



Published in final edited form as:

IEEE Trans Hum Mach Syst. 2025 April ; 55(2): 309–314. doi:10.1109/thms.2025.3539187.

A Single-Camera Method for Estimating Lift Asymmetry Angles Using Deep Learning Computer Vision Algorithms

Zhengyang Lou,

Department of Electrical and Computer Engineering, University of Wisconsin-Madison, Madison, WI 53706 USA

Zitong Zhan,

Department of Electrical and Computer Engineering, University of Wisconsin-Madison, Madison, WI 53706 USA. He is now with the Department of Computer Science and Engineering, University at Buffalo, Buffalo, NY 14260 USA

Huan Xu,

Department of Electrical and Computer Engineering University of Wisconsin-Madison, Madison, WI 53706 USA. He is now with the School of Computer Science, Georgia Institute of Technology, Atlanta, GA 30332 USA

Yin Li,

Department of Biostatistics & Medical Informatics and Department of Computer Sciences, University of Wisconsin-Madison, Madison, WI 53706 USA

Yu Hen Hu [Life Fellow, IEEE],

Department of Electrical and Computer Engineering, University of Wisconsin-Madison, Madison, WI 53706 USA

Ming-Lun Lu,

CDC National Institute for Occupational Safety and Health, Division of Field Studies and Engineering, Cincinnati, OH 45226 USA

Dwight M. Werren,

CDC National Institute for Occupational Safety and Health, Taft Laboratories, Cincinnati, OH 45226 USA

Robert G. Radwin [Senior Member, IEEE]

Department of Industrial and Systems Engineering and the Department of Biomedical Engineering, University of Wisconsin-Madison, Madison, WI 53706 USA

Abstract

A computer vision (CV) method to automatically measure the revised NIOSH lifting equation asymmetry angle (A) from a single camera is described and tested. A laboratory study involving

Personal use is permitted, but republication/redistribution requires IEEE permission. See <https://www.ieee.org/publications/rights/index.html> for more information.

Corresponding author: Robert G. Radwin. rradwin@wisc.edu.

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/THMS.2025.3539187>.

ten participants performing various lifts was used to estimate A in comparison to ground truth joint coordinates obtained using 3-D motion capture (MoCap). To address challenges, such as obstructed views and limitations in camera placement in real-world scenarios, the CV method utilized video-derived coordinates from a selected set of landmarks. A 2-D pose estimator (HR-Net) detected landmark coordinates in each video frame, and a 3-D algorithm (VideoPose3D) estimated the depth of each 2-D landmark by analyzing its trajectories. The mean absolute precision error for the CV method, compared to MoCap measurements using the same subset of landmarks for estimating A , was 6.25° ($SD = 10.19^\circ$, $N = 360$). The mean absolute accuracy error of the CV method, compared against conventional MoCap landmark markers was 9.45° ($SD = 14.01^\circ$, $N = 360$).

Keywords

Computer vision (CV); low back pain; manual materials handling; NIOSH lifting equation; risk assessment

I. INTRODUCTION

The revised NIOSH lifting equation (RNLE) is a widely employed tool for assessing ergonomic risks associated with manual lifts [1], [2]. It has been validated through numerous epidemiological studies, demonstrating its effectiveness in predicting various outcomes related to low back pain [3]. Incorporated into ISO standard 11228-1 and numerous industry consensus standards [4], the RNLE calculates the recommended weight limit (RWL) based on a load constant and six multipliers derived from several task variables, including horizontal and vertical distances between hands and body, lift asymmetry angle (A), load displacement, hand coupling, and lifting task frequency [1].

Nonetheless, manual measurement of these task variables is arduous and time-consuming, particularly when multiple tasks are involved during a shift [5], [6]. To address this, on-screen posture specifications from video recordings have been utilized to determine posture related RNLE task variables, such as hand locations and lift asymmetry angle. That method took approximately 200 seconds per posture specification [6]. Additionally, challenges such as load projection and workplace obstructions further complicate manual measurements of the asymmetric angle.

Measuring A presents the greatest challenge among the RNLE task variables due to the intricate identification of the asymmetry line and sagittal line during a lift [5]. The angle is crucial for determining the asymmetry multiplier (AM)

$$AM = 1 - .0032A. \quad (1)$$

Angle A quantifies the displacement of a lifted object from the worker's mid-sagittal plane at the start or end of a lift, measured in degrees at both points. Specifically applicable to asymmetric lifts at the origin or destination (see Fig. 1), A is operationally defined as the

angle between the asymmetry line and the mid-sagittal line without any pivoting or stepping [1]. Given that the mid-sagittal plane is referenced to the neutral anatomical position rather than the actual position assumed during the lift's origin or destination, this measurement necessitates identifying the sagittal plane while the body deviates from neutral.

Researchers, driven by the challenges of manual measurement of the RNLE, have explored automated solutions utilizing sensors like 3-D motion capture (MoCap) or inertial measurement units (IMUs) to extract RNLE parameters. Shammaa et al. [7] outlined a method for measuring RNLE parameters, including horizontal location (H), vertical location (V), vertical travel distance (D), and A , by calculating A from MoCap marker positions on the body's surface. This method involves placing markers on both sides of the ankles to establish the mid-sagittal plane, assuming aligned ankles and neutral foot positions, which may not always hold true. While highly precise, 3-D MoCap systems are impractical for field use due to their reliance on multiple cameras and limited visibility in obstructed workplace environments.

With the improvement in accuracy and reliability of IMU sensors [8], [9], [10], their potential for estimating RNLE parameters were studied. Unlike MoCap systems, IMU sensors offer simplicity and freedom from visual obstructions. However, they have several drawbacks; they are invasive, must be attached to the body, possibly interfere with worker movements, and their accuracy may degrade over time from drift [11].

Computer vision (CV) provides a practical, non-invasive method for automated and precise lifting monitoring solely using video data. In [12], a modified twin Gaussian algorithm [13] was employed to estimate 3-D postures and joint kinematics. Mehrizi et al. [14] investigated 3-D posture estimation feasibility using deep models and a multi-view setup, adapting a classic 2-D human pose stacked hourglass network [15].

