Retraction

Retracted: Construction of Virtual Piano Performance System Based on Visual Gesture Recognition

Mobile Information Systems

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope

(2) Discrepancies in the description of the research reported

(3) Discrepancies between the availability of data and the research described

(4) Inappropriate citations

(5) Incoherent, meaningless and/or irrelevant content included in the article

(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] K. Tian, "Construction of Virtual Piano Performance System Based on Visual Gesture Recognition," *Mobile Information Systems*, vol. 2022, Article ID 9446905, 12 pages, 2022.

Research Article

Construction of Virtual Piano Performance System Based on Visual Gesture Recognition

Ke Tian ⓘ

Conservatory of Music, Nanjing Xiaozhuang University, Nanjing, Jiangsu 211171, China

Correspondence should be addressed to Ke Tian; z2008004@njxzc.edu.cn

In order for computers to understand people's gestures through cameras and react accordingly, through in-depth research, gesture recognition technology in natural human-computer interaction is proposed. Combining natural human-computer interaction technology with music performance, using computer vision-based gestures, music is played in a virtual environment. Experiments show that the virtual piano has 14 piano keys. For the realization of piano performance, it is each piano key; once the piano key is greater than the set value, just call m_Wave.Load() to make a sound. According to the CWave class in the object-oriented MFC class library under VC++, create an object m_Wave of class CWave. Then, according to the m_Wave.Load() function of the CWave class, the connection of the sound is realized. It successfully solves the troubles of music lovers, enriches people's spiritual life, and has certain practicability and scalability.

1. Introduction

Human-computer interaction technology is the interaction between human and computer [1]. This technology is increasingly becoming an important part of people's lives; in particular, computer and information technology, intelligent technology, biotechnology, and other technical fields have developed rapidly, as shown in Figure 1. Virtual technology started in the 1990s; the latest development is advanced human-computer interaction technology, which is an effective simulation of human activities in the real world. Advanced human-computer interaction includes techniques for behaviors, such as listening, speaking, and grasping. It is a collection of various technologies, such as artificial intelligence, computer graphics, human-machine interface, and network technology [2]. Virtual technology is a technology of great significance; it is important for defense, military, aerospace, medical, education, manufacturing, art, entertainment, and even people's daily lives, where all have great connections and effects. Computer vision is also a subject area at the forefront of knowledge. During the past thirty or forty years of development, it has been applied to

important fields, such as robot navigation, industrial inspection, and medical image processing [3]. It has become one of the important and widely applied disciplines in people's daily life, national economy, national defense science and technology research, and other fields. More than 85% of human perception of external information is obtained through vision. Therefore, in the application of virtual reality technology, visual communication has become the most important sensory interaction link in virtual reality technology, and visual technology has naturally become an important support technology for virtual reality [4]. In the early days of virtual reality technology research, the pioneer of computer graphics, Ivan Sutherland, in the Sword of Damocles system, realizes 3D stereoscopic display; the suspended objects in the air were observed with the human eye, which was very noticeable. Many mainstream virtual reality systems now support visual technology without exception. In the technology of virtual reality visual interaction, using noncontact methods, for example, the camera lens, and so on, to observe the user's actions to realize human-computer interaction, has become an increasingly promising 3D interactive technology [5].
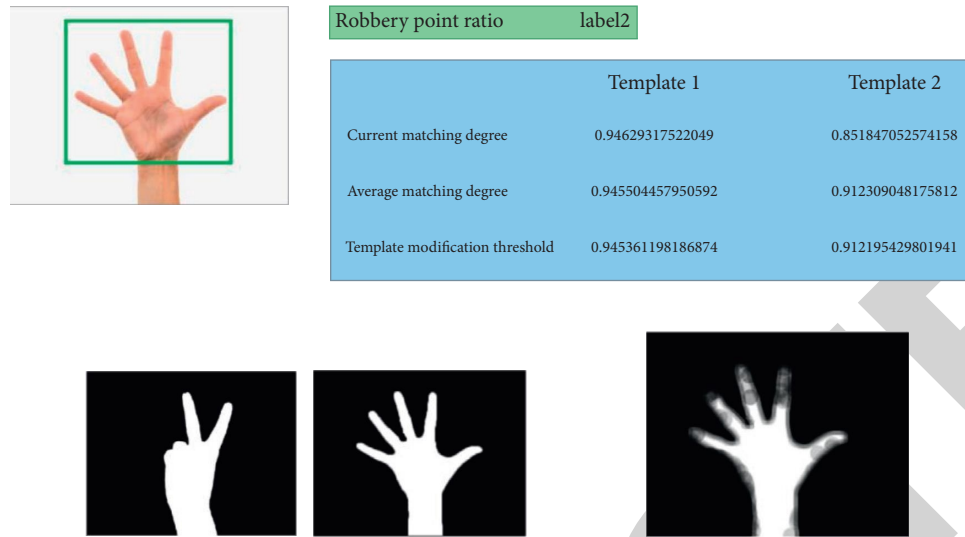
Robbery point ratio	label2	
	Template 1	Template 2
Current matching degree	0.94629317522049	0.851847052574158
Average matching degree	0.945504457950592	0.912309048175812
Template modification threshold	0.945361198186874	0.912195429801941

FIGURE 1: Human-computer interaction technology.

2. Literature Review

Music is the language of the human mind. Every country and every nation have their own unique and wonderful music. After thousands of years of development, music has developed into a systematic, complete, and diverse discipline. Lee and others found that in order to play music well, it is necessary to accept certain training and guidance [6]. Zhu and others found that in this era of extremely rich material life and spiritual life, more and more people like music, like to play music, and like to express their feelings through music. In traditional music performance, people usually get all kinds of wonderful music by playing various musical instruments [7]. Khan et al. found that due to the economy, for various reasons, such as playing skills and types of instruments, many people cannot get their favorite instruments to play. The authors propose to combine computer gesture-driven technology with the music performance process; the user only needs to step into the virtual scene, move the palm left and right to simulate plucking the strings, just can play the basic scales of various tones: do, re, mi, fa, so, la, si, and do. These basic scales are connected to make beautiful music, without direct contact with various musical instruments [8]. Kim et al. found that this method not only meets the needs of music fans for music performance, but also solves the inconvenience caused by playing various musical instruments, and has high usability and practical significance; it can be applied to some entertainment venues as well as music academies, music museums, and other places [9]. Traver et al. found that computer trend recognition technology is a new type of natural human-computer interaction. Different from the traditional machine-centered human-computer interaction method, natural human-computer interaction is a multimedia and multimode interaction technology, which emphasizes more on the user-centered, using computer-controlled technology that conforms to natural communication habits, and provides users with a natural and effective human-

machine communication interface [10]. Zheng et al. found that in the process of research and development of natural human-computer interaction systems, human-virtual device interface become an important part of the research center. Particularly in recent years, with the rapid development of computer technology, research into new human-computer interaction technology that conforms to human-computer communication habits become extremely active and also made gratifying progress [11]. Joglekar et al. found that, among them, computer gesture recognition technology is a natural, intuitive, and easy-to-learn means of human-computer interaction [12]. Perrett et al. found that in the process of gesture recognition, the human hand is directly used as the input device of the computer, and the communication between the human and the computer will no longer require an intermediate medium; the user can simply define an appropriate gesture in order to control the surrounding machines [13]. Wibisono found that current computer gesture recognition technology is mainly used in multichannel user interface, postoperative human rehabilitation, virtual environment interaction, intangible heritage protection, blind sign language recognition, and so on [14]. In the field of music digitization and virtual performance, gesture recognition also has great usability and advantages. Mo and Sun found that applying gesture recognition technology to virtual performance technology, the virtual performance of music can be realized, allowing more music lovers to engage in music performance and research [15].

