Influence of a marker-based motion capture system on the performance of Microsoft Kinect v2 skeleton algorithm

Naeemabadi, Mohammadreza; Dinesen, Birthe; Andersen, Ole Kæseler; Hansen, John

Published in:
IEEE Sensors Journal

DOI (link to publication from Publisher):
10.1109/JSEN.2018.2876624

Publication date:
2019

Document Version
Accepted author manuscript, peer reviewed version

Influence of a marker-based motion capture system on the performance of Microsoft Kinect v2 skeleton algorithm

MReza Naeemabadi, Birthe Dinesen, Ole Kæseler Andersen and John Hansen

***Abstract*— Microsoft Kinect sensors are being widely used as low-cost marker-less motion capture systems in various kinematic studies. Previous studies investigated the reliability and validity of Microsoft Kinect sensors by employing marker-based motion capture systems. Both systems employ IR emitters and detectors to track human posture and physical activities. This paper hypothesizes that the motion capture systems may interfere with Microsoft Kinect One sensor and influence the sensor's performance in tracking the skeleton. Hence, this study investigated the impact of a motion capture system on the Microsoft Kinect v2 skeleton algorithm using a mannequin in the presence of eight Qualisys Oqus 300/310 cameras and retroreflective markers. It was found that the motion capture system introduced a destructive impact on the Microsoft Kinect v2 skeleton tracking algorithm. In addition, it was observed that retroreflective markers placed near the joints caused the Microsoft Kinect v2 to give an incorrect reading of estimate the joint position. The motion capture cameras thus caused a time-varying distortion of the Microsoft Kinect estimate of the joint position. It is believed that the inference can be reduced by decreasing the number of markers and avoiding facing the motion capture cameras in sight of Microsoft Kinect v2.**

***Index Terms*— Microsoft Kinect sensor, Qualisys motion capture system, RGB-D sensor, mutual interference, experimental noise analysis.**

I. Introduction

MICROSOFT Kinect sensors are considered to be one of the most successful marker-less human physical activity trackers. The Microsoft Kinect sensors are low-cost and portable compared to marker-based motion capture systems. Consequently, they have attracted researchers' and developers' interests to use these sensors in various research areas, such as robotics and navigation [1], [2], animal studies [3], agriculture [4], physical activity and rehabilitation [5]–[7], fall detection [8], gaming and virtual reality [9]. In addition, Microsoft Kinect sensors have been proposed for use in a telerehabilitation program to track users' physical activities [10], [11]. Previous studies introduced a post-processing algorithm to generates a 3D model of the tracked skeletons based on the acquired depth images [12]–[14].

Several SDKs were developed for Microsoft Kinect sensors to estimate the position of each joint derived from the mentioned algorithms [15]–[17]. The Microsoft Kinect SDK human joint skeleton algorithm replicates 20 positions in each detected skeleton in Kinect v1 (Kinect Xbox 360) [18] and 25 positions in Kinect v2 (Kinect Xbox One) [15].

Although several studies evaluated the validity and reliability of the Microsoft Kinect v2 skeleton estimation using marker-based motion capture systems as a golden standard system [19]–[27], none of these studies investigated or reported any distortion in the estimated joint positions. The majority of these studies employed Vicon [19], [20], [33], [34], [21], [26]–[32], BTS SMART-DX [22], [35], or Qualisys [24], [36], [37] motion capture systems that all utilize reflective markers.

Microsoft Kinect v2 estimates the depth information based on time-of-flight (ToF) principle using infrared (IR) emitters and detector [38]. While the majority of marker-based motion capture systems estimate the position of retroreflective markers in the space using captured IR images from multiple cameras surrounding the region of interest [39]. Hence, different IR sources might cause cross-system interference.

Naeemabadi et al. [40] showed Microsoft Kinect v2 is using IR projectors emitting ray with 850nm wavelength. They remarked Qualisys motion capture system interferes the depth images captured by the Microsoft Kinect v2, while they are utilized simultaneously.

Therefore, this study is aimed to investigate the possible impact of the motion capture systems on the Microsoft Kinect 2 skeleton algorithm and providing possible solutions to reduce the potentially destructive impact on the Microsoft Kinect v2 recordings.

This work was supported in part by the Aage and Johanne Louis-Hansen Foundation and Aalborg University. For further information, see http://www.labwellfaretech.com/fp/kneeortho/?lang=en."

M. R. Naeemabadi and B. Dinesen are with Laboratory of Welfare Technologies - Telehealth and Telerehabilitation, Integrative Neuroscience Research Group, SMI®, Department of Health Science and Technology, Faculty of Medicine, Aalborg University, Denmark (e-mail: reza@hst.aau.dk; bid@hst.aau.dk).

O.K. Andersen is with Integrative Neuroscience Research Group, SMI®, Department of Health Science and Technology, Faculty of Medicine, Aalborg University, Denmark (e-mail: oka@hst.aau.dk).

J. Hansen is with Laboratory for Cardio-Technology, Medical Informatics Group, Department of Health Science and Technology, Faculty of Medicine, Aalborg University, Denmark (e-mail: joh@hst.aau.dk).

II. Materials and Methods

A. Procedure

In this study, a steady human posture was provided using a mannequin with bendable joints. The mannequin surface was made of polyurethane foam with the colored plush surface. It was dressed in black tight training clothes and placed at the center of the motion lab, surrounded by eight Qualisys Oqus 300/310 cameras. A Microsoft Kinect v2 sensor was mounted on a tripod (1.2m from the floor) and placed approximately 2.5m from the mannequin, which is within the recommended range for tracking human skeleton. The Kinect v2 sensor viewing area was adjusted to where the mannequin was placed at the center of the RGB camera and facing the camera (see Fig. 1).

A set of retroreflective markers were attached to the mannequin. The landmarks were chosen based on previous studies, focusing on tracking the chest, spine, head, hip, upper and lower extremities kinematics [41], [42]. An additional 20 markers were placed on the estimated joint positions by the Kinect sensor in order to track the position of the joint using the motion capture system.

