Hand Gesture Recognition Based on Near-Infrared Sensing Wristband

ABSTRACT

Wrist-worn gesture sensing systems can be used as a seamless interface for AR/VR interactions and control of various devices. In this paper, we present a low-cost gesture sensing system that utilizes near Infrared Emitters (600 - 1100 nm) and Photo-Receivers encompassing the wrist to infer hand gestures. The proposed system consists of a wristband comprising Infrared emitters and receivers, data acquisition hardware, data post-processing software, and gesture classification algorithms. During the data acquisition process, 24 near Infrared Emitters are sequentially switched on around the wrist, and twelve Photo-diodes measure the light reflected, refracted, and scattered by the tissues inside the wrist. The acquired data corresponding to different gestures are labeled and input into a machine learning algorithm for gesture classification. To demonstrated the accuracy and speed of the proposed system, real-time gesture sensing user studies were conducted. As a result of this comparison, we obtained an average accuracy of 98.06% with standard deviation of 1.82%. In addition, we evaluated that the system can perform six-eight gestures per second in real time using a desktop computer operating with Core i7-7800X CPU at 3.5GHz and 32 GB RAM.

Author Keywords

Hand Gesture; NIR; Human-machine Interaction(HCI); Bio-sensing; Virtual-Reality; Wearable sensing

INTRODUCTION

Hand gesture recognition refers to the problem of identifying hand gestures executed by a user at a specific time. Humans naturally gesticulate with their hands forming both static hand poses and dynamic gestures to deliver information. For this reason, hand gesture sensing or recognition have long been studied for intuitive control of interactive systems, as well as in many other engineering and medical applications [1, 2, 3]. Some typical such applications include human-machine interaction interfaces, control of hand prostheses and rehabilitation devices, sign language interpretations [4, 5, 6].

In order to design all-day-wearable gesture sensing devices, the following requirements are usually considered: the devices should be non-obtrusive, they should not cause physical discomfort or encumbrances to the natural hand movement; moreover, they should also be intuitive and easily accessible. To meet all such criteria, a wrist-worn device is a great candidate. Therefore, there has been several bio-sensing research that seeks to infer gestures from tracking the anatomical changes within the wrist.

Several wrist-worn devices have been proposed for hand gesture recognition. The sensing modalities include camera-based systems[7], inertial motion sensing [8, 9], Electromyography(EMG) [10, 11, 12, 13], Electrical Impedance Tomography (EIT) [14, 15], and capacitive and resistive pressure sensing systems [16, 11, 17]. Each modality has its own merits and limits. **Wearable camera systems** attach small cameras near the wrist to recognize different hand shapes. For example, Digits [7] uses a 3D infrared camera to identify gestures using machine vision systems. However, some significant limitations of this type of sensing include line-of-sight occlusions, ambient light noise, and higher computational cost associated with more complicated imaging processing algorithms. **Inertial motion sensing systems** employ inertial measurement units (IMUs), which consist of accelerometers, gyroscopes, and magnetometers, to measure arm and finger orientations [8]. Accelerometers data has also been used to recognize different activities by sensing dynamic features related to hand motions [9]. However, Inertial motion sensors are very limited at detecting static hand gestures. **Electromyography (EMG)** estimates the myoelectric potential generated during hand and finger movements by attaching electrodes to the upper part of the forearm. It has been extensively explored for static and dynamic gesture detection [10, 12]. Compared to other sensing techniques, the limitations of EMG systems include the requirement of massive datasets and the heavy computation burden associated with extensive signal processing. In addition, they also require careful initial configuration and calibration for adequate performance. **Electrical impedance tomography (EIT)** is another well-studied method for hand gesture recognition [15]. It measures the impedance changes between pairs of electrodes to track the wrist tissue changes. However, this type of method is susceptible to resistance coupling between the electrode and skin, and sometimes require electric-conductive gel for stable coupling. Other sensing modalities include **Force sensing resistors (FSRs) system**, which measures the pressure distribution around the wrist to identify different static gestures [16], and Capacitive pressure sensors, such as GestureWrist [2, 17], which measures capacitive changes around the wrist are also used. Most of the gesture sensing methods mentioned above are unsuitable for practical use due to either low accuracy, high cost, poor ergonomics, portability, or ease of use.

