

Research Article

Jiayue Yan*

The reform of the teaching mode of aesthetic education for university students based on digital media technology

<https://doi.org/10.1515/comp-2025-0026>

received October 12, 2023; accepted February 11, 2025

Abstract: In response to the current problems in teaching aesthetic education, such as a single teaching method and lack of innovation, the study uses digital media technology to improve it. First, Maya software is used to construct a three-dimensional (3D) model of ancient architecture, and then the UNity3D engine is used to create a roaming scene for students to experience the heritage and elegance of ancient architecture in the aesthetic education classroom. To enable students to achieve an immersive teaching experience, the study uses Kinect somatosensory technology to enable students to interact with the simulated scenes. The students were first tracked by a combination of random forest algorithm and mean shift algorithm, and then the dynamic time warping algorithm was used for dynamic gesture recognition to ensure that students could move forward and backward through the gestures while using the simulation to navigate through the building. Through the above operations, the study completes the design of a virtual recreation system for ancient cultural buildings based on digital media technology. Through the experiment, it was obtained that 94.81% of the total number of students surveyed believed that this method could stimulate interest in learning about aesthetic education classes. The system designed by the research provides new ideas for the reform of the teaching mode of aesthetic education in universities.

Keywords: aesthetic education teaching, paradigm reform, digital media technology, dynamic time regularization algorithm

1 Introduction

The status and role of aesthetic education in the cultivation of talents in colleges and universities is becoming more and more prominent. Aesthetic education is education of sentiment, education of the mind, as well as education that enriches the imagination and cultivates the sense of innovation [1]. At the present stage, aesthetic education is mainly a fusion and innovation of art and aesthetic teaching experience. However, this purely skill-based education has greatly reduced the space for aesthetic education in time and has “flattened” the richness of aesthetic education [2–4]. Digital media is a product of the integration of computer technology, network technology, digital communication technology, and culture, art, and business. It is an integrated processing of text, audio, graphics, images, and other media information, using network communication and computer technology as the main means. It is a software and hardware technology that realizes the representation, recording, processing, storage, transmission, display, and management of images, making abstract information perceptible, manageable, and interactive [5]. The advantages of introducing digital media into higher education are mainly reflected in the following aspects:

1. Improve students' learning interest and participation. By providing more vivid, vivid, and interesting learning content, stimulate students' learning interest and enthusiasm through visual, auditory, and other sensory stimulation, and improve their participation.
2. Adapt to the individual needs of students, so that students can get a more comprehensive and in-depth learning experience and understanding.
3. Promoting students' active and independent learning can make students more actively participate in the learning process and promote students' active learning and independent learning through interaction and experience.
4. Expand students' learning forms and ways. Digital media technology can integrate learning into all aspects of students' daily life through the Internet, multimedia, virtual

* **Corresponding author: Jiayue Yan**, Engineering and Technology College, Hubei University of Technology, Wuhan, 430068, P. R. China, e-mail: yjy012023zn@163.com

reality (VR), and other ways, making learning more interesting and effective.

5. Digital media technology can provide more flexible and diversified learning methods and improve students' innovative ability and comprehensive quality through innovative practice and exploration. The use of digital media technology makes the aesthetic education of college students more interesting and participatory, including multimedia, interactive and personalized, providing personalized educational content and ways, so that every student can get a more comprehensive and in-depth experience and understanding in the process of receiving aesthetic education. In addition, there is situational, digital media technology can create a more real and vivid aesthetic education scene through VR and other technologies. Through innovative practice and exploration, digital media can help to improve innovative ability and comprehensive literacy.

To this end, the study uses digital media technology to model ancient cultural buildings in three dimensions and construct virtual roaming scenes. This will enable students to experience the heritage and elegance of ancient cultural buildings more deeply in the classroom and enhance their aesthetic skills as well as their morale.

This teaching method breaks through the originally simplified teaching mode, providing students with a richer learning experience and effectively enhancing their ability to appreciate beauty. Aesthetics education can promote the all-around development of college students' intelligence, cultivate their positive and healthy aesthetics, promote the formation and development of students' mental health, and cultivate their creativity and imagination. Promote the comprehensive development of people through beautiful education and provide strong talent support for society. The innovation points and main problems solved in this study are as follows:

1. Breaking traditional teaching models and achieving innovation in aesthetic education teaching. The education of architectural aesthetics has been achieved through three-dimensional (3D) modelling and virtual interaction.
2. Changes in lighting conditions or skin-like backgrounds in the environment can make gesture extraction difficult. Research has solved this difficulty by using median filtering to denoise depth images.
3. Dynamic gesture recognition is difficult to recognize in real time because of the excessive speed of human hand movement or wrist rotation. The research combines the random forest algorithm and mean shift algorithm to track human bones, accurately locate human joints, and obtain

the 3D coordinates of joint points. Then, dynamic time-warping algorithm is used to recognize dynamic gestures.

4. The research combines static and dynamic gesture recognition to achieve high precision gesture recognition, bring a more immersive roaming experience to students, and provide a high-quality aesthetic appreciation environment.
5. Due to the differences in height, height, obesity, and thinness of each individual, their movement trajectories and timing may vary when performing the same gesture behavior. Therefore, research and explore dynamic gesture recognition based on bone tracking technology to ensure high precision recognition and tracking.

2 Related works

With the full spread of 5G network technology and the rapid development of science, digital media technology is being used in an increasing number of scenarios. Mills and Brown designed a 3D virtual drawing program by investigating the creativity of users. Using VR head-mounted displays and sensor analysis focused on how students delivered the same story in written, oral, and virtual drawing modes. The lack of reciprocity between the three modes is addressed by creating stories across modes [6]. Jiang et al. analysed the characteristics of digital media and applied digital media technology using bi-directional technological innovation, integrating it into 3D animation design and virtual simulation to promote technological upgrading in animation creation [7]. Liu found that traditional advertising models could not adapt to the needs as the times evolved. For this reason, its analysis investigated the application of digital media technology in advertising art [8]. Lu studied the impact of digital media technology on the design of forest scene animation and built a framework for managing the visualization of virtual forest scenes using scene map technology [9]. Zhu based on the present situation of the digital media art design education analyzes the influence of qing digital media technology in art design education and the existing problems, from the angle of creative education, and puts forward some countermeasures and suggestions [10]. Xu and Guan analysed the differences in academic performance between students who studied in the "four-in-one" blended learning mode (official website, WeChat public platform, official microblog, and cloud database) and those who studied through traditional learning methods [11]. Inej and Ogar examined the impact of digital media on effective healthcare delivery in Cross River Island and concluded that the exponential growth of

digital media has led to more efficient, transparent, and faster healthcare delivery in Cross River [12]. Gragorios and Herron designed a teaching video system for classroom video teaching with digital media technology in the context of media integration, combining digital media with other media according to the principle of media integration. Check the classroom video instruction and provide subtitles to help students learn in the boring instructional design [13]. Mao and Jiang explored the path of digital media technology for exploring the development of UI, visual sensing image technology, and digital media technology based on the means of visual sensing technology after achieving a close connection between digital media and art [14].

