Virtual Reconstruction of Museum Spatial Information Based on Unity3D

Zi Yang[1(✉)] and He Wang[2]

[1] Hubei Institute of Fine Arts, Wuhan 430060, China
yangzi1232022@163.com
[2] Changchun Humanities and Sciences College, Changchun 130051, China

Abstract. Museum space scene structure is complex, color information contains more, it is difficult to obtain complete and accurate three-dimensional scene information. Based on Unity3D, a virtual reconstruction method of museum spatial information is designed. Aiming at the notable structured information in museums, the three-dimensional space coordinates of museums are calculated. After the scale space is established, the extreme points in the scale space are detected, and the image sequences are detected and matched. The surface information of the point cloud model is obtained by solving Poisson's equation, and the surface with geometric entity information is obtained by fusing the point cloud. The museum 3D virtual reconstruction model is built in Unity3D to realize the transformation from 2D image to 3D scene. The test results show that the proposed method improves the similarity of museum structure to a certain extent, and can restore the spatial information of museum more truly.

Keywords: Unity3D · Museum · Spatial reconstruction · Spatial information · Virtual reconstruction · 3D reconstruction

1 Introduction

It is well known that humans perceive their surroundings mainly through vision. Through the eyes to obtain, transmission, memory and understanding of the environment around the external information, thus understanding the world, so the importance of vision for human beings is self-evident. But in some high-risk or precision and other special places, human incompetence. Therefore, with the development of computer technology, the computer vision technology has emerged, which simulates the human vision digitally, acquires and interprets the information needed in the work. With the development of artificial intelligence industry, the ability to perceive the space and process the information of the surrounding environment has been greatly improved. Nowadays, we can often see the scene of applying the 3D reconstruction technology of indoor scene [1]. For example, interior architects can measure the whole room size in advance, do layout planning and furniture customization in advance by using 3D virtual reconstruction technology. The museum can use the indoor scene 3D virtual reconstruction technology to

W. Fu and L. Yun (Eds.): ADHIP 2022, LNICST 468, pp. 648–659, 2023.
https://doi.org/10.1007/978-3-031-28787-9_48

build an online museum to meet the needs of more people. Mobile phone manufacturers can reconstruct the 3D model of indoor scene by using the 3D virtual reconstruction technology, generate various annotated images more conveniently and train convolution neural network.

A museum is an institution, building, place or public institution that collects, collects, displays and studies objects representing the natural and human cultural heritage, classifies those objects of scientific, historical or artistic value, and provides the public with knowledge, education and appreciation of culture and education. Aiming at indoor scene 3D reconstruction is one of the most valuable directions in 3D reconstruction. How to make 3D reconstruction more effective, and how to make machine obtain all kinds of 3D information by image is of great significance. Virtual museum is a kind of digital exhibition hall constructed by computer 3D modeling technology and virtual reality technology. It is a kind of 3D interactive experience mode. Based on the traditional museum, virtual museum and its exhibits are transplanted onto the Internet by using virtual technology. It breaks through the limitation of time and space, enriches the exhibition mode of museum and enhances the publicity and education functions of museum. Three-dimensional data, the computer to its acquisition, cognition, transformation, there is still a lot of room for improvement. It can be seen that the computer has the same powerful function as the human eye, that is, it can get the information from the three dimensional world at anytime and anywhere, and directly process, describe and judge the obtained information. The process of obtaining 3D information by means of computer or digital sensor is called 3D virtual reconstruction. The research goal of 3D virtual reconstruction is to input one or several images to computer for analysis and processing, so as to restore the 3D spatial information of reconstructed objects.

Unity3D, a multi-platform integrated development tool for building visualizations, real-time 3D animation and other types of interactive content developed by Unity3D Technologies, is a fully integrated professional engine that provides a fully fledged interface, one of its greatest features being support for cross-platform development. The Unity3D engine has a rich library of tutorials and descriptions of physics engines, animations, sounds, timelines, scripts, and sample code for developers to use to get a faster understanding of the interface. Based on Unity3D, a virtual reconstruction method of museum spatial information is proposed in this paper, and the information from different angles can be seen directly and concretely.