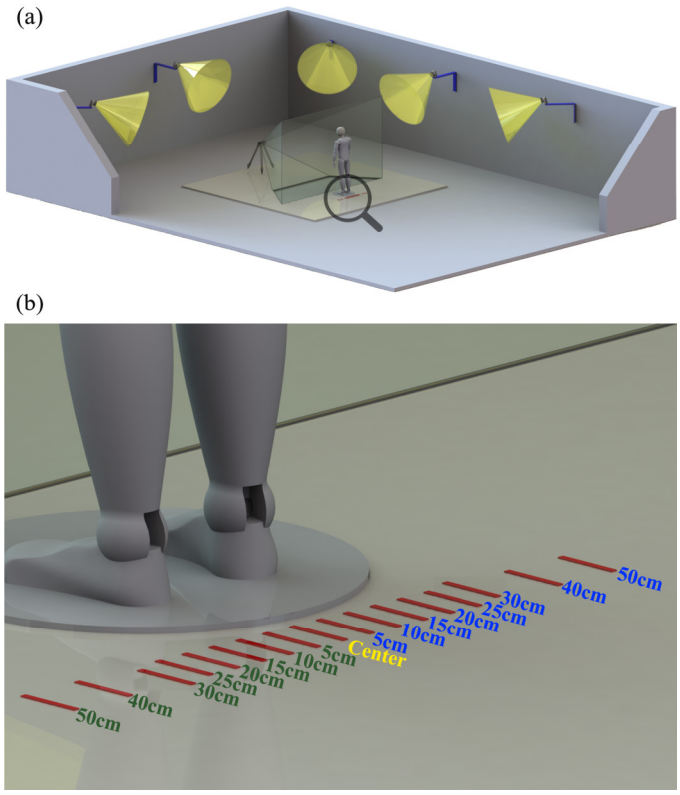


Fig. 1. Experiment's setup. (a) Laboratory dimensions and position of Oqus cameras, mannequin and Kinect sensor, are drawn in the figure. The Kinect sensor was mounted on a tripod 1.2m from the floor and placed approximately 2.5m from the mannequin. The central position is defined where the mannequin is standing in the center of the Kinect RGB camera. (b) The central position of the mannequin is shown with yellow color. Eight horizontal translations of the mannequin in each side (right and left) are shown with green and blue colors.

The mannequin posture remained fixed during all recordings, while 16 horizontal movements (in the coronal plane of the mannequin) were performed, moving away from the center of the room (seen by both the Kinect and the Qualisys motion capture system) in both right and left directions.

In each mannequin position, the recordings were repeated in the absence and presence of the Qualisys Oqus cameras (active noise source) and with and without the reflective markers (passive noise source) in order to evaluate the impact of the possible noise sources on the performance of the skeleton algorithm. Hence, the data were collected for 17 different mannequin positions, and in each position, four different data recordings were performed. Consequently, a total of 68 different combinations of mannequin position-configuration were recorded.

In this study, the possible noise sources were divided into passive and active noise sources in order to estimate the impact and behavior of each noise source individually. Retroreflective markers were assumed to be the source of passive noise, while the motion capture cameras were regarded as active noise sources.

B. Data Collection

The Qualisys motion capture system was utilized as a golden standard marker-based motion capture system. The recordings were captured using Qualisys Track Manager (QTM) 2.9 (build 1697) with a sampling frequency of 250 Hz and exposure time of 200 µs (Qualisys AB, Gothenburg, Sweden). The 25 joint positions (estimated by Microsoft Kinect SDK v2.0) were captured and stored using a Microsoft Kinect v2 sensor with a sampling frequency of 30 Hz [15]. A custom-built TCP/IP based software program was developed to trigger simultaneous data acquisition on both systems.

C. Data Processing

The absolute position of mannequin joints may vary slightly between each pair of recordings at the same position; therefore, an alternative estimation was utilized. Although the absolute position of each joint might change, the relative positions of each pair of joints were expected to remain intact. By assuming that the mannequin posture did not change during the recordings, it can be concluded that the length of each bone between the two joints and the angle of each joint remained constant. Accordingly, the bone length and joint angle were assumed to be independent of absolute joint position and constant.

Seventeen joints out of 25 joints from Microsoft SDK version 2.0 skeleton algorithm were utilized to assess the effect of the noise source on the estimated joint positions. The head, neck, spine, shoulders, elbows, hip, knees, ankles and wrist joints were used in the analysis, while the feet, hands and thumb joints were excluded. Fig. 2 shows the joints included and the corresponding virtual bones between each joint pair in this study.

The length of each bone ($B^i_{a,b}$) can be calculated by measuring the Euclidean distance between each two joints. Similarly, the angle between each pair of bones ($\theta_{(B^i_{a,b},B^i_{b,c})}$) can be estimated by calculating the angle between each of the two bones. In this study, neck, shoulder, hip, and knee joint angles

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2018.2876624, IEEE Sensors Journal

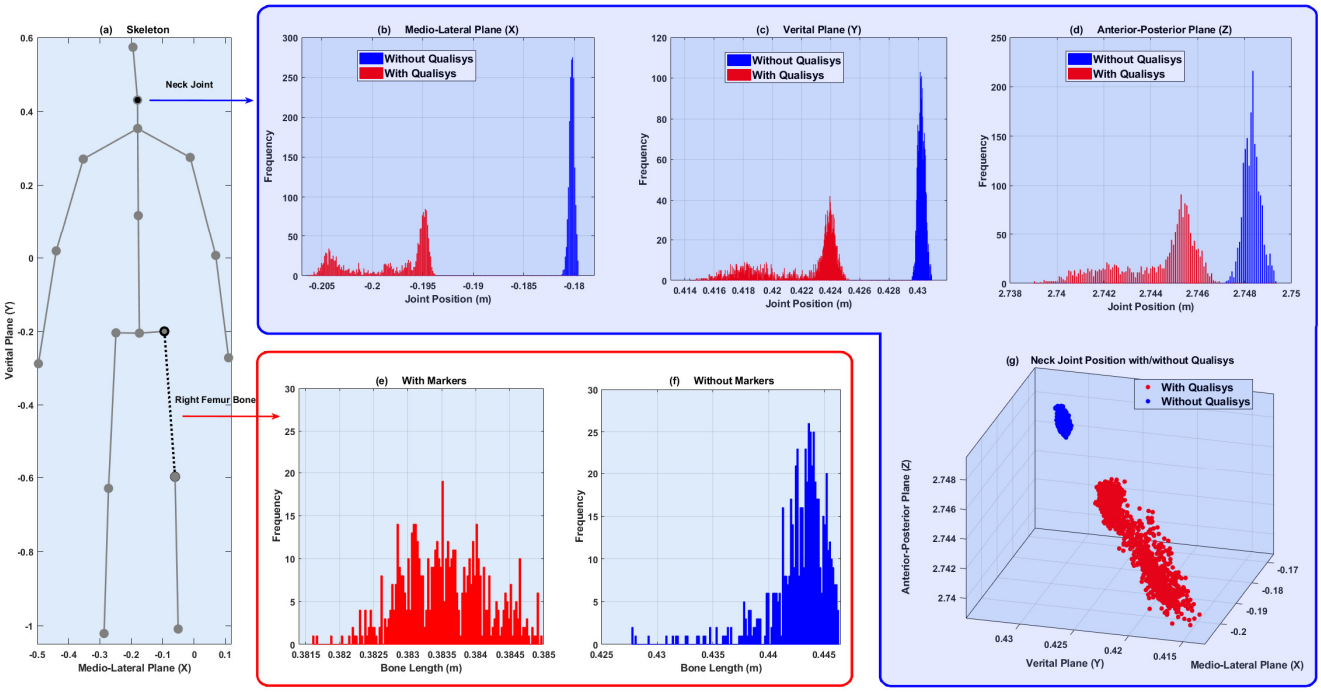


Fig. 3. Estimate skeleton captured by the Kinect v2 sensor. (b-d) Distribution of the neck joint in the recording with and without the contribution of the active noise, represented in red and blue colors, respectively. The planes reflect anatomical human body planes, and the axis reflects the Microsoft Kinect local coordinates. (e) Distribution of right femur bone length in the absence of noise sources. (f) Distribution of right femur bone length where reflective markers involved as a passive noise source. (g) Corresponding point cloud of neck joint position in the space with and without the contribution of the active noise.

