GestureTank: A gesture detection water vessel for foot movements

K.S. Lasith Gunawardena, Masahito Hirakawa

Abstract— **Computers have become ubiquitous and integrated into our day-to-day activities. Researchers have been exploring mechanisms for interacting with computers using natural means in natural environments. Water interaction is a perfect example. This paper presents our attempt to use foot gestures performed in water as an interaction mechanism. It is an extension of our previous study for detecting objects in a water vessel. An experiment was performed to determine which foot-based gestures are suitable for implementation, and we proceeded to recognize a selected set of gestures using machine-learning techniques. We present our findings regarding which algorithms provide the best recognition rates.**

Index Terms— **Gesture Recognition, 3D Interaction, Machine Learning.**

I. INTRODUCTION

The need to use a keyboard and mouse limits our interactions with computers. Novel interaction techniques have become necessary because of the need to be connected on a continuous basis while engaging in day-to-day activities. This is particularly true because computing devices are no longer seen as simply data processing devices but as gateways to activities such as knowledge acquisition and entertainment. The recent introduction of wearable computing devices renders such devices invisible, resulting in ubiquitous computing.

Vision is a sensory facility that allows us to gather information from the external environment. While substantial communication proceeds in spoken and written languages, gestures are a more subtle form of interpersonal communication. Given its naturalness, many researchers have attempted to develop schemes to detect human gestures, particularly regarding vision-based research using one or more cameras as sensors. However, in the past decade, there has been 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 the Microsoft Kinect [1], SoftKinetic's DepthSense [2], Leap Motion's Leap Motion Controller [3], Ring by Logbar Inc. [4], and Myo by Thalmic Labs [5].

Most of these devices are targeted at detecting either finger motion, hand/arm movements, head movements, or body movements. Only a few devices have been designed for use with foot-based interactions. This might reflect the fact that foot movements are less accurate, involve further execution time, and are probably less fulfilling than hand movements for the same task [6], [7]. However, it is possible to use foot interaction in non-accurate spatial tasks [6]. Furthermore, since human locomotion is essentially bipedal, investigating foot-based interaction is beneficial.

Meanwhile, the use of technology for public social interactions is increasing in popularity. 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 cultural activity. In such environments, a strong potential exists for highly user-friendly (or invisible) interfaces that use water as an interface medium. Water creates unique sensations that can have a soothing effect on the body. Hydrotherapy is a well-known treatment method that uses the physical properties of water for wellbeing. Fatigue attributed to gestural movements in air and on surfaces has been identified as an issue that requires investigation [8]–[11]

This study introduces GestureTank, a tangible user interface that enables interaction in a volume of water. A unique feature of this study is that it treats a volume of water as a three-dimensional (3D) interaction space, in contrast to previous studies [12]–[15], which have concentrated on interactions in air (using various body parts or walking movements) or on the surface of a volume of water. The GestureTank detects the 3D positions of objects (e.g., feet and hands) inserted in water and detects seven gestures performed by the human foot.

This study was conducted on the basis of observations obtained through our previous studies [16]–[19] and is an improvement of our SensorTank architecture [20]. GestureTank is organized as a water vessel wherein laser and phototransistor combinations are arranged as sensing units. GestureTank provides visual, auditory, and thermal feedback to the user through a liquid crystal display (LCD) monitor mounted below the vessel, embedded speakers, and a heating element, respectively, as proposed for SensorTank. A touch frame has been implemented in GestureTank over the vessel to assist in the detection of multiple objects inserted in water. Although the sensing resolution of laser-phototransistor pairs is rather coarse, the system can successfully identify foot gestures with reasonable performance using machine-learning techniques.

The remainder of this paper is organized as follows. Section II provides an overview of existing research, which is the background for this study. In addition, the limitations of previous studies when applied to our problem domain are discussed. In Section III, we explain our previous study, which has led to the present research, followed by an explanation of the hardware and software designs, as well as prototype

applications of the proposed system. Section IV discusses experiments conducted for usability testing of gestures and an evaluation of system performance. Section V provides results of the gesture recognition performance and further elaborates on the limitations of the proposed system. Section VI provides conclusions and outlines potential future research.

II. BACKGROUND AND RELATED WORK

A. Foot-based gestures

Previous studies in gesture detection have focused on the hands and body, and most applications target hand and sometimes body movement. As suggested by Alexander et al. [13], the human foot is a highly dexterous system with advanced movements using multiple joints that increase in movement complexity from the hip to ankle.

However, as observed by Scott et al. [12], the foot does not offer the same degree of precision and dexterity for selection as the wrist and hand. In their study, four motions were considered for foot-based interaction, as illustrated in Fig. 1a–1d.

a. dorsiflexion

b. heel rotation

c. plantar flexion

d. toe rotation
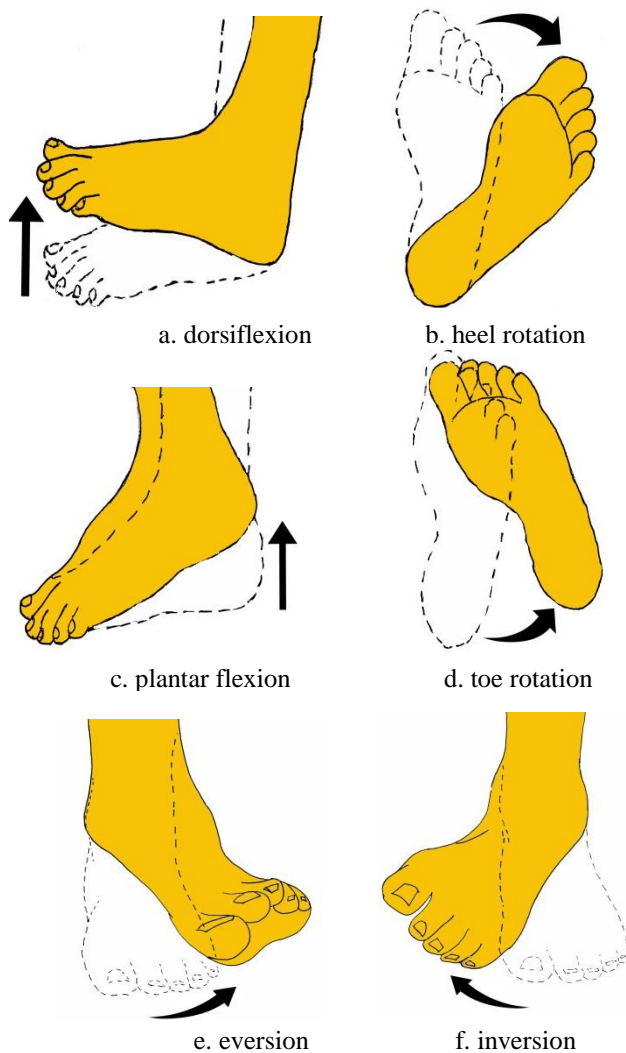
e. eversion

f. inversion

Fig. 1. Six basic foot movements: (b) and (d) right foot; (a), (c), (e), and (f) left foot.

