Automation in Construction 94 (2018) 1–10

Contents lists available at ScienceDirect

Automation in Construction

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

Ergonomic posture recognition using 3D view-invariant features from single ordinary camera

Hong Zhang[a], Xuzhong Yan[a,b,*], Heng Li[b]

[a] Institute of Construction Management, College of Civil Engineering and Architecture, Zhejiang University, Hangzhou, China

[b] Department of Building and Real Estate, Faculty of Construction and Environment, the Hong Kong Polytechnic University, Hong Kong, China

ARTICLE INFO

Keywords:
Ergonomics
Person posture recognition
3D view-invariant
Relative 3D joint position
Joint angle
Convolutional neural network
Construction worker
Safety and health

ABSTRACT

Manual construction tasks are physically demanding, requiring prolonged awkward postures that can cause pain and injury. Person posture recognition (PPR) is essential in postural ergonomic hazard assessment. This paper proposed an ergonomic posture recognition method using 3D view-invariant features from a single 2D camera that is non-intrusive and widely installed on construction sites. Based on the detected 2D skeletons, view-invariant relative 3D joint position (R3DJP) and joint angle are extracted as classification features by employing a multi-stage convolutional nerual network (CNN) architecture, so that the learned classifier is not sensitive to camera viewpoints. Three posture classifiers regarding arms, back, and legs are trained, so that they can be simultaneously classified in one video frame. The posture recognition accuracies of three body parts are 98.6%, 99.5%, 99.8%, respectively. For generalization ability, the relevant accuracies are 94.9%, 93.9%, 94.6%, respectively. Both the classification accuracy and generalization ability of the method outperform previous vision-based methods in construction. The proposed method enables reliable and accurate postural ergonomic assessment for improving construction workers' safety and healthy.

1. Introduction

Work-related musculoskeletal disorders (WMSDs) are common occupational hazards in the manually demanding construction industry. In Hong Kong, the Pilot Medical Examination Scheme (PMES) for Construction Workers revealed that 41% of registered workers have musculoskeletal pain [1]. In the United States, 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 to 64 years doubled [2]. In an ergonomic perspective, it is suggested the frequency and duration of awkward postures regarding trunk, upper and lower limbs be controlled within acceptable ranges [3–6]. Traditional observational methods require safety personnel to collect posture data through site observations and questionnaires, which may be inaccuracy and inefficient due to subjective bias [7]. Such limitations are significant on construction sites due to insufficient capable workforce and continuously changing environments [8].

To address the problem of manual observation methods, there are currently two technical streams in person posture recognition (PPR), one is wearable sensor-based methods, and the other is computer vision-based methods. Compared with the wearable sensor system (e.g. YEI 3-Space Sensor), optical system (e.g. Vicon) and depth cameras (e.g., Microsoft Kinect), the ordinary surveillance camera (common camera) is more practical for ergonomic posture capture in construction, because it is non-intrusive and has been widely installed in the construction industry for surveillance [9]. In this study, the videos captured by a single 2D ordinary camera are depended for ergonomic posture recognition.

One challenge in PPR from a single ordinary camera is huge variances in the projection of intra-class postures, and similar projection of inter-class postures while being viewed from different 2D camera viewpoints [10]. Therefore, the objective of this study is to propose an ergonomic posture recognition method using view-invariant features, i.e. relative 3D joint positions (R3DJPs) [11] and joint angles that are estimated from 2D video frames captured by a single ordinary camera for field surveillance. To achieve the research objective, view-invariant 3D skeletons are estimated by lifting 2D coordinates into 3D using a multi-stage convolutional neural network (CNN). The captured body is divided into 16 3D skeletons with 12 joints and 5 end points. Then discriminative R3DJPs and joint angle features are extracted for arms, back and legs respectively for ergonomic posture classification based on quantitatively defined ergonomic postures according to a classical ergonomic rule, the Ovako Working Posture Analysis System (OWAS) [4]. Using the view-invariant features extracted from video frame samples,

* Corresponding author at: Institute of Construction Management, College of Civil Engineering and Architecture, Zhejiang University, Hangzhou, China.
E-mail address: 11512075@zju.edu.cn (X. Yan).

https://doi.org/10.1016/j.autcon.2018.05.033
Received 6 March 2018; Received in revised form 30 May 2018; Accepted 31 May 2018
0926-5805/

this study tested different machine learning methods in terms of classification performances. According to the average classification accuracy in 5-fold cross-validations, three optimal classifiers are selected for ergonomic posture classification regarding arms, trunk and legs respectively.

The contributions of this study are threefold. First, the paper proposes a view-invariant ergonomic posture recognition approach using 3D body skeletons and joints estimated from single 2D camera. Second, view-invariant R3DJPs and joint angle features are used as the discriminative representation of ergonomic postures in each 2D video frame. Third, trained by 3D view-invariant features, the three classifiers regarding arms, back and legs are tested to be effective in various camera viewpoints, also outperform previous vision-based ergonomic posture recognition methods in terms of average accuracy and generalization ability [12–15].