In our previous work [16], we demonstrated the feasibility of extracting RNLE parameters—such as vertical lifting heights, horizontal distances, and lifting frequencies and automatically detecting lift origin and destination using a single video camera. However, the method for estimating A required multiple cameras, which presents challenges in calibration and practical implementation in the workplace due to space constraints and computational complexity. To address these limitations, our current study presents a novel approach to estimate the RNLE asymmetry angle A , using only a single camera.

This new method leverages deep-learning-based single-view 3-D human pose estimation algorithms [17], [18], [19], [20] to predict 3D skeletal joint coordinates and estimate asymmetry angle A based on video from a monocular RGB camera. These 3-D human pose estimation algorithms have shown high accuracy in standard CV datasets like Human3.6M [21]. A recent work also studied their application in ergonomic assessments [40]. However, their suitability for automatic asymmetry angle estimation remains unexplored.

Our method considers alternative, more visible body landmarks for estimating the sagittal plane, diverging from the ankle markers used in previous research [7], which may not consistently measure the midsagittal plane and are impractical for CV field application. We compare the single-camera CV method for estimating the asymmetry angle against

asymmetry angles measured using multicamera 3-D MoCap data obtained simultaneously as ground truth measures [8], [9], assessing the accuracy of our approach relative to conventional laboratory methods. Additionally, we compare the precision of CV-measured asymmetry angles against those computed using the same alternative landmarks measured by MoCap.

II. METHODS

A. Asymmetry Angle Using Conventional 3-D MoCap Joint Coordinates (A_{MoCap})

The asymmetry angle is the angle between the sagittal line and the asymmetry line (see Fig. 1), measured at both the start and end of a lift [1]. It is assumed that at the start, the body aligns with the asymmetry line, and the feet remain stationary before the torso rotates. Similarly, at the end, the body aligns with the asymmetry line, and the feet remain stationary before releasing the object. The mid-sagittal line is defined as the line passing through the midpoint between the inner ankle bones (considered the body center) and lying in the mid-sagittal plane, based on the neutral body position in the *applications manual for the RNLE* [1] for the ease of operational measurements. Trunk rotation (lift asymmetry) occurs after lifting the object with both hands. The asymmetry line, defining the asymmetry plane, is determined by projecting the object's center on the floor (point of projection) to the body center.

Calculation of the asymmetry angle (A) requires determining the body center (O), the point of projection (P), and a point (Q) as shown in Fig. 1. According to [6], the x - y coordinates of the body center O were estimated as the mid-point between the two ankle joints, and the point of projection P was estimated using the x - y coordinates of the mid-point of the knuckles of both hands. A second point on the mid-sagittal line (point Q) is estimated as the mid-point of the second distal phalanges (second toes) of both feet of the subject. Given the x - y coordinates of O , P and Q , A may be computed using the following formula:

$$A_{\text{MoCap}} = \cos^{-1} \frac{\overrightarrow{OP} \cdot \overrightarrow{OQ}}{\|\overrightarrow{OP}\| \cdot \|\overrightarrow{OQ}\|} \quad (2)$$

where $\overrightarrow{OP}$ and $\overrightarrow{OQ}$ are two dimensional vectors in the x - y plane from body center O to P and Q , respectively. $\|\overrightarrow{OP}\|$ and $\|\overrightarrow{OQ}\|$ are the magnitudes of the corresponding vectors.

B. Asymmetry Angle Using Computer Vision Joint Coordinates (A_{CV})

We redefine the sagittal plane as the vertical plane intersecting the midpoint between hip landmarks instead of the traditional ankle joints. From an ergonomics perspective, the rationale is that the sagittal plane is perpendicular to the hips when the upright body is in the neutral anatomical position and these landmarks are not affected by rotation of the feet. For CV applications, hips are preferable to ankles due to their more reliable estimations in the field. This is because ankles, as extremity joints, present fewer constraints and are therefore harder to estimate [38], and are more prone to occlusion, which degrades accuracy

[39]. Additionally, we opt for wrist joints over knuckles, as they are anatomically close and commonly detected by CV algorithms [21].

A skeletal joint body model of a worker performing a lift is shown on the left-hand side of Fig. 2. Denote the projection of the 3-D coordinates of the hip landmarks on the floor as H_L and H_R , the projection of the 3-D coordinates of the wrist landmarks on the floor as W_L and W_R as shown in the right side of the Fig. 2. The body center here is denoted by m_h , the point of projection is now denoted by m_c . A unit-vector $\vec{v}_s$ that is perpendicular to the line segment $\overline{H_R H_L}$ and originates from m_h gives the direction of the mid-sagittal line. Similarly, a unit-vector $\vec{v}_a$ from m_h to m_c gives the direction of the asymmetry line. Then, the CV angle of asymmetry (A_{CV}) can be estimated as the arc cosine of the inner product of these two unit-vectors

$$A_{CV} = \cos^{-1}(\vec{v}_a^T \vec{v}_s). \quad (3)$$

Given a single view video, the 3-D coordinates will need to be estimated from the video using another approach, which will be described in Section II-D.

C. Asymmetry Angle Using 3-D MoCap Joint Coordinates and CV Landmarks ($\bar{A}_{CV}$)

The same landmarks used for A_{CV} were also calculated using the corresponding 3-D MoCap measurements. We term this quantity $\bar{A}_{CV}$. Specifically, $\bar{A}_{CV}$ was calculated using the hip landmarks on the floor as H_L and H_R , the projection of the 3-D coordinates of the wrist landmarks on the floor as W_L and W_R measured from the MoCap system.

D. Deep-Learning-Based 3-D Landmark Estimation

CV algorithms enable automatic estimation of a 3-D human skeleton from monocular video. Initially, body location is detected in each video frame using a human detector, resulting in a bounding box of the body region [24]. Next, a 2-D pose estimator, HR-Net, processes each cropped human image to estimate 2-D coordinates for 17 skeletal joints [25]. Tracking then links 2-D landmarks belonging to the same individual across frames [26]. Within a local window around the lifting frame, VideoPose3D [20], a 3-D pose estimator, processes tracked 2-D landmarks to estimate the relative depth of each 3-D landmark for the lifting video frame, as depicted in Fig. 3.

