






Gesture Recognition Using Reflected Visible and Infrared Lightwave Signals

Li Yu, *Student Member, IEEE*, Hisham Abuella , *Student Member, IEEE*,
Md Zobaer Islam , *Student Member, IEEE*, John F. O'Hara , *Senior Member, IEEE*,
Christopher Crick , *Member, IEEE*, and Sabit Ekin , *Member, IEEE*

Abstract—In this article, we demonstrate the ability to recognize hand gestures in a noncontact wireless fashion using only incoherent light signals reflected from a human subject. Fundamentally distinguished from radar, lidar, and camera-based sensing systems, this sensing modality uses only a low-cost light source (e.g., LED) and a sensor (e.g., photodetector). The lightwave-based gesture recognition system identifies different gestures from the variations in light intensity reflected from the subject's hand within a short (20–35 cm) range. As users perform different gestures, scattered light forms unique, statistically repeatable, time-domain signatures. These signatures can be learned by repeated sampling to obtain the training model against which unknown gesture signals are tested and categorized. These time-domain variations of the lightwave signals reflected from hand are denoised, standardized, and then classified by using machine learning classification tools such as K -nearest neighbors and support vector machine. Performance evaluations have been conducted with eight gestures, five subjects, different distances and lighting conditions, and visible and infrared light sources. The results demonstrate the best hand gesture recognition performance of infrared sensing at 20 cm with an average of 96% accuracy. The developed gesture recognition system is low-cost, effective, and noncontact technology for numerous human–computer interaction applications.

Index Terms—Gesture recognition, human–computer interaction (HCI), LiDAR, lightwave sensing (LWS), noncontact sensing, RADAR, signal classification, visible light sensing.

I. INTRODUCTION

WITH the growth of the computer and communication industries, Internet of Things, and the application of computers in medicine, human–computer interaction (HCI) is becoming an increasingly important technological discipline. HCI research is crucial for creating complex computerized systems that can be operated intuitively and efficiently by people without any formal training. Ideally, it leverages existing and familiar human experiences to make software and devices more

comprehensible and usable. Well-designed HCI interfaces make it convenient to control machines for education, labor, communication, and entertainment environments [1]. Such efforts have gained much attention in recent years. For example, virtual reality allows employees to better understand the nature of their work, especially when it is in an unfamiliar domain. Speech, gesture, and handwriting recognition are also highly effective, since they leverage common activities of everyday life. As such, they are important topics in applied HCI research [2], [3]. Hand gesture recognition is another natural choice for HCI. Simple movements of the hand can represent a type of sign language to machines resulting in the execution of complex actions. As a result, the recognition of hand gestures as a connection between humans and computers is now an active research area [4].

Existing hand gesture recognition techniques can be classified into two groups: wearable sensing and remote (noncontact) sensing. In wearable sensing, the user literally wears the sensor(s), which may be installed on a glove or otherwise attached to the hand. While this sensing mode is both stable and responsive, the sensor(s) must be worn whenever hand movement is to be detected. This inconvenience strips away some of the advantages of wearable sensing. In general, although wearable sensing has higher accuracy, it is simply too inconvenient for many potential users [5]–[8].

In remote sensing, hand gestures are perceived without any special hardware attached to the hand. The most frequently used sensors utilize radio frequency (RF) waves, cameras, and sound waves. The hand gesture or body motion can be identified by monitoring changes in received signals such as Doppler shifts, signal intensities, or image sequences [9]. Research on using reflected RF signals (radar) for gesture recognition is relatively mature. In [10]–[12], the authors use the received signal strength along with measured phase differences of the received signals as features to identify gestures. In [12], Google presents Soli, the first end-to-end fine gesture recognition and tracking system for HCI using a millimeter-wave radar. Soli consists of a system of multiple millimeter-wave radar transmitters and receivers. The RF-based gesture recognition method is prone to have electromagnetic interference and electromagnetic compatibility issues [13], [14].

In imaging-based gesture recognition systems, the input data are images (2-D or 3-D) and/or videos. The main challenges are separating the objects from the background and feature extraction [15], [16]. Deep-learning-based image classification

Manuscript received July 22, 2020; revised October 22, 2020; accepted November 27, 2020. Date of publication January 1, 2021; date of current version January 13, 2021. This article was recommended by Associate Editor Z. Wang. (Corresponding author: Sabit Ekin.)

Li Yu, Hisham Abuella, Md Zobaer Islam, John F. O'Hara, and Sabit Ekin are with the School of Electrical and Computer Engineering, Oklahoma State University, Stillwater, OK 74078 USA (e-mail: li.yu10@okstate.edu; hisham.abuella@okstate.edu; zobaer.islam@okstate.edu; oharaj@okstate.edu; sabit.ekin@okstate.edu).

Christopher Crick is with the Computer Science Department, Oklahoma State University, Stillwater, OK 74078 USA (e-mail: chriscrick@cs.okstate.edu).

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/THMS.2020.3043302>.

Digital Object Identifier 10.1109/THMS.2020.3043302

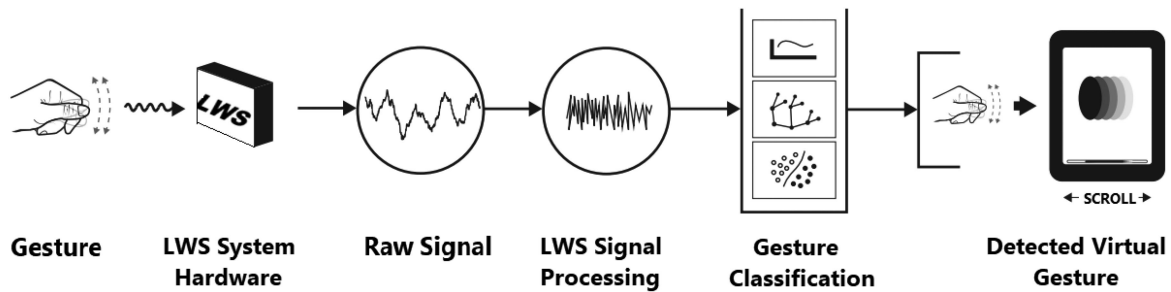


Fig. 1. Functional overview of LWS-based gesture recognition system.

has attracted much attention with the development of ubiquitous computing power over the past several years. However, deep learning methods usually work with a large amount of training samples and the data need to be labeled [17], [18]. Coupled with the large storage and processing requirements of images and videos, this increases the difficulty and complexity of this method. Meanwhile, security and privacy issues also must be taken into consideration [19], [20].

Sound-based sensing systems utilize ultrasonic waves and measure the Doppler shift of those waves reflected by the objects. The velocity of a moving hand, for example, causes characteristic Doppler shifts that serve as a signature to identify activities. The sound-based system is not susceptible to environmental noise and has good accuracy even using an uncomplicated classification algorithm [21], [22]. Adults cannot hear the ultrasonic frequency and, therefore, will not be disturbed. However, the frequencies employed may harm or perturb children and pets [23].

The general strategy of using light for sensing has attracted much attention recently due to the advancements in light-emitting diodes (LEDs), which now provide unprecedented illumination efficiency and lifetime [24]. In addition, light can be sensed using simple and inexpensive photodetectors or solar cells. Light signals, in general, require less processing capability and system complexity, compared to RF systems. They also suffer far less from cross-technology interference, owing to the increasing number of RF appliances sharing the same standardized spectrum. Visible light has already been applied to occupation estimation by analyzing the distribution of reflected and shaded signals [25]. Objects cast shadows by blocking parts of the light beams from light sources. The shape of a shadow can also be regarded as the pattern in gesture identification [9], [26]. In [27], Li *et al.* propose a shadow-based hand pose reconstructing system. There are multiple photodiodes placed in the bottom; the binary blockage maps are obtained when the hand gesture blocked the light signal right above the sensors. The hand features are extracted from the blockage maps to build the hand skeleton model to realize the hand gesture recognition and tracking. Visible light sensing can be applied to detect and identify body and arm gestures based on placing multiple receivers (photodiodes) on the floor or ceiling. However, the interference from obstacles between the body and the receiver becomes a critical issue in the shadow-based implementation. Therefore,

lightwave sensing (LWS) over shorter distances becomes attractive. Indeed, it appears in many ways that the analysis of signals from reflected light is better suited to application involving small distances [28], [29]. In [30], Gong *et al.* utilize the infrared light sensing to recognize six gestures within 0.5–7 cm. In [31], Li *et al.* propose a self-powered gesture recognition system, which combines the received visible light signals with setting position. It is a low-cost, highly accurate, and stable gesture recognition system within a 0.5–3 cm sensing distance.

