Safety Science 135 (2021) 105130

Contents lists available at ScienceDirect

Safety Science

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

Computer vision technologies for safety science and management in construction: A critical review and future research directions

Brian H.W. Guo[a,*], Yang Zou[b], Yihai Fang[c], Yang Miang Goh[d], Patrick X.W. Zou[e,f,*]

[a] Dept. of Civil & Natural Resources Engineering, University of Canterbury, 69 Creyke Road, Christchurch 8140, New Zealand
[b] Dept. of Civil and Environmental Engineering, University of Auckland, 20 Symonds Street, Auckland 1010, New Zealand
[c] Dept. of Civil Engineering, Monash University, 23 College Walk, Clayton 3800, Australia
[d] Safety and Resilience Research Unit (SaRRU), Dept. of Building, School of Design and Environment, National Univ. of Singapore, 4 Architecture Dr., Singapore 117566, Singapore
[e] Engineering Research Center of Highway Infrastructure Digitalization, Ministry of Education, Chang'an University, Xi'an 710064, China
[f] School of Economics and Management, Chang'an University, Xi'an, Shanxi 710064, China

A R T I C L E I N F O

Keywords:
Computer vision
Construction health and safety
Safety science
Safety culture
Safety Climate, Hazard
Safety management system
Digital technologies
Automation

A B S T R A C T

Recent years have seen growing interests in developing and applying computer vision technologies to solve safety problems in the construction industry. Despite the technological advancements, there is no research that exams the theoretical links between computer vision technology and safety science and management. Thus, the objectives of this paper are to: (1) investigate the current status of applying computer vision technology to construction safety, (2) examine the links between computer vision applications and key research themes of construction safety, (3) discuss the theoretical challenges of applying computer vision to construction safety, and (4) recommend future research directions. A five-step review approach was adopted to search and analyze peer-reviewed academic journal articles. A three-level computer vision development framework was proposed to categorized computer vision applications in the construction industry. The links between computer vision and three main safety research traditions: safety management system, behavior-based safety program, and safety culture, were discussed. The results suggest that the majority of past efforts were focused on object recognition, object tracking, and action recognition, with limited research focused on recognizing unsafe behavior. There are even fewer studies aimed at developing vision-based safety assessment and prediction systems. Based on the review findings, four future research directions are suggested: (1) develop and test a behavioral-cues-based safety climate measure, (2) develop safety behavior datasets, (3) develop a formal hazard identification and assessment model, and (4) develop criteria to evaluate the real impacts of vision-based technologies on safety performance.

1. Introduction

Construction is a pillar industry for economic development and employment worldwide. However, construction sites are hazardous in nature. The industry has been one of the top contributors to workplace injuries and fatalities in many countries and regions and therefore construction safety remains as one of the major issues in academic research and in practice. Traditionally research effort has been focused the policy, management, human and cultural issues of safety and there have been critical reviews on these dimensions, for example, human factors in construction safety (Goh et al., 2018; Guo et al., 2015) and strategic safety management (Zou and Sunindijo, 2015). Effective safety planning and hazard analysis are an essential prerequisite to accident prevention. It has been argued that traditional safety management practices in the construction industry have been manual, time-consuming, selective, and therefore inefficient and error-prone (Zhang et al., 2015b; Zhang et al., 2013). For example, behavior-based safety (BBS) programs still rely on manual observations to collect unsafe behaviors (Guo et al., 2018). Manual observations and inspections are difficult to cover the whole site and monitor all workers. In addition, paper-based hazard identification systems would impede timely risk communication (Zou et al., 2017a).

The past two decades have seen increasing applications of digital technologies and techniques to help improve construction health and

* Corresponding authors.
E-mail addresses: brian.guo@canterbury.ac.nz (B.H.W. Guo), yang.zou@auckland.ac.nz (Y. Zou), yihai.fang@monash.edu (Y. Fang), bdggym@nus.edu.sg (Y.M. Goh), pxwzou@outlook.com (P.X.W. Zou).

https://doi.org/10.1016/j.ssci.2020.105130
Received 12 June 2019; Accepted 16 December 2020
Available online 6 January 2021
0925-7535/

safety management (Guo et al., 2017a). Examples of digital technologies include, but are not limited to, building information modelling (BIM) (Zou et al., 2017b), tracking and positioning technologies (Teizer et al., 2013), augmented reality and virtual reality (Wang et al., 2013), and artificial intelligence (Goh and Guo, 2018; Guo and Goh, 2017). These digital technologies have demonstrated great potential to improve safety planning, hazard management, and safety training and education. Comprehensive summaries of the research in this topic area can be found in (Guo et al., 2017a; Guo et al., 2017b; Li et al., 2018).

Recent years have also seen growing interests in computer vision applications to address the problems in the architecture, engineering, and construction (AEC) industry. Computer vision is an interdisciplinary area that deals with how computers can provide enriched information to support and achieve a high-level understanding of objects and events present in a scene through the analysis of digital images or videos. In recent years, significant research efforts have been made to recognize and track physical elements of engineering projects (e.g., building elements, tools and equipment, materials, and workers) (Teizer, 2015). These efforts have laid a strong foundation for more advanced analysis and assessment, like understanding of construction activities (Seo et al., 2015), productivity assessment (Ibrahim et al., 2009), quality analysis (Akinci et al., 2006), and real-time structural health of bridges and highways (Fraser et al., 2009). These efforts also made the applications of computer vision to construction health and safety (H&S) possible, and there have been important initiatives to automate H&S processes. For example, Yu et al. (2017) tested the feasibility and accuracy of using computer vision for recognizing three types of unsafe construction behaviors. Han and Lee (2013) proposed a new computer vision-based framework for unsafe action detection and behavior monitoring. Similarly, Fang et al. (2018d) developed a set of computer vision algorithms to detect workers not wearing harnesses.

