

사람의 움직임에 대한 Optical Flow 데이터 세트 생성 방법에 관한 연구

배은지, 윤현세, 김도영, 송혜원, 김재경, 이경준, 이상훈

연세대학교

{eo_oej, hsyoon97, tnyffx, shw3164, jkkproject, naive2kj90, slee} @ yonsei.ac.kr

Generating and Augmenting Human Optical Flow Dataset

Bae Eunjee, Yoon Hyunse, Kim Doyoung, Song Hyewon, Kim Jaekyung, Lee Kyungjune,

Lee Sanghoon

Yonsei Univ.

요약

인간을 훈련 및 테스트 대상으로 하는 광학 흐름 데이터 세트가 부족하다. 결과적으로 광학 흐름은 현장에 참여한 인간과의 실제 상황에 적용하기 어려워졌다. 따라서 본 논문에서는 광범위한 합성 인간 모션, 카메라 각도 및 카메라 모션으로 구성된 광학 흐름 데이터 세트를 생성하는 방법을 제안한다. 데이터 세트는 오픈 소스 소프트웨어, 합성 인간 및 인간 모션 데이터의 조합을 사용하여 생성된다. 제안하는 방법을 통해 우리는 데이터 세트의 양과 품질을 더욱 향상시킬 수 있도록 동일한 동작에서 데이터를 지속적으로 증가시킬 수 있다.

I. 서론

심층 신경망을 훈련할 때는 작업에 맞는 데이터 세트를 제공하는 것이 필수적이다. 테스트용 데이터 세트가 학습용 데이터 세트와 유사성을 공유하지 않는 경우 네트워크가 결합 있는 결과를 생성하여 정확도가 떨어질 가능성이 크다. 따라서 심층 신경망의 목적에 맞는 데이터 세트를 제공하는 것이 필수적이다.

심층 신경망을 통해 물체의 Optical Flow를 추정하기 위해, 그 효과가 입증된 많은 합성 데이터 세트가 있다. KITTI를 제외한 다른 데이터 세트는 컴퓨터 소프트웨어를 사용하여 생성된 3차원 물체의 움직임을 녹화하고 그 시퀀스에서 캡처된 움직임을 이미지로 구성한다. 다른 데이터 세트와 달리 KITTI는 도로를 주행하는 실제 자동차의 RGB 시퀀스로 구성된다. 대부분의 심층 신경망은 이러한 데이터 세트를 사용하여 Optical Flow를 추정하기 위한 모델을 학습시켜 이러한 모델의 성능은 각 데이터 세트의 주어진 데이터로 측정된다. 그러나 실제 인간 모션 시퀀스로 이러한 훈련된 모델을 테스트할 때 정확도가 크게 떨어진다. 이는 이러한 모델을 학습하는데 사용되는 데이터 세트에 테스트된 데이터와 유사한 기능을 가진 데이터가 포함되어 있지 않기 때문이다.

따라서 본 논문은 실제 인간 데이터로 테스트할 때 모델의 정확도를 향상시키기 위해 RGB 데이터, Depth map, Optical Flow 및 카메라 매개 변수와 같은 다양한 데이터 유형으로 구성된 인간 Optical Flow 데이터를 생성하는 방법을 제안한다. 우리는 이 데이터 세트를 사용하여 기존 데이터 세트로 pretrained 모델을 훈련하고 추정된 Optical Flow의 정확도를 우리 데이터 세트로 훈련되지 않은 모델과 비교할 것이다.



그림 1. 2차원 데이터 취득을 위한 3차원 합성 데이터 생성

II. 본론

본 논문에서 제안한 방법은 3차원 인간 모델에서 파생된 다양한 2차원 데이터를 생성하는 것을 목표로 한다. 제안된 방법은 원하는 데이터를 출력하기 위해 파이썬 프로그래밍을 지원하는 주요 소프트웨어로 Blender 3.0을 사용하였다. Blender 3.0은 3차원 객체, 시각 효과 (VFX) 및 3차원 애니메이션을 모델링하는데 특화된 오픈 소스 소프트웨어이다. 사용자에게 더 많은 자유를 주기 위해 소프트웨어를 파이썬 프로그래밍과 함께 사용하여 맞춤형 알고리즘 또는 수학 문제를 해결할 수 있다. 본 논문에서 제안하는 방법을 통해 얻을 수 있는 데이터는 RGB 데이터와 그에 해당되는 Depth Map, Optical Flow 및 카메라 파라미터이다.

먼저 Blender 3.0의 가상 공간에 SMPL-X 모델을 불러온다. SMPL-X 모델은 맥스 플랭크 연구소가 설계한 3차원 합성 사람 모델이다 [6]. SMPL-X 모델은 사용자가 신체 형태, 얼굴 모양, 신체 포즈, 손 포즈 및 얼굴 표정을 제어할 수 있도록 한다. 신체 형태와 얼굴 형태에 대한 Blendshape 파라미터를 제어하여 신체 형태와 얼굴 형태를 랜덤하게 생성할 수 있다. 신체 형태와 얼굴 형태는 임의로 설정되는 반면, 각각의 신체 자세는 주어진 동작 시퀀스에 따라 설정된다. 이러한 몸 동작 시퀀스는

3차원 관절 시퀀스로 저장된 다양한 유형의 모션을 포함하는 AMSASS 데이터 세트에서 얻는다. Handcrafted 방법과 Deep Learning 모델은 Optical Flow를 구할 때 이미지의 RGB feature를 사용하여 두 사진 간의 매칭되는 픽셀을 찾아 매칭되는 픽셀의 거리를 구하기 때문에 사람의 Optical Flow를 구하기 위해 3차원 인간 모델에도 Texture를 적용한다. 3차원 합성 사람 생성, Textre 적용 및 AMASS 모션 시퀀스 데이터를 적용해 3차원 합성 인간에 대한 설정이 끝난다.

