
5	Walking	Walking
6	Stay	Walking
7	Walking	Walking
8	Walking	Walking
9	Walking	Walking
10	Walking	Walking
11	Walking	Walking
12	Stay	Walking
13	Stay	Stay
14	Stay	Stay
15	Walking	Stay
16	Walking	Walking
17	Bus	Walking
18	Bus	Bus
19	Bus	Bus
20	Bus	Bus
21	Bus	Bus
22	Stay	Bus
23	Stay	Bus
24	Bus	Bus
25	Bus	Bus
26	Bus	Bus
27	Stay	Bus
28	Stay	Bus
29	Bus	Bus
30	Bus	Bus

Table 6-8. Heuristic-based result revision, example case 2.

Case 2	CR	RR
…		Walking
133	Walking	Walking
134	Stay	Walking
135	Stay	Stay
136	Stay	Stay
137	Stay	Stay
138	Subway	Stay
139	Subway	Subway
140	Stay	Subway
141	Stay	Subway
142	Walking	Subway
143	Subway	Subway
144	Stay	Subway
145	Subway	Subway
146	Stay	Subway
147	Stay	Subway
148	Subway	Subway
149	Stay	Subway
150	Stay	Subway
151	Stay	Subway
152	Stay	Subway
…		Subway
167	Stay	Subway
168	Stay	Subway
169	Stay	Stay

Figure 6-5 provides a visualization of the refined result (Table 6-7 and 6-8). There is a few enhancement on physical movement such as stay, walking and jogging. But in case of the activities riding a bus or subway, revised results are much enhanced. Base on our experiments, these activities are more sensitive than physical movements and many unexpected situations are easy to happen.

Figure 6-5. Visualization of heuristic-based revision result, (up) Example case 1, (down) Example case 2.

6.3 Lightweight Activity Recognizer

As described in chapter 5 about A-NB algorithms and HARF framework, a real-time activity recognition system has been implemented in the form of smartphone application using the Android OS. The application which uses the Android OS can be installed easily on a Smartphone or mobile devices. In addition, a variety of Smartphone UI (touch screen, keyboard, sound, etc.) enables to the users to add or model their own activities by themselves. The Smartphone application is developed based on the version of Android 2.3.3 (API level 10), as shown in Figure 6-6, the Samsung Nexus S and Galaxy S3 model.

(a) (b) (c)

Figure 6-6. Smartphone application which implementing real-time activity recognition framework. (a) Initial state of application. Sensor monitoring, accelerometer visualization, UI for activity training & recognition. (b) Visualizing 3-axis accelerometer values of walking activity. (c) A screenshot of activity recognition results.

Table 6-9 shows the results of the activity recognition using the developed pARNL Smartphone application. The experiments was conducted on 15 activities including 4 location-base activities (Waiting bus at bus stop, Having a meal at cafeteria, Exercising at gym, Visiting a park) but the result of recognizing activities, only uses GPS location information, shows more than 99.5% accuracy. In the real-world experiment, there is only 1 mis-recognized case was found. But if the GPS on Smartphone is guaranteed to work well, location based activities are well recognized in proposed system. Therefore, the experimental results in Table 6-9 are the accuracy table of 11 activities without visiting a specific location activity.

Table 6-9. Activity recognition accuracy table of 11 activities for validating proposed HARF.

Location		Home			Office			Outdoor				
	Activity	Standing	Walking	Sitting	Standing	Walking	Sitting	Standing	Walking	Sitting	Jogging	Car
Home	Standing	90.32	-	9.68	-	-	-	-	-	-	-	-
	Walking	10.43	83.47	6.1	-	-	-	-	-	-	-	-
	Sitting	2.56	-	98.44	-	-	-	-	-	-	-	-
Office	Standing	-	-	-	95.2	-	4.8	-	-	-	-	-
	Walking	-	-	-	4.84	94.35	0.81	-	-	-	-	-
	Sitting	-	-	-	1.2	0.61	98.19	-	-	-	-	-
Outdoor	Standing	-	-	-	-	-	-	94.34	-	5.66	-	-
	Walking	-	-	-	-	-	-	12.77	80.85	6.38	-	-
	Sitting	-	-	-	-	-	-	2.5	-	97.5	-	-
	Jogging	-	-	-	-	-	-	2.17	10.86	1.47	85.5	-
	Car	-	-	-	-	-	-	16.25	6.25	1.25	-	76.25

The recognition result of 15 activities shows the high accuracy of 92.96% and the result of 11 activities without activities based on only location is 90.4%. There are several cases which shows different accuracy on same activities. It shows that the types of the activity can be changed depending on where the activities are occurred. For example, walking activities in home or outdoor are seldomly recognized as standing activity because the user is frequently stopped or turned in environmental reason. But the recognition accuracy of both sitting and standing activities are relatively higher than others because of their static characteristic. In case of jogging and car, there are some mis-recognition results because a jogging activity is similar to walking and a car is frequently stopped or drove slowly. Figure 6-7 depicts True Positive and False Negative of the 11 activities based on the Table 6-9.