Aesthetic education is the training of one's sense of beauty and cognition, cultivating children's awareness of the pursuit of beauty as well as cultivation and artistic qualities. Li et al. shape contemporary Chinese aesthetic education from the perspective of policies and concerns, including the concept and functions of aesthetic education, a review of aesthetic education policies, the effects of aesthetic education, and the evaluation of aesthetic education and its development trends [15]. Wang et al. use aesthetics as the basis, using cases of traditional classical aesthetics to build a multidisciplinary fusion of creativity and aesthetics as a body of knowledge [16]. Shi and Cheng studied the relationship between aesthetic education in higher education and classroom teaching in primary and secondary schools to enrich the primary and secondary school classrooms and enlighten students' wisdom [17]. Wen et al. used the online aesthetic education courses offered by the Chinese "iCourse" platform and the American "edX" platform to compare the courses and their contents, identify the characteristics and problems of each, and provide new ideas for future teaching of aesthetic education [18]. Khasanova developed ethnic and intercultural competencies by providing theoretical information in the classroom, supporting student interaction and peer review to promote students' overall development in the field of music [19]. He and Luo constructed an intelligent recognition system based on the advantages of convolutional neural networks (CNN) in image processing to assess the evolutionary characteristics of extreme rainfall weather and to post-evaluate music teaching. It makes the assessment of the effectiveness of music education more efficient [20].

A synthesis of the above-mentioned literature shows that the teaching reform of aesthetics courses in higher education institutions is gradually being emphasized. However, not many studies have applied digital media technology to the teaching reform of aesthetics. To this end, the study modelled famous architectural sites around the world and used VR technology to create 3D virtual

scenes of the buildings, followed by gesture recognition based on Kinect body capture technology to realize human-computer interaction. Students are made to experience the cultural heritage in each building in an immersive way.

3 Design of a virtual reproduction system for ancient cultural buildings based on digital media technology

This study introduces VR technology into aesthetic education courses in colleges and universities. Taking music appreciation class as an example, teachers use digital media technology to break through practice and space restrictions and create ancient architectural scenes such as "Dunhuang Flying Sky" and "Silk Road" by using virtual wandering scenes while playing Chinese classical music "High Mountain" and "Flowing Water" in class. In the process of learning, students can truly feel the media that music, art, film, and television want to integrate. To stimulate students' curiosity and divergent thinking. At the same time, students can also use VR painting software to create in the learning process and express the melody trend through lines. This learning mode, which combines "listening," "seeing," and "manual," creates a situational atmosphere for students to appreciate music, so that students can better understand the connotation of classical music. This kind of teaching mode changes the traditional teaching mode which mainly imparts theoretical knowledge and provides a new and innovative way to reform the teaching mode of college students' aesthetic education in an innovative way. The research uses 3D modelling technology, VR technology, somatosensory interaction technology, and artificial intelligence algorithms to realize the construction of the virtual reproduction system.

3.1 Unity3D-based virtual roaming scene creation

Digital media technology can provide more intuitive, vivid, and vivid aesthetic education content, which makes it easier for college students to accept and understand. At the same time, digital media technology can also make college students more actively participate in the process of aesthetic education through interaction and experience and improve their learning interest and enthusiasm. The

application of digital media technology can expand the form and way of college students' aesthetic education. The traditional aesthetic education mainly focuses on classroom teaching, which is monotonous in form and boring in content, and is difficult to arouse the interest of college students. Digital media technology can integrate aesthetic education into all aspects of college students' daily life through the Internet, multimedia, VR, and other ways to make aesthetic education more interesting and effective. Digital media technology can provide personalized educational content and methods according to different aesthetic needs and characteristics of college students, so that college students can get more comprehensive and in-depth experience and understanding in the process of receiving aesthetic education. The theoretical framework of the research includes the following four aspects: (1) Digital media technology: This article studies the 3D modelling of ancient cultural buildings and the construction of virtual roaming scenes by using digital media technology. Digital media technology involves many fields such as computer graphics, VR, and human-computer interaction, which provides a new way for the protection and inheritance of ancient cultural architecture. (2) VR technology: Through VR technology, students can feel the heritage and style of ancient cultural buildings more deeply in the classroom and improve students' aesthetic ability. VR technology can create an immersive interactive experience, so that students can feel the charm of ancient cultural buildings. (3) Artificial intelligence algorithm: The dynamic time warping (DTW) algorithm and grid collider algorithm used in the research are both classic algorithms in the field of artificial intelligence. The DTW algorithm is used for gesture recognition and can deal with time series matching of gesture sequences effectively. The grid collider algorithm is used to detect collision events in gesture sequences, which provides important supplementary information for gesture recognition. (4) Teaching reform concept: The research aims to explore the method of using digital media technology and VR technology to realize the reform of aesthetic education teaching methods in universities. Through innovative teaching methods, stimulate students' interest in learning and improve students' learning effect. At the same time, the method designed by the institute can also enhance students' learning interaction and cultivate students' innovative thinking and practical ability.

In university aesthetic education classes, it is essential to develop students' appreciation of the beauty of various types of architecture around the world, but simply using videos or pictures does not provide students with a more realistic and stunning experience. To this end, the study uses VR technology to create roaming scenes of architecture and uses somatosensory technology for human-

computer interaction to bring a richer teaching experience to students. The research uses geometric modelling as the main approach and other modelling methods as a supplement to create 3D models of cultural buildings. Geometric modelling techniques model data by representing, controlling, analysing, and outputting geometric entities with geometric and topological information reflecting the shape, position and presentation of the structural drawing. Maya modelling software was chosen for the study to model the interior and exterior of the building. Similar and repetitive details are often found in the external structures of buildings, for which a modular modelling approach was used to simplify the work and improve modelling efficiency. Many ancient buildings come with frescoes as well as inscriptions inside the building. In the modelling process, it is necessary to consider not only the structure of the model, but also the seamless stitching of textures such as frescoes at a later stage, and the study unfolds the UV of the model into a whole on the basis of a single-sided infill, which is able to achieve perfect stitching of the internal textures.

The 3D model alone is far from satisfying the user's need for realism in VR, and to have a better visual space, you need to add mapping textures to the model. The first element of mapping is the UV spread, also known as UV mapping, and Maya software comes with a variety of mapping methods that use a predetermined rule to project UV texture coordinates onto the surface of the model, automatically creating an association between the texture image and the surface. The results of UV mapping using different mapping methods are shown in Figure 1.

