Accuracy evaluation of two markerless motion capture systems for measurement of upper extremities: Kinect V2 and Captiv

Tim Steinebach1 | Eric H. Grosse2 | Christoph H. Glock2 | Jurij Wakula1 | Alexander Lunin2

1Institute for Ergonomics and Human Factors, Technische Universität Darmstadt, Darmstadt, Germany
2Institute of Production and Supply Chain Management, Technische Universität Darmstadt, Darmstadt, Germany

Correspondence
Tim Steinebach, Institute for Ergonomics and Human Factors, Technische Universität Darmstadt, Otto-Bernd-Straße 2, 64287 Darmstadt, Germany.
Email: t.steinebach@iad.tu-darmstadt.de

Abstract
Motion capturing is a promising method to assess working postures and human movements and, therewith, the risk of musculoskeletal injuries that could occur while performing manual tasks in industrial settings. To obtain a reliable risk assessment, the motion capture system used has to accurately measure body postures adopted by the worker during the task. This study evaluates the accuracy to measure joint angles of upper extremities of two different motion capture systems, namely the Microsoft Kinect V2 and the Captiv system, for angles of upper extremities. For this purpose, an experimental study was conducted involving 12 subjects performing preset static postures and basic movements, including elbow flexion, shoulder flexion, and shoulder abduction. In addition, to examine whether self-occlusion or occlusion of body parts by work equipment has an impact on the accuracy of the Kinect V2, the subjects handled boxes during some of the tests. As a gold standard, a goniometer for static and an angle scale for dynamic exercises was used. The Captiv system shows high correlation coefficients ($r > .93$) and small mean absolute errors (<5°) for all movements except for elbow flexion. The Kinect V2 has sufficient results for joint angles captured without occlusion as well, but the accuracy significantly decreases when occlusion occurs.

KEYWORDS
accuracy, Captiv, Kinect V2, occlusion, upper extremities

1 | INTRODUCTION

Despite advances in technology and automation, manual human work is still a key factor in many industries. In the logistics sector, for example, around 80% of the warehouses in Western Europe are still mainly operated with manual labor (De Koster, Le-duc, & Roodbergen, 2007). Manual material handling is often performed in awkward postures including stretching and bending of the back with repetitive movements. This can lead to work-related musculoskeletal disorders (MSD), which are one of the main causes for sick-leave and incapacity for work in the industry (BKK Gesundheitsreport, 2013). In Germany, for example, the cost of sick-leave add up to approximately €133 billion per year due to the loss of labor productivity. About 22.8% of these absences are due to MSDs (BAuA, Federal Institute for Occupational Safety & Health, 2018). To reduce the risk for developing MSDs, an ergonomic analysis of industrial workplaces is recommended (and in many countries required by law) to create an appropriate workplace design preventing high physical stresses or overexertions, especially for jobs characterized by frequent manual handling of loads.
Recently, motion capture technologies have been used for an ergonomic assessment of physical work loads on industrial workplaces more frequently. These systems allow the automatic recording of postures and movements during manual working tasks. The collected data can be digitized, graphically visualized and analyzed using statistical computing methods. Although marker-based systems are considered to be the gold standard in laboratory environments (Lai et al., 2013), markerless systems are much better suited for practitioners because they can be used with less elaborate preparation and are significantly cheaper (e.g., Corazza, Mündermann, Gambaretto, Ferrigno, & Andriacchi, 2010). This study evaluates the suitability of two markerless motion capture systems for an ergonomic assessment of working postures and movements of upper extremities: the optical sensor Microsoft Kinect V2 and the body-mounted inertial sensors of the Captiv L7000 system provided by TEA (http://teaergo.com). This study presents an important first step in understanding the applicability of these technologies for upper limb movements and provides insights into the accuracy of the Captiv system since there is little evidence on the system’s reliability in the literature so far.

In the past, the Kinect system has frequently been studied for its accuracy in the assessment of postures and specific movements, usually in laboratories. In the majority of studies, the parameters measured by the Kinect were compared with the data of a gold standard. Most often, marker-based optical systems were used as a gold standard (e.g., Bonnechère et al., 2014; Clark et al., 2015; Dutta, 2012; Kuster, Heinlein, Bauer, & Graf, 2016; Plantard, Müller et al., 2017). Body-mounted sensors (e.g., Huber, Seitz, Leeser, & Sternad, 2015; Romero et al., 2017), goniometers (e.g., Hawi et al., 2014; Lee et al., 2015) and orthogonal reference photographs (e.g., Matsen, Lauder, Rector, Keeling, & Cherones, 2016) have been used as gold standards for postures and motions as well. Building on these findings, the Kinect has already been used for the automation of ergonomic assessments by calculating joint angles, for example using the Ovako Working Posture Assessment (Diego-Mas & Alcaide-Marzal, 2014) or the Rapid Upper Limb Assessment (RULA; Manghisi et al., 2017). The results of these tools calculated by the Kinect data, compared with expert ratings, are sufficient with over 70% correctly assessed RULA scores for the upper body, even under suboptimal conditions in real working environments (Plantard, Shum, Le Pierre, & Multron, 2017). Moreover, some researchers used raw data from multiple Kinect sensors to capture and consolidate movements from different perspectives. Most of them focused on the analysis of human gait (Dolatabadi, Taati, & Mihailidis, 2016; Geerse, Coolen, & Roerdink, 2015; Müller, Ilg, Giese, & Ludolph, 2017). Chen, Lee, and Lin (2015), for example, developed an approach with multiple Kinect V1 sensors to evaluate the accuracy in range-of-motion experiments of the upper extremities. However, researchers agree that the newer generation (Kinect V2) is more accurate than the Kinect V1 for motion capture of upper extremities (Mishra, Skubic, & Abbott, 2015; Wang, Kurillo, Ofli, & Bajcsy, 2015). Therefore, it is recommended to use the current Kinect V2 model for further studies.

