

Context-Aware Kinect Sensor based PC Control Interface for Handicapped Users

Seung-Hyun Oh

*Dept. of Computer Engineering, Gyeongju Campus, Dongdae-ro 123,
Gyeongju-si, KOREA, 780-714
shoh@dongguk.ac.kr*

Abstract

It is widely considered that it is very difficult to use a personal computer for people with disabilities. The mouse and keyboard based interfaces assumes that all people have hands and arms. System interface that will allow handicapped users without the hand and arm having a user experience similar to the general users is required. In this paper, by using the Kinect sensor, which is supplied with the MS X-box, with speech and motion recognition, author will offer a new computer interface for handicapped users with no hands and arms to be able to control personal computers at a level similar to the general user. It is shown that implemented application with the new interface system enables users to input user data, information and easily control the execution of the program.

Keyword: *Context-aware, X-Box, Kinect sensor, pc control, motion recognition, speech recognition, handicapped user, handicap interface system*

1. Introduction

Recently, many interfaces for interaction with the human-computer has been developed. Initially, graphics based interface for batch processing base were mostly used, and now, the human computer interaction trend is switching to the natural input method that uses motion recognition and speech recognition. The purpose of these interfaces is to allow the user to communicate input and output directly to a computer. Kinect [1] it was made by MS's UI research as a low cost performance device, the best next-generation interface for the interaction of current is on the rise. Kinect is sensor device that comes by default with the X-Box gaming console has two depth perception camera sensors, one video camera and a multi-array microphone. Microsoft supports the development of the Kinect SDK continuously for various Kinect applications.

With the help of the MS's Kinect game console and game accessories that comes with the X-Box device the user is able to control speech and motion controlled games. MS Windows operating systems can support Kinect sensor device independently without X-Box and provide Kinect camera sensor and a microphone sensor implementation through a variety of functions to support and take advantage of these features. Many useful Kinect applications have been already developed. For example, Kinect sensor using the camera sensor monitoring and corrective training already helps in the medical profession to provide training for people after the accident. In addition, the camera sensor is utilized in many areas where the depth of a very low-cost three-dimensional data needed to be recognized.

In this paper, using the Kinect sensor device, a new personal computer interface is proposed for handicapped users with physical disabilities who will be able to experience the similar level of the user interface as non-handicapped people. Handicapped users via the new

interface, without the help of the mouse and the keyboard capable to execute various applications, input user data, and also send e-mail, enjoy movies and music. This paper is organized as follows. Section 2 introduces a related research to Kinect, and Section 3 is a brief introduction to the Kinect sensor itself. Section 4 a new interface for the handicapped user using the Kinect sensor system is introduced. Finally in Section 5 experimental results using the implemented system are presented. Conclusions are made in Section 6.

2. Related Work

Currently a variety of user interfaces using Kinect sensors have been developed. Extraction and recognition of the skeleton image was achieved by the Kinect's depth sensor to create the motion-based and speech recognition interface [2]. This interface is used to build a home based shopping experience using the Kinect sensor to allow users to easily shop online in the real-time, select product with hand motions. Another new working mechanism using the Kinect depth camera for efficient gesture recognition using the video information has been proposed [3]. Mobile devices like the smartphones by incorporating hand gestures were turned into real-time control devices for interactive presentation. In addition, preliminary results of the construction of a surveillance system were presented [4]. Two subsystems were developed. The reaction system was effective to provide a safe navigation for the robot and the classification system for indoor situations. Kinect's depth sensor and an infrared camera measures the distance of the object, and identifies the location of the object is to provide security.

In addition, Kinect any other device be used in place of areas where there are many. Kinect goal of any devices with the human body alone operation aimed at, because multiple devices can be replaced. In [5] the real-time computer interaction system using Hand Data Glove has been introduced. It is found that the static keyboard and mouse are having many limitations with them, while in the case with data glove can be used for the same purpose without any limitations. Although these gloves are costly, this experiment proves that such devices are a good technological for interacting and controlling the devices, software or hardware. The air writing and sketching with 3D game are the software applications that involved data glove as input device. IPTV research [6] in the Kinect field has been written. Speech recognition was used to switch different channel or operate your motion by manipulating parts, etc. Kinect research in the healthcare sector was well studied in [7], where a simple and cost effective robot was developed, which is multi-purpose tele-present in nature. Mainly it is designed for hospital environments but can be extended to other applicable areas such as security, industrial inspection, etc. The robot can monitor the patients and Kinect sensors successfully monitor the patient's condition. All these related researches are proving the fact that Kinect can improve or replace existing technology and equipment in many areas.