After UV mapping, the images were studied for rendering in Photoshop to complete the model, Unity3D is an emerging engine with a strong presence in the film and television industry and has great strengths in creating roaming scenes. In order to achieve a realistic roaming effect, characters must react when they encounter objects in the scene, and characters stop moving forward when they encounter buildings and are not able to walk through walls. The study uses the Unity3D engine to add colliders to the model. Mesh colliders are complex colliders that fit into the structure of the model and can achieve good interaction. However, it consumes a lot of computational resources and affects the performance of the scene, so other colliders need to be selected depending on the situation. In some cases where interaction is not required, prototype colliders can be used instead, saving resources and optimizing scene performance. The external environment of the virtual scene mainly includes the sky, mountains, and features such as flowers and trees. The study uses Unity's own sky box material to simulate the virtual sky

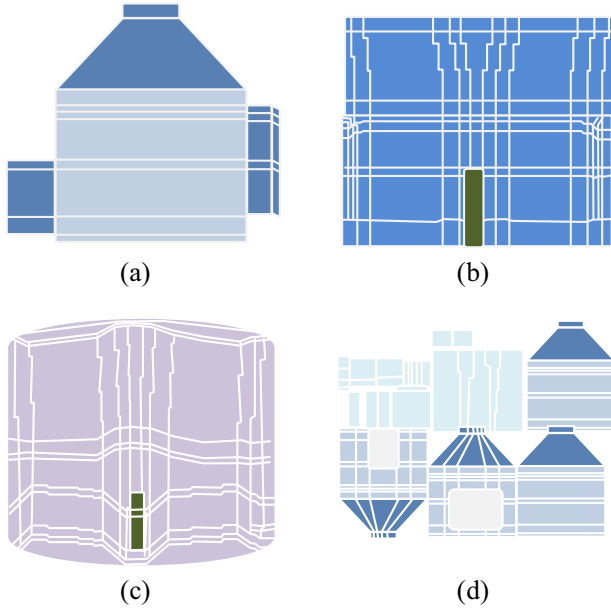


Figure 1: UV mapping with different mapping methods. (a) Planar mapping results. (b) Cylinder mapping results. (c) Spherical mapping results. (d) Automatic mapping results.

environment. The study also uses a manual approach for the external terrain of the building. Combining the above operations, the study created a roaming 3D scene of a cultural building based on Unity3D.

3.2 Kinect-based depth image pre-processing in roaming scenes

Kinect is a 3D body camera that not only acquires colour images of targets but also captures depth images of targets and has features such as human skeleton, face recognition, and voice recognition. The main methods of environmental depth measurement are triangulation, time-of-flight, and structured light measurement. The structured light measurement method is mainly based on optical coding

technology, which is at the heart of Kinect's depth image acquisition. The light source for optical encoding is a laser scatter, the main characteristic of which is that the scatter is highly random and changes pattern depending on the distance and size of the Kinect. the Kinect laser scatter imaging principle is shown in Figure 2.

Before recognizing the scattered images of a spatial object, the light source is calibrated: assuming that a reference plane is taken every 10 cm, 30 scattered images can be obtained with this calibration. When the sensor is used to take a picture of the object to be measured, the scattered image of the object to be measured is obtained, and this image is correlated with the previous 30 images to obtain 30 correlated images. The peaks on the correlation images are the position of the object to be measured, and all peaks are superimposed and differenced to obtain a 3D image of the object to be measured, i.e., a depth image. This results in the height and width of the object being imaged and the actual physical position of the object not corresponding to each other. The original depth information captured by Kinect is not the actual spatial distance of the object. Therefore, it is necessary to correct for errors in the depth data and the actual distance to the object. The formula for converting the depth value to the actual distance is based on the Kinect calibration principle

$$d = K \tan (Hd_{\text{raw}} + L) - O. \quad (1)$$

In equation (1), d_{raw} is the depth value of a pixel point on the depth image, $H = 3.5 \times 10^{-4}$ rad, $K = 12.36$ cm, $L = 1.18$ rad, $O = 3.7$ cm. The actual depth distance of the pixel is obtained from equation (1) at d . The 3D coordinates of the pixel point in real space are calculated from the depth distance in equation (2).

$$\begin{cases} x = \left(x_d - \frac{w}{2}\right)(z_d - 10)F\frac{w}{h} \\ y = \left(y_d - \frac{w}{2}\right)(z_d - 10)F\frac{w}{h} \\ z = z_d = d. \end{cases} \quad (2)$$

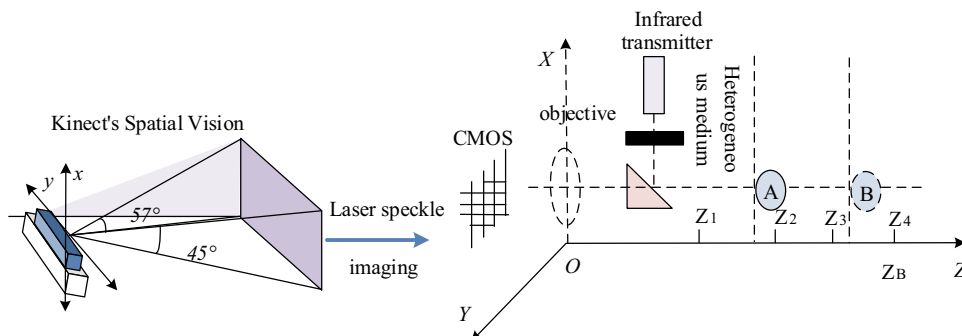


Figure 2: Kinect laser speckle imaging principle.

In equation (2), $F = 0.0021$ and $w \times h$ are the resolutions of the Kinect. The RGB images acquired by the Kinect sensor have some visual deviations from the depth images, and to correct for these deviations, the depth data is converted to the corresponding RGB data. When capturing depth images with the Kinect, the infrared component of the strong light finder will cover the laser scatter on the surface of the target to be measured, and the CMOS sensor is not sensitive to the scatter pattern in these areas, resulting in a hollow noise in the final depth image. For this reason, the study uses median filtering to filter out the noise. The median filtering principle is shown in Figure 3.

The median filter for noise reduction of the data is calculated as

$$f(i, j) = \text{median}_w \{f(r, t) | f(r, t) \in N_{f(i, j)}\}. \quad (3)$$

In equation (3), $f(r, t)$ represents the grey value of any pixel point within the domain window w , $N_{f(i, j)}$ is the solid field of grey values, and median_w is the intermediate value for finding that taken grey in the window. Combining the above operations, the study converts the depth information into real distances and calibrates and denoises the image.

3.3 Human gesture recognition during roaming

When performing dynamic recognition, it is first necessary to track the various movements of the user. The study uses Kinect to capture the depth data of the human body and extract the morphology of the body to obtain a depth image. The depth image features are then extracted from the depth image, and then pixel-by-pixel information is inferred based on the depth image features using machine learning algorithms for body part recognition. All pixel information is aggregated using clustering algorithms to form a reliable prediction of 3D skeletal joint positions as

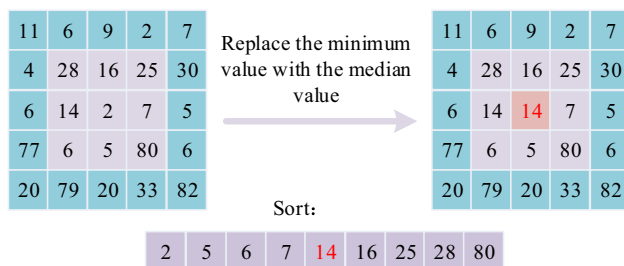


Figure 3: Schematic diagram of median filtering principle.

the final output, and then the evaluation system uses this for each pixel of the output to determine the position of the joint point. An illustration of the depth difference feature in the depth image of the human skeletal nodes is shown in Figure 4.

