




## Article

# Designing an Interactive Communication Assistance System for Hearing-Impaired College Students Based on Gesture Recognition and Representation

Yancong Zhu <sup>1,2</sup>, Juan Zhang <sup>2,3</sup>, Zhaoxi Zhang <sup>4</sup>, Gina Clepper <sup>5</sup>, Jingpeng Jia <sup>6,\*</sup> and Wei Liu <sup>2,\*</sup> 

<sup>1</sup> School of Biological Science and Medical Engineering, Beihang University, Beijing 100191, China; by1410104@mail.buaa.edu.cn or yancong.zhu@bnu.edu.cn

<sup>2</sup> Faculty of Psychology, Beijing Normal University, Beijing 100875, China; 201828061061@mail.bnu.edu.cn or zhangjuan01@crhms.cn

<sup>3</sup> Human Resources Department, China Reform Health Management and Services Group Co., Ltd., Beijing 100028, China

<sup>4</sup> School of Engineering and Applied Sciences, University at Buffalo, Buffalo, NY 14260, USA; zhaoxizh@buffalo.edu

<sup>5</sup> College of Engineering, Purdue University, West Lafayette, IN 47907, USA; gclepper@alumni.purdue.edu

<sup>6</sup> College of Special Education, Beijing Union University, Beijing 100075, China

\* Correspondence: tjtingpeng@bnu.edu.cn (J.J.); wei.liu@bnu.edu.cn (W.L.)



**Citation:** Zhu, Y.; Zhang, J.; Zhang, Z.; Clepper, G.; Jia, J.; Liu, W. Designing an Interactive Communication Assistance System for Hearing-Impaired College Students Based on Gesture Recognition and Representation.

*Future Internet* **2022**, *14*, 198.

<https://doi.org/10.3390/fi14070198>

Academic Editors: Christos Troussas, Akrivi Krouska, Cleo Sgouropoulou and Jaime Caro

Received: 25 May 2022

Accepted: 28 June 2022

Published: 29 June 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Abstract:** Developing a smart classroom can make the modern classroom more efficient and intelligent. Much research has been conducted pertaining to smart classrooms for hearing-impaired college students. However, there have been few significant breakthroughs in mobilizing students' learning efficiency as measured by information transmission, communication, and interaction in class. This research collects data through nonparticipatory observation and in-depth interviews and analyzes available data on classroom interaction needs of these students. We found that diversified explanations, recordable interactive contents, and interaction between teachers and students could improve the learning effects in the classroom. We also propose a tracking-processing method based on gesture recognition and representation and present a design for a processing system based on AT89C52 microcontroller and Kinect. In this way, sign language can be translated into text and all students can receive the information and participate in the interaction, which greatly improves students' autonomy and enthusiasm of learning. This design enables deaf students to fully use classroom learning resources, reduces learning time costs, and improves learning efficiency. It can assist teachers in teaching and tutoring students to enhance their experience.

**Keywords:** hearing-impaired college student; multimedia-equipped classroom; qualitative analysis; gesture recognition and representation; tracking processing

## 1. Introduction

China's Ministry of Education has issued guidelines for integrated education for the disabled ([gov.cn/zhengce/content/2017-02/23/content\\_5170264.htm](http://gov.cn/zhengce/content/2017-02/23/content_5170264.htm), accessed on 20 June 2022), drawing on the basic ideas of Vygotsky, a special educator [1]. Subsequently, Chinese institutions of higher education for the disabled are exploring the possibility and methods of integrated education to different degrees and in different ways [2,3]. One of the questions is whether and how to establish learning environments that can improve learning resources for the disabled. In general, colleges and universities have been supported to increase total enrollment to develop higher education for the disabled at a steady pace. Hearing-impaired students account for a certain proportion of college students. Such students often attend classes in multimedia-equipped classrooms, and most of their teachers have normal hearing. The shortage of both teachers and accessible communication methods is one of the major limitations on the learning effects of the hearing-impaired college students. In addition,

these classrooms are important but sometimes inaccessible places to study, raise questions, and have discussions. Although multimedia facilities can help students understand and communicate better with teachers and improve learning efficiency, traditional classrooms and facilities are neither distinctively ‘personalized’ nor well-targeted for them [4–6]. As current classrooms and facilities maintain no significant difference from those of normal-hearing students, the existing classrooms and facilities cannot match their internal learning needs. As a result, they have been trapped in a predicament with low efficiency where they have trouble in understanding what they have learned and become frustrated in learning.

Sign language is the dominant language for hearing-impaired college students. The development of sign-language recognition technology has a direct impact on their learning efficiency. Sign-language recognition technology mainly includes sensor-based recognition technology and vision-based recognition technology. Sensor-based sign-language recognition technologies are typically data gloves, which collect data by an optical fiber sensor on the dorsum of the fingers. The change of the fingers causes the rotation of the optical fiber, with alternating lights generated. The analog signal transmits the information of the finger change to the processor [7–9]. The vision-based sign-language gesture-recognition method consists of acquiring digital video-stream information for sign-language gestures using a camera, and then processing the information in a video frame by using a computer digital image-processing technology to acquire characteristics and complete gesture recognition. For example, Chang [10] proposed a sign-language recognition method based on curvature scale space (CSS) and hidden Markov model (HMM), and Geetha [11] proposed a new method to recognize gestures with static gesture language letters. These technologies are of great help for single-sign language recognition. However, there are still problems in the application in the classroom, and advanced recognition technology has not been introduced into the classroom.

To sum up, there have been many studies to help these students solve communication problems. However, these studies mainly focus on the intelligence of technology, which is not combined with the classroom situation and real needs of students. Therefore, the learning effect cannot be well-improved. Based on nonparticipatory observation and in-depth discussions, the paper adopts qualitative analysis to find the problems among them in the classroom interaction, especially in communication. An interactive communication assistance system is established through the teachers’ master terminal, students’ deputy terminal, depth camera, and master controller, and sign language can be translated into text using a PAJ7620 sensor, Atmega328P, AT89C52 microcontroller, and Kinect. In this way, teachers and students in the classroom can participate in two-way interaction, which would greatly improve students’ autonomy and enthusiasm of learning.

