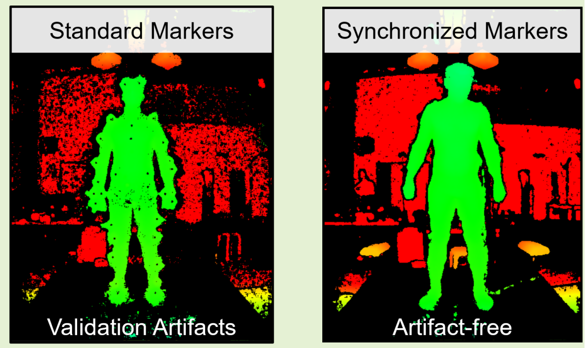
This article has been accepted for publication in IEEE Sensors Journal. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2025.3531760

Minimizing Validation Artifacts in Azure Kinect Caused by Marker-Based Systems using Synchronized Markers

Nikolai Hepke, Moritz Scherer, Benjamin Weyers, Steffen Müller, Jörg Lohscheller

***Abstract*—The Azure Kinect is a human pose tracking device suitable for a wide range of applications. To evaluate its accuracy, it is frequently compared to marker-based motion tracking systems, which are regarded as the gold standard in human pose tracking. However, the infrared light emitted by this system, as well as the reflective markers attached to the body, significantly impair the Kinect's ability to track the body, thus hindering a meaningful validation of the system. To address this issue, we developed a novel set of active markers synchronized with the Kinect's shutter, effectively eliminating these sources of interference. The system's efficacy was assessed through a study involving 10 participants, who performed 10 distinct exercises using both the newly developed synchronized markers and the standard Vicon markers. When using the synchronized markers, as opposed to the standard ones, the median positional error across all joints decreased from 2.4 cm to 0.96 cm, and the Pearson correlation across all joints increased from a mean of 0.48 to 0.70. The results demonstrate that the synchronized markers do not cause interference and are not detrimental to the Kinect's capacity for accurate and reliable individual tracking. Therefore, the synchronized markers are considerably better suited for validating the Kinect against a marker-based reference system. To enable others to benefit from these improvements, we have published an assembly manual alongside the necessary resources.**



***Index Terms*—Artifact Reduction, Azure Kinect, Human Motion Capturing, Simultaneous Measurement, Synchronized Markers, Validation, Vicon**

I. Introduction

PRECISE, low-cost human motion capturing (MoCap) technology has extensive applications across various fields, including human-robot interaction [1], fall detection [2], sports [3] and gaming [4]. Although numerous methods rely on RGB videos to estimate human poses, certain sensors utilize depth images instead, such as the Microsoft Kinect.

With the release of the Azure Kinect DK, Microsoft introduced the third iteration of the Kinect. Unlike its predecessors, this version is specifically designed for AI applications rather than gaming. It is equipped with an infrared Time of Flight (ToF) depth sensor, which enables it to create an accurate depth image of the scene, and an RGB camera, which captures synchronized color images. The Azure Kinect Body Tracking SDK employs machine learning algorithms to accurately determine the position and pose of multiple individuals simultaneously[1].

Certain fields place special demands on the quality of MoCap systems. Thus, to utilize the Azure Kinect for physiotherapy [5], gait analysis [6], or tele-rehabilitation [7], it is crucial to quantify the accuracy of the tracked pose as precisely as possible.

Therefore, numerous studies have compared the Azure Kinect to a marker-based motion capture (MoCap) system (e.g. Vicon by Vicon Motion Systems, Oxford, UK or the Qualisys by Qualisys AB, Sweden), which are considered to be the gold standard in the field. These systems employ retro-reflective spherical markers, strategically positioned at specific anatomical locations on the human body. Multiple synchronized infrared cameras, equipped with strobing infrared LEDs, are arranged around the subject. This configuration ensures that the light emitted by each camera is reflected back by the markers, facilitating precise tracking of each marker.

It is well documented that both the markers and the strobing infrared light negatively affect the Kinect V2's ability to accurately assess human poses [8] [9]. This interference appears to be even stronger on the Azure Kinect [10] [11] [12]. The markers intensely reflect the infrared light emitted by the depth

Manuscript received 24. September 2024; This work was funded in part by the Ministry for Science and Health of Rhineland-Palatinate as part of the research training group (Forschungskolleg) "Immersive Extended Reality for Physical Activity and Health" (XR-PATH).

Nikolai Hepke, Moritz Scherer, Steffen Müller and Jörg Lohscheller are with the Trier University of Applied Sciences, 54293 Trier, Germany (e-mail: hepken@hochschule-trier.de)

Benjamin Weyers is with the Trier University, 54296 Trier

[1]https://learn.microsoft.com/de-de/azure/kinect-dk/body-sdk-setup

This article has been accepted for publication in IEEE Sensors Journal. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2025.3531760

sensor, causing the infrared camera's pixels to oversaturate. This oversaturation leads to the invalidation of specific pixels in the depth image[2]. The interference caused by the markers is referred to as *passive noise*. Furthermore, the strobing infrared light introduces noise in the depth image, which is known as *active noise* [8]. The Azure Kinect Body Tracking SDK relies on infrared and depth images for pose tracking. Consequently, any noise affecting these images negatively impacts the accuracy of the estimated pose.

To effectively evaluate the accuracy of the Azure Kinect, it is crucial to devise a method that allows for comparison with a gold standard without influencing the Kinect's performance. After reviewing existing approaches, we propose an alternative technique called *Synchronized Markers* that addresses this need. To assess the effectiveness of this solution, a study was conducted. Differences in artifacts are highlighted, and the tracking results are presented and discussed, culminating in a comprehensive conclusion.

II. Related Works

There are multiple ways to reduce or eliminate the interferences between the two systems. For highly repeatable movements, a common approach is to separately record the trials of the Kinect and the ground truth reference system. By conducting multiple runs and calculating the average across these trials, any potential interference between the two systems can be effectively eliminated. This was used, for instance, to evaluate walking trials on a treadmill [10], single joint movements [13], [14] and, with the help of a metronome and precise instructions, complex movements like a squat or a lunge [15]. However, due to non-simultaneous data collection, this method is inherently prone to error. It relies heavily on participants being consistent in their repetitions, as any variance in movements between trials will be indistinguishable from measurement error.

Interference can be completely eliminated by substituting the participant with a mannequin. Since the mannequin can hold its pose indefinitely, it allows for independent measurements of exactly the same pose. This method was used to measure the effects of the marker-based system on the Kinect [8] and to compare the Kinect with its predecessors [16] or with itself in different operating modes [17] without the need for a marker-based system. However, only limited information about the system's performance can be obtained, as only a narrow range of static body poses can be examined. Hence, no studies have been found that used this method for anything other than a normal standing pose.

To eliminate the effects of active noise caused by the reference system, one study used the external synchronization port of the Kinect to trigger the Qualisys system in sync with the Kinect's shutter [12]. This solution limits the frame rate of the reference system to that of the Kinect but it enables the recording of complex movements undisturbed by active noise. Smaller markers, with a diameter of 2.5 mm, were employed to reduce the effects of the passive noise. However, the findings of [10] reveal that even markers with a diameter as small as 3 mm can significantly impact the tracking accuracy of the Azure Kinect. The study concluded that the Azure Kinect struggled to consistently generate a reliable skeleton during concurrent measurements.