3. Methods

Virtual technology is essentially a human-computer interaction technology, and it is an advanced human-computer interaction technology (because it requires interaction in a purely natural way and for multichannel information). The virtual reality system is a concrete embodiment of the application of virtual technology; it is input and output device
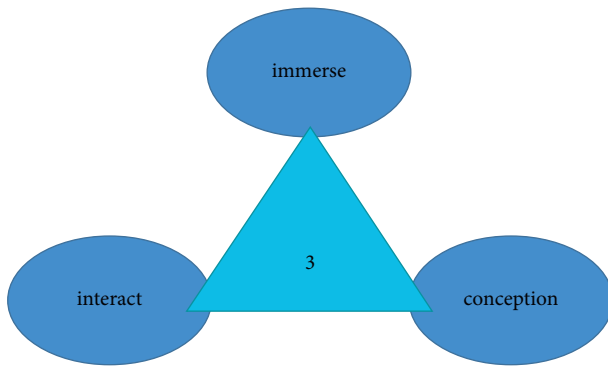
Figure 2: Three "I" diagrams.

by various sensors, and computer-built virtual scenes, a system implementation of virtual technology composed of virtual objects. The usual virtual reality system has three basic characteristics, namely, "Immersion-Interaction-Imagination"; in 1993, Burdea proposed the "Triangle of Spiritual Technology," and it is as shown in Figure 2.

Immersion is the operator in the virtual reality system, immersion when interacting with computer-generated virtual environments; Interaction is the interaction between input and output devices and the virtual environment; and Imagination is a variety of "ideas" based on computer simulations.

These three characteristics are the basic characteristics of virtual reality system, the role of humans in virtual reality systems is emphasized. In the past, people could only observe the results of calculation and processing from outside the computer system so that people can walk into the virtual scene and immerse themselves in the environment created by the computer system. In the past, people could only use the keyboard, mouse, interact with single-dimensional digitized information in a computer environment; now people use virtual reality interactive technology, being able to use a variety of sensors to interact with a diverse information environment. The purpose of people is that, in a virtual reality system, information processing systems consist of computers and other equipment in order to try to "meet" people's needs, rather than forcing people to serve those computer systems. The feeling that people can get through the virtual reality system, feel as close as possible to what they get in the real world. Enabling people to immerse themselves in virtual reality systems, and through the virtual reality system, we can get the feeling that cannot be experienced in the real world. Breaking through the limitations of physical space and time and avoid dangers involving life, it makes the virtual reality system have the virtuality of "beyond reality." The basic means and fundamental purpose of constructing a virtual reality system, it is to use high-performance computer software and hardware and integrate various advanced equipment, in order to achieve an immersive sense of immersion for the operator, a virtual reality system with perfect interaction capability [16]. At present, in countries all over the world, researchers are making efforts and attempts [17].

Generally speaking, to build a complete virtual reality system, it is necessary to have the overall division of

hardware and software in terms of hardware, such as a tracking system that can detect the operator's head, hands, body, and other positions' information. Force feedback system that can provide force sense and touch; audio system that can provide sound; image generation and display systems capable of producing three-dimensional stereoscopic images; and tracking systems that provide vision, as the author did. In terms of software, there is generally a software support environment, there are also toolsets that can generate virtual scenes and virtual objects; it also receive various high-performance sensor information functions (e.g., helmet tracking information, visual image processing information, etc.). The ability to generate and display three-dimensional graphics (such as virtual hands, virtual piano, etc.); with the development of virtual reality technology today, a lot of applicable results have been achieved, and virtual reality technology is popularizing to the market; with the multidimensional information space, it can gradually provide an application environment for virtual reality systems. Future virtual reality technology will be like other current science and technology; it has become the most common and best method and tool for human beings to understand and transform the world [18].

(1) If decomposed from the perspective of the system, the virtual reality system mainly includes VR scene observation system, VR scene generation and accelerated display system, high-performance computer processing system, audio system, tracking system, and haptic and force feedback system, as shown in Figure 3:

(1) VR scene observation system: it is used to observe the VR graphics scene output by the computer, such as helmet display and stereo glasses.

(2) VR scene generation and accelerated display system: it generates visual images and stereographic display, such as graphics subsystems in workstations and professional stereoscopic 3D graphics accelerators.

(3) High-performance computer processing system: it is a computer system with high processing speed, large storage capacity, and strong networking characteristics, such as a high-performance PC.

(4) Audio system: it provides stereo source and determine spatial position, such as stereo, speakers, headphones, and so on.

(5) Tracking system: it is used in order to determine the position of the participant's head, hands, and body, such as space tracking locator and space trackball.

(6) Haptic and force feedback system: it provides force and pressure feedback, such as data gloves with force or haptic feedback, robotic arms, and so on.

