Interactive Design of 3D Dynamic Gesture Based on SVM-LSTM Model

Tao Wang, College of Physics, Chongqing University, Chongqing, China
Xiaolong Cai, Chongqing University, Chongqing, China
Liping Wang, Collage of Mathematics and Statistics, Chongqing University, Chongqing, China
Haoye Tian, College of Software, Chongqing University, Chongqing, China

ABSTRACT

Visual hand gesture interaction is one of the main ways of human-computer interaction, and provides users more interactive degrees of freedom and more realistic interactive experience. Authors present a hybrid model based on SVM-LSTM, and design a three-dimensional dynamic gesture interaction system. The system uses Leap Motion to capture gesture information, combined with SVM powerful static gesture classification ability and LSTM powerful variable-length time series gesture processing ability, enabling real-time recognition of user gestures. The gesture interaction method can automatically define the start and end of gestures, recognition accuracy reached 96.4%, greatly reducing the cost of learning. Experiments have shown that the gesture interaction method proposed by authors is effective. In the simulated mobile environment, the average gesture prediction only takes 0.15 seconds, and ordinary users can quickly grasp this method.

KEYWORDS

3D, Hand Gesture, Human-Computer Interaction, Leap Motion, LSTM, Neural Networks, Real-Time, SVM

INTRODUCTION

With the development of human-computer interaction technology, a variety of new means of interaction are emerging, so that human-computer interaction is moving towards a more natural, efficient and intelligent direction. Visual hand gesture interaction is one of the main ways of human-computer interaction. Compared with the popular keyboard, mouse and interactive physical touch screen, visual gesture interaction can help users get rid of those constraints, and provide users greater interactive space, more interactive degrees of freedom and more realistic interactive experience. Therefore, the gesture interactive technology has attracted increasing attention around the world, and quickly became a hot research field of human-computer interaction. It has been widely used in virtual / augmented reality, pervasive computing, smart spaces, computer-based interactive games and other fields, such as Leap Motion, Microsoft Kinect and Intel RealSence have quietly stepped into the lives of consumers (Si & Hou, 2016; Wang et al., 2016; Chao et al., 2017).

At the same time, the rapid development of augmented reality is more realistic for the interaction of gestures. At present, augmented reality mainly takes the mobile phone and the head-mounted display as a hardware carrier. The head-mounted display has a stronger immersion than the mobile phone, and the overall experience is even better. Different from the mature touch screen interaction
mode on the mobile phone, the head-mounted display will use the natural gesture interaction as the main interaction mode. Because, as compared with conventional virtual reality head-mounted display handle operation, the use of an augmented reality scene is more diverse, more portable equipment requirements. After all, people can accept playing virtual reality games with handle in living room, but do not want to use the handle when operating an augmented reality device outdoors. At the same time, 3D dynamic gestures are one of the most natural gestures of human interaction. Obviously, users want to spend a lower learning costs to use augmented reality device with the natural gestures (Yaoyuneyong, Foster, & Flynn, 2014; Waghmare Amit, Sonawane Kunal, Chavan Puja, & Kadu Nanasaheb, 2014).

Traditional gesture interaction techniques are often based on a statistical learning method or neural network technology, most of these methods have some problems such as poor recognition accuracy, high learning cost and so on. Li et al. (2012) uses Three-point Alignment Algorithm in Kinect to recognize planar gestures with accuracy 84%; Hasan et al. (2014) presents a method for hand gesture recognition using neural networks, which obtain 86.3% accuracy; Dardas et al. (2011) presents a real time system, which includes detecting and tracking bare hand in cluttered background using skin detection and hand postures contours comparison algorithm after face subtraction, and recognizing hand gestures using Principle Components Analysis (PCA) and obtain 90% accuracy; Trigueiros et al. (2013) tries a few classifiers, the highest accuracy classifier is neural networks (accuracy 91%); Lai et al. (2014) presents a method based on convex defect character points of the hand contour, recognizing planar number gestures in real-time with 95.1% accuracy, but planar digital gestures are not natural human-computer interaction gestures; Wang et al. (2013) uses PCA+LDA algorithm in 3d laser scanner to obtain 96% accuracy, but cannot inference in real time.

In view of this, authors use Leap Motion as a hand data collection device, and propose a novel method of 3D dynamic gesture interaction based on SVM-LSTM model. Compared with the traditional method of gesture recognition, this method has the following characteristics: 1. It recognizes dynamic gestures, which is one of the most natural human-computer interaction modes; 2. It collects 3D hand data to recognize gestures with an accuracy of 96.8%, and the accuracy is higher than the other methods that use plane data; 3. It uses SVM-LSTM hybrid model, automatically determining the start and end gestures, and recognition gesture in real-time (about 0.15s); 4. It costs less to learn and common users can quickly grasp this method.

OVERVIEW OF LEAP MOTION

Leap Motion Fundamentals

Leap Motion is a body-sensing controller for PCs and Macs, developed by the American Leap Company. Leap Motion uses a multiple-angle imaging technique, the basic principle of which is to use multiple cameras at different locations to capture object images at the same time, and then use algorithms to process the pictures to get the depth information of the object at that time. Leap Motion’s surface is an infrared filter that filters out visible light interference. Below the filter are two 120-frame-rate cameras with infrared LED in the middle. The light of infrared LED slips through the infrared filter to the hand, captured by the camera. Thus an almost perfect hand image masking the complex background comes into being, as shown in Figures 1 and 2. The detection range of Leap motion is roughly 25 mm to 600 mm above the sensor, and the detected space is generally an inverted four-sided cone, as shown in Figure 3 (Jin, Chen, Chen, Hu, & Zhang, 2016; Leapmotion, 2016).

With the rapid development of virtual / augmented reality, Leap Motion has also started to adapt to a variety of head-mounted displays. Because of its compact size, the Leap Motion needs only to be attached to the head mounted display to track gestures in a mobile environment, as shown in Figure 4 (Leapmotion, 2016).
Happy Measure: Augmented Reality for Mobile Virtual Furnishing
[www.igi-global.com/article/happy-measure-augmented-reality-mobile/76333?camid=4v1a](www.igi-global.com/article/happy-measure-augmented-reality-mobile/76333?camid=4v1a)

Talking Avatar: An Intelligent Mobile Application Based on Third-Party Cloud Services
Feng Ye, Qian Huang, Shengyan Wu and Yong Chen (2019). *International Journal of Technology and Human Interaction* (pp. 1-13).
[www.igi-global.com/article/talking-avatar/227398?camid=4v1a](www.igi-global.com/article/talking-avatar/227398?camid=4v1a)