The study combines the random forest algorithm and the mean shift algorithm to track the human skeleton and accurately locate the human joints to obtain the 3D coordinates of the joints. To reduce the computational complexity and improve the recognition accuracy, six joints of the left hand, right hand, left wrist, right wrist, left elbow, and right elbow are selected as feature vectors. The forearm movement of the hand can be represented by the wrist joint as well as the elbow key data, and a feature vector consists of the 3D coordinates of the six joint points, which can be written as

$$F_n = \{(x_1, y_1, z_1), (x_2, y_2, z_2), \dots, (x_6, y_6, z_6)\}. \quad (4)$$

In equation (4), (x, y, z) is the coordinate of the corresponding joint point and n is the human skeletal joint data at the time of t_n . Considering that the elbow and wrist joints as well as the hand joints move excessively during the gesture movement, while the shoulder joints contribute little value to the gesture movement. Therefore, the centre of the left/right shoulder was chosen as the reference point for the study, which can be written as

$$A(x_c, y_c, z_c) = \frac{1}{2}(x_{ls} + x_{rs}, y_{ls} + y_{rs}, z_{ls} + z_{rs}). \quad (5)$$

After obtaining the reference point A by means of equation (5), the data is normalized using

$$O_{\text{norm}} = \frac{O - A}{|LS - RS|}. \quad (6)$$

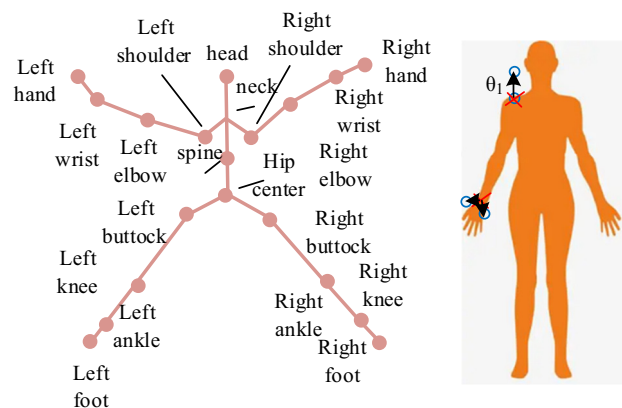


Figure 4: Schematic diagram of depth difference features in bone node depth images.

In equation (6), O is the coordinate vector corresponding to the skeletal joint point before normalization, O_{norm} is the vector after normalization, LS is the coordinate vector corresponding to the left shoulder joint point, and RS is the coordinate vector corresponding to the right joint point. The position of the shoulder joint does not change significantly during the gesture recognition process, and by using it as a reference point, the variation in feature vectors caused by different user sizes and distance sensors and thus by the gesture recognition process is eliminated. Invalid gesture frame sequences can interfere with the gesture recognition effect during dynamic recognition. The study improves the real-time performance of the dynamic gesture recognition system by detecting the gesture starting point and focus, segmenting the before and after dynamic gesture frame sequences, extracting the gesture frames with high correlation Xu Lei, and removing the invalid frame sequences.

The study uses the DTW algorithm for dynamic gesture recognition. The DTW algorithm is a typical template matching algorithm that finds an effective time-calibrated matching path between a sequence of gestures and a sequence of template gestures and solves for the regularization function corresponding to the minimum cumulative distance when the sequence is calibrated. $W(n)$. The DTW regularization path is shown in Figure 5.

The path formed by the grid points passing between the starting and ending points is the regularized path, as follows

$$W = \{w_1, w_2, \dots, w_k\}, \max(N, M) \leq K \leq M + N - 1. \quad (7)$$

In equation (7), the k -th element of W w_k represents the path of the k -th matching. The regularized paths satisfy the three constraints of continuity and monotonicity and

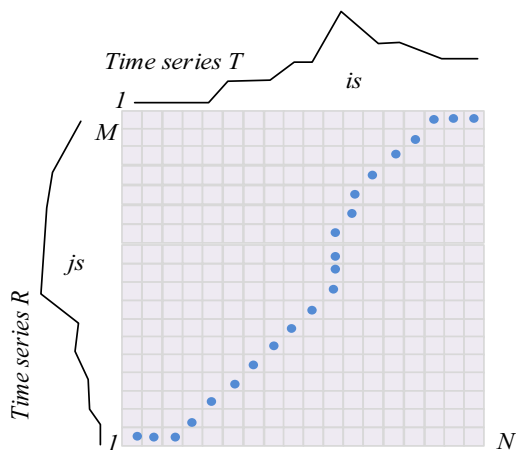


Figure 5: Tuning path of the DTW algorithm.

the same starting point and end point. To determine the optimal regularization path, the minimum cost function of the regularization path is defined as follows:

$$DTW(R, T) = \min \left\{ \frac{1}{K} \sqrt{\sum_{k=1}^K w_k} \right\}. \quad (8)$$

The distance at each point on the matching path is accumulated. When performing dynamic gesture recognition, the gesture sequence to be measured is matched with the reference gesture sequence by the DTW algorithm, and the gesture corresponding to the minimum distance value output is the final recognition result. The cumulative distance is shown as

$$\gamma(i, j) = D(R_i, T_j) + \min \{ \gamma(i-1, j-1), \gamma(i-1, j), \gamma(i, j-1) \}. \quad (9)$$

However, the DTW algorithm is an iterative computation of the data, and as the sequence of gestures increases, the complexity of the algorithm and the computational effort increases, resulting in a lower recognition rate, and the need to set a reference gesture template before each recognition also increases the computational effort. To address this, the study uses a diamond shape to constrain the global path, and the slope of the path curve is controlled to be in the range of 0.5–1. This way, when using the DTW algorithm for gesture matching, only the matching distance of the corresponding frame of the grid point within the diamond needs to be calculated. The global path constraint is shown in Figure 6.

In addition, the study sets a distortion threshold to solve the problem of too many reference gesture templates leading to increased matching complexity. Combining the above operations, the study uses the DTW algorithm to achieve dynamic recognition of gestures and realize

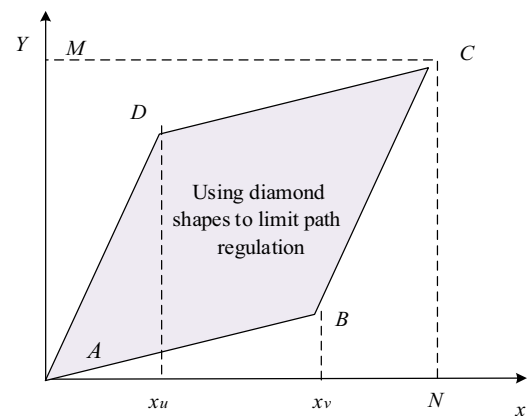


Figure 6: Global path restrictions.

forward, left, backward, and right turn operations using gestures in roaming scenes.

