Science Arts & Métiers (SAM)
is an open access repository that collects the work of Arts et Métiers Institute of Technology researchers and makes it freely available over the web where possible.

This is an author-deposited version published in: https://sam.ensam.eu
Handle ID: .http://hdl.handle.net/10985/20156

To cite this version :

Saman VAFADAR, Wafa SKALLI, Aurore BONNET-LEBRUN, Marc KHALIFÉ, Mathis RENAUDIN, Amine HAMZA, Laurent GAJNY - A novel dataset and deep learning-based approach for marker-less motion capture during gait - Gait and Posture - Vol. 86, p.70 - 76 - 2021

Any correspondence concerning this service should be sent to the repository Administrator : scienceouverte@ensam.eu

A novel dataset and deep learning-based approach for marker-less motion capture during gait

Saman Vafadar[1, *], Wafa Skalli[1, b], Aurore Bonnet-Lebrun[1, c], Marc Khalifé[1,2,d], Mathis Renaudin[1,e], Amine Hamza[1,3,f], Laurent Gajny[1, *]

[1] Institut de Biomecanique Humaine Georges Charpak, Arts et Metiers, Institute of Technology, Paris, France.

[2] Orthopedic surgery unit, Georges Pompidou European Hospital, Paris, France.

[3]CHU de Rouen, Department of Orthopedic Surgery, Rouen, France.

[*]Corresponding authors: Saman Vafadar, saman.vafadar@yahoo.com; Laurent Gajny, laurent.gajny@ensam.eu

[b] E-mail address: Wafa Skalli, wafa.skalli@ensam.eu

[c] E-mail address: Aurore Bonnet-Lebrun, aurore.bonnet-lebrun@ensam.eu

[d] E-mail address: Marc Khalifé, marckhalife@icloud.com

[e] E-mail address: Mathis Renaudin, mathis.renaudin@ensam.eu

[f] E-mail address: Amine Hamza, aminehamza@hotmail.com

Acknowledgments

The authors thank the ParisTech BiomecAM chair program, on subject-specific musculoskeletal modelling and in particular Société Générale and COVEA. We also thank Matthieu Boëssé for his technical help.

1. Introduction

Quantitative gait analysis is an effective tool for diagnosis and prognosis in clinical issues like cerebral palsy [1]–[3]. Also, it may assist in the diagnosis and succeeding therapy of individuals at risk of falls [4], [5]. The current established gold standard systems for gait analysis are marker-based motion capture systems. However, their recognized interest, their routine clinical use is limited to a few highly specialized centers for various reasons: labor costs for data acquisition and processing, equipment costs, the need for a controlled laboratory environment, and the time-consuming testing procedure [6].

Because of these limitations, marker-less motion capture systems are proposed as cost-effective and easy-to-use alternatives. The marker-less systems mainly utilize two types of cameras, either depth cameras or RGB cameras, generally based on three-dimensional (3D) human pose estimation methods. Such methods refer to processing two-dimensional (2D) images, captured by a single (or multiple) depth camera or RGB camera, to locate a set of anatomical landmarks (or joint centers) in the three-dimensional capture volume.

Clark et al. [7] authored a recent literature review on the uses, validity, and developments of depth cameras, especially the Kinect camera. For example, for the overground gait, it was reported that the Kinect's validity for most of the kinematic gait parameters was restricted ($r<0.75$) [8]. Also, based on several studies [9], [10], substantial errors were found in assessing kinematic gait parameters (RMS error > 10 degrees or Bland-Altman limits of agreement > 10 degrees).

Concerning RGB cameras, in recent years, human pose estimation methods based on artificial intelligence, especially deep learning, have demonstrated significant progress [11]–[14]. These methods estimate the three-dimensional (3D) position of joint centers (e.g., ankles, knees) using a single or a multi-view (captured by several synchronized cameras) RGB image.

The deep learning-based human pose estimation methods are mostly trained using publicly available three-dimensional human pose datasets, mainly HumanEva [15] and Human3.6M [16] datasets. In these datasets, the reference values for joint centers were obtained by marker-based motion capture systems. However, although very rich in terms of the number of gestures and images, they may be less adapted to train the human pose estimation methods for clinical gait study for the following reasons.

First, the subjects wore their regular clothing, and the placement of markers on the regular clothing would reduce the accuracy of joint centers' reference values. Second, the number of subjects is limited. HumanEva and Human3.6M datasets contain the reference values of only four and seven

subjects. Moreover, these datasets are limited to adults and do not include the motion of pathological cases. Therefore, a human pose dataset for the specific aim of gait study should be collected.

