International Journal for Housing Science and Its Applications
Publish August 3, 2025. Volume 46, Issue 3 Pages 478-494

https://doi.org/10.70517/ijhsa46334

Virtual Reality Technology and Multiple Perceptual Model Fusion in Stage Character Design: A Computational and Algorithmic Approach to Reinventing Artistic Representation

Jing Zeng[1,*]
[1] Fashion Art (Fashion Performance and Design), School of Design, Jianghan University, Wuhan, Hubei, 430000, China
Corresponding authors: (e-mail: 18086117820@163.com).

Abstract With the continuous maturity of deep learning technology, its role in stage character design is becoming more and more important. The article proposes a deep neural network-based multi-view human tracking method for stage environment and designs the stage lighting control system. And on this basis, it designs an automatic stage lighting tracking system, including control and management module, actor identification and localization module, data processing module, and speed interpolation module, which facilitates the control of lighting on the stage. Finally, a series of experimental tests are conducted to verify the effectiveness of the method of this paper. The experimental results on simulated and real labeled datasets show that the binary cyclic code of this paper's method can still achieve more than 91% recognition accuracy under 60% occlusion rate, which has a very good anti-occlusion performance. Without affecting the stage performance and viewing experience, this paper's method solves the tracking instability problem caused by stage darkness and actor's apparent similarity, which is highly feasible and has a wide range of application prospects in the stage performance industry.

Index Terms deep neural network, human tracking, lighting control system, stage character design

I. Introduction

Nowadays, mankind is facing a change that has not been seen in a hundred years, the rapid development of science and technology, and the iteration of science and technology has had different degrees of impact on all kinds of fields and industries in the world. With the progress of computer technology and video technology, the field of application is becoming more and more extensive, and gradually developing into an interdisciplinary, cross-media combination of art [1]-[3]. Digital virtual characters in the virtual image technology support into the modern stage performance, although the research on digital virtual characters in the field of theater stage is still lack of in-depth understanding, but the field is constantly being explored, with the intervention of virtual reality technology (VR), to a certain extent, changed the traditional stage performance and the form of expression [4]-[7]. At the same time, it enriches the audiovisual language of modern and contemporary stage art, increases the narrative space, and changes the relationship between viewing and acting [8].

The application of virtual characters on the stage, compared with the traditional stage art performance, is a novel and unique form of expression, the application of this VR technology makes the performing arts more diversified while meeting the aesthetic needs of the audience [9]-[11]. This trend in the current domestic and international stage design field, has been widely concerned and discussed. The appearance of virtual characters on the stage also better adapts to the aesthetic needs of modern audiences [12]. In the early stage of application, the function of VR technology is only a simple scene display, and there are strong limitations in the scope of application. After years of practice and innovation, VR technology has become an important auxiliary means in the process of stage performance and artistic creation [13]-[15]. Especially in the field of stage performance, the application of this technology can break the limitations of space and time [16]. With the assistance of relevant equipment, it can make the audience produce “immersive experience”, so that they can produce physical and mental pleasure in a specific situation and obtain interactive experience [17], [18].

The article firstly developed a deep neural network-based multi-view human body tracking method for stage environment, using multiple cameras and YOLO algorithm to detect and locate the target, then applying KNN and RANSAC algorithms to extract the human body feature points, calculating the real-time position of the target and converting it into coordinates, and sending the results to the console. Then the advantages of IoT technology are fully utilized, and the system architecture containing perception layer, network layer, platform layer and application layer is constructed, and its functions are designed from three aspects of automatic dimming, scene switching and actor tracking respectively, and the stage lighting automatic tracking system is designed. Finally, the method

proposed in this paper is tested in simulation and real labeled dataset for performance experiments, and a questionnaire survey is carried out with 500 audience members as the research object to explore the application effect of this paper's method in stage character design.

II. Stage character tracking and lighting control algorithm design

II. A. Deep neural network based localization design

In order to realize the stage lighting localization and tracking function of deep neural network technology, it is necessary to detect and identify the target by YOLO algorithm firstly, and then extract the key points and feature points of the human body in each image with key point detection algorithm to form a matching point group for matching. If the process involves only one protagonist, the protagonist will be tracked by multiple camera devices to ensure the accuracy of the coordinates. If more than one protagonist is involved, the system will assign 2~3 cameras to each protagonist for tracking by multi-camera multi-target tracking method. To prevent too many cameras from tracking the same protagonist, the system will assign a unique identifier to the protagonist through pre-set parameters to ensure the uniqueness of the protagonist, and also ensure that the cameras can continuously track and locate the protagonist.

After the matching point groups have been matched, the 3D coordinates of the matching point groups and the corresponding 2D coordinates are calculated, so that the stage lights can be continuously illuminated on the corresponding coordinates. In order to avoid visual blind spots, all the cameras and stage lights are set up on the high side of the stage to ensure maximum coverage of the entire stage. All cameras are infrared cameras, which work with short-wave infrared to see details in low light. In addition, these cameras are equipped with automatic gain control (AGC), which ensures that the target can be detected and tracked even in low light conditions, meeting the requirement of accurate positioning under any conditions.

II. A. 1) Human body detection based on YOLOv5 algorithm

In order to ensure that the detection speed and detection quality of the system designed in this paper can meet the standards, the YOLOv5 algorithm in the convolutional neural network is selected. The YOLOv5 network architecture is shown in Fig. 1, which consists of four parts: input, backbone, neck, and head. From the architecture shown in Fig. 1, the input plays the key role of data preprocessing, which integrates Mosaic data enhancement, adaptive anchor frame adjustment, and image size normalization [19]. Overall, the connection between each module together enhances the performance of the deep learning model in the target detection task, realizing highly accurate and efficient target detection. The system designed in this paper is able to detect and recognize targets quickly and accurately in dynamic environments, making the system much more efficient.

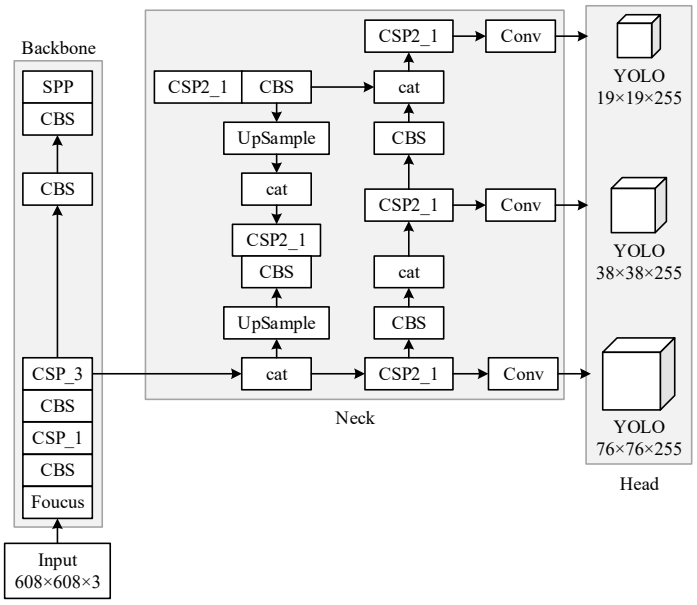

Figure 1: YOLOv5 Network Architecture

II. A. 2) Feature point extraction based on KNN and RANSAC algorithms

After recognizing and detecting the human body, in order to further improve the accuracy and stability when extracting feature points, the process of extracting feature points focuses on the key parts of the human body. In addition, the system automatically detects the features of each human body and generates feature points to assist detection. Subsequently, it is processed and recognized by image processing algorithms. The key points of the human body and the image feature points are extracted from the captured images, and through the respective matching point groups and 3D points, comprehensive processing is carried out so as to obtain the accurate distribution of the human body's ground projection.

