applied sciences

MDPI

Article

Finger-Gesture Recognition for Visible Light Communication Systems Using Machine Learning

Julian Webber [1,*], Abolfazl Mehbodniya [2], Rui Teng [3], Ahmed Arafa [2] and Ahmed Alwakeel [2]

[1] Graduate School of Engineering Science, Osaka University, Toyonaka 560-8531, Japan
[2] Department of Electronics and Communication Engineering, Kuwait College of Science and Technology, 7th Ring Road, Doha 20185145, Kuwait; a.niya@kcst.edu.kw (A.M.); a.arafa@kcst.edu.kw (A.A.); a.alwakeel@kcst.edu.kw (A.A.)
[3] Organization for Research Initiatives and Development, Doshisha University, Kyoto 610-0394, Japan; dr.r.teng@ieee.org
* Correspondence: webber@ee.es.osaka-u.ac.jp

Abstract: Gesture recognition (GR) has many applications for human-computer interaction (HCI) in the healthcare, home, and business arenas. However, the common techniques to realize gesture recognition using video processing are computationally intensive and expensive. In this work, we propose to task existing visible light communications (VLC) systems with gesture recognition. Different finger movements are identified by training on the light transitions between fingers using the long short-term memory (LSTM) neural network. This paper describes the design and implementation of the gesture recognition technique for a practical VLC system operating over a distance of 48 cm. The platform uses a single low-cost light-emitting diode (LED) and photo-diode sensor at the receiver side. The system recognizes gestures from interruptions in the direct light transmission, and is therefore suitable for high-speed communication. Gesture recognition accuracies were conducted for five gestures, and results demonstrate that the proposed system is able to accurately identify the gestures in up to 88% of cases.

Keywords: visible light communications (VLC); gesture recognition (GR); human-computer interaction (HCI); human activity recognition (HAR); machine learning (ML); neural network; long short-term memory (LSTM); photo-diode (PD)

Citation: Webber, J.; Mehbodniya, A.; Teng, R.; Arafa, A.; Alwakeel, A. Finger-Gesture Recognition for Visible Light Communication Systems Using Machine Learning. *Appl. Sci.* **2021**, *11*, 11582. https://doi.org/10.3390/app112411582

Academic Editor: Grzegorz Dudek

Received: 2 November 2021
Accepted: 2 December 2021
Published: 7 December 2021

1. Introduction

Gesture recognition (GR) systems can greatly assist the elderly or infirm as well as persons unable to control equipment through speech. Meanwhile the growth of Internet of Things (IoT) propelled the need for improved human-computer interaction (HCI) to enable control of devices inthe areas of work, play, health, communication, and education. For real-world application, a GR system should require modest computing resources and be implementable with low-cost. While proprietary GR systems are emerging, they tend to be expensive, single-task oriented, and application-specific.

Gesture recognition systems can be classified into contact or contactless types. The most common contact type is the accelerometer or inertial sensor, while the contactless types include (i) ultrasound-, (ii) mm-wave radar-, (iii) video camera-, and (iv) photo-diode (PD)-based units. An accelerometer consists of multiple motion sensors in order to detect movement in the three cardinal directions. A wrist-strapped accelerometer is a low-cost GR solution in which the sensor directly tracks the hand gesture. Although research benefited from analysis of accelerometer data collected by smartphones, such systems are still impractical. Short-range frequency-modulated continuous wave (FMCW) radar was recently used in movement and gesture detection, as well as monitoring vital-signs (breathing and heart rates), based on measuring the Doppler shifts. Similarly, GR can also be achieved by measuring the Doppler from ultrasonic waves reflected by limb movement. However,

these approaches are prone to clutter between the Tx and Rx reducing the resolution, and ultrasounds can also cause stress to pets and infants who can hear the low-frequency waves. Unlike visible light, some radio-frequency systems are precluded from use in hospitals, aircraft, or mines due to electromagnetic compatibility issues. One issue with video-based GR is that the foreground limb image needs to be distinguished from nearby clutter and background objects. As deep-learning algorithms became more powerful, the ability to delineate these images increased. However, deep learning often necessitates a high degree of storage and processing power, such as from a desktop computer. Although recent development kits including the Nvidia Jetson and Microsoft Kinect [1] greatly facilitated AI-based image processing, the hardware and computational costs can still be prohibitive. Another disadvantage of using video cameras for GR is due to privacy concerns and laws. Meanwhile, interest in photo-diode (PD)-based GR will increase with the emerging visible light communication (VLC) systems, which can be made with light emitting diodes (LEDs) at a fraction of the cost.

Gesture recognition is a related field of human activity recognition (HAR), and recent developments are briefly described here. Two common methods for HAR are those based on video scene extraction and that of indirect sensing using wireless signals. Indirect sensing involves the analysis of the received signal strength signature from Wi-Fi signals that are blocked or reflected by human movement. Researchers demonstrated accuracies above 90% using support vector machines (SVM) machine learning (ML) [2–4]. However, it is currently very difficult to classify the subtle finger gestures using the wireless signals in a practical setting with a wall-mounted access-point, and it becomes harder with several people in the room. Physical activity recognition system using wrist-band based sensors were designed for wheelchair-bound patients with spinal cord injuries [5]. Smart healthcare systems are increasingly employing neural networks to categorize and automate functions [6]. Estimation of the number of people in a room was made through an analysis of reflection and blocking of visible light [7]. The long short-term memory (LSTM) algorithm is a type of recurrent neural network that can efficiently learn time-series sequences that are increasingly used in ML-based HAR systems, such as [8], for wearable activity recognition [9] and sign language translation [10].

Meanwhile, visible light communication systems exploit the existing lighting infrastructure to provide high-speed and secure data communication [11–13] and are expected to become commonplace in homes and office following the release of the IEEE 802.11bb [14] Standardization currently scheduled for 2022. VLC leverages the huge bandwidth available in the nonionizing visible electromagnetic spectrum [15]. Light is a suitable communication medium in medical environments [16–18] where there are strict electromagnetic compatibility conformance standards. VLC-based health monitoring [19] and notification systems were developed for the blind [20]. VLC systems can be built with very low-cost [21] using standard light emitting diodes (LEDs) and photodiodes (PDs), such as those commonly used in DVD players. High-speed VLC systems direct the transmission of focused light between the transmitter (Tx) LED and receiver (Rx) PD. On the other hand, currently, most GR systems for visible light operate on reflected light captured by multiple PDs. A non-ML-based motion detection system using VL comprising multiple PDs was proposed in [22]. The work focused on communications performance, and there were no gesture classification accuracy results.

Gesture patterns are statistically repeatable and can be learned by repeated sampling using ML. A summary of recent hand GR research using ML is tabulated in Table 1. Infrared (IR) systems are less affected by ambient light and can generally achieve higher classification accuracies. However, most IR systems do not achieve the high visible light (VL) data-rates and at the same price-point. Using the decision-trees algorithm, authors reported a 98% classification accuracy using IR proximity sensors [23]. Feature extraction using SVM achieved 95% accuracy on data collected from an accelerometer [24]. Back-propagation was used to track hand trajectories using an inertial sensor with 89% accuracy [25]. A smart electronic-skin comprising an array of detectors and LSTM processing was proposed [26].

