PPG-based Finger-level Gesture Recognition Leveraging Wearables

Tianming Zhao*, Jian Liu†, Yan Wang*, Hongbo Liu‡ and Yingying Chen†
*Binghamton University, Binghamton, NY 13902 tzhao7,yanwang@binghamton.edu
†Rutgers University, Piscataway, NJ 08854 jl2210,yingche@scarletmail.rutgers.edu
‡Indiana University-Purdue University Indianapolis, Indianapolis, IN 46202 hl45@iupui.edu

***Abstract*—This paper subverts the traditional understanding of Photoplethysmography (PPG) and opens up a new direction of the utility of PPG in commodity wearable devices, especially in the domain of human computer interaction of fine-grained gesture recognition. We demonstrate that it is possible to leverage the widely deployed PPG sensors in wrist-worn wearable devices to enable finger-level gesture recognition, which could facilitate many emerging human-computer interactions (e.g., sign-language interpretation and virtual reality). While prior solutions in gesture recognition require dedicated devices (e.g., video cameras or IR sensors) or leverage various signals in the environments (e.g., sound, RF or ambient light), this paper introduces the first PPG-based gesture recognition system that can differentiate fine-grained hand gestures at finger level using commodity wearables. Our innovative system harnesses the unique blood flow changes in a user's wrist area to distinguish the user's finger and hand movements. The insight is that hand gestures involve a series of muscle and tendon movements that compress the arterial geometry with different degrees, resulting in significant motion artifacts to the blood flow with different intensity and time duration. By leveraging the unique characteristics of the motion artifacts to PPG, our system can accurately extract the gesture-related signals from the significant background noise (i.e., pulses), and identify different minute finger-level gestures. Extensive experiments are conducted with over 3600 gestures collected from 10 adults. Our prototype study using two commodity PPG sensors can differentiate nine finger-level gestures from American Sign Language with an average recognition accuracy over 88%, suggesting that our PPG-based finger-level gesture recognition system is promising to be one of the most critical components in sign language translation using wearables.**

I. Introduction

The popularity of wrist-worn wearable devices has a sharp increase since 2015, an estimation of 101.4 million wrist-worn wearable devices will be shipped worldwide in 2019 [1]. Such increasing popularity of wrist-worn wearables creates a unique opportunity of using various sensing modalities in wearables for pervasive hand or finger gesture recognition. Hand and finger gestures usually have a diverse combinations and thus present rich information that can facilitate many complicated human computer interaction (HCI) applications, for example wearable controls, virtual reality (VR)/augmented reality (AR), and automatic sign language translation. Taking the automatic sign language translation as an example illustrated in Figure 1, a wrist-worn wearable device (e.g., a smartwatch or a wristband) could leverage its sensors to realize and convert sign language into audio and text and back again, which will greatly help people who are deaf or have difficulty hearing to communicate with those who do not know the sign language.

Existing solutions of gesture recognition mainly rely on cameras [2]–[4] microphones [5], [6], radio frequency (RF) [7]–[9] or special body sensors (e.g., Electromyography (SEMG) [10], Electrical Impedance Tomography (EIT) sensor [11], and electrocardiogram (ECG) sensor [12]). The approaches using cameras face occlusion and privacy issues.

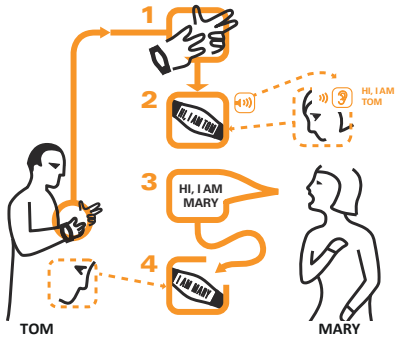


Fig. 1. Illustration of the automatic sign language translation using wearables in daily communications.

Microphones are vulnerable to ambient acoustic noises. The RF-based approaches are usually known to be device-free, but they are very sensitive to indoor multipath effects or RF interference. Using special body sensors for gesture recognition is more robust to environmental noises but requires extra cost and manpower of installation. Recently, motion sensors in wearables present their great potential in hand and finger gesture recognition on the wrist [13], [14], but motion sensors are sensitive to body motions and are thus difficult to identify fine-grained finger-level gestures, such as sign language gestures.

In this work, we propose to recognize the fine-grained finger-level gestures such as sign language using low-cost PPG sensors in wearable devices. We study the unique PPG features resulted from finger-level gestures, and carefully devise a system that can effectively detect, segment, extract, and classify finger-level gestures based on only PPG measurements. The basic idea of our system is examining the blood flow changes resulted from finger-level gestures based on the PPG measurements, which are collected by low-cost PPG sensors in wrist-worn wearable devices. The advantages of our approach are two-fold. First, our system could be easily applied to billions of existing wrist-worn wearable devices without extra cost, enabling every wrist-worn wearable device to recognize fine-grained gestures on users' fingers (e.g., sign language). Second, our system only relies on wrist-worn PPG sensors, which directly obtain gesture related features without the impact of environmental changes (e.g., ambient light, sound, RF) and moderate body movements (e.g., walking, turning body, slow arm movements), thus is more robust in practical scenarios. Our main contributions are summarized as follows:

- We demonstrate that PPG sensors in commodity wrist-worn wearable devices can be utilized to recognize fine-grained finger-level gestures. We develop a machine-learning approach by leveraging the unique gesture-related PPG patterns captured by wearables on the wrist. To our best knowledge,

978-1-5386-4128-6/18/$31.00 ©2018 IEEE

this is the first work recognizing finger-level gestures using commodity PPG sensors that are readily available in wrist-worn wearable devices.
- We explore the physical meaning and characteristics of PPG measurements collected from the PPG sensor on the wrist and develop a novel data extraction method that can precisely separate the PPG measurements caused by minute finger movements from the continuous background noise caused by human pulses.
- We show that it is possible to accurately identify complicated finger-level gestures with minute differences (e.g., sign language gestures) by exploiting various types of features extracted from the unique gesture-related PPG patterns in different signal spaces (e.g., dynamic time warping, wavelet transform, Fourier transform).
- We conduct experiments with 10 participants wearing our prototype consisting of two off-the-shelf PPG sensors and an Arduino board. We show that our system can achieve over 88% average accuracy of identifying 9 finger-level gestures from American Sign Language, suggesting that our PPG-based finger-level gesture recognition system is promising to be one of the most critical components in sign language translation using wearables.

II. Related Work

In general, current techniques for gesture recognition can be broadly categorized into four categories (i.e., vision-based, RF-based, acoustic-based and body sensor based) as follows:

Vision Based. A couple of vision-based approaches have been developed to recognize hand/body gestures with the help of cameras (e.g., Microsoft Kinect [2] and leap motion [3]) or visible light (e.g., LiSense [4]). However, these approaches are sensitive to the ambient light and may require users to have the line of sight to the cameras or need specific light sensing equipment installation.

