Video-Based Automatic Wrist Flexion and Extension Classification

Cheng-Hsien Lee, Yu Hen Hu, *Life Fellow, IEEE*, Stephen Bao, and Robert G. Radwin, *Senior Member, IEEE*

***Abstract*—A computer vision method was developed to automatically measure wrist flexion and extension from a 2-D video for occupational health and safety research. Marker-less tracked skeletal joints of the elbow, wrist, and hand estimated the wrist flexion/extension angle between the hand and forearm. Based on the estimated angles, wrist posture was classified as flexion (palmar bending), neutral (no bending), or extension (dorsal bending) for each cycle of hand movement. Applying to a set of laboratory videos of a simulated repetitive motion task, we demonstrated the feasibility of using this algorithm for assessing the state of hand activities during manual work. Tested on 1464 frames from 61 recorded videos for 16 participants, the algorithm achieved an average performance of 72.40% correct, per-class accuracy. The sensitivity and specificity for flexion were 66.16% and 91.47%, respectively. The sensitivity and specificity for extension were 77.12% and 89.72%, respectively. This compared favorably against a previously reported consistency rate of 57% between human analyst estimates and wrist electrogoniometer measured wrist flexion/extension angles. We also applied this technique to 262 video frames of hand flexion instances selected from industrial field video data. For these videos, the average correct per-class accuracy was 76.03% in comparison to human observers. The sensitivity and specificity for flexion were 69.23% and 94.17%, respectively, and the sensitivity and specificity for extension were 91.95% and 80.57%, respectively.**

***Index Terms*—Classification, ergonomics, human factors, risk assessment, wrist posture.**

I. Introduction

WORK-related upper extremity injuries are prevalent in jobs involving repetitive and forceful exertions with awkward postures [1]. A significant portion of those claims is attributed to the hand and wrist [2]. The ability to perform quick and accurate assessment of workplace injury risk related to joint postures is important for research and injury prevention programs [3].

The joint angles are a critical factor for assessing work-related musculoskeletal risk [4]. Studies on upper extremity distal wrist disorders have garnered great importance [5]–[7]. Strenuous and rapid wrist-bending while gripping is an important risk factor for carpal tunnel syndrome. Wrist posture is one of the six variables, including intensity of exertion, duration of exertion per cycle, efforts per minute, wrist posture, speed of exertion, and duration of task per day, for estimating the strain index [8]. Research focusing on methods to assess the physical exposure to work-related musculoskeletal risk has been reported, including wrist posture [9]–[12].

Postural risk assessment methods include both subjective and objective approaches. Subjective methods, such as analyst observation and worker's self-report, have limited reliability and validity due to interrater variability [14]–[16]. Objective approaches, including direct measurements, video recording, and computer-aided analysis [13], may offer more accurate assessments [17]–[19] at expense of intrusive measurement methods [20]. A recent study [3] compared observer estimated wrist angles from video recordings to those measured using electrogoniometers. The agreement of estimated wrist angles between the two methods was 57% for flexion/extension classification using side-view video taken from the sagittal plane.

Recent research has also applied the Kinect range sensor for assessing postures. The influence of sensor view angle relative to the worker on identifying the risk level of recorded postures was studied [21]. Rapid upper limb assessment in real work conditions using Kinect skeleton data was also described [22]. However, in these works, the resolution was limited, and detailed hand joint position, such as the knuckle, was not provided. As such, the wrist flexion angle cannot be computed. In [23], a novel intelligent system for the rapid entire body assessment was applied based on convolutional pose machines. The reported deviations of the wrist angles were small with the chosen working postures. However, the experiment performed was quite different from the current work. First, in [23], the focus was on lifting. For each of the 12 lifting poses, the subject was instructed to maintain a stationary pose for 15 s. Moreover, for these 12 lifting poses, the ground truth wrist flexion angles were relatively small. Hence, there was no motion blur in the video captured. Although they did compute wrist angles, the flexion and extension of the wrist were not the focus.

Manuscript received 1 April 2021; revised 22 July 2021, 19 October 2021, and 25 February 2022; accepted 3 March 2022. Date of publication 22 April 2022; date of current version 15 September 2022. This work was supported by the National Institute for Occupational Safety and Health under Grant R01OH011024 (Radwin). *(Corresponding author: Cheng-Hsien Lee.)*

Cheng-Hsien Lee and Yu Hen Hu are with the Department of Electrical and Computer Engineering, University of Wisconsin–Madison, Madison, WI 53706 USA (e-mail: lee.chenghsien@gmail.com; yhhu@wisc.edu).

Stephen Bao is with the Washington State Department of Labor and Industries, SHARP Program, Olympia, WA 98504-4330 USA (e-mail: baos235@lni.wa.gov).

Robert G. Radwin is with the Department of Industrial and Systems Engineering and the Department of Biomedical Engineering, University of Wisconsin–Madison, Madison, WI 53706 USA (e-mail: rradwin@wisc.edu).

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, Federalwide under Application No. FWA00005399.

Color versions of one or more figures in this article are available at https://doi.org/10.1109/THMS.2022.3164776.

