

3D 콘텐츠 제어를 위한 키넥트 기반의 동작 인식 모델

Kinect-based Motion Recognition Model for the 3D Contents Control

최한석

목포대학교 멀티미디어공학과

Han Suk Choi(chs@mokpo.ac.kr)

요약

본 논문에서는 키넥트 적외선 프로젝터를 통해 깊이를 감지할 수 있는 카메라를 이용하여 사람 움직임을 추적하고 본 논문에서 제안한 몸동작 모델 인식을 통하여 3D 콘텐츠를 제어하는 기법을 제안 한다. 본 논문에서 제안하는 사람의 동작 인식 모델은 사람의 오른팔과 왼팔의 손목, 팔꿈치, 어깨 움직임의 거리를 계산하여 좌, 우, 상, 하, 확대, 축소, 선택 등의 7가지 동작 상태를 인식한다. 본 연구에서 제안한 키넥트 기반의 동작 인식 모델은 기존의 접촉식 방식의 인터페이스와 비교할 때 특정센서 또는 장비 부착에 대한 불편함을 없애고 고비용의 하드웨어 시스템을 이용하지 않음으로서 사람의 자연스런 몸동작 이동에 따른 저 비용 3D 콘텐츠 제어 기술을 보여준다.

■ 중심어 : | 3D 콘텐츠 | 키넥트 | 동작인식 | 사용자 체험형 인터페이스 | 제어 모델 |

Abstract

This paper proposes a kinect-based human motion recognition model for the 3D contents control after tracking the human body gesture through the camera in the infrared kinect project. The proposed human motion model in this paper computes the distance variation of the body movement from shoulder to right and left hand, wrist, arm, and elbow. The human motion model is classified into the movement directions such as the left movement, right movement, up, down, enlargement, downsizing, and selection. The proposed kinect-based human motion recognition model is very natural and low cost compared to other contact type gesture recognition technologies and device based gesture technologies with the expensive hardware system.

■ keyword : | 3D Contents | Kinect | Motion Recognition | User Experienced Interface | Control Model |

1. 서론

최근 들어 사람과 디지털 기기간의 상호작용에 의한 인터랙션 기술이 실감콘텐츠 기술 발전과 매우 중요한 기술로 인식되고 있다. 각종 센서, 컴퓨팅 파워, 인터페이스 기술 발전은 인간과 디지털 기기간의 스마트한 인터랙션을 보여 주고 있고, 이에 대한 구체적인 사례들은 안경, 시계, 의복과 같은 착용형 단말 인터페이스 방법과 손동작, 몸동작, 얼굴 추적 등 인간의 신체 움직임

을 인식하고, 이를 해석하여 디지털 디바이스들을 제어하는 동작 인식 인터페이스 기술로 구분 된다[1].

사람의 동작인식 기술은 센서나 특정 장치를 사람의 신체에 부착하여 신체의 움직임으로 감지한 데이터를 동작인식에 활용하는 접촉식 방식과 카메라를 이용하여 사람의 움직임 영상을 추적하여 동작인식 정보를 추출하는 비 접촉식 방식으로 구분 된다. 접촉식 방식은 사용자의 움직임을 감지할 수 있는 센서나 장치를 사람의 몸에 직접 부착해 정보를 획득하는 방식으로 직접

센서를 통하여 비교적 정확한 정보를 얻을 수 있는 장점이 있지만, 사용자가 센서나 장비를 착용해야하는 불편함이 매우 큰 단점이다. 비 접촉식 방식은 대부분 사람의 움직임을 영상으로 촬영한 후 움직임 정보를 추출한 비전 기술을 이용 한다[2][3][5].

본 논문에서는 키넥트 적외선 프로젝터를 통해 깊이를 감지할 수 있는 카메라에 찍힌 영상속의 사람 움직임을 추적하고 본 논문에서 제안한 몸동작 모델 인식을 통하여 3D 콘텐츠를 제어하는 기법을 제안 한다. 본 논문에서는 키넥트 SDK를 이용하여 사람 신체 부위의 3D 위치를 분석하여 사람의 골격(Skeleton)을 생성한 후, 신체 길이와 관련된 연결 포인트의 거리를 계산하여 사람의 동작 모델을 인식한다. 본 논문에서 사람의 동작 모델은 오른팔과 왼팔의 손목, 팔꿈치, 어깨 움직임의 거리를 계산하여 좌, 우, 상, 하, 확대, 축소, 선택 등의 7가지 동작 상태를 파악한다. 이러한 7가지 동작 상태의 값의 변화를 추적하여 3D 콘텐츠 제어를 위한 동작 인식 모델을 설계하고, 이를 사전에 제작된 3D 콘텐츠 제어를 위한 사용자 체험형(User eXperienced : UX) 인터페이스로 활용한다.