A different study identified the interference caused by the Qualisys markers and developed a new marker set configuration [18]. They modeled the shanks as a rigid body with six markers and removed those that were facing the Kinect. As the position of all markers on the rigid body was previously recorded, the position of the missing markers could be estimated based on the locations of the remaining markers. However, this solution is ineffective if a person needs to turn around or to be recorded from multiple angles simultaneously. In contrast to other studies, they found no effects of active noise and concluded that the interference from markers on the hips, shoulders, feet, and wrists was insignificant.

Several studies suggest that the impact of the interference is smaller. One study found that the effects of the Qualisys might not affect the Kinect's measurements, as long as the subject is closer than 2.5 m to the Kinect, while the Kinect V2 was not affected at all [11]. Further studies do not discuss any interference between the two systems, nor any measurements taken to mitigate their effects [5], [19], [6], [20].

Given the varied and sometimes contradictory reports on the severity of artifacts, our goal is to quantify the impact of standard markers on the tracking performance of the Kinect Body Tracking SDK. We compare this with the tracking performance achieved using our solution: the custom-built, synchronized markers.

III. Methods

The following section describes the design of the synchronized markers and the study conducted to quantify their effects on the measurements compared to those of the standard markers.

A. The Kinect Body Tracking SDK

Microsoft releases very little information about the inner workings of the Kinect Body Tracking SDK. However, Microsoft held a presentation at ICIP2019 and at CVPR2019[3] where the core mechanics of the Body Tracking SDK were explained. According to that presentation[4], the SDK employs a CNN to estimate the 2D joint positions of the skeleton in the infrared images. Subsequently, a 3D kinematic model is fitted to the depth image by re-projecting the 2D skeleton onto the depth image and adjusting the joints according to anatomical limits. The SDK outputs the joint positions and orientations of the kinematic model structured as a skeleton. However, as both the infrared image and the depth image are disturbed by the noise, a difference in tracking quality is hypothesized, as illustrated in Fig. 1.

For each individual joint, the Body Tracking SDK offers four levels of confidence[5]:

[2]https://learn.microsoft.com/en-us/azure/kinect-dk/depth-camera

[3]https://www.microsoft.com/en-us/research/project/skeletal-tracking-on-azure-kinect/

[4]https://www.microsoft.com/en-us/research/uploads/prod/2020/01/AKBTSDK.pdf

[5]https://microsoft.github.io/Azure-Kinect-Body-Tracking/release/1.1.x/namespace_microsoft_1_1_azure_1_1_kinect_1_1_body_tracking_adfff503ebc1491373c89e96887cad226.html

This article has been accepted for publication in IEEE Sensors Journal. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2025.3531760

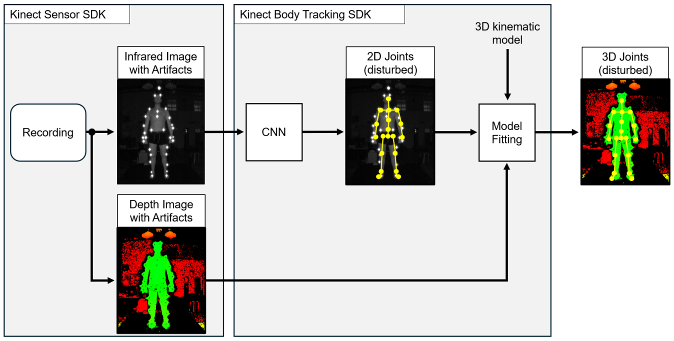


Fig. 1. Process of skeleton generation by the SDK, with noise introduced by the tracking system (based on Microsoft's presentation[4]).

1) None: Joint out of range
2) Low: Joint occluded
3) Medium: Joint observed
4) High: Currently not in use

The position of each joint identified with a confidence level of *Low* or *None* is determined exclusively by the position of other joints, rather than by actual depth data. Consequently, these joints are classified as *untracked*, whereas the joints that receive a confidence level of *Medium* are referred to as *tracked*.

B. Synchronized markers

Since any retro-reflective surface introduces noise within the Kinect recordings, we opted to use active instead of passive markers. Active markers emit infrared light themselves instead of reflecting an external light source, eliminating the need for the Vicon's infrared emitters. Thus, a set of active markers can eliminate both active and passive noise. Constantly glowing active markers would still appear as noise in the Kinect's infrared and depth images. Therefore, the Kinect's synchronization ports are used to strobe the markers with the shutter of the Kinect, illuminating them only between the Kinect's shutters.

1) Marker control: The Arduino Micro[6] is used to control the timing of marker illumination. For every shutter of the color camera, the master Kinect sends a high signal to the 3.5 mm sync-out port. The signal's duration is about 8 microseconds with an electric potential of 5V TTL/CMOS[7]. To ensure consistent recognition, it is connected through an audio jack to one of the Arduino's interrupt pins.

When a signal is received, the Arduino waits for the Kinect's depth shutter to complete before illuminating the markers for 10 milliseconds. Given that the Vicon records at 120 Hz, capturing a frame every 8.3 milliseconds, this ensures that the markers are visible to the Vicon for at least one frame per Kinect frame.

When no signal is received for more than one second at a time, the marker set switches to the *always-on* operation mode to continuously illuminate the markers. This allows for an easy differentiation between a problem with the power supply and a problem with the generation or transmission of the synchronization signal. Furthermore, it facilitates the recording of an uninterrupted *functional calibration trial*, which is required for the Nexus software[8]. This mode deactivates automatically once a synchronization signal is received and the markers return to the default operation mode.

[6]https://store.arduino.cc/products/arduino-micro
[7]https://learn.microsoft.com/en-en/azure/kinect-dk/multi-camera-sync

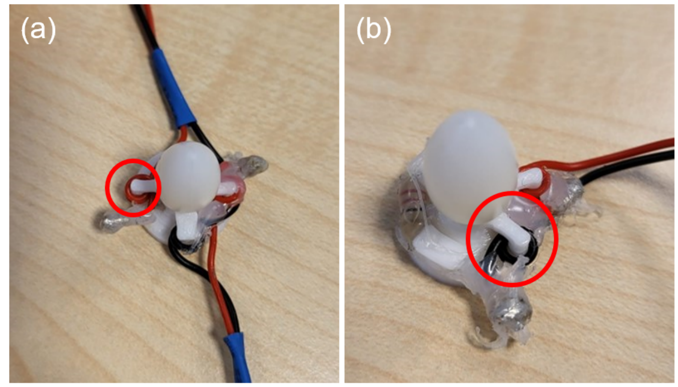


Fig. 2. Two types of marker bases. a) Connecting marker without resistor. b) End marker with resistor. All wires are spun around the strain relieves (red circles). Soldering is coated with hot glue for isolation and to cover sharp edges.

2) Markers: In order for the markers to be sufficiently bright to be detected by the Vicon, 100 mA 850 nm 1.45 V LEDs were used. Two LEDs were daisy-chained and connected to a 20 Ω resistor. In certain positions within the marker layout, the daisy-chain configuration proved impractical due to the absence of nearby markers. These LEDs were soldered to a 35 Ω resistor. Since the LEDs had a narrow beam angle of 17.5°, an additional diffuser was used to spread the light in all directions.

A custom-designed base was used to keep the LEDs in place and to provide a sufficient surface area for it to be fixed on the skin with tape. The base was designed in two different variants: one with the ability to carry a resistor (*end marker*) and one without (*connecting marker*), both shown in Fig. 2. As the participants had to wear the harness of markers for multiple hours, it was to be expected that they exert some strain on the cables. In order to protect the soldering connection, the most vulnerable part of the cabling, each base was equipped with strain reliefs. The wires were wrapped around each handle of the 3D-printed base to ensure that any force applied to the cable was transferred to the handles rather than the solder joint.

