ISSN: 2754-4516

Journal of Medicine and Healthcare

SCIENTIFIC Research and Community

Research Article

Open Access

Sign Language Recognition System Using Cnn

Rung-Shiang, Lee, Hung-Yuan Huang, Pei-Zhen Lin, Hao-Rui,Wu, Ting-Shyuan Lu and I-Te Chen*

Department of Healthcare Administration and Medical Informatics, Kaohsiung Medical University

ABSTRACT

In this study, we have built an automatic sign language translation system for deaf and dumb persons to communicate with ordinary people. According to the Statistics Department of the Taiwan Ministry of Health and Welfare, there are 119,682 hearing impaired persons, and 14,831 voice function or language dysfunctions. The Deaf and dumb persons' account for 11.7% of the population with physical and mental disabilities. However, there are only 488 qualified people with the sign language translation skill certificate, which shows the importance of automatic sign language translation systems.

This system collects 11 signals including five fingers' curvature, 3-axis gyroscope and 3-axis accelerometer from left and right hand separately. In addition, a total of 22 signals are collected by the two sensors, Flex sensor and GY-521 six-axis with single-board computer Arduino MEGA 2560; and then uploaded to server via ESP-01S Wi-Fi module. While server receives the 22 signals, it converts to a RGB picture using PHP program. As a result, we can compare the picture with the model trained by TensorFlow and the compared result is stored in the database. Meanwhile, the comparison stored in database which can be accessed by APP programs would be displayed on the screen of the mobile device and be read aloud.

The TensorFlow training model collects 25 sign language gestures, each based on 100 training gesture pictures, and a sign language recognition training model is Convolutional Neural Network (CNN). In this study, the results of the sign language recognition training model are further confirmed by 10 people other than those in training database. So far, the indeed recognition rate of sign language is about 84.4%, and the system response time is about 2.243 seconds.

***Corresponding author**

I-Te Chen, Department of Healthcare Administration and Medical Informatics, Kaohsiung Medical University. Email: itchen@kmu.edu.tw

Received: December 15, 2020; **Accepted:** December 21, 2020; **Published:** December 30, 2020

Keywords: Deaf and dumb persons, Sign language recognize, TensorFlow, CNN

Introduction

Sign language is an essential communication language for deaf and dumb persons. However, ordinary people are not familiar with sign language and are unable to communicate with them. Although the Taiwanese government subsidizes deaf and dumb persons to apply for translators, the ratio of sign language translators to deaf and dumb persons is roughly 1: 300 and subsidies are only 30 hours per month in the workplace [1, 2]. Non-workplace or overtime service must be paid at their own expenses. Communication barriers seriously affect the life and survival of deaf and dumb persons. Therefore, this study hopes to develop a sign language recognition system to eliminate the gap between deaf and dumb persons and ordinary people and to improve the quality of life and socioeconomic status of deaf and dumb persons.

Due to the rapid development of science and technology, devices that controlled by gestures have increased. For example: doors or windows can be commanded by serial gestures [3]. and LED lighting can be switched and brightness to be tuned by hand-sensing gloves [4].

In recent years, the maturity of machine learning such as TensorFlow and the improved computing power have applied to recognize gestures. We also hope to develop a sign language recognition and communication system so as to display the translation results on the mobile device; and be read by devices themselves. Therefore, this system can reduce the general public's inconvenience to communicate with deaf and dumb persons due to lack of knowledge of sign language. This study has three contributions as followed:

Firstly, a real-time sign language recognition communication system can be used to improve the life quality of deaf and dumb persons. Secondly, compared with text-based communication software such as Telegram, Skype, Line, and Messenger, real-time and effective sign language translation can not only reduce input process, but also let us to communicate more intuitively. Thirdly, the Taiwanese government provides deaf and dumb persons to apply 30 service-hour of sign language translators for free each month in the workplace. However, 30 service-hour is not enough for translators to accompany them in daily life. Hence, this system can prevent translators from reducing their enthusiasm and willingness because of overtime working to achieve a win-win situation.

