Journal of Biomechanics 129 (2021) 110860

Contents lists available at ScienceDirect

Journal of Biomechanics

journal homepage: www.elsevier.com/locate/jbiomech

A computer-vision method to estimate joint angles and L5/S1 moments during lifting tasks through a single camera

Hanwen Wang, Ziyang Xie, Lu Lu, Li Li, Xu Xu *

Edward P. Fitts Department of Industrial and Systems Engineering, North Carolina State University, Raleigh, NC 27695, USA

ARTICLE INFO

Keywords:
Musculoskeletal disorders
Low-back injuries
Markerless motion tracking
Joint kinematics
L5/S1 joint moment

ABSTRACT

Weight lifting is a risk factor of work-related low-back musculoskeletal disorders (MSD). From the ergonomics perspective, it is important to measure workers' body motion during a lifting task and estimate low-back joint moments to ensure the low-back biomechanical loadings are within the failure tolerance. With the recent development of advanced deep neural networks, an increasing number of computer vision algorithms have been presented to estimate 3D human poses through videos. In this study, we first performed a 3D pose estimation of lifting tasks using a single RGB camera and VideoPose3D, an open-source library with a fully convolutional model. Joint angle trajectories and L5/S1 joint moment were then calculated following a top-down inverse dynamic biomechanical model. To evaluate the accuracy of the computer-vision-based angular trajectories and L5/S1 joint moments, we conducted an experiment in which participants performed a variety of lifting tasks. The body motions of the participants were concurrently captured by an RGB camera and a laboratory-grade motion tracking system. The body joint angles and L5/S1 joint moments obtained from the camera were compared with those obtained from the motion tracking system. The results showed a strong correlation ($r > 0.9$, RMSE $< 10°$) between the two methods for shoulder flexion, trunk flexion, trunk rotation, and elbow flexion. The computer-vision-based method also yielded a good estimate for the total L5/S1 moment and the L5/S1 moment in the sagittal plane ($r > 0.9$, RMSE < 20 N·m). This study showed computer vision could facilitate safety practitioners to quickly identify the jobs with high MSD risks through field survey videos.

1. Introduction

Manual materials handling (MMH) is considered one of the work-related risk factors of low-back musculoskeletal disorder (da Costa & Vieira, 2010; U.S. Department of Labor, 2016; Yang et al., 2016). From the ergonomics perspective, it is critical to measure joint kinematics and evaluate L5/S1 joint moment during a lifting task, and ensure the low-back joint loadings are within the failure tolerance to avoid low-back injuries (Skals et al., 2021; Coenen et al., 2014)". One valid method to capture workers' body motion is to use an optical marker-based motion tracking system. Such a system is capable of obtaining three-dimensional coordinates of markers that are attached to the workers' bodies in a laboratory environment. The dynamic moments at L5/S1 joint are then calculated using workers' body motion together with body segment inertial properties (Kingma et al., 1996). To date, numerous studies have used an optical motion tracking system to investigate the L5/S1 joint moment for identifying the risks associated with a variety of lifting tasks (Desjardins et al., 1998, Kingma et al., 1998, Larivière & Gagnon, 1998). However, applying an optical motion tracking system is less practical for field studies due to its bulky size, high cost, and required expertise.

To overcome these limitations, a few studies sought to develop video-based coding systems that use human raters to observe workers' postures from the videos recorded in field studies. Raters estimate body pose in selected keyframes extracted from the recorded videos by fitting the poses to a predefined digital manikin. The workers' motion trajectories are then reconstructed by interpolating the rater-identified poses in the keyframes. In previous studies, L5/S1 joint moments were further estimated by combining the reconstructed motion and a biomechanics model (Coenen et al., 2011, Xu et al., 2012). While this method does not rely on a laboratory-based motion tracking system for capturing workers' body motion, it remains labor-intensive as raters would need to observe a large number of video frames. In addition, the accuracy of the reconstructed body motion heavily relies on the experience of raters as well as the view angle of the videos.

Researchers have also sought to apply depth sensors, such as

* Corresponding author.
E-mail address: xxu@ncsu.edu (X. Xu).

https://doi.org/10.1016/j.jbiomech.2021.110860
Accepted 2 November 2021
Available online 8 November 2021

Microsoft Kinect and Intel RealSense, to track and assess body motion. In one study, a depth sensor was used to capture shoulder kinematics during computer use for office ergonomics assessment (Xu et al., 2017). In another study, lumbosacral (L5/S1) load during static load-handling activities was estimated by a Kinect-driven model (Asadi and Arjmand, 2020). Depth sensors are low-cost, portable, and can provide a reasonable accuracy on human pose reconstruction. On the other hand, their coverage area is quite limited (Shum et al., 2013, Han et al., 2013), and the accuracy can be substantially affected by the illumination conditions of the environment (Azzari et al., 2013).

With the recent development of advanced deep neural networks, an increasing number of computer vision algorithms have been presented to estimate 3D human pose. For example, Openpose, an open-source system to detect the human body from single images, implemented 3D poses reconstructions using multiple 2D calibrated images (Cao et al., 2021). By using Openpose, a recent study used two synchronized videos during walking to compute lower limb joint kinematics by applying a triangulation algorithm on the 2D joint center coordinates derived from videos (D'Antonio et al., 2020). While Openpose can yield 3D poses from synchronized videos, camera calibration among multiple cameras is time-consuming and requires expertise in computer vision, which could be a technical burden for ergonomics practitioners. Another recent study trained three artificial neural networks using the data obtained by a motion tracking system to predict 3D postures, segmental orientations, and lumbosacral moments during load-handling activities (Aghazadeh et al., 2020). While the proposed method in this study achieved promising results, its efficiency for field use remains unclear since the hand locations need to be manually input for pose estimation and moment prediction.

In recent years, a single-camera-based 3D pose reconstruction algorithm named VideoPose3D was developed (Pavllo et al., 2019). A semi-supervised approach was introduced to process unlabeled video without any 2D ground truth annotations. This approach is able to estimate 3D poses by using a fully convolutional model generated by dilated temporal convolutions over 2D joint points. Because this algorithm only relies on the video captured from a single camera, it has a good potential for ergonomists to investigate workers' body postures and the associated joint loadings in the field through the recorded video.

In this study, we developed a computer-vision-based method for estimating joint kinematics and analyzing low-back joint moments during lifting tasks. Particularly, Detectron2 (Wu et al., 2019) was adopted for 2D key-point detection, and VideoPose3D (Pavllo et al., 2019) was applied to process the unlabeled video data and reconstruct workers' 3D poses. A top-down inverse dynamic biomechanical model was then applied to the estimated 3D pose for estimating the joint angles and the moment at L5/S1 joint. To examine the validity of this proposed method, we conducted an experiment where participants perform a variety of lifting tasks, and their motion were concurrently captured by a camera and a laboratory-based motion tracking system. The joint angular trajectories and peak L5/S1 joint moment derived from the proposed method were compared against those derived from the motion tracking system.

2. Methods

2.1. Experiment design

After a discussion with a few safety practitioners, we conducted the following experiment to mimic commonly observed lifting tasks in industry. Twelve male participants (age 47.50 $\pm$ 11.30 years; height 1.74 $\pm$ 0.07 m; weight 84.50 $\pm$ 12.70 kg) were asked to lift a 10 kg crate (39 $\times$ 31 $\times$ 22 cm) and place it on a shelf. The crate was filled with sponge in which 10 1-kg metal bars were evenly distributed to ensure the crate gravity center is at the geometric center. The participants were asked to stand in front of the crate and finish the lifting tasks without moving the feet. The initial distance between a participant and the crate as well as the lifting speed were chosen by the participants. According to the NIOSH lifting equation (Waters et al., 1993), a number of factors, such as lifting heights and asymmetric lifting angles, commonly exist in industrial lifting applications (e.g. warehouses, packing houses). The experiment consists of three vertical lifting ranges: floor to knuckle height, floor to shoulder height, and knuckle to shoulder height. Additionally, each lifting range was combined with three asymmetric angles (0°, 30° to the right, and 60° to the right), which is the angle of the end position relative to the starting position of the crate. Each lift condition was repeated twice in a full-factorial randomized design. Considering the balance among storage volume, reconstruction accuracy and processing efficiency, all lifting trials were captured by a camcorder (GR-850U, JVC) with a resolution of 720 $\times$ 480 pixels. The camera was placed on the rear-right side (135 degrees from the sagittal plane). Participants' body motion was also concurrently recorded by a motion tracking system (Motion Analysis, Santa Rosa, CA) through 45 reflective markers (Cappozzo et al., 1995) attached to the bony landmarks of the participants at 100 Hz.