One major disadvantage of the Kinect V2 that has been mentioned in the literature is the capturing of work processes in environments where occlusion occurs. Occlusion, in this context, could either be caused by an object (e.g., a box) handled by the subject or by the subject itself, since the field of view of the camera is blocked. This may reduce the accuracy of the optical system (i.e., the Kinect V2). Especially, this is important for practical cases in real work settings, as such occlusion situations often occur (Dzeng, Hsueh, & Ho, 2017; Kuster et al., 2016; Müller et al., 2017). Although Plantard, Shum et al. (2017) indicated that the accuracy of the Kinect V2 is sufficient to calculate RULA values even in situations where occlusion occurs, an accuracy evaluation of joint angle measurement in authentic working environments using multiple Kinect V2 sensors is still lacking. To the best of the authors’ knowledge, prior research has so far not used multiple Kinect V2 systems to capture human motions from different perspectives to reduce the influence of occlusion of upper extremities and to improve the accuracy of the system in real working tasks.

In contrast to the Kinect V1 and V2, the Captiv system has not yet been analyzed for its accuracy at all. So far, this system has only been used to investigate other topics and thus some degree of system accuracy has been assumed. Bartnicka, Ziętkiewicz, and Kowalski (2015), for example, used Captiv to capture body postures of surgeons during surgeries. Furthermore, Bartnicka et al. (2017) examined the motions of orthopedists and medical staff to improve ergonomic working conditions. Taber, Sweeney, Bishop, and Boute (2017) used Captiv to measure the movements of limb, torso, and head to jettison a helicopter push-out window. In another study, Vignais, Bernard, Touvenot, and Sagot (2017) established a continuous RULA computation by supplying a biomechanical model with motion data from Captiv. Furthermore, Captiv is based on inertial sensors, and hence occlusion does not influence the accuracy of the system.

The aim of this study is to quantitatively evaluate and compare the accuracy of a motion capture system consisting of two Kinect V2 sensors and the Captiv system using a subject study including a defined working task (the lifting of boxes). Here, the measured joint angles of the upper extremities are compared with a gold standard for static postures (a goniometer, e.g., Hawi et al., 2014) and movements (an angle scale). The following research questions are addressed in the work at hand.

1. How accurate are the Captiv system and a Kinect V2 system with two sensors in capturing basic postures and movements, both with respect to a gold standard and to one another?

2. How is the accuracy of the two-sensor Kinect V2 system influenced by occlusion of upper extremities?

2 | MATERIALS AND METHODS

2.1 | Analyzed systems

The Kinect V2 is the second generation of an inexpensive, depth-sensing device provided by Microsoft and originally developed for interaction with the Xbox (see Figure 1 left). The time-of-flight technology included in the Kinect V2 allows capturing 30 frames per
second with a depth resolution of 512 × 426 pixels. An algorithm is able to detect a total of 25 joints of the analyzed subject. In this study, the software provided by iPi Soft (http://ipisoft.com/) was used for the data integration of both Kinect sensors and the transformation of the raw data. This program was analyzed and rated more accurate than similar software in earlier studies (e.g., Choppin, Lane, & Wheat, 2014; Choppin & Wheat, 2012). Moreover, other researchers frequently based their findings on calculations obtained with iPi Soft (e.g., Cha et al., 2015; Fornaser, Tomasin, De Cecco, Tavernini, & Zanetti, 2017; Skals, Rasmussen, Bendtsen, Yang, & Andersen, 2017; Zhou, Li, & Bai, 2017).

The second motion capture system used in this study is the Captiv L7000 system provided by the TEA group (see Figure 1). This system uses inertial sensors with integrated magnetometers, accelerometers, and gyroscopes for motion detection. The sensors are fixed on the subject with the help of straps, where a correct attachment of the sensors according to the recommendations of the manufacturer is mandatory for an accurate measurement. A total of 15 sensors are placed on the right and left upper and lower arm, hand (fixed by gloves), thigh and lower leg, foot and centrally on the forehead, back and hip. It is also possible to use fewer sensors if only certain joints are examined—only the sensor of the back is to be applied with each measurement. The battery-powered sensors are wireless and communicate with a recording device worn on the subject’s body. The sampling rate is either 32, 64, or 128 Hz. With the help of the corresponding software provided by TEA, the data from the accelerometers, gyroscopes, magnetometers are transferred into quaternions and then combined by an algorithm into joint angles.

These two systems were chosen for the purpose of our study because, on the one hand, Kinect is a quite inexpensive and widely available system that has often been evaluated in the literature as having sufficient performance. Therefore, it could be especially suitable for companies that do not have the financial capacity for high-priced systems. In addition, unlike many alternative motion capture techniques, the operator is unaffected by markers or cables in his/her work. Captiv, on the other hand, is evaluated because it uses a different approach to capture motion than the Kinect and since it has not been validated yet. In addition, compared with optical motion capture systems, inertial measurement unit-based systems are more suitable for field studies due to their quick setup, which makes the Captiv system attractive for practical analyses.

Both systems are thus well-established complete solutions consisting of hardware and software. Our focus is not on improving the algorithms used by the systems. Instead, the whole system is understood as an integrated measurement tool whose output we analyze. By comparing both systems, we can evaluate two fundamentally different motion capture technologies to provide recommendations for action regarding which system is suited for capturing postures and motion for industrial working tasks.

