
저자 (Authors)	신기철, 이승훈, 권재영, 장현국, 이희연 Shin Kichul, Lee Seunghoon, Kwon Jaeyoung, Jang Hyunkook, Lee Heeyoun
출처 (Source)	한국HCI학회 학술대회 , 2019.2, 645-649(5 pages)
발행처 (Publisher)	한국HCI학회 The HCI Society of Korea
URL	http://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE08008190
APA Style	신기철, 이승훈, 권재영, 장현국, 이희연 (2019). 영상인식과 립모션을 활용한 한국수어 번역. 한국HCI학회 학술대회, 645-649
이용정보 (Accessed)	동신대학교 220.95.42.*** 2020/08/18 22:24 (KST)

저작권 안내

DBpia에서 제공되는 모든 저작물의 저작권은 원저작자에게 있으며, 누리미디어는 각 저작물의 내용을 보증하거나 책임을 지지 않습니다. 그리고 DBpia에서 제공되는 저작물은 DBpia와 구독계약을 체결한 기관소속 이용자 혹은 해당 저작물의 개별 구매자가 비영리적으로만 이용할 수 있습니다. 그러므로 이에 위반하여 DBpia에서 제공되는 저작물을 복제, 전송 등의 방법으로 무단 이용하는 경우 관련 법령에 따라 민, 형사상의 책임을 질 수 있습니다.

Copyright Information

Copyright of all literary works provided by DBpia belongs to the copyright holder(s) and Nurimedia does not guarantee contents of the literary work or assume responsibility for the same. In addition, the literary works provided by DBpia may only be used by the users affiliated to the institutions which executed a subscription agreement with DBpia or the individual purchasers of the literary work(s) for non-commercial purposes. Therefore, any person who illegally uses the literary works provided by DBpia by means of reproduction or transmission shall assume civil and criminal responsibility according to applicable laws and regulations.

영상인식과 립모션을 활용한 한국수어 번역 Interpretation of Sign Language Using Image Recognition & Leap Motion

신기철
Shin Kichul
건국대학교
Konkuk University
skchope71@naver.com

장현국
Jang Hyunkook
건국대학교
Konkuk University
hxjang6@naver.com

이승훈
Lee Seunghoon
건국대학교
Konkuk University
asia_a@cclab.konkuk.ac.kr

이희연
Lee Heeyoun
건국대학교
Konkuk University
heeyounlee2@gmail.com

권재영
Kwon Jaeyoung
건국대학교
Konkuk University
pandos_112@naver.com

요약문

이번 연구는 영상인식과 립모션을 활용하여 농인들의 언어인 수어를 한국어로 번역하여 출력하였습니다. 영상인식을 활용함으로써 사용자의 접근성을 높였고, 립모션을 사용하여 손동작 인식의 정확도를 높였습니다. 뿐만 아니라 기존의 한국어 형태소만을 번역하는 원론적인 번역이 아닌 실제로 농인들의 일상생활에서 자주 쓰이는 수화동작들을 추출하여 번역함으로써 실제 생활에서 사용할 수 있는 실용적인 번역 시스템을 구축하였습니다

주제어

영상인식, 립모션, Opencv, 수어인식, 수화동작

1. 서론

현재 우리나라의 듣지 못하는 농인 및 언어 장애인의 숫자를 합치면 총 30 만 명에 달한다고 한다 그리고 조기 청각장애와 미등록 언어 장애인을 합치면 그 수가 무려 35 만 명에 달할 것이라 보건복지부는 추산하고 있다

이들은 비율로 치면 인구의 1%가 되지 않아 사회의 관심을 받지 못하고 있지만, 이들에게 한국어 - 수어 번역기는 이들이 실생활을 살아갈 때 필요한 필수품이라 할 수 있다

국립국어원과 한국농아인협회의 조사에 따르면 농인들의 80%가 한국수어를 주요 의사소통 수단으로 사용하고 있고 나머지 10%도 손동작을 통해 본인의 의사를 전달한다고 답하여 사실상 대부분의 농인이 수화를 사용하여 본인의 의사 표현을 하고 있다고 볼 수 있다. 게다가 국어를 읽지 못하거나 제대로 이해하지 못하는 청인들이 대부분이기 때문에 글로써 의사표현을 하는 것도 농인들에게는 상대적으로 어려운 상황이라고 볼 수 있다 이런 상황에서 농인들이 의사소통이나 정보취득 면에서 다른 청인들과 비교하였을 때 뒤지지 않기 위해서는 수어 -> 한국어로 번역해 줄 수 있는 수화통역사가 필요한데 현재 우리나라는 수화통역사의 숫자가 매우 적다.