In the matching process, each feature point is assigned the corresponding 3D coordinates one by one. By applying the Nearest Neighbor Rule Classification Matching Algorithm (KNN Algorithm), the distance between each sample, as well as the difference in their directions, can be calculated to minimize the error in the matching process.

The calculation of the distances follows equation (1):

$$d = \sqrt{\sum_{i=1}^{n} (A_i - B_i)^2} \tag{1}$$

where: n is the number of dimensions. A_i and B_i are the values of the two feature points in the i-dimensional space, respectively. d is the distance between the two samples.

In order to measure the difference in direction, equation (2) can be utilized to calculate the direction cosine value between two feature points:

$$\cos\theta = \frac{x_1 x_2 + y_1 y_2}{\sqrt{x_1^2 + x_2^2} \times \sqrt{y_1^2 + y_2^2}} \tag{2}$$

where: X_1, Y_1 is the coordinates of the 1st actor. X_2, Y_2 is the coordinates of the 2nd actor. $\cos\theta$ is the cosine of the angle formed by the two actors with the origin of the coordinates [20].

II. A. 3) Cross-dataset training and motion pose prediction

Cross-dataset training enables the stability and extensiveness of the system to be effectively improved by collecting data in different environments, different lighting conditions and different actor postures as a supplementary training set. And the motion pose prediction can be categorized into two cases. The first case is uniform motion, i.e., the actor's route in a certain time is uniform, in this case, the system can calculate the next coordinate point based on the 2D projection coordinates of previous feature points, as shown in Eq. (3) and Eq. (4):

$$x_{next} = x_{current} + \frac{t \times (x_{current} - x_{last})}{t_p} \tag{3}$$

$$y_{next} = y_{current} + \frac{t \times (y_{current} - y_{last})}{t_p} \tag{4}$$

where: $(X_{current}, Y_{current})$ is the current coordinate. (X_{last}, Y_{last}) is the last coordinate. (X_{next}, Y_{next}) is the coordinate of the next time. t is the next time to be predicted. t_p is the time sampling period.

II. B. System workflow

This paper proposes a multi-view human body tracking system, the system structure is shown in Figure 2, the multi-view visual tracking technology system is mainly composed of multiple professional-grade camera devices, servers with deep neural network processing capabilities, stage lighting, lighting consoles and through such as DMX512 or Artnet and other control buses work together.

First, the server receives images from multiple synchronized cameras, locates the human body using a human detection algorithm and creates a human detection frame. Then, combining these features, the algorithm server calculates the 3D coordinates of the key points of the human body and corrects for reprojection errors to improve accuracy. Subsequently the 2D stage coordinates of the performer are calculated. Finally these coordinates are converted into DMX commands, which the console receives and directs the lights to accurately illuminate the performer, the system workflow is shown in Figure 3.

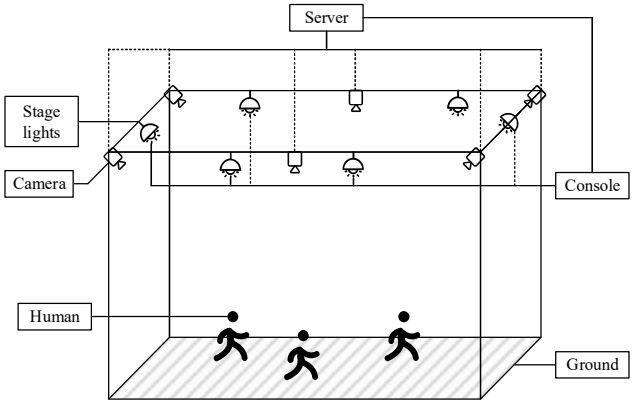

Figure 2: System Structure

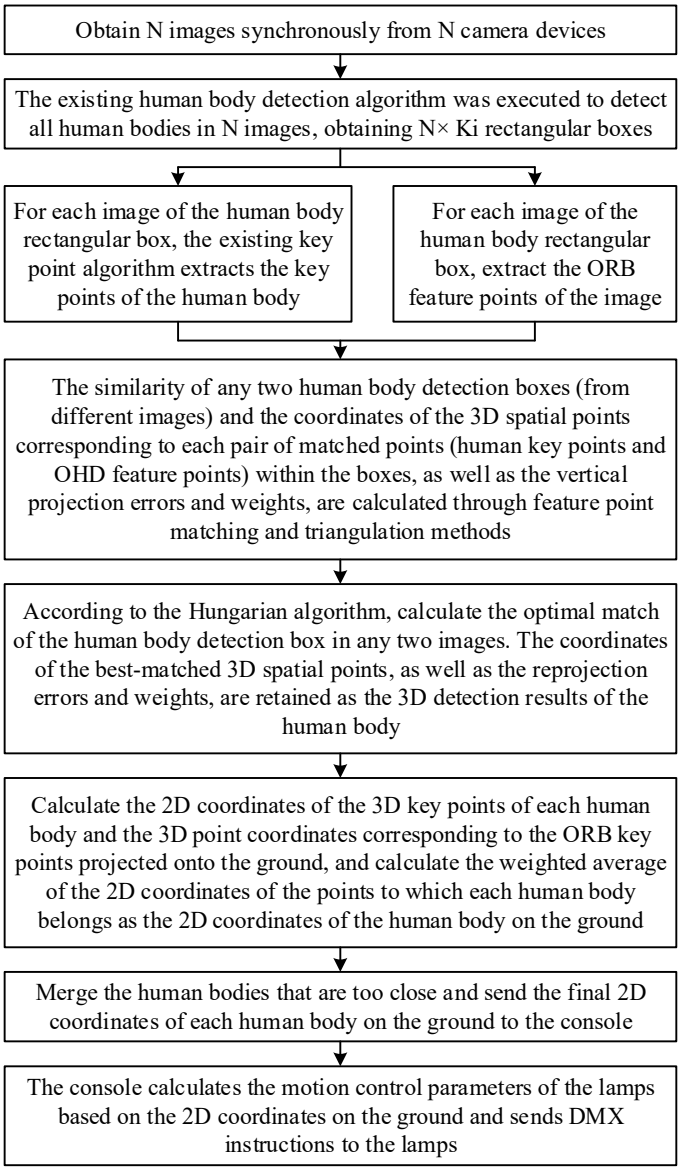

Figure 3: System Workflow

On the basis of the above workflow, the camera technology of the system also applies a flexible mechanism to cope with the challenges in the complex environment of the stage. In the working process, it is not possible to obtain

the pixel depth information through a single image, so it is necessary to have two or more cameras focusing on the same target to take a picture, and at this time, the system will accurately locate the target through the method of multi-view triangulation.

