WBFS: An Interactive Virtual Simulation Landscape System Based on Kinect

Yuecheng Peng*and Qiuhai He
Beijing Forestry University
Beijing, China
*Corresponding author

Xinyuan Huang
Communication University of China
Beijing, China

Abstract—**Using gesture to implement 3D navigation still exists many problems to be solved. This paper introduces a solution about gesture interaction in 3D virtual scenes. In order to protect the intangible culture heritage of Dai Ethnic, a nationality of China, we designed and developed the interactive virtual simulation landscape system "Within the Bamboo Forest (WBFS)", where anyone can enjoy a virtual experience of travelling in a village of Dai Ethnic with a free tourist guide. At the respects of technology, we used a Kinect sensor as the bottom hardware layer to collect depth data, OpenNI and NITE plugins as the middle layer to analyze user's gestures, Unity3D Game Engine as the system top layer to integrate art resources and develop interfaces. This solution mainly achieves three aspects of interacting with Kinect sensor, they are sliding main-menu selection, roaming in 3D scenes and clicking 3D objects and elements in scenes.**

Keywords-gesture interaction; virtual reality; Kinect; Unity3D

I. INTRODUCTION

Since the release of the Microsoft Kinect device, there have been several efforts towards integrating art and motion-tracking technology. ROCCETTI et al. describe the design and implementation of a multimedia system that interactively illustrates the process of preparing one of the masterpieces of the culinary heritage of the Italian city of Bologna, the Tortellino [1]. With their system, anyone can enjoy a virtual experience learning how to prepare a Tortellino. What's more creative? JongHwan Oh et al. presented a Hands-Up system [4], where they utilized ceiling to display information and to interact with computers by using Microsoft Kinect, which could give commands to the computer through a few hands motion.

At the respects of interaction with gestures, Gilles Bailly and his colleges introduced Finger-Count gestures in 2012 [2], which was a coherent set of multi-finger and two-handed gestures. And also they proposed the Finger-Count menu, a menu technique and teaching method of implicitly learning Finger-Count gestures. In modern gesture recognition research, Kinect sensor is widely used. David Fleer and Christian Leichsenring presented an unobtrusive multimodal interface for smart objects (MISO) in an everyday indoor environment [3]. MISO uses pointing for object selection and context-sensitive arm gestures for object control. Finger snaps are used to confirm object selections and to aid with gesture segmentation. This system mainly relies on the Kinect depth camera.

II. DESIGNE OF THE SYSTEM

A. Background

The Intangible Cultural Heritage of Dai Ethnic contains many kinds of culture, such as, local folk architecture, the unique Peacock Dance, the traditional custom, manual papermaking technology and so on [9]. They are listed in the first China Intangible Cultural Heritage List. However, in the context of economic globalization, like other minorities, the abundant intangible culture heritage of Dai Ethnic faces confusion of survival and development. In this historic moment, protection has to be put on the agenda. The significance of this project is to create a 3D village of Dai Ethnic. Tripping in this village through large screen, users can be interested and immersed in it. At the same time, they can easily get the knowledge of local culture.

B. Design of the Dai Village

According to the local architecture style and natural landscape feature, we re-designed and built the virtual simulation village of Dai Ethnic. This village is assembled by many key elements culture of Dai Ethnic, such as the residents' houses, Octagonal Pavilion, water tower, central stone of village with totem, the main temple of local religious named Dragon Tower and a number of local life and production tools. There are eight interactive points in the scene, which is used to introduce the related culture. In addition, villages of Dai Ethnic in real-world are located in unique natural environment. We design this village with a river across it and bamboo forests around it. In order to make the village more realistic, we built it according to the proportion of 1:1. Figure I shows the concept map of the village.
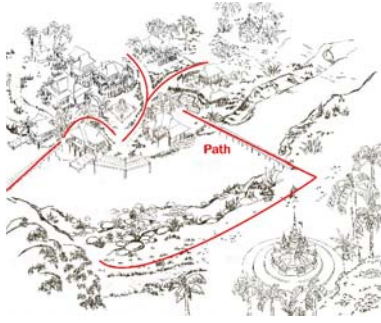
FIGURE I. THE CONCEPT MAP OF THE VILLAGE

At the beginning, we designed two shapes of the trip routes. One is a line shape and the other is a circle shape. In order to test which one is better, we use four parameters to evaluate these two programs. They are total time of roaming, average time of meeting one interactive point, total length of roaming and natural landscape contents. Table I shows the result of this experiment.

