

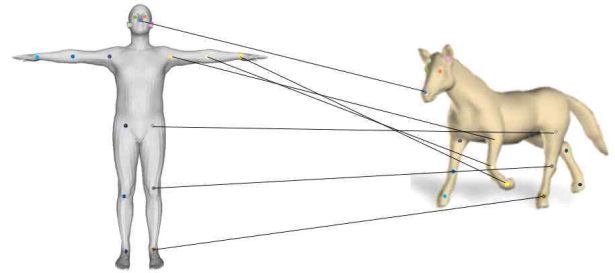
유사참값을 이용한 사람과 사족 동물의 삼차원 자세 추정

민지홍*, 최지훈*, 심성대*, 김태원**
국방과학연구소*, 위드로봇**

3d Pose Estimation of Human and Quadruped Mammals using Pseudo Ground Truth

Jihong Min*, Jihoon Choi*, Sungdae Sim*, Taewon Kim**
Agency for Defense Development*, Withrobot**

Abstract - 본 연구에서는 단안 영상으로부터 사람과 사족 보행 동물의 삼차원 자세를 추정하는 기법을 제안한다. 이를 위해 사람과 동물의 공통형상모델을 활용하여 삼차원 자세에 대한 유사참값을 만드는 기법을 제안하고, 제안된 유사참값을 이용하여 이차원 자세로부터 삼차원 자세를 추정하는 방법을 제안한다. 제안된 연구는 실제 확보한 데이터로부터 정성적으로 결과를 확인하였으며, 나아가 학습되지 않은 종의 동물에서도 삼차원 자세가 추정됨을 확인하였다.



<그림 1> 공통 형상 모델

1. 서 론

모션캡처장치, 영상 및 라이다 등 다양한 장치로부터 추정된 사람의 삼차원 자세는 주로 영화, 메타버스, 헬스케어, 치안, HCI 등 여러 분야에서 사용되고 있다. 자율주행 분야에서도 조급씩 사람의 삼차원 자세를 이용하려는 추세가 있다. 예를 들면 차량 전방의 자전거를 탄 사람의 손짓을 인식하여 차량의 속도를 줄이거나 경로를 변경할 수 있고, 주행로 주변 보행자의 행동을 예측하여 현재 지역 경로계획에 반영할 수 있다. 이러한 목적으로 올해 공개된 Waymo 데이터셋[1]에서는 도심 보행자의 키포인트 레이블이 추가되었다.

본 연구에서도 이와 같은 맥락으로 국방의 자율주행차량이 야지에서 주행 중에 만날 수 있는 사람, 동물의 자세를 추정하고 미래의 움직임을 예측하여 경로계획에 반영하고자 한다.

대부분 사람의 삼차원 자세를 측정하기 위해서는 실내에서 모션캡처 장비를 이용하고 있으며, 실외의 경우 사람 몸에 관성항법장치를 장착하여 드론과 같이 이용하는 연구[2]를 제외하고는 거의 유사참값[3]이나 그에 준하는 데이터를 만들어 활용한다.

동물의 자세 추정 연구와 관련하여 데이터 셋[4]들이 최근에 공개되고 있으나 레이블된 데이터양이 사람 데이터셋보다 매우 적다. 동물 자세에 대하여 레이블링을 하기 위해 많은 양의 레이블된 사람데이터, 적은 양의 레이블된 동물데이터 및 레이블이 안된 동물데이터를 가지고 domain adaptation[4][5]기법을 적용하는 연구들이 종종 있으나 이러한 연구조차 이차원에 관한 연구들이 대부분이다. 나아가 동물에 대한 삼차원 자세 데이터는 극히 드물고 삼차원 자세 추정연구 역시 사람 연구 대비 매우 적은 실정이다.

본 연구에서는 동물의 삼차원 자세 데이터가 없고, 사람과 동물의 영상 데이터량 크게 다른 상황에서 각 자세를 추정하기 위해 domain adaptation보다는 사람과 동물을 하나의 모델로 표현하는 공통 형상 모델[6][7] 연구결과를 활용한다. 공통 형상 모델을 이용하여 사람과 동물의 삼차원 자세 유사참값을 작성한 후, 공통 형상 모델로 표현된 이차원 자세로부터 삼차원 자세를 추정하는 방법을 제안한다.

2. 본 론

사람의 삼차원 모델(SMPL)[8]과 일부 동물의 삼차원 모델(SMAL)[9]을 이용하여 영상으로부터 삼차원 외형과 자세 등을 추정하는 연구들이 있으나, 사람, 동물 종류에 따라 많은 모델과 영상 데이터가 요구된다.