2.2. Computer vision method

The workflow of the proposed video-based L5/S1 joint moment estimation method includes three major steps: 2D key-point detection, 3D reconstruction, and joint angle and moment calculation (Fig. 1). The input is the videos of each participant, and the output is the joint angle and the L5/S1 joint moment.

2.2.1. 2D key-point detection and 3D reconstruction

The recorded videos are first processed in Detectron2 (Wu et al., 2019) to estimate 2D key-points in each frame. In the input layer, the estimated 2D (x, y) coordinates of the J joints in each frame are applied in a temporal convolution with C output channels and W kernel size. B ResNet-style blocks surrounded by a skip-connection (He et al., 2016) first perform a 1D convolution with kernel size W and dilation factor $D = W^B$, followed by a convolution with kernel size = 1. In this study, J = 17, C = 1024, W = 3 and B = 4. The model parameters follow the choice of the original VideoPose3D architecture based on three primary considerations (Pavllo et al., 2019). First, the model must avoid overfitting with the selected parameters. Second, the test error must saturate quickly, eliminating the need to model long-term dependencies. Third, 17 key points including shoulder, elbow, and wrist are used to articulately represent the human pose under different conditions. Each convolution process is followed by batch normalization (Ioffe & Szegedy, 2015), rectified linear units (Nair & Hinton, 2010), and a dropout layer except the last layer (Srivastava et al., 2014). The receptive field of each block increases exponentially by a factor of W , while the quantity of parameters increases linearly. Thus, the receptive field for any output frame will include information extracted from all input frames.

A semi-supervised training method introduced in VideoPose3D (Pavllo et al., 2019) is then applied to the last layer for predicting the 3D poses in video frames. Since we do not include any ground truth pose data or the camera extrinsic parameters for the recorded videos, this method does not train a traditional supervised loss where the ground truth 3D poses data is set as the target. A projection layer is added after 3D pose estimation, and the 3D predicted poses are regressed and projected back to 2D coordinates. A penalty is applied if the 2D coordinates from the projection process are far from the 2D data input. As the global position of key joints can be arbitrary for human kinetics analysis, the coordinates of the reconstructed key joints are transformed in a way that the coordinates of the mid-hip joint are considered as the origin.

2.2.2. Angle and moment calculation

It should be noted that most of the output "joints" in VideoPose3D, such as "Shoulder" and "Hip", do not have a practical anatomical meaning. Thus, we used 9 of 17 output "key joints" from VideoPose3D to

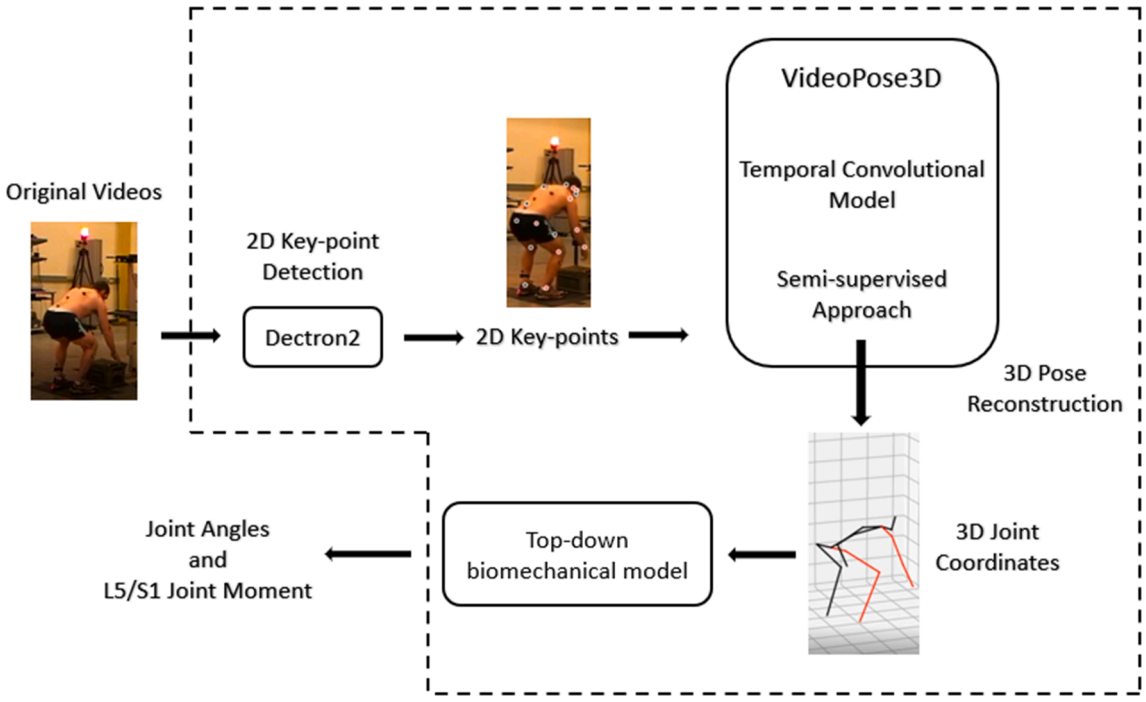

Fig. 1. Workflow of the single camera-based computer vision method.

estimate the positions of 10 anatomical joint centers or bony landmarks that were later used for biomechanical analysis (middle column in Table 1). For those joints that are not identified in VideoPose3D, such as C7 and L5/S1 joint, the anthropometric parameters were used to estimate their positions (Chaffin & Anderson, 1991). For validation purposes, these joint centers and bony landmarks were also estimated by the marker set of the motion tracking system (right column in Table 1). The locations of hip joint center, shoulder joint center, and L5/S1 are based on Seidel et al. (1995), de Leva (1996), and Reynolds (1982), correspondingly. The coordinates from both systems were first filtered by a fourth-order Butterworth low-pass filter at 8 Hz. Body segments including upper arms, forearms, hands, head, trunk above L5/S1 joint, were then defined according to a top-down inverse dynamic biomechanics model (Kingma 1996).

Body segment inertial properties, including mass (m) and moment of inertia (I), were estimated based on a previous anthropometry study (Zatsiorsky, 2002) as well as participants' weight and stature. The center of mass location (CoM_i) of each body segment, i , was determined as a proportional location of the segment length, which can be determined from the distal and proximal joint center location. Angles of the right elbow flexion, angles of the right shoulder flexion, abduction and rotation, together with the angles of trunk flexion, lateral bending and rotation (described in Table 2 and Fig. 2) across a lifting task were calculated based on the instantaneous orientations of the anatomical axes of the body segments following the ISB recommendation (Wu et al., 2005, Wu et al., 2002). The L5/S1 joint moments (M_{L5S1}) were calculated by an inverse dynamics model (Eq. (1)) (de Leva, 1996).

Table 1
The joint centers and bony landmarks derived by VideoPose3D and a motion tracking system.