Digital Object Identifier 10.1109/THMS.2022.3164776

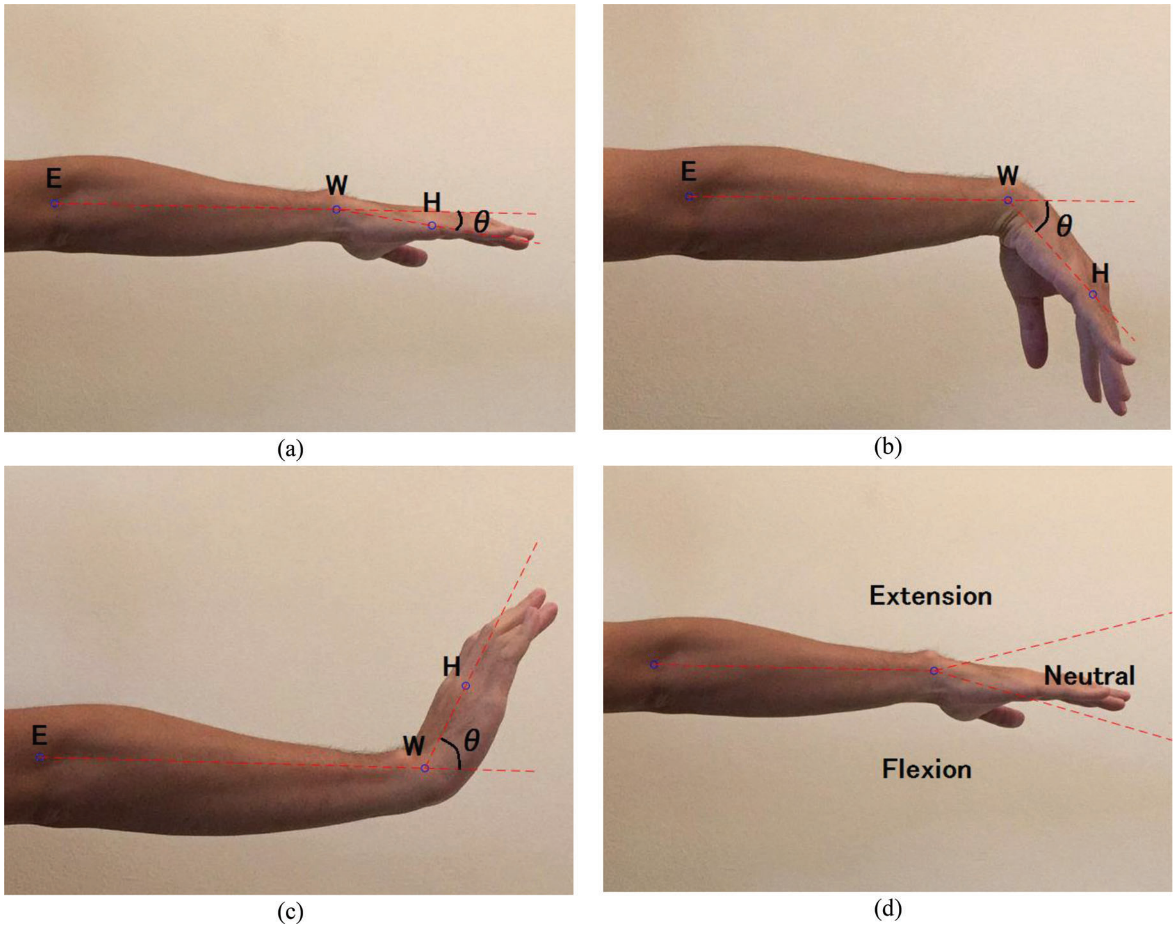

Fig. 1. Types of wrist postures. (a) Neutral. (b) Flexion. (c) Extension. (d) Ranges of the three postures.

In this study, a computer vision method is developed to automatically measure the wrist flexion/extension angle from a video recording and classify the wrist flexion/extension state using a 3-bin scale [24]. By automating the process of estimating the wrist flexion/extension angle, the method promises greater accuracy in accessing physical exposure of the upper extremity during work. Tested on 1464 frames from 61 recorded videos for 16 participants, the algorithm achieved an average performance of 72.40% correct, per-class accuracy. This compares favorably against the 57% consistency rate reported in [3] where the human observer estimated wrist flexion/extension states from video recordings are compared against those measured using electrogoniometers.

The same approach can be extended to automate the measurement of other joint angles for more accurate assessment of work-related musculoskeletal risks [4]. This automated assessment method also incurs lower cost and can be scaled up to more job types and more observations.

II. Methods

The method assumes that each video frame is viewed from a direction perpendicular to the sagittal plane of the hand. Hence, the 2-D coordinates of three upper limb joints, the elbow, wrist, and the back of the hand (knuckle) were measured using an open-source human body key point estimation package, OpenPose [25]. From these three key points, the wrist flexion angle was computed. These estimated angles were classified into three hand-posture classes: flexion, neutral, or extension. From a laboratory experiment, we collected 1464 video frames from 61 videos of controlled repetitive hand tasks performed by 16 participants. Each video frame yielded a wrist flexion angle estimate. Motion capture (MoCap) IR markers of the Optotrak motion capture system [26] were attached to the upper limb of the participants to provide ground truth measurements of the three key points. The same wrist flexion angle was estimated from these ground truth keypoint locations and similarly classified into one of the three classes as ground truth. Finally, the computer-vision hand state classification results were compared against the ground truth label. In addition to the laboratory experiment data, we also applied this algorithm to a set of videos acquired from a field study (the SHARP study [27]).

A. *Wrist Flexion/Extension Angles*

Referring to Fig. 1, when the palm faces downward, the angle between the elbow (E), wrist (W), and the knuckle of the middle finger on the back of the hand (H) may be used to define three states of hand postures: flexion, neutral, and extension.

Fig. 2. Three tracked points and the hand posture angle θ.

These classifications are determined by the wrist flexion/extension angle θ. Assume the vector $\overrightarrow{EW}$ points to the positive direction of the reference axis. θ is the angle between the vector $\overrightarrow{WH}$ and the reference vector $\overrightarrow{EW}$. If the hand bends toward the back of the hand, θ is positive. If the hand bends toward the palm, θ is negative. The magnitude of θ should be smaller than 90°. For example, the angle θ in Fig. 1(a) is approximately zero, Fig. 1(b) is negative, and Fig. 1(c) is positive (or extension).

Based on widely used posture observational methods (e.g., [15]), we define the wrist posture as neutral if θ is between –15° to 15°. If $\theta > 15^\circ$, the hand posture state is extension. If $\theta < -15^\circ$, the hand posture state is flexion.

B. Wrist Flexion Angle Estimation Algorithm

Denote the image coordinates of three manually selected landmark points H, W, and F as (x_h, y_h), (x_w, y_w), and (x_E, y_E), respectively. The angle θ is the angle between line segments $\overrightarrow{WH}$ and $\overrightarrow{EW}$. An example of the three landmark points is shown in Fig. 2.