Wherein, when the 2 camera devices O_1 and O_2 are aligned to 1 target, there will exist two frames I_1 and I_2 at a certain point in time, R,t is the rotation matrix and translation vector of the second frame with respect to the first frame, K is the internal parameter matrix of the camera devices, and s_1 and s_2 are the depth information, i.e., the distance between the target and the camera devices. Assuming that there is a feature point P_1,I_2 in I_1 there is a feature point P_2, based on the above information it can be derived that the feature point P_1,P_2 is at I_1,I_2, points x_1 and x_2 in the plane on the frame image, respectively, with the relation Eq:

$$x_1 = K^{-1} \times p_1 \tag{5}$$

$$x_2 = K^{-1} \times p_2 \tag{6}$$

According to the opposite pole constraint, the following relation is available:

$$s_1 x_1 = s_2 R x_2 + t \tag{7}$$

The left and right sides of the equation are multiplied by $\hat{x}_1$ at the same time to obtain:

$$s_1 x_1 \hat{x}_1 = 0 = s_2 \hat{x}_1 R x_2 + \hat{x}_1 t \tag{8}$$

Finally, s_1 can be calculated from the equation in the left half of the equation, and s_2 can be calculated from the equation in the right half. However, in the actual pose estimation, it is also necessary to consider the error caused by noise, and the result in Eq. (8) is not necessarily 0, so in order to ensure the accuracy of the result, the system will calculate s_1 and s_2 by least squares.

When there are more than 2 cameras focused on a target, the target Y is observed in multiple key frames $k_1,k_2 \ldots k_n$ at a given moment. Similarly, the coordinates in the plane of the feature point taken from the observation in each frame are x_k, such that $x_k = [u_k \quad v_k \quad 1]^T$, and the projection matrix is $[R_k \quad T_K] \in R^{3\times4}$. where u_k, v_k when are the internal references of the camera device. Thus, when the real coordinates are swapped with the camera coordinates, the projection relationship is as follows:

$$\forall k, \lambda_k x_k = P_k Y \tag{9}$$

Where: Y is the location of the target to be solved.

Expanding the equation yields:

$$\lambda_k = \begin{bmatrix} u_k \\ v_k \\ 1 \end{bmatrix} = \begin{bmatrix} P_{k,1}^T \\ P_{k,2}^T \\ P_{k,3}^T \end{bmatrix} Y \tag{10}$$

This can be derived from the third line of the above equation:

$$\lambda_k = P_{k,3}^T Y \tag{11}$$

Where: $P_{k,3}^T$ means the third line of P_k.

Substituting the result of the third line for the first two lines subsequently gives:

$$u_k P_{k,3}^T Y = P_{k,1}^T Y \tag{12}$$

$$v_k P_{k,3}^T Y = P_{k,2}^T Y \tag{13}$$

The above two equations are the result of one observation. However, in actual performances, multiple observations are required, so after assuming that n observation has been made (i.e., $k = 1,2,\cdots,n$), the resulting equation is:

$$\begin{bmatrix} u_1 P_{1,3}^T - P_{1,1}^T \\ v_1 P_{1,3}^T - P_{1,2}^T \\ \vdots \\ u_n P_{n,3}^T - P_{n,1}^T \\ v_n P_{n,3}^T - P_{n,2}^T \end{bmatrix} Y = 0 \tag{14}$$

Let $D = \begin{bmatrix} u_1 P_{1,3}^T - P_{1,1}^T \\ v_1 P_{1,3}^T - P_{1,2}^T \\ \vdots \\ u_n P_{n,3}^T - P_{n,1}^T \\ v_n P_{n,3}^T - P_{n,2}^T \end{bmatrix}$, then you get $DY = 0$.

II. C. Software design

II. C. 1) IoT-based distributed lighting control system

For the stage distributed lighting control system architecture, this paper introduces the Internet of Things (IoT) technology in the design process, and divides the system architecture into four parts: perception layer, network layer, platform layer and application layer, and the stage distributed lighting control system architecture is shown in Fig. 4.

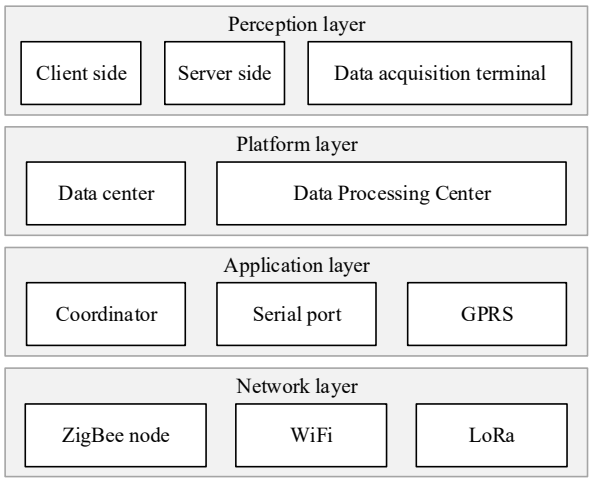

Figure 4: Framework of the distributed lighting control system

In the stage distributed lighting control system architecture shown in the figure, the perception layer is the data input source of the whole stage distributed lighting control system, and it monitors the stage lighting, sound effects, actors' positions and other information in real time through various sensors and devices. The network layer is responsible for transmitting the data collected in the perception layer to the data center. In order to ensure the stability and security of the data, this paper adopts the communication protocol mode of ZigBee+WiFi+LoRa. At the same time, in order to avoid data loss or large time delay, this paper adopts multi-channel backup transmission technology. The platform layer is the core of the stage distributed lighting control system, which mainly includes two parts: the data center and the data processing center. Among them, the data center is mainly responsible for receiving and storing the data transmitted from the network layer, and the data processing center processes, analyzes and mines the received data to extract the lighting control strategy and actor position information. In the stage distributed lighting control system designed in this paper, the application layer controls specific execution units according to the actual needs, including automatic dimming, scene switching, and remote control.

According to the above approach, the design of the distributed lighting control system architecture is realized to provide a reliable guarantee for its stable operation.

