RGB-D Sensors as Marker-Less MOCAP Systems: A Comparison Between Microsoft Kinect V2 and the New Microsoft Kinect Azure

Benedetta Rosa, Filippo Colombo Zefinetti, Andrea Vitali(✉), and Daniele Regazzoni

University of Bergamo, Bergamo, Italy
b.rosa1@studenti.unibg.it, {filippo.colombozefinetti, andrea.vitali1,daniele.regazzoni}@unibg.it

Abstract. Marker-less motion capture (MOCAP) systems based on consumer technology simplify the analysis of movements in several research fields such as industry, healthcare and sports. Even if the marker-less MOCAP systems have performances with precision and accuracy lower than the marker-based MOCAP solutions, their low cost and ease of use make them the most suitable tools for full-body movements analysis. The most interesting category is relative to the use of RGB-D devices. This research work aims to compare the performances of the last two generations of Kinect devices as marker-less MOCAP systems: Microsoft Kinect v2 and Azure devices. To conduct the tests, a list of specific movements is acquired and evaluated. This work measures the improvements of the Azure in tracking human body movements. The gathered results are presented and discussed by evaluating performances and limitations of both marker-less MOCAP systems. Conclusions and future developments are shown and discussed.

Keywords: Kinect V2 · Kinect Azure · Marker-less MOCAP systems · Accuracy

1 Introduction

A motion capture (MOCAP) system is able to detect one (or multiple) human body shapes used to reconstruct the corresponding virtual avatar that mimics the real movements. The animated virtual avatar can be recreated by means of different innovative technologies [1]. The market offers a wide range of technologies with different technical specifications, purposes, costs, limits and advantages. It is important to understand which device better fits with any specific purposes and budget. Among the main MOCAP solutions, the marker-less MOCAP systems based on RGB-D sensors are playing an important role to evaluate human motions in several research fields, such as industry, healthcare and sports.

Even if this family of MOCAP systems does not guarantee as accurate and precise acquisitions as the most powerful marker-based MOCAP systems, the low-cost and the ease of use of RGB-D sensors is preferred in all those contexts in which an error of measurement of several millimeters can be considered negligible. These are the cases of the motion evaluation of actions in which the analysis is usually based on wide

J. L. Wright et al. (Eds.): AHFE 2021, LNNS 264, pp. 359–367, 2021.
https://doi.org/10.1007/978-3-030-79763-8_43

movements of limbs or part of them. The most important RGB-D devices are relative to the family of Microsoft Kinect. In the last decade, Microsoft has released 2 devices (i.e., Kinect v1 and Kinect v2) for gaming, which have been exploited by scientific researchers to develop marker-less MOCAP solutions in several research fields.

In 2020, Microsoft released the Kinect Azure that has been totally designed as RGB-D device for research and development. The new Kinect Azure has to be investigated about its accuracy and precision. The aim of this paper is the comparison of the Kinect Azure with the well-known Microsoft Kinect v2 to understand which are the real potentialities as RGB-D device for a marker-less MOCAP system. The comparison is performed by the definition of a specific list of movements. The movements known a-priori allow a person to reach predefined positions useful to evaluate the accuracy of the MOCAP systems exploited. This approach is useful to avoid the introduction of a more precise and costly marker based Mocap system for validating the measurements relative to the tracked movements.

First, the paper presents the scientific background relative to the Kinect Azure device and the Kinect v2 device with particular emphasis on their use as low-cost marker-less MOCAP systems. Then, the reached results are compared and discussed to evaluate performances and limitations. Finally, conclusions are discussed to understand which features of the Kinect Azure will really increase the accuracy of a future marker-less MOCAP systems.

2 Scientific Background

MOCAP devices have been developed since the 1970s, when they were first created for military use, and have been developed for the entertainment industry since the mid-1980s [2]. Concerning the most used optical systems, the academic literature defines the accuracy of a MOCAP system as the comparison between the positions tracked from the investigated MOCAP system with the same motions simultaneously tracked by a gold standard solution, which is usually a marker-based mocap system, such as Vicon, Optitrack and Qualisys [3–7]. A maker-based MOCAP system is composed by retro-reflective markers and is considered the solution with the higher accuracy, even though costly, available in the market. For example, the accuracy of Vicon system is between 0,15 mm and 2 mm [8]. Several research works have been done using Vicon as reference for evaluating marker-less system based on one or multiple RGB-D devices. Scano et al. [9] evaluated the accuracy and reliability of the Microsoft sensor by means of a gold standard marker-based MOCAP system. Each subject performs two upper-limbs movements in three different orientation. Results confirm that RGB-D sensors can track upper-limbs movement for rehabilitation, but only in specific devices orientations. Cai et al. [10] quantify the accuracy and the reliability of Kinect V2 by means recording and assessing certain upper limbs movements. In particular, the Microsoft sensors are compared to Vicon MOCAP system as gold standard. The four performed movements lead to confirm that the Kinect V2 has potential in upper limbs assessment in patients' rehabilitation analysis.