Denote *x*–*y* image coordinates of vectors $\overrightarrow{WH}$ and $\overrightarrow{EW}$ as

$$\overrightarrow{WH} = (x_h - x_w, y_h - y_w) \tag{1}$$

and

$$\overrightarrow{EW} = (x_w - x_e, y_w - y_e). \tag{2}$$

Using trigonometry, one has

$$|\theta| = \cos^{-1}\left(\frac{\overrightarrow{EW} \cdot \overrightarrow{WH}}{\overrightarrow{EW}\overrightarrow{WH}}\right). \tag{3}$$

θ is negative if the rotation is toward the palm direction (flexion) and positive otherwise (extension).

C. Skeletal Joint Position Estimation

To estimate skeletal joint locations in the frame of the elbow (E), the wrist (W), and the knuckle on the back of the hand (H), an open-source 2-D human pose estimation software package Openpose [25] was applied. Openpose takes a video frame as the input to a two-branch convolutional neural network and predicts a set of 2-D confidence maps of body part locations and a set of 2-D vector fields of part affinities simultaneously. A greedy inference is then applied to parse the confidence map and affinity field to output 2-D coordinates of 25 key points corresponding to 25 body skeletal joints of the subject. In this work, we use the key point locations of the elbow, wrist, and hand (knuckle) of one hand. The OpenPose algorithm can automatically start when presented with a video clip and will estimate key point locations for each video frame. The accuracy of the joint position estimation is compromised if part of the limb is occluded. In this laboratory experiment, the camera is placed facing the sagittal plane to minimize occlusion. However, for videos taken from factories, manual screening of the videos is required to select segments of frames where the arm and hand movements are always visible. A diagram of the proposed wrist flexion angle estimation algorithm is summarized in Fig. 3.

III. Experiment

A. Data Collection of Laboratory Videos

The dataset consists of 61 videos from 16 participants (3 males and 13 females). These videos were selected among videos recorded in the laboratory of a simulated repetitive motion task [28]. The subjects were recruited with informed consent and IRB approval. In each video, the participant stands in front of an apparatus and moves tennis balls from one location to another for 15 cycles, as demonstrated in Figs. 2 and 4.

The apparatus (see Fig. 4) is comprised of an 840-mm travel length linear belt drive actuator (Misumi MSS-625) driven by a bipolar stepper motor (ElectroCraft Model TPP34 with 560-N·cm torque) and controlled by a stepper motor controller (IMS MX-CS101-401). The device can move a 2-kg object every 0.5 s across the actuator length of travel.

The hands were videotaped from the side view (sagittal plane) with the palm of the right-hand facing downward. The selected videos included two kinds of tasks, moving the tennis balls from location A to 1 and from location A to 2 (see Fig. 4). Videos were selected in which the hand of the subject was always in the sagittal plane (location from A to 1 and A to 2).

An Optotrak 3-D infrared motion tracking system (MoCap) was used to track IR beacons attached to the hands and arms of the subject, as shown in Fig. 2. The MoCap system returns accurate estimates of 3-D coordinates of the skeletal joints of the elbow, wrist, and hand based on the tracked trajectory of these markers. These 3-D coordinates are specified relative to a world coordinate system whose *x*–*y* plane is not guaranteed to be parallel to the sagittal plane of the subject. To remedy this problem, we choose an initial video frame before the subject started moving the hand for camera calibration. We used six pairs of 3-D coordinates from the MoCap system and corresponding 2-D image coordinates to align the MoCap system relative to the video camera. The MoCap estimated 3-D joint coordinates of the elbow, wrist, and hand were projected to give corresponding 2-D image coordinates. The same formulas in (1)–(3) were used to calculate the ground truth hand-pose angles.

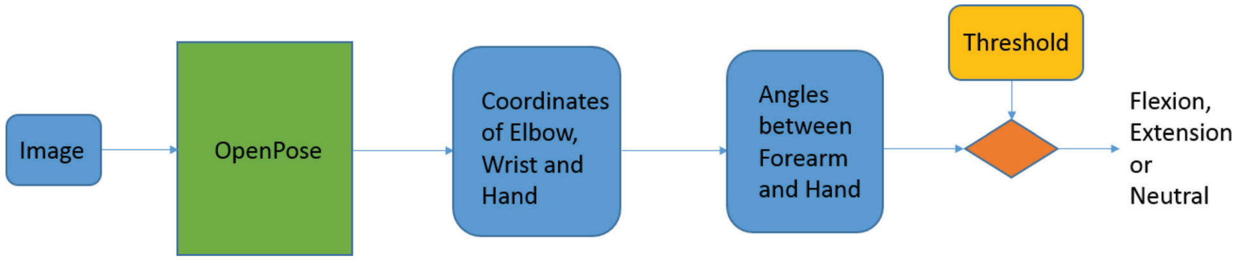

Fig. 3. Diagram of our method.

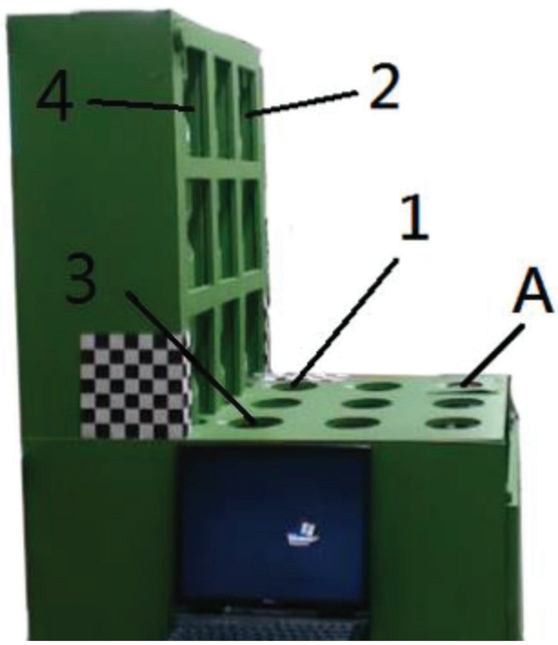

Fig. 4. Laboratory task simulation apparatus. The participant grasps a ball from location A and moves it to either locations C or 2, depending on the specified task.

