손목 부착형 웨어러블 RGB 카메라에 최적화된 손 자세 추정기술

이정호[O], 최창환[*], 민재은[**], 최용근[*], 최상일[*]
[O]단국대학교 컴퓨터공학과,
[*]단국대학교 컴퓨터공학과,
[**]단국대학교 응용컴퓨터공학과
e-mail: 72210297@dankook.ac.kr[O], ho03206@naver.com[*], min2ndid@naver.com[**],
{younggch, choisi}@dankook.ac.kr[*]

An Optimized Hand Pose Estimation in Wearable Wrist-Attached RGB Camera

Jeongho Lee[O], Changhwan Choi[*], Jaeeun Min[**], Younggeun Choi[*], Sang-Il Choi[*]
[O]Dept. of Computer Science and Engineering, Dankook University,
[*]Dept. of Computer Science and Engineering, Dankook University,
[**]Dept. of Applied Computer Science and Engineering, Dankook University

● 요 약 ●

본 논문에서는 손목 부착형 웨어러블(Wearable) RGB 카메라를 통해 취득한 손 이미지에 최적화된 손 자세 추정모델과 학습방법을 제안한다. 최근 의료분야에서 활발하게 인공지능이 사용되고 있으며 그 중 이미지 인식을 중심으로 하는 진단 분야[1]가 괄목할만한 성과를 보인다. 본 연구에서는 웨어러블 카메라를 통해 얻은 손 자세를 활용하여 질병 진단에 적용할 계획이다. 또한, 본 연구수행을 통해 질병진단에 필요한 데이터 측정비용 절감 및 개인 맞춤형 진단서비스를 제공할 것으로 기대된다.

키워드: 손 자세 추정(Hand Pose Estimation), 의료 AI(Medical AI), 웨어러블 기기(Wearable Device)

I. Introduction

고령화 시대에 접어들면서 양질의 헬스케어 서비스에 대한 관심이 늘어나고 있고 이에 인공지능을 도입한 헬스케어 시장 규모가 급성장하고 있다. 인공지능 헬스케어 기술은 다량의 데이터를 인간수준의 지능을 활용하여 정밀 진단 및 치료, 개인별 맞춤형 질병 예측 및 예방, 시공간의 제약이 없는 측정, 진료 등의 특징을 가진다. 본 연구에서는 웨어러블 카메라를 통해 촬영한 손 이미지를 분석하여 자세를 추정하고 추정한 데이터를 활용해 류마티스 관절염과 같은 질병을 진단 및 재활에 활용할 예정이다. 하지만 새롭게 웨어러블 카메라 손 이미지에 맞는 데이터셋을 제작하는 것은 큰 비용이 발생하게 된다. 그래서 본 논문에서는 기존 학습 데이터에 가상의 손 이미지를 추가하여 모델을 학습하는 방법과 손 자세 추정 딥러닝 모델을 제안한다. 또한, 본 연구의 영역을 수화, 증강현실(AR) 등으로 확장할 수 있을 것으로 기대된다.

II. Preliminaries

1. Related works

손 자세 추정[2]은 이미지나 영상 속에서 손 관절이 어떻게 구성되어 있고 해당 관절의 위치를 추정하는 문제이다. 기존의 자세 추정은 사람에게 센서와 같은 다양한 장비를 부착하여 실시간으로 정교하게 움직임을 파악할 수 있지만 높은 비용이 들어가며 한정된 영역에서만 가능한 방법이다. 최근 딥러닝을 자세 추정에 사용하며 고가의 센서 장비 없이 RGB 카메라로 찍은 사진을 통해 손 자세 추정을 할 수 있어 비용을 절감할 수 있고 응용 분야인 수화[3], 손짓 인식[4], 증강현실(AR)[4] 등으로 확대되고 있다.

