

HOME HEALTHCARE USING UBIQUITOUS COMPUTING AND  
ROBOT TECHNOLOGIES

By

MINH PHAM

Bachelor of Computer Science  
Hanoi University of Science and Technology  
Ha Noi, Vietnam  
2007

Master of Science in Management Information System  
Oklahoma State University  
Stillwater, Oklahoma, US  
2012

Submitted to the Faculty of the  
Graduate College of  
Oklahoma State University  
in partial fulfillment of  
the requirements for  
the Degree of  
DOCTOR OF PHILOSOPHY  
December, 2018

COPYRIGHT ©

By

MINH PHAM

December, 2018

HOME HEALTHCARE USING UBIQUITOUS COMPUTING AND  
ROBOT TECHNOLOGIES

Dissertation Approved:

Dr. Weihua Sheng

---

Dissertation Advisor

Dr. Yanmin Gong

---

Dr. Sabit Ekin

---

Dr. Alex Bishop

Name: Minh Pham

Institution: Oklahoma State University

Location: Stillwater, Oklahoma

Title of Study: HOME HEALTHCARE USING UBIQUITOUS COMPUTING AND  
ROBOT TECHNOLOGIES

Pages in Study: 131

Major Field: Electrical Engineering

The rapid increase of senior population worldwide is challenging the existing healthcare and support systems. Recently, smart home environments are utilized for ubiquitous health monitoring, allowing patients to stay in the comfort of their homes. In this dissertation, a Cloud-based Smart Home Environment (CoSHE) for home healthcare is presented, which consists of ambient intelligence, wearable computing, and robot technologies. The system includes a smart home which is embedded with distributed environmental sensors to support human localization. Wearable units are developed to collect physiological, motion and audio signals through non-invasive wearable sensors and provide contextual information in terms of the resident's daily activity and location in the home. This enables healthcare professionals to study daily activities, behavioral changes and monitor rehabilitation and recovery processes. The sensor data are processed in a smart home gateway and sent to a private cloud, which provides real-time data access for remote caregivers. Our case studies show that contextual information provided by ubiquitous computing can help better understand the patient's health status. With a robot assistant in the loop, we demonstrated that the CoSHE can facilitate healthcare delivery via interaction between human and robot.

## TABLE OF CONTENTS

Chapter	Page
<b>I INTRODUCTION</b>	<b>1</b>
I.1 Motivation . . . . .	1
I.2 Challenges . . . . .	4
I.2.1 Hardware setup . . . . .	4
I.2.2 Physiological data collection . . . . .	4
I.2.3 Robot development . . . . .	5
I.3 Objectives . . . . .	5
I.4 Contributions . . . . .	6
I.5 Outlines . . . . .	7
<b>II SMART GARMENT DEVELOPMENT</b>	<b>9</b>
II.1 Introduction . . . . .	9
II.2 Related Work . . . . .	9
II.3 Garment prototype 1 . . . . .	10
II.3.1 ECG Sensor . . . . .	11
II.3.2 SPO <sub>2</sub> Sensor . . . . .	11
II.3.3 Respiration Sensor . . . . .	12
II.3.4 Airflow Sensor . . . . .	13
II.3.5 Garment design . . . . .	13
II.4 Garment prototype 2 . . . . .	15
II.5 Signal Acquisition Result . . . . .	17
II.5.1 Signals collected using standard electrodes . . . . .	17

II.5.2	Signals collected using textile electrodes . . . . .	19
II.6	Summary . . . . .	20
<b>III</b>	<b>CLOUD-BASED SMART HOME ENVIRONMENT (CoSHE)</b>	<b>21</b>
III.1	Introduction . . . . .	21
III.2	Related Work . . . . .	21
III.3	System Architecture . . . . .	25
III.3.1	Smart Home . . . . .	26
III.3.2	Cloud Infrastructure . . . . .	27
III.4	Summary . . . . .	30
<b>IV</b>	<b>CONTEXT AWARENESS IN CoSHE</b>	<b>31</b>
IV.1	Body Activity Recognition . . . . .	31
IV.1.1	Related Work . . . . .	31
IV.1.2	Hardware Setup . . . . .	32
IV.1.3	Methodology . . . . .	33
IV.1.4	Experiment Results . . . . .	34
IV.2	Indoor Human Localization . . . . .	36
IV.2.1	Related Work . . . . .	36
IV.2.2	Methodology . . . . .	40
IV.2.3	Experiment Results . . . . .	51
IV.3	Summary . . . . .	55
<b>V</b>	<b>WELL-BEING MONITORING</b>	<b>56</b>
V.1	Body Hydration Monitoring . . . . .	56
V.1.1	Introduction . . . . .	56
V.1.2	Related Work . . . . .	57
V.1.3	AutoHydrate Application . . . . .	58
V.2	Negative Emotion Recognition . . . . .	65

V.2.1	Introduction . . . . .	65
V.2.2	Negative Emotion Recognition using ECG . . . . .	65
V.2.3	Negative Emotion Recognition using Facial Expression . . . . .	78
V.2.4	Multimodal Fusion Framework . . . . .	85
V.2.5	Summary . . . . .	97
<b>VI HEALTHCARE DELIVERY THROUGH A ROBOT ASSISTANT</b>		<b>98</b>
VI.1	Introduction . . . . .	98
VI.2	Robot Assistant . . . . .	99
VI.3	Hydration Reminder . . . . .	102
VI.3.1	Hydration Recommendation . . . . .	102
VI.3.2	Interactive Conversation . . . . .	102
VI.4	Emotion Regulation . . . . .	104
VI.4.1	Distraction method . . . . .	104
VI.4.2	Interactive Conversation . . . . .	107
VI.4.3	Result . . . . .	108
VI.5	Summary . . . . .	110
<b>VII CONCLUSIONS AND FUTURE WORKS</b>		<b>111</b>
VII.1	Conclusion . . . . .	111
VII.2	Future Work . . . . .	112
<b>BIBLIOGRAPHY</b>		<b>115</b>

## LIST OF TABLES

Table		Page
III.1	Smart home projects. . . . .	25
III.2	PIR Panasonic EKMC1601111 specification. . . . .	27
III.3	Data stored in the CoSHE database. . . . .	30
IV.1	Indoor localization methods . . . . .	40
IV.2	Summary statistics of distance error (D). . . . .	52
V.1	Body Activity Classification . . . . .	61
V.2	Estimated Energy Requirement . . . . .	62
V.3	Drinking Detection Results using 8-Fold Cross-Validation . . . . .	63
V.4	Result of Body Activity Level Recognition using 6-Fold Cross-Validation . . . . .	64
V.5	Final Detection Results . . . . .	64
V.6	Result of negative emotion classification. . . . .	76
V.7	CNN architectures of ILSVRC top competitors. . . . .	80
V.8	Result of transfer learning using GoogLeNet. . . . .	85
V.9	Yaw angle categories and corresponding model accuracies. . . . .	89
V.10	Fusion Rule based on two individual models. . . . .	93
V.11	Confusion matrices of negative emotion recognition models. . . . .	96



## LIST OF FIGURES

Figure		Page
I.1	U.S. Population 60+ by Age: 1960-2050. . . . .	2
I.2	System overview. . . . .	6
II.1	Health kit. . . . .	11
II.2	The positions of the electrodes. . . . .	12
II.3	How the health kit works. . . . .	13
II.4	Smart garment prototype 1. . . . .	14
II.5	Wearable unit components. . . . .	15
II.6	Smart garment prototype 2. . . . .	16
II.7	Physiological signals collected using wearable sensors. . . . .	17
II.8	ECG signal affected by artificial noise. . . . .	18
II.9	ECG signals collected by using standard electrodes and textile electrodes. . . . .	19
III.1	The architecture of the Cloud-based smart home healthcare system. . . . .	24
III.2	Smarthome testbed setup. . . . .	26
III.3	The principle of PIR sensors. . . . .	27
III.4	Private cloud infrastructure . . . . .	28
IV.1	An inertial measurement unit (IMU). . . . .	32
IV.2	Activity recognition result. . . . .	35
IV.3	Overview of the approach. . . . .	41

IV.4	Top view of the mock apartment. Dotted circles denote the field of view of the PIRs. . . . .	42
IV.5	PIR sensing area. (a) PIR configuration; (b) One PIR is active; (c) Two PIRs are active; (d) No PIR is active. . . . .	43
IV.6	Sensing regions of a PIR sensor. . . . .	44
IV.7	(a) Sit and lie map. (b) Walk and stand map. . . . .	47
IV.8	Particle filter diagram. . . . .	48
IV.9	Result of localization and tracking. Solid lines denote ground truth and dashed lines denote estimated trajectories. . . . .	53
IV.10	(a) Localization and tracking without activity recognition. (b) Localization and tracking when using only PIRs. . . . .	54
IV.11	Another result of localization and tracking. . . . .	55
V.1	Hydration monitoring algorithm. . . . .	59
V.2	The setup of the AutoHydrate system. . . . .	63
V.3	A two-dimensional representation of emotional state. . . . .	67
V.4	Emotion elicitation using Oculus Rift. . . . .	70
V.5	An ECG signal of 4 seconds in length. . . . .	70
V.6	ECG time-delayed mutual information. . . . .	71
V.7	False nearest neighbors test used for choosing dimension. . . . .	71
V.8	ECG time-delayed embedding state space. . . . .	72
V.9	Recurrence plot of ECG signal with non-negative emotion. . . . .	73
V.10	Recurrence plot of ECG signal with negative emotion. . . . .	73
V.11	(a) Original valence values; (b) New label with two categories: Negative and Non-negative. . . . .	74
V.12	Classification result on our own dataset. . . . .	77
V.13	Classification result on the RECOLA dataset. . . . .	77
V.14	Classification result on the DECAF dataset. . . . .	78

V.15	A traditional facial expression recognition system. . . . .	79
V.16	Sample images of facial expression databases. . . . .	82
V.17	Transfer learning method. . . . .	83
V.18	Transfer learning using GoogLeNet for 7-emotion recognition. . . .	84
V.19	Transfer learning using GoogLeNet for negative emotion recognition.	86
V.20	A setup to simulate human movement while human face is captured.	88
V.21	Examples from the test set including facial images with different yaw angles. . . . .	89
V.22	Sample dataset of ECG signals and human activities. . . . .	90
V.23	Normal heart rate ranges corresponding to activity levels. . . . .	90
V.24	Labeling the data based on the normal ranges of heart rate. . . . .	91
V.25	Decision fusion diagram used to recognize the negative emotion. . .	92
V.26	A summary of the fusion rule. . . . .	96
VI.1	Closing the loop with a robot assistant. . . . .	99
VI.2	Robot assistant. . . . .	100
VI.3	Interactive conversation method. . . . .	101
VI.4	Natural language interpreter. . . . .	101
VI.5	The loop of hydration reminder. . . . .	103
VI.6	Hydration reminder application of the robot assistant. . . . .	104
VI.7	Emotion regulation with the robot assistant. . . . .	105
VI.8	Emotion regulation application of the robot assistant. . . . .	106
VI.9	Interactive conversation evaluation. . . . .	108
VI.10	Experimental result. . . . .	109

## CHAPTER I

### INTRODUCTION

In this chapter, we first present the motivation of this work. Then, the challenges, objectives, and contributions are described. The organization of this dissertation is outlined at the end of this chapter.

#### I.1 Motivation

Based on the statistics from the National Institute on Aging [1], the world population is aging rapidly. There are around 125 million people aged 80 and older in 2015 all over the world. This number is expected to triple by 2050. And according to the U.S. Bureau of the Census [2], by 2020 there will be more than 20 million people whose age is over 75 as shown in Fig. I.1. However, human functionalities are more likely to deteriorate at an older age, which can lead to diseases such as heart problems, stroke, respiratory disorder, dementia, etc. With the demographic shift, there is a need to keep older adults healthy, functionally able and independent. Elderly people are an important asset to society. The life experience and wisdom they have gained over the years make them a vital social resource. Providing older people with age-friendly physical and social environments helps them live an independent life and also improves the active participation, thus maximizes their contribution to the society [3].

Elders have the option of going to adult day care, long-term care, nursing homes, hospice care, and home care. Even though all these options support the health, nutritional, social support, and daily living needs of older adults, the feeling of in-

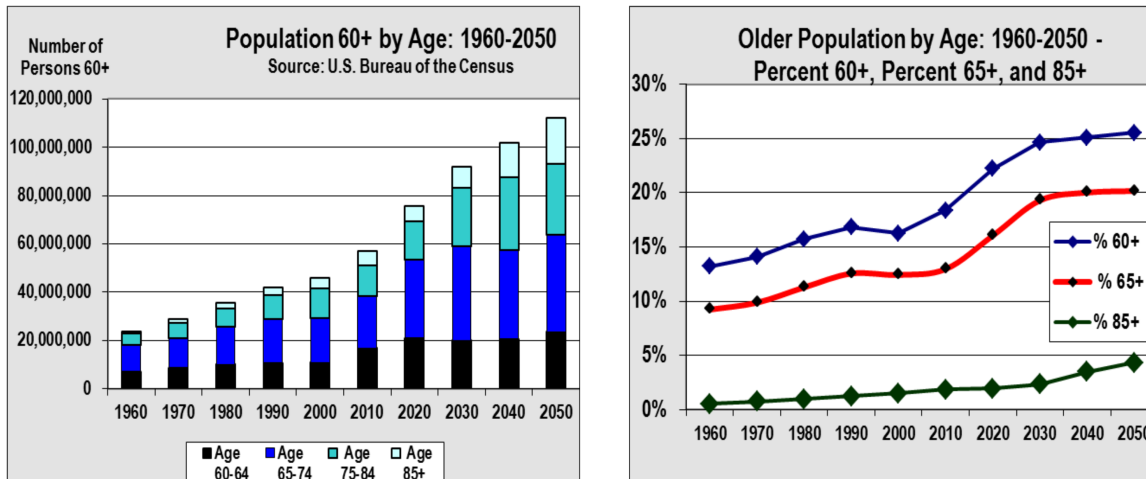


Figure I.1: U.S. Population 60+ by Age: 1960-2050.

dependence is lost. Elders would prefer to stay in the comfort of their home where they feel more confident than moving to expensive adult care or healthcare facilities as revealed in the surveys conducted in the United States [4] and Europe [5]. Hence if older adults are able to complete self-care activities on their own, maintaining independence can provide them with a sense of accomplishment and ability to maintain independence longer [6]. The best way to support them is to provide a physical environment that promotes the development and use of innovative technologies that encourage active aging [3]. As a result, building a Smart Home Environment has recently become a main research focus. Home automation that provides security, entertainment, and energy conservation and tailored to the elderly would be the perfect use of technology to achieve this. A smart home for elderly has to satisfy not only those demands but also the need for health monitoring.

According to [7], there were more than 360,200 cases of out-of-hospital cardiac arrest in the US. Seventy percent of the cases occurred in homes or residence, but only about 46% of people get the immediate help before professional help arrives. There is obviously a need for health monitoring that can continuously measure physiological signals, especially heart signal, of human and give timely alert.

In order to continuously collect vital signals such as electrocardiogram (ECG),

blood oxygen level (SpO<sub>2</sub>), or breath signal, while the older adults stay at home, a solution of using smart garments has been proposed by researchers in both academia and industry. A smart garment should deliver the functionality while ensuring comfortability. Its functionality can be achieved by using medical sensors and a small computing platform, which are integrated in the garment, to collect, process, store and transmit the data. To ensure comfortability, the garment should be designed in such a way that it operates seamlessly, dons and doffs easily, and is washable. Among the signals collected, ECG is a crucial one that represents the electrical activity of the heart. It is possible to utilize ECG for diagnosis in terms of physical health problems such as cardiovascular diseases, disorders of heart rhythm, sudden cardiac arrest, etc. [8, 9], and even mental health problems such as anxiety disorders, mood disorders, depression, etc. [10, 11]. The ECG signal is sensitive to motion-induced noises, therefore, if unreliable signals are used for processing, there is a high possibility of getting false alarms. If human body activity can be recognized and human location can be estimated, useful context information will be provided and may be used as prior knowledge for health diagnosis. For example, it is normal that the heart rate increases when the human is running. But it may be abnormal if the heart rate suddenly increases while taking a rest, and in this case, medical intervention should be considered.

Therefore, to enable context-aware health monitoring, it is essential that the location of the older adult can be tracked. By knowing the human position along with daily activity, we can achieve a better understanding of the older adult regarding where he/she is and what he/she is doing. Based on that, more accurate health monitoring can be achieved. The combination of a smart home and a smart garment can help achieve this goal. In the case of anomalies, medical interventions should be provided by a caregiver or a doctor. For those older adults who live alone at home, a robot assistant, which is aware of the human health status and able to interact with

the human, can be used to provide timely intervention.

## **I.2 Challenges**

In this section, we discuss some research challenges in building a home healthcare system that is able to monitor the human context information and health status while providing interventions.

### **I.2.1 Hardware setup**

The smart home environment should consist of low cost and unobtrusive sensors, with a capability of human localization without violating the resident's privacy. The localization system needs to be easy to use and maintain.

The wearable unit, including non-invasive physiological sensors and a computing engine, should be well designed to be embedded in a garment. The garment should be able to provide reliable vital signs while making the wearer feel comfortable. Since the battery life is a critical issue for embedded systems, power awareness should be taken into consideration in hardware design.

### **I.2.2 Physiological data collection**

It may be difficult to recruit older adults to participate our experiments, in which they have to wear a smart garment and perform daily activities in our smart home environment. The hesitation to wear a garment with embedded electrical wires and circuits and the health condition of the older adults are the main obstruction. Moreover, useful datasets of older adults regarding their health concerns need to be continuously collected using environmental and/or wearable sensors, from which the data can be used to perform early diagnosis of medical problems. In order to build such datasets, it requires a large population of older adults from whom the physiological data can be collected.

### **I.2.3 Robot development**

A robot assistant should be developed and equipped with the capabilities of talking with the human, approaching to the human, and understanding the human context information and health status to have proper responses. This requires the robot appearance and human-robot interaction to be carefully designed. Also, it needs a system to support efficient information transmission, which can enable the integration of the robot into the smart home environment.

## **I.3 Objectives**

In this dissertation, we focus on developing a Cloud-Based Smart Home Environment (CoSHE) which can provide both physiological and contextual information of the resident, while a robot assistant can provide healthcare services. The objectives of this proposal consist of:

- Developing a smart garment that is comfortable to wear for continuous collection of physiological signals.
- Developing a cloud-based system to store and process data collected from both home sensors and wearable sensors.
- Developing algorithms to provide context information such as body activity and indoor human location.
- Developing algorithms to monitor human hydration status and recognize negative emotion.
- Developing an interactive conversation mechanism of the robot to assist human when health issues are detected.



## I.4 Contributions

The contributions of our work are depicted in Fig. I.2 and summarized as follows:

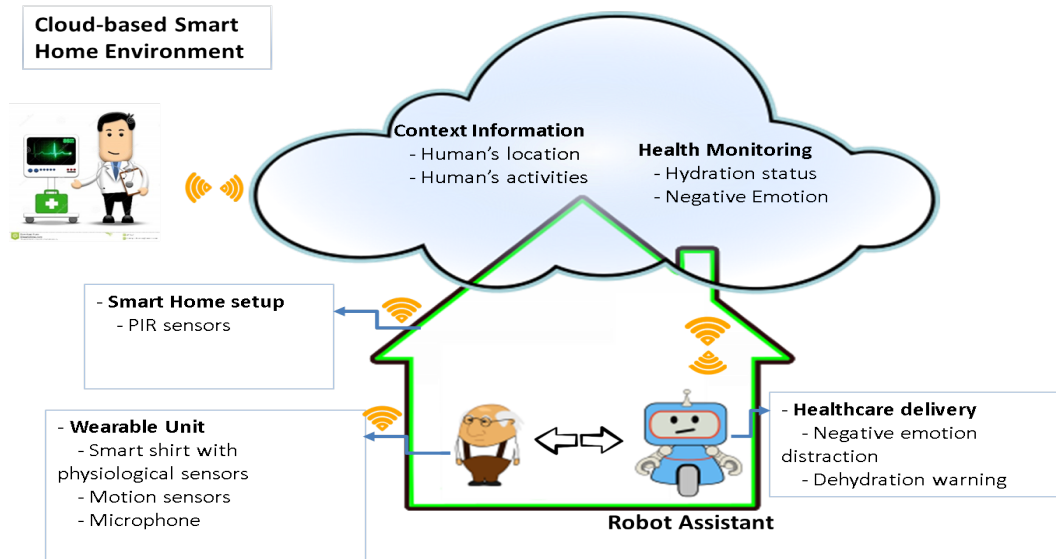


Figure I.2: System overview.

1. We have developed several wearable health-kits consisting of a sensor platform and microprocessors such as Arduino [12] and Intel Edison [13]. The sensors can collect physiological signals including electrocardiography (ECG), oxygen saturation (SPo2), and respiration. The kit is embedded in our own garment designs in which textile electrodes and conductive fabric and threads are used. The garment prototypes are washable and comfortable to wear for long-term data collection.
2. We have developed a cloud-based smart home environment consisting of three major components: (1) a smart home with passive infrared (PIR) sensors, (2) a wearable unit that consists of the smart garment and other sensors such as an Inertial Measurement Unit (IMU), a smartwatch and a microphone, and (3) a private cloud infrastructure. With the private cloud, information is stored and shared among different components, which makes it a complete ecosystem

supporting healthcare applications. Remote caregivers and the robot can access the human physiological data along with the contextual information.

3. We have developed algorithms to extract context information such as human body activity and human location from home and wearable sensors data. Data from the IMU is used to detect body activity and estimate the velocity and heading angle. The array of PIR sensors on the home ceiling is used to roughly estimate the human location. Along with a behavior-based map, human motion data and home sensor data are then fused in a particle filter to derive the human location.
4. We have built a health monitoring system based on the CoSHE environment, which is able to recognize negative emotion using the fusion of ECG signal and facial expression, and monitor hydration based on drinking sound recognition and activity level estimation.
5. We have built a healthcare delivery system using a robot assistant that proactively engages the human in conversation when the health monitoring system raises a red flag. Through interactive conversations, the robot can remind the human to drink water and help improve the human emotion when the human is in negative emotion.

## I.5 Outlines

The organization of this dissertation is as follows:

- Chapter 1 presents the motivation and contribution of this research.
- Chapter 2 describes the design of the smart garment.
- Chapter 3 introduces the CoSHE system and its components.

- Chapter 4 presents the context information that CoSHE provides and the methods of recognizing human body activity and tracking human location in an indoor environment.
- Chapter 5 introduces healthcare applications enabled by the CoSHE system, including hydration monitoring and emotion recognition.
- Chapter 6 introduces opportunities for healthcare delivery through a robot assistant.
- Chapter 7 presents the future works and concludes the dissertation.

## CHAPTER II

### SMART GARMENT DEVELOPMENT

This chapter first presents the smart garment we develop and the wearable sensors we use to collect physiological signals. Next, it describes the data acquisition mechanism and shows some acquired data.

#### II.1 Introduction

As mentioned in the previous chapter, the adults are likely to have more health-related issues when aging. The age factor often affects their mobility and reduces the frequency of clinic visits, not to mention difficulties they may confront with if they live in rural areas. Apparently, there is a demand for a medical device that can continuously measure physiological signals, process and give alerts to the older adults while they stay at their own home. In addition, the device needs to be easy to use and comfortable to don, doff and wear for a long time without affecting their daily activities. To satisfy those requirements, a smart garment seems to be one of the best solutions. In the next section, a literature review of smart garments is presented.

#### II.2 Related Work

There are currently a few products available on the market, such as Healthwatch [14] and Hexoskin [15], which can be used in medical or sports applications. Wearable electrodes are embedded in those smart garments in order to continuously monitor physiological signals while maintaining wear comfort. The signals collected normally include ECG (Electrocardiogram) and respiration, from which features such as heart

rate and respiration rate can be extracted. There are also research projects performed by universities and institutes, which focused on developing textile sensors using conductive fabrics and conductive inks. In [16], Paul et al. developed biomedical sensors using dry electrodes and conductive ink to collect ECG and EMG (Electromyography) signal. In [17], recent advanced techniques in smart textile was surveyed thoroughly, and both advantages and disadvantages of each technique were discussed, for example flexibility, ergonomics, low power consumption, integration and autonomy. In [18], Ankhili et al. developed a washable ECG monitoring underwear able to record and transmit wirelessly the ECG signal. However, in order to make those products or prototypes to be widely used, there are improvements the need to be made. To ensure comfort, the ECG electrodes in such products are designed so that they do not stick on human skin, and do not tightly press the skin. The ability of electrodes to conduct the signal from the body to the health monitoring system is very sensitive to body movement, and variation in the space between the electrode and the skin creates artificial noise within the signal. Therefore, the artificial noise in these signals might hide important features for analysis or could easily lead to a false alarm. Classifying movement is an essential component of the process of determining the relationship between textile electrode signal quality and the subject's physical movements, and provides an important foundation for the development of viable medical-quality textile-based sensing systems. In the next sections, we present two smart garment prototypes and the acquired data from the garments.

### **II.3 Garment prototype 1**

The health kit, as shown in Fig. II.1, consists of a sensor platform and an Arduino with Bluetooth integrated. The sensor platform is a circuit board which physiological sensors such as ECG, SPo2, air flow, and respiration and accelerometer sensors can connect to. This figure also shows several kinds of sensor that can collect different

physiological data from the human body.

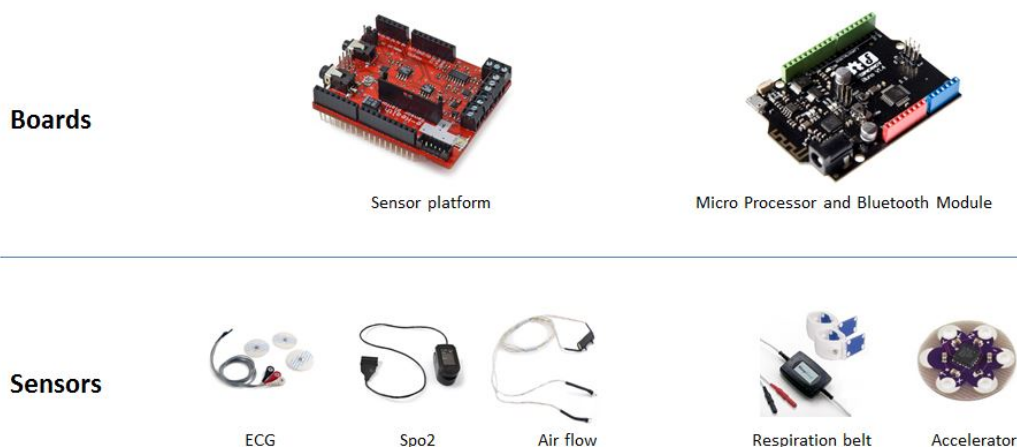


Figure II.1: Health kit.

### II.3.1 ECG Sensor

The ECG sensor includes three disposable electrodes which often include a metal snap, an Ag/AgCl sensing element and an adhesive hydrogel. One end of the electrode is attached to the human skin at locations such as under right clavicle (positive electrode), under left clavicle (neutral electrode) and at lower left abdomen (negative electrode) as shown in Fig. II.2. The other is connected to the sensor platform where the analog ECG signal will be amplified and transmitted to the Arduino board where the signal will be converted from analog into digital. By using these electrodes, ECG lead-II is collected and it reflects the electrical activity of the heart.

### II.3.2 SPo2 Sensor

The SPo2 sensor measures the amount of oxygen dissolved in the blood, based on the detection of Hemoglobin and Deoxyhemoglobin. Two different light wavelengths, which are the red light and infrared light, are used to measure the actual difference in the absorption spectra of HbO<sub>2</sub> and Hb. Most SPo2 sensors work on extremities

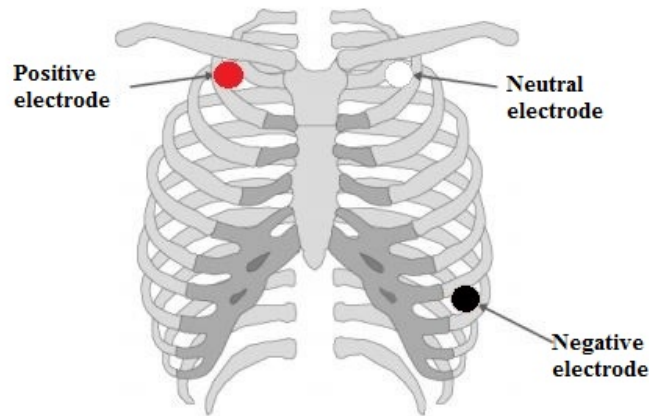


Figure II.2: The positions of the electrodes.

such as fingertip, toe, or earlobe. The SPO<sub>2</sub> sensor in our setup is used on the finger, and its outputs include two values which are the percentage of hemoglobin molecules in the arterial blood and heart rate. That percentage is considered normal when it is in the range of 95 to 99. If the value is between 88 and 94 percent, the hypoxic drive problem may be present. And if the value is 100 percent, it can indicate carbon monoxide poisoning. These digital values are then transmitted to the sensor platform through wires.

### II.3.3 Respiration Sensor

The respiration sensor is an inductive plethysmography belt that is worn over the chest to measure chest movement when human breaths in and out. The belt goes with an inductive interface cable which is able to convert inductance changes into voltage. The mechanism is based on Farada's Law in which a current applied through a loop of wire generates a magnetic field normal to the orientation of the loop, and that a change in the area enclosed by the loop creates an opposing current within the loop directly proportional to the change in the area (Lenz's Law). The output voltage is weak so it needs to be amplified for the sensor platform to acquire [19]. By using the respiration sensor, changes in chest volume and respiratory rate can be estimated.

Thus, abnormality in respiratory rate can be detected, which is one of the earliest indicators of physiological instability.

### II.3.4 Airflow Sensor

The air flow sensor is used to measure nasal thermal airflow changes as well as the nasal temperature air. This sensor has a flexible thread which fits behind the ears, and two prongs which are placed in the nostrils to measure breathing. The airflow sensor and respiration sensor both measure breath but in different places of the body and different phases of respiration circulation. Normally, there should be a constant delay between those two output signals. By using both, we can validate the signals as well as detect abnormality in some senses.

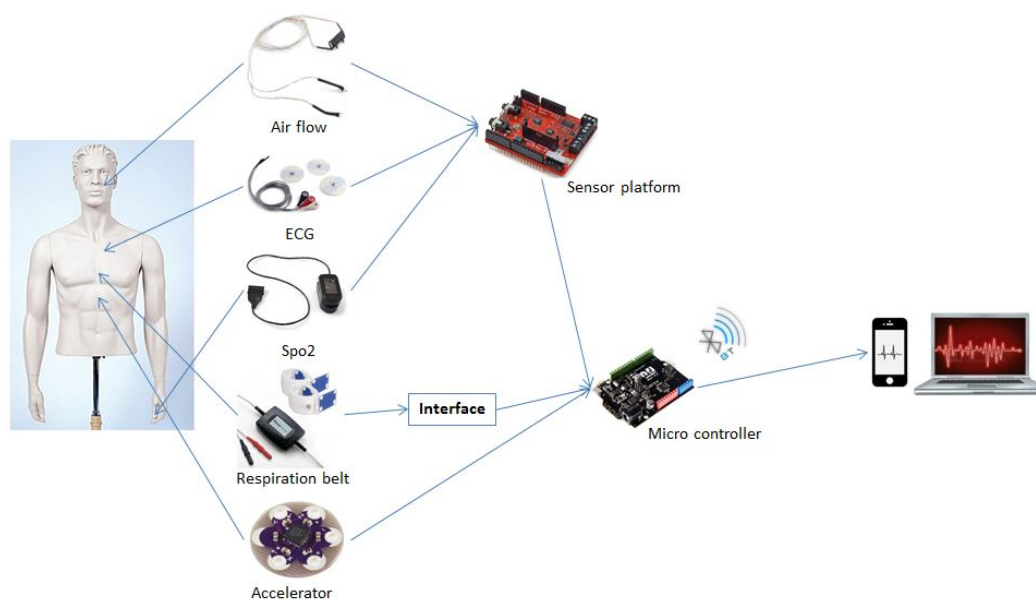


Figure II.3: How the health kit works.

