Automation in Construction 91 (2018) 53–61

Contents lists available at ScienceDirect

Automation in Construction

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

Falls from heights: A computer vision-based approach for safety harness detection

Weili Fang[a,b], Lieyun Ding[a,b,*], Hanbin Luo[a,b], Peter E.D. Love[c]

[a] Dept. of Construction Management, School of Civil Engineering and Mechanics, Huazhong University of Science and Technology, Wuhan, Hubei, China
[b] Hubei Engineering Research Center for Virtual, Safe and Automated Construction, (ViSAC), HUST, China
[c] Dept. of Civil Engineering, Curtin University, Perth, Western Australia, Australia

ARTICLE INFO

Keywords:
Convolution neural network
Falls from height
Harness
Unsafe behavior

ABSTRACT

Falls from heights (FFH) are major contributors of injuries and deaths in construction. Yet, despite workers being made aware of the dangers associated with not wearing a safety harness, many forget or purposefully do not wear them when working at heights. To address this problem, this paper develops an automated computer vision-based method that uses two convolutional neural network (CNN) models to determine if workers are wearing their harness when performing tasks while working at heights. The algorithms developed are: (1) a Faster-R-CNN to detect the presence of a worker; and (2) a deep CNN model to identify the harness. A database of photographs of people working at heights was created from activities undertaken on several construction projects in Wuhan, China. The database was then used to test and train the developed networks. The precision and recall rates for the Faster R-CNN were 99% and 95%, and the CNN models 80% and 98%, respectively. The results demonstrate that the developed method can accurately detect workers not wearing their harness. Thus, the computer vision-based approach developed can be used by construction and safety managers as a mechanism to proactively identify unsafe behavior and therefore take immediate action to mitigate the likelihood of a FFH occurring.

1. Introduction

Falls from heights (FFH) are a major problem in construction [1–6]. Research has revealed that FFH account for approximately 48% of serious injuries and 30% of fatalities [7]. Numerous safety policies and procedures have been established to protect people working at heights in construction [8]. For example, scaffolds/platforms and the use fall prevention solutions such as travel restraints systems (e.g. lines and belts) are required when working above a certain height [9].

Yet, despite the considerable amount of research that has been undertaken and the implementation of policies, procedures and the development of protection measures, FFH remain a pervasive problem, particularly for scaffolders and roofers [10]. In China, for example, people working above a height of two metres are required by law to use fall arrest equipment [10]. There has, however, been a reluctance from scaffolders to use a harnesses, in spite of its use being a legal requirement and workers being cognizant of their exposure to a fall [10]. Reasons for such non-compliance have been found to be attributable to discomfort while wearing the harness and the restrictions it place on movement [10]. While such reasons may well have a degree of validity, such workers tend to have a poor awareness and risk perception. Thus, good communication, effective consultation, improved training and reasonable adjustments can often be enough to head off objections to wearing a harness.

But more fundamentally, behavioral and cultural change is required to address the reluctance to wear personal protective equipment (PPE), but this can take a considerable amount of time to implement. To expedite and enact behavioral change, it is suggested that real-time monitoring of harnesses that are worn by people working at heights can contribute to preventing falls. Construction and safety managers require practical methods to monitor and ensure workers are using their harnesses, particularly scaffolders. However, the safety inspection process can be toilsome and is often undertaken intermittently [11]. As a result, safety compliance is unable to be assured and therefore the likelihood a FFH remains a risk.

To address this problem, the research presented in this paper develops an automatic and non-invasive approach using a computer-vision-based method to monitor the use of harnesses. Computer vision-based methods have been widely used in construction. For example, to track workers on-site [12,13], progress monitoring [14], productivity

* Corresponding author at: Dept. of Construction Management, School of Civil Engineering and Mechanics, Huazhong University of Science and Technology, Wuhan, Hubei, China.
E-mail address: dly@hust.edu.cn (L. Ding).

https://doi.org/10.1016/j.autcon.2018.02.018
Received 15 September 2017; Received in revised form 21 January 2018; Accepted 7 February 2018
0926-5805/

analysis [15], safety and health monitoring [5], automated documentation [5], and postural ergonomic assessment [16].

In comparison with sensing techniques (e.g., Radio Frequency Identification (RFID), Geographical Positioning Systems (GPS) and Ultra-Wide Band (UWB)), which tend to be limited to providing location data for a specific entity being monitored, computer vision can provide a rich set of information (e.g., locations and behaviors of project entities and site conditionss) by analyzing images or videos [5]. While technologies, such as RFID, for instance, have been widely used in construction and applied to an array of PPE types [17], there has been an absence of research that has monitored the use of harnesses. The paper commences by reviewing existing methods that have been used to prevent FFH in construction and then introduces a novel approach based on convolutional neural network (CNN) that can be used to monitor the use of safety harnesses worn by people working at heights. The technical challenges of the developed CNN approach are presented and the implications for future research are identified.

2. Falls from heights

The unique, dynamic, and complex working environment of construction sites and non-standardized design and work procedures can increase workers' exposure to hazards [5]. The prevention of FFH has received a significant amount of attention from construction safety and health management researchers and professionals [18]. Undertaking regular safety inspections and risk assessments to identify hazards has been repeatedly identified in the literature as a core activity to preventing the occurrence of falls [7,19]. A comprehensive review of the FHH literature is therefore eschewed, as this can be found in Nadhim et al. [7]. But for the purposes of brevity key studies that are aligned with the research presented in this paper are drawn upon.

