베이지안 분류를 이용한 입 모양 기반 발음 인식 방법

(특허등록번호 : 10-2405570)

키워드 : 입모양 인식, 베이지안 분류, 비음성 기반 음성인식, 저연산 AI, 실시간 영상 분석, 음성 대체 인터페이스

사업화 적용 가능 분야

① 소음 환경 대응 음성 인식 솔루션
공장, 지하철, 야외 현장 등 음성 인식이 어려운 환경에서 카메라 기반 발음 인식 인터페이스 구현 가능

② 청각장애인 보조 커뮤니케이션 시스템
입 모양 기반 자막 자동 생성, 무음 커뮤니케이션 시스템, 보조 의사소통 기기 적용 가능

③ 엣지 디바이스·저사양 기기 AI 탑재
딥러닝 대비 연산 복잡도가 낮아 키오스크, 모바일, 스마트글래스, 웨어러블 기기 등 경량 디바이스에 최적화

④ AI자막·영상 콘텐츠 자동 분석
실시간 자막 생성, 교육 플랫폼, 영상회의 시스템, 보안 감시 영상 분석 등 다양한 B2B 확장 가능

기술 성숙도(TRL) – 6~7단계 (시제품 실증 완료 단계)

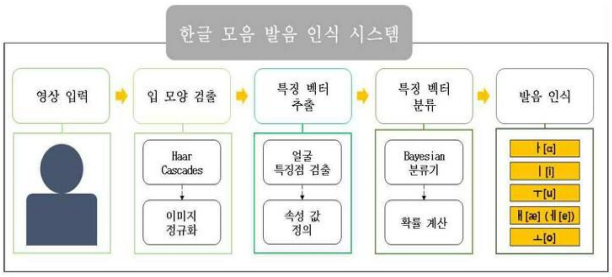
- 500개 영상 데이터 기반 실험 수행 완료
- 화자 종속 최대 94%, 전체 평균 약 85~93% 인식률 확보
- 실시간 30초 영상 환경에서 발음 인식 검증 완료

▶ **연구실 수준을 넘어 실제 환경에서 적용 가능한 단계**

기술 개요

- **영상 정보만을 활용하여 사람의 입 모양을 분석하고, 입의 가로·세로 길이(Mx, My), 입술 두께(La, Lb), 볼 및 턱 거리(Fx, Fy) 등 6개 특징 벡터를 정의하여 베이지안 분류기로 한글 모음(ㅏ, ㅣ, ㅜ, ㅐ, ㅗ)을 인식하는 기술임**

- **Harr-Cascades 기반 얼굴 검출과 이미지 정규화 기법을 통해 얼굴 크기와 상관없이 일정한 특징값을 추출하고, 확률 기반 모델을 통해 발음을 분류함**



[한글 모음 발음 인식방법 개념도]

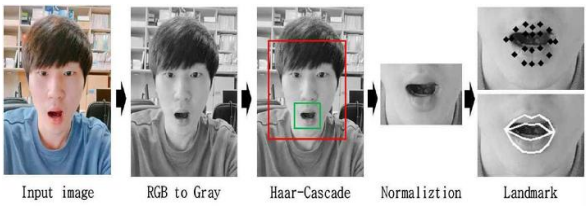
기술 특장점

① **음성 정보 없이 영상으로만 발음 인식 가능**
▶ 기존 시스템은 음성+영상 결합 또는 CNN/SVN 기반 고연산 구조, 본 기술은 영상 단독+확률 기반 분류 구조

② **저연산·고효율 구조**
▶ 딥러닝 대비 낮은 복잡도, 고사양 CPU 불필요
적은 학습 데이터로도 높은 성능 확보

③ **경량 학습 데이터 요구**
▶ CNN 대비 소량 데이터 환경에서 더 높은 인식률 확보

④ **실시간 처리 가능**
▶ 정적 이미지뿐 아니라 실시간 영상에서도 발음 추출 가능



[얼굴 영역 검출과 랜드마크 부여 방법]

※ 기대효과

1. 고성능 서버·GPU 의존도 낮음 → SaaS 및 임베디드 제품 용이하여 **저비용 AI 솔루션 상용화 가능**
2. 청각 보조시장, 무소음 인터페이스 시장, 스마트 헬스케어·보조 공학 시장 등 **글로벌 특수 시장 공략 가능**
3. 현재 모음 5개에서 향후 자음, 단어, 문장 인식 모델로 확장이 가능하여 **영상 기반 AI 엔진 확장성 보유**
4. 교육 플랫폼, 영상회의 시스템, 자막 생성 솔루션, 스마트 디바이스 제조사 등 **B2B 라이선스 모델 구축 가능**

연구자 | 차경애 교수
대구대학교 AI학과
chaka@daegu.ac.kr

RISE
대구대학교RISE사업단