Hindawi
Mobile Information Systems
Volume 2022, Article ID 7539240, 13 pages
https://doi.org/10.1155/2022/7539240

Research Article

Metaverse Space Ecological Scene Design Based on Multimedia Digital Technology

Xuheng Xu,[1] Guiheng Zou,[2] Lifeng Chen ,[3,4] and Ting Zhou[5]

[1]*Theatre and Film Institute, Zhejiang Vocational Academy of Art, Hangzhou 310053, Zhejiang, China*
[2]*Security and Protection Department, Zhejiang Police Vocational Academy, Hangzhou 310018, Zhejiang, China*
[3]*School of Public Affairs, Zhejiang University, Hangzhou 310058, Zhejiang, China*
[4]*School of Business, Zhejiang University City College, Hangzhou 310015, Zhejiang, China*
[5]*College of International Business, Zhejiang Yuexiu University of Foreign Languages, Shaoxing 312000, Zhejiang, China*

Correspondence should be addressed to Lifeng Chen; chenlifeng@zucc.edu.cn

Received 29 March 2022; Revised 28 April 2022; Accepted 6 June 2022; Published 6 July 2022

Academic Editor: Liming Chen

Copyright © 2022 Xuheng Xu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Yuan cosmos is a virtual world linked and created by scientific and technological means, which is mapped and interacted with the real world, and a digital living space with a new social system. With the increasing popularity of data acquisition and production equipment, people are increasingly convenient to produce multimedia data such as images, graphics, audio, video, animation, and three-dimensional models. In addition to the rapid development of digital technology itself, the biological information technology related to digital technology also greatly promotes the emergence of the metauniverse. This paper aims to study the application of multimedia digital technology to the ecological scene design of metauniverse space, introduces the related concepts of metauniverse and multimedia digital technology, expounds the related methods of multimedia digital technology and neural network related algorithms, and then takes the three-dimensional simulation of the auditory system in the interactive multisensory simulation system of the constituent elements of metauniverse as an example. The mel-frequency cepstrum coefficient (MFCC) is used to simulate the auditory characteristics of the auditory periphery (cochlea) as the perceptual end of the model. A variety of bionic mechanisms are used in the model, such as designing the connection mode of neurons, learning state and release effect, and the regeneration mechanism of neurons. For the verification of the performance of the model, the speech sample database, including English words and phrases, is recorded and the speech content information recognized by the model by means of speech recognition is experimented. The experimental results show that, in terms of phrase accuracy, the DN-1 model improves 2.59% and the DN-2 model improves 2.77% compared with MFCC feature on the basis of mixed features. When only DBN features are used, the performance improvement rate of the developmental network model is small.

1. Introduction

Multimedia digitization is an innovative technology brought about by the development of modern computer technology. Digital media technology has not only changed the original mode of radio and television production and broadcasting but also spawned many new production and broadcasting platforms involved in collecting, editing, and broadcasting. The application of digital multimedia is suitable for platforms for the development of modern network environments. With the existence of mobile Internet and growing ubiquitous perception systems, such as cameras and voice controllers citywide, the entire city can be effectively collected and processed in real time. The emergence of smart phones has greatly promoted the digital mapping process of individuals. Therefore, smart phones have greatly promoted the collection of personal data. Using smart phones can accurately make data portraits for users, thus forming massive individual uploads of natural information, thereby building the foundation of Metaverse Space for personal information and social interaction. Big data has also promoted the development of artificial intelligence accordingly. The rapid development of artificial intelligence in the 20th century is essentially due to the rapid accumulation of data

materials, thus forming an evolutionary path with "deep neural network + big data" as the core method and changing the long-standing limitations of symbolic logic methods on the development of artificial intelligence.

Judging from the status quo of artificial intelligence, although it still lags behind humans in overall cognition and adaptability, it has approached or surpassed the human levels in most subfields. The development of artificial intelligence has laid the foundation of intelligent control technology for the development of the Metaverse Space. In this context, the traditional concept of "interface design" has been broken and the interactive display experience design in a brand-new multimedia environment is carried out. Through the research of the paper, it can promote further development of the comprehensive application of multimedia interaction technology in the Metaverse Space ecological scene and quickly improve their ability to use interactive technology in multimedia interface design. By in-depth research on the realization and research of various elements in multimedia interactive production technology, the paper explores the application of designing a brand-new interactive display and experience environment through the effective use of multimedia interactive technology.

The innovation of this paper is that (1) it takes digital media art as the research object, deeply studies and explores the radar application cases of digital media art in virtual reality technology, and summarizes the relevant theoretical exploration and forward-looking prediction for the future. (2) In this paper, multimedia digital technology is applied to the ecological scene design of metaspace, which is innovative and practical.

2. Related Work

With the passage of time, digital media art has gradually infiltrated our daily lives. It is not only a new technology but also a new way of life and aesthetic form. A wealth of research has been carried out by scholars on digital media arts. Under the background of contemporary cultural protection and dynamic inheritance, the interpretation and re-expression of the artistic connotations of Chinese literati paintings have become the main direction of heritage research. Digital technology and multimedia expression have become important means of cultural expression and dissemination. Chinese literati paintings are mostly ink paintings. The particularity of ink paintings makes it difficult to decompose and extract the content of the pictures in a simple way, which to a certain extent causes difficulties in digitization, re-expression, and public interpretation. To solve this problem, Zhang et al. proposed a new robust multiview fuzzy clustering algorithm through multimedia digital technology for image segmentation of Chinese literati paintings to achieve effective decomposition and extraction of ancient paintings. In this way, the electronic and digital conversion and preservation of literati paintings could be realized. This preservation method could preserve the artistry of literati paintings better than the traditional scanning method and is of great value to the re-expression and dissemination of cultural heritage [1]. The world of communications and computing has changed dramatically over the past few years. The advent of social media has made data transfer easier, which has raised issues of unauthorized use and redistribution of digital content. This copyright scheme has hardly affected the publishing rights of authors and publishers. Hassan et al. proposed a robust framework to protect copyright property and prevent illegal use or copying of data in the event that only authorized users could legally use the data [2]. Evolved Multimedia Broadcast Multicast Service (eMBMS) is a technology in Long-Term Evolution (LTE) that provides a broadcast bearer to deliver video content and files to an unlimited number of users. The bearer utilizes multiple cell sites to build a "Single-Frequency Network" (SFN) area with identical downlink transmissions over a portion of the LTE Orthogonal Frequency-Division Multiplexing (OFDM) waveform. The main data needs are the Internet, DVB, and higher-speed cellular broadcasting. The resulting signals are combined at the antenna of the user equipment in such a way that adjacent cell sites, which are usually interfering, become sources of useful signals, thereby improving the overall signal-to-interference ratio and spectral efficiency. Luo introduced eMBMS technology and architecture and evaluated its performance and impact on wireless network engineering [3]. Dongmei believed that compared with new media, the biggest advantage of paper media was that the content was authentic and professional. At the same time, paper media is more in line with people's reading habits. He analyzed the path of the integration of broadcast television and digital network technology based on the multimedia editing platform. By analyzing the background and motivation of the integration of radio and television and new media, he analyzed the reasons for the successful transformation and development of related enterprises, explored the path of radio and television and new media integration, and provided strategic suggestions for the development of China's radio and television [4]. Meng et al. believed that the application of digital multimedia technology has been the inevitable result of the social development of higher education, and it was also an inevitable choice to achieve innovation and development. Therefore, colleges and universities should strengthen the emphasis and investment on digital multimedia technology in the actual teaching process. The national education department should strengthen support and guidance and strengthen the integration of traditional teaching methods and digital multimedia technology [5]. Digital multimedia is a computer-based graphics and image application technology, which is widely used in the field of design engineering. In his research, Li analyzed the optimization and development of urban landscape design under the influence of digital multimedia technology. In actual work, a large number of display drawings and information are usually required to express the intention of landscape design. Li's research found that digital technology has strong practical value in garden design [6]. With the development of new media, student management faces more challenges. New media has a larger network system, including digital technology and mobile technology. Wang and Wang innovated the optimization mode of college student management based on multimedia network platform. The diversification of new media has brought difficulties to the management of college students, and the immediacy of