II. C. 2) Distributed lighting control system function design

In the stage of designing the function of distributed lighting control system, this paper, based on data monitoring and analysis, divides the function of a theater's stage distributed lighting control system into three main aspects, which are automatic dimming, scene switching and actor tracking [21]. In the automatic dimming function, the light brightness, color temperature and other parameters are automatically adjusted mainly according to the real-time monitoring of the stage light and performance requirements. In the specific realization process, first determine the difference between the stage light intensity and color temperature parameters monitored by the lighting sensor in real time and the preset standard values, the specific calculation is shown in Equation (15):

$$f(a) = a(t) - \frac{dF(t)}{d\Omega(t)} \tag{15}$$

where: $f(a)$ indicates the difference between the stage lighting status data monitored by the lighting sensor in real time and the preset standard value. $a(t)$ represents the stage lighting status data monitored by the lighting sensor in real time at the t moment. $F(t)$ and $\Omega(t)$ denote the brightness perceived by a person to the stage lighting stimulus at the t moment and used as an estimated radiant flux parameter, and the corresponding stereo angle parameter when the lighting sensor acquired the stage lighting status data, respectively. Combined with the calculation results of Eq. (1), the brightness, color temperature and other parameters of the lighting equipment are automatically adjusted according to the magnitude and direction of the difference, so that the lighting effect matches the stage light.

In the scene switching function, different lighting scenes are quickly switched mainly according to the performance requirements. In the specific realization stage, first, according to different performance needs, preset different lighting scenes, including lighting brightness, color temperature, and effect parameters, and then through the sensor real-time monitoring of the performance status on the stage, the plot change parameters, and comparing the collected data with the preset performance needs, judging whether the scene needs to be switched. The specific judgment method is shown in equation (16):

$$f(E) = E(t) - \frac{x(t)\cos\theta}{r^2} \tag{16}$$

where: $f(E)$ represents the fitting function between the collected performance state, plot change parameters and the current execution of the preset lighting scene. $E(t)$ represents the preset t-moment lighting scene information. θ and r denote the angle between the normal of the surface element and the beam propagation direction and the distance between the point light source and the surface element, respectively. When the judgment result outputted by Eq. (2) is that the scene needs to be switched, the parameters of the lighting equipment are quickly adjusted to realize the switching of the scene, and the corresponding adjustment target is the parameter under the preset scene that is consistent with the current performance progress.

Actor tracking function mainly through infrared or ultrasonic technology to monitor the position of the actor in real time, and automatically adjust the lighting to follow the movement of the actor. In the specific realization process, the actor position sensor is first used to monitor the position coordinates of the actor in real time, and then by comparing the position relationship between the collected position coordinates and the preset lighting area, it is judged whether it is necessary to adjust the position of the lights. The specific judgment method is shown in equation (17):

$$l_j^i = \frac{E_{\max} h}{\left[\sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + h^2}\right]^3} \tag{17}$$

where: l_j^i denotes the positional deviation between the acquired positional coordinates and the preset lighting area. x_i and y_i denote the actor position coordinates monitored by the actor position sensor in real time, respectively. x_j and y_j represent the preset position coordinates of the lighting area, respectively. $E_{\max}$ represents the maximum light source illuminance parameter. h represents the corresponding light source height information of the theater stage space. According to the analysis results, the angle and brightness and other parameters of the lighting equipment are automatically adjusted so that the corresponding lights follow the movement of the actors. In the stage of designing the function of the above theater stage distributed lighting control system, this paper invokes a feedback mechanism to continuously optimize the control parameters, and in this way improves the precision and

stability of the function implementation to provide a reliable guarantee for ensuring the best stage performance effect and good audience viewing experience.

III. Design and realization of automatic tracking system for stage lighting

In the previous sections, the related technologies of stage human body tracking and lighting control have been thoroughly studied and analyzed. In this section, the development and realization of the stage lighting automatic tracking servo system is completed based on the kernel function adaptive Mean-Shift target tracking algorithm, the speed interpolation algorithm based on the iterative inequality model, and the Huazhong 8 numerical control system platform.

III. A. Introduction to the system framework

The stage lighting automatic tracking servo system mainly contains four modules: control and management module, actor identification and positioning module, data processing module and speed interpolation module. The main processing flow of the stage lighting auto-tracking servo system is as follows: firstly, after the lower computer is ready, the data processing module sends the status information of the servo system and servo motors to the upper computer, which receives the information and analyzes it, and after confirming that the lower computer is ready, it initializes the system by means of the control and management module, and at the same time, turns on the system camera to display the stage image. Secondly, select the working mode of the system and configure the corresponding parameters, send the data under the corresponding mode to the lower computer through the data processing module, the lower computer receives the data through the conversion and processing, and then calculates the speed of the servo motor in the cycle through the speed interpolation algorithm. Finally, through the interface function, the speed value is sent to the servo system, which drives the servo motor to drive the lighting equipment to rotate, and at the same time reads the actual position of the motor and feeds back to the lower computer system.

III. B. Control and Management Module

III. B. 1) System initialization

The system initialization function is used to initialize all the parameters of the whole system. After checking the readiness of the lower computer, the height of the lighting equipment, the coordinate position on the stage plane and other data will be sent to the lower computer through the data processing module, which is used to initialize the lower computer system, and at the same time, the camera will be turned on by the camera driver and the video image will be displayed in the video window in real time.

III. B. 2) Model selection

Stage lighting automatic tracking servo system has three working modes: manual mode, semi-automatic mode and automatic mode. The main role of the mode selection is to make the staff can switch between the three modes, to achieve a variety of lighting equipment control, which can reflect more stage art effects. System switching between the three modes of operation should pay attention to check whether the servo motor is in a stopped state, because the system in the realization of the three modes of operation of the servo system selected by the control mode is not exactly the same, so in the change of the system mode of operation must be checked before the current control mode of the servo system and the system's working mode to match. The parameters required for each mode are then selected and sent to the lower computer in accordance with the data protocol format for each mode.

III. B. 3) Stopping operations

The main function of the stop function is to stop the movement of the lighting equipment in the current mode, generally the stop command will be frequently used in the process of system mode switching. The main process of the stop command is that after the lower computer analyzes the stop command, the speed interpolation module makes the rotation speed of the two axis motors gradually decrease to zero. There is no requirement for the final stopping position of the lighting equipment when the stop command is executed.

III. B. 4) Reset operation

The main function of the reset function is to make the lighting equipment move to the origin of the coordinates of the lower unit, i.e. the coordinate position sent to the lower unit when the system is initialized. The difference between the reset operation and the stop operation lies in the coordinate position of the last lighting device. The stop operation has no requirement for the last coordinate position, while the reset operation requires that the coordinate position of the last lighting device must be the coordinate origin in the lower computer subsystem.

III. C. Actor Identification and Localization Module

The main function of the actor recognition and localization module is to identify the tracking target and determine the target's current positional coordinates in all subsequent images captured by the camera after the actor to be tracked is selected by the staff in the start image, so that among the three working modes of the system, the module works only in the automatic mode.

