MedCrave Step into the World of Research
MedCrave Online Journal of Applied Bionics and Biomechanics

Research Article

Open Access

CrossMark

Testing the Microsoft kinect skeletal tracking accuracy under varying external factors

Abstract

Focusing on its possible use in motion analysis, the accuracy of the Microsoft Kinect was investigated under various external factors including relative position, external IR light, computational power and large nearby surfaces. Two different experiments were performed that either focused on a general situation in an open room or when seated at a table. Results indicated that a large number of factors significantly affect the measurement error, but with only minor effect sizes, where the relative position and orientation have shown to be most influential. Additionally, body movement and increased depth contrast (i.e. isolation from surrounding objects) are believed to increase the accuracy of the skeletal tracking process.

Keywords: kinect, skeletal, tracking, accuracy, motion analysis

Volume 6 Issue 1 - 2022

Carlos Diaz Novo,[1] Ronald Boss,[2] Peter Kyberd,[3] Ed Norman Biden,[4] Joyce Eduardo Taboada Diaz,[5] Maylin Hernández Ricardo[6]

[1]Universidad Tecnológica del Uruguay UTEC, Carrera de Ingeniería Biomédica y Mecatrónica. Uruguay
[2]Facultad de Mecánica Universidad Tecnológica de Delft. Holanda, Netherlands
[3]Escuela de energía y electrónica. Universidad de Portsmouth. Inglaterra, UK
[4]Facultad de Mecánica Universidad de New Brunswick, Canadá
[5]Universidad Tecnológica de La Habana "José Antonio Echeverría", CUJAE, Cuba
[6]Universidad de Ciencias Pedagógicas "Enrique J.Varona", Cuba

Correspondence: Joyce Eduardo Taboada Diaz, La Habana, Cuba, Tel 535252955; Email joyceetd77@gmail.com

Received: June 15, 2022 | **Published:** July 06, 2022

Introduction

Ever since its introduction in late 2010, the Microsoft Kinect has been an extremely popular electronic device. Aside from its application in gaming and entertainment, it has enjoyed a substantially increased interest in the academic world.[1–3] Its ability to perceive depth and track kinematic data has opened doors for low-cost and portable alternatives to otherwise expensive and elaborate methods.[4–7] Marker less motion analysis is one of such applications, wherein the Kinect has already been implemented for its potential characteristics.[3,8]

The Kinect is equipped with an infrared light (IR) emitter and sensor, a colour camera and an array of microphones. The IR emitter sends out a pseudo-random point cloud which, in turn, is read by the IR sensor. They provide two different perspectives on the scene much like human vision and triangulation is used to estimate the distance of each point from the Kinect, thus providing a stream of depth information.[7] This allows the Kinect to map out the scene and, more specifically, identify human shapes and its correspondent joints. As a result, a final skeleton can be used to recognize body postures and movements, ultimately used for motion analysis[5] and on physical therapies and rehabilitation monitoring process.[9–11]

The average noise levels of the obtained depth data are reported to be of similar magnitude with soft-tissue artifacts in motion capture systems such as Vicon,[12] indicating that the Kinect is a viable tool for motion analysis. However, occlusion of body parts,[7] poor IR-reflectivity in clothing and relative subject movements outside the field of view[4] are reported to decrease the Kinect's accuracy. Furthermore, it is seemingly unable to correctly assess joint angles[13,14] small range of hand movements[15] and maintain a constant anthropometry per subject.

Clearly, the skeletal tracking is a complex process, based on a series of assumptions. The precise details of these tracking algorithms, how they affect kinematic data and how various external factors affect their capabilities, are not revealed.[14] A recent study revealed the accuracy of Kinect V2 landmark movements was moderate to excellent and depended on movement dimension, landmark location and performed task.[16]

In this study, a pair of experiments was performed that attempts to investigate the accuracy of the skeletal tracking process of the Kinect device under various conditions. Based on its working principle, factors such as relative position, external IR light, computational power and reflectivity of nearby surfaces are believed to influence the performance. As a result, more insight is provided on the underlying processes that make up kinematic data and how to minimize measurement error. Accordingly, this can be used to find the ideal conditions for its application in motion analysis.

Materials and methods

In order to investigate the limits and possibilities of the Kinect device, two different experiments were performed. The first experiment included a more general approach, where a mannequin was tracked by a Kinect device under a large number of varying factors. In the second experiment, the same mannequin and several human subjects were tracked while seated at a table under static conditions. This was performed in order to test the influence of large nearby surfaces and to validate the use of the mannequin.