- 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 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 rotation on the toe.

Eversion and inversion, as illustrated in Fig. 1e and 1f, are other possible foot movements documented in biomechanics [21]. Inversion and eversion are movements to turn the sole of the foot inwards and outwards, respectively. In terms of dexterity, some movements have limits in their degree of flexibility. The typical limits for inversion and eversion are 20°–30° and 5°–15° [21], respectively. Similarly, the ranges of motion for dorsiflexion and plantar flexion are 10°–20° and 40°–55°, respectively [22].

Foot gestures are of interest for a number of reasons. Feet can be used to supplement gestures performed by hands when operating a vehicle or in situations wherein the hands are occupied or unavailable due to disability. For example, operating a mobile phone with a hands-free kit is useful when carrying goods in both hands. Another situation is the use of ordinary electronic appliances in a kitchen or another environment in which water splashing can affect appliance operation.

B. Gesture detection technologies

Many commercial gesture detection products have been introduced in the past several years. Based on our review of previous studies, these can be classified by several aspects.

1) Two-dimensional (2D) gestures vs. three-dimensional (3D) gestures

2D gestures performed by the hands, feet, or body can be detected using cameras ranging from basic webcams [23] to more sophisticated high-resolution or high-speed cameras [24]. In addition, devices such as pressure sensing pads can be used to track human gait. Touch sensitive surfaces, such as those on tablets, mobile phones, and touch screens, have evolved; they can now detect not only positional touch but also strokes made using one or more fingers. Stereo cameras used with software-based techniques [25] and combinations of cameras and depth sensing devices are two approaches for 3D gesture detection. One common technique used by depth sensing cameras is known as time-of-flight. Experiments have also been conducted on 3D gestural interaction in the proximity of tabletop surfaces [26].

2) Camera vs. sensor-based approaches

Camera-based approaches use images or video streaming and perform image/video processing to highlight required objects and extract positional data. The first products for gesture detection exclusively used regular red-green-blue (RGB) video, and special cameras that sense other supplemental features, such as heat and infrared (IR), are recent additions. The Microsoft Kinect uses IR laser projection and a combination of two types of camera, i.e., a conventional RGB

camera and an IR-sensing camera, to detect 3D movements of a human body. Camera-based gesture recognition can be affected by a variety of noises from external/environmental factors, such as lighting, and require high processing power.

On the other hand, sensors that detect valuable information, such as relative position and 3D acceleration, show promise. One advantage of such sensors over camera systems is that the user does not have to face a particular direction. The accelerometer sensor contained in the Nintendo Wii Remote [27] has been used for gesture detection [28], [29]. Furthermore, accelerometers embedded in watches [13], [30] and mobile phones [12] have also been used for gesture detection. The Myo armband [5] recently introduced by Thalmic Labs uses electromyography data pertaining to arm movement to control digital devices. Soundwave [31] uses the speaker and microphone combination embedded in most computing devices to sense in-air gestures around the computer using the Doppler effect. Touché [14] enables gesture recognition using a swept frequency capacitive testing technique with only a single electrode.

3) Body worn devices vs. environmental gesture detection devices

The data glove was one of the first body worn (wearable) gesture devices. Active data gloves comprise sensors that measure the flexing of joints and acceleration. Although some researchers have proposed body worn cameras, smaller devices, such as rings [4], armbands, leg bands, and shoe-embedded sensors, have also been introduced. One disadvantage of such devices is that the natural movement of the user can be impeded because they "wear" the device throughout the gesture detection process. Environmental devices, such as cameras and depth sensors, are advantageous because they do not require wearing cumbersome items.

C. Raw point data generators vs. classified output generators

Most gesture detection devices, such as cameras and the Microsoft Kinect, provide data pertaining to the gesture that can then be processed by software for training and testing gestures. However, embedded-type devices, such as Myo, with built-in classifiers that enable the devices to directly output the gesture type have recently been introduced.

D. Machine learning algorithms for gesture detection

Machine learning is an extensive discipline with applications in different domains. Machine learning can be used for two types of problems: classification and regression. The output takes discrete values in classification and continuous values in regression. Spatial gestures form a static classification problem for which algorithms, such as naïve Bayes [12], [32], [33], k-nearest neighbor (k-NN) [34], [35], adaptive boosting (AdaBoost) [29], [36], support vector machines (SVM) [14], [37], and decision trees [38], have been used. On the other hand, a temporal classification problem wherein real-time tracking is performed requires different algorithms, such as hidden Markov models [28], [39], and dynamic time warping [30].

A supervised approach has commonly been applied to detecting gestures. This requires a training phase in which gestures are recorded and trained using a learning algorithm to build a model. In the training phase, possible gestures are divided into classes. Once a set of gestures is recorded and classified, the model can predict which class a new incoming input value belongs to. In this scenario, each gesture is recorded using a feature vector, i.e., a predetermined set of features.

E. Water interaction

Most existing experiments have considered hand and/or body gestures in air as input to interactive systems. However, in this form of interaction, the user lacks proper feedback from the physical world. In contrast, interaction with water can provide the user tangible and tactile feedback. The Wii Remote has been used as a sensor for water level measurement [40]. The Microsoft Kinect motion detector device was originally designed for use with full body gaming but has been used for scanning a dynamic water surface [41] and depth (up to 0.203 m) [42]. However, these experiments were not intended to be used to explore techniques for human–computer interaction. AquaTop [43] uses the Kinect to detect gestures performed at the surface of cloudy water.

When considering research in liquid interaction, a number of applications have been presented [16]–[20], [44], [45]. Koh et al. presented a tangible and malleable interface that allowed the user to produce a 3D response using ferromagnetic fluids [46]. However, in most previous trials, interaction with water (or another liquid) occurred at the surface level because the researchers were interested in the dynamic aspects of water rather than human gestures. Although Touché [14] enables the detection of gestures performed in water and in air, it was not designed to provide gesture-related positional data. Gurgle [15] is a public space that augments an existing water fountain with watery reflections and sound to motivate change in human behavior.

The sonar technique is a well-known and established methodology that has been used for object detection in water. The marine industry uses sonar instruments, such as hydrophones; however, they are costly and not designed for use over distances less than 1 m. Most low-cost acoustic sensors that use ultrasonic sound waves have narrow detection angle (20°–30°) and minimum detection distance (10–20 cm). Furthermore, acoustic sensors detect only the distance to an object; thus, to generate information about the shape of an object, multiple sensors operating at different frequencies might be required. In comparison, low-power laser modules are quite reasonably priced, and when paired with phototransistors, they provide a simple detection mechanism that can be scaled according to the detection space.

