Journal of Image Processing Theory and Applications (2023)
Clausius Scientific Press, Canada

DOI: 10.23977/jipta.2023.060114
ISSN 2560-6239 Vol. 6 Num. 1

Development of Web3D education platform suitable for schoolchild

Yanyu Liu[1,a], Donghui Bao[1,b], Lili Ye[2,c], Yanyan Chen[1,d]

[1]School of Electronic Information Engineering, Beihai Vocational College, Beihai, 536000, China
[2]Yifu Primary School of Haicheng District, Beihai City, Beihai, 536000, China
[a]liuyanyu@bhzyxy.edu.cn, [b]bhbdh@126.com, [c]liliye2023@163.com, [d]chenyy202312@163.com

Keywords: Education platform; Schoolchild; Facial expression recognition; Action recognition

Abstract: 3D education has vivid three-dimensional expression and powerful interactive functions, which is in line with the situational teaching and cognitive requirements of schoolchild. It is an excellent platform for displaying teaching content, and more importantly, it can improve schoolchildren's learning enthusiasm. This article implements a three-dimensional curriculum teaching mode application, breaking away from the singularity of two-dimensional view teaching mode in presenting information. The educational scene is built on the WebGL framework based on Three.js, and a Web3D education platform suitable for young children is constructed to achieve three-dimensional teaching scene. At the same time, the facial expression recognition algorithm and human action recognition algorithm based on deep learning are applied to the Web3D education platform to achieve natural and real-time human-computer interaction in the teaching process without manual input instructions, so that learners can feel the sense of participation and immersion brought by intelligent interaction, and increase the fun of learning.

1. Introduction

In the field of basic education, schoolchild-centered teaching model is mainly adopted. In this model, whether the learning enthusiasm of schoolchildren, especially schoolchild aged 6-12, can be effectively mobilized and whether creative thinking can be reasonably developed will directly affect the teaching effect. Therefore, higher requirements are put forward for teachers' teaching. In this context, higher requirements have been put forward for the design and production of teaching courseware. Excellent teaching materials should be able to fully mobilize schoolchildren's learning enthusiasm and enable them to better participate in teaching. However, in the actual teaching process, some teaching content that cannot be expressed well with pictures and text, such as space and graphics in primary school mathematics, animal cognition in primary school share reading, scene dialogues in primary school English, etc. These abstract knowledge can be expressed in an immersive manner Interactive multimedia courseware for teaching presentation will definitely achieve better teaching results.

At present, the multimedia courseware making tools commonly used by teachers mainly include

PPT, whiteboard, Flash and so on. These tools cannot form a three-dimensional effect when expressing teaching scenarios such as geometric construction, animal and plant cognition, role-playing dialogue, etc. In these application scenarios, 3D courseware is significantly better than 2D courseware made of text, images, audio and video, and animation, making it more attractive to schoolchild.

3D courseware has strong interactive ability, realistic immersive experience, and vivid 3D expression, which conforms to the cognitive laws of young children and the requirements of situational teaching. It can better improve schoolchildren's learning enthusiasm and is a teaching display platform suitable for young children. But currently, many 3D engines rely on plugins or rely on a huge support environment, which greatly limits the development of 3D courseware.

European and American countries were the first to apply 3D technology to the field of basic education. Due to its intuitive teaching environment, 3D technology provides schoolchildren with a better teaching experience and provides a new model for the development and reform of basic education. In recent years, China has also accelerated research on the application of 3D technology in the field of basic education. In order to promote the development of 3D education in China, the Ministry of Education has cooperated with relevant domestic enterprises to establish 3D Education Research Center, which provides a large number of 3D teaching courseware for primary and secondary school teachers. These resources provide reform samples for traditional teaching models, increase the interactivity and innovation of teaching. Domestic research mainly focuses on the following areas [1, 2]:

1) 3D courseware, such as the design and research of primary school English courseware based on Flash and Papervision 3D technology;

2) Virtual reality and augmented reality applications, such as the design and development of X3D based exploratory virtual learning environments, research on the design and application of primary school English teaching resources based on augmented reality, research on the application of immersive virtual reality in primary school science classes, and practical research on augmented reality technology in primary school art teaching;

3) 3D educational games, such as the design and development of primary school mathematics education game "Clever Play Shape" based on the Cocos2d-X engine, and the design and development of a gamified learning environment based on Unity3D under the STEM concept.