Strategies to prevent and mitigate the severity of injuries can be categorized as being passive or proactive [7]. Strategies of a passive nature are based on analyzing fall accident data to develop future prevention plans. For example, identifying those factors that have contributed to fatal occupational falls from accident reports and acquiring data from regular safety inspections [20]. FFH preventive measures that have been derived from an analysis of accident records and autopsy records include [20]: (1) fixed barriers; (2) travel restraint systems (e.g., belts), fall arrest systems (e.g., harness); and (3) fall containment systems (e.g., nets). Factors that have been found to contribute to roofers FFH include cognitive slips and lapses, weather, and schedule demands [21]. The emergent risk factors contributing to FFH are often prioritized and then used to develop mitigation strategies [22,23]. For example, an automated Building Information Modeling-based safety checking platform that is integrated with safety risks has been developed, which supports fall prevention planning prior to the commencement of construction [24,25].

Proactive strategies are precautionary measures that place emphasis on safety training and education. For example, the implementation of specific fall protection training programs [19]; and the design of short courses, seminars and talks that focus on working the risks of working at heights with the aim to improve people's safety behavior. While enforcement of regulations may increase the use of PPE [1], this is a reactive approach to addressing the issue of safety and does not necessarily change people behaviors [5]. Hence, it is more important to influence the mind-sets, attitudes and culture (i.e. values and beliefs) of workers than solving specific violations [26]. It has been suggested that effective measures to enhance the use of PPE are needed, especially in the context of FFH, as scaffolders, are often reluctant use their harness [10]. The purpose of harness monitoring is to ensure that it is being used correctly by workers and to ensure an organization's safety and health plans and standards are being met.

3. Computer vision-based approaches

Computer vision is an interdisciplinary field of endeavour that deals with how computers can acquire a high-level of understanding from digital images or videos. From an engineering perspective, it seeks to automate tasks that the human visual system is unable to do. Vision-based applications have been developed to capture and process video [27–29]. This had been aided by the development of new algorithms (e.g. Faster R-CNN) that can be used to detect and track resources (e.g., people, plant and equipment), as well as identify the unsafe behavior of workers [30–36].

A fundamental tenet of computer vision-based is action recognition, which is used to exploit handcrafted features (e.g., shapes) from images or videos. To extract features of workers' actions, descriptors such as Histogram of Oriented Gradients (HOG) [37], Histogram of Optical Flow (HOF) [38], and Bag-of-features (Bof) [39] have all been employed to compute on the image or videos. Hand-crafted feature-based methods usually employ a three-stage procedure, which consists of: (1) extraction; (2) representation; and (3) classification.

Image representation that is used to recognize human actions can extract features such as shapes and temporal motions from images. Action recognition features, however, need to contain rich information so that a wide range of actions can be identified and analyzed. Techniques that can be used to analyze such features include classifier tools (e.g. Support Vector Machine (SVM)), temporal state-space models (e.g., Hidden Markov models (HMM), conditional random fields (CRF)), and detection-based methods (e.g., bag-of-words coding). However, the use of these approaches may lead to overfitting and therefore weaken the ability to derive generalizations from a dataset.

Another approach that is often used to collect motion data from stereo videos and to reconstruct a three-dimensional (3D) skeleton model are depth sensors (Kinect™) [5,40–45]. Kinect™ and multiple video cameras have been used to monitor the behavior of workers by estimating the positioning of individual joints in 3D [41–45]. This method provides a useful way to obtain accurate motion data. But more specifically, it provides the ability to record, model, and analyze the human motion that has occurred from the performing an unsafe act. However, monitoring the positioning of workers using 3D can require lengthy computational periods and the line of motion may also be hampered by sensitivities in lighting [46,47].

4. Convolutional neural networks in construction

Deep learning methods that incorporate CNNs have been demonstrated to be effective for computer vision and pattern recognition [48,49]. LeCun et al. [50] developed the LeNet-5 (a CNN model), which recognizes handwritten numbers, based on a dataset created by the Mixed National Institute of Standards and Technology. CNN models can effectively and automatically recognize features from static images by stacking multiple convolutional and pooling layers.

Krizhevsky et al. [49] was the first to achieve substantially high levels of image classification accuracy at the ImageNet Large Scale Visual Recognition Challenge (LSVRC) by training a deep CNN. Since the inception of Krizhevsky's [41] deep CNN, almost all the effective algorithms used for image classification, object recognition, and visual tracking that have been developed are based on this fundamental work. Krizhevsky et al. [49] used deep CNNs to classify 1.2 million high-resolution images in the ImageNet LSVRC-2010 contest into a thousand different classes. Hong et al. [51] proposed an online visual tracking algorithm by learning a discriminative saliency map using a CNN, which provided superior results compared to other state-of-the-art tracking algorithms (e.g., discrete fourier transform (DSK), local sparse and K-selection (LSK), and circulant tructure of tracking-by-detection with kernels (CSK)).

The success of region-based CNNs and region proposal methods has prompted advancements in object detection and their use in