The closed virtual reality system does not directly interact with the real world, and any operation does not have a direct effect on the real world [19]. The closed virtual reality system consists of three parts: modeling module, 3D model library, and interaction module. The modeling module can use knowledge base, pattern recognition,
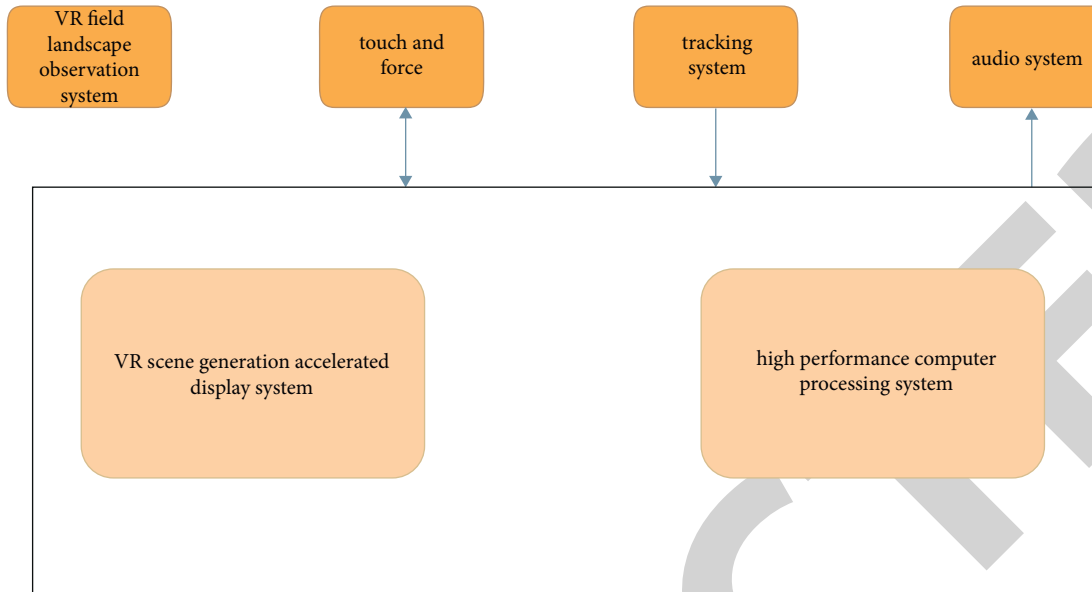
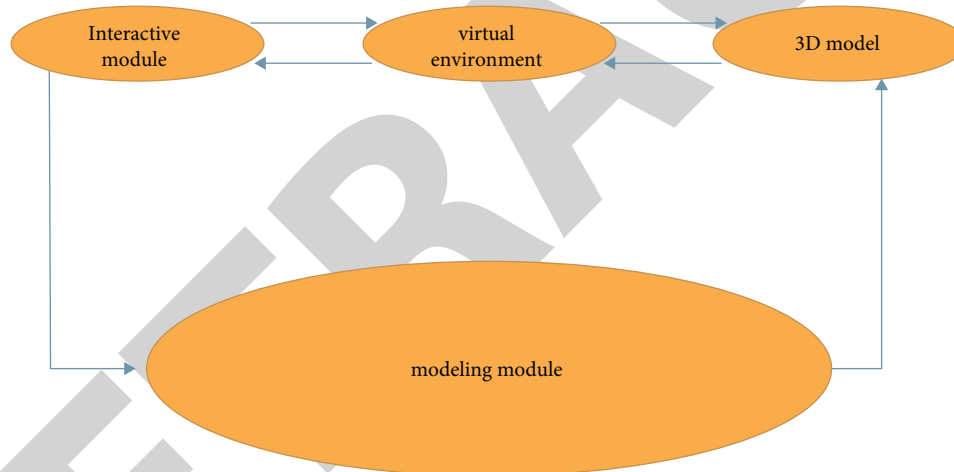Figure 3: Schematic diagram of virtual reality system.

Figure 4: Composition of a closed virtual system.

artificial intelligence, and other technologies to build models, visual simulation of virtual scene through 3D animation, and sound simulation through sound production. The 3D model library is the 3D representation of the components of the real world, which can constitute the corresponding virtual environment. The interaction module contains several submodules, such as sensor, signal detection and control, and signal feedback and control. Human movements can be detected by sensors, and then the virtual environment can be operated by controlling submodules; at the same time, the feedback gives people the feeling of movement, touch, force, and so on. The composition of the closed virtual reality system is shown in Figure 4.

Open virtual reality system; then a feedback closed loop can be formed through sensors and the real world; therefore, the virtual environment can be used for the purpose of direct operation and remote operation of the real world. Sensor devices include the abovementioned visual sensor devices and auditory and tactile system sensors.

Given the virtual reality system hardware equipment, virtual worlds are often created and drawn using a variety of applications and toolboxes, and an interface with the virtual world is realized [20]. Drawing is the process of creating a sensory image that describes a virtual world and belongs to the software-generated virtual environment, build virtual worlds, and virtual scenes and virtual objects. First of all, the model of its virtual object must be established; when the virtual object is created according to various tools such as CAD, OpenGL, and so on, generate virtual pairs that can be viewed in virtual scenes; then store them as separate files in the model library. In virtual reality technology, the modeling and rendering of virtual scenes and virtual objects has always been the focus of research; it is also the primary core issue. At present, the world is around the virtual environment modeling problem. There are mainly two solutions: one is
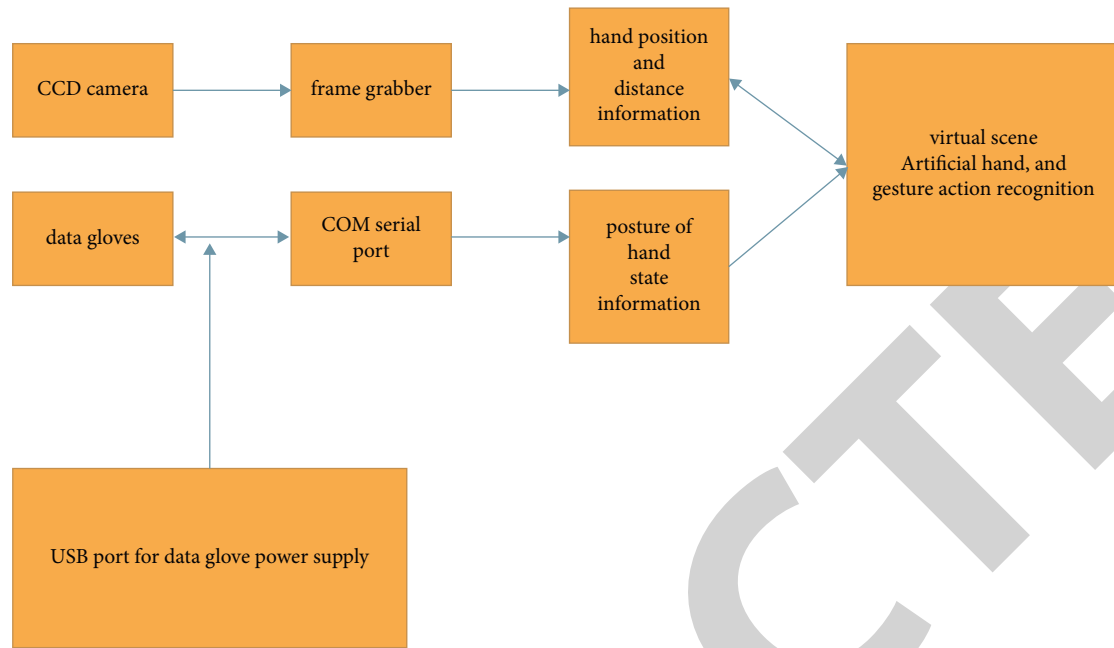
Figure 5: Schematic diagram of virtual system composition.

the modeling and drawing based on geometry (graphics) based on traditional computer graphics; the second is image-based modeling and rendering based on image sampling of the 3D environment. Both methods have advantages and disadvantages. Each object in the virtual environment contains two aspects: shape and appearance. Model files used to store geometric models in virtual environments should be able to provide information on both. At the same time, it must meet three common indicators of virtual modeling technology: interactive display ability, interactive manipulation ability, the ability to easily construct, and the requirements for virtual object models. Open GL is a common tool software for drawing virtual devices; the author's virtual scene and virtual hands, as well as the virtual piano, are drawn with Open GL. Open GL is a set of graphics rendering algorithms, providing a standard cross-platform programming interface. It is easy to implement various transformations, shading, lighting, textures, interaction operations, and animations of models in Open GL, but it can only provide modeling functions of basic geometric elements, which makes the modeling of complex models relatively difficult. 3D graphics modeling tools such as 3DMAX can be used. It is convenient to establish various complex special body models, but it is difficult to carry out program control. After the complex model is established, it can be easily controlled and transformed in Open GL.