### 2.2 | Subjects

Twelve young and apparently healthy individuals (age: 23.8 ± 2.6 years, height: 177.3 ± 9.4 cm, weight: 70.9 ± 12.3 kg, seven males and five females) participated in the study, which is comparable to similar studies (e.g., Choppin & Wheat, 2012). All subjects read an ethics statement and signed an informed written consent before the experiments. The subjects wore tight-fitting clothes that allowed an appropriate placement of inertial sensors.

### 2.3 | Experimental procedure

The experiment was set up by placing two Kinect V2 sensors 2.5 and 2.7 m, respectively, in front of the subject and 1.5 m above the ground; see Figure 2. One of the sensors was facing the subject, as this has often been described as an optimal sensor setup (e.g., Skals et al., 2017; Plantard, Muller et al., 2017). In preliminary tests, the second sensor was placed at an angle of 80° to the first sensor (iPi Soft, 2019). This setup was also used in previous studies (e.g. Skals et al., 2017; Stone & Skubic, 2011) and found to be particularly robust before our tests, even for complex movements, and therefore adopted for the experiments. For all experiments in this study, attention was paid to the best possible exclusion of all potential influences, such as ferromagnetic metals, electromagnetic radiation, or inaccurate alignment of the sensors.

To assess the accuracy of the two systems, static postures and basic movements (with and without holding a box in both hands) of 12 volunteers were recorded. The focus was on the joint angles of upper extremities: (a) shoulder abduction/adduction, (b) shoulder

---

**FIGURE 1** Kinect V2 (left), Captiv sensor (middle), and Captiv recording device (right)
Anatomical planes were determined to calculate shoulder angles. For the frontal plane, the connecting vector of the right and left shoulder and the trunk vector, which consists of the connection between the spine base and neck, were selected. Based on this, the sagittal plane was defined. The first support vector was also described via the trunk vector. The second support vector was defined as the normal vector \( \vec{n} \) of the frontal plane (Diego-Mas &...
Alcaide-Marzal, 2014). The following approaches were used to determine the abduction or flexion angle of the shoulder.

1. The angle between the projected vector of the upper arm on the frontal plane and the sagittal plane is consistent with the abduction angle of the shoulder.
2. The angle between the projected vector of the upper arm on the sagittal plane and the frontal plane is consistent with the flexion angle of the shoulder (Lee et al., 2015).

For automation of the described procedure, a visual basic for applications macro was developed. Once the angles measured by the two motion capture systems had been calculated, synchronized and collected in a dataset, the corresponding values of the gold standard were manually added. For this purpose, the values measured with the goniometer in Phase 1 were used. For Phases 2 and 3, time-synchronous measuring points per movement and test person were read off the angle scale and added to the dataset.

2.5 | Statistical analysis

To determine the deviation of the two motion capture systems from the gold standard, the mean absolute error (MAE) was calculated. Assuming that the gold standard represents the true value, the MAE indicates the measurement error of the systems. This type of error calculation has frequently been used for evaluating the Kinect V2 sensor in the literature (e.g., Cippitelli et al., 2015; Eltoukhy, Kuenze, Oh, Wooten, & Signorile, 2017; Nixon, Howard, & Chen, 2013; Schmitz et al., 2015).

The MAE was calculated as follows in this study, where \( n \) is the number of measuring values.

\[
\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |\varphi_{\text{KinectV2/Captiv}} - \varphi_{\text{gold standard}}|
\]  

We followed the recommendation of Willmott and Matsuura (2005) by choosing the MAE instead of the root mean square error (RMSE) as a measure of deviation, because RMSE gives a relatively high weight to large errors and is therefore hard to interpret. However, the RMSE is also reported below to facilitate comparing our results to those of earlier studies that relied on the RMSE. To analyze the differences between the two systems, the Wilcoxon signed-rank tests, was performed (\( \alpha = .05 \)) because the corresponding data were not normally distributed.

In addition, correlation coefficients between the two systems and the gold standard were calculated by comparing the maximum angle, minimum angle and the midpoint angle (angle that results at \( 0.5 \times t \), with \( t = t_{\text{max angle}} - t_{\text{min angle}} \)) of each movement repetition. It should be noted that a high correlation does not necessarily imply a high quantitative agreement of the measured values (Dolatabadi et al., 2016). Two different methods were applied, depending on whether or not the dataset was normally distributed, namely Pearson’s and Spearman’s correlation coefficients. All necessary calculations were performed using the IBM SPSS. Bland–Altman plots were generated for illustration purposes. They investigate the presence of absolute agreement between the measures of the two systems and the gold standard (Bland & Altman, 1986). The average values of the gold standard and Kinect or Captiv measurement \( \frac{x_{\text{gold standard}} - x_{\text{Kinect/Captiv}}}{2} \) are plotted on the x-axis and the differences between the two measured values \( x_{\text{Kinect/Captiv}} - x_{\text{gold standard}} \) are plotted on the y-axis.

3 | RESULTS

This section presents the descriptive results of the subject study concerning the accuracy of both systems for each phase.
3.1 | Results for Phase 1

The MAEs of the systems’ measurements compared with the goniometer are between 2.5° and 12.9° for the Kinect V2 and between 1.9° and 14.9° for the Captiv for static postures (see Table 2). Especially for the elbow, the Captiv system shows high errors, while the angles of the shoulder, were captured very accurately in all postures with deviations <4°.

In general, the correlation coefficients for both systems are very high and significant, except for the Kinect’s measurements of abduction (Posture 1) and flexion (Posture 3) or the angle of the elbow in Captiv, respectively.