## 2. Related Works

Traditional dynamic gesture-recognition methods mainly include peripherals with acquisition functions (e.g., keyboard, mouse, etc.), wearable data gloves, and ordinary RGB depth-acquisition cameras. However, these kinds of hardware lack human factors, and it is difficult to realize harmonious and natural human–computer interaction.

Gesture recognition based on body-mounted sensor equipment [9] can obtain characteristic parameters of hand posture information such as finger bending through corresponding sensors and special material magnets, and then transfer them into digital data for transmission. Commonly used sign-language recognition methods include finger recognition, word-level sign-language recognition, and continuous sign-language gesture recognition. The recognition of finger language is static, relying on advanced machine learning and neural network algorithms, including New ID, HVC and RIEVL [12]. Neural network algorithms mainly include the radial basis function (RBF) [13] and minimax fuzzy neural network algorithm [14]. Gestures represent a dynamic time series. VPL data gloves are used as input devices, and the neural network is used as a gesture classifier to recognize gestures based on hand shape, hand movement direction, hand trajectory, hand speed, and other features [15,16]. Continuous sign-language gesture recognition is also the recognition

of multiple dynamic signed word gestures, so this kind of development is also through data gloves and a neural network model to identify features. Gesture-recognition technologies based on wearable sensors have a wide range of applications. The primary advantage of such systems is that they are less affected by the external environment and can accurately obtain action information. The disadvantage is that multiple sensors must cooperate to obtain hand data, and such wearable devices are complex and affect the user's movements, so the cost of data gloves is high.

Dynamic gesture recognition based on Kinect depth information is a noncontact interactive gesture-recognition method. The advantages of this kind of noncontact interactive gesture recognition are as follows: (1) Noncontact gesture recognition can avoid the health risks of contacting devices and the interference of irrelevant people to users, bringing a better interactive experience to users and making human-machine interaction more natural [17–22]. (2) Gesture data can be obtained through 3D point-cloud depth-information filtering with good robustness and a high recognition rate [23–29].

Based on the precise algorithm of Kinect, this paper captures the gestures of the hearing-impaired college students in class, to improve the efficiency of classroom interaction.

### 3. Materials and Methods

#### 3.1. Nonparticipatory Classroom Observations

To ensure the integrity of the multimedia-equipped classroom scenario and obtain a real situation of the hearing-impaired students in the classroom, this study adopts the method of nonparticipatory observation. The observation is to understand how they behave themselves and what problems they may encounter in the classroom. The observation subjects are 25 hearing-impaired sophomores of one class and one teacher of a 'Design Thinking' course. The observation was conducted with a lectern, a blackboard, a projector, and a computer. We observed the behaviors of teachers and students and recorded the data, including:

1. When did the interaction between students and teachers occur in class?
2. What was the context when the interaction occurred?
3. What was the behavior of the observed subjects during the interaction?
4. How many times did that behavior occur during the interaction?
5. How effective the communication was between student and teachers?
6. Was there someone else involved except the hearing-impaired college students?

Teachers for hearing disabled students include teachers with normal hearing and hearing-impaired teachers, among which those with normal hearing account for a majority among the faculties. Therefore, with the aim of understanding the characteristics of different types of teachers, this study conducted four classroom observations: two for teachers with normal hearing and two for hearing-impaired teachers.

Hearing-impaired teachers and teachers with normal hearing are different in lecturing and interactions within the class. These teachers can communicate more smoothly with students through sign language. They do not rely on excessive assistance of mouth shape and text. The efficiency of learning is higher, and students listened to teachers more attentively and answered questions more actively and conveniently. However, in the classroom of teachers with normal hearing, due to their lack of proficiency in sign language, teachers have less communication with students in sign language. Instead, they conduct the teaching through text and pictures. PowerPoint slides, words on the blackboard, and voice translation software (e.g., WeChat or specialized translation Apps) are used to convey information, thus failing to satisfy communication needs., and the focus and involvement of students become lower.

Meanwhile, no matter which classroom it was, the spatial form of a classroom tends to be a seedling nursery pattern or rectangular pattern. Some students find it hard to see gestures and interactions of teachers and other students because of the blocked sight,

e.g., body occlusion. As a result, they can neither understand the interaction clearly nor participate, which greatly reduces the classroom effectiveness.

### 3.2. Interviews and Qualitative Analysis

According to the problems of classroom observation, we go further to explore the problems of classroom interaction and the real needs of the hearing-impaired college students through semistructured interviews.

### 3.3. Participants

A total of 12 hearing-impaired college students, including 8 females and 4 males, from freshman to senior, were recruited for this interview. All participants were interviewed by writing. Since the interviewees did not agree to take photos, we did not take any photos during the interview.

### 3.4. Procedure

The questions in this interview were divided into four parts. The first part includes some basic information of deaf college students and teachers. The second part involves the discussions about the classroom interaction and communication of deaf college students in the multimedia-equipped classroom, with its purpose to explore in depth what the classroom interaction and communication problems are about, how the problems come into being, and what the current solutions are. The third part is about supplementary questions, aiming to understand whether the communication problems they have in class will affect their learning and processing after class. The last part provides some open questions to supplement the aspects ignored during the interview, to have a more comprehensive understanding of the problems of deaf college students in classroom communication.

