Journal of the Ergonomics Society of Korea
Vol. 31, No. 4 pp.577-583, August 2012 http://dx.doi.org/10.5143/JESK.2012.31.4.577

Gesture based Natural User Interface for e-Training

C. J. Lim[1], Nam-Hee Lee[2], Yun-Guen Jeong[2], Seung-Il Heo[3]

[1]Department of Game and Multimedia Engineering, Korea Polytechnic University, Siheung, 429-793
[2]R&D Team, Media Interactive Incorporate, Seoul, 152-780
[3]R&D Team, Media4th Incorporate, Seoul, 137-825

ABSTRACT

Objective: This paper describes the process and results related to the development of gesture recognition-based natural user interface(NUI) for vehicle maintenance e-Training system. **Background:** E-Training refers to education training that acquires and improves the necessary capabilities to perform tasks by using information and communication technology(simulation, 3D virtual reality, and augmented reality), device(PC, tablet, smartphone, and HMD), and environment(wired/wireless internet and cloud computing). **Method:** Palm movement from depth camera is used as a pointing device, where finger movement is extracted by using OpenCV library as a selection protocol. **Results:** The proposed NUI allows trainees to control objects, such as cars and engines, on a large screen through gesture recognition. In addition, it includes the learning environment to understand the procedure of either assemble or disassemble certain parts. **Conclusion:** Future works are related to the implementation of gesture recognition technology for a multiple number of trainees. **Application:** The results of this interface can be applied not only in e-Training system, but also in other systems, such as digital signage, tangible game, controlling 3D contents, etc.

Keywords: Gesture recognition, NUI(Natural User Interface), e-Training, Gesture protocol, Tracking system

1. Introduction

e-Training이란 업무에 필요한 수행능력을 습득, 향상시키기 위하여 정보통신 기술(시뮬레이션, 3D 가상현실, 증강현실 등), 장비(PC, 태블릿, 스마트 폰, HMD 등), 환경(유무선 인터넷, 클라우드 컴퓨팅 등)을 활용하여 실시하는 교육훈련을 말한다. e-Training의 대표적인 예로는 장비형 시뮬레이션을 이용한 의료실습이 있다. 이 훈련시스템은 사용자가 시뮬레이션 되는 가상의 수술환경을 화면으로 보면서 동기화된 수술 장비를 조작하여 가상의 환자를 치료하는 방식이다(Smith, Ian F. et al., 2008).

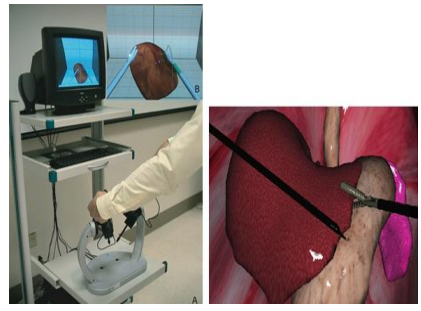

Figure 1. Application of e-Training for medical exercise in simulation environment

Figure 1과 같이 가상의 수술환경을 화면으로 보면서 동기화된 수술 장비를 조작하여 가상의 환자를 치료하는 방식이다. 하지만 이러한 방식은 해당 시스템을 구축하기 위한 비용이 많이 들며, 고가의 장비이므로 유지보수가 어려우며

Corresponding Author: C. J. Lim. Department of Game and Multimedia Engineering, Korea Polytechnic University, Siheung, 429-793.
E-mail: scjlim@kpu.ac.kr
Copyright@2012 by Ergonomics Society of Korea(pISSN:1229-1684 eISSN:2093-8462). All right reserved.
ⓒThis is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License(http://creativecommons.org/licenses/by-nc/3.0/), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited. http://www.esk.or.kr

시스템에 설치되어 있는 장치들을 조작하는데 있어서 새롭고 어려운 인터페이스를 배워야 하는 단점이 있다. 이러한 단점을 개선하기 위한 방법으로 최근에는 장치 인터페이스가 필요 없는 무구속 동작인식 인터페이스가 도입되고 있다. 게임분야에서는 이미 Xbox 360 kinect 인터페이스 등에서 상용화되어 서비스 되고 있고 향후 모든 산업분야와 일상생활에 빠르게 확산될 것이다.