3.2 | Results for Phases 2 and 3

After manually evaluating the joint angles during movements using the angle scale, the MAE, the RMSE and the correlation coefficients were calculated again. For Phase 2 (without box) and Phase 3 (with box), these values are summarized in Tables 3 and 4, respectively.

The measurement errors during the abduction of the shoulder are relatively low for the Kinect V2 sensor as well as for the Captiv system with MAE < 3.9° in Phase 2. Adding the box in Phase 3 results in significantly higher values for this movement for the Kinect V2 in contrast to Phase 2 (the Wilcoxon signed-rank test; left: \( p = .01 \)). Overall, however, the mean errors of both systems remain below 5.0° for shoulder abduction. Consequently, the correlations with \( r > .89 \) in the experiments with and without a box are very high and statistically significant. Looking at the abduction in Phase 2, the correlation with coefficients \( r > .97 \) is almost perfect for both motion capture systems.

The flexion of the shoulder results in a notably higher MAE for the Kinect V2 than for the Captiv L7000. Errors of the Captiv system remain below 4.2°. The Kinect V2 has higher deviations, in particular for the left arm that shows larger mean absolute errors during flexion than the right arm in Phase 2 (MAE_{right} = 9.5°, MAE_{left} = 20.9°, the Wilcoxon signed-rank test; \( p = .068 \)) and Phase 3 (MAE_{right} = 10.2°, MAE_{left} = 26.7°, the Wilcoxon signed-rank test; \( p < .001 \)). A discussion on this follows in Section 4.

For flexion of the elbow, the Kinect system also has higher MAEs than Captiv, especially for the detection of the left arm. This is also reflected in the correlations, which are lower for Kinect (\( r < .72 \)) than for Captiv (\( r > .93 \)).

Finally, it can be observed that the measurement deviations of the Kinect system in Phase 3 compared to those of Phase 2 for the flexion of the shoulders and the elbows are significantly higher. The MAE of the left elbow, for example, increases from a deviation of 22.5° in Phase 2 to a very high error of 40.2° in Phase 3 (the Wilcoxon signed-rank test; \( p < .001 \)). The differences in the Captiv measurements between those two phases, however, are much smaller or negligible.

On the basis of the calculated data, Bland-Altman plots are generated, as shown in Figures 4–6.

4 | DISCUSSION

4.1 | Static postures

As can be seen in Table 2, especially the static abduction of the shoulders was captured with very small mean absolute errors by both systems. This applies to both smaller and larger abduction angles,

### TABLE 2  MAE, RMSE, and correlation coefficients—Phase 1

| Phase 1 | MAE | RMSE | Correlation |
|---------|-----|------|-------------|
|         | Kinect V2 | Captiv | Kinect V2 | Captiv | Kinect V2 | Captiv |
| Abduction/adduction shoulder | Posture 1 | Right | 6.5° | 3.3° | 10.0° | 4.5° | 0.146 | 0.984** |
| | Left | 4.9° | 2.8° | 8.2° | 3.4° | 0.120 | 0.947** |
| | Posture 2 | Right | 2.5° | 3.1° | 3.2° | 4.9° | 0.815** | 0.660* |
| | Left | 2.9° | 2.1° | 4.7° | 2.7° | 0.658* | 0.945** |
| Flexion/extension shoulder | Posture 3 | Right | 12.9° | 2.1° | 15.0° | 2.7° | 0.300 | 0.957** |
| | Left | 7.9° | 2.1° | 10.0° | 2.6° | 0.554 | 0.967** |
| | Posture 4 | Right | 10.9° | 1.8° | 13.3° | 2.5° | 0.848** | 0.989** |
| | Left | 8.8° | 1.9° | 11.1° | 2.5° | 0.854** | 0.989** |
| Flexion/extension elbow | Posture 2 | Right | 7.0° | 13.3° | 8.5° | 15.1° | 0.861** | 0.737* |
| | Left | 4.4° | 12.1° | 7.0° | 14.0° | 0.751** | 0.494 |
| | Posture 4 | Right | 4.4° | 13.8° | 5.6° | 16.8° | 0.923** (S) | 0.650° (S) |
| | Left | 4.6° | 14.9° | 5.0° | 15.9° | 0.979** | −0.343 |

Note: (S): Spearman’s correlation coefficient, otherwise Pearson’s correlation coefficient.
Abbreviations: MAE, mean absolute error; RMSE, root mean square error.
\*p < .05.
\**p < .01.
with the errors of Captiv being slightly smaller at ≤3.3°. These values are consistent with the studies of Chen et al. (2015), who obtained a mean error of 4.4° in abduction using two Kinect sensors. It should be noted that the arms of one subject in Posture 1 were not correctly recognized by the Kinect V2. For this reason, the errors in this exercise are higher than for Posture 2, which also has an impact on the correlation coefficients.

Looking at the corresponding Bland–Altman plots (Figure 4), the mean difference is close to zero for Captiv in many cases. Furthermore, the two dashed lines (95% limit of agreement) are closer to each other for the Captiv measurement. The narrower the limit of the agreement is, the more practical is the use of this measuring method (Dolatabadi et al., 2016).

For the flexion of the shoulders, however, we observed a significant accuracy difference (the Wilcoxon signed-rank test) between the two systems. While Captiv delivered very accurate values with MAE ≤ 2.1°, average errors between 7.9° and 12.9° for the individual postures were found for the Kinect V2. Fernandez-Baena, Susin, and Lligadas (2012) observed similar mean errors for the shoulder flexion (7°–10°) measured with Kinect V1. Huber et al. (2015), on the other hand, report mean deviations >15° for the flexion when detected with a V1 model placed in front of the subject. Especially the right side, which can only be captured by one of the two Kinect sensors because it is occluded by the body of the subject, shows very high errors. The Bland–Altman plots (Figure 5) show that the mean difference between the Kinect and the goniometer is negative. Thus, the Kinect V2 underestimates the flexion angle of the shoulder.