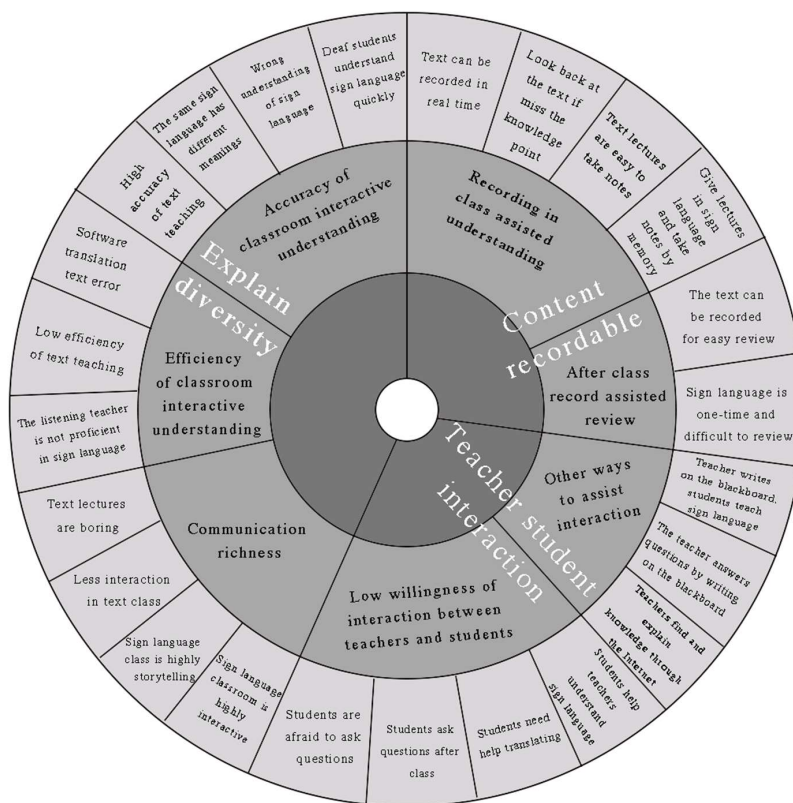
After transcribing, the research applies qualitative-analysis research methods with the help of NVivo11 qualitative-analysis software, analyzes the transcribed interview text word by word and completes the three-level coding, thus building a basic framework for analyzing the in-depth needs of deaf college students in classroom communication.

In the qualitative research, coding is the process of redefining the contents of data, and classifying, summarizing, and interpreting the data of each part at the same time. Through the method of extracting information vocabulary, it is the key link between collecting data and forming the theory of interpreting data [30]. Firstly, the sentences related to the topic are screened and extracted, given their concepts, and encoded at the first level in NVivo11 software. These concepts are classified and summarized to form different genera and establish nodes. At this stage, 37 generic categories were established. For example, some students said: 'I could not fully understand the questions or the teacher's words during the interaction when the teacher was using hand gestures. I sometimes think that I understand, but I am wrong when the teacher writes on the blackboard. I realized I had misunderstood'. However, some students said: 'I think some students do not understand the teacher's question because of different understanding ability. I think there are many mistakes in understanding writing words. It is better to use gesture language because writing words will be very troublesome.' Such content is classified as the individual ability to understand. Therefore, we classify these two parts of the interview content as having different understanding accuracy. Secondly, the relevance among 37 categories in the primary code is mined through the secondary code, and seven tree nodes are established in NVivo11; that is, seven main categories are established, which are the accuracy of understanding the lecture, the efficiency of understanding, the richness of interactive communication, in class records, after class records, the willingness of teachers and students to interact, and the assistance of interaction. Finally, through the induction and summary of the seven main categories three core categories are created, namely three core nodes: the diversity of explanation, the recordable contents, and teacher-student interaction. The three-level coding process is shown in Table 1.

**Table 1.** The interview coding.

Selective Coding	Axial Coding	Reference Points	%
Variety of interactive lectures	Different ways of interaction in class have different understanding accuracy	84	43.3
	Different ways of interaction in class have different understanding efficiency		
	There are differences in communication richness		
Class interactions can be recorded	You need to take notes in class to help you understand	72	37.1
	You need to take notes after class to help you review		
Teacher–student interaction	The willingness of teacher–student interaction is low	38	19.6
	Interaction needs to be assisted in other ways		

Figure 1 shows a hierarchy diagram generated by NVivo11 to clarify the hierarchical status of each node in the coding system. The frequency and percentage of each core node in the subjects’ views are shown in Table 1: among the three influencing factors, the subjects believe that the most relevant factor to classroom interaction is the diversity of lectures, with 84 reference points, accounting for 43.3%; then, the contents can be recorded, with 72 reference points, accounting for 37.1%; the last is teacher–student interaction, with 38 reference points, accounting for 19.6%.



**Figure 1.** Hierarchical content of each node in the coding system.

## 4. Results

### 4.1. Findings

According to the coding results, in the scenario of multimedia-equipped classroom, the key points to solve the classroom interactive communication can mainly be divided into four points:

1. In terms of teacher–student interaction efficiency, when college students attend classes in the classroom, different class methods initiate different class efficiency. In the non-signed classroom, the students' learning efficiency is not high because they are not familiar with words or the presentation of words is too slow, such as typing on the spot, writing on the blackboard, etc. The teaching method based on sign language is easier to understand and obtain access into, which makes the classroom efficiency higher. Therefore, sign language should be the major measure to improve the classroom interaction efficiency.
2. As for the recording of the classroom interaction contents, sign language cannot achieve a good recording function because of its transience. Therefore, they can understand and forget their knowledge points easily as well, but written words can be recorded and saved in real time, which solves this problem. At the same time, the recording function of the text can help them check repeatedly and understand the knowledge points better, hence, it enables them to keep up with the tempo of the teacher's lecture and facilitate the review after class. Therefore, the need for written words plays an essential role in the classroom.
3. When it comes to the accuracy of interactive understanding, there is no unified standard for sign language. There are different sign languages used by normal hearing teachers and the hearing-impaired college students, which leads to inaccurate transmission of sign language or different meanings of a gesture, and gives rise to mutual incomprehension or even misunderstanding between teachers and students. Compared with sign language, the accuracy of text is higher in conveying information. In addition, some mentioned that through textual communication, they can improve their communication level to communicate with hearing teachers more efficiently in the future and improve their adaptability into the society. Therefore, during the process of design, much emphasis will be laid on how to combine sign language with text and how to convert sign language accurately into text.
4. In terms of classroom interaction efficiency, teacher–student communication directly affects the efficiency in the classroom. Due to the communication problems between the hearing-impaired college students and normal-hearing teachers, when asking or answering questions, students dare not to ask or answer for fear of misunderstanding, incomprehension, or embarrassment, thus lowering the interaction willingness. During the interaction, they often choose to use sign language repeatedly or slow down the speed of sign language to confirm whether they have understood accurately, which will also reduce the efficiency of the class. Furthermore, most interactions only happen between teachers and students, and other students remain passively receiving. They had no idea about what the interactive students had answered, which proves the lack of attraction in the classroom. Therefore, the classroom auxiliary system should have the function of improving students' interactive participation. The construction of auxiliary system and terminal equipment is also an important part of the design.