The hardware environment in which the author works is a Pentium four-processor PC, one-color CCD camera, one CG300 Daheng frame grabber, CAS-GLOVE data gloves, and a virtual software platform developed by the Chinese Academy of Sciences. The software development tool is VC++6.0, and the system composition diagram is shown in Figure 5.

Software development interface and schematic diagram of virtual world, as shown above, the composition principle of the virtual system software development platform of this department, it consists of the following modules:

(1) Initialization (OpenGL, communication port, etc.) module: complete the automatic initialization before running.

(2) Control module: the overall control system operation and the coordination and scheduling between each module.

(3) Communication module: organize the computer to communicate with the DSP controller through the serial port.

(4) Drawing output module: control the real-time drawing of virtual hands.

(5) Data processing module: the sensor information sent from the serial port of the computer, converted to bending angle information.

(6) Force feedback output module: set the force feedback mode and output port.

(7) Calibration module: calibrate the data glove before use and determine the maximum/minimum bend angle of the sensor.

(8) Data acquisition module: after the sensor information of the data glove is processed by the DSP signal, it is sent to the serial port of the computer.

(9) Angle calibration module: in the debugging stage, it is used to calibrate the value of each bending angle.

The interconnection between them is shown in Figure 6.

The author, on the basis of the original system software platform, shows that the introduction of visual target tracking and recognition system enables the system to process information from the visually collected images; fused with the object parameters of the virtual scene, it can
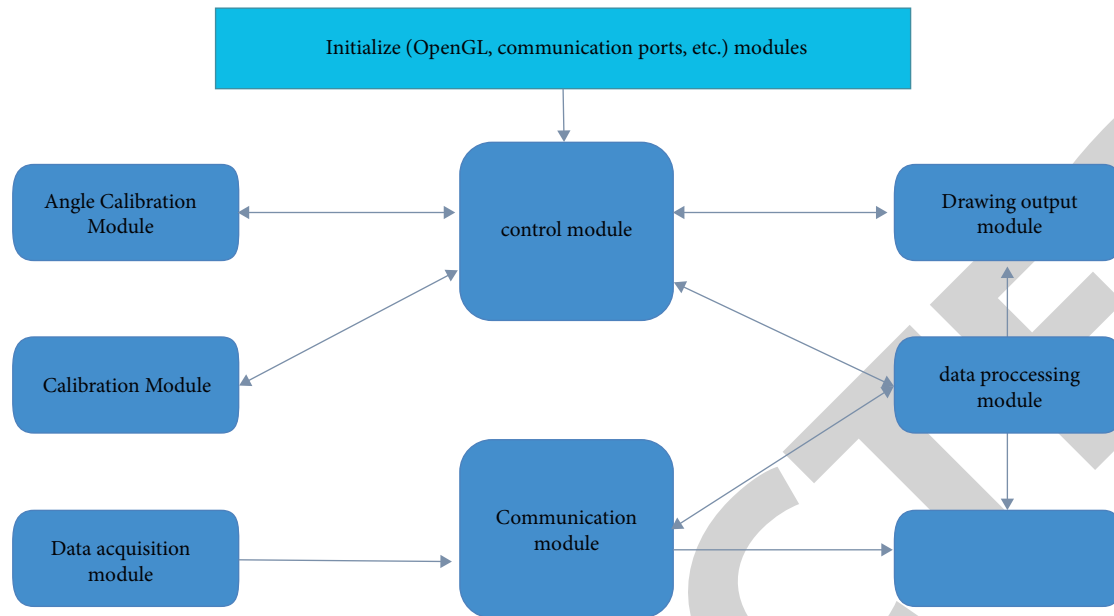
FIGURE 6: Schematic diagram of virtual development software platform modules.
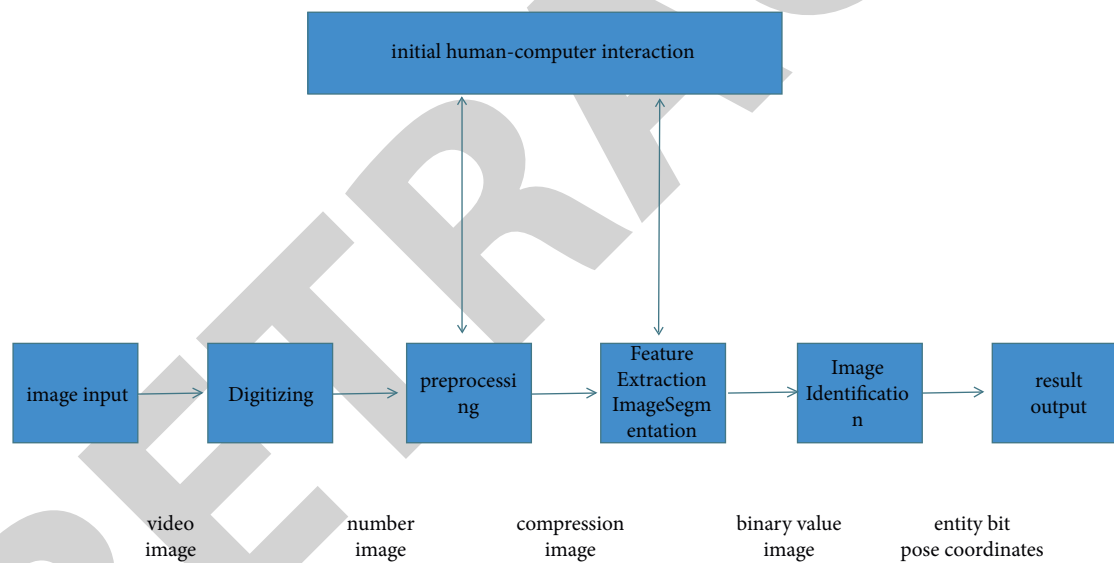

FIGURE 7: The working process of the vision system.

effectively realize the tracking of real-time motion positioning of virtual objects under vision. The process is as follows: the camera captures the target image to the frame acquisition card, the image information is converted from analog signal to digital signal, and the digital signal is sent to virtual software platform for processing. The data of the data glove passes through the PC COM port, the virtual hand that transmits data to the virtual world in the virtual platform, the sensor data of the data glove is transmitted to the virtual hand. The author's work is to use a video camera and a frame grabber in order to obtain the image information required for the three-dimensional spatial positioning of the virtual hand, image processing, target color recognition, state filtering; then according to the image centroid and distance information, real-time tracking of the coordinates and state of the hand under vision, and according to the real-time tracking, the virtual hand is moved above the virtual piano to realize the virtual piano performance. A typical working process of a vision system is shown in Figure 7.

