

# 머신러닝과 립 모션을 활용한 지화 번역 시스템 구현에 관한 연구

손다은 (수원대학교 정보통신공학과 [thsek096@naver.com](mailto:thsek096@naver.com))

고형민 (수원대학교 정보통신공학과 [95gudals@naver.com](mailto:95gudals@naver.com))

신행용 (수원대학교 정보통신공학과 [yeh0103@naver.com](mailto:yeh0103@naver.com))

## A Study on Finger Language Translation System using Machine Learning and Leap Motion

Da Eun Son (Dept. of ICT, University of Suwon [thsek096@naver.com](mailto:thsek096@naver.com))

Hyeong Min Go (Dept. of ICT, University of Suwon [95gudals@naver.com](mailto:95gudals@naver.com))

Haeng yong Shin (Dept. of ICT, University of Suwon [yeh0103@naver.com](mailto:yeh0103@naver.com))

### 요 약

농아는 청각장애인과 언어장애인을 말하며 청각장애인과 언어장애인은 음성으로 의사소통하는 것에 어려움이 있기 때문에 수화나 구화 등을 이용하여 의사소통을 한다. 그러나 수화는 모든 사람들이 사용하는 통신 수단이 아니기 때문에 수화를 알지 못하는 사람과의 의사소통하는 데 있어 한계가 생길 수밖에 없다. 본 논문에서는 장애인과 비장애인이 어려움 없이 의사소통할 수 있는 수단으로 지화 번역 시스템을 제안하고 설계 및 구현하였다. 립 모션으로 지화를 인식하였고 인식률을 높이기 위해 머신 러닝 기술을 이용하여 지화 데이터를 스스로 학습시켰다. 구현 및 실험 결과를 통해 제안한 알고리즘 적용으로 인식률 개선이 이뤄졌음을 확인하였다.

### ABSTRACT

Deaf mutism (a hearing-impaired person and speech disorders) communicates using sign language. There are difficulties in communicating by voice. However, sign language can only be limited in communicating with people who know sign language because everyone doesn't use sign language when they communicate. In this paper, a finger language translation system is proposed and implemented as a means for the disabled and the non-disabled to communicate without difficulty. The proposed algorithm recognizes the finger language data by leap motion and self-learns the data using machine learning technology to increase recognition rate. We show performance improvement from the simulation results.

### 1. 서론

전 세계에는 약 3 억 6 천만 명의 농아들이 있고, 이들 중 약 7 천만 명이 수화를 이용하여 다른 사람들과 대화하고 있다. 수화는 농아들에게는 매우 유용한 소통 방식이지만, 모든 사람이 사용하는 통신 수단이 아니라 사용법을 알고 있는 사람들에게만 적용되는 제한적인 언어이다. 그래서 수화를 알지 못하는 사람들과 수화를 사용하는 사람들 사이에 의사소통의 한계가 생기고 이를 해결하기 위해서는 수화 번역기가 필요하다고 생각하였다. 이러한 문제를 해결하기 위해 프로젝트를 진행하게 되었고, 앞에서 말했듯이 수화는 제한적인 언어로 의사소통을 하는 데 한계가 있는 것이 사실이다. 이러한 한계로 인해 농아들은 사회적으로 어려움을 겪고 있으며 사회활동을 하는데

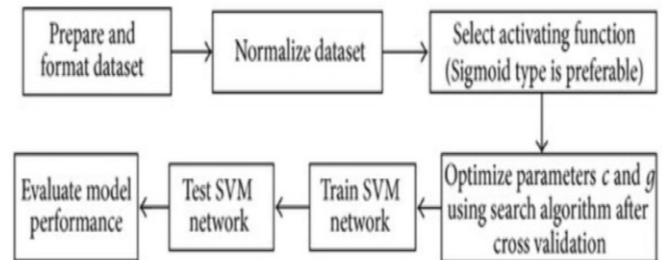
많은 제한이 있다. 수화 번역기를 사용한다면 수화 사용자끼리만 의사소통이 가능한 기존의 생활 환경을 극복하고 경제, 사회활동이 가능해짐에 따라 소비의 수준이 향상될 것이다. 또한 수화 번역기가 있다면 언제 어디서든 수화를 하더라도 원활하게 의사소통이 가능해질 것이다. 이를 위해 본 논문에서는 머신 러닝 기술 및 립 모션 기반의 지화 번역 시스템을 제안하고 설계 및 구현을 통해 인식률 등의 성능 개선을 확인하고자 한다.

