CO7080

Joint-Level Vision-Based Ergonomic Assessment Tool for Construction Workers

Yantao Yu, Xincong Yang, Heng Li, Xiaochun Luo, Hongling Guo, Qi Fang

Yantao Yu: Ph.D. Student, Dept. of Building and Real Estate, Hong Kong Polytechnic Univ., Hung Hom, Kowloon, Hong Kong. E-mail: yt.yu@connect.polyu.hk

Xincong Yang: Ph.D. Candidate, Dept. of Building and Real Estate, Hong Kong Polytechnic Univ., Hung Hom, Kowloon, Hong Kong. E-mail: xincong.yang@outlook.com

Heng Li: Chair Professor, Dept. of Building and Real Estate, Hong Kong Polytechnic Univ., Hung Hom, Kowloon, Hong Kong. Email: bshengli@polyu.edu.hk

Xiaochun Luo: Senior Research Fellow, Dept. of Building and Real Estate, Hong Kong Polytechnic Univ., Hung Hom, Kowloon, Hong Kong. E-mail: bsericlo@polyu.edu.hk

Hongling Guo: Associate Professor, Dept. of Construction Management, Tsinghua University, Beijing, China. E-mail: hlguo@tsinghua.edu.cn

Qi Fang: Ph.D. Student, School of Civil Engineering &Mechanics, Huazhong University of Science & Technology, Wuhan, China. E-mail: fangqi_hust@hotmail.com

Abstract

Construction workers are commonly subjected to ergonomic. Accurate ergonomic assessment is needed to reduce ergonomic risks. However, the diverse and dynamic nature of construction site makes it difficult to collect workers posture data for ergonomic assessment without intrusiveness. The Joint-level Vision-based Ergonomic assessment tool for Construction workers (JVEC), therefore, is proposed to provide automatic and detailed ergonomic assessment for construction workers based on construction videos. JVEC extracts construction workers' skeleton data from videos with advanced deep learning methods, then Rapid Entire Body Assessment (REBA) is used to conduct the joint-level ergonomic assessment. This approach was demonstrated and tested with a laboratory experiment and an on-site experiment, which indicates that the accuracy of ergonomic risk scores (70-96%) and the feasibility of construction sites. This research contributes to an accurate and non-intrusive ergonomic assessment method for construction workers. In addition, this research, for the first time, introduces the 2D video-based 3D pose estimation algorithms to the construction industry, which may benefit research on construction health, safety, and productivity by providing long-time and accurate behavior data.

Keyword: construction, worker, ergonomic risks, computer vision, deep learning, occupational safety and health, 3D posture estimation.

Introduction

The construction industry is one of the most dangerous industries due to the high rate of work-related injuries and death. Compared with other industries, construction workers suffer two times more work-related injuries (Entzel et al., 2007). The incident rate of nonfatal occupational injuries and illnesses involving days away from work in the US construction sector reached up to 34.6% in 2015 (U.S. Bureau of Labor Statistics, 2016). Work-related musculoskeletal disorders (WMSD) are the biggest work-related health issues (Kim, 2017). In the US, WMSDs are the main causes of absence from work (Punnett et al., 2004). Sweden and Brazil also share the same trends (Kim, 2017). Besides, aging of the labor force, rising labor wages and a shortage of manpower have been becoming new challenges to the construction industry. Working ergonomically can help improve productivity and relieve labor force issues. Therefore, it is necessary to make the current construction workforce sustainable by improving occupational health.

Ergonomic interventions have been proved effective to prevent work-related injuries (Entzel et al., 2007). Efficient ergonomic assessment is the prerequisite for ergonomic interventions. However, due to the large variety of construction activities and the complexity of construction sites, it is difficult to get posture data accurately and continuously, let alone provide an accurate ergonomic assessment (Wang et al., 2015). Manual observation and 3D pose estimation sensors have been applied to capture working posture data. Although these studies have proven the concept, they might be inaccurate and intrusive in nature. In addition, the result of manual observation is subjective. Different observers might give different assessments based on the same posture. The results of 3D pose estimation sensors are more objective and accurate, but this method requires multiple sensors to be attached to the body of construction workers, which might result in uncomfortableness and irritation.

In order to achieve nonintrusive and accurate ergonomic assessment for construction workers, this research aims to develop a joint-level vision-based ergonomic assessment tool for construction workers (JVEC). JVEC contains a 3D pose estimator and a REBA (Rapid Entire Body Assessment)-based ergonomic risk score module. The 3D (three-dimensional) pose estimator is composed of two deep learning networks, which are able to get 3D joint coordinates from the videos related to construction workers. Such a data collection method makes it possible to collect construction workers' postures on sites instead of in laboratories, providing a solid foundation for ergonomic assessment and improvements. REBA-based ergonomic risk score module transfers the 3D joint coordinates to joint angles, then utilizes REBA to provide a joint-level ergonomic risk score for the whole body. By combining the deep learning 3D pose estimator and REBA rules, JVEC realizes joint-level ergonomic assessments based on the data from real construction sites.

The remainder of this paper is arranged as follows. First, a review is given to provide a summary and comparison of previous ergonomic assessment methods from the perspectives of data collection and ergonomic assessment rules, revealing the current research gaps of ergonomic assessment for construction workers. Then the research methodology is explained. Thirdly, the design and results of a laboratory experiment and an on-site experiment are given, followed by a discussion about the contributions and limitations of the methodology. Finally, a conclusion is drawn.

Literature review

Previous ergonomic assessment methods can be divided into the self-report-based methods and data-based analysis. Data-based analysis methods can be further divided into optical and non-optical (Nunes et al., 2015). These methods have various strengths and weakness. The aim of this review is to compare the above methods and select the one that best fit the data requirements of ergonomic assessment as well as the harsh environments on construction sites.

Self-report-based ergonomic assessment method

The self-report-based ergonomic assessment method focuses on workers' subjective ergonomic feelings and self-assessment on physical discomfort. For example, Corlett and Bishop (1976) developed a scale named "body map" to score the discomfort level of each body part. (Borg, 1998) introduced the Borg RPE scale (Borg Rating of Perceived Exertion Scale) to describe the workers' perceived workload. (Åhsberg and Gamberale, 1998) used Swedish Occupational Fatigue Inventory (SOFI) for workers to assess the fatigue level. Above the methods provide effective tools for workers' ergonomic-assessment. However, as these methods mainly focus on workers' subjective feelings, making the results not general enough to support ergonomic improvements. Besides, questionnaires and interviews are the main data collection approaches in these methods (Wiktorin et al., 1993), making data collection time- and effort- consuming.

