Advanced Engineering Informatics 34 (2017) 152–163

Contents lists available at ScienceDirect

Advanced Engineering Informatics

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

Full length article

Development of ergonomic posture recognition technique based on 2D ordinary camera for construction hazard prevention through view-invariant features in 2D skeleton motion

Xuzhong Yan[a,c], Heng Li[a], Chen Wang[b,*], JoonOh Seo[a], Hong Zhang[c], Hongwei Wang[d]

[a] Department of Building and Real Estate, Faculty of Construction and Environment, The Hong Kong Polytechnic University, Hong Kong, China
[b] Centre for Construction Automation & Project Management, College of Civil Engineering, Huaqiao University, 361021 Xiamen, China
[c] Institute of Construction Management, College of Civil Engineering and Architecture, Zhejiang University, Hangzhou 310058, China
[d] School of Management, Huazhong University of Science and Technology, Wuhan 430074, China

ARTICLE INFO

Keywords:
Ergonomics
Person posture recognition
RGB camera
2D skeleton
View-invariant
Construction worker

ABSTRACT

Outdoor tasks operated by construction workers are physically demanding, requiring awkward postures leading to pain, injury, accident, or permanent disability. Ergonomic posture recognition (EPR) technique could be a novel solution for ergonomic hazard monitoring and assessment, yet non-intrusiveness and applicability in complex outdoor environment are always critical considerations for device selection in construction site. Thus, we choose RGB camera to capture skeleton motions, which is non-intrusive for workers compared with wearable sensors. It is also stable and widely used in an outdoor construction site considering various light conditions and complex working areas. This study aims to develop an ergonomic posture recognition technique based on 2D ordinary camera for construction hazard prevention through view-invariant features in 2D skeleton motion. Based on captured 2D skeleton motion samples in the test-run, view-invariant features as classifier inputs were extracted to ensure the learned classifier not sensitive to various camera viewpoints and distances to a worker. Three posture classifiers regarding human back, arms, and legs were employed to ensure three postures to be recognized simultaneously in one video frame. The average accuracies of three classifiers in 5-fold cross validation were as high as 95.0%, 96.5%, and 97.6%, respectively, and the overall accuracies tested by three new activities regarding ergonomic assessment scores captured from different camera heights and viewpoints were 89.2%, 88.3%, and 87.6%, respectively. The developed EPR-aided construction accident auto-prevention technique demonstrated robust accuracy to support on-site postural ergonomic assessment for construction workers' safety and health assurance.

1. Introduction

Work-related musculoskeletal disorders (WMSDs) are common occupational hazards, especially in the labor-intensive and manually demanding construction industry [10,22]. In the US construction industry, it is reported that the median days away from work due to WMSDs increased from 8 days in 1992 to 13 days in 2014; the proportion of WMSDs among workers aged 55–64 years doubled [55]. They also reported that more than 40% of WMSDs among workers are back injuries that are mainly caused by overexertion. In Hong Kong, the Pilot Medical Examination Scheme (PMES) for construction workers revealed that 41% registered workers reported musculoskeletal pain [56]. As the aging of construction workers becomes a critical problem in Hong Kong [33], the prevalence of WMSDs in construction would be even higher due to decreased functional capacity of aged workers [34]. As a group of painful disorders of muscles, tendons, nerves, cartilage, and joints, WMSDs among construction workers are mainly caused by compression, shear, tensile stress, and muscle force repeatedly acting on load bearing tissues [8]. The majority of these loadings in the construction industry can be attributed to repetitive works in prolonged awkward working postures, including overhead reaching, neck and back bending, and squatting, for example stooping posture in rebar tying work can significantly reduce the electromyographic activity of lumbar muscles a reduction of 60–80%) compared with one-legged kneeling and squatting posture [49]. The reduced electromyographic activity of muscles may shift the loading to ligaments and joint capsules of load bearing structures, which is known as a risk factor for WMSDs. In ergonomic perspective, it is suggested the frequency and duration of

* Corresponding author.
E-mail address: wch@hqu.edu.cn (C. Wang).

https://doi.org/10.1016/j.aei.2017.11.001
Received 19 June 2017; Received in revised form 2 October 2017; Accepted 8 November 2017
Available online 13 November 2017
1474-0346/

awkward postures of trunk, upper and lower limbs be controlled within acceptable ranges (ISO, 2003) [46,27]. For example, the Ovako Working Posture Analysis System (OWAS) classified the posture combinations of back, arms, and legs, and relative proportions of certain postures during work time into four action categories based on the risk assessment of musculoskeletal disorders [25,31]. In the OWAS system, the action categories range from "1, no actions required" to "4, corrective measures needed immediately". When the proportion of a certain posture during the observation period is larger than the fixed frequency limits defined by OWAS, the action category changes from lower to higher, which indicates the urgency of corrective actions is increasing. Adequate ergonomic training is also helpful to make the construction workers aware of ergonomically hazardous postures on the job site [41]. Therefore, it is essential to identify, to define, and to classify working postures without interfering construction workers so that the duration and frequency of awkward working postures can be monitored and controlled based on ergonomic rules. The information could also enable project safety inspectors to reduce or to prevent WMSDs hazards by protective actions such as site layout rearrangement [14].

This study aims to develop an awkward posture recognition technique using view-invariant features from 2D skeleton motion data captured by a single ordinary RGB camera. Compared with wearable sensor, RGB camera is less intrusive for construction workers. Compared with stereo camera or RGB-D camera, ordinary RGB camera has been widely used in the construction industry so that this technique is not device-intensive for most construction sites. The human body captured was divided into 13 2D skeletons with 14 joints using an architecture of the two-branch multi-stage convolutional neural network (CNN). Since there are huge variances in the projection of intra-class postures and similar projection of inter-class postures while being viewed from different camera viewpoints, a view-invariant representation of postures using probability density as discriminative features was developed in this study. Using this technique, the descriptors of the same type of postures were similar despite different viewpoints captured by a monocular RGB camera. Using the view-invariant features extracted from video frame samples, different machine learning methods on classification performance were tested. According to the average classification accuracy in 5-fold cross-validations, three optimal classifiers from training datasets for postures classification regarding arms, trunk, and legs were validated with good classification performance. The contribution comes twofold. First, a technique for ergonomic posture classification and recognition from 2D motion data of body skeleton and joint was developed, where three ergonomic postures regarding arms, trunks, and legs could be recognized simultaneously in one video frame. Second, view-invariant features were used as the representation of ergonomic postures and proven effective at posture classification in 2D imagery, so that the classifiers were effective in various camera viewpoints and distances (within its filming range) relative to an object. The outcomes of this study could lead to more realistic postural ergonomic assessment in an outdoor construction environment where ordinary RGB cameras could be easily installed.

2. Ergonomic posture recognition systems and ergonomic posture assessment

2.1. Ergonomic posture assessment in construction

The construction sector is labor-intensive and physically demanding and it has always been prone to be afflicted by WMSDs. According to previous statistics found in Wang et al. [51], seven tasks, namely concrete pouring, drywall installing, flooring, framing, masonry, plumbing, and roofing, hold a WMSDs incidence rate of more than 50%. WMSDs among construction workers are mainly caused by compression, shear, tensile stress, and muscle force repeatedly acting on load bearing tissues [8]. Most of these loadings exposed by construction workers can be attributed to repetitive works in prolonged awkward working postures [2]. For example, prolonged, repetitive awkward postures like stooping, squatting and kneeling, frequently confronted by construction workers, can cause overexertion in the spine and muscle of lower back and neck, which are common causes of WMSDs around lower back and neck.

To identify the ergonomically hazardous working pattern and job site layout, frequency and duration of ergonomically hazardous postures of the neck, shoulder, trunk, upper and lower limbs should be monitored and controlled [46,27], and it is essential to recognize ergonomic and non-ergonomic working postures as one of the foundations of ergonomic analysis in construction. Early studies regarding awkward posture assessment methods include the "Rapid Entire Body Analysis" REBA) for the entire body posture analysis [21]; "Posture, Activity, Tools and Handling", an ergonomic analysis of non-repetitive work [5]; "Rapid Upper Limb Assessment" RULA, an ergonomic assessment tool focusing on upper limbs [32]; and "Ovako Working Posture Analyzing System" OWAS for identifying and evaluating working postures [24]. The above-mentioned ergonomic rules were to reduce worker physical exposure to specific workplace hazards that could cause or aggravate WMSDs. They provide a quantitative and systematic assessment of ergonomic postural hazards in major body parts. However, these methods require site observations and questionnaires by safety personnel to collect data, which are inaccurate and inefficient due to subjective bias [47]. Such limitations are significant on construction sites due to the continuously changing environments and insufficient capable workforce [28]. Although limited by the data capture technology available at the time when these ergonomic evaluation methods were developed, the quantitative and systematic ergonomic rules provided valuable insights for awkward posture classification.

2.2. Person posture recognition systems

Person posture recognition (PPR) is normally based on wearable motion sensors or external sensors. The most popular wearable sensor-based system is inertial measurement units IMUs). Each IMU consists of a triaxial accelerometer, a triaxial gyro, and a triaxial compass, by which real-time positions of body parts can be translated from the angular data, the acceleration due to gravity, and magnetic direction. Many researchers used IMUs as motion capture method to identify and track human motion for ergonomic assessment. For example, Chen et al. [7] employed an IMU-based system to recognize awkward postures from sequencing actions for ergonomic interventions in construction. Based on the capture of motion data from wearable IMUs integrated into a body area network, Valero et al. [50] characterized unsafe postures regarding WMSDs of construction workers. Although wearable IMUs have satisfactory performance in accuracy and repeatability in the field experiment [42,55], there are some problems of using them in motion capture in the context of construction work. First, errors in the output of IMUs accumulate over time to the point that the output becomes invalid [16,39]. It may take longer for error to accumulate in units that contain error filters, but the error still builds up. Thus, the reliability of the IMU-based system is vulnerable to long working time and the inefficiency of error filters. Second, IMUs must be attached tightly to the human body to prevent rotational noise caused by unstable adherence. However, it was revealed by the front-line safety inspectors and construction workers that wearing tight sensor-based equipment is a physical burden for workers who should conduct manual work for a prolonged period, especially in hot and humid outdoor environment [16,39]. This burden may interfere with construction activity and reduce productivity [17]. Due to these problems of intrusive IMU-based motion capture system in construction, we focus on the development of computer vision based external sensing system for motion capture and assessment.

