

# Sign\_Language.txt

## 요구사항분석서

Date  
2022-04-06

---

### Developer Information

201815004 이찬민

201611205 박승민

201814119 문지영

# 목차

## 1. 개요

- 1.1 프로젝트 기획 배경
- 1.2 기술동향
  - 1.2.1 컴퓨터 비전 (Computer Vision)
    - 1.2.1.1 의미
    - 1.2.1.2 대표적인 기술
    - 1.2.1.3 딥러닝을 활용한 컴퓨터비전
  - 1.2.2 딥러닝 기반 제스처 인식 시스템
    - 1.2.2.1 랜드마크 생성 단계
    - 1.2.2.2 제스처를 분류하는 단계
- 1.3 프로젝트 주요 기능 및 특징
  - 1.3.1 수어 입력 기능
  - 1.3.2 수어를 글로 번역 기능
  - 1.3.3 글로 출력 기능
- 1.4 조원 구성 및 역할 분담
- 1.5 일정

## 2. 기능적 요구사항

- 2.1 Top Level Use Case Diagram
- 2.2 각 기능별 동작 시나리오

## 3. 비기능적 요구사항

- 3.1 사용편리성
- 3.2 신뢰성
- 3.3 성능
- 3.4 이식성
- 3.5 유지관리
- 3.6 구현상 제약사항
- 3.7 인터페이스
- 3.8 법적 제약사항

# 1 개요

## 1.1 프로젝트 기획 배경

저희가 기획한 서비스는 공공장소에서 청각장애인과 한국수어를 모르는 사람 간의 원활한 의사소통을 위한 서비스입니다. (여기서 수어란 수화 언어를 줄여 이르는 말입니다.)

현재 영어 수어(American Sign Language; ASL)에 대한 데이터셋은 잘 구축되어 있고 ASL에 대한 연구는 활발하게 이루어지고 있습니다. 하지만 한국수어는 관련 데이터셋이 많이 없고, 이에 대한 연구도 많이 없습니다. 그래서 저희는 국내 사용을 목적으로 한국수어를 이용하여 서비스를 만들고자 합니다. 또한, 존재하는 모든 한국수화를 학습하는 것은 시간, 용량 등의 물리적 제약으로 거의 불가능합니다. 그래서 병원, 공항, 도서관 등 공공장소로 상황을 제한하고 키워드 중심으로 데이터셋을 수집 및 활용할 것입니다.

수어 관련 어플리케이션을 찾아보면, 지문자 학습 기능, 검색한 단어를 수어 영상으로 보여주거나, 수어 단어 학습 기능 등을 가진 앱이 많고 수어 관련 서비스를 찾아보면, 대부분 텍스트를 지문자로 변환하거나 지문자를 인식해 텍스트로 보여주는 것이 많습니다. 여기서 지문자란 자음과 모음으로 이루어져있고 수화에서 표현할 수 없는 말을 보충하기 위해 문자언어를 손모양이나 손동작으로 표현하는 것입니다. 지문자로 의사소통을 하면 메시지 전달이 느려지고 의미가 명확하지 않을 수 있습니다. 그래서 저희는 단어 단위의 한국수어 데이터셋을 이용하여 학습모델을 만들고자 합니다.

따라서, 저희는 청각장애인이 공공장소를 이용할 때, 한국수어를 모르는 상대방과 원활한 의사소통이 가능하도록 하고 딥러닝을 통해 빠른 인식속도와 인식율을 가진 한국수어 번역 서비스를 개발하려고 합니다.

## 1.2 기술동향

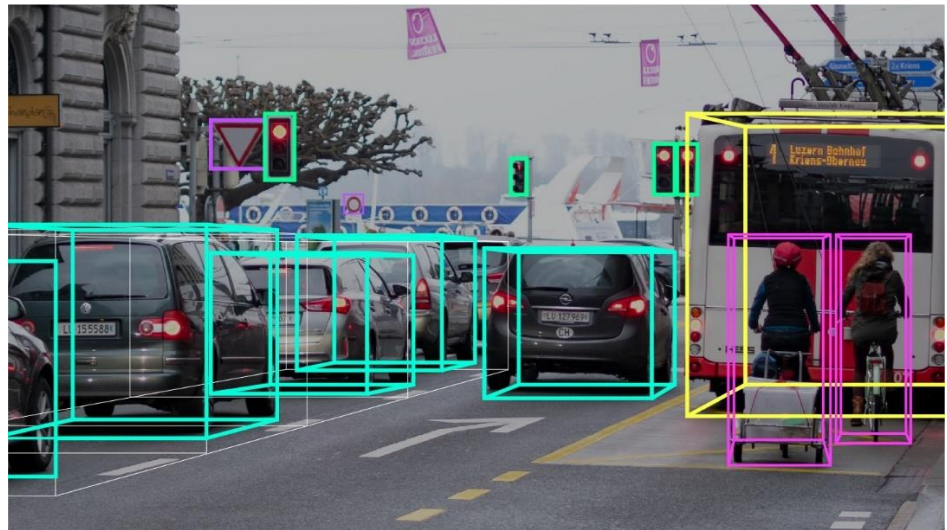
### 1.2.1 컴퓨터 비전 (Computer Vision)

