Foot-based static gesture interaction in water

By

Gunawardena Kennantudawage Sanjeewa Lasith

A dissertation submitted in partial fulfillment
of the requirements for the
degree of Doctor of Engineering

in the

Interdisciplinary Graduate School of Science and
Engineering
of
Shimane University
Japan

January 2016

Declaration

I hereby certify that this dissertation entitled **“**Foot-based static gesture interaction in water**”** is entirely my own work. Wherever other sources of information have been used, they have been acknowledged

This dissertation has not been accepted for any degree and is not being submitted for any other degree.

Gunawardena Kennantudawage Sanjeewa Lasith	Signature	:
(S 119804)	Date	:

Acknowledgements

This dissertation is the outcome of a study carried out over several years in the selected domain of Human Computer Interaction. I record my gratitude to all the personalities who assisted me in numerous ways to accomplish this research in Shimane University.

My research would not be in the present qualified state if not for the direct supervision of my supervisor, Professor Masahito Hirakawa. Sensei, with his proven track of guiding research, ensured completion of my research project with improvements necessary for a doctoral research study. He advised how to handle challenges with confidence. Sensei's attitude and advice to foreign students specially will receive my highest regard. He also ensured our social life to progress well in Matsue city.

Incidentally I was motivated to carry out higher studies in Japan after experiencing the traits of few Sri Lankan Academics who followed their higher studies in Japan. I specially extend my gratitude to Prof. K.P. Hewagamage, for guiding me through my Masters research at University of Colombo, Sri Lanka, and further for connecting me up with Prof. Masahito Hirakawa.

I also wish to thank academic staff members of the Department of Mathematics and Computer Science for their contribution to my studies at Shimane University. My colleagues at Hirakawa Lab whom I have worked since 2010 should receive a special comment. I appreciate their collaboration, friendship and assistance extended to me for lab-related research work and activities. My sincere thanks are registered for my friends in Shimane University, Associations and friends in Matsue who assisted us always, and for their indirectly contribution to my research work to progress smoothly. Although I regret not naming them individually, I value their helpful outlook.

Currently, I am on leave from my workplace, University of Sri Jayewardenepura, Sri Lanka to pursue my doctoral studies. I wish to acknowledge and thank the assistance provided by the Vice Chancellor, Dean of the Faculty of Management Studies & Commerce, and Head, Department of IT in granting me leave, as well as my departmental colleagues for covering up my duties during my absence.

This study would have not materialized if not for the MEXT scholarship I received. I honor the Ministry of Education, Culture, Sports, Science, and Technology, Government of Japan for granting me the MEXT Scholarship 2010. I pay special gratitude to the University Grants Commission, Ministry of Higher Education, Government of Sri Lanka, and Embassy of Japan in Sri Lanka for selecting me for this scholarship.

Finally, I record my deepest gratitude to my family for the support they provided at all times: My loving wife Prabhani, who has been encouraging and supporting me in every possible way. Taking care of our daughter Senuli, and pursuing her research as well was truly a sacrifice. I owe a deep debt of gratitude to my beloved parents, whose unwavering love and care prompted me to continue studies on this research. Their wishes are to progress education as a global social responsibility. Especially this research is dedicated to them for the encouragement, love and binding guidance.

Abstract

This dissertation investigates how human foot gestures performed in a water medium can be detected using a computer. The feet play a significant role in the human gait cycle. Foot operated pedals are ubiquitous in vehicles and found in musical devices. As an interaction body part, it has received minimal attention in the study of gesture recognition. This is probably because foot movements are considered less accurate when compared to the hand.

In the sphere of computing, gesture recognition has recently been a topic of interest to many researchers, with many detection devices in the market. In the absolute majority of research and devices, the interaction takes place within the air medium. Further, there has been a recent trend to investigate interaction in public locations and natural mediums. Water is one of the closely associated mediums with day to day human life. The sensation provided by water has a relaxing effect on the human body. Consequently, it is opportune to investigate foot gestures in a natural liquid such as water. The scope of the research limits to detecting static gestures, where the gesture performed at a single instance of time is analyzed. Static gestures can also be referred to as poses when considering the human body.

The research initially attempted to review the literature and existing technologies to get a firm understanding of existing techniques and technologies for object detection, and to study the practicality with relation to water. One observation from existing research was that most gesture movements focused on the hand, finger, and full body movement and less significance was on the feet. Another was that only a few researchers have investigated immersive interaction within a three-dimensional volume of water. Further study of previous research examined the software process for gesture detection, which included the protocols, toolkits, and machine learning algorithms.

On the completion of the literature survey, experiments with existing gesture recognition devices revealed their detection abilities deteriorated during water interaction. This deterioration was due to the low-intensity energy used, as well as using dispersive beams. Using vision-based configurations contribute to increasing in system setup space. Experiments which used laser technology in water were promising and required less space. Similarly, phototransistors suited the requirement

to detect laser beams. An acrylic tank mounted on top of a display was used as the primary interaction space that contained water. 78 laser-phototransistor pairs are used and connected to an Arduino Uno Microcontroller via a multiplexer. The temperature of the tank is read using a thermistor. A heating element is connected to increase the temperature of the tank when necessary. A water faucet connects to the tank using a water pump for demonstration purposes. Using a touch frame above the water surface improves object detection. The hardware framework developed as part of this research is called SensorTank and is one of a kind in its detection approach. It constitutes a significant contribution of this research. Its functionality and robustness were tested under different test conditions including murky water, temperature, light level, as well as examining the effect of ripples and air bubbles. This framework is a principal contribution of this research. The literature survey did not find a device that was capable of detecting locations of immersed object in limited spaces such as a foot interaction tank.

The software framework was developed using Processing language. This program interfaces with applications in MaxMSP, Adobe Flash, to provide audio and visual output. Open Sound Control and Tangible User Interface Objects protocols are used to communicate with applications and hardware. Point cloud data about the object is filtered to detect foot sized objects using the Connected Component Labelling Algorithm. An experiment conducted revealed seven gestures that are suitable for feet movement in the water. Finally, a gesture recognition analysis was performed using 11,036 samples to evaluate the machine learning algorithm that recognized the gestures with a high level of precision. Overall recognition rates over 90% were achieved, with the best recognition suitable for real-time usage provided by the Adaboost algorithm at 96.64%. The gesture analysis uses the open sourced Gesture Recognition Toolkit. Identification of the gestures is significant as it permits application developers the opportunity to develop applications based on the identified gestures.

The prototype applications built as part of this research focus on foot bath (Ashiyu) environments, home relaxation, and automation. This research study contributes not only to the sphere of human-computer interaction; its applications extend to rehabilitative medicine as well as entertainment in public spaces.

List of Figures

List of Tables

Abbreviations

1. CLI – Command Line Interface
2. GUI – Graphical User Interface
3. 3D – 3 Dimension
4. HCI – Human Computer Interaction
5. 2D – 2 Dimension
6. GIS – Geographical Information Systems
7. TUI – Tangible User Interface
8. LED – Light Emitting Diode
9. RFID – Radio Frequency Identification
10. RGB – Red Green Blue
11. CPU- Central Processing Unit
12. IR – Infra Red
13. FTIR - Frustrated Total Internal Reflection
14. TIR – Total Internal Reflection
15. EMG – Electromyography
16. IMU - Inertial Measurement Unit
17. USB – Universal Serial Bus
18. SDK – System Development Kit
19. OSC – Open Sound Control
20. MIDI – Musical Instrument Digital Interface
21. TCP – Transmission Control Protocol
22. UDP – User Datagram Protocol
23. TUIO – Tangible User Interface Objects
24. SVM – Support Vector Machines
25. WEKA – Waikato Environment for Knowledge Analysis
26. MOOC – Massive Open Online Course
27. GRT – Gesture Recognition Toolkit
28. ROC – Receiver Operating Characteristic
29. ANBC – Adaptive Naïve Bayes Classifier
30. k-NN – K-Nearest Neighbor
31. DT – Decision Tree

32. DTW – Dynamic Time Warping
33. PCB – Printed Circuit Board
34. DC – Direct Current
35. IC – Integrated Circuit
36. LCD – Liquid Crystal Display
37. HDMI – High Definition Multimedia Interface
38. API – Application Program Interface
39. MIT – Massachusetts Institute of Technology
40. bps – Bits per Second
41. NTU – Nephelometric Turbidity Units
42. ID - Identification

Table of Contents

Chapter 1

Introduction

1.1 Background of the research

The introduction of the computer as a data processing device from the middle of the last century has led to numerous productivity enhancements for humankind. From the initial days up until the 1980's, the keyboard was considered as the primary input device used with the Command Line Interface (CLI). The mouse and Graphical User Interface (GUI) changed this scenario and are a mainstay in desktop computing together with the keyboard even today. Technological developments and human perception on computing have changed the user interface further. As the costs of ownership in computing devices have fallen, the device size has also shrunken and permits us to own and use more than one computing device. As a result, it is found that the mobile phone in use today is effectively is a multifunction appliance. It has also given rise to the concept of mobile computing.

With smaller mobile devices, novel interaction techniques are required to replace the keyboard and mouse. The Apple iPhone heralded the touch revolution that was further expanded by the introduction of the iPad multi-touch sensitive tablet. Thus, the Touch interface was born, resulting in hard buttons disappearing and replaced with larger screens.

Multi-touch interfaces, interactive surfaces, and miniature computing platforms have enabled the path to Weiser's vision of ubiquitous computing (Mark Weiser, 1991) to be realized. As devices miniaturization accelerates, the Internet of things (Ashton, 2009) becomes a reality with day to day objects connected with each other on a continuous basis. The trends above have influenced to change the perception on the role played by computing devices. Computing devices are no longer seen simply as data processing devices, but as gateways to such activities as knowledge acquisition and entertainment. Further, as computers become ubiquitous, they have also empowered social interaction and collaboration. Novel interaction mechanisms have debuted in public spaces (Häkkilä, Koskenranta, Posti, & He, 2014;

Ventä-Olkkonen, Akerman, Puikkonen, Colley, & Häkkilä, 2014; Virolainen, Puikkonen, Kärkkäinen, & Häkkilä, 2010).

While the display is still seen by many as the central interaction space, miniaturization has driven it to a minimum point bordering the invisible. As envisaged by Weiser, "a good computing tool is one which is invisible and permits one to focus on the task at hand, while not intruding on the consciousness of the user" (Marc Weiser, 1994). The recent introduction of smart watches, devices such as Google Glass and other wearable computing devices demands natural interaction mechanisms such as speech and gestures.

Out of the sensing performed by humans, vision is a key sensory facility that permits us to gather the majority of information from the external environment. While formal human communication takes place mainly in spoken and written languages, gestures constitute a more subtle form of interpersonal communication.

A "gesture" can be defined as a movement or position of a part of the body to communicate or express an idea or meaning. Movement of the body part with respect to time results in dynamic gestures, while positioning the body part for a single instance in time results in static gestures. Although most gestures are voluntary, some gestures such as throwing up one's arms in joy or kicking the ground in anger can be involuntary. It is more natural for humans to use gestures while performing day to day tasks as well as communication. Therefore, interaction with computing devices by way of gesturing can be more intuitive over the use of keyboard and mouse.

However, recognizing human gestures is a potential barrier to this interaction, and requires technology solutions. The barrier has led to researchers attempting to develop techniques for detecting human gestures. The initial attempts used cameras and image processing techniques. The past decade has seen an increased focus on detecting gestures using non-vision-based sensors or hybrid devices using a combination of cameras and other sensors. A few examples of such devices are Microsoft Kinect[1], Softkinetic's DepthSense[2], Leap Motion's Leap Motion Controller[3], Ring from Logbar Inc. and Myo of Thalmic Labs[4]. To enable a computer to recognize gestures, the preferred mode of operation is to apply techniques referred

[1] https://www.microsoft.com/en-us/kinectforwindows
[2] http://www.softkinetic.com/Products/DepthSenseCameras
[3] https://www.leapmotion.com/product
[4] https://www.thalmic.com/en/myo

to as machine learning on data produced by cameras and sensors. Machine learning has been successfully used for different complicated pattern recognition problems. These problems include, but not limited to speech recognition, handwriting recognition, fingerprint recognition, and face recognition.

Despite the decreasing cost of sensor technology and rapid advances in computing technology, most of the existing devices are targeted at detecting either finger motion, hand/arm movements, head movements or body movements, while only a few devices have been designed for use with foot-based interactions. In fact, this is evidenced by the small number of publications dealing with foot-based interactions. Further proof is on this is found in Karam & Schrafel's taxonomy of gestures in HCI, which considers 128 papers from the past 40 years (1965-2005). In their summarization of gestures according to a body part, the feet are not even classified as a unique category but is probably classified under "body part" category where fewer than ten papers have been published (Karam & Schraefel, 2005). Researchers have attributed this to the fact that foot movements are less accurate, require more execution time and are probably less satisfying than hand movements for the same task (Pakkanen & Raisamo, 2004; Pearson & Weiser, 1986). Others have proposed that it is possible to use foot interaction in non-accurate spatial tasks (Pakkanen & Raisamo, 2004). Foot pedals have been extensively available to support musicians when playing instruments such as guitars, pianos, organs, drums. These facts combined with the fact that human beings are bipedal leads to the possibility that it is opportune to investigate foot-based interaction.

Similar to natural gestures such as hand and foot movements, interests in natural and organic interfaces have increased recently. Natural and organic interfaces can provide tangible interaction. Water is a medium that is naturally intertwined with human life since birth. Our innate affinity with water is frequently visible in the way people come together around fountains, commonly located in city centers. Ashiyu, public places where people can bathe their feet, are quite common in Japan and are a regular part of the cultural activity. In these environments, a strong potential exists for highly user-friendly (or invisible) interfaces that use water as an interface medium. Interaction with water is an example for the merging of two not so friendly mediums – water and computing technology. Further, interaction with water can provide auditory, visual as well as tactile feedback. The unique sensation provided by water can have a relaxing effect on the body; in fact, fatigue attributed to gestural movements in the air and on surfaces has been identified as an issue that requires investigation (Yee, 2009).

1.2 Research overview

The Human Computer Interaction discipline contains aspects from different domains. In this research the hardware aspect considers detection devices which can detect objects embedded in water for interaction, while from the software aspect detecting foot gestural interaction poses is considered. This research concentrates on the intersection between these two domains of water interaction and foot gestures (See Figure 1.1).

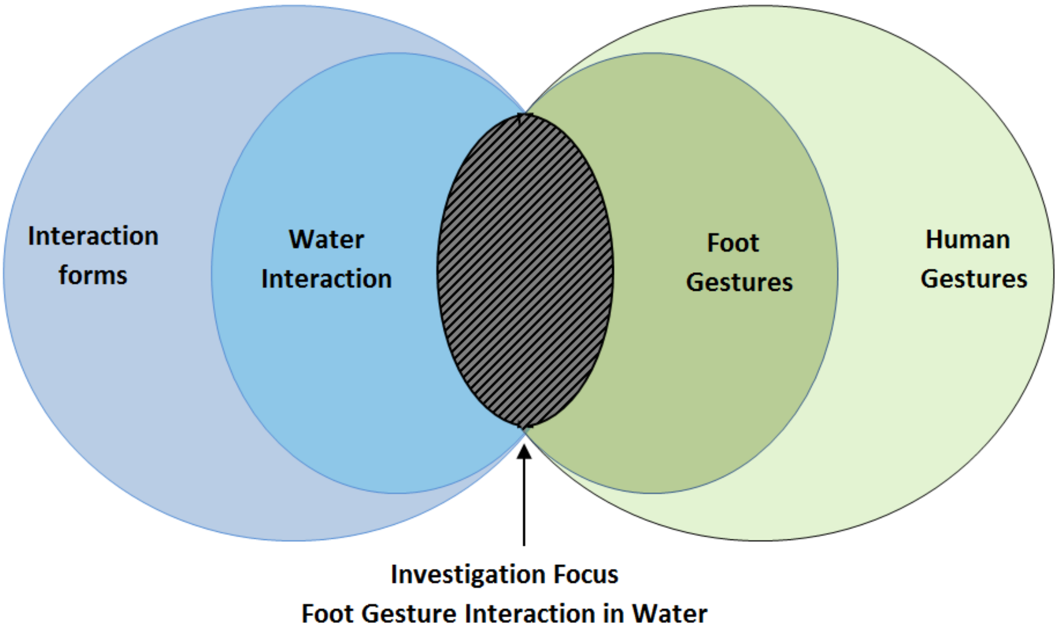


Figure 1.1. Investigation focus of the research

This dissertation investigates how human foot gestures performed in the medium of water can be detected and used as a viable interaction method between humans and computers. Whereas previous research has dealt with water interaction at the water surface, our research deals with three-dimensional interactions with a volume of water in a water vessel. It analyzes the existing technologies used for interaction with water, and identifies their limitations and further enhances them to present an improved system. The dissertation further studies natural forms of interaction with water and examines which static gestures are practically feasible for recognition.

1.3 Research objectives and scope

The main topic of this dissertation is detecting foot gestures in water and the main research question is:

How is it possible to detect gesture interaction performed by human feet in water?

This research question formed the foundation for a set of aims and objectives upon which this dissertation is based. These are to:

1. Evaluate existing hardware devices and techniques currently available and documented in research for the purpose of detecting objects in water.
2. Evaluate existing software techniques currently available and documented in research for the purpose of detecting static foot gestures.
3. Develop a robust hardware framework that can detect foot movement in water.
4. Test the hardware framework against external environmental factors
5. Develop a software framework that can detect the static gestures using objects detected by the hardware
6. Evaluate software framework static gesture detection performance.

The interaction technique applies to a single person using the system at a time.

1.4 Research contributions

The contributions made by this dissertation to the body of knowledge can be explained under several categories.

1.4.1 Framework for object detection in water

This dissertation introduces a novel approach for detecting physical objects within a 3D volume of water. In the literature survey, it was not possible to locate any complementary research that has addressed this issue of detecting objects of the order of a foot which is submerged in water. The detection is even possible if the object is completely submerged, which although not applicable to the human experiences discussed, can have other applications.

1.4.2 Exploration of foot movement in water as an interaction technique

The dissertation presents an analysis of several foot movements on interaction in water. The user feedback can be helpful for future researchers to design further novel interaction techniques for foot gestures beyond those presented in this dissertation.

1.4.3 Applications of foot movement in water for Infotainment

The dissertation proposes several novel applications that make an ordinary foot bath to be a musical instrument and relaxing pool. These examples of ubiquitous computing can apply to a home environment as well as public spaces.

1.4.4 Medical – rehabilitation

Since water is closely associated with life from the time of birth, the medical-rehabilitation domain has many applications that are closely related to our proposed system. Water therapy or hydrotherapy is a well-known treatment method that uses the physical properties of water for wellbeing.

Some disorders affecting children create difficulties in the development of motor skills. These include motor skills disorder, autism spectrum disorder, dyspraxia, cerebral palsy. The effect of water via hydrotherapy as a treatment for children suffering from these disorders has been documented (Mortimer, Privopoulos, & Kumar, 2014).

Some illnesses that affect humans during adulthood including arthritis, as well as movement disorders after illnesses such as strokes also, can benefit from rehabilitation activities carried out in water. In physical therapy, to evaluate the ankle joint swelling and Edema measurement on ankle injuries a water volumeter is used (Petersen et al., 1999; Reis et al., 2004). The system proposed in this research can be of this extended to measure the volume of water displaced when an object such as the hand or foot is immersed.

As humans age, the perceptual-motor (Activities that involve the interaction and integration of perceptual processes and voluntary physical movement) adaptability declines (Guan & Wade, 2000). The ankle often the source of problems involving gait in this situation, and a water based monitoring system can be an effective medium for rehabilitation. Hot water based balneotherapy has been found to be useful in these circumstances (Berger, Klein, & Commandeur, 2008). However, as people age, sensory acuity diminishes (Edelstein, 1988). Aging can lead to accidental hot water burns, which can be avoided by using our system where the temperature of the water body can be illustrated using color infographics.

1.5 Chapter organization

The next chapter describes the comprehensive bibliographic literature survey carried out to investigate the prior work conducted in interaction mediums, water interaction, gesture detection and related domains. The chapter also provides an overview of technologies used for gesture detection, machine learning, and machine learning algorithms.

In the third chapter, the hardware system developed is explained in detail together with previous experiments performed in water to test the viability of same techniques used for gesture detection in the air. It also describes challenges experienced in the development, and how the challenges were overcome.

Chapter four describes the software architecture of the developed system, the main gesture detection process and application scenarios for our system. A number of experiments were carried out to evaluate the robustness of detection as well as the suitability of chosen foot gestures. They are detailed in chapter five together with the experiments conducted to assess the gesture detection ability of our system. This chapter concludes by discussing the results of our experiments.

The final chapter discusses the summary of findings concerning the environment and previous research available. Limitations of this study and suggestions on ways to overcome them are further discussed. Finally, insights on future research directions are stated.

Chapter 2

Review of literature

The first part of the Literature Survey aimed to analyze the existing research carried in Hardware related domains to conceptualize a solution to detect objects in water. The subsequent section covers the types of gestures relevant for the study and the software approaches to detect gestures.

2.1 Interaction mediums in Human Computer Interaction

Interaction forms in HCI can be classified according to different factors. One method could be based on the medium of interaction.

2.1.1 Air interaction

A popular medium of interaction has been air. Actions performed in air have contributed to a majority of this category. A number of studies have been carried out in a variety of related mediums which include fog (Rakkolainen & Palovuori, 2004), Soap bubbles (Döring, Sylvester, & Schmidt, 2012) without touching a device or surface. 3D interaction on air in the proximity of tabletop surfaces has also been experimented (Takeoka, Miyaki, & Rekimoto, 2010). While interaction performed in air by the hand, or any other part of the body is common, it generally lacks feedback from the physical world.

2.1.2 Tangible interaction

A Tangible user interface (TUI) provides users the ability to interact with the digital system and its information using manipulation of physical objects which are related to the system. This concept was introduced by Prof. Hiroshi Ishii in 1997 (Ishii & Ullmer, 1997). TUIs provide the graspable functionality to ubiquitous systems.

Tangible interaction can provide the user with tangible and tactile feedback. As unlike interaction in air, the user receives the feeling of pressure, friction or resistance. The interaction can take place at the surface level of a medium such as a sensor, touch sensitive surface, or even ice (Ventä-Olkkonen et al., 2014; Virolainen et al., 2010) or it can be immersive in a medium such as water, or sand or mud (Gerhardt, 2009).

2.1.3 Liquid interaction

In liquid interaction, developing tangible and flexible interfaces using ferromagnetic fluids, has been explored (Koh et al., 2011). Yet, out of the liquids, water being a natural medium which is close to human life is probably the most commonly cited in interaction research. Unlike the previously discussed categories, interaction with water provides not only the feeling of temperature, liquidness, and fluidity but the sensation persist longer than other elements as the interaction object such as hands or feet gets wet (Pier & Goldberg, 2005). Further water intensifies the experience with the system as it wraps around the fingers and hand (Pier & Goldberg, 2005). The physical properties of water have been utilized for many types of research. These properties include water flux, water movement, and water pressure. While water interaction is not universally applicable to all situations as actions in air, there are specific locations where they may be more relevant.

Raffe et al. categorized water contact in player-computer interaction into six groups (Raffe et al., 2015):

1. Vicinity of water
2. Sporadic contact
3. On top of water
4. Partially submerged
5. Floating
6. Underwater

A few devices that are designed for on-air interaction have been used in water, although not directly for interaction. They include the use of the Wii Remote for water level measurement (Hut, Weijs, & Luxemburg, 2010). The Kinect has been used for measuring water level at shallow depths up to 0.203 meters (Mankoff et al., 2011).

