



HAL
open science

Comparison of two confidencebased methods for markerless motion capture

Anaïs Chaumeil, Antoine Muller, Raphaël Dumas, Thomas Robert

► To cite this version:

Anaïs Chaumeil, Antoine Muller, Raphaël Dumas, Thomas Robert. Comparison of two confidence-based methods for markerless motion capture. 48e Congrès de la Société de Biomécanique, Oct 2023, Grenoble, France. pp.S172-S174. hal-04289910

HAL Id: hal-04289910

<https://hal.science/hal-04289910v1>

Submitted on 16 Nov 2023

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.

ABSTRACTS 48th Congress of the Society of Biomechanics

Yohan Payan & Lucie Bailly

To cite this article: Yohan Payan & Lucie Bailly (2023) ABSTRACTS 48th Congress of the Society of Biomechanics, Computer Methods in Biomechanics and Biomedical Engineering, 26:sup1, S1-S341, DOI: [10.1080/10255842.2023.2246304](https://doi.org/10.1080/10255842.2023.2246304)

To link to this article: <https://doi.org/10.1080/10255842.2023.2246304>



© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 24 Oct 2023.



Submit your article to this journal [↗](#)



Article views: 214



View related articles [↗](#)



View Crossmark data [↗](#)

Comparison of two confidence-based methods for markerless motion capture

A. Chaumeil, A. Muller, R. Dumas and T. Robert

LBMC UMR_T 9406, Univ Lyon, Univ Eiffel, Univ Claude Bernard Lyon 1, Lyon, France

1. Introduction

Most video-based human pose estimation software provide results in the 2D coordinate system of the input image or video. However, studies in biomechanics often require 3D information. One commonly used method to obtain 3D points from 2D points and calibrated cameras is the direct linear transform (DLT) (Abdel-Aziz and Karara 1971).

Most pose estimation software associates each pixel with a value between 0 and 1, which corresponds to the confidence that the investigated keypoint is located at this particular pixel. It transforms the image into a confidence heatmap. The most common approach uses this confidence information in a binary way: the 2D point of interest corresponds to the pixel with the maximal confidence over the whole heatmap. Yet, Iskakov et al. (2019) obtained promising results by using the whole confidence information: both temporal continuity and accuracy increased, especially when using a small set of video cameras. However, these results were obtained using specifically trained neural networks. It would then be interesting to test if such improvement could also be obtained using the same idea applied to more generic tools, e.g. using confidence heatmaps generated by OpenPose (Cao et al. 2019). Two approaches can be envisaged, according to Iskakov et al. (2019): one consists in using a classical binary 2D point estimation for each camera view and then to compute the 3D point using a weighted DLT based on their confidence given by the heatmaps. The second approach, which is recommended by Iskakov et al. (2019), consists in projecting the 2D heatmaps to obtain 3D confidence volumes, from which a 3D point is extracted.

In this study, we propose an approach that allows exploiting the whole information from the confidence heatmaps provided as output of OpenPose. The goal of this study is therefore to compare the two previously presented confidence-based methods for markerless 3D reconstruction.

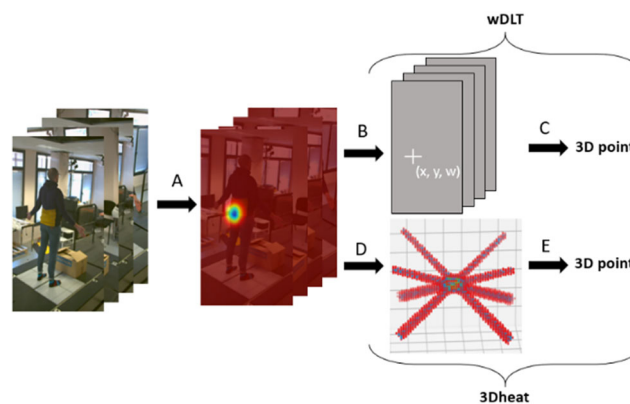


Figure 1. Visual representation of the confidence-based methods.

