

User Interface Construction for Recognition of Gesture and Voice

Sun-Myung Hwang¹ and Ji-Hyun Lee^{2*}

^{1,2}Computer Engineering Dept. Daejeon University
¹sunhwang@dju.kr, ²yeom@dju.kr

Abstract

In this paper, we propose an efficient computing environment which implements the NUI (Natural User Interface) for computer operation by motion and voice and a foundation which can provide for contents of new form. The pc operation system and other computer become possible if the key necked technology is applied. If the KINECT technology is applied different pc operation is possible. Also, we show through the simulation that CUI (Computer User Interface) interacts naturally than the computer operation of the mouse and keyboard base.

Keywords: *computing environment, NUI, KINECT, simulation*

1. Introduction

Recently, by spreading the smartphone and also touch screen which is the most convenient usage for user interfaces, it becomes the spotlight [2]. But the critical point of touch screen is also clearly existence. Among the critical points, one of it is the unchangeable method by whatever you have to do is pressing in order to operate. Also, for the touch screen, the screen keyboard and mouse button also need to press in order to operate. Even it is pressed on screen it is just the button pressing only. That critical point is being examined [1].

By highly persisted concern of voice recognition technology and motion recognition technology to understand the behavior of people, there will no touch of the new construction computing environment without pressing closer on the button just the construction of functional computing environment only. Also, the foundation of content fits its environment is provided and the existence of the user interface is not receiving that just the variety creation of the contents framework only.

2. Requirements for Better Performance

All Not only the simply hand Figure, the human standing behavior capture but also by following capture of the head, hand, leg, *etc.*, and location portion, the other motion for recognizing, KINECT [3, 4] is being used and the simply 2D camera the unable received input of in depth information cannot be received.



Figure 1. RGB Camera Image (left) and Depth Value

* Corresponding Author

3. KINECT Construction and Principle of Operation

KINECT is composed of RGB camera, an infrared recognition sensor, an infrared laser, microphone array and tilt motor. Here the microphone array, the microphone array configuration and tilt motor is the motor that allow you to adjust the angle of the KINECT [5].

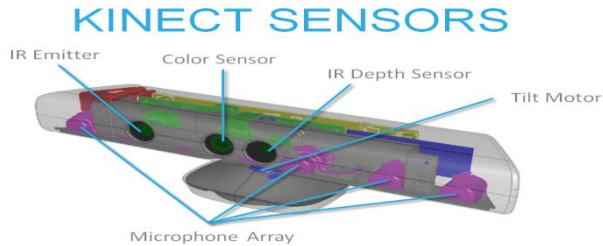


Figure 2. KINECT Component

Most of the human motion recognition is RGB camera, infrared recognition sensor, and infrared laser. RGB is simply problem of color and big size also getting the number and conduct two-dimensions. Infrared recognition sensor and an infrared laser, as shown in Figure 4, and the reflection of an object by using the infrared laser sensor infrared recognized as 3D depth information of the object read and processed. The Processing of 3D depth is a key feature of the KINECT.



Figure 3. Infrared Laser Motive Feature



Figure 4. Depth Information Processing

4. System Construction

Our system, KINUPS (KInect NUi Project System) program features action sequence is show in Figure 5. All by voice command is recognized. Mouse Mode is to perform the

functions of the mouse. For example, mouse click, mouse double click and wheel *etc.*, voice recognition function is voice commands execution.

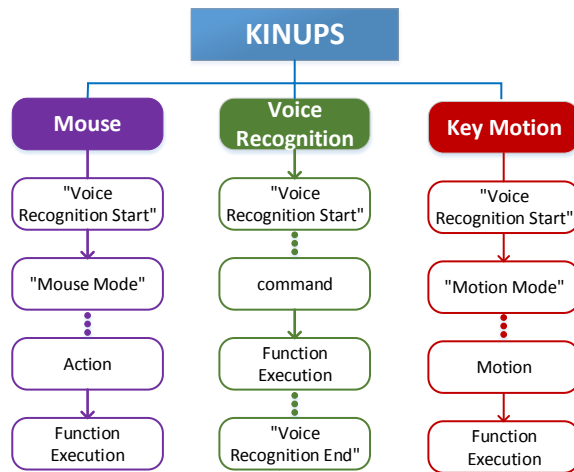


Figure 5. Program Features Action Sequence

Our system, KINUPS can map to mouse actions as follows. With your left hand, the left and right click and double click and wheel operation is performed. By your right hands, the cursor movement and drag operation is performing. Mouse handling method is as shown in Figure 6.