Public spaces provide people to socialize with each other and relax. Often, water fountains are artificially created in public squares, and water provides itself to be a soothing substance, cooling the air, while the sound of water is falling acting as a calming catalyst. Researchers have enhanced such locations using technology. Gurgle (Arroyo, Bonanni, & Valkanova, 2012) is an interactive location in a public space which augments an existing water fountain with watery reflections and sound to motivate people to pause and take a drink of water. The city mouse (Häkkilä et al., 2014) is an interactive location where the participants rotate a 3D model of the earth presented on a screen by using a stone ball resting on a water fountain. Rainterior (Okude, 2011) is an interactive display based entertainment system that detects raindrops are falling on a water surface.

2.1.3.1 Sporadic contact with water

Mann's hydraulophone (Mann, Georgas, & Janzen, 2006) uses an array of water jets as a haptic surface that also functions as a musical keyboard. The Tantalus Fountain (Dietz, Han, Westhues, Barnwell, & Yerazunis, 2006) utilizes the electro-optical properties of water to use it as a capacitive proximity electrode. The AquaHarp is a musical instrument that looks like, and is operated like a harp but has streams of flowing water instead of strings. (Dietz et al., 2006).

Kitchens and Bathrooms are two other locations in which human-water interaction occurs on a daily basis. This provides researchers with an ideal opportunity to introduce water interaction technologies. Smart Sinks uses a webcam, LEDs, RFID reader to enhance interaction with the water faucet (Bonanni, Arroyo, Lee, & Selker, 2005). TubTouch provides an integrated user interface for entertainment in a bathtub using capacitive sensing (Hirai, Sakakibara, & Hayashi, 2013). Vortexbath provided tangible interaction on a water bowl by detecting ripples and vortex and playing audio-visual media (Watanabe, 2007).

2.1.3.2 Interaction at surface level of water

Touch surfaces such as tablets and screens provided the physical sensation of touch. TouchPond (Dietz et al., 2006) operates much like a liquid touch screen using water. The Kinect has been used for scanning dynamic water surface (Hut, 2011). Aquatop (Matoba et al., 2013) provides an immersive experience to detect gestures performed at the surface of cloudy water by using a Kinect.

2.1.3.3 Immersive interaction with water

While research involving water interaction has been present in numbers, object detection in water has been a relative challenge. Movement of hand gestures performed in water tub has been detected by making use of Total Internal Reflection and cameras (Ikeda & Hirakawa, 2010).

2.2 Devices for interaction detection

Interaction detection devices, which can be considered as the forerunner of modern gesture detection devices, have a longer history than even the modern computer. The Theremin patented in 1928 by Leon Theremin is an electronic musical instrument controlled without physical contact by a performer. The device consists of two antennas that can estimate the relative position of the performer's hands. The performer can control the frequency with gestures performed using one hand and the amplitude with the other hand. At no time do the performer's hands touch the Theremin.

Devices which can detect objects in an interaction setting used today can be divided into two main categories; vision based devices which use optical cameras for detection and sensor-based devices.

2.2.1 Vision-based detection

In their study of literature spanning 40 years, Karam & Schrafel identified Cameras as the most widely used devices for gesture recognition (Karam & Schraefel, 2005). They have been used for both static gesture detection via pictures or single video frames, or for temporal gesture detection via streaming video. The initial attempts used grayscale images on black and white cameras with low frame rates and frame resolutions. In the recent past, the introduction of inexpensive red-green-blue (RGB) video cameras, especially "webcams" with high resolution used for video conferencing via personal computers has created an increase in research which uses Image processing. Other reasons that have contributed to this situation are the significant decrease in CPU cost involved in processing video, as well as the development of new data compression techniques for video, and powerful algorithms.

In the marine industry, automated sorting of fish and measuring them is a very useful feature. Camera-based approaches have been proposed with satisfactory results (White, Svellingen, & Strachan, 2006; B. Zion, Shklyar, & Karplus, 1999; Boaz Zion, Alchanatis, Ostrovsky, Barki, & Karplus, 2007).

Camera-based approaches do have their limitations especially with poor performance in low light conditions. One approach to improving performance was to use markers (Dutta, Aparna, Sridharan, & Vipin, 2013). Other approaches include multiple cameras (Malik & Laszlo, 2004), high-resolution cameras (Sangsuriyachot & Sugimoto, 2012), high frame rate cameras (Takeoka et al., 2010), and special cameras such as heat and infrared sensing. Another complementary approach has been to illuminate the object involved in the gesture using infra-red (IR) light and use normal RGB cameras retrofitted with IR filters (Mistry & Maes, 2010).

The Microsoft Kinect uses IR laser projection and a combination of two cameras, a conventional RGB camera, and an IR-sensing camera, to detect 3D movements of a human body. It has been used to detect hand gestures (Oh, Kim, & Hong, 2013; Ren, Meng, Yuan, & Zhang, 2011), posture (Visutarrom, Mongkolnam, & Chan, 2014), as well as combinations of the above and foot gestures (Hoste & Signer, 2014). Nevertheless, its limitation on use in direct sunlight has also been documented (Hoste & Signer, 2014) The leap motion device is another compact sensing device that is used to capture hand and finger movement (Schmidt, Araujo, Pappa, & Nascimento, 2014). It utilizes two IR cameras supported by three IR LEDs as its sensing technology ("Leap Motion Patent Application," 2014).

While sensor based multi-touch screens are popularly found in smartphones and tablets, their cost rises with the dimensions of the display, making it very expensive for large-sized screens. The Frustrated Total Internal Reflection (FTIR) technique introduced by Han (J. Y. Han, 2005) enabled larger shared displays with multi-touch – multi-user functionality. The touch surface is edge lit using IR LEDs, and the diffused IR light is captured using a Camera, which contains an IR pass filter. Image processing carried out on a PC detects the touch points. This scheme is coupled with rear projection screens to create the multi-touch enabled display. This technique has been modified for presenting floors that detect feet (Sangsuriyachot & Sugimoto, 2012) and gestures performed on the water surface (Dietz et al., 2006).

Nevertheless, vision-based gesture recognition can be affected by a variety of noises because of external/environmental factors, such as lighting and a need for high processing power in the case of multiple cameras.

2.2.2 Sensor-based detection

Depending on their purpose, hardware sensors can acquire different physical characteristics. One of the earliest sensor based devices with a history spanning 30 years in use is Data gloves or wired gloves. Karam and Schrafel's literature survey reveals gloves as the second most widely used in gesture research (Karam & Schraefel, 2005).

They consist of a sensor-embedded glove to capture the flexing of joints in the hand and individual fingers. A motion tracker is also embedded to capture the position and acceleration of the glove. The tracking is conducted using magnetic tracking or inertial tracking. Data gloves have been used for sign language hand gesture dataset recording (Kadous & Sammut, 2005). While data gloves have been expensive and aimed in general at professional users such as those in the computer graphics and movie industry, they have provided high capture resolutions.

Within the last decade, the use of sensor-based gesture research has increased. The "smart" mobile phone has often been used as the sensing device for gesture recognition (Crossan, Brewster, & Ng, 2010; T. Han, Alexander, Karnik, Irani, & Subramanian, 2011; Scott, Dearman, Yatani, & Truong, 2010). While the size and weight have been decreasing, Smartphones today contain multiple sensors such as accelerometers, GPS, magnetometers. They can be used to detect information, such as relative position and 3D acceleration. Compared with vison-based systems, one advantage is that the user is free to move in either direction, rather than face a camera for accurate detections.

The Nintendo Wii is a popular gaming console. The Wii Remote Controller (Wiimote) which can detect movement in 3D has been used for hand gesture detection (Schlömer, Poppinga, Henze, & Boll, 2008). The key components in the Wiimote for movement detection are accelerometers and IR sensors.

Accelerometers have been embedded not only in the mobile phones and Wiimotes, but also in watches (Alexander, Han, Judd, Irani, & Subramanian, 2012; Mace, Gao, & Coskun, 2013) and shoes (Paradiso, Hsiao, & Benbasat, 2000) for the purpose of gesture detection.

Using electromagnetic waves for detecting objects has a long history dating back to the 2nd world war. RADAR (RAdio Detection And Ranging) was the only technique at the time. LIDAR (Light Detection And Ranging) is a complementary technology that uses lasers to determine the object location. While these technologies are applicable to large objects and longer distances, Google's project Soli is aimed at detecting micro gestures performed using the hands. The Soundwave technique detects dynamic hand gestures using electromagnetic waves generated using a personal computer speaker and sensed by built-in microphones using the Doppler effect (Gupta, Morris, Patel, & Tan, 2012).

Touch screens are part of many consumer electronic devices today, it enables making the device compact, as space for physical buttons is eliminated by the touch input enabled display. The initial displays used resistive technologies that needed pressure to be applied to the selected region and used stylus pens for location selection. Capacitive sensing is a more recent technology used popularly in touch screen based smartphones and tablets today. A related technique known as swept frequency capacitive sensing has been used to detect gestures performed in Air, on surfaces and water using only a single electrode (Sato, Poupyrev, & Harrison, 2012). Although initial sensing was limited to one touch point, multi-touch became popular usage with the launch of the Apple iPhone.

Gesture detection mechanisms have been worn on different parts of the body. iRing is an intelligent input ring device that can detect finger gesture via and IR reflection sensor (Ogata, Sugiura, Osawa, & Imai, 2012). Also, devices have been mounted on legs, arms, neck (Mistry & Maes, 2009), head (Mistry, Maes, & Chang, 2009) and on feet or footwear (Bailly, Müller, Rohs, Wigdor, & Kratz, 2012; Crossan et al., 2010).

Electromyography (EMG) sensing is another latest addition to the sensor technology. The muscles in the human body are controlled by motor neurons that transmit electrical signals. The signals cause the muscles to contract. An EMG sensor uses electrodes to detect and measure the electric signals. The MYO armband by Thalmic Labs uses this technology and contains eight medical grade electrodes that touch the skin but do not require conductive gel to get good readings. Also, the MYO encloses an Inertial Measurement Unit (IMU) which comprises of a 3D Gyroscope, 3D accelerometer, and a magnetometer. The gestures performed using this armband has been experimented for musical interaction (Nymoen, Romarheim, Alexander, & Jensenius, 2015). The MYO is designed to be worn on the thickest part of the forearm muscle.

One negative aspect of body worn sensors is that due to the detection mechanism being strapped to the body, the user’s natural movement can be impeded.
On the other hand, external detection is non-intrusive as there is no need to wear cumbersome devices. These can be in the form of Cameras, sensor-based devices such as Kinect, Depthsense or even floor mats.

2.2.3 Hybrid detection devices

Since the vision based approach and sensor based approach both have their own limitations and advantages, merging the two approaches can result in hybrid detection schemes. Digits is a wrist-worn device that can generate the full 3d pose of the user's hand (Kim et al., 2012). It contains an IR camera as well as an IMU. Commercial Hybrid devices that combine both sensor-based and vision-based methods are yet to be made available.

2.3 Supplementary technologies for interaction detection

While section 2.2 described the main technologies used for gesture recognition, a few supplementary technologies exist. They are not used exclusively for gesture detection, but rather find their way into some of the devices mentioned in section 2.2. A few others technologies are used to integrate each other.

2.3.1 Light Emitting Diodes (LEDs)

A LED is a two lead semiconductor device which emits light in the form of photons from its p-n junction when activated. While LEDs have been used in many electronic devices as indicator lamps, the use of them as an compact, low power, reliable light source has been explored in interaction detection. Visible light in the form of Blue LEDs have been used for mixed reality Art installations (Rudomin, Diaz, Hernández, & Rivera, 2005), to detect water waves in Virtual Reality Applications (Pier & Goldberg, 2005), and to create a Total Internal Reflection (TIR) based multi-dip interface (Ikeda, Nagira, & Hirakawa, 2009).

Visible light illumination can clash with existing illumination or sunlight. Therefore Infra-Red (IR) light emitting LEDs have more commonly been used to illuminate objects for interaction detection. They emit IR with wavelengths of 700nm -1050nm. IR LEDs have often been used to illuminate FTIR based interactive surfaces (J. Y. Han, 2005; Okude, 2011) including ones which use water (Dietz et al., 2006).

2.3.2 Lasers

Lasers provide more focused beams of light when compared to LEDs. But due to the risk of eye damage due to the intensity of the beams, the use has been limited in most cases to surfaces where eye contact is not frequent. An exception is the popular

Microsoft Kinect Device, which uses an IR laser with a 780nm wavelength in a diffused speckle pattern. Due to the speckle pattern, the intensity of a speck is not strong enough to cause any damage to the eye. Lasers use diffraction gratings to produce different patterns. One pattern which has been used for detecting objects is the line pattern, which when combined with the laser is referred to as a Line Laser. IR line lasers have been used to detect gestures for moving a virtual mouse (Mistry & Maes, 2010).

2.3.3 Acoustics

Low-cost acoustic sensors that use ultrasound have been used for gesture detection successfully (Gupta et al., 2012). A special category of devices referred to as Acoustic Cameras such are used in the marine industry for fish sorting and counting. The DIDSON (Dual Frequency Identification SONar) is one such product (J. Han, Honda, Asada, & Shibata, 2009).

2.3.4 Distance sensors

Distance Sensors or Proximity sensors can be used to measure the presence of nearby objects without physical contact. They often operate using infrared beams

2.3.5 Floor sensors

Floor Sensors have been used to measure human gait, which can also be linked to body pose. One type of sensor element that is used in such devices is Force Sensing Resistors. Commercial mats with active areas of 6mm x 6mm have been used in research (Srinivasan, Birchfield, Qian, & Kidané, 2005). The Nintendo Wii Balance Board is a commercial product consisting of pressure sensors and has been used for navigating spatial data with feet (Schöning, Daiber, Krüger, & Rohs, 2009).

2.3.6 Capacitive sensing

Capacitive Sensing is based on capacitive coupling where the human body capacitance is taken as an input. This technique is commonly available in Touchpads of laptop computers and screens of tablets. Capacitive Sensing has been used to detect bed postures (Rus, Grosse-Puppendahl, & Kuijper, 2014) as well as Touch Sensors in Bathrooms (Hirai et al., 2013). A related technique referred as swept frequency capacitive sensing has been used to provide touch interaction to everyday objects (Sato et al., 2012).

2.3.7 3D Scanners

3D Scanners are used to analyze real world objects and collect data pertaining to its shape and appearance. The geometric shape of an object scanned is generated as a point cloud, which can be used to reconstruct the shape. Non-contact active scanners use a time of flight laser range finder technique or a triangulation based lasers scanning technique.

2.3.8 Microcontrollers

In the last decade, miniaturization of computing devices has given birth to a new generation of devices which can fit on the human palm while being cost effective. This has enabled hitherto unknown electronic sensors to be interfaced with computers. The Arduino range of microcontrollers is one of the most popular small factor devices. The Arduino Uno is one of the most popular models. It is based on the ATmega328P microcontroller. It has 14 digital input/output pins and 6 analog inputs. USB is the preferred connectivity option to a computer. It operates on 5VDC.

The Arduino is not directly used for interaction detection, but has been used with many different sensors and different situation such as, when interfacing with solenoid modules for controlling water output (Hoste & Signer, 2014; Richter, Manke, & Seror, 2013), measuring the amount of pressure under an array of force sensing resistors (Gerhardt, 2009), controlling robotic arms (Richter et al., 2013).

The Raspberry Pi ("Rasberry Pi," 2015) is a similar, more powerful microcontroller which can emulate most basic applications run on a personal computer.

2.4 Types of gestures

Gestures can be classified according to different characteristics. Regarding the part of the body involved, gestures can involve fingers, hands, arms, head, feet or full body motion.

2.4.1 Hand gestures

Perhaps one of the first forms of non-verbal communication would have sign language as observed from the drawings from different parts of the world. In today's context, while there are thousands of languages spoken, for the deaf and the mute or people with similar disabilities, they cannot be of much help. Therefore, sign language and its use will continue to exist as long as humanity finds such disabilities.

Technology plays an increasingly valuable role for physically challenged individuals, and while there are sign languages being used in the various countries. The American Sign Language contains the American Manual Alphabet consisting of English alphabet and ten digits ("American Manual Alphabet," 2012) and this was exclusively expressed using fingers and single hand gesture.

Decoding such country-specific sign language has been carried out (Lamari, Bhuiyan, & Iwata, 1999; G. W. Wang, Zhang, & Zhuang, 2012) while many others have considered limited gesture subsets (Dhawale, Masoodian, & Rogers, 2006; Francke, Ruiz-del-Solar, & Verschae, 2007; Gupta et al., 2012; Lenman, Bretzner, & Thuresson, 2002; Malik & Laszlo, 2004; Ren et al., 2011; Sato et al., 2012; Schmidt et al., 2014; Zeiß, Marinc, Braun, Große-Puppendahl, & Beck, 2014).

Gestures have been performed using one hand as well as both. In certain cases, the gesturing hand contains additional objects by way of tracking devices (Schlömer et al., 2008) or markers (Mistry et al., 2009).

Displacing water from the surface in a throwing action using hands is often carried out by children who play with each other in the water. Scooping (Figure 2.1a & 2.1b), Paddling (Figure 2.1c), Twirling (Figure 2.1d) can further be identified as hand gestures that are associated exclusively with water interaction. Scooping has been used to pick up virtual objects floating on water (Tanabe & Hirakawa, 2006) as well as for detection as a gesture(Ikeda & Hirakawa, 2010).

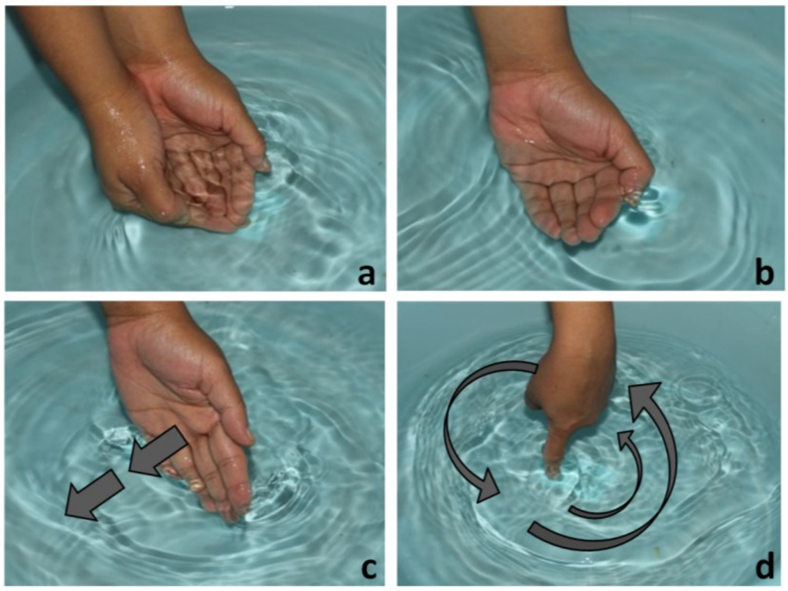


Figure 2.1. Natural hand gestures performed with water

2.4.2 Foot gestures

In comparison with hand movements, foot movements have had limited research as well as applications. One reason is that the foot does not offer the same precision and dexterity as hands (Scott et al., 2010). They identified four movements as possible foot gestures (Figure 2.2a – 2.2d).

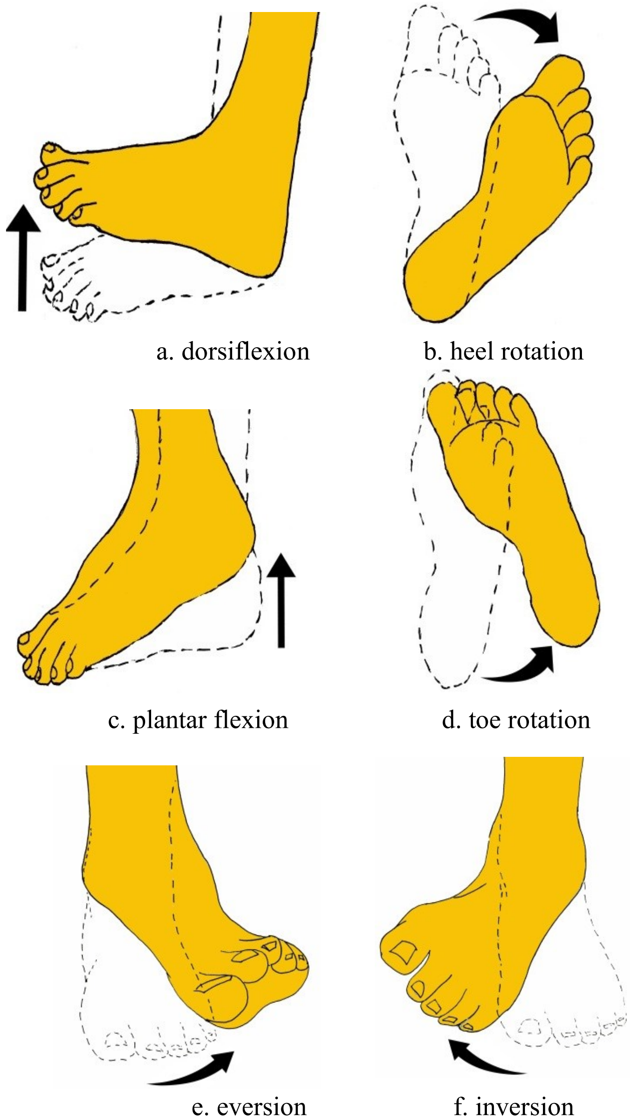


Figure 2.2. Six basic foot movements

(Perspectives for (b) and (d) are from the right foot. Perspectives for (a), (c), (e), and (f) are from the left foot. (Gunawardena & Hirakawa, 2015))

- Dorsiflexion: rotation of the ankle that decreases the angle between the shin and foot.
- Heel rotation: internal and external rotation of the foot and leg with respect to the midline of the body, while pivoting the rotation on the heel.

- Plantar flexion: rotation of the ankle that increases the angle between the shin and foot.
- Toe rotation: internal and external rotation of the foot and leg, while pivoting the rotation on the toe.

The movement of the feet is a research area in biomechanics, and two other motions are described in the literature (Floyd, 2008). Inversion (Figure 2.2.e) and eversion (Figure 2.2.f) are movements to face the sole of the foot inwards and outwards, respectively. These movements have limits on the degree of flexibility, in terms of dexterity. The typical limits for inversion and eversion are 20–30 degrees and 5–15 degrees (Floyd, 2008), respectively. Similarly, the ranges of motion of dorsiflexion and plantar flexion are 10–20 degrees and 40–55 degrees, respectively (Nordin & Frankel, Victor, 2012).

Kicking (T. Han et al., 2011) and Foot Tapping (Crossan et al., 2010) have also been foot gestures that have been experimented on for mobile interaction.

Paddling with the feet is a movement that can be natural with water. It can be performed sideways (Figure 2.3a), up and down in a tapping arrangement (Figure 2.3b) or back and forth (Figure 2.3c). In the last movement, it is possible to create a bubbling effect by moving the legs in and out of the water.

Figure 2.3. Natural foot gestures performed in water

Some advantages of using the foot over hands for Geographical Information System (GIS) applications in surfaces have been documented (Schöning et al., 2009).

- More intuitive for entering continuous data in situations such as navigation.
- Physically less exhausting than using one or both hands when manipulating and application on a surface.
- Provides additional mappings for iconic gestures for single commands

2.4.3 Gait and full body

The locomotion achieved by moving human limbs is referred to as human gait. Gait movement has been tracked using cameras, footwear mounted sensors and pressure floors. High-resolution pressure sensing floors have been used to help study human dance movement (Srinivasan et al., 2005).