ACM ISBN 978-1-4503-2138-9.
DOI: 10.1145/1235

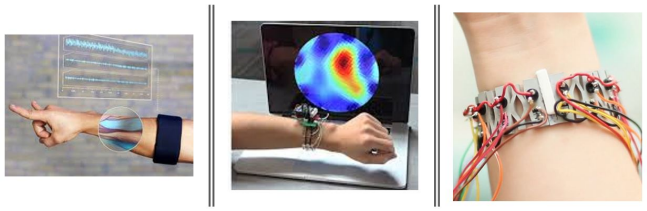

Figure 1. Gesture sensing technologies: Electromyography (EMG), Electrical Impedance Tomography (EIT)[15], Near Infrared Sensing(NIR) [18], From Left to Right

NEAR INFRARED(NIR) SYSTEM

Near-infrared (NIR) systems have been long investigated in the area of medical applications, because of their ability to image changes in tissues [19, 20]. The near-infrared window is the light spectrum in the wavelength range between 600nm to 1100nm. As shown in Figure 2, Light has maximum penetration in tissues in Near Infrared window. The absorption coefficient of water, melanin, de-oxygenated, and oxygenated blood are low in NIR window. Photons that enter human tissue typically undergo absorption, scattering, and reflection. Changes in the anatomical structures of the wrist while performing different gestures will result in changes in how injected light interacts with tissues, including the three types of incidents mentioned above. These changes can be utilized to detect hand gestures by measuring the transmission and diffraction of light through the wrist.

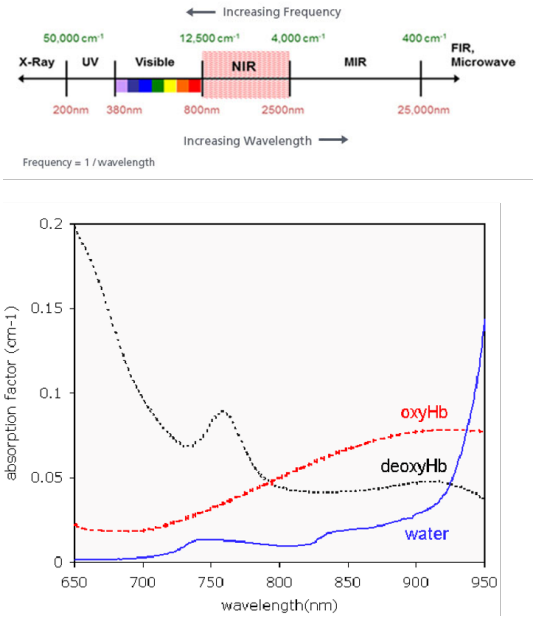

Figure 2. Absorption rate of NIR for different tissues [21]

This project uses wristband which consists of pairs of infrared emitters and receivers to measure the light reflected or scattered through/off the wrist. There have been some previous works [18], that demonstrated NIR's potentials for non-invasive and accurate gestures recognition. In our study, We explore further to improve the accuracy of the NIR system in detecting gestures by developing a system which is robust against skin-sensor coupling. This study also explores the impact of the number and configuration of the sensors on the overall accuracy of the system. Two different wavelengths were also tested to see the variation in accuracy at different light absorption rates of tissues.

NIR WRISTBAND IMPLEMENTATION

Wristband Hardware

The NIR wristband (shown in Figure 4) is composed of 24 near-infrared emitters, 12 photo-diodes with an on-chip trans-impedance amplifier. A controller is used to switch the IR LEDs and sample the photodiodes output sequentially. The sampled data is then used by a machine learning algorithm to classify features into different gestures.