TABLE I. RESULT OF SELECTING ROUTE EXPERIMENT

	Line	Circle
Total time[s]	125	197
Average time of one point[s]	15.6	24.5
Total length[m]	150	236
Natural landscape contents	Bamboo forests	River, wood bridge, bamboo forests

As we can see in Table 1, the line route shape takes less total time and average time of meeting one interactive point than the circle route. WBFS will be used in a museum, there may be more people to use it in one day. However, users need time to read the culture introduction. In the line route, the interactive points are too compacted and this will reduce interesting of roaming. The circle route contains more natural landscape. Finally, we decided to use the circle route as Figure I shows.

C. Setup of WBFS

WBFS is an interactive virtual simulation landscape system, which is based on Kinect. It consists of Kinect hardware (Version 1.0), one PC (Intel® CoreTM i5- 3470 CPU @ 3.20 GHz × 2; 4 GB RAM; GeForce GTX 670 1 GB) and one projector. The Kinect points at users and captures depth data. Within the Bamboo Forest software application is installed on PC, it processes user's control and renders scenes in real-time. Figure II shows the setup of WBFS.
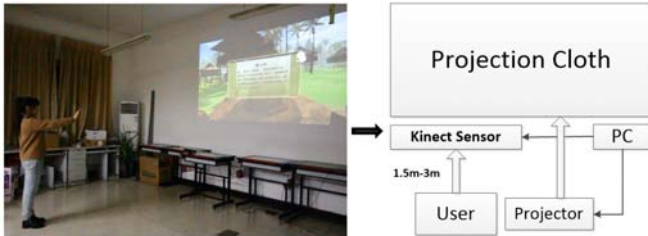

FIGURE II. SETUP OF WBFS

III. IMPLEMENTATION

A. WBFS Development Framework

The development framework of WBFS mainly includes three layers. Figure III displays a three-layered view of the WBFS development concept with each layer representing an integral element. In the bottom layer is Kinect hardware, which is used to capture the visual and audio elements of the real world. In the middle layer is OpenNI framework [10], which defines APIs for writing applications utilizing Natural Interaction [6]. In the top layer is Unity3D Game Engine, which is used to integrate models, materials, lights and graphical user interface, to render pictures and to develop the final application [8].
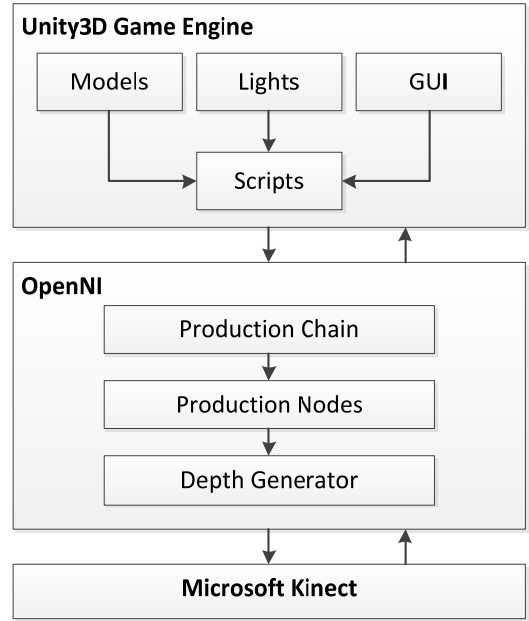

FIGURE III. WBFS DEVELOPMENT FRAMEWORK

When WBFS works, firstly, the application sends a specific message to OpenNI. This message calls a top Production Node of a Production Chain. This is the node that outputs the required data on a practical level, for example, a hand point generator. Secondly, sensor module and middleware components in Production Chain will call a series of corresponding Production Nodes. Then, these Production Nodes will call the corresponding Data Generators, such as DepthGenerator, IRGenerator, etc [5]. These Data Generators analyze original data Kinect hardware captures, and return the processed data to Production Chain. For example, if the Data Generator is DepthGenerator, it will obtain X frames stream data per second and then export depth image to Production Chain. The middleware components in Production Chain are used to package these data and wait for application calling it. So we don't need to process these data manually and we build the Production Chain and call the specific Production Node needed.