Literature review

Sign language recognition is mainly divided into two categories: one is based on image recognition, which uses a camera to capture gestures and facial expressions as images for recognition. The other one is to formulate spacial information with contact points or sensors (such as Flex Sensor, Six-axis sensor) on wearable devices (such as gloves) which can capture user's hand movement for the server to recognize the signal. The two type of recognition we discussed as following:

Citation: I-Te Chen, et al (2020) Sign Language Recognition System Using Cnn. Journal of Medicine and Healthcare. SRC/JMHC-158.
DOI: doi.org/10.47363/JMHC/2020(2)138

Sign language recognition based on photography

Shi processed series of images of the gestures taken by the camera through image background subtraction, noise removal, feature extraction; and then put them into a neural network for identification [5]. However, Shi 's proposed scheme will be affected by the light source in the environment, the background, and the angle of the gesture offset. As a result, Shi 's proposed scheme will reduce the recognition rate and cause misjudgment of gestures. Figure 1 is the interface of the scheme:

Figure 1: Shi 's Interface of static gesture recognition scheme

In Zhang et al. established a gesture password unlocking system [3]. They used Kinect to execute skeletal tracking. When the initial gesture is detected, the system can quickly know the current coordination of the hand and use the skeletal tracking within a certain period of time as the basis for future comparisons. Zhang et al. only recognize 5 gestures, and the shape of gestures may cause recognition errors due to changes in angle and distance.

Sign recognition based on gloves

Huang proposed a gesture recognition system with sensors and gloves [6]. The system uses the Kalman filter detection method to distinguish the sign language and the results from sensor to make a preliminary decision tree (Figure 2(b)). Then perform probability neural network to recognize sign language. In the sign language for recognition is the alphabet system (Figure 2(a)) which presented by one singal hand [6]. Except J and Z have to be recognized with gesture and movement, all other 24 alphabets would be recognize by still gesture, which is based on the degree of fingers bending and the direction of the wrist.

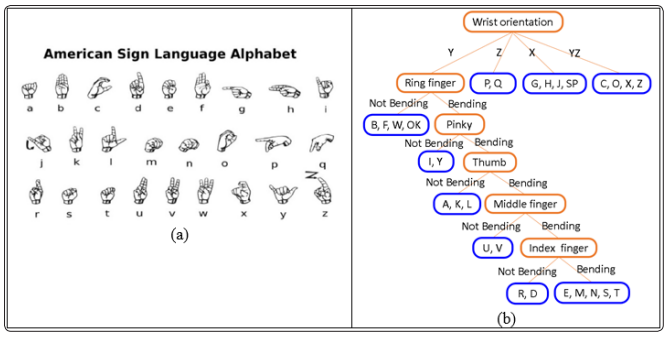

Figure 2: (a) American sign Language Alphabet (b) Preliminary Decision tree

Chen established a set of 5DT Data glove 5 Ultra gesture to control a single role in virtual system [7]. The 5DT Data glove 5 Ultra is connected with a wearable optical fiber as a measurement sensor to detect the bending angle of the fingers. Moreover, using Inter Trax2, the 8 sensors in the Inter Trax2 provide stable and rapid angular velocity which can locate the gloves spacial inforation. Finally, the signals collected by 5DT Data glove 5 Ultra and Inter Trax2 were used to identify different gestures to control the movement and behavior of the virtual role.

Figure 3: (a) 5DT Data glove 5 Ultra (b) InterTrax2

The recognition method based on image processing is convenient for users because wearing additional devices is unnecessary. However, before recognition, the image must be pre-processed to obtain skeletal information. And the complicated steps result in high recognition error rate. In addition, if the camera that captures the hand image is insufficient, there may be blind spots in the image, which means complete hand information cannot be obtained to cause recognition errors.

Although this problem can be solved by increasing the number of cameras, the number of images to be processed will increase and the amount of calculations will increase consequently. Finally, it is necessary to consider that the images with hand overlapping or with different environment background factors will result in misjudgment to increase recognition errors.