In this paper, we propose many kinds of Kinect sensors applications that without using the hands and arms and other part of the body allow people with disabilities to use a computer system via the new a new type of motion-based interfaces and speech recognition interfaces. It also provides a application program categories for controlling the PC, a handicapped user to be able to control the PC in the same manner as general users.

3. Introduction to Kinect Sensor

Kinect is a UI device made by MS. First, it was introduced as replacement of the usual game controller by making user's body to serve as a controller and operate for different purposes. Now after the introduction of the Kinect for Windows, with the use of available Kinect SDK anyone is able to develop new interfaces and applications of the program.

Kinect has two depth perception camera sensors, one video camera and a multi-array microphone. Camera sensor captures video, while the depth perception sensors using the infrared thermo graphic camera recognizes the depth of object in 3D dimension. As shown in Figure 1, Kinect depth sensor using the SDK, provides a skeleton coordinates.

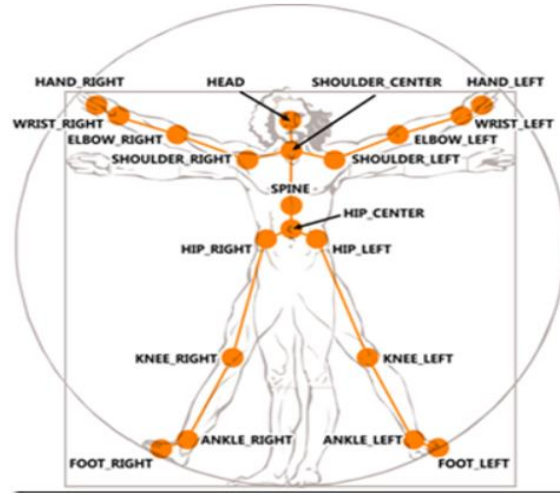


Figure 1. Skeleton Joint Elements

Only 6 people can be detected by Kinect sensors, but at any given time only two people's skeleton joints position movements can tracked and recognized. After the release of the Kinect's SDK 1.7 for Windows OS, the depth sensor started to sense objects using the 3D imaging, and this technology was called Kinect Fusion. Kinect's multi-array microphone has a support of four microphones. Through the usage of four microphones located in the device, Kinect has the ability to detect the position of the source object and movements that occur. Kinect with the use of speech recognition in the available SDK supports various languages. In order to facilitate handicapped users using personal computers and to experience similar level of human-computer interaction, it is required to explicitly define the various commands to control Kinect's speech recognition feature.

4. Context-aware Kinect Sensor based PC Control Interface

4.1. Decision of Mouse Interface Level

Mouse interface replaces the input interface of PC such as keyboard and mouse. The handicap users can control the mouse pointer by using their arms, head, and voice recognition in order to play music/video, surfing the internet, edit the document, etc. This paper proposes an interface which can automatically decide the mouse level control depend on the degree of disability by tracking all parts of the body (such as head, shoulder, elbow, wrist, etc.). In the initial phase, the software checks whether if the users have arms or not. If the users have their arms, it replaces the PC's mouse by tracking the user's elbow. In contrast, it replaces the PC's mouse by tracking the user's head. Figure 2 shows the decision-making process mouse interface. After check the skeleton and decide the interface level, the software supports voice recognition for the "mouse click" event. The software invokes the mouse click event when the user say "select".

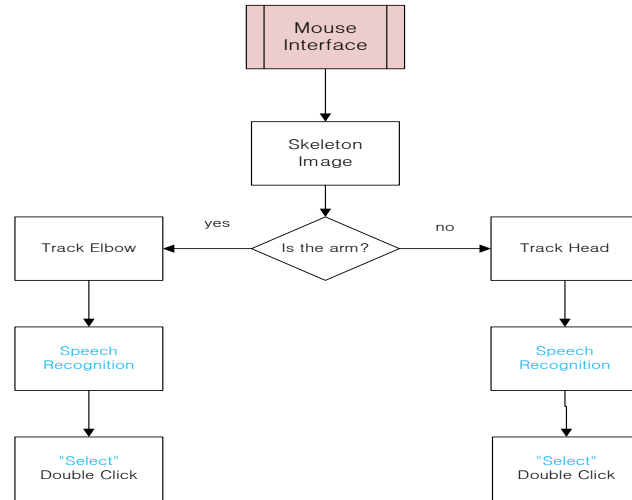


Figure 2. Decision-making Process Mouse Interface

In our Kinect-based interface system, the mouse coordinates (X, Y) will be recognized by calculating the skeleton coordinates (point.X, point.Y). The coordinates of the mouse equals to the skeleton coordinates multiple by the layout, then divide for the depth of the frame, equaltion (1).