Seo et al. (2015) reviewed computer vision techniques for the construction health and safety monitoring. The study was focused on evaluating unsafe behavior from a technical and practical perspective (e.g., image sensing devices, camera position, viewpoints, etc.). However, safety is a multi-faceted concept and consists of other important dimensions, like safety culture, safety climate, and safety management system (Cooper, 2000a,b), and it relates to not only human behavior but also the surrounding environmental conditions such as moving machines, equipment and objects. To the best of the authors' knowledge, there is no research that exams the theoretical links between computer vision technology and safety science and management. To further reap the benefits of computer vision for construction health and safety, it is important to review and evaluate the theoretical links to these key dimensions.

The objectives of this paper are to: (1) investigate the current status of applying computer vision to construction safety, (2) examine the links between computer vision applications and key research themes of construction safety (i.e., safety management system, behavior-based safety program, and safety culture), (3) discuss the theoretical challenges of applying computer vision to construction safety, and (4) recommend future research directions.

2. Methodology

This study applied the systematic review approach adopted by Zhou et al. (2015). The approach consists of five main steps: (1) literature search, (2) literature selection, (3) literature coding, (4) data analysis, and (5) discussion.

2.1. Literature search

Scopus is chosen as the database in this study, as it covers a wider journal range (i.e. over 22,000 journals) than Web of Science and Google Scholar (Falagas et al., 2008). Search attributes and their values are presented in Table 1.

Table 1
First search methods.

Search attributes	Values used in the search
Database	Scopus
Keywords and Boolean operators	"Computer vision" OR "object recognition", OR "object tracking", OR "action recognition" AND "construction", OR "health and safety", OR "risk assessment", OR "hazard", OR "accident", OR "incident", OR "safety"
Search scope	Article title, abstract, or keywords
Published year	from all years to present
Subject area	Engineering; Computer Science; Social Science; Psychology;
Source /Document type	Journal article
Language	English

As this study focuses only on the applications of computer science to construction health and safety management, only four relevant subject areas were considered, including Engineering, Computer Science, Social Science, and Psychology. Other subject areas like Mathematics, Physics and Astronomy, Material Science, Medicine, Earth, and Planetary Sciences were excluded. Only journal articles were considered in this study. The initial search has resulted in a total of 1711 journal papers.

Following the systematic review approach (Zhou et al., 2015), a secondary search was conducted by picking up journals that are relevant to the construction industry and safety science. Eleven journals were selected to perform the secondary search, including Automation in Construction, Journal of Computing in Civil Engineering, Advanced Engineering Informatics, Journal of Construction Engineering and Management, Journal of Management in Engineering, Accident Analysis and Prevention, Computer-Aided Civil and Infrastructure Engineering, Canadian Journal of Civil Engineering, and Electronic Journal of Information Technology in Construction, and Safety Science.

Keywords and Boolean operators, "computer vision" AND "construction", were implemented to search relevant papers in these journals. The secondary search has identified 893 journal papers.

2.2. Literature selection

Note that the secondary search did not focus only on the applications of computer vision to construction health and safety, rather it considered general applications of computer vision to construction. This consideration is based on the fact that vision-based health and safety management largely relies on object recognition and tracking. Including the foundational works can help understand the evolution of computer vision for construction health and safety. A preliminary review was performed to determine if the 893 papers should be kept for in-depth review and analysis based on the following filter criteria:

Table 2
Development levels of computer vision.

Development Level		Function	Key research questions
L1: Detection, recognition, and tracking	L1.1	Object detection and recognition	Is something there? What is the object?
	L1.2	Object tracking	Where is the object? Where is the object headed?
	L1.3	Action recognition	What is the object doing?
L2: Assessment	L2.1	Object assessment	Is the object a hazard? Is the object in an unsafe or unhealthy state?
	L2.2	Behavior assessment	Is the action unsafe or unhealthy?
	L2.3	Condition assessment	Is the working condition (scenario) unsafe?
L3: Prediction	L3.1	Behavior prediction	How will the object behave?
	L3.2	Incident prediction	Will the next incident occur?

Table 3
Relevant paper number by journals.

Journal title	Number of papers
Automation in Construction	69
Journal of Computing in Civil Engineering	27
Advanced Engineering Informatics	23
Journal of Construction Engineering and Management	9
Accident Analysis and Prevention	8
Canadian Journal of Civil Engineering	4
Computer-Aided Civil and Infrastructure Engineering	4
Safety Science	3
Electronic Journal Of Information Technology In Construction	3
Construction Innovation	3

(1) Only research articles, technical papers, case study, review papers were kept, while all book reviews, editorials, and conference papers were excluded.
(2) Articles that only mentioned "computer vision" but do not focus were removed.
(3) Articles that apply computer vision to the robot, manufacturing, structural assessment, inspection, and defect and crack detection were also excluded, as they do not represent the foundational works for vision-based health and safety management.

As a result, a total of 165 papers were retained for in-depth review and analysis.

2.3. Literature coding

All remaining papers were coded according to (1) title, (2) publication year, (3) journal title, (4) country or region (all authors were counted), (5) development level.

Based on the literature review, we classified the development and application of computer vision in the construction industry into three different levels: L1 detection, recognition and tracking, L2 assessment, and L3 prediction. As shown in Table 2, Level 1 can be further decomposed into three sub-levels: L1.1 object detection and recognition, L1.2 object tracking, and L1.3 action recognition. These three levels were proposed mainly based on a health and safety perspective, as applications at different levels have different implications on construction health and safety management.

The three sub-levels ask three key research questions (1) is something there? (2) what is the object? (3) where is the object? and (4) what is the object doing? respectively. Computer vision applications at the Level 2 aim to determine if an object is a hazard (i.e., L2.1 object assessment) and an action is safe or not (i.e., L2.2 behavior assessment) or assess the risk level of the work conditions which consist of building elements, workers, equipment, tools, and materials (i.e., L2.3 condition assessment). Key research questions at this level include: (1) is the object a hazard? (2) is the action unsafe or unhealthy? and (3) is the working condition (scenario) unsafe? At Level 3, computer vision techniques are used to predict behavior (i.e., L3.1 behavior prediction) or generate early warnings of incidents (i.e., L3.2 incident prediction). Key research questions at this level are: (1) how will the object behave, and (2) will the next incident occur?