E. Perspective-n-Point (PnP) for Coordinate Alignment and Refinement of the Model

The VideoPose3D algorithm [20] needs to be fine-tuned with domain-specific data to enhance its performance [14], [15], [27]. This dataset, described in detail in the subsequent section, comprises video clips of subjects performing lifting tasks and corresponding 3-D joint coordinates measured using MoCap. Initially, HR-Net is employed on each frame to estimate the 2-D poses (joint locations). Then, the pre-trained VideoPose3D model is refined using HR-net estimated 2-D poses as the input and the corresponding MoCap-derived 3-D joint locations as targets.

Since MoCap-derived 3-D landmarks are in a different coordinate system from that of the video camera, alignment is necessary using a PnP algorithm [23]. The PnP algorithm estimates the camera pose, namely the extrinsic parameters of the video camera, using the set of 2-D and corresponding 3-D joint coordinates. Then, the MoCap-measured 3-D joint positions are transformed into the camera coordinate system for refinement training. To validate the alignment process, MoCap-derived 3-D joint coordinates are back-projected onto the video camera's image plane using the estimated camera pose. The normalized reprojection error, indicating the difference between back projected 2-D joint coordinates and those estimated using HRNet, divided by the diagonal length of video frames, is 0.02. Given a video frame resolution of 640 by 480 pixels, the average back projection error is approximately 16 pixels. During refinement of the VideoPose3D model, we employed leave-one-subject-out cross-validation (see Section II-G.) to mitigate concerns of overfitting. Hyperparameters selected for fine-tuning include a learning rate of 0.001, batch size of 64, training epochs of 20, and the AMSGrad optimizer [28].

F. Data Acquisition

The video data utilized in this article originated from a previous NIOSH study conducted in Cincinnati, Ohio [8]. Ten NIOSH associates participated in this prior study, performing 12 lifting tasks under IRB approval, with informed consent obtained from each participant. Hand locations for each lifting task were adjusted based on the 12 lifting zones defined by the American conference of governmental industrial hygienists (ACGIH) threshold limit value (TLV) for lifting to encompass a wide range of lifting motions [29]. Each subject repeated each lifting task three times, resulting in a total of 36 randomly assigned lifting trials completed in one session. During each trial, subjects walked from a starting point to one of 12 marked locations, lifted a wire grid frame weighing 0.45 kg, turned around, walked to shelf height to drop off the frame, and returned to the finish line. The lightweight frame ensured clear visibility of MoCap markers.

Among the 360 video recordings, 147 exhibited trunk rotation before stepping or significant hip movement, while the remaining videos had near-zero asymmetry angles A . These videos were used for analysis as a control condition for zero asymmetry angles. Asymmetry angle A was measured for three types of lifts (symmetric, asymmetric toward the camera, and asymmetric away from the camera) after lift initiation (BOL), as depicted in Fig. 4.

Ethical approval was obtained from the appropriate local ethics committee or Institutional Review Board (University of Wisconsin-Madison Federal Wide Assurance no. #FWA00005399), protocol number 2016-0297-CR006, and informed consent has been obtained.

G. Evaluation With Cross-Validation

The machine-learning algorithm underwent leave-one-subject-out cross-validation to evaluate its accuracy and robustness while mitigating potential overfitting [27]. The dataset was divided into ten subsets, and this process was repeated ten times. In each iteration, one subject was designated as the test set, and the remaining nine subjects constituted the training set. This led to the creation of ten models, each trained and evaluated using a

different test set. The overall performance, measured by angle estimation error in degrees, was averaged across these ten models.

Evaluation metrics included the mean absolute error and standard deviation of the estimated asymmetry angle compared to ground truth MoCap values [8], [9]. To complement the mean absolute error, a threshold accuracy with a 15° error interval was introduced (see Section III-A), following the RNLE Applications Manual [22]. Precision was assessed by comparing the 2-D CV-measured asymmetry angle (A_{CV}) against the asymmetry angle computed using the same MoCap-measured landmarks ($\bar{A}_{CV}$). Accuracy was evaluated by comparing the 2-D CV-measured asymmetry angle (A_{CV}) against the conventional 3-D MoCap-derived angle (A_{MoCap}).

III. Results

A. Hip-to-Hip Error ($A_{CV} - \bar{A}_{CV}$)

The CV-measured asymmetry angle (A_{CV}) was compared against the asymmetry angle, computed using the same landmarks measured by MoCap ($\bar{A}_{CV}$). We term this hip-to-hip error as both used hip joints. The overall mean absolute angular error was 6.25° (SD = 10.19°). Across all subjects, the minimum mean error was 4.58° (SD = 5.86°) and the maximum mean error was 8.59° (SD = 11.28°).

From (1), a 15° increment of A results in a 5% reduction of AM and hence a 5% change in RWL. We consider less than a 5% change in the RWL within the practical limits of precision of any of the variables included in the RNLE. Consequently, we evaluated the percentage of samples whose absolute error was less than 15° . The results showed that 91% of samples had an error of less than 15° across ten subjects.

A histogram across all subjects, for the number of estimated hip-to-hip error cases plotted against error range, is shown in Fig. 5. The histogram confirmed that most of the errors were small and within the range of $[-15^\circ, 15^\circ]$ (91%), yet had a few outliers having errors as large as -60° . We summarize the percentiles of precision error above specific thresholds in Table I using a 15° interval following the RNLE guidelines. The percentiles for errors beyond 30° and 45° were 3.1% and 1.1%, respectively.

As shown in Fig. 4, there were three types of lifts: symmetric lift (SL); turn toward the camera (TT); and turn away from the camera (TA). The mean absolute angular error and percent within 15° precision and percent within 30° precision for each type of lift are summarized in Table II. In state-of-the-art 2-D pose estimation algorithms, the 2-D coordinates of joints that are not observed directly are often estimated using prior information, leading to lower accuracy of the 2-D body joint positions. This leads to inaccurate 3-D body joint position estimates and hence higher angle estimation error. This explains why in Table II, the “turn away” lift scenario exhibits higher error rate compared to the SL and TT scenarios.

B. Hip-to-Ankle Error ($A_{CV} - A_{MoCap}$)

The CV measured asymmetry angle A_{CV} was further compared to the conventional 3-D MoCap derived angle (A_{MoCap}). We term this quantity hip-to-ankle error as A_{CV} used hip while A_{MoCap} used ankles. The overall mean absolute accuracy error was 9.45° (SD = 14.01°). Across all subjects, the minimum mean error was 5.31° (SD = 7.34°) and maximum mean error was 16.11° (SD = 16.62°).

