IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received July 7, 2020, accepted July 28, 2020, date of publication August 3, 2020, date of current version August 14, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3013903

Research on Intelligent Experimental Equipment and Key Algorithms Based on Multimodal Fusion Perception

BOTAO ZENG, ZHIQUAN FENG, (Member, IEEE), TAO XU, MENGTING XIAO, AND RUI HAN
School of Information Science and Engineering, University of Jinan, Jinan 250022, China
Shandong Provincial Key Laboratory of Network Based Intelligent Computing, Jinan 250022, China

Corresponding author: Zhiquan Feng (ise_fengzq@ujn.edu.cn)

This paper was supported by the National Key R&D Program of China (No. 2018YFB1004901), and the Independent Innovation Team Project of Jinan City (No. 2019GXRC013).

ABSTRACT The application of virtual reality technology in science experiment education is a research with practical significance and value in human-computer interaction. However, in some existing education tools based on virtual reality, due to the single interaction mode, the complexity of user intention and the non-physical interaction characteristics brought by virtualization, their experimental teaching ability is limited, resulting in the lack of practical value and popularity. In order to solve these problems, a multimodal interaction model is constructed by fusing gesture, speech and pressure information. Specifically, our tasks include: 1) collecting user input information and time series information to construct basic data input tuples. 2) The basic interaction information is used to identify the user's basic intention, and the correlation degree between the user's intentions is considered to determine the correctness of the current identification intention. 3) It allows users to alternate between multi-channel and single channel interaction. Based on this model, we build a multi-modal intelligent interactive virtual experiment platform (MIIVEP), and design and implement a kind of dropper with strong perception ability, which has been verified, tested, evaluated and applied in the intelligent virtual experiment system. In addition, in order to evaluate this work more effectively, we developed a fair scoring criterion for the virtual experimental system (Evaluation scale of virtual experiment system, ESVES), and invited middle school teachers and students to participate in the verification of the results of this work. Through the user's actual use effect verification and result research, we prove the effectiveness of the proposed model and the corresponding implementation.

INDEX TERMS Virtual experiment, intelligent dropper, multimodal fusion, pressure sensors, human-computer interaction.

I. INTRODUCTION

The development and innovation of science and technology have changed people's way of life and influenced how people think and learn. These changes have led to constant adjustments and improvements to educational methods, from words to images, and reality to virtual reality. In traditional education, for subjects centered on scientific education, frequent experiments require educators to spend considerable energy preparing experimental equipment and tutoring students. The one-to-many model can deteriorate the energy levels of teachers and lead to unsatisfactory experimental results. Additionally, the huge demand for experimental materials can result in a large economic burden. As a result, the educational tools based on virtual reality technology began to receive attention. Virtual reality technology has strong simulation ability. Researchers hope to develop more intelligent and effective experimental teaching tools with this advantage, which is virtual teaching experiment system.

The early auxiliary experimental systems were designed based on two-dimensional graphics [1]. Although experiments could be performed, the results could be verified, and large quantities of materials were not needed, these methods lacked authenticity related to teacher supervision and experience. Three-dimensional virtual experimental systems have significantly improved the presentation quality.

The associate editor coordinating the review of this manuscript and approving it for publication was Xiaogang Jin.

Scene simulation enhancements have improved the visual rendering and immersion effects of virtual experiments [2]. However, the dependence on traditional input devices, such as mouse and keyboard, weakens the effective interaction experience in the experimental process. Thanks to the development of science and technology, the advent of virtual reality equipment provides a more natural interactive way for many interactive projects, such as sensing the three-dimensional space position of human and human hands through depth camera. This article [3] uses this kind of equipment to improve the interaction mode of the experiment, and applies it to the visual teaching interaction experiment, so that the sense of operation experience and immersion of the virtual teaching experiment is significantly improved. But in the current virtual teaching system, there are the following problems: first, based on the visual or single-mode interaction mode, users need to memorize complex operation commands, which will increase the workload of users. Secondly, these single-mode commands cannot accurately convey the user's real operation intention, resulting in the inconsistency between the user's intention and the actual behavior results, thus reducing the actual teaching effect of the virtual experimental system. Finally, the fully virtualized experimental equipment hinders the user's cognition of the experimental equipment to a certain extent, which makes the user's cognition of the use value and function of the experimental equipment fuzzy, resulting in the user's cognition deviation of the experimental content in principle. In response to the aforementioned limitations, we propose a reliable and effective interaction framework: on the one hand, using the characteristics of multi-modal information such as redundancy and complementarity, and the characteristics of single-modal information such as modal characteristics, we use more diversified user information as the basis for inferring user interaction intention, to achieve accurate judgment of user intention. On the other hand, using the characteristics of different modes to simplify the user's intention expression and reduce the user's interaction load will help the user focus on the exploration and understanding of the experimental content. At the same time, based on the concept of virtual reality integration, we designed and implemented an intelligent dropper, which can simulate a more real interaction experience through intelligent devices, which is helpful for users to understand the experimental principle and the experimental process, and master the actual related experimental skills.

II. RELATED WORK

Virtual reality technology is a research field that has attracted considerable attention in recent years. This technology can provide users with realistic visual, auditory and other sensory feelings through interactive devices such as virtual helmets and data gloves, giving people a real experience in the scene. Its unique virtual characteristics and intelligent interaction mode make virtual reality technology widely used in military, medical, training, education and other fields [4].

The intrinsic and immersive interactions provided by virtual reality technology allow such technology to be applied in cognitive rehabilitation training [5], and related research has made substantial progress. Rizzo *et al.* successfully applied virtual reality technology in PTSD assessment and the treatment of soldiers [6]. In the field of higher education, virtual teaching laboratory is a new teaching tool based on the development of virtual experiment technology. It has many significant advantages: first, the traditional experimental teaching method has notable environmental and material requirements, especially for chemistry and physics experiments. The cost of experimental material consumption and laboratory maintenance is high, and virtual experimental systems provide a cheaper choice than traditional laboratories [7]. Second, some experiments have potential dangers, which virtual experiment teaching has closed in the virtual environment, and solved the problem that traditional teaching is difficult to repeat experiments. Third, this kind of teaching method helps to improve the focus of their opponents' tasks [8], which is the problem that teachers pay attention to [9]. Fourth, in the science teaching course, the dynamic visualization of virtual experiment can make students get better results in acquiring knowledge [10]. In addition, the teaching objects of virtual experiment system are diverse, which not only helps normal students, but also helps students with physiological disorders to learn. Balado *et al.* designed and developed an electronic circuit laboratory (VLEC) based on a remote Internet mode and provided electronic circuit course assistance for hearing-impaired students [11], and effectively improve the scores of the hearing impaired students in the electronic circuit related courses. Additionally, a practical study showed that students prefer computer-aided tools to textbooks [12].

The current virtual teaching laboratories are all aimed at a specific population or limited to a certain institution. For example, Shin *et al.* designed and implemented a web-based interactive virtual laboratory system for unit operation and process system engineering education to reduce the cost of experimental operation and improve the efficiency of education [13]. Naranjo *et al.* designed a photon virtual laboratory for LED research and provided a verification mechanism to evaluate the contribution of new materials to traditional teaching methods and identify student learning outcomes [14]. Duarte *et al.* designed a general virtual laboratory for electrical engineering [15]. The laboratory provided a real and enhanced learning experience for students with poor upper body mobility and was equipped with an intelligent laboratory assistant. Kim *et al.* designed a virtual laboratory system for electronic and digital circuit experiments [16]. Through this system, virtual experimental data similar to real experimental data can be obtained, thereby improving the efficiency of learners and educators. Gustavsson [17] built a remote virtual laboratory that provides students with remote experimental opportunities and teaches them to effectively use experimental equipment. These virtual laboratories generally serve higher education and ignore the experimental requirements of chemistry and physics in secondary education. Additionally, experimental control still relies on mouse and keyboard devices, which are obviously not sufficiently intelligent in the current era of human-computer intelligent interaction. Based on Kinect device, Liao *et al.* Implemented a virtual experiment system for electrical training, which enables people and virtual objects to interact directly through gestures [18], rather than through mouse and keyboard, which is obviously more suitable for users' natural interaction than mouse interaction. However, considering the current situation of scientific experiments in secondary schools, chemistry and physics experiments are the main body of middle school experiments, which should pay more attention to the authenticity and experience of the operation process. Therefore, single modal interaction is obviously unable to satisfy this situation. Multimodal interaction can enhance the performance of virtual teaching experiment system more effectively.

