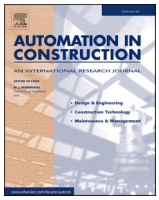
Automation in Construction 164 (2024) 105452

Contents lists available at ScienceDirect

Automation in Construction

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



3D pose estimation dataset and deep learning-based ergonomic risk assessment in construction

Chao Fan [a], Qipei Mei [b], Xinming Li [a,*]

[a] Department of Mechanical Engineering, University of Alberta, 116 St NW, Edmonton, Alberta T6G 2E1, Canada
[b] Department of Civil Engineering, University of Alberta, 116 St NW, Edmonton, Alberta T6G 2E1, Canada

ARTICLE INFO

Keywords:
3D human pose estimation
Motion capture data
Computer vision
Construction safety
Ergonomic risk assessment
REBA
RULA

ABSTRACT

Pose estimation of construction workers is critical to ensuring safe construction and protecting construction workers from ergonomic risks. Computer vision (CV)-based 3D pose estimation for construction workers is increasingly used in ergonomic risk assessment (ERA) due to its considerable practicability and accuracy. Currently, the deficiencies of (1) dedicated datasets for construction activities and (2) informative 3D biomechanical models to both Rapid Entire Body Analysis (REBA) and Rapid Upper Limb Analysis (RULA) impede the performance of CV-based ERA in construction sectors. Therefore, this study introduces a deep learning-based ERA by introducing a new dataset, ConstructionPose3D (CP3D), that follows a proposed 3D biomechanical skeletal model to support REBA and RULA. This dataset contains approximately 421,000 accurate 3D poses and annotations for construction activities. The results indicate that the proposed deep learning ERA models trained with CP3D outperform those without CP3D in accurately estimating the poses of construction workers, leading to improved ERA.

1. Introduction

Timely identification and response to workplace risks are essential to ensure workplace safety, health, and productivity [1,2]. Nearly 80% of worksite injuries are caused by unsafe operations [3]. Meanwhile, construction workers are often exposed to operations that can lead to forceful exertions, repetitive movements, and awkward body postures that often have an imperceptible but detrimental effect on their health [4]. The negative impact of these operations often leads to work-related musculoskeletal disorders (WMSDs) [5,6]. The Association of Workers' Compensation Boards of Canada reports that the manufacturing and construction sectors had the second and third-highest number of lost-time injury claims in 2021, accounting for 13.6% and 10.4%, respectively [7]. The report also shows that the construction and manufacturing sectors had the highest numbers of fatalities in 2021, accounting for 19.6% and 16.7%, respectively [7]. The United States Bureau of Labor Statistics indicates that 30% of occupational injuries and illnesses were WMSDs in 2018 [8]. Therefore, proactive identification and prevention of WMSDs and health risks are profoundly beneficial.

Traditional methods for worker safety and health management rely on human observation, self-reporting, and direct measurement [9–12], which are inevitably subjective, error-prone, invasive, and time-consuming [4,13]. For example, there are studies involving the direct measurement of workers through objective and responsive inertial measurement units (IMUs). Clearly, this invasive approach of equipping workers with additional devices amplifies the psychological load on them. Alternatively, indirect measures such as CV are preferable to alleviate the burden imposed on workers by invasive direct measurement methods, whereas monocular cameras are more feasible, considering their lower price than depth cameras [14].

CV-based methods for ERA and WMSDs prevention are robust and cost-effective. Due to the objectivity, time-saving, and cost-effectiveness of CV-based methods, they have received much attention in recent years, especially those based on deep learning [10,15–20]. These methods have also become increasingly popular in the construction industry. However, deep learning-based methods come with a problem that must be addressed: to get high-accuracy models, extensive amounts of high-quality data are needed for training. As a result, researchers have created several 3D human pose datasets, such as the Human3.6 M [21], MuCo [22], and NTU RGB + D 120 [23]. Meanwhile, researchers have also created 2D datasets, including COCO [24] and MPII [25]. However,

* Corresponding author.
E-mail address: xinming.li@ualberta.ca (X. Li).

https://doi.org/10.1016/j.autcon.2024.105452
Received 12 October 2023; Received in revised form 2 February 2024; Accepted 30 April 2024

most publicly available datasets primarily concentrate on the human pose estimation (HPE) of everyday activities and are not designed for construction activities. In addition, current publicly available datasets do not have comprehensive body joint information, such as details about hands. Therefore, ERA methods, such as REBA and RULA, which require such joint information to calculate risks contributed by strains from hands and wrists, cannot be applied accurately. In the context of CV-based deep learning methodologies, the utilization of a comprehensive dataset yields notable enhancements in terms of overall model generalization [26,27], overfitting prevention [28,29], and the effective handling of variations [30,31].

To address the lack of construction activity-specific 3D datasets with all coordinates required for REBA and RULA assessments, we introduce CP3D, a dataset obtained with an accurate retro-reflective marker-based motion capture (MoCap) system specifically designed for construction tasks. Unlike other datasets focusing on daily activities, CP3D focuses specifically on the poses of construction workers. With a high-performance motion capture system, the graphical and kinematic data were acquired by recording the activities of 5 male and 2 female subjects from 4 different perspectives (front, back, left, and right). With consent, we recorded and analyzed the workers' activities for >50 h from multiple angles and different work areas (University of Alberta ethics approval: Pro00111404) at several large construction facilities. As a result, the fourteen most frequent activities were selected and included in CP3D. CP3D fills a gap in the CV and ERA research communities that lack adequate 3D datasets designed for construction activities. For comprehensive data collection that fulfills thorough assessments of postural ergonomic risks, we additionally created a biomechanical model for the MoCap system. Datasets obtained from MoCap systems can represent human activities aggregated by a series of three-dimensional skeleton models [2]. With the help of the construction activity dataset, CV-based approaches [22,32–41] can extract human poses and activities from RGB videos or images [42] with higher precision, thus further achieving the goal of assessing worker safety and health. Differing from the majority of existing ERA methods, our approach employs a 3D dataset to achieve greater accuracy, which is a rationale behind our proposal of CP3D as a dedicated 3D dataset. We trained and tested models based on ResNet-50 using CP3D and other publicly available datasets. The training of the models is based on 18 specific 3D human joints to meet the needs of ERA tools. Both 3D and 2D datasets are used so that the model can infer 3D joint information from 2D images.

Furthermore, we introduce a CV-based ERA method that estimates REBA and RULA from monocular cameras. To the best of our knowledge, no CV-based ERA approach utilizes the same level of detailed 3D joint information as ours. The CV-based REBA approach [15] and CV-based RULA method are used to verify the performance of the CP3D benchmark dataset. Our CV-based ERA method is meticulously tested on CP3D and other publicly available datasets, achieving comparable performance. The results suggest that the model, trained on a generic activity dataset, demonstrates lower accuracy in joint recognition and ERA for construction activity workers compared to our model's performance. Our model, jointly trained using both CP3D and the generic activity dataset, exhibits better results in this context. The model trained with CP3D benchmark dataset achieves state-of-the-art performance with 35% accuracy improvement and, equally importantly, improves the comprehensiveness and performance of existing publicly accessible generic datasets. CP3D offers a distinctive compilation of construction activities tailored for future deep learning methodologies within the realms of HPE and ERA in construction-related contexts. Encompassing approximately 421,000 samples with various construction activities, performers, and captured perspectives, the dataset is designed to elevate performance within this specific domain. Additionally, its integration with other datasets contributes to augmenting dataset diversity, specifically catering to HPE and ERA. The inclusion of a wide range of construction scenarios aims to enhance overall model generalization, reduce the risk of overfitting, and adeptly handle variations, thereby facilitating future research across various industries.

The study aims to address three major challenges. Firstly, existing datasets predominantly concentrate on daily human activities, lacking the essential support for enabling CV algorithms to learn distinctive features relevant to the activities of construction workers. This impedes further improvement of CV methods for the construction sector. Secondly, there is a deficiency in biomechanical models for MoCap systems, specifically crafted to capture sufficient 3D joint information simultaneously applicable to both REBA and RULA methodologies. Thirdly, prevailing CV-based REBA and RULA methods, which utilize monocular cameras, often rely on 2D joint information and lack the comprehensive joint details mandated by REBA and RULA, thus necessitating a more holistic approach. Hence, the contributions include (1) the development of a comprehensive 3D HPE/ERA dataset with around 421,000 frames designed explicitly for construction activities; (2) the development of a biomechanical skeletal model specifically for collecting 3D human pose data on construction activities; (3) the development of CV-based ERA approaches using REBA and RULA. The contributions are further discussed in section 3.

Table 1 shows the abbreviations and their corresponding definitions used in this study.

2. Related work

2.1. 3D HPE

In the past decades, HPE based on CV has received compelling attention in the human pose recognition research field, and most of the progress has come from deep learning-based methods. HPE tasks can be broadly divided into 2D and 3D categories. Comprehensive reviews of 2D and 3D HPE methods can be found in [43–50]. Since ERA tools such as REBA and RULA require a 3D posture representation of persons, 3D HPE is required. This can be implemented with different approaches [44], such as multiple monocular cameras or IMUs. In this study, we conduct multi-person 3D HPE from a monocular camera with a single view, representing the most common scenario in construction facilities. Approaches for multi-person 3D HPE from a monocular camera with a single view are generally classified into top-down and bottom-up approaches [44–46,49,50].

- Top-down approaches: This strategy involves using human detectors to identify the positions of individuals, followed by the separate detection of their joints. The first step in the top-down 3D multi-person HPE approach is to detect the presence of each person in the image. The 3D HPE deep learning method estimates the absolute root (pelvis) position of each detected person and their 3D root-relative pose. The complete pose is then converted to world coordinates using the estimated 3D pose and its corresponding root position. In top-down HPE methods, the ability of person detectors to

Table 1
Abbreviations and definitions.

Abbreviations	Definitions
CV	Computer Vision
REBA	Rapid Entire Body Analysis
RULA	Rapid Upper Limb Analysis
CP3D	ConstructionPose3D
HPE	Human Pose Estimation
ERA	Ergonomic Risk Assessment
WMSDs	Work-related Musculoskeletal Disorders
IMUs	Inertial Measurement Units
MoCap	Motion Capture
CVRE	Computer Vision-based REBA Estimation
MPJPE	Mean Per Joint Position Error
DRWPA	A deep learning-based RULA method for working posture assessment

recognize the edges of extensively overlapped human forms could be compromised. In 2017, Rogez et al. [51] proposed LCR-Net, an innovative end-to-end localization classification-regression framework. LCR-Net demonstrated robustness in controlled environment datasets but underperformed in in-the-wild datasets. This limitation was addressed by Rogez et al., using synthetic data for training in their LCR-Net++, which utilizes a pose proposal generator to produce candidate poses for further classification [51]. The classified poses are then refined by a regressor in both 2D and 3D. Moon et al. [41] introduced a camera distance-aware framework comprising a human detector, a human root estimator, and a root-relative pose estimator. The framework can be fused with other human detectors and 3D human estimators.

- Bottom-up approaches: Contrary to top-down approaches, bottom-up approaches first identify the depth map and each joint independently. Each joint is then associated with the corresponding human based on the depths of the root and the joints, enabling complete pose estimation. A primary challenge of bottom-up approaches is accurately grouping joints to their respective person. Furthermore, these approaches face significant difficulties in effectively connecting key joints within occluded scenes, which increases the complexity of key joint association. A multi-stage bottom-up framework proposed by Zanfir et al. [52] begins by estimating volumetric heatmaps to predict 3D joint locations. Limbs are then connected based on the confidence scores of the connections between joints. In the final stage, skeleton grouping is employed to assign the limbs to the corresponding people. The approach introduced by Mehta et al. [22] utilizes the Occlusion-Robust Pose Map, which is capable of estimating human posture with high precision even in the presence of occlusions. They also introduced MuCo-3DHP and MuPoTS-3D, the first multi-person 3D dataset, for training and testing their method.

Regarding performance, the top-down approach generally outperforms the bottom-up approach. Firstly, the bottom-up approach struggles to generate precise heat maps, leading to diminished posture estimation performance when the image is scaled. Additionally, it often fails to accurately predict the posture of individuals with smaller body sizes. Moreover, the bottom-up approach faces challenges in associating key points to specific human instances in confined environments. Consequently, this paper incorporates a top-down 3D HPE method into our ERA approach.