On the other hand, the wearable device based recognition method requires additional equipment, but the information obtained by the sensor is relatively reliable. Moreover, there is no blind spots problem and less calculation amount. Therefore, the wearable device based recognition method is suitable as a real-time identification system; and this study choose the glove device as the method for sign language recognition. Nevertheless, the 5DT Data glove 5 Ultra plus InterTrax2 mentioned in is expensive [7]. In this study, we will explore the feasibility of less expensive devices.

Sign language recognition system architecture and implementation

This session is divided into two parts: system architecture and implementation. And the implementation is divided into hand detection, server calculation, and model building.

Architecture

This system collects 22 signals including five fingers' curvature, 3-axis gyroscope and 3-axis accelerometer by the two sensors, Flex sensor and GY-521 six-axis from left and right hand respectively and all signals are processed by single-board computer Arduino MEGA 2560; and then uploaded to server via ESP-01S Wi-Fi module. While server receives the 22 signals, it converts to a RGB picture using PHP program. As a result, we can compare the picture with the model trained by TensorFlow and the compared result is stored in the database. Meanwhile, the comparison stored in database which can be accessed by APP programs would be displayed on the screen of the mobile device and be read aloud. The system architecture shows as Figure 4.

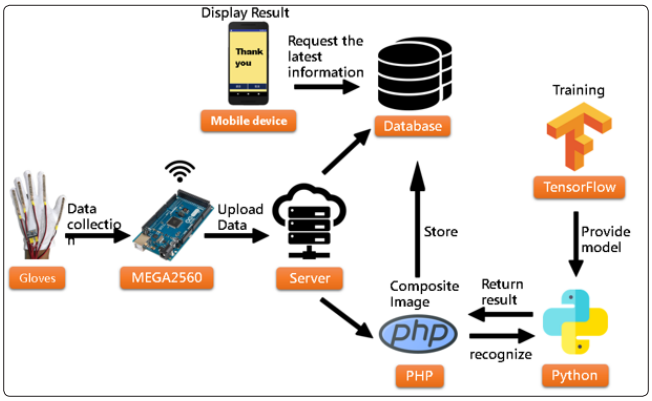

Figure 4: Sign language recognition system architecture

Implementation

In this study, cotton gloves are used, and the sensor is stitched onto the gloves (Figure 5(a)) to read the hand gesture signals.

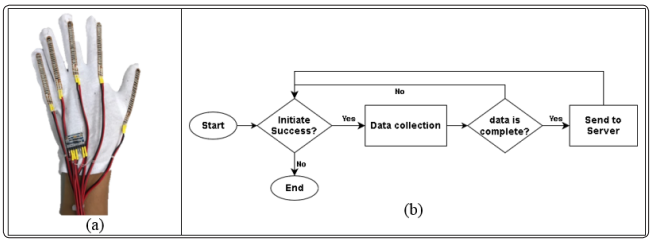

Figure 5: (a) Gloves with sensor (b) Signal collection flowchart

We choose the Arduino MEGA 2560 as the development board, and use the Arduino IDE to develop the hand signal collection system which collects the signal of the curvature sensor and the GY-521 six-axis sensor. Accordingly, the collected signals are uploaded to the server through the ESP-01S Wi-Fi module.

Our ten fingers require ten PWM interface pins to collect the curvature. The commonly used Arduino UNO board only has 6 PWM pins, so we choose Arduino MEGA 2560 with 15 PWMs for development. In addition, ESP-01S requires serial port, and MEGA 2560 provides 3 serial ports to better control ESP-01S Wi-Fi module.

