

공항 내 시설 안내 서비스를 위한 마커리스 한국 수어 인식 기술

박한무*, 정윤영

Korea Electronics Technology Institute(KETI)

*hanmu@keti.re.kr, yoonyj@keti.re.kr

Keypoint-based Korean sign language recognition for guiding airport facilities

*Han-Mu Park, Yoon Young Jeong

Korea Electronics Technology Institute (KETI)

요 약

공항은 다양한 시설들이 혼재되어 있는 복합 공간으로 원하는 장소로 이동하기 위해서는 지도나 표지판뿐만 아니라 유·무인 안내 서비스를 이용하는 것이 필수적이다. 그러나 이러한 시설 안내 서비스는 이용객의 대다수를 차지하는 비장애인을 기준으로 설계되어 있는 것이 대부분이며, 특히 청각장애인은 음성 언어인 한국어를 통한 의사소통이 어려워 이러한 안내 서비스의 사각 지대에 놓여있다고 볼 수 있다. 본 논문에서는 청각장애인을 대상으로 공항 내 시설 위치 안내 서비스를 제공하는 키오스크를 구현하기 위해 마커리스(markerless) 한국 수어 인식 기술을 적용하는 방법을 제안한다. 기존의 알고리즘들이 수어를 인식하기 위해 사용자의 신체에 마커(marker)를 부착하거나 사용에 제약이 있는 깊이, 자기장 센서 등을 활용한 데 반해, 특징점 기반의 인식 기술은 별도의 마커 없이 RGB 영상만을 활용하여 수어를 인식하기 때문에 사용자의 편의성을 극대화할 수 있는 방법이다. 영상 기반의 마커리스 수어 인식 기술을 위치 안내 키오스크에 적용함으로써 청각장애인이 다른 사람의 도움 없이 공항 시설을 쉽고 안전하게 이용할 수 있도록 지원할 수 있으며, 나아가 배리어프리 (barrier-free) 공항을 구현하는 데에 큰 도움이 될 수 있을 것으로 기대된다.

1. 서론

공항은 항공기에 탑승하기 위한 시설 외에도 음식점, 편의점, 의무실, 라운지 등 다양한 시설들이 혼재되어 있는 복합 시설이다. 공항과 같이 복잡한 공간에서 원하는 장소로 이동하기 위해서는 지도나 표지판뿐만 아니라 유·무인 안내 서비스의 도움을 받는 것이 필수적인데, 이러한 안내 서비스는 비장애인을 기준으로 설계되어 있기 때문에 장애인들의 이용이 어려운 상황이다[1]. 특히, 한국 수어는 한국어와 문법 체계가 상이한 별개의 언어로, 농인들은 한국어로 필답을 하는 것에 익숙하지 않아 문자 기반 안내 서비스에 대한 만족도가 높지 않았다. 수어 통역사를 통한 안내 서비스를 제공하는 항공사도 있으나, 간단한 용무임에도 통역사를 대동해야 하고 통역사의 일정에 따라 대기 시간이 존재하는 등, 청각장애인의 이동권 증진 측면에서 스스로 원하는 장소로 이동할 수 있는 데에 도움을 줄 수 있는 안내 서비스의 개발이 필요한 상황이다.

이러한 수어 통역사의 역할을 일부 분담하고 청각장애인의 기본권을 증진하기 위한 목적으로 다양한 수어 번역 기술들이 연구되었다. 기존의 연구 방향은 화자의 동작을 인식하기 위한 마커(marker)의 사용 여부에 따라 크게 두 가지로 구분할 수 있다. 첫번째 방식은 화자의 신체에 마커를 부착[2]하거나 센서가 장착된 장갑[3,4] 등을 활용하는 방식이다. 이러한 방식은 화자의 몸동작을 매우 정확하게 인식할 수 있지만, 인식을 위해 마커를 부착하거나 장비를 착용해야 하기 때문에 공항과 같은 장소에서 활용하기에는 어려움이 있다. 두 번째 방식은 마커를 사용하지



