

Retraction

Retracted: Embedded VR Video Image Control System in the Ideological and Political Teaching System Based on Artificial Intelligence

Mobile Information Systems

Received 8 August 2023; Accepted 8 August 2023; Published 9 August 2023

Copyright © 2023 Mobile Information Systems. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

- [1] F. Jia, S. Xu, and J. Huo, "Embedded VR Video Image Control System in the Ideological and Political Teaching System Based on Artificial Intelligence," *Mobile Information Systems*, vol. 2021, Article ID 7839821, 12 pages, 2021.

Research Article

Embedded VR Video Image Control System in the Ideological and Political Teaching System Based on Artificial Intelligence

Fengzhen Jia,¹ Shiqiang Xu ,² and Jiaofei Huo³

¹Marxist Academy, Xijing University, Xi'an 710123, Shaanxi, China

²School of Sciences, Chang'an University, Xi'an 710064, Shaanxi, China

³Mechanical and Electrical, Xijing University, Xi'an 710123, Shaanxi, China

Correspondence should be addressed to Shiqiang Xu; xsq005596@chd.edu.cn

Received 9 August 2021; Revised 16 September 2021; Accepted 23 September 2021; Published 15 November 2021

Academic Editor: Sang-Bing Tsai

Copyright © 2021 Fengzhen Jia et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the increasing development of multimedia teaching, the combination of virtual reality (VR) and video image control has very attractive development prospects in ideological and political teaching, for example, the use of virtual technology in games and so on. However, most virtual reality environments are currently built, and the functional development of artificial intelligence multimedia teaching systems is not comprehensive. An artificial intelligence VR video image control system is constructed for the multimedia teaching system. This article analyzes the development of artificial intelligence multimedia teaching systems and compares the detection performance and efficiency of traditional methods and artificial intelligence multimedia VR ideological and political teaching. Research shows that, in the use of VR to control the images of ideological and political teaching, the average accuracy of these ten video images is 75.68%. This shows that the video image classification and detection algorithm model based on artificial intelligence in this paper can extract deeper and more abstract features to classify the target. The artificial intelligence VR video image control algorithm constructed in this paper can reduce the maximum failure rate by 49.16%, 61.02%, and 66.94%, respectively. Compared with the traditional algorithm, the artificial intelligence VR video image control algorithm constructed in this paper can reduce the storage access delay time of 10 different video images by an average of 15.93%, can obtain about 9.37% performance optimization, and can reduce the video image control time by 7.28% and 10.63%, respectively. For pictures, the artificial intelligence VR video image control system in this article can increase the performance by up to 28.49%.

1. Introduction

At present, related researches on cognitive education focus on plane and three-dimensional graphics, and there are little researches on graphics processing in virtual environments. In the virtual reality learning environment, the sensory experience of the learner is richer and more immersive than the plane and three-dimensional space, and the cognitive load may also change. As a result, the conclusions drawn in the plane and three-dimensional learning environment may not be applicable to the virtual reality environment, which is the first influencing factor concerned in cognitive education.

As an important carrier of virtual reality experience, virtual reality simulation platform has gradually matured and popularized and is widely used in industry, science and

education, medical treatment, entertainment, and other fields, as well as manufacturing and other fields [1]. Somatosensory design refers to the design of human somatosensory, and the somatosensory design for the simulation platform is to simulate the movement state felt in the user experience scene. Virtual reality in the learning environment can fully “occupy” students’ visual, auditory, tactile, and other perception channels. This multisensory learning method can not only improve learning efficiency but also cause cognitive education overload and hinder learning progress. Stepan proposed that cognitive education is caused by learning tasks, learners, and the interaction between the two [2]. Guo believes that cognitive education has many influencing factors, such as the number of items, presentation time, and text clues, which will have an impact

on cognitive education in virtual environments [3]. However, some studies have found that text clues increase cognitive education. Berghea investigated the impact of text cues on the cognitive education of graphics processing in a virtual environment [4]. Cognitive education research is inseparable from the investigation of individual working memory. Konrad proposed a knowledge management system based on a dimensional model [5]. Lee used the DKHM model to build an evolutionary knowledge tree to verify the hot issues of topic documents and designed a method for automatically constructing a knowledge tree based on text classification for the creation and maintenance of the knowledge tree and constructed a system knowledge modeling framework. The knowledge structure is further disassembled [6]. Wei-Tse proposes an intelligent algorithm based on binary tree and knowledge tree reasoning to build a new class arrangement system. Its flexibility and scalability greatly improve the efficiency of class arrangement [7].

There are still certain limitations in the action design of the existing VR simulation platform. Somatosensory designers often spend more energy in the action editing process, and the design quality depends more on the designer's technical level. Aline uses these two types of indicators for cognitive education evaluation. The task performance method measures the objective cognitive education value, and the subjective measurement method measures the cognitive education subjectively felt by the subjects. At the same time, it examines the sensitivity of the evaluation indicators; that is, the indicators reflect the project whether the number, presentation time, and text cues bring about significant changes in cognitive education and whether they are consistent [8]. Jain studied the influencing factors and mechanisms of visual information processing in the virtual environment from the perspective of cognitive education, which has theoretical significance and applications for enhancing the credibility and authenticity of the construction of virtual learning environments and optimizing the control of virtual perception cognitive education and learning costs [9, 10]. In the immersive virtual learning environment, based on the dual-task experimental paradigm, using virtual space three-dimensional graphics as experimental materials, focusing on the impact of the number of memorization items, presentation time, and text cues on virtual cognitive education, using task performance methods and subjectivity, the measurement method is used to evaluate cognitive education. Jerald proposed that, in an immersive virtual learning environment, a large number of memorization items and a short presentation time will increase cognitive education. Text clues increase the amount of information processing and cognitive education. The number of memorization items, presentation time, and text clues will also increase. It has an interactive impact on virtual cognitive education [11]. However, the design method using analog input in the above research will cause errors in the simulator input action and the response of the VR panoramic film [12, 13]. At the same time, the postediting process is cumbersome, and the technical threshold of the somatosensory designer is high, and then the VR

somatosensory design appears. Some problems are poor quality controllability, low operability, and high workload.