B. OpenNI Interface

As Figure III shows, we need to develop an interface between Unity3D and OpenNI. This interface will be called as a plugin in Unity3D. As what are needed in WBFS, we create DepthGenerator to read the depth image, UserGenerator to detect or delete image, GestureGenerator to call functions about body skeleton data and Hands Generator to call functions about hand point tracking [7]. We use callback function to read data from generators. The most important function of this plugin is to recognize user's hand and return hand's 3D position. To do this, we need to recognize some hand gestures to activate the hand tracking. The gestures recognized by OpenNI are: Wave, Swipe, RaiseHand and Click (push forward and back). When a new hand is created, mark down the first hand point then keep updating the hand point. The most two important callback functions are Hand_Create and Hand_Update. Moreover, Registrations for callback functions have to be made to different generators. For example, Gesture_Recognized and Gesture_Process are registered to GestureGenerator, Hand_Create, Hand_Update and Hand_Destroy are registered to HandGenerator.

C. Main-menu Selection

The main-menu is separated into four parts. Each part represents different menu content, such as outdoor roaming, indoor navigation, operating instructions and animation show. We use a slider to switch in each part. In order to avoid improper operation of using this system, we use the push-in gesture to unlock control and the put-down gesture to release control. Once users unlock control by push-in gesture, the slider control will be active at the same time. Users can very easily control the slider by just waving their hands. We filter the hand point data in vertical and depth direction, leaving only the displacement value of horizontal direction. Therefore, the slider can move in the horizontal direction in a scroll bar. There are three triggers at the both ends and middle in it. When the slider touches the triggers at the both ends, the menu will be switched. If the slider stops on the middle trigger longer than two seconds in outdoor roaming or indoor navigation, it means go into 3D scenes and start roaming. Figure IV shows the main-menu.

FIGURE IV. MAIN-MENU SELECTION

D. Curser Moving Control and Click

The windows default curser is very small and tedious. Users stand about two meters or more before the Kinect sensor, so that it is very inconvenient using the small curser. We use a 20*20 pixels picture of a hand instead of the default curser (from here on, referred to as hand-curser). Although we gather the displacement data of users' hand in real world, we can't use the data directly in Unity3D Game Engine. As the real world Units and virtual world Units are different, the displacement data captured in real world need to be transformed. For better user experience, we did a lot experiments to confirm an appropriate scaling variable. With the help of Update Function, hand-curser will move on screen per frame according to users' hand movement in real time.

As Kinect is the only input hardware in WBFS, users can't use mouse to click button. In order to solve this problem, we design a click response mechanism. When hand-curser moves on a button, there will be a halo around hand-curser. After 1.5 seconds, it indicates clicking this button. However, hand-curser is just a texture and the buttons are GUI components. It can't be used as default curser to execute mouse related functions, such as OnMouseOver Function. The solution is to transform the virtual world coordinate of hand-curse to the screen coordinate in real time. Use its moving region insisting of four angles to touch the button on screen.

E. 3D Scene Roaming and Objects Click

In order to achieve the purpose of both roaming and clicking, we design four response regions on screen. The up and down regions indicates scene camera moving forward and backward. The left and right regions indicates scene camera rotating left and right. The moving and rotating speed of camera is two meters per second and five degree per second respectively. At this speed, users can control comfortable. Generally, users can't see these four regions on screen, only after hand-curser moving on them (as Figure VI shows).

The 3D objects in virtual world are not points. They are rendered as a region on screen and this region is continuously changing with the scene camera moving. Therefore, we can't use the same method as clicking button on screen to click 3D objects. The solution is to create a ray between camera and center of gravity of object. And then monitor the collision between hand-curser and the ray. When the collision occurs, a pop-up window will display on screen, which is an introduction of the culture about the objects in the village of Dai Ethnic.