The performance of person posture recognition in computer vision is

based on the reliability of person detection and tracking [43], which is subject to the performance of both data processing methods and industrial computer vision hardware. We reviewed the performance of three widely used commercial cameras, namely stereo camera (e.g. Bumblebee XB3™), RGB-D camera (e.g. Microsoft Kinect™), and 2D RGB camera in PPR.

By using stereo vision based 3D human body model, Pellegrini and Iocchi [36] tracked posture model points with temporal filters and recognized principal postures based on hidden Markov models. Although the accuracy of a stereo camera can be gained using the correct base length given a focal length of the lens, it suffers from a pixel dropout problem. Furthermore, due to the sensitivity to light changes and complexities of stereo geometry calculation, the reconstruction of a depth map from stereo images is not practical enough for real-world applications in the construction industry. An RGB-D camera uses an infrared ray (IR) structured light projector along with a depth sensor in addition to an ordinary RGB camera. As a commercial RGB-D camera, the development of the Kinect application using an OpenNI software development kit (SDK) also allows for the extraction of human body skeleton models from depth video frames. The Kinect can detect and track human body skeletons by a deep randomized decision forest classifier [45], based on which PPR can be achieved using a different selection of posture features [19,20]. For instance, Escorcia et al. [12] used pose-codebook based on body joint angles to generate a bag of poses histogram as posture features, based on which a multiclass support vector machine SVM) action classifier was learned. Ray and Teizer [41] chose depth values of a person in a rescaled grayscale image as posture features. In their research, linear discriminant analysis LDA) was applied to classify ergonomic postures in four categories, sitting, stooping, standing, and crawling. Though accurate enough in person posture recognition problems [18], there is a severe problem in Kinect that it is vulnerable to outdoor conditions where a certain amount of solar IR and ferromagnetic radiation can introduce significant noises [52], even wash out the scene generated by the Kinect. Compared with RGB-D camera, ordinary RGB camera has been widely used in the construction industry so that this technique is not device-intensive for most construction sites. To this end, we revisited different representations of person posture for PPR using an RGB camera.

Approaches for person posture analysis in 2D images captured by the ordinary RGB camera can be categorized by different representations [1]. Typical representations include appearance-based [23], shape-based [3], volume-based [30], interest-point-based [29], optical flow-based [11], and kinematic feature-based [1]. Since body joint angles or locations are *rich representations* for PPR in computer vision [38], they are selected as basic features of human posture in 2D images for the purpose of classification. We localize anatomical thirteen 2D skeletons with 14 joints by using the architecture of the two-branch multi-stage convolutional neural network (CNN). The first branch of this architecture predicts confidence maps for body part detection; the second predicts Part Affinity Fields (PAFs) for body part association. The confidence map is a 2D representation of the belief that a specific body segment occurs in the image. The PAFs is a 2D vector field that points from one part of a limb to the other so that two key points of the limb are associated. Finally, the affinity fields and confidence maps are parsed by greedy inference to output the 2D anatomical points for a person in the image.

2.3. View-invariant feature for classification in 2D image

As major challenges in computer vision-based PPR given the location of body skeletons and joints, inter-class similarity and intra-class variability of postures due to view variance require view-invariant posture features [54]. The reason is that a subject can perform a posture in various directions with distinctive characteristics of body skeletons and joints movements, and there may be huge differences in the details of the same posture captured from different viewpoints, while subtle distinguishable 2D details in two different posture categories. For example, if the subject holds a horizontal straight arm posture, the camera will capture a point in front of the subject, while capturing a line on the side of the subject. Geometrically, posture features is a projected set of points and line segments. The utility of a feature of the projection is a function of how it varies with different viewpoints. Though it was proved by Burns et al. [6] that general-case view-invariants do not exist for a given number of points given true perspective, weak perspective or orthographic projection models, there are approximate view-invariant representations of postures, based on which 2D posture classification can be conducted in a principled manner. For example, to get a view-invariant representation of action, Kong et al. [26] proposed a sample-affinity matrix approach to measuring the similarities between different video samples in various views, which allows them to learn distinguishable features robust to view variations. Another work in action recognition [44] viewed an action as a set of pose transitions defined by triplets of points. They decomposed posture into a set of point-triplets and defined invariant representations of rigid motion by triplets across two frames. In this way, they can measure the similarity of pose transitions instead of matching score based on individual postures. Different from view-invariant action recognition that focuses on temporal and sequential features among video frames of an action; view-invariant posture recognition requires static features in each video frame of a posture. For example, Wu and Huang [53] used both physical and mathematical features in view-independent posture recognition. The physical features include texture, shape, and contour length, density, and edge length and 2-order moments of edge distribution. As for the mathematical features, principal component analysis (PCA) is used to capture a lower-dimensional feature space. In our case, based on the pixel locations of skeletons and joints, we propose a view-invariant representation of ergonomic postures using probability density of joint angle and length ratio of two skeletons.

3. Posture definition, data sampling and procedures

3.1. Quantitative ergonomic posture definition

We defined typical awkward postures based on the OWAS criteria [25,31] that have been validated in many jobs in different industrial contexts. Based on the risk assessment of musculoskeletal disorders, the OWAS define the action categories range from 1, no actions required, to 4, corrective measures needed immediately. When the proportion of a certain posture during the observation period is larger than the fixed frequency limits defined by OWAS, the action category changes from lower to higher, which indicates the urgency of corrective actions is increasing. Furthermore, based on the definition of flexion-extension angle (F), lateral bending angle (L), and rotation angle (R) of body parts by Goodvin et al. [15] and Yan et al. [55], nine postures that were frequently encountered by construction workers were quantitatively defined in Table 1.

We used three classifiers to recognize back, arms, and legs postures respectively. All the three posture classifiers were applied in each video frame captured by an ordinary RGB camera so that different combinations of back, arms, and legs postures could be recognized simultaneously. For instance, if a subject performs a squatting posture with his back bent over 20°, both knee bent and back bent posture should be captured and detected based on the classifier of back and leg postures, respectively.

We focused on three body parts: back, arms, and legs. For back, we focused on: **A.** straight posture, **B.** back bent slightly, and **C.** back bent heavily. For arms, we focused on **A.** both arms below shoulder level, **B.** one arm at or above shoulder level, and **C.** both arms at or above shoulder level. For legs, we focused on **A.** standing, **B.** knees bent, and **C.** squatting. In our approach, postures of all the three body parts could

Table 1
Quantitative definition of ergonomic postures.

Body part	Posture description	Quantitative definition
Back	**A.** Straight Back	$\boldsymbol{F}_{\mathbf{back}} \in [-20^{\circ}, +20^{\circ}]$ and $\boldsymbol{L}_{\mathbf{back}}, \boldsymbol{R}_{\mathbf{back}} \in [-10^{\circ}, +10^{\circ}]$
	B. Back Bent	$\boldsymbol{F}_{\mathbf{back}} \in (-60^{\circ}, -20^{\circ})$ or $(+20^{\circ}, +60^{\circ})$
	C. Back Bent Heavily	$\boldsymbol{F}_{\mathbf{back}} \in (-90^{\circ}, -60^{\circ})$ or $(+60^{\circ}, +180^{\circ})$
Arms	**A.** Both below Shoulder Level	Both $\boldsymbol{F}_{\mathbf{arms}} \in [-90^{\circ}, +90^{\circ}]$
	B. One at or above Shoulder Level	One $\boldsymbol{F}_{\mathbf{arm}} \in (+90^{\circ}, +180^{\circ})$
	C. Both at or above Shoulder Level	Both $\boldsymbol{F}_{\mathbf{arms}} \in (+90^{\circ}, +180^{\circ})$
Legs	**A.** Standing	One or Both $\boldsymbol{F}_{\mathbf{legs}} \in [0^{\circ}, +20^{\circ})$
	B. Knees Bent	One or Both $\boldsymbol{F}_{\mathbf{leg}} \in [+20^{\circ}, +150^{\circ})$
	C. Squatting	Both $\boldsymbol{F}_{\mathbf{leg}} \in [+150^{\circ}, +180^{\circ})$

Note: $-\boldsymbol{F}_{\mathbf{back}}$ indicates backward extension; $+\boldsymbol{F}_{\mathbf{back}}$ indicates forward flexion; $-L_{\text{back}}$ indicates left lateral bending; $+L_{\text{back}}$ indicates right lateral bending; $-\boldsymbol{R}_{\mathbf{back}}$ indicates right rotation; $+\boldsymbol{R}_{\mathbf{back}}$ indicates left rotation; $-\boldsymbol{F}_{\mathbf{arm}}$ indicates backward stretch; $+\boldsymbol{F}_{\mathbf{arm}}$ indicates forward stretch; $\boldsymbol{F}_{\boldsymbol{leg}}$ indicates knee bending angle

be recognized in one video frame captured by the ordinary RGB camera so that different combinations of back, arms, and legs postures could be classified simultaneously. Frequency and duration of different postures were monitored during an observation period. The frequency values were ranked as four corrective action categories as shown in Fig. 1. Once the frequency value of a recognized posture during a working period exceeds a fixed limit, the corresponding corrective action category would change from lower to higher, indicating the urgency of corrective ergonomic interventions.

