# Design and Usability Analysis of Gesture-Based Control for Common Desktop Tasks

Farzin Farhadi-Niaki[1], S. Ali Etemad[2], and Ali Arya[3]

[1] School of Electrical Engineering and Computer Science, University of Ottawa, Ottawa, Canada
ffarh101@uottawa.ca
[2] Department of Systems and Computer Engineering, Carleton University, Ottawa, Canada
ali_etemad@carleton.ca
[3] School of Information Technology, Carleton University, Ottawa, Canada
arya@carleton.ca

**Abstract.** We have designed and implemented a vision-based system capable of interacting with user's natural arm and finger gestures. Using depth-based vision has reduced the effect of ambient disturbances such as noise and lighting condition. Various arm and finger gestures are designed and a system capable of detection and classification of gestures is developed and implemented. Finally the gesture recognition routine is linked to a simplified desktop for usability and human factor studies. Several factors such as precision, efficiency, ease-of-use, pleasure, fatigue, naturalness, and overall satisfaction are investigated in detail. Through different simple and complex tasks, it is concluded that finger-based inputs are superior to arm-based ones in the long run. Furthermore, it is shown that arm gestures cause more fatigue and appear less natural than finger gestures. However, factors such as time, overall satisfaction, and easiness were not affected by selecting one over the other.

**Keywords:** Usability study, human factors, arm/finger gestures, WIMP.

## 1 Introduction

The new wave of input systems in video game consoles (such as Nintendo Wii, Xbox Kinect, and PlayStation Move) is leading the new generation of Human-Computer Interaction (HCI) systems to focus on creating interfaces that are more intuitive and user-friendly. While the gaming industry is currently leading the way using the aforementioned consoles, it will not be long before users will be controlling advanced and simple Virtual Reality (VR) and computer systems using body gestures that feel intuitive. Having an HCI system designed intuitively, thus, can provide higher user satisfaction and better performance. Development of reliable gesture recognition algorithms, choosing appropriate gestures, and studying the usability of these gesture-based methods are among topics that require the attention of researchers.

In this paper, we describe the design and implementation of a system capable of interacting with natural gesture inputs through computer vision methods. The vision

M. Kurosu (Ed.): Human-Computer Interaction, Part IV, HCII 2013, LNCS 8007, pp. 215–224, 2013.

sub-system is based on the Kinect depth-based camera. Recognition algorithms have been studied and implemented to form a robust system. The operating environment is a simulated computer desktop containing several objects such as windows and icons that simplifies the user interface and allows better control over test sessions. Gestures, both using full arm and only using fingers, have been carefully designed and assigned to replace mouse inputs for common desktop tasks. Successive to implementation, the system is tested with multiple users which provide the feedback needed to analyze the usability of such systems with respect to factors such as precision, efficiency, ease-of-use, fun-to-use, fatigue, naturalness, and overall satisfaction.

The major contributions of this study are: a) choice of natural gestures, b) usability study for gesture-based input, and c) system design (UI and gesture recognition) and relatively novel use of existing API's to implement gesture recognition method. This method conserves the developing time (no need for making samples and perform training and testing sessions) and running time for gesture recognition and user interaction compared to learning-based traditional method.

## 2 Related Work

Employing natural arm and finger gestures for VR applications can be studied in two different aspects: technical design and implementation, and usability. Through the following we review some relevant vision-based gesture systems in the two mentioned fields.

Detecting hand gestures has been subject to extensive research. Hidden Markov models (HMM) are one of the popular classifiers for this purpose. Marcel et al. [1] employed input-output HMMs for tracking variations in the skin color of the human body. Similarly, Chen et al. [2] employed HMMs for detecting hand postures. The AdaBoost algorithm was revised and used by Liu et al. [3] to automatically recognize users' hand from the video stream. Yu et al. [4], proposed a hand gesture feature extraction method using multi-layer perceptrons. Raheja et al. [5] used principal component analysis (PCA) for hand pattern matching. Other techniques such as cascade classifiers often used for face tracking applications have also been utilized to recognize hands and various parts of the human body [6].