Multimodal information interaction is based on voice, vision, touch and other multimodal information sharing. The key to solve the problem is to build multimodal fusion model. Multimodal fusion is generally the integration of related features or intermediate decisions of multiple media [19], which can provide complementary information, so as to improve the accuracy of decision-making process. Among them, fusion level and fusion method are the main research directions. From the fusion level, the most extensive strategy is to fuse information in the feature layer, also known as early fusion. Another method is decision level fusion or post fusion [20], [21]. Among the fusion methods, there are rule-based fusion method [22], [23], classification based fusion method [24], [25] and estimation based fusion method [26], [27]. The most widely used method is classification based fusion method, which includes a series of classification techniques, which is used to classify the observed multimodal information into a predefined class. These classification techniques include Bayesian inference [20], [28], support vector machine (SVM) [29], [30], dynamic Bayesian network (DBN) [31], and maximum entropy model. In addition, neural network (NN) is also a method to integrate multimodal data, which is similar to a nonlinear black box, and can be trained to solve the problem of unclear definition and complex calculation [32]. These methods provide researchers with models to solve practical problems, such as SVM based identification [33], Bayesian based speech recognition [28] and speech digital recognition [34], dynamic Bayesian based shot classification of moving video [35]. Information fusion can also effectively enhance the anti-interference ability and robustness of algorithms or systems. For example, Du *et al.*, for robot control, combined Kalman filter and particle filter, combined with leap motion and Kinect [36], improved the stability and reliability of robot control, and proposed a marker free human-machine interface [37].

To sum up, this article takes the actual teaching needs as the motivation, and analyzes the general defects and weaknesses of the current virtual teaching experiment system from the perspective of intelligent interaction based on multimodal information fusion. This article constructs a multi-modal information interaction framework for virtual teaching, which can deduce the real intention of users according to the subtasks contained in the subtasks by analyzing the different modal information in the process of user interaction, and realize the intelligent and natural human-computer interaction. On this basis, this article designs and implements an intelligent experimental device based on virtual reality fusion technology, which simulates the interactive operation of real devices and enhances the user's understanding and memory of the experimental principle and content. In order to evaluate our work, we developed an effective grading rule (ESVES) based on the characteristics of virtual experimental system, which will be used by 40 volunteers to make objective use evaluation of our system.

III. DESIGN AND IMPLEMENTATION OF INTELLIGENT EXPERIMENT

In this part, we introduce the implementation algorithm and details of the interactive framework of the intelligent experimental system. At the same time, a kind of virtual experimental equipment is designed and implemented: intelligent dropper.

A. INTERACTION FRAMEWORK

The virtual experiment system can provide 3-D virtual scientific experiment environment, and can realize the experiment phenomenon and operation experience with strong sense of reality. In order to meet the needs of users for low load, high efficiency and high experience, this article takes multimodal information interaction as the core to improve the interactive experience and learning efficiency of users, and reduce the interaction load. The specific interaction scheme is shown in Fig. 1.

The whole process of virtual experiment platform can be divided into multi-modal information input, multi-modal information processing and identification, and interactive application with processing results. In the process of multimodal information input, the microphone obtains the user's voice information, and the tactile sensor and depth camera capture the user's operation behavior. Visual information mainly includes image information and depth information, which is used to detect the user's gesture and the spatial position of virtual hand, and further analyze the user's operation intention. The acquisition of tactile information is mainly used to analyze the user's operation behavior on intelligent devices, such as determining the time, frequency and measurement of the user's using the dropper. The processing and identification of multimodal information is the key to infer the user's intention. In this part, multimodal information is classified and sorted into different subtasks, and the way to integrate these subtasks is decided to infer the user's real intention, so as to realize the intelligent interaction between the system and the user. In the whole multi-modal

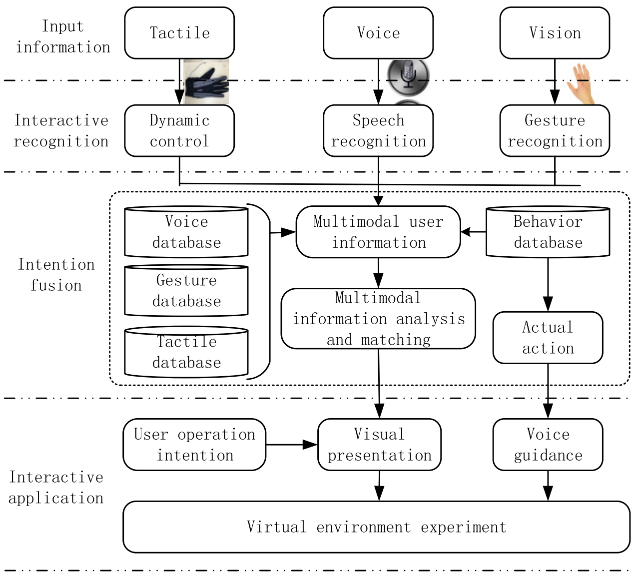


FIGURE 1. Overall framework.

information processing, we use the modal information as much as possible to analyze the context of user operation behavior, and realize the understanding and feedback of user operation reasonably and effectively. The significance of multimodal information processing lies in: on the one hand, it is more complete to obtain the user's operation information to avoid the error of understanding the user's intention caused by the lack of information; on the other hand, it provides more diversified ways for the expression of user's intention, which can minimize the user's operation load and meet the interaction needs of users with different expression habits Intelligent interaction effect.

In the interactive application, voice and vision are the main ways of information transmission. The system will show the experimental content to the user in a visual way, and give the user effective guidance and prompt through voice. In general, the multimodal information processing framework considers the behavior of the system and the user in the interactive application.

B. DESIGN OF INTELLIGENT EXPERIMENTAL EQUIPMENT

The dropper is responsible for the quantitative control of reaction reagents and the connection of the experimental process in scientific experiments. In the general virtual experiment, the full virtual setting mode deprives the responsibility of the dropper. In the experiment, users can only choose to use the "add" and "do not add" functions of the dropper. They can't control the size of the drop and the amount of reagent added. In order to achieve more real operation experience and intelligent interaction, it is necessary to construct a kind of dropper with virtual and real integration and perception ability.

In order to ensure that the real meaning of virtual devices is not weakened, the original value in the experiment is lost. It is very important to endow the virtual device with the same ability and experience as the real device. This article takes the intelligent dropper as an example to verify the importance of this work.

1) STRUCTURAL DESIGN

We use the micro sensor and glove to realize the intelligent dropper. There is a micro pressure sensor at the index finger of the glove. When the user uses the dropper, the sensor receives the user behavior information. This design can provide interactive simulation with minimal user operation burden (Fig. 2).

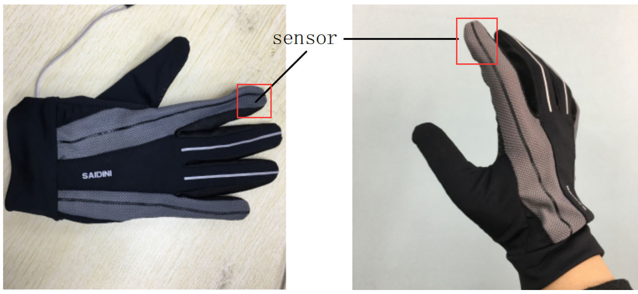


FIGURE 2. Intelligent dropper hardware.

2) PRESS DYNAMIC SIMULATION

In practice, there is a deformation process of the drop at the mouth of the dropper. When the drop is large enough, the liquid will fall. We designed and implemented this process. With the change of force, the drop size and drop trend will also change. If the user finds that the reagent added is incorrect in the process of adding reagent, he can immediately cancel the dripping operation. For the convenience of description, the definition of droplet morphology is given as $Df = \{sl, sv\}$, where sl represents the spatial position vector of the droplet, and sv represents the deformation vector of the droplet on each direction axis. We use S_push to represent the sequence of user pressed values over a period of time S_pt. $\mathrm{D}(S\_push) = d(S\_push)/d(S\_pt)$ represents the pressure change at the current moment, corresponding to the actual droplet morphology change, as shown in TABLE 1.

Combined with formula (1), the dynamic presentation of droplet changes can be realized and more real interaction effect can be provided.

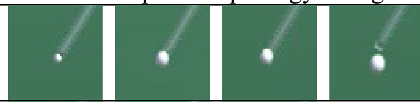
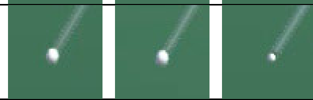
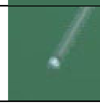
$$Df' = Df + \frac{s_{\max} - D(S\_push)}{s_{\max}} * Tra^T \tag{1}$$

Tra is a transfer vector. When the shape of the droplet changes, the relative position of the droplet will shift to a certain extent. Tra is used to correct the position of the droplet, $s_{\max}$ represent $\max(S\_push)$.