2. Web3d technology

The emergence of Web3D technology has injected new vitality into 3D education and provided new research ideas for the field of 3D interaction. Web3D is a new technology that combines web technology with 3D technology, which is a networked extension of 3D technology. Its main features include real-time rendering, infinite interactivity, virtual reality, network optimization and compression, etc. Based on this, this article will use a WebGL 3D world construction technology [3] applied to the web without relying on any browser plugins to build a Web3D education platform suitable for schoolchild.

At present, the mainstream technologies of Web3D at home and abroad mainly include VRML, X3D, Viewpoint, Flash Player, Shockwave3D, GoogleO3D, Java Applet, etc. However, early technologies were not mature. In 2014, the World Wide Web Alliance established a new HTML5 standard, and WebGL, as an important technology in the HTML5 standard, solved two challenges for us: firstly, it can achieve web interactive 3D animation production through JavaScript scripting programs without the support of any web browser plugins; Secondly, it can achieve graphic rendering through the OpenGLES2.0 graphics rendering library. Web3D has been widely developed and applied in the field of constructing virtual learning environments [4]. Based on this, this article

mainly adopts WebGL construction technology to build a Web3D education platform suitable for schoolchild.

3. Web3D education platform construction

Vivid 3D models have an innate advantage for cognitive stimulation in schoolchild aged 6-12 years. At the same time, compared with the two-dimensional courseware commonly used at present, three-dimensional immersive virtual interaction scenarios can improve schoolchildren's learning enthusiasm and develop schoolchildren's creative thinking. When choosing to build teaching scenes, it mainly aims at the teaching contents that two-dimensional courseware cannot express well, such as the cognitive teaching of animals and plants that schoolchildren can hardly access in real life and the interactive Q&A based on this. Firstly, 3D models corresponding to the different animal and plant species displayed in the Web3D teaching scene are generated for teaching demonstration. Secondly, various difficult test questions are generated to match the 3D models of animals and plants according to the content of the questions, and various vivid 3D virtual models presented in the Web3D teaching scene are used. Compensate for deviations in schoolchildren's imagination and understanding.

The construction of Web3D education platform mainly consists of two parts: first, to build an immersive 3D virtual teaching environment for schoolchildren to use; secondly, establish a 3D answering environment for schoolchildren to use. The overall implementation process of the system is shown in Figure 1. Select Three.js as the WebGL framework used in the research institute, and use the model loading program provided by THREE.js to introduce the 3D model into the scene interface and present the 3D animation effect. By adjusting the lighting of the page scene, setting up the renderer, and calling the model loading function, a three-dimensional virtual scene is ultimately constructed on the web end. Modularize the functionality of virtual classrooms by constructing reusable components. The overall implementation plan of the system is shown in Figure 2.