3) Power supply and cabling: The set of synchronized markers was designed for the Plug-In Gait full body marker set[9], which was supplemented by two additional markers on the spine to measure the curvature of the back. Of the 41 markers, 38 were daisy-chained in pairs. The resulting 19 pairs and 3 single markers were connected in parallel to each other. The schematics of the cabling are provided in the assembly manual. As each of these marker strings consumes 100 mA at 5 V, the total power consumption of the entire harness is 11 W (2.2 A at 5 V). For safety reasons, the entire system was powered by two power banks (one for the upper body and one for the

[8]https://www.vicon.com/software/nexus/
[9]https://help.vicon.com/space/Nexus216/11607226/Full+body+modeling+with+Plug-in+Gait

This article has been accepted for publication in IEEE Sensors Journal. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2025.3531760

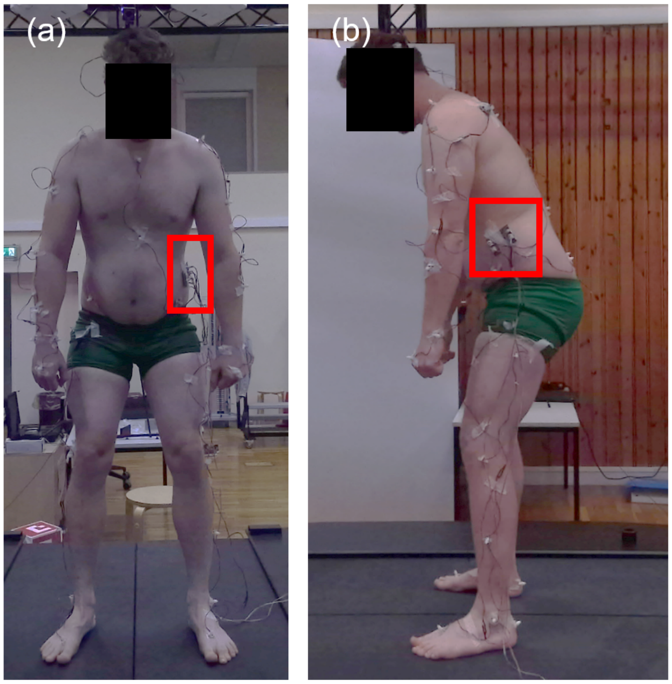


Fig. 3. Subject wearing the synchronized markers in a) frontal view and b) side view. Red Rectangle: Power distributor. Excess wire is spun in a loop and taped to the skin. For privacy reasons, the image is anonymized.

lower body), rather than connecting it to the main grid. Since both power banks are equipped with short circuit protection, any mechanical failure causing a short circuit would shut off the power supply, preventing the cabling from heating up. As the markers are never turned off for more than one second at a time, they always draw enough current for the power bank to stay active.

The Arduino controls two MOSFET modules, which are used as electronic switches to route the power from the supply to the markers. Both are connected by long cables to the power distributor of the harness. This setup allows for the power banks, the Arduino and the MOSFETs to be placed well away from the participant, protecting both the control electronics from disturbances caused by movements of the subject and the subject from the very unlikely event of one of the power banks overheating or catching fire.

The harness, which is shown in Fig. 3, is split into seven parts to facilitate the process of the marker placement on the subject's body: arms, legs, spine and head, lower back and frontal torso. Each part is connected to the main power distributor, located over the left hip, via a plug, except for the arms, which are connected to the cabling of the spine via a custom connector over the shoulder blades. Each of the connections on the shoulder blades carries a colored LED, which serves as a visual indicator for the correct functionality of the harness. The cables connecting the markers were designed to be long enough to accommodate a person up to 2 meters tall. Excess wire was spun together and taped to the skin to prevent entanglement.

4) Assembly: We have published a comprehensive step-by-step manual for the assembly of the harness on Zenodo [21]. This resource includes supplementary diagrams, all required 3D models, and the Arduino code for the control logic. Additionally, it features a Python program designed to synchronously manage the recordings across all four Kinects. We invite others to adopt this system for their research endeavors and to further build upon and improve it.

C. Experimental setup

The following section describes the experiment that was conducted to assess the feasibility of the markers.

1) Population: Ten participants (5w/5m, 28.4 ± 3.5 years, 174.9 ± 9.3 cm, 73.9 ± 8.3 kg) performed the exercise protocol with both the synchronized and the standard markers. Further ten participants (5w/5m, 26.7 ± 2.8 years, 174.6 ± 9.8 cm, 71.9 ± 16.8 kg) performed the protocol with just the synchronized markers to collect further data about the reliability of the markers. None of the participants had musculoskeletal, cardiovascular, or neurological disorders or surgical procedures within 6 months leading up to the study. This study was approved by the local ethics committee. All participants provided written informed consent.

2) Experimental protocol: We used the Vicon built into a GRAIL System (Motek Medical B.V., Netherlands)[10], which is a setup with 10 Vero 1.3 cameras[11]. The treadmill was locked in a horizontal position to be used as a solid platform and the railings were removed to avoid obstructing the view onto the subject. To minimize the risk of a subject falling off the platform, gym mats were placed on the floor. Four Kinects were positioned 1 meter above the platform, each placed 2.5 meters from the center and arranged around the platform to capture footage of the subject from the front, back, and both sides. Each Kinect was connected to its own computer and synchronized in a daisy-chain configuration, with the synchronized markers at the end of the chain. Each PC ran the Azure Kinect recording software provided by Microsoft, which was managed by a custom Python program. This Python program could be operated by a host PC to simultaneously start all recordings with a fixed exposure value and an appropriate head start for the subordinate cameras in order for the synchronization to work properly. Each Kinect was set up to record in the *NFOV Unbinned* mode, while recording 4K RGB video and IMU data. The Vicon was calibrated according to the instructions of the manufacturer.

The participants were asked to wear short, tight-fitting pants (e.g. bicycle pants) and a tight fitting upper garment. Male participants were asked to abstain from wearing any upper garment, as this allows for every marker to be taped directly to the skin. As dark fabrics can have a negative effect on the performance of the IR sensor [22], light clothing was preferred. In preparation for the measurement, the 41 synchronized markers were placed on the body using double-sided tape. The markers for the head were secured with Velcro to a properly fitting headband. Instructors demonstrated the required movements for the functional calibration trial, which participants then practiced without recording. This practice

[10]https://www.motekmedical.com/solution/grail/
[11]https://www.vicon.com/hardware/cameras/vero/

session allowed participants to familiarize themselves with the movements and ensured that the markers remained securely in place. Following this, each subject stood within the tracking area while the markers were connected to the power banks. After confirming that all markers were registered by the Vicon, the functional trial was recorded. Then the Kinects and the Vicons simultaneously recorded 4 movements of the lumbar spine (flexion, extension, rotation, and side-movement) as well as 6 exercises (shoulder press, squat, deadlift, hip abduction, bent-over row, and lunge). Each movement was repeated 10 times. If fatigue interfered with the correct execution of an exercise, the subject was asked to stop earlier. To ensure the safe and correct execution, each exercise was first demonstrated through a video, followed by a trial run for practice. A physical therapist was present at all times to provide correction and ensure the physical well-being of the participants. The pace of each exercise was directed by a metronome. Based on recommendations from other studies [23], [24], [25], [26], we instructed participants to perform a synchronization impulse at the start of each recording. This impulse is a distinct movement that stands out from the exercise movements. Since the exercises involve relatively slow movements, participants were asked to rapidly abduct and adduct their left arm three times. After a pause, participants were asked to repeat the protocol for a second time. The first 10 participants had their synchronized markers replaced with standard ones before repeating the 10 exercises for a third time.