We use two important sensors to collect signals: the first one is a curvature sensor that can digitize the degree of finger bending into a value in the range from 0 ~ 1023; the second one is GY-521, which provides three-axis acceleration and three-axis angular velocity to understand hand direction and movement. The signal collection process is shown in Figure 5 (b):

The software and hardware specifications of the server are shown in Table 1:

Table 1: Server software and hardware specifications

Hardware	Software
CPU: Xeon E3-1230 v3 3.30GHz **RAM:**16 G	**OS:** CentOS 7.6 (1810) **Software:** Apache 2.4.6 PHP 7.2.18 Maria DB 5.5.60 Tensor Flow 1.13.1 Python 3.7.3 Anaconda 4.6.14

CentOS is a stable OS, and Apache + PHP is a bridge to communicate with the front-end gloves. When the glove data collection is completed then uploaded, and stored in the Maria DB database; the data would be combined into pictures. Furthermore, the method of composing pictures is to convert the original data into the range from 0 ~ 255, which means the left-hand data stored into the R layer, the right-hand data into the G layer, and we fix the B layer value as 0 to generate a RGB picture. The width of the picture is 11 data of one hand, which are the curvature of five fingers, tri-axial acceleration, and tri-axial angular velocity. Figure 6 is a composing method of generating pictures.

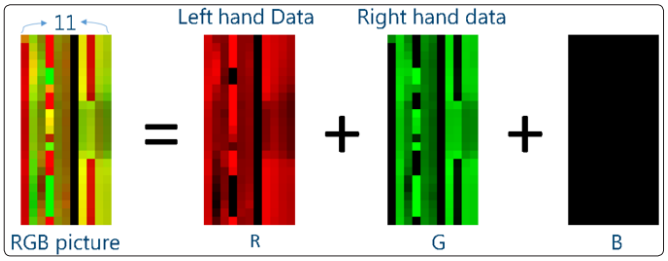

Figure 6: composing method of generating pictures

After composing the pictures, we can compare and classify the pictures with the model trained by CNN in next step. Then, the results will be stored in the database and be provided to the APP for query and display.

We choose the Maria DB database of Open source, which can reduce the system construction cost. There are three tables in the database (Figure 7), which are the **User** table for user's information storage, the **Translate** table for comparison results storage, and the **Data** table for the original 22 signals storage that come from the gloves.

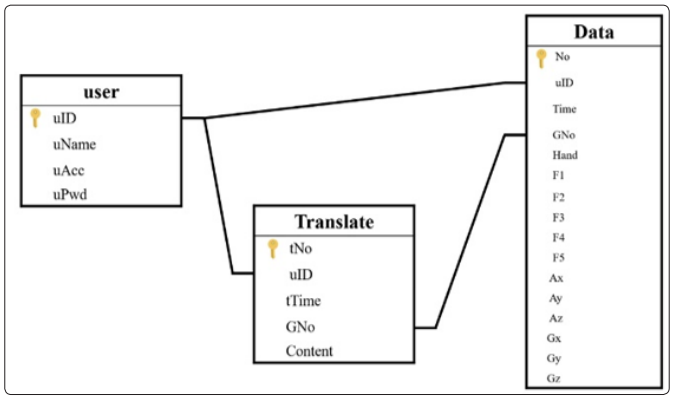

Figure 7: Database Entity-Relationship Diagram

Neural network (NN) is the most important technology of this system. We use TensorFlow training CNN model because of stronger fault tolerance due to potential gestures variances from each person. Besides, CNN is insensitive to these variances and can correctly classify pictures. In addition, CNN convolutional layer can capture features more precisely. The CNN architecture of this system is shown in Figure 8:

Citation: I-Te Chen, et al (2020) Sign Language Recognition System Using Cnn. Journal of Medicine and Healthcare. SRC/JMHC-158.
DOI: doi.org/10.47363/JMHC/2020(2)138

