International Journal for Housing Science and Its Applications
Publish August 6, 2025. Volume 46, Issue 3 Pages 4745-4757

https://doi.org/10.70517/ijhsa463406

# Exploration of virtual scene creation and its interactive teaching mode of music performance based on augmented reality technology

**Jingjin Zhang**[1,*]
[1] Luoyang Weishusheng Middle School, Luoyang, Henan, 471000, China
Corresponding authors: (e-mail: zhangjingjinyx@126.com).

**Abstract** The progress of education has made higher education workers realize the importance of educational reform, and the teaching of music performing arts has an important role in cultivating students' music, performance and many other skills. In this paper, under the guidance of augmented reality key technology, the music performance scene image is calibrated in the pixel domain, and the image is further preprocessed using real-time human body keying principle based on HSV to meet the requirements of three-dimensional scene construction. Combined with the principle of three-dimensional reality fusion, double-sided camera and related software, the virtual scene of music performance is created. In view of the current situation of music performance teaching in colleges and universities, by integrating the music performance virtual scene into the traditional teaching mode, an interactive teaching mode based on augmented reality technology is designed, and the mode is verified and analyzed. After the intervention of the teaching mode in this paper, the teaching effect ($P=0.005<0.05$), participation ($P=0.004<0.05$), and positivity ($P=0.008<0.05$) show significant differences, while there is no significant difference in the traditional teaching mode, which indicates that the addition of the music performance virtual scene in the traditional teaching mode is more favorable to the improvement of the level of students' music performance.

**Index Terms** augmented reality technology, virtual scene, HSV, teaching mode

## I. Introduction

With the continuous progress of people's life, people's demand for the spiritual world is getting higher and higher. In order to satisfy people's ever-increasing spiritual needs, China's cultural industry is constantly enriched and developed, especially in today's cultural industry, which has been constantly utilizing a variety of emerging technologies to obtain breakthrough progress [1]-[3]. The performing arts are the foundation of China's cultural industry, and the performing arts can continuously transmit the social value of art to the audience through various forms [4]. Therefore, China is trying to make breakthroughs in performing arts through the integration of resources between multicultural fields [5]. Stage is the carrier of artistic performance, a good artistic performance can not be separated from a good lighting set and stage construction technology [6]. The modernized stage construction of music performance has continuously integrated many new media technologies, especially the integration of virtual reality (VR) technology into music performance, which has become the trend of stage development in the new era [7]-[9].

The application of virtual reality and augmented reality technologies in the field of music performance is creating a new way of music experience [10]. Analyzed from the perspective of musicology, these immersive technologies not only change the form of music presentation, but also redefine the interaction between performers and audiences [11]. In virtual reality concerts, the audience can immerse themselves in the three-dimensional sound field and even freely choose the viewing angle and position, which greatly enhances the sensory experience of music [12], [13]. Augmented reality technology, on the other hand, allows virtual elements to be superimposed onto the real world, such as adding visualized musical elements or virtual musicians to a live performance, thus enriching the visual effect of the performance [14]-[16]. The application of these technologies has opened up new possibilities for musical performances, such as cross-regional virtual ensembles and interactive musical installations [17]. However, such technology-mediated musical experiences have also raised some thoughts, such as the problem of maintaining the realism and emotional delivery of musical performances in virtual environments, and the question of whether technological enhancements distract listeners from the music itself [18]-[21]. Scholars of musicology need to investigate how these new technologies affect the creative thinking, performance forms, and reception aesthetics of music, as well as how to balance technological innovations with the essential needs of musical art [22]-[25].

In this paper, the images of the music performance scene are calibrated, and in order to ensure the usability of the images, the real-time human body keying method based on HSV is used to process them. After completing the image processing work, a three-dimensional stereo vision based on binocular simulation camera is proposed to perfect the high-precision three-dimensional virtual reality fusion visual effect. In order to improve the teaching quality of music performance in colleges and universities, the interactive teaching mode based on augmented reality technology is designed from the virtual scene of music performance and the traditional teaching mode, and the teaching mode validation program is formulated. According to the research program of this paper, the interactive teaching mode is validated and explored and analyzed.

## II. Music Performance Virtual Scene Creation

### *II. A.Augmented Reality Key Technologies*

To establish an augmented reality environment, it is necessary to comprehensively consider human perception, graphic generation and many other aspects, but the basic principles of the design have the following aspects:

(1) Maintain the original shape information of computer-generated graphics.

(2) Use high simulation graphics to make the augmented effect more realistic and natural.

(3) Give full consideration to the registration accuracy of the augmented reality system.

(4) The rationality of the design for specific tasks.

Therefore, augmented reality includes the following key technologies.

#### II. A. 1) Display technology

The most basic problem of augmented reality system design is to realize the integration of virtual information and the real world, and display technology is one of the basic technologies of augmented reality system [26], [27]. In general, display technologies for augmented reality can be categorized as follows:

Helmet monitor display, ordinary monitor display, handheld monitor display, and projection display.

(1) Helmet monitor display (HMD)

A helmet display can be viewed as a special kind of graphics rendering display. Its point of view is determined by a tracking system. The following requirements are to be fulfilled by using a helmet display:

a) The position and orientation of the helmet display is determined by the tracking system.

b) Parameters for the position of the viewpoint in the simulation environment relative to the position of the manipulated object need to be obtained.

c) The position of the manipulated object in the simulation environment needs to be obtained.

The helmet display can be divided into optical perspective display mode and video perspective display mode. Optical perspective helmet display places an optical synthesizer in front of the user's eyes. The optical synthesizer is partially transparent and allows the user to see directly into the real world. The synthesizer is also partially reflective, allowing the user to see a virtual image of the real world as it reflects off of a display worn on the head. A video see-through helmet display closes off the view with one or two cameras that capture real-world scenes. The scene synthesizer is responsible for synthesizing the camera video and graphics and sending the result to the display in front of the user's eyes.

