Construction of English Classroom Situational Teaching Mode Based on Digital Twin Technology

Shan Gu (shangu748@gmail.com)
North China University of Science and Technology
Zhenjing Da
North China University of Science and Technology

Research Article

Keywords: Digital twin technology, English classroom, situational teaching, model construction

Posted Date: October 12th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-941109/v1

Construction of English Classroom Situational Teaching Mode Based On Digital Twin Technology

Shan Gu, Zhenjing Da*
Faculty of International Languages, Qingong College, North China University of Science and Technology, Tangshan, Hebei, 063000, China
**Email: happybabybear@126.com*

Abstract. In order to improve the effect of English classroom teaching, this paper combines the digital twin technology to construct the English classroom situational teaching mode. The system uses advanced virtual reality technology and computer image technology, and combines with video and audio synchronization processing technology to provide a new set of methods for students' language learning. The graphics rendering server of the scene interactive teaching system renders and generates 3D virtual scenes or real images in real time. Moreover, this paper constructs the functional modules of the situational teaching system according to the English classroom teaching situation, and conducts an in-depth analysis of the system implementation methods, expresses the system core algorithm flow in the form of diagrams and tables, and obtains the overall system framework. Finally, this paper evaluates the effect of the English classroom situational teaching model proposed in this paper through experimental research. From the experimental results, it can be seen that the teaching model proposed in this paper is very effective.

Keywords: Digital twin technology; English classroom; situational teaching; model construction

1. Introduction

With the rapid development of video capture, graphics and image processing technology, network technology and other technologies, it has become a reality to simulate real life scenes on a computer by constructing virtual reality scenes. Through this system, students can practice language dialogue as if they are in a real scene, so that their listening and speaking ability in foreign languages can be effectively exercised. At the same time, the system can use cartoon characters to communicate with users. In addition, it improves the interest of elementary and middle school students in learning foreign languages, and enables students to learn languages in an atmosphere of interest. The system is to use virtual technology to simulate teaching scenes that are difficult to explain, and to visualize it so that users can better learn related skills under visualization and participation [1].

With the development of science and technology, multimedia technology has risen rapidly, and its application has spread to all corners of the national economy and social life, which has brought tremendous changes to human production methods, working methods and even lifestyles. Similarly, multimedia technology has a positive effect on teaching and can provide students with the most ideal teaching environment. The situational interactive teaching system is the multimedia technology used in teaching. The system can carry out virtual situational teaching, exercise language listening and speaking ability in actual situations, and has many features and functions that are particularly valuable for education and teaching processes [2].

In our country, English classroom is the main practice place for students. In the teaching process, how teachers can effectively organize classroom practice activities and provide as many opportunities for language practice as possible plays a very important role in improving students' English proficiency, especially listening and speaking. For this reason, the situational interactive teaching system provides a good practice environment in the English classroom, simulating teaching scenes that are difficult to explain, so that students can practice English listening and speaking in a virtual setting to exercise their listening and speaking more effectively. Human-computer interaction and immediate feedback are the salient features of multimedia technology, which are not available in any other media. The multimedia computer further combines the audio-visual function of the TV with the interactive function of the computer to produce a new and colorful human-computer interaction method with both pictures and texts. This interactive method is of great significance to the teaching process, and it can effectively stimulate students' interest in learning, make students have a strong desire to learn, and form learning motivation. Interactivity is unique to multimedia computers. It is precisely because of this characteristic that multimedia computers are not only a means of teaching, but also an important factor in changing the traditional teaching mode and even teaching ideas [3].

This article combines the digital twin technology to construct the English classroom situational teaching mode, and the English classroom is the main practice place for students. In the teaching process, this paper introduces or creates vivid and concrete scenes, which is conducive to the construction of meaning for students to generate communicative motivation.

2. Related work

The situational teaching method was born in the United Kingdom. The situational teaching method requires teachers to construct purposeful situational scenes, situational cases, and situational tools for the teaching content before teaching [4]. In the situational teaching scene, the instructor needs to use simple or complex “scene props” (teaching tools) according to the course to

induce learners to think independently, and make them enter the role in the built environment to deepen their understanding of teaching content [5].

The same is true for situational MOOC videos. Among them, the instructor can combine the content displayed in text, pictures, and videos with real work scenes, life scenes, and learning scenes, which is beneficial to improve the learning efficiency of learners. Emotional experience is another feature of the situational teaching method. The application of emotional experience in the situational teaching method is to eliminate learners' resistance and fatigue caused by long-term boring learning. The sublimation of emotional experience will enable learners to connect with their own reality in the learning environment, thereby gaining the driving force for learning [6].