$$\begin{aligned}
 X &= (int)((point.X * LayoutRootWidth / sensor.DepthStream.FrameWidth) \\
 Y &= (int)((point.Y * LayoutRoot.Height / sensor.DepthStream.FrameHeight).
 \end{aligned}
 \tag{1}$$

In Win32 API, SetCursorPos(x, y) function is used to set the moving of mouse cursor, and Mouse_event () function is used for the double-click event.

4.2. Context-Aware Korean Speech Recognition Interface

In this paper, we use speech recognition interface to replace the keyboard input with the help of Kinect microphone. First, the interface recognizes 2350 Korean syllables. Then, it combines those syllables and forms Korean words. Recently, Kinect does not officially support for Korean language recognition, we must use a special method to recognize Korean words. Because the recognition of Korean words is restricted, we propose to use the combination of syllables to form a word. Besides, we also propose to combine among words to form a sentence. In order to form a sentence, we use timers to decide where the end of the sentence is and insert the punctuation.

Figure 3 shows our speech recognition interface process in order to recognize Korean voice. Our speech recognition interface consists of two modes: data mode and command mode. In data mode, speech recognition is used to input the text in word processor or the content of email. In command mode, speech recognition is used to invoke the exist function to run an application, or execute a command. For example, when user says “Notepad”, it will open the notepad application. When user says “Close”, it will end the program. We also define a sub-command mode for each separate application when it is opened.

4.2.1 Context Recognition and Mode Conversion: There are two ways to distinguish, and change between data mode and command mode. Firstly, speech recognition uses the support of Win32 API. In Win32, each application has a process identify (p_id). When an application is called, the speech recognition seeks for the p_id then decides the next current mode. For example, if the speech recognition found that the p_id is Notepad or MS Word, it will set the

state and enter to the data mode. In contrast, if the p_id is Video Player, it will enter the command mode. Another one, we define our own method to switch between two modes. From command mode to data mode, for example, when internet explorer is running, we say “address” to switch from command mode to data mode. And the pointer will be placed in the address bar for inputting website address. Then, we can say the name of the website. From data mode to command mode, for example, in MS Word we call a statement by saying statement name twice: copy-copy to do the copy statement, close-close to exit the MS Word. Besides, we also say “command mode” or “data mode” in order to switch between them.

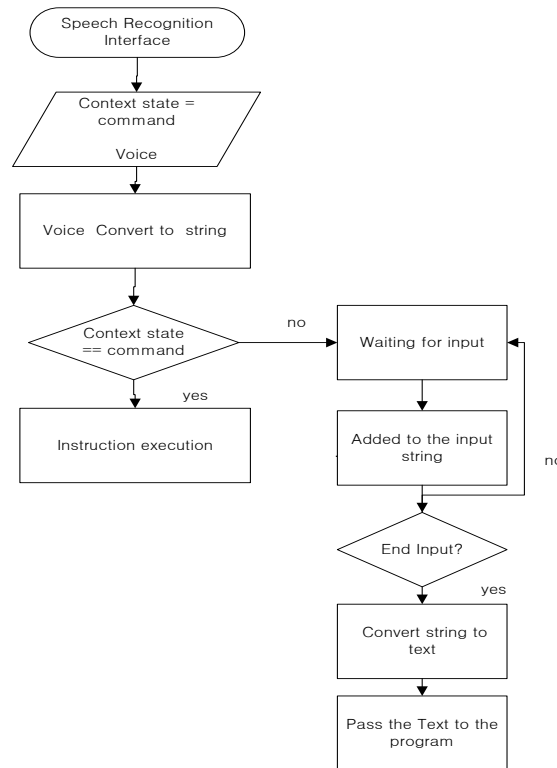


Figure 3. Speech Recognition Command Line Interface

4.2.2. Data Input mode: Data input mode is explained follows the right side of Figure 3. All syllables, words will be recognized during the waiting time. After the waiting time expires, all recognized data is added in StringType. When the input phase finished, the string is converted to text, and then passes to the target program. As we discuss above, we need specific method to combine among Korean syllables to form a word, and combine among words to form a sentence. Because of specific characteristic of Korean language, we use two time intervals: shorter time interval (word timer) for word recognition, and another longer time interval (sentence timer) for sentence recognition.

