

인공지능 데이터 구축·활용 가이드라인

- 수어영상 AI 데이터 구축 분야 -

인공지능 데이터 구축		데이터 가공 및 주관기업
		데이터 공급 담당
		가상데이터 담당
		모델 개발 담당
	 사단법인 한국농아인협회 KOREA ASSOCIATION OF THE DEAF	클라우드소싱 담당
가이드라인 작성	테스트웍스	금효영
	테스트웍스	김효경
	테스트웍스	최유종
가이드라인 버전	Ver 2.4 2020. 12. 2.	

목 차

1. 데이터 구축 개요	1
2. 임무 정의	1
2.1 임무 정의	1
2.2 데이터 구축 유의사항	1
3. 획득·정제	2
3.1 원시데이터 선정	2
3.2 획득·정제 절차	2
3.3 획득·정제 기준	2
3.4 획득·정제 조직	2
3.5 획득·정제 도구	2
4. 어노테이션/라벨링	2
4.1 어노테이션/라벨링 절차	2
4.2 어노테이션/라벨링 기준	3
4.3 어노테이션/라벨링 조직	4
4.4 어노테이션/라벨링 도구	4
5. 검수	5
5.1 검수 절차	5
5.2 검수 기준	5
5.3 검수 조직	5
5.4 검수 도구	6
5.5 기타 품질관리 활동	6
6. 활용	6
6.1 활용 모델	6
6.1.1 모델 학습	6
6.1.2 서비스 활용 시나리오	6
6.2 접근	6
6.3 유지보수	6
붙임1 인공지능 데이터 명세서 양식	34
붙임2 인공지능 데이터 명세서 작성 예시	36

1. 데이터 구축 개요

1.1 구축 개요

- 본 과제를 통해 구축되는 수어영상 AI 데이터셋은 한국수어의 동작을 인식하고 한국어 문장으로 번역하는 인공지능 개발을 위해 활용되는 데이터셋임
- 사람의 동작인식(Human Pose Estimation) 및 자연어 처리 분야의 인공지능 기술력 발전에 따라 이를 활용한, 수어 영상 인식을 통한 농인-청인 간 커뮤니케이션 장벽을 해소할 수 있는 기술적 가능성이 확대됨
- 이에, 한국수어 동작인식 및 수어 번역 AI 시스템 개발을 위한 수어영상 AI 데이터셋 구축 및 시범서비스를 통한 데이터셋 활용 가능성을 검증하고자 함
- 본 데이터셋의 구축 및 공개를 통해 수어 인식 인공지능 개발의 토대를 마련함으로써 농인이 겪는 의사소통 어려움 극복 및 배리어 프리(Barrier Free) 실현에 기여할 수 있을 것으로 기대됨

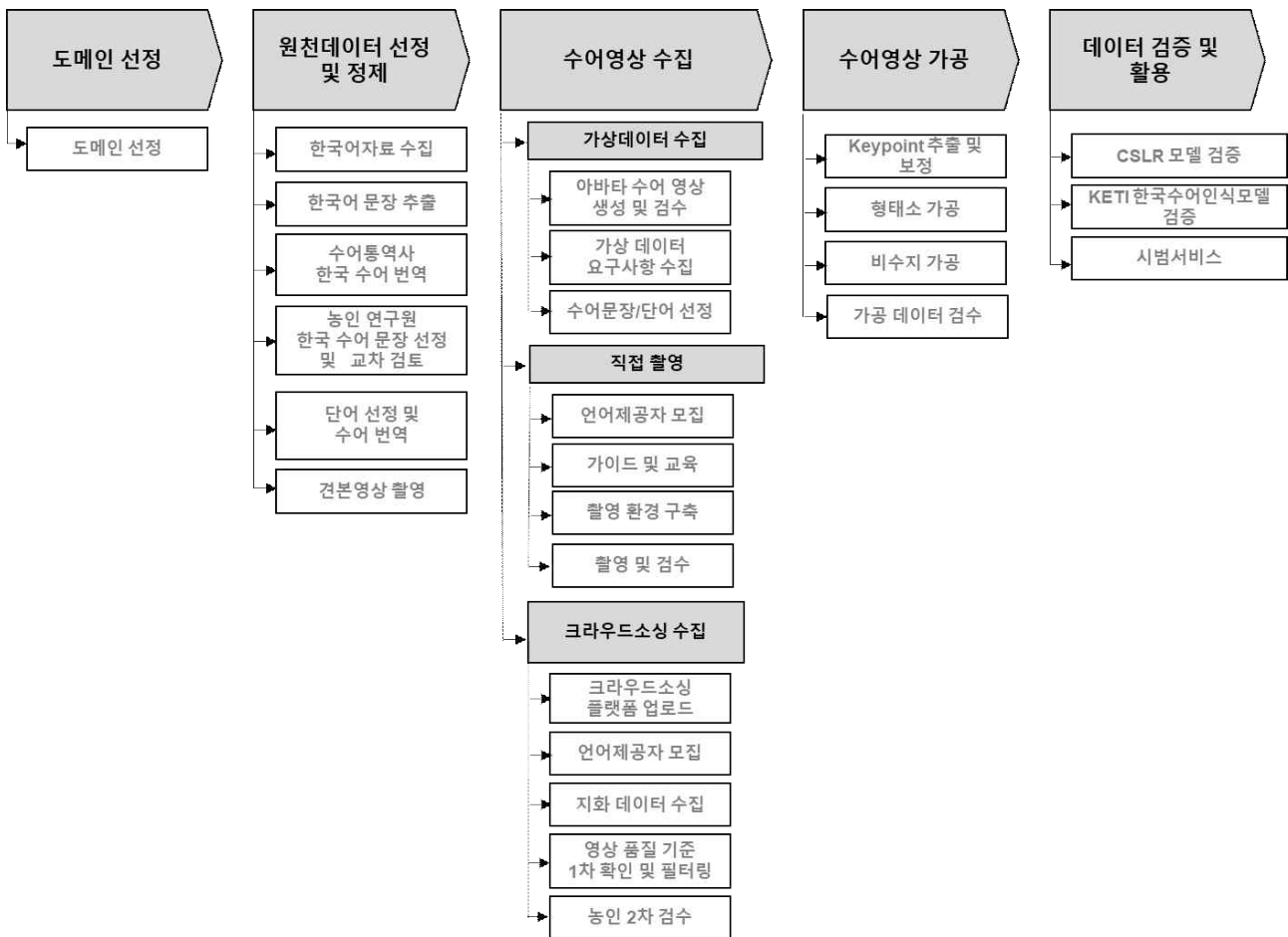


그림1 수어영상 데이터셋 구축 공정

1.2 수어영상 수집방법 및 수집분량

1.2.1 클라우드소싱 : 테스트웍스의 클라우드소싱 플랫폼 aiWorks를 이용하여, 21명의 언어제공자(농인)로부터 지숫자/지문자 1000개의 클립을 제공받음.

1.2.2 직접촬영(나사렛대학교) : 직접촬영 방법으로 촬영소에 5개의 카메라를 5개의 각도로 세팅하여 20명의 언어제공자로부터 단어 3000개, 문자 2000개 클립을 제공 받음

1.2.3 가상데이터(이큐포울) : 직접촬영으로 수집한 데이터를 바탕으로 단어 2000개, 문자 1000개의 클립을 선별하여 데이터를 수집하고, 해당 데이터를 가상의 데이터로 변경 직접촬영의 데이터를 선별하여 사용하기 때문에, 조건은 직접촬영과 동일

수집방법	구분	한국수어 수	영상 클립 수(NN명*5각도)	
직접촬영	단어	3,000	300,000 클립	
	문장	2,000	200,000 클립	
클라우드소싱	지숫자	2,00	4,200	21,000 클립
	지문자	800	16,800	
가상데이터 생성	단어	2,000	10,000 클립	
	문장	1,000	5,000 클립	

표 1

2. 임무 정의

2.1 임무 정의

2.1.1 정성적 임무 정의 및 목표

- 1) 수어를 인식할 수 있는 AI 기술 및 응용 서비스 개발, 인간 행동 인식 인공지능개발을 위한 공개 AI 데이터셋 구축
- 2) 세계 최대 규모 고품질 수어 영상 데이터셋 구축
- 3) 데이터셋 활용 방안 제시 및 성능 검증을 위한 수어 인식 인공지능 모델 개발
- 4) 수어 인식 인공지능 개발을 위한 가상 데이터 활용 가능성 검증
- 5) 수어 인식 인공지능을 활용한 향후 서비스 개발 방향 제시
- 6) 한국수어 영상 데이터셋 구축을 통한 수어인식 인공지능 개발 및 서비스 개발 환경의 기틀을 마련하고 장기적으로 인공지능을 활용한 농인의 접근성 향상 및 배리어 프리 실현에 기여

2.1.2 정량적 임무 정의 및 목표

- 1) 구축 목표
- 2) 품질관리 목표 - openpose에서 추출된 데이터를 가공하기 때문에, 기존 pose estimation 모델에서 측정된 데이터를 보정하여 품질을 높임

2.2 데이터 구축 유의사항

2.2.1 개인정보 처리 방침

- 1) 본 과제를 통해 구축되는 데이터셋은 구축 완료 이후, AIHub를 통해 공개 예정
- 2) 수어의 경우, 얼굴 표정으로 전달되는 메시지(비수지 요소)가 있기 때문에 비식별화가 불가하며 이에 따라 수어 제공자의 얼굴 공개가 불가피
- 3) 이에, 촬영 이전 데이터셋 구축 목적 및 향후 활용 방침에 대해 상세한 설명을 제공하고 아래와 같은 개인정보 활용 동의서에 동의를 받음으로써 법적 이슈 해결
- 4) 개인정보 활용 동의 관련해 법무법인의 자문을 통해 데이터셋 구축 및 공개 과정에서 발생할 수 있는 법적 이슈에 대비
- 5) 본 데이터셋은 한국정보화진흥원에서 운영하는 AIHub를 통해 공개될 예정이며, 플랫폼 관리자에게 신청 및 승인 절차를 통해 데이터 활용 가능하며, 공개 데이터 내 초상권 등 민감정보가 포함되어 있으므로 연구 목적의 공개 신청 시, 심사를 통해 승인
- 6) 개인정보 처리 방침 가이드 등 별도 문서로 첨부

2.3 수어 사용자 요구사항 정의

2.3.1 수어 사용자가 겪는 어려움



< 농인이 겪는 사회 적응 어려움 >

그림 2

- 1) 사회 전반에 걸친 청인 중심의 정보 전달 체계로 인해 일상생활, 경제활동, 긴급상황 등 거의 모든 분야에 걸쳐 다양한 고충을 겪으며 최종적으로 사회 통합과 참여에 있어 제한을 겪음

- 2) 2008년, '장애인차별금지 및 권리구제 등에 관한 법률'(이하 장애인차별금지법)이 제정 및 시행되고 있으며 다양한 분야의 서비스 제공에 있어 차별을 금지하는 조항이 명시되어 있으나 이 조항을 지키기 위한 조치는 거의 이뤄지지 않고 있음
- 3) 재해, 재난 상황 등에서 자유롭지 않은 정보 접근으로 인해 경제적 피해, 생명 피해 등 차별에 노출 빈도 및 확률이 청인에 비해 더욱 높음
- 4) 사회 전반에 걸쳐 소통 방식의 차이로 인한 접근성이 제한되고 있는 상황에서 인공지능 기술 활용의 사회문제 해결 가능성 확인을 통해 농민들에게 필요한 각종 정보에 대한 접근성을 확보하는 다양한 방법 도출 필요

2.3.2 수요 관점

- 1) AI 기술을 활용한 수어영상 인식 서비스로 커뮤니케이션 장벽을 해소할 수 있는 기술적 가능성 확대
- 2) 수어 영상 인식이란 농민의 수어를 인공지능 기술을 통해 의미를 정확히 이해하고 청인이 인지할 수 있는 형태(텍스트 및 오디오)로 출력하는 것을 의미함
- 3) 인공지능 기술이 앞선 미국, 중국, 유럽 국가 중심으로 Computer Vision, NLP 등을 결합 및 활용, 수어를 인식하고 농민-청인 간 의사소통의 장벽 극복을 위한 다양한 시도들이 진행 중
- 4) 한국 수어의 모든 발화를 데이터화 하는 작업은 현실적으로 불가능하며 본 사업의 기간, 예산, 인력 등을 고려한 명확한 목적 도메인의 설정이 필요하며, 이에, 실수요자인 농민 및 전문가 회의를 통해 농민의 편의 증진 기여 가능성, 범용적 활용가능성, 실현 가능성을 도메인 선정 기준으로 길찾기/교통/주소를 도메인을 선정하여 데이터셋을 구축함
- 5) 수요 관점에서 길찾기/교통/주소는 목적지를 향해가는 과정에서 의사소통이 필요할 경우 겪는 불편함이 매우 큰 것으로 자체 인터뷰를 통해 확인함
예를 들어, 정확한 목적지 설명을 못해 택시에서 내려서 1KM 넘게 걸어가거나 익숙하지 않은 목적지를 찾아가는 과정에서 물어볼 수 없어 한참 헤메는 경우 등
- 6) 또한 수요 관점의 요구사항으로, 기존 정부 및 기관 등에서 수어 및 농문화가 고려되지 않은 유사 사업이 많았던 만큼, 한국어 대응 수어가 아닌, 실제 농민들이 사용하는 수어를 인식하고 이를 통해 삶의 질을 높일 수 있는 기술 및 서비스 개발을 요구

