AZURE KINECT BODY TRACKING UNDER REVIEW FOR THE SPECIFIC CASE OF UPPER LIMB EXERCISES

EUGENIO IVORRA, MARIO ORTEGA, MARIANO ALCANIZ

Institute for Research and Innovation in Bioengineering, Universitat Politecnica de Valencia, Valencia, Spain

DOI: 10.17973/MMSJ.2021_6_2021012

e-mail: euivmar@i3b.upv.es

A tool for human pose estimation and quantification using consumer-level equipment is a long-pursued objective. Many studies have employed the Microsoft Kinect v2 depth camera but with recent release of the new Kinect Azure a revision is required. This work researches the specific case of estimating the range of motion in five upper limb exercises using four different pose estimation methods. These exercises were recorded with the Kinect Azure camera and assessed with the OptiTrack motion tracking system as baseline. The statistical analysis consisted of evaluation of intra-rater reliability with intra-class correlation, the Pearson correlation coefficient and Bland–Altman statistical procedure. The modified version of the OpenPose algorithm with the post-processing algorithm PoseFix had excellent reliability with most intra-class correlations being over 0.75. The Azure body tracking algorithm had intermediate results. The results obtained justify clinicians employing these methods, as quick and low-cost simple tools, to assess upper limb angles.

KEYWORDS

Human pose estimation, Microsoft Azure Kinect, Upper limb exercises, OptiTrack system .

1 INTRODUCTION

Human balance disturbances are common disorders in different populations. Previous reports have shown the contribution of the upper limb to human postural control; for example, upper limb immobilisation or fatiguing arm exercises have shown that decreases in upper limb function negatively affect postural control [Souza et al., 2016].Therefore, the assessment of upper limb disorders is crucial for elaborating proper rehabilitation or for treatment diagnosis.

Upper limb movements have usually been assessed through traditional tests and scales. The Wolf Motor Function Test, the Action Research Arm Test and the Melbourne Assessment are valid tools for measuring the quality of upper limb movement [Randall et al., 2001, van Wegen et al., 2010]. However, the majority of these scales may be biased and could introduce a subjective component due to the therapist's level of experience and are frequently time-consuming. In order to overcome the limitations of the traditional scales, instrumented systems have been developed and have appeared in the market. In clinical practice, the goniometer is a widely used instrument to measure range of motion. It is regarded as a simple, versatile, and easy to-use instrument. Reports indicate that its accuracy is highly dependent on the level of assessor experience and the anatomical joint being measured. It is also limited to measuring joint angles in single planes and static positions [Walmsley et al., 2018].

Currently, quantitative measurements of upper limb functions are normally performed using optical markerbased motion capture systems [Cai et al., 2019]. They use reflective landmarks positioned on the body that are detected by optical cameras located in the tracking area. Optical systems are considered gold standard due to their high accuracy in detecting human poses and movements [Aurand et al., 2017].

Nonetheless, the high cost (€20,000), the amount of space required, and, in some cases, the substantial amount of time spent preparing the subject for assessment make the introduction of these systems in clinics and daily routines difficult [Oh et al., 2018]. Cameras with the ability to detect depth and colour (RGB-D cameras), one of the earliest being Microsoft Kinect v1 released in 2011, have previously been used and validated for human pose estimation (HPE) [Cai et al., 2019, Eltoukhy et al., 2017, Oh et al., 2018]. In this study, a Microsoft Azure Kinect camera is employed because it is one of the most modern and well-known RGB-D cameras.

The solution to the problem of HPE, i.e. the problem of localisation of human joints, has recently made significant progress as a result of the use of convolutional neural networks (CNN) in images. The state-of-the-art method OpenPose [Cao et al., 2017] is able to perform 8.8 FPS HPE of 19 persons in the image by employing the Nvidia 1080 GTX GPU. In fact, this method has already been validated as a motion analysis method, but only for bilateral squat exercise [Ota et al., 2020]. Two subsequent methods, Mask RCNN [He et al., 2017] and AlphaPose [Fang et al., 2017], have made small improvements to the mean average precision (mAP) metric, but at the cost of slower runtime. Osokin modified the CNN of the OpenPose algorithm for making it more computationally efficient [Osokin, 2018]. This modified OpenPose algorithm is employed in this paper and is referred to as OpenPoseMod.

Obtaining the 3D skeletal pose from a monocular RGB image is a much harder challenge than 2D attempted by fewer methods [Bogo et al., 2016, Martinez et al., 2017]. Unfortunately, these methods are typically offline or do not provide predictions in real world units [Mehta et al., 2017a]. This could be solved by an additional depth channel provided by RGB-D sensors that overcomes forward–backward ambiguities in monocular pose estimation. The best-known studies are those based on Microsoft Kinect software development kit (SDK) [Wang et al., 2015] that exploits temporal and kinematic constraints to result in a smooth output skeleton. This system is popular in clinical practice because it gives good results and is easy to use due to its free SDK which is able to track the movement of human joints without markers. This tool has been employed and validated in numerous clinical applications such as gait and motion analysis and shoulder joint angle or jump-landing kinematics [Asaeda et al., 2018, Eltoukhy et al., 2017, Valevicius et al., 2019].

To the best of our knowledge, this is the first study to analyse the new Microsoft Azure Kinect Body tracking technology with two recent 3D human pose algorithms using depth information and only colour information for upper limb movements (OpenPoseMod and RGB-3DHP). For this comparison, a baseline test was performed with the optical marker-based OptiTrack system.