A histogram across all subjects, for the number of estimated accuracy error cases plotted against error range, is shown in Fig. 6. The histogram confirmed that most of the errors were small and within the range of $[-15^\circ, 15^\circ]$ (79%), yet had a few outliers having errors as large as -80° .

We summarize the percentiles of accuracy error above specific thresholds in Table III using a 15° interval. The percentiles for errors beyond 30° and 45° were 6.7% and 1.9% respectively. The mean absolute accuracy error and percent within 15° accuracy and percent within 30° accuracy for each type of lift are given in Table IV.

IV. DISCUSSION

The findings highlight the efficacy of the CV method in estimating A using a single camera. The mean error for each subject remained below 15° , with minimal outliers, indicating the algorithm's accuracy for the dataset.

A. Impact of Different Estimates of Mid-Sagittal Line

The primary distinction between A_{CV} and A_{MoCap} lies in the method used to define the mid-sagittal line. In the CV method, hip orientation determines this line, while the MoCap method relies on the toes of the feet. Given that the hips initiate movement before the feet during the lifting trials' turning phase, the difference between A_{CV} and A_{MoCap} would likely be smaller than between A_{CV} and A_{MoCap} . If the hips and toes rotate at the same rate, A_{CV} and A_{MoCap} angles would become the same. The vector perpendicular to the hip, which CV employs to calculate the sagittal line, closely follows the point of projection, offering a more realistic representation than the method used in A_{MoCap} .

B. Accuracy of Manual Measurements

In [30], manual measurements of the asymmetry angle showed a mean error of -4.5° (SD = 4.8°), compared to our method's -2.34° (SD = 10.19°). Despite the slightly lower SD in manual measurements, the error difference is within 15° , likely having minimal impact on RWL calculations. The key benefit of our approach over manual measurements is not accuracy, but its non-intrusive nature, enabling potential for fully autonomous monitoring and broadening its use beyond standard workplace tasks.

V. CONCLUSION

This article demonstrated that modern CV techniques can estimate the RNLE asymmetry angle A from monocular videos with an average error below 7° . By using a single camera, this approach provides a simpler and potentially more practical method for automatically

determining this lifting risk factor in industrial settings, avoiding the need for multiple cameras.

Acknowledgments

This work was supported in part the National Institute for Occupational Safety and Health (NIOSH/CDC), R01OH011024 (Radwin), in part by the National Institute for Occupational Safety and Health, Centers for Disease Control and Prevention (NIOSH, CDC) R01OH011024 (Radwin), and in part by NIOSH Pilot Project Research Training Program from the University of Michigan Center for Occupational Health and Safety Engineering under Grant T42 OH008455. This article was recommended by Associate Editor D. Kaber.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the University of Wisconsin-Madison Institutional Review Board under Application No. 2020-0625, and performed in line with the Federal wide Assurance from the Office of Human Research Protection at the United States Department of Health and Human Services, #FWA00005399.

REFERENCES

- [1]. Water TR, Putz-Anderson V, and Garg A, “Applications manual for the revised NIOSH lifting equation,” U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health. 1994. [Online]. Available: 10.26616/NIOSH PUB941110revised092021
- [2]. Lowe B, Dempsey P, and Jones E, N. I. for Occupational Safety and H. (NIOSH), “Assessment methods used by certified ergonomics professionals,” in Proc. Hum. Factors Ergonom. Soc. Annu. Meeting, 2018, pp. 838–842.
- [3]. Fox RR, Lu M-L, Occhipinti E, and Jaeger M, “Understanding outcome metrics of the revised NIOSH lifting equation,” Appl. Ergonom, vol. 81, 2019, Art. no. 102897.
- [4]. Lu M-L, Putz-Anderson V, Garg A, and Davis KG, “Evaluation of the impact of the revised national institute for occupational safety and health lifting equation,” Hum. Factors, vol. 58, pp. 667–682, 2016. [PubMed: 26822795]
- [5]. Dempsey PG, “Usability of the revised NIOSH lifting equation,” Ergonomics, vol. 45, pp. 817–828, 2002. [PubMed: 12487684]
- [6]. Waters T, Krieg E, and Werren D, “Comparisons of the accuracies of two methods for assessing trunk posture associated with low back disorders,” in Proc. 17th World Congr. Ergonom., 2009.
- [7]. Shammaa MH, Kawachi K, and Suzuki H, “Accurate measurements for calculating lifting index of NIOSH lifting equation using motion capture technology,” in Proc. JSPE Semestrial Meeting 2007 JSPE Spring Meeting, 2007.
- [8]. Barim MS, Lu M-L, Feng S, Hughes G, Hayden M, and Werren D, “Accuracy of an algorithm using motion data of five wearable IMU sensors for estimating lifting duration and lifting risk factors,” in Proc. Hum. Factors Ergonom. Soc. Annu. Meeting, 2019, pp. 1105–1111.
- [9]. Lu M-L, Feng S, Hughes G, Barim MS, Hayden M, and Werren D, “Development of an algorithm for automatically assessing lifting risk factors using inertial measurement units,” in Proc. Hum. Factors Ergonom. Soc. Annu. Meeting, 2019, pp. 1334–1338.
- [10]. Hlucny SD and Novak D, “Characterizing human box-lifting behavior using wearable inertial motion sensors,” Sensors, vol. 20, 2020, Art. no. 2323.
- [11]. Kok M, Hol JD, and Schön TB, “Using inertial sensors for position and orientation estimation,” 2017, arXiv:1704.06053.
- [12]. Mehrizi R, Peng X, Xu X, Zhang S, Metaxas D, and Li K, “A computer vision-based method for 3D posture estimation of symmetrical lifting,” J. Biomech, vol. 69, pp. 40–46, 2018. [PubMed: 29398001]
- [13]. Bo L and Sminchisescu C, “Twin Gaussian processes for structured prediction,” Int. J. Comput. Vis, vol. 87, pp. 28–52, 2010.
- [14]. Mehrizi R, Peng X, Metaxas DN, Xu X, Zhang S, and Li K, “Predicting 3-D lower back joint load in lifting: A deep pose estimation approach,” IEEE Trans. Hum.-Mach. Syst, vol. 49, no. 1, pp. 85–94, Feb. 2019.