Contrary results were found for the flexion of the elbow in static postures. The mean absolute errors of the Kinect measurement were below 7° for both sides and in both postures. The errors of the Captiv system compared with the goniometer measurements, however, were significantly higher with errors ≥12°.

This error may have been caused by the initialization (manual alignment of the sensors to minimize the impact of environmental influences) of the Captiv system, which only had a “moderate” quality according to the software, although the initialization was carried out exactly as recommended by the manufacturer. A moderate initialization quality may not be sufficient for the highest accuracy requirements and has to be taken into account when interpreting the results.

In sum, the Kinect V2 is able to deliver results that are almost as accurate as those obtained by Captiv L7000 in very simple, static situations without occlusion of body parts.

### 4.2 | Basic movements

This subsection evaluates dynamic experiments. The following types of occlusion are of particular interest.

| TABLE 3 | MAE, RMSE, and correlation coefficients without box—Phase 2 |
|---------|---------------------------------|----------------|----------------|----------------|
| Phase 2 | **MAE** | **RMSE** | **Correlation** |
|         | Kinect V2 | Captiv | Kinect V2 | Captiv | Kinect V2 | Captiv |
| Abduction/adduction shoulder | Right | 3.9° | 2.6° | 5.1° | 3.1° | 0.978** | 0.992** |
| | Left | 2.8° | 2.8° | 4.1° | 4.0° | 0.972** | 0.977** |
| Flexion/extension shoulder | Right | 9.5° | 4.1° | 10.9° | 5.0° | 0.972** | 0.991** |
| | Left | 20.9° | 3.0° | 31.6° | 3.9° | 0.494** | 0.990** |
| Flexion/extension elbow | Right | 14.8° | 6.9° | 32.8° | 8.9° | 0.645** | 0.979** |
| | Left | 22.5° | 7.2° | 40.7° | 8.9° | 0.519** | 0.975** |

Note: (S): Spearman’s correlation coefficient, otherwise Pearson’s correlation coefficient.

Abbreviations: MAE, mean absolute error; RMSE, root mean square error.

**p < .01.

### TABLE 4 | MAE, RMSE, and correlation coefficients with box—Phase 3

| Phase 3 | **MAE** | **RMSE** | **Correlation** |
|---------|---------------------------------|----------------|----------------|
|         | Kinect V2 | Captiv | Kinect V2 | Captiv | Kinect V2 | Captiv |
| Abduction/adduction shoulder | Right | 5.0° | 2.0° | 6.5° | 2.5° | 0.891** | 0.996** |
| | Left | 4.3° | 2.8° | 5.2° | 3.6° | 0.977** | 0.993** |
| Flexion/extension shoulder | Right | 10.2° | 4.2° | 11.0° | 5.2° | 0.926** | 0.940** |
| | Left | 26.7° | 3.4° | 35.7° | 4.4° | 0.161 (S) | 0.931** |
| Flexion/extension elbow | Right | 13.8° | 10.7° | 20.5° | 14.1° | 0.721** | 0.939** |
| | Left | 40.2° | 9.4° | 49.1° | 12.5° | 0.013 (S) | 0.952** |

Note: (S): Spearman’s correlation coefficient, otherwise Pearson’s correlation coefficient.

Abbreviations: MAE, mean absolute error; RMSE, root mean square error.

**p < .01.
Self-occlusion (occlusion of the left arm by the right arm).

Occlusion by tools/equipment (body parts covered by the box held between the hands).

As mentioned in the previous section, the mean absolute deviations of both systems for the abduction of the shoulders in Phase 2 with MAE < 4.4° are very low and the correlation coefficients with \( r > .9 \) are high. For these movements without occlusion, both systems can be recommended for use.

In Phase 3, MAEs were higher for the Kinect system. A statistical comparison of the averages shows that the mean absolute errors of abduction in the left shoulder were significantly higher in Phase 3 than in Phase 2 (the Wilcoxon signed-rank test; \( p = .01 \)). This could be due to the fact that the view of the laterally placed sensor during the abduction or flexion is occluded by the box. Although the camera facing the subject should allow a good detection of the movement, the errors increased when the box is held by a subject. The second Kinect sensor (with the software iPi Soft) seems not to be able to fully compensate the occluded view.

With respect to the capturing of the shoulder flexion, Captiv has a maximum MAE of 4.2° in Phases 2 and 3. The significant correlation coefficients of \( r > .93 \) indicate almost perfect accordance to the gold standard. In contrast, the errors of the Kinect system are, similar to the static postures, significantly higher.

To gain insights into the effects of self-occlusion on the results obtained by the Kinect sensors, the errors of the right and left joints of the subjects are compared. The left shoulder has an MAE of 20.9° in Phase 2, while the right arm only has an MAE of 9.5°. Nevertheless, these differences are barely significant for Phase 2 with \( p = .068 \) (\( p < .001 \) in Phase 3). This effect can be explained by the fact that during the simultaneous shoulder flexion of both sides, the left arm is occluded by the right one. The second Kinect sensor, which can capture both arms uncovered due to the different positioning, is not able to compensate the occlusion of the first sensor—neither in the phase with nor without box.