RF Based. RF-based approaches have become increasingly important due to the prevalent wireless environments. For instance, WiDraw [7] and Wisee [8] propose to use channel state information(CSI) and Doppler shifts of wireless signals to achieve fine-grained gesture recognition, respectively. WiTrack [15] and WiTrack2.0 [16] can track multiple users by examining the multi-path effects of Frequency-Modulated Continuous-Wave (FMCW) signals. These approaches, however, either can be easily affected by environmental changes such as people walking or require dedicated and costly devices such as the Universal Software Radio Peripheral (USRP).

Acoustic Based. Acoustic-based approaches are also explored by several studies. For instance, CAT [5] and FingerIO [6] track a smartphone's motion and a finger's dynamics by using audio components (e.g., multiple external speakers, device's microphones), respectively. However, these approaches need occupy device's speaker/microphone and external audio hardware (e.g., nearby speakers), which is not always available in many real-world scenarios.

Body Sensor Based. In addition, several customized wearable devices, which can be worn on users' forearm or wrist, are designed to capture the hand gesture by capturing either Surface Electromyography (SEMG) signals [10], Electrical Impedance Tomography (EIT) [11] or electrocardiogram (ECG) [12]. However, these solutions need extra hardware supports and are not compatible with existing mobile/wearable devices. Another body of related work is using motion sensors in wrist-worn wearables to achieve hand and finger gesture recognition [13], [14]. The solution, however, is sensitive to body/arm motions and cannot identify fine-grained finger-level gestures, such as sign language gestures.

Fig. 2. Example of PPG sensors in wrist-worn wearable devices.

Different from previous work, we propose to innovatively use the photoplethysmogram (PPG) sensor, which is originally used for heart rate detection in the most of the commodity wearable devices (e.g., smartwatch and wristband), to perform fine-grained finger-level gesture recognition and detection. To the best of our knowledge, it is the first wrist-worn PPG sensor based gesture recognition system. With the proposed scheme, we envision that most wearable device manufacturers would open the interface of PPG raw readings to developers soon.

III. Preliminaries & Feasibility Study

In this section, we discuss the preliminaries, design intuitions and feasibility studies.

A. Wearable Photoplethysmography (PPG) Sensor

During the past few years, more and more commodity wrist-worn wearables (e.g., smartwaches and activity trackers) are equipped with Photoplethysmography (PPG) sensors on their back as illustrated in Figure 2. These wrist-worn PPG sensors are mainly designed to measure and record users' heart rates. Specifically, a typical PPG sensor consists of a couple of LEDs and a photodiode/photodetector (PD), which detects the light reflected from the wrist tissue. The principle of PPG is the detection of blood volume changes in the microvascular bed of tissue. When light travels through biological tissue, different substances (e.g., skin, blood and blood vessel, tendon, and bone) have the different absorptivities of light. Usually, blood absorbs more light than the surrounding tissue. Therefore, by utilizing a PD to capture the intensity changes of the light reflected from the tissue, the wearable device can derive the blood flow changes in the wrist-area tissue and calculate the pulse rate or even blood pressure [17].

It is important to note that most PPG sensors embedded in commodity wearable devices use green LEDs as light source has much greater absorptivity for oxyhemoglobin and deoxyhemoglobin compared to other light sources (e.g., red or infrared light) [18]. We thus use green-LED PPG sensors in this work to study and evaluate PPG based gesture recognition.

B. Intuition of Gesture Recognition Using PPG Sensors

The current use of PPG in wearables is limited to heart rate, pulse oximetry and blood pressure monitoring. Such applications only focus on examining regular blood flow changing patterns in the radial artery and the ulnar artery, and consider mechanical movement artifacts as noise [18]. In this work, we put forward an innovative idea of using readily available PPG in wearables for finger-level gesture recognition. We show that hand gestures, especially finger gestures (i.e., flexion, extension, abduction, and adduction), result in significant motion artifacts to PPG. The reason behind this is that the two major muscles controlling hand gestures [19], namely flexor digitorum superficialis and flexor

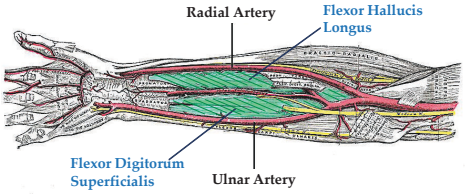


Fig. 3. Illustration of the finger movement related muscles in the anatomy of a human forearm.

hallucis longus, are right beside the radial artery and the ulnar artery as illustrated in Figure 3. Any hand or finger gestures would involve a series of complicated muscle and tendon movements that may compress the arterial geometry with different degrees. Since the blood absorbs most of the green light, the changes of the light reflected from the wrist area present different degrees of disturbances of the flood flow in terms of the shapes and duration of PPG waveforms. Current PPG sensors in off-the-shelf wearables are usually equipped with two green LED and photodiodes to ensure accurate pulse estimation by increasing the diversity (i.e., monitoring blood flow changes at different locations on the wrist). In this work, we mimic this approach and utilize two separated PPG sensors at close but different locations on the wrist to ensure our gesture recognition accuracy.

C. Feasibility Study

In order to explore the feasibility of using PPG sensors in commodity wearables for finger-level gesture recognition, we conduct five sets of experiments on a sensing platform prototyped with two off-the-shelf PPG sensors (i.e., a photodiode sensor and a green LED) connecting to an Arduino UNO (Rev3) board, which continuously collects PPG readings at 100Hz and save them to a PC. During the experiments, a user wears a wristband to fix two off-the-shelf PPG sensors on the inner side of the wrist, and respectively bends each of his fingers as illustrated in Figure 4 to emulate the simplest elements of sign language gestures. Specifically, in each set of the experiments, the user bends one of his finger 10 times with 8s between each bending. We record the process of the experiments using a video camera synchronized with the PPG measurements to determine the starting and ending time of each finger bending gesture.

We extract the PPG sensor readings within the time window between the starting and ending points identified in the video footage of each gesture and examine their changing patterns. As we expected, bending different fingers result in different unique patterns in PPG readings. Figure 4 presents an example of the unique patterns in PPG that correspond to bending different fingers, which is from one out of the two sensors. Moreover, we notice that same finger movements generate similar patterns, which demonstrates that it is possible to utilize readily available PPG sensors in wearables for fine-grained gesture recognition.

IV. Challenges & System Design

A. Challenges

In order to build a system that can recognize fine-grained finger-level gestures (e.g., sign language) using PPG sensors in wearable devices, a number of challenges need to be addressed.