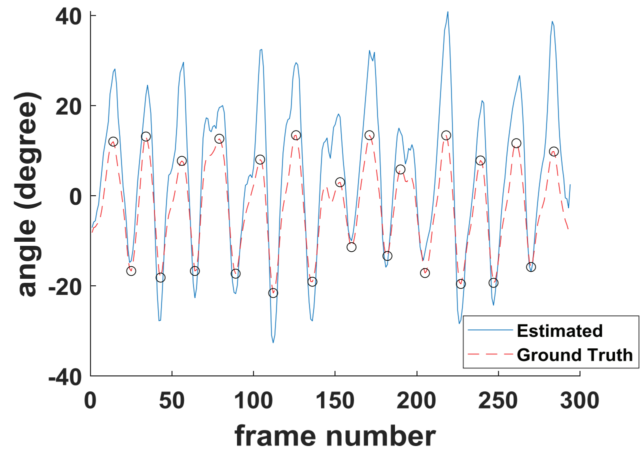

Fig. 5. Angle prediction by the algorithm.

Although the hand flexion angles are estimated for each video frame, the hand state needs to be estimated only at *target frames* corresponding to the local maximum and the local minimum ground truth angles. These target frames where the extreme postures occurred were marked with circles in Fig. 5. There were a total of 1464 target frames extracted from 61 laboratory videos. The ground truth angles were obtained from MoCap measurements and the estimated angles were obtained from the videos using our computer vision algorithm.

B. Synchronization for Laboratory Videos

Because the initialization of the MoCap system and the video recorder were independent, there was a time offset between the recorded signals. This is shown in Fig. 6(a) where the solid line and the dashed line represent the MoCap (ground truth) and our algorithm estimated angles, respectively. While these signals have the same frequency, the blue curve (estimated angle) lags behind the red dotted curve (ground truth angel) by about four cycles. To synchronize these two signals, we evaluated the cross-correlation of them and search for the position D of the global maximum of the cross-correlation coefficient. This is shown in Fig. 6(b). D is the estimated timing offset (in unit of frames) between the ground truth and the estimated angles. The ground truth angle curve then will be shifted by D frames so that it is aligned temporally with the video observation, as shown in Fig. 6(c). The relationships between the two signals before and after shifting are shown in Fig. 6(d).

C. Industrial Field Videos

In addition to the videos recorded in the laboratory, we also used industry field videos recorded on-site in the State of Washington Department of Labor and Industries SHARP [27] project. Four kinds of tasks performed by different workers were video-recorded. The tasks included a laundry handler in a commercial laundry facility, a lumber handler in a sawmill, an assembler in an electronics plant, and a pharmacist in a large hospital pharmacy. These jobs had large variations in work posture, and each of them was performed in multiple locations in the same facility. Each job was recorded from two different angles using two synchronized camcorders. A time-sample observation method was used in which human analysts observed the hand postures at preselected frames during a task. Custom data processing software was used for posture estimation by analysts to obtain a continuous angular scale. The software showed the preselected

