International Journal for Housing Science and Its Applications
Publish August 6, 2025. Volume 46, Issue 3 Pages 6892-6905

https://doi.org/10.70517/ijhsa463591

A Study on the Enhancement of Digital Modeling on Character Emotional Presentation Ability in Film and Television Art

Zhe Xu[1,*]
[1] China-Korean Institute of New Media, Zhongnan University of Economics and Law, Wuhan, Hubei, 430000, China
Corresponding authors: (e-mail: waibiwaibixz@163.com).

Abstract In the era of experience economy, film and television art, as an important way to satisfy users' entertainment needs, has a direct impact on the communication effect of its character emotion expression ability. This paper constructs a digital character emotion expression framework by taking the PAD three-dimensional emotion space as the theoretical basis and combining the first-order motion model and generative adversarial network. The study adopts the MATLAB development environment to verify the effectiveness of the emotion calculation method, evaluates the quality of digital character synthesis through experiments based on the TensorFlow framework in the Linux environment, and conducts an experience test of emotional characteristics among 80 college students. The results show that: the accuracy of the emotion calculation method proposed in this paper reaches 84.78%, which is 14.86% and 5.16% higher than that of OCC and fuzzy inference methods, respectively; the peak signal-to-noise ratio of the constructed FOMM-GAN digital character synthesis model reaches 35.63, with an average gradient of 5.96, and the synthesis time is 94.27ms, which is significantly better than the comparison method; the subjects' perception of the digital character's emotion The mean values of the subjects' experience of the degree of emotional perception, adaptability, anthropomorphism, engagement and initiative were 1.654, 1.688, 1.538, 1.513 and 1.484, respectively, indicating that the digital characters have a high ability to express emotions. The study confirms that the digital character modeling method based on PAD space and FOMM-GAN can effectively enhance the emotional expressiveness of film and television characters, and provide new ideas for film and television art creation.

Index Terms Digital modeling, Film and television art, Emotion presentation, PAD emotion space, First-order motion model, Generative adversarial network

I. Introduction

Traditional film and television production methods have been unable to meet the growing visual aesthetic needs of modern audiences, and computer digital modeling technology has injected new vitality into film and television works with its unique artistic expression and shocking visual effects [1]-[3]. Digital modeling technology integrates many fields such as computer graphics, animation technology, visual effects, etc. Through a series of processes such as 3D modeling, animation production, and special effects synthesis, it seamlessly combines the virtual digital world with real images to create a visual spectacle beyond reality [4]-[6]. Digital modeling, as an important part of producing visual effects for film and television, is crucial to improving the quality of film and television works [7].

For character modeling technology, film and television creators, with the help of 3D modeling software, ensure that every part of the character model can achieve natural and expressive deformation driven by bones through fine weight drawing and adjustment [8], [9]. At the same time, applying motion capture data to the corresponding 3D characters, we can strive to achieve the best effect in time rhythm and kinetic rhythm through fine adjustments, so that the character's movements are more smooth and natural, rich in character emotion [10]-[12]. In addition, for the scene modeling technology, skillfully designed scenes can highlight the character image characteristics and personality style, making it more vitality and distinctive levels [13], [14]. For example, in animated films, the character's identity mark and social status can be effectively highlighted through the carefully planned characteristic architecture and interior decoration [15]. And in the science fiction theme of the film, with the help of building the future city, cosmic spacecraft and other scenes, can strengthen the characters of science fiction color and the spirit of adventure [16].

Today's world is in the era of experience economy, people's demand for entertainment and services has become the dominant factor in production and consumption. As an important way to meet users' entertainment needs, the development goal of film and television art is to provide users with unforgettable experiences. In the creation of film and television art, the character is the carrier of the story, and its ability to express its emotions directly affects the infectious force and dissemination effect of the work. Traditional movie and television character emotional expression mainly relies on actor performance or animator manual design, facing limited expressive power, production costs and high problems. The development of digital technology provides new possibilities for the emotional expression of film and television characters, especially the digital modeling technology can give the characters rich emotional expression ability. However, the existing digital character models are still insufficient in terms of naturalness, coherence and realism of emotional expression, and it is difficult to meet the audience's expectations for character emotional expression. At present, academic research on the application of digital modeling in the emotional expression of film and television characters is still in its infancy, lacking a systematic theoretical framework and experimental verification. How to use digital modeling technology to enhance the emotional expression of film and television characters has become a key problem to be solved in the development of film and television art. Affective computing, as an important branch in the field of artificial intelligence, provides a theoretical basis for solving this problem, while advanced technologies such as generative adversarial networks and first-order motion models provide technical support for realizing high-quality emotional expression of digital characters. Exploring the application mechanism and effect of digital modeling technology in the emotional expression of film and television characters will not only help to enhance the artistic expression of film and television works, but also provide an important reference for the development and application of digital human technology.

Based on the PAD 3D emotion space theory, this study constructs a digital character emotion expression framework for film and television art by combining the first-order motion model and generative adversarial network technology. Firstly, the PAD 3D emotion space is used to quantitatively describe the emotion of the character, and indicators such as emotional intensity, fluctuation amount and fluctuation rate are introduced to realize the accurate calculation of the emotional state of the digital character; secondly, the first-order motion model is combined with the generative adversarial network, and the FOMM-GAN digital character synthesis model is proposed to realize high-quality expression of emotion of the digital character; lastly, the 3D virtual environment is designed to construct the overall realization framework of digital characters, and verify the effectiveness of the proposed method through simulation experiments and user experience tests. By systematically studying the effect of digital modeling on the emotional expression of film and television characters, this study aims to provide new technical methods and theoretical guidance for the creation of film and television art, and to promote the innovative development of film and television art in the era of experience economy.

II. Digital Modeling Methods for Film and Television Characters

The 21st century is the era of experience economy, people's demand for entertainment and services will become the dominant factor in production and consumption. Among them, the development of film and television art is to better meet the user's entertainment needs and provide users with unforgettable experiences. Digital modeling in film and television art can give the character emotional ability, further enhancing the emotional expression of film and television art. Based on this, exploring the effect of digital modeling on the enhancement of the character's emotional display ability can effectively further enhance the emotional expression ability of the character in film and television art.

II. A. Methods of Presenting Emotions in Digital Characters

II. A. 1) PAD three-dimensional emotional space

To introduce human emotions and feelings into artificial intelligence, it is necessary to quantify subjective emotions based on certain psychological theories of emotions. Emotions are difficult to be precisely defined due to uncertainty, sensuality, and abstraction.The PAD 3-Dimensional Emotion Model suggests that emotions can be composed of three cognitively meaningful dimensions, namely, the pleasantness, activation, and dominance of emotions [17].

Pleasurability refers to the degree of pleasure/displeasure, positive/negative of an emotional state. For example, happiness and disgust are two typical emotions with high and low pleasantness, respectively; happiness is pleasurable and disgust is unpleasant. Activation is the degree of physiological activity of an emotion, for example, anger and sadness are two typical emotions with high and low activation, respectively. Dominance is the ability of an emotion to control and influence the external environment and others. Emotions produced autonomously by an individual and those produced by the influence of the external environment are different in terms of emotional dominance. For example, contempt is subjectively generated by the individual and has a high degree of dominance. Fear, on the other hand, is generated by the individual being influenced by the external environment/others and has a relatively low degree of dominance.