### II.3.5 Garment design

After the hardware is tested, a smart garment prototype is developed which integrates all electronic parts so that physiological signals can be collected when it is worn. To make it more comfortable, the standard ECG electrodes are replaced by textile ones,



and the normal electric wires are replaced by conductive threads. Several types of conductive fabric are tested such as conductive felt fabric and conductive metalized nylon fabric. There is a trade-off between picking up good signals and the comfort. Because good signals are acquired if the contact between electrodes and skin is durable, which can be achieved by using the sticky material or pressure to hold the electrode, it may cause red itchy or rash on the skin. After several testing, conductive metalized nylon fabric is chosen because it has great conductivity and is easy to sew by the fusing method. The smart garment prototype is shown in Fig. II.4.

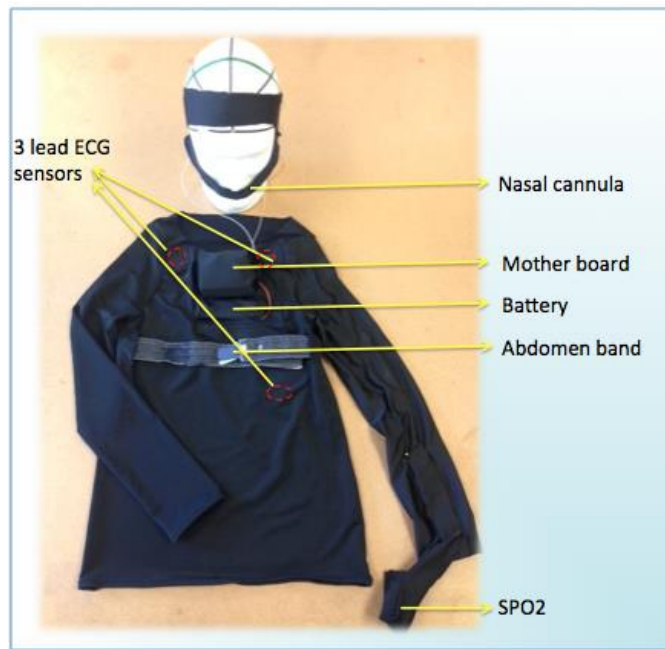


Figure II.4: Smart garment prototype 1.

With this prototype, a single lead ECG signal is collected from the garment using Lead-II configuration. The use of single lead ECG is shown to provide sufficient information for detecting heart rate, studying arrhythmias and heart rate variability which can be used by doctors for monitoring and diagnosis. The signal is amplified and sampled at an optimal sampling rate of 200Hz. The respiration belt is an inductive transducer which measures the changes in thoracic or abdominal circumference during respiration. These measurements can indicate inhalation, expiration and breathing

strength which can be used to derive the respiratory rate. The respiration data is sampled at 20Hz.

## II.4 Garment prototype 2

Another smart garment prototype is designed to embed additional sensors. Its main purpose is similar to the previous version, i.e. to pervasively collect physiological signals and activity data. The data collected from this system includes Electrocardiogram ECG, Acoustic Activity and Body Activity. The signals are sampled and framed using The Intel Edison compute module. The data is timestamped to allow synchronization with ambient sensors in the smart home. Our wearable system components are shown in Fig. II.5. The additional components are described as follows.

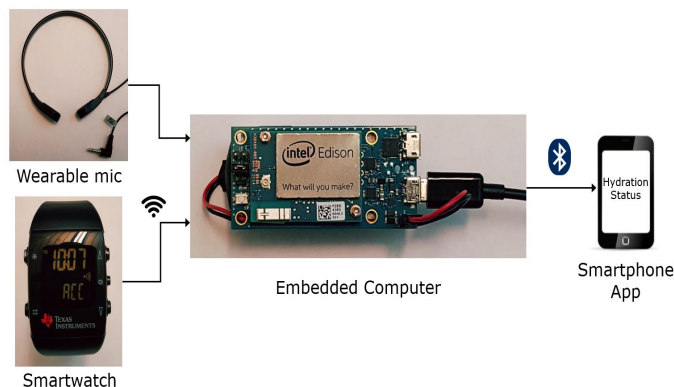


Figure II.5: Wearable unit components.

### Smartwatch:

The eZ430-Chronus smartwatch is used to collect hand activity data. It uses the CC430F6137 MCU with a 915MHz wireless transceiver. The watch has integrated 3-axis accelerometer with a measurement of  $\pm 2g$ . The accelerometer data is wirelessly transmitted to the home gateway at 33Hz.

### Acoustic Sensor:

A throat microphone is used to record audio signals from the throat area. The



Figure II.6: Smart garment prototype 2.

throat microphone absorbs vibrations from the wearer's throat instead of sound signals which allows picking up sounds in extremely noisy and windy environments. It has a sensitivity of  $-66\text{dB} \pm 3\text{db}$  and can pick up sound signals from 20Hz-16000Hz which is sufficient for detecting drinking and other activities. The acoustic signal is sampled at a rate of 16000Hz.

#### **Embedded Computer:**

The Intel Edison compute module is used to collect and process signals, classify drinking and body activities. It is a high performance embedded platform with a dual-core 500MHz Intel Atom CPU and WiFi and Bluetooth support with minimal power consumption [13]. In addition to its higher performance compared to other embedded devices, its low profile makes it an ideal solution for wearable applications. A USB sound adapter with the 48K/44.1KHz sampling rate and 16bit ADC/DAC is used to interface the throat microphone to the embedded controller. A 9V alkaline battery is also used to power the system.

#### **Garment design:**

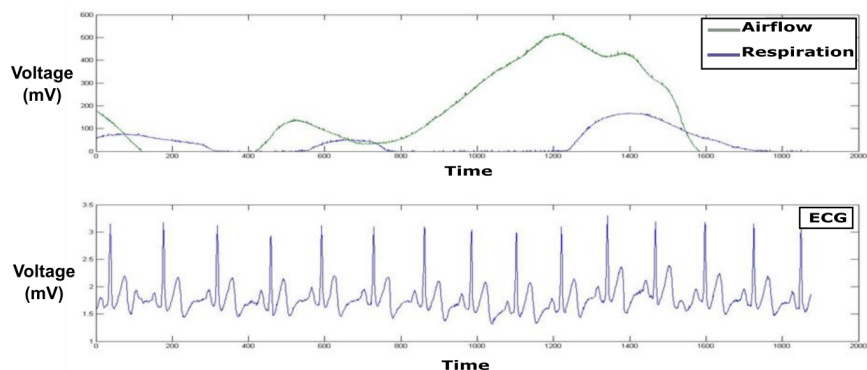
The prototype is designed as a slimming compression tank top, which allows the textile electrodes to have a better contact with the human skin. It has a pocket in

front of the chest to store a 3D box containing the electrical board as shown in Fig. II.6.

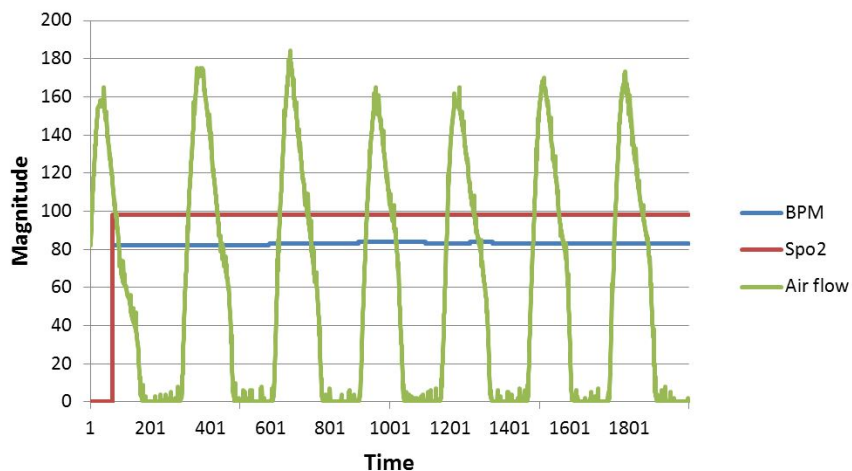
## II.5 Signal Acquisition Result

### II.5.1 Signals collected using standard electrodes

All sensors are connected to the sensor platform which is a conditioning circuit, and the analog signals are converted to digital signals on the Arduino board and transmitted to a mobile device or computer using Bluetooth as shown in Fig. II.3. These physiological data are synchronized with data from smart home in the server. In Fig.



(a) 10-second length signal of ECG and respiration.



(b) 10-second length signal of heart rate, Spo2, and airflow.

Figure II.7: Physiological signals collected using wearable sensors.

II.7a, physiological signals such as the ECG signal, airflow signal and respiration sig-

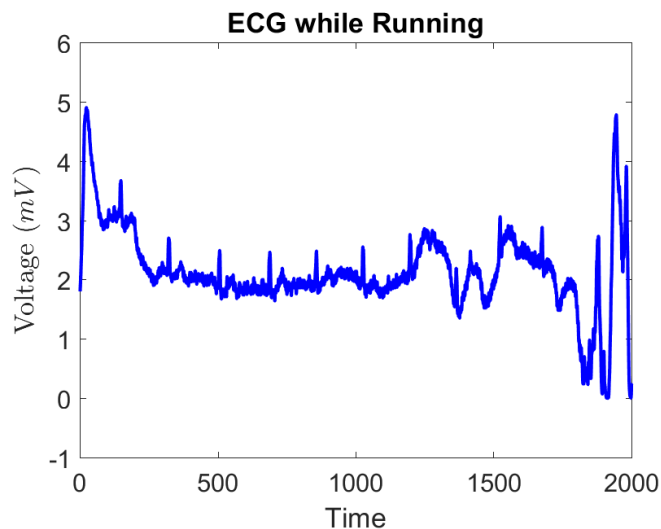


Figure II.8: ECG signal affected by artificial noise.

nal in 10 seconds frame are shown in real-time in Matlab. These signals are collected when the subject is in sitting position and at a sampling rate of  $250\text{ Hz}$ . Fig. II.7b displays the sample result showing breath, heart rate and oxygen concentration in blood while the subject is moving. The ECG signal quality is easy to get affected by artificial noise. For example, when the human walks, jogs or runs, the collected ECG becomes worse as shown in Fig. II.8. It is because the contact between ECG electrodes and human skin is affected by the movement. All other sensors return relatively good signals regardless of movement intensity. ECG, however, has more potential for analysis, and many studies have been done using ECG for health problem diagnosis. Most of them are conducted when the subject is resting, e.g. lying or sitting, partly because of ECG signal's sensitivity to movement noise. Therefore, analyzing and processing ECG signal as well as other physiological data when human is moving, e.g. doing daily activities, is a challenge for researchers. An idea to overcome this problem is to combine context information. By knowing where the human is and what the human is doing, we can somehow adaptively filter physiological signals to reduce the impact of the movement noises.

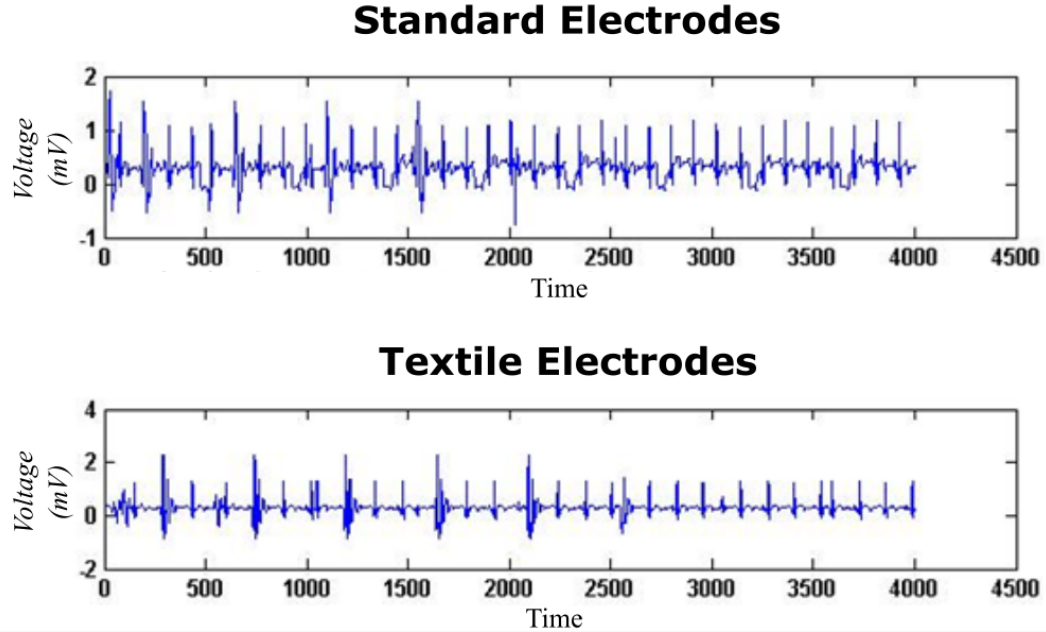


Figure II.9: ECG signals collected by using standard electrodes and textile electrodes.

### II.5.2 Signals collected using textile electrodes

An experiment is performed to compare the ECG signal collected using the garment with embedded textile electrodes to the one collected by standard electrodes on the same subject at the same time. The data are collected from 3 male subjects sitting on a chair for multiple trials, and each trial lasts 5 minutes. Sample acquired data are shown in Fig. II.9. We use a Pan Tompkins algorithm [20] to recognize the R peaks of the signals. The heart rate (beat per minute) can be extracted by using the formula  $HeartRate = (60 \times fs)/RR$ , where  $fs$  is the sampling frequency of the ECG signal and  $RR$  is the R-R peak interval. Heart rate variability (HRV) is another important feature extracted from ECG signal. It measures the changes in time between R-R intervals and can be given an estimate of a person's autonomic nervous system. The HRVs are calculated by taking the root mean sum of the squared differences between RR intervals in a time frame of 4-second from the both systems. The Pearson's correlation is used to measure the strength of the linear relationship between those two HRVs. The correlation coefficient is calculated to be 0.98, which means the HRVs

from both systems have almost perfectly positive linear relationship. The change in the mean is calculated to be 5 ms. This preliminary result shows that the ECG signals collected from the textile electrodes are acceptable in terms of HRV extraction.

## II.6 Summary

In this chapter, we introduced two garment prototypes to collect physiological signals, body movement signal, and throat sound. The textile sensors and conductive threads are embedded in the garment to facilitate its functionality and comfort. An Arduino and an Intel Edison are used to process and transmit the acquired signals. The garments with textile electrodes can be used in experiments while keeping the subjects comfortable when wearing them. The physiological signals collected from the garments include ECG, respiration, airflow, and Spo2. In addition, motion signal and throat sound can also be collected. It was demonstrated that the ECG signals collected using textile electrodes are reliable to extract the HRV, which is an essential feature used in many studies. Since the garment prototypes are designed for human to wear for long term health monitoring, more tests should be conducted to make sure the comfort is maintained while good signals are continuously acquired.

## CHAPTER III

### CLOUD-BASED SMART HOME ENVIRONMENT (CoSHE)

In this chapter, we present a Cloud-Based Smart Home Environment (CoSHE) for home healthcare. A smart home environment is a setup with ambient sensors to provide context information. The cloud infrastructure is then described as a means to provide a data storage function and real-time data access for remote caregivers.

#### III.1 Introduction

Besides dealing directly with physiological information, recognizing the context of residents, such as what they do, or where they are in the home can provide important information for health status evaluation. For example, time spent in sedentary behaviors, such as sitting or lying down, may be an indicator of cardiovascular disease [21]. Another application can be for patients recovering from a stroke where monitoring of re-learning everyday living skills such as eating, going to the bathroom, and dressing can facilitate evaluation of rehabilitation and recovery process [22]. The CoSHE system is built to enable health monitoring by providing physiological signals with context recognition. This enables healthcare professionals to study daily activities, behavioral changes and monitor rehabilitation and recovery processes.

#### III.2 Related Work

Advances in the development of telehealth and portable hospital-grade technologies allow healthcare providers to deliver in-home medical services that were previously available only in hospitals [23]. According to Grönvall et al. [24], older adults with



severe heart conditions can conduct self-heart diagnosis by measuring their weight, blood pressure, pulse, and cardiac activity and report to the hospital through online questionnaires. Hospital nurses will contact patients if something unexpected happens. However, the significant amount of work required from the nurses reduces the system efficiency. The ACCENT (Advanced Component Control Enhancing Network Technologies) Home Care System [25] is another system for telehealth and telecare. Even though this system has been originally designed for general applicability, its idea and architecture cover most of the essential components needed for telecare services in home environments. This system was developed with the following issues in mind: management, conflicts, policies, and interface of home care. The recent technical innovations have led to the development of smart sensors, cloud computing, and big data analysis, which have been applied to home healthcare. Real-time telemonitoring of diverse bio-signals and video consultations are two main growing approaches for home-based care, both in research and commercial systems [26]. The existing smart home projects often combine environmental and wearable sensors. For example, environmental sensors may include PIR (Passive Infrared Sensor), RFID (Radio Frequency Identification), ultrasonic, camera, microphone, or pressure sensors. Wearable sensors are often used in smart healthcare applications for measuring body temperature, respiration, cardiac activity (ECG - Electrocardiogram), blood oxygen saturation (SpO<sub>2</sub>), perspiration (GSR - Galvanic Skin Response), muscle activity (EMG - Electromyogram), brain activity (EEG-Electroencephalogram) and motion (accelerometer and gyroscope).

Numerous smart home research projects have been conducted in the US, Europe, and Japan with many of them focusing on different approaches to providing healthcare [27, 28, 29, 30, 31, 32, 33]. The Aging In Place project at the University of Missouri started in 2000, and it has been constantly developed with the main goal of meeting the desire of older adults to remain independent in their own home [27].

An integrated sensor network is installed to monitor health and enhance the safety of older adults. Additionally, functionalities are added to supplement registered nurses by giving an alert when abnormal changes in the health signals of an older adult are detected [34]. Smart Medical Home is another project conducted by the University of Rochester, which is set up in a five-room house equipped with environmental sensors, bio-sensors, and a video camera system. It is designed to maintain health, diagnose and treat disease. Other projects such as the Aware Home at Georgia Institute of Technology [28], the Gator Tech Smart House at University of Florida [29] and the CASAS at Washington State University [30] are all aimed at designing a smart home-based health monitoring platform. These kinds of projects are often equipped with typical devices such as a motion tracking system, an ultrasonic location tracking system, a smart floor, and a smart bed, along with a software system providing artificial intelligence to assist the elderly. They can track the location, recognize the activities of the resident, assist those with cognitive difficulties, or help with diabetes management. There are many other smart home projects including the CareLab in Germany, ProSafe in France, ENABLE in the Netherlands, CareNet in the UK and Ubiquitous Home in Japan [31]. Recently a project called VictoryaHome was funded by the European Ambient Assisted Living Joint Program [32] which is focused on assisting older adults and caregivers in their homes. This system can help older adults live more independently by providing services including monitoring health, safety and facilitating social contact. The system can serve as a medication reminder, check activity, and detect falls by using a wearable smart belt-clip. Other functions such as virtual visit are achieved using a robot named Giraff [33]. This enables older adults and their remote caregivers or relatives to make video calls, and the caregivers can remotely control the robot to move around and check the older adult through the robot's camera. The VictoryaHome Cloud was established to facilitate telecommunication and data transmission for all the previously mentioned functions. However, no

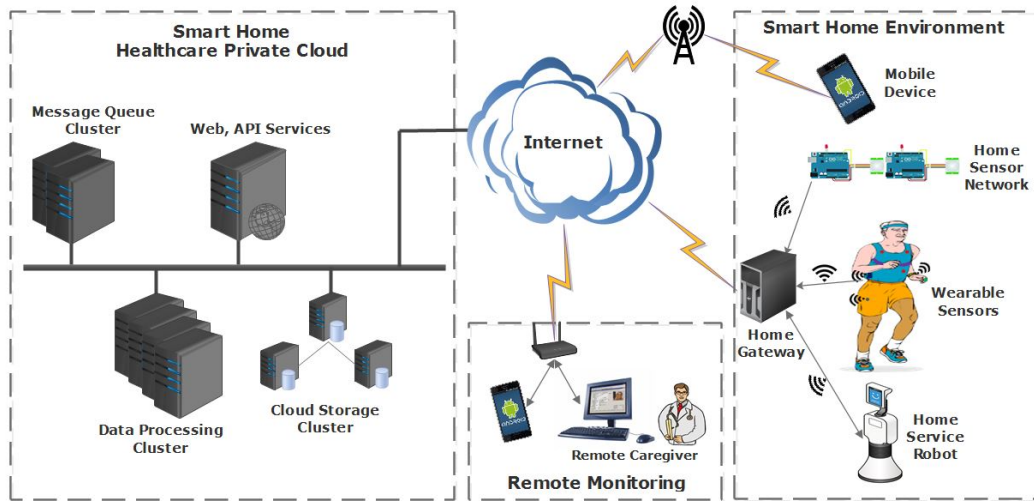


Figure III.1: The architecture of the Cloud-based smart home healthcare system.

environmental and medical sensors are used in the VictoryaHome, and as a result, it cannot monitor real-time health status and context information.

The above review of existing systems, which is summarized in TABLE III.1, shows that many of them can only provide a limited amount of information regarding the resident's health condition. Some of them lack the ability to provide health monitoring or healthcare delivery. Some projects do not equip either physiological or home sensors. Essentially, complete information is critical to accurately assessing the patient's health condition. This can be achieved by developing a comprehensive system which integrates wearable physiological sensors to collect bio-signals, a smart home environment with unobtrusive environmental sensors to understand the person's daily activities and a cloud platform to aggregate the data and provide analytics. CoSHE is built as a comprehensive system which supports all of the above functions and further provides a platform for health monitoring, early detection, and prevention of different health problems.

Table III.1: Smart home projects.

Project	Wearable sensor	Home sensor	Robot assistant	Remote monitoring
The Aging In Place [27]	-Accelerator -No physiological sensors	-Mat, bed -Cameras	No	No
Gator Tech Smart House [29]	-Actigraph watch (accelerator and pulse sensor)	-Door sensor -Infrared motion -Cameras	No	No
CACAS [30]	No	-Door, bed sensor -Infrared motion	No	Yes
Victorya Home[32]	-Smart belt-clip (accelerator) -No physiological sensors	No	Yes	Yes
CoSHE	-Smart shirt (ECG and respiration sensor) -Smartwatch, IMU (accelerator and gyroscope)	-Passive Infrared (PIR)	Yes	Yes

### III.3 System Architecture

The cloud-based smart home environment consists of three major components: a smart home setup, a wearable unit, and a private cloud infrastructure. The wearable units are used to collect physiological and body activity information through non-invasive, wearable sensors. Smart home ambient sensors are used for collecting motion and activity information from a human subject as discussed in Chapter 2. Data from the wearable units and environmental sensors are processed by a home gateway where pre-processing, indoor localization and activity recognition algorithms are implemented. The data is then sent to our private cloud infrastructure which provides users and remote caregivers with comprehensive physiological data with context information. The following subsections detail the three major components of our sys-

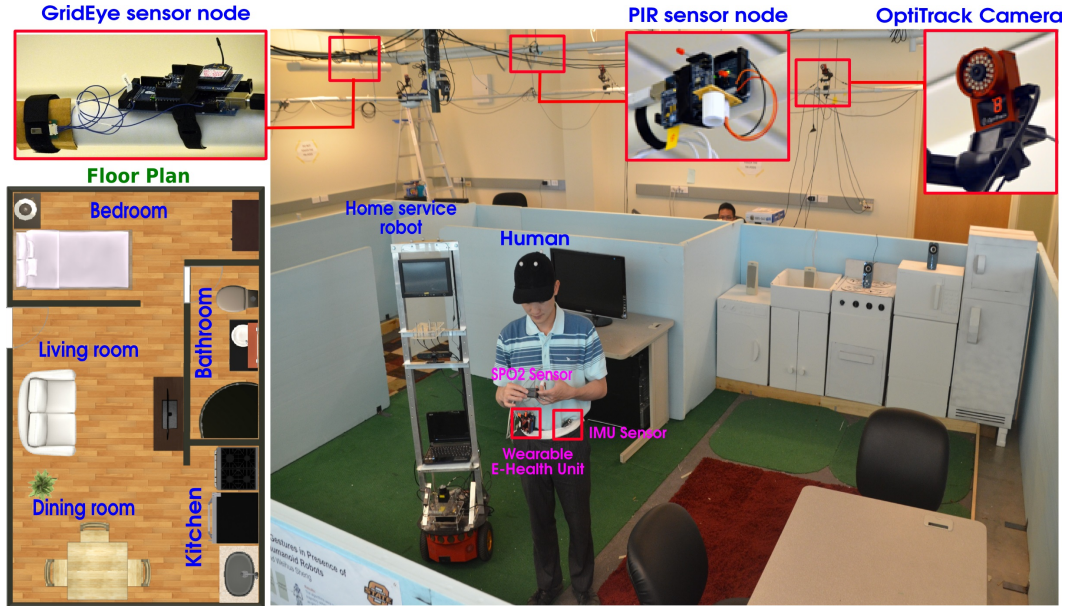


Figure III.2: Smarthome testbed setup.

tem architecture as shown in Fig. III.1.

### III.3.1 Smart Home

The Smart Home environment consists of a 16x22 ft. testbed equipped with environmental sensors, a human subject with wearable sensors, mobile devices such as smartphone and a home gateway. The floor layout and an overview of the smart home testbed are shown in Fig. III.2.

**Home Sensor Network:** The environmental sensors include a network of PIR sensors and Grid-EYE thermopile array sensors. These sensors are distributed on the testbed ceiling in a configuration and the data is transmitted to the home gateway through the ZigBee protocol. PIR sensors detect the IR radiations emitted by the heat of objects. The sensor in the device is split into two parts such that they detect the change in the radiation caused by motion in its field of view. The changes in the amount of IR radiation are reflected by the voltage generated. It is demonstrated in Fig. III.3, when a human walks in the field of view of the PIR sensor. The PIR sensor used in our research is Panasonic EKMC1601111 [35], and its specifications

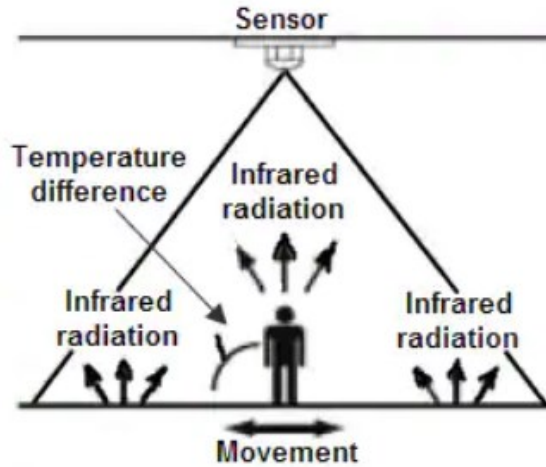


Figure III.3: The principle of PIR sensors.

are shown in Table. III.2

Our experimental setup uses the OptiTrack Camera system as shown in Fig. III.2 for evaluating our context recognition and classification algorithms against the ground truth. The smart home setup and sensor placement, sensor calibration and operation are presented in [36].

### III.3.2 Cloud Infrastructure

The Smart Home Healthcare Cloud is implemented using the Private Cloud Infrastructure we built in our lab. This system is set up using the open source Cloud Orchestration Software, OpenStack Juno. Three server nodes and a storage server

Table III.2: PIR Panasonic EKMC1601111 specification.

Item	Performance characteristics
Detection performance	Standard detection type
Current consumption (in standby mode)	170A
Lens color	White
Detection distance	5m
Output type	Digital
Output current (during detection period)	100A
Detection area (Horizontal)	82
Detection area (Vertical)	94
Detection zone	64

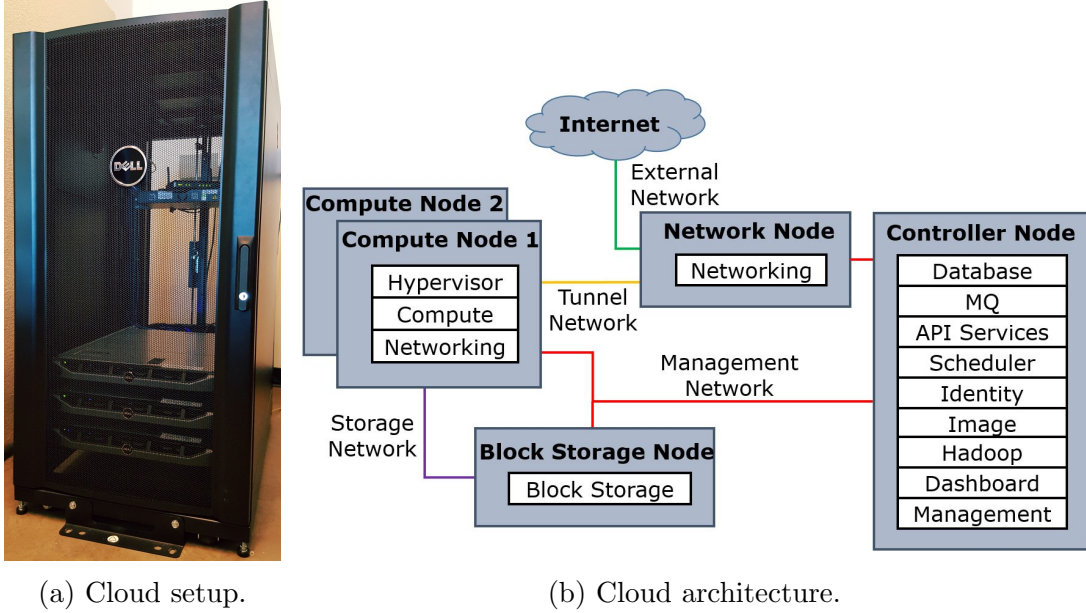


Figure III.4: Private cloud infrastructure

are used in the setup as shown in Fig. III.4. Most of the cloud management services including the message queue (MQ), authentication, databases, and networking are implemented on the Controller node. Two Compute nodes host KVM hypervisors and client services which will create a virtualized environment for instances. Block storage is implemented using the storage server to provide a persistent storage for running instances. Our cloud provides an Infrastructure-as-a-Service (IaaS) solution for different research and educational purposes. OpenStack Sahara project is also deployed to allow for rapid configuration, reliable auto-deployment and scaling of Hadoop Clusters in our cloud infrastructure. The Hadoop framework allows distributed processing of large data sets across clusters.

