An Application of Somatosensory Interaction for 3D Virtual Experiments

Si Chen[1(✉)], Chenqing Wang[2], and Jianping Huang[3]

[1] Fuzhou University of International Studies and Trade,
Fuzhou 350202, Fujian, China
`731553247@qq.com`
[2] Fuzhou Institute of Technology, Fuzhou 350506, Fujian, China
[3] Graduate School of Creative Industry Design,
National Taiwan University of Arts, New Taipei City, Taiwan

Abstract. Interaction based on natural limb movements is the so-called Somatosensory Interaction, which has multi-faceted and continuous features at the information transmission level. The dependence of the keyboard will be significantly weakened, and the learning subject will pay more attention to the corresponding learning tasks in this interactive mode, so that the learning comfort can be significantly improved in the three-dimensional virtual system. Situated in such technical context, more and more teachers and students are actively applying virtual experiment technology in their teaching and learning environment. human-computer interaction mode can directly affect the user's sense of sensation and immersion. In recent years, with the continuous development of digital media technology, the use of Somatosensory Interaction design and three-dimensional virtual experiment has practical improved the teaching effect and promoted transformation of teaching interaction mode. It not only saves a lot of animation problems involved in traditional virtual experiments, but also save costs significantly, and provide relevant innovation direction for the in-depth application of virtual experimental technology.

Keywords: Somatosensory Interaction design · 3D virtual experiment · Kinect

1 Preface

With the rapid development of information technology, the original two-dimensional plane has developed into a three-dimensional system, and further developed to virtual and augmented reality. In this technical context, more and more teachers and students are actively applying virtual experiment technology. The use of multimedia simulation and virtual reality (VR) and other related technologies to conduct analog simulation of traditional experiments through computer is the so-called virtual experiment. The technology of interacting with computer by means of a mouse, a remote control and other devices and through user gestures, body language, etc., is a so-called somatosensory interaction. Taking Kinect as an example, Kinect can capture and track user gestures and actions in real time. The main recognition modules of Kinect are infrared cameras, cameras, microphone arrays and color cameras [1]. Kinect emits

infrared rays by an infrared emitter. After these infrared rays are reflected by the human body, a random reflection spots are captured and analyzed by the infrared camera, and then the depth image of the human body and the object in the visible region can be created, thus achieving the recognition of the human movement. The color camera can capture the object in the visible region and correct the human body movement. The function of the microphone array is to collect the sound, and provide noise filtering and sound source localization functions to realize language recognition. At present, the role of virtual experiment is becoming increasingly significant in the process of teaching and research. However, the application of this technology has certain limitations in some aspects. For example, the truthfulness of the simulation is insufficient, the interactive operation is not user-friendly, and the virtual experimental materials are not readily available, and so on. This will inevitably make it difficult for teachers and students to immerse themselves in the experiment, and the related effects will be greatly reduced. However, with the combination of somatosensory interaction and the virtual experiment, and through the somatosensory interaction mode, the students can directly use their hands to manipulate the corresponding virtual instrument, which can improve the truthfulness of the virtual experiment, and also exercise students' flexibility in operation.

2   Principle of Somatosensory Interaction Design

The basis of somatosensory interaction is the real and natural interaction, while the core is the natural human-computer interaction technology. In this interactive mode, a high-simulation learning environment can be constructed by means of a three-dimensional virtual learning system, so that the corresponding situations can be easily created in traditional classroom, and thus the learning subject can learn through the problem situation. Generally, the main principles adopted in the design of the interactive technology are: first, consistency. That is, the design style of the relevant somatosensory interaction should be consistent, and the knowledge and concepts conveyed need to be consistent. In the specific design process, the designer should ensure the internal unity, so that the operator can understand thoroughly in the use of interaction mode, and the efficiency of learning the related interaction method will be significantly improved. Second, immediate feedback. This principle requires the device to feedback the relevant information involved in each step of the operation in real time, so that the user can quickly understand the results of his or her action, thereby the user interaction can be standardized and the learning effect can be improved. And the more complete the feedback information, the better the effect than the local feedback. Third, active reflections and guidance to migration of knowledge and techniques. Somatosensory interaction mainly covers three levels: reflection, action and senses. In the specific design process, it is necessary to pay attention to the psychological experience of the reflective level, strengthen the real problem situation design, guide the operator to self-think and explore, and thus realizing the migration of knowledge and skills.