3차원 합성 인간 모델에 대해 설정한 후 합성 인간을 둘러싼 가상의 환경을 만든다. 합성 인간 모델 주위에 합성 공간 혹은 배경을 추가하지 않을 경우 Blender 3.0에서 추출된 이미지의 배경은 검은색으로 표현이 된다. 하지만 합성 인간을 제외한 공간이 검은색으로 채워져 있을 경우 배경에 대한 두 이미지 간의 모션 관계성을 구하기 어렵다. 또한 합성이 아닌 현실에서 찍은 사진에도 얻고자 하는 인간 데이터 외에 복잡한 배경이 존재하기에 실제와 차이가 생겨난다. 이러한 차이를 없애기 위해 Blender 3.0에서 배경 데이터도 같이 나오도록 설정해야 한다. 배경을 더하기 위해 Blender 3.0에서 제공하는 톨로 간단한 벽과 바닥을 생성하여 Texture 역할을 할 그림을 추가한다.

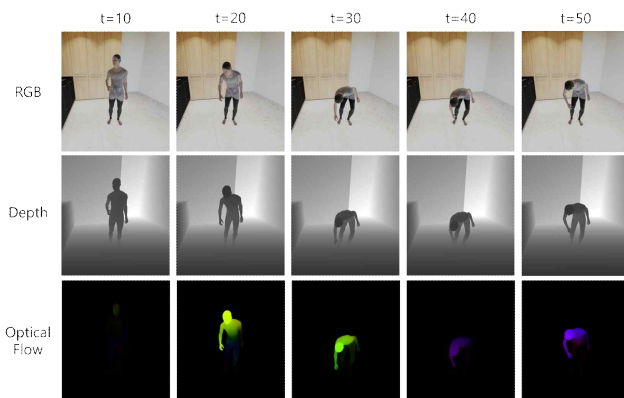


그림 2. 제안된 방법으로 생성된 2D 합성 데이터 샘플

학습 데이터의 다양성과 양을 높이기 위해 3차원 합성 공간에서 렌더링을 통해 데이터 취득을 도와주는 Blender 3.0 카메라 톨의 숫자를 늘릴 수 있다. 카메라의 숫자를 늘림으로써 같은 모션에서 각 카메라가 고유의 데이터를 얻을 수 있다. 카메라 수를 늘리는 것 외에도 데이터 세트의 크기를 증가시키지만, RGB 데이터에 아티팩트를 추가하고 카메라 움직임에 동요를 주는 것도 데이터의 다양성을 증가시킨다.

III. 결과 및 분석

본 논문에서 생성된 학습용 데이터를 사용하여 공개되어 있는 State-of-the-art Optical Flow 모델인 GMA 모델에 학습을 시켜 보았다. 본 논문에서 제안하는 데이터 세트를 사용하여 학습시킨 모델과 학습시키지 않은 모델을 비교하였다. 두 모델의 성능을 테스트용 Optical Flow 데이터를 기준으로 비교해 보았다. 실험한 결과 본 논문에서 제안한 데이터 세트를 사용하여 finetune 한 모델의 성능이 finetune 하지 않은 모델의 성능보다 월등히 좋았다. 이러한 결과에 대한 이유는 이전 모델을 학습할 때 사용한 데이터 세트에는 사람에 특화된 데이터가 없었기 때문이다.



그림 3. 사람에 대한 Optical Flow 결과 비교

그림 3에서 표기된 결과를 가지고 비교를 했을 때 finetune 하지 않은 모델은 사람의 큰 윤곽은 잘 유추하지만 세세한 모션의 디테일에 대해 부정확한 Optical Flow 값을 볼 수 있다. 하지만 finetune한 모델에서 사람의 손과 다리의 모션의 Optical flow가 정답과 제일 가까운 결과를 얻을 수 있다.

IV. 결론

본 논문에서는 Deep Learning 모델을 기반으로 두 이미지 간의 사람 Optical Flow Estimation 취약점을 보완하기 위하여 사람 Optical Flow 데이터 세트 생성 방법을 제안했다. 이 방법을 통해 여러 합성 데이터를 얻을 수 있으며 Blender 3.0에서 제공되는 여러 기능을 활용하여 데이터의 양과 질을 높일 수 있다. 생성된 데이터 세트를 사용하여 Deep Learning 모델을 학습시킨 결과 사람의 움직임에 대해 정확한 Optical Flow를 얻을 수 있었다.

ACKNOWLEDGMENT

이 성과는 2020년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (No. 2020R1A2C3011697)

참 고 문 헌

- [1] Mayer, Nikolaus, et al. "A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.
- [2] Dosovitskiy, Alexey, et al. "Flownet: Learning optical flow with convolutional networks." Proceedings of the IEEE international conference on computer vision. 2015.
- [3] Butler, Daniel J., et al. "A naturalistic open source movie for optical flow evaluation." European conference on computer vision. Springer, Berlin, Heidelberg, 2012.
- [4] Geiger, Andreas, et al. "Vision meets robotics: The kitti dataset." The International Journal of Robotics Research 32.11 (2013): 1231-1237.
- [5] Pavlakos, Georgios, et al. "Expressive body capture: 3d hands, face, and body from a single image." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019.
- [6] Jiang, Shihao, et al. "Learning to estimate hidden motions with global motion aggregation." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021.