Full body gestures can be considered as a more natural and intuitive way to interact with video games. The motions could be movements which can be performed independently such as kicking and jumping, or assisted with the use of a prop such as a ball or a wand. Researches have compiled databases of such full body gestures (Bon-Woo Hwang, Sungmin Kim, & Seong-Whan Lee, 2006).

2.5 Gesture detection process

Gesture detection utilizes the similar techniques as pattern recognition in image or video processing. The stages involved in an application that detects gestures can be illustrated as follows (Figure 2.4).

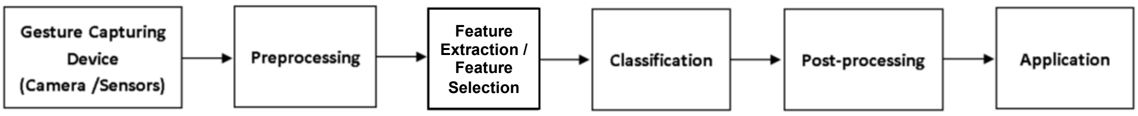


Figure 2.4. Stages in gesture detection

2.5.1 Gesture capturing devices

The devices covered under devices for interaction detection in Chapter 2.2 are the same for this section.

2.5.2 Pre-processing

This is an initial stage which uses established techniques to prepare data for another process. The objective is to ensure the next stage of processing is performed easily and effectively. In the case of gesture recognition filtering relevant data or removing irrelevant data using algorithms is beneficial. Algorithms/Filters such as Moving Average, Double Moving average, Low Pass, High Pass, or taking the derivative or second order derivative of data can be useful for preprocessing depending on the task at hand.

2.5.3 Feature extraction / Feature selection

Gesture detection either by vision or sensor techniques generates large volumes of data. Even after filtering, processing such large number of data is complicated. If all the data is used for gesture detection, the gesture training process would require a large number of data or overfitting issue can arise. This phenomenon is known as the "Curse of dimensionality" which refers to difficulties of managing data in a large number of dimensions. Two approaches commonly used to resolve this problem.

In the feature extraction process, "features" which are suitable for detecting the gestures are extracted from the data. Feature extraction can be performed by specialized techniques such as Principal Component Analysis and Linear Discriminant Analysis, which result in lowering of the dimensions. However, the original data would be transformed during this process.

In feature selection, a subset of the data which is most useful is selected to represent the gesture. In feature extraction the meaning of the original data is lost during transformation, feature selection avoids this issue. Features are selected in a way that they optimize an objective function such as correct classification of gestures.

2.5.4 Gesture classification

Using the features selected, a model needs to be trained with each gesture labeled as belonging to a class. Once sufficient numbers of samples have been trained, new gesture samples can be processed for classification. Algorithms used for gesture classification are separately discussed in section 2.7.1.

2.5.5 Post-processing

Once the classifier has marked a gesture as belonging to one class, it further can be processed by setting filters such as class label filters to enable a minimum consecutive number of samples to be detected before the post-processing module to output the classifier output. Another post processor provides a timeout period where if no labels are present the post-processing module does not generate output. The output of the gesture recognition process can be channeled to an external program which uses the gestures identified to produce an output from the computer system.

2.6 Protocols associated with gesture detection

2.6.1 Open Sound Control Protocol

Open sound control (OSC) is a protocol used for communication among computers and musical devices such as synthesizers in a networked environment. Its origins are closely tied to the MIDI standard used in hardware synthesizers. A connection to a device requires opening a Transmission Control Protocol (TCP) or User Datagram Protocol (UDP) connection using an IP address or hostname and a port number. OSC implementations are available as libraries in many modern programming languages. One reason for its popularity in the in the musical industry has been the open-ended support for data formats. While it has not been used exclusively for gesture detection related software, since gestures have been closely associated with multimedia presentations, such research has been documented (Gillian, 2011).

2.6.2 Tangible User Interface Objects Protocol

Tangible User Interface Objects (TUIO) is a protocol designed especially to deal with the needs of tangible user interfaces such as touch tables. It is based on OSC, and, therefore, can be easily implemented on any platform that supports OSC (Kaltenbrunner, Bovermann, Bencina, & Costanza, 2005). It sends UDP packets via port 3333. TUIO Server implementations are supported by many major multi-touch hardware vendors such as Reactable[5] and PQ Labs[6]. There are client implementations in different programming languages as well as application frameworks. Although many implementations are on 2D interactive surfaces, the protocol defines profiles that support 2.5 D as well as 3D (Kaltenbrunner et al., 2005).

2.7 Machine learning approach for gesture detection

2.7.1 Algorithms for gesture recognition

Machine learning is an extensive discipline with applications in different domains. Therefore, a large number of algorithms have been developed to suit various types of problems. There are two kinds of problem for which machine learning is

---

[5] http://www.reactable.com/products
[6] http://www.multitouch.com/product.html

commonly applied: classification and regression. The output takes discrete values in classification, but continuous values in regression.

The algorithms need to “learn” and model accordingly from data, and this gives rise to categories of machine learning algorithms based on the learning style.

1. Supervised Learning
2. Unsupervised Learning
3. Semi-supervised Learning
4. Reinforcement Learning

Spatial gestures require static classification for which algorithms such as naïve Bayes (Mace et al., 2013; Scott et al., 2010; Ziaie, Müller, Foster, & Knoll, 2009), k-nearest neighbor (k-NN) (Nimbalkar, Karhe, & Patil, 2014; Vafadar & Behrad, 2008), adaptive boosting (AdaBoost) (Hoffman, Varcholik, & LaViola, 2010; C. C. Wang & Wang, 2008), support vector machines (SVMs) (Dardas & Georganas, 2011; Oh et al., 2013; Sato et al., 2012) and decision trees (Oh et al., 2013) have been used. On the other hand, a temporal classification problem in which real-time tracking is to be performed requires different algorithms, such as hidden Markov models (Chen, Fu, & Huang, 2003; Schlömer et al., 2008) and dynamic time warping (Barczewska & Drozd, 2013).

For gesture recognition, the supervised approach has commonly been applied. In this approach, the first task is to identify the different gesture categories or classes involved and record gestures belonging to each. The next step is to isolate a training set in selected gestures from all classes are recorded and used for building a prediction model. This also requires the selection of a suitable learning algorithm to construct the model. Once a set of gestures are recorded and classified, the model can predict which class a new incoming input value belongs to. For the purpose of training a model, the gesture is defined using a number of relevant feature vectors.

2.7.2 Toolkits for machine learning

2.7.2.1 Weka

Waikato Environment for Knowledge Analysis (WEKA) is a machine learning software was developed at the University of Waikato, New Zealand. It was started as a project in 1992, at a time when learning algorithms were not unified and available for use on one platform. The forerunner to its current versions was developed in 1997 using Java language.

Apart from supporting a large number of existing algorithms, WEKA enables the addition of new algorithms by way of its framework and, therefore, permits researchers and developers to concentrate on the new algorithms itself, rather than having to focus on the supporting infrastructure and evaluation mechanisms (Hall et al., 2009). The publication of a series of books (Witten, Frank, & Hall, 2011) together with the support mailing list and a WEKA e-learning course (WEKA MOOC), have added to its popularity.

Further, the ability for researchers to use the functionality of WEKA using a GUI is also a plus point. A non-technical person could use the WEKA Explorer GUI option from the initial screen (Figure 2.5) to easily analyze data.

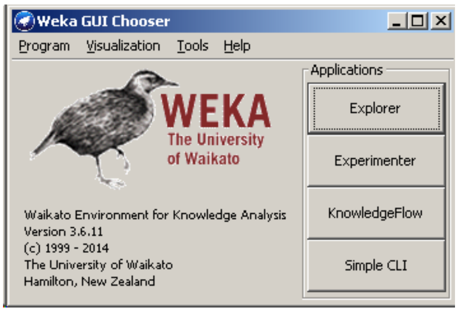


Figure 2.5. Weka GUI Chooser

This has resulted in many researchers selecting WEKA as their chosen machine learning software for applications in gesture recognition (Francke, Ruiz-del-Solar, & Verschae, 2007; Sato, Poupyrev, & Harrison, 2012; Schmidt, Araujo, Pappa, & Nascimento, 2014; Visutarrom, Mongkolnam, & Chan, 2014)

When classifying data using Weka, its output (Figure 2.6) provides a number of performance measures.

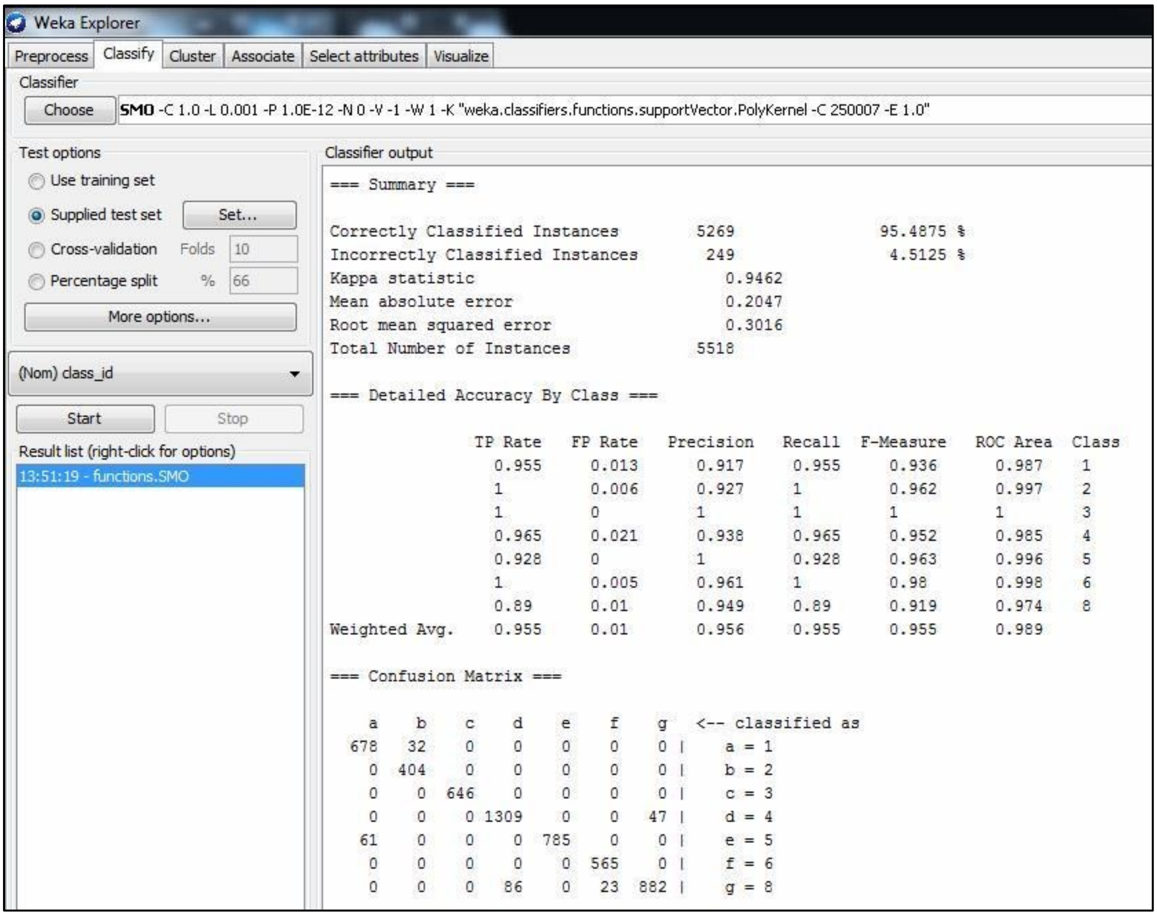


Figure 2.6. Weka classifier output

They can be explained as follows.

1. Correctly classified instances – the number of instances correctly classified. This is indicated by a number as well as a percentage of the total instances submitted to classification. This has certain disadvantages as a performance estimate as it is not sensitive to class distribution.
2. Incorrectly classified instances - the number of instances incorrectly classified. This is indicated by a number as well as a percentage of the total instances submitted to classification.
3. Kappa statistic – used to measure the agreement between predicted and observed categorisation of the dataset, while correcting for an agreement that occurs by chance (Witten et al., 2011). A value of 1 indicates perfect agreement while 0 indicates a chance agreement.

4. Mean absolute error – Is the average of the absolute errors, where an absolute error is the absolute difference value between the prediction and the corresponding true value
5. Root mean squared error – is the square root of the mean squared error, where the mean squared error is the average of the square of every absolute difference value.
6. Total number of instances – the no of samples in the training / test dataset
7. Confusion matrix – (Table 2.1) contains information about actual and predicted classifications done by a classifier such as WEKA.

Table 2.1. Confusion matrix for two class variable

		Predicted Class	
		Positive	Negative
Actual Class	Positive	True Positive	False Negative
	Negative	False Positive	True Negative

- true positive (TP): predicted to be positive and the actual value is also positive
- false positive (FP): predicted to be positive, but the real value is negative
- true negative (TN): predicted to be negative and the actual value is also negative
- false negative (FN): predicted to be negative, but the actual value is positive

8. TP rate - Positives correctly classified (as a given class) calculated as a fraction of the total positives = TP / (TP+FP)
9. FP rate – Negatives incorrectly classified (as a given class) calculated as a fraction from the total negatives = FP / (FP+TN)
10. Precision – Proportion of instances that are truly of a class divided by the total instances classified as that class = TP / (TP+FP)
11. Recall – Proportion of instances classified as a given class divided by the actual total in that class (equivalent to TP rate)
12. F-measure – Is a variant of accuracy that is not affected by negatives.
    Calculated as 2 (Precision) (Recall) / (Precision + Recall)

13. Receiver Operating Characteristic (ROC) Area – ROC is a two-dimensional graph in which the false positive rate is plotted on the X axis, and the true positive rate is plotted on the Y axis. The ROC curve is considered to be a good evaluator for comparing classifiers. An optimal classifier will have an ROC area value approaching 1 with 0.5 being comparable for random guessing.
14. Class – is the class label under consideration.

2.7.2.2 Gesture recognition toolkit

While there is software such as WEKA, Matlab, and R for machine learning, they are mainly used for offline gesture analysis. The gesture recognition toolkit (GRT) is aimed at supporting researchers, technologists, artists and similar interest individuals who are not hardcore programmers for the purpose of real-time gesture recognition (Gillian & Paradiso, 2014). The software is an open source product developed using C++ under the MIT license and has been available since 2012. Apart from a comprehensive C++ API, it has an easy to use graphical user interface (GRT – GUI).

A few more advantages of using the GRT over the mainstream machine learning software is that it can be easily integrated into C++ projects, as well as flexible enough to be used for image processing, video processing, sensor inputs or a hybrid detection system. It supports a large number of preprocessing, feature extracting, classifying and post processing algorithms. A few of the classifier algorithms it currently supports include – Adaptive Naive Bayes Classifier (ANBC), AdaBoost, K-Nearest Neighbor (k-NN), Decision Tree (DT), Dynamic Time Warping (DTW), MinDist, Support Vector Machine (SVM), and Softmax Classifier. The Null Rejection coefficient parameter in GRT is a useful one which is not found in other toolkits. This enables rejection of movements which do not belong to any trained class.

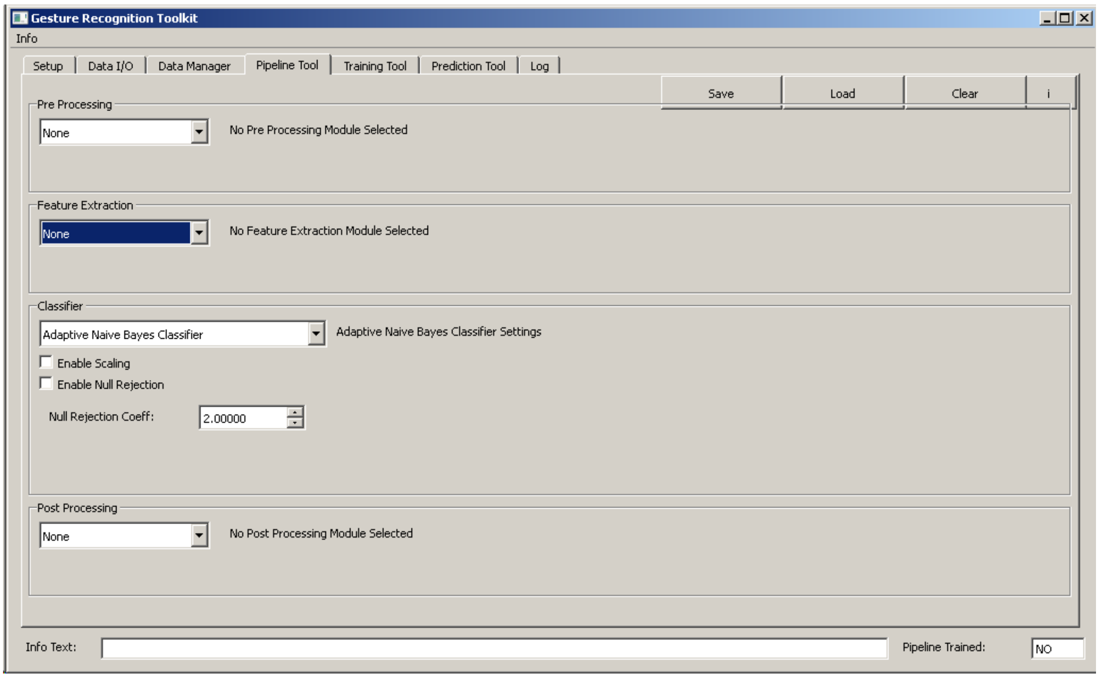


Figure 2.7. GRT GUI classifier selection menu

When using the GRT-GUI for classification, the first options in the setup table should be set. They include the type of task involved (Classification or Regression), number of input feature vectors, and the number of output feature vectors. The data I/O tab is used to configure OSC ports and IP address for sending out data. It also supports controlling the GUI using commands issued via OSC. The data manager enables the OSC data to be recorded in its proprietary file format (in a text format with a header plus tab delimited feature data), or load a pre-recorded file into the system. This tab also provides summarized information on the dataset including class counts and time series graphs.

The pipeline tool tab (Figure 2.7) controls the main processes handled by the GRT. The data fed into the GRT can be pre-processed using a number of built-in filters. The feature extraction option can be enabled if a custom developed module is connected but is not used for regular operations. The classifier selection tool permits some parameters to be tweaked, depending on the selected classifier. For example, “k” in k-NN algorithm. Two parameters are common for all the classifiers. The enable scaling parameter allows min-max scaling on the trained data, and also on real time data input to the system (using trained data). Null rejection is used to reject data that do not belong to any of the trained classes. The null rejection coefficient is a threshold that is used to control Null rejection. The post processing stage contains several filters that can be used to control the output. For example, the class label filter can be used to ensure that within a gesture movement, accidental detection of one sample belonging to another gesture class does not get displayed, once a minimum count of samples is set.

In the training tool tab, the selected classifier can be trained, and once this is done, using the pipeline tool it is even possible to export the trained model for embedding in any C++ application that uses the GRT. Further k-fold cross validation and validation using an external dataset can be performed using the training tool tab.

The prediction tool tab is used for real-time prediction and provides not only the predicted class label, but likelihood probabilities, and a graphing facility.

2.8 The future with gesture recognition

While the number of gesture detection devices which interface with computing devices has recently seen a marked increase, another trend has been the embedding of gesture recognition technology on standalone devices. This requires that the gesture classification engine be embedded in firmware, and provide a classified output. Even with the MYO this technique is used, where it classifies the arm and hand movement based on its sensor readings and provides the gesture class to the SDK. Following this trend, a number of notebook computers introduced within 2014-15 have introduced hand gesture controlling using the webcam built into the system.

Today gesture recognition is not limited to computer interaction. Hyundai's HCD-14 Genesis concept car ("Hyundai HCD-14," 2013) incorporates 3-D hand gesture recognition for controlling the in-car navigation, HVAC (heating, ventilation, and air conditioning) and infotainment system. Samsung's Smart TV ("Samsung Smart TV," 2013) has the ability to detect 13 different hand gestures. Google's Project Soli (Google, 2015) uses radar technology embedded in a sensor chip to detect micro gestures performed close to the sensor surface.

While gesture recognition may not replace the keyboard or mouse as a day to day input device, the operation of computing and other devices with no direct touch has its own applications, especially in the medical field. For example with the introduction of electronic devices in a clean environment such as a surgery room, the risk of harming patients by way of germs in input devices can be eliminated using gesture controls.

2.9 Summary

This chapter presented the results of a literature review which Types of Gestures performed in HCI research, spatial and temporal gesture forms, 2D and 3D gesture forms. While most gestures were completed in an air medium, most interaction that has some degree of liquid interaction was mostly at surface level, and almost all of it was dealing with water. Regarding detection technique, a number of complementary methods have been used. They include body worn vs. external detection, vision vs. sensor based detection. The chapter further presented technologies, protocols and machine learning techniques for gesture recognition.

Chapter 3

Hardware system overview

This chapter explains the hardware technologies developed and experimented and for gesture detection in water.

3.1 Experiments with existing interaction detection devices

While research on water or liquid interaction has been performed by many, most of the experiments, have not actively considered the human gestures in water. Nevertheless, the hardware used is examined for their suitability together with hardware commercially available for gesture detection in the air.

3.1.1 Microsoft Kinect

Previous researchers that utilized the Kinect motion detector device and described in Chapter 2 were not used for immersive interaction with water. In the case of Aquatop (Matoba et al., 2013), its interaction was at the level of the water surface. In the experiments with the Kinect, it was found that when the unit is positioned above the water surface and gestures are performed within water, the ripples that are generated act as a barrier to successful detection of objects beneath the surface. Moreover, when the Kinect was mounted on the side of the water body using a clear acrylic tank, detection is only successful within 5 cm of the tank wall.

3.1.2 Softkinetic Depthsense

Similar to the Kinect, ripples affected detection when positioned above the water surface and when positioned on a site, the sensing was only near the wall of the tank. Leap Motion Device

3.1.3 Multi-Touch frames

Multi-touch frames are a product designed to provide multi-touch input functionality to a standard LCD Screen, by overlaying above the screen. It uses IR LED and Phototransistor technology to resolve objects. The product usually has a glass to protect the display, but it is not a functional requirement. An experimented was conducted by using a touch frame without its glass positioned around an acrylic water tank. The experiment revealed that the detection is inaccurate and unsuccessful in general.

3.1.4 Camera-based approaches

Web cameras are a low cost yet effective technique for gesture detection in the case of hand or body interaction. However, in the event of foot interactions, camera-based approaches require significant space to set up. For example, when considering hand interaction in a water vessel of 50 cm × 37 cm dimension, the distance from the cameras to the vessel is 64 cm (Ikeda & Hirakawa, 2010). For foot interactions, the dimensions of the water vessel need to increase further. Therefore unlike in (Kimura, Gunawardena, & Hirakawa, 2013) the use of Web cameras mounted on the sides is infeasible. Further capturing foot interactions using a camera based system while allowing free movement to the user demand that the interaction space be elevated or floor to be modified to embed devices. (Sangsuriyachot & Sugimoto, 2012). As such the use of web cameras is infeasible for the domain of the research problem.

3.2 Experiments with existing technologies

Since the devices mentioned in section 3.1 cannot immediately be applied for gesture recognition in water, it is desirable to examine the underlying detection technologies and consider their suitability for gesture detection in water.