Figure 1 shows the PAD 3D emotion space framework. In the PAD emotional space, each coordinate corresponds to an emotional state. In other words each emotional state can be represented by a point in this space, which can also be viewed as a vector ($e=[P,A,D]$, e is the PAD emotion vector). Each dimension has a range of [-1,1] after normalization, and the point (0,0,0) indicates a calm state of emotion.

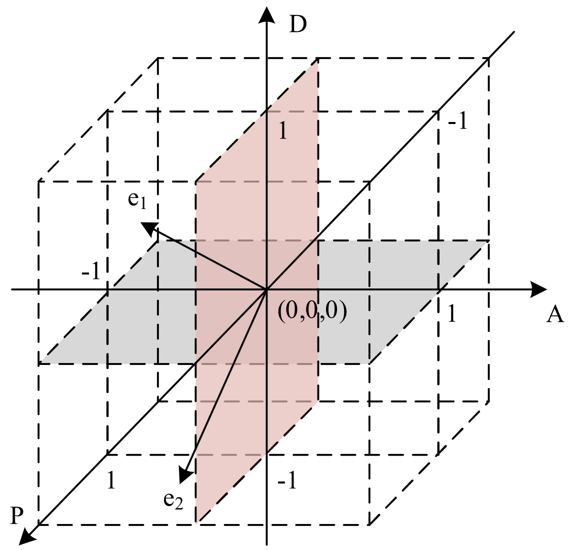


Figure 1: PAD 3D emotional space framework

II. A. 2) PAD Spatial Affective Metrics

Each PAD point in the PAD 3D emotion space has a unique emotion state corresponding to it, let point X be any point in the PAD 3D emotion space, the emotion state corresponding to X is recorded as emotion X, the coordinates of X are $\left(e_P^X, e_A^X, e_D^X\right)$, e_P^X denotes the size of emotion X on pleasure P, e_A^X denotes the size of emotion X on activation A, and e_D^X denotes the size of emotion X on dominance D. To characterize the strength of emotion X, this paper introduces the concept of emotional intensity.

(1) Emotion intensity. For any point X in the PAD emotion space, the vector $\overrightarrow{OX}$ constituted by it and the coordinate origin O is the emotion intensity M_X of emotion X. Its physical meaning is the degree of deviation of emotion X relative to the calm emotional state. Let the value of the emotional intensity M_X be r, the angle of direction of M_X be $\theta=\left(\theta_P,\theta_A,\theta_D\right)$, and the vectors $\theta_P,\theta_A,\theta_D$ are the angles of the vector $\overrightarrow{OX}$ with the P-axis, A-axis and D-axis, respectively, then we have:

$$r=\left|\overrightarrow{OX}\right|=\sqrt{\left(e_P^X\right)^2+\left(e_A^X\right)^2+\left(e_D^X\right)^2} \tag{1}$$

$$\begin{gathered}\theta=\frac{180^{\circ}}{\pi}\left(\arccos\frac{e_P^X}{r},\arccos\frac{e_A^X}{r},\arccos\frac{e_D^X}{r}\right)\\ \theta_P,\theta_A,\theta_D\in\left[-180^{\circ},180^{\circ}\right]\end{gathered} \tag{2}$$

Since human emotions are time-varying, in order to be able to quantitatively describe the emotional fluctuations, this paper utilizes the above concept of emotional intensity and gives the following definitions:

(2) Mood fluctuation quantity. In the PAD emotion space, the amount of change in emotional intensity caused by the transformation of any emotion A into emotion B is called the amount of fluctuation from emotion A to emotion B, denoted as F_{AB}.

Let $\left|F_{AB}\right|$ be the amount of fluctuation of emotion value of F_{AB}, which indicates the change in the magnitude of emotion intensity value. Taking into account the direction of change of emotional intensity value, we can construct $\left|F_{AB}\right|$ to calculate the formula as:

$$\left|F_{AB}\right|=\lambda\left|M_B-M_A\right| \tag{3}$$

where M_A and M_B are the emotional intensity of emotions A and B, respectively, which can be had according to the emotional intensity:

$$|M_B - M_A| = |\overrightarrow{OB} - \overrightarrow{OA}| = \sqrt{\left(e_P^B - e_P^A\right)^2 + \left(e_A^B - e_A^A\right)^2 + \left(e_D^B - e_D^A\right)^2} \tag{4}$$

λ is the sentiment value fluctuation direction factor, which takes the value of:

$$\begin{cases} \lambda = 1, & |\overrightarrow{OB}| \geq |\overrightarrow{OA}| \\ \lambda = -1, & |\overrightarrow{OB}| < |\overrightarrow{OA}| \end{cases} \tag{5}$$

That is, λ is positive when the emotion with low emotional intensity value turns to the emotion with high emotional intensity value, and vice versa is negative.

Let α be the amount of emotional angle fluctuation of F_{AB}, which indicates the size of emotional intensity angle change. Taking the pleasantness in the emotion PAD as a reference, the formula of α can be constructed as:

$$\begin{cases} \alpha = \gamma \cdot \dfrac{180^\circ}{\pi} \cdot \beta \\ \beta = \arccos \dfrac{\overrightarrow{OA} \cdot \overrightarrow{OB}}{|\overrightarrow{OA}||\overrightarrow{OB}|}, \beta \in \left[0, 180^\circ\right] \end{cases} \tag{6}$$

where $\overrightarrow{OA} \cdot \overrightarrow{OB}$ is the dot product of the two vectors, β is the angle between $\overrightarrow{OA}$ and $\overrightarrow{OB}$ and γ is the direction of the mood angle fluctuation factor, which takes the value of:

$$\begin{cases} \gamma = 1, & e_P^B \geq e_P^A \\ \gamma = -1, & e_P^B < e_P^A \end{cases} \tag{7}$$

That is, γ is positive when the mood of low pleasantness turns into a mood of high pleasantness, and vice versa is negative.

(3) Mood volatility. In the PAD emotion space, the amount of change in emotional intensity caused by the transformation from any emotion A to emotion B per unit of time is called the volatility of emotion A to emotion B, denoted as K_{AB}.

Let K_r be the volatility of the emotional value of K_{AB}, K_α be the volatility of the emotional angle of K_{AB}, and t be the time elapsed from the conversion of emotion A to emotion B, then we have:

$$K_r = \frac{|F_{AB}|}{t} \tag{8}$$

$$K_\alpha = \frac{\alpha}{t} \tag{9}$$

II. A. 3) 3D Emotional Space Realization

According to the classification and characteristics of personality, human personality is divided into introvert and extrovert. Introverted people are quiet, emotions are not easy to show, good at patience, emotional stability and difficult to transfer, the absolute value of their emotional threshold is larger. Extroverted people like to socialize, emotionally exposed to the outside, lack of endurance, impulsive, mind change drastically, the absolute value of the emotional threshold is smaller.

