자폐 스펙트럼 장애의 손동작 분석 알고리즘 개발

성은서[1], 윤우승[1,2], 이석[3], 한경림[1,4]*

[1]한국과학기술연구원 뇌과학연구소 뇌융합기술연구단, [2]서울대학교 물리천문학부, [3]크리모,
[4]국가연구소대학교 KIST스쿨 바이오-메디컬 융합전공

esther98@kist.re.kr, riowoosung@kist.re.kr, slee@kist.re.kr, *khan@kist.re.kr

Development of a hand motion analysis algorithm for autism spectrum disorder

Eunsuh Sung[1], Wooseung Yoon[1,2], Seok Lee[3], Kyungreem Han[1,4]*

[1]Center for Brain Technology, Brain Science Institute, Korea Institute of Science and Technology, Seoul 02792, Korea, [2]Department of Physics & Astronomy, Seoul National University, Seoul 08826, Korea, [3]CREAMO Inc., Seoul 02792, Korea, [4]Division of Bio-Medical Science & Technology, KIST School, Korea National University of Science and Technology, Seoul 02792, Korea

요 약

의학적 관점에서 소근육 운동 조절 능력은 자폐 스펙트럼 장애를 포함한 신경 발달 장애 진단과 평가에 중요한 지표가 될 수 있다. 자폐 스펙트럼 장애 (ASD) 아동의 소근육 운동에 대한 정확한 평가는 장애 아동의 진단과 치료에 중요한 정보를 제공한다. 본 연구에서는 컴퓨터 비전 기술을 이용하여 정교한 손동작 평가를 통해 아동의 소근육 운동 발달 수준 및 특징을 추출하는 방법을 고찰한다. 기존 2차원 및 2.5차원 랜드마크 기반 손 자세 추정의 경우 손이 가려지거나 복잡한 자세에서 비논리적인 추정이 일어날 수 있다. 본 연구에서는 이를 극복하기 위해 메쉬 기반 추정을 통해 손의 공간적 특성을 잘 나타내도록 하고, 4차원 정보를 이용하여 연속성이 유지되도록 손동작 분석을 개선하였다. 실제 ASD 아동의 놀이 치료 과정 영상에 적용하였을 때 버텍스를 이용한 메쉬 기반 추정 기법이 기존 2.5차원 랜드마크 기반 손 자세 추정이나 3차원 파라메트릭 모델 기반 손 자세 추정 방식보다 주변 물체나 신체 부분과의 간섭이 적고 손의 표현력이 더 우수하게 나타났다. 이러한 정교한 분석 기법은 자폐 스펙트럼 아동의 진단, 예후예측, 반응예측 지표 개발 및 치료 프로그램 고도화에 기여할 것이다.

Ⅰ. 서 론

소근육 운동은 아동 발달 수준을 평가하는 발달 이정표(Developmental milestones) 중 하나이다. 소근육 운동 기술은 아동의 향후 지적 능력 및 자기돌봄역량과 연관성이 보고되었으며, 특히 자폐 스펙트럼 장애(Autism Spectrum Disorder, ASD) 등 신경발달장애(Neurodevelopmental Disorder)를 예측, 진단, 및 치료하는 데 큰 도움이 될 수 있다.[1] 자폐 스펙트럼 장애는 아동기에 발병하여 향후 환자의 생애에 지대한 영향을 끼친다. 자폐 아동의 운동 발달에서는 발달 지연과 상동행동이 나타날 수 있으며, 대근육 운동 발달 문제보다 소근육 운동 발달 문제가 더 자주 나타나기에 [2] 소근육 운동 기술을 객관적으로 평가하는 것은 중요하다.

본 연구에서는 중요한 소근육 조절 체계인 손 운동에 대한 객관적이고 수치화된 특징을 얻기 위해 3D 손 자세 추정 컴퓨터 비전 분석법을 제안한다. 인간 손의 경우는 물체나 손끼리 상호작용하는 경우가 많아 손이 가려지는 경우가 다수 발생한다. 따라서 손의 랜드마크를 추출할 때 비가시적인 랜드마크의 위치를 추정하는 것은 필수적이나 이 과정에서 다양한 오류가 발생한다.

손의 자세를 추정하는 방법은 크게 두 가지로 나눌 수 있다. 하나는 손의 관절 위치를 랜드마크로 지정하여 추정하는 방법[3, 4]이고, 나머지 하나는 손이 차지하는 공간 그 자체를 입체적으로 재구성하는 방법[5, 6]이다. 한편, 단일 RGB 이미지 기반의 2차원 랜드마크 손 자세 추정의 경우 손이 가려진 자세나 복잡한 자세에서 비논리적인 예측이 일어날 수 있다.[6] 기존 방법들은 멀티뷰 방식을 통해 비가시적 랜드마크에 대한 정보와 3차원 정보를 확보한다. 손을 3차원 객체로 재구성할 경우 멀티뷰 방식에서 손의 입체성을 고려할 수 있듯 손의 3차원 공간적 특성을 고려하기 때문에 물체와의 간섭을 줄일 수 있다. 또한 4차원 정보를 사용하면 비가시적 랜드마크 정보를 얻을 수 있다. 따라서 본 연구에서는 입체적 재구성 방법을 채택하고, 4차원 정보를 통해 2차원 관절 토큰을 개선한다. 그 후 개선된 토큰을 바탕으로 공간적 정보를 고려한 손의 메쉬를 재구성하는 전략을 취한다.