2.3.3 문제 정의 정량적 목표

- 1) 정량적 목표 : 단어 2,000개, 문장 3,000개, 지숫자/지문자 1,000개

3. 획득·정제

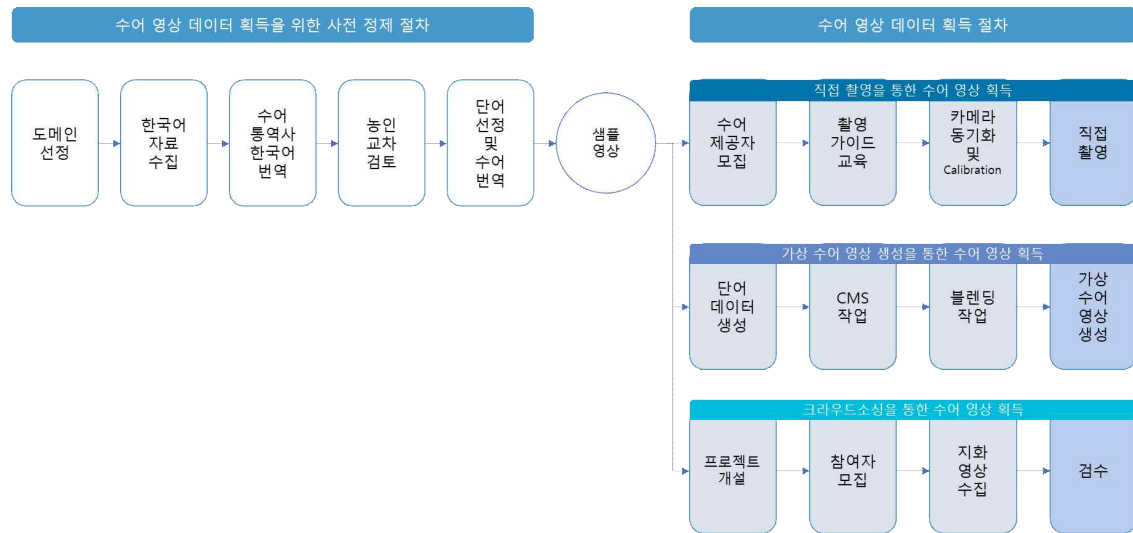


그림 3

3.1 수어 영상 획득을 위한 사전 정제 절차

3.1.1 도메인 선정

- 1) 한국수어의 모든 발화를 데이터화 하는 작업은 현실적으로 단기간에 실행하기에 불가능하며, 성공적인 본 과제수행을 위해 사업의 기간, 예산, 인력, 데이터셋의 향후 활용 가능성과 현실적인 실현 가능성을 고려한 명확한 목적 도메인의 설정이 필수
- 2) 더불어 농인의 편의를 증진할 수 있고, 범용적 활용 가능성을 선정 기준으로 실수요자인 농인 대상 수요 조사와 전문가 회의를 통해 길찾기/교통/주소 도메인을 최종 선정하여 데이터셋 구축 수행

도메인 선정 기준

①농인의 편의 증진에 기여 가능성

- 구축된 데이터셋이 농인이 겪는 다양한 어려움을 해소하는데 기여하는 서비스와 연계되어야 함
- 농인이 겪는 어려움의 심각성 및 일상에서의 빈도 등을 함께 고려해야 함

②범용적 활용 가능성

- 향후 다양한 서비스와의 연계를 통한 활용 가능성이 높은 데이터셋 구축 필요
- 시간의 흐름에 따라 기구축된 데이터의 가치가 훼손되지 않고 오랜 기간 유효성을 유지할 수 있는 데이터셋 구축 필요
- 한정된 용도로 사용되는 데이터셋이 아닌 범용적으로 활용 가능한 공공성을 띤 데이터셋 구축이 필요

③실현 가능성

- 원천 데이터 선정 및 정제부터 응용서비스 개발 단계까지 기술적 실현 가능성이 담보되어야 함
- 동일한 의미를 가진 복수의 수어 영상을 최대한 균일한 품질로 수집/가공하는 과정에서 현실적 제약사항이 없어야 함
- 전문용어 등 한국수어에 없는 단어가 많은 것으로 예상되는 도메인이거나 현재 인공지능 수어 인식 성능 수준에 비해 지나치게 복잡한 수어의 인식을 필요로 하는 도메인은 선정 제외

3.1.2 한국어 자료 수집

- 한국어 자료 수집 기준은 단순 문장이 아니라 문장이 사용되는 상황을 종합적으로 파악해 의도 및 상황 목적에 맞는 문장을 선정하는 것임
- 농인들의 이동권을 보장하기 위해 서울교통공사, T-Map 등 수어인식 인공지능의 실제 수요가 예상되는 기관 및 기업을 도메인으로 선정해 해당 기관들에서 빈도 높게 사용되는 한국어 자료를 협조 받아 한국 수어 문장 및 단어, 지문자와 지숫자 선정 과정의 기초자료로 활용
- 도메인과 관련한 한국어 담화 자료는 관련 기관 홈페이지 검색 및 협조 요청, 문헌 조사 등의 방법으로 수집

3.1.3 수어 통역사 한국어 번역

- 해당 도메인에서 빈도 높게 사용되어 수집된 한국어 자료를 기반으로 한국어와 한국수어에 능통한 수어통역사가 농인과 청인의 담화를 예상하여 한국어 문장 2,000개를 선정
- 선정한 한국어 문장 2,000개의 의미를 분석하여 번역
- 각각 한국수어 영상으로 녹화하여 기록
- 한국어로 표현된 문장의 이해에 있어서 농인과 청인 간 차이가 존재하며 선정한 한국어 문장이 가지는 정확한 의미 전달을 위해 농인의 참여에 앞서 한국어와 한국수어에 능통한 수어 통역사의 번역과정이 필요

3.1.4 한국 수어 문장선정 및 농인 교차 검토

- 수어를 제1언어로 하는 농인 연구원은 (나)에서 촬영된 한국수어 영상 자료를 보고 현장에서 실제 사용되는 한국수어 어휘와 문형을 고려하여 수정 및 확정하며, 수정된 내용은 영상으로 촬영
- 한국수어도 독립적 언어이므로 같은 의미가 발화자마다 어휘나 문형이 조금씩 달라 다양한 문장으로 표현될 수 있음

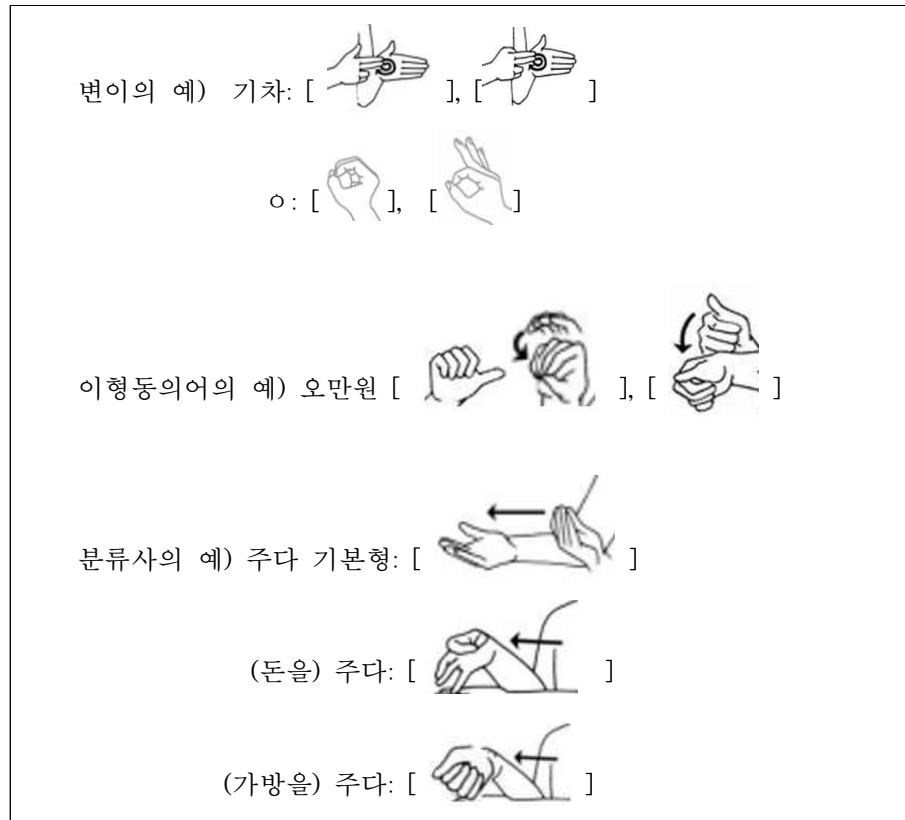
예) 한국어: 서울역으로 가주세요.
한국수어1: [서울] [기차①] [곳]
한국수어2: [서울] [기차②] [곳]
한국수어3: [서울] [KTX] [곳]
한국수어4: [서울] [ㅇ ㄱ ㄱ ①]
한국수어5: [서울] [ㅇ ㄱ ㄱ ②]
한국수어6: [기차①/②] [서울] [곳/ㅇ ㄱ ㄱ ①/②]
한국수어7: [KTX] [서울] [곳/ㅇ ㄱ ㄱ ①/②]

3.1.5 한국수어 단어 선정 및 수어 번역

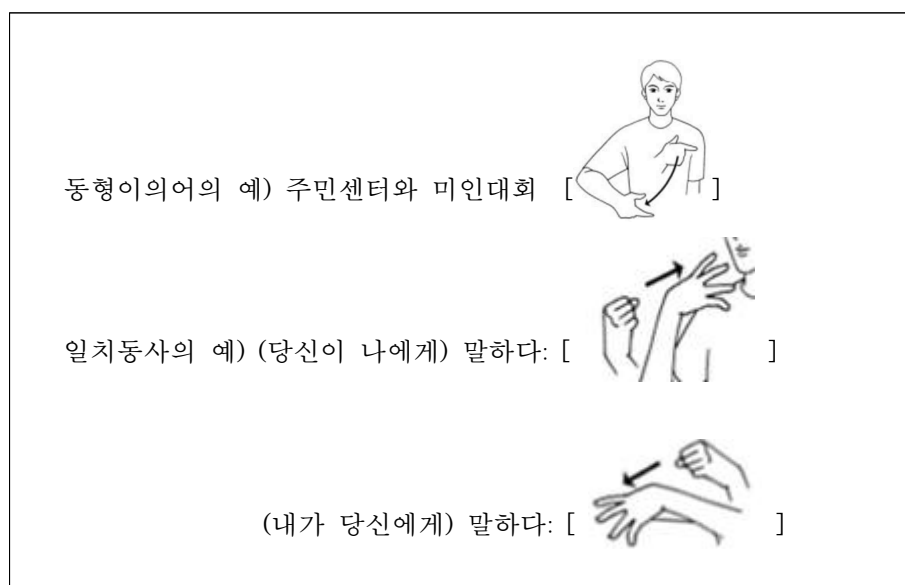
- 한국수어 문장 2,000개에서 단어를 추출하며, 변이형과 일치동사, 분류사도 포함하여 선정

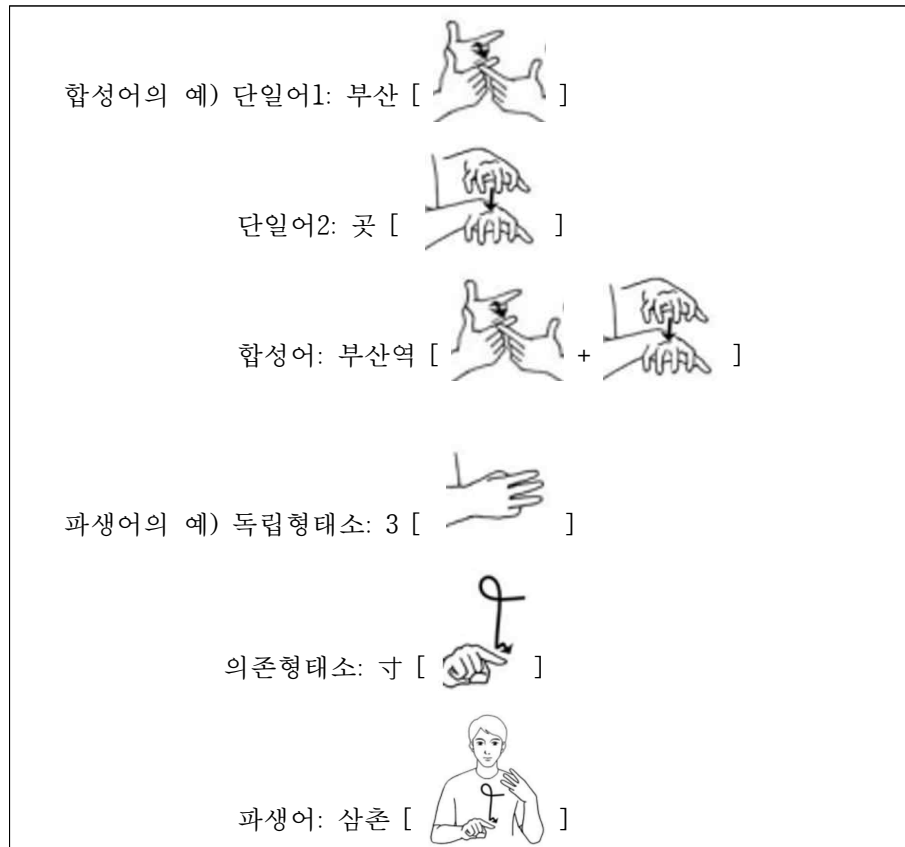
예) 한국어: 서울역으로 가주세요.
한국수어1: [서울] [기차①] [곳]
한국수어2: [서울] [기차②] [곳]
한국수어3: [서울] [KTX] [곳]
한국수어4: [서울] [ㅇ ㄱ ㄱ ①]
한국수어5: [서울] [ㅇ ㄱ ㄱ ②]
한국수어6: [기차①/②] [서울] [곳/ㅇ ㄱ ㄱ ①/②]