By tracking the shape of shadows cast through hand-blocking using a 32-sensor array, researchers achieved 96% accuracy [27]. Although the system achieved good performance, the large 6 × 6 ft array is rather impractical, and additionally, not aimed at communications. Classification performance is generally improved by deploying multiple PDs on the ceiling and floor. As the cost and computational complexity generally scale with the number of detection chains, these should be kept to a minimum. K-nearest neighbors (KNN) is a low-complexity, nonparametric algorithm that can distinguish gesture classes based on the Euclidean distances between samples. An accuracy of 48% was achieved using KNN with a single PD and increased to 83% by employing two PDs [28]. Classification of reflected IR waves was achieved using a hybrid KNN and SVM [29]. The researchers used the THORLABS PDA100 PD module (currently cost about $430) to capture a wide range of wavelengths with design ease. When the separation was 20 cm, the average denoised accuracy was 96% for IR and 85% for VL. The performance decreased with increasing Tx-Rx distance due to the lower received light intensity. When the separation increased to 35 cm, the performance decreased to 91% for IR and 73% for VL. The use of reflected light generally requires additional postprocessing to remove artifacts generated by multipath reflections from surrounding clutter and is sensitive to thresholding. This makes building a practical low-cost system challenging, and these systems offer lower data rates. The FingerLight system employs 8 spatially separated PDs and a recurrent neural network to learn the gestures from measured light intensities. When a hand is carefully positioned in front of the sensor array, a 99% classification accuracy was reported possible [30]. Short-range millimeter wave radar has provided a 98% classification accuracy for hand gesture recognition using LSTM [31]. Image processing-based techniques generally exhibit the highest performance but require very high computing resources, and hence, are less suitable for low-cost, portable-use cases. GR using captured video is often implemented using CNNs, and researchers reported a 97% classification accuracy using this technique [32]. Recurrent neural networks are able to extract auto-correlations in sequential data and were particularly successful with speech- and hand-writing recognition. The LSTM recurrent network contains gates that allow it to operate on relatively long time sequences. Multimodal gesture recognition using 3D convolution and convolutional LSTM was described in [33]. Tracking of hand-joint movements using the unscented Kalman filter [34] with LSTM and dynamic probabilities [35] was reported.

Our proposed GR solution is part of a wider VLC-capable system, and therefore the GR capability comes at almost no additional cost. The system learns to associate finger movements with the pattern of light directly impinging on the PD in the absence of obstruction by fingers. This method is unaffected by nearby clutter or by the light-reflecting properties of a subjects skin, which can depend on their age and gender. This enables us to employ a low-cost PD (about $8 in small volumes) and the approach is compatible with high-speed VLC systems targeted for communications. We employ the LSTM algorithm for the gesture classification which requires considerably lower complexity than than that of the CNN algorithm for video processing. Despite the modest complexity, the gesture recognition performs well (88%) and can be used within a communications-based VLC system.

Our contributions can be summarized as follows:

1. Provided a review of contemporary gesture recognition systems.
2. Developed a practical GR methodology that can be integrated with a VLC system. The technique uses common off-the-shelf components with full part numbers provided.
3. Developed a system using a single PD that receives direct light from the transmitting LED.
4. Demonstrated an efficient LSTM-based GR system with limited computational complexity.
5. Achieved high classification accuracy under natural settings: gestures made at natural speed and visible light.
6. Confirmed the system performance at different sampling rates and complexities.

In this paper, we focus on describing the operation of the GR module, which uses the same components as the VLC system for compatibility. The scope of this paper is limited to the gesture recognition system, and a full description of the communication

operation will be described separately. The context switching between the sensing and communications systems is an implementation issue and outside the scope of this paper. However, we considered a method based on halting the communications as soon as the hand is inserted between the Tx and Rx. This would be detected by a significant dip in the received signal power. Communications would then resume a short period after the signal blocking finishes.

The organization of this paper is as follows. Section 2 describes the VLC channel model, and Section 3 discusses the activity recognition concept and our proposed solutions for a VLC system. Section 4 details the system implementation and experiment setup, while Section 5 describes the performance results. Discussions on areas for future work and a conclusion is drawn in Sections 6 and 7, respectively.

Table 1. Gesture recognition systems using machine learning.

Reference	Processing	Sensor	Accuracy (%)	VLC
[23]	Decision-trees	IR proximity	98	No
[24]	SVM	Accelerometer	95	No
[25]	BP-NN	Inertial sensor	89	No
[26]	LSTM	5×7 sensor array	85	No
[27]	PCA	32 PDs	96	No
[28]	KNN	3×3 PD array	48 (single PD)	No
[29]	KNN/SVM	IR/VL (PDA100A)	73(VL@35 cm)	No
[30]	RNN	8 PDs	99 (10 cm)	No
[31]	LSTM	FMCW radar	98	No
[32]	CNN	RGB Camera	97	No
[33]	LSTM	RGB/depth Camera	98	No
[34]	LSTM	RGB Camera (dataset)	85	No
[35]	DP-LSTM	RGB Camera	83	No
This work	LSTM	Single PD (low-cost)	88	Yes

2. VLC Channel Model

Assume a channel model between a Tx (LED) and an Rx (PD), and consider only the line-of-sight (LOS) path. The channel impulse response of this LOS component is deterministic and given by Equation (1) [36].

$$h^{LOS}(t) = I(\phi)\frac{g(\psi)A_{PD}}{d^2}\delta(t - d/c), \tag{1}$$

where A_{PD} is the photo-diode surface area, ϕ is the angle from the Tx to Rx, ψ is is the angle of incidence with respect to the axis normal to the receiver surface, d is distance between Tx and Rx, c is the speed of light, $g(\psi)$ is the Rx optical gain function, and $I(\phi)$ is the luminous intensity.

At the Rx, the received optical power can be expressed as (2).

$$P_R = H(0)P_E, \tag{2}$$

where $H(0)$ is the channel DC gain, and P_E is the emitted optical intensity.

It is common to model the emitted signal by a generalized Lambertian pattern, and the DC channel gain can be expressed as [37].

$$H(0) = \frac{(m+1)A_{PD}}{2\pi d^2}\cos^m(\phi)T_s(\psi)g(\psi)\cos(\psi), \tag{3}$$

for $0 \leq \psi \leq \Psi_c$ where Lambertian order is denoted by (4)

$$m = \frac{-\ln(2)}{\ln(\cos(\Phi_{1/2}))}, \tag{4}$$

where $\Phi_{1/2}$ is the semiangle at half-illuminance of the Tx. $T_s(\psi)$ is the optical filter gain, Ψ_c is the Rx field of view (FOV) semi-angle.

The illuminance at a point on the receiving plane is described by $I(\psi)\cos(\psi)/d^2$ [38]. The total received power with lens is plotted in Figure 1. This figure shows that the power is greatest directly below the LED and falls off greatest at the corners. The Rx power is sufficiently high in all directions within 2 m of the center, and therefore photo-detectors receive sufficient illuminance in a typical small room or office setting.