Kinect data capture

All experiments were performed while using the Kinect's ability to search for a seated skeleton. It is assumed that the presence or absence of lower limbs in the field of view did not affect the tracked skeleton. Before each trial, the Kinect is reset by blinding the cameras, forcing it to restart the skeletal tracking process. All measurements were performed at 30 Hz, using both a colour and depth resolution of 640 x 480 pixels. Data was captured using Kinect for Windows SDK v1.7 and an altered version of the Kinect Explorer as provided in the Kinect Developer Toolkit.

Data analysis

Because of the lack of information on limb orientation from Kinect data, joint coordinates are most representative for the Kinect's

accuracy. More specifically, shoulder joints are considered to represent the accuracy of proximal body parts, whereas wrist joints represent the distal body parts. For this reason, the absolute distance error between both shoulders and wrists were taken as the dependent variables.

As the Kinect is believed to make anthropometric assumptions from its built-in database, the real shoulder distance that is used for determining the measurement error is determined from anthropometric relations as described in.[17] For the mannequin, a thorax length of 31cm is measured, resulting in a shoulder distance of approximately 36cm. For human subjects, their body lengths were measured, such that their shoulder distance is equal to 18.34% of this parameter. All data analysis is performed using MATLAB 2009b (Math Works, Natick, MA).

Statistical analysis

Statistical tests were performed using SPSS Statistics version 19.0 (IBM Corp., Armonk, NY), where significance was determined at $\alpha = 0.05$ (two- sided) and a Bonferroni correction is used to adjust for multiple comparisons. Apart from the level of significance, the estimate of the partial effect sizes is also taken into account using Cohen's guidelines (i.e. 0.1-0.3 small effect, 0.3-0.5 medium effect and 0.5-1 large effect.[18] In the case of taking absolute errors, the distributions are positively skewed. For this reason, a logarithmic data transformation is performed for the statistic procedures in order to increase normality and homogeneity of variance.

Experiments

Because each experiment serves a different purpose, different procedures and techniques in data processing were required and are described in this section.

Experiment I: Multi-factor analysis

In this experiment, the upper half of a mannequin is animated by moving both arms up and down simultaneously. This movement is obtained by using a 12V DC motor outside of the Kinect's field of view that winds up a thin rope connected to the mannequin's wrists. A rigid link between both arms made ensure that the distance between the wrists remained constant and the movement symmetric along the sagittal plane. The mannequin was equipped with a black sweater and black gloves in order to retain a constant colour. Under these conditions, the following factors were varied:

- 4 different positions inside the Kinect's Field of View (FoV)
- 2 different positions in height
- 2 different orientations of the Kinect sensor
- 2 different movement speeds
- 2 computers with different capabilities
- presence of external infrared light

Figure 1 shows how the different positions and orientations relate to each other. The FoV, centered at 2.8m from the Kinect, was divided into four equal squares of each 1.40x1.40m. Height was varied between 0.20 and 0.65m and Kinect orientation was varied between a frontal position and a 45 angle, both at 1m from the floor.

The difference in movement speed was obtained by controlling the input voltage of the motor at two distinct levels, resulting in 0.2 and 0.4m/s. One computer was equipped with 8GB of RAM and a 3.40GHz quad-core processor, while a second was used with 4GB RAM and 2.10GHz dual-core processor. Throughout the rest of this paper, they will be referred to by their CPU. The external infrared light was introduced by turning on an eight camera Vicon motion capture system.

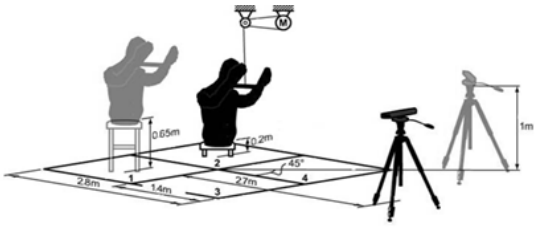

Figure 1 Experimental set-up of the multi-factor analysis. The mannequin is placed in four different positions (numbered 1 - 4) and two different heights (0.20m, 0.65m) with the Kinects in two different orientations (front, 45°). The wrists are fixed at a certain distance and moved upwards by an external motor, out of the field of view, at two different speeds. Dimensions are not in scale.