2.2. 3D HPE datasets

Owning to the advent of high-precision MoCap systems, the generation of large-scale 3D HPE benchmark datasets has become more feasible. Table 2 below lists datasets that are widely used in 3D HPE. Human3.6 M is the most commonly used single-view 3D HPE dataset among the datasets. The dataset was collected from 6 professional male performers and 5 professional female actresses. The dataset consists of 3.6 million human pose samples with 17 daily activities from 4 different views in a controlled lab environment. Although more 3D HPE datasets have been published after Human3.6 M and HumanEva, Human3.6 M and HumanEva are still the most widely used and standard benchmark datasets for 3D HPE [45,49]. It is worth mentioning that MuCo-3DHP is the first large publicly accessible multi-person 3D HPE dataset with over 400,000 annotated frames. MuPoTS-3D is a relatively small dataset for testing. They were proposed by Mehta et al. [22] for training and testing their single-view 3D HPE method. CML is a relatively new 3D HPE dataset, and to the best of our knowledge, it is the only publicly available dataset centered on construction activities. However, CML is a combination of other existing datasets, and that makes CML not as diverse and abundant as it should be.

In contrast to daily activities, construction activities are more physically demanding and intricate and possess distinct human movement characteristics that set them apart from daily activities. In addition, construction activities include varying movements with a wide range and even a lot of movements with self-occlusions. Hence, applying generic daily activity datasets to construction scenarios will not bring optimal performance due to the lack of unique data patterns and characteristics for construction activities. Taking the widely used Human3.6 M as an example, it has 15 activities in the entire dataset, which consists of 3.6 million annotated images. The 14 activities include 'Directions,' 'Discussion,' 'Greeting,' 'Posing,' 'Purchases,' 'Taking Photo,' 'Waiting,' 'Walking,' 'Walking Dog,' 'Walking Pair,' 'Eating,' 'Phone Talk,' 'Sitting,' 'Smoking,' and 'Sitting Down,' which are not directly related to construction activities.

Therefore, there is a need for a comprehensive dataset that emphasizes the most frequent activities and adequate data that can cover the variations in the construction activities. The CV-based ERA is closely related to CV-based HPE. In this sense, the CP3D dataset developed in this study can be used alone or together with other datasets to train deep learning methods, thereby improving the action type breadth and accuracy of HPE. In addition, CP3D contains essential joint information (i.e., hand coordinates) that some datasets [21] might not have for more comprehensive ERAs.

Table 2
Datasets for 3D HPE.

Dataset	Year	#Frames	#Subjects	Resolution	Context and Characteristics
CML [2]	2022	> 146,000	10	N/A	Lab environment
AMASS [53]	2019	9 million	346	variable	Unified parametrization of 15 datasets
MuPoTS-3D [22]	2018	8000	8	2048 × 2048, 1920 × 1080	Multi-person, "In the wild"
3DPW [54]	2018	> 50,000	7	N/A	"In the wild," single moving camera & IMUs
DIP-IMU [55]	2018	330,000	10	N/A	Inertial Measurement Units
MuCo-3DHP [22]	2018	400,000	8	1024 × 1024	Multi-person
MPI-INF-3DHP [25]	2017	1.3 million	8	N/A	"In the wild" & lab, marker-less ground truth
Total Capture [56]	2017	1.9 million	5	1920 × 1080	Lab environment, Inertial Measurement Units
NTU + RGBD 120 [23]	2016	> 114,000	40	1920 × 1080	Office environment
CMU Panoptic [57]	2016	1.5 million	8	1920 × 1080	Lab environment
TNT15 [58]	2016	13,000	4	800 × 600	Office environment, Inertial Measurement Units
Human3.6 M [21]	2014	3.6 million	11	1000 × 1000	Lab environment
HumanEva [59]	2010	40,000	4	660 × 500	Lab environment

2.3. CV-based ERA

According to the U.S. Occupational Safety and Health Administration, ergonomic risks include repetition, awkward posture, forceful motion, stationary position, direct pressure, vibration, extreme

temperature, noise, and work stress [60]. The advantages of CV in visual information perception fit well with the visual ERA [15]. Although ERA has a long history of exploration, CV-based ERA has recently emerged due to advancements in the CV field. Similar to CV-based HPE, the CV-based ERA methods can also be categorized into 2D and 3D, depending on the category of human key point coordinates used to perform the ERA. These approaches focus only on ergonomic risks that can be visually identified.

For accurate ERA, 3D information on workers' joints is required. Table 3 shows some of the most recent CV-based 3D ERA approaches. It's worth noting that only [15] can estimate ergonomic risks for multiple people. Fan et al. [15] proposed the CV-based REBA Estimation (CVRE) that is based on 3D poses, which is further developed in this study for the ERA.

REBA [66] and RULA [67] are the two commonly used quantitative observational ERA tools. RULA was originally proposed for rapid assessment of the strain/load imposed on the musculoskeletal system by postures, muscle functions, and external loads on the neck, trunk, and upper limbs. REBA, a similar posture analysis and ergonomic risk assessment tool, is sensitive to musculoskeletal risks of work tasks in a variety of occupations. A few researchers have applied their CV-based pose detections for ERA: CVRE is capable of CV-based 3D multi-person ERA [15]. CVRE includes a human detection module, a human key point estimation/detection module, a joint angle calculation module, a REBA score calculation module, and a feedback module. CVRE adopts the human key point estimation module from 3DMPPE [41]. Because 3DMPPE is a 3D multi-person pose estimation approach, CVRE can inherently estimate ergonomic risks with 3D joint data for multiple people. Like CVRE, the CREBAS [16] utilizes CV to estimate ergonomic risks with 3D joint data. CREBAS uses the body detection model of MediaPipe, and the method sets body areas based on the positions of faces, which are detected by the face tracking model. The joint positions are estimated using the posture tracking model of MediaPipe. Finally, based on the 3D joint positions, CREBAS calculates the joint angles and uses these angles to calculate the REBA scores. In addition, a deep learning-based RULA method for working posture assessment (DRWPA) is a CV-based 3D ERA approach. DRWPA adopts the 3D HPE method [38] proposed by Martinez et al., which is a single-person-based method. Hence, DRWPA can only estimate ergonomic risk for single-person scenarios. Meanwhile, the dataset DRWPA uses Human3.6 M for training. Since Human3.6 M does not contain annotations of the hands, the accuracy of DRWPA-based RULA scores is sacrificed because of the loss of hand information.

Therefore, the current CV-based ERA methods either rely solely on 2D human pose information [19,68], lack hand coordinates, or are restricted to single-person scenarios. However, the nature of construction work is closely tied to manually handling materials or using tools. Ergonomic risk assessment algorithms such as REBA and RULA necessitate the utilization of hand coordinates to compute parameters such as hand bending, hand side bending, and wrist twisting [66,67,69]. Consequently, datasets offering solely 2D joint coordinates or lacking sufficient hand coordinates fall short of fulfilling the criteria for a thorough ERA, especially using REBA and RULA [70]. More importantly, to test the performance of the developed dataset in this paper within the ERA domain, it is necessary to employ a 3D ERA method that incorporates hand information. As a result, we propose a 3D ERA approach incorporating REBA and RULA for construction activities.

Table 3
Approaches for CV-based ERA.

Approach	Year	2D or 3D Pose	Multi-person	ERA method
Jeong et al. [16]	2023	3D	N/A	REBA
Fan et al. [15]	2022	3D	Yes	REBA
Ciccarelli et al. [61]	2022	3D	N/A	RULA
Li et al. [62]	2019	3D	N/A	RULA
Yu et al. [63]	2019	3D	N/A	Workload analysis
Parsa et al. [64]	2021	3D	N/A	REBA
Yu et al. [65]	2018	3D	N/A	Workload analysis

3. Methodology

Based on the aforementioned research gaps, this paper introduces a new dataset called CP3D, a new biomechanical skeletal model specifically for collecting 3D human body data on construction activities, and an ERA system for REBA and RULA estimations. As an overview, in step 1, the MoCap system captures 3D coordinates and videos of the 14 construction activities of the performers in the lab, and the CP3D dataset is created by using these collected 3D coordinates along with their video frames. In step 2, deep learning models are trained with the CP3D dataset. To test our method's performance in identifying construction workers' joints, the CP3D is divided into two parts for training and testing. In step 3, the deep learning models are used for HPE. To assess the impact of the dataset on estimating the poses of construction workers, we employ the methodology introduced in [15]. Specifically, two deep learning models are trained within this framework; one exclusively utilizes publicly available datasets for training, while the other incorporates CP3D in addition to the same public datasets. Subsequently, in step 4, the estimated poses from these two models are utilized for the computation of REBA and RULA risk scores. In step 5, the results of the risk assessment are derived by utilizing the REBA and RULA risk scores obtained in step 4. Additionally, the trained models are applied to the test data separately for method validation and performance testing.

In the following steps, validation of the method and testing of the model performance is executed; the validation and testing procedures are threefold: (1) validation of HPE: the estimated joint angles (from CV-based CP3D-trained models) in the test partition and their corresponding ground truth joint angles (from MoCap system) are compared; (2) trained model testing: with MoCap as the ground truth, the Mean Per Joint Position Errors (MPJPE) for the test partition from deep learning trained models with and without CP3D dataset are obtained; (3) performance improvement validation: to validate that CP3D can improve the performance of models, we also compare the joint angle estimation performance of a deep learning model trained only using public datasets with a model trained jointly with CP3D.

Fig. 1 shows our study's methodological steps to estimate ergonomic risks, which will be further explained in the following sections.

3.1. CP3D dataset

Optical marker-based human MoCap systems are widely used to capture ground truth 3D coordinates for HPE datasets [47]. In this paper, we adopt a non-invasive marker-based MoCap for data collection, as such an approach does not impose additional psychological burdens on the worker/subject and has less interference with the worker's tasks [15]. The data was collected using the Vicon MoCap system [71], consisting of eight Vicon Vero cameras and one Vicon Vue camera. The Vicon Vero cameras capture the 3D coordinates of the reflective markers by capturing the light signals reflected by the markers. Eight Vicon Vero cameras are used to ensure that the markers can be captured from multiple locations/directions in the lab. A Vicon Vue camera captures footage of subjects engaging in activities, and these recorded activities serve as the basis for generating the video frames in the CP3D dataset. The lab layout is shown in Fig. 2.

The Vicon Vue and Vero cameras are set to a frequency of 100 Hz for synchronization. The Vicon MoCap system keeps track of the 48 retro-reflective markers on the performers. The biomechanical model used in CVRE for data collection is Plugin-Gait, but it does not contain all the joint information required by REBA and RULA. Therefore, we have

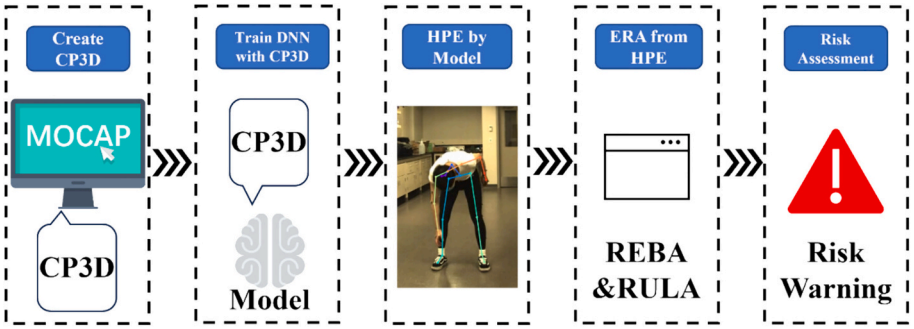


Fig. 1. Methodological steps to estimate ergonomic risks in our study.

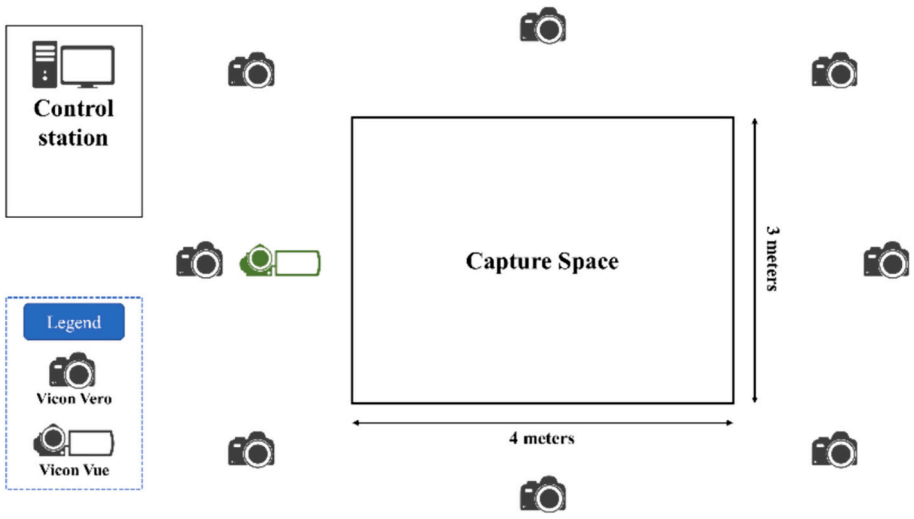


Fig. 2. The lab layout for data collection.

created a new biomechanical model for adequate and accurate data collection and deployed it to the Vicon MoCap system. Furthermore, we also created a new algorithm to extract joint information. Fig. 3 shows the creation processes of the proposed CP3D, which corresponds to the first step in Fig. 1. Therefore, one of the objectives to develop a CV HPE/ERA dataset designed explicitly for construction activities can be achieved, which fills in the dataset gaps in construction fields.

