

Research Article

Digital Media Design for Dynamic Gesture Interaction with Image Processing

Xiaoli Xiong  and Yongguang Hou

Education Science College, Weinan Normal University, Weinan 714099, Shanxi, China

Correspondence should be addressed to Xiaoli Xiong; hyg@wnu.edu.cn

Received 18 February 2022; Revised 15 March 2022; Accepted 24 March 2022; Published 27 April 2022

Academic Editor: Wei Liu

Copyright © 2022 Xiaoli Xiong and Yongguang Hou. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the rapid development of human-computer interaction technology and research in the fields of ergonomics and user experience, people have increasingly demanded robot availability and ease of use. The rapid development of modern computer computing and digital media technology has made people's contact and interaction with computers more and more frequent, and they have appeared more in people's daily lives. Therefore, the convenience and freedom of computer communication have been proposed. Based on the above background, the research content of this paper is the design of dynamic gesture interactive digital media based on image processing. This paper proposes a design scheme of an interactive touch system based on image processing and a dynamic gesture recognition method and uses the time difference method to experimentally simulate the system proposed in this paper. Image processing is the technique of analyzing images with a computer to achieve the desired results. Image processing technology generally includes three parts: image compression, enhancement and restoration, and matching, description, and recognition. The experimental results show that, for the same motion trajectory, the accuracy depends on the complexity of the gesture degree. And the system's recognition accuracy rate is always maintained above 98%, confirming that the system has the performance requirements of high recognition accuracy, fast response speed, and stable work. Finally, the system was tested for performance and function, which verified that the system meets the real-time requirements. For a particular gesture operator, the accuracy rate has a great relationship with the standard degree of its operation. If the similarity with the model is high, you can achieve a high recognition rate.

1. Introduction

Human-computer interaction has increasingly become an important part of people's daily life. Especially in recent years, with the rapid development of computer technology, the research on novel human-computer interaction technology that conforms to the habits of interpersonal communication has become very active, and gratifying progress has also been made. The extreme improvement of technology has made people begin to process digital image information by using computers. In the early days of technology formation, the objects of digital image processing were mainly humans. Researchers spent most of their energy processing images so that the image quality could meet the needs of human vision.

Research on gesture recognition as a 3D interaction means for virtual environment interaction, considering that

people have a lot of empirical knowledge of gestures. If people can translate these skills from everyday experience and use them in human-computer interaction, then we can expect intuitive, easy-to-operate, and powerful human-computer interfaces. For example, when people wave their hands at the robot, the robot can know that people are calling it. The study of human language understanding, that is, the perception of human language, and the information fusion of human language and natural language is of great significance for improving the computer's human language understanding level and enhancing the practicability of human-machine interfaces.

With the continuous progress of human civilization, the vigorous development of scientific and technological production, and the further development of digital image processing technology, the era of digital media has quietly

arrived [1, 2]. And with the research of computer science and artificial intelligence technology and thinking science, the research of digital image processing has entered a brand new stage [3, 4]. Researchers hope to be able to use computers to accurately explain images and fully simulate the functions of the human visual system [5]. Although the digital image understanding technology has achieved great results in the research of theoretical methods, it is a relatively difficult research field in itself, and there are many research difficulties [6, 7]. Moreover, because humans have little understanding of the human head's visual understanding process, the current digital image processing technology is still an important research area for further exploration [8, 9]. In addition, from the beginning of computer design, human-computer interaction technology has attracted people's attention [10, 11]. From the most primitive paper tapes, keyboards, mechanical mice, and optical mice to wireless keyboard and mouse devices used today, their used methods are also getting closer to the natural use habits of humans [12, 13]. But with the popularization of computers and the increase in the number of people, the traditional human-computer interaction interface is increasingly criticized by people [14, 15]. Its single operation mode, fixed input mode, and obvious human-machine isolation gradually make people expect to change this unfriendly working method [16, 17]. In such an environment, more people start to pay attention to the development and reform of human-computer interaction, and they expect that human-computer interaction is as natural, simple, and true as human communication with each other. Humans use their own habits to control computers and communicate with the computer [18, 19].

Lv et al. proposed a comprehensive automatic sign language recognition technology using multiple methods, including analysis of semantic grammar, vision-based tracking technology, feature extraction and parameter estimation, gesture recognition scheme and classification method, gesture segmentation, and automatic grammar processing; it is independent of the detection of gestures [20]. The recognition of gestures integrates other modules such as face detection and speech recognition. The Mitra and Acharya survey analyzed gesture recognition and its relationship to facial head posture. They analyzed some of the main methods that have been successful in gesture recognition, such as Hidden Markov Models, example (particle) filtering techniques, functional state machines, and neural network techniques. Ramu et al. proposed a method for tracking virtual objects, especially human hands, in video images. They focus on the use of color segmentation-based tracking algorithms and contour-based algorithms [21, 22]. Many subsequent researchers have adopted the above rules as assumptions in their gesture recognition methods. Pavlovic divides gesture recognition methods into two categories: gesture recognition based on 3D models and apparent gesture recognition technology. The former completely describes the motion parameters of gestures from the perspective of space and time, but due to the complexity of its own spatial properties, it is difficult to simplify the computational efficiency and algorithm description. The latter is based on the apparent method, which

increases the limitation of the application space and ensures that its own complexity will not be too large from the scope of application. It provides a simpler choice for gesture recognition [23].

Gatteschi et al. proposed a deviceless gesture recognition system GRfid based on COTS-RFID device output phase information. RFID phase information can capture the spatial characteristics of various gestures with low-cost commodity hardware. In GRfid, after collecting data through hardware, it processed the data through a series of functional modules such as data preprocessing, gesture detection, contour training, and gesture recognition and achieved good gesture recognition performance [24, 25]. Plouffe discussed the development of a natural gesture user interface for real-time gesture tracking and recognition based on depth data collected by Kinect sensors. First, based on the assumption that the user's hand is the object closest to the camera in the scene, the space of interest corresponding to the hand is segmented [26]. Plouffe proposed a new algorithm to increase the scan time to identify the first pixel on the contour of the hand in this space. Starting from this pixel, the directional search algorithm allows us to recognize the entire hand contour. Then, the K-curvature algorithm is used to determine the position of the fingertip on the contour, the candidate gesture is selected using the Dynamic Time Warping (DTW) method, and the gesture is identified by comparing the observed gesture with a series of prerecorded reference gestures [27]. Comparing the results with the latest methods, the results show that the proposed system is superior to most solutions in static recognition of symbol numbers and has similar performance in static and dynamic recognition of popular symbols and sign language alphabet [28, 29]. Hsu proposed a digital pen (inertial pen) based on inertial sensors and its related handwriting and gesture recognition algorithms based on Dynamic Time Warping (DTW). Hsu developed a template selection method (minimum-maximum template selection method) based on the smallest to the largest class for the DTW recognizer to obtain a superior classification for improving recognition. Experimental results have successfully verified the effectiveness of DTW-based recognition algorithms for online handwriting and gesture recognition using inertial pens [30, 31]. Despinoy proposed a new unsupervised algorithm that can automatically segment kinematic data in robot training sessions. Without relying on any prior information or model, the algorithm can detect critical points that define relevant spatiotemporal segments in motion data. Despinoy uses data sets recorded during practical expert training for advanced analysis and evaluation of its algorithms. After comparing the method proposed by Despinoy with the manual annotation of surgical gestures, it was found that the accuracy of the learning purpose was 97.4% [32, 33]. Hong has designed and implemented an acceleration gesture recognition method based on random projection (RP). He made statistics on 2,400 gesture trajectories, and the final experimental results showed that the accuracy of the algorithm for specific individuals reached 98.41%. For non-specific individuals, this ratio is 67%, and it can effectively recognize acceleration gestures [34–36].