We acquired the 2D locations of 13 skeletons, which comprised a head-neck part, left/right shoulder, left/right upper arms, left/right forearms, left/right half part of trunk, left/right thigh, left/right lower leg, as shown in Fig. 2. Skeletons and joints were encoded in the corresponding colors. In some cases, appearance and viewpoint variation, crowding, occlusion, range limits of the camera, and other common imaging artifacts could cause motion information losses. Taking the picture on the left side of Fig. 2 as an example, due to the occlusion of the subject's right lower leg, the location of the ankle joints of his right lower leg losses, this might affect the accuracy of posture classification. Therefore, considering motion information losses and viewpoint variance, we collected a large 2D dataset of ergonomic postures from a variety of viewpoints. We also selected distinguishable view-invariant features for person posture classification and recognition via an ordinary RGB camera.

Fig. 2. 2D motion capture containing viewpoint variation, crowding, occlusion, and range limits of camera (Author Captured).

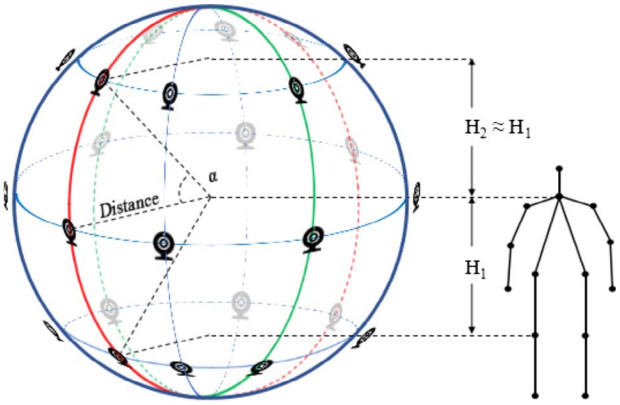


Fig. 3. Heights, distances, and viewpoints of an RGB camera subject to body joints.

3.2. Posture data sampling

To collect objective and reliable samples for ergonomic posture classification based on the quantitative definition, we established a real world "quasi-observability sphere" by choosing different camera viewpoints, distances, and heights to capture each defined ergonomic posture, as shown in Fig. 3. First, we selected three distances regarding 2 m, 4 m, and 6 m from the camera lens to the joint of neck and trunk. For each distance, we captured eight camera viewpoints by dividing the

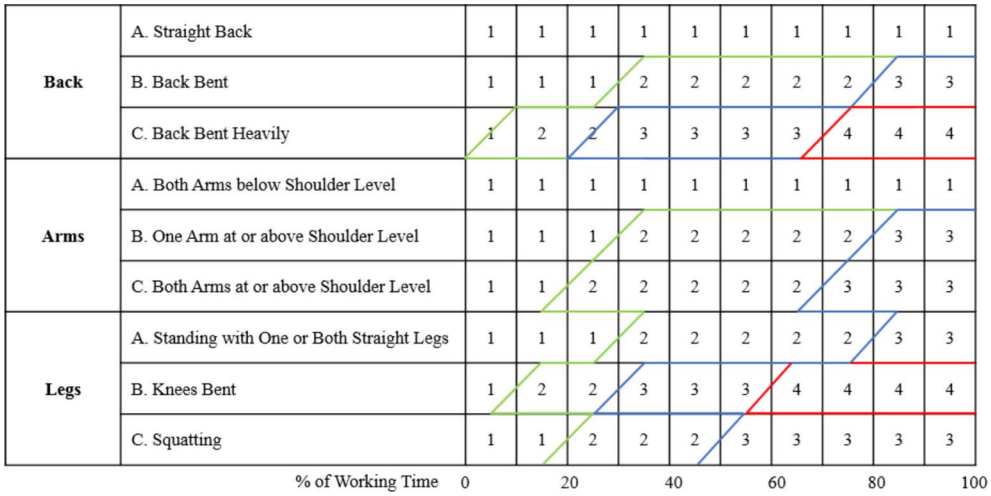


Fig. 1. Corrective action categories considering frequency of postures (based on the OWAS).

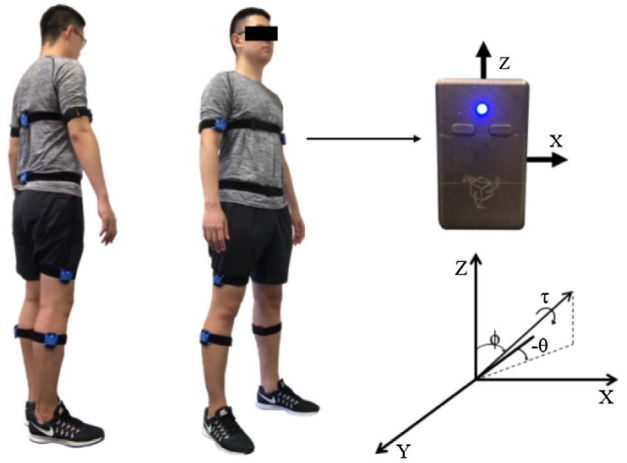


Fig. 4. Calibration state (left); Illustration of tilt angle (ϕ), tilt azimuth angle (θ) and twist angle (τ).

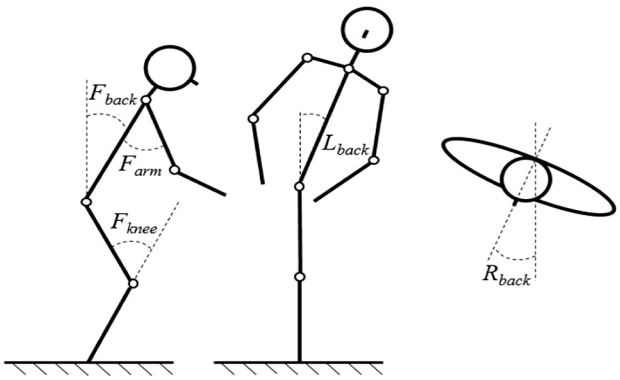


Fig. 5. Flexion-extension angle, lateral bending angle, and rotation angle of targeted body parts.

circle that has a radius of the current distance and a center point of neck-trunk joint into eight equal areas. Three heights of camera lens above level ground were selected subject to the length between neck and knee (H_1, H_2). The camera pitch angle α was determined by the distance. The reason we designed the real world "quasi-observability sphere" for sampling was that these camera locations were "general viewpoints" which granted isotropic probabilities to the viewing orientations around the observed scene [4].

To increase the representativeness of posture data sample, we took individual differences into consideration and recruited six people in different body weights and heights. To quantitatively distinguish each ergonomic posture, they were asked to wear IMUs (YEI 3-Space SensorTM Bluetooth) on targeted body parts in the process of video recording, which could capture real-time quaternions to represent body movements. The outputted quaternion ($q = [w,xi,yj,zk]$) was adjusted to be more accurate and reliable by a built-in Kalman filter. Meanwhile, each IMU was firmly mounted on the targeted body parts using a strap and rubber sleeve to hold the sensor, which could avoid rational noises caused by unstable adherence. They were also asked to recalibrate the sensors frequently during posture sampling, which can reduce the accumulated errors in IMUs. The location of each IMU was selected according to anatomical landmarks as suggested by Plagenhoef et al. [37] and Faber et al. [13]. To capture the flexion-extension angle, lateral bending angle, and rotation angle of each targeted body part, we adopted the algorithm developed by Yan et al. [55]. We located one IMU on upper trunk to capture trunk flexion angle; one on upper trunk and one on lower back to capture rotation angle of trunk; one on upper arm and one on upper trunk to capture the flexion angle of upper arm; one on thigh, and one on lower leg to capture leg flexion angle. For each IMU sensor, Two initial unit vectors were utilized to represent the calibrated state of each sensor: one is a positive unit vector in a quaternion format with w = 0 along the left handed Y-axis denoted p_1 to determine tilt (ϕ) and tilt azimuth (θ) angles, the other is a negative unit vector in the same quaternion format along the Z-axis, denoted p_2, to estimate twist (τ) angles. The calibration state of the IMUs is shown in Fig. 4. Taking the IMU that is located on upper trunk as an example for illustration, the coordinates after rotation of both p_1 and p_2 in quaternion formats can be represented in Eqs. (1) and (2):

$$p_1' = [0,(-2w_1{\cdot}z_1 + 2x_1{\cdot}y_1)i,(w_1^2{-}x_1^2 + y_1^2{-}z_1^2)j,(2w_1{\cdot}x_1 + 2y_1{\cdot}z_1)k] \quad (1)$$

$$p_2' = [0,(-2w_2{\cdot}y_2{-}2x_2{\cdot}z_2)i,(2w_2{\cdot}x_2{-}2y_2{\cdot}z_2)j,(-w_2^2 + x_2^2 + y_2^2{-}z_2^2)k] \quad (2)$$

Trunk inclination tilt (ϕ), tilt azimuth (θ) and twist (τ) angles can then be determined by Eqs. (3)–(5):

$$\phi = \arccos(w_1^2{-}x_1^2 + y_1^2{-}z_1^2) \quad (3)$$

$$\theta = \arctan\left(\frac{2w_1{\cdot}z_1{-}2x_1{\cdot}y_1}{-2w_1{\cdot}x_1 + 2y_1{\cdot}z_1}\right) \quad (4)$$

$$\tau = \arctan\left(\frac{2w_2{\cdot}y_2 + 2x_2{\cdot}z_2}{-2w_2{\cdot}x_2{-}2y_2{\cdot}z_2}\right) \quad (5)$$

Flexion-extension, lateral bending, and rotation angle of targeted body part can be calculated from each sensor's tilt angle (ϕ), tilt azimuth angle (θ), and twist angle (τ): $F = \phi{\cdot}\cos(\theta)$, $L = \phi{\cdot}\sin(\theta)$, $R = \tau$. The predefined body angles for posture classification were approximately calculated: $\boldsymbol{F_{back} = F_{upper\ trunk\ IMU}}$; $\boldsymbol{R_{back} = R_{upper\ trunk\ IMU} - R_{lower\ back\ IMU}}$; $\boldsymbol{F_{arms} = F_{upper\ trunk\ IMU} + F_{upper\ arm\ IMU}}$; $\boldsymbol{F_{legs} = 180° - (L_{thigh\ IMU} - L_{lower\ leg\ IMU})}$, as shown in Fig. 5. Flexion angles of legs were determined by the lateral angle of IMUs on the thigh and lower leg because the sensors were located on the side of thigh and lower leg.

