

# 키넥트 환경에서 데이터 마이닝을 이용한 수화 번역기

이상준\*, 우태호\*, 김지아\*, 박선영\*, 이수원\*, 김계영\*

\*송실대학교 컴퓨터학부

e-mail:{donalpacino, etut, jia0106, twtw9003, swlee, gykim11}@ssu.ac.kr

## A Sign Language Translator using Data Mining in Kinect Environment

Sang-Jun Lee\*, Tea-Ho Woo\*, Jia Kim\*, Seon-Yeong Park\*,  
Gye-Young Kim\*, Soo-Won Lee\*,

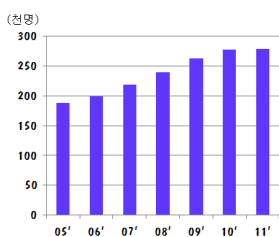
\*School of Computer Science & Engineering, Soong-Sil University

### 요 약

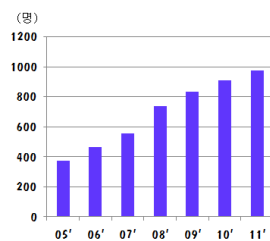
본 연구에서는 키넥트(Kinect) 센서를 통해 수화 동작에서 손의 좌표와 이동방향을 추출하여 속성으로 하고, 데이터 마이닝의 분류 기법을 통해 수화를 인식하여 그 결과를 한글 텍스트로 번역해주는 소프트웨어를 개발한다. 제안 방법의 1단계에서는 0.05초 단위로 추출한 손의 좌표만을 속성으로 한다. 2단계에서는 개개인의 특성 및 화면상의 위치와 같은 요소에 따라 좌표 값이 달라지기 때문에, 손의 움직임에서 변위를 추출하여 손이 움직이는 방향을 속성으로 한다. 하지만 비슷한 방향으로 움직이는 수화가 있을 경우 수화의 구분이 어려우므로 3단계에서는 손의 좌표, 방향 두 가지를 분류하는 속성으로 사용한다. 향후 연구 방향은 수화의 중요한 요소인 손의 위치를 속성으로 추가시키고, 데이터 마이닝의 부스팅(Boosting) 기법을 적용하여 인식률을 높이는 것이다.

### 1. 서 론

현재 우리나라의 청각·언어 등록 장애인의 수는 2011년 12월 말 기준으로 27만 8530명이다([그림 1]). 하지만 2011년까지 등록된 수화통역사의 수는 975명으로 청각·언어 장애인 285명당 수화통역사 1명꼴이다([그림 2]). 따라서 청각·언어 장애인과 비장애인의 편의를 위한 수화 번역 서비스가 필요한 상황이다.



[그림 1] 청각언어 등록 장애인 수



[그림 2] 수화통역사 수

기존에는 말이나 문자를 수화로 번역하는 제품과 관련 연구가 많이 있었던 반면, 수화를 비장애인이 이해할 수 있도록 음성이나 문자로 번역하는 제품은 국내에 상용화된 것을 찾아보기 힘들며 관련 연구도 많지 않다. 청각·언어 장애인과 비장애인의 의사소통을 돕는 연구는 기존의 비장애인의 말이나 문자가 수화로 번역되는 단방향적인 구조에서 비장애인의 말과 문자는 청각·언어 장애인

에게 수화로 번역되고 청각·언어 장애인의 수화는 비장애인이 이해할 수 있는 음성과 문자로 번역되는 쌍방향적인 구조로 전환되어야 한다. 따라서 본 프로젝트는 현재 진행되고 있는 연구와 시중 제품들에 비해서 수화를 문자로 번역한다는 차별성이 있으며 청각·언어 장애인과 비장애인 간의 의사소통을 쌍방향적으로 지원하는 시스템이 개발되는데 시발점이 될 것이다.