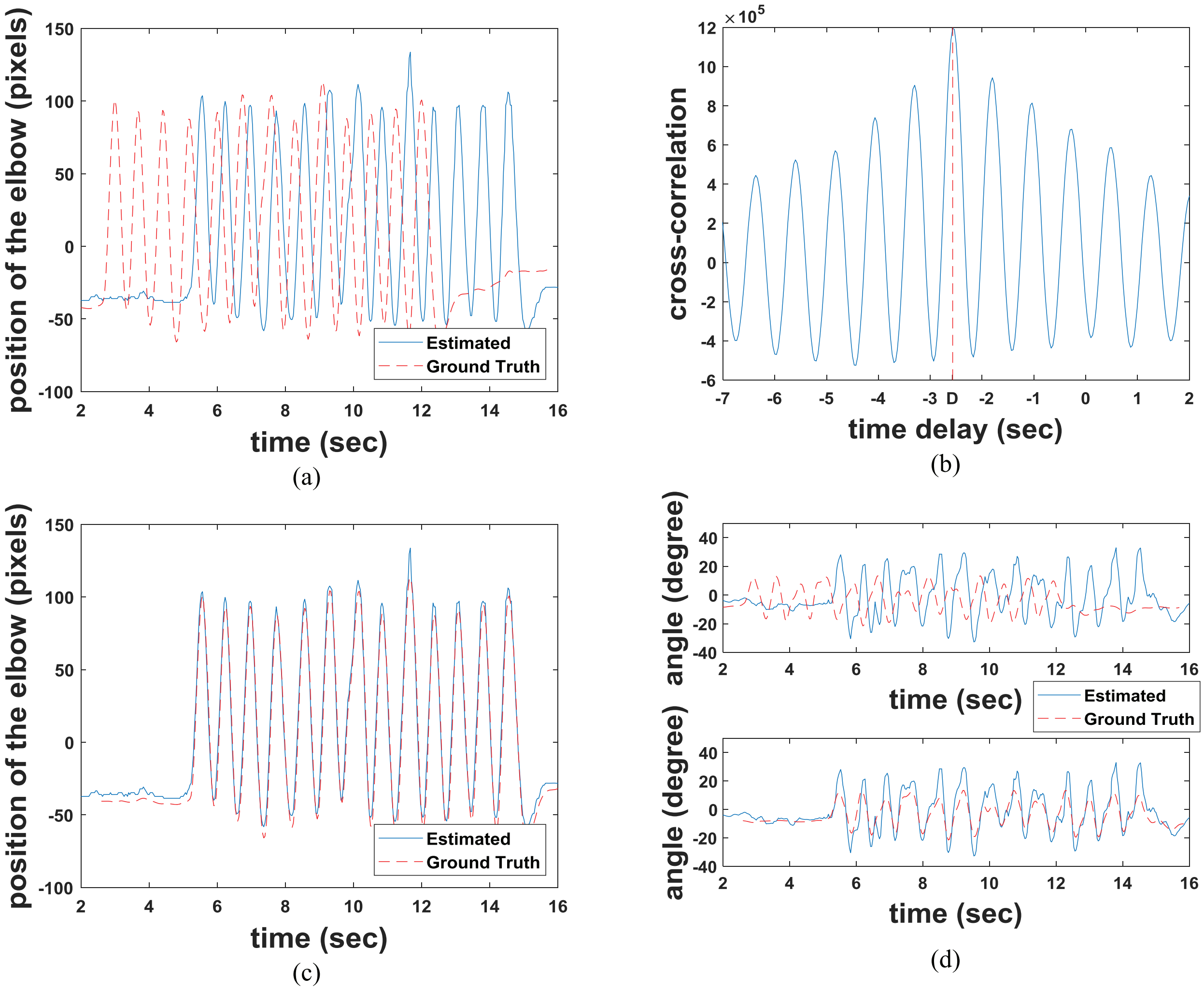

Fig. 6. (a) Position-time curves of estimated and ground truth data. (b) Cross-correlation of estimated and ground truth signals. (c) Aligned position-time curves of estimated and ground truth data. (d) (Top) Angle-time curves of estimated and ground truth data; (Bottom) Two curves are aligned after shifting one signal by the time delay D.

frame from two camera angles and asked the analyst to reproduce the wrist angle they see by clicking on the posture diagram.

We applied our algorithm to the same preselected frames and compared the posture estimate results. Due to the low resolution of the videos from the field studies, sometimes our algorithm failed to detect a correct key-point location. We manually eliminated the frames in which the estimated joint locations were not detected. There were 262 side-view video frames selected. We considered the angles estimated observation by analysts as the ground truth and the angles estimated by our algorithm as the prediction.

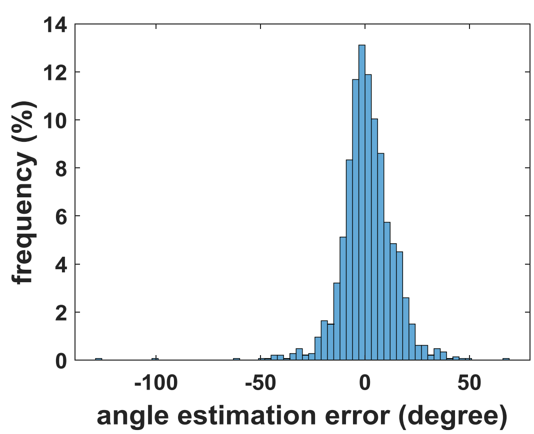

Fig. 7. Histogram of the angle estimation errors for laboratory videos.

IV. Results

A. Angle Estimation Error

We defined an angle estimation error as the difference between the predicted angle and the ground truth (MoCap estimated, or human analysts estimated) angle. In Fig. 7, a histogram of the angle estimation errors for the laboratory videos is plotted. The corresponding histogram of angle estimation errors for the industrial videos is plotted in Fig. 8. The root-mean-square angle estimation error is the standard deviation of these distributions. In the case of laboratory videos, the rms angle estimation error

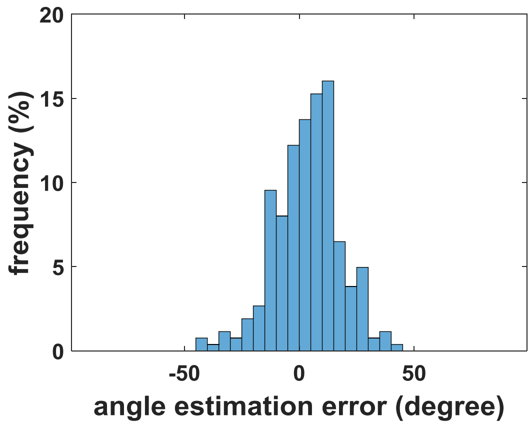

Fig. 8. Histogram of the angle estimation errors for industrial videos.

TABLE I
CONFUSION MATRIX FOR LABORATORY VIDEOS

Ground Truth \ Prediction	$\theta < -15^\circ$ Flexion	$15^\circ > \theta \geq -15^\circ$ Neutral	$\theta \geq 15^\circ$ Extension
$\theta < -15^\circ$ Flexion	**131 (8.95%)**	**67 (4.58%)**	**0 (0.00%)**
$15^\circ > \theta \geq -15^\circ$ Neutral	**105 (7.17%)**	**587 (40.10%)**	**102 (6.97%)**
$\theta \geq 15^\circ$ Extension	**3 (0.20%)**	**105 (7.17%)**	**364 (24.86%)**

was 12.64°. For the industrial videos, the rms angle estimation error was 14.70°.

B. Averaged Per-Class Classification Rate

The confusion matrix for classifying the 1464 target frames from the 61 laboratory videos is shown in Table I. The (*i,j*)th element of the confusion matrix, C_{ij} is the number of samples (video frames) that belong to the (ground truth) *i*th class but were classified by our algorithm as the *j*th class. Thus, the overall classification rate is calculated as

$$r = \sum_{i=1}^{3} C_{ii} \Big/ \sum_{i=1}^{3} \sum_{j=1}^{3} C_{ij} = \sum_{i=1}^{3} C_{ii} \Big/ N$$

where $N = 1464$ is the total number of samples used. The entries in the parentheses below C_{ij} are C_{ij}/N (see Table I).

Using the confusion matrix, one may also define a per-class classification accuracy

$$r_i = C_{ii} \Big/ \sum_{j=1}^{3} C_{ij} \quad i = 1, 2, 3$$

and averaged per-class classification rate

$$r_{APC} = (r_1 + r_2 + r_3)/3 \tag{4}$$

TABLE II
CONFUSION MATRIX FOR INDUSTRIAL VIDEOS

Ground Truth \ Prediction	$\theta < -15^\circ$ Flexion	$15^\circ > \theta \geq -15^\circ$ Neutral	$\theta \geq 15^\circ$ Extension
$\theta < -15^\circ$ Flexion	**27 (10.31%)**	**11 (4.20%)**	**1 (0.38%)**
$15^\circ > \theta \geq -15^\circ$ Neutral	**12 (4.58%)**	**91 (34.73%)**	**33 (12.60%)**
$\theta \geq 15^\circ$ Extension	**1 (0.38%)**	**6 (2.29%)**	**80 (30.53%)**

TABLE III
SENSITIVITY AND SPECIFICITY FOR LABORATORY VIDEOS

Class	Flexion	Neutral	Extension
Sensitivity	66.16%	73.93%	77.12%
Specificity	75.12%	73.88%	72.38%

r_{APC} tends to compensate for the imbalanced distributions of samples in different classes.