Re-using the PPG Sensors in Wearables for Finger-level Gesture Recognition. The PPG sensors in commodity wearable devices are specifically designed for monitoring pulse rate

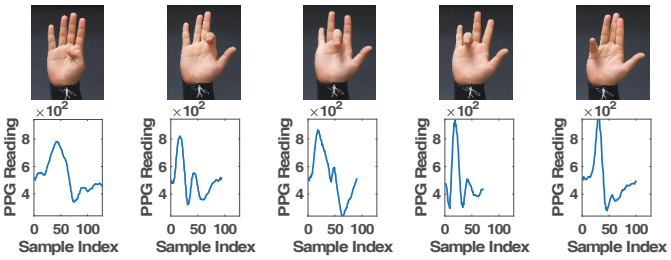


Fig. 4. Example of PPG readings associated with five finger-bending gestures in the feasibility study.

or blood pressure. The blood flow changes associated with finger-level gestures have much shorter duration and do not have repetitive patterns compared to those caused by pulses. Our system thus needs to detect and discriminate the unique PPG patterns of different finger movements by re-using the low-cost PPG sensors in commodity wearable devices.

Gesture-related PPG Readings Interfered by Pulses. In this work, PPG readings corresponding to finger-level gestures are treated as target signals that our system wants to identify and examine. Therefore, the PPG readings resulted from pulses are considered to be the noise. Such noise always exists and sometimes has intensity comparable to that of the signals caused by finger-level gestures. Our approach should be intelligent enough to separate relevant useful signals from the complicated noise caused by pulses.

Accurate Finger Gesture Recognition Using Single Sensing Modality on the Wrist. It is also challenging to achieve high accuracy in fine-grained finger-level gesture recognition by using the readily available but coarse-grained wrist-worn PPG technique. Commodity wearable devices usually only have few PPG sensors that are placed very close to each other. Such layout limits the coverage of the PPG sensors on the wrist and the diversity of sensor readings, which could significantly impact the performance of gesture recognition. We thus need to explore the critical features of PPG readings in different domains to achieve accurate finger-level gesture recognition.

B. System Overview

The basic idea of our system is examining the blood flow changes collected by readily available PPG sensors in commodity wrist-worn wearable devices to differentiate different fine-grained finger-level gestures. Toward this end, we design a machine-learning approach that mainly contains two major steps: *Training Phase* and *Classification Phase*. As illustrated in Figure 5, our system first takes as inputs the PPG measurements from wrist-worn PPG sensors. Then it conducts *Coarse-grained Gesture Detection and Reference Sensor Determination* to determine whether there is any gesture being performed based on the signal energy after mitigating the noise from pulses. Then the system automatically determines the *Reference Sensor*, which is the sensor presenting significant (i.e. containing more energy) gesture-related PPG patterns compared to those related to pulses. The system will keep monitoring the PPG sensor if there is no gesture detected. Otherwise, it will further process the raw PPG measurements depending on whether it is in the *Training Phase* or *Classification Phase*.

Training Phase. In the *Training Phase*, we collect labeled PPG measurements for each gesture and build binary classifiers for each user. Specifically, we perform *Fine-grained*

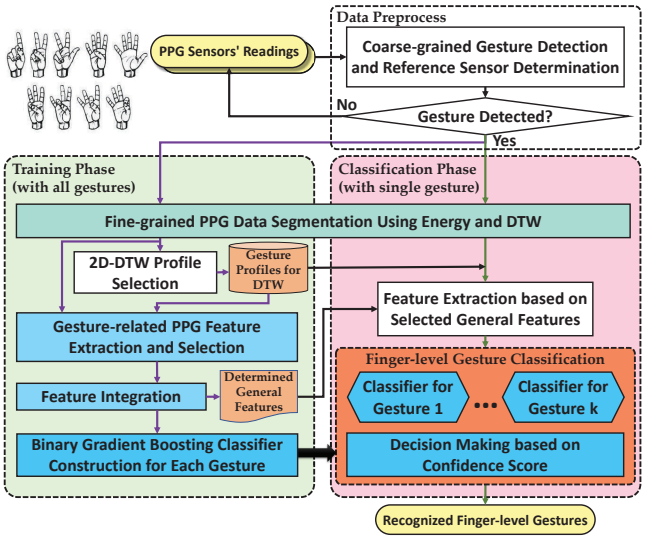


Fig. 5. Overview of the PPG-based finger-level gesture recognition system.

PPG Data Segmentation Using Energy and DTW to accurately extract the segments containing gesture-related PPG measurements between the estimated starting and ending points of gestures, which are obtained by examining the energy and dynamic time warping (DTW) distance to pulse profiles in a sliding window.

After segmentation, our system calculates the 2D-DTW distances between every two segments for every gesture in *2D-DTW Profile Selection* and selects three profile segments that are most representative for each gesture (i.e., having the minimum average 2D-DTW distance to other segments of the same gesture). The selected profile segments will be used to calculate the DTW features in the *Classification Phase*. Meanwhile, the system performs *PPG Feature Extraction and Selection* to derive a variety of features in different signal spaces (e.g., discrete wavelet transform, fast Fourier transform) and selects the critical features that can effectively capture the unique gesture-related PPG patterns for each gesture. Because the selected critical feature sets are optimized for each gesture, the system further derives a super set of the selected critical features in *Feature Integration* to ensure the system performance. Next, we perform *Binary Gradient Boosting Classifier Construction for Each Gesture* to train a binary classifier for each target gesture using Gradient Boosting.

Classification Phase. In the *Classification Phase*, our system collects testing PPG measurements in real time and determines which finger gesture has been performed based on the classification results. The system extracts the selected critical features from the PPG data segments in *Feature Extraction based on Selected General Features* and performs *Finger-level Gesture Classification* to determine which target gesture has been performed. Specifically, the system processes the extracted features using the binary gradient boosting classifiers built for target gestures in parallel. Each classifier generates a confidence score, and the system takes the target gesture having the highest confidence score as the recognized gesture.

V. Fine-grained PPG Data Segmentation

In this section, we discuss how to achieve fine-grained data segmentation based on the raw PPG data segments that have been determined to contain significant gesture-related PPG patterns through the *Data Preprocess* discussed in Section VII.

A. Energy-based Starting Point Detection

Due to the consistent existence of pulse signals in PPG measurements, it is difficult to remove the pulse signals without jeopardizing the details of the gesture-related readings, which are critical to characterizing the starting and ending points of a specific gesture. In order to accurately determine the starting point, we seek an effective detection approach to mitigate the impact of pulse signals. We find that the gesture-related PPG signals are usually stronger than those caused by pulses as illustrated in Figure 6(a), because gestures usually involve dynamics of major forearm muscles/tendons close to the sensor on the wrist. Inspired by the above observation, we design an energy-based starting point detection scheme to effectively estimate the starting of gesture-related PPG signals without removing the interference of pulses.