In [14], the author introduced detailed information about the framework and development methods related to VR-based training simulators used in advanced cardiac life support medical schemes. In [15], the authors found that virtual reality can be used to assess neurocognitive dysfunction and study related clinical symptoms. In addition, preliminary findings suggest that it can be used for cognitive rehabilitation of mental illness, social skills training interventions, and virtual reality assisted therapies. In [16], the authors used immersive virtual reality to trigger an illusion of full-body ownership, putting the perpetrator inside the victim of domestic abuse. According to research at home and abroad, the task of completing image processing and control in video images is a new demand in the field of interaction.

Because the objects in the video image have many uncertain deformations due to the demand of the full perspective, this brings about great difficulty to the recognition task [17]. However, evaluating performance in multimedia teaching is not easy, as it involves artificial decisions that are inaccurate, fuzzy, and uncertain [18, 19]. In [20], the authors designed conceptualization using UML tools. In [21], the authors reported the setup and evaluation of robust speech recognition system components designed to generate subtitles for real-life media sets. The findings show that immersive multimedia learning conducted with peer support groups reported that immersive multimedia learning performed significantly better in all spoken expressions used for reading and speaking. In [22], the authors introduced the concept of Multimedia Software as a Service (MSaaS) and proposed an experience quality paradigm based on general priority, which is used in wireless large data volume multimedia communication and has significant energy-saving effect [23]. It has the potential of future multimedia Internet of Things devices and systems [24]. The multimedia evaluation benchmarking program organizes an annual scientific evaluation task cycle in the field.

Based on the detailed analysis of VR technology, this paper constructs an embedded VR video image control system for multimedia teaching systems and uses the systemic teaching design ideas to analyze the development of multimedia teaching. This paper introduces the construction of video image datasets, analyzes the influence of network parameters on the final detection accuracy, and compares the detection performance and efficiency of traditional methods and the design method of this paper. The embedded VR video image control algorithm constructed in this paper can reduce the maximum failure rate by 49.16%, 61.02%, and 66.94%, respectively. Compared with the traditional algorithm, the embedded VR video image control algorithm constructed in this paper can reduce the storage access delay time of 10 different video images on average by 15.93% and can obtain about 9.37% performance optimization and reduce video image control time by 7.28% and 10.63%, respectively. For picture, the embedded VR video image control system in this article can improve performance by up to 28.49%.

2. Method

2.1. Video Image Classification Detection Algorithm Based on Image Control

2.1.1. Selective Search.

- (1) Color similarity: use the Manhattan distance to obtain the color histogram of each channel in each of the 25 directions. In this way, each region can be represented by a 75-dimensional vector $C_i = \{c_i^1, \dots, c_i^{75}\}$. The calculation of the color similarity between regions is shown in the following formula:

$$S_{\text{color}}(r_i, r_j) = \sum_{k=1}^n \min(c_i^k, c_j^k), \quad (1)$$

where r_i is the initial region; $r_i = \{r_1, r_2, \dots, r_i, r_j, \dots, r_n\}$, in which (r_i, r_j) is the neighbouring region pair.

- (2) Texture similarity: the texture feature uses the SIFT feature here. For each color channel in each region, 8 directions are calculated, and the variance $\sigma = 1$ is Gaussian differential. Each channel takes 10 histograms in each direction and finally obtains a 240-dimensional calculation of the feature vector, $T_i = \{t_i^1, \dots, t_i^{240}\}$, and the texture similarity between regions is consistent with the color similarity. The Manhattan distance is also used; see the following formula:

$$S_{\text{texture}}(r_i, r_j) = \sum_{k=1}^n \min(t_i^k, t_j^k). \quad (2)$$

- (3) Size similarity: the size here refers to the total number of pixels in the area. The size similarity is used in order to merge the small area lines as much as possible. The calculation of dimensional similarity is shown in the following formula:

$$S_{\text{size}}(r_i, r_j) = 1 - \frac{\text{size}(r_i) + \text{size}(r_j)}{\text{size}(im)}, \quad (3)$$

where im is the entire color area. Although the selection search algorithm can find the candidate region of the target well, it makes use of the traditional features and adopts additional feature extraction process, which becomes a part of the time required by the target detection algorithm.

2.1.2. Target Frame Position Return. In the training process of target frame regression network, N data pairs are input. In the data pair $\{(P^i, G^i)\}_{i=1, \dots, N}$, here $P^i = (P_x^i, P_y^i, P_w^i, P_h^i)$ represents the center point coordinates of the prediction target frame and the length and width of the target frame, which also corresponds to each truth target frame $G^i = (G_x^i, G_y^i, G_w^i, G_h^i)$. The purpose of learning is to find the mapping relationship between the candidate target box and the real target box. At the same time, four equations are

defined to express this transformation relationship: $t_x(P), t_y(P)$ is the scale invariant transformation of the central coordinates of the proposed target frame P , and $t_w(P), t_h(P)$ is the logarithmic space transformation of the length and width of the proposed target frame P . In this way, we can calculate the candidate target frame P close to the real target frame \hat{G} through the following formula:

$$\begin{cases} \hat{G}_x = P_w t_x(P) + P_x, \\ \hat{G}_y = P_h t_y(P) + P_y, \\ \hat{G}_w = P_w \exp(t_w(P)), \\ \hat{G}_h = P_h \exp(t_h(P)). \end{cases} \quad (4)$$

In each of the above equations, $t_w(P)$ is represented as a linear equation of the candidate region feature vector φ ; see $t_*(P) = W_*^T \varphi$. That is, the objective function we want to learn represents the position mapping relationship between the predicted target frame and the real target frame.