Word timer and sentence timer always start at the same time depend on the start of speech input. Within the specified time, they restart continuously without interrupted by any else input event. If there is recognition of a new word within a word timer, it will be added to the old word and form a new word with two syllables, not a sentence. If there is recognition of new word when the word timer expires and sentence timer is still not expire, it will be added as a new word and distinguish with the old word by a space. The sentence is formed when the

sentence timer expires or when user say “Enter”. The punctuation is also added into the end of the sentence.

For the better recognition system, we intend to optimize the continuous idle time so that we can recognize the Korean words in natural speaking.

4.2.3. Command Mode: In command mode, recognized commands should be predefined. For each active application, specific commands are set based on the base hierarchical of application shortcuts. Figure 4 shows the hierarchical of video/music application for controlling by speech recognition command.

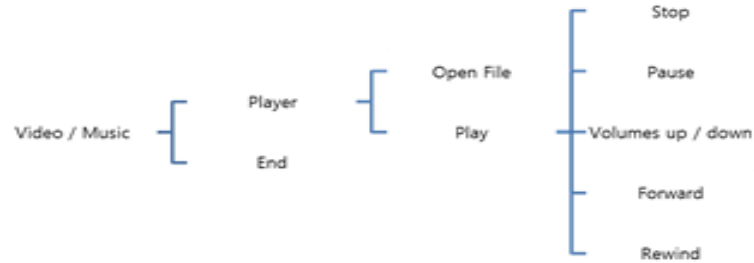


Figure 4. Video / music Program Configuration Command Speech Recognition

4.3. Kinect-based System Interface Configuration

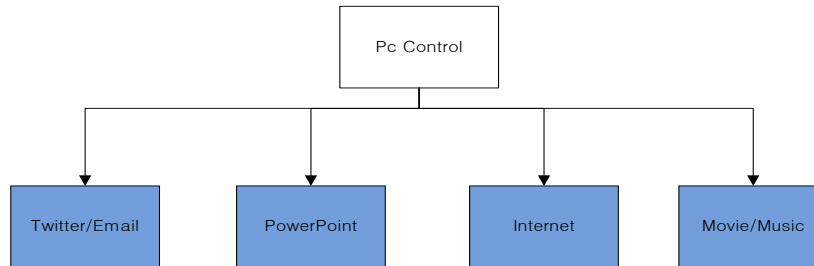


Figure 5. The Entire System Category

Figure 5 shows the category of our provided service for handicap users. Though the system, handicap user can launch the email, open power point, edit document, surfer the internet, etc. The system is used as the similar level for all users. In this section, we explain more details in some services of our system.

First is the video/music application. When user says “players”, our system will open video/music application. In video/music application, say “Open” to open the open file screen. User uses arms or head to control the mouse cursor. User selects the file to open by pointing the mouse cursor on the filename and say “Select”. Also, user can pause, stop, volumes up/down, fast forward, rewind, and exit the application using the same actions.

Secondly, user says “Power Point” to open power point application. User says “File Open” to open file open screen, and choose a file by pointing mouse on the file name and say “Select”. User says “Slide show” to switch to the slide show screen. In the slide show screen, user says “Next Page”, “Previous Page”, and “End” to move to next slide, previous slide or exit power point application, respectively. Also, user can copy, paste, or delete a slide.

Another one is Internet browser. User says “Google” or “Naver” to open the internet explorer which the address of Google/ Naver website. User says “address”, the pointer will

be pointed on the address bar so that user can say the website name they want to browse, for example twitter.com or yahoomail.com. When user is in the login window, say “Username”, “Password” to fill the username and password, respectively. Then, user says “Enter” to login. Also, user says “Exit” to close the internet explorer.