FIGURE V. FOUR REGIONS AFTER HAND-CURSER MOVES ON
THEM

IV. USER EXPERIENCE STUDY

The user experience study attempted to answer the question of whether our system allows people to comfortably and successfully roaming in virtual 3D scenes and have a general understanding of culture about Dai Ethnic. Ten tourists were recruited from a tour team visiting in Yunnan, the place Dai Ethnic people lives. Subjects were required to walk freely in the virtual 3D scenes and find as many interactive points as they want. After the roaming was finished, subjects were directed to an online survey which they completed after the experimenter left. With the feedback of using system, we evaluate it through how many interactive points they found, the feeling of controlling the system, how much culture have they learned and how about the natural environment.

V. RESULTS AND DISCUSSION

Subjects in the survey reported feeling successful using WBFS, and reported high levels of ease and pleasure, and low levels of frustration. They also felt the current implementation, provided it were able to roam and click what they are interested, was nearly as helpful as they could imagine the interaction style being generally.

Since more and more game engines are free to developers, more application in the field of 3D scenes roaming are developed. With the help of Microsoft Kinect and OpenNI plugin, independent developers could try to make more imaginative interactive applications. WBFS can be seen as a kind of serious game. It has greater social significance than its technical contributions.

VI. CONCLUSIONS AND FUTURE WORK

This paper proposes a solution about developing 3D scenes roaming based on Kinect. In order to achieve the targets of the project, lots of experiments are done to improve user experience in this solution. By developing Within the Bamboo Forest System, this paper elaborates many design details in controlling with Kinect. It transforms the original scene roaming by mouse and keyboard input to large screen interaction. However, there are still many respects to be improved in interaction methods, such as the more easily and abundant gestures.

As an achievement of project Application and Research of New Media Technology on the Protection of Intangible Cultural Heritage, this application will cooperate with local government and the tourism industry in the future. It will face to more people to show the digital Intangible Cultural Heritage of Dai Ethnic. This novel interaction will be applied in more projects and generate more social value.

REFERENCES

[1] Roccetti M, Marfia G, Zanichelli M. The art and craft of making the Tortellino: playing with a digital gesture recognizer for preparing pasta culinary recipes[J]. Computers in Entertainment (CIE), 2010, 8(4): 28.

[2] Bailly G, Müller J, Lecolinet E. Design and evaluation of finger-count interaction: Combining multitouch gestures and menus[J]. International Journal of Human-Computer Studies, 2012, 70(10): 673-689.

[3] Fleer D, Leichsenring C. MISO: a context-sensitive multimodal interface for smart objects based on hand gestures and finger snaps[C] // Adjunct proceedings of the 25th annual ACM symposium on User interface software and technology. ACM, 2012: 93-94.

[4] Oh J H, Jung Y, Cho Y, et al. Hands-up: motion recognition using kinect and a ceiling to improve the convenience of human life[C] // CHI'12 Extended Abstracts on Human Factors in Computing Systems. ACM, 2012: 1655-1660.

[5] Panger G. Kinect in the kitchen: testing depth camera interactions in practical home environments[C] // CHI'12 Extended Abstracts on Human Factors in Computing Systems. ACM, 2012: 1985-1990.

[6] Kin K, Miller T, Bollensdorff B, et al. Eden: a professional multitouch tool for constructing virtual organic environments[C] // Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, 2011: 1343-1352.

[7] Alexiadis D S, Kelly P, Daras P, et al. Evaluating a dancer's performance using kinect-based skeleton tracking[C] // Proceedings of the 19th ACM international conference on Multimedia. ACM, 2011: 659-662.

[8] Indraprastha A, Shinozaki M. The investigation on using Unity3D game engine in urban design study[J]. ITB Journal of Information and Communication Technology, 3 (1), 1-18., 2009.

[9] Bo Dao. Introduction of the Intangible Cultural Heritage of Dai Ethnic. The Ethnic Publishing House, 2010.

[10] OpenNI Organization, Programmer Guide [EB/OL]. http://www.openni. org.