2.1.3. Regional Classification Detection. During the training process, each target area participating in the training has two labels: one is the category information u , and the other is the real target frame v of the target. Here we use the loss function L of the multitasking target to uniformly train the classification task and the target frame position regression task, so the advantage is that the network parameters can be optimized for both targets at the same time. The loss function is designed as

$$L(p, u, t^u, v) = L_{\text{cls}}(p, u) + \lambda[u \geq 1]L_{\text{loc}}(t^u, v), \quad (5)$$

where $L_{\text{cls}}(p, u) = -\log(p_u)$ represents the logarithmic loss of class prediction probability, $v = (v_x^u, v_y^u, v_w^u, v_h^u)$ represents the real target frame position of class u , and $t^u = (t_x^u, t_y^u, t_w^u, t_h^u)$ represents the position of network prediction $[u \geq 1]$, which means that when $u \geq 1$, it is 1 and 0, and only the position loss of foreground category is calculated.

2.2. Embedded VR Video Image Processing and Control Algorithm

2.2.1. Automatic White Balance Technology. Assume that each color adjustment is an independent gain adjustment process. The gain coefficient is expressed by the following formula:

$$\begin{aligned} La &= K_L L, \\ Ma &= K_M M, \\ Sa &= K_S S. \end{aligned} \quad (6)$$

In the above formula, $L, M,$ and S represent the original initial values of the three colors, and $K_L, K_M,$ and K_S represent gain adjustment coefficients that increase the original signal to the three color adjustment values $La, Ma,$ and Sa . The above formula becomes the following formula:

$$\begin{aligned}
Ra &= K_R R, \\
Ga &= K_G G, \\
Ba &= K_B B.
\end{aligned} \tag{7}$$

The corresponding R/G/B tricolor gain adjustment coefficients are obtained through different white balance algorithms, and the adjusted adjustment values are calculated according to the formula to realize the adjustment of the image collection white balance.

2.2.2. Gray World Algorithm. The calculation method of the white balance algorithm based on the gray world is as follows: statistically obtain the corresponding average values R_{av} , G_{av} , and B_{av} of the three RGB component values, and calculate the gain values GainR, GainG, and GainB of each channel by the following formula:

$$\begin{cases}
\text{GainR} = \frac{\max(R_{av}, G_{av}, B_{av})}{R_{av}}, \\
\text{GainG} = \frac{\max(R_{av}, G_{av}, B_{av})}{G_{av}}, \\
\text{GainB} = \frac{\max(R_{av}, G_{av}, B_{av})}{B_{av}},
\end{cases} \tag{8}$$

where R' , G' , and B' obtained after formula calculation and processing are the three-channel component values of the final pixel:

$$\begin{aligned}
R' &= \begin{cases} R * \text{GainR} (R * \text{GainR} < 255) \\ 255 (R * \text{GainR} > 255), \end{cases} \\
G' &= \begin{cases} G * \text{GainG} (G * \text{GainG} < 255) \\ 255 (G * \text{GainG} > 255), \end{cases} \\
B' &= \begin{cases} B * \text{GainB} (B * \text{GainB} < 255) \\ 255 (B * \text{GainB} > 255). \end{cases}
\end{aligned} \tag{9}$$

This calculation method is relatively simple and suitable for application in embedded systems. It reduces the occupied time of limited memory and CPU (Central Processing Unit) during the calculation process and is more suitable for use in scenes with richer colors.

2.2.3. GW and PR Orthogonal Combination Algorithm. The implementation steps of the gray world algorithm and the perfect reflection (PR) orthogonal combination algorithm are as follows:

First, the mean values $(R_{ave}, G_{ave}, B_{ave})$ and $(R_{max}, G_{max}, B_{max})$ of the three color components of the image are maximized, and then the quadratic equation formula of the R component is established as follows:

$$\begin{aligned}
u^r R_{ave}^2 + v^r R_{ave} &= K_{ave}, \\
u^r R_{max}^2 + v^r R_{max} &= K_{max} \\
K_{ave} &= \frac{R_{ave} + G_{ave} + B_{ave}}{3}, \\
K_{max} &= \frac{R_{max} + G_{max} + B_{max}}{3}.
\end{aligned} \tag{10}$$

The u and v values in the equation are calculated according to the above formula, and then the pixel value of the R component after the calculation and adjustment of the white balance parameter is $R_A = u^r R_B^2 + v^r R_B$. R in the formula is the pixel component value of the original image, and R is the pixel value obtained by the algorithm calculation.

2.3. Multimedia Teaching Theory Based on Virtual Reality (VR) Technology

2.3.1. VRML Technology Concept. Based on comprehensive consideration of many VR software platforms, this paper chooses VRML as the design technology platform. It is mainly based on the fact that VRML itself appears as the core of the second-generation web browsing technology and is closely integrated with HTML, JAVA, and multimedia information. VRML is a description language. Describe the real world and objective world in an intuitive way to form an easy-to-understand virtual scene in line with the real world.

2.3.2. Image Control Theory. Image control theory emphasizes practical experience and believes that the basic level of knowledge acquisition is practical experience. Virtual experiment of multimedia system based on VR technology is to use VR technology to build a virtual learning scene and simulate the entire process of actual teaching operations on the computer screen, so as to directly provide students with an atmosphere of doing. Although this process is only done on a computer, the actual experimental teaching requirements are followed during this operation. The functions and characteristics of the virtual experimental equipment fully reflect the performance and form of the experimental equipment, so that the learner can achieve the experience of doing.

2.3.3. Behaviorist Theory. Behaviorism theory emphasizes learners' specific behavioral performance and advocates measuring the effect of teaching from learners' actual behaviors, which has important implications for the development of virtual experiments. In practical experiment teaching, it is difficult to meet the needs of students' experimental operation due to limited conditions. In the actual experimental teaching at this stage, it is difficult to meet the

experimental teaching requirements that students form operation skills through repeated practice. The development of virtual experiment provides another experimental operation environment for learners.

3. Experiment

3.1. Data Source and Experimental Equipment. When training candidate area networks and object classification and detection networks in artificial intelligence ideological and political teaching, this paper uses a unified image scale as input. On this basis, the proportion of nearly 30000 objects in 2000 video images and 10000 ordinary images is calculated, and the statistical histograms of average height are drawn. Through statistics, the minimum values of the average width and height of the target 130 are obtained. The video image size is 9000×4500 ; this is bigger. The video image is uniformly downsampled to 2500×1250 , and the compression ratio is 3.6 so that the smallest target also has a resolution of 36×36 . Finally, we ensure that the length of the shortest side of all images is 1250 during training. The specifications and parameters of conventional images and video images are shown in Table 1.