### 4.2. Design Transformation

In the phase of user investigation, we learned that although sign language is the first language for the hearing-impaired college students, lectures given by sign language are temporary for them and cannot be well-recorded. This undoubtedly makes it more challenging for deaf college students to memorize pertinent information. Hence, the contents of the lecture should be recorded in the form of text. Furthermore, the Q&As between teachers and students become the exploration of key knowledge points, which also shows the knowledge points that deaf college students have a poor understanding of

or that teachers have not explained clearly. The recording of the interactive part can help deaf college students to focus on recording their own problems or problems with other classmates, to find out the omissions and fill in the gaps in time.

When the hearing-impaired college students are in class, due to different ways of communication with non-hearing-impaired teachers, they have more difficulty in understanding—or even misunderstand—some of the knowledge points. This requires deaf college students to ask and answer questions actively and have a deeper understanding of knowledge points in the process of interaction. It also requires the designation of asking or answering questions to help deaf college students to interact in time. In addition, considering that some deaf college students might be embarrassed by communication problems, setting up an after-school questioning function can be a solution. Furthermore, the interaction in class usually centers on difficult points of lectures. If deaf college students did not sort out complete Q&As in class, they need to supplement them after class. This requires setting up an entrance for supplementary contents after class. Table 2 shows the specific transformation from needs to functional design of deaf college students.

**Table 2.** Specific transformation from needs to functional design.

Requirements	Transformation	Function Design
Variety of interactive lectures	Interactive content presentation	Live captioning Multiple definition of words Sign language is translated in literal form
Class interactions can be recorded	Interactive content recording	Classroom implementation records Interactive content is saved in the cloud Text interaction facilitates notetaking
Teacher–student interaction	Interactions	Remind when answering questions Sight and hearing draw attention to each other Choose answers

## 5. Assistant System Design

### 5.1. Overall Design

In recent years, researchers have spent much time on recognizing sign language through machines. Raheja [31] used a camera to capture real-time video at 30 frames per second and then analyzed dynamic gestures frame by frame. By extracting skin regions as an auxiliary evaluation process, the system converts each frame into an HSV color space and filters out skin-color pixel regions. Kishore collected data using four cameras, each angled from the center to the  $-20^\circ$ ,  $-10^\circ$ ,  $10^\circ$ , and  $20^\circ$  positions [32]. Angle changes can better capture the dynamic effect. However, in the actual scene, gestures will be affected by the surrounding environment, especially the background, such as the background and skin color close to the existence of dynamic objects. Some researchers make sensory gloves using flexible acceleration sensors and collect palm-movement data through three-axis acceleration sensors. Thang [33] used a five-dimensional tracker that can track 11 finger features, including the flexion of the thumb, index finger, middle finger, ring finger, and little finger. The use of these flexible sensors is compassionate, and the recognition of motion is quick and accurate. However, because these sensor devices are installed in hand, they do not consider the impact of comfort on hand movement and the deviation brought by it. Therefore, it is necessary to use a device that can efficiently complete the background segmentation and naturally capture the hand joints' position to study gestures. At present, the technology in common use is to recognize sign language by a depth camera, e.g., Kinect and Leap Motion. To recognize sign language by depth camera is to emit infrared light through the infrared light source and receive the infrared light reflected by the environment and the user.

By detecting the phase differences, we carry out 3D imaging of the user, and then recognize human contour and joint positioning. Users can define the basic actions as sign

language words through the actions read by the depth camera. The depth camera is fixed above the display screen, and a user makes the required action or sign language in front of the display screen.

At present, when using in the classroom of teaching college students with hearing impairment, the depth camera is fixed; thus, the range for recognition is limited, which affects the teaching quality. In addition, the current teaching system does not have the function of sending swift messages through simple actions, which is vital; thus, the existing gesture-recognition methods are not suitable in class. Based on these above needs, we propose the following design for a tracking-processing system for gesture recognition and representation, including a main terminal, subterminal, gesture sensor, PTZ, and main controller.

1. The main terminal is a teacher's computer, as an output unit, including a sign-language recognition module, display module, and voice module. The main terminal is used to track the joint position, display the text converted by the action of college students with hearing impairment, convert the text or action into voice information, and recognize sign language according to the action and the video picture obtained by the depth camera.
2. The subterminal is mainly used to transmit the information ID number of the student's name and the angle of the seat and the podium, which is used for the adjustment of the camera direction and the student information displayed on the teacher's computer. The control part of the subterminal adopts the AT89C52 single-chip microcomputer as the processor, and the ID unit is allocated on the AT89C52 single-chip computer and is connected with the AT89C52 through the wireless serial port LC2S. In addition, the subterminal is equipped with a lighting module, used to indicate whether the student is actively participating. The indicator of the lighting module can convey common information, to intuitively attract the attention of teachers and students.
3. The gesture sensor PAJ7620, fixed on the student's desk, is connected with the deputy terminal to obtain and identify the movement information. Depending on the student's action, the PAJ7620 can be used to identify the corresponding message and determine whether to activate the signal of the sensor, trigger quick messages, or activate the gesture-recognition system by the basic action that the deputy terminal gesture sensor made. The gesture sensor can improve the communication efficiency between these students and teachers by combining the sign-language recognition.
4. The depth camera is connected to the master terminal and its lens faces the deputy terminal. The depth camera here is Kinect, with 1080P HD video recording and the ability to recognize human body movements. The angle between the depth camera and the deputy terminal should be no more than 10 degrees, to ensure that the master terminal can acquire the gestures captured by the depth camera, then convert the gestures to sign language, or trigger a swift message according to the gestures. The student could turn the depth camera on or off by performing a specific action to the gesture sensor.
5. The tripod head is a rotatable structure on which the depth camera is installed. The tripod head is designed with single axis camera and is controlled by the master controller.
6. The master controller is used to connect the master terminal and the deputy terminal as well as the tripod head. The master controller selects Arduino UNO development board based on Atmega328P as the core processor.

The overall system design is shown in Figure 2. The prototype design is shown in Figure 3.