Generally, in a visual system, we collect images through cameras; the camera can convert the light signal it feels into the corresponding electrical signal. This electrical signal is sent to the TV screen, and the ingested image can be presented. Such images are generally referred to as video images. When the camera acts as the "eye" of the visual system, when the ingested image is sent to the "brain" computer that processes the image for recognition, the first problem to be solved is how to turn a video image into a digital image that a computer can process. Currently in image input and digitization, the commonly used method is a CCD camera plus

an image acquisition card (grabber) [21]. The camera obtains the analog video image, and the image acquisition card completes the analog/digital (AD) conversion, digitizing the video signal. It can be seen that ordinary vision is composed of three parts: camera, frame grabber, and computer.

In the whole process of vision system work, feature extraction, image segmentation, and image recognition are the core tasks, in order to carry out these tasks quickly and efficiently, we can take advantage of the unique functions of the image device in the system, properly adjusting the volume and quality of the captured image. This process belongs to the preprocessing of the image, such as denoising, filtering, histogram averaging, and so on. In order to separate the target from the image background, the process of finding differences between targets and nontargets is called feature extraction. There are many ways to extract image features; the two main ones are image-based brightness features and image-based texture features; brightness-based feature extraction methods include histogram feature extraction. The texture-based feature extraction method is more flexible in the specific method, such as the contour features of the specific target image, the region growth of the target image, fractal method of image splitting and merging, and so on. And according to some feature differences between the different extracted targets, the area division of an image is called image segmentation. There are many methods of image segmentation, among which the simplest image segmentation is the binarized image; that is, all pixels in the image are divided into two categories: target and nontarget. If the point corresponding to the target pixel is set to 1, and the other pixels are set to 0, then the image of the target area is white, and the other areas are black. Initialize the human-computer interaction process, mainly complete the process of image preprocessing, feature extraction and image segmentation, and some parameter selection or parameter passing. Like: Modify the parameters of the chromaticity and brightness of the frame acquisition card through the man-machine interface; the capture effect of the image can be changed. Another example is using a keyboard or a mouse to draw a target image area in the image.

The image acquisition system is the first step in the work of the robot vision system; it is also a step that has an important impact on the subsequent processing effect.

The quality of the collected images is good or bad; it directly affects the correctness of the final image processing and 3D reconstruction results. It consists of camera, frame grabber, computer, and image collection software. Its main function is to convert the analog video signal acquired by the vision sensor in real time, convert it into a digital image signal, and directly transmit the image to the computer for display and processing, or transfer the digital image to a dedicated image processing system, perform real-time front-end processing of visual signals.

Color space is a three-dimensional linear space, any kind of color light with a certain brightness, both are a point or a vector in a space, and this space is the color space. Modern chromatic theory states that a person's perception of any color can be weighted by three monochromatic colors of red, green, and blue; therefore, the three monochromatic colors of red, green, and blue are called three primary colors. This is the three-color principle; our color threshold is carried out in a three-dimensional color space. There are many types of color spaces; the more commonly used are Hue-Saturation-Intensity space (HIS for short), YUV space, and Red-Green-Blue space (RGB).

When we choose the three primary colors of red, green, and blue with definite luminous flux as the basis of the three-dimensional color space, it constitutes the RGB color space. But the RGB color mode is susceptible to lighting effects. In visual color recognition, color recognition is done through different color markers. However, the light intensity of different positions on the playing field varies greatly, making the R, G, and B values of a color vary greatly at different positions on the field. In this way, the color cannot be judged by a single set of thresholds, the robustness of the identification procedure is also out of the question. Its color space model is shown in Figure 8.

The HSI color space model is a color representation system corresponding to the characteristics of human perception; it has a directional arrangement for color awakening, which is very beneficial to the extraction of color information. In the HIS color model, H (hue) is the essential characteristic of color in hue, corresponding to the dominant wavelength of light; S (saturation) is saturation, indicating the degree to which a color is blended with white, that is, the purity of the color; and I (intensity) is the light intensity, the brightness of the light. The HIS color space model is shown in Figure 9.

The YUV model is the color space used by European television systems (PAL format). Under normal circumstances, the camera takes the captured color image signal, after color separation, amplification, and correction, to obtain RGB signal, after being transformed by the matrix transformation circuit, obtain the luminance signal Y and the two color-difference signals R-Y and B-Y. Finally, the sender encodes the luminance and color-difference signals respectively, sent on the same channel.

Unlike the RGB space, the HSI and YUV color spaces represent the spectrum in two dimensions; the third dimension is used to represent the intensity of the color; for example, in HSI, H and S represent color information, I represents intensity. In YUV, U and V represent color, and Y represents intensity. Therefore, these two color spaces, compared to the RGB mode, are more suitable for occasions where the light intensity changes. In practical applications, RGB signals can be converted into HSI signals or YUV signals. In order to do this, coordinate transformation is needed to find an easy-to-analyze color space; it can exclude the influence of lighting and extract the nonluminance information of the color.

In addition to the need for purpose, the conversion of the color space model must follow certain principles.

First, for the reversibility of the conversion, there are color space models S and D. If there is a conversion $D = f(S)$, there must be $s = f^{-1}[D]$; the transformation f can be a transformation between the color space models S and D.

