Perceiving Motion Cues Inspired by Microsoft Kinect Sensor on Game Experiencing

Jiawei Xu¹ and Shigang Yue²
School of Computer Science
University of Lincoln
Brayford Pool, Lincoln, United Kingdom
+44(0)1522837301
jxu@lincoln.ac.uk¹
syue@lincoln.ac.uk²

Ruisheng Wang
Department of Geomatics Engineering,
Schulich School of Engineering,
University of Calgary
Calgary, Canada
+ (403)210-9509
ruiswang@ucalgary.ca

Loo Chu Kiong
Advanced Robotics Lab
Department of Artificial Intelligence Computer Science and Information Technology, University of Malaya,
Kuala Lumpur, Malaysia
+(603)79676314
ckloo.um@um.edu.my

Abstract This paper proposed a novel method to replace the traditional mouse controller by using Microsoft Kinect Sensor to realize the functional implementation on human-machine interaction. With human hand gestures and movements, Kinect Sensor could accurately recognize the participants’ intention and transmit our order to desktop or laptop. In addition, the trend in current HCI market is giving the customer more freedom and experiencing feeling by involving human cognitive factors more deeply. Kinect sensor receives the motion cues continuously from the human’s intention and feedback the reaction during the experiments. The comparison accuracy between the hand movement and mouse cursor demonstrates the efficiency for the proposed method. In addition, the experimental results on hit rate in the game of Fruit Ninja and Shape Touching proves the real-time ability of the proposed framework. The performance evaluation built up a promise foundation for the further applications in the field of human-machine interaction. The contribution of this work is the expansion on hand gesture perception and early formulation on Mac iPad.

Keywords Microsoft Kinect Sensor; Motion perception; Human-machine Interaction; Hand Gesture

I. INTRODUCTION

Mouse controller, no matter wire or wireless, receives user’s instructions, such as click, drag or border selection in a 2-dimensional planar interface. Then the computer receives the signal and appears the feedback on the monitor or screen. Another case is the touch screen on the mobile, whether it is Apple iOS, Android or other intelligent systems, we touch the screen or click the logo via fingers. Both of them have the same feature, the user should touch the mouse or mobile then the agents can perceive the instructions. However, Microsoft Kinect sensor changed the definition of concept [1][2][3][4][5].

Microsoft Kinect SDK is developed by computer scientists or engineers to realize the preliminary human machine interaction application. They could sense our human body movement by perceiving the distance change in depth, which is actually a depth camera. Many open source code realized some basic ideas, such as skeleton recognition, human face surveillance, body tracking or even speech recognition. While in UK or North American market, the price of Kinect is relatively acceptable to customer, about 208 GBP or 330 US dollars, which is an ideal tool for further application.

Fig.1. Replacing the tradition mouse cursor by our hand gesture provides the intuitive feeling of human-machine interaction. We designed the module by Microsoft Kinect which can capture human hand movements correctly.

II. RELATED WORK AND METHODOLOGY

First, we analog the mouse controller in the real 3D world by sensing the human body motion, however, traditional computer mouse controller is a two-dimensional device as it can reflect users control and instruction. A systematic review on gesture recognition or motion perception can be traced to [7], where the authors summarized and categorized previous approaches nicely. In [8], dynamic hand gestures are treated as the motion cues for visual understanding. Later on, recognition of dynamic hand gestures is realized on the conventional camera as shown in [9]. Another way, early formulation of 3D information was combined into the traditional 2D images which lead to the active appearance model in [10]. Our idea is converting the mouse into hand operation while it is free of constraints. This improves the concept on the human-machine interaction, which upgrades the mouse control by using various types of hand movements. Second, motion perception is a key point in the dynamic world. The physical motion draws our visual attention and human body responses the actions to a moving object or a dynamic agent. For example, a frisbee is flying over our head and we will try to pick it. Thus we have a corresponding reaction to the response or decision while our body moves and adjusts to a correct pose to catch it. It is a real complex process and we need to finish it in a very short second. Then the second motivation for this paper is used to realize our hand movements and computer can be perceptible by this kind of motion using Microsoft Kinect sensor. In general, our work can be summarized as following: We replace the traditional mouse controller by our direct hand actions under the real 3D world, and these actions can be perceived by laptop or desktop.
A. Depth Perception

The Kinect contains three vital pieces that work together to detect your motion and create your physical image on the screen: an RGB color VGA video camera, a depth sensor, and a multi-array microphone. The camera detects the red, green, and blue color components as well as body-type and facial features. It has a pixel resolution of 640x480 and a frame rate of 30 fps. This helps in facial recognition and body recognition.