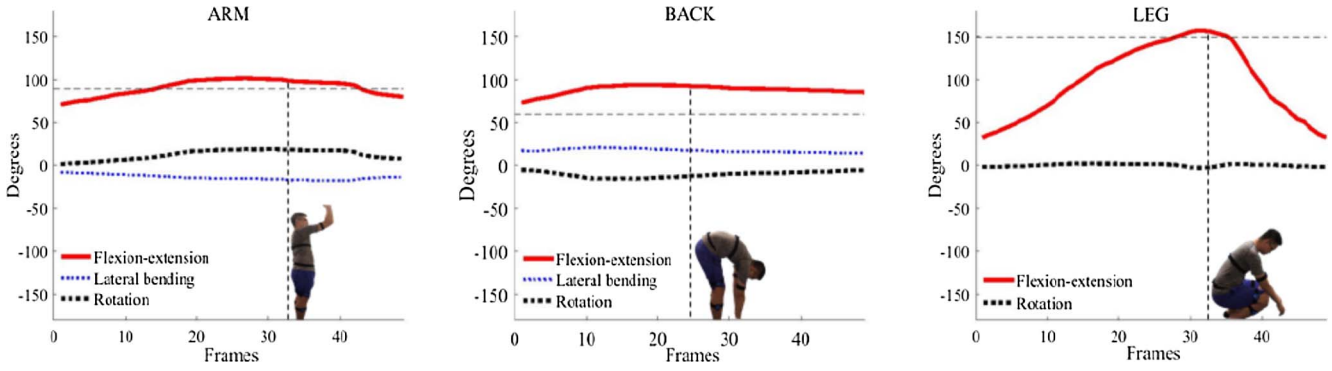


Fig. 6. IMU-based posture data sampling.

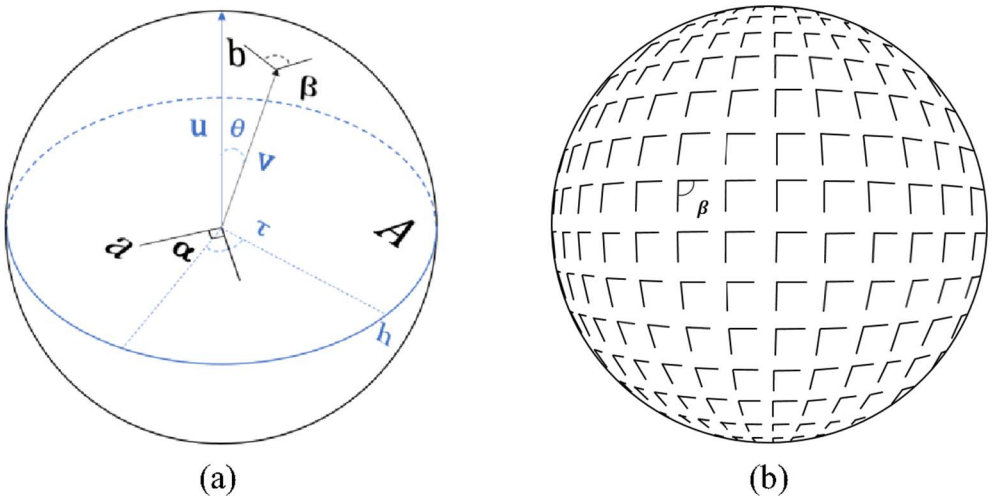


Fig. 7. (a) Sphere plot of β (α = 90°) and skeleton length; (b) Projections of 3D arm posture ($\boldsymbol{F}_{\mathbf{arm}=90^\circ}$) from different viewpoints in the same distance.

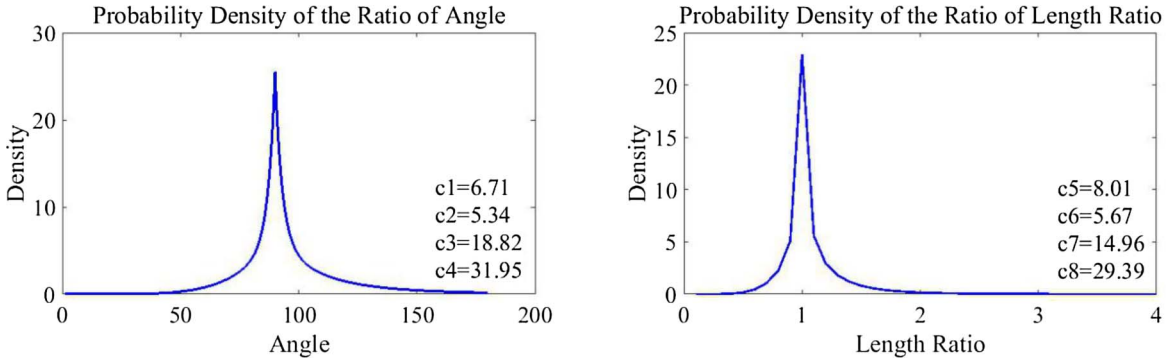


Fig. 8. Approximation of probability density of angle and length ratio.

Table 2
Feature extraction from viewed angle.

Classifier	Feature of viewed angle (β)	Feature of viewed length ratio
Back	$\log(\alpha_i/\angle\overrightarrow{h}\overrightarrow{j})$,$\log(\alpha_i/\angle\overrightarrow{h}\overrightarrow{k})$, $\log(\alpha_i/\angle\overrightarrow{i}\overrightarrow{j})$,$\log(\alpha_i/\angle\overrightarrow{i}\overrightarrow{k})$, i = 1,2,3. α_1 = 10°,α_2 = 40°,α_3 = 120°	$\log\left[(R_a)/\left(\frac{\lvert h\rvert}{\lvert j\rvert}\right)\right]$,$\log\left[(R_a)/\left(\frac{\lvert h\rvert}{\lvert k\rvert}\right)\right]$ $\log\left[(R_a)/\left(\frac{\lvert i\rvert}{\lvert j\rvert}\right)\right]$,$\log\left[(R_a)/\left(\frac{\lvert i\rvert}{\lvert k\rvert}\right)\right]$
Arms	$\log(\alpha_i/\angle\overrightarrow{i}\overrightarrow{e})$,$\log(\alpha_i/\angle\overrightarrow{h}\overrightarrow{e})$, $\log(\alpha_i/\angle\overrightarrow{i}\overrightarrow{d})$,$\log(\alpha_i/\angle\overrightarrow{h}\overrightarrow{d})$, i = 1,2. α_1 = 45°,α_2 = 135°	$\log\left[(R_a)/\left(\frac{\lvert c\rvert}{\lvert e\rvert}\right)\right]$,$\log\left[(R_a)/\left(\frac{\lvert e\rvert}{\lvert g\rvert}\right)\right]$,$\log\left[(R_a)/\left(\frac{\lvert e\rvert}{\lvert i\rvert}\right)\right]$, $\log\left[(R_a)/\left(\frac{\lvert b\rvert}{\lvert d\rvert}\right)\right]$,$\log\left[(R_a)/\left(\frac{\lvert d\rvert}{\lvert f\rvert}\right)\right]$,$\log\left[(R_a)/\left(\frac{\lvert d\rvert}{\lvert h\rvert}\right)\right]$
Legs	$\log(\alpha_i/\angle\overrightarrow{j}\overrightarrow{l})$,$\log(\alpha_i/\angle\overrightarrow{k}\overrightarrow{m})$, i = 1,2,3 α_1 = 10°,α_2 = 85°,α_3 = 165°	$\log\left[(R_a)/\left(\frac{\lvert i\rvert}{\lvert k\rvert}\right)\right]$,$\log\left[(R_a)/\left(\frac{\lvert k\rvert}{\lvert m\rvert}\right)\right]$ $\log\left[(R_a)/\left(\frac{\lvert h\rvert}{\lvert j\rvert}\right)\right]$,$\log\left[(R_a)/\left(\frac{\lvert j\rvert}{\lvert l\rvert}\right)\right]$

Note: R_a is the ratio of real-world individual length ratio.

After above preparation, each person was instructed to perform the predefined postures freely in situ and each person's movements for approximately 60 minutes were recorded (excluding rest time, 5 s for each posture at each viewpoint). To match video data and IMU output, we set the data capture frequency of all the IMUs equal to the frame number of the camera (30 frames/second). Finally, the image samples of different ergonomic postures were collected based on the quantitative definition of awkward postures. Fig. 6 shows three posture examples in 50 frames, including both arms above shoulder level, back bent heavily, and squatting. The solid red lines indicate the flexion angle; the blue dotted fine lines indicate the lateral bending angle; the black dotted thick lines indicate the rotation angles. Three predefined angle thresholds of different ergonomic postures were used to label each posture as a predefined posture. It should be noted that in the right subfigure labeled as "LEG", the lateral bending angle was not illustrated due to the limited degree of freedom of the joint of human knees.