Bland–Altman plots in Figure 6 graphically illustrate these errors. The dotted limit-of-agreement lines in the right plots (no self-occlusion) are much closer to each other than in the left plots. Self-occlusion also becomes evident from the high number of measurements of the left shoulder located in the negative range of the differences (y-axis) of the left diagrams. These high negative differences occurred when the Kinect was unable to detect the joints of the occluded arm. The corresponding iPi Soft visualizations show that, in these cases, the arm points down parallel to the upper body, even though the arm is actually flexed at 90°. In addition, the influence of the occlusion of the box can also be recognized for shoulder flexion. The MAEs in the third phase is again higher than in the second phase.

**Figure 4** Bland–Altman plot—Phase 1: Shoulder abduction
The detection of joint angles during elbow flexion is not completely satisfactory for both systems. The problem of the initialization of Captiv, which has already been described in the static phase, is also recognizable during movements. Mean absolute errors between 6.9° and 10.7° could be insufficient for specific clinical studies or ergonomic analyses. Whether a system can still be used in investigations depends ultimately on the context of the analysis and on whether or not the conclusion of an analysis changes due to the measurement error (Myles & Cui, 2007).

The detection of the joint angles of the elbow by the Kinect V2 sensor has even higher errors than the measurement with Captiv. In contrast to the relatively accurate results of the static motion capture of the elbow flexion, the detection of dynamic movements by the Kinect sensor failed for some subjects. Again, the self-occlusion of body parts seems to be problematic. The mean absolute error of the left elbow in Phase 2 is higher (MAE = 22.5°) than the one obtained for the right elbow (MAE = 14.8°).

In addition to the mathematical error calculations presented in this section, qualitative observations were made repeatedly with Kinect. A systematic error occurred in nearly half of all movements where a box was handled. The iPi Soft visualization shows that Kinect falsely identifies the box that was held by the subjects between the hands as a leg of the subject (see Figure 7). So, Kinect may incorrectly identify a work item as a body part and thus record erroneous movements. The error pattern described here had only a marginal influence on the joint angle of the upper body, as only one of the legs was affected. Nevertheless, this misinterpretation of the system makes examinations difficult or impossible in industrial practice, especially if the angles of the lower body are of relevance.

The results of this study are subject to limitations. First, it has to be taken into account that the two motion capture systems were only evaluated in this study with respect to three different joint angles. Furthermore, the Kinect system can face further challenges in a real work environment. Daylight, moving objects in the background, suboptimal sensor placement and larger areas of the subject’s movements may further reduce the accuracy of the motion capturing. In addition, the experiments were carried out only with relatively young participants (20–27 years), which does not necessarily correspond to the average age of persons performing manual operations, for example, in logistics.

5 | CONCLUSION

The aim of this study was to quantitatively compare the accuracy of the Kinect V2 and the Captiv L7000 motion capture systems in a subject experiment study. The joint angles of the upper extremities measured by both systems were compared with a reference system for static postures and movements. A central objective of this
experimental study was to investigate if a second Kinect V2 sensor, which detects the motions from another angle, is able to compensate the problem of occlusion that results, for example, from the handling of a box. This disadvantage has been described earlier in the literature. The results of the study showed that the angles of the upper extremities have significantly higher measurement errors when occluded by either a work item or the subjects themselves. This especially applies for the flexion of the shoulders and elbows.

We conclude that the Kinect V2 sensor and the Captiv L7000 system are both able to obtain accurate results in situations without occlusion for joint angles of upper extremities. However, the flexion angle of the shoulder is underestimated during measurement with the Kinect V2. In addition, Captiv is less accurate in determining elbow angles; in the case of this study possibly due to an insufficient initialization quality. For many clinical studies, these characteristics are sufficient as long as there are no occluded extremities while capturing with Kinect. In the next step, motion data captured by Captiv in a real manufacturing environment could be integrated in ergonomic risk assessment, for example, RULA or the European Assembly Worksheet.

Moreover, it may be worthwhile to test the Captiv system in further studies for its accuracy in more complex movement sequences to further demonstrate its applicability to real working environments. In addition, it would be desirable to assess the effect of initialization quality on the accuracy of the Captiv system. Also, the capturing with more than two Kinect sensors or the integration of raw data with another software than iPi Soft could be evaluated in an extension of this study to analyze if this leads to higher accuracy.
CONFLICT OF INTERESTS
The authors declare that there are no conflict of interests.

ORCID
Tim Steinebach http://orcid.org/0000-0003-1224-9121

REFERENCES

Aptel, M., Aublet-Cuvelier, A., & Cnockaert, J. C. (2002). Work-related musculoskeletal disorders of the upper limb. Joint, Bone, Spine, 69, 546–555.

Bartnicka, J., Piedrabuena, A., Portilla, R., Mleczko, K., Moyano, C., Pagador, J.,... Tokarczyk, J. (2017). Ergonomic education in orthopedic surgery. LIFE: International Journal of Health and Life Sciences, 3, 194–215.

Bartnicka, J., Ziżtikiewicz, A., & Kowalski, G. (2015). Ergonomic analysis of surgeries with the use of wireless body postures measurement system. Innovations in Manufacturing Management and Engineering, 2, 361–371.

BAuA–Federal Institute for Occupational Safety and Health (2018). Economic Costs due to Incapacity for Work 2016. Retrieved from https://www.baua.de/DE/Themen/Arbeitswelt-und-Arbeitsschutz-im-Wandel/Arbeitsweltberichterstattung/Kosten-der-AU/pdf/Kosten-2016.pdf?__blob=publicationFile&v=4

BKK Gesundheitsreport (2013). BKK Gesundheitsreport 2013–Gesundheit in Bewegung. Berlin, Germany: BKK Dachverband e. V.

Bland, J. M., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. Lancet, 1, 307–310.