3. Results

3.1. Journal sources

Table 3 presents the distribution of the 153 papers in the 10 journals. Automation in Construction has 69 papers, accounting for almost half (45%) of all selected papers. Journal of Computing in Civil Engineering and Advanced Engineering Informatics have 27 and 23 papers, respectively, followed by the Journal of Construction Engineering and Management and Accident Analysis and Prevention. Only 11% of the selected papers were published in the other 5 journals.

3.2. Year profile of publications

The earliest relevant paper was published in the Journal of Computing in Civil Engineering in 1997. The paper presented a vision-based interactive control technology to support the operation of bridge paint removal (Moon and Bernold, 1997). As shown in Fig. 1, the period between 1997 and 2007 had seen very few (only 10) studies of computer vision for construction. Starting from 2008, the topic of computer vision had received increasing attention since 2008 in the construction industry and safety science and management. Despite the fluctuations between 2008 and 2019, there has been an increasing trend during the period.

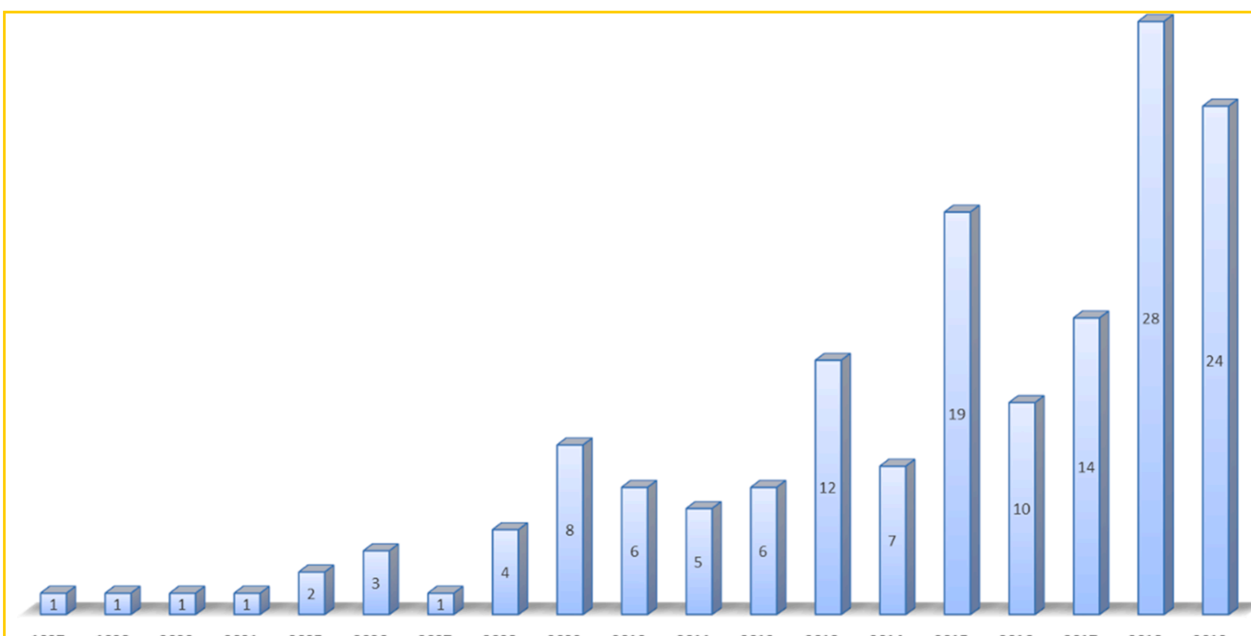

Fig. 1. Annual distribution of publications from 1997 to 2019.

3.3. Publications distributed by country/region

The selected publications were then coded based on all the authors' affiliation. As shown in Fig. 2, authors from the USA wrote about 41% (62 papers) of all papers selected, which are followed by authors from Canada (24%), South Korea (18%), China (14%), the UK (11%), Hong Kong (10%), Australia (8%), and Germany (5%). Authors from other countries and regions were responsible for 14% of the selected papers.

3.4. L1: Detection, recognition, and tracking

Construction sites involve various project-related objects (e.g., workers, equipment, tools, and resources) and a wide range of activities (e.g., earthmoving, lifting and hoisting). Safety risks and hazards are closely related to site objects' physical characteristics, location, moving path, activities, and spatial and temporal relationships between them. Thus, recognizing and tracking site objects of interest and recognizing activities are essential for computer systems to understand the complex scenes of construction and perform a risk assessment and hazard management.

3.4.1. L1.1 object detection and recognition

In general, object recognition aims to recognize the objects of interest on site from images and video frames. Object detection and recognition are concerned with the questions: (1) is something there, and (2) what is the object? Based on their recognition cues, the methods can be generally classified into three main categories: (1) geometry-based, (2) appearance-based, and (3) feature-based (Tajeen and Zhu, 2014). Early object recognition studies in the construction industry tended to apply geometry-based and appearance-based methods. For example, Chi and Caldas (2011a) selected four main geometric and appearance features (i.e., aspect ratio, height-normalized area size, percentage of occupancy of the bounding box, and average gray-scaled color) to classify mobile heavy equipment and workers.

The period between 2011 and 2016 has seen more applications of visual feature detectors and descriptors to object detection in the construction industry (Azar, 2015; Azar and McCabe, 2012; Dimitrov and Golparvar-Fard, 2014; Memarzadeh et al., 2013; Park et al., 2015; Rezazadeh Azar and McCabe, 2011; Soltani et al., 2016). They include the Scale-Invariant Feature Transform (SIFT) (Lowe, 1999), Histogram of Oriented Gradients (HOG) (Dalal and Triggs, 2005), Haar-like features (Lienhart and Maydt, 2002), and the Speeded Up Robust Features (SURF) (Bay et al., 2008). These methods capture a set of local visual features to represent an object. The significant advantage of these methods is that they are robust in partial occlusion due to the fact that they capture scale-illumination- and affine transformation- invariant features (Carr et al., 2012; Tajeen and Zhu, 2014).