The literature [7] carried out an educational project that spread the concept of wave energy to high school students based on physics, and developed a virtual reality system combining software and hardware to simulate the interaction between buoys and waves for students to experience. The literature [8] used experiential learning to explain the differences in subject interest. The research results show that male students are more interested in experiential learning in physics, while female students are more interested in experiential learning in biology and chemistry. Therefore, future teaching can be considered to match the students' interests and deepen the connection with students' lives. The literature [9] conducted research on experiential learning in English subject based on digital software. The research results prove that experiential learning can increase students' interest in English learning through rich language experiments, and students will use a pleasant way to practice language and grammar, thereby improving learning efficiency.

The literature [10] conducted a research on the influence of experiential learning on teachers' classroom practice. The results show that, through personal experience, teachers have a deeper understanding of things and a more thorough understanding of themselves. When they explain to students, they often bring different learning effects and create an extraordinary experience effect for themselves and others. The literature [11] designed experiential learning courses for students and witnessed the changes of students in the whole process. Specifically, students are able to apply knowledge from other courses to the real world, while improving their writing and communication skills, as well as the ability to analyze and synthesize information. These skills are essential for success in a variety of studies. The literature [12] introduced a method for pre-service teachers to participate in experiential learning activities to enrich their knowledge and skills of teaching content. The proposed framework can be applied to the teaching of a wide range of subjects in different contexts. The literature [13] conducted project research on experiential learning in games, where students participate in games to experience and learn. The research results show that the experiential learning model can make the process of teaching and learning games more effective and full of enjoyment. The literature [14] analyzed the current situation and reasons for the application of experiential learning in the classroom. Moreover, it believed that teachers do not have enough ability to take care of everything in the classroom, teachers do not have enough skills to use learning aids, teachers cannot fully understand the spirit of the new curriculum, and teachers always abuse various experience methods.

3. Mapping rules from image to sound

The conversion from image to sound is mainly reflected in the mapping of image features to sound parameters. Image features usually refer to the significant basic features or characteristics in the image. Feature extraction is to extract the physical characteristics, geometric characteristics and other information of the target from the image, such as color, brightness, shape, area, curvature, distance and so on. The sound parameters usually include frequency, amplitude, tone, duration and stereo position. Usually, one dimension of image information is mapped to a certain dimension parameter of sound, or several dimensions of image parameters are mapped to sound output at the same time. When the users with visual impairment are in the mobile state, they often need to quickly understand the surrounding environment, especially the situation in the direction of advance, in order to make the choice of avoiding or advancing. In this case, the demand for specific target recognition is not high, and the speed requirement is put in the first place, that is to quickly reflect the information collected by the camera to the user with some simple prompt effect[15].

(1) Mapping from image to sound

In order to obtain high resolution, the image is expressed as sound in the form of time division multiplexing. Whenever the first $(k-1)$ images are processed, the new k-th image is sampled, digitized and cached as an $M \times N$ pixel matrix P^k. This process takes τ seconds. In this process, a recognizable logo is placed at the beginning of a new image or at the end of the previous image. The value $P_{ij}^{(k)}$ of each element in the pixel matrix is a value of an image gray tone G, such as[16]:

$$P^{(k)} = \left(P_{ij}^{(k)}\right), P_{ij}^{(k)} \in \left(g_1, \mathrm{L}\ , g_G\right), i = 1, \mathrm{L}\ , M, j = 1, \mathrm{L}\ , N \quad (1)$$

Therefore, when the image starts to be converted to sound, it starts from one of the N columns at the same time, and starts from the first column $j = 1$ on the far left. Figure 1 describes the conversion principle of a simple example[17]. The image in the example is a 8×8 picture with 3 kinds of gray tones $\left(M = N = 8, G = 3\right)$.

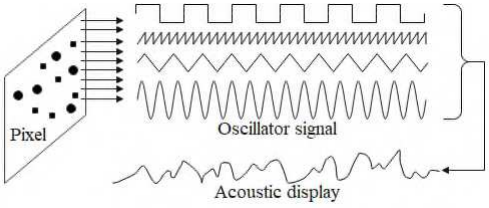

Figure 1 Mapping from image to sound

During mapping conversion, for each pixel, the vertical axis position corresponds to the frequency, the horizontal position corresponds to the time axis, and the brightness corresponds to the oscillation amplitude. It takes T seconds to convert the entire N-column pixel matrix to sound. For a given column j, each pixel in the column excites a corresponding sinusoidal oscillation in the frequency range of the sound. On the basis of different forms of quadrature sinusoidal oscillators, we assume that their frequencies are all integer multiples of some fundamental frequencies. This will ensure that the information of these sinusoidal oscillation waves is well preserved in the conversion from geometric space to Hilbert space. A pixel i at a higher longitudinal position corresponds to a higher frequency oscillating wave f_i. The greater the brightness of the pixel expressed in the form of $P_{ij}^{(k)}$, the higher the amplitude of the corresponding oscillating wave. When M oscillating signals in the same column are superimposed together, the mapped sound is defined in T/N seconds. Then, the $(j+1)$-th column is converted into sound, and this process is repeated until the Nth column of the rightmost column is converted into sound. From the beginning of the conversion, the total time is T seconds[18]. Subsequently, it still takes τ seconds to obtain a new pixel matrix $P^{(k+1)}$. At the same time, the image separation system continues to work to prepare material for the next mapping. Make sure that the time τ to acquire the image is much less than the conversion time T, that is, $\tau = T$. Once a new pixel matrix is buffered, the image-to-sound mapping conversion immediately starts from the leftmost column. Therefore, every $\tau + T$ seconds, this conversion returns to a specific column[19].