Bonnechère, B., Jansen, B., Salvia, P., Bouzahouene, H., Sholukha, V., & Cornelis, J. (2014). Determination of the precision and accuracy of morphological measurements using the Kinect™ sensor. Comparison with standard stereophotogrammetry. Ergonomics, 57, 622–631.

Brosseau, L., Balmer, S., Tousignant, M., O’Sullivan, J. P., Goudreault, C., Bland, J. M., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. Lancet, 1, 307–310.

Buckley, P. W., & Devereux, J. J. (2002). The nature of work-related neck and upper limb musculoskeletal disorders. Applied Ergonomics, 33, 207–217.

Cha, K., Lee, E. Y., Heo, M. H., Shin, K. C., Son, J., & Kim, D. (2015). Analysis of climbing postures and movements in sport climbing for realistic 3d climbing animations. Procedia Engineering, 112, 52–57.

Chapleau, J., Canet, F., Petit, Y., Laflamme, G. Y., & Rouleau, D. M. (2011). Validity of goniometric elbow measurements: Comparative study with a radiographic method. Clinical Orthopaedics and Related Research, 469, 3134–3140.

Chen, Y. C., Lee, H. J., & Lin, K. H. (2015). Measurement of body joint angles for physical therapy based on mean shift tracking using two low cost Kinect images. 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 703–706.

Choppin, S., Lane, B., & Wheat, J. (2014). The accuracy of the Microsoft Kinect in joint ang-le measurement. Sports Technology, 7, 98–105.

Choppin, S., & Wheat, J. (2012). Marker-less tracking of human movement using Microsoft Kinect. 30th Annual Conference of Biomechanics in Sports—Conference Proceedings Archive, 86, 231–234.

Cippitelli, E., Gasparrini, S., Spinsante, S., Gambi, E., Verdini, F., Burattini, L., & Fioretti, S. (2015). Validation of an optimized algorithm to use Kinect in a non-structured environment for Sit-to-Stand analysis. Engineering in Medicine and Biology Society (EMBC), 5078–5081.

Clark, R. A., Pua, Y.-H., Oliveira, C. C., Bower, K. J., Thilagarajah, S., & McGaw, R. (2015). Reliability and concurrent validity of the Microsoft Xbox One Kinect for assessment of standing balance and postural control. Gait & Posture, 42, 210–213.

Corazza, S., Mündermann, L., Gambaretto, E., Ferrigno, G., & Andriacchi, T. P. (2010). Markerless motion capture through visual hull, articulated ICP and subject specific model generation. International Journal of Computer Vision, 87, 156–169.

De Koster, R., Le-duc, T., & Roodbergen, K. J. (2007). Design and control of warehouse order picking: A literature review. European Journal of Operational Research, 182, 481–501.

Diego-Mas, J. A., & Alcaide-Marzal, J. (2014). Using Kinect™ sensor in observational methods for assessing postures at work. Applied Ergonomics, 45, 976–985.

Dolatabadi, E., Taati, B., & Mihailidis, A. (2016). Concurrent validity of the Microsoft Kinect for Windows v2 for measuring spatiotemporal gait parameters. Medical Engineering and Physics, 38, 952–958.

Dutta, T. (2012). Evaluation of the Kinect™ sensor for 3-D kinematic measurement in the workplace. Applied Ergonomics, 43, 645–649.

Dzeng, R. J., Hsueh, H. H., & Ho, C. W. (2017). Automated Posture Assessment for construction workers. 40th International Convention on Information and Communication Technology, Electronics and Microelectronics, 1027–1031.

Eloukhly, M., Kuenze, C., Oh, J., Wooten, S., & Signorile, J. (2017). Kinect-based assessment of lower limb kinematics and dynamic postural control during the star excursion balance test. Gait & Posture, 58, 421–427.

Fernandez-Baena, A., Susin, A., & Liligadas, X. (2012). Biomechanical Validation of Upper-Body and Lower-Body Joint Movements of Kinect Motion Capture Data for Rehabilitation Treatments. 4th International Conference on Intelligent Net-working and Collaborative Systems, 656–661.

Fornar, A., Tomasini, P., De Cecco, M., Tavarnini, M., & Zanetti, M. (2017). Automatic graph based spatiotemporal extrinsic calibration of multiple Kinect V2 ToF cameras. Robotics and Autonomous Systems, 98, 105–125.

Geerse, D. J., Coolen, B. H., & Roerdink, M. (2015). Kinematic validation of a multi-Kinect v2 instrumented 10-meter walkway for quantitative gait assessments. PLOS One, 10, 1–15.

Hawi, N., Liodakis, E., Musolf, D., Suero, E. M., Stuebig, T., Claassen, L., & Cikat, M. (2014). Range of motion assessment of the shoulder and elbow joints using a motion sensing input device: A pilot study. Technology and Health Care, 22, 289–295.

Huber, M. E., Seitz, A. L., Leeser, M., & Sternad, D. (2015). Validity and reliability of Kinect skeleton for measuring shoulder joint angles: A feasibility study. Physiotherapy, 101, 389–393.

Huber, M. E., Seitz, A. L., Leeser, M., & Sternad, D. (2015). Validity and reliability of Kinect skeleton for measuring shoulder joint angles: A feasibility study. Physiotherapy, 101, 389–393.

Koster, R. P., Heinlein, B., Bauer, C. M., & Gra, E. S. (2016). Accuracy of KinectOne to quantify kinematics of the upper body. Gait & Posture, 47, 80–85.

Lai, C. L., Tseng, S. Y., Huang, C. H., Pei, C., Chi, W. M., Hsu, L. C., & Sun, T. L. (2013). Fun and accurate static balance training to enhance fall prevention ability of aged adults: A preliminary study. Human Factors and Ergonomics in Manufacturing & Service Industries, 23, 517–527.