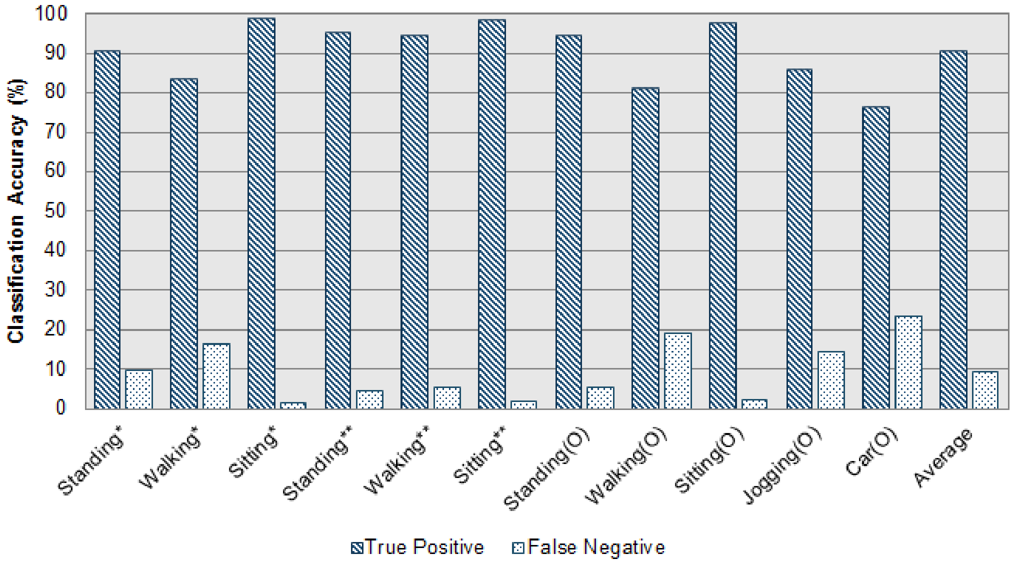


Figure 6-7. Activity recognition accuracy graph of 11 activities.

Figure 6-8. Accuracy comparison between Naïve Bayes and HARF for 15 activities.

Performance comparison of the HARF and the Naïve Bayes algorithm is shown in Figure 6-8. The experiment was performed on 16 activities in a same environment and the result of the Naïve Bayes and the proposed HARF are 81.17% and 89.88% respectively. Recognition results of Standing, Jogging, Sitting and visit specific location are fine with around 90% accuracy in both the HARF and a Naïve Bayes. In case of recognizing transportation activity(Car Driving), HARF showed 76% and but the Naïve Bayes showed the low accuracy of around 50%.

Chapter 7.
Conclusion and Future Directions

In this work, multimodal approach by utilizing the set of embedded sensors on smartphones is proposed in order to recognize different user contexts, such as walking, jogging, riding on a bus, or taking a subway. Overall, demonstration shows that the proposed approach was able to recognize eight contexts, including ambulatory activities and other particular contexts while on a bus or subway. Additionally, it was able to recognize these activities regardless of what the user was doing with his or her smartphone, such as making a phone call, using applications, playing games, or listening to music. Accordingly, the author designed and implemented the proposed system, which enabled position-free recognition and was able to recognize activities wherever the smartphone was attached on the body.

Performance evaluations of the accelerometer and audio data classification schemes showed that the proposed algorithm and system performed better than existing approaches. Proposed system is tested by implementing a smartphone application running on an Android OS. These evaluations also showed that the system works well in real-world environments with the accuracy of 92.43%. Nevertheless, the current system is still limited to a small number of contexts. Further research efforts are necessary to extend the target context category. In addition, the current system is not able to provide detailed information about the recognized contexts, such as bus number, subway line number, or street name while walking. These challenges motivate future research that seeks to utilize other kinds of sensory data to construct a more integrative context-aware system.

References

[1] J. Kwapisz, G. Weiss and S. Moore, "Activity recognition using cell phone accelerometers", ACM SIGKDD Explor. Newsletter, vol. 12, pp. 74-82, 2010

[2] J. Ward, P. Lukowicz, G. Troster, and T. Starner, "Activity recognition of assembly tasks using body-worn microphones and accelerometers", IEEE Trans. Pattern Anal. Mach. Int., vol. 28, pp. 1553-1567, 2006

[3] L. Bao and S. S. Intille, "Activity recognition from user-annotated acceleration data", Pervasive 2004, pp. 1-17, April 21-23, 2004

[4] L. Liao, D. Fox, H. Kautz, "Extracting places and activities from GPS traces using hierarchical conditional random fields", Int. J. Rob. Res., vol. 26, pp. 119-134, 2007

[5] A. Eronen, V. Peltonen, J. Tuomi, A. Klapuri, S. Fagerlund, T. Sorsa, G. Lorho, and J. Huopaniemi, "Audio-based context recognition", IEEE Trans. Audio Speech Lan. Proc. vol. 14, pp. 321-329, 2006

[6] A. Khan, Y. Lee, S. Lee, T. Kim, "A triaxial accelerometer-based physical-activity recognition via augmented-signal features and a hierarchical recognizer", IEEE Trans. Inf. Technol. Biomed., vol. 14, pp. 1166-1172, 2010

[7] N. Ravi, N. Dandekar, P. Mysore, and M. L. Littman, "Activity recognition from accelerometer data", In Proceedings of the 7th Conference on Innovative Applications of Artificial Intelligence, Pittsburgh, PA, USA, pp. 1541-1546, July 9-13, 2005

