

Cascades 방법을 이용한 합성곱 신경망 기반 사용자 동작 추정 방법 연구

최룡, 지수미, 성연식*

동국대학교 멀티미디어공학과

choi.si@mme.dongguk.edu, sumi@dongguk.edu, sung@dongguk.edu

Cascades of CNN-Based Human Pose Estimation Method Study

Ryong Choi, Sumi Ji, Yunsick Sung*

Dept. of Multimedia Engineering, Dongguk University-Seoul, South Korea

요 약

사용자 동작 추정이란 이미지 또는 비디오에서 사용자의 관절 위치를 추정하는 과정을 말한다. 기존의 연구들은 사용자의 몸에서 관절의 큰 부분(어깨, 무릎, 골반, 손, 발 등)만을 추정하거나 손의 세부 관절을 별도로 추정 했다. 하지만 특정 분야(수화, 댄스 등)에선 몸짓과 손을 함께 사용하기에 우리는 사용자 몸의 큰 관절과 손의 세부 관절을 같이 추정하는 방법에 대한 연구를 제안한다. 본 논문에서 제안하는 사용자 동작 추정 방법은 Cascades 방법을 이용한 합성곱 신경망 기반 회귀모델을 적용한 방식이다. 손의 관절들은 다른 큰 관절들(어깨, 무릎, 골반 등)보다 작아서 정밀한 추정을 요구하기에 Cascades 방법을 사용해 보다 정밀하게 추정할 수 있다.

1. 연구 필요성 및 문제점

많은 사람들이 여가 시간을 통해 다양한 취미 활동을 하고 있다. 그 중에도 몸을 사용하는 스포츠, 댄스, 운동 등은 일반인들이 전문가들의 수준만큼 도달하기 쉽지 않아, 전문가에게 직접 교육을 받는 등 많은 시간과 비용을 투자한다. 최근 딥러닝 기술의 발전으로 딥러닝 기반 사용자 동작 추정 기술을 통해 일반인들이 학습된 전문가의 동작을 보고 따라하는 플랫폼들이 개발되고 있다. 이는 비대면 방식으로 시간과 장소에 제약 없이 이용할 수 있는 방식으로 많은 사람들의 관심을 끌고 있다.

사용자 동작 추정이란 이미지에서 사용자의 동작을 추론하는 과정에 필요한 기술을 포함하며, 자율주행, 스포츠, 게임, 증강현실 등 다양한 분야에서 움직임 인식에 활용되고 있다. 기본적으로 이미지 또는 비디오에서 사용자의 관절 위치(어깨, 무릎, 허리, 손목, 팔꿈치, 손, 발 등)를 추정하는 작업이 수반된다.

사용자 몸의 큰 관절(머리, 어깨, 팔꿈치, 손, 골반, 무릎, 발)을 추정하는 방법[1-2]이나 손의 세부 관절을 추정하는 방법[3]에 대한 연구는 지속적으로 진행되어 왔지만 몸의 큰 관절과 손의 세부 관절을 같이

추정하는 연구[4]는 거의 없어왔다.

수화는 손의 움직임과 비수지 신호(얼굴표정과 몸짓)를 통한 시각적인 언어로 농인들의 일차언어이다. 수화의 동작을 추정하려면 손의 세부 관절과 몸짓을 같이 추정해야한다. 댄스 동작도 마찬가지로 몸의 동작도 중요하지만 손의 섬세한 표현도 중요하다.

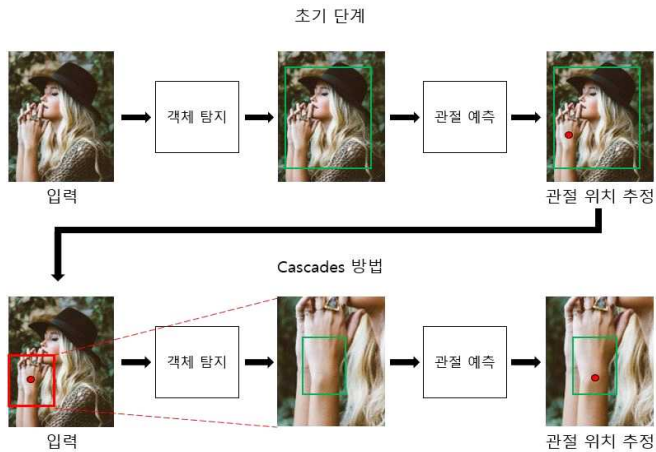
따라서 본 논문에서는 사용자 동작 추정분야에서 사용자 몸의 큰 관절과 손의 세부 관절을 함께 추정하는 방법을 제안한다. 이는 Cascades 방법을 이용한 합성곱 신경망 기반 회귀모델을 적용한 방식으로, 이를 통해 사용자 동작 추정이 가능하다.