2. Methods

2.1. Participants

For this study, eight participants were recruited (3 women), with age: 28.3 ± 2.8 years old, height: 1.70 ± 0.09 m and mass: 63.1 ± 6.36 kg. Participants had no history of musculoskeletal problems. They signed an informed consent form and the study was approved by our institutional review board.

2.2. Experimental setup

Data was collected using 10 Qualisys Miquis Video cameras at 60 Hz which were synchronized and calibrated using Qualisys Track Manager (v2021.1.2 – average residual after calibration: 0.38 mm). Participants wore their usual clothes. A box (approximately 60x40x30 cm) was put in front of them on the ground. Participants were asked to stand still and then squat, grab the box, stand up with the box, squat again and then put the box back in front of them. Three repetitions per participant were performed.

2.3. Data processing

Videos were processed using Theia3D (v 2022.1), which was used as reference in the absence of markers. A visual explanation of the two confidence-based methods is displayed in Figure 1. Both methods were used to obtain 3D position of the main joint centres of interest in this study (left (L) and right (R) shoulders, elbows, wrists, hips, knees and ankles). OpenPose (v 1.7.0) was applied to a subset of four evenly distributed cameras. It provided the confidence heatmap of each keypoint (Figure 1A) from which the position (x, y) and associated confidence (w) of the keypoints were extracted (Figure 1B).

First (Figure 1C), 3D positions were estimated using a confidence-based weighted DLT (wDLT)

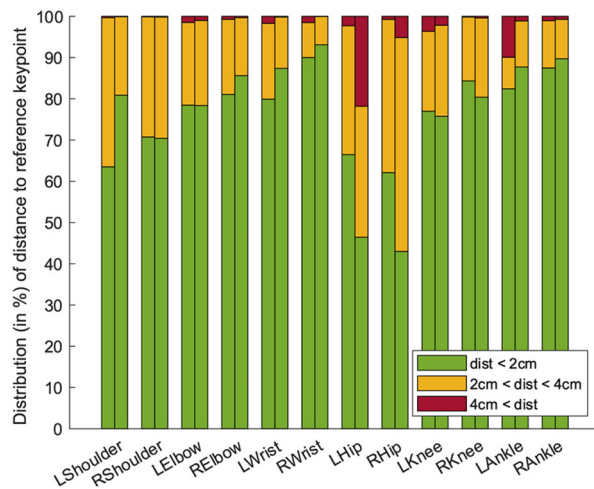


Figure 2. Distribution, in percent, of the Euclidean distance between wDLT and Theia3D (left bar) and between 3Dheat and Theia3D (right bar), for each keypoint.

(Pagnon et al. 2021). Second, 3D heatmaps (3Dheat) were built for each point (Figure 1D). 3D space was discretized with a point every 3 mm in all 3 directions. Then, using the intrinsic and extrinsic parameters of the cameras, the points were projected on the heatmaps for each camera and associated with the corresponding confidence value. The 4 confidence values associated with each camera were then summed. The 3D point with the maximum summed confidence was selected as the position of the keypoint for the studied frame (Figure 1E).

2.4. Data analysis

Euclidean distances between joint centre positions estimated with wDLT and 3Dheat and Theia3D were computed. Associated root mean square errors (RMSE) and standard deviation were computed for each keypoint for both methods. Euclidean distances were classified according to their values (under 2 cm, between 2 cm and 4 cm and above 4 cm). Threshold values were selected based on the differences observed in the literature between marker-based and markerless for keypoint detection (Kanko et al. 2021): below 2 cm, the distance is considered to be due to measurement noise, and above 4 cm it is considered to be a major difference.