2 METHODS

2.1 Participants

Thirty healthy individuals, 20 men and 10 women, participated in this study. They had no known musculoskeletal or vestibular disease. The Universitat Politècnica de València granted ethical approval for the study. All participants in the study signed an informed consent form. The mean age of the participants was 31.5 ± 10.3 years. The participants did not present any mobility impairment. Their average height was 1.7 meters and the average weight was 70.6 kg, and 22 were righthanded.

2.2 *Instrumentation and procedures*

2.2.1 Human pose estimation methods

Four different HPE methods were investigated and tested in this study: OptiTrack as the baseline; Azure Kinect body tracking method as the modern update of the commonly used Kinect v2 employed in numerous rehabilitation studies and clinical trials; OpenPoseMod as a CNN that employs an RGB-D camera for obtaining 3D human poses; and an RGB-based human pose estimation that leverages state-of-the-art algorithms in the field.

The first HPE method employed was the OptiTrack motion capture system. OptiTrack typically generates less than 0.2 mm of measurement error; thus, it is considered the gold standard [Nagymáté and Kiss, 2018]. The setup employed included 28 Prime 13 cameras distributed equally along two levels of height. This setup is able to capture a volume of 12 × 6 × 2 meters and costs around €23,000. The skeleton obtained with the OptiTrack human motion tracking proprietary algorithm corresponds to the Rizzoli marker set protocols [Leardini et al., 2007].

The second HPE method was the Microsoft Azure Kinect body tracking method (Azure). Azure is able to obtain a 3D human skeleton composed of 32 3D joints with the coordinates. It presumably provides better tracking results compared to Kinect v1 and v2 and can record more joints (from 20–25 to 32). This technology employs a deep-learning algorithm (not published) to estimate the 3D joints for each RGB-D image [Shotton et al., 2013] using only depth information. It works in real time with consumer equipment with a dedicated graphical processing unit.

The third HPE method was a modified OpenPosebased algorithm (OpenPoseMod) that estimates first a 2D human pose using a CNN called Lightweight OpenPose [Osokin, 2018] and then depth information and the corresponding 3D human pose for each person. In order to enhance the precision and accuracy of OpenPoseMod, after estimation of the 2D human pose, the next step is a post-processing filter using the state-of-the-art method PoseFix [Moon et al., 2019]. PoseFix reported an increase from 64.2% average precision of the original OpenPose up to 76.7 mean average precision with the public Human Pose Microsoft COCO dataset [Lin et al., 2014].

Once the 2D human points of each joint have been extracted and refined, the 3D points of each can be easily and efficiently extracted with the information from the depth image. The complete procedure for obtaining the 3D human pose with OpenPoseMod is synthesised in Fig 1.

The fourth HPE method was RGB 3D human pose estimation (RGB-3DHP). As shown in Fig 1, it is composed of two parts. The first infers the 2D joint locations and joint detection confidences (heatmap) using the Lightweight OpenPose CNN and the second estimates occlusion-robust pose maps and infers the 3D pose of the joints. This method, which is defined in [Mehta et al., 2018], is a further development of the Vnect [Mehta et al., 2017b] algorithm.

These 3D human pose estimation techniques were calculated offline using the recorded media. However, the processing time was evaluated in order to determine if it was possible to employ the particular technology in real time. A personal computer with an Intel Xeon W3225 CPU, Nvidia Titan RTX and 128G RAM memory was employed for testing. The results shown in Table 1 were obtained by averaging the processing time of all the subjects and exercises. According to these results and the assumption that real time was more than 20 fps, we can affirm that the only method that does not work in real time is OpenPoseMod with PoseFix. However, OpenPoseMod without PoseFix can work up to 35 FPS vs the 8.8 FPS of the original OpenPose implementation.

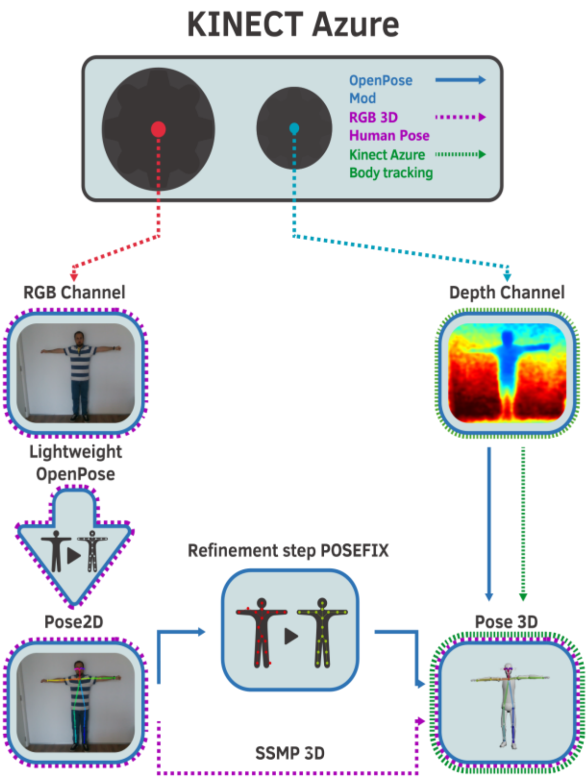


Figure 1. Schematic diagram of the procedure for estimating the 3D human skeleton with three methods

A summary of the comparison of these four 3D HPE methods is shown in Tab 1.