3) Data processing: All recorded Kinect videos are processed with the Kinect Body Tracking SDK v1.1.2 in DirectML mode on the same computer. This configuration was proposed by [17], as it yields the most reproducible results.

a) Spatial registration: The Kinects were spatially aligned with each other in two steps. First, a coarse registration using Aruco markers, followed by a fine registration with an iterative closest point (ICP) algorithm, similar to the methods suggested by [27], [28], or [29]. The Aruco markers were placed on a 3D-printed cube, which also serves as a digital twin to align the coordinate systems with. For fine registration we used a four degrees of freedom ICP by CloudCompare[12], leveraging the IMU data.

The Aruco cube was equipped with infrared LEDs, which could be registered as markers by the Vicon software. As their position in the 3D print is known, this was used to align the Vicon to the digital twin and to register the Vicon in the same coordinate system as the Kinect.

b) Temporal synchronization: The Kinects are synchronized to one another by their external synchronization system. To align the Vicon with the Kinect, the synchronization impulse was utilized. By performing a cross-correlation between the velocities of the Vicon and Kinect keypoints of the left wrist along the vertical axis, a dependable and accurate alignment was achieved. Each correlation was manually inspected.

c) Correction of the Vicon data: The markers are only active in one or two out of every four Vicon frames. During the remaining frames, the markers are inactive and therefore not recorded. Since the Nexus software's labeling operations depend on directly consecutive frames, this intermittent marker inactivity disrupts its proper functionality. Without the labeling of markers, the positions in the empty frames cannot be properly interpolated. To address this, each frame where the markers are visible (active frames) is cloned to replace the subsequent inactive frames. While this enables proper labeling, it leads to jagged movements, as the position of the markers are updated only once for every four frames. However, once the markers are correctly labeled in the active frames, the previously duplicated frames can be replaced by spline-interpolated frames. This approach ensures smooth motion and allows subsequent Nexus pipelines to operate correctly.

d) Interpolation and filtering: The positional data from the Vicon and from the Kinect was spline interpolated to exactly 30 Hz. In order to remove noise from the dataset, a low-pass filter was used on both the Kinect and the Vicon data, as suggested by [30]. For this dataset, a 5th order 5 Hz Butterworth low-pass filter was employed.

D. Statistical analysis

The objective of this work is to quantify the disturbances introduced by a marker-based system and to evaluate the effectiveness and reliability of synchronized markers during simultaneous recording of Kinect and Vicon systems.

1) Reliability of signal detection: To ensure correct registration of the synchronized markers, they must be active in at least one Vicon frame for every 2-3 frames in which they are inactive. Due to slight variations in the Kinect's frame rate, the markers may activate or deactivate very close to the Vicon's shutter. Since LEDs require time to reach their maximum brightness, they may appear dimmer in these instances. This results in only a subset of markers being recognized in certain Vicon frames. These frames are excluded from further analysis, and only frames containing more than 30 markers are considered *valid*. When two consecutive frames each contain more than 30 markers, the frame with the higher marker count is accepted as valid, while the other is discarded. A sync signal is deemed unrecognized if no valid frame is detected after six or more consecutive invalid frames.

2) Marker detection: To verify whether the synchronized markers are properly recognized by the Vicon system, we calculate the average number of markers detected per valid frame for both marker types. As different subjects tend to occlude different markers during the execution of an exercise, only the trials of those subjects who performed the exercises with both types of markers were assessed.

3) Skeleton availability: We compare the number of detection failures by the SDK on recorded data from each of the four points of view. For each frame in which a skeleton could be found, the confidence level of each of the tracked joints is compared. This analysis is conducted once on the data recorded with the synchronized markers and once on the data recorded with the standard markers.

4) Tracking accuracy: For all tracked joints, the spatial difference between the Kinect joints and the Vicon markers is calculated. However, as the Vicon markers are placed on the skin while the Kinect joints are positioned within the body,

[12]https://www.danielgm.net/cc/

This article has been accepted for publication in IEEE Sensors Journal. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2025.3531760

TABLE I

APPROXIMATION OF KINECT KEYPOINTS WITH VICON MARKERS.

Kinect Joint	Vicon Markers
Pelvis	[LASI, RASI, LPSI, RPSI]
Spine Navel	[[LASI, RASI, LPSI, RPSI], [STRN, T10]]
Spine Chest	[STRN,T10]
Neck	[CLAV, C7]
Head	[RFHD, LFHD, RBHD LBHD]
Elbow Left	LELB
Wrist Left	[LWRA, LWRB]
Hip Left	[LASI, LPSI]
Knee Left	LKNE
Ankle Left	LANK
Foot Left	LTOE

To eliminate redundancies, only the left extremities are detailed. Keypoints on the right are calculated accordingly. For each group of markers, the mean position within the bracket is calculated. Hands, Hand-tips, Thumbs, Nose, Eyes, and Ears are excluded.

a systematic offset between the two occurs. To mitigate this issue, certain joints are approximated using a single Vicon marker, whereas others are represented by the mean of a set of Vicon markers, as demonstrated in studies [5], [11], [24], [31], and others. The mapping of the markers to the joints is shown in Table I. To address the remaining offset, the mean position of each keypoint is subtracted from each frame's position data within every recording, resulting in a *zero-mean shifted* position [31]. The accuracy of the Kinect is measured by the absolute difference between the Kinect's and the Vicon's positions after adjusting for zero-mean shift.

5) Correlation: The Pearson correlation coefficient is computed for each joint across every dimension with the following defined intervals: poor ($r < 0.4$), moderate ($r = 0.4 - 0.7$), good ($r = 0.7 - 0.9$), and excellent ($r > 0.9$) [31].

IV. RESULTS

Initially, the noise associated with the standard markers is analyzed, and subsequently, the detection of the synchronized markers by the Vicon system is evaluated. Finally, the differences in joint tracking accuracy and availability are quantified.

A. Mechanical failures

During the entire study, just two mechanical failures occurred. On one occasion, an improperly soldered joint caused an intermittent connection. However, since each circuit board was prepared in duplicate, this issue was resolved without interrupting the study. The second failure occurred when a wire was pulled out of the plug connecting the arm markers to the harness on the back. This caused only a brief interruption, as the loose cable was quickly re-soldered to the board. However, this incident highlighted the need for adding strain relief on the connections in future studies. There were no problems with any of the LED markers, the Arduino, or the MOSFET modules.

B. Interference of the Vicon-system

There are two types of noise that impact the Kinect's ability to generate accurate depth and infrared images: the *passive noise* and the *active noise*.