Because of the constant nature of the mannequi´s movement no repetitions were necessary, resulting in a total of 128 different trials. All trials were performed in a randomized order. A 6-way MANOVA on log-transformed data is performed to test for significance and to determine corresponding effect size.

Experiment II: Tabletop interference

A static mannequin is placed at a table with its arms resting on a tabletop. By varying the Kinect´s position in both height and orientation, the effect of the tabletop on the skeletal tracking accuracy is examined from different points of view. The different locations of the Kinect are kept in the same horizontal distance from the mannequin, being 1.30m. The height is varied between 0.7, 1.45 and 2.15m, where the lowest height is equal to the height of the tabletop. For each height, the vertical inclination angle of the Kinect is set to respectively 0°, -30° and 50°, such that the centre of the field of view resemble the same spot.

Orientation is varied between an angle of 0°, 45° and 90° with respect to the mannequin´s sagittal plane. Figure 2 shows the overall set-up of this experiment and how position and orientation relate to the seated mannequin. Before each subject is tracked by the Kinect device, the distance between wrist and elbow are measured. The subjects are specifically asked to keep these distances constant over the course of the experiment. All subjects were wearing clothing with long sleeves, such that a uniform surface reflectivity of the arms and torso is established. Because the static posture of the mannequin, this experiment allowed for tracking human subjects which are asked to sit as still as possible in the same posture, allowing for a validation of using the mannequin.

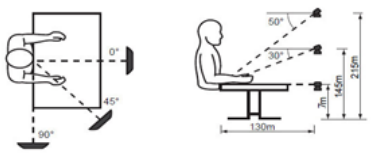

Figure 2 Experimental set-up for investigating table top interference. The Kinect position and orientation is varied between three different heights and three different angles with respect to the subject's sagittal plane. Dimensions are not in scale.

Each trial is repeated three times and a total of five subjects were measured. A 3-way MANOVA on log-transformed data is performed to test for significance and to determine corresponding effect sizes.

Results

Experiment I: Multi-factor analysis

The grand mean for the shoulder distance error is 0.018m and 0.013m for the wrist distance error. Multivariate test show that all main factors, except for computer CPU, and the majority of interactions are significant between the dependent variables, were Kinect position shows the largest effect size (0.30). Figure 3 & 4 show a collection of the mean absolute error and their standard deviations for the shoulders and wrist, respectively.

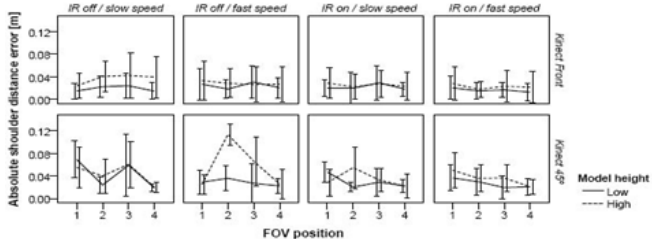

Figure 3 Means of absolute shoulder wrist distance error as measured by the Kinect. Error bars represent ± 1 standard deviation.

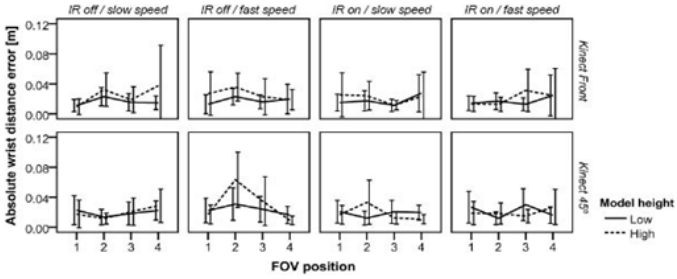

Figure 4 Means of absolute wrist distance error as measured by Kinect. Error bars represent +/- 1 standard deviation.

The back-transformed means of the main factors with corresponding effect sizes and p-value are shown in Table 1. Several interactions of factors are also found significant (not shown), but they rarely show an effect size grater that 0.1 and never greater than 0.15.

A Bonferroni post-hoc analysis within the FoV-group shows that no significant difference is observed between positions 1 & 2 ($p = 0.204$), 1 & 3 ($p = 0.418$) and 2 & 3 ($p = 1.000$) for the shoulder error. Position 4 shows the lowest error over all other positions ($p < 0.001$) and no position with significant highest error is found. For the wrist error, all positions differ significantly, where position 1 shows the lowest error ($p < 0.001$) and 2 shows the highest ($p < 0.001$).