Since 2017, deep learning techniques have become primary object detection and recognition methods. For example, Kim et al. (2017) applied the region-based fully convolutional network (R-FCN) (Dai et al., 2016) as a classifier to recognize heavy equipment on site (e.g., dump truck, excavator, loader, concrete mixer truck, and road roller). The method has achieved a high level of precision and recall rate. In addition, another deep learning technique, Faster R-CNN (region-based convolutional neural networks), has also gained popularity. For example, it has been utilized to detect non-hardhat-users (Fang et al., 2018b), worker (Fang et al., 2018d; Son et al., 2019), nails and screws (Wang et al., 2019). The major advantage of Faster R-CNN, compared to R-CNN, is that it is much faster and enables real-time object detection (Ren et al., 2015). It is also powerful to deal with occlusion (Fang et al., 2018b). Tajeen and Zhu (2014) developed and evaluated a dataset of five classes of construction equipment (excavator, loader, dozer, roller, and backhoe) by using two well-known object recognition methods developed by Torralba et al. (2004) and Felzenszwalb et al. (2009). Results suggested that the methods demonstrated strengths in different aspects (i.e., correctness, robustness, and speed). A summary of object detection and recognition studies in the construction industry is presented in Table 4.

Research has been conducted to recognize special "objects", such as workspaces and trades. For example, Luo et al. (2019) identified four types of workspaces (i.e., working areas, paths, laydown areas, and rest areas) by integrating object detection, multiple object tracking, action recognition, and reasoning. Due to the temporal and spatial nature of site hazards, vision-based workplace identification can help safety planning. Fang et al. (2018c) proposed a novel framework to recognize trades by analyzing the dynamic spatiotemporal relevance between workers and non-worker objects.

3.4.2. L1.2 object tracking

Object tracking is another important research topic in computer vision. Monitoring workers and equipment is of great importance for site safety. The goal of object tracking is to locate a moving object of interest over time. In general, it consists of detecting the object, creating a unique ID for the object, and tracking the object as it moves around frames in a video. In a tracking task, objects of interest can be represented by shapes (e.g., skeleton, points, geometric shapes, and contour) and appearances (e.g., color, edges, and texture).

2D object tracking methods can be classified into three categories: (1) Silhouette (or Contour) tracking, (2) kernel tracking, and (3) Point tracking (Park et al., 2011; Yilmaz et al., 2006). Silhouette tracking

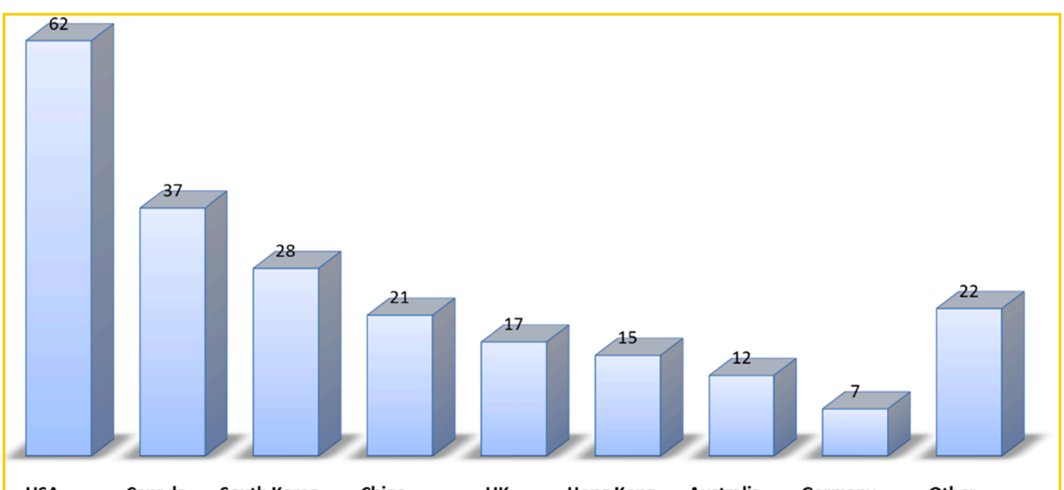

Fig. 2. Geographical distribution of publications.

Table 4
Object detection and recognition studies in the construction industry.

Paper	Object	Detectors/classifiers	Performance (correctness)
Chi and Caldas (2011a)	Mobile heavy equipment; workers	Two classifiers: Normal Bayes classifier and Neural network	Accuracy: 96%
Rezazadeh Azar and McCabe (2011)	Dump trucks	Haar–histogram of oriented gradients (HOG) and Blob–HOG	Detection: 91%
Park and Brilakis (2012)	Worker	HOG, HSV color histogram, support vector machine (SVM), and k-NN classifier	Precision: 99.0% Recall: 81.4%
Azar and McCabe (2012)	Hydraulic excavators	HOG, latent support vector machine (SVM) and, spatial–temporal reasoning	Accuracy: 95.2%
Memarzadeh et al. (2013)	Standing workers; excavators; dump trucks	Histograms of Oriented Gradients and Colors (HOG + C), and binary SVM	Accuracy: 98.83% (standing workers), 82.10% (excavators), and 84.88% (dump trucks)
Dimitrov and Golparvar-Fard (2014)	Construction materials	Hue-Saturation-Value (HSV) color values, kernel Support Vector Machine	Accuracy: 97.1%
Azar (2015)	Dump trucks; excavators	HOG, Linear SVM, and AprilTag	Precision: 100%, Recall: 64.6% (excavators) and 77.1% (dump trucks)
Park et al. (2015)	Human body; hardhat	HOG and SVM	Precision: 99.6% Recall: 96.8%
Kim et al. (2016)	Construction site objects (e.g., loader, dust, crane, etc.)	Data-driven scene parsing method	Recognition: 81.48%
Soltani et al. (2016)	Excavators	HOG, 3D model with 16 backgrounds	Recall: 98%; Accuracy: 75%
Kim et al. (2017)	Dump truck, excavator, loader, concrete mixer truck, and road roller	Region-based fully convolutional network (R-FCN)	Average Precision: 96.33%, Average Recall: 91.94%
Hamledari et al. (2017)	Drywall, insulation, stud, electrical outlet	SVM	Precision: from 80.21 to 92%, Recall: from 80.50% to 93.43%
Kim and Kim (2018)	Concrete mixer truck	HOG and SVM	Precision: 77.27%, Recall: 75.56
Fang et al. (2018b)	Non-hardhat-users	Faster R-CNN (region-based convolutional neural networks)	Precision: 95.7%, Recall: 94.9%
Fang et al. (2018d)	Worker, safety harness	Faster R-CNN and a deep CNN	Precision: 99%, Recall: 95%
Kolar et al. (2018)	Safety guardrail	Convolutional neural network (CNN)	Accuracy: 96.5%
Fang et al. (2018e)	Excavator, worker	Improved Faster Regions with Convolutional Neural Network Features (IFaster R-CNN)	Accuracy: 91% (worker), 95% (excavator)
Mneymneh et al. (2018)	Worker, hardhat	HOG and SVM	Human: precision: 98.82%, recall of 86.41%; Hardhat: precision: 94.65%, recall of 93.48%
Son et al. (2019)	Worker	Faster R-CNN	Precision: 96.03%, Recall: 98.13% Accuracy: 94.3%
Wang et al. (2019)	Nails and screws	Faster R-CNN	Average Precision: 89.1%