Emotional state $e_s^n = [x_s, y_s, z_s]$ in 3D emotional space, emotional threshold is T_i and emotional intensity $\gamma_s = \sqrt{x_s^2 + y_s^2 + z_s^2}$. Based on the external stimuli, let the probability matrix of the change in emotional state be:

$$p_s = \begin{bmatrix} p_{xx} & p_{xy} & p_{xz} \\ p_{yz} & p_{yy} & p_{yz} \\ p_{zx} & p_{zy} & p_{zz} \end{bmatrix} \tag{10}$$

Then the amount of emotional change is $\Delta e_s = e_s^n p_s$ and the new emotional state after the stimulus is:

$$\begin{aligned} e_s^{n+1} &= e_s^n + \Delta e_s = e_s^n + e_s^n p_s \\ &= [x_s, y_s, z_s] + [x_s, y_s, z_s]\begin{bmatrix} p_{xx} & p_{xy} & p_{xz} \\ p_{yz} & p_{yy} & p_{yz} \\ p_{zx} & p_{zy} & p_{zz} \end{bmatrix} \\ &= [x_l, y_l, z_l] \end{aligned} \tag{11}$$

(1) If the original emotional intensity $\gamma_l = \sqrt{x_l^2 + y_l^2 + z_l^2} > T_i$, the new emotional intensity $\gamma_l' = \gamma_l$. The new emotional state behaves as the one with the largest absolute value among the 3 coordinates, otherwise $\gamma_l' = \gamma_s$ and the new emotional state behaves as the original emotional state.

(2) If $\gamma_l' > 1$, then let $\gamma_l' = 1$, and the new emotional state is the intersection of the line connecting the origin and the point of the newly found state with the sphere.

The characteristics of the personality-based and PAD emotion models in the 3D emotion space are that emotional interaction systems with different personality traits can be created according to individual preferences or the requirements of the application, as a way to enhance the effect of the presentation of the emotional capabilities of the digitized characters in the virtual scene.

II. B. Digital character modeling incorporating emotions

II. B. 1) Generating Adversarial Networks

Generative Adversarial Network (GAN) is a new neural network framework that consists of two main components, the generator (G) and the discriminator (D) [18]. The generator takes random noise as input and generates corresponding “fake” data, while the discriminator evaluates whether the input data is “fake” or real data generated by the generator. The process can be likened to a game between counterfeiters and the police, with both sides constantly improving their abilities to achieve better results. The generator needs to continuously improve its ability to generate “fake” data in order to achieve the effect of fake to real. The discriminator needs to continuously improve the ability to judge the authenticity of the input data to achieve the effect of correctly distinguishing between “true” and “false” data. In the process of dynamic game, both sides continue to improve their respective capabilities, and ultimately reach a state of equilibrium, when both sides can not let the other party fail. the training process of GAN is such a dynamic game process, the generator's goal is to generate the most realistic data possible, and the discriminator's goal is to correctly differentiate between the real data and the generator-generated “fake” data as much as possible. The goal of the discriminator is to correctly distinguish between the real data and the “fake” data generated by the generator. Both sides of the dynamic game will eventually converge to a Nash equilibrium, at which time the generator can generate very realistic “fake” data, and the discriminator can also accurately distinguish between the real data and the “fake” data generated by the generator.

In this paper, the proposed method adopts the idea of adversarial generative network, which is applied to the simplification of skeleton topology and extraction of key information of digitized characters. Through the process of dynamic gaming, the generator can continuously improve the ability of generating realistic data, so as to be able to generate more realistic human skeletal structure. At the same time, the discriminator can continuously improve its ability to distinguish the real data from the “fake” data generated by the generator, so that it can more accurately determine the degree of realism of the skeletal structure generated by the generator. In this dynamic game, the generator and the discriminator can cooperate with each other to achieve better results.

(1) The training process of the generator

First, the parameters of the discriminator D are fixed when the generator is trained, and then the generator G will attach the label G(z) of the “fake” sample to the random vector z from the prior distribution through the neural network. Afterwards, the “fake” sample G(z) is passed along with the true sample to the discriminator D, which calculates the error based on the difference between the true label and the predicted value of the input sample. Finally, the backpropagation algorithm will update and optimize the parameters of the generator G in conjunction with the error.

(2) The training process of the discriminator

First, the parameters of the generator G need to be fixed, and then, the generator G will convert random vectors from the prior distribution into negatively labeled “fake” samples G(z) through the neural network, and samples x, which are collected in the real data, are recorded as the real samples with positive labels. The positive sample x and the negative sample G(z) are passed together to the discriminator D, which calculates the error based on the difference between the predicted values of the real label and the input sample label. Finally, the backpropagation algorithm updates and optimizes the parameters of the discriminator D based on the previously calculated errors [19].

There are two sources of the discriminator D, the "fake" data produced by the generator G and the real data of the original sample, the discriminator needs to fulfill a binary classification task, i.e., to identify the positives and negatives of the samples, and the D loss function is defined as:

$$L(D) = -\int p(x)\left[p(data \mid x)\log D(x) + p(g \mid x)\log(1 - D(x))\right]dx \tag{12}$$

where D is the discriminator, for the input sample data, D needs to discriminate the probability $D(x)$ that it is the correct sample. $p(g \mid x)$ denotes the probability of the sample data generated by the generator. Assume that half of the input sample x comes from real data and half from G -synthesized "fake" data, i.e., $p_{scr}(data) = p_{scr}(g) = 1/2$. The probability of the input sample x is:

$$p(x) = p_{scr}(data)p(x \mid data) + p_{scr}(g)p(x \mid g) \tag{13}$$

Again, based on the equation:

$$p_{scr}(data)p(x \mid data) = p(x)p(data \mid x) \tag{14}$$

$$p_{scr}(g)p(x \mid g) = p(x)p(g \mid x) \tag{15}$$

Substituting the above equation into the loss function yields:

$$L(D) = -\frac{1}{2}\left(E_{x\sim p(x|data)}\left[\log D(x)\right] + E_{x\sim p(x|g)}\left[\log(1 - D(x))\right]\right) \tag{16}$$

Further, the objective function of the generative adversarial network can be derived as:

$$V(G,D) = -\frac{1}{2}\left(E_{x\sim p(x|data)}\left[\log D(x)\right] + E_{x\sim p(x|g)}\left[\log(1 - D(x))\right]\right) \tag{17}$$