3 Motion Capture Algorithm Design and Route Development Based on Somatosensory Interaction

3.1 The Working Principle of Kinect

Under this somatosensory interaction technology, the camera plays an important role in capturing the operator's limb movements, which can then be identified, stored, and processed using a computer program. The Kinect camera also captures the operator's gestures and then uses the background program to convert these gestures into manipulation commands. The somatosensory interaction technology discriminates and captures the operator's gestures, mainly using the camera and the Prime Sense system [2]. The captured dynamic video is then compared to the already stored mannequin. If the two human models are identical, the corresponding gestures can be created into a skeleton model that is consistent with them. With the support of this model, it is possible to identify up to 25 key parts of the human body. At present, based on this, the identification technology of human sitting and standing posture is further introduced [3] With the depth of field and RGB camera, the Kinect somatosensory interaction technology can project the recognized 3D object onto the corresponding display screen.

3.2 Human Motion Capture Algorithm

The motion capture implementation mechanism involved in Kinect somatosensory interaction technology is as follows: First, the infrared emitter will emit corresponding infrared light when the object is encountered, and then the light will be captured by the infrared camera integrated in the somatosensory interaction device, so that it can form the corresponding mathematical model. See Fig. 1 below for details. The purpose of constructing this model is to obtain the corresponding Zo, which is to calculate the distance between the object and the infrared emitter.

In Fig. 1, the focus of the infrared camera and the infrared emitter correspond to point C and point L, respectively, and perpendicular to the image depth direction, corresponding to the Z axis, and the farther apart from the camera, the corresponding Z axis value is larger. The X and Z axes are perpendicular to each other, which represent the baseline of the infrared camera and emitter, and their baseline distance is represented by b. It is assumed that there is an O reference plane in the specific space, and the distance between it and the infrared camera in the Z- axis is represented by Z_0, and that the object surface in the space is represented by K, and its distance from the infrared camera in the Z- axis is represented by Z_k. The distance from point C to the Z-axis is denoted by f, which corresponds to the focal length of the corresponding camera. For the same infrared light, it is reflected at points K and O, respectively, and projected on the depth image based on the point C, and then the measurement gap d can be obtained. When combined with principle of similar triangles, the following equation can be obtained:

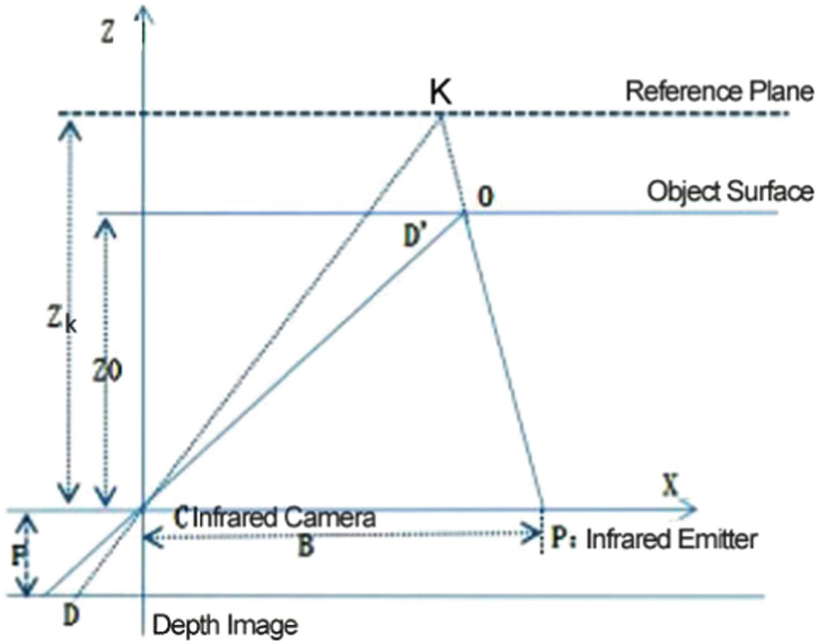$$\frac{D}{b} = \frac{Z_o - Z_k}{Z_o} \tag{1}$$

Fig. 1. Calculation of the spatial depth Z_0 based on the infrared spot

$$\frac{d}{f} = \frac{D}{Z_k} \tag{2}$$

Combining the two equations, the D parameter can be eliminated, so that:

$$Zk = \frac{Z_o}{1 + \frac{Z_o}{fb}d} \tag{3}$$

With this formula, the Zk mathematical model corresponding to the point K can be calculated, and the relevant parameters such as d, f, and Zo and b need to be corrected with the depth information. For the two parameters f and b, the relevant hardware can be provided. In the specific practice, based on the different spatial environment, the selection of O reference plane is often dynamic. Therefore, the Zo value is usually determined in conjunction with the correction. The d value is closely related to the Zo value. Therefore, it is necessary to perform linear regression of the above 3 formula, that is, linearly normalize the d parameter, and indicate this parameter by $md' + n$, wherein m and n are the parameters corresponding to the normalization process, and substitute it into the above formula (3), the following formula (4) can be obtained:

$$Zk^{-1} = (\frac{m}{fb})d' + (Z_o^{-1} + \frac{n}{fb}) \tag{4}$$

This formula shows the linear relationship between Zk and Zo. By means of a large number of samples and obtaining the corresponding regression coefficients by least square method, the spatial depth value of point k can be obtained.