In summary, limited research that focuses on the issues related to object tracking or gesture detection within a limited aquatic space, such as a water vessel, has been conducted.

III. PROTOTYPE SYSTEM

A. Background

One of the authors has presented interactive systems using water as a medium [16]–[18]. O-Key [16] uses a web camera, video projector, tub, and personal computer to detect the movement of hands at a horizontal level (2D) to identify a scooping gesture. Subsequent experiments [17], [18] made use

of frustrated total internal reflection [47] as a technique for sensing hands submersed in a tub. The system comprises an acrylic tub, two web cameras, a video projector, and a personal computer. The depth positions of the hands in a 3D volume (i.e., water) and their spatial positions can be obtained using the cameras.

However, the setup space required for camera-based approaches is significantly large because cameras must be positioned away from targets so that the field of view can cover the entire search space. For a tub with a base size of 50 cm × 37 cm [17], the distance from the cameras to the tub is 64 cm. If a larger water tub (interaction area) is required, the camera must be positioned further away. Therefore, the practical implementation of camera-based detection in water is limited, particularly when considering foot interactions. Implementing foot interaction using camera-based systems requires the interaction area to be considerably elevated or the floor to be suitably modified to embed devices [26].

Experiments conducted using the Kinect motion detector [1] revealed that, when the unit is positioned above the water surface and gestures are performed underwater, the ripples generated act as an obstacle to successful detection. Moreover, when the Kinect was mounted on the side of the water body using a clear acrylic tank, detection is successful within only 5 cm from the tank wall. Further experiments conducted using the SoftKinetic DepthSense 311 Camera [2] provided similar results. One reason why these devices do not perform as expected in water can be attributed to the use of low power IR illumination. IR is attenuated in water; therefore, it can be difficult to detect objects as depth and distance increase.

B. Hardware design

In a previous study on the SensorTank [20], the issues explained in the previous section were overcome using a sensor array comprising red lasers and phototransistors. The proposed GestureTank is an improvement over the SensorTank and uses the same tank constructed using transparent acrylic panels of 1.5 cm thickness with tank dimensions of 20 cm × 88.4 cm × 50 cm (H × L × W). A photograph of GestureTank is shown in Fig. 2.

The red lasers and phototransistors are arranged such that they face each other at the sides of the tank, as illustrated in Fig. 3. When an object, such as a hand or foot, is inserted into the tank, it blocks the path of one of the red lasers, and this can be detected by the associated phototransistor (Fig. 4). A total of 78 sensors are mounted at horizontal intervals of 5 cm and at vertical intervals of 3 cm. This separation was based on hand and foot anthropometric data, as explained in the literature [20].
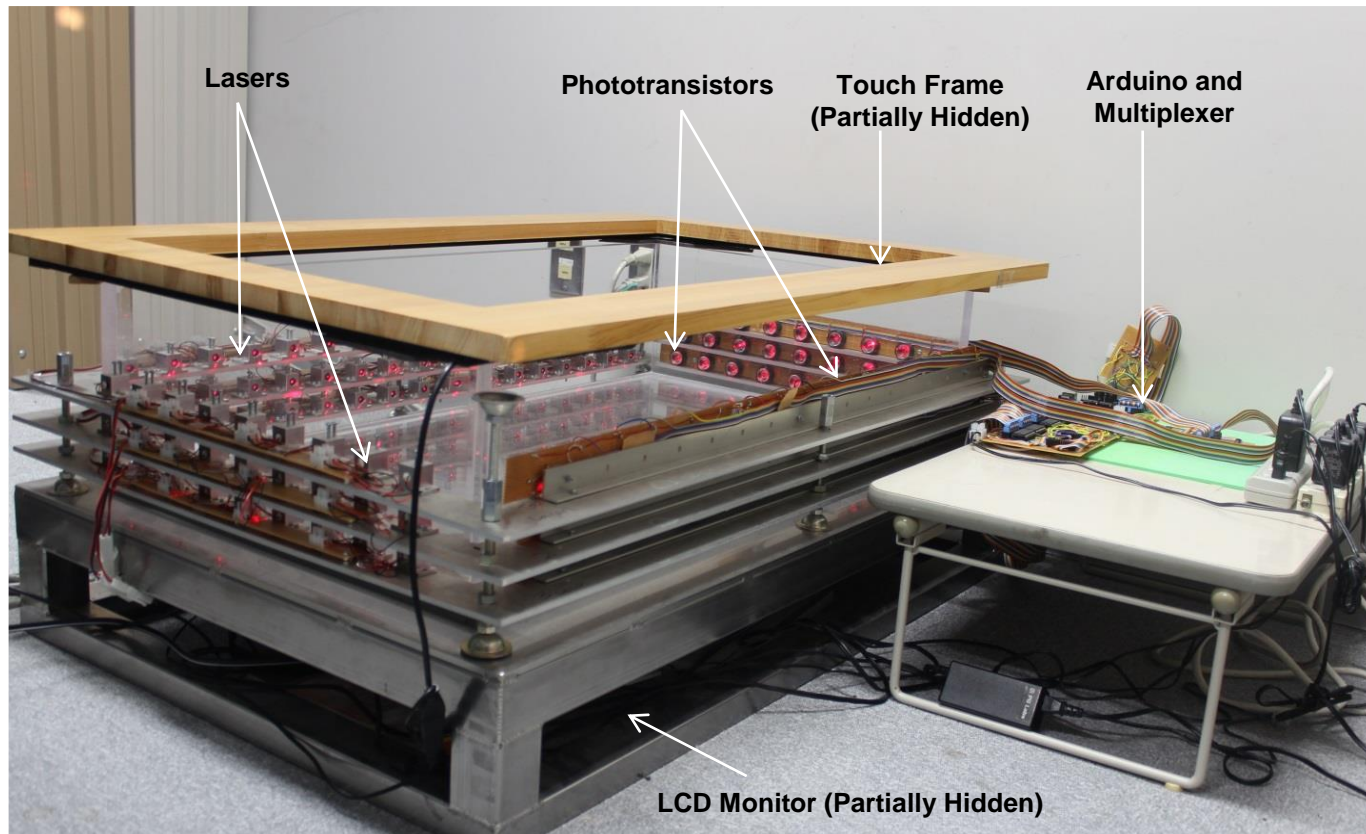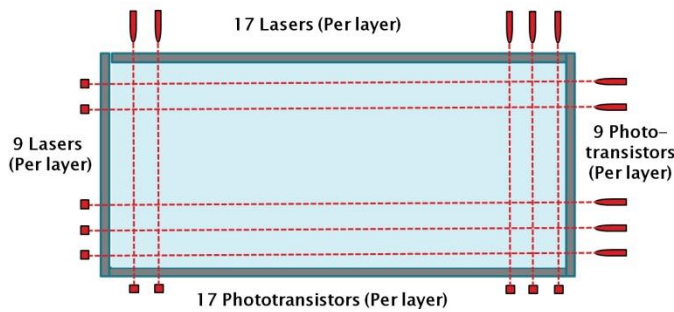