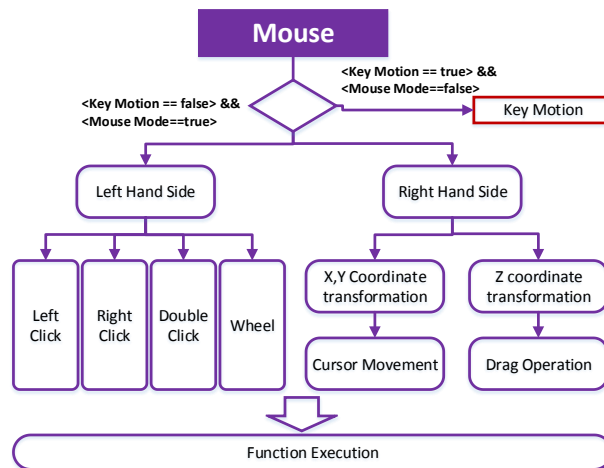


Figure 6. Mouse Handling Method

4.1 Motion Recognition Principle

KINECT recognizes objects and people by the joints, as shown in Figure 7. We can specify the track joints obtained, as shown in Figure 8 follow-ups, joint or virtual image at a specific location, or the skeleton of a man, and draw around objects. Values are specified x, y, and z coordinates of each joint every value of x, y, and z coordinates of the joints of the receiving value. By using that special feature, it is mapping the left hand joint, left hand joint coordinates and the location of the cursor on the screen so that by easily to replace the mouse , the human body can be controlled by computer.

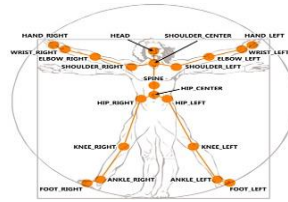


Figure 7. The Joints of Each Part of the Body

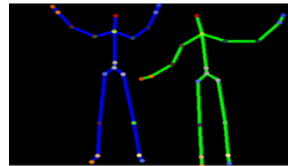


Figure 8. A Track Joints using Skeleton Figure Via Actual KINECT

4.2 Shortening Motion

By comparing the receiving each joints of coordinate values and the other joints values. Moreover, if there is an input a condition of other values, the special function at the specific motion can be inputted. If it is assumed that there is the right hand joint above the shoulder joint, it will be recognized that the user now is also holding up the right hand. However, when the right hand elbow joint is on above the shoulders, the lift up of a moment of arms state can be known.

Therefore, using each joint of other coordinate values, several actions can be recognized precisely. Also, while it has the utilization of the recognition behavior, the variety of functions can be embodied.

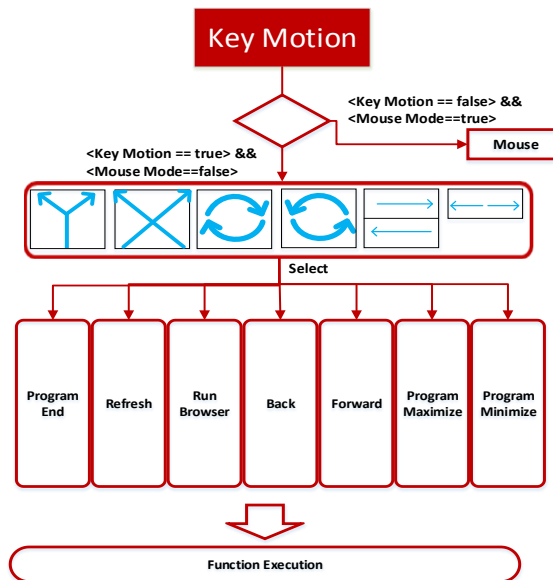


Figure 9. Shorten Motion's Parts Handling Method

Control, Return, Switch between Korean and English, letter mode, number mode, clear all, clear, End of program, close tab, new window, new folder, delete, cancel, maximize click, double click, drag, right click
Shut down, restart, sleeping mode, screen lock, internet, Chrome, IE
On-screen keyboard, Control Panel, My computer, Library, Task Manager, Music Player, Video Player
Select all, select mode, yes, all yes, no, all no
Undo, redo
Up, down
Play/pause, stop, next, Previous, Volume down, Volume up, Mute
Save, Replace, Unsaved, Revert, Re-run
Cut, Copy, Paste
Switch Mode, Previous Screen, Next Screen
Mouse Mode
Help (This window shows. If one more of talking, will be closed.)

