

AR Interaction System Is Recognized Based on Kinect's Motion Depth

Shuwen Wang, Zhang Lu*, Xie Xiang

College of Electrical and Information Engineering, Quzhou University, Quzhou, Zhejiang, 324000, China

Abstract

Kinect-based motion depth recognition AR interaction system is designed for Kinect's motion depth recognition AR interaction management system, the main functions of the system are AR scanning requests, feedback, etc., users can use the system to the motion depth recognition AR interaction of all aspects of information to view, but also can carry out other setup operations.

Keywords

Kinect; Interactive system; AR scan; Depth recognition.

1. Introduction

With the development of computer vision technology, the visual-based human-computer interaction intelligence continues to improve, the use of human gesture movements, speech, and computer environment to establish a similar human-human human-computer interaction mode, compared with the traditional mouse, keyboard and other equipment communication, can reduce the hardware to human-computer interaction constraints and limitations, improve the simplicity of information exchange, natural. In recent years, many scientific research institutions have carried out in-depth research on visual-based contactless gesture perception technology, and there are new modes of manipulation in the fields of body-based entertainment games.

At present, contactless human-computer interaction in medical, educational and entertainment and many other aspects have many needs and applications, human-computer interaction exploration and research will continue to deepen. This topic is mainly based on Kinect sensor acquisition and processing of human body color data, depth data and bone data and other motion information, gesture recognition and human motion recognition research, and the above research results are used for Kinect-based panoramic virtual scene human-computer interaction system implementation. This topic takes gesture and human body movement recognition as input information, applies to the display and introduction of Huizhou Cultural Pavilion, as well as the roaming medium of Huishang Mansion VR.

2. Introduction to Kinect

Kinect is a self-developed by Microsoft, a body-feeling peripheral product, officially launched in 2010, the original purpose of the development is to allow player users to have a new human-computer interaction experience. Use the user's own body movements to interact with the device or scene around the target object. Kinect was mainly used in previous years for body games, but as the need for people's lives grew, researchers and amateurs began to work on implanting sensory technologies in industry, medicine, and everyday life, such as robotic control through body-feeling technology, da Vinci surgical robots, and so on.

The Kinect v2 features human features with "three eyes" and "four ears": three eyes are a color camera, a depth (infrared) camera, and an infrared projector. The four ears are a quad linear

microphone array with sound from four microphones, built-in components such as digital signal processor DSP, and filter background noise to locate the sound source direction.



Figure 1. Kinect hardware construction

Kinect's bone tracking is one of its core technologies, which accurately calibrates 25 joint points in the human body and tracks the location of joint points in real time. Bone tracking is implemented step by step using machine learning methods based on deep images. The first step is to split the outline of the human body, determine whether each pixel on the depth image belongs to a user, filter the background pixels. The second step is to identify human body parts, from the contours of the human body to identify different parts, such as the head, torso, limbs and other limbs. The third step is joint positioning, locating 25 joint points from human parts to produce bone diagrams as shown in Figure 2.

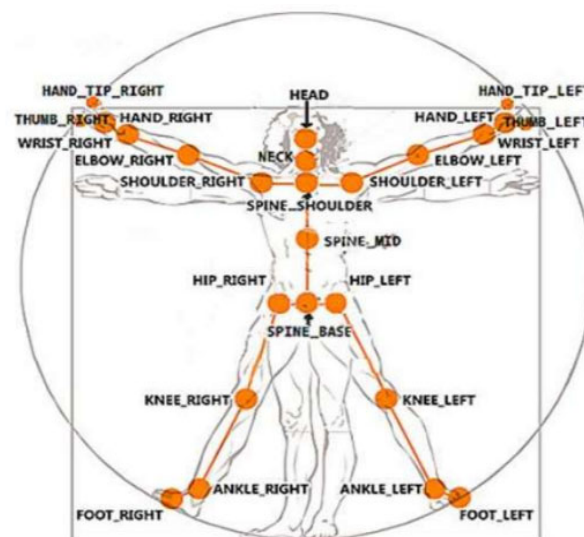


Figure 2. Bone Picture

3. Action Recognition Scheme

3.1. Second Detection Feature - Distance from Ground

The rate of decline at the center of the target human body is the first detection feature, and if the first detection feature is detected, the second feature, the height of the two hip center points from the ground, is further detected. Ground detection is difficult, by querying Kinect SDK can be found in the IBodyFrame class there is a get_FloorClipPlane method, the parameters of the function is vector4 splaclipPlane, the return value is HRESULT type. There are four float-type data members in the vector4 structure, x, y, z, w, the four parameters are the coefficients of the ground equation (x, y, z) and constant (w), and the four numbers are assigned to A, B, C, D, respectively, and the ground equation is $Ax-By-Cz-D=0$

The ground equation here is obtained under Kinect's depth camera coordinate system, and the constant items represent the distance from the center point of Kinect's depth camera to the ground. Knowing the three-dimensional coordinates of the bone point and the ground equation, calculate the distance between the center of the two hips from the ground.

$$baseH = \frac{|Ax_i + By_i + Cz_i + D|}{\sqrt{A^2 + B^2 + C^2}}$$

Tests have shown that these four parameters are not always accurately obtained, constants (w) are often not obtained correctly, and the system automatically replaces 0, which is clearly not correct and can lead to fatal errors in high-level feature detection.

Therefore, this project approximates the use of the right foot to represent the ground as an alternative, because in daily life, the human foot is generally on the ground, so as to calculate the height of the two hip center point from the right foot baseH. When $baseH < HT$ (critical value), a second detection feature is considered detected.

