Camera-Based Tracking and Evaluation of the Performance of a Fitness Exercise

Linda Büker[1][a], Dennis Bussenius[1], Eva Schobert[2], Andreas Hein[1][b] and Sandra Hellmers[1][c]

[1]*Assistance Systems and Medical Device Technology, Carl von Ossietzky University Oldenburg, Oldenburg, Germany*

[2]*Herodikos GmbH, Hans-Schütte-Str. 20, 26316 Varel, Germany*

Keywords: Azure Kinect DK, Health, Human Pose Analysis, Exercise Analysis.

Abstract: Back pain is a significant condition worldwide that is becoming more common over the years. Although individually adapted exercise therapy would be very successful, the problem is - due to lack of capacity - it is rarely prescribed by physicians. We analyse if it is feasible to automatically track and evaluate the performance of an exercise used in a fitness check to support physicians in adapting exercise therapy and thereby providing that kind of treatment to more patients. A depth camera with body tracking is used to detect the pose of the subjects. We have developed a system that evaluates the execution of an exercise in terms of correct performance based on the recognized joint positions. The feasibility study conducted shows, that it is important to avoid as many occlusions of any kind as possible to get the best achievable body tracking. Then, however, the evaluation of the performance of the tested finger-floor-distance exercise seems feasible.

1 INTRODUCTION

Lower back pain is a significant epidemiological burden worldwide, especially in countries with a high socio-demographic index, with a remarkable increase in the last years (Mattiuzzi et al., 2020). The number of hospital cases with a primary diagnosis of "back pain" among the 2.5 million members of the German health insurance "DAK-Gesundheit" has increased by 80% from 2007 to 2016 (Marschall et al., 2018). While the 2003/2006 German Back Pain Study found that up to 85% of the population experience back pain at least once in their lifetime (Schmidt et al., 2007), 61.3% of the German population in 2021 reported having suffered from it in the past 12 months (von der Lippe et al., 2021). Therefore back pain is one of the most common physical complaints (Saß et al., 2015).

Long-term success with back pain is achieved through exercise therapy (Bredow et al., 2016), for which an individual adaptation is mandatory, in view of the patients needs (Hoffmann et al., 2010). Thereby, information regarding the patient's fitness, for instance, is required for individual adaptation (Barker and Eickmeyer, 2020) and can be assessed by physicians in a fitness check. However, physicians often do not have sufficient time with every patient (Irving et al., 2017), which means that a fitness check can not be conducted and thus exercise therapy is rarely integrated into the patient's treatment plan. As a result, in 2017, only less than one third of American population with back pain got exercise therapy as treatment (statista, 2022).

To open exercise therapy to a large number of patients, a technology-based approach could be used to automate the tracking and evaluation of the aforementioned fitness check. Hereby the physician would be supported in terms of time, which enables more frequent application of these checks. First, however, the feasibility of such an approach needs to be tested. Therefore, we developed a system to automatically track and evaluate a fitness exercise, the finger-floor-distance exercise, and tested it in a preliminary study.

2 STATE OF THE ART

To the best of our knowledge, there is only one study evaluating the finger-floor-distance, among other exercises, automatically. For this, a marker-based optical motion capture system was used (Garrido-Castro et al., 2012). In general, to evaluate an exercise, the movement of the person has to be tracked first.

[a] https://orcid.org/0000-0002-6129-0940
[b] https://orcid.org/0000-0001-8846-2282
[c] https://orcid.org/0000-0002-1686-6752

Büker, L., Bussenius, D., Schobert, E., Hein, A. and Hellmers, S.
Camera-Based Tracking and Evaluation of the Performance of a Fitness Exercise.
DOI: 10.5220/0011755700003414
In *Proceedings of the 16th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2023) - Volume 5: HEALTHINF*, pages 489-496
ISBN: 978-989-758-631-6; ISSN: 2184-4305