new media communication has also brought challenges to higher education. Therefore, Wang proposed a student management optimization platform based on information management technology. While using the emerging multimedia education platform, teachers also need to continuously improve management model innovation [7]. The disadvantages of these research studies are that digital media technology is still a relatively new art and technology industry, some theoretical studies are not perfect, and there are still many practical problems that need to be solved.

3. The Method of Multimedia Digital Technology for Metaverse Space Scene Design

The concept of digital media is based on the leading role and method of digital technology in information dissemination. The concepts of "digital media" and "multimedia" are also different. The full name of multimedia should be multisensory media, which is not a kind of media but a technology and method for digital media to encode, process, store, and present information [8].

3.1. The Application Concept of Digital Media Technology in Public Art Creation. The continuous development of digitization has put the current human civilization in a major historical transformation stage, that is, it is facing the singularity of civilization development [9]. The so-called singularity often refers to discontinuous points with important special properties in mathematics and physics. In the social sciences, it refers to those important historical nodes in the evolution of civilization. Since the evolution of human civilization, it has a history of tens of thousands of years. In the long history of human evolution, there are often multiple significant historical moments, which lead human civilization from primitive, backward, and barbaric to a relatively prosperous human civilization [10].

3.1.1. Digital Holographic Projection Technology. Holographic projection technology, also known as holographic 3D technology, is a recording and reproduction technology that uses the principles of interference and diffraction to record all information in the reflected light waves of the object and reproduce the real three-dimensional image of the object [11]. The related pictures are shown in Figure 1.

Holographic projection technology has a wide range of applications and has bright development prospects in education, aerospace, national defense, film and television, entertainment, and art. Many countries and institutions are vigorously researching and developing, and it also has very broad application prospects in the field of public art [12, 13].

3.1.2. VR and AR Technology

(1) Virtual Reality Technology. Virtual reality technology (abbreviated as VR), also known as spiritual environment technology, is a computer system that can create and experience the virtual world. In recent years, medical design, art, real estate, archaeology, military, entertainment, and other fields have begun to apply VR technology to their respective industries, creating large wealth for the society (Figure 2).

Virtual reality technology integrates digital images, sensors, multimedia technology, artificial intelligence, computer graphics, network, and other information technologies, gradually establishes its brand-new development achievements, provides a certain degree of support for the creation and experience of the virtual world, and greatly promotes the rapid development of information technology. The biggest feature of VR technology is to establish an artificial virtual environment through the computer. This environment is to copy other real environments, apply them to the computer, and produce a new "virtual environment" or a three-dimensional space formed only through the computer, and make the user feel immersive.

(2) Augmented Reality Technology. Augmented reality (AR) superimposes the real environment and virtual objects on the same screen or space in real time [14]. It does not present a completely virtual world to users, but superimposes our real world with virtual objects to produce an experience that we cannot obtain in our normal state. It can be said that it is the virtualization and expansion of the real world.

3.2. The Origin of the Metaverse Space and Its Components. With the "Internet+" thinking, wisdom and technology extend their influence on politics, economy, culture, and life. With the development and popularization of mobile Internet and the formation of users' user habits, the growth momentum of mobile terminal users has slowed down and the dividend of consuming Internet is gradually decreasing. The outbreak of the pandemic in 2020 has caused heavy damage to the tourism industry, and the consumption and travel modes of tourists are undergoing changes. "Cloud tourism" and "cloud live broadcast" further expand the online digitization process of users. With the advent of the concept of "meta," the Internet has become a hot spot in the industry.

3.2.1. From the Internet to the Universe. The development of artificial intelligence has laid the foundation of intelligent control technology for the development of metauniverse. In addition to the rapid development of digital technology itself, the biological information technology related to digital technology also greatly promotes the emergence of the metauniverse. In recent years, a series of developments in human-computer interaction of biological and digital technology, especially the coupling development of the nervous system and electronic systems, are of great help to the development of metauniverse. It mainly includes three categories: one is the motion perception system [15]. The whole body sensors can accurately perceive human actions by digital processing. At present, they have been fully applied in the field of entertainment, especially in the field of film. The second is organ perception and feedback system, such as

(a) (b) (c) (d)

Figure 1: Digital holographic projection.

(a) (b)

Figure 2: VR experience.

artificial cochlea, electronic skin, and tactile gloves. Artificial cochlea has been widely used in medicine. In essence, it is an artificial organ that directly acts on the nervous system. Electronic skin and tactile gloves can more accurately measure and perceive fine hand movements and imitate various feelings to feed back to the skin. At present, it is applied in some very high-end manufacturing and medical fields [16, 17]. The third is brain wave control and brain interface: one is not deep into the skin and uses brain waves emitted by the brain to control digital devices; the other is to directly connect the nerve to the electrode to form a biological electronic nervous system. The former has been applied in many fields. For example, there are civilian products for UAV control, while the latter still has ethical and technical difficulties.

In addition, the expansion of a series of digital technologies in all aspects of economy and society, including blockchain, e-commerce, online games, digital economy, and smart cities, has made the corresponding technical preparations for the emergence of the metauniverse, which

makes the birth of the met-universe a natural and desirable product [18].