[8] H. Lu, W. Pan, N. Lane, T. Choudhury, and A. Campbell, "Soundsense: Scalable Sound Sensing for People-Centric Applications on Mobile Phones", In Proceedings of the 7th International Conference on Mobile Systems, Applications, and Services, Krakow, Poland, pp. 165-178, June 22-25, 2009

[9] L. Liao, D. Fox, and H. Kautz, “Location-based activity recognition”, Adv. Neural Inf. Process. Syst. vol. 18, pp. 787-794, 2005

[10] Y. Zheng, and X. Xie, “Learning travel recommendations from user-generated GPS traces”, ACM Trans. Intell. Syst. Technol., 2011

[11] A. Thiagarajan, J. Biagioni, T. Gerlich, and J. Eriksson, “Cooperative Transit Tracking Using Smart-Phones”, In Proceedings of the 8th ACM Conference on Embedded Networked Sensor Systems, Zurich, Switzerland, pp. 85-98, November 3-5, 2010

[12] C. Yiqiang, Y. Qiang, Y. Jie and C. Xiaoyong, “Power-efficient access-point selection for indoor location estimation”, IEEE Trans. Knowl. Data Eng., vol. 18, pp. 877-888, 2006

[13] S. Pekka, and R. Juha, “Recognizing Human Activities User independently on Smartphones Based on Accelerometer Data”, International Journal of Artificial Intelligence and Interactive Multimedia, vol. 1, no. 5, 2012

[14] J. Frank, S. Mannor, and D. Precup, “Activity recognition with mobile phones”, In: Machine Learning and Knowledge Discovery in Databases, Lecture Notes in Computer Science, vol. 6913, pp. 630-633, 2011

[15] J. Gomes, S. Krishnaswamy, Gaber, Mohamed, P. Sousa and E. Menasalvas, “MARS: a personalised Mobile Activity Recognition System”, In: Proceedings of the IEEE International Conference on Mobile Data Management, July 23-26, 2012

[16] Y. He, Y. Li, "Physical Activity Recognition Utilizing the Built-In Kinematic Sensors of a Smartphone", International Journal of Distributed Sensor Networks, Volume 2013, Article ID 481580, 10 pages, 2013

[17] A.M. Khan, Y.K. Lee, and S.Y. Lee, "Accelerometer's position free human activity recognition using a hierarchical recognition mode", e-Health Networking Applications and Services (Healthcom), 2010 12th IEEE International Conference on, pp. 296-301, July 1-3, 2010

[18] B. N. Schilit, N. I. Adams, and R. Want, “Context-aware computing applications”, in Mobile Computing Systems and Applications, pp. 85-90, Dec. 1994

[19] R. Hull, P. Neaves, and J. Bedford-Roberts, “Towards situated computing”, in ISWC '97: Proceedings of the 1st IEEE International Symposium on Wearable Computers. Washington, DC, USA: IEEE Computer Society, p. 146, 1997

[20] P. J. Brown, “The stick-e document: a framework for creating context-aware applications”, in Proceedings of EP'96, Palo Alto., pp. 259-272, January 1996

[21] N. S. Ryan, J. Pascoe, and D. R. Morse, “Enhanced reality fieldwork: the context-aware archaeological assistant”, in Computer Applications in Archaeology 1997, October 1998

[22] A. K. Dey and G. D. Abowd, “Towards a better understanding of context and context-awareness”, CHI 2000 Workshop on the What, Who, Where, When, and How of Context-Awareness, 2000

[23] K. Henricksen, “A framework for context-aware pervasive computing applications”, Dissertation, School of Information Technology and Electrical Engineering, The University of Queensland, September, 2003

[24] R. Want, A. Hopper, V. Falcao, and J. Gibbons, “The active badge location system”, ACM Trans. Inf. Syst., vol. 10, no. 1, pp. 91-102, 1992

[25] S. Sigg, “Development of a novel context prediction algorithm and analysis of context prediction schemes”, Dissertation, University of Kassel, 2008

[26] R. Mayrhofer, H. Radi, and A. Ferscha, Recognizing and predicting context by learning from user behavior”, Radiomatics: Journal of Communication Engineering, special issue on Advances in Mobile Multimedia, vol. 1, no. 1, May 2004

[27] T. Strang and C. Linnhoff-Popien, “A context modeling survey”, in First International Workshop on Advanced Context Modeling, Reasoning And

Management, Nottingham, England, September 2004

[28] H. Brashear, T. Starner, P. Lukowicz, and H. Junker, “Using multiple sensors for mobile sign language recognition”, in Wearable Computers, 2003. Proceedings. Seventh IEEE International Symposium on, pp. 45-52, 2003

[29] J. Liu, Z. Wang, L. Zhong, J. Wickramasuriya, and V. Vasudevan, “uwave: Accelerometer-based personalized gesture recognition and its applications”, Pervasive Computing and Communications, IEEE International Conference on, vol. 0, pp. 1-9, 2009

[30] A. Ferscha, S. Vogl, B. Emsenhuber, and B.Wally, “Physical shortcuts for media remote controls”, in Adjunct Proceedings of the 2nd International Conference on Intelligent Technologies for Interactive Entertainment (INTETAIN 2008), January, 2008

[31] G. D. Abowd, C. G. Atkeson, J. Hong, S. Long, R. Kooper, and M. Pinkerton, “Cyberguide: a mobile context-aware tour guide”, Wirel. Netw., vol. 3, no. 5, pp. 421-433, 1997