Optical posture data collection methods for ergonomic assessment

Manual observation

The observation-based ergonomic assessment method classifies working postures through manual observation, then rates ergonomic rates based on posture category. Relevant ergonomic assessment tools provide objective and general approaches to assess workers' ergonomics. RULA (Rapid Upper Limb Assessment), OWAS (Ovako Working Posture Analyzing System) and REBA (Rapid Entire Body Assessment) are three popular ergonomic assessment tools (Hignett and McAtamney, 2000; Karhu et al., 1981; McAtamney and Nigel Corlett, 1993). RULA provides upper limb ergonomic assessment based on force and muscle activities. It is easy and convenient for ergonomic assessment, but not suitable enough for the construction industry, because construction activities are not only limited to upper limb postures and lifting tasks. OWAS is an assessment tool covering the whole body, which provides score based on the posture of each body part, such as "back bent slightly" or "back bent heavily" (Yan et al., 2017). It is easy and convenient but is not objective enough. OWAS describes postures with ambiguous words such as "slightly" and "heavily" without providing a clear boundary. REBA provides ergonomic risk scores according to objective joint angles and benefits the sequent automation process. Besides, REBA provides multi-level results, including joint level, body-part level, whole-body level. Such results are more comprehensive for ergonomic improvements. Based on the above strengths, REBA is selected as the ergonomic assessment rule. However, visual observations are imprecise and subjective, leading to less dependable fatigue evaluation (Alavinia et al., 2007; Valero et al., 2016). So, a more reliable data collection methods is required to provide accurate REBA-based ergonomic risk assessment.

Marker-based 3D pose estimation methods

Marker-based 3D pose estimation system requires certain markers (usually Light-emitting diode markers) attached to the subjects to emit signals and devices placed in the surrounding spaces to receive signals (Nunes et al., 2011). The marker-based method can provide highly accurate 3D pose estimation results, but not suitable for construction site environments due to three reasons: 1) the system requires careful calibration and highly controlled laboratory condition. 2) the system requires expensive multi-camera, which will increase the cost of the ergonomic assessment. 3) the markers are intrusive in nature and might lead to construction workers' uncomfortableness and irritation.

Markerless 3D pose estimation methods

Considering the intrusiveness of marker-based 3D pose estimation system, markerless 3D pose estimation methods have been proposed to get accurate posture data in a non-intrusive way. These methods generate accurate data by identifying surface features. According to different vision data sources, this section divides the markerless 3D pose estimation methods into depth camera methods and RGB camera methods.

Depth camera. Depth cameras have also been successfully applied in ergonomic risk evaluation (Dutta, 2012; Ning and Guo, 2013; Seo et al., 2016). More specifically, several previous studies have applied RGBD (RGB-Depth) sensors in construction management, for example, to use to recognize workers action (Ray and Teizer, 2012) to use depth cameras to capture worker joint angles (Yu et al., 2017; Zhu et al., 2014). Depth camera is turned out to be an efficient tool to collect posture data for ergonomic evaluation but is not appropriate for construction sites because depth cameras cannot work under direct sunlight.

RGB camera. Other research extracted posture information based on RGB pictures from ordinary cameras. For example, S. Han et al. (2013) has successfully classified workers' postures with a double lens camera; Seo et al. (2015) proposed computer vision-based framework to identify construction activities from 2D image sequences. S. U. Han et al. (2012) and Seo, Yin, et al. (2016)

applied cameras to capture the worker's posture. However, since the pictures just contain 2D information, one can only identify worker's postures such as standing and squatting, rather than joint angles, therefore making it difficult to assess ergonomics more detailed. Yan et al. (2017) successfully applied view-invariant features to classify postures as "slightly", "normally" and "heavily" danger, and provided ergonomic assessments with OWAS. However, the ambiguous classification standards may result in subjective ergonomic risk assessment results. As mentioned previously, REBA is more objective than OWAS, because it scores the ergonomic risks based on joint angle values. However, the posture capture method in (Yan et al. 2017) cannot provide data accurate enough for REBA. A new research also shows the necessity of applying 3D posture-based ergonomic assessment (Li et al., 2018), which visualized the 3D posture of construction workers and combined it with REBA for ergonomic assessment. The visualization process making the ergonomic assessment more accurate and intuitional. The methodology was proved to be feasible in the laboratory. However, as construction activities are of great variety and different workers have different work habits, it is necessary to provide ergonomic assessments based on data from real construction sites.

Non-optical posture data collection methods for ergonomic assessment

A variety of sensors have been applied for ergonomic assessments, including posture-based approaches and physical indicator approaches. Posture-based approaches collect motion data through sensors and use the above ergonomic tools to find out ergonomic risk; physical indicator approaches utilize sensors to collect workers physical signals that can reflect fatigue or nonergonomic situations, such as electromyography (EMG) and heart rate (HR).

Posture-based approaches. Wearable sensors were used to collect more precise worker's posture data (Nath et al., 2017; Yan et al., 2017). One of the widely-used 3D pose estimation sensor systems is inertial measurement unit (IMU) (Alwasel et al., 2017; Sedighi Maman et al., 2017; Valero et al., 2016). If attached to key joints, IMU sensors are able to capture the location and acceleration of the joints, and the human body motion data can thus be retrieved (Yan et al. 2017). Previous research has successfully used IMU to collect posture data and use it in ergonomic assessment (Valero et al. 2016; Nath et al. 2017). IMU can collect motion data automatically and continuously, making the ergonomic evaluation more convincing. The main disadvantage is the intrusiveness. IMU sensors are required to be tied tightly to the human body, but from the view of the application, workers may reject wearing sensors so tightly. Such sensors are feasible for the short-period track but may instigate irritation if used for long-time application (Golabchi et al., 2016; Valero et al., 2016).

Physical indicator-based approaches. Physical indicator-based approaches analyze ergonomics by directly measuring the physical indicators during work. Electromyography (EMG) and heart rate (HR) are two frequently used indicators (Antwi-Afari et al., 2017; Hwang et al., 2016; Umer et al., 2017). Besides, if combined with work-related postures, the physical indicators can reflect ergonomic situations more accurately (Cheng et al., 2013). These indicators are easy to be measured with commercially available sensors. However, because of the intrusiveness, these assessments can only be performed in laboratories, focusing on certain body parts and certain work postures. As a result, these methods are too microcosmic to assess a worker's ergonomic status of the whole body. Besides, EMG and HR sensors require direct contact with skin, limiting the application on construction sites.

To conclude, the above 3D pose estimation methodologies are not suitable for ergonomic assessment in the construction industry. The non-optical methods and the marker-based methods are intrusive and may result in uncomfortableness. As far as the markerless methods, depth cameras can provide 3D posture data but cannot work in the outdoor environment. Previous RGB camera-based methods can only generate 2D posture data. A recent progress in computer vision shed lights on RGB camera-based 3D joint level ergonomic assessment. The computer vision algorithm can estimate 3D human skeleton from 2D video frames, making it possible to collect joint angles just based on videos (Zhou et al., 2017). By combining the state-of-art computer vision algorithm and the widely-used REBA, this research aims at developing an accurate, real-time, and non-intrusive ergonomic assessment method which is suitable for both indoor and outdoor environments.

Methodology

JVEC consists of a 3D posture estimator and a REBA score calculator. Figure 1 illustrates the framework of the methodology. First, a 3D pose estimator is trained to estimate workers' joint 3D coordinates from 2D images. The method is based on an advanced deep-learning algorithm for human-body detection (Zhou et al. 2017). Then, joint angles are calculated based on the 3D coordinators. Finally, by comparing the joint angles and the REBA rules, the ergonomic status of a posture during construction tasks can be evaluated (Hignett and McAtamney, 2000). An experiment is designed to test the accuracy through the comparing with IMU.