The second parameter in need for in depth study, as mentioned earlier, is the usability aspect of natural arm and finger gestures for practical and VR system inputs. In this regard, Cabral et al. [7] discussed numerous issues associated with the use of gestures as input modes. Their studies showed that both time and fatigue increases when gestures are used for simple pointing tasks. Villaroman et al. [8] show that Kinect-assisted instruction can be utilized to accomplish certain learning results in HCI courses. Moreover, through their study, it is confirmed that OpenNI, a system that is also employed in this research, is a reliable and effective tool to be utilized along with Kinect. It was shown that when the two are used together, students are provided with a hands-on experience on gesture based natural user interaction systems and technologies. Through another study on using Kinect for VR interaction, Kang et al. [9] used distance information and joints' location information and achieved higher recognition rates. They also showed that their system was 27% faster than the mouse device. Bragdon et al. [10] developed a system that combines touch and air gesture hybrid interactions for small developer group meetings. Their proposed system proves applicable with different devices such as multi-touch screens, mobile touch devices, and Kinect. The use and usability of hand gestures for tasks such as making telephone calls, operating the television, and executing mathematical calculations has been studies by Bhuiyan and Picking [11]. Their study suggests that such technologies can benefit the elderly and the disabled users by causing more independence while some challenges still remain to overcome. Applications of gestures as inputs for medical systems have also been explored. In a study conducted by Ebert et al. [12] rebuilding images from a CT data was tested and it was shown that image recreation time using gestures was longer than using mouse/keyboard. The system, however, maintained certain advantages such as reducing the potential for infection, for both patients and staff.

The provided review on the relevant literature shows that the use of natural arm/hand/finger gestures for interaction with different systems has been growing. This trend is especially increasing as new generation game consoles such Kinect are becoming more available and convenient to purchase and develop by researchers. Free developing and computer vision tools such as OpenNI and OpenCV also aid and accelerate the process.

## 3 Methodology

In this section we describe the design and methodology used for developing our gesture-based simulated desktop interaction system. The three different steps of the system design are described through the following sections: user interface design, natural gesture selection, and gesture recognition module.

### 3.1 User Interface Design

The simulated desktop for the system was developed using the Allegro library. Allegro is an open source library used for game and multimedia programming. Its cross-platform nature makes it easy to integrate with other modules of the system. The Post-WIMP (windows-icons-menus-pointers) [13] design is adapted for the desktop while neutral colors are utilized to reduce user error or bias. Novice users can learn WIMP user interfaces easily, as they are very good at abstracting workplaces due to their analogous paradigm to documents like paper sheets or folders. Having a rectangular region on a 2D flat screen makes them preferable to system developers while their generality also makes them a good fit in multitasking environments.

### 3.2 Natural Gesture Selection

We first studied several possible natural gestures suitable for a Post-WIMP user interface [14] [15], and then defined the best matches of the predefined gestures to our prototype. One determining criterion for the selected gestures is the intuitive naturalness of the actions and effects. Tables 1 and 2 present the final selected arm and finger gestures for the proposed system and corresponding descriptions. Table 3 presents the analogies for the three input mechanisms with respect to one another and corresponding actions.

**Table 1.** Final design for arm gesture set

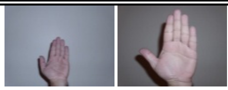
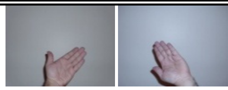
| Description | Hand pushing | Hand moving | Hand circling |
|---|---|---|---|
| Hand gestures |  |  |  |

**Table 2.** Final design for finger gesture set

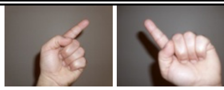
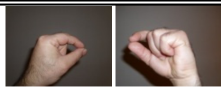
| Description | Finger tapping | Finger moving | Pinching |
|---|---|---|---|
| Finger gestures |  |  |  |