Ancient Chinese architecture is the oldest and most complete architectural system in the world and has long adhered to the principle of combining mechanics and aesthetics. As an important component of art, architecture provides a valuable resource for aesthetic education, and architectural art education is an important way to implement aesthetic education. When students use the institute to design a virtual scene reenactment system, the class begins with the teacher organizing students to enter the scene and acquire human skeletal data through a Kinect body sensor to connect the body camera to the Unity3D game engine. The students enter the external scene and use the movement of the hand skeleton points to control the virtual mouse to move forward, backward, left, and right in the scene. This allows them to tour the interior of the building and experience the aesthetic structure of the building. If you exit the building after the tour, when you hit the trigger collider placed at the entrance, you will recede to the outer scene interface. This allows students to experience the beauty of architecture up close and personal without having to leave home, enhancing their appreciation of beauty and thus promoting their interest in learning about art and bringing them a new teaching experience beyond the books themselves.

By designing and developing a virtual scene reproduction system and using Kinect and Unity3D game engine, this study enables students to experience the beauty of ancient architecture in an immersive way, thereby improving their ability to appreciate beauty and stimulating their interest in art learning. In the traditional education mode, students can only learn about ancient buildings through books or pictures. Although such teaching methods can provide certain information, they cannot make students feel the physical beauty and spatial sense of buildings. Therefore, through VR technology, this study allows students to personally enter the virtual ancient architecture scene, control the movement of virtual characters through their own actions, and further understand

the structure and design concept of ancient architecture. At the same time, in other aesthetic education classes such as music and art, students can feel the beauty more directly through these 3D fields, and their perception of the learning content has been further improved.

For the educational opportunities at the university level, the significance of this study lies in providing a new way and means of education. Due to the influence of many factors such as region, funds, and time, many students may not be able to visit the ancient architecture in person, and the virtual scene reproduction system developed in this research can provide an effective alternative for these students. In addition, this research can also promote interdisciplinary cooperation and exchange. Architectural art education is an important way to implement aesthetic education, which needs the cross-cooperation of many disciplines such as aesthetics, architecture, and computer science. Through the development of such a VR system, cooperation and exchange between different disciplines can be promoted, and the comprehensive quality and innovative ability of students can be improved. To sum up, this study is of great significance for improving the aesthetic education received by college students and can also expand educational opportunities at the university level and promote interdisciplinary cooperation and exchange.

4 Performance analysis of a virtual reproduction system for cultural buildings based on digital media technology

In order to examine the rendering effect of the 3D models built according to different building types, the study introduced first input delay (FID), time to interactive (TTI), total blocking time (TBT), and cumulative layout shift (CLS) as

Table 1: Rendering effects of virtual reproduction of 3D models of cultural buildings

Building type	Experiment 1				Experiment 2			
	FID (ms)	TTI (s)	TBT (ms)	CLS	FID (ms)	TTI (s)	TBT (ms)	CLS
1	65	4.7	212	0.058	64	4.8	211	0.057
2	71	5.1	223	0.071	70	5.2	224	0.070
3	68	5.0	218	0.062	69	5.1	219	0.063
4	59	4.3	204	0.050	58	4.0	201	0.043
5	78	5.6	234	0.080	79	5.7	238	0.079

evaluation metrics to evaluate the rendering effect of the models. CLS was used as an evaluation metric to evaluate the rendering effect of the model. Five different buildings were selected as modelling targets, and the rendering metrics are recorded in Table 1.

In Table 1, the average FID value obtained by the modelling method studied is 68.1ms, which is less than the maximum threshold of 100 ms for this metric; the average TTI value is 4.5, which is less than the maximum threshold of 5 s for this metric; the average TBT value is 218 ms, which is less than the maximum threshold of 300 ms for this metric; and the average CLS score is 0.063, which is less than the maximum threshold of 0.1 for this metric. The data analysis shows that after repeated testing and with different types of building modelling situations, the rendering of the model has shown satisfactory results and has high stability and safety reliability.

To test the practicality of the virtual roaming scene, the study conducted corresponding collision tests at four different locations in the same building virtual scene and recorded the change in error rate as the number of tests

increased. To further compare and analyse the reasonable-ness of the colliders chosen for the study, collision tests were also conducted using box colliders, sphere colliders, and capsule colliders in the same experimental environment. The final results are shown in Figure 7.

In Figure 7, the collision error rate in the roaming scene increases as the number of tests increases. The average error rate of the grid collider selected for the study is 9.3%, and the variation is the smallest among the four colliders as the number of tests accumulates; the average error rate of the box collider is 14.3%, 5.0% higher than that of the grid collider, the average error rate of the sphere collider is 21.4%, 12.1% higher than that of the grid collider, and the average error rate of the capsule collider is 16.8%, 7.5% higher than the grid collider. The grid collider chosen for the study provides a more realistic roaming learning experience for the user, as shown in Figure 7. To test the effectiveness of noise reduction using depth images in somatosensory interaction, the algorithm was used to recognize human poses before and after noise reduction for different signal-to-noise images, and the comparison of

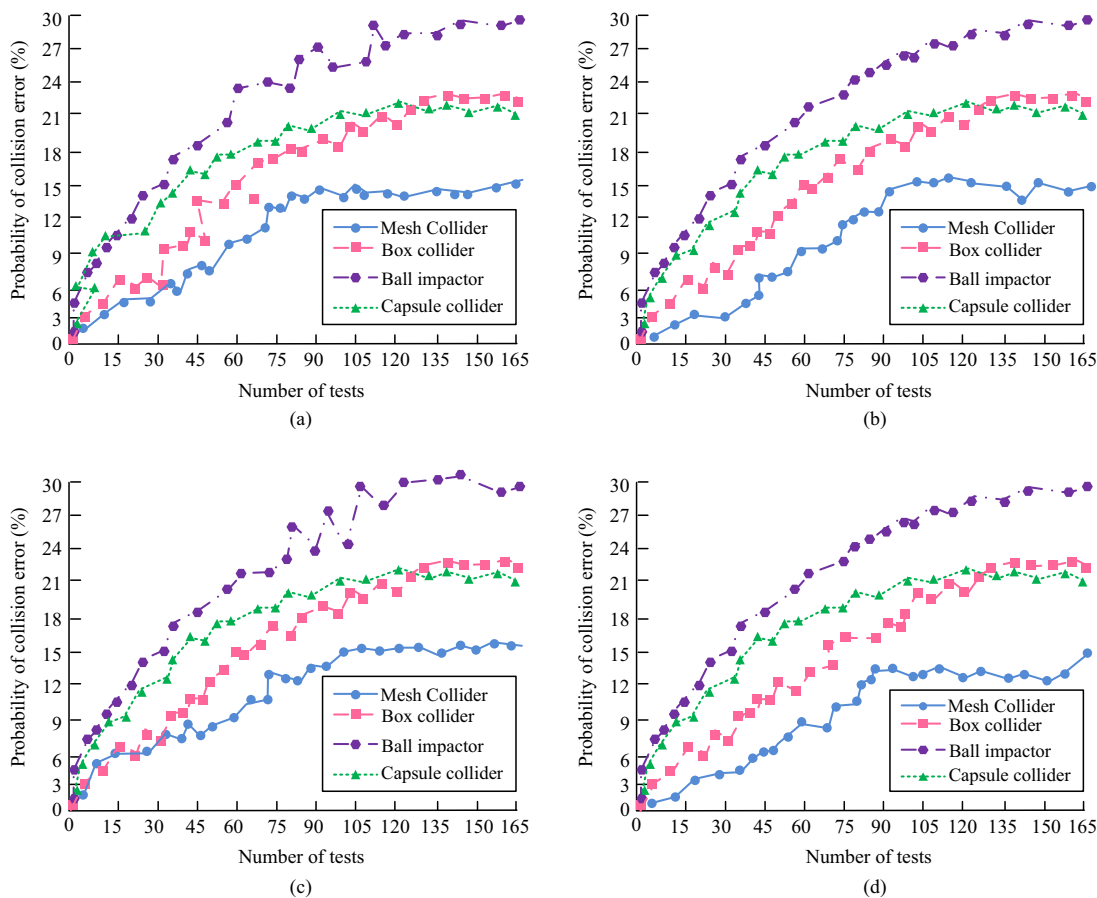


Figure 7: Comparison of impact tests of different colliders at different locations. (a) Location 1. (b) Location 2. (c) Location 3. (d) Location 4.

the recognition accuracy of the algorithm was recorded as the recognition data increased in Figure 8.