- 또한 한국수어의 변이형과 이형동의어, 분류사 등 표현 중 사용 빈도가 높은 것들을 중심으로 문장에 적용하여 추가 추출



- 도메인과 관련한 그 밖의 한국수어 단어와 지문자, 지숫자도 추가 추출하며, 추출한 단어에서 동형이의어와 이형동의어, 합성어와 파생어를 구분하여 표기





- 모든 한국수어 단어는 영상과 문서로 기록하며, 농인 연구원의 교차 검토를 통해 단어 3,000개와 지문자 800개, 지숫자 200개를 최종 선정
- 지명, 지하철역명 등과 같은 고유명사는 한국 수어로 모두 표현되지 않고 주로 지문자로 표현되고 있음. 또한 일상생활에서 날짜, 시간, 금액, 개수 등을 표현하기 위해 지숫자는 빈번하게 사용되는 수어임. 이에, 지문자 800개와 지숫자 200개를 본 과제의 수어 단어 범위에 포함
- 택시와 지하철 상황을 중심으로 빈도 높게 사용될 수 있는 지숫자 200개와 지문자 800개를 추출하여 영상으로 기록
- 획득한 문장 2,000개와 단어 3,000개, 지숫자 200개, 지문자 800개는 농인 연구진이 교차 검토하여 최종 확정하고, 총 6,000개의 샘플 영상을 제작하는 것으로 정제

3.1.6 샘플 영상 촬영 절차

- 위 절차를 거쳐 선정된 한국수어 문장, 단어, 지화를 샘플영상으로 구축함으로써 영식 수집 및 획득 시의 기준을 제시해 균질한 수어영상을 획득함

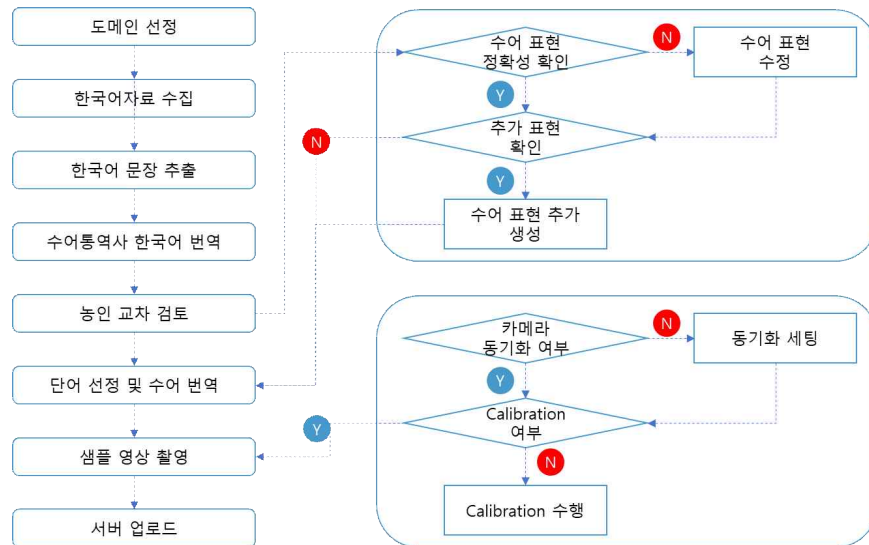


그림 23

직접촬영 영상 획득 절차

- 직접촬영을 통한 영상 획득은 샘플 영상 획득 절차와 유사하며, 한국 수어로 번역할 문장 및 단어 선정이 완료된 이후, 각 문장/단어 당 언어제공자 20명이 참여하여 총 500,000 영상 클립 획득을 목표로 함

순번	절차	내용
1	수어제공자 모집	정확한 한국 수어 영상 획득을 위해 아래 조건을 충족하는 수어제공자 20명 모집 - 19세 이상의 농인 - 한국수어를 제1언어로 사용하는 농인 - 초·중·고등학교 모두 농학교를 졸업한 농인
2	촬영 가이드 교육	영상의 균질성 확보를 위해 수어제공자 대상 촬영 시 주의사항 안내
3	카메라 동기화	카메라 5대간 동기화 여부 확인
4	Calibration	Calibration 확인
5	촬영	수어 영상 촬영 (총 500,000 클립)

3.2 가상 데이터 획득 절차

3.3.1 가상데이터 획득 필요성 및 목적

- 클라우드소싱 및 수어가 등장하는 매체(뉴스)에서 수집한 영상에 비해 스튜디오에서 촬영한 영상 (Real Data)은 품질 보장하고 데이터 셋에 필요한 단어, 문장에 대한 영상 수집이 용이(ex: KETI Sign language dataset)
- 하지만 통제된 환경에서 수집된 데이터는 실제 서비스 환경의 배경 등 데이터 다양성 측면에서 한계가 있고 이는 인공지능 모델 학습에 편향 및 실제 테스트 단계에서 인식을 저하로 연결될 가능성이 있음
- 본 과제를 통해 실제 도메인에서 취득한 수어 영상과 시뮬레이션을 통해 생성한 아바타 수어 영상을 동시에 사용하여 수어 인식모델을 학습함으로써 수어 인식 모델 개발에서 가상 데이터의 활용 가능성 검증을 연구 목표로 설정
- 실수집 데이터와 동일한 샘플 영상을 기반으로 가상 데이터를 생성하여 직접 촬영 데이터와 아바타 데이터 간의 일관성 확보
- 직접 촬영 시의 카메라 각도 및 조명(광량, 방향) 등 촬영 조건을 아바타 생성 시 최대한 유사하게 반영함으로써 직접 촬영 데이터와 아바타 데이터 간의 일관성 확보

3.3.2 가상데이터 획득 절차

- 나사렛대학교에서 선정 및 제공하는 샘플 영상의 범위 내에서 가상데이터를 생성
- 지역이나 그룹 혹은 개인에 따라 동일 단어나 문장에 대해서 서로 다른 형식의 수어 발화의 가능성을 피하기 위해, 가상데이터는 농인 연구원이 직접 촬영한 각각의 표제단어와 표제문장에 대한 샘플 영상을 기반으로 생성
- 농인 연구원이 직접 모션 캡처 의상을 착용하여, 샘플 영상을 기반으로, 표제단어와 표제문장의 움직임을 데이터화

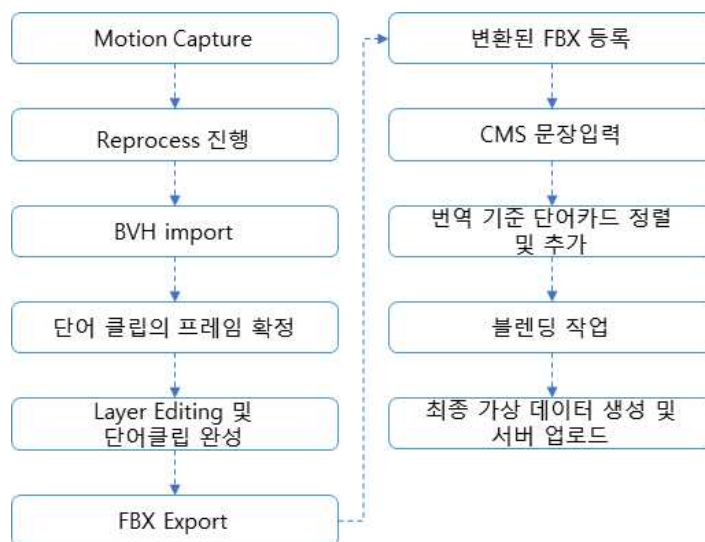


그림 25

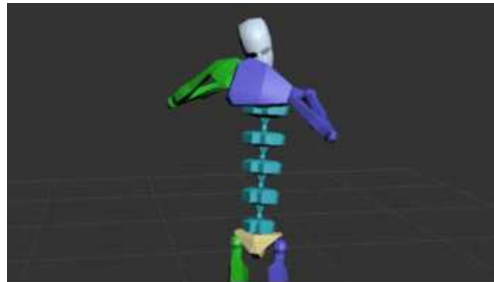
순번	절차	내용
1	단어 데이터 생성	Xsens 장비 및 Xsens MVN을 이용한 Motion Capture 및 파일 변환






3D 프로그램을 통한 단어 클립 완성 (사용 프로그램 : 3D Max)



3D 프로그램을 통한 BVH import 및 단어 클립의 프레임 확정



Layer를 통한 Editing 및 단어 클립 완성 및 FBX Export

2	CMS 작업	<p>변환된 FBX를 등록 후 CMS를 이용한 문장 입력</p> 
		<p>번역 기준 단어카드 정렬 및 추가 등을 통한 CMS 작업</p> 
3	블렌딩 작업	<p>각 단어 별 자연스러운 연결을 위한 블렌딩 작업</p> 
4	생성 데이터 확인	<p>최종 가상데이터 추출</p>



구분	용어	정의 및 상세정보
1	Xsens/Xsens MVN	소형화 및 집적화된 MEMS 관성 측정 센서(IMU) 기술을 기반으로 한 Motion Capture 시스템.
2	Motion Capture	몸에 센서를 부착시키거나, 적외선을 이용하는 등의 방법으로 인체의 움직임을 디지털 형태로 기록하는 작업
3	3D Max	3차원 모델링 프로그램
4	CMS	콘텐츠 관리 시스템
5	블렌딩 작업	영상 병합 작업

3.2.1 클라우드소싱 영상 획득 절차

1) 클라우드 소싱을 통한 영상 획득 목적

<클라우드 소싱 Occlusion 이슈>

1-1 멀티캠의 장점

- 1) 멀티캠 세팅은 여러각도에서 촬영되기 때문에 이를 기반으로 모델을 학습하면 viewpoint variation에 강인한 모델이 학습 가능함.
- 2) pose ground truth를 정확히 어노테이션 하기 위해서는 occlusion 핸들링이 되어야 하고 이를 해결하기 위해서는 멀티캠이 대안이 될 수 있음.

1-2 클라우드 소싱 데이터 의의

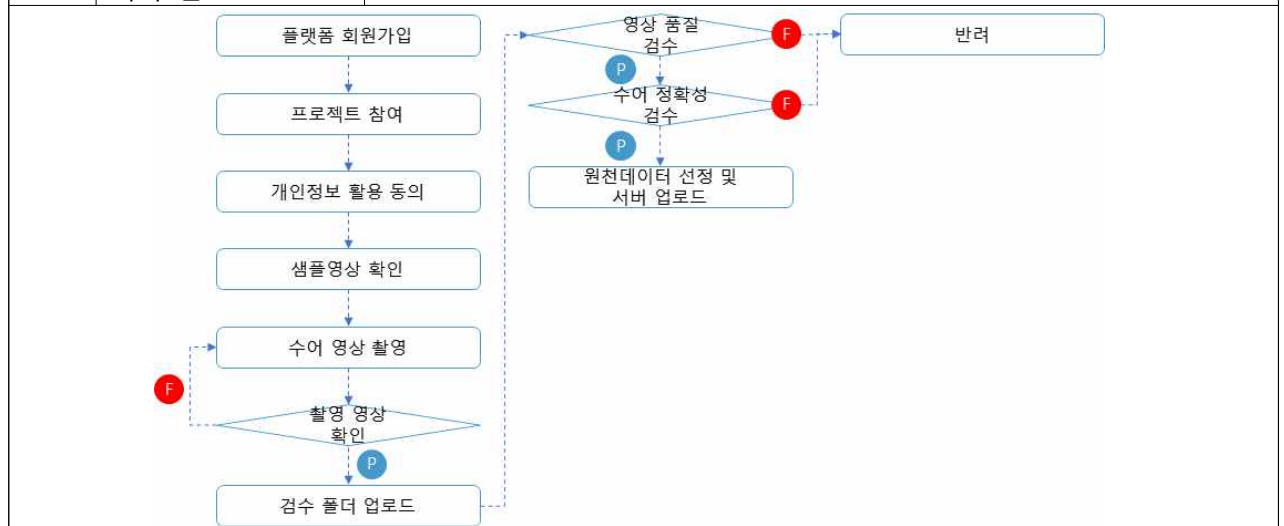
클라우드 소싱 데이터는 멀티캠과 같이 복잡한 셋업을 사용할 수 없음. 하지만 여러 사람, 다양한 배경 데이터를 수집하기에 용이함. 데이터의 특성상 멀티캠보다는 ground truth 퀄리티에 노이즈가 있으므로 클라우드 데이터를 단독으로 학습하기에는 부적합할 수 있음. 하지만 데이터가 촬영된 환경이 application이 적용될 환경과 유사하다는 점에서 두가지 활용 방안에 의의가 있음.