For laboratory videos corresponding to Table I, the average per-class accuracy was 72.40%. For industrial videos, the corresponding confusion matrix is shown in Table II. The average per-class accuracy was 76.03%.

C. Per-Class Sensitivity and Specificity

For each class (i = 1, 2, 3), we may convert the confusion matrix into a 2×2 matrix. For example, let $i = 1$, the confusion matrix in Table I can be converted into

$$\begin{bmatrix} TP_1 & FN_1 \\ FP_1 & TN_1 \end{bmatrix} = \begin{bmatrix} C_{11} & C_{12} + C_{13} \\ C_{21} + C_{31} & C_{22} + C_{23} + C_{32} + C_{33} \end{bmatrix} = \begin{bmatrix} 131 & 67 \\ 108 & 1158 \end{bmatrix}.$$

Then, the sensitivity and specificity for class i may be defined as

$$\text{Sen}_i = TP_i/TP_i + FN_i = C_{ii} \Big/ \sum_{j=1}^{3} C_{ij}$$

$$\text{Spe}_i = TN_i/TN_i + FP_i = \sum_{m \neq i} \sum_{n \neq i} C_{mn} \Big/ \sum_{m \neq i} \sum_{n=1}^{3} C_{mn}$$

The per-class sensitivity and specificity for the laboratory videos are listed in Table III.

The per-class sensitivity and specificity for the SHARP industrial videos are listed in Table IV.

TABLE IV
SENSITIVITY AND SPECIFICITY FOR INDUSTRIAL VIDEOS

Class	Flexion	Neutral	Extension
Sensitivity	69.23%	66.91%	91.95%
Specificity	76.68%	84.92%	67.43%

D. Nonsagittal Plane Tasks

The direction of the video viewpoint in this work was perpendicular to the sagittal plane of the subject. We also conducted experiments in which the camera viewpoint was not directly perpendicular to the plane of hand movement. This allowed us to investigate the difference between 2-D and 3-D angles estimated. We collected 1167 frames from 43 recorded videos for 15 participants (2 males and 13 females). As shown in Fig. 4, two kinds of tasks, moving a tennis ball from location A to 3 and from location A to 4 (nonsagittal plane tasks), were studied. As discussed earlier, for the sagittal plane tasks, the rms angle estimation error was 12.64° and the average per-class accuracy was 72.60%. For the nonsagittal plane task, the rms angle estimation error was 23.25° and the average per-class accuracy was 74.26%. Although the nonsagittal plane task had a larger angle estimation error, the per-class accuracy was slightly better.

V. DISCUSSION

In this work, we present a computer-vision-based method, utilizing keypoint detection, to measure the joint angle of the hand posture. This measurement can be used as an objective indicator to assess the physical exposure of workers in a factory. Our posture estimation measurements were automatic, therefore more objective and consistent.

From Table I, note that there are $C_{21} = 105$ (7.17%) neural wrist angles incorrectly classified by our algorithm for the flexion. As shown in Fig. 5, the computer vision estimated angles seem to be larger than those estimated using MoCap. This discrepancy may be due to the misalignment of the joint positions estimated by OpenPose and by the MoCap sensors. Reducing these estimation discrepancies may in turn reduce the angle classification error. In Table II, note that there are 33 cases (12.6% of the samples) of neutral wrist angles misclassified as extensions. However, since the ground truth was obtained from human observers, we attribute this discrepancy to human subjective judgment variations.

Applying the proposed method in the real work environment might be promising. Not limited to the hand posture, our method can also be extended to the application of measuring neck, shoulder, and elbow postures. For example, we can use the three landmarks: ear, the center of shoulders, and the center of hips detected by Openpose to replace the hand, wrist, and elbow in this experiment. The new three landmarks viewed from the sagittal plane of the subject will form the angle of the neck.

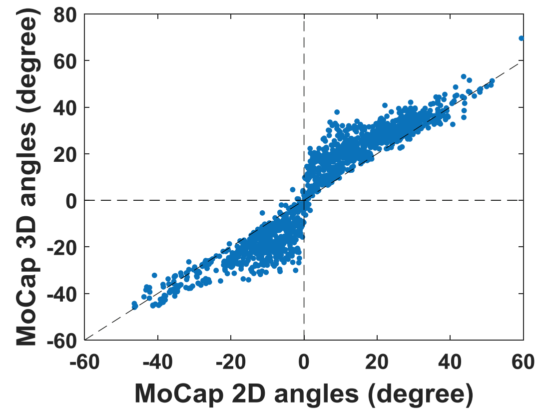

Fig. 9. MoCap estimated 3-D angles versus 2-D angles.

Due to the limitation of single-view videos, the assumption of the side view is necessary. In future work, depth image videos might be applied to remove this restriction.

A study in [3] compared the results between the video analysis and electrogoniometer method. They classified the wrist-flexion-extension posture into five categories: $\theta \geq 45°$, $45° > \theta \geq 15°$, $15° > \theta \geq -15°$, $-15° > \theta \geq -45°$, and $\theta < -45°$, and achieve 57% agreement. For comparison, we use the same five-class categories, and the resulting confusion matrix is shown in Table V. The agreement between MoCap and the computer vision method is 70.42%, which is 13% higher than the agreement between the electrogoniometer and observation method. Since the electrogoniometer provided 3-D angles ground truth, we also compare the angles computed by original 3-D coordinates of MoCap with the 2-D angles of the computer vision method, and the confusion matrix is shown in Table VI. The overall agreement is 56.97%, which is consistent with the results in [3]. With similar performance, our method saves much time and labor since it is automatic.