were compared. A preliminary investigation showed that the positions of the joint and bone length and joint angles in the presence of noise source did not have a normal distribution (Shapiro-Wilk normality test $p< 0.05$). Therefore, the median position (M) of the joints and interquartile range (IQR) were utilized to measure the influence of noise (see Fig. 3a, 3b, 3c).

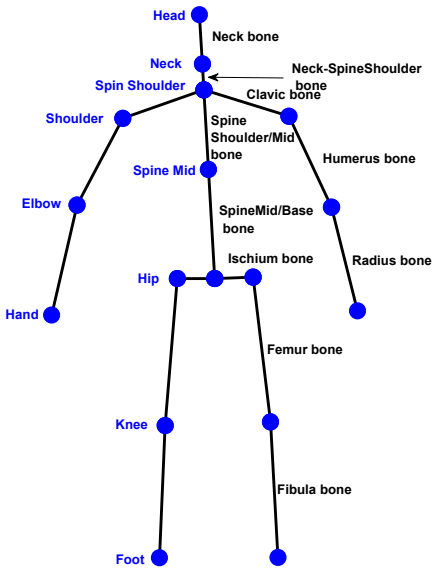


Fig. 2. The seventeen chosen joints and corresponding sixteen virtual bones. The joints' names are shown with blue circles, while the name of each bone is shown in black lines

Each pair of recordings was compared by calculating the difference between the median (M) bone length and the ratio of interquartile ranges (IQR) of bone length. The difference of bone lengths in each paired recording is shown as $Bias\left(B^i_{a,b}\right)$ (see equation 1), and the ratio of IQRs in the presence and absence of noise in shown as $Ratio\left(B^i_{a,b}\right)$ (see equation 2).

$$Bias\left(B^i_{a,b_i}\right) = \left\|M_{Noise}\left(B^i_{a,b}\right) - M\left(B^i_{a,b}\right)\right\| \quad (1)$$

$$Ratio\left(B^i_{a,b}\right) = \frac{\left\|IQR^N\left(B^i_{a,b}\right)\right\|}{\left\|IQR\left(B^i_{a,b}\right)\right\|} \quad (2)$$

Where, $M_{Noise}\left(B^i_{a,b}\right)$ and $M\left(B^i_{a,b}\right)$ are median length of the bone in the presence and absence of noise source, while $IQR_{noise}\left(B^i_{a,b}\right)$ and $IQR\left(B^i_{a,b}\right)$ represent the IQR of the bone length in each recording whether or not noise source is involved. Accordingly, the difference of medians $Bias\left(\theta_{\left(B^i_{a,b},B^i_{b,c}\right)}\right)$ and ratio IRQs $Ratio\left(\theta_{\left(B^i_{a,b},B^i_{b,c}\right)}\right)$ can be measured for the angle of the joints.

III. Results

A. Impact of Passive Noise Sources

The impact of the passive noise source in 17 positions and 16 bones was evaluated. A total of 272 bone-positions were compared in the presence and absence of the passive noise.

The results showed that the femur bone was more influenced by the passive noise sources than other bones. However, the unexpected variation of bone (ratio of IQRs) was low (except for the right radius at 50cm right translation and left radius at

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2018.2876624, IEEE Sensors Journal