The accuracy has been also evaluated by comparing movements and positions measured in the real world and compered with the virtual joints of the avatar generated by a

marker-less MOCAP system. Vitali et al. [11] have already assessed the accuracy performance of Kinect V2 under a double depth sensor configuration and they compared the empirical values with ground-truth values known a priori. This work evaluates and compares the performances of the two systems used to track the position of human articulations of both upper and lower limbs during the execution of predefined movements. They found that the 8 GoPro system is more accurate than the Kinect V2 double configuration for joints of both the upper and lower body.

The Microsoft Kinect Azure device was released in March 2020 and, differently from its predecessors, this version of the Kinect is less focused on gaming and more oriented towards logistics, robotics, healthcare and retail [12]. The main study conducted until now about Kinect Azure performances led by Albert and his research group [13]. The aim of the study is to assess motion tracking performance of a single Kinect Azure device compared to a single Kinect V2 device as MOCAP systems by evaluating a treadmill walking. The tracked movements have been compared with the gold standard Vicon multi-camera motion capturing. The main results show a higher accuracy of Kinect Azure with respect to its predecessor Kinect V2 regarding the spatial parameter, while no significant improvements were detected for the temporal parameters. A better tracking accuracy of Kinect Azure was found for the foot marker trajectories while the Kinect V2 gave more positive results for the mid and upper body region.

In this paper, the well-known Kinect V2 sensor is compared with the new Kinect Azure. Both RGB-D sensor performances are assessed by means a series of specific movements. In particular, the exercises involve not only upper limbs but also legs in vertical and horizontal directions. The acquisition campaign permits to evaluate the accuracy of the innovated Microsoft sensors compared to the previous version.

3 Data Acquisition Planning

The methodology of the research has been chosen by taking as reference approach the research work by Vitali et al. [11]. The best plan for the acquisition has been evaluated in order to fit the purpose of the research and to limit systematic error for both systems. In order to maintain continuity with the previous research and given the similarity of the devices under consideration the same movements and layout configurations are considered.

3.1 Scene Set-Up

The system layouts have been designed to optimize the acquisition of lower and upper limbs by considering the presence of a staircase and the subject position during the established movements (Fig. 1). Both MOCAP markerless systems are composed by two RGB-D devices. All the sensors are mounted on tripods at 120 cm of height. Both the Kinect V2 and Azure devices have been placed to guarantee a frontal view of the person as well as an opposite lateral perspective. For each MOCAP system, RGB-D has been placed in front of the first step of the staircase, while the other sensors are in the direction of the upper left corner of the staircase. The frontal Kinects are both 200 cm far from the centre of the staircase while their counterparts are 230 cm $\times$ 170 cm distant.

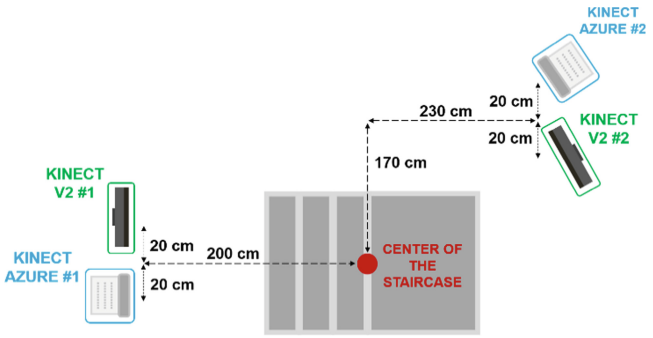


Fig. 1. Lay-out configuration of the two marker-less MOCAP systems.

3.2 Identification of Movements

Three specific movements have been identified to estimate the accuracy of the marker-less devices: climb and descent the staircase (Fig. 2a), free vertical movement of arms (Fig. 2b) and vertical movement of a hand following a predefined trajectory (Fig. 2c) [11]. In order to follow a common guideline and to make a consistent comparison, the subject is asked to reach the specified positions whose ground truth value is known a priori.