Based on the existing light-based gesture recognition research, we have developed, for the first time to our knowledge, hand gesture recognition utilizing reflected lightwave (infrared and visible) signals. Fig. 1 depicts our LWS system. The main functional components include the LWS hardware, signal processing algorithms, and classification algorithms. LED light sources are used to illuminate the hand. The reflected intensity of the light varies with the movement of the hand and is captured and converted into an electrical current by a commercially available photodetector. The time-domain variations of the received (raw) signals are: 1) denoised by using discrete wavelet transform (DWT); 2) standardized by using Z score standardization; and 3) classified by using machine learning tools such as K -nearest neighbors (KNN) and support vector machine (SVM), which are trained by prior captured datasets. With this modality, we can distinguish different hand gestures with accuracy up to 96%.

Our contributions in this study can thus be summarized as follows.

- 1) A novel LWS-based hand gesture recognition system has been developed.
- 2) The gesture recognition system has been implemented in hardware and software subassemblies.
- 3) The performance comparison of machine learning classification methods KNN and SVM has been conducted.
- 4) A comparison between LWS using visible and infrared light has been performed.
- 5) A system performance summary (classification confusion matrix) has been generated for different distances and environmental lighting conditions.

The remainder of this article is organized as follows. Section II presents the principles and system design, both in terms of hardware and software algorithms utilized. Section III presents the

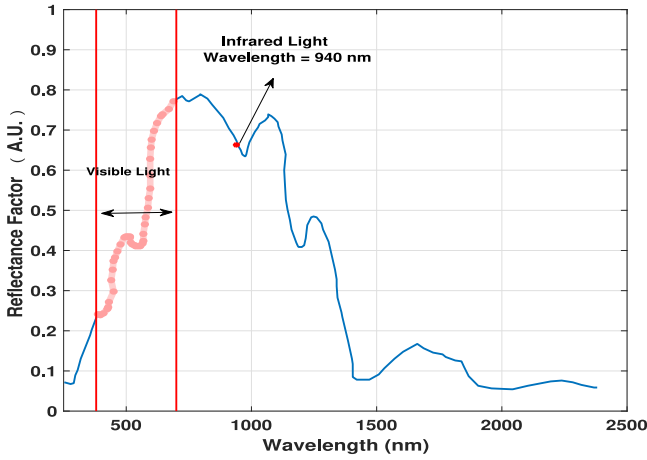


Fig. 2. Spectral reflectance variability of human skin (regenerated from NIST study [32]).

evaluation of the LWS system and some brief related discussion. Finally, Section IV concludes this article.

II. SYSTEM DESIGN AND IMPLEMENTATION

A. Influence of the Reflectance Spectrum

The spectral reflectance of human skin offers unique opportunities for noncontact sensing applications. Indeed, it can serve as an identifying signature. For example, imaging a human face with hyperspectral cameras provides very broad reflectance spectra, which can be divided into numerous narrow bands. Each of these can be used to increase the accuracy of face recognition. A National Institute of Standards and Technology (NIST) project collected measurements from 28 human subjects and calculated the spectra for their reflectance measurements over the 250–2500 nm wavelength range [32]. Fig. 2 shows the reflectance spectrum for the mean of all the samples. The spectrum exhibits the variations and scales of reflectance factors, which are critical for the aforementioned applications. While hyperspectral information is not required for hand gesture recognition, it does indicate what wavelengths LWS might be implemented at most effectively. In hand gesture recognition, higher reflectance translates to a larger reflected light intensity for a fixed power level. This feature is beneficial to achieving greater dynamic range in sensing and to reducing power consumption. From the NIST spectrum, the visible/infrared wavelength range of 600–1200 nm has the most significant reflectance. Hence, we began LWS using visible and invisible infrared light sources.

B. Sensing Hardware

Our hardware schematic is shown in Fig. 3, and the experimental setup is shown in Fig. 4. The LWS hardware consisted of one photodetector as a receiver, a digital signal processing unit to convert analog data to digital modality, light sources (visible or infrared) as transmitters, and finally an electronic unit that could process and store the received digital data. The light sources consisted of visible and infrared LEDs. The infrared light source was invisible 940 nm IR lamp board with light

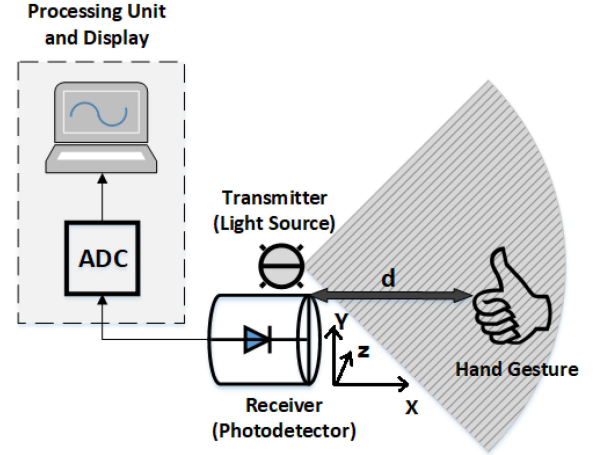


Fig. 3. Hardware overview of the LWS-based system.

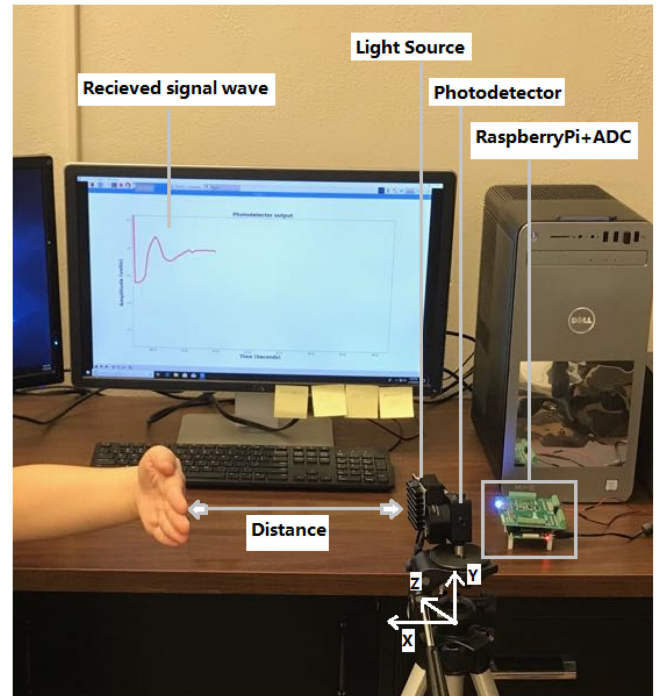


Fig. 4. Experimental setup of our LWS system.

sensor (48 black LED illuminator array) having 30 ft range and 120° wide angle beam [33]. The visible light source was 25 white 5 mm LED arrays [34]. A Raspberry Pi miniature computer with a PiPlate ADC circuit handled the data collection and digitization. A commercial photodetector served as the detector. The photodetector was Thorlabs PDA100A with spectrum responsivity 340–1100 nm, bandwidth 2.4 MHz, and area 100 mm² [35]. The gain of the PDA100A was the same in the visible and infrared measurements (20 dB) in this article.

C. Measurement Procedure Overview

In operation, volunteers perform gestures at a distance d in front of the receiver (and transmitter). As the hand makes different gestures, the photodetector records light intensity waves that are unique according to the changing distance, shape, and

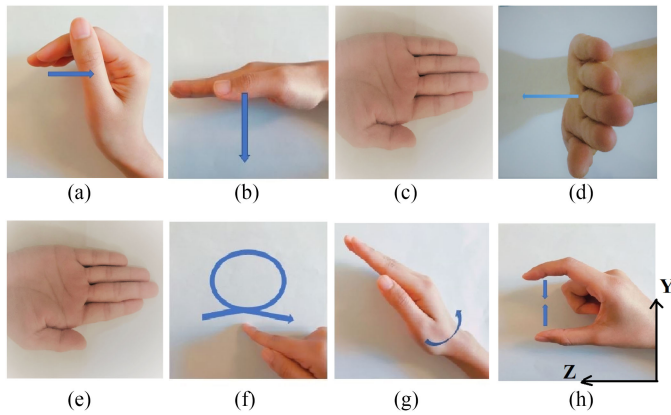


Fig. 5. Gesture set. (a) Finger slide. (b) Push. (c) Far away from the sensor. (d) Sweep from right to left. (e) Palm hold. (f) Circle. (g) Palm tilt. (h) Pinch pinky.

scattering cross section of the subject's hand. The transmitter is pointed such that the hand is centered in the brightest part of the transmitter's beam. This was visibly obvious in case of the visible light source, whereas an infrared monitor was used to optimize aiming while using the infrared source.