The algorithm for recognizing human motion has been continuously optimized with the development of camera technology, which provides an important technical basis for the application of somatosensory interaction technology. Motion recognition can be completely converted into shape recognition, which obviously provides a new way of thinking. The core algorithm of this technology is mainly: in the following formula, the pixel point in the image is represented by x, and its eigenvalue f_θ is [4]:

$$f_\theta(I,x) = d_I(x + \frac{u}{d_I(x)}) - d_I(x + \frac{v}{dI(x)})$$ (5)

Then use $d_I(x)$ to indicate the depth value of the pixel. When the user enters the image, the coordinates of his own attributes can be obtained by combining the position of user and the distance from the camera. Then you can start normalization by multiplying $d_I(x)$ and the offset values of the above coordinates to obtain two different depth values, which are then subtracted to obtain the eigenvalues f_θ. If this pixel belongs to the background image, then the value of $d_I(x)$ will be a maximum positive value, so that the shape of the human motion can be separated from the background image. After acquiring the silhouette of the human body, the motion represented by the silhouette can be recognized. Currently, the algorithm database has integrated up to 500,000 frames of human silhouette images, covering driving, dancing, walking, running and many other actions. Based on the database information, the corresponding classification device can be trained and the human skeleton can be further subdivided into 31 types of labels. Then, according to the Decision Tree algorithm, the x pixel points in the I image are assigned to the corresponding bone labels. The formula (6) indicates the decision forest consisting of T decision trees. In this decision tree, different leaf nodes exist f_θ and τ. The former corresponds to the eigenvalue and the latter to the threshold. And these decision trees can, based on $P_t(C|I, x)$, express associated pixel as the probability of belonging to a certain bone tag C.

$$P(C|I,x) = \frac{1}{T}\sum_{t=1}^{T} P(C|I,x)$$ (6)

In the following formula, the weight value of the pixel on the tag C is represented by w_{ic}. The pixel depth value $d_I(x)$ can ensure the stability of the corresponding weight value depth. Using the above formula, the correlation between the pixel and the tag C can be obtained:

$$w_{ic} = P(C|I,x) \cdot d_I(x)^2$$ (7)

When many such pixels on the human silhouette are clearly identified and have their own structure tag, the mean shift algorithm can be used to integrate the pixels of different tags. When the weight value is greater than a certain threshold of the pixel

point, it will be divided above the joint point to which this label is mapped. The total value of these weight values is the predicted value of the joint point confidence. When the joint point confidence exceeds a certain predicted value, the algorithm result can be included in the corresponding human silhouette image. With a large number of joint points, a complete human skeleton can be derived so as to identify the human movement.

3.3   Somatosensory Interactive Development Technology Route

Open natural interaction, the so-called OpenNI, the essence of it is openness. The application of this technology in the field of somatosensory interaction has been relatively mature and has been used widely. The development of this system is also the development of the overall architecture with this platform.

The following figure shows the interaction framework, where the top layer is the application layer, which corresponds to the application; the bottom layer is the hardware layer, represented by Hardware, which covers depth, color image cameras, and sound input; between the application layer and the hardware layer is the corresponding interface layer. The OpenNI, between applications and the hardware devices, provides the corresponding interfaces and middleware [5] (Fig. 2).
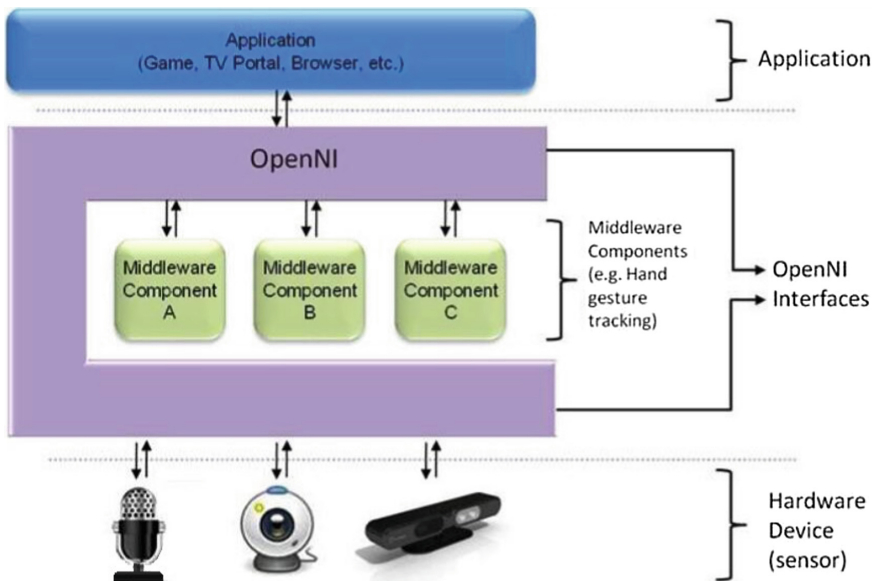