	Opti-Track	Azure Kinect body tracking	Open Pose-Mod	OpenPose Mod + PoseFix	RGB-3DHP
Algorithm	Infrared marker tracking	CNN using depth image	CNN using RGB and depth image	CNN using RGB and depth image	CNN using RGB
2D keypoints	No	Yes, project the 3D to 2D	Yes	Yes	Yes
3D keypoints	Yes	Yes	Yes, infers 3D using a depth image	Yes, infers 3D using a depth image	Yes, infers the 3D using a CNN
Calib required	Yes	No	No	No	No
Cost €	20.000	399	180	180	50
Hardware	NIR	Microsoft	RGB-	RGB-D	RGB

	camera	Kinect Azure	D camera	camera	camera
FPS	120 ±1	26 ±2	35 ±4	8 ± 2	21 ±3

Table 1. A summary of the comparison of 3D human pose estimation methods

2.2.2 Procedure

Participants were asked to wear a black suit on which optical markers were positioned in order to obtain the 3D position of these landmarks in a virtual world through OptiTrack technology. These subjects performed an exercise set to build an average model of the exercises that could be used for checking the results of other technologies. Later, participants were asked to perform the same exercise set in their own clothes. The participants were assessed in four different areas without obstacles and repeated the exercises three times.The exercises were recorded using the Microsoft Azure Kinect RGB-D camera and the recorded media were processed with the three different HPE methods (OpenposeMod, Azure and RGB-3DHP). This way, the coordinate system, timeline and conditions were equal for all of them. The participants did not perform the exercises simultaneously with OptiTrack because the other three technologies lost accuracy significantly due to the black suit with optical markers. A manual temporal and spatial synchronisation were performed to transform the OptiTrack results with the other three HPE methods. Microsoft Azure Kinect camera was located at 2.5 meters from the participant and 0.9 m over the floor during the experimental tests and was configured to record at 1280 × 720 colour resolution, narrow field-of-view unbinned mode with 640 × 576 for the depth mode and at 30 FPS framerate.

The exercise set was composed of five exercises designed for assessment of numerous rehabilitation parameters, especially for the joints of upper body parts. Exercises lasted between 20 and 40 seconds each with 30 seconds of break between them. Please check supplementary materials for a graphical description. These exercises were

1. Shoulder abduction in the frontal plane (shABfp).
2. Flexion of the shoulder in the sagittal plane (shFLsg).
3. Flexion of the elbow in the sagittal plane (elFLsg).
4. External rotation of the shoulder in the zenith plane (shROTzp).
5. Horizontal flexion of the shoulder in the zenith plane (shFLzp).

2.3 Statistical and data analysis

The raw 3D joint positions were smoothed for all HPE methods with a rolling window median of two seconds. Then, kinematic parameters were calculated as the average performance of all the subjects for each technology. The kinematic parameters defined the range of motion (minimum, maximum and difference) of angles between the joints involved in the movement. Shoulder and elbow angles were calculated following the international standards defined by the International Society of Biomechanics [Wu et al., 2005].

Pearson correlation coefficients were calculated to determine the concurrent validity of the three technologies with the OptiTrack baseline at an alpha value of 0.05. In addition, intra-class correlation (ICC) for model (2,k) was calculated to consider the consistency of the within-subject agreement between systems, taking into account possible systematic errors. ICC was considered poor if it was < 0.4, fair if 0.4–0.6, good if 0.6–0.75 and excellent if ≥ 0.75 [McGinley et al., 2009].

Finally, Bland–Altman analysis [Bland and Altman, 2010] was also performed. It is commonly employed to compare two methods of measurement and interpret findings to determine whether a new method of measurement could replace an existing accepted 'gold-standard' method [Ota et al., 2020]. The statistical analyses were performed using the Matlab R2019B computational environment and Microsoft Excel 2016.

3 RESULTS

The ICC obtained for each HPE method and for each exercise are shown in Tab. 2. To determine which exercises were better calculated and which worst for each HPE method, the difference from OptiTrack was also determined (Fig. 2).

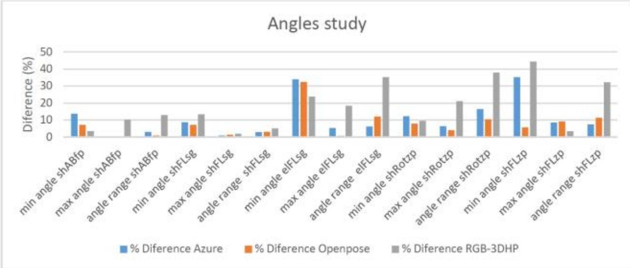


Figure 2. Percentage difference in the upper link angle between OptiTrack and other technologies. The lower the percentage the better the technology.