3.2.1 Infra-Red Light Emitting Diodes (IR LEDs)

Since IR LEDs have been used in selected gesture detection approaches mentioned in chapter 2, several experiments were conducted to examine the practicality of using the same technology for water interaction. The procedure adopted was to use an IR LED array on one side of a water vessel and examined the output from the opposite side using an IR Sensitive Camera. Two types were considered:

1. Osram Opto SFH 4550 IR LED (850nm wavelength, 700mW/sr intensity 6-degree viewing angle).
2. Osram Opto SFH 4511 IR LED (950nm wavelength, 1200mW/sr intensity 8-degree viewing angle).

The resulting image proved that IR LEDs have two issues that prevent their successful use. One is that unlike in air, IR light is attenuated in water, and, therefore, makes it harder to detect as the distance to travel within water increases. The intensity provided by LEDs was insufficient. Another related issue was that the LEDs do not produce a coherent beam; i.e. the intensity disperses at a short radius around the center. Therefore, it is unsuitable as a point source for a matrix in illuminating objects.

3.2.2 Lasers modules

Standard laser modules produce a point or single dot patterned beam that is coherent. The following products were examined:

1. EGISMOS Red Laser #S836501D-AL01A (650nm, <1mW)
2. Red Laser (650nm, 5mW)
3. AIXIZ IR Laser (780nm, 5mW)

Beams produced from all of these products were successfully detected on the opposite side of the water vessel. Since the bean was coherent, a phototransistor with a matching pass filter was used to evaluate the amount of light received. However, one concern in using lasers was the effect of laser radiation, and accidental exposure to the eye. According to the ANSI Z 136.1 Standard ("American National Standard for Safe Use of Lasers," 2007) Laser#1 to class II and Laser #2 & #3 belongs to class III. While class II lasers are relatively safe to use due to the normal human eyeblink reflex, class III can produce injuries to the eye, and may require additional safety procedures.

3.2.3 Line lasers

The beam divergence when using a line laser is an issue when used in the proposed system, as the resulting silhouette is a function of the distance between the laser and the object. Therefore, calculation of the object size based on the silhouette can be complicated. A solution for this method was suggested in (Kimura et al., 2013) but for a foot interaction space, this setup is infeasible due to space requirements.

3.2.4 Acoustics detection

Low-cost acoustic sensors that use ultrasound have been used for gesture detection successfully (Gupta et al., 2012). However, the distances they operate and the detection angles are not sufficient by itself for our testing environment. As explained in Chapter 2, acoustics is a well know methodology for object detection in oceanography and marine domains. However, this technology by itself is too costly to be applied for detection in a smaller area. Low-cost acoustic sensors for in air detection are available, but most of them have narrow detection angle such as 20—30 degrees and a minimum detection distances in the order of 10—20cm. Further, they only detect the distance to an object: therefore estimating the shape of the object may require multiple sensors operating at different frequencies. Further, reflections created by the sides of the tank collide with each other which complicate object resolving.

3.2.5 Distance sensors

While distance sensors have been used in systems that use water (Yabu, Kamada, Takahashi, Kawarazuka, & Miyata, 2005) they do not come into contact with water and further have not been used to extract positional data, but only for velocity and slope angle.

3.2.6 Floor sensors

While floor sensors can be used to detect gait, the sensor detects pressure, and in an environment where water is present, the circuitry can fail. Therefore to detect objects in an immersive water environment, this approach is infeasible. Further if the object is not touching the bottom of the water vessel, no pressure is detected, and this scheme again becomes unsuitable.

3.2.7 Capacitive sensing

Although capacitive base technologies are used in touch surfaces to extract positional data, applications with water introduce obstacles to evaluating the location. Even Touché (Sato et al., 2012) only facilitated the detection of gestures performed in water and was not designed to provide any information on the 3D space (positional data) in which the gesture is performed.

3.2.8 3D Scanners

3D scanners are capable of providing a very accurate (<1mm) details on an object to be scanned. Yet they are used mainly for static objects, and may take several minutes to complete a single scan. While the resolution is suitable for a very accurate representation of a hand or foot, using it in a real time environment is not practical due to the time taken for a single scan. Further, they infer the points making up the object based either on features which are designed to work in air medium. When the object is immersed in water the scanners may need to be re-calibrated to provide proper measurements. Therefore using a 3D scanner to find the position of an object immersed in water is deemed infeasible.

3.3 System architecture: SensorTank

After carefully analyzing the strengths and weaknesses of using the existing hardware and technology mentioned in sections 3.1 & 3.2 for water based gesture detection, it was decided to use a sensor array. This approach is similar to the light grid approach suggested by Pearson and Weiser for foot movement detection (Pearson & Weiser, 1986). Each detection unit in the array consists of a red dot laser and a phototransistor. Dot lasers provided a coherent source and can be arranged compactly unlike the line lasers. Further attenuation in water is relatively negligible. The phototransistor sensing enabled the tank to be compact and hassle-free use for foot movement, unlike the case if cameras were mounted on the sides. We introduce this tank as SensorTank (Figure 3.1).

The tank was designed for use by a single person. The width and length of the tank provided and obstruction free movement of the foot. The height was intended to accommodate the layers of sensors covering the ankle of the foot up to the *fibula* bone. It was also needed to ensure that the height of the tank should accommodate any water waves that may be generated when moving the foot in the tank. The tank was built using transparent acrylic panels of 1.5 cm thickness and had dimensions of 20 cm × 88.4 cm × 50 cm (H × L × W).

For interacting with water, a visual stimulus was by way of a display. While previous water interactions had used projectors mounted overhead (Ikeda & Hirakawa, 2010; Ikeda et al., 2009; Matoba et al., 2013; Tanabe & Hirakawa, 2006), it was decided to use an LCD monitor (Panasonic Viera TH-L32DT3, 32 Inch) at the bottom of the tank. It had inbuilt stereo speakers that enabled to provide audio stimulus when required.

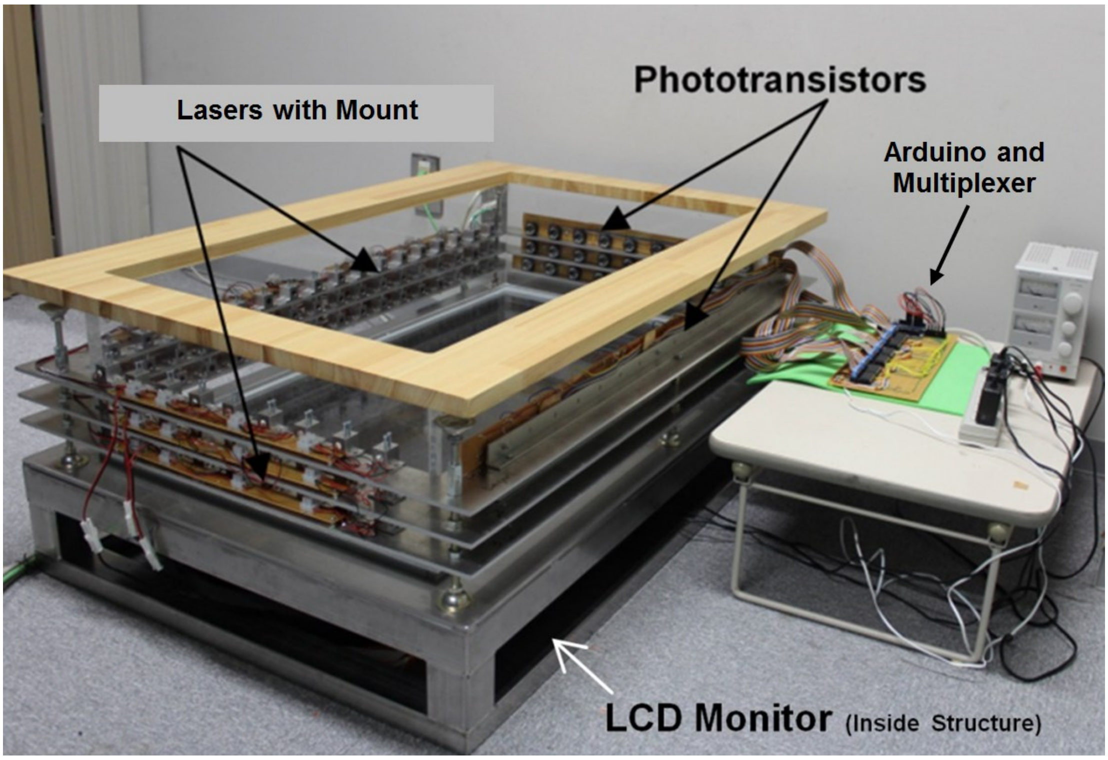


Figure 3.1. SensorTank system

When an object such as a foot is inserted into the tank, one or more laser beams are blocked by the object. This change can be sensed by the associated phototransistors (Figure 3.2). Since the phototransistors can trigger false positives, a load resistor value (10K Ω) that was not triggered by external light sources was selected.

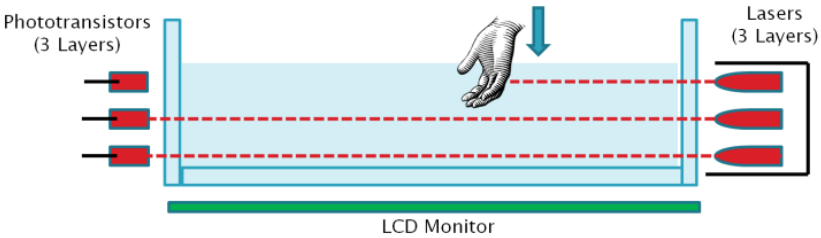


Figure 3.2. Cross-section diagram of SensorTank
from (Gunawardena, Kimura, & Hirakawa, 2014)

78 sensor pairs were arranged in a matrix arrangement at a separation of 5cm horizontal and 3 cm vertical between modules on all four walls of the tank using three mounting layers. The lowest layer was positioned 1.5cm from the bottom of the tank. A single layer consisting of 26 units is illustrated below in figure 3.3.

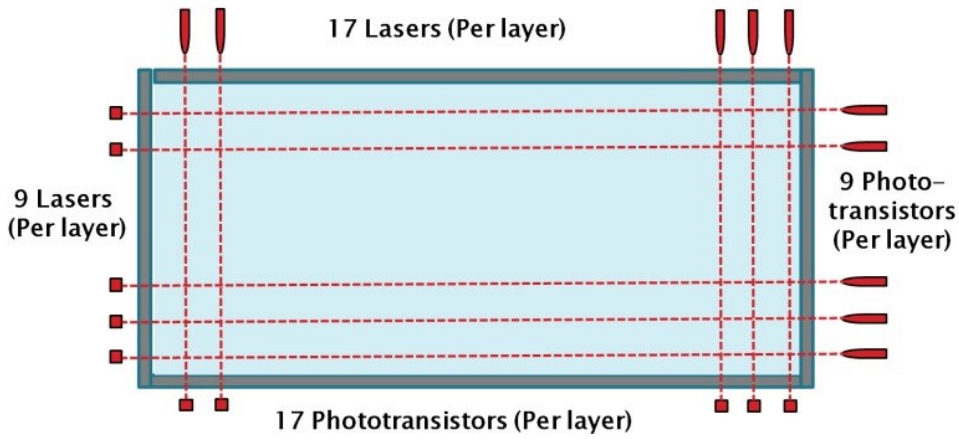


Figure 3.3. Overhead view of SensorTank
from (Gunawardena et al., 2014)

The sensing resolution is rather coarse compared to other existing touch sensing devices as the objective is not to identify the position of a toe tip. The scope defined the application domain that is aimed by the system; i.e. use of a body part, such as a foot or hand with 3D volume for interaction with a computer. The hand and foot anthropometry data (*Japan Body Size Data 1992-1994*, 1994) was considered to determine the laser spacing (points). The lower fifth percentile of the population was deemed to validate our resolution as practiced in ergonomics.

1. The fifth percentile foot length of males and females at nine years of age is 18.5 cm for each; this distance occupies 3 to 4 laser points.
2. The fifth percentile foot breadth of males and females at nine years of age is 7 and 6.9 cm, respectively; these distances occupy 1 to 2 laser points.
3. The fifth percentile hand length of males and females at nine years of age is 12.81 and 12.97 cm, respectively; these distances occupy 2 to 3 laser points.
4. The fifth percentile handbreadth of males and females at nine years of age is 5.36 and 5.39 cm, respectively; these distances occupy 1 to 2 laser points.

Since the horizontal separation between lasers is 5cm, theoretically it is possible that a part of the foot is inserted between two adjacent lasers. They would not be detectable at that instant. However, since the interaction is expected to be dynamic in nature, an object such as a foot would be detected without trouble once they are moved in either direction.

The chosen laser module for the SensorTank prototype was a class II Red Laser (650nm, <1mW). This was chosen for safety reasons, as IR lasers fell into class III and require further safety procedures. The laser was positioned on a mount using two screws. The mount was affixed using two screws to the metal plate that formed the layer. (Figure 3.4)

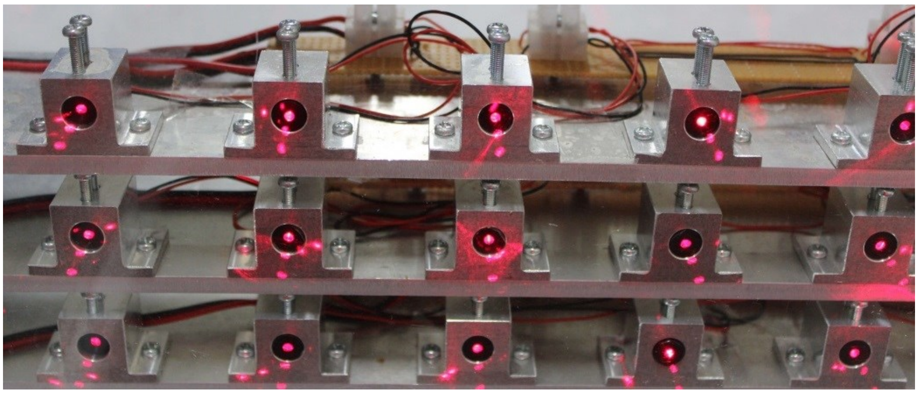


Figure 3.4: Lasers with their mounts
from (Gunawardena et al., 2014)

The phototransistor chosen was Silicon NPN Phototransistor (Vishay BPW77NB). The common collector amplifier configuration is used to connect the phototransistors. They transition from a low state to high in the presence of external illumination. Originally the phototransistors were soldered onto the PCB. (Figure 3.5)

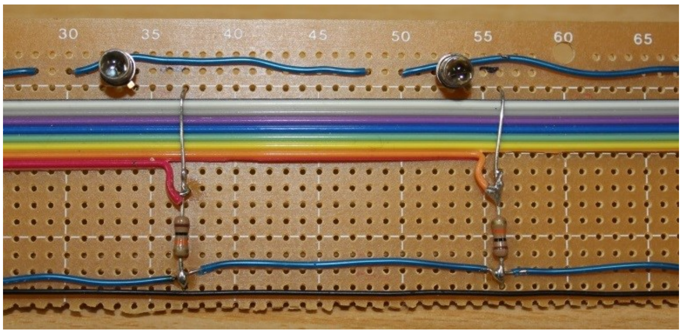


Figure 3.5. Original mounting for phototransistors

Since the beam from each laser module is focused to a very tiny spot (ø=2 mm) and the receptacle area of the phototransistor is almost of equal size (ø =2.54 mm) the positioning has to be of very high precision. However, since the laser is mounted using two screws at the top, and the mount uses two screws on the sides, a slight misalignment is possible. To reduce the error that could be caused by misalignment, a focusing lens (ø =20 mm) is attached in front of each phototransistor, so that even If the laser beam is slightly misaligned, it is focused to the phototransistor (Figure 3.6).

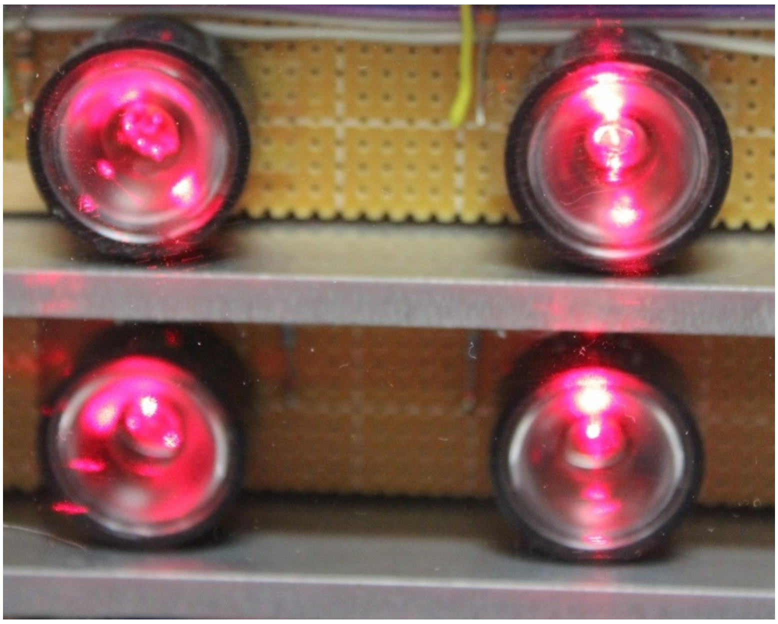


Figure 3.6. Lens to focus laser beam to phototransistor

The beam matrix is not visible under normal circumstances unless an object is immersed in the tank. For illustration purposes, a clouding agent is mixed with the water to illuminate it. (Figure 3.7)

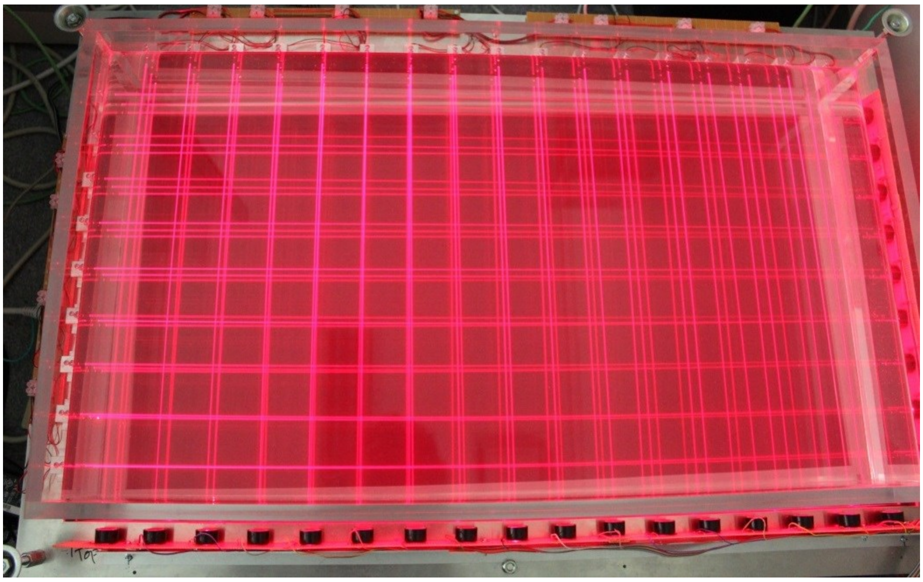


Figure 3.7. Illuminated laser matrix
from (Gunawardena et al., 2014)

3.3.1 Data acquisition hardware configuration

The 28 lasers in each layer are collectively powered by a single 3.3 V DC power supply. The phototransistors on each side of each layer are connected by a ribbon cable which carries 5V DC power as well as the positional data. Therefore, each layer provides two data streams for the length (17 signals) and width (9 signals) of the tank. (Refer Appendix A for circuit diagrams)

The Arduino Uno was selected as the Microcontroller board to stream the data via USB to a Computer. However, since the Arduino had only 14 digital input/output pins and six analog input pins to connect the 78 data sources, a multiplexer was required. A 16-Channel Analog Multiplexer / De-multiplexer (74HC4067) Integrated Circuit was used for this purpose. In total six multiplexer ICs were used (The PCB board in figure 3.8 displays the complete circuit with two extra IC sockets).

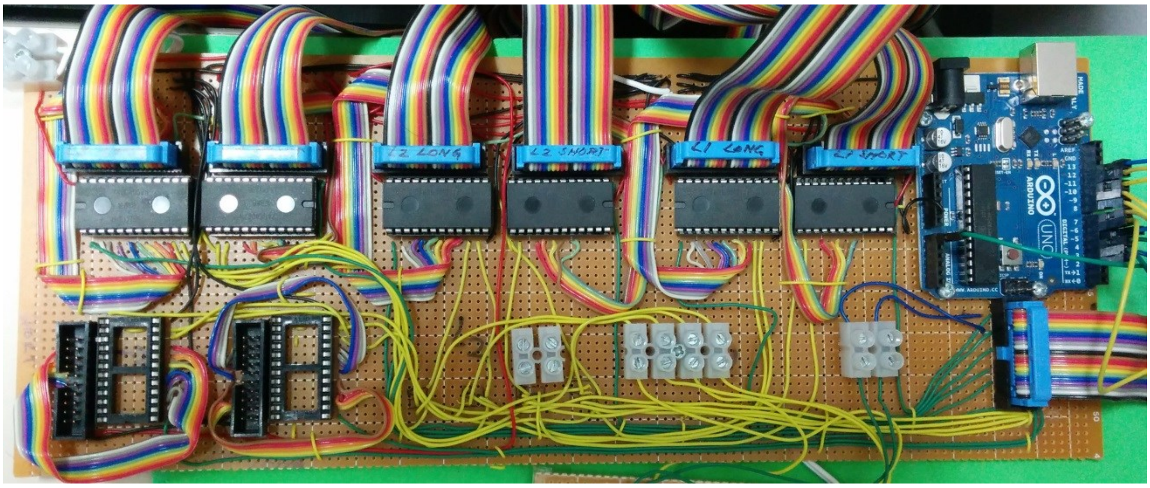


Figure 3.8. Multiplexer board with Arduino

The Arduino has its proprietary language, and the coding was performed to read all 78 inputs while switching multiplexers.

3.3.2 Data acquisition algorithm

Once the data from a sensor is read, it is stored in a 2D Integer array. The subscripts of the array are the data about the sensor numbers in the x-axis (length) and y-axis (width). In each array position, the values written into the array correspond to a weight assigned to the layer, which refers to the depth of the tank. The values that represent layer 1, 2 and 3 are marked as 1, 2 or 4 respectively.

For example, if the position corresponding to sensor five on the length and sensor nine on the width has an object on layer 3, the value stored at [4, 8] would be 4 (note that the array subscript starts from zero, so element 5 is addressed by subscript 5). However each element in the 2D array is really storing three values about the three layers, and as such during a single data acquisition cycle, each array element is added data up to three times depending on the depth of the object. Therefore, If one considers the same situation as in the previous example, If the position corresponding to sensor 5 (length), sensor 9 (width) has objects in all three layers, the final value stored at [4, 8] would be 1+2+4 = 7.

Finally, the data sent via the Arduino contains the only the x, y coordinate where an object is present and the corresponding depth. This is performed to speed up the data transfer process that is carried out at 115200 bps.

3.3.3 Pilot proof of concept visualization and demonstration

A simple visualization of the object immersed in the tank was completed by mapping the depth to a circular graphic with the radius and color changing according to the depth using a the processing language program. (Figure 3.9) This system was demonstrated at a Science Exhibition held at Kunibiki Messe, Matsue City, Shimane Prefecture, Japan on 15th March 2012. (Figure 3.10)

Figure 3.9. Proof of concept visualization
from (Gunawardena, Kimura, & Hirakawa, 2012)

Figure 3.10. Demonstration of system

One of the issues identified during this initial experimenting with the system was the behavior exhibited by the system when two or more objects were immersed simultaneously. This issue, which is a discussion topic in certain multi-touch screen technologies, is referred to as ghosting.