The innovations of this paper are as follows: (1) Research and design a multitouch detection and positioning algorithm and a gesture tracking and recognition algorithm based on optical multitouch technology, digital image processing technology, and gesture tracking recognition technology. Thereby, multitouch positioning and gesture recognition of the whole system are realized. (2) Establish a human-computer interaction optical touch system that can obtain multicontact coordinate information stably, quickly, and accurately, and track and recognize user gestures in real time. The system uses the time difference method of motion monitoring and the algorithm based on the HSV color space model to realize the positioning of the gesture and obtains a more accurate gesture positioning image. (3) Combine the improved CamShift algorithm with Kalman filtering to achieve gesture tracking. While ensuring the real-time performance of calculation, the problem of occlusion of external objects in the process of gesture tracking is well solved, and experiments are carried out on the designed system. (4) The final simulation results verify the effectiveness and practicability of the system.

2. Dynamic Gesture Interaction Method Based on Image Processing

2.1. Interaction Design Based on Image Processing. In order to avoid the phenomenon of occlusion and strange points, it is calculated based on geometric principles that four optical sensors are required, which are distributed at four right angle points on the screen. In this way, when the occluder blocks the propagation of infrared light, the four around the image information received by the optical sensor determine the true position of the obstruction. The infrared emitter's emission angle must be greater than 90 degrees to cover the entire screen surface, eliminating possible gray areas, ensuring that the contacts can be detected at any position on the screen, and ensuring the accuracy of the platform. The detection principle of the four-eye sensor touch platform is shown in Figure 1.

As shown in Figure 1, the sensor is distributed at the four vertices of the touch screen. The infrared light emitted by the optical sensor is reflected back to the optical sensor by the reflective bars around the screen, forming a light band in the optical sensor. When there is a contact in front of the screen, the contact will block the infrared light emitted by the infrared transmitter. The four optical sensors at the right angle point could originally receive the infrared light. When there is a contact, it will not receive the infrared light. The infrared light is blocked by the contacts. At the same time, the signal waveform image collected by the processor in the platform will show a corresponding depression in the waveform, indicating that the infrared is blocked and a contact appears. However, when two contacts A and B appear in the middle of the screen, they block the light that should have been reflected back into the optical sensor, forming a shadow in the optical band of the optical sensor. The dotted line in the figure indicates the original light path. The sensor passes the collected information to the processor, and the system uses

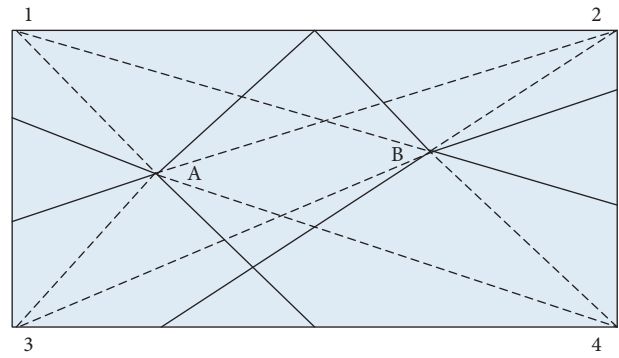


FIGURE 1: Detection principle of four-eye sensor touch platform.

the waveform received by the detection processor to set the corresponding detection contact algorithm.

2.2. Dynamic Gesture Interaction. No matter whether the gesture is static or dynamic, the sequence of recognition should firstly perform image acquisition, hand detection, and segmentation gesture analysis and then perform static or dynamic gesture recognition. The key technologies in gesture recognition are gesture segmentation, gesture analysis, and gesture recognition. Dynamic gesture recognition is based on the different shapes and movements of the hand at different times. The recognition process includes the motion trajectory of the gesture and the movement of the hand. This method can accurately determine the user's intention. Gesture recognition is a topic in computer science and language technology that aims to recognize human gestures through mathematical algorithms. Gestures can originate from any body movement or state but usually originate from the face or hands. Dynamic gesture recognition can be understood as a series of continuous actions, which is the main difference from static gesture recognition. Dynamic gesture recognition includes hand rotation, shape change, and motion trajectory. There are three main recognition methods: template-based methods, grammar-based methods, and statistics-based methods.

2.2.1. Template-Based Method. The template matching algorithm is similar to the static matching method, comparing and discriminating each gesture and template with the gesture to be recognized. This method has a small amount of calculation, but this method usually cannot rule out many interference factors, which may affect the accuracy of gestures. The calculation amount for complex gesture recognition is large, which affects the accuracy and real time of the recognition. For example, a dynamic programming algorithm, which solves the effect of time on the template and the image to be tested, compares the features of each frame in the image to be tested with any moment in the template and finds the matching path between them; this method has a high accuracy rate. However, considering that the two samples need to be uniquely identified, it is susceptible to environmental interference to cause matching failures, and with the improvement of the sample library and gestures, the

matching speed will become slower and slower. The initial gesture recognition mainly uses machine equipment to directly detect the angle and spatial position of each joint of the hand and arm. Most of these devices connect the computer system and the user through wired technology so that the user's gesture information can be transmitted to the recognition system completely and without error.

2.2.2. Grammar-Based Approach. The grammar-based method is to describe the dynamics of gestures as a specific dynamic attribute and transform the data into a grammar knowledge base, which can describe many dynamic details that cannot be expressed by the original data information and can accurately express what it really means. The grammar-based method is mainly a finite state machine (FSM) model method. This method has many disadvantages; for example, it cannot accurately describe a dynamic system, and the system is less robust. It is not mature enough for human recognition and needs further development.