Fig. 2. GestureTank system
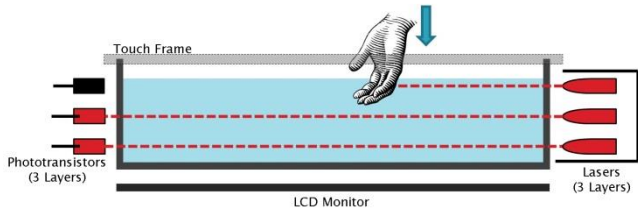
Fig. 3. Overhead view of the GestureTank

Fig. 4. Cross-section diagram of the GestureTank

One issue with our previous version [20] is the appearance of ghost objects in detection due to occlusion. In our previous study, a separate laser-phototransistor layer was used for ghost cancelation. However, this can also be affected by occlusion. In this study, a touch frame is introduced for ghost cancelation.

The touch frame (PQ Labs G3) is a commercially available device with a diagonal dimension of 40 inches for multi-touch detection on a screen. The glass affixed to the touch frame is discarded and the frame is positioned above the water surface of our tank for ghost elimination.

The algorithm used to eliminate ghost points is explained as follows. Consider two objects inserted into water, as illustrated in Fig. 5.

$\bigotimes$ Object position A_i taken from touch sensing frame

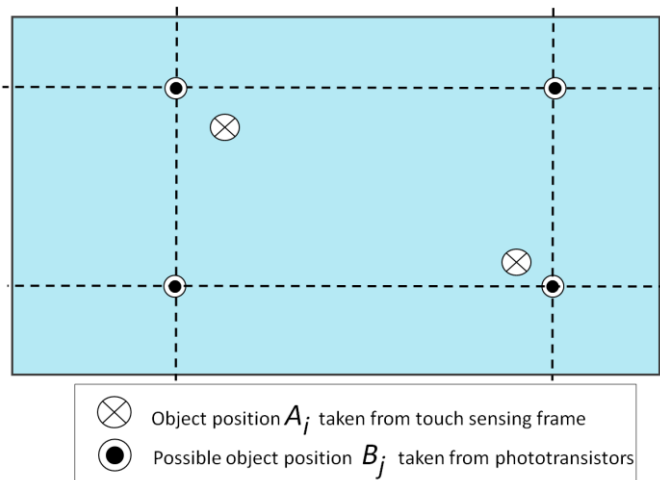$\odot$ Possible object position B_j taken from phototransistors

Fig. 5. Scheme for the resolution of ghost points

Two sets of associated positional data are provided by the touch sensing frame device. Two sensors along the width and breadth of the tank simultaneously detect objects via the laser-phototransistor array, resulting in four possible objects. Note that the object position taken from the touch sensing frame device might not be identical to that taken by the sensors because objects can be inserted at a slant.

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 Fig. 5, the system calculates the Euclidian distance from the weight center B_j ($1 \leq j \leq 4$) of each possible object region, indicated by a black-centered round symbol, and identifies the one with the shortest distance from A_i as the actual object region to be associated with A_i.

The architecture of the GestureTank system is illustrated in Fig. 6. To estimate the temperature of the water, a waterproof thermistor is connected to an Arduino Uno Microcontroller [48]. An aquarium water heater is installed in the tank to heat the water as required. One of the application scenarios discussed in Section III.D utilizes a water faucet connected to the Arduino. Visual interaction is performed using the LCD monitor placed at the bottom of the tank. Built-in audio speakers are also used for our applications.
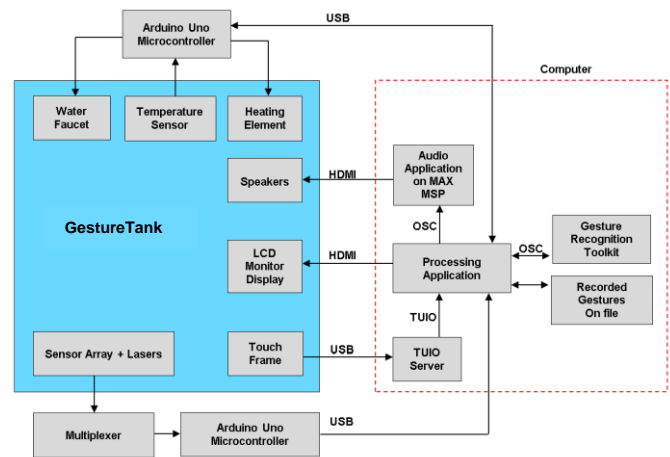

Fig. 6. GestureTank system architecture

C. Software design

The Arduino is programmed using its proprietary programming language to provide a stream of serial data corresponding to the coordinates of 3D space in the GestureTank. This data is received via USB to our main application, which was developed using processing language.

The positional information is coded by the length (x-axis) and width (y-axis) coordinates of the tracked object, along with a depth value that corresponds to the sum of weighted values given for the layers at which the object is located. Consider that values of 1, 2, and 4 are assigned to layers 1, 2, and 3, respectively. This ensures that the summation of any weight combination can be easily decoded to determine which layers contain an object. For example, when an object covers layers 1 and 3, the depth information is expressed as 5 (1 + 4). This ensures that the data stream is optimised for performance. Therefore, the entire 3D point space is compressed into a 2D matrix with 17 columns and nine rows. The elements in the matrix can have a value of zero if no objects are present at that location, or values of one to seven depending on whether the object is present in one, two, or three layers. This 2D matrix is transformed into a 2D matrix with binary data, with zero indicating no object at the given length and width in any laser

layer (1–3). A connected component analysis (blob detection) runs on the 2D matrix to identify different objects in the tank.

Noise filtering is performed to extract objects suitable for further analysis. Ultimately, each gesture object forms a 3D point cloud.

Data sent from the touch frame are received via the Tangible User Interface Objects (TUIO) protocol by our processing application. The ghost cancelation algorithm is then executed to filter only real objects in the gesture tank.

As explained by Gillian [49], despite powerful sensors and rapid prototyping tools, performing real-time gesture recognition has some challenges because some of the existing powerful machine-learning tools are better suited to offline analysis. The open source gesture recognition toolkit (GRT) [49] is used for our analysis because it employs a selection of machine-learning algorithms, including adaptive naïve Bayes classifiers (ANBC), AdaBoost, k-NN, minimum distance (MinDist), Softmax classifiers, and SVMs, to be integrated seamlessly. The GRT graphical user interface (GUI) allows us to focus on fine-tuning our feature vectors without concerning ourselves with the technical aspects of selecting an appropriate machine-learning algorithm.