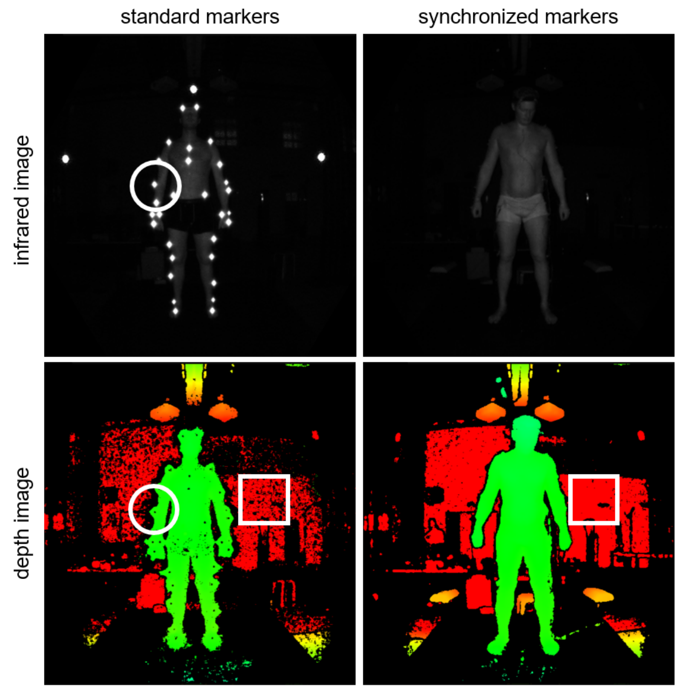


Fig. 4. Comparison of infrared (upper row) and depth image (lower row) of the Kinect when using the standard markers (left column) or the novel synchronized markers (right column). White circle: highlighting of a marker. White square: highlighting of the active noise.

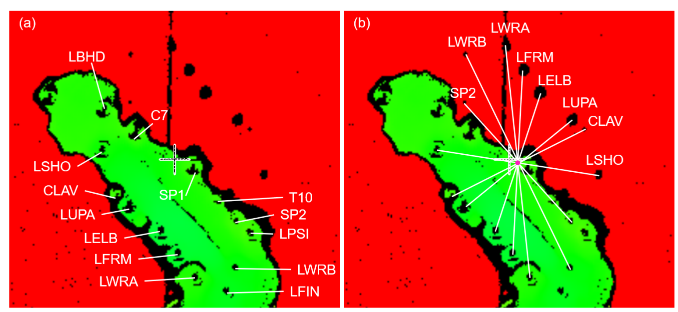


Fig. 5. Cropped depth image of a person doing a squat (side view). Image center is highlighted by a cross. a) Names of markers corresponding to artifacts (SP1 and SP2 are added to spine). b) Markers connected to their corresponding point reflection artifacts. Purple dot: center of reflection.

a) Passive noise: The markers used by the Vicon system are highly reflective and, as the ToF sensor emits infrared light, are therefore brightly visible in the IR image. In the depth image, the markers appear as black spots with no depth information. However, with the Azure Kinect, the markers also impact their surroundings in the depth image, creating a circular distortion around each marker, as shown in Fig. 4. Furthermore, the depth images contain a type of artifact that, to the best of our knowledge, has not been reported on yet. Markers close to the center do not only affect their immediate surroundings but also the mirrored point around a position close to the image center, as illustrated in Fig. 5. These *point-reflection artifacts* vanish with increasing distance to the center.

b) Active noise: Given that the shutter speed of the Azure Kinect's NFOV unbinned mode is 12.8 ms, it is guaranteed

This article has been accepted for publication in IEEE Sensors Journal. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2025.3531760

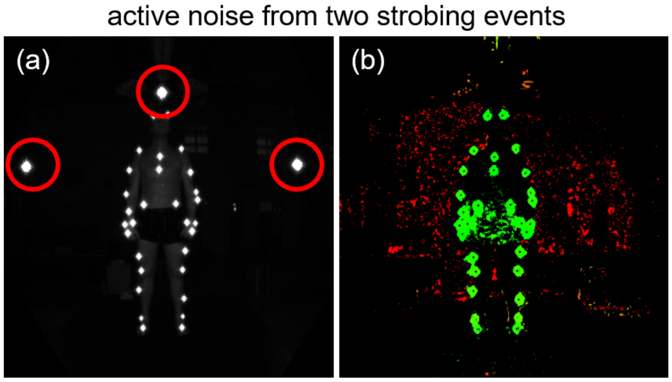


Fig. 6. Impact of active noise generated on image quality. a) Infrared image. Red circle: Vicon camera. b) Almost completely corrupted corresponding depth image

that at least one Vicon strobe event will occur during one Kinect shutter period. The infrared light emitted by the Vicon cameras impacts both the infrared image and the depth image. The infrared images show bright spots where the cameras are positioned, while the depth images display increased noise compared to those captured with synchronized markers, as shown in Fig. 4. However, due to slight shifts in the Kinect's timing, two strobe events sometimes affect a single shutter. In those cases, the white spots of the cameras appear larger in the infrared image, while the depth image is almost completely corrupted, as shown in Fig. 6.

c) Interference of simultaneous measurement on the Vicon: The simultaneous recording of the Kinect and the Vicon has no noticeable impact on the Vicon, aside from the Kinect appearing as an unlabeled marker. This problem can be resolved by adjusting the region of interest to exclude the Kinect.

C. Interference caused by synchronized markers

As shown in the right half of Fig. 4, the infrared image recorded with the synchronized markers does not show any white spots around the body. This indicates that no external infrared light source is present. As the markers are physically present, they appear as small bumps around the silhouette, particularly noticeable around the person's shoulders. However, this distortion is minor compared to the standard markers.

D. Recognition of sync-signals and markers

Out of the entire dataset of 400 recordings, every synchronization signal was detected in 397 instances without exception. In the three remaining recordings, the markers failed for 471, 675, and 543 consecutive frames. Every recording contains more than 10,000 frames. In all three instances, the entire marker set failed to illuminate. Of a total of 41 markers, the Vicon recognized on average 40.636 ($\pm$ 0.86) when using the synchronized markers, while it detected an average of 40.630 ($\pm$ 0.95) for the standard markers.

E. Availability of the skeletons

The Kinect Body Tracking SDK was utilized to estimate the skeleton in videos recorded from the front, both sides,

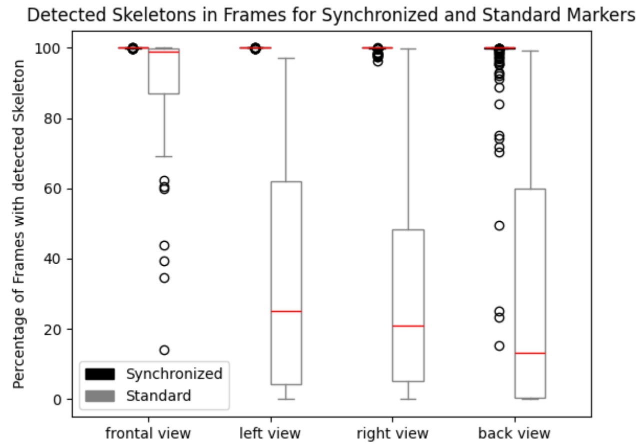


Fig. 7. Percentage of frames where a skeleton could be extracted. Comparison for the front, both sides, and the back between synchronized and standard markers.

and the back. The percentage of frames in which a body was found is shown in Fig. 7. In the frontal view, the SDK could detect a body for a median of 99.95% ($\pm$ 0.01%) of frames for the synchronized markers, while the median for the standard markers was 98.86% ($\pm$ 15.6%). When viewed from the left, the SDK could retrieve a skeleton with a median of 99.95% ($\pm$ 0.01%) for the synchronized markers, while the standard markers facilitated a detection rate of 25% ($\pm$ 31.08%). For the right side, the median for the synchronized markers was 99.95% ($\pm$ 0.49%) and the median of the standard markers was 21.04% ($\pm$ 30.65%). When viewed from behind, for the synchronized markers, many outliers were observed, particularly in the *shoulder press* and *deadlift* exercises. However, the median was over 99% ($\pm$ 10.87%), while the median of the standard markers dropped to 13.23% ($\pm$ 33.8%).