3.4 Ghosting and ghost cancellation

In its prototype version - 1 form, SensorTank can detect a single object position accurately, and if two objects such as two feet are inserted in the same row position or same column position.

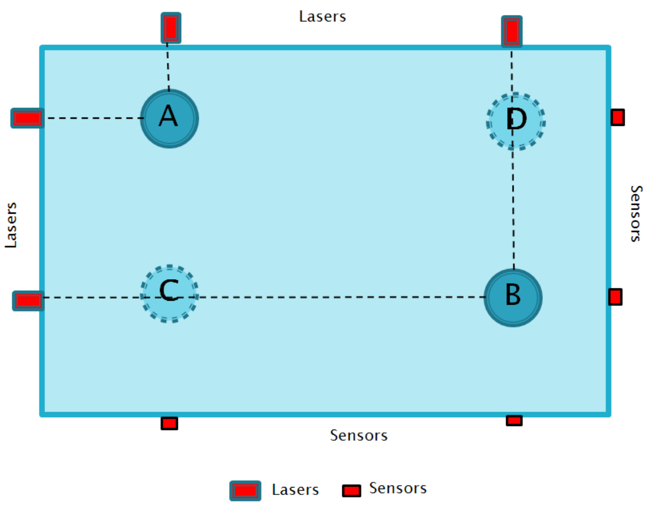


Figure 3.11. Ghost object appearance (overhead view)

In figure 3.11, objects A and B are real objects, and are positioned at two random locations where they do not share a common x, or y coordinate. During detection, the detection process detects ghost objects C & D due to occlusion. This phenomenon is commonly referred to as ghosting.

3.4.1 Ghost cancellation using additional laser-phototransistor layer

One approach to resolving this issue is by introducing a new laser-phototransistor layer. Unlike the previous layers, the lasers in this layer emit beams at a 45-degree angle to the regular layer laser beams (Figure 3.12). This layer has 25 lasers-phototransistor pairs and is positioned above the acrylic tank. If this layer were to be placed around the acrylic tank similar to the regular layers, laser light may partly reflect out of the tank due to reflection and further may not escape the tank due to refraction.

Figure 3.12. Ghost cancellation layer

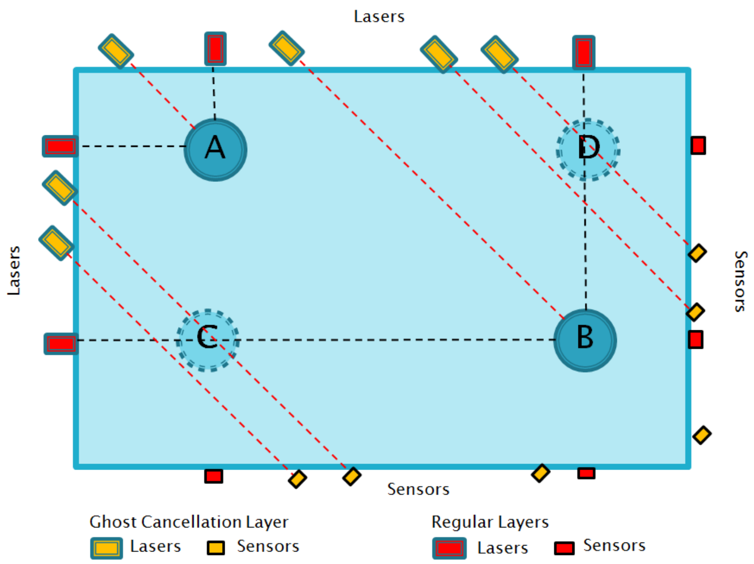


Figure 3.13. Ghost cancellation layer operation from (Gunawardena et al., 2014)

The diagonal layer is positioned in such a way that the each intersection point in the regular layers is connected with a particular diagonal laser beam of the new layer. In this scenario, when the tank detects more than two objects in its original scheme, the corresponding locations status is checked using the diagonal layer. For the example in figure 3.13, the diagonal layer check will reveal that there are no objects at locations C and D and that the real objects are at location A and B. This elimination process is carried out at the Arduino microcontroller itself.

While this new layer is useful for eliminating ghosting, it uses the same technique as the existing laser-phototransistor layers, and, therefore, it too can introduce ghost points. Although this is suitable for clearing ghosting in two objects, if the number of objects increases, the possibility of ghosting exists. Another issue is that the new layer is positioned above the water surface, so when detecting ghost points, it has to consider positions in a short radius for ghost elimination. This situation arises if the objects are positioned in a slanted orientation.

3.4.2 Ghost cancellation using touch frame

In Section 3.1.3, an experiment with multi-touch frames as a direct method for detecting positions in water was carried out. While this was unsuccessful, the touch frame mounted above the water surface over the acrylic tank provides beneficial to eliminate ghosting (Figure 3.14). The PQ Labs range of touch frames, which were used for our experiments, has their proprietary method of eliminating ghosting.

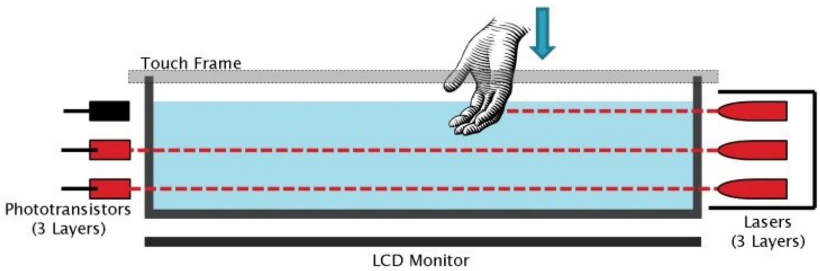


Figure 3.14. GestureTank system side view

In this scenario too, if the regular layers detect more than two points, the algorithm used to eliminate ghost points is executed. If we consider the situation where two objects inserted into water, as illustrated in figure 3.15.

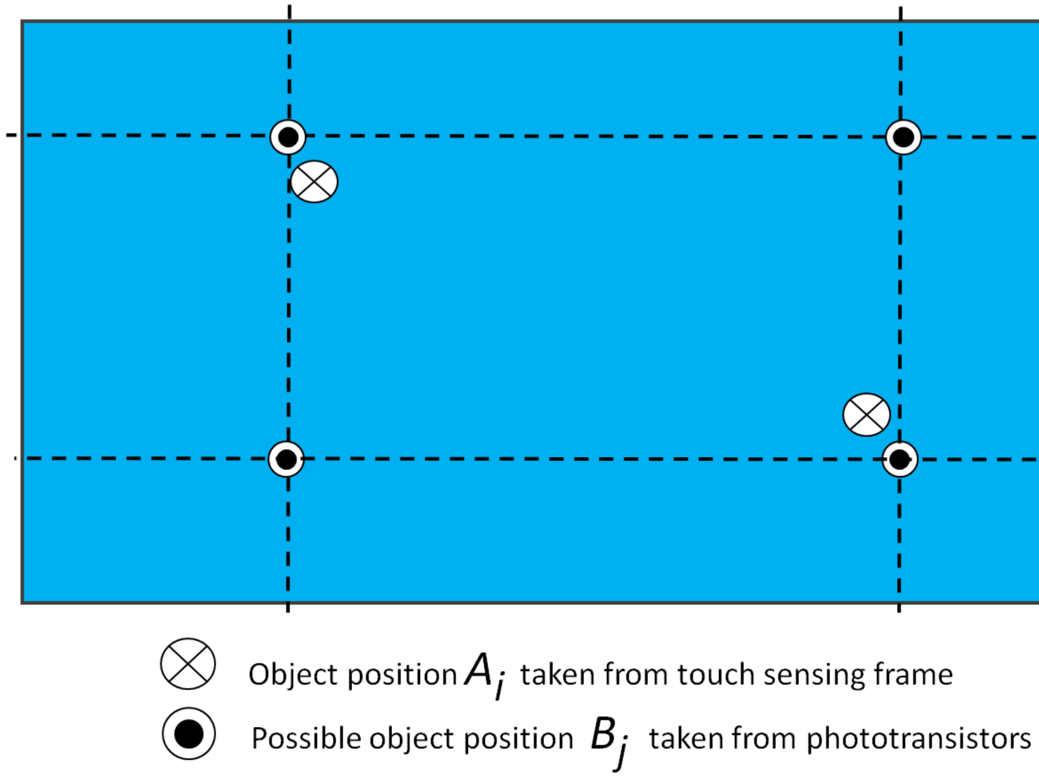


Figure 3.15. Ghost point resolution using a touch frame (from (Gunawardena & Hirakawa, 2015))

The regular layers detect four positions (B_1-B_4) The touch frame provides two positional data (A_1 and A_2). The positions detected using the two methods may not be identical since they could be inserted in a slanting position.

For every point A_i ($1 \leq i \leq 2$) detected by the touch sensing frame, denoted by a round symbol with a cross mark in the figure, the system calculates the distance from the weight centre B_j ($1 \leq j \leq 4$) of each possible object region, indicated by a black-centred round symbol, and then identifies the one having the shortest distance from A_i as the actual object region to be associated with it. Unlike the ghost cancellation layer operation, in this approach the ghost cancellation is performed not at the Arduino level, but at the Personal computer to which the touch frame and Arduino are connected. This improved method is the basis for the next version of SensorTank, which is named as GestureTank.

3.5 System architecture: GestureTank

In GestureTank, using the touch frame meant that the system was slimmer than using the ghost cancellation layer introduced in section 3.4.1. Apart from that, in this version, the sensing technique and data acquisition method has had no significant change from the previous SensorTank. Figure 3.16 introduces the general appearance of GestureTank. In addition to the main components, a water faucet, temperature sensor and heating element is connected via secondary Arduino microcontroller to the main program (Figure 3.17). These ancillary components are used only for an application scenario, and their detailed description is provided in chapter 4.

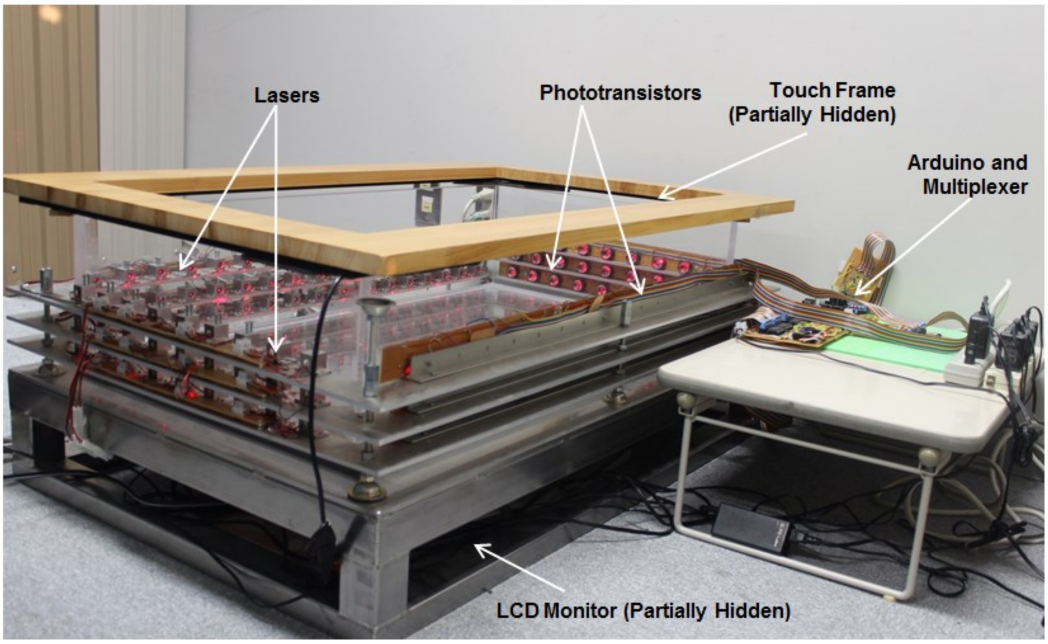


Figure 3.16. GestureTank system

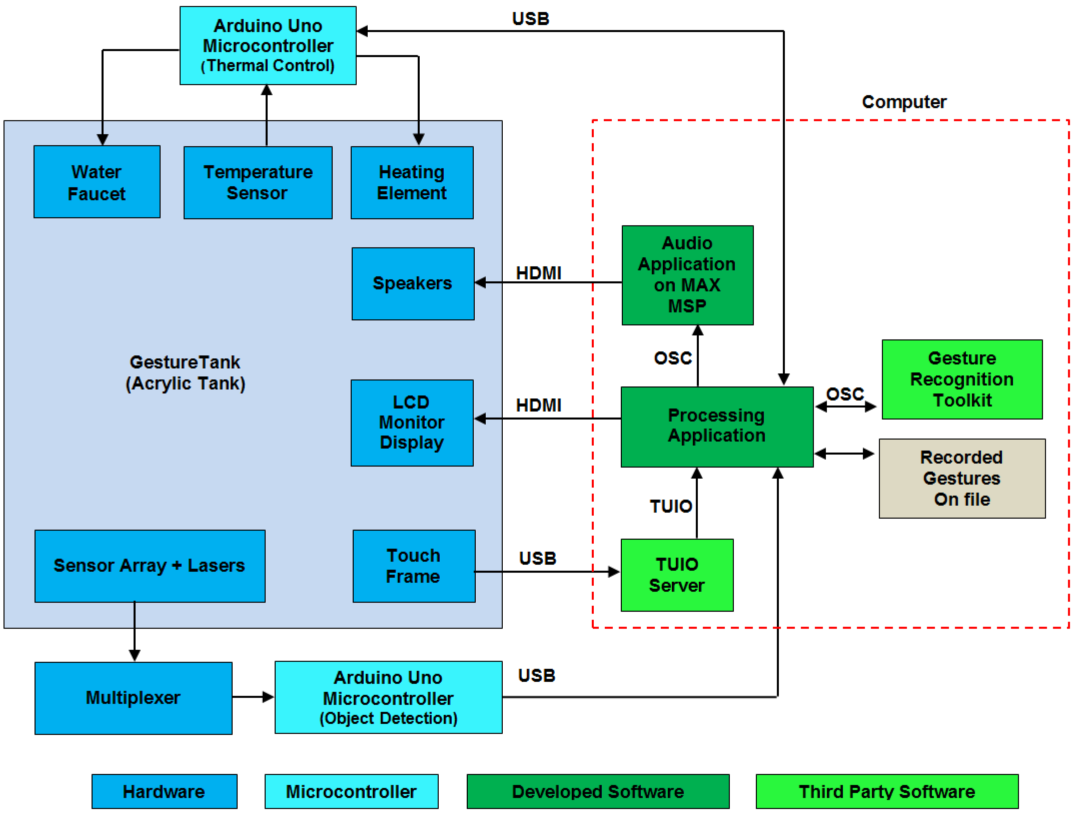


Figure 3.17. GestureTank system architecture

The PQ Labs G3 40 Inch touch frame is connected via USB to the Computer. The touch point data from the frame is received by the master application via the TUIO protocol. The HDMI Interface is used for LCD monitor connectivity and audio output. The remaining software components are explained in detail in the next chapter.

3.6 Summary

The first topics covered in this section discussed the experiments carried out with existing devices and existing technologies that have been cited in previous research. On gesture detection in water, most of them did not provide satisfactory object detection. Therefore based on some of the strong points of the technologies, a new architecture was presented. SensorTank was the first version developed and during testing it was discovered that an effect of ghosting was observed in multiple object detection. Version two: GestureTank was built to overcome this issue.

Chapter 4

Gesture detection – software approach

This chapter explains the algorithms and software components created as well as interconnected with as part of the research study, to detect gestures in the water vessel.

4.1 Software architecture

The main application was written in processing language. A block diagram of the steps carried out within the processing program is displayed in figure 4.1. Note that the modules marked in grey were programmed using Processing language.

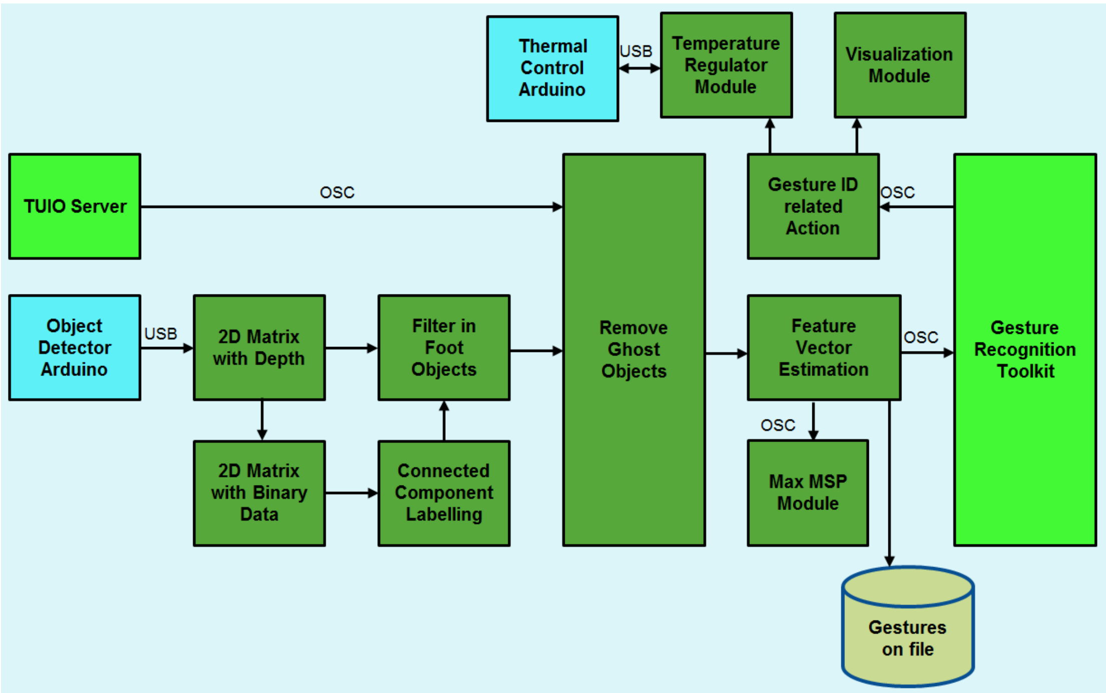


Figure 4.1. Algorithmic architecture for gesture detection

The main input data for this program arrived from the three laser-phototransistor layers via the multiplexed Arduino output. The first step involved is pre-processing this data, to eliminate any object that does not fit the form of a foot. Generally although random noise was not an issue, filtering out objects was required to increase the accuracy of recognition later. The original data received was stored in a 2D Matrix with Depth similar to the process described in section 3.3.2. However, to identify connected points, the connected component labelling algorithm (blob detection algorithm) is run on a 2D matrix that replaces the depth information with 1 (i.e. Matrix with Binary data). Once this process is complete it is possible to filter out any blob that does not fit the form of a foot. The output of this process creates a 2D matrix (with depth information) for each foot shaped object.

The second step involved is to remove ghost objects. The ghost cancellation algorithm mentioned in section 3.4.2 is performed. The output from this module contains a point cloud pertaining to a foot stored in the form of a 2D Matrix.

4.1.1 Feature selection

The next step is to extract feature vectors from the matrix of points. To avoid the curse of dimensionality, we need to select the optimum features that can be used to identify gestures correctly. Basic features that can be considered include the length of the foot and width of the foot. However, since the foot movement gestures considered are three dimensional – i.e. involves movement both sideways and up-down, we experimented with a number of features.

Finally, five features are selected as being suitable for gesture analysis: PointsinL3Front, PointsinL3Back, L1L2Right, L1L2Left, and CheckLR. PointsinL3Front is the percentage of the number of foot regions in the topmost row (L3Front) to the total number of foot regions in Layer 3 (NL_3). PointsinL3Back is the percentage of the number of foot regions in the bottommost row (L3Back) to NL_3. L1L2Right is the percentage of the total number of foot regions in the rightmost column in Layer 1 (L1Right) and Layer 2 (L2Right), compared to the total number of foot regions in all three layers (N). L1L2Left is the percentage of the number of foot regions in the leftmost column in Layer 1 (L1Left) and Layer 2 (L2Left). CheckLR provides a score of either 1, 0 or -1 depending on the balance of the foot calculated by considering the summation of all foot regions in the leftmost and rightmost columns, where a negative value indicates that the balance is tilted to the left. Figure 4.2 shows the overhead and side view of a foot with the selected features highlighted. A graph showing the features themselves is given in figure 4.3.

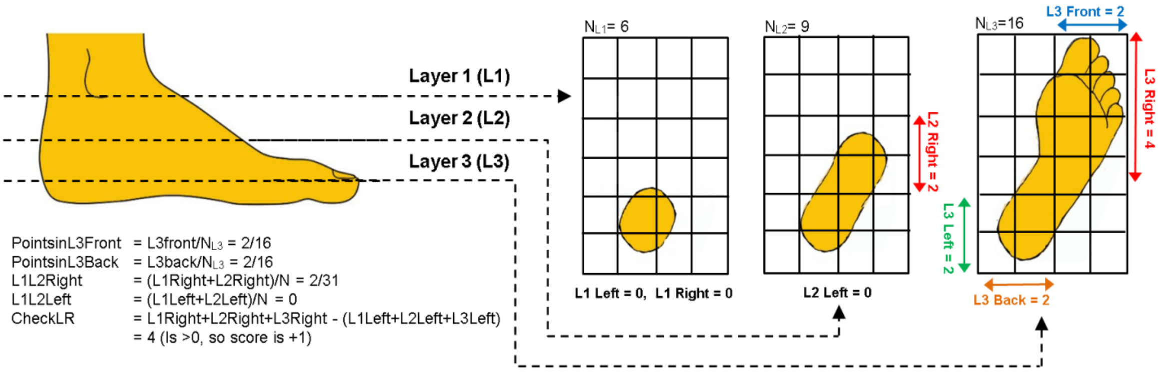


Figure 4.4. Illustration of selected feature vectors

(From (Gunawardena & Hirakawa, 2015))

Figure 4.3. Plot of the selected feature vectors

(From (Gunawardena & Hirakawa, 2015))

The next step was to send the feature vectors to the gesture recognition toolkit for training and testing purposes. Apart from this facility, the feature vectors were stored in a log file for future reference. This file also permitted storing the class of the gesture (Gesture ID, See section 5.2) when recording the training dataset.

4.2 Gesture detection process

For training and testing gestures, we selected the Gesture Recognition Toolkit (GRT) as a feasible option. Some reasons contributed to this. One was that the fact that GRT supported a large number of algorithms, which could be quickly evaluated. Another was the facility provided for real-time as well as offline analysis of gestures. The main processing program sent the features pertaining to a foot via OSC protocol to the GRT, which is listening on port 5000. GRT Version 0.1.14 was the selected version.

Our gesture recognition problem was a classification problem (See Section 5.2 & 5.3 for details of classes). For our analysis in section 5.3, we used the offline facility whereby we provided a testing data set and a training dataset any compared the classification performance. For Real-time classification, the GRT was trained using a selected classification algorithm using the training dataset and used for prediction, where the predicted class and probabilities were returned via OSC to the processing application for visualization and any other form of response.

4.3 Applications

SensorTank and GestureTank can be used in different environments that deal with water. Most of the prototypes are built around a foot bath (Ashiyu) activity, although they need not be always for foot interaction. Ashiyu is a public foot bath, which has a long cultural history in Japan. Since they are often free, it is a public space where people would come and relax while talking to each other. While the foot bath provides a relaxing, calming sense to the body, the interactive capabilities provided by our system can provide the mind a calming atmosphere. We introduced here several scenarios that have been tested as well as few proposed.