Experiment II: Tabletop interference

For all performed trials, no skeleton was tracked when the Kinect was in the high position. The same situation occurred for the mid position and 90 orientations, except for 3 subjects. Figure 5 shows line charts for the shoulder and wrist distance error means and their standard deviations, where all subject measurements are bulked together in comparison to the mannequin data.

A full factorial MANOVA test on the log-transformed data shows that all observed differences for both shoulder and wrist distance errors are of significant order ($p < 0.001$). The grand mean for the shoulder distance error is 0.027m and 0.098m for the wrist distance error. In general, the mannequin showed higher levels of error compared to the human subjects with an effect size reaching 0.3. Further effect sizes indicate that the shoulder distance error is most influenced by orientation angle (0.387), whereas wrist distance error is most influenced by height (0.392) and its interaction with orientation angle (0.457).

Discussion

Experiment I: Multi-factor analysis

In general, the error at the wrists is significantly lower than the error measured at the shoulders. This is most probably because the Kinect does not place the shoulder joints at the same locations as defined by the used anthropometric relations, providing for an offset in measurement error.

Table 1 Table showing back-transformed means (in meters) of the main factors following from a 6-way MANOVA test performed on log-transformed data. Effect sizes and p-values are shown to indicate importance and significance of found results, respectively.

		Absolute shoulder distance error			Absolute wrist distance error		
		Mean	Effect size	p- value	Mean	Effect size	p-value
FoV	1	0.019	0.134	<0.001	0.012	0.095	<0.001
	2	0.020			0.015		
	3	0.018			0.012		
	4	0.014			0.013		
Height	Low High	0.015 0.020	0.123	<0.001	0.013 0.013	0.017	0.023
Kinect position	Front	0.013	0.291	<0.001	0.012	0.047	<0.001
	45°	0.024			0.014		
Speed	Slow Fast	0.019 0.016	0.06	<0.001	0.012 0.014	0.057	<0.001
Computer CPU	2.10GHz	0.018	0.003	0.667	0.013	0.014	0.07
	3.40Ghz	0.017			0.013		
IR light	Off	0.020	0.111	<0.001	0.014	0.057	<0.001
	On	0.016			0.012		

Citation: Novo CD, Boss R, Kyberd P, et al. Testing the Microsoft kinect skeletal tracking accuracy under varying external factors. *MOJ App Bio Biomech.* 2022;6(1):7–11. DOI: 10.15406/mojabb.2022.06.00160

When looking at the raw data in Figures 3 and 4, most values of mean error are only a few centimeters and are therefore within the range of the Kinect's average noise level. Cases that exceed this and reach values around 10 cm with small standard deviations are more important, and seem to occur the most when the Kinect is in the 45º position. It immediately follows from the performed statistical analysis, however, that a lot of factors appear to have a significant effect on the error that is made.

Due to these many factors that have even the slightest effect-and the fact that large sample sizes are used-it is more important to look at effects size. Here, largest effect sizes are observed in the factors that change the relative position between the mannequin and Kinect. Ideal conditions would be to have the mannequin close to the Kinect (FoV = 3 or 4), slightly lower (height = low) and the Kinect itself right in front of the mannequin (Kinect position = Front). The presence of external IR light only has a small effect on the shoulder error and even seems to improve the accuracy.

Other main and interaction factors very rarely exceed the threshold of introducing a small effect (>0.1). This can also be seen from the mean values in Table 1, which only shows very small differences. This indicates that the measurement error of the Kinect will always remain with the range of only several centimeters when the subject is in the Kinect´s working field of view. This coincides with results as reported by Dutta (2012).

Traces of these minor effects seem to disappear when looking at the wrist distance error. This implies that the measurement error at distal body parts is more dependent on the Kinect´s intrinsic tracking capabilities, and less on external factors. The effect of computer CPU is also slightly related to this effect, as it appears to be the only main factor that increases in effect size and significance towards the wrist distance error-indicating that tracking distal body parts requires more computational effort. Moreover, the fact that the arms were moving also allowed the Kinect to see the wrists from multiple perspectives, reducing the effects of factors that change relative position. All these factors are believed to contribute to the highly significant differences between the two dependent variables.

Experiment II: Tabletop interference

From Figure 5, a trend can be seen that shows higher levels of error when orientation angle increases. This is due to the fact that the Kinect de vice assumes the subject is facing the camera, as would be the case when playing games with it being the primary purpose of the device. This also coincides with the results from Experiment I.