Ⅱ. 손동작 분석 모델

본 연구에서는 손의 이미지에서 임베딩을 추출한 후 4차원 정보를 반영하여 2차원 랜드마크를 추출하고 2차원 랜드마크를 바탕으로 손을 입체적 객체로 재구성하여 최종적으로 손의 3차원 랜드마크를 계산하고 이를 바탕으로 손의 운동 특성을 감지하였다.

1. 4차원 정보를 이용한 동작의 연속성 보정

손의 연속적인 움직임에서는 손 전체나 대부분이 가려지는 상황이 발생한다. 따라서 전 프레임들의 손과 후 프레임들의 손을 통해 4차원 정보를 고려하면 그 중간 프레임들의 손 자세를 추정할 수 있다. 따라서 본 연구

에서는 임베딩에 시간적인 정보를 반영한 후 손 자세를 추정한다.

이미지에서 임베딩을 추출하기 위해서 많은 손 자세 추정 기법들에서는 CNN을 사용한다[4, 6, 8]. 본 연구에서는 CNN 대신에 고속 처리와 경량화를 특징으로 하는 FastViT 백본을 이용하여 손 이미지에서 임베딩을 추출하였다. 추출한 임베딩에 시간적 정보를 반영하기 위해 [8]의 계층적 시계열 트랜스포머 기술에서 짧은 기간 동안의 시계열 정보를 이용한 손 자세 추정을 이용하였다.

2. 2차원 랜드마크 기반 손 자세 추정

입체적 재구성 방법에서도 2차원 랜드마크 기반 손 자세 추정은 필요하다. 입체적 재구성 방법에 백본 모델에서 나오는 임베딩을 바로 3차원 객체로 재구성할 경우 과적합 문제가 발생할 수 있고, 실제 사진 같은 제어되지 않고 많은 변수가 발생할 수 있는 환경에서 모델이 잘 작동하지 않을 수 있다. 계산 리소스가 많이 든다는 단점도 있는데, FastViT의 경우 경량화된 ResNet 모델인 ResNet-18 모델보다 많은 임베딩을 출력하여 2차원 랜드마크를 거치지 않고 바로 3차원으로 재구성할 시 계산 리소스의 문제가 현저하게 나타난다. 따라서 입체적 재구성에 앞서 시계열 트랜스포머에서 출력된 가공된 피쳐를 이용하여 손의 2차원 랜드마크를 추정한다.

3. 3차원 객체로 재구성

고차원 피쳐로부터 2차원 랜드마크를 추출한 이후에는 [5]의 방법을 이용해 2차원 랜드마크와 고차원 피쳐를 합쳐 관절 토큰을 생성하고, 토큰을 바탕으로 메쉬를 생성한다. 손의 메쉬 생성 방법에는 2가지 방법론이 존재한다. 하나는 파라메트릭 모델을 사용하는 것인데, 각 관절의 회전각으로 구성된 포즈 파라미터나 포즈의 차원을 축소한 PCA에 기반하여 손의 모양을 생성한다. 전자의 경우 공간적인 상관관계를 무시하여 불가능한 포즈가 발생하는 경우가 있어 추가 작업이 필요하며 후자는 손의 표현력이 줄어든다.

두번째는 버텍스 기반 방법으로 직접 메쉬를 생성하는 방식이다. 버텍스 기반 방법은 손의 표현력 제한으로부터 자유롭고 공간적인 상관관계를 고려하여 손을 3차원 객체로 재구성할 수 있다. 이 3차원 객체로부터 3D 랜드마크 좌표를 구할 수 있으며, 구해진 랜드마크 좌표 궤적에 대해 분류 모델을 적용할 수 있다. 손동작 분석 모델의 구조는 그림 1과 같다.

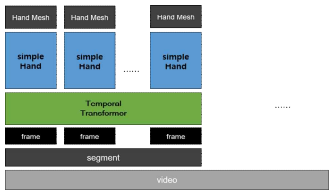

그림 1. ASD 손동작 분석 모델의 구조

Ⅲ. 모델의 임상적 적용

실제 촬영 데이터에 대해서 2.5차원 랜드마크 기반 추정법 (그림 2), 파라메트릭 모델 기반 추정법[9] (그림 3), 버텍스 기반 추정법 (그림 4)을 적용해 3차원 랜드마크를 표기한 결과는 다음과 같다.