Computer vision-based 3D posture estimator

To make the proposed method non-intrusive, the motion data must be collected automatically without interfering with work activities. This involves the use of computer vision. The workflow of the 3D pose estimator is shown in Figure 2. First, RGB images are collected from construction video clips. A deep leaning architecture, named hourglass network (Newell et al. 2016), is trained to estimate the 2D coordinates of joints. Then the joint length ratio constraints (set as the average of all subjects of Human 3.6M dataset), are used to estimate the 3D coordinates (Zhou et al. 2017). The following is a more detailed explanation.

2D joint position estimation. Stacked hourglass network architecture is applied to estimate the 2D joint capacity position. The framework is presented in Figure 2. The architecture contains repeated hourglass structures. The first hourglass structure takes an RGB image as input and joint heat maps as output. As for the other hourglass structure, both the input and output are joint heat maps. The goal of each heat map is to minimize the difference between the estimated joint positions and the ground truth, as shown in Eq.1.

$$L_{2D}(x,y) = \sum\sum\left(\hat{P}(x,y) - P_0(x,y)\right)^2, x \in [0,W], y \in [0,H] \quad \text{Eq.1}$$

where

$\hat{P}(x, y)$ represents the estimated heat map, i.e. the possibility distribution of the joint on each pixel;

$P_0(x, y)$ represents the real possibility distribution;

W and H is the width and height of a picture, and (x, y) represents pixel position.

Figure 3 illustrates the detailed structure of each hourglass structure. In an hourglass structure, the input, an RGB image, is first compressed to extract features and then decompressed to consolidate the features for inferring human joint positions. Input compression includes convolution and max-pooling, as shown in Figure 3. An RGB image is essentially a matrix denoted as $\boldsymbol{M}_{\boldsymbol{0}(w_0 * h_0 * 3)}$. The features can be extracted through multiplying a weight matrix $\boldsymbol{W_1}$, and different features can be extracted with different matrices. In Figure 3(a), for example, k features are extracted by multiplying $(\boldsymbol{W_1}, \boldsymbol{W_2}, \dots, \boldsymbol{W_k})$. To improve the non-linearity of the model and decrease the amount of calculation, max pooling is used to reduce the size of convolutional layer. Figure 3(a) provides an intuitive explanation of max pooling. The value of the four pixels on the top left corner is 5,1,9,2. In the max pooling layer, the four pixels are represented by only one pixel, and the value is set as the maximum of the original four pixels, that is, 9. The picture will be ultimately compressed to lowest resolution at 4×4 pixels. In the decompression period, the low-resolution results will be restored to the original size through nearest neighbor up-sampling and adding features across scales. Figure 3(b) shows the up-sampling process. The method first enlarges the image size by, for example, three times.

Then, the convolutional layer of the same size will be added to the up-sampled layer. This step makes sure that the model involves into not only the partial features of the image through convolution but also the features of the whole image through adding the features across scales. After repetitive down-sampling and up-sampling, each pixel has its own score describing the possibility of whether the pixel belongs to a joint, for example, the right knee joint. The score is basically a weighted average of the image information of each pixel. The training process is used to calculate the weights by minimizing the differences between the predicted score (0~1) with the ground truth score (0 or 1). After being trained, the network could be used to calculate the right knee score for each pixel in a new picture, then the heap map of the right knee can be retrieved (Newell et al., 2016).

3D joint position estimation. Based on the estimated 2D joint heat map, a model is trained to estimate the depth of each joint according to the 2D joint positions and the geometry constraints of the joints. The objective of the model training is to minimize the following loss function Eq.2. The basic idea of the geometry constraints is that the bone length of a person should be fixed. Based on the idea, the method aims to minimize the length difference between the bones of an identified skeleton and the ones of a standard skeleton. Eq.2 is the loss function of the arm group A, including left/right upper/lower arm bone, represented by $bone \in A$. n represents the number of elements belonging to A, which equals to 4. For each bone, the estimated length is $\hat{l}_{bone}$. The standard length is l_{bone}, which is calculated as the average of the length of all the samples in the database. r_{bone} is the ratio between $\hat{l}_{bone}$ and l_{bone}.

$$L_{depth}\left(\hat{Y}_{depth}\middle|\hat{P}(x,y)\right)=\frac{1}{n}\sum_{bone\in A}\left(r_{bone}-\overline{r}_{bone}\right)^2 \qquad \text{Eq.2}$$

$$r_{bone} = \frac{\hat{l}_{bone}}{l_{bone}}, \overline{r}_{bone} = \frac{1}{n} \sum_{bone \in A} r_{bone}$$

In this step, 16 key joints will be selected as the research subjects, i.e. J=16, as shown in Figure 2. There are totally 18 joints in Figure 2 because joint 3 neck and joint 13 hip are the middle points of join2/4 and 12/16 respectively.

Validation of the posture capture method

An experiment was enacted to validate the accuracy of the above method. IMU sensors were applied to collect the ground truth of joint positions. During the experiments, 13 sensors were tied to the subject's key joints. At the same time, a common RGB camera collected videos simultaneously. Then, the IMU data was transferred to skeleton data, and the skeleton data was extracted from the RGB video with the 3d posture estimator method. The skeleton data from IMU at time t is denoted as $\boldsymbol{X}(t)$**,** where $\boldsymbol{X} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_{13})$ is a 13×3 matrix representing the 3D coordinates of 13 joints. The predicted skeleton is denoted as $\boldsymbol{Y}(t)$**,** which is a similar 13×3 matrix. The prediction accuracy is defined as the average of Euclidean distance between **X** and ***Y***, i.e.

$$Dist = \frac{1}{13t} \sum_{t} \left\| \mathbf{X}(t) - \mathbf{Y}(t) \right\| \qquad \text{Eq.3}$$

The REBA-based WMSD risk score

REBA is an ergonomic assessment tool mainly based on joint angles. To utilize REBA in this research, the captured 3D joint coordinates must be transferred to joint parameters as required by REBA. Then according to these parameters, REBA can provide joint level, body part level, and whole-body level ergonomic assessments.

Joint angle calculation

The joint parameters required by REBA are presented in Figure 4. REBA divides the human body into five parts including trunk, neck, leg, upper arm and lower arm. The following part explains how to calculate the above parameters.

Trunk parameters. Trunk parameters include trunk flexion angle, side angle and twist angle. The calculation of trunk flexion and side is shown in Figure 5a. The numbers represent the corresponding joints in Figure 2. Plane ***a*** represents the upper body, which is defined by the neck and two hips. Plane ***b*** is the lower body plane, which is defined by waist and two knees. α is the angle between plane ***a*** and plane ***b***. $\beta = 90 - \alpha$ is the trunk flexion angle. γ is the trunk lateral flexion angle, which is defined by the angle between line 13-3 and line 13-3', where 13-3 represents the current spine position and 13-3' represents the neutral spine position. If we denote the norm vector of plane a as the $\boldsymbol{n}$, vector 12-16 as $\boldsymbol{v}$, then vector 13-3' = $\boldsymbol{n} \times \boldsymbol{v}$. Trunk twist is defined as the angles between shoulders vector 5-8 and hip vector 12-16, as shown in Figure 5b.

Neck parameters. Neck parameters include neck flexion angle and neck side angle. The calculation is similar with trunk flexion angle and trunk side angle. The two planes are replaced with upper shoulder plane defined by point 1, 2, 4 and lower shoulder plane defined by point 2,4,11.