e-Training 분야 중 자동차 정비 실습훈련에 동작인식 기반 인터페이스를 적용할 수 있다. 이를 통해 기존 이러닝(e-Learning)의 이론형 교육에서 벗어나 실재감, 몰입감을 향상시킬 수 있는 가상현실, 증강현실 기반의 체험형 훈련시스템을 구축할 수 있다. 본 논문에서는 자동자 정비 실습훈련을 자기주도적으로 수행할 수 있는 가상현실 시스템을 위한 동작인식(제스쳐)기반 인터페이스를 제안하고 개발하여 적용해 봄으로써 가능성과 한계점을 분석해 보았다(Park, Do. Yun. and Lee, Ji. Hyun., 2010).

해당 서비스에 적용되는 관련 기술은 Figure 2와 같이 머리와 손의 제스처(움직임)를 이용한 콘텐츠를 제어하는 것으로 머리 움직임(Head Gesture)과 손 움직임(Hands Gesture) 정보를 활용하여 자동차 정비훈련에 관련된 콘텐츠(자동차 엔진 3D 모델 등)를 조작해 볼 수 있고 조립/분해 과정을 경험해 볼 수 있도록 하는 것이다(Park, K. S. and Lim, C. J., 1999).

Figure 2. Content control technology using multimodal user interface

Figure 3과 같이 기존의 훈련모형을 가상현실 형으로 재구성하여 원격지에서도 자기주도적으로 자동차 정비훈련을 수행할 수 있는 시스템의 개념을 보여준다.

Figure 3. Self-directed e-Training service for vehicle maintenance training

2. Depth Camera

머리와 제스쳐 인식을 이용한 콘텐츠의 제어를 위해서는 동작인식 모듈 및 디바이스(센서)가 필요하다.

Prime-Sense는 독자적인 적외선 기반의 3차원 깊이 감지 기술을 이용하여 영상기반의 고속 Feature 추출, Skeleton Tracking이 가능한 PS1080 SOC 및 동작인식용 Open Source API solution(Open NI)을 출시하였으며, 해당 칩이 마이크로소프트의 키넥트 센서에 적용되었다(Figure 5). 최근 영상(Vision)기반 인식시스템을 구현할 때 Prime-Sense의 Depth 카메라를 활용하는 것이 영상처리 과정을

Figure 4. Gesture recognition by depth

단순화, 효율화하여 처리하기 때문에 많이 활용되고 있다 (Figure 4).

관절인식은 관절추적 시스템에 의해 사람이 동작인식 디바이스의 인식 범위(1.2~3.5m)에 있을 경우 Figure 6과 같이 손바닥, 머리, 목, 발, 팔 등의 뼈대의 좌표가 실시간으로 생성되는 방식을 채택하였다(Adam Haber, 2011).

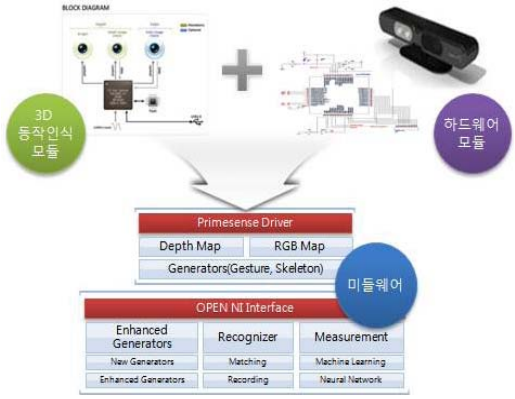

Figure 5. Natural user interface of gesture recognition device and its middleware