methods represent the object by edges, contours or silhouettes. The goal of Silhouette tracking is to find the object region in each frame based on the representation. Kernel-based trackers compute the motion of the kernel of the object (i.e., shape and appearance) in each frame. Point-based methods detect objects that are represented by points in consecutive frames (Yilmaz et al., 2006).

Table 5 summarizes the studies of object tracking in the construction industry. It is evident that construction workers are the main object of interest. Different tracking methods (e.g., Kernel tracking, Point tracking, and Silhouette tracking) were applied to track workers (Park and Brilakis, 2016; Teizer and Vela, 2009; Yang et al., 2010; Zhu et al., 2016b). Despite these advances, there are a number of significant challenges when tracking construction site resources, including (1) scale variations, (2) occlusions, (3) appearance similarity, (4) abrupt movement, (5) background clutter (Konstantinou et al., 2019; Teizer, 2015). Park et al. (2011) conducted experiments to compare these 2D vision trackers' effectiveness in tracking construction resources. Results indicated that the kernel-based method was stable and insensitive to illumination conditions, illumination variation, and scale variations and that the point-based method is effective to deal with occlusions. They suggested that overall the kernel-based methods were the most appropriate for tracking construction site resources. In order to address these challenges, Konstantinou et al. (2019) proposed a novel 2D tracking method that consists of three models, including an adaptive model, a prediction model, and an appearance model, to track multiple workers in complex environments (e.g., occlusions, illumination variations, congested environment, and abrupt changes of workers' motion).

2D tracking results are not always adequate for comprehensive construction-related assessment and analysis (e.g., productivity or safety assessment) and hence the attempts in acquiring a 3D position, direction, and speed of construction objects are important. Brilakis et al. (2011) proposed a vision-based tracking method that provides 3D positions of wheel loaders and trucks by correlating two camera views. Konstantinou and Brilakis (2018) adopted a motion-based matching method in conjunction with geometric restrictions. In addition, Lee and Park (2019) developed a framework that can track multiple workers' 3D positions based on stereo vision.

3.4.3. L1.3 action recognition

Human action recognition is an active research topic in the computer vision community (Vrigkas et al., 2015). It aims to recognize the actions of an agent of interest. The goal of action recognition is to correctly classify input data (e.g., video sequences or still images) into the underlying action category. Research has suggested that most of the construction accidents are caused by human errors and human-equipment interactions. Thus, an effective vision-based safety monitoring system requires to recognize not only human actions but also equipment operations. Edwards et al. (2016) classified human behavior into five categories with different levels of abstraction and complexity: (1) pose, (2) gesture, (3) action, (4) interaction (human-to-human and human-to-object), and (5) activity.

Human action recognition methods can be grouped into two main categories: (1) unimodal and (2) multimodal methods (Vrigkas et al., 2015). Unimodal methods use the data of a single modality to represent actions and they can be further categorized into four methods (1) space–time, (2) stochastic, (3) rule-based, and (4) shape-based methods. In contrast, multimodal methods integrate features obtained from

Table 5
Object tracking studies in the construction industry.

Paper	Object tracked	Description	Tracking method applied
Teizer and Vela (2009)	Workers	Compared four tracking techniques: Density Mean-shift, Bayesian Segmentation, Active Contours, and Graph-cuts, The Bayesian method with temporal averaging performed the best overall.	Kernel tracking
Chi et al. (2009)	Construction resources	Probabilistic Hausdorff image matching	Silhouette tracking
Yang et al. (2010)	Multiple workers	Kernel covariance tracking	Kernel tracking
Brilakis et al. (2011)	Project-related entities (e.g. wheel loaders and trucks)	Compared the three categories of 2D vision-based tracking methods – contour-based, kernel-based, and point-based methods.	Kernel tracking, Silhouette tracking, and Point tracking
Park et al. (2011)	Construction site resources	Compared the three categories of 2D vision-based tracking methods – contour-based, kernel-based, and point-based methods.	Kernel tracking, Silhouette tracking, and Point tracking
Teizer (2015)	Temporary resources on infrastructure construction sites	Reviewed the status quo and open challenges in vision-based tracking of temporary resources on infrastructure construction sites	/
Park and Brilakis (2016)	Workers	Eigen-images and particle filtering (Ross et al., 2008)	Point tracking and Kernel tracking
Yuan et al. (2016)	Excavator	Optical flow estimation and 3D triangulation	
Zhu et al. (2016b)	Workers, roller, truck, and dozer	Particle filtering	Point tracking
Kim and Chi (2017)	Construction equipment	A median-flow algorithm (Kalal et al., 2011) and a pyramidal Lucas-Kanade algorithm (Bouguet, 2001)	Point tracking
Lee and Park (2019)	Workers	3D tracking based on stereo vision	3D tracking
Konstantinou et al. (2019)	Multiple workers	Track multiple workers in complex environments (e.g., occlusions, illumination variations, congested environment, and abrupt changes of workers' motion)	/

different sources (e.g., visual and audio data) and they can be classified into three categories: (1) affective, (2) behavioral, and (3) social networking methods. It is beyond the scope of this paper to introduce each method. Details can be found in work by Vrigkas et al. (2015).