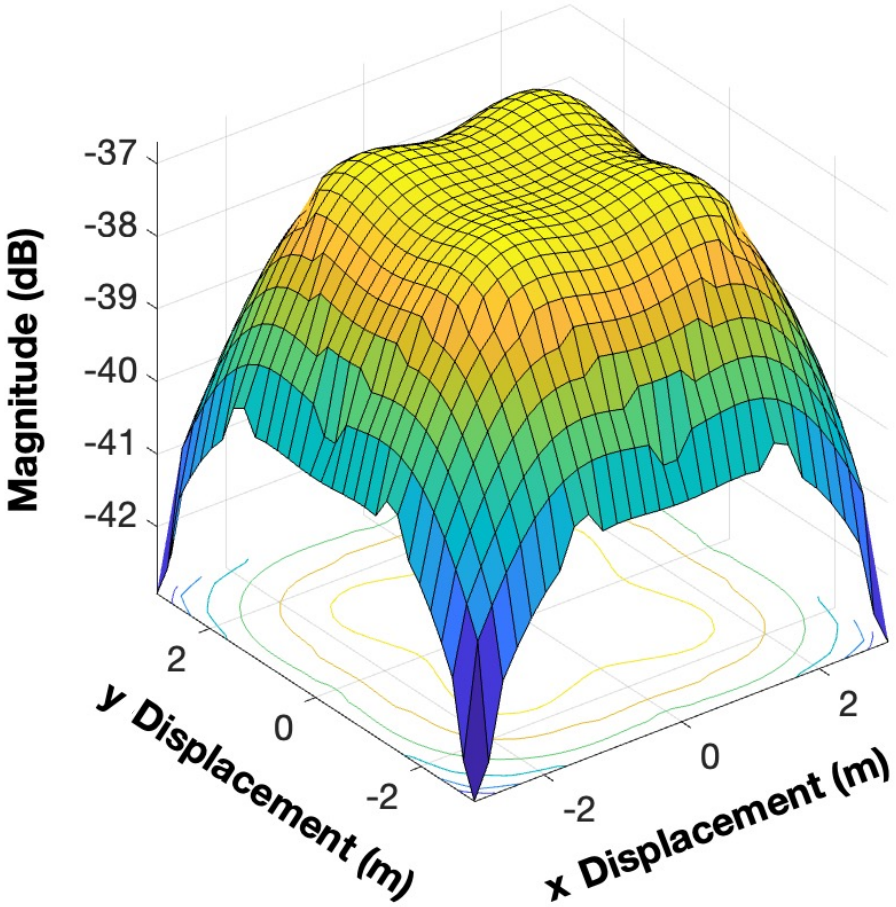

Figure 1. Lambertian simulation for total Rx power for $\phi = 30°, \psi = 30°$ FOV.

3. Gesture Recognition System with LSTM Network

A typical HAR system comprises data acquisition, segmentation, feature extraction, and classification stages. The categorization is based on an analysis of the pattern activity sensed on each PD. Through training, the system learns to associate the sequences with each activity.

The concept of the hand movement recognition system is shown in Figure 2. The identification activity takes place between the LED and PD. Unobstructed light from the LED is incident on the photo-diode sensor and, as an object moves in between the two, light can become blocked. The task is to associate the sequence of incident light with the particular gesture. Typically, a hand may move at about 1 m/s or 1000 mm/s. The distance between fingers is up to about 10 mm, and therefore periods of activity and inactivity will typically last for about 10 ms. To reliably capture these movements the symbol sensing slot-time should be at least 0.1 ms. The slot time depends on the underlying use of the VLC system and is a trade-off between VLC data rate requirements, prediction accuracy, and computational complexity. The signaling rate is typically easily satisfied by modern VLC systems that operate above 1 Mbit/s.

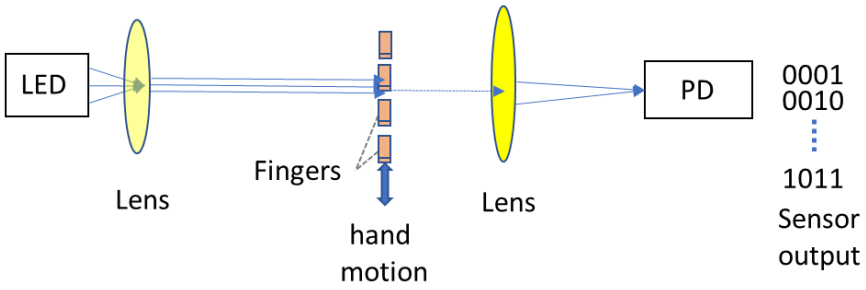

Figure 2. Concept of finger movement recognition system based on received patterns of light on a photo-diode sensor.

LSTM is a type of recurrent network that learns patterns embedded in time-series data [39] and has complexity proportional to the number of time-steps. The network is applied here to predict the finger gesture on a per time-step basis. The network comprises a sequence layer for handling the series input data, an LSTM layer for computing the learning, a fully-connected layer, a softmax layer, and finally, a classification layer. The size of the fully connected layer determines how well the network can learn the dependencies but care is required to avoid problems associated with over-fitting. The LSTM block diagram is shown in Figure 3 in which $\mathbf{x}_t$ represents the input data. The hidden-state and cell-states at time t are termed $\mathbf{h}_t$ and $\mathbf{c}_t$, respectively. The current state and the next sequence data samples will determine the output and updated cell state. The cell state is given by Equation (5)

$$\mathbf{c}_t = f_t \odot \mathbf{c}_{t\text{-}1} + i_t \odot g_t \tag{5}$$

The hidden-state is given by Equation (6)

$$\mathbf{h}_t = \mathrm{o}_t \odot \sigma_c(\mathbf{c}_t), \tag{6}$$

where σ_c represents the state activation function. Control gates allow data to be forgotten or remembered at each iteration.

The forget, cell-candidate, input, and output-states at time step t are given by Equations (7)–(10) respectively:

$$f_t = \sigma_c(W_f \mathbf{x}_t + R_f \mathbf{h}_{t-1} + b_f), \tag{7}$$
$$g_t = \sigma_c(W_g \mathbf{x}_t + R_g \mathbf{h}_{t-1} + b_g), \tag{8}$$
$$i_t = \sigma_c(W_i \mathbf{x}_t + R_i \mathbf{h}_{t-1} + b_i), \tag{9}$$
$$o_t = \sigma_c(W_o \mathbf{x}_t + R_o \mathbf{h}_{t-1} + b_o), \tag{10}$$

where W_f, W_g ,W_i, W_o represent the forget, cell-candidate, input, and output weights. R_f,R_g,R_i, and R_o are the forget, cell-candidate, input, and output recurrent weights. b_f, b_g,b_i, and b_o are the forget, cell-candidate, input, and output biases.

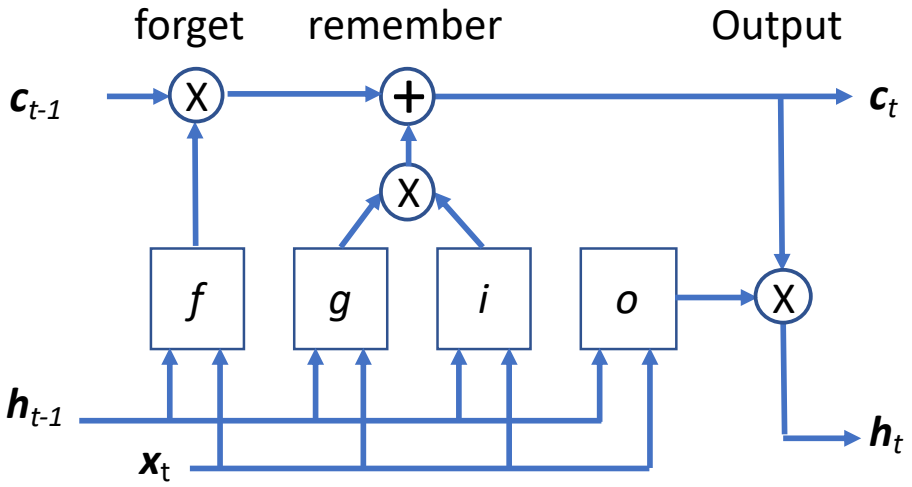

Figure 3. LSTM algorithm unit structure.