3) ENHANCED FUNCTIONALITY

In fact, there are different models of droppers. According to different experimental requirements, users often use different types of droppers.

TABLE 1. Success rate of user intent of T1.

Change of pressure value	User intention	Effect of droplet morphology change
D(S_push) > 0	Dropping reagent	
D(S_push) < 0	Cancel current drip operation	
D(S_push) → 0	No operation	

If more reagents are used in an experiment, more than one dropper or only one dropper will be used, but frequent cleaning is required. In the virtual experiment, using multiple droppers is obviously not suitable, which will increase the user's operation burden. It is a good choice to use only one dropper, but frequent cleaning operation will also increase the user's burden. Therefore, we use voice recognition to assist users. Users can clean the dropper through voice instructions. Of course, this is an optional user operation, and the cleaning operation is more in line with the actual use criteria of the dropper. In addition, in general, we also consider other situations. For example, the user presses on the dropper, which causes the droplet to spray out. The simulation of this process is helpful to help users understand the use of dropper. We do this with the following formula

$$Df_{sp} = Dv * \frac{s_{\max} - s_{\min}}{s_{\max}} \tag{2}$$

Df_{sp} is the drop velocity, and its maximum velocity is Dv, $s_{\min}$ represent min(S_push).

C. MULTIMODAL INTENTION UNDERSTANDING

In this article, a multi-modal intelligent interactive virtual experiment platform is constructed based on the integration of vision, touch and voice. The basic processing framework is shown in Fig. 3.

After the experiment starts, the user operation information is acquired and processed in the input layer. The pre-processing results of each modal information are matched with the sub intention database in the perceptual layer, and the fuzzy sub intention set is obtained by combining the interactive target information transmitted by the virtual hand. In the fusion layer, the multi-modal intention fusion algorithm uses the sub intention set to input each modal time series transmitted by the layer to obtain the temporary user intention. When the previous operation intention exists and is valid, the temporary user intention combines the information of previous user intention to get the final user intention of this operation. The final intention obtained is directly applied to MIIVEP, and then the current final intention will replace the position of the previous operation intention as one of the conditions for the next intention fusion.

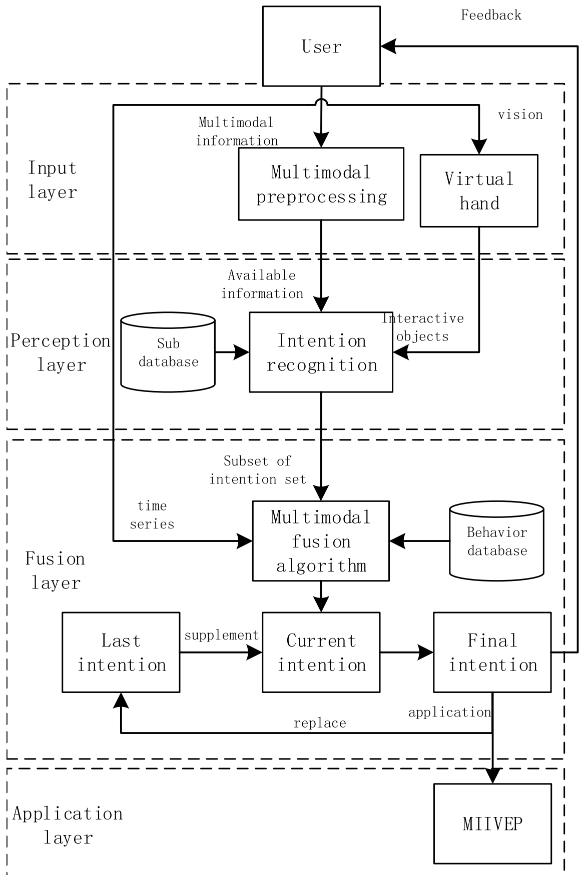


FIGURE 3. Fusion model framework.

1) USER INTENTION AND BEHAVIOR DATABASE

Firstly, the sub intention database and experimental behavior database of different modes are established (TABLE 2). According to the number of modes, three sub intention databases Db_Ta, Db_Vi, Db_Ge and behavior database Db_Ac are established. At the same time, a target database Db_Tg is established according to the experimental content.

Among them, Db_Vi is a speech sub intention database based on speech, which contains intention numbers n_2 determined by the actual experimental content, other n_i is

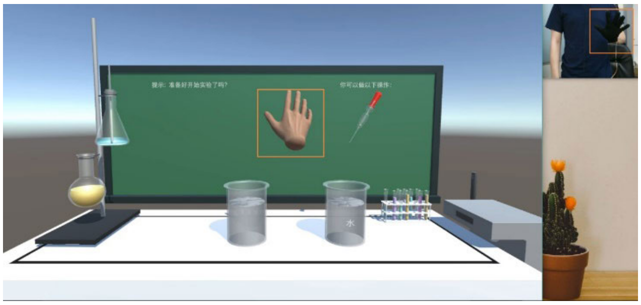


FIGURE 4. **Initial interface of experiment.**

TABLE 2. **Database.**

Database name	Set	Time label
Db_Ta	$Db\_Ta = \{ta_1, ta_2, ta_3, \ldots ta_{n1}\}$	1
Db_Vi	$Db\_Vi = \{vi_1, vi_2, vi_3, \ldots vi_{n2}\}$	2
Db_Ge	$Db\_Ge = \{ge_1, ge_2, ge_3, \ldots ge_{n3}\}$	3
Db_Ac	$Db\_Ac = \{ac_1, ac_2, ac_3, \ldots ac_{n4}\}$	-
Db_Tg	$Db\_Tg = \{tg_1, tg_2, tg_3, \ldots tg_{n5}\}$	-
N	$N = \{n_1, n_2, n_3, n_4, n_5\}$	-

determined in the same way. In addition, according to the analysis, the order of each mode instruction has a great influence on the accurate understanding of users, so the time series $Time\_sq$ is considered, tactile, speech and visual are respectively time tagged.

For example, $Time\_sq = [2,1,3]$ indicates that the user sends out the voice instruction first and then the visual instruction. If $Time\_sq = [3,0,0]$, which means that the user only issued visual instructions.

2) MULTIMODAL INFORMATION PERCEPTION AND RECOGNITION

In the interaction, the user information is processed by the perception layer to get the initial single-mode user intention.

Among them, Kinect depth sensing device is used for gesture perception. Tactile sensing uses a thin-film pressure sensor, which presses the pressure sensor on the intelligent dropper to transmit the touch information to the system. Voice information is obtained by microphone. After preprocessing, the obtained modal information is matched with the existing database to get the sub intention contained in the current user information.

In addition, the user intention in virtual experiment usually contains an operation object Tag. For example, in the intention of "grabbing a test tube", the operation object Tag = testtube. Therefore, in the recognition process, the operation object should be one of the factors to understand the intention. In the virtual experiment, the collision between the virtual hand and the target is regarded as the selected target Tag.

$$Tag = \{tg_i | Min\left(D\left(Vr_{hand}, tg_i\right)\right), \quad Tou = 1\} \tag{3}$$

where $Tou = 1$ indicates that the virtual hand has collided with a target, and the Euclidean distance between the virtual hand and the potential target is calculated by function D

We define a 5-tuple to describe the user operation information $U_{info} = (Voi, Tac, Ges, Tag, Tim)$, where Voi is the user's and voice information, Tac is the tactile information, Ges is the visual information, Tag is the target information, and Tim is the time series of three modal information input.

When the user sends a command, the system receives the user information $U_{info} = (Voi, Tac, Ges, Tag)$. After matching with the sub intention database, a user sub intention triple$U_{inten} = (A_1, A_2, A_3)$ is obtained, and then input U_{inten} into the intention fusion model.

Where $A_1 = (Db\_Vi \cap Voi|Tag), A_2 = (Db\_Ta \cap Tac|Tag), A_3 = (Db\_Ge \cap Ges|Tag)$, the information of different modes is intersected with the corresponding database respectively. If it is not empty, it indicates that the sub intention of the current information expression exists. In addition, the operation target Tag may be empty. If it is empty, it means that the currently executed user command does not involve the experimental equipment in the virtual experiment.

3) MULTIMODAL INTENTION FUSION ALGORITHM

In the actual virtual chemical experiment system, the way that users send out an instruction is multimodal. During the period of instruction sending, the information of different modes is input into the system in different order, which indicates that users have different priority on mode selection, and also contains some potential information. For example, the mode that users choose first contains more real intention of users. Therefore, we take the time series $Time\_sq$ was included in the scope of identification.

