RELIABILITY AND VALIDITY OF A DEEP LEARNING ALGORITHM BASED MARKERLESS MOTION CAPTURE SYSTEM IN MEASURING SQUATS

Sakurako Musha[1], Daishiro Kobayashi[1], Yuya Takaku[1], Yasuko Hirono[2], Takuya Otsu[2], and Norihisa Fujii[2]

[1]Sportip, Inc., Tokyo, Japan
[2]Faculty of Health and Sport Sciences, University of Tsukuba, Tsukuba, Japan

This study aimed to compare the performance of a traditional marker-based motion capture system and a video-based markerless system in analyzing squats and to determine the reliability and validity of the markerless system. Twenty-one squats were recorded using a marker-based motion capture system and a 2D video camera. We analyzed the 2D video data using Sportip Motion 3D, a deep learning-based 3D human pose estimation algorithm based specifically on sports activities, and the peak lower limb joint angles were calculated by both systems. There was an excellent agreement between VICON and Sportip Motion 3D for all joint angles (hip intraclass correlation coefficient (ICC) = 0.96, knee ICC = 0.92, ankle ICC = 0.86), with average differences of less than 1.3°. These results indicate that squat analysis using Sportip Motion 3D is equally reliable and accurate as the conventional marker-based method.

KEYWORDS: human pose estimation, motion analysis, deep learning

INTRODUCTION: In sports and clinical situations, coaches and therapists assess motions by observing athletes to identify problems. In other words, motion analysis in the current coaching field is subjective and not objective, as it depends on the skills of the evaluator. For detailed motion assessment, a 3D motion analysis system with markers, such as VICON, was used. While this is the gold-standard approach, it has some limitations. Firstly, the equipment is extremely expensive, causing a significant economic cost would be inevitable. Secondly, sensors and markers must be used, increasing the cost of time and the need for specialized knowledge. Therefore, it is difficult to incorporate marker-based motion analysis into various coaching fields, and it can only be used in special environments, including laboratories and hospitals (Mundermann, Corazza, & Andriacchi, 2006).

In recent years, markerless motion analysis systems have been developed to address these issues. Some deep learning-based methods have been published for a task called “human pose estimation,” which estimates the position of body landmarks in an image (Sun et al., 2019), and their accuracy has been improving. They can be used to estimate the subject's posture at each frame and analyze the motion in the video. Several studies on motion analysis using open-source software for pose estimation, such as Open Pose, have been conducted, but just a few in sports science. Furthermore, the reliability and validity of sports motion analysis using such software have not been established (Ota et al., 2021). Sportip Motion 3D (Sportip, Inc., Tokyo, Japan) is a deep learning-based 3D human pose estimation model that learns mainly on sports activities from VICON and incorporates its methods into model architecture, loss function, and post-processing. Thus, the performance of this human pose estimation model in fitness and sports activities is thought to be excellent. However, it has not been investigated in academic studies yet.

Therefore, this study aimed to clarify the relevance and agreement of squatting analysis using VICON, a marker-based motion capture system, and Sportip Motion 3D, a markerless system, and to confirm the validity of Sportip Motion 3D as a measurement method.

METHODS:

Subjects and Motion task: Seven healthy university adults (height 174.1 ± 4.5 cm, mass 64.2 ± 9.2 kg) performed a bilateral squat, squatting from a standing position with their hands extended horizontally until their thighs were parallel to the ground. The subjects repeated the movement three times, for a total of 21 squats to be measured.

Data collection: The positions of markers attached on the body were recorded using a 14-camera VICON motion capture system (VICON MX+, VICON Motion Systems Ltd., Oxford, UK), sampled at 250 Hz, and a Butterworth low-pass filter with a 5–22 Hz cut off frequency (Winter, 2004). A video camera (LUMIX FZ-300, Panasonic, Japan, 120 Hz) was set up to record the subject's left-side view. The time synchronization between the VICON and the video camera was performed by projecting the lighting of the LED (PH-145, DKH, Japan) on the screen of the video camera and capturing the trigger signal of the lighting in the VICON. Using the data from the video camera, Sportip Motion 3D estimated 24 feature points: head, ears, suprasternal, shoulders, elbows, wrists, hands, lower ribs, hips, knees, ankles, heels, and toes (Fig. 1).

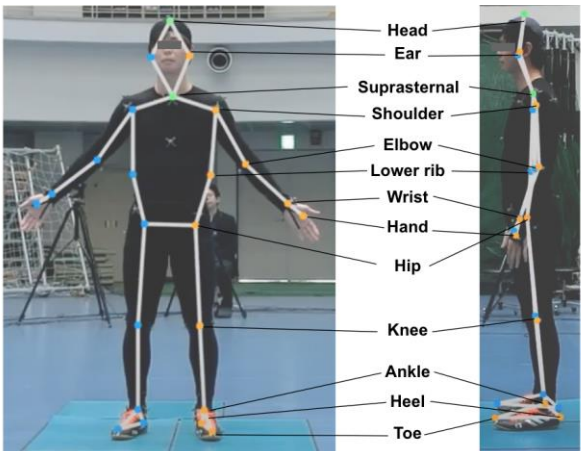

Figure 1: Placement of each feature point in Sportip Motion 3D.

