

가상현실 실감체험을 위한 인터렉션 지원 방법

홍성진, 신희숙*, 길연희**
한국전자통신연구원

sjhong0117@etri.re.kr, *hsshin8@etri.re.kr, **yhgil@etri.re.kr

Interaction Method for Realistic Experience of Virtual Reality

Sungjin Hong, Heesook Shin*, Youn-Hee Gil**
Electronics and Telecommunications Research Institute

요약

기존 HMD 컨트롤러 기반의 인터렉션은 버튼으로 사전에 정의된 동작 표현하는 방식으로 가상현실 체험의 몰입감을 떨어뜨린다. 본 논문은 HMD 와 깊이 카메라를 이용해 가상현실에서의 실감 인터렉션 지원 방법을 제안한다. 실제 손을 이용한 인터렉션을 지원하기 위해 HMD 와 깊이 카메라의 좌표계를 일치시켜 손의 움직임을 표현한다. 또한, CNN 기반의 모델을 활용하여 손 동작을 인식하고, 표현하여 실제 손을 이용한 가상현실 체험을 가능하게 한다.

I. 서론

최근 가상현실 기술은 관람, 스포츠, 게임, 교육 등 많은 분야에서 적용되어 다양한 서비스를 제공한다. 장소에 제약없이 몰입형 환경을 제공할 수 있는 장점으로 루지, 봅슬레이, 행글라이더 등과 같은 평소에 경험하기 어려운 체험을 간접적으로 제공한다[1]. 또한, 위험한 상황을 반복숙달하고 실패에도 추가비용없이 반복훈련이 가능하여 비행사 훈련[2], 바리스타 훈련[3] 등 직업훈련에도 많이 활용되고 있다.

HMD(Head Mounted Display)는 사용자의 두 눈에 몰입형 가상환경을 출력하고, 손에 움직일 수 있는 컨트롤러를 이용하여 가상환경의 인터렉션을 지원한다[4]. 하지만 대부분의 HMD 컨트롤러는 게임패드처럼 손가락으로 버튼을 누르는 형태로 제작되어 사용자 손의 제스처와 다르게 표현되고 몰입감을 떨어뜨린다. 또한, 일부 장애를 가진 사람들은 해당 버튼을 누르는 동작 조차도 어려움을 겪고 있어 다른 방식의 인터페이스 지원이 필요하다.

사용자 동작분석은 최근 깊이 카메라가 많이 활용되고 있다. 실시간으로 3 차원 정보를 획득하고, 이를 분석하여 저비용으로 객체 검출 및 인체 관절 검출이 가능하다. 또한, 마커를 부착하는 등 특정 장비 착용을 요구하지 않아 사용에 용이하다. Microsoft 는 xbox 에 키넥트를 지원하며 사용자 동작인식 기반의 콘텐츠 체험을 지원했고[5], Ultraleap Ltd.는 leap motion 을 통해 사용자 손가락의 관절정보를 제공하며 제스처를 분석을 용이하게 했다[6].

Y. S. Lee and B.S. Sohn 은 HMD 와 깊이카메라를 이용하여 사용자 가상환경을 탐험할 수 있게 했다. 이동, 회전 등의 동작을 정의하고 동작인식을 통해 가상공간 내의 캐릭터를 조작했다[7]. J.Y Ahn, et al.은 HMD 위에 RGB 카메라를 부착하여 책상 위에 놓인 손을 검출했고,

마우스의 클릭, 줌인과 같은 조작을 위한 손 제스처 인식 모델을 제안했다[8]. 하지만 두 방법 모두 카메라의 설치된 위치와 상관없이 사용자의 특정 동작을 조작하고자하는 물체의 특정 속성과 맵핑시키는 형태로 사용자의 자연스러운 움직임을 반영하지 못한다.

본 논문에서는 앞선 방법들의 문제를 해결하기 위해 HMD 와 깊이 카메라의 로컬 좌표계를 하나의 좌표계로 일치시키고, 손 동작 인식을 통한 인터렉션 지원 방법을 제안한다.