Using the standard markers, the SDK was able to consistently generate the model only when the Kinect was recording from the front. Therefore, all subsequent comparisons will be limited to the frontal views. Any frame in which the SDK failed to provide tracking information will be excluded from the comparisons.

F. Self-reported confidence

The SDK-reported confidence levels for tracking each joint are illustrated in Fig. 8. Almost every joint in the runs with the synchronized markers shows higher tracking confidence. The joints *hand*, *hand-tip*, and *thumb* were rarely tracked, and were therefore excluded from all subsequent calculations. The joints *pelvis*, *spine navel*, and *spine chest* have been tracked with perfect confidence for all runs.

G. Mean absolute error

For each tracked joint, we calculated the absolute distance between the zero-mean shifted Vicon markers and the zero-mean shifted Kinect joints. The results are shown in Fig. 9 on

[14] https://learn.microsoft.com/en-us/previous-versions/azure/kinect-dk/body-joints

This article has been accepted for publication in IEEE Sensors Journal. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2025.3531760

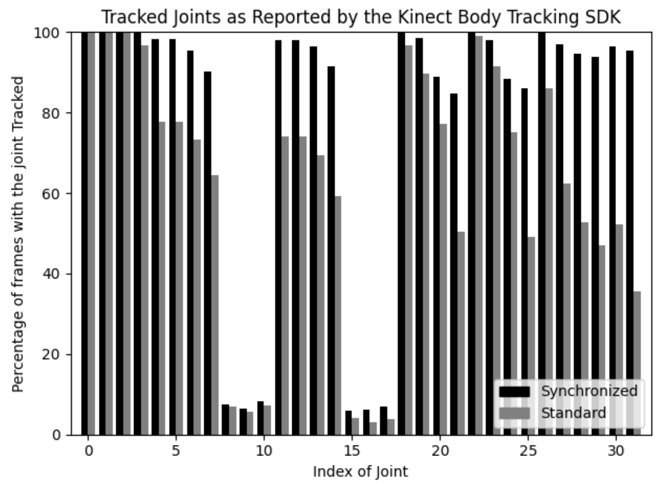


Fig. 8. Percentage of frames in which each joint is tracked. Indices of joints are specified by the body tracking SDK[14]. Small columns: confidence levels for the hand, hand-tip, and thumb.

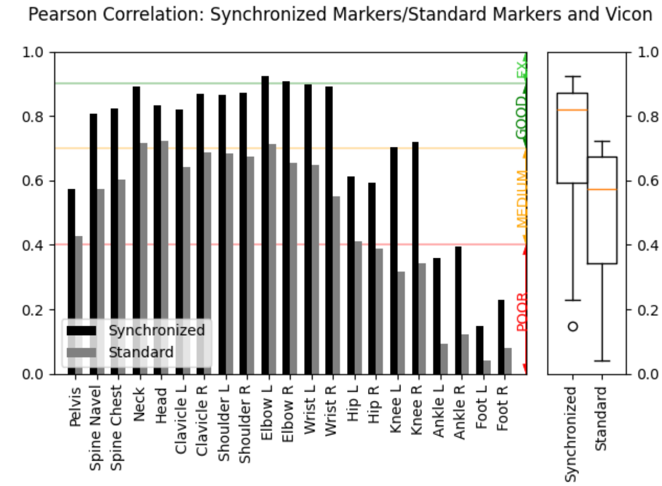


Fig. 10. Pearson correlation coefficient for each joint. Thresholds poor: $r < 0.4$, moderate: $r = 0.4 - 0.7$, good: $r = 0.7 - 0.9$, and excellent $r > 0.9$. Right: Boxplot of all means of every joint.

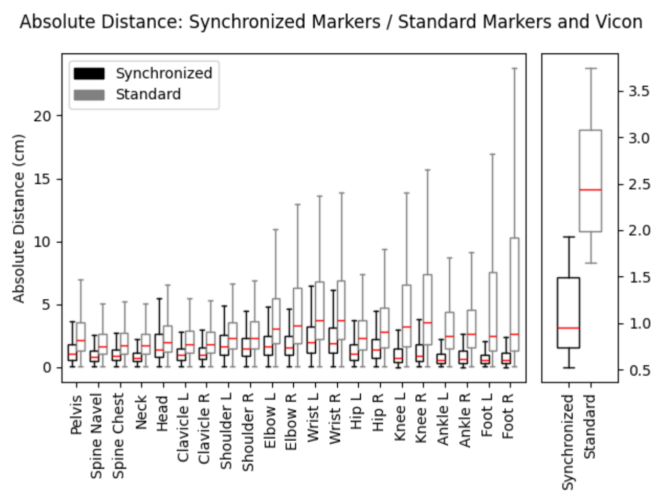


Fig. 9. Mean absolute error of every zero-mean shifted joint in every recording for synchronized and standard markers. Right: Boxplot of all medians. Outliers were removed to improve visibility.

the left. For the synchronized markers, the median errors range from 0.5 cm to 1.9 cm, whereas for the standard markers, the errors range from 1.7 cm to 3.7 cm. The standard deviations for the synchronized markers range from 0.7 cm to 3.3 cm with a mean of 1.4 cm, while they range for the standard markers from 2.0 cm to 25.3 cm with a mean of 6.1 cm. For each joint, the error was considerably lower for the synchronized markers compared to the error of the standard markers. The distances between the two runs were greater in the extremities than in the torso.

The boxplots of all medians are shown in Fig. 9 on the right. With a median absolute error of 0.96 cm, the synchronized markers are achieving considerably more accurate results than the standard markers with a median of 2.4 cm. The Wilcoxon rank-sum test of all medians results in a p-value of $1.27e^{-7}$, indicating a significantly greater accuracy for the synchronized markers.

H. Pearson correlation

For each spatial dimension and each recording, the Pearson correlation coefficient was calculated separately. The mean correlation across all recordings and dimensions is visualized for each joint in Fig. 10. When using the synchronized markers, the Kinect Body Tracking SDK generates a skeleton that achieves an *excellent* rating for 2 joints, a *good* rating for 12 joints, a *medium* rating for 3 joints, and a *poor* rating for 4 joints. When using the standard markers, the SDK achieves no *excellent* rating, a *good* rating for 3 joints, a *medium* rating for 12 joints, and a *poor* rating for 6 joints. For every single joint, the correlation is higher for the synchronized markers. The mean Pearson correlation over all joints is 0.70 for the synchronized and 0.48 for the standard markers. The Wilcoxon rank-sum test across all joint results in a p-value of 0.0018, indicating a significantly higher correlation for the synchronized markers.

V. Discussion

This study found that the simultaneous recording of the Vicon and the Azure Kinect has a severe impact on the tracking accuracy of the Kinect. The standard markers appear as bright white spots in the infrared image. On the depth images, they appear as black points surrounded by a circle with a depth similar to the silhouette of the person. These findings are similar to those reported in [8] for the Kinect v2. Furthermore, we found that markers close to the center can cause point-reflection artifacts.