$$\omega_i = 2\pi f_i \quad (2)$$

The conversion formula can be expressed as:

$$s(t) = \sum_{i=1}^{M} P_{ij}^{(k)} \sin\left(\omega_i t + \phi_i^{(k)}\right) \quad (3)$$

Within t unit time that satisfies the condition, the j column of the pixel matrix $P^{(k)}$ is converted, such as[20]:

$$t_k + (j-1)\cdot\frac{T}{N} \le t \le t_k + j\cdot\frac{T}{N} \quad (4)$$
$$j = 1,\mathrm{L}\ ,N, k = 1,2,\mathrm{L}$$

In the formula, t_k is the moment when the first column in the pixel matrix $P^{(k)}$ starts to transform. Therefore, if the time when the first image $(k=1)$ is captured is recorded as $t=0$, then[21]:

$$t_k = (k-1)(\tau + T) + \tau \quad (5)$$

In addition, it needs to meet the monotonicity and separability requirements:

$$\omega_i > \omega_{i-1}, i = 1,2,\mathrm{O}\ ,M \quad (6)$$

As mentioned earlier, a synchronized identifiable identification confirmation is completed at the same time. $\phi_i^{(k)}$ n is a random constant in the image-to-sound conversion process, but it may change when the synchronization confirmation is generated.

For a simple image environment, the corresponding sound conversion is easy to complete. For example, a bright line in a dark background, extending from the lower left corner to the upper right corner, will easily be mapped as a single sound with gradually increasing pitch until the confirmation mark for the image is marked. Similarly, a bright rectangle will be mapped to a certain bandwidth of sound, with duration corresponding to its width and bandwidth corresponding to its height. The convenience of mapping to simple shapes is very important, because the good expression of simple shapes means that there will be no insurmountable obstacles for more complex image processing. In reality, images are often more complex and difficult to express with simple voice, but after long-term training and adaptation, the visually impaired can still reflect the rough information of the environment image naturally and quickly.

One of the major defects of the above image to sound conversion is that it can't make good use of the difference in sound acquisition time between the left ear and the right ear. In fact, this slight difference in auditory time is an important basis for human beings to judge the origin of students. Although the CCD camera still provides many application functions, the delay function has not been well applied, because the current timing method can meet the basic requirements without more hardware investment[22].

The $M \times N$ brightness value is transmitted in T seconds, and each brightness value is selected from the possible set of G. The transmission speed per second reaches Ibits, which is expressed by the formula:

$$I = \frac{MN \cdot^2 \log(G)}{T} \quad (7)$$

In order to prevent the loss of information during the conversion process, we need to strictly limit the values of M, N, G, and T, and even we need to take the human ear into consideration.

The brightness value of G is allowed to be superimposed in the process of corresponding to the amplitude of the signal. The amplitude conveys the information of the image, and these periodic signals themselves and their superposition together reflect the given amplitude, even if their number is large or even unlimited. When monitoring the information of the Fourier parameters, the brightness range of G can be reproduced within the range that the human ear can recognize. For the convenience of analysis, we will consider a very simple image sequence: all images are black, and only a bright pixel (i', j') exists in the k-th image, such as:

$$P_{ij}^{(k)} = 0, \quad \forall i, j \setminus \{i = i', j = j', k = k'\} \quad (8)$$

By lowering the pitch:

$$s(t) = \begin{cases} P_{ij}^{(k)} \sin(\omega_i t) & t^{(1)} \le t < t^{(2)} \\ 0 & otherwise \end{cases} \quad (9)$$

Among them,

$$t^{(2)} - t^{(1)} = \frac{T}{N} \quad (10)$$

When ignoring the influence of horizontal synchronization confirmation, the Fourier transform can significantly eliminate the crosstalk between multiple sinusoidal signals that occurs within T/N seconds, and the isolation frequency step Δf between the signals can be set to $2/(T/N)$ Hz. In this way, the oscillation values of two adjacent pixels in the vertical position can be well represented. For the equidistant frequency stepwise setting $\Delta f = B/(M-1)$, the bandwidth is B Hz. At this time, the crosstalk limit becomes:

$$(M-1)N \le \frac{B \cdot T}{2} \quad (11)$$

$B = 5kHZ, T = 1s$, $M = N, M \le 50$.