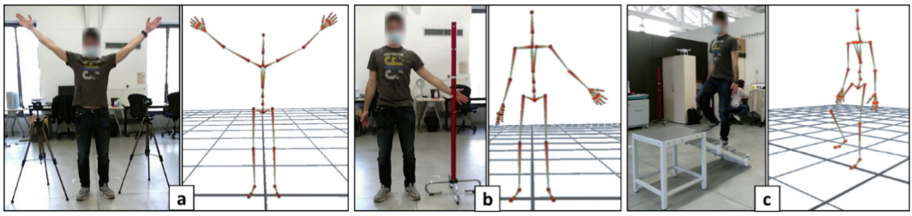


Fig. 2. Performed movement and the relative virtual avatar: climbing the staircase (a), putting hands up (b) and Vertical movement of hand by following a predefined trajectory (c).

Climb and Descent a Staircase. A rehabilitation staircase composed by 3 steps is used. The steps are 16 cm height except for the first one that is 17 cm height. Users have to climb and descent all the 3 steps. The subject climbs the first step with the right foot then he/she starts the descent with the left foot on the third step. The virtual joint evaluated are the Left and the Right Toe that are relative to the metatarsal of subject's feet. The measured parameter is the height of Left and Right Toe when the foot touches each step of the staircases.

Free Vertical Movement of Arms. A strand is horizontally placed at a height of 177 cm. Users abducts their arms until they touch the strand (i.e., the maximum height reached). The strand has to be touched for 5 s using the wrists before lowering the arms and placing wrists on the upper parts of two tripods placed under the strand. Both tripods

have a height of 105 cm (i.e., the minimum height reached). The virtual joints under consideration are the Left Hand and Right Hand, while the measured parameters are the maximum and minimum height of the wrist human joints.

Vertical Movement of Hand by Following a Predefined Trajectory. A vertical stick is used as aid to the users who move the left hand along the pipe, top down. Users place their wrists on a marker on the fixed stick at a height of 165 cm (i.e., the maximum height reached) and then they lower the hand keeping the contact with the stick since it reaches the second marker at the height of 100 cm (i.e., the minimum height reached). The virtual joints under consideration are the Left Hand and Right Hand and the measured parameter is the height of the virtual joint considered.

4 Acquisition Campaign

The acquisition campaign involved 11 subjects, 8 males and 3 females. A subject has been randomly chosen to perform the vertical movement of hand by following a predefined vertical trajectory 14 times. This task of acquisition permits to assess the influence of the measurement of different testers to the accuracy of the MOCAP systems. The translation from RGB-D sensors output data to virtual human avatar and the kinematic data are provided by iPiSoft MOCAP suite. The accuracy is evaluated by the Mean Absolute Error (MAE) and its standard deviation (SD), that is the average value of the differences between the virtual joint measurement and the real values.

4.1 Climb and Descent a Staircase

Table 1 is relative to the MAE and its SD for each step of the staircase. Microsoft Kinect V2 has an accuracy for the feet tracking on the ground (i.e., 0.96 cm) and a higher MAE for the other steps. The MAE reaches a value of 1.50 cm for the first step and the second step with the latter also having the higher SD of 1.31 cm. Microsoft Kinect Azure has a different behaviour: it shows the higher MAE for the ground position (i.e., 1.46 cm) and reaches its lower value in correspondence of the third step (i.e., 0.32 cm). With the exception of the ground value, the Kinect Azure has significantly lower MAE for each of the staircase step.

Table 1. MAE and the standard deviation for each step of the staircase using Kinect Azure and Kinect V2.

Step	MAE Kinect Azure [cm]	Std. Dev. Kinect Azure [cm]	Step	MAE Kinect V2 [cm]	Std. Dev. Kinect V2 [cm]
Ground	1.46	0.84	Ground	0.96	0.75
1st Step	0.74	0.83	1st Step	1.45	0.90
2nd Step	0.47	0.56	2nd Step	1.48	1.31
3rd Step	0.32	0.53	3rd Step	1.37	0.93

4.2 Free Vertical Movement of Arms

Table 2 shows the values for MAE and SD for the maximum and the minimum height, whose ground truth values are 177 and 105 cm respectively. The MAE of V2 highlights the lower accuracy in the higher position of the arms (i.e., 2.12 cm). It is meanly caused by the proximity to the end of the vertical field of view of the device [11]. This feature has been improved with the Kinect Azure that has vertical FOV (V-FOV) of 65°. The MAE in the maximum height for Kinect Azure is lower than of the V2 one (i.e., 0.98 cm). The minimum height is strongly improved with the Kinect Azure with a MAE value of 0.77 cm that is lower than Kinect v2 (i.e., 1.41 cm). The maximum height with the Kinect V2 also has a high standard deviation (i.e., 1.68 cm).