Volunteers were asked to perform gestures (see Fig. 5) in the designated area in front of the photodetector and light source. Each gesture could be finished in 2–3 s but was recorded for 6 s at a sampling rate of 100 Hz, resulting in individual gesture datasets of approximately 600 bytes (single precision). The digital data were then processed and classified offline using the algorithms mentioned in the next subsections. A video that shows the environments, experimental setup and the data collection process was taken (see video link at <https://youtu.be/OSTciFfvZa0>).

D. Signal Processing

In order to extract the patterns and important signal features resulting from different gestures, and to remove redundant information and noise from the received signal, multiple signal processing algorithms are needed. A flow diagram of the sequential operations performed on the raw data is presented in Fig. 6. First, discrete wavelet denoising was used for noise and interference removal. Then, a simple thresholding scheme was used to segment the long received data stream and mark the beginning and the end of a hand gesture. Then, Z scores were used to standardize the signal amplitude based on the variance of the data. This normalized the data collected in varying distance and background lighting conditions. Finally, the data were classified to a certain gesture based on the database of available gestures. Each algorithm is explained in detail as follows.

1) *System Noise and Denoising*: The raw signal is usually corrupted by noise that distorts the significant signal features, especially when the amplitude of the reflected signal is relatively low. Two noise features that were obvious in the frequency domain appear to be due to the flicker of ceiling lights, evident at 120 Hz, and the flicker of nearby computer monitors, evident at 60 Hz. The magnitude of the Fourier transform of the received signal is shown in Fig. 7. The flicker noise sources appear as

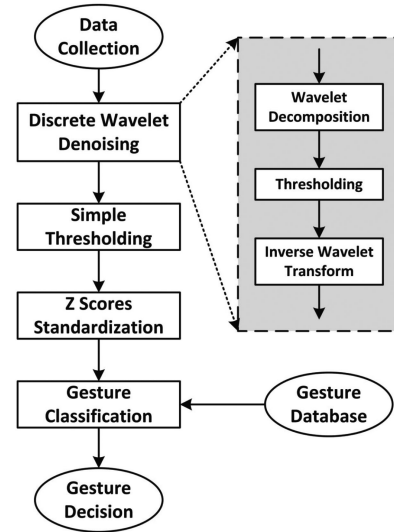


Fig. 6. Flow diagram of the steps and algorithms used to prepare measured data for gesture classification.

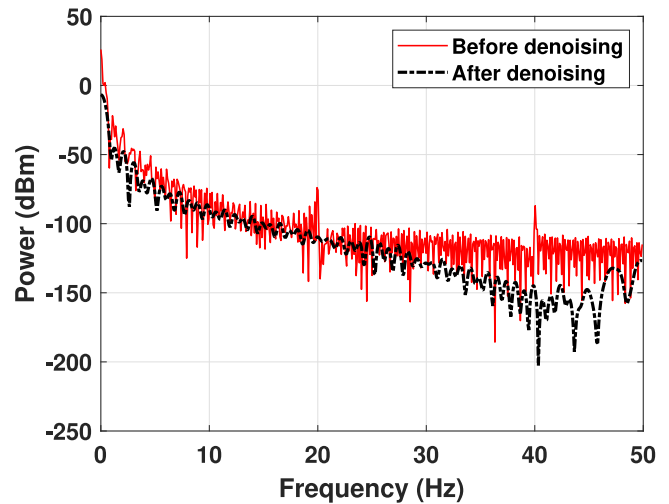


Fig. 7. Power spectrum of the received visible light signal at 20 cm before and after discrete wavelet denoising block.

peaks at 20 Hz (for the 120 Hz signal) and 40 Hz (for the 60 Hz signal) due to spectral folding caused by the 100 Hz sampling rate being below the Nyquist criterion.

In order to denoise the signal, the DWT was used [36]. The wavelet thresholding method has been proven remarkably adept in signal denoising in various published research works, including electrocardiogram (ECG) denoising [37]. The ECG signal is similar to our gesture data in that the heart activities are variable with time on similar scales and indicate different health states based on the various time-domain waves. The DWT is highly useful in analyzing nonstationary signals since it provides both a time- and frequency-domain representation of the signal [38].

2) *Discrete Wavelet Denoising*: The wavelet thresholding method removed the noise by forcing the DWT coefficients of noise to zero. Coefficients of noise were distinguished from those of the meaningful part of the signal by their relatively

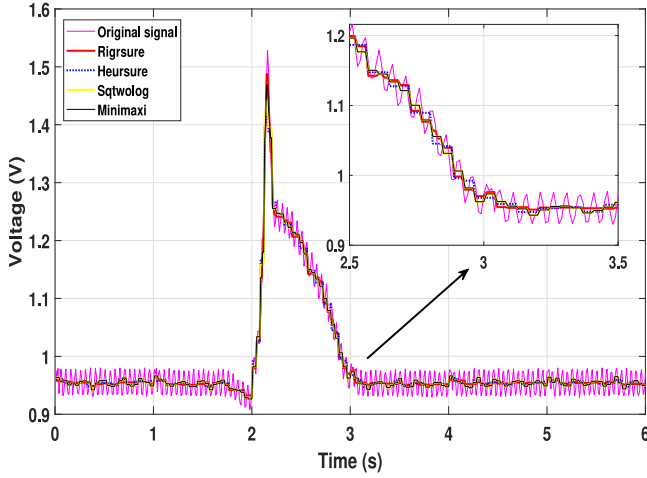


Fig. 8. Denoised wave of received visible light signal at 20 cm with four thresholding methods.

small magnitude. Wavelet thresholding means that each DWT coefficient is compared to a threshold to determine whether it is a part of the desired signal or not. Thresholding is usually applied to the DWT detail coefficients, which are related to high-frequency noise. When the value of a coefficient was found to be less than the threshold, it was forced to zero. Following this operation, the inverse wavelet transformation used the remaining nonzero coefficients to produce the denoised signal. Coefficients larger than the threshold were dealt with differently according to the thresholding scheme employed, whether it was “soft” and “hard.” More details of these two thresholding methods are available in [39]. We applied soft thresholding in our method. We have empirically observed that soft thresholding produces a more mathematically tractable signal, and also one that is easier to interpret visually. Specifically, the soft method eliminates certain signal “blips” that survive with hard thresholding.

Determining the proper threshold value is an important detail in the denoising process. A large threshold may oversmooth the recovered signal, losing important time-domain detail information. A threshold that is too small cannot effectively eliminate the noise. An appropriate threshold should be selected to balance these competing requirements. Donoho and Johnstone [40] have done a large amount of research in this area. There are two main categories: global and level-dependent thresholding. Global thresholding employs the same threshold value for all coefficients at every decomposition level. In level-dependent thresholding, it is necessary to find a suitable and possibly different threshold for each decomposition level. In [41], four thresholding techniques were analyzed and evaluated for denoising performance, based on the calculation of mean square error: *rigrsure*, *heursure*, *sqtwolog*, and *minimaxi*. We found that the *rigrsure* showed an optimum performance in our case. As shown in Fig. 8, the red wave with *rigrsure* principle is the smoothest one and provides the foundation of subsequent operations. The denoising effectiveness is demonstrated with another signal in Fig. 9.

3) *Gesture Detection*: Since gestures generally lasted only 2–3 s, much of the 6 s worth of collected data could be discarded.

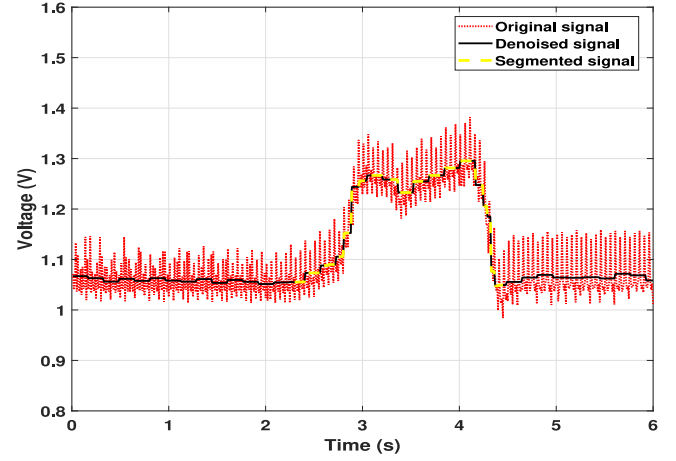


Fig. 9. Effect of discrete wavelet denoising and simple thresholding blocks on the time domain from received visible light signal at 20 cm.