- 1) 멀티캠 데이터에서 학습된 모델을 클라우드 소싱 데이터에 검증함으로써 domain gap이 있을 경우 얼마나 performance gap 이 발생하는지 검증 가능함.
- 2) Domain adaptation 방법들을 사용하여 클라우드 소싱 된 데이터와 멀티캠을 함께 학습. 추후에 application에 적용될 때 클라우드 소싱 데이터의 다양한 정보들이 domain gap에서 오는 performance 하락을 줄이는데 사용 가능함.

2) 클라우드소싱 영상 획득 절차

- 수어 중 지화로 표현되는 단어를 별도로 추출해 클라우드소싱을 통해 획득
- 전국 수어 사용이 가능한 인원을 대상으로 진행
- 지문자 800개, 지숫자 200개를 대상으로, 각 지화별 21명의 수어제공자 참여를 유도해 최종 21,000개 클립의 영상 획득

순번	절차	내용
1	플랫폼 회원가입	농인 및 수어 사용 가능자 모집 홍보를 통해 aiWorks 회원으로 유치
2	프로젝트 참여	aiWorks 내 개설된 수어 영상 수집 프로젝트 참여
3	개인정보 활용 동의	개인정보 활용 동의 등 향후 데이터 공개 및 활용을 위한 법적 동의 절차 진행
4	샘플영상 확인	촬영할 샘플 수어 영상 확인
5	수어 영상 촬영	플랫폼에서 제공하는 UI 에 따라 수어 영상 촬영 시작
6	촬영영상 확인	촬영한 수어 영상 확인, 기준에 어긋날 시 재촬영 수행
7	검수 폴더 업로드	촬영된 수어 영상 업로드 검수 폴더에 업로드, 이후 4 - 7 과정 반복
8	영상 품질 검수	영상의 화질, 초점, 손동작의 화면 밖 이탈 등 영상 품질 중점으로 검수 절차 진행, 품질 기준 미달 시 데이터 반려
9	수어 정확성 검수	품질 이상이 없는 영상 데이터일 경우, 정확한 수어 표현이 이뤄졌는 지 수어 전문가의 검수 수행, 표현이 정확하지 않을 경우, 데이터 반려
10	원천데이터 선정 및 서버 업로드	검수가 완료된 데이터의 경우, 가공 절차를 위해 서버 업로드



3.3 획득·정제 기준

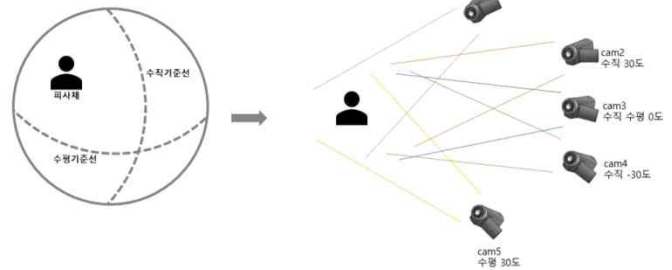
3.3.1 직접 촬영 영상 획득 및 정제 기준

1) 언어 제공자 기준

- 양질의 한국수어 문장 및 단어 영상을 추가 획득하기 위해 참여하는 언어제공자 자격 요건을 다음과 같이 성별과 연령, 지역, 출신 학교 등을 고려하여 선정
 - 19세 이상의 농인
 - 한국수어를 제1언어로 사용하는 농인
 - 초·중·고등학교 모두 농학교를 졸업한 농인
- 수어 가독성을 높이기 위해 촬영용 검정색 상의를 치수별로 비치하고, 언어제공자에게 다음과 같이 사전 안내를 하여 영상을 통일성 있게 수집
 - 검정색 상/하의 착용
 - 안경 미착용

2) 영상 품질 기준

기준		내용		
해상도		1920*1080p (16:9)		
거리		2 M		
화각		약 30°		
카메라 각도 (구도)	정면	수직	0°	
		수평	0°	
	우측면	수직	30°	
		수평	0°	
	좌측면	수직	-30°	
		수평	0°	
	정면상단	수직	0°	
		수평	30°	
	정면하단	수직	0°	
		수평	-30°	
저장형태		.mp4		
Shutter Speed		1/2000s(모션블러 이슈가 발생하지 않는 셔터 스피드)		
A/F		Off		
Frame Rate		30FPS		
캘리브레이션		유		



3.3.2 클라우드소싱 영상 획득 및 정제 기준

- 샘플 촬영된 지문자 및 지숫자의 경우, 주관기관 테스트웍스 자체 클라우드소싱 플랫폼인 aiWorks를 통해 영상을 수집
- 영상의 화질 및 참여의 편의성 등을 고려해 모바일 앱을 통해 지문자 및 지숫자 수어 영상을 수집
- 균일한 품질의 원시 데이터 획득을 위해 아래 기준 설정 및 가이드

기준	내용
디바이스	스마트폰
해상도	1080*1920p (9:16)
거리	1.5 M
저장형태	.mp4

- 클라우드소싱의 경우, 수어 사용이 가능한 불특정 다수의 참여가 예상되고 다양한 촬영 환경이 있으므로 균질한 데이터 획득을 위해 아래와 같은 별도의 검수 절차 시행

순번	절차	내용
1	영상 확인	정상적으로 제출이 완료되었는지 확인
2	영상 품질 검수	영상의 화질, 초점, 손동작의 화면 밖 이탈 등 영상 품질 중점으로 검수 절차 진행, 품질 기준 미달 시 데이터 반려
3	수어 정확성 검수	품질 이상이 없는 영상 데이터일 경우, 정확한 수어 표현이 이뤄졌는지 수어 전문가의 검수 수행, 표현이 정확하지 않을 경우, 데이터 반려
4	반려	검수 기준 미충족 영상은 반려 처리되어 사용자에게 해당 정보 전달
5	해당단어 재수집	영상 분량 확보를 위해 수집되지 못한 단어에 대해 Slot을 재생성해 재수집


```

graph TD
    A[영상 확인] --> B{영상 품질}
    B -- F --> C[반려]
    B -- P --> D{수어 정확도}
    D -- F --> C
    D -- P --> E[서버 업로드]
    C -.-> F[해당단어 재수집]
  
```

3.3.3 가상 데이터 영상의 획득 및 정제 기준

- 직접촬영으로 수집한 데이터를 바탕으로 단어 2000개, 문자 1000개의 클립을 선별하여 데이터를 수집하고, 해당 데이터를 가상의 데이터로 변경 직접촬영의 데이터를 선별하여 사용하기 때문에, 조건은 직접촬영과 동일 따라서, 정제가 추가적으로 필요하지 않음

3.4 획득·정제 조직

- 가공 대상이 되는 원시 데이터의 종류는 샘플 영상, 직접 촬영 획득 영상, 클라우드소싱 획득 영상, 가상 생성을 통한 영상으로 구분
- 각 원시 데이터는 주관 및 참여 기관의 전문 역량에 따라 역할 분배
- 원시 데이터 획득을 위한 일정, 획득 과정에서의 이슈 도출 및 해결, 품질 확보 등 PMO 역할은 주관기관 테스트웍스에서 수행

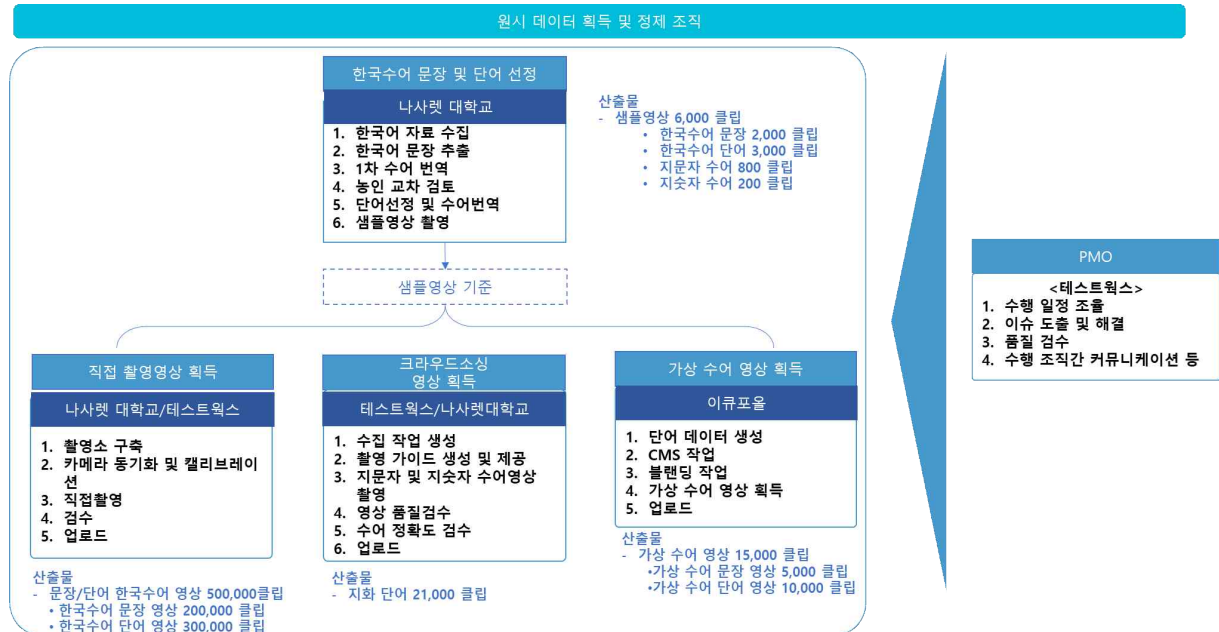


그림 37 원시 데이터 획득 및 정제 조직

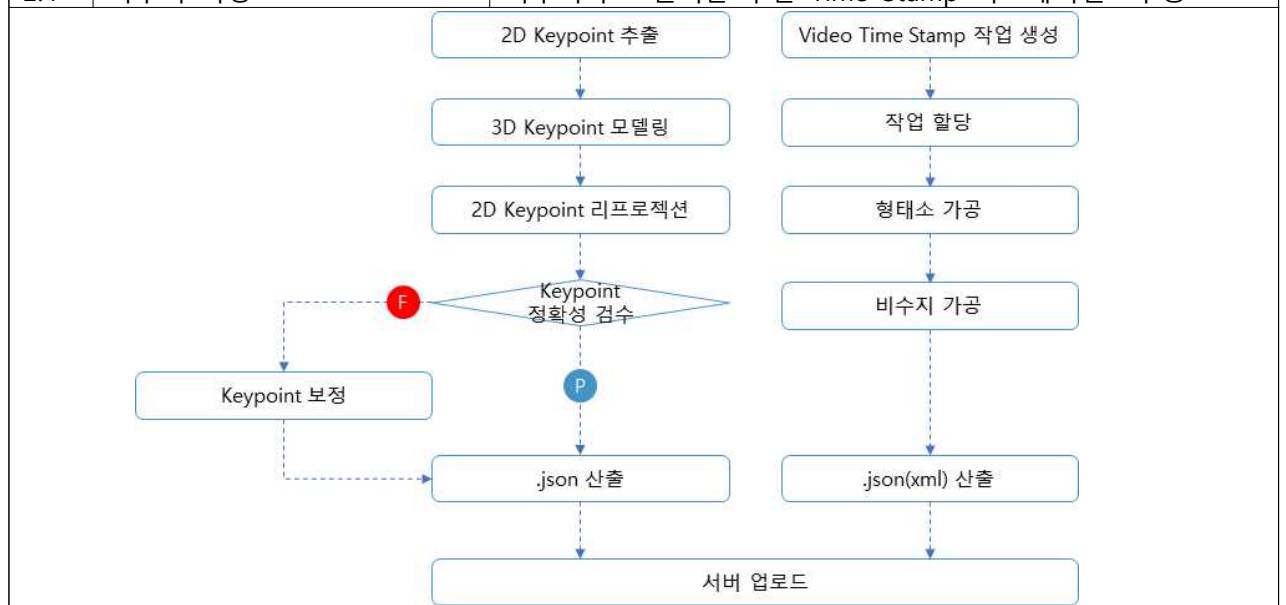
3.5 획득·정제 도구

활용 단계	구분	도구
샘플영상 및 직접촬영 영상획득	HW	BFS-U3-31S4C-C 
	SW	aiWorks 
가상 수어영상 획득	HW	Xsens 장비
	SW	Xsens MVN Animate Pro
		3D Max
		Web기반 CMS 시스템
		블렌딩 툴 Unity