The user intention triple U_{inten} and time series ime_{sq} are combined to form a quad $I\text{nput}_{\text{Data}} = \left(A_1, A_2, A_3, Time_{sq}\right)$, and then input it into the intention understanding model $F_1\left(A_1, A_2, A_3, Time_{sq}\right)$,

$$F\left(A_1, A_2, A_3, Time\_sq\right) = \begin{cases} (Tim\left[A_1\right] * A_1) \cap (Tim\left[A_2\right] * A_2) \cap (Tim\left[A_3\right] * A_3), \\ A_1 \neq \emptyset or A_2 \neq \emptyset or A_3 \neq \emptyset \\ A_1 + A_2 \cap A_3, \quad A_1 \neq \emptyset, \ Time\_sq[1] = 2 \end{cases} \tag{4}$$

where $Time\_sq[1]$=2 means that the first mode of input is speech, and the expression intention of speech is more direct than gesture and touch. In the system test of this article, the intention expressed by user voice is the closest to the real intention. Therefore, when the first input mode is voice, the user's intention is specified as speech sub intention. $Tim\left[A_i\right]$ indicates whether sub intention A_i exists or not.

$$Tim\left[A_i\right] = \begin{cases} 0, & i \notin Time\_sq \\ 1, & i \in Time\_sq \end{cases} \tag{5}$$

$Cur\_i = F_1\left(A_1, A_2, A_3, Time\_sq\right)$ to get the user's current intention is still not the final intention. In the virtual chemistry experiment, the user carries out the experiment according to certain steps, and there is a certain correlation between these steps.

Even if the degree of these associations is different, we think it is still necessary to consider the impact of the previous user's intention on the current user.

Therefore, we need to further fuse the previous user intention to infer the authenticity of the current intention.

According to the experience and the actual experimental steps, we set up the correlation matrix $\boldsymbol{\varepsilon} = \left\{\varepsilon_{i,j}\right\}, i, j \in N$, $\varepsilon_{i,j}$ between the two intentions. $\varepsilon_{i,j}$ represents the correlation degree between the two intentions i and j, and N represents the total number of user intentions in the virtual experiment. The ultimate intention can be expressed as $Final\_i$,

$$Final\_i = \begin{cases} Cur_i, & \varepsilon_{Cur\_i,La\_i} \geq \sigma \\ 0, & \varepsilon_{Cur\_i,La\_i} < \sigma \end{cases} \tag{6}$$

where σ is a threshold. When the correlation degree ε_{Cur_i,La_i} between the current intention and the current intention is greater than the threshold, the final intention is the current intention, otherwise the current intention is invalid. In order to accurately determine whether the processing result of current user intention is accurate, when $Final\_i = 0$, the system will ask the user again whether the current intention is feasible, and then obtain the user voice instruction V_{cmd}, if $V_{cmd} = Yes$, the final intention is still the current intention, otherwise it is empty, waiting for the user to re-enter the instruction.

The feedback of the system to the user's intention is presented in the virtual scene. After the user's final intention $Final\_i$ is confirmed, it is matched with the behavior database. If $Final\_i \cap Db\_Ac = \emptyset$, it prompts the user that the current intention cannot be implemented and waits for the user to re issue the instruction. Otherwise, the system will execute the operation and feed back to the user.

The algorithm of multimodal intention understanding is as follows. According to the algorithm, we can infer the user's interaction intention in the virtual experiment.

Algorithm 1 Multimodal Intention Understanding Algorithm

Input: $Voi, Tac, Ges, Tag, Time_{sq}, \boldsymbol{\varepsilon}, La_i, Vr_{hand}$
Output: $Final\_i$

1. According to the results of Vr_{hand} collision detection, the operation target Tag of current user intention is determined;
2. Determine the multimodal operation information tuple $U_{info} = (Voi, Tac, Ges, Tag, Tim)$ of the current user;
3. ***while***$(U_{info} = (Voi, Tac, Ges, Tag, Tim) \neq \emptyset)$
4. ***do***
5. By matching with the current single-mode intention databases, a new user intent triplet $U_{inten} = (A_1, A_2, A_3)$ is obtained. Combined with the time series $Time_{sq}$, the input information Quad $I\text{nput}_{\text{Data}} = \left(A_1, A_2, A_3, Time_{sq}\right)$ is formed;
6. Input the Input_Data into the multimodal fusion model $F\left(A_1, A_2, A_3, Time\_sq\right)$;
7. ***if*** $A_1 \neq \emptyset, Time_{sq[1]} = 2$ ***then***
8. $Cur_i = A_1 + A_2 \cap A_3$
9. ***Else*** $Cur_i = (Tim\left[A_1\right] * A_1) \cap (Tim\left[A_2\right] * A_2) \cap (Tim\left[A_3\right] * A_3)$
10. ***Endif***
11. The former user intention is combined with the current intention to determine the credibility of the current intention;
12. ***if*** $\varepsilon_{Cur\_i,La\_i} \geq \sigma$ ***then***
13. Determine the current intention as the final executable intention $Final\_i = Cur\_i$;
14. ***Else***
15. The current intention is judged to be unenforceable, and the voice asks the user whether to still carry out the current intention;
16. ***if*** $V_{cmd} = Yes$ ***then***
17. $Final\_i = Cur\_i$
18. ***Else*** The current intention is not executable, waiting for the user to re-enter the instruction;
19. ***Endif***
20. ***Endif***
21. Matching final intention $Final\_i$ and behavior database Db_{Ac};
22. ***if*** $Final\_i \cap Db\_Ac = \emptyset$ ***then***
23. The system does not perform the user operation and prompts the user to re-enter the command;
24. ***Else*** Execute the final intention and feedback user results;
25. **E*ndif***
26. ***Endwhile***

IV. DESIGN AND IMPLEMENTATION OF INTELLIGENT EXPERIMENT

A. BASIC SETTINGS

In order to test the performance of the algorithm in this article, the virtual experimental platform is implemented on the ordinary desktop computer with the configuration of inter i5-6500, 3.20ghz main frequency, 8GB memory and independent graphics card. The software development environment is unity3d 5.4, and Kinect V2 is used as video input device.

B. EXPERIMENT OF ACID AND ALKALI DETECTION

The user controls the virtual hand to grasp the dropper, and chooses to load phenolphthalein reagent or purple litmus reagent into the dropper. Then the user drops the liquid from the dropper into two beakers and observes the color change of the liquid in the beaker. In order to reduce the influence of irrelevant variables and increase the credibility of the experiment, we have verified that it is most appropriate for each participant to conduct six experiments.

1) EXPERIMENTAL PROCESS

The user sends out the voice command to "start the experiment". According to the experimental navigation, the user grabs the dropper through the combination of gesture and voice.

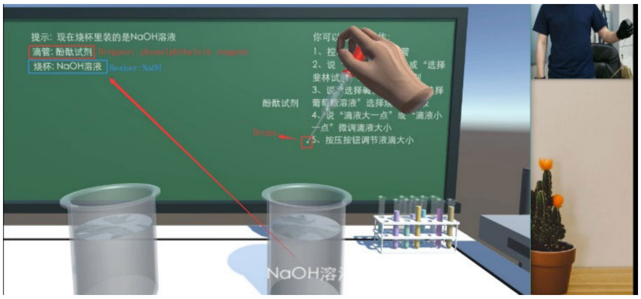


FIGURE 5. **Selecting reagents.**

According to the steps, the user selects the reagent in the dropper by voice, as shown in Fig. 5, the user selects phenolphthalein reagent. It can be seen that the user's intention result "dropper: phenolphthalein reagent" is displayed in the yellow box of the virtual blackboard. Then, the user continues to select the solution to be verified. When the user expresses his choice of alkaline solution, the yellow box position of the virtual blackboard displays the current user intention result "beaker: NaOH solution".

FIGURE 6. **Intention feedback.**

Next, the user drops the reagent into the beaker according to the prompt (Fig. 6). At this time, the system will obtain the user's voice information, gesture information and touch information, and infer the user's intention according to the multimodal fusion algorithm. The user sends out the intention of "increasing the amount of single drop added" according to the needs. It can be seen that the volume of the drop increases, and the amount of single addition increases correspondingly. In order to speed up the experiment, the user adjusts the drop speed. After the instruction is given, the system recognizes the user's intention and gives a voice feedback to the user that "the dropping speed of the drop is accelerated".

During the operation, the user mistakenly drops the drops onto the desktop. You can see that "you dropped the drops outside the beaker" at the prompt position. At the same time, the voice will further remind the user of the operation error (Fig. 7).