This can be done via body-worn systems such as motion capture suits or marker-based camera systems. However, these systems have the disadvantage that they are usually time-consuming and require expertise (Carse et al., 2013). In addition, the systems can restrict the person's freedom of movement, thus hindering the execution of a fitness check. Therefore, the use of external marker-less systems, such as body tracking systems using camera images, is reasonable. There are multiple approaches to body tracking systems via RGB-images (Cao et al., 2019; Bulat et al., 2020; Liu et al., 2021). Although these systems achieve very accurate results, they only track people in two dimensions, which means that depth information cannot be represented well. Therefore, there are also some approaches that perform body tracking in three dimensions (Ye et al., 2011; Chun et al., 2018; Büker et al., 2021). One of these approaches was developed by Microsoft: the Azure Kinect Body Tracking SDK (Microsoft Inc., 2020), which can track multiple people (with 32 joint points each) in real time on the depth images of the Azure Kinect DK camera. Previous analyses have shown the resulting accuracy of the SDK, in comparison to the Vicon system (Vicon Motion Systems, Oxford, UK) as gold standard: depending on the joint considered, the mean euclidean distance between the two systems is between 10 mm and 60 mm (Albert et al., 2020); the root mean square error (RMSE) of the joint angles is between 7.2° and 32.3° for the lower extremities (Ma et al., 2020).

For the considered exercise, not only the tracking of the person is relevant, but also the relation of finger to floor plane. There are multiple approaches using depth images for fall detection by calculating the distance of the persons head and/or persons centroid to the floor. The required plane of the floor is either determined by a predefined area/points (Yang et al., 2015), or planes are detected in the depth image by using a RANSAC-based approach (Diraco et al., 2010).

After the person's pose and the floor are detected, it must be evaluated whether the exercise is performed correctly. A classification model can be used to detect, for example, whether a ski jumper on a video performs a bad pose while jumping (Wang et al., 2019). Several approaches compare body angles between the user and a professional who has previously performed the exercise once. Here, the quality of execution is either solely classified (Alabbasi et al., 2015) or, in addition, a continuous color coding of the individual joint angles is visually displayed (Thar et al., 2019). A comparison of joint angles between several persons can also be used to compare multiple dancers to each other for synchronization by continuously color coding the considered parts of displayed skeletons (Zhou et al., 2019). Another possibility is the use of rules, that state how various body angles, distances, or positions should be for a correct execution and classify each rule into satisfied or not satisfied (Conner and Poor, 2016; Rector et al., 2013; Chen and Yang, 2018). The result is either output visually or auditory, giving correction suggestions in case the rules are not fulfilled.

3 MATERIALS AND METHODS

In the following, the considered exercise, study design, used sensor system, as well as the used rules and analysis are described.

3.1 Exercise

The finger-floor-distance exercise has excellent metric properties for patients with lower back pain (Perret et al., 2001) and is used in multiple back pain related studies (Gurcay et al., 2009; Olaogun et al., 2004). Therefore, this paper analyzes the feasibility of automatic exercise evaluation using the finger-floor-distance exercise as an example.

For the execution of the finger-floor-distance exercise, the knees, arms and fingers must be fully extended. The feet should be close together. The subject then bends forward and tries to touch the floor with the fingertips as close as possible. The distance of the fingertips to the floor is measured for evaluation. (Perret et al., 2001)

Figure 1 shows the performance of the finger-floor-distance exercise.

Figure 1: Finger-Floor-Distance exercise.

3.2 Study Design

To analyze the feasibility of a camera-based automatic evaluation of the finger-floor-distance exercise, we conducted a study with 10 subjects between the ages of 23 and 30 (4 women, 6 men) with a height between 168 cm and 191 cm. The subjects performed the exercise finger-floor-distance twice with different finger to floor distances and/or different knee angles. This ensured that a wide variety of positions were tested. The subjects were filmed by two depth cameras. The knee angle and the distance between fingertips, as well as hands, and floor were measured manually with a goniometer and a tape measure as ground truth. The test procedures were approved by the local ethics committee (ethical vote: Carl von Ossietzky University Oldenburg, Drs.EK/2021/067) and conducted in accordance with the Declaration of Helsinki.

3.3 Applied Sensor System

In order to evaluate the finger-floor-distance exercise, the subject is recorded during the exercise performance. Two Azure Kinect DK RGB-D cameras are used for this purpose to generate twice as much data. The Kinect measures the depth with a time-of-flight camera. It has a typical systematic error of $<$11 mm + 0.1% of the objects distance and a random error with a standard deviation of $\leq$17 mm (Microsoft Inc., 2021). Both cameras are positioned at an angle of about 45° to the persons front (left and right respectively) in order to try to avoid as much occlusion of the subjects legs by the subjects arms (see Figure 2). The cameras are started with the following configuration parameters:

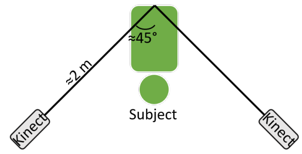
- color resolution: 1536p (2048 pixel x 1536 pixel),
- depth mode: narrow field of view binned,
- frame rate: 30 frames per second,
- inertial measurement unit (IMU) enabled.