Upper arm parameters. Upper arm parameters include the flexion and abduction angles of both upper arms. Figure 6 illustrates the calculation process. Plane ***a*** is the frontal plane. Plane ***b*** is a sagittal plane, which is perpendicular to plane ***a***. Line ***l*** is the intersecting line of the plane ***a*** and plane ***b***. Line 8-9 represents the left upper arm. Line 8-9' is the projection on plane ***a***, and line 8-9'' is the projection on plane ***b***. Then the upper arm abduction is defined by angle β, i.e. the angle between line ***l*** and line 8-9'; the upper arm flexion is defined by angle α, i.e. the angle between line ***l*** and line 8-9''.

Lower arm parameters. Lower arm parameters include the flexion angles of both elbows. The calculation is similar with leg flexion angles.

Leg parameters. Leg parameters include the flexion angles of both knees and the balance of legs. As shown in Figure 6b, the flexion angles equal to the supplement of the angle between vector 12-14 and vector 14-15. The balance of legs is defined by the difference between two knee flexion angles.

REBA score

REBA provides ergonomic risk score rules based on the above parameters. Figure 4 illustrates the calculation process of the REBA score. Firstly, a joint-level score is given based on joint parameters. Take the trunk score for instance. A base trunk score is given based on trunk flexion angles; then the score will be added by 1 if trunk twist or side is found. Secondly, a body segment level score is calculated according to the joint-level score. REBA divides the human body into two parts. Part A includes trunk, neck, and legs, and part B includes upper arms and lower arms. The scores of part A and part B are separately regulated by table A and table B in REBA. Finally, the whole-body level score, which also provides the urgency of ergonomic improvements, is calculated according to the scores of part A and part B based on Table C in REBA. Table C in REBA is a score matrix, where the row represents score A and the column represents score B. Each element in Table C is the whole-body score according to score A and score B. In this research, the posture-based score was completed by JVEC, while the load- and work pattern-related score was completed manually. For a more detailed explanation, please refer to (Hignett and McAtamney, 2000).

Experiments, results, and analysis

Two experiments were conducted to validate the proposed methodology. The first one is a laboratory experiment aiming at testing the accuracy of the methodology. While the second one is an on-site experiment for the validation of the whole ergonomic assessment methodology.

Laboratory experiment for accuracy assessment

Experiment dataset

In the experiment, the subject enacted three construction activities including rebar, bricklaying, and plaster. The subject was wearing IMU sensors during the whole experiment process. The IMU sensor has an accuracy of 1° (3-Space™ Wireless 2.4GHz DSSS) (Yost Labs, 2017). To ensure the accuracy of IMU sensors, the researchers (1) calibrated the IMU sensors with static postures for 1 minute and rotated each IMU sensor for 30 times to eliminate the initial error; and (2) controlled the duration of each task and calibrated the IMU sensors with static postures before each task to mitigate the time-accumulated error. A common camera and a set of IMU sensors recorded the subject's motions simultaneously. The frequency of camera data is 25 fps (frames per second), and the frequency of IMU data is 30 fps. Figure 7 shows the representative frames of the experiment, Table 1 presents the raw data.

Camera data. The camera data was captured with the 3D-pose estimator. Figure 8 shows one of the captured frames. The data of each frame was represented by a 16×3 matrix composed of the 3D coordinates of 16 joints.

IMU data. The IMU data was captured with IMU sensors. These IMU sensors were tied to the key joints of the subject. The sensor data was recorded in BVH (Biovision Hierarchy) format, which contains the joints' three dimensions relative to original position and the offset of one child joint to its parent joint. The BVH structure is presented in Figure 9. The root joint is hip joint, which contains six parameters in raw data, including three-dimensional positions and three-dimensional rotations. In this research, we only focus on the posture of the subject, rather than the orientation and location, so all the six parameters of the hip joint were set as zero during the experiment. The arrows start at parent joints and end at child joints. The chest point gives an example of the data structure of all the joints except for the hip/root joint. Each joint contains the offsets along three dimensions relative to the original posture and three-dimensional rotation angles relative to the original posture at time t.

Data processing

The camera data and the IMU data are different in joint description, skeleton description, orientation, and frequency, thus the data must be preprocessed to eliminate the differences.

Unify skeleton description. For the camera data, a skeleton includes 16 joints, while an IMU skeleton includes only 13 joints. Consequently, when validating the camera data with the IMU data, only the positions of 13 joints, including the head and the 12 joints of four limbs, were adopted.

Unify joint description by translating IMU data from Euler angles to positions. It is a general practice to evaluate the accuracy of a pose estimation algorithms with the distance between predicted joint locations with the ground truth (Grinciunaite et al., 2016). In this study, camera data corresponded to the predicted joint locations and was represented by 3D joint positions. IMU data corresponded to the ground truth and was recorded by the rotation and offset of each joint. To make these two kinds of data comparable, the IMU data was transformed to joint positions based on the Denavit-Hartenberg matrix. The matrix separates a screw displacement into the product of a pure translation along a line and a pure rotation about the line (Legnani et al., 1996). In this study, the motion of each joint was seen as a screw displacement relevant to its parent joint, which was separated to pure translations and pure rotation along x, y and z-axis. Eq.4 illustrates the calculation process. Figure 10 provides the positive directions of the rotation systems and the rotation angles.

$$\boldsymbol{v_{joint}} = \boldsymbol{v_{root}} \cdot \boldsymbol{DH_{joint}} \qquad \text{Eq.}$$

$$DH_c = DH_P \cdot R \cdot T$$

$$R = R_z R_y R_x$$

$$= \begin{bmatrix} \cos(r_z) & -\sin(r_z) & 0 & 0 \\ \sin(r_z) & \cos(r_z) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos(r_y) & 0 & -\sin(r_y) & 0 \\ 0 & 1 & 0 & 0 \\ \sin(r_y) & 0 & \cos(r_y) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(r_x) & -\sin(r_x) \\ 0 & \sin(r_x) & \cos(r_x) \\ 0 & 0 & 0 \end{bmatrix}$$

$$T = \begin{bmatrix} 1 & 0 & 0 & \Delta x \\ 0 & 1 & 0 & \Delta y \\ 0 & 0 & 1 & \Delta z \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

where

$\boldsymbol{v_{root}}$ is the position of the root joint; $\mathbf{v_{joint}}$ is the position of the target joint;

$\boldsymbol{DH}$ means the Denavit-Hartenberg matrix from one joint to another; $\boldsymbol{DH_c}$ means the Denavit-Hartenberg matrix of a child joint; $\boldsymbol{DH_p}$ means the Denavit-Hartenberg matrix of a parent joint;

$\boldsymbol{R_x}, \boldsymbol{R_y}, \boldsymbol{R_z}$ are the rotation matrices around Zr, Yr, Xr axis; **R** is the rotation matrix; **T** is the transformation matrix;

$\mathrm{r_x}, r_y. r_z$ are the rotation angles around the Xr, Yr, Zr axis;

$\Delta x,\ \Delta y,\ \Delta z$ are the offsets of a child joint to its parent joint along X, Y, Z axis.

Unify joint description by translating camera data from positions to Euler angles. Since REBA score is based on joint angles, a comparison was also made according to the differences between joint angles generated from camera data and IMU data. Since the IMU data only contains 12 limb joints and 1 head joint, only joint angles on limbs could be compared, including elbow angles and knee angles.