### 2. 립 모션 구성

립 모션이란 Michael Buckwald 와 David Holz 가 2010

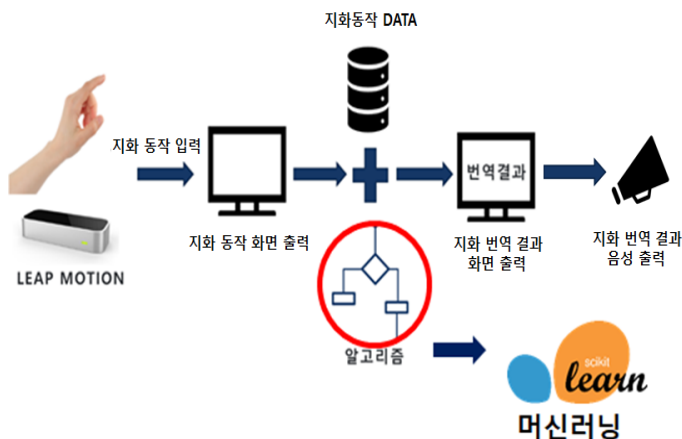
년 설립한 3D 모션컨트롤 기기 개발업체이다 [1]-[2]. 3D 모델링 하는데 키보드와 마우스를 사용한다는 불편을 해소하기 위해 만들었다. 립 모션은 사람의 두 손, 열 개의 손가락의 동작을 감지하여 그 동작을 통해 컴퓨터를 제어할 수 있는 장치이다. 립 모션은 높이 12.7 mm, 너비 80 mm 의 크기를 가지는 USB 방식의 장치로, 내부에 적외선 LED 와 카메라를 내장하고 있어 150 도 시야각과 반경 50cm 이내의 손가락 관절의 움직임을 초당 200 프레임의 속도로 정밀하게 인식할 수 있다[13]. 10 개의 손가락의 독립적인 움직임을 모두 감지하며, 손끝의 3 차원 위치 정보와 방향 정보, 손바닥의 방향 정보까지 인식한다. 또한, 대부분의 운영체제 및 프로그래밍 언어를 위한 SDK 를 지원하고 있기 때문에, 다양한 플랫폼 및 응용 분야에서 새로운 입력 장치로서 활용될 수 있는 잠재력을 가지고 있다.

SVM 기술이 데이터를 학습하는 과정은 아래와 같다. Input dataset 을 준비하고 dataset 을 정규화한 다음 활성화 기능을 선택한다. 그 후에 교차 검증을 하고 검색 알고리즘을 사용하여 매개변수 cost 와 gamma 를 최적화한다. 최적화된 다음 SVM 학습모델을 훈련하고 테스트하는 과정을 거치고 나서 학습모델 성능을 평가한다.



<그림 2> SVM 학습모델 알고리즘

### 3. 지화 번역 시스템 구성



<그림 1> 지화 번역 시스템 구조

립 모션은 손과 손가락의 움직임을 인식하는 소형 모션 컨트롤러이다. 사용자가 수화를 하게 되면 Leap Motion 이 손과 손가락을 인식하여 수화 동작을 입력 받는다. 그 후에 수화 동작 데이터와 수화 번역 알고리즘을 통해 수화 번역 결과를 화면에 출력한다. 수화번역 결과를 음성으로 출력해준다 [3].

### 4. 인식을 최적화

인식률을 높이기 위해 machine learning 기술을 이용하였다. 손과 손가락의 좌표 값에 해당하는 데이터를 스스로 학습하고 학습과정을 거쳐 학습모델을 생성하는데 지도학습의 분류에 해당하는 SVM(Support Vector Machine) 모델을 선택하였다. SVM 학습모델은 test data 를 토대로 분류하면서 학습하는 형태로 일반화의 오류가 낮고 계산 비용이 적게 들며 결과 해석이 쉽고, 저차원과 고차원의 데이터에 모두 잘 작동한다는 장점이 있다 [4]-[6].

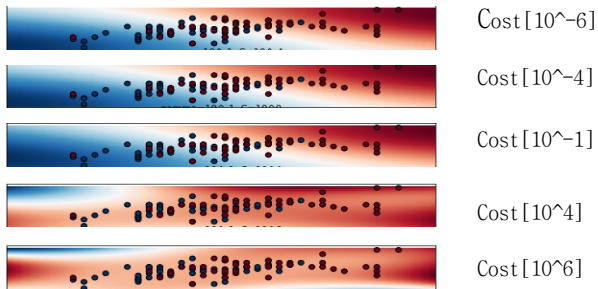
### 5. 구현 결과

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	FRC Area	Class
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	A
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	B
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	C
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	D
0.955	0.000	0.997	0.955	0.975	0.975	0.978	0.953	E
1.000	0.000	0.990	1.000	0.995	0.995	1.000	0.990	F
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	G
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	H
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	I
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	J
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	K
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	L
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	M
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	N
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	O
1.000	0.001	0.958	1.000	0.979	0.978	0.999	0.958	P
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	Q
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	R
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	S
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	T
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	U
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	V
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	W
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	X
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	Y
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	Z
1.000	0.007	0.819	1.000	0.900	0.901	0.996	0.819	AA
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	BB
0.782	0.000	1.000	0.782	0.877	0.881	0.891	0.789	CC
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	DD
0.989	0.000	1.000	0.989	0.994	0.994	0.994	0.989	EE
0.991	0.000	0.992	0.991	0.991	0.991	0.995	0.984	FF