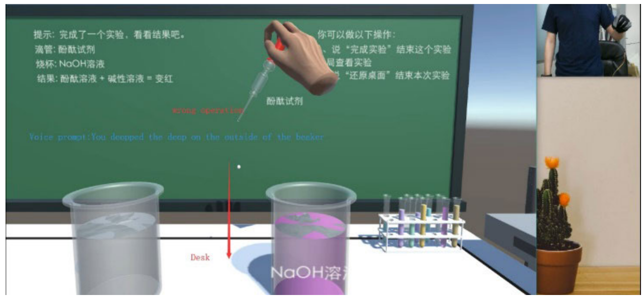


FIGURE 7. **Fig. 7. Error operation feedback.**

In the process of experiment, users interact with experimental equipment with virtual hand, and need to grasp different devices. In order to achieve more realistic interaction effect, different grabbing objects have different grasping effects, which makes the virtual experiment closer to the real interaction (Fig. 8).

(a) (b) (c)

FIGURE 8. **Grasping experimental equipment.**

When the user drops phenolphthalein reagent into the beaker, it can be observed that the color in the beaker is a dynamic change process (Fig. 9). This is to enhance the experimental effect and experience, and enhance the user's impression of the experiment, so as to facilitate the user to understand the experiment.

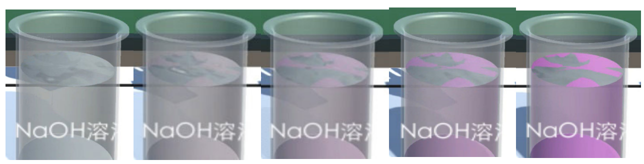


FIGURE 9. **Experimental phenomena.**

After the final experimental results are generated, the virtual blackboard shows the relevant knowledge points (Fig. 10), and prompts the user to conduct the experiment again by instructing "clean the desktop" and "start the experiment".

2) EXPERIMENTAL TASK

In order to evaluate the effectiveness of the model and the designed intelligent dropper in improving the teaching ability of virtual experiment, and to compare the performance of

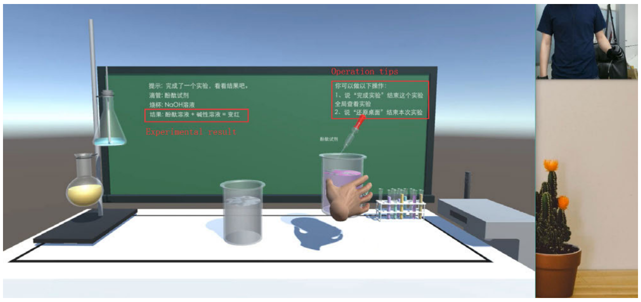


FIGURE 10. Experimental result.

other virtual experimental systems under different experimental tasks. We invited 40 volunteers to participate in the experiment, including 10 middle school teachers and 30 middle school students, with a male to female ratio of 1:1. The students were between 8 and 16 years old, and none of them had ever used the virtual experimental system.

The experiment can be divided into two parts: 1. Experiment on MIIVEP; 2. Experimental verification on SIVE, OpenGL [2], NOBOOK [38] and real experiment.

3) RESULT ANALYSIS

a: VERIFICATION OF MIIVEP

In the experiment, all participants used at least two modes of interaction, and most participants chose three modes of interaction. Especially in the interaction related to the depth of virtual space, users prefer to express their intention by voice rather than visual gesture. Participants indicated that visual based interaction was too laborious to control objects in the virtual space, and the voice expression intention was more direct, which indicated that providing alternative interaction modes could effectively reduce the interaction load of users. The increase of the types of interaction modes will lead to the diversification of the expression of the same user's intention. Although it can meet the needs of the diversity of intention expression and reduce the interaction load, it may cause the user's intention difficult to express correctly and increase the user's use difficulty.

We asked 20 participants to carry out six acid-base detection experiments on MIIVEP, and counted the success rate of each user's intention execution.

These intentions are as follows: I-T11: holding the dropper, I-T12: selecting the reagent in the dropper, I-T13: adjusting the dropping speed of the drop, I-T14: dropping reagent into the beaker, I-T15: changing the acidity and alkalinity of the solution in the beaker, and I-T16: adjusting the single dosage of the drop.

As shown in TABLE 3, the success rate of intention expression in the first experiment was not high. It is not difficult to infer that this is due to the inadaptability and diversity of intention expression caused by the first use of virtual experimental system. After many experiments, the participants gradually adapted to the use of the system and the expression of intention, and the success rate of intention expression began to increase rapidly (from the first to the third

TABLE 3. Intention execution success rate.

Succeed Rate (20 people)						
Times	1	2	3	4	5	6
I-T11	20.41	37.03	55.56	83.33	86.96	100
I-T12	17.24	32.79	50.00	76.92	95.24	100
I-T13	18.52	31.75	48.78	76.92	95.24	95.24
I-T14	21.73	33.90	62.50	80.00	100	100
I-T15	24.69	41.67	64.52	83.33	100	100
I-T16	25.64	44.45	66.67	86.96	100	100

experiment, the average increase was 47.69%). In the fourth experiment, the rising speed of the participants' intention expression success rate began to slow down. In the fifth and sixth experiments, participants had basically mastered the use of virtual experimental system, and the average success rate of each intention expression reached 99.2%.

Then, we asked the experimenters to carry out six acid-base detection experiments on the virtual experiment system (SIVE) based on single-mode interaction (only gesture), and counted the sixth intention success rate. The results are shown in TABLE 4.

TABLE 4. Comparison results.

Intention	I-T11	I-T12	I-T13	I-T14	I-T15	I-T16
MIIVEP	100	100	100	95.24	100	100
SIVE	83.33	76.92	80	76.92	83.33	86.96

As shown in the table, the success rate of user intention expression in MIIVEP (99.2%) is better than that in SIVE (81.24%). Through the analysis of the two tables, we can see: on the one hand, the integration of multimodal information helps to solve the incompleteness of user's expression information and enhance the expression of intention. On the other hand, multimodal interaction will increase the user's previous use difficulty and reduce the success rate of operation. However, the results also show that users can eliminate this effect with less use times. In a word, the multi-modal understanding algorithm proposed in this article can effectively enhance the ability of the system to understand the user's intention, thus improving the experimental efficiency and interactive experience. Fig. 11 shows the students who are experimenting on MIIVEP.

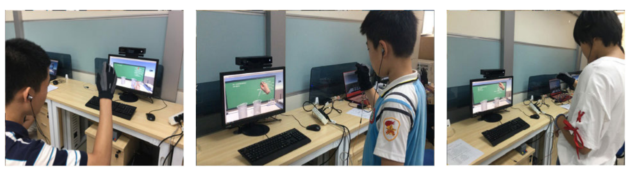


FIGURE 11. Middle school teachers and students participate in experimental verification.

b: VERIFICATION OF SMART DEVICES

We mainly verify the design cost (DC), display effect (DE), functional integrity (FP), operational authenticity (OA) and low load (LL) of the intelligent dropper. Therefore, it is more reasonable to evaluate these items by means of user experience survey. First, we invited 10 middle school students to carry out real acid-base detection experiments. According to the use of droppers in real experiments, we scored each item of intelligent dropper. The higher the score is, the closer the performance is to the real dropper.

As shown in TABLE 5, in the aspect of effect display, participants thought that the intelligent dropper has achieved good results, but there is still a certain gap between the effect and the real experiment. For the function of the dropper, the intelligent dropper not only has the ability of the real dropper, but also adds additional functions with strong practicability. Therefore, most participants gave a high score. A few participants thought that the additional functions were not provided by the real dropper, and they did not agree with such addition. In terms of operation authenticity, the operation experience of the intelligent dropper is very close to that of the real dropper, with a score of 8.48.

In terms of low load, the glove wearing mode of intelligent dropper is almost the same as the interactive load of real operation. Obviously, the participants have a high acceptance of this method. It is worth mentioning that the design cost is low. Even if a complete set of ordinary experimental equipment is purchased, it only costs about 29rmb, while the design cost of intelligent dropper is about 79.65 RMB, which is far higher than the price of a set of real experimental equipment. However, the value of virtual experiment lies in saving experimental materials and infinite repeated experiments. Compared with the cumulative cost of real experiment, the cost of designing intelligent dropper is obviously worthwhile, so participants still give an average score of 8.0. Finally, the participants generally agreed that the intelligent dropper designed by us was practical and interesting, and strongly supported us to continue to improve the equipment and continue to design new intelligent equipment (TABLE 5).

TABLE 5. Comparison results.