2 Virtual Reconstruction of Museum Spatial Information Based on Unity3D

2.1 Computing Museum Three-Dimensional Space Coordinates

Virtual museum is the use of virtual reality technology, simulation models or scenes, used to show the museum as a whole, or the history of real existence, has now disappeared or endangered scene. Virtual museums are generally presented as a separate roaming system for visitors to browse, learn to use, can also be embedded in the web page, as part of the digital museum functions to display to visitors. Three-dimensional virtual reconstruction is the process of obtaining 3D scene from the matching disparity map. The image taken by binocular camera is processed, and the 3D information of spatial objects is calculated by parallax map. The information that the camera simulates as the imaging model is the camera parameters, including the internal parameters and external parameters. Only when the specific parameters of the camera are determined, the spatial geometry of the target can be further reconstructed. Among them, internal parameters include the optical characteristics, scale factor, focal length and lens distortion of the camera itself, while external parameters mainly include the camera's position in the world coordinate system. Projecting a 3-D point onto a normalized image plane, then calculating the radial and tangential distortions of the points on the normalized plane, and projecting the de-distorted points onto the pixel plane through the internal parameter matrix, the correct position of the point on the image is obtained [2]. In the following discussion, we assume that all images have been de-distorted and that there is no need to include distortion parameters in the calculation. Firstly, the museum image taken by binocular camera is transformed to coordinate, and the three-dimensional space coordinate is obtained. Any spatial object point in a real-world scene is projected simultaneously in two images. According to the principle of binocular imaging system, the coordinate of image points in 3D space is solved by triangular geometry. Assuming that the parallax of a matching point A_1 and A_2 on the left and right image is h, the following relationships can be deduced according to the imaging scale relationship:

$$\begin{cases} p_1 - p_0 = g_a \dfrac{a_1}{c_1} \\ q_1 - q_0 = g_b \dfrac{b_1}{c_1} \\ p_2 - p_0 = g_a \dfrac{a_1 - h}{c_1} \\ q_2 - q_0 = g_b \dfrac{b_1}{c_1} \end{cases} \quad (1)$$

In Formula (1), (p_0, q_0) represents the central coordinate value of the left and right images; (p_1, q_1) and (p_2, q_2) represent the pixel coordinates of a matching point on the left and right images respectively; (a_1, b_1, c_1) represents the three-dimensional space coordinates of the pixel in the left camera coordinate system; and g_a and g_b represent the focal length of the left and right cameras on the horizontal and vertical axes. The expression of orientation in the position and pose of a camera is often described indirectly by the expression of rotation. The pointing of the camera to the Z axis in the current camera coordinate system (generally default to camera orientation) is expressed through a rotation transformation that completes the conversion of the camera from the Z axis in the world coordinate system to the Z axis in the current camera coordinate system [3]. Epipolar alignment of the image, the matching points are on the same line, so the pixels are the same vertical. Therefore, the spatial coordinate data of any point on the image is obtained, and the calculation formula is as follows:

$$\begin{cases} a_1 = \dfrac{h(p_1 - p_0)}{p_1 - p_2} \\ b_1 = \dfrac{hg_a(q_1 - q_0)}{g_b(p_1 - p_2)} \\ c_1 = \dfrac{hg_a}{p_1 - p_2} \end{cases} \tag{2}$$

According to the location of two points, the camera calibration result is obtained. The 3D coordinates of the point in the camera coordinate system in space can be solved by the inner and outer parameter matrices.

2.2 Detection of Spatial Information Characteristics of Museums

After calibrating and correcting the museum 3D space coordinates, it is necessary to detect and match the features of the images taken by the museum scenes. In the work of 3D reconstruction of image sequence, a key part is to find out the relationship between images in image sequence, and then get the 3D information according to the mutual information of these images. Although there is not much noise and weather in the indoor scene, the scene structure is complex and the color information contains more. Direct grayscale algorithm will lose part of the original color information of the image. This paper proposes a grayscale compensation algorithm to extract more effective feature points in feature detection. The resulting gray scale value is based on the original gray scale value to increase color compensation, the formula is as follows:

$$d = \gamma + \gamma' + \varphi \tag{3}$$

In Formula (3), d represents the gray value after compensation; γ and γ' represent the gray value and color compensation amount of the original image respectively; and φ represents illumination compensation amount. The formula for the amount of color compensation is as follows:

$$\gamma' = \delta_1 \mathrm{sgn}(200\beta_1)|200\beta_1|^{\delta_2} + \delta_1 \mathrm{sgn}(50\beta_2)|50\beta_2|^{\delta_2} \tag{4}$$