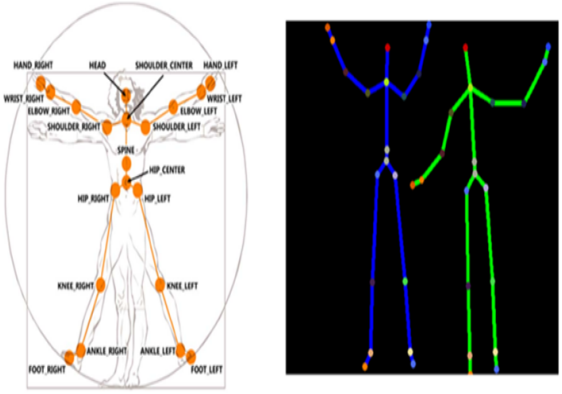

Figure 6. Skeletal tracking system

이러한 방식을 살펴보면 Figure 7처럼 각 뼈대 별로 고유 명칭이 정해져 있고, 각 명칭마다 고유번호가 부여되어 있어서(예: 머리 부분은 0번) 동작인식 API(Application Programming Interface)를 이용하여 실시간으로 해당하는 뼈대의 좌표를 추출할 수 있다. e-Training을 위한 머리, 손의 제스처를 사용하기 위해 머리, 목, 왼손, 오른손 부분의 좌표를 추출하여 사용하였다(Manin, 2011).

본 논문에서는 가상현실 시스템 상에서 입력장치로 키보드/마우스 대신 사용자가 맨손으로 화면 상의 3D 이미지를

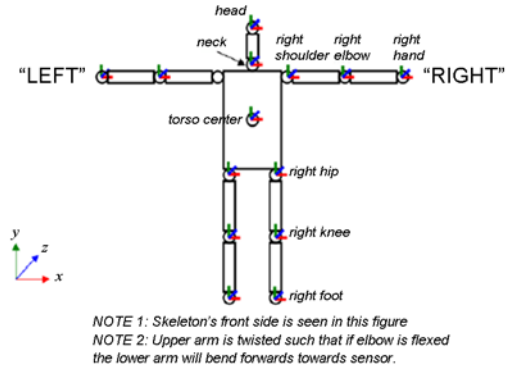

Figure 7. Joint configuration for gesture recognition

조작할(만지고 옮길) 수 있는 인터페이스를 NUI(Natural User Interface)라고 정의하였다. 이를 구현하고 자동차 정비 훈련시스템에 활용하기 위해서는 손바닥 및 손가락의 움직임에 대한 인식 및 처리가 필요하고 이를 사용자의 의도와 매칭시키는 과정이 필요하다.

3. Gesture based NUI for e-Training

Figure 8은 자기주도 e-Training 서비스의 정비 교육 훈련용 증강현실 시스템 구성도를 나타낸 것이다. 작업자는 정비 도구를 가지고 자동차 부품들을 직접 분해하고 조립하는 실습 외에도 e-Training을 이용하여 가상현실 상에서 동작인식으로 해당 동작을 반복하여 훈련할 수 있는 장점이 있다. e-Training에서 제공해야할 동작 제스처는 먼저 훈련자가 실시간으로 도구를 집는 것이다. 이를 위해 사용자가 도구를 집기 위한 손가락 좌표들이 필요하지만, 현재 일반적으로 사용되는 동작인식 모듈에서는 해당 뼈대 위치의 중심 좌표를

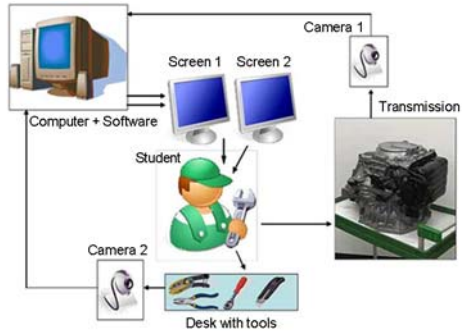

Figure 8. System configuration of augmented reality-based vehicle maintenance training

추출할 뿐, 도구를 잡는 제스처를 위한 손가락 좌표를 추출하지 못하는 기술적인 한계가 있다. 손가락 위치를 파악하기 위해서 손바닥 영상을 Open CV 이용하여 처리하였다. 그리고 사람 몸을 나타내는 깊이 정보(Depth) 영상의 색상 값을 배열로 저장하고, 저장된 배열을 영상처리를 위한 이미지(IplImage) 구조체에 저장하였다.