[32] K. Cheverst, N. Davies, K. Mitchell, A. Friday, and C. Efstratiou, “Developing a context-aware electronic tourist guide: some issues and experiences”, in CHI '00: Proceedings of the SIGCHI conference on Human factors in computing systems. New York, NY, USA: ACM, pp. 17-24, 2000

[33] A. Schmidt, M. Beigl, and H.-W. Gellersen, “There is more to context than location”, Computers & Graphics, vol. 23, no. 6, pp. 893-901, 1999

[34] N. Roy, A. Roy, and S. K. Das, “Context-aware resource management in multi inhabitant smart homes: A nash h-learning based approach”, in PERCOM '06: Proceedings of the Fourth Annual IEEE International Conference on Pervasive Computing and Communications. Washington, DC, USA: IEEE Computer Society, pp. 148-158, 2006

[35] M. Sutterer, O. Droegehorn, and K. David, “Upos: User profile ontology with situation-dependent preferences support”, in ACHI '08: Proceedings of the First

International Conference on Advances in Computer-Human Interaction, Washington, DC, USA: IEEE Computer Society, pp. 230-235, 2008

[36] J. E. Bardram and N. Nørskov, “A context-aware patient safety system for the operating room”, in UbiComp '08: Proceedings of the 10th international conference on Ubiquitous computing. New York, NY, USA: ACM, pp. 272-281, 2008

[37] A. Matic, P. Mehta, J. M. Rehg, V. Osmani, and O. Mayora, “aid-me: Automatic identification of dressing failures through monitoring of patients and activity evaluation”, in 4th International Conference on Pervasive Computing Technologies for Healthcare 2010 (Pervasive Health 2010), March, 2010

[38] A. K. Dey, G. D. Abowd, and D. Salber, “A conceptual framework and a toolkit for supporting the rapid prototyping of context-aware applications”, Hum.-Comput. Interact., vol. 16, no. 2, pp. 97-166, 2001

[39] P. Korpipaa, J. Mantyjarvi, J. Kela, H. Keranen, and E.-J. Malm, “Managing context information in mobile devices”, IEEE Pervasive Computing, vol. 2, pp. 42-51, 2003

[40] H. Chen, T. Finin, and A. Joshi, “An Intelligent Broker for Context-Aware Systems”, Adjunct Proceedings of Ubicomp 2003, pp. 183-184, October, 2003

[41] P. Fahy and S. Clarke, “Cass: a middleware for mobile context-aware applications”, in Workshop on Context Awareness, MobiSys, 2004

[42] T. Gu, H. K. Pung, and D. Zhang, “A service-oriented middleware for building context-aware services”, J. Network and Computer Applications, vol. 28, no. 1, pp. 1-18, 2005

[43] S. L. Lau, “Towards a user-centric context aware system”, Ph. D. Thesis, Kassel University, November 22, 2011

[44] L. Barkhuus and A. Dey, “Is Context-Aware Computing Taking Control Away from the User? Three Levels of Interactivity Examined”, in In Proceedings of

Ubicomp 2003. Springer, pp. 149-156, 2003

[45] G. Chen and D. Kotz, “A survey of context-aware mobile computing research”, Dartmouth College, Hanover, NH, USA, Tech. Rep., 2000

[46] M. Weiser, “The computer for the 21st century”, SIGMOBILE Mob. Comput. Commun. Rev., vol. 3, no. 3, pp. 3-11, 1999

[47] A. Pentland, “Smart rooms, smart clothes”, in Pattern Recognition, 1998. Proceedings. Fourteenth International Conference on, vol. 2, pp. 949-953, 1998

[48] T. Starner, J. Weaver, and A. Pentland, “A wearable computer based American sign language recognizer”, in ISWC '97: Proceedings of the 1st IEEE International Symposium on Wearable Computers. Washington, DC, USA: IEEE Computer Society, 1997

[49] N. Eagle and A. Pentland, “Wearables in the workplace: sensing interactions at the office”, in Wearable Computers, 2003. Proceedings. Seventh IEEE International Symposium on, 2003

[50] Z. Zhou, X. Chen, Y.-C. Chung, Z. He, T. Han, and J. Keller, “Activity analysis, summarization, and visualization for indoor human activity monitoring”, Circuits and Systems for Video Technology, IEEE Transactions on, vol. 18, no. 11, pp. 1489-1498, Nov. 2008

[51] Z. He, L. Jim “Activity Recognition from Acceleration Data Using AR Model Representation and SVM”, In Proceeding of the International Conference on Machine Learning and Cybernetics 2008, Kunming, China, pp. 2245–2250, July 12-15, 2008

[52] N. Kern, B. Schiele, and A. Schmidt, “Recognizing context for annotating a live life recording”, Personal Ubiquitous Comput., vol. 11, no. 4, pp. 251-263, 2007

[53] La The Vinh, Sungyoung Lee, Young-Tack Park and Brian J. d'Auriol, "A Novel Feature Selection Method Based on Normalized Mutual Information",

Applied Intelligence, ISSN: 0924-669X, Vol.37, No. 1, pp.100-120, August 23, 2011