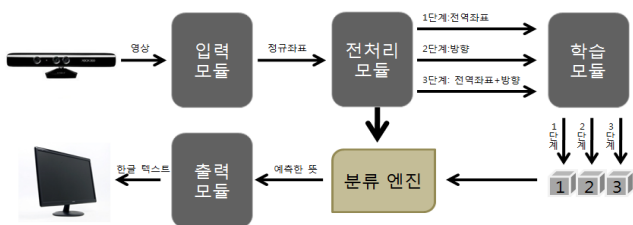
마이크로소프트의 동작인식 센서인 키넥트는 RGB 카메라, 깊이 측정기, 마이크로폰으로 이루어져 있으며 신체의 48곳을 감지해 동작을 추적한다. 센서가 손과 발을 포함해 몸 전체의 움직임을 실시간으로 읽어내고 앞뒤로의 움직임도 파악한다. 따라서 손의 x, y좌표와 손이 키넥트 센서로부터 얼마나 떨어졌는지를 나타내는 깊이 값을 알 수 있다.

[1]에서는 키넥트 센서를 이용하여 깊이 값을 추출하고, 클러스터링 기반 파티클 필터를 이용하여 손 포즈를 추정한다. [2]에서는 동영상에서 손과 얼굴 영역을 추출하고, 시간에 따라 변하는 움직임의 반경 범위를 검출하여 영역의 크기를 정규화한다. 이 과정에서 영상에 나타난 색상을 분석하여 영역을 분할하고 이를 통해 움직임 정보를 추출하여 3차원 볼륨 형태의 데이터로 표현한다. 본 연구에서는 수화에서 중요 요소인 손바닥의 방향을 의미하는 수향과 손의 위치를 알기 위한 손의 좌표를 속성으로 하여 분류작업을 통해 해당 수화 동작에 대한 뜻을 번역하는 소프트웨어를 제작한다[3].

본 연구에서 개발하고자 하는 소프트웨어는 키넥트 환경에서 구동하기 때문에, 컴퓨터 USB에 연결하는 것만으로도 손쉽게 사용이 가능하므로 사용이 편리하다. 또한 수화 장갑 형식의 제품들은 주로 개인이 소유하여 혼자 사용하는 것에 비해 본 연구에서 개발하는 소프트웨어는 공공장소와 같은 사람이 자주 찾는 장소에 설치해두면 그곳을 이용하는 많은 수의 사람들을 대상으로 활용이 가능하므로 타 제품, 기술에 비해 범용적이다. 그러므로 인식률을 높여서 상용화가 가능하게 된다면 과급력이 클 것으로 예상된다.

본 논문의 2장에서는 본 연구에서 개발한 소프트웨어의 시스템 구조도에 대해 서술하고, 이론적 접근 방법에 대해 서술한다. 본 논문의 3장에서는 실험을 서술하고 4장에서는 결론 및 향후 연구에 대해 서술한다.

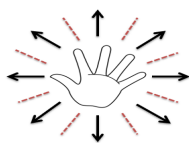
## 2. 데이터 마이닝을 이용한 수화 번역



[그림 3] 시스템 구조도

[그림 3]은 본 연구에서 제안하는 시스템 구조도이다. 입력 모듈에서는 SDK가 제공하는 NUI Library를 이용하여 키넥트 센서를 통해 영상을 인식하여 두 손의 정규좌표를 추출한다. 전처리 모듈에서는 입력 모듈로부터 받은 정규좌표를 학습 데이터의 속성 형태로 변환한다. 학습 모듈은 단계별로 속성이 다른 학습 데이터를 분류기 제작에 사용한다. 1단계에서는 0.2초 단위로 추출한 ‘일련의 손의 좌표’를 학습 데이터의 속성으로 사용하고, 2단계에서는 0.2초 단위로 이전 좌표와 현재 좌표를 비교해 삼각함수를 적용하여 얻어낸 ‘손 움직임의 방향’을 학습 데이터로 사용한다. 3단계에서는 상호보완을 위해 두 가지를 모두 학습 데이터로 한다. 단계별로 추출한 학습 데이터로 각각 세 개의 분류기를 생성하고 이를 바탕으로 분류엔진을 제작하여 수화 동작을 인식하여 뜻을 예측하는데 사용한다.