Figure 9. Left: depth information of image using pixel buffer; Right: output image after image processing

Open NI가 제공하는 API 함수를 이용하여 body tracking을 하여 실시간으로 3차원 손바닥 중심 좌표를 얻었고, 깊이 정보 좌표인 Z좌표에 따라 해당 손바닥의 크기를 측정하여 손바닥 영상의 영역에 해당하는 폭(width)과 넓이(height)를 구하였다.

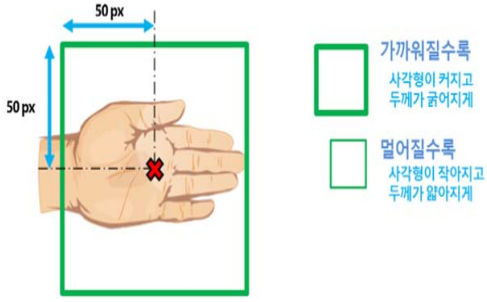

Figure 10. Variation in palm region according to depth information in real-time

본 논문에서는 손의 영역만을 빨리 처리하기 위해 영상처리의 ROI 함수를 사용하여 실시간으로 변동되는 손바닥 영역을 지정하였다. 그리고 손의 윤곽선 추출을 위해 Contour 알고리즘을 사용하였으며, 추출된 윤곽선의 좌표로 Convex Hull 알고리즘 처리를 하여 손 끝 좌표를 얻었고, 손 끝 좌표들을 배열에 저장하여 해당 좌표의 평균을 내서 실시간으로 손바닥 중심 좌표를 구하였다(Figure 11).

엄지손가락 위치를 파악하여 시계 방향(CLOCK_WISE), 혹은 반 시계 방향으로 해당 손가락 끝 좌표의 순서를 정해주었다(Figure 12).

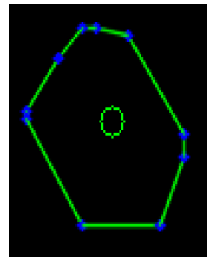

Figure 11. Center of palm after combining coordinates obtained from convex hull

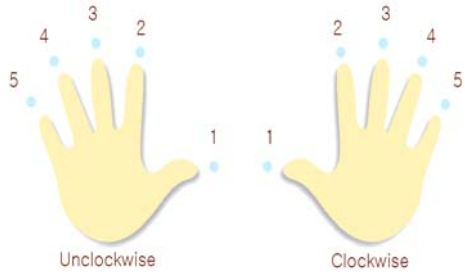

Figure 12. Order of fingertip according to thumb position

그리고 5개의 손가락 끝 좌표와 손바닥 중심 좌표를 검사하고 해당 좌표 중 가장 거리가 긴 좌표를 계산하면 검지손가락의 좌표가 나오게 된다(Figure 13).

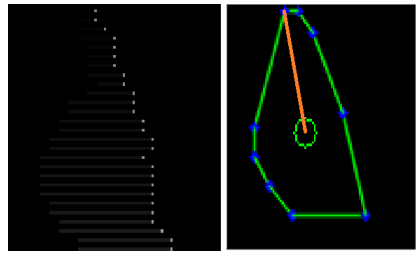

Figure 13. Output image when index finger is stretched

또한 Figure 14와 같이 손바닥 중심 좌표를 중심으로 한 원의 영역을 선정하여 손가락들이 원의 영역 안으로 모두 들어올 경우 손을 쥐었음을 구분하며, 영역 밖으로 나가게 되면 손이 펴짐을 구분할 수 있는데 이러한 영역에 해당하는 원의 반지름 길이는 중심 좌표와 손가락 끝 점 사이 거리 중 가장 긴 거리의 80% 정도로 정하였다. 향후, 이 비율이 정확도에 미치는 영향 등은 사용자 테스트를 통해 분석할 필요가 있다.