본 연구에서 제안한 키넥트 기반의 동작 인식 모델은 기존의 접촉식 방식의 인터페이스와 비교할 때 특정센서 또는 장비 부착에 대한 불편함을 없애고 고비용의 하드웨어 시스템을 이용하지 않음으로서 사람의 자연스런 몸동작 이동에 따른 저 비용 3D 콘텐츠 제어 기술을 보여준다.

본 논문의 구성은 2장은 관련 연구로서 접촉식 방식의 3D 콘텐츠 제어 인터페이스 기술과 키넥트 기반의 비 접촉식 인터페이스 기술 특성을 기술한다. 3장에서는 키넥트 기반의 동작 인식모델 기능을 설계하고, 키넥트 기반의 동작 모델 인식 알고리즘을 제안한다. 4장에서는 구현환경 및 실험 결과를 제시하고, 5장에서는 본 논문의 결론 및 향후 연구 방향을 기술한다.

II. 관련 연구

2.1 접촉식 방식의 3D 콘텐츠 제어 인터페이스

접촉식 방식의 3D 콘텐츠 제어 방법은 사용자자의

움직임 감지가 가능한 센서나 장치를 사용자가 직접 착용하여 3D 콘텐츠와 상호작용하는 것이다. 접촉식 방식으로 사용되는 대표적인 인터페이스는 광섬유, 마크 네틱센서, 자이로 센서를 이용한 데이터 글로브, 모션 트랙커 등이 있다[2]. 접촉식 방식에 의한 입력 장치들은 대부분 상용화 되어 있다. 예를 들어 3D로 제작된 네비게이션 콘텐츠에서 터치기반 3D 인터랙션 조작 인터페이스, 닌텐도의 Wii Remote 모션 컨트롤러 인터페이스, 소니사의 PlayStation Move 인터페이스, 스페이스 마우스를 대처하는 GlobalFish 인터페이스, 조이스틱 기능을 탄성체 링으로 만들어 터치패드를 내장한 Groove Pad 인터페이스 등 다양한 3D 물리적 인터페이스가 소개 되고 있다[3-5][7].

2.2 키넥트 기반의 비접촉식 인터페이스

비 접촉식 방식 기반의 3D 콘텐츠 제어 방법은 대부분 카메라를 통하여 획득한 영상정보로부터 사람의 움직임을 추적하고 인식한다. 이러한 컴퓨터 비전 기반의 비 접촉식 기법은 초기화(전 처리), 추적, 사용자 동작 예측, 동작 인식 과정을 통하여 사용자의 동작 모델을 알아내고 이러한 동작 인식을 통하여 3D 콘텐츠와 상호작용한다. 비 접촉식 방식은 신체 부위에 특정 마커를 부착하여 시각 기술을 이용하여 신체 움직임을 추적하는 하마 마커 기반 방식과 전혀 마커를 사용하지 않은 마커리스 방식 두 가지가 있다[5].

마커 방식은 영상의 컬러, 형태, LED, 적외선 등을 마커 속성에 미리 정해두고 그 마커를 추적하는 것으로 비교적 쉽고 빠르게 사용자의 동작을 추적할 수 있다. 이 방법은 영상의 특징점 추출을 상대적으로 빠르게 할 수 있으나 사용자는 여전히 부자연스러운 마커를 부착해야 하는 단점이 있다. 마커리스 방식은 옵티컬 플로우, 배경분리 기법, 모션 히스토리이미지 등의 기법을 이용하여 사용자 움직임의 방향 및 속도 등을 감지하여 움직임을 실시간으로 빠르게 추적할 수 있고 광원의 밝기를 추적하여 움직임을 추적하기도 한다. 마커리스 방식은 카메라의 위치, 광원의 간섭, 그림자 등과 같은 외부 요인에 취약한 단점이 있다.