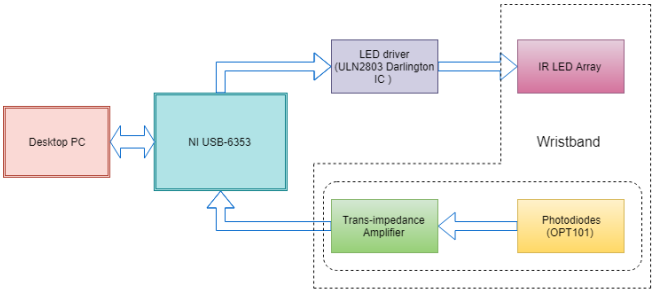

Figure 3. Block Diagram of NIR wristband

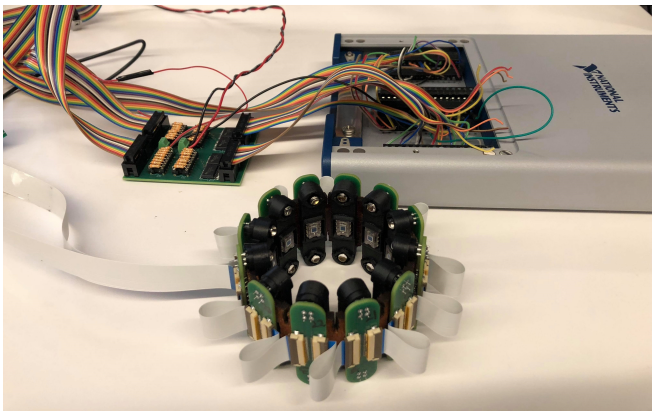

Figure 4. Wristband hardware setup: The data acquisition and IR LED control hardware

The selection of emitters consider factors including power consumption, optical power, and emitter-detector spacing on the wristband. After testing different IR emitters to see the effect of radiant intensity (12 - 1100 mW/sr) and light beam angle ($6-130°$) and emission wavelength on the received signal, we chose MTMD7885NN24 multi-chip IR emitters that have peak emission wavelengths at 770nm, 810nm, and 850nm. We chose the photodiodes to be TI's OPT101 with an on-chip trans-impedance amplifier and 1 Mohm feedback resistor creating a bandwidth of 14kHz. The built-in trans-impedance amplifier reduces noise pick-up and space requirement, compared to an external amplifier.

To generate a timed and synchronized PWM control signal sequence, a NI DAQ(USB-6353) device was employed. The

Figure 5. The NIR Wristband

Figure 6. Single IR emitter and receiver module

PWM sequence was generated at a 20Hz frequency, with 4% duty cycle and 15 degrees phase shift of consecutive channels. It creates a sequential switching of all emitters in 50 ms duration. A Darlington array ICs, ULN2803a (capable of sinking up to 500 mA, 50V) was used to amplify the control signal. The cascade connection of two transistors in Darlington arrays creates the effect of a single transistor with a very high current gain. The very high β allows for high output current drive with a very-low input current, essentially equating to operation with low GPIO voltages. Current limiting resistors were utilized to control the amount of current drawn by the emitters. The acquired analog signal from each photodiode amplifier was sampled at 16-bit resolution and 9600 samples per second.

Compared to SenseIR [18], this wristband hardware is developed with more IR emitters, i.e., 24 IR emitters and a better mechanical wristband design. This improvement increases the relative prediction accuracy because the rigidity if the wristband improves the robustness of the system to misalignment errors. A careful analysis of the optical power with respect to emitter-receiver spacing and IR beam angle has also been done. The wristnband hardware also features PCB connectors (as shown in Figure 6) which are modular and durable. Not only this reduces the cumbersome cabling needed but also reduces noise pickups.

Machine Learning Software

The software of this study consists of four major components: data collection, data processing, model training, and both offline and real-time evaluation.