#### 1.2.1.1 의미

컴퓨터 비전은 기계의 시각에 해당하는 부분을 연구하는 컴퓨터 과학의 연구 분야 중 하나입니다. 과학적 관점에서 컴퓨터 비전은 이미지에서 정보를 추출하는 인공 시스템을 연구합니다.

#### 1.2.1.2 대표적인 기술

컴퓨터 비전 분야에 대표적인 기술에는 이미지 내 객체를 인지하여 그 클래스를 분류하는 기술인 객체 분류, 이미지나 비디오 영상에서 객체를 식별해내는 기술인 객체 탐지 및 위치 식별, 이미지나 비디오 영상 프레임 내에서 객체를 따로 분할하여 의미 있는 부분만 분석할 수 있게 하는 기술인 객체 분할, 비디오 영상 내의 객체의 행동을 인식하여 분류하는 기술인 행동 분류가 등이 있습니다.

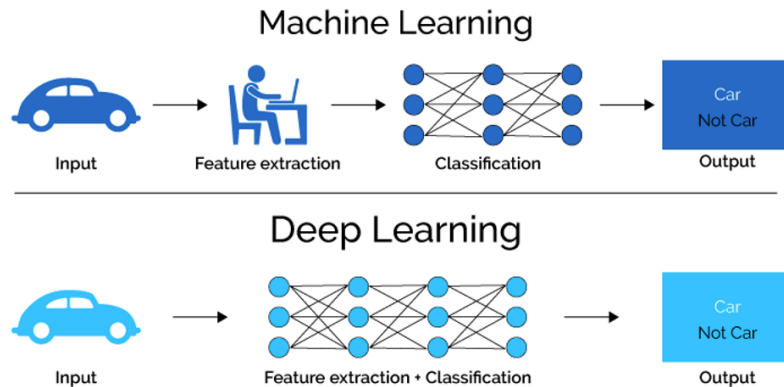


출처: <https://www.superb-ai.com/ko-blog/everything-about-computer-vision-data>

#### 1.2.1.3 딥러닝을 활용한 컴퓨터비전

딥러닝은 Neural network 라고 하는 인간의 뇌에서 특정 부위의 뉴런 간의 상호 연결이 어떻게 동작하는가 에서 영감을 받은 알고리즘을 사용하여 데이터 샘플에서 패턴을 추출하는데 사용합니다.

현재 컴퓨터 비전은 딥러닝을 활용하여 더 좋은 결과물을 만들어 냅니다. 아래 그림은 머신 러닝은 알고리즘을 이용하여 데이터 구문 분석을 하지만 딥러닝 알고리즘은 인공신경망 계층에 의존함을 보여주고 있습니다.



출처: <https://xd.adobe.com/ideas/principles/emerging-technology/what-is-computer-vision-how-does-it-work/>

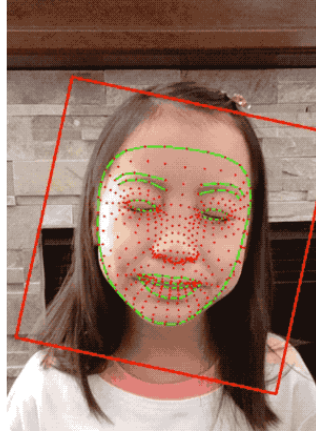
### 1.2.2 딥러닝 기반 제스처 인식 시스템

이전에는 대부분 손의 색깔을 지우고 거기서부터 나오는 음영의 대비된 이미지에 라벨을 붙이고 학습시켰습니다. 이와 같은 방식은 인식률이 좋지 않았고 피부의 색깔을 어떻게 바꾸냐에 따라서도 많은 차이를 보였습니다. 그 이후 손에 랜드마크를 세워 각도와 거리를 계산하는 방식을 이용하여 손의 움직임을 인식하고 라벨을 붙여 학습시키기 시작하였습니다. 하지만 이 방식도 상체를 거의 다 사용하는 수어를 학습시키기에는 부족하였습니다.