4. 어노테이션/라벨링

4.1 어노테이션/라벨링 절차

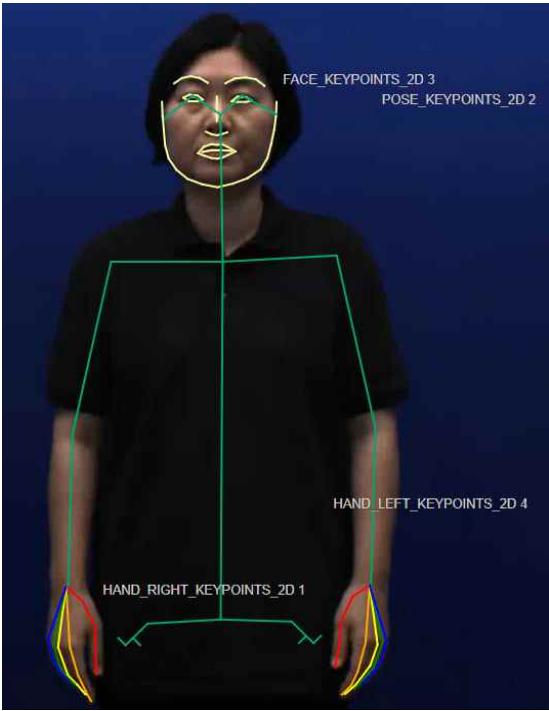
순번	절차	내용
1.1	2D Keypoint 추출	개별 원천 영상의 2D Keypoint 추출
1.2	3D Keypoint 모델링	1.1 결과물을 활용한 3D Keypoint 모델링
1.3	2D Keypoint 리프로젝션	1.2 결과물을 활용, 각 시점 영상 기준으로 리프로젝션
1.4	Keypoint 정확성 검수	주요 특징점이 미추출 혹은 오추출이 된 지점 확인
1.5	Keypoint 보정	이미지 기반 미추출 혹은 오추출 된 특징점을 입력
2.1	Video Time Stamp 작업 생성	웹 기반 Video Time Stamp 도구 내 작업 생성
2.2	작업 할당	농인 Annotator 투입 및 작업 할당
2.3	형태소 가공	형태소 단위로 시작점 및 끝점 Time Stamp 어노테이션 수행
2.4	비수지 가공	비수지가 표현되는 구간 Time Stamp 어노테이션 수행



4.2 어노테이션/라벨링 기준

4.2.1 특징점(Keypoint) 어노테이션 - Blackolive

- 수집된 원천 영상에 대해 Pose Estimation 도구를 활용해, 수어 동작인식에 필요한 신체 주요 특징점(Keypoint)의 좌표를 테스트웍스의 저작도구인 Blackolive를 사용하여 연속적으로 추출


예시 이미지	구분	레이블 내용
	people	"person_id": -1,
	face_keypoints_2d	"face_keypoints_2d": [945.106524633349, 369.30099638632925, 1, 945.79140397612866, 379.82651915646989, 1, 946.75593994246071, 394.40340281496083, 1, 951.84201421005116, 406.03426897797556, 1, 956.92827521557751, 417.67416127035023, 1, 967.1470432340082, 427.55347631801692, 1, ...], 생략
	pose_keypoints_2d	"pose_keypoints_2d": [1086.71, 273.642, 1, 1089.81, 468.385, 1, 942.449, 471.388, 1, 892.071, 719.451, 1, 870.271, 939.214, 1, 1224.98, 468.248, 1, 1268.96, 713.287, 1, 1290.76, 932.954, 1, 1074.24, 942.353, 1, 980.032, 942.374, 1, 970.032, 952.374, 1, 960.032, 962.374, 1, 1174.67, 942.34, 1, 1184.67, 952.34, 1, 1194.67, 962.34, 1, 1045.95, 248.553, 1, 1115.13, 242.346, 1, 1014.53, 292.545, 1, ...], 생략
	hand_left_keypoints_2d	"hand_left_keypoints_2d": [1280.5045545870355, 1015.8415741979892, 1, 1253.984261707637, 1031.6531239532305, 1, 1231.5923998996557, 1065.2369589029993, 1, 1231.2966959528753, 1099.1527836794719, 1, 1226.9137676770551, 1122.1774062036852, 1, 1265.1763462323, 1073.7067287412242, 1, 1259.17441685543, 1118.2948841436787, 1, ...], 생략
	hand_right_keypoints_2d	"hand_right_keypoints_2d": [768.657489274291, 1051.9308453519459, 1, 796.49482053203451, 1067.6195129391749, 1, 815.764352285363, 1097.7834591692806, 1, 810.52363664804443, 1134.1819898254566, 1, 805.96088932735518, 1165.737979178642, 1, 768.23848820580486, 1111.3200680070361, 1, 774.86301138961392, 1156.3914566446329, 1, ...], 생략
	face_keypoints_3d	"face_keypoints_3d": [-133.10242469397861, -359.31445155826992, 2410.4162433018564, 1, -131.91170078967917, -337.68953485645096, 2405.8934943148024, 1, -128.28661318184086, -316.2945375145901, 2404.7252221587, 1, -126.22214262054986, -295.68691204662292, 2397.6841726218363, 1, -120.40788354764592, -276.35291919503072, 2384.2395545835489, 1, ...]
	pose_keypoints_3d	"pose_keypoints_3d": [-52.771458322391133, -332.59267331726875, 2329.141697068816, 1, -52.026553281348086, -138.07851045585682, 2433.9003995113289, 1, -202.95883245005521, -137.62576660994716, 2436.1597705742765, 1, -247.4446448826603, 101.4382364761112, 2408.6096117794327, 1, -261.64607650603523, 299.78648224831704, 2326.4199956624288, 1, 110.18765450680199, ...]

	hand_left_keypoints_3d	"hand_left_keypoints_3d": [167.70571574820769, 300.59105449762296, 2373.4162925146688, 1, 143.51218280928583, 321.41310212473252, 2362.686444827018, 1, 126.87220967535845, 354.62560950317953, 2343.6475963499179, 1, 125.82595313856309, 388.9476364566691, 2339.6469583895887, 1, 126.63112531657312, 413.13320553844136, 2338.0533935438793, 1, ...
	hand_right_keypoints_3d	"hand_right_keypoints_3d": [-268.08337967559987, 298.310620050278, 2324.9367195542727, 1, -244.94653907648529, 320.42053557023371, 2316.2879895183573, 1, -226.85341372738495, 353.99123210111827, 2305.6311993155027, 1, ...
	camparam	"camparam": {"Intrinsics": {"data": "2433.30857499696 0.315735925171355 1008.170460144 0 2433.84664809364 591.098126931379 0 0 1"}, "CameraMatrix": {"data": "1 0 0 0 0 1 0 0 0 1"}, "Distortion": {"rows": "5", "data": "-0.0687674961028772 -0.17948623902024 -0.00234814090149838 -0.00217784771007804 2.69579858267317"}}}
<p>데이터는 각 포인트 별로 아래와 같이 (좌표, confidence)의 형태로 저장됨</p> <p>2D : x1,y1,confidence1,x2,y2,confidence2,</p> <p>3D : x1,y1,z1,confidence1,x2,y2,z2,confidence2,</p>		

- 수어 영상 중 왼손, 오른손의 키포인트 중 누락 및 이탈한 포인트가 존재하는지 확인하여 키포인트 수정 어노테이션 수행
- 3D 키포인트 정보와 카메라 파라미터 정보 포함

4.2.2 형태소 및 비수지 어노테이션

- 수집된 원천 수어 영상의 수어 동작별 의미(형태소)를 해당 동작 시작 시점부터 마지막 시점까지 수어를 제1 언어로 사용하는 농인이 식별하여 어노테이션함
- 수어는 이형동의어, 동형이의어 등 문맥과 상황에 따라 파악되는 의미가 있으며 이로 인한 가공의 혼란을 줄이기 위해 나사렛 대학교의 한국수어 번역 과정에서 수어 Glossary를 구축해 이를 형태소 어노테이션의 Ground Truth 값으로 활용
- 한국 수어의 비수지 요소 관련 자료 및 표준화가 완료되지 않은 관계로 수어 전문가들과의 협의를 바탕으로, 1형태소에 대해 이견의 여지가 없이 명확하게 구분할 수 있는 '의문사가 없는 의문형'을 비수지 어노테이션 요소로 선정하여 어노테이션 함
- 의문사가 없는 의문형의 경우, 눈을 동그랗게 뜨고, 눈썹이 올라가며, 고개가 살짝 앞으로 기울어지는 형태로 비수지 표현이 공통적으로 나타남
- 복합명사와 같은 특수한 형태소의 경우, TTA 표준 가이드라인을 바탕으로 수어전문가들과의 협의를 바탕으로 명확한 기준을 수립한다.

이미지	구분	레이블 내용
	start	해당 형태소 구간의 시작 시각
	end	해당 형태소 구간의 종료 시각
	attributes-name	해당 형태소의 표제어
	attributes-attribute	해당 형태소의 비수지 요소

4.3 어노테이션/라벨링 조직

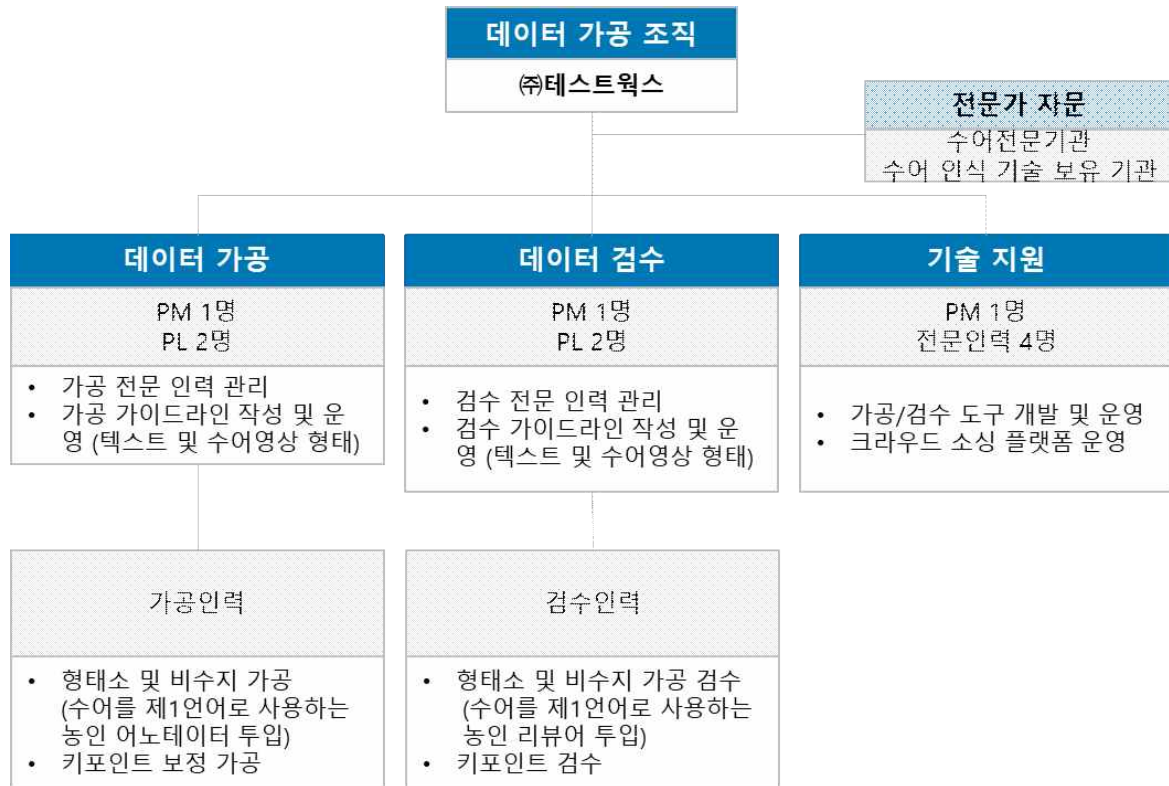


그림 43 데이터 가공 조직

데이터 검수 도식(예시)



그림 44 검수 프로세스

4.4 어노테이션/라벨링 도구

- 수어영상인식 모델 개발에 필요한 주요 특징점의 좌표 및 구간별 의미를 데이터셋으로 구축하기 위해 단계에 적합한 도구를 선정 및 자체 개발을 통해 활용

4.4.1 Human Pose Estimation 도구

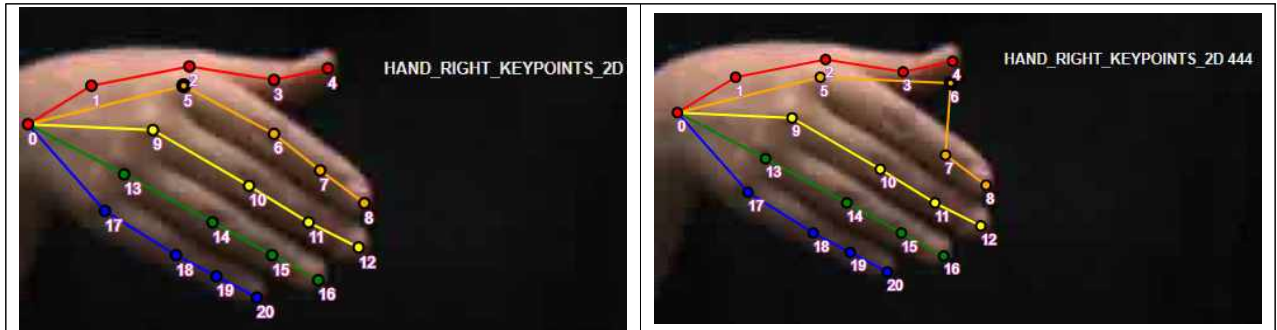
- openpose

openpose란 이번 수어 영상에서 Keypoint를 검출해주는 오픈소스로

- 사진 또는 영상에서 신체의 특정 부분에 대한 특징점(Keypoint)를 탐지(Human Pose Estimation 과정에서 활용되는 오픈소스 도구는 라이선스 이슈로 본 가이드라인에 언급하지 않음)
- 합성곱 신경망(Convolutional Neural Network, CNN)을 기반으로 사람의 몸, 손, 얼굴에 대한 특징점(Keypoint) 데이터를 추출할 수 있음
- 수어 영상을 Human Pose Estimation 도구를 활용하여 2D 키포인트 데이터(JSON) 추출