키넥트는 이러한 단점을 보완하기 위하여 다수 카메

라 및 적외선 카메라를 이용하여 사람의 움직임에 대한 3차원 깊이 정보를 추적한다. 키넥트를 이용한 사람의 동작인식은 카메라를 통한 사람의 동작을 RGB 화상으로 입력을 받는 카메라와 Laser-Pointing 후 깊이를 감지하는 카메라로 구성하고, 입력된 2개의 화상 정보를 디지털 필터링을 통해 가공 후 사람의 동작 인식 정보를 3D Skeleton 데이터로 생성 한다. 생성된 사람의 동작 3D Skeleton 정보는 키넥트의 제스처 인식 알고리즘을 통해 선형데이터로 재가공 하여 UX 기반 비 접촉식 인터페이스로 3D 콘텐츠를 제어한다. 또한, Kinect 센서는 레이저를 이용해 사용자에게 적외선을 투사하고 연결점(dot)들의 반사파 강도를 측정하고, 반사강도를 통해 거리를 측정하고 반사강도가 약한 점들은 멀리서부터, 강도가 높은 점들은 전면에 있는 사용자로부터 온 것으로 추정한다. 이를 통해 사람의 주요 뼈마디를 인식해 내고 신체 움직임을 감지한다[3][5].

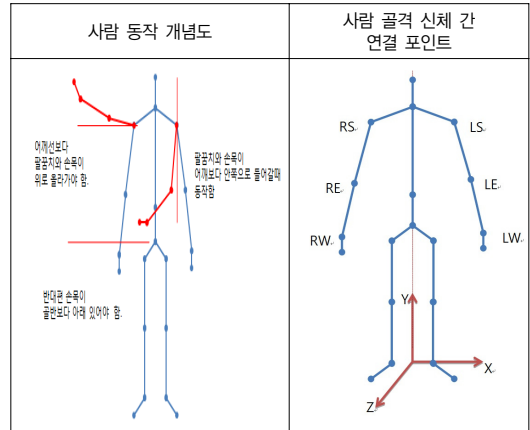


그림 1. 키넥트 기반의 사람 동작 모델 기본 개념도

III. 키넥트 기반의 동작 인식 모델 설계

3.1 동작 모델 기능 설계

키넥트 적외선 카메라가 사람의 골격을 인지하는 기법을 이용하여 사람의 손목, 팔목, 어깨 움직임에 기반한 7가지 동작 기능 모델을 설계한다[6][8]. [그림 1]은 키넥트를 이용한 사람의 동작 모델 기본 개념도 이다. 사람의 동작 모델 기능은 손목, 팔꿈치, 어깨선을 중심으로 좌로 이동, 우로 이동, 위로 이동, 아래로 이동, 확대, 축소, 선택 등 7가지 기능으로 구분한다.

좌로 이동은 오른팔을 이용하여 어깨선을 중심으로 손목, 팔꿈치 변화가 좌측으로 이동하는 것을 말하고, 우로 이동은 좌측 팔이 어깨선을 기준으로 손목, 팔꿈치가 우측으로 이동하고, 위로(상) 이동은 어깨선 보다 위로, 아래로(하) 이동은 어깨선보다 밑으로, 확대는 양손이 기준보다 밖으로 이동, 축소는 양손이 기준 보다 안쪽으로 이동하는 것이고, 선택은 한손을 정면으로 이동하는 것을 말한다. [그림 2]는 사람의 신체 움직임을 7가지 동작 모델을 기능별로 설계 한 것이다.