Data Collection

During this phase, the PWM waveform generation and emitter control session are run continuously in the background and scan the full 360 degrees of the wrist at an effective frequency of 20Hz. During each scanning of 360 degrees, 24 emitters are pulsed in sequence, each generating a finite width rectangular waveform.

The rising edge of emitter 1 triggers the analog data acquisition and lasts for exactly 50ms. Since the analog signal is sampled at 9600 samples/sec, 480 samples are acquired per channel in every full wrist scan, which is referred to as one 'frame'. An example of such a frame of data is shown in Figure 7

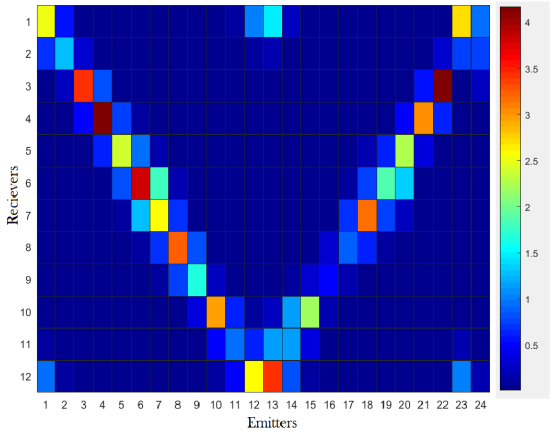

Figure 7. Matrix of Emitter-receiver measurements of a single frame

Data Processing

The data collected from each photodiode (channel) are first filtered using a median filter of width five, to suppress some narrow-width noises from external near-infrared interference. The resulting data for each channel, which is 480 samples, are then segmented into 24 segments, each corresponding to the time when a specific emitter has been switched on. Because we have observed that each 20-sample segment appear to be almost flat in our experiment, we computed the mean value as a representation for each segment to reduce the size of data by a factor of 20. In this way, the 480 samples per channel can be reduced to 24 samples, and a total of 12 channels sum up to 288 samples per frame. These processed data were input into both the model training and real-time evaluation steps in this study.

Model Training

For several applications, wearable gesture recognition systems are required to function in real time with comparable accuracy to those in offline modes. For a gesture recognition system to operate in real-time, it has to recognize a gesture in less than 300 ms equivalent to 3Hz update rate [10]. Since wearable systems run on systems which needs low computational complexity and low power systems, the primary challenge here is choosing a classifier which can exhibit good performance using less complicated recognition models.

The labeled and processed data from the Data Processing step are employed to train various machine learning models. We have explored the following supervised learning classifiers: k-Nearest Neighbors (kNN), Support Vector Machine (SVM), Linear Discriminant Analysis(LDA), and Neural Networks (NN). After a series of accuracy comparison for different models, a shallow neural network worked best for our context. Therefore we will focus on describing this model. The shallow neural network that we chose was a single-layer fully connected network, with the hidden layer consisting of 56 hidden nodes. The activation function was set to be the rectified linear unit (ReLU) function, and the cost function was a negative log-likelihood function. During training, one epoch of data was divided into batches of the size of 600, and a total of 1000 epochs were trained, after which the changes in both accuracy and cost-function value falls below a small threshold. No dropout or batch normalization was applied during training, because this model is relatively simple, and it already generalized pretty well without these tricks.

Model Evaluation

We performed both offline evaluation of the trained model using the collected data, also real-time evaluation by applying the trained model on real-time streamed sensor data.

1. For **offline evaluation**, a typically collected dataset usually consists of 10 trials. 10-fold cross-validation was performed by training on any nine trials and validating on the other one trial of data. Both validation accuracy and confusion matrices were averaged over all ten folds and served as a metric of gesture classification performance.

2. For **real-time evaluation**, a real-time software streamed frames of raw sensor data from the wristband hardware, and performed the same data-processing procedure on each frame, and invoked model inference employing the trained model to predict the gesture of the current frame. The real-time evaluation software can update a new gesture at 6-8Hz.