Unify frequency. In the experiment, both the camera and IMU sensors could record the start time automatically. Then the time of each camera frame and IMU frame was calculated based on the start time, a number of frames and the data frequency (Eq.5). Then the IMU frame and camera frame with the same or nearest time were matched.

$$t_f = t_0 + f/v \qquad \text{Eq.5}$$

where

t_f is the time of the frame f. The format of t_f is hh: mm: ss.000;

t_0 is the start time. The format of t_0 is hh: mm: ss.000;

v is the data frequency. $v = 25$ fps for camera data; $v = 30$ fps for IMU data.

Unify orientation. As previously mentioned, the orientation of the IMU skeleton, i.e. the rotation of the hips/root joint, was set as zero. However, the camera skeleton's orientation varied over time. To make the orientation consistent, firstly, the origin of the camera data was set as the hip joint, which was defined by the middle point of the left and right hip. Then the two skeletons were aligned based on the hip vector and the left upper leg vector. The hip vector starts from left hip joint and middle hips joint; the left upper leg vector starts from the left hip joint and ends at left knee joint. Rodrigues matrix is applied to align the vectors. The Rodrigues matrix is usually used to calculate the DH matrix (Denavit-Hartenberg matrix) given the rotated vector and the original vector. For example, the hip vector in an IMU skeleton is denoted as $\mathbf{v_0}$, the hips vector in the corresponding camera skeleton is $\mathbf{v}$. Then the Rodrigues matrix is defined as:

$$\boldsymbol{v} = \boldsymbol{v_0} cos\theta + (\boldsymbol{k} \times \boldsymbol{v}) sin\theta + \boldsymbol{k}(\boldsymbol{k} \cdot \boldsymbol{v})(1 - cos\theta) \qquad \text{Eq.6}$$

where

θ is the rotation angle, $\theta = arccos(\frac{\boldsymbol{v} \cdot \boldsymbol{v_0}}{|\boldsymbol{v}||\boldsymbol{v_0}|})$;

$\boldsymbol{k}$ is the rotation axis, $\boldsymbol{k} = \frac{\boldsymbol{v} \cdot \boldsymbol{v_0}}{|\boldsymbol{v}||\boldsymbol{v_0}| sin\boldsymbol{\theta}}$

Other joints rotated with the same rotation angle θ and axis $\boldsymbol{k}$.

Results of accuracy assessment by comparing camera data with sensor data

Joint coordinates estimation accuracy. The estimated joint locations based on camera data were compared with the joint locations measured with IMU sensors. The comparison results are shown in Table 2 and Figure 11. The average error is 4.10 cm per joint. The nearer the joint is to the hips, the smaller the error is. The reason is that when transforming IMU data from angles to positions, the child joint position was calculated based on its parent joint's position, so the error was accumulated.

Joint angles estimation accuracy. The joint angles, including the elbow angles and knee angles on both sides, were calculated based on camera data and IMU data respectively. The joint angle estimation error was evaluated by the difference between camera joint angel results and IMU joint angle results. The distribution of the error is given in Figure 12. The mean of the error is -0.70°. The standard deviation is 8.21°.

Joint score estimation accuracy. The aim of this research is to provide ergonomic risk scores based on joint angels, so the REBA-based ergonomic risk score was also used as an indicator to provide accuracy. In this experiment, the ergonomic risk scores were calculated based on camera data and IMU data. The camera data-based score is compared with IMU data-based score to assess the accuracy of the ergonomic evaluation. As aforementioned, only 4 joint angles (the angles of both elbows and both knees) could be generated from IMU data, so only the elbow and knee score were compared. Figure 13 is the result of each frame and Figure 14 is the confusion matrix of the score result.

According to REBA, knee angles were divided into 3 categories (0-120°,120-150°,150-180°), which correspond to score 2, 1, 0 respectively. Elbow angels were divided into 2 categories (80-120°, and else), which correspond to score 2 and score 1. It could be observed from Figure 13 that most of the joint angles were classified into the right category, which means the method could provide the right ergonomic risk score. Figure 14 provided a more quantitative description about the accuracy. For the knee joint, the accuracy of score 0, 1 and 2 were 70%, 93%, and 85% respectively; for the elbow joint, the accuracy of score 1 and 2 were 96% and 75% respectively.

Site experiment

Experiment dataset

The site experiment was conducted to validate the feasibility of the JVEC methodology. Six trades of construction workers were involved, including bricklayer, concreter, pipe layer, bar fixer, scaffolder and formwork erector. The authors shot a ten-minute video for each worker and then applying JVEC to performer ergonomic assessments based on the videos. The video frequency is 25 fps.

Results of 3D pose estimator

The 3D pose estimator was applied to get the worker's motion data from the video clips. Figure 15 shows the results of 3D pose estimation. It could be observed that the 3D pose estimation could generate 3D skeleton based on video frames in a real construction site. Compared with depth camera-based methods, the 3D pose estimation worked well in both indoor (the 1st and 2nd rows in Figure 15) and outdoor environments (the 3rd-12th rows in Figure 15). In addition, the 3D pose estimation also worked well in a dark environment such as the bricklayer images (the 1st row). However, the joint position estimation results on construction sites were not so accurate as those in the laboratory. For example, the estimated right knee angle of frame 9 in bricklayer images (the 2nd row), the left elbow angle of frame 9 in scaffolder images (the 10th row), and the left knee in frame 9 in form worker images (the 12th row) are obviously different with the corresponding video frames. The reasons might be 1) the obstructions between the human body and the camera such as the worker's body segments in the bricklaying task (the 1st row), the tubes in the scaffolding task (the 5th row) and the forms in the form task (the 11th row); 2) the differences in the lengths of human body segments between the workers and the ones used to train the 3D pose estimator. The 3D pose estimator estimates the depth of joints based on body segment lengths. If the workers' body segment lengths are different with the body segment lengths in the training dataset, estimation error will occur.

Results of the ergonomic assessment

Figure 16 and Figure 17 represent the results of the ergonomic assessment. Figure 16 shows the whole body ergonomic risk score of each frame, which demonstrates that the proposed methodology could provide a quantitative ergonomic risk score for each frame of videos on real construction site. Most of the scores were between 10 and 13, which is consistent with the observation results of the REBA scores of construction workers (Kulkarni and Devalkar, 2018). It could be observed from Figure 16 that the pipe layer, bar fixer, form worker and bricklayer seemed to have higher ergonomic risk than concreter and scaffolder during the on-site experiment, which means, during this experiment, the former four trades of workers were faced with higher ergonomic risk than concreters and scaffolders.

Figure 17 shows more detailed results, which consist of 30 frequency histograms (5 body segment REBA score items × 6 construction trades). The pipe layer had the highest trunk and leg ergonomic risk score because the pipe layer was continuously squatting or bending during the experiment. The comparison of each column in Figure 17 suggested the ergonomic risk of each body segment. In the first and last column, the trunk score, upper arm score, and lower arm score tend to be higher than other scores, which suggested that the bricklayer and form erector should pay attention to their arms and trunks. Similarly, for the concreter and scaffolder, both neck and lower arms deserved more attention.

Discussion

Construction workers faced with high ergonomic risks, resulting in a negative influence on the workers' well-being and productivity. It is important, therefore, to assess the workers' ergonomic risk accurately. Observations, sensors, and depth cameras are three main posture data collecting methods but are faced with challenges of low accuracy, uncomfortableness, and unsuitability to outdoor environments. This research intends to solve the issues by blending a video-based 3D pose estimation algorithm with REBA ergonomic risk score.