4.3.1 Visually stimulating patterns with soft music

In this application, using the feet or hands, the user paddles the water. The system calculates the centroid of the 3D Object immersed. A glowing pattern that changes color and spread radius depending on the depth of the centroid and area of the object is displayed (Figure 4.4). Further, the location (x, y, z) of the centroid is used to play a soothing musical tone.

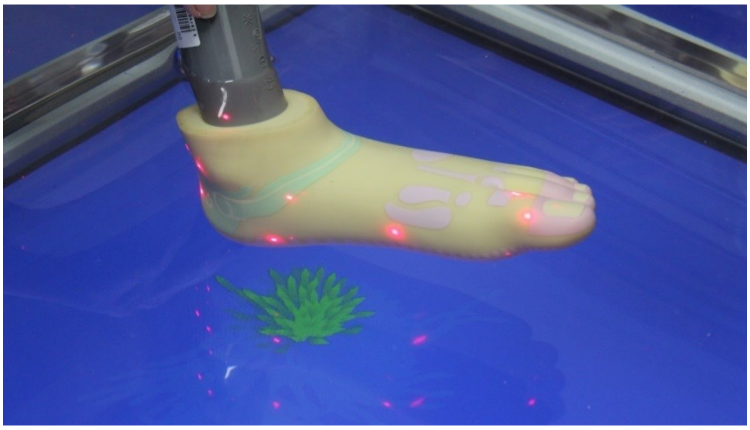


Figure 4.5. Glowing pattern following foot

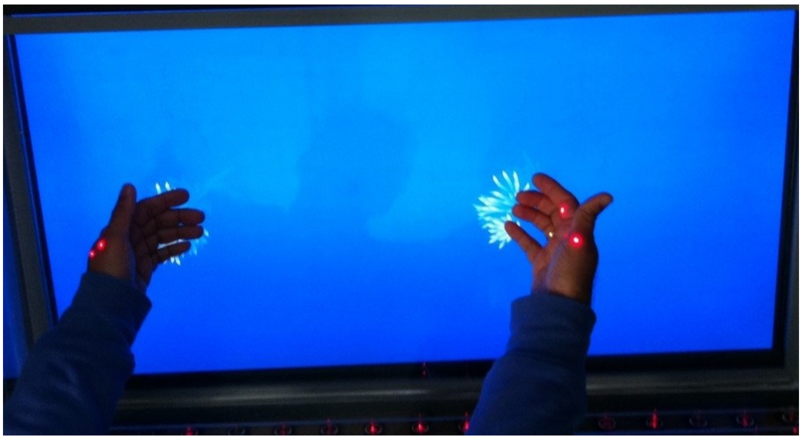


Figure 4.6. Glowing pattern following the two hands

The visualization is developed in processing language and supports both hands (Figure 4.5). Musical tones generated in MIDI are based on the Centroid Coordinate. This application is developed using MAX/MSP, which is a visual programming language. In this case, the depth (z) is mapped to the duration of the MIDI tone and x, y coordinates are mapped to MIDI pitch and MIDI velocity respectively.

4.3.2 Fish following the foot

In this demonstration, the system calculates the centroid of an immersed foot and displays a fish so that it follows the foot movements, as shown in figure 4.6. The fish changes its direction and depth position depending on the inserted foot position. This animation uses Adobe flash. Further, the location (x, y and z) of the centroid is used to play a soothing musical tone using the same technique mentioned in section 4.3.1.

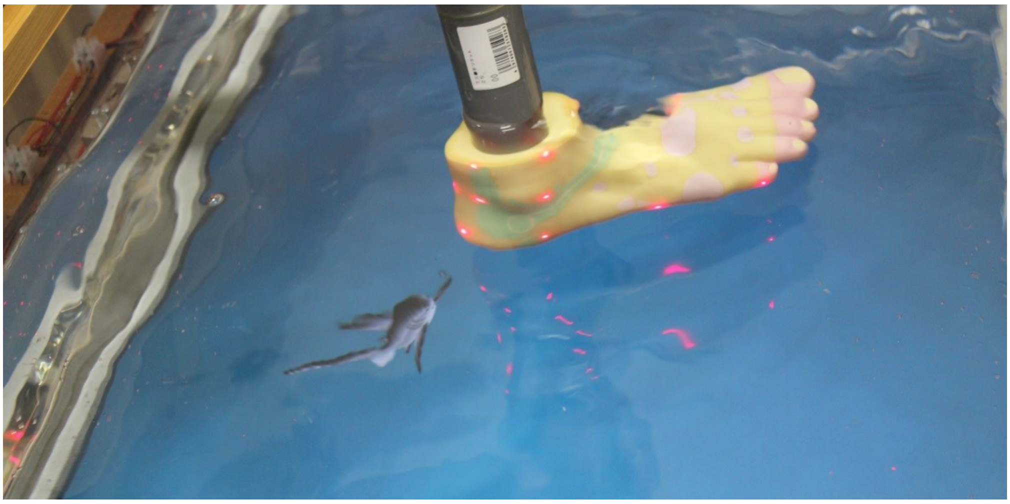


Figure 4.7. Fish following the foot

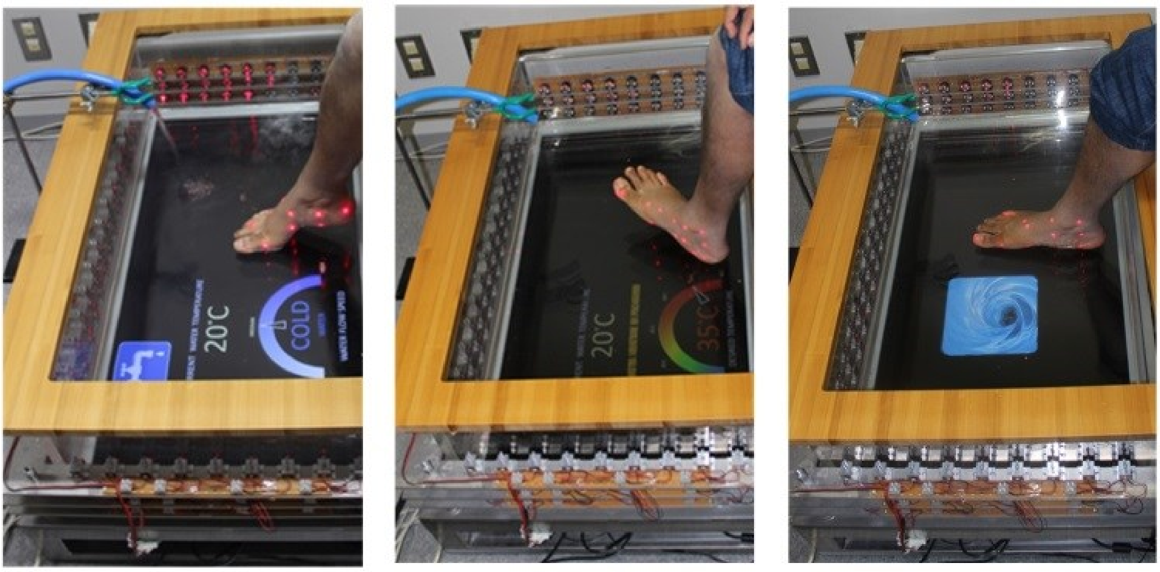
from (Gunawardena et al., 2014)

4.3.3 Bathtub operations

When using a bathtub, a number of operations are carried out; this is an ideal situation to realize the full extent of gestures in water. In this scenario, faucets, draining and temperature control can be facilitated by our system. The foot gestures used in this section are illustrated in section 5.3.

The water inlet (faucet) can be controlled using two foot gestures: The raised-heel gesture with the foot facing forward sends cold water to the tub (Figure 4.7(a)). Once the foot is brought back to the resting position (foot-resting gesture), the faucet is closed. Warm water can be sent into the tub by utilizing the raised toes gesture with the foot facing forward. For this demonstration, a water inlet solenoid valve is the ideal controlling element to be connected to our system. However, in this pilot demo we have used a small aquarium water pump. This water pump is controlled using an electrical relay connected to the Arduino for this purpose. If a solenoid valve were used, it would have been possible to control the speed of the water flow, which could be set to indicate the degree of foot tilt.

In the same system, it is possible to control the temperature of the tub, for which we use a small electric water heater used for aquariums. The temperature can be increased by moving the raised toes foot to the right (Figure 4.7(b)) to match a figure displayed on Screen (such as 35°C). This would trigger a relay connected to the heater to be switched on. A separate temperature sensing thermistor (Ishizuka Electronics 103AT-11, 10KΩ thermistor) connected to the tub provides temperature readings to the Arduino. When the temperature in the tub matches the user set temperature, the relay is switched off. Similarly, the temperature can be lowered by moving the foot to the left. However, we did not implement any technique to cool the tub in this study. Finally, to drain the tank it would be possible to operate a valve triggered by maintaining the foot-resting position for ten seconds (Figure 4.7(c)), although this is not implemented in our prototype model.



(a) Faucet sending out Cold Water (raised heel gesture)

(b) Increasing temperature of the tank with heating element (raised toes, foot move to right)

(c) Draining the tank (Simulation only, not implemented in hardware)

Figure 4.8. Gestures used in the operation of a bathtub from (Gunawardena & Hirakawa, 2015)

4.3.4 Bathtub/footbath music player

Apart from controlling the operations in a bathtub, entertainment for relaxation is also a possibility. Listening to music or music video is suited for this type of environment or even for a foot bath. We propose a foot gesture controlled music player to be used in such an environment.

The raising of the toes with the foot facing forward can be set to indicate starting play. The raised-heel gesture can be configured to indicate pausing play. Movement of the foot to the right and left with the foot touching the surface can be configured to indicate skipping a track forward and backwards, respectively. Similarly, foot movement to the right and left with raised toes can be configured to indicate increasing and decreasing volume, respectively. Finally, resting the foot in position for ten seconds can be configured to indicate stopping play.

4.3.5 Applications in medical field

We further propose that system is used in the medical field. Patients who are undergoing rehabilitation after an injury to feet or arms need to carry out movements gradually. Performing such actions in water (hydrotherapy) or liquids like paraffin is one approach currently used. One reason for this is the hydrostatic pressure provided by a liquid helps to reduce swelling. Patients with knee pain due to illnesses like arthritis may also benefit from this approach. Chinn & Hertel emphasise that for lateral ankle sprains suffered by athletes, regaining regains full motion, strength and neuromuscular coordination is important. They further explain that actions such as dorsiflexion and plantar flexion motion in a controlled environment are required for rehabilitation and that Hydrotherapy is a recommended course of action to improve the range of motion in such situations. (Chinn & Hertel, 2010)

One way we propose this can be done is by introducing a target symbol on the display at the bottom of the tank. The patient has to the move his/her hand or foot in the appropriate direction. The speed of the target symbol can be changed according to the state of rehabilitation, and it would be possible using the system to estimate the speed and direction of movement performed by the patient. A performance score could then be calculated and presented so that the patient is motivated. Further, this can be used by the medical professional monitoring the patient.

Another advantage is that if required, the temperature of the liquid can be changed so that it will be more comfortable to use. Further, replacing water with a higher viscosity liquid this could add more resistance to the movement that may be beneficial to those who seek gradual movement of the limbs.

4.4 Summary

This chapter explains the software aspects of the research, starting with the software development carried out for the data acquired from sensors. To apply machine learning techniques, features vectors were chosen, to enable feature selection. A few prototype applications that used the detected gestures, as well as other system features, were also explained in the same chapter. The application scenarios provide evidence of the usage in different domains that the system can be beneficial.

Chapter 5

Experiments and results

This chapter describes the tests performed to prove the validity of the system, and the metrics used to evaluate its performance.

5.1 Robustness of gesture detection hardware technique

5.1.1 Effect of murky water on the system

As the system is designed with foot gestures in mind, and taking to account that our feet may collect a considerable amount of dust and particles, it is inevitable that the particles may find their way into the water vessel. This is especially the case of a public space such as a foot bath. While sand or other heavier particles sink to the bottom of the tank, it had no significant effect on the detection since the detection plane was 1.5 cm above the bottom of the tank. The only effect if could have was to disturb the display viewer ability, but under experimental conditions, it was not significant.

Therefore, an important consideration should be how the system performs in the cloudy (i.e. murky) water. Water can get murky when it contains suspended solid particles, which are too light and small to settle down rapidly in the bottom of the tank. To simulate murky water, a commercial bath salt was used to cloud the tank, and an experiment was conducted to evaluate the gesture detection performance.

Turbidity is measured by a nephelometer, and the results are given in NTUs (Nephelometric Turbidity Units). Starting with clear water (0 NTU), we gradually added bath salts to the water and measured the number of phototransistors that provided false positives. A false positive response occurs when the laser beam is sufficiently blocked by murky water so that it does not trigger the phototransistor on. We considered the phototransistors positioned across the length and the width of the tank as two datasets, as the distance travelled by the laser was different depending on the direction. Dataset 1 contained 27 values (Distance travelled was 88.4 cm), and Dataset II contained 51 values (Distance travelled was 50cm). Figure 5.1 shows the experimental results.

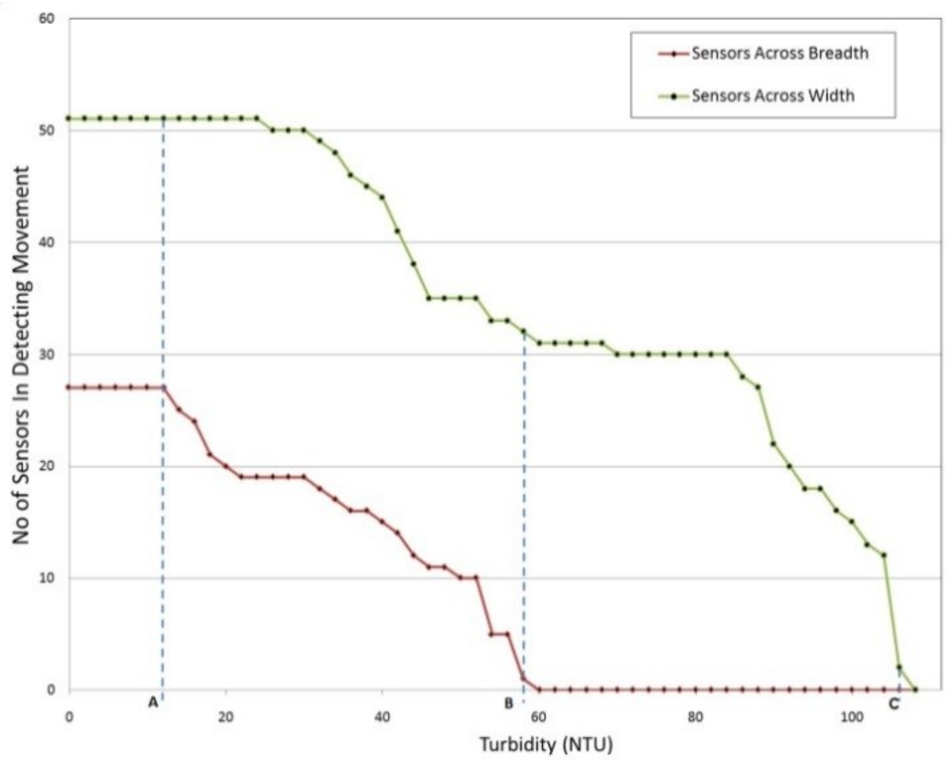


Figure 5.1. Detection threshold for phototransistors in murky water from (Gunawardena et al., 2014)

Until the turbidity reached 12 NTU (defined as state A), all sensors performed as expected. The first false positive detection was observed at 13 NTU. I.e. the system would detect that an object was placed between the laser and the phototransistor, as sufficient laser illumination was not received by one phototransistor across the length. At 58 NTU (defined as state B), all sensors except one across the length were providing false positives. At 60 NTU all phototransistors across the length failed. At 106 NTU (defined as state C), only one sensor across the width was functioning as expected while the remaining 77 sensors provided false positives.

Since the lasers were of the same power (<1mW) and the distance travelled was the same within each dataset, the expectation was that all of the phototransistors on each side to give false positives at a particular turbidity. However, the variations in the graph can be attributed to possible minute alignment differences between the laser and phototransistor, as well as uneven distribution of the clouding agent. Nevertheless, it is possible to assume that if we eliminate the variations, our system should be able to detect objects successfully up to a turbidity level of 58 NTU.

The turbidity of states A, B, and C are visually exhibited in figure 5.2. To obtain these images, visually a camera was placed on the outer side of the tank on its longer dimension, while and a hand was positioned at 1 cm, 43 cm, and 85 cm distances from the camera as illustrated in figure 5.3.

(a) Turbidity at state A (12 NTU)

(b) Turbidity at state B (56 NTU)

(c) Turbidity at state C (106 NTU)

Figure 5.2. Turbidity at three reference points
from (Gunawardena et al., 2014)

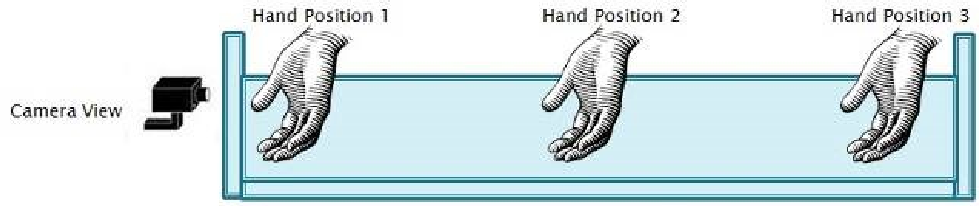


Figure 5.3. Reference points for visual comparison
from (Gunawardena et al., 2014)

Another concern is that apart from the turbidity increase affecting detection it also could also influence the visibility of the display. Figure 5.4 shows the states A & B together with the default state of 0 NTU.

(a) Display at Turbidity 0 NTU

(b) Display at Turbidity 12 NTU (State A)

(c) Display at Turbidity 56 NTU (State B)

Figure 5.4 Clarity of the displays at different turbidity levels from (Gunawardena et al., 2014)

At state A the display visibility is not affected significantly to disturb following of the visuals Even when reaching state B, the displayed image is still visible. However, for better visibility, the display monitor may be replaced with an overhead projector as used in (Matoba et al., 2013)

5.1.2 Effect of ripples and bubbles on the system

Ripples occur when gestures are performed in the tank. An experiment was conducted to examine the generation of ripples and bubbles. A visual inspection was carried out to examine the highest waves created during the foot movements indicated in section 5.2, and whether the system detected false positives or false negatives. The no. of ripples had no bearing on the detection. The waves created generated troughs and peaks of up to a maximum of 2.5cm. However, since the standard water level was maintained at a height of 145mm, and the height of the topmost laser layer from the bottom of the tank was 10.5 cm, we did not observe any detection failures. As long as the laser-phototransistor path was not exposed to air to create total internal reflections and refractions into air detection was unaffected. Bubbling was not a significant factor for the gestures experimented either.

5.1.3 Effect of temperature on the system

An experiment was carried out to evaluate whether the system produced false positives or false negatives while the temperature of the water in the vessel was varied between 10°C and 40°C. The results revealed that there was no effect on the sensing performance.

5.1.4 Effect of lighting level on the system

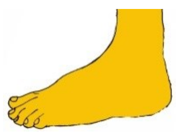
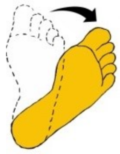
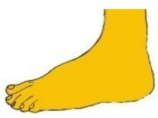
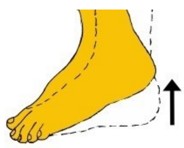
The experiment was carried out in a laboratory environment under ambient lighting conditions, and the system was not exposed to direct sunlight. In the case of a lab experiment, the light falling on the detection area was tested across light levels of 1 lux to 270 lux (Lumens), with the detector being placed at the surface level of the detection tank. The detector used was a Google Nexus 5 smartphone light sensor using CPU-Z software (Android 5.1.1). If the effect of sunlight is significant, it is possible to implement a filter in front of the phototransistors to permit only the frequency range of the lasers to negate the influence of environmental lighting conditions.

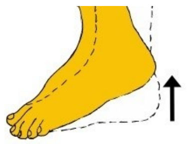
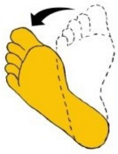
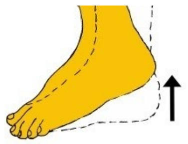
5.2 Gesture usability testing

In chapter 2, some foot gestures that have been documented in research were illustrated. However, the movements documented in the images are 2D motions. Because long-term use of body gestures may cause fatigue (Yee, 2009) and that there was no previous water interaction using foot research found during the literature survey, we conducted an experiment to investigate the usability of a set of proposed gestures based on previous studies. Firstly we prepared a set of 11 gestures (Gesture IDs) by using previously identified 2D gestures as well as combining two 2D gestures.

Table 5.1. List of possible gestures investigated

(from (Gunawardena & Hirakawa, 2015))

Gesture ID	Perspective view of gesture	
1		
	Resting foot, Facing forward	
2		
	Resting foot, Pivot at heel, Move clockwise	
3		
	Resting foot, Pivot at heel, Move anti-clockwise	
4		
	Raised toes, Facing forward	
5		
	Raised heel, Facing forward	
6		
	Raised toes, Pivot at heel, Move clockwise	

7		
	Raised heel, Pivot at toes, Move anti-clockwise	
8		
	Raised toes, Pivot at heel, Move anti-clockwise	
9		
	Raised heel, Pivot at toes, Move clockwise	
10		
	Moving foot in a way that the sole faces outwards / Foot moved outwards horizontally (Eversion)	
11		
	Moving foot in a way that the sole faces inwards / Foot moved inwards horizontally (Inversion)	

Next the comfortability of performing these gestures in our environment was evaluated. The evaluation instrument had two responses per gesture. We also recorded the foot anthropometry data for each subject's feet.

17 participants (4 female, 13 male between 20 – 60 years) participated in the test. They were asked to place a foot in the tank and perform the gestures listed in Table 5.1 twice in a sequence of their preference and comment on the whether they considered it to be a natural and comfortable gesture in a water environment. The Gesture ID was observed, and the response (Yes/No) was recorded. At the time of performing the gestures, they were assisted by the on the screen display at the bottom of the tank guiding on how the gesture selected should be performed. The results are recorded in table 5.2.

Table 5.2. Results of pilot user testing

Subject	Gender	Foot Anthropology (cm)	Agreement Y/N of whether the gesture is suitable (ID)										
			1	2	3	4	5	6	7	8	9	10	11
A	M	26.4	Y	Y	Y	Y	N	Y	N	N	N	N	N
B	M	24	Y	Y	Y	Y	Y	Y	N	Y	N	N	N
C	M	26.5	Y	Y	Y	Y	Y	Y	N	N	N	N	N
D	M	24.8	Y	Y	Y	Y	N	Y	N	N	Y	N	N
E	M	23.6	Y	Y	Y	Y	N	Y	N	Y	N	N	N
F	M	24.1	Y	Y	Y	Y	Y	Y	N	N	Y	N	N
G	M	24.5	Y	Y	Y	Y	Y	N	N	Y	N	N	N
H	M	25.9	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	N
I	F	23.5	Y	Y	Y	Y	N	Y	N	Y	N	N	N
J	F	22	Y	Y	Y	Y	Y	Y	N	Y	N	N	N
K	M	22	Y	Y	Y	Y	Y	Y	N	Y	N	N	N
L	M	26	Y	Y	Y	Y	Y	Y	N	Y	N	N	N
M	M	27	Y	Y	Y	Y	Y	Y	N	Y	N	N	N
N	M	25	Y	Y	N	Y	Y	Y	N	N	Y	N	N
O	M	25.5	Y	Y	Y	Y	Y	Y	N	Y	N	N	N
P	F	22.5	Y	Y	Y	Y	Y	Y	N	Y	N	N	N
Q	F	25.5	Y	Y	Y	Y	Y	Y	N	Y	N	N	N
Total			17	17	16	17	13	16	1	12	4	0	0
Percentage (Y)			100%	100%	94%	100%	76%	94%	7%	71%	24%	0%	0%

We considered a threshold of 70% as a reasonable comfort percentage. Accordingly seven (gesture IDs: 1, 2, 3, 4, 5, 6, and 8) of the 11 gestures can be considered as suitable for further experimentation with our system.

