PAPER • OPEN ACCESS

Research on Virtual Roaming Technology of Urban Scene Based on Multi-mode Human-Computer Interaction

To cite this article: Chang Zhang *et al* 2020 *J. Phys.: Conf. Ser.* **1570** 012096

View the article online for updates and enhancements.

Research on Virtual Roaming Technology of Urban Scene Based on Multi-mode Human-Computer Interaction

Chang Zhang[1][*], Rumeng Lv[1], Xiaobing Chen[1,2] and Yujin Jiang[3]

[1]Faculty of Computer and Software Engineering, Huaiyin Institute of Technology, Huai'an, Jiangsu Province, 223003, China

[2]Key Lab of Advanced Manufacturing Technology of Jiangsu Province, Huaiyin Institute of Technology, Huai'an, Jiangsu Province, 223003, China

[3]School of management and engineering, Huaiyin Institute of technology, Huai'an City, Jiangsu Province, 223003

[*]Corresponding author's e-mail: 139940116@qq.com

Abstract. Virtual reality technology refers to the use of computer multi-body technology to simulate the environment from many aspects, such as human smell, vision, touch and hearing, so that people can have a sense of being in it in the virtual environment constructed by computer, and can obtain all kinds of information with the help of human-computer interaction.By combining virtual reality with somatosensory interaction using gesture recognition, speech recognition and other modes, this paper develops an immersive urban scene roaming system based on Unity3D multi-mode human-computer interaction, which gives users a near-real scene roaming and somatosensory interaction experience.

1. Introduction

Virtual reality (VR) technology has also received great attention and development. This technology is to combine computer, multimedia, simulation and other technologies to construct a virtual 3D environment of simulation, which can make people produce a sense of reality on the spot, and play an important role in stimulating people's innovative thinking and imagination. At present, virtual visual system has been widely used, and its value has become increasingly prominent.Compared with the traditional VR project, the Unity3D multi-mode urban scene virtual roaming system studied in this paper adds the functions of gesture recognition, speech recognition and so on. It uses a variety of input methods to interact with the virtual scene, combines virtual reality with somatosensory interaction, and gives users a near-real scene roaming and somatosensory interaction experience. This project is based on the existing VR equipment, using 3dsMax to build 3D scenes and restore scene and character movement and interaction based on Unity3D engine. Unity3D is a mature 3D interactive engine that can calculate many kinds of texture and support many kinds of dynamic 3 effects.

In 3DMAX modeling,the use of editable polygon modeling improves the editability, practicability and compatibility of modeling;the use of standard switching makes the model more round. Using the relevant knowledge of human body structure, the proportion of human body and the morphological characteristics of each part are constructed to make the model more in line with the aesthetic sense of human body shape. In terms of accuracy, there is no excessive pursuit of high accuracy, nor too rough,

in the two take a trade-off, that is, to save a certain degree of accuracy, so that the model is complete, angular, vivid, and will not let the model have too many faces to lead to loading difficulties, in practical use to make the whole carton off frames, and so on.

2. Virtual scene roaming and interaction

When the 3D model is created, it is necessary to use the VR engine to quickly build the scene, carry on the scene drive, render the model and scene in real time, add all kinds of special effects, create VR interactive operation, and control the display of the 3D scene[1].Unity 3D and Unreal Engine 4. are the most popular VR engines in the market Unreal Engine engine and Unity engine are the mainstream virtual reality development tools in virtual reality industry development, with a total market share of up to 95%[2].

2.1Real-time rendering

Real-time rendering is to make the 3D scene real and fast display on the screen, the picture is smooth, interactive, strong sense of reality. 3D scene rendering is a more complex programming work, in order to reduce the difficulty of programming, need a good 3D graphics rendering engine.

2.2Collision detection

In the virtual scene, each model object is independent of each other, and it is necessary to simulate the reaction between characters, objects, and objects in the real scene when roaming, which is called collision detection. For example, when people encounter obstacles such as trees, walls, doors, or close contact with obstacles in real scenes, they will stop moving forward and react. Similarly, after setting up collision detection in the virtual scene, it should also produce such an effect, if there is no setting of collision detection, there will be a phenomenon of passing through the wall, which will be distorted. Therefore, collision detection technology is very critical, collision detection effect will affect the virtual scene roaming reality and lifelike.

3. Virtual Scene Roaming Technology for Multi-mode Man-machine Interaction

3.1Menu control based on gesture recognition

Gesture recognition belongs to a topic of computer science, which is mainly realized by mathematical algorithm. Gesture recognition can be defined to recognize the movement of various parts of the body, but in the current practical aspect, it is mainly suitable for the movement of the hand and face. If speech recognition is to make the machine understand what you are saying, then gesture recognition is to make the machine understand what you are doing. In the current time, the main use is the combination of hardware to achieve gesture recognition. This main introduction mainly introduces the calculation of the coordinate address of the opponent's knuckles in the three-dimensional coordinates to judge the bending degree of the fingers and to judge the simple gesture posture. Gesture control menu control. Further gesture operation is supported with Leap Motion hardware support. Gesture operation is developed separately. Image analysis of gesture actions recognized by hardware so as to map gesture corresponding actions.

Fig.1 Menu control of gesture recognition

Above for instance operation, in the case of open palm, the mapping action is returned to the superior menu.

In this paper, the gesture recognition is mainly used to recognize the static gesture, and the image acquisition of the gesture is carried out in each frame, and the judgment recognition is made according to the constructed algorithm[6].

The main steps are as follows.