The main process of the actor recognition and localization module is to first read the first frame image from the camera, and after manually selecting the tracking target, the kernel function bandwidth adaptive Mean-Shift target tracking algorithm proposed in this paper is applied to calculate the position of the target in the subsequent images and mark it with a rectangular box in the video window, and at the same time, the positional coordinates information is sent to the lower computer through the data processing module for computation [22]. The above process is repeated until a stop signal is detected.

III. D. Data-processing modules

III. D. 1) Data interaction

The data interaction function is mainly realized in the form of CAN bus, defining the basic format of the data storage area in all information frames on the CAN bus as shown in Table 1. Among them, the identification bit indicates the classification of the data in this frame, 0 indicates the system control command, and 1 indicates that the information in this frame is data. The command identification bit indicates the serial number of various commands in the command represented by the identification bit. The data area represents the system data to be transmitted via the CAN bus, and in order to minimize the error during data transmission, the data in the data area is basically transmitted in the form of 16-bit complementary code.

Table 1: Basic format

0	1	2	3	4	5	6	7
Identification bit	Command identifier	Data area					

III. D. 2) Data conversion

The data conversion function consists of two main parts, namely, the coordinate conversion between the position coordinate system of the stage and the angle coordinate system of the servo motor in the automatic and semi-automatic modes, and the conversion between the joystick analog and the speed of the servo motor in the manual mode.

(1) Coordinate system conversion

The upper computer will be the actor recognition and positioning module calculated by the actor's coordinate position through the data processing module through the CAN bus to the lower computer, at this time the coordinate position is in the video image captured by the camera to prevail in the stage coordinate system calculated, however, in the lower computer subsystem, the origin of the coordinates for the lighting equipment in the stage plane of the projection point, that is, initialization of the system is set up when the initial coordinate point of the lighting equipment.

Therefore, the first step in the transformation of the coordinate position is to carry out a coordinate system translation, assuming that the position coordinates of the actor on the stage calculated by the actor identification and positioning module in the host computer is (x,y), and the coordinates of the projection point of the lighting equipment on the stage is (x_0, y_0), then the real position coordinates of the actor's position coordinates (x,y) in the lower computer subsystem coordinate system corresponds to the position of the coordinates of the position of the coordinates of the (x', y'), and to meet the following relationship:

$$x' = x - x_0, \quad y' = y - y_0 \tag{18}$$

Since the servo system control of the motor is based on the angle of rotation of the motor, the position coordinate (x', y') needs to be converted to a coordinate system, i.e., the plane coordinates are converted to the angle coordinates of the motor.

The light irradiation point P in the lower computer subsystem in the plane coordinates (x', y') corresponding to the angle coordinates of (α, β), and to meet:

$$\alpha = \arctan(x' / h), \beta = \arctan(y' / h) \tag{19}$$

where h represents the vertical distance of the lighting equipment from the stage plane.

The first equation in the formula (in the case of the x-axis) is obtained by simultaneously deriving both sides of the equation with respect to time t:

$$\frac{d\alpha}{dt}=\frac{d\alpha}{dx^{'}}\cdot\frac{dx^{'}}{dt}=\frac{d}{dt}\left[\arctan\left(\frac{x^{'}}{h}\right)\right]\cdot\frac{dx^{'}}{dt}=\frac{h}{h^2+x^{'2}}\frac{dx^{'}}{dt} \tag{20}$$

In the formula remember $\frac{d\alpha}{dt}$ as v_{α} and $\frac{dx^{'}}{dt}$ as $v_{x^{'}}$, respectively, represent α the angular size and x the rate of change of axis coordinates with time, then the formula can be expressed as:

$$v_{\alpha}=\frac{h}{h^2+x^{'2}}v_{x^{'}} \tag{21}$$

Therefore, you can adjust the speed size of the servo motor according to the formula, so that the light spot of the light irradiation on the stage along the x-axis with a speed of $v_x^{'}$ uniform motion, the same reason can be obtained from the v_{β} and $v_y^{'}$ of the change of the relationship between the formula:

$$v_{\beta}=\frac{h}{h^2+y^{'2}}v_{y^{'}} \tag{22}$$

On the contrary, the lower computer, after getting the angular coordinates of the motor position fed back by the servo system, can derive the position coordinates of the current light irradiation point on the stage as (x,y) according to the formula and satisfy:

$$x=h\cdot\tan\alpha+x_0,\ y=h\cdot\tan\beta+y_0 \tag{23}$$

(2) Semi-automatic mode speed conversion

In manual mode, the servo motor rotation speed is determined by the analog of the joystick, assuming that the maximum value of the absolute value of the joystick analog is $rocker\_\max$, the maximum rate at which the servo motor can run is $v\_\max$, then when the lower computer receives and analyzes the analog of the joystick is $rocker$ (plus or minus on behalf of the positive and negative direction of the servo motor movement), the servo motor should rotate at a rate of:

$$v=rocker\frac{v\_\max}{rocker\_\max} \tag{24}$$

III. E. Velocity interpolation module

The main function of the speed interpolation module is to use the speed interpolation algorithm to calculate the speed of the servo motor in the current cycle according to the data after passing through the data processing module, and then control the speed of the motor through the function interface provided by the servo system, so as to make the lighting equipment rotate to the specified position at the fastest speed or move according to the specified trajectory.

IV. Role-tracking performance test experiment

It includes anti-masking experiments and tracking experiments. The anti-obscuration experiments are carried out on simulated and real labeled datasets to verify the anti-obscuration performance of the binary cyclic code. The tracking experiments are carried out on the actor tracking dataset to verify the effectiveness of the proposed data association algorithm based on intersection ratio and the stable actor tracking algorithm based on infrared ink markers.

IV. A. Data sets

The datasets used in the experiments of this section include the simulation dataset and the real dataset. In both the simulation dataset and the real dataset, the images have complete coding information, which can satisfy the recognizable requirements of ring code classification.

IV. A. 1) Simulation labeling dataset

The simulation dataset is used to validate the anti-obscuration performance of binary cyclic codes with different percentages of cropping to simulate the occurrence of occlusion, the sample images of the training set are generated with 0%, 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65% occlusion rates, and the sample images of the test set are generated with 2%, 12%, 22%, 32% , 42%, 52%, and 62% occlusion rates were generated. Image processing operations such as affine transformation, blurring, noise, sharpening relief, randomly discarding pixels, and color change are used to do data augmentation on the training set and test set sample images. 20522 and 8956 training set and test set images containing ring codes with different occlusion rates are obtained respectively.

IV. A. 2) Real-labeled datasets

The real dataset was acquired by printing the 16 binary ring codes shown in Fig. 5 on 16 identical pieces of clothing using NIR ink. The room was lighted with NIR fill-in lamps, and an infrared camera was used to capture videos of the experimental subjects performing with predetermined movements. The video was captured with a frame rate of 50 fps and a resolution of 1,920 × 1,080 pixels. Again, to ensure that the training data and the test data have different distributions, the types of movements for the training data include standing horizontally in a group of four people doing upward jumps, forward clapping of the hands and holding the chest, squatting, and multi-directional chest expansion, etc., and the types of movements for the test data include walking around in a circle in a group of four people with randomized markers for masking. Marker frame labeling and actor frame labeling were done on the captured video frames, and finally four training video data with a duration of 2 min and four test video data with a duration of 1 min were obtained.