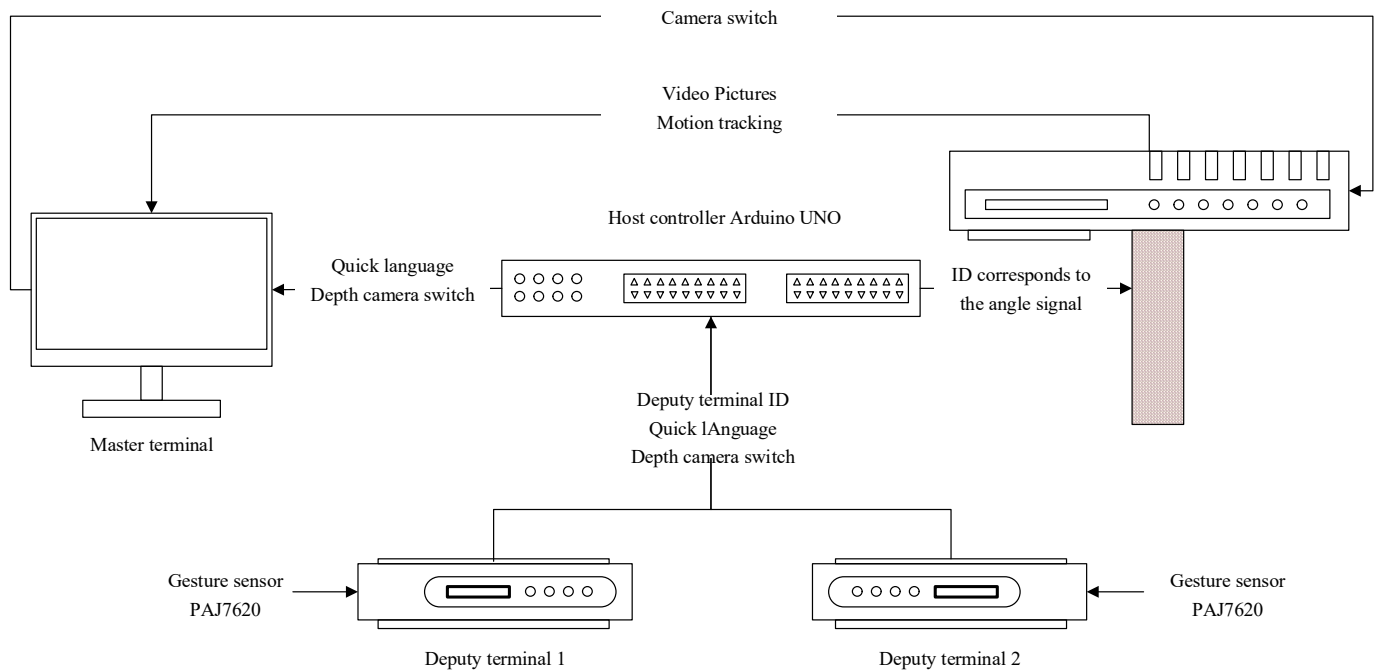


Figure 2. The system.

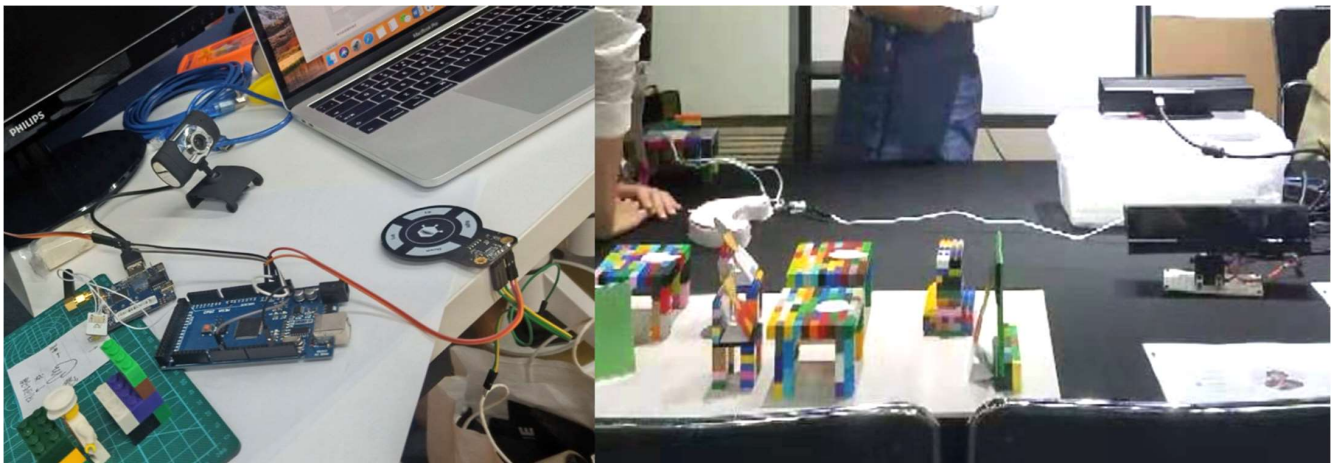


Figure 3. The prototype.

### 5.2. Operation Principle

The system reserves the way the deputy terminal numbers ID information in advance and stores it in the master terminal. When the gesture sensor of the deputy terminal detects the action information, the depth camera will be triggered to work, and the master terminal will display the picture obtained by the depth camera. The master terminal judges whether its location is in the center of the depth camera according to the angle data corresponding to the ID number information of the deputy terminal saved in advance. If not, the signal will be sent to the tripod head through the master controller of the master terminal, then the tripod head drives the depth camera to rotate, and then the position of the deputy terminal can be locked no matter whether the depth camera is aligned with the master terminal or the deputy terminal. At this point, the depth camera can identify the gesture language of students in any designated position in the classroom. Therefore, when these students encounter problems, the system can respond timely, effectively, and quickly without students leaving their seats and stepping in front of the camera when signing, thus improving students' learning efficiency.

After the recognition of the position, the depth camera identifies students' gestures and transmits them to the master controller. During the processing, the master controller converts gestures into text and transmits them to the master terminal, and the display screen of the master terminal will show the text after conversion, while other students will be able to see the interactive contents and participate in the classroom interaction, thus improving the understanding and efficiency of the interaction.

### 6. Gesture Recognition and Representation

In this design, gesture recognition mainly relies on Kinect and the embedded system based on Arduino with the Atmega32P as the core processor. The running process of the identification system is shown in Figure 4.