Index	DC	DE	FP	OA	LL
Avg(score)	8.0	8.2	9.5	8.48	9.2

C. COMPARATIVE EXPERIMENT

1) TIME EFFICIENCY ANALYSIS

In order to better verify the effectiveness of the proposed framework, we compared the experimental time with NOBOOK [38] and OpenGL [2]. At the same time, in order to avoid the influence of different systems, we also implemented single-mode (gesture) interaction (SIVE) on the virtual experimental system established in this article.

As shown in Fig. 12, a is the experimental platform and demonstration of MIIVEP designed in this article, and b represents the NOBOOK experimental platform. b uses the mouse interaction mode, through the mouse can quickly select reagents, and get the results. The multimodal interaction in this article provides rich interaction modes, and users can issue instructions in different ways. Compared with the two methods, it is not difficult to see that mouse interaction is simple and fast, but it lacks interaction and can't exercise the user's experimental ability and provide real operation experience. The purpose of multimodal interaction is to improve the practical ability of users, increase the interest of experiments, stimulate the learning enthusiasm of users, and ultimately achieve the purpose of improving the learning efficiency of users and the popularization ability of virtual teaching.

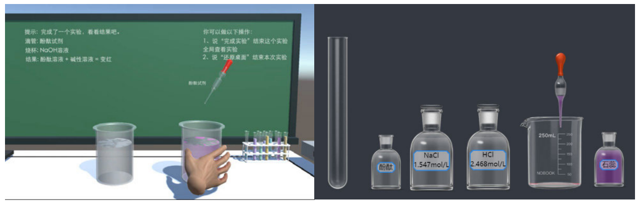


FIGURE 12. Experimental platform.

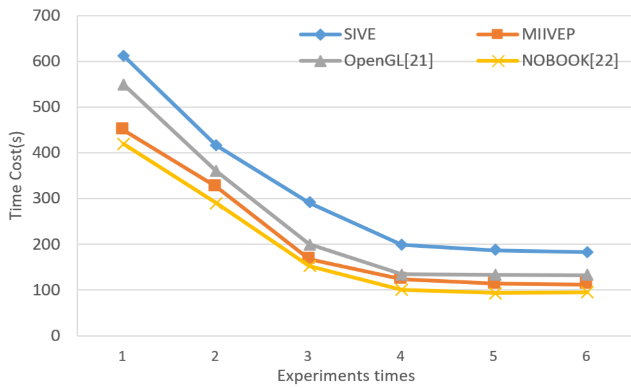


FIGURE 13. Comparison of experimental time spent.

A total of 20 participants were invited to carry out acid-base verification experiments on different experimental platforms. In Fig. 13, the total experimental time spent in completing the acid-base detection experiment under different methods was recorded in Fig. 13, including the system time, trial and error time, exploration time and understanding time spent by users in the experiment process.

The results showed that with the increase of the number of experiments, the experimental time gradually decreased, and in the last three experiments, the experimental time gradually stabilized. Analysis of the data shows that mouse interaction has a greater advantage in reducing time cost. However, with the increase of the number of experiments, the time cost of MIIVEP is gradually approaching that of NOBOOK and OpenGL based on mouse interaction, which shows that MIIVEP can also approach mouse interaction in terms of time consumption while providing more natural interactive means. In addition, the analysis of the time cost of SIVE and MIIVEP shows that MIIVEP takes much less time to complete the experiment than SIVE. According to the feedback from the experimenters, this is because multimodal interaction can better meet the interaction habits of different users. Each user can interact smoothly according to his own habits, while SIVE only provides a unified interaction mode, and users need to spend time to get used to and memorize.

In order to verify the effect of MIIVEP on different experiments, we continued to invite 20 participants to carry out silicic acid preparation experiment (see section 4.5.1 for the experimental process), and counted the time spent in the last three experiments (TABLE 6). We only count the time spent in the last three experiments and the final average time spent. All the data only keep the integer part.

TABLE 6. Time spent in preparation of silicic acid.

Time Cost(s)	SIVE	MIIVEP	OpenGL[2]	NOBOOK[38]
4^{th}	512	344	314	261
5^{th}	496	336	308	256
6^{th}	507	339	311	253
AVG	456	340	311	256

2) ANALYSIS OF LEARNING EFFECT

We divided 40 participants into two groups and completed three acid-base tests under MIIVEP and NOBOOK. And the two groups were tested in class, and the score distribution of each group was counted. In order to facilitate statistical analysis, we divide students' scores into four grades: a (score >= 90), level B (score >= 80), level C (score >= 60), and level D (score < 60). The score statistics are shown in Fig. 14.

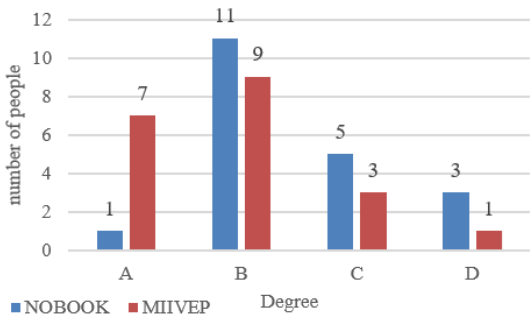


FIGURE 14. Distribution of test scores.

The results showed that among the 20 participants using MIIVEP, the A-level rate was 35%, the (A+B) rate was 80%, and the failure rate was only 5%. Among the 20 participants who used NOBOOK, the A-grade rate was only 5%, the (A+B) rate was 60%, and the unqualified rate was 15%. This verifies the ability of MIIVEP in enhancing students' learning effect. According to the feedback from participants, the interaction mode provided by MIIVEP is more realistic and intuitive, focusing on the experimental effect and interactive experience, while NOBOOK ignores these aspects. In addition, members of the MIIVEP team said that the intelligent dropper made the experimental process more interesting and gave them a clearer understanding of the intelligent dropper.

3) EVALUATION SCALE OF VIRTUAL EXPERIMENT SYSTEM

It is found that the performance of virtual experiment itself is the key to affect teachers' acceptance of this teaching method. According to the teachers' requirements for the performance of virtual experiment, we have developed a virtual experiment system evaluation standard ESVES. It includes authenticity (A): whether the interaction process is realistic or not, and whether the effect is realistic. Safety (S): whether the experimental process is safe. Explorability (E): whether the experiment can support the operation beyond the preset, like the actual experiment, there can be inquiry operation. Repeatability (R): whether the experiment can be repeated many times. Device cognition (DA): enhance users' cognition of experimental equipment.

Intelligence (I): whether the experiment interaction process can accurately analyze the user's intention and make reasonable response. A total of six indicators, the full score of each index is 100. We compared the ESVES of MIIVEP, OpenGL and NOBOOK (Fig. 15).

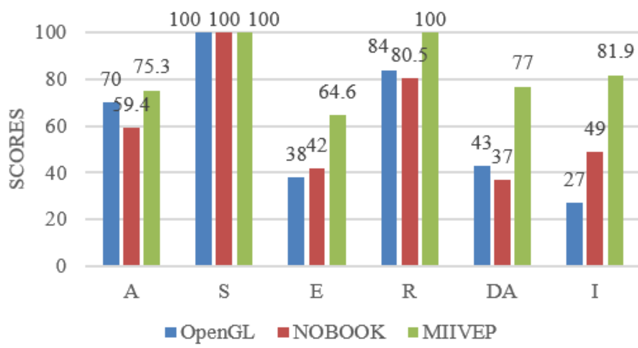


FIGURE 15. ESVES.

According to the score results of 40 participants, the safety ability of each virtual experimental system was recognized. In terms of authenticity, participants believe that the main difference lies in the interaction mode and effect presentation, such as the intelligent dropper in this article, which improves the operation authenticity of experimental equipment. In general, the scores of MIIVEP on ESVES are higher than those of OpenGL and NOBOOK, which proves that MIIVEP is more in line with the needs of users than other systems and is recognized by users.

Furthermore, we analyzed the ESVES data of NOBOOK and MIIVEP by ANOVA (TABLE 7). The results of six indicators are as follows: SS is the sum of squares of deviation, DF is the degree of freedom, MS is the mean square value, F is the test value, P-value is the adjoint probability, and F CRIT is the critical value of quantile value of F distribution at significance level of 0.05.

TABLE 7. Time spent in preparation of silicic acid.