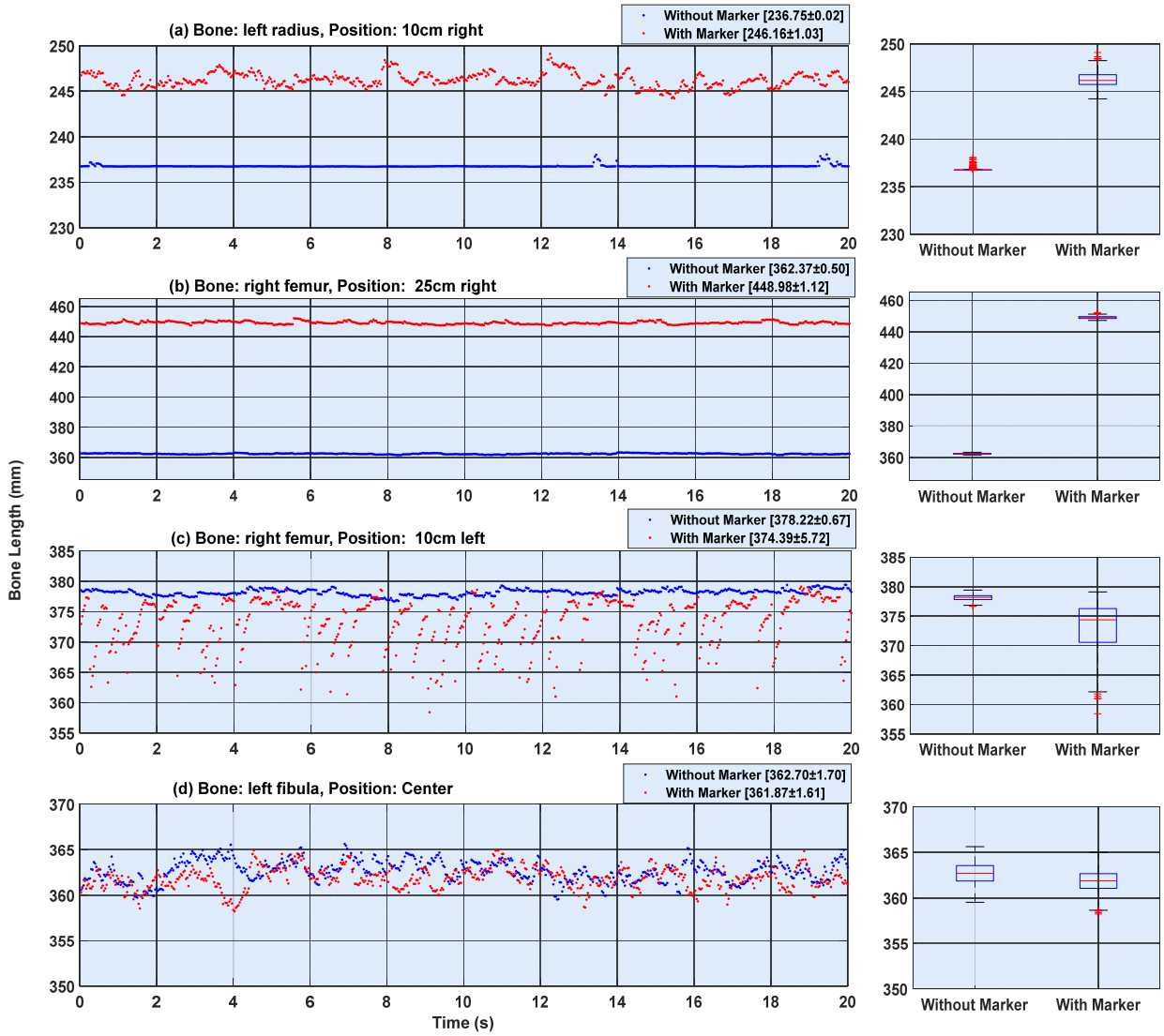


Fig. 4. Comparison of the impact of the passive noise sources on bone length in four different bone-positions. Blue dots show bone length where the noise sources were not involved in the recordings, and red dots represent the bone length in the presence of the passive noise sources. The values in the square brackets represent the median and IQR of each recording. (a) Left radius bone in 10cm right translation, where the difference in bone length is 9.4mm, and the ratio of IQR is 44.2. (b) Right femur in 25cm right translation, where the difference in bone length is 86.6 mm and the ratio of IQR is 2.2 (c) Right femur in 10 cm left translation, where the difference in bone length is 3.9 mm, and the ratio of IQR is 8.6 (d) Left fibula in central position, where the difference in bone length is 0.7 mm and the ratio of IQR is 0.9.

10cm right translation, where the ratios of IQRs for these bone-positions were 108 and 44, respectively). Fig. 4 compares bone length for four different bone-positions in the presence and absence of passive noises.

Fig. 4 represents four different possible effects of passive noise on bone length. In general, the passive noise may cause both considerable bone change and variation, as shown in Fig. 4a. In the majority of bone-positions and mostly in femur bones, only considerable steady change in bone length due to passive noise has been observed (see Fig. 4b). In a few bone-positions, the bone lengths did not significantly change, but very high variation in bone length were observed (see Fig. 4c). Some of the bone-positions were not influenced by the passive noise, such as the left fibula in the center position, shown in Fig. 4d.

Similarly, the impact of the passive noise source on seven joint angles in 17 positions were evaluated. The results indicate that the upper limb joint angles were more affected by the presence of the passive noise source than lower limb joint angles. The left shoulder joint in center position showed the greatest change in the estimated angle, and the right hip angle at 50cm right translation showed the highest variation compared to the absence of passive noise sources.

B. Impact of Active Noise Sources

Fig. 5 illustrates four different effects of the active noise on the bone length in different bone-positions.

As depicted in Fig. 5, neck, clavicle, and humerus bones were more affected by the active noise source (Fig. 5a). Moreover, in the central positions of the mannequin (5 cm and 10 cm translations), the higher impact of active noise sources were recorded. The results indicated that the highest distortion was