The labeled dataset used for the anti-obscuration experiments requires further processing of the video data, i.e., cropping the binary cyclic coding frames in each frame of the video and manually sifting out the sample data that are fuzzy, don't have complete coding structure, and duplicated, to finally get the training dataset containing 14,500 labeled images and the test dataset with 65,000 labeled images.

IV. A. 3) Actor tracking dataset

To verify the effectiveness of the algorithm based on this paper, a test dataset was collected for actor tracking. A near-infrared fill light was shone indoors to capture six actors walking around the stage randomly accompanied by entering and exiting the scene. The captured videos have a frame rate of 25 fps, a resolution of 3560×1440 pixels, a duration of about 2 min, and are labeled according to the mot format standard.

IV. B. Anti-masking experiments

Verify the anti-obscuration performance of ring markers with simulated data and real marker data. The anti-obscuration performance of the markers under different occlusion rates is measured by training a generalized classifier model and using the classification accuracy of the trained classifier on the test set. Obviously, the markers with good occlusion resistance can still maintain high classification accuracy at higher occlusion rates. The lightweight MobileNet-V2 classification network is used without any special qualification on the classifier model.

IV. B. 1) Simulation Demonstration

The results of the anti-obscuration experiments on the simulated dataset are shown in Fig. 5. As can be seen from the figure, when the occlusion rate is 60%, the classification accuracy is still above 91%, which indicates that the method in this paper has good anti-occlusion ability.

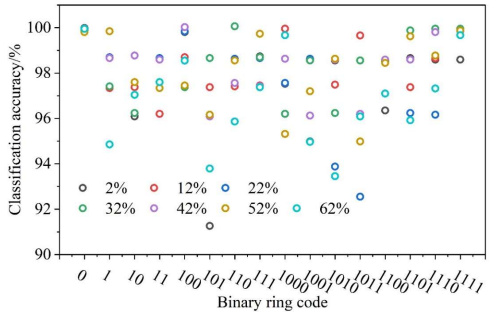

Figure 5: The experimental results of the anti-occlusion in the simulation data set

IV. B. 2) Real-world scenario validation

The results of the anti-obscuration experiments on the real dataset are shown in Fig. 6. Different from the simulation experiments, the figure only gives the classification accuracy of different binary cyclic codes without giving the results under different occlusion rates because people are randomly occluded in real scenarios cannot be quantized. As can be seen from the figure, the classification accuracies for binary cyclic codes on real datasets are all above 90%, which indicates that the method in this paper has good occlusion-resistant ability in practical applications.

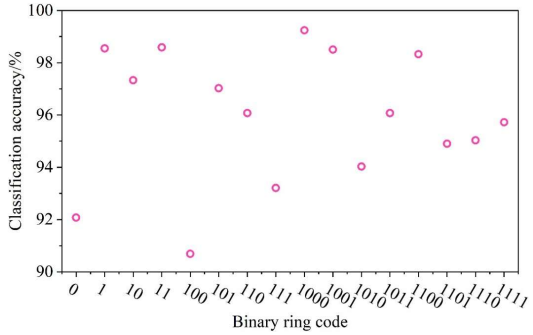

Figure 6: Anti-occlusion results in real data sets

IV. B. 3) Follow-up experiments

FairMOT tracking model is used to verify the effectiveness of this paper's algorithm with real video data.

(1) Comparison of association methods

The generation value and matching results under different association indexes are shown in Fig. 7. The comparison metrics used in this paper's algorithm include L2-Norm, IoU, GIoU. As shown in Fig. If the metrics such as L2-Norm, IoU, GIoU are used to measure the correlation between actor frames and marker frames, and then the actor frames belonging to the marker frames are selected according to this correlation, it will result in the wrong matching of marker frames. And using the intersection ratio proposed in this paper for the correlation metric, the accuracy of the association between actor frames and marker frames can be improved and correct matching can be performed.

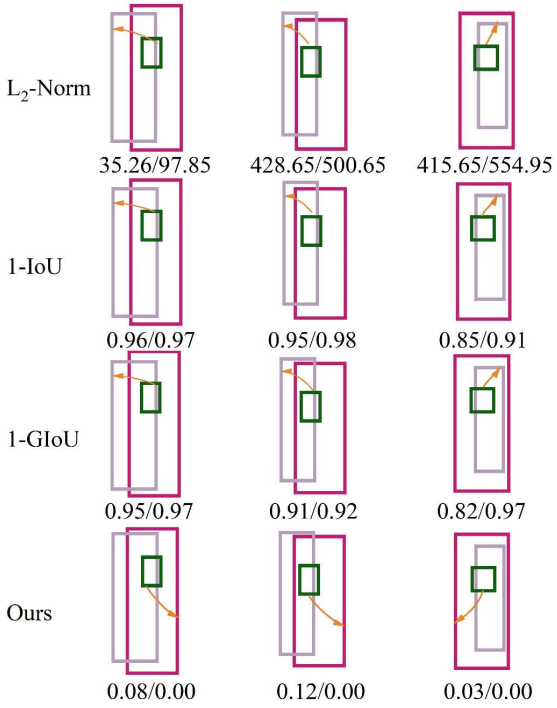

Figure 7: Cost values and matching results under different correlation indicators

(2) Quantitative Estimation of Tracking Performance

To further validate the effectiveness of this paper's algorithm, the performance of using this paper's algorithm was compared on the actor tracking dataset. The tracking algorithms include ByteTrack, BoT_SORT, and FairMOT, and the performance comparison of each algorithm is shown in Table 2 (the results after using this paper's algorithm are in parentheses). The experiments use metrics such as IDP, IDR, and IDF1, where IDP is used to measure accuracy, IDR is used to measure recall, and IDF1 combines accuracy and recall, i.e., the reconciled mean of the two.

The experimental results show that after using the algorithms in this paper, the accuracy and recall of each tracking algorithm for stage actor ID tracking have been greatly improved, which verifies that the algorithms in this paper can greatly improve the stability of tracking stage actors.