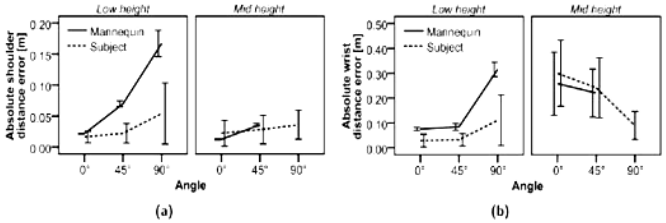

Figure 5 Means of (a) absolute shoulder and (b) wrist distance error as measured by the Kinect. No data was obtained for the mannequin at the 90° orientation at mid height due to the lack of a tracked skeleton. Error bars represent +/- 1 standard deviation. Note the differences in vertical axis scaling.

The fact that no skeletons were tracked at all by the Kinect at the high position, indicates that height also increases the error. This was also observed during the measurements of these trials, where these positions decrease the perceived depth contrast between the subject's arms and the table. As a result, the Kinect lumps these two together into a single object and is having trouble finding a human shape in it.

At mid height, the shoulder error seems lower and the wrist error shows a reversed trend. This can also be seen in the observed effect sizes which indicate that shoulder distance error is mostly affected by angle and not by height, while the opposite occurs with the wrist distance error. This strange behavior is probably a result due to a combination of the following effects: when increasing height, tabletop interference increases especially at the wrists, as they are closest to the surface; at the lowest position, the hand, wrist and elbow joint positions are in line from the Kinect's perspective, forcing it to infer these joint positions and bone lengths and introducing higher levels of error; and at increasing orientation angle, the Kinect would have to rely more and more on anthropometric as sumptions of a seated skeleton when determining the wrist and shoulder distance (and less on actual tracked data), perhaps this is more accurate due to the generalized posture of a seated subject. The latter, however, remains a supposition as to the exact reason why the wrist error would decrease with increasing orientation angle. The differences between the mannequin and human subjects and also within human subjects are all reported significant.

These differences are believed to originate from a large number of factors such as body posture, body shape, hair, clothing, watches, earrings and so on, which affect the comparison with Kinect's built-in trained dataset. The higher levels of error for the mannequin also suggests that a moving subject is assumed when searching for a skeleton, where even subtle movements such as a slight repositioning of one's self could make the difference.

The ideal position when measuring tabletop tasks would be to have the Kinect slightly higher from the table surface. This would decrease the size of the table as observed by the Kinect, decreasing its interference, and also increase the depth contrast between the arms and tabletop surface. Additionally, this position avoids the hand, wrist and elbow joint to be in the same line, preventing these joints to be inferred and therefore decreasing accuracy.

Overall

The most important observation from both experiments is that the Kinect's accuracy is dependent on numerous factors and sometimes even unpredictable. Even though they do contribute towards certain measures of error, a high error cannot always be deduced towards a set of external factors. For example, once a skeleton is tracked, the Kinect seems to show a certain bias in assuming that this initial guess is correct, sometimes resulting in skeletons being tracked at places where there is no person at all. In other words, one can only maximize accuracy by making the tracking process as easy as possible for the Kinect, i.e. being in positions and orientations as they occur in the database and by maximizing depth contrast between the subject and surrounding background/objects.

In general, the observed levels of error are mainly because the Kinect fails to keep a constant anthropometric mapping of a single subject. It is always trying to find the best possible solution for the placement of each joint that make up the skeleton, which might be a different decision every time due to both random and non-random processes. It is possible that post-processing of the data may ameliorate some of the inaccuracies. For example, fixing the segment sizes of the skeleton to known lengths, filtering to remove noise and jitter, etc. It is unknown how this would improve the capabilities, and if the precision

is sufficient to use the device in assessment of outcome measures.

Conclusion

Based on the performed experiments using the Kinect device, several conclusions and recommendations can be made concerning its use in motion analysis. As for relative position and orientation, the Kinect should be right in front of the subject and as close as possible while still being able to capture the whole body. Also, it should be slightly higher than the subject, with an exception for when the subject is seated at a table, in which case it should be slightly higher from the table surface. The movement of limbs allows the Kinect to see the moving joints from different perspectives, thus decreasing the effect of position and orientation. By increasing the depth contrast between the subject and the surroundings, i.e. clear isolation from objects, the Kinect is able to perform a better segmentation of the human shape which positively affects the eventual skeleton.