3.1.1. *Activity selection*

Common construction activities are selected by observing workers in construction facilities for the selection step. To select the most frequent candidate activities, we conducted observations in a large-scale modular construction facility and a thorough analysis of the surveillance videos inside their facility. The analysis involves manual observation of randomly chosen videos from a pool of 70 h of surveillance footage, with a particular emphasis on 10 h of production activities. This analysis

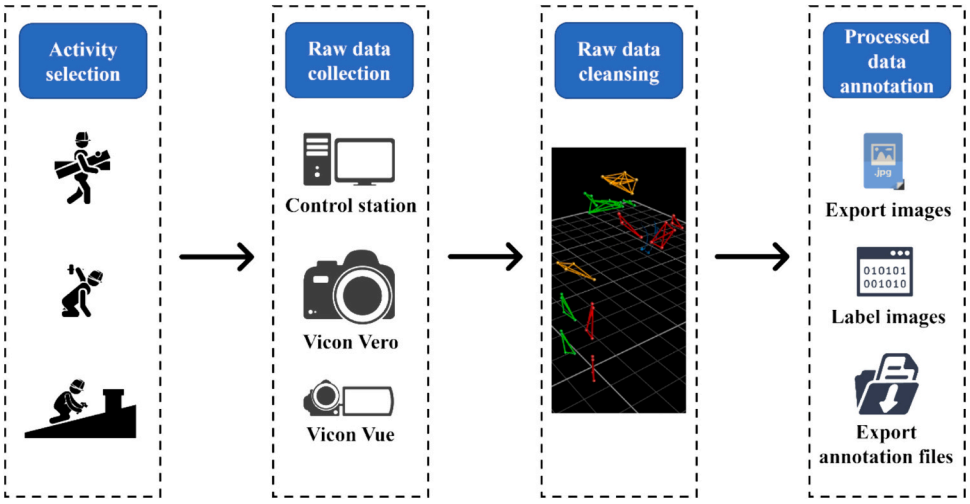


Fig. 3. Creation processes of CP3D, and it corresponds to the first step in Fig. 1.

entails documenting various types of construction activities and tracking the frequency at which these activities occur, noting consistent repetition of actions as if they were performed once. From the video analysis, we identified the 14 most frequent activities and included their number of occurrences in Table 4. Fig. 4 gives examples of those activities. To ensure that daily activities with low physical demand are not mistakenly identified as construction activities, the physical demand analysis [72] is employed to classify construction activities. As a result, the 14 most common work activities are simulated by a trained group of five males and two females.

3.1.2. Raw data collection

For the data collection step, performers carry out those common construction activities for the MoCap system data collection. According to the U.S. Bureau of Labor Statistics [73], female workers numbered 10.9% of the entire construction workforce in 2020 in the U.S. Therefore, to better represent the actual gender demographic, two female and five male performers are recruited for construction activity simulations. Each subject performs at least four trials for each of the 14 activities, and the trials with the best quality (i.e., low number of missing markers and smooth marker trajectories) are used to create the CP3D dataset. It is worth noting that the Vicon Vue camera captures 4 different perspectives in different trials of the same activity for dataset diversification. CP3D contains a total of 7 subjects' simulations of construction activities. Each simulated trial includes approximately 10 to 14 s of MoCap data. The demographic information of the trained performers is listed in Table 5. The average height, weight, and BMI are 173 cm, 64 kg, and 21.2, respectively. The standard deviation of the height, weight, and BMI are 8.2 cm, 10.6 kg, and 1.6, respectively.

The data acquisition frequencies of the Vero and Vue cameras in the motion capture system are adjustable through the Vicon Nexus. With the MoCap system's camera set at 100 Hz and each trial lasts around 10 to 14 s, each trial has 1000 to 1400 frames. In this setting, each subject simulates 4 trials for every 14 activities; the entire dataset contains about 421,000 annotated frames.

3.1.3. Raw data cleansing

For the data cleansing step, irrelevant and incorrect data in the simulated construction activity data is processed. To perform data cleansing, the beginning and end of each trial are trimmed in the dataset because they contain actions irrelevant to construction tasks. In preparation for annotation, any missing or mislabeled reflective markers are manually corrected as part of the data cleansing phase. The built-in functions in the Vicon MoCap software are used to process the missing or mislabeled markers. Since the Vicon MoCap system sometimes fails to capture some retro-reflective markers, the trajectories/3D coordinates of those markers that are not captured are not shown in the captured trial. The Vicon Nexus Software provides the ability to fill in gaps for the retro-reflective marker trajectories. This study uses "spline fill," "pattern fill," and "rigid body fill" to fill in gaps for more accurate results. However, since the way for gap-filling is to estimate the trajectories of an uncaptured marker based on the trajectories of captured markers, errors are inherent in estimations.

Table 4
The 14 most common construction activities with their number of occurrences.

Construction activity	Description	Number of occurrences	Example
1	kneeling and working	59	Fig. 4 (1)
2	carrying heavy objects over the worker's head	23	Fig. 4 (2)
3	dragging	27	Fig. 4 (3)
4	bending over and working with a power drill	82	Fig. 4 (4)
5	holding and moving heavy objects with both hands	71	Fig. 4 (5)
6	installing windows	37	Fig. 4 (6)
7	kneeling on one knee and nailing nails	41	Fig. 4 (7)
8	dragging and backing up	28	Fig. 4 (8)
9	bending and grabbing an object on the ground	84	Fig. 4 (9)
10	pushing forward	48	Fig. 4 (10)
11	pushing objects above the worker's shoulders	33	Fig. 4 (11)
12	extending arms on a ladder	43	Fig. 4 (12)
13	reaching above the worker's head	89	Fig. 4 (13)
14	sitting with trunk bending	135	Fig. 4 (14)

3.1.4. Processed data annotation

Lastly, for the annotation step, we developed a software tool for the automated conversion of annotation files, which was utilized to convert MoCap data into the MSCOCO [24] format annotation. FFmpeg is employed to convert the trial videos into images, with a frame rate of 100 frames per second, in order to synchronize with the MoCap data [74]. To facilitate the annotation process, we created a software tool in Python to convert Vicon MoCap data in csv format into the MSCOCO JSON format annotation. The software will be made public together with the dataset at https://github.com/xinmingliUofA/CP3D. The data for each subject is composed of four JSON files. These files encompass the 3D and 2D joint coordinates corresponding to the subject's stance in each image, along with specific image information such as image size, filename, and action type. Additionally, the JSON files include Vicon Vue camera details like rotation matrix, world coordinates, focal length, and principal point.

3.2. A new biomechanical model

Vicon's motion-capture system involves attaching retro-reflective markers to subjects' skin and clothing. Vicon Nexus provides a function to label markers automatically based on biomechanical models. However, none of the model templates provided by Vicon Nexus meets ERA's needs. Thus, a template containing all the markers required for the REBA and RULA is created for the Vicon Nexus motion capture software, using a total of 48 retro-reflective markers. The template and name of abbreviations are shown in Fig. 5 and Table 6. Unlike the widely used and well-known Plugin-Gait model provided in the Vicon Nexus, our template provides the 3D coordinates of hands and other joints for ERA, which is suitable for direct 3D coordinate conversions.

The motion capture software for collecting and processing data is the Vicon Nexus. Besides the algorithms mentioned in the data cleansing process, the other gap-filling algorithm we use in the Vicon Nexus is specifically designed for markers on rigid body segments. This algorithm only works for gap-filling with >3 markers on the segment. To reduce the error, we add redundant markers to our Vicon biomechanical model to ensure the rigid body segments have enough markers to use the gap-filling algorithm. As shown in Fig. 5, the markers for head, knees, and ankles are 'rigid body,' with each of these segments having >3 markers around them. For a detailed example, only two markers on the medial and lateral sides of the left knee are required to capture the coordinates of the left knee. However, markers on performers may be blocked by obstacles or their body segments (self-occlusion). The rigid body algorithm can estimate the blocked markers based on the other unblocked ones. As a result, including redundant markers on rigid bodies improves the accuracy of data collection.

3.3. 3D ERA approach incorporating REBA and RULA

Similar to CVRE, our CV-based REBA method also has a modular design, which means that each module can be replaced with software/program that functions similarly to it. Fig. 6 shows the workflow of our

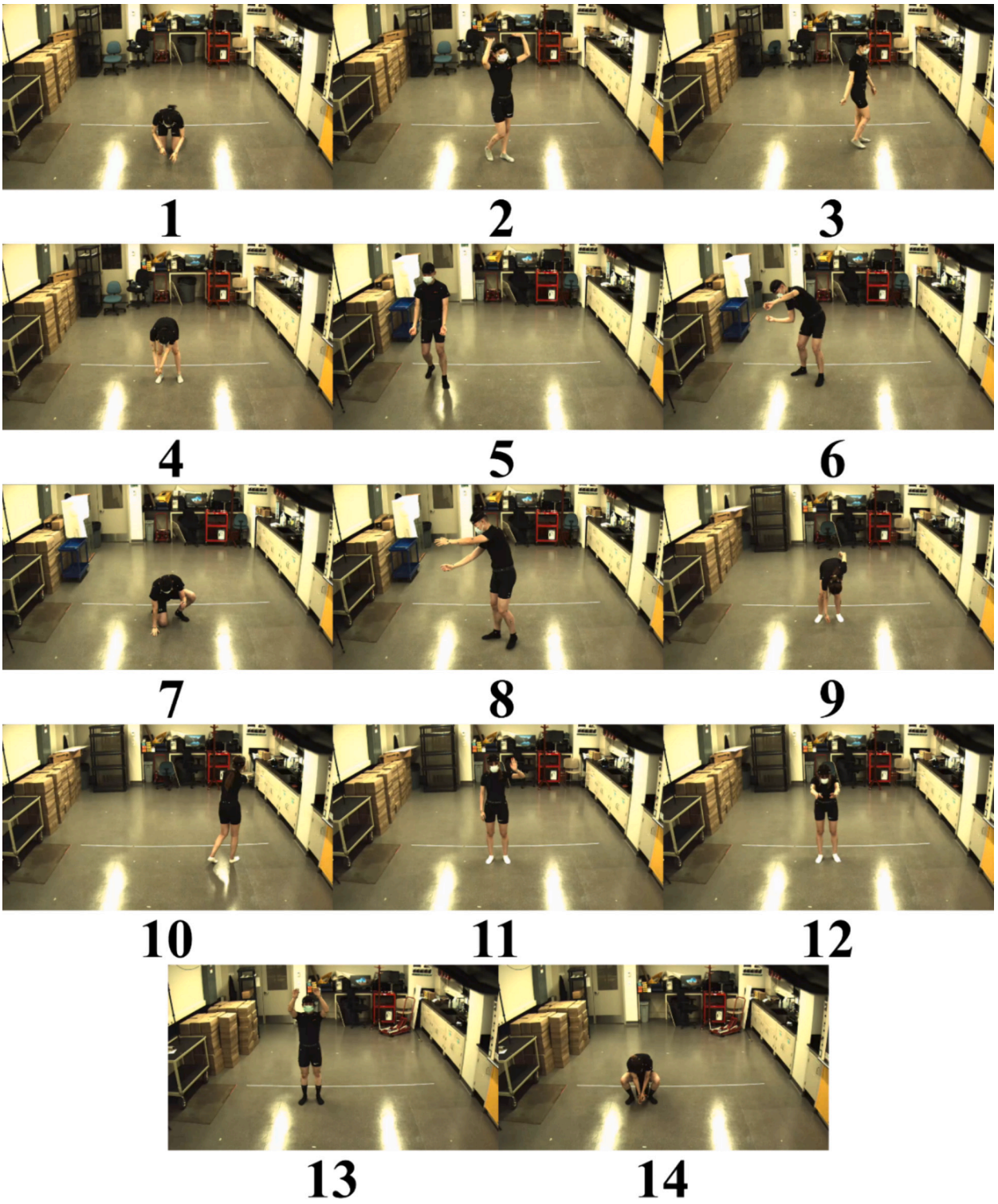


Fig. 4. Examples of the 14 most common construction activities.

modular REBA and RULA methods. Our CV-based REBA consists of four functional modules: human body detection, 3D HPE, joint angle estimation, and REBA risk score estimation. Our method uses YOLOv7 [75] as the human body detection module. For the 3D HPE module, our approach adopts the 3D pose estimation method [41]. The articulated pose is selected for the joint angle estimation module because it does not contain unnecessary texture/body shape and background information [62]. The joint angle estimation module uses an 18-joint representation because the representation includes all the necessary joints for REBA and RULA. The 18 joints representation are head, nose, thorax, right/left shoulder, right/left elbow, right/left wrist, right/left hand, right/left hip, pelvis, right/left knee, and right/left ankle. CVRE employs a 16-joint representation due to the utilization of the Human3.6 M training dataset, which lacks any hand-related information. Fig. 7 shows the 18 joint articulated pose representations in our method, with one side labeled for better clarity. The REBA score estimation module follows the calculation rules of REBA [66].