The basic idea of our energy-based starting point detection method is to determine the time corresponding to the local maximum of the short-time energy of PPG signals. The reason behind this is that when using a sliding window with the same length of a signal to calculate the short-time energy of the signal, the energy reaches its maximum value when the signal entirely falls into the window. Therefore, by carefully choosing the size of the sliding window (e.g., the average length of target gesture-related signals), the starting point of the gesture-related signals would be the same time when the short-time energy of the signals reaches its maximum. In particular, given the data segment containing gesture-related PPG signals $P(t)$ from the *Coarse-grained Gesture Segmentation* (Section VII), the starting point detection problem can be formulated as the following objective function:

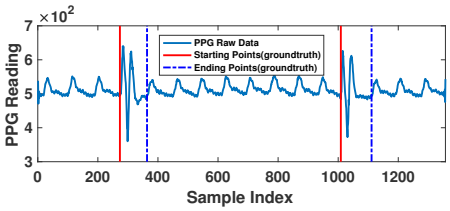
$$\arg\max_{\tau} \quad (P(\tau) - \mathbf{1}\theta)P(\tau)^T, \tag{1}$$

where $P(\tau) = [p(\tau), p(\tau+\delta)\cdots, p(\tau+W)]$, $p(\tau)$ denotes the amplitude of the PPG signal at time τ, δ represents the PPG sensor sampling interval, W is the length of the sliding window, θ is the threshold used to avoid finding the local maximum energy resulted from pulse signals, $\mathbf{1}$ is an all-one vector of the same length as $P(\tau)$, and T indicates the transpose operation. The above problem can be easily solved through simple 1-D searching within the period derived from coarse-grained gesture segmentation.

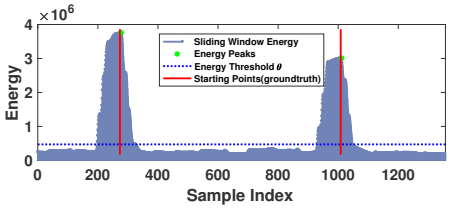
Through our preliminary study on the time length of 1080 finger gestures performed by three users as shown in Figure 7, we find that the length of gesture-related signals has the range between 0.7s and 1.4s with an average of 1.2s. Therefore, we empirically determine the length of the sliding window as 1.2s to ensure the accuracy of our energy-based starting point detection. Note that the threshold θ is user-specific and needs to be dynamically determined by the maximum short-time energy of the PPG signals when there is no gesture detected in the *Coarse-grained Gesture Detection*. Figure 6(b) illustrates the short-time energy corresponding to the PPG signals in Figure 6(a). We can clearly see that the energy peaks in Figure 6(b) are very close to the ground truth observed from the synchronized video footage, suggesting that our algorithm could promisingly capture the starting point of gestures in the PPG measurements.

B. DTW-based Ending Point Detection

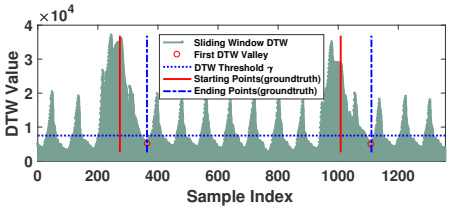
Detecting the ending point of a gesture-related signal is more challenging than detecting the starting point because the muscles are more relaxed at the end of the gesture and the



(a) PPG raw data with gesture ground truth



(b) Short-time energy of PPG



(c) Short-time DTW distance to pulse profile

Fig. 6. Example of detecting starting and ending point of a gesture-related PPG measurements using energy and DTW.

corresponding gesture-related PPG signals are usually weaker than those at the beginning of the gesture. As illustrated in Figure 6(a), the PPG measurements around the ending point do not have significant patterns that can facilitate the ending point detection. However, we find that gesture-related PPG signals are usually immediately followed by pulse signals, which are very clear and easy to identify. Hence, instead of directly locating the ending point based on PPG readings, we design a DTW-based ending point detection scheme, which aims to identify the starting time of the first pulse signal following the gesture-related signal. We employ the dynamic time warping (DTW) to measure the similarity between the user's pulse profile P_{pulse} and the PPG measurements collected after the already-detected starting point of the gesture.

Intuitively, the time when the DTW value reaches the minimum is the starting time of the pulse signals and also the ending point of the gesture-related signals. We adopt DTW because it can stretch and compress parts of PPG measurements to accommodate the small variations in the pulse signals. To summary up, this ending point detection problem is defined as follows:

$$\arg\min_{t} \quad DTW(P(t), P_{pulse}), \quad s.t., \tau < t \leq \tau + W_p, \qquad (2)$$

where $DTW(\cdot,\cdot)$ is the function to calculate the DTW distance, $P(t)$ has the same definition as $P(\tau)$ in Equation 1, W_p is the time duration for the gesture, and τ is the detected starting point. After searching the DTW distances for all $P(t)$, we find the time index of the first local minimum in the DTW distances (i.e., the starting time of the first pulse after the gesture) as the ending point of the gesture-related signals. Figure 6(c) presents the DTW between a selected pulse profile and the raw PPG measurements in Figure 6(a) with $W_p = 0.88$s. From the figure, we can observe that the time indexes of the detected first local minimum DTW values are very close to the ground truth of the ending time of the two gestures, which demonstrates the effectiveness of the DTW-based ending point detection scheme.

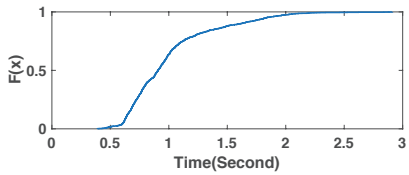


Fig. 7. Preliminary study: CDF of the duration of 1080 gestures from 3 users.

Extracting Pulse Profiles. The pulse profile P_{pulse} can be extracted from the PPG measurements that are collected when there is no gesture performed (e.g., at the beginning of the training phase). In particular, we first detect the pulse signal peaks in the PPG measurements. Given the fact that a typical PPG pulse signal always has a peak, if the pulse signal peak is located at t_p, so the PPG measurements between $[t_p - t_d, t_p + t_s]$ are identified as the user's pulse profile. In this work, we respectively choose $t_d = 0.2$s and $t_s = 0.6$s based on the duration of diastole (i.e., 0.15s$\sim$ 0.26s) and systole (i.e., 0.44s$\sim$ 0.74) phases of the vascular system reflected in a typical PPG pulse signal [20], which can effectively extract all users' pulse profile.

C. Segmentation on Inconspicuous Gesture-related Patterns

Our DTW-based ending point detection can accurately determine the ending point if the gesture-related PPG pattern has significant amplitudes compared to those of the pulse-related patterns. However, in rare cases, the gesture-related PPG patterns may not have significant amplitudes when the sensor is at the locations far away from the arteries. Note that such inconspicuous patterns are not easy to be extracted as their boundaries with pulse-related patterns are very vague, but they still contain rich information that could greatly facilitate gesture recognition. In this work, we find that when using two PPG sensors close to each other on the wrist, at least one of the sensors can generate gesture-related PPG patterns with significant amplitudes. Inspired by this observation, we adopt a reference-based approach to accurately determine the ending point for the inconspicuous gesture-related PPG patterns.