4. System Implementation

4.1. Design Approach

Two design approaches were considered for the gesture sensing operation. Approach (i): the mark-space waveform generated by all fingers is encoded. As a finger cuts the light beam, it results in a space period where the received light intensity on the PD sensor is low. In the period where light can pass between the fingers, the received intensity is high. Approach (ii): the PD output is summed over the duration of the whole gesture. The total light incident on the PD from the first to last finger cutting the light beam is recorded. The first approach was selected after an initial study showed it was more reliable, and in particular, is less dependent on the hand-speed. A minimum and maximum threshold is set, and the on-off signal is passed to the LSTM algorithm.

4.2. VLC Transceiver

The VLC system is implemented with real-time transmission and reception of symbols using an arbitrary waveform generator (AWG) and digital storage oscilloscope (DSO) as depicted in Figure 4. VLC data modulation/demodulation and activity recognition tasks are computed off-line using a personal computer with Matlab software.

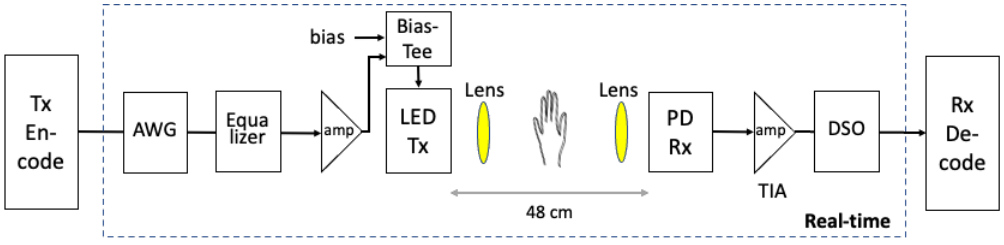

Figure 4. VLC for HAR system block diagram.

The Tx signal was generated with amplitude 1.80 V at 100 kHz in real-time using an arbitrary waveform generator Tektronix AWG 710B (max. 2.1 GHz bandwidth, 4.2 Gsa/s). An amplitude equalizer was inserted to counteract the low-pass frequency response of the LED. The amplitude equalizer provides about 7 dB loss at DC and the normalized gain rises to unity in the high-pass region at around 100 MHz. A Mini-Circuits ZHL-500 (0.1 MHz to 500 MHz) 17-dB gain-block is employed as a preamplifier to increase the small signal-level. The amplified data signal is added to a LED bias voltage of 4.2 v using a Mini-Circuits Bias-T ZFBT-4R2GW-FT+ (0.1–6000 MHz bandwidth) and the output connected to a Luxeon Rebel LED via a standard SMA connector. The LED was selected as it is

capable of supporting a data-rate in the order of 100 Mbit/s for communications. However, many other LEDs can also be used for the purpose of gesture recognition. The bias-T and amplifier had minimum operating frequency around 50 kHz. The bias voltage is adjusted to maximize the amplifier output power but backed off to avoid distortion. The amplifier, bias-T and LED were mounted onto a movable micro-stage platform to facilitate the alignment of the Tx.

To increase the communication distance, a focusing-lens of diameter 40 mm was placed at both the Tx and Rx sides with a separation of 30 cm as shown in Figure 5. The focusing lens produces a narrow beam with optimum focus at the region where the hand is placed which is at the half-distance between Tx LED and RX PD. The required distance can be easily adjusted by increasing or decreasing the lens focal-range. In the current set-up if the hand is positioned away from the center-point then the signal-to-noise ratio (SNR) is reduced and therefore estimation accuracy will be degraded. A focusing lens is also an integral and necessary component in all VLC systems and so is not an additional cost. A consumer VLC system may likely employ directional Tx/Rx or an adaptive lens mechanism.

A standard PD (Hamamatsu S10784 commonly used in DVD laser-discs) was employed at the receiver. The PD output was amplified by an OPA 2356 based low-noise amplifier (LNA) circuit that has a BW of about 200 MHz and was used here as a trans-impedance amplifier (TIA). The Rx waveform is detected by a PD and amplified by the LNA. LEDs generate incoherent light, which can be detected using simple direct or envelope detection circuitry. The Rx DSO was set at 2 Msa/s with a total 3.2 Mpoints stored after peak sampling.

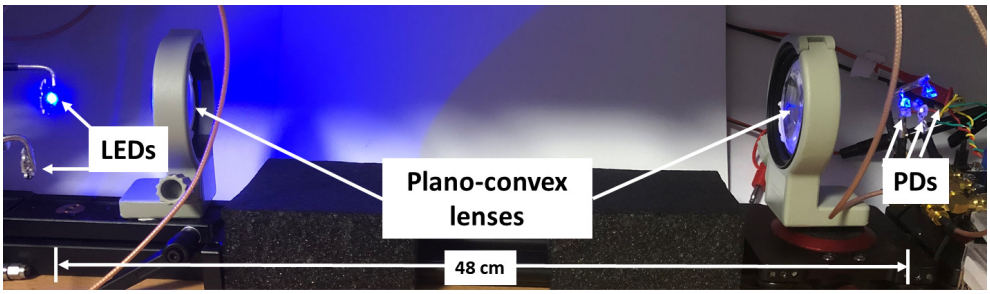

Figure 5. Photograph of optical component section.

4.3. Gesture Waveform Capture

As a proof of concept, the system was trained with five gestures with an increasing number of fingers as follows:

- Reference Rx signal (absence of movement),
- pointing up-down with 1 finger,
- pointing up-down with 2 fingers,
- pointing up-down with 3 fingers, and
- pointing up-down with 4 fingers.

The hand was moved up and down over a period of two seconds at a steady-rate corresponding to a natural hand gesture. As the separation between each finger is only about 3–5 mm, the sampling rate needs to be sufficiently high to capture the correspondingly short duration of light. The Rx signal is first down-sampled as the sampling rate is higher than the modulated light signal. The modulation is removed by finding the signal maxima and the resultant signal corresponding to 1–4 fingers present is shown in Figure 6 (top) to (bottom). The small peaks at the start of each cycle are due to the combined filtering response of the analogue and sample and hold circuitry in the digital storage oscilloscope. The response quickly decays and does not affect the operation of the system. The blocking of light by each finger results in low amplitudes and can be seen in each capture. In part, the accuracy can decrease as the number of fingers increase due to the re-

duced clarity of the raw signal. This reduction is partly offset however as classification improves when a signal has more unique features.