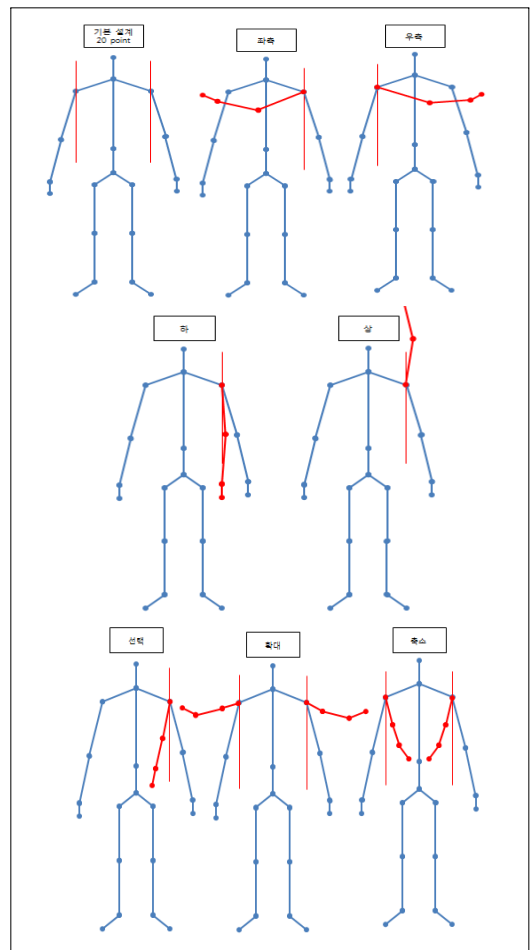


그림 2. 7가지 동작 모델 기능 설계

3.2 키넥트 기반의 사람 동작모델 인식 알고리즘

본 논문에서는 [그림 2]에서 설계한 사람의 동작 모델에서 7가지 기능별 동작 인식 알고리즘을 [표 1] 같이 제안한다. [표 1]에서는 [그림 1]의 사람 신체 간 연결 포인트 위치를 설정하고, 연결 포인트 간 이동거리 값의 X, Y, Z 값 변화량을 구하여 사람의 동작 모델을 인식 한다.

표 1. 키넥트 기반의 동작 모델 인식 알고리즘

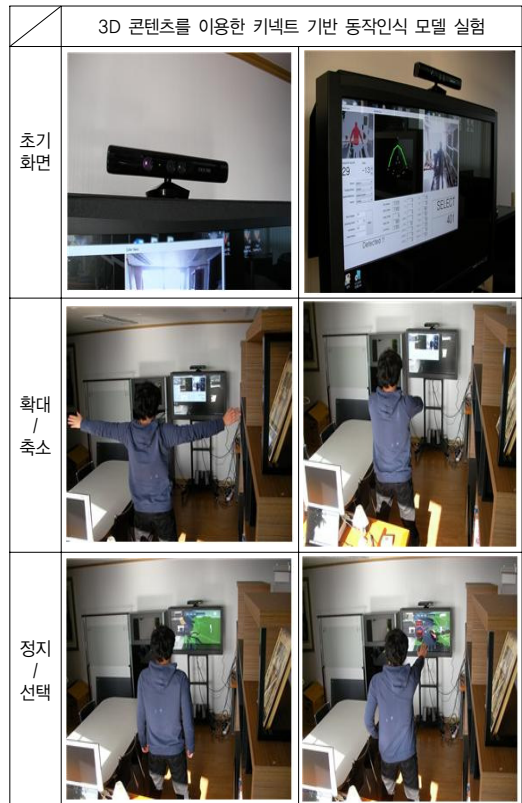
<p>단계 1) 손목, 팔꿈치, 어깨 위치에서 연결 포인트별 X, Y, Z 좌표값 계산</p> <p>손목의 Z축 위치 Z_{LW} 및 Z_{RW}, 어깨의 위치 Z_{LS} 및 Z_{RS}, 팔꿈치의 위치 Z_{LE} 및 Z_{RE} 에 대하여 X, Y, Z 좌표값을 계산. 만일, $Z_{LW} = \sqrt[3]{Z_{LS}^2 + Z_{LE}^2}$ (왼쪽 팔인 경우) 이고, $Z_{RW} = \sqrt[3]{Z_{RS}^2 + Z_{RE}^2}$ 인 (오른팔목일 경우)인 경우에 대하여,</p>	
<p>단계 2) 좌/우/상/하 이동 인식 판단</p> <p>손목의 Z축 위치 Z_{LW} 및 Z_{RW} 어깨의 위치 Z_{LS} 및 Z_{RS}, 팔꿈치의 위치 Z_{LE} 및 Z_{RE} 에 대해서, X, Y 변화량 $d\Delta = \frac{ dY_{LW} }{ dX_{LW} }$ 가 $d\Delta < 1, dX_{LW} > 0$ 이면 좌로 이동으로 판단 $d\Delta < 1, dX_{LW} < 0$ 이면 우로 이동으로 판단 $d\Delta > 1, dY_{LW} > 0$ 이면 위로(상) 이동으로 판단 $d\Delta > 1, dY_{LW} < 0$ 이면 아래로(하) 이동으로 판단</p>	
<p>단계 3) 확대 / 축소 인식 판단</p> <p>$Z_{LW} = \sqrt[3]{Z_{LS}^2 + Z_{LE}^2}$ 이고 $Z_{RW} = \sqrt[3]{Z_{RS}^2 + Z_{RE}^2}$ 인 조건에서 왼팔목(LW)의 X,Y 변화량 $d\Delta_{LW} = \frac{ dY_{LW} }{ dX_{LW} }$, 오른팔목(RW)의 X,Y 변화량 $d_{RW} = \frac{ dY_{RW} }{ dX_{RW} }$ 일때, $0 \leq d\Delta_{LW} < 1$ 이고 $0 \leq d\Delta_{RW} < 1$ 인 경우, X축 변화량 $dX_{LW} > 0, dX_{RW} < 0$ 인 경우, 확대판단 X축 변화량 $dX_{LW} < 0, dX_{RW} > 0$ 인 경우, 축소판단</p>	
<p>단계 4) 선택 인식 판단</p> <p>$Z_{LW} = \sqrt[3]{Z_{LS}^2 + Z_{LE}^2}$ 또는 $Z_{RW} = \sqrt[3]{Z_{RS}^2 + Z_{RE}^2}$ 인 조건에서 $d\Delta = \frac{ dY_{LW} }{ dX_{LW} }$ 에서, $0.9 < d\Delta < 1.1$ 경우, 선택 동작으로 인식</p>	