The connotation of metauniverse is to absorb the achievements of information revolution (5g/6g), Internet Revolution (Web3.0), artificial intelligence revolution, and virtual reality technology revolution including VR, AR, and MR, especially game engine, and show mankind the possibility of building a holographic digital world parallel to the traditional physical world.

The technology group supporting the metauniverse includes five parts: (1) network and computing technology; (2) artificial intelligence; (3) video game technology; (4) display technology: VR, AR, and MR; (5) blockchain technology.

3.2.2. Elements of the Yuan Universe

(1) Highly Immersive 3D Visual Experience. As a high degree of simulation of the real world and the formation of individual immersive experience, the first thing to form is the natural substitution of the visual system, which requires a higher degree of accuracy of the 3D visual system.

(2) Interactive Multisensory Simulation System. Simulating vision alone is far from forming a metauniverse. Human exploration in the field of nature depends not only on vision but also on hearing, smell, taste, touch, and other sensory organs to perceive the world [19]. Therefore, the highly realistic metauniverse must form an all-round perceptual simulation, not just a visual simulation. At present, the stereo simulation of the auditory system has developed to a high level, and its algorithm is relatively simple, but for the simulation of multiple sound sources in a complex scene, a more powerful algorithm simulation is still needed. The simulation technology of taste and smell system is relatively slow to develop, mainly because the human taste and smell system is too complex to form a compound taste and smell, which needs the support of more complex chemical technology. For example, it is still difficult to form a complex and accurate flavor through a simple concentration of the proportion of chemical raw materials. There are far from enough scenes of smell and taste in the existing entertainment scenes; therefore, it is difficult to form a commercial development [20].

This paper will take the stereo simulation of the auditory system as an example.

(3) Simulation of Natural Systems. To achieve more realistic scenario simulation, metauniverse must be able to accurately simulate natural systems; otherwise, it is an upgraded online game. This requires that the metauniverse can accurately simulate the natural systems and phenomena on the earth, including pure natural systems such as weather systems, ocean system, and biological systems, as well as the simulation of physical and chemical changes caused by human operations. In the current 3D system, preliminary simulation has been roughly achieved. For example, with the change of time, there will be changes in the rise and fall of the sun, wind, frost, rain, and snow. The recent illusory engine can even make the light and shadow close to the real picture. However, these are only preliminary visual simulations. To accurately simulate the objects in the metauniverse, we need to model a large number of objects at different levels.

(4) Intelligent Scenes with Strong AI. Whether it is the entertainment purpose of metauniverse or other life and work purposes, its value is not only to provide a pure 3D immersion digital analog place but also to provide each user with greater freedom, richer event activities, and more convenient service perception. This requires a variety of highly anthropomorphic social scenes in the metauniverse. Of course, in the early stage, the metauniverse can integrate different special AI algorithms through the support of powerful computing power to form the early general intelligence. With the further development of the general AI model, the digital subject formed by AI can evolve by itself in the metauniverse to form a digital subject in line with the internal situational characteristics of the metauniverse.

(5) Participation and Entry of a Large Number of Social Subjects and Behaviors. The above elements are actually the technical preparation or objective elements of the metauniverse, forming the environmental basis of the metauniverse. The final formation of the metauniverse requires the entry of more abundant natural persons, to form various rich practical scenes similar to the real society. This mainly includes several categories: first, entertainment scenes. This is the original motivation of the metauniverse. The interaction between humans and AI is always a lack of realism. This feeling not only comes from the rigidity of AI but also produces digital estrangement due to the inability to integrate it with offline even after AI is highly intelligent. Therefore, the integration of more social real individuals will make the metauniverse more vivid and entertaining. The second is the working scene. Various current online communication mechanisms, such as online meetings, can be nested into the scene of the metauniverse to replace most of the offline actual communication. The third is the consumption scenario. The vast majority of today's e-commerce systems can display and interact in a more realistic form in the metauniverse. The fourth is the social scenes. Metauniverse is an immersive virtual space in which users can carry out cultural, social, and entertainment activities. Its core lies in the bearing of virtual assets and virtual identity. Also, its four technical pillars are blockchain, game, network computing power, and VR [21]. As a high degree of simulation of the real world and the formation of individual immersive experience, the first thing to form is the natural substitution of the visual system, which requires a higher degree of accuracy of the 3D visual system. The seven elements that make up the metauniverse are shown in Figure 3.

In short, the value of the metauniverse is to enable the vast majority of human activities to reproduce in the metauniverse, to form a more convenient digital twin system. This requires that the metauniverse can be supported by

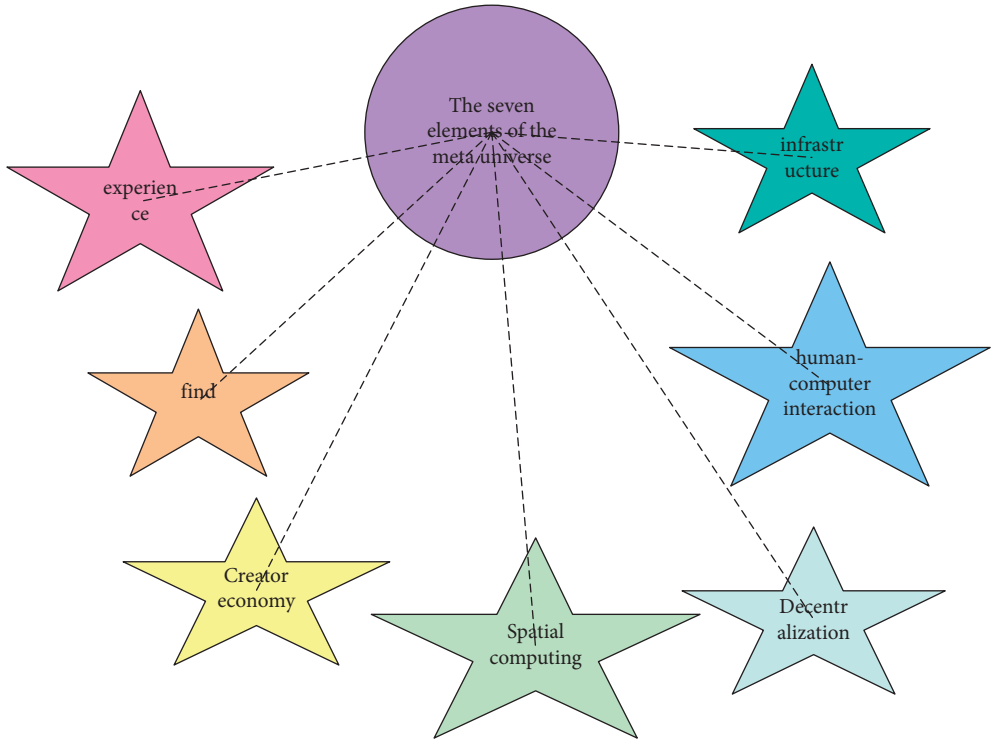

Figure 3: The seven elements of the metauniverse.

the vast majority of natural people and participate in its construction. It is not just a unilateral passive park. This means that the metauniverse will be an open digital space system with three-dimensional depth. The ultimate prosperity of this space system depends on the participation and construction of countless subjects, and metauniverse is a more infrastructure developer [22].