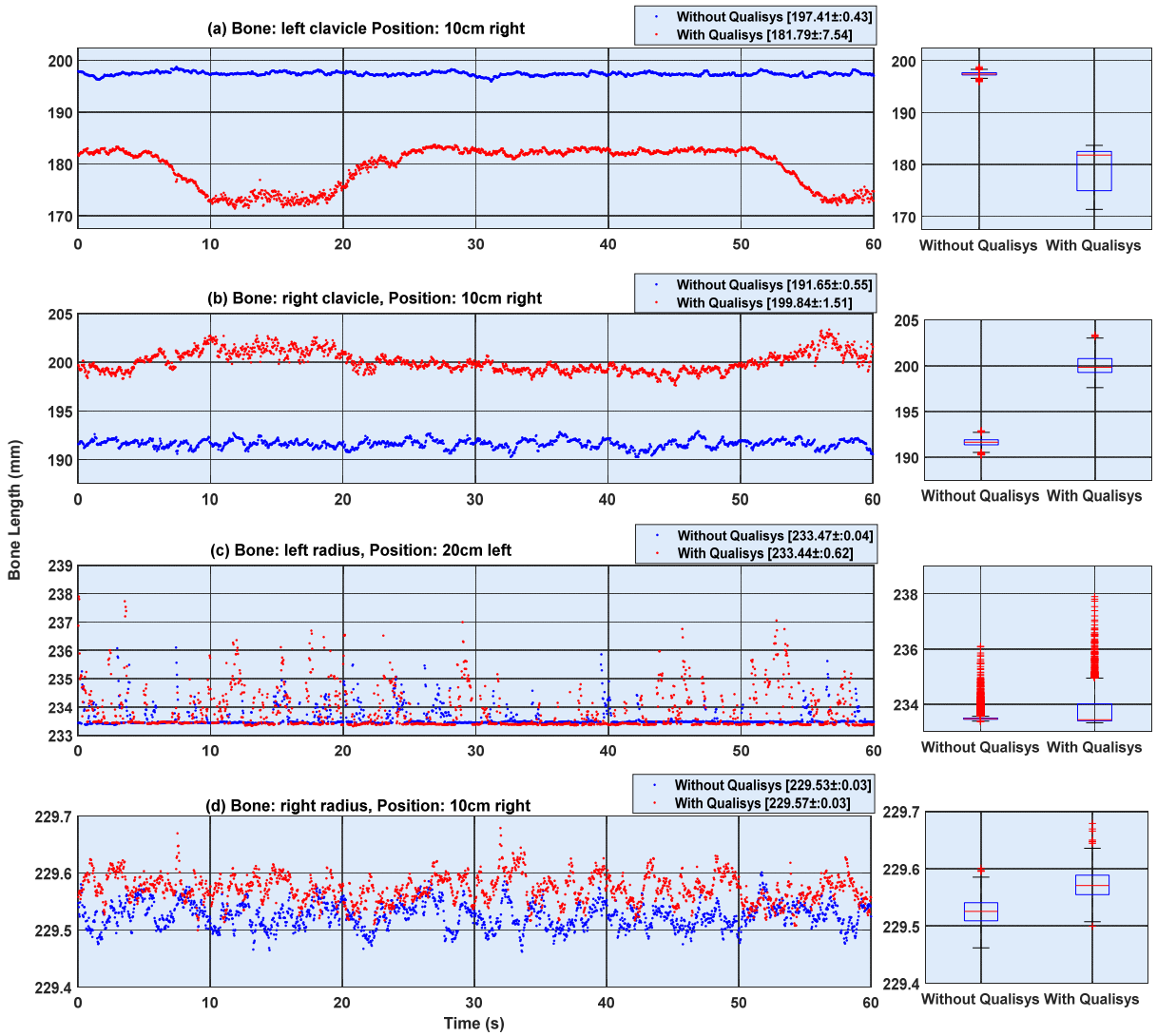


Fig. 5. Comparing the impact of active sources on bone length in four different bone-positions. Blue dots show bone length where the noise sources were not involved in the recordings, while red dots represent the bone length in the presence of active noise sources. The values in the square brackets represent the median and IQR of each recording. (a) Left clavicle bone in 10cm right translation, where the difference in bone length is 15.8 mm, and the ratio of IQR is 17.6. (b) Right clavicle in 10 cm right translation, where the difference in bone length is 8.3 mm, and the ratio of IQR is 2.8. (c) Left radius in 20 cm left translation, where the difference in bone length is 2.0 mm, and the ratio of IQR is 15.5. (d) Right radius in 10 cm right translation, where the difference in bone length is 0.0 mm, and the ratio of IQR is 1.1.

in the neck and left clavicle bones in 10 cm right translation.

The results showed that the upper limb joints were more affected by the active noise source rather than the lower limbs. The left shoulder at 20 cm left translation had the highest change and the neck joint in the 10 cm right translation had the highest variation in terms of measurement in the absence of active noise.

The results show that the neck, neck-spine shoulder, spine shoulder/mid and clavicle bone length had the same periodic behavior, with roughly 45-second intervals (see Fig. 5a). However, the intensity of variation might be lower, as can be seen in Fig. 5b. Very few recordings introduced a high variation in bone length and a low bone length change (see Fig. 5c). Based on the results, the remaining bone-positions were less affected by the active noise (see Fig. 5d).

C. Investigating Recorded IR and Depth Images

The captured IR images showed the retroreflective markers, and the Qualisys Oqus camera caused the bright regions around the object (see Fig. 6c). The live video stream of IR images indicated changes in the brightness levels of pixels around the camera, while the brightness of those pixels representing positions of retroreflective markers remained constant.

The corresponding depth images indicated that the Microsoft Kinect SDK v2.0 was not able to estimate the depth of those bright areas. As a result, several areas of unknown depth were observed due to the noise sources (see Fig. 6d). The area of unknown depth surrounding the Qualisys Oqus camera (next to the mannequin's neck) was altered in terms of a number of pixels and area size.

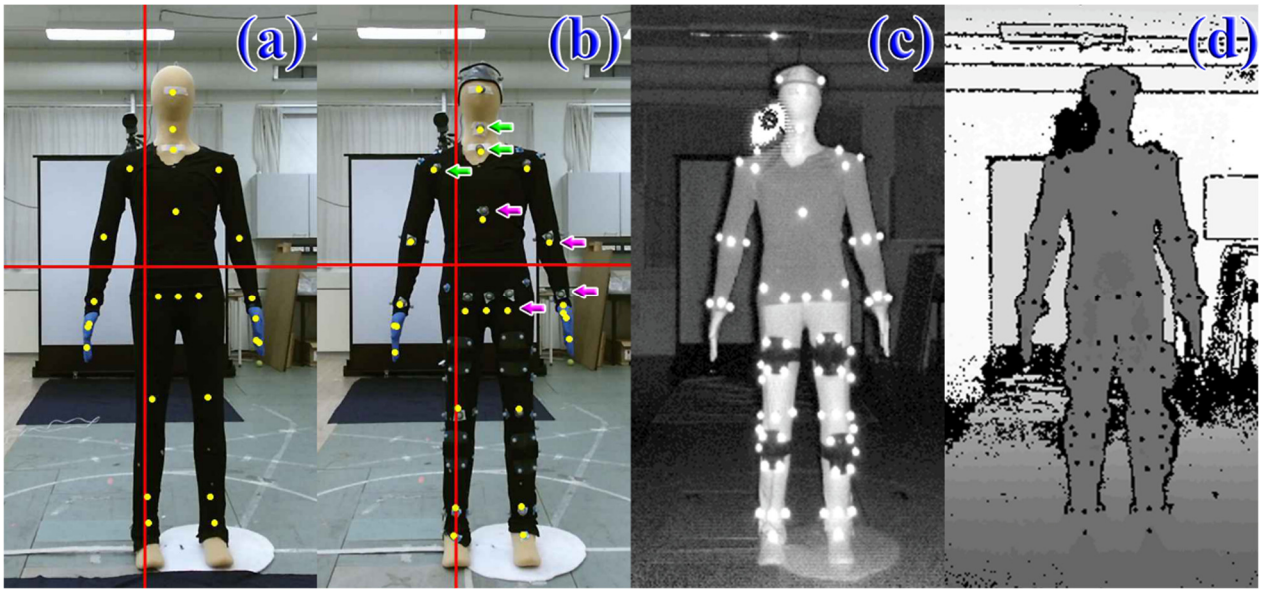


Fig. 6. A record of Microsoft Kinect v2 in absence and presence of noise sources while mannequin was placed in 10 cm translation to the left. (a) RGB record in the absence of active and passive noise sources, yellow circles represent the center of estimated joints position by Microsoft Kinect SDK v2.0 skeleton algorithm. (b) RGB record in the presence of the active and passive noises were interfering with the Microsoft Kinect sensor and the estimated joint positions. The green arrows show the joints are affected by the active noise and the purple arrows show the joint positions are affected by the passive noise. (c) the corresponding IR record in the presence of the active and passive noises. (d) corresponding estimated depth map in the presence of the active and passive noises. (The black pixels represent areas of unknown depth).

IV. Discussion

This study investigated the effect of active and passive noise sources introduced by a motion capture system on the Microsoft Kinect SDK v2.0 skeleton algorithm, and the impact of distortion were assessed.

The findings show that the active noise source (Qualisys Oqus cameras) may change the Euclidian distance between two estimated joints up to 15 mm, while the changes in bone length due to the passive noise source (retroreflective markers) may exceed 80 mm. As a result, we may assume that the passive noise sources cause a greater change between two joint distances (bone length) than do the active noise sources.

The results showed that spine shoulder-mid, spine mid-base, femurs and fibular bones were more influenced by the passive noise in almost all the positions. Having examined the results appears that the Microsoft skeleton algorithm is trying to compensate for the induced error and avoid propagation error to the other estimated joints. In addition, it seems that the Microsoft Kinect SDK v2.0 skeleton algorithm is trying to avoid estimating the joints' position precisely onto the area of unknown depth. Therefore, the joints were shifted from the unknown depth area into the nearest possible area (see Fig. 6b, where the joints are shown with purple arrows). Consequently, length of spine shoulder-mid and spine mid-base bones were distorted.