In formula (4), δ_1 is the color contrast parameter, the value of which is 2.3; δ_2 is the range parameter, the value of which is 0.6; sgn is the sign function; β_1 is the chroma signal; and β_2 is the saturation signal. The extraction equation for the amount of light compensation is as follows:

$$\varphi = \varepsilon \gamma' e \left| -\frac{\gamma}{2\eta^2} \right| \tag{5}$$

In formula (5), the value of the contrast parameter of ε illumination is 1.8; the value of e is the natural constant; and the value of η is the standard deviation, which is 0.25. Because there are many pixels involved in the 3D virtual reconstruction of spatial features of museums, and because of occlusion and other problems, it is easy to track the end points of line segments in images because of different viewpoints, the end points of line segments do not belong to the corresponding situation, but deteriorate the performance [4]. So the descriptor is constructed here, and then the descriptor matching method is used to extract the line feature descriptor. Feature points are extracted using Harris corner extraction algorithm. After extraction, it is necessary to check the quality of its checkpoints, judge whether it contains the required information, and it can not be the edge of the image.

Image points are different from other points in the image, that is, some special image points can provide more useful information, such points are called feature points. Obtaining the 2-D coordinates of such feature points by using the relevant feature point detection technology is the foundation and key step of 3-D reconstruction. Generally speaking, the size of the image entropy value is proportional to the size of the information contained in the image. After the scale space is established, the extreme points in the space are detected. Then, the resulting point is chosen as the key candidate point. Specifically, in the detection, each pixel is compared with a certain point. Taking a pixel as an example, this paper compares the pixel with 8 neighboring pixels in the same scale and 18 neighboring pixels in the corresponding position of upper and lower neighboring scale. At the same time, because the feature corners are reduced, the matching time can be reduced, which improves the real-time of the algorithm. The feature corners extracted by Harris replace the feature points extracted by Gaussian convolution algorithm in SIFT algorithm. SIFT algorithm has a certain stability, its extraction operator in the brightness, scale changes, such as interference, can still detect a large number of accurate feature points. Markov distance is used to judge the similarity between the two eigenvectors. At the same time, the correlation between feature vectors is eliminated, and the matching precision is improved.

2.3 Point Cloud Fusion of Museum Spatial Information

Point cloud fusion of museum spatial information is to eliminate redundant observation points of the same name. By point cloud fusion, data redundancy can be reduced and dense sampling points on the surface of objects can be obtained. The flow of the point cloud fusion is shown in Fig. 1.

It can be seen from Fig. 1 that first, it is necessary to collect point cloud data from multiple viewpoints, and then perform point cloud filtering. On this basis, multi-view point cloud splicing and matching are performed, and the integration of point cloud data is completed. Keep the equipment in the structured light system fixed and set the number of shots in advance according to the complexity of the target model. The scanned object is placed on the rotating table, and the rotating table drives the target model to rotate at a certain speed. The camera collects 8 point cloud data with different viewing angles. The key point of point cloud fusion is how to distinguish the same name from multiple observations. According to the spatial resolution of the reference image under the light beam of the same name in the image to be matched, determine whether the reference image is a point of the same name. The formula is as follows:

$$w(n) = \frac{|\kappa_1(n) - \kappa_2(n)|}{\kappa_2(n)} \tag{6}$$

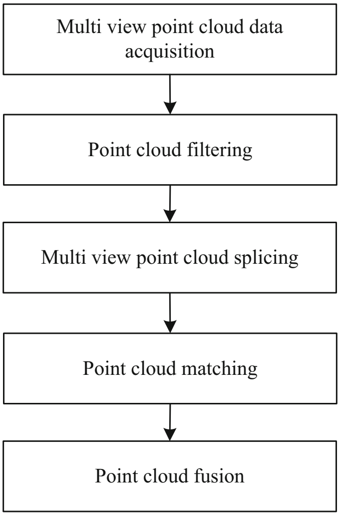

Fig. 1. Flow of point cloud merging

In Type 6 (6), n represents spatial resolution; $w(n)$ represents judgment function; $\kappa_1(n)$ represents the corresponding depth value in the image to be matched; and $\kappa_2(n)$ represents the depth value of the reference image guided by object points in the image to be matched. In 3D space, the distribution of point cloud is scattered and disordered, and it is difficult to establish a connection between them, but there is a topological relationship between points and points in space. At this point, all the points calculated by the search are Gaussian normal distribution, the rest are removed as outliers by filtering algorithm. If the point points in the space of the current processing pixel have the same names in the fused point cloud, they are directly ignored, otherwise, the fused point cloud data are added, and the same names in the reference image and the image to be matched are marked as processed [6]. The plane information is detected from the depth map, and then sparsely sampled in the plane to reduce the data representation redundancy. The surface information of point cloud model is obtained by solving Poisson's equation, and the surface model with geometrical entity information is obtained by reconstructing the target model. According to local information to evaluate the smoothness of object points, the potential plane position of object is given, and then sparse sampling is carried out in object plane to reduce the redundancy of expression in dense point cloud [7]. The formula for calculating the mean value of the difference between objects is:

$$\vartheta = \frac{\sum_{N=1}^{n} N(\tau_1 - \tau_2)}{|\tau_2|} \tag{7}$$

In Formula (7), ϑ indicates the mean difference of the points; N indicates the total number of points; τ_1 indicates the neighborhood depth value in the depth chart; and τ_2 indicates the depth value of the current point. After obtaining the probability of each point in the object plane in the depth map, the object plane can be changed into a connected region by the way of region growth. Then we can use the grid in the image plane to sample the connected area of the depth map.

2.4 Establishment of 3D Virtual Reconstruction Model of Museum Based on Unity3D

It has become a reliable method to obtain 3D models of building level objects by 3D virtual reconstruction technology. The good user experience in the 3D virtual reconstruction of museum spatial information mainly depends on the immersion, and the immersion mainly depends on the reality of the 3D scene. Therefore, the fineness of the 3D model in the scene, the mapping and the light efficiency all play a very important role. In this paper, Unity3D is used to complete the three-dimensional virtual reconstruction of museum spatial information, that is, to achieve the transformation from two-dimensional image to three-dimensional scene. The reconstruction method can obtain the 3D structure of the object quickly without the limitation of equipment and environment, and the point cloud is rich in information and the triangulated model has strong stereo. The 3D model is too rough and crude, the reality of the whole scene will be reduced, the user's immersion will be reduced or even all disappeared. However, if the scene is too complex, too refined, the higher the requirements for computer performance, the average user's computer system will run faster, real-time will also be greatly reduced, resulting in a direct impact on the user's experience. Therefore, it is necessary to consider the trade-off in modeling, so as to enhance the authenticity of the model. In Unity3D, the rotation of the camera enables the rotation of the 3D spatial field of view for viewing museum models in 3D space. To achieve the rotation function, the following scheme is set. The camera position shall remain unchanged, and the camera angle shall be changed correspondingly only by moving the mouse (PC end) or sliding gestures (mobile end). The implementation steps for this scenario are shown in Fig. 2.

The code is simple and easy to maintain, only dynamic change the camera angle, rotation speed and angle easy to control, the user experience is better. The 3D virtual reconstruction model of museum spatial information is the most suitable choice because of the limitation of space occupied and the actual operation and interaction. So the interface design is also based on the PC. In Unity3D, the translation of 3D field of vision is realized by the translation of the camera, mainly by the direction vector of the camera left and right, and the direction vector of camera up and down. The translation of the viewport in this way requires special attention to the speed control of the translation, which should be related to the size of the viewport to achieve a comfortable translation effect. According to the rules, location, layout, quantity, visual rendering, associated business items, constraints and other structured processing of node attributes. Then it realizes visualization display and 3D engine rendering. Compared with other prototyping software, Unity3D is more convenient and easy to modify and export. It can also set up connections and actions and interactive demo. Low-fidelity prototypes can efficiently test the usability of a system, improve the user experience, and reduce the cost of program development. Three-dimensional zoom is achieved by controlling the viewport size of the camera in Unity3D and dynamically changing the viewport size by sliding the mouse wheel or two fingers. Increase your camera viewport by a multiple of 1.1 as the mouse wheel scrolls forward or slides outwards, and decrease your camera viewport by a