Exer	Kinematic param	Azure			OpenposeMod			RGB-3DHP		
		L	I	H	L	I	H	L	I	H
1	min angle shABfp	0.63	0.80	0.83	0.71	**0.84**	0.89	0.49	0.79	0.91
	max angle shABfp	0.74	**0.90**	0.97	0.76	**0.90**	0.95	0.47	0.75	0.86
	angle range shABfp	0.72	0.89	0.94	0.77	**0.92**	0.97	0.46	0.74	0.85
2	min angle shFLsg	0.55	0.72	0.72	0.61	0.72	0.76	0.46	**0.74**	0.85
	max angle shFLsg	0.74	0.91	0.97	0.78	**0.92**	0.97	0.50	0.81	0.93
	angle range shFLsg	0.70	0.86	0.91	0.74	**0.88**	0.93	0.49	0.80	0.91
3	min angle elFLsg	0.63	**0.79**	0.82	0.46	0.54	0.57	0.44	0.70	0.81
	max angle elFLsg	0.71	0.88	0.93	0.77	**0.91**	0.96	0.44	0.71	0.81
	angle range elFLsg	0.73	**0.89**	0.95	0.65	0.77	0.81	0.38	0.61	0.70
4	min angle shRotzp	0.64	0.81	0.84	0.76	**0.90**	0.95	0.46	0.74	0.84
	max angle shRotzp	0.60	0.77	0.79	0.69	**0.82**	0.87	0.38	0.61	0.70
	angle range shRotzp	0.48	0.63	0.65	0.64	**0.76**	0.80	0.29	0.47	0.54
5	min angle shFLzp	0.45	0.6	0.63	0.68	**0.81**	0.86	0.33	0.52	0.60

max angle shFLzp	0.69	0.86	0.90	0.73	**0.87**	0.92	0.49	0.79	0.91
angle range shFLzp	0.62	**0.79**	0.81	0.63	0.75	0.79	0.35	0.57	0.65

Table 2. Intra-class correlation (ICC) for kinematic parameters calculated for upper limb joints. In each row, the best value is in bold. (L- Lower, I-ICC,H-Higher)

The five tasks results average of the % difference for Azure, OpenPoseMod and RGB-3DHP were 10.7%, 7.6% and 18.2%, respectively. These results are consistent with the Root Mean Square Deviation (RMSD) of 10 degrees for Azure, 8 for OpenPoseMod and 22 for RGB-3DHP and Pearson correlation coefficients of 0.979, 0.988 and 0.941 (all p<0.05), respectively. These values show that there was a strong correlation, but there was still some difference in accuracy, between OptiTrack and the RGB-D camera-based methods.

Finally, the results of the Bland–Altman analysis are shown in Fig. 3. In these three methods, 95% of the measures were inside the limits of agreement (LOA). A perfect match between methods would give a mean of 0, and a smaller LOA means better adjustment.

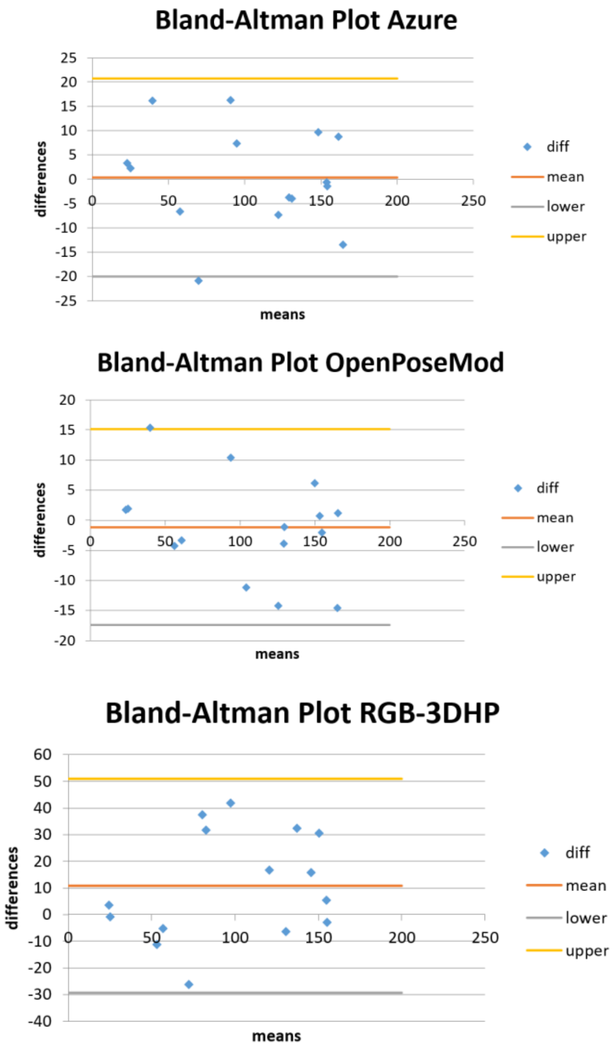


Figure 3. Bland–Altman plot for upper-limb exercises for each technology.

4 DISCUSSION

4.1 Discussion of results

The purpose of this study was to evaluate and compare the performance of the Azure body tracking algorithm with two alternative human pose estimation algorithms using OptiTrack system as benchmark. From the analysis shown in Fig 2, it can be concluded that the results greatly depend on the type of exercise. These measures are used for example in rehabilitation exercises to assess the degree of recovery from injuries that limit the range of motion. Currently, these angles are clinical measured using goniometers; therefore, a fast and accurate camera-based method would be a significant improvement. Exercises with smaller differences from OptiTrack are those in the frontal plane, parallel to the camera sensor and those with wide movements. The algorithms had worse pose estimations when own subjects' clothes happened to be black, had their arms tightly against the body or perfectly aligned with a normal vector from the camera. Exercise 4 had two of these problems which could be rectified by changing the camera setup position (or using another camera) so the movement was on the frontal plane. Moreover, previous studies such as [Bonnechere et al., 2014] have reported poor results when using the Kinect for measuring the elbow angle on the sagittal plane.