The depth sensor contains a monochrome CMOS sensor and infrared projector that help create the 3D imagery throughout the room. It also measures the distance of each point of the player's body by transmitting invisible near-infrared light and measuring its "time of flight" after it reflects off the objects [6] [11].

B. Hand Gesture Recognition

During this process, an important factor is the motion perception. A sense of motion is crucial for the perception of our own motion in relation to other moving and static objects in the environment. While the motion is mainly composed of moving direction and motion intensity, a complete model for the motion perception can be modulated by the top-down input, which is a kind of prior knowledge, and bottom-up features, such as intensity, depth or noise[12]. For an example, while we are doing experiment on human-machine interaction, like Fruit Ninja, the fruits are coming up from the bottom line of the screen, then this prior information helps our decision and improves our estimation of the emerging objects. Another, the Kinect sensor could perceive our hand actions in the different conditions if we put ourselves under various environment, i.e., indoor or outdoor, high brightness or looming background.

Table I defines our main functions of the hand gesture, each hand is responsible for particular aims.

![Snapshot of the hands gesture during the experiments](image)

**Table I. Categorization and Main Functions of Our Hands**

| Left hand       | Right hand       |
|-----------------|------------------|
| Click/Confirmation | Drag/Localization |
| Cutting | Keeping balance |
| Moving | Rotation |

A framework illustrated the flowchart of our coding module is shown in figure 2. Figure 3 indicate our flowchart scheme in general. While Table II reflects the basic order by our hand gestures, the reference data is the traditional mouse cursor.

**Table II. Accuracy Percentage by the Hand Movements Compared with the Traditional Mouse (Ground-Truth)**

|   | Click | Cutting | Drag | Balancing | Rotation |
|---|-------|--------|------|-----------|---------|
| A | 98.75 | 78.43  | 81.62| 83.34     | 85.24   |
| B | 84.27 | 86.37  | 88.36| 84.41     | 72.42   |
| C | 80.70 | 82.40  | 84.99| 83.88     | 78.83   |
| D | 99.27 | 97.28  | 97.29| 94.47     | 83.67   |
| E | 99.18 | 97.78  | 94.57| 84.84     | 71.13   |
| F | 99.27 | 98.75  | 97.96| 94.77     | 87.23   |
| G | 99.18 | 97.94  | 96.26| 84.67     | 68.27   |
| H | 99.27 | 97.75  | 96.46| 93.86     | 86.38   |
| I | 99.18 | 97.46  | 94.24| 87.23     | 73.92   |
| J | 99.27 | 77.12  | 97.96| 94.56     | 87.23   |
| K | 99.21 | 98.32  | 97.71| 94.03     | 82.32   |
| L | 99.29 | 98.00  | 96.56| 91.05     | 80.02   |
| M | 98.71 | 97.92  | 97.92| 93.72     | 82.33   |
| N | 80.84 | 79.91  | 78.44| 70.33     | 68.51   |

**Table III. Core Code on Mouse Controller**

```csharp
using System;
using System.Collections.Generic;
using System.Linq;
using System.Text;
using System.Threading.Tasks;
namespace AnalogController
{
    public class MouseContorller
    {
        // <summary>
```
The code performs relatively robustly and we have tested it on Fruit Ninja and Shape games. The results are illustrated as below. The precision is an important factor to measure our work. After testing on these games, we find the localization and sensitivity are acceptable, the hit rate is almost same comparing with the traditional mouse controller, which are 78 and 80, respectively. (Using Kinect Sensor/ Using mouse controller). While A detailed performance evaluation will be discussed in section IV. Table IV and Table V gives us the quantitative analysis.