알고

그림 1 마커리스 한국 수어 인식 기술을 적용한
공항 내 시설 안내 키오스크

화자의 자세를 추정하여 인식하는 마커리스(markerless) 방식이다. 마커리스 방식은 수어 인식을 위해 화자가 부착하거나 착용해야 하는 장비가 없기 때문에 편의성이 극대화되는 방식이지만, 상대적으로 자세 추정 정확도가 낮기 때문에 마커 기반 방식에 비해 인식 성능이 좋지 않았다. 이러한 자세 추정 정확도 측면에서의 단점을 보완하기 위해 Microsoft의 Kinect 나 Ultraleap의 Leap motion 과 같이 적외선 영상 기반의 능동 깊이 센서를 활용한 방식들이 제안[5,6,7,8]되었으며, 최근에는 인공지능 기술의 발전과 함께 RGB 영상만으로도 높은 자세 추정 정확도를 보이는 Openpose[9,10]와 같은 알고리즘을 활용한 딥러닝 기반의 인식 기술[11]이 제안되었다. 적외선 센서 기반의 인식 기술은 실외에서는 사용이 어려운 반면, RGB 영상 기반의 인식 방식은 이러한 제약이 없어 범용성이 매우 높은 방식이다.

본 논문에서는 청각장애인이 공항의 시설들을 다른 사람의 도움 없이 스스로 편리하게 이용할 수 있도록 지원하기 위한 공항 내 시설 위치 안내 서비스를 제안한다. 제안하는 서비스는 영상 기반 마커리스 수어 인식 기술을 그림 1과 같은 키오스크에 적용하여 편의성이 극대화된 방식으로 사용자의 의도를 파악하고, 목적지의 위치를 약도에 표기하여 원하는 장소로 안내하는 방식으로 구현되었다. 본 논문의 2 장에서는 키오스크에 적용하기 위한 영상 기반 마커리스 한국 수어 인식 기술에 대하여 설명하고, 이를 학습하기 위한 데이터셋 구축 과정에 대해 3 장에서 설명한다. 4 장에서는 영상 기반 수어 인식 기술이 적용된 안내 서비스에 대해 설명하고, 마지막으로 향후 연구 방향을 제시한다.

2. 영상 기반 마커리스 한국 수어 인식 기술

본 논문에서 공항 내 시설 안내 서비스를 위해 적용하고자 하는 영상 기반 한국 수어 인식 모델[11]은 그림 2와 같이 인코더 파트와 디코더 파트로 구성되어 있다. 인코더 파트에서는 Openpose[9,10] 알고리즘을 활용하여 상반신과 손가락, 얼굴 부위의 특징점들을 매 프레임 추출한 후, 각 특징점들의 좌표를 Gated recurrent units (GRUs)[12]에 입력 특징값으로 하여 컨텍스트 벡터(context vector)를 생성한다. 디코더 파트에서는 인코더 파트에서 생성된 컨텍스트 벡터를 입력으로 받아 한국어 문장을 단어 단위로 생성한다. 이 때 입력 시퀀스가 길어짐에 따라 발생 가능한 정보 손실을 최소화 하기 위하여 어텐션 매커니즘(attention mechanism)[13]을 적용하여, 입력 수어 영상의 길이 변화에 강건한 방식을 사용하였다.

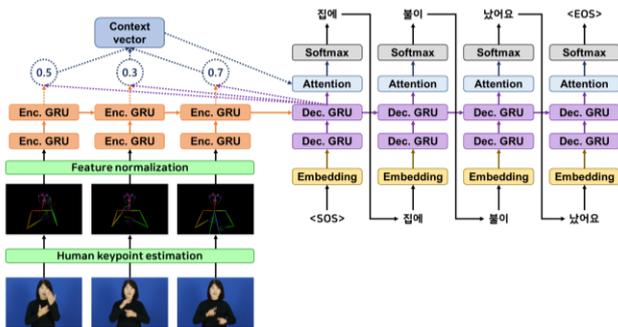


그림 2 영상 기반 마커리스 수어 인식 모델 구조도[11]