3.3. Fundamentals of a Metaverse Space-Based Interactive Virtual Auditory Environment Real-Time Rendering System

3.3.1. Neuronal Competition Mechanism

(1) Related Concepts of Neural Network. Artificial neural networks (abbreviated as ANNs) are also referred to as neural networks (NNs) or connection model. It is an algorithmic mathematical model that imitates the behavior characteristics of animal neural networks and carries out distributed parallel information processing. Neural network is a technology to simulate human intelligent behavior. The structure of a single neuron is shown in Figure 4.

It is similar to a nonlinear threshold device with multiple inputs and unique outputs. Define the input vector of the neuron:

$$A = [A_1, A_2, A_3, \ldots, A_n]^T. \quad (1)$$

Define weight vector ε:

$$\varepsilon = [\varepsilon_1, \varepsilon_2, \varepsilon_3, \ldots, \varepsilon_n]^T. \quad (2)$$

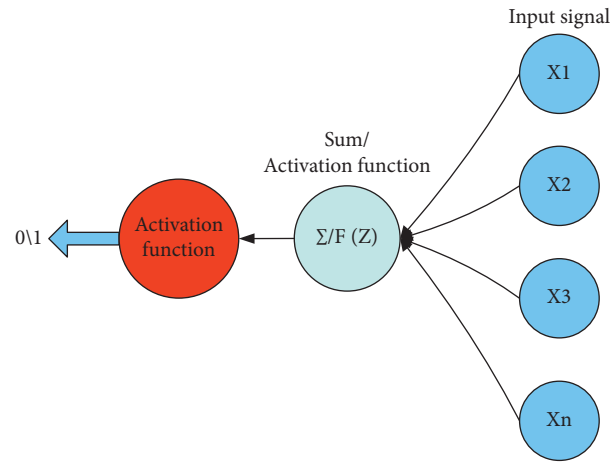

Figure 4: Neuron structure.

∂ is the threshold of neurons, and F is the activation function of neural elements. Then, the neuron output vector B is

$$B = \left(\sum_{j}^{N} A_j \varepsilon_j + \partial \right). \quad (3)$$

(2) Neural Network Type. According to the network architecture, neural networks can be divided into feedforward neural networks and recursive neural networks [23]. The neurons in the feedforward network are arranged in layers, and each neuron is only connected to the neurons in the

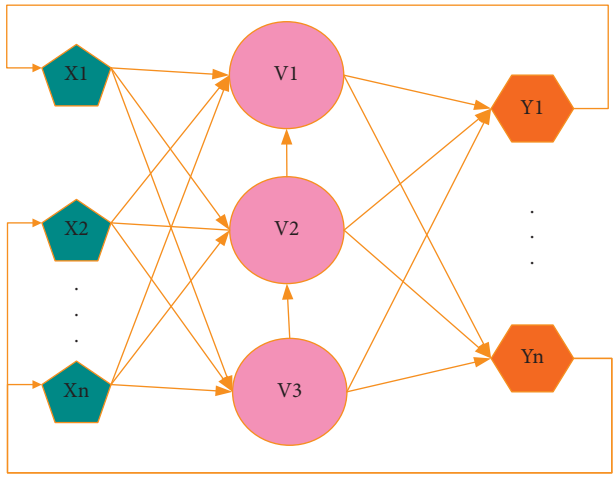

Figure 5: Schematic diagram of the forward network structure with feedback from input to output.

upper layer. Figure 5 shows an example of a feedforward neural network.

3.4. Related Technologies of Deep Neural Network. In real life, the deep neural network is widely used to extract and analyze image semantic features, which lays a solid foundation for the research of image classification technology.

3.4.1. Development of Deep Neural Network Learning. People began to study neural networks very early. The perceptron model is a basic criterion based on modern neural networks. It is a network model that can realize the function of classification and recognition through training [24]. It is a real neural network.

3.4.2. Deep Neural Network Model. Deep learning technology originated from neuroscience, also known as deep neural network. Because the current deep neural network mainly uses convolution structure, deep neural network is sometimes called deep convolution neural network. Machine learning is a method to realize artificial intelligence. Machine learning is the process of analyzing and training a large amount of data by using algorithms. Deep learning originated from a perceptron-based artificial neural network. The first-generation neural network perceptron model is shown in Figure 6.

(1) RBM Neural Network. Further research on the RBM model is one of the core contents of deep learning, which is of great significance. The RBM energy model is shown in Figure 6.

RBM is an undirected graph probability model, which is based on energy. We define the joint probability distribution by combining the energy function of the input layer vector x and the hidden layer vector h as follows:

$$p(x,h) = \frac{e^{-\text{energy}(x,h)}}{z}, \tag{4}$$

where the normalization constant $Z = \sum_{x,h} e^{-\text{energy}(x,h)}$. The marginal probability distribution of the observable input data X is

$$p(x) = \sum_h p(x,h) = \sum_h \frac{e^{-\text{energy}(x,h)}}{z}. \tag{5}$$

Introduce free energy to change equation (5) into

$$p(x) = \frac{e^{-\text{freeEnergy}(x)}}{z}. \tag{6}$$

$Z = \sum_x e^{-\text{freeEnergy}(x)}$ in equation (6), i.e.,

$$\text{freeEnergy}(x) = -\log \sum_h e^{-\text{energy}(x,h)}. \tag{7}$$

θ represents the parameters of the model, which can be obtained by taking the logarithm and derivation of equation (6):

$$\begin{aligned} \frac{\partial \log\ p(x)}{\partial \theta} &= -\frac{\partial \text{freeEnergy}(x)}{\partial \theta} \\ &\quad + \frac{1}{z} \sum_{\tilde{x}} e^{-\text{freeEnergy}(\tilde{x})} \frac{\partial \text{freeEnergy}(\tilde{x})}{\partial \theta} \\ &= -\frac{\partial \text{freeEnergy}(x)}{\partial \theta} + \sum_{\tilde{x}} p(\tilde{x}) \frac{\partial \text{freeEnergy}(\tilde{x})}{\partial \theta}. \end{aligned} \tag{8}$$