The objective of this study is to investigate the added value of collecting a dedicated human pose dataset and comparing the performance of a marker-less motion capture system based on the state-of-the-art human pose estimation method when trained on the Human3.6M dataset alone and after finetuning on the new dataset.

2. Materials and Methods

2.1. Setup and calibration of the marker-less system

The gait analysis was focused on the walking part of a 4 m stand up and go test that is routinely used in clinics [17]. The proposed marker-less motion capture system was composed of four RGB cameras (GoPro Hero 7 Black) – resolution 1920x1080 pixels, 100 frames per second, and linear field of view. The cameras were fixed using specific 3D printed fasteners on two aluminum bars. The distance and angle between two cameras on a bar were respectively 75cm and 15°. Each aluminum bar was mounted on a one-meter height tripod. As shown in Figure 1, the first couple of cameras was placed in front of the subject (referred to as frontal cameras) and the second on its right side (referred to as lateral cameras).

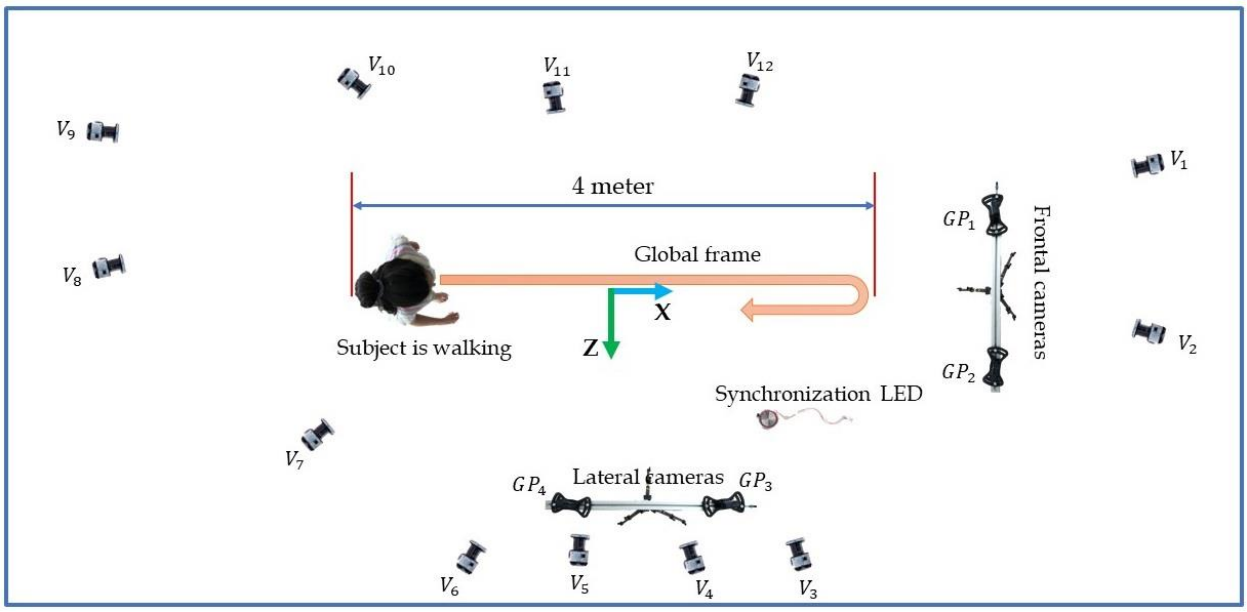

Figure 1. The experimental setup showing the placement of the VICON system's cameras (Vi) and the GoPro cameras (GPi)

A calibration checkerboard (calib.io, 9x14, 40mm) was used for intrinsic and extrinsic calibration of the marker-less system. The classical pinhole model [18] was adopted due to the selection of linear field of view. Around 40 images of the checkerboard were acquired at various positions in the

experiment room. The intrinsic (focal lengths, principal points, skew coefficient, and distortion parameters), extrinsic (relative position and orientation of each camera with respect to a global 3D frame) were estimated using the MATLAB Stereo Camera Calibrator App.

At this stage, a 3D frame was defined for the marker-less motion capture system, ensuring its potential for use in a realistic clinical environment. However, for validation purposes, we needed to define a common 3D frame with a reference system.

2.2. Setup and calibration of the reference system

A marker-based motion capture system (VICON system, Oxford Metrics, UK) was used as a reference for both training and evaluation. It was composed of twelve cameras. The frequency of data acquisition for data collection of the ENSAM pose dataset was 100 Hz to simplify the synchronization of the marker-less and marker-based systems. The VICON system was calibrated according to the recommendations of the supplier. The VICON 3D frame was adopted as the experiment room frame for both systems (Figure 1).