IV. 실험 및 평가

4.1 실험 환경 및 구현 결과

본 논문에서 키넥트 기반의 동작 모델 인식 알고리즘의 성능 실험을 위하여 H/W는 팬티엄 4 듀얼 코어 64Bit CPU i7 2600 및 16GB RAM, 라데 온 HD 6950 그래픽 카드를 가진 PC에서 키넥트 설치하여 실험하였다. S/W는 64Bit Windows 7 운영체제에서 벤투즈 3 Designer 설치하여 사전에 미리 제작된 전남 관광 문화 3D 콘텐츠를 제어하는 실험을 하였다. 키넥트 기반의 동작 모델 인식 알고리즘 개발 언어는 C#을 이용하여 비주얼 스튜디오에서 구현하였다.

[그림 3]는 키넥트기반의 동작 인식 모델을 이용한 3D 콘텐츠 제어 구현 결과이다. 본 연구에서 실험한 키넥트 기반의 사람 동작 인식 모델은 [그림 3]에서와 같이 초기화면에서 7가지 동작 모델에 따라 실험하였다.



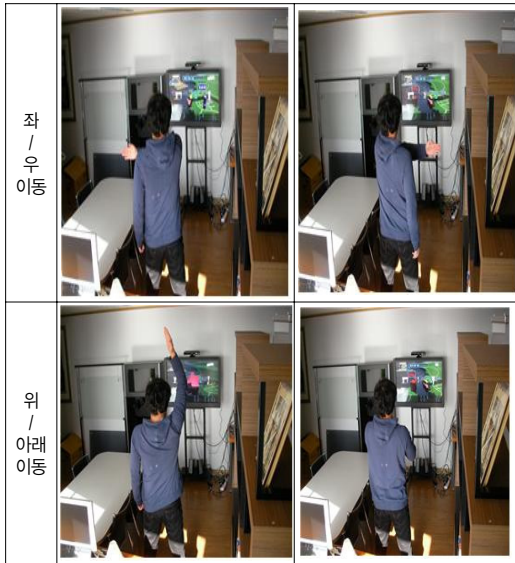


그림 3. 키넥트 기반 동작 인식 모델을 이용한 3D 콘텐츠 제어 구현 결과

4.2 평가

본 실험에서는 5명의 실험대상 학생이 10차례씩 7가지 모델에 따른 각각의 제스처를 3-4 미터, 5-6미터, 7미터 이상으로 구분하여 실험 하였다. 실험 결과 키넥트 적외선 프로젝트는 카메라로부터 3-4 미터 내에서 7가지 동작 모델별로 훈련된 사용자는 매우 자연스럽게 효과적으로 사람 동작을 99.8% 정도 인식 하였고, 5-6미터에서는 90.6% 정도 인식 하였다. 7미터 이상에서는 인식율이 좋지 않았다.

실험 결과 3-4미터 이내에서는 훈련된 사람의 동작에 의거 매우 효과적으로 3D 콘텐츠가 제어 되었고, 동작이 훈련되지 않거나 거리가 카메라로부터 멀어질 때는 동작이 효과적으로 인식되지 않았다. 따라서 수차례 실험 결과 사람의 동작 모델의 인식 거리 및 외부 조명 상태를 정립하였다. 본 연구에서는 7가지 동작모델을 중심으로 3D 콘텐츠 제어 기술을 집중 실험 하였으나, 향후 에는 현 동작 모델을 확장하여 좀 더 세부적인 손동작과 다른 몸동작 인식 모델을 제정의 하고, 해당 동작에 따른 실감형 3D 콘텐츠 제어기술을 구현할 예정이다.