Azure had an excellent ICC (Tab. 2) except for measuring shoulder rotation in zenith plane which was rated good (range 0.63–0.81). These results are lower to those obtained by [Cai et al., 2019], with ratings of 0.59–0.96 for shoulder motions measured by Kinect v2. The absolute mean error for Azure was 10.7 degrees, higher than the 6 degrees reported by [Shotton et al., 2013]for different upper limb movements measured with Kinectv2 or the 7.6 degrees of error of [Wiedemann et al., 2015] also with Kinect v2. These results were recently corroborated by Albert et al [Albert et al., 2020] who reported during a gait analysis that the Kinect v2 performed better than the Azure in the mid and upper body region, especially in the upper extremities. However, a specific study should be performed with the same conditions and exercises in order to conclude if it is meaningful to upgrade to Kinect Azure from Kinect v2. Moreover, overall Azure results are lower than those obtained for OpenPoseMod with PoseFix, but at the cost of offline processing. With the overall results reviewed, it can be concluded that the best RGB-D camera method of the three analysed in this study is OpenPoseMod. This method can be employed as an effective alternative to the traditional goniometer or the expensive OptiTrack.

Based on the results of this study, several recommendations can be made. It is important to be aware of the advantages and disadvantages of each technology. For example, if the illumination cannot be controlled or the scenario is cluttered, it is a good idea to employ the Azure algorithm because it is only based on depth information calculated using the infrared spectrum. On the other hand, OpenPoseMod showed better performance, mainly due to the employment of PoseFix, when the acquisition scenery was controlled. RGB-3DHP could be also employed, at a very low cost, for some exercises because it only needs a common RGB camera, although its accuracy is lower than that of the other two methods. This method is recommended for ludic applications or motivational games in rehabilitation with animated avatars like [Tannous et al., 2016]. Fig 4 summarises the recommendations depending on application requirements. It is important to remark that this diagram shows our recommended method for each context but it does not imply that other method is not valid, for example, Kinect Azure could be also employed when there is good illumination and real time is required.

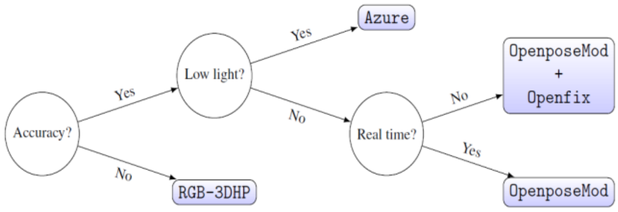


Figure 4. Decision tree for recommendations from the RGB-D methods for upper limb joint estimation

When using these technologies, we advise averaging multiple repetitions of the same subject, avoiding direct sunlight and asking the subjects not to wear dark or reflective clothing. Another important recommendation is to set the camera up so its sensor is parallel as much as possible to the plane of the movement, or to use two or more cameras to acquire images from the best angle possible.

4.2 Limitations and future works

Our results are specific to a population of healthy participants so results may vary with subjects with mobility restrictions. Although OptiTrack was considered as the baseline because it is a reference measurement system in clinical context these systems have potential sources of errors. The main potential source of error is soft tissue artefacts. This could happen when markers on the skin moves and no longer match the underlying anatomical bone landmarks. However, there are algorithms that could be applied in future works that can decrease these errors like the work of [Cutti et al., 2006].

As part of future work, this study will be extended to analyse temporal kinematics parameters and compare the full trajectory and not only ROM. Also, we will explore the influence of a multicamera synced setup. This would lower the error due to self-occlusion and give more reliable estimations averaging the measures from different cameras as done by the OptiTrack system.

We have also planned to explore the potential use of Kinect Azure for real time lower limbs measuring. There are many hand tracking possibilities but the number of methods to track lower limbs is very low in comparison and could be very useful for example for using VR in real time.

Another important point to explore is the use of tracking algorithms like the Kalman filter [Rodriguez, 1987] in order to improve temporal stability and give smoother trajectories without the need of a post-processing step. Finally, a Kinematic model could be introduced, as it is done in OptiTrack system, to take into consideration bone segment of constant lengths and joints with limited number of degrees of freedom.

5 CONCLUSION

RGB-D-based 3D human pose estimation has advanced significantly due to new deep-learning algorithms and more accurate depth and colour sensors. Optical markers are still the gold-standard method but this study shows that the distance between technologies has been reduced. This original research should be of interest to a broad readership, including those interested in methods of human pose estimation, innovative solutions with the Microsoft Kinect Sensor and rehabilitation monitoring sensors.OpenPoseMod, Microsoft Azure body tracking and RGB-3DHP algorithms could be used as low-cost alternatives to laboratory-grade systems under certain circumstances for multiple practical applications like rehabilitation exercises of the upper limb, physiotherapy or monitoring postural hygiene in fitness centres.

6 CONFILICST OF INTEREST

The authors declare that there are no conflicts of interest.