II. 본론

사용자 친화적인 가상현실 인터렉션 지원을 위한 장치 구성은 HMD 와 깊이 카메라로 구성된다. 체험 공간 내에서의 HMD 와 컨트롤러를 추적하기 위해 베이스 스테이션을 설치하고, 사용자의 동작을 추적하기 위해 정면에 깊이 카메라를 설치한다.

다음으로 가상공간내에서 컨트롤러로 활용되는 부분을 사용자의 동작으로 대체하기 위해 HMD 와 카메라 간의 로컬 좌표계를 일치시키는 작업을 수행한다. 본 논문에서는 HMD 컨트롤러가 가상공간내에서 손 역할을 한다고 가정한다. 카메라에서 검출한 손의 위치정보는 베이스 스테이션에서 바라본 컨트롤러의 위치와 대응된다. 동일한 시점에 손의 위치 데이터를 총 $t \geq 3$ 개 수집하고, SVD(Singular value decomposition)와 ICP(Iterative Closest Point)를 이용하여 대응되는 점들간의 거리 오차를 최소화하는 변환행렬을 획득한다.

다음으로는 사용자의 손을 표현하기 위한 상태 분석을 실시한다. 최근 딥러닝 기법은 객체 검출 및 인식 분야에서 안정적인 성능으로 많이 활용되고 있다[9]. 그 중 CNN(Convolutional Neural Networks)기반 모델은 공간정보를 고려할 수 있는 특징이 있다. 본 논문에서는

Layer (type)	Output Shape	Param #
input_1 (InputLayer)	(None, 36, 36, 1)	0
conv2d_1 (Conv2D)	(None, 34, 34, 64)	640
conv2d_2 (Conv2D)	(None, 32, 32, 64)	36928
max_pooling2d_1 (MaxPooling2D)	(None, 16, 16, 64)	0
conv2d_3 (Conv2D)	(None, 14, 14, 64)	36928
conv2d_4 (Conv2D)	(None, 12, 12, 64)	36928
max_pooling2d_2 (MaxPooling2D)	(None, 6, 6, 64)	0
conv2d_5 (Conv2D)	(None, 4, 4, 64)	36928
flatten_1 (Flatten)	(None, 1024)	0
dropout_1 (Dropout)	(None, 1024)	0
dense_1 (Dense)	(None, 4)	4100

Total params: 152,452
Trainable params: 152,452
Non-trainable params: 0

그림 1. 손 상태 분석을 위한 CNN 구조

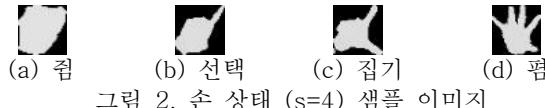


그림 2. 손 상태 (s=4) 샘플 이미지

이미지에 나타난 각 손가락 간의 위치 정보를 고려하여 CNN 기반의 모델을 정의하고 활용한다. 그림 1은 3 차원 손 이미지의 상태를 분석하기 위한 CNN 구조를 나타낸다. 가로 w와 세로 h 크기의 3 차원 정보를 가진 손 깊이 이미지를 입력받아 s 개의 손 상태를 인식한다.

III. 실험

HMD 와 깊이 카메라는 각각 HTC VIVE PRO 와 키넥트 V2 를 활용했다. 깊이 카메라에서 나타나는 손의 위치는 SDK 에서 제공되는 관절의 위치를 활용했다. HMD 와 카메라의 좌표계를 일치시키기 위해 t=100 개의 점을 위치 좌표를 이용해 변환행렬을 획득했다. 입력된 좌표계의 평균 오차는 1.78cm 로 나타났다.