V. 결론

본 논문에서는 키넥트 적외선 프로젝터를 통해 깊이를 감지할 수 있는 카메라에 찍힌 영상속의 사람 움직임을 추적하고 본 논문에서 제안한 동작 모델 인식을 통하여 3D 콘텐츠를 제어하는 기법을 제안 하였다. 본 논문에서는 키넥트 SDK를 이용하여 사람 신체 부위의 3D 위치를 분석하여 사람의 골격(Skeleton)을 생성한 후, 신체 길이와 관련된 연결 포인트의 거리를 계산하여 사람의 동작 모델을 인식한다. 본 논문에서 사람의 동작 모델은 오른팔과 왼팔의 손목, 팔꿈치, 어깨 움직임의 거리를 계산하여 좌, 우, 상, 하, 확대, 축소, 선택 등의 7가지 동작 상태를 파악하였다.

본 논문에서 제안한 키넥트 기반의 동작 인식 모델은 기존의 접촉식 방식의 인터페이스와 비교할 때 특정센서 또는 장비 부착에 대한 불편함을 없애고 고비용의 하드웨어 시스템을 이용하지 않음으로서 사람의 자연스런 몸동작 이동에 따른 저 비용 3D 콘텐츠 제어 기술을 보여준다. 본 논문에서 제안한 사용자 체형형 3D 콘텐츠 제어 기법은 가상현실, 증강현실, 3D 게임, 3D 교육용 실감 콘텐츠 등 다양한 응용 콘텐츠에 활용 가능하다. 또한, 실감형 스마트 TV 사용자 인터랙션 핵심 기술로 활용 할 수 있어 차세대 실감 산업 기술 발전에 크게 활용 가능하다.

본 연구의 향후 계획은 본 논문의 동작 모델을 확장하여 좀 더 세부적인 손동작이나 다른 몸동작 인식 모델을 정의 하고, 해당 동작에 따른 실감형 3D/4D 콘텐츠 제어기술을 더욱 심도 있게 연구할 예정이다.

참고 문헌

- [1] “심층 리포트:휴먼-디바이스 인터랙션 기술”, CT 인사이트, 통권 제27호 pp.2-39, 한국콘텐츠진흥원, 2012(12).
- [2] 김준우, 맹준희, 주지영, 임광진, “멀티터치 스크린과 실감형 인터페이스를 적용한 과학실험 학습 시스템”, 한국콘텐츠학회논문지, 제10권, 제8호,

- pp.461-471, 2010(8).
- [3] 신유경, 최종훈, “개별 사용자 환경을 위한 스마트 TV 리모트 콘트롤 인터랙션 방식 제안”, 한국콘텐츠학회논문지, 제11권, 제11호, pp.70-78, 2011(11).
- [4] 이상현, 김대진, 최홍섭, “제스처 인식을 이용한 DID 인터페이스 구현”, 한국디지털콘텐츠학회논문지, 제13권, 제3호, pp.343-352, 2012(9).
- [5] 홍동표, 우운택 “제스처기반 사용자 인터페이스에 대한 연구 동향”, Telecommunications Review, 제18권, 제3호, pp.403-413, 2008.
- [6] 조선영, 변혜란, 이희경, 차지훈, “키넥트 센서 데이터를 이용한 손 제스처 인식,” 방송공학회논문지, 제17권, 제3호, pp.447-458, 2012.
- [7] 김용원, 조동식, 김영희, 김혜미, 김기홍, “실감형 가상현실 상호작용 기술동향”, 전자통신동향분석, 사용자중심체세대콘텐츠기술특집, pp.62-72, 2012.
- [8] H. S. Choi and M. U. Kang, “A User Experienced 3D Contents Control Using the Face Tracking,” International Conference on Convergence Content 2012, pp.89-90, 2012(12).

저 자 소 개

최한석(Han Suk Choi)



정회원

- 1980년 2월 : 전남대학교 수학교육과(이학사)
 - 1983년 8월 : 미국 웨스턴일리노이 대학원 컴퓨터과학과(이학석사)
 - 1987년 2월 : 전북대학교 대학원 컴퓨터과학과(이학박사)
- 1989년 3월 ~ 현재 : 목포대학교 멀티미디어공학과 교수
- <관심분야> : UX기반 3D 콘텐츠 제어, 인터랙티브 실감 콘텐츠, 스마트 웹 컴퓨팅, 빅 데이터 분석 및 가시화 서비스