The images used in this experiment are assumed to be "frozen" state, which can be converted at least in the t time range without losing the current image content due to the new scene transformation. However, the human brain also has limitations in the time domain of receiving and understanding information. A good method is to get a good understanding of the previous information, And when the subsequent image comes, it is only sensitive to the changed part of the image. For normal people, although most of the time for visual response time is only a few percent of a second, but enough to let him remember important details. For example, when a door is opened or a coffee cup is picked up, the bystander will immediately know. If there is a longer time, people will consider more environmental information. Of course, the brain still tends to learn the corresponding information in a few seconds, such as speaking or moving. Therefore, our conversion time unit t is about 1 second, and the corresponding time is much less than 1 second, which can not only avoid the generation of blurred images, but also have enough time to transfer image information to the sound channel. The human auditory bandwidth is about 20kHz, but the available bandwidth is usually no more than 5-6khz.

In addition, it should be noted that the quality of the conversion also depends on the situation of the image. The situation corresponding to the above formula is the situation where there are some bright pigments in each column assuming a completely black background. The greatest crosstalk must occur between the closest points, but it is actually found that crosstalk will also occur between two bright spots with a large distance. Therefore, it is easier to find small bright spots on a dark background than dark spots on a bright background.

In order to meet the higher-level needs of users, another part of the experiment is to add part of the content of pattern recognition in a static state to help users understand the shape characteristics of the target object. For different situations in a changeable environment, for convenient and reliable recognition, image features must be invariant to their translation, rotation, and scale transformation. That is, no matter what kind of translation, rotation, and zooming in or zooming out of the recognized image, it can still be correctly recognized.

For the continuous gray function $f(x,y)$, the $(p+q)$-order two-dimensional origin moment M_{pq} is defined as:

$$M_{pq} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x^p y^q f(x,y)dxdy, p,q = 0,1,2 \quad (12)$$

$f(x,y)$ is a piecewise continuous bounded function, and has a non-zero value in a finite area on the x and y planes. According to the uniqueness theorem, each moment is uniquely determined by $f(x,y)$. Correspondingly, $f(x,y)$ is uniquely determined by its moments. In addition, the $(p+q)$-th order central moment μ_{pq} of $f(x,y)$ can be defined.

$$\mu_{pq} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} (x-\bar{x})^p (y-\bar{y})^q f(x,y)dxdy, p,q = 0,1,2 \quad (13)$$

For a $M \times N$ discrete digital image $f(i,j)$, its pq-order geometric moment and central moment are respectively:

$$m_{pq} = \sum_{i=1}^{M}\sum_{j=1}^{N} i^p j^q f(i,j) \quad (14)$$

$$\mu_{pq} = \sum_{i=1}^{M}\sum_{j=1}^{N} (i-\bar{x})^p (j-\bar{y})^q f(i,j), p,q = 0,1,2 \quad (15)$$

In the formula,

$$\bar{x} = m_{10}/m_{00}, \bar{y} = m_{01}/m_{00} \quad (16)$$

is the center of gravity of the image, and $m_{00} = \mu_{00}$ can be obtained. For grayscale images, μ_{00} is equivalent to the quality of the image. However, for a binary image, μ_{00} is equivalent to the area of the image.

The geometric moment and central moment of the image can describe the shape of the image, and the central moment has nothing to do with the translation of the image. The zero-order central moment μ_{00} is used to normalize the other central moments, and the normalized central moment of the image can be obtained[23]:

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^{\gamma}}, \gamma = (p+q+2)/2 \quad (17)$$

4. Construction of English classroom situational teaching mode based on digital twin technology

This paper uses digital twin technology to construct the English classroom scene teaching mode. Defocus 3D measurement directly uses the relationship between object depth, camera parameters and image ambiguity to measure object depth. Figure 2 below shows the principle of 3D measurement.

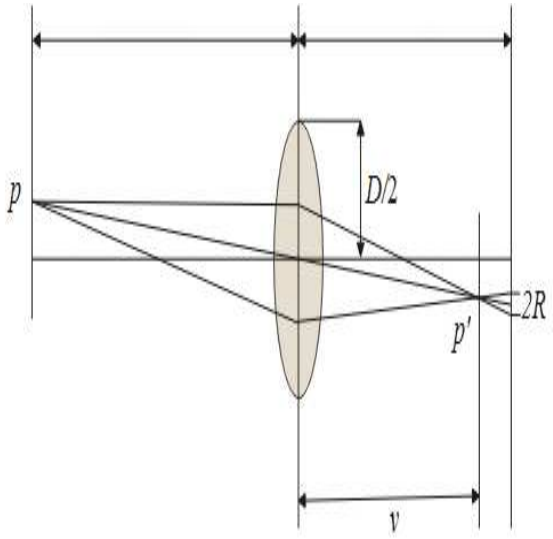

Figure 2 Schematic diagram of defocus 3D measurement

Because of the optical defocus, the imaging system becomes the result of the scene and the imaging system. Fuzzy edges in the image may be the result of the sharp edge of the scene being defocused and blurred by the imaging system, or the result of the soft edge of the scene being sharply focused and imaged by the imaging system. Therefore, it is necessary to capture at least two images with different defocus degrees of the scene at the same time, and then solve the defocus value u to eliminate the influence of the unknown light intensity distribution of the scene on the defocus value. By changing the distance s from the image detector to the lens, two images with different degrees of triangulation are obtained. Figure 3 shows the imaging principle of a telecentric lens.