The Smart Home Healthcare Cloud is set up using a four-layered cloud architecture by adopting the Software as a Service (SaaS) cloud resource provisioning model based on our system requirements. The four layers and their functions are described as follows:

- Service Presentation Layer: is the interface for the Wireless Application Protocol WAP, Application Programming Interface API and WEB services. This

layer provides data access for remote users [37].

- Cloud Engine Layer: uses the Message Queue (MQ) cluster to distribute and manage jobs among the data processing and cloud storage clusters using message-driven scheduling [37].
- Data Processing Layer: uses the Data Processing cluster for implementing data mining algorithms on the physiological and context data. This will allow medical professionals and researchers to study relationships between daily activity and health condition, understand behavioral changes and cause-and-effect studies. The algorithms process sensor data transmitted from the front end and write it on the cloud storage using the MQ.
- Cloud Storage Layer: serves data resources for the cloud using the Storage Cluster [37]. These resources include user information, physiological and environmental data, and context information.

We used a hybrid data storage model, which is comprised of a relational database and non-relational database. MySQL is used as a Relational database for storing all structured data such as user information, sensor type, data format, and sampling rate. On the other hand, the Non-relational database MongoDB (NoSQL) is used to store sensor data. We used NoSQL because medical sensor data such as ECG and respiration have much higher temporal resolution than other sensors like PIR, Accelerometer, Spo2, etc. In order to have rapid access to these data for a real-time application, NoSQL provides the ability to adapt with distributed storage which is taken care of by Hadoop File System (HDFS) and MapReduce in the cloud. Our database tables include the information as shown in Table. III.3. The demographic and diagnosis information can be filled in by the care giver and the care taker will be provided access to the database through the WEB API.



Table III.3: Data stored in the CoSHE database.

<b>Demographic information</b>	<b>Context information</b>	<b>Health Status</b>
Full name	Location in home	ECG raw data
Date of birth	Body activity	Heart rate
Sex		Respiration rate
		Blood oxygen concentration
		Negative emotion
		Hydration status

### III.4 Summary

In this chapter, we presented our CoSHE system for home healthcare, which consists of a smart home and a cloud infrastructure. The smart home is equipped by a network of environmental sensors and a motion capture and tracking system. It provides an environment to conduct and validate experiments including but not limited to human localization and tracking, walking pattern classification, and abnormal patterns of daily activities detection. The cloud infrastructure provides a vast storage, large memory, fast processing units and a bunch of services, which supports IoT (Internet of Thing) and big data applications. In our research, the data from the environmental and wearable sensors can be sent to the cloud where they are stored and processed to provide meaningful information. The smart home, cloud infrastructure, and wearable units make CoSHE to be a comprehensive system that is able to provide context information and health status of the human and the real-time access from remote caregivers.

## CHAPTER IV

### CONTEXT AWARENESS IN CoSHE

In this chapter, we describe two types of context information provided by CoSHE, which are human body activity and human location.

#### IV.1 Body Activity Recognition

Human daily activity is a key factor that affects directly the accuracy of health status inference. Body movement can cause artificial noise while recording physiological signals thereby requiring data preprocessing. On the other hand, human activity should be taken into account while making decisions based on recorded physiological signals. For example, fast heart rate inferred by ECG signal is not necessarily considered tachycardia, if at the current moment human activity is recognized as vigorous running. But, if the heart rate is suddenly high while the human is still, there might be a health problem and further assistance may be needed.

##### IV.1.1 Related Work

Using wearable sensors to recognize body activity has been studied by many researchers for years and the most broadly used sensor are accelerometers [38]. The possible reason is that the accelerometers are inexpensive, low power consumption and it is now embedded in many gadgets such as smartwatches, smartphones, and other wearable technology devices.

In [39], Bao and Intille recognized body activities using five biaxial accelerometers, which were worn on different parts of the body such as the thigh, hip, and wrist. Af-

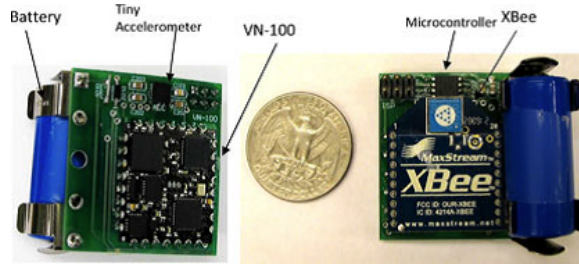


Figure IV.1: An inertial measurement unit (IMU).

ter accelerometer data were collected from 20 subjects, features such as mean energy, frequency-domain entropy and correlation of acceleration were extracted. Totally, 20 activities like walking, sitting, scrubbing, standing still, watching TV, etc. were recognized by using decision tree algorithm with the average accuracy of 80%. Kwapisz et al. [40] used phone-based accelerometers to classify body activities. Data was collected from 29 people with a cell phone was in the front pants leg pocket, while they performed different activities like walking, jogging, ascending stairs, and descending stairs. The sampling rate was 20 samples per second. The data was divided into 10-second segments to extract features such as mean, standard deviation, the time between peaks, etc. The classification accuracy of their body activity recognition was 90%. In [41], acceleration data from a chest sensor strap was used to classify 5 human activities such as walking, running, sitting, ascending, and descending. It resulted in 95.7% overall accuracy.

### IV.1.2 Hardware Setup

Our body activity recognition system uses a smartwatch and thigh worn IMU. The smartwatch was used since it can be conveniently worn by the user in a daily scenario. However it does not provide enough information for classifying complex activities as the thigh worn sensor and as a result, we used it for classifying basic daily activities whereas the thigh worn IMU was used for our location tracking application.

The IMU node we developed consists of a VN-100 orientation sensor module from

VectorNav, Inc.[28], for motion sensing, an XBee RF module for wireless communication, and a power management unit to prolong the battery life. The picture of the IMU node is shown in Fig. IV.1. The motion data include orientations (roll, pitch, and yaw), 3-dimension (3D) accelerations, 3D angular velocities, and 3D magnetic field. The sensor has a very small footprint, similar to a quarter. The typical operating voltage range is from 3.1 to 5.5V, and the power supply current is 65mA in the normal operation mode. This IMU is attached to the right thigh of the human subject to sense the human movement.

### IV.1.3 Methodology

A set of experiments is carried out on 6 subjects whose age ranges from 25 to 35 for classifying 10 activities. Each subject wears the IMU on the right thigh and performs different activities for 15 minutes including static postures (lie, sit, and stand) and dynamic activities (walk, turn left, turn right, stand-to-sit, sit-to-stand, sit-to-lie, and lie-to-sit), which are labeled by numbers from 0 to 9. 3-axial linear acceleration and 3-axial angular rate at a constant rate of 20 *Hz* are captured. The data are labeled manually during the experiments by using a wireless keypad with each key being assigned to a corresponding activity.

The raw signal is put through a filter to reduce noise and then features are extracted using a sliding window method with 2-second window width and 1-second step size. Both time domain and frequency domain features are extracted for the acceleration and angular rate data. The time domain features included mean, standard deviation, median, maximum and minimum values of magnitude, squared sum of magnitude angular rate below 25 percentile and 75 percentile. Features in frequency domain include spectrum peak frequency below 5 Hz, number of peaks in magnitude spectrum of angular rate below 5 Hz, and integral of magnitude spectrum from 0 to 5 Hz. The data consists of 6000 observations, which were split into 70% train and

30% validation sets. Several features having high skewness and kurtosis were Log and Square Root transformed.

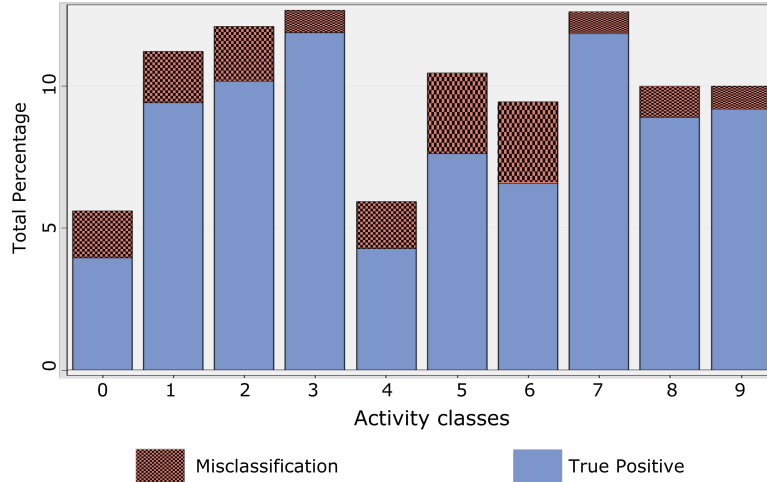
From the total set, seven important features are selected using the stepwise regression. These features include the mean and the variance of magnitude of acceleration, the number of peaks in acceleration magnitude spectrum below 5 Hz, the squared sum of magnitude of acceleration below 25 percentile, the number of peaks in the magnitude spectrum of angular rate below 5 Hz, the squared sum of magnitude angular rate below 75 percentile and 25 percentile. Several models are then applied to classify activities, including k-Nearest Neighbor, Decision Trees, and Neural Networks. Gradient Boosting Decision tree results in the highest classification accuracy. The configuration of the Gradient Boosting model has training proportion 60, the number of iteration 50, shrinkage parameter 0.1 and Square Loss Function with the base learner of maximum branch 2 and maximum depth 2.

#### **IV.1.4 Experiment Results**

As a result, the misclassification rate is 0.1 in training data and 0.13 in validation data. In other words, the accuracy of classification is 90% in training data and 87% in validation data. Fig. IV.2 shows the classification rate and confusion matrix of 10 activities using the validation data, in which the blue (lightly shaded) region represents correct classification and the remaining part represents misclassification. This result is then used to support human localization in the next subsection.

Another experiment is done with the smartwatch, and the accelerometer data from the smartwatch are used to classify activities such as Sitting, Standing, Lying, Walking, Climbing Up and Downstairs, Jogging and Running. The data were first filtered using a moving average filter and framed into a window size of 2s with 50% overlap. 13 time and frequency domain features are extracted from each frame and 89.12% body activity level classification accuracy is achieved using Gradient Boosting

Decision Tree. A sample application of this contextual information is shown using the Hydration Monitoring Application on section V.



(a) Activity classification chart for 10 activities.

		Predicted classes									
		0	1	2	3	4	5	6	7	8	9
Actual classes	0	0.78	0.02	0.15	0.00	0.00	0.00	0.03	0.00	0.00	0.01
	1	0.01	0.91	0.01	0.01	0.00	0.03	0.01	0.02	0.00	0.01
	2	0.12	0.04	0.82	0.00	0.00	0.01	0.01	0.00	0.00	0.01
	3	0.00	0.01	0.00	0.96	0.01	0.01	0.00	0.01	0.00	0.00
	4	0.01	0.00	0.00	0.00	0.86	0.04	0.05	0.02	0.01	0.00
	5	0.02	0.01	0.03	0.02	0.03	0.78	0.07	0.02	0.03	0.00
	6	0.01	0.02	0.00	0.00	0.08	0.06	0.78	0.04	0.01	0.01
	7	0.01	0.03	0.01	0.02	0.01	0.05	0.03	0.81	0.01	0.04
	8	0.00	0.00	0.00	0.00	0.01	0.05	0.01	0.02	0.91	0.00
	9	0.01	0.01	0.00	0.00	0.00	0.00	0.02	0.04	0.00	0.92

(b) Confusion matrix.

Figure IV.2: Activity recognition result.

The models are implemented on the gateway to classify daily activity and the labels are continuously sent to the cloud for online monitoring purpose. All raw data including timestamp, acceleration, and angular rate are uploaded to the cloud every 24 hours for storage and future analysis.

## IV.2 Indoor Human Localization

Indoor human monitoring plays an important role in living assistance and emergency detection. For example, to save energy and provide human comfort, human location can be used to trigger home automation tasks such as automatically controlling lights and air conditioners. Knowledge of human location can also help domestic service robots to better assist people.

### IV.2.1 Related Work

Numerous researches have been conducted in developing smart homes, social robots and assistive technology for elderly care [42]. This section reviews research in indoor localization, which is a key technology to develop such elderly care applications.

Researchers have developed various techniques to localize residents in indoor environments [43]. Basically, we can classify them into infrastructure based methods and non-infrastructure based methods. In infrastructure based methods, either active beacons or passive sensors like cameras/motion detection sensors are deployed in the environment to help localize the human. In non-infrastructure based methods, only wearable body sensors are used for localization.

Active infrastructure based methods can provide a good tracking accuracy. Active Badge [44], Active Bat [45] and Cricket [46] are three examples. Active Badge was developed at Olivetti Research Laboratory and it used the diffuse infrared technology. They used a network of sensors placed around the building with each room having at least one sensor. The human carried a tag that could emit a unique code for approximately a tenth of a second every 15 seconds (a beacon) in order for the sensors to detect it. A master station, also connected to the network, polled the sensors for badge 'sightings', processed the data, and then made it available to clients that could display it in a useful visual form. When a sensor received signals from the tag, the human was known to be in that room. The Active Bat system developed

by AT&T uses the ultrasonic location technique to locate the resident based on the principle of trilateration. The Bat (a transmitter) emitted a short pulse of ultrasound signal to the receivers mounted at known locations on the ceiling. Using the time of flight of the pulse, the distance of the Bat from each receiver was calculated and the 3D position of the Bat was determined [45]. Cricket [46] is another example that used the ultrasound technology to localize a human. Cricket indoor location system consisted of location beacons and listeners. Beacons were transmitters being attached to the ceiling of buildings and receivers called listeners were attached to the devices that required location information. Active beacons transmitted location information and an ultrasonic pulse. The passive listener captured this information to calculate its distance from the beacon [46]. In [47], WiFi RSS (Received Signal Strength) fingerprints were used to detect person location and estimate the velocity. Based on the 802.11n channel state information, a method was proposed to estimate the moving speed and distance of a subject. It would work for a big area such as a large office environment where a network of access points is deployed. The localization accuracy is within  $2m$ . However, this method is not suitable for an apartment where only one or two access points are available.

Passive infrastructure based methods use cameras or simple motion detection sensors. Cameras can be used to locate humans more accurately. For example, Easy Living used two sets of stereo color cameras for tracking multiple humans [48]. Even though this method provided a practical person-tracking solution, there is always privacy concern when cameras are used. In [49] multiple cameras and floor sensors were used to localize a human in a smart home environment. The system consisted of three components: camera localization, sensory floor localization, and condensation tracker. The presence of a person was detected by the sensory floor by measuring the pressure variation. The cameras detected the person by using background subtraction and human template matching. The condensation tracker was used to improve the



accuracy of localization. Recently, passive infrared (PIR) sensors have been used to detect and localize humans because of their simplicity and less privacy concern. In [50], a method was proposed to detect and localize the human inside homes using PIR sensor networks. Multiple PIR sensors were installed on the ceiling. In the initial offline training phase, a unique identifying set for each of the location regions was saved as a fingerprint. Using this fingerprint database as a lookup table, the identifying set from the online testing phase was mapped, from which the target location was recognized. Luo and Chen [51] proposed a technique to track the path of a human using PIR sensors. An array of PIR sensors was placed in a room so that the sensing areas of two or more sensors overlap. The experiment was conducted on a test bed in which a total of 12 sensors are placed to ensure the maximum overlap and the full coverage of the test area. A location accuracy of  $0.5m$  was achieved by the system. In [52, 51], PIR sensors were used to localize humans and track human movement. The accuracy of these PIR-based location systems greatly depends on the distribution density of sensors and the intersection of the field of views (FOVs) of the sensors. To reduce sensory uncertainty, multi-sensor techniques and data fusion algorithms are usually adopted. In [53], Yang et al. developed a low-cost and small-size human tracking system based on PIRs and wireless sensor network. A localization method based on detecting angle bisectors of PIR sensors and data fusion was presented. Then, Kalman filtering and particle filtering were employed for target tracking. In [54], Ibrahim Al-Naimi presented an automatic identification and tracking method by combining data from PIR sensors and floor pressure sensors. In [55], a method using PIR and RF for localization was proposed to deal with the RSS variation issue that caused a significant localization error. PIRs were used to identify the rough area, then the position of the person was estimated by applying the K Nearest Neighbor (K-NN) algorithm to the fingerprints inside that area. That method can reduce the localization error; However, the accuracy has not been validated through real

experiments.

Non-infrastructure based methods do not rely on any environmental devices or sensors. Instead, body-worn sensors are typically used. Location tracking based on wearable sensors has been investigated by several researchers. Personal dead reckoning (PDR) [56, 57] is the most common technique. Foxlin [58] developed an inertial sensor based NavShoe system which utilizes a complementary extended Kalman filter (EKF) to dynamically estimate an error state vector consisting of accelerometer and gyro biases, velocity, location, and attitude errors. His method results in an error of about 1% of the walking distance. Ojeda et al. [57] developed a Zero Velocity Update (ZUPT) based method that compensates the accelerometer bias and achieves an error of about 2% for short walks. Stirling et al. [59] built a PDR system that uses foot acceleration to estimate the stride length and the detection of the stance phase to resolve the step heading. Jirawimut et al. [60] presented a PDR method where the compass bias error and the step size error were estimated online using a Kalman filter. However, due to the nature of dead reckoning, PDR-based human tracking suffers unbounded growing errors. Mezentsev et al.'s study [56] showed that for medium accuracy PDR systems the two-dimensional location error grows proportionally to the square root of time.

There has been also existing work that combines both environmental infrastructure and wearable sensors for improved location tracking. Some researchers proposed a hybrid solution in which IMUs are added to aid localization. The gyroscope, accelerometer, and magnetometer in the IMU are used to yield relative positioning information. Meanwhile the absolute reference is obtained using other methods such as light sensing [61] and WiFi signal strength [62, 63]. For example, in [62], researchers used WiFi signal strength to determine the absolute reference and use a dual-axis accelerometer, a gyroscope, and a pressure sensor to provide additional information. A method based on a Kalman filter and a particle filter was proposed which fuses the

heterogeneous information from two independent data sources. However, in indoor environments, the heading angle inferred by the magnetometer in an IMU sensor is not reliable due to the magnetic disturbances caused by electronic devices, metal structures of the building, etc. Also because of the slowly increasing drift of a gyroscope, a small drift may result in big errors after a sequence of integration processes [64]. A summary of the indoor localization methods is described in table IV.1.

Table IV.1: Indoor localization methods

Technology	Pros	Cons
<b>Infrastructure based methods</b>		
IR [44]	Low cost	Low accuracy (at room level)
Ultrasonic [45]	High accuracy, low energy consumption	Affected by noise, expensive
RF and ultrasound [46]	Low cost	Signals may be reflected or blocked by walls
WiFi [47]	Low cost	Not good for small area
Camera [48, 49]	Multi-person tracking	Privacy violated
PIR [50, 51, 52, 53]	Low cost	Low accuracy
PIR and floor sensor [54]	High accuracy	Expensive
PIR and RF [55]	RSS variation is reduced	The accuracy is not much improved
<b>Non-infrastructure based methods</b>		
IMU [56, 57, 58]	Small design, one single unit	Not stable due to drift issue, heading errors, magnetic errors
<b>Hybrid methods</b>		
IMU and Wifi [62, 63]	Good for large environment	Not good for small apartment

In this research, we adopt a sensor fusion based approach to integrate both environmental sensors and wearable motion sensors. We use PIR sensors as the environmental sensors and IMU as the wearable motion sensor. More importantly, in order to significantly improve the localization accuracy, we utilize the human activity/location correlation, based on a Bayesian filtering framework. In other words, by recognizing human activity and integrating with a pre-known map of furniture in the testing environment, the accuracy of localization and tracking is greatly improved.

## IV.2.2 Methodology

This section presents the detailed methodology for the proposed indoor human localization system.

## A. Overview

The proposed approach adopts a sensor fusion strategy that combines motion data from the human body and PIR sensor data from the environment for human localization. In addition, this approach takes advantage of the correlation between the human locations and their activities in home environments for improved accuracy.

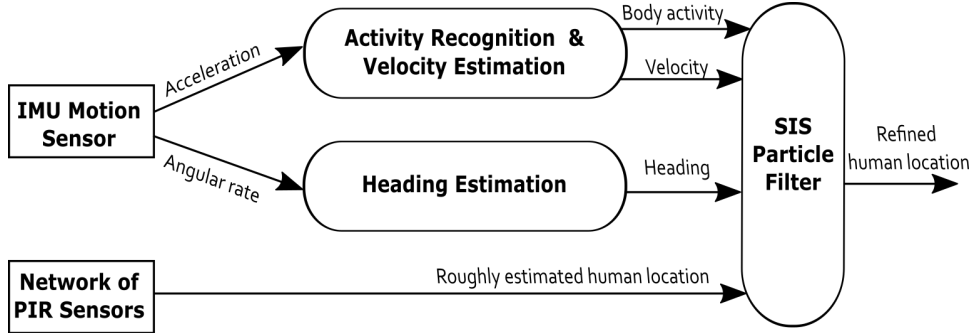


Figure IV.3: Overview of the approach.

The overview of the approach is depicted in Fig. IV.3. The data from the IMU sensor, which includes 3D acceleration and angular rate, are inputs to the following three modules: Activity Recognition, Velocity Estimation, and Heading Estimation. From these modules the human body activity, velocity, and heading are derived, respectively. Meanwhile, the data from the array of PIR sensors are also collected. These two channels of information are fused through a Particle Filter module to estimate the location of the human. The behavior-based map, as prior knowledge, represents the correlation between human’s location and his activity. In this map, the position of walls and furniture (such as table, chair, and bed), as well as other facilities, are pre-known. This map basically encodes the location probability of the resident when he is conducting certain activities such as walking, sitting and lying. This is important information that can be used to improve the localization accuracy in a Bayesian filtering framework.

In the next sections, we will first introduce the observation model of the PIR sensors and the IMU based motion model. Then we will discuss the behavior-based

map. Finally, we describe the sensor fusion algorithm that combines the PIR and IMU data for human localization.

## B. Observation Model of PIR Sensor

In our Smart Home setup, eight PIR sensors are deployed on the ceiling in a configuration as shown in Fig. IV.5. This configuration is determined to ensure that the sensors can cover the majority of the apartment area. The detecting area of the PIR sensors is circular on the floor and their radius can be changed by adjusting the cylindrical lens covers. In this study, the radius is set to be  $1.1m$ .

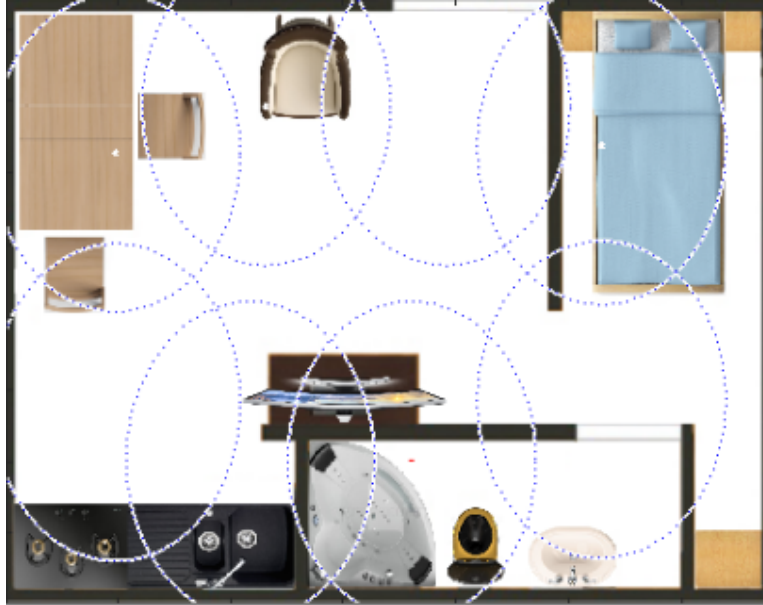


Figure IV.4: Top view of the mock apartment. Dotted circles denote the field of view of the PIRs.

The array output of 8 PIR sensors is one of the following cases: (1) Only one PIR is active. The human should be in one of the color areas as shown in Fig. IV.5(b). (2) Two PIRs are active. The human should locate in the overlap area of those 2 sensors as shown in Fig. IV.5(c). (3) No PIR is active. The human is outside of all PIR sensing areas as shown in Fig. IV.5(d).

Based on the PIR sensor observation model mentioned in [65], our PIR sensor

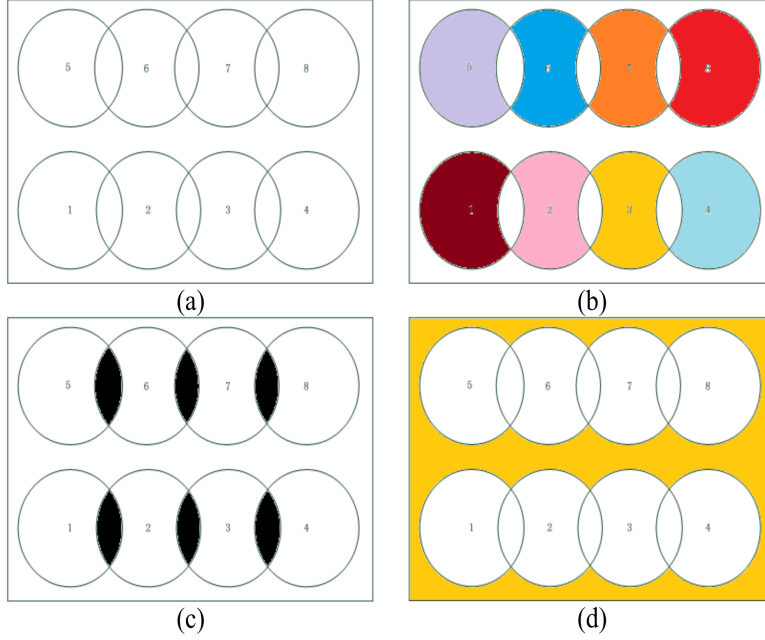


Figure IV.5: PIR sensing area. (a) PIR configuration; (b) One PIR is active; (c) Two PIRs are active; (d) No PIR is active.

model is expressed by Equation (IV.1)

$$P(z_k^{PIR,i} | s_k) = \begin{cases} p^{z_k^{PIR,i}} (1-p)^{1-z_k^{PIR,i}} & \text{if } |s_k - C_i| \leq r_i \\ q^{z_k^{PIR,i}} (1-q)^{1-z_k^{PIR,i}} & \text{if } r_i \leq |s_k - C_i| \leq r_i + \epsilon \\ 1 - z_k^{PIR,i} & \text{if } |s_k - C_i| \geq r_i + \epsilon \end{cases} \quad (\text{IV.1})$$

where  $p$  is the probability of detection,  $q$  is the probability of false alarm.  $z_k^{PIR,i}$  is the binary output from PIR sensor  $i$  at time  $k$ , which takes value from  $\{0,1\}$ .  $s_k$  is the human state which is the 2D location. According to [65], the parameter  $p$  and  $q$  are estimated from measurements ( $p = 0.7$ ,  $q = 0.05$ ), but it is not realistic when we consider the distance between the human and the sensor. We discovered that false alarm may occur when the human is not in the sensing range, but not too far away from the sensor, which is depicted by the gray area as shown in Fig. IV.6. If the human is far away (out of the dashed circle), the false alarm rate  $q$  becomes 0. It is possible that false negative (miss) may occur because of the sensor's failure. In our

setup, the probability of miss can be reduced because we can use the wearable IMU to detect if there is a human motion. If there is human motion from the IMU and the PIR sensor is off, that implies a false negative and which can be corrected. If there is no human motion from the IMU, there is no need to update the human location. Therefore, the final values of those probabilities are  $p = 0.9$  and  $q = 0.05$ .

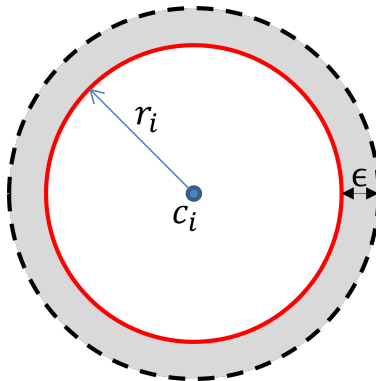


Figure IV.6: Sensing regions of a PIR sensor.

### C. IMU-Based Motion Model

The IMU sensor attached to the human body can provide important information regarding the activity and motion of the wearer. Here we derive the motion model of the wearer using the data obtained from the IMU. We first recognize human's activity, based on that the moving velocity is estimated. Then we compute the heading based on the gyro output.

**Velocity estimation:** Once the walking activity is detected, the velocity  $v_k$  is estimated by integrating the acceleration during the swing phase of each step, which includes a stance phase and a swing phase [58]. When the stationary stance is detected, or the activity is recognized as standing, sitting, or lying the velocity is reset to zero to eliminate accumulated errors.  $v_k$  is estimated by using Equation (IV.2) as

follows

$$v_k = \begin{cases} a_{xy}T_s + v_{k-1} & \text{if swing phase} \\ 0 & \text{if stance phase;} \\ & \text{or standing, sitting, lying} \end{cases} \quad (\text{IV.2})$$

where  $a_{xy} = a_x + a_y$  is the horizontal acceleration,  $T_s$  is the sampling period. To detect the stationary stance, we apply thresholds to both the magnitude of the acceleration  $|a_m| = \sqrt{(a_x^2 + a_y^2 + a_z^2)}$  and the magnitude of angular rate  $|w_m| = \sqrt{(w_x^2 + w_y^2 + w_z^2)}$ . The stance phase is recognized if  $9m/s^2 < |a_m| < 11m/s^2$  and  $|w_m| < 30^\circ/s$ .

**Heading estimation:** Heading provides walking direction and this information can be directly read from the outputs (yaw, pitch, and roll) from the IMU. However, in indoor living environments, due to the magnetic disturbance caused by many devices such as computers, microwave, and other electrical appliances, the accuracy of heading read from the IMU is neither accurate nor reliable. This problem has been observed in many experiments we conducted inside our testbed.

Therefore, we developed a different approach to estimating the walking direction. In our approach, the angular rate from the IMU is used to estimate the heading changes when the human is walking. Similar to velocity estimation, the heading angle change can be estimated by integrating the angular rate over time. It is clear that the drifts of the gyroscope may lead to poor results since the errors are also integrated. In order to overcome this problem, we utilize the result from activity recognition, in which turning right and turning left are detected. The time  $t_0$  when the human starts turning, and the time  $t_1$  when the human finishes that turning are recorded. Then the estimated heading  $\theta_k$  is calculated by adding the amount of angle change to the previous heading angle  $\theta_{k-1}$  and the measurement noise  $N(0, \sigma_\theta)$  which has zero mean and a standard deviation value of  $\sigma_\theta$ ). The standard deviation is set



to  $\pi/6$  based on the measurements obtained in our experimental testing.

$$P(\theta_k) = N(\theta_{k-1} + w_z T_s, \sigma_\theta) \quad (\text{IV.3})$$

Now we can derive the motion model of the human,

$$s_k = f(s_{k-1}) + n_k \quad (\text{IV.4})$$

where the state vector  $s_k = [x_k, y_k]^T$ ,  $n_k$  is the process noise. The propagated position can be expressed by the following equation

$$\begin{bmatrix} x_k \\ y_k \end{bmatrix} = \begin{bmatrix} x_{k-1} + T_s v_k \cos(\theta_k) + n_k \\ y_{k-1} + T_s v_k \sin(\theta_k) + n_k \end{bmatrix} \quad (\text{IV.5})$$

where  $v_k$  and  $\theta_k$  are the velocity and heading at time  $k$  sampled from normal distribution  $N(v_k, \sigma_v)$  and  $N(\theta_k, \sigma_\theta)$  with velocity mean  $v_k$  and heading mean  $\theta_k$  which are estimated from the acceleration and angular rate information as mentioned earlier. The standard deviation  $\sigma_v$  and  $\sigma_\theta$  are set to 20% of the mean velocity and  $\pi/6$  rad respectively.