The effect of viewing angle on wrist posture estimation has been addressed in [29]. A scatter plot of the angle computed using the 3-D coordinates of MoCap versus the angle computed using the transformed 2-D image coordinates is represented in Fig. 9. We observe that the 3-D angle is larger than the 2-D angle in most cases. Furthermore, a large range of 3-D angles possibly appear as 0° in the 2-D image. This observation may be explained using sketches in Fig. 10. Note that due to projection, $\angle EW_{2D}H$ will be larger than the $\angle EW_{3D}H$. The flexion angle defined in this work (c.f. Figs. 1 and 2) is the complementary angle of the $\angle EWH$. $\theta = 180° - \angle EWH$. Hence, $\theta_{2D} \leq \theta_{3D}$. Although we assume the camera is photographing from the side view, this phenomenon is mainly due to the supination and pronation of the wrist. It is also a main source of error.

We conclude that using a computer-vision-based method to estimate wrist flexion/extension provides a quick assessment and requires no equipment attached to the subject. This nonintrusive, automated observation method allows long-term, large-scale collection of exposure data in longitudinal studies in the future. Combined with health outcomes, such studies will provide data-centric validation of exposure models developed to prevent job-related hand injuries.

TABLE V
CONFUSION MATRIX FOR LABORATORY VIDEOS WITH FIVE CATEGORIES

Ground Truth

$\theta < -45^{\circ}$ Flexion	**0 (0.00%)**	**3 (0.20%)**	**0 (0.00%)**	**0 (0.00%)**	**0 (0.00%)**
$-15^{\circ} > \theta \geq -45^{\circ}$ Flexion	**5 (0.34%)**	**123 (8.40%)**	**67 (4.58%)**	**0 (0.00%)**	**0 (0.00%)**
$15^{\circ} > \theta \geq -15^{\circ}$ Neutral	**3 (0.20%)**	**102 (6.97%)**	**587 (40.10%)**	**100 (6.83%)**	**2 (0.14%)**
$45^{\circ} > \theta \geq 15^{\circ}$ Extension	**1 (0.07%)**	**2 (0.14%)**	**104 (7.10%)**	**317 (21.65%)**	**37 (2.53%)**
$\theta \geq 45^{\circ}$ Extension	**0 (0.00%)**	**0 (0.00%)**	**1 (0.07%)**	**6 (0.41%)**	**4 (0.27%)**
	$\theta < -45^{\circ}$ Flexion	$-15^{\circ} > \theta \geq -45^{\circ}$ Flexion	$15^{\circ} > \theta \geq -15^{\circ}$ Neutral	$45^{\circ} > \theta \geq 15^{\circ}$ Extension	$\theta \geq 45^{\circ}$ Extension

Prediction

TABLE VI
CONFUSION MATRIX FOR LABORATORY VIDEOS WITH FIVE CATEGORIES AND 3-D MOCAP ANGLES

Ground Truth

$\theta < -45^{\circ}$ Flexion	**2 (0.14%)**	**3 (0.20%)**	**0 (0.00%)**	**0 (0.00%)**	**0 (0.00%)**
$-15^{\circ} > \theta \geq -45^{\circ}$ Flexion	**5 (0.34%)**	**191 (13.05%)**	**217 (14.82%)**	**7 (0.48%)**	**0 (0.00%)**
$15^{\circ} > \theta \geq -15^{\circ}$ Neutral	**1 (0.07%)**	**30 (2.05%)**	**272 (18.58%)**	**40 (2.73%)**	**2 (0.14%)**
$45^{\circ} > \theta \geq 15^{\circ}$ Extension	**1 (0.07%)**	**6 (0.41%)**	**268 (18.31%)**	**366 (25.00%)**	**38 (2.60%)**
$\theta \geq 45^{\circ}$ Extension	**0 (0.00%)**	**0 (0.00%)**	**2 (0.14%)**	**10 (0.68%)**	**3 (0.20%)**
	$\theta < -45^{\circ}$ Flexion	$-15^{\circ} > \theta \geq -45^{\circ}$ Flexion	$15^{\circ} > \theta \geq -15^{\circ}$ Neutral	$45^{\circ} > \theta \geq 15^{\circ}$ Extension	$\theta \geq 45^{\circ}$ Extension

Prediction

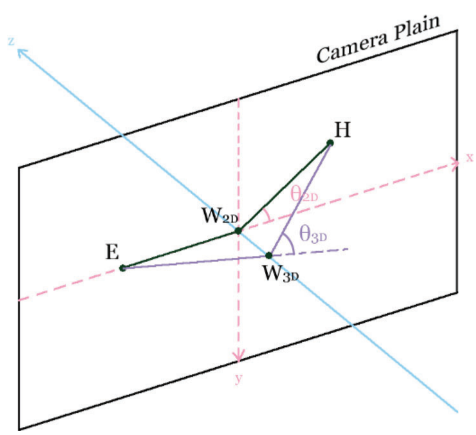

Fig. 10. Comparison between the 3-D and 2-D angles. The 3-D angle is larger than the 2-D angle in this case.

REFERENCES

[1] J. A. Barondess, "Epidemiologic evidence," in *Musculoskeletal Disorders and the Workplace: Low Back and Upper Extremities*. Washington, DC, USA: Nat. Acad. Sci., 2001, ch. 4, pp. 85–183.

[2] "By the numbers: 2017 WSIB (Ontario) statistical report," WSIB, Toronto, ON, Canada, 2017.

[3] C. D. McKinnon, S. Ehmke, A. M. Kociolek, J. P. Callaghan, and P. J. Keir, "Wrist posture estimation differences and reliability between video analysis and electrogoniometer methods," *Hum. Factors*, vol. 63, pp. 1284–1294, 2020.

[4] B. Juul-Kristensen, N. Fallentin, and C. Ekdahl, "Criteria for classification of posture in repetitive work by observation methods: A review," *Int. J. Ind. Ergonom.*, vol. 19, no. 5, pp. 397–411, 1997.

[5] A. M. Dale *et al.*, "Prevalence and incidence of carpal tunnel syndrome in US working populations: Pooled analysis of six prospective studies," *Scand. J. Work, Environ. Health*, vol. 39, no. 5, 2013, Art. no. 495.