2. Background

2.1. Postural ergonomic assessment rules

WMSDs in construction are mainly caused by compression, shear, tensile stress and muscle force repeatedly acting on load bearing tissues [16]. Most of these loadings in the construction industry can be attributed to repetitive works in prolonged awkward working postures, including overhead reaching, stooping and squatting [17, 18]. From a perspective of ergonomic assessment, the frequency and duration of non-ergonomic postures of targeted body parts should be monitored and controlled so as to identify the hazardous working pattern and job site layout [3–6]. Accordingly, many ergonomic assessment rules have been proposed for postural ergonomic hazards monitoring and control. Representative research on awkward posture assessment rules include the “Rapid Upper Limb Assessment” (RULA), an ergonomic assessment tool focusing on upper limbs [6]; the “Ovako Working Posture Analyzing System” (OWAS) for identifying and evaluating working postures [4]; the ISO 11226:2000 for determining the acceptable angles and holding times of working postures [5], and the EN 1005-4 as a guidance when designing machinery component parts in assessing machine-related postures and body movements, i.e. during assembly, installation, operation, maintenance, repair and dismantlement [19].

The objective of these ergonomic assessment rules is to provide a quantitative and systematic criterion to identify and control postural ergonomic hazards in workplace. For example, the OWAS classified the posture combinations of arms, back and legs and relative proportions of certain postures during work time into four action categories based on the risk assessment of WMSDs [4, 20]. In the OWAS, the action categories range from 1 that requires no corrective actions to 4 that needs corrective measures immediately. When the proportion of a certain posture during the observation period is larger than the frequency threshold defined by the OWAS, the action category changes from lower to higher, which indicates the urgency of corrective actions is increasing. Similar to monitoring non-ergonomic posture frequency, some ergonomic rules provide acceptable thresholds of angles and holding times of targeted body parts that are prone to WMSDs. For example, the ISO 11226:2000 provides ergonomic guidelines for workplace design or redesign of jobs and products with the basic concepts of ergonomics in general and working postures in particular. The international standard provides recommended limits for working postures without any, or only with minimal external force exertion, while considering joint angles and holding time aspects.

The ergonomic assessment rules require posture data as input for ergonomic hazards analysis and control in workplace. At the time when the ergonomic assessment rules were developed, posture data were mainly collected through observation and questionnaires [8]. These manual data collection methods are labor-intensive and time-consuming, inconsistent and unreliable due to subjective bias and inactive workers' participation [8, 21, 22], which impose practical constraints on ergonomic assessment in construction. With the fast development and iterations of motion data acquisition technology, the traditional manual observation and questionnaire methods can be replaced by many advanced motion capture technologies. However, postural ergonomic assessment rules do not fade and still play an essential role in safety and health management in construction. For example, Xinming, et al. [23] created a 3D model to imitate and animate manual construction tasks in a virtual environment based on the RULA. They analyzed body joint angles from 3D visualization based on the traditional ergonomic assessment rule to identify postural ergonomic hazards. Some researchers also applied the RULA to establish a virtual 3D workplace for proactive ergonomic design of construction workplace [24]. Based on the ISO 11226:2000, Yan, et al. [25] developed a real-time motion warning system that enables workers' self-management of ergonomic hazards in operational pattern using wearable Inertial Measurement Units (WIMUs). Compared with traditional work-related postural ergonomic assessment methods that focus on the design of ergonomic rules in workplace, some studies in construction focus more on the combination of the well-developed ergonomic assessment rules and the advanced motion capture technology considering specific industrial contexts [12, 25, 26], based on which automated, accurate and reliable ergonomic assessment can be performed for the monitoring and control of work-related ergonomic hazards.

In this research, typical awkward postures are defined based on the OWAS that have been validated in many jobs in different industrial contexts [4], as shown in Table 1. The frequency of each posture during work time is defined as different action categories ranging from 1 (no actions required) to 4 (corrective measures needed immediately). Once the frequency value of a recognized posture during a working period exceeds its limit, the corresponding action category will change from lower to higher, indicating the urgency of corrective ergonomic interventions.

2.2. Ergonomic posture capture systems

Ergonomic posture capture is a process used to detect, track and record an object's postures that involve ergonomic hazards, based on which automated postural ergonomic assessment and control can be conducted. Various motion capture technologies have been actively developed that can serve as the ergonomic posture capture system, including wearable sensor-based system, depth sensor-based system and ordinary surveillance camera-based system.

A wearable sensor-based system captures motion data by using a set of portable sensors attached on the targeted body parts of a wearer. The most popular wearable sensor in ergonomic assessment is the wearable Inertial Measurement Unit (WIMU) sensor. The WIMUs have satisfactory performance in accuracy [25, 27–29]. However, they need to be attached tightly to the wearer's body to prevent output noise caused by unstable adherence. It has been revealed by the front-line construction

Table 1
Action category considering frequency of postures [4].

% of working time		10	20	30	40	50	60	70	80	90	100
Arms	A. Both arms below shoulder level	1	1	1	1	1	1	1	1	1	1
	B. One arm at or above shoulder level	1	1	1	2	2	2	2	2	3	3
	C. Both arms at or above shoulder level	1	1	2	2	2	2	2	3	3	3
Back	A. Straight back	1	1	1	1	1	1	1	1	1	1
	B. Back bent	1	1	1	2	2	2	2	2	3	3
	C. Back bent heavily	1	2	2	3	3	3	3	4	4	4
Legs	A. Standing with one or both straight legs	1	1	1	2	2	2	2	2	3	3
	B. Knees bent	1	2	2	3	3	3	4	4	4	4
	C. Squatting	1	1	2	2	2	3	3	3	3	3