Evaluating indicator	Interactive methods	SS	df	MS	F	P-value	F CRIT
A	NOBOOK	5056.2	1	5056.2	468.3891	1.04E-34	3.963472
	MIIVEP						
E	NOBOOK	10215.2	1	10215.2	912.0714	8.56E-45	3.963472
	MIIVEP						
R	NOBOOK	7605	1	7605	941.5714	2.72E-45	3.963472
	MIIVEP						
DA	NOBOOK	32000	1	32000	5804.651	5.44E-75	3.963472
	MIIVEP						
I	NOBOOK	21648.2	1	21648.2	1748.715	3.55E-55	3.963472
	MIIVEP						

In Table 7, there is a lack of analysis on the safety index (s), because the virtual teaching for users, the degree of risk is almost no, and users also give consistent evaluation. According to the analysis of the other five indicators, F value is greater than F CRIT value, which indicates that the system performance of MIIVEP based on multimodality is significantly different from that based on mouse NOBOOK, which verifies the effectiveness and feasibility of multimodal intention understanding model in improving learning efficiency and system performance, and reducing interaction load.

D. SYSTEM USABILITY SCALE

Subsequently, we invited all participants to complete a sus questionnaire [39], which was designed to investigate the user's evaluation of MIIVEP. The questionnaire contained 10 questions, and the evaluation of each question ranged from very agree to very disagree, with a total of 5 grades. The users chose different levels according to the content of the questions to express their attitude towards MIIVEP. Referring to the original SUS, we developed a sus questionnaire suitable for this article, as shown in TABLE 8.

Sus is widely used in the system usability questionnaire survey. Its question design has good objectivity, and it is easy to quantify the score and convert it into a percentage system. After research, Sus can get a real evaluation of the system in no more than 15 samples. All the scales have considerable sensitivity, and the number of samples in this article is 40, which is more helpful to evaluate the usability of the system using this algorithm. The SUS score of MIIVEP is 80 points. According to Bangor's interpretation of SUS [40], MIIVEP's grade is B among the seven grades. This shows that MIIVEP has good usability and easy to learn, but there are still some deficiencies in some aspects.

E. PREPARATION EXPERIMENT OF SILICIC ACID

In order to verify the universality and inclusiveness of the framework and algorithm in this article, and to verify the function of intelligent dropper. We used MIIVEP to carry out a relatively complex experiment on the preparation of silicic acid

TABLE 8. The MIIVEP of the System Usability Scale (SUS), showing the minor modifications to the original Brookes instrument.

Order	Questions
1	I think that I would like to use MIIVEP frequently
2	I found the MIIVEP unnecessarily complex
3	I thought the MIIVEP was easy to use
4	I think that I would need the support of a technical person to be able to use MIIVEP
5	I found the various functions in MIIVEP were well integrated
6	I thought there was too much inconsistency in MIIVEP
7	I would imagine that most people would learn to use MIIVEP very quickly
8	I found the MIIVEP very cumbersome to use
9	I felt very confident using the MIIVEP
10	I needed to learn a lot of things before I could get going with MIIVEP

1) EXPERIMENTAL PROCESS

The process of silicic acid preparation experiment is relatively complex, but its application on MIIVEP is similar to that of acid-base test. Fig. 16 shows the process of silicic acid preparation experiment.

V. SUMMARY

This article in view of the existing virtual experiment system are widespread the interaction of a single, the actual teaching effect is poor, complex user intent and so on, puts forward a kind of information based on multimodal interaction model, the model integration modal information, such as voice, gestures, touch different modal analysis, the relationship between the subset can be divided into different intentions,

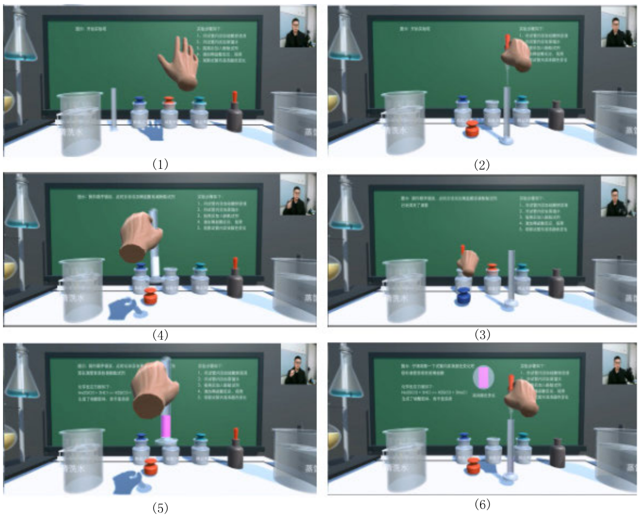


FIGURE 16. Experiment T2: preparation of silicic acid.

and will build the user intent set intentions subset fusion, and formed a set with the matching operation behavior, has been clear about the different user input the mapping relationship between information and user intent, and the mapping relationship between user intent and operation behavior. Then, based on the model, the multi-mode intelligent sensing system is realized. At the same time, an intelligent dropper based on the fusion of virtual and real is proposed, which simulates the real dropper on the basis of the user's understanding of intention, and finally realizes the virtual teaching environment of natural interaction and intelligent teaching. The main contribution of this article lies in that it not only solves some bottleneck problems and key problems that hinder the popularization of virtual experiment teaching in middle schools, but also enhances the intelligence and feasibility of virtual experiment system. On the other hand, the design of virtual and real fusion of intelligent devices, through virtual and real interaction, to meet the experimental interaction experience and interaction authenticity needs, so that users can quickly grasp the experimental content.

However, the model proposed in this article still has some shortcomings: on the one hand, the mode type limits the wider interaction behavior. For example, in the experiment of producing gas, the user cannot judge the category of gas by smell and the reaction degree of experiment, which hinders the user's understanding of relevant knowledge to some extent. On the other hand, more convenient and effective intelligent devices should be realized. The integration of virtual and real devices is too single, which limits the generalization ability of the virtual experimental system to some extent. More intelligent devices can be designed, such as intelligent beakers and intelligent liquid separation funnels. To solve these problems is the focus of the future research, but also to further enhance the portability of the model, so that it can have good performance and robustness in different types of virtual experimental systems.

REFERENCES

[1] H. B. Wang, M. Yang, and Y. Q. Ye, "ModelChemlab: A software for simulated chemistry experiments," *Comput. Appl. Chem.*, vol. 18, no. 1, pp. 95–96, 2001.

[2] F. Q. Zhu, L. P. Yang, and C. G. Li, "Research on virtual chemical experiment system based on OpenGL," *Comput. Eng. Des.*, vol. 29, no. 3, pp. 723–772, 2008.

[3] J. Lee, H. Gu, H. Kim, J. Kim, H. Kim, and H. Kim, "Interactive manipulation of 3D objects using kinect for visualization tools in education," in *Proc. 13th Int. Conf. Control, Autom. Syst. (ICCAS)*, Oct. 2013, pp. 1220–1222.

[4] D. G. Hawkins, "Virtual reality and passive simulators: The future of fun," in *Communication in the Age of Virtual Reality*. 1995, pp. 159–189.

[5] R. I. Garcia-Betances, V. Jiménez-Mixco, M. T. Arredondo, and M. F. Cabrera-Umpiérrez, "Using virtual reality for cognitive training of the elderly," *Amer. J. Alzheimer's Disease Other Dementias*, vol. 30, no. 1, pp. 49–54, 2015.

[6] A. Rizzo, M. J. Roy, A. Hartholt, M. Costanzo, K. B. Highland, T. Jovanovic, S. D. Norrholm, C. Reist, B. Rothbaum, and J. Difede "Virtual reality applications for the assessment and treatment of PTSD," in *Handbook of Military Psychology*. Cham, Switzerland: Springer, 2017, pp. 453–471.

[7] D. Fernandez-Aviles, D. Dotor, D. Contreras, and J. C. Salazar, "Virtual labs: A new tool in the education: Experience of technical University of Madrid," in *Proc. 13th Int. Conf. Remote Eng. Virtual Instrum. (REV)*, Feb. 2016, p. 272, doi: 10.1109/REV.2016.7444480.

[8] T. Wolf, "Assessing student learning in a virtual laboratory environment," *IEEE Trans. Educ.*, vol. 53, no. 2, pp. 216–222, May 2010, doi: 10.1109/te.2008.2012114.

[9] T. Lynch and I. Ghergulescu, "Newton virtual labs: Introduction and teacher perspective," in *Proc. IEEE 17th Int. Conf. Adv. Learn. Technol. (ICALT)*, Jul. 2017, pp. 343–345, doi: 10.1109/ICALT.2017.133.

[10] N. R. Herga, B. Čagran, and D. Dinevski, "Virtual laboratory in the role of dynamic visualisation for better understanding of chemistry in primary school," *EURASIA J. Math., Sci. Technol. Edu.*, vol. 12, no. 3, pp. 593–698, Jul. 2016, doi: 10.12973/eurasia.2016.1224a.