Table 2. MAE and SD for the free vertical movement of arms with both MOCAP system.

Microsoft Kinect Azure				**Microsoft Kinect V2**			
Max. Height [cm]		Min. Height [cm]		Max. Height [cm]		Min. Height [cm]	
MAE	Std. Dev.	MAE	Std. Dev.	MAE	Std. Dev.	MAE	Std. Dev.
0.98	0.95	0.77	0.58	2.12	1.68	1.41	0.87

The left wrist values are better than the right wrist ones even though the configuration of the Kinects leave the left side of the body at the extremity of the FoV. Kinect Azure and Kinect V2 have a quite stable MAEs for both arms. Also in this case, the Kinect V2 shows limitations during the movement tracking near the limit of the V-FOV. Its MAE in the maximum position reaches 2.00 cm for the right hand and 2.25 cm for the left hand. The MAE of the upper position for Kinect Azure has a value of 1.29 cm, with a SD of 1.09 cm.

4.3 Vertical Movement of Hand by Following a Predefined Trajectory

Table 3. MAE and SD for the movement of the left arm following a predefined trajectory using Microsoft Kinect Azure and Microsoft Kinect V2.

Microsoft Kinect Azure				**Microsoft Kinect V2**			
Max. Height [cm]		Min. Height [cm]		Max. Height [cm]		Min. Height [cm]	
MAE	Std. Dev.	MAE	Std. Dev.	MAE	Std. Dev.	MAE	Std. Dev.
0.99	0.79	1.58	1.55	2.59	1.70	1.47	1.27

The left-hand joint has been measured with respect to a minimum and a maximum height (i.e., 100 cm and 165 cm respectively). Table 3 shows the MAE and the SD for the two devices in both positions. The Kinect V2 has the higher MAE for the upper position

of the hand. The Kinect Azure has a MAE of 0.99 cm for the maximum position and 1.58 cm for the minimum, with a standard deviation having a high value for the minimum height (1.55 cm) and 0.79 cm for the maximum. In this case, the accuracy of Kinect V2 at the minimum position is better (MAE value of 1.47 cm) with respect to the one of the Kinect Azure (MAE value of 1.58 cm).

5 Repeatability Analysis

A repeatability study has been planned in order to evaluate the impact of the users' actions, that affect the measurements of motions. Hence, a single tester has been chosen to repeat the same movement for 14 times. Table 4 shows the MAE and the SD for both devices. The Kinect Azure has a MAE of 0.86 cm and a SD of 0.56 cm and the Kinect V2 has a MAE of 0.92 cm and a SD of 0.69 cm. These values demonstrate a very low difference of accuracy among the Kinect Azure and Kinect v2.

Table 4. MAE and SD for the repeatability analysis using Microsoft Kinect Azure.

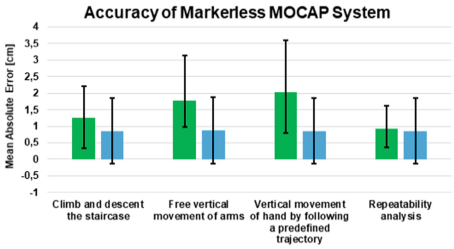
Microsoft Kinect Azure				**Microsoft Kinect V2**			
Max. Height [cm]		Min. Height [cm]		Max. Height [cm]		Min. Height [cm]	
MAE	Std. Dev.	MAE	Std. Dev.	MAE	Std. Dev.	MAE	Std. Dev.
0.80	0.61	0.92	0.54	0.42	0.40	1.43	0.53

The repeatability test gives positive result: the variations found between different testers are also due to the subject performing the exercise and not only to the sensors.