**Table 3.** Arm and finger gestures, and mouse analogies

| Finger gestures | Arm gestures | Mouse | Actions |
|---|---|---|---|
| Tapping | Pushing | Left click | Selecting/Opening/Closing/Dropping |
| Moving | Moving | Moving | Pointer movement |
| Pinching | Circling | Drag | Grabbing/Resizing |

### 3.3 Gesture Recognition Module

For arm gesture detection and recognition, some predefined features (circle and push) of applied APIs (OpenNI and NITE) are utilized, and a very efficient and accurate system is developed. Such pre-defined functionalities, however, do not exist for finger gestures (pinch and tap) detection and classification in OpenNI and NITE. Therefore we have implemented the algorithm in OpenCV (Fig. 1) to detect the fingers (Fig. 2).

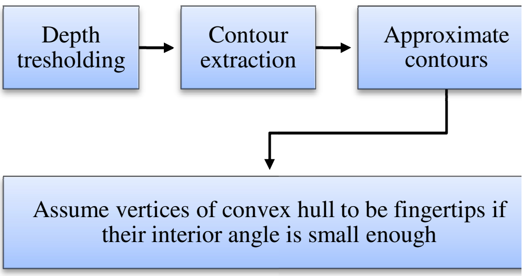



**Fig. 1.** Fingertip detection algorithm

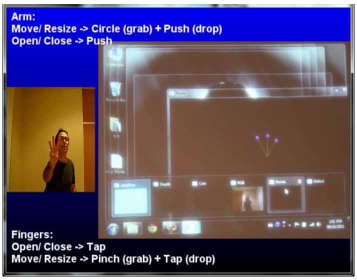



**Fig. 2.** Finger detection outputs

The "tapping" gesture has been defined based on the depth change of fingertip ($z_p$) comparing to the depth of hand/palm's centre ($z_h$), and a proper threshold ($D$) as shown below:

$$\text{Tapping happens when: } |\boldsymbol{z_h} - \boldsymbol{z_p}| > D \tag{1}$$

The "pinching" gesture has been defined based on the distance between the point finger's tip ($x_p$, $y_p$, $z_p$) and thumb's tip ($x_t$, $y_t$, $z_t$) based on the following:

$$\text{Pinching happens when: } \begin{cases} |\boldsymbol{x_p} - \boldsymbol{x_t}| < \varepsilon \\ |\boldsymbol{y_p} - \boldsymbol{y_t}| < \varepsilon \\ |\boldsymbol{z_p} - \boldsymbol{z_t}| < \varepsilon \end{cases} \tag{2}$$

We have designed an Algorithm (similar to our previous work for arm gestures in [16]) to control our user interface objects utilizing the recognized finger/arm gestures (circle and push are replaced by pinch and tap). This algorithm also recognizes index finger and thumb, with a possibility of orderly detecting all five fingers.

## 4 User Experiments

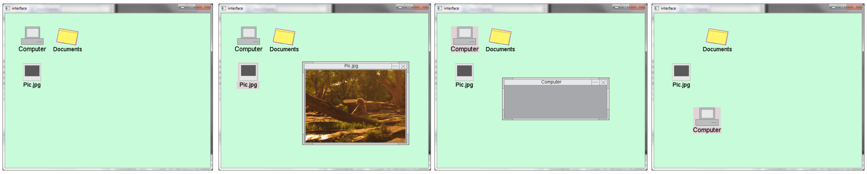
In this study two different interactive variables namely arm and finger gestures are employed and compared. A set of usability parameters are analyzed for each input method when performing combined tasks with two difficulty levels (simple and complex), on big-screen display. According to our previous study (gestures vs. mouse) [16], using gestures on big-screen was proved to be superior to using gestures on desktop-screen. Therefore, we have chosen solely big-screen display for this inter-gestures study. The usability parameters consist of human factors such as ease of use, fatigue, naturalness, pleasantness, and overall satisfaction, as well as performance factors such as efficiency and effectiveness.