5.3 Gesture recognition test using machine learning

5.3.1 Experimental setup

In this experiment, the features chosen are examined by way of machine learning algorithms. The first step is to record gestures about the seven gesture IDs. This experiment too was conducted in the laboratory environment with 17 participants (4 female and 13 male, between 20–60 years and foot anthropometry range of 22–27 cm). Each participant attempted the experiment and tried the gestures belonging to the seven gesture IDs in a random order of their choice, twice. The participant was expected to indicate which gesture he/she was performing, and this data was logged on into the processing application for later analysis, and use for training. A total of

17,177 samples were recorded. After pre-processing to remove duplicate values, our dataset contained 11,036 samples belonging to the seven gesture IDs.

5.3.2 Parameter selection

The dataset was split 50%-50% using random selection. It was decided to use six algorithms that support supervised learning present in the GRT. The parameter selection phase used the training dataset on each of the algorithms and evaluated the Accuracy percentage. In Table 5.3, the highest accuracy received by each algorithm is displayed together with the key parameters for which the accuracy was obtained. (Null rejection coefficient was set to 3 for all algorithms)

Table 5.3. Algorithms used for tenfold cross-validation

Classifying Algorithm	Highest Accuracy Percentage	Parameters
ANBC	83.58	Null Rejection Coefficient = 3
AdaBoost	96.77	Boosting Iterations = 20, Null Rejection Coefficient = 3
k-NN	92.13	K = 3, Null Rejection Coefficient = 3
MinDist	89.12	No. of Clusters = 2, Null Rejection Coefficient = 3
Softmax	91.98	Null Rejection Coefficient = 3
SVM	94.02	Linear Kernel, $\gamma = 0.1$, Null Rejection Coefficient = 3

5.3.3 Testing phase

Finally, the trained model was tested on the test dataset using the selected parameters. This dataset contained the remaining 50% of pre-processed data with 5518 samples. Table 5.4 provides the recognition rate obtained for each gesture ID class and the total recognition accuracy with the presents a summary of the recognition rates obtained using GRT for each gesture class in the test dataset.

Table 5.4. Gesture recognition rates

Gesture ID	1	2	3	4	5	6	8	Total Recognition Rate %	Time taken for Training (ms)	Time taken for Testing test dataset (ms)
No. of Samples	710	404	646	1356	846	565	991			
% ANBC	94.78	97.52	92.1	99.33	92.78	76.99	49.30	88.5	16	51
%AdaBoost	95.49	100	100	96.53	92.78	100	95.45	96.64	12310	48
% k-NN	88.12	94.5	92.32	86.84	90.85	96.1	97.00	88.12	5448	6505
% MinDist	87.74	97.52	87.61	95.87	92.78	100	93.13	93.43	29	43
% Softmax	94.78	89.35	87.61	96.09	92.78	100	89	93.05	10944	37
% SVM	95.49	98.26	100	95.87	92.78	100	89	95.19	1519	392

5.4 Results & discussion

During the gesture usability test, the use of Eversion (Gesture 10) and Inversion (Gesture 11) as suitable gestures is rejected by all participants. There seems to be a general consensus that moving the foot in such ways is not comfortable at all. The number of users who agree for using Gestures 7 and 9 is also low with only one and four users agreeing respectively. The common feature between these two is that they involve the raised heel. In fact, out of all the gestures, the ones in which the heel is raised has a lower agreement than resting foot gestures and raised-toes gestures. One possible reason behind this could be that the weight of the foot bears down on the toe in Gestures 5, 7 and 9. Moving a foot anticlockwise or clockwise further in this situation can be strenuous.

From our analysis of the gestures, it is evident that the best performing classifier for our data is AdaBoost with 20 boosting iterations and a null rejection coefficient of 3. The Adaboost algorithm classified gestures 2, 3 and 6 with a 100% accuracy while giving the lowest performance for gesture 5 at 92.78%. Additionally, SVM algorithm with a linear kernel, gamma of 0.1 and null rejection coefficient of 3 gave the second best performance. The SVM algorithm classified gestures 3 and 6 with 100% accuracy, with the lowest performance for gesture 7 at 89%. It must be noted that the k-NN algorithm detects gesture 8 with a higher recognition rate (97%), but has a far lower recognition rate for all of the other gestures, with the lowest being 86.84% for gesture 4.

Since the system applications expect the detection to be performed in real time, we evaluated the time that was taken to train the system using 5518 samples, as well as test the dataset of 5518 samples. In this situation too, AdaBoost performs acceptably although to train it takes 12310 milliseconds while SVM takes only 1519 milliseconds. However, for testing the testing dataset of 5518 samples, AdaBoost takes only 48 milliseconds while SVM takes 392 milliseconds.

5.5 Summary

This chapter first describes the tests carried out to evaluate the robustness of the hardware detection system. This included effect of murky water, ripples and bubbles, temperature and lighting level. Possible foot gestures identified in the literature survey and experimentations were tested with a pilot group. This enabled to develop a more manageable set of gestures for detection experiments. In the final stage, gesture detection was attempted using several machine learning algorithms. After selecting the parameters for each algorithm that provided the highest accuracies using cross-validation, detection of the same dataset was compared with six algorithms. Adaboost provided the best overall recognition rate with 96.64% recognition.

Chapter 6

Conclusion

6.1 Research summary

The work presented in this dissertation has set out to investigate how static foot gestures performed in water can be detected and recognized using Human Computer Interaction techniques. To find answers to the main topic, the following research objectives were formulated.

1. Evaluate existing hardware devices and techniques currently available and documented in research for the purpose of detecting objects in water.
2. Evaluate existing software techniques currently available and documented in research for the purpose of detecting static foot gestures.
3. Develop a robust hardware framework that can detect foot movement in water.
4. Test the hardware framework against external environmental factors
5. Develop a software framework that can detect the static gestures using objects detected by the hardware
6. Evaluate software framework static gesture detection performance.

During a comprehensive literature review, efforts were made to compile the list of techniques and technologies used for gestures in the air. During this process, it was noted that foot-based gestures were performed only in limited research, compared to the hand, arm and full body gesture research. Another realization was that only a few researchers had studied interactions with liquids and most of them dealt with water. This can be attributed to the fact that it is the most commonly interacted liquid in day to day life, so developing computer interfaces takes that into consideration. This research aimed to investigate the immersive interactions in water: i.e. three-dimensional interactions and no previous research had dealt with gesture recognition in such a situation.

After considering the viability of existing research techniques and technologies for water based interaction using feet, a novel method was proposed and implemented. The hardware system consisted of an acrylic water tank that had an LCD at its bottom. Around the tank layers that contained phototransistors on two adjacent sides and red lasers on the opposite sides were mounted to detect objects within the tank. The sensor output was channeled via a multiplexer and processed using an Arduino Uno microcontroller before reading via USB to a personal computer. Due to an optical phenomenon known as ghosting this system could only successfully detect one object immersed at a time, and if exceeded can detect ghost objects. This system was further improved once by using an extra layer of lasers, and finally using a touch frame to detect up to two objects successfully. The system's robustness was verified using several tests.

In the next stage, a gesture analysis was performed to evaluate the appropriate foot gestures for detection. Machine learning was used to train and detect seven static foot gestures. A total of 11,036 samples were recorded with 50% used for testing and 50% for training. The best results were obtained for the AdaBoost algorithm with a 96.64% total recognition rate.

The research findings suggest that detecting foot gestures can be performed with satisfactory performance rates.

6.2 Research implications

Due to its dexterity, the hand-based gesture movement possibilities are wide, which result in large application scenario possibilities. However, when it comes to the feet, we do not have the same level of possibilities. Nevertheless, we simulated several scenarios where the foot-based gestures can be used. This included the infotainment domain with applications in public spaces and home automation. Rehabilitative therapy in the medical domain is a sector where hydrotherapy is currently in use, and the proposed system can be used to improve or monitor patient progress.

6.3 Research limitations

From the literature survey, two distinct approaches to detect gestures were identified: body worn devices and external detection. This system falls into the latter category. While this enables the user to be free of any devices that may impede his/her natural movement, it makes the system setup more complex and detecting specific gestures more challenging.

In the selected approach, the foot is detected using three layers (5cm horizontal and 3 cm vertical separation between laser modules). In its current setup, it is theoretically possible that a foot could be placed between two lasers as so to avoid detection. However, since we consider gestures, during the foot movement, it would cross through many laser-phototransistor pairs. It is also possible to place a middle layer in such a way that its lasers are placed horizontally right between the lasers at the upper and lower layers.

When compared with optics-based approaches the output shape of the foot is coarse, and detecting fingers is not possible. However, the aim of this research is not to accurately detect the shape of the object immersed in water. Ideally, by adding more layers and/or modules, higher resolution can be obtained. Using the components we used in this model, it is physically possible to have maximum horizontal and vertical resolutions of 2 cm. Further reduction may be possible using more integrated laser modules. However, the hardware cost of such an effort can be overwhelming. Therefore, this system is ideally suited for interaction tanks that are in the order of few meters at maximum.

The ghosting issue mentioned in section 3.4 was overcome using a touch frame in the final version of the system. However, this does not mean that the system is multi-user friendly. The touch frame helps to eliminate ghosting when there are two real objects in the tank, whereas when the number of real objects increase, detection and elimination of ghost objects is more complicated. Additionally when the two real objects are close to each other, there is a possibility that the detection can be affected by the system could estimate it to be a single object rather than two feet.

Another issue with the touch frame is that it can introduce a slight positional error as it is located on the surface of the water (for example if the foot is inserted at a slant). While the touch frame and other electronic devices do not come into close contact with water during lab experiments, further padding and waterproofing may be required before the setup is used in a public space. On the same topic of public spaces, the system configuration is determined mainly by the tank size, and while a personal computer or laptop was used for the experiments, it would be possible to make the system more compact by utilizing a small factor computing device such as the Raspberry Pi.

6.4 Future work

To increase the accuracy of detected objects (i.e. increase resolution) one approach to replace the dot lasers modules and replace them with a line laser that outputs a parallel, non-diverging beam. While existing commercial line lasers have a beam with is diverging, they can be modified as presented in (Kimura et al., 2013) to enable this. On the opposite end to receive the bean, one would have to use a scanning element such as that utilized in a flatbed scanner) this should enable a high resolution although the cost has not been evaluated. While ultrasound detection is currently used for underwater detection, they are used for detecting larger objects than a human foot, and for larger volumes of water. Further, we did not experiment using low-cost acoustic sensors available for use in air detection, although it is feasibility for detection in water can be investigated. Another more practical possibility to consider a multi-sensory detection; i.e. detecting the 3D position using the SensorTank architecture while estimating the 2D shape via a camera based solution.

While external sensing devices such as sensors and cameras provide obstruction free movement for the user, the cost of implementation increases as the area of interaction increases. This is especially true when we consider the use of water. In this situation, another approach that would be worth to consider is using a body-worn (yet waterproof) contraption to detect the movement. The Thalmic Labs MYO worn on the thigh muscle is a candidate for this activity.

References

Alexander, J., Han, T., Judd, W., Irani, P., & Subramanian, S. (2012). Putting your best foot forward: investigating real-world mappings for foot-based gestures. In *Proceedings of the 30th international conference on Human factors in computing systems (CHI 2012)* (pp. 1229–1238). Austin, Texas: ACM. http://doi.org/10.1145/2208516.2208575

American Manual Alphabet. (2012). Retrieved May 19, 2015, from http://www.iidc.indiana.edu/cedir/kidsweb/amachart.html

American National Standard for Safe Use of Lasers. (2007). Retrieved May 12, 2015, from https://www.lia.org/PDF/Z136_1_s.pdf

Arroyo, E., Bonanni, L., & Valkanova, N. (2012). Embedded interaction in a water fountain for motivating behavior change in public space. In *Proceedings of the 30th ACM Conference on Human Factors in Computing Systems (CHI 2012)* (pp. 685–688). http://doi.org/10.1145/2207676.2207773

Ashton, K. (2009). That "Internet of Things" Thing. Retrieved from http://www.rfidjournal.com/articles/view?4986

Bailly, G., Müller, J., Rohs, M., Wigdor, D., & Kratz, S. (2012). ShoeSense: A New Perspective on Gestural Interaction and Wearable Applications. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1239–1248). New York, NY, USA: ACM. http://doi.org/10.1145/2207676.2208576

Barczewska, K., & Drozd, A. (2013). Comparison of methods for hand gesture recognition based on Dynamic Time Warping algorithm. In *Proceedings of 2013 Federated Conference on Computer Science and Information Systems (FedCSIS)* (pp. 207–210).

Berger, L., Klein, C., & Commandeur, M. (2008). Evaluation of the immediate and midterm effects of mobilization in hot spa water on static and dynamic balance in elderly subjects. *Annales de Réadaptation et de Médecine Physique : Revue Scientifique de La Société Française de Rééducation Fonctionnelle de Réadaptation et de Médecine Physique*, *51*(2), 84–95. http://doi.org/10.1016/j.annrmp.2007.10.007

Bonanni, L., Arroyo, E., Lee, C., & Selker, T. (2005). Smart sinks: real-world opportunities for context-aware interaction. In *extended abstracts on CHI'05* (pp. 1232–1235). New York, NY, USA: ACM. http://doi.org/10.1145/1056808.1056884

Bon-Woo Hwang, Sungmin Kim, & Seong-Whan Lee. (2006). A Full-Body Gesture Database for Automatic Gesture Recognition. In *7th International Conference on Automatic Face and Gesture Recognition (FGR06)* (pp. 243–248). IEEE. http://doi.org/10.1109/FGR.2006.8

Chen, F.-S., Fu, C.-M., & Huang, C.-L. (2003). Hand gesture recognition using a real-time tracking method and hidden Markov models. *Image and Vision Computing*, *21*(8), 745–758. http://doi.org/10.1016/S0262-8856(03)00070-2

Chinn, L., & Hertel, J. (2010). Rehabilitation of Foot and Ankle Injuries in Athletes. *Clinical Sports Medicine*, *29*(1), 157–167. http://doi.org/10.1016/j.csm.2009.09.006.Rehabilitation

Crossan, A., Brewster, S., & Ng, A. (2010). Foot Tapping for Mobile Interaction. In *Proceedings of the 24th BCS Interaction Specialist Group Conference* (pp. 418–422). Swinton, UK, UK: British Computer Society. Retrieved from http://dl.acm.org/citation.cfm?id=2146303.2146366

Dardas, N. H., & Georganas, N. D. (2011). Real-Time Hand Gesture Detection and Recognition Using Bag-of-Features and Support Vector Machine Techniques. *IEEE Transactions on Instrumentation and Measurement*, *60*(11), 3592–3607. http://doi.org/10.1109/TIM.2011.2161140

Dhawale, P., Masoodian, M., & Rogers, B. (2006). Bare-hand 3D gesture input to interactive systems. In *Proceedings of the 6th ACM SIGCHI New Zealand chapter's international conference on Computer-human interaction design centered HCI (CHINZ '06)* (pp. 25–32). http://doi.org/10.1145/1152760.1152764

Dietz, P. H., Han, J. Y., Westhues, J., Barnwell, J., & Yerazunis, W. (2006). Submerging Technologies. In *ACM SIGGRAPH 2006 Sketches* (p. 11). New York, NY, USA: ACM. http://doi.org/10.1145/1179849.1179863

Döring, T., Sylvester, A., & Schmidt, A. (2012). Exploring Material-centered Design Concepts for Tangible Interaction. In *CHI '12 Extended Abstracts on Human Factors in Computing Systems* (pp. 1523–1528). New York, NY, USA: ACM. http://doi.org/10.1145/2212776.2223666

Dutta, P., Aparna, B., Sridharan, B., & Vipin, T. (2013). Efficient Marker based Gesture Recognition Technique. *International Journal of Computer Applications*, *81*(November), 11–15. http://doi.org/10.5120/13974-196

Edelstein, J. E. (1988). Foot care for the aging. *Physical Therapy*, *68*(12), 1882–1886.

Floyd, R. T. (2008). *Manual of Structural Kinesiology* (17th ed.). McGraw-Hill. Retrieved from http://www.amazon.com/Manual-Structural-Kinesiology-T-Floyd/dp/0078022517%3FSubscriptionId%3D0JYN1NVW651KCA56C102%26tag%3Dtechkie-20%26linkCode%3Dxm2%26camp%3D2025%26creative%3D165953%26creativeASIN%3D0078022517

Francke, H., Ruiz-del-Solar, J., & Verschae, R. (2007). Real-time hand gesture detection and recognition using boosted classifiers and active learning. *Advances in Image and Video Technology*, 533–547. Retrieved from http://www.springerlink.com/index/hv32142088720j12.pdf

Gerhardt, T. J. (2009). *The Mud Tub : Computer , Meet Mud*. New York University. Retrieved from https://itp.nyu.edu/projects_documents/1241120984_TomGerhardt_ThesisDoc.pdf

Gillian, N. E. (2011). *Gesture Recognition for Musician Computer Interaction*. *Social Sciences*. Queen's University, Belfast. Retrieved from http://www.nickgillian.com/papers/NicholasGillianThesis.pdf

Gillian, N. E., & Paradiso, J. A. (2014). The Gesture Recognition Toolkit. *Journal of Machine Learning Research*, *15*, 3483–3487. Retrieved from http://jmlr.org/papers/v15/gillian14a.html

Google. (2015). Google Project Soli. Retrieved August 20, 2015, from https://www.google.com/atap/project-soli/

Guan, J., & Wade, M. G. (2000). The effect of aging on adaptive eye-hand coordination. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences*, *55*(3), P151–P162.

Gunawardena, K. S. L., & Hirakawa, M. (2015). GestureTank : A gesture detection water vessel for foot movements. *International Journal on Advances in ICT for Emerging Regions*, *8*(2), 1–11.

Gunawardena, K. S. L., Kimura, K., & Hirakawa, M. (2012). SensorTank : A Water Vessel for Interaction. In *Proceedings of the 10th Asia Pacific Conference on Computer Human Interaction (apCHI'12);* (pp. 759–760).

Gunawardena, K. S. L., Kimura, K., & Hirakawa, M. (2014). SensorTank: An Interactive Water Vessel. *Journal of Information Processing*, *22*(3), 547–555. http://doi.org/10.2197/ipsjjip.22.547

Gupta, S., Morris, D., Patel, S., & Tan, D. (2012). SoundWave: Using the Doppler Effect to Sense Gestures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1911–1914). New York, NY, USA: ACM. http://doi.org/10.1145/2207676.2208331

Häkkilä, J., Koskenranta, O., Posti, M., & He, Y. (2014). City Landmark As an Interactive Installation: Experiences with Stone, Water and Public Space. In *Proceedings of the 8th International Conference on Tangible, Embedded and Embodied Interaction* (pp. 221–224). New York, NY, USA: ACM. http://doi.org/10.1145/2540930.2540980

Hall, M. A., National, H., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA Data Mining Software: An Update. *SIGKDD Explor. Newsl.*, *11*(1), 10–18. http://doi.org/10.1145/1656274.1656278

Han, J., Honda, N., Asada, A., & Shibata, K. (2009). Automated acoustic method for counting and sizing farmed fish during transfer using DIDSON. *Fisheries Science*, *75*(6), 1359–1367. http://doi.org/10.1007/s12562-009-0162-5

Han, J. Y. (2005). Low-cost multi-touch sensing through frustrated total internal reflection. In *Proceedings of the 18th Annual ACM Symposium on User Interface Software and Technology (UIST '05)* (pp. 115–118). http://doi.org/10.1145/1095034.1095054

Han, T., Alexander, J., Karnik, A., Irani, P., & Subramanian, S. (2011). Kick: Investigating the Use of Kick Gestures for Mobile Interactions. In *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services* (pp. 29–32). New York, NY, USA: ACM. http://doi.org/10.1145/2037373.2037379

Hirai, S., Sakakibara, Y., & Hayashi, H. (2013). Enabling Interactive Bathroom Entertainment Using Embedded Touch Sensors in the Bathtub. In D. Reidsma, H. Katayose, & A. Nijholt (Eds.), *Advances in Computer Entertainment SE - 49*

(Vol. 8253, pp. 544–547). Springer International Publishing. http://doi.org/10.1007/978-3-319-03161-3_49

Hoffman, M., Varcholik, P., & LaViola, J. J. (2010). Breaking the status quo: Improving 3D gesture recognition with spatially convenient input devices. In *Proceedings of IEEE Virtual Reality (VR) Conference 2010* (pp. 59–66). http://doi.org/10.1109/VR.2010.5444813

Hoste, L., & Signer, B. (2014). Water Ball Z: An Augmented Fighting Game Using Water As Tactile Feedback. In *Proceedings of the 8th International Conference on Tangible, Embedded and Embodied Interaction* (pp. 173–176). New York, NY, USA: ACM. http://doi.org/10.1145/2540930.2540946

Hut, R. W. (2011). Scanning dynamic water surfaces using a Kinect as a sensor. *Geophysical Research Abstracts*, *13*(EGU2011-10096). Retrieved from http://meetingorganizer.copernicus.org/EGU2011/EGU2011-10096.pdf

Hut, R. W., Weijs, S. V., & Luxemburg, W. M. J. (2010). Using the Wiimote as a sensor in water research. *Water Resources Research*, *46*(12), 1–5. http://doi.org/10.1029/2010WR009350

Hyundai HCD-14. (2013). Retrieved May 9, 2015, from http://worldwide.hyundai.com/WW/Experience/ConceptCar/2013/HCD-14/index.html

Ikeda, M., & Hirakawa, M. (2010). A Multi-Dip Interface with Water. *Journal of Information Processing*, *51*(11), 2103–2111.

Ikeda, M., Nagira, N., & Hirakawa, M. (2009). A Multi-Dip Interface with Water. In *Proceedings of the International Conference on Complex, Intelligent and Software Intensive Systems* (pp. 169–176). Ieee. http://doi.org/10.1109/CISIS.2009.149

Ishii, H., & Ullmer, B. (1997). Tangible bits: towards seamless interfaces between people, bits and atoms. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (p. 241). http://doi.org/http://doi.acm.org/10.1145/604045.604048

Japan Body Size Data 1992-1994. (1994). Research Institute of Human Engineering for Quality Life.

Kadous, M. W., & Sammut, C. (2005). Classification of Multivariate Time Series and Structured Data Using Constructive Induction. *Machine Learning*, *58*(2-3), 179–216. http://doi.org/10.1007/s10994-005-5826-5

Kaltenbrunner, M., Bovermann, T., Bencina, R., & Costanza, E. (2005). TUIO - A Protocol for Table Based Tangible User Interfaces. In *Proceedings of the 6th International Workshop on Gesture in HumanComputer Interaction and Simulation GW 2005* (pp. 1–5).