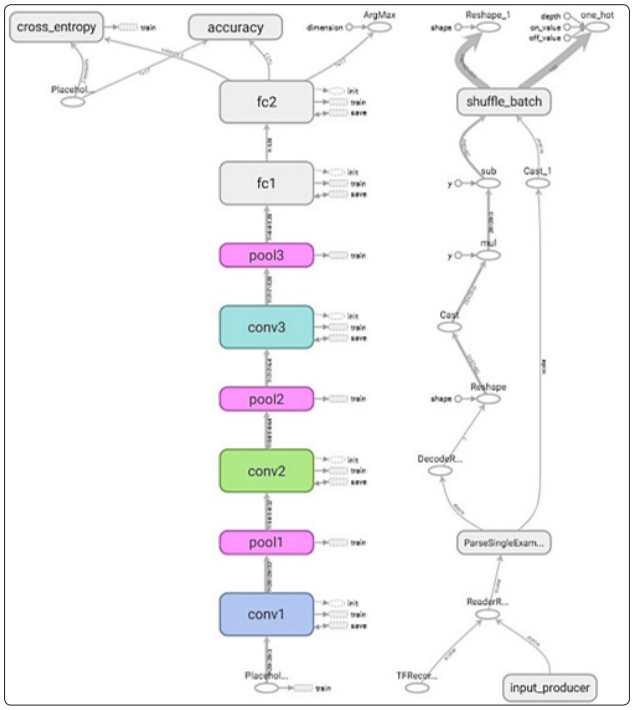

Figure 8: System CNN architecture

In Figure 8, we use three convolutional layers, three pooling layers, and two fully connected layers. Furthermore, the number of neuron nodes is 32, 64, and 128 to capture features in more detail. The input image size is 44 * 44 * 3, and the output is 25 categories; the label index is converted into one-hot form. The convolutional layers use a 3 * 3 filter and the padding parameter is set to **SAME** to retain the boundary information. Activation function uses **ReLU** instead of **Sigmoid**. Since in the NN back-propagation architecture, the derivative of Sigmoid is 0.25; there will be a problem of gradient disappearance, resulting in low learning efficiency. Moreover, Sigmoid needs to be normalized, otherwise it will lose its characteristic performance.

The pooling layers use **Max Pooling** to extract the maximum value in a 2 * 2 filter; thereby reducing the size of the matrix and the amount of calculation in order to prevent the problem of over-fitting. Next layers are two fully connected layers; the first layer converts a two-dimensional matrix into a one-dimensional vector to be input into the second fully connected layer, which outputs comparison result. Comparison results are actually expressed in a percentage form with **Softmax** where the largest value is also the highest probability, then generated the final output result.

The system currently trains 25 sign language gestures, each based on 100 training gesture pictures. It is important to prevent over-fitting when building CNN model. Over-fitting means that the model adapt uniquely one training data set and might not be able to recognize general sign language correctly in real world. The most common cause of over-fitting is insufficient model samples. In order to detect whether the model is over-fitting in our study, the training pictures are divided into 80% training data and 20% verifying data during the training process for a total of 500 times. If there is over-fitting, the accuracy of the training set will be high, but the accuracy of the verifying set will show an incorrect downward trend, resulting in poor sign language recognition. After the above operations, the accuracy rate of this model training set is 94.66%. Figures 9 is the loss functions and accuracy rates, in which the verifying set accuracy rate is 93.33%, where there is no much difference from the training set. Thus we could claim our study is a generalized model without overfitting problem.

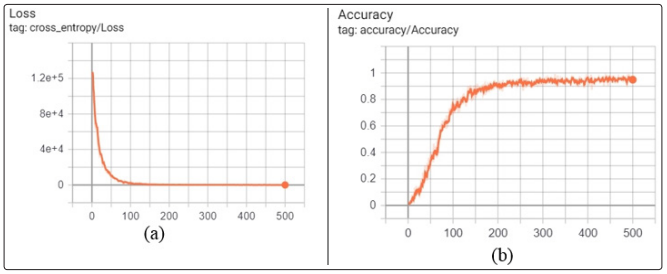

Figure 9: (a) Loss functions (b) Accuracy rates

We using Android Studio develop an APP to display the recognition results (Figure 10 (a)). After logging in to the mobile device and pressing the "Get Data" button, the system will request the server for the latest translation data of the user within 30 seconds to be displayed on the screen and read it aloud. APP flowchart is shown in Figure 10 (b):