4.4.2 Keypoint가공 도구 - Blackolive

Blackolive : 테스트웍스 내 데이터 저작도구로 Keypoint 가공 관련 도구로 사용



- 어노테이터(Annotator)의 보정 작업으로, 인공지능 모델 자동화 시 누락되었거나 오류를 포함한 키포인트 수정을 키포인트 저작도구를 이용하여 수정을 통해 데이터셋 품질 확보
- openpose를 이용하여 키포인트를 검출한 후 수동으로 수정하는 경우는 키포인트 객체의 위치 조정, 키포인트의 이탈의 경우임
- 키포인트 저작도구인 Blackolive 내 오른손(Right hand), 왼손(Left hand) 타입 별 템플릿(Template)을 제공하여(선택적으로 보여줄 수 있는 체크박스 및 손가락 별 고유의 속성값을 지님 : 색상), openpose 자동검출 후 수정에 용이하도록 보조하여 수정 작업을 신속하게 할 수 있음, 또한 가공된 키포인트 값에 대한 검수 시 해당기능을 이용하여 검수에 보조
- 생성된 객체는 전체 선택과 마우스 드래그로 이동할 수 있으며, 각 포인트의 위치를 개별적으로 수정할 수 있음
- 어노테이션 작업이 완료되면 어노테이션 다운로드 버튼을 통해 xml 형태로 데이터 추출, 이 후 json 후처리 기능 지원

4.4.3 Time Stamp형태소 가공 도구 - Time Stamp Tool

Time Stamp Tool : 테스트웍스 내 저작도구로 형태소 가공 도구로 사용



→ 형태소의 어노테이션은 영상 전체가 아닌, 수어 표현을 하는 핵심 구간만 가공(레이블링) 한다.

- 자체 개발한 비디오 어노테이션(어노테이션) 도구를 통해, 영상의 특정 구간을 마우스 드래그로 설정
- 구간을 설정하면 해당 구간의 시작 / 끝 타임 스탬프(Time stamp)를 포함한 카드가 자동으로 생성되며, 카드에 수어의 의미(예: 인사)를 입력함. 선택적으로 세부 속성(예: 1형태소 의문사 없는 의문형) 기록 가능
- 복수의 영상을 한 작업으로 리스트화 생성 가능
- 마우스 클릭을 통해 이전/다음 영상으로 넘어갈 수 있음
- 어노테이션 작업이 완료된 프로젝트는 상단 '다운로드'버튼을 통해 JSON 형태로 데이터 추출
- 각 구간별 수어 의미를 도구 내 플레이어에서 자막으로 실시간 확인 가능
- 회원가입, 프로젝트 생성, 데이터베이스를 통한 작업 내용 백업 등 프로젝트 운영에 필요한 전반적인 기능 제공

5. 검수

5.1 검수 절차

5.1.1 데이터 가공 시 검수 절차

- 본 과제에서 데이터 가공 시 필요한 검수는 1) 특징점(Keypoint) 추출의 수정 값 정확성 검수, 2) 형태소 구간 분할의 정확성 검수가 있음
- 각 검수 절차 및 검수 기준 미달 데이터 처리 절차는 아래 표와 같음

순번	절차	내용
1	특징점(Keypoint) 정확성 검수	Pose Estimation 도구의 경우, 성능 개발이 현재 진행 중이기에 완벽한 특징점(Keypoint) Detection 이 현실적으로 되기 어려움 이에, 자체 개발한 도구를 활용해 Frame 단위로 영상 및 Keypoint 좌표를 육안으로 검수해 Keypoint가 신체 부위에 정확하게 추출되지 않은(미추출, 오추출) 이미지 검출
2	특징점(Keypoint) 보정	자체 개발한 Keypoint 보정 도구를 활용해, 신체 해당 지점에 맞도록 좌표 추가 혹은 수정 작업 수행
3	구간 분할 정확성 검수	수어 동작이 연속적으로 진행되는 경우, Transition 발생 등, 특정 의미를 갖는 수어의 시작과 끝을 정확히 구분하기가 어려움 데이터셋의 품질 확보를 위해 분할된 구간이 정확한지 검수 수행
4	다수결 원칙 적용, 구간 수정	수어를 제1언어로 사용하는 농인 어노테이터 3인이 구간을 동시에 확인하고 다수결 원칙에 따라 구간 수정
5	서버 업로드	보정된 Keypoint 좌표 및 구간 정보는 학습용 데이터로 저장되어 서버 업로드

Keypoint 정확성 검수 절차	형태소 구간 분할 정확성 검수
<div>키포인트 보정 도구 접속</div> <div>원천 영상 및 키포인트 데이터(json) 로드</div> <div>Keypoint 정확성 검수</div> <div>키포인트 보정</div> <div>서버 업로드</div>	<div>형태소 가공 도구 접속</div> <div>원천 영상 및 형태소 데이터(json) 로드</div> <div>구간분할 정확성 검수</div> <div>다수결 원칙 적용 구간 수정</div> <div>서버 업로드</div>

5.2 검수 기준

검수절차	통과 조건	중점 확인 내용	구분
키포인트 정확성	주요 관절 지점에 키포인트의 정확한 추출	누락 포인트 및 인체 구조상 불가능한 포인트가 추출될 시, 좌표 스크립트 분석을 통해 검수 대상 이미지 확인	자동
		모든 키포인트가 추출되어 있는지 확인(미추출) 키포인트가 정확한 위치에 추출되어 있는지 확인(오추출)	육안
구간분할	형태소 발화 구간의 정확한	분할 구간에 대해 3인의 어노테이터의 의견이 일	육안

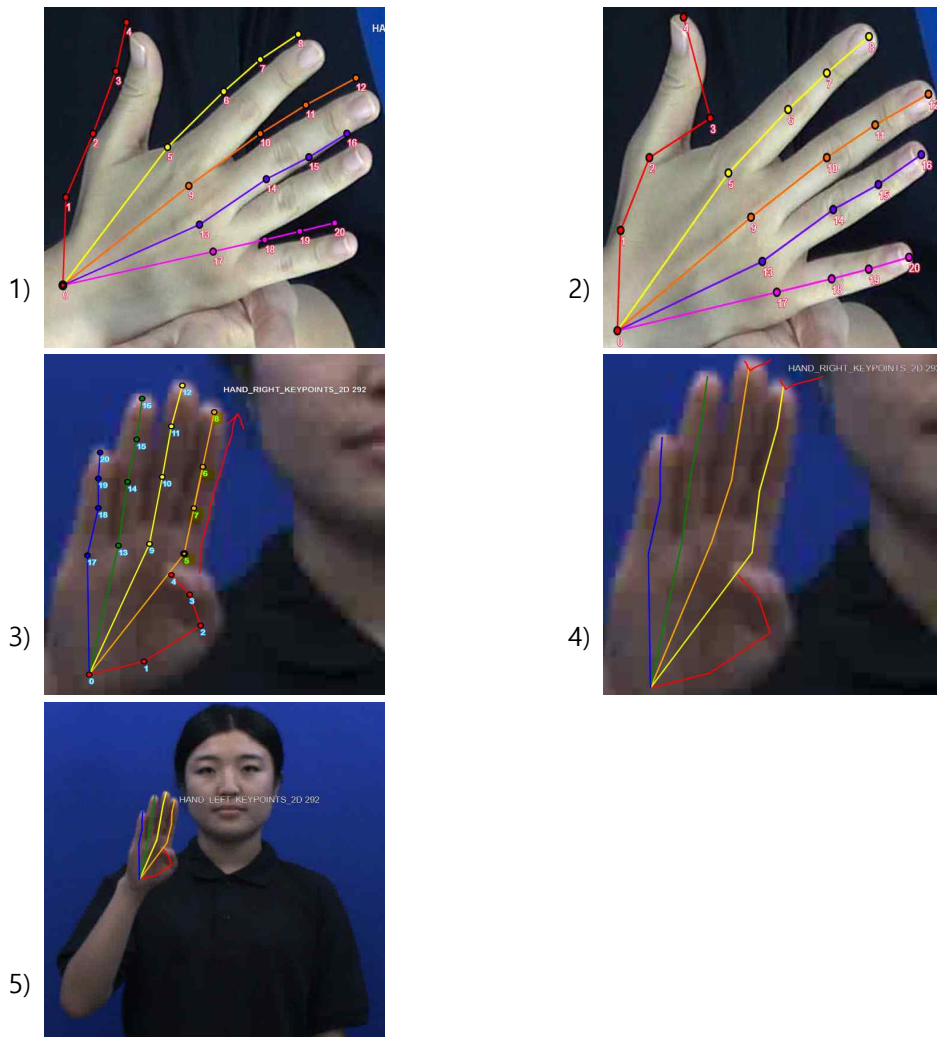
정확성	구분	치하는 지 확인	
형태소 정확성	정확한 형태소 입력	수어 형태소 표제어 Glossary 대로 가공이 되었는 지 확인	육안

5.3 검수 방법

5.3.1 키포인트 검수방법

수어영상 키포인트 가공에서 검수의 기준은 하기와 같다.

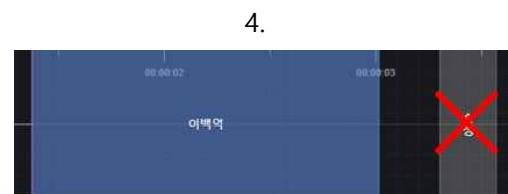
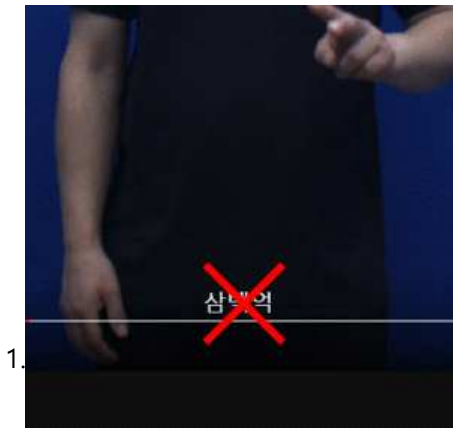
- 1) 전체 이탈 : 키포인트 객체가, 해당 위치가 아닌 다른 곳으로 이탈을 했을 경우, 반려처리를 한다.
- 2) 일부 이탈 : 키포인트 객체 중 일부분이 객체가 아닌 곳에 위치했을 경우, 반려처리를 한다.
- 3) 포인트 순서 값 오류 : 손가락 별 포인트 순서 값이 순서대로(오름차순) 위치하지 않았을 경우, 반려처리를 한다.
- 4) 손가락 위치(고유 색상 값) 오류 : 지정된 색별로 키포인트 값이 위치하지 않았을 경우, 반려처리를 한다,
손가락 별 지정된 색상(엄지 : 빨강, 검지 : 주황, 중지 : 노랑, 약지 : 초록, 소지 : 파랑)
- 5) 객체 라벨링의 오류 : 왼손에 hand_right 혹은 오른손에 hand_left의 객체가 위치 했을 경우 반려처리를 한다.



5.3.2 형태소 검수방법

수어영상 형태소 가공에서 검수의 기준은 하기와 같다.