### 4.1 Training Session

In this session participants are asked to practice primitive tasks for a period of 30 minutes in order to get acquainted with the system prior to participating in the main test. Furthermore, this phase plays a major role in validity of particular satisfaction criteria such as fatigue and naturalness.

### 4.2 Test Session

During the test session, the different variety of tasks using each input method on big-screen is examined. As described earlier, the two variables, input method and task difficulty, combine to generate four states. Each state is examined independently for each participant. Figure 3 presents snapshots of some activities performed during the test sessions.




(a) (b) (c) (d)

**Fig. 3.** User interface: (a) initial configuration, (b) object opened, (c) object resized, and (d) object moved

### 4.3 Questionnaire and Observations

Following the test sessions, participants are asked to provide their feedback through a questionnaire from which user satisfaction criteria are extracted. Ratings are scaled from 1 to 5 (1 for absolutely unsatisfied and 5 for extremely satisfied). Extra written feedback is also acquired for further information regarding both participants and system. During the different states of the test sessions, time and error are observed and recorded for each user.

## 5 Results and Discussions

This study is conducted using 10 participants (5 males and 5 females) and in the age range of 26 to 36 (average of 30 years old).

### 5.1 Hypotheses and Analyses

For the different factors being studied, two-way repeated analysis of variances (ANOVA) is carried out for two independent variables:

1. Difficulty (simple task vs. complex task)
2. Input method (finger gestures vs. arm gestures)

All experiments were carried out on the big-screen and at $p < 0.05$ significance level and for 10 participants.

Notation: In our analyses, we calculate the mean and standard deviation for different variables in the forms of $M_{variable}$ (e.g. $M_{simple}$ is the mean for simple task) and $SD_{variable}$ (e.g. $SD_{finger}$ is the standard deviation for finger gestures). Moreover, F(df,MS) is the test statistic (F-ratio) in which df and MS are the degree of freedom and mean square respectively for the variables (within variables when more than one, and within subjects). The F-ratio is calculated using $MS_{variable(s)}/MS_{error(s)}$ and P is the probability value.

Table 4 presents our statistical analyses, hypotheses, and results for different factors.

**Table 4.** ANOVA analyses, hypotheses, and results for different factors

| | Hypotheses | Variable 1 | Variable 2 | Variables 1 & 2 | Results |
|---|---|---|---|---|---|
| **Time** | Using finger gestures is faster than using arm gestures as inputs. | $F(1,617.01) = 202.5$<br>**$P = 0.00$**<br>($M_{simple} = 13.35$ $SD_{simple} = 2.30$)<br>vs.<br>($M_{complex} = 21.21$ $SD_{complex} = 3.43$) | $F(1,2.86) = 0.31$<br>**$P = 0.59$** | $F(1,0.13) = 0.05$<br>**$P = 0.82$** | Rejected |
| **Easiness** | Using finger gestures is easier than using arm gestures as inputs. | $F(1,0) = 0$<br>**$P = 1$** | $F(1,0) = 0$<br>**$P = 1$** | $F(1,0.10) = 2.25$<br>**$P = 0.16$** | Rejected |
| **Fatigue** | Using finger gestures causes less fatigue than using arm gestures as inputs. | $F(1,0) = 0$<br>**$P = 1$** | $F(1,4.90) = 12.25$<br>**$P = 0.00$**<br>($M_{finger} = 4.50$ $SD_{finger} = 0.60$)<br>vs.<br>($M_{arm} = 3.80$ $SD_{arm} = 0.69$) | $F(1,0) = 0$<br>**$P = 1$** | Confirmed |
| **Naturalness** | Using finger gestures is more natural than using arm gestures as inputs. | $F(1,0.02) = 1.00$<br>**$P = 0.34$** | $F(1,11.02) = 441.0$<br>**$P = 0.00$**<br>($M_{finger} = 5$ $SD_{finger} = 0$)<br>vs.<br>($M_{arm} = 3.95$ $SD_{arm} = 0.22$) | $F(1,0.02) = 1.00$<br>**$P = 0.34$** | Confirmed |
| **Pleasantness** | Using finger gestures is more pleasant than using arm gestures as inputs. | $F(1,0.40) = 6.00$<br>**$P = 0.03$**<br>($M_{simple} = 4.80$ $SD_{simple} = 0.41$)<br>vs.<br>($M_{complex} = 4.60$ $SD_{complex} = 0.50$) | $F(1,0) = 0$<br>**$P = 1$** | $F(1,0.40) = 6.00$<br>**$P = 0.03$**<br>($M_{finger-simple} = 4.90$ $SD_{finger-simple} = 0.31$)<br>vs.<br>($M_{finger-complex} = 4.50$ $SD_{finger-complex} = 0.52$) | Rejected |
| **Overall Satisfaction** | Overall, using finger gestures as inputs is a more popular experience compared to arm gestures. | $F(1,0.40) = 3.27$<br>**$P = 0.10$** | $F(1,0.40) = 3.27$<br>**$P = 0.10$** | $F(1,0.10) = 2.25$<br>**$P = 0.16$** | Rejected |