As shown in Fig. 7, 18 joints are articulated and used for ERA. The 18

Table 5
Demographic information of simulated construction activity performers.

Performer	Height (cm)	Weight (kg)	BMI	Gender
1	163	52	19.6	F
2	184	76	22.4	M
3	164	54	20.1	F
4	170	60	20.8	M
5	181	80	24.4	M
6	171	61	20.9	M
7	178	65	20.5	M

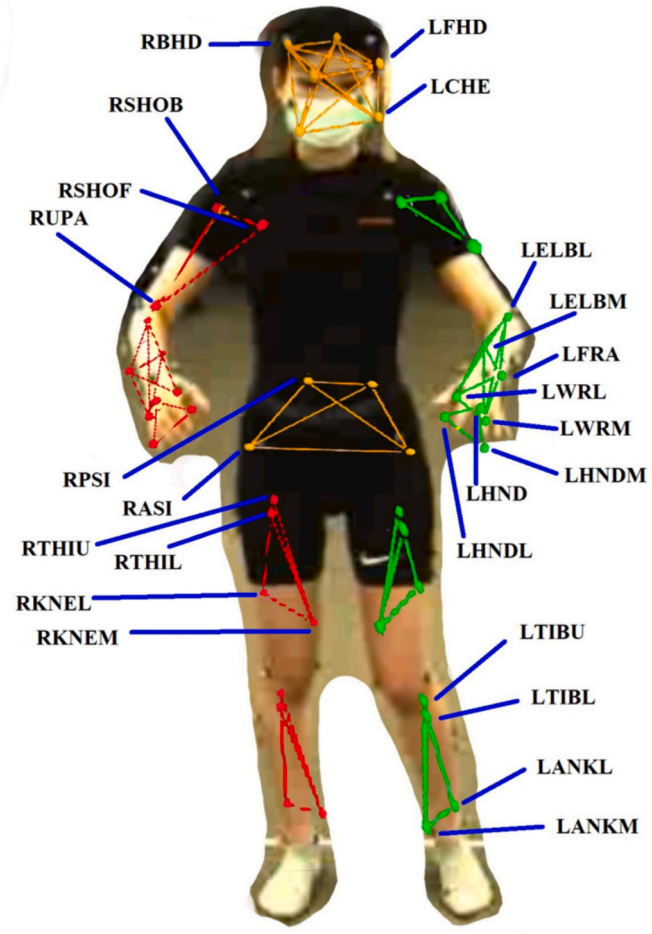


Fig. 5. The layout of the customized biomechanical model.

joints are derived from the 48 retro-reflective markers. The head coordinate is calculated as the center of the RFHD and the LBHD. The left hip and right hip are calculated as the center of the R(*L*)ASI and R(*L*)PSI. The pelvis is calculated as the midpoints of the left and right hips. The thorax is calculated as the center of the left and right shoulder. The nose is calculated as the midpoint of the head and thorax. All the other joints are calculated as the center of their corresponding 2 retro-reflective markers.

Our CV-based RULA approach uses a top-down CV HPE structure to estimate workers' joints. Our approach comprises four modules: human detection, posture estimation, joint angle estimation, and RULA score calculation, where the human detection and 3D HPE models are identical to our REBA approach. The RULA score calculation module is developed based on the algorithm defined in [67]. The joint angle calculations of RULA are the same as those of REBA, except that RULA doesn't include joint angles for the lower body. It is worth noting that the load and coupling scores are not considered in our REBA and RULA because they are not easily accessible from visual information.

3.3.1. Joint angle calculation

For joint angle calculations, the sagittal plane (P_s), the frontal plane (P_f), and the horizontal plane (P_h) are required. Fig. 8 demonstrates the positions of those three planes. Body segments are represented by 2 end joints, and they are in the vector form:

$$S_{a-b} = J_b - J_a \tag{1}$$

where S_{a-b} represents the body segment vector pointing from joint a to joint b; J_a represents the 3D coordinate of joint a. The projected segment vector is calculated with the following equation:

$$S_{a-b}^{P_i} = S_{a-b} - S_{a-b} \bullet \frac{P_i}{\|P_i\|^2} P_i \tag{2}$$

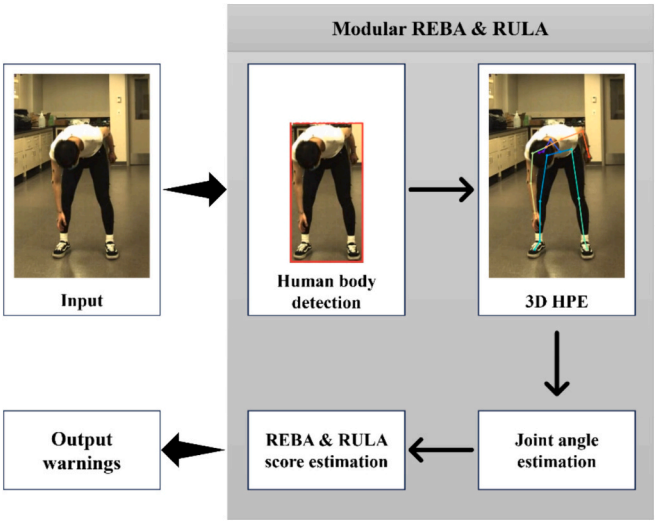


Fig. 6. Workflow of our modular REBA and RULA methods.

Table 6
Names and abbreviations of markers.

Abbreviation	R(*L*)FHD	R(*L*)BHD	R(*L*)CHE	R(*L*)SHOF	R(*L*)SHOB	R(*L*)UPA
Name	Right(left) forehead	Right(left) back head	Right(left) cheek	Right(left) shoulder front	Right(left) shoulder back	Right(left) upper arm
Abbreviation	R(*L*)ELBL	R(*L*)ELBM	R(*L*)FRA	R(*L*)WRL	R(*L*)WRM	R(*L*)HND
Name	Right(left) elbow lateral	Right(left) elbow medial	Right(left) front arm	Right(left) wrist lateral	Right(left) wrist medial	Right(left) hand
Abbreviation	R(*L*)HNDL	R(*L*)HNDM	R(*L*)ASI	R(*L*)PSI	R(*L*)THIU	R(*L*)THIL
Name	Right(left) hand lateral	Right(left) hand medial	Right(left) anterior superior iliac	Right(left) posterior superior iliac	Right(left) thigh upper	Right(left) thigh lower
Abbreviation	R(*L*)KNEL	R(*L*)KNEM	R(*L*)TIBU	R(*L*)TIBL	R(*L*)ANKL	R(*L*)ANKM
Name	Right(left) knee lateral	Right(left) knee medial	Right(left) tibia upper	Right(left) tibia lower	Right(left) ankle lateral	Right(left) ankle medial

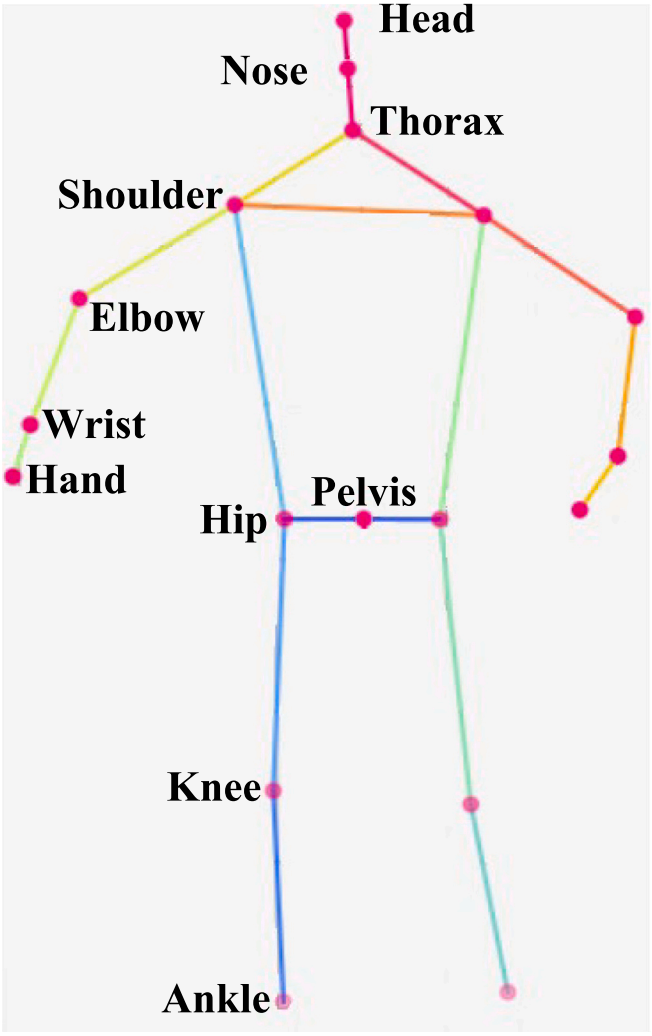


Fig. 7. The 18 joint articulated pose representations.

where $S^{P_i}_{a-b}$ is the projection of the segment vector on the plane P_i. Joint angles between projected segment vectors are calculated with the following equation:

$$\theta^{P_i}_{a-b\wedge c-d} = arccos\left(\frac{S^{P_i}_{a-b} \bullet S^{P_i}_{c-d}}{|S^{P_i}_{a-b}||S^{P_i}_{c-d}|}\right) \quad (3)$$

where $\theta^{P_i}_{a-b\wedge c-d}$ represents the angle between projected segment vectors $S^{P_i}_{a-b}$ and $S^{P_i}_{c-d}$. A detailed definition of angle calculations is demonstrated in [66]. The calculations of side bending and twisting of some segments are worth mentioning. The trunk side bending angle is the angle between $S_{pelvis-thorax}$ and $S^{P_s}_{pelvis-thorax}$. The trunk twisting angle is the angle between $S^{P_h}_{L\ shoulder-R\ shoulder}$ and $S^{P_h}_{L\ hip-R\ hip}$. The neck side-bending angle is the angle between $S^{P_f}_{pelvis-thorax}$ and $S^{P_f}_{head-thorax}$. The neck twisting angle is the angle between $S^{P_h}_{head-nose}$ and $S^{P_h}_{L\ shoulder-R\ shoulder}$. The shoulder-raised parameter of REBA is defined by the angle between $S^{P_f}_{L\ shoulder-R\ shoulder}$ and $S^{P_h}_{head-thorax}$. The upper arm abducted parameter is defined by the angle between $S^{P_f}_{L\ elbow-L\ shoulder}$ and $S^{P_f}_{L\ hip-L\ shoulder}$ for the left side. Similarly, the angle between $S^{P_f}_{R\ elbow-R\ shoulder}$ and $S^{P_f}_{R\ hip-R\ shoulder}$ defines the right side.

3.3.2. Network and training details

The 3D HPE module uses 3DMPPE PoseNet, a state-of-the-art CV algorithm capable of accurately estimating 3D joint coordinates. The backbone architecture of PoseNet is ResNet-50 [76] and PoseNet outperforms most state-of-the-art methods in terms of MPJPE [41]. PoseNet requires extensive data for training to achieve satisfactory accuracy. It needs at least one 2D and one 3D dataset for training because it estimates the camera-centered joint coordinates by separately estimating the 2D image coordinates and 3D depth value. Using the estimated depth value, the 2D image coordinates are projected back into the camera-centered coordinate space. Fig. 9 shows a visualized qualitative result of a scenario from our approach.

At present, there are some training datasets for HPE, but most of these datasets only contain 2D coordinates of human key points, which cannot be used to estimate the 3D human joint angles accurately. As mentioned earlier, the Human3.6 M dataset does not meet the requirement for precise ERA because it does not contain annotations for 3D coordinates of the human hands. CP3D in this study includes not only the 3D annotations of human hands but also the annotations of all the key points needed for REBA and RULA. The HPE module in our approach allows its deep learning method to be trained using multiple datasets. To enable our method to detect necessary human joints for ERA and 3D HPE, the training of deep learning methods includes at least one 3D dataset and one 2D dataset. In order to compare the performance between the models trained with and without CP3D, the first model is trained with MuCo-3DHP and MSCOCO datasets, while the second model is trained with CP3D, MuCo-3DHP, and MSCOCO datasets.