Joint centers of bony landmarks for biomechanical analysis	VideoPose3D-based counterparts	Motion tracking system-based counterparts
Left hip joint center	Left hip	Left hip joint center
Left shoulder joint center	Left shoulder	Left shoulder joint center
Left elbow joint center	Left elbow	(Left lateral humeral epicondyle + left medial humeral epicondyle)/2
Left wrist joint center	Left wrist	(Left radial styloid + left ulnar styloid)/2
Right hip joint center	Right hip	Right hip joint center
Right shoulder joint center	Right shoulder	Right shoulder joint center
Right elbow joint center	Right elbow	(Right lateral humeral epicondyle + right medial humeral epicondyle)/2
Right wrist joint center	Right wrist	(Right radial styloid + right ulnar styloid)/2
Mid hip	(Left hip + right hip)/2	(Left hip + right hip)/2
Mid shoulder	(Left shoulder + right shoulder)/2	(Left shoulder + right shoulder)/2
C7	Mid shoulder + (mid shoulder - mid hip) × 0.2248	C7
L5/S1	Mid hip + (mid shoulder – mid hip) × 0.1934	L5/S1

Table 2
List of the angles to be estimated by VideoPose3D and motion tracking system.

Angle	Angle Definition
Shoulder flexion	The angle between the projection of upper arm and the projection of trunk on global Z-X plane
Shoulder abduction	The angle between the projection of upper arm and the projection of trunk on global Y-Z plane
Shoulder rotation	The angle between the projection of upper arm and the projection of trunk on global X-Y plane
Elbow flexion	The angle between the projection of forearm and the extension line of upper arm on global Z-X plane
Trunk flexion	The angle between the projection of trunk on global Z-X plane and the global Y-Z plane
Trunk lateral bending	The angle between the projection of trunk on global Y-Z plane and the global Z-X plane
Trunk rotation	The angle between the projection of trunk on global X-Y plane and the global Y-Z plane

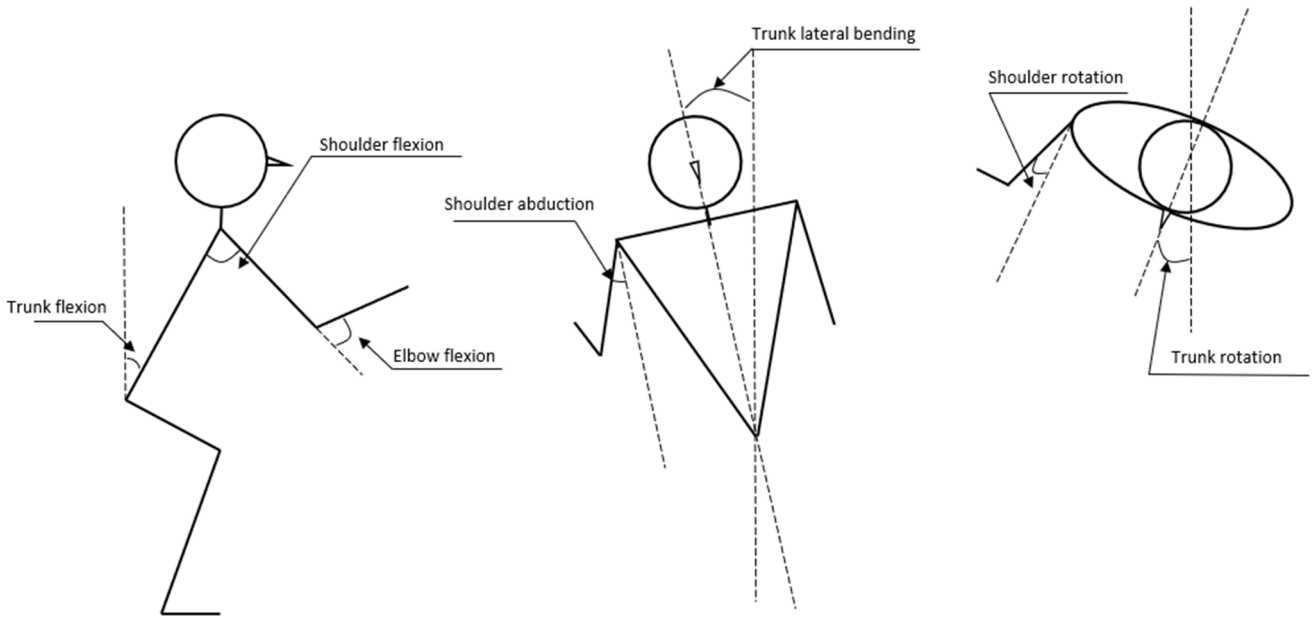

Fig. 2. The angles to be estimated.

$$M_{L5S1} = -(r_r - r_{L5S1}) \times F_r - \sum_{i=1}^{k}[(r_i - r_{L5S1}) \times m_i g] + \sum_{i=1}^{k}[(r_i - r_{L5S1}) \times m_i a_i] + \sum_{i=1}^{k}(I_i \alpha_i) \quad (1)$$

where F_r is the external force applied on the hands; $m_i g$, a_i and $I_i \alpha_i$ are gravity, acceleration and angular momentum of body segment i that are above the L5/S1 joint; r_r , r_i and r_{L5S1} are the position vectors of the external force, center of segment mass and the L5/S1 joint, k is the number of segments included in this model (upper arms, forearms, hands, head and trunk). Note that VideoPose3D is not capable to provide the 3D locations of an object beyond 17 human body key points. Thus, we assume that the weight of the crate is equally distributed on both hands during lifting tasks. The load applied on hands is then estimated based on the hand acceleration and the mass of the crate. This assumption was also applied to calculate L5/S1 moment by the motion tracking system.

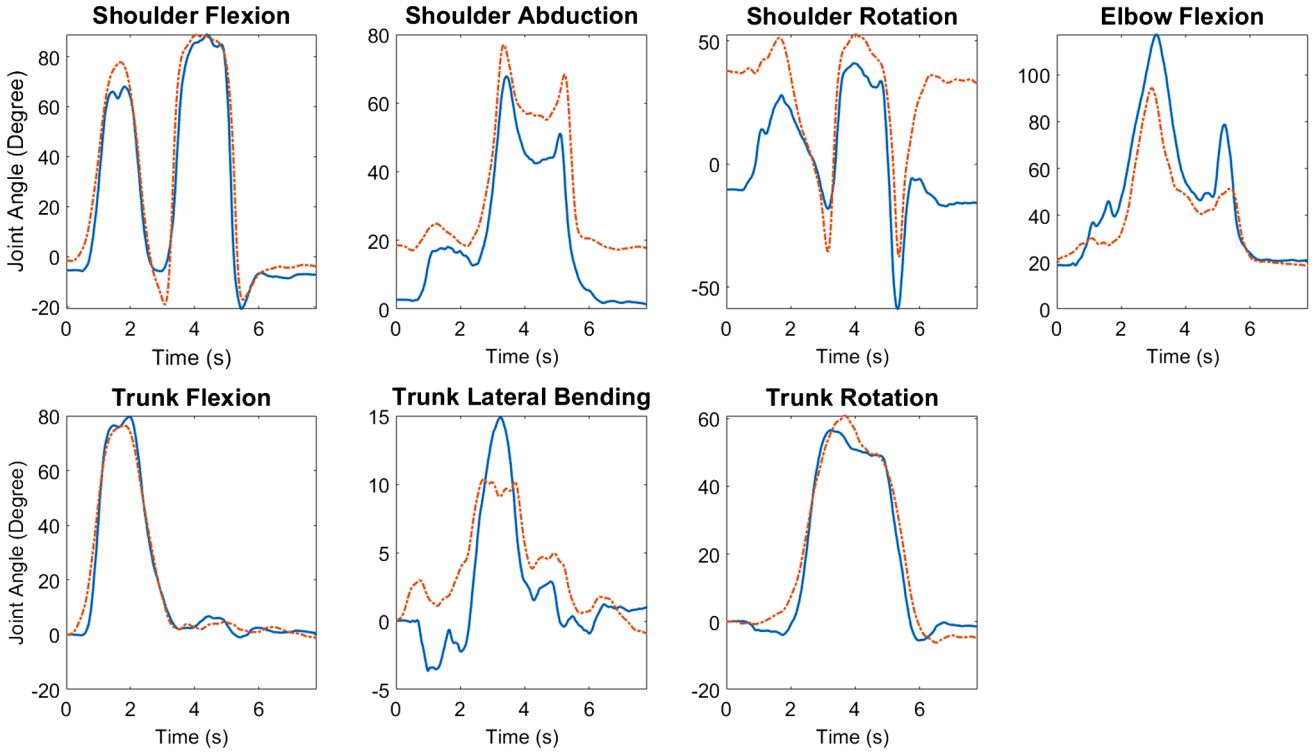

Fig. 3. The angle of major joints above the L5/S1 for a representative lifting trial (a 60° asymmetric lift with floor-to-shoulder height). The dash lines represent the computer-vision-based method and the solid lines represents the motion tracking system-based method.