The GRT can accept a real-time data stream from another application via the Open Sound Control (OSC) protocol. Furthermore, once training data are recorded, configured, and trained to perform gesture recognition, the real-time prediction results can be streamed back to our application via OSC.
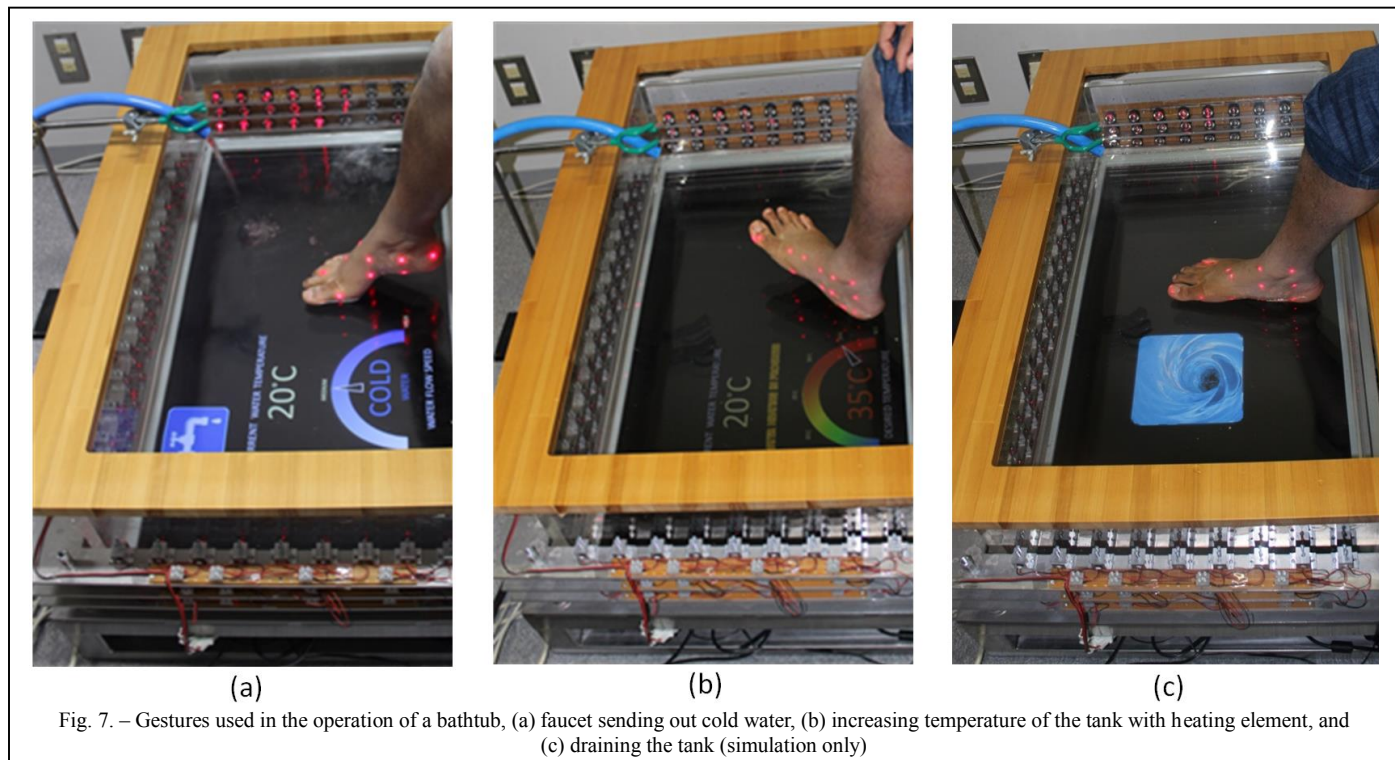
D. Application scenarios

One application is the operation of a bathtub. Foot gestures can be used to control a faucet using two gestures. A raised-heel gesture with the foot facing forward sends cold water to the tub (Fig. 7a). Once the foot is brought back to the resting position (foot-resting gesture), the faucet is closed. Similarly, a raised-toe gesture with the foot facing forward sends warm water into the tub. Note that in this demonstration, a water inlet solenoid valve is not implemented. The use of a solenoid valve would enable us to control the speed of the water flow, and the speed could be set to indicate the degree of foot tilt if desired. Another operation is to control the temperature of the water, for which a small electric water heater is used. The maximum temperature can be set by moving the raised toes to the right (Fig. 7b). Similarly, the maximum temperature can be lowered by moving the foot to the left. However, in this study, a technique to cool the water is not implemented. Finally, to drain the tank it would be possible to operate a valve triggered by maintaining the foot-resting position (Fig. 7c) for ten seconds, although this is not implemented in our GestureTank.

A bathtub or foot tub is an environment for relaxation; thus, the ability to play and listen to music is our second application. In this scenario, playing pre-recorded tracks such as MP3s or even music videos is envisaged. Another application scenario is the control of a music player in a bathtub or foot bath environment. Raising the toes with the foot facing forward can be used to indicate starting play. A raised-heel gesture can be set to indicate pausing play. Movement of the foot to the right and left with the foot touching the surface can indicate skipping a track forward and backward, respectively. Similarly, foot movement to the right and left with raised toes can indicate increasing and decreasing volume, respectively. Finally, resting the foot in position for ten seconds can stop play.

Our final application scenario allows a user to play his or her own songs by considering the tub as a musical instrument digital interface (MIDI) device. A note is controlled by the centroid position of the feet in the water. The pitch can be controlled by moving a foot to the right and left while touching the surface of the water. Similarly, foot movement to the right

(a) (b) (c)

Fig. 7. – Gestures used in the operation of a bathtub, (a) faucet sending out cold water, (b) increasing temperature of the tank with heating element, and (c) draining the tank (simulation only)

and left with raised toes can indicate increasing and decreasing volume, respectively. MIDI tones are generated via the MAX MSP application.
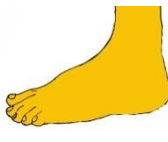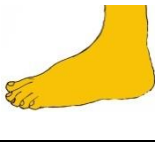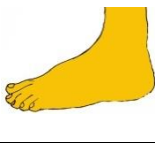
IV. TESTING
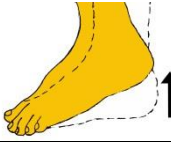
A. Usability Testing of Gestures

When considering day-to-day interactions with water, a number of gestures that are unique to interaction with water can be found. Scooping, paddling, and twirling are basic hand gestures, while washing hands is a more complex gesture [20]. When considering foot interaction, paddling [20] is a possible unique gesture.