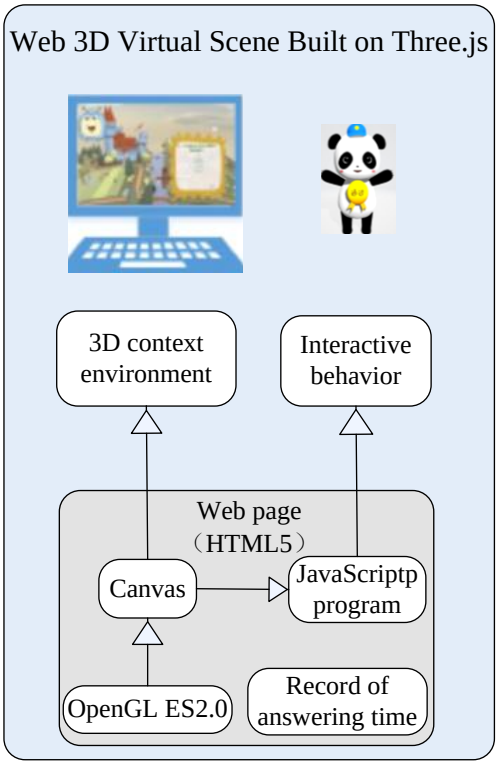

Figure 1: Overall implementation process of the system

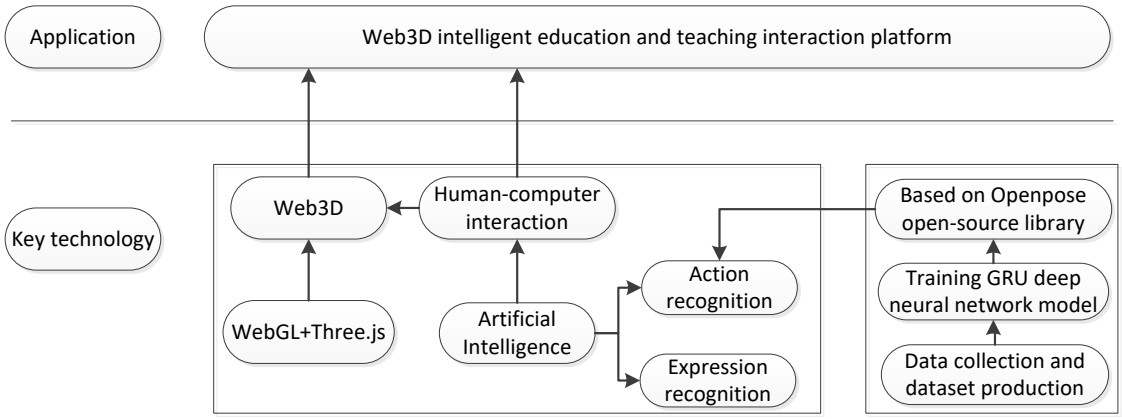

Figure 2: Overall experimental plan of the system

According to the development requirements of the intelligent interaction framework, this framework divides the business logic functions into four modules.

The first module is to build the classroom environment, which is the cornerstone of the framework and is used for the initialization and construction of the entire system. In order to reduce code redundancy and lightweight the entire project, this article designs four basic components, encapsulating scene, renderer, camera, and lighting one by one. Scene is the foundation of the entire virtual classroom, and all objects to be presented can only be presented under the influence of the camera and lighting by adding them to the Scene. The simulation of the entire virtual classroom is determined by the lighting, and the perspective seen by the user is determined by the camera. The function of the renderer is to project the entire rendering effect onto the browser for presentation. So when you want to create another classroom content scene, the framework in this article can quickly initialize and construct the scene, and you only need to change the corresponding peripheral parameters to obtain a variety of scenes.

The second module is to load course scene content. After the system initializes the classroom scene, this article will present the corresponding scene content based on the user's scene selection (if the user does not select the corresponding scene button, this article will prioritize displaying indoor scenes). In order to quickly load the model to be presented, this article encapsulates the commonly used Mesh objects, such as the Geometry and Material components contained in the Mesh objects. For individual complex models that need to be loaded, the framework also provides additional loaders to ensure that special models and additional textures are pre-loaded from non-internal sources, enabling these complex models to achieve fast automatic loading. Virtual teachers are the key to virtual classrooms, and all settings in virtual classrooms are designed for the construction of virtual teachers. For this framework, two interfaces are set: attribute and interactive behavior settings. Attribute settings include basic teacher parameter settings (position, size, lighting rendering), and interactive behavior settings are used to engage in corresponding dialogue and interaction when receiving feedback from the user or server, playing a role in motivating the situation.

The third module is human-computer interaction, and a good teaching experience depends on the degree of interactivity of the system. According to the corresponding design, this article divides the human-computer interaction module into three parts: ①voice API, camera controller, and mouse selector. Voice API: Google's voice API is used to obtain voice and convert it into text for transmission to the server, and all output languages of the intelligent teacher are also played by this voice API; ②Camera controller: The scenes in this article are divided into two categories: indoor and outdoor. Regardless of whether it is an indoor or outdoor scene, users can freely control the angle change of the camera through the mouse to act as their own eyes. The system can also switch

different perspectives according to the change of problem type; ③Mouse controller: In the Web3D environment, the mouse is equivalent to the user's hands. Except for the interaction with the intelligent teacher that is not controlled by the mouse, all other interactive objects on the interface are completed by the user's mouse. This article can also use event callback functions to achieve the interaction behavior between the user and other objects.