USER STUDY DESIGN

To assess the performance of the built wristband, hand gesture sensing studies were conducted across multiple subjects. To better compare our results with a previous study done by SenseIR [18], we designed the studies to replicate the set-ups in SenseIR as much as possible. Specifically, a total of four pinch gestures (Index Pinch, Middle Pinch, Ring Pinch, and Little Pinch), six common gestures (Fist, Spread, Call, Gun, Index Point, and Thumbs up), and three wrist gestures (Wrist Flexion, Wrist Extension, and Wrist Abduction) were chosen, as shown in Figure 8.

Figure 8. Set of gestures used in the user study

All participant wore the wristband on their right arm. Before the test, each subject was given some time to practice the gestures. During a single trial, the subject was prompt to perform all 13 gestures in a random sequence specified by the software, each gesture for 5 seconds. The randomness in gesture sequence can enhance the generalization of the machine learning model.

A total of 10 trials of data were collected for each subject. Data corresponding to the first 2 seconds of a new gesture was truncated, because we observed that the raw sensor signals would fluctuate to difference extents during this transitioning time, this truncation can provide more stabilized signals for each gesture. The labels of the data were simultaneously created and time stamped.

RESULTS

In our study, the confusion matrix was computed to serve as the metric of performance. The confusion matrix is a $N_{gesture}$ by $N_{gesture}$ matrix, each entry is strictly between 0 and 100, with the i-th column and j-th row indicating the percentage of i-th gesture (true) being classified as the j-th gesture, during validation. Its diagonal elements imply the percentage of each gesture being classified correctly, whose average value is a quantitative measure for the performance of the trained model, therefore denoted as "validation accuracy" for the rest of the study. The confusion matrix is always averaged over all folds during the 10-fold cross-validation. The confusion matrix (shown in Figure 9) corresponds to the confusion matrix averaged over all ten subjects, and all ten folds for each subject. The validation accuracy over all subjects is 98.06% with the standard deviation of $\pm$ 1.82%.

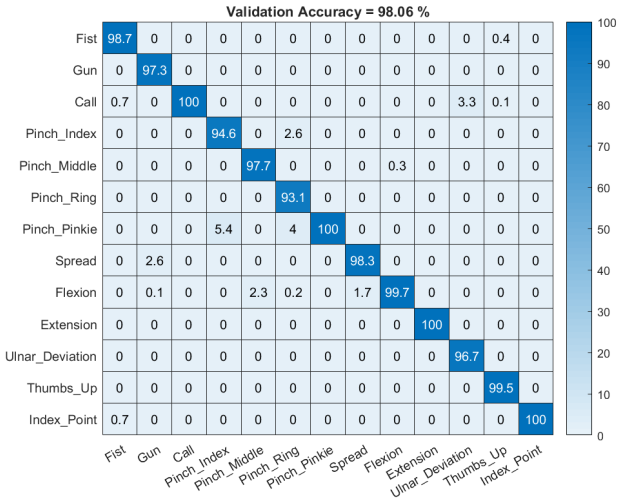

Figure 9. Confusion Matrix - The mean validation accuracy across 10 participants for each gesture

The prediction accuracy is observed to be less for less-pronounced gestures. i.e., pinch gestures and other gestures which only involve the movement of a single finger. This is because the pinch gestures use common muscle group. However, wrist gestures, gestures which rotate the whole hand around the wrist joint are easily recognized by the ML algorithms. In addition, other experiments were designed to investigate the impact of the number of sensors, sensor coverage area and density, peak emission wavelength of the IR emitters, and motion artifacts (particularly arm rotations) on the sensing performance:

1. Number of sensors: From a dataset collected on the full setup (24 emitters and 12 detectors), we removed some data to create combination setups with different number of emitters (6, 12, 24) and detectors (6, 12) as shown in Figure 10, Where the blue circles indicate the active IR emitters and the blue rectangles indicate the active IR receivers. We only reduce the numbers of emitters and detectors in the new setups, both are still equally spaced around the wrist. The results show a small decrease in accuracy for less number of sensors. However, the small variation in accuracy for different emitter and receiver combination shows that it is still possible to get a reasonably high accuracy with less number of sensors.