Based on the results, it seems that the passive noise source, in general, has a stationary impact on the estimated joint position. As a result, most of the estimated bone length did not vary during the recordings (both in presence and absence of the passive markers).

The raw data indicated that elbow and wrist joints in the presence of the passive noise have either an increase or decrease in length variation ratio. This is most likely a result of the fact that the joints were surrounded by multiple areas of unknown distance (see Fig. 6d); therefore, the ratio of variation in radius bone length is estimated as either too high or too low due, respectively, to fluctuation between boundaries of unknown depth areas or restricting between unknown depth areas. The same results were also obtained for knee joints. Therefore, it can be concluded that the reflective markers which surround joints (purposed in some landmarks, such as the Helen-Hayes marker set [43]) might lead the Kinect skeleton algorithm to provide an inaccurate estimate of corresponding joints.

By examining the impact of active noise, it can be concluded that the noise influence depends on the position of the mannequin. The result illustrates that in the near central translations (5 and 10 cm horizontal translations) head, neck, spine-mid, spine-shoulder, and shoulder joints had higher displacement and variation. As a result, length of neck, clavicle and humerus bones were highly influenced by the active noise source.

Having examined Fig. 6a, it can obviously be noticed that one of the active noise sources (which was behind the mannequin neck in the nearly central position) appeared on the right side of the mannequin neck leading to a slightly left-move in the detection of the mannequin. As a result, the joints' position (which were nearby the unknown depth areas) were shifted (see Fig. 6b, where the joints are marked by green arrows.

The impact of interference on the mediolateral plane (x) and vertical plane (y) were increased when one of the Qualisys Oqus cameras was in direct sight of the Kinect sensor. Evidently,

light interference appears to be the main reason for the distortion, and the interference may stem from identical light wavelengths. Hence, both Microsoft Kinect v2.0 sensor and Qualisys Oqus cameras use 850 nm wavelength. The Microsoft Kinect v2.0 uses intensity modulation with three different frequencies [44], [45], while the Qualisys motion capture system employs time-division multiplexing between several cameras. Therefore, the destructive noise seems to be a result of the timing between two systems, which may also explain why the distortion seemed to vary over time.

The results showed that the passive noise source caused lower bone length variation (RoIQR) in comparison to the active noise, while larger changes in bone length (Bias) were observed.

This study showed the presence of a marker-based motion capture system interfere with the Microsoft Kinect v2 skeleton algorithm. However, Giblin et al. [46] reported that they did not observe any interaction between Microsoft Kinect v2 and the Vicon motion capture system. However, it remains unknown how interference between the two systems was evaluated.

Vicon cameras can be equipped with 875 nm, 780 nm, or 623 nm LED strobes; therefore; it is difficult to evaluate the exact distortion on Kinect recordings in the individual reports. However, the passive noise source may have shown its influence on the recordings. Hotrabhavananda et al. [33] evaluated the Kinect v2 depth and skeleton data, while the subjects were asked to perform three clinical evaluation tasks. Gaddam et al. [47] used three Vicon MX cameras to evaluate the Kinect SDK v2.0 and with the Vicon cameras in direct sight of the Kinect. Moreover, Muller et al. [29] introduced a marker-less gait assessment system with six Microsoft Kinect v 2.0 sensors. They did not report any distortion introduced by the motion capture system.

Eltoukhy et al. and Oh et al. [22], [35] used a Smart DX-7000 BTS motion capture system that used an 850 nm LED strobe, and this might have had a negative impact on the Kinect accuracy. It might be argued that there was a low consistency between the two systems in the study. Woolford et al. [24] used the Qualisys ProReflex system and reflective markers to assess Microsoft Kinect v2 accuracy for monitoring physical activities. However, Qualisys ProReflex cameras emit 880 nm strobe, which might still contribute as an active noise source in the Kinect v2 recordings. Alessandro et al. [36], [37] employed 12 Qualisys Oqus cameras and 39 passive retro-reflective markers in two studies to investigate the performance of the Microsoft Kinect v2. It seems that their results may be influenced by investigated interference, as they did not propose a solution to avoid it.

The current study had two methodological limitations. First only the effect of noise sources on the Microsoft Kinect SDK v2.0 skeleton algorithm was evaluated; however, Sarbolandi et al. [44] showed that the Microsoft SDK had overall better performance. Second, the limited positions and static postures might have impacted interference, giving different in dynamic postures.

V. Conclusion

In this study, the influences of passive and active noise sources on the Microsoft Kinect SDK v2.0 skeleton algorithm was evaluated. The findings indicate that the estimated position of joints was sensitive to the projected IR lights from the Qualisys motion capture system and retroreflective markers. In addition, the results showed that the presence of the active noise sources in proximity to any of the estimated joints caused inaccuracy in finding the optimized position. Moreover, the markers misled the skeleton algorithm, causing it to inaccurately estimate the position of those joints surrounded by the retroreflective markers. However, the accuracy and precision of the Microsoft Kinect v2 was frequently evaluated by a marker-based motion capture system; none of the previous studies reported any interference between the two systems. It can be concluded that the accuracy and precision of the Microsoft Kinect v2 might be higher in the absence of the interferences. Therefore, it can be hypothesized that the Microsoft Kinect v2 skeleton tracking algorithm might have higher accuracy than the reported value in the previous studies.

The impact of interference might be reduced in the following ways.

- First, avoiding facing the active noise source insight of Microsoft Kinect.
- Second, reducing the number of the active sources (IR strobes)
- Third, decreasing the number of reflective markers to the possible minimum number,
- Finally avoid placing the markers around or on the Microsoft Kinect joints.

Acknowledgment

This study is supported by the Aage and Johanne Louis-Hansen Foundation and Aalborg University. For further information, see http://www.labwellfaretech.com/fp/kneeortho/?lang=en

References

[1] K. B. Cho and B. H. Lee, “Intelligent lead: A novel HRI sensor for guide robots,” *Sensors*, vol. 12, no. 6, pp. 8301–8318, 2012.

[2] O. M. Mozos, H. Mizutani, R. Kurazume, and T. Hasegawa, “Categorization of indoor places using the Kinect sensor,” *Sensors*, vol. 12, no. 5, pp. 6695–6711, 2012.

[3] B. Lee, M. Kiani, and M. Ghovanloo, “A Smart Wirelessly Powered Homecage for Long-Term High-Throughput Behavioral Experiments,” *IEEE Sens. J.*, vol. 15, no. 9, pp. 4905–4916, 2015.

[4] G. Azzari, M. L. Goulden, and R. B. Rusu, “Rapid characterization of vegetation structure with a microsoft kinect sensor,” *Sensors*, vol. 13, no. 2, pp. 2384–2398, 2013.