In Figure 8, the recognition accuracy of the algorithm gradually decreases as the amount of test data continues to increase, with relatively little change in the algorithm after noise reduction. The initial accuracy of the pre-noise reduction algorithm for gesture recognition gradually decreases as the signal-to-noise ratio increases. When the number of tests was 30, the recognition accuracy of the pre-noise reduction algorithm was 85.32% at a signal-to-noise ratio of 30 dB, 80.31% at a signal-to-noise ratio of 50 dB, 79.98% at a signal-to-noise ratio of 70 dB and 70.48% at a signal-to-noise ratio of 100 dB. The post-noise reduction algorithm does not change much with the signal-to-noise ratio. The comprehensive content of Figure 8 shows that the noise reduction algorithm chosen for the study has a better noise reduction effect and can make the algorithm recognition independent of the original image signal-to-noise ratio. In order to test the improvement effect of the research-improved DTW algorithm, the two algorithms were used for recognition before and after

the improvement, and the change in time and accuracy of recognition is recorded in Figure 9.

In Figure 9 (a), the gesture recognition accuracy of the algorithm gradually decreases as the test data gradually increases, with a higher decrease for the pre-improvement algorithm. At a data volume of 20, the recognition accuracy of the pre-improvement algorithm was 90.63%, and at a data volume of 200, the recognition accuracy was 78.43%, a decrease of 12.20%. In contrast, the improved algorithm only decreased by 5.41%. In Figure 9 (b), the running time of the algorithm gradually increases as the amount of data increases. The average running time of the algorithm before the improvement was 1.18 s, and the average running time of the improved algorithm was 0.62 s, which was 0.56 s less than that of the algorithm before the improvement.

To further validate the gesture recognition effectiveness of the algorithms, the research design algorithm (Algorithm 1) is compared with several more popular recognition algorithms under different lighting conditions. These include a gesture recognition algorithm that

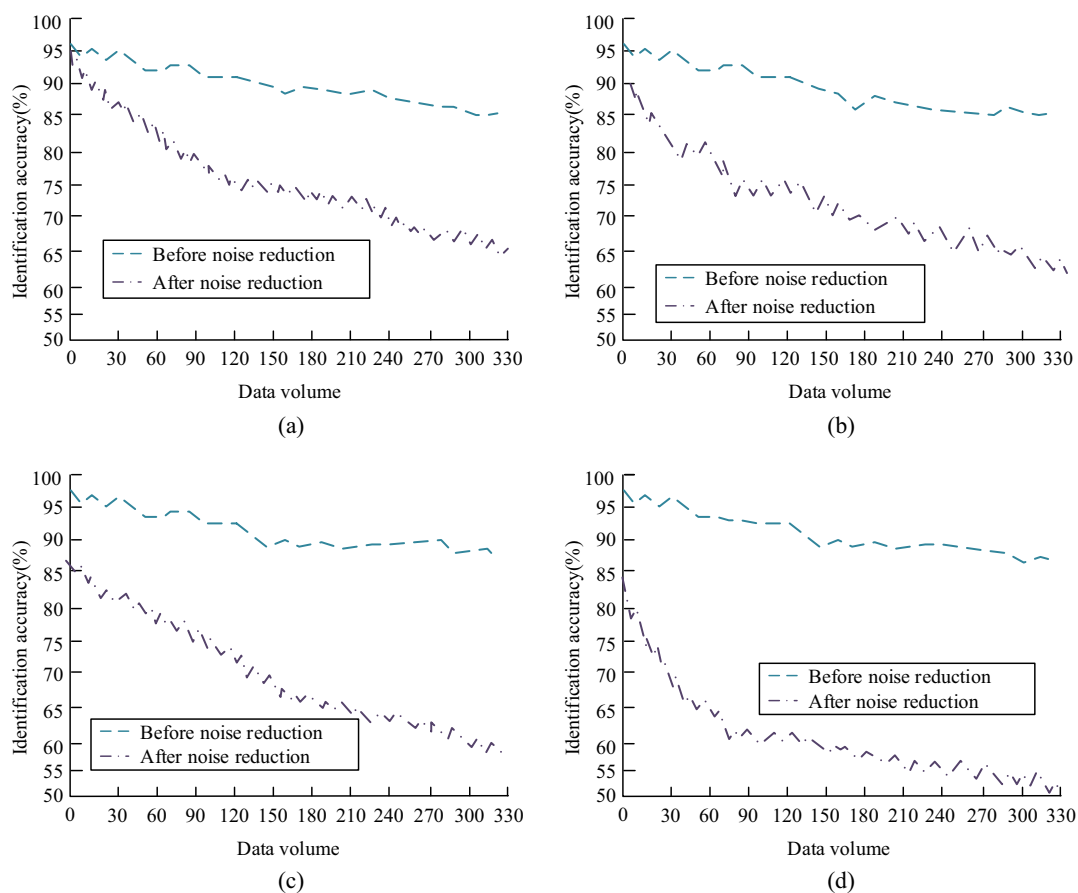


Figure 8: Comparison of recognition effects before and after noise reduction under different signal-to-noise ratios. (a) Signal-to-noise ratio = 30 dB. (b) Signal-to-noise ratio = 50 dB. (c) Signal-to-noise ratio = 70 dB. (d) Signal-to-noise ratio = 100 dB.