(2) Normal monitor display

The use of ordinary monitor display is the simplest display in augmented reality, but also the most used way, the principle of using this way to the camera to obtain the real world image information and computer-generated virtual image information splicing, synthesis, and then output on the monitor.

(3) Handheld display

This is a flat LCD monitor, the use of bundled cameras to provide video-based perspective enhancement. The handheld display acts as a window or magnifying glass to display real objects covered with AR.

(4) Projection Display

Projective displays project the desired virtual information directly onto the real object for enhancement, in the simplest case projecting the enhancement information directly onto the object surface.

#### II. A. 2) Registration techniques

Registration is one of the key technologies of augmented reality, and registration is actually the process of stitching computer-generated virtual objects with the real environment obtained from the camera [28]. Registration first determines the positional relationship between the virtual object and the observer, and then projects the virtual object into the observer's field of view through projection transformation. Registration is generally categorized into dynamic registration and static registration. Dynamic registration is to determine the relative position of the camera and the real object in the case of relative motion. Static registration is the case where the camera and the real object are relatively stationary to determine the relative position of the two. In current augmented reality systems, registration techniques can be generally categorized into three types, tracker-based registration techniques, cognitive-based registration techniques and computer vision-based registration techniques.

**II. A. 3) Interface and visualization techniques**

There are two trends in the research of AR interaction means: first, using different devices and taking the best of each. The second is to make virtual objects and the real world a whole through practical interfaces [29]. For displaying information in AR displays, the following problems are usually addressed:

(1) Positioning errors in AR technology are hard to avoid. There are two ways to solve this: one is to visualize an area in the screen to draw the object based on the predicted tracking and measurement errors. The second is to gradually fade out the hidden virtual objects along the edges of the occluded area when drawing virtual objects that are occluded by real objects so that the localization errors are reduced.

(2) An excessive amount of virtual information data can make the enhanced display appear confusing and reduce readability.

(3) Improve the drawing quality of virtual objects to increase their realism.

In AR applications, people often aspire to more natural interaction methods, but it is very difficult to interact with virtual information in the real world, and now the following 3 main methods are used.

(1) Menus: mostly used in Pocket PC AR applications.

(2) Special markers: used to fix special markers on the device used for interaction.

(3) Special tools: generally simple in shape, easy to recognize, and can trigger some system events by pressing keys.

## *II. B.Scene Design*

**II. B. 1) Multi-sensor pixel domain calibration**

Real Interaction Scene and Multi-Sensor System As shown in Fig. 1, the real interaction scene described in this paper consists of a foreground entry body, a multi-sensor system, and a backdrop green screen, and the designed multi-sensor system is able to capture real-time RGB-D-based binocular real-scene information in real time, including a Kinect with an infrared sensor and two high-definition cameras.

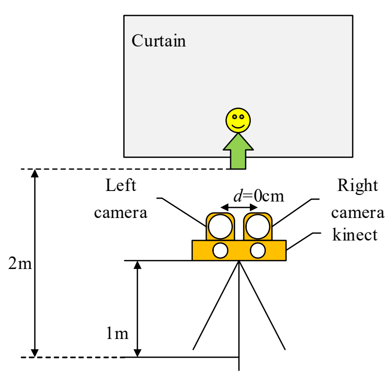



Figure 1: Real interaction scenarios and multi-sensor systems

It can be noted that the pixel domain conversion between different sensors is actually the conversion of different projection matrices $M$, and the conversion relationship between two projection matrices can be considered as a linear transformation relationship. As a result, this paper skips the complicated process of solving the internal and external camera parameters, and extracts the pixel coordinate transformation relationship between two different pixel domains $I_1$ and $I_2$ as Eq. (1):

$$\begin{bmatrix} u_1 \\ v_1 \\ 1 \end{bmatrix} = M_T * \begin{bmatrix} u_2 \\ v_2 \\ 1 \end{bmatrix} = \begin{bmatrix} a & b & c \\ d & e & f \\ 0 & 0 & 1 \end{bmatrix} * \begin{bmatrix} u_2 \\ v_2 \\ 1 \end{bmatrix} \quad (1)$$

The transformation matrix $M_T$ is a three-dimensional matrix whose third row is the vector $[0,0,1]$ unchanged after actual parsing, and $a$, $b$, $c$, $d$, $e$, and $f$ are the six unknown transformation parameters to be solved.

The prediction pixel point error is further reduced during the scaling process of the pixel conversion, which can be reduced to one pixel gap. Accurate multi-sensor calibration can delineate a more accurate foreground region for subsequent human keying and facilitate the data extraction of the real scene during the 3D fusion process.

### II. B. 2) HSV-based real-time human keying

HSV-based real-time human keying, as a widely adopted human keying technique in virtual scenes of music performances, is characterized by easy construction as well as simple computation.

As a result, the green screen keying algorithm proposed in this paper converts the image processing space to the HSV space that has a clear definition of lighting, a hexagonal color model consisting of hue (H), saturation (S), and brightness (V). The process of RGB and HSV color model conversion in this paper is as follows:

(1) Firstly, the $(R,G,B)$ three-channel data are normalized to obtain $(R_0,G_0,B_0)$ and define $C_{\max}$ and $C_{\min}$ as the maximum and minimum values in $(R_0,G_0,B_0)$ and $C_0$ as the difference between $C_{\max}$ and $C_{\min}$;

(2) For $H$, as in Eq. (2) is calculated as:

$$H=\begin{cases}0^{\circ}, & C_0=0\\ 60^{\circ}*\left(\dfrac{G_0-B_0}{C_0}\right), & C_{\max}=R_0\\ 60^{\circ}*\left(\dfrac{B_0-R_0}{C_0}+2\right), & C_{\max}=G_0\\ 60^{\circ}*\left(\dfrac{R_0-G_0}{C_0}+4\right), & C_{\max}=B_0\end{cases} \tag{2}$$

(3) For $S$, as in Eq. (3) is calculated as:

$$S=\begin{cases}0, & C_{\max}=0\\ \dfrac{C_0}{C_{\max}}, & C_{\max}\neq 0\end{cases} \tag{3}$$

(4) For $V$, as in Eq. (4) is calculated as:

$$V=C_{\max} \tag{4}$$

In HSV color space can be more directly on the shadow region to determine the rejection, this is because the shadow region did not change the physical material information of the scene, the scene before and after the shadow of the hue information has not been altered, the brightness of the obvious changes, and for the saturation, purely the intensity of the light changes will not directly affect the saturation, the saturation of the change will be stable in a small range.

