Accuracy Evaluation of Computer Vision-based Markerless Human Pose Estimation for Measuring Shoulder Range of Motion

Thomas Hellstén and Jonny Karlsson
Arcada University of Applied Sciences
School of Engineering, Culture and Wellbeing, Helsinki, Finland
emails: thomas.hellsten@arcada.fi, jonny.karlsson@arcada.fi

***Abstract*— TeleRehabilitation (TR) requires precise joint Range Of Motion (ROM) measurement methods. This study assessed the accuracy of a Computer Vision (CV)-based markerless Human Pose Estimation (HPE) application for active shoulder ROM by comparing it with Universal Goniometry (UG) in 20 healthy volunteers. The correlation coefficients between the two methods were 0.94 for shoulder extension, 0.83 for adduction, 0.76 for abduction, and 0.67 for flexion, with mean differences ranging from 3.6° in flexion to -7° in adduction. These findings indicate that the markerless CV application is a moderately accurate tool for measuring shoulder ROM.**

Keywords-computer vision; markerless; range of motion; shoulder.

I. INTRODUCTTION

Measuring joint Range Of Motion (ROM) is essential for healthcare professionals to evaluate and treat patients with joint disorders, as it quantifies flexibility and function to establish baseline mobility, monitor progress, and tailor interventions for rehabilitation [1]. Traditionally, this evaluation is done by professionals with a Universal Goniometry (UG) at clinics [2]. ROM assessment methods are evolving with TeleRehabilitation (TR) approaches [3] that show the feasibility of CV-based Human Pose Estimation (HPE) for at-home use via a computing device with an integrated camera. This study builds on earlier testing of CV-based HPE for shoulder ROM [5] by assessing its accuracy in measuring active shoulder flexion, extension, abduction, and adduction compared to manually measured UG results, with the extended abstract structured into Section II (methods), Section III (accuracy results), Section IV (discussion), and Section V (conclusions and work).

II. METHODS

A prototype application was developed to evaluate the accuracy of CV-based marklerless HPE for shoulder ROM measurements. The application uses You Only Look Once, version 8 (YOLOv8) pose estimation to detect the shoulder, elbow, and hip joints and the law of cosines to calculate the shoulder angle [5]. Prior to the study, written consent was obtained from participants after the research protocol was approved by the Arcada University of Applied Science Research Committee (March 2024). Joint angles measured automatically by the CV-based markless HPE application were compared with manually measured UG results on 20 healthy participants (students and personnel aged 18 or older) by two near-graduating physiotherapy students using standardized assessments of the left shoulder (active flexion, extension, abduction, and adduction) with instructions provided from a test manual and demonstration by the test leader. All measurements were recorded in a blinded manner. The prototype registered joint angles with a timestamp that was concealed until manual UG measurements, and the results were collected into a single file for analysis. Environmental factors were standardized by performing all measurements in the same room with consistent lighting, temperature, and computer and webcam placement (55 cm high and approximately 2 m from the participant).

A. *Statistical analysis*

Descriptive statistics for the voluntary test participants were computed and are reported in the results section. Accuracy was assessed with Pearson's correlation (*r*) analysis to calculate the correlation between the CV-based markerless HPE application and the manually UG measured shoulder ROMs. Bland–Altman plot analysis was used to estimate the agreement between the two methods. For correlation analysis, the following classification was used: 1.00–0.90 as very strong, 0.89–0.70 as strong, 0.69–0.50 as moderate, 0.49–0.30 as weak, and 0.29–0 as very weak [6].

III. RESULTS

The study participants (N=20) were healthy young adults (10 females, 20 males) aged between 20 and 33 years (mean: 22.9 years), as shown in Table 1.