Both cameras are evaluated individually. However, they are synchronized with an aux cable to ensure that they do not interfere with each other. The cameras are connected to the same computer (Windows 10 operating system) and can be started via console call. According to Tölgyessy et al. (Tölgyessy et al., 2021), the cameras were warmed up for one hour before the actual recording.

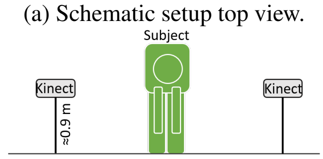
To track the subject's pose, the Azure Kinect Body Tracking SDK is used. The Azure Kinect with the body tracking SDK has the additional advantage that it also works in real-time and thus a direct evaluation of the exercise is possible. Body Tracking is running with CUDA as its processing mode.

The *floor_detector_sample* from Microsofts Azure Kinect Samples[1] is used to detect the floor plane, which is required for calculating the distance from the fingers and hands to the floor. The floor detector needs IMU-data of the camera as well as the depth image to generate a floor plane whose equation is provided in normal form (see Section 3.5).

[1]https://github.com/microsoft/Azure-Kinect-Samples



(a) Schematic setup top view.



(b) Schematic setup frontal view.

Figure 2: Overview of the experimental setup.

3.4 Rules

We decided to use rules instead of a machine learning approach to detect whether the finger-floor-distance exercise was performed correctly, and if so, how well. The use of rules has the advantage that we don't need a large amount of data, as is usually needed for machine learning approaches. Additionally, we already know the exact logic to decide whether a performance is correct or not, which makes a rule-based approach useful and explainable.

To evaluate whether the execution of the finger-floor-distance exercise is correct, the compliance of the following rules must be observed.

1. **Feet Distance:** The feet of the subject need to be close together. Therefore, the system checks whether the distance of the subjects feet is less than the distance of the subjects hips. If the feet are more than hip-width apart, the exercise is classified as incorrect.
2. **Knee Angle:** The knees need to be fully extended as bending the knees give the subject an advantage and leads to an incorrect performance. The knees are fully extended when the knee angle (angle between lower leg and thigh) is 180°. Since not every person can extend their knees to 180°, it is checked whether the knee angle is greater than or equal to 170°.

During the exercise, the arms and fingers must also be fully extended. Since not extending the arms or fingers would lead to a worse result, this is not tested at this point.

After evaluating whether the exercise was performed correctly, it is necessary to check how well the exercise is executed. For this purpose, the following is analyzed:

3. **Finger Floor Distance:** The closest distance of the fingertip (and hand) position to the plane defining the floor, needs to be calculated. If rule 1 and 2 are fulfilled, this distance is the result of the finger-floor-distance exercise.

3.5 Analysis

To analyze the performance of the finger-floor-distance exercise, the previous described three rules need to be checked for every frame of the video.

For rule 1, the $correct\_feet\_distance$ at timestamp t need to be measured as following:

$$correct\_feet\_distance_t = d\_feet_t < d\_hips_t, \quad (1)$$

where d_feet and d_hip are calculated as

$$d\_\{feet, hips\}_t(p_t, q_t) = \sqrt{(p_{t,x} - q_{t,x})^2 + (p_{t,y} - q_{t,y})^2 + (p_{t,z} - q_{t,z})^2}, \quad (2)$$

where

$p_{t,\{x/y/z\}}$ = position of the left ankle/hip at frame t for axis $x/y/z$ respectively,

$q_{t,\{x/y/z\}}$ = position of the right ankle/hip at frame t for axis $x/y/z$ respectively.

