

입체 영상 내 사용자 인터랙션 활용을 위한 사용자 손 제스처 인식 기술 연구 동향

송복득, 최홍규, 김성훈
한국전자통신연구원

bdsong@etri.re.kr, hk-choi@etri.re.kr, steve-kim@etri.re.kr

Research Trends of User Hand Gesture Recognition Technologies for Utilizing User Interaction in Stereoscopic Images

Bok Deuk Song, HongKyw Choi, Sung-Hoon Kim

Intelligent Convergence Research Laboratory
Electronics and Telecommunications Research Institute(ETRI).

요 약

사용자 손 제스처 인식 기술은 영상 인식 분야에서 오랫동안 연구된 분야로 오늘날까지 카메라, 적외선 센서, 레이더 등 다양한 디바이스를 활용하여 연구되어 왔다. 최근에는 인공 지능 기술과 결합하여 자연스러운 사용자 손 제스처의 인식율을 높이는 방향으로 연구되고 있다. 본 논문은 사용자 선택형 UHD 입체 미디어 서비스내 사용자 인터랙션 인식을 위하여 다양한 디바이스 환경에서의 인공 지능 기술을 활용한 사용자 손 제스처 인식 기술에 대한 연구 동향을 간략히 소개한다.

I. 서 론

컴퓨팅 장치의 성능 발전과 컴퓨터와 인간과의 사용자 인터페이스 관련 다양한 센서가 개발되면서 컴퓨터를 사용할 때 마우스에서만 제어되던 사용자 인터페이스에서 사용자의 자연스러운 움직임을 감지하여 컴퓨터와의 인터페이스를 제공하는 NUI(Natural User Interface: NUI) 기술이 활발히 연구되고 있다.

다양한 NUI 기술 중에 사용자의 자연스러운 움직임을 제공하는 사용자의 손 제스처 인식 기술이 있다. 사용자 손 제스처 인식 기술은 사용자가 수행한 제스처를 입력 받을 수 있는 센서가 필요한데 센서 입력 방식이 사람의 신체를 접촉하는 방식과 비접촉으로 인식하는 기술이 있다. 각각의 방식에 따라 사용자 손 제스처 인식 범위 및 형태가 정해지고 사용자 제공 GUI(Graphic User Interface)도 다양하게 제공된다. 예를 들어 입체 영상에서 키넥트(Kinect)와 같은 3D 깊이 값을 제공하는 비접촉식 센서를 기반으로 3D 안경을 착용한 사용자가 가상 인터페이스를 이용하여 클릭, 슬라이드, 홀드, 드래그 인/줌 아웃 등의 손 제스처를 사용하여 3D TV를 감상하는 기술도 연구되고 있다[1].

그리고 인공 지능 기술이 활발히 연구되어 인공 지능 기술과 사용자 손 제스처 인식 기술이 결합되어 다양한 사용자 인터페이스를 제공하고 있다.

본 논문은 입체 영상 내 사용자 인터랙션에 활용할 수 있는 다양한 디바이스 환경 중 비접촉식 센서와 인공 지능 기술이 결합된 사용자 손 제스처 인식 기술에 대한 간략한 연구 동향을 소개한다.

II. 본론

본 논문에서는 사용자의 손 제스처 인식 기술에 많이 사용되고 있는 카메라와 립 모션을 이용하여 인공 지능 기술을 활용한 사용자 손 제스처 인식 기술에 대하여 소개한다. 그리고 최근에 비접촉식 센서로 연구되고 있는 레이더 센서를 활용한 사용자 손 제스처 인식 기술에 대해서 간략히 소개한다.

1) 카메라 기반 사용자 손 제스처 인식 기술

오래전부터 단일 카메라로부터 사용자 손 제스처 인식에 대한 연구는 활발히 진행되고 있었다. 대표적으로 영상 인식 알고리즘을 활용한 수화 인식 기술과 로봇제어를 위한 손 제스처 인식, 가상 현실 게임을 위한 손 제스처 인식 등이 연구되어 왔다. 그리고 인공 지능 기술의 발전으로 손 제스처 인식율이 높아지고 있으며 활발히 연구되고 있다. 예를 들어 MANO 모델을 활용하여 손 모델의 파라미터를 추정하여 3 차원 손 모양 템플릿에 적용시키는 연구를 수행하여 여러 개의 보조 네트워크를 활용하여 실세계의 3 차원 손 모양 데이터를 적용할 수 있는 연구가 수행되고 있다[2].

그림 1 은 tensorflow.js 를 기반으로 웹 환경에서 딥러닝 기반의 손 제스처 인식 예제[3][4]를 구현한 결과 화면으로 카메라는 실내의 경우 외부 조명 및 환경에 영향을 받지 않고 근거리에서 사용자의 손 제스처 인식이 높은 것을 확인할 수 있다. 이러한 점은 입체 영상과 연동 시 고려해서 사용할 수 있을 것으로 예상된다.