### II. B. 3) Three-dimensional reality fusion

Before constructing stereoscopic vision, the first place is to realize the basic three-dimensional virtual-reality fusion in augmented reality. In the research scenario of this paper, the interaction space of augmented reality mainly consists of three parts: the human body in the real scene, the virtual interaction scene, and the virtual interaction object, and the three-dimensional fusion process is shown in Figure 2. The three-dimensional fusion process is actually the definition of the interaction space and the process of three-dimensional registration of the virtual and real scenes, and the real scene state is highly restored in order to effectively ensure that the three-dimensional fusion of fidelity, which requires a more accurate data analysis of the real scene.

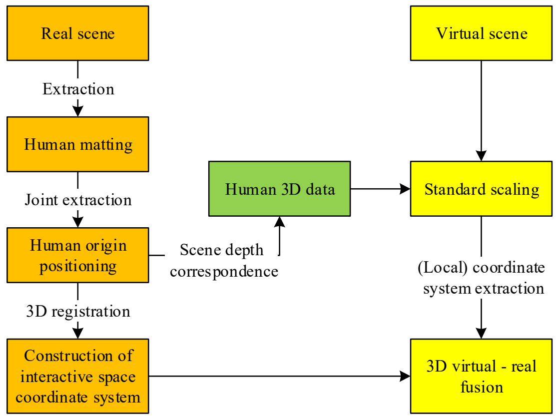



Figure 2: The process of virtual-real three-dimensional fusion

(1) Three-dimensional registration of real scenes

Compared with the two-dimensional real scene projection, the interaction scene studied in this paper is obviously a three-dimensional broadcast space. In order to fully restore the user's interaction effect and visual experience in the real scene, this paper introduces the depth information in the real scene, expands the local coordinate system of the planar real scene projection into a three-dimensional spatial coordinate system and takes it as the interaction spatial coordinate system. As mentioned before, in the interaction scene in this paper, the human interaction area is located between the background green screen and the multi-sensor system, with reference to the human body positional relationship, this paper defines the depth direction of the multi-sensor system to the human body in the real scene as the direction of the positive $z$ axis. Therefore, the 3D registration coordinate of the real scene in the interaction space is the spatial origin, which is a projection plane perpendicular to the positive $z$ axis of the space.

(2) Virtual scene 3D registration

Previously proposed Kinect color joint pixel coordinates can be easily converted to the human body keying map, this paper will be the head joints to the human body keying local coordinate system origin (the midpoint of the joints of the feet) distance is defined as the image of the human body height $d$. Based on the camera projection imaging principle and combined with the actual music performance scene in the human-computer distance relationship and camera attributes can be derived from the actual human body height, as shown in equation (5):

$$h=\frac{d_{is}}{h_c+d_c*\tan\theta/2}*\left(h_c+d_c*\tan\theta/2\right)=\frac{d_{is}*2d*\tan\theta/2}{hei} \tag{5}$$

where, hei is the height of the human body image; $d_c$ is the standard human-computer distance, which can be obtained by Kinect real-time acquisition of human head depth information, in this paper the value is at most 2m; $h_c$ is the height of the binocular camera system off the ground, which is taken to be 1m in this paper; and $\theta$ is the camera's viewing angle, which is 110 degrees in this paper.

The formula for calculating the virtual scene object size that is positively and linearly correlated with the actual human height can be simply expressed as equation (6):

$$s_{space}=\gamma h \tag{6}$$

$\gamma$ is the coefficient, which is 0.8 in this paper, which relies on the standard scene model library constructed in this paper, which contains hundreds of common broadcast space and broadcast interactive object models, uniformly produced in accordance with the standard human height ratio, and can maximally restore the physical properties of the space and interactive objects in the real scene and ensure a high degree of fidelity after adjusting the dimensions according to the actual height of the human body.

#### II. B. 4) 3D stereo vision based on binocular camera

In this paper, a 3D stereo vision based on binocular simulation camera is proposed as a visual complement to high-precision 3D reality-virtual fusion, which optimizes the 3D registration of the real scene elements with the virtual viewpoints, and 3D fuses the user's vision with the interaction space in order to achieve a high simulation of visual simulation.

(1) Real Scene in Stereoscopic Vision

Under stereoscopic vision, the left and right real scenes in the interactive space actually exist as the projection plane of visual imaging, and the width of the projection plane changes with the distance between the human and the machine in the actual broadcast scene, which can be deduced by the principle of camera imaging, as shown in equation (7):

$$w_{project} = 2 * d_c * \tan \theta/2 \quad (7)$$

where $d_c$ is the actual human-computer distance in the real scene. The projection plane height can be easily obtained based on the camera imaging aspect ratio, as shown in equation (8):

$$h_{project} = w_{project} * \frac{hei}{wid} \quad (8)$$

In this paper wid is the resolution width taken as 1920 and hei is the resolution height taken as 1080.

(2) Point-of-view simulation in stereo vision

The most important thing in the construction of stereo vision is to realize the accurate simulation of the real human eye, in order to simulate the real human eye imaging, the virtual viewpoints must be set as the left and right viewpoints respectively, and the three-dimensional registered coordinates of the left and right virtual viewpoints must meet the human-computer relationship in the real scene. The simulation of left and right viewpoints under stereoscopic vision constructs binocular vision for the user in the interaction space, thus increasing the visual and interaction space fusion on the basis of virtual-reality fusion, which enables the user to have a similar sense of presence with virtual reality. At the same time, three-dimensional vision can provide users with more observation details, which improves the visual logic of the virtual-reality scene in the interaction space and the simulation of three-dimensional fusion.