Data analysis: The peak flexion angles of the hip, knee, and ankle joints in the sagittal plane were calculated from the Sportip Motion 3D and the marker positions from VICON, respectively. The hip joint angle was defined as the angle formed by the straight line connecting "left hip" and "left knee" to the straight line connecting "suprasternal" and "left hip". The knee joint angle as the angle formed by "left hip", "left knee", and "left ankle". The ankle joint angle was defined as the angle formed by the straight line connecting "left heel" and "left toe" to the straight line connecting "left knee" and "left ankle". Sportip Motion 3D uses deep convolutional neural network, a deep learning approach, to recognize image features, identify anatomical landmarks, and estimate human poses in 3D. This pose estimation model learned more than 700,000 images from Sportip, Inc.'s original dataset and the Microsoft COCO dataset (Lin et al., 2014), which included different individuals performing several actions in various clothing.

Statistical analysis: The intraclass correlation coefficient [ICC (2, 1)] confirmed the validity of the peak lower limb joint flexion angles generated by Sportip Motion 3D and VICON. Furthermore, the Bland–Altman analysis was used to assess the data from both systems. We performed statistical analysis using IBM SPSS Statistics, v.28 (IBM Japan Ltd., Tokyo, Japan).

RESULTS: Figure 2 depicts a representative example of temporal variations in joint angles during squatting. A strongly significant association was discovered between the data obtained using Sportip Motion 3D and those using VICON (R^2 = 0.86–0.94, $p < 0.05$). According to the criteria proposed by Landis and Koch (1977), the ICC (2.1) of Sportip Motion 3D and VICON were determined to be "almost perfect" for all joint angles (hip ICC = 0.96, knee ICC = 0.92, ankle ICC = 0.86). The flexion angles measured by both systems in a squat differed by less than 1.3° on average (hip 1.1° ± 3.5°, knee −1.8° ± 3.2°, ankle 1.0° ± 2.0°).

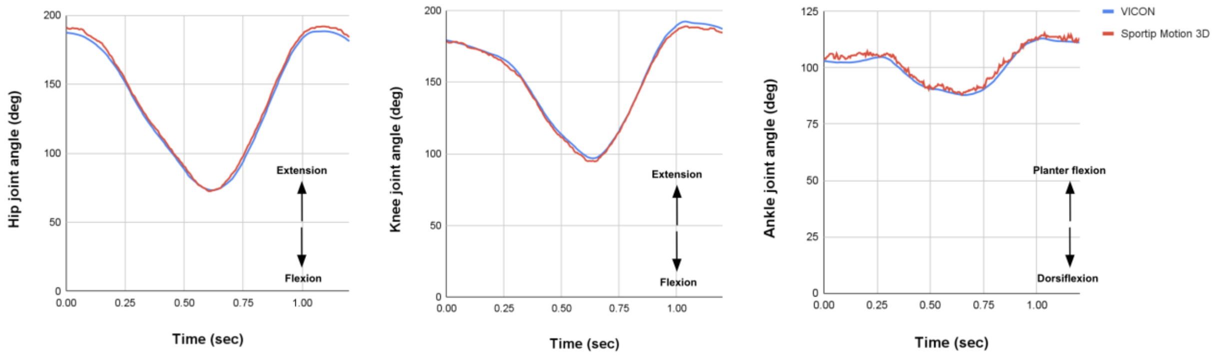

Figure 2: A representative example of temporal variations during the squat.

Figure 3 depicts the Bland–Altman plot of each joint angle measured using Sportip Motion 3D and VICON. There was a slightly proportional bias toward hip and knee joint angles, with tendencies for smaller flexion to produce smaller values and larger flexion to produce bigger values than VICON.

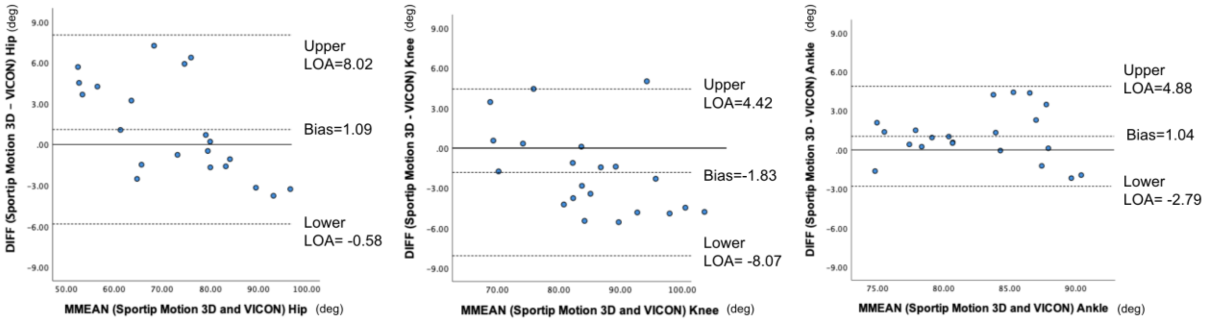

Figure 3: The Bland–Altman plot for each angle measured using Sportip Motion 3D and VICON.