Our findings, illustrated in Fig. 4, show that the synchronized markers completely eliminate the artifacts in the infrared image. In the depth image, the typical artifacts caused by passive noise do not appear. This shows that the synchronized markers are a viable option to reduce the detrimental effects of the standard markers.

The strobing infrared light introduces noise in the depth image and can corrupt it entirely if two strobe events occur within a single Kinect shutter cycle, as shown in Fig. 6. It is therefore advisable to reduce the Vicon's frame rate to a maximum of 60 Hz when using the standard markers with the Kinect, eliminating the possibility of two Vicon strobes appearing in a single Kinect shutter. As the synchronized markers do not require the strobing infrared light of the Vicon, they eliminate the active noise entirely.

When investigating the recognition of synchronization signals, we found that they could be detected very reliably. There were only three failures. Given the extensive duration of the outages, it is hypothesized that the problem stemmed from an improperly closed connection, either between the Kinect and the Arduino or between the power supply and the markers. The detection rate of the synchronized markers is very similar to that of the standard markers, suggesting that the standard markers can effectively be replaced by the synchronized ones, as indicated by [9] for active markers.

A comparison of skeleton tracking capabilities reveals that the SDK produces more reliable results when synchronized markers are utilized, regardless of the point of view of the Kinect. When using synchronized markers, the SDK successfully generates a skeleton in nearly every frame, with occasional challenges only occurring for the rear-facing Kinect. With standard markers, the front-facing Kinect achieves a high median detection rate, but it exhibits larger standard deviation. When recording the person from the side or back, a skeleton could be generated for only a fraction of the frames. These findings show that the standard markers are detrimental to the reliability of the SDK, consistent with the results reported in [10]. They recorded the subjects from different angles and found that the SDK failed to consistently deliver the skeleton model, even when employing smaller markers.

For almost every joint, the tracking confidence is higher for the synchronized markers. When comparing only those joints that are considered tracked, the mean positional error of the synchronized markers is considerably lower than for the standard markers. The Wilcoxon rank-sum test of all medians confirms a significantly greater accuracy for the synchronized markers.

The Pearson correlation is significantly higher for every joint of the synchronized markers as well. Additionally, the data suggest that lower body tracking is considerably less accurate than upper body tracking. However, the correlation is tightly linked to the extent of movement exerted by the specified joint. Since most exercises involved minimal lower body movement, only minor involuntary movements were captured. In cases of minimal movement, the random noise in pose detection constitutes a larger proportion of the total recorded movement, leading to a lower correlation. It should therefore be interpreted as a relative value.

The synchronized marker set proved to be reliable. As each of the 20 participants wore the markers for at least two hours, it is reasonable to assume that this design (with additional strain relief for the connections) appears sufficiently robust to be used for future studies.

A. Advantages of the synchronized markers

Beyond eliminating interference, the synchronized markers offer additional advantages over standard ones. As the Vicon does not emit infrared light itself, each camera is not susceptible to reflective surfaces or to other cameras. This means participants can wear glasses if needed. Furthermore, the Vicon cameras do not have to be masked. Standard markers require the practitioner to wear gloves, as fat and moisture from the hand can dampen the reflective properties of the coating. This is not the case with the diffusers, which makes this process easier. Additionally, once the Kinects stop recording, the markers turn off for a second. Using this information, the temporal synchronization between the Kinect and the Vicon becomes trivial, eliminating the need for a synchronization impulse. As only cheap off-the-shelf parts were used, the cost of the entire marker set, the controller, and the power supply does not exceed $150.

B. Disadvantages of the synchronized markers

The greatest disadvantage of the synchronized markers is the time required to design and build them. To accelerate the process for others, the assembly instructions and 3D designs for this marker set are publicly available at Zenodo. Applying the marker set to a subject is more time-consuming because, in addition to the markers, the circuit boards and excess wires also need to be fixed to the person. The cable connecting the MOSFETs to the harness, as well as the harness itself, could be considered a tripping hazard; however, no incidents of subjects tripping were reported. The process of correcting the invalid Vicon frames is time-consuming and could be eliminated by triggering the Vicon with the Kinect's sync signal, as suggested by [12].

C. Limitations

Participants were instructed to perform the exercises in a slow and controlled manner. Apart from the synchronization impulse, the exercises did not include any rapid movement. The proposed approach may encounter challenges during fast-paced activities such as running, jumping, or other quick body movements, as these could place additional strain on the wires, increase the risk of tripping, or cause the markers to detach from the skin.

The distance between the Vicon cameras and the person did not exceed 4 meters due to the compact setup of the GRAIL system. A wide tracking space with further distances to each camera might require stronger LEDs to be reliably tracked.

Additionally, while the synchronized markers did not emit infrared light that interferes with the Kinect, the physical presence of the markers, power distributor, and wires introduced minor alterations to the silhouette. Such altered conditions could potentially impact the SDK's pose-tracking accuracy.

VI. Conclusion

The results of the study show that the proposed novel set of custom-made synchronized markers is an effective solution for the simultaneous measurement of the Kinect and the Vicon without interference. The noise introduced by the standard markers and the Vicon greatly impacts the SDK's ability to detect a skeleton in the recordings, especially for the side and rear views. While the frontal view with standard markers allows the SDK to generate a skeleton in most frames, the accuracy of each joint is significantly lower compared to the same recording when using synchronized markers.

The synchronized markers can be reliably tracked by the Vicon. We therefore believe we have found a solution for the interference problem between the Kinect and the reference system. We invite others to implement it in their studies as well.

REFERENCES

[1] G. Du and P. Zhang, "Markerless human–robot interface for dual robot manipulators using Kinect sensor", *Robot. Comput.-Integr. Manuf.*, vol 30, no. 2, pp. 150-159, doi: 10.1016/j.rcim.2013.09.003.

[2] W. Chen, Z. Jiang, H. Guo and X. Ni, "Fall detection based on key points of human-skeleton using openpose", *Symmetry*, vol. 12 no.5: 744, May 2020, doi: 10.3390/sym12050744.

[3] X. Suo, W. Tang, and Z. Li, "Motion Capture Technology in Sports Scenarios: A Survey" *Sensors (Basel)*, vol. 24, no. 9: 2947, May 2024, doi: https://doi.org/10.3390/s24092947.

[4] R. Okada, N. Kondoh and B. Stenger, (2007, May). "A Video Motion Capture System for Interactive Games", *MVA2007*, pp. 186-189, May 2007.

[5] M. Antico, N. Balletti, G. Laudato, A. Lazich, M. Notarantonio, R. Oliveto, et al., "Postural control assessment via Microsoft Azure Kinect DK: An evaluation study," *Comput. Methods Programs Biomed.*, vol. 209, p. 106324, Sep. 2021., doi: 10.1016/j.cmpb.2021.106324.

[6] J. A. Albert, V. Owolabi, A. Gebel, C. M. Brahms, U. Granacher, and B. Arnrich, "Evaluation of the Pose Tracking Performance of the Azure Kinect and Kinect v2 for Gait Analysis in Comparison with a Gold Standard: A Pilot Study" *Sensors (Basel)*, vol. 20, no. 18, p. 5104, Sep. 8 2020., doi: 10.3390/s20185104.

[7] A. Cattaneo, M. Fragasso, M. Magni, I. Mostachetti, E. Perri, and A. Vitali, "Assessing the Accuracy of the Azure Kinect for Telerehabilitation After Breast Cancer Surgery", dHealth, pp. 83-88, 2023, doi: 10.3233/SHTI230017.