[11] S. M. Baladoh, A. F. Elgamal, and H. A. Abas, "Virtual lab to develop achievement in electronic circuits for hearing-impaired students," *Edu. Inf. Technol.*, vol. 22, no. 5, pp. 2071–2085, Sep. 2017, doi: 10.1007/s10639-016-9532-7.

[12] L. Rajendran, V, Ramachandran, and J. Divya, "A study on the effectiveness of virtual lab in E-learning," *Int. J. Comput. Sci. Eng.*, vol. 2, no. 6, 2010.

[13] D. Shin, S. Yoon, K. Y. Lee, and E. S. Lee, "A Web-based, interactive virtual laboratory system for unit operations and process systems engineering education: Issues, design and implementation," *Comput. Chem. Eng.*, vol. 26, no. 2, pp. 319–330, 2010, doi: 10.1016/s0098-1354(01)00749-9.

[14] F. L. Naranjo, G. Martínez, Á. L. Pérez, and P. J. Pardo, "Design, development, testing and validation of a photonics virtual laboratory for the study of LEDs," *Educ. Training Opt. Photon.*, Jul. 2013, doi: 10.1353/ect.2014.0057.

[15] M. Duarte, A. Mahalingam, and B. P. Butz, "The intelligent universal virtual laboratory (UVL)," in *Proc. Frontiers Edu. 35th Annu. Conf.*, Oct. 2005, doi: 10.1109/fie.2005.1611987.

[16] D. Kim, K. Choi, C. Jeon, J. Lim, S. Kim, S. Seo, and J. Yoo, "A Web-based virtual laboratory system for electronic and digital circuits experiments," in *Hybrid Learning and Education*. Berlin, Germany: Springer, 2009, doi: 10.1007/978-3-642-03697-2_8.

[17] I. Gustavsson, "A remote access laboratory for electrical circuit experiments," *Int. J. Eng. Edu.*, vol. 19, no. 3, pp. 409–419, 2013.

[18] H. Liao and Z. Qu, "Virtual experiment system for electrician training based on Kinect and Unity3D," in *Proc. Int. Conf. Mech. Sci., Electr. Eng. Comput. (MEC)*, Dec. 2013, pp. 2659–2662, doi: 10.1109/MEC.2013.6885480.

[19] P. K. Atrey, M. A. Hossain, A. El Saddik, and M. S. Kankanhalli, "Multimodal fusion for multimedia analysis: A survey," *Multimedia Syst.*, vol. 16, no. 6, pp. 345–379, Nov. 2010.

[20] D. L. Hall and J. Llinas, "An introduction to multisensor data fusion," *Proc. IEEE*, vol. 85, no. 1, pp. 6–23, Jan. 1997.

[21] C. G. M. Snoek, M. Worring, and A. W. M. Smeulders, "Early versus late fusion in semantic video analysis," in *Proc. 13th Annu. ACM Int. Conf. Multimedia (MULTIMEDIA)*, Singapore, 2005, pp. 399–402.

[22] M.-T. Yang, S.-C. Wang, and Y.-Y. Lin, "A multimodal fusion system for people detection and tracking," *Int. J. Imag. Syst. Technol.*, vol. 15, no. 2, pp. 131–142, 2005.
[23] C. Neti, B. Maison, A. Senior, G. Iyengar, P. Cuetos, S. Basu, and A. Verma, "Joint processing of audio and visual information for multimedia indexing and human-computer interaction," in *Proc. Int. Conf. RIAO*, Paris, France, 2000, pp. 294–301.
[24] P. K. Atrey, M. S. Kankanhalli, and J. B. Oommen, "Goal-oriented optimal subset selection of correlated multimedia streams," *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 3, no. 1, p. 2, Feb. 2007.
[25] P. K. Atrey, M. S. Kankanhalli, and A. El Saddik, "Confidence building among correlated streams in multimedia surveillance systems," in *Proc. Int. Conf. Multimedia Modeling*, Singapore, 2007, pp. 155–164.
[26] R. E. Kalman, "A new approach to linear filtering and prediction problems," *J. Basic Eng.*, vol. 82, no. 1, pp. 35–45, Mar. 1960.
[27] R. C. Luo, C. C. Yih, and K. L. Su, "Multisensor fusion and integration: Approaches, applications, and future research directions," *IEEE Sensors J.*, vol. 2, no. 2, pp. 107–119, Apr. 2002.
[28] V. Pitsikalis, A. Katsamanis, G. Papandreou, and P. Maragos, "Adaptive multimodal fusion by uncertainty compensation," in *Proc. 9th Int. Conf. Spoken Lang. Process*. Pittsburgh, PA, USA, 2006, pp. 1–4.
[29] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining Knowl. Discovery*, vol. 2, no. 2, pp. 121–167, 1998.
[30] H. Bredin and G. Chollet, "Audio-visual speech synchrony measure for talking-face identity verification," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Paris, France, vol. 2, Apr. 2007, pp. 233–236.
[31] T. Choudhury, J. M. Rehg, V. Pavlovic, and A. Pentland, "Boosting and structure learning in dynamic Bayesian networks for audio-visual speaker detection," in *Proc. Object Recognit. Supported Interact. Service Robots*, Quebec City, QC, Canada, vol. 3, Aug. 2002, pp. 789–794.
[32] H. Wu, "Sensor data fusion for context-aware computing using Dempster-Shafer theory," Ph.D. dissertation, Robotics Inst., Carnegie Mellon Univ., Pittsburgh, PA, USA, 2003.
[33] H. Li, P. Zhang, S. Al Moubayed, S. N. Patel, and A. P. Sample, "ID-match: A hybrid computer vision and RFID system for recognizing individuals in groups," in *Proc. CHI Conf. Extended Abstr. Hum. Factors Comput. Syst. (CHI EA)*, 2016, pp. 4933–4944.
[34] G. F. Meyer, J. B. Mulligan, and S. M. Wuerger, "Continuous audio–visual digit recognition using N-best decision fusion," *Inf. Fusion*, vol. 5, no. 2, pp. 91–101, Jun. 2004.
[35] Y. Ding and G. Fan, "Segmental hidden Markov models for view-based sport video analysis," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Minneapolis, MN, USA, Jun. 2007, pp. 1–8.
[36] G. Du, P. Zhang, and D. Li, "Human–manipulator interface based on multisensory process via Kalman filters," *IEEE Trans. Ind. Electron.*, vol. 61, no. 10, pp. 5411–5418, Oct. 2014, doi: 10.1109/TIE.2014.2301728.
[37] G. Du, P. Zhang, and X. Liu, "Markerless human–manipulator interface using leap motion with interval Kalman filter and improved particle filter," *IEEE Trans. Ind. Informat.*, vol. 12, no. 2, pp. 694–704, Apr. 2016, doi: 10.1109/TII.2016.2526674.
[38] P. R. Cohen and D. McGee, "Tangible multimodal interfaces for safety-critical applications," *Commun. ACM*, vol. 47, no. 1, pp. 1–46, 2004.
[39] J. Brooke, "SUS-A quick and dirty usability scale," *Usability Eval. Ind.*, vol. 189, no. 194, pp. 4–7, 1996.
[40] A. Bangor, P. Kortum, and J. Miller, "Determining what individual SUS scores mean: Adding an adjective rating scale," *J. Usability Stud.*, vol. 4, no. 3, pp. 114–123, 2009.

ZHIQUAN FENG (Member, IEEE) received the master's degree from Northwestern Polytechnical University, China, in 1995, and the Ph.D. degree from the Computer Science and Engineering Department, Shandong University, in 2006. He is currently a Professor with the School of Information Science and Engineering, University of Jinan. He has published more than 50 articles on international journals, national journals, and conferences in recent years. His research interests include human hand tracking/recognition/interaction, virtual reality, human–computer interaction, and image processing.

TAO XU received the M.S. degree in computer science and application from the China University of Petroleum, in 2007, and the Ph.D. degree from the Laboratory of Intelligent Recognition and Image Processing, School of Computer Science and Engineering, Beihang University. He is currently a Lecturer with the School of Information Science and Engineering, University of Jinan. His research interests include gesture recognition, image processing, and human–computer interaction.

MENGTING XIAO is currently pursuing the master's degree with the School of Information Science and Engineering, University of Jinan. Her main research interest includes human–computer collaboration and interaction.

BOTAO ZENG is currently pursuing the master's degree with the School of Information Science and Engineering, University of Jinan. His main research interest includes human–computer collaboration and interaction.

RUI HAN is currently pursuing the master's degree with the School of Information Science and Engineering, University of Jinan. Her research interests include human–computer interaction and virtual reality.