Table 2: Performance comparison of each algorithm

Tracking algorithm	IDP	IDR	IDF1
ByteTrack	50.6(62.9)	49.5(59.9)	50.9(61.8)
BoT_SORT	62.3(68.2)	58.5(64.3)	60.6(65.8)
FairMOT	25.9(36.5)	20.9(29.5)	23.2(33.6)

V. The application effect of this paper's method in stage character design

V. A. Study design

In order to explore the effectiveness of this paper's methodology in the application of stage character design, a questionnaire was administered to 500 audience members in this section, of which 286 were from field research and 214 were from e-mail. The experiment establishes the following hypotheses:

H1: Audience members prefer the theme of the stage performance more
H1a: It is entertaining
H1b: have intellectual nature
H1c: able to fulfill the audience's emotional needs
H1d: is attractive
H1e: higher likelihood of audience acceptance
H2: The audience prefers the graphic look of the stage more
H2a: Visual effects can reflect the standard of stage character production
H2b: Reflects the imagination and creativity of the author.
H2c: More diverse and exquisite costumes
H2d: Characters with distinctive and individualized language, behavior, or dress, thus making a strong impression.
H3: Audience preference for vocalization of stage performances
H3a: Opening, closing and interlude songs are more appreciated and collectible.
H3b: Scene soundtracks blend more perfectly with the atmosphere or characters' state of mind
H3c: Fusion of character's emotion and voice actor's emotion

V. B. Overall description of the questionnaire

Before conducting the formal study, the questionnaires were divided into two groups, the field research group and the e-mail group, according to the different sources of data sources, and an independent samples test was conducted to determine whether all the questionnaires could be combined into one for subsequent analysis by examining the differences between the two groups of questionnaires, and the results of t-tests of the data sources are shown in Table 3.

As can be seen from the table, the F-value of the indicator “Entertainment” has a Sig. value of 0.072, which is greater than the expected significance level of 0.05, so the hypothesis that the variance of the indicator is equal is accepted, and the t-value of the corresponding value has a Sig. value of 0.149, which is greater than 0.05, so that the data obtained from the field research and e-mail channels are not significantly different in terms of the indicator. There is no significant variability in the data obtained from the two channels of research and e-mail.

Using the above methodology it was possible to determine the variability of the data sources on the other indicators. The results showed that for all indicators, the two sources of data did not contribute to their variability, so that the two sets of questionnaires could be combined into one for subsequent analysis.

Table 3: T test results of data sources

		Levene test of the variance equation		T test of the mean equation						
		F	Sig.	t	df	Sig.(Double side)	Mean interpolation	Standard error value	95% confidence interval of the difference	
									Lower limit	Upper limit
Entertainment	Let's say the variance is equal	3.152	0.072	1.469	422	0.149	0.206	0.144	-0.079	0.487
	Let's say that the variance is not equal			1.479	345.366	0.144	0.202	0.143	-0.063	0.483
Knowledgeability	Let's say the variance is equal	16.296	0.000	1.312	422	0.196	0.155	0.121	-0.074	0.385
	Let's say that the variance is not equal			1.244	289.652	0.211	0.159	0.123	-0.087	0.383
Satisfaction of emotional needs	Let's say the variance is equal	4.665	0.036	-1.832	422	0.072	-0.266	0.145	-0.543	0.023
	Let's say that the variance is not equal			-1.777	303.321	0.075	-0.265	0.149	-0.555	0.023
Attraction	Let's say the variance is equal	6.922	0.008	1.598	422	0.112	0.145	0.089	-0.036	0.322
	Let's say that the variance is not equal			1.692	352.155	0.106	0.141	0.085	-0.031	0.312
The possibility of being accepted by the audience	Let's say the variance is equal	0.041	0.839	1.559	422	0.123	0.149	0.096	-0.035	0.322
	Let's say that the variance is not equal			1.622	378.62	0.106	0.146	0.088	-0.032	0.322
Visual effect	Let's say the variance is equal	30.321	0.000	-3.895	422	0.395	0.188	0.216	-0.822	-0.362
	Let's say that the variance is not equal			-3.556	275.65	0.375	0.188	0.215	-0.836	-0.241
Dress feature	Let's say the variance is equal	17.855	0.000	1.622	422	0.105	0.241	0.163	-0.042	0.522
	Let's say that the variance is not equal			1.567	286.954	0.119	0.236	0.155	-0.066	0.536
Imagination and creativity	Let's say the variance is equal	10.120	0.002	0.542	422	0.588	0.085	0.155	-0.225	0.532
	Let's say that the variance is not equal			0.531	304.356	0.596	0.085	0.162	-0.221	0.396
Personage of roles	Let's say the variance is equal	32.992	0.000	1.755	422	0.088	0.212	0.126	-0.023	0.466

	Let's say that the variance is not equal			1.623	256.32	0.102	0.095	0.133	-0.451	0.111
The pleasure of music	Let's say the variance is equal	2.795	0.095	-1.212	422	0.226	-0.179	0.149	-0.463	0.112
	Let's say that the variance is not equal			-1.225	338.95	0.222	-0.179	0.145	--0.46	0.109
Scene dubbing	Let's say the variance is equal	2.563	0.112	-3.175	422	0.968	0.011	0.223	-0.815	-0.193
	Let's say that the variance is not equal			-3.206	336.65	0.968	0.011	0.212	-0.812	-0.195
Dubbing effect	Let's say the variance is equal	0.088	0.769	-1.523	422	0.135	-0.236	0.155	-0.569	0.071
	Let's say that the variance is not equal			-1.523	331.261	0.135	-0.236	0.155	-0.569	0.075

V. C. Perceptual difference analysis

Descriptive statistics for each perceived difference are shown in Table 4. It includes the sample size, the mean value, and the standard deviation. From the mean value, the audience's score for the imagination and creativity of the characters in the stage is the highest at 4.11, which leads to the conclusion that using the method of this paper for the stage character design can bring the audience a good effect of characterization and presentation.

Table 4: Descriptive statistics of perceptual differences

	N	mean	Standard deviation	Standard error of mean
Entertainment	500	3.65	0.992	0.102
Knowledgeability	500	3.19	0.957	0.1
Satisfaction of emotional needs	500	3.21	1.008	0.088
Attraction	500	3.85	0.952	0.088
Acceptance possibility	500	3.12	1.21	0.058
Visual effect	500	3.8	1.181	0.069
Dress feature	500	3.2	1.143	0.069
Imagination and creativity	500	4.11	1.235	0.09
Personage of roles	500	3.35	1.04	0.12
The pleasure of music	500	3.57	1.07	0.082
Scene dubbing	500	4.07	1.077	0.068
Dubbing effect	500	3.67	1.052	0.075

The results of the analysis of the differences in the perceptions are shown in Table 5. From the table, the Sig. value of the T-test of the knowledge index is 0.005, which is smaller than the expected significance level of 0.05, and it can be assumed that there is a significant difference between the audience's perceptions of the stage character design under this index, which indicates that the audience recognizes more that the character after designing by using the method of this paper is knowledgeable on this index, and therefore the hypothesis H1b is valid.

In summary, the validation of each hypothesis in this study can be obtained, and the validation of each hypothesis in this study is shown in Table 6. Overall, the audience has a good evaluation of the stage character design based on the method of this paper.