Combining the basic foot gestures listed in Section II.A can result in a number of possible complex 3D foot gestures. Body gestures can cause fatigue over long-term use [11]; the literature survey did not find previous research that uses complex foot gestures. Therefore, an experiment was conducted to investigate the usability of the proposed gestures. The experiment was conducted in a laboratory environment with 17 participants (four female and 13 male) aged 18–60 years with foot anthropometric range of 22–27 cm. The participants were asked to place a foot in the tank and perform the gestures listed in Table 1 twice in a sequence of their own preference. When the task was completed, the participants were queried about whether they agree that performing each gesture is suitable, comfortable, and natural while relaxing the foot in water. The response was recoded using a binary response scale (yes/no).

TABLE 1 – POSSIBLE GESTURES INVESTIGATED

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)	

B. Feature Selection

In the next experiment, our objective was to select the features of the foot that result in the best classification of gestures. Again, the experiment was conducted in a laboratory environment with 17 participants (four female and 13 male) aged 18–60 with foot anthropometric range of 22–27 cm. Each participant was asked to perform the gestures discussed in Section IV.A in random order.

3D positional data pertaining to the foot gesture recognition were captured by the Arduino and sent as serial data via USB to the processing application. At the time of recording, the user informed the tester what gesture they were performing. A pre-processing module checked whether the data

contained any object matching the dimensions of a foot. If so, the data were recorded in a proprietary file format that stores them as a 2D matrix together with the gesture class given by the user. The same data expression scheme explained in Section III was applied; however, in this case, the matrix contains only the positional data pertaining to the foot and not data pertaining to the entire tank. This data was used by the GRT for training and testing.

As gesture recognition relies on the quality of the data input to the gesture recognition system, a key goal is to avoid the curse of dimensionality, i.e., to select the optimum number of features that can be used to identify the gestures uniquely. Considering the domain knowledge of foot shape and structure, 24 feature vectors pertaining to foot gesture recognition, such as length, width, and height, are considered. From the recorded gesture sample set, each feature vector was plotted against the classifying instances to estimate the optimal feature vectors that provide the best classification.

C. Gesture Recognition Performance

Different algorithms can be considered for recognition of gestures using features. As the GRT supports a number of algorithms, six algorithms were used for evaluation (ANBC, Adaboost, k-NN, MinDist, Softmax, and SVM). A ten-fold cross-validation accuracy test was performed on our training dataset. This testing was further beneficial to evaluate the parameters that provided the highest accuracy for each algorithm.

In the next stage, the models trained using the selected parameters were tested against the test dataset to evaluate which algorithms and gestures demonstrated the highest recognition rates.

V. Results and Discussion

A. Usability Testing of Gestures

The number of times a user agreed that each gesture was suitable for water interaction is summarized in Table 2. The agreement level is also indicated as a percentage. Seven out of eleven gestures were accepted as suitable for inclusion in our recognition system considering an agreement threshold of 70%.

TABLE 2 – USABILITY TESTING OF GESTURES

Gesture ID	1	2	3	4	5	6	7	8	9	10	11
No. of users who agree that the gesture is suitable (17 Users)	17	17	16	17	13	16	1	12	4	0	0
Agreement Level (%)	100	100	94	100	76	94	7	71	24	0	0

Eversion (Gesture 10) and inversion (Gesture 11) were rejected as suitable gestures by all participants. There seems to be a general consensus that moving the foot in such ways is not comfortable. The number of users who agreed that using Gestures 7 and 9 was also low, with only one and four users agreeing, respectively. The common feature between these two gestures is that they involve a raised heel. In fact, out of all the gestures, the ones in which the heel is raised had 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 anti-clockwise or clockwise further in this situation can be strenuous.

B. Feature Selection

A total of 17,177 samples representing the seven gestures selected were recorded. After pre-processing to remove duplicate values, our dataset contained 11,036 samples. When considered individually, none of the feature vectors provided satisfactory classification for the seven classes. By considering five features as the stopping criterion, a subset containing five features that combine to provide the best classification was chosen: 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 (N_{L3}). PointsinL3Back is the percentage of the number of foot regions in the bottommost row (L3Back) to N_{L3}. L1L2Right is the percentage of the total number of foot regions in the rightmost column in Layer 1 (L1Right) and Layer 2 (L2Right) compared with 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) to N. CheckLR provides a score of 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 8 shows the overhead and side views of a foot with the selected features highlighted. A graph showing the features is shown in Fig. 9.
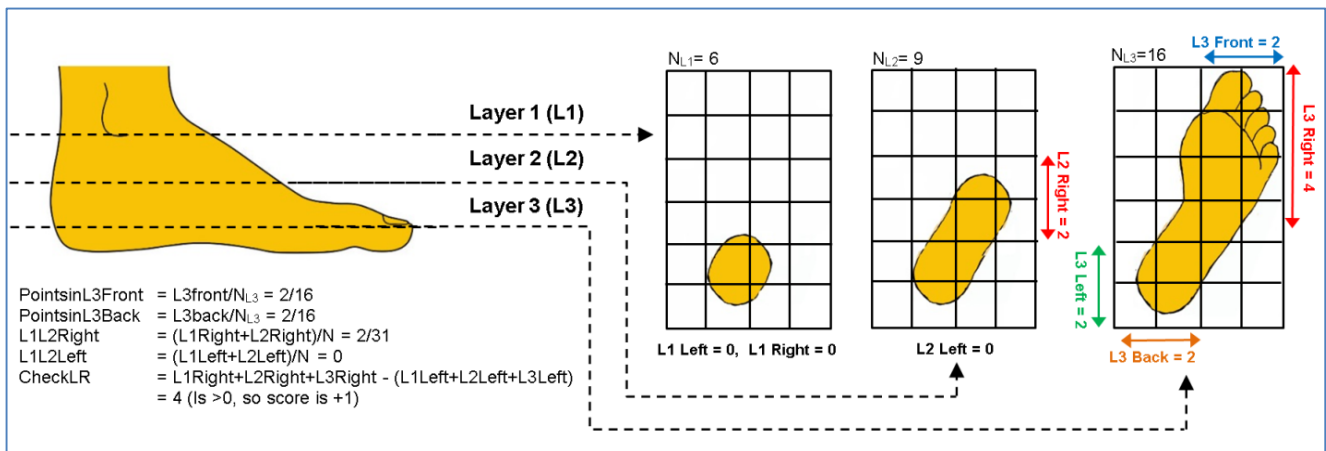
Fig. 8. Selected feature vectors

Fig. 9. Plot of the selected feature vectors

C. Gesture Recognition Performance

The same dataset used for feature selection was applied for this experiment. A ten-fold cross-validation accuracy test was conducted using a training dataset of 5,518 samples (50% of data). Table 3 shows the highest accuracy obtained and the key parameters for which the accuracy was obtained.