We, therefore, needed to detect the beginning and end for every gesture, in between which the received signal had large changes in magnitude and shape due to hand movement. Only data between these points were forwarded for classification. Generally, the reflected intensity becomes much larger when the user is performing a gesture. Therefore, we used a thresholding scheme to detect the start and end points of each gesture *after* the denoising operation. Note that this is now a time-domain thresholding, entirely separate from the thresholding operations used in denoising. We found that suitable thresholds are slightly greater ($\sim 10\%$) than the averaged measured signal intensity *before the gesture begins*. The suitable thresholds were obtained based on the received signal value without any initial hand movements because the fluctuation of waveform could be very tiny, as shown in Fig. 13(d), while sweeping from right to left. If we chose the thresholds greater than 10%, the details like this would be removed. With this threshold, we did not lose any gesture related data, but got rid of redundant data. Thus, it was the optimal option for the recognition accuracy. The absolute scale of this threshold is variable based on the environmental situation and signal intensities observed during the measurement. For example, the average background intensity measured in a light room is larger than in a dark room. An example of this gesture detection process is shown on measured data by the amber dashed curve in Fig. 9.

After deleting the unusable data, we obtain a time series of the gesture with an unpredictable number of data points. Obviously, the filtered signals will not always have the same length at this stage. However, the follow-on classification algorithms require that all the input data have a vectorized representation of the *same* length, which allows the algorithms to efficiently execute matrix operations in batch. Signals can be time-scaled or zero-padded to cause their vector lengths to match. Dynamic time warping (DTW) is a method that aligns two time series to measure the similarities between them [42]. For example, two signals for a same gesture with different performance speed have similar shapes but different magnitudes or lengths. The DTW can make them match perfectly. We can get two new time series of same length after applying the DTW. Meanwhile, we find that

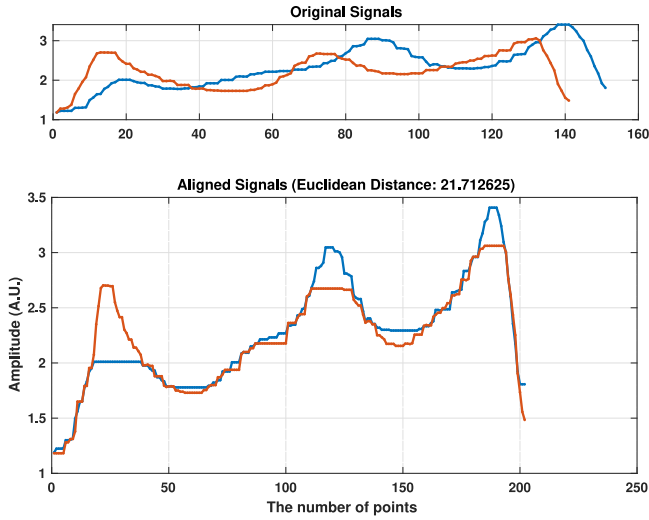


Fig. 10. DTW result for two samples of infrared light signal at 20 cm with different duration from the same gesture.

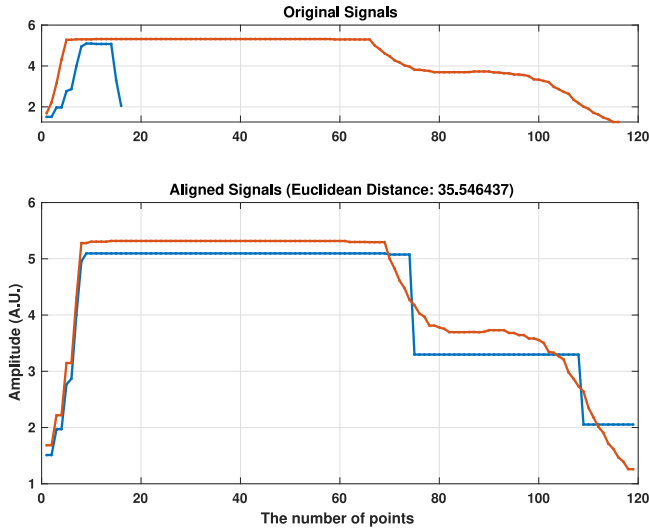


Fig. 11. DTW result for two samples of infrared light signal at 20 cm with different duration from different gesture with similar wave shape.

it makes two waves with different duration from a same gesture match very well, as shown in Fig. 10. This approach adjusts the shape of waves to an identical one for one gesture to eliminate the diversity from various users. This can be an advantage in the recognition process. However, it reduces the distinctiveness for different gestures, which have similar wave shapes because it aligns all the waves in a similar shape, as shown in Fig. 11. In this figure, those two waves are from two different gestures with similar wave shapes. As can be seen, the shape of the blue wave is changed with the red one. The dissimilarity among different gestures is reduced through this procedure. This operation will make all the waves, no matter whether they come from same gestures or different gestures, have a similar shape, resulting in error classification. Furthermore, it will damage the recognition accuracy significantly. Therefore, we found that the most

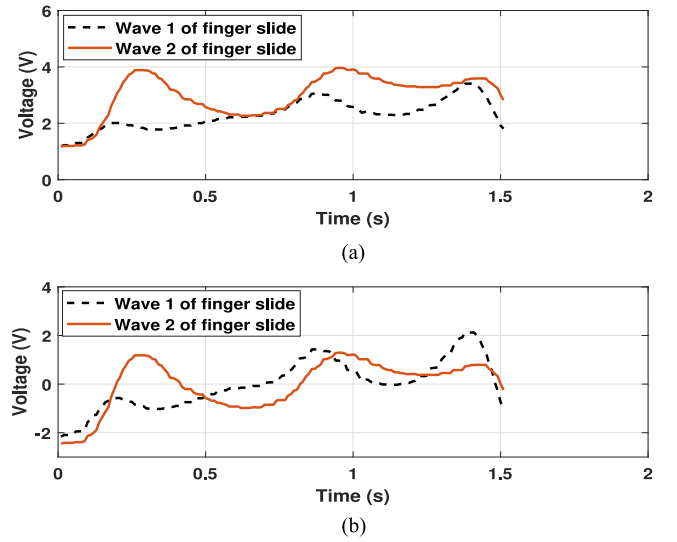


Fig. 12. Effect of Z scores normalization block on the time-domain signal from received infrared light signal at 20 cm. (a) Waveform before standardization. (b) Waveform after standardization.

reliable gesture recognition was accomplished by employing zero padding to match their lengths.

4) *Standardization*: Due to the diversity of factors involved when different users perform gestures, the magnitudes of measured light signals from the *same* gesture may be highly variable, even though their wave shapes are generally similar. For example, the distance d between a user's hand and the receiver affects the magnitude of the reflected signal. These variations deteriorate the recognition accuracy. An example is shown in Fig. 12(a). Two denoised signal plots are shown, one each of the same gesture from two different volunteers. While the waves have similar shapes, they have quite different magnitudes. To solve this problem, the denoised signals must be made to have similar magnitudes. This can be done with the Z scores method, which compares the standard deviations (SDs) of the compared signals. By visual inspection of the signals in Fig. 12(a), the one with a larger magnitude also has a larger SD. Signals can, therefore, be scaled by their SD to get similar magnitudes. Z scores is a standardization method that simply converts a dataset to a distribution with zero mean and unity SD. As shown in Fig. 12(b), the two previously described waves are shown again, after applying the Z scores standardization; they now exhibit similar magnitudes. This method made our system agnostic to different users and measurement conditions.

E. Classifier Training

After the signal processing is completed, all of the datasets were used as feature vectors in a training set to build a gesture classification method using the KNN method. We applied KNN and SVM in our study. The recognition accuracy showed that the KNN behaved better than SVM in the recognition of eight gestures. Based on the data analysis, the variances of our feature vectors were small. Due to the tiny difference and much overlap of all the waves (as shown in Fig. 13), KNN was found to be more