ACKNOWLEDGMENTS

THE OPTITRACK 3D CAPTURE MOVEMENT SYSTEM WAS FUNDED BY THE EUROPEAN UNION THROUGH THE ERDF (EUROPEAN REGIONAL DEVELOPMENT FUND) PROGRAM OF THE VALENCIAN COMMUNITY 2014-2020 (IDIFEDER/2018/029)

REFERENCES

[Albert et al., 2020] Albert, J. A., Owolabi, V., Gebel, A., Brahms, C. M., Granacher, U., Arnrich, B.: 'Evaluation of the pose tracking performance of the azure kinect and kinect v2 for gait analysis in comparison with a gold standard: A pilot study'; Sensors, 20, 18 (2020), 5104.

[Asaeda et al., 2018] Asaeda, M., Kuwahara, W., Fujita, N., Yamasaki, T., Adachi, N.: 'Validity of motion analysis using the Kinect system to evaluate single leg stance in patients with hip disorders'; Gait & Posture, 62 (2018), 458–462. https://doi.org/10.1016/j.gaitpost.2018.04.010

[Aurand et al., 2017] Aurand, A. M., Dufour, J. S., Marras, W. S.: 'Accuracy map of an optical motion capture system with 42 or 21 cameras in a large measurement volume'; J Biomech, 58 (2017), 237–40.

[Bland and Altman, 2010] Bland, J. M., Altman, D. G.: 'Statistical methods for assessing agreement between two methods of clinical measurement'; International Journal of Nursing Studies, 47, 8 (2010), 931–936.

[Bogo et al., 2016] Bogo, F., Kanazawa, A., Lassner, C., Gehler, P., Romero, J., Black, M. J.: 'Keep it SMPL: Automatic estimation of 3D human pose and shape from a single image'; Presented at the European Conference on Computer Vision, Springer (2016), 561–578.

[Bonnechere et al., 2014] Bonnechere, B., Jansen, B., Salvia, P., Bouzahouene, H., Omelina, L., Moiseev, F., et al.: 'Validity and reliability of the Kinect within functional assessment activities: comparison with standard stereophotogrammetry'; Gait & Posture, 39, 1 (2014), 593–598.

[Cai et al., 2019] Cai, L., Ma, Y., Xiong, S., Zhang, Y.: 'Validity and Reliability of Upper Limb Functional Assessment Using the Microsoft Kinect V2 Sensor'; Applied Bionics and Biomechanics, 2019 (2019), 7175240. https://doi.org/10.1155/2019/7175240

[Cao et al., 2017] Cao, Z., Simon, T., Wei, S.-E., Sheikh, Y.: 'Realtime multi-person 2d pose estimation using part affinity fields'; Presented at the Proceedings of the IEEE