TABLE 3 – ALGORITHMS USED FOR TEN-FOLD 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

In the next stage, the remaining 50% of the samples were used as testing data, and the experiment was conducted for each algorithm to evaluate the recognition rates obtained using the GRT.

TABLE 4 – GESTURE RECOGNITION RATES

Gesture ID	1	2	3	4	5	6	8
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
%AdaBoost	95.49	100	100	96.53	92.78	100	95.45
% k-NN	88.12	94.5	92.32	86.84	90.85	96.1	97.00
% MinDist	87.74	97.52	87.61	95.87	92.78	100	93.13
% Softmax	94.78	89.35	87.61	96.09	92.78	100	89
% SVM	95.49	98.26	100	95.87	92.78	100	89

The total recognition accuracies for the algorithms were 88.5%, 96.64%, 88.12% 93.43%, 93.05%, and 95.19% for

ANBC, AdaBoost, k-NN, MinDist, Softmax, and SVM, respectively (Table 4). The best performing classifier for our data was AdaBoost with 20 boosting iterations and a null rejection coefficient of 3. Out of the gestures performed, AdaBoost classified Gestures 2, 3, and 6 with 100% accuracy, with the lowest performance for Gesture 5 at 92.78%. The SVM with a linear kernel, gamma of 0.1, and null rejection coefficient of 3 showed the second best performance. Of the gestures performed, the SVM classified Gestures 3 and 6 with 100% accuracy, with the lowest performance for Gesture 7 at 89%. Note that the k-NN algorithm detected Gesture 8 with a higher recognition rate (97%); however, it demonstrated a far lower recognition rate for all other gestures, with the lowest being 86.84% for Gesture 4. Furthermore, when considering the time required for each algorithm to train the 5,518 samples in our training dataset, AdaBoost took 12,100 ms, and the SVM required only 1,464 ms. However, for testing the testing dataset with 5,518 samples, AdaBoost required only 47 ms, while the SVM took 1,391 ms.

D. Discussion

It should be mentioned that, although the resolution of the sensing in our system setup is rather coarse (laser-phototransistor spacing of 5 cm horizontally and 3 cm vertically), the gesture recognition performance is fairly good. The spacing can be further improved using the current equipment. The resolution can be improved up to a maximum of 1 cm horizontal and 2 cm vertical spacing. Such improvement could allow us to evaluate the degree of movement for eversion, inversion, dorsiflexion, and plantar flexion. While the touch frame enables ghosting resolution, a slight positional error can be introduced if the foot is inserted at a slant due to the touch frame being located above the tank and the surface of the water. Although our centroid and other feature calculations are performed using the laser-phototransistor array, the positional error from the touch frame can affect detection if the feet are very close to each other.

VI. SUMMARY AND FUTURE WORK

This paper has presented the structure of the interactive GestureTank system, wherein water is used as a medium for interaction between the user and the system. Combinations of lasers and phototransistors are arranged at the four sides of a water tank, and the system detects the positions of objects, such as feet and hands, inserted into the water. Visual feedback is provided by an LCD monitor placed at the bottom of the tank. Auditory feedback is provided through speakers embedded in the LCD monitor. A thermal sensor and regulator are employed to detect and regulate the water temperature in the vessel. A touch frame is mounted at the top of the tank to assist in the elimination of ghost points, which can affect detection when two or more objects are present.

Several experiments were conducted as part of this study. During a user evaluation to select gestures for foot-based water interaction, seven gestures were selected with a minimum acceptance level of 70%. Using gestures recorded in the system from a user sample, five features pertaining to the foot were selected to represent the gesture. A subsequent experiment

tested the detection performance for each of the seven gesture classes. Despite the coarseness of the system resolution, good detection performance was obtained for AdaBoost recognition. Further study of object position detection and gesture recognition remain to make the system more practical.

At present, there are no readily available gesture detection products for use in water environments. However, arm-worn devices such as Myo can be affixed to the calf muscle to detect foot movements. This study can be extended to recognize more complex foot gestures, such as paddling and tapping.

REFERENCES

[1] "Microsoft Kintect," 2015. [Online]. Available: http://www.xbox.com/en-US/kinect. [Accessed: 09-May-2015].

[2] "SoftKinectic Depthsense Camera," 2015. [Online]. Available: http://www.softkinetic.com/Products/DepthSenseCameras . [Accessed: 09-May-2015].

[3] "Leapmotion Hardware Sensor," 2015. [Online]. Available: https://www.leapmotion.com/product . [Accessed: 09-May-2015].

[4] "Ring Control Device," 2015. [Online]. Available: http://logbar.jp/ring/en . [Accessed: 05-Apr-2015].

[5] "MYO Armband," 2015. [Online]. Available: https://www.thalmic.com/en/myo . [Accessed: 15-Apr-2015].

[6] T. Pakkanen and R. Raisamo, "Appropriateness of Foot Interaction for Non-accurate Spatial Tasks," in *Proceedings of the CHI '04 Extended Abstracts on Human Factors in Computing Systems*, 2004, pp. 1123–1126.

[7] G. Pearson and M. Weiser, "Of moles and men: the design of foot controls for workstations," *ACM SIGCHI Bulletin*, vol. 17, no. 4, pp. 333–339, 1986.

[8] S. Lenman, L. Bretzner, and B. Thuresson, "Computer vision based hand gesture interfaces for human-computer interaction," 2002.

[9] T. Ito, "Hand Gesture-Based Manipulation of a Personalized Avatar Robot in Remote Communication," in *Human Interface and the Management of Information. Interacting with Information*, vol. 6771, M. Smith and G. Salvendy, Eds. Springer Berlin Heidelberg, 2011, pp. 425–434.

[10] M. H. P. H. van Beurden, W. A. Ijsselsteijn, and Y. A. W. de Kort, "User experience of gesture based interfaces: A comparison with traditional interaction methods on pragmatic and hedonic qualities," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 7206, Springer Berlin Heidelberg, 2012, pp. 36–47.

[11] W. Yee, "Potential Limitations of Multi-touch Gesture Vocabulary: Differentiation, Adoption, Fatigue," in *Human-Computer Interaction. Novel Interaction Methods and Techniques SE - 32*, vol. 5611, J. Jacko, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 291–300.

[12] J. Scott, D. Dearman, K. Yatani, and K. N. Truong, "Sensing Foot Gestures from the Pocket," in *Proceedings of the 23rd Annual ACM Symposium on User Interface Software and Technology(UIST '10)*, 2010, pp. 199–208.

[13] J. Alexander, T. Han, W. Judd, P. Irani, and S. Subramanian, "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)*, 2012, pp. 1229–1238.

[14] M. Sato, I. Poupyrev, and C. Harrison, "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)*, 2012, pp. 483–492.