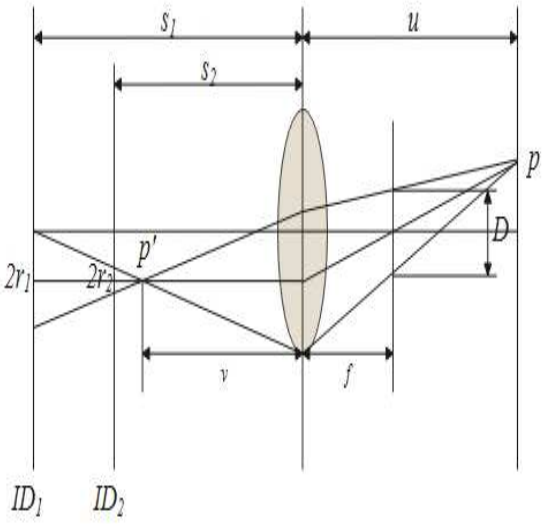

Figure 3 Imaging principle of telecentric lens

The English scene interactive teaching system needs to load three-dimensional virtual scenes or real-life images. The system host performs digital real-time synthesis of the performance image and the three-dimensional virtual scene or real image after the matting process to achieve the integration of the two. Finally, it outputs the synthesized video images to a streaming media publishing system or to a large screen for on-site teaching. At the same time, the learning content in the classroom can also use the local collection unit of the situational interactive teaching system, and it can be stored in the system in real time for later use. The system structure diagram is shown in Figure 4.

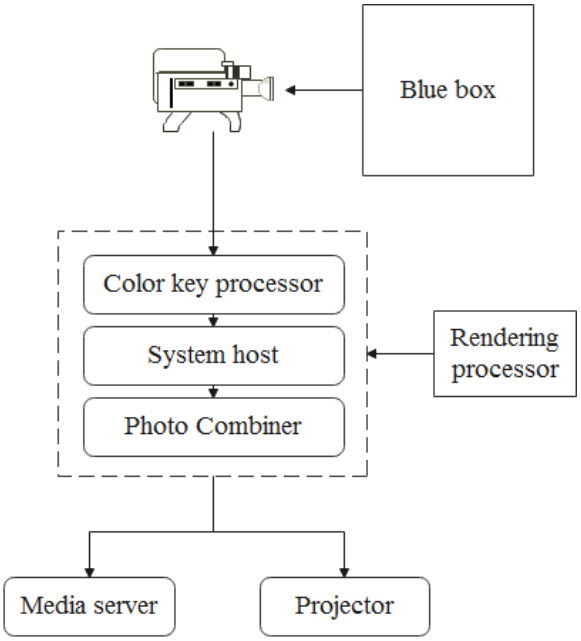

Figure 4 System structure diagram

The main function of the video capture module is to use the camera to collect real-time video, real-time sound collection, or select the recorded video to add to the selected scene to perform video synthesis. This module has some controls to control the playback, pause, and stop of the loaded video. When we click the keying button, we can click the background color that needs to be keyed out in the capture video frame. At the same time, we can drag the slider or manually enter the parameters to adjust the threshold of the keying, and crop the input video. This module is mainly used to collect and process video and audio. The design of the video acquisition module is shown in Figure 5.

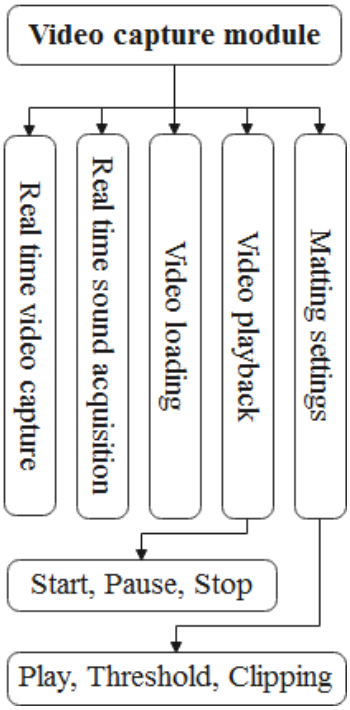

Figure 5 Design of video capture module

The function of the camera animation is to set the camera animation in the virtual scene, push and pull the lens, manually set the key frame, and create the camera animation to make the synthesized video more colorful. The design of the control module is shown in Figure 6.