The discriminator wants to maximize Eq. $V(G,D)$ with $D(x)$ close to 1 and $D(G(z))$ close to 0, i.e., the discriminator can accurately identify the sample data as true or false, while the generator wants to minimize Eq. $V(G,D)$ so that $D(G(z))$ is close to 1, when the discriminator is deceived into believing that the generated data $G(z)$ are all true data. Thus, the optimization problem of generating adversarial networks is converted into a min-max problem, i.e.,:

$$\min_G \max_D V(G,D) \tag{18}$$

In the actual training process of GAN, the optimal discriminator D_G^* under the current conditions is first found when the parameter conditions of G are fixed. For a single sample, the optimization problem at this point is:

$$\max_D p(x \mid data)\log D(x) + p(x \mid g)\log(1 - D(x)) \tag{19}$$

The optimal solution can be solved as:

$$\hat{D}(x) = \frac{p(x \mid data)}{p(x \mid data) + p(x \mid g)} \tag{20}$$

Integrating over x yields the optimal solution of $\max_D V(G,D)$ as:

$$D_G^* = \frac{p_{data}}{p_{data} + p_g} \tag{21}$$

Then, the parameter conditions of the discriminator D are fixed to optimize the generator G and the optimization problem becomes:

$$\min_G V(G, D_G^*) = \min_G \left\{2 * JSD(p_{data} \parallel p_g) - \log 4\right\} \tag{22}$$

where $JSD(\cdot)$ is the JS scatter between the two distributions. From Eq. It can be seen that the generator G is minimizing the JS scatter between the two distributions p_{data} and p_g, i.e., the JS scatter between the sample distribution of the real data and the synthetic sample distribution. Finally, the GAN will reach the Nash equilibrium of the generator and the discriminator at the minimum point of $JSD(p_{data} \| p_g)$, after which the generator and the discriminator will not continue to optimize.

II. B. 2) First-order motion model

First-order motion modeling (FOMM) learns target object keypoints in an image in a self-supervised manner without relying on a priori information about the target, such as the pose point information of a human body or the landmark point information of a face, which allows FOMM to drive the motion of arbitrary target objects [20].

The keypoint detection module in FOMM uses a standard U-Net network structure to extract sparse keypoints of interest in the image, as well as a local affine transformation with respect to the reference frame. The last layer of the U-Net decoder uses SoftMax activation to predict a heat map that can be interpreted as a keypoint detection confidence map, and then the keypoint location is obtained by summing the product of the coordinates of the individual pixel points of the image and the heat map weights. Then:

$$u^k = \sum_{z \in Z} M^k(z) z \tag{23}$$

$\sum_{z \in Z} M^k(z) z = 1$, z is the pixel coordinate (x, y) in the image and the set of all pixel coordinates is Z. $M^k(z)$ is the weight value of the k th heat map at pixel coordinate z.

In addition the keypoint detection module uses the heat map of the keypoints to predict the affine transformation $A^k_{S \leftarrow R}$ from the reference frame to the source frame and the affine transformation $A^k_{D \leftarrow R}$ from the reference frame to the driving frame, and thus predicts the affine transformation between the keypoint transitions of the driving frame and the source frame $A^k_{S \leftarrow D}$, computed as:

$$A^k_{S \leftarrow D} = A^k_{S \leftarrow R} \left(A^k_{D \leftarrow R} \right)^{-1} \tag{24}$$

The dense network prediction module also follows the standard U-Net network structure, and for each keypoint, the heat map $H_k(z)$ is computed for each transformation according to Eq. (25), where var is a hyperparameter that is usually, empirically, taken to be 0.01. The sparse optical flow field $\mathrm{F}_k(z)$ is then computed according to Eq. (26) for each keypoint. The optical flow field prediction module outputs an intermediate feature map ξ by combining the heat map $H_k(z)$ and the distortion result after distorting the source frames using $\mathrm{F}_k(z)$, and finally outputs the network-predicted dense optical flow field $\mathrm{F}(z)$ and the occlusion map O_{s+D}, as shown in Eqs. (27) and (28). Namely:

$$H_k(z) = e^{-(z-u^k_{D \leftarrow R})^2/(2*var)} - e^{-(z-u^k_{S \leftarrow R})^2/(2*var)} \tag{25}$$

$$\mathrm{F}_k(z) = u^k_{S \leftarrow R} + A^k_{S \leftarrow D}(z - u^k_{D \leftarrow R}) \tag{26}$$

$$\mathrm{F}(z) = M_0 z + \sum_{k=1}^{K} Conv_{7\times7}(\xi) * \mathrm{F}_k(z) \tag{27}$$

$$\mathrm{O} = \mathrm{Conv}_{7\times7}(\xi) \tag{28}$$

Finally, the generator module follows the codec structure, and the source frame image is fed into the generator module, which undergoes two downsampling convolution modules to obtain a feature mapping $\zeta \in \mathbb{R}^{H'\times W'}$ of the source frame of dimension $H' \times W'$. However, the source frame at this point is not subjected to a pixel-to-pixel alignment operation with the driver image to be generated, and in order to deal with this misalignment, the FOMM uses a dense optical flow field $\mathrm{F}(z)$ to perform a distortion operation on the feature mapping ζ of the source frame. In addition, the FOMM considers that when there is occlusion in the source frame S, the optical flow field may not be sufficient to generate the driving frame D from the source frame S. Since, the occluded portion of the source frame S cannot be repaired by image warping, the FOMM introduces an occlusion map $\mathrm{O}_{S \leftarrow D} \in (0,1)^{H'\times W'}$ to mask the feature map region that should be patched. The processed feature maps are as follows:

$$\zeta' = \mathrm{O}_{S \leftarrow D} \odot f_w(\zeta, \mathrm{F}(z)) \tag{29}$$

where $f_w(\cdot,\cdot)$ denotes the backward warping operation and denotes the dot product operation. Finally, the processed ζ' is fed into the subsequent network layer of the generator module to generate the final driving frame D.

II. B. 3) Digital character synthesis

In order to realize the digital emotion modeling of film and television art characters, this paper combines FOMM and GAN to establish a first-order digital character synthesis model, and its specific framework is shown in Fig. 2. The inputs of the first-order motion model are the source photo and the driving video, using the key point detection network for image feature extraction, and combining the SoftMax function to generate a number of key points, which are computed to obtain their Jacobi matrix, and in order to synchronize the processing of the source photo and the driving video, an abstract reference frame is designed so that the network can synchronize the detection of the key points as well as the generation of the Jacobi matrix. Generation. The sparse motion field is solved using the Jacobi matrices of the source photos and videos. The dense optical flow field is then generated using the source image combined with the sparse motion field to generate a mask. The generator network uses the source photo combined with the dense optical flow motion field and the mask to synthesize and output the source photo person picture containing the motion poses that drive the video.