For the $correct\_knee\_angle$ at time t (rule 2), the knee angle α needs to be greater or equal than 170°:

$$correct\_knee\_angle_t = \alpha_t \geq 170^\circ \quad (3)$$

with

$$\alpha_t = 180 - (\frac{180}{\pi} \cdot \arccos(\vec{AB}_{norm,x} \cdot \vec{BC}_{norm,x} + \vec{AB}_{norm,y} \cdot \vec{BC}_{norm,y} + \vec{AB}_{norm,z} \cdot \vec{BC}_{norm,z})), \quad (4)$$

where

$$\vec{AB}_{norm,x} = \frac{\vec{AB}_x}{\sqrt{\vec{AB}_x^2 + \vec{AB}_y^2 + \vec{AB}_z^2}} \quad (5)$$

and $\vec{AB}_{norm,\{y,z\}}$, as well as $\vec{BC}_{norm,\{x,y,z\}}$ respectively. Here $\vec{AB}$ is the vector from the hip position to the knee position at frame t, while $\vec{BC}$ is the vector from the knee position to the ankle position at frame t. The angle for the right and left half of the body is calculated in each frame.

For rule 3, the distance between a point p (position of left and right fingertip and hand respectively) and the floor plane f needs to be calculated for every frame t. The floor plane is given in normal form and therefore consists of a normal vector N and a point on the plane, the origin, O, as well as the constant C:

$$C = (N_x \cdot O_x + N_y \cdot O_y + N_z \cdot O_z) \cdot -1 \quad (6)$$

The distance d_f then is calculated as:

$$d\_f_t(p_t, f_t) = \frac{|(p_{t,x} \cdot N_{t,x} + p_{t,y} \cdot N_{t,y} + p_{t,z} \cdot N_{t,z} + C_t)|}{\sqrt{N_{t,x}^2 + N_{t,y}^2 + N_{t,z}^2}} \quad (7)$$

4 RESULTS

This section presents striking aspects found while conducting the study, as well as a comparison between automatic evaluation and ground truth.

4.1 Striking Aspects

While analyzing the results, we noticed that some results are considerably worse than others. When visualizing these results as point clouds together with the detected skeleton using the *simple_3d_viewer* from Microsofts Azure Kinect Samples[1], it has been observed that the body tracking has detected the skeleton incorrectly in these cases. This occurred mainly when the subjects' hair was worn open. An example frame is shown in Figure 3. It can be seen that, on the one hand, the confidence with which a joint position was detected is low for the posterior arm and leg. This can be identified from the only slight coloring of these joints and bones. On the other hand, it can be observed that the skeleton was not detected at the exact position, but was partially located outside the subject's body (also the posterior arm and leg).

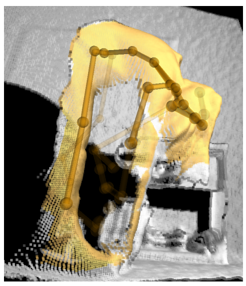


Figure 3: Incorrect body tracking at arms, presumably due to occlusion by openly worn hair. The point cloud was rotated for a better view of the incorrect body tracking.

Upon further observation, it was noticed that some of the subjects who were able to touch the floor with their hands leaned slightly forward with their hands, whereby their fingers could no longer be seen in the image frame. This also seems to negatively influence the body tracking and thus leads to poorer results.

Figure 4 shows how the body tracking does not recognize the position of the hands and fingers, and thus has to estimate them.

In the results of the analysis, it was also observed that the knee angle sometimes varies greatly within a recording (see Figure 5). On closer visual inspection, it is noticeable that the legs are in some cases very poorly recognized and "float" in the air. Figure 6 shows the subject from Figure 5 with floating legs for illustration.

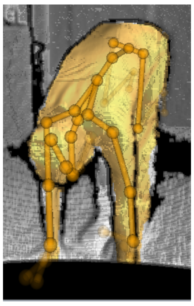


Figure 4: Incorrect body tracking at hands, presumably due to the fact that the hands are outside the frame. The positions of the hand on the left side of the frame is assumed to be outside the point cloud by the body tracking.

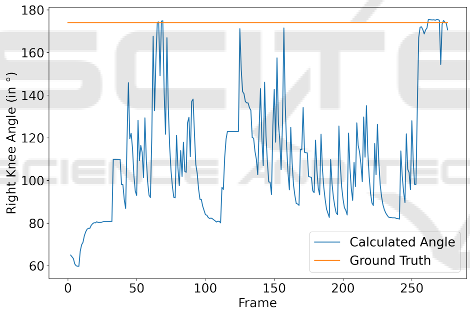


Figure 5: Great variations in knee angle over time.

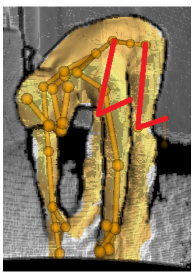


Figure 6: Incorrect body tracking at legs. The skeleton legs are floating in the air, the subject's legs are on the ground.