[6] J. M. Kapellusch *et al.*, "Exposure-response relationships for the ACGIH threshold limit value for hand-activity level: Results from a pooled data study of carpal tunnel syndrome," *Scand. J. Work, Environ. Health*, vol. 40, no. 6, 2014, Art. no. 610.

[7] C. Harris-Adamson *et al.*, "Biomechanical risk factors for carpal tunnel syndrome: A pooled study of 2474 workers," *Occupat. Environ. Med.*, vol. 72, no. 1, pp. 33–41, 2015.

[8] J. S. Moore and A. Garg, "The strain index: A proposed method to analyze jobs for risk of distal upper extremity disorders," *Amer. Ind. Hyg. Assoc. J.*, vol. 56, no. 5, pp. 443–458, 1995.

[9] B. A. Silverstein, L. J. Fine, and T. J. Armstrong, "Occupational factors and carpal tunnel syndrome," *Amer. J. Ind. Med.*, vol. 11, no. 3, pp. 343–358, 1987.

[10] L. Punnett and D. H. Wegman, "Work-related musculoskeletal disorders: The epidemiologic evidence and the debate," *J. Electromyography Kinesiology*, vol. 14, no. 1, pp. 13–23, 2004.

[11] G. C. David, "Ergonomic methods for assessing exposure to risk factors for work-related musculoskeletal disorders," *Occupat. Med.*, vol. 55, no. 3, pp. 190–199, 2005.

[12] G. David, V. Woods, G. Li, and P. Buckle, "The development of the quick exposure check (QEC) for assessing exposure to risk factors for work-related musculoskeletal disorders," *Appl. Ergonom.*, vol. 39, no. 1, pp. 57–69, 2008.

[13] G. Li and P. Buckle, "Current techniques for assessing physical exposure to work-related musculoskeletal risks, with emphasis on posture-based methods," *Ergonomics*, vol. 42, no. 5, pp. 674–695, 1999.

[14] R. J. Shephard, "Limits to the measurement of habitual physical activity by questionnaires," *Brit. J. Sports Med.*, vol. 37, no. 3, pp. 197–206, 2003.

[15] S. Bao, N. Howard, P. Spielholz, B. Silverstein, and N. Polissar, "Inter-rater reliability of posture observations," *Hum. Factors*, vol. 51, no. 3, pp. 292–309, 2009.

[16] W. Lee, J.-H. Lin, and S. Bao, "Inter-rater reliability of an inertial measurement unit sensor-based posture-matching method: A pilot study," *Int. J. Ind. Ergonom.*, vol. 80, 2020, Art. no. 103025.

[17] B. Juul-Kristensen, G. Å. Hansson, N. Fallentin, J. H. Andersen, and C. Ekdahl, "Assessment of work postures and movements using a video-based observation method and direct technical measurements," *Appl. Ergonom.*, vol. 32, no. 5, pp. 517–524, 2001.

[18] A. Hua *et al.*, "Evaluation of machine learning models for classifying upper extremity exercises using inertial measurement unit-based kinematic data," *IEEE J. Biomed. Health Inform.*, vol. 24, no. 9, pp. 2452–2460, Sep. 2020.

[19] W. Yang, D. Yang, Y. Liu, and H. Liu, "Decoding simultaneous multi-DOF wrist movements from raw EMG signals using a convolutional neural network," *IEEE Trans. Human-Mach. Syst.*, vol. 49, no. 5, pp. 411–420, Oct. 2019.

[20] M. C. Schall, Jr., R. F. Sesek, and L. A. Cavuoto, "Barriers to the adoption of wearable sensors in the workplace: A survey of occupational safety and health professionals," *Hum. Factors*, vol. 60, no. 3, pp. 351–362, 2018.

[21] J. A. Diego-Mas and J. Alcaide-Marzal, "Using Kinect^{TM} sensor in observational methods for assessing postures at work," *Appl. Ergonom.*, vol. 45, no. 4, pp. 976–985, 2014.

[22] P. Plantard, H. P. Shum, A.-S. L. Pierres, and F. Multon, "Validation of an ergonomic assessment method using Kinect data in real workplace conditions," *Appl. Ergonom.*, vol. 65, pp. 562–569, 2017.

[23] Z. Li, R. Zhang, C.-H. Lee, and Y.-C. Lee, "An evaluation of posture recognition based on intelligent rapid entire body assessment system for determining musculoskeletal disorders," *Sensors*, vol. 20, no. 16, 2020, Art. no. 4414.

[24] A. M. Kociolek and P. J. Keir, "Reliability of distal upper extremity posture matching using slow-motion and frame-by-frame video methods," *Hum. Factors*, vol. 52, no. 3, pp. 441–455, 2010.

[25] Z. Cao, G. H. Martinez, T. Simon, S. Wei, and Y. A. Sheikh, "OpenPose: Realtime multi-person 2D pose estimation using part affinity fields," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 1, pp. 172–186, Jan. 2021.

[26] J. Schmidt, D. R. Berg, H.-L. Ploeg, L. Ploeg, J. Schmidt, and D. R. Berg, "Precision, repeatability and accuracy of Optotrak optical motion tracking systems," *Int. J. Exp. Comput. Biomech.*, vol. 1, no. 1, pp. 114–127, 2009.

[27] S. Bao, B. Silverstein, N. Howard, and P. Spielholz, "The Washington state SHARP approach to exposure assessment," *Fundam. Assessment Tools Occupat. Ergonom.*, vol. 44, no. 1, pp. 22–44, 2006.

[28] O. Akkas, C.-H. Lee, Y. H. Hu, T. Y. Yen, and R. G. Radwin, "Measuring elemental time and duty cycle using automated video processing," *Ergonomics*, vol. 59, no. 11, pp. 1514–1525, 2016.

[29] M. H. Lau and T. J. Armstrong, "The effect of viewing angle on wrist posture estimation from photographic images using novice raters," *Appl. Ergonom.*, vol. 42, no. 5, pp. 634–643, 2011.