Table 6 lists previous human and equipment action recognition studies in the construction industry sorted by year. Construction involves a wide range of heavy equipment (e.g., tower crane, dump trucks and excavators, etc.). Thus, recognizing equipment operations has also attracted much attention from researchers.

Since 2007, attention has been paid on recognizing basic equipment operations, such as hydraulic excavator actions (Zou and Kim, 2007) and excavator and truck actions (Golparvar-Fard et al., 2013). The information of basic actions is of less usefulness to productivity analysis and schedule assessment. Thus, in recent years, researchers have made efforts to recognize actions with a higher level of abstraction and

Table 6
Action recognition studies in the construction industry.

Study	Year	Action type	Level of action
Zou and Kim (2007)	2007	Hydraulic excavator actions	Action
Gong and Caldas (2009)	2010	Concrete column pour	Activity
Gong et al. (2011)	2011	Both worker actions (traveling, transporting, bending down, aligning, and nailing) and equipment actions (relocating, excavating, and swing)	Action
Bai et al. (2011)	2012	Worker back bending	Pose
Ray and Teizer (2012)	2012	Worker standing, bending, sitting, crawling	Pose
Han et al. (2013)	2013	Reaching too far in ladder climbing	Pose
Golparvar-Fard et al. (2013)	2013	Excavator and truck actions (digging, hauling, dumping, swinging, etc.)	Action
Azar et al. (2012)	2013	Excavators and dump trucks actions during dirt loading	Object-to-object interaction
Ranaweera et al. (2012)	2013	Liner-lowering activity in tunnel construction	activity
Yang et al. (2012)	2014	Tower crane activity (loading, lifting, and unloading materials)	Activity
Khosrowpour et al. (2014)	2014	picking up, holding, walking, putting down, measuring and cutting, breaking the gypsum board, and idling	Action
Han et al. (2014)	2014	Safe and unsafe actions	Action
Yang et al. (2016)	2016	lay brick, transporting, cut plate, drilling, tie rebar	Action
Bügler et al. (2017)	2017	Earthmoving operations	Object-to-object interaction
Yu et al. (2017)	2017	Ladder climbing, leaning on handrails, dumping waste from height	Pose
Yan et al. (2017)	2017	Ergonomic postures	Pose
Azar (2015)	2017	Earthwork operation (e.g., backfill)	Object-to-object interaction
Luo et al. (2018b)	2018	17 types of construction activities and actions (e.g., placing concrete and fixing rebar)	Activity and action
Zhang et al. (2018)	2018	Both arms below, etc.	Pose
Soltani et al. (2018)	2018	Excavator pose	Pose
Kim et al. (2018a)	2018	Mixed activities of construction equipment	Activity
Kim et al. (2018b)	2018	Earthmoving operations	Object-to-object interaction
Ding et al. (2018)	2018	Normal ladder and abnormal ladder climbing	Action
Luo et al. (2018a)	2018	Steel bending, transporting, walking	Action
Luo et al. (2018c)	2018	16 classes of activities of rebar and formwork	Activity and action
Fang et al. (2018a)	2018	Aerial operation scenario	Human-object interactions
Xu and Yoon (2019)	2019	Excavator manipulator pose	Pose
Kim and Chi (2019)	2019	excavator earthmoving	Activity
Liang et al. (2019)	2019	Excavator manipulator pose	Pose

complexity. For example, earthmoving activities involve the interactions between excavators and dump trucks. Efforts were made to recognize the interactions (Bügler et al., 2017; Kim and Chi, 2019; Rezazadeh Azar, 2017; Rezazadeh Azar et al., 2012). Event/Activity classification usually requires more information/features. For example, liner-lowering events were recognized based on the tunnel concrete liner recognition and descending trajectories of the liner (Ranaweera et al., 2012).

In addition, Kim et al. (2018b) developed a vision-based activity recognition framework that can recognize the interaction between excavators and dump trucks in earthmoving operations. They classified equipment behavior into low-level individual actions (e.g., load soil, travel to dumping area, and dump soil) and high-level activities (idle, travel, and work). High-level activities were recognized by interpreting the interactions based on a knowledge-based system that evaluates individual actions and proximity between excavators and dump trucks. Interaction analysis was conducted based on co-existence and proximity of equipment and its action consistency.

Early efforts of vision-based human action recognition were mainly focused on the pose and basic action (e.g., traveling, bending down, standing, sitting, and crawling) (Bai et al., 2011; Gong et al., 2011; Ray and Teizer, 2012). Inter-class similarity and intra-class variance are major challenges faced with vision-based human action recognition researchers (Gong et al., 2011). In construction, the inter-class similarity of different worker actions may be big, while the intra-class variance of the same action might also be significant.

Action recognition can also be broken down into two processes: action detection and action recognition. For example, Yu et al. (2017) detected unsafe behaviors (i.e., is there an unsafe behavior?) by determining the value ranges of key joint parameters based on the theory of human skeleton model, while recognized unsafe behaviors (i.e., which type of unsafe behavior) by selecting and measuring the parameters that can distinguish one behavior with the others. Luo et al. (2018b) proposed a two-step method to recognize diverse construction activities in still site images. It first detects construction-related objects using Faster R-CNN. The information was further processed based on semantic relevance rules to classify construction activities.

It is worth noting that the main purpose of these studies is to facilitate productivity analysis and optimization, rather than safety assessment and monitoring. Despite this, equipment operation recognition forms an essential part of vision-based safety assessment and monitoring system, as many site hazards are equipment-related.

3.5. L2: Assessment

High-level vision-based assessment is usually based on object detection and recognition, object tracking, and action recognition. For example, Fang et al. (2018c) demonstrated that detecting non-certified work on site was based on a combination of object detection, object tracking, face recognition, and trade recognition techniques. Assessment can be data-driven (e.g., using machine learning techniques), knowledge-driven (e.g., using rules), and model-driven (e.g., actual against planned schedule). It can be focused on objects, behaviors of objects, and conditions of the construction site.