Conference on Computer Vision and Pattern Recognition (2017), 7291–7299.
[Cutti et al., 2006] Cutti, A. G., Cappello, A., Davalli, A.: 'In vivo validation of a new technique that compensates for soft tissue artefact in the upper-arm: preliminary results'; Clinical Biomechanics, 21 (2006), S13–S19.
[Eltoukhy et al., 2017] Eltoukhy, M., Kuenze, C., Oh, J., Wooten, S., Signorile, J.: 'Kinect-based assessment of lower limb kinematics and dynamic postural control during the star excursion balance test'; Gait & Posture, 58 (2017), 421–427. https://doi.org/10.1016/j.gaitpost.2017.09.010
[Fang et al., 2017] Fang, H.-S., Xie, S., Tai, Y.-W., Lu, C.: 'Rmpe: Regional multi-person pose estimation'; Presented at the Proceedings of the IEEE International Conference on Computer Vision (2017), 2334–2343.
[He et al., 2017] He, K., Gkioxari, G., Dollár, P., Girshick, R.: 'Mask r-cnn'; Presented at the Computer Vision (ICCV), 2017 IEEE International Conference on, IEEE (2017), 2980–2988.
[Leardini et al., 2007] Leardini, A., Sawacha, Z., Paolini, G., Ingrosso, S., Nativo, R., Benedetti, M. G.: 'A new anatomically based protocol for gait analysis in children'; Gait & Posture, 26, 4 (2007), 560–571.
[Lin et al., 2014] Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., et al.: 'Microsoft coco: Common objects in context'; In European conference on computer vision (2014), 740–755.
[Martinez et al., 2017] Martinez, J., Hossain, R., Romero, J., Little, J. J.: 'A simple yet effective baseline for 3d human pose estimation'; Presented at the Proceedings of the IEEE International Conference on Computer Vision (2017), 2640–2649.
[McGinley et al., 2009] McGinley, J. L., Baker, R., Wolfe, R., Morris, M. E.: 'The reliability of three-dimensional kinematic gait measurements: a systematic review'; Gait & Posture, 29, 3 (2009), 360–369.
[Mehta et al., 2017a] Mehta, D., Rhodin, H., Casas, D., Fua, P., Sotnychenko, O., Xu, W., Theobalt, C.: 'Monocular 3d human pose estimation in the wild using improved cnn supervision'; Presented at the 3D Vision (3DV), 2017 International Conference on, IEEE (2017a), 506–516.
[Mehta et al., 2018] Mehta, D., Sotnychenko, O., Mueller, F., Xu, W., Sridhar, S., Pons-Moll, G., Theobalt, C.: 'Single-shot multi-person 3d pose estimation from monocular rgb'; Presented at the 2018 International Conference on 3D Vision (3DV), IEEE (2018), 120–130.
[Mehta et al., 2017b] Mehta, D., Sridhar, S., Sotnychenko, O., Rhodin, H., Shafiei, M., Seidel, H.-P., et al.: 'Vnect: Real-time 3d human pose estimation with a single rgb camera'; ACM Transactions on Graphics (TOG), 36, 4 (2017b), 44.
[Moon et al., 2019] Moon, G., Chang, J. Y., Lee, K. M.: 'Posefix: Model-agnostic general human pose refinement network'; Presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2019), 7773–7781.
[Nagymate and Kiss, 2018] Nagymate, G., Kiss, R. M.: 'Application of OptiTrack motion capture systems in human movement analysis'; Recent Innovations in Mechatronics, 5, 1. (2018), 1–9.
[Oh et al., 2018] Oh, J., Kuenze, C., Jacopetti, M., Signorile, J. F., Eltoukhy, M.: 'Validity of the Microsoft Kinect™ in assessing spatiotemporal and lower extremity kinematics during stair ascent and descent in healthy young individuals'; Medical Engineering & Physics, 60 (2018), 70–76.
[Osokin, 2018] Osokin, D.: 'Real-time 2D Multi-Person Pose Estimation on CPU: Lightweight OpenPose'; ArXiv Preprint ArXiv:1811.12004 (2018).
[Ota et al., 2020] Ota, M., Tateuchi, H., Hashiguchi, T., Kato, T., Ogino, Y., Yamagata, M., Ichihashi, N.: 'Verification of reliability and validity of motion analysis systems during bilateral squat using human pose tracking algorithm'; Gait & Posture, 80 (2020), 62–67. https://doi.org/10.1016/j.gaitpost.2020.05.027
[Randall et al., 2001] Randall, M., Carlin, J. B., Chondros, P., Reddihough, D.: 'Reliability of the Melbourne assessment of unilateral upper limb function'; Developmental Medicine & Child Neurology, 43, 11 (2001), 761–767.
[Rodriguez, 1987] Rodriguez, G.: 'Kalman filtering, smoothing, and recursive robot arm forward and inverse dynamics'; IEEE Journal on Robotics and Automation, 3, 6 (1987), 624–639.
[Shotton et al., 2013] Shotton, J., Sharp, T., Kipman, A., Fitzgibbon, A., Finocchio, M., Blake, A., et al.: 'Real-time human pose recognition in parts from single depth images'; Communications of the ACM, 56, 1 (2013), 116–124.
[Souza et al., 2016] Souza, L., Lemos, T., Silva, D. C., de Oliveira, J. M., Guedes Correa, J. F., Tavares, P. L., et al.: 'Balance Impairments after Brachial Plexus Injury as Assessed through Clinical and Posturographic Evaluation'; Frontiers in Human Neuroscience, 9 (2016), 715–715. https://doi.org/10.3389/fnhum.2015.00715
[Tannous et al., 2016] Tannous, H., Istrate, D., Tho, M. H. B., Dao, T.: 'Feasibility study of a serious game based on Kinect system for functional rehabilitation of the lower limbs'; European Research in Telemedicine/La Recherche Européenne En Télémédecine, 5, 3 (2016), 97–104.
[Valevicius et al., 2019] Valevicius, A. M., Boser, Q. A., Lavoie, E. B., Chapman, C. S., Pilarski, P. M., Hebert, J. S., Vette, A. H.: 'Characterization of normative angular joint kinematics during two functional upper limb tasks'; Gait & Posture, 69 (2019), 176–186.
[van Wegen et al., 2010] van Wegen, E., Nijland, R., Verbunt, J., van Wijk, R., van Kordelaar, J., Kwakkel, G.: 'A comparison of two validated tests for upper limb function after stroke: The Wolf Motor Function Test and the Action Research Arm Test'; Journal of Rehabilitation Medicine, 42, 7 (2010), 694–696. https://doi.org/10.2340/16501977-0560
[Walmsley et al., 2018] Walmsley, C. P., Williams, S. A., Grisbrook, T., Elliott, C., Imms, C., Campbell, A.: 'Measurement of Upper Limb Range of Motion Using Wearable Sensors: A Systematic Review'; Sports Medicine - Open, 4, 1 (2018), 53. https://doi.org/10.1186/s40798-018-0167-7

[Wang et al., 2015] Wang, Q., Kurillo, G., Ofli, F., Bajcsy, R.: 'Evaluation of pose tracking accuracy in the first and second generations of microsoft kinect'; Presented at the 2015 international conference on healthcare informatics, IEEE (2015), 380–389.

[Wiedemann et al., 2015] Wiedemann, L., Planinc, R., Nemec, I., Kampel, M.: 'Performance evaluation of joint angles obtained by the Kinect v2'; (2015).