In order to adapt to common camera interfaces, such as HD-SDI interface, USB, and IEEE 1394 interface, three types of acquisition devices, HD camera, standard definition camera, and camera, are selected for testing. The test of each acquisition device requires the use of teacher computer and acquisition computer. The teacher computer adopts 32-bit Windows 7 system and 3.0 GHz dual-core CPU, and the computer screenshot software realized by image driving technology is installed; the acquisition computer uses a 64-bit Windows 7 system and a 2.66 GHz quad-core CPU to install the recording and broadcasting system described in this paper. The acquisition equipment used in the test is shown in Table 2.

3.2. Overall Design of the Multimedia Teaching System. Modular design is adopted to improve the efficiency of program development and reduce the difficulty of code maintenance. This also facilitates code reuse. It is also more conducive to code reuse. The system is divided into console module, teacher-side image acquisition module, audio-video acquisition module, picture-in-picture module, video switching module, and audio-video compression coding module. The overall functional module diagram of the system is shown in Figure 1.

The console module mainly implements interface design and image display functions, mainly displaying high-definition camera video, computer screen, picture-in-picture, and switched pictures. The teacher-side image acquisition module mainly completes the teacher-side image sending, receiving, format conversion, and BMP image synthesis functions. The audio and video acquisition module mainly realizes the collection of video and audio of HD cameras and also supports SD cameras, cameras, and microphones. The picture-in-picture module implements functions such as image scaling and picture-in-picture synthesis. The image

switching module mainly completes switching, image format conversion, and other functions. The audio and video compression and coding module mainly realizes the compression and coding of video stream and audio format conversion.

3.3. Image Acquisition Process Based on Embedded VR Video Processing Control System

3.3.1. Screen Image Sending. Now we introduce the functions that need to be used in the system. The sending process of the screen image on the teacher side is shown in Figure 2.

Firstly, the image acquisition of the changing area of the teacher's computer screen is completed with the image driver. Then the collected image is compressed into JPEG format to ensure high definition and high compression ratio, which is convenient for network transmission. Then the JPEG image data is packaged and the header format information is added to ensure that the receiver can accurately obtain the image capture sequence, the coordinates on the computer screen, and the length of the data. Finally, the image data is transmitted to the network through the socket.

3.3.2. Screen Image Reception. The teacher-side image acquisition module needs to receive the screen image sent by the teacher and synthesize the changed image part into a complete computer screen image. The receiving process of the screen image is shown in Figure 3.

First use socket to receive the screen image; then extract JPEG data according to the packet header format defined in Figure 3; JPEG image synthesis often has problems such as inconsistent block positions or quality factors. In order to ensure that the image can be accurately restored to the computer screen image and reduce the complexity of the operation, the JPEG image is converted to the BMP format. Finally, according to the image coordinate information in the packet header format, the converted BMP image is synthesized into a complete computer screen image, and the complete computer screen image is sent to the console module.

3.3.3. Screen Image Synthesis. After converting the JPEG format image into a BMP image, the image needs to be synthesized into a complete computer screen image. Since the full screen image is received for the first time, the resolution of the desktop of the teacher can be obtained according to the coordinate information described in Figure 4, and then a screen display buffer is established according to the resolution to store the full screen image in BMP format. When the image of the changed part is received, the converted BMP image is filled to the specified position of the buffer according to the image coordinate information.

3.4. Functional Test of Embedded VR Video Image Control System. This section will introduce the main functions and corresponding functional test methods of the embedded VR

TABLE 1: Specifications and parameters of conventional images and video images.

Image type	Image size	Resolution	Number
Regular image	9000 × 4500	130 × 130	5000
Video image	9000 × 4500	130 × 130	1000
Regular image	2500 × 1250	36 × 36	5000
Video image	2500 × 1250	36 × 36	1000

TABLE 2: Acquisition equipment used for testing.

Collection device type	Collection device model	Collection device interface
HD camera	Panasonic AJ-HPX2100 MC and HD USB external capture card TC-UB100SDI	HD-SDI
SD camera	SONY DSR-250p	1394
Camera	Rainbow flying, dual-core USB 2.0, no driver	USB 2.0

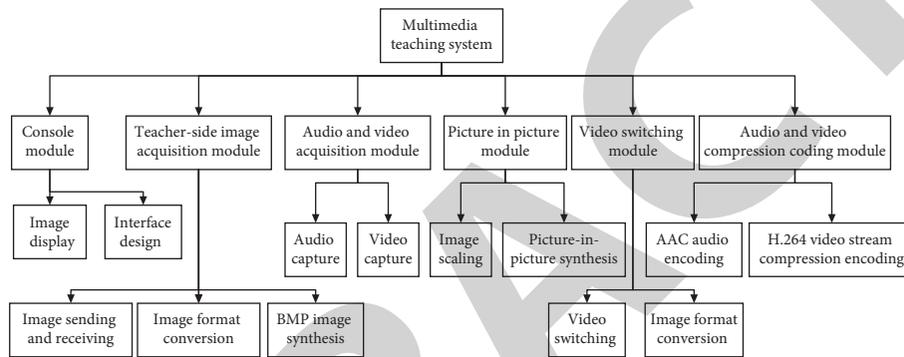


FIGURE 1: The overall function module of the system.

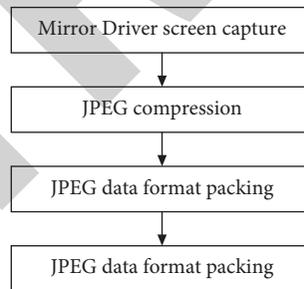


FIGURE 2: Screen image sending process.

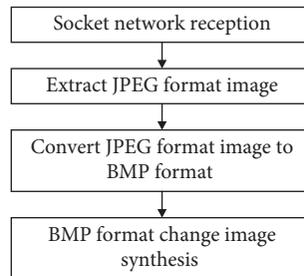


FIGURE 3: Screen image receiving process.

