



HAL
open science

Investigating a hybrid markerless method based on DeepLabCut

Alexandru Ghitu, Benjamin Balonchard, Theophile Cocquerez, Khalil Ben
Mansour

► **To cite this version:**

Alexandru Ghitu, Benjamin Balonchard, Theophile Cocquerez, Khalil Ben Mansour. Investigating a hybrid markerless method based on DeepLabCut. 2024. hal-04737933

HAL Id: hal-04737933

<https://hal.science/hal-04737933v1>

Preprint submitted on 15 Oct 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Investigating a hybrid markerless method based on DeepLabCut

Alexandru Ghitu^a, Benjamin Balonchard^a,
Theophile Cocqueret^a, Khalil Ben Mansour^{a*}

*a. Laboratoire de Biomécanique et Bio-ingénierie -
UMR 7338 CNRS UTC - Alliance Sorbonne Université
Centre d'Innovation, 60200 Compiègne*

*Corresponding author. Email: khalil.ben-mansour@utc.fr

Keywords: Motion capture; pose estimation, key point, Artificial intelligence

1. Introduction

Marker-based motion capture systems are widely regarded as the most accurate method for capturing motion data (Miller et al., 2016). However, these systems have several limitations, including high costs due to the large number of cameras, the need to be checked by an expert and the limits of installation in outdoor, aquatic or industrial environments.

Currently, Artificial Intelligence has made significant progress, resulting in Markerless Motion Capture solutions (MMCs) becoming a more cost-effective and lightweight alternative to traditional motion capture methods. MMCs require only RGB-calibrated cameras (Escalera et al., 2010), which are now widely available, making them easier to transport and to integrate into workplaces or training facilities. Nevertheless, the accuracy of these systems can be affected by various factors, such as the quality of the video cameras, the lighting conditions, the complexity of the movement, and the algorithms used for tracking and reconstruction.

This study presents the initial results of a benchmark conducted under controlled conditions. The benchmark used a model trained on blue fabric markers, referred to as DIY. The aim was to explore if DIY could improve accuracy when using Markerless models such as ResNets (Kaiming H. et al, 2015). The purpose of using invariant markers was to track consistent points across subjects, which theoretically leads to enhanced accuracy.

The exploration of this alternative was motivated by the difficulty in discerning points of interest due to subjects wearing various clothing, which complicates bone observation. The DIY method was used to mitigate operator-dependent ambiguity during the labelling stage and provide invariant points of interest during the AI detection phase. The training and detection were both done using the DIY fabric.

The study was carried out using DeepLabCut, a toolbox based on transfer learning (Mathis A. et al., 2018). It focused mainly on the detection of points of interest in 2D, which is critical for accurate 3D reconstruction.

2. Methods

Eleven participants were involved in this study, each completing identical tasks under two distinct conditions: Markerless and DIY. This facilitated the creation of the two models for identifying key anatomical points. Additionally, five experienced operators annotated two sets of recorded data from a twelfth participant, not part of the training set. The outcomes of this manual annotation serve as the reference against which the other two methods are compared.

2.1 Setup

Data were collected using a single fixed and calibrated camera (Apple iPhone 13) at a frame frequency of 24Hz. The training and evaluation videos were 450L x 800H pixels in size, with the subject's height being approximately 300 pixels. We tried to keep the frame size small to avoid increasing the training and inference time (Nath T et al, 2019).

To collect data on the subject in the DIY setup, blue fabric of the same size was taped to the points of interest: the left and right wrists (WristL and WristR), the left and right elbows (ElbowL and ElbowR), the left and right shoulders (ShoulderL and ShoulderR), the left and right hips (HipL and HipR) and the left and right eyes (HeadL and HeadR). Note: In the DIY setup, the fabric was placed on the forehead, just above the eyes. In addition, the fabric was wrapped around the subjects' wrists and elbows so that they were visible even during complex internal/external arm rotation movements.

Other than the differences listed above, both setups were identical. The subject was asked to place chess pieces on a chessboard (Fig.1).