To deal with the difficult calculation of RBM partition function, the approximate value of log likelihood gradient $\partial \log\ p(x)/\partial\theta$ is usually used for training. The model parameter update rule is defined by the free energy gradient of the sample $x \sim p\ (x)$ subject to the data distribution and the sample $\tilde{x} \sim p(\tilde{x})$ subject to the model distribution as follows:

$$\begin{aligned} E_{\hat{p}}\left[\frac{\partial \log\ p(x)}{\partial \theta}\right] &= -E_{\hat{p}}\left[\frac{\partial \text{freeEnergy}(x)}{\partial \theta}\right] \\ &\quad + E_p\left[\frac{\partial \text{freeEnergy}(\tilde{x})}{\partial \theta}\right], \end{aligned} \tag{9}$$

where p is the model probability distribution, $E_{\hat{p}}$ and E_P are the expected values under the corresponding probability distribution, and $\hat{p}$ is the empirical probability distribution of the training dataset. The first term of equation (9) is relatively simple, which is generally replaced by the expectation of training samples; the second item contains the samples obtained from model P. Generally, samples are sampled by Monte Carlo Markov chain (MCMC) algorithm.

(2) Self-Coding Network. Automatic coder is an unsupervised learning algorithm, which adopts a backpropagation algorithm and is mainly used for high-dimensional complex data processing or feature extraction. The self-coding network is a special feedforward neural network, which is mainly used in dimensionality reduction, nonlinear feature extraction, expression learning, and other tasks [25]. The structure is shown in Figure 7.

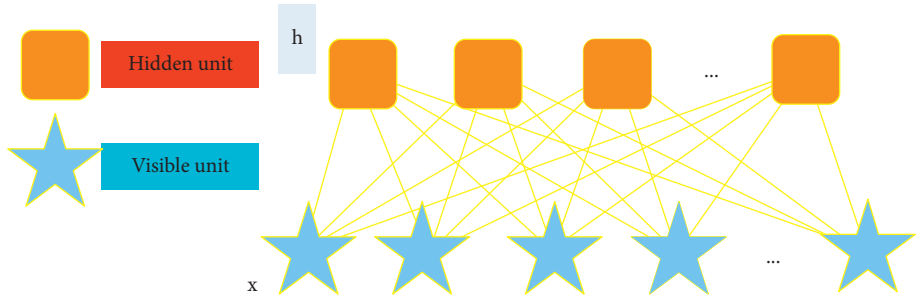

Figure 6: RBM energy model.

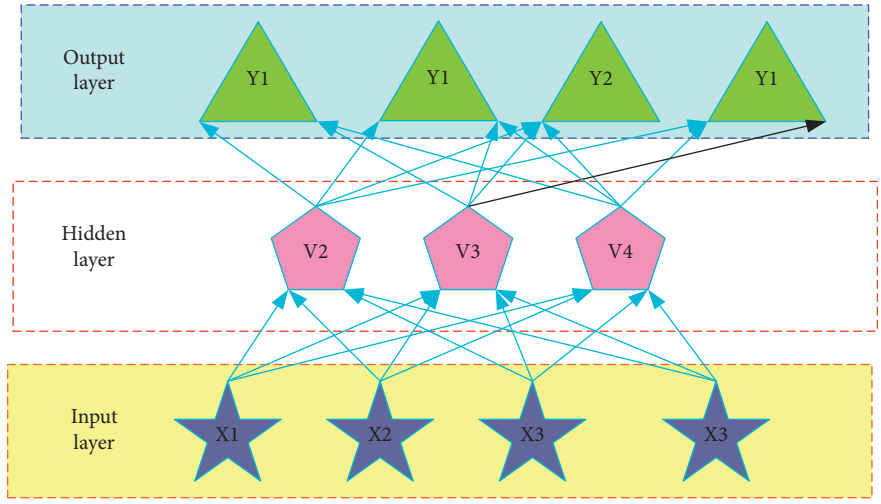

Figure 7: Self-coding network structure with only one hidden layer.

The automatic encoder is mainly composed of an encoder and a reconstructed decoder. The encoder can be represented by function $D = F(x)$, and the decoder can be represented by function $S = J(D)$. The main function of the encoder is to extract features from the input data, while the function of the decoder is to restore the features extracted by the encoder to the original information. The relevant formula is as follows:

$$\begin{aligned} D_i &= F\left(Qx_i + B\right), \\ \hat{x}_i &= F\left(Q'x + A\right), \end{aligned} \tag{10}$$

where F represents the nonlinear activation function, which is generally a sigmoid function, and $Q = \sum_{i=1}^{N} Q_{ij}x_i$ and $Q' = \sum_{i=1}^{M} Q_{ij}x_i$ are the weight matrix of the encoder and decoder, respectively. B and A are the deviations of hidden layer D and input layer x, respectively. Because of the limited ability of single-layer automatic encoders to extract signal features, SVM, softmax, and other classifiers are generally added to the automatic encoder.

(3) DBN Model Structure and Training. Deep belief network (DBN) is a typical deep learning network model, which is mainly stacked by a restricted Boltzmann machine (RBM) or automatic encoder. The RBM structure and its learning algorithm have been introduced in the previous two sections. In this section, multiple RBMs are stacked to form a deep belief network, and the DBN with three-tier structure is selected as the speech feature of this chapter. Through layer-by-layer greedy and training, a deep belief network with three-layer structure can be obtained. The model structure is shown in Figure 8. In this paper, the output eigenvector of DBN is taken as the input eigenvector of the developmental network model, and the speech data in this paper are one-dimensional [26].

As can be seen from Figure 8, the DBN is formed by the superposition of multiple RBMs, with one visible layer and the others as a hidden layer or output layer. The layer nodes are connected to each other, and there is no connection between the nodes in the layer. The training and learning of DBN network mainly have two processes: training process and fine-tuning process. The training of DBN is carried out layer by layer, that is, the output of the previous layer is used as the input of the next layer.

(4) DN Learning Algorithm. Developmental network (DN) is a new intelligent neural network proposed for simulating the autonomous mental development of the human brain. The DN model is evolved from the LCA network, and the learning algorithm of DN is based on the LCA algorithm. The leaf component analysis method can be understood as a biology-based *in situ* learning algorithm, which is possible.