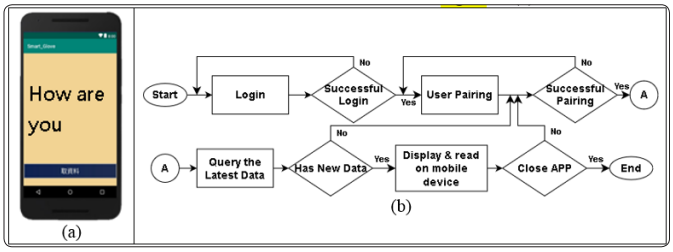

Figure 10: (a) Results show on the mobile device via APP (b) APP flowchart

Experimental Results

In this study, we collect the signals of 25 sign language generated by ten persons respectively and the accuracy rates are shown in Table 2, which were used to test the trained model described in the previous session. We find that the accuracy of the word "Taipei" is low; preliminary estimates may be result from the characteristics of "Taipei" are not obvious and the current number of samples leaves room for improvement. Overall, the average accuracy rate is 84.4%, which is different from the model prediction accuracy rate of 93.33%. In the future, we will increase the number of samples in the training set to achieve sample diversity, thereby improving the accuracy of recognition.

Table 2: Test results: O: correct & X: incorrect

Gesture	Can		Complete		Teacher		But		How are you	
	O	X	O	X	O	X	O	X	O	X
Frequency	10	0	10	0	9	1	9	1	9	1
Recognition rate	100%		100%		90%		90%		90%	
Gesture	Hope		I		Friend		Many		Yes	
	O	X	O	X	O	X	O	X	O	X
Frequency	10	0	9	1	6	4	10	0	9	1
Recognition rate	100%		90%		60%		100%		90%	
Gesture	Personality		Kaohsiung		Great		Get rich		Happy	
	O	X	O	X	O	X	O	X	O	X
Frequency	10	0	6	4	7	3	8	2	8	2
Recognition rate	100%		60%		70%		80%		80%	
Gesture	Understanding		Environment		Deaf		Work		Go home	
	O	X	O	X	O	X	O	X	O	X
Frequency	8	2	7	3	8	2	7	3	10	0
Recognition rate	80%		70%		80%		70%		100%	
Gesture	Good Morning		Taipei		Exhausting		Think		Thank you	
	O	X	O	X	O	X	O	X	O	X
Frequency	10	0	4	6	10	0	10	0	7	3
Recognition rate	100%		40%		100%		100%		70%	
Average recognition rate: 84.4%										

The response time of our system is the duration between the data uploaded to mobile device and the how long it takes to display the results. We record the time required for 25 sign language words to be translated by the system in Table 3. The results show that the average response time of this system is about 2.243 seconds.

Table 3: System Response Time

Gesture	Can	Complete	Teacher	But	How are you
Time	2.209 s	2.19 s	2.231 s	2.272 s	2.236 s
Gesture	Hope	I	Friend	Many	Yes
Time	2.222 s	2.245 s	2.293 s	2.2 s	2.249 s
Gesture	Personality	Kaohsiung	Great	Get rich	Happy
Time	2.222 s	2.239 s	2.272 s	2.23 s	2.271 s
Gesture	Understanding	Environment	Deaf	Work	Go home
Time	2.255 s	2.255 s	2.214 s	2.272 s	2.249 s
Gesture	Good Morning	Taipei	Exhausting	Think	Thank you
Time	2.257 s	2.211 s	2.22 s	2.359 s	2.21 s
Average recognition time: 2.243 s					

Conclusions and Discussion

In this study, we proposed and implemented sign language recognition system using Flex sensor and GY-521 six-axis sensors. And we collected 25 sign language gestures, 100 training pictures for each gesture, and trained by CNN. In addition, we found a user whose gestures are not in the training set to test the accuracy of the system. The test results show that we achieved 84.4% accuracy, and the response time of the recognition system was about 2.243 seconds [8-11].

However, we used a lot of device wires currently, which might reduce user mobility and make it difficult to put on and take off. In the future, this system will be improved by using two Arduino Nano on the back of the hand to reduce the wire. And the data of the two hands are sent to the database separately to improve the transmission speed. Moreover, expanding the word bank is also the future work we need to do.