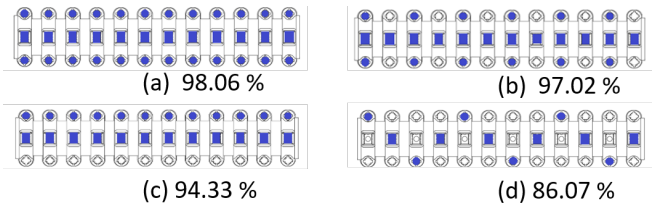

Figure 10. Number of sensors vs Validation accuracy

2. Peak emission wavelengths: Two wavelengths, 770 nm, and 850 nm, are tested in an experiment with a 12-emitter-12-detector configuration. The resulting validation accuracies for the two wavelengths are 94.33% and 93.37%, respectively. The changes in accuracy are not significant between the chosen two wavelengths.

3. Sensor coverage area: For hand gestures, more useful signals may concentrate on specific regions of the wrist. Sensor configuration, particularly sensor density in some regions of the wrist and the overall coverage area of the wristband should also be carefully designed to keep the trade-off between the accuracy of the system and wristband complexity or cost. In Figure 11, we showed the variation in accuracy for different sensor configuration of 12 emitters and 6 receivers wristband. This result shows that the sensors can be concentrated to a particular side of the wrist without a significant variation in accuracy. The sensors can also be arranged in a low-density configuration to include other complementary types of sensor such as EMG, inertial and pressure sensors which can be used to reduce the practicality issues related with the NIR wristband. However, the results also show covering more wrist area with sensors can give slightly better results than having sensors concentrated in a particular area.

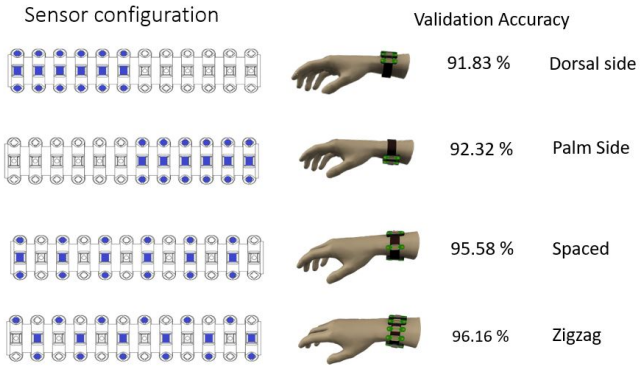

Figure 11. Different sensor configurations Vs Validation accuracy

4. Arm rotation: We have found that the rotation of the arm introduces a relatively significant change in the near-infrared signals. Therefore a significant error may be observed if the arm rotation angles during training and validation are different. This problem can be solved by introducing various arm orientations during training or including the orientation information from additional sensors during both training and validation. We performed a preliminary study in which data corresponding to three different arm rotations (0°, 45°, 90°) were collected (as shown in Fig. 12) Validation accuracies were computed by training on the 0° data and validating on the 45 ° and 90° data. The accuracy of the system reduces to 74% and 68% for arm rotation of 45° and 90° respectively.

DISCUSSION

From the experiments conducted and observations during those experiments, we distinguished some factors that affect the model training and recognition prediction algorithms.