![Image](https://example.com/image.png)

Fig. 4. Our testing is implemented on Fruit Ninja and Shape Game. We stand before the laptop about 5 meters away and replace the traditional mouse controller by our hands gestures.

### IV. DISCUSSION

The primary goal of the current research is to explore the further application on the human-machine interaction. We notice that, a similar technology has already been built and released recently by Leap Motion, Inc (https://www.leapmotion.com/product). However, there is a difference in that Leap Motion created his own hardware, while our project is an adaptation of existing hardware. Hence there is the novelty and potential for this interesting research and technology. We further compare our method with Android mobile application by taking average scores over 100 games of Fruit Ninja and Shape Game.

| Participants | Average Score by Hand Gestures | Average Score Testing on iPad |
|--------------|-------------------------------|-------------------------------|
| A            | 178                           | 159                           |
| B            | 185                           | 167                           |
| C            | 169                           | 178                           |
| D            | 192                           | 179                           |
| E            | 178                           | 191                           |
| F            | 193                           | 175                           |
| G            | 191                           | 194                           |
| H            | 197                           | 175                           |
| I            | 187                           | 196                           |
| J            | 179                           | 172                           |
| K            | 191                           | 198                           |
| L            | 179                           | 188                           |
| M            | 182                           | 189                           |
| N            | 180                           | 179                           |

![Image](https://example.com/image.png)

TABLE IV. Average scoring over 100 games of Fruit Ninja, the middle column indicates the module developed by our method; right column is the score on Mac iPad by 14 voluntary participants.
To summarize, the performance of our module outperforms the average score testing on iPad. The experiment results positively reflect the superiority hand gesture with respect to Fruit Ninja. We also compared our method on Kinect Shape Game by using our module and iPad, experiment results also promise a higher performance. Here we illustrated the hit rate in table V.

TABLE V. AVERAGE HIT RATE OVER 100 GAMES OF SHAPE GAME. THE MIDDLE COLUMN INDICATES THE MODULE DEVELOPED BY OUR METHOD; RIGHT COLUMN IS THE SCORE ON MAC IPAD BY 14 VOLUNTARY PARTICIPANTS.

| Participants | Average Hit Rate Testing on Desktop | Average Hit Rate Testing on Mobile |
|--------------|-------------------------------------|-----------------------------------|
| A            | 76%                                 | 79%                               |
| B            | 68%                                 | 65%                               |
| C            | 74%                                 | 72%                               |
| D            | 85%                                 | 81%                               |
| F            | 91%                                 | 95%                               |
| G            | 72%                                 | 84%                               |
| H            | 87%                                 | 77%                               |
| I            | 75%                                 | 76%                               |
| J            | 82%                                 | 85%                               |
| K            | 67%                                 | 65%                               |
| L            | 74%                                 | 73%                               |
| M            | 68%                                 | 69%                               |
| N            | 78%                                 | 75%                               |

Furthermore, we compared our method with other state-of-the-art methods to evaluate this method’s advantage and disadvantage. Compared with [13], the author advised a method based on the ZCam and an SVM-SMO classifier. Furthermore, the authors in [14] proposed the hand gesture using a range camera with a satisfactory real-time ability. In [15], motion tracking using a range camera and the mean-shift algorithm was combined together to capture the hand gestures. There are still a few methods but hereby we compared these 3 approaches as we simulated them thoroughly on 14 subjects by giving the same game experiencing. Here below is our result indicated in table VI.