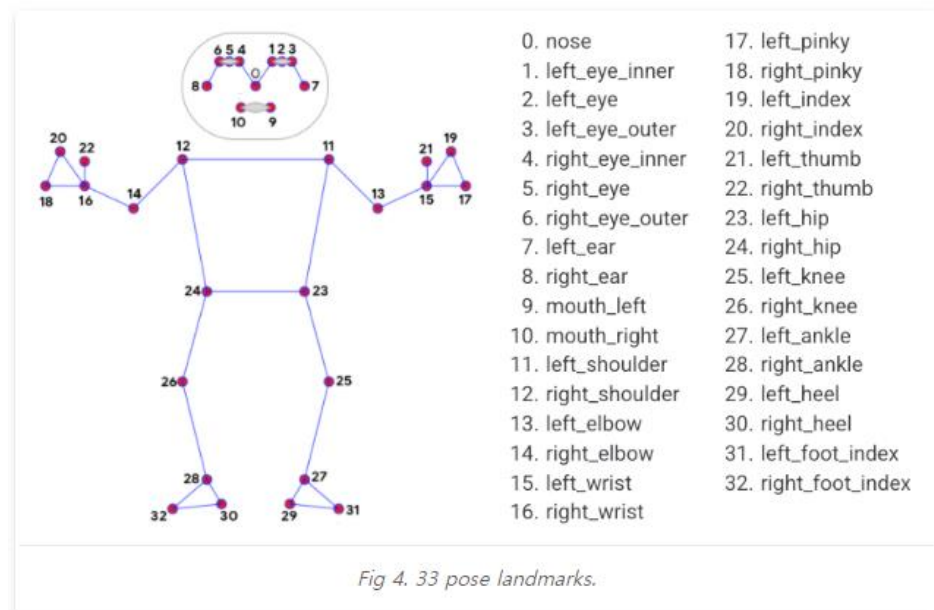
그리고 가장 최근에 구글에서 얼굴, 손에 이어 몸 전신의 움직임을 인식하여 학습시킬 수 있도록 만들어 수어를 학습시키기에 더욱 적합해졌습니다. 수어를 인식하기 위한 제스처 인식은 영상에서부터 제스처를 계산하기 위한 랜드마크를 만드는 단계, 제스처를 분류하는 단계가 있습니다.

#### 1.2.2.1 랜드마크 생성 단계

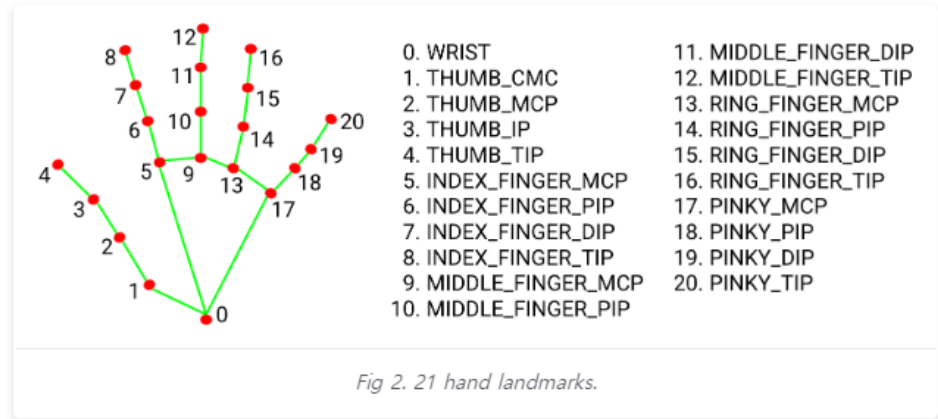
구글에서 제공하는 오픈소스 MediaPipe에서 얼굴은 MediaPipe FaceMesh를 통해 미리 학습된 얼굴인식 기술을 이용하여 얼굴에 468개의 랜드마크를 만들어줍니다.