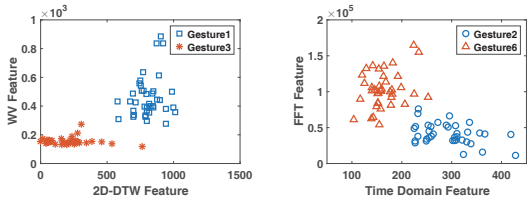
In particular, assuming our system identifies the ending point t_R on the sensor R with significant gesture-related PPG patterns (i.e., *Reference Sensor* discussed in Section VII) using our DTW-based method, the system further derives the ending point at the other sensor D as $t_D = t_R + \Delta T$, where ΔT is the time delay of the ending point on sensor R. According to our empirical study, ΔT is nonzero and stable between two sensors across different gestures. Since muscles and tendons at different locations of the forearm compress the arteries with different pressure and duration when performing a gesture, the gesture-related patterns captured by the PPG sensors at different locations will last different time periods. Because the system can always find multiple gestures that generate significant PPG patterns on both sensors, ΔT can be easily estimated in the *Training Phase* by calculating the average time difference of the ending points from the gestures where both sensors are determined to be *Reference Sensors*.

VI. Finger-level Gesture Classification

In this section, we explore the PPG features that could facilitate gesture recognition and discuss how to build the binary classifier using Gradient Boosting and perform gesture classification in the *Training Phase* and *Classification Phase*.

TABLE I
LIST OF EXTRACTED FEATURES.

Category	Features (# of features)	Description
Time Domain	**Classic Statistics** (4): mean, peak-to-peak, RMS, variance	Descriptive statistics of each segment, reflecting the statistical characteristics of the unique gesture-related patterns.
	Cross Correlation between Sensors (9)	A vector of cross correlation coefficients between the segments from two PPG sensors based on a sequence of the lag values, characterizing the relationship between two PPG sensors in a gesture.
	2D-DTW to Gesture Profiles (9)	Similarity between PPG measurements from two sensors (i.e., 2D) and the corresponding gesture profiles, directly capturing the temporal shape characteristics of the unique gesture-related patterns.
Frequency Domain	**Fast Fourier Transform (< 5Hz) (6)**: skewness, kurtosis, mean, median, var, peak-to-peak	Statistics of frequency components in the specific low frequency band, analyzing the unique PPG patterns in frequency domain.
Time-frequency Domain	**Discrete Wavelet Transform (4)**: mean, peak-to-peak, RMS, variance	Statistics of the third level decomposition of the wavelet transform using the Harr wavelet, revealing the details of gesture-related patterns at interested time and frequency scale.
	Wigner Ville Distribution [21] (13): first-order derivative, frequency and time when the signal reaches the maximum, maximum energy (E^i_{max}) / minimum energy (E^i_{min}), differential energy ($E^i_{max} - E^i_{min}$), STD^i and AV^i of the energy within the i^{th} sliding window	Fine-grained time-frequency features with high resolutions, capturing details of gesture-related patterns having short time duration.
	Autoregressive Coefficients [22] (9)	Time variant coefficients that can capture the characteristics of gesture-related patterns independent of the patterns' time scales.



(a) Gesture 1 & gesture 3 (b) Gesture 2 & gesture 6

Fig. 8. Example of different finger-level gestures and corresponding features.

A. PPG Feature Extraction and Selection

Feature Extraction. To capture the characteristics of unique gesture-related PPG patterns, we explore the efficacy of different kinds of features including typical temporal statistics (e.g., mean, variance, standard deviation (STD)), cross-correlation, autoregressive (AR), dynamic time warping (DTW), fast Fourier transform (FFT), discrete wavelet transform (DWT), and Wigner Ville distribution as listed in Table I. The features can be categorized into three types: *Time Domain*, *Frequency Domain*, and *Time-Frequency Domain*, which are designed to capture the detailed characteristics of the gesture-related PPG patterns across different frequency and time resolutions. While the *AR Coefficients*, *FFT*, *DWT*, *WVD*, and most of the *Classic Statistics* are all focusing on analyzing an individual sensor's measurements, the *Cross Correlation* and *2D-DTW* are promising for characterizing the unique gesture-based PPG patterns in terms of the relationship between a pair of sensors. Moreover, our *Time-Frequency (TF) Domain* features include three major TF types (i.e., non-parametric linear TF analysis (DWT), non-parametric quadratic TF analysis (WVD), and parametric time-varying based metric (AR)), which can well capture the dynamics of gestures in PPG measurements. In total, we extract 54 different features from each PPG sensor. Note that in order to calculate the 2D-DTW feature, our system first performs *2D-DTW Profile Selection* in the *Training Phase*, which calculates the 2D-DTW distance between every two segments for every gesture in the training data and selects three segments that have the minimum average 2D-DTW distance to other segments of the same gesture as the profile for later use in the *Classification Phase*.

Feature Selection. Our system further employs the elastic net feature selection method [23] in the *Training Phase* to automatically choose the most discriminative ones from our extracted features. In particular, the system respectively performs the elastic net feature selection on the PPG features corresponding to every target gesture. Based on the one-stand-deviation rule [24], our system keeps the most significant highly correlated features and eliminates noisy and redundant features to shrink the feature set and avoid overfitting. Next, in order to generalize the features set for classifying all target gestures, our system integrates the features selected for each target gesture and generates a general feature set $\mathbb{F}$ as follows:

$$\mathbb{F} = F(g_1) \cup \ldots \cup F(g_n), \tag{3}$$

where $F(g_n)$ is the selected feature set of the n_{th} target gesture g_n. After the feature selection and integration, we keep 46 *Determined General Features* in $\mathbb{F}$, which will be used in the *Classification Phase*. Figure 8 illustrates that our features can effectively capture different characteristics of PPG patterns for distinguishing different gestures.

B. Gradient Boosting Tree based Classification

Next, we build a binary classifier for each target gesture by using the Gradient Boosting Tree (GBT). We choose GBT mainly because 1) GBT is famous for its robustness to various types of features with different scales, which is the exact case in our project (e.g., the mean value of the PPG signal reading of the gesture period is around 500, and the autoregressive coefficients are the numbers fluctuated around 0 with value less than 1). 2) GBT classifier is robust to the collinearity of feature data. Because our features are heterogeneous across different domains, it may result in unexpected correlation or unbalance ranges that possess the collinearity. Therefore, GBT would eliminate the efforts to normalize or whiten the feature data before classification [25]. We note that among all the machine learning methods, such as Random Forest (RF) and Support Vector Machine (SVM), adopted for our classifier implementation, GBT has the best performance.