Figure 4 shows the times taken to complete simple and complex tasks using arm and finger gestures. Analysis of users' feedback regarding the primitive tasks is shown in Fig. 5, where pinching to resize and pushing to close a window were the most difficult gestures, due to the relatively small control access area of the objects.

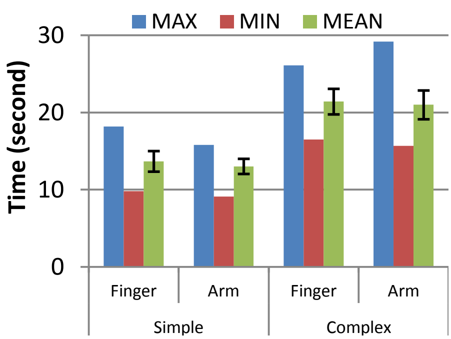



**Fig. 4.** Temporal statistical factors

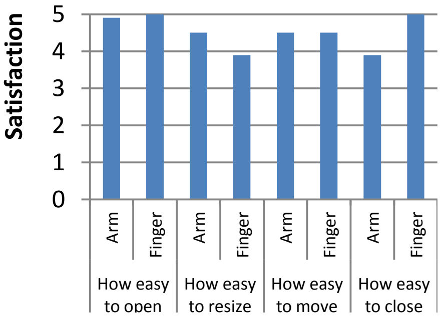



**Fig. 5.** Satisfaction on primitive tasks

As the following figures show, the Tapping was the smoothest gesture with the least errors (average number of trials) in both simple and complex tasks, while the Circling had the highest error in the simple task. However, the most errors happened with the Pushing during the complex task on the action of closing a window.

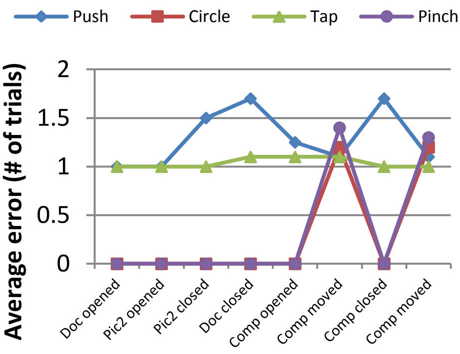



**Fig. 6.** Gestures errors in complex task

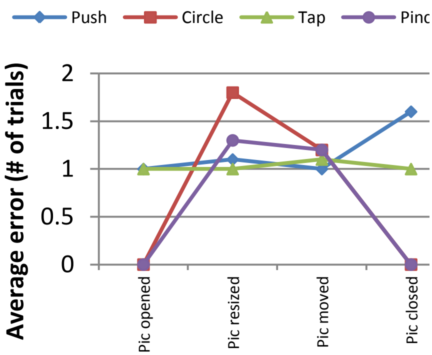



**Fig. 7.** Gestures errors in simple task

### 5.2 Discussion

Before disclosing the defined gestures to the participants, we asked them to try grabbing an object (here an icon) naturally and based on their common sense. Interestingly, 85% of participants in their first guess could correctly pick an object on screen by Pinching gesture. This anecdotal evidence of natural grabbing indicates that we have been successful in defining our finger gestures in a natural way.