Table 7 presents vision-based assessment studies in the construction industry. It shows that past efforts were focused on vision-based productivity and safety assessment. In general, productivity analysis was performed based on object recognition, tracking, and/or action recognition. For example, Gong and Caldas (2011) developed a prototype system which uses objects' (e.g., column, slab, scaffold, bobcat loader, etc.) spatial positions and moving trajectories as inputs to determine schedule progress by comparing the inputs against defined spatial regions. Bai et al. (2011) applied artificial neural networks (ANN) to determine if the working status is effective or not. In addition, Turkan et al. (2012) combined computer vision and building information modeling (BIM) technologies to assess productivity by measuring earned value (EV).

Table 7
Vision-based productivity and safety assessment studies in the construction industry.

Study	Assessment type	Assessment method
Ibrahim et al. (2009)	Productivity assessment	Model-driven
Gong and Caldas (2011)	Productivity assessment	Knowledge-driven
Chi and Caldas (2011b)	Safety assessment	Knowledge-driven
Bai et al. (2011)	Productivity assessment	Data-driven
Turkan et al. (2012)	Productivity assessment	Model-driven
Seo et al. (2014)	Ergonomic assessment	Data-driven
Han et al. (2014)	Safety assessment	Data-driven
Liu et al. (2015)	Productivity assessment	Knowledge-driven
Kim et al. (2015)	Safety assessment	Knowledge-driven
Bügler et al. (2017)	Productivity assessment	Data-driven
Kropp et al. (2018)	Productivity assessment	Model-driven
Chen et al. (2019)	Safety assessment	Model-driven
Kim et al. (2019a)	Safety assessment	Model-driven
Kim et al. (2019b)	Productivity assessment	Data-driven

From a health and safety perspective, object assessment refers to assessing if the object of interest is a hazard or not, or if the object is in an unsafe and unhealthy state. Vision-based site condition assessment is a higher level of application than object recognition and tracking and action recognition. The basic information extracted from object recognition and tracking and action recognition are often further assessed based on rule-based or ontology-based knowledge models. For example, Chi and Caldas (2011b) demonstrated how the data acquired from object recognition and tracking could be utilized for automatic safety assessment. They designed safety rules to detect three types of safety violation (i.e., speed limit, dangerous access, and close proximity). Similarly, Kim et al. (2015) developed an on-site safety assessment system which incorporates two modules: vision processing module (VPM) and safety assessment module (SAM). VPM collects the spatial information about workers and equipment using computer vision, while SAM assesses the safety levels using IF-THEN rules and fuzzy inference. Ray and Teizer (2012) applied to a linear discriminant analysis method to classify worker posture and then used predefined rules to determine if the posture is ergonomic or not. Chen et al. (2019) evaluated workers' safety risk based on the fusion of position and posture information. However, the way that the authors measured the risk level without linking it to a specific hazard and the construction context is confusing and arbitrary. Kim et al. (2019a) used a camera-mounted unmanned aerial vehicle to evaluate the safety risk of struck-by hazard based on proximity measurement and monitoring. The system will visualize the hazard when a worker is close to mobile equipment. However, vision-based safety assessment system which is only based on proximity tends to generate false alarms. For example, it is safe when a worker is close to an excavator that is not operating. It would be more reliable if reasoning and inference are based on a combination of proximity measurement and action recognition.

3.6. L3: Prediction

From a proactive safety management perspective, being able to predict a dangerous situation before it occurs is far more important than recognizing it thereafter. This task is referred to as vision-based prediction where computer vision techniques anticipate "when will the object(s) do what". Most of the existing action recognition methods, as reviewed in Section 3.4.3, were developed and applied to recognize complete actions. They are unable to recognize unfinished action videos and therefore cannot be used for action prediction.

The literature review reveals that applications of computer vision to predict the motion of workers and mobile equipment in the construction industry have been rather limited. As an early attempt, Zhu et al. (2016a) used Kalman filters for predicting movements of workers and mobile equipment on the construction site in order to prevent, potential struck-by accidents. This method involves two main steps: (1) estimate 3D positions of workers and mobile equipment through visual detection, tracking, and triangulation and (2) feed the estimated positions into a Kalman filter to predict the positions of workers and equipment. The performance of the Kalman filter can be further enhanced by learning from past prediction.

Despite the scarcity of relevant studies in the construction domain, action prediction is gaining increased attention in the computer vision community. For example, Ryoo (2011) presented a human activity prediction method, a combination of integral bag-of-words and dynamic bag-of-words, for activity prediction. Lan et al. (2014) proposed a new representation, hierarchical movements, to describe human movements at multiple levels of granularities and developed a max-margin learning framework for action prediction. Koppula and Saxena (2016) used an anticipatory temporal conditional random field (ATCRF) to anticipate human activities. The study obtained an activity anticipation accuracy of 84.1%, 74.4%, and 62.2% for an anticipation time of 1, 3, and 10 s respectively. It is clear that the method is less powerful to predict human activities for a long time horizon. To address this problem, Farha et al. (2018) proposed two novel approaches to improve prediction duration. The first approach is based on a recurrent neural network (RNN), while the second approach builds on a convolutional neural network (CNN). The work can predict video content of up to several minutes' length. It is beyond the scope of this paper to review human activity prediction studies in the computer vision community. Readers are referred to a comprehensive survey by Trong et al. (2017).

4. Discussion

In a comprehensive survey, Jin et al. (2019) identified five main research topics within the theme of construction safety, including (1) safety climate and safety culture, (2) information and communication technology (ICT) in safety management, (3) workers' safety perception and behavior, (4) safety management system, and (5) hazard identification, accident causation, and risk management in safety. Considering the fact that computer vision represents an ICT and that hazard identification and risk management can be integrated into safety management systems, this section discusses three main aspects to demonstrate how computer vision technologies may be applied in improving construction safety management. These are (1) vision-based safety management system, (2) vision-based behavior-based safety (BBS) program, and (3) vision-based safety culture sensing system.