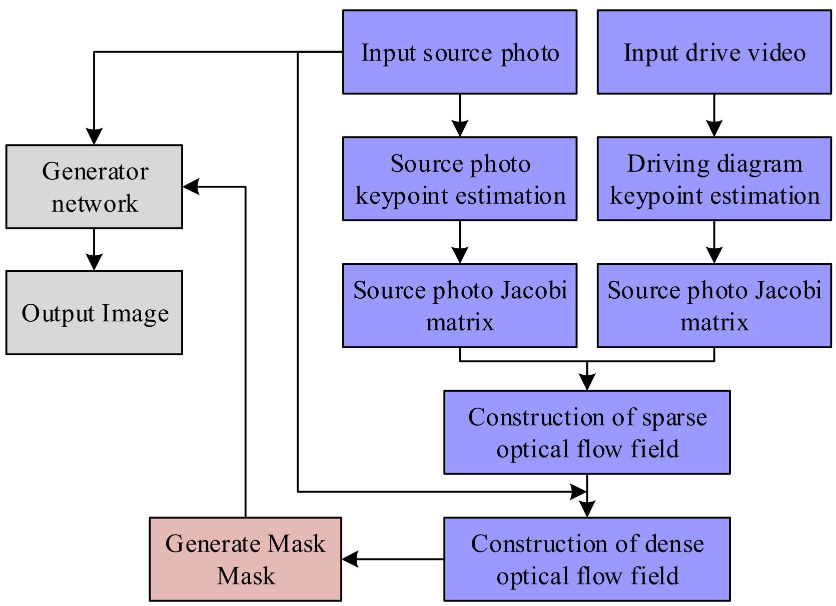


Figure 2: Order number character synthesis model

FOMM introduces a occlusion-aware generator, which uses an automatically estimated occlusion mask to indicate the portion of the object in the source image that is not visible and should be inferred from the context. It uses a first-order Taylor expansion to record changes in motion that reflect changes in different motions on the image, and FOMM estimates changes in motion trajectories over two images that are not necessarily consecutive.

II. C.Framework for Emotional Digital Persona Realization

II. C. 1) 3D Virtual Environment Design

In order to better show the ability of character emotion expression supported by digital modeling, this paper adopts the ECS architecture of Unity3D, which provides basic design guidelines for the application and improves the development efficiency of the application. At the same time, we refer to OSG to modularize the functions of the engine, and the modules are independent of each other. Each module provides an interface to the outside world, and other modules only need to call the interface of this module without knowing the specific logic, so each module only needs to deal with its own business logic, which is convenient for the implementation of the function and the subsequent update and maintenance. As the rendering engine may run on multiple platforms, it should support a certain degree of portability. After a series of analysis of this engine, the engine's architectural design is shown in Figure 3.

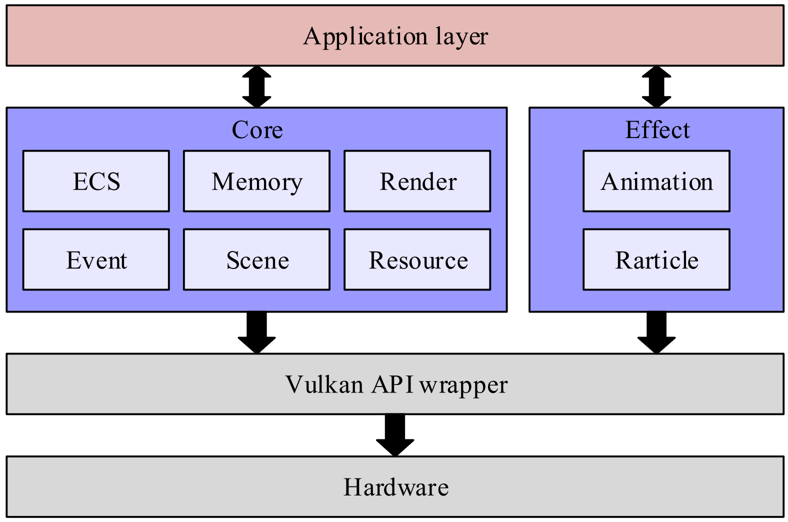


Figure 3: Three-dimensional virtual environment design

The bottom layer of the rendering engine is the Vulkan API wrapper layer. The API layer encapsulates Vulkan-related APIs and is responsible for interacting directly with the hardware and providing interfaces to the upper layers. The core layer of the engine is located between the Vulkan API wrapper layer and the application layer, and is responsible for calling the Vulkan API layer and providing interfaces to the application layer. The engine core layer is mainly divided into ECS module, memory management module, rendering module, event management module, scene management module, resource management module and other functional modules.

The architectural design of the engine greatly decouples the various modules of the engine, and the modules are independent of each other, which is convenient for subsequent development and maintenance. The engine implements the ECS module, which provides design specifications for the upper layer applications and can well realize the required functions. In addition, the graphics API layer of the engine is encapsulated separately and can be replaced by other graphics APIs in a relatively simple way.

II. C. 2) Overall consideration of digital personas

For the characteristics of the 3D virtual environment, we have the following considerations for the design of the digital character:

The response of the digital character to the stimulation of the surrounding environment should be rapid and stable. Because the digital character is in a dynamic environment, its fast and reasonable response to external stimuli is an important reflection of its intelligence. The virtual environment is a dynamic environment where new situations are constantly occurring. In this way, the digital character has to be able to deal with unexpected events that interrupt the original behavior. Therefore, the digital persona cannot adopt the traditional “sense-plan-act” cycle. The digital persona should have a wide range of capabilities. She should be able to perceive her surroundings, have beliefs about her surroundings and her own state of being, and have a certain internal state of mind. She should be able to make behavioral choices and take certain actions in response to goals, and interact with the user.

Based on the above considerations, we constructed the overall framework of the digital character with aggregate modifiers and geometric modeling as actuators. The geometry modifier receives control commands and parameters from the action engine, and directly modifies the attributes in VRML so that the geometry of the character changes to represent certain actions. The geometry modifier is implemented by EAI and is embedded in a web page through an Applet and downloaded to the client browser. The characters are geometrically modeled in the VRML language and conform to the H-Anim character modeling standard, which is very generic and guarantees many aspects of realism for the characters. The successful completion or failure of an action should play a feedback message to the server-side digital character controller so that the behavior engine can make a decision on whether or not to proceed to the next action, and reflect the results of that action to the environment as well as to its own internal state of mind.

III. Evaluation and analysis of emotional performance of digital characters

In the era of experience economy, consumers' demand for film and television art and culture is gradually growing, and digital characters, as an important part of film and television art, have a direct impact on the communication effect of film and television art in terms of their ability to express their emotions. Therefore, it is of great theoretical and practical significance to explore the emotional expression design of virtual digital characters to provide guidance and optimization suggestions for the expression of character emotions in film and television art.

III. A. Validation of digital character synthesis effects

III. A. 1) Experiments with affective computing methods

In the PAD 3D emotion space, the main emotions included are pleasure, surprise, disgust, fear, sadness and anger, and simulation experiments on the effectiveness of different emotion calculation methods are conducted in the MATLAB development environment. In order to verify the accuracy of the emotion calculation methods in this paper compared with the emotion calculation methods based on OCC and fuzzy reasoning (FR), 30 groups of data labeled with six basic emotions are selected in this paper, and the PAD space vector values are obtained by emotion cognitive reasoning with two emotion calculation methods respectively, and then the accuracy rate is calculated based on the table of the correspondence relationship between the PAD space quadrant and the emotion. The accuracy rates of different calculation methods are shown in Figure 4.