Since ESP-01S uploads need to wait for server feedback; and the two ESP-01S upload processes need to wait about 1.5 seconds. And the APP requests data from the server every 0.3 seconds; so the average response time is 2.243 seconds. Thus, how to quickly upload hand signal data is our future work. In terms of improving the recognition rate, the APP will be designed to allow users feedback whether the translation is correct. Moreover, the correct data of more users should be added to the training set, so that the accuracy and recognition rate of the model can be better.

Although modern communication tools are developed, deaf and dumb persons can also use Skype, LINE, IG, Telegram, Messenger

and other communication software to communicate. But compared to text messages, being able to communicate face to face is the most humane. The use of our equipment allows ordinary people to communicate with deaf and dumb persons without the need the knowledge of sign language. Consequently, it can reduce the translation fee for deaf and dumb persons in Taiwan.

Our current understanding of sign language is still insufficient. If we can cooperate with deaf and dumb persons more in the future, we will have a deeper understanding of sign language and hope to translate sign language more accurately. At present, our system can only translate separate words. In the future, it is expected that without pause between words, the entire sentence can be translated instantly, and making communication more efficient. In addition, sign language is often used with face expressions to express different meanings. We hope that in the future, under the premise of privacy permission, other image sensors can be combined to make the translation more accurate.

Author Contributions

This article was proposed by Rung-Shiang, Lee & I-Te Chen; software, Hung-Yuan Huang, & Hao-Rui,Wu; validation, Pei-Zhen Lin & Ting-Shyuan Lu; writing—original draft preparation, Rung-Shiang, Lee, Hung-Yuan Huang; writing—review & editing and project administration, I-Te Chen.

Acknowledgments

This work was supported in part by Kaohsiung Medical University under grant: KMU-M109007, and Department of Medical Research, Kaohsiung Medical University Hospital.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ministry of Health and Welfare, Statistics on the Number of People with Physical Disabilities, Available online: https://dep.mohw.gov.tw/dos/cp-2976-13815-113.html, accessed on: 20 October,2020.
2. Ministry of Labor, Qualified Skills, Available online: https://statdb.mol.gov.tw/statis/jspProxy.aspx?sys=210&kind=21&type=1&funid=q12023, accessed on: 20 October,2020.
3. Jue-wei Zhang , Yu-hao Lin (2015) A Hand Gesture and Shape Matching Access Control System Using Depth Point Cloud Technology, JITAS, 15: 79-96.
4. Yung-Tse Wu (2010) Hand Sensation LED Lighting Control System, Master Thesis, National Cheng Kung University.
5. Che-Yu Shih (2012) Computer-Vision Based Real-Time Hand Gesture Recognition and Its Application, Master Thesis, Yuan Ze University.
6. Pin-Hsun Huang (2015) Dynamic hand gesture recognition based on adaptive Kalman filter, Master Thesis, Tamkang University.
7. Min-Ting Chen (2007) The Development of Hand Gesture Control Avatar System by Using Data Glove, Master Thesis, Chung Yuan Christian University.
8. Qi Sheng Chen (2020) Shuming Zhang, Zhang Xiang Zhang, Yu Kai Wang, Smart Gloves, Available online: https://www.easonchang.com/2018/04/01/smart-gloves.
9. UW News(2016), UW undergraduate team wins $10,000 Lemelson-MIT Student Prize for gloves that translate sign language, Available online: https://www.washington.edu/news/2016/04/12/uw-undergraduate-team-wins-10000-lemelson-mit-student-prize-for-gloves-that-translate-sign-language , accessed on: 20 October,2020.
10. Bai Xuan Wu, Sign language instant translation gloves, You can also say Hello, Available online: http://news.ltn.com.tw/news/life/paper/922837, accessed on: 20 October,2020.
11. Taiwan Sign Language Online Dictionary, Available online: http://lngproc.ccu.edu.tw/TSL , accessed on: 20 October,2020.