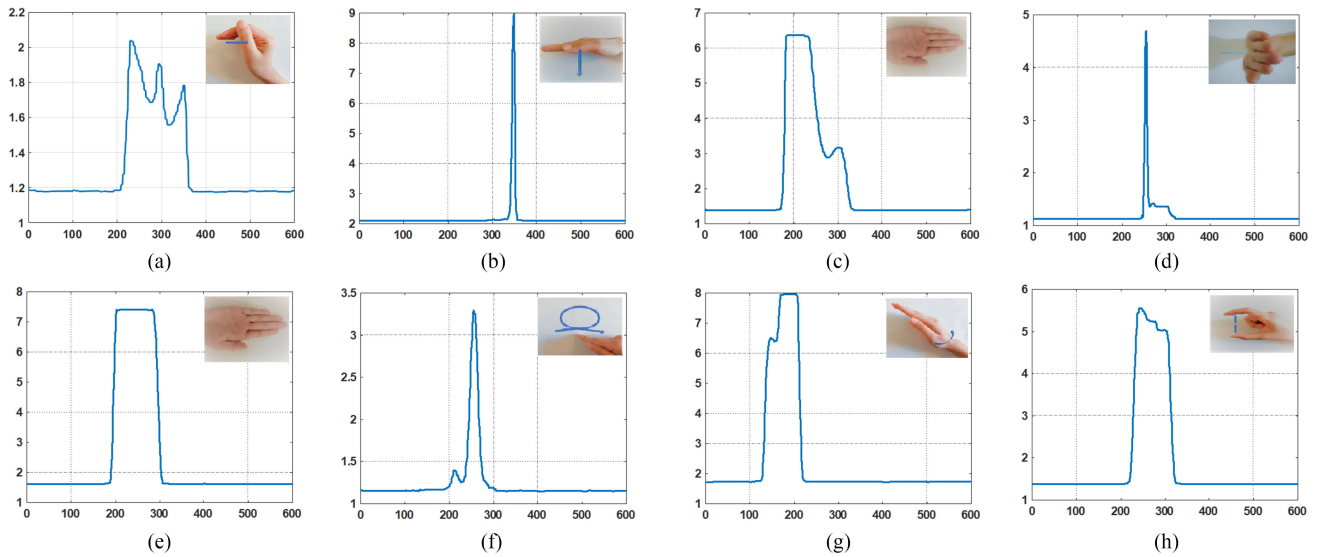


Fig. 13. Time-series waves of received infrared light signal at 20 cm with the corresponding gesture. (a) Finger slide. (b) Push. (c) Far away from the sensor. (d) Sweep from right to left. (e) Palm hold. (f) Circle. (g) Palm tilt. (h) Pinch pinky.

suitable for our proposed system. Hand gestures were selected based on the common HCI tasks [26], [43]. The waves shown in Fig. 13 are the denoised time-domain signals. After data analysis, we found that the variance of our feature vectors was small. This could also be observed visually as relatively small differences in the eight gestures in the time-domain waves. KNN is a nonparametric lazy method that utilizes the distance between each sample to separate them into several classes without any assumptions in the classification and regression applications [44]. Therefore, due to the small difference and strong overlap of all the waves, KNN was found to be more suitable for our method.

To build a training model, we collected several waves for each gesture. As mentioned earlier, statistical results showed that the volunteers completed the gesture movements shown in Fig. 13 in about 2–3 s. When the data collection was completed, the raw received data were denoised, trimmed, normalized, and padded as required and as discussed previously. Using these data, all the feature vectors were used to train a KNN classification model and get the best parameters for our gesture recognition system. New waves with accurate record of gesture type were used to test the trained model.

III. EVALUATION, RESULTS, AND DISCUSSION

In this section, we evaluate our gesture recognition system using the collected data from real volunteers with K -fold validation. First, we discuss the data collection method and software used. Then, we describe a K -fold validation procedure used to evaluate the gesture classification performance with less bias. We then present a calculated confusion matrix that classifies the accuracy for eight different gestures and visualizes the classification error. Furthermore, we present analysis of the gesture classification accuracy when using different light sources at different distances. Finally, the impact of environment lighting conditions in gesture classification accuracy is presented.

A. Data Collection

Referring to the setup shown in Fig. 4, volunteers made hand gestures in front of the light source. For each gesture, the 600 points were saved in the Raspberry Pi. The data collection used Python scripting language run in PiPlate and Raspberry Pi. The signal processing and classification algorithm using MATLAB was applied in the PC offline based on the saved data in Raspberry Pi.

We instructed volunteers how to perform each gesture and gave them several minutes to practice until they were comfortable with each gesture. We recorded data from five volunteers (two females of 25 and 27 years old and three males, two of 27 years old and one of 24 years old) performing 24 repetitions of the eight gestures (960 waves) shown in Fig. 13 at distance 20 cm from the sensor for infrared source in normal indoor room lighting conditions. There was only one window in the room, and the curtains were always closed. The color of the wall was white. The light of environment was from fluorescent lights in the ceiling and computer screen. The background color, material characteristics, distance from the gesture scene, etc., are not expected to impact the results significantly. The background will affect only the initial received signal value when there is no hand movement. The signal that was useful for gesture identification was the reflected light by hand movement between the photodetector and the background. When volunteers performed the gesture, they could sit in front of the photodetector or stand by the side of the photodetector and then perform the gesture in front of the photodetector at the specific distance. The received signals of those two situations would have small difference in the reflected intensity. To observe the impact of ambient light, the same number of waves was recorded at a distance of 20 cm infrared sensing in the dark indoor room (no lighting) conditions. Furthermore, to obtain the effect of sensing distance on accuracy, the same number of waves was collected at a distance of 35 cm

TABLE I
TENFOLD VALIDATION CONFUSION MATRIX OF INFRARED LWS AROUND
20 cm (AMBIENT LIGHT IS ON)

Estimating Gesture									
		(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)
Performing Gesture	(a)	0.93	0	0.04	0	0	0.01	0	0.01
	(b)	0	0.92	0	0.06	0	0.02	0	0
	(c)	0	0	1	0	0	0	0	0
	(d)	0.01	0.01	0	0.97	0	0.01	0	0
	(e)	0	0	0	0	0.98	0	0.02	0
	(f)	0.01	0	0	0.03	0	0.96	0	0
	(g)	0	0	0.03	0	0	0	0.97	0
	(g)	0.02	0	0	0.01	0	0.01	0	0.96

from the sensor for infrared sources in normal indoor room lighting conditions. There were three datasets with same number of waves for infrared sensing. The same three datasets under same conditions were collected using the visible light source to determine the effect of different sources on gesture recognition performance.

B. Gesture Recognition Accuracy

In this section, we evaluate the accuracy of our system at different distances, with different sources, and in different ambient lighting conditions. All the results were obtained with tenfold cross validation to use all the data and make the parameters fine-tuned. In tenfold cross validation, all the waves were divided into ten equal size subsets randomly. One of the ten subsets was taken as testing data, the remaining nine subsets were used as training data. Then, the cross validation was repeated ten times, and each of the ten subsets was used as the testing data only once. Note that the datasets were divided randomly, and the model was not supposed to be subject dependent. The training datasets were overlapped with each other. From our results, the testing accuracy results of ten models were not significantly different (the SD range of those ten matrices were within 5%). Finally, ten confusion matrices were averaged to obtain a final result. To measure the reproducibility of the system, the accuracy result of leave-one-subject-out validation was calculated using the same samples. The confusion matrix of tenfold cross validation and leave-one-subject-out validation results for infrared light sensing around 20 cm are shown in Tables I and II. The overall accuracy of cross-validation result was 96.13% (SD = 2.59%). And the overall recognition result of leave-one-subject-out validation was 92.13% (SD = 3.14%). This result represented a more realistic status compared with cross-validation result.

1) *Accuracy of Infrared and Visible LWS Compared at the Same Distance ($d = 20$ cm and $d = 35$ cm):* The dataset of infrared LWS had 960 gesture waves named as infrared dataset around 20 cm. The infrared dataset around 35 cm had the same number of waves. The forms of the datasets for visible LWS at 20 and 35 cm were identical.

TABLE II
LEAVE-ONE-SUBJECT-OUT VALIDATION CONFUSION MATRIX OF INFRARED
LWS AROUND 20 cm (AMBIENT LIGHT IS ON)

Estimating Gesture										
		(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	
Performing Gesture	(a)	0.87	0	0.01	0.01	0	0.02	0.02	0.07	
	(b)	0	0.89	0	0.04	0	0.02	0	0.05	
	(c)	0	0	0.95	0	0	0	0	0.5	
	(d)	0.01	0.03	0	0.93	0	0.03	0	0	
	(e)	0	0.01	0	0.05	0.91	0	0.02	0.01	
	(f)	0.02	0.01	0	0.01	0	0.94	0.01	0.01	
	(g)	0	0.01	0.01	0	0	0	0.96	0.02	
	(g)	0.02	0	0.02	0.03	0	0.01	0	0.92	

TABLE III
CONFUSION MATRIX OF VISIBLE LWS AROUND 20 cm (AMBIENT LIGHT IS ON)

Estimating Gesture										
		(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	
Performing Gesture	(a)	0.81	0	0	0.02	0	0.09	0.09	0	
	(b)	0	0.91	0.03	0.04	0.02	0	0	0	
	(c)	0	0.00	0.85	0	0.02	0	0.05	0.08	
	(d)	0.01	0.02	0.02	0.86	0.01	0.01	0.05	0.04	
	(e)	0	0	0.08	0.01	0.85	0	0.02	0.05	
	(f)	0.03	0	0	0	0	0.83	0.12	0.02	
	(g)	0.02	0	0.03	0	0.01	0.09	0.83	0.02	
	(h)	0.01	0	0.06	0	0.05	0.01	0.03	0.84	

As shown in Tables I and III, the accuracy of gesture recognition at $d = 20$ cm is with using the infrared light than that with the visible light.