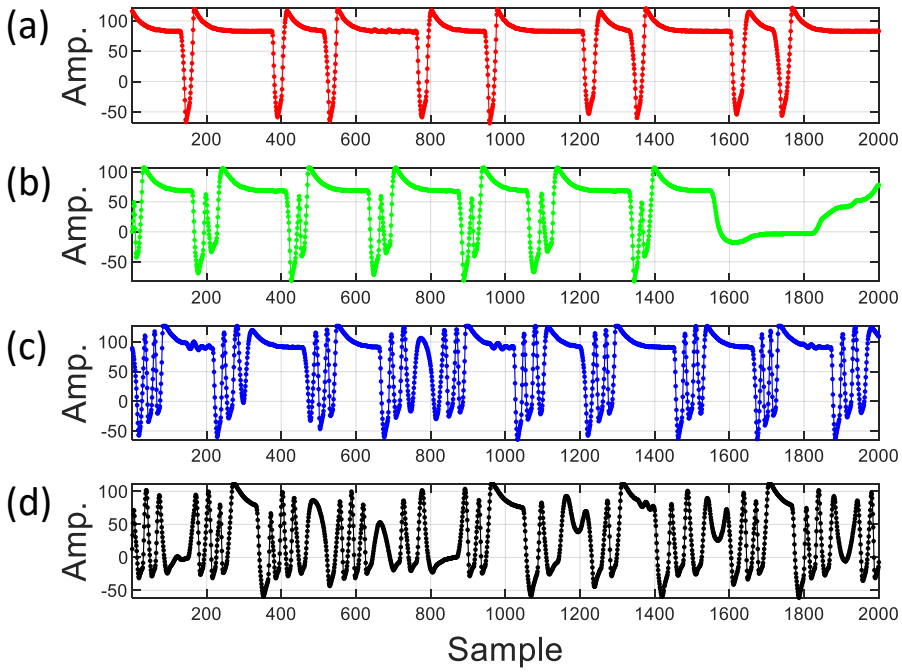

Figure 6. Received signal captured on VLC photodiode corresponding to (from top): (**a**) 1 finger; (**b**) 2 fingers; (**c**) 3 fingers; and (**d**) 4 fingers gestures.

4.4. Process Flow

There are three processing stages: signal-conditioning, training, and classification. Signal conditioning: The waveform sampled by the photo-diode undergoes signal conditioning prior to the identification. The signal magnitude is normalized so that the maximum value for each gesture is one. Gesture training: Data are collected for each of the 5 gestures. For each gesture, multiple frames are collected by repeating the movement over a period of two seconds. The data are then randomly split into two sets one for training and one for classification. This needs to be performed once on first use for each user, as they may have different movement styles and speed for the same gesture type. Gesture classification: The gestures are classified by ML. A practical gesture recognition system should be able to operate in real-time. Therefore a trade-off can be met between computational complexity and accuracy. We selected the LSTM algorithm as it offers a good performance to complexity ratio and is suitable for the repetitive sequential waveforms generated by hand gestures.

4.5. Signal Conditioning

The signal for training and categorization should encode the finger gesture and the performance should be relatively unaffected by the level of ambient light. Any reflected light from an object near to the PD should not result in a high amplitude signal that cannot be recognized from the same motion without reflection. Therefore, the signal should be normalized such that all signals have the same amplitude regardless of the ambient light intensity. The normalization scales the signal according to the minimum and peak signal level recorded over the measurement period. As the ambient light changes more slowly than the direct LED light across a measurement frame, this is a simple and efficient step. The recorded gesture features may vary slightly between each motion and also due to environment. Each user also presents their hands at a slightly different angle and moves them at a variable speed, and there will be temporal variations and potentially irregular random reflex movements. The natural light present in the morning will be different to

the artificial light in the evening and can vary if it is cloudy or sunny. All PDs exhibit a noise floor, and the TIA has a noise figure which contributes to a lowering of the signal integrity. Signal conditioning is required to manage these effects and to provide a clean representative signal which contains the essential features of each gesture to the ML algorithm. After conditioning, the Rx signal has range −1/+1 and is processed by the LSTM algorithm.

4.6. Training and Evaluation

As a proof of concept, data were collected for four different hands. The smallest span (from extended little finger to thumb) was measured as 16.3 cm and the largest hand had a span of 21.4 cm. Data were collected for the four hands on two separate measurement campaigns. During a first session, data were collected for training the neural network algorithm. A second validation session was conducted on the same day for evaluating the performance of the trained neural network. The data were divided equally into training and verification sets; that is, the training to verification ratio was 50% of all data. This figure is common in ML research and some systems use higher amounts of training to achieve high accuracies. Over-fitting can occur if the system is trained with too much data, and conversely, under-fitting if the training ratio is too low. The LSTM algorithm predicts the next sample in a sequence, and hence the most likely gesture classification, subject to the noise, variation, and irregularities present in human movement. The LSTM was trained using the stochastic gradient descent with momentum (SGDM) optimizer. This is a commonly applied solver with accelerated gradients to reduce the solving time [40]. After training, the LSTM was switched to validation mode in which a section from the nontraining set is evaluated. The output of the stochastic gradient solver can be sensitive to the initial random seed used and, therefore, a Monte Carlo type simulation was set-up averaging results over 50 cycles each with a different random seed. The accuracy and loss versus iteration performance for one of the random seed settings is shown in Figure 7.

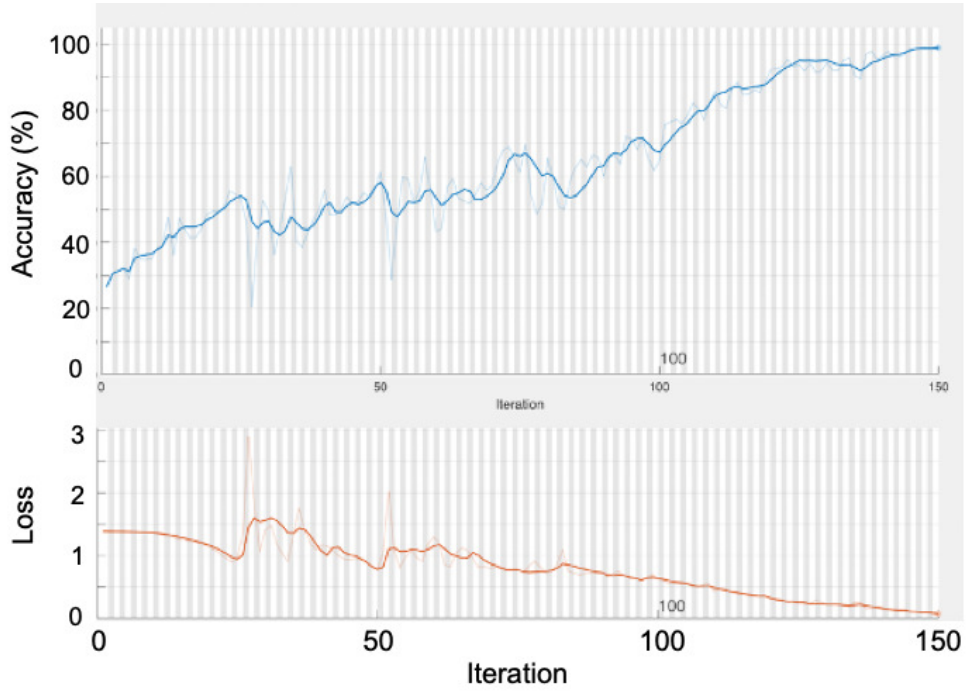

Figure 7. Performance of LSTM algorithm (**top**) Accuracy versus iteration and (**bottom**) Loss versus iteration.

5. Performance Evaluation

The VLC testbed was positioned square to a window with center at a diagonal distance of 4.65 m. The light through the window would enter the room in the direction of the VLC

receiver unit. There was no direct sunlight impinging on the Rx in this experiment due to an office-divider positioned between the window and the Tx unit.

A correct classification is determined when the actual and estimated gesture is identical. An average accuracy is computed for all gestures, users and tests per user. An example of predicted versus actual gesture accuracy is shown in Figure 8, for the case of a low number of iterations and sample-rate and demonstrates the frequency and duration of observed errors. There is good agreement between the actual and estimated gesture, and in this example, most errors occurred between the transition from two to three fingers.