이 방식을 e-Training 시스템에 적용하면 손 모양의 커

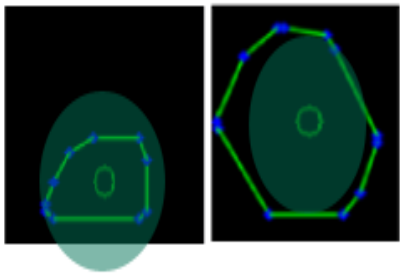

Figure 14. Output images when hand is fisted(left) and released(right)

서가 해당 도구에 다가갔을 때, 손을 움켜쥐게 되면 도구가 선택이 되고, 손을 움켜쥔 채로 움직임에 따라 도구도 같이 움직이는 과정을 구현할 수 있다(Figure 15). 반대로 손을 펼치면 손 모양 커서는 도구를 놓고 더 이상 커서와 도구가 같이 이동하지 않게 된다. 현재의 동작인식 기술과 영상처리 과정을 조합시키면, 기존 RGB 카메라를 이용한 방법에서 영상처리 부분에서 발생한 속도저하 문제를 크게 개선시킬 수 있다. 실제로 동작인식 디바이스에서 영상처리만을 이용할 경우 평균 fps는 15~20fps였지만, 영상처리와 동작인식 SDK의 장점을 활용한 결과 평균 fps가 45~50fps로 처리 속도가 향상되었다. 그 이유는 기존 RGB 카메라 방식의 경우 실시간으로 피부색을 검출하고 손 및 손 중심 좌표를 위한 프로세스를 거쳐야 하지만 동작인식 디바이스에서는 이 과정을 SOC로 처리하기 때문이다.

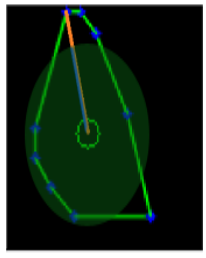

Figure 15. Output image when radius of circle is applied in real-time

Figure 16. Output images when hands are fisted(top) and tool is moved(bottom)

4. Gesture based NUI Protocols

e-Training 시스템의 훈련 시나리오는 먼저 대형 스크린에서 강사가 자동차 부품 등을 시각화하여 보여주는 단계, 조립/분해 동작을 사용자 별로 숙지하고 직접 체험하는 단계 등으로 구성하였다.

다음은 대형 스크린에서 강사가 자동차 부품 등을 시각화하여 보여주는 동작에 관한 프로토콜에 대한 설명이다. 제스처 프로토콜에 관한 연구는 여러 분야에서 연구되어 왔지만 적용분야에 따라 사용편의성 등에 차이가 있기 때문에 본 논문에서는 직접 단순하고 직관적인 동작 프로토콜을 설계를 하고 이를 적용시켰다. Figure 17의 Volume Up / Down은 손을 쥔 상태로 1초간 대기 후 위로 올리면 볼륨이 증가하며, 반대의 경우는 볼륨이 감소하는 제스처를 적용한 모습이다.

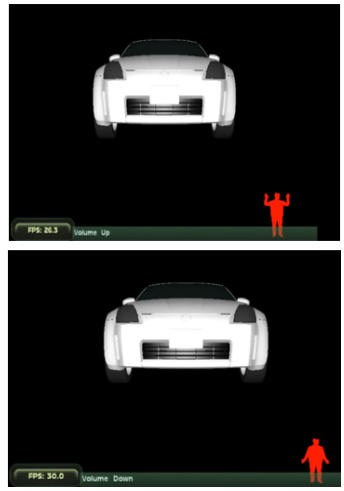

Figure 17. Output images when volume is increased(top) and decreased(bottom)

객체 선택의 경우 양 손의 좌표를 실시간으로 충돌처리하여 겹치게 될 경우 객체를 선택하는 방식을, 객체 Drag의 경우 손을 쥔 상태로 1초간 대기하고 몸을 기준으로 손을 앞으로 내밀게 되면, 해당 객체가 실시간으로 해당 손의 움직임에 따라 움직이게 되는 방식을 적용하였다(Figure 18).

X, Y축 회전은 Figure 18과 같이 손을 쥔 상태로 1초간

대기 후 머리 위로 올리면 객체의 X축 회전이 가능하고, Y축 회전은 1초 대기 후 몸을 경계로 왼쪽, 혹은 오른쪽 범위에 손이 있으면 Y축 회전이 되게 하는 방식으로 하였다.