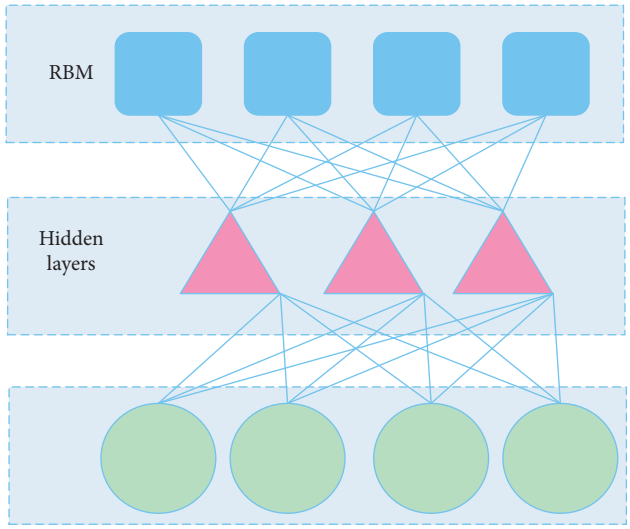

Figure 8: DBN model structure diagram.

The meaning of the concept of "leaf" can be understood through two simple concepts: leaf region and leaf component. The leaf region refers to several components of the mapping space, and each part represents a leaf region. The leaf component is discussed for each leaf region, which means that in a leaf region, a vector is used to approximate the surface of the leaf region, and the vector used as the representation is the leaf component of the region. From the concept of leaf region and leaf component, we can see that the leaf is related to a certain region. Region refers to a certain region, and the component is the component of the region. It is conceivable that leaves represent regions with certain characteristics. For several regions divided in the mapping space, the selection principle of the region optimal component is to make all or more samples in the region fall on this vector or distributed on both sides of this vector as much as possible.

3.5. Establishment of an Interactive Multisensor Simulation System for Metauniverse Based on Convolutional Neural Network. Convolutional neural network is essentially an end-to-end mapping from one end to the other, and the content of the network model is this end-to-end mapping rule. There is no explicit mathematical expression for this rule. It comes from learning by recognizing a large number of training data. In the process of training, all weights need to be initialized. Generally, small random numbers are used to initialize the network. The purpose of doing so is to avoid reaching saturation in advance due to the influence of some weight abnormal items in the process of training because too large weight items will lead to training failure. Second, because of the random nature of random numbers, different values will be generated, which can ensure the normal training and learning of the model. The training process is described below.

The training process includes four steps and can be divided into two main stages:

(1) The first stage is the forward communication stage.

(1) Take a sample (A, B_v) from the dataset and input it into the network.

(2) Calculate its corresponding actual output U_v.

At this stage, the sample information is transformed from the input layer to the output layer. The expression of the execution process of the network is

$$U_v = g_n\left(\ldots\left(g_2 g_1\left(A_v T^1\right)T^2\right)T^n\ldots\right). \tag{11}$$

(2) The second stage is the backpropagation stage.

(1) The difference between the actual output value U_v and the ideal output value B_v is taken as the error value.

(2) Adjust the weight coefficient by minimizing the error.

The above two stages of operation have accuracy requirements. Here, E_v is defined as the error size of the vth training sample in the model. The overall error of the whole network model is defined as E, and the mathematical description is as follows:

$$\begin{aligned} E_v &= \frac{1}{2}\sum_{j=1}^{n}\left(B_{vj} - U_{vj}\right)^2, \\ E &= \sum E_v. \end{aligned} \tag{12}$$

As can be seen from the above description, the input sample data first carry out forward propagation to calculate the error, then carry out backpropagation to transfer the error layer by layer, and then adjust and update the weight. To fully describe the training process, the number of input layer, hidden layer, and output layer are o, P, and Q, respectively.

Set separately

$$A = (a_1, a_2, \ldots, a_o), \tag{13}$$

which is the input vector added to the network,

$$R = \left(r_1, r_2, \ldots, r_p\right), \tag{14}$$

which is the middle-layer output vector,

$$B = \left(b_1, b_2, \ldots, b_q\right), \tag{15}$$

which is the actual output vector of the network, and

$$C = \left(c_1, c_2, \ldots, c_q\right), \tag{16}$$

which represents the target output vector of each module of training data. The weight from output unit I to intermediate unit J is set as W_{ij}, and the weight from intermediate unit J to output unit m is set as T_{jm}. The thresholds of the output unit and the intermediate unit are represented by ε_m and φ_j.

Thus, the output of the middle layer is

$$S_j = g\left(\sum_{i=0}^{o} W_{ij}A_i + \varphi_j\right). \tag{17}$$

The output of the output layer is

$$B_m = g\sum_{j=0}^{p} T_n R_j + \varepsilon_m, \tag{18}$$

where $g(\bullet)$ is the excitation function, and the sigmoid function adopted above is expressed as follows:

$$g(A) = \frac{1}{1 + E^{-ma}}. \tag{19}$$

4. Experiment and Analysis of Auditory System in an Interactive Multisensory Simulation System Based on Metauniverse

In this paper, the auditory system in the interactive multisensory simulation system based on the metauniverse is experimented. Combined with the developmental network, the developmental network auditory model of the human auditory system is roughly constructed. The auditory characteristics of the auditory periphery (cochlea) simulated by mel-frequency cepstrum coefficient (MFCC) are used as the perceptual end of the model. A variety of bionic mechanisms are used in the model, such as designing the connection mode of neurons, learning state and release effect, and the regeneration mechanism of neurons. For the verification of the performance of the model, record the speech sample database, including English words and phrases, and experiment the speech content information recognized by the model by means of speech recognition.

4.1. Experiment and Development Network Analysis. To test the influence of DBN features on the performance of the developmental network model, the parameters of DBN feature extractor are reasonably set and the matching threshold of the developmental network model is set to 0.95 to make the model in the optimal state. On the basis of MFCC features, the DBN feature extractor is used to extract depth features (mixed features) for the recognition experiment of English words and phrases in the developmental network model. In the experiment, MFCC features, DBN features, and the combined features (mixed features, i.e., MFCC + DBN features) are compared in the recognition performance of the developmental network model. The DBN feature in the experiment refers to the feature obtained by the DBN feature extractor after the simple preprocessing described in the previous chapter. The specific experimental results are shown in Tables 1 and 2.

From the experimental results in Tables 1 and 2, it can be seen that the new features extracted by the DBN feature extractor based on MFCC features have significantly improved the recognition performance of English words and phrases compared with MFCC features. Among them, for the correct recognition rate of English words, the DN-1 model is 1.66% higher than MFCC feature on the basis of mixed features and the DN-2 model is 1.86% higher than MFCC feature. There is little difference in the performance change rate between the two. In terms of phrase accuracy, the DN-1 model improves 2.59% and the DN-2 model improves 2.77% compared with MFCC feature on the basis of mixed features. When only DBN features are used, the performance improvement rate of the developmental network model is small. The reason for this phenomenon is that, on the basis of MFCC features, the multilayer structure of deep belief network is conducive to the decomposition and reconstruction of speech features and can improve the correlation of speech features. The correlation of speech conceptual features is improved, and the performance of the developmental network model will naturally be improved. Based on the analysis of the above results, the performance comparison results of the two types of developmental network models with mixed features can be obtained, as shown in Table 3.