#### D. Behavior-Based Map

A human's location and behavior in a home environment are highly correlated. For example, when the human is detected as sitting, it means that he may be sitting on a chair in the dining area, on an armchair in the living room, on the bed, or on the toilet. Based on the location estimate in the most recent time steps, we can find which furniture is actually involved. Therefore the position of that furniture, which is pre-known, can be used to determine the human position and the error accumulation can be prevented, allowing the location accuracy to be significantly improved. In this sense, the activities serve as virtual landmarks in human localization.

To facilitate such behavioral landmark-assisted localization, we introduce the concept of the behavior-based map and conduct the location inference in a Bayesian

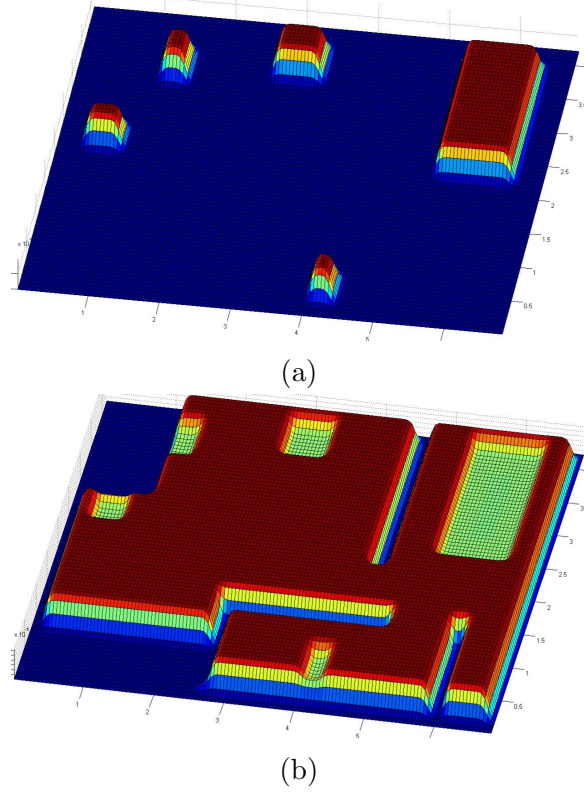


Figure IV.7: (a) Sit and lie map. (b) Walk and stand map.

framework through particle filtering. Basically, the behavior-based map can be represented by an accessibility probability function (APF) which can be defined as

$$P(s_k|a_k) = \varphi$$

here  $\varphi$  is the probability of being at location  $s_k$  when the human is conducting activity  $a_k$  which can be either sit, lie or walk. We assume the furniture location in the testbed is fixed as shown in Fig. IV.4, and the human can walk anywhere except places occupied by furniture and walls. Fig. IV.7(a) shows the sit and lie map which is applied when the human activity is recognized as sitting or lying. In this map, the red color means high probability for sitting or lying, and the blue color means extremely low probability for sitting or lying. For the walk and stand map as shown in Fig. IV.7(b), if the place is occupied by the furniture such as the table or the wall,  $P(s_k|a_k)$  is set to almost zero, which implies that a human will not walk in that area. The behavior map will be utilized to improve the location estimate, as explained in

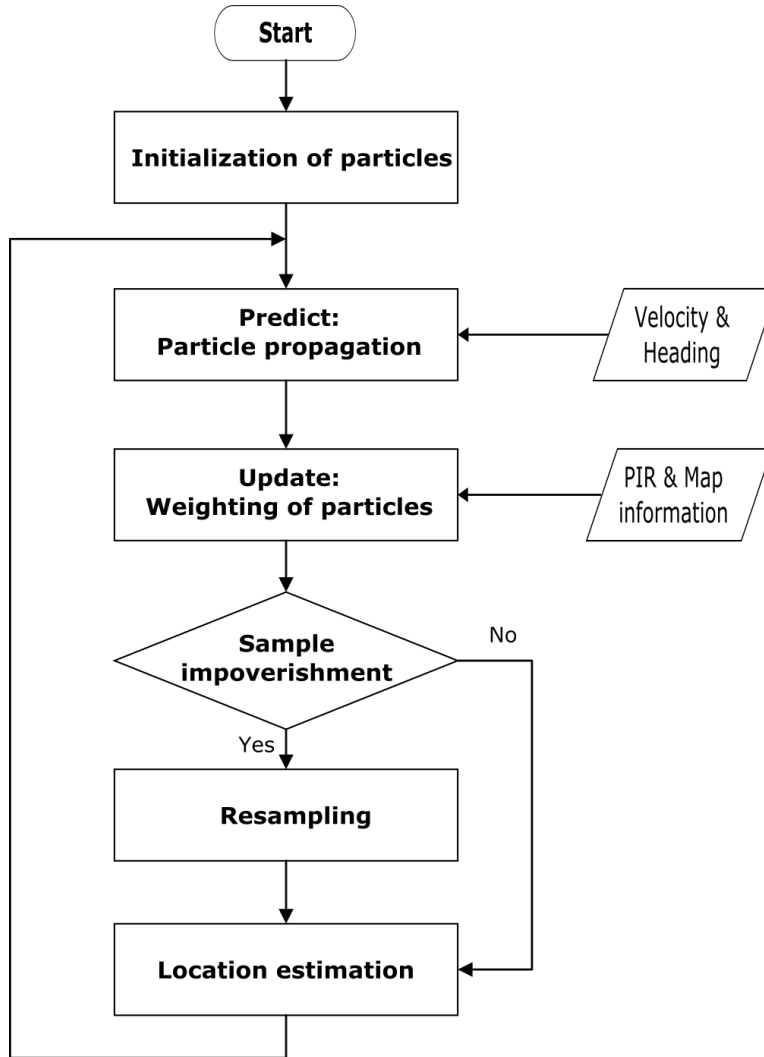


Figure IV.8: Particle filter diagram.

the next section.

### E. Human Localization through Sensor Fusion

By fusing the two channels of information: PIR data and IMU data, we can derive a more accurate location estimate than using only one of them. The fusion is done in a Bayesian filtering framework. Although Kalman filtering is a popular method used in many navigation systems, the requirement of the linear model and Gaussian noise is not satisfied in our case. Therefore, we choose particle filtering for human localization with multiple data sources.

The particle filter, or Sequential Monte Carlo, is one of the numerical methods to estimate posterior density function  $P(s_k|z_{1:k})$  of system state  $s_k$  given the observation data up to time  $k$ ,  $z_{1:k}$ . It is an iterative process consisting of two main steps, prediction and update, which employs two models, system model (or motion model) and measurement model. Each particle represents a possible state, and all particles can approximate the posterior probability distribution of the state. The prediction step is based on the state motion model  $P(s_k|s_{k-1})$ , in which the propagation of particles utilizes velocity and heading angle estimates provided by the IMU. The propagation of particles is expressed by Equation (IV.5). The update step is based on the observation model  $P(z_k|s_k)$ , in which the PIR data and activity information derived from the IMU data are used to compute the weight of each particle. Locations of particles with more weight mean they are closer to the true state. At the end of the loop is the resampling step which removes low-weight particles and generates new particles with normalized weight. The whole process of particle filtering is described in Fig. IV.8.

Here we explain how to update the weight of particles. We need to calculate the likelihood of the observations  $z_k^i = [z_k^{PIR,i}, z_k^{IMU,i}]$  given a human location  $s_k$

$$\begin{aligned} P(z_k^i|s_k) &= P(z_k^{PIR,i}, z_k^{IMU,i}|s_k) \\ &= P(z_k^{PIR,i}|s_k) \cdot P(z_k^{IMU,i}|s_k) \end{aligned}$$

Here

$$\begin{aligned} P(z_k^{IMU,i}|s_k) &= \sum_{a_k} P(z_k^{IMU,i}, a_k|s_k) \\ &\propto \sum_{a_k} P(z_k^{IMU,i}, s_k|a_k) \cdot P(a_k) \\ &= \sum_{a_k} P(z_k^{IMU,i}|a_k) \cdot P(s_k|a_k) \cdot P(a_k) \\ &= \sum_{a_k} P(z_k^{IMU,i}|a_k) \cdot P(a_k|s_k) \end{aligned} \tag{IV.6}$$

Therefore we have

$$P(z_k^i | s_k) = P(z_k^{PIR,i} | s_k) \sum_{a_k} P(z_k^{IMU,i} | a_k) \cdot P(a_k | s_k) \quad (\text{IV.7})$$

The above equation calculates the likelihood of observing the PIR data  $z_k^{PIR,i}$  and IMU data  $z_k^{IMU,i}$  given that the location  $s_k$  of the human is known. With the correlation between activity and location  $P(a_k | s_k)$ , the behavior-based map is integrated into the update step. The weight of each particle is updated to generate the posterior distribution  $P(s_k | z_{1:k})$ . Let  $w_k^{PIR,i} \approx P(z_k^{PIR,i} | s_k)$  be the weight of  $i$ th particle at time  $k$  based on the PIR sensors, and  $w_k^{IMU,i} \approx P(z_k^{IMU,i} | s_k)$  be the weight based on the IMU sensor.  $P(z_k^{IMU,i} | s_k)$  can be calculated according to Equation (IV.6). The updated weight should be the product of these two weights and the previous weight

$$w_k^i = w_k^{PIR,i} \cdot w_k^{IMU,i} \cdot w_{k-1}^i \quad (\text{IV.8})$$

If the number of particles with low weight reaches a certain threshold, the resampling should be conducted. Otherwise, the particle filter will degenerate when there are only a few high weighted particles, which leads to a poor approximation of the state estimate. The effective number of particles  $N_{eff}$  is used as the indicator of degeneracy [66], which is calculated as follow

$$N_{eff} = \frac{1}{\sum_1^N (w_k^i)^2} \quad (\text{IV.9})$$

where  $N$  is the total number of particles. If  $N_{eff}$  is less than a threshold  $N_t$ , the resampling will kill particles with low weight and high-weight particles are replicated. After resampling, the particles become more concentrated in higher probability area of the posterior, and the state is estimated based on the mean of the posterior distribution.

$$\hat{s}_k = E [s_k^i] = \frac{\sum_{i=1}^{N_k} s_k^i \times w_k^i}{\sum_{i=1}^{N_k} w_k^i} \quad (\text{IV.10})$$

and the covariance matrix is

$$P_k = \sum_{i=1}^N w_k^i (s_k^i - \hat{s}_k) \quad (\text{IV.11})$$

---

**Algorithm 1** The Particle filtering for human localization Initial

---

- 1: Particles' parameters  $\{s_1^i, w_1^i, \theta_1^i, N\}, i = 1, \dots, N$
- 2:  $T$ : total number of observations
- 3: **for**  $k = 2 : T$  **do**
- 4:     Estimate  $v_{k, k}$  and recognize human activity  $a_k$  based on the observation data  $z_k^{IMU}$
- 5:     *Prediction Step:*
- 6:     Propagate particles according to Eq. (IV.5).
- 7:     *Update Step:*
- 8:     Assign the particle a weight  $w_k^i$  according to the observation  $z_k^{PIR}$ , and  $z_k^{IMU}$ .

$$\begin{aligned} w_k^i &= w_k^{PIR,i} \times w_k^{IMU,i} \times w_{k-1}^i \\ &= \sum_{a_k} [P(a_k|s_k)P(z_k^{IMU}|a_k)] \times P(z_{i,k}^{PIR}|s_k)w_{k-1}^i \end{aligned}$$

- 9: **end for**
- 10: Normalization  $w_k^i = \frac{w_k^i}{\sum_{i=1}^N w_k^i}$
- 11: Estimate: calculate  $\hat{s}_k$  according to

$$\hat{s}_k = E [s_k^i] = \frac{\sum_{i=1}^{N_k} s_k^i \times w_k^i}{\sum_{i=1}^{N_k} w_k^i}$$

where:  $N_k$ : Number of remaining particles at step  $k$

- 12: **procedure** RE-SAMPLING
  - 13:     Calculate  $N_{\text{eff}} = \frac{1}{\sum_{i=1}^N (w_k^i)^2}$
  - 14:     Set  $N_t = \text{resample percentage} \times N$
  - 15:     **if**  $N_{\text{eff}} < N_t$  **then**
  - 16:          $s_k^t, w_{k,i=1:N}^{t'} = \text{resample } s_k^t, w_{k,i=1:N}^t$
  - 17:     **end if**
- 

Algorithm 1 shows how the particle filtering algorithm is employed for human localization.

## IV.2.3 Experiment Results

### A. Environment Setup

For the human localization, experiments are conducted with six human subjects, who are asked to follow the same trajectories marked on the floor in a mock apartment

built in our lab. Each human subject is asked to move around the mock apartment. They wear a cap with reflective markers on it, enabling the head location to be tracked by the OptiTrack system as the ground truth. Regular daily activities are performed: standing, walking, sitting, and sleeping. One IMU motion sensor is used, which is attached on the right thigh of the human subject. Eight PIR Sensors are placed on the ceiling at a standard ceiling height of  $2.45m$ . The PIR data is read every  $0.05s$  from each sensor (sampling rate of  $20Hz$ ). The radius of each PIR sensor is restricted to be  $1.10m$  by a cylindrical lens cover. The server PC is responsible for collecting the data from the PIR sensors and the IMU sensor. Based on these data, the server PC estimates the human’s location in the apartment and derives the movement trajectory of the human.

## B. Evaluation

An experiment is conducted by following a scenario in which the human starts at the entrance door area, and we assume that their initial position and heading are known. The human subject enters the apartment triggering the PIR at the front door and the system starts performing localization and tracking. The human is asked to go to the dining table area and sit on the chair, go to the kitchen area, go to the living room and sit on the armchair, and stop on the way to the bedroom. In Fig. IV.9, the result of localization and tracking of six subjects is shown. The green line is the ground truth obtained from the OptiTrack system, and the red dashed line is the estimated trajectory. The length of the trajectory is approximately  $35m$ . The

Table IV.2: Summary statistics of distance error (D).

	<b>S1</b>	<b>S2</b>	<b>S3</b>	<b>S4</b>	<b>S5</b>	<b>S6</b>
<b>Max of D (m)</b>	0.592	0.631	0.601	0.552	0.612	0.582
<b>Mean of D (m)</b>	0.141	0.155	0.130	0.121	0.139	0.135
<b>Std of D (m)</b>	0.105	0.110	0.118	0.101	0.113	0.102

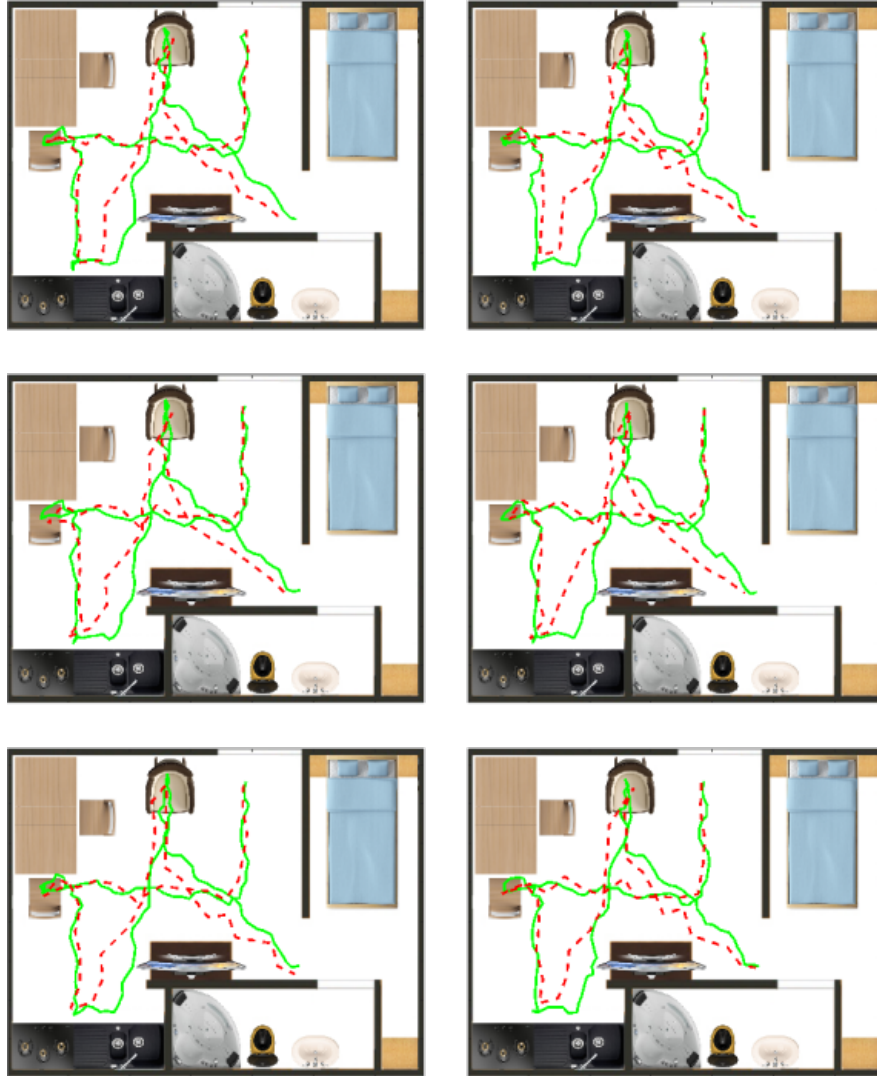


Figure IV.9: Result of localization and tracking. Solid lines denote ground truth and dashed lines denote estimated trajectories.

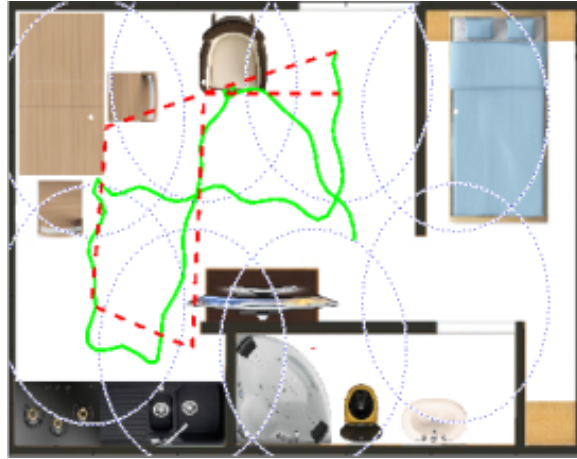
distance error between the ground truth and the estimated lines is calculated for each human subject as shown in Table. IV.2. On average, the mean of distance error is  $0.137m$ .

In Fig. IV.10(a), with the same trajectory, the level of accuracy is reduced if the human subject just walks and no sitting or lying is performed. And Fig. IV.10(b) shows the result of localization and tracking when only PIR sensors are used. In this case, the accuracy depends only on the sensor distribution, the number of sensors, and sensor field of view. And those parameters are already preset in our testbed, thus





(a)



(b)

Figure IV.10: (a) Localization and tracking without activity recognition. (b) Localization and tracking when using only PIRs.

the accuracy is extremely reduced. The mean of distance error is around  $2m$ .

Another experiment was conducted, in which the human subject was asked to wake up, go to the restroom, have breakfast and go out. A typical result of a human subject is picked up and shown in Fig. IV.11. The length of the trajectory is approximately  $45m$  and the mean of distance error is  $0.125m$ .

Obviously, by employing activity recognition and taking advantage of the correlation between activity and furniture locations, the accuracy of localization and tracking is significantly improved.

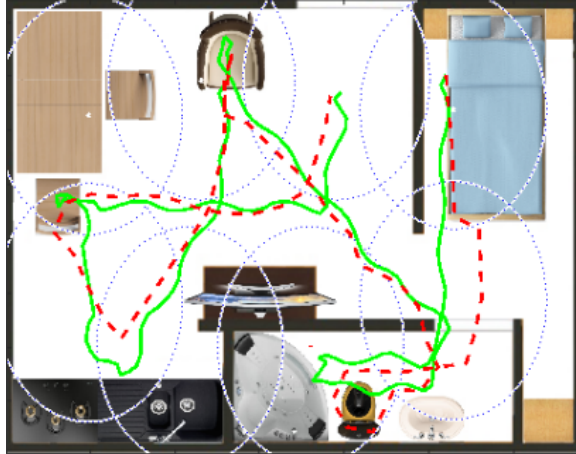


Figure IV.11: Another result of localization and tracking.

### IV.3 Summary

In this chapter, we demonstrated that the context information can be generated from our CoSHE system based on the home and wearable sensors. We studied the problem of human localization in an indoor environment. We proposed a method that uses a particle filter to fuse data from both environmental sensors (PIR) and wearable sensor (IMU) for human localization. Then acceleration and angular rate data from the IMU sensor worn on the human body are used for body activity recognition, velocity estimation, and heading estimation. The PIR sensors provide external rough tracking of the human location. The human body activity is used as virtual landmarks that help reduce location errors. Since we only use one wearable motion sensor, our method has the advantage of reducing the obtrusiveness to the wearer, while maintaining high accuracy of indoor localization. We conducted experiments in a mock apartment and the accuracy of the proposed method is evaluated. The context information does not only provide insights of human daily activity, movement pattern, human tracking but also can be supportive information for healthcare inferences and services.

## CHAPTER V

### WELL-BEING MONITORING

In this chapter, two case studies are conducted to evaluate the usability of our system in real time monitoring for both physical and mental well-beings. Firstly, we present the capability of the system in body hydration monitoring. Secondly, the ability to detect negative emotions is demonstrated.

#### V.1 Body Hydration Monitoring

##### V.1.1 Introduction

Water is an essential component of the human body with numerous functions including nutrient transport, body temperature regulation, and solvent for different substances. The average healthy adult has 45-75% body water composition and the body adapts to changes in water intake and losses to keep its water balance. Consuming too much water causes Hyponatremia (water intoxication) and a very minimal consumption results in Dehydration which in critical cases can both result in severe illness including swelling in the brain and loss of short-term and long-term memory [67],[68]. Healthy hydration is, therefore, essential for the normal functioning of the body. This can be achieved by using a cumulative analysis of drinking habits, activity context, and environmental conditions.

In recent years a number of hydration recommendation systems and gadgets have been developed in the form of smartwatches, mobile apps and smart garments. However, most of them are not able to measure the actual daily fluid intake and also rely on user input to understand daily activity. Moreover, some systems use accelerometer

data to recognize physical activities. Their activity classification, however, relies on set cut-points which cannot efficiently generalize to different demographic groups and a variety of physical activities [69].

In this case study, we present a hydration monitoring application called Auto-Hydrate which overcomes the above shortcomings through automatic detection of drinking and physical activities. The system is comprised of an embedded computer, a throat microphone, and a smartwatch. The embedded computer is used to collect and process the signals in real time. The throat microphone records sound from the throat area and the smartwatch is used to collect body activity data. The hydration monitoring algorithm runs on the embedded computer and real-time feedback is transmitted to our private cloud database, which can be accessed via Internet-connected devices. The system has a very minimal form factor and can be worn conveniently for longer periods.

### **V.1.2 Related Work**

#### **A. Acoustic Sensors for Tracheal Activity Recognition**

Acoustic sensor based systems use a microphone worn around the throat to collect acoustic signals. The signals are transmitted to a controller hardware through wired or wireless communication [70]. A range of time, frequency and cepstral domain features are then extracted for classification. Yin et al. [71] had achieved an average accuracy of 84.9% for food type recognition, 97.6% for liquid food intake and 99.7% for solid food intake using Hidden Markov Models HMM and Decision Tree. Koji et al. [70] classified 12 activities including eating, drinking, speaking and laughing using SVM and obtained F-measure of 79.5%. The authors also validated the real world application of their method with a small user study. Temiloluwa et al. [72] achieved an F1 score of 86.6% using 5-Nearest Neighbor 5-NN and 86.7% using Naive Bayes. Yin et al. [73] also classified eating activities with an accuracy of 86.82% and 98.35%

using KNN and SVM respectively.

## **B. Estimated Energy Requirement**

The human body fluid requirement varies depending on the body activity level, environmental and medical conditions. It is difficult to set a single level of intake requirement that would assure adequate hydration in all conditions. The Institute of Medicine, Dietary Reference Intakes committee has therefore established an Adequate Intake AI for different sex and age groups [74]. The study is conducted on more than 1200 individuals using doubly labeled water technique and a series of regression equations are established for Estimated Energy Requirement EER in Kcal per Day [67]. These equations take into account the sex, age, weight, height and physical activity of the subject.

### **V.1.3 AutoHydrate Application**

#### **A. Hardware Description**

AutoHydrate consists of an acoustic sensor for recording sounds from the throat area, a smartwatch for sensing body activity and an embedded computer for collecting and processing the data. The embedded computer continuously collects the signals and classifies drinking and body activities in real time. The system architecture and hardware components are illustrated in Fig. II.5, which includes a high sensitivity throat microphone to record audio signals from the throat area, a smartwatch is to collect activity data, and an Intel Edison compute module to collect and process signals, detect drinking sound, and classify body activities.

#### **B. Drinking and Body Activity Recognition**

Two types of activity recognition are carried out in this study, drinking activity recognition and body activity recognition. Our system uses the algorithm shown

in Fig. V.1 to provide real-time hydration information to the user. The following subsections will detail each stage of the algorithm.

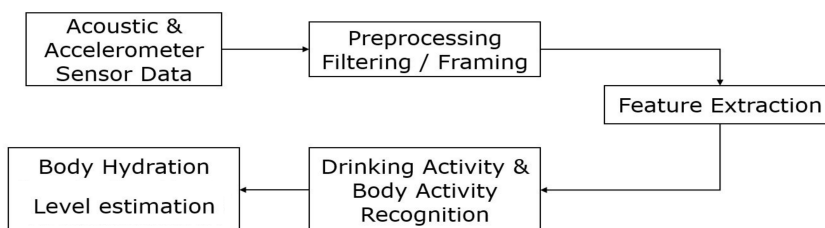


Figure V.1: Hydration monitoring algorithm.

- Preprocessing

The preprocessing step has two stages. In the first stage, the audio signal is sampled at 16000Hz and filtered using a low pass filter. Based on our analysis and work done in [70] most of the energy of a drinking sound lies below 5000Hz and therefore we have used a low pass FIR filter with 5500Hz cutoff frequency. The audio signal is framed into a window of size 500ms with 50% overlap. We have chosen this window size to include a maximum drinking segment.

A weighted moving average filter is used to remove noise from the accelerometer data. This is then framed into 66 (2 seconds) samples with 50% overlap.

- Feature Extraction

A careful selection of features is critical for accurate classification. Based on our analysis and review of work done in [71], [70] and [72] we extract time domain, frequency domain and cepstral domain features from the acoustic signal. These features are selected optimally based on our evaluation of the system’s recognition performance while using the minimum number of features to minimize computation for real-time operation.

- Time Domain Features: Zero-Crossing Rate

- Frequency Domain Features: Total Spectrum Power, Sub-band Powers, Spectral Centroid, Spectral Spread, Spectral Flux, Spectral Rolloff
- Cepstral Domain Features: Mel Frequency Cepstral Coefficients (MFCC)

For classifying body activities from the accelerometer data, we extract time domain and frequency domain features the same way as we do in session 4.1.2

- Drinking Activity Recognition

Support Vector Machines have been widely used in audio classification problems [70], [75], [76]. SVM classifies data by constructing a hyperplane  $wx + b = 0$  for a given set of training vectors belonging to separate classes  $(x_1, y_1), \dots, (x_i, y_i)$  where  $x_i \in R^n$  and  $y_i \in \{-1, +1\}$ . It also uses linear models to implement non-linear class boundaries by transforming the input space using a nonlinear mapping into a new feature space. A linear model is then constructed in the feature space which represents a nonlinear decision boundary in the original space. We have used Gaussian Radial Basis function as our mapping function. The SVM algorithm provides a binary classification which is +1 for drinking activity detected or -1 for drinking not detected.

- Body Activity Recognition

The use of Decision tree for body activity recognition is discussed in [77]. Decision Tree is a hierarchical model that recursively separates the input space into class regions. Its aim is to find the smallest tree possible. In order to achieve that it looks for the best attribute that would make the data as pure as possible after the split. Gradient Boosting is a form of ensemble model which is a sequential iteration of several decision trees. It starts with developing a decision tree as a base learner and the second base learner is built to fit the pseudo residuals of the previous base learner. It readjusts weights for misclassified observations at each step which are then given higher weight than correctly

classified ones. The readjustment is carried out until its loss function reaches a threshold. We have used the Gradient Boosting Decision Tree algorithm to classify activities into four categories as summarized in Table V.1.

Table V.1: Body Activity Classification

<b>Activity</b>	<b>Description</b>
Sedentary	Sitting on chair, Lying, Standing
Light	Walking
Moderate	Going Up/Down stairs
Vigorous	Jogging, Running

- Hydration characteristics Based on the result of drinking sound recognition, we can track water intake with an assumption of the average size of one gulp of water. With activity level estimation, the EER can be calculated based on the daily EER calculation published by Institute of Medicine, Dietary Reference Intakes committee as shown in the Table V.2 [67]. Body hydration requirement and energy intake are related by the relationship: 1mL of fluid per kcal [68]. Some characteristics of hydration monitoring are as follow:

$$Fl_{AI} = c_1y + PA(c_2w + c_3h) + c_4 \quad (V.1)$$

$$Fl_{in} = 20.09 \sum_{i=1}^N h_i(n) \quad (V.2)$$

$$PA = P * argmax\left(\frac{1}{M} \sum_{j=1}^M g_j(m)\right) \quad (V.3)$$

Where  $Fl_{AI}$  is the adequate fluid intake based on  $y$  (age),  $w$  (weight),  $h$  (height) and  $PA$  (physical activity coefficient), The coefficients  $c_1 - c_4$  are included in Table. V.2 for different demographics.  $Fl_{food}$  is the amount of fluid consumed from solid food and  $Fl_{met}$  is the metabolic fluid.  $Fl_{in}$  is the amount of fluid taken for  $N$  number of 500ms audio recordings in an hour period and  $h_i(n)$  represents detected drinking sounds.  $PA$  represents the physical activity coefficient



Table V.2: Estimated Energy Requirement

Sex	Age	Estimated Energy Requirement in Kcal/day
M	9-18	$88.5 - 61.9y + PA(26.7w + 903h) + 25$
F	9-18	$135.3 - 30.8y + PA(10.0w + 934h) + 25$
M	19+	$662 - 9.53y + PA(15.91w + 539.6h)$
F	19+	$354 - 6.91y + PA(9.36w + 726h)$

where  $g_j(m)$  represents our Decision Tree classifier for M number of 2s activities in an hour. The activity with maximum detections in the 1 hour period is scaled by the activity coefficient P which varies for different age groups. For example males 19+ years old have P = 1.00 for Sedentary, 1.11 for Light, 1.25 for Moderate and 1.48 for Vigorous activity based on calculations by [67].

We can utilize this information to give a further recommendation to the human as mentioned in Section 6.3.