Karam, M., & Schraefel, M. C. (2005). *A Taxonomy of Gestures in Human Computer Interactions*. Retrieved from http://eprints.soton.ac.uk/261149/

Kim, D., Hilliges, O., Izadi, S., Butler, A. D., Chen, J., Oikonomidis, I., & Olivier, P. (2012). Digits: Freehand 3D Interactions Anywhere Using a Wrist-worn Gloveless Sensor. In *Proceedings of the 25th Annual ACM Symposium on User*

Interface Software and Technology (pp. 167–176). New York, NY, USA: ACM. http://doi.org/10.1145/2380116.2380139

Kimura, K., Gunawardena, K. S. L., & Hirakawa, M. (2013). A framework for detecting positions of objects in water. In *2013 International Joint Conference on Awareness Science and Technology and Ubi-Media Computing (iCAST-UMEDIA),* (pp. 679–685). http://doi.org/10.1109/ICAwST.2013.6765524

Koh, J. T. K. V., Karunanayaka, K., Sepulveda, J., Tharakan, M. J., Krishnan, M., & Cheok, A. D. (2011). Liquid Interface: A Malleable, Transient, Direct-touch Interface. *Computers in Entertainment*, *9*(2), 7:1–7:8. http://doi.org/10.1145/1998376.1998378

Lamari, M. V., Bhuiyan, M. S., & Iwata, A. (1999). Hand alphabet recognition using morphological PCA and neural networks. In *IJCNN'99. International Joint Conference on Neural Networks. Proceedings (Cat. No.99CH36339)* (Vol. 4). http://doi.org/10.1109/IJCNN.1999.833533

Leap Motion Patent Application. (2014). Retrieved May 19, 2015, from https://community.leapmotion.com/t/leap-patent-application-and-other-references/717

Lenman, S., Bretzner, L., & Thuresson, B. (2002). *Computer vision based hand gesture interfaces for human-computer interaction*. *CID, Stockholm, Sweden*. Retrieved from http://cid.nada.kth.se/pdf/CID-172.pdf

Mace, D., Gao, W., & Coskun, A. (2013). Accelerometer-based Hand Gesture Recognition Using Feature Weighted Naïve Bayesian Classifiers and Dynamic Time Warping. In *Proceedings of the Companion Publication of the 2013 International Conference on Intelligent User Interfaces Companion (IUI '13 Companion)* (pp. 83–84). New York, NY, USA: ACM. http://doi.org/10.1145/2451176.2451211

Malik, S., & Laszlo, J. (2004). Visual Touchpad: A Two-handed Gestural Input Device. In *Proceedings of the 6th International Conference on Multimodal Interfaces (ICMI '04)* (pp. 289–296). New York, NY, USA: ACM. http://doi.org/10.1145/1027933.1027980

Mankoff, K. D., Russo, T. A., Norris, B. K., Hossainzadeh, S., Beem, L., Walter, J. I., & Tulaczyk, S. M. (2011). Kinects as sensors in earth science : glaciological , geomorphological , and hydrological applications. In *American Geophysical Union, Fall Meeting 2011, abstract #C41D-0442*.

Mann, S., Georgas, M., & Janzen, R. (2006). Water Jets as Pixels: Water Fountains as Both Sensors and Displays. In *Multimedia, 2006. ISM'06. Eighth IEEE International Symposium on* (pp. 766–772). http://doi.org/10.1109/ISM.2006.158

Matoba, Y., Takahashi, Y., Tokui, T., Phuong, S., Yamano, S., & Koike, H. (2013). AquaTop Display: A True "Immersive" Water Display System. In *ACM SIGGRAPH 2013 Emerging Technologies* (pp. 4:1–4:1). New York, NY, USA: ACM. http://doi.org/10.1145/2503368.2503372

Mistry, P., & Maes, P. (2009). SixthSense: A Wearable Gestural Interface. In *ACM SIGGRAPH ASIA 2009 Sketches* (pp. 11:1–11:1). New York, NY, USA: ACM. http://doi.org/10.1145/1667146.1667160

Mistry, P., & Maes, P. (2010). Mouseless. In *Adjunct proceedings of the 23nd annual ACM symposium on User interface software and technology - UIST '10* (pp. 441–442). New York, New York, USA: ACM Press. http://doi.org/10.1145/1866218.1866261

Mistry, P., Maes, P., & Chang, L. (2009). WUW - Wear Ur World: A Wearable Gestural Interface. In *CHI '09 Extended Abstracts on Human Factors in Computing Systems* (pp. 4111–4116). New York, NY, USA: ACM. http://doi.org/10.1145/1520340.1520626

Mortimer, R., Privopoulos, M., & Kumar, S. (2014). The effectiveness of hydrotherapy in the treatment of social and behavioral aspects of children with autism spectrum disorders: A systematic review. *Journal of Multidisciplinary Healthcare*, *7*, 93–103. http://doi.org/10.2147/JMDH.S55345

Nimbalkar, A. K., Karhe, R. R., & Patil, C. S. (2014). Face and Hand Gesture Recognition using Principle Component Analysis and kNN Classifier. *International Journal of Computer Applications*, *99*(8), 26–28.

Nordin, M., & Frankel, Victor, H. (2012). *Basic Biomechanics of the Musculoskeletal System* (4th ed.). LWW. Retrieved from http://www.amazon.com/Biomechanics-Musculoskeletal-System-Margareta-Nordin/dp/1609133358%3FSubscriptionId%3D0JYN1NVW651KCA56C102%26tag%3Dtechkie-20%26linkCode%3Dxm2%26camp%3D2025%26creative%3D165953%26creativeASIN%3D1609133358

Nymoen, K., Romarheim, M., Alexander, H., & Jensenius, R. (2015). MuMYO — Evaluating and Exploring the MYO Armband for Musical Interaction. In *New Interfaces for Musical Expression (NIME'15)* (p. Paper 179). Retrieved from https://nime2015.lsu.edu/proceedings/179/index.html

Ogata, M., Sugiura, Y., Osawa, H., & Imai, M. (2012). iRing: Intelligent Ring Using Infrared Reflection. In *Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology* (pp. 131–136). New York, NY, USA: ACM. http://doi.org/10.1145/2380116.2380135

Oh, J., Kim, T., & Hong, H. (2013). Using binary decision tree and multiclass SVM for human gesture recognition. In *2013 International Conference on Information Science and Applications, (ICISA)* (pp. 1 – 4). IEEE. http://doi.org/10.1109/ICISA.2013.6579388

Okude, E. (2011). Rainterior: an interactive water display with illuminating raindrops. In *ACM ITS* (pp. 270–271). Retrieved from http://dl.acm.org/citation.cfm?id=2076410

Pakkanen, T., & Raisamo, R. (2004). Appropriateness of Foot Interaction for Non-accurate Spatial Tasks. In *Proceedings of the CHI '04 Extended Abstracts on Human Factors in Computing Systems* (pp. 1123–1126). New York, NY, USA: ACM. http://doi.org/10.1145/985921.986004

Paradiso, J. A., Hsiao, K. Y., & Benbasat, A. (2000). Interfacing to the Foot: Apparatus and Applications. In *CHI '00 Extended Abstracts on Human Factors in Computing Systems* (pp. 175–176). New York, NY, USA: ACM. http://doi.org/10.1145/633292.633389

Pearson, G., & Weiser, M. (1986). Of moles and men: the design of foot controls for workstations. *ACM SIGCHI Bulletin*, *17*(4), 333–339. http://doi.org/10.1145/22339.22392

Petersen, E. J., Irish, S. M., Lyons, C. L., Miklaski, S. F., Bryan, J. M., Henderson, N. E., & Masullo, L. N. (1999). Reliability of Water Volumetry and the Figure of Eight Method on Subjects with Ankle Joint Swelling. *Journal of Orthopaedic & Sports Physical Therapy*, *29*(10), 609–615.

Pier, M. D., & Goldberg, I. R. (2005). Using Water As Interface Media in VR Applications. In *Proceedings of the 2005 Latin American Conference on Human-computer Interaction* (pp. 162–169). New York, NY, USA: ACM. http://doi.org/10.1145/1111360.1111377

Raffe, W. L., Tamassia, M., Zambetta, F., Li, X., Pell, S. J., & Mueller, F. F. (2015). Player-Computer Interaction Features for Designing Digital Play Experiences across Six Degrees of Water Contact. In *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '15* (pp. 295–305). New York, New York, USA: ACM Press. http://doi.org/10.1145/2793107.2793134

Rakkolainen, I., & Palovuori, K. (2004). Interactive Digital FogScreen. In *Proceedings of the Third Nordic Conference on Human-computer Interaction* (pp. 459–460). New York, NY, USA: ACM. http://doi.org/10.1145/1028014.1028096

Rasberry Pi. (2015). Retrieved May 9, 2015, from https://www.raspberrypi.org/

Reis, F. A. dos, Ribeiro, E. A., Carvalho, P. de T. C. de, Belchior, A. C. G., Arakaki, J. C., & Vasconcelos, R. A. de. (2004). Analysis of the Figure-of-Eight method and volumetry reliability for ankle edema measurement. *Revista Brasileira de Medicina Do Esporte*, *10*, 468–471. http://doi.org/http://dx.doi.org/10.1590/S1517-86922004000600003

Ren, Z., Meng, J., Yuan, J., & Zhang, Z. (2011). Robust Hand Gesture Recognition with Kinect Sensor. In *Proceedings of the 19th ACM International Conference on Multimedia* (Vol. 15, pp. 759–760). New York, NY, USA: ACM. http://doi.org/10.1145/2072298.2072443

Richter, H., Manke, F., & Seror, M. (2013). LiquiTouch: Liquid As a Medium for Versatile Tactile Feedback on Touch Surfaces. In *Proceedings of the 7th International Conference on Tangible, Embedded and Embodied Interaction* (pp. 315–318). New York, NY, USA: ACM. http://doi.org/10.1145/2460625.2460678

Rudomin, I., Diaz, M., Hernández, B., & Rivera, D. (2005). Water, Temperature and Proximity Sensing for a Mixed Reality Art Installation. In M. Maybury, O. Stock, & W. Wahlster (Eds.), *Intelligent Technologies for Interactive Entertainment SE - 16* (Vol. 3814, pp. 155–163). Springer Berlin Heidelberg. http://doi.org/10.1007/11590323_16

Rus, S., Grosse-Puppendahl, T., & Kuijper, A. (2014). Recognition of Bed Postures Using Mutual Capacitance Sensing. In E. Aarts, B. de Ruyter, P. Markopoulos, E. van Loenen, R. Wichert, B. Schouten, … G. O'Hare (Eds.), *Ambient Intelligence* (Vol. 8850, pp. 51–66). Springer International Publishing. http://doi.org/10.1007/978-3-319-14112-1_5

Samsung Smart TV. (2013). Retrieved May 9, 2015, from http://www.samsung.com/ph/smarttv/common/guide_book_3p_si/waving.html

Sangsuriyachot, N., & Sugimoto, M. (2012). Novel interaction techniques based on a combination of hand and foot gestures in tabletop environments. In *Proceedings of the 10th Asia Pacific Conference on Computer Human Interaction (apCHI'12)* (pp. 21–28). http://doi.org/10.1145/2350046.2350053

Sato, M., Poupyrev, I., & Harrison, C. (2012). Touché: Enhancing Touch Interaction on Humans, Screens, Liquids, and Everyday Objects. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems (CHI '12)* (pp. 483–492). New York, NY, USA: ACM. http://doi.org/10.1145/2207676.2207743

Schlömer, T., Poppinga, B., Henze, N., & Boll, S. (2008). Gesture Recognition with a Wii Controller. In *Proceedings of the 2nd International Conference on Tangible and Embedded Interaction (TEI '08)* (pp. 11–14). New York, NY, USA: ACM. http://doi.org/10.1145/1347390.1347395

Schmidt, T., Araujo, F. P., Pappa, G. L., & Nascimento, E. R. (2014). Real-Time Hand Gesture Recognition Based on Sparse Positional Data. In *Proceedings of WVC 2014, Brazilian Workshop on Computer Vision* (pp. 243–248). Retrieved from http://www.wvc2014.facom.ufu.br/sites/wvc2014.facom.ufu.br/files/IMCE/proceedings_WVC_2014_final.pdf

Schöning, J., Daiber, F., Krüger, A., & Rohs, M. (2009). Using Hands and Feet to Navigate and Manipulate Spatial Data. In *CHI '09 Extended Abstracts on Human Factors in Computing Systems* (pp. 4663–4668). New York, NY, USA: ACM. http://doi.org/10.1145/1520340.1520717

Scott, J., Dearman, D., Yatani, K., & Truong, K. N. (2010). Sensing Foot Gestures from the Pocket. In *Proceedings of the 23rd Annual ACM Symposium on User Interface Software and Technology(UIST '10)* (pp. 199–208). New York, NY, USA: ACM. http://doi.org/10.1145/1866029.1866063

Srinivasan, P., Birchfield, D., Qian, G., & Kidané, A. (2005). A Pressure Sensing Floor for Interactive Media Applications. In *Proceedings of the 2005 ACM SIGCHI International Conference on Advances in Computer Entertainment Technology* (pp. 278–281). New York, NY, USA: ACM. http://doi.org/10.1145/1178477.1178526

Takeoka, Y., Miyaki, T., & Rekimoto, J. (2010). Z-touch : An Infrastructure for 3D gesture interaction in the proximity of tabletop surfaces. In *ACM International Conference on Interactive Tabletops and Surfaces (ITS '10)* (pp. 90–94). New York, NY, USA: ACM. http://doi.org/10.1145/1936652.1936668

Tanabe, C., & Hirakawa, M. (2006). O-Key: A Scoopable Interface. In *Proceedings of the Eighth IEEE International Symposium on Multimedia (ISM'06)* (pp. 185–192). Ieee. http://doi.org/10.1109/ISM.2006.117

Vafadar, M., & Behrad, A. (2008). Human hand gesture recognition using spatio-temporal volumes for human-computer interaction. In *International Symposium on Telecommunications, 2008 (IST 2008)* (pp. 713–718).

Ventä-Olkkonen, L., Akerman, P., Puikkonen, A., Colley, A., & Häkkilä, J. (2014). Touching the Ice: In-the-wild Study of an Interactive Icewall. In *Proceedings of the 26th Australian Computer-Human Interaction Conference on Designing Futures: The Future of Design* (pp. 129–132). New York, NY, USA: ACM. http://doi.org/10.1145/2686612.2686630

Virolainen, A., Puikkonen, A., Kärkkäinen, T., & Häkkilä, J. (2010). Cool Interaction with Calm Technologies: Experimenting with Ice As a Multitouch Surface. In *ACM International Conference on Interactive Tabletops and Surfaces* (pp. 15–18). New York, NY, USA: ACM. http://doi.org/10.1145/1936652.1936656

Visutarrom, T., Mongkolnam, P., & Chan, J. H. (2014). Postural classification using Kinect. In *Computer Science and Engineering Conference (ICSEC), 2014 International* (pp. 403–408). http://doi.org/10.1109/ICSEC.2014.6978231

Wang, C. C., & Wang, K. C. (2008). Hand Posture Recognition Using Adaboost with SIFT for Human Robot Interaction. In S. Lee, I. Suh, & M. Kim (Eds.), *Recent Progress in Robotics: Viable Robotic Service to Human SE - 25* (Vol. 370, pp. 317–329). Springer Berlin Heidelberg. http://doi.org/10.1007/978-3-540-76729-9_25

Wang, G. W., Zhang, C., & Zhuang, J. (2012). An application of classifier combination methods in hand gesture recognition. *Mathematical Problems in Engineering*, *2012*, 1–17. http://doi.org/10.1155/2012/346951

Watanabe, J. (2007). VortexBath: Study of Tangible Interaction with Water in Bathroom for Accessing and Playing Media Files. In J. Jacko (Ed.), *Human-Computer Interaction. Interaction Platforms and Techniques SE - 136* (Vol. 4551, pp. 1240–1248). Springer Berlin Heidelberg. http://doi.org/10.1007/978-3-540-73107-8_136

Weiser, M. (1991). The Computer for the 21st Century. *Scientific American Special Issue on Communications, Computers, and Networks*, *265*(3), 94–104. http://doi.org/http://doi.acm.org/10.1145/329124.329126

Weiser, M. (1994). The World is Not a Desktop. *Interactions*, *1*(1), 7–8. http://doi.org/10.1145/174800.174801

White, D. J., Svellingen, C., & Strachan, N. J. C. (2006). Automated measurement of species and length of fish by computer vision. *Fisheries Research*, *80*(2-3), 203–210. http://doi.org/10.1016/j.fishres.2006.04.009

Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data Mining: Practical Machine Learning Tools and Techniques, Third Edition (The Morgan Kaufmann Series in Data Management Systems)* (3rd ed.). Morgan Kaufmann.

Yabu, H., Kamada, Y., Takahashi, M., Kawarazuka, Y., & Miyata, K. (2005). Ton2: A VR Application with Novel Interaction Method Using Displacement Data. In *ACM SIGGRAPH 2005 Emerging Technologies*. New York, NY, USA: ACM. http://doi.org/10.1145/1187297.1187322

Yee, W. (2009). Potential Limitations of Multi-touch Gesture Vocabulary: Differentiation, Adoption, Fatigue. In J. Jacko (Ed.), *Human-Computer Interaction. Novel Interaction Methods and Techniques SE - 32* (Vol. 5611, pp. 291–300). Berlin, Heidelberg: Springer Berlin Heidelberg.

http://doi.org/10.1007/978-3-642-02577-8_32

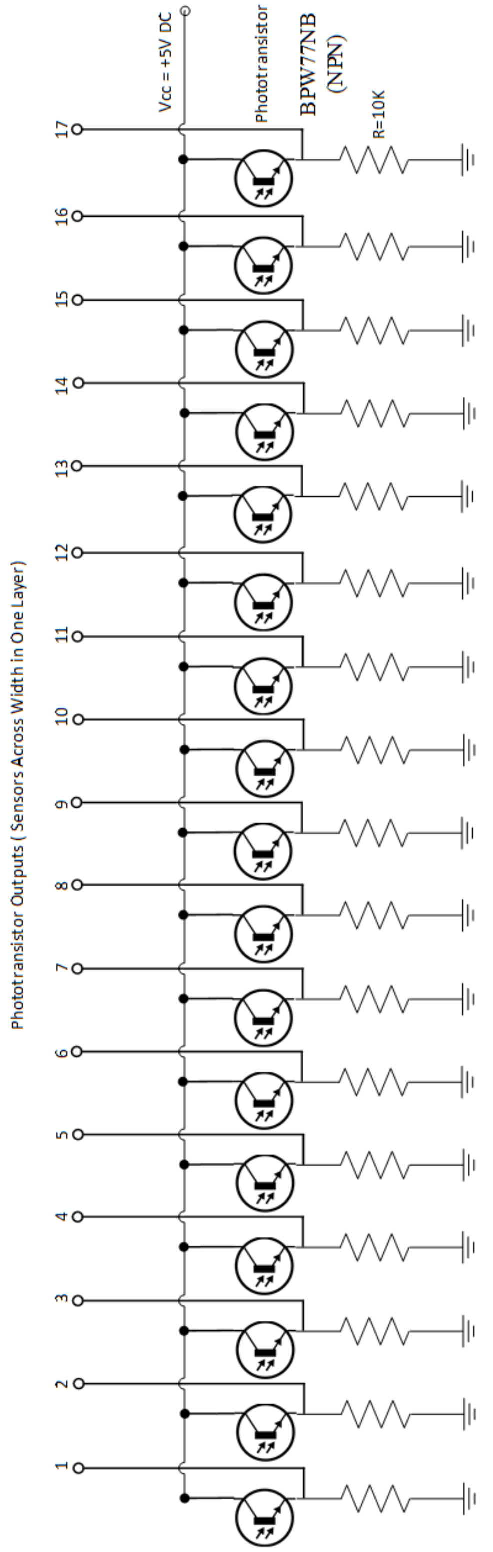
Zeiß, S., Marinc, A., Braun, A., Große-Puppendahl, T., & Beck, S. (2014). A gesture-based door control using capacitive sensors. *Lecture Notes in Computer Science (including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *8530 LNCS*, 207–216. http://doi.org/10.1007/978-3-319-07788-8_20

Ziaie, P., Müller, T., Foster, M. E., & Knoll, A. (2009). A Naïve Bayes Classifier with Distance Weighting for Hand-Gesture Recognition. In H. Sarbazi-Azad, B. Parhami, S.-G. Miremadi, & S. Hessabi (Eds.), *Advances in Computer Science and Engineering SE - 38* (Vol. 6, pp. 308–315). Springer Berlin Heidelberg. http://doi.org/10.1007/978-3-540-89985-3_38

Zion, B., Alchanatis, V., Ostrovsky, V., Barki, A., & Karplus, I. (2007). Real-time underwater sorting of edible fish species. *Computers and Electronics in Agriculture*, *56*(1), 34–45. http://doi.org/10.1016/j.compag.2006.12.007

Zion, B., Shklyar, a., & Karplus, I. (1999). Sorting fish by computer vision. *Computers and Electronics in Agriculture*, *23*(3), 175–187. http://doi.org/10.1016/S0168-1699(99)00030-7

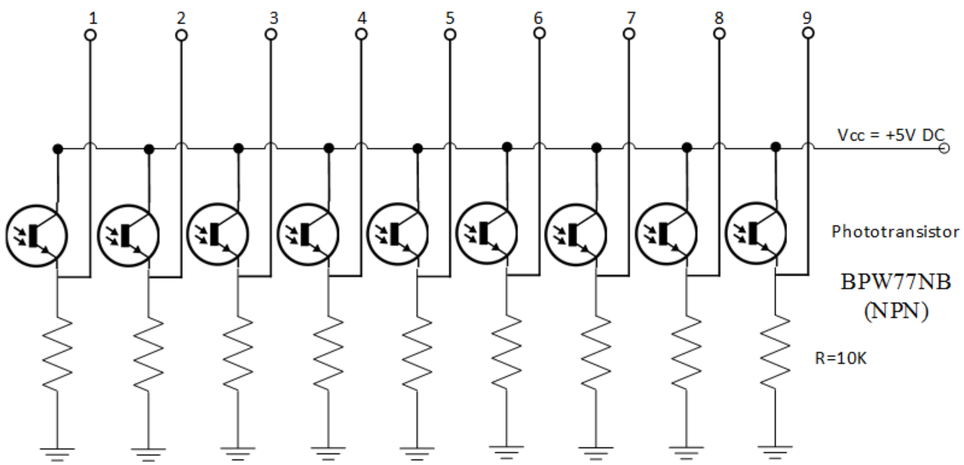
Appendix A: Circuit diagrams for Hardware



Ribbon Cable Header Pin Layout

Header Connector Number	Connection	Color
1	GND	Black
2	Output 1	Grey
3	Output 2	Blue
4	Output 3	Yellow
5	Output 4	Red
6	Output 5	Black
7	Output 6	Grey
8	Output 7	Blue
9	Output 8	Yellow
10	Output 9	Red
11	VCC	White
12	Output 10	Purple
13	Output 11	Green
14	Output 12	Orange
15	Output 13	Brown
16	Output 14	White
17	Output 15	Purple
18	Output 16	Green
19	Output 17	Orange
20	No Connection	Brown

Phototransistor Outputs (Sensors Across Breadth in One Layer)



Ribbon Cable Header Pin Layout

Header Connector Number	Connection	Color
1	GND	Black
2	Output 1	Grey
3	Output 2	Blue
4	Output 3	Yellow
5	Output 4	Red
6	No Connection	Black
7	No Connection	Grey
8	No Connection	Blue
9	No Connection	Yellow
10	No Connection	Red

Header Connector Number	Connection	Color
11	VCC	White
12	Output 5	Purple
13	Output 6	Green
14	Output 7	Orange
15	Output 8	Brown
16	Output 9	White
17	No Connection	Purple
18	No Connection	Green
19	No Connection	Orange
20	No Connection	Brown