The CV-based HPE module is implemented with PyTorch, and its backbone is ResNet-50, pre-trained with the ImageNet dataset [77]. The deep learning method is trained with a start learning rate of 0.001, and

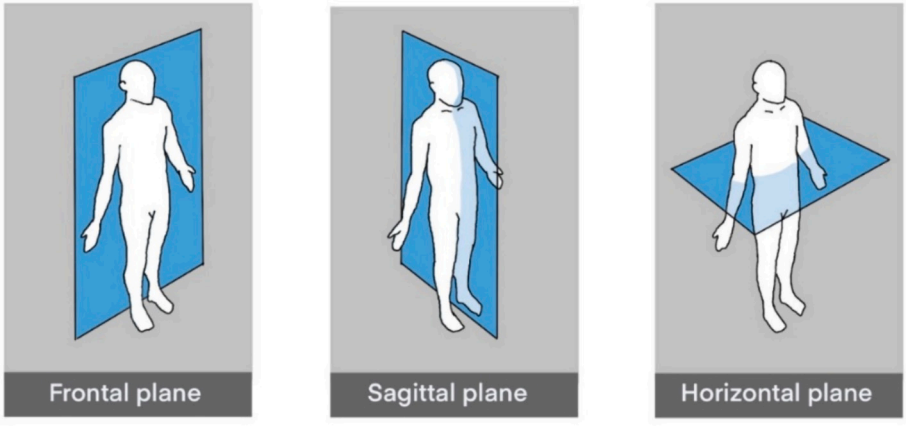


Fig. 8. The positions of frontal, sagittal, and horizontal planes.

Fig. 9. A visualized qualitative result of a scenario from our approach.

the Adam optimizer [78] updates the learning rate with a mini-batch size of 128. Table 7 shows the training hyper-parameters and environment of the models. In addition to CP3D, our method uses MSCOCO as a 2D training dataset and MuCo-3DHP as another 3D training dataset for performance comparison. Around 86% of CP3D is used for training and the remaining 14% for testing. Since subject 7 has a BMI and height that are close to the average values of all the subjects, data from subject No. 7 is used for testing, and data from subjects No. 1 through 6 is used as the training partition.

3.4. Validation

Since 3D HPE is an intermediate process of our CV-based ERA approach, further evaluation of our CV-based ERA is needed. Hence, our ERA module's joint angles and other parameters are compared to their corresponding ground truth values. The evaluations are carried out for both deep learning models to evaluate the performance and improvement that CP3D brings to the generic daily activity dataset. Two models are used in the HPE module of our ERA method, and the effect of those two HPE models on ERA performance is compared. The validation process can be divided into three steps.

Step 1 (validation of HPE): We compare the joint angles of a model trained using both MuCo-3DHP and CP3D to the ground truth joint angles. For CP3D, 86% (the training partition) of our dataset was used for training, and another 14% (the test partition) was used for testing the performance. The test partition we use in CP3D is independent of the training partition by segregating the partitions with different performers; thus, the similarity between the training and testing partition is minimized.

The arithmetic average and the standard deviation of the absolute difference between the ground truth and estimated ERA parameters are also calculated for the 14 actions. The arithmetic average of absolute differences for an ERA angle/parameter is in Eq. (4).

$$\mathrm{M}(\theta^e, \theta^{gt}) = \frac{1}{N}\sum_{i=1}^{N} |\theta_i^e - \theta_i^{gt}| \tag{4}$$

where N is the number of frames, θ_i^e and θ_i^{gt} are the estimated angle/parameter and the ground truth angle/parameter of the i_{th} frame. Hence, the average of differences for all the ERA angles/parameters is calculated.

The mathematical representation of the standard deviation of absolute differences for a REBA or RULA angle/parameter is in the form of Eq. (5).

$$\sigma(\theta^e, \theta^{gt}) = \sqrt{\frac{1}{N}\sum_{i=1}^{N} \left(|\theta_i^e - \theta_i^{gt}| - \mu\right)^2} \tag{5}$$

where N is the number of frames, $|\theta_i^e - \theta_i^{gt}|$ is the absolute value of the difference between the estimated angle/parameter and the ground truth angle/parameter of the i_{th} frame. μ is the average of the absolute value of differences.

Step 2 (trained model testing): The MPJPE metric is used for the HPE performance evaluation of those two models by comparing the estimated 3D coordinates of human key points with ground truth. Eq. (6) is the mathematical representation of the MPJPE and is the most common evaluation metric used in 3D HPE.

$$\mathrm{E}(\mathrm{x}, \tilde{\mathrm{x}}) = \frac{1}{N}\sum_{i=1}^{N} \|x_i - \tilde{x}_i\|_2 \tag{6}$$

where N is the number of joints, x_i and $\tilde{x}_i$ are the ground truth coordinate and the estimated coordinate of the i_{th} joint. Finally, the MPJPEs are averaged over all frames. The MPJPE is meant to evaluate the performance and improvement of the CP3D dataset concerning existing generic 3D datasets, MuCo-3DHP, in this case.

Step 3 (performance improvement validation): We compare the performance of a deep learning model trained individually using public datasets with a model trained jointly with CP3D. Specifically, we compare the model trained using both MuCo-3DHP and CP3D (Model 1) with the model trained using only MuCo-3DHP (Model 2). For CP3D, the same training partition of our dataset in Step 1 was used to train Model 1, and the same test partition was used to test the performance of both Models 1 and 2.

Table 7
Hyper-parameters and training environment of the models.

Training parameters/environment	Values
CV framework	PyTorch
CV backbone	ResNet-50 pre-trained with ImageNet dataset
start learning rate	0.001
Optimization algorithm	Adam optimizer
Mini-batch size	128
Epochs	25
GPU	2 NVIDIA GA102GL RTX A5000
CPU	Intel Xeon W-2295
Operating system	Ubuntu 22.04 64-bit
GPU driver	Nvidia 530.30 with CUDA 11.8
Programming language	Python 3.9

4. Results and discussion

For the first step of the validation procedure (validation of HPE), results are listed in Table 9 and Table 10. The performance of the model trained with CP3D is robust because the estimations are close to the ground truth values. Fig. 10 shows a sample frame's estimated pose and ground truth pose in the test partition.

For the second step (trained model testing), the MPJPE of the test partition is listed in Table 8 for the 14 activities. From Table 8, the MPJPE from the deep learning model jointly trained with CP3D and MuCo-3DHP outperforms the MPJPE from the deep learning model solely trained with MuCo-3DHP for all 14 activities. Given that the average MPJPE of the model trained with CP3D only has 60 mm, the average MPJPE for the model trained with CP3D is 30 mm less than the model without CP3D. This means a performance increase of about 35% when using CP3D together with MuCo-3DHP for construction activity posture estimation. Hence, the results indicate that CP3D can improve HPE performance for construction activities. As the results suggest, the differences between the MPJPEs obtained from the deep learning model with CP3D and those without CP3D are around 50%. Hence, the HPE module with CP3D performs significantly better than the HPE module

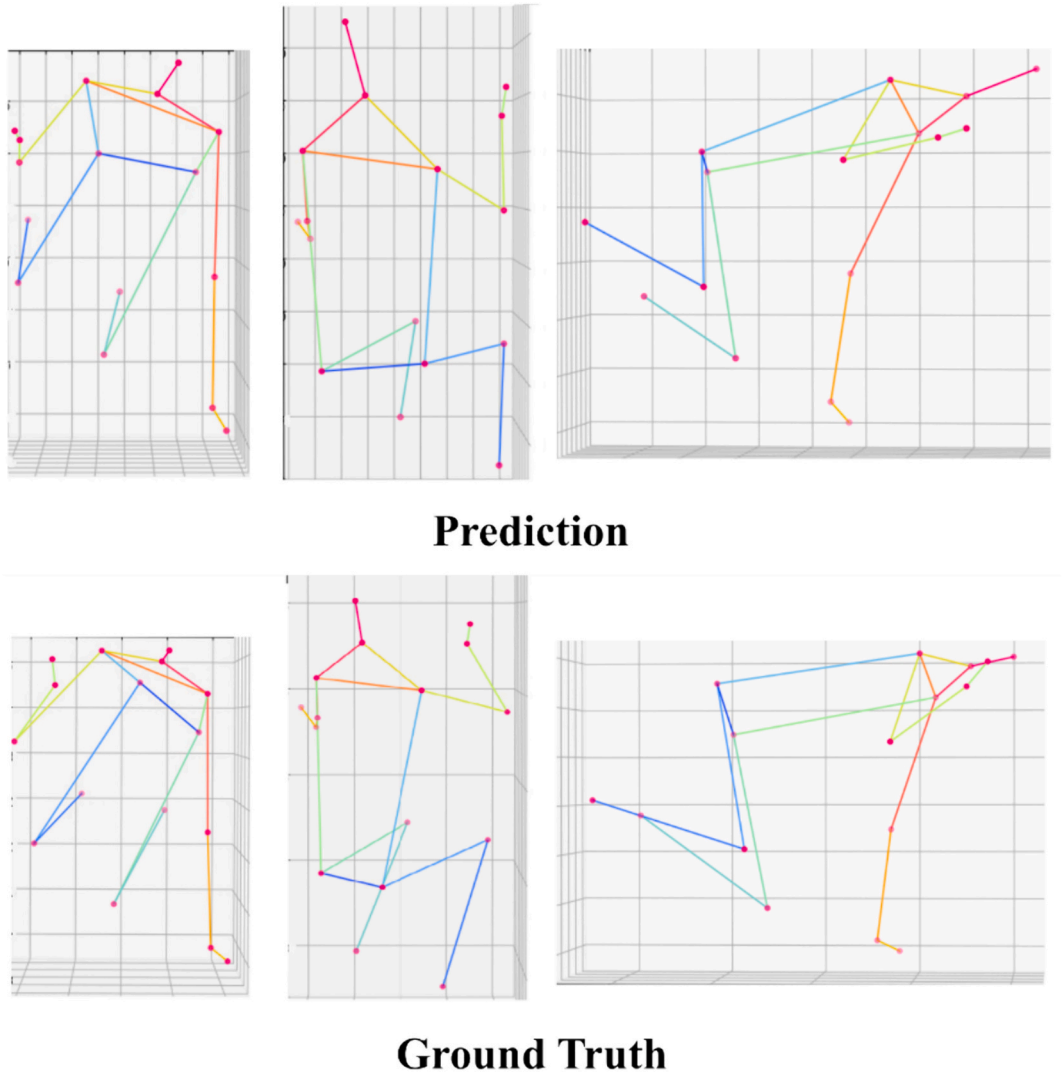


Fig. 10. Estimated pose and the ground truth pose of a sample frame in the test partition. The predictions are results from the model trained with CP3D.

Table 8
MPJPE (millimeters) of performance of deep learning models with and without CP3D.

Dataset	CP3D + MuCo+COCO (mm)	MuCo+COCO (mm)
Act. 1	54.87	103.50
Act. 2	60.67	99.85
Act. 3	53.28	88.52
Act. 4	51.79	94.71
Act. 5	57.75	96.80
Act. 6	67.50	103.03
Act. 7	76.88	106.39
Act. 8	49.78	81.35
Act. 9	55.14	78.99
Act. 10	57.48	88.29
Act. 11	55.68	90.55
Act. 12	56.83	81.75
Act. 13	66.04	79.44
Act. 14	81.74	98.78
Average	60.39	92.28

without CP3D.

For the third step (performance improvement validation), the performance of models 1 and 2 are compared. The entire CP3D dataset consists of 421,000 annotated frames/images. Since the dataset has 14% as a test partition, around 52,600 images are used for testing. With the help of the training partition, two deep learning models were trained separately with and without CP3D. Both deep learning models are tested with the test partition, and the results are compared with the corresponding ground truth values. The discrepancies between the ground truth and the estimated values from model 1 are used to compare to those from model 2.

In terms of the performance improvement of CP3D over a generic daily activity dataset, we can see the results of the average in Table 9 and Table 11. From the average of the sample-to-sample difference for REBA trained with and without CP3D, results for all the activities demonstrate performance improvement for the CP3D-trained model, except for activities 6 and 7. Similarly, in terms of RULA, the CP3D-trained model exhibits better performance across all activities except for activity 9. Since all the models use COCO as part of the training datasets, MuCo-3DHP and CP3D are used to refer to models.

