

골격 정보 기반 수어의 수동 단위 시간 분할

이원재, 이한규
한국전자통신연구원

russell@etri.re.kr, hkl@etri.re.kr

Temporal Segmentation of Movement in Sign Language using Skeleton Information

Wonjae Lee, Han-kyu Lee
Electronics and Telecommunications Research Institute

요 약

본 논문은 골격 정보 기반 수어의 수동 단위 시간 분할 시스템에 대해 논의한다. Azure Kinect Body Tracking SDK 를 사용하여 골격 정보를 얻었으며, 우세손의 손끝 속도 변화를 분석하여 시간 분할을 수행하였다. 수어 단어의 정지 분절에 대응하는 구간을 검출하기 위해 속도 그래프의 극소점을 파악하였다. 시간 분할 결과를 평가하기 위해 3 차원 애니메이션으로 결과를 시각화 하였다.

I. 서론

수어는 농인(deaf)과의 의사소통에 필수적인 언어이다. 수어를 구사할 수 있는 청인(聽人, a hearing person)이 소수이기 때문에, 농인이 의료기관을 방문한 상황 같은 경우 수어 통역사가 필요하다. 하지만 수어 통역사의 수가 한정되어 있고 비용 문제로 인해 청인과의 소통이 필요한 모든 상황에서 수어 통역사를 대동하기는 힘들다.

최근 심층 학습(deep learning) 기술의 급속한 발전으로 수어 인식 관련 많은 성과가 있었다. 본 논문에서는 골격 정보에 기반한 수어의 수동(손의 움직임) 단위 시간 분할 시스템에 대해 기술한다. 개발된 시스템은 심층 학습을 위한 수어 동영상 주석(annotation) 작업, 수어 인식 등에 활용될 수 있다.

관련 선행 연구로는 Kinect 이용 손 모양과 제스처 인식 프레임워크 [1], 관절을 통한 학습을 위한 동영상 자막 기반 연속 동작 이해 [2]가 있다.

II. 본론

본 연구에서는 선행 연구 [1][2]와 일부 유사한 방식으로 손의 속도 변화를 분석하여 시간 분할(temporal segmentation)을 수행한다. Azure Kinect Body Tracking SDK 를 사용하여 각 프레임의 골격 정보를 얻고, 이를 통해 각 프레임 별 손끝의 위치 정보를 파악한다. 프레임 간 손끝의 이동거리, 즉 속도를 계산하고 Gaussian filter 를 적용한다. 손끝 속도의 극소점들(local minima)을 파악한다.

1. 손끝 속도 계산

수어에서 수동은 크게 경로 수동과 내부 수동으로 나뉜다 [3]. 경로 수동은 수형의 변화 없이 수위(손의 위치)만 변하는 수동을 일컫는다. 내부 수동은 손 위치의 변화 없이 손목과 손가락 관절을 움직여 수형과 수향을 전환하는 것을 말한다. 수동소가 경로 수동 또는 내부 수동 중 하나로만 이루어져 있으면 단순 수동, 경로 수동과 내부 수동 두 성분 모두를 갖고 있으면 복합 수동이 된다.

Azure Kinect Body Tracking SDK 는 손, 손목, 손끝, 엄지 손가락과 같은 수어 인식과 관련된 골격 정보를 제공한다. 본 연구에서는 주로 손목과 손끝 위치 정보로 실험을 진행했으며, 경로 수동과 내부 수동을 모두 인식하기 위해 손끝을 기준으로 시간 분할을 수행하였다.

수어는 양손을 사용하므로 어떠한 조합으로 분석해야 하는 지에 대한 문제가 있다. 양손 손끝의 속도를 더하여 분석한 경우와 우세손 손끝 속도에 대해서만 분석한 경우를 비교해보니, 많은 경우 검출되는 주요 극소점들은 유사했고 일부 사례에서 우세손에 대해서만 분석한 경우 추가적인 극소점들을 검출했다. 이에 따라 우세손 손끝의 속도만을 분석하기로 결정하였다.

시간 분할 시스템은 왼손과 오른손 중 어느 손이 우세손인지에 대한 입력을 받을 수도 있고, 분석 구간에 대해 각 손의 총 이동거리를 계산하여 이동거리가 더 긴 손을 우세손으로 추정하는 기능도 제공한다.

시스템에서는 각 프레임 간 우세손 손끝의 이동거리를 계산하고, 속도의 변화를 분석하여 시간 분할을 수행한다.

2. 극소점 파악 및 필터링

리들 앤 존슨(Liddell & Johnson)의 움직임-정지(movement-hold) 모델은 수어 단어의 동시적이면서도 순차적인 구조를 잘 설명하고 있다 [3].

이 모델에 따르면 움직임과 정지의 순차적 구성으로 수어 단어가 이루어진다. 이에 따라 본 시스템에서는 정지 분절(hold)에 대응하는 손끝 속도의 극소점을 검출하여 시간 분할을 수행한다. 이때, 노이즈 등으로 인해 필터를 적용하지 않으면 지나치게 많은 극소점이 나타나므로 sigma 값 2의 Gaussian filter를 적용하였다.

실험 결과 검출된 극소점 중 상당수는 정지 분절에 대응하지 않는 것으로 추정되어 2 가지 방법으로 필터링을 수행했다.

첫번째, 분석 대상 구간의 최대 손끝 속도를 max_velocity 라고 할 때, 극소점에서의 속도가 $\text{max_velocity} \times \text{min_threshold}$ 와 같거나 작아야 유효한 극소점이다. min_threshold 값으로 0.35 를 사용했다.