2.2.3. Statistics-Based Method. The statistical-based method is mainly applied to the analysis and identification of complex dynamic systems with multiple targets. It has a good scope of application. Commonly used algorithms include Deep Confidence Network (DBN) and Hidden Markov Model (HMM). Deep belief network is more flexible and extensible. It can integrate dynamic systems with multiple information and multiple objects. However, due to its complex structure and many configuration parameters, deep belief network is rarely used in gesture recognition. Hidden Markov Model can be applied to continuous gesture recognition, but because its training requires a large amount of data to support, it is more complicated, and it is difficult to apply it to the recognition scene with a large amount of data. Hidden Markov Models are statistical models that describe a Markov process with hidden unknown parameters. The difficulty is to determine the implicit parameters of the process from the observable parameters. These parameters are then used for further analysis, such as pattern recognition.

2.3. Digital Media Design for Dynamic Gesture Interaction Based on Image Processing

2.3.1. Overall Design. Image processing refers to the technique of analyzing images with a computer to achieve the desired results. It is also called image processing. Image processing generally refers to digital image processing. The interactive touch system based on image processing is mainly divided into three large modules: multitouch detection and positioning module, multitouch digital image processing module, and multitouch gesture tracking and recognition module. The system research and implementation flow is shown in Figure 2. Among them, the commonly used methods of image processing are image transformation, image coding and compression, image enhancement and restoration, image segmentation, image description, and image classification.

As shown in Figure 2, the system research and implementation process of this paper is shown. In-depth research is performed in the order of the three modules mentioned above. Each module will carry out detailed algorithm design and implementation according to the expected goals of the system. Following the pace of technological development, in order to enhance the friendliness of human-computer interaction, this paper is based on optical multitouch technology and digital image processing technology, mainly including infrared sensing technology, image denoising, background differential processing, image feature extraction, edge detection, and algorithms related to motion tracking. According to the system requirements of this paper, we started to research and build a human-computer interactive touch system that can obtain multicontact coordinate information stably, quickly, and accurately and track and recognize user gestures in real time.

Using the visual gesture recognition theory, this paper preliminarily designs a dynamic gesture interactive recognition system based on image processing, which mainly includes several parts as shown in Figure 3.

As shown in Figure 3, the gesture modeling module is a basic part of the overall design of a dynamic gesture interactive recognition system based on image processing and plays a key role in determining the recognition gesture range. The success of the gesture modeling module depends on the gesture environment. Generally speaking, in less demanding situations, only a simple model can complete the system functions. However, the human hand is generally in a complex scene, so it is necessary to build a fine and effective gesture model. The design basis of the gesture modeling module is shown in Figure 4.

As shown in Figure 4, two methods of gesture modeling based on a three-dimensional model and gesture modeling based on gesture appearance are comprehensively used to design a gesture modeling module. First, the three-dimensional model is used to establish the texture, network, geometry, and skeleton models of the gestures. Then, the performance of gestures and the appearance of the opponent's gestures are used to estimate the parameters. The Haar-like feature algorithm and AdaBoost face detection algorithm are used to model the gestures.

2.3.2. Gesture Motion Detection Based on Time Difference Method. The time difference method is a method that extracts the time difference value of pixels between adjacent frames in the continuous image sequence as the measurement value and, after thresholding the measurement value, obtains the moving area of the target object from the image with the processed data. When the gesture is moving in the image sequence, the measurement value changes, and the threshold data processing result in the motion estimation is the degree of change in the image intensity so as to describe the image intensity by the difference between the adjacent frames of the image sequence. Formula (1) defines the calculation of the phase difference of the graph:

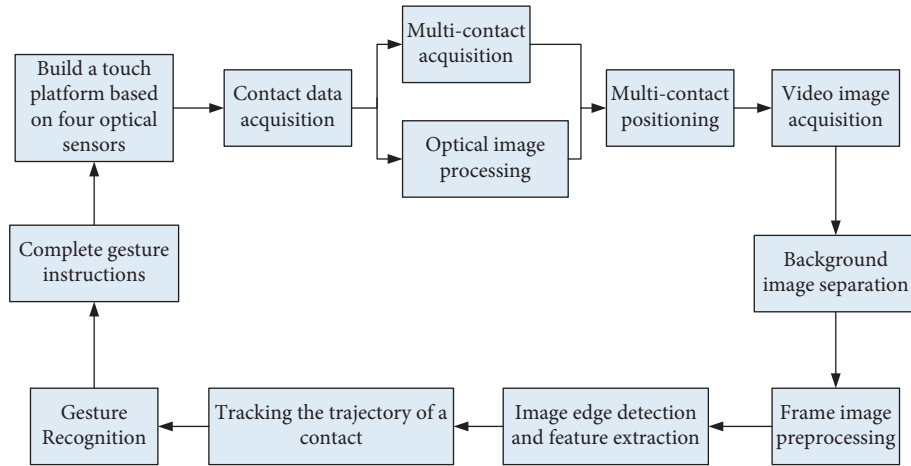


FIGURE 2: System design overall process.

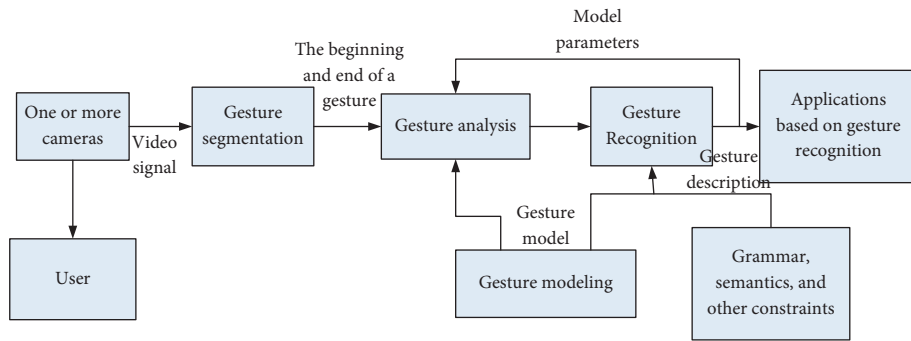


FIGURE 3: Overall design of dynamic gesture recognition system based on image processing.