TABLE I. CHARACTERISTICS OF THE STUDY PARTICIPANTS

Participants (n[a])	**Age, years; mean (SD[b])**	**Length; (cm[c]); mean (SD[b])**	**Weight (kg[d]); mean (SD[b])**	**BMI[e]; mean (SD[b])**
Total (20)	23.4 (1.9)	178.7 (7.9)	78.0 (12.8)	24.3 (2.3)
Female (4)	23.5 (1.9)	170.3 (3.0)	64.3 (4.4)	22.1 (0.8)
Male (16)	23.1 (2.0)	180.8 (7.3)	81.4 (11.8)	24.8 (2.3)

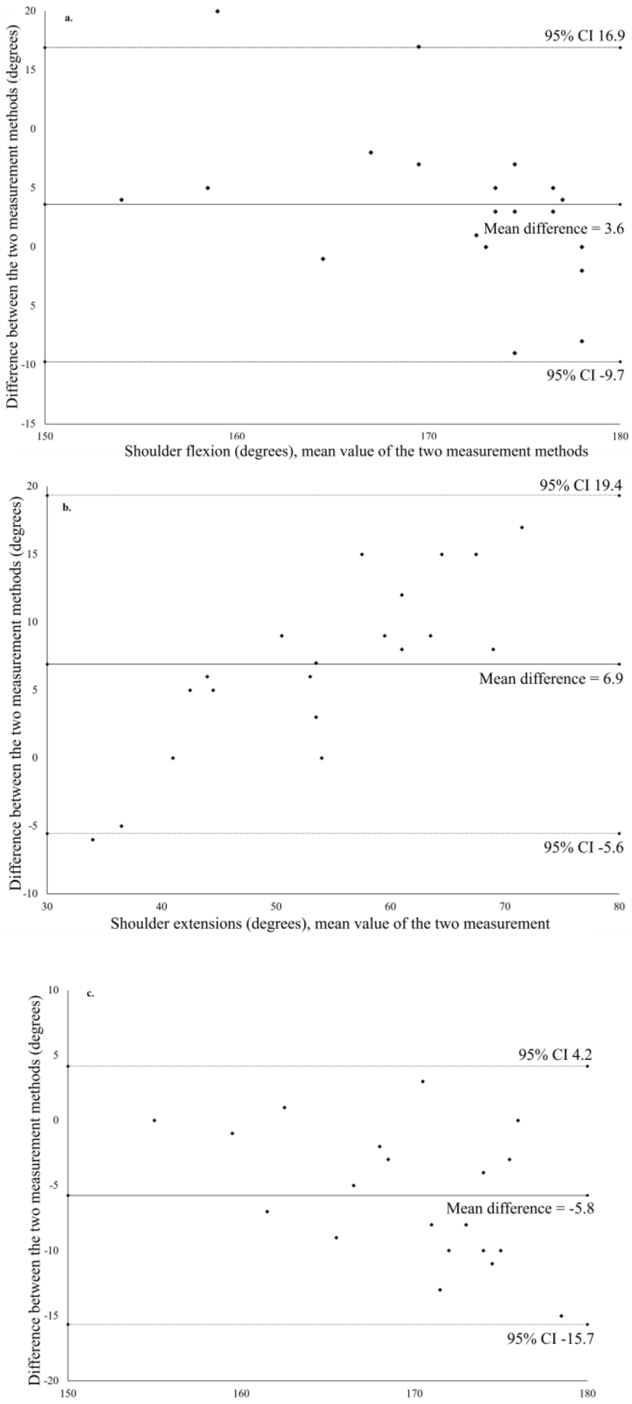
a. n: number of participants, b. SD: standard deviation, c. cm: centimeter, d. kg: kilogram, e. BMI: body mass index (kg/m2)

The results showed a very strong correlation (Persons *r* value) in active extension 0.94, strong correlation in active

Courtesy of IARIA Board and IARIA Press. Original source: ThinkMind Digital Library https://www.thinkmind.org

Copyright (c) IARIA, 2025. ISBN: 978-1-68558-270-8

adduction 0.83, abduction 0.76 and moderate correlation in active flexion 0.67. The mean difference (degrees) between the two methods was lowest in active shoulder flexion (3.6°). The highest mean difference (-7°) was in active shoulder adduction. Detailed results are shown in Figure 1.



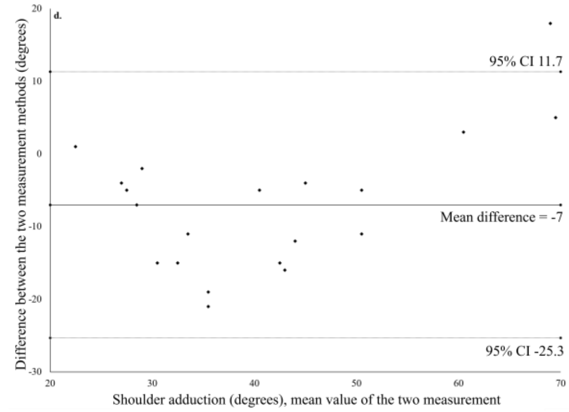


Figure 1. Bland–Altman plots for (a) shoulder flexion, (b) shoulder extension, (c) shoulder abduction, (d) shoulder adduction. The outer lines represent 95% limits of agreement. The middle line represents the mean of the differences between the two measurement methods.