[54] K. Van Laerhoven and O. Cakmakci, “What shall we teach our pants?”, in ISWC '00: Proceedings of the 4th IEEE International Symposium on Wearable Computers. Washington, DC, USA: IEEE Computer Society, 2000

[55] J. Mantyjarvi, J. Himberg, and T. Seppanen, “Recognizing human motion with multiple acceleration sensors”, in Systems, Man, and Cybernetics, 2001 IEEE International Conference on, vol. 2, pp. 747-752, 2001

[56] J. R. Quinlan, C4.5: programs for machine learning. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1993

[57] E. M. Tapia, S. S. Intille, W. Haskell, K. Larson, J. Wright, A. King, and R. Friedman, “Real-Time recognition of physical activities and their intensities using wireless accelerometers and a heart rate monitor”, in Wearable Computers, 2007 11th IEEE International Symposium on, 2007

[58] J. Lester, T. Choudhury, and G. Borriello, “A practical approach to recognizing physical activities”, in Pervasive, ser. Lecture Notes in Computer Science, vol. 3968. Springer, pp. 1-16, 2006

[59] Y. Cho, Y. Nam, Y. Choi, and W. Cho, “SmartBuckle: human activity recognition using a 3-axis accelerometer and a wearable camera”, in Proceedings of the 2nd International Workshop on Systems and Networking Support for Health Care and Assisted Living Environments. Breckenridge, Colorado: ACM, 2008

[60] M. Mladenov and M. Mock, “A step counter service for java-enabled devices using a built-in accelerometer”, in CAMS '09: Proceedings of the 1st International Workshop on Context-Aware Middleware and Services. New York, NY, USA: ACM, 2009

[61] G. Bieber, J. Voskamp, and B. Urban, “Activity recognition for everyday life on mobile phones”, in HCI (6), pp. 289-296, 2009

[62] T. Brezmes, J. Gorricho, and J. Cotrina, "Activity recognition from accelerometer data on a mobile phone", in Proceedings of the 10th International Work-Conference on Artificial Neural Networks, Salamanca, Spain: Springer-Verlag, pp. 796-799, 2009

[63] C. Mattmann, O. Amft, H. Harms, F. Clemens, and G. Troster, "Recognizing upper body postures using textile strain sensors", in Proc. 11th International Activity recognition system using unobtrusive sensor devices Symposium on Wearable Computers (ISWC07), 2007

[64] H. Harms, O. Amft, D. Favre, C. Liesen, D. Roggen, and G. Troster, "Motion sensitive clothing", in Adjunct Proceedings of Pervasive 2010 Conference, 2010

[65] J. Cheng, O. Amft, and P. Lukowicz, "Active capacitive sensing: Exploring a new wearable sensing modality for activity recognition", in Pervasive Computing, pp. 319-336, 2010

[66] H. W. Gellersen, A. Schmidt, and M. Beigl, "Multi-sensor context-awareness in mobile devices and smart artifacts", Mob. Netw. Appl., vol. 7, no. 5, pp. 341-351, 2002

[67] K. V. Laerhoven, K. Aidoo, and S. Lowette, "Real-time analysis of data from many sensors with neural networks", in In ISWC. IEEE Press, pp. 115-123, 2001

[68] R. Want, A. Hopper, V. Falcao, and J. Gibbons, "The active badge location system", ACM Trans. Inf. Syst., vol. 10, no. 1, pp. 91-102, 1992

[69] M. Beigl, H.-W. Gellersen, and A. Schmidt, "Mediacups: experience with design and use of computer-augmented everyday artifacts", Computer Networks, vol. 35, no. 4, pp. 401-409, 2001

[70] A. Ferscha, S. Vogl, B. Emsenhuber, and B.Wally, "Physical shortcuts for media remote controls", in Adjunct Proceedings of the 2nd International Conference on Intelligent Technologies for Interactive Entertainment

(INTETAIN 2008), January, 2008

[71] G. H. Jin, S. B. Lee, and T. S. Lee, "Context awareness of human motion states using accelerometer", J. Med. Syst., vol. 32, no. 2, pp. 93-100, 2008

[72] L. C. Jatoba, U. Grossmann, C. Kunze, J. Ottenbacher, and W. Stork, "Context-aware mobile health monitoring: Evaluation of different pattern recognition methods for classification of physical activity", in Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE, pp. 5250-5253, Aug. 2008

[73] M. Jehn, A. Schmidt-Trucksaess, T. Schuster, H. Hanssen, M. Weis, M. Halle, and F. Koehler, "Accelerometer-based quantification of 6-minute walk test performance in patients with chronic heart failure: Applicability in telemedicine", Journal of Cardiac Failure, vol. 15, no. 4, pp. 334-340, 2009

[74] S. Helal, W. Mann, H. El-Zabadani, J. King, Y. Kaddoura, and E. Jansen, "The Gator Tech Smart House: A programmable pervasive space", Computer, vol. 38, no. 3, pp. 50-60, 2005

[75] D. J. Cook, M. Youngblood, E. O. Heierman, III, K. Gopalratnam, S. Rao, A. Litvin, and F. Khawaja, "MavHome: An agent-based smart home", in PERCOM '03: Proceedings of the First IEEE International Conference on Pervasive Computing and Communications. Washington, DC, USA: IEEE Computer Society, 2003