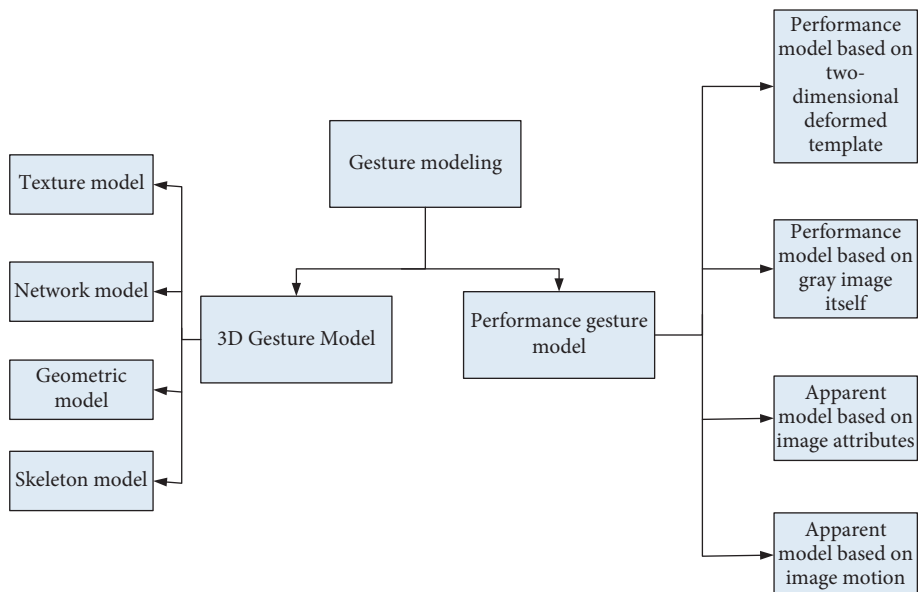


FIGURE 4: Gesture modeling module design basic classification.

$$fd(x, t1, t2) = f(x, t2) - f(x, t1). \quad (1)$$

fd in the formula represents the difference image. In the calculation process, because it only involves subtraction, the algorithm's calculation process is efficient and simple, and the algorithm can be applied to parallel calculation. The difference image can reflect the shape and movement of the target object at a higher level. However, in actual applications, the position of the camera in the gesture recognition system and the different positions or states of the background of the target object should be fully taken into account. Therefore, the difference in the hand must be retained by means of differential calculation to remove the area of the scene that changed. In this case, this paper uses formula (2) to calculate the difference or sum of adjacent images:

$$S = |R1 - R2| + |G1 - G2| + |B1 - B2|. \quad (2)$$

Among them, R1, G1, and B1 and R2, G2, and B2, respectively, represent the RGB value of each pixel in the target image and the previous image of the target image. By calculating the value of S, the calculation result is compared with the set threshold k , and judgment is made according to the comparison result. If $S > k$, the pixel is considered to be a changed area, and if $S \leq k$, the pixel is considered to have no movement, so the pixel is filtered. A new output image can be obtained that reflects the characteristics of the target image. At the same time, considering the interaction between the RGB values of adjacent pixels in the moving process of the target object, this paper introduces a 3×3 auxiliary matrix to carry out difference statistics on the change of RGB values of adjacent frames of the target image, as shown in Table 1:

$$\Delta r(i, j) = \frac{(\sum_s |r_{t2}(i, j) - r_{t1}(i, j)|)}{9}, \quad (3)$$

$$\Delta g(i, j) = \frac{(\sum_s |g_{t2}(i, j) - g_{t1}(i, j)|)}{9}, \quad (4)$$

$$\Delta b(i, j) = \frac{(\sum_s |b_{t2}(i, j) - b_{t1}(i, j)|)}{9}. \quad (5)$$

Formulas (3)–(5) calculate the average value of the statistical change of the RGB component of a pixel point between adjacent frames in the S range. Although the calculation process based on the time difference method based on motion detection is efficient and simple, and it is relatively easy to implement, the regional “hollow” phenomenon of the moving part of the image may occur during this type of calculation, which results in the calculation result being only a moving target part of the edge data information. This needs to be further addressed. After the gesture is generated, the ratio of the person's head to other parts of the body will still produce a certain degree of position change, but the position change relative to the hand is small, so you can perform a cluster analysis on the head or other body parts. It can be objectively considered that after extracting the motion characteristics of other parts of the target object, there are only a small number of target object motion components.

TABLE 1: Auxiliary matrix.

$i-1, j-1$	$i-1, j$	$i-1, j+1$
$i, j-1$	i, j	$i, j+1$
$i+1, j-1$	$i+1, j$	$i+1, j+1$



FIGURE 5: Kinect somatosensory device and PC.

3. Experiments

3.1. Experimental Environment. Kinect was developed by Microsoft Corporation. It was originally applied to the Xbox 360 game console as a game peripheral. It only requires players to complete the game operation through actions and voice and does not require the physical buttons of traditional game consoles, so it is also called a somatosensory device. Due to Microsoft's global popularity and the reliability of the device, it has been loved and favored by many researchers. Kinect can provide deep image information, which can be used in many fields such as behavior recognition, face recognition, and 3D modeling. In addition, it can also capture the movements of the whole body, use the body to play games, and realize body perception recognition and equipment control through the human body. Kinect has achieved great success in the market and has successively released Kinect for Windows v1, Kinect for Windows 1, and Kinect for Windows v2, including open-source software installation packages. Kinect somatosensory equipment and PC are shown in Figure 5.

Four linear infrared sensors, in order to ensure that the technical requirements of this system are wide angle and free of distortion, the Lanzhou TSL1401CL module is used to emit and collect infrared light, and the sensor module automatically transmits the collected data to the data processor for further processing. Touch the reflective strip of the screen to reflect the infrared light emitted by the sensor. The infrared camera has a USB transmission interface to record the user's gesture operation track directly to the data processor via USB for subsequent system processing; a short-throw wide-angle projector projects the image to be displayed on a glass screen with a diffuse reflective coating. The algorithms studied in this paper mainly include multicontact coordinate

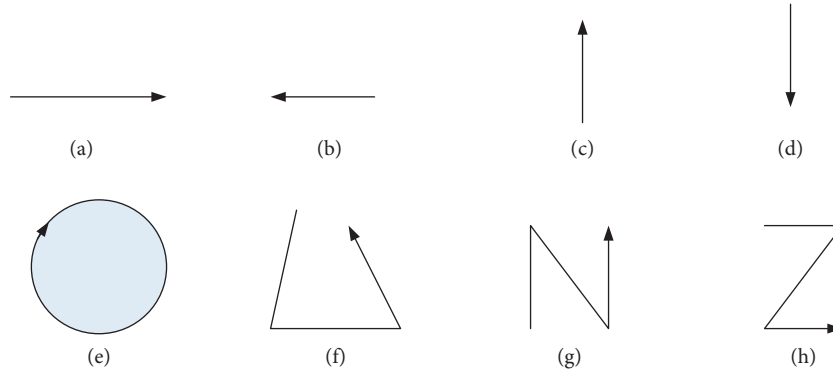


FIGURE 6: Defining 8 trajectories.