When calculating the distance from joint positions to the ground, the result depends not only on body tracking, but also on the quality of floor plane detection, since the floor is detected anew in every frame. Therefore, we compared the floor planes of each video with one another in terms of position and orientation. Accordingly, the difference between the closest distance of the camera to floor plane of all frames and the angles between the normal vector of all frames is calculated. For all videos, the difference in position is between 0.0 cm and 4.35 cm (mean: 0.64 cm). The orientation differs between 0.0° and 1.13° with a mean of 0.244°. This suggests that the detection of the floor may have an impact on the results, and therefore must be taken into consideration.

Due to the floating legs and the differences in floor detection, the knee angle is only calculated in the following if the feet are a maximum of 5 cm from the ground. This ensures that frames with strongly incorrect detection of the legs are disregarded.

4.2 Comparison with Ground Truth

In the subsequent analysis, the smallest and the mean knee angle is calculated for all considered frames of each video. Concurrently the smallest distance between fingertip, as well as hand, and floor is determined for every video.

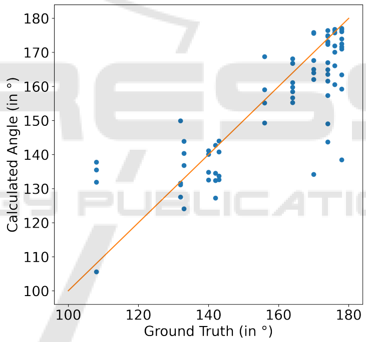


Figure 7: Mean of the calculated knee angle vs. ground truth. Visualized are all 80 data points (10 subjects with two executions, two cameras and left and right side).

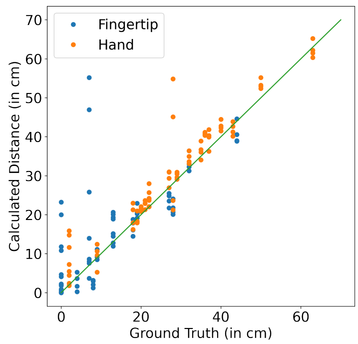


Figure 8: Minimal calculated distance of fingertip and hand to floor vs. ground truth. Visualized are all 80 data points (10 subjects with two executions, two cameras and left and right side) each.

The distribution between the mean of the calculated knee angle per video and the ground truth, as well as the distribution between the minimal calculated distance of the fingertips and hands to the floor and the ground truth are shown in Figures 7 and 8. One can see that although many calculated angles and distances are very close to the ground truth, some values show very strong differences. The difference of the distance from fingertip to floor, for example, ranges up to 48.2 cm, while the distance of the hand has a maximum difference of 26.84 cm (see Table 1). Not only the maximum difference of the fingertip is bigger than of the hand, but Figure 8 also shows the number of prominent outliers is bigger when looking at the fingertips. It is also noticeable when looking at Figures 7 and 8, that the distances of the fingertips and hands were calculated often larger than the ground truth, while the average calculated knee angle is mostly smaller than the ground truth.

Table 1: Minimal, maximal and mean difference as well as RMSE of calculated and ground truth values for all 80 data points (10 subjects, 2 executions, 2 cameras, left and right side) each.

	Distance (in cm)		Knee Angle (in °)	
	Fingertip	Hand	Smallest	Mean
Max	48.20	26.84	66.47	39.56
Min	0.00	0.04	0.14	0.04
Mean	4.33	3.21	18.54	8.08
RMSE	8.86	5.09	23.48	11.76

However, when leaving out the videos with an incorrect body tracking due to occlusion by the subject's hair as well as the videos with the subjects hands outside of the frame, the maximum differences between calculated values and ground truth are considerably lower (see Table 2). The maximum difference of the smallest distance from fingertip to floor is now approximately 40 cm smaller while the maximum difference of the hand is approximately 20 cm smaller. The maximum differences of the knee angle on the other side has remained (almost) the same. The mean of the differences between calculated values and ground truth values is in all cases slightly smaller when looking only at the videos with good body tracking.