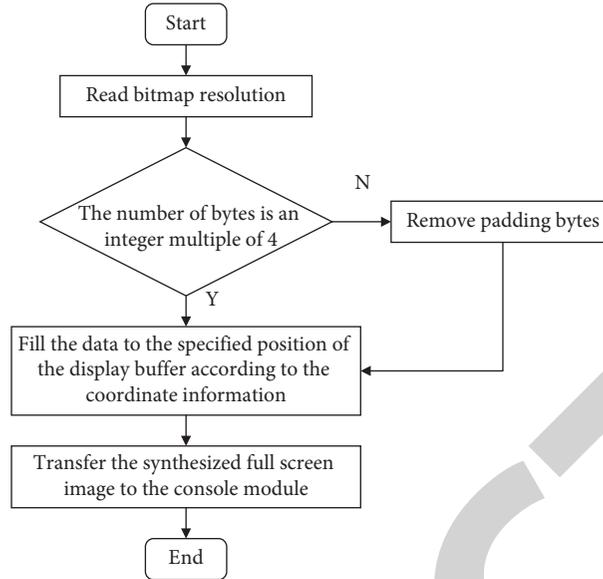


FIGURE 4: Image synthesis process.

video image control system. The test data are shown in Table 3. The test methods and tools are as follows:

- (1) VLC multimedia player, test image compression method, resolution, frame rate, and audio parameters
- (2) Kelai Network Analysis System (CSNAS), used to test support for network protocols and network delays
- (3) Browser for testing system configuration functions
- (4) ONVIF protocol analysis tool to test the interconnection protocol between devices
- (5) Voltmeter and power meter, used to test the electrical parameters of the equipment
- (6) Temperature and hygrometer, measuring the working environment parameters of the equipment
- (7) Related tools needed to set up the experimental environment

4. Results and Discussion

4.1. Analysis of Training Results of Video Image Classification Detection. This article has carried out experiments for different initial scales and initial aspect ratios. Experiments were also performed on a single scale with multiple aspect ratios. When the scale is 1282 and the aspect ratio is $\{1:1, 1:2, 2:1\}$, the average detection accuracy is 76.46%. When the scale is 2562 and the aspect ratio is $\{1:1, 1:2, 2:1\}$, the average accuracy of detection is 77.11%. Then, a multiscale single-width-to-height ratio comparison experiment was performed. For a scale size of $\{1282, 2562, 5122\}$ and an aspect ratio of 1:1, the average detection accuracy was 77.92%.

The experiments show that the composition of the predicted regions with different ratios has a certain impact

on the final cache rate. It can be clearly seen that the detection category of single-scale single-aspect ratios will decrease. The data samples are again sampled with the target area scale and aspect ratio after downsampling. For the distribution of object scales in video images, we obtain the histogram of the scale distribution by counting the square root of the target area, as shown in Figure 5. It can be seen that the scale size of the object is mainly distributed between 0 and 500, and a small proportion of objects with a scale above 500 has been occupied.

Further, the width to height ratio of the object in the video image is sampled, and the statistical results are shown in Figure 6. The aspect ratio is distributed from 1:3 to 3:1, mainly around $\{1:1, 1:2, 2:1, 3:2\}$. According to the statistical results and previous experiments, we chose four scale units $\{642, 1282, 2562, \text{ and } 5122\}$ and four aspect ratios $\{1:1, 1:2, 2:1, \text{ and } 3:2\}$ as the reference area size of the candidate area. The average accuracy of these scale and aspect ratio experiments is 79.28%.

In the same way, the data set used in this method training is used. We train the algorithm model of video image classification and detection based on image control in 10 categories by testing 5000 conventional images and 1000 video images that are not involved in the training in the data set, as shown in Figure 7.

It can be seen from Figure 7 that, in the case where the shape of the object is relatively fixed and not sensitive to changes in the viewing angle, the gradient features in this paper can be well represented, so the detection rate is not bad. The detection accuracies of bed, table, monitor, curtain, chair, door, window, picture, mirror, sofa, and mAP are 63.15%, 59.38%, 60.11%, 60.85%, 61.37%, 59.33%, 65.94%, 64.73%, 60.17%, 63.22%, and 62.94%. The average accuracies of these ten video images are 75.68%, 73.19%, 78.28%, 74.29%, 75.84%, 73.29%, 76.81%, 75.08%, 71.98%, 74.20%, and 76.18%. This also shows that the algorithm model of

TABLE 3: System function test.

Image	Maximum image size	3648 × 2752
	Frame rate	H.264 60 fps (1920 × 1080), H.264 30 fps D1 (720 × 480), MJPEG 5 fps (3644 × 2748)
Storage, networking, and interfaces	Storage function	Support SD
	Audio input and output	1 audio input (MIC IN/Line IN) 3.5 mm interface, 1 3.5 mm audio output interface
	Interface protocol	ONVIF
Environmental adaptability and power test	Power supply	DC12 V ± 10%/PoE
	Power consumption	6 W MAX (9 W MAX when ICR is switched)

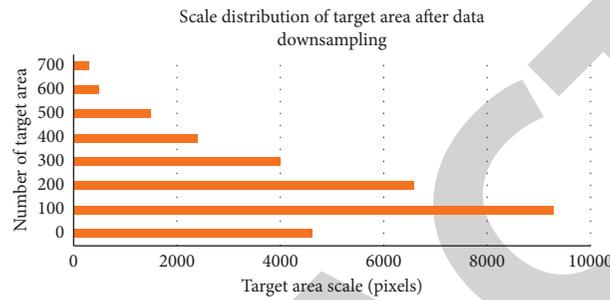


FIGURE 5: Scale distribution of target area after data downsampling.