분류 엔진은 전처리 모듈로부터 학습 데이터를 받아서 해당 수화가 무슨 뜻을 가졌는지 출력 모듈에 넘겨준다. 출력 모듈은 사용자 화면에 그 결과를 출력한다.



[그림 4] 손의 8방향

손 움직임의 방향은 삼각함수를 적용하여 구한다. [그림 4]의 굵은 화살표처럼 8방향으로 나누어서 손의 움직임이 8방향 중 어느 방향에 해당하는 지를 알아낸다. 방향은 각도로 구분하여 알 수 있는데, 실제 손의 움직임은 [그림 4]의 굵은 화살표와 정확하게 일치하지 않음으로 대략적으로 그 방향인지 구별하기 위하여 점선을 경계로 하여 영역을 나눈다. 인식한 손의 좌표를 구하면 이전 좌표를 중심으로 하여 삼각함수 값을 구할 수 있다. 이 값을 통해 움직인 방향이 두 점선 안의 영역에 속하는지를 알 수 있다. 해당 방향이 두 점선 안의 영역에 있으면 두 점선의 중선인 굵은 화살표 방향으로 이동하는 것으로 정한다.

수화 번역을 위한 수화 패턴은 데이터 마이닝의 분류기법을 통해 해결한다. 여러 분류기들 중 일반적으로 인식률이 높은 의사결정나무를 분류기로 선정한다. 오픈 소스인 Weka에서 제공하는 J48 알고리즘을 사용하여 분류기를 제작하고 그 성능을 평가한다.

일련의 손의 좌표, 움직임의 방향을 학습 데이터의 속성으로 하여 정보 이익이 가장 큰 속성을 기반으로 하여 의사 결정 나무를 생성하거나, 각 요소에 대한 분류기를 제작하여 인식률을 비교하는 방식의 부스팅 기법을 적용하여 인식률을 높인다.

사용자가 본 연구에서 개발한 소프트웨어를 실행한 이후 수행하는 동작 중에는 수화 외에 무의미한 동작도 존재한다. 따라서 사용자가 수행하는 동작이 수화인지 무의미한 동작인지를 구분해 내는 작업이 필요하다. 현재는 버튼을 이용하여 수화가 시작되는 순간부터 끝나는 순간만의 동작을 가져오도록 하여 의미 없는 동작을 분리했다. 차후, 시작과 끝 동작을 정하여 의미 없는 동작의 분리가 버튼 없이 가능하도록 한다.

```
@relation sign
|
@attribute ld1 {-1, 0, 1, 2, 3, 4, 5, 6, 7, 8}
@attribute lx1 numeric
@attribute ly1 numeric
@attribute rd1 {-1, 0, 1, 2, 3, 4, 5, 6, 7, 8}
@attribute rx1 numeric
@attribute ry1 numeric
@attribute ld2 {-1, 0, 1, 2, 3, 4, 5, 6, 7, 8}
@attribute lx2 numeric
@attribute ly2 numeric
```

[그림 5] 3단계 Training Data의 내부 구조  
relation - 데이터의 이름,  
attribute - 각 속성의 이름과 유형

[그림 5]는 3단계에서 손의 좌표와 손의 움직임의 방향 모두를 속성으로 한 학습 데이터이다. ld1은 첫 번째 프레임에서 왼손의 방향을 나타내는 속성이고, rd1은 첫 번째 프레임에서 오른손의 방향을 나타내는 속성이다. lx1, ly1은 첫 번째 프레임에서 왼손의 x, y좌표를 나타내는 속성

이고 rx1, ry1은 첫 번째 프레임에서 오른손의 x, y좌표를 나타내는 속성이다. 1단계에서는 0.2초 단위로 손의 좌표를 인식했지만, 3단계에서는 0.2초 단위로 손의 좌표와 손의 방향을 인식한다. 0.2초 단위로 손의 좌표를 인식하여 학습 데이터를 만들어 평가할 때, 최대 3초 동안 수화를 인식하기 때문에 프레임의 개수는 15개이다. 또한 속성은 한 프레임 당 좌표에 해당하는 4개의 속성(양 손의 x, y좌표)과 방향에 해당하는 2개의 속성(왼손, 오른손의 방향)이 프레임마다 저장되므로 총 90개이다.