2.3. Joint angle and L5/S1 moment validation

The performance of our proposed computer-vision-based lifting task assessment method is validated against the motion tracking system-based method. Linear regressions between the two methods were performed on the extracted joint angles and the peak L5/S1 joint moment for all lifting trials. The root-mean-square error (RMSE), the average absolute error (AAE) with absolute percent error, and correlation coefficient (r) were also calculated to describe the performance of the proposed method. The estimation errors were calculated by subtracting the reference values from the estimated values. Histograms of the estimation error across all trials were constructed to indicate the degree of underestimation and overestimation.

3. Results

As a representative trial illustrated in Fig. 3 and Fig. 4, all participants finished their lifting tasks in less than eight seconds. Good consistency was observed between the estimated values and the references for joint angle trajectories and the L5/S1 joint moment. Across all lifting trials, strong correlations ($r > 0.9$) between the estimated value and the references were found on shoulder flexion, trunk flexion, trunk rotation, and elbow flexion. For shoulder abduction, shoulder rotation and trunk lateral bending, the corresponding correlation coefficients were lower (0.590, 0.627 and 0.504, respectively). The average absolute error was within 10° for all estimated angles except for shoulder abduction (16.16°), shoulder rotation (15.95°) and elbow flexion (10.37°) (Fig. 5). Based on the histograms of the estimation errors for joint angle trajectories across all the lifting conditions (Fig. 6), the computer-vision-based method overestimated the joint angles for shoulder abduction, shoulder rotation and trunk rotation. The error distributions of shoulder flexion, trunk flexion, lateral bending and elbow flexion were approximately symmetric and zero-centered.

For the L5/S1 peak moment, the computer-vision-based method yielded a good estimate of the moment in the sagittal plane and the total moment. The corresponding correlation coefficients were above 0.9, RMSE were below 20 N·m and the absolute percent errors were less than 12% (Fig. 7(a)). The correlation coefficients of the L5/S1 moment in the

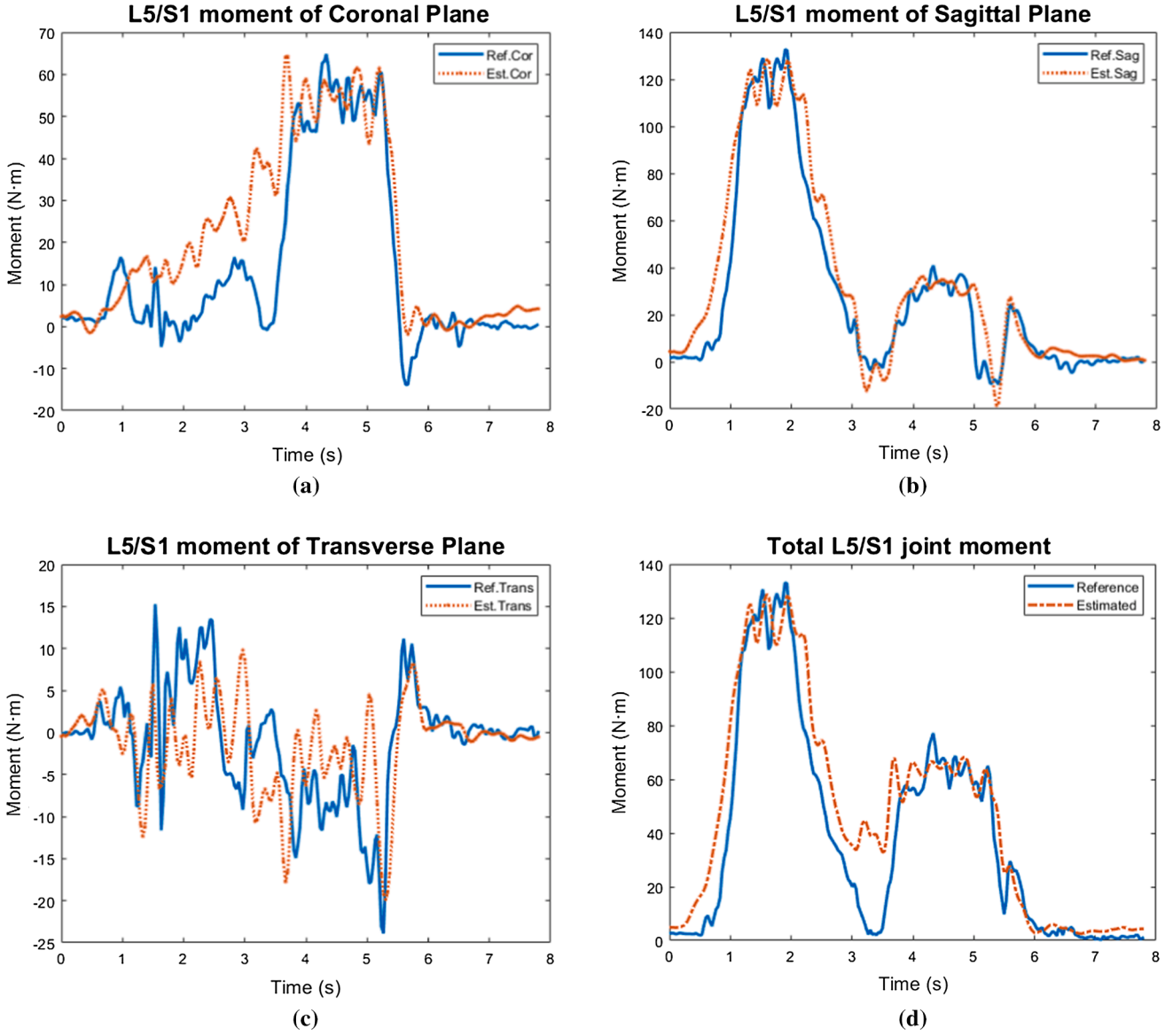

Fig. 4. (a) The estimated L5/S1 moment vs. reference L5/S1 moment in the coronal plane. (b) The estimated L5/S1 moment vs. reference L5/S1 moment in the sagittal plane. (c) The estimated L5/S1 moment vs. reference L5/S1 moment in the transverse plane (d) Total estimated L5/S1 moment (vector summation of 3D moments) vs. total reference moment. All figures are from the same representative lifting trial (a 60° asymmetric lift with floor-to-shoulder height).