몸을 인식하기 위한 포즈 랜드마크 모델인 BlazePose GNUM 3D를 이용하여 포즈에 대한 33개의 랜드마크를 만들어줍니다. 아래의 BlazePose GNUM 3D 모델을 사용하면 실시간으로 몸 전체에서 랜드마크를 자동으로 잡아줍니다. 그리고 손과 같은 경우는 MediaPipeHands를 이용하여 각 손에 21개의 랜드마크를 잡아줍니다.



마지막으로 아래의 3가지 모델을 모두 통합한 MediaPipe Holistic은 약 540개의 랜드마크를 만들어 냅니다.



#### 1.2.2.2 제스처를 분류하는 단계

얼굴, 포즈(몸), 손에서부터의 540개 이상의 랜드마크를 이용하여 움직임을 인식하고 각 판단 기준을 이용하여 제스처를 분류합니다.

### 1.3 프로젝트 주요 기능 및 특징

#### 1.3.1 수어 입력 기능

청각장애인이 카메라 앞에서 수화를 하는 것을 의미합니다. 청각장애인은 인식률을 위해 카메라 앞에서 정확하게 수화를 하는 것이 의사소통에 큰 도움이 됩니다.

#### 1.3.2 수어를 글로 번역 기능

앞서 받은 영상 데이터를 실시간으로 학습된 모델을 통해 수어를 예측하는 기능입니다. 오픈소스 활용 및 수집한 데이터로 학습된 모델이 있고, 이것을 통해 영상 처리 및 분석을 하여, 그 결과를 텍스트로 저장합니다.

구체적으로 이 학습모델을 만들기 위해서 먼저, 오픈소스를 활용하거나 Python와 OpenCV, mediapipe을 이용하여 데이터를 수집합니다. 그리고 RNN(LSTM)를 이용하여 수집된 데이터를 학습하여 번역을 위한 모델을 만듭니다.

모든 단어를 학습시키기에는 너무 광범위하고 예측율이 떨어질 것을 감안하여, 특정 공공장소에서 자주 사용할 것 같은 단어를 위주로 학습 모델을 만들 계획입니다. 가장 먼저 병원을 시작으로 환자가 의사에게 자주 사용할 것 같은 수화 데이터를 이용하여 학습모델을 만들 것입니다.

### 1.3.3 글로 출력 기능

실시간으로 수어를 예측하여 수어를 모르는 사람(이하 상대방)에게 글로 보여주는 기능입니다. 수어에는 조사가 없기 때문에 직역한 수어를 글로 상대방에게 보여주게 됩니다. 또한, 상대방은 글을 통해 대답을 할 수 있고, 이 과정에서 청각장애인의 이해를 위해 조사를 지운 글을 보여줄 것입니다.

## 1.4 조원 구성 및 역할 분담

이찬민	1. 딥러닝을 통한 수어 학습 2. 모델 생성 3. 서버에서 모델을 통한 예측이 가능하도록 하기
박승민	1. 학습된 모델로부터 예측결과를 받아오기 2. 웹 상에 내용을 뿌려주기
문지영	1. 청각장애인을 위한 UI/UX 2. 비장애인을 위한 UI/UX 3. 자주 사용하는 표현 UI/UX 4. 웹페이지 생성
공통	1. 데이터 수집 2. 데이터 구축