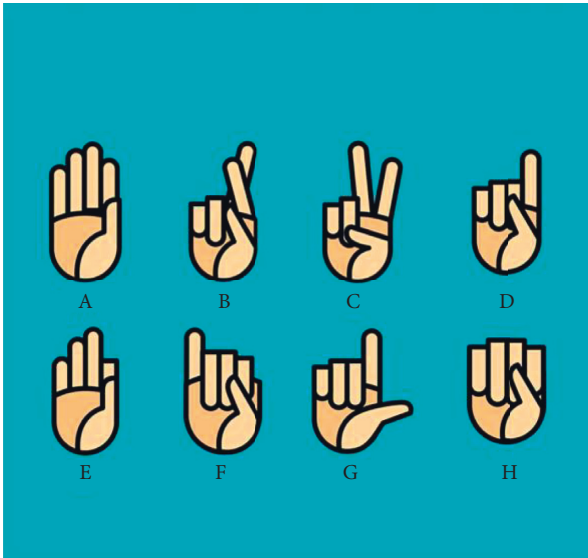


FIGURE 7: Defining 8 gestures.

positioning algorithms, digital image processing, and gesture recognition algorithms that are implemented on the PC side, based on Microsoft under the development environment of Visual Studio 2019; it is realized by C++ coding with the help of OpenCV library functions. This experiment is based on an image recognition dynamic gesture recognition system hardware PC and a web camera. To ensure the smooth running of the system, the minimum configuration of the PC is as follows: CPU running speed is 2.0 GHz, and memory is not less than 1 G. Ordinary H103 G network camera can be used as the camera.

3.2. Experimental Settings. This paper sets 8 kinds of motion trajectories and 8 kinds of gesture images. The combination of static gesture + gesture dynamic trajectory, plus gesture changes, is combined into hundreds of gestures. The static gesture map and trajectory are shown in Figures 5 and 6.

As shown in Figures 6 and 7, the eight types of motion trajectories defined in this paper, numbered a to h, are up, down, left, right, counterclockwise circle, counterclockwise triangle, N-shaped, and Z-shaped. The eight gestures defined are numbered A to H. In the experiment, a motion trajectory

TABLE 2: Dynamic gesture setting table.

Dynamic gesture type	Motion track type	Gesture change type
Dynamic gesture 0	Trajectory a	Gesture A-C
Dynamic gesture 1	Trajectory b	Gesture G-D
Dynamic gesture 2	Trajectory c	Gesture E
Dynamic gesture 3	Trajectory d	Gesture C-A
Dynamic gesture 4	Trajectory e	Gesture A
Dynamic gesture 5	Trajectory f	Gesture B-C-H-E
Dynamic gesture 6	Trajectory g	Gesture C-B-D-G
Dynamic gesture 7	Trajectory g	Gesture C-A-E-F
Dynamic gesture 8	Trajectory h	Gesture A-F
Dynamic gesture 9	Trajectory h	Gesture G-F-E-D

and a gesture are set as a dynamic gesture, and 10 different dynamic gestures are set as shown in Table 2.

3.3. Experimental Data. When the gesture sample database was established, due to the lack of smooth movements of individuals, the joints in the input samples often jittered, causing a lot of interference and useless information. Therefore, this paper uses the Transform Smooth Parameters function provided by Kinect SDK to filter the skeletal nodes to eliminate jitter and calculates the movement of more than 4 pixels in the trajectory as a valid point. The dynamic gesture sample library contains a large number of samples. 2000 samples are collected for each dynamic gesture, and a total of ten kinds of dynamic gestures and a large number of samples will be input and processed. The operation process is tedious and time-consuming; therefore, this paper uses Kinect. The device collects samples and saves the depth images, human bones, and motion trajectories in Kinect. After obtaining the raw data from the software that comes with Kinect, the collected raw data is processed by image preprocessing and K-curvature extraction to obtain gestures. According to the feature node of the extracted gesture, the extraction action is repeated 2,000 times continuously, and the fragments containing the dynamic gesture samples are intercepted to obtain 200 dynamic gesture samples. In this way, for 10 types of dynamic gestures, 200 samples of 10 types of dynamic gestures can be obtained, and a sample database is established.

TABLE 3: Dynamic gesture recognition result table.

Dynamic gesture type	Number of test set samples	Correct number	Accuracy (%)
Dynamic gesture 0	100	96	96
Dynamic gesture 1	100	94	94
Dynamic gesture 2	100	98	98
Dynamic gesture 3	100	96	96
Dynamic gesture 4	100	100	100
Dynamic gesture 5	100	94	94
Dynamic gesture 6	100	84	84
Dynamic gesture 7	100	100	100
Dynamic gesture 8	100	98	98
Dynamic gesture 9	100	86	76

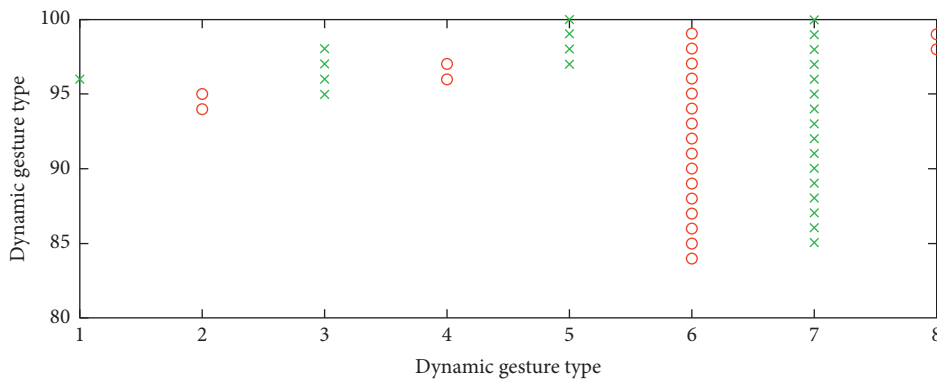


FIGURE 8: Dynamic gesture recognition accuracy.

4. Discussion

4.1. Dynamic Gesture Recognition Results. After the gesture sample database is established, 2000 samples of 10 dynamic gestures are trained and identified. The process is as follows: 200 samples of each dynamic gesture are divided into a training set and a test set, and the training set is used to train the set model and adjust the model parameters; the test set is to check the accuracy of the trained model and determine whether the trained model has been trained. In this paper, 100 samples are selected as the training set of the HMM-NBC model, and the remaining samples are used as the test set of the model. First, the motion trajectory HMM model and the gesture HMM model are trained. The parameters in the HMM need to be set. In this paper, the number of hidden states S in the gesture HMM model is set to 10, and the observation state value M is set to 11. The S value is set to 10, and the M value is set to 13. After the HMM model is initialized, the HMM model training can be started. After all the 10 dynamic gestures are trained, the test set is input to complete the dynamic gesture recognition. The experimental results are shown in Table 3.

As can be seen from Table 3, for the set of 10 kinds of dynamic gestures, the average accuracy rate of the 10 kinds of gestures is 93%, and the comparison effect is shown in Figure 8.

It can be seen from Figure 8 that the accuracy of gestures 0 to 5 is more than 90%, the accuracy of gestures 0, 2, 3, 4, and 8 is as high as 96%, and the accuracy of other gestures is also more than 86%, due to some gestures. The trajectory is

TABLE 4: System test-click gesture accuracy.