[76] F. Rivera-Illingworth, V. Callaghan, and H. Hagras, "Automated discovery of human activities inside pervasive living spaces", in Pervasive Computing and Applications, 2006 1st International Symposium on, pp. 77-82, 2006

[77] V. L. Erickson, Y. Lin, A. Kamthe, R. Brahme, A. Surana, A. E. Cerpa, M. D. Sohn, and S. Narayanan, "Energy efficient building environment control strategies using real-time occupancy measurements", in BuildSys '09: Proceedings of the First ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings. New York, NY, USA: ACM, 2009

[78] A. Ferscha, B. Emsenhuber, S. Gusenbauer, and B. Wally, “PowerSaver: Pocket-worn activity tracker for energy management”, Innsbruck, Austria, pp. 321-324, September, 2007

[79] W. Lin, M.-T. Sun, R. Poovandran, and Z. Zhang, “Human activity recognition for video surveillance”, in Circuits and Systems, 2008. ISCAS 2008, IEEE International Symposium on, pp. 2737-2740, 2008

[80] S. Agrawal, I. Constandache, S. Gaonkar, and R. R. Choudhury, “Phone Pointpen: using mobile phones to write in air”, in MobiHeld '09: Proceedings of the 1st ACM workshop on Networking, systems, and applications for mobile handhelds. New York, NY, USA: ACM, 2009

[81] D. T. G. Huynh, “Human activity recognition with wearable sensors”, Dissertation, Technische Universitat Darmstadt, 2008

[82] T. Gu, S. Chen, X. Tao, and J. Lu, “An unsupervised approach to activity recognition and segmentation based on object-use fingerprints”, Data Knowl. Eng., vol. 69, no. 6, pp. 533-544, 2010

[83] K. Van Laerhoven, N. Kern, H.-W. Gellersen, and B. Schiele, “Towards a wearable inertial sensor network”, in Eurowearable, 2003. IEE, 2003

[84] K. S. Kunze, P. Lukowicz, H. Junker, and G. Troster, “Where am i: Recognizing on-body positions of wearable sensors”, in LoCA, ser. Lecture Notes in Computer Science, vol. 3479. Springer, pp. 264-275, 2005

[85] C. Lombriser, N. B. Bharatula, D. Roggen, and G. Troster, “On-body activity recognition in a dynamic sensor network”, in BodyNets '07: Proceedings of the ICST 2nd international conference on Body area networks. ICST, Brussels, Belgium, pp. 1-6, 2007

[86] J. Yang, “Toward physical activity diary: motion recognition using simple acceleration features with mobile phones”, in IMCE '09: Proceedings of the 1st international workshop on Interactive multimedia for consumer electronics, New York, NY, USA: ACM, pp. 1-10, 2009

[87] S. L. Lau and K. David, “Movement recognition using the accelerometer in smartphones”, in Future Network & Mobile Summit, 2010

[88] C. Doukas and I. Maglogiannis, “Enabling human status awareness in assistive environments based on advanced sound and motion data classification”, in PETRA '08: Proceedings of the 1st international conference on Pervasive Technologies Related to Assistive Environments, New York, NY, USA: ACM, pp. 1-8, 2008

[89] S. Zhang, P. McCullagh, C. Nugent, and H. Zheng, “Activity monitoring using a smart phone's accelerometer with hierarchical classification”, in The 6th International Conference on Intelligent Environments, Kualar Lumpar, Malaysia, 2010

[90] H. Junker, O. Amft, P. Lukowicz, and G. Troster, “Gesture spotting with body-worn inertial sensors to detect user activities”, Pattern Recognition, vol. 41, no. 6, pp. 2010-2024, 2008

[91] E. Kim, S. Helal, and D. Cook, “Human activity recognition and pattern discovery”, IEEE Pervasive Computing, vol. 9, no. 1, pp. 48-53, 2010

[92] Parkka, Juha., M. Ermes, P. Korpipaa, J. Mantyjarvi, J. Peltola, and I. Korhonen, “Activity classi cation using realistic data from wearable sensors”, Information Technology in Biomedicine, IEEE Transactions on, vol. 10, no. 1, pp. 119-28, 2006

[93] N. Wang, E. Ambikairajah, and B. G. Lovell, N.H.and Celler, “Accelerometry based classification of walking patterns using time-frequency analysis”, in Engineering in Medicine and Biology Society, 2007

[94] K. Van Laerhoven and H.-W. Gellersen, “Spine versus porcupine: a study in distributed wearable activity recognition”, in Wearable Computers, ISWC 2004, Eighth International Symposium on, vol. 1, pp. 142-149, 2004

[95] S. L. Lau, I. Konig, K. David, B. Parandian, C. Carius-Dussel, and M. Schultz, “Supporting patient monitoring using activity recognition with a

smartphone", in The 7th International Symposium on Wireless Communication Systems(ISWCS), 2010, York, United Kingdom, September 19-22, 2010

[96] T. Huynh and B. Schiele, "Unsupervised discovery of structure in activity data using multiple eigenspaces", in 2nd International Workshop on Location- and Context- Awareness (LoCA 2006), Dublin, Ireland, May, 2006

[97] D. Minnen, T. Starner, I. Essa, and C. Isbell, "Improving activity discovery with automatic neighborhood estimation", in IJCAI'07: Proceedings of the 20th international joint conference on Artificial intelligence. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2007