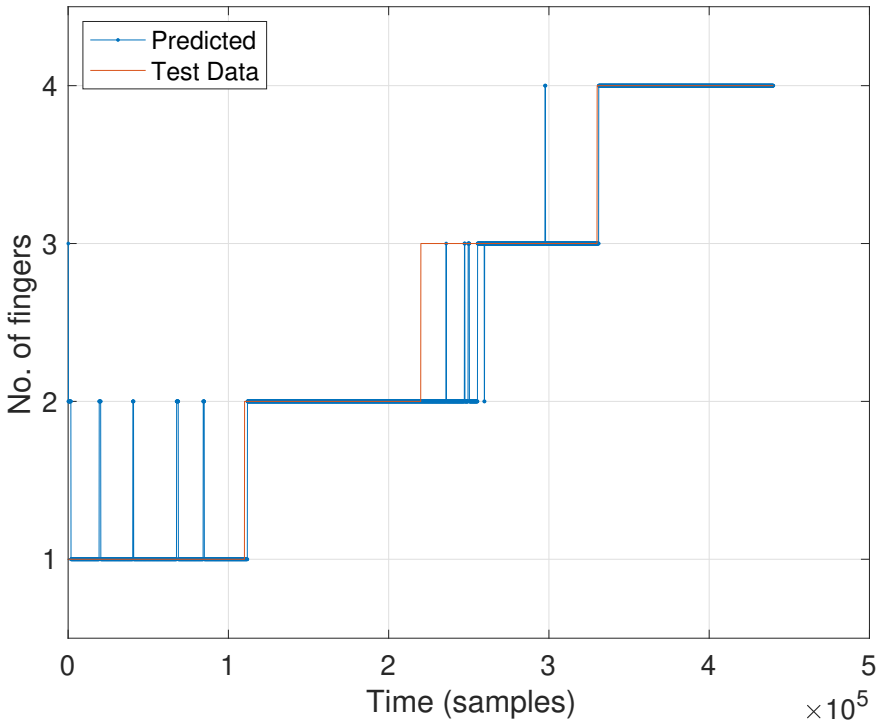

Figure 8. Predicted versus actual number of fingers in gesture.

Classification accuracy versus number of LSTM hidden-units is tabulated in Table 2 and plotted in Figure 9. The performance peaked at 75% accuracy for 50 hidden-units and gradually decreased as the number of units increased. The number of units should not be too large to avoid over-fitting. The performance is limited by the resolution of the input waveforms but can be improved by over-sampling the Rx signal in the presence of sampling and receiver noise. The classification accuracy increased to 88% when the number of samples per symbol increased by a factor of two and is due to the reduction in noise through averaging. We can compare this performance with other GR systems employing visible light using a single PD. Classification accuracies of 85% and 73% were achieved when the Tx-Rx separation was 20 cm and 35 cm, respectively, ref [29] with reflected light. Our accuracy could be further improved by employing a moving-average filter or wavelet denoising. Our performance may also increase by shortening the Tx-Rx separation from 48 cm. However, this is considered a realistic separation for a practical VLC system.

Table 2. Accuracy versus number of LSTM hidden-units.

Hidden-Units	Accuracy (%)
25	72
50	75
75	72
100	71
125	69
150	68
175	70
200	64
225	62

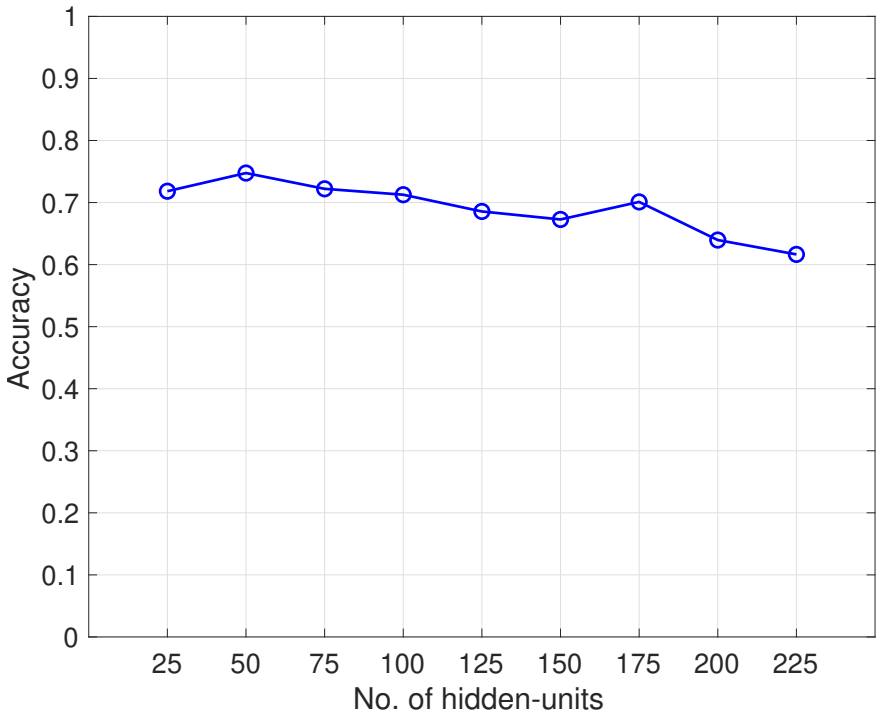

Figure 9. Accuracy versus number of LSTM hidden-units.

The speed of making a hand gesture depends on each individual. If the Rx is tracking, say, a robot arm, one could expect a highly regular pattern with near constant time intervals between blocking. However, there is a relatively large time variation with human gestures. Hand movements, even by the same person, move at a slightly different angle, speed, and position relative to the sensor. Therefore, the performance can depend on the sample-rate, and a system should be capable of increasing this to capture patterns from subjects who make very fast hand movements. Figure 10 shows the normalized performance figure-of-merit versus the sensor sample-rate. The normalized performance figure-of-merit in Figure 10 is computed by dividing the classification accuracy by the processing time and normalized to the highest value. From this result, we could select 0.25 MHz sampling-rate as providing a good performance to processing-time ratio. These results show that there are diminishing performance benefits from over-sampling when considering the added processing complexity. There are a number of VLC parameters that can affect the overall accuracy of the GR system. In particular, performance is sensitive to LED bias-voltage, which should be set high enough to enable communication over the required distance but not so high as to distort the waveform.

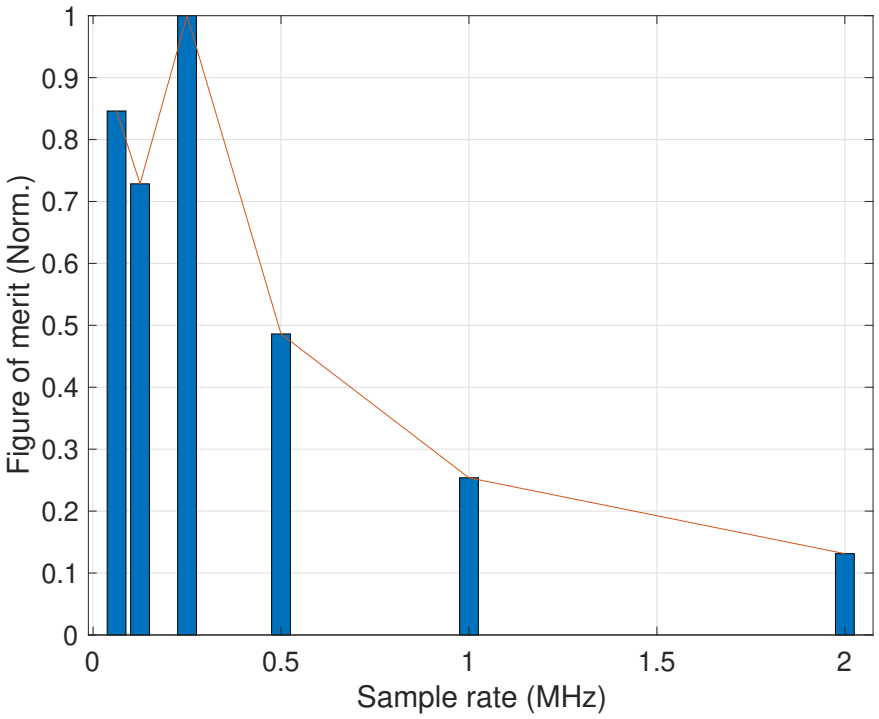

Figure 10. Normalized figure-of-merit versus sensor sample-rate.