### C. Experimental Result

Three different sets of experiments are conducted to evaluate our system. These include drinking activity recognition, body activity recognition, and hydration monitoring. The participants are instructed to wear our system as shown in Fig. V.2.

We have 10 healthy and normal weight participants of average age 28.2 in our study, and the data from 8 subjects is used in our analysis. All participants are allowed to drink 500ml of water under a normal pace. The participants take an average of 24.89 gulps to drink 500ml water and the average amount of water per gulp is 20.09ml. For training the drinking detection algorithm, 213 drinking sounds are used and the classification is evaluated using the following metrics.



Figure V.2: The setup of the AutoHydrate system.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (V.4)$$

$$Precision = \frac{TP}{TP + FP} \quad (V.5)$$

$$Recall = \frac{TP}{TP + FN} \quad (V.6)$$

$$F - measure = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (V.7)$$

Table V.3: Drinking Detection Results using 8-Fold Cross-Validation

		Prediction		Recall
		Drinking	Not Drinking	
<b>Actual</b>	Drinking	263	49	<b>84.3</b>
<b>Activity</b>	Not Drinking	41	706	<b>94.5</b>
<b>Precision</b>		<b>86.5</b>	<b>93.5</b>	

Where TP is True Positive, TN is True Negative, FP is False Positive, FN is False Negative. We have used 8-fold cross-validation or leave-one-participant-out method for evaluating our classifier. The detection results are included in TABLE V.3.

The data from 6 subjects were used for evaluating the body activity recognition algorithm. The subjects were instructed to perform a set of activities which belong to one of the four activity categories. The activities include Sitting, Standing, Lying,

Walking, Going Downstairs and Upstairs, Jogging and Running. Our detection results using 6-fold cross-validation are shown in TABLE V.4.

The combined system has a classification accuracy of 91.5% and 89.12% for Drinking Activity and Body Activity respectively as can be seen from Table. V.5. The use of K-fold cross-validation shows that our system can detect a new user input at the high accuracy level. This will make it easily adaptable for new users.

Table V.4: Result of Body Activity Level Recognition using 6-Fold Cross-Validation

		<b>Prediction</b>				<b>Recall</b>
		Sedentary	Light	Moderate	Vigorous	
<b>Actual Activity</b>	Sedentary	721	10	19	48	<b>90.35</b>
	Light	84	142	40	28	<b>48.29</b>
	Moderate	63	18	229	52	<b>63.26</b>
	Vigorous	57	6	25	443	<b>83.43</b>
<b>Precision</b>		<b>77.95</b>	<b>80.68</b>	<b>73.16</b>	<b>77.58</b>	

Table V.5: Final Detection Results

	<b>AC</b>	<b>PR</b>	<b>RE</b>	<b>F</b>
Drinking recognition	91.5	86.5	84.3	85.4
Body Activity recognition	89.12	77.58	83.43	80.39

A small qualitative survey is conducted on all the participants, which include 10 graduate students with the age from 26 to 37 years old. The results show that 87.5% of the users believe a wearable reminder will help them stay hydrated. While more than 55% of the users rated 5 for the ease of use, 33% of the users rated the system's comfort 3 on a scale of 1(poor) - 5(best). Moreover, around 50% of the users believe this system can benefit anyone.

## V.2 Negative Emotion Recognition

### V.2.1 Introduction

Along with physical health, mental health is also crucial to human life quality. Mental health includes human's emotional, psychological, and social well-being. According to [78], positive emotional states have a relationship with healthy patterns of responding in cardiovascular activity and the immune system. On the other hand, negative emotional states influence unhealthy patterns of physiological functioning and lowered immune activity. Many studies focusing on negative affect have revealed that negative affect including emotional states such as nervous, upset, angry, anxious, hostile, or depressed may increase human's susceptibility to illness [79, 80]. For example, hostility, anger, and aggressiveness are considered risk factors for coronary heart disease [81]. Apparently, it is crucial to help human get out of negative emotions to reduce the chance of developing physical health problems. In this case study, we develop an application which can recognize negative emotions by using ECG data collected from our smart shirt and human face expression. The result of negative emotions detection is uploaded in real-time to our private cloud and accessible to remote caregivers.

### V.2.2 Negative Emotion Recognition using ECG

#### A. Related Work

**Emotion Elicitation Methods:** Human's emotions are not constant, as they may change according to environmental conditions, memories, mood, etc. Therefore, it is certain that using the same stimuli techniques may elicit different emotions for different individuals. Even for the same person but in different situations, different emotions may be elicited. Furthermore, the same emotion may have different intensities. Therefore, it is not a trivial to verify what emotion is felt by the subject in an experiment [82]. Stimuli techniques used by researchers often ask participants to

watch film clips, look at pictures, listen to music, or participate in a computer game.

In [83], the elicitation method used in experiments is to let participants watch a series of short movie clips with easy-to-understand content to induce emotions such as anger, frustration, amusement, fear, surprise, and sadness. Between each clip, they have to give feedback about the emotion elicited, and the intensity of that emotion.

In [84], a study is conducted to evaluate the level of anxiety of three patients who are classified as euthymic, depressed or mixed-state. They first experience a resting state by closing their eyes for 5 minutes, then opening for another 5 minutes. Then, a sequence of images with different arousal and valence is shown to the subjects in 6 minutes and they have to give comments on the images. The images are from the IAPS (International Affective Picture System) database, which is widely used for eliciting emotions.

In [85], a study is conducted on male subjects who love to listen to music. The experiment is done in a quiet room to make sure that the subjects can experience the emotions. They are asked to pick four songs that are able to evoke emotional memories and certain moods corresponding to the four target emotions.

Another experiment, in [86], is conducted to elicit frustration by asking participants to play a computer game. They are supposed to complete the game in a certain amount of time to win a prize. However, at some points in time, the computer mouse is interfered to not follow the subjects' intention without their awareness. The frustration is induced after each failed mouse-click.

**Emotion Data Collection:** In [87], it is claimed that all emotions can be characterized in terms of judged valence (pleasant or unpleasant) and arousal (calm or aroused). Figure V.3 shows some named emotions as coordinates in the arousalvalence space [88]. The relation between physiological signals and arousal/valence is established in psychophysiology that argues that the activation of the autonomic nervous

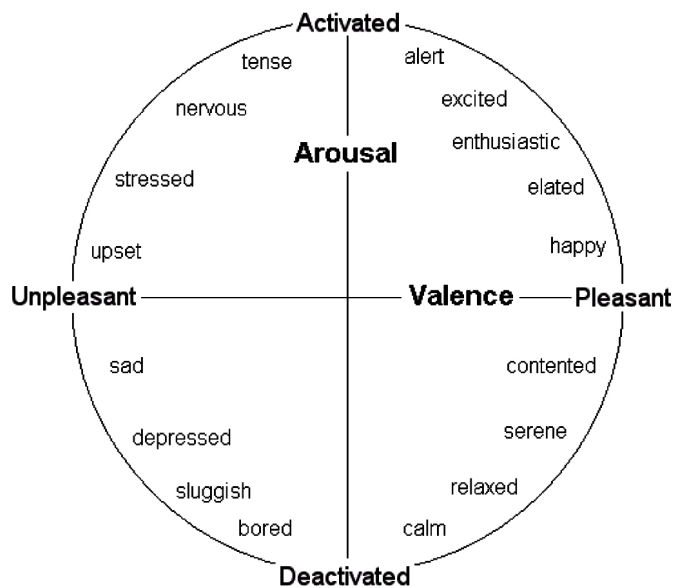


Figure V.3: A two-dimensional representation of emotional state.

system (ANS) changes while emotions are elicited [89]. Electrocardiogram (ECG) [90, 91, 11], Electroencephalography (EEG) [92, 93, 94], Electromyogram (EMG) [95, 96], or Skin Conductivity [96, 97] are physiological signals mostly used for emotion recognition. Another reason why physiological signals are chosen is that it is impossible or extremely difficult to control the natural physiological reactions of our body, such as changes in temperature, skin conductance or heart rate. On the other hand, it is possible to recognize emotions using facial expression, audio signals, body pose, gestures, etc. However, people are able to consciously control their own expression. Thus, physiological signals seem to be more reliable sources of information than human's expressions.

There are several public datasets that can be used for emotion recognition, for example, RECOLA [98] and DECAF [99]. The RECOLA dataset is obtained through experiments on 23 participants, in which their audiovisual and physiological data were collected. The video and audio data are recorded at 25fps and 44.1  $kHz$ , respectively. The bio-signals include ECG and Electrodermal Activity (EDA). In addition, the participants' personal information, such as age, gender, and mother tongue are also

recorded. The annotations are performed using the ANNEMO web-based annotation tool by six people, three females and three males. The label of emotions, represented as valence and arousal, is marked with a frame rate of 40ms. The DECAF dataset is collected from 30 participants (14 females and 16 males). The recorded biosignals include brain signal (Magnetoencephalogram -MEG), ECG, Bi-polar horizontal Electro-oculogram (hEOG) and Bi-polar trapezius Electro-myogram (tEMG). The Near-Infra-Red (NIR) facial data are also recorded. The emotional status is elicited by letting the participants watch different music videos and movie clips.

**Emotion Recognition:** In general, physiological signals are involuntary reactions of the human body and they are difficult to be faked. Thus, recognizing emotion based on physiological signals tends to have more reliability than external behaviors like facial expression or speech. Such signals as electrical activities of human brain and heartbeat have been mostly used in research of emotion recognition [100], [101], [102]. Compared to the brain signal, which needs a complicated system to acquire, the means used to collect the heartbeat signal are much simpler and easier to apply in practice. Nowadays, ECG signal can be acquired from the human body through wearable ECG sensors in the form of a smart shirt.

ECG-based emotion recognition has recently been adopted in areas such as human health and human-computer interaction. The ECG signal can be collected and processed in many different ways. For example, in [10], the authors propose a method to perform ECG synthesis, signal decomposition, and feature extraction based on the Empirical Mode Decomposition (EMD) method. They use the International Affective Picture System (IAPS) dataset and their own dataset for emotion recognition. Classification of arousal and valence is conducted but the accuracy is not high. In [90], based on ECG, Heart Rate Variability (HRV) is calculated. Time and frequency domain analysis is applied for feature extraction. Several nonlinear indices are ex-

tracted based on Approximate Entropy, Lagged Pointcare Plot and the Detrended Fluctuation Analysis (DFA). Their experimental results show a recognition accuracy of 84.72% on the valence dimension, and 84.26% on the arousal dimension. In [103], the non-linear feature Hurst is computed using Rescaled Range Statistics (RRS) and Finite Variance Scaling (FVS) methods. Several new Hurst features are proposed by combining the existing RRS and FVS methods with Higher Order Statistics (HOS). The accuracy of classifying six emotional states (happiness, sadness, fear, disgust, surprise and neutral) is 92.87%, and 76.45% using random and subject independent validation respectively.

## **B. Acquisition Setup for Emotion Elicitation**

Recent developments of Virtual Reality (VR) technology provides 3D immersive environment and has more impact on user emotion reactions. By wearing the VR goggle to watch videos, the wearer is almost fully isolated from the ambient environment, so it can minimize unwanted intervention from outside affecting his/her emotion. This equipment works with 3D videos, 360° videos, and even normal 2D videos. In our research, we use the Oculus Rift headset, as shown in Fig. V.4, and a set of videos to conduct experiments to induce participants' emotion. Compared to other studies using the traditional way of watching 2D videos, our method is expected to help participants induce their emotions easier. Videos are manually classified into different categories corresponding to emotional states such as anger, disgust, fear, sadness, neutral, calm, and happiness.

In the next section, several models are built to classify negative emotions based on two datasets, one from our experiment and one from DECAF dataset, which was conducted by researchers at University of Trento [99].





Figure V.4: Emotion elicitation using Oculus Rift.

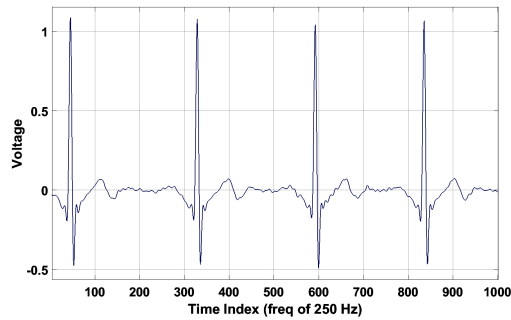


Figure V.5: An ECG signal of 4 seconds in length.

### C. Methodology

We use the smart shirt to collect the ECG data from the human. The raw ECG signal is filtered using a bandpass filter within the frequency band 5-15  $Hz$  to remove unwanted noise, for example, electrical and muscle noise. Then, features are extracted using a sliding window of 4-second width and 2-second step size. The ECG signal ready for analysis has a sampling rate of 250  $Hz$  as shown in Fig. V.5. From the ECG data, the heart rate and heart rate variability can be derived by detecting R peaks. Features in time and frequency domain, as well as nonlinear features, are extracted as the input to the classification models. However, in this paper, we focus on recurrence quantitative analysis (RQA), which is a method of nonlinear data analysis, to investigate the dynamics of heart's electronic reactivity to the emotion changes. There are only a few studies applying RQA on ECG signal to recognize

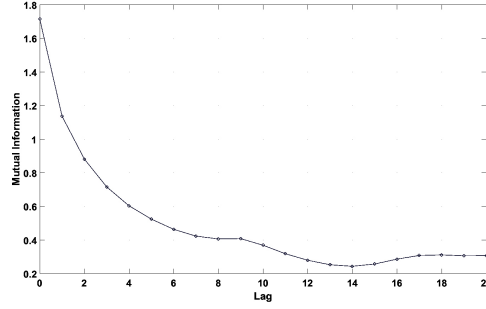


Figure V.6: ECG time-delayed mutual information.

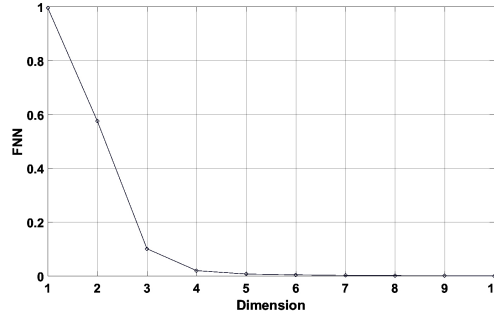


Figure V.7: False nearest neighbors test used for choosing dimension.

human emotion. Recently, Goshvarpour et al. [104] indicated that the RQA is one of the most significant features to differentiate two groups, men or women, based on ECG responses while watching sad images. It has also been shown that nonlinear features have the advantage in physiological signals processing to deal with negative affect.

**Recurrence Plot (RP):** The RQA measures are calculated based on recurrence plots, which is a graph visualizing the times that a phase space trajectory travels to the same place again [105]. The definition of the RP is shown below

$$R_{i,j} = \Theta(\epsilon_i - \|\vec{x}_i - \vec{x}_j\|), \quad \vec{x}_i \in \mathbb{R}^m, \quad i, j = 1, \dots, N, \quad (\text{V.8})$$

where  $N$  is the number of considered states  $x_i$ ,  $\epsilon_i$  is a threshold distance,  $\|\cdot\|$  a norm and  $\Theta(\cdot)$  the Heaviside function. In our case, only one time series of ECG signal is available, the phase space can be reconstructed by embedding dimension and time

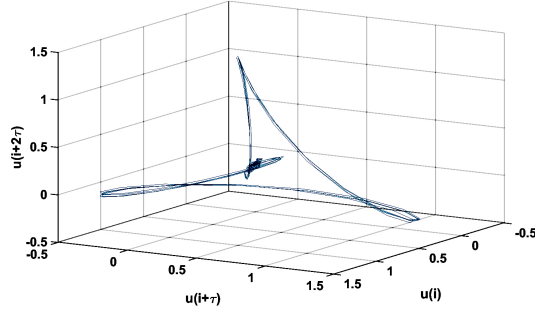


Figure V.8: ECG time-delayed embedding state space.

delay.

$$\vec{x}_i = (u(i), u(i + \tau), \dots, u(i + \tau(m - 1))), \quad (\text{V.9})$$

where  $u(i)$  is the time series,  $\tau$  is the embedding time delay and  $m$  is the embedding dimension. To estimate the time delay, we use mutual information analysis [106], and the time delay  $\tau$  which minimizes the mutual information is selected. As shown in Fig. V.6, the delay time  $\tau = 14$  is chosen. The embedding dimension  $m$  of the ECG signal is determined by using the False Nearest Neighbors (FNN) method [107]. An appropriate dimension is selected in such a way that most of the nearest neighbors do not move apart significantly in the next higher dimension. As shown in Fig. V.7, the embedding dimension  $m = 5$  is selected. The ECG time-delay embedding state space is shown in Fig. V.8. Each point of the phase space trajectory  $x_i$  is evaluated if it is close enough to another point of the trajectory  $x_j$  (less than a specified threshold  $\epsilon$ ). In the 2-D recurrence plot, the states at time  $i$  and time  $j$  are denoted by black points if  $R_{i,j} \equiv 1$  and white points if  $R_{i,j} \equiv 0$ . Fig. V.9 and Fig. V.10 show the recurrence plots of 4-seconds ECG data corresponding to non-negative and negative emotion label, respectively, with  $\tau = 14$ ,  $m = 5$  and  $\epsilon = 0.07$ .

**RQA measures:** Based on the RP, the RQA measures such as RR, DET, LMAX, VMAX, ENT, TND, LAM, and TT are extracted[108].

- Recurrence rate RR: The percentage of recurrence points in an RP:  $RR =$

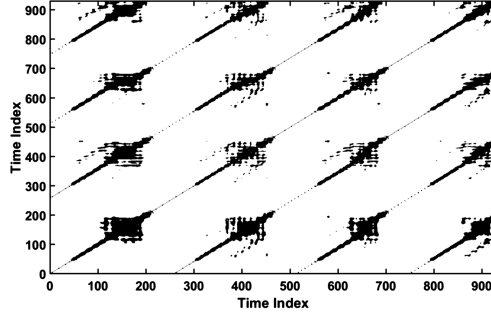


Figure V.9: Recurrence plot of ECG signal with non-negative emotion.

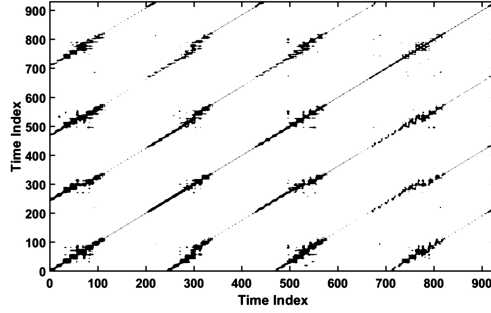


Figure V.10: Recurrence plot of ECG signal with negative emotion.

$$\frac{1}{N^2} \sum_{i,j=1}^N R_{i,j}$$

- Determinism DET: The percentage of recurrence points which form diagonal lines:

$$DET = \frac{\sum_{l=l_{\min}}^N lP(l)}{\sum_{l=1}^N lP(l)}$$

- Longest diagonal line LMAX: The length of the longest diagonal line:

$$LMAX = \max(\{l_i; i = 1, \dots, N_l\})$$

- Longest vertical line VMAX: The length of the longest diagonal line:

$$VMAX = \max(\{v_i; i = 1, \dots, N_v\})$$

- Entropy ENT: The Shannon entropy of the probability distribution of the diagonal line lengths  $p(l)$ :

$$ENT = - \sum_{l=l_{\min}}^N p(l) \ln p(l)$$

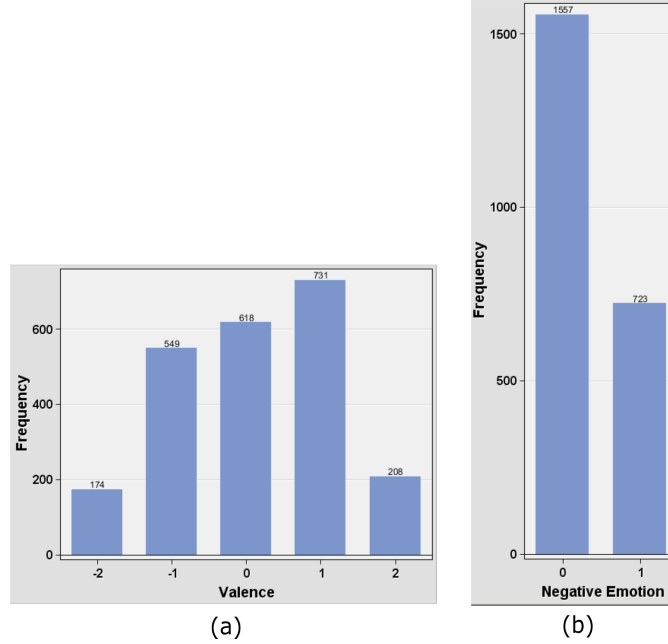


Figure V.11: (a) Original valence values; (b) New label with two categories: Negative and Non-negative.

- Trend TND: The paling of the RP towards its edges:

$$TREN D = \frac{\sum_{i=1}^{\tilde{N}} (i - \tilde{N}/2)(RR_i - \langle RR_i \rangle)}{\sum_{i=1}^{\tilde{N}} (i - \tilde{N}/2)^2}$$

- Trapping time TT: The average length of the vertical lines:

$$TT = TT = \frac{\sum_{v=v_{\min}}^N vP(v)}{\sum_{v=v_{\min}}^N P(v)}$$

- Laminality LAM: The percentage of recurrence points which form vertical lines:

$$LAM = \frac{\sum_{v=v_{\min}}^N vP(v)}{\sum_{v=1}^N vP(v)}$$

Where

$N$  the number of points on the phase space trajectory.

$N_l$  - the number of diagonal lines in the recurrence plot.

$N_v$  the number of vertical lines in the recurrence plot.

$P(l), P(v)$  - the histogram of the line lengths of diagonal/vertical lines.

$\tilde{N}$  - the maximal number of diagonals parallel to the LOI which will be considered for the calculation of TND.

**Classification:** The emotional state labels include negative (1) and non-negative (0). Negative emotions include anger, disgust, fear, and sadness, and non-negative emotions include calm, happiness and excitement. In terms of valence and arousal, we consider valence less than 0 negative emotion and valence greater than or equal to 0 non-negative emotion.

We apply the same methodology of data preprocessing and feature extraction to the RECOLA and DECAF dataset for comparison purpose. However, we need to do some extra data processing on emotion labels. In the RECOLA dataset, an emotion in terms of valence and arousal is labeled by six annotators every 40 ms. The valence value fluctuates significantly in small amounts of time, which is not feasible for the heart signal to react properly to the emotion changes. To deal with this issue, first, we take the average of all valence values evaluated from six annotators. A median filter is applied to smoothen the valence, and then the valence values are converted to a binary format with 0 as the threshold point. A sliding window with 4-second width and 2-second step size is used to loop through the labeled data. A window threshold  $\sigma$  is used, which determines the unique label of the whole window size of the data, i.e. 1000 data points. If there are more than  $\sigma\%$  of 1000 data points with the same label, then that label is assigned to the window. Otherwise, the window of data is not used for training. The  $\sigma$  value is chosen at 70%, 80% and 90%. In the DECAF dataset, an emotion label is assigned to a whole session of the experiment in which a person watches a movie clip. In this case, the label is kept unchanged and the preprocessing method is applied the same way as we do on our dataset.

An issue common to all three datasets after the binary label is created, as shown in Fig. V.11, is that the label becomes imbalanced, i.e. the observations with negative emotion are significantly less than the ones with non-negative, which may cause bias in the classification. Therefore, we perform under-sampling to make two classes of the label similar to each other. The emotion label and the RQA measures extracted from

Table V.6: Result of negative emotion classification.

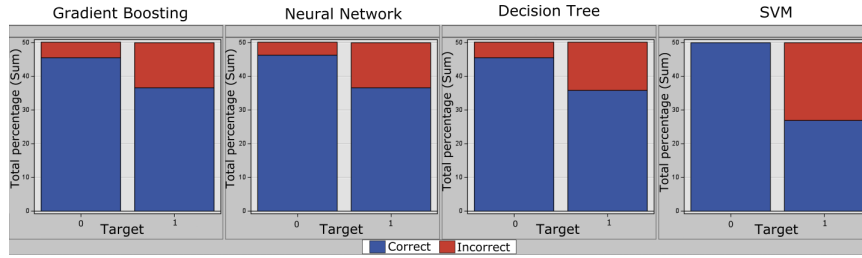
	Our Dataset	RECOLA			DECAF
		70%	80%	90%	
Neural Network	0.821	<b>0.771</b>	0.65	0.599	0.53
Decision Tree	0.813	0.747	0.638	0.63	<b>0.60</b>
Gradient Boosting	<b>0.828</b>	0.741	0.7	0.598	0.533
SVM RBF	0.769	0.671	0.639	0.609	0.529

the ECG data are used as input to the classification models. Multiple classification models such as Logistic Regression, Gradient Boosting, and Neural Network have been built.

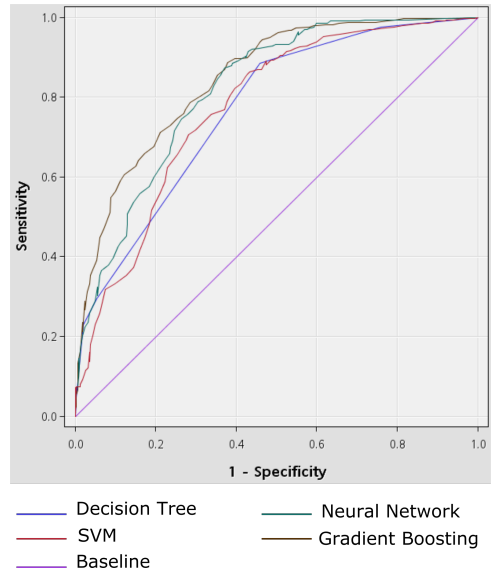
#### D. Experiments and Results

**Experimental Setup and Data Collection:** The experiment is conducted in a dark room so that the participants can be immersed in the virtual environment as much as possible. There are totally 10 participants, who are students in the age between 26 and 40. Each participant wears our smart shirt with textile electrodes integrated to collect the ECG signal while watching the videos.

To make sure they really experience the desired emotion, on one hand, we use different categories of video. On the other hand, a post-interview is conducted after each video to ask about their feeling. We also equip them with a wireless keypad so that they can press a button to mark the moment when they feel they are experiencing a certain emotion during experiments. The 1-minute ECG data around that moment is marked with the label of that emotion. All data from the shirt and the keypad are synchronized and sent to a computer wirelessly for storing and data analysis purposes.



(a) Classification result on our own dataset.



(b) ROC chart.

Figure V.12: Classification result on our own dataset.

**Results:** A summary of classification result on our dataset is shown in Fig. V.12. Fig. V.13 and Fig. V.14 show the classification results when applied on the RECOLA and DECAF datasets, respectively. As shown in Table. V.6, for our dataset, the Gradient Boosting model has the highest accuracy of classification, which is around 82.8%. The Neural Network results in the highest accuracy when it is applied to the RECOLA dataset with the window threshold  $\sigma = 70\%$ . When applying classification

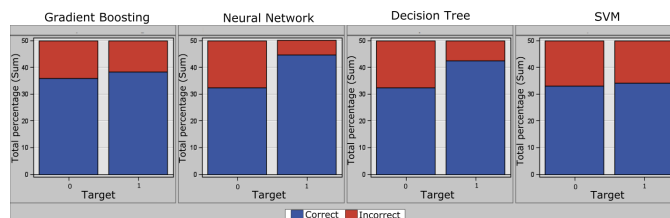


Figure V.13: Classification result on the RECOLA dataset.



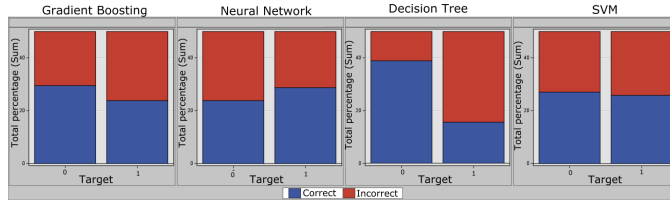


Figure V.14: Classification result on the DECAF dataset.

models on the DECAF dataset, however, the highest accuracy is only 60%.

The difference in classification accuracy among the three datasets may be caused by the difference in the way of eliciting emotions and labeling the emotions. In our data collection process, using the Oculus Rift offers an immersive VR environment to elicit expected emotions. An emotion label is marked by the participant while watching movie clips, and the corresponding 1-minute ECG signal is marked too. That 1-minute ECG reflects the emotion changes, and the training data built up from such ECG signals has more chances to result in higher accuracy.

### V.2.3 Negative Emotion Recognition using Facial Expression

ECG is a physiological signal, therefore it is believed to be a reliable data source to detect human emotion. However, during the process of data acquisition, the ECG signal collected does not always have a good quality since it is sensitive to the noise caused by the body movement. The bad quality of the signal will affect the model's accuracy. In this section, we take into account the facial expression, which is a supplemental source of data for emotion recognition.

## A. Review of Image Recognition Systems

Automated and real-time facial expression recognition has played an important role in human-computer interaction, social robots, and healthcare [109]. Based on modern psychology, six basic facial expressions are defined as universal emotions, including happiness, surprise, sadness, anger, fear, and disgust. Facial expression is generated

by movements of the muscles on eyebrow, mouth, nose, and eyes. Many studies have been conducted to recognize the human emotion based on the features extracted from those parts of the face. Traditionally, a typical facial expression recognition system consists of steps as shown in Fig. V.15. In the image acquisition step, a single image or image sequences are taken as input of the system, and they can be either grayscale or color images. The pre-processing is to enhance the quality of the input image by reducing noise, smoothing, filtering and normalizing the image. The segmentation step divides the image into homogeneous, self-consistent regions corresponding to different objects in the image on the bases of texture, edge, and intensity. The feature extraction step is to extract meaningful information from shape, motion, color, texture of the facial image. In the classification step, the features extracted from the previous step become the input of classification models, which classify the image into a certain category of facial expression. For example, in [110], k-nearest neighbor classifiers and multilayer perceptron neural networks are used as classification models.

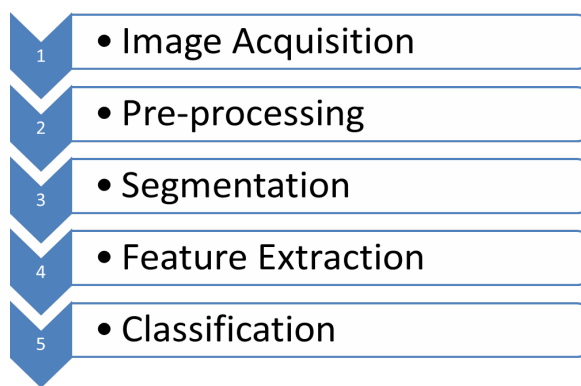


Figure V.15: A traditional facial expression recognition system.

In the past few years, Deep Neural Networks (DNNs) has become a prominent technique in the field of computer vision [111]. Among DNNs, Convolutional Neural Network (CNN) is a special kind of multi-layer neural networks, which is used to recognize visual patterns directly from pixels of 2D or 3D images [112]. And the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) [113] came to be a

Table V.7: CNN architectures of ILSVRC top competitors.