각 추정법을 적용한 결과, 2.5 차원 랜드마크 모델의 경우 물체와 옆 손과 간섭을 보이는 오류가 발생하고 (그림 2), 3차원 파라메트릭 모델의 경우는 손가락 위치가 어긋나는 경우가 자주 나타난다 (그림 3). 버텍스 모델의 경우는 물체나 다른 손과의 간섭이 현저히 감소하며 (그림 4), 메쉬 모델을 적용하였을 때 파라메트릭 모델에 비해서 손의 입체적인 모양을 더 잘 반영하며 표현력이 우수하다(그림 5).

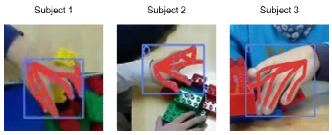

그림 2. ASD 아동 놀이 치료 영상의 2.5차원 랜드마크 모델 적용

그림 3. ASD 아동 놀이 치료 영상의 3차원 파라메트릭 모델 적용

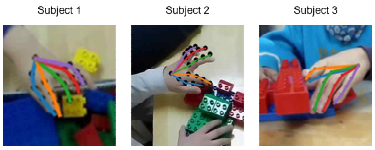

그림 4. ASD 아동 놀이 치료 영상의 버텍스 파라메트릭 모델 적용

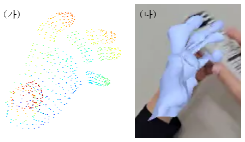

그림 5. 3차원 파라메트릭 (가) 및 버텍스 모델 (나)의 메쉬 모델 적용 예시

Ⅳ. 결론 및 전망

본 연구에서 제안하듯 3차원 버텍스 기반 모델을 사용하여 공간적 특성을 고려하면 2차원 랜드마크 기반으로 손 자세를 추정하는 경우보다 잘못된 손동작 추정과 물체와의 간섭이 현저히 감소하고 손의 표현력이 향상될 수 있다. 이를 자폐 아동의 소근육 운동 연구에 적용하면 더 객관적인 특징을 추출할 수 있으며, 분류 모델을 통해 분류하면 ASD의 자동화된 진단 및 평가 과정에 적용할 수 있다.

본 연구의 초기 모델을 DexYCB 데이터셋에 대해 학습시키고 MPJPE(Mean Per Joint Position Error) 및 MPVPE(Mean Per Vertex Position Error) 평가 지표로 타 모델과 비교한 결과이다. (표 1)

표 1. DexYCB 기반 정확도 평가

	HandOccNet	MobRecon	H2ONet	simpleHand	Ours
MPJPE	14.0	14.2	14.0	12.4	33.4
MPVPE	13.1	13.1	13.0	12.1	34.9

이러한 정교한 분석 기법과 평가 지표의 개발은 ASD 아동의 특징적인 소근육 운동 기술을 정상 발달 아동(Typical Development, TD) 및 기타 신경발달장애 아동과 비교하기 위한 기반이 될 수 있고, 궁극적으로 ASD 아동의 진단, 예후예측, 반응예측 지표 개발 및 치료 프로그램 고도화에 기여할 것이다.

ACKNOWLEDGMENT

이 연구는 과학기술정보통신부의 재원으로 한국지능정보사회진흥원의 지원을 받아 개발 중인 '장애인 소통 지원 초거대 AI 멀티모달 기반 서비스 개발 (기여율 50%)', 정보통신기획평가원의 지원 (No.2022-0-00375-001, 디지털 치료 활성화를 위한 XR트윈 핵심 기술 개발, 30%), 및 한국과학기술연구원 미래원천 뇌과학 기술개발사업 (2E32921, 고효율 예측 뇌 기능 모사 알고리즘 개발, 20%)의 지원을 받아 수행되었습니다.

참 고 문 헌

[1] Smits-Engelsman BC, Wilson PH, Westenberg Y, Duysens J. Fine motor

deficiencies in children with developmental coordination disorder and learning disabilities: an underlying open-loop control deficit. Hum Mov Sci. 2003 Nov

[2] Melo C, Ribeiro TP, Prior C, Gesta C, Martins V, Oliveira G, Temudo T. Motor stereotypies in autism spectrum disorder: Clinical randomized study and classification proposal. Autism. 2023 Feb

[3] Zhang, Fan, et al. "Mediapipe hands: On-device real-time hand tracking." arXiv preprint arXiv:2006.10214 (2020).

[4] Z. Cao, G. Hidalgo, T. Simon, S. -E. Wei and Y. Sheikh, "OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 43, no. 1, pp. 172-186, 1 Jan. 2021,

[5] Zhou, Zhishan, et al. "A Simple Baseline for Efficient Hand Mesh Reconstruction." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024.

[6] Zhang, Xiong, et al. "End-to-end hand mesh recovery from a monocular rgb image." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019.

[7] Liu, Ruicong, et al. "Single-to-Dual-View Adaptation for Egocentric 3D Hand Pose Estimation." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024.

[8] Wen, Yilin, et al. "Hierarchical temporal transformer for 3d hand pose estimation and action recognition from egocentric rgb videos." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2023.

[9] “EasyMocap“ https://github.com/zju3dv/EasyMocap