수화통역 센터의 조사에 따르면 수화통역사 1 인당 담당하는 농인의 수는 전국을 평균으로 했을 때 327 명이고, 광역시 기준으로는 인당 500 명에 달하는 농인들을 담당한다고 한다. 그렇기 때문에 대부분의 농인들은 질 좋은 통역 서비스를 제공받지 못해서 사회에서 정보취득이나 의사소통에서 문제가 발생하고 그에 따라 다른 사람들과 농인 들의 사회적 격차를 초래하고 있다. 따라서 정부는 이런 상황을 방치하지 않고 농인들의 생활편의를 위해 적극적으로 이 문제에 대처할 필요가 있다

실제로 법률 13978 호 한국수화언어 법에도 국가는 한국수어의 정보화를 통하여 지식과 정보를 활용할 수 있도록 각종 사업을 적극적으로 장려하고 시행하여야 한다고 적시되어 있다

다행히 최근 기술의 발전으로 수어 동작을 인식하여 이를 한국어로 출력하려는 다양한 시도들이 지속되고 있다. 키넥트를 활용한 수화 동작 인식, 컴퓨터 영상을 통한 수화 인식 등 다양한 시도들이 있었고 최근에 나온 립모션 컨트롤러(Leap Motion Controller)를 이용한 수화 동작 인식도 활발히 이루어지고 있다.

영상인식을 통한 수화 인식은 실생활에서 사용자들이 편하게 접근할 수 있는 장점이 있고, 수화 동작뿐만 아니라 얼굴 표정도 함께 인식할 수 있는 장점이 있어 립모션과 함께 최근에 가장 활발히 연구되고 있는 분야이다

본 논문에서는 영상인식과 립모션을 활용한 수어 동작 인식 방법과 기존 수어 인식 관련 논문을 소개하고 그것과 비교하여 본 연구와의 차별점과 분석 결과에 관해 설명할 것입니다

2. 관련 연구

최근 ICT 기술의 발전으로 인해 영상 인식 및 수화 동작 인식을 통한 수어 번역에 관한 연구가 활발히 이루어지고 있다.

기존 연구들을 살펴보면 국립국어원에 등재되어있는 한국수어의 단어 단위로 번역이 아닌 수어를 한국어와 같다고 생각되어 가장

기초 단위인 형태소만을 번역한 연구[1]
미국 수어(ASL)를 번역할 때도 단어가 아닌
가장 기초형태인 알파벳만을 번역한 연구[2]
가 있다

그런데 이는 수어에 대한 무지로 발생한 현상으로
수어는 한국어와 전혀 다르기 때문에 이는 한국어
-> 한국어 로 번역한 형태이고 형태소를 조합한다
하더라도 그게 단어가 되는 것이 아니기 때문에
이는 수어에 대한 특성을 전혀 파악하지 못한
잘못된 방향의 연구라고 볼 수 있다.

그리고 이뿐만 아니라 지금까지의 연구의 대부분
은 데이터를 받아 그것을 학습하여 동작을 스스로
인식하는 것이 아닌 제작자가 특정 동작들에 관해
값의 범위를 정해줌에 따라 실제 사용할 때 조금의
오차가 발생했을 경우 실효성이 떨어지는 특성이
있다. 그래서 손가락의 개수, 손 위치각도 및
손가락 좌표값등을 받아 분석하는 연구[1], 손의
앞과 뒤, 개수를 받아 수어를 구분하는 연구[3]
등이 있었다. 하지만 이는 제작자가 매번 스스로
동작을 실행함으로써 값을 찾아야 하고,
실험단계에서 정확한 결과가 나오지 않으면 이후
사용 단계에서 실용성이 낮다는 평가가 받을 수
있기 때문에 이를 해결하기 위해서는 데이터를
받아 딥러닝을 하여 학습시키는 방향으로 인식을
해야 한다.

이와 같은 문제를 해결하기 위해서 본
연구에서는 딥러닝을 통해 수화 기본 동작들을
학습시키고, 형태소가 아닌 수어의 단어 문장
단위로 번역하는 시스템에 대해 고찰하였다.

3. 연구 방법

본 연구는 수화 인식을 하기 위해 2 가지 방법을
사용하였다. Opencv 를 활용한 영상인식, 립모션
을 활용한 손동작 인식 이렇게 2 가지 방법을 사용
한 이유는 접근성이 좋고 손과 표정을 모두 인식할
수 있는 영상인식, 손가락 길이, 회전 방향 등의
데이터를 정확한 값으로 빠르게 받을 수 있는
립모션의 장점을 모두 활용하고 이를 통하여 기존
방법과 비교하였을 때 보다 효과적으로 수어를 번
역하고자 함이다.