6. Discussion

Human limbs generally do not move with a constant velocity, and different users may move their hands at a different speed. Depending on the point of capture, the finger may be accelerating or decelerating. To compensate, the signal can be time-scaled as a function of the finger velocity. For example, a person who moves their hand at half the speed of another person would have their signal sampled at half the rate. The duration of shadows generated by their fingers should then be approximately the same. Hand-speed could be determined by a variety of means offline or during a calibration, such as by mm-wave radar. It would also be possible to identify an individual from their unique finger signature, and this is an interesting area for future work.

6.1. Calibration

Light-intensity distribution may vary at different locations within a room. The natural changes in the ambient light level within limits should be managed by the amplitude normalization step. For optimized performance, a calibration should be made if the system is moved to a new location where the ambient light range may be different. The calibration routine which could quickly cycle through parameters such as Tx LED amplitude, equalizer coefficients, Tx-amp bias, and Rx TIA tuning to find optimized values. Alternatively, a look-up table can supply the coefficients based on the location, time of day, and season. Aging of components and heating may also result in drift, which can be resolved by a relatively in-frequent calibration once a week. The calibration routine could also be executed automatically once the system is first switched on.

6.2. Sensitivity to Hand Movement

Practical VLC systems require lenses to focus beams of light on the small photo-diode. If the hand is placed off-center, the Rx beam will be slightly off-focus and the accuracy may be reduced. This issue can be solved using an automatic lens or by employing multiple spatially separated PDs. An interesting alternative solution would be to employ the neural network to learn and predict gestures in cases where the beam is defocused. A study on the performance as a function of hand-offset position is considered as part of the future work.

6.3. Competing Systems and Cost

Assuming that a VLC infrastructure was established, the additional cost for the GR subsystem would mainly be due to the software development time. The cost of a dedicated gesture system is worth consideration. In our work, we employed relatively expensive and bulky AWG and DSO. The off-line processing could be conducted in real-time using a low-power microprocessor, such as the MSP430 from Texas Instruments, which includes built-in signal converters. One competitor to the optical system is an accelerometer based design that could be positioned on the wrist by a strap or as part of a smartwatch. However, a wrist-based transmitter unit would also be needed for relaying the accelerometer data to a receiving unit for further processing. A VLC-based system is still preferable in a hospital environment or for the elderly who may not own a smartwatch or smartphone.

6.4. Areas for Future Work

The number of recognizable gestures could be increased to include common sign-language ones. The system could be developed for general human activity recognition by extending the distance between LEDs and PD with their placement on the ceiling and/or wall. The duration of each shadow cast could be encoded as a binary sequence, and this could enable a probabilistic neural network to be employed for gesture pattern recognition, applying a similar approach to [41], where binary bits encoded a communications busy-idle state. We will investigate if there is any variability in the performance with different directions of sunlight and placement. However, this should not impact the system design. Finally, an automated VLC system should include an initial detection block which would be intermittently polled to recognize when a finger gesture is deliberately being performed.

7. Conclusions

This work described the design and implementation of a finger-gesture recognition system for visible light communication systems. The system employs a single low-cost LED at the Tx and a single photo-diode at the Rx and operates on the patterns of blocking of direct light by the finger motion. The LSTM algorithm can correctly categorize the finger gestures with an average accuracy of 88%, and the optimized number of hidden units was 50. A good performance-to-complexity state could be achieved by sampling the light at 250 kHz. The system has many applications in human-computer interaction, including health-care, commerce, and in the home. Our further work will focus on increasing the number of gestures and tasking the system with recognizing individuals from their gesture signatures.

Author Contributions: All authors contributed to the paper. Conceptualization & methodology, J.W., A.M. and R.T.; software, J.W.; validation, investigation, formal analysis, all authors; writing—original draft preparation, J.W., A.M. and R.T.; writing—review and editing, all authors; funding acquisition, A.M. All authors have read and agreed to the published version of the manuscript.

Funding: This work was partially supported by the Kuwait Foundation for Advancement of Sciences (KFAS) under Grant PR19-13NH-04.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

Conflicts of Interest: The authors declare that there are no conflict of interest regarding the publication of this paper.