Given N training samples $\{(x_i, y_i)\}$, where x_i and y_i represent the gesture-related feature set and corresponding label with respect to one specific gesture (i.e., y_i = 1 or -1 represents whether x_i is from this gesture), GBT seeks a function $\phi(x_i) = \sum_{m=1}^{M} \omega_m h_m(x_i)$ to iteratively select weak learners $h_j(\cdot)$ and their weights ω_j to minimize a loss function as follows:

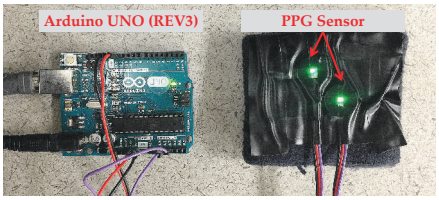


Fig. 9. Prototype wrist-worn PPG sensing platform.

$$\mathbf{L} = \sum_{i=1}^{N} L(y_i, \phi x_i). \quad (4)$$

Specifically, we adopt the GBT implementation from the library of SQBlib [26] for gesture-related feature training. Specifically, the loss function $L(\cdot)$ is chosen as the exponential loss $L = e^{y_i \phi(x_i)}$ that applies enough shrinkage (i.e., 0.1) and number of iterations (i.e., $M = 2000$), and the sub-sampling of the training dataset is a fraction of 0.5. The above parameters adopted in GBT are optimized in terms of the speed and accuracy based on our empirical study. Once the loss function is determined, we next will build a binary gradient classifier $b_k(\cdots)$ for each profiled gesture $g_k, k = 1, \cdots, K$ to complete the *Training Phase*, and each binary gradient classifier will output a score for the testing feature set. The reason of using binary classifier is that binary classifier has high accuracy with distinguishing one gesture from other gestures, whereas a multi-classifier has relative lower accuracy when performs the same classification task [27].

In *Classification Phase*, our system uses the binary classifiers for all the gestures in parallel to classify previously unseen gesture-related feature set x. Specifically, we sum the stage score [28] of each binary classifier, and choose the label k of binary classifier $b_k(x)$ with highest score as the final classification.

VII. Data Preprocess

In this section, we present two components that are critical to fine-grained data segmentation.

Coarse-grained Gesture Detection and Segmentation. To facilitate the fine-grained data extraction, our system preprocesses the raw PPG measurements to 1) determine whether there is a gesture performed or not based on the short-time energy of the PPG measurements; 2) and extract the PPG measurements that surely include the whole gesture-related PPG pattern. Specifically, the system first applies a high-pass filter to the raw PPG measurements to mitigate the interference of pulses. The reason to use the high-pass filter is that the finger-level gestures have more high-frequency components compared to the pulses, which are usually under 2Hz [29]. In this work, we build a Butterworth high-pass filter with the cut-off frequency at 2Hz. Note that we only use the filtering technique in the coarse-grained gesture detection, because the filter removes the low-frequency components of both pulse and gesture-related signals, which negatively impact the gesture recognition accuracy. Then the system decides whether there is a gesture performed or not depending on if the short-time energy of the filtered PPG measurements crosses a threshold τ or not. We set the threshold to $\tau = \mu + 3\delta$, where μ and δ are the mean and standard deviation of the short-time energy of the filtered PPG measurements collected during the time when the user is asked to be static (i.e., at the beginning of the training phase). When the system detects a gesture at t_g, we employ a fixed time window W_c to extract the raw PPG measurements within $[t_g, t_g + W_c]$ for the fine-grained segmentation. We set

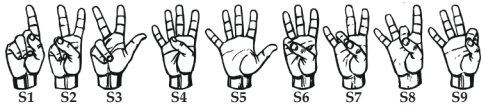


Fig. 10. American Sign Language of number one to nine.

$W_c = 4.5$s to ensure the window can cover all possible duration of gestures that we have observed in our preliminary study as shown in Figure 7.

Reference Sensor Determination. Intuitively, significant gesture-related PPG patterns could result in accurate data segmentation. However, we notice that the intensity of gesture-related PPG patterns is sensitive to the locations of sensors on the wrist, thus it may not be significant enough for segmentation. The insight is that the PPG sensors can capture more significant changes of reflected light when they are closer to the arteries that are directly compressed by muscles and tendons. Through our extensive tests, we find that two PPG sensors at a close distance on the wrist can already provide good diversity, and at least one of them can provide gesture-related PPG signals that have the stronger intensity than that of pulse-related signals. Therefore, in this work, we employ a two-sensor approach and determine which sensor could be the *Reference Sensor* having the significant gesture-related PPG patterns, which will be taken as the input for the fine-grained data segmentation. Specifically, we examine the short-time energy of the extracted PPG measurements and determine whether a sensor is a *Reference Sensor* or not depending on if its short-time energy exceeds the threshold θ, which has been defined in Section V.

VIII. Experiment and Evaluation

A. Experimental Methodology

Wearable Prototype. We notice that existing manufacturers do not provide direct access to raw PPG readings; instead, they only provide computed heart rate. Therefore, we design a wearable prototype that mimics the layout of PPG sensors in commodity wearables to demonstrate that our system can be applied to the existing wearable products without extra efforts. Our prototype consists of two commodity PPG sensors (with single green LED) and an Arduino UNO (REV3) as shown in Figure 9. The sensors are closely placed to each other and fixed on the inner side of a wristband, so that it reduces the movements of sensors and ensures our system to take sensor measurements at similar locations in different experiments. In the experiments, we adopt various sampling rates (i.e., $30Hz$ to $100Hz$) to evaluate the system. Unless mentioned otherwise, the default sampling rate is set to 100Hz.

Data Collection. We recruit 10 participants including 9 males and 1 female whose ages are between 20 to 30 to perform finger-level gestures for evaluation. We focus on the elementary gestures from American Sign Language involving movements of fingers from a single hand as shown in Figure 10. The participants are asked to respectively perform the nine finger-level gestures for 40 times while wearing our wearable prototype on the right wrist. Note that our system can be applied to other more complicated finger-level gestures on whichever wrist regardless of the posture of the hand since different gestures involve different combinations of muscle movements that can be captured by our system. In total, we collect 3600 PPG segments for the experimental evaluation. Unless mentioned otherwise, our results are derived from 20 rounds Monte Carlo cross-validation using 50% of our data set

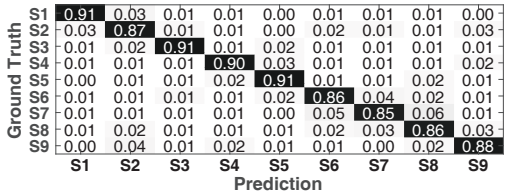


Fig. 11. Confusion matrix of recognizing nine finger-level gestures among ten participants.

for training and the rest for validation. The data are processed by our system implemented by MATLAB, which is run on an ASUS Q324U notebook.