Table 1: Correct rate of English word recognition under different features of the developmental network model based on metauniverse context (%).

Model type	DN-1 model	DN-2 model
MFCC	92.11	95.32
DBN	93.01	96.02
MFCC + DBN	94.55	97.52

Table 2: Correct rate of phrase recognition based on the developmental network model in metauniverse context under different features (%).

Model type	DN-1 model	DN-2 model
MFCC	85.44	91.88
DBN	86.88	92.54
MFCC + DBN	88.06	93.99

Table 3 shows the comparison results of the recognition performance of DN-1 and DN-2 models for English words and phrases based on the mixed features. In this paper, the mixed features refer to the new features extracted by the DBN feature extractor based on MFCC features. It can be seen that the deep belief network is introduced into the model x area to deeply extract the input information features. Whether it is the DN-1 model or DN-2 model, compared with the MFCC features in Section 4, the performance of the development network model has been greatly improved, which shows that it is feasible to introduce the deep belief network as the feature extractor.

4.2. Performance Comparison of Various Speech Recognition Models. To reflect the performance of the model in speech recognition, the recognition rate of this model and other models under the same experimental conditions is compared through experiments. The same experimental conditions refer to the use of the self-recorded speech database, the same training sample set and test sample set, and the same new speech feature extraction method. Other models are dynamic time warping (DTW), hidden Markov model (HMM), improved backpropagation neural network (BP), and convolutional neural network (CNN). DTW matches by calculating the minimum cumulative distance between the two vectors, which requires the training sample set as the

Table 3: Comparison of DN-1 and DN-2 model performance in mixed characteristics (%).

Identify object	English words	Phrase
DN-1 model	92.66	89.10
DN-2 model	95.58	93.01

template. The HMM model sets six states, and each state has three mixed Gaussian probability density functions. In the BP neural network, the selection of the number of hidden layer neurons is essential. In the experiment, the number of hidden layer neurons is 35 and the structure is 3 layers. The selection principle of DN-1 and DN-2 model parameters is to optimize the recognition performance of the model. CNN selects the maximum pool size of 3, and the step size of the pool window is set to 3. Experiments are carried out on the above speech recognition models and repeated 10 times to obtain the steady-state value or average value as the recognition result, as shown in Figure 9.

It can be seen intuitively from Figure 9 that, for simple English words and phrases in low-noise environment, compared with the traditional speech recognition model, the developmental network model constructed in this paper generally has high recognition rate, strong antinoise ability, and small difference in recognition rate between English words and phrases. Compared with CNN in neural network, the recognition rate of the model constructed in this paper is slightly lower than that of CNN. The reason for this phenomenon does not rule out the influence of model structure parameters.

5. Discussion

As an important historical node of human civilization, the Metaverse Space requires the high attention and vigilance of the whole society at the beginning of its birth. The Metaverse Space itself has three natural drawbacks, namely, knowingness, certainty, and nonpractice. Therefore, the governance of the Metaverse Space must adhere to the pregovernance orientation of embedding the Metaverse Space in the real society and forming a symbiotic civilization. This requires the common vigilance and awareness of all human beings, from the elite to the public, as well as the restriction and guidance of national policies and regulations. The idea that the Metaverse Space will form a completely self-circulating and unregulated virtual universe is very dangerous and wrong. The return of human civilization must be the extension of civilization based on the real world, rather than the self-limitation of the digital world.

Virtual reality technology uses computers to virtualize real objects to construct a simulated world. With the continuous extension of the network, corresponding digital technologies are also advancing simultaneously, such as big data and artificial intelligence technology. The full penetration of the network in society is bound to form a continuous digital mapping of the real world because only through digital collection and construction can interactive transmission be carried out on the network. It can be said that networking is the main thread of the digital process.

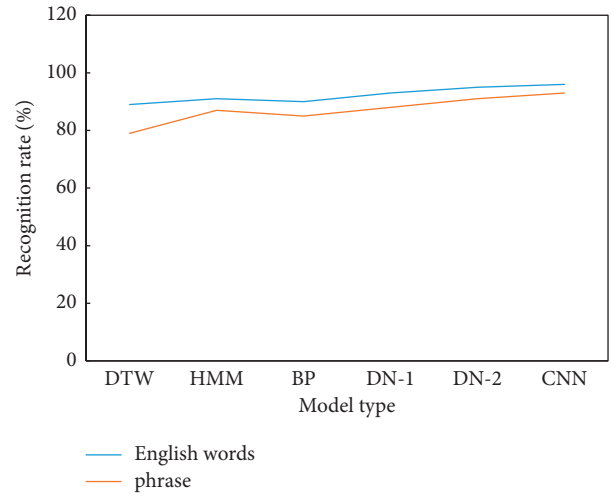

Figure 9: Comparison results of recognition rates of different speech recognition models.

Enhancing visual authenticity and vividness enables people to feel a stronger visual experience and psychological experience. In strengthening the information conveyed, we show design pays more attention to the relationship between virtual objects, the relationship between reality and objects in the virtual world, and the study of the way information is conveyed, rather than focusing on the relationship between real objects and people.

6. Conclusion

At present, the research of machine vision has not become a perfect subject, which has great significance compared with the research of human hearing. Human auditory system is an important channel of information second only to the visual system. It has superior performance in listening, sound and object discrimination, which is unmatched by computers. Aiming at the channel structure of the human auditory system for speech information processing, combined with the basic theory of developmental networks, this paper roughly simulates the human auditory system and establishes an artificial auditory model. Before the establishment of the model, this paper briefly introduces the structure and information processing pathway of the human auditory system, expounds the basic theory of developmental networks in detail, and finally explores the establishment of the model and the performance of the model.

The emergence and development of digital media have promoted the expansion of information communication from one-way communication to interactive two-way communication. Therefore, information communication design should not only pay attention to the link of information expression but also pay attention to the interactive process of information communication. In this way, as a part of the visual expression of information, visual communication design should not only express information through design forms but also consider the experience and emotion of information recipients or audiences to achieve the effectiveness of information transmission. The media

represented by the network has been popularized and applied, and all kinds of multimedia technology, interactive technology, and virtual reality technology have begun to enter people's lives, which provides a broader platform and unpredictable possibilities for visual design. Digital design software also provides more convenience for design. However, no matter how many technical forms are, they are just the carriers of ideas, and the blind worship of technology is not what we pursue. Therefore, designers need to travel with things not tired of technology, open up design thinking, pay attention to the thinking of the essence of design, and make meaningful design to promote the progress of human thought.