## III. Interactive teaching model and validation program design

With the advancement of the tide of information revolution, augmented reality technology is more and more applied to all walks of life in the society, which also drives the education industry to develop better and faster. Along with the universal application of augmented reality technology in the field of education, it will lead to the overall changes in the field of education from the goal to the mode. To respond to the direction of teaching change and development, on the basis of augmented reality technology, the interactive teaching mode based on virtual scene is proposed, and a responsive validation program is developed. Details are as follows:

### *III. A. Interactive teaching model*

#### III. A. 1) Creating a learning environment

In order to create different learning atmospheres for students according to their different learning stages, and to make the greatest support in the external environment, so that students can thoroughly enter the learning atmosphere from their senses and be completely immersed in the learning of music performance knowledge. Augmented reality technology can clearly show all kinds of music scenes in the virtual space, music learners can always interact with a variety of musical instruments in the virtual music performance environment, to obtain the maximum control and operation of the entire music learning environment. Comparing with the existing traditional music teaching forms, it seems that the application of augmented reality technology in teaching provides more learning opportunities for learners. No longer limited to a fixed time and a fixed classroom, in the virtual performance environment, learners can freely choose the time and place to conduct learning activities. In the era of continuous dissemination and development of new music knowledge, augmented reality technology provides a broad way and space for the renewal and development of music education.

#### III. A. 2) Enrichment of teaching methods

Augmented reality technology can improve the existing teaching mode, turn students' passive learning into active learning, and enrich the teaching methods. Augmented reality technology also provides deep interactive teaching, opening up the connection between students and teachers, between students and students, and between students and society and nature, so that students are not just learning knowledge, but also can apply it flexibly in communication and interaction, and students can motivate each other, help each other, and experience the fun of helping each other in learning, for example, to produce accurate musical performances or demonstrations, and to present to each learner the art of musical performances that are difficult to see in daily life by means of virtual reality technology to present the art of music performance, which is difficult to see in daily life, to every learner. This interactive and intuitive teaching method is conducive to stimulating learners' interest and enthusiasm in learning music, thus improving the quality and effect of teaching and learning.

#### III. A. 3) Subjective initiative

In traditional teaching, due to the limitation of teaching conditions and resources, teachers can only take the public as the standard in designing teaching contents and teaching methods, and cannot fully consider the real situation of each person, and the concept of teaching according to students' abilities can only remain in theory. Augmented reality technology can create personalized teaching, different students customize different learning modes, choose the most suitable for themselves and the most effective learning methods, for example, different colors of light-sensitive images will flow with the music playback, learners in the process of listening to the music to accurately capture the different colors of the image to make clear the characteristics of the music, so as to achieve the virtual music performance scene to improve the training of music listening skills. Students as the main body to actively choose to learn, rather than passive acceptance. Augmented reality technology will be a great help for the future development of education, and its advanced technical features and advantages can improve the shortcomings of traditional teaching, effectively through the information technology teaching means, mobilize the students' enthusiasm for learning, play the students' subjective initiative, enrich the teaching mode and teaching methodology, and guide the direction of the development of education reform.

### *III. B. Verification Program Design*

With the continuous development of information technology, augmented reality technology has become a widely used technology. Augmented reality technology has also begun to be widely used in college music performance teaching and training. Augmented reality technology brings many benefits to college music performance teaching and training, such as improving teaching effect, increasing students' participation and enthusiasm. The following experiment is designed to verify the application effect of the teaching mode in this paper.

#### III. B. 1) Purpose of the study

The purpose of this study is to investigate the application effect of reality enhancement technology in college music performance teaching and training, to compare the learning effectiveness of students under the use of reality enhancement technology and traditional teaching methods, and to verify the advantages of virtual simulation technology in college music performance teaching by comparing the learning effectiveness of traditional teaching with that of reality enhancement technology teaching, so as to provide reference for college dance teaching.

#### III. B. 2) Research subjects

The research subjects of this experiment are students majoring in music performance in a university, a total of 40, and the 40 students are randomly divided into 20 each in the experimental group and the control group. The experimental group uses the teaching mode of this paper for music performance training, and the control group uses the traditional teaching mode for music performance training. In the experimental design, the augmented reality technology used in the experimental group mainly includes display technology, three-dimensional scene modeling technology, animation rendering technology, etc., and the traditional teaching group uses the traditional face-to-face teaching method for teaching. At the end of the experiment, the experimental data will be analyzed and compared to determine the application effect of augmented reality technology in music performance training in colleges and universities.

#### III. B. 3) Experimental materials

(1) Music performance teaching materials
(2) Reality enhancement technology software
(3) Sensors
(4) 3D scene modeling software
(5) Animation rendering software

#### III. B. 4) Experimental Procedures

The specific processes involved in this experiment are as follows:

(1) Pre-experiment questionnaire.
(2) Students practiced basic dance movements.
(3) Students in the experimental group performed the interactive teaching mode based on augmented reality technology.
(4) Students in the control group were trained in the traditional teaching mode

(5) Collect training data.
(6) Learning effect test and questionnaire survey after the experiment.
(7) Data analysis and comparison.