Contributions of JVEC

Higher ergonomic estimation accuracy. According to the laboratory experiment, the joint position error was 4.10cm per joint; the mean of joint angle error was -0.70°; and the ergonomic risk score accuracy was 70-96%. The probability of misclassifying was 4-30%. Previous posture-based ergonomic assessment methods are mainly based on manual-observation to collect the joint angles data, of which the accuracy is about 54% (Lowe, 2004). Thus the accuracy has been increased a lot, benefiting efficient ergonomic assessment.

Vision-based ergonomic assessments. JVEC collects posture data from videos or images, which makes it more applicable for construction sites. The vision-based method can capture workers posture without any sensors. Besides, compared with the depth camera, such as Kinect, the method can work under direct sunlight, which makes it more suitable for construction sites.

Joint-level ergonomic assessments. Different from previous vision-based methods (Yan, Li et al.

2017), the method can provide joint-level ergonomic assessments, which makes the assessments results in more accurate and more benefit to specific ergonomic interventions. The reason lies in the new posture data collection method. This study uses the deep-learning-based computer vision algorithm to provide 3D joint location data with monocular RGB camera (such as cameras in smart phones). The benefits include 1) collecting posture data and assessing ergonomic risk in a non-intrusive way without inferring workers' normal construction tasks, and 2) providing 3D posture data, instead of 2D posture data, so that the joint angles can be measured more accurately.

Near real-time ergonomic assessments. In the experiments, JVEC performed 3D pose estimation, joint angle calculation and REBA scoring for each video frame. The whole JVEC process for one frame took about 0.2 s on one Titan X GPU with CUDA 8.0 and cudnn 6, which means that the processor frequency is 5 fps.

Limitations of JVEC

Higher 3D motion estimation accuracy. Figure 18 provides two failure cases of the on-site experiment. The reason lies in visual obstacles. In Figure 18a, the worker was squatting, and most of the body parts were invisible. In Figure 18c, the worker's right arm was blocked by the form, leading to the high error of the left arm. The 3D pose estimation could be improved by adding more pictures with obstructions to the training dataset so that the blocked body segment could be inferred. In addition, the current accuracy evaluation assumes that the IMU could provide the ground truth joint angles. However, the IMU sensors have an error of 1 degree (3-Space™ Wireless 2.4GHz DSSS) (Yost Labs, 2017). This may lead to a joint location error positively related to the segment length. The accuracy assessment could be more objective if there are more accurate 3D joint location methods suitable for outdoor environment.

Multi-worker motion estimation. The current version of JVEC can only be applied on frames containing only one worker. However, in most of the case, one supervision camera can record the activities of several workers. If JVEC can recognize all the workers within one frame, the efficiency could be increased a lot.

Automatic external load and repetitiveness identification. REBA provides ergonomic risk score based on posture, external load, and repetitiveness. However, JVEC could only generate posture information automatically. In this research, the external load and repetitiveness score were still given by manual observation.

More on-site data. Though the experiment demonstrates the feasibility of the methodology, the on-site experiment only records a ten-minute video for each worker. For data-based ergonomic improvement suggestions, it is necessary to take longer-period video records for more construction workers.

Suggestions for future research

Based on the above limitations of JVEC, extensive research should be conducted in the future to make the method more applicable for ergonomic assessments on construction sites.

1) Ergonomic assessments considering external loads

As one of the ergonomic risks, external loads should be considered in the future version of JVEC.

Two methods could be used to collect the external load data:

Inferring external load data from videos. This method will first identify the objects carried by the worker, then infer the external load based on the object category. Two assumptions are involved here: 1) Most of the construction workloads are located at hands, and 2) the objects of the same category are of the same weight.

Inferring external load data with pressure sensors. This method will involve insole-shaped pressure sensors, which can measure workers' ground reaction forces during work. The difference between the total ground reaction force and the worker's self-weight is the external load. The assumption is that the construction workloads are located at hands.

2) Identifying motion repetitiveness with activity recognition algorithms

The time sequence of each joint angle could be generated based on the joint positions of each frame. The time sequence data contains information such as motion repetitiveness and duration, which could contribute to a more automatic and accurate ergonomic risk assessment method. To reach the goal, machine learning and deep learning algorithms for sequence data could be applied, such as Long Short-Term Memory (LSTM) (Yoo, 2017).

3) Inferring multiple construction workers' 3D postures and those behind visual obstacles

The 3D pose estimator can be improved by integrating with two advanced deep learning algorithms for multiple worker recognition and vision-obstacle remove. The algorithm in (Chu et al., 2017) can remove obstacles, and the open pose algorithm can identify 2D skeletons from images (Cao et al., 2017).

4) Collecting more site data for data-based ergonomic improvement suggestions

The authors will collect more videos from construction sites for comprehensive ergonomic assessment results, which may support the following suggestions including the most high-risk construction trades, the heavy-load joints of each trade, and the work-rest schedule for each trade and even each worker.

Conclusion

This research presents an automatic ergonomic assessment tool for construction workers based on on-site data to prevent ergonomic risks. It can provide accurate and timely ergonomic assessment based on 2D videos by using a state-of-art 3D posture estimation algorithm to capture 3D joint positions from 2D images as well as adopting the REBA rule to get multi-level ergonomic risk scores. The laboratory experiments show that the proposed method can successfully and efficiently get workers' 3D joint positions and practice accurate enough for the sequent REBA-based ergonomic assessment. The site experiment demonstrates that JVEC is workable on construction sites as well. The proposed methodology contributes to both construction workers behavior data collection and ergonomic assessment. Compared with previous behavior data collection methods (manual observation, 3D pose estimation sensors, and depth cameras), this methodology could provide accurate posture data with a less intrusive way in both indoor and outdoor environments, which makes it suitable for construction sites. By blending the data collection methods and REBA score, JVEC has a potential to provide comprehensive ergonomic assessment results and suggestions such as the most high-risk construction trades, the heavy-load joints of each trade, and the work-rest schedule for each trade or worker, for decision making on ergonomic risk management in the future.

Data Availability Statement

Data generated or analyzed during the study are available from the corresponding author by request. Information about the *Journal*'s data-sharing policy can be found here: http://ascelibrary.org/doi/10.1061/(ASCE)CO.1943-7862.0001263.

References

Åhsberg, E., & Gamberale, F. (1998). Perceived fatigue during physical work: an experimental evaluation of a fatigue inventory. *International Journal of Industrial Ergonomics*, *21*(2), 117–131. https://doi.org/https://doi.org/10.1016/S0169-8141(96)00071-6

Alavinia, S. M., van Duivenbooden, C., & Burdorf, A. (2007). Influence of work-related factors and individual characteristics on work ability among Dutch construction workers. *Scandinavian Journal of Work, Environment & Health*, *33*(5), 351–357. Retrieved from http://www.jstor.org/stable/40967665

Alwasel, A., Abdel-Rahman, E. M., Haas, C. T., & Lee, S. (2017). Experience, Productivity, and Musculoskeletal Injury among Masonry Workers. *Journal of Construction Engineering and Management*, *143*(6), 5017003. https://doi.org/doi:10.1061/(ASCE)CO.1943-7862.0001308

Antwi-Afari, M. F., Li, H., Edwards, D. J., Pärn, E. A., Seo, J., & Wong, A. Y. L. (2017). Biomechanical analysis of risk factors for work-related musculoskeletal disorders during repetitive lifting task in construction workers. *Automation in Construction*, *83*, 41–47. https://doi.org/10.1016/j.autcon.2017.07.007

Borg, G. (1998). *Borg's perceived exertion and pain scales*. Champaign, IL, US: Human Kinetics.