The fourth module is to display virtual scenes, and all dynamic effects in the browser are achieved through animation mechanisms. The virtual scenes in the virtual classroom mainly exist in two parts: some animation settings when the virtual teacher interacts with the user, and the animation processing of various objects in the image scene of the 3D course scene.

4. Natural human-computer interaction based on facial expression recognition in Web3D education platform

Emotion is a fundamental component of every individual, which can influence behavior, thinking ability, decision-making, cognition, adaptability, happiness, and the way humans communicate with each other [5]. We hope that computers can understand us humans and interact with them more naturally. Therefore, computers must be able to recognize and express emotions, and possess "emotional intelligence". For educational scenarios, learners' emotions are first expressed through facial expressions [6]. Charles Darwin [7] explored the universal meaning of facial expressions and their association with a specific emotion. He classified human emotions and proposed six basic emotions: happiness, sadness, surprise, fear, disgust, and anger, each of which corresponds to its unique facial expression characteristics. There are two main methods for extracting facial expressions: dynamic sequences and static images [8, 9]. As an effective expression of emotions, expressions are intuitive and easily perceivable explicit features. Minor changes in key facial regions can reflect an individual's immediate mentality. Therefore, effectively capturing the features and changes of facial information is an important prerequisite for analyzing individual emotional dynamics. The deep learning based expression recognition algorithm uses neural networks to classify and learn a large number of facial emotion images, ultimately achieving high efficiency. Common methods include [10, 11, 12, 13, 14, 15, and 16]. Real time facial expression analysis, multi frame image emotion judgment, and learner's partial operational behavior data are used as the basis for emotion classification, ultimately achieving personalized changes in teaching strategies during the answering process intelligent interactive technologies such as real-time emotional interaction.

To achieve rapid detection and real-time analysis of emotional analysis in the education process, this paper proposes an emotional analysis model based on deep learning. The core content of this model is expression detection and recognition.

1) Constructing an expression dataset for educational scenarios used in the project: The facial expression dataset produced needs to be diverse, and a more refined classification is needed based on the seven common expression datasets (currently the selected dataset is fer2013, and a more comprehensive emotionet dataset will be adopted in the future). Initially, expressions are classified into three categories: positive, non-positive, and neutral. In addition to common expressions, expression samples that meet educational scenarios also need to be defined. Based on the study of the existing database recording process, a detailed expression database construction plan is specified, and a relatively complete expression database is classified and recorded for verification based on this plan. The database includes a two-dimensional RGB image sequence, depth images of corresponding frames, and three-dimensional feature point data of the entire face. The construction of this database can provide data support for subsequent expression detection and recognition.

2) Algorithm related: Based on a deep understanding of the theoretical framework of deep

learning, learn and study facial expression recognition methods based on deep learning. Put self-made facial expression training samples into a convolutional neural network (CNN) built by a deep learning framework to extract deep features of the image, and then use a softmax classifier to classify facial expression features, The real-time transmission of facial expression information is ultimately achieved through the socket interface between the web end and the server. Different human-machine interaction modes that are suitable for educational scenarios need to be defined for different expression classifications.

3) The combination of Web3D interaction engine and facial expression recognition algorithm, applying familiar knowledge of web graphics engine development, builds a simple education platform, and adds deep learning based facial expression recognition algorithm to the interface, which is used to extract features, analyze the psychological characteristics of current schoolchildren, and make natural interaction behaviors that are suitable for educational scenarios in a timely manner.

4) Implementation process:

● Create diverse facial expression datasets suitable for educational scenarios.