#### III. B. 5) Experimental indicators

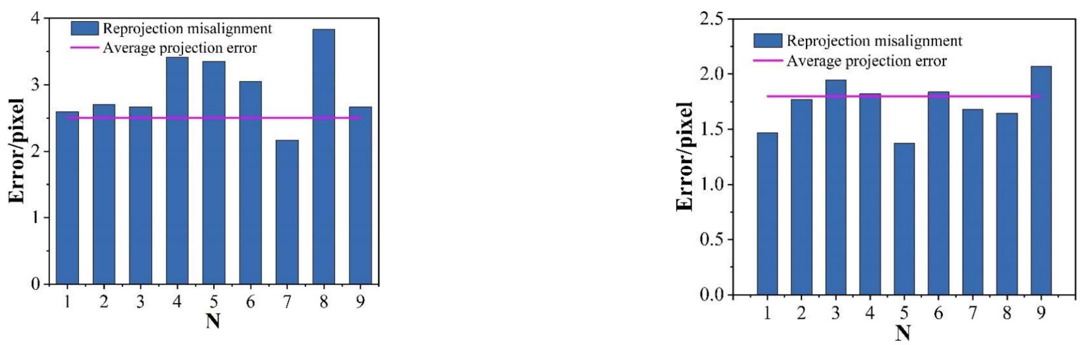
The indicators of this experiment include students' learning effectiveness, participation and learning interest, etc. Learning effectiveness is mainly assessed by scoring, and participation and learning interest are evaluated by questionnaire, based on the results of data analysis, which confirms the practical application effect of the teaching model in this paper.

## IV. Exploration of interactive teaching mode based on augmented reality technology

### *IV. A. Analysis of virtual scene simulation for music performance*

#### IV. A. 1) Simulation analysis of multi-sensor pixel domain calibration

Based on the previous research on multi-sensor pixel domain calibration, this section performs pixel domain calibration experiments by installing sensors, and then improves the error of projected fusion of images of music performance scenes based on this. In order to prevent the accident of the experiment, a total of 8 batches of experiments are carried out, and the final average projection error is taken, and the result of the reprojection error is shown in Figure 3, wherein (a) ~ (b) are the joint calibration of sensor external parameters and the method in this paper (multi-sensor pixel domain calibration). It can be concluded that multi-sensor pixel-domain calibration has significantly smaller reprojection errors than sensor-external parameter co-calibration, which provides a guarantee for the subsequent construction of virtual scenes for musical performances.




(a)External joint calibration of sensors (b)The method of this article

Figure 3: Reprojection error result

#### IV. A. 2) Real-time human keying simulation analysis

(1) Subjective evaluation

The human eye observation makes the image processing the most direct method of good or bad, commonly used subjective evaluation methods subjective score method and differential subjective score method. The average subjective score method is a number of observations which are made for the image quality evaluation of an average, there are generally five standards: excellent, good, medium, poor, poor corresponds to 5, 4, 3, 2, 1 a few points, respectively. Differential subjective score method is based on the average subjective score method, which:

$$d_{i,j} = MOS_{dehaze} - MOS_{haze} \tag{9}$$

$$d_{i,j} = \frac{d_{i,j} - \min\left(d_{ij}\right)}{\max\left(d_{i,j}\right) - \min\left(d_{i,j}\right)} \tag{10}$$

$MOS_{dehaze}$ is the subjective score of the processed image, $MOS_{haze}$ is the subjective score of the pre-processed image, $d_{i,j}$ represents the difference in the observer's subjective scores for the post-processed and pre-processed images, and $d_{i,j}$ represents the average of the differences in the subjective scores, i.e. the difference subjective score method results.

From the virtual scene of the music performance, 10 images are selected as the research object of this subjective evaluation analysis, and the subjective evaluation analysis of real-time human body keying is carried out under the role of the above formula, and the results of the subjective evaluation analysis are shown in Fig. 4, in which HF stands for the homomorphic filtering method, and CLAHE stands for the contrast-constrained histogram equalization method. It can be clearly seen that the subjective score of this paper's method is 4~5, indicating that the image quality reaches the needs of music performance scene design under the role of HSV-based real-time human keying.

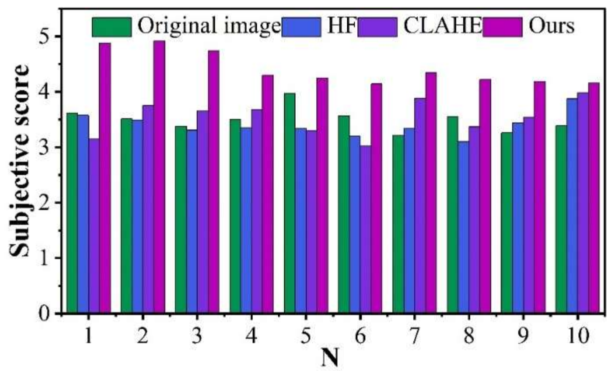



Figure 4: Subjective evaluation and analysis results

(2) Objective Evaluation

Objective evaluation of image quality refers to the use of several metrics related to image parameters, the establishment of relevant mathematical models, automatic calculation and derive the quality of the image, the common objective evaluation of image quality indicators are:

The larger the information entropy of an image is, the larger the amount of information contained in the image is, and the process of calculating the information entropy of a digital image is as follows:

$$S=-\sum_{i=1}^{256} p(i)\log p(i) \tag{11}$$

$i$ is accumulated from 1 to 256 and the entropy value is calculated. If the selected image is a color image, calculate the grayscale value of each channel on the HVB color space, count the number of pixels on a certain HVB color point, and find the probability of the appearance of the HVB color point.

The average gradient of the image indicates the size of the gray level change rate near the boundary or both sides, the larger the gray level change rate, the larger the rate of change of the image tiny detail contrast, and vice versa, the smaller the rate of change of the tiny detail contrast, which to a certain extent describes the degree of clarity of the image, so the larger the average gradient is, the clearer the image is, and the expression for the average gradient of the image is shown in Eq. (12):

$$\bar{g}=\frac{\sum_{x=1}^{M}\sum_{y=1}^{N}\sqrt{\frac{F^2(x)+F^2(y)}{2}}}{(M-1)\times(N-1)} \tag{12}$$

where $M$ and $N$ are the width and height of the image, respectively, $f(x,y)$ is a point on the image, $F(x)$ is the gradient in the horizontal direction, and $F(x)$ matrix has the first column as the difference between the second column and the first column of the original matrix, and the second column as the difference between the third and the half the difference of the first column, and the last column is the difference between the last two columns. The gradient in the vertical direction is $F(y)$ and so on.