하지만 사람과 동물의 외형적인 유사점을 이용하여 공통형상 모델을 제안하여 네트워크의 크기를 줄이고 성능을 높이는 연구[6]가 있으며 본 연구에서는 이를 이용하여 사람과 동물의 삼차원 자세 유사참값을 작성하였다. 이를 통해 영상으로부터 삼차원 자세를 바로 추정할 수 있지만, 영상이 흐릿해지거나 객체의 윤곽이 잘 안 나오는 경우 성능이 잘 나오지 않는 단점이 존재한다.

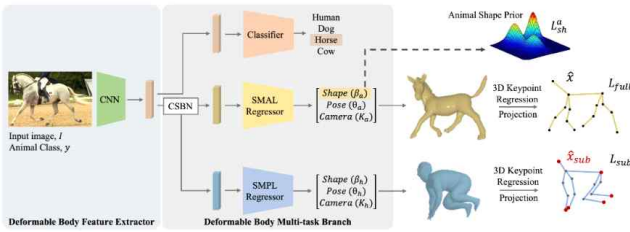
이와는 달리 키포인트 방식의 이차원 자세 추정은 영상의 왜란에 강인하여 더 나은 자세추정 결과를 만들어낸다. 이런 장점을 이용하여 이차원 자세로부터 삼차원 자세 추정하는 방법을 도입하였다. 따라서 이차원 키포인트와 공통형상모델로 작성된 삼차원 자세를 이용하여 자세추정의 차원을 올리는 모델을 학습시켰다.

2.1 공통형상모델

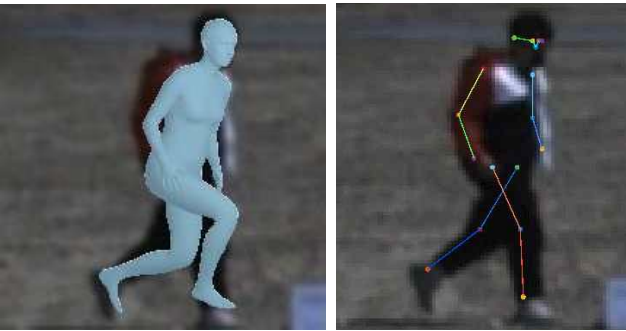
사람과 사족보행 동물은 그림 1과 같이 외형적으로 상당히 유사한 특징이 있으며 이를 기반으로 공통적인 조인트(키포인트)와 골격(트리구조)을 만든다. 삼차원 자세 유사 참값을 만들기 위해서는 10개의 키포인트 레이블을 주로 사용하였으며, 이차원 자세 추정을 위해서는 17개의 키포인트 레이블을 사용하였다.

2.2 삼차원 자세 유사 참값

삼차원 자세의 유사참값을 만들기 위해 공통형상모델 기반의 자세추정 모델[6]을 사용하였다. 그림 2는 [6] 논문에서 소개된 네트워크 구성도를 나타낸다. 네트워크는 기본적으로 SMPL과 SMAL모델을 주로 사용하고 있으며, 크게 세 파트로 구성된다. 첫 번째는 영상으로부터 특징을 추출하는 모델, 두 번째는 입력된 영상의 종을 재분류하는 세부 모델, SMAL의 외형과 자세를 추정하는 세부 모델, SMPL의 외형과 자세를 추정하는 세부 모델로 multitask branch로 구성되며, 마지막은 각 branch마다 삼차원 자세를 추정하는 모델로 구성된다.



<그림 2> 삼차원 자세 유사 참값 작성을 위한 공통 형상 모델 기반 삼차원추정 모델



<그림 3> (a) 외형 추정 결과 (b) 이차원 자세 추정결과

2.3 공통 형상 모델 기반 이차원 자세 추정

그림 3(a)과 같이 객체가 작거나 영상이 흐릿한 경우 외형추정이 어려우나 그림 3(b)과같이 이차원 자세 추정은 더 나은 성능을 보여준다. 하지만 각 종에 따른 모델이 별도로 필요한 단점이 발생할수 있다. 따라서 공통 형상 모델 기반의 사람과 동물을 같이 학습하여 하나의 네트워크에서 여러 종의 이차원 자세를 추정하는 기법[7]을 적용하였다. 공통형상모델 기반의 이차원 자세추정 기법에서는 종의 차이로 인해 발생하는 domain gap을 줄이기보다는 공통된 키포인트의 특징적인 다양성 즉 분산을 조금 더 넓히는 방향으로 학습하였다. 또한, 여러 이차원 키포인트 추정방식 중 top-down방식의 alphapose[10] 모델을 이용하였다. 이를 통해 하나의 네트워크로 사람 및 사족 동물의 이차원 자세 추정을 한번에 하고 사람의 경우 인식률도 조금 더 높이는 결과를 얻었다.