Active shoulder flexion had two outliers outside the limits of agreement, and active shoulder adduction had one outlier outside the limits of agreement.

IV. DISCUSSION

The study shows a strong correlation between UG and the CV-based markerless HPE method for nearly all shoulder ROMs, extension, abduction, and adduction, with the exception of shoulder flexion. In abduction and adduction, the CV-based markerless HPE application tended to yield slightly lower values than UG. Despite some variation between the two methods, the results remain clinically acceptable given that UG, the standard for professionals, has its own measurement error [7]. However, since accuracy was assessed under standardized clinical conditions using a small group of healthy young adults, it is uncertain if these findings would be equally valid for older individuals or those with various diseases. Consequently, the prototype should be evaluated on more diverse populations and under different background and lighting conditions. Preliminary tests indicate that the prototype is sensitive to darker environments and inconsistent backgrounds, suggesting that adjustments to the underlying YOLOv8 pose estimation model may be needed for reliable home use. Overall, our results demonstrate that CV-based markerless HPE has great potential.

V. CONCLUSION AND FUTURE WORK

Our study indicates that CV-based markerless HPE applications show great potential as a promising means for implementing automatic real-time TR. However, they must be rigorously tested and developed collaboratively before becoming a standard tool in daily healthcare practice.

ACKNOWLEDGMENT

This work was supported by the “Fonden för teknisk undervisning & forskning” (TUF) under Grant 1580, a fund for supporting education and technical research at Arcada University of Applied Sciences.

Courtesy of IARIA Board and IARIA Press. Original source: ThinkMind Digital Library https://www.thinkmind.org

REFERENCES

[1] V. Lind, M. Svensson, and M. L. Harringe, "Reliability and validity of a digital goniometer for measuring knee joint range of motion," Measurement in Physical Education and Exercise Science, vol. 26, no. 3, pp. 191-198, Nov. 2022, doi: 10.1080/1091367X.2021.2004150.

[2] T. G. Russell, G. A. Jull, and R. Wootton, "Can the Internet be used as a medium to evaluate knee angle?," Manual Therapy, vol. 8, no. 4, pp. 242-246, Nov. 2003, doi: 10.1016/S1356-689X(03)00016-X..

[3] T. Hellstén, J. Arokoski, T. Sjögren, A.-M. Jäppinen, and J. Kettunen, "The Current State of Remote Physiotherapy in Finland: Cross-sectional Web-Based Questionnaire Study," JMIR Rehabil Assist Technol, vol. 9, no. 2, p. e35569, Jun. 2022, doi: 10.2196/35569.

[4] E. Hannink et al., "Validity and feasibility of remote measurement systems for functional movement and posture assessments in people with axial spondylarthritis," Healthcare Technology Letters, vol. 9, no. 6, pp. 110-118, Dec. 2022, doi: 10.1049/htl2.12038.

[5] T. Hellstén and J. Karlsson, "Using Computer Vision based Markerless Pose Estimation for Measuring Shoulder Range of Motion," presented at the eTELEMED 2024, The Sixteenth International Conference on eHealth, Telemedicine, and Social Medicine, Barcelona, Spain, 2024. [retrieved: February, 2025] https://www.thinkmind.org/library/eTELEMED/eTELEMED_2024/etelemed_2024_1_60_40016.html.

[6] D. E. Hinkle, S. G. Jurs, and W. Wiersma, Applied statistics for the behavioral sciences / Dennis E. Hinkle, William Wiersma, Stephen G. Jurs, 5th ed. Boston: Houghton Mifflin, 2003.

[7] M. J. Kolber and W. J. Hanney, "The reliability and concurrent validity of shoulder mobility measurements using a digital inclinometer and goniometer: a technical report,", Int J Sports Phys Ther, vol. 7, no. 3, pp. 306-13, June 2012.

Courtesy of IARIA Board and IARIA Press. Original source: ThinkMind Digital Library https://www.thinkmind.org