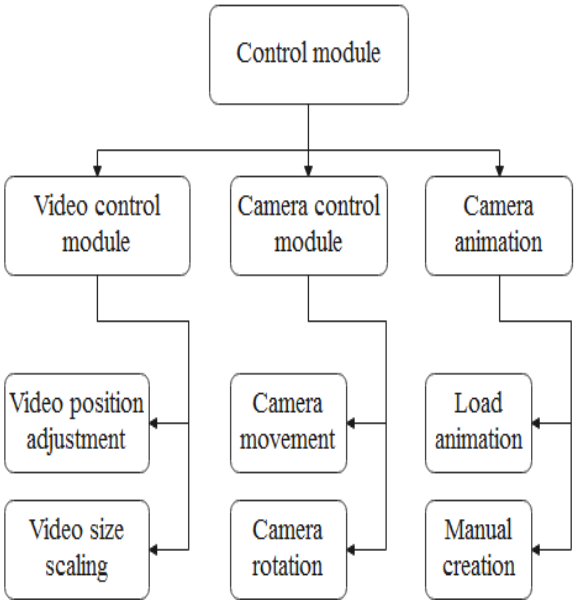

Figure 6 Design of control module

The English scene interactive teaching system calls the standard virtual scene model sequence file established by the digital twin technology in the background, and performs real-time three-dimensional filling and rendering on the Open-GL graphics platform according to the camera's parameter changes. The design of the scene management module is shown in Figure 7.

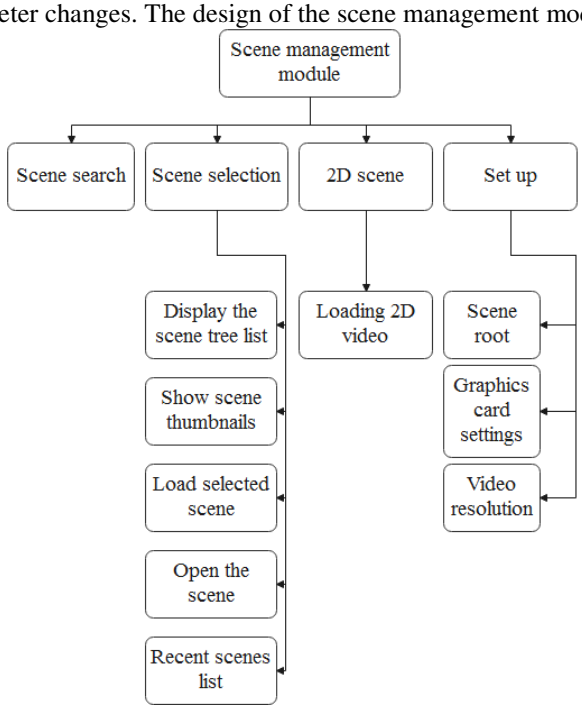

Figure 7 Design of the scene management module

5. System test

This paper constructs an English classroom situational teaching model based on digital twin technology, and constructs a corresponding system. On this basis, the performance of the system is verified. Therefore, this paper verifies the digital transformation effect of the system constructed in this paper, and evaluates the quality of English classroom situational teaching on this basis. First of all, this paper carries out the evaluation of the digital processing effect of the English teaching resources of the system constructed in this paper, as shown in Table 1 and Figure 8.

Table 1 Statistical table of the digital effect of the English classroom situational teaching model based on the digital twin technology

NO.	Situational evaluation	NO.	Situational evaluation	NO.	Situational evaluation
1	78.47	28	93.92	55	79.16
2	90.65	29	90.11	56	84.87
3	87.34	30	77.95	57	88.73
4	86.82	31	87.44	58	84.68
5	89.52	32	77.76	59	84.58
6	85.58	33	77.63	60	81.84
7	86.32	34	83.17	61	92.03

8	86.45	35	87.37	62	82.95
9	92.88	36	78.50	63	82.76
10	92.09	37	92.58	64	85.64
11	87.56	38	76.61	65	82.42
12	77.28	39	87.55	66	86.86
13	92.80	40	91.08	67	80.34
14	87.11	41	78.20	68	88.42
15	85.69	42	80.88	69	81.59
16	90.33	43	85.75	70	84.01
17	88.64	44	89.77	71	77.39
18	86.00	45	79.77	72	91.54
19	92.87	46	84.85	73	89.53
20	81.98	47	77.96	74	93.95
21	92.09	48	79.84	75	90.42
22	80.67	49	92.19	76	78.11
23	93.95	50	90.92	77	77.06
24	80.77	51	93.18	78	86.98
25	82.99	52	77.62	79	77.42
26	79.74	53	84.30	80	80.28
27	79.16	54	81.08	81	91.92