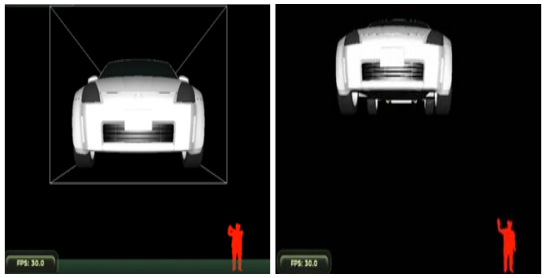

Figure 18. Left: selecting object; Right: dragging object

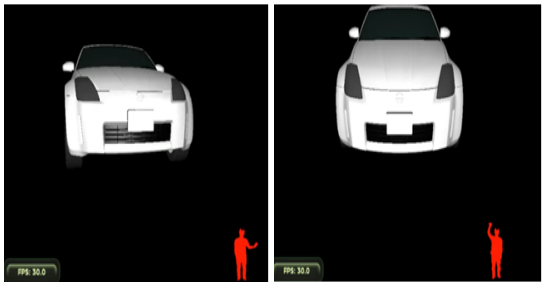

Figure 19. Left: x-axis rotation; Right: y-axis rotation

Figure 20 및 21과 같이 머리(고개)를 좌측, 혹은 우측으로 기울이게 되면 객체를 중심으로 카메라가 Y축 회전을 하게 되며, 머리를 앞, 혹은 뒤로 기울이면 객체를 중심으로 카메라가 앞, 뒤로 이동하게 되는데 전자의 경우 훈련자가 객체의 옆면, 혹은 앞, 뒤로 볼 때의 제스처를 매핑하여 적용하였다.

객체 확대, 축소의 경우는 Figure 22에서와 같이 양 손을 몸 앞으로 1초간 대기하면 양 손의 간격에 따라 해당 객체의 확대, 축소가 이루어지는 멀티터치 방식을 사용했다.

Figure 20. Rotating camera about y-axis according to head movement

Figure 21. Moving camera front(left) and back(right) according to head movement movement

Figure 22. Output images when object is zoom in(left) and zoom out(right) using multi-touch application

5. Conclusion

본 논문에서는 동작인식(제스처)기반 NUI를 자동차 정비 훈련시스템에 적용할 때 제스처 인식시스템 개발 방법, 동작 프로토콜 등을 제안하였고, 실제 구현을 통하여 가능성과 한계점을 파악하였다.

기존의 2D RGB 카메라를 이용한 영상처리에 비해 3D Depth Camera를 이용한 방법은 인식정확도와 처리속도의 향상을 기대할 수 있었고, 본 논문에서는 손가락 인식을 포함하여 자동차 부품 조립/분해 훈련시스템에 적용 가능한 NUI 시스템을 개발하였다. 본 연구결과를 실제 상용화 적용이 가능한 수준으로 높이기 위해서는 다양한 사용자에 대한 User Test를 수행하여 그 결과를 사용자 프로토콜 및 개발 방향에 피드백해야 할 것으로 판단된다. 이와 관련하여 추후 연구할 것은 ISO 9241-9 '키보드 방식이 아닌 입력장치에 대한 인간공학적 요구사항'에 따라 Fitt's Law 및 Tapping Test 등이다. 또한 자동차 정비 훈련모형 등을 반영하여 반복훈련 테스트/평가 등 사용자 동작 평가에도 적용할 수도

있다고 생각된다.