Based on the data distribution in the figure, it can be seen that the average accuracy rate of the emotion calculation method proposed in this paper is 84.78%, which is 14.86% and 5.16% higher than the average accuracy rate of the emotion calculation methods based on OCC and fuzzy reasoning (FR), respectively. Since the PAD mind space is not an isotropic homogeneous Euclidean metric space, the fuzzy inference-based affective calculation method is not the most appropriate to calculate the affective distance directly with the formula of the distance between two points in the space, and the OCC model-based affective calculation method has the advantage of being simple to calculate, but increases the risk of incorrectly judging the unknown affective points when the basic affective points are close to the coordinate axes. For example, sadness and anger are in the same quadrant in the PAD emotion reference point, but the value difference is large, and in the quadrant correspondence table, sadness is classified into another quadrant, which is also a limitation of the method. In this paper, the method takes into account the influence of the variance of the basic emotional data on the calculation of emotional distance, and also introduces variables such as personality and emotional intensity, thus avoiding the controversial part of the method of quadrant correspondence table, effectively improving the accuracy of the calculation of emotions, and making it feasible to apply it to the calculation of emotions in digital characters.

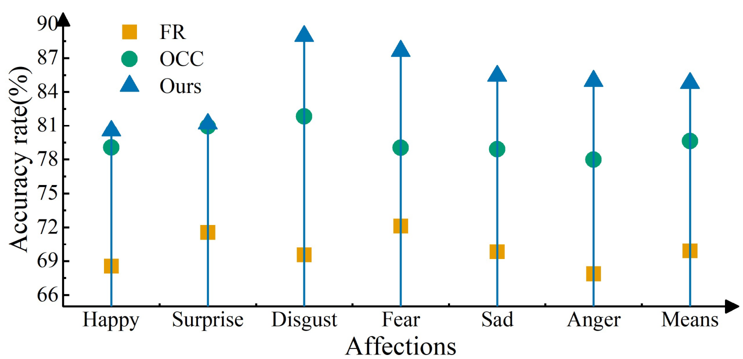


Figure 4: The accuracy rates of different calculation methods

III. A. 2) Digital character synthesis effects

In order to realize the modeling of digital characters in 3D virtual space, this paper proposes a digital character synthesis model (FOMM-GAN) based on the first-order motion model and generative adversarial network, and the design and implementation of the model as well as the attack comparison experiments are done on Linux operating system. The software environment is python and TensorFlow deep learning framework, and NVIDIA GeForce GTX 3090 Ti graphics card is used to accelerate the training of the experiments.

This section first presents the stability comparison between the FOMM-GAN based digital character synthesis model and the WGAN based generative model. Then peak signal-to-noise ratio (PSNR), average gradient (AVG), and runtime are used as evaluation metrics to measure the quality of digital character synthesis by comparing multiple models.

Figure 5 shows the comparison of the total loss of different models. The total generator loss consists of generator loss, identity loss, and perturbation loss. Among them, the generator loss indicates the loss of synthesizing mask, the identity loss indicates the loss of the target model when synthesizing, and the perturbation loss indicates the loss of adding perturbation when synthesizing confrontation samples, and the smaller the total loss is, the better the model is. From the curve change in the figure, the total loss of this paper's model changes smaller and faster, which also indicates that the stability of this paper's model is better compared to the WGAN model. When realizing the synthesis of digital characters, it can effectively solve the problem of the poor effect of digital character synthesis caused by the gradient explosion of the model.

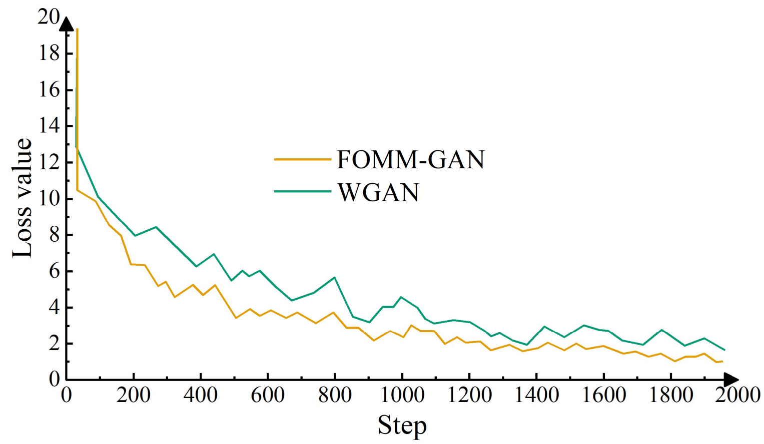


Figure 5: The comparison of total losses of different models

In order to further verify the cleaning degree and quality of the digital characters synthesized by the FOMM-GAN model, this paper chooses PSNR, AVG and running time as evaluation indexes, and selects CycleGAN, DiscoGAN, ICGAN, StarGAN, and GANImation as comparisons, and obtains the comparative results of different methods as shown in Table 1.

From the data in the table, the digital characters synthesized using the method of this paper have PSNR and AVG values of 35.63 and 5.96, respectively, which are significantly higher than the other five comparison methods. This fully demonstrates that the digital character synthesis model designed in this paper can better maintain the details and clarity of the original photos, and the synthesized digital character images are higher in clarity and better in quality. In addition, in the comparison results of running test time, under 1000 training images, the algorithms based on DiscoGAN model and based on IcGAN model have a slightly smaller test time difference between them, which is only 0.18%. The algorithm based on StarGAN model and the algorithm based on GANImation model, they have relatively long test times, both exceeding 100 ms. while the running test time of this paper's model when performing digital character synthesis (94.27 ms) is significantly less than the other five algorithms. In addition, the training time belongs to the preprocessing stage, which can be performed offline. Therefore, the FOMM-GAN digital character synthesis model constructed based on the first-order motion model and generative adversarial network in this paper has a faster running speed and can achieve real-time synthesis of digital characters.

Table 1: The comparison results of different methods

Model	PSNR	AVG	Test time	
			Quantity	Time(ms)
CycleGAN	19.31	2.31	1000	97.89
DiscoGAN	20.48	2.74	1000	96.42
IcGAN	20.24	3.57	1000	96.25
StarGAN	29.57	3.42	1000	102.38
GANImation	31.19	3.28	1000	106.54
FOMM-GAN	35.63	5.96	1000	94.27