In terms of the standard deviation of our REBA and RULA ERA approaches, the results can be obtained from Table 10 and Table 12. By analyzing the standard deviation of the sample-to-sample differences between REBA models trained with and without CP3D, it is evident that the CP3D-trained model performs superiorly in all activities except for activities 6 and 7. Similarly, when examining RULA, the performance of the CP3D-trained model surpasses that of the non-CP3D model in all activities, with the exception of activities 5 and 9. The results prove the feasibility and performance of our CV-based ERA approaches because

Table 9
The average of the sample-to-sample difference for REBA & RULA, CP3D + MuCo-3DHP.

Task Number	Neck (degree)	Neck Twist	Neck Side Bend	Trunk (degree)	Trunk Twist	Trunk Side Bend	Leg (degree)	Upper Arm (degree)	Shoulder Raised	Upper Arm Abducted	Lower Arm (degree)	Wrist (degree)	Wrist Bend	REBA	RULA
1	7.04	0.36	0.00	4.87	0	0.12	5.53	8.15	0	0.25	8.39	11.65	0.12	1.17	0.01
2	9.83	0.70	0	4.25	0	0.02	5.67	11.95	0	0	14.62	29.33	0.71	0.98	0.12
3	9.36	0.70	0	3.20	0	0.15	5.32	5.53	0	0.17	11.37	6.08	0.01	1.01	0.70
4	8.78	0.52	0.03	6.57	0.00	0.03	3.92	6.43	0.00	0.31	6.86	11.90	0.12	1.64	0.01
5	11.92	0.68	0.00	4.16	0	0.12	4.89	7.95	0.00	0.22	10.54	11.79	0.08	0.68	0.60
6	11.62	0.01	0.32	7.05	0.14	0.05	4.21	19.81	0.04	0.00	15.62	13.47	0.06	2.35	0.01
7	5.96	0.39	0.01	7.97	0.00	0.20	7.47	9.61	0.00	0.29	18.41	25.96	0.30	1.40	0.05
8	4.21	0.08	0.05	4.00	0	0.36	4.83	10.45	0	0	10.17	16.78	0.13	1.07	0.19
9	7.88	0.30	0.14	6.16	0.00	0.03	4.00	10.16	0.07	0.08	8.87	13.55	0.14	1.60	0.33
10	5.60	0.16	0.00	3.18	0	0.32	6.17	13.83	0	0.21	19.05	15.89	0.14	1.10	0.44
11	7.02	0.58	0.00	4.13	0	0.1	4.32	9.37	0.00	0.20	11.20	13.29	0.18	1.13	0.71
12	10.15	0.59	0.00	6.03	0	0.17	2.69	8.76	0	0.00	14.92	17.56	0.03	1.47	0.06
13	8.57	0.38	0.02	4.36	0	0.06	3.84	7.57	0.00	0.00	14.16	17.01	0.12	0.88	0.02
14	10.84	0.47	0.00	11.10	0	0.14	14.82	11.02	0	0.20	9.50	14.45	0.17	1.20	0.00

Table 10
The standard deviation of the sample-to-sample difference for REBA & RULA, CP3D + MuCo-3DHP.

Task Number	Neck (degree)	Neck Twist	Neck Side Bend	Trunk (degree)	Trunk Twist	Trunk Side Bend	Leg (degree)	Upper Arm (degree)	Shoulder Raised	Upper Arm Abducted	Lower Arm (degree)	Wrist (degree)	Wrist Bend	REBA	RULA
1	7.10	0.48	0.01	4.60	0	0.32	5.41	8.31	0	0.43	6.83	11.67	0.32	1.05	0.34
2	4.74	0.45	0	2.87	0	0.15	4.94	9.68	0	0	11.52	15.80	0.45	0.95	0.32
3	4.74	0.45	0	2.17	0	0.36	4.30	4.23	0	0.38	8.80	5.58	0.08	0.80	0.64
4	5.51	0.50	0.19	5.31	0.02	0.16	2.46	5.63	0.04	0.46	5.35	11.46	0.32	1.29	0.11
5	6.02	0.46	0.05	3.42	0	0.33	4.26	7.19	0.04	0.42	11.59	9.17	0.28	0.70	0.60
6	7.70	0.10	0.47	5.63	0.36	0.23	3.17	19.59	0.21	0.02	13.47	14.51	0.25	1.31	0.13
7	4.49	0.49	0.10	5.92	0.06	0.40	7.63	8.20	0.01	0.45	23.12	18.82	0.46	1.33	0.23
8	3.46	0.27	0.23	2.60	0	0.48	3.72	8.32	0	0	7.52	13.99	0.34	0.94	0.40
9	5.54	0.46	0.35	6.53	0.03	0.18	3.33	8.31	0.26	0.29	8.47	16.13	0.36	1.61	0.47
10	3.21	0.37	0.02	3.17	0	0.47	5.33	13.43	0	0.41	20.26	12.67	0.35	0.90	0.50
11	4.12	0.49	0.05	2.44	0	0.30	3.77	8.96	0.05	0.41	7.96	11.64	0.38	0.93	0.62
12	5.21	0.49	0.02	4.35	0	0.38	2.63	6.10	0	0.03	11.36	16.41	0.18	0.99	0.25
13	5.67	0.49	0.15	3.02	0	0.24	3.67	6.54	0.01	0.02	11.11	14.37	0.33	0.81	0.14
14	6.61	0.50	0.01	9.52	0	0.35	7.91	8.82	0	0.41	7.74	12.17	0.38	1.02	0.07

the estimated parameters and risk scores are close to ground truth values. More importantly, the model trained with CP3D outperforms the model without CP3D in the case of construction activity ERA. Compared to the generic daily activity dataset MuCo-3DHP, the improvement that CP3D brings to ERA for construction activities is significant. Despite some minor discrepancies, the estimated joint angles follow the same pattern as the ground truth. This means that our CV-based method trained with CP3D can estimate the 3D joint angles, including the wrist angle. The estimated ERA scores also follow the same pattern as the ground truth. In addition, the deep learning model jointly trained with our dataset and MuCo-3DHP is more robust than the one trained only with MuCo-3DHP. For example, Fig. 11 shows the estimated (prediction) and ground truth (GT) neck angle for 3000 consecutive frames of a trial from activity 1.

According to Tables 4, 5, and 6, the deep learning model trained with our dataset performs better than the model trained without our dataset. Meanwhile, the feasibility and robustness of our ERA approach are demonstrated.

5. Conclusion

In this study, a new CV dataset, called CP3D, is created based on the 14 most common construction tasks observed in several construction facilities. This dataset fills in the research gap where no comprehensive dataset is available for construction activities. CP3D can be used with other public datasets to improve the performance of HPE and ERA methods. It is worth noting that our dataset contains annotated hand information, unlike the most commonly used 3D dataset, Human3.6 M. The inclusion of hand annotations allows CP3D to be applied to automated ERA tools such as REBA and RULA. The study also proposes a biomechanical skeletal model to ensure precise human body data acquisition concerning postural ergonomic hazards. Consequently, with the support of CP3D and precise human pose data acquisition, a CV-based ERA method, which incorporates REBA and RULA, is developed

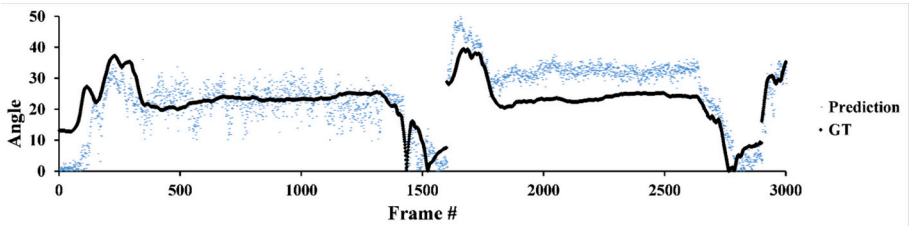


Fig. 11. The estimated and ground truth neck angle for a trial from activity 1.

Table 11
The average of the sample-to-sample difference for REBA & RULA, MuCo-3DHP.

Task Number	Neck (degree)	Neck Twist	Neck Side Bend	Trunk (degree)	Trunk Twist	Trunk Side Bend	Leg (degree)	Upper Arm (degree)	Shoulder Raised	Upper Arm Abducted	Lower Arm (degree)	Wrist (degree)	Wrist Bend	REBA	RULA
1	8.81	0.351	0.031	12.67	0	0.35	19.79	12.78	0	0.28	15.83	14.06	0.10	1.48	0.04
2	9.68	0.72	0.00	3.59	0.01	0.08	6.21	11.32	0.00	0.00	24.12	37.53	0.84	1.04	0.16
3	9.21	0.75	0.00	2.89	0.00	0.22	6.88	8.02	0.00	0.32	18.84	10.24	0.05	1.32	0.82
4	9.82	0.56	0.01	17.22	0.01	0.21	5.07	10.86	0.01	0.29	11.11	19.08	0.21	2.33	0.03
5	12.42	0.71	0.00	3.84	0.00	0.26	6.39	11.10	0.00	0.27	26.32	13.74	0.09	0.89	0.60
6	15.63	0.04	0.45	13.17	0.00	0.27	5.78	41.91	0.08	0.00	26.60	24.17	0.15	2.12	0.02
7	7.12	0.38	0.15	10.43	0.00	0.30	9.84	13.26	0.00	0.49	24.27	29.11	0.33	1.33	0.08
8	5.26	0.09	0.15	6.42	0.00	0.35	6.90	13.29	0	0.06	16.97	8.97	0.06	1.09	0.30
9	10.03	0.34	0.01	12.97	0.00	0.30	4.11	14.23	0.00	0.21	12.65	25.86	0.21	1.71	0.31
10	6.12	0.16	0	3.39	0.00	0.39	6.82	20.06	0	0.36	26.72	11.65	0.06	1.31	0.45
11	7.83	0.58	0.00	3.32	0.00	0.25	5.60	14.24	0.00	0.32	25.32	14.55	0.20	1.27	0.72
12	10.36	0.69	0.01	8.45	0	0.34	7.02	11.18	0	0.01	17.05	14.41	0.05	1.83	0.18
13	10.89	0.36	0.05	2.70	0	0.12	5.67	10.05	0.00	0.01	19.42	12.04	0.05	0.89	0.05
14	10.82	0.52	0	16.38	0.00	0.14	12.64	13.08	0	0.33	12.33	15.99	0.25	1.48	0.01

Table 12
The standard deviation of the sample-to-sample difference for REBA & RULA, MuCo-3DHP.

Task Number	Neck (degree)	Neck Twist	Neck Side Bend	Trunk (degree)	Trunk Twist	Trunk Side Bend	Leg (degree)	Upper Arm (degree)	Shoulder Raised	Upper Arm Abducted	Lower Arm (degree)	Wrist (degree)	Wrist Bend	REBA	RULA
1	10.38	0.48	0.18	9.01	0	0.48	19.20	13.65	0	0.45	14.18	13.32	0.31	1.36	0.35
2	4.97	0.45	0.06	2.42	0.11	0.28	5.44	10.20	0.03	0.06	17.44	14.08	0.37	0.96	0.37
3	4.58	0.43	0.02	2.14	0.06	0.42	5.53	7.91	0.02	0.47	28.19	15.38	0.22	0.99	0.67
4	6.86	0.50	0.11	11.77	0.10	0.40	10.63	0.11	0.46	10.17	15.56	0.41	10.63	1.68	0.18
5	5.25	0.45	0.06	3.52	0.07	0.44	5.43	9.34	0.05	0.44	27.02	11.89	0.29	0.89	0.59
6	10.05	0.19	0.50	8.33	0.06	0.45	3.89	42.05	0.28	0.04	17.43	23.84	0.36	1.23	0.15
7	4.79	0.49	0.36	10.78	0.06	0.46	9.46	10.43	0.03	0.50	30.75	17.96	0.47	1.16	0.27
8	4.48	0.29	0.36	4.31	0.01	0.48	5.14	12.57	0	0.24	13.40	9.44	0.24	0.99	0.46
9	8.44	0.47	0.09	9.67	0.01	0.46	3.08	12.80	0.02	0.41	13.47	27.47	0.41	1.40	0.46
10	3.72	0.37	0	2.93	0.04	0.49	5.86	17.14	0	0.48	26.09	10.19	0.23	1.11	0.50
11	6.16	0.49	0.06	2.20	0.03	0.43	4.74	14.46	0.04	0.47	25.88	13.84	0.40	1.15	0.67
12	5.15	0.46	0.08	5.40	0	0.47	3.92	9.64	0	0.11	15.31	12.95	0.21	1.22	0.39
13	6.90	0.48	0.21	2.36	0	0.33	4.31	10.07	0.05	0.10	19.52	10.13	0.22	0.83	0.21
14	6.45	0.50	0	10.54	0.01	0.34	10.12	10.08	0	0.47	12.01	14.73	0.43	1.18	0.08

and validated.

