궤도로봇의 직관적 제어를 위한 음성-제스쳐 인식 모델

Multimodal Intelligence and Interaction Laboratory@Hanyang University, Republic of Korea
김태현, 정재준, 남서용, 지도교수 유용재

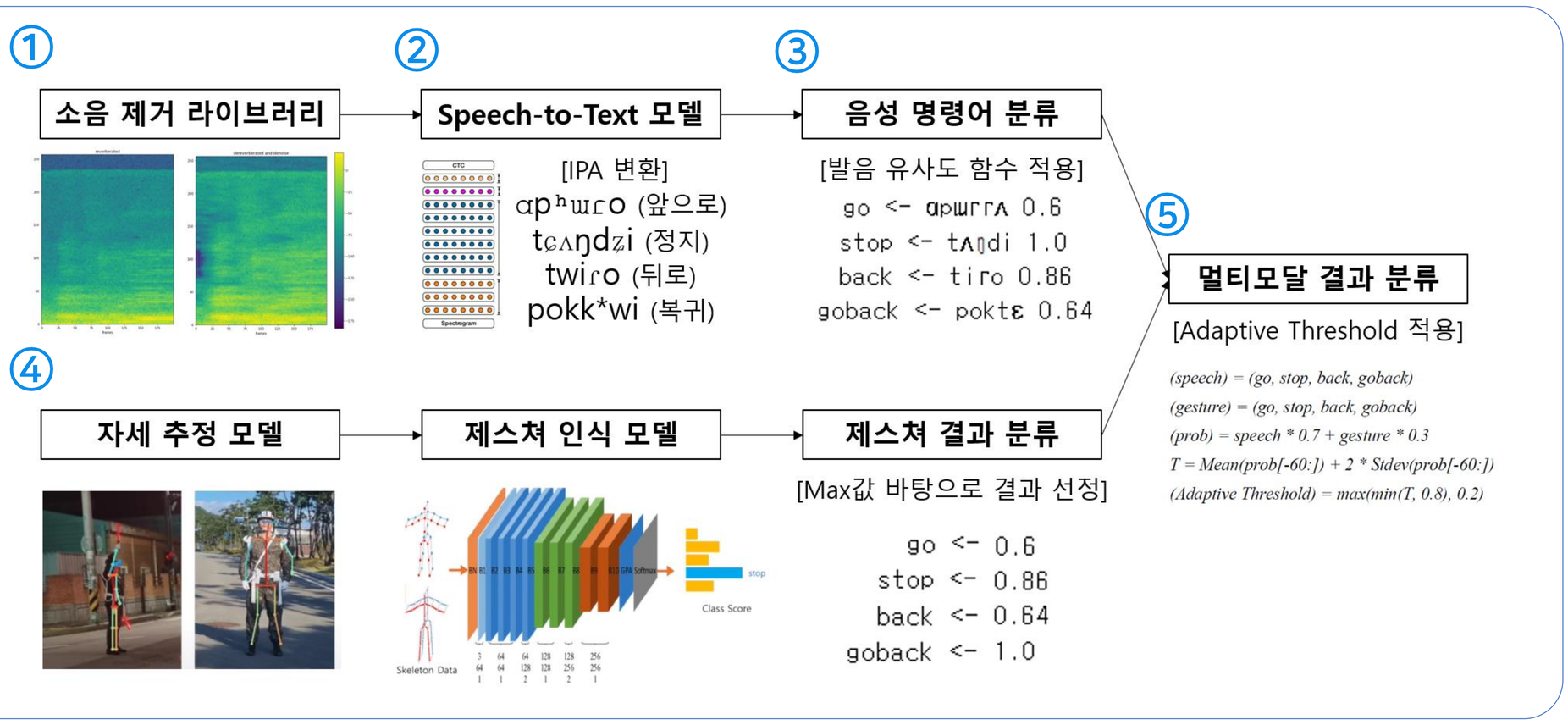
Introduction

- 사회 기반 시설물 점검을 위한 로봇의 도입이 계속 이루어지고 있지만, 실제 현장에서의 저조도 및 소음 환경에 의한 인식의 어려움이 있었음.
- 사용자의 의도를 적절하게 파악하기 위하여, 제스쳐 및 음성신호를 바탕으로 멀티모달 인공지능 알고리즘을 구성하고자 하였음.
- 극한 상황에서의 강건한 인식을 위하여 데이터 증강 전략을 적용하였음.

System Pipeline

- ① 정적인 소음을 제거하는 Stationary Noise Reduction (noisereduce) 적용 및 소음 대비 음성 강조
- ② 국제발음기호(IPA)로 라벨 값 재구성 및 Deep Speech 2 기반 STT(Speech-to-Text) 모델 학습 진행
  - 소음 환경 데이터 학습 및 정적 소음 제거 전처리 증강 시도
- ③ Levenshtein Distance 를 활용한 음성 유사도 기반 분류 알고리즘 구현
  - 자음 조음 위치 및 자음 조음 방법에 대한 Deletion 연산
  - 조음 위치, 조음 방법, 조음 강도, 유성음 여부, 입술 모양에 대한 Substitution 연산 진행

  위 알고리즘 적용을 바탕으로 직접 녹음한 데이터셋에 대하여 약 93% 의 정확도를 보였다.



- ④ 자세 추정 모델을 바탕으로 스켈레톤 포인트를 포착 및 시계열 데이터를 제스쳐 인식 모델 바탕으로 분류
  - 밝은 환경, 저조도 환경 등 다양한 수신호 데이터셋으로 학습
- ⑤ 음성 인식과 제스쳐 인식의 이전 결과 값들에 대한 평균과 표준편차 바탕으로 Adaptive Threshold 를 구성
  - 큐(Queue) 에 들어온 결과 값이 위 Threshold 를 넘기면 명령을 실시간으로 인식하여 작동하도록 구현

Conclusion

- 소음 및 저조도 환경에서의 데이터셋을 이용, 실험 및 성능 평가를 수행한 결과, 높은 인식률 및 실시간에 근접한 짧은 계산 시간을 보여주었음.
- 향후 실 작업 환경에서의 데이터셋으로 Training 하여 더 높은 성능을 달성하고자 함.

HANYANG UNIVERSITY