For that purpose, the calibration checkerboard was equipped with six reflective markers. The relative position of markers and checkerboard corners was measured. To this end, this calibration tool was scanned using a biplane X-ray system (EOS system, EOS imaging Paris France). It enabled the 3D reconstructions of the checkerboard and markers, thanks to an in-house software (Figure 2).

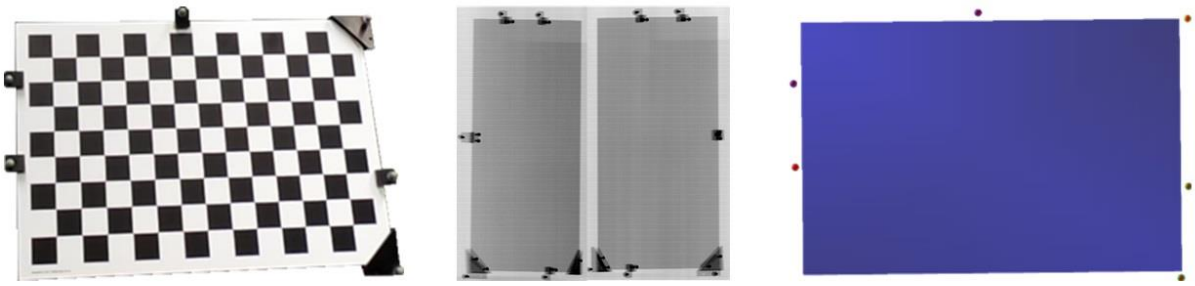

Figure 2. **Left:** The calibration device equipped with six markers. **Middle:** The biplanar X-ray images of the calibration device. **Right:** The 3D reconstruction for measuring the relative position of markers and checkerboard corners.

2.3. Synchronization

An Arduino board connected to the VICON-based system was used to synchronize the GoPro-based and the VICON-based systems. The Arduino generated a time-dependent signal that turned ON and OFF a set of LEDs placed in the capture space (Figure 1). The LEDs' status, recorded in every frame of the videos, was determined using image processing techniques. Cross-correlation between the

retrieved signals from the GoPro cameras and VICON data helped find the temporal offset between the recorded videos and data.

2.4. Constitution of the new database

Thirty-one subjects (19 males and 12 females) participated in this study after informed consent. They all had to perform Vicon based gait analysis for research previously approved by the relevant ethics committee (CCPP 06036 and CPP 06001, Paris VI). Subjects were on average 25 years old with a large variation range (range: 6-44 years), mean height was 168 cm (125-199 cm), body mass was 65 kg (range: 30-90 kg), and Body Mass Index (BMI) was 22.5 kg/m^2 (range: 16.9-29.4 kg/m^2). Among this population, 22 subjects were asymptomatic, one subject had moderate scoliosis, another had a spondylolisthesis. The remaining seven subjects were children or teenagers (range: 6-15 years) suffering from X-linked hypophosphatemia (XLH). This pathology can cause short stature and bone deformities.

2.5. In-vivo data acquisition

The Vicon based gait analysis protocol, as published in [19], [20], was used. Subjects were asked to be with their underwear. Fifty-one reflective markers were placed by an orthopedic surgeon directly on the subjects' skin, at well-defined anatomical landmarks locations. The marker-set was based on [21], [22]. Subjects equipped with markers first underwent a biplanar X-ray examination using the EOS system (EOS Imaging, Paris, France) in a standard standing posture (Figure 3) to get accurate registration between anatomic and markers coordinate systems. Then, subjects were asked to walk for ten walking trials at their comfortable self-selected walking speed, recorded by both motion capture systems.