Second, for the lossless conversion, because RGB space is a set of integers obtained by a sample quantizer, any
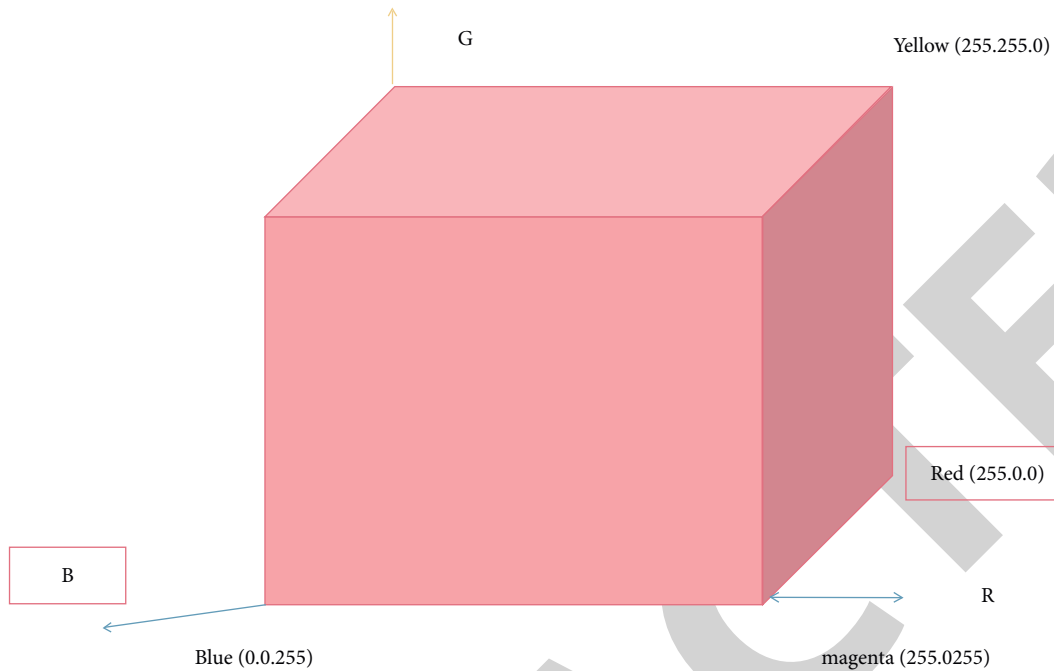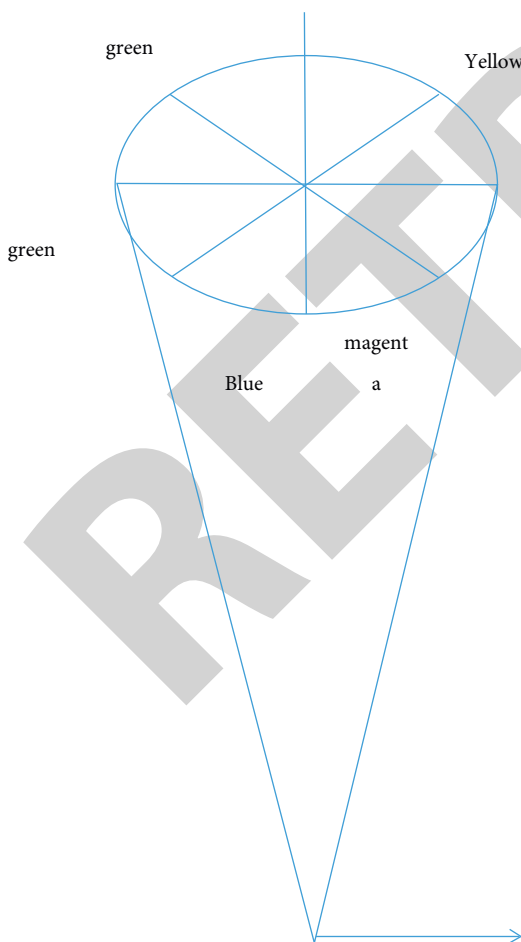
FIGURE 8: RGB color model.

FIGURE 9: HIS color model.

conversion to and from RGB space, both must be invertible transformations from integers to integers.

RGB to HIS conversion:

$$\begin{cases} H = \mathrm{arctg}\left[\dfrac{\sqrt{3}\,(G-B)}{2R-G-B}\right], \\[2mm] S = I - 3\min(R, G, B)(R+G+B), \\[2mm] I = \dfrac{1}{3}(R+G+B). \end{cases} \tag{1}$$

RGB to YUV conversion:

RGB image obtained from the camera is converted to YUV values in the program. Here is a more recognized conversion formula:

$$\begin{bmatrix} Y \\ U \\ V \end{bmatrix} = \begin{bmatrix} 0.299 & 0.586 & 0.113 \\ -0.168 & -0.332 & -0.5 \\ 0.5 & -0.418 & 0.082 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}. \tag{2}$$

4. Experiments and Discussion

The images collected by the DH-QP300 image acquisition card selected in the experiment are color images, usually developed by image processing programs, both requiring that the object to be processed is a grayscale image; color images should be converted to grayscale images before image preprocessing [22].

Color images are obtained by mixing red (R), green (G), and blue (B) as primary colors in different proportions.

Grayscale is the process of making the R, G, and B components of a color equal in value. The value range of R, G, and B is 0–255, so the grayscale image can only represent 256 colors.

There are three main methods of grayscale processing:

Maximum method: Make the value of the R, G, and B components equal to the largest of the 3 values; namely,

$$R = G = B = \max(R, G, B). \tag{3}$$

The maximum method results in a very bright grayscale image.

Average method: Make the value of R, G, B components equal to the average of the three; that is,

$$R = G = B = \frac{(R + G + B)}{3}. \tag{4}$$

The averaging method produces a softer grayscale image.

Weighted average method: Assign different values to R, G, B components according to importance or other indicators, and make a weighted average of the three values; namely,

$$R = G = B = \frac{(W_R R + W_G G + W_B B)}{3}. \tag{5}$$

Among them, W_R, W_G, and W_B are the weights of R, G, and B components, respectively. W_R, W_G, and W_B take different values; the weighted average method forms different grayscale images. Experiments and theoretical derivations prove that $W_R = 0.30$, $W_G = 0.59$, and $W_G = 0.11$; that is,

$$R = G = B = V_{\text{gray}} = 0.3R + 0.59G + 0.11B. \tag{6}$$

The most reasonable grayscale image can be obtained.

According to the target requirements of image processing, it is necessary to turn the image into an image with only two gray levels; that is, the image is binarized. Let the gray value range of the image $f(x, y)$ be in $[(a, b]$, and the binarization threshold is set to $T(a \le Tsb)$; then, the general formula for binarization is

$$g(x.y) = \begin{cases} 1 = f(x, y) \ge T, \\ 0 = f(x, y) \le T. \end{cases} \tag{7}$$

$g(x, y)$ is a binary image; usually, we use 1 to represent the object (black area) and use 0 to represent the background area (white area). There are many ways to choose the threshold T; it determines the quality of the binary image. According to the selection method of the threshold T, there are many methods for image binarization, such as mode method and threshold method. For example, the pattern method is that when the grayscale histogram has bimodality, the grayscale of the object and the background, generally, they are near the two peaks, so the center point of the valley can be taken as the threshold; this method is called the mode method. However, in reality, the grayscale histogram is not

very smooth, and there will be local minima generated by some small bumps, which will cause great inconvenience to automatic judgment [23]. The simpler method is to first smooth the grayscale histogram and then select the value or take the selected grayscale as the center to examine the two points of k; use these two points to judge the point of the selected gray value, whether it is a maximum or minimum point. Such a processing method will generate some noise, but it will not have a great impact on judgment.

According to the pointer to the image, the image data are taken out for processing, and the process is shown in Figure 10.