3.2. Bone Imaging and Sensing Technology

Skeleton tracking relies on Kinect v2's infrared sensor, which is primarily a unique black-and-white spectroscopic technology to sense changes in the outside environment and to represent the distance between the target object and the camera in different colors. When the Kinectv2 device is operating normally, the entire visual range of the Kinect v2 camera will be fully covered by the laser emitted by the infrared transmitter, and then use optical coding technology to encode the visible range of space, optical coding is in fact to the space to be measured, generally using the method of lighting, in essence, is still a structural light technology.

Kinect v2 can detect bone point information for up to six people at the same time, including the identification of two people's movements, each person can record 25 bone points in detail, the camera obtains data range including limbs, torso and fingers, etc., to build a full body-feeling operating system. In order to be able to see the user's movements more carefully, the system uses machine learning methods, on the basis of which the Kinect v2 camera can identify the user's behavior information to the maximum extent possible.

3.3. AR Takes Color Data and Displays

Select "AR get color data and display" to enter AR to get color data and display theme, the same, you can move the phone and zoom in screen, switch browsing to get a point on the background two-value map, this point contains its corresponding coordinates to the depth map, photographic function, right hand forward 0.4m can trigger the camera, and can slide right hand to switch background map.



Figure 3. Gets color data

3.4. Identify Bone Points

Using Kinect's real-time bone tracking technology, select the three key bone points of the human body's center point (spinemid), two hip centers (spinebase), right foot (rightfoot), and calculate the spatial position, speed of motion, and the spatial position of the center point of the two hips, the height of the ground, and the time to stay at lower altitudes.

By monitoring the target human body, the returned bone data is processed in real time, and the drop speed spine V is found at the center point of the human body once every 10 adjacent frames. The speed threshold V is then set for the corresponding movement behavior. If the threshold for an action is exceeded, the related event is triggered. In order to detect accuracy, try to stand in the right place and let Kinect get the bone points of the whole body.

3.5. Real-time Processing of Deep Images

Coloring depth means that different colors represent different distance information so that you can see it at a glance. The conversion of color space is involved in the conversion of RGB space and HSV space. The image's expressiveness is improved after processing, but it would be better if you could represent a 16-bit grayscale in 32-bit color. When the RGB value is the same, it appears gray. Grayscale values range from 0 to 255, 0 is black, 255 is white, and the color between is gray. The gray values are now shown in RGB mode;

Select "AR Image Processing" in the main page, go to the theme tour, display AR screen details can click on the corresponding "enter the selective viewing module, you can see the expressiveness of different images to display."

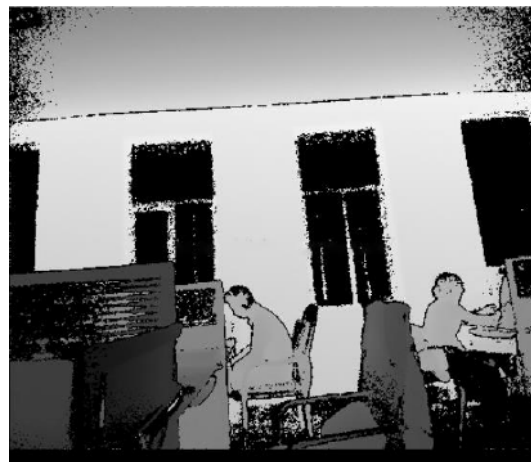


Figure 4. Gets the depth map

3.6. AR Requests A Scan and Save

Click "AR Scan" in the navigation menu bar to first enter the AR scan request, show the status in the request, and after the request is successful, you can begin performing the AR scan. Once the request is successful, you can begin an AR scan, get the currently selected path from the selection folder, and save the depth image. As shown above, through the mobile phone, scanning environment, when moving the phone, will form a scanned motion recognition system track, motion recognition system track is convenient to distinguish and save.



Figure 5. AR requests a scan

Acknowledgments

This work was partially supported by Zhejiang Province Science and Technology Innovation Program for College Students (No. 2020R471005).

References

- [1] Ni Xiaogang, Tian Zhongyang, Li Wangbo. Robotic Control Methods Based on Kinect Vision Functions, Journal, Beijing University of Technology, 2016, 42 (4): 487-491
- [2] Feng Zhiquan, Yang Bo, Xu Tao, etc. Summary Action-Based 3D Human-Computer Interaction Paradigm based on Natural Gesture Tracking, Journal of Computer Science, 2014, 37 (6): 1309-1323.
- [3] Lu Lei, Zhang Jinling, Zhu Yingjie, etc. A static gesture recognition method based on data gloves. Journal of Computer Aided Design and Graphics, 2015, 27 (12): 2410-2418.
- [4] Cai Linxuan, Zhang Jianrong, Liu Binbin. Virtual Environment Sense Interaction Control Based on Gesture Recognition (J. Journal of Huazhong University of Science and Technology (Natural Science Edition), 2015, 43 (S1): 136-139-165.
- [5] Zhang Xuehua, Liu Huaping, Sun Fuchun, etc. Mobile Robotic Target Tracking with Kinect, Journal of Intelligent Systems, 2014, 9 (1): 34-39.
- [6] Chen Hanxiong, Huang Yayun, Liu Yu, etc. Research and implementation of Kinect-based aerial gesture tracking recognition, TV technology, 2015, 39 (21): 91-94.
- [7] Yu Naigong, Wang Jin. Non-contact gesture recognition method based on human arm and joint information. Journal of Beijing University of Technology, 2016, 42 (3): 362-368.
- [8] Tian Congress, Yan Jianqin, Han Xu, etc. A new method of human behavior recognition based on joint point information. robot, 2014, 36 (3): 285-292.
- [9] Yuan Fangjian, Wang Yixuan, Wang Yigang, etc. Fingertip recognition and gesture determination based on Kinect depth images . . . Electronic Technology, 2014, 27 (8): 6-10.