- [15]. Mehrizi R, Peng X, Xu X, Zhang S, and Li K, "A deep neural network-based method for estimation of 3D lifting motions," *J. Biomech*, vol. 84, pp. 87–93, 2019. [PubMed: 30587377]
- [16]. Wang X, Hu YH, Lu M-L, and Radwin RG, "The accuracy of a 2D video-based lifting monitor," *Ergonomics*, vol. 62, pp. 1043–1054, 2019. [PubMed: 31092146]
- [17]. Chen T, Fang C, Shen X, Zhu Y, Chen Z, and Luo J, "Anatomy-aware 3D human pose estimation in videos," 2020, arXiv:2002.10322.
- [18]. Lin J and Lee GH, "Trajectory space factorization for deep video-based 3D human pose estimation," 2019, arXiv:1908.08289.
- [19]. Xu J, Yu Z, Ni B, Yang J, Yang X, and Zhang W, "Deep kinematics analysis for monocular 3D human pose estimation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 896–905.
- [20]. Pavllo D, Feichtenhofer C, Grangier D, and Auli M, "3D human pose estimation in video with temporal convolutions and semi-supervised training," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 7745–7754.
- [21]. Ionescu C, Papava D, Olaru V, and Sminchisescu C, "Human3.6m: Large scale datasets and predictive methods for 3D human sensing in natural environments," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 7, pp. 1325–1339, Jul. 2014. [PubMed: 26353306]
- [22]. Waters TR, Putz-Anderson V, Garg A, and Fine LJ, "Revised NIOSH equation for the design and evaluation of manual lifting tasks," *Ergonomics*, vol. 36, pp. 749–776, 1993. [PubMed: 8339717]
- [23]. Andrew AM, *Multiple View Geometry in Computer Vision*, Bingley, United Kingdom: Emerald Group Publishing Limited, 2001.
- [24]. He K, Gkioxari G, Dollár P, and Girshick R, "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2980–2988.
- [25]. Sun K, Xiao B, Liu D, and Wang J, "Deep high-resolution representation learning for human pose estimation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 5686–5696.
- [26]. Xiao B, Wu H, and Wei Y, "Simple baselines for human pose estimation and tracking," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 466–481.
- [27]. Bishop CM and Nasrabadi NM, *Pattern Recognition and Machine Learning*, vol. 4. Berlin, Germany: Springer, 2006.
- [28]. Reddi SJ, Kale S, and Kumar S, "On the convergence of Adam and beyond," 2019, arXiv:1904.09237.
- [29]. Armstrong TJ, "ACGIH TLV for hand activity level," in *Biomechanics in Ergonomics*. Boca Raton, FL, USA: CRC Press, 2007, pp. 377–390.
- [30]. Waters TR, Baron SL, and Kemmlert K, "Accuracy of measurements for the revised NIOSH lifting equation," *Appl. Ergonom.*, vol. 29, pp. 433–438, 1998.
- [31]. Dempsey PG and Fathallah FA, "Application issues and theoretical concerns regarding the 1991 NIOSH equation asymmetry multiplier," *Int. J. Ind. Ergonom.*, vol. 23, pp. 181–191, 1999.
- [32]. Sárádi I, Linder T, Arras KO, and Leibe B, "How robust is 3D human pose estimation to occlusion?," 2018, arXiv:1808.09316.
- [33]. Cheng Y, Yang B, Wang B, and Tan RT, "3D human pose estimation using spatio-temporal networks with explicit occlusion training," in *Proc. AAAI Conf. Artif. Intell.*, 2020, pp. 10631–10638.
- [34]. Zhang J, Nie X, and Feng J, "Inference stage optimization for cross-scenario 3D human pose estimation," 2020, arXiv:2007.02054.
- [35]. Ilie A and Welch G, "Online control of active camera networks for computer vision tasks," *ACM Trans. Sensor Netw.*, vol. 10, pp. 1–40, 2014.
- [36]. Waters TR, Lu M-L, Piacitelli LA, Werren D, and Deddens JA, "Efficacy of the revised NIOSH lifting equation to predict risk of low back pain due to manual lifting: Expanded cross-sectional analysis," *J. Occup. Environ. Med.*, vol. 53, pp. 1061–1067, 2011. [PubMed: 21866048]
- [37]. Bradski G, "The OpenCV library," *Dr. Dobb's J. Softw. Tools*, vol. 120, pp. 122–125, 2000.

- [38]. Särändi I, Linder T, Arras KO, and Leibe B, “Synthetic occlusion augmentation with volumetric heatmaps for the 2018 ECCV PoseTrack challenge on 3D human pose estimation,” 2018, arXiv:1809.04987.
- [39]. Wang J, Huang S, Wang X, and Tao D, “Not all parts are created equal: 3D pose estimation by modeling bi-directional dependencies of body parts,” in Proc. IEEE/CVF Int. Conf. Comput. Vis., 2019, pp. 7770–7779.
- [40]. Murugan AS et al. , “Optimising computer vision-based ergonomic assessments: Sensitivity to camera position and monocular 3D pose model,” *Ergonomics*, vol. 68, pp. 120–137, 2025. [PubMed: 38293749]