The $Sobel$ operator is used to do edge detection on the image edges. The $Sobel$ operator is a discrete difference operator containing $3\times3$ matrices, which are plane-convolved in the image respectively, to obtain the edge detection images in the horizontal and vertical directions, with the following expressions:

$$G=\sqrt{G_x^2+G_y^2} \tag{13}$$

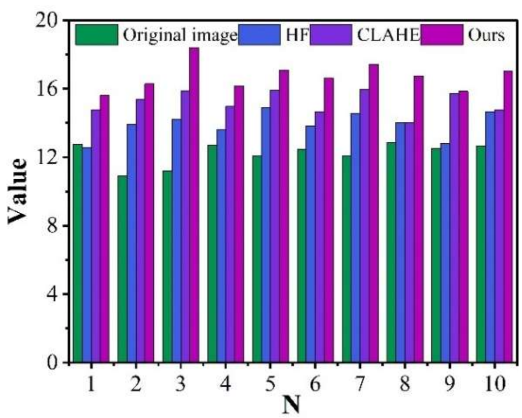
$G_x$ and $G_y$ represent the edge detection images in the horizontal and vertical directions, respectively.

The variance can reflect the degree of dispersion of the image gray level distribution, for a foggy image, the gray level distribution of the gray level image is often several in a certain region, while the gray level distribution of the fog-free image is more dispersed, in terms of this feature, the variance of the image can be a certain degree of response to the degree of image gray level dispersion. The expression of image variance calculation is shown in equation (14):

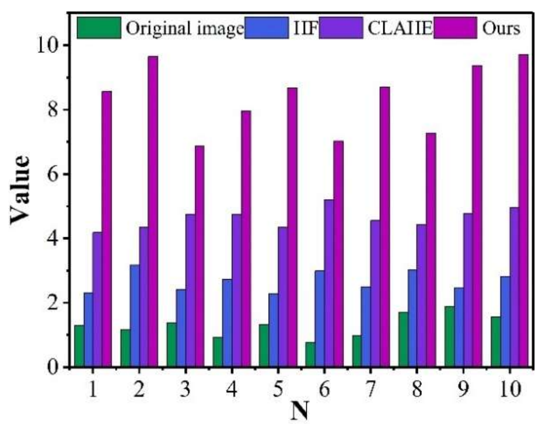
$$V = \frac{\sum_{i=1}^{M}\sum_{i=1}^{N}\left[f(x,y)-\overline{f}(x,y)\right]^2}{M \times N - 1} \tag{14}$$

where $\overline{f}(x,y)$ denotes the average gray value of the image.

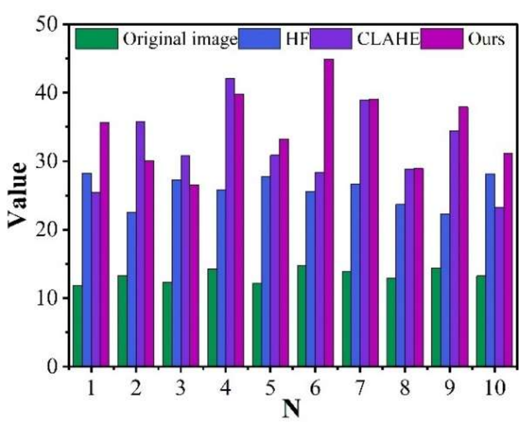
The information entropy, average gradient, edge strength and time-consuming are used as objective evaluation indexes to analyze the real-time human keying of different methods, and the results of the objective evaluation analysis are shown in Fig. 5, in which (a)~(d) represent the information entropy, average gradient, edge strength and time-consuming, respectively. It can be seen through the histogram performance, compared with other methods, the real-time human body keying based on HSV has the best performance in terms of information entropy, average gradient, edge strength, and time consuming, so that the processed image meets the requirements for the construction of three-dimensional virtual scenes of music performances.



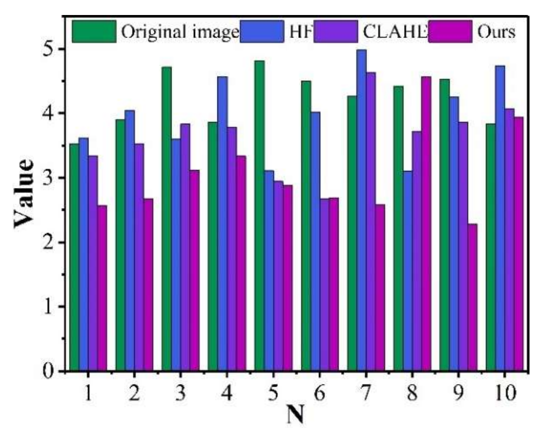

(a)Information entropy




(b)Average gradient




(c)Edge strength




(d)Time-consuming

Figure 5: Objectively evaluate and analyze the results

**IV. A. 3) Analysis of the Experiential Effects of Virtual Scenes for Music Performance**

Through the way of volunteers experiencing the virtual scene of music performance, the experience effect of the virtual scene of music performance constructed in this paper is scored, and the experience effect is divided into six dimensions Interactivity (it refers to the degree of human-computer interaction in the experimental process), sense of operation (it refers to whether or not there is a sense of operation in the experimental process using the real music equipment), observation of the effect (it refers to the degree of realism of the virtual music performance scene), intelligibility ( The degree of intelligence of the equipment of the virtual music performance scene), Exploratory (it refers to whether the virtual music performance scene supports exploratory experiments), and Interesting (whether the volunteers find it interesting). Adopting a 5-point system, the higher the score represents the better the evaluation of the indicator, the experiential effect of the virtual music performance scene is analyzed as shown in Figure 6. According to the numerical performance in the figure, it can be seen that the six dimensions of the virtual scene in this paper are evaluated as 3.515, 3.491, 3.453, 3.493, 3.507, 3.549, and all the data are above 3, which indicates that the volunteers show a positive and acceptable attitude towards the experiential effect of the virtual scene of music performance.