향후, 동작인식 칩의 성능향상이 이루어지고, 이에 수반된 관련 SDK도 버전업(Version Up)이 되면서, 현 동작인식 기술의 단점으로 지적되고 있는 깊이 정보 카메라의 음영지역(적외선이 투사되지 못하여 생기는 겹침 현상)이나 동작인식 시 일시적인 떨림 현상, Latency(사람의 동작속도 대비 90%, 사람의 동작속도를 동작인식에서 조금 뒤처지는 현상)의 문제 등이 개선될 것이다. 다시 말해 e-Training에 적용되는 현재의 동작인식 기술로는 훈련자의 세부적인 동작을 인식하지 못 하는 한계가 있다. 그러나 단순히 눈으로 보고 듣고 하는 멀티미디어 교육의 한계점을 보완할 수 있는 대안이 가상현실이며, 동작인식 NUI 기술을 적용함으로써 기존 단방향의 멀티미디어 교육의 한계를 일부 충족시켜 줄 수 있다. 기존 e-Training 시스템은 초기 구축비용이 높고, 직업훈련 수단으로서 성공사례가 부족하여 산업현장의 교육훈련에 활용하는 경우는 미미한 수준이지만, 본 논문에서 새롭게 제안된 시스템은 산업현장의 인력훈련 수요에 부합하면서도 고가 장비에 대한 투자의 최소화 및 훈련 위험의 회피에 적합한 새로운 e-Training NUI 시스템 개발이 이루어질 수 있다는 점에서 의미가 있다.

Acknowledgements

This research was supported by grants from the R&D Program (Industrial Strategic Technology Development) funded by Ministry of Knowledge Economy(MKE), and the Culture Technology Research & Development Program funded by Ministry of culture, Sports and Tourism(MCST) and Korea Creative Content Agency (KOCCA).

References

Adam, Haber., Visualization and Simulation of 3D depth data from the Kinect Camera, http://cgi.cse.unsw.edu.au/~cs4411/wiki/index.php?title=Adam, (retrieved June 3, 2011).

Broekman, B., et al. Embedded Software Testing, Hongrung publishing company, 2008.

Fukuhara, et al. 3D-Motion Estimation of Human Head for Model-Based Image Coding. IEE Proc., Vol. 140, no. 1, 26-35, 1993.

Gwon, Wonil., Testing Process in ISO/IEC 29119-2, Software Engineering: Software Testing, 2008.

IEEE Standard for Software Testing Documentation, http://blog.naver.com/raindrap?Redirect=Log&logNo=40048822447, (retrieved March 12, 2008).

Manin., Kinect for Windows SDK, http://www.cnblogs.com/aawolf/archive/2011/06/21/2086139.html, (retrieved June 21, 2011).

Park, Doyun and Lee, Ji-Hyun., Investigating the Affective Quality of Motion in User Interfaces to Improve User Experience, 67-78, 2010.

Park, K. S. and Lim, C. J., "An efficient camera calibration method for vision-based head tracking", Man-Machine Production Systems Laboratory, Department of Industrial Engineering, Korea advanced Institute of Science and Technology, 1999.

Smith, Ian F., et al. Test automation for embedded products, 2004.

The innovation business of the knowledge economic technique, e-training service for practicing automobile maintenance, (pp. 3-20), 2011.

The Ministry knowledge economy Technology and Standard, Remote control equipment based on motion, 2012.

Author listings

C. J. Lim: scjlim@kpu.ac.kr

Highest degree: PhD, Department of Industrial Engineering, KAIST
Position title: Professor, Department of Game and Multimedia Engineering, Korea Polytechnic University
Areas of interest: HCI/VR, Serious Game Design

Nam-Hee Lee: nutrient2002@hanmail.net

Highest degree: Bachelor's degree, Department of Game and Multimedia Engineering, Korea Polytechnic University
Position title: Researcher, R&D Team, Media Interactive Incorporate
Areas of interest: Gesture Recognition, System Integration

Yun-Guen Jeong: sadmyth7@gmail.com

Highest degree: MD, Department of Information and Communication Engineering, Soon Chun Hyang University
Position title: Researcher, R&D Team, Media Interactive Incorporate
Areas of interest: HCI, Serious Game

Seung-Il Heo: rookeh1@media4th.co.kr

Highest degree: MD, Department of Visual design, Kookmin University
Position title: CEO, R&D Team, Media4th Incorporate
Areas of interest: UI/UX design, Digital Marketing

Date Received : 2012-07-17

Date Revised : 2012-07-23

Date Accepted : 2012-07-23