The binarized image formed after the image is segmented, according to the author's needs, set the color value of the target area as 255, the color value of the background area is 0, and its center point is the center of mass of the target; the calculation method is

$$S(x, y) = \begin{cases} 1, \\ 0. \end{cases} \tag{8}$$

Then, determine the position of the centroid of the hand:

$$x = \frac{\sum_{i=1}^{N} \sum_{j=1}^{M} S(i, j) \times i}{\sum_{i=1}^{N} \sum_{j=1}^{M} S(i, j)},$$

$$y = \frac{\sum_{i=1}^{N} \sum_{j=1}^{M} S(i, j) \times j}{\sum_{i=1}^{N} \sum_{j=1}^{M} S(i, j)}. \tag{9}$$

Among them, i and j are the number of pixels in the row and the number of pixels in the column of the determined hand area, respectively.

As shown in the figure, d is the distance of the target image from the camera, h is the side height of the target under vision, H is the height of the target image, D is the distance from the actual scene plane to the center of the camera, f is the focal distance of the camera, so the distance d of the target image from the camera can be reflected by the following formula, the distance between the target and the camera: the focal distance = the height of the side of the target: the height of the target image, as shown in the following formula.

$$d : f = h : H. \tag{10}$$

According to the above formula, we can calculate $d = f * h/H$. In the formula, we can measure h, and we can also obtain H according to the information of the target image segmented by the collection, we also calculate the value f of the focal distance, the calculation method of the author's camera focal distance f; what is used is to set the angle of view of the camera to 16 to get the image as shown in Figure 11; according to the triangle shown in Figure 11, the focal distance is calculated to be 2579 using the formula $f = 384/\tan 8$. According to the actual fixed camera height, the image resolution is $768 * 576$.

Among them, the kinematic relationship of each virtual hand joint and other parts, it can be explained by mathematical formula, since the mathematical models of the
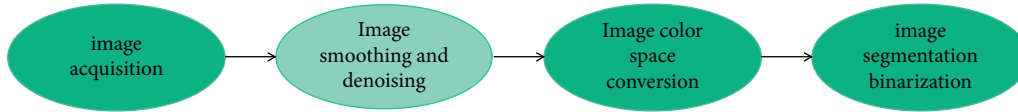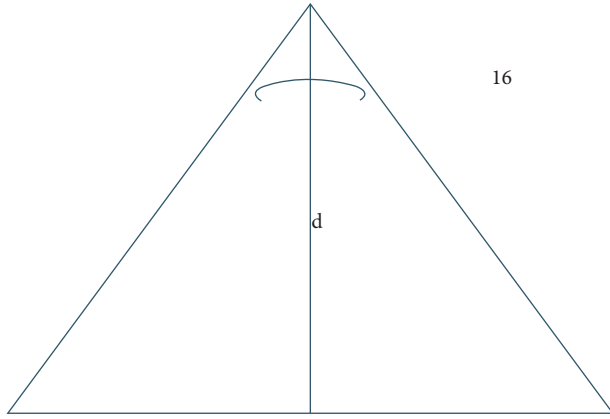
FIGURE 10: Schematic diagram of image processing.

FIGURE 11: The focal length of the camera.

FIGURE 12: Schematic diagram of two-dimensional coordinate system.

movements of the thumb, index finger, middle finger, ring finger, and little finger are similar, the author only explains the wrist joint, and the base coordinates defined on the wrist joint are $(X0, Y0, Z0)$; its $X0$-axis points along the forearm to the metacarpal bone corresponding to the index finger. The $Z0$-axis is parallel to the rotation axis of the wrist joint, and the $Y0$-axis is defined according to the right-hand rule. Define the wrist joint coordinate system WR on the carpal bone as (XWR, YR, ZWR), where m is the rotation angle of the wrist joint around the XWR axis, when $\varphi = 0°$, the palm and the forearm are in a straight line, and the base coordinate system coincides with the wrist coordinate system. From this, the homogeneous transformation matrix from the wrist joint coordinate WR to the base coordinate can be derived:

$$m^0R = \begin{bmatrix} \cos(\varphi R) & -\min(\varphi) & 0 & 0 \\ \min(\varphi) & \cos(\varphi) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad (11)$$

The pixel coordinates of the visual image and the distance information to the virtual hand, information mapping of states in the 3D world, is the coordinate point information and distance information of the image, converted to the virtual world coordinate system according to the coordinate transformation. According to the knowledge of the previous section, points (x, y) are all pixel coordinates, which are the coordinate points that reflect the two-dimensional array image collected into the memory, is a two-dimensional coordinate point, as shown in Figure 12 is the coordinate system of the image coordinate.

Introduce visual target tracking into virtual software platform, the specific work of its introduction is briefly introduced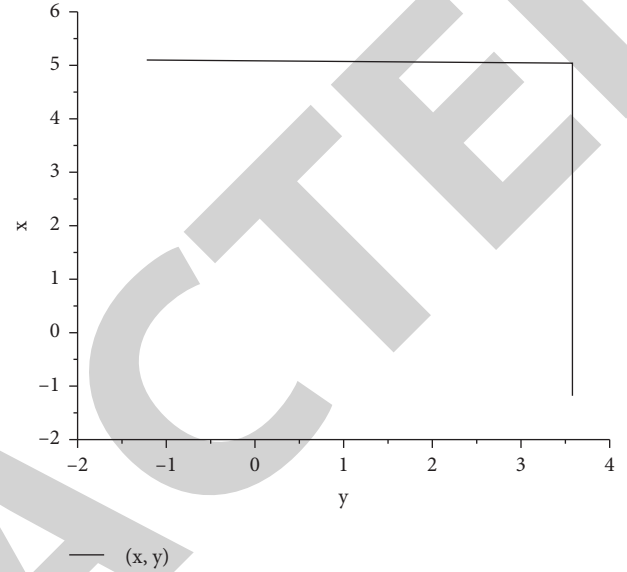, and a theoretical analysis is made on the image location and the distance and depth reflection. And according to the six set gestures, the feature values are given, as a fixed gesture template; therefore, through the recognition of the set virtual gesture, the corresponding piano key pronunciation is realized, which achieves the purpose of the research work in this paper [24].