<그림 3> SVM 을 통한 지화 번역 결과 인식률

TP Rate 는 실제 값을 예측한 값이 정답일 경우를 나타내는데 자음 중에는 ‘ㄱ’ 모음에는 ‘ㅋ, ㆁ, ㆁ’ 에서 예측 값이 1.000 이 아닌 0.955, 0.782, 0.989, 0.991 이 나온 것을 볼 수 있다. 또한 Precision(정밀도) 측면에서 보면 ‘ㄱ, ㆁ, ㆁ, ㆁ, ㆁ’ 에서 학습 모델(SVM)이 올바르게 예측했다고 분류한 것 중에서 오류가 있는 것을 알 수 있으며 Recall(재현율)은 올바르게 예측한 결과 중에 SVM 이 올바르게 예측한 비율을 나타내는데 ‘ㄱ, ㆁ, ㆁ, ㆁ’ 에서 0.955, 0.782, 0.989, 0.991 의 비율이 나온 것을 볼 수 있다.

## &lt; Cost, gamma optimization &gt;



&lt;그림 4&gt; Cost, gamma 최적화

교차 검증 및 검색 알고리즘을 통해 Gamma 값과 Cost 값의 최적값으로 다음과 같이 결정하였다.

- $\text{Gamma} = \frac{1}{31}$
- $\text{Cost} = 10^{-4}$

	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	aa	ab	ac	ad	ae		<-- classified as
411	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	a = A
0 308	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	b = B
0 0 323	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	c = C
0 0 0 331	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	d = D
0 0 0 0 298	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	e = E
0 0 0 0 0 310	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	f = F
0 0 0 0 0 0 296	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	g = G
0 0 0 0 0 0 0 289	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	h = H
0 0 0 0 0 0 0 0 287	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	i = I
0 0 0 0 0 0 0 0 0 328	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	j = J
0 0 0 0 0 0 0 0 0 0 273	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	k = K
0 0 0 0 0 0 0 0 0 0 0 318	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	l = L
0 0 0 0 0 0 0 0 0 0 0 0 365	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	m = M
0 0 0 0 0 0 0 0 0 0 0 0 0 300	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	n = N
0 0 0 0 0 0 0 0 0 0 0 0 0 0 342	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	o = O
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 322	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	p = P
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 287	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	q = Q
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 352	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	r = R
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 328	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	s = S
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 416	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	t = T
0 317	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	u = U
0 362	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	v = V
0 351	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	w = W
0 351	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	x = X
0 346	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	y = Y
0 321	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	z = Z
0 334	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	aa = Aa
0 317	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	ab = Bb
0 265	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	ac = Cc
0 321	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	ad = Dd
0 355	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	ae = Ee

&lt;그림 5&gt; SVM 학습모델을 통한 Confusion Matrix

립 모션으로 ‘ㄱ’을 인식했을 때 SVM 학습모델을 통해 ‘ㄱ’으로 예측되는 값을 나타내는 표이다. Confusion Matrix를 보면 ‘ㄱ, ㅂ, ㄴ, ㅋ’에서 인식이 떨어진다.

## 6. 결론 및 향후계획

제한적인 언어이기 때문에 생기는 의사소통의 한계를 해결하고 원활하게 의사소통이 가능해질 수 있게 하기 위해 머신 러닝과 립 모션을 활용한 지화 번역기를 구현하였다. 인식률이 낮아 머신 러닝 기술을 활용하여 지화 번역의 인식률을 높였으며 특히 구분이 힘든 ‘ㄱ, ㅂ’, ‘ㄴ, ㅋ, ㄷ, ㅌ’ 부분에 인식률이 개선되었음을 확인하였다. 향후 계획으로는 현재 지화 번역 정도만 번역이 가능했다면 단어나 문장도 번역할 수 있게 발전시킬 것이다. 또한 부족한 수화 데이터 셋을 추가시켜 실용성을 높일 계획이고, 더 나아가 수화 번역 결과를 동작인식 기술을 사용하여 음성

으로 알려주는 기능도 추가할 계획이다.

## 참고문헌

[1] 신인호, 천동훈, 박한훈, "슈팅 게임의 현실감 개선을 위한 립 모션 기반 인터페이스 구현", Journal of the HCI Society of Korea (한국 HCI 학회 논문지)

[2] 정필성, 조양현, "라즈베리 파이와 립 모션을 이용한 지화 번역 시스템 설계 및 구현", 한국정보통신 학회논문지

[3] <https://www.leapmotion.com/>

[4] 양희덕, "딥러닝과 제스처 인식 기술", 한국방송 · 미디어공학회

[5] 오일석, "기계 학습 MACHINE LEARNING", 한빛아카데미

[6] <https://scikit-learn.org/stable/index.html>

본 논문은 과학기술정보통신부 정보통신창의인재양성 사업의 지원을 통해 수행한 ICT 멘토링 프로젝트 결과물입니다.