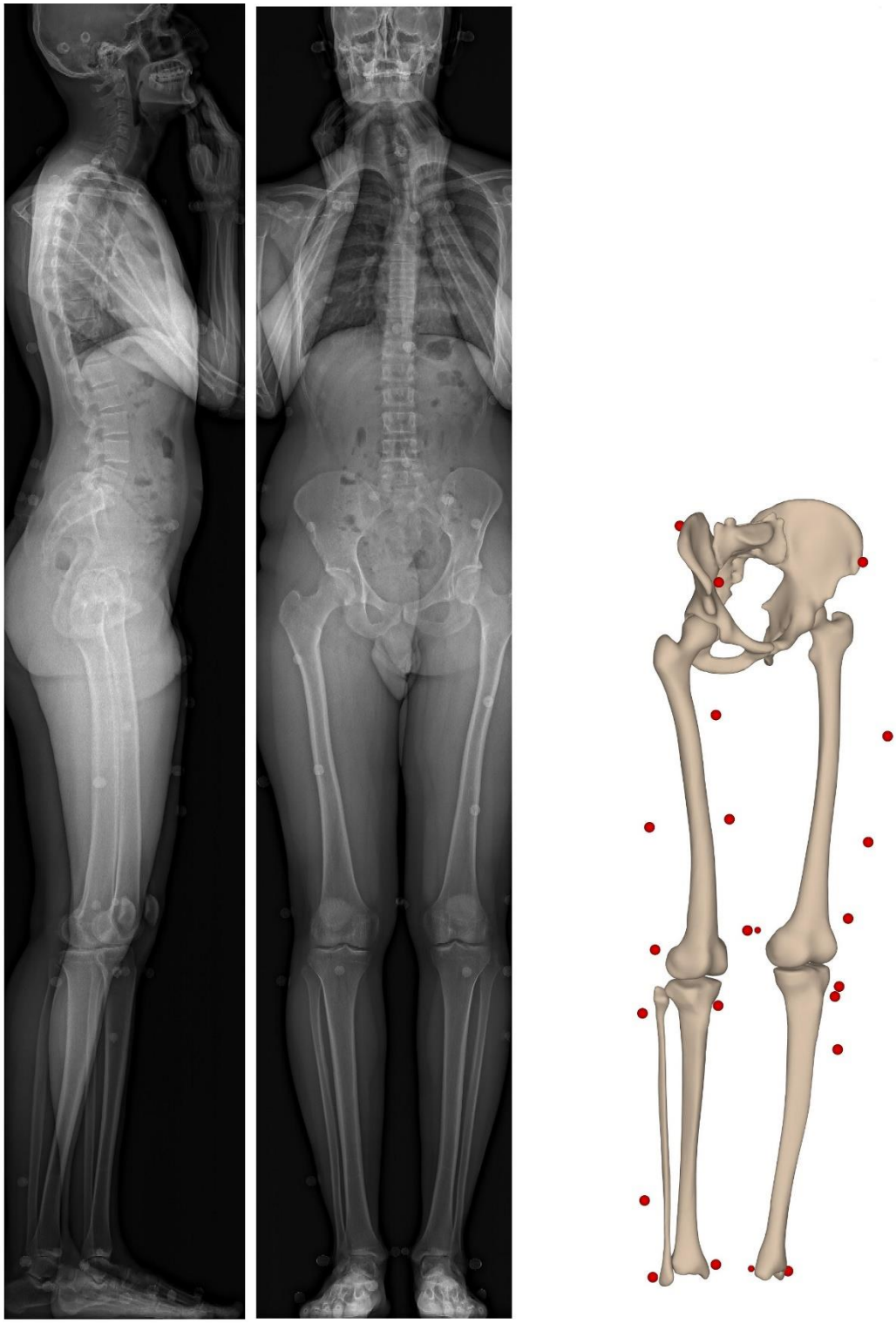

Figure 3. The 3D reconstruction of pelvis, femurs, tibias, fibulas, and markers using the biplanar X-ray images.

2.6. Data labeling

The annotation data consisted of the 3D position of 15 joint or segment centers (hereafter, for simplicity called joint centers) – head, neck, shoulders, elbows, wrists, pelvis, hips, knees, and ankles – and subjects' bounding boxes in multi-view images. The annotation data were obtained from the recorded data by the marker-based motion capture system and the EOS imaging system.

The bi-planar radiographs acquired by the EOS system were then processed as described in [23]. First, 3D shape models of the lower limbs and pelvis were reconstructed [24], [25]. Second, the reflective markers were reconstructed by fitting markers' sphere models to the markers contours visible on the bi-planar X-ray images. These two steps helped register the bony segments' anatomical frames, constructed using the biplanar X-ray images, to the marker-based motion capture system's environment (Figure 3).

The upper limbs centers, including head, neck, shoulders, elbows, and wrists, were defined based on the markers in the marker-based motion capture environment. On the other hand, the hips, knees, and ankles centers were determined using the 3D reconstructed shape models obtained using medical imaging. These key points were first defined on the 3D model then registered in the marker-based system's environment.

Once the position of the body joint centers was entirely determined with the marker-based system, the joint centers were projected into the image planes of the marker-less system's cameras. Then, for each frame of each camera, a subject' bounding box was determined using the projected joint centers.

2.7. 3D joint center detection using a convolutional neural network

The implemented deep learning-based human pose estimation method for the proposed marker-less motion capture system was presented by Iskakov et al. [13] that achieved state-of-the-art results on the Human3.6M dataset [16]. The pose estimation method consisted of two main parts, 2D pose estimation and 3D pose reconstruction. The input to the 2D pose estimation – based on a convolutional neural network ResNet-152 [26] – was a multi-view RGB image. The 3D pose reconstruction took the output of the 2D estimation, aggregated in a 3D volume, processed by a convolutional neural network called V2V-PoseNet [27], and estimated the 3D joint centers.