5. Conclusion

According to what the author wants to achieve is to use a monocular CCD camera to take pictures of the target, the collected image is sent to the computer for processing, and the feature information of the object is obtained by using the principle of camera imaging, providing information for the position and depth of virtual objects. Therefore, the author first realized the construction of image acquisition and processing hardware; on the basis of the development kit provided by the image acquisition card, the image acquisition is realized by using VC++ programming, and preprocessing and image segmentation of the collected images, the characteristic information of the image is obtained, and the color plane tracking is realized. Visual acquisition and image processing is introduced into the virtual reality development platform, and the integration with the original system data information is realized. The scene reflects the status information of the data glove. The virtual piano has 14 piano keys; for the implementation of piano performance, it is each piano key; once the piano key is greater than the set value, just call m_Wave.Load() to make a sound. According

to the CWave class in the object-oriented MFC class library under VC++, create an object m_Wave of class CWave; then according to the m_Wave.Load () function of the CWave class, the connection of the sound is realized. This work involves visual images and virtual reality technology and involves a lot of knowledge content. The research time and the level of knowledge acquired are limited, only for basic knowledge and visual applications under the virtual reality system, make tentative learning and exploration, so there are some deficiencies in the realization of the work of this paper. Because the camera used is often unstable, the captured image will be discolored and disturbed by noise; it has a bad influence on the image processing work. Because the image plane segmentation adopted by the author is relatively simple background, it needs to be improved for complex background. In addition, there are still inherent deficiencies in the image feature information of monocular vision; it is faced with the occlusion problem when the three-dimensional object rotates. Since the author is in the process of mapping the image feature information to the virtual scene, the target image distance reflection information under monocular vision is calculated according to the image pixels, so the depth reflection in the virtual scene can only take a similar value, and the accuracy is not high.

Data Availability

No data were used to support this study.

Conflicts of Interest

The author declares that there are no conflicts of interest with any financial organizations regarding the material reported in this manuscript.

Acknowledgments

References

[1] M. Xue, X. Liu, H. Wu, B. Yu, and F. Meng, "Construction of cu2+-doped ceo2 nanocrystals hierarchical hollow structure and its enhanced photocatalytic performance," *Journal of Materials Science: Materials in Electronics*, vol. 32, no. 23, pp. 27576–27586, 2021.

[2] Y. Zhang, "Vr technology to the adjustment of piano playing mentality," *Mathematical Problems in Engineering*, vol. 2021, no. 1, 12 pages, Article ID 5844118, 2021.

[3] A. Ramírez, "Broken things, or silent suffering, and: the doctor, the curandero, and the virgin, or, is magic possible?" *Dialogos*, vol. 21, no. 1, pp. 123–127, 2018.

[4] X. Zhang, L. Geng, Y.-Z. Zhang et al., "Construction of cu-based mofs with enhanced hydrogenation performance by integrating open electropositive metal sites," *CrystEngComm*, vol. 21, no. 36, pp. 5382–5386, 2019.

[5] A. Samsudeen and M. M. Sulphey, "Construction of a tool to examine the attitude of employees towards performance management system," *Journal of Advanced Research in Dynamical and Control Systems*, vol. 9, no. 14, pp. 2249–2259, 2017.

[6] M. Lee, G. Jeon, I. Lee, T. Kim, B. Kim, and T. Lim, "A study on hil-based simulation environment construction for evaluation of autonomous driving performance," *The Journal of Korean Institute of Communications and Information Sciences*, vol. 46, no. 1, pp. 95–101, 2021.

[7] D. Zhu and X. J. M. Z. Zhan, "Efficient whole-cell biocatalyst for Neu5Ac production by manipulating synthetic, degradation and transmembrane pathways," *Biotechnology Letters*, vol. 39, no. 1, pp. 55–63, 2017.

[8] F. Khan, S. Leem, and S. Cho, "Hand-based gesture recognition for vehicular applications using ir-uwb radar," *Sensors*, vol. 17, no. 4, p. 833, 2017.

[9] S. Kim, Y. Ban, and S. Lee, "Tracking and classification of in-air hand gesture based on thermal guided joint filter," *Sensors*, vol. 17, no. 12, p. 166, 2017.

[10] V. J. Traver, P. Latorre-Carmona, E. Salvador-Balaguer, F. Pla, and B. Javidi, "Three-dimensional integral imaging for gesture recognition under occlusions," *IEEE Signal Processing Letters*, vol. 24, no. 99, 1 page, 2016.

[11] Q. Zheng, X. Tian, S. Liu, M. Yang, and J. Yang, "Static hand gesture recognition based on Gaussian mixture model and partial differential equation," *IAENG International Journal of Computer Science*, vol. 45, no. 4, pp. 569–583, 2018.

[12] S. Joglekar, H. Sawant, A. Jain, P. Dhadda, and P. Sonawane, "A multi-modular approach for gesture recognition and text formulation in american sign language," *International Journal of Computer Applications Technology and Research*, vol. 9, no. 7, pp. 217–224, 2020.

[13] T. Perrett, M. Mirmehdi, and E. Dias, "Visual monitoring of driver and passenger control panel interactions," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 2, pp. 321–331, 2017.

[14] A. Wibisono, "A training gesture-based-scroll visual artificial intelligence and measuring its effectiveness using hidden-markov modeling methods," *Journal of Information Systems and Informatics*, vol. 2, no. 1, pp. 163–168, 2020.

[15] T. Mo and P. Sun, "Research on key issues of gesture recognition for artificial intelligence," *Soft Computing*, vol. 24, no. 8, pp. 5795–5803, 2020.

[16] J. Zhang, K. Shao, and X. Luo, "Small sample image recognition using improved convolutional neural network," *Journal of Visual Communication and Image Representation*, vol. 55, no. AUG, pp. 640–647, 2018.

[17] J. Wan, Q. Ruan, W. Li, and S. Deng, "One-shot learning gesture recognition from rgb-d data using bag of features," *Gesture Recognition*, vol. 14, no. 1, pp. 329–364, 2017.

[18] B. Bansal, "Gesture recognition: a survey," *International Journal of Computer Application*, vol. 139, no. 2, pp. 8–10, 2016.

[19] A. I. Maqueda, C. R. del-Blanco, F. Jaureguizar, and N. Garcia, "Temporal pyramid matching of local binary subpatterns for hand-gesture recognition," *IEEE Signal Processing Letters*, vol. 23, no. 8, pp. 1037–1041, 2016.

[20] T. Song, H. Zhao, Z. Liu, H. Liu, Y. Hu, and D. Sun, "Intelligent human hand gesture recognition by local-global fusing quality-aware features," *Future Generation Computer Systems*, vol. 115, no. 7043, pp. 298–303, 2021.

[21] J. Yang, H. Zou, S. Cao, Z. Chen, and L. Xie, "Mobileda: towards edge domain adaptation," *IEEE Internet of Things Journal*, vol. 71 page, 2020.

[22] S. M. Aslam and S. Samreen, "GEsture recognition algorithm for visually blind touch interaction optimization using crow search method," *IEEE Access*, vol. 8, no. 99, 1 page, 2020.

[23] G. Panis, A. Lanitis, N. Tsapatsoulis, and T. F. Cootes, "Overview of research on facial ageing using the fg-net ageing database," *IET Biometrics*, vol. 5, no. 2, pp. 37–46, 2016.

[24] B. A. Delail, H. Bhaskar, M. J. Zemerly, and M. Al-Mualla, "Robust likelihood model for illumination invariance in particle filtering," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, no. 99, 1 page, 2017.