There are multiple possible thresholds for classifying the quality of the exercise. Often, distances of the fingertips of 0-10 cm are taken as typically (Janka et al., 2019). The confusion matrix for a threshold of 10 cm is seen in Table 3. With $\leq$ 10 cm as the positive value, precision of these results is 1.0 and recall is 0.94. The knees should be fully extended. As described in Section 3.4, the threshold should therefore be 170°. The confusion matrix for this threshold for the mean of the knee angles is seen in Table 4. Here, precision is also 1.0, while recall is 0.54, with $>$ 170° as the positive value.

Table 2: Minimal, maximal and mean difference as well as RMSE of calculated and ground truth values for videos with good body tracking and visible fingertips (48 data points each).

	Distance (in cm)		Knee Angle (in °)	
	Fingertip	Hand	Smallest	Mean
Max	7.93	5.18	50.27	39.56
Min	0.00	0.11	0.14	0.04
Mean	2.82	2.41	15.47	7.14
RMSE	3.75	2.77	20.47	10.71

Table 3: Confusion matrix for the distance of the fingertips for videos with good body tracking and visible fingertips.

		Calculated Distance	
		$>$ 10 cm	$\leq$ 10 cm
Ground Truth	$>$ 10 cm	32	0
	$\leq$ 10 cm	1	15

Table 4: Confusion matrix for the mean knee angle for videos with good body tracking and visible fingertips.

		Calculated Mean Angle	
		$<$ 170°	$\geq$ 170°
Ground Truth	$<$ 170°	20	0
	$\geq$ 170°	13	15

5 DISCUSSION

With the described study we tested if automated tracking and evaluation of an exemplary exercise with a camera-based system is feasible. In this process we noticed, that there are some major inaccuracies in the results of the automatic tracking. The quality of the body tracking seems to have a considerable influence on the results. Poor body tracking is mainly caused by hair/upper body occlusions or by body parts outside the image frame. When excluding the results caused by these low quality body tracking outcomes, the maximal differences between automatic evaluation and ground truth are considerably reduced. This showed, that the quality of the body tracking has an impact on the results of an automatic exercise tracking and evaluation. When automatically analyzing an exercise for a fitness check, possible causes for occlusions should be taken into consideration and - if identified- removed. Specifically, this means that the subject should tie up his/her hair and wear tighter clothing. Even though the influence of clothing was not analyzed in this paper, it is reasonable to assume that loose clothing is more likely to cause occlusion than tight clothing. In addition, the camera's angle of view must be chosen so that the camera sees as many parts of the subject's body as possible and there is as little self-occlusion as possible. It also has to be ensured that the subject is completely within the camera's field of view. In addition, the floor detection needs to be taken into account when analysing distances between joint positions and the floor, as the position and orientation of the floor plane varies between different frames.

A direct comparison with the systems described in Section 2 is not possible because the focus of the evaluation of the other approaches was different. However, when looking at the accuracy of the position of the hands of the body tracking during a gait analysis compared to the accuracy of the hand/finger to floor distances, our accuracy is very good. While the average euclidean distance between detected hand position and ground truth is about 4.5 cm and 5.5 cm (Albert et al., 2020), our average difference between minimal calculated hand-floor-distance and ground truth is only about half as big. The RMSE of 11.9° of the knee angle calculated in literature (Ma et al., 2020) is very similar to the RMSE of the mean of our calculated knee angle. Our RMSE of the smallest knee angle, however, is about twice as high. Although all frames with the feet further away from the floor than 5 cm were excluded, the smallest calculated angle is probably still strongly influenced by the faulty body tracking of the legs (such as the floating legs). This can probably also be seen in the detected average knee angle often being smaller than the ground truth. The confusion matrix of the mean knee angle for videos with good body tracking and visible fingertips shows this as well: Almost half of all actual angles $\geq 170°$ are detected incorrectly. The recall of the fingertip distance, on the other hand, is considerably higher. Only one > 10 cm calculated distance is actually $\leq$ 10 cm. For fingertips and knees, all positive classified values are actually positive. Considering these comparisons and the suggestions above, an analysis of the fingertip distance is feasible. The average knee angle is often smaller than the ground truth, but if this error is reduced, for example with better positioning of the camera, the calculated angle should probably also provide appropriate results.

In future work, attention should be paid to adjusting the camera's angle of view to optimize body tracking results and avoid occlusions. It can also be tested whether other body tracking methods give better results, and thus more accurate knee angles. Subsequently, more exercises will be looked at to evaluate if the automatic analysis of multiple exercises is feasible. If this is the case, the software will then be adapted to enable subjects to perform exercises independently and have their execution tracked and evaluated automatically to help physicians to individually adapt patient's exercise therapies.