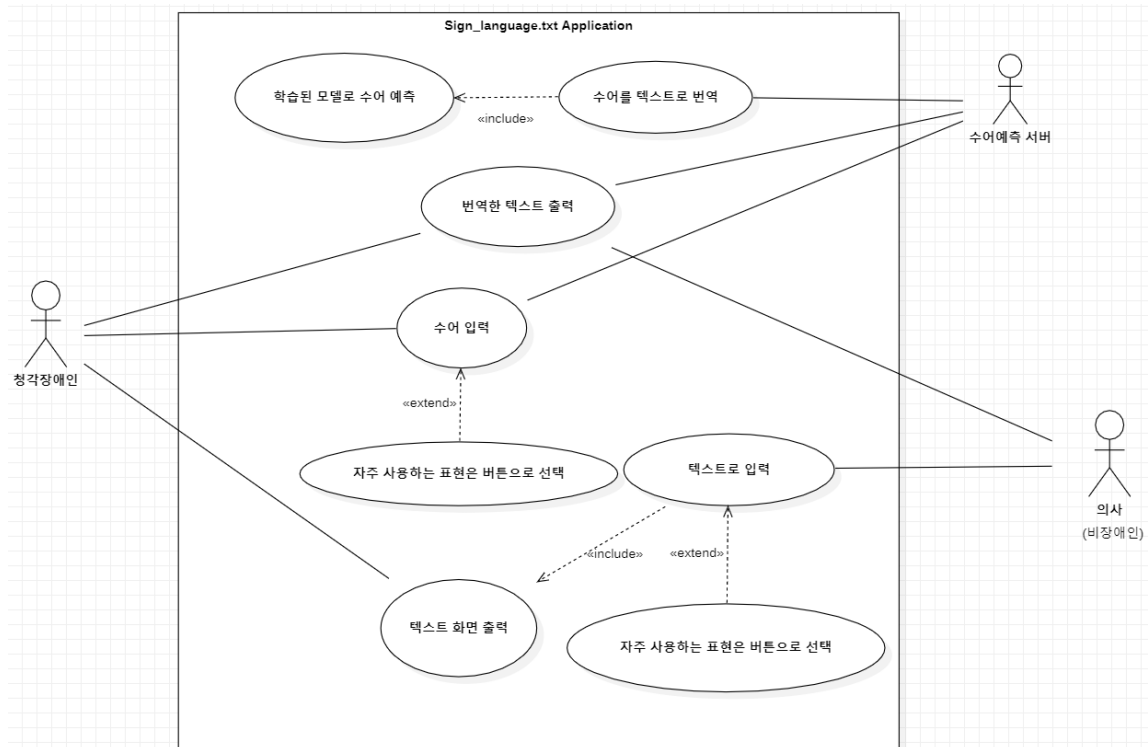
## 1.5 일정

	프로젝트 진행 계획	보고서 제출 일정
3월	- 수어를 글로 번역하는 주제 선정	- 프로젝트 신청서 제출
4월	- 특정 목적(예, 병원)을 위한 수어 데이터를 수집 - 사용할 오픈소스를 탐색	- 요구사항 분석 - 요구사항 분석서 작성 - 시스템 설계
5월	- 프로토타입용 웹페이지, 서버 개발 - 딥러닝을 이용한 영상 처리	- 시스템 설계서 작성 - 프로토타입 설계 및 구현 - 프로토타입 테스트
6월	- 딥러닝을 이용한 영상 처리	- 프로토타입 구현결과 발표 - 기말보고서 작성
7월	- 웹페이지, 서버 개발	
8월	- 웹페이지, 서버 개발 - 인식속도 및 인식률 개선	



## 2 기능적 요구사항

### 2.1 Top Level Use Case Diagram



### 2.2 각 기능별 동작 시나리오

<b>Name of Use Case:</b>	텍스트로 입력
<b>Description:</b>	의사(비장애인)측이 웹의 입력 폼에 텍스트 출력을 위한 텍스트를 입력
<b>Actors:</b>	Initiated by 의사(비장애인) communicated with 청각장애인
<b>Pre-conditions:</b>	1. 서비스에 접속 2. 모니터와 키보드, 마우스가 준비되어야 함
<b>Post-conditions:</b>	1. 의사(비장애인) 화면과 청각장애인 쪽 화면에 내용이 출력된다
<b>Flow</b>	
<b>Actor Actions</b>	<b>System response</b>
1. 의사(비장애인)가 키보드를 이용하여 화면에 내용을 입력한다.	
	1.1 입력된 내용을 의사(비장애인)측 화면과 청각장애인측 화면에 표시한다.
<b>Alternative Path:</b>	
<b>Actor Actions</b>	<b>System response</b>
1. 의사(비장애인)가 자주 사용하는 내용을 버튼 클릭을 이용하여 화면에 내용을 입력한다.	

	1.1 입력된 내용을 의사(비장애인)측 화면과 청각장애인측 화면에 표시한다.
<b>Exception Path:</b>	
<b>Actor Actions</b>	<b>System response</b>
1.a 화면에 내용이 출력되지 않을 경우 새로 고침을 수행한다.	

<b>Name of Use Case:</b>	수어 입력
<b>Description:</b>	청각장애인 측이 화면의 카메라를 향해 수어를 수행
<b>Actors:</b>	Initiated by 청각장애인 communicated with 수화예측서버
<b>Pre-conditions:</b>	1. 서비스에 접속 2. 웹캠 또는 카메라가 탑재된 모니터, 마우스 필요 3. 수어를 예측하기 위한 서버와의 연결
<b>Post-conditions:</b>	1. 수어 정보가 서버로 보내진 상태
<b>Flow</b>	
<b>Actor Actions</b>	<b>System response</b>
1. 청각장애인이 카메라를 향해 수어를 한다.	
	1.1 수어 정보를 서버로 전송한다.
<b>Alternative Path:</b>	
<b>Actor Actions</b>	<b>System response</b>
1.a 청각장애인이 자주 사용하는 내용을 버튼 클릭을 이용하여 화면에 내용을 입력한다.	
	1.a.1 입력된 내용을 의사(비장애인)측 화면과 청각장애인측 화면에 표시한다.
<b>Exception Path:</b>	
<b>Actor Actions</b>	<b>System response</b>