- 1) 형태소 의미 오류 : 본 수어영상이 담고 있는 형태소가 아닌, 다른 형태소의 정보로 라벨링 되었을 경우
- 2) 형태소 범위 및 구간 오류 : 수어의 의미를 담고있는 구간이 아닌 그 외의 범위, 구간에 라벨링 되어있을 경우
- 3) 비수지 표현 오류 : 비수지 표현이 아닌 수어영상에 비수지 정보가 들어가있을 경우
- 4) 형태소 수량 오류 : 수어영상에서 표현하는 형태소 수량보다 형태소가 많을 경우



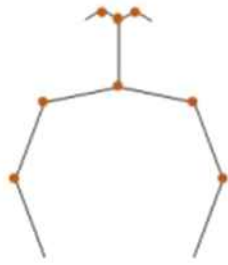
5.4 검수 도구

5.4.1 자동 검수 도구

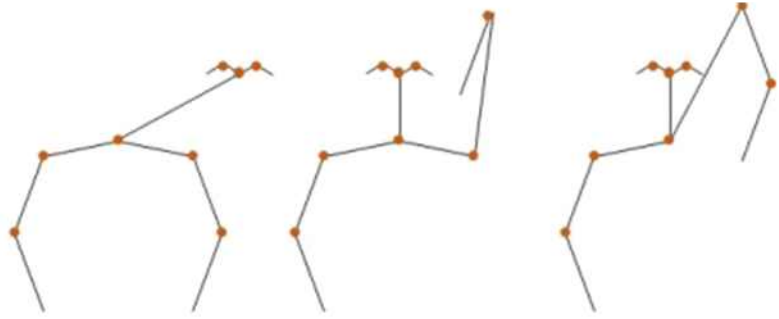
1) 특징점(keyponit) 자동 검사 도구

- 연구개발에서 지원이 가능한, 필터링 기능으로 자동 검수 도구 내용 보완
- 해당 기능으로는 키포인트의 값이 앞, 뒤 프레임에 비해 변경이 많은 케이스를 알려주는 기능으로, 수정주의 구간을 알려주는 도구임
- 누락된 포인트 및 인체 구조상 불가능한 포인트 검사로 가공데이터 무결성 확보
- 수어 영상의 특성 상 왼손, 오른손 등의 포인트 중 누락된 포인트가 존재하면 잘못 가공된 데이터일 확률이 높음

정상 구조



비정상 구조



이에 수어 영상에서 불가능한 인체 구조를 자동으로 인식하여 수집 및 가공 단계에서 오류 데이터의 효율적 제거

6. 활용

6.1 활용 모델

6.1.1 모델 학습

- 본 과제에서 구축되는 데이터셋 검증을 위해 아래 그림과 같이 Continuous Sign Recognition Model을 활용
- CNN (Convolutional Neural Networks)은 수어 영상을 입력으로 받아서 학습에 유용한 정보를 포함하고 있는 피쳐(Feature) 형태로 인코딩하는 역할 수행
- LSTM 모듈이 피쳐에서의 각 프레임에 대한 시간적 (Temporal) 정보를 압축하여 입력받은 수어 영상에 대응되는 글로스 집합을 추정

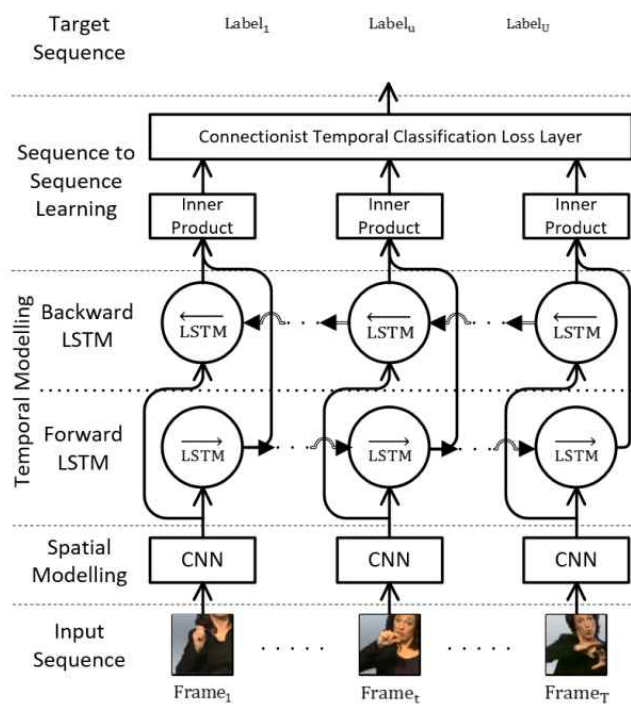


그림 59 모델 학습(CNN)

- 기술한 바와 같이 기존의 실제 도메인에서 학습한 수어 인식 모델의 추가적인 성능 향상을 위해, 시뮬레이션을 이용해 생성된 가상 아바타 수어 영상을 실제 영상과 함께 사용하여 수어 인식 모델을 학습
- 실제 도메인의 데이터와 가상 도메인의 데이터를 혼합하여 학습하는 모델 개요도는 아래의 그림 참고

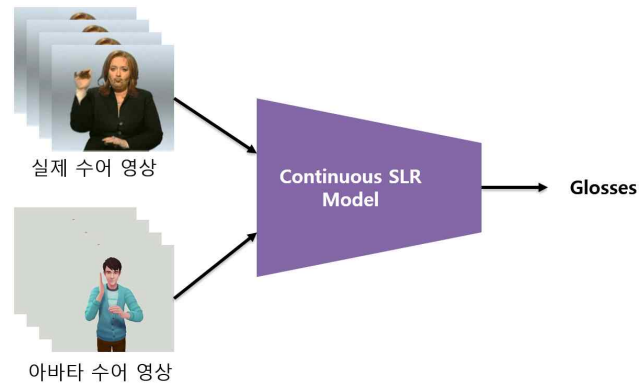


그림 60

- 구체적인 공개 학습 모델의 종류는 추후 변경 가능
- 데이터셋 분할(Split) 비율은 학습(8):검증(1):시험(1)을 권장

6.1.2 서비스 활용 시나리오

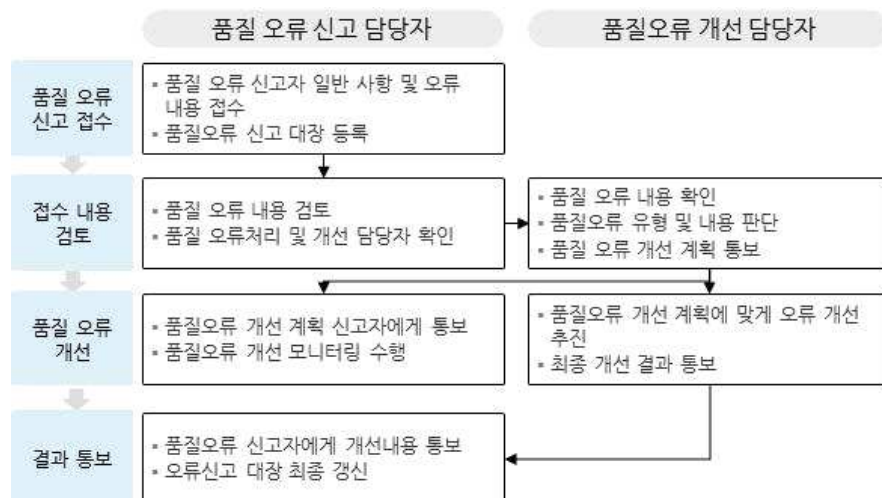
- 아래 두가지 시나리오를 바탕으로 시범서비스 어플리케이션을 개발 예정
- 농인 탑승객일 경우, 택시 내 수어 통역 시스템
- 지하철 역사 내 수어 통역 시스템

6.2 접근

- 본 데이터셋은 한국정보화진흥원에서 운영하는 AIHub를 통해 공개될 예정이며, 플랫폼 관리자에게 신청 및 승인 절차를 통해 데이터 활용 가능
- 공개 데이터 내 초상권 등 민감정보가 포함되어 있으므로 연구 목적의 공개 신청 시, 심사를 통해 승인
- Github 오픈 소스 공유 플랫폼을 활용하여 학습 모델 코드 공개, 공개 범위는 전체를 대상으로 공개.(변경 가능)

6.3 유지보수

- 구축된 수어영상 데이터를 이용하는 외부의 이해관계자(일반국민, 유관기관)로부터 품질 오류에 대한 신고를 접수 받아 확인·조치 및 개선 결과를 통지하는 일련의 과정으로 구성
- 품질오류 신고는 구축 시 인지하지 못한 데이터 오류를 인지하고 이를 개선함으로써 수어영상 AI 학습데이터 품질을 지속적으로 향상



붙임1

인공지능 데이터 명세서 양식

데이터 이름	수어영상 AI데이터				
데이터 포맷	데이터셋 구성 - 수어 영상 (.mp4) - 특징점(keypoint) 가공 데이터 (.json) - 형태소 및 비수지 가공 데이터 (.json)				
활용 분야	사회문제 해결 분야 동작인식 분야				
데이터 요약	수어 동작인식 및 한국수어-한국어 간 번역 인공지능 개발을 위해 활용되는 학습용 데이터셋				
데이터 출처	1. 수어를 사용하는 농인이 직접 촬영을 통한 원시데이터 획득 2. 아바타를 통한 가상 데이터 생성을 통하여 원시데이터 획득				
데이터 이력	배포버전	Ver 1.0			
	개정이력	신규			
	작성자/ 배포자	작성자 최유종 / 배포자 금효영			
데이터 구성	붙임2 참고				
어노테이션 포맷	Keypoint 어노테이션	pose_keypoints_2d	25개 키포인트 각각의 2D x,y,confidence값		
		hand_left_keypoints_2d	21개 키포인트 각각의 2D x,y,confidence값		
		hand_right_keypoints_2d	21개 키포인트 각각의 2D x,y,confidence값		
		face_keypoints_2d	70개 키포인트 각각의 2D x,y,confidence값		
		pose_keypoints_3d	calibration된 3D 값		
		hand_left_keypoints_3d	calibration된 3D 값		
		hand_right_keypoints_3d	calibration된 3D 값		
		face_keypoints_3d	calibration된 3D 값		
		camparam	카메라 파라미터 값		
	형태소/비수지 어노테이션	구분	Tag	Attributes	설명
		Meta	URL	-	수어동영상 경로
		Meta	NAME	-	매칭되는 동영상 이름
		Meta	Duration	-	수어동영상 전체 구간
		Meta	ExportedOn	-	json 다운로드 받은 날짜
		data	start	-	해당 형태소 구간의 시작 시각
		data	end	-	해당 형태소 구간의 종료 시각
		data	attributes	name	해당 형태소의 표제어
		data		attribute	해당 형태소의 비수지 요소
데이터 통계	데이터 구축 규모	데이터셋 종류		동영상 수량	
		수어문장2,000_직접촬영		200,000클립	
		수어문장1,000_가상데이터		5,000클립	
		수어단어3,000_직접촬영		300,000클립	
		수어단어2,000_가상데이터		10,000클립	
		지수어1,000_클라우드소싱		21,000클립	
		총합		536,000클립	
	데이터 분포	'붙임2' 데이터 분포 참조			
	기타 활용 통계	해당 사항 없음			

기타 정보	대표성 (Coverage)	수어를 제1언어로 하는 20대 이상 성인을 수어제공자 대상으로 함
	독립성	원시 데이터 취득 과정에서 개인 민감정보(초상권) 포함이 불가피 취득 과정에서 개인정보 활용동의 및 연구 목적에 한해 개방 필수
	유의사항	개인민감정보의 유출 방지를 위해 목적 확인을 통한 승인 등 사용자 통제 방안 필요
	관련 연구	없음

□ 데이터 포맷

분류	예시	데이터항목			json 형식
영상		구분	Tag	Attributes	{ "0": { "metaData": { "url":"http://52.231.159.158 :{"start":"2.027","end":"3.648","attributes": [{"name":"어디","attribute":["1형태소 의문사 없는 의문형 "]}]}} } } }
		Meta	URL	-	
		Meta	NAME	-	
		Meta	Duration	-	
		Meta	ExportedOn	-	
		data	start	-	
		data	end	-	
		data	attributes	name	
		data		attribute	
이미지		pose_keypoints_2d			,"hand_left_keypoints_2d":[1143.58, 974.645,0.654592,1114.86,992.833, 0.735167,1097.63,1027.29,0.826771 ,1095.71,1060.8,0.792472,1097.63,1 087.6,0.794426,1123.47,1052.18,0.8 39221,1120.6,1091.43,0.734 ... (생략) ,"hand_right_keypoints_2d":[732.28 1,968.632,0.606077,756.648,995.92 3,0.733027,773.218,1033.94,0.8469 74,779.066,1072.92,0.688546,777.1 16,1098.27,0.535953,741.053, ... (생략)
		hand_left_keypoints_2d			
		hand_right_keypoints_2d			
		face_keypoints_2d			
		pose_keypoints_3d			
		hand_left_keypoints_3d			
		hand_right_keypoints_3d			
		face_keypoints_3d			
		camparam			

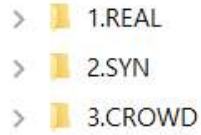
※ 위 json 은 실제 데이터셋 포맷 중 일부 발췌

[현재 디렉토리 구조 및 depth]

- 대분류/중분류/소분류로 디렉토리를 구분

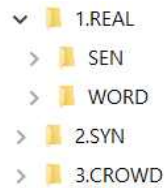
1) 대분류 : 데이터 수집의 종류(직접촬영, 가상데이터, 클라우드소싱)

직접촬영 데이터의 수집은 나사렛대학교, 가상데이터는 EQ4ALL, 클라우드데이터는 aiWorks로 이루어진다.



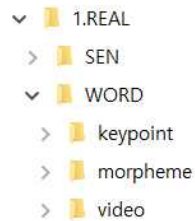
2) 중분류 : 데이터의 정보(문장, 단어, 지수어)

직접촬영 데이터와 가상 데이터의 경우 문장과 단어의 5각도 구성 되지만, 클라우드소싱 데이터 경우 지수어 표현으로, 7각도 만으로 구성이 된다.

