Research on Natural Human-Computer Interaction in Virtual Roaming

Fuyun Li, Jianping Feng and Mousong Fu *
School of Information Science and Technology, Hainan Normal University, Haikou, Hainan, China
E-mail: *450073933@qq.com

Abstract. With the development of computer technology, human-computer interaction technology has gone through the development process of "command line interface", "graphical user interface" and "natural user interface". This paper explores the use of low-cost Kinect body sensor, in close to the premise of people's habits, let users interact with the virtual scene roaming system in a natural way. Using the Kinect skeleton tracking technology, human posture is recognized through the relative position between joint points, and four kinds of self-defined postures are used to control the change of camera's view angle to realize virtual scene roaming. The algorithm of this method is simple, it realizes the natural interaction between human and computer, and provides reference for the development of related systems in the future.

1. Introduction

As a new technology in the research frontier, virtual reality technology integrates multimedia, computer graphics, sensors, networks and other technologies. Because of its strong visual impact and rich interaction, virtual reality system is widely used in industrial simulation, virtual roaming, industrial design, education and training, medicine, military, aerospace and other fields [1].

As an important branch of virtual reality technology, virtual roaming has been widely used in many fields. Various kinds of virtual scene roaming have sprung up and achieved fruitful results, such as campus virtual roaming, museum virtual roaming, tourist attractions virtual roaming, forest landscape virtual roaming, building roaming and so on, all of which show that virtual scene roaming is immeasurable future and development trend.

With the development of virtual reality technology and the breakthrough of graphics software and hardware technology, human-computer interaction technology also has new content. The current human-computer interaction technology has gradually transferred from the past computer-centric to user-centric. Through the development process from "command line interface" to "graphical user interface" and then to "natural user interface", users are no longer satisfied with the traditional interaction mode of constantly moving mouse and operating keyboard to control the computer, so a new reality based interaction has emerged [2]. At present, the well-known ones include Kinect body sensor of Microsoft, Leap Motion gesture controller of Leap Company, etc.

Kinect body sensor developed by Microsoft can complete natural human-computer interaction without any hand-held or wearable devices. It can interact with the terminal through body movements, gestures, postures, voice and other natural ways. This natural human-computer interaction will bring natural and real feelings to users.

*Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.
Published under licence by IOP Publishing Ltd
This paper explores the use of low-cost Kinect body sensor, in close to the premise of people's habits, let users interact with the virtual scene roaming system in a natural way.

2. Natural human-computer interaction design

2.1. Kinect body sensor
As shown in figure 1, Kinect body sensor includes Colour Camera, Infrared Camera and Microphone. The Colour Camera collects video data with a resolution of 1920 × 1080. Infrared Camera is divided into infrared transmitting and receiving parts. Microphone Array can record language data and locate sound source. Colour image data, depth image data, bone data and audio data can be obtained by Kinect.

![Figure 1. Kinect body sensor.](image)

2.2. Skeleton tracking
Skeletal tracking is one of the core technologies of Kinect body sensor. It can accurately calibrate 25 joint points and track the position of the joint in real time. Skeleton tracking is realized step by step by using machine learning method on the basis of depth image. The first step is human body contour segmentation, which determines whether each pixel on the depth image belongs to a user or not, and filters the background pixels. The second step is the recognition of human body parts. Different parts, such as head, trunk and limbs, are identified from human body contours. The third step is to locate the joints. 25 joints are located from the human body, and the skeleton map is generated as shown in figure 2 [3].

![Figure 2. The skeleton map.](image)
2.3. Posture design

The design of natural human-computer interaction shall follow the following principles:

- Maximum fit action habits: This is the requirement of natural interaction, so that every interaction action is in line with daily action habits, such as raising legs to show walking, left and right waving to show switching, etc.
- Minimum mutual interference: No matter how accurate the algorithm is, there will be some false recognition phenomenon, especially the similar movement posture. Therefore, we should try to avoid similar actions in interaction design, especially avoid using similar actions to represent different operations at the same time.
- Minimum recognition time: For the high error recognition rate of gesture recognition and speech recognition, if it is turned on for a long time, many small actions and small sounds will be recognized as target operations. Therefore, we should minimize the duration of this recognition, preferably using posture recognition, and only turn on gesture recognition when gesture is needed [4].
- Make the movement range of users' limbs as small as possible, and reduce the fatigue and boredom brought by interaction.

Posture is a static relationship expressed by the spatial arrangement of different parts of the body. The position of the body and each joint point defines a posture, that is, the relative position between the nodes can determine a posture. The type and complexity of the posture is directly proportional to the complexity of the recognition algorithm [5]. For a simple posture, by tracking the spatial position between the joint points of the limbs, combined with triangle geometry, it can be matched quickly [6].