Click gesture samples	Accuracy (%)	Response time ms/each
$N = 100$	98.24	≤ 140
$N = 150$	98.68	≤ 135
$N = 200$	98.77	≤ 130
$N = 250$	98.85	≤ 135
$N = 300$	99.23	≤ 135

relatively simple, and the hand changes significantly, so the accuracy is higher; the movement trajectories of gesture 5, gesture 6, gesture 7, and gesture 9 are more complex and the hand changes into 4 types, resulting in lower accuracy. Gesture 0 and gesture 6 contain the same movement trajectory, gestures are different, gesture 0 contains two gestures, and gesture 6 contains four gestures. Therefore, it can be concluded that the accuracy of the same movement trajectory depends on the complexity of the gesture.

4.2. Gesture Instruction Implementation. System test-click gesture accuracy is shown in Table 4.

As shown in Table 4, this system repeatedly tests the click gesture. When the finger touches the screen and presses the icon for more than 1 S, the system can successfully recognize the gesture as a click operation and feed it back to the system to execute the click instruction. The corresponding icon is selected, and the icon is lit. The experimental results show that when the number of samples is larger, the system's recognition accuracy is generally higher, the response speed

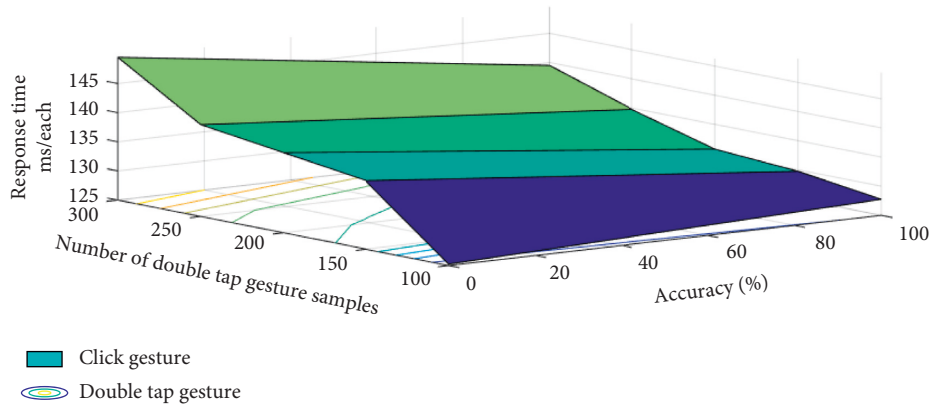


FIGURE 9: System test-click gesture accuracy.

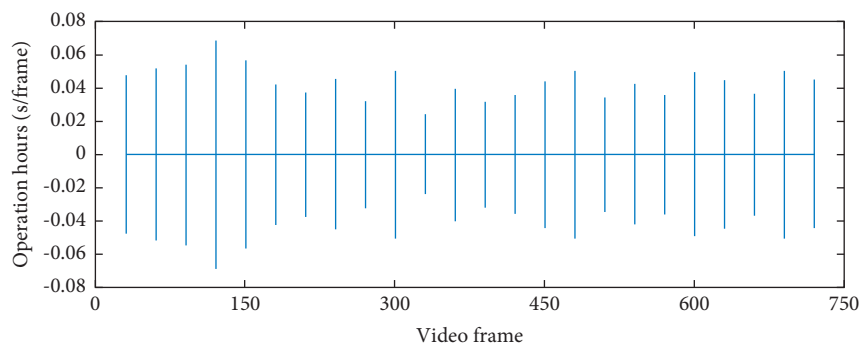


FIGURE 10: Kinect imaging extraction and computational performance test.

is stable at 130 ms, and the system's recognition accuracy is always maintained above 98%. It is confirmed that the system has the expected performance requirements such as high recognition accuracy, fast response speed, and stable operation. System test: the comparison between the accuracy of double-click gestures and the accuracy of single-click gestures is shown in Figure 9.

It can be seen from Figure 9 that the test-double-tap gesture is the same as the single-tap gesture. Double-tap the finger on the same icon on the screen within 1S to display the content of the icon after double-clicking. The icon corresponding to the icon is displayed, and the double-click instruction is completed. It proves that the system has completed the recognition of the double-click gesture and completed the opening instruction of the response. The data shows that the system repeatedly recognizes the double-click gesture with a higher recognition rate and a higher response rate and reaches the system's expected goal.

4.3. Performance Test Analysis. Under this test platform, the performance test is divided into two parts: the performance of Kinect's own imaging and skeletal computing; the performance of two-handed hand extraction and fingertip recognition. Kinect's own imaging and depth data and other information frame rates are related to the set resolution. In this paper, the resolution of the color map and the depth map is $640 * 480$. The comprehensive imaging and calculation performance tests are shown in Figure 10.

As shown in Figure 10, the running time of each frame is about 40 ms, and the peak is about 65 ms, which meets the real-time requirements. The running time of each frame is about 40 ms, and the peak value is about 65 ms, which meets the real-time requirements. Through experiments on different genders, heights, and weights, the accuracy of the system is tested. The test results are shown in Table 5.

As shown in Table 5, the system has little impact on different people. Considering factors such as gender, height, and body shape, the accuracy obtained by the experiment is similar to the dynamic gesture recognition rate. This experiment collected 10 human gestures. Gestures 0–9 are recognized, and it can be obtained that the system is related to the complexity of gestures and is less affected by human differences. This experiment also records the recognition time. It can be obtained that the gesture is more complicated, and the recognition time is relatively long. This system fully meets the requirements of real-time performance. For a specific gesture operator, the accuracy rate has a great relationship with the standard degree of its operation. If the similarity with the model is high, a high recognition rate can be achieved.

4.4. Functional Test Analysis. The purpose of the experiment is to control the playback of the PPT through the recognition of gestures. Therefore, for the defined 8 gesture commands, the traditional way of matching is to directly calculate the forward probability of the observed value sequence. In the

TABLE 5: Human-computer interaction system test result.

Frequency	Gender	Body type (cm, kg)	Gesture	Accuracy (%)	Average recognition time
1	Male	164, 56	Dynamic gesture 0	93	1.088
2	Male	169, 84	Dynamic gesture 1	92	1.153
3	Male	178, 67	Dynamic gesture 2	95	0.994
4	Male	167, 58	Dynamic gesture 3	87	1.575
5	Male	172, 82	Dynamic gesture 4	89	1.599
6	Female	156, 43	Dynamic gesture 5	95	0.985
7	Female	167, 49	Dynamic gesture 6	91	1.043
8	Female	173, 68	Dynamic gesture 7	89	1.077
9	Female	161, 46	Dynamic gesture 8	87	1.466
10	Female	166, 53	Dynamic gesture 9	93	1.573

TABLE 6: Comparison between the original method and the proposed sliding window method.