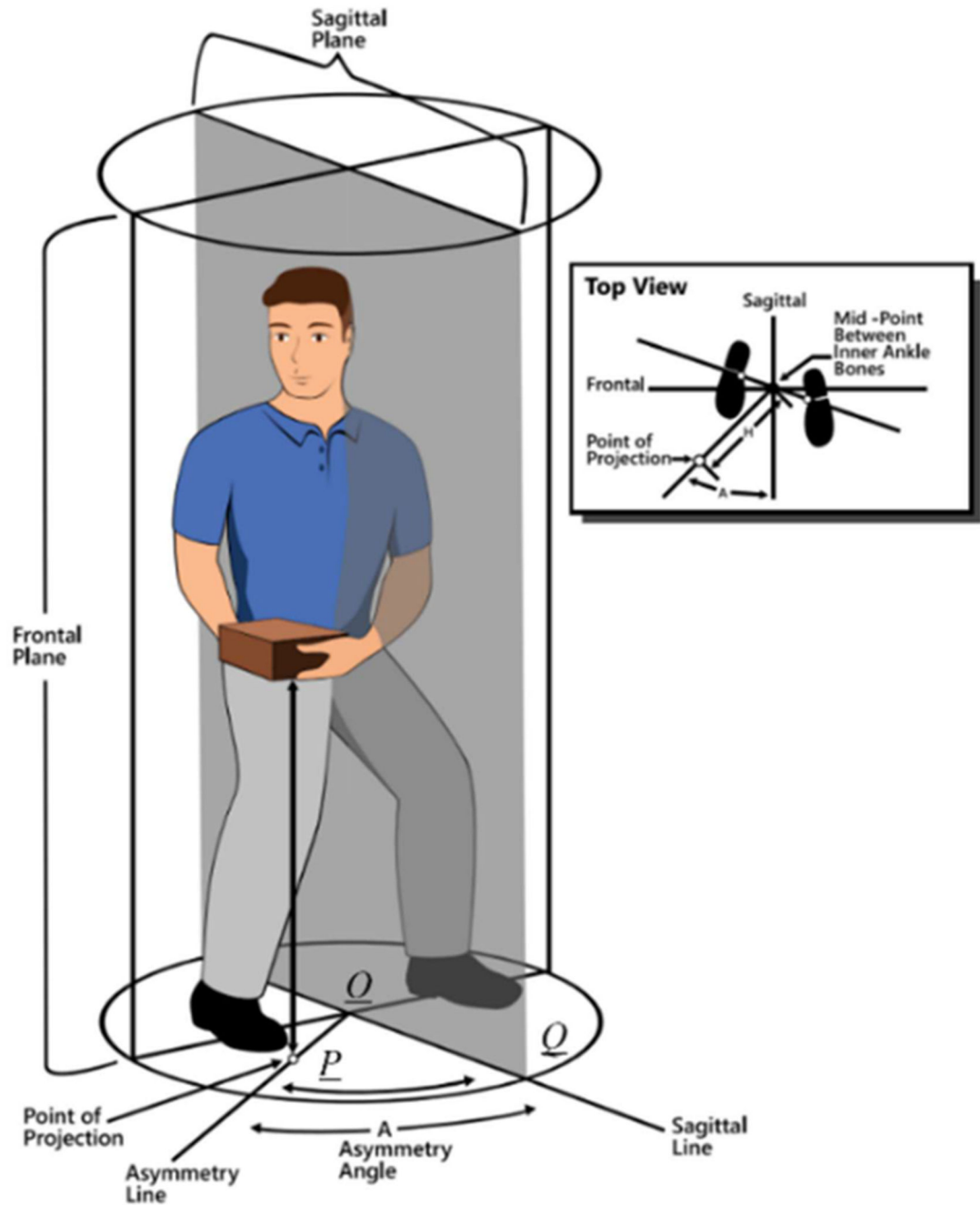


Fig. 1. Illustration of asymmetry angle. (a) From “Applications Manual for the Revised NIOSH Lifting Equation” by Waters, T. R., Putz-Anderson, V., Garg, A., Application Manual for the Revised NIOSH Lifting Equation, DHHS (NIOSH) Publication No. 94-110. (Revised 9/2021). In the public domain. Alt Text: Person lifting object while twisting to the right.

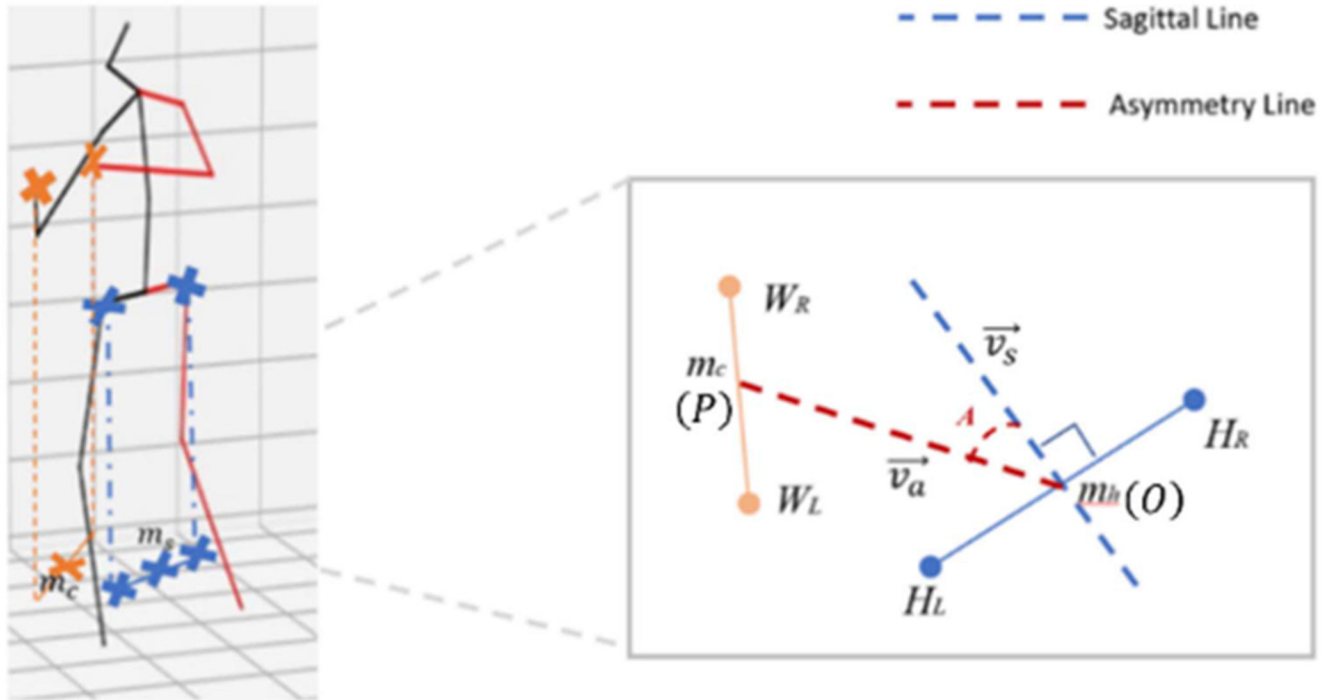


Fig. 2.

Left: 3-D alternative landmarks, either from computer-vision or MoCap. Right: top-down view demonstrating how body asymmetry angle is derived from the 3-D landmarks. Alt Text: Stick figure body linkages and projection of angles on the horizontal plane.