- Misalignment of the IR sensors: This problem mainly happens for multiple session tests when the user takeoff the

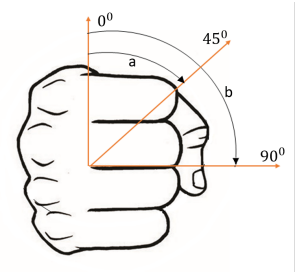

Figure 12. Hand or fore-arm Rotation Test

wristband and wear it again. However, the shift can also happen while the user wears the wristband in single session tests. The misalignment can happen either by rotational displacement or longitudinal displacement on the wrist. The position of the sensors will be displaced from the training position causing the shift for the order of the features. Since the wristband is sensitive for a shift in millimeters, manual calibration is very difficult. Every time the wristband shifts, the position/orientation of the wristband with respect to a reference position should be calculated. The reference position can be recognized using a specific gesture which shows strong features or a significant change in the signal. After the position is detected, reassignment of the channels can be used to maintain the order of the features without rotating the wristband.

- Arm movement: motion artifacts are the other major problem in NIR systems. The motion of the hand (arm rotation, elongation, elevation) causes blood movement and muscle/tendon movement. These movements cause an unwanted change in the signal. It is difficult to solve this problem entirely by including the arm movement information in the study or machine learning training procedures. However, it can be reduced by training the different arm rotations and elevations for the same gesture. An additional sensors output such as IMU orientation data can also be used to compensate for the arm rotation problem.

- Skin-sensor Coupling: Even though NIR wristband does not suffer from any electrical sensor-skin coupling unlike EMG and EIT techniques, They still suffer from problems caused by mechanical coupling between the sensor and the skin. This is mainly because light can be reflected directly from the skin without entering the skin tissue, thereby causing the saturation of photo-diodes. This problem can be reduced by using an appropriate mechanical wristband design which can keep the coupling constant.

- IR interference: IR light from external IR sources, e.g., The sun, IR illuminating cameras adds noise on the main signal. In order to remove these noises, enclosing the outer part of the wristband with IR block film can be a potential solution. The other solution can be a differential measurement of the external IR and subtraction from the main signal.

CONCLUSION

This paper presents a wrist-worn gesture sensing system that consists of an array of an Infrared Emitter and Photo-Receivers that are used to detect gestures by measuring reflected and refracted light from tissues in or under the skin. In this study, we have demonstrated that Near-infrared wristbands can offer a low-cost and high accuracy gesture sensing possibility. With the advancement of ultra-miniature SMD IR emitters and receivers, these techniques can be easily integrated to wrist-worn devices such as smart-watches and Fitbit monitors. The system's software consists of data acquisition, preprocessing and classification stages. Thirteen classes of gestures were analyzed to validate the accuracy of the classification algorithm. Future works include data collection and processing for cross-session and cross-user performance. Integration with orientation sensors and pressure sensors should be investigated as a potential solution to enable recognition of arm rotation and pressure distribution around the wrist. Development of embedded data acquisition and wireless data transfer methods can also be implemented for entirely wearable and wireless wristband gesture sensing system.

ACKNOWLEDGMENTS

REFERENCES

1. Rui Fukui, Masahiko Watanabe, Masamichi Shimosaka, and Tomomasa Sato. Hand-shape classification with a wrist contour sensor: Analyses of feature types, resemblance between subjects, and data variation with pronation angle. *The International Journal of Robotics Research*, 33(4):658–671, 2014.

2. Jun Rekimoto. Gesturewrist and gesturepad: Unobtrusive wearable interaction devices. In *Wearable Computers, 2001. Proceedings. Fifth International Symposium on*, pages 21–27. IEEE, 2001.

3. Frederic Kerber, Markus Löchtefeld, Antonio Krüger, Jess McIntosh, Charlie McNeill, and Mike Fraser. Understanding same-side interactions with wrist-worn devices. In *Proceedings of the 9th Nordic Conference on Human-Computer Interaction*, page 28. ACM, 2016.

4. Charles C Peck. Gesture sensing split keyboard and approach for capturing keystrokes, October 7 2003. US Patent 6,630,924.

5. Andrea Colaço, Ahmed Kirmani, Hye Soo Yang, Nan-Wei Gong, Chris Schmandt, and Vivek K Goyal. Mime: compact, low power 3d gesture sensing for interaction with head mounted displays. In *Proceedings of the 26th annual ACM symposium on User interface software and technology*, pages 227–236. ACM, 2013.