두번째, 극소점과 인접한 극대점(local maximum)의 속도 값 차이가 $\text{max_velocity} \times \text{minor_threshold}$ 보다 작고, $\text{adj_local_threshold}$ 프레임 수 이하로 인접한 다른 극소점의 속도가 해당 극소점보다 작으면 무효한 극소점이다. 이는 사소한 극소점은 무효화하면서도 넓은 시간 구간에서 보았을 때 유효한 극소점은 유지하는 기법이다. minor_threshold , $\text{adj_local_threshold}$ 값으로 각각 0.1, 15 를 사용했다.

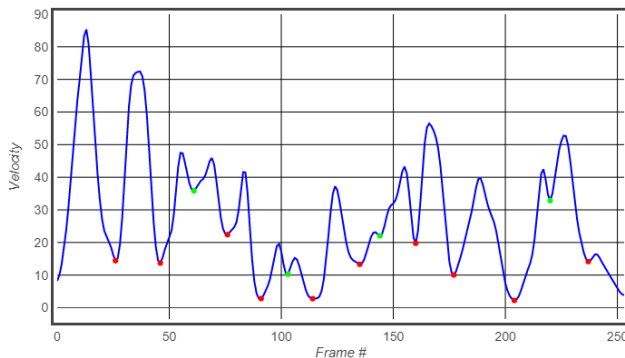


그림 1 ‘아침 일찍 산에 올라 산 아래 풍경을 바라보니 내 마음이 차분해졌다.’에 대응하는 수어 문장에 대한 시간 분할 예. 빨간색 점은 정지 분절에 대응하는 것으로 추정되는 유효한 극소점이며, 녹색점은 무효 극소점이다.

3. 시간 분할 결과 시각화

시간 분할 결과를 평가하기 위해 골격정보를 바탕으로 3 차원 인체 모형 애니메이션을 생성하고, 손끝의 궤적을 표출했다. 각 구간별 손끝 궤적의 색을 다르게 입히고, 현재 재생 중인 구간의 궤적은 불투명도를 0.7 로, 인접한 전후 구간 궤적은 불투명도를 0.3 으로, 다른 구간 궤적은 불투명도를 0 으로 설정하여 분할 결과의 시각적 확인을 도모하였다. 3 차원 그래픽 생성에는 VPython 을 사용하였다.

이러한 궤적 시각화 외에도, 분할 결과 구간별 정지 영상과 수어사전 상의 대표 스케치와의 비교를 통해서도 시간 분할 결과를 평가할 수 있을 것으로 예상된다.

4. 어깨와 팔꿈치 각도 실험

손목이나 손끝의 속도 분석 외에도 어깨와 팔꿈치 각도 변화를 분석하여 시간 분할을 수행하는 실험을 진행하였다. 이러한 각도 분석 기반 시간 분할은 사람에 따른 팔 길이 변화에도 일관된 결과를 도출할 것으로 예상하였다.

실험 결과 각도 분석을 통한 시간 분할은 원형 수동 외에는 손목 속도 분석을 통한 시간 분할 결과와

유사하였다. 손을 사용하여 일정속도로 원을 그릴 때 손끝/손목 속도 분석에서는 하나의 움직임 분절(movement)로 인식하나, 각도 분석에서는 여러 개의 움직임 분절과 정지 분절로 인식하는 단점이 있었다. 이는 신체 구조상 원형 수동 시 어깨와 팔꿈치가 일정한 각속도를 유지하지 못하기 때문으로 추정된다.

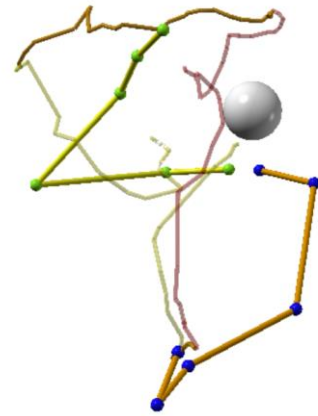


그림 2 시간 분할 결과 시각화 예. 흰색 공이 머리이고 오른팔 관절은 녹색 공으로, 왼팔 관절은 파란색 공으로 나타났다.

III. 결론

본 논문에서는 골격 정보 기반 수어의 수동 단위 시간 분할 시스템에 대해 기술하였다. 이 시스템에서는 경로 수동과 내부 수동을 인식하기 위해 우세손 손끝의 속도를 분석하여 극소점을 정지 분절로 추정한다. 이때, 2 가지 기준으로 극소점을 필터링하여 사소한 극소점에 대해서는 정지 분절로 인식하지 않는다. 분할 결과 평가를 위해 3 차원 애니메이션을 생성하였다.

ACKNOWLEDGMENT

이 논문은 2022 년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임. (2022-0-00010, 청각장애인을 위한 의료현장 양방향 수어 동시통역 서비스 기술 개발)

참 고 문 헌

- [1] Pedersoli F, Benini S, Adami N, Leonardi R, "XKin: An open source framework for hand pose and gesture recognition using Kinect," Vis Comput 30(10), 2014, pp. 1107- 1122.
- [2] Iori Yanokura, Naoki Wake, Kazuhiro Sasabuchi, Katsushi Ikeuchi, and Masayuki Inaba. "Understanding action sequences based on video captioning for learning-from-observation." arXiv preprint arXiv:2101.05061, 2020.
- [3] 원성욱, 김유미, 남기현, 김성완, 한국수어 문법, 국립국어원, 2021.