<b>Name of Use Case:</b>	수어를 텍스트로 번역
<b>Description:</b>	인식된 수어로부터 수어의 내용을 예측해 텍스트로 번역
<b>Actors:</b>	Initiated by 수화예측 서버 communicated with 청각장애인
<b>Pre-conditions:</b>	1. 미리 학습된 모델 2. 수어 정보가 들어오면 바로 예측이 가능한 상태
<b>Post-conditions:</b>	1. 청각장애인이 수행한 수어의 의미가 나온다. 2. 의사(비장애인)측 화면과 청각장애인 쪽 화면에 수어의 의미가 출력된다.

Flow	
Actor Actions	System response
1. 청각장애인으로부터 얻어진 수어 입력 정보를 받는다.	
	1.1 수어를 학습된 모델로 예측을 한다.
	1.2 예측된 결과를 출력할 준비를 한다.
Exception Path:	
Actor Actions	System response
	1.1.a 수어를 학습된 모델로 예측 시 학습되지 않은 수어이거나 결과가 번역이 안되는 경우 청각 장애인에게 수어의 재수행을 요구한다.

<b>Name of Use Case:</b>	번역한 텍스트 출력
<b>Description:</b>	서버로부터 받은 수어의 의미를 양측 화면에 출력
<b>Actors:</b>	Initiated by 수화예측서버 communicated with 청각장애인, 의사(비장애인)
<b>Pre-conditions:</b>	1. 번역이 완료된 텍스트 형태의 수어 의미 2. 출력기능에 문제가 없어야 함
<b>Post-conditions:</b>	1. 의사(비장애인)쪽 화면과 청각장애인 쪽 화면에 수어의 의미가 출력된다.
Flow	
Actor Actions	System response
1. 수어 번역이 완료되어 출력한다.	
	1.1 의사(비장애인) 쪽 화면과 청각장애인 쪽 화면에 수어의 의미가 출력된다.
2. 청각장애인 actor가 수어의 의미를 확인한다.	
Alternative Path:	
Actor Actions	System response
Exception Path:	
Actor Actions	System response
	1.1.a 출력이 되지 않는다면 다시 번역된 수어의 의미를 서버로부터 요구한다.
2.a 청각장애인actor가 수행한 수어의 의미가 화면에 표시된 내용과 다르다면 다시 수어를 수행할 수 있도록 한다.	

### 3 비기능적 요구사항

#### 3.1 사용 편리성

- 청각장애인이 잘 사용할 수 있도록 텍스트 중심의 사용 안내할 것입니다.
- 수화를 번역한 텍스트가 소통이 가능한 수준이어야 됩니다.
  - > 문장의 조사를 표현하지는 못하더라도 각 문장성분을 나열해서 소통할 수는 있는 수준이 되어 합니다.
- ex) 나, 학교, 가다 식으로 최소한 SOV순서

#### 3.2 신뢰성

- 사용자의 수화를 인식하는 데 있어서 약 90%의 일치도를 일정시간 이상 내면 그 수화를 인식하여 텍스트로 번역하도록 합니다.

#### 3.3 성능

- 수어 인식시간을 최대 6초로 합니다.
- 수어 예측이 틀린 경우, 재시도 하는 횟수를 최대 3회로 합니다.

#### 3.4 이식성

#### 3.5 유지관리

- 새로운 기능 추가, 성능개선 등의 시스템 업그레이드 등 git을 이용하여 버전관리를 합니다.
- 클린코드로 작성합니다.

#### 3.6 구현상 제약사항

- 딥러닝 기반이므로 개발에 사용되는 하드웨어 성능이 일정수준 이상 요구됩니다.
  - > 하드웨어 성능이 부족할 경우 모션인식에서의 딜레이가 심하게 됩니다.

#### 3.7 인터페이스

#### 3.8 법적 제약사항

- 미리 오픈된 수화기반 데이터 셋 활용함에 있어서 데이터셋의 출처를 밝혀야합니다.