5. Experiments and Analysis

In order to implement the new user interface the latest the Kinect SDK version 1.7 was used with the Visual Studio 2010 C # programming. The developed interface was tested in Windows 7 32bit operating system. The entire system provides support of the program for each category like the GOM media player and notepad. Korean speech language pack was used to ensure that the two programs run in the background. During the experiments such modes as the mouse interface mode, data input mode and command mode were tested to control the music and video viewing.

5.1. Mouse Interface

Figure 6 shows the configuration screen of the mouse interface to help to run the program. The configuration screen captures the user's head and arms to recognize the video shown and also consists of a TextBlock and TextBlock2 text labels at the bottom left and bottom right of the screen for speech recognition commands. There are two types of timers in the test program, words timer and sentences timer. Using the head and elbows a handicapped user can control the mouse cursor movements. For precise gesture recognition some adjustment and program training is required. And after successful mouse cursor movements controlling using the head, a user is able to manipulate a cursor over the icon and select it with the special speech recognition command.

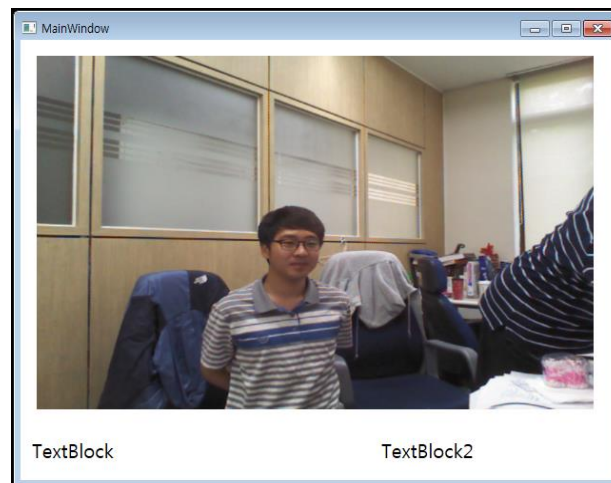


Figure 6. Execution Display Interface Recognizes Mouse Head

5.2. Data Input Mode

Figure 7 shows the data the input mode by using the NotePad application the text typing. The left bottom of the window is the word timer and the right bottom of the window is sentence start timer. In the data input mode by using a phrase "너는 누구" (“Who are you” in English) the sentence will be written in the NotePad window. Once a text is pronounced, the left bottom of the window will show the text while passing the same data to the NotePad

application. By running many experiments and test runs an author has identified an optimal timer value. During experiments the word Timer was set to 3 seconds and the sentence Timer was set to 5 seconds. Without NotePad running, the data input mode will not work due to context aware nature of the proposed entire system.