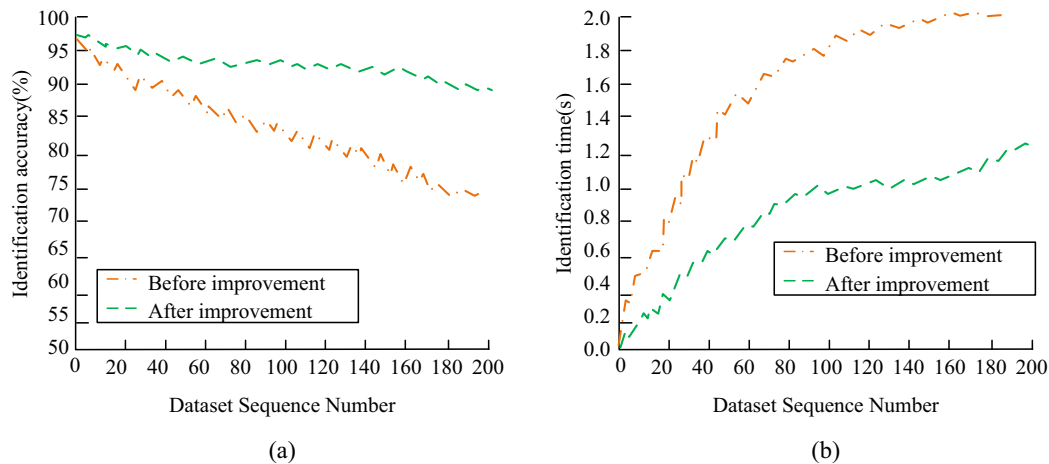


Figure 9: Comparison of recognition time and recognition accuracy of the DTW algorithm before and after improvement. (a) Run-time comparison. (b) Comparison of recognition accuracy.

incorporates attention and time-domain multi-scale features (Algorithm 2), a gesture recognition algorithm based on improved YOLOv3 (Algorithm 3), a gesture recognition algorithm based on multi-scale deep neural networks (Algorithm 4), and a recognition algorithm based on CNNs (Algorithm 5). The details are shown in Table 2.

In Table 2, the average running time of Algorithm 1 is 0.59 s, with an average accuracy rate of 96.60%; the average running time of Algorithm 2 is 0.74 s, 0.15 s more than Algorithm 1, with an average accuracy rate of 92.22%, 4.38% lower than Algorithm 1; the average running time of Algorithm 3 is 1.24 s, 0.65 s more than Algorithm 1, with an average accuracy rate of 82.24%, 14.38% lower than the average running time of Algorithm 4 was 0.88 s, 0.29 s more than Algorithm 1, and the average accuracy rate was 89.45%, 7.15% lower than Algorithm 1; the average running time of Algorithm 5 was 1.75 s, 1.16 s more than Algorithm 1, and the average accuracy rate was 76.16%, 20.44% lower than Algorithm 1. Combining the contents of Table 1, it can be seen that the gesture recognition

performance of Algorithm 1 is better than the other four algorithms and more efficient.

To test the application effect of the cultural architecture scene designed by the institute in the reform of aesthetic education teaching methods in universities, the research department was divided into four batches into practical teaching experiments, and the experience survey was conducted among students who participated in the teaching. The questionnaire investigates students' learning outcomes from four aspects: experience, stimulation of learning interest, comparison of teaching methods, and degree of learning assistance, and records the specific data in Table 3.

As shown in Tables 3, 94.81% of the surveyed students believe that they can stimulate their interest in aesthetic education classroom learning, and 95.12% of the surveyed students believe that there is a big difference with the previous teaching mode. The students who are more willing to use this teaching method for classroom learning account for 91.05% of the total number surveyed. The students who think that this kind of teaching method can

Table 2: Comparison of gesture recognition of various algorithms under different lighting conditions

Project		Algorithm 1	Algorithm 2	Algorithm 3	Algorithm 4	Algorithm 5
Backlight	Time (s)	0.58	0.75	1.21	0.88	1.76
	Accuracy (%)	96.84	92.43	82.15	89.47	76.23
Natural light	Time (s)	0.59	0.74	1.22	0.89	1.74
	Accuracy (%)	96.72	92.11	82.41	89.52	76.21
Strong light	Time (s)	0.61	0.73	1.23	0.90	1.77
	Accuracy (%)	96.43	92.01	82.11	89.44	75.98
Dim light	Time (s)	0.58	0.74	1.26	0.87	1.73
	Accuracy (%)	96.41	92.32	82.30	89.36	76.22

Table 3: Survey results of the application of research design methods to the reform of university aesthetic education teaching models

Batch	Can stimulate learning interest (%)			Whether there is a difference in degree (%)			Hope the classroom format is combined with VR (%)			This teaching method is helpful to aesthetic education learning (%)		
	A	B	C	A	B	C	A	B	C	A	B	C
1	96.42	2.13	1.45	94.77	3.26	1.97	90.45	7.81	1.74	91.58	6.88	1.54
2	95.23	3.48	1.29	96.48	2.78	0.74	91.75	6.99	1.26	92.05	6.59	1.36
3	92.47	5.62	1.91	95.23	3.64	1.13	90.84	7.48	1.68	92.11	6.94	0.95
4	95.12	3.12	1.76	93.99	6.11	0	91.15	7.05	1.80	91.75	6.77	1.48

A: Very recognized, B: Generally recognized, C: Not recognized.

bring significant help to students' learning account for 91.87% of the total students surveyed. Based on the contents in the table, it can be seen that the method designed by the research institute can effectively improve the teaching effect of college aesthetic education and enhance the learning interaction of students.

In order to further study the practical application effect of the aesthetic education teaching reform method designed, 100 sophomore and junior college students were selected from a number of colleges and departments who voluntarily chose aesthetic education courses. They were divided into two groups: test group (A) and control group (B). There were 50 people in each group, and the mean age was 20.5 years. The students in the control group were delayed to the second semester to use the reformed aesthetic education teaching method, while the experimental group was taught the reformed mode in the first semester. The primary effect of aesthetic education is the formation and improvement of aesthetic ability or aesthetic humanistic accomplishment. As a post-test tool, the questionnaire of "Aesthetic ability Test" is composed of eight questions. The subjects' answers to each question were scored by two professional teachers, respectively, and the average score was taken as the actual score of the subjects. A1 and B1 represent two groups of students in the first semester, and A2 and B2 represent two groups of students in the second semester. The test results of the two groups of students are shown in Table 4.

As can be seen from Table 4, in the first period of learning, there was a significant difference in the aesthetic

ability between group A1 and group B1. However, it can be seen that both groups of students have improved their aesthetic ability after taking the course. In the learning process of the second semester, the aesthetic ability of group B2 was significantly improved compared with group B1. Based on the data in Table 4, it can be seen that the teaching method proposed by the research can effectively improve the teaching effect and enhance the learning ability of students.

According to the practical application results and student questionnaire survey, besides the above-mentioned methods, there are some other ways to make the proposed teaching method more effective:

1. Combine a variety of digital media technologies: In addition to using 3D modelling and virtual roaming technology, you can also consider combining other digital media technologies to provide a more authentic and immersive learning experience and better stimulate students' learning interest.
2. Introduction of gamification elements: Integrating gamification elements into the teaching process can increase the interest and interaction of learning. For example, we can design some games related to ancient cultural buildings, so that students can learn knowledge in the game and improve the learning effect.
3. Carry out cooperative learning: Introducing cooperative learning into classroom teaching can make students better play their own advantages, learn from each other, help each other, and improve the learning effect. For example, students can be organized to have group

Table 4: Comparison of the results of aesthetic ability in the two groups after two semesters of study

Project	Mean (md)	Average (X)	Standard error (s)	t	P	Differences between groups	Concentration ratio (%)
A2	82.0	80.4	6.8	42.83	0.04	3.52	48.41
B2	70.5	70.2	14.0			61.00	
A1	80.4	80.1	7.0	32.84	0.00	3.85	-30.3
B1	41.8	40.5	12.5			27.01	

- discussions, team projects, and other activities to cultivate students' sense of cooperation and team spirit.
4. Establish a scientific evaluation system: The establishment of a scientific evaluation system can timely understand the learning situation and learning needs of students and provide references for the subsequent teaching process. At the same time, the teaching method and content can be optimized and the teaching quality can be improved through the evaluation results.