The new human pose dataset (hereafter, referred to as ENSAM dataset) was split randomly into training (19 subjects – 4 XLH, 1 Scoliosis) and test sets (12 subjects – 3 XLH, 1 Spondylolisthesis). The deep learning-based pose estimation method was first trained on the Human3.6M dataset, then finetuned on the ENSAM training set. Before and after finetuning, the positions of joint centers were estimated for the ENSAM test set. Agreement between the estimated joint centers and the

corresponding reference values was assessed using the Bland-Altman method by determining the mean difference (bias) and 95% limits of agreement along the axes of the global coordinate system (Figure 1). The joint position error, which determines the Euclidean distance between the estimated and reference values, was also measured. In very few cases (less than 0.5% of the total number of frames), clear outliers were detected, probably due to left-right flipped information. As these events were rare and could further be automatically detected and removed, they were arbitrarily removed.

3. Results

First, the human pose estimation was trained on the Human3.6M dataset and evaluated on the ENSAM test set. Then, the human pose estimation method was finetuned on the ENSAM training set and evaluated on the ENSAM test set. The difference between the estimated joint centers and the corresponding reference values, along the axes of the global coordinate system, is presented in Table 1. The mean differences (biases) were relatively high (within 49.8 mm) and became close to zero (within 3.1 mm) across all joints after finetuning on the ENSAM training set. Also, the 95% confidence intervals (2 standard deviations) were significantly reduced across all joints. For instance, along the Z-axis for the hip joint, the confidence interval was reduced from 106.9 mm to 17.4 mm.

Table 1. Bland-Altman bias and 95% confidence interval (2SD: 2 standard deviation) between the estimated and corresponding reference joint centers on the ENSAM test set

Axis		ankles	knees	hips	pelvis	neck	head	wrists	elbows	shoulders
		Trained on the Human3.6M dataset								
X	mean	4.6	3.4	4.7	6.3	6.3	1.3	3.8	4.9	5.4
	2SD	68.5	50.6	38.1	24.0	25.5	41.8	55.4	67.6	42.9
Y	mean	25.6	-30.5	-48.9	-49.8	-20.3	-14.4	-5.3	-19.4	34.9
	2SD	36.9	47.4	36.5	28.9	44.4	40.9	38.1	33.4	20.8
Z	mean	4.0	3.0	-0.6	-0.1	1.3	-3.8	5.1	2.7	-0.7
	2SD	30.3	29.3	106.9	11.7	12.1	14.1	90.8	37.0	75.0
		Trained on the Human3.6M dataset and fintuned on the ENSAM training set								
X	mean	0.1	-0.1	0.3	0.4	-0.3	0.6	-0.1	0.3	0.1
	2SD	13.3	16.3	26.5	18.2	18.6	31.3	27.1	34.8	31.6
Y	mean	-0.2	-1.2	-1.6	-3.1	1.5	-0.2	1.4	-1.5	-0.1
	2SD	6.6	14.1	21.9	21.8	14.8	2.9	21.1	14.0	13.4
Z	mean	-0.2	-0.1	0.3	0.4	0.4	0.3	-1.1	0.5	0.2
	2SD	5.2	8.1	17.4	7.6	8.6	9.4	24.4	12.5	39.0

There was no statistically significant difference among the joint position error of twelve subjects (p=0.62). Also, there was no statistically significant difference between the joint position error of pathological subjects and asymptomatic subjects (p=0.58). The joint position errors across all subjects are presented in Table 2.

After finetuning on the ENSAM test set, the ankles were the most accurate (mean error = 6.2 mm). The second and third most accurate joints were the knees (mean error = 10.3 mm) and the neck (joint position error = 11.3 mm). The least accurate joints were the shoulders (mean error = 21.1 mm).

Table 2. Joint position error (Euclidean distance between the estimated joint centers and corresponding reference values) on the ENSAM test set.

	ankles	knees	hips	pelvis	neck	head	wrists	elbows	shoulders
	Trained on the Human3.6M dataset								
mean	46.7	40.8	75.2	52.0	31.7	28.0	33.7	39.5	55.4
2SD	31.0	53.1	35.5	28.2	23.3	37.2	92.2	49.7	24.6
	Trained on the Human3.6M dataset and fintuned on the ENSAM training set								
mean	6.2	10.3	17.7	14.0	11.3	14.5	15.9	16.8	21.1
2SD	9.7	10.4	15.7	11.2	11.6	15.2	27.9	21.0	30.4

4. Discussion

While gait analysis is of high clinical relevance, there is a lack of methods based on RGB cameras that could be accurate enough to be considered in clinical routine. While deep learning-based pose estimation methods appear attractive, accuracy is still an issue that may be related to the non-specificity of currently used training datasets.