● Put the prepared training set into the designed CNN network for training. The initial neural network structure includes two convolutional layers, two pooling layers, and two fully connected layers, and improve the recognition accuracy of the algorithm; Convolution operation is the foundation of CNN model construction. In the CNN image processing process, the convolution kernel sequentially performs convolution operation with different position blocks of the image to obtain the output. The basic definition of convolution operation (two-dimensional) is:

$$o[n] = (X * W)(m,n) = f[m,n] * g[m,n] = \sum_{u=-\infty}^{\infty} \sum_{v=-\infty}^{\infty} f[u,v] g[m-u, n-v]$$

Where W is our convolution kernel and X is our input. Overall, CNN is essentially a variant of Multi-Layer Perception (MLP): based on full connectivity, it combines the unique advantages of sparse connectivity by using local connections between neurons in adjacent layers (convolutional layers and pooling layers) to discover the interrelationships between input features at different levels, while reducing the number of parameters that need to be learned through weight sharing.

● Apply the trained deep learning based facial expression recognition algorithm to the Web3D end to achieve a WebGL multimedia intelligent interaction engine that utilizes facial expression recognition algorithms to control the interaction between the page end and the user.

● Calling tensorflow.js on the web side reduces the transmission latency between the web side and the server.

The natural human-machine interaction technology roadmap based on facial expression recognition in Web3D education platform is shown in Figure 3.

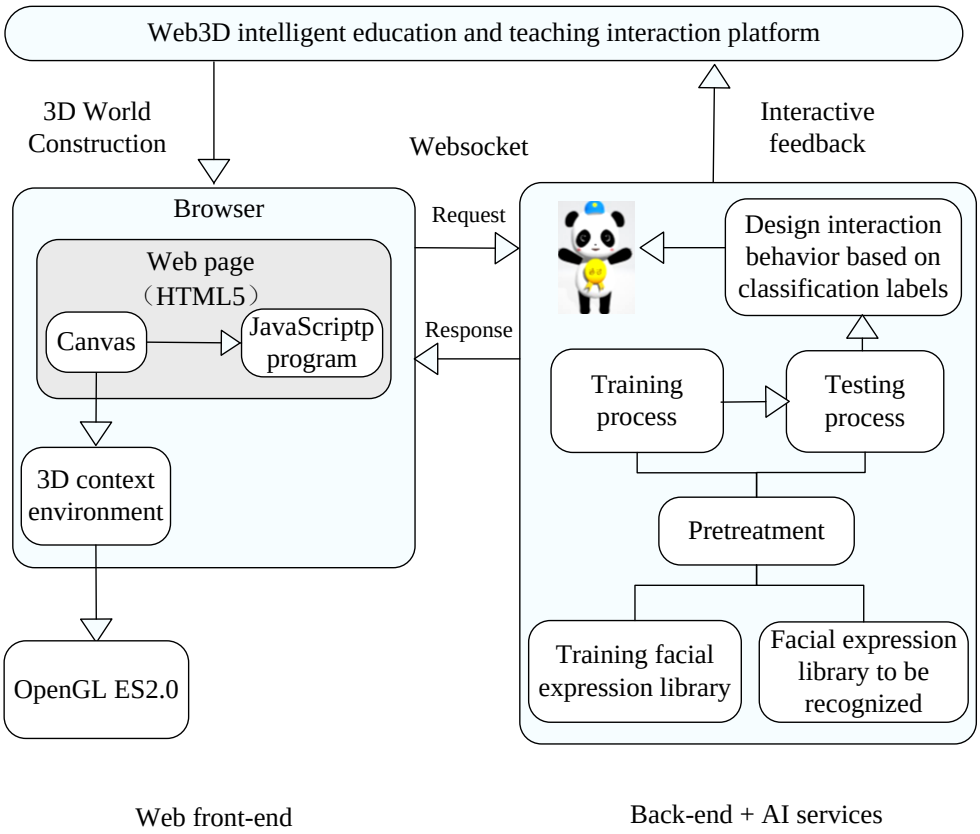

Figure 3: Roadmap of natural Human computer interaction technology based on expression recognition in Web3D education platform

5. Natural human-computer interaction based on Action recognition in Web3D education platform

Action recognition is currently one of the main technologies in human-computer interaction. Scholars and researchers have attempted to use deep learning to address challenges such as human action recognition, advanced image processing, and human tracking, and have achieved good results [17, 18, 19, 20, and 21].