[98] J. Lin, E. Keogh, S. Lonardi, and B. Chiu, "A symbolic representation of time series, with implications for streaming algorithms", in DMKD '03: Proceedings of the 8th ACM SIGMOD workshop on Research issues in data mining and knowledge discovery. New York, NY, USA: ACM, 2003

[99] D. Pyle, "Data preparation for data mining", San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1999

[100] E. J. Keogh, S. Chu, D. Hart, and M. J. Pazzani, "An online algorithm for segmenting time series", in ICDM '01: Proceedings of the 2001 IEEE International Conference on Data Mining. Washington, DC, USA: IEEE Computer Society, pp. 289-296, 2001

[101] J. Mäntyjärvi, J. Himberg, and T. Seppänen, "Recognizing human motion with multiple acceleration sensors", in Systems, Man, and Cybernetics, 2001 IEEE International Conference on, vol. 2, pp. 747-752, 2001

[102] T. Huynh and B. Schiele, "Towards less supervision in activity recognition from wearable sensors", in Proceedings of the 10th IEEE International Symposium on Wearable Computing (ISWC 2006), Montreux, Switzerland, pp. 3-10, October, 2006

[103] K. Van Laerhoven and H.-W. Gellersen, "Spine versus porcupine: a study in distributed wearable activity recognition", in Wearable Computers, 2004,

ISWC 2004. Eighth International Symposium on, vol. 1, pp. 142-149, 2004

[104] C. Bouten, K. Koekkoek, M. Verduin, R. Kodde, and J. Janssen, “A triaxial accelerometer and portable data processing unit for the assessment of daily physical activity”, Biomedical Engineering, IEEE Transactions on, vol. 44, no. 3, pp. 136-147, Mar. 1997

[105] T. Choudhury, M. Philipose, D. Wyatt, and J. Lester, “Towards activity databases: Using sensors and statistical models to summarize people's lives”, IEEE Data Eng. Bull., vol. 29, no. 1, pp. 49-58, 2006

[106] F. Ichikawa, J. Chipchase, and R. Grignani, “Where's the phone? a study of mobile phone location in public spaces”, in 2nd International Conference on Mobile Technology, Applications and Systems 2005, Guangzhou, November, 2005

[107] Rabiner, L.R. “A tutorial on hidden markov models and selected applications in speech recognition”, Proceeding of the IEEE, vol. 77, no. 2, pp. 257-286, 1989

[108] C. Kim, K.-B. Hwang, “Comparative Study of Machine Learning Techniques for Spammer Detection in Social Bookmarking Systems,” Journal of KIISE : Computing Practices and Letters, vol. 15, no. 5, pp. 345-349, May, 2009

[109] C.-H. Lee, “An Information-theoretic Approach for Value-Based Weighting in Naïve Bayesian Learning,” Journal of KIISE : Databases, vol. 37, no. 6, pp. 285-291, December, 2010.

[110] Acoustic Environment Classification. Available online: http://lemur.cmp.uea.ac.uk/Research/noise_db/

[111] Demonstration Movie of Real-Time Activity Recognition Using Smartphone. Available online: http://www.youtube.com/watch?v=Zg9KVtGEDD4/

[112] Ma, L.; Smith, D.; Milner, B. “Environmental Noise Classification for Context-Aware Applications”, In Proceeding of the EuroSpeech, Geneva,

Switzerland, pp. 2237–2240, September 1-4, 2003

[113] P. Srivastava, W.-C. Wong, "Hierarchical Human Activity Recognition using GMM", AMBIENT 2012 : The Second International Conference on Ambient Computing, Applications, Services and Technologies, Barcelona, Spain, pp. 32-37, Sept. 23, 2012

[114] J.A. Stork, L. Spinello, J. Silva, K.O. Arras, "Audio-based human activity recognition using Non-Markovian Ensemble Voting", published in RO-MAN, 2012 IEEE, Paris, France, pp. 509-514, Sept. 9-13, 2012

[115] H. Wu, "Sensor Data Fusion for Context-Aware Computing Using Dempster-Shafer Theory", Ph. D. Thesis, CMU-RI-TR-03-52, Carnegie Mellon University, December, 2003

[116] David Lee Hall, "Mathematical Techniques in Multisensor Data Fusion", Artech House Inc., ISBN 0-89006-558-6, 1992

[117] Lawrence A. Klein, "Sensor and Data Fusion Concepts and Applications" (second edition), SPIE Optical Engineering Press, ISBN 0-8194-3231-8, 1999

[118] Shulin Yang and Kuo-Chu Chang, "Modular Neural Net Architecture for Automatic Target Recognition", SPIE Proceedings Vol.2755, Signal Processing, Sensor Fusion, and Target Recognition V, Orlando, Florida, pp. 166-177, April 8-10, 1996

Appendix: List of Publications

A.1 Journal Papers

[1] **한만형**, 이승룡, "스마트폰 멀티모달 센서 기반 개인화 행위모델링 및 실시간 행위인지 알고리즘", 정보과학회논문지 : 소프트웨어 및 응용, 제40권, 제6호, 발간예정일: 2013년 6월 15일

[2] **Manhyung Han**, La The Vinh, Young-Koo Lee and Sungyoung Lee, "Comprehensive Context Recognizer Based on Multimodal Sensors in a Smartphone", Journal of Sensors (SCIE), ISSN: 1424-8220, Vol. 12, No. 9, pp. 12588-12605, IF: 1.77, September 17, 2012