Indeed, publicly available datasets like the Human3.6M dataset have contributed significantly to advancing the human pose estimation methods. However, these datasets are addressed to solve computer vision problems in various situations and contexts, and they are not sufficient for particular applications. Thus, a new human pose dataset (ENSAM dataset), well-adapted for gait study, was collected. In the ENSAM dataset, despite the Human3.6M dataset, the markers were placed directly on subjects' anatomical landmarks by orthopedic surgeons. The bi-planar X-ray images from subjects helped reconstruct the 3D coordinates of the bony segments and the reflective markers, thus allowing accurate registration of joint centers' coordinates to marker-based coordinate systems. This technique reduced the errors of determining joint centers, which were mostly attributed to the marker misplacement on the anatomical landmarks [23]. Also, while the Human3.6M dataset contains the reference data for only seven young adults, the ENSAM dataset contains the walking trials of thirty-one symptomatic and pathological subjects of different age groups.

This study aimed to assess the role of the ENSAM pose dataset in the advancement and evaluation of marker-less motion capture systems. Comparing the performance of the human pose estimation method on the ENSAM test set, trained on the Human3.6M dataset, before and after finetuning on the ENSAM training set, showed a drastic improvement. For instance, for the ankle joint center, the 95% confidence intervals along the axes of the global coordinate system were within 68.5 mm.

Finetuning the human pose estimation method on the ENSAM training set reduced these values to be less than or equal to 13.3 mm. These improvements may highlight the need for a well-adapted human pose dataset for training the human pose estimation methods which are the basis for the marker-less motion capture systems.

The outliers, which were excluded, were found in three walking trials of one subject. In these three walking trials, the subject was walking in the backward direction (opposite direction of walking toward frontal cameras). These high errors mostly (but not limited to) occurred because the human pose estimation method estimated the left shoulder instead of the right shoulder and vice versa. We arbitrarily manually removed this data because, in future developments, these inconsistencies in the trajectory of a joint center could be easily identified by graphical visualization and filtered out. Except for this case, results appeared as robust. There was no statistically significant difference among the 95% confidence interval of the joint position error of subjects ($p=0.57$).

As shown in Table 2, the joint centers were not estimated with the same level of accuracy. For instance, the ankles position error was 6.2 mm, whereas the shoulder position error was 21.1 mm. Several factors may affect the estimation accuracy of different joint centers, including occlusion and the accuracy of joint centers' reference values.

Occlusion occurs when a joint is hidden by another body segment or joint. In the ENSAM pose dataset, in five out of ten walking trials, subjects were walking toward the frontal cameras. Thus, from the lateral cameras' view, the subjects' left elbow and wrist were mostly occluded by the subject's body and could be seen only by the frontal cameras. When a joint center was occluded, the accuracy of estimation was decreased. The mean joint position error of the right elbow and wrist was 14.1 mm and 12.8 mm, whereas the left elbow and wrist's joint position error was 18.8 mm and 16.2 mm.

The accuracy of joint centers' reference values can be interpreted as the consistency of joint centers' reference values across different subjects. The shoulders' reference values were retrieved from the markers placed on the acromion that were affected by the marker placement error. The human pose estimation method was trained and assessed using these reference values. Therefore, the less accurate the reference values were, the less accurate the estimations were. The shoulders' joint position errors were the highest. Also, hip joint center error (mean: 17.7 mm) was higher than pelvis center error (mean: 14.0 mm) because the hip joint centers were retrieved from the reconstructed femurs registered to the marker-based motion capture environment using the markers placed on the thighs. The soft tissue artifact associated with the thigh was greater than the pelvis [28].

This study had some limitations. The number of subjects is still limited and not representative of the range of subjects with or without musculoskeletal troubles (age <= 44 years, particularly lack of older adults). However, the protocol of data collection is now well defined and further data collection is in progress. Also, in this study, the performance of the proposed marker-less system was only assessed in terms of joint center positions. Work is in progress to evaluate the motion capture system in terms of clinically relevant gait parameters. Another limitation of this study was that the implemented deep learning-based human pose estimation could not estimate the foot centers' position, which is relevant in gait study. The development of the pose estimation method is in progress. Also, in this study, four RGB cameras were used. A fewer number of cameras may make the system easier-to-use. In future studies, various pose estimation methods with a fewer number of cameras should be analyzed.