Applying deep learning based human action recognition algorithms to WebGL based Web3D intelligent education and teaching interaction platforms, abandoning the rigid human-machine interaction method of mouse and keyboard in the past, and instead using human pose estimation to achieve human-machine interaction. Through GRU (Gated Recurrent Unit) training and open-source human skeleton library OpenPose and other related technologies, the recognition and tracking of human torso and limbs can be achieved, thus achieving a natural human-computer interaction in the teaching process that does not require manual input of instructions, allowing learners to feel the sense of participation and immersion brought by intelligent interaction, and increasing learning pleasure.

1) Deep action recognition: We trained a deep neural network by utilizing the GRU (Gated Recurrent Unit) and TensorFlow framework, and then implemented human pose recognition using the open-source human skeleton library OpenPose.

2) The integration of Web3D interaction engine and action recognition algorithm: We apply action recognition to control the 3D model and scene on the web in real-time, achieving human-machine intelligent interaction. The Web3D education platform enables students to directly

engage in interactive learning in a virtual teaching environment, experiencing the immersion and participation brought by intelligent interaction.

3) Implementation process:

- Create a dataset suitable for educational scenarios.
- Put the prepared training set into the GRU deep neural network for training.
- Apply the trained deep learning based action recognition algorithm to the Web3D end to achieve a Web3D intelligent education and teaching interaction platform that utilizes action recognition algorithms to control the interaction between the front-end and users.

The human-machine interaction technology roadmap based on action recognition in Web3D education platform is shown in Figure 4.