4. Experiments and data interpretation

4.1. View-invariant 2D feature extraction

As one essential step in machine learning classification problem,

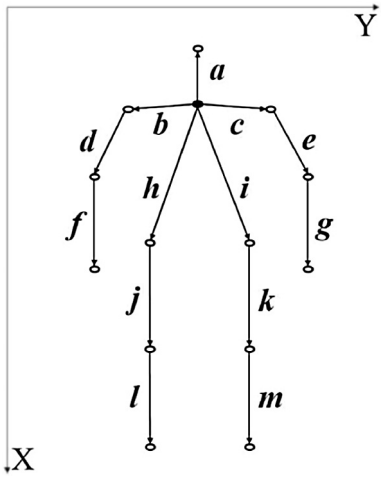


Fig. 9. Representing body angle and length ratio using skeleton vectors.

selection and manipulation of the features of raw data might significantly affect the accuracy of a learned classifier [9,48]. The selected raw data features should carry the common characteristics of each predefined category, while shared by all the categories. In this study, the raw data set consisted of locations of pixels detected as human body skeletons and joints in 2D images. It was not proper to employ them as features to learn the person posture classifiers due to the variation of 2D features with respect to different camera viewpoints. This problem involved the variation in the 2D projection as a function of viewpoint and camera distances with respect to an object. Take the recognition of a 3D upper arm posture ($\boldsymbol{F}_{\mathbf{arm}} = 90^{\circ}$) for example, as shown in Fig. 6. With the distance between the camera and the arm posture being constant, Fig. 7(a) depicts the observability sphere of the orthographic projection as an approximation of the actual central projection [4]. By this approximation, the functions between a real-world 3D angle/length and the corresponding projected 2D angles/lengths are in Eqs. (6) and (7):

$$\text{Angle: } \tan\beta = \cos\theta\tan\alpha \Big/ \left[1-\frac{1}{2}\sin^2\theta\left(1-\frac{\cos 2\tau}{\cos\alpha}\right)\right] \tag{6}$$

$$\text{Length: } \boldsymbol{b} = s\left[\boldsymbol{a}-\boldsymbol{v}(\boldsymbol{a}\cdot\boldsymbol{v})\right] \tag{7}$$

where α, β are 3D and 2D angles respectively. The slant angle, denoted θ, is between the normal $\boldsymbol{v}$ to the imaging plane and the normal $\boldsymbol{u}$ to the angle plane A. The tilt angle τ is between the bisector of α and the parallel line of intersection between the imaging plane and the angle plane. $\boldsymbol{a}$ and $\boldsymbol{b}$ indicate 3D and 2D skeleton lengths respectively. In Eq. (7), the image scale coefficient s changes with the camera distance with respect to the subject, which could be approximately eliminated by calculating the proportion of a skeleton length to the length captured at a special viewpoint ($\theta = \tau = 0$). These estimated functions describe the projections of a real-world posture that vary with different cameral viewpoints, as intuitively shown in Fig. 7b**)**.

Besides arm postures, view variance produced three challenges to PPR of all body parts. The first challenge was inter-class similarity of postures from distinct categories. Workers could perform actions in different directions and distances relative to a camera with distinctive characteristics of body segments movements, thus two different postures were only distinguished by subtle details. The second challenge was intra-class variability of postures in the same category. Homogeneous postures might be perceived as heterogeneous geometric structure as perceived in different directions and distances with respect to the camera. Third, the number of viewpoints was theoretically infinite. It was impossible to cover all real-world training samples. Compared with 3D recognition developed by Xia et al. [54] that contained depth information, or 2D action recognition proposed by Rao et al. [40] and Parameswaran and Chellappa [35] that contained continuous time sequencing characteristics of human activities, the recognition performance of static postures in 2D imagery developed in this study was more vulnerable to those challenges due to a lack of additional temporal or spatial information.

Table 3
Comparison of different supervised classification learner (* denotes maximum classification accuracy).

Body part	Classification learner	Optimal parameters and functions	Accuracy
Back	BP-ANN	I. 4 hidden layers II. 10 neurons/layer III. Transfer: tansig IV. Training: Levenberg-Marquardt V. Performance: MSE	93.0%
	DT	I. Maximum number of splits: 1000 II. Split criterion: Maximum deviance reduction III. Surrogate decision splits: On IV. Maximum number of surrogate: 5	93.2%
	KNN	I. Number of neighbors: 11 II. Distance metric: City block III. Distance weight: Squared inverse IV. Data standardized	93.6%
	EC	I. Ensemble method: Bag II. Learner type: Decision trees III. Maximum number of splits: 8 IV. Number of learners: 30 V. Learning rate: 0.1 VI. Subspace dimension: 1	**95.0%** *
Arms	BP-ANN	I. 4 hidden layers II. 15 neurons/layer III. Transfer: Tansig IV. Training: Levenberg-Marquardt V. Performance: MSE	90.7
	DT	I. Maximum number of splits: 145 II. Split criterion: Maximum deviance reduction III. Surrogate decision splits: Off	93.5%
	EC	I. Ensemble method: Bag II. Learner type: Decision trees III. Maximum number of splits: 5 IV. Number of learners: 29 V. Learning rate: 0.1 VI. Subspace dimension: 1	**96.5%***
Legs	BP-ANN	I. 3 hidden layers II. 10 neurons/layer III. Transfer: Tansig IV. Training: Levenberg-Marquardt V. Performance: MSE	91.6%
	DT	I. Maximum number of splits: 300 II. Split criterion: Maximum deviance reduction III. Surrogate decision splits: Off	96.2%
	SVM	I. Kernel: Gaussian II. Box constraint level: 4 III. Kernel scale: Heuristic subsampling IV. Multiclass method: one-vs-one V. Data standardized	90.1%
	KNN	I. Number of neighbors: 4 II. Distance metric: City block III. Distance weight: Squared inverse IV. Data standardized	95.1%
	EC	I. Ensemble method: Bag II. Learner type: Decision trees III. Maximum number of splits: 8 IV. Number of learners: 28 V. Learning rate: 0.1 VI. Subspace dimension: 1	**97.6%***

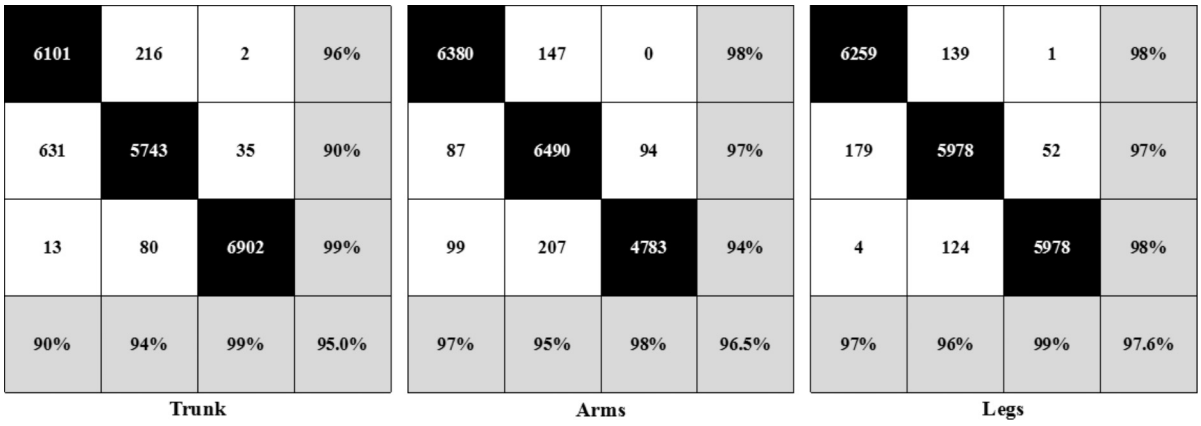


Fig. 10. Total confusion matrix of 5-fold cross validation (left column: true positive rate; bottom row: positive predictive value; bottom right: average accuracy).

4.2. Projection probability density

Although it was proved that general-case view-invariant did not exist for any number of points, given weak perspective, true perspective, or orthographic projection, there were types of features whose view variations were sufficiently low to make discrimination training feasible. In this study, we extracted probability density features of the logarithmic ratio of angle and the logarithmic ratio of length ratio from detected articulated skeletal system represent and discriminate different postures.

To introduce the probability density of the ratio of the original real angle (captured with $\theta = \tau = 0$) to all the potential projected angles in different viewpoints; and the ratio of the original length ratio (captured with $\theta = \tau = 0$) to all the potential length ratios in different viewpoints, we used the observability sphere, denoted O, with a radius R, as shown in Fig. 7(a). The ratios of an angle and its two arms were defined by Eqs. (8) and (9):

$$R_{angle} = \alpha/\beta \tag{8}$$

$$R_{length} = \left(\frac{|\boldsymbol{a_2}|}{|\boldsymbol{a_1}|}\right) \Big/ \left(\frac{|\boldsymbol{b_2}|}{|\boldsymbol{b_1}|}\right) \tag{9}$$

In an orthographic projection, the image scale coefficient denoted s, and length of $\boldsymbol{a_1}$ or $\boldsymbol{a_2}$ does not influence R_{length}.