3. 공항 내 시설 안내 서비스를 위한 한국 수어 데이터셋

공항 내 시설 안내 서비스의 핵심이 되는 수어 인식 모델을 학습하기 위해서는 목적에 맞게 설계된 학습용 데이터셋이 구축되어야 한다. 이를 위해 본 연구에서는 한국공항공사와 협업하여 그림 4와 같이 공항 내 시설 안내를 위한 질의응답 문장셋을 선정하였으며(질문용 140 문장, 응답용 24 문장), 선정된 문장셋을 바탕으로 전문 수어 언어 제공자 20 명을 섭외하여 데이터셋을 구축하였다. 수어 영상 획득 시 언어 제공자의 자세를 정확하게 추출하기 위해 그림 3과 같이 배경을 제거하였으며, 촬영 각도에 따른 영향을 학습할 수 있도록 20 도 간격으로 세 방향에(정면, 좌측, 우측) 카메라를 배치하여 데이터를 취득하였다. 이러한 방식으로 약 140 시간 분량의 수어 영상 데이터셋을

구축하였으며, 구축된 데이터를 모두 활용하여 한국 수어 인식 모델을 학습하였다.



그림 3 한국 수어 데이터 예시

순번	구분	내용	순번	수어 표현
1	관행시절	편의점은 어디에 있습니까?	A 1	편의점 어디 곳
2	관행시절	여기 휴게실은 어디에 있습니까?	A 1	여기 어디 곳
3	관행시절	여기 화장실은 어디에 있습니까?	A 1	화장실 어디 곳
4	관행시절	여기 출구는 어디에 있습니까?	A 1	출구 어디 곳
5	관행시절	여기 손화장품 어디에 있습니까?	A 1	손화장품 어디 곳
6	관행시절	여기 식당은 어디에 있습니까?	A 1	식당 어디 곳
7	관행시절	여기 공항 내 손화장품 어디에 있습니까?	A 1	공항 내 손화장품 어디 곳
8	관행시절	여기 공항 내 샤워실은 어디에 있습니까?	A 1	공항 내 샤워실 어디 곳
9	관행시절	여기 기동차들 어디에 있습니까?	A 1	기동차 어디 곳
10	관행시절	여기 출구문을 열려면 손화장품 어디에 있습니까?	A 1	출구문 열려면 어디 곳
11	관행시절	여기 안내선 일목종의 어디에 있습니까?	A 2	안내선 일목종 어디 곳
12	관행시절	여기 국제선 일목종의 어디에 있습니까?	A 1	국제선 일목종 어디 곳
13	관행시절	여기 국제선 출구문의 어디에 있습니까?	A 2	국제선 출구문 어디 곳

그림 4 한국 수어 데이터 문장셋 예시

4. 공항 내 시설 안내 키오스크

본 논문에서 제안하는 키오스크를 활용한 공항 내 시설 안내 프레임워크는 그림 5와 같이 서버와 클라이언트로 역할이 구분되어 있는데, 이는 여러 장소에 키오스크를 배치하여 접근성을 높이고 주된 연산은 중앙 서버에서 하는 방식으로 서비스 운영 비용을 절감하기 위한 설계이다. 제안하는 서비스 프레임워크에서 클라이언트는 키오스크를 의미하며, 상단에 배치된 RGB 카메라를 이용하여 사용자의 영상을 서버로 전달하고 인식 결과에 따른 안내 정보를 표출하는 역할을 한다. 다음으로 서버는 딥 러닝 모델을 직접 실행하여 추론하는 역할을 하며, 크게 자세를 추정하는 모델과 문장을 인식하는 모델이 순차적으로 실행되는 구조로 설계되었다.