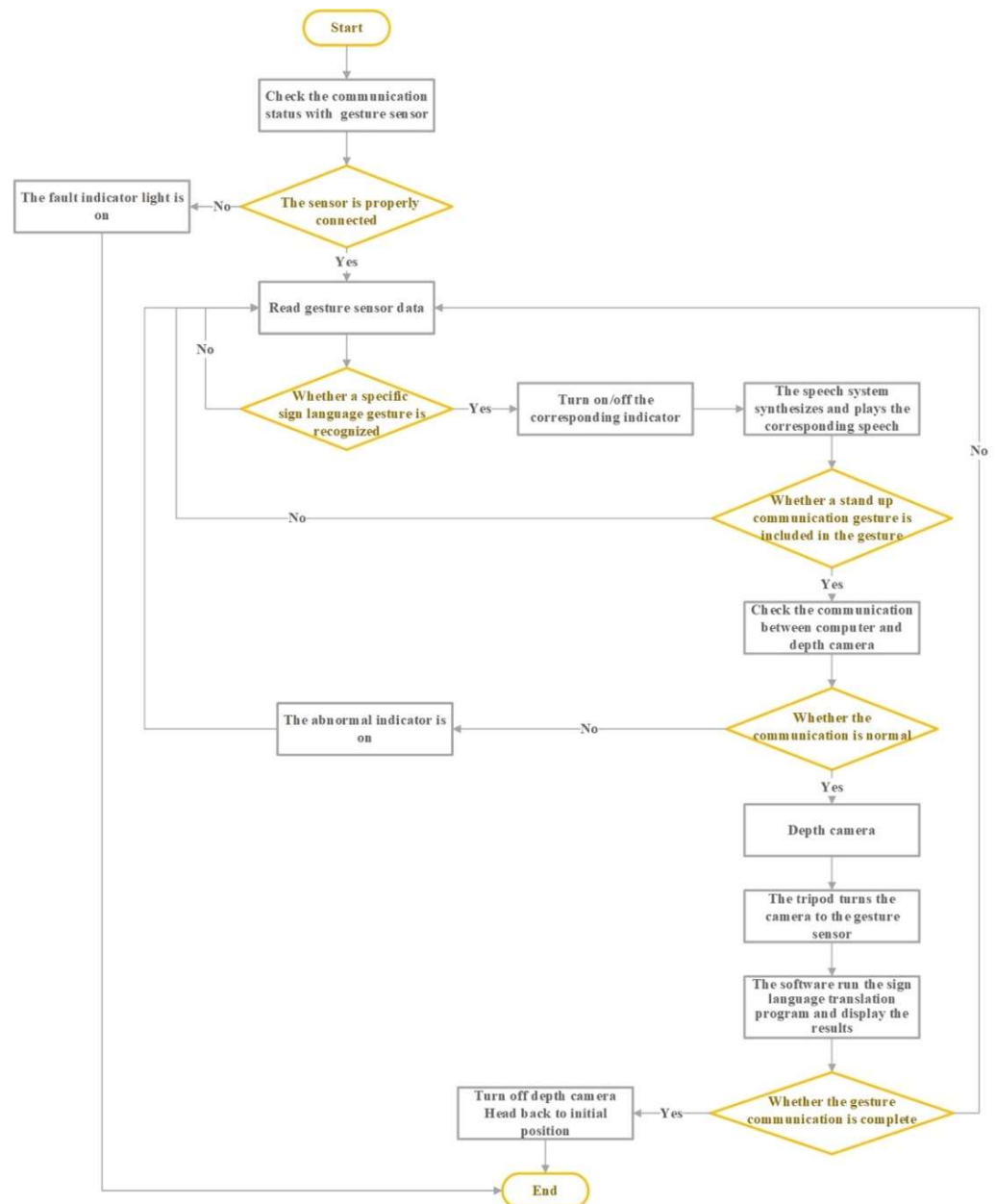


Figure 4. The flow of the sign-language recognition system.

### 6.1. Basic Gesture Model

Kinect uses a technique called Light Coding. Its light-encoded infrared transmitter emits a ‘stereo code’ with three-dimensional depth. Firstly, the space was calibrated: within the effective range (0.5 m to 4.5 m) that Kinect could recognize, a reference plane was taken every 10 cm, a total of 40 reference planes were taken, and then the speckle patterns on the 40 reference planes were saved. When the object coordinates in space need to be measured, the speckle pattern of the object to be measured is captured by an infrared camera and compared with speckle images of 40 reference planes to obtain the image. The first level of gesture size is input by sequential model, then through the functional model and the loading mode function. The linear stack of multiple network layers is built to construct the gesture model, and the model is trained in depth.

### 6.2. Gesture Recognition and Tracking

Gesture recognition and tracking is mainly divided into gesture feature extraction and feature judgment. The depth image obtained by Kinect is extracted with coarse granularity to obtain preliminary hand-gesture information. The depth image  $f(x, y, z)$  obtained by Kinect is mapped to the color image  $f(x, y)$ ; then, the threshold method is used to segment gestures, as shown in expression (1):

$$g(x, y, z) = \begin{cases} 1 & f(x, y, z) \in (T, T + s) \\ 0 & f(x, y, z) \notin (T, T + s) \end{cases} \quad (1)$$

where  $S$  is a distance constant and  $g(x, y, z)$  is the segmented gesture region. After segmentation of the gestures, feature extraction is carried out on the contour of the gesture. The classical extraction methods of the gesture contour include the Roberts operator, Sobel operator, Prewitt operator, Laplace operator, Log operator, and Canny operator [34–36]. In this design, the Canny operator is used to extract the gesture contour, which can extract a more complete gesture contour. Here, Canny performs denoising and looks for image brightness and image edge. It uses the mask function and operation to obtain the maximum value of gesture-edge points and gesture-contour direction, and then obtains the brightness image of the gesture edge and sorts contour values through the shuffle function to complete the gesture extraction. Then, a convolutional neural network model is used to train the extracted features. The captured gesture and recognition model are matched based on judgement of the gestures, and output those with a matching degree higher than 60%.

## 7. Discussion and Conclusions

This study is not to design a product that can identify straightforwardly and translate semantics, but rather to improve classroom efficiency and mobilize the participation of all classroom participants. We designed a system to cover all the relevant stakeholders in the classroom. This design has been patented as a new utility model by the state intellectual property office of China (ID: ZL 2021 2 0026852.3.).