5. Conclusion

This study proposed a new dataset that was used for training and evaluation of a marker-less motion capture system based on RGB cameras and deep learning-based human pose estimation methods. Even with a limited number of subjects in the training set (nineteen subjects), the proposed approach demonstrated the relevance of considering

a dedicated thoroughly constituted dataset. The evaluation of the marker-less system on the test set obtained promising results (mean joint position error = 14.4 mm), with potential improvements with the future extension of this dataset to a broader range of populations. Work is in progress for extensive evaluation in terms of clinically relevant gait parameters, thus paving the way for using RGB cameras-based gait analysis.

6. References

[1] M. G. Benedetti *et al.*, “SIAMOC position paper on gait analysis in clinical practice: General requirements, methods and appropriateness. Results of an Italian consensus conference,” *Gait & Posture*, vol. 58, pp. 252–260, Oct. 2017, doi: 10.1016/j.gaitpost.2017.08.003.

[2] M. Ferrarin *et al.*, “Does gait analysis change clinical decision-making in poststroke patients? Results from a pragmatic prospective observational study,” *Eur J Phys Rehabil Med*, vol. 51, no. 2, pp. 171–84, 2015.

[3] T. A. L. Wren, G. E. Gorton, S. Õunpuu, and C. A. Tucker, “Efficacy of clinical gait analysis: A systematic review,” *Gait & Posture*, vol. 34, no. 2, pp. 149–153, Jun. 2011, doi: 10.1016/j.gaitpost.2011.03.027.

[4] K. Paterson, K. Hill, and N. Lythgo, “Stride dynamics, gait variability and prospective falls risk in active community dwelling older women,” *Gait & Posture*, vol. 33, no. 2, pp. 251–255, Feb. 2011, doi: 10.1016/j.gaitpost.2010.11.014.

[5] M. J. P. Toebes, M. J. M. Hoozemans, R. Furrer, J. Dekker, and J. H. van Dieën, “Local dynamic stability and variability of gait are associated with fall history in elderly subjects,” *Gait & Posture*, vol. 36, no. 3, pp. 527–531, Jul. 2012, doi: 10.1016/j.gaitpost.2012.05.016.

[6] S. R. Simon, “Quantification of human motion: gait analysis—benefits and limitations to its application to clinical problems,” *Journal of Biomechanics*, vol. 37, no. 12, pp. 1869–1880, Dec. 2004, doi: 10.1016/j.jbiomech.2004.02.047.

[7] R. A. Clark, B. F. Mentiplay, E. Hough, and Y. H. Pua, “Three-dimensional cameras and skeleton pose tracking for physical function assessment: A review of uses, validity, current developments and Kinect alternatives,” *Gait & Posture*, vol. 68, pp. 193–200, Feb. 2019, doi: 10.1016/j.gaitpost.2018.11.029.

[8] S. Springer and G. Yogev Seligmann, “Validity of the Kinect for Gait Assessment: A Focused Review,” *Sensors*, vol. 16, no. 2, Art. no. 2, Feb. 2016, doi: 10.3390/s16020194.

[9] A. Pfister, A. M. West, S. Bronner, and J. A. Noah, “Comparative abilities of Microsoft Kinect and Vicon 3D motion capture for gait analysis,” *Journal of Medical Engineering & Technology*, vol. 38, no. 5, pp. 274–280, Jul. 2014, doi: 10.3109/03091902.2014.909540.

[10] X. Xu, R. W. McGorry, L.-S. Chou, J. Lin, and C. Chang, “Accuracy of the Microsoft Kinect™ for measuring gait parameters during treadmill walking,” *Gait & Posture*, vol. 42, no. 2, pp. 145–151, Jul. 2015, doi: 10.1016/j.gaitpost.2015.05.002.

[11] Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei, and Y. Sheikh, “OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields,” *arXiv:1812.08008 [cs]*, May 2019, Accessed: Nov. 12, 2020. [Online]. Available: http://arxiv.org/abs/1812.08008.

[12] N. B. Gundavarapu, D. Srivastava, R. Mitra, A. Sharma, and A. Jain, “Structured Aleatoric Uncertainty in Human Pose Estimation.,” in *CVPR Workshops*, 2019, vol. 2.

[13] K. Iskakov, E. Burkov, V. Lempitsky, and Y. Malkov, “Learnable Triangulation of Human Pose,” 2019, pp. 7718–7727, Accessed: Nov. 12, 2020. [Online]. Available: https://openaccess.thecvf.com/content_ICCV_2019/html/Iskakov_Learnable_Triangulation_of_Human_Pose_ICCV_2019_paper.html.