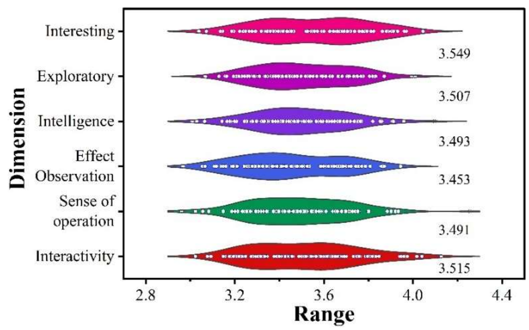



Figure 6: Analysis of Experience Effect

### *IV. B. Validation Analysis of Interactive Teaching Models*

#### IV. B. 1) Homogeneity test

The independent samples t-test was conducted to compare the pre-test values of each dimension index between the experimental and control groups to examine whether there was a significant difference between the experimental and control groups before the intervention, and the results of the homogeneity test are shown in Fig. 7, in which A1~A3 denote the teaching effect, participation, and positivity, respectively, while CG and EG denote the control and experimental groups. The data in the figure show that there is no significant difference (P>0.05) between the experimental group and the control group in teaching effect (P=0.208>0.05), participation (P=0.706>0.05), and positivity (P=0.722>0.05), so it can be regarded as subjects are homogeneous and experimental comparisons can be made.

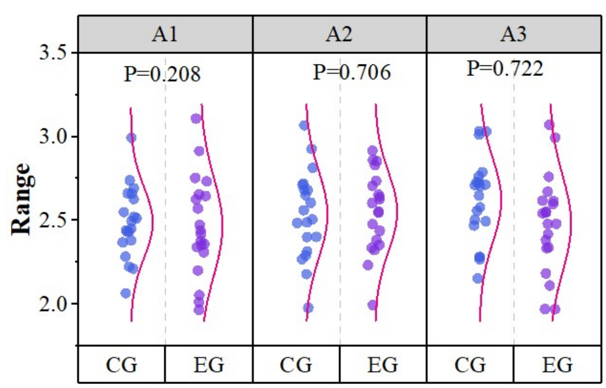



Figure 7: Homogeneity test results

#### IV. B. 2) Post-intervention difference tests

Independent samples t-test was used to analyze the post-intervention data of the experimental group and the control group for the test of difference, to examine whether there was a significant difference between the experimental group and the control group in each dimension index after intervention, and the results of the post-intervention test of difference between the experimental group and the control group are shown in Figure 8. As can be seen from the P-value in the figure, after the intervention, the experimental group and the control group showed significant differences (P<0.05) in teaching effect (P=0.003<0.05), participation (P=0.009<0.05), and positivity (P=0.001<0.05), which indicates that the teaching mode of this paper is more superior than the traditional teaching mode in disguise, and it has a facilitating effect on the teaching of music performance in colleges and universities.

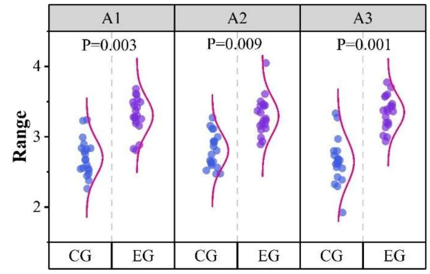



Figure 8: Post-intervention difference test

**IV. B. 3) Tests of within-group variability**

(1) Experimental group

The paired-samples t-test was used to examine the differences between the dimensional indicators of the members of the experimental group before and after the intervention of the teaching model in this paper, and the analysis of the difference test before and after the intervention of the experimental group is shown in Figure 9. After the intervention of the teaching model in this paper, the paired samples t-test of the experimental group in the pre- and post-tests of teaching effectiveness (P=0.005<0.05), engagement (P=0.004<0.05), and motivation (P=0.008<0.05) showed significant differences (***P<0.05). It proves the application value of interactive teaching mode based on augmented reality technology in the improvement of students' music performance level.

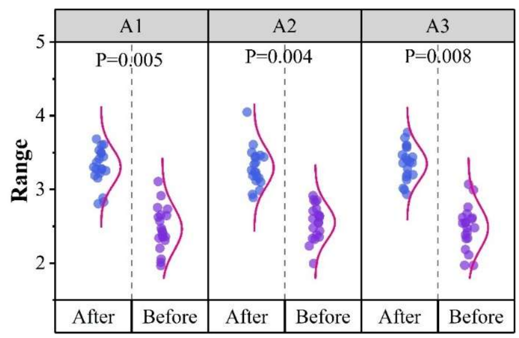



Figure 9: Difference test before and after the intervention in the experimental group

(2) Control group

Paired-samples t-test was used to examine the differences between the pre-test and post-test indicators of each dimension of the students in the control group, and the analysis of the difference test before and after the intervention in the control group is shown in Figure 10. Under the intervention of traditional teaching mode, the control group did not show significant differences in teaching effect (P=0.076<0.05), participation (P=0.058<0.05), and motivation (P=0.062<0.05), which indicates that the traditional teaching mode does not have a significant effect on the improvement of students' music performance level.

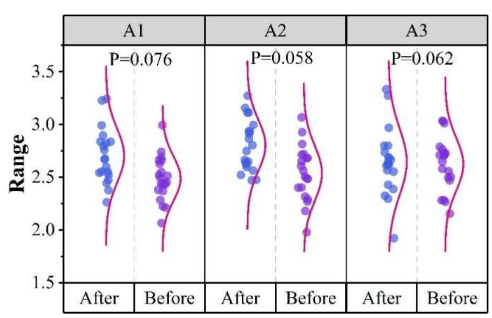



Figure 10: Difference test before and after the intervention in the control group

## V. Conclusion

In this paper, the virtual scene of music performance is designed with the help of augmented reality technology (AR), and in order to improve the level of students' music performance, the designed virtual scene of music performance is introduced into the teaching curriculum of colleges and universities, thus, an interactive teaching mode based on augmented reality technology is proposed. The actual classroom performance effect of the teaching mode in this paper is verified through scale testing and statistical analysis methods. Under the intervention of this teaching mode, it is found that the three dimensions of students' music performance level before and after the intervention show a significant difference, which meets $P<0.05$. When the traditional teaching mode is used for the experimental intervention, there is no significant difference in the students' music performance level, which can be summarized that compared with the traditional teaching mode, the interactive teaching mode based on the augmented reality technology has a more significant effect on the enhancement of the students' music performance level. It can be summarized that compared with the traditional teaching mode, the interactive teaching mode based on augmented reality technology has a more significant effect on improving students' music performance level.