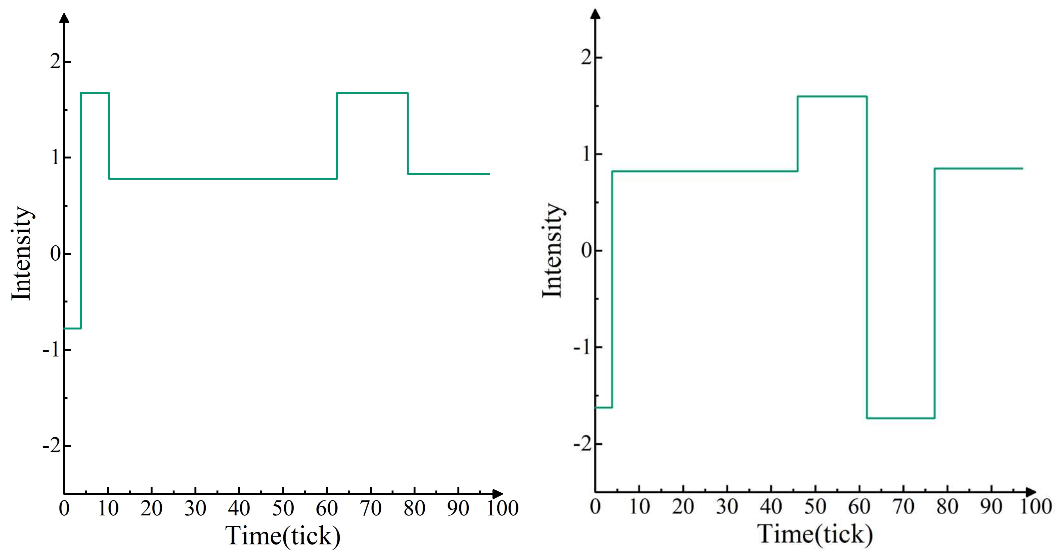
III. B. Analysis of Digital Character Emotional Performance

III. B. 1) Changes in the character's emotional state

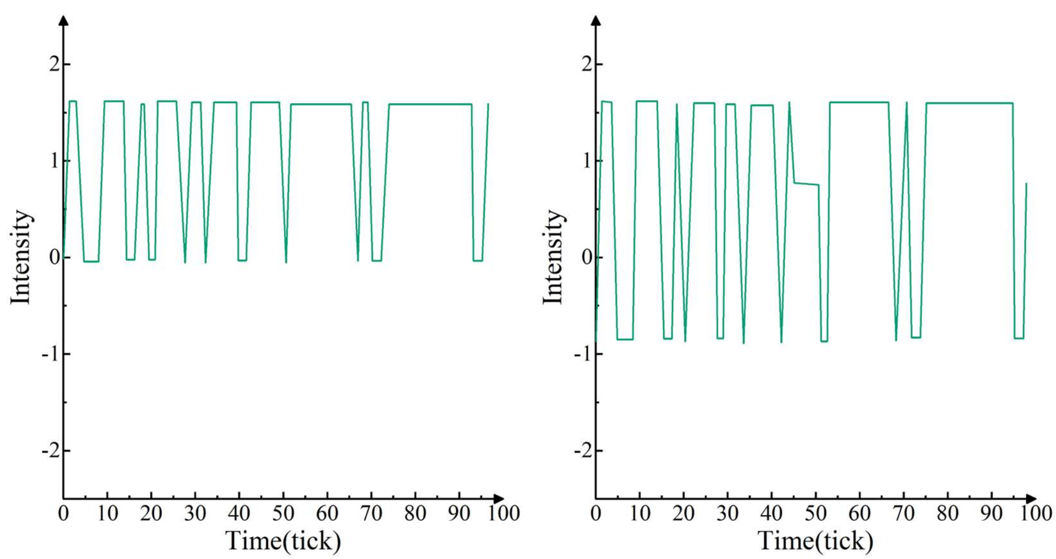
After establishing the basic model of the digital character using PAD emotion calculation and FOMM-GAN model, the 3D model construction of the digital character is realized under the 3D virtual engine constructed in this paper, and it is mapped and simulated on the simulation platform Netlogo to verify and explore the change of the emotional state of the digital character.

Based on the previous definition of the emotional intensity of digital characters, this paper sets up two different types of simulation scenarios on the simulation platform, i.e., learning scenarios and entertainment scenarios, and in order to be closer to the real scenarios, this paper assumes that the emotional abilities in the learning scenarios and the entertainment scenarios have a certain degree of delay. After setting the relevant parameters, the emotional state and intensity changes of digital characters in different scenes are counted, and the results are shown in Fig. 6. Among them, Fig. 6(a)~(b) shows the change of emotional state and intensity under the learning scene, and Fig. 6(c)~(d) shows the change of emotional state and intensity under the entertainment scene, respectively.

From the figure, it can be seen that the emotional intensity value under the learning scene is still 1.7 from tick=10 to tick=63, while the two corresponding emotional intensity values under the entertainment scene change from 0 at tick=10 to about 1.6 at tick=63, which is obviously stronger than the corresponding emotional intensity under the learning scene, and the span of the entertainment scene's own emotional intensity is This indicates that the digital characters will choose the entertainment scene as their final choice based on their own emotional factors. This fully demonstrates that digital characters have strong emotion calculation and expression ability, and can effectively realize their own emotion expression.



(a) Learning scene - Affective state (b) Learning scene - Intensity variation



(c) Entertain scene - Affective state (d) Entertain scene - Affective state

Figure 6: Changes in the emotional state of the characters

III. B. 2) Level of emotional characterization experience

In this study, 80 students were randomly selected from all the students majoring in Artificial Intelligence related majors in a university in City H to conduct the experiment, which mainly aimed to explore the influence of the emotional performance of digital characters on their emotional experience level. A total of 80 scales were distributed and 80 scales were retrieved, and the statistics of the experimental data as well as the subsequent analysis of the data were carried out using SPSS software. In order to assess the quality of the emotional performance of the digital characters, this study measured the emotional perception (Q1), emotional adaptability (Q2), emotional anthropomorphism (Q3) and emotional involvement (Q4) and initiative (Q5) of the digital characters, and collected the data by using a five-point Likert scale (-2~2), and the mean statistic is now utilized to target the level of the subject's experience of the emotional characteristics of the digital characters. Descriptive analysis is carried out, for each emotional characteristic dimension, the higher the mean value, the higher the level of the subjects' experience of the emotional characteristic, Table 2 shows the statistical results.

On the whole, the mean values of the subjects' experience level of each dimension of emotional characteristics of the digital character have reached a high level, and most of the subjects' experience scores are concentrated on 1 and 2, which indicates that the subjects' perception level of the emotional characteristics of the digital character is high. That is, the degree of recognition of the digital character's ability to express emotion is high, and the overall level of embodiment of the emotional characteristics of the digital character is high. From the various dimensions of the emotional characteristics of the digital character respectively, first of all, in terms of the subjects' perception of the digital character and the adaptability of the two aspects, basically all the subjects very much agree that the digital character has a basic emotion, indicating that they can obviously feel the emotional changes of the digital character. And they experienced that the emotional changes of the digital characters were in line with the law of change of emotions in life, and could follow the development of the process, while only a few subjects did not have obvious feelings about the emotional changes of the digital characters. In terms of the anthropomorphism of the digital character's emotion, most of the subjects thought that the digital character's emotion change had human-like nature and could change their emotion naturally and smoothly. A small number of subjects questioned the anthropomorphic nature of the emotions of digital characters, and some subjects did not recognize this feature of digital characters, and the overall mean value of this feature was slightly lower than that of the other feature dimensions, indicating that the anthropomorphic nature of the emotional changes of digital characters still needs to be further optimized and improved. In terms of the degree of emotional involvement of digital characters, most of the subjects think that digital characters are very expressive of emotions, have a strong infectious power, and can be infected and inspired by the emotions of digital characters. The higher ratings of the digital character's proactivity indicate that the digital character can actively communicate and interact with the subjects, providing them with a friendly and warm emotional interaction experience.