As a view orientation, the vector $\boldsymbol{v}$ represents each point on sphere O. The imaging plane (camera scene) is normal to $\boldsymbol{v}$, which is also tilted by τ and slant by θ relative to the equator plane A. The observation probability, denoted $\Delta p(dO)$, of a sphere area element dO on the overall observability sphere surface is proportional to (θ, τ) where:

$$\Delta p(dO) = dA/\boldsymbol{4}\pi R^2 = \frac{1}{4\pi}\sin(\theta)d\theta d\tau \tag{10}$$

The probability density of a ratio is the summation of all $\Delta p(dO)$ that belong to a subinterval of R_{angle} or R_{length}, which can also be approximated by Eqs. (11) and (12):

$$p_{angle} = c_1 e^{(-c_2|\log R_{angle}|)} + c_3 e^{(-c_4|\log R_{angle}|)} \tag{11}$$

$$p_{length} = c_5 e^{(-c_6|\log R_{length}|)} + c_7 e^{(-c_8|\log R_{length}|)} \tag{12}$$

where coefficients c_1–c_8 were optimized by minimum square error procedures. Taking $\alpha = 90^\circ$ and $|\boldsymbol{a_2}|/|\boldsymbol{a_1}| = 1$ as two examples, the probability density curves of p_{angle} and p_{length} are depicted in Fig. 8.

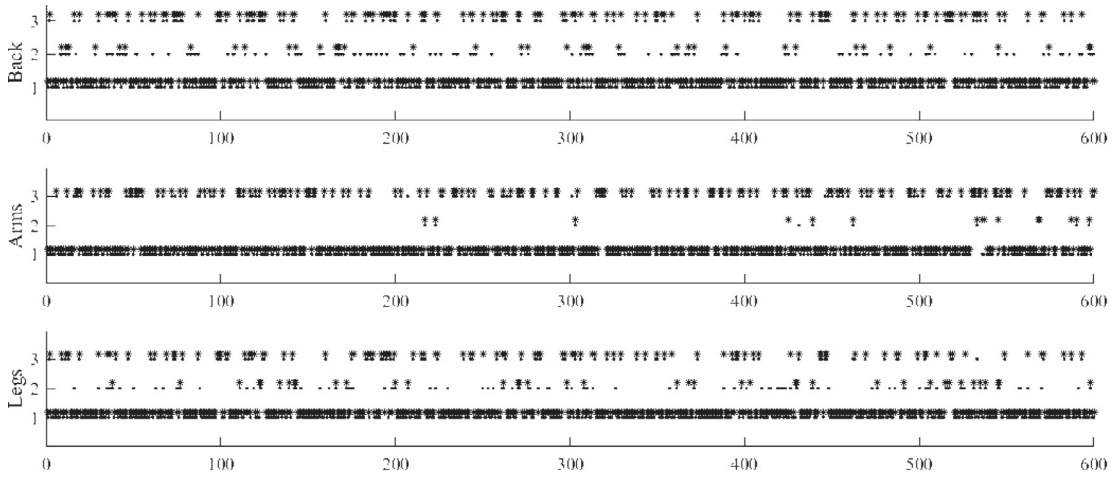
The probability density function derived from an original angle or an original length ratio was view-invariant with respect to most real-world camera viewpoints ($0^\circ \le \theta < 90^\circ$, $0 \le \tau < 180^\circ$). In addition, the peaking effect of the probability density function guaranteed a high probability (nearly 90%) that a viewed angel/length ratio would have values between half to double of the original angle/length ratio. Hence, given a viewed angle or viewed length ratio with the projection orientation being a prior unknown, the probability density value of the ratio of a predefined angle/length ratio over the viewed angle/length ratio were unique features whose view variations were sufficiently low to make discrimination training feasible.

4.3. Angle and length ratio feature extraction

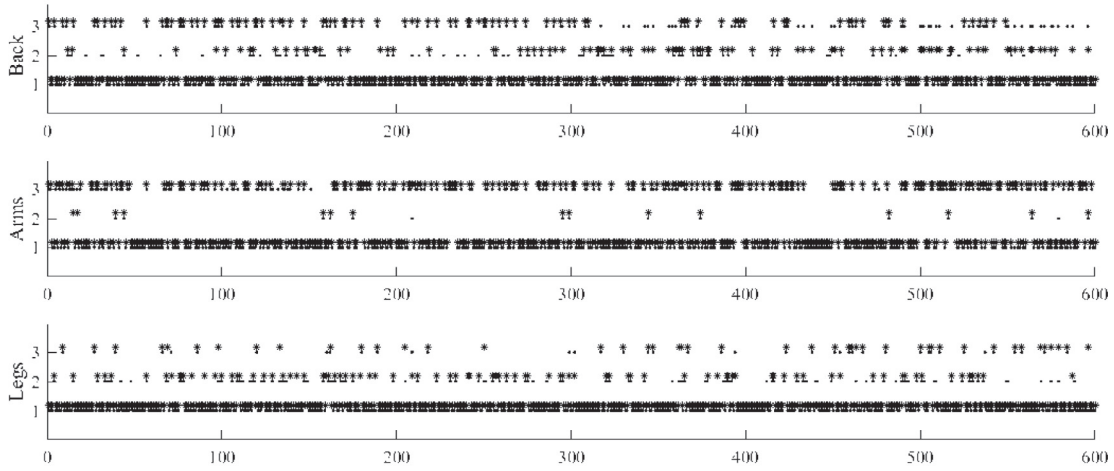
For the angle, because of the orientation was a priori unknown, the original value of a projected angle was unknown. Therefore, we defined the original value by choosing them from each posture category. We extracted angle features using the probability density of the ratio of the predefined original angle over the viewed angle. Because of the symmetry of the predefined probability density function, we chose the mean values of an angle interval that represented a posture category as shown in the second column of Table 2. For length ratio, since the length of body skeleton is a constant, the original ratio captured with $\theta = \tau = 0$ is also approximately a constant. Therefore, we used real-world length ratios as the original ratio. To increase the representativeness of posture data sample, we took individual differences into consideration by recruiting 6 people in different body weights and heights. All the features derived from length ratio were listed in the third column of Table 2. For each targeted body part, we only selected a set of refined features so that three classifications regarding back, arms, and legs postures were conducted simultaneously in a video frame. Each viewed Angle β between corresponding skeletons was represented by corresponding vectors from $\boldsymbol{a}$-$\boldsymbol{m}$, as shown in Fig. 9. Assuming occlusion was also a type of features in 2D posture recognition, if the endpoints of a skeleton vector were blocked and not detected by the camera, the corresponding viewed angle and length ratio would be set to 0.

4.4. Supervised ergonomic posture classification

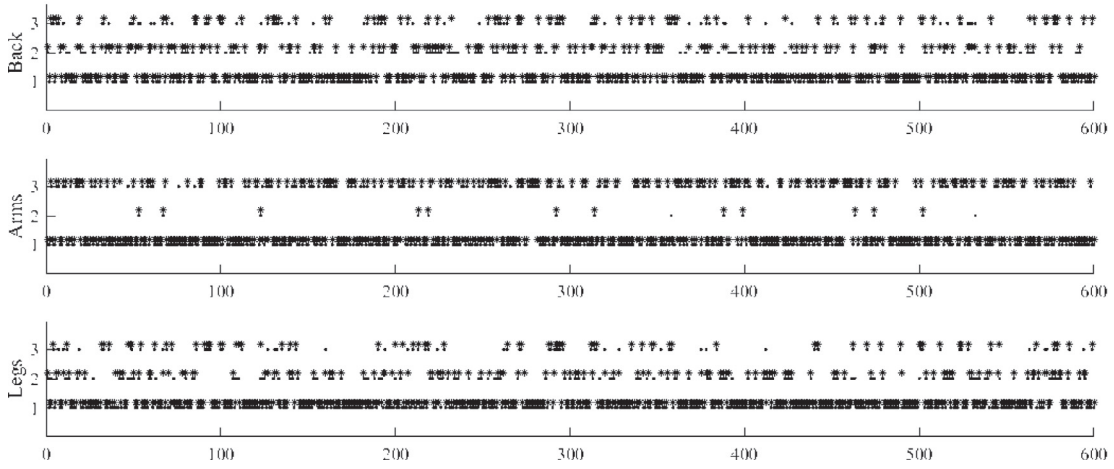
Since different supervised classification learners had various performance on the different data structures, we learned the classifiers by testing several types of mature machine learning models, including: back-propagation artificial neural network (BP-ANN), decision trees (DT), support vector machines (SVM), K-nearest neighbor classifiers (KNN), and ensemble classifiers (EC). For each supervised learning method, we selected model parameters and functions by comparing intragroup performance on the same feature set. The final accuracy of each classification learner was tested by a 5-fold cross validation (7:3). The optimal classifier was determined by comparing intergroup accuracy. The selected classification learner and corresponding parameters, functions, and accuracy (> 90%) are listed in Table 3. According to this



(a) Activity one



(b) Activity two



(c) Activity three

Fig. 11. Prediction vs. ground truth.

procedure, the optimal classifier for human back was the bagged decision trees with an average classification accuracy at 95.0% (Maximum number of splits: 8, Number of learners: 30, Learning rate: 0.1, Subspace dimension: 1). For human arms, the bagged decision trees outperformed others with and average classification accuracy at 96.5% (Maximum number of splits: 5, Number of learners: 29, Learning rate: 0.1, Subspace dimension: 1). For human legs, the bagged decision trees outperform others with an average classification accuracy of 97.6% (Maximum number of splits: 8, Number of learners: 28, Learning rate: 0.1, Subspace dimension: 1). The corresponding total confusion matrices of 5-fold cross validation are shown in Fig. 10.