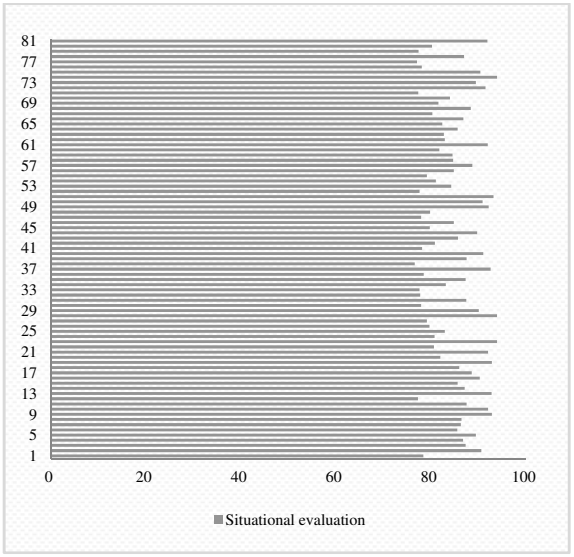

Figure 8 Statistical diagram of the digital effect of the English classroom situational teaching model based on the digital twin technology

Through the above analysis, it can be seen that the English classroom situational teaching model based on digital twin technology constructed in this paper can effectively transform the traditional teaching model into a digital three-dimensional teaching model to create a spatial teaching experience for students and make students have a certain sense of immersion. On this basis, this paper evaluates the teaching effect of the English classroom situational teaching model of this paper by means of scoring. The results are shown in Table 2 and Figure 9.

Table 2 Statistical table of the teaching effect of the English classroom situational teaching model based on the digital twin technology

NO.	Teaching Evaluation	NO.	Teaching Evaluation	NO.	Teaching Evaluation
1	91.73	28	89.82	55	86.72
2	81.85	29	90.07	56	85.86
3	85.40	30	76.95	57	91.36
4	91.07	31	91.01	58	91.23
5	81.68	32	89.40	59	85.67
6	81.63	33	93.72	60	76.00
7	81.62	34	82.32	61	91.33
8	89.98	35	90.53	62	77.68
9	88.78	36	87.32	63	81.11
10	86.36	37	76.91	64	89.60
11	87.43	38	77.39	65	84.07
12	87.41	39	83.70	66	88.55

13	86.22	40	87.48	67	78.75
14	91.01	41	85.37	68	80.25
15	87.83	42	83.12	69	79.20
16	77.54	43	77.52	70	78.89
17	89.72	44	84.80	71	78.13
18	78.86	45	88.94	72	80.45
19	86.81	46	80.30	73	82.24
20	80.53	47	92.15	74	81.48
21	91.53	48	89.10	75	79.87
22	80.91	49	76.44	76	83.87
23	87.65	50	76.06	77	86.51
24	90.91	51	86.13	78	76.08
25	77.55	52	92.95	79	77.10
26	89.11	53	92.16	80	92.15
27	92.30	54	91.92	81	89.56

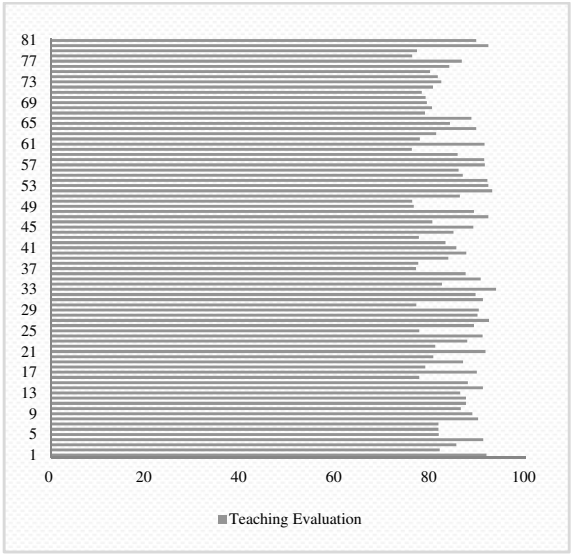

Figure 9 Statistical diagram of the teaching effect of the English classroom situational teaching model based on the digital twin technology

From the above analysis results, the English classroom situational teaching model based on digital twin technology constructed in this paper has certain effects and can effectively improve the quality of English teaching.

6. Conclusion

This article combines the digital twin technology to construct the English classroom situational teaching mode, and the English classroom is the main practice place for students. Introducing or creating vivid and concrete scenes in the teaching process is conducive to the construction of meaning for students to generate communicative motivation. Moreover, the virtual scene provided by the situational interactive teaching system can provide an intuitive and vivid image, and the performance in the virtual scene reproduced by a large screen or projection can allow students to feel the scene and produce imagination and association through vision and hearing, and it can stimulate students' interest in learning. At the same time, the situational interactive teaching system uses advanced virtual reality technology and computer image technology, and it combines video and audio synchronization processing technology to provide a set of new methods for students' language learning. In addition, the graphics rendering server of the scene interactive teaching system renders and generates three-dimensional virtual scenes or real-life images in real time. Finally, this paper combined with experimental analysis to further prove the effectiveness of the method proposed in this paper.