B. Evaluation Metrics

Precision. Given N_g segments of a gesture type g, precision of recognizing the gesture type g is defined as $Precision_g = N_g^T/(N_g^T + M_g^F)$, where N_g^T is the number of gesture segments correctly recognized as the gesture g. M_g^F is the number of gesture segments corresponding to other gestures which are mistakenly recognized as the gesture type g.

Recall. Recall of the gesture type g is defined as the percentage of the segments that are correctly recognized as the gesture type g among all segments of the gesture type g, which is defined as $Recall_g = N_g^T/N_g$.

C. Finger-level Gesture Recognition Performance

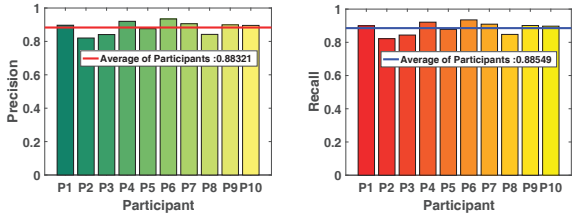
Figure 11 depicts the confusion matrix for the recognition of the nine American Sign Language gestures. Each entry C_{ij} denotes the percentage of the number of gesture segments i was predicted as gesture type j in the total number of i. The diagonal entries show the average accuracy of recognizing each gesture, respectively. Specifically, the average accuracy is 88.32% with standard deviation 2.3% among all the 9 gestures. We find that the recognition results of the gesture $S2, S6, S7, S8$ are relatively low (i.e., around 86%). This is because those gestures have more subtle differences in the tendon/muscle dynamics than other gestures. Overall, the results confirm that it is promising to use commodity wrist-worn PPG sensors to perform finger-level gesture recognition.

D. Impact of Different Users

Figure 12(a) and (b) present the average precision and recall of recognizing each finger-level gesture across different participants. We observe that all participants have high accuracy on recognizing these finger-level gestures. Specifically, the average precision and recall of all the 10 participants are 88% and 89%, respectively, and the lowest average value of the precision and recall among all the participants is still above 80%. The results show the robustness and scalability of our proposed system across different users, and demonstrate the system is promising to act as an integrated function in commodity wearables once the interface of PPG raw signals to developers is open.

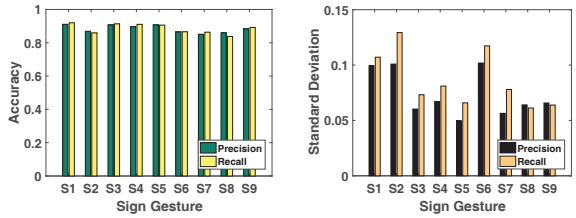
E. Impact of Different Gestures

We next study the impact of different sign gestures on the performance of the proposed system and show the average precision/recall for each sign gesture. As shown in Figure 13(a), it is encouraging to find that all those gestures can be recognized well with the lowest average precision and recall as 85% and 84%, respectively. Furthermore, Figure 13(b) shows the standard deviation of the precision and recall of recognizing each gesture. The gestures $S1$, $S2$ and $S6$ have relatively high standard deviation. This is because participants $P2$ and $P3$ inconsistently perform $S1$, $S2$ and $S6$ respectively based on our observation. Overall, our system is robust on recognizing different finger-level gestures.



(a) Average precision (b) Average recall

Fig. 12. Participant study: comparison of gesture recognition performance among ten participants.



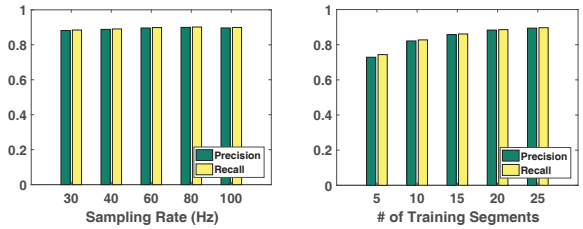
(a) Average precision & recall (b) STD of precision & recall

Fig. 13. Gesture study: Comparison of gesture recognition performance between nine gestures.

F. Various Impacts on the System

Impact of Sampling Rate. The sampling rate of sensing hardware is one of the critical impact factors on affecting the power consumption of wearables, thus we study the performance of the proposed system with different sampling rates on PPG sensors. Most of the commodity wearables have around $100Hz$ PPG sampling rate. For instance, Samsung Simband [30] configures its PPG sensor to $128Hz$ to perform time-centric tasks (e.g., Pulse Arrival Time calculations). Therefore, we set our wearable prototype to collect PPG measurements with different sampling rates (i.e., 30Hz to 100Hz with a step size of 20Hz) in our experiments. Figure 14(a) shows the average precision and recall of the gesture recognition under different sampling rates. We find that the precision/recall increases with the increased sampling rate, however, the precision/recall still maintain as high as 87% at the lowest sampling rate (i.e., $30Hz$). As the results implied, our system is compatible to commodity wearables and can provide high recognition accuracy with lower PPG sampling rate in terms of the power consumption.

Impact of Training Data Size. We change the percentage of data used for training in the Monte Carlo cross-validation to study the performance of our system under different training data size as shown in Figure 14(b). In particular, we choose the percentages $12.5\%, 25\%, 37.5\%, 50\%$, and 62.5%, which correspond to $5, 10, 15, 20$, and 25 PPG segments with respect to each gesture for training, and use the rest of our data for validation. We observe that our system can achieve an average precision of 75% for recognizing nine finger-level gestures using only 5 PPG segments of each gesture for training. As the size of the training data grows, the system performance improves significantly. More specifically, the average precision and recall can achieve 89% and 90% respectively when 25 segments of training data for each gesture are collected in the training phase. The above results indicate our system can achieve good recognition performance with a limited size of training data (e.g., 5 sets per gesture), which ensures great convenience for practical usage on commodity wearables.



(a) Impact of sampling rate (b) Impact of # of training segments

Fig. 14. Impact factor study: average precision and recall of recognizing nine gestures with different sampling rates and # of training segments.

IX. Conclusion and Future Work

As an important means for human-computer interactions, gesture recognition has attracted significant research efforts in recent years. This paper serves as the first step towards a comprehensive understanding of the PPG-based gesture recognition. We made a novel proposition to recognize the fine-grained finger-level gestures such as sign language using low-cost PPG sensors in wearables. In particular, we develop a fine-grained data segmentation method that can successfully separate the unique gesture-related patterns from the PPG measurements that are continuously interfered by pulses. Additionally, we study the unique PPG features resulted from finger-level gestures in different signal domains and devise a system that can effectively recognize finger-level gestures by only using PPG measurements. Our experiments with over 3600 gestures collected from 10 demonstrate that our system can differentiate nine elementary finger-level gestures from American Sign Language with an average recognition accuracy over 88%. We are aware of the intensity of reflected light captured by PPG sensors are sensitive to different skin colors (e.g., light colored skin reflects more light) and locations (e.g., outer side of the wrist has weaker signals); the PPG signals could be significantly impacted by strenuous exercises. We would like to present our findings and seek solutions (e.g., including motion sensors) for these potential impact factors in our future work.