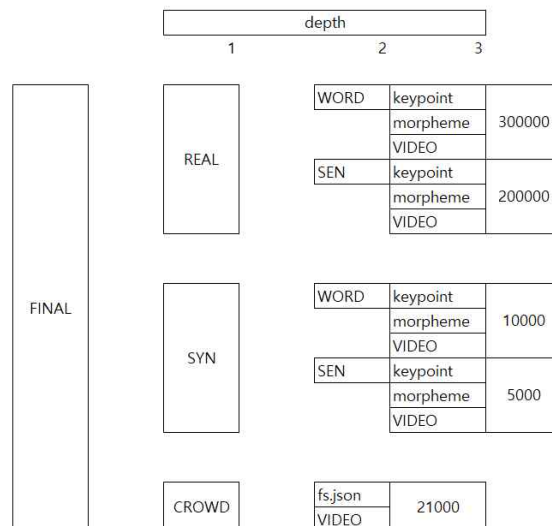


3) 소분류 : 데이터 output의 유형(keypoint, morpheme, video)

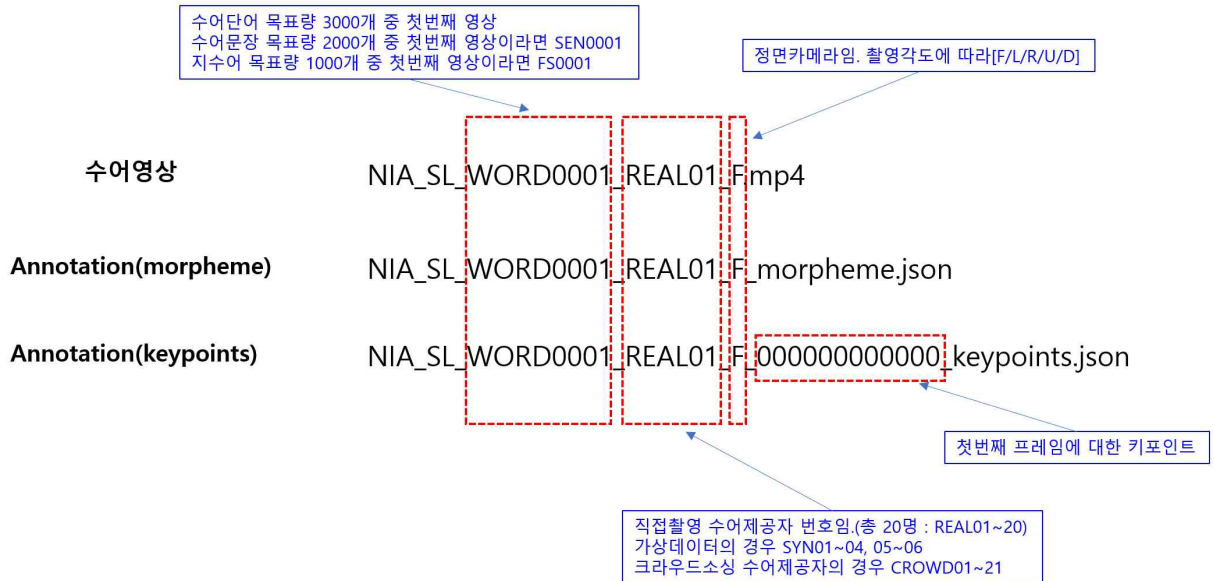
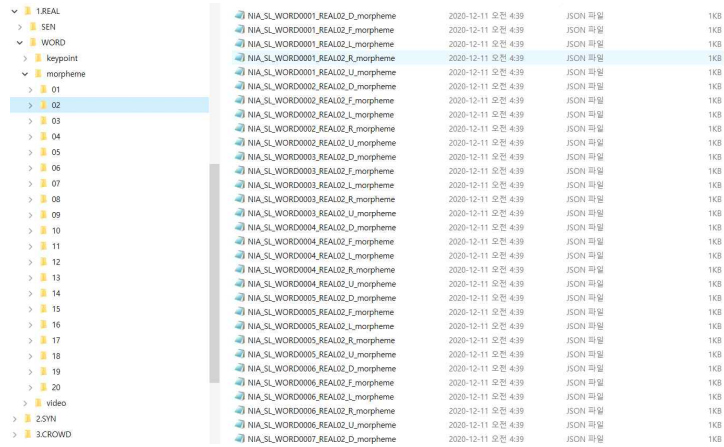
video의 경우 mp4, keypoint와 morpheme은 json으로 구성이 된다.



4) depth



[실제 디렉토리 적용]



수어동영상 파일이름

수어문장/단어/지수어 [NIA_SL_SEN/WORD/FS.XXX]	-	촬영방법 [REAL/SYN/CROWD.XX]	-	촬영각도 [F/U/D/R/L]	.mp4
Required 수어문장2,000개, 단어3,000개, 지문자/숫자1,000개 구별자		Required 촬영방법에 따라 실촬영/가상 데이터/크라우드소싱 수어제공자 20/6/21명 구별자		Required 촬영각도 표식자 크라우드소싱 촬영의 경우 단방향 촬영이어서 "F"만 존재	

가공 파일 이름

수어문장/단어 [NIA_SL_SEN/WORDXXXX]	-	촬영방법 [REAL/SYN/CROWDXX]	-	촬영각도 [F/U/D/R/L]	-	가공 방법 [morpheme/keypoints]	.json
Required 수어문장2,000개, 단어3,000개, 지문자/숫자1,000개 구별자		Required 촬영방법에 따라 실촬영/가상데이터/크라 우드소싱 수어제공자 20/6/21명 구별자		Required 촬영각도 표식자 크 라 우 드 소싱의 경 우 단방향 촬영이 어 서 "F"만 존재		Required Morpheme : 형태소/비수 지요소 가공 keypoints : 동영상의 해 당 프레임의 키포인트 가 공	

□ 어노테이션 포맷

1. 특징점(Keypoint) 어노테이션 파일(json)

Keypoint 어노테이션	pose_keypoints_2d	25개 키포인트 각각의 2D x,y,confidence값
	hand_left_keypoints_2d	21개 키포인트 각각의 2D x,y,confidence값
	hand_right_keypoints_2d	21개 키포인트 각각의 2D x,y,confidence값
	face_keypoints_2d	70개 키포인트 각각의 2D x,y,confidence값
	pose_keypoints_3d	calibration된 3D 값
	hand_left_keypoints_3d	calibration된 3D 값
	hand_right_keypoints_3d	calibration된 3D 값
	face_keypoints_3d	calibration된 3D 값
	camparam	카메라 파라미터

```
{
  "version": 1.3,
  "people": [
    {
      "person_id": -1,
      "face_keypoints_2d": [
        924.372,266.914,1.928,544,285.83,1.932,732,303.772,1.938,865,322.453,1.947,694,340.396,1.959,489,355.626,1,,생략]
      "pose_keypoints_2d": [
        999.286,284.957,1,1001.13,457.677,1,856.06,443.951,1,781.407,615.686,1,889.43,522.544,1,1157.18,465.576,1,,생략]
      "hand_left_keypoints_2d": [
        1189,943.647,1,1165.59,957.199,1,1148.34,984.612,1,1143.72,1016.33,1,1141.57,1038.82,1,1174.22,1012.94,1,,생략]
      "hand_right_keypoints_2d": [
        903.352,525.443,1,924.003,526.178,1,942.236,524,1,950.201,525.559,1,957.454,531.231,1,948.261,478.169,1,,생략]
      "face_keypoints_3d": [
        -0.05331848354191994,-0.307761028483601,2.2448629348987725,1,-0.05423536118312393,-0.2879220150221957,2.2481657673845308,1,-0.05010230242437036,
        -0.26869776419549596,2.249881642986661,1,-0.04403089179581396,-0.24911386580168768,2.2491945131553064,1,-0.03526261505991139,-0.23152802488646446,
        2.242671127325956,1,-0.02357977578998643,-0.21672341222028876,2.22795758110144,1,,생략]
      "pose_keypoints_3d": [
        0.014403601703736208,-0.28331108434798313,2.170256349999683,1,0.017341783764269557,-0.12358898955997814,2.271601801205036,1,-0.12664165323081705,
        -0.1356501844628309,2.279214977911347,1,-0.18615361064500485,0.019356051122002834,2.1196769996893474,1,-0.08007272874450884,-0.06149237900647961,
        1.9559448963165105,1,0.17083326812140107,-0.1143037175495055,2.2594510289906347,1,,생략]
      "hand_left_keypoints_3d": [
        0.19640496536484364,0.34592518801894623,2.186213304283462,1,0.17359734933923943,0.36100523125031486,2.1822197580400755,1,0.15623969655520956,
        0.38559021008434935,2.169292630436767,1,0.1510525994625324,0.4132366928017908,2.1570444991250466,1,0.1487007193569105,0.43386811198658326,
        2.152444172733391,1,0.1790159484789374,0.4110766516145837,2.151235721406932,1,,생략]
      "hand_right_keypoints_3d": [
        -0.06874325510885065,-0.05977030466891,1.9715142412563065,1,-0.0505937175965543,-0.05503007428259695,1.955634799567237,1,-0.034576898871321045,
        -0.05491047520437559,1.9280352936392025,1,-0.027513080836885168,-0.051968217824926485,1.9021371880331717,1,-0.021223938495403234,
        -0.04643703104668351,1.8785174651957919,1,-0.029583923893924224,-0.09452782617340688,1.9237309186384584,1,,생략]
    }
  ],
  "camparam": {
    "Intrinsics": {
      "data": "2266.79453416316 1.09503716185646 955.347282058466 0 2267.20420101965 570.233341394511 0 0 1"
    },
    "CameraMatrix": {
      "data": "1 0 0 0 1 0 0 0 1 0"
    },
    "Distortion": {
      "rows": "5",
      "data": "-0.100686987342219 0.62252535834329 -0.000521413674628704 -0.00124490857307188 -7.20449461083176"
    }
  }
}
```


2. 형태소/비수지 어노테이션(json)

형태소/비수지 어노테이션	구분	Tag	Attributes	설명
	Meta	URL	-	수어동영상 경로
	Meta	NAME	-	매칭되는 동영상 이름
	Meta	Duration	-	수어동영상 전체 구간
	Meta	ExportedOn	-	json 다운로드 받은 날짜
	data	start	-	해당 형태소 구간의 시작 시각
	data	end	-	해당 형태소 구간의 종료 시각
	data	attributes	name	해당 형태소의 표제어
			attribute	해당 형태소의 비수지 요소이며, 해당 속성값이 존재하는 경우에만 존재

```
{
  "metaData": {
    "url": "https://blackolivevideo.blob.core.windows.net/sign-language/1029_moyu0229/NIA_SL_SEN1034_REAL01_F.mp4",
    "name": "NIA_SL_SEN1034_REAL01_F.mp4",
    "duration": 5.117,
    "exportedOn": "2020/12/10"
  },
  "data": [
    {
      "start": 1.517,
      "end": 3.364,
      "attributes": [
        {
          "name": "돈얼마",
          "attribute": [
            "1형태소 의문사 있는 의문형"
          ]
        }
      ]
    }
  ]
}
```



- ※ 수어의 언어 구성 요소에 얼굴 표정이 들어가므로 수어 인식을 위해서는 표정 변화를 확인할 수 없는 얼굴 비식별화 과정을 거칠 수 없지만, 이 가이드라인에서는 개인정보 보호를 위해 비식별화 처리하였음.
- ※ 다양한 비수지요소에 대해 표준안이 마련되어 있지 않은 상황이고, 비수지요소가 수어영상 데이터셋에서 고려된 사례도 세계적으로 전무한 상황 아래 공통성을 발견할 수 있는 "1형태소 의문사가 없는 의문형"에 대한 비수지 가공을 우선적으로 수행함

□ 데이터 통계

○ 데이터 구축 규모

데이터셋 종류	동영상 수량	json 수량	output data구성
수어문장2000_직접촬영	200,000클립	키포인트 가공 200,000 * 30FPS * 동영상재생시간 만큼의 json	SENXXXX_REALXX 5개 각도의 동영상 5개(.mp4) + 키포인트 가공 30fps*동영상재생시간*5각도 json files + 형태소/비수지 가공 5각도 json files + 캘리브레이션 정보
		형태소/비수지 가공 200,000 json	
수어문장1000_가상데이터	5,000클립	키포인트 가공 5,000 * 30FPS * 동영상재생시간 만큼의 json	SENXXXX_SYNXX 5개 각도의 동영상 5개(.mp4) + 키포인트 가공 30fps*동영상재생시간*5각도 json files + 형태소/비수지 가공 5각도 json files
		형태소/비수지 가공 5,000 json	
수어단어3000_직접촬영	300,000클립	키포인트 가공 300,000 * 30FPS * 동영상재생시간 만큼의 json	WORDXXXX_REALXX 5개 각도의 동영상 5개(.mp4) + 키포인트 가공 30fps*동영상재생시간*5각도 json files + 형태소/비수지 가공 5각도 json files + 캘리브레이션 정보
		형태소/비수지 가공 300,000 json	
수어단어2000_가상데이터	10,000클립	키포인트 가공 10,000 * 30FPS * 동영상재생시간 만큼의 json	WORDXXXX_SYNXX 5개 각도의 동영상 5개(.mp4) + 키포인트 가공 30fps*동영상재생시간*5각도 json files + 형태소/비수지 가공 5각도 json files
		형태소/비수지 가공 10,000 json	
지수어1000_크라우드소싱	21,000클립	키포인트 가공 21,000 * 30FPS * 동영상재생시간 만큼의 json	FSXXXX_CROWDXX 동영상 1개(.mp4) + 키포인트 가공 30fps*동영상재생시간*1각도 json files + 형태소/비수지 가공 F각도(1) json files
		형태소/비수지 가공 21,000 json	
총합	536,000클립		

○ 데이터 분포

- 구축 데이터 영상 클립 수: 536,000개
- 직접촬영 언어 제공자 성별 및 연령대 인원 및 비율

성별 인원(명)	
남자	여자
15	25

연령대 인원(명)			
20대	30대	40대	50대
9	13	16	2