This study is a supplementary work to the authors' previous research, comparing the arm gestures to mouse/keyboard [16], using the same user interface. The results in [16] are summarized as follows:

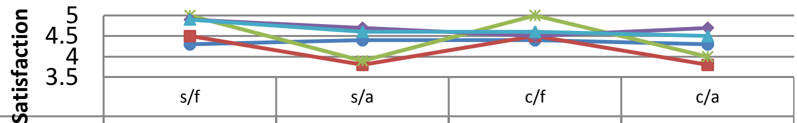
The gesture inputs are significantly slower and more fatiguing than using a mouse. Moreover, using a mouse is significantly easier than using arm gestures whiles neither inputs hold a significant popularity over the other. For the naturalness and the pleasure factors, the arm gestures as inputs do not feel significantly more natural or more fun to use compared to mouse. However, it is revealed that using arm gestures on big-screen is significantly more natural and more pleasant than using a mouse on both the desktop and the big-screen. Also it is shown that arm gestures used on big-screen is significantly more pleasant compared to when it is used on desktop.

According to the provided statistical analyses in the present study, we summarize our hypotheses verification as follows:

In general, the main result of this experiment is that fatigue is less for finger compared to arm gestures. To elaborate on, the naturalness and the fatigue factors analyses support our initial hypotheses, meaning the finger gestures significantly are more natural and cause less fatigue as inputs compared to the arm gestures. The initial hypotheses in terms of time, overall satisfactory, and easiness are rejected, implying that finger and arm maintain similar performances and popularities among participants, and neither finger gestures nor arm gestures are significantly easier than the other. Moreover, for the pleasure factor, the initial hypothesis is rejected as well, meaning the finger gestures compared to the arm gestures are not significantly more pleasant to employ as inputs. However, it is revealed that finger gestures are significantly more pleasant for simple tasks rather than complex ones.

Finally, through written feedback, most participants preferred “mostly finger” as their preferred combination of using arm and/or finger gestures, which is in positive correlation with the findings of this study.

As shown in Fig. 8, using arm is easier in short term (simple tasks). However, it is easier to use finger in the long run (complex tasks). In addition, using finger in the short term is the most pleasant, and in the long term is the least pleasant. The overall satisfaction had its highest level on the simple task using finger gestures, and its lowest level on the complex task using arm gestures.




| | s/f | s/a | c/f | c/a |
|---|---|---|---|---|
| Easy | 4.3 | 4.4 | 4.4 | 4.3 |
| Fatigue | 4.5 | 3.8 | 4.5 | 3.8 |
| Natural | 5 | 3.9 | 5 | 4 |
| Pleasant | 4.9 | 4.7 | 4.5 | 4.7 |
| Overall | 4.9 | 4.6 | 4.6 | 4.5 |

**Fig. 8.** Satisfaction comparison (s: simple, c: complex, f: finger, a: arm)

## 6 Conclusion

Using Kinect depth-based cameras along with OpenNI, NITE, and OpenCV, a gesture detection and recognition system has been developed and linked to a simulated desktop environment. Gesture studies were carried out and natural and intuitive gestures were chosen and utilized for performing various tasks in the environment. Two sets of gestures were designed, one using the arm and one using fingers. The performed tasks were designed with different difficulty levels for extraction of more information regarding the system at hand.

A comprehensive usability study indicated that finger-based gestures appeared more natural and less tiring for participants. However, this variable showed no significant effect on time, easiness, and overall satisfaction. Finally, through written feedback, most participants indicated that they would prefer a combination of fingers and arm gestures with “mostly fingers” as their method of choice for such applications.