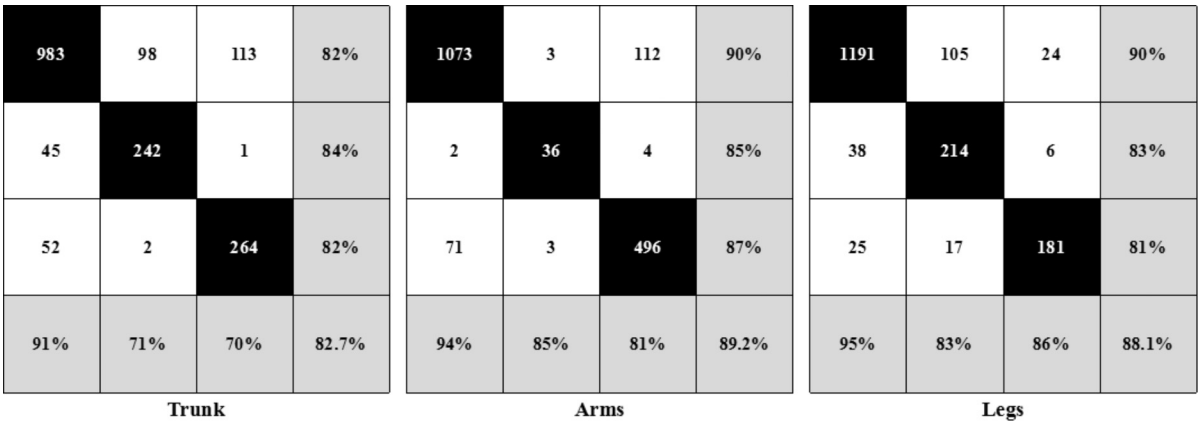


Fig. 12. Confusion matrix of all the new testing data (left column: true positive rate; bottom row: positive predictive value; bottom right: average accuracy).

Table 4
Performance on ergonomic hazard assessment.

	Body part	Overall accuracy	A				B				C			
			p	p-s	r	r-s	p	p-s	r	r-s	p	p-s	r	r-s
1	Back	89.2%	0.61	1	0.71	1	0.20	1	0.07	1	0.19	2	0.22	2
	Arms		0.70	1	0.72	1	0.02	1	0.02	1	0.28	2	0.26	2
	Legs		0.69	2	0.76	2	0.16	2	0.06	1	0.15	1	0.18	2
2	Back	88.3%	0.64	1	0.67	1	0.13	1	0.16	1	0.23	2	0.17	2
	Arms		0.59	1	0.61	1	0.03	1	0.03	1	0.38	2	0.36	2
	Legs		0.75	2	0.78	2	0.19	1	0.15	2	0.06	1	0.07	1
3	Back	87.6%	0.55	1	0.61	1	0.24	1	0.25	1	0.21	3	0.14	2
	Arms		0.62	1	0.65	1	0.02	1	0.02	1	0.36	2	0.33	2
	Legs		0.65	2	0.66	2	0.21	2	0.22	2	0.14	1	0.12	2

Notes: Overall accuracy indicates the percentage of the situation when 'p-s = r-s' during the observational period. **p**: percentage in prediction; **p-s**: predicted ergonomic hazard score; **A, B, C**: posture categories; **r**: percentage in real postures; **r-s**: ergonomic hazard score of real postures.

5. Validation and discussions

To further validate the developed View-Invariant 2D feature extraction method and the learned posture classifier for ergonomic assessment in construction sites, three activities were recorded in a laboratory experiment to examine the accuracy of posture recognition and predicted ergonomic hazard score of each targeted body part. We recruited a new participant to conduct the validation. The participant was instructed to wear IMU sensors as shown in Fig. 4. All the postures in the experiment were assigned posture labels by the IMU outputs according to the predefined posture categories. These labels served as the ground truth for the validation. To test the ability of generalization of the three classifiers trained by View-Invariant features, three new camera locations that had not been included in the sampling procedure were chosen. The first camera was located between the ground level and the kneel level; the second camera was located between the kneel level and the head level; the third camera was located above the head level (the ground, kneel, and head level is shown in Fig. 3). During each activity, the participant was instructed to lift a box using back bending or squatting postures from the ground and to put it on an overhead surface by one hand or both hands. The participant could move freely within the camera range to do the task using different postures. The sequence and duration of postures were decided by the participant. Three 10-minute estimated postures and real postures during the experiment were illustrated in Fig. 11. Index 1–3 indicate posture A, B, and C in each body part as listed in Table 1. The middle frame was selected to represent each second in the video records. The blue dots indicate the predictions, while the red asterisks indicates the real postures. The confusion matrices of all the new testing data regarding three body parts are shown in Fig. 11 (see Fig. 12).

Based on the predictions of three activities, the overall performance of the three trained classifiers regarding ergonomic assessment based on OWAS criteria were calculated, as shown in Table 4. Once these ergonomic hazard score during the observation period exceeded fixed limits, the urgency of corrective protective ergonomic actions would increase.

The overall accuracy of the three classifiers regarding ergonomic hazard assessment scores based on the OWAS were 89%, 88%, and 87% respectively. There was no significant difference among the classification performance to recognize postures from different camera viewpoints in the experiment, which indicated that the classifiers based on View-Invariant features can be generalized to different viewpoints within camera range. Based on a comparison between the prediction and the ground truth, we found that there were some frequently occurred scenes leading to misclassification. For the arm classifier, the recognition of arm posture A was vulnerable when the participant squatted or bended back heavily facing the camera. For the leg classifier, the discrimination between leg posture A and B was vulnerable when the participant bended back heavily. For the back classifier, the major misclassifications derived from the recognition of back posture B. It was vulnerable when the participant bended legs or squat. Our primary diagnosis of these misclassifications based on the captured video frames was that the feature of occlusion was not distinguishable regarding back and legs postures recognition, which was a major obstacle affecting the classification performance of the three trained classifiers. During the feature extraction procedure, the angle and length ratio of occluded skeletons were simplified into 0, which might affect the accuracy of the posture classifier due to a lack of skeleton position information. One possible solution is to estimate possibilities of the potential areas where both endpoints of the occluded skeletons locate in a

2D image. Such estimation should base on structural information regarding line segments network and metric information such as orientation, relative position, and length on one skeleton with respect to another as suggested by Burns et al. [6].

6. Conclusions

We developed a posture classification method based on 2D ordinary camera to monitor postural ergonomic hazards of construction workers. On the basis of OWAS posture assessment criteria, we classified working postures and collected posture samples using a set of predefined ergonomic rules designed to determine whether or not a worker needs corrective actions for ergonomic purposes. Considering view variance of projection during video recording, we used probability density features of angle and length ratio extracted from 2D skeleton motion data captured by a monocular RGB camera. These features' view variations were sufficiently low to make discrimination training feasible. Three classifiers in terms of back, arms, and legs were trained using different machine learning methods. According to the average classification accuracy derived from a 5-fold cross-validation, three bagged trees classifiers in terms of arms, trunk, and legs with different parameters outperformed other classifiers. Finally, an experiment was conducted to validate the portability and extensibility of the developed classifiers, whose performance could satisfy prospective postural ergonomic monitoring and assessment to improve workers' safety and health in construction sites. The contributions of this study are twofold. First, our approach enables three ergonomic postures regarding arms, trunks, and legs to be recognized simultaneously in one video frame. Second, we used view-invariant features as the representation of ergonomic postures and proved it effective at posture classification in 2D imagery so that the classifiers were effective in various viewpoints and distances of the ordinary camera. The outputs of this study could lead to more realistic postural ergonomic assessment in outdoor construction environments where ordinary RGB cameras are widely installed. The developed technique could also be employed in other worker-behavior-based analysis by fine-tuning posture definitions. Investigations on the feasibility and effectiveness of this technique for the development of more representative and distinguishable features of occluded 2D skeletons are recommended for future study.

Acknowledgements

The work was supported by the Hong Kong Construction Industry Council (CIC) project "Waistband Enabled Construction Workers Low Back Health Monitoring System" (No. K-ZB64) and the Research Grants Council of Hong Kong with the grant titled "Proactively Monitoring Construction Progress by Integrating 3D Laser-scanning and BIM" (PolyU 152093/14E) as well as the National Natural Science Foundation of China Grant (Grant no. 71390524). This project was also supported by the Specialized Research Fund for the Doctoral Program of Higher Education of China (No. 20130101110058).