[Wu et al., 2005] Wu, G., van der Helm, F. C. T., (DirkJan) Veeger, H. E. J., Makhsous, M., Van Roy, P., Anglin, C., et al.: 'ISB recommendation on definitions of joint coordinate systems of various joints for the reporting of human joint motion—Part II: shoulder, elbow, wrist and hand'; Journal of Biomechanics, 38, 5 (2005), 981–992. https://doi.org/10.1016/j.jbiomech.2004.05.042

7 SUPPLEMENTARY MATERIALS

(a) Dedicated room for OptiTrack

(b) Subject in motion capture suit

Figure S1. Optitrack setup

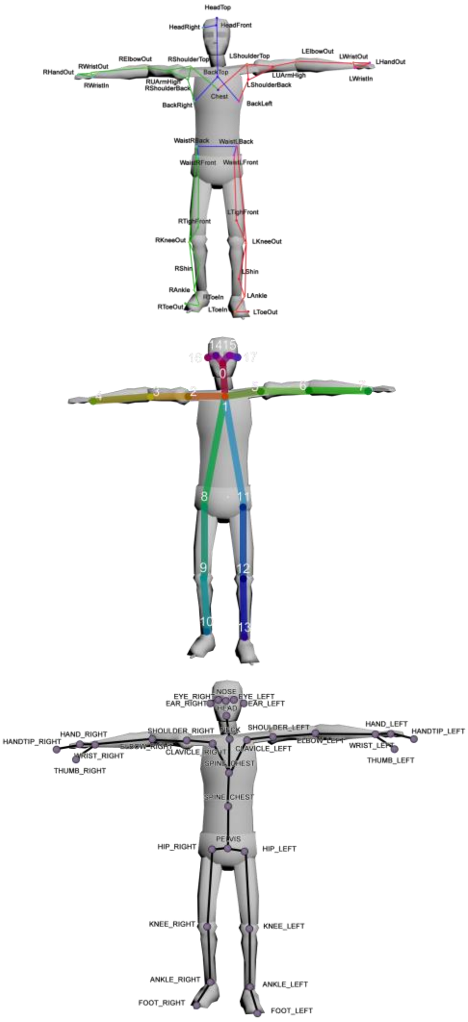


Figure S2. Human pose skeleton obtained by: Optitrack, OpenposeMod and RGB-3DHP and Azure

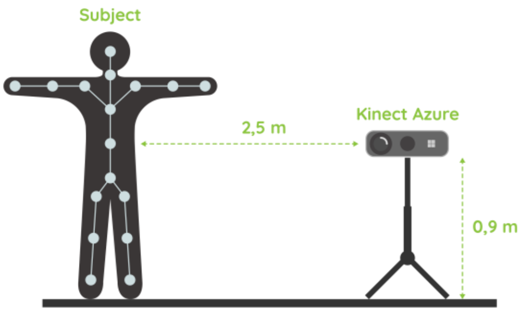


Figure S3. Acquisition setup with the RGB-D camera

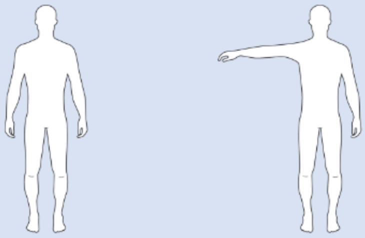
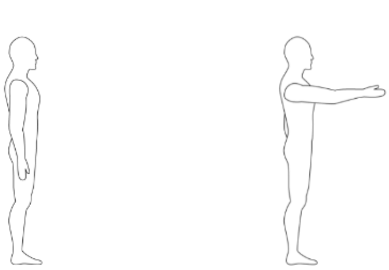
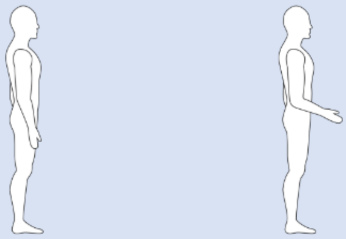
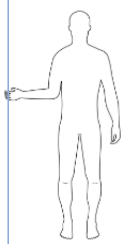
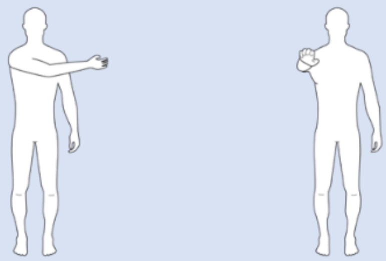
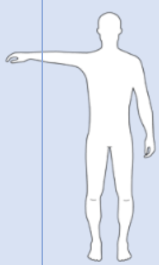
Exercise	Name of the measured angle	Initial	Final
1	shABfp		
2	shFLsg		
3	elFLsg		
4	shROTzp		
5	shFLzp		

Table S1. Graphical explanation of the exercise set for the upper limb

Exercise	Kinematic parameter	OptiTrack	Azure	OpenPoseMod	RGB-3DHP
1	min angle shABfp	24.58	21.22	22.82	25.45
	max angle shABfp	153.51	154.11	152.84	137.68
	angle range shABfp	128.93	132.89	130.02	112.23
2	min angle shFLsg	26.19	23.91	24.31	22.69
	max angle shFLsg	153.58	155.00	155.63	156.55
	angle range shFLsg	127.39	131.09	131.32	133.87
3	min angle elFLsg	47.45	31.32	32.08	58.75
	max angle elFLsg	165.77	157.02	164.63	135.30
	angle range elFLsg	118.32	125.70	132.56	76.55
4	min angle shRotzp	54.08	60.68	58.36	59.23
	max angle shRotzp	153.12	143.44	146.99	120.72
	angle range shRotzp	99.04	82.76	88.63	61.48
5	min angle shFLzp	59.23	80.09	62.60	85.51
	max angle shFLzp	157.65	171.15	172.21	152.22
	angle range shFLzp	98.41	91.05	109.61	66.70

Table S2. Kinematic parameters calculated for upper limb joints (in degrees)