ACKNOWLEDGEMENTS

Funded by the German Federal Ministry of Education and Research (Project No. 16SV8580), as well as by the Lower Saxony Ministry of Science and Culture (grant number 11-76251-12-10/19 ZN3491) within the Lower Saxony "Vorab" of the Volkswagen Foundation and supported by the Center for Digital Innovations (ZDIN).

REFERENCES

Alabbasi, H., Gradinaru, A., Moldoveanu, F., and Moldoveanu, A. (2015). Human motion tracking & evaluation using kinect v2 sensor. In *2015 E-Health and Bioengineering Conference (EHB)*, pages 1–4.

Albert, J. A., Owolabi, V., Gebel, A., Brahms, C. M., Granacher, U., and Arnrich, B. (2020). Evaluation of the pose tracking performance of the azure kinect and kinect v2 for gait analysis in comparison with a gold standard: A pilot study. *Sensors*, 20(18).

Barker, K. and Eickmeyer, S. (2020). Therapeutic exercise. *The Medical clinics of North America*, 104(2):189–198.

Bredow, J., Bloess, K., Oppermann, J., Boese, C., Löhrer, L., and Eysel, P. (2016). Konservative therapie beim unspezifischen, chronischen kreuzschmerz. *Der Orthopäde*, 45:573–578.

Bulat, A., Kossaifi, J., Tzimiropoulos, G., and Pantic, M. (2020). Toward fast and accurate human pose estimation via soft-gated skip connections.

Büker, L. C., Zuber, F., Hein, A., and Fudickar, S. (2021). Hrdepthnet: Depth image-based marker-less tracking of body joints. *Sensors*, 21(4).

Cao, Z., Hidalgo Martinez, G., Simon, T., Wei, S., and Sheikh, Y. A. (2019). Openpose: Realtime multi-person 2d pose estimation using part affinity fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Carse, B., Meadows, B., Bowers, R., and Rowe, P. (2013). Affordable clinical gait analysis: An assessment of the marker tracking accuracy of a new low-cost optical 3d motion analysis system. *Physiotherapy*, 99(4):347–351.

Chen, S. and Yang, R. (2018). Pose trainer: Correcting exercise posture using pose estimation.

Chun, J., Park, S., and Ji, M. (2018). 3d human pose estimation from rgb-d images using deep learning method. In *Proceedings of the 2018 International Conference* *on Sensors, Signal and Image Processing*, SSIP 2018, page 51–55, New York, NY, USA. Association for Computing Machinery.

Conner, C. and Poor, G. M. (2016). Correcting exercise form using body tracking. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '16, page 3028–3034, New York, NY, USA. Association for Computing Machinery.

Diraco, G., Leone, A., and Siciliano, P. (2010). An active vision system for fall detection and posture recognition in elderly healthcare. In *2010 Design, Automation & Test in Europe Conference & Exhibition (DATE 2010)*, pages 1536–1541.

Garrido-Castro, J. L., Medina-Carnicer, R., Schiottis, R., Galisteo, A. M., Collantes-Estevez, E., and Gonzalez-Navas, C. (2012). Assessment of spinal mobility in ankylosing spondylitis using a video-based motion capture system. *Manual Therapy*, 17(5):422–426.

Gurcay, E., Bal, A., Eksioglu, E., Hasturk, A. E., Gurcay, A. G., and Cakci, A. (2009). Acute low back pain: clinical course and prognostic factors. *Disability and rehabilitation*, 31(10):840–845.

Hoffmann, M. D., Kraemer, W. J., and Judelson, D. A. (2010). Therapeutic exercise. In Frontera, W. R. and DeLisa, J. A., editors, *DeLisa's Physical Medicine and Rehabilitation*, chapter 61, page 1654. Lippincott Williams & Wilkins Health, 5 edition.

Irving, G., Neves, A. L., Dambha-Miller, H., Oishi, A., Tagashira, H., Verho, A., and Holden, J. (2017). International variations in primary care physician consultation time: a systematic review of 67 countries. *BMJ open*, 7(10):e017902.

Janka, M., Merkel, A., and Schuh, A. (2019). Diagnostik an der lendenwirbelsäule. *MMW-Fortschritte der Medizin*, 161(1):55–58.