Regarding HPE for construction workers, the deep learning model jointly trained with CP3D and a popular generic daily activity dataset (MuCo-3DHP) performs better than the deep learning model solely trained with the generic daily activity datasets. The MPJPE results suggest a 35% performance increase when using CP3D together with MuCo-3DHP for construction activity posture estimation. Our dataset can be applied to ergonomic risks and safety hazard identifications in construction facilities. It is complementary to the generic daily activity datasets because it produces higher accuracy when our dataset is used with those datasets.

Regarding ERA, our CV-based REBA and RULA approaches achieve comparable results. In the meantime, compared to a deep learning model solely trained with MuCo-3DHP, our ERA approach performs better with a deep learning model jointly trained with CP3D and MuCo-3DHP. Hence, CP3D can improve the accuracy of CV-based ergonomic risks and safety hazard identification.

To further promote the research of three-dimensional pose estimation, ergonomic risk assessment, and productivity analysis for construction workers, we have made CP3D and associated software code publicly available at https://github.com/xinmingliUofA/CP3D.

6. Limitations and future work

Although the deep learning models containing our dataset improve the performance of HPE and ERA in construction activities compared to the deep learning model trained with generic activities, the accuracy can be improved. There are three main reasons for this drawback. First, the comprehensiveness of our dataset can be further enhanced by including more activities and more performers with more diverse demographic backgrounds. Second, inevitable errors are introduced in the MoCap data gap-filling process because gap-filling is an estimation instead of actual values. Third, occlusion is a challenging issue that the CV-based single monocular camera HPE/ERA method cannot overcome because occlusion causes information loss of the occluded body parts. Another limitation is related to the MoCap setup for data collection; the test data is collected in a lab environment, and the environment cannot fully represent actual construction facilities. CP3D was acquired within a laboratory setting due to illumination variations, the complexity of the work environment leading to occlusions, and the physical and psychological burdens associated with MoCap systems during data capture on a real construction facility, all hindering the creation of 'in-the-wild' datasets. It is worth noting that insufficient "in-the-wild" datasets for 3D HPE are still a challenge that needs to be solved. Similarities exist between CP3D training and testing partitions (they both consist of 14 activities), so the performance improvement might not be as noticeable when detecting other activities.

To overcome the limitations and further improve our dataset and ERA method, four improvements can be added to the study. First, we can include more motion capture cameras to reduce the number of occluded markers. Hence, reducing the number of gap fillings leads to fewer errors. More video cameras can improve the diversity of the dataset by capturing more images from different directions. The loss of information due to occlusion can also be solved by incorporating more monocular cameras into the CV-based ERA system because using other intrusive devices to solve this problem violates the objectives of this study.

Second, the test data can be collected on an actual construction facility. However, a MoCap system needs to be deployed on the construction facilities. For future data collection on construction facilities, it is preferable to use IMUs-based MoCap because the complex environment of construction facilities can cause severe occlusions and tripping hazards. Third, the quantitative performance of multi-person scenarios has not been obtained due to the lack of a ground truth multi-person construction dataset; only qualitative results have been obtained for the verification of multi-person ERA capability. Therefore, a multi-person construction activity benchmark dataset with more testing activities should be created in the future. Fourth, CP3D will be further enriched with more annotated data in the future.

CRediT authorship contribution statement

Chao Fan: Writing – original draft, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Qipei Mei:** Writing – review & editing, Supervision, Methodology, Investigation, Funding acquisition, Conceptualization. **Xinming Li:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization.

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.

Data availability

Data will be made available on request.

Acknowledgment

Funding: This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) through the Alliance Grant with Alberta Innovates [File No. ALLRP 561120 - 20].

References

[1] V. Patel, A. Chesmore, C.M. Legner, S. Pandey, Trends in workplace wearable technologies and connected-worker solutions for next-generation occupational safety, health, and productivity, Adv. Intell. Syst. 4 (2022) 2100099, https://doi.org/10.1002/aisy.202100099.
[2] Y. Tian, H. Li, H. Cui, J. Chen, Construction motion data library: an integrated motion dataset for on-site activity recognition, Sci. Data 9 (2022) 726, https://doi.org/10.1038/s41597-022-01841-1.
[3] H. Li, M. Lu, S.-C. Hsu, M. Gray, T. Huang, Proactive behavior-based safety management for construction safety improvement, Saf. Sci. 75 (2015) 107–117, https://doi.org/10.1016/j.ssci.2015.01.013.
[4] X. Li, S. Han, M. Gul, M. Al-Hussein, Automated ergonomic risk assessment based on 3D visualization, in: 34th International Symposium on Automation and Robotics in Construction, 2017, pp. 380–387, https://doi.org/10.22260/ISARC2017/0052.
[5] D. Wang, F. Dai, X. Ning, Risk assessment of work-related musculoskeletal disorders in construction: state-of-the-art review, J. Constr. Eng. Manag. 141 (2015) 04015008, https://doi.org/10.1061/(ASCE)CO.1943-7862.0000979.
[6] J. Seo, M. Moon, S. Lee, Construction operation simulation reflecting workers' muscle fatigue, in: Computing in Civil Engineering 2015, American Society of Civil Engineers, Austin, Texas, 2015, pp. 515–522, https://doi.org/10.1061/9780784479247.064.
[7] Association of Workers' Compensation Boards of Canada, National Work Injury/Disease Statistics Program (NWISP). https://awcbc.org/en/statistics/, 2023.
[8] U.S. Bureau of Labor Statistics, Occupational Injuries and Illnesses Resulting in Musculoskeletal Disorders (MSDs). https://www.bls.gov/iif/factsheets/msds.htm, 2020 (accessed on December 3, 2023).
[9] G.C. David, Ergonomic methods for assessing exposure to risk factors for work-related musculoskeletal disorders, Occup. Med. 55 (2005) 190–199, https://doi.org/10.1093/occmed/kqi082.
[10] M. MassirisFernández, J.Á. Fernández, J.M. Bajo, C.A. Delrieux, Ergonomic risk assessment based on computer vision and machine learning, Comput. Ind. Eng. 149 (2020) 106816, https://doi.org/10.1016/j.cie.2020.106816.
[11] P. Plantard, H.P.H. Shum, A.-S. Le Pierres, F. Multon, Validation of an ergonomic assessment method using Kinect data in real workplace conditions, Appl. Ergon. 65 (2017) 562–569, https://doi.org/10.1016/j.apergo.2016.10.015.
[12] N. Vignais, F. Bernard, G. Touvenot, J.-C. Sagot, Physical risk factors identification based on body sensor network combined to videotaping, Appl. Ergon. 65 (2017) 410–417, https://doi.org/10.1016/j.apergo.2017.05.003.
[13] S.Y. Guo, L.Y. Ding, H.B. Luo, X.Y. Jiang, A big-data-based platform of workers' behavior: observations from the field, Accid. Anal. Prev. 93 (2016) 299–309, https://doi.org/10.1016/j.aap.2015.09.024.
[14] A. Pidurkar, R. Sadakale, A.K. Prakash, Monocular camera based computer vision system for cost effective autonomous vehicle, in: In: 2019 10th International Conference on Computing, Communication and Networking Technologies, IEEE, Kanpur, India, 2019, pp. 1–5, https://doi.org/10.1109/ICCCNT45670.2019.8944496.
[15] C. Fan, Q. Mei, Q. Yang, X. Li, Computer-vision based rapid entire body analysis (REBA) estimation, in: Modular and Offsite Construction Summit Proceedings, 2022, pp. 90–97, https://doi.org/10.29173/mocs269.
[16] S. Jeong, J. Kook, CREBAS: computer-based REBA evaluation system for wood manufacturers using MediaPipe, Appl. Sci. 13 (2023) 938, https://doi.org/10.3390/app13020938.
[17] W. Fang, L. Ding, P.E.D. Love, H. Luo, H. Li, F. Peña-Mora, B. Zhong, C. Zhou, Computer vision applications in construction safety assurance, Autom. Constr. 110 (2020) 103013, https://doi.org/10.1016/j.autcon.2019.103013.
[18] E. Barberi, M. Chillemi, F. Cucinotta, D. Milardi, M. Raffaele, F. Salmeri, F. Sfravara, Posture interactive self evaluation algorithm based on computer vision, in: Proceedings of International Joint Conference on Mechanics, Design Engineering & Advanced Manufacturing, Springer International Publishing, 2023, pp. 1516–1526, https://doi.org/10.1007/978-3-031-15928-2_132.
[19] G.K. Nayak, E. Kim, Development of a fully automated RULA assessment system based on computer vision, Int. J. Ind. Ergon. 86 (2021) 103218, https://doi.org/10.1016/j.ergon.2021.103218.
[20] J. Seo, K. Yin, S. Lee, Automated postural ergonomic assessment using a computer vision-based posture classification, in: Construction Research Congress 2016, American Society of Civil Engineers, San Juan, Puerto Rico, 2016, pp. 809–818, https://doi.org/10.1061/9780784479827.082.
[21] C. Ionescu, D. Papava, V. Olaru, C. Sminchisescu, Human3.6M: large scale datasets and predictive methods for 3D human sensing in natural environments, IEEE Trans. Pattern Anal. Mach. Intell. 36 (2014) 1325–1339, https://doi.org/10.1109/TPAMI.2013.248.
[22] D. Mehta, O. Sotnychenko, F. Mueller, W. Xu, S. Sridhar, G. Pons-Moll, C. Theobalt, Single-shot multi-person 3D pose estimation from monocular RGB, in: 2018 International Conference on 3D Vision, IEEE, Verona, 2018, pp. 120–130, https://doi.org/10.1109/3DV.2018.00024.
[23] J. Liu, A. Shahroudy, M. Perez, G. Wang, L.-Y. Duan, A.C. Kot, NTU RGB+D 120: a large-scale benchmark for 3D human activity understanding, IEEE Trans. Pattern Anal. Mach. Intell. 42 (2020) 2684–2701, https://doi.org/10.1109/TPAMI.2019.2916873.
[24] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, C. L. Zitnick, Microsoft COCO: common objects in context, in: Proceedings of European Conference on Computer Vision 2014, Springer International Publishing, 2014, pp. 740–755, https://doi.org/10.1007/978-3-319-10602-1_48.
[25] D. Mehta, H. Rhodin, D. Casas, P. Fua, O. Sotnychenko, W. Xu, C. Theobalt, Monocular 3D human pose estimation in the wild using improved CNN supervision, in: 2017 International Conference on 3D Vision, IEEE, Qingdao, 2017, pp. 506–516, https://doi.org/10.1109/3DV.2017.00064.
[26] C. Zimmermann, D. Ceylan, J. Yang, B. Russell, M.J. Argus, T. Brox, FreiHAND: a dataset for markerless capture of hand pose and shape from single RGB images, in: 2019 IEEE/CVF International Conference on Computer Vision, IEEE, Seoul, Korea (South), 2019, pp. 813–822, https://doi.org/10.1109/ICCV.2019.00090.
[27] Y. Zhang, H. Wu, H. Liu, L. Tong, M.D. Wang, Improve model generalization and robustness to dataset bias with bias-regularized learning and domain-guided augmentation, arXiv preprint (2019), https://doi.org/10.48550/ARXIV.1910.06745 (accessed on July 9, 2023).
[28] C.F.G.D. Santos, J.P. Papa, Avoiding overfitting: a survey on regularization methods for convolutional neural networks, ACM Comput. Surv. 54 (2022) 1–25, https://doi.org/10.1145/3510413.
[29] S. Salman, X. Liu, Overfitting mechanism and avoidance in deep neural networks, arXiv preprint (2019), https://doi.org/10.48550/ARXIV.1901.06566 (accessed on July 12, 2023).
[30] D. Hendrycks, N. Mu, E.D. Cubuk, B. Zoph, J. Gilmer, B. Lakshminarayanan, AugMix: a simple data processing method to improve robustness and uncertainty, arXiv preprint (2019), https://doi.org/10.48550/ARXIV.1912.02781 (accessed on July 20, 2023).
[31] D. Hendrycks, S. Basart, N. Mu, S. Kadavath, F. Wang, E. Dorundo, R. Desai, T. Zhu, S. Parajuli, M. Guo, D. Song, J. Steinhardt, J. Gilmer, The many faces of robustness: A critical analysis of out-of-distribution generalization, in: 2021 IEEE/CVF International Conference on Computer Vision, IEEE, Montreal, QC, Canada, 2021, pp. 8320–8329, https://doi.org/10.1109/ICCV48922.2021.00823.
[32] P. Wang, W. Li, C. Li, Y. Hou, Action recognition based on joint trajectory maps with convolutional neural networks, Knowl.-Based Syst. 158 (2018) 43–53, https://doi.org/10.1016/j.knosys.2018.05.029.
[33] C. Li, Y. Hou, P. Wang, W. Li, Joint distance maps based action recognition with convolutional neural networks, IEEE Sign. Process. Lett. 24 (2017) 624–628, https://doi.org/10.1109/LSP.2017.2678539.
[34] S.E. Wei, V. Ramakrishna, T. Kanade, Y. Sheikh, Convolutional pose machines, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,