5 Limitations and future directions

Kinect devices can be blocked by obstacles, and the placement of the device can be optimized to ensure that obstructions can be minimized in the classroom or learning environment. The current research has demonstrated the potential of gesture recognition in aesthetic education teaching, and the future can further explore the deep integration of gesture recognition and speech recognition technology to achieve a more natural and smooth interactive experience. For example, students can freely explore the virtual roaming scene through voice commands and gestures, enhancing the interactive and immersive feeling of learning. In the later research work, aesthetic education teaching can be integrated with history, literature, philosophy, and other disciplines to further enrich the teaching model of aesthetic education.

6 Conclusion

At the present stage, the teaching of aesthetic education is mainly a combination of simple art skills teaching and aesthetic education, and the teaching effect achieved is not ideal. To this end, the study makes use of digital media technology to model ancient cultural buildings in three dimensions and construct virtual roaming scenes. This enables students to experience the heritage and elegance of ancient cultural architecture more deeply in the classroom and enhances their aesthetic ability. The average runtime of the DTW algorithm is 0.62 s, which is 0.56 s less than that of the algorithm before the improvement, and the improvement is obvious. The recognition algorithm designed for the study had an average running time of 0.59 s and an average accuracy rate of 96.60%, and its gesture recognition performance was better than the other four algorithms. The Kinect also has a speech

recognition function, which can be considered in future research to combine speech recognition with dynamic gesture recognition in architectural virtual roaming to bring a more interactive learning experience to students. The depth image data and bone joint data used by the research institute are obtained through Kinect. If the device's implementation is obstructed by obstacles, it will inevitably affect the final recognition effect. Research results can be achieved by participating in academic conferences, presenting our research results at conferences, and exchanging and discussing with relevant researchers. In addition, we can also cooperate with educational practitioners to apply the research results to actual educational practice, and test the effectiveness and operability of our research results through practice. Finally, through cooperation with the media, the research results can be disseminated to the public, so as to improve the public's awareness of and attention to aesthetic education teaching and further promote the development of aesthetic education teaching.

Funding information: The research is supported by Provincial Teaching Reform Research Project for Undergraduate Universities in Hubei Province in 2024: "Research on the Theory and Practice of Digital Resources of Intangible Cultural Heritage in Jingchu Area Enabling Intelligent Aesthetic Education" (Project No. 2024574).

Author contributions: Jiayue Yan: conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, project administration, resources, software, supervision, validation, visualization, writing – original draft, writing – review & editing.

Conflict of interest: The author declares that there is no conflict of interest.

Data availability statement: The data used to support the findings of the research are available from the corresponding author upon reasonable request.

References

- [1] C. Li and M. K. Saat, "The development of aesthetic education: A perspective of calligraphy and painting theory teaching for Chinese institutions of higher learning in new era," *Cross-Cultural Commun.*, vol. 18, no. 1, pp. 83–91, 2022.

- [2] M. He, "Innovative construction of aesthetic education curriculum system in engineering college," *Adv. Educ. Humanit. Soc. Sci. Res.*, vol. 2, no. 1, p. 146, 2022.
- [3] L. Bi, "Research on the teaching reform of business English elective courses under the background of the integration of specialty," *Innov. Curric. Teach. Methodol.*, vol. 5, no. 6, pp. 5–9, 2022.
- [4] M. G. Amanbaevich and S. G. Sarsenovna, "Education of ethics in fine arts," *Cent. Asian J. Arts Des.*, vol. 4, no. 1, pp. 15–18, 2023.
- [5] H. Ma and J. Li, "An innovative method for digital media education based on mobile internet technology," *Learning (IJET)*, vol. 16, no. 13, pp. 68–81, 2021.
- [6] K. A. Mills and A. Brown, "Immersive virtual reality (VR) for digital media making: transmediation is key," *Learn. Media Technol.*, vol. 47, no. 2, pp. 179–200, 2022.
- [7] R. Jiang, L. Wang, and S. B. Tsai, "An empirical study on digital media technology in film and television animation design," *Math. Probl. Eng.*, vol. 2022, pp. 1–10, 2022.
- [8] Z. Liu, "Application analysis of computer digital media technology in advertising art," *For. Chem. Rev.*, vol. 3, pp. 301–308, 2022.
- [9] Y. Lu, "Influence of digital media technology on forest scene animation design," *For. Chem. Rev.*, vol. 1, pp. 54–61, 2021.
- [10] W. Zhu, "Study of creative thinking in digital media art design education," *Creative Educ.*, vol. 11, no. 2, pp. 77–85, 2020.
- [11] P. Xu and Y. Guan, "The design and implementation of "four-in-one" blended learning model in digital media technology classroom," *Scholar: Hum. Sci.*, vol. 12, no. 1, p. 404, 2020.
- [12] P. U. Ineji and I. P. Ogar, "Impact of digital media on effective healthcare delivery in Cross River State," *Int. J. Commun. Res.*, vol. 11, no. 1, pp. 73–79, 2021.
- [13] V. Gragorious and D. Herron, "Digital media design instruction in relation to media integration," *Glob. Media J.*, vol. 20, no. 55, pp. 1–3, 2022.
- [14] T. Mao and X. Jiang, "The use of digital media art using UI and visual sensing image technology," *J. Sens.*, vol. 2021, pp. 1–11, 2021.
- [15] J. Li, E. Xue, J. Li, and E. Xue, "Shaping the aesthetic education in China: Policies and concerns," In *Shaping Education Reform in China*, vol. 9, Springer, Singapore, 2020, pp. 127–153.
- [16] L. Wang, Y. Li, and M. Fang, "Aesthetic Education Curriculum Construction and Innovation In The New Engineering Education: An example of teaching an art appreciation aesthetics education cours," *Adv. Educ. Humanit. Soc. Sci. Res.*, vol. 1, no. 1, p. 369, 2022.
- [17] Y. Shi and D. Cheng, "Study on the connection of aesthetic education and classrooms in primary and secondary schools," *For. Chem. Rev.*, vol. 11, pp. 957–966, 2021.
- [18] X. Wen, J. Shen, and X. Gao, "A comparative study of online aesthetic education courses in Chinese and American Colleges: Based on the "iCourse" and "edX" platforms," *Int. J. Educ. Humanit.*, vol. 5, no. 2, pp. 103–109, 2022.
- [19] N. Khasanova, "The role of music lessons in the formation of national and intercultural competence in students," *Ment. Enlight. Sci.-Methodol. J.*, vol. 2020, no. 2, pp. 130–139, 2020.
- [20] D. He and N. Luo, "Spatiotemporal evolution characteristics of extreme rainfall based on intelligent recognition and evaluation of music teaching effect in colleges and universities," *Arab. J. Geosci.*, vol. 14, pp. 1–13, 2021.