Data Availability

The data underlying the results presented in the study are available within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was financially supported by the Science Foundation of Zhejiang Vocational Academy of Art (QNTD2020006).

References

[1] J. Zhang, Y. Zhou, K. Xia, Y. Jiang, and Y. Liu, "A novel automatic image segmentation method for Chinese literati paintings using multi-view fuzzy clustering technology," *Multimedia Systems*, vol. 26, no. 1, pp. 37–51, 2020.

[2] E. R. Hassan, M. Tahoun, and G. S. ElTaweel, "A robust computational DRM framework for protecting multimedia contents using AES and ECC," *Alexandria Engineering Journal*, vol. 59, no. 3, pp. 1275–1286, 2020.

[3] Y. Luo, "Evolved multimedia broadcast Multicast service in LTE along with cognitive radio on TV bands," *International Journal of Electronics Engineering Research*, vol. 11, no. 2, pp. 171–182, 2019.

[4] Y. Dongmei, "Fusion path of radio TV and digital network technology based on multimedia editing platform and paper media innovation," *Paper Asia*, vol. 2, no. 2, pp. 172–175, 2019.

[5] L. Meng, X. Zhao, and Y. Sun, "An advanced investigation and analysis of teaching innovation in universities based on digital multimedia technology," *Boletin Tecnico/Technical Bulletin*, vol. 55, no. 18, pp. 458–462, 2017.

[6] X. Li, "Optimization and development of city landscape design under the influence of digital multimedia technology," *Boletin Tecnico/Technical Bulletin*, vol. 55, no. 18, pp. 108–113, 2017.

[7] B. Wang and A. Wang, "An optimization model innovation for college student management based on multimedia network platform," *Boletin Tecnico/Technical Bulletin*, vol. 55, no. 8, pp. 195–202, 2017.

[8] X. Tang, J. Bian, and S. Liu, "Application of multimedia digital technology in traditional residence architectural analysis and protection," *Revista de la Facultad de Ingenieria*, vol. 32, no. 4, pp. 796–803, 2017.

[9] C. Inquimbert, P. Tramini, O. Romieu, and N. Giraudeau, "Pedagogical evaluation of digital technology to enhance dental student learning," *European Journal of Dermatology*, vol. 13, no. 01, pp. 053–057, 2019.

[10] E. Bonacini, "A survey on the digital enhancement of the archaeological sites on Google and a multimedia pilot project in the Agrigento Valley of the Temples in Sicily (Italy)," *International Journal of Internet Technology and Secured Transactions*, vol. 7, no. 1, p. 28, 2017.

[11] H. Yu, "Application analysis of new Internet multimedia technology in optimizing the ideological and political education system of college students," *Wireless Communications and Mobile Computing*, vol. 2021, no. 4, Article ID 5557343, pp. 1–12, 2021.

[12] E. Borba, A. G. Corrêa, R. D. D. Lopes, and M. Zuffo, "Usability in virtual reality: evaluating user experience with interactive archaeometry tools in digital simulations," *Multimedia Tools and Applications*, vol. 79, no. 5-6, pp. 3425–3447, 2020.

[13] Y. Zhang, X. Zhang, T. Zhang, and B. Yin, "Crowd motion editing based on mesh deformation," *International Journal of Data Mining and Bioinformatics*, vol. 2020, no. 3,Article ID 3634054, pp. 1–13, 2020.

[14] S. Bratt and L. Hodgins, "Towards the design of a digital fluency course - an exploratory study," *Journal of Educational Multimedia and Hypermedia*, vol. 28, no. 1, pp. 21–38, 2019.

[15] Y. Li, "Research and application of the teaching mode with the integration of multimedia technology and teaching management," *Agro Food Industry Hi-Tech*, vol. 28, no. 1, pp. 2764–2768, 2017.

[16] Q. Wang and X. Ji, "Research on the 3d animation design and model simulation optimization based on multimedia technology," *Boletin Tecnico/Technical Bulletin*, vol. 55, no. 6, pp. 541–547, 2017.

[17] K. Khamadi and A. Senoprabowo, "Adaptasi permainan tradisional mul-mulan ke dalam perancangan game design document," *ANDHARUPA: Jurnal Desain Komunikasi Visual & Multimedia*, vol. 4, no. 01, pp. 100–118, 2018.

[18] C. Yufu, "Application and value analysis optimization of multimedia virtual reality technology in urban gardens landscape design," *Boletin Tecnico/Technical Bulletin*, vol. 55, no. 15, pp. 219–226, 2017.

[19] A. Saad, K. Robin, S. B. O. Khan, and N. H. Ubaidullah, "The evaluation of user acceptance of an iban digital story telling (IDST) application among iban language teachers," *The International Journal of Multimedia & Its Applications*, vol. 9, no. 4/5/6, pp. 01–14, 2017.

[20] N. Alherbawi, Z. Shukur, and R. Sulaiman, "JPEG image classification in digital forensic via DCT coefficient analysis," *Multimedia Tools and Applications*, vol. 77, no. 10, pp. 12805–12835, 2018.

[21] X. Gong, W. Li, and W. Jing, "Blind extraction of digital watermarking algorithm for color images," *The International Journal of Multimedia & Its Applications*, vol. 13, no. 2, pp. 15–26, 2021.

[22] A. O. Isah, J. K. Alhassan, S. S. Olanrewaju, and E. F. Aminu, "Enhancing AES with time-bound and feedback artificial agent algorithms for security and tracking of multimedia data on transition," *International Journal of Cyber-Security and Digital Forensics*, vol. 6, no. 4, pp. 162–178, 2017.

[23] S. Zhang, L. Zhou, M. Lu, and Y. Wang, "Design and simulation of a target scene generator with a telecentric structure

in the image space," *Applied Optics*, vol. 58, no. 9, p. 2394, 2019.

[24] J. K. H. Martin, A. M. Stromberg, M. Chen, and M. I. Mizuko, "Comparing embedded and non-embedded visual scene displays for one adult diagnosed with autism spectrum disorder: a clinical application of single case design," *Child Language Teaching and Therapy*, vol. 36, no. 1, pp. 3–18, 2020.

[25] Y. Xiaozhou, B. Fan, and P. Jones, "Recognition method of outdoor design scene based on support vector machine and feature fusion," *Journal of Intelligent and Fuzzy Systems*, vol. 39, no. 6, pp. 8757–8766, 2020.

[26] F. Aliyu and C. A. Talib, "Virtual reality technology," *Asia Proceedings of Social Sciences*, vol. 4, no. 3, pp. 66–68, 2019.