[15] E. Arroyo, L. Bonanni, and N. Valkanova, "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), 2012, pp. 685–688.

[16] C. Tanabe and M. Hirakawa, "O-Key: A Scoopable Interface," in *Proceedings of the Eighth IEEE International Symposium on Multimedia (ISM'06)*, 2006, pp. 185–192.

[17] M. Ikeda, N. Nagira, and M. Hirakawa, "A Multi-Dip Interface with Water," in *Proceedings of the International Conference on Complex, Intelligent and Software Intensive Systems*, 2009, pp. 169–176.

[18] M. Ikeda and M. Hirakawa, "A Multi-Dip Interface with Water," *Journal of Information Processing*, vol. 51, no. 11, pp. 2103–2111, 2010.

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

[20] K. S. L. Gunawardena, K. Kimura, and M. Hirakawa, "SensorTank: An Interactive Water Vessel," *Journal of Information Processing*, vol. 22, no. 3, pp. 547–555, 2014.

[21] R. T. Floyd, *Manual of Structural Kinesiology*, 17th ed. McGraw-Hill, 2008.

[22] M. Nordin and H. Frankel, Victor, *Basic Biomechanics of the Musculoskeletal System*, 4th ed. LWW, 2012.

[23] P. Dhawale, M. Masoodian, and B. Rogers, "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), 2006, pp. 25–32.

[24] Y. Takeoka, T. Miyaki, and J. Rekimoto, "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)*, 2010, pp. 90–94.

[25] S. Malik and J. Laszlo, "Visual Touchpad: A Two-handed Gestural Input Device," in *Proceedings of the 6th International Conference on Multimodal Interfaces (ICMI '04)*, 2004, pp. 289–296.

[26] N. Sangsuriyachot and M. Sugimoto, "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)*, 2012, pp. 21–28.

[27] "Nintendo Wii," 2015. [Online]. Available: http://www.nintendo.com/wii . [Accessed: 09-May-2015].

[28] T. Schlömer, B. Poppinga, N. Henze, and S. Boll, "Gesture Recognition with a Wii Controller," in *Proceedings of the 2nd International Conference on Tangible and Embedded Interaction (TEI '08)*, 2008, pp. 11–14.

[29] M. Hoffman, P. Varcholik, and J. J. LaViola, "Breaking the status quo: Improving 3D gesture recognition with spatially convenient input devices," in *Proceedings of IEEE Virtual Reality (VR) Conference 2010*, 2010, pp. 59–66.

[30] K. Barczewska and A. Drozd, "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)*, 2013, pp. 207–210.

[31] S. Gupta, D. Morris, S. Patel, and D. Tan, "SoundWave: Using the Doppler Effect to Sense Gestures," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2012, pp. 1911–1914.

[32] D. Mace, W. Gao, and A. Coskun, "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)*, 2013, pp. 83–84.

[33] P. Ziaie, T. Müller, M. E. Foster, and A. Knoll, "A Naïve Bayes Classifier with Distance Weighting for Hand-Gesture Recognition," in *Advances in Computer Science and Engineering SE - 38*, vol. 6, H. Sarbazi-Azad, B.

Parhami, S.-G. Miremadi, and S. Hessabi, Eds. Springer Berlin Heidelberg, 2009, pp. 308–315.

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

[35] A. K. Nimbalkar, R. R. Karhe, and C. S. Patil, "Face and Hand Gesture Recognition using Principle Component Analysis and kNN Classifier," *International Journal of Computer Applications*, vol. 99, no. 8, pp. 26–28, Aug. 2014.

[36] C. C. Wang and K. C. Wang, "Hand Posture Recognition Using Adaboost with SIFT for Human Robot Interaction," in *Recent Progress in Robotics: Viable Robotic Service to Human SE - 25*, vol. 370, S. Lee, I. Suh, and M. Kim, Eds. Springer Berlin Heidelberg, 2008, pp. 317–329.

[37] N. H. Dardas and N. D. Georganas, "Real-Time Hand Gesture Detection and Recognition Using Bag-of-Features and Support Vector Machine Techniques," *IEEE Transactions on Instrumentation and Measurement*, vol. 60, no. 11, pp. 3592–3607, Nov. 2011.

[38] J. Oh, T. Kim, and H. Hong, "Using binary decision tree and multiclass SVM for human gesture recognition," in *2013 International Conference on Information Science and Applications, (ICISA)*, 2013, pp. 1 – 4.

[39] F.-S. Chen, C.-M. Fu, and C.-L. Huang, "Hand gesture recognition using a real-time tracking method and hidden Markov models," *Image and Vision Computing*, vol. 21, no. 8, pp. 745–758, Aug. 2003.

[40] R. W. Hut, S. V. Weijs, and W. M. J. Luxemburg, "Using the Wiimote as a sensor in water research," *Water Resources Research*, vol. 46, no. 12, pp. 1–5, Dec. 2010.

[41] R. W. Hut, "Scanning dynamic water surfaces using a Kinect as a sensor," *Geophysical Research Abstracts*, vol. 13, no. EGU2011–10096, 2011.

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

[43] Y. Matoba, Y. Takahashi, T. Tokui, S. Phuong, S. Yamano, and H. Koike, "AquaTop Display: A True 'Immersive' Water Display System," in *ACM SIGGRAPH 2013 Emerging Technologies*, 2013, pp. 4:1–4:1.

[44] S. Mann, R. Janzen, and J. Huang, "'WaterTouch': An Aquatic Interactive Multimedia Sensory Table Based on Total Internal Reflection in Water," in *Proceedings of the 19th ACM International Conference on Multimedia*, 2011, pp. 925–928.

[45] J. Watanabe, "VortexBath: Study of Tangible Interaction with Water in Bathroom for Accessing and Playing Media Files," in *Human-Computer Interaction. Interaction Platforms and Techniques SE - 136*, vol. 4551, J. Jacko, Ed. Springer Berlin Heidelberg, 2007, pp. 1240–1248.

[46] J. T. K. V. Koh, K. Karunanayaka, J. Sepulveda, M. J. Tharakan, M. Krishnan, and A. D. Cheok, "Liquid Interface: A Malleable, Transient, Direct-touch Interface," *Computers in Entertainment*, vol. 9, no. 2, pp. 7:1–7:8, Jul. 2011.

[47] J. Y. Han, "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)*, 2005, pp. 115–118.

[48] "Arduino Uno Microcontroller," 2015. [Online]. Available: http://arduino.cc/en/Main/ArduinoBoardUno . [Accessed: 09-May-2015].

[49] N. E. Gillian and J. A. Paradiso, "The Gesture Recognition Toolkit," *Journal of Machine Learning Research*, vol. 15, pp. 3483–3487, 2014.