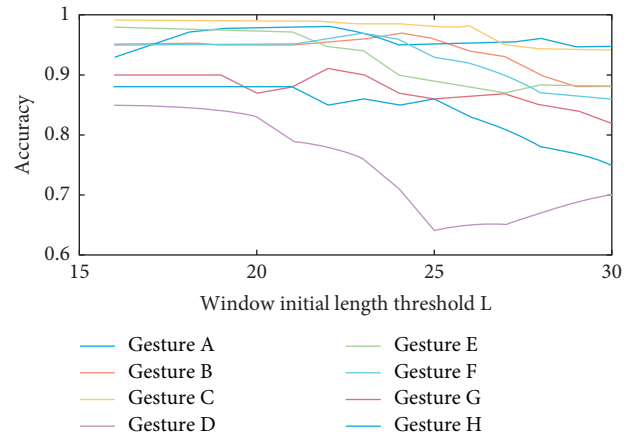
Gesture	Output command	Number of training samples	The correct number of traditional matches	Sliding window matches the correct number
A	The last page	190	171	189
B	First page	250	151	240
C	Start playing	230	155	215
D	End playback	220	150	211
E	Next page	210	145	205
F	Previous page	420	190	410
G	Autoplay	210	180	201
H	Return	220	177	213

experiment, we take traditional matching and use sliding window to determine the matching model of the sequence for comparison and get the data results in Table 6 (initial length threshold $L = 19$).

As shown in Table 6, from the experimental results, it is found that the result of the horizontal left in the correct rate is lower than that of other models. The main reason after analysis is that, in the positioning of the human hand, Kinect will shake when the hand passes through the chest area, affecting the experiment. On the basis of increasing the number of models and the test sample size, the built model still tends to be about 86% correct. But on the whole, the accuracy of the algorithm in each model is higher than that of the traditional algorithm, and it has achieved good experimental results. In terms of time complexity, the calculation amount when the observation sequence length reaches 36 is about 133 ms, which meets the requirements of real time in the experiment of the application of this system-PPT courseware playback.

The first effect of the length threshold L on the experiment is the efficiency problem, if the efficiency is reduced. If the value of L is too large, then the amount of calculation increases, which reduces the efficiency; if the value of L is too small, it will affect the accuracy of the calculation and will inevitably affect the function of the system. This paper tested L during the experiment and calculated the effect of L on the experimental results when it is between 15 and 29. The results are shown in Figure 11.

As shown in Figure 11, it can be seen from the figure that as L increases, the accuracy rate generally increases, but all of them decline after reaching a certain peak. The results are obtained by selecting different L values for several models.

FIGURE 11: The effects of the initial window's length L on the results.

This paper chooses $L = 19$ as the optimal initial window threshold of the experiment, which has better adaptability, and this value has been verified in subsequent tests.

5. Conclusions

Since the 1980s, the software and hardware technology of computers have made great progress, and at the same time, the users of computers have rapidly expanded from computer experts to ordinary users who have not received special training. This greatly increases the importance of user interface in system design and software development and strongly stimulates the progress of the human-computer interface.

Gesture is a highly natural and intuitive mode of communication. The human hand is directly used as the input device of the computer, eliminating the intermediate medium existing in traditional human-computer interaction, and the user can realize simple and direct interaction with the computer. In this paper, some methods of each step of gesture recognition are roughly researched and introduced, and some methods used in this paper and the results of experiments are focused on.

In this paper, the related algorithms of the interactive touch system based on image processing are analyzed and researched, and the implementation of the algorithm is completed on the PC. This system is mainly divided into three modules: multicontact detection and positioning module, multicontact digital image processing module, and multitouch gesture tracking and recognition module. In this paper, by writing a contact positioning algorithm, a contact edge extraction algorithm, and a gesture recognition algorithm, the detection and positioning of the contacts and the recognition of gesture instructions are completed.

With the rapid development of digital media technology, great progress has been made in digital media technology based on the errors, but there are still shortcomings in terms of ease of use and robustness. This paper has done a lot of work on gesture recognition in gesture segmentation and extraction, which improves the accuracy and natural experience of gesture recognition, but there are still simple gesture commands, and command recognition uses probability statistics to achieve the level of semantic recognition. In addition, fingertip recognition stays in a two-dimensional space, and inaccurate recognition occurs when the hand is not facing the camera device. However, due to the limitation of time and technology, we have not explored it in depth, and we will conduct further experimental research in the follow-up work.

Data Availability

This paper does not cover data research. No data were used to support this study.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the 2018 Special Scientific Research Project of the Shaanxi Education Department "Application Research of Digital Animation in the Protection and Inheritance of Intangible Cultural Heritage in Eastern Qin" (no. 18JK0267); the 2021 Shaanxi Higher Education Teaching Reform Research Project "Research on the Cultivation and Improvement Path of Informatization Teaching Ability of Normal University Students Based on Professional Certification" (no. 21BY149); and the 2021 Higher Education Scientific Research Project of Shaanxi Higher Education Association "Research and Practice on the Cultivation System of Normal University Students' Informatization Teaching Ability under the Context of New Liberal Arts" (no. XGH21212).