[3] Yongkoo Han, **Manhyung Han**, Sungyoung Lee, A. M. Jehad Sarkar and Young-Koo Lee, "A Framework for Supervising Lifestyle Diseases Using Long-Term Activity Monitoring", Journal of Sensors (SCIE), ISSN: 1424-8220, Vol. 12, No. 5, pp. 5363-5379, IF: 1.77, April 26, 2012

[4] Asad Masood Khattak, Phan Tran Ho Truc, Le Xuan Hung, La The Vinh, Viet-Hung Dang, Donghai Guan, Zeeshan Pervez, **Manhyung Han**, Sungyoung Lee and Young-Koo Lee, "Towards Smart Homes Using Low Level Sensory Data", Journal of Sensors (SCIE), ISSN: 1424-8220, Vol. 11, pp. 11581-11604, IF: 1.77, December 12, 2011

[5] La The Vinh, Hung Xuan Le, Hung Quoc Ngo, Hyoung Il Kim, **Manhyung Han**, Young-Koo Lee and Sungyoung Lee, "Semi-Markov Conditional Random Fields for Accelerometer-Based Activity Recognition", Applied Intelligence, ISSN: 0924-669X, IF: 0.775, 2010

[6] Xuan Hung Le, Sungyoung Lee, Ismail Butun, Murad Khalid, Ravi Sankar, Miso (Hyoung-IL) Kim, **Manhyung Han**, Young-Koo Lee and Heejo Lee, “An Energy-Efficient Access Control Scheme for Wireless Sensor Networks based on Elliptic Curve Cryptography”, Journal of Communications and Networks, Vol. 11, No. 6, pp. 599-606, December 2009

A.2 Conference Papers

[1] **Manhyung Han**, Yong-Koo Han, Hyoung-Il Kim and Sungyoung Lee, “Mobile Activity Sensor Logger for Monitoring Chronic Disease Patients”, The 8th International Conference on Wearable Micro and Nano Technologies for Personalized Health (pHealth 2011), Lyon, France, June 29 - July 1, 2011

[2] **한만형**, 이승룡, “토픽 모델 기반의 단일 행위들로부터 행위 패턴 인지 기술”, 한국통신학회 2010년도 하계종합학술발표회, 2010년 6월 23일 ~ 25일

[3] Asad Masood Khattak, Zeeshan Pervez, **Manhyung Han**, Chris Nugent and Sungyoung Lee, “DDSS: Dynamic Decision Support System for Elderly”, The 25th IEEE International Symposium on Computer-Based Medical Systems (CBMS 2012), Rome, Italy, June 20-22, 2012

[4] Md. Kamrul Hasan, **Manhyung Han**, Sungyoung Lee and Young-Koo Lee, “A Sensor Network System for Event Attribution in Multi-user Home Environment”, IC-AI'10 - The 2010 International Conference on Artificial Intelligence, Las Vegas, USA, July 12-15, 2010

[5] Asad Masood Khattak, La The Vinh, Dang Viet Hung, Phan Tran Ho Truc, Le Xuan Hung, D. Guan, Zeeshan Pervez, **Manhyung Han**, Sungyoung Lee and Young-Koo Lee, “Context-aware Human Activity Recognition and Decision

Making", 12th International Conference on e-Health Networking, Application Services(IEEE HealthCom 2010), Lyon, France, July 1-3, 2010

[6] Xuan Hung Le, Sungyoung Lee, Phan Tran Ho Truc, La The Vinh, Asad Masood Khattak, **Manhyung Han**, Dang Viet Hung, Mohammad M. Hassan, Miso (Hyung-Il) Kim, Kyo-Ho Koo, Young-Koo Lee and Eui-Nam Huh, "Secured WSN-integrated Cloud Computing for u-Life Care", 7th Annual IEEE CCNC 2010 Conference, Las Vegas, USA, January 9-12, 2010

[7] Asad Masood Khattak, Khalid Latif, **Manhyung Han**, Sungyoung Lee, Young-Koo Lee and Hyoung-Il Kim, "Change Tracer: Tracking Changes in Web Ontologies", 21st IEEE International Conference on Tools with Artificial Intelligence, Newark, NJ, USA, November 2-4, 2009.

A.3 Patents

[1] 이승룡, 이영구, 라더빈, 르산흥, 노콕흥, 김형일, **한만형**, "세미 마르코프 조건부 랜덤 필드 모델 기반의 행동 인식 방법", 대한민국 특허, 등록번호: 10-1163834, 등록일자: 2012년 7월 2일

[2] Sungyoung Lee, Young-Koo Lee, La The Vinh, Le Xuan Hung, Ngo Quoc Hung, Hyoung Il Kim and **Manhyung Han**, "Method of Recognizing Activity on Basis of Semi-Markov Conditional Random Field Model", Appl. No. 12/886,800, U.S. Patent, Sep. 21, 2010

[3] Sungyoung Lee, Young-Koo Lee, Asad Masood Khattak, Hyoung Il Kim and **Manhyung Han**, Method for Reconciling Mappings in Dynamic/Evolving Web-Ontologies Using Change History Ontology, Appl. No. 12/576,342, U.S. Patent, Oct. 9, 2009