Based on the current education situation of deaf college students, this study finds that the hearing-impaired college students, as a special educational group, still lack educational resources through data research. Further user research was carried out to explore the communication problems of deaf college students in class and the reasons behind them. In the user research phase, through nonparticipatory observation and user interviews, we better understood the needs of deaf college students in class, including lecture diversity, recordable content, and teacher–student interaction. After understanding the classroom communicative needs of deaf college students, we carried out demand transformation, function design, and product design. We produced a system to assist the communication of deaf college students in class.

We found that these students are eager to use the classroom and improve learning efficiency effectively. Classroom efficiency is the result of collective action. In raising and answering questions, the teacher and the students in the whole class should be participants in order to arouse the group’s enthusiasm. However, the existing smart classroom can only

realize information transmission through the intelligent part, and information transmission is limited to the interaction between teachers and individual students. Such neglect can give other students the illusion of not participating in the interaction process. Through the preliminary investigation, it is also known that teachers are mainly teachers with normal hearing, and there are specific communication problems between deaf college students and ordinary-hearing teachers. In the classroom context, multimedia facilities still cannot meet the communicative needs of deaf college students in the classroom.

Based on Kinect and embedded systematic design, the system for assisting the communication in the class produced through research is designed according to the communicative problems of these students in the classroom. The product has yet to be used in a real classroom. Hence, its effectiveness has yet to be proven. Due to the epidemic, we could not enter relevant schools, so we did not test the system in the classroom. We showed this system to three students online (i.e., the study could not be carried out offline due to the COVID-19 epidemic). We asked the three students to put forward suggestions after the simulation class. Three students believe this system is highly effective and will attract all students to participate in classroom interaction. However, the speed of gesture recognition is not very efficient.

Many of the requirements of this study were concluded through qualitative analysis, without a corresponding questionnaire survey and quantitative-data analysis. In the future, we will use questionnaires and quantitative methods to analyze students' understanding of sign-language translation and their needs in group interaction to improve the existing system.

In addition, the amount of isolated and continuous-word sign-language recognition is insignificant in terms of the amount of system recognition. Realizing intelligent sign-language recognition requires more artificial-intelligence algorithms and a lot of data training, which is also the direction of our future work. As a special educational group, college students with hearing impairments have always been concerned. From the user's point of view, this study designs a product for the communicative problems of deaf college students in the classroom. To some extent, this study can help deaf college students improve the efficiency and quality of classroom communication to obtain better education outcomes.

**Author Contributions:** Conceptualization, Y.Z.; methodology, J.Z.; validation, G.C.; formal analysis, Z.Z.; data curation, J.J.; writing—original draft preparation, Y.Z.; writing—review and editing, W.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** We would like to express our gratitude to the hearing-impaired students for their meaningful support of this work. We would like to thank the China–US Young Maker Competition organized by the Ministry of Education of People's Republic of China.

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

## References

1. Gindis, B. Vygotsky's vision: Reshaping the practice of special education for the 21st century. *Remedial Spec. Educ.* **1999**, *20*, 333–340. [[CrossRef](#)]
2. Zhang, Y.; Rosen, S.; Cheng, L.; Li, J. Inclusive higher education for students with disabilities in China: What do the university teachers think? *High. Educ. Stud.* **2018**, *8*, 104–115. [[CrossRef](#)]
3. Li, H.; Lin, J.; Wu, H.; Li, Z.; Han, M. "How do I survive exclusion?" Voices of students with disabilities at China's top universities. *Child. Youth Serv. Rev.* **2021**, *120*, 105738. [[CrossRef](#)]
4. Chadwick, D.; Wesson, C.; Fullwood, C. Internet Access by People with Intellectual Disabilities: Inequalities and Opportunities. *Futur. Internet* **2013**, *5*, 376–397. [[CrossRef](#)]
5. Chen, Y.-T. A study to explore the effects of self-regulated learning environment for hearing-impaired students. *J. Comput. Assist. Learn.* **2014**, *30*, 97–109. [[CrossRef](#)]
6. Xu, B. Using New Media in Teaching English Reading and Writing for Hearing Impaired Students—Taking Leshan Special Education School as an Example. *Theory Pract. Lang. Stud.* **2018**, *8*, 588–594. [[CrossRef](#)]