Based on the above considerations, the four postures designed in this paper are as follows: left hand to left, right hand to right, left foot forward, left foot backward. The effect of the corresponding scene is shown in Table 1.

| Posture          | Scene effect     |
|------------------|------------------|
| Left hand to left| Towards the left |
| Right hand to right| Towards the right |
| left foot forward| Move forward     |
| left foot backward| Move backward   |

3. Realization of natural human-computer interaction

3.1. Development of somatosensory control program

By analyzing the collected position data of the key joints of the user's skeleton, this paper studies how these key joints move. After establishing the mathematical model in the program, carrying out simple mathematical calculation, the data is finally transferred to the data analysis and processing module to process the movement data of the key joints of the user. Through C# programming, we can achieve the result of roaming in Unity3D with the first person perspective through Kinect. As shown in Figure 3, it is the flow chart of the whole program.
3.2. Posture recognition algorithm

In order to adapt to the users of different height and body shape, and make the software self-adaptive, the left and right shoulder joint points are used to calculate the user's shoulder width, and half of the shoulder width is taken as the threshold value.

"Left hand to left, right hand to right" these two postures use the left shoulder joint point and the left wrist joint point, the right shoulder joint point and the right wrist joint point.

If the difference between "X coordinate of right wrist joint point" and "X coordinate of right shoulder joint point" is greater than the threshold value, the code is:

\[
\text{body.Joints}\[\text{JointType.Wristright}\].\text{Position.X} - \text{body.Joints}\[\text{JointType.Shoulderright}\].\text{Position.X} > \text{Threshold}, \text{it is judged as "right hand to right".}
\]

If the difference between "X coordinate of left shoulder joint point" and "X coordinate of left wrist joint point" is greater than the threshold value, the code is:

\[
\text{body.Joints}\[\text{JointType.Shoulderleft}\].\text{Position.X} - \text{body.Joints}\[\text{JointType.Wristleft}\].\text{Position.X} > \text{Threshold}, \text{it is judged as "left hand to left".}
\]

"Left foot forward, left foot backward" these two postures use the left and right ankle joint points.

If the difference between "Z coordinate of right ankle joint point" and "Z coordinate of left ankle joint point" is greater than the threshold value, the code is:

\[
\text{body.Joints}\[\text{JointType.Ankleright}\].\text{Position.Z} - \text{body.Joints}\[\text{JointType.Anklleleft}\].\text{Position.Z} > \text{Threshold}, \text{it is judged as "left foot forward".}
\]

If the difference between "Z coordinate of left ankle joint point" and "Z coordinate of right ankle joint point" is greater than the threshold value, the code is:

\[
\text{body.Joints}\[\text{JointType.Anklleleft}\].\text{Position.Z} - \text{body.Joints}\[\text{JointType.Ankleright}\].\text{Position.Z} > \text{Threshold}, \text{it is judged as "left foot backward".}
\]

3.3. Roaming control

In the roaming control, four user postures are used to control the camera angle of view. When the "left foot forward, left foot backward" posture is detected, the camera's movement is set by the position attribute; when the "left hand to left, right hand to right" posture is detected, the camera's rotation is set by the rotation method. When the camera rotates, set the camera to rotate along the Y axis to achieve horizontal rotation.
4. User experience and result analysis
During the test, the user makes four postures towards the camera, and roams the virtual scene from the first person perspective, the effect is shown in Figure 4.5.6.7.8. In this paper, posture interaction is used to replace the traditional mouse keyboard interaction. The experimenter evaluates the scene roaming interaction design from the accuracy of action recognition, the rationality of action design, and the real-time nature of interaction, and feels the interaction effect in the process of virtual scene roaming. As a whole, the comprehensive experience effect of the software is relatively ideal.
5. Conclusion
In this paper, we design and implement a virtual roaming system based on virtual reality technology and Kinect body sensor. The posture interactive control method is added to the virtual roaming system innovatively, which makes the system separate from the external control hardware and truly realizes the natural interaction between human and computer, which provides a reference for the development of related systems in the future. The human-computer interaction method of virtual roaming system developed in this paper can be widely used in all kinds of virtual scene roaming, and has a certain application prospect. The next step is to add voice interaction on the basis of posture interaction, and combine the two interaction methods to explore more natural human-computer interaction.

Acknowledgments
Thanks to the support by Hainan Province Natural Science Foundation Project (No. 617124), National Natural Science Foundation of China (No. 61562022), Hainan Province Natural Science foundation(No. 617119), Scientific Research Projects of Higher Education Institutions in Hainan(No. Hnky2017-17).

Reference
[1] Wang Xinglu 2017, Application Research on Somatosensory Interaction Technology Base on Unity3D and Kinect [D]. Lanzhou Jiaotong University.
[2] Wang Liguo 2016, Research on virtual scene construction and interactive technology of cultural theme park [D]. North University of China.
[3] LiuKaiyu, Xia Bin. Real-time human posture recognition based on Kinect. Electronic Design Engineering, 2014,22(19): 31-34.
[4] Wu Guobin, Li Bin,Yan Jizhou. KINECT human-computer interaction development practice. Beijing: POSTS & TELECOM PRESS, 2013: 95-98.
[5] Xie Liang, LiaoHongjian, YangYubao. Recognition and Application Research of Kinect-based Gesture. Computer Technology and Development, 2013, 23(5): 258-260.
[6] Yu Tao. Kinect application development practice (talk to machines in the most natural way). Beijing: China Machine Press, 2014: 242-253.