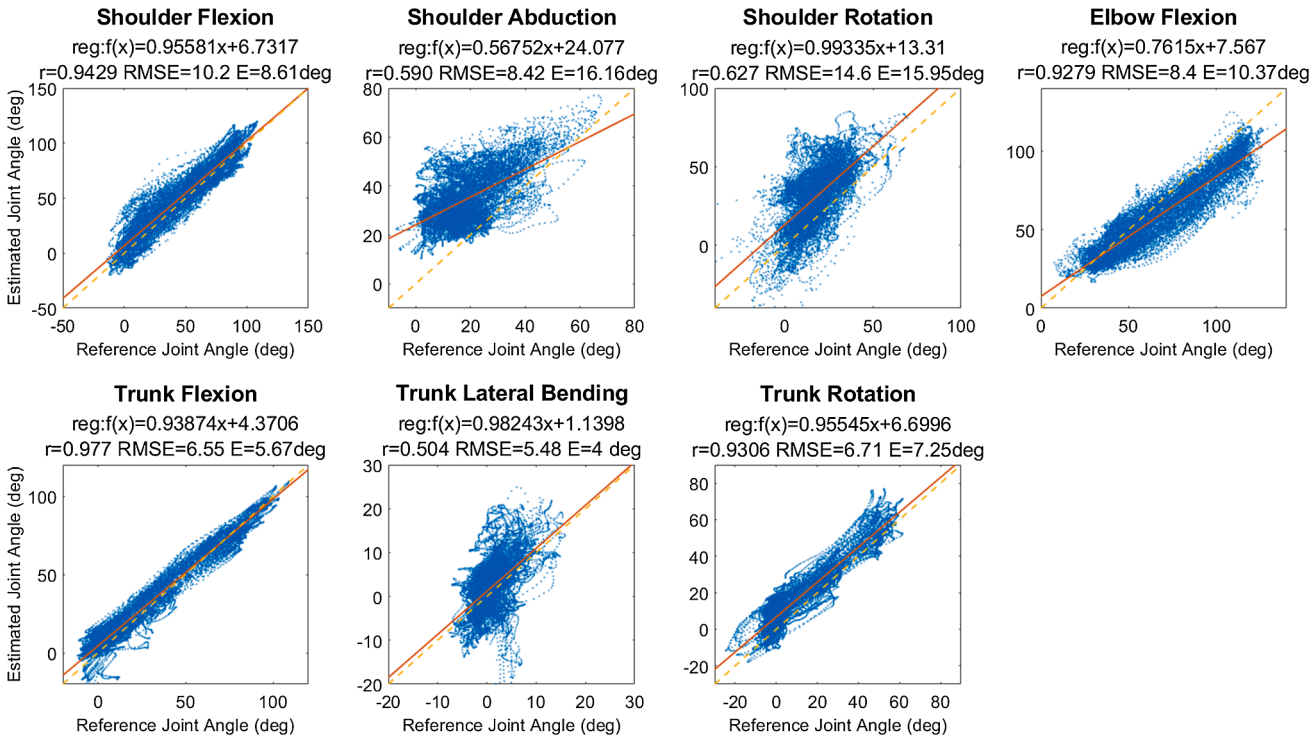

Fig. 5. The comparison between the estimated joint angles and the reference joint angles. *r* is the correlation coefficient. *RMSE* is the root-mean-square error. *E* indicates the average absolute error (AAE). *Reg* refers to the linear regression between the estimated and reference joint angles. The solid line is the linear regression line that generated from the data points and the dashed diagonal line is the identity line.

coronal plane and the transverse plane were smaller (0.785 and 0.199). Although the average absolute errors of the coronal plane and the transverse plane seemed to be lower than that of the sagittal plane, the absolute percentage errors in those two planes were much larger due to the small magnitude of the moments in these two planes. The histograms (Fig. 7(b)) showed that the proposed computer-vision-based method overestimated the peak moment in the coronal plane and underestimated the peak moment in the transverse plane.

4. Discussion

In this study, a computer-vision-based method was proposed to estimate joint angular trajectories and 3D L5/S1 joint moment during lifting tasks. The input of this method is the videos captured by a single camera. This method was validated against the references derived from a laboratory-based motion tracking system. Compared with the results from several previous computer-vision-based works (Aghazadeh et al., 2020, Mehrizi et al., 2018, Mehrizi et al., 2019), our method showed comparable efficiency and accuracy in the prediction of the joint kinematic and the moment at L5/S1 joint. The correlation coefficient and the linear regression outcomes indicated that the proposed method could provide a reasonable estimate on joint kinematics of shoulder flexion, trunk flexion, trunk rotation, and elbow flexion. The total L5/S1 peak moment and the peak moment in the sagittal plane were strongly correlated with the references obtained by the motion tracking system. The histograms of estimation errors showed that the frequencies of overestimation and underestimation were approximately identical for the peak moment in the sagittal plane.

On the other hand, the estimated angular trajectories of shoulder abduction, shoulder rotation and trunk lateral bending were not well correlated with the references. Due to the resolutions of the input videos, the computer-vision-based method has a limited sensitivity compared to a motion tracking system. In other words, our proposed method is less sensitive to small joint angle variations. Since the majority of the body movements occurs in the sagittal plane, the small joint motions in the transverse and coronal planes, such as shoulder rotation, were captured with less accuracy. Although our proposed method was less accurate than a previous study (Mehrizi et al., 2018) in terms of joint angle prediction, the current results could be less overfitted. This is because the deep neural network adopted in this study was trained by a completely independent multimodal dataset (Human3.6 M) (Ionescu et al., 2011, 2014), rather than a dataset created under the same laboratory conditions.

The correlations of the estimated peak moments in the coronal and transverse planes were not as good as the ones in the sagittal plane. Similarly, this could be explained by the limited sensitivity of the computer-vision-based method. The small movements of lateral bending and rotation were difficult to be precisely estimated. In turn, the accuracies of the moments in the coronal and the transverse planes were affected by these errors in 3D pose reconstructions. Yet, the moments of the coronal and the transverse planes were not dominant in the total L5/S1 moment. Therefore, the correlation coefficient and the absolute percentage errors of the total L5/S1 moment were close to those of the L5/S1 moment in the sagittal plane.

To further evaluate whether different asymmetric lifting angles (0°, 30° to the right, and 60° to the right) could affect the estimation error of the total moment at L5/S1 joint, a post-hoc Tukey test was further performed. Fig. 8 shows that the trials an asymmetric angle of 60 °had the smallest estimation error, since trials with 60° asymmetric angle contained more body segment movements in the coronal plane and transverse plane. As the camera is set at 135°, asymmetric lifting tasks allow camera to observe more body movement without view obstacle, and thus provide a more robust estimate on peak total moment.

It has also been found that for most trials, shoulder abduction and shoulder rotation had a significant angle difference at the start and end of the lifting when a participant was in a standing posture. Such discrepancy is attributable to the different definitions between the motion tracking system-based joint center locations and the VideoPose3D identified “joint centers”. For example, in the motion tracking system-based method, the shoulder joint center was identified based on de