III. The Proposed Scheme

1. 가상 이미지

웨어러블 카메라로 취득한 손 이미지는 [Fig. 3.]와 같이 일반적으로 사용하는 손 이미지와 다른 화각, 카메라 각도로 촬영이 되어있어

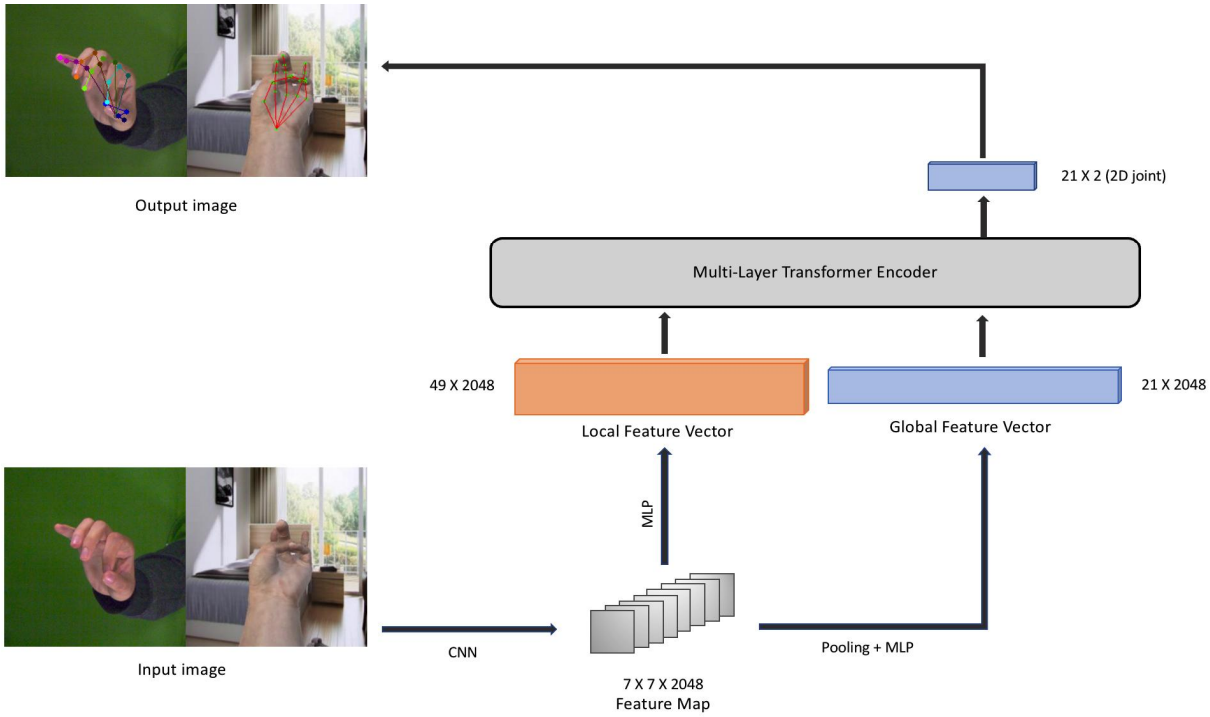

Fig. 1. 제안한 손 자세 추정 딥러닝 모델

모델을 학습하기 위해 새로운 손 이미지가 필요하다. 하지만 손 관절 21개를 라벨링 하는 것은 많은 비용이 발생한다. 본 연구에서는 Unity라는 프로그램을 통해 가상 손 이미지를 제작하여 웨어러블 카메라로 찍은 손 이미지와 유사하게 제작하였으며 2차원 관절 좌표뿐만 아니라 3차원 관절 좌표도 손쉽게 얻을 수 있다. 제작한 가상 이미지는 [Fig. 4.]와 같다.

Fig. 2. 손목 부착형 웨어러블 카메라

Fig. 3. **왼쪽:** 일반적인 손 이미지. **오른쪽:** 웨어러블 카메라로 촬영한 손 이미지

Fig. 4. 프로그램으로 제작한 가상의 손 이미지

2. 데이터셋 구성

손 자세 추정 데이터셋으로 주로 사용되는 FreiHAND[5]와 제작한 가상 손 이미지를 적정 비율로 섞어 학습 데이터셋을 구성하였다. 기존 데이터셋에 가상의 이미지를 추가해줌으로써 라벨링 비용을 절감하였고 동작하기 어려운 이미지도 제작이 가능하였다.

3. 딥러닝 모델

본 논문에서는 인코더 3개를 이어 붙여서 만든 다층 순차적 인코더 모델[6]을 제안한다. [Fig. 1.]을 살펴보면, 인코더의 입력토큰을 만들기 위해 백본 네트워크인 HRNet[7]에서 피처맵을 추출해준다. 그리고 추출한 피처맵을 두 가지 방식으로 임베딩을 하여 입력 토큰으로 만들어준다. 두 가지 방식의 임베딩은 다음과 같다. 첫 번째, 지역 특징 벡터(Local Feature Vector)는 7x7 크기의 피처맵을 평탄화하여 49개의 벡터로 만들어주면서 손의 부분적인 정보를 가지고 있다. 두 번째, 전역 특징 벡터(Global Feature Vector)는 피처맵을 전역 평균 풀링(Global Average Pooling)을 하여 7x7 크기의 피처맵을

1개의 벡터로 만들어주면서 전체적인 손의 정보를 가지고 있다. 1개의 벡터인 전역 특징 벡터를 MLP에 넣어 우리가 구하고자 하는 손 관절 개수에 해당하는 21개의 벡터를 얻은 후 위 두 가지 벡터를 함께 모델의 입력으로 넣어주게 된다. 마지막으로 2차원 손 관절 좌표 21개의 x, y 좌표가 모델의 출력으로 나오게 된다.

IV. Experimental Results

본 논문에서 제안하는 방법들을 검증하기 위해 3가지 종류의 실험을 진행하였다. 비교 실험을 (3) 가상 이미지 추가, (4) 가상 이미지 비율, (5) 다른 모델과의 성능 비교로 나누어 소개하고자 한다.

1. 실험환경

본 연구에서는 분석을 위해 Python 기반의 Pytorch 프레임워크를 사용했으며, Ubuntu 18.04 운영체제와 GeForce RTX 3090 GPU를 사용했다. 각 실험의 검증을 위해 웨어러블 카메라로 촬영한 1,000장을 동일하게 검증 데이터로 사용했으며 모델은 총 50 epoch 학습을 하였다.