Figure 10. Voice Instruction Guide Help

4.3 Speech Recognition

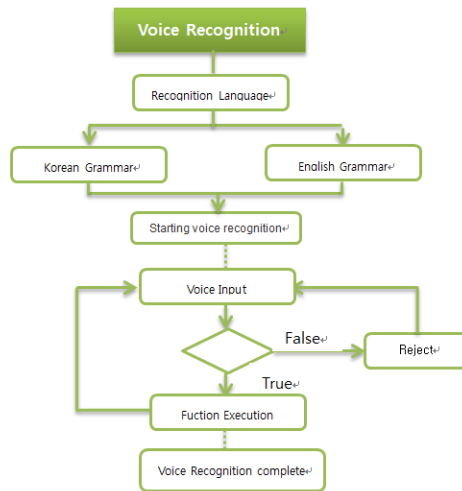


Figure 11. Speech Recognition Process

Because of the Speech recognition that analyze of the received input micro voice source that's why if the pre-configuration of voice source became matching, the understanding and the accomplishment principle is in progress. Therefore, in order to improve the recognition rate which is associated with speech recognition, a lot of effort is required. However, still far that speech recognition matches the voice source with data that we have received as above whether can divide to some extent between 'recognized.', 'not recognized'. Currently, in up forward if there is a usage of the elimination method of many data, noise voice source, and howling voice source, the better speech recognition results can be obtained. Overall, if there is a new method or a new algorithm, the recognition rate will be a lot better for sure.

English recognition and Korean recognition are awareness to work with a similar principle but KINECT does not formally support Korean voice. Because of that Korean Recognition, the Microsoft provides the Speech Platform of Korean to use for KINECT separately to operate set.

Table 1. Part of the Application using the KINECT

KINECT Function	Practical Section
Joint Tracking	Control the movement of the cursor by tracking the hand joints
	Swipe motion by calculating the left-to-right movement of the joint application
	Implement an abbreviated motion by calculating the relative joint coordinate values
Voice Capture	Voice input (voice source analysis and post-processing software, etc.) can be interlocked with implementation
Depth Information	Implementation can be used in a variety of motion by applying a three-dimensional rather than two-dimensional motion (including shortening motion)

5. Conclusion

In this paper, we propose a NUI by voice and motion recognition. Instead of keyboard and mouse, our method provides an easy controlling user interface. The tedious manual works are sure replaced by the method. While keyboard and mouse occupied the space so the flat surface should be installed that is the point. Also, as a large screen that cannot be seen as a whole from the short distance would be inconvenient. But at a relatively long rang distance if motion and voice can be operated, it will have more diverse and improve of the user interface of the existing user-friendly interface

In the future, we will apply to the more intelligent presentation method by only gesture and voice

Acknowledgement

This research was supported by a grant of Daejeon University in 2014.

References

- [1] S. Mitra and T. Acharya, "Gesture recognition: a survey," *IEEE T. Systems, Man and Cybernetics*, vol. 37, no. 3, (2007), pp. 311-324.
- [2] C. Pheatt and A. Wayman, "Using the Xbox Kinect sensor for gesture recognition," *Journal of Computing Sciences in Colleges*, vol. 28, Issue, 5, (2013), pp. 226-227.
- [3] J. S. Park, "Three-dimensional game programming is using DirectX 10," Green Press, (2009).
- [4] H. C. Yoon and J. S. Park, "Gesture-based Avatar Animation Using Kinect Sensor," *The 3rd ICCT 2013 Proceeding*, (2013), pp. 356-357.
- [5] C. Endres and S. Dimitrov, "Using a Theremin for Micro-Gesture Recognition an Automative Environment", *Proc. of the 2nd Int. conf. on Automotive User Interface*, Pennsylvania, USA, (2010) November 11-12.
- [6] A. Rieneer, M. Rossibory and A. Forscha, "Natrak DVI based on intuitive hand gesture," *Springer LNCS*, (2011) September, pp. 62-66.