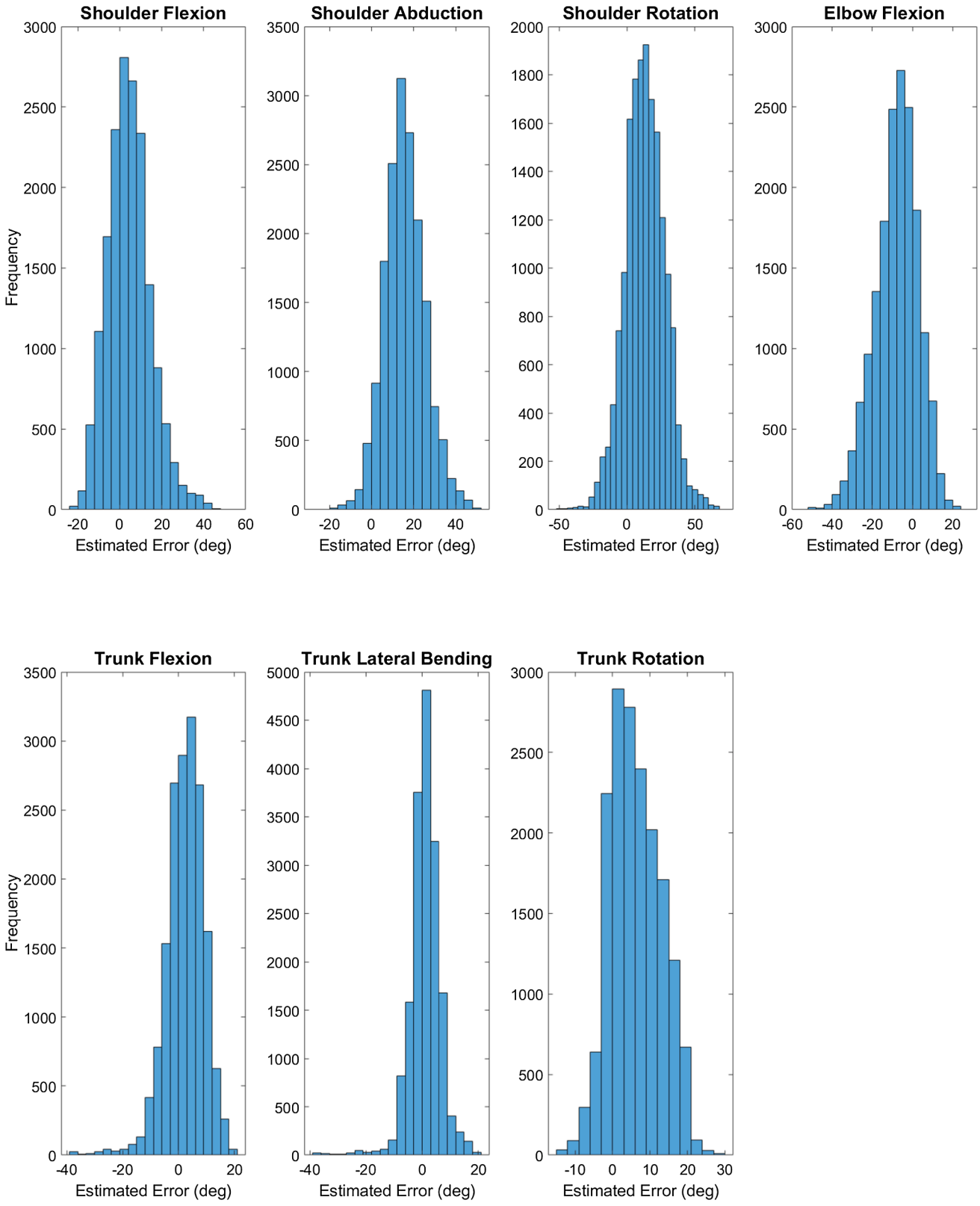

Fig. 6. Histograms of the estimation error of the joint angles. "Frequency" indicates the number of trials whose error is within a specific error range.

Leva (1996), which was described by the positions of acromion and elbow. The hip joint was identified by Seidel et al. (1995), which was calculated based on the right and left PSIS and L5/S1 joint. In the computer-vision-based method, the "joint centers" for a frame are predicted as the pixels with the highest confidence yielded by a temporal convolutional model, which has no practical anatomical meaning.

It should be noted that joint detection can be affected when the camera view of a body segment is blocked by other objects or other body segments for a short moment. Disturbance in the estimated body joint location will lead to substantial errors in body joint acceleration, which

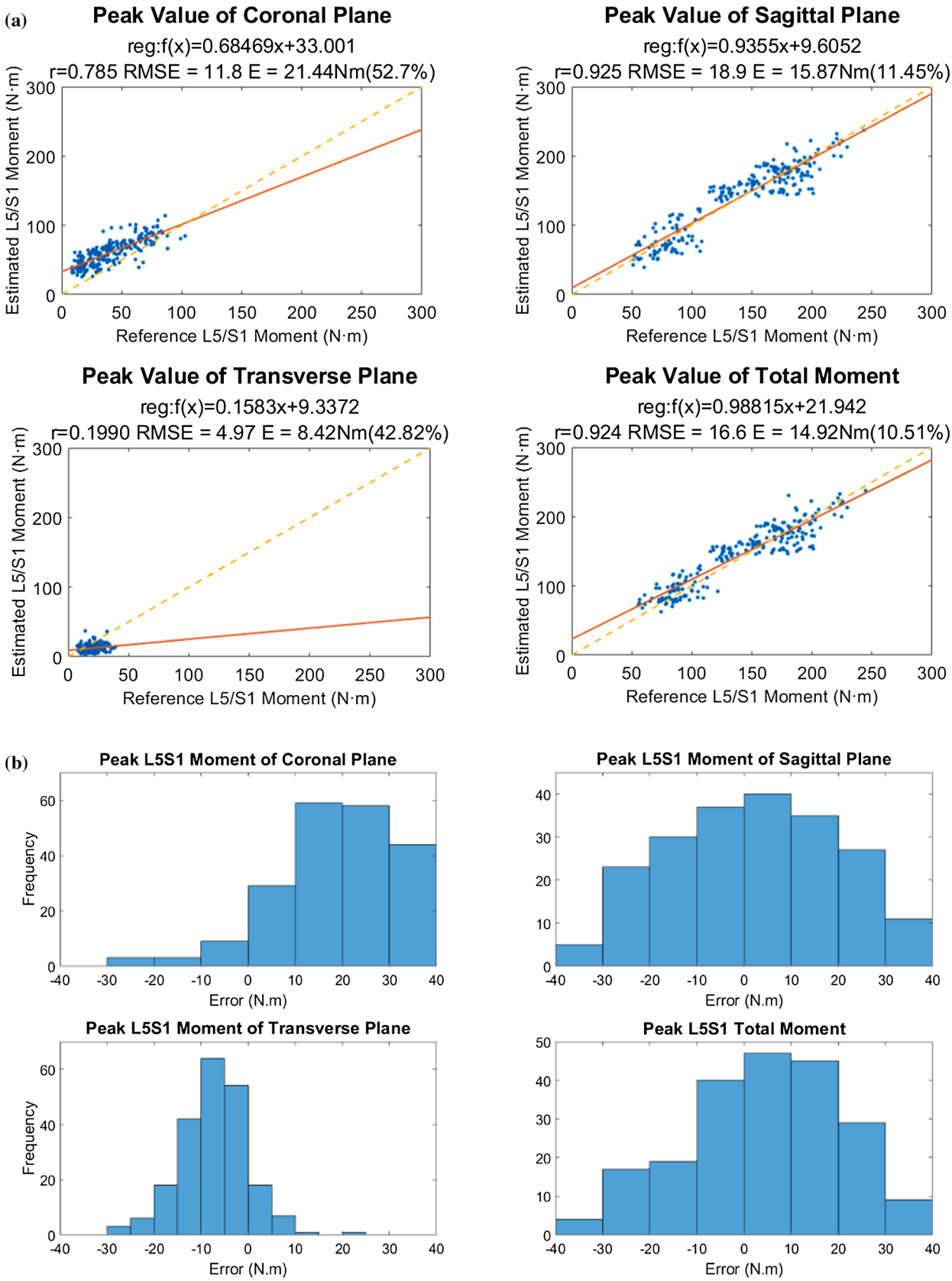

Fig. 7. (a) The comparison between the estimated peak 3D L5/S1 moment and the reference peak 3D L5/S1 moment for all lifting trials. *r* is the correlation coefficient. *RMSE* is the root-mean-square error. *E* indicates the average absolute error (AAE). *Reg* refers to the linear regression between the estimated and reference moments. The solid line is the linear regression line that generated from the data points and the dashed diagonal line is the identity line. (b) Histograms of the estimation error of the peak 3D L5/S1 moment.