[5] K. J. Bower, J. Louie, Y. Landesrocha, P. Seedy, A. Gorelik, and J. Bernhardt, “Clinical feasibility of interactive motion-controlled games for stroke rehabilitation.,” *J. Neuroeng. Rehabil.*, vol. 12, no. 1, p. 63, 2015.

[6] Y.-J. J. Chang, S.-F. F. Chen, and J.-D. Da Huang, “A Kinect-based system for physical rehabilitation: A pilot study for young adults with motor disabilities,” *Res. Dev. Disabil.*, vol. 32, no. 6, pp. 2566–2570, Jan. 2011.

[7] A. González, P. Fraisse, and M. Hayashibe, “Adaptive Interface for Personalized Center of Mass Self-Identification in Home

Rehabilitation," *IEEE Sens. J.*, vol. 15, no. 5, pp. 2814–2823, 2015.

[8] M. Li, G. Xu, B. He, X. Ma, and J. Xie, "Pre-impact fall detection based on a modified zero moment point criterion using data from Kinect sensors," *IEEE Sens. J.*, vol. 18, no. 13, pp. 5522–5531, 2018.

[9] S. Gaukrodger *et al.*, "Gait tracking for virtual reality clinical applications: A low cost solution," *Gait Posture*, vol. 37, no. 2013, p. S31, 2013.

[10] S. H. Lee *et al.*, "Measurement of shoulder range of motion in patients with adhesive capsulitis using a Kinect," *PLoS One*, vol. 10, no. 6, 2015.

[11] G. Blumrosen, Y. Miron, N. Intrator, and M. Plotnik, "A real-time kinect signature-based patient home monitoring system," *Sensors*, vol. 16, no. 11, 2016.

[12] J. Shotton *et al.*, "Real-time human pose recognition in parts from single depth images," *Stud. Comput. Intell.*, vol. 411, pp. 119–135, 2013.

[13] W. He, R. Girshick, J. Shotton, P. Kohli, A. Criminisi, and A. Fitzgibbon, "Efficient Regression of General-Activity Human Poses from Depth Images," *Proc. IEEE Int. Conf. Comput. Vis.*, pp. 415–422, 2011.

[14] W. Shen, K. Deng, X. Bai, T. Leyvand, B. Guo, and Z. Tu, "Exemplar-based human action pose correction," *IEEE Trans. Cybern.*, vol. 44, no. 7, pp. 1053–1066, 2014.

[15] Microsoft, "Kinect for Windows SDK 2.0," 2014. [Online]. Available: https://www.microsoft.com/en-us/download/details.aspx?id=44561. [Accessed: 31-Jan-2017].

[16] Open Natural Interaction, "OpenNI 2.2.0.33," 2015. [Online]. Available: https://structure.io/openni. [Accessed: 31-Jan-2017].

[17] OpenKinect, "Libfreenect v0.5.6," 2016. [Online]. Available: https://openkinect.org. [Accessed: 31-Jan-2017].

[18] Microsoft, "Kinect for Windows SDK v1.8," 2013. [Online]. Available: https://www.microsoft.com/en-us/download/details.aspx?id=40278. [Accessed: 31-Jan-2017].

[19] R. A. Clark *et al.*, "Reliability and concurrent validity of the Microsoft Xbox One Kinect for assessment of standing balance and postural control," *Gait Posture*, vol. 42, no. 2, pp. 210–213, Apr. 2015.

[20] R. P. Kuster, B. Heinlein, C. M. Bauer, and E. S. Graf, "Accuracy of KinectOne to quantify kinematics of the upper body," *Gait Posture*, vol. 47, pp. 80–85, 2016.

[21] M. Mcgroarty, D. Meldrum, S. Giblin, H. French, and F. Wetterling, "Variations in knee flexion measurements for overhead squat as measured with marker-based and markerless motion capture systems," *Gait Posture*, vol. 49, pp. 89–90, 2016.

[22] M. Eltoukhy, J. Oh, C. Kuenze, and J. Signorile, "Improved kinect-based spatiotemporal and kinematic treadmill gait assessment," *Gait Posture*, vol. 51, pp. 77–83, 2017.

[23] L. G. Wiedemann, R. Planinc, I. Nemec, and M. Kampel, "Performance evaluation of joint angles obtained by the kinect V2," in *Technologies for Active and Assisted Living (TechAAL), IET International Conference on*, 2015, pp. 1–6.

[24] K. Woolford, "Defining Accuracy in the Use of Kinect V2 for Exercise Monitoring," in *MOCO '15 Proceedings of the 2nd International Workshop on Movement and Computing*, 2015, pp. 112–119.

[25] J. Darby, M. B. Sánchez, P. B. Butler, and I. D. Loram, "An evaluation of 3D head pose estimation using the Microsoft Kinect v2," *Gait Posture*, vol. 48, pp. 83–88, 2016.

[26] E. Auvinet, F. Multon, V. Manning, J. Meunier, and J. P. Cobb, "Validity and sensitivity of the longitudinal asymmetry index to detect gait asymmetry using Microsoft Kinect data," *Gait Posture*, vol. 51, pp. 162–168, 2017.

[27] B. F. Mentiplay *et al.*, "Gait assessment using the Microsoft Xbox One Kinect: Concurrent validity and inter-day reliability of spatiotemporal and kinematic variables," *J. Biomech.*, vol. 48, no. 10, pp. 2166–2170, May 2015.

[28] K. Otte *et al.*, "Accuracy and reliability of the kinect version 2 for clinical measurement of motor function," *PLoS One*, vol. 11, no. 11, pp. 1–18, 2016.

[29] B. Müller, W. Ilg, M. A. Giese, and N. Ludolph, "Validation of enhanced kinect sensor based motion capturing for gait assessment," *PLoS One*, vol. 12, no. 14, pp. 14–16, 2017.

[30] A. Lioulemes, M. Theofanidis, and F. Makedon, "Quantitative analysis of the human upper-limb kinematic model for robot-based rehabilitation applications," *IEEE Conf. Autom. Sci. Eng. (CASE), Fort Worth TX*, pp. 1061–1066, 2016.

[31] A. K. Mishra, M. Skubic, and C. Abbott, "Development and preliminary validation of an interactive remote physical therapy system," *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS*, vol. 2015–Novem, pp. 190–193, Aug. 2015.

[32] M. Ma, R. Proffitt, and M. Skubic, "Validation of a Kinect V2 based rehabilitation game," *PLoS One*, vol. 13, no. 8, p. e0202338, 2018.

[33] B. Hotrabhavananda, A. K. Mishra, M. Skubic, N. Hotrabhavananda, and C. Abbott, "Evaluation of the microsoft kinect skeletal versus depth data analysis for timed-up and go and figure of 8 walk tests," *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS*, vol. 2016–Octob, pp. 2274–2277, 2016.