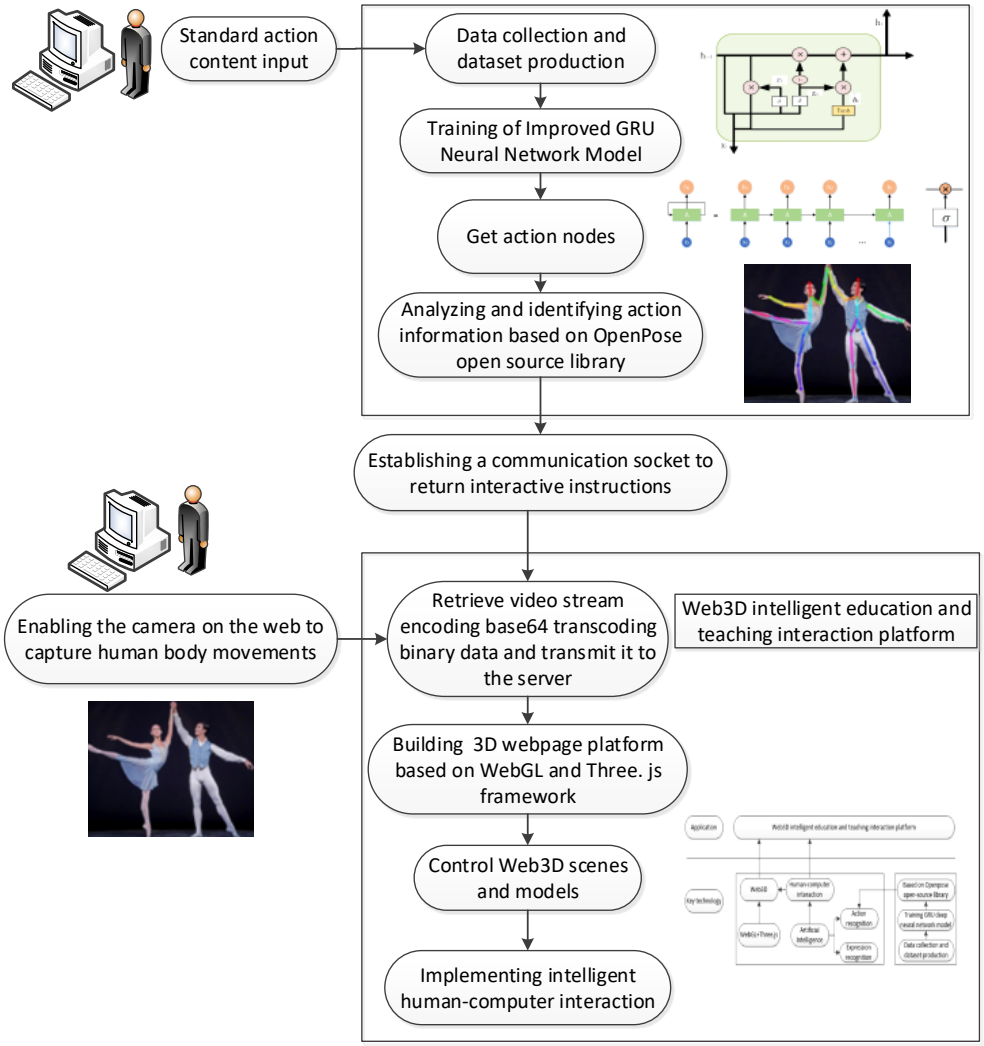

Figure 4: Roadmap of human computer interaction technology based on action recognition in Web3D education platform

The final 3D scene presentation of the schoolchild interaction interface is shown in Figure 5. This mainly includes teaching scenarios and question and answer scenarios. The teaching scenario interface involves the introduction of relevant models, which also serves as the knowledge foundation for subsequent answering scenarios. The entire answering scenario, except for the answer box, is composed of 3D models. Schoolchildren can interact with the machine through mouse clicks and keyboard input, or use action recognition to directly control the web-based 3D scene and model, achieving human-machine interaction. Facial expression recognition algorithms

accurately analyze the psychological and emotional changes of learners, and take corresponding measures to achieve intelligent human-machine interaction teaching.

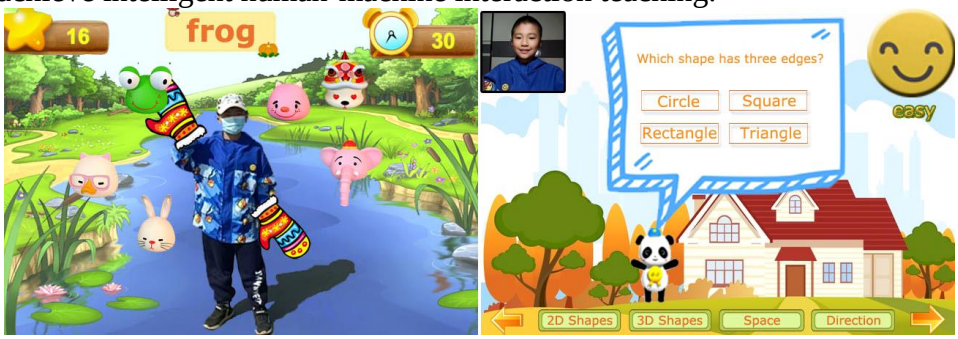

Figure 5: 3D scene of schoolchild interaction

6. Conclusions

This article implements a 3D based course teaching mode application, breaking away from the singularity of 2D view teaching mode in presenting information. The educational scene is built on the WebGL framework based on Three.js, and a Web3D education platform is constructed to achieve three-dimensional teaching scene. At the same time, the focus will be on integrating natural interaction technologies based on facial expression recognition and action recognition into existing platforms, capturing the learning status of learners in real time, analyzing changes in schoolchildren's psychological emotions, and making corresponding intervention adjustments to achieve intelligent human-machine interaction functions, effectively mobilizing schoolchildren's learning enthusiasm, and developing their creative thinking.

Acknowledgements

2020 Vocational Education Informatization Construction Research Project "Research on Natural Human Computer Interaction Based on Expression Recognition in Web3D Education Platform", Project Number: XXHJS20-0045

2021 Guangxi University Young and Middle aged Teachers Research Basic Ability Enhancement Project "Research on Natural Human Computer Interaction Based on Action Recognition in Web3D Education Platform", Project Number: 2021KY1439

References

[1] Sujie Wu: Cross-platform Application Research on Three Dimensional Interactive Courseware (Master, hebei normal university, China 2017).