Cao, Z., Simon, T., Wei, S.-E., & Sheikh, Y. (2017). Real-time multi-person 2d pose estimation using part affinity fields. In *CVPR* (Vol. 1, p. 7).

Cheng, T., Teizer, J., Migliaccio, G. C., & Gatti, U. C. (2013). Automated task-level activity analysis through the fusion of real-time location sensors and worker's thoracic posture data. *Automation in Construction*, *29*, 24–39. https://doi.org/http://dx.doi.org/10.1016/j.autcon.2012.08.003

Chu, X., Yang, W., Ouyang, W., Ma, C., Yuille, A. L., & Wang, X. (2017). Multi-Context Attention for Human Pose Estimation. Retrieved from http://arxiv.org/abs/1702.07432

Corlett, E. N., & Bishop, R. P. (1976). A technique for assessing postural discomfort. *Ergonomics*, *19*(2), 175–182. https://doi.org/10.1080/00140137608931530

Dutta, T. (2012). Evaluation of the Kinect™ sensor for 3-D kinematic measurement in the workplace. *Applied Ergonomics*, *43*(4), 645–649. https://doi.org/10.1016/j.apergo.2011.09.011

Entzel, P., Albers, J., & Welch, L. (2007). Best practices for preventing musculoskeletal disorders in masonry: Stakeholder Perspectives. *Applied Ergonomics*, *38*(5), 557–566. https://doi.org/10.1016/j.apergo.2006.08.004

Golabchi, A., Han, S., & Fayek, A. R. (2016). A fuzzy logic approach to posture-based ergonomic analysis for field observation and assessment of construction manual operations. *Canadian Journal of Civil Engineering*, *43*(4), 294–303. https://doi.org/10.1139/cjce-2015-0143

Grinciunaite, A., Gudi, A., Tasli, E., & Uyl, M. den. (2016). Human Pose Estimation in Space and Time using 3D CNN. https://doi.org/10.1007/978-3-319-49409-8_5

Han, S., Lee, S., & Pena-Mora, F. (2013). Vision-based detection of unsafe actions of a construction worker: a case study of ladder climbing. *Journal of Computing in Civil Engineering*, *27*(6), 635–644. https://doi.org/10.1061/(ASCE)cp.1943-5487.0000279

Han, S. U., Lee, S. H., & Pena-Mora, F. (2012). *A Machine-learning Classification Approach to Automatic Detection of Workers' Actions for Behavior-based Safety Analysis*. (R. R. Issa & I. Flood, Eds.), *Proceedings of the 2012 ASCE International Conference on Computing in Civil Engineering*. https://doi.org/10.1061/9780784412343.0009

Hignett, S., & McAtamney, L. (2000). Rapid Entire Body Assessment (REBA). *Applied Ergonomics*, *31*(2), 201–205. https://doi.org/10.1016/S0003-6870(99)00039-3

Hwang, S., Seo, J., Jebelli, H., & Lee, S. (2016). Feasibility analysis of heart rate monitoring of construction workers using a photoplethysmography (PPG) sensor embedded in a wristband-type activity tracker. *Automation in Construction*, *71*(Part 2), 372–381. https://doi.org/10.1016/j.autcon.2016.08.029

Karhu, O., Härkönen, R., Sorvali, P., & Vepsäläinen, P. (1981). Observing working postures in industry: Examples of OWAS application. *Applied Ergonomics*, *12*(1), 13–17. https://doi.org/https://doi.org/10.1016/0003-6870(81)90088-0

Kim, I.-J. (2017). The Role of Ergonomics for Construction Industry Safety and Health Improvements. *Journal of Ergonomics*, *7*(2), 1000e166. https://doi.org/10.4172/2165-7556.1000e166

Kulkarni, V. S., & Devalkar, R. V. (2018). Postural analysis of building construction workers using ergonomics. *International Journal of Construction Management*, 1–8. https://doi.org/10.1080/15623599.2018.1452096

Legnani, G., Casolo, F., Righettini, P., & Zappa, B. (1996). A homogeneous matrix approach to 3D kinematics and dynamics — I. Theory. *Mechanism and Machine Theory*, *31*(5), 573–587. https://doi.org/10.1016/0094-114X(95)00100-D

Li, X., Han, S., Gül, M., Al-Hussein, M., & El-Rich, M. (2018). 3D Visualization-Based Ergonomic Risk Assessment and Work Modification Framework and Its Validation for a Lifting Task. *Journal of Construction Engineering and Management*, *144*(1), 4017093. https://doi.org/doi:10.1061/(ASCE)CO.1943-7862.0001412

Lowe, B. D. (2004). Accuracy and validity of observational estimates of wrist and forearm posture. *Ergonomics*, *47*(5), 527–554. https://doi.org/10.1080/00140130310001653057

McAtamney, L., & Nigel Corlett, E. (1993). RULA: a survey method for the investigation of work-related upper limb disorders. *Applied Ergonomics*, *24*(2), 91–99. https://doi.org/10.1016/0003-6870(93)90080-S

Nath, N. D., Akhavian, R., & Behzadan, A. H. (2017). Ergonomic analysis of construction worker's body postures using wearable mobile sensors. *Applied Ergonomics*, *62*, 107–117. https://doi.org/10.1016/j.apergo.2017.02.007

Newell, A., Yang, K., & Deng, J. (2016). Stacked Hourglass Networks for Human Pose Estimation. *European Conference on Computer Vision*, 483–499. Retrieved from arxiv.org/abs/1603.06937

Ning, X., & Guo, G. (2013). Assessing spinal loading using the Kinect depth sensor: A feasibility study. *IEEE Sensors Journal*, *13*(4), 1139–1140. https://doi.org/10.1109/JSEN.2012.2230252

Nunes, J. F., Moreira, P. M., & Tavares, J. M. R. S. (2011). Human Motion Analysis and Simulation Tools. In O. T. Institute (Ed.), *Handbook of Research on Computational Simulation and Modeling in Engineering* (pp. 359–388). 200 Constitution Ave., NW, Washington, DC 20210. https://doi.org/10.4018/978-1-4666-8823-0.ch012

Nunes, J. F., Moreira, P. M., & Tavares, J. M. R. S. (2015). Human Motion Analysis and Simulation Tools. In *Handbook of Research on Computational Simulation and Modeling in Engineering* (pp. 359–388). https://doi.org/10.4018/978-1-4666-8823-0.ch012

Punnett, L., Wegman, D. H., Ã, Laura, P., & Wegman, David, H. (2004). Work-related musculoskeletal disorders: The epidemiologic evidence and the debate. *Journal of Electromyography and Kinesiology*, *14*(1), 13–23. https://doi.org/10.1016/j.jelekin.2003.09.015

Ray, S. J., & Teizer, J. (2012). Real-time construction worker posture analysis for ergonomics training. *Advanced Engineering Informatics*, *26*(2), 439–455. https://doi.org/10.1016/j.aei.2012.02.011

Sedighi Maman, Z., Alamdar Yazdi, M. A., Cavuoto, L. A., & Megahed, F. M. (2017). A data-driven approach to modeling physical fatigue in the workplace using wearable sensors.