2.2 Model training

Both models (Markerless and DIY) were trained with the ResNet-50 convolutional neural network for 1030000 iterations on a dataset containing 200 uniformly extracted frames from the acquired videos (as recommended in Mathis A. et al., 2018): approximately 20 frames from 11 different subjects (some frames were omitted because they were too similar).

In the DIY setup, the centre of each observed fabric was marked if visible. In the classic markerless setup, the centre of the wrists, elbows and eyes were labelled if visible. In the case of shoulders and hips, the edges of the subject's shoulders and hips were labelled if visible.

2.3 Benchmark

Both methods were tested using human-labelled data. Five experienced operators manually labelled 20 frames for both setups, using the two methods as described in Section 2.2, on a subject that was not used to train the model. Finally, to assess the accuracy of the two models, the results obtained were compared to the median of the coordinates of each manually labelled point.



Figure 1 Markerless and DIY setups, before and after point detection.

3. Results and discussion

Table 1 presents the benchmark results for both Markerless and DIY methods. The DIY setup resulted in better accuracy, except for WristL and ShoulderL.

Table 1. Median (M), 1st quartile ($Q1$) and 3rd quartile ($Q3$) distance (in pixels) between AI-labelled points of interest and the median of operator labels.

	Markerless			DIY		
	Q1	M	Q3	Q1	M	Q3
WristR	1.49	2.32	3.39	1.50	1.73	2.01
WristL	1.15	1.88	2.63	1.45	1.95	2.19
ElbowR	2.48	6.76	8.10	1.74	2.12	2.41
ElbowL	3.29	4.60	5.94	1.99	2.25	2.66
ShoulderR	1.14	2.46	3.68	1.58	2.07	2.30
ShoulderL	1.41	2.66	3.64	1.95	2.71	3.44
HeadR	2.03	2.43	2.83	0.41	0.89	1.15
HeadL	2.05	2.50	3.19	1.16	1.68	2.15
HipR	3.52	4.91	6.62	1.85	2.7	4.49
HipL	1.37	2.84	5.46	2.22	2.49	3.78

Additionally, in the DIY setup, HeadL detection was less accurate due to the subject's hair covering the blue fabric. In the markerless setup, HipR detection was almost twice as inaccurate as HipL detection.

Furthermore, ElbowR detection was less accurate than ElbowL detection. These contradictory results may be due to chance, as the randomly selected frames presented a more complex pattern to detect for these two points.

The DIY setup could potentially enhance accuracy in benchmarks and provide a more precise labelling experience for operators, thereby reducing ambiguity. It was observed during a subsequent benchmark of the Markerless dataset that the dispersion between operator labelling can reach almost 3% of the subject's height in pixels. This observation may be attributed to the reduced visibility of certain joints, such as an extended elbow, resulting in more ambiguous labelling.

Furthermore, it may be advisable to consider a fabric colour that offers more contrast with the subject's clothing.

4. Conclusions

Accurate definition of key anatomical landmarks of the human body is essential for robust motion biomechanics analysis. This study aimed to explore if DIY could improve the precision compared to a conventional markerless method based on the DeepLabCut toolbox.

The benchmark showed that the DIY setup performed better overall. It is worth considering that these are preliminary results and that training with more frames may enhance the accuracy of the MMCs. It could be argued that the DIY approach offers a compromise between the highly accurate marker-based motion capture and the less accurate MMCs.

Acknowledgements

The authors would like to thank the BMBI – TSS platform team for their assistance during the experiment, as well as all the subjects who volunteered to help gather data for training.

Conflict of Interest Statement

All authors have no conflicts of interest to declare.

References

- Miller, E. et al. (2016). *Mechanical testing for three-dimensional motion analysis reliability*. Gait Posture 50, 116–119. doi.org/10.1016/j.gaitpost.2016.08.017
- Escalera D. L et al. (2010). *Automatic chessboard detection for intrinsic and extrinsic camera parameter calibration*. Sensor 10, 2027-2044. doi:10.3390/s100302027

Kaiming H., et al. (2015). *Deep residual learning for image recognition*, IEEE.
doi:10.1109/CVPR.2016.90

Received date: 28/03/2024

Accepted date: 28/06/2024

Published date: XX/XX/2024

Volume: 1

Publication year: 2024