Liu, H., Liu, F., Fan, X., and Huang, D. (2021). Polarized self-attention: Towards high-quality pixel-wise regression.

Ma, Y., Sheng, B., Hart, R., and Zhang, Y. (2020). The validity of a dual azure kinect-based motion capture system for gait analysis: a preliminary study. In *2020 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, pages 1201–1206.

Marschall, J., Hildebrandt, S., Zich, K., Tisch, T., Sörensen, J., and Nolting, H.-D. (2018). Gesundheitsreport 2018. beiträge zur gesundheitsökonomie und versorgungsforschung (band 21). https://www.dak.de/dak/download/gesundheitsreport-2108884.pdf. page 134.

Mattiuzzi, C., Lippi, G., and Bovo, C. (2020). Current epidemiology of low back pain. *JJournal of Hospital Management and Health Policy*, 4.

Microsoft Inc. (2020). About Azure Kinect DK — Microsoft Docs.

Microsoft Inc. (2021). Azure kinect dk hardware specifications.

Olaogun, M. O., Adedoyin, R. A., Ikem, I. C., and Anifaloba, O. R. (2004). Reliability of rating low back pain with a visual analogue scale and a semantic differential scale. *Physiotherapy theory and practice*, 20(2):135–142.

Perret, C., Poiraudeau, S., Fermanian, J., Colau, M. M. L., Benhamou, M. A. M., and Revel, M. (2001). Validity, reliability, and responsiveness of the fingertip-to-floor test. *Archives of physical medicine and rehabilitation*, 82(11):1566–1570.

Rector, K., Bennett, C. L., and Kientz, J. A. (2013). Eyes-free yoga: An exergame using depth cameras for blind & low vision exercise. In *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '13, New York, NY, USA. Association for Computing Machinery.

Saß, A.-C., Lampert, T., Prütz, F., Seeling, S., Starker, A., Kroll, L. E., Rommel, A., Ryl, L., and Ziese, T. (2015). Gesundheit in deutschland. gesundheitsberichterstattung des bundes. gemeinsam getragen von rki und destatis. page 69.

Schmidt, C. O., Raspe, H., Pfingsten, M., Hasenbring, M., Basler, H. D., Eich, W., and Kohlmann, T. (2007). Back pain in the german adult population: prevalence, severity, and sociodemographic correlates in a multiregional survey. *Spine*, 32(18):2005–2011.

statista (2022). Percentage of u.s. respondents that were prescribed select treatments for their back pain as of 2017, by age.

Thar, M. C., Winn, K. Z. N., and Funabiki, N. (2019). A proposal of yoga pose assessment method using pose detection for self-learning. In *2019 International Conference on Advanced Information Technologies (ICAIT)*, pages 137–142.

Tölgyessy, M., Dekan, M., Chovanec, L., and Hubinský, P. (2021). Evaluation of the azure kinect and its comparison to kinect v1 and kinect v2. *Sensors*, 21(2).

von der Lippe, E., Krause, L., Porst, M., Wengler, A., Leddin, J., Müller, A., Zeisler, M.-L., Anton, A., Rommel, A., and study group, B. . (2021). Journal of health monitoring. prävalenz von rücken- und nackenschmerzen in deutschland. ergebnisse der krankheitslast- studie burden 2020.

Wang, J., Qiu, K., Peng, H., Fu, J., and Zhu, J. (2019). Ai coach: Deep human pose estimation and analysis for personalized athletic training assistance. In *Proceedings of the 27th ACM International Conference on Multimedia*, MM '19, page 2228–2230, New York, NY, USA. Association for Computing Machinery.

Yang, L., Ren, Y., Hu, H., and Tian, B. (2015). New fast fall detection method based on spatio-temporal context tracking of head by using depth images. *Sensors*, 15(9):23004–23019.

Ye, M., Wang, X., Yang, R., Ren, L., and Pollefeys, M. (2011). Accurate 3d pose estimation from a single depth image. In *2011 International Conference on Computer Vision*, pages 731–738.

Zhou, Z., Tsubouchi, Y., and Yatani, K. (2019). Visualizing out-of-synchronization in group dancing. In *The Adjunct Publication of the 32nd Annual ACM Symposium on User Interface Software and Technology*, UIST '19, page 107–109, New York, NY, USA. Association for Computing Machinery.