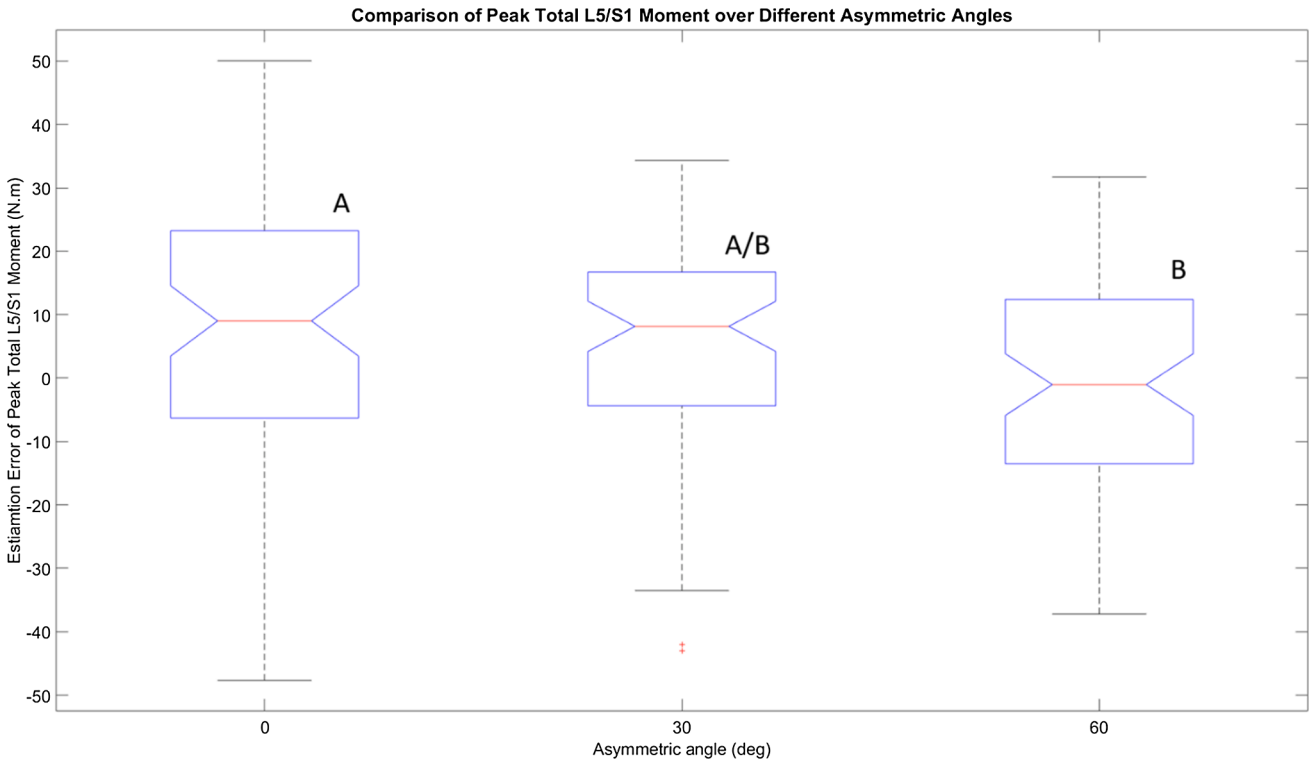

Fig. 8. Post-hoc comparison of the estimation errors of peak L5/S1 joint moment over different asymmetric lifting angles.

in turn results in errors in kinetics analysis and moment estimations. The underestimation of the peak moment in the transverse plane and the overestimation of the peak moment in the coronal plane can be partially attributed to view occlusion during the lifting tasks. In the current study, the camera was placed at the rear-right side (135 degrees from the sagittal plane). In a few lifting conditions, when the participants placed the crate on the shelf, their left forearms were blocked by the trunk and/or the right arms. Consequently, the left wrist was reconstructed at a location closer to the trunk compared to the reality, which resulted in an overestimated moment in the coronal plane. In addition, the incorrect reconstructed position of the left wrist also resulted in a smaller angular acceleration, which led to underestimated moments in the transverse plane.

There are a few limitations that need to be addressed. First, the generalizability of the proposed method should be further investigated by introducing multiple camera view angles and placement locations. Our results indicates that the peak L5/S1 moment error can be affected by the interaction between lifting conditions and the camera placement. Therefore, there might exist optimal camera view angle and placement for an individual lifting task. Second, our proposed method was validated for lifting tasks without moving the feet. In a previous study, significant errors were found on the peak L5/S1 joint moments if fictitious force was ignored when the body-centered reference frame was moving (Xu et al., 2013). Whether a computer-vision-based method can estimate joint moment for other common occupational tasks, such as pushing and pulling, where foot movements exist should be further investigated. Third, as the experiment was performed in a laboratory environment, the lifting tasks were performed within confined ranges. For example, all asymmetric lifting trials started from symmetric standing positions and were towards the right side, only two asymmetric angles were included in the experiment design, and the lateral bending angle in each task was relatively small. In addition, only one lifting weight was adopted in the current study. Such limited lifting conditions may result in less complete body motion patterns and thus limit the generalizability of the outcomes. Fourth, in our top-down inverse dynamics model, we assumed an equal weight distribution on both hands across the lifting tasks. If the weight in a crate is not well balanced, this assumption may lead to an underestimated lateral bending moment.

CRediT authorship contribution statement

Hanwen Wang: Methodology, Software, Validation. **Ziyang Xie:** Software. **Lu Lu:** Software. **Li Li:** Software. **Xu Xu:** Supervision, Methodology.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

The authors are grateful to Dr. Jacob Banks, Niall O'Brien and Amanda Rivard for assistance in data collection and data post-processing. This manuscript is based upon work supported by the National Science Foundation under Grant # 2013451.

References

Aghazadeh, F., Arjmand, N., Nasrabadi, A.M., 2020. Coupled artificial neural networks to estimate 3D whole-body posture, lumbosacral moments, and spinal loads during load-handling activities. J. Biomech. 102, 109332. https://doi.org/10.1016/j.jbiomech.2019.109332.

Asadi, F., Arjmand, N., 2020. N. Marker-less versus marker-based driven musculoskeletal models of the spine during static load-handling activities. J. Biomech. 112, 110043. https://doi.org/10.1016/j.jbiomech.2020.110043.

Azzari, G., Goulden, M.L., Rusu, R.B., 2013. Rapid characterization of vegetation structure with a microsoft kinect sensor. Sensors (Switzerland) 13 (2), 2384–2398. https://doi.org/10.3390/s130202384.

Cappozzo, A., Catani, F., Della Croce, U., Leardini, A., 1995. Position and orietnation in space of bones during movement. Clin. Biomech. 10(4), 171–178. pdf Aha.

Coenen, P., Gouttebarge, V., Van Der Burght, A.S.A.M., Van Dieën, J.H., Frings-Dresen, M.H.W., Van Der Beek, A.J., Burdorf, A., 2014. The effect of lifting during