TABLE VI. COMPARISONS WITH OTHER STATE-OF-THE-ART METHODS.

| Participants | Average Recognition Accuracy | Miss Rate |
|--------------|-----------------------------|-----------|
| [13]         | 0.17                        | 0.19      |
| [14]         | 0.74                        | 0.21      |
| [15]         | 0.79                        | 0.17      |
| Proposed method | 0.85            | 0.14      |

V. CONCLUSION

The idea of our experiments is straightforward and successful by using Microsoft Kinect SDK skillfully. Actually, application development can be more sophisticated than this and we are still on the primary stage of human-machine interaction. There exist more high level and more cognitive processes in future [16] [17]. We have also tried other applications in the lab, human face tracking, and online fitting room by Unity 3D or the latest innovation, Leap Motion. Anyway, intelligent human-machine interaction by using advanced equipment is the state-of-the-art trend in the upcoming years [18]. Another, as motion cues are the key factor dominated our visual perception by our previous research, it is essential for us to develop new modules by integrating motion cues [19].

Future work will be on the combination of motion cues which can be extracted quickly and robustly to identify attentive objects that are most relevant to the entity that possess the vision system [20] [21] [22].

REFERENCES

[1] http://en.wikipedia.org/wiki/Kinect
[2] http://www.microsoft.com
[3] http://www.openkinect.org
[4] http://www.ros.org/wiki/kinect
[5] K. Berger, K. Ruhl, C. Brummer, Yannic Schr oder, Alexander Scholz, and Marcus Magnor. Markerless motion capture using multiple color-depth sensors. In Proc. Vision, Modeling and Visualization (VMV) 2011, pages 317–324, October 2011.
[6] S. Mitra, T. Acharya, Gesture recognition: A survey, IEEE TRANSACTIONS ON SYSTEMS, MAN AND CYBERNETICS - PART C, (Volume:37 , Issue: 3, pp311 - 324, 2007.
[7] M. Yeasin and S. Chaudhuri, “Visual understanding of dynamic hand gestures,” Pattern Recogn., vol. 33, pp. 1805–1817, 2000.
[8] M. Yeasin and S. Chaudhuri, “Visual understanding of dynamic hand gestures,” Pattern Recogn., vol. 33, pp. 1805–1817, 2000.
[9] J. Xiao, S. Baker, I. Matthews, and T. Kanade, “Real-time combined 2D+3D active appearance models,” Comput. Vis. Pattern Recogn., vol. 2, pp. 538–542, 2004.
[10] T. Dutta. Evaluation of the kinect sensor for 3-d kinematic measurement in the workplace. Applied Ergonomics, October 2011.
[11] B. Freedman. Method and system for object reconstruction. US Patent US 2010/0118123 A1.
[12] B. Lucas. “3D Hand gesture recognition using a ZCam and an SMM-SMO classifier”. Graduate Theses and Dissertations, Paper 10829, 2009.
[13] L. Li, Ray Jarvis, Real time Hand Gesture Recognition using a Range Camera. Australasian Conference on Robotics and Automation (ACRA), December 2–4, 2009.
[14] H. Lahamy, and D.D. Lichti, Evaluation of real-time hand motion tracking using a range camera and the mean-shift algorithm. The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences. 38 (Part 5/W12), 6 pp., 2011.
[15] J. Shotton, A. Fitzgibbon and M. Cook, etc. “Real-time human pose recognition in parts from single depth images,” In Proc. of IEEE Conf. on CVPR, 2011, pp.1297-1304.
[16] K. Khoshelham. Accuracy analysis of kinect depth data. GeoInformation Science, 38(5/W12):1(6, 2010).
[17] Mikkel Viager. Analysis of kinect for mobile robots. Technical report, Department of Electrical Engineering, Technical University of Denmark,march 2011.
[18] J. Xu, S. Yue and Y. Tang, A Motion Attention Model Based on Rarity Weighting and Motion Cues in Dynamic Scenes, International Journal of Pattern recognition and Artificial Intelligence, Vol. 27, No. 6, 2013.
[19] J.Xu and S. Yue, "Mimicking Visual Searching with Integrated Top cues and Low-level Features, Neurocomputing, Volume 133, pp 1-18, 10 June 2014.
[20] J.Xu and S. Yue, Visual Saliency Adaptive Segmentation of Attentive Objects in Dynamic Scenes, accepted by Neurocomputing, In press.
[21] W. Luo, H. Li, G. Liu and K.N. Ngan, Global salient information maximization for saliency detection, Signal Processing: Image Communication 27 238-248, 2012.