The findings of this study, we believe, can be widely used for designing gesture-based systems, especially for WIMP interfaces. In general, finger gestures would be preferred, especially for longer lasting application which can cause more fatigue. For more rare functionalities, however, arm gestures would also be a valid choice.

## References


1. Marcel, S., Bernier, O., Viallet, J.E., Collobert, D.: Hand Gesture Recognition Using Input-Output Hidden Markov Models. In: Conference on Automatic Face and Gesture Recognition (2000)
2. Chen, F., Fu, C., Huang, C.: Hand Gesture Recognition Using a Real-Time Tracking Method and Hidden Markov Models. Image and Vision Computing, 745–758 (2003)
3. Liu, Y., Zhang, P.: Vision-Based Human-Computer System Using Hand Gestures. International Conference on Computational Intelligence and Security, 529-532 (2009)
4. Yu, C., Wang, X., Huang, H., Shen, J., Wu, K.: Vision-Based Hand Gesture Recognition Using Combinational Features. In: Sixth International Conference on Intelligent Information Hiding and Multimedia Signal Processing, pp. 543–546 (2010)
5. Raheja, J.L., Shyam, R., Kumar, U., Prasad, P.B.: Real-Time Robotic Hand Control Using Hand Gestures. In: Second International Conference on Machine Learning and Computing, Bangalore, India, pp. 12–16 (2010)
6. Chen, Q., Cordea, M.D., Petriu, E.M., Varkonyi-Koczy, A.R., Whalen, T.E.: Human-Computer Interaction for Smart Environment Applications Using Hand-Gesture and Facial-Expressions. International Journal of Advanced Media and Communication 3(1/2), 95–109 (2009)
7. Cabral, M.C., Morimoto, C.H., Zuffo, M.K.: On the Usability of Gesture Interfaces in Virtual Reality Environments. In: Proc. of the Latin American Conf. on Human-Computer Interaction, Cuernavaca, Mexico, pp. 100–108 (2005)
8. Villaroman, N., Rowe, D., Swan, B.: Teaching Natural User Interaction Using OpenNI and the Microsoft Kinect Sensor. In: Proc. of the 2011 Conference on Information Technology Education, West Point, New York, USA, pp. 227–232 (2011)
9. Kang, J., Seo, D., Jung, D.: A Study on the Control Method of 3-Dimensional Space Application Using KINECT System. International Journal of Computer Science and Network Security (2011)
10. Bragdon, A., DeLine, R., Hinckley, K., Morris, M.R.: Code Space: Touch+Air Gesture Hybrid Interactions for Supporting Developer Meetings. In: Proc. of ACM International Conference on Interactive Tabletops and Surfaces, Kobe, Japan (2011)
11. Bhuiyan, M., Picking, R.: A Gesture Controlled User Interface for Inclusive Design and Evaluative Study of Its Usability. Journal of Software Engineering and Applications (2011)
12. Ebert, L.C., Hatch, G., Ampanozi, G., Thali, M.J., Ross, S.: You Can't Touch This: Touch-free Navigation Through Radiological Images. SAGE Publications (2011)
13. Van Dam, A.: Post-WIMP User Interfaces. Communications of the ACM 40(2), 63–67 (1997)
14. Vatavu, R., Pentiuc, S., Chaillou, C.: On Natural Gestures for Interacting in Virtual Environments. Advances in Electrical and Computer Engineering 5(24), 72–79 (2005)
15. Pavlovic, V., Huang, T.: Visual Interpretation of Hand Gestures for Human-Computer Interaction: A Review. IEEE Transactions on Pattern Analysis and Machine Intelligence 19(7), 667–693 (1997)
16. Farhadi-Niaki, F., GhasemAghaei, R., Arya, A.: Empirical Study of a Vision-based Depth-Sensitive Human-Computer Interaction System. In: 10th Asia Pacific Conference on Computer Human Interaction, Matsue, Japan, pp. 101–108 (2012)