6 Discussion and Conclusions

The paper presents a research work about the evaluation of the accuracy of a marker-less Mocap system based on two Microsoft Kinect Azure device compared with a Mocap system based on two Microsoft Kinect V2. The main aim is evaluating if Kinect Azure really represents a step forward in the realization of competitive and reliable RGB-D sensors. The analysis was carried out through the execution of three different predefined movements as well as a repeatability analysis of a specific movement to evaluate intra-operator errors. The comparison concerned the evaluation of the virtual joints position compared to ground truth values. According to the results, the Kinect Azure presents generally better tracking capabilities than the older Kinect v2 device. The MAEs of Kinect Azure are generally lower with a clear worst case about the tracking of the lower human body (Fig. 3). The chosen layout is meant to create challenging conditions for the Mocap systems (e.g. acquisition close to the border of the field of view or occlusions) to better highlight performances and differences between the two. Occlusions penalized more Kinect V2 acquisitions among which the MAEs are over 1.7 cm, while the wider V-FoV of Kinect Azure allowed avoid tracking problems. As a matter of fact, during the

vertical movement of arms we reached a height of 177 cm with no issue for the Kinect Azure devices, that has a MAE lower than 1 cm. Kinect v2 presents a MAE of 1.77 cm that cannot be accepted for wide movement performed with arms. At present, the main lack of the Kinect azure is the tracking of the feet on the ground. This feature, on the other hand, is accurately calculated by the Kinect V2. As far as it concerns the limits of the study, we considered a sample of only 11 subjects. Despite of the aforementioned open issues, the presented results allow us to conclude that the Kinect Azure presents several relevant features, which will be exploited to further improve the quality of acquisitions.



	Microsoft Kinect V2	Microsoft Kinect AZURE
Climb and descent the staircase	1.25	0.86
Free vertical movement of arms	1.77	0.88
Vertical movement of hand by following a predefined trajectory	2.03	0.86
Repeatability analysis	0.92	0.86

Fig. 3. Accuracy measured for each movement tracked by means both MOCAP systems.

References

1. Zhang, Z.: Microsoft kinect sensor and its effect. IEEE Multimedia **19**(2), 4–10 (2012)
2. Furniss, M.: Motion Capture (1999). http://web.mit.edu/m-i-t/articles/index_furniss.html
3. Fernandez-Baena, A., Susın, A., Lligadas, X.: Biomechanical validation of upper- body and lower-body joint movements of kinect motion capture data for rehabilitation treatments. In: Proceedings of the 4th International Conference on Intelligent Networking and Collaborative Systems, INCoS 2012, pp. 656–661 (2012)
4. Chen, C., Jafari, R., Kehtarnavaz, N.: Improving human action recognition using fusion of depth camera and inertial sensors. IEEE Trans. Hum.-Mach. Syst. **45**(1), 51–61 (2015)
5. Kotsifaki, A., Whiteley, R., Hansen, C.: Dual Kinect V2 system can capture lower limb kinematics reasonably well in a clinical setting: concurrent validity of a dual camera markerless motion capture system in professional football players. BMJ Open Sport Exerc. Med. **4**(1) (2018)
6. Stone, E.E., Butler, M., McRuer, A., Gray, A., Marks, J., Skubic, M.: Evaluation of the microsoft kinect for screening ACL injury. In: Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS, vol. 2013, pp. 4152–4155 (2013)
7. Bernardina, G.R.D., Monnet, T., Pinto, H.T., De Barros, R.M.L., Cerveri, P., Silvatti, A.P.: Are action sport cameras accurate enough for 3D motion analysis? A comparison with a commercial motion capture system. J. Appl. Biomech. **35**(1), 80–86 (2019)
8. Vicon: Vicon Award Winning Motion Capture Systems. https://www.vicon.com/
9. Scano, A., Mira, R.M., Cerveri, P., Molinari Tosatti, L., Sacco, M.: Analysis of upper-limb and trunk kinematic variability: accuracy and reliability of an RGB-D sensor. Multimodal Technol. Interact. **4**(2), 14 (2020)

10. Cai, L., Ma, Y., Xiong, S., Zhang, Y.: Validity and reliability of upper limb functional assessment using the Microsoft Kinect V2 sensor. Appl. Bionics Biomech. (2019)
11. Vitali, A., Regazzoni, D., Rizzi, C., Lupi, G.: Low cost markerless motion capture systems: a comparison between RGB cameras and RGB-D sensors. In: Proceedings of the ASME IMECE 2020. American Society of Mechanical Engineers (ASME) (2020)
12. Microsoft: Azure Kinect DK. https://azure.microsoft.com/it-it/services/kinect-dk/
13. 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), 5104 (2020)