References

[1] S. Ali, M. Shah, Human action recognition in videos using kinematic features and multiple instance learning, IEEE Trans. Pattern Anal. Mach. Intell. 32 (2) (2010) 288–303.
[2] M.F. Antwi-Afari, H. Li, D.J. Edwards, E.A. Pärn, J. Seo, A.Y.L. Wong, Biomechanical analysis of risk factors for work-related musculoskeletal disorders during repetitive lifting task in construction workers, Automat. Constr. 83 (2017) 41–47.
[3] S. Baker, G.K.M. Cheung, T. Kanade, Shape-from-silhouette of articulated objects and its use for human body kinematics estimation and motion capture, in: 2003 IEEE Conference on Computer Vision and Pattern Recognition, vol. 01, 2003, pp. 77, doi: 10.1109/CVPR.2003.1211340.
[4] J. Ben-Arie, The probabilistic peaking effect of viewed angles and distances with application to 3-D object recognition, IEEE Trans. Pattern Anal. Mach. Intell. 12 (8) (1990) 760–774.
[5] B. Buchholz, V. Paquet, L. Punnet, D. Lee, S. Moir, Path: a work sampling-based approach to ergonomic job analysis for construction and other non-repetitive work, Appl. Ergon. 27 (3) (1996) 177–187.
[6] J.B. Burns, Richard S. Weiss, E.M. Riseman, View variation of point-set and line-segment features, IEEE Trans. Pattern Anal. Mach. Intell. 15 (1) (1993) 51–68.
[7] J. Chen, J. Qiu, C. Ahn, Construction worker's awkward posture recognition through supervised motion tensor decomposition, Automat. Constr. 77 (2017) 67–81.
[8] J.H.V. Dieen, M.J. Hoozemans, H.M. Toussaint, Stoop or squat: a review of biomechanical studies on lifting technique, Clin. Biomech. 14 (10) (1999) 685–696.
[9] P. Domingos, A few useful things to know about machine learning, Commun. ACM 55 (10) (2012) 78–87.
[10] S. Eaves, D.E. Gyi, A.G.F. Gibb, Building healthy construction workers: their views on health, wellbeing and better workplace design, Appl. Ergon. 54 (2016) 10–18.
[11] A.A. Efros, A.C. Berg, G. Mori, J. Malik, Recognizing action at a distance, IEEE Int. Conf. Comput. Vis. 2 (1) (2003) 726–733.
[12] V. Escorcia, M.A. Dávila, M. Golparvar-Fard, J.C. Niebles, Automated vision-based recognition of construction worker actions for building interior construction operations using RGBD cameras, in: Construction Research Congress 2012: Construction Challenges in a Flat World, 2012, pp. 879–888.
[13] G.S. Faber, I. Kingma, S.M. Bruijn, J.H. Van Dieën, Optimal inertial sensor location for ambulatory measurement of trunk inclination, J. Biomech. 42 (14) (2009) 2406–2409.
[14] A. Golabchi, S. Han, J. Seo, S. Han, S. Lee, M. Al-Hussein, An automated biomechanical simulation approach to ergonomic job analysis for workplace design, J. Constr. Eng. Manage. 141 (8) (2015) 04015020.
[15] C. Goodvin, E.J. Park, K. Huang, K. Sakaki, Development of a real-time three-dimensional spinal motion measurement system for clinical practice, Med. Biol. Eng. Comput. 44 (12) (2006) 1061–1075.
[16] M.S. Grewal, V.D. Henderson, R.S. Miyasako, Application of Kalman filtering to the calibration and alignment of inertial navigation systems, IEEE Trans. Autom. Control 36 (1) (1991) 4–13.
[17] H. Guo, Y. Yu, M. Skitmore, Visualization technology-based construction safety management: a review, Automat. Constr. 73 (2017) 135–144.
[18] S. Han, M. Achar, S. Lee, F. Peña-Mora, Empirical assessment of a RGB-D sensor on motion capture and action recognition for construction worker monitoring, Vis. Eng. 1 (1) (2013) 1–13.
[19] S. Han, S. Lee, F. Peña-Mora, Vision-based detection of unsafe actions of a construction worker: case study of ladder climbing, J. Comput. Civ. Eng. 27 (6) (2013) 635–644.
[20] S. Han, S. Lee, F. Peña-Mora, Comparative study of motion features for similarity-based modeling and classification of unsafe actions in construction, J. Comput. Civ. Eng. 28 (5) (2014) A4014005.
[21] S. Hignett, L. Mcatamney, Rapid entire body assessment (REBA), Appl. Ergon. 31 (2) (2000) 201–205.
[22] A.M. Jarkas, Effect of buildability on labor productivity: a practical quantification approach, J. Constr. Eng. Manage. 142 (2) (2016) 06015002.
[23] H. Jiang, M.S. Drew, Z.N. Li, Successive convex matching for action detection, in: Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference, 2006, pp. 1646–1653.
[24] I. Kant, J.H.V. Notermans, P.J.A. Borm, Observations of working postures in garages using the ovako working posture analysing system (OWAS) and consequent workload reduction recommendations, Ergonomics 33 (2) (1990) 209–220.
[25] O. Karhu, P. Kansi, I. Kuorinka, Correcting working postures in industry: a practical method for analysis, Appl. Ergon. 8 (4) (1977) 199–201.
[26] Y. Kong, Z. Ding, J. Li, Y. Fu, Deeply learned view-invariant features for cross-view action recognition, IEEE Trans. Image Process. (2017), http://dx.doi.org/10.1109/TIP.2017.2696786.
[27] S. Kraatz, J. Lang, T. Kraus, E. Munster, E. Ochsmann, The incremental effect of psychosocial workplace factors on the development of neck and shoulder disorders: a systematic review of longitudinal studies, Int. Arch. Occup. Environ. Health 86 (4) (2013) 375–395.
[28] H. Laitinen, M. Marjamäki, K. Päivärinta, The validity of the tr safety observation method on building construction, Accid. Anal. Prev. 31 (5) (1999) 463–472.
[29] I. Laptev, M. Marszalek, C. Schmid, B. Rozenfeld, Learning realistic human actions from movies, in: IEEE Conference on Computer Vision and Pattern Recognition, 2008, doi: 10.1109/CVPR.2008.4587756.
[30] J. Liu, S. Ali, M. Shah, Recognizing human actions using multiple features, in: IEEE Conference on Computer Vision and Pattern Recognition, 2008, doi: 10.1109/CVPR.2008.4587527.
[31] M. Mattila, M. Vilkki, OWAS methods, in: W. Karwowski, W.S. Marras (Eds.), The Occupational Ergonomics Handbook, second ed., CRC Press, 1998.
[32] L. Mcatamney, E.N. Corlett, Rula: a survey method for the investigation of work-related upper limb disorders, Appl. Ergon. 24 (2) (1993) 91–99.
[33] J.Y.K. Ng, H.S.C. Alan, The ageing construction workforce in Hong Kong: a review, Lecture Notes Eng. Comput. Sci. 2216 (1) (2015) 1–7.
[34] O. Okunribido, T. Wynn, Ageing and work-related musculoskeletal disorders: a review of the recent literature. Health and Safety Executive Research Report RR799. Health and Safety Laboratory, Buxton: Health and Safety Executive, 2010.
[35] V. Parameswaran, R. Chellappa, View invariance for human action recognition, Int. J. Comput. Vis. 66 (1) (2006) 83–101.
[36] S. Pellegrini, L. Iocchi, Human posture tracking and classification through stereo vision and 3D model matching, Eurasip J. Image Video Process. 1 (2008) 1–12.
[37] S. Plagenhoef, F.G. Evans, T. Abdelnour, Anatomical data for analyzing human motion, Res. Q. Exerc. Sport 54 (2) (1983) 169–178.
[38] R. Poppe, A survey on vision-based human action recognition, Image Vis. Comput. 28 (6) (2010) 976–990.

[39] W. Premerlani, P. Bizard, Direction Cosine Matrix IMU: Theory, Diy Drone, NY, 2009.
[40] C. Rao, A. Yilmaz, M. Shah, View-invariant representation and recognition of actions, Int. J. Comput. Vis. 50 (2) (2002) 203–226.
[41] S.J. Ray, J. Teizer, Real-time construction worker posture analysis for ergonomics training, Adv. Eng. Inform. 26 (2) (2012) 439–455.
[42] M.C. Schall, N.B. Fethke, H. Chen, S. Oyama, D.I. Douphrate, Accuracy and repeatability of an inertial measurement unit system for field-based occupational studies, Ergonomics 59 (4) (2016) 591–602.
[43] J. Seo, S. Han, S. Lee, H. Kim, Computer vision techniques for construction safety and health monitoring, Adv. Eng. Inform. 29 (2) (2015) 239–251.
[44] Y. Shen, H. Foroosh, View-invariant action recognition from point triplets, IEEE Trans. Pattern Anal. Mach. Intell. 31 (10) (2009) 1898–1905.
[45] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, A. Blake, Real-time human pose recognition in parts from single depth images, in: IEEE Conference on Computer Vision and Pattern Recognition, vol. 56(1), 2011, pp. 1297–1304.
[46] J.S. Silva Jr., L.R. Correa, L.C. Morrone, Evaluation of lumbar overload in hotel maids, Work 41 (Suppl. 1) (2012) 2496–2498.
[47] L. Straker, A. Campbell, J. Coleman, M. Ciccarelli, W. Dankaerts, In vivo laboratory validation of the physiometer: a measurement system for long-term recording of posture and movements in the workplace, Ergonomics 53 (5) (2010) 672–684.
[48] H.A.L. Thi, H.M. Le, P.D. Tao, Feature selection in machine learning: an exact penalty approach using a Difference of Convex function Algorithm, Mach. Learn. 101 (1–3) (2014) 163–186.
[49] W. Umer, H. Li, G.P.Y. Szeto, A.Y.L. Wong, Identification of biomechanical risk factors for the development of lower-back disorders during manual rebar tying, J. Constr. Eng. Manage. 143 (1) (2016) 04016080.
[50] E. Valero, A. Sivanathan, F. Bosche, M. Abdel-Wahab, Musculoskeletal disorders in construction: a review and a novel system for activity tracking with body area network, Appl. Ergon. 54 (2016) 120–130.
[51] D. Wang, F. Dai, X. Ning, Risk assessment of work-related musculoskeletal disorders in construction state-of-the-art review, J. Constr. Eng. Manage. 141 (6) (2015) 04015008.
[52] I.P.T. Weerasinghe, J.Y. Ruwanpura, J.E. Boyd, A.F. Habib, Application of microsoft kinect sensor for tracking construction workers, in: Construction Research Congress 2012: Construction Challenges in a Flat World, 2012, pp. 858–867.
[53] Y. Wu, T.S. Huang, View-independent recognition of hand postures, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2000, doi: 10.1109/CVPR.2000.854749.
[54] L. Xia, C.C. Chen, J.K. Aggarwal, View-invariant human action recognition using histograms of 3d joints, in: Comput Vision & Pattern Recognition Workshops, 2012, pp. 20–27.
[55] X. Yan, H. Li, A.R. Li, H. Zhang, Wearable IMU-based real-time motion warning system for construction workers' musculoskeletal disorders prevention, Automat. Constr. 74 (2017) 2–11.
[56] W. Yi, A. Chan, Health profile of construction workers in Hong Kong, Int. J. Environ. Res. Publ. Health 13 (12) (2016) 1–15.