In order to quantify this difference, we measured the light power as a function of distance from the source using a calibrated optical power meter. Using Tables IV and V, “measured radiated power” refers to power measured directly in front of the light source at different distances. “Measured power at detector reflected from flat hand” refers to the power measured after reflection from a hand-held flat and normal to the source beam. In this case, the power meter probe was placed directly in front of the usual system photodetector. The distance refers to where the hand was held during the power measurement. One can observe from the tables that the infrared light power is much larger than the visible light power. As shown in Fig. 14, the wavelength of our visible light source is from 400 to 900 nm, and the wavelength of our infrared light source is 940 nm. The spectrum was obtained using an Ocean Insight USB2000+ spectrometer [45]. During measurements, the light sources, both visible and infrared, were pointed directly at the tip of the optical fiber that conveys light to the spectrometer. The distance between the light sources and fiber tip, as well as pointing of the fiber tip, was adjusted to avoid saturating the spectrometer’s detector. Spectral data were stored by the spectrometer control software during acquisition.

TABLE IV
POWER MEASUREMENTS FOR INFRARED LIGHT SETUP

Distance (cm)	5	10	15	20	25	30	35
Measured radiated power (mW)	15.6	5.9	3.05	1.86	1.22	0.87	0.65
Measured power at detector reflected from a flat hand (mW)	1.09	0.415	0.248	0.193	0.065	0.053	0.048

TABLE V
POWER MEASUREMENTS FOR VISIBLE LIGHT SETUP

Distance (cm)	5	10	15	20	25	30	35
Power measured by power meter(mW)	3.02	1.46	0.750	0.460	0.320	0.240	0.190
Reflected power measured by power meter (mW)	0.0512	0.0412	0.0364	0.0337	0.0315	0.0296	0.0286

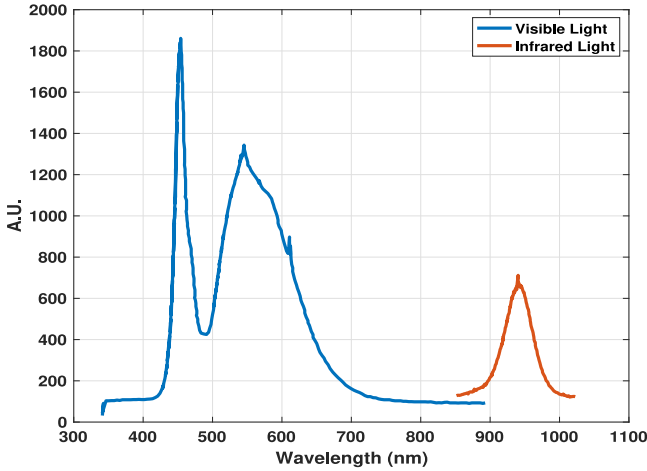


Fig. 14. Measured spectrum of our infrared and visible light sources. (The different peak levels of infrared and visible light are not necessarily indicative of relative light intensity. The visible and infrared spectra were scaled separately and then combined in the figure plot.)

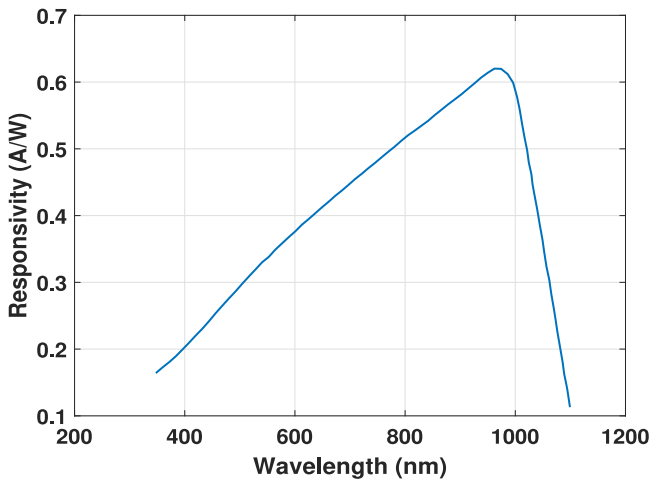


Fig. 15. PDA100A spectrum responsivity [46].

The spectrum responsivity of the photodetector is shown in Fig. 15, which is regenerated from the PDA100A manual [46]. Because of the larger light power and responsivity of the photodetector, the infrared light generates larger voltages at the detector. The larger voltage signals generally seem to raise the signal-to-noise ratio (SNR) of the system, resulting in higher

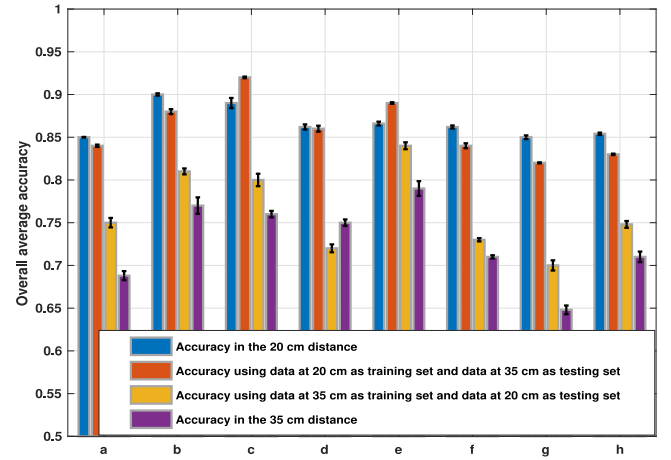


Fig. 16. Accuracy results of visible LWS at 20 and 35 cm (ambient light is ON; mean = 84.75%, SD = 2.96% for the accuracy at 20 cm; mean = 86.00%, SD = 3.42% for the accuracy using the dataset at 20 cm as training set and dataset at 35 cm as testing set; mean = 76.13%, SD = 4.94% for the accuracy using the dataset at 35 cm as training set and dataset at 20 cm as testing set; and mean = 71.13%, SD = 4.91% for the accuracy at 20 cm) with confidence interval = 0.95.

recognition precision. The results in Figs. 16 and 17 at $d = 35$ cm reveal the same conclusion.

2) *Accuracy at Different Distances:* Figs. 16 and 17 reveal the effect of distance on gesture recognition accuracy. Accuracy decreases with increasing distance due to the lower reflected light intensity. The reduced accuracy at increased distance with both infrared and visible light sources can also be attributed to lower light intensities reducing the overall system SNR. The results of 20 cm sensing data for training with 30 cm data for testing and *vice versa* are shown in Figs. 16 and 17, to show the sensitivity on different training and testing distances. The effect of distance for collecting testing data had negligible impact on the accuracy of the system. The accuracies of the same training dataset with different testing sets were found to be similar.

However, the standardization discussed in Section II-C was meant to reduce the impact of absolute light intensity on performance. The above result was, therefore, somewhat unexpected and compelled us to compare the recognition accuracy with and without standardization. As shown in Fig. 18, the standardization process does not make the system agnostic to the impact of different sources and distances, but it does systematically increase the recognition accuracy.

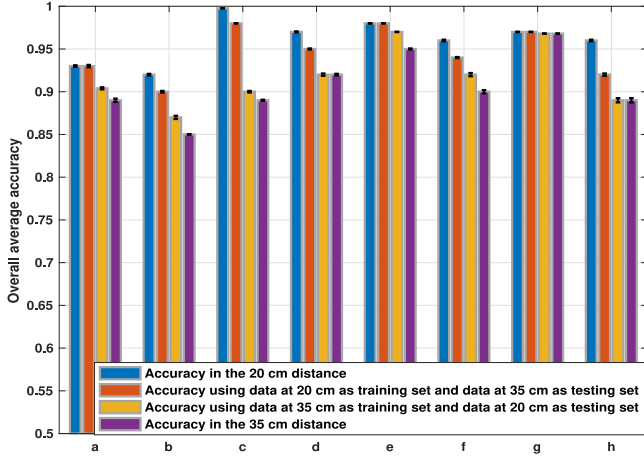


Fig. 17. Accuracy results of infrared LWS at 20 and 35 cm (ambient light is ON; mean = 96.13%, SD = 2.59% for the accuracy at 20 cm; mean = 94.62%, SD = 2.92% for the accuracy using the dataset at 20 cm as training set and dataset at 35 cm as testing set; mean = 92.00%, SD = 3.55% for the accuracy using the dataset at 35 cm as training set and dataset at 20 cm as testing set; and mean = 84.75%, SD = 2.96% for the accuracy at 20 cm) with confidence interval = 0.95.