2. 연구내용과 방법

일반적으로 사용자 동작 추정 방법에서 큰 분류는 Top-Down, Bottom-Up 방식 두 가지로 나뉜다. Top-Down 방식[1-2]은 이미지에서 사용자에게 해당하는 부분을 탐지한 후 신체의 동작을 추정하는 방식이고 Bottom-Up 방식[5]은 신체 주요 부위의 관절을 우선 검출한 후 사용자의 신체에 표시하는 방식이다. Bottom-Up 방식은 탐지과정이 생략되기 때문에 검출 속도가 빠르지만 Top-Down 방식에 비해 정확도가 낮다. 우리가 제안하는 방법은 작은 손가락 관

*교신저자 : 성연식(sung@dongguk.edu)

절도 찾아내야하기 때문에 정확도가 높은 Top-Down 방식을 사용한다.



(그림 1) Cascades 방법을 이용한 합성곱 신경망 기반 사용자 동작 추정 방법의 전체 동작 과정

그림 1은 본 논문에서 제안하는 방법의 동작과정을 보여준다. 첫 입력 이미지는 고정된 크기로 조정되어 입력된다. 하지만 전체 이미지에서 손가락의 관절을 찾는 과정을 수행하면 광범위한 이미지에서 작은 손가락 관절을 찾아내야하기 때문에 정확하지 않을 수 있다. 따라서 세밀한 관절 추정을 위한 Cascades 방법을 적용하고, 관절들의 상관관계 학습을 위한 합성곱 신경망을 통한 움직임 추정 방법을 제안한다.

Cascades 방법은 처음 입력 이미지에 대해 관절을 추정한 후 추정된 관절 중심으로 바운딩 박스를 그려 해당 부분의 이미지를 추출하고 그 다음 추출된 이미지의 크기를 기존 입력 이미지 크기와 동일하게 재조정 후 재입력하여 다시 관절을 추정한다. 이러한 과정을 반복해 관절의 위치를 추정해야하는 범위를 점점 축소하여 추정함으로써, 손가락 같이 추정하기 어려운 작은 관절이어도 위치를 보다 정밀하게 추정 해낼 수 있다.

더불어 합성곱 신경망을 사용하여 관절들 사이의 상관관계를 학습한다. 이를 통해 각 관절들의 전체적인 문맥을 고려하기 때문에 이미지에서 관절이 사물에 가려지거나 몸에 가려져 보이지 않더라도 관절 위치 추측이 가능함에 따라, 따로 관절들 사이의 상관관계를 정의하지 않아도 된다.

주어진 입력에 대해서 합성곱 신경망을 통해 관절이 있을 법한 위치를 탐지하고 이를 토대로 각 관절의 x, y 좌표를 찾아내는 합성곱 신경망 기반의 회귀모델을 학습시켜 이미지속의 모든 관절을 찾아내

최종적으로 사용자의 동작을 추정한다.

3. 결론 및 향후 연구

본 논문에선 Cascades 방법을 이용한 합성곱 신경망 기반 사용자 동작 추정 방법을 제안하였다. 사용자의 몸의 관절과 손의 세부 관절을 동시에 추정하기 위해 기존의 방법(손을 한 개의 부분으로 취급)으로는 동작을 추정할 수 없었던 수화나 댄스 동작도 추정할 수 있어 보다 광범위한 분야에서 유용하게 사용될 것이다.

제안하는 연구 방법을 기반으로 동작 추정 알고리즘을 개발하여, 제안하는 방법의 이를 추가로 보완하는 연구를 진행할 예정이다.

사사표기

"본 연구는 과학기술정보통신부 및 정보통신기획평가원의 글로벌핵심인재양성지원사업의 연구결과로 수행되었음" (2020-0-01576)

참고문헌

- [1] Alexander Toshev, Christian Szegedy, "DeepPose: Human Pose Estimation via Deep Neural Networks," IEEE Conference on Computer Vision and Pattern Recognition(CVPR), Columbus, USA, 2014.
- [2] Shih-En Wei, Varun Ramakrishna, Takeo Kanade, Yaser Sheikh, "Convolutional Pose Machines," IEEE Conference on Computer Vision and Pattern Recognition(CVPR), LAS VEGAS, USA, 2016.
- [3] Tomas Simon, Hanbyul Joo, Iain Matthews, Yaser Sheikh, "Hand Keypoint Detection in Single Images using Multiview Bootstrapping," IEEE Conference on Computer Vision and Pattern Recognition(CVPR), Honolulu, Hawaii, 2017.
- [4] Sang-Ki Ko, Changjo Kim, Hyedong Jung, Choongsang Cho, "Neural Sign Language Translation Based on Human Keypoint Estimation," Applied Sciences-Basel, vol.9, no.13, 2019.
- [5] Zhe Cao, Tomas Simon, Shih-En Wei, Yaser Sheikh, "Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields," IEEE Conference on Computer Vision and Pattern Recognition(CVPR), 2017.