립모션을 활용한 수어인식과정은 다음과 같다. 본
연구가 진행된 실험환경은 Ubuntu18.04
Desktop 에서 진행하였고, 시스템 전체 구성도는
아래 그림 1(하드웨어 구성도), 그림 2 (소프트웨어
구성도)와 같다. PC 와 유선으로 연결된 립모션을
통해 사람의 손동작 제스처를 인식한다. 이후
PC 로 프레임 데이터가 전송 및 수집되면서,
데이터를 병합시키는 과정을 거친다. 쌓인
데이터를 이용하여 학습 후, 학습률을 측정한다.



그림 1. 하드웨어 구성도



그림 2. 소프트웨어 구성도

영상인식을 통한 수어인식과정은 다음과 같다.
웹캠 or 영상을 통한 데이터 인식 -> CNN 을 통한
수화 동작 분류 학습 -> 수화 동작 인식의 단계를
통해 수화동작을 학습하는 과정으로 진행된다.

그 중 핵심은 CNN 을 통한 수화동작 분류 학습
과정인데 이는 위와 같은 구성도 (그림 2. 소프트
웨어 구성도)로 이루어져 있다.

수어에서 공통으로 나타나는 기본 수형 동작의
수집 및 추출 -> 데이터 전처리 -> CNN 을 통한
수화동작 학습 -> 수화동작 인식 이렇게 구분할
수 있다. 이 과정에서 가장 중요한 것은 수어에서
공통으로 나오는 수형동작을 선정하여 추출하는
과정이다. 가장 자주 쓰이는 수어를 선정하는
기준과 수어의 표준동작이 있는 데이터베이스를
찾는 과정이 필요한데 한국에는 국립국어원이
있어 그곳을 통해 수어의 표준 동작과 수어중에
어떤 동작이 가장 많이 쓰이는지를 확인할 수 있다.

수형	해당 단어
	27(가슴터치) 30(세로이동) 33(손목회전) 21,38(한손)
	1,2,4,5,6,7,8,9,11,13,24, 25,32,36,41,42,49
	8,14,15,28,31(양손) 17,18,23,46,47(오른손)
	12,26(양손) 19,22,34(오른손)
	29,48(양손) 4,10,16,20,35,42,43(오른손)
해당사항없음	3,37,39,40,44,45,50

표 1. 수형동작 - 단어 매칭 정렬

본 연구에서는 국립국어원에 나오는 2 만 개가
넘는 수어에서 나오는 모든 수형 동작들을
분석하는 것은 어렵다고 생각되어 수어 중에
일상생활에서 가장 많이 쓰이는 단어 50 개를

선정하여 그중에서 공통으로 나오는 수형동작을 추출하고 그것들의 인식을 1 차 목표로 설정하였습니다.

위의 단어 50 개는 국립국어원에서 조회 수와 사용빈도를 기준으로 가장 빈번하게 사용된 단어들을 추출한 것입니다. 위의 표를 보면 50 개의 단어에서 공통으로 나오는 수형동작은 양손의 엄지와 검지를 펴다, 손바닥을 펴다, 주먹을 쥐다, 검지를 펴다, 동그라미를 만들다 이렇게 5 가지로 나타낼 수 있다. 그리고 양손과 한 손을 구분하면 총 10 가지의 손동작이 공통적으로 포함된다고 볼 수 있다. 그런데 수어에는 수형(손의 형태) 뿐만 아니라 수위(손의위치),수동(손움직임)이라는 요소가 있다.

수위	위치
기본동작	위치와 무관
얼굴상단	뺨,이마
얼굴하단	턱,입
왼손	손등,주먹,팔

표 2. 수위 - 몸의 위치 배치 정렬

수동(손움직임)은 모션이기 때문에 가로,세로, 회전 3 가지 경우의 수가 나왔지만 인식의 편의성을 위해 가장 나중 순서로 두고 수형과 수위를 합쳐 수어의 기본동작을 만들었다.