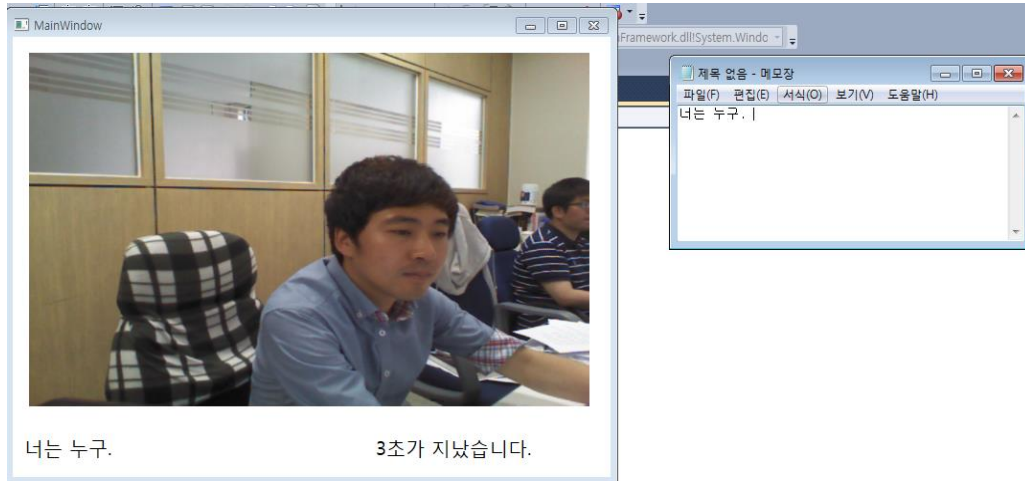


Figure 7. The Data Input Mode Screen

5.3. Watching Movies via the Command Mode

Figure 8 shows the command mode screen of the system to watch the movie. In Figure 8, verbal command 'player' executes a stylized Gom media player and runs a floating select file window after saying the 'Open' verbal command. Selection of the file is accomplished using the gesture recognition and moving the mouse cursor. And after that using the special commands such as 'play', 'sound up', 'sound down', 'stop' etc. it is possible to control everything in an easy way without using a physical mouse or keyboard. All these commands have been defined in the developed program to make it possible. All these video playing related commands such as 'sound up', 'sound down' and 'stop' will not work in any other simultaneously running application for two main reasons. First, it is because the system is context aware. And second, because the upper and lower commands are differentiated in the video and music commands hierarchy tree.

6. Conclusion

This study of the MS Kinect sensor provides a foundation for the convenient use of a PC system by handicapped users. This Paper provided two interfaces, the motion sensor with the mouse interface and the speech recognition interface that recognizes and controls the situation. Motion sensor-based mouse interface allows users without hands or arms to use their head and elbows to control the mouse cursor. Controlling the different applications using the speech recognition is very convenient due to context aware nature of the system. The proposed Kinect based interface system enables handicapped users to use the computer system at the same level as compared to normal people facilitating to perform such tasks as video and music playback, internet surfing, document editing, sending an e-mail. In future works the deeper potential of the Kinect sensors should be explored and other problems of the human computer interaction by the handicapped users must be addressed.

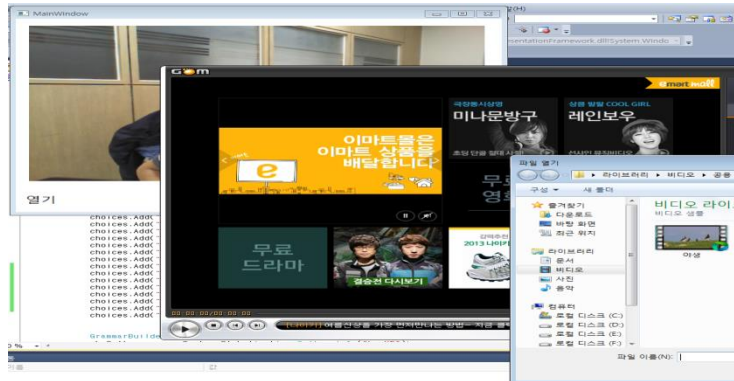


Figure 8. Watching Movies via the Command Mode

References

- [1] Kinect for Window, <http://www.microsoft.com/en-us/kinectforwindows/>
- [2] T. Arici, "Introduction to programming with Kinect: Understanding hand / arm / head motion and spoken commands," Signal Processing and Communications Applications Conference (SIU) 2012, (2012) April, pp. 18-20.
- [3] V. Tam and L.-S. Li, "Integrating the Kinect camera, gesture recognition and mobile devices for interactive discussion," Teaching, Assessment and Learning for Engineering (TALE), 2012 IEEE International Conference, pp. H4C-11, H4C-13, 20-23 August (2012)
- [4] D. S. O. Correa, D. F. Sciotti, M. G. Prado, D. O. Sales, D. F. Wolf and F. S. Osorio, "Mobile Robots Navigation in Indoor Environments Using Kinect Sensor," Critical Embedded Systems (CBSEC), 2012 Second Brazilian Conference on (2012) May 20-25, pp. 36, 41.
- [5] P. Kumar, J. Verma and S. Prasad, "Hand Data Glove: A Wearable Real-Time Device for Human-Computer Interaction", IJAST, vol. 43, (2012) May & June.
- [6] M.-H. Lee, "The Service Generation Apparatus for IPTV Interactive Digital Channel", IJAST, vol. 10, (2009) September, pp. 37-51.
- [7] Z.-u. Haque, R. Siddiqui, R. Javaid and I. M. Hussain, "Tele-presence Based Patient Monitoring Robot", IJAST, vol. 42, (2012) May, pp. 45-60.

Author



Seung-Hyun Oh, He received the B.S. degree in computer science from Dongguk University, Seoul, Korea, in 1988, and the M.S. and Ph.D. degrees in computer engineering from Dongguk University, Seoul, Korea, in 1998 and 2001, respectively.

He has been a Professor with the Department of Computer Engineering, Dongguk University Gyeongju Campus since 2002. His current research interests include wireless communications and sensor networking system.