Step 1: Gesture Recognition Algorithm

Step 2: Determine finger status

Step 3: Judge palm gestures

Step 4: Pass the identifying success information to other functions

3.2 Scene interaction based on speech recognition

Interaction has good real-time performance, VRML language has extremely strong interaction performance, can be divided into two categories: First, interactive nodes constitute field of view, sensors and interpolation nodes, and so on. Second, programming nodes can be formed by means of programming. In the former interaction link, the events formed by the system are captured by the interaction nodes and then processed. Route/T subsequently passed to the scene so that the latter forms a change[3].

This project mainly uses the interaction way is the speech interaction and the gesture recognition two ways. We all know that sound is actually a wave, common audio files in various formats such as MP3 and other formats through the use of AU and other modification software for editing, are presented a waveform state for processing. In speech recognition, the moving window function is used to divide the sound file into frames, and the sound file is divided into many small segments called frames. Then the waveform transformation is carried out, and the waveform transformation without descriptive ability is changed into different forms according to different waveform transformation modes. A common waveform transformation is MFCC ,M FCC transforms the waveform of each frame fragment into a multidimensional vector according to the physiological characteristics of the human ear, and store the information of the sound contained in this frame

segment in this vector. In this transformation, sound is converted from a waveform to a matrix of multiple rows N columns (N as the total number of frames).

Speech recognition used in this paper is to mount baidu speech recognition REST API to transform speech to text. Baidu's speech recognition technology adopts a simpler and more effective method than the current mainstream speech recognition system[4]. They adopt a neural network-like deep learning algorithm to replace the previous recognition module, thus greatly improving the recognition efficiency. Speech recognition of the prescribed direction of automatic walking for roaming experience.

For studying the application of speech recognition in virtual reality, we adopt the speech recognition of baidu cloud API, in the case of networking, we can use space bar to record recording and recognition function combined with fixed direction walking to achieve voice control function.
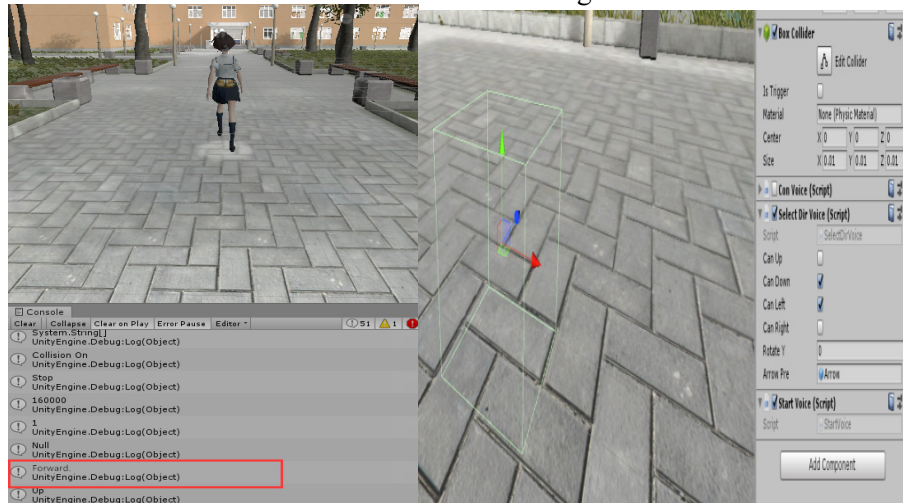

Fig.2 Scene interaction of voice recognition

At each node that needs to judge the walking direction, we set some nodes with the same direction, and modify the Selection Direction code on them to support the input of voice control[5]. The startvoice script is added for speech recognition, and the Con Voice script is used to control the voice recording function of the Start Voice script.

When we reach the direction node, we only need to press the space bar to make speech recording, and then release the space bar to make speech recognition. When the recognition is successful, Start Voice pass the recognized speech text to the function in the Selection Directon to judge and complete the operation of speech control.

4. Conclusion

Scene virtual roaming system is an extension of VR technology that can create and experience virtual scenes. It uses multi-source information fusion interactive and 3D dynamic scene simulation, can use VR professional equipment to let users interact with the scene, real-time perception and operation of various objects in the virtual world, so as to obtain real experience through vision, touch and hearing.

The scene virtual roaming system can create access ports through the Internet, so that users can experience the target scene remotely. This technology can be applied to real estate, tourism, exhibition, and even entertainment industries. Domestic support for high-tech enterprises has increased in recent years, people's living standards have improved, the high-speed popularity of the Internet of things is conducive to the development of virtual reality technology, VR application technology is also destined to embed more life scenes.

References

[1]   Xu, J.F. (2018) Research and practice of scene interactive roaming based on Unity3D. J.
       Science and Technology Innovation Bulletin .15(34):103+105.

[2]   Kang ,Q. (1999) VR Technology in Simulation Visual Environment. professional committee of marine ship driving, china navigation society. Collection of Maritime Safety Papers of the Marine Ship Driving Committee of the Chinese Maritime Society [C]. China Navigation Society Marine Ship Driving Committee: China Navigation Society ,5.

[3]   High ,V.Z., Li ,J.L., Xie, J.C. (2017) The Design of Panoramic Virtual Reality Roaming System Based on Unity3D. J. Mining Research and Development .37(12):68-72.

[4]   Chen ,C.T. (2017) A study on multi-feature gesture recognition and tracking algorithm in human-computer interaction based on virtual reality . J.Software Engineering ,20(12):23-25.

[5]   Xu, Y. (2013) Design of virtual roaming system based on natural human-computer interaction technology.D. Central South University.27-30.

[6]   Wang, X.L. (2010) Human computer interaction technology of gesture recognition based on computer vision. D.Zhejiang University of technology.9-11.