work on low back pain: A health impact assessment based on a meta-analysis. Occup. Environ. Med. 71 (12), 871–877. https://doi.org/10.1136/oemed-2014-102346.
Coenen, Pieter, Kingma, Idsart, Boot, Cécile R.L., Faber, Gert S., Xu, Xu, Bongers, Paulien M., van Dieën, Jaap H., 2011. Estimation of low back moments from video analysis: a validation study. J. Biomech. 44 (13), 2369–2375.
Chaffin, D.B., Anderson, C.K., 1991. Occupational biomechanics. Wiley, New York, NY.
Cao, Z., Hidalgo, G., Simon, T., Wei, S. E., Sheikh, Y., 2021. OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields. IEEE Trans. Pattern Anal. Mach. Intell. 43(1), pp. 172–186. https://doi.org/10.1109/TPAMI.2019.2929257.
D'Antonio, E., Taborri, J., Palermo, E., Rossi, S., Patane, F., 2020. A markerless system for gait analysis based on OpenPose library. In: I2MTC 2020 - International Instrumentation and Measurement Technology Conference, Proceedings, 19–24. https://doi.org/10.1109/I2MTC43012.2020.9128918.
da Costa, B.R., Vieira, E.R., 2010. Risk factors for work-related musculoskeletal disorders: A systematic review of recent longitudinal studies. Am. J. Ind. Med. 53 (3), 285–323. https://doi.org/10.1002/ajim.20750.
Desjardins, P., Plamondon, A., Gagnon, M., 1998. Sensitivity analysis of segment models to estimate the net reaction moments at the L5/S1 joint in lifting. Med. Eng. Phys. 20 (2), 153–158. https://doi.org/10.1016/S1350-4533(97)00036-2.
Han, J., Shao, L., Xu, D., Shotton, J., 2013. Enhanced computer vision with microsoft kinect sensor: A eeview. IEEE Trans. Cybernetics 43 (5), 1318–1334.
He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Conference on Computer Vision and Pattern Recognition (CVPR), pages 770–778, 2016. 3.
Ioffe, S., Szegedy, C., 2005. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: International Conference on Machine Learning (ICML), pages 448–456, 2015. 3, 5.
Ionescu, C., Li, F., Sminchisescu, C., 2011. Latent structured models for human pose estimation. In: Proceedings of the IEEE International Conference on Computer Vision, 2220–2227. https://doi.org/10.1109/ICCV.2011.6126500.
Ionescu, C., Papava, D., Olaru, V., Sminchisescu, C., 2014. Human3.6M: Large scale datasets and predictive methods for 3D human sensing in natural environments. IEEE Trans. Pattern Anal. Mach. Intell. 36 (7), 1325–1339. https://doi.org/10.1109/TPAMI.2013.248.
Kingma, I., De Looze, M.P., Van Dieën, J.H., Toussaint, H.M., Adams, M.A., Baten, C.T. M., 1998. When is a lifting movement too asymmetric to identify lowback loading by 2-D analysis? Ergonomics 41 (10), 1453–1461. https://doi.org/10.1080/001401398186207.
Kingma, I., Toussaint, H.M., De Looze, M.P., Van Dieen, J.H., 1996. Segment inertial parameter evaluation in two anthropometric models by application of a dynamic linked segment model. J. Biomech. 29 (5), 693–704. https://doi.org/10.1016/0021-9290(95)00086-0.
Larivière, C., Gagnon, D., 1998. Comparison between two dynamic methods to estimate triaxial net reaction moments at the L5/S1 joint during lifting. Clin. Biomech. 13 (1), 36–47. https://doi.org/10.1016/S0268-0033(97)00021-1.
Leva, P. De., 1996. "Adjustments to Zatsiorsky-Seluyanov's segment inertia parameters. J. Biomech., 29(9), pp. 1223–1230, 1996.
Mehrizi, R., Peng, X., Xu, X., Zhang, S., Metaxas, D., Li, K., 2018. A computer vision based method for 3D posture estimation of symmetrical lifting. J. Biomech. 69, 40–46. https://doi.org/10.1016/j.jbiomech.2018.01.012.
Mehrizi, R., Peng, X., Metaxas, D.N., Xu, X., Zhang, S., Li, K., 2019. Predicting 3-D lower back joint load in lifting: A deep pose estimation approach. IEEE Trans. Hum.-Mach. Syst. 49 (1), 85–94. https://doi.org/10.1109/THMS.2018.2884811.
Nair, V., Hinton, 2010. G. E. Rectified linear units improve restricted boltzmann machines. In: International Conference on Machine Learning (ICML), pages 807–814, 2010. 3.
Pavllo, D., Feichtenhofer, C., Grangier, D., Auli, M., 2019. 3D human pose estimation in video with temporal convolutions and semi-supervised training. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2019-June, 7745–7754. https://doi.org/10.1109/CVPR.2019.00794.
Reynolds, H.M., 1982. Spatial geometry of the human pelvis/H.M. Reynolds, C.C. Snow, J.W. Young; prepared for U.S. Department of Transportation, Federal Aviation Administration, Office of Aviation Medicine. The Office; National Technical Information Service [distributor, Washington, D.C.: Springfield, Va.
Seidel, G.K., Marchinda, D.M., Dijkers, M., Soutaslittle, R.W., 1995. Hip-joint center location from palpable bony landmarks - a cadaver study. J. Biomech. 28, 995e998.
Shum, Hubert P.H., Ho, Edmond S.L., Jiang, Yang, Takagi, Shu, 2013. Real-Time Posture Reconstruction for Microsoft Kinect. IEEE Trans. Cybern. 43 (5), 1357–1369.
Skals, S., Bláfoss, R., Andersen, L.L., Andersen, M.S., de Zee, M., 2021. Manual material handling in the supermarket sector. Part 2: Knee, spine and shoulder joint reaction forces. Appl. Ergon. 92 https://doi.org/10.1016/j.apergo.2020.103345.
Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, 2014. I, and Salakhutdinov. R. (2014). Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. 15(1):1929–1958, 2014. 3.
U. S. Department of Labor, 2016. Nonfatal Occupational Injuries and Illnesses Requiring Days Away from Work. News Release: U.S Bureau of Labor Statistics. USDL-16-2130), 1–32.
Waters, T.R., Putz-Anderson, V., Garg, A., Fine, L.J., 1993. Revised NIOSH equation for the design and evaluation of manual lifting tasks. Ergonomics 36 (7), 749–776. https://doi.org/10.1080/00140139308967940.
Wu, G., Siegler, S., Allard, P., Kirtley, C., Leardini, A., Rosenbaum, D., Whittle, M., D'Lima, D.D., Cristofolini, L., Witte, H., Schmid, O., Stokes, I., 2002. ISB recommendation on definitions of joint coordinate system of various joints for the reporting of human joint motion - Part I: Ankle, hip, and spine. J. Biomech. 35 (4), 543–548. https://doi.org/10.1016/S0021-9290(01)00222-6.
Wu, G., Van Der Helm, F.C.T., Veeger, H.E.J., Makhsous, M., Van Roy, P., Anglin, C., Nagels, J., Karduna, A.R., McQuade, K., Wang, X., Werner, F.W., Buchholz, B., 2005. ISB recommendation on definitions of joint coordinate systems of various joints for the reporting of human joint motion - Part II: Shoulder, elbow, wrist and hand. J. Biomech. 38 (5), 981–992. https://doi.org/10.1016/j.jbiomech.2004.05.042.
Wu, Y., Kirillov, A., Massa, F., Lo, W.-Y., Girshick, R., 2019. Detectron2. https://github.com/facebookresearch/detectron2.
Xu, X., Chang, C.C., Faber, G.S., Kingma, I., Dennerlein, J.T., 2012. Estimating 3-D L5/S1 moments during manual lifting using a video coding system: Validity and interrater reliability. Human Factors 54 (6), 1053–1065. https://doi.org/10.1177/0018720812441945.
Xu, X., Faber, G.S., Kingma, I., Chang, C.C., Hsiang, S.M., 2013. The error of L5/S1 joint moment calculation in a body-centered non-inertial reference frame when the fictitious force is ignored. J. Biomech. 46 (11), 1943–1947. https://doi.org/10.1016/j.jbiomech.2013.05.012.
Xu, Xu, Robertson, Michelle, Chen, Karen B., Lin, Jia-hua, McGorry, Raymond W., 2017. Using the Microsoft Kinect™ to assess 3-D shoulder kinematics during computer use. Appl. Ergon. 65, 418–423. https://doi.org/10.1016/j.apergo.2017.04.004.
Yang, H., Haldeman, S., Lu, M.L., Baker, D., 2016. Low Back Pain Prevalence and Related Workplace Psychosocial Risk Factors: A Study Using Data From the 2010 National Health Interview Survey. J. Manipulative Physiol. Ther. 39 (7), 459–472. https://doi.org/10.1016/j.jmpt.2016.07.004.
Zatsiorsky, V.M., 2002. Kinetics of human motion. Human Kinetics, Champaign, IL.