References

1. Wang, C.; Liu, Z.; Chan, S. Superpixel-based hand gesture recognition with Kinect depth camera. *IEEE Trans. Multimed.* **2015**, *17*, 29–39. [CrossRef]
2. Li, W.; Xu, Y.; Tan, B.; Piechocki, R. Passive wireless sensing for unsupervised human activity recognition in healthcare. In Proceedings of the International Wireless Communications and Mobile Computing Conference (IWCMC), Valencia, Spain, 26–30 June 2017; pp. 1528–1533.
3. Bhat, S.; Mehbodniya, A.; Alwakeel, A.; Webber, J.; Al-Begain, K. Human Motion Patterns Recognition based on RSS and Support Vector Machines. In Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC), Seoul, Korea, 25–28 May 2020; pp. 1–6.
4. Bhat, S.; Mehbodniya, A.; Alwakeel, A.; Webber, J.; Al-Begain, K. Human Recognition using Single-Input-Single-Output Channel Model and Support Vector Machines. *Int. J. Adv. Comput. Sci. Appl. (IJACSA)* **2021**, *12*, 811–823. [CrossRef]
5. Alhammad, N.; Al-Dossari, H. Dynamic Segmentation for Physical Activity Recognition Using a Single Wearable Sensor. *Appl. Sci.* **2021**, *11*, 2633. [CrossRef]
6. Mucchi, L.; Jayousi, S.; Caputo, S.; Paoletti, E.; Zoppi, P.; Geli, S.; Dioniso, P. How 6G Technology Can Change the Future Wireless Healthcare. In Proceedings of the IEEE 2nd 6G Wireless Summit (6G SUMMIT), Levi, Finland, 17–20 March 2020; pp. 1–5.
7. Yang, Y.; Hao, J.; Luo, J.; Pan, S.J. Ceilingsee: Device-free occupancy inference through lighting infrastructure based led sensing. In Proceedings of the IEEE International Conference on Pervasive Computing and Communication (PerComs), Kona, HI, USA, 13–17 March 2017; pp. 247–256.
8. Xia, K.; Huang, J.; Wang, H. LSTM-CNN architecture for human activity recognition. *IEEE Access* **2020**, *8*, 56855–56866. [CrossRef]
9. Ordóñez, F.J.; Roggen, D. Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition. *Sensors* **2016**, *16*, 115. [CrossRef] [PubMed]
10. Guo, D.; Zhou, W.; Li, H.; Wang, M. Hierarchical lstm for sign language translation. In Proceedings of the AAAI Conference on Artificial Intelligence, New Orleans, LA, USA, 2–7 February 2018; Volume 32.
11. Du, C.; Ma, S.; He, Y.; Lu, S.; Li, H.; Zhang, H.; Li, S. Nonorthogonal Multiple Access for Visible Light Communication IoT Networks. *Hindawi Wirel. Commun. Mob. Comput.* **2020**, *2020*, 5791436. [CrossRef]
12. Kim, B.W. Secrecy Dimming Capacity in Multi-LED PAM-Based Visible Light Communications. *Hindawi Wirel. Commun. Mob. Comput.* **2017**, *2017*, 4094096. [CrossRef]
13. Wang, Z.; Chen, S. A chaos-based encryption scheme for DCT precoded OFDM-based visible light communication systems. *Hindawi J. Electr. Comput. Eng.* **2016**, *2016*, 2326563. [CrossRef]
14. Purwita, A.A.; Haas, H. Studies of Flatness of LiFi Channel for IEEE 802.11 bb. In Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC), Seoul, Korea, 25–28 May 2020; pp. 1–6.
15. Ghassemlooy, Z.; Alves, L.; Zvanovec, S.; Khalighi, M. (Eds.) *Visible Light Communications: Theory and Applications*; CRC Press: Boca Raton, FL, USA, 2017.
16. Ding, W.; Yang, F.; Yang, H.; Wang, J.; Wang, X.; Zhang, X.; Song, J. A hybrid power line and visible light communication system for indoor hospital applications. *Comput. Ind.* **2015**, *68*, 170–178. [CrossRef]
17. An, J.; Chung, W. A novel indoor healthcare with time hopping-based visible light communication. In Proceedings of the IEEE 3rd World Forum on Internet of Things (WF-IoT), Reston, VA, USA, 12–14 December 2016; pp. 19–23.
18. Lim, K.; Lee, H.; Chung, W. Multichannel visible light communication with wavelength division for medical data transmission. *J. Med. Imaging Health Inform.* **2015**, *5*, 1947–1951. [CrossRef]
19. Tan, Y.; Chung, W. Mobile health–monitoring system through visible light communication. *Bio-Med. Mater. Eng.* **2014**, *24*, 3529–3538. [CrossRef] [PubMed]
20. Jerry Chong, J.; Saon, S.; Mahamad, A.; Othman, M.; Rasidi, N.; Setiawan, M. Visible Light Communication-Based Indoor Notification System for Blind People. In *Embracing Industry 4.0*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 93–103.
21. Zhang, C.; Tabor, J.; Zhang, J.; Zhang, X. Extending mobile interaction through near-field visible light sensing. In Proceedings of the ACM International Conference on Mobile Computing and Networking, MobiCom '15, Paris, France, 7–11 September 2015; pp. 345–357.
22. Sewaiwar, A.; Vikramaditya, S.; Chung, Y.-H. Visible light communication based motion detection. *Opt. Express* **2015**, *23*, 18769–18776. [CrossRef] [PubMed]
23. Cheng, H.; Chen, A.M.; Razdan, A.; Buller, E. Contactless gesture recognition system using proximity sensors. In Proceedings of the IEEE International Conference on Consumer Electronics (ICCE), Berlin, Germany, 6–8 September 2011; pp. 149–150.
24. Wu, J.; Pan, G.; Zhang, D.; Qi, G.; Li, S. Gesture recognition with a 3-d accelerometer. In *Proceedings of the International Conference on Ubiquitous Intelligence and Computing*; Springer: Berlin/Heidelberg, Germany, 2009; pp. 25–38.
25. Wang, Z.; Chen, B.; Wu, J. Effective inertial hand gesture recognition using particle filtering based trajectory matching. *Hindawi Wirel. Commun. Mob. Comput.* **2018**, *1*, 1–9. [CrossRef]
26. Liu, G.; Kong, D.; Hu, S; Yu, Q.; Liu, Z.; Chen, T. Smart electronic skin having gesture recognition function by LSTM neural network. *Appl. Phys. Lett.* **2018**, *113*, 084102. [CrossRef]
27. Venkatnarayan, R.H.; Shahzad, M. Gesture recognition using ambient light. *ACM Interact. Mob. Wearable Ubiquitous Technol.* **2018**, *2*, 1–28. [CrossRef]

28. Kaholokula, M.D.A. Reusing Ambient Light to Recognize Hand Gestures. Undergraduate Thesis, Dartmouth College, Hanover, NH, USA, 2016.
29. Yu, L.; Abuella, H.; Islam, M.; O'Hara, J.; Crick, C.; Ekin, S. Gesture Recognition using Reflected Visible and Infrared Light Wave Signals. *arXiv* **2020**, arXiv:2007.08178.
30. Huang, M.; Duan, H.; Chen, Y.; Yang, Y.; Hao, J.; Chen, L. Demo Abstract: FingerLite: Finger Gesture Recognition Using Ambient Light. In Proceedings of the INFOCOM 2020-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), Toronto, ON, Canada, 6 July 2020; pp. 1268–1269.
31. Choi, J.W.; Ryu, S.J.; Kim, J.H. Short-range radar based real-time hand gesture recognition using LSTM encoder. *IEEE Access* **2019**, *7*, 33610–33618. [CrossRef]
32. Pinto, R.F.; Borges, C.D.; Almeida, A.; Paula, I.C. Static hand gesture recognition based on convolutional neural networks. *Hindawi Wirel. Commun. Mob. Comput.* **2019**, *2019*, 4167890. [CrossRef]
33. Zhu, G.; Zhang, L.; Shen, P.; Song, J. Multimodal gesture recognition using 3-D convolution and convolutional LSTM. *IEEE Access* **2017**, *5*, 4517–4524. [CrossRef]
34. Ma, C.; Wang, A.; Chen, G.; Xu, C. Hand joints-based gesture recognition for noisy dataset using nested interval unscented Kalman filter with LSTM network. *Vis. Comput.* **2018**, *34*, 1053–1063. [CrossRef]
35. Jian, C.; Li, J.; Zhang, M. LSTM-based dynamic probability continuous hand gesture trajectory recognition. *IET Image Process.* **2019**, *13*, 2314–2320. [CrossRef]
36. Barry, J.R. *Wireless Infrared Communications*; Kluwer Academic Publishers: Norwell, MA, USA, 1994.
37. Komine, T.; Nakagawa, M. Fundamental analysis for visible-light communication system using LED lights. *IEEE Trans. Consum. Electron.* **2004**, *50*, 100–107. [CrossRef]
38. Do, T.; Junho, H.; Souhwan, J.; Yoan, S.; Myungsik, Y. Modeling and analysis of the wireless channel formed by LED angle in visible light communication. In Proceedings of the International Conference on Information Networking (ICOIN2012), Bali, Indonesia, 1–3 February 2012; pp. 354–357.
39. Greff, K.; Srivastava, R.; Koutník, J.; Steunebrink, B.; Schmidhuber, J. LSTM: A search space odyssey. *IEEE Trans. Neural Netw. Learn. Syst.* **2016**, *28*, 2222–2232. [CrossRef] [PubMed]
40. Postalcıoğlu, S. Performance analysis of different optimizers for deep learning-based image recognition. *Int. J. Pattern Recognit. Artif. Intell.* **2020**, *34*, 2051003. [CrossRef]
41. Webber, J.; Mehbodniya, A.; Hou, Y.; Yano, K.; Kumagai, T. Study on Idle Slot Availability Prediction for WLAN using a Probabilistic Neural Network. In Proceedings of the IEEE Asia Pacific Conference on Communications (APCC'17), Perth, Australia, 11–13 December 2017.