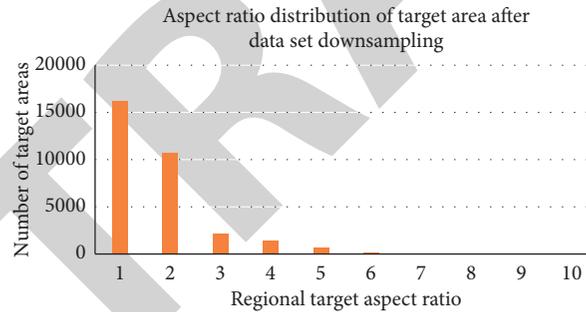


FIGURE 6: Aspect ratio distribution of the target area after data downsampling.

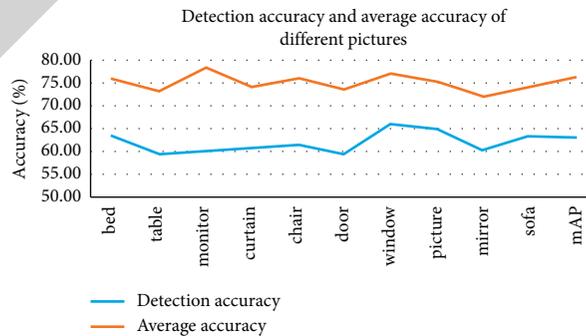


FIGURE 7: Control accuracy and average accuracy of video images.

video image classification and detection based on image control in this paper can extract deeper and more abstract features to classify targets.

4.2. Performance Test of Embedded VR Video Image Control System. To fulfill the needs of video image acquisition, the acquisition equipment selects high-definition cameras. The

TABLE 4: CPU usage at different frequencies.

Resolution	CPU (frequency 10)			CPU (frequency 20)			CPU (frequency 30)		
	Min (%)	Intermediate (%)	Max (%)	Min (%)	Intermediate (%)	Max (%)	Min (%)	Intermediate (%)	Max (%)
350 × 288	42.15	42.99	43.68	42.39	43.12	44.05	42.91	43.39	44.22
840 × 480	45.24	46.29	46.98	45.61	46.85	47.21	46.04	47.31	48.05
720 × 576	48.94	49.58	50.27	49.11	50.08	50.96	50.07	51.28	51.99
1280 × 720	52.67	53.66	54.27	53.12	54.33	54.97	53.69	55.02	56.31
1920 × 1080	55.28	57.30	58.96	56.06	58.44	60.17	57.83	60.33	62.09

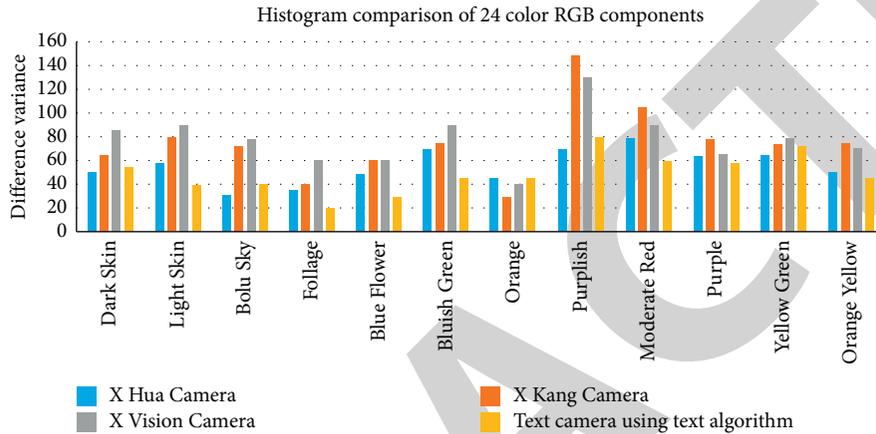


FIGURE 8: Histogram comparison of 24-color RGB components.

way of playing videos on the teacher's computer is adopted here to ensure the continuous transmission of changing screen images. By setting different output video resolutions and video capture frame rates to the recording and playback system and running for 1 hour, record the minimum, maximum, and average CPU usage. The resolutions are mainly 352×288 , 840×480 , 720×576 , 1280×720 , and 1920×1080 . Frame rates include 10, 20, and 25. The following will test the CPU usage in the case of a resolution corresponding to different sampling frame rates. Table 4 shows the CPU cache of different frequencies.

The test results show that, under the same resolution, the CPU usage increases as the frequency increases; at the same frequency, the CPU usage increases as the resolution increases. At present, the performance of computers on the market is different. The recording and broadcasting system with software processing as the core does not have a uniform standard on CPU usage. At the current resolution and frequency, the CPU usage meets the needs of multimedia teaching systems; and, after running for 12 hours on the laptop, the system can still run smoothly and meet the requirements.

After analyzing the output data, it was found that the intermediate value of the 24-color card difference of the X Hua camera was 53.2916, the average value of the 24-color card difference of the X Kang camera was 56.3394, and the average value of the 24-color card difference of the X Vision camera was 60.2937. The average color difference of the 24-color card taken by the camera of the VR video image control algorithm is 40.1168. To make the output results look

more intuitive, Figure 8 is a histogram comparison of the 24-color RGB components of the images taken by four different cameras. Based on the experimental data, it can be seen that the algorithm improves the control effect of video image processing.

4.3. Performance Analysis of Video Image Control Based on Multimedia Teaching System. This article evaluates and analyzes the performance improvement of the multimedia teaching system from the aspects of video image control failure rate, average storage access delay, and video image control time.

Figure 9 illustrates the comparison of the failure rates of the embedded VR video image control algorithm and the original algorithm constructed in this paper. Through experimental analysis, it is found that, compared with the original algorithm model, the embedded VR video image control algorithm constructed in this paper can reduce the average failure rate by 16.27% and can obtain an average failure rate compression of 11.08%. Except for "table," the failure rates of the other 9 groups of video images have been compressed to varying degrees. For mirror, chair, and door, the embedded VR video image control algorithm constructed in this paper can reduce the maximum failure rate by 49.16%, 61.02%, and 66.94%, respectively.