References

- [1] H. J. Wood and R. Albrecht, "Digital image processing of Apstar coude Zeeman plates," *Asia Europe Journal*, vol. 13, no. 2, pp. 163–174, 2015.
- [2] G. T. Herman, R. Marabini, J.-M. Carazo, E. Garduno, and R. M. Lewitt, "Image processing approaches to biological three-dimensional electron microscopy," *International Journal of Imaging Systems and Technology*, vol. 11, no. 1, pp. 12–29, 2015.
- [3] R. T. Mullapudi, V. Vasista, and U. Bondhugula, "PolyMage," *ACM SIGARCH - Computer Architecture News*, vol. 43, no. 1, pp. 429–443, 2015.
- [4] R. Membarth, O. Reiche, F. Hannig, J. Teich, M. Körner, and W. Eckert, "HIPAcc: a domain-specific language and compiler for image processing," *IEEE Transactions on Parallel and Distributed Systems*, vol. 27, no. 1, pp. 210–224, 2015.
- [5] D. Mortari, N. D. Christopher, and R. Zanetti, "Image processing of illuminated ellipsoid," *Journal of Spacecraft and Rockets*, vol. 53, no. 3, pp. 1–9, 2016.
- [6] C.-H. Chen, C.-H. Ting, and Y.-P. Huang, "3D interactive system using integral photography by embedded optical sensors for portable devices," *Journal of Display Technology*, vol. 12, no. 11, pp. 1329–1334, 2016.
- [7] Y. Chen, L. J. Xu, and Z. L. Gong, "Application of image processing and analysis techniques in molecular imaging," *Journal of Shanghai Jiaotong University*, vol. 35, no. 4, pp. 605–610, 2015.
- [8] J. Li, Y. Xia, W. Wang, F. Wang, W. Zhang, and W. Zhu, "Dipping process characteristics based on image processing of pictures captured by high-speed cameras," *Nano-Micro Letters*, vol. 7, no. 1, pp. 1–11, 2015.
- [9] X. Hu, G. Feng, S. Duan, and L. Lu, "A memristive multilayer cellular neural network with applications to image processing," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 8, pp. 1–13, 2016.
- [10] J. Hegarty, R. Daly, Z. Devito, J. Ragan-Kelley, M. Horowitz, and P. Hanrahan, "Rigel," *ACM Transactions on Graphics*, vol. 35, no. 4, pp. 1–11, 2016.
- [11] X. Lu and J. Fei, "Velocity tracking control of wheeled mobile robots by iterative learning control," *International Journal of Advanced Robotic Systems*, vol. 13, no. 3, p. 1, 2016.
- [12] S. Lemeš, D. Štrbac, and M. Cabaravdic, "Using industrial robots to manipulate the measured object in CMM," *International Journal of Advanced Robotic Systems*, vol. 10, no. 281, pp. 1–9, 2013.
- [13] P. Forbrig, F. Paternó, and A. M. Pejtersen, "Human-computer interaction," *Encyclopedia of Creativity Invention Innovation & Entrepreneurship*, vol. 19, no. 2, pp. 43–50, 2017.
- [14] E. Churchill, A. Bowser, and J. Preece, "Teaching and learning human-computer interaction," *Concurrency and Computation: Practice and Experience*, vol. 28, no. 4, pp. 1291–1309, 2016.
- [15] F. Schwenker, S. Scherer, and L. P. Morency, "Multimodal pattern recognition of social signals in human-computer interaction: third IAPR TC3 workshop, MPRSS 2014, stockholm, Sweden, august 24, 2014, revised selected papers," *Lecture Notes in Computer Science*, vol. 7742, no. 2, pp. 123–141, 2015.
- [16] A. Ramkumar, P. J. Stappers, W. J. Niessen et al., "Using GOMS and NASA-TLX to evaluate human-computer interaction process in interactive segmentation," *International Journal of Human-Computer Interaction*, vol. 33, no. 2, pp. 1–12, 2016.

- [17] A. Bulling and K. Kunze, "Eyewear computers for human-computer interaction," *Interactions*, vol. 23, no. 3, pp. 70–73, 2016.
- [18] J. O. Wobbrock and J. A. Kientz, "Research contributions in human-computer interaction," *Interactions*, vol. 23, no. 3, pp. 38–44, 2016.
- [19] T. Gross, J. Gulliksen, P. Kotzé, L. Oestreicher, and P. Palanque, "Human-computer interaction – interact 2009," *Lecture Notes in Computer Science*, Springer Nature, vol. 5726, no. 2, , pp. 131–141, Switzerland, 2017.
- [20] Z. Lv, A. Halawani, S. Feng, H. Li, and S. U. Réhman, "Multimodal hand and foot gesture interaction for handheld devices," *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 11, no. 1, pp. 1–19, 2014.
- [21] N. Ramu, V. Pandi, J. D. Lazarus, and S. Radhakrishnan, "A novel trust model for secure group communication in distributed computing," *Journal of Organizational and End User Computing*, vol. 32, no. 3, pp. 1–14, 2020.
- [22] L. M. T. Pham, L. T. T. Tran, P. Thipwong, and W. T. Huang, "Dynamic capability and organizational performance," *Journal of Organizational and End User Computing*, vol. 31, no. 2, pp. 1–21, 2019.
- [23] Y. Hong, M. Wan, and Z. Li, "Understanding the health information sharing behavior of social media users," *Journal of Organizational and End User Computing*, vol. 33, no. 5, pp. 180–203, 2021.
- [24] V. Gatteschi, F. Lamberti, P. Montuschi, and A. Sanna, "Semantics-based intelligent human-computer interaction," *IEEE Intelligent Systems*, vol. 31, no. 4, pp. 11–21, 2016.
- [25] Y. Zou, J. Xiao, J. Han, K. Wu, Y. Li, and L. M. Ni, "GRfid: a device-free RFID-based gesture recognition system," *IEEE Transactions on Mobile Computing*, vol. 16, no. 2, pp. 381–393, 2017.
- [26] R. Zinko, H. de Burgh-Woodman, Z. Z. Furner, and S. J. Kim, "Seeing is believing," *Journal of Organizational and End User Computing*, vol. 33, no. 2, pp. 85–104, 2021.
- [27] J. R. Drake and C. P. Furner, "Screening job candidates with social media," *Journal of Organizational and End User Computing*, vol. 32, no. 4, pp. 63–84, 2020.
- [28] S. S. Rautaray and A. Agrawal, "Vision based hand gesture recognition for human computer interaction: a survey," *Artificial Intelligence Review*, vol. 43, no. 1, pp. 1–54, 2015.
- [29] G. Plouffe and A. M. Cretu, "Static and dynamic hand gesture recognition in depth data using dynamic time warping," *IEEE Transactions on Instrumentation and Measurement*, vol. 65, no. 2, pp. 305–316, 2015.
- [30] M. Devi, S. Saharia, and D. K. B. D.K.Bhattacharyya, "Dance gesture recognition: a survey," *International Journal of Computer Application*, vol. 122, no. 5, pp. 19–26, 2015.
- [31] Y. L. Yu-Liang Hsu, C. L. Cheng-Ling Chu, Y. J. Yi-Ju Tsai, and fnm Jeen-Shing Wang, "An inertial pen with dynamic time warping recognizer for handwriting and gesture recognition," *IEEE Sensors Journal*, vol. 15, no. 1, pp. 154–163, 2015.
- [32] R. Craciunescu, A. D. Mihovska, and S. Kyriazakos, "Non Audio-Video gesture recognition system," *Acta Veterinaria Scandinavica*, vol. 8, no. 2, pp. 98–110, 2016.
- [33] F. Despinoy, D. Bouget, G. Forestier et al., "Unsupervised trajectory segmentation for surgical gesture recognition in robotic training," *IEEE Transactions on Bio-Medical Engineering*, vol. 63, no. 6, pp. 1280–1291, 2015.
- [34] F. Pla, P. Latorrecarmona, E. Salvadorbalaguer, and B. Javidi, "Integral imaging acquisition and processing for human gesture recognition," *Amphibia-Reptilia*, vol. 12, no. 1, pp. 15–20, 2015.
- [35] L. Hong, L. Rong, and L. I. Shuling, "Acceleration gesture recognition based on random projection," *Journal of Computer Applications*, vol. 35, no. 1, pp. 189–193, 2015.
- [36] K. M. Sagayam and D. J. Hemanth, "Hand posture and gesture recognition techniques for virtual reality applications: a survey," *Virtual Reality*, vol. 21, no. 2, pp. 1–17, 2016.