Table 5: Analysis of various perceptual differences

		Levene test of the variance equation		T test of the mean equation		
		F	Sig.	t	df	Sig.(Double side)
entertainment	Let's say the variance is equal	1.543	0.214	1.155	422	0.259
	Let's say that the variance is not equal			1.155	226.359	0.251
knowledgeability	Let's say the variance is equal	1.543	0.214	3.202	422	0.005
	Let's say that the variance is not equal			3.202	188.624	0.005
Satisfaction of emotional needs	Let's say the variance is equal	4.625	0.036	3.105	422	0.005
	Let's say that the variance is not equal			3.006	185.652	0.005
attraction	Let's say the variance is equal	1.546	0.221	3.169	422	0.004
	Let's say that the variance is not equal			3.206	293.658	0.004
The possibility of being accepted by the audience	Let's say the variance is equal	0.942	0.336	0.836	422	0.403
	Let's say that the variance is not equal			0.825	255.62	0.402
Visual effect	Let's say the variance is equal	3.605	0.06	2.133	422	0.032
	Let's say that the variance is not equal			2.116	185.65	0.031
Dress feature	Let's say the variance is equal	9.956	0.001	1.388	422	0.033
	Let's say that the variance is not equal			1.356	216.34	0.271
Imagination and creativity	Let's say the variance is equal	14.658	0	3.775	422	0.275
	Let's say that the variance is not equal			3.731	195.64	0.000
Personage of roles	Let's say the variance is equal	15.21	0	3.775	422	0.000
	Let's say that the variance is not equal			3.761	193.65	0.001
The pleasure of music	Let's say the variance is equal	2.653	0.105	3.551	422	0.001
	Let's say that the variance is not equal			3.52	229.65	0.001
Scene dubbing	Let's say the variance is equal	0.074	0.782	0.923	422	0.001
	Let's say that the variance is not equal			0.923	229.65	0.362
Dubbing effect	Let's say the variance is equal	4.305	0.05	2.55	422	0.013
	Let's say that the variance is not equal			2.513	255.61	0.013

Table 6: This study shows the validation of each hypothesis

Research hypothesis	Hypothesis validation
H1	
H1a	Out of reach
H1b	
H1c	
H1d	
H1e	Out of reach
H2	Set up
H2a	Set up
H2b	Set up
H2c	Out of reach
H2d	Set up
H3	
H3a	Set up
H3b	Out of reach
H3c	Set up

VI. Conclusion

This article explores the related applications of virtual reality technology and multiple perceptual modeling fusion in human body tracking and lighting control in stage character design. Through experimental tests, the article draws the following conclusions:

(1) In the results of anti-obscuration experiments on real datasets, the classification accuracies of binary cyclic codes on real datasets are all above 90%, from which it can be concluded that the method of this paper has good anti-obscuration ability in practical applications.

(2) In the aesthetic difference of stage art, the audience's mean score for the imagination and creativity of the characters in the stage is the highest at 4.11, that is, the use of this paper's method for stage character design can reflect rich imagination and creativity, bringing good stage presentation effects to the audience.

References

[1] Jia, B. (2023). The Cross-Media and Cross-Cultural Communication of Art—Using Video Games as the Carrier. Studies in Art and Architecture, 2(2), 52-55.

[2] Long, D. (2019). " Andersstreben" and Cross-media Narrative. Theoretical Studies in Literature and Art, 39(3), 184.

[3] Li, S., & Li, J. (2022). Construction of Interactive Virtual Reality Simulation Digital Media System Based on Cross‐Media Resources. Computational Intelligence and Neuroscience, 2022(1), 6419128.

[4] 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.

[5] Baía Reis, A., & Ashmore, M. (2022). From video streaming to virtual reality worlds: an academic, reflective, and creative study on live theatre and performance in the metaverse. International Journal of Performance Arts and Digital Media, 18(1), 7-28.

[6] Iseli, C. (2023). Double Trouble. Digital Avatars on Stage. ACTOR & AVATAR, 114.

[7] Ma, M., & Yuping, K. (2025). Revolutionizing the stage: exploring the multidimensional landscape of digital theater. Digital Scholarship in the Humanities, fqaf015.

[8] Lamberti, F., Gatteschi, V., Sanna, A., & Cannavò, A. (2019). A multimodal interface for virtual character animation based on live performance and natural language processing. International Journal of Human–Computer Interaction, 35(18), 1655-1671.

[9] Kadry, A., & Hussien, E. (2023). Applications of Virtual Reality Technologies in the Field of Design and Arts. International Journal of Multidisciplinary Studies in Art and Technology, 6(2), 44-70.

[10] Liu, F., Gao, Y., Yu, Y., Zhou, S., & Wu, Y. (2021). Computer aided design in the diversified forms of artistic design. Computer-Aided Design and Applications, 19(3).

[11] Wang, Y., & Hu, X. B. (2022). Three‐dimensional virtual VR technology in environmental art design. International Journal of Communication Systems, 35(5), e4736.

[12] Cmentowski, S., Karaosmanoglu, S., Kievelitz, F., Steinicke, F., & Krüger, J. (2023). A Matter of Perspective: Designing Immersive Character Transitions for Virtual Reality Games. Proceedings of the ACM on Human-Computer Interaction, 7(CHI PLAY), 73-103.

[13] Wilson, H. R. (2020). New ways of seeing, feeling, being: intimate encounters in virtual reality performance. International Journal of Performance Arts and Digital Media, 16(2), 114-133.

[14] Melki, H. (2021). Stage-directing the Virtual Reality Experience: Developing a Theatrical Framework for Immersive Literacy. International Journal of Film and Media Arts, 6(2), 129-147.

[15] Demetriou, P. A. (2018). 'Imagineering'mixed reality (MR) immersive experiences in the postdigital revolution: innovation, collectivity, participation and ethics in staging experiments as performances. International Journal of Performance Arts and Digital Media, 14(2), 169-186.

[16] 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.

[17] Qian, J. (2020, April). Application of VR in art design. In Journal of Physics: Conference Series (Vol. 1533, No. 2, p. 022004). IOP Publishing.

[18] An, R. (2024). Art curation in virtual spaces: The influence of digital technology in redefining the aesthetics and interpretation of art. Humanities, Arts and Social Sciences Studies, 503-518.

[19] BangjvHuang,BolinMa & LiLu. (2025). Research for the Detection of Aircraft Target in Real-Time Based on YOLOv5.Advances in Civil Engineering,2025(1),9521952-9521952.

[20] Bohang Chen,Jun Ma,Lingfei Zhang,Zhuang Xiong,Jinyu Fan & Haiming Lan. (2023). An improved weighted KNN fingerprint positioning algorithm. Wireless Networks,30(6),6011-6022.

[21] Chiradeja Pathomthat & Yoomak Suntiti. (2023). Development of public lighting system with smart lighting control systems and internet of thing (IoT) technologies for smart city. Energy Reports,10,3355-3372.

[22] Jianqing Gao,Haiyang Zou,Fuquan Zhang & Tsu Yang Wu. (2022). An intelligent stage light-based actor identification and positioning system. International Journal of Information and Computer Security,18(1-2),204-218.