앞에서 추출한 수형 동작을 순서대로 A,B,C,D,E 로 설정하고 손의 위치(수위)를 1,2,3,4 숫자로 표현하였다. 그리하여 A1 = 엄지와 검지를 펴다 + 기본동작, A2 = 엄지와 검지를 펴다 + 얼굴 상단 이렇게 손동작과 위치를 합쳐 기초 동작을 생성하였다. 이는 수어의 공통동작을 추출한 결과이기 때문에 이후 모션을 넣게 되면 그것이 수어가 되어 일상생활에서 쓰이는 수어를 인식할 수 있게 된다. 그렇게 하여 [A1 ~ A4], [B1 ~ B4],[C1 ~ C4], [D1 ~ D4], [E1 ~ E4] 총 20 개의 기초수형동작의 인식을 1 차목표로 잡고 연구를 하였다. 이 후 기초수형동작 데이터를 수집한 다음 배경차분과 CNN 을 활용하여 수화동작을 인식한다. 배경차분에서는 Running Average 를 사용하여 현재 프레임과 배경모델 프레임을 계산한 다음 그차이로부터 사용자가 원하는 전경객체를 구하는 방식이다. 그 차이 이미지를 임계값으로 설정한 뒤 Thresholding 을 하면 구하고자 하는 전경 객체는 하얀색으로 표시되고 그 외의 배경이미지는 검은색으로 표현되어 구하고자 하는 전경 객체를 인식할 수 있게 된다. 이렇게 배경차분을 통해 기초수형동작 데이터를 얻은 후에 CNN 분석을 위해 데이터를 전처리하는 과정이 필요하다. 그 이유는 수집한 수형동작의 픽셀크기가 너무 크면 처리해야 할 데이터의 양이 너무 많아 학습시키기에 적절하지 않을 수 있기에 모든 사진을 89x100 픽셀 사이즈로 크기를 전처리한다.

그렇게 한 다음 이미지 학습을 위해 이를 CNN 을 통해 학습시키는 과정이 필요하다.

CNN 네트워크는 7 개의 hidden convolution layer 가 있고 활성화 기능으로 Relu 와 1 개의 완전 연결된 레이어로 구성되어 있다. 그리고 이 네트워크는 배치크기 64 로 50 회 반복하여 학습시킵니다. 반복횟수가 50 회인 이유는 그 이상 반복하여 학습시켜도 유효성 검사 정확도가 증가하지 않기 때문에 그렇게 횟수가 정해져 있다. 이 CNN 모델은 트레이닝 세트 1000 개, 테스트 세트 100 개로 이루어져 있으며 유효성 검사 정확도에서 96.6%의 정확도를 기록했습니다. 그렇게 하여 기초수형동작의 인식결과는 다음 그림 3 ~ 그림 7 와 같이 나타났습니다.



그림 3. 주먹(Fist)손동작 인식



그림 4. 손바닥(Palm)손동작 인식



그림 5. Swing 손동작 인식



그림 6. 손가락 1 개 손동작 인식



그림 7. 손가락 2 개 손동작 인식

위에서 배경차분과 CNN 학습모델을 사용하여 영상인식을 통해 수어동작을 인식하였다면, 이번에는 립모션을 활용하여 손바닥 회전, 각도, 위치 등의 데이터 수집을 활용하여 딥러닝 학습을 통해 수어 동작을 인식해보았습니다.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	-0.2751	-0.1743	0.1028	-49	170	-12	-39	154	-107	-43	144	-106	18	136	-110	44	112	-88
2	-0.2897	-0.1701	0.0999	-49	170	-11	-40	155	-104	-12	145	-108	18	136	-112	47	110	-84
3	-0.2646	-0.1637	0.0905	-49	170	-39	-41	156	-105	-12	145	-109	19	136	-114	50	102	-85
4	-0.2854	-0.1600	0.0916	-19	170	-29	-42	157	-106	-12	144	-100	20	136	-115	52	101	-86
5	-0.2589	-0.1648	0.0837	-19	170	-28	-42	158	-107	-11	144	-102	22	136	-117	54	101	-87
6	-0.255	-0.1699	0.0833	-11	177	-27	-41	159	-108	-11	143	-103	23	136	-119	57	101	-87

그림 8. 데이터 수집

데이터 수집은 손바닥의 회전 각도 및 위치를 알기 위해서 pitch, yaw, roll (x 축, y 축, z 축) 을 수집하고, 추가로 각 손가락의 위치를 알기 위해서 position (x, y, z) 를 수집하였다

```
Epoch 98/100
311/311 [=====] - 7s 22ms/step - loss: 0.0933 - acc: 0.9614
Epoch 99/100
311/311 [=====] - 7s 21ms/step - loss: 0.0926 - acc: 0.9614
Epoch 100/100
311/311 [=====] - 7s 21ms/step - loss: 0.0929 - acc: 0.9614
12.3096678495 Minutes of Execution
Model Saved
acc: 91.03%
```