그림 5 공항 내 시설 안내 서비스 프레임워크 구조도

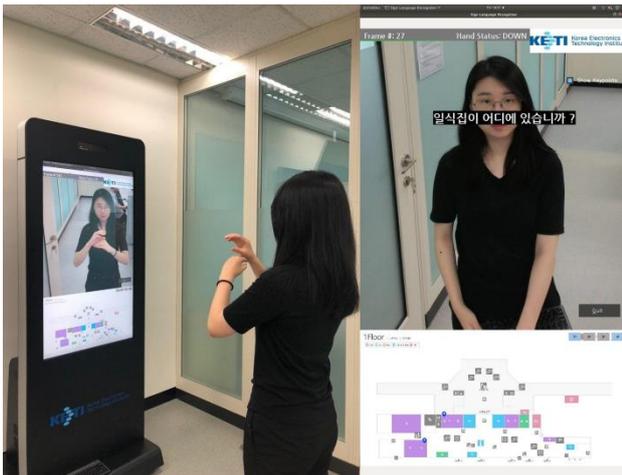


그림 6 공항 안내 키오스크 및 실행 화면

5. 결론

본 논문에서는 특징점 기반의 한국 수어 인식 기술을 키오스크에 적용하여 청각장애인을 위한 공항 내 시설 위치 안내 서비스를 구현하는 방법을 제안하였다. 제안하는 방법은 설치 공간에 제약이 적은 RGB 영상만을 이용한 방식이며, 신체에 별도의 마커를 부착하지 않아도 사용이 가능하여 편의성이 좋은 방식이다. 제안한 방식이 적용된 키오스크가 국내의 공항 내 여러 곳에 설치된다면 청각장애인의 이동권을 증진시키고, 나아가 배리어프리 공항을 구현하는 데에 많은 도움이 될 것으로 기대된다.

Acknowledgement

이 논문은 2020 년도 정부 (과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (2017-0-00255, 자율지능디지털동반자 프레임워크 및 응용 연구 개발)

참고문헌

[1] 국가인권위원회, “항공기 이용에 있어 장애인 차별 개선을 위한 정책 권고,” 정책권고, 15 직권 0001300, 2016.

[2] 조재현, 엄성용, “마커셋을 이용한 지문자 인식 기법,” 한국정보과학회 학술발표논문집, 26(1B), 671-673, 1999.

[3] S. A. Mehdi and Y. N. Khan, “Sign language recognition using sensor gloves,” ICONIP, 2002.

[4] T. T. Swee, A. K. Ariff, S.-H. Salleh, and S. K. Seng, “Wireless data gloves Malay sign language recognition system,” ICICSP, 2007.

[5] Z. Zafrulla, H. Brashear, T. Starner, H. Hamilton, and P.

Presti, “American sign language recognition with the kinect,” ICMI, 2011.

[6] X. Chai, G. Li, Y. Lin, Z. Xu, Y. Tang, and X. Chen, “Sign Language Recognition and Translation with Kinect,” CS, 2013.

[7] M. Mohandes, S. Aliyu, and M. Deriche, “Arabic sign language recognition using the leap motion controller,” ISIE, 2014.

[8] C. H. Chuan, E. Regina, and C. Guardino, “American Sign Language Recognition Using Leap Motion Sensor,” ICMLA, 2014.

[9] Z. Cao, G. H. Martinez, T. Simon, S. Wei, and Y. A. Sheikh, “OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields,” IEEE TPAMI, 2019.

[10] T. Simon, H. Joo, I. Matthews, and Y. A. Sheikh, “Hand Keypoint Detection in Single Images using Multiview Bootstrapping,” CVPR, 2017.

[11] S.-K. Ko, C. J. Kim, H. Jung, and C. Cho, “Neural Sign Language Translation Based on Human Keypoint Estimation,” AS, 2019.

[12] K. Cho, B. V. Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, “Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation,” EMNLP, 2014.

[13] D. Bahdanau, K. Cho, and Y. Bengio, “Neural Machine Translation by Jointly Learning to Align and Translate,” ICLR, 2015.