Year	CNN	Top-5 error rate	Number of parameters
2012	AlexNet	15.3%	60 million
2014	GoogLeNet	6.67%	4 million
2014	VGG Net	7.3%	138 million
2015	ResNet-50	3.6%	25.6 million

benchmark in object category classification on hundreds of object categories and millions of images. The ImageNet dataset consists of photographs collected from Flickr and other search engines. They are manually labeled with the presence of one of 1000 object categories. Many studies have been conducted using this dataset to build and evaluate their CNN models. AlexNet, GoogleNet, VGGNet and ResNet are CNN architectures of ILSVRC top competitors and an analysis was performed to compare those models in terms of accuracy, memory footprint, parameters, operations count, inference time and power consumption. Table V.7 [114] shows a summary of those models with their Top-5 error rate and the number of parameters.

## B. Facial Expression Databases

There are several online databases that provide collections of facial images along with emotion labels. The labels are annotated manually or automatically. In this study, we collect facial images databases that have the emotion annotation being done manually in discrete emotion labels, for example, sadness, happiness, fear, disgust, etc. Below is a brief description of the datasets we use for training and testing our negative emotion detection models.

**AffectNet:** AffectNet database [115] contains more than 1,000,000 facial images collected from the Internet by querying three major search engines using 1250 emotion-related keywords in six different languages. About half of the retrieved images were manually annotated for the presence of seven discrete facial expressions and the intensity of valence and arousal. Up to now, AffectNet is the largest database of facial

expression, valence, and arousal in the wild enabling research in automated facial expression recognition.

**MUG:** MUG (Multimedia Understanding Group) is a research group of the Information Processing Laboratory of the Electrical and Computer Engineering Department of the Aristotle University of Thessaloniki. MUG facial expression database [116] includes the recordings of 77 Caucasian subjects. There are totally and 1462 labeled facial images. The subjects were asked to perform the six basic expressions, which are anger, disgust, fear, happiness, sadness, and surprise.

**JAFFE:** Japanese Female Facial Expression (JAFFE) database [117] contains 213 images of 7 facial expressions (6 basic facial expressions + 1 neutral) posed by 10 Japanese female models. Each image has been rated on 6 emotion adjectives by 60 Japanese subjects.

**ISED:** Indian Spontaneous Expression Database (ISED) [118] contains facial images of 26 Indian participants, which are categorized into 4 emotion classes: happiness, surprise, sadness, and disgust.

**CK+:** The Extended Cohn-Kanade Dataset [119] includes facial behavior of 210 adults recorded using two hardware synchronized Panasonic AG-7500 cameras. Participants were 18 to 50 years of age, 69% female, 81% Euro-American, 13% Afro-American, and 6% other groups.

**RFD:** Radboud Faces Database [120] contains portrait images of 49 models in two subsets: 39 Caucasian Dutch adults (19 female), and 10 Caucasian Dutch children (6 female). All models showed eight facial expressions including neutral, anger, sadness, fear, disgust, surprise, happiness, and contempt, with three gaze directions. In this

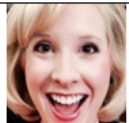
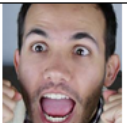






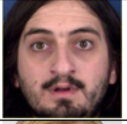
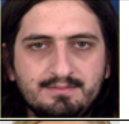
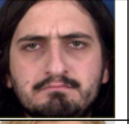
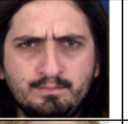
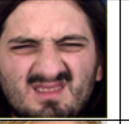
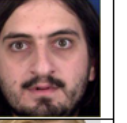
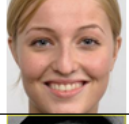










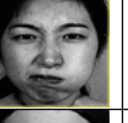





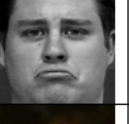


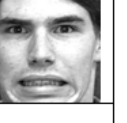

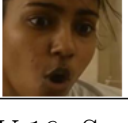
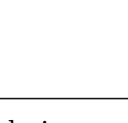
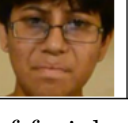
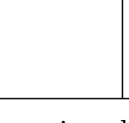

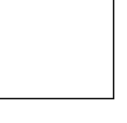
	Happiness	Surprise	Neutral	Sadness	Anger	Disgust	Fear
Affect-Net (~400K)							
MUG (401)							
RFD (1407)							
JAFFE (214)							
CK+ (309)							
ISED (428)							

Figure V.16: Sample images of facial expression databases.

dataset, faces are captured by five cameras with different angles ( $0^\circ$ ,  $45^\circ$ ,  $90^\circ$ ,  $135^\circ$ , and  $180^\circ$ ). Sample images of facial expression databases are shown in Fig. V.16.

### C. Methodology

We follow a transfer learning approach to develop our CNN network, i.e. starting from a network pre-trained on the generic ImageNet dataset. We choose GoogLeNet to be the CNN architecture to apply on our datasets. The GoogLeNet was the winning architecture on ImageNet 2014 and it has much less number of parameters compared to other deep neural network architectures. Fine-tuning a network with transfer learning is faster and easier than training a network from scratch with randomly initialized weights. Learned features can be transferred to a new task of emotion classification using a smaller number of training facial images. Our transfer learning method is shown in Fig. V.17.

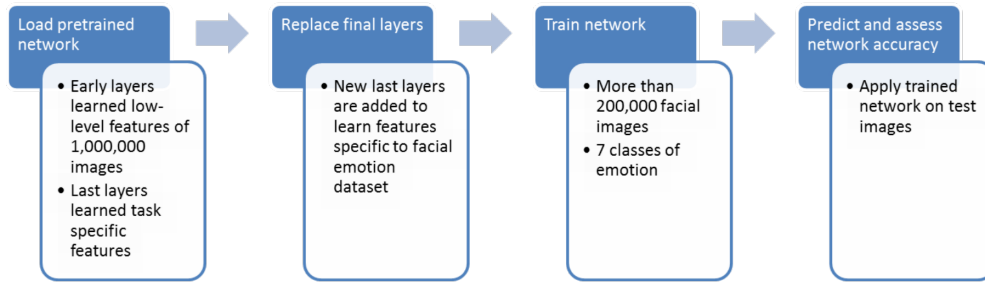


Figure V.17: Transfer learning method.

The facial expression dataset used in the transfer learning is AffectNet. However, we just consider the portion that has been manually labeled and matched with the labels when classified by Microsoft Azure [121]. This helps increase the reliability of the ground truth of the dataset. The number of images used for transfer learning is 17,534 and the number of images for testing is 1,427. The image files are located in 7 different folders corresponding to 7 classes of emotion. Out of 17,534 images used for building models, 70% and 30% of the images are used for training and validation, respectively.

GoogLeNet has used over a million images to train, learned rich feature representations for a wide range of images and it can classify images into 1000 object categories. GoogLeNet takes an image of size 224x224x3 as input and then generates a label for the object in the image with the probabilities for each of the object categories. In order to retrain the GoogLeNet to classify facial expression images, the last three layers of the GoogLeNet should be replaced. These three layers, 'loss3-classifier', 'prob', and 'output', contain information on how to combine the features that the GoogLeNet extracts into class probabilities and labels. Three new layers added to the layer graph include a fully connected layer, a softmax layer, and a classification output layer. The output size of the final fully connected layer is to 7, which is the number of emotion classes in the new dataset.



(a) The training process and result of the facial expression classification model.



(b) Examples of faces classified by the facial expression classification model.

Figure V.18: Transfer learning using GoogLeNet for 7-emotion recognition.

## D. Results

**Seven emotions recognition** Before training, some parameters need to be specified. When performing transfer learning, a large number of epochs is not necessary, so we set the number of epochs to 6. The mini-batch size is 32, the initial learning rate is 0.0005, the validation frequency is 5. We also freeze the weights of earlier layers in the network by setting the learning rates in those layers to zero, which helps speed up network training. The number of frozen layers is set to be 0, 10, 110, and 141 for each training. The network is configured and trained by the Matlab Deep Learning Toolbox with the usage of a single GPU.

A result of the network trained with first 10 layers frozen is shown in Fig. V.18. The validation accuracy is 90.06%. More details of the result are shown in Table

Table V.8: Result of transfer learning using GoogLeNet.

Accuracy	141 Layers Frozen	110 Layers Frozen	10 Layers Frozen	No Layers Frozen	Microsoft Azure
Validation accuracy	54.63%	79.58%	90.06%	96.58%	
Applied on MUG	30.42%	42.89%	59.60%	66.08%	61.34%
Applied on JAFFE	15.57%	26.42%	42.45%	38.68%	31.13%
Applied on ISED	33.41%	44.39%	54.44%	57.94%	57.24%
Applied on CK+	35.92%	56.63%	66.67%	76.05%	84.78%
Applied on RFD	46.13%	53.30%	71.86%	82.90%	80.06%

V.8. We test our final model on other datasets such as MUG, JAFFE, ISED, CK+, and RFD. Microsoft Azure, which is considered a benchmark, is also used to classify facial images in those datasets. The network with no layers frozen turns out to be better than any network trained with layers frozen. Its validation accuracy is 96.58%. Compared to Microsoft Azure, it is a little bit better on all test datasets except the CK+ dataset.

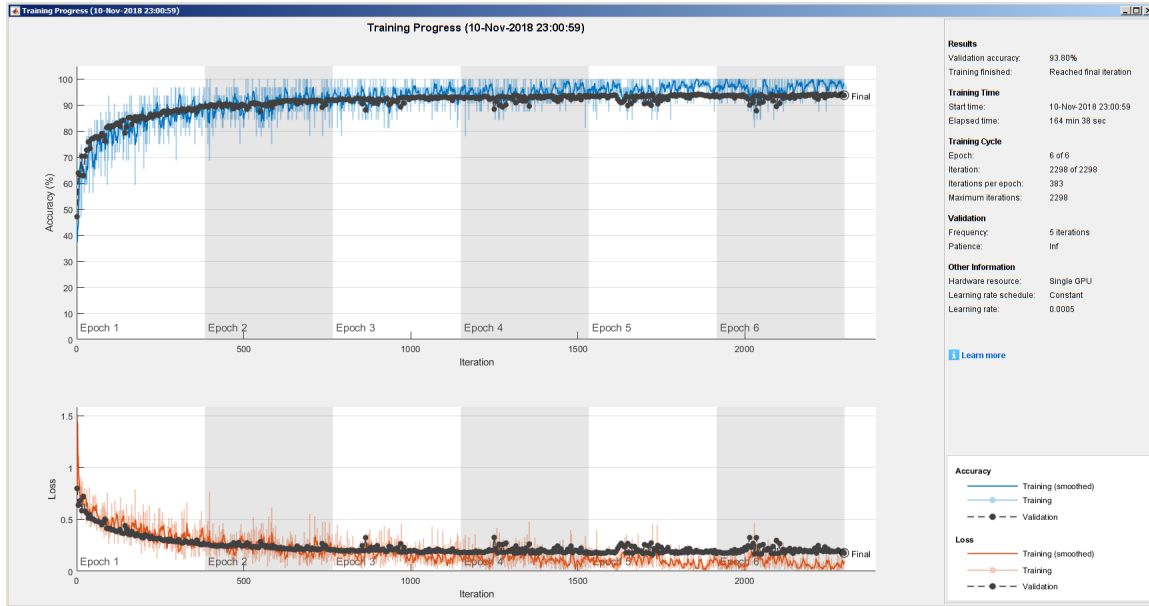
**Negative Emotion Recognition:** Another model is trained using the transfer learning method based on the previous model trained with seven emotions. The difference of this model is in the final fully connected layer, in which the parameter is set to 2 for binary classification purpose. The dataset is regrouped in 2 classes, which are Negative containing Sadness, Anger, Disgust and Fear emotion and Non-negative containing Happiness, Neutral, and Surprise emotion.

The result of negative emotion detection model is shown in Fig. V.19, which has a classification accuracy of around 93.8% on the validation set. In the next section, this Face-based negative emotion detection model along with the ECG-based model built in the previous part are combined together in a fusion model.

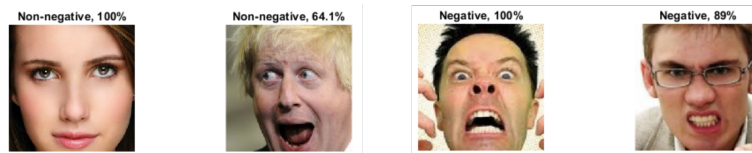
#### V.2.4 Multimodal Fusion Framework

Our robot assistant is equipped with cameras for human interaction purpose, so it is able to capture images of the environment. We can utilize images of the human





(a) Result of the Face-based negative emotion classification model.



(b) Examples of faces classified by the Face-based negative emotion classification model. Figure V.19: Transfer learning using GoogLeNet for negative emotion recognition.

face for emotion recognition. Facial expression can be integrated along with ECG signals to complement each other in order to recognize human emotion more reliably. For example, when the robot is too far away from the human to collect the voice, or the lighting condition is poor then the physiological signals acquired from the shirt can be used. However, the shirt cannot be worn all the time, or in some cases in which the human has vigorous activities such as doing exercises, the signals collected from the shirt may not be good for emotion detection, then the information from facial expression and voice should be used instead. Therefore, the approach of fusing multiple sensors to detect emotion may result in higher accuracy in emotion detection.

There are two approaches of fusion: Fusion at feature level and fusion at decision level. The feature-level fusion requires all data sources to be synchronized during the data acquisition step so that separated feature vectors of each single model can be

combined together to build a feature vector to be the input of a final model. The decision-level fusion, on the other hand, lets each model do its own classification task, then their outputs are fused to give a final result. The decision-level fusion is suitable to our negative emotion detection. In our system, ECG signals are collected and processed using a smart garment, whereas face images are captured and processed using the robot camera. We can reduce the workload and communication bandwidth of raw sensor data transmission, thus prolong the battery life of the smart garment and allow physiological data and human health status to be updated seamlessly.

There are several decision rules for binary classification problem, for example maximum likelihood (ML), maximum a posterior probability (MAP), Bayes' minimum average cost rule [122], and Hidden Markov models (HMM) decision fusion. The concepts of likelihood ratio test (LRT), the probabilities of false positive and false negative, and Bayes theorem are often employed by those decision rules. We choose the Bayes' minimum average cost rule for the decision fusion in the negative emotion recognition.

### **A. Contextual Factors**

We need to consider different scenarios that affect the data acquisition and subsequent classification result. Since the ECG signal is collected by using skin-contact textile electrodes, it is sensitive to the body movement which causes the loose contact between the electrodes with the human skin. The quality of the ECG signal has an impact on the accuracy of the ECG model. If the ECG signal is significantly degraded due to the human activities, the ECG-based model should be ignored. Regarding the negative emotion detection based on face image, in the case of the camera being unable to capture the face, the Face-based model should not be employed. If an image of a portion of the face is captured, the subsequent detection output of the model should have a lower accuracy compared to a captured image of a full face. An experiment



Figure V.20: A setup to simulate human movement while human face is captured.

was conducted to quantify the relationship between the accuracy of the Face-based model and the yaw angle of the face, which is considered an input for the decision rules. Another purpose of this experiment was to find out if human activities have impact on the ECG signal, which assists a decision to reject or not the ECG-based model.

In the experiment, we used a setup of a treadmill to simulate the human movement and three cameras in front of the treadmill to capture human face at different angles. The setup is shown in Fig. V.20. Six people were recruited to perform the experiments. They were asked to stand still, walk and run on the treadmill while showing 7 emotional expressions on their face, including happiness, surprise, neutral, sadness, fear, disgust and anger. The motion data, ECG signals, and face images were collected.

We build a test set of facial expressions captured at different angles from 6 people.

Table V.9: Yaw angle categories and corresponding model accuracies.

Yaw Angle	Accuracy of Face-based negative emotion detection model
$ YA  \leq 10^\circ$	88.53%
$10^\circ <  YA  \leq 25^\circ$	85.61%
$25^\circ <  YA  \leq 40^\circ$	83.64%
$ YA  > 40^\circ$	70.59%



Figure V.21: Examples from the test set including facial images with different yaw angles.

The data set of facial images are grouped into 2 classes: Negative (sadness, fear, disgust and anger) and Non-Negative (happiness, surprise, and neutral). We further categorize the data set based on the yaw angle of the face, which is calculated based on the head pose estimation using Convolutional Neural Networks and adaptive gradient methods [123]. An example is shown in Fig. V.21, in which the negative degrees mean the head turning to the right, and the positive degrees mean the head turning to the left. The yaw angle (YA) categories are  $|YA| \leq 10^\circ$ ,  $10^\circ < |YA| \leq 25^\circ$ ,  $25^\circ < |YA| \leq 40^\circ$ , and  $|YA| > 40^\circ$ . The Face-based model, which is trained in the previous section, is applied on each category and the classification accuracies are shown in Table V.9. It shows that this model is affected by the yaw angle of the face when recognizing the negative emotions. There is a variation of the accuracy level among the categories of the yaw angle. The accuracy is significantly lowered when  $|YA| > 40^\circ$ .

We also build a dataset including ECG signal and motion data to investigate the relationship between human activity levels and the quality of the ECG acquired. Each person performs activities such as standing, walking (velocity less than 4 mph)

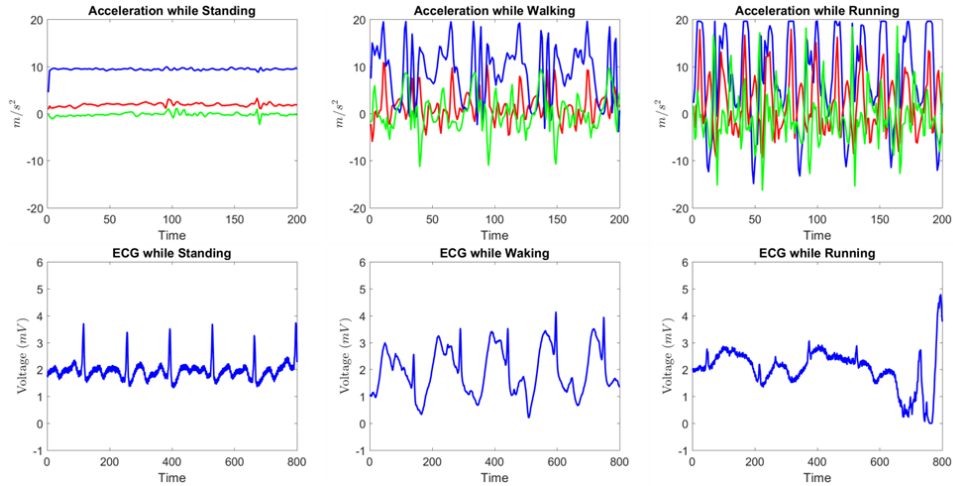


Figure V.22: Sample dataset of ECG signals and human activities.

Age (Year)	MHR (bpm)	Normal range of Heart Rate (bpm)	
		Moderate activities	Vigorous activities
30	186	93 – 130.2	130.2 – 158.1
60	165	82.5 – 115.5	115.5 – 140.3
90	144	72 – 100.8	100.8 – 122.4
<b>Regardless of the age</b>		<b>72 – 130.2</b>	<b>100.8 – 158.1</b>

Figure V.23: Normal heart rate ranges corresponding to activity levels.

and jogging/running (velocity greater than 4 mph) on the treadmill while the ECG is collected. The sample data collected is shown in Fig. V.22. The ECG and the motion data are divided into small observations which comprise of 4-second chunks of data. In order to label the ECG data as good or bad, we calculate the heart rate from the ECG signal and compare to its normal range corresponding to each activity level as shown Fig. V.23 [124, 125]. The Maximum Heart Rate (MHR) is calculated by using the formula  $MHR = 207 - 0.7 \times age$ . The normal ranges of heart rate are 60 - 100 bpm (beat per minute), 50% - 70% MHR, and 70% - 85% MHR for sedentary(e.g. standing), moderate (e.g. walking) and vigorous activity (e.g. jogging,running), respectively. In our dataset, the heart rate is calculated based

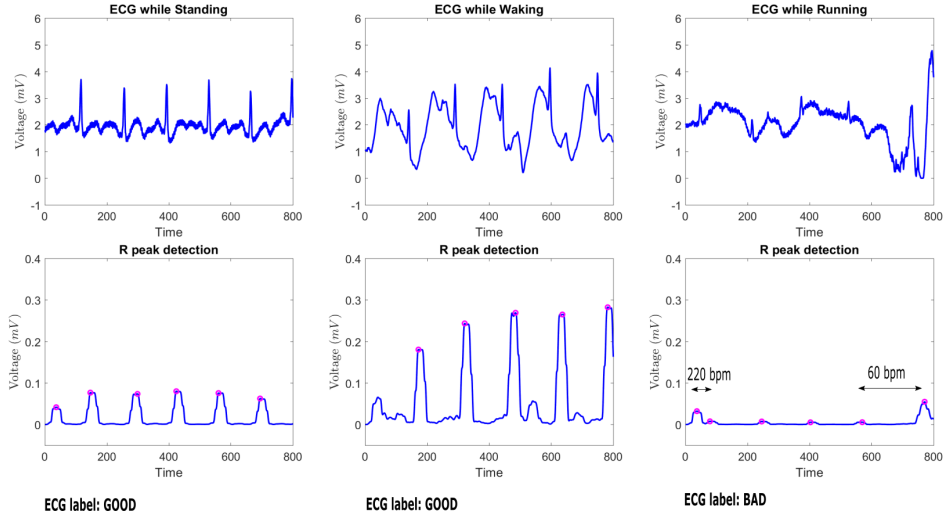


Figure V.24: Labeling the data based on the normal ranges of heart rate.

on R-R intervals detection using Pan Tompkins algorithm [20]. If the heart rate is in the normal range regardless of the age, the ECG is labeled as good; Otherwise, the ECG is labeled as bad as shown in Fig. V.24. The results show that in the category of vigorous level, 98% of observations have the heart rate out of the normal range (100.8 - 158.1 *bpm*). It means the ECG is unacceptable if activity level is vigorous. In the categories of moderate and sedentary level, the heart rates are in the normal ranges (72 - 130.2 *bpm*, and 60 -100 *bpm*, respectively).

## B. Methodology

If the ECG signal is good and the face is captured, a decision rule is employed in consideration of the weights determined by the probability of false alarm (false positive rate) and the probability of miss (false negative rate) of each individual models. The weight of the Face-base model is also affected by another parameter, which is the yaw angle of the face. The methodology overview is described in Fig. V.25.

Let us consider two hypotheses:

- $H_0$ : The observation is labeled as Non-negative emotion.
- $H_1$ : The observation is labeled as Negative emotion.

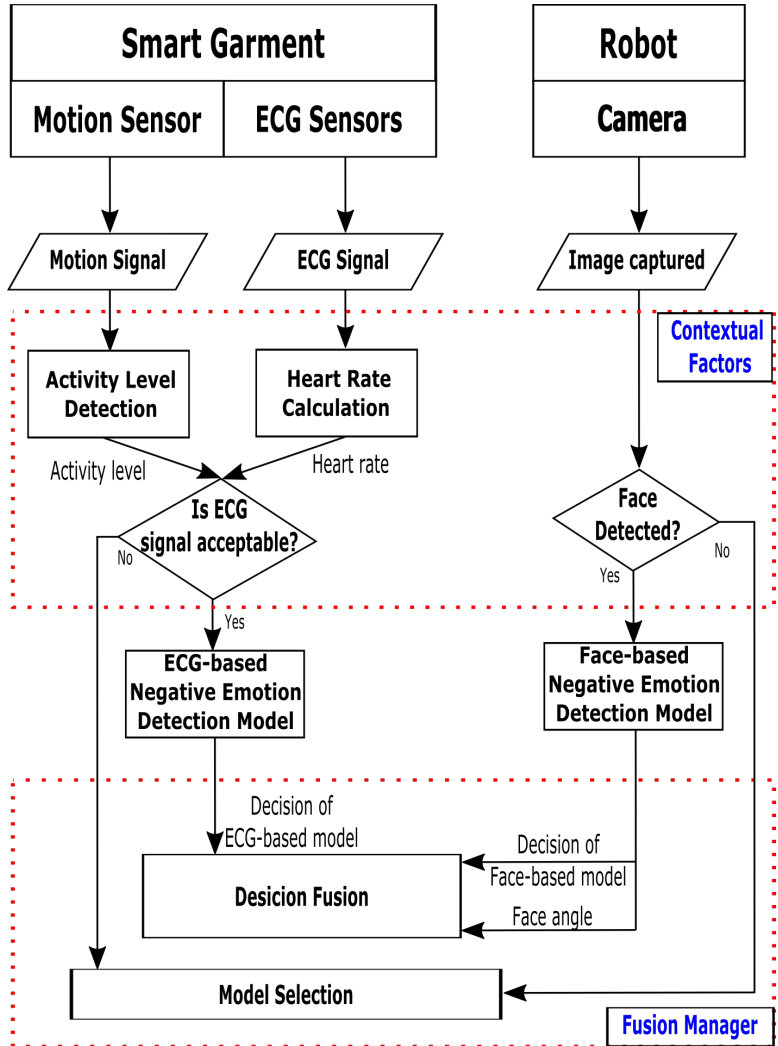


Figure V.25: Decision fusion diagram used to recognize the negative emotion.

The priori probabilities of the two hypotheses are denoted by  $P(H_0) = P_0$  and  $P(H_1) = P_1$ . The decision of the ECG-based and Face-based negative emotion detection model is denoted by  $E$  and  $F$ , respectively and they can output 0 or 1, for example

- $E = 0$ : The observation is detected as Non-negative emotion using ECG-based model.
- $E = 1$ : The observation is detected as Negative emotion using ECG-based model.

Table V.10: Fusion Rule based on two individual models.

<b>E (ECG-based model)</b>	<b>F (Face-based model)</b>	<b>Fusion</b>
0	0	Fu1
0	1	Fu2
1	0	Fu3
1	1	Fu4

The final decision after fusing the two models is denoted as  $Fu = f(E, F)$ , which also can be 0 or 1 as shown in Table. V.10.

The probability of false alarm is denoted as  $P_{FE}$  and  $P_{FF}$ , and the probability of miss is denoted as  $P_{ME}$  and  $P_{MF}$  for the ECG-based and Face-based model respectively.

- ECG-based model:  $P_{FE} = P(E = 1|H_0)$  and  $P_{ME} = P(E = 0|H_1)$
- Face-based model:  $P_{FF} = P(F = 1|H_0)$  and  $P_{MF} = P(F = 0|H_1)$

Based on [122], the optimum decision rule is given by the likelihood ratio test as follow

$$\frac{P(E, F|H_1)}{P(E, F|H_0)} \underset{H_1}{\overset{H_0}{>}} \frac{P_0 \cdot (C_{10} - C_{00})}{P_1 \cdot (C_{01} - C_{11})} \quad (\text{V.10})$$

in which,  $C_{01}$  is the cost of misses and  $C_{10}$  is the cost of false alarms. The values of  $C_{01}$  and  $C_{10}$  can be selected due to a variety of criteria. In this case study, we consider the false alarm not a costly problem. We use a robot, which is mentioned in the next chapter, to interact with the human when a negative emotion is detected. The false alarms can increase the frequency of the robot interaction with the human. It may cost more power and other resources, but it is not a big concern. On the other hand, getting the older adults, especially who are living alone, involved in conversations with the robot is crucial for them to maintain an active social life. In the case of misses, when the human truly has a negative emotion which is not detected, there is no interaction from the robot. We definitely want to reduce the chance of its occurrence. Therefore, we evaluate the cost of misses ( $C_{01}$ ) higher than the cost of



false alarms ( $C_{10}$ ). Also, there is no cost when the classification is correct, which is  $C_{00} = C_{11} = 0$ . The Equation V.10 becomes

$$\frac{P(Fu|H_1)}{P(Fu|H_0)} \underset{H_1}{\overset{H_0}{\gtrless}} \frac{P_0 \cdot C_{10}}{P_1 \cdot C_{01}}$$

Using Bayes rule, we have

$$\frac{C_{01} \cdot P(H_1|Fu) \cdot P(Fu)}{C_{10} \cdot P(H_0|Fu) \cdot P(Fu)} \underset{H_1}{\overset{H_0}{\gtrless}} 1$$

The log-likelihood ratio test is

$$\log \frac{C_{01} \cdot P(H_1|Fu)}{C_{10} \cdot P(H_0|Fu)} \underset{H_1}{\overset{H_0}{\gtrless}} 0$$

$$\log \frac{C_{01} \cdot \frac{P(H_1) \cdot P(Fu|H_1)}{P(Fu)}}{C_{10} \cdot \frac{P(H_0) \cdot P(Fu|H_0)}{P(Fu)}} \underset{H_1}{\overset{H_0}{\gtrless}} 0$$

$$\log \frac{C_{01} \cdot P_1}{C_{10} \cdot P_0} + \log \frac{P(Fu|H_1)}{P(Fu|H_0)} \underset{H_1}{\overset{H_0}{\gtrless}} 0$$

Under the assumption of conditional dependence among local sensor observations (face images and ECG signals), we have

$$\log \frac{C_{01} \cdot P_1}{C_{10} \cdot P_0} + \log \frac{P(E|H_1) \cdot P(F|H_1)}{P(E|H_0) \cdot P(F|H_0)} \underset{H_1}{\overset{H_0}{\gtrless}} 0$$

$$\log \frac{C_{01} \cdot P_1}{C_{10} \cdot P_0} + \log \frac{(1 - P_{ME})^E \cdot (P_{ME})^{1-E} \cdot (1 - P_{MF})^F \cdot (P_{MF})^{1-F}}{(P_{FE})^E \cdot (1 - P_{FE})^{1-E} \cdot (P_{FF})^F \cdot (1 - P_{FF})^{1-F}} \underset{H_1}{\overset{H_0}{\gtrless}} 0$$

$$\begin{aligned} \log \frac{C_{01} \cdot P_1}{C_{10} \cdot P_0} + E \cdot \log \frac{1 - P_{ME}}{P_{FE}} + (1 - E) \cdot \log \frac{P_{ME}}{1 - P_{FE}} \\ + F \cdot \log \frac{1 - P_{MF}}{P_{FF}} + (1 - F) \cdot \log \frac{P_{MF}}{1 - P_{FF}} \end{aligned} \underset{H_1}{\overset{H_0}{\gtrless}} 0$$

$$\begin{aligned} \log \frac{C_{01} \cdot P_1 \cdot P_{ME} \cdot P_{MF}}{C_{10} \cdot P_0 \cdot (1 - P_{FE}) \cdot (1 - P_{FF})} + E \cdot \log \frac{(1 - P_{ME}) \cdot (1 - P_{FE})}{(P_{FE}) \cdot (P_{ME})} \\ + F \cdot \log \frac{(1 - P_{MF}) \cdot (1 - P_{FF})}{(P_{FF}) \cdot (P_{MF})} \end{aligned} \underset{H_1}{\overset{H_0}{\gtrless}} 0$$

Therefore, we have a final decision rule which can be expressed as

$$\mathbf{F}\mathbf{u} = \begin{cases} \mathbf{1}, & \text{if } \mathbf{a}_0 + \mathbf{a}_1 \cdot \mathbf{E} + \mathbf{a}_2 \cdot \mathbf{F} > \mathbf{0} \\ \mathbf{0}, & \text{otherwise} \end{cases}$$

Where, the weight  $a_0, a_1$  and  $a_2$  are given by

$$\begin{aligned} a_0 &= \log \frac{C_{01} \cdot P_1 \cdot P_{ME} \cdot P_{MF}}{C_{10} \cdot P_0 \cdot (1 - P_{FE}) \cdot (1 - P_{FF})} \\ a_1 &= \log \frac{(1 - P_{ME}) \cdot (1 - P_{FE})}{(P_{FE}) \cdot (P_{ME})} \\ a_2 &= \alpha \cdot \log \frac{(1 - P_{MF}) \cdot (1 - P_{FF})}{(P_{FF}) \cdot (P_{MF})} \end{aligned}$$

Because the Face-based model has different accuracy levels when applying into facial images with different yaw angles of the face, it yields a better classification accuracy for images with small yaw angle of the face. The weight  $a_2$  should be adjusted by a penalty parameter  $\alpha$  according to the yaw angle of the face. We have

$$\alpha = \begin{cases} 0.8853, & |YA| \leq 10^\circ \\ 0.8561, & 10^\circ < |YA| \leq 25^\circ \\ 0.8364, & 25^\circ < |YA| \leq 40^\circ \\ 0.7059, & |YA| > 40^\circ \end{cases} \quad (\text{V.11})$$

### C. Results

Assuming there are outputs from ECG-based and Face-based models, to validate the decision fusion, we picked RECOLA dataset which includes both ECG signal and facial images. The test dataset includes 2560 observations, in which there are 422 negative emotion observations and 2138 non-negative emotion observations. The ground truth valence labels measured by 6 annotators are grouped in every 4-second, averaged out and compared with a threshold to determine a label in the test dataset as non-negative or negative emotion.