Within this skeleton, proximal body parts are most affected by external factors, whereas distal body parts are most affected by intrinsic tracking algorithms and computational power. An additional motion capture system using infrared light can be used simul-tenuously without introducing large errors in the skeletal tracking process.

When one is planning to use the Microsoft Kinect as a tool in motion analysis, these conclusions should be used as guidelines in order to be able to use its full potential, but also to provide for knowledge on its working principle and corresponding limits.

Acknowledgments

The authors would like to thank the Government of Canada's Canadian International Development Agency (CIDA) for funding through the Students for Development (SFD) and University Partnerships in Cooperation and Development (UPCD) programs.

Conflicts of interest

The authors declare that there are no conflicts of interest.

References

1. Ma Y, Mithraratne K, Wilson N, et al. Kinect V2-Based gait analysis for children with cerebral palsy: validity and reliability of spatial margin of stability and spatiotemporal variables. *Sensors*. 2021;21(6):2104.
2. t Özsoy U, Yıldırım Y, Karaşin S, et al. Reliability and agreement of Azure Kinect and Kinect v2 depth sensors in the shoulder joint ROM estimation. *Journal of Shoulder and Elbow Surgery*. 2022.
3. Guess T, Bliss R, Hall J, et al. Comparison of Azure Kinect overground gait spatiotemporal parameters to marker based optical motion capture. *Gait & Posture*. 2022;96,130–136.
4. Allahyari T, Sahraneshin A, Khalkhali H. Validity of the Microsoft Kinect for measurement of neck angle: comparison with electrogoniometry. *International Journal of Occupational Safety and Ergonomics*. 2017;23(4):524–532.
5. Wu E, Ma Ch, Shi X, et al. Imputing missing indoor air quality data with inverse mapping generative adversarial network. *Building and Environment*. 2022;215:108896.
6. Benedetti E, Ravanelli R, Moroni M, et al. Exploiting performance of different low-cost sensors for small amplitude oscillatory motion monitoring: preliminary comparisons in view of possible integration. *J Sensors.* 2016:1–10.
7. Ripic Z, Kuenze CH , Andersen M, et al. Ground reaction force and joint moment estimation during gait using an Azure Kinect-driven musculoskeletal modeling approach. *Gait & Posture*. 2022;95:49–55.
8. Stack E, King R, Janko B, et al. Could In-Home Sensors Surpass Human Observation of People with Parkinson's at High Risk of Falling? An Ethnographic Study. Hindawi Publishing. Corporation *BioMed Research International*. 2016.
9. Zhang MW, Ho RC. Harnessing the potential of the Kinect sensor for psychiatric rehabilitation for stroke survivors. *Technol Health Care*. 2016;24(4):599–602.
10. Guo J, Zhang Q, Chai H, et al. Obtaining lower-body Euler angle time series in an accurate way using depth camera relying on Optimized Kinect CNN. *Measurement*. 2022;188.
11. Antico M, Ballett N, Laudato G, et al. Postural control assessment via Microsoft Azure Kinect DK: An evaluation study. *Computer Methods and Programs in Biomedicine.* 2021;209.
12. Woldegiorgis B, Lin Ch, Sananta R. Using Kinect body joint detection system to predict energy expenditures during physical activities. *Applied Ergonomics*. 2021;97.
13. Hu G, Wang W, Chen B, et al. Concurrent validity of evaluating knee kinematics using Kinect system during rehabilitation exercise. *Medicine in Novel Technology and Devices*. 2021;11:100068.
14. Basha M, Aboelnour N, Aly S, et al. Impact of Kinect-based virtual reality training on physical fitness and quality of life in severely burned children: A monocentric randomized controlled trial. *Annals of Physical and Rehabilitation Medicine*. 2022;65(1).
15. Summa S, Tartarisco G, Favetta M, et al. Spatio-temporal parameters of ataxia gait dataset obtained with the Kinect. *Data in Brief*. 2020;32.
16. Pachoulakis I, Xilourgos N, Papadopoulos N, et al. A. Kinect-Based Physiotherapy and Assessment Platform for Parkinson's Disease Patients. *Journal of Medical Engineering* 2016;2:1437–1454.
17. Smith S, Bull A. Rapid calculation of bespoke body segment parameters using 3D infra-red scanning. *Medical Engineering & Physics*. 2018;6236–6245.
18. Cohen J. A power primer. *Psychol Bull*. 1992;112(1):155–159.