손 상태는 그림 2 와 같이 총 4 개(s=4)의 상태를 정의하고 인식에 활용했다. 가상객체를 쥐고 펴거나 집거나 버튼을 누르는 행위를 고려해 상태를 정의했다. 데이터 수집은 깊이 카메라를 통해 검출된 손의 위치를 중심으로 가로, 세로, 깊이를 각각 15cm 이내인 픽셀들을 획득했고, 36x36 (w x h) 픽셀 사이즈로 이미지를 변환했다. 각 상태별로 500 개의 데이터를 수집했고, 400 개는 학습 데이터, 100 개는 테스트 데이터로 활용했다. 그림 3 은 손 상태 인식의 혼동행렬(Confusion Matrix)로 평균 95%의 인식률을 나타냈다. 카메라로부터의 손의 위치와 회전의 변화에도 손의 움직임은 대부분 유사한 형태로 안정적인 인식률을 보였으나 그 외 선택, 집기, 팔 동작들은 형태의 변화가 나타나 다른 손 동작으로 인식하는 결과를 보여주었다.

제안하는 방법을 적용하여 사용자의 손 움직임과 모양에 따라 가상 손을 표현할 수 있었고, 가상 객체를 잡고, 놓고, 버튼을 누르는 동작을 컨트롤러 없이 손을 이용해 수행할 수 있었다.

IV. 결론

본 논문에서는 HMD 와 깊이 카메라를 이용하여 가상환경에서의 사용자 친화적인 인터랙션 방법을 제시하였다. 칼리브레이션을 통해 각 기기들의 로컬 좌표계를 하나로 통합하고, 사용자의 손 움직임을 반영했다. 또한, CNN 기반의 모델을 활용하여 손 상태를

	쥔	선택	집기	펴
쥔	1.00	0	0	0
선택	0	0.92	0.08	0
집기	0	0.07	0.90	0.03
펴	0	0	0.02	0.98

그림 3. 손 상태 인식 혼동행렬 (행:예측, 열:정답)

인식하고 실제 손을 활용한 인터랙션을 가능하게 했다. 제안하는 방법은 손의 인터랙션을 대체했지만 무릎이나 발과 같은 주요관절을 추가적으로 활용한다면 캐릭터 리깅(rigging)을 통해 사용자의 동작으로 캐릭터의 전신 움직임을 표현할 수 있을 것으로 예상된다. 또한, 단일 카메라를 다중 카메라로 확장한다면 관절의 가려짐을 최소화하고 추적 성능을 높이며 실감 인터랙션 체험을 지원할 수 있을 것으로 예상된다.

ACKNOWLEDGMENT

본 연구는 한국전자통신연구원 주요사업의 일환으로 수행되었음. [20ZH1200, 초실감 입체공간 미디어·콘텐츠 원천기술 연구]

참 고 문 헌

- [1] 홍성진 외 5 명, 청소년 체육교육을 위한 가상증강현실 스포츠 통합플랫폼 기술 개발, 정보과학회지, Nov. 2018.
- [2] 정구문 외 4 명, VR HMD 기반 모의 비행 훈련 장치의 조종사 훈련 효과 평가, 한국항공운항학회지, Dec. 2018.
- [3] Daydream Labs: Teaching Skill in VR, (<https://www.blog.google/products/daydream/daydream-labs-teaching-skills-vr/>, accessed 29 Jun. 2020).
- [4] HTC Vive Pro, (<https://www.vive.com/kr/product/vive-pro-full-kit/>, accessed 29 Jun. 2020).
- [5] Microsoft Kinect, (<https://developer.microsoft.com/en-us/windows/kinect>, accessed 29 Jun. 2020).
- [6] Leap Motion (<https://developer.leapmotion.com/>, accessed 29. Jun. 2020).
- [7] Yea Som Lee and Bong-Soo Sohn "Immersive Gesture Interfaces for Navigation of 3D Maps in HMD-based Mobile Virtual Environments," Mobile Information Systems, Vol.2018, 2018.
- [8] Joon Young Ahn, et al. "A VR/AR Interface Design based on Unaligned Hand Position and Gaze Direction," IEIE Transactions on Smart Processing and Computing, Vol. 9, No. 2, April 2020.
- [9] Xiao Yan Wu "A hand gesture recognition algorithm based on DC-CNN, " Multimedia Tools and Applications, Vol.79, pp.9193-9205, 2020.