Figure 10 illustrates the comparison of memory access performance improvement between the embedded VR video image control algorithm and the traditional algorithm constructed in this paper. Compared with the

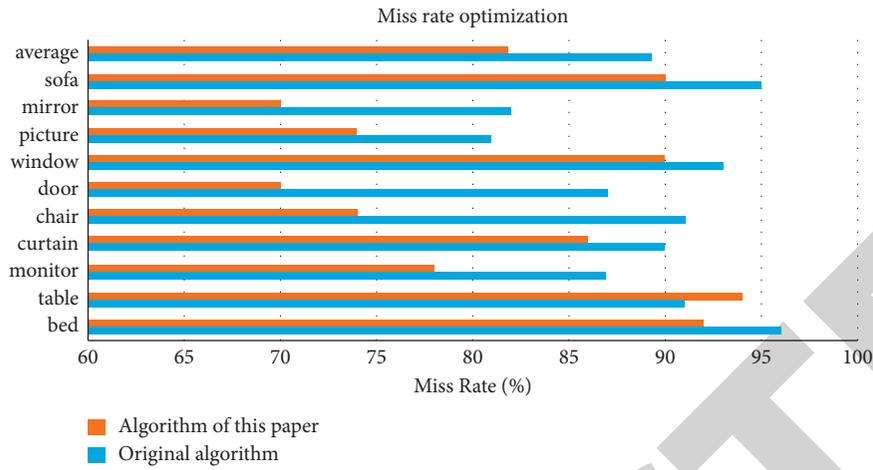


FIGURE 9: Optimization of failure rate.

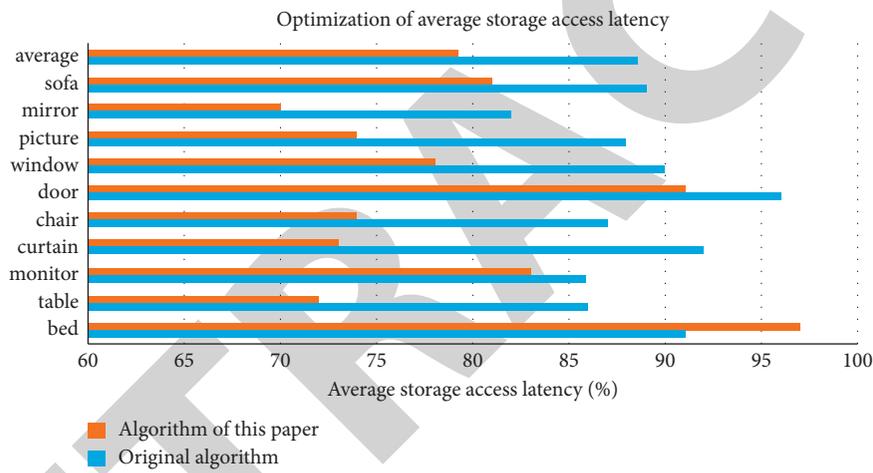


FIGURE 10: Optimization of average storage access latency.

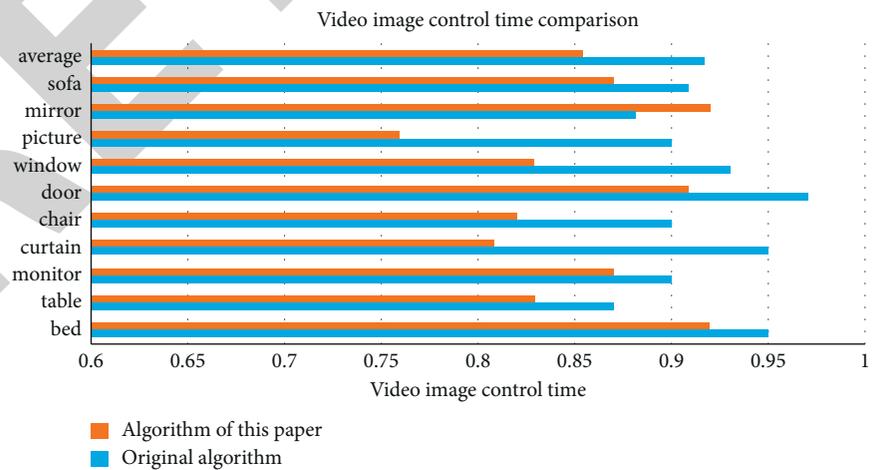


FIGURE 11: Video image control time comparison.

traditional algorithm, the embedded VR video image control algorithm constructed in this paper can reduce the storage access delay time of 10 different video images by

an average of 15.93% and can achieve a performance optimization of about 9.37%. For mirror images such as mirror, picture, chair, curtain, and table, the degree of

optimization is obvious; for video images such as sofa, door, monitor, and bed, the performance improvement of overflow operation is very limited, so it fails to show good performance.

The performance improvement of overflow operation mainly comes from two aspects. First, the fine-grained pseudodivision method is used to adjust the private and shared quotas between each set in the same aggregate set and can perceive the dynamic changes of the memory access characteristics of the program at different stages of operation. Second, by sacrificing the cooperation between block replacement, overflow, and receive operations, it is possible to dynamically control the proportion of capacity space occupied by local blocks and overflow blocks in each set.

Figure 11 illustrates the comparison of the overall performance improvement of the multimedia teaching system between the embedded VR video image control algorithm and the traditional algorithm constructed from the perspective of video image control time. For 10 different video images, the embedded VR video image control system in this paper can reduce the video image control time by 7.28% and 10.63%, respectively. For picture, the embedded VR video image control system in this article can improve performance by up to 28.49%.

5. Conclusion

Based on the detailed analysis of artificial intelligence VR technology, this paper constructs an embedded artificial intelligence VR video image control thinking cognitive teaching system and uses a systematic teaching design to analyze the development of thinking cognitive teaching. This paper selects VRML in virtual reality technology to model the ideological cognitive teaching scene and analyzes the specific implementation and system performance in the experiment. The embedded VR video image control algorithm constructed in this paper can reduce the maximum failure rate by 49.16%, 61.02%, and 66.94%, respectively. Compared with traditional algorithms, the artificial intelligence embedded VR video image control algorithm proposed in this paper can reduce the average storage access time of 10 different video images by 15.93%, the performance can reach 9.3%, and the video image control time can be reduced by 7.28% and 10.63%, respectively. For pictures, the VR video image control system embedded in this article can improve performance by up to 28.49%. Therefore artificial intelligence VR technology can boost the progress of teaching and improve the learning efficiency of students.