[표 1] 인식에 사용된 수화 동작(8가지)

뜻	수화	뜻	수화
가다		먹다	
달리다		붓다	
시원하다		덜다	
사과		오늘	

현재 본 연구에서 개발하고자 하는 수화 번역기 소프트웨어의 인식 범위는 [표 1]과 같이 8가지이다.

### 3. 실험

키 160cm, 평균 체중의 여성 2명과 키 180cm, 평균 체중의 남성 2명, 총 4명의 사람이 각 수화에 대해 10번씩 학습한다. 각 단어 당 총 40개의 행으로 구성된 학습 데이터를 만든다.

[표 2] 분류 방법에 따른 3단계 평가 결과(Weka에서 실행)

동작	인식률		
	J48	NaiveBayes	LWL
가다	88.0%	78.6%	69.0%
먹다	88.0%	78.0%	38.0%
달리다	76.0%	78.3%	37.0%
붓다	84.8%	87.0%	69.6%
시원하다	80.0%	90.0%	55.0%
덜다	95.6%	97.8%	88.9%
사과	86.4%	81.8%	65.9%
오늘	86.3%	82.4%	84.3%
평균 인식률	85.7%	84.1%	63.5%

[표 2]는 Weka에서 학습 데이터를 교차 검증으로 평가한 결과이다. [표 2]와 같이 J48이 가장 높은 인식률을 보여 의사결정나무의 C4.5버전인 J48을 분류기로 선택하게 되었다.

[표 3] 단계별 평가 결과

동작	인식률		
	좌표 (1단계)	방향 (2단계)	좌표+방향 (3단계)
가다	70.0%	40.0%	100.0%
먹다	80.0%	50.0%	70.0%
달리다	70.0%	10.0%	90.0%
붓다	80.0%	80.0%	90.0%
시원하다	60.0%	60.0%	60.0%
덜다	70.0%	20.0%	100.0%
사과	70.0%	60.0%	80.0%
오늘	80.0%	30.0%	100.0%
평균 인식률	72.5%	43.8%	86.3%

$$\text{인식률} = \frac{\text{정확히 인식된 동작}}{\text{총 동작}}$$

[수식 1] 인식률 계산식

그 후, 키 160cm의 평균 체중의 여성이 새로운 사례를 제공한 것을 바탕으로 테스트한다. 각 단어 당 10번씩 테스트하여 [수식 1]로 인식률을 계산한 결과가 위 [표 3]과 같다.

### 4. 결론 및 향후 연구

본 연구에서는 키넥트 환경에서 좌표를 추출하여 데이터 마이닝 기법 중 하나인 분류를 통해 수화 동작을 한글 텍스트로 번역하는 소프트웨어를 개발하였다. [표 3]과 같이 손의 좌표, 방향 두 가지를 분류하는 속성으로 사용하였을 때 인식률은 86.3%이다.

향후 연구로는 OpenCV를 사용하여 손의 형상에 대한 정보를 추출한 후 속성에 추가하여 인식할 수 있는 수화의 범위를 확장할 수 있을 것이다. 수화에서는 어떤 손가락을 구부리고 있는지, 어떤 손가락을 펴고 있는지와 같은 손가락의 모양, 형태에 따라 수화의 내용이 달라지기 때문이다. 뿐만 아니라 손의 위치가 수화에서 중요한 요소이므로 손의 위치를 속성으로 했을 때, 인식률을 평가하여 인식률을 높이기 위한 방법을 연구한다.

### 참고문헌

- [1] 유수곤, 신봉기, 이성환, 깊이 영상에서 클러스터링 기반 파티클 필터를 이용한 3차원 손 포즈 추정, 정보과학회논문지 : 컴퓨팅의 실제 및 레터 제18권 제9호, 2012.
- [2] 김호준, 수화 패턴 인식을 위한 2단계 신경망 모델, 한국지능시스템학회 논문지, 제22권 제3호, 2012.