2.4 실험 결과

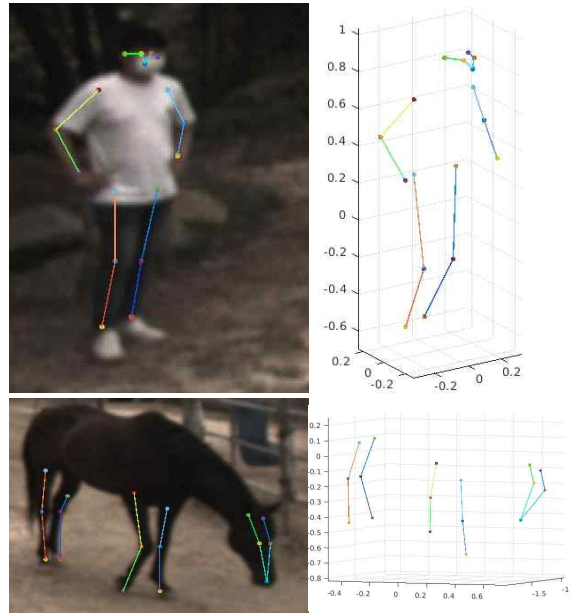
공통 형상 모델 기반 삼차원 자세 유사 참값을 만들기 위해 LSP, LSP-extended MS COCO, MPII, Stanford Extra, Animal Pose 데이터 셋에서 사람, 개, 소, 말의 클래스를 이용하였고, 이차원 자세 추정을 위해 MS COCO, Animal Pose 데이터 셋에서도 동일한 클래스를 이용하였다. 이차원 키포인트에서 삼차원 자세 키포인트로 차원을 올리기 위해서 네트워크가 간단하지만 빠르고 성능 좋은 모델[11]을 이용하였다.

그림 4는 사람과 동물에 대해 각각 이차원 추정한 결과와 그에 따른 삼차원 추정 결과를 보여준다. 그림 5는 학습에 사용되지 않았던 종류의 클래스에서도 이차원 자세와 삼차원 자세를 추정한 결과를 보여준다.

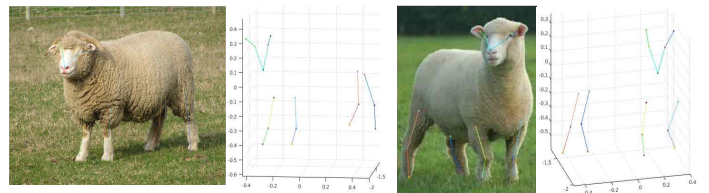
3. 결 론

본 연구에서는 단안 영상으로부터 공통 형상 모델을 이용한 사람과 동물의 삼차원 자세 추정에 대한 기법을 제안하였다. 공통 형상 모델을 통해 적은 양의 네트워크로 여러 종의 자세 추정이 가능하고 성능도 나아짐을 확인하였고, 나아가 제한적이지만 학습되지 않은 종의 자세 추정 가능성을 확인하였다.

좀더 성능을 향상하기 위해 동물에 대해 이차원 영상과 레이블을 지속적으로 획득하고 있으며, 향후엔 정량적인 평가가 가능할 것이다.



<그림 4> 이차원 및 삼차원 자세추정 결과 (첫 번째줄) 사람, (두번째줄) 말



<그림 5> 이차원 및 삼차원 자세추정

[참 고 문 헌]

[1] <https://waymo.com/open>
 [2] Timo von Marcard, et al, "Recovering Accurate 3D Human Pose in the Wild Using IMUs and a Moving Camera", ECCV 2018
 [3] Kanazawa, et al, "Learning 3D Human Dynamics from Video", CVPR 2019
 [4] Cao. Jinkun, et al, "Cross-Domain Adaptation for Animal Pose Estimation", ICCV 2019
 [5] Alexander Mathis, et al, "Pretraining boosts out-of-domain robustness for pose estimation", WACV 2021
 [6] Kim Youwang, et al, "Unified 3D Mesh Recovery of Humans and Animals by Learning Animal Exercise", BMVC 2021
 [7] 민지홍 외, "사람과 사족 동물의 공통형상 모델 기반 자세추정기법", ICROS 2022
 [8] Loper, et al, "SMPL: A Skinned Multi-Person Linear Model", SIGGRAPH 2015
 [9] Zuffi, et al, "3D Menagerie: Modeling the 3D Shape and Pose of Animals", CVPR 2017
 [10] H. Fang, et al., "RMPE: Regional Multi-person Pose Estimation," ICCV 2017
 [11] Julieta Martinez, et al, "A simple yet effective baseline for 3d human pose estimation", ICCV2017