[2] Ke Liu. Design and Application of 3D Visualization Resources in Teaching Scene (Master, Central China Normal University, China 2017).

[3] Qiang Fang. The Research and Implementation of WebGL Based 3D Graphics Engine (Master, Anhui University, China 2013).

[4] X. Kang, Q. Peng. Integration of CAD models with product assembly planning in a Web-based 3D visualized environment [J]. International Journal on Interactive Design and Manufacturing (IJIDeM), 2014, 8(2): 121-131.

[5] L. Morrish, N. Rickard, T. C. Chin, et al. Emotion regulation in adolescent well-being and positive education [J]. Journal of Happiness Studies, 2018, 19(5): 1543-1564.

[6] J. T. Cacioppo, L. G. Tassinary. Inferring psychological significance from physiological signals [J]. American psychologist, 1990, 45(1): 16.

[7] C. Darwin, P. Prodger. The expression of the emotions in man and animals[M]. USA: Oxford University Press, 1998.

[8] I. Cohen, N. Sebe, A. Garg, et al. Facial expression recognition from video sequences: temporal and static

modeling[J]. Computer Vision and image understanding, 2003, 91(1-2): 160-187.
[9] Z. Zeng, M. Pantic, G. I. Roisman, et al. A survey of affect recognition methods: Audio, visual, and spontaneous expressions [J]. IEEE transactions on pattern analysis and machine intelligence, 2008, 31(1): 39-58.
[10] C. W. Hsu, C. C. Chang, C. J. Lin. A practical guide to support vector classification[J]. BJU International, 2008, 101(1): 1396-1400.
[11] P. Viola, M. Jones. Rapid object detection using a boosted cascade of simple features[C]. 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2001, with CD-ROM, Kauai, HI, USA, December 8-14, 2001, 511-518.
[12] M. Osadchy, Y. L. Cun, M. L. Miller. Synergistic face detection and pose estimation with energy-based models[J]. Journal of Machine Learning Research, 2007, 8(1): 1197-1215.
[13] P. Liu, S. Han, Z. Meng, et al. Facial expression recognition via a boosted deep belief network[C]. 2014 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2014, Columbus, OH, USA, June 23-28, 2014, 1805-1812.
[14] X. Chen, X. Yang, M. Wang, et al. Convolution neural network for automatic facial expression recognition[C]. 2017 International conference on applied system innovation (ICASI), Sapporo, Japan, May 13-17, 2017, 814-817.
[15] A. T. Lopes, D. E. Aguiar, T. Oliveira-Santos. A facial expression recognition system using convolutional networks[C]. 28th SIBGRAPI Conference on Graphics, Patterns and Images, SIBGRAPI 2015, Salvador, Bahia, Brazil, August 26-29, 2015, 273-280.
[16] I. Choi, H. Ahn, J. Yoo. Facial expression classification using deep convolutional neural network [J]. Journal of Electrical Engineering&Technology, 2018, 13(1): 485-492
[17] Liu C, Freeman W T, Adelson E H, et al. Human-assisted motion annotation[C]. //Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on. IEEE, 2008:1-8
[18] Akihiro I, Nobutaka S, Yoshiaki S. Computer Vision - ACCV 2007[M]. Springer Berlin Heidelberg, 2007
[19] Abd-Almageed W, Davis L. Human detection using iterative feature selection and logistic principal component analysis[C]. //Robotics and Automation, 2008. ICRA 2008. IEEE International Conference on. IEEE
[20] Xie L, Pan W, Tang C, et a1. A pyramidal deep learning architecture for humanaction recognition [J].Journal of Modelling Identification and Control,2014, 21(2): 139.146
[21] Gaves A, Mohamed AR, Hinton G. Speech recognition with deep recurrent neural networks. 2013 IEEE International Conference on Acoustics, Speech and Signal Processing. Vancouver: IEEE, 2013. 6645–6649.