[14] H. Qiu, C. Wang, J. Wang, N. Wang, and W. Zeng, “Cross View Fusion for 3D Human Pose Estimation,” 2019, pp. 4342–4351, Accessed: Nov. 12, 2020. [Online]. Available: https://openaccess.thecvf.com/content_ICCV_2019/html/Qiu_Cross_View_Fusion_for_3D_Human_Pose_Estimation_ICCV_2019_paper.html.

[15] L. Sigal, A. O. Balan, and M. J. Black, “Humaneva: Synchronized video and motion capture dataset and baseline algorithm for evaluation of articulated human motion,” *International journal of computer vision*, vol. 87, no. 1–2, p. 4, 2010.

[16] C. Ionescu, D. Papava, V. Olaru, and C. Sminchisescu, “Human3.6M: Large Scale Datasets and Predictive Methods for 3D Human Sensing in Natural Environments,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 7, pp. 1325–1339, Jul. 2014, doi: 10.1109/TPAMI.2013.248.

[17] G. Kojima *et al.*, “Does the timed up and go test predict future falls among British community-dwelling older people? Prospective cohort study nested within a randomised controlled trial,” *BMC Geriatrics*, vol. 15, no. 1, p. 38, Apr. 2015, doi: 10.1186/s12877-015-0039-7.

[18] J. Heikkila and O. Silven, “A four-step camera calibration procedure with implicit image correction,” in *Proceedings of ieee computer society conference on computer vision and pattern recognition*, 1997, pp. 1106–1112.

[19] C. Sauret, H. Pillet, W. Skalli, and M. Sangeux, “On the use of knee functional calibration to determine the medio-lateral axis of the femur in gait analysis: Comparison with EOS biplanar radiographs as reference,” *Gait & Posture*, vol. 50, pp. 180–184, Oct. 2016, doi: 10.1016/j.gaitpost.2016.09.008.

[20] M. Sangeux, H. Pillet, and W. Skalli, “Which method of hip joint centre localisation should be used in gait analysis?,” *Gait & Posture*, vol. 40, no. 1, pp. 20–25, May 2014, doi: 10.1016/j.gaitpost.2014.01.024.

[21] A. Leardini, Z. Sawacha, G. Paolini, S. Ingrosso, R. Nativo, and M. G. Benedetti, “A new anatomically based protocol for gait analysis in children,” *Gait & Posture*, vol. 26, no. 4, pp. 560–571, Oct. 2007, doi: 10.1016/j.gaitpost.2006.12.018.

[22] R. B. Davis, S. Õunpuu, D. Tyburski, and J. R. Gage, “A gait analysis data collection and reduction technique,” *Human Movement Science*, vol. 10, no. 5, pp. 575–587, Oct. 1991, doi: 10.1016/0167-9457(91)90046-Z.

[23] H. Pillet, M. Sangeux, J. Hausselle, R. El Rachkidi, and W. Skalli, “A reference method for the evaluation of femoral head joint center location technique based on external markers,” *Gait & Posture*, vol. 39, no. 1, pp. 655–658, Jan. 2014, doi: 10.1016/j.gaitpost.2013.08.020.

[24] Y. Chaibi *et al.*, “Fast 3D reconstruction of the lower limb using a parametric model and statistical inferences and clinical measurements calculation from biplanar X-rays,” *Computer Methods in Biomechanics and Biomedical Engineering*, vol. 15, no. 5, pp. 457–466, May 2012, doi: 10.1080/10255842.2010.540758.

[25] D. Mitton *et al.*, “3D reconstruction of the pelvis from bi-planar radiography,” *Computer Methods in Biomechanics and Biomedical Engineering*, vol. 9, no. 1, pp. 1–5, Feb. 2006, doi: 10.1080/10255840500521786.

[26] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” *arXiv:1512.03385 [cs]*, Dec. 2015, Accessed: Nov. 13, 2020. [Online]. Available: http://arxiv.org/abs/1512.03385.

[27] G. Moon, J. Y. Chang, and K. M. Lee, “V2V-PoseNet: Voxel-to-Voxel Prediction Network for Accurate 3D Hand and Human Pose Estimation From a Single Depth Map,” 2018, pp. 5079–5088, Accessed: Nov. 12, 2020. [Online]. Available: https://openaccess.thecvf.com/content_cvpr_2018/html/Moon_V2V-PoseNet_Voxel-to-Voxel_Prediction_CVPR_2018_paper.html.

[28] A. Leardini, L. Chiari, U. D. Croce, and A. Cappozzo, “Human movement analysis using stereophotogrammetry: Part 3. Soft tissue artifact assessment and compensation,” *Gait & Posture*, vol. 21, no. 2, pp. 212–225, Feb. 2005, doi: 10.1016/j.gaitpost.2004.05.002.