Data Availability

This article does not cover data research. No data were used to support this study.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] N. Baskaran and R. Eswari, "Efficient VM selection strategies in cloud datacenter using fuzzy soft set," *Journal of Organizational and End User Computing*, vol. 33, no. 5, pp. 153–179, 2021.
- [2] K. Stepan, J. Zeiger, S. Hanchuk, A. Signore, and A. Iloreta, "Immersive virtual reality as a teaching tool for neuroanatomy: immersive VR as a neuroanatomy teaching tool," *International Forum of Allergy & Rhinology*, vol. 7, no. 3, pp. 1006–1013, 2017.
- [3] C. Guo, Y. Cui, and Z. Liu, "Optimal multicast of tiled 360 VR video," *IEEE Wireless Communication Letters*, vol. 22, no. 12, pp. 2563–2566, 2018.
- [4] F. Berghea, V. Vlad, L. Palanciuc et al., "Quantitative Doppler in musculoskeletal ultrasonography-suboptimal performance of both experienced and in-training sonographers in selection of the highest Doppler signal image from cine-loops," *Medical Ultrasonography*, vol. 19, no. 3, p. 282, 2017.
- [5] R. Konrad, D. G. Dansereau, A. Masood, and G. Wetzstein, "Spin VR: towards live-streaming 3D virtual reality video," *ACM Transactions on Graphics*, vol. 36, no. 6, 2017.
- [6] H. T. Lee and Y. S. Kim, "The effect of sports VR training for improving human body composition," *EURASIP Journal on Image and Video Processing*, vol. 1, pp. 1–5, 2018.
- [7] H.-I. C. Wei-Tse, M.-S. Chen, I.-C. Shen, and B.-Y. Chen, "High-resolution 360 video foveated stitching for real-time VR," *Computer Graphics Forum*, vol. 36, no. 7, pp. 115–123, 2017.
- [8] M. Aline, T. Rafael, and N. Luciana, "An analysis of VR technology used in immersive simulations with a serious game perspective," *IEEE Computer Graphics & Applications*, vol. 38, no. 2, pp. 57–73, 2018.
- [9] N. Jain, A. Wydra, W. Hai, N. M. Thalmann, and D. Thalmann, "Time-scaled interactive object-driven multi-party VR," *The Visual Computer*, vol. 34, no. 9, pp. 1–11, 2018.
- [10] K. Y. Chau, K. M. Y. Law, and Y. M. Tang, "Impact of self-directed learning and educational technology readiness on synchronous e-learning," *Journal of Organizational and End User Computing*, vol. 33, no. 6, pp. 1–20, 2021.
- [11] J. Jerald, "Human-centered VR design: five essentials every engineer needs to know," *IEEE Computer Graphics and Applications*, vol. 38, no. 2, pp. 15–21, 2018.
- [12] L. P. Berg and J. M. Vance, "Industry use of virtual reality in product design and manufacturing: a survey," *Virtual Reality*, vol. 21, no. 1, pp. 1–17, 2017.
- [13] L. Pengsen, H. Cui, Y. Cao, X. Hou, and L. Zou, "A method of multimedia teaching evaluation based on fuzzy linguistic concept lattice," *Multimedia Tools and Applications*, vol. 78, pp. 30975–31001, 2019.
- [14] R. E. Mayer, "Using multimedia for e-learning," *Journal of Computer Assisted Learning*, vol. 33, no. 5, pp. 403–423, 2017.
- [15] M. Huijbregts, R. Ordelman, and F. D. Jong, "Annotation of heterogeneous multimedia content using automatic speech recognition," *Lecture Notes in Computer Science*, vol. 4816, no. 6, pp. 78–90, 2017.
- [16] M. Asnawi and A. Merza, "The effectiveness of immersive multimedia learning with peer support on English speaking and reading aloud," *International Journal of Instruction*, vol. 10, no. 1, pp. 203–218, 2017.
- [17] Y. Hou and Q. Wang, "Research and improvement of content-based image retrieval framework," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 32, no. 12, Article ID 1850043, 2018.

- [18] Q. Zhang, L. T. Yang, Z. Chen, and F. Xia, "A high-order possibilistic-means algorithm for clustering incomplete multimedia data," *IEEE Systems Journal*, vol. 11, no. 4, pp. 2160–2169, 2017.
- [19] Y. Zhang, Q. He, Y. Xiang et al., "Low-cost and confidentiality-preserving data acquisition for internet of multimedia things," *IEEE Internet of Things Journal*, vol. 5, no. 5, pp. 3442–3451, 2017.
- [20] B. Hossein and F. A. Ebrahim, "Multimedia listening comprehension: metacognitive instruction or metacognitive instruction through dialogic interaction," *Recall*, vol. 30, no. 1, pp. 131–152, 2018.
- [21] R. Werner, A. Arslanm, A. K. Peter, A. Luigi, and L. Sun, "Challenges of future multimedia QOE monitoring for internet service providers," *Multimedia Tools & Applications*, vol. 76, no. 3, pp. 1–24, 2017.
- [22] R. Michael, P. Konstantin, L. E. Sigrun, T. S. Peter, and D. Thomas, "From annotation to computer-aided diagnosis: detailed evaluation of a medical multimedia system," *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 13, no. 3, pp. 1–26, 2017.
- [23] B. Susanne, E. Touradj, G. Cathal, J. Laleh, E. Noel, and O. Connor, "Mm health 2017: workshop on multimedia for personal health and health care," *IEEE Multimedia*, vol. 25, no. 1, pp. 7–11, 2018.
- [24] M. Larson, M. Soleymani, G. Gravier, B. Ionescu, and G. J. F. Jones, "The benchmarking initiative for multimedia evaluation: mediaeval 2016," *IEEE Multimedia*, vol. 24, no. 1, pp. 93–96, 2017.