[34] M. R. Kharazi *et al.*, "Validity of microsoft kinectTM for measuring gait parameters," in *Biomedical Engineering (ICBME), 2015 22nd Iranian Conference on*, 2015, no. November, pp. 375–379.

[35] J. Oh, C. Kuenze, M. Jacopetti, J. F. Signorile, and M. Eltoukhy, "Validity of the Microsoft Kinect™ in assessing spatiotemporal and lower extremity kinematics during stair ascent and descent in healthy young individuals," *Med. Eng. Phys.*, vol. 0, pp. 1–7, 2018.

[36] A. Napoli, S. Glass, C. Ward, C. Tucker, and I. Obeid, "Performance analysis of a generalized motion capture system using microsoft kinect 2.0," *Biomed. Signal Process. Control*, vol. 38, pp. 265–280, 2017.

[37] A. Napoli, S. M. Glass, C. Tucker, and I. Obeid, "The Automated Assessment of Postural Stability: Balance Detection Algorithm," *Ann. Biomed. Eng.*, pp. 1–10, 2017.

[38] C. S. Bamji *et al.*, "A 0.13 μm CMOS System-on-Chip for a 512 × 424 Time-of-Flight Image Sensor with Multi-Frequency Photo-Demodulation up to 130 MHz and 2 GS/s ADC," *IEEE J. Solid-State Circuits*, vol. 50, no. 1, pp. 303–319, 2015.

[39] E. Ceseracciu, Z. Sawacha, and C. Cobelli, "Comparison of markerless and marker-based motion capture technologies through simultaneous data collection during gait: Proof of concept," *PLoS One*, vol. 9, no. 3, pp. 1–7, 2014.

[40] Mr. Naeemabadi, B. Dinesen, O. K. Andersen, and J. Hansen, "Investigating the impact of a motion capture system on Microsoft Kinect v2 recordings: A caution for using the technologies together," *PLoS One*, vol. 13, no. 9, p. e0204052, 2018.

[41] T. D. Collins, S. N. Ghoussayni, D. J. Ewins, and J. A. Kent, "A six degrees-of-freedom marker set for gait analysis: Repeatability and comparison with a modified Helen Hayes set," *Gait Posture*, vol. 30, no. 2, pp. 173–180, 2009.

[42] S. van Sint Jan, *Color Atlas of Skeletal Landmark Definitions*. Elsevier Health Sciences, 2007.

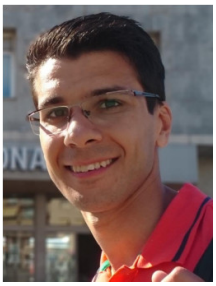
[43] M. P. Kadaba, H. K. Ramakrishnan, and M. E. Wootten, "Measurement of lower extremity kinematics during level walking," *J. Orthop. Res.*, vol. 8, no. 3, pp. 383–392, 1990.

[44] H. Sarbolandi, D. Lefloch, and A. Kolb, "Kinect range sensing: Structured-light versus Time-of-Flight Kinect," *Comput. Vis. Image Underst.*, vol. 139, pp. 1–20, 2015.

[45] J. Sell and P. O'Connor, "The xbox one system on a chip and kinect sensor," *IEEE Micro*, vol. 34, no. 2, pp. 44–53, 2014.

[46] S. Giblin *et al.*, "Bone length calibration can significantly improve the measurement accuracy of knee flexion angle when using a marker-less system to capture the motion of countermovement jump," *3rd IEEE EMBS Int. Conf. Biomed. Heal. Informatics, BHI 2016*, pp. 392–397, 2016.

[47] S. P. R. Gaddam, M. K. Chippa, S. Sastry, A. Ange, V. Berki, and B. L. Davis, "Estimating forces during exercise activity using non-invasive kinect camera," *Proc. - 2015 Int. Conf. Comput. Sci. Comput. Intell. CSCI 2015*, pp. 825–828, 2016.



MReza Naeemabadi received the B.Sc. and M.Sc. degrees in Biomedical Engineering, Bioelectric from Azad University Mashhad and Isfahan University of Medical Science, respectively. His background is mainly focused on biomedical signal processing and telehealth services. Currently, he is pursuing the Ph.D. degree in biomedical engineering at SMI, Health Science and Technology Department, Aalborg University and also affiliated with Laboratory of Welfare Technologies - Telehealth and Telerehabilitation, Aalborg University. His current PhD is focusing on employing human motion trackers in telerehabilitation applications.

Birthe Dinesen is Professor and Head of the Laboratory of Telehealth & and Telerehabilitation at Aalborg University Denmark. Since 2004, Dr. Dinesen has carried out research within telehealth and telerehabilitation. She has been the principal investigator on several groundbreaking telehealth and - rehabilitation studies within chronic disease management. She has been author or co-author of more than 70 peer-reviewed papers in international journals. Dr. Dinesen's research focuses on future technology-enabled care delivery models. Her focus is on participatory design, engagement methods and the use of information technology to improve health, including mobile, social, and distributed computing systems. In 2012, Dr. Dinesen founded the Transatlantic Telehealth Research Network (TTRN) between Danish and American research and health institutions, including UC Berkeley, UC Davis and the Cleveland Clinic. With an international perspective on technology-enabled healthcare delivery, Dr Dinesen has carried outPage 12 of 14 research in Australia, Norway, Sweden, China and the USA. You can read about her research at this link: https://www.labwelfaretech.com/?lang=en.

Ole Kæseler Andersen received the M.Sc. degree in electrical engineering with specialization in biomedical engineering in 1992, and the Ph.D. degree in biomedical engineering in 1996, both from Aalborg University, Aalborg, Denmark. He defended his Dr. Scient. (doctoral degree in natural sciences) thesis in 2007 on the nature and organization of the withdrawal reflex sbmitted to Aalborg University. He is currently a Professor in biomedical engineering at SMI®, Department of Health Science and Technology, Aalborg University. He is head of the Doctoral School in Medicine, Biomedical Science, and Technology, Aalborg University. His research interests include rehabilitation engineering, sensor technology, thermal and electrical stimulation techniques, biomarkers for spinal nociception, and sensory-motor interaction related to the withdrawal reflex.

John Hansen received the M.Sc. degree in signal processing with a specialization in biomedical engineering in 2002 and the Ph.D. degree in biomedical science and engineering in 2007, both from Aalborg University. He is currently an Associate Professor at the Laboratory for Cardio-Technology, Medical Informatics Group, Department of Health Science and Technology, Aalborg University. His position comprises research and teaching within the area of embedded technologies with focus on telemonitoring and health diagnostics. He has authored or co-authored over 60 scientific publications.