## References


[1] Xie, X., Xie, X., & Martínez-Climent, C. (2019). Identifying the factors determining the entrepreneurial ecosystem of internet cultural industries in emerging economies. International Entrepreneurship and Management Journal, 15, 503-522.

[2] Bakhshi, H., & Throsby, D. (2012). New technologies in cultural institutions: theory, evidence and policy implications. International journal of cultural policy, 18(2), 205-222.

[3] Peukert, C. (2019). The next wave of digital technological change and the cultural industries. Journal of Cultural Economics, 43(2), 189-210.

[4] Huizhu, S. (2009). Performing arts and cultural identity in the era of interculturalism. TDR: The Drama Review, 53(2), 7-11.

[5] Wu, J., Jiang, K., & Yuan, C. (2020). Market-Oriented Reforms, Economic Efficiency and Commercial Performing Arts in China. The Journal of Arts Management, Law, and Society, 50(1), 33-51.

[6] Yang, H. (2019). Policies and Measures Related to China's Cultural Industry. Development of China's Cultural Industry, 19-48.

[7] Liu, W. (2020). Influence of Virtual Reality Technology on Audience's Performance Behavior of Stage Music Performance Art. International Journal of Multimedia Computing, 1(2), 1-12.

[8] Loveridge, B. (2020). Networked music performance in virtual reality: current perspectives. Journal of Network Music and Arts, 2(1), 2.

[9] Zyl, M. V. (2020). The effects of virtual reality on music performance anxiety among university-level music majors. Visions of Research in Music Education, 35(1), 15.

[10] Park, J., Choi, Y., & Lee, K. M. (2024). Research trends in virtual reality music concert technology: A systematic literature review. IEEE Transactions on Visualization and Computer Graphics.

[11] Loveridge, B. (2023). An overview of immersive virtual reality music experiences in online platforms. Journal of Network Music and Arts, 5(1), 5.

[12] Onderdijk, K. E., Bouckaert, L., Van Dyck, E., & Maes, P. J. (2023). Concert experiences in virtual reality environments. Virtual Reality, 27(3), 2383-2396.

[13] Charron, J. P. (2017). Music audiences 3.0: Concert-goers' psychological motivations at the dawn of virtual reality. Frontiers in psychology, 8, 800.

[14] Lee, J., Kim, Y., Heo, M. H., Kim, D., & Shin, B. S. (2015). Real-time projection-based augmented reality system for dynamic objects in the performing arts. Symmetry, 7(1), 182-192.

[15] Nagele, A. N., Bauer, V., Healey, P. G., Reiss, J. D., Cooke, H., Cowlishaw, T., ... & Pike, C. (2021). Interactive audio augmented reality in participatory performance. Frontiers in Virtual Reality, 1, 610320.

[16] Selfridge, R., & Barthet, M. (2019). Augmented live music performance using mixed reality and emotion feedback. In 14th Int. Symposium on Computer Music Multidisciplinary Research (Vol. 210).

[17] Li, W. (2022, February). Musical instrument performance in augmented virtuality. In Proceedings of the 6th International Conference on Digital Signal Processing (pp. 91-97).

[18] Cook, M. J. (2019). Augmented reality: Examining its value in a music technology classroom. Practice and potential. Waikato Journal of Education, 24(2), 23-38.

[19] Kiefer, C., & Chevalier, C. (2018). Towards new modes of collective musical expression through audio augmented reality. In Proceedings of the International Conference on New Interfaces for Musical Expression (pp. 25-28).

[20] Turchet, L., Hamilton, R., & Çamci, A. (2021). Music in extended realities. IEEE Access, 9, 15810-15832.

[21] Morrow, G., & Morrow, G. (2020). Conclusions: virtual reality, augmented reality and mixed reality. Designing the Music Business: Design Culture, Music Video and Virtual Reality, 177-201.

[22] Petrović, N. (2020, September). Augmented and virtual reality web applications for music stage performance. In 2020 55th international scientific conference on information, communication and energy systems and technologies (ICEST) (pp. 33-36). IEEE.

[23] Guclu, H., Kocer, S., & Dundar, O. (2021). Application of augmented reality in music education. The Eurasia Proceedings of Science Technology Engineering and Mathematics, 14, 45-56.

[24] Nijs, L., & Behzadaval, B. (2024). Laying the foundation for augmented reality in music education. Ieee Access.

[25] Lee, G. J. (2016). Research of Space Augmented Reality based Convergence Performance-Focused on Images used for Performing Arts. Journal of Broadcast Engineering, 21(5), 672-688.

[26] Permatasakti Damba & Basiroen Vera Jenny. (2018). Experimental Design: Javanese Folktale Preservation Using Augmented Reality Technology. Humaniora,9(3),265-265.

[27] Souad Rabah,Ahlem Assila,Elio Khouri,Florian Maier,Fakreddine Ababsa,Valéry bourny... & Frédéric Mérienne. (2018). Towards improving the future of manufacturing through digital twin and augmented reality technologies. Procedia Manufacturing,17,460-467.

[28] Zhang Xiaoying,Zhu Yonggang,Chen Lumin,Duan Peng & Zhou Meijuan. (2024). Augmented reality navigation method based on image segmentation and sensor tracking registration technology. Scientific Reports,14(1),15281.

[29] Jingang Jiang,Zhiyuan Huang,Wei Qian,Yongde Zhang & Yi Liu. (2019). Registration Technology of Augmented Reality in Oral Medicine: A Review. IEEE Access,7,53566-53584.