Fig. 2. OpenNI development framework

OpenNI defines the human skeleton frame, and each joint has two parameters, direction and position. Currently, in OpenNI, a total of 24 joint points are defined. With the middleware NITE, the human body motion can be captured, and at most 15 joint points of them can be obtained [5], as shown in detail in Fig. 3 below.
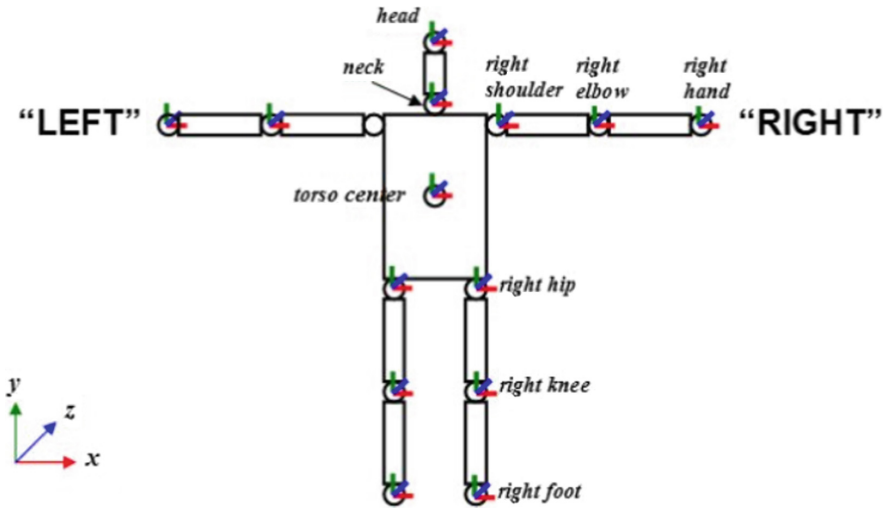
Fig. 3. Schematic diagram of joint points currently supported by OpenNI

In addition, the interactive technology also builds a category to serve the communication of software, which is called the production node category. To build an OpenNI framework, the role of production nodes is extremely important, that is, the development by means of OpenNI needs to apply production nodes. In addition, the technology defines other categories to make it more functional. These features are not necessary; developers can use them according to their own needs. Such categories are often referred to collectively as capability category. The technology uses OpenNI to realize the recognition of motions, and the core technology of it is the production node category, skeleton tracking and posture detection capability category.

4   Application of Somatosensory Interaction Design in 3D Virtual Experiment

4.1   Control of the Experimental Interface

The somatosensory interaction mode needs to be applied throughout the virtual experiment. Operators can use gestures to manipulate related windows and objects. For example, the selection and manipulation of menus, the choice of experiments, etc. At this point, the gesture can completely replace the operation of the mouse. In the Unity3D system, the gesture can replace the mouse for input by displaying the student's skeleton points recognized by Kinect and binding the coordinates x1, y1 and z1 of the mouse to the coordinates of the student's left hand or right hand (x2, y2, z2). For example, the algorithm of the abscissa: $x1 = x2 * a + b$, where b and a correspond to the position difference and the activity proportional coefficient of the arm relative to the screen area, respectively. In this way, the manipulation of the menu can be realized by means of gestures.

4.2   Selection and Assembly of Experimental Instruments

Students can conduct hydrogen production experiments in virtual experiments. They can capture the corresponding experimental equipment via gestures from the instrument library and select the corresponding reagents and experimental materials from the drug warehouse. The system will based on the experimental content to identify whether the experimental equipment combination selected by the student is accurate, and then give corresponding prompts. If the selected equipment and hydrogen preparation conditions are inconsistent, then the corresponding error will be prompted. Only when the selected instrument and reagent information are accurate, will it prompt the correct message and subsequent experiments will be carried out. If the logic judgment condition is not accurate, then the corresponding error will be prompted and the corresponding experimental preparation knowledge will be given. The corresponding virtual operating environment is consistent with the real environment in which the students are located. Students only need to use gestures to manipulate the relevant experimental equipment in the actual environment, thus producing extremely realistic experimental results. The animation interaction is detected by the vision of the distance between the bone nodes and devices and the collision between the object and the instrument. Students can use their hands to produce actions as performed in real experiments, complete the design of the relevant experimental principles, and add appropriate reagents, such as water and dilute sulfuric acid, into the corresponding containers, as shown in detail in Fig. 4 below.