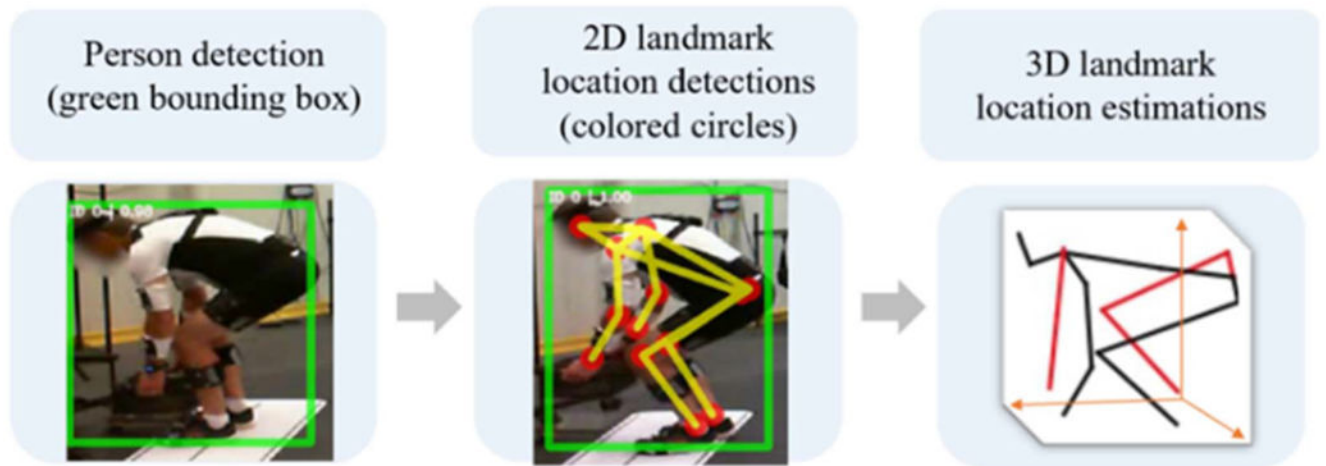


Fig. 3.

Flow chart for video-based 3-D landmark using deep learning. Left: person bounding box (green box) from the human detector [24]. Middle: 2-D landmarks (colored circles) detected by HR-Net [25] and linked over-frame by the tracking module (Xiao et al., 2018). Left and middle used slightly different frames to illustrate that successive 2-D key points are needed as inputs to VideoPose3D. Right: 3-D landmarks estimated by VideoPose3D [20].
 Alt Text: Left: Person lifting, Middle: Person lifting with stick figure superimposed. Right: Stick figure of person lifting.



Fig. 4.

An illustration of measuring the angle of lift asymmetry in this article (Left: an example of symmetric lift; Middle: an example of turning toward the camera; and Right: an example of turning away from the camera). Source: NIOSH. Alt Text: Left: Person lifting, Middle: Person turning left while lifting. Right: Person turning right while lifting.

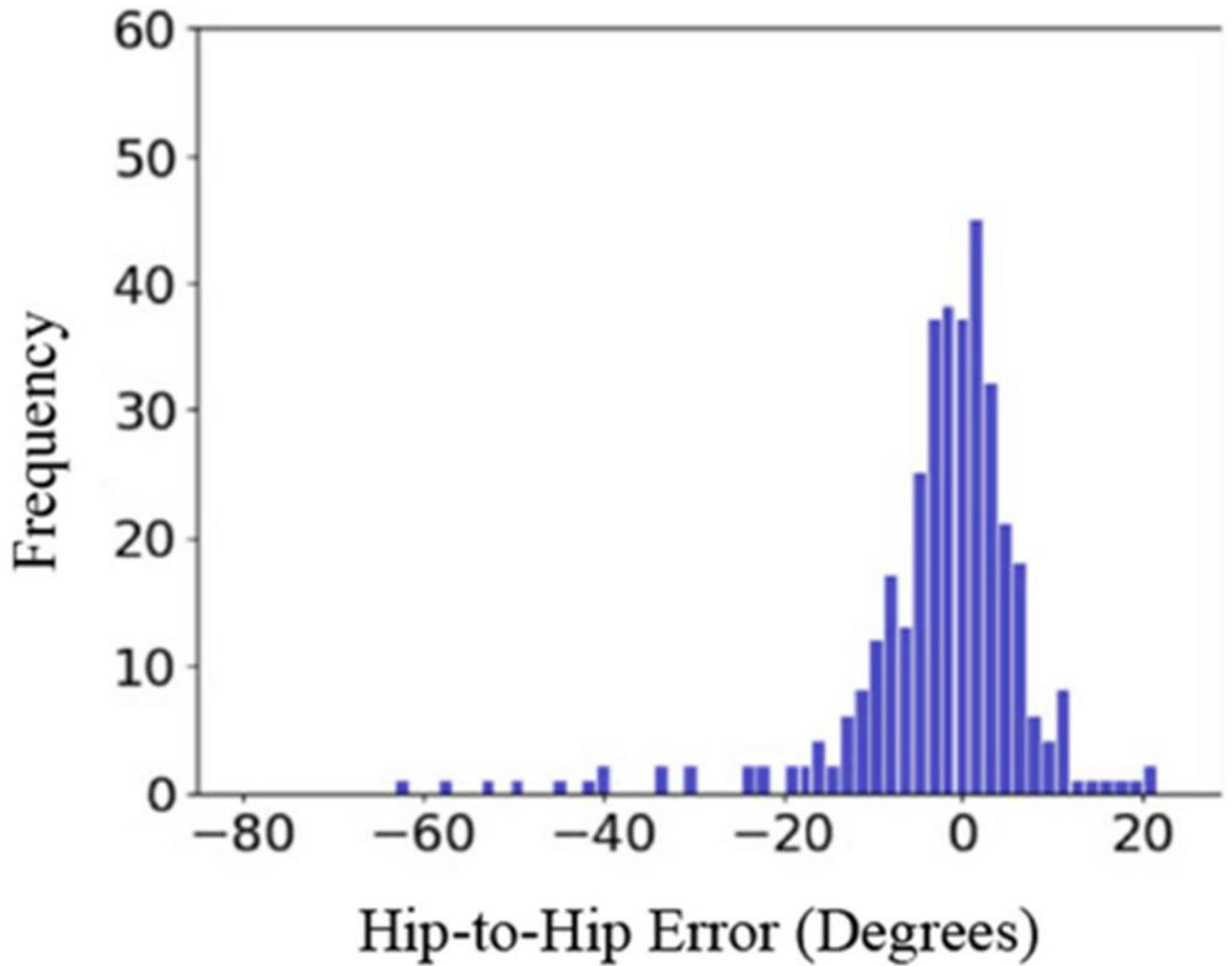


Fig. 5.