6. William T Freeman and Craig D Weissman. Hand gesture machine control system, January 14 1997. US Patent 5,594,469.

7. David Kim, Otmar Hilliges, Shahram Izadi, Alex D Butler, Jiawen Chen, Iason Oikonomidis, and Patrick Olivier. 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*, pages 167–176. ACM, 2012.

8. Hongyi Wen, Julian Ramos Rojas, and Anind K Dey. Serendipity: Finger gesture recognition using an

off-the-shelf smartwatch. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 3847–3851. ACM, 2016.

9. Gierad Laput, Robert Xiao, and Chris Harrison. Viband: High-fidelity bio-acoustic sensing using commodity smartwatch accelerometers. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, pages 321–333. ACM, 2016.
10. Marco E Benalcázar, Andrés G Jaramillo, A Zea, Andrés Páez, Víctor Hugo Andaluz, et al. Hand gesture recognition using machine learning and the myo armband. In *Signal Processing Conference (EUSIPCO), 2017 25th European*, pages 1040–1044. IEEE, 2017.
11. Jess McIntosh, Charlie McNeill, Mike Fraser, Frederic Kerber, Markus Löchtefeld, and Antonio Krüger. Empress: Practical hand gesture classification with wrist-mounted emg and pressure sensing. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 2332–2342. ACM, 2016.
12. Ali Boyali and Naohisa Hashimoto. Spectral collaborative representation based classification for hand gestures recognition on electromyography signals. *Biomedical Signal Processing and Control*, 24:11–18, 2016.
13. Yangjian Huang, Weichao Guo, Jianwei Liu, Jiayuan He, Haisheng Xia, Xinjun Sheng, Haitao Wang, Xuetao Feng, and Peter B Shull. Preliminary testing of a hand gesture recognition wristband based on emg and inertial sensor fusion. In *International Conference on Intelligent Robotics and Applications*, pages 359–367. Springer, 2015.
14. Yang Zhang, Robert Xiao, and Chris Harrison. Advancing hand gesture recognition with high resolution electrical impedance tomography. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, pages 843–850. ACM, 2016.
15. Yang Zhang and Chris Harrison. Tomo: Wearable, low-cost electrical impedance tomography for hand gesture recognition. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*, pages 167–173. ACM, 2015.
16. Artem Dementyev and Joseph A Paradiso. Wristflex: low-power gesture input with wrist-worn pressure sensors. In *Proceedings of the 27th annual ACM symposium on User interface software and technology*, pages 161–166. ACM, 2014.
17. Hoang Truong, Phuc Nguyen, Nam Bui, Anh Nguyen, and Tam Vu. Low-power capacitive sensing wristband for hand gesture recognition. In *Proceedings of the 9th ACM Workshop on Wireless of the Students, by the Students, and for the Students*, pages 21–21. ACM, 2017.
18. Jess McIntosh, Asier Marzo, and Mike Fraser. Sensir: Detecting hand gestures with a wearable bracelet using infrared transmission and reflection. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*, pages 593–597. ACM, 2017.
19. Kristian P Nielsen, Lu Zhao, Jakob J Stamnes, Knut Stamnes, and Johan Moan. The optics of human skin: Aspects important for human health. *Solar radiation and human health*, 1:35–46, 2008.
20. Joseph Chaiken, Bin Deng, Jerry Goodisman, George Shaheen, and Rebecca J Bussjaeger. Analyzing near-infrared scattering from human skin to monitor changes in hematocrit. *Journal of biomedical optics*, 16(9):097005, 2011.
21. Meltem Izzetoglu, Scott C Bunce, Kurtulus Izzetoglu, Banu Onaral, and Kambiz Pourrezaei. Functional brain imaging using near-infrared technology. *IEEE Engineering in Medicine and Biology Magazine*, 26(4):38, 2007.