Acknowledgment

We sincerely thank Dr. Gokhan Ersan with the Department of Art in Binghamton University for creating the illustration figure. Yan Wang and Hongbo Liu are corresponding authors. This work is supported in part by the National Science Foundation Grants CNS-1566455, CNS-1820624, and CNS-1514436.

References

[1] J. Klostermann Klostermann. (2015) Blinging up fashionable wearable technologies. [Online]. Available: https://cloudtweaks.com/2015/11/blinging-up-fashionable-wearable-technologies/

[2] Z. Ren, J. Yuan, J. Meng, and Z. Zhang, "Robust part-based hand gesture recognition using kinect sensor," *IEEE transactions on multimedia*, vol. 15, no. 5, pp. 1110–1120, 2013.

[3] G. Marin, F. Dominio, and P. Zanuttigh, "Hand gesture recognition with leap motion and kinect devices," in *Image Processing (ICIP), 2014 IEEE International Conference on*. IEEE, 2014, pp. 1565–1569.

[4] T. Li, C. An, Z. Tian, A. T. Campbell, and X. Zhou, "Human sensing using visible light communication," in *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*. ACM, 2015, pp. 331–344.

[5] W. Mao, J. He, and L. Qiu, "Cat: high-precision acoustic motion tracking," in *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking*. ACM, 2016, pp. 69–81.

[6] R. Nandakumar, V. Iyer, D. Tan, and S. Gollakota, "Fingerio: Using active sonar for fine-grained finger tracking," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 2016, pp. 1515–1525.

[7] L. Sun, S. Sen, D. Koutsonikolas, and K.-H. Kim, "Widraw: Enabling hands-free drawing in the air on commodity wifi devices," in *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*. ACM, 2015, pp. 77–89.

[8] Q. Pu, S. Gupta, S. Gollakota, and S. Patel, "Whole-home gesture recognition using wireless signals," in *Proceedings of the 19th annual international conference on Mobile computing & networking*. ACM, 2013, pp. 27–38.

[9] P. Asadzadeh, L. Kulik, and E. Tanin, "Gesture recognition using rfid technology," *Personal and Ubiquitous Computing*, vol. 16, no. 3, pp. 225–234, 2012.

[10] Z. Lu, X. Chen, Q. Li, X. Zhang, and P. Zhou, "A hand gesture recognition framework and wearable gesture-based interaction prototype for mobile devices," *IEEE transactions on human-machine systems*, vol. 44, no. 2, pp. 293–299, 2014.

[11] Y. Zhang and C. 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*. ACM, 2015, pp. 167–173.

[12] X. Zhang, X. Chen, Y. Li, V. Lantz, K. Wang, and J. Yang, "A framework for hand gesture recognition based on accelerometer and emg sensors," *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol. 41, no. 6, pp. 1064–1076, 2011.

[13] C. Xu, P. H. Pathak, and P. Mohapatra, "Finger-writing with smartwatch: A case for finger and hand gesture recognition using smartwatch," in *Proceedings of the 16th International Workshop on Mobile Computing Systems and Applications*. ACM, 2015, pp. 9–14.

[14] H. Wang, T. T.-T. Lai, and R. Roy Choudhury, "Mole: Motion leaks through smartwatch sensors," in *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*. ACM, 2015, pp. 155–166.

[15] F. Adib, Z. Kabelac, D. Katabi, and R. C. Miller, "3d tracking via body radio reflections." in *NSDI*, vol. 14, 2014, pp. 317–329.

[16] F. Adib, Z. Kabelac, and D. Katabi, "Multi-person localization via rf body reflections." in *NSDI*, 2015, pp. 279–292.

[17] J. Allen, "Photoplethysmography and its application in clinical physiological measurement," *Physiological measurement*, vol. 28, no. 3, p. R1, 2007.

[18] T. Tamura, Y. Maeda, M. Sekine, and M. Yoshida, "Wearable photoplethysmographic sensors-past and present," *Electronics*, vol. 3, no. 2, pp. 282–302, 2014.

[19] INNERBODY.COM. (2017) Flexor digitorum superficialis muscle. [Online]. Available: http://www.innerbody.com/image_musc05/musc50.html

[20] T. Bombardini, V. Gemignani, E. Bianchini, L. Venneri, C. Petersen, E. Pasanisi, L. Pratali, D. Alonso-Rodriguez, M. Pianelli, F. Faita *et al.*, "Diastolic time–frequency relation in the stress echo lab: filling timing and flow at different heart rates," *Cardiovascular ultrasound*, vol. 6, no. 1, p. 15, 2008.

[21] E. Ebrahimzadeh and M. Pooyan, "Early detection of sudden cardiac death by using classical linear techniques and time-frequency methods on electrocardiogram signals," *Journal of Biomedical Science and Engineering*, vol. 4, no. 11, p. 699, 2011.

[22] B. Vuksanovic and M. Alhamdi, "Ar-based method for ecg classification and patient recognition," *International Journal of Biometrics and Bioinformatics (IJBB)*, vol. 7, no. 2, p. 74, 2013.

[23] S. Noei, P. Ashtari, M. Jahed, and B. V. Vahdat, "Classification of eeg signals using the spatio-temporal feature selection via the elastic net," in *Biomedical Engineering and 2016 1st International Iranian Conference on Biomedical Engineering (ICBME), 2016 23rd Iranian Conference on*. IEEE, 2016, pp. 232–236.

[24] J. Friedman, T. Hastie, and R. Tibshirani, "Regularization paths for generalized linear models via coordinate descent," *Journal of statistical software*, vol. 33, no. 1, p. 1, 2010.

[25] T. Hastie, R. Tibshirani, and J. Friedman, "The elements of statistical learning new york," *NY: Springer*, pp. 115–163, 2001.

[26] C. Becker, R. Rigamonti, V. Lepetit, and P. Fua, "Supervised feature learning for curvilinear structure segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2013, pp. 526–533.

[27] M. Galar, A. Fernández, E. Barrenechea, H. Bustince, and F. Herrera, "An overview of ensemble methods for binary classifiers in multi-class problems: Experimental study on one-vs-one and one-vs-all schemes," *Pattern Recognition*, vol. 44, no. 8, pp. 1761–1776, 2011.

[28] R. Sznitman, C. Becker, F. Fleuret, and P. Fua, "Fast object detection with entropy-driven evaluation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 3270–3277.

[29] A. Ahmed. (1998) Frequency of a beating heart. [Online]. Available: https://hypertextbook.com/facts/1998/ArsheAhmed.shtml

[30] Simband. (2017) Why is 128 hz used as a sampling frequency for the ppg signals? [Online]. Available: http://www.simband.io/documentation/faq.html