[8] M. Naeemabadi, B. Dinesen, O. K. Andersen and J. Hansen, "Influence of a Marker-Based Motion Capture System on the Performance of Microsoft Kinect v2 Skeleton Algorithm," *IEEE Sensors Journal*, vol. 19, no. 1, pp. 171-179, 1 Jan.1, 2019, doi: 10.1109/JSEN.2018.2876624.

[9] M. 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, Sep. 14 2018., doi: 10.1371/journal.pone.0204052.

[10] L. F. Yeung, Z. Yang, K. C. Cheng, D. Du, and R. K. Tong, "Effects of camera viewing angles on tracking kinematic gait patterns using Azure Kinect, Kinect v2 and Orbbec Astra Pro v2," *Gait and Posture*, vol. 87, pp. 19–26, Jun. 2021., doi: 10.1016/j.gaitpost.2021.04.005.

[11] J. Bertram, T. Krüger, H. M. Röhling, A. Jelusic, S. Mansow-Model, R. Schniepp, et al., "Accuracy and repeatability of the Microsoft Azure Kinect for clinical measurement of motor function," *PLoS One*, vol. 18, no. 1, Jan. 26 2023., doi: 10.1371/journal.pone.0279697.

[12] J. Colombel, D. Daney, V. Bonnet, and F. Charpillet, "Markerless 3D Human Pose Tracking in the Wild with Fusion of Multiple Depth Cameras: Comparative Experimental Study with Kinect 2 and 3", *Activity and Behavior Computing*, 2021, pp. 119–134, doi: 10.1007/978-981-15-8944-7_8.

[13] E. Ivorra, M. Ortega, and M. Alcaniz, "AZURE KINECT BODY TRACKING UNDER REVIEW FOR THE SPECIFIC CASE OF UPPER LIMB EXERCISES", *MM Sci. J*, vol. 2021, no. 2, pp. 4333–4341, 2021., doi: 10.17973/MMSJ.2021_6_2021012.

[14] U. Özsoy, Y. Yıldırım, S. Karaşin, R. Şekerci, and L. B. Süzen, "Reliability and agreement of Azure Kinect and Kinect v2 depth sensors in the shoulder joint range of motion estimation", *J. Shoulder Elb. Surg.*, 2022, vol. 31, no. 10, pp. 2049-2056, Oct. 2022, doi: 10.1016/j.jse.2022.04.007.

[15] S. Jo, S. Song, J. Kim, and C. Song, "Agreement between Azure Kinect and Marker-Based Motion Analysis during Functional Movements: A Feasibility Study", *Sensors (Basel)*, vol. 22, no. 24, p. 9819, Dec. 14 2022., doi: 10.3390/s22249819.

[16] M. Tölgyessy, M. Dekan, and L. Chovanec, "Skeleton Tracking Accuracy and Precision Evaluation of Kinect V1, Kinect V2, and the Azure Kinect," *Appl. Sci. (Basel)*, vol. 11, no. 12, p. 5756, 2021., doi: 10.3390/app11125756.

[17] L. Büker, V. Quinten, M. Hackbarth, S. Hellmers, R. Diekmann, and A. Hein, "How the Processing Mode Influences Azure Kinect Body Tracking Results,"*Sensors (Basel)*, vol. 23, no. 2, p. 878, Jan. 12 2023., doi: 10.3390/s23020878.

[18] K. Yoshimoto and M. Shinya, "Use of the Azure Kinect to measure foot clearance during obstacle crossing: A validation study," *PLoS One*, vol. 17, no. 3, Mar. 11 2022., doi: 10.1371/journal.pone.0265215.

[19] T. M. Guess, R. Bliss, J. B. Hall, and A. M. Kiselica, "Comparison of Azure Kinect overground gait spatiotemporal parameters to marker based optical motion capture," *Gait and Posture*, vol. 96, pp. 130–136, Jul. 2022., doi: 10.1016/j.gaitpost.2022.05.021.

[20] J. Thomas, J. B. Hall, R. Bliss, and T. M. Guess, "Comparison of Azure Kinect and optical retroreflective motion capture for kinematic and spatiotemporal evaluation of the sit-to-stand test," *Gait and Posture*, vol. 94, pp. 153–159, May 2022., doi: 10.1016/j.gaitpost.2022.03.011.

[21] N. Hepke, M. Scherer, B. Weyers, S. Müller, and J. Lohscheller, "Minimizing Validation Artifacts in Azure Kinect Caused by Marker Based Systems using Synchronized Markers: Instruction Manual", *Zenodo*, 2024, doi: https://doi.org/10.5281/zenodo.13789699

[22] M. Tölgyessy, M. Dekan, Ľ. Chovanec, and P. Hubinský, "Evaluation of the Azure Kinect and Its Comparison to Kinect V1 and Kinect V2," *Sensors (Basel)*, vol. 21, no. 2, p. 413, Jan. 8 2021, doi: 10.3390/s21020413.

[23] D. J. Geerse, B. H. Coolen, and M. Roerdink, "Kinematic Validation of a Multi-Kinect v2 Instrumented 10-Meter Walkway for Quantitative Gait Assessments," *PLoS One*, vol. 10, no. 10, Oct. 13 2015., doi: 10.1371/journal.pone.0139913.

[24] M. Ma, R. Proffitt, and M. Skubic, "Validation of a Kinect V2 based rehabilitation game," *PLoS One*, vol. 13, no. 8, Aug. 24 2018, doi: 10.1371/journal.pone.0202338.

[25] 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., doi: 10.1016/j.bspc.2017.06.006.

[26] X. Xu, R. W. McGorry, L. S. Chou, J. H. Lin, and C. C. Chang, "Accuracy of the Microsoft Kinect for measuring gait parameters during treadmill walking," *Gait and Posture*, vol. 42, no. 2, pp. 145–151, Jul. 2015., doi: 10.1016/j.gaitpost.2015.05.002.

[27] B. Müller, W. Ilg, M. A. Giese, and N. Ludolph, "Improved Kinect sensor based motion capturing system for gait assessment", *bioRxiv*, 2017, doi: 10.1101/098863.

[28] S. Bu and S. Lee, "Easy to Calibrate: Marker-Less Calibration of Multiview Azure Kinect," *Comput. Model. Eng. Sci.*, vol. 136, no. 3, pp. 3083–3096, 2023., doi: 10.32604/cmes.2023.024460.

[29] L. Romeo, R. Marani, A. G. Perri, and T. D'Orazio, "Microsoft Azure Kinect Calibration for Three-Dimensional Dense Point Clouds and Reliable Skeletons," *Sensors (Basel)*, vol. 22, no. 13, p. 4986, Jul. 1 2022., doi: 10.3390/s22134986.

[30] R. A. Clark, Y. H. Pua, K. Fortin, C. Ritchie, K. E. Webster, L. Denehy, et al., "Validity of the Microsoft Kinect for assessment of postural control," *Gait and Posture*, vol. 36, no. 3, pp. 372–377, Jul. 2012., doi: 10.1016/j.gaitpost.2012.03.033

[31] K. Otte, B. Kayser, S. Mansow-Model, J. Verrel, F. Paul, A. U. Brandt, et al., "Accuracy and Reliability of the Kinect Version 2 for Clinical Measurement of Motor Function," *PLoS One*, vol. 11, no. 11, Nov. 18 2016., doi: 10.1371/journal.pone.0166532.