Table 2: The level of experience of emotional characteristics

Index		-2	-1	0	1	2	Means	SD
Q1	Q11	0	1	1	22	56	1.654	0.346
	Q12	0	0	1	30	49		
	Q13	0	0	0	24	56		
Q2	Q21	0	1	2	21	56	1.688	0.358
	Q22	0	0	2	18	60		
	Q23	0	1	3	16	60		
Q3	Q31	0	0	5	15	60	1.538	0.416
	Q32	0	0	10	27	43		
	Q33	0	0	4	31	45		
Q4	Q41	0	1	8	11	60	1.513	0.407
	Q42	0	2	2	23	53		
	Q43	0	1	4	43	32		
Q5	Q51	0	2	9	15	54	1.484	0.379
	Q52	0	1	11	26	42		
	Q53	0	0	8	18	54		

IV. Conclusion

Digital modeling technology provides important support for the improvement of character emotion expression ability in film and television art. Through experimental verification, the emotion calculation method based on PAD 3D emotion space shows obvious advantages in emotion cognitive reasoning, and its accuracy rate reaches 84.78%, which is 14.86% and 5.16% higher than that of the emotion calculation method based on OCC and fuzzy reasoning, respectively. The constructed FOMM-GAN digital character synthesis model outperforms the existing models in terms of image quality and processing efficiency, with a peak signal-to-noise ratio of 35.63, an average gradient of 5.96, and a test time of only 94.27ms for processing 1000 images, which saves 8.11ms and 12.27ms compared to the models such as StarGAN and GANImation, respectively, for the processing time. Verified by the simulation platform, it is found that the digital character can autonomously adjust the emotional state according to the emotional factor in different scenes, and the emotional intensity value in the entertainment scene can reach up to 1.6, showing good emotional calculation and expression ability. The user experience evaluation shows that the mean value of the subjects' emotional perception of the digital characters reaches 1.654, and the mean value of the adaptability is 1.688, indicating that the constructed digital characters can be effectively perceived and emotionally resonated by the users.

The enhancement of the ability of digital modeling to show the emotion of the characters provides a new technical path for the creation of film and television art, which helps to improve the immersion and infectiousness of the film and television works, and promotes the sustainable development of film and television art in the era of experience economy. Future research should further explore the multimodal emotion expression mechanism to enhance the anthropomorphism and naturalness of digital character emotion expression.

References

[1] Jiang, R., Wang, L., & Tsai, S. B. (2022). An empirical study on digital media technology in film and television animation design. Mathematical Problems in Engineering, 2022(1), 5905117.

[2] Tan, S. (2022). Animation image art design mode using 3D modeling technology. Wireless Communications and Mobile Computing, 2022(1), 6577461.

[3] Du, J. (2024). The Application of Computer-Aided Film and Television Art Design Elements in Exhibition Space Design. International Journal of High Speed Electronics and Systems, 2540113.

[4] Rubio-Tamayo, J. L., Gertrudix Barrio, M., & García García, F. (2017). Immersive environments and virtual reality: Systematic review and advances in communication, interaction and simulation. Multimodal technologies and interaction, 1(4), 21.

[5] Li, L. (2022). The influence of digital twins on the methods of film and television creation. Computers and Electrical Engineering, 103, 108314.

[6] Sun, L. (2022). Research on Digital Media Art Film and Television Special Effects Technology Based on Virtual and Reality Algorithm. Scientific programming, 2022(1), 4424772.

[7] Garstki, K. (2017). Virtual representation: the production of 3D digital artifacts. Journal of Archaeological Method and Theory, 24, 726-750.

[8] Keller, M., Werling, K., Shin, S., Delp, S., Pujades, S., Liu, C. K., & Black, M. J. (2023). From skin to skeleton: Towards biomechanically accurate 3d digital humans. ACM Transactions on Graphics (TOG), 42(6), 1-12.

[9] Liu, L., Xu, W., Zollhoefer, M., Kim, H., Bernard, F., Habermann, M., ... & Theobalt, C. (2019). Neural rendering and reenactment of human actor videos. ACM Transactions on Graphics (TOG), 38(5), 1-14.

[10] Xia, S., Gao, L., Lai, Y. K., Yuan, M. Z., & Chai, J. (2017). A survey on human performance capture and animation. Journal of Computer Science and Technology, 32, 536-554.

[11] Sharma, S., Verma, S., Kumar, M., & Sharma, L. (2019, February). Use of motion capture in 3D animation: motion capture systems, challenges, and recent trends. In 2019 international conference on machine learning, big data, cloud and parallel computing (comitcon) (pp. 289-294). IEEE.

[12] Habermann, M., Liu, L., Xu, W., Zollhoefer, M., Pons-Moll, G., & Theobalt, C. (2021). Real-time deep dynamic characters. ACM Transactions on Graphics (ToG), 40(4), 1-16.

[13] Kumaran, V., Rowe, J., Mott, B., & Lester, J. (2023, October). Scenecraft: automating interactive narrative scene generation in digital games with large language models. In Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment (Vol. 19, No. 1, pp. 86-96).

[14] Kuang, R. (2021). Design and implementation of 3D film and television scene production algorithm based on the internet of things. Wireless Communications and Mobile Computing, 2021(1), 1219849.

[15] Yu, H., Wang, C., Zhuang, P., Menapace, W., Siarohin, A., Cao, J., ... & Lee, H. Y. (2024). 4real: Towards photorealistic 4d scene generation via video diffusion models. Advances in Neural Information Processing Systems, 37, 45256-45280.

[16] Wu, Z., Rubanova, Y., Kabra, R., Hudson, D., Gilitschenski, I., Aytar, Y., ... & Kipf, T. (2024). Neural assets: 3d-aware multi-object scene synthesis with image diffusion models. Advances in Neural Information Processing Systems, 37, 76289-76318.

[17] Xu Wuhuan,Yao Zhong,He Dandan & Cao Ling. (2025). Understanding online review helpfulness: a pleasure-arousal-dominance (PAD) model perspective. Aslib Journal of Information Management,77(2),391-412.

[18] Fiammetta Marulli,Fabio Lancellotti,Pierluigi Paganini,Giovanna Dondossola & Roberta Terruggia. (2025). Towards a novel approach to enhance cyber security assessment of industrial energy control and distribution systems through generative adversarial networks. Journal of High Speed Networks,31(2),105-122.

[19] Aman Kumar,Afshin Marani & Moncef L. Nehdi. (2025). Generative adversarial network approach for predicting tensile behavior and failure pattern of fiber-reinforced cementitious matrices. Engineering Structures,335,120276-120276.

[20] Yu Xu,Feng Xu,Qiang Liu & Jianwen Chen. (2023). Improved First-Order Motion Model of Image Animation with Enhanced Dense Motion and Repair Ability. Applied Sciences,13(7).