7. Deb, S.; Bhattacharya, P. Augmented Sign Language Modeling (ASLM) with interaction design on smartphone—An assistive learning and communication tool for inclusive classroom. *Procedia Comput. Sci.* **2018**, *125*, 492–500. [[CrossRef](#)]
8. Bragg, D.; Koller, O.; Bellard, M.; Berke, L.; Boudreault, P.; Braffort, A.; Caselli, N.; Huenerfauth, M.; Kacorri, H.; Verhoef, T.; et al. Sign language recognition, generation, and translation: An interdisciplinary perspective. In Proceedings of the 21st International Conference on Computers and Accessibility, New York, NY, USA, 28–30 October 2019; pp. 16–31.
9. Zimmerman, T.G.; Lanier, J.; Blanchard, C.; Bryson, S.; Harvill, Y. A hand gesture interface device. *ACM SIGCHI Bull.* **1986**, *18*, 189–192. [[CrossRef](#)]
10. Chang, C.-C.; Pengwu, C.-M. Gesture recognition approach for sign language using curvature scale space and hidden Markov model. In Proceedings of the 2004 IEEE International Conference on Multimedia and Expo (ICME), Taipei, Taiwan, 27–30 June 2004; Volume 2, pp. 1187–1190.
11. Geetha, M.; Menon, R.; Jayan, S.; James, R.; Janardhan, G.V.V. Gesture Recognition for American Sign Language with Polygon Approximation. In Proceedings of the IEEE International Conference on Technology for Education, Tamil Nadu, India, 14–16 July 2011; pp. 241–245.
12. Zhao, M.; Quek, F.; Wu, X. RIEVL: Recursive induction learning in hand gesture recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **1998**, *20*, 1174–1185. [[CrossRef](#)]
13. Avola, D.; Bernardi, M.; Cinque, L.; Foresti, G.L.; Massaroni, C. Exploiting Recurrent Neural Networks and Leap Motion Controller for the Recognition of Sign Language and Semaphoric Hand Gestures. *IEEE Trans. Multimed.* **2018**, *21*, 234–245. [[CrossRef](#)]
14. Rautaray, S.S.; Agrawal, A. Vision based hand gesture recognition for human computer interaction: A survey. *Artif. Intell. Rev.* **2015**, *43*, 1–54. [[CrossRef](#)]
15. Fels, S.; Hinton, G. Glove-Talk: A neural network interface between a data-glove and a speech synthesizer. *IEEE Trans. Neural Netw.* **1993**, *4*, 2–8. [[CrossRef](#)] [[PubMed](#)]
16. Glauser, O.; Wu, S.; Panozzo, D.; Hilliges, O.; Sorkine-Hornung, O. Interactive hand pose estimation using a stretch-sensing soft glove. *ACM Trans. Graph.* **2019**, *38*, 1–15. [[CrossRef](#)]
17. Thieme, A.; Belgrave, D.; Doherty, G. Machine learning in mental health: A systematic review of the HCI literature to support the development of effective and implementable ML systems. *ACM Trans. Computer-Hum. Interact.* **2020**, *27*, 1–53. [[CrossRef](#)]
18. Desmet, P.; Xue, H.; Xin, X.; Liu, W. Emotion deep dive for designers: Seven propositions that operationalize emotions in design innovation. In Proceedings of the International Conference on Applied Human Factors and Ergonomics. AHFE International, New York, NY, USA, 24–28 July 2022.
19. Zhu, Y.; Jing, Y.; Jiang, M.; Zhang, Z.; Wang, D.; Liu, W. A Experimental Study of the Cognitive Load of In-vehicle Multiscreen Connected HUD. In *International Conference on Human-Computer Interaction*; Springer: Cham, Switzerland, 2021; pp. 268–285.
20. Liu, W. Designing Generation Y Interactions: The Case of YPhone. *Virtual Real. Intell. Hardw.* **2022**, *4*, 132–152. [[CrossRef](#)]
21. Desmet, P.; Overbeeke, K.; Tax, S. Designing products with added emotional value: Development and application of an approach for research through design. *Des. J.* **2001**, *4*, 32–47. [[CrossRef](#)]
22. Gray, C.M.; Chivukula, S.S.; Lee, A. What Kind of Work Do “Asshole Designers” Create? Describing Properties of Ethical Concern on Reddit. In Proceedings of the ACM Designing Interactive Systems Conference, Eindhoven, The Netherlands, 6–10 July 2020; pp. 61–73.
23. Xin, X.; Wang, Y.; Xiang, G.; Yang, W.; Liu, W. Effectiveness of Multimodal Display in Navigation Situation. In Proceedings of the Ninth International Symposium of Chinese CHI, Online, 16–17 October 2021; pp. 50–62.
24. Kang, H.; Hur, N.; Lee, S.; Yoshikawa, H. Horizontal parallax distortion in toed-in camera with wide-angle lens for mobile device. *Opt. Commun.* **2008**, *281*, 1430–1437. [[CrossRef](#)]
25. Li, W.; Jin, C.-B.; Liu, M.; Kim, H.; Cui, X. Local similarity refinement of shape-preserved warping for parallax-tolerant image stitching. *IET Image Process.* **2018**, *12*, 661–668. [[CrossRef](#)]
26. Chen, Y.; Zhang, B.; Zhou, J.; Wang, K. Real-time 3D unstructured environment reconstruction utilizing VR and Kinect-based immersive teleoperation for agricultural field robots. *Comput. Electron. Agric.* **2020**, *175*, 105579. [[CrossRef](#)]
27. Roy, P.P.; Kumar, P.; Kim, B.-G. An Efficient Sign Language Recognition (SLR) System Using Camshift Tracker and Hidden Markov Model (HMM). *SN Comput. Sci.* **2021**, *2*, 1–15. [[CrossRef](#)]
28. Zhu, Y.; Tang, G.; Liu, W.; Qi, R. How Post 90’s Gesture Interact with Automobile SkyLight. *Int. J. Hum.-Comput. Interact.* **2022**, *38*, 395–405. [[CrossRef](#)]
29. Sagayam, K.M.; Hemanth, D.J. ABC algorithm based optimization of 1-D hidden Markov model for hand gesture recognition applications. *Comput. Ind.* **2018**, *99*, 313–323. [[CrossRef](#)]
30. Charmaz, K.; Belgrave, L.L. Thinking about Data with Grounded Theory. *Qual. Inq.* **2019**, *25*, 743–753. [[CrossRef](#)]
31. Raheja, J.L.; Mishra, A.; Chaudhary, A. Indian sign language recognition using SVM. *Pattern Recognit. Image Anal.* **2016**, *26*, 434–441. [[CrossRef](#)]
32. Kishore, P.V.V.; Prasad, M.V.; Prasad, C.R.; Rahul, R. 4-Camera model for sign language recognition using elliptical fourier descriptors and ANN. In Proceedings of the International Conference on Signal Processing and Communication Engineering Systems, Vijayawada, India, 2–3 January 2015; pp. 34–38.
33. Thang, P.Q.; Dung, N.D.; Thuy, N.T. A comparison of SimpSVM and RVM for sign language recognition. In Proceedings of the International Conference on Machine Learning and Soft Computing, Ho Chi Minh City, Vietnam, 13–16 January 2017; pp. 98–104.

34. Deng, C.X.; Wang, G.B.; Yang, X.R. Image edge detection algorithm based on improved canny operator. In Proceedings of the International Conference on Wavelet Analysis and Pattern Recognition, Tianjin, China, 14–17 July 2013; pp. 168–172.
35. Zhao, H.; Qin, G.; Wang, X. Improvement of canny algorithm based on pavement edge detection. In Proceedings of the 3rd International Congress on Image and Signal Processing, Yantai, China, 16–18 October 2010; Volume 2, pp. 964–967.
36. Saxena, S.; Singh, Y.; Agarwal, B.; Poonia, R.C. Comparative analysis between different edge detection techniques on mammogram images using PSNR and MSE. *J. Inf. Optim. Sci.* **2022**, *43*, 347–356. [[CrossRef](#)]