DISCUSSION: This is the first study to validate the performance of a markerless system based on Sportip Motion 3D using complex and fast movements, such as sports activities, as learning data. The results indicate that it can measure the joint positions of lower limbs during a squat with an excellent agreement with VICON, a marker-based system.

Recently, research on markerless motion analysis has received a lot of attention. For example, systems for 3D human pose estimation using a depth sensor, such as Kinect, have been developed, and numerous studies have been conducted to verify the reliability and validity of assessment using them. In a study, Hu et al. (2021) discovered that Kinect had high agreement with VICON in squat analysis (ICC > 0.8). However, although Kinect is less expensive than a 3D motion analysis system, such as VICON, it is not cheap enough for everyone to use. Additionally, it requires the installation of a dedicated camera, thereby complicating the setup process and limiting its application to a few facilities. Therefore, the problem of being unable to provide motion assessment using objective measures remains in the coaching field.

Several types of research have been conducted using software that can perform pose estimation from images taken with a smartphone without the need for special equipment (e.g., Open Pose). However, only a few studies have verified the reliability and validity, and the accuracy for sports activities, such as strength training, running, throwing, and others, has not been established yet. The results of this study indicated that Sportip Motion 3D is a highly relevant and valid measurement system for VICON. Ota et al. (2020) discovered that hip angle inaccuracy and bias were the main issues for analyzing a squat in their investigation of the agreement with VICON using Open Pose (hip ICC = 0.37, knee ICC = 0.83, ankle ICC = 0.75). In this study, Sportip Motion 3D revealed better ICCs in all joint angles with VICON than Open Pose, which is an open-source software, and it is considered a motion analysis system that solves open-source software problems.

Additionally, for the hip and knee joint angles, the larger the flexion angle, the greater the value tended to be compared to that of VICON, and the smaller the flexion angle, the lesser the value tended to be. Thus, this suggests that during flexion, the estimated feature points of the knee or hip joint have a proportional bias. Therefore, the bias of the feature points themselves should be clarified and reduced in the future.
In this study, the comparison was based on the joint angle, but it is necessary to confirm the difference in coordinate values and other variables in the future. In addition, the fact that motion tasks were relatively simple movements and analyzed in a single plane may have enabled Sportip Motion 3D to demonstrate high accuracy. Therefore, it is imperative to verify the performance of the system in multiplex movements, including sports activities. With additional validation, the system can further accelerate the generalization of motion analysis and solve problems in the coaching field.

CONCLUSION: This study indicates that the markerless motion capture system, Sportip Motion 3D, can measure lower limb joint angles during a squat with a high agreement with common marker-based systems (hip ICC = 0.96, knee ICC = 0.92, ankle ICC = 0.86). Also, the software can measure flexion angles with a difference of less than 1.3° on average. These findings reveal that squat analysis using Sportip Motion 3D has a performance comparable to that of conventional marker-based methods. Thus, in the future, it is necessary to confirm the difference in coordinate values and other variables.

REFERENCES

Hu, G., Wang, W., Chen, B., Zhi, H., Li, Y., Shen, Y., & Wang, K. (2021). Concurrent validity of evaluating knee kinematics using Kinect system during rehabilitation exercise. *Medicine in Novel Technology and Devices*, 11, 100068.
Landis, J.R., & Koch, G.G. (1977). The Measurement of observer agreement for categorical data. *Biometrics*, 33, 159-74.
Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., & Zitnick, C.L. (2014). Microsoft coco: Common objects in context. *In European Conference on Computer Vision* (pp. 740-755).
Mundermann, L., Corazza, S., & Andriacchi, T.P. (2006). The evolution of methods for the capture of human movement leading to markerless motion capture for biomechanical applications. *Journal of Neuroengineering and Rehabilitation*, 3, 6.
Ota, M., Tateuchi, H., Hashiguchi, T., Kato, T., Ogino, Y., Yamagata, M., & Ichihashi, N. (2020). Verification of reliability and validity of motion analysis systems during bilateral squat using human pose tracking algorithm. *Gait & posture*, 80, 62-7.
Sun, K., Xiao, B., Liu, D., & Wang, J. (2019). Deep high-resolution representation learning for human pose estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 5693-5703).
Winter, D. A. (2009). *Biomechanics and motor control of human movement*. John Wiley & Sons.

ACKNOWLEDGEMENTS: We would like to thank Prof. Fujii for the insightful discussion. We would also like to appreciate the Laboratory of Sports Biomechanics at the University of Tsukuba for the data collection.