2. 평가지표

본 논문에서는 모델의 성능을 검증하기 위해 자세 추정의 평가지표로 주로 사용되는 MPJPE를 사용하였다. MPJPE는 모든 관절의 추정 좌표와 정답 좌표의 거리(단위 : mm)를 평균하여 산출되는 지표이며 값이 낮을수록 좋은 성능을 보인다. 계산하는 수식은 다음과 같다.

$$MPJPE = \frac{1}{T}\frac{1}{N}\sum_{t=1}^{T}\sum_{i=1}^{N}\left\|(J_i^{(t)} - J_{wrist}^{(t)}) - (J_i^{\hat{(t)}} - J_{wrist}^{\hat{(t)}})\right\|_2$$

3. 가상 이미지 추가에 따른 성능 비교 실험

본 논문에서 제안하는 가상 이미지 추가에 따른 성능향상를 검증하기 위해 데이터셋을 3가지로 구성하였다.가상 이미지 2만장과 FreiHAND는 약 13만장을 가지고 비교 실험한 결과는 아래의 [Table 1]과 같다.

Table 1. 가상 이미지 추가에 따른 비교 실험

데이터셋 구성	MPJPE(mm)
가상 이미지로만 구성	33.60
FreiHAND로만 구성	4.73
FreiHAND + 가상 이미지 구성	**2.92**

4. 가상 손 이미지 비율에 따른 비교 실험

학습 데이터로 FreiHAND 약 13만장과 제작한 가상 이미지 2, 4, 6만장을 각각 섞어 총 15, 17, 19만장의 학습 데이터를 제작하였고 각 학습 데이터로 모델을 학습시켰다. 비교 실험한 결과는 [Table 2]과 같다.

Table 2. 가상 이미지 비율에 따른 모델 성능 비교 실험

가상 이미지의 비율	MPJPE(mm)
15%, 20,000장	**2.92**
30%, 40,000장	3.07
45%, 60,000장	4.84

5. 다른 손 자세 추정 모델과 비교 실험

손 자세 추정 모델로 주로 사용하는 OpenPose[8], Mediapipe[9], MMPose[10]를 가지고 손목 부착형 카메라에서 촬영한 손 이미지에서의 성능 비교를 진행하였다. 비교 실험 결과는 [Table 3]와 같다.

Table 3. 다른 손 자세 추정 모델과의 성능 비교 실험

손 자세 추정 모델	MPJPE(mm)
OpenPose (2018)	16.46
MMPose (2019)	6.43
MediaPipe (2019)	5.75
Ours	**2.92**

V. Conclusions

본 논문에서는 손목 부착형 웨어러블 카메라의 학습을 위한 새롭게 이미지를 촬영하여 만드는 대신 프로그램을 사용하여 가상 이미지를 제작함으로써 비용 절감과 더불어 기존 가지고 있던 데이터셋에 가상 이미지를 추가하여 본 연구의 테스트에서 성능향상을 보여줬다. 본 연구의 응용범위를 수화나 증강현실(AR) 등으로 나아갈 것으로 기대된다.

ACKNOWLEDGEMENT

이 논문은 2022년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원 (IITP-2022-00155227, 문맥정보를 이용한 딥러닝 기반의 의료 진단에 활용 가능한 ICT-BIO 융합 기술 개발)과 2021년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임(IITP-2017-0-00091, 멀티 모달 딥러닝 기반의 바이오 헬스케어 데이터 분석 기술 개발/ IITP-2021-0-01061, 멀티 모달 딥러닝 모델 기반 한국수어 통역 시스템 개발)

REFERENCES

[1] G.Litjens et al., “A survey on deep learning in medical image analysis”, in *Medical Image Analysis, vol. 32,* 2017.

[2] W. Chen et al., “A Survey on Hand Pose Estimation with Wearable Sensors and Computer-Vision-Based Methods”, in *Sensors, vol. 20, no. 4*, 2020.

[3] M. Jun et al., "Development of Korean Sign Language Production Avatars with Transformers", in *Korean Institute of Next Generation Computing*, 2022.

[4] L. Yong et al., "Hand Gesture Recognition in the Virtual Space based on Deep Learning", in *Journal of Digital Contents Society Vol.21, No.3, pp. 471-478*, 2020.

[5] Zimmermann, Christian, et al. "Freihand: A dataset for markerless capture of hand pose and shape from single rgb images." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019.

[6] Devlin, Jacob, et al. "Bert: Pre-training of deep bidirectional transformers for language understanding." arXiv preprint arXiv:1810.04805 (2018).

[7] Sun, Ke, et al. "Deep high-resolution representation learning for human pose estimation." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019.

[8] Cao, Zhe, et al. "Realtime multi-person 2d pose estimation using part affinity fields." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.

[9] Lugaresi, Camillo, et al. "Mediapipe: A framework for building perception pipelines." arXiv preprint arXiv:1906.08172 (2019).

[10] Chen, Kai, et al. "MMDetection: Open mmlab detection toolbox and benchmark." arXiv preprint arXiv:1906.07155 (2019).