그림 9. 데이터 학습

학습은 RNN-LSTM 기법을 이용하여 데이터를 학습시키고, 이를 모델로 변환하는 작업을 진행하였다

```
Model Loaded
Hit Enter to quit!
Connected
Hand Detected
Record now
Try Again
Ready
Hand Detected
Record now
Swipe Right
Ready
Hand Detected
Record now
Swipe Left
```

그림 10. 제스처 예측

예측 과정은 손동작 데이터를 수집 및 학습을 통해 만들어진 모델을 불러와서 학습된 모델을 현재 들어오는 프레임 데이터와 비교하는 작업을 진행하였다.

4. 결과 및 향후 연구 방향

본 연구에서는 농인들의 정보취득과 의사소통에서의 문제점을 해결하고 사회적 격차를 해소하고자 수어를 청인이 들을 수 있는 한국어로 번역하고자 하였다. 영상인식을 통한 수어인식 같은 경우 배경차분과 CNN 을 통해 수형동작을 학습시켜 96.6%의 정확도로 기초수형동작을 인식하여 이후 수동(움직임)을 더하여 온전한 수어를 번역할 수 있는 의의를 가졌다.

립모션을 통한 수어인식의 경우 기존 연구에서 사용된 데이터의 좌표값을 이용하여 정해진 동작만을 인식하는 것이 아니라 사용자 스스로 새로운 동작을 학습시킴으로써 동작 개수의 한계점을 줄였다.

영상을 통한 수어인식은 현재 농인들의 통역을 도와주는 손말이음센터와 같이 표정과 수화를 같이 볼 수 있기 때문에 농인들의 감정표현이나 의사를 더 잘 반영하여 대화를 할 수 있을 뿐더러 스마트폰을 통해 언제 어디서든 소통이 가능하다는 접근성과 편의성이 있기 때문에 농인들이 생활에서 필요할 때 언제든지 사용할 수 있다는 장점이 있다. 립모션을 통한 수어인식은 립모션이 손가락 움직임을 0.01mm 까지 추적하는 고정밀도와 초당 200 프레임의 데이터를 받는 반응속도가 있기 때문에 이런 장점을 활용하여 수어동작인식의 정확도를 높이고 반응속도를 높일 수 있어 농인들이 집에 있거나 특정 장소에 있을 때 조금 더 빠르고 높은 정확도로 의사를 전달할 수 있다. 본 연구에서는 이 2 가지를 통해 수어동작을 인식함으로써 장소와 시간에 구애받지 않고 수어통역을 받을 수 있는 편의성과 함께 기존의 방법보다 보다 빠르고 정확도 높은 수어번역을 할 수 있게 되었다.

향후 연구에서는 수어의 수형 동작 데이터를 더 많이 수집하여 이후 일상생활에서 쓰이는 수어들을 전부 다 해석할 수 있는 데이터베이스를 구축하는 것이 중요하다. 그리고 수어에는 조사와 같이 문장을 매끄럽게 해주는 요소가 없기 때문에 해석할 때 그런 부분을 어떻게 해결할지에 관한 깊은 고찰이 더 필요하다. 또한 현재의 시스템은 PC 와 유선 연결을 통해서 진행하기 때문에 이동성이 부족하다. 더 많은 사람이 편리한 형태로 사용하기 위해서는 모바일 환경에서 활용할 수 있는 방안이 필요하다.

사사의 글

이 논문은 2018 년도 산업통상자원부의
‘창의산업융합 특성화 인재양성사업’의 지원을
받아 연구되었음. (과제번호: N0000717)

참고문헌

- [1] 2016 년도 대한전자공학회 정기총회 및 추계
학술대회, 2016.11, 772-775(4 pages)
- [2] 양희덕, 이성환 (2011). 연속된 수화 인식을
위한 자동화된 수화 및 지화 적출.
정보과학회논문지 : 소프트웨어 및 응용, 38(2),
102-107
- [3] 이부형, 송필재 (2006). 수화소 분석을 통한
손동작 움직임 표현방법. 멀티미디어학회논문지,
9(8), 1067-1075
- [4] Danilo Avola, Marco Bernardi, Luigi
Cinque, Gian Luca Foresti, Cristiano
Massaroni (2018). Exploiting Recurrent Neural
Networks and Leap Motion Controller for Sign
Language and Semaphore Gesture
Recognition. arXiv:1803.10435v1