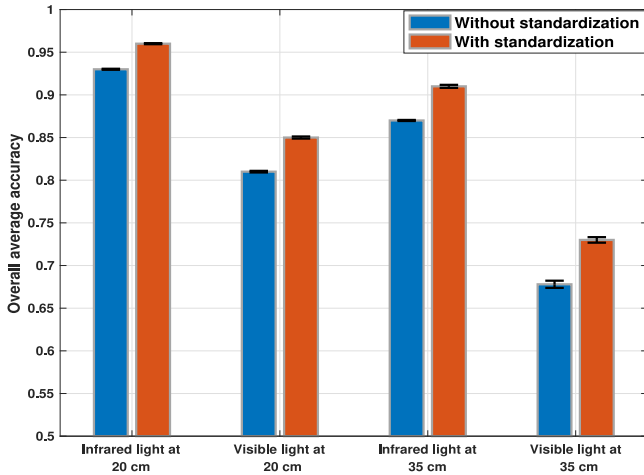


Fig. 18. Effect of standardization step on the average accuracy for both infrared and visible light source at 20 and 35 cm (ambient light is ON) with confidence interval = 0.95.

3) *Accuracy With Different Ambient Lighting:* To determine the effect of ambient light on recognition accuracy, we compared the classifications result with ambient lights ON and OFF using both infrared and visible LWS at $d = 20$ cm. The ambient light is ON means the fluorescent lights in the ceiling and computer screen light are included in the environment light. The ambient light is OFF means that there is no light in the room and the light sources are our infrared light or visible light source.

It is seen from the results in Tables I and VI that, when using the infrared source, ambient lighting conditions have little, if any, significant impact. This is consistent with the fact that the power of the reflected infrared light is much greater than the ambient contribution. However, Tables III and VII show that for visible LWS, recognition is slightly better when ambient lighting is OFF.

TABLE VI
CONFUSION MATRIX OF INFRARED WAVE SENSING AROUND 20 cm IN THE DARK (AMBIENT LIGHT IS OFF)

		Estimating Gesture							
Performing Gesture		(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)
	(a)	0.94	0	0	0	0	0	0	0.06
	(b)	0	0.97	0	0	0	0.03	0	0
	(c)	0	0	0.99	0	0	0	0.01	0
	(d)	0	0.02	0	0.98	0	0	0	0
	(e)	0	0	0	0	0.98	0	0.02	0
	(f)	0.03	0.02	0	0	0	0.92	0	0.03
	(g)	0	0	0.03	0	0	0	0.97	0
	(h)	0	0	0	0	0	0	0	1.00

TABLE VII
CONFUSION MATRIX OF VISIBLE LWS AROUND 20 cm IN THE DARK (AMBIENT LIGHT IS OFF)

		Estimating Gesture							
Performing Gesture		(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)
	(a)	0.85	0	0	0.03	0	0.1	0.02	0
	(b)	0	0.90	0	0.1	0	0	0	0
	(c)	0	0	0.95	0	0	0	0.05	0
	(d)	0	0.03	0.06	0.88	0	0.011	0	0
	(e)	0	0.03	0	0	0.89	0	0.08	0
	(f)	0	0.04	0	0	0	0.88	0.08	0.02
	(g)	0	0.01	0	0.01	0.03	0	0.85	0.1
	(h)	0	0	0	0.07	0.01	0	0.06	0.86

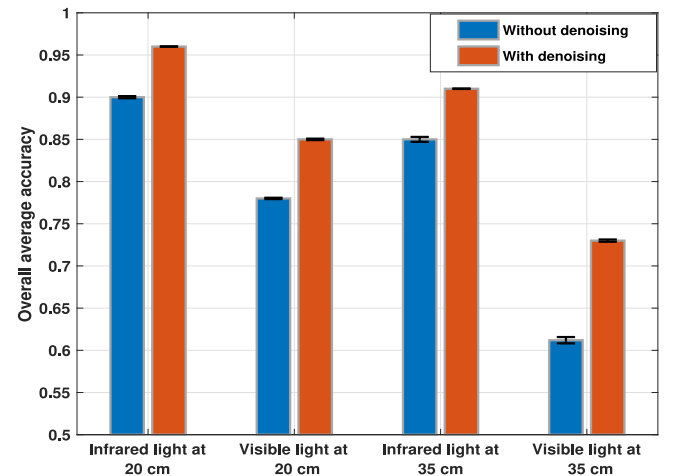


Fig. 19. Effect of denoising step on the average accuracy for both infrared and visible light source at 20 and 35 cm (ambient light is ON) with confidence interval = 0.95.

This is attributed to the fact that the visible light source along with signal reflections is weaker. The weaker reflected signals are impacted more by the environment noise compared with the stronger one.

As a related topic, we also show the performance of the denoising operation since this was meant to reduce the impact

of noise level on recognition accuracy. As shown in Fig. 19, the accuracy systematically improves by including denoising, although the benefit is not uniform to all conditions. It appears to have the greatest positive impact as signal levels approach the environmental noise levels.

IV. CONCLUSION

In this article, we presented a light-based hand gesture recognition system that utilized incoherent light reflection signals to accomplish hand gesture recognition in a short range between 20 and 35 cm. The main innovation was the exploitation of ubiquitous light, which is safe, low-cost, and easily generated and analyzed. We have shown how we employ a series of signal processing steps and use machine learning such that this sensing modality can achieve high recognition accuracy for eight gestures (in the case of infrared light sensing) in common ambient lighting conditions.

In order to verify the limitations and capabilities of our system, more measurements and testing should be done on more subjects and in different lighting conditions. The age, gender, and skin complexion of subjects have to be taken into consideration. Moreover, it was observed that recognition accuracy improved with transmitter power, which suggested that improved performance can be achieved with greater SNR and/or dynamic range in the system. More photodetectors are needed to be applied to achieve the tracking of hand or body movement at large distances. Future studies will be conducted to better quantify and model the operation of the system, to further verify this method's practicality and limitations, and to improve system performance.

ACKNOWLEDGMENT

The authors would like to thank anonymous reviewers for their constructive comments, which helped in improving this manuscript.