We apply both ECG-based and Face-based emotion recognition models on that test dataset. The probability of false alarm and probability of miss of each model are

Table V.11: Confusion matrices of negative emotion recognition models.

		ECG-based model		Face-based model		Fusion model	
		0	1	0	1	0	1
Actual Emotion	0	1719	419	1680	389	1668	470
	1	247	175	200	291	93	329

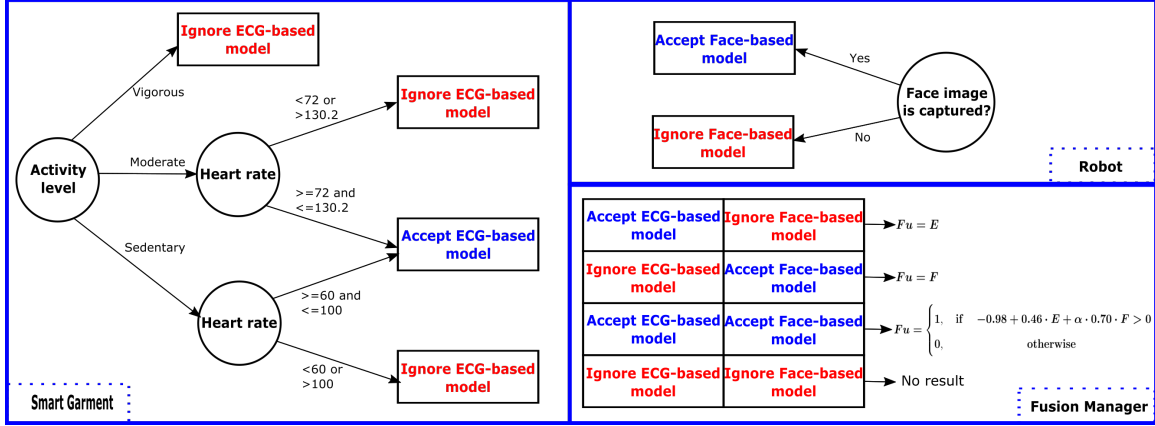


Figure V.26: A summary of the fusion rule.

calculated. Since we evaluate the cost of false alarm  $C_{10}$  less than the cost of miss  $C_{01}$ , we assume  $C_{10} = 2$  and  $C_{01} = 2.5$ . By applying Equation V.11, we can calculate the final decision of the fusion model. The confusion matrices of 3 models are shown in Table V.11. The weight  $a_0, a_1, a_2$  are calculated, in which  $a_0 = -0.98$ ,  $a_1 = 0.46$  and  $a_2 = \alpha \cdot 0.70$ . The accuracy of the fusion model is 78%, compared to 74% of the the ECG-based and 77% of the Face-based model, when they are applied on the RECOLA test set.

In other scenarios where contextual information affects the quality of the ECG signal and the availability of the facial image, the Model Selection module is used to select one model to result the output. If the ECG is not acceptable, Face-base model is selected. If human face is not captured, ECG-based model is selected. If the ECG is unacceptable and the face is not captured, no output is generated from the system. A summary of the fusion is shown in Fig. V.26.

### V.2.5 Summary

In this chapter, we described two case studies of how to utilize CoSHE in well-being monitoring, the body hydration monitoring and negative emotion detection. The body hydration monitoring was done by detecting drinking sound and body activity levels with the accuracy of 91.5% and 89.12%, respectively. It can provide the information of the amount of fluid taken, daily activity, and amount of fluid that should be taken for the day. However, the comfort of the throat microphone needs to be improved and the system power consumption needs to be optimized. Regarding the negative emotion, it can be detected by ECG signal and facial expression with the accuracy of 82.8% and 93.8%. A decision fusion model was built from ECG-based and Face-based model. It can expand the capability of recognizing the negative emotion in more scenarios of daily life. In the decision rule, the human activity levels can assist the decision whether or not the ECG is good to be used in recognizing the negative emotion. The face angle is another factor included in the decision rule to adjust the weight of the Face-based negative emotion recognition model. In the next chapter, we introduce a robot assistant that can interact with the human via conversations in case of anomalies.

## CHAPTER VI

### HEALTHCARE DELIVERY THROUGH A ROBOT ASSISTANT

This chapter introduces a robot assistant, its hardware components and functionalities. Then, two use cases demonstrating how the robot is used in hydration reminder and emotion regulation applications are presented.

#### VI.1 Introduction

The CoSHE itself can provide human context information and human health status, which can be accessed in real-time by a remote caregiver or doctor. Then, the remote caregiver should contact the older adult or if needed to double check or give out necessary instructions, for instance. However, there are cases when the caregiver cannot reach out to the older adult via phone or video chat, especially when the adult has an issue, for example, an injurious fall or having symptoms of a stroke in a private area. A robot assistant is a good solution in such cases. The robot assistant is useful if it is able to understand the human context information and health status, move to the human location, and interact with the human. In our study, the robot assistant is integrated into the CoSHE system to form a closed-loop healthcare system as depicted in Fig. VI.1. The robot assistant is able to access the cloud database and query the context information as well as the health status of the human. It is responsible for delivering healthcare services whenever the health monitoring system raises a red flag, i.e., when dehydration status is detected or negative emotions are recognized.

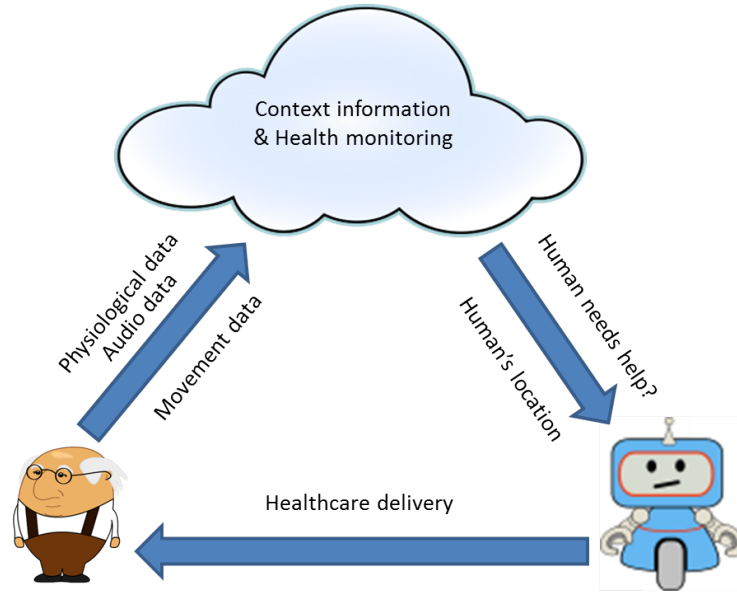


Figure VI.1: Closing the loop with a robot assistant.

## VI.2 Robot Assistant

The robot assistant, as shown in Fig. VI.2, is built on a Pioneer P3-DX base with an approximately 1.5 m-long aluminum frame holding up a touchscreen monitor which is used for video communication and graphic user interface [126]. The robot is equipped with various sensors and devices which include a laser rangefinder (LRF), a vision system, an auditory system, an Intel NUC minicomputer, and batteries. The LRF, a Hokuyo URG-04LX-UG01 [127], is a low-power LRF with a distance range up to 5600 *mm*, an angle range up to 240<sup>0</sup>, and an accuracy of 30 *mm*. The vision system is built using an Asus Xtion Pro Live RGB and Depth (RGB-D) camera [128], a pan-tilt unit, and a control board. The auditory system is built by extending the built-in microphone array of the PS3eye camera [129]. It features four microphones and employs technologies for echo cancellation and background noise suppression. This allows the auditory system to be used for speech recognition, sound localization, and sound separation in noisy environments. The microphone array operates with each channel processing 16-bit samples at a sampling rate up to 48 kHz per channel and a large dynamic range of signal-to-noise ratio up to 90 dB.

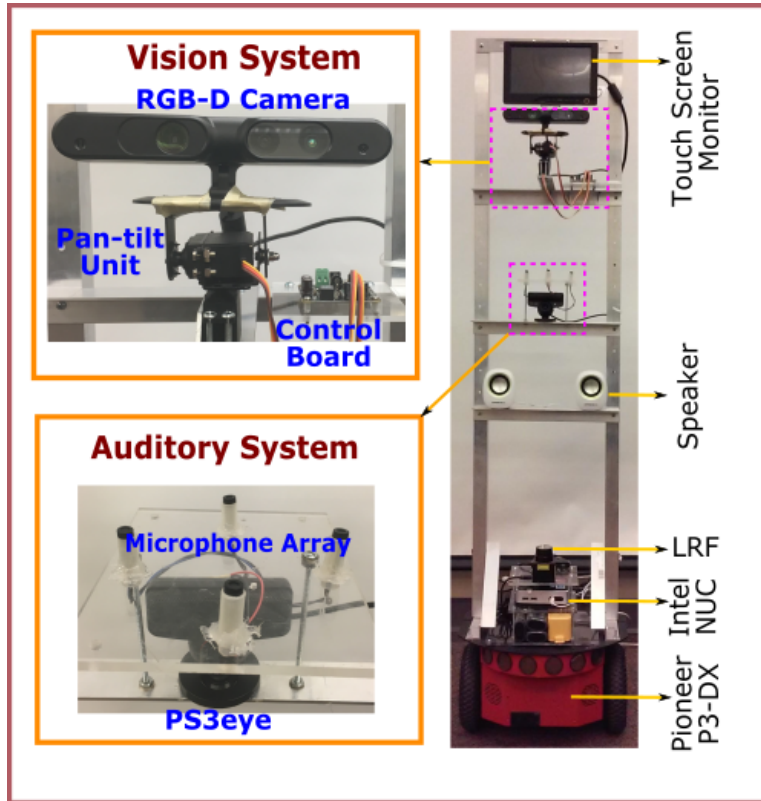


Figure VI.2: Robot assistant.

This robot delivers healthcare services by engaging the human in interactive conversations in order to help the human get out of negative emotions if the negative emotions are detected, and reminding the human of drinking more water if dehydration status is detected. To facilitate such interactive conversations, we utilize Google APIs to provide functions such as converting voice to text, analyzing intents and actions, etc. A general diagram of the interactive conversation method is shown in Fig. VI.3.

- Automatic Speech Recognizer is the component used to convert human voice to a sequence of words. The robot assistant is programmed to utilize the functionalities of Google Cloud Speech API, in which neural network models are used to recognize over 80 languages and variants, with transcripts available immediately after speaking.

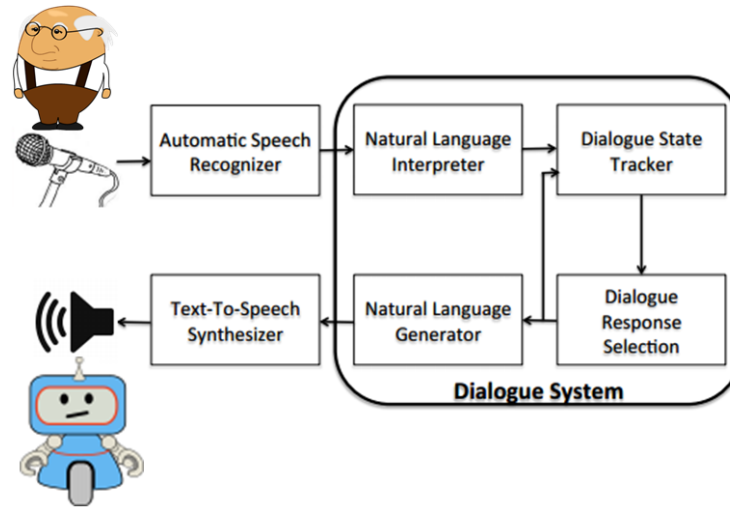


Figure VI.3: Interactive conversation method.

- Natural Language Interpreter is responsible for extracting the structure and meaning of the input text, which is enabled by Google Cloud Natural Language API. An example of text structure analyzing is shown in Fig. VI.4.

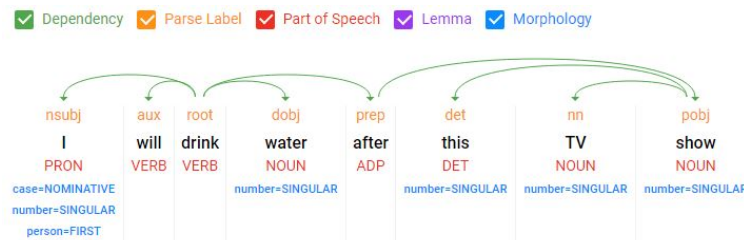


Figure VI.4: Natural language interpreter.

- Dialogue State Tracker and Dialogue Response Selection are supported by the Google API AI. We combine the API AI with the expression evaluation in our emotion regulation system, which is implemented inside the robot assistant. This allows the conversations to be tracked properly.
- Natural Language Generator is used to generate textual summaries from the input data set. In our case, the input data is only from what the human says during the conversation. Other factors, such as human emotion or context information, are not considered. This output response text is generated through



templates, rules-based workflows, and intent-driven approaches.

- Text-to-Speech Synthesizer is used to convert normal language text into speech with proper tones of voice. Google text-to-speech is used in the robot assistant.

## VI.3 Hydration Reminder

### VI.3.1 Hydration Recommendation

Based on [67], the recommended fluid amount is calculated as follow.

$$Fl_{recom} = Fl_{AI} - Fl_{in} - Fl_{food} - Fl_{met} \quad (VI.1)$$

Where  $Fl_{AI}$  and  $Fl_{in}$  are provided by the health monitoring service as mentioned in Section 5.  $Fl_{AI}$  is calculated based on the human's activity level and BMI index such as weight and height.  $Fl_{in}$  is calculated based on drinking sound recognition. Of the total fluid requirement, we account 1000mL intake from solid foods and 350ml from metabolic water per day. The difference between the adjusted fluid requirement and amount of water drunk is recommended to the human every 1 hour to keep a healthy hydration state.

### VI.3.2 Interactive Conversation

Knowing the human's dehydration status, the robot assistant can approach the human location to start a conversation. The conversation, in this case, is to remind the human of drinking more water if a necessary amount of water intake is recommended. When talking with the human, the robot tries to encourage the human to drink at that moment. If the human is reluctant to drink, the robot will persuade the human to drink water.

The content of the conversations is stored in the database, which can be used to improve the conversations in the future. For example, as shown in Fig. VI.6 (b),

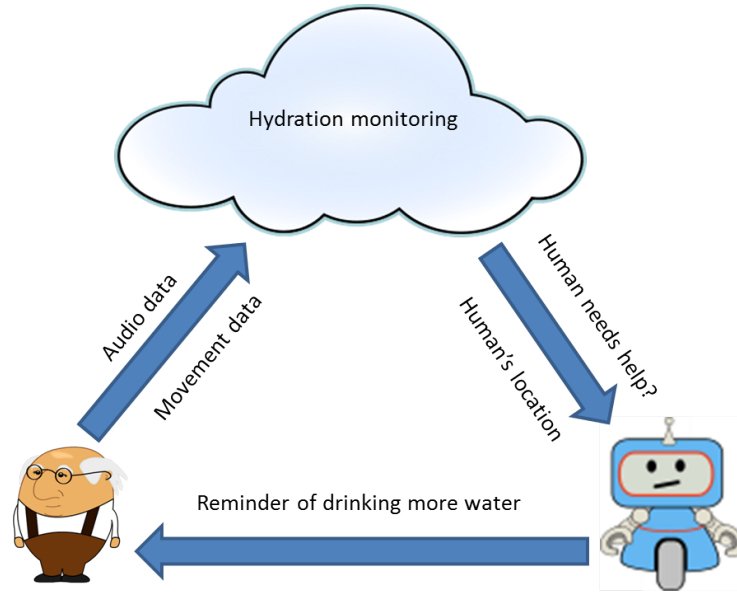
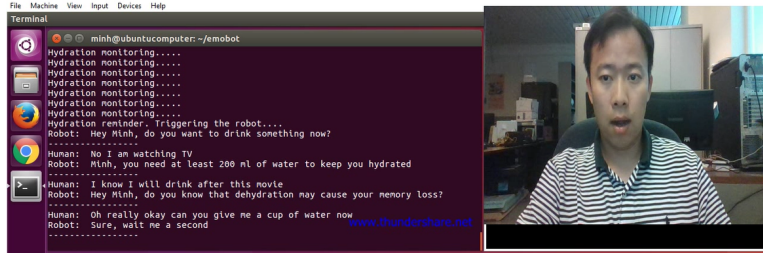


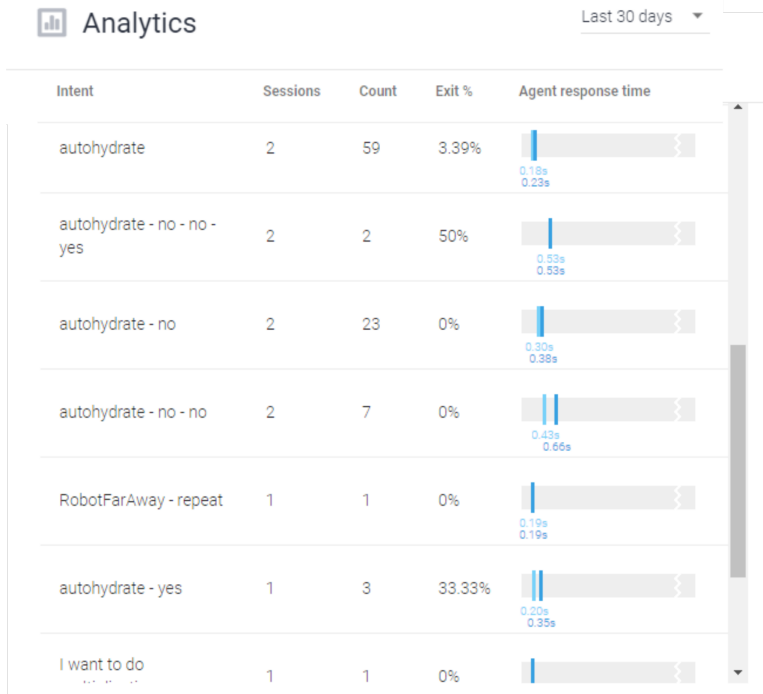
Figure VI.5: The loop of hydration reminder.

based on the history of conversations, we may know that in the last 30 days, there are 3 times in which the human agrees to drink right after the first time of reminder from the robot, and there are 2 times that the human agrees to drink right after the third time of reminder from the robot along with dehydration knowledge provided. A sample conversation between the robot and the human is described as follows.

- *Robot: Hi Minh, Would you like to drink something now?*  
*Human: No I am watching TV*
- *Robot: Minh, you need at least 300 ml of water to keep you hydrated?*  
*Human: I know I will go grab a cup of water after this TV show.*
- *Robot: Hey Minh, do you know that dehydration may cause diabetes?*  
*Human: Really yes I will go drink now thank you.*
- *Robot: You are welcome Minh.*



(a) Interactive conversation for reminding hydration.



(b) Analytics of conversations.

Figure VI.6: Hydration reminder application of the robot assistant.

## VI.4 Emotion Regulation

### VI.4.1 Distraction method

As already mentioned, negative emotional states can be considered risk factors for physical diseases and illness. Therefore, it is desirable to get the human out of negative emotion once it occurs. One of the solutions is to distract them. In [130], the laboratory studies show that directing depressed people to focus on positive distractions can reduce their negative affect. Those studies also show that inducing depressed people to distract from negative thoughts leads to relatively more positive

appraisals of situations, better problem solving, and less distress. According to [131],

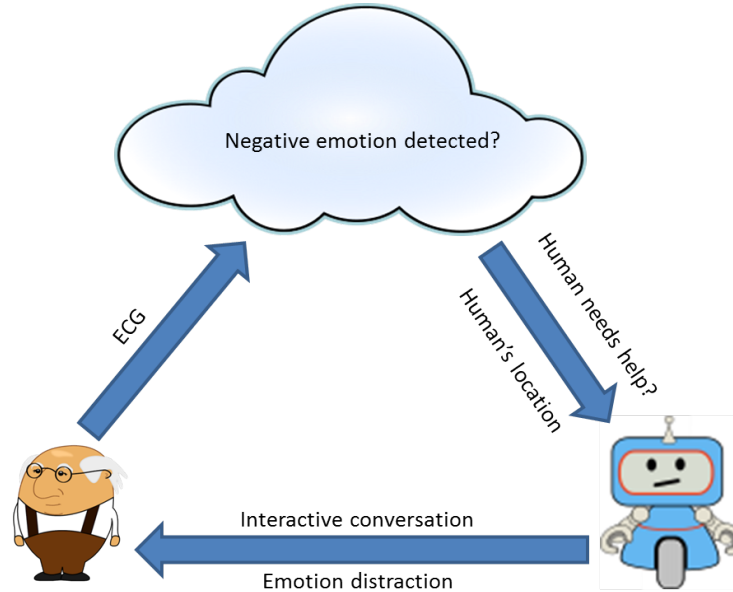
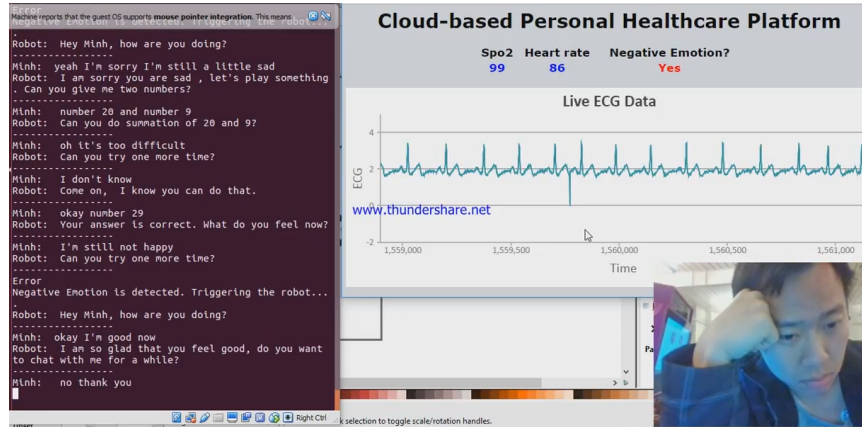


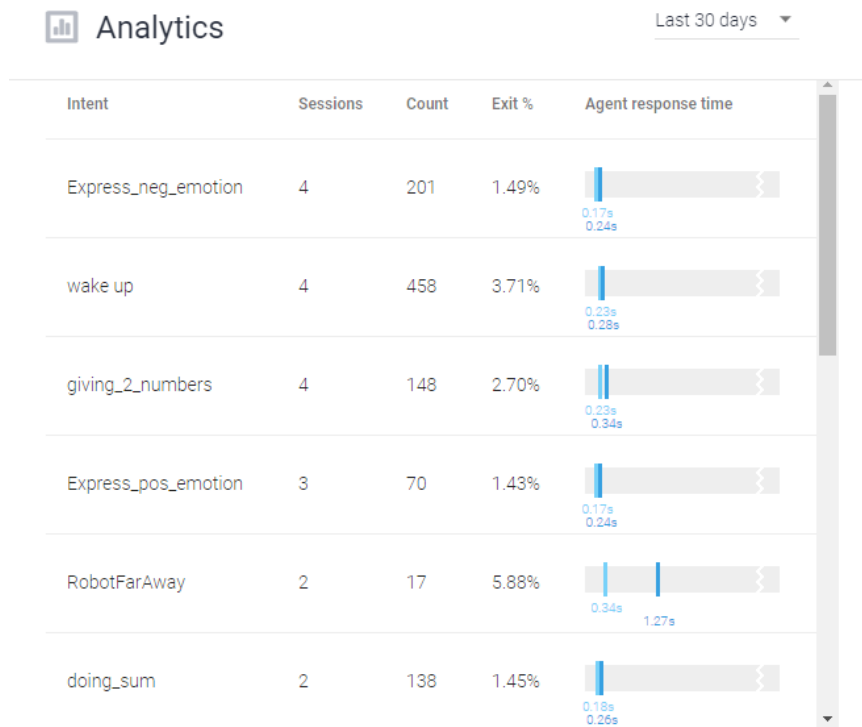
Figure VI.7: Emotion regulation with the robot assistant.

through experiments conducted, the authors claim that people can be distracted from negative emotions by loading their working memory. The more working memory is being used by a distracting activity, the less room will remain for negative emotions to persist. Experimental results from [131] demonstrate that solving a math problem is an effective way to load working memory. The participants in those experiments report that their negative emotion is reduced after solving math problems.

As shown in Fig. VI.7, through the wearable unit, the human's ECG and movement signal are processed and transmitted to the cloud to update negative emotion status and the human latest location. This information is shared with the robot assistant. If the negative emotion is detected, it can approach the human and start talking to distract the human from the negative emotion. In addition to health monitoring, human also receives the healthcare delivery from the robot assistant, which forms a closed-loop healthcare system. Based on the above theory, we implement our robot assistant in such a way that it can talk with the human and encourage them to do some math problems.



(a) Interactive conversation for distraction.



(b) Analytics of conversations that include several intents and their frequency of being invoked and likelihood of ending the conversation.

Figure VI.8: Emotion regulation application of the robot assistant.

## VI.4.2 Interactive Conversation

According to the literature above, one of the best ways to distract humans is to make them actively use their working memory area, for example solving a simple mathematics problem. Therefore, our idea is to design a robot that is able to talk with the human through interactive conversations. We develop several types of conversation with different complexity level to adapt to different intensities of the negative emotions the human may have. The conversations between the human and the robot are stored in the cloud, which may be used for analysis and improving the interactive conversations in the future as shown in VI.8. A sample conversation between the robot and the human is described as follows.

- *Robot: Hi Minh, how are you doing now?*

*Human: I am a little bit sad*

- *Robot: I am sorry that you feel sad. Can you join me playing something? Can you give me two numbers, please?*

*Human: It's OK. Number 7 and number 3*

- *Robot: What is summation of 7 and 3?*

*Human: It is 10.*

- *Robot: Awesome, it's correct. What are you feeling now, Minh?*

*Human: I am still sad*

- *Robot: I am sorry that you are still sad. Can you do multiplication of 7 and 3 and plus 5*

*Human: It is 24, correct?*

- *Robot: No, it's wrong. Can you try one more time?*

*Human: Oh, it is 25.*

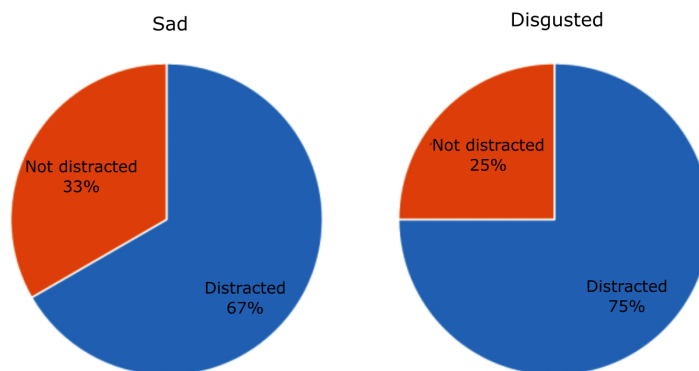


Figure VI.9: Interactive conversation evaluation.

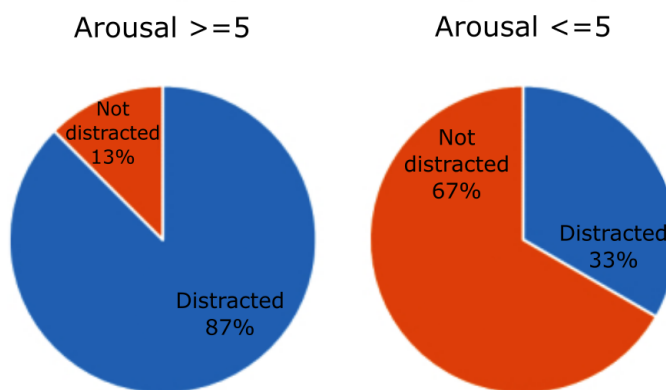
- *Robot: Awesome, it's correct. How do you feel now, Minh?*  
*Human: I feel better now.*
- *Robot: I am glad that you feel better now. Do you want to chat with me for a while?*  
*Human: No, I have to go out.*
- *Robot: Ok, bye for now.*

### VI.4.3 Result

An experiment is conducted with ten subjects to evaluate the interactive conversation. Participants' emotion is induced by letting them watch emotional videos through Oculus or normal screens. Negative emotions such as sadness and disgust are focused in the experiment. During the experiment, ECG signal from the participants is collected using our smart shirt. They are also required to use a wireless keypad to give feedback when they feel a certain emotion as shown in Fig. VI.9. This data may be used for later analysis. After finishing watching and the negative emotions being induced, the individual goes through our interactive conversation system. Finally,



(a) Human with negative emotions being distracted by interactive conversation.



(b) Arousal level being distracted by interactive conversation.

Figure VI.10: Experimental result.

a survey is conducted in which they are asked if they really experience a certain emotion with a level of intensity (from 1 to 10), and if they feel distracted or not after attending the conversation, including following options:

- Not distracted at all
- Distracted (or somewhat distracted)

The survey result shows that out of ten people, six people feel sad and four people feel disgusted with 5.3 and 7.3 as the mean of arousal respectively. Our experiment shows that 90% of people get somewhat distracted or distracted and that people who have higher intensity seem to be more likely to be distracted. The result is shown in Fig. VI.10.



## VI.5 Summary

In this chapter, we introduced a robot assistant which makes the CoSHE a closed-loop healthcare system. Via the cloud system, it is aware of context information and health status of the human. It can perform hydration reminder and emotion regulation via interactive conversations with the human. The conversations are built by utilizing Google services such as speech recognition, dialogue building and text-to-speech. They are designed to help the human keep hydrated if the dehydration is detected, and distract the human if he or she falls in the negative emotion. However, the interactive conversation still needs to improve. One existing problem, which is common for most computer dialogue systems, is that the robot cannot deal with any arbitrary input from human. It can only respond to expected human's inputs, which are included in the robot program. Another problem is that the robot has limited understanding of the human's contextual information, which leads to poor conversations. The human emotion is an important contextual input that may help the robot produce more human-like conversations, which is considered a future work of this research.