IEEE, Las Vegas, NV, USA, 2016, pp. 4724–4732, https://doi.org/10.1109/CVPR.2016.511.
[35] A. Newell, K. Yang, J. Deng, Stacked hourglass networks for human pose estimation, in: Proceedings of European Conference on Computer Vision, Springer International Publishing, 2016, pp. 483–499, https://doi.org/10.1007/978-3-319-46484-8_29.
[36] A. Toshev, C. Szegedy, DeepPose: human pose estimation via deep neural networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, IEEE, Columbus, OH, USA, 2014, pp. 1653–1660, https://doi.org/10.1109/CVPR.2014.214.
[37] W. Yang, W. Ouyang, X. Wang, J. Ren, H. Li, X. Wang, 3D human pose estimation in the wild by adversarial learning, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, IEEE, Salt Lake City, UT, USA, 2018, pp. 5255–5264, https://doi.org/10.1109/CVPR.2018.00551.
[38] J. Martinez, R. Hossain, J. Romero, J.J. Little, A simple yet effective baseline for 3d human pose estimation, in: 2017 IEEE International Conference on Computer Vision, IEEE, Venice, 2017, pp. 2659–2668, https://doi.org/10.1109/ICCV.2017.288.
[39] G. Pavlakos, X. Zhou, K.G. Derpanis, K. Daniilidis, Coarse-to-fine volumetric prediction for single-image 3D human pose, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, Honolulu, HI, 2017, pp. 1263–1272, https://doi.org/10.1109/CVPR.2017.139.
[40] J. Tompson, A. Jain, Y. LeCun, C. Bregler, Joint training of a convolutional network and a graphical model for human pose estimation, in: Proceedings of the 27th International Conference on Neural Information Processing Systems, 2014, pp. 1799–1807, https://doi.org/10.48550/ARXIV.1406.2984.
[41] G. Moon, J.Y. Chang, K.M. Lee, Camera distance-aware top-down approach for 3D Multi-person pose estimation from a single RGB image, in: 2019 IEEE/CVF International Conference on Computer Vision, IEEE, Seoul, Korea (South), 2019, pp. 10132–10141, https://doi.org/10.1109/ICCV.2019.01023.
[42] X. Yang, Y. Tian, Effective 3D action recognition using EigenJoints, J. Vis. Commun. Image Represent. 25 (2014) 2–11, https://doi.org/10.1016/j.jvcir.2013.03.001.
[43] J. Yan, M. Zhou, J. Pan, M. Yin, B. Fang, Recent advances in 3D human pose estimation: from optimization to implementation and beyond, Int. J. Pattern Recognit. Artif. Intell. 36 (2022) 2255003, https://doi.org/10.1142/S0218001422550035.
[44] C. Zheng, W. Wu, C. Chen, T. Yang, S. Zhu, J. Shen, N. Kehtarnavaz, M. Shah, Deep learning-based human pose estimation: a survey, ACM Comput. Surv. 56 (2024) 1–37, https://doi.org/10.1145/3603618.
[45] X. Ji, Q. Fang, J. Dong, Q. Shuai, W. Jiang, X. Zhou, A survey on monocular 3D human pose estimation, Virtu. Realit. Intellig. Hardw. 2 (2020) 471–500, https://doi.org/10.1016/j.vrih.2020.04.005.
[46] M. Ben Gamra, M.A. Akhloufi, A review of deep learning techniques for 2D and 3D human pose estimation, Image Vis. Comput. 114 (2021) 104282, https://doi.org/10.1016/j.imavis.2021.104282.
[47] W. Liu, Q. Bao, Y. Sun, T. Mei, Recent advances of monocular 2D and 3D human pose estimation: a deep learning perspective, ACM Comput. Surv. 55 (2023) 1–41, https://doi.org/10.1145/3524497.
[48] Y. Desmarais, D. Mottet, P. Slangen, P. Montesinos, A review of 3D human pose estimation algorithms for markerless motion capture, Comput. Vis. Image Underst. 212 (2021) 103275, https://doi.org/10.1016/j.cviu.2021.103275.
[49] J. Wang, S. Tan, X. Zhen, S. Xu, F. Zheng, Z. He, L. Shao, Deep 3D human pose estimation: a review, Comput. Vis. Image Underst. 210 (2021) 103225, https://doi.org/10.1016/j.cviu.2021.103225.
[50] N. Sarafianos, B. Boteanu, B. Ionescu, I.A. Kakadiaris, 3D human pose estimation: a review of the literature and analysis of covariates, Comput. Vis. Image Underst. 152 (2016) 1–20, https://doi.org/10.1016/j.cviu.2016.09.002.
[51] G. Rogez, P. Weinzaepfel, C. Schmid, LCR-net: Localization-classification-regression for human pose, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, Honolulu, HI, 2017, pp. 1216–1224, https://doi.org/10.1109/CVPR.2017.134.
[52] A. Zanfir, E. Marinoiu, M. Zanfir, A.-I. Popa, C. Sminchisescu, Deep network for the integrated 3D sensing of multiple people in natural images, in: Advances in Neural Information Processing Systems, 2018, in: https://proceedings.neurips.cc/paper_files/paper/2018/file/6a6610feab86a1f294dbbf5855c74af9-Paper.pdf (accessed on July 9, 2023).
[53] N. Mahmood, N. Ghorbani, N.F. Troje, G. Pons-Moll, M. Black, AMASS: archive of motion capture as surface shapes, in: 2019 IEEE/CVF International Conference on Computer Vision, IEEE, Seoul, Korea (South), 2019, pp. 5441–5450, https://doi.org/10.1109/ICCV.2019.00554.
[54] T. Von Marcard, R. Henschel, M.J. Black, B. Rosenhahn, G. Pons-Moll, Recovering accurate 3D human pose in the wild using IMUs and a moving camera, in: Proceedings of the European conference on computer vision 2018, Springer International Publishing, 2018, pp. 614–631, https://doi.org/10.1007/978-3-030-01249-6_37.
[55] Y. Huang, M. Kaufmann, E. Aksan, M.J. Black, O. Hilliges, G. Pons-Moll, Deep inertial poser: learning to reconstruct human pose from sparse inertial measurements in real time, ACM Trans. Graph. 37 (2018) 1–15, https://doi.org/10.1145/3272127.3275108.
[56] M. Trumble, A. Gilbert, C. Malleson, A. Hilton, J. Collomosse, Total capture: 3D human pose estimation fusing video and inertial sensors, in: Proceedings of the British Machine Vision Conference 2017, British Machine Vision Association, London, UK, 2017, p. 14, https://doi.org/10.5244/C.31.14.
[57] H. Joo, H. Liu, L. Tan, L. Gui, B. Nabbe, I. Matthews, T. Kanade, S. Nobuhara, Y. Sheikh, Panoptic studio: A massively multiview system for social motion capture, in: 2015 IEEE International Conference on Computer Vision, IEEE, Santiago, Chile, 2015, pp. 3334–3342, https://doi.org/10.1109/ICCV.2015.381.
[58] T.V. Marcard, G. Pons-Moll, B. Rosenhahn, Human pose estimation from video and IMUs, IEEE Trans. Pattern Anal. Mach. Intell. 38 (2016) 1533–1547, https://doi.org/10.1109/TPAMI.2016.2522398.
[59] L. Sigal, A.O. Balan, M.J. Black, HumanEva: synchronized video and motion capture dataset and baseline algorithm for evaluation of articulated human motion, Int. J. Comput. Vis. 87 (2010) 4–27, https://doi.org/10.1007/s11263-009-0273-6.
[60] Occupational Safety and Health Administration, Identifying and addressing Ergonomic Hazards Workbook. https://www.osha.gov/sites/default/files/2018-12/fy15_sh-27643-sh5_ErgonomicsWorkbook.pdf, 2023 accessed on April 27.
[61] M. Ciccarelli, A. Papetti, C. Scoccia, G. Menchi, L. Mostarda, G. Palmieri, M. Germani, A system to improve the physical ergonomics in human-robot collaboration, Proced. Comput. Sci. 200 (2022) 689–698, https://doi.org/10.1016/j.procs.2022.01.267.
[62] L. Li, X. Xu, A deep learning-based RULA method for working posture assessment, Proceed. Human Fact. Ergonom. Soc. Ann. Meet. 63 (2019) 1090–1094, https://doi.org/10.1177/1071181319631174.
[63] Y. Yu, H. Li, W. Umer, C. Dong, X. Yang, M. Skitmore, A.Y.L. Wong, Automatic biomechanical workload estimation for construction workers by computer vision and smart insoles, J. Comput. Civ. Eng. 33 (2019) 04019010, https://doi.org/10.1061/(ASCE)CP.1943-5487.0000827.
[64] B. Parsa, A.G. Banerjee, A multi-task learning approach for human activity segmentation and ergonomics risk assessment, in: 2021 IEEE Winter Conference on Applications of Computer Vision, IEEE, Waikoloa, HI, USA, 2021, pp. 2351–2361, https://doi.org/10.1109/WACV48630.2021.00240.
[65] Y. Yu, H. Li, X. Yang, W. Umer, Estimating construction workers' physical workload by fusing computer vision and smart insole technologies, in: Proceedings of the International Symposium on Automation and Robotics in Construction, Taipei, Taiwan, 2018, pp. 1212–1219, https://doi.org/10.22260/ISARC2018/0168.
[66] S. Hignett, L. McAtamney, Rapid entire body assessment (REBA), Appl. Ergon. 31 (2000) 201–205, https://doi.org/10.1016/S0003-6870(99)00039-3.
[67] L. McAtamney, E. Nigel Corlett, RULA: a survey method for the investigation of work-related upper limb disorders, Appl. Ergon. 24 (1993) 91–99, https://doi.org/10.1016/0003-6870(93)90080-S.
[68] A. Altieri, S. Ceccacci, A. Talipu, M. Mengoni, A low cost motion analysis system based on RGB cameras to support ergonomic risk assessment in real workplaces, in: 40th Computers and Information in Engineering Conference, American Society of Mechanical Engineers, Virtual, 2020, https://doi.org/10.1115/DETC2020-22308 p. V009T09A067. Online.
[69] W. Kim, J. Sung, D. Saakes, C. Huang, S. Xiong, Ergonomic postural assessment using a new open-source human pose estimation technology (OpenPose), Int. J. Ind. Ergon. 84 (2021) 103164, https://doi.org/10.1016/j.ergon.2021.103164.
[70] D. Kee, Systematic comparison of OWAS, RULA, and REBA based on a literature review, Int. J. Environ. Res. Public Health 19 (2022) 595, https://doi.org/10.3390/ijerph19010595.
[71] Vicon Motion Systems Ltd, Vicon Nexus. https://www.vicon.com/software/nexus/, 2023 accessed on July 10.
[72] R. Aliasgari, A framework to automate physical demand analysis based on artificial intelligence and motion capture for workplace safety improvement, Univ. Alberta 2022 (2023), https://doi.org/10.7939/r3-az1s-n184 (accessed on July 10.
[73] U.S. Bureau of Labor Statistics, Construction Industry Employees by Sex, 2003 to 2020. https://www.bls.gov/spotlight/2022/the-construction-industry-labor-force-2003-to-2020/home.htm, 2022 (accessed on May 4, 2023).
[74] FFmpeg Team, FFmpeg. https://ffmpeg.org/, 2023 accessed on July 11.
[75] C.Y. Wang, A. Bochkovskiy, H.Y.M. Liao, YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, IEEE, Vancouver, BC, Canada, 2023, pp. 7464–7475, https://doi.org/10.1109/CVPR52729.2023.00721.
[76] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, IEEE, Las Vegas, NV, USA, 2016, pp. 770–778, https://doi.org/10.1109/CVPR.2016.90.
[77] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A.C. Berg, L. Fei-Fei, ImageNet large scale visual recognition challenge, Int. J. Comput. Vis. 115 (2015) 211–252, https://doi.org/10.1007/s11263-015-0816-y.
[78] D.P. Kingma, J. Ba, Adam: a method for stochastic optimization, arXiv preprint (2014), https://doi.org/10.48550/ARXIV.1412.6980 (accessed on July 20, 2023).