Lee, S. H., Yoon, C., Chung, S. G., Kim, H. C., Kwak, Y., Park, H. W., & Kim, K. (2015). Measurement of shoulder range of motion in patients with adhesive capsulitis using a kinect. PLOS One, 10, 1–12.

Manghisi, V. M., Uva, A. E., Fiorentino, M., Bevilacqua, V., Trotta, G. F., & Corazza, S., Mündermann, L., Gambaretto, E., Ferrigno, G., & Andriacchi, T. P. (2010). Markerless motion capture through visual hull, articulated ICP and subject specific model generation. International Journal of Computer Vision, 87, 156–169.

Manshadi, V. M., Uva, A. E., Fiorentino, M., Bevilacqua, V., Trotta, G. F., & Corazza, S., Mündermann, L., Gambaretto, E., Ferrigno, G., & Andriacchi, T. P. (2010). Markerless motion capture through visual hull, articulated ICP and subject specific model generation. International Journal of Computer Vision, 87, 156–169.

Matsen, F. A., Lauder, A., Rector, K., Keeling, P., & Cherones, A. L. (2016). Measurement of active shoulder motion using the Kinect, a commercially available infrared position detection system. Journal of Shoulder and Elbow Surgery, 25, 216–223.
Milanese, S., Gordon, S., Buettner, P., Flavell, C., Ruston, S., & Coe, D. (2014). Reliability and concurrent validity of knee angle measurement. Smart phone app versus universal goniometer used by experienced and novice clinicians. *Manual Therapy*, 19, 569–574.

Mishra, A. K., Skubic, M., & Abbott, C. (2015). Development and preliminary validation of an interactive remote physical therapy system. *37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 190–193.

Müller, B., Ilg, W., Giese, M. A., & Ludolph, N. (2017). Validation of enhanced Kinect sensor based motion capturing for gait assessment. *PLOS One*, 12, 1–18.

Myles, P. S., & Cui, J. (2007). Using the Bland–Altman method to measure agreement with repeated measures. *British Journal of Anaesthesia*, 99, 309–311.

Nixon, M. E., Howard, A. M., & Chen, Y. (2013). Quantitative evaluation of the Microsoft Kinect for use in an upper extremity virtual rehabilitation environment. *2013 International Conference on Virtual Rehabilitation*, 222–228.

Plantard, P., Muller, A., Pontonnier, C., Dumont, G., Shum, H. P., & Multon, F. (2017). Inverse dynamics based on occlusion-resistant Kinect data: Is it usable for ergonomics? *International Journal of Industrial Ergonomics*, 61, 71–80.

Plantard, P., Shum, H. P., Le Pierres, A. S., & Multon, F. (2017). Validation of an ergonomic assessment method using Kinect data in real workplace conditions. *Applied Ergonomics*, 65, 562–569.

Romero, V., Amaral, J., Fitzpatrick, P., Schmidt, R. C., Duncan, A. W., & Richardson, M. J. (2017). Can low-cost motion-tracking systems substitute a Polhemus system when researching social motor coordination in children? *Behavior Research Methods*, 49(2), 588–601.

Rothstein, J. M., Miller, P. J., & Roettger, R. F. (1983). Goniometric reliability in a clinical setting. *Physical Therapy*, 63, 1611–1615.

Schmitz, A., Ye, M., Bogess, G., Shapiro, R., Yang, R., & Noehren, B. (2015). The measurement of in vivo joint angles during a squat using a single camera markerless motion capture system as compared to a marker based system. *Gait & Posture*, 41, 694–698.

Skals, S., Rasmussen, K. P., Bendtsen, K. M., Yang, J., & Andersen, M. S. (2017). A musculoskeletal model driven by dual Microsoft Kinect Sensor data. *Multibody System Dynamics*, 41, 297–316.

Stone, E. E., & Skubic, M. (2011). Evaluation of an inexpensive depth camera for passive in-home fall risk assessment. *5th International ICST Conference on Pervasive Computing Technologies for Healthcare*, 71–77.

Tabor, M. J., Sweeney, D., Bishop, N., & Boute, R. (2017). Factor effecting the capability to jettison an S92 push-out window. *International Journal of Industrial Ergonomics*, 58, 79–89.

Vignais, N., Bernard, F., Touvenot, G., & Sagot, J. C. (2017). Physical risk factors identification based on body sensor network combined to videotaping. *Applied Ergonomics*, 65, 410–417.

Wang, Q., Kurillo, G., Ofli, F., & Bajcsy, R. (2015). Evaluation of pose tracking accuracy in the first and second generations of Microsoft Kinect. *Proceedings of the 2015 International Conference on Healthcare Informatics*, 380–389.

Willmott, C. J., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research*, 30, 79–82.

Xu, X., Robertson, M., Chen, K. B., Lin, J. H., & McGorry, R. W. (2017). Using the Microsoft Kinect™ to assess 3-D shoulder kinematics during computer use. *Applied Ergonomics*, 65, 418–423.

Zhou, L., Li, Y., & Bai, S. (2017). A human-centered design optimization approach for robotic exoskeletons through biomechanical simulation. *Robotics and Autonomous Systems*, 91, 337–347.

How to cite this article: Steinebach T, Grosse EH, Glock CH, Wakula J, Lunin A. Accuracy evaluation of two markerless motion capture systems for measurement of upper extremities: Kinect V2 and Captiv. *Hum. Factors Man.* 2020;30:291–302. https://doi.org/10.1002/hfm.20840