## CHAPTER VII

### CONCLUSIONS AND FUTURE WORKS

#### VII.1 Conclusion

This dissertation has presented my work in developing a cloud-based smart home environment supporting health care applications. Below is a summary of the main contributions.

- We have developed smart garments with embedded electrical circuits to collect physiological signals such as electrocardiography (ECG), oxygen saturation (SPo2), and respiration. Textile electrodes and conductive threads are used.
- A cloud-based smart home environment has been developed, which provides a complete ecosystem that supports health care applications. Remote caregivers are able to keep track health status and contextual information of the human living inside.
- We have developed algorithms to recognize human body activity and track of human location using information from an IMU and an array of PIR sensors, which are fused in a particle filter.
- We have developed a health monitoring system that can recognize negative emotions of the human based on the ECG signal. It also can monitor dehydration status and detect negative emotion.
- A healthcare delivery service has been implemented on the robot assistant that can distract the human from negative emotions and remind to drink more water.

- With the robot assistant and smart garment, the CoSHE becomes a closed-loop healthcare system. It will be a potential application if the system is implemented in places where elderly people reside, especially for those who want to live independently or have difficulties in traveling to clinic.

## VII.2 Future Work

Our work can be extended in the following directions in the future.

- Emotion fusion

Human face captured by the robot camera and the ECG acquired by the smart garment are already fused in a decision fusion. We can extend the capability of recognizing negative emotions in even more scenarios if a model using the human voice is included. It is a valuable data source when the ECG quality is bad and the human face is not captured.

What we need to do next is to build a setup to perform new experiments in which we can collect ECG signal, the human face expression and the human voice at the same time when certain emotions are induced. The current setup using Oculus covering a part of the human face will not let the facial expression to be perfectly captured. One option is not using the Oculus to induce emotion while the experiment is conducted. This way we can capture the whole human face, but we need to build a new environment, which should be isolated, to induce emotion by letting human watch videos. The other option is to still keep the Oculus. Using Oculus helps human get better isolated from the ambient environment and is easier to induce emotion, but only the human's mouth area is exposed to be captured, and it may cause difficulty in facial expression detection.

Furthermore, we would like to improve the emotion recognition system so that

it can detect more emotions of the human. Once different types of emotion have been detected, they will be crucial input for the interactive conversation model, thus help the robot appropriately react with more contexts based on human emotions.

- Cognitive assessment

A big problem of older adults when aging is mental disorders such as dementia, cognitive impairment, etc., which cause disability. Developing an automatic cognitive assessment application utilizing our robot assistant's functionalities can be considered in near future. Furthermore, through interactive conversations, not only cognitive assessment can be implemented, but the robot assistant can also provide cognitive training to the older adults to help them enhance their cognitive abilities.

Cognitive assessments are valuable tools in assessing neurological conditions. They are critical in measuring deficits in cognitive function in an array of neurological disorders and during the ageing process. Automation of cognitive assessments is one way to address the increasing burden on medical resources for an ever increasing ageing population.

Walking speed is considered an indicator of mild cognitive impairment [132]. Our home and wearable sensors are able to collect motion data of human in the daily life. The data of each elderly person collected in long-term will establish a large dataset stored in the cloud. The cognitive assessment tool can provide ground truth label for the dataset. This may enable predictive models for early detection of mild cognitive impairment using walking speed.

- Other issues and directions

The smart home and smart garments can be considered IoT (Internet of Things) devices that are connected to the Internet to facilitate the communication with

remote caregivers. However, the daily life activities at home and health status of the older adults are sensitive information and need to be protected from hackers. Privacy and security issues during the process of storing and transmitting human data should be addressed.

Power consumption is a common problem in systems where data is frequently transmitted wirelessly between a network of sensor nodes and a center node where data is processed. A solution for that is to embed more tasks in the sensor nodes where algorithms should be implemented to process data and only send out the inferred output.

Robot emotion model should be developed to allow the robot to express its emotion, for example by showing a smiling face, or a sad face, by changing the tone of voice, or by performing physical movements. The robot emotion model would take all contextual information provided or sensed from both environment and human. Input data may be video signal, audio signal, human actions, human language, human emotion, human health state, etc. Output of the robot emotion model is the robot's reaction.

## BIBLIOGRAPHY

- [1] N. I. on Aging, “Worlds older population grows dramatically,” 2016.
- [2] “U.S. Census Bureau.” <https://www.census.gov>, Oct 2015.
- [3] J. Guzman, A. Pawliczko, S. Beales, C. Till, and I. Voelcker, “Ageing in the twentyfirst century: A celebration and a challenge,” *New York: United Nations Population Fund*, 2012.
- [4] N. Farber, D. Shinkle, J. Lynott, W. Fox-Grage, and R. Harrell, *Aging in place: A state survey of livability policies and practices*. 2011.
- [5] G. Gesano, F. Heins, and A. Naldini, “Regional challenges in the perspective of 2020, regional disparities and future challenges,” *ISMERI Europa*, 2009.
- [6] Wikipedia, “Elderly care,” 2016.
- [7] A. H. Association, “Aha releases 2015 heart and stroke statistics,” 2015.
- [8] S. Gradl, P. Kugler, C. Lohmüller, and B. Eskofier, “Real-time ecg monitoring and arrhythmia detection using android-based mobile devices,” in *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 2452–2455, IEEE, 2012.
- [9] C. D. Galloway, D. E. Albert, and S. B. Freedman, “iphone ecg application for community screening to detect silent atrial fibrillation: a novel technology to prevent stroke,” *Int. J. Cardiol.*, vol. 165, pp. 193–194, 2013.

- [10] F. Agrafioti, D. Hatzinakos, and A. K. Anderson, "Ecg pattern analysis for emotion detection," *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 102–115, 2012.
- [11] H. W. Guo, Y. S. Huang, J. C. Chien, and J. S. Shieh, "Short-term analysis of heart rate variability for emotion recognition via a wearable ecg device," in *2015 International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS)*, pp. 262–265, IEEE, 2015.
- [12] D. Robot, "Bluno - an arduino bluetooth board." [https://www.dfrobot.com/?route=product/product&product\\_id=1044](https://www.dfrobot.com/?route=product/product&product_id=1044). Accessed: 2016-07-01.
- [13] Intel, "Intel edison compute module." <https://software.intel.com/en-us/iot/hardware/edison>. Accessed: 2016-07-01.
- [14] "HealthWatch Technologies, LLC." <https://healthwatchtech.com>, Oct 2018.
- [15] "Carre Technologies inc (Hexoskin)." <https://www.hexoskin.com>, Oct 2018.
- [16] G. Paul, R. Torah, S. Beeby, and J. Tudor, "The development of screen printed conductive networks on textiles for biopotential monitoring applications," *Sensors and Actuators A: Physical*, vol. 206, pp. 35–41, 2014.
- [17] M. Stoppa and A. Chiolerio, "Wearable electronics and smart textiles: a critical review," *sensors*, vol. 14, no. 7, pp. 11957–11992, 2014.
- [18] A. Ankhili, X. Tao, C. Cochrane, D. Coulon, and V. Koncar, "Washable and reliable textile electrodes embedded into underwear fabric for electrocardiography (ecg) monitoring," *Materials*, vol. 11, no. 2, p. 256, 2018.
- [19] G. G. Mazeika and R. Swanson, "Respiratory inductance plethysmography an introduction," *URL: http://www.aastweb.org/resources/focusgroups/rip\_intro.pdf*, 2007.

- [20] H. Sedghamiz, “Complete pan tompkins implementation ecg qrs detector,” *MATLAB Central, Mathworks*, 2014.
- [21] C. English, P. J. Manns, C. Tucak, and J. Bernhardt, “Physical activity and sedentary behaviors in community-dwelling stroke survivors: a systematic review,” *Physical therapy*, 2013.
- [22] N. S. Association *et al.*, “Hope: The stroke recovery guide,” *Retrieved July*, vol. 25, p. 2010, 2007.
- [23] B. Leff, “Hospital at home,” in *Geriatrics Models of Care*, pp. 163–171, Springer, 2015.
- [24] E. Grönvall and N. Verdezoto, “Beyond self-monitoring: Understanding non-functional aspects of home-based healthcare technology,” in *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*, pp. 587–596, ACM, 2013.
- [25] K. Turner, “Accent (advanced component control enhancing network technologies). retrieved august 2013,” 2013.
- [26] A. Holzinger, M. Ziefle, and C. Röcker, *Pervasive Health*. Springer, 2014.
- [27] G. Demiris, B. K. Hensel, M. Skubic, and M. Rantz, “Senior residents perceived need of and preferences for smart home sensor technologies,” *International journal of technology assessment in health care*, vol. 24, no. 01, pp. 120–124, 2008.
- [28] T. L. Chen, C.-H. King, A. L. Thomaz, and C. C. Kemp, “Touched by a robot: An investigation of subjective responses to robot-initiated touch,” in *Proceedings of the 6th international conference on Human-robot interaction*, pp. 457–464, ACM, 2011.



- [29] A. Helal, D. J. Cook, and M. Schmalz, “Smart home-based health platform for behavioral monitoring and alteration of diabetes patients,” *Journal of diabetes science and technology*, vol. 3, no. 1, pp. 141–148, 2009.
- [30] D. J. Cook, A. S. Crandall, B. L. Thomas, and N. C. Krishnan, “Casas: A smart home in a box,” *Computer*, vol. 46, no. 7, 2013.
- [31] P. Rashidi and A. Mihailidis, “A survey on ambient-assisted living tools for older adults,” *Biomedical and Health Informatics, IEEE Journal of*, vol. 17, no. 3, pp. 579–590, 2013.
- [32] J. A. Serrano, H. Van Den Heuvel, P. Björkman, S. Von Rump, and I. Bierhoff, “One year of victoryahome,” in *AAL Forum*, 2014.
- [33] “Giraff Technologies AB.” <http://www.giraff.org/?lang=en>, Sep 2016.
- [34] M. J. Rantz, M. Skubic, R. J. Koopman, L. Phillips, G. L. Alexander, S. J. Miller, and R. D. Guevara, “Using sensor networks to detect urinary tract infections in older adults,” in *e-Health Networking Applications and Services (Healthcom), 2011 13th IEEE International Conference on*, pp. 142–149, IEEE, 2011.
- [35] P. Corporation, “Pir ekmc1601111 data sheet.”
- [36] D. C. Kallur, “Human localization and activity recognition using distributed motion sensors,” 2014.
- [37] O. Terzo and L. Mossucca, *Cloud Computing with E-science Applications*. Crc Press, 2015.
- [38] O. D. Lara and M. A. Labrador, “A survey on human activity recognition using wearable sensors,” *IEEE Communications Surveys & Tutorials*, vol. 15, no. 3, pp. 1192–1209, 2013.

- [39] L. Bao and S. S. Intille, “Activity recognition from user-annotated acceleration data,” in *International Conference on Pervasive Computing*, pp. 1–17, Springer, 2004.
- [40] J. R. Kwapisz, G. M. Weiss, and S. A. Moore, “Activity recognition using cell phone accelerometers,” *ACM SigKDD Explorations Newsletter*, vol. 12, no. 2, pp. 74–82, 2011.
- [41] O. D. Lara, A. J. Pérez, M. A. Labrador, and J. D. Posada, “Centinela: A human activity recognition system based on acceleration and vital sign data,” *Pervasive and mobile computing*, vol. 8, no. 5, pp. 717–729, 2012.
- [42] W. E. Forum, “Global population ageing: Peril or promise,” 2012.
- [43] J. Hightower and G. Borriello, “Location systems for ubiquitous computing,” *Computer*, no. 8, pp. 57–66, 2001.
- [44] R. Want, A. Hopper, V. Falcao, and J. Gibbons, “The active badge location system,” *ACM Transactions on Information Systems (TOIS)*, vol. 10, no. 1, pp. 91–102, 1992.
- [45] A. Ward, A. Jones, and A. Hopper, “A new location technique for the active office,” *Personal Communications, IEEE*, vol. 4, no. 5, pp. 42–47, 1997.
- [46] N. B. Priyantha, A. Chakraborty, and H. Balakrishnan, “The cricket location-support system,” in *Proceedings of the 6th annual international conference on Mobile computing and networking*, pp. 32–43, ACM, 2000.
- [47] Z.-P. Jiang, W. Xi, X. Li, S. Tang, J.-Z. Zhao, J.-S. Han, K. Zhao, Z. Wang, and B. Xiao, “Communicating is crowdsourcing: Wi-fi indoor localization with csi-based speed estimation,” *Journal of Computer Science and Technology*, vol. 29, no. 4, pp. 589–604, 2014.

- [48] J. Krumm, S. Harris, B. Meyers, B. Brumitt, M. Hale, and S. Shafer, "Multi-camera multi-person tracking for easy living," in *Visual Surveillance, 2000. Proceedings. Third IEEE International Workshop on*, pp. 3–10, IEEE, 2000.
- [49] C.-R. Yu, C.-L. Wu, C.-H. Lu, and L.-C. Fu, "Human localization via multi-cameras and floor sensors in smart home," in *Systems, Man and Cybernetics, 2006. SMC'06. IEEE International Conference on*, vol. 5, pp. 3822–3827, IEEE, 2006.
- [50] R.-S. Hsiao, D.-B. Lin, H.-P. Lin, S.-C. Cheng, and C.-H. Chung, "Indoor target detection and localization in pyroelectric infrared sensor networks," in *Proc. the 8th IEEE VTS Asia Pacific Wireless Communications Symposium (APWCS 2011), Singapore*, 2011.
- [51] R. C. Luo and O. Chen, "Wireless and pyroelectric sensory fusion system for indoor human/robot localization and monitoring," *IEEE/ASME Transactions on mechatronics*, vol. 18, no. 3, pp. 845–853, 2013.
- [52] B. Shen and G. Wang, "Distributed target localization and tracking with wireless pyroelectric sensor networks," *International Journal on Smart Sensing and Intelligent Systems*, vol. 6, no. 4, pp. 1400–1418, 2013.
- [53] B. Yang, J. Luo, and Q. Liu, "A novel low-cost and small-size human tracking system with pyroelectric infrared sensor mesh network," *Infrared Physics & Technology*, vol. 63, pp. 147–156, 2014.
- [54] I. Al-Naimi, C. B. Wong, P. Moore, and X. Chen, "Advanced approach for indoor identification and tracking using smart floor and pyroelectric infrared sensors," in *Information and Communication Systems (ICICS), 2014 5th International Conference on*, pp. 1–6, IEEE, 2014.

- [55] M. A. Bitew, R.-S. Hsiao, H.-P. Lin, and D.-B. Lin, “Hybrid indoor human localization system for addressing the issue of rss variation in fingerprinting,” *International Journal of Distributed Sensor Networks*, vol. 2015, p. 3, 2015.
- [56] O. Mezentsev, J. Collin, and G. Lachapelle, “Pedestrian dead reckoning—a solution to navigation in gps signal degraded areas?,” *Geomatica*, vol. 59, no. 2, pp. 175–182, 2005.
- [57] L. Ojeda and J. Borenstein, “Non-gps navigation with the personal dead-reckoning system,” in *Defense and Security Symposium*, pp. 65610C–65610C, International Society for Optics and Photonics, 2007.
- [58] E. Foxlin, “Pedestrian tracking with shoe-mounted inertial sensors,” *IEEE Computer graphics and applications*, vol. 25, no. 6, pp. 38–46, 2005.
- [59] R. Stirling, J. Collin, K. Fyfe, and G. Lachapelle, “An innovative shoe-mounted pedestrian navigation system,” in *proceedings of European navigation conference GNSS*, pp. 110–5, 2003.
- [60] R. Jirawimut, P. Ptasinski, V. Garaj, F. Cecelja, and W. Balachandran, “A method for dead reckoning parameter correction in pedestrian navigation system,” *Instrumentation and Measurement, IEEE Transactions on*, vol. 52, no. 1, pp. 209–215, 2003.
- [61] A. D. Cheek and Y. Li, “Ubiquitous interaction with positioning and navigation using a novel light sensor-based information transmission system,” *Personal and Ubiquitous Computing*, vol. 12, no. 6, pp. 445–458, 2008.
- [62] F. Evennou and F. Marx, “Advanced integration of wifi and inertial navigation systems for indoor mobile positioning,” *Eurasip journal on applied signal processing*, vol. 2006, pp. 164–164, 2006.

- [63] G. Retscher, “Test and integration of location sensors for a multi-sensor personal navigator,” *Journal of Navigation*, vol. 60, no. 01, pp. 107–117, 2007.
- [64] D. Roetenberg, *Inertial and magnetic sensing of human motion*. University of Twente, 2006.
- [65] S. Oh, L. Schenato, P. Chen, and S. Sastry, “Tracking and coordination of multiple agents using sensor networks: system design, algorithms and experiments,” *Proceedings of the IEEE*, vol. 95, no. 1, pp. 234–254, 2007.
- [66] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, “A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking,” *IEEE Transactions on signal processing*, vol. 50, no. 2, pp. 174–188, 2002.
- [67] K. L. Marcia Nelms, Kathryn P. Sucher and S. L. Roth, *Nutrition Therapy and Pathophysiology 2e*. Cengage Learning, 2010.
- [68] B. M. Popkin, K. E. D’Anci, and I. H. Rosenberg, “Water, hydration, and health,” *Nutrition reviews*, vol. 68, no. 8, pp. 439–458, 2010.
- [69] K. B. Watson, S. A. Carlson, D. D. Carroll, and J. E. Fulton, “Comparison of accelerometer cut points to estimate physical activity in us adults,” *Journal of sports sciences*, vol. 32, no. 7, pp. 660–669, 2014.
- [70] K. Yatani and K. N. Truong, “Bodyscope: a wearable acoustic sensor for activity recognition,” in *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, pp. 341–350, ACM, 2012.
- [71] Y. Bi, M. Lv, C. Song, W. Xu, N. Guan, and W. Yi, “Autodietary: A wearable acoustic sensor system for food intake recognition in daily life,” *IEEE Sensors Journal*, vol. 16, no. 3, pp. 806–816, 2016.

- [72] T. Olubanjo and M. Ghovanloo, “Tracheal activity recognition based on acoustic signals,” in *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 1436–1439, IEEE, 2014.
- [73] Y. Bi, W. Xu, N. Guan, Y. Wei, and W. Yi, “Pervasive eating habits monitoring and recognition through a wearable acoustic sensor,” in *Proceedings of the 8th International Conference on Pervasive Computing Technologies for Healthcare*, pp. 174–177, ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2014.
- [74] M. Sawka, “Dietary reference intakes for water, potassium, sodium, chloride, and sulfate,” 2005.
- [75] G. Guo and S. Z. Li, “Content-based audio classification and retrieval by support vector machines,” *IEEE transactions on Neural Networks*, vol. 14, no. 1, pp. 209–215, 2003.
- [76] H. Lu, W. Pan, N. D. Lane, T. Choudhury, and A. T. 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*, pp. 165–178, ACM, 2009.
- [77] S. Chernbumroong, A. S. Atkins, and H. Yu, “Activity classification using a single wrist-worn accelerometer,” in *Software, Knowledge Information, Industrial Management and Applications (SKIMA), 2011 5th International Conference on*, pp. 1–6, IEEE, 2011.
- [78] P. Salovey, A. J. Rothman, J. B. Detweiler, and W. T. Steward, “Emotional states and physical health.,” *American psychologist*, vol. 55, no. 1, p. 110, 2000.
- [79] S. Cohen, W. J. Doyle, D. P. Skoner, P. Fireman, J. M. Gwaltney Jr, and J. T. Newsom, “State and trait negative affect as predictors of objective and

- subjective symptoms of respiratory viral infections.,” *Journal of personality and Social Psychology*, vol. 68, no. 1, p. 159, 1995.
- [80] S. Booth-Kewley and H. S. Friedman, “Psychological predictors of heart disease: a quantitative review.,” *Psychological bulletin*, vol. 101, no. 3, p. 343, 1987.
- [81] T. W. Smith, K. Glazer, J. M. Ruiz, and L. C. Gallo, “Hostility, anger, aggressiveness, and coronary heart disease: An interpersonal perspective on personality, emotion, and health,” *Journal of personality*, vol. 72, no. 6, pp. 1217–1270, 2004.
- [82] L. Li and J.-h. Chen, “Emotion recognition using physiological signals,” in *Advances in Artificial Reality and Tele-Existence*, pp. 437–446, Springer, 2006.
- [83] F. Nasoz, K. Alvarez, C. L. Lisetti, and N. Finkelstein, “Emotion recognition from physiological signals using wireless sensors for presence technologies,” *Cognition, Technology & Work*, vol. 6, no. 1, pp. 4–14, 2004.
- [84] A. Greco, A. Lanata, G. Valenza, G. Rota, N. Vanello, and E. Scilingo, “On the deconvolution analysis of electrodermal activity in bipolar patients,” in *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 6691–6694, IEEE, 2012.
- [85] J. Kim and E. André, “Emotion recognition based on physiological changes in music listening,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 30, no. 12, pp. 2067–2083, 2008.
- [86] J. Scheirer, R. Fernandez, J. Klein, and R. W. Picard, “Frustrating the user on purpose: a step toward building an affective computer,” *Interacting with computers*, vol. 14, no. 2, pp. 93–118, 2002.

- [87] P. J. Lang, “The emotion probe: Studies of motivation and attention.,” *American psychologist*, vol. 50, no. 5, p. 372, 1995.
- [88] J. Posner, J. A. Russell, and B. S. Peterson, “The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology,” *Development and psychopathology*, vol. 17, no. 3, pp. 715–734, 2005.
- [89] R. W. Levenson, “Emotion and the autonomic nervous system: A prospectus for research on autonomic specificity.,” 1988.
- [90] M. Nardelli, G. Valenza, A. Greco, A. Lanata, and E. P. Scilingo, “Recognizing emotions induced by affective sounds through heart rate variability,” *IEEE Transactions on Affective Computing*, vol. 6, no. 4, pp. 385–394, 2015.
- [91] G. Valenza, L. Citi, A. Lanatá, E. P. Scilingo, and R. Barbieri, “Revealing real-time emotional responses: a personalized assessment based on heartbeat dynamics,” *Scientific reports*, vol. 4, 2014.
- [92] Y.-P. Lin, C.-H. Wang, T.-P. Jung, T.-L. Wu, S.-K. Jeng, J.-R. Duann, and J.-H. Chen, “Eeg-based emotion recognition in music listening,” *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 7, pp. 1798–1806, 2010.
- [93] D. Nie, X.-W. Wang, L.-C. Shi, and B.-L. Lu, “Eeg-based emotion recognition during watching movies,” in *Neural Engineering (NER), 2011 5th International IEEE/EMBS Conference on*, pp. 667–670, IEEE, 2011.
- [94] Y. Liu, O. Sourina, and M. K. Nguyen, “Real-time eeg-based human emotion recognition and visualization,” in *Cyberworlds (CW), 2010 International Conference on*, pp. 262–269, IEEE, 2010.



- [95] B. Cheng and G.-Y. Liu, “Emotion recognition from surface emg signal using wavelet transform and neural network,” in *Proceedings of the 2nd international conference on bioinformatics and biomedical engineering (ICBBE)*, pp. 1363–1366, 2008.
- [96] A. Nakasone, H. Prendinger, and M. Ishizuka, “Emotion recognition from electromyography and skin conductance,” in *Proc. of the 5th International Workshop on Biosignal Interpretation*, pp. 219–222, Citeseer, 2005.
- [97] L. M. Williams, M. L. Phillips, M. J. Brammer, D. Skerrett, J. Lagopoulos, C. Rennie, H. Bahramali, G. Olivieri, A. S. David, A. Peduto, *et al.*, “Arousal dissociates amygdala and hippocampal fear responses: evidence from simultaneous fmri and skin conductance recording,” *Neuroimage*, vol. 14, no. 5, pp. 1070–1079, 2001.
- [98] F. Ringeval, A. Sonderegger, J. Sauer, and D. Lalanne, “Introducing the recola multimodal corpus of remote collaborative and affective interactions,” in *Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on*, pp. 1–8, IEEE, 2013.
- [99] M. K. Abadi, R. Subramanian, S. M. Kia, P. Avesani, I. Patras, and N. Sebe, “Decaf: Meg-based multimodal database for decoding affective physiological responses,” *IEEE Transactions on Affective Computing*, vol. 6, no. 3, pp. 209–222, 2015.
- [100] P. C. Petrantonakis and L. J. Hadjileontiadis, “Emotion recognition from brain signals using hybrid adaptive filtering and higher order crossings analysis,” *IEEE Transactions on affective computing*, vol. 1, no. 2, pp. 81–97, 2010.

- [101] S. Nasehi, H. Pourghassem, and I. Isfahan, “An optimal eeg-based emotion recognition algorithm using gabor,” *WSEAS Transactions on Signal Processing*, vol. 3, no. 8, pp. 87–99, 2012.
- [102] M. Murugappan, K. Wan, S. Yaacob, *et al.*, “Electrocardiogram-based emotion recognition system using empirical mode decomposition and discrete fourier transform,” *Expert Systems*, vol. 31, no. 2, pp. 110–120, 2014.
- [103] J. Selvaraj, M. Murugappan, K. Wan, and S. Yaacob, “Classification of emotional states from electrocardiogram signals: a non-linear approach based on hurst,” *Biomedical engineering online*, vol. 12, no. 1, p. 1, 2013.
- [104] A. Goshvarpour, A. Abbasi, and A. Goshvarpour, “Do men and women have different ecg responses to sad pictures?,” *Biomedical Signal Processing and Control*, vol. 38, pp. 67–73, 2017.
- [105] J.-P. Eckmann, S. O. Kamphorst, and D. Ruelle, “Recurrence plots of dynamical systems,” *EPL (Europhysics Letters)*, vol. 4, no. 9, p. 973, 1987.
- [106] A. M. Fraser and H. L. Swinney, “Independent coordinates for strange attractors from mutual information,” *Physical review A*, vol. 33, no. 2, p. 1134, 1986.
- [107] M. B. Kennel, R. Brown, and H. D. Abarbanel, “Determining embedding dimension for phase-space reconstruction using a geometrical construction,” *Physical review A*, vol. 45, no. 6, p. 3403, 1992.
- [108] N. Marwan, N. Wessel, U. Meyerfeldt, A. Schirdewan, and J. Kurths, “Recurrence-plot-based measures of complexity and their application to heart-rate-variability data,” *Physical review E*, vol. 66, no. 2, p. 026702, 2002.
- [109] P. Werner, A. Al-Hamadi, R. Niese, S. Walter, S. Gruss, and H. C. Traue, “Towards pain monitoring: Facial expression, head pose, a new database, an

- automatic system and remaining challenges,” in *Proceedings of the British Machine Vision Conference*, pp. 119–1, 2013.
- [110] P. Tarnowski, M. Kołodziej, A. Majkowski, and R. J. Rak, “Emotion recognition using facial expressions,” *Procedia Computer Science*, vol. 108, pp. 1175–1184, 2017.
- [111] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in *Advances in neural information processing systems*, pp. 1097–1105, 2012.
- [112] S. Lawrence, C. L. Giles, A. C. Tsoi, and A. D. Back, “Face recognition: A convolutional neural-network approach,” *IEEE transactions on neural networks*, vol. 8, no. 1, pp. 98–113, 1997.
- [113] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, *et al.*, “Imagenet large scale visual recognition challenge,” *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, 2015.
- [114] A. Canziani, A. Paszke, and E. Culurciello, “An analysis of deep neural network models for practical applications,” *arXiv preprint arXiv:1605.07678*, 2016.
- [115] A. Mollahosseini, B. Hasani, and M. H. Mahoor, “Affectnet: A database for facial expression, valence, and arousal computing in the wild,” *arXiv preprint arXiv:1708.03985*, 2017.
- [116] N. Aifanti, C. Papachristou, and A. Delopoulos, “The mug facial expression database,” in *Image analysis for multimedia interactive services (WIAMIS), 2010 11th international workshop on*, pp. 1–4, IEEE, 2010.

- [117] M. J. Lyons, S. Akamatsu, M. Kamachi, J. Gyoba, and J. Budynek, “The japanese female facial expression (jaffe) database,” in *Proceedings of third international conference on automatic face and gesture recognition*, pp. 14–16, 1998.
- [118] S. Happy, P. Patnaik, A. Routray, and R. Guha, “The indian spontaneous expression database for emotion recognition,” *IEEE Transactions on Affective Computing*, vol. 8, no. 1, pp. 131–142, 2017.
- [119] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, “The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression,” in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*, pp. 94–101, IEEE, 2010.
- [120] O. Langner, R. Dotsch, G. Bijlstra, D. H. Wigboldus, S. T. Hawk, and A. Van Knippenberg, “Presentation and validation of the radboud faces database,” *Cognition and emotion*, vol. 24, no. 8, pp. 1377–1388, 2010.
- [121] A. Del Sole, “Introducing microsoft cognitive services,” in *Microsoft Computer Vision APIs Distilled*, pp. 1–4, Springer, 2018.
- [122] P. K. Varshney, *Distributed detection and data fusion*. Springer Science & Business Media, 2012.
- [123] M. Patacchiola and A. Cangelosi, “Head pose estimation in the wild using convolutional neural networks and adaptive gradient methods,” *Pattern Recognition*, vol. 71, pp. 132–143, 2017.
- [124] R. L. Gellish, B. R. Goslin, R. E. Olson, A. McDONALD, G. D. Russi, and V. K. Moudgil, “Longitudinal modeling of the relationship between age and

maximal heart rate.,” *Medicine and science in sports and exercise*, vol. 39, no. 5, pp. 822–829, 2007.

- [125] M. Clinic, “Exercise intensity: How to measure it.” <https://www.mayoclinic.org/healthy-lifestyle/fitness/in-depth/exercise-intensity/art-20046887>. Accessed: 2018-07-01.
- [126] H. M. Do, W. Sheng, M. Liu, and S. Zhang, “Context-aware sound event recognition for home service robots,” in *Automation Science and Engineering (CASE), 2016 IEEE International Conference on*, pp. 739–744, IEEE, 2016.
- [127] “Hokuyo laser.”
- [128] ASUS, “Xtion pro live.”
- [129] Sony, “Ps3 eye camera.”
- [130] S. Nolen-Hoeksema, B. E. Wisco, and S. Lyubomirsky, “Rethinking rumination,” *Perspectives on psychological science*, vol. 3, no. 5, pp. 400–424, 2008.
- [131] L. F. Van Dillen and S. L. Koole, “Clearing the mind: a working memory model of distraction from negative mood.,” *Emotion*, vol. 7, no. 4, p. 715, 2007.
- [132] T. Buracchio, H. H. Dodge, D. Howieson, D. Wasserman, and J. Kaye, “The trajectory of gait speed preceding mild cognitive impairment,” *Archives of neurology*, vol. 67, no. 8, pp. 980–986, 2010.

VITA

Minh Pham

Candidate for the Degree of

Doctor of Philosophy

Dissertation: HOME HEALTHCARE USING UBIQUITOUS COMPUTING AND  
ROBOT TECHNOLOGIES

Major Field: Electrical Engineering

Biographical:

Education:

Completed the requirements for the degree of Doctor of Philosophy in  
Electrical Engineering at Oklahoma State University in December, 2018

Received the M.S. degree in Management Information System from Ok-  
lahoma State University, Stillwater, Oklahoma, 2012

Received the B.S. degree in Computer Science from Hanoi University of  
Science and Technology, Hanoi, Vietnam, 2007