Applied Ergonomics, *65*, 515–529. https://doi.org/10.1016/j.apergo.2017.02.001

Seo, J., Han, S., Lee, S., & Kim, H. (2015). Computer vision techniques for construction safety and health monitoring. *Advanced Engineering Informatics*, *29*(2), 239–251. https://doi.org/http://dx.doi.org/10.1016/j.aei.2015.02.001

Seo, J., Lee, S., & Seo, J. (2016). Simulation-Based Assessment of Workers' Muscle Fatigue and Its Impact on Construction Operations. *Journal of Construction Engineering and Management*, *142*(11), 04016063. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001182

Seo, J., Yin, K., & Lee, S. (2016). Automated Postural Ergonomic Assessment Using a Computer Vision-Based Posture Classification. In *Construction Research Congress 2016* (pp. 809–818). Reston, VA: American Society of Civil Engineers. https://doi.org/10.1061/9780784479827.082

U.S. Bureau of Labor Statistics. (2016). Nonfatal occupational injuries and illnesses resulting in days away from work in 2016 : The Economics Daily: U.S. Bureau of Labor Statistics. Retrieved April 27, 2018, from https://www.bls.gov/opub/ted/2017/nonfatal-occupational-injuries-and-illnesses-resulting-in-days-away-from-work-in-2016.htm?view_full

Umer, W., Li, H., Szeto, G. P. Y., & Wong, A. Y. L. (2017). Identification of Biomechanical Risk Factors for the Development of Lower-Back Disorders during Manual Rebar Tying. *Journal of Construction Engineering and Management*, *143*(1), 04016080. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001208

Valero, E., Sivanathan, A., Bosché, F., & Abdel-Wahab, M. (2016). Musculoskeletal disorders in construction: A review and a novel system for activity tracking with body area network.

Applied Ergonomics, *54*, 120–130. https://doi.org/https://doi.org/10.1016/j.apergo.2015.11.020

Wang, D., Dai, F., & Ning, X. (2015). Risk Assessment of Work-Related Musculoskeletal Disorders in Construction: State-of-the-Art Review. *Journal of Construction Engineering and Management*, *141*(6), 04015008. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000979

Wiktorin, C., Karlqvist, L., & Winkel, J. (1993). Validity of self-reported exposures to work postures and manual materials handling. Stockholm MUSIC I Study Group. *Scandinavian Journal of Work, Environment & Health*, *19*(3), 208–214. https://doi.org/10.5271/sjweh.1481

Yan, X., Li, H., Wang, C., Seo, J., Zhang, H., & Wang, H. (2017). Development of ergonomic posture recognition technique based on 2D ordinary camera for construction hazard prevention through view-invariant features in 2D skeleton motion. *Advanced Engineering Informatics*, *34*, 152–163. https://doi.org/10.1016/j.aei.2017.11.001

Yoo, J. H. (2017). Large-scale Video Classification guided by Batch Normalized LSTM Translator. Retrieved from http://arxiv.org/abs/1707.04045

Yost Labs. (2017). 3-Space^{TM} Sensors. Retrieved April 19, 2018, from https://yostlabs.com/3-space-sensors/

Yu, Y., Guo, H., Ding, Q., Li, H., & Skitmore, M. (2017). An experimental study of real-time identification of construction workers' unsafe behaviors. *Automation in Construction*, *82*, 193–206. https://doi.org/10.1016/j.autcon.2017.05.002

Zhou, X., Huang, Q., Sun, X., Xue, X., & Wei, Y. (2017). Towards 3D Human Pose Estimation in the Wild: a Weakly-supervised Approach. *IEEE International Conference on Computer*

Vision. Retrieved from http://arxiv.org/abs/1704.02447

Zhu, G., Cao, L. (2014). Human Motion Recognition Based on Skeletal Information of Kinect Sensor. *Computer Simulation*, *31*(12), 329–333,345.

Table 1 The raw data of the indoor experiment

Activity	Duration (second)	Camera data		IMU data	
		Number of frames	Number of joints	Number of frames	Number of joints
Rebar	114	2853	16	3424	13
Bricklaying	22	550	16	660	13
Plaster	19	475	16	570	13
Total	155	3878	-	4654	-

Table 2 The errors of the 3D **joint** position capture from 3D video frames (Unit: cm)

Task	Statisitc	Head	LS	RS	LE	RE	LW	RW	LH	RH	LK	RK	LA	RA	Mean
Rebar	Mean	6.12	4.17	3.11	4.76	4.38	5.40	5.79	0.88	0.88	0.88	1.86	3.62	4.10	3.53
	Std.	1.99	1.22	1.19	1.15	1.11	1.24	1.10	0.32	0.32	0.32	0.89	2.21	2.45	1.40
Bricklayer	Mean	7.94	5.69	5.58	4.23	3.71	6.38	5.57	1.13	1.13	1.13	3.02	9.10	9.47	4.93
	Std.	1.94	1.36	1.25	0.88	0.86	1.02	1.27	0.35	0.35	0.35	1.20	1.62	1.66	0.93
Plaster	Mean	10.38	8.85	8.77	7.58	8.98	10.31	12.97	0.58	0.58	0.58	2.11	2.74	3.14	3.24
	Std.	1.45	1.01	1.31	0.93	1.95	1.07	2.00	0.17	0.17	0.17	1.00	1.22	1.24	1.32
Mean	Mean	8.14	6.24	5.82	5.52	5.69	7.36	8.11	0.87	0.87	0.87	2.33	5.16	5.57	3.90
	Std.	2.26	1.70	1.93	1.30	1.70	1.71	2.27	0.34	0.34	0.34	1.03	2.84	2.96	1.59

Note: Std. represents standard deviation.

Figure Captions

Figure 1 The framework of the ergonomic assessment methodology

Figure 2 The framework of the 3D pose estimator

Figure 3 The process of compression and decompression: (a) Convolution and max pooling in compression period; (b) Up-sampling in decompression period

Figure 4 The structure of REBA

Figure 5 The calculation of trunk flexion, trunk and twist side angle: (a) Trunk flexion and side; (b) Trunk twist

Figure 6 The calculation of upper arm flexion, abduction and knee flexion: (a) Upper arm flexion and abduction; (b) Knee flexion

Figure 7 Representative frames of the experiment

Figure 8 The captured 3D joint data from camera frame

Figure 9 The BVH data structure

Figure 10 The calculation of joint positions

Figure 11 The distribution of joint position estimation error

Figure 12 The distribution of joint angle estimation error

Figure 13 The distribution of joint score based on camera and IMU data

Figure 14 The confusion matrix to assess the ergonomic assessment accuracy

Figure 15 Part results of the 3D motion capture

Figure 16 The whole body ergonomic risk scores

Figure 17 Eight score items of six construction trades

Figure 18 Two failure cases of motion capture: (a) The video frame of first failure case; (b) The 3D pose estimation result of first failure case; (c) The video frame of second failure case; (d) The 3D pose estimation result of second failure case