7. DECLARATIONS

Funding: Not applicable

Conflicts of interest: The author has no conflicts of interest

Availability of data and material: Not applicable

Code availability: Not applicable

References

[1]. Dong H , Gong S , Liu C , et al. Large margin relative distance learning for person re-identification[J], Iet Computer Vision, 2017, 11(6):455-462.
[2]. Delgaty, Laura. Twelve tips for academic role and institutional change in distance learning[J], Medical Teacher, 2015, 37(1):41-46.
[3]. Stefanovic M , Tadic D , Nestic S , et al. An Assessment of Distance Learning Laboratory Objectives for Control Engineering Education[J], Computer applications in engineering education, 2015, 23(2):191-202.
[4]. BABU P. REMESH. Developing Open and Distance Learning Programme in Labour and Development: Results of a Needs Assessment Study[J], journal of natural history, 2015, 196(29):265-291.
[5]. Wu P , Low S P , Liu J Y , et al. Critical Success Factors in Distance Learning Construction Programs at Central Queensland University: Students' Perspective[J], Journal of Professional Issues in Engineering Education and Practice, 2015, 141(1):05014003.
[6]. Willis E A , Szabo-Reed A N , Ptomey L T , et al. Distance learning strategies for weight management utilizing social media: A comparison of phone conference call versus social media platform. Rationale and design for a randomized study[J], Contemporary Clinical Trials, 2016, 47(6):282-288.
[7]. Ye H J , Zhan D C , Jiang Y . Fast generalization rates for distance metric learning: Improved theoretical analysis for smooth strongly convex distance metric learning[J], Machine Learning, 2019, 108(2):267-295.
[8]. Kim H K, Park J, Choi Y, et al. Virtual reality sickness questionnaire (VRSQ): Motion sickness measurement index in a virtual reality environment[J], Applied ergonomics, 2018, 69: 66-73.
[9]. Yung R, Khoo-Lattimore C. New realities: a systematic literature review on virtual reality and augmented reality in tourism research[J], Current Issues in Tourism, 2019, 22(17): 2056-2081.
[10]. Makransky G, Terkildsen T S, Mayer R E. Adding immersive virtual reality to a science lab simulation causes more presence but less learning[J], Learning and Instruction, 2019, 60: 225-236.
[11]. Yiannakopoulou E, Nikiteas N, Perrea D, et al. Virtual reality simulators and training in laparoscopic surgery[J], International Journal of Surgery, 2015, 13: 60-64.
[12]. Jensen L, Konradsen F. A review of the use of virtual reality head-mounted displays in education and training[J], Education and Information Technologies, 2018, 23(4): 1515-1529.
[13]. Howard M C. A meta-analysis and systematic literature review of virtual reality rehabilitation programs[J], Computers in Human Behavior, 2017, 70: 317-327.
[14]. Serino M, Cordrey K, McLaughlin L, et al. Pokémon Go and augmented virtual reality games: a cautionary commentary for parents and pediatricians[J], Current opinion in pediatrics, 2016, 28(5): 673-677.
[15]. Smith M J, Ginger E J, Wright K, et al. Virtual reality job interview training in adults with autism spectrum disorder[J], Journal of autism and developmental disorders, 2014, 44(10): 2450-2463.
[16]. Farshid M, Paschen J, Eriksson T, et al. Go boldly!: Explore augmented reality (AR), virtual reality (VR), and mixed reality (MR) for business[J], Business Horizons, 2018, 61(5): 657-663.
[17]. Alhalabi W. Virtual reality systems enhance students' achievements in engineering education[J], Behaviour & Information Technology, 2016, 35(11): 919-925.
[18]. Muhanna M A. Virtual reality and the CAVE: Taxonomy, interaction challenges and research directions[J], Journal of King Saud University-Computer and Information Sciences, 2015, 27(3): 344-361.
[19]. Valmaggia L R, Latif L, Kempton M J, et al. Virtual reality in the psychological treatment for mental health problems: an systematic review of recent evidence[J], Psychiatry research, 2016, 236: 189-195.
[20]. Slater M. Immersion and the illusion of presence in virtual reality[J], British Journal of Psychology, 2018, 109(3): 431-433.
[21]. Dobkin B H. A rehabilitation-internet-of-things in the home to augment motor skills and exercise training[J]. Neurorehabilitation and neural repair, 2017, 31(3): 217-227.
[22]. Yao J, Ansari N. Caching in energy harvesting aided Internet of Things: A game-theoretic approach[J]. IEEE Internet of Things Journal, 2018, 6(2): 3194-3201.
[23]. Siegel J E, Kumar S, Sarma S E. The future internet of things: Secure, efficient, and model-based[J]. IEEE Internet of Things Journal, 2017, 5(4): 2386-2398.

Shan Gu Shan Gu is a teacher graduated from China University of Mining and Technology (Beijing) with a master degree working on foreign language teaching in Faculty of International Languages, Qingong College, North China University of Science and Technology, China, Her research interests include E-C and C-E translation theories and practice, English learning and teaching and , more than 7 papers published and 1 book published.

Zhenjing Da works on Faculty of International Languages in Qinggong College, North China University of Science and Technology. His research interests include syntax, translation skills and English teaching. More than 8 papers published and 1 book published.