Fig. 4. Selection and assembly of experimental instruments

4.3   Operations in the Experimental Process

In the virtual experiment process, the instrument can be operated and corresponding experimental reagents can be added by means of gestures. When adding the corresponding reagents, it needs to be implemented as much as possible by animation. Using the arm bones, as well as the collision detection of the reagents and the calculation of

the distance to stimulate the generation of related operation animations, and then the corresponding reagent weights are added according to the experimental requirements in order to complete the subsequent animation interactions, and display the corresponding experimental phenomena. For the demonstration of experimental animation, you can use the conditional statement to set up, so you can complete the virtual experiment of hydrogen production, as shown in detail in Fig. 5 below.
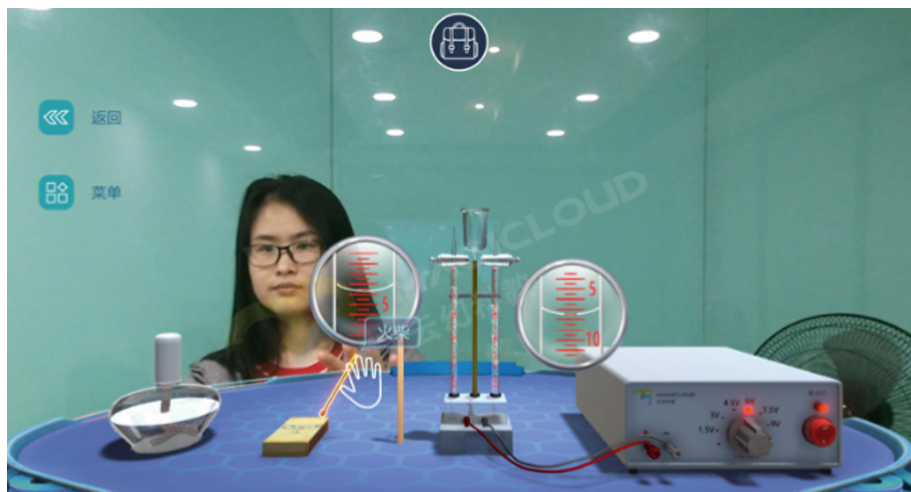

Fig. 5. Operations in the experimental process

4.4   Experimental Test

When the experiment is completed, the system will generate the corresponding experimental test results, and ask questions by taking into account of the experimental cautions, operational safety, experimental phenomena and equipment assembly based on experimental principles, while students can also use the virtual environment to answer questions by gestures. The system evaluates the results according to the students' answers and gives corresponding learning suggestions.

5   Conclusion

The implementation of virtual experiments using VR technology through the Kinect somatosensory technology allows students to perform experiments more realistically in a virtual environment. In this way, many sensory organs of students can be mobilized to participate in experiments and learning, so that it is better to help students improve their experimental ability. This technology can also save a lot of animation production problems involved in traditional virtual experiments, so it can greatly improve the efficiency of the Kinect-based virtual experiment, and save significant cost, providing

corresponding direction of innovation on the in-depth application of virtual experimental technology.

References

1. Chen, G.: Virtual display system of process design based on somatosensory interaction technology. J. Xi'an Univ. Posts Telecommun. **21**(5), 76–79 (2016)
2. Zhang, Y., Cheng, X., Chen, H., et al.: Design and implementation of middleware based on Kinect somatosensory interactive development. Comput. Appl. Softw. **4**, 10–14 (2018)
3. Zhang, R., Liu, L., Zhang, G., et al.: Research on Kinect's method of accurately tracking multi-objectives in somatosensory interaction process. Mod. Manuf. Eng. **10**, 76–80 (2014)
4. Sun, B., Zhang, J., Cai, Y., et al.: Multiplayer online virtual experiment system based on Kinect somatosensory interaction. Comput. Sci. **43**(9), 284–288 (2016)
5. Gu, R., He, Y., Jin, L.: Design and implementation of intelligent teaching control system based on Kinect skeleton tracking technology. Comput. Appl. Softw. **30**(6), 276–280 (2013)