<그림 1> 카메라를 이용한 딥 러닝 기반의 손 제스처 인식 화면

2) 립 모션기반 사용자 손 제스처 인식 기술

립 모션은 비접촉식으로 사용자 손 제스처 인식에 특화된 장비로 열 개 손가락 추적 및 인식 기능을 제공한다. 특히 공간 제약을 받지 않고 가격이 저렴한 장점을 제공하며 가상 현실 게임, 원격 제어 분야 등에 활발히 적용되고 있다. 최근에 딥 러닝 기술을 활용하여 립 모션에서 추출되는 손 좌표 값을 활용하여 학습한 데이터를 이용해 만들어진 학습 모델을 활용하여 실시간 사용자 손 제스처 인식 기술이 연구되고 있다. 예를 들어 립 모션을 이용하여 3 차원 손 좌표 정보를 검출하고 각 정보를 합쳐 2 차원의 RGB 이미지로 생성한 후 만들어진 이미지를 CNN 모델 중 하나인 SSD 모델을 통해 학습시킴으로써 손 제스처를 분류하고 인식하는 기술[5]을 연구하거나 딥 러닝의 RNN(Recurrent Neural Networks)을 활용하여 Shape Retrieval Counter(SHREC) 데이터 세트를 사용하여 수화 인식 분야에 적용되고 있다[6]. 립 모션은 카메라와 달리 외부 조명 환경의 제약이 적은 장점을 제공하며 딥 러닝 기술을 접목하여 사용자 손 제스처 인식을 향상해 기여하고 있는 점을 고려해서 카메라와 같이 입체 영상 내 사용자 인터랙션 인식에 사용할 수 있을 것으로 예상된다.

3) 레이더(Radar) 센서를 활용한 사용자 손 제스처 인식 기술

사용자 손 제스처를 인식하기 위해서는 디바이스 특성에 대한 환경 요소 및 손 제스처의 특성에 영향을 미치는 요소들을 고려해야 한다. 이러한 다양한 요소에서 안정적인 특징을 추출하기 위해서 효과적인 사용자 손 제스처 인식을 위하여 레이더(Radar) 센서를 활용한 사용자 손 제스처 인식 기술 연구되고 있다.

레이더 센서는 손에 특별한 장치를 장착하지 않고 전파를 이용하여 사용자 손 제스처를 인식하는 기술이다. 구글의 솔리(Soli)의 경우 밀리파 레이더로 떨어진 물체 움직임을 인식하는 기술로 기기의 화면을 직접 터치하지 않고도 손짓으로 디바이스의 기능 및 화면을 제어하는 기술이다. 그리고 물체가 움직이는 경우에 비트 주파수는 도플러 주파수의 함수로서 표현할 FMCW 레이더를 활용하여 손 제스처를 인식하는 기술[7]이 있다. 전파의 특성상 레이더 센서의 경우 인신 거리 및 설치 환경을 고려하여 사용할 수 있을 것으로 예상된다.

III. 결론

디지털 콘텐츠 산업의 발전과 모션 센서 보급화에 따라 사용자와 미디어 사이의 인터랙션을 통한 몰입형 콘텐츠 시장이 발전하고 있다. 그리고 인공지능 기술 발전을 통해 사용자의 자연스러운 제스처 인식이 높아지면서

사용자 편의 인터페이스가 제공되어 다양한 산업에서 활용하고 있다. 특히 사람 몸에 센서를 부착하지 않고 사람의 손 제스처 인식 기술은 입체 영상과 연동하여 기존 2 차원적인 영상 보다 몰입감을 증대시킬 것으로 예상된다.

본 논문에서는 사용자 선택형 UHD 입체 미디어 서비스의 사용자 인터랙션 인식 기술 적용을 위한 비접촉식 사용자 손 제스처 인식 기술에 대한 동향을 간략히 소개하고 있다. 이 연구 동향을 바탕으로 기술의 장단점을 분석하여 입체 영상 내 사용자 인터랙션 인식 개발에 사용자 선택형 UHD 입체 미디어 서비스 내 사용자 인터랙션 인식 기술에 적용 및 확장 개발이 가능하도록 연구를 수행할 예정이다.

ACKNOWLEDGMENT

This work was supported by Electronics and Telecommunications Research Institute (ETRI) grant funded by the Korean government. [21ZH1200, The research of the basic media • contents technologies]

참 고 문 헌

- [1] Shun Zhang, Shizhou Zhang, "A Novel Human-3DTV Interaction System Based on Free Hand Gestures and a Touch-Based Virtual Interface", IEEE Access 7: 165961-165973 (2019)
- [2] Yuxiao Zhou, Marc Habermann, Weipeng Xu, Ikhsanul Habibie, Christian Theobalt, Feng Xu, "Monocular Real-Time Hand Shape and Motion Capture Using Multi-Modal Data", Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, pp. 5346-5355.
- [3] <https://github.com/andypotato/fingerpose>
- [4] <https://github.com/dredwardhyde/tensorflow-js-example>
- [5] Gyu-Min Kim, Joong-Hwan Baek, "Real-Time Hand Gesture Recognition Based on Deep Learning", Journal of Korea Multimedia Society Vol. 22, No. 4, April 2019(pp. 424-431)
- [6] Avola, D., et al., Exploiting recurrent neural networks and leap motion controller for the recognition of sign language and semaphoric hand gestures. IEEE Transactions on Multimedia, 2018. 21(1): p. 234-245.
- [7] Yong Wang, Aihu Ren, Mu Zhou, Wen Wang, Xiaobo Yang, "A Novel Detection and Recognition Method for Continuous Hand Gesture Using FMCW Radar", IEEE Access 8: 167264-167275, 2020