REFERENCES

- [1] J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami, "Internet of Things (IoT): A vision, architectural elements, and future directions," *Future Gener. Comput. Syst.*, vol. 29, no. 7, pp. 1645–1660, 2013 (including Special Sections: Cyber-Enabled Distributed Computing Ubiquitous Cloud Network Services & Cloud Computing Scientific Applications—Big Data, Scalable Analytics, Beyond).
- [2] R. Aigner *et al.*, "Understanding mid-air hand gestures: A study of human preferences in usage of gesture types for HCI," Microsoft Research, Redmond, WA, USA, Rep. MSR-TR-2012-111, Nov. 2012.
- [3] T. Seehapoch and S. Wongthanavasu, "Speech emotion recognition using support vector machines," in *Proc. 5th Int. Conf. Knowl. Smart Technol.*, 2013, pp. 86–91.
- [4] S. S. Rautaray and A. Agrawal, "Vision based hand gesture recognition for human computer interaction: A survey," *Artif. Intell. Rev.*, vol. 43, no. 1, pp. 1–54, Jan. 2015.
- [5] C. Zhu and W. Sheng, "Wearable sensor-based hand gesture and daily activity recognition for robot-assisted living," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 41, no. 3, pp. 569–573, May 2011.
- [6] A. Nelson *et al.*, "Wearable multi-sensor gesture recognition for paralysis patients," in *Proc. IEEE SENSORS Conf.*, Nov. 2013, pp. 1–4.
- [7] Z. Lv, "Wearable smartphone: Wearable hybrid framework for hand and foot gesture interaction on smartphone," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops*, Dec. 2013, pp. 436–443.
- [8] M. Caputo, K. Denker, B. Dums, and G. Umlauf, "3D hand gesture recognition based on sensor fusion of commodity hardware," in *Proc. Mensch Comput. Conf.*, Jan. 2012, pp. 293–302.
- [9] R. H. Venkatnarayan and M. Shahzad, "Gesture recognition using ambient light," in *Proc. ACM Interact. Mobile Wearable Ubiquitous Technol.*, vol. 2, no. 1, Mar. 2018, Art. no. 40.
- [10] F. Adib, Z. Kabelac, D. Katabi, and R. C. Miller, "3D tracking via body radio reflections," in *Proc. 11th USENIX Symp. Netw. Syst. Des. Implementation*, 2014, pp. 317–329.
- [11] S. Sen, J. Lee, K.-H. Kim, and P. Congdon, "Avoiding multipath to revive inbuilding WiFi localization," in *Proc. 11th Annu. Int. Conf. Mobile Syst., Appl., Serv.*, 2013, pp. 249–262.
- [12] J. Lien *et al.*, "Soli: Ubiquitous gesture sensing with millimeter wave radar," *ACM Trans. Graph.*, vol. 35, no. 4, Jul. 2016, Art. no. 142.
- [13] Z. Chi *et al.*, "EAR: Exploiting uncontrollable ambient RF signals in heterogeneous networks for gesture recognition," in *Proc. 16th Conf. Embedded Netw. Sens. Syst.*, Nov. 2018, pp. 237–249.
- [14] Z. Tian, X. Yang, and M. Zhou, "WiCatch: A Wi-Fi based hand gesture recognition system," *IEEE Access*, vol. 6, 16911–16923, 2018.
- [15] L. Chen, H. Wei, and J. Ferryman, "A survey of human motion analysis using depth imagery," *Pattern Recognit. Lett.*, vol. 34, no. 15, pp. 1995–2006, 2013 (Smart Approaches for Human Action Recognition).
- [16] M. A. R. Ahad, J. K. Tan, H. Kim, and S. Ishikawa, "Motion history image: Its variants and applications," *Mach. Vis. Appl.*, vol. 23, no. 2, pp. 255–281, Mar. 2012.
- [17] M. Asadi-Aghbolaghi *et al.*, "A survey on deep learning based approaches for action and gesture recognition in image sequences," in *Proc. 12th IEEE Int. Conf. Automat. Face Gesture Recognit.*, Washington, DC, USA, 2017, pp. 476–483.
- [18] O. Oyedotun and A. Khashman, "Deep learning in vision-based static hand gesture recognition," *Neural Comput. Appl.*, vol. 28, pp. 3941–3951, Apr. 2016.
- [19] S. Oprisescu, C. Rasche, and B. Su, "Automatic static hand gesture recognition using ToF cameras," in *Proc. 20th Eur. Signal Process. Conf.*, Aug. 2012, pp. 2748–2751.
- [20] T. Plotz, C. Chen, N. Y. Hammerla, and G. D. Abowd, "Automatic synchronization of wearable sensors and video-cameras for ground truth annotation—A practical approach," in *Proc. 16th Int. Symp. Wearable Comput.*, Jun. 2012, pp. 100–103.
- [21] Y. Qifan, T. Hao, Z. Xuebing, L. Yin, and Z. Sanfeng, "Dolphin: Ultrasonic-based gesture recognition on smartphone platform," in *Proc. IEEE 17th Int. Conf. Comput. Sci. Eng.*, 2014, pp. 1461–1468.
- [22] A. Mujibiya, X. Cao, D. S. Tan, D. Morris, S. N. Patel, and J. Rekimoto, "The sound of touch: On-body touch and gesture sensing based on transdermal ultrasound propagation," in *Proc. ACM Int. Conf. Interact. Tabletops Surfaces*, 2013, pp. 189–198.
- [23] H. Watanabe, T. Terada, and M. Tsukamoto, "Ultrasound-based movement sensing, gesture-, and context-recognition," in *Proc. Int. Symp. Wearable Comput.*, 2013, pp. 57–64.
- [24] C. Zhang, J. Tabor, J. Zhang, and X. Zhang, "Extending mobile interaction through near-field visible light sensing," in *Proc. 21st Annu. Int. Conf. Mobile Comput. Netw.*, 2015, pp. 345–357.
- [25] Y. Yang, J. Hao, J. Luo, and S. J. Pan, "CeilingSee: Device-free occupancy inference through lighting infrastructure based led sensing," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun.*, Mar. 2017, pp. 247–256.
- [26] M. Kaholokula, "Reusing ambient light to recognize hand gestures," Dept. Comput. Sci., Undergraduate thesis, Dartmouth College, Hanover, NH, USA, 2016.
- [27] T. Li, X. Xiong, Y. Xie, G. Hito, X.-D. Yang, and X. Zhou, "Reconstructing hand poses using visible light," in *Proc. ACM Interactive, Mobile, Wearable Ubiquitous Technol.*, Sep. 2017, Art. no. 71.
- [28] T. Hao, R. Zhou, and G. Xing, "COBRA: Color barcode streaming for smartphone systems," in *Proc. Int. Conf. Mobile Syst., Appl. Services*, 2012, pp. 85–98.
- [29] H. Cheng, A. M. Chen, A. Razdan, and E. Buller, "Contactless gesture recognition system using proximity sensors," in *Proc. IEEE Int. Conf. Consum. Electron.*, Jan. 2011, pp. 149–150.
- [30] J. Gong, Y. Zhang, X. Zhou, and X.-D. Yang, "Pyro: Thumb—Tip gesture recognition using pyroelectric infrared sensing," in *Proc. 30th Annu. ACM Symp. User Interface Softw. Technol.*, 2017, pp. 553–563.
- [31] Y. Li, T. Li, R. A. Patel, X.-D. Yang, and X. Zhou, "Self-powered gesture recognition with ambient light," in *Proc. 31st Annu. ACM Symp. User Interface Softw. Technol.*, 2018, pp. 595–608.
- [32] C. C. Cooksey and D. W. Allen, "Reflectance measurements of human skin from the ultraviolet to the shortwave infrared (250 nm to 2500 nm)," *Proc. SPIE*, vol. 8734, 2013, Art. no. 87340N.

- [33] *940 nm IR Lamp Board with Light Sensor (48 Black LED Illuminator Array)*, Amazon.com, Inc., Seattle, WA, USA. Accessed on: Jul. 11, 2020. [Online]. Available: <https://www.amazon.com/gp/product/B0785W2RQQ>
- [34] *Super Bright White 5 mm LED (25 Pack)*, Adafruit Industries, New York, NY, USA. Accessed on: Jul. 11, 2020. [Online]. Available: <https://www.adafruit.com/product/754>
- [35] *PDA100A*, Thorlabs, Inc., Newton, NJ, USA. Accessed on: Jul. 11, 2020. [Online]. Available: <https://www.thorlabs.com/thorproduct.cfm?partnumber=PDA100A>
- [36] S. Lahmiri and M. Boukadoum, "Physiological signal denoising with variational mode decomposition and weighted reconstruction after DWT thresholding," in *Proc. IEEE Int. Symp. Circuits Syst.*, 2015, pp. 806–809.
- [37] P. Singh, G. Pradhan, and S. Shah Nawazuddin, "Denoising of ECG signal by non-local estimation of approximation coefficients in DWT," *Biocybern. Biomed. Eng.*, vol. 37, no. 3, pp. 599–610, 2017.
- [38] J. P. Amezcua-Sanchez and H. Adeli, "A new music-empirical wavelet transform methodology for time-frequency analysis of noisy nonlinear and non-stationary signals," *Digit. Signal Process.*, vol. 45, pp. 55–68, 2015.
- [39] A. Kashaf, N. Javaid, Z. A. Khan, and I. A. Khan, "TSEP: Threshold-sensitive stable election protocol for WSNs," in *Proc. 10th Int. Conf. Front. Inf. Technol.*, 2012, pp. 164–168.
- [40] J. Lord, M. Rast, C. Mckinlay, and P. Mininni, "Wavelet decomposition of forced turbulence: Applicability of the iterative Donoho-Johnstone threshold," *Phys. Fluids*, vol. 24, Feb. 2012, Art. no. 025102.
- [41] D. Valencia, D. Orejuela, J. Salazar, and J. Valencia, "Comparison analysis between rigsure, sqtwolog, heursure and minimaxi techniques using hard and soft thresholding methods," in *Proc. 21st Symp. Signal Process., Images Artif. Vis.*, Aug. 2016, pp. 1–5.
- [42] F. Zhou and F. de la Torre, "Generalized time warping for multi-modal alignment of human motion," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2012, pp. 1282–1289.
- [43] W. T. Freeman and M. Roth, "Orientation histograms for hand gesture recognition," Mitsubishi Electr. Res. Labs., Cambridge, MA, USA, Tech. Rep. TR94-03, 1994.
- [44] Q. Chen, D. Li, and C. Tang, "KNN matting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 9, pp. 2175–2188, Sep. 2013.
- [45] *USB Ser. UV-NIR Spectrometers*, Ocean Insight, Duiven, The Netherlands. Accessed on: Jul. 11, 2020. [Online]. Available: <https://www.oceaninsight.com/products/spectrometers/usb-series/usb-uv-n%ir/>
- [46] *PDA100A(-EC) Si Switchable Gain Detector User Guide*, Thorlabs, Inc., Newton, NJ, USA. Accessed on: Jul. 11, 2020. [Online]. Available: <https://www.thorlabs.com/drawings/4db5f4acdbba82df-2EE78459-B895-BFEB-3%C3F1B3F7149ECE/PDA100A-Manual.pdf>