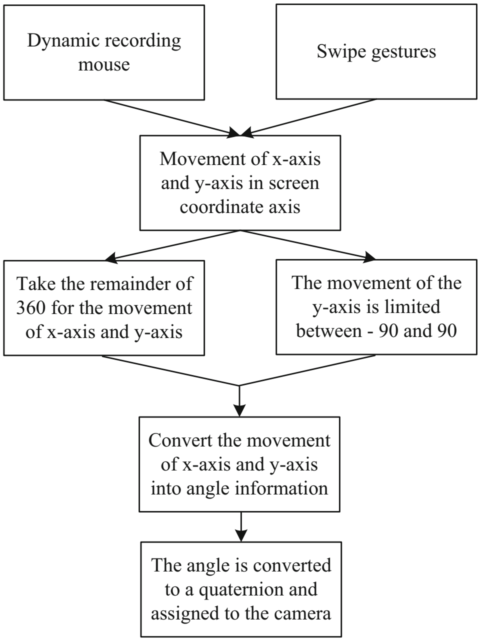

Fig. 2. Realization steps of 3D visual field rotation

multiple of 1.1 as the mouse wheel scrolls backward or slides inward. By encapsulating the JSON format string data, the node attributes in the JSON string are filtered, and the complex multi-layer topology is parsed and filtered. The pre-processing data (multi-layer, processing, filtering) is input, and then the algorithm is invoked to calculate the layout of the network topology. According to the constraints of the node attributes, the layout strategy, the layout algorithm, the layout rules are associated with nodes, and the nodes are displayed, which makes the visualization model of museum spatial information clear and flexible. Thus, the design of 3D virtual reconstruction method of museum spatial information based on Unity3D is completed.

3 Experimental Study

3.1 Experimental Preparation

In order to evaluate the performance of 3D virtual reconstruction method more accurately, the performance of the method is tested on the open dataset. The dataset used is ICLNUIM dataset, it is the dataset that reconstructs the indoor scene as the target. It has the true value of the point cloud reconstruction model, which is mainly used for the evaluation of 3D reconstruction. The test images selected for this article are the same scene shot from the front, left, right, and above four directions, corresponding to four sequences, each with the number of frames and sequence length shown in Table 1.

In order to quantitatively analyze the results of spatial 3D virtual reconstruction, structural similarity is used to evaluate the results. Structural similarity is a measure

Table 1. Scene sequence features

Sequence	Direction	Frame number	Binocular image sequence
1	Front face	1623	32
2	Left lateral surface	1027	34
3	Facies lateralis	1236	42
4	Above	982	38

of the similarity between two images, which is symmetrical, bounded and uniquely maximal. The brightness is estimated by mean, the contrast is estimated by standard deviation, and the similarity is estimated by covariance.

3.2 Results and Analysis

A 3D virtual reconstruction model is constructed after the same angle is obtained in the front, left, right and above four directions, and the structure similarity between the output image and the scene image is calculated. The measurement results of 3D virtual reconstruction of museum spatial information based on Unity3D proposed in this paper are compared with those based on SFM and BIM. The experimental results are shown in Tables 2–5.

Table 2. Structural similarity comparison of sequence 1

Number of tests	Virtual reconstruction of museum spatial information based on unity3D	Virtual reconstruction of museum spatial information based on SFM	Virtual reconstruction of museum spatial information based on BIM
1	0.753	0.683	0.674
2	0.762	0.685	0.678
3	0.775	**0.698**	0.665
4	0.786	0.675	0.662
5	**0.788**	0.682	0.673
6	0.775	0.686	0.675
7	0.782	0.693	**0.682**
8	0.766	0.682	0.681
9	0.752	0.685	0.664
10	0.781	0.674	0.671

In the front modeling of indoor scene, the similarity of 3D virtual reconstruction method based on Unity3D is 0.772, which is higher than that based on SFM and BIM.

Table 3. Structural similarity comparison of sequence 2

Number of tests	Virtual reconstruction of museum spatial information based on unity3D	Virtual reconstruction of museum spatial information based on SFM	Virtual reconstruction of museum spatial information based on BIM
1	0.747	0.644	0.644
2	0.744	0.657	0.638
3	0.738	0.668	0.646
4	0.746	**0.676**	0.655
5	0.745	0.652	0.662
6	0.732	0.665	0.663
7	0.753	0.658	0.656
8	**0.759**	0.666	0.652
9	0.745	0.652	0.641
10	0.742	0.643	**0.667**