Histogram plot for the hip-to-hip error. The horizontal is the error in degrees. A negative error indicates that the estimation is smaller than ground truth. Alt Text: Histogram plot of hip-to-hip error (degrees) on the x -axis and count on the y -axis. Most error occurs around zero degrees.

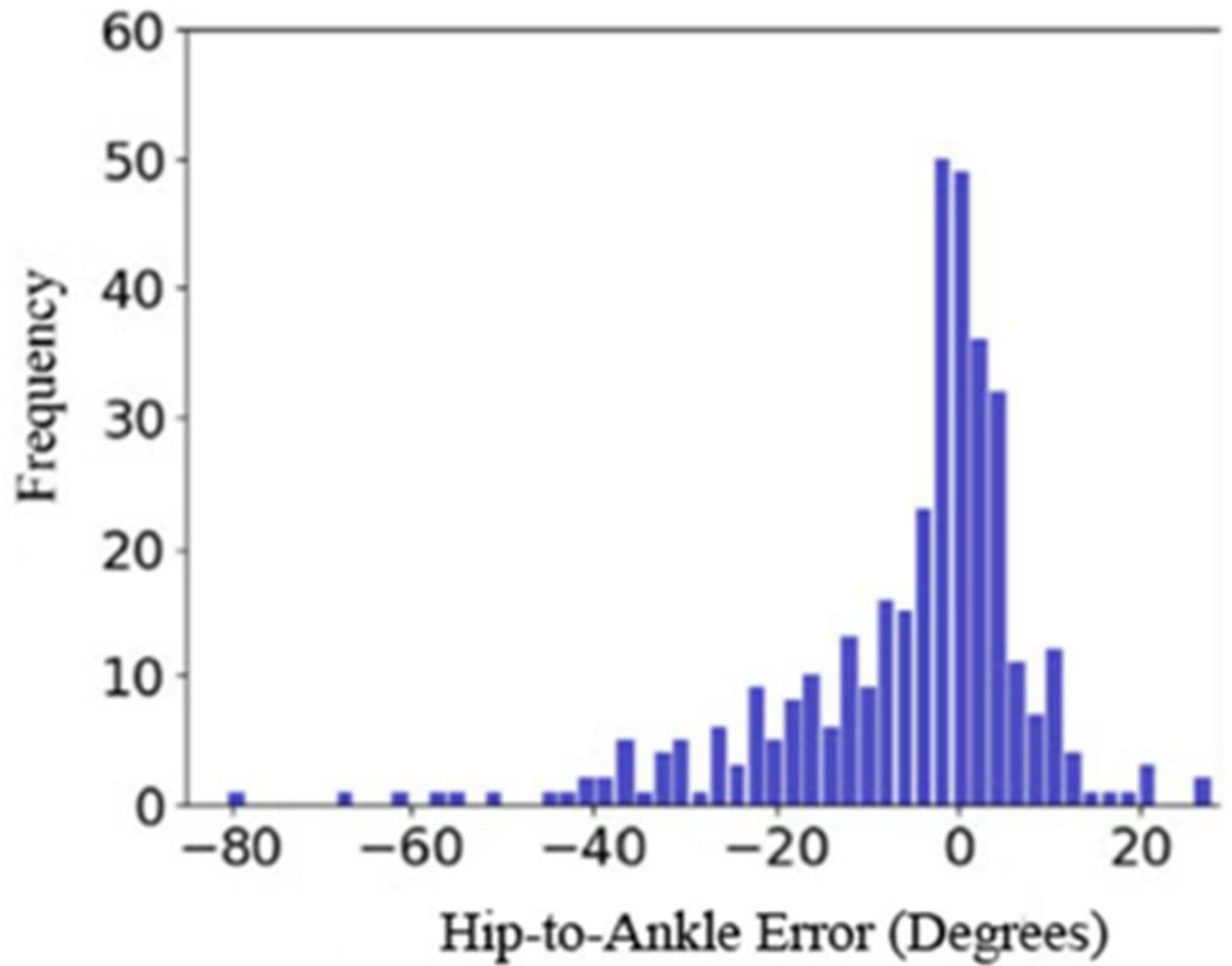


Fig. 6.

Histogram plot for the hip-to-ankle error. The horizontal axis is the error in degrees. A negative error indicates that the estimation is smaller than ground truth. Alt Text: Histogram plot of hip-to-ankle error (degrees) on the x -axis and count on the y -axis. Most errors occur around zero degrees.

TABLE I

SUMMARY OF ASYMMETRY ANGLE HIP-TO-HIP ERROR PERCENTILES

Hip-to-Hip Error	Error Percentiles
15°	8.6%
30°	3.1%
45°	1.1%
60°	0.3%

TABLE II

SUMMARY OF ASYMMETRY ANGLE HIP-TO-HIP ERROR ACROSS THREE TYPES OF LIFTS

Type of Lift	Mean±SD (°)	Percent within 15° Precision	Percent within 30° Precision
SL	4.28±5.61	98%	100%
TT	7.13±10.25	90%	97%
TA	14.30±18.10	63%	80%

TABLE III

SUMMARY OF ASYMMETRY ANGLE HIP-TO-ANKLE ERROR PERCENTILES

Hip-to-Ankle Error	Error Percentiles
15°	21.1%
30°	6.7%
45°	1.9%
60°	0.8%

TABLE IV

SUMMARY OF ASYMMETRY ACCURACY ERROR ACROSS THREE TYPES OF LIFTS

Type of Lift	Mean±SD (°)	Percent within 15° Accuracy	Percent within 30° Accuracy
SL	4.78±8.19	96%	99%
TT	15.75±14.84	53%	87%
TA	17.21±19.50	58%	83%