3. Results and discussion

The RMSE is below 2 cm for all keypoints except 3Dheat LHip (3.3 ± 1.9 cm), wDLT and 3Dheat RHip (2.0 ± 0.7 cm and 2.4 ± 0.9 cm respectively) and wDLT LAnkle (2.4 ± 1.7 cm). For both hips, RMSE

and standard deviation are higher for 3Dheat than wDLT (LHip: 1.9 ± 0.9 cm for wDLT). For wDLT LAnkle, RMSE is higher than 2 cm with a bigger standard deviation than for 3Dheat (1.59 ± 0.80 cm for 3Dheat).

Figure 2 displays the distribution of differences. For the hips, there are fewer differences that are below 2 cm and more differences above 4 cm for 3Dheat than for wDLT. For LAnkle, the proportion of differences above 4 cm is higher for wDLT (9.9%) than for 3Dheat (1.1%), while the proportion of differences below 2 cm remains similar (87.5% for wDLT and 89.6% for 3Dheat). For the other keypoints, there is not a clear trend for one method generating smaller differences than another.

For the hips, the observed differences between the methods can be explained by several factors. Firstly, the movement itself makes it difficult to estimate the position of the hip joint centers, especially during the squat movement. Hip keypoints obtained with OpenPose thus have a medium or low confidence. Moreover, differences in hip joint center definition and detection between Theia3D and OpenPose could generate systematic errors between the two confidence-based 3D reconstruction methods and Theia3D. However, both Theia3D and OpenPose provide limited information about biomechanical definition of the keypoints, thus systematic errors can only be evaluated experimentally. Secondly, 3Dheat is a completely different approach. For wDLT, the point that is reconstructed is the point that is closest to all projection lines. For 3Dheat, the selected point is the one with maximum confidence. Heatmaps create beams in 3D space by projecting zones of confidence instead of points, and the point with the maximum summed confidence is at the crossing of two or more beams. For example, for the hips, beams do not often cross all together at some point in space, or they only cross two by two. Thus, the maximum summed confidence is not in the middle of a beam-crossing area (where the point given by the wDLT results lies) but somewhere on the periphery of this area.

For LAnkle, 3Dheat provides better results, especially regarding differences above 4 cm. One factor explaining this difference is the high number of inconsistent detections generated by OpenPose, which sometimes detects LAnkle where it is absolutely not (typically when LAnkle is occluded). In this case, the inconsistent beam crosses the other beams far from where they intersect and has thus little influence on the maximum summed confidence. However, the

point that is selected by wDLT is influenced by the inconsistent point, even if it has low confidence.

4. Conclusions

Comparison of two confidence-based methods for markerless motion capture has been performed. Overall, the two methods provide similar results. Yet, they seem to be robust against different situations: 3Dheat in the case of one wrongly detected point, wDLT in the case of uncertain points. Nonetheless, both approaches rely on OpenPose keypoint detection quality.

References

Abdel-Aziz YI, Karara HM. 1971. Direct linear transformation from comparator coordinates into object space coordinates in close-range photogrammetry. *Photogramm Eng Remote Sens.* 81(2):103–107.

Iskakov K, Burkov E, Lempitsky V, Malkov Y. 2019. Learnable triangulation of human pose. 2019 IEEE/CVF International Conference on Computer Vision (ICCV). doi:10.1109/ICCV.2019.00781.

Cao Z, Hidalgo G, Simon T, Wei S-E, Sheikh Y. 2019. OpenPose: realtime multi-person 2D pose estimation using part affinity fields.

Kanko RM, Laende EK, Davis EM, Selbie WS, Deluzio KJ. 2021. Concurrent assessment of gait kinematics using marker-based and markerless motion capture. *J Biomech.* 127:110665. doi:10.1016/j.jbiomech.2021.110665.

Pagnon D, Domalain M, Reveret L. 2021. Pose2Sim: an end-to-end workflow for 3D markerless sports kinematics—part 1: robustness. *Sensors.* 21(19):6530. doi:10.3390/s21196530.

KEYWORDS Markerless motion capture; confidence heatmaps; triangulation; 3D point reconstruction

 thomas.robort@univ-eiffel.fr