Table 4. Structural similarity comparison of sequence 3

Number of tests	Virtual reconstruction of museum spatial information based on unity3D	Virtual reconstruction of museum spatial information based on SFM	Virtual reconstruction of museum spatial information based on BIM
1	0.747	0.650	**0.672**
2	0.738	0.647	0.642
3	0.746	**0.678**	0.666
4	0.759	0.666	0.655
5	0.733	0.652	0.668
6	0.732	0.640	0.655
7	0.745	0.673	0.642
8	0.749	0.636	0.653
9	**0.763**	0.665	0.660
10	0.751	0.652	0.631

In the left side modeling of indoor scene, the similarity of 3D virtual reconstruction method based on Unity3D is 0.745, which is 0.087 and 0.093 higher than that based on SFM and BIM.

In the right side modeling of indoor scene, the similarity of 3D virtual reconstruction method based on Unity3D is 0.746, which is 0.090 and 0.092 higher than that based on SFM and BIM.

Table 5. Structural similarity comparison of sequence 4

Number of tests	Virtual reconstruction of museum spatial information based on unity3D	Virtual reconstruction of museum spatial information based on SFM	Virtual reconstruction of museum spatial information based on BIM
1	0.713	0.604	0.590
2	0.704	0.609	0.597
3	0.718	0.622	0.588
4	**0.726**	**0.623**	0.616
5	0.702	0.615	**0.625**
6	0.703	0.612	0.593
7	0.715	0.598	0.592
8	0.702	0.583	0.611
9	0.701	0.591	0.624
10	0.714	0.612	0.602

In the modeling of indoor scene, the similarity of 3D virtual reconstruction method based on Unity3D is 0.710, which is 0.103 and 0.106 higher than that based on SFM and BIM. From the above results, we can see that the structure similarity of the proposed method is improved to a certain extent, which shows that the proposed method has higher precision and can more accurately restore museum 3D virtual scene information. Because the method in this paper detects the extreme points of the space on the basis of the scale space. On this basis, the image sequence is matched, and the surface with geometric entity information is obtained by fusing point clouds. This can further improve the virtual matching accuracy. Finally Unity3D built a 3D virtual reconstruction model of the museum.

4 Conclusion

Museum virtual space display is a new experience learning method for visiting learners, which can greatly improve learners' immersion and experience. Virtual museums provide learners with rich and real information and sensory senses, enable more people to receive cultural knowledge anytime, anywhere, enrich the methods of experiential learning, and make the educational function of museums play the greatest role and effect. In this paper, the three-dimensional spatial coordinates of the museum are calculated based on the significant structured information in the museum. The extreme points in the scale space are detected, and the image sequence is detected and matched, which improves the accuracy of 3D virtual reconstruction. The built-in 3D virtual reconstruction model of the museum in Unity3D realizes the conversion from 2D images to 3D scenes. In the process of feature extraction and matching, there are few feature points on the surface of the cabinet with smooth surface and weak texture representation, which will cause hollows in the subsequent reconstruction. Experiments show that the method in this paper improves the similarity of the museum structure and can restore the spatial information of the museum more realistically. How to extract features from smooth and weak texture surfaces and fill in the voids smoothly and without distortion is the future research direction.

References

1. Yu, Z., Peng, X., Qiu, C., et al.: An improved approach of indoor space three-dimensional reconstruction with RGB-D SLAM. Power Syst. Big Data **23**(5), 30–37 (2020)
2. Chen, L., Yu, L.: Visual rationality design of interior three-dimensional space based on virtual optics. Laser J. **41**(11), 183–187 (2020)
3. Wang, Z., Peng, M., Xu, H.: Research on Information Visualization design of bronze ware in museums. HuNan BaoZhuang **35**(5), 57–61 (2020)
4. Cao, J., Ye, L.-Q.: Multi view 3D reconstruction method of virtual scene in building interior space. Comput. Simul. **37**(9), 303–306, 381 (2020)
5. Zhang, S., Zhao, W., Peng, J., et al.: Augmented reality museum display system based on object 6D pose estimation. J. Northwest Univ. (Natural Science Edition), **51**(5), 816–823 (2021)
6. Yang, K.: Design of digital 3D panoramic super-resolution reconstruction system based on virtual reality. Modern Electron. Technique **43**(10), 145–147, 152 (2020)
7. Chao, Y.: Design of museum space multi?element interactive system based on three?dimensional representation. Modern Electron. Technique **44**(12), 155–158 (2021)