MorekerLess – Better pose estimation with less work

David Pagnon[1], Dario Cazzola[1], Steffi Colyer[1]
[1]University of Bath, Bath, United Kingdom
Email: dp2032@bath.ac.uk

Summary

Markerless motion capture is becoming the new norm for kinematic analysis, but some challenges remain. One of them lies in 2D pose estimation: anatomical keypoints could be labeled more accurately and should be more numerous for certain parts of the body, such as the spine. However, manually creating a new dataset or modifying annotations is extremely costly, as hundreds of thousands of images are generally required. We introduce MorekerLess, a pipeline for automatically labeling and training on any keypoint set. Results show that additional keypoints are correctly detected, and joint center estimation is improved by 0.2-0.4 cm over the current state-of-the-art. Our pipeline and models will be released at: http://github.com/davidpagnon/morekerless.

Introduction

Creating a new dataset for training markerless algorithms usually requires experts to manually annotate thousands of images. Moreover, if keypoints need to be moved or added, labeling must start over. This study aimed to explore an alternative approach leveraging an existing dataset: one single body mesh is labelled, given different body shapes, clothed, animated, put in different environments, and rendered to create synthetic but realistic videos. We then automatically obtain a large and diverse dataset of images and labels, which is used to train a pose estimation model.

Methods

BEDLAM [1] provides synthetic but realistic videos, consisting of animated SMPL meshes in virtual environments. We select vertices on one single reference mesh, propagated to the whole dataset. These vertices are treated as skin markers and used to calculate joint centers via regression equations.

37 points were chosen relating to 8 mid-segment vertices, 3 for the spine, 2 for the head, 2 per hand, 4 per foot, and 12 calculated joint centers, for extensive motion analysis with minimal keypoints (Figure 1). We reprojected the points on images and retrained for 30 epochs with RTMPose-m [2].

The resulting model, MorekerLess, was evaluated using bioCV [3], a dataset consisting of 9-camera video and optical motion capture. We took a subset of 10 trials from 10 participants, performing one of 5 tasks. We ran our model on video data, and triangulated results with Pose2Sim [4]. We also ran the same markerless pipeline with RTMPose-m, which was previously trained for 700 epochs on 8 datasets with Halpe26 points, for state-of-the-art comparison. 3D markerless joint centers were compared with marker-based ones, and the mean and standard-deviation of Euclidian distances were calculated.

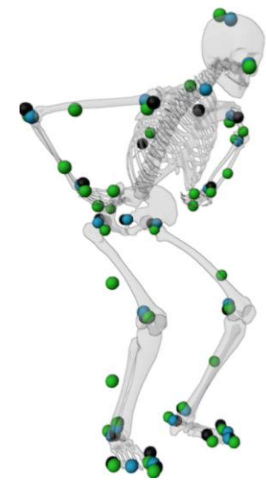

Figure 1: Green dots represent MorekerLess keypoints, blue dots Halpe26, and black dots joints computed from markers. OpenSim inverse kinematics was automatically run with Pose2Sim [4].

Results and Discussion

All MorekerLess points were triangulated with a reprojection error of 3 to 9 pixels, matching the Halpe26 reprojection errors. A small improvement to the 3D joint center estimations was exhibited compared to the over Halpe26 results (Table 1).

Table 1: Mean difference (cm) with marker-based joint centers

Mean ± SD	Halpe26	**MorekerLess**
Ankle	2.6 ± 1.5	**2.2 ± 1.4**
Knee	2.8 ± 1.2	**2.4 ± 1.1**
Hip	4.0 ± 1.1	**3.8 ± 1.1**

Conclusions

This approach shows promising results and could enable the estimation of custom keypoints without costly manual labeling. Moreover, it could be further developed to train for monocular 3D pose estimation, direct joint angle prediction, or even force estimation from videos.

Acknowledgments

This research was funded by CAMERA, the RCUK Centre for the Analysis of Motion, Entertainment Research and Applications, EP/M023281/1 and EP/T022523/1. Special thanks to Murray Evans and Jake Willott for their help.

References

[1] Black et al. (2023) *IEEE/CVF*, 8726-8737.
[2] Jiang et al. (2023) *arXiv*, 2303.07399.
[3] Evans et al. (2024) Scientific Data, **11**: 1300.
[4] Pagnon et al. (2022) *JOSS*, 4362.