4.1. Vision-based safety management system

A safety management system (SMS) integrates a combination of activities and functions to identify hazards and manage risks in the workplace (Guo and Yiu, 2016). Common safety management activities in the construction industry include, but are not limited to, health and safety policy, safety planning, hazard management, workplace inspections, incident reporting and investigation, and training and supervision. Safety (or accidents) can be understood as an outcome of the quality of the implementation and monitoring of these integrated activities and processes (Le Coze, 2013). In general, a safety management system approach focuses on three main aspects: physical workplace, people, and organization issues (Makin and Winder, 2008; Zou and Sunindijo, 2015). The links between computer vision and SMS can be discussed based on these three aspects.

4.1.1. Safe workplace

One purpose of SMS is to create a safe workplace for workers by identifying hazards, evaluating and controlling risks. Current hazard identification practices in the construction industry are largely paper-based, manual, and inefficient (Zhang et al., 2015a). Research has suggested that a large proportion of safety hazards remain unrecognized and unmanaged in complex and dynamic construction environments (Carter and Smith, 2006; Jeelani et al., 2016). Past studies have demonstrated that computer vision can be applied to support hazard identification, such as Kim et al. (2019a) and Kim et al. (2015). To be useful in practice, a vision-based hazard identification system must have a satisfactory level of comprehensiveness and correctness. First, it is crucial for the system to identify most, if not all, hazards. To achieve this, a comprehensive hazard profile and knowledge base need to be developed. Human experts recognize hazards by utilizing their knowledge and experience. Such a knowledge base needs to be generalized and formalized so that the computer system can understand and apply it to different scenarios and situations. Note that hazard identification is more than common sense, because hazards can be created and emerged by spatial and temporal dynamics between construction site objects (Sacks et al., 2009). Therefore, it is inadequate to define and model a hazard only by its source. As construction is dynamic in nature, the mechanism for a hazard to take place needs to be modeled and formalized. An important assumption can be made that the nature of the hazard is closely associated with spatial and temporal relationships between building elements, materials, temporary equipment and tools, operations, and human workspace. By developing such a spatial and temporal model, a hazard can be re-defined using geometric, spatial, and temporal features, and specific patterns could be identified to distinguish the hazard from others.

Once a formal hazard model is integrated into a vision-based system, it is essential that the system understands the 'meaning" of the whole site scene. Being able to recognize all objects and actions does not necessarily enable the system to achieve a higher level of scene understanding. From a health and safety perspective, scene understanding involves the following main hierarchical functions:

(1) Understand what the objects are (including humans);
(2) Understand their roles and relationships;
(3) Understand their actions;
(4) Understand the interactions between their actions;
(5) Understand the activities, scenarios, or work packages they participate in;
(6) Understand the interactions between different activities, scenarios, or work packages.

This higher level of scene understanding can generate richer information that can drive more powerful reasoning based on the formal hazard model, which improves the system's capability to identify unexpected hazards and unknown potential hazard sets. Once hazards are identified, safety risk evaluation can be performed based on rules and/or knowledge models.

4.1.2. Safe people

The key concept that links computer vision with the strategy is "*situational awareness (SA)*". SA also refers to "*the ability of maintaining an appropriate picture of situations to perform safe operations*" (Le Coze, 2016). Construction is highly dynamic in nature, and it is beyond even experienced workers' ability to identify all hazards present on site. As a result, there are often discrepancies between the human operators' (i.e., workers and managers) understanding of system status and actual system status. Endsley developed a model of situational awareness (Endsley, 1995), in which SA is defined at three ascending levels, perception, comprehension, and projection:

- Level 1: Perception, the perception of the elements in the environment,
- Level 2: Comprehension, comprehension of the current situation,
- Level 3: Projection, prediction of future status.

As illustrated in Fig. 3, there exists a correspondence between the three-level of SA and the three levels of computer vision development proposed earlier in Section 2.3.

On the left-hand side, developing and maintaining situation awareness requires site personnel to constantly perceive the dynamic site environment, understand the safety impacts, and predict the future hazard profile. Note that Level 1 SA requires high attentional demands and a mental model that directs attention; Levels 2 and 3 SA are developed based on valid safety knowledge that can interpret the information obtained at Level 1. Given the linkage between goals and mental models (Endsley, 2015), it is likely that workers ignore some safety-related information when they are under production pressure. In addition, it is challenging for workers, especially inexperienced novices, to obtain a quick and reliable understanding of safety impacts on site. It is even difficult for experienced workers to understand and predict the interactions between different objects and activities. As a result, it is unlikely the workers are able to develop and maintain a high level of SA on their own.

There is a significant potential for a vision-based system to reduce the cognitive burden and improve workers' SA at all three levels. Computer vision, combined with other information technologies (e.g., wearable technologies and wireless sensors), is capable of switching workers' information processing fashion. Without information technologies, workers' information processing is largely goal-driven and top-down, that is, a worker's goals direct his/her attention to the site environment. Due to the goal conflicts (e.g., production vs safety), goal-driven and top-down information processing and decision making can deactivate safety-oriented goals and mental models. By providing timely cues, hazard information, assessment, and prediction, vision-based systems are able to switch the fashion to bottom-up and data-driven information processing. This can considerably help workers develop and maintain SA.

4.1.3. Safe organization

Vision-based safety management systems can help build a learning organization. Efforts can be made to create video and image databases of unsafe behavior and incidents. Computer vision techniques can be applied to annotate accident-related videos and images. Information retrieval applications can be developed to facilitate the retrieval of relevant information for safety training, education, and planning. Data-driven accident classification and analysis are not new to safety research (Goh and Ubeynarayana, 2017; Tixier et al., 2016). Nevertheless, vision-based accident classification and analysis is still in its infancy. In the long term, once the database reaches a suitable size, machine learning algorithms could be applied for useful and legitimate accident classification and prediction. This can significantly help construction companies become true learning organizations if accident-related sensitivity and bureaucracy can be well managed.

4.2. Vision-based behavior-based safety (BBS) program

In the evolution of safety theories over the past decades, understanding and managing safety behavior has been an important and popular research topic. Promoting safe behavior is a crucial factor in safety (Guo et al., 2016a). For instance, BBS has received significant attention from researchers since the 1970 s. In general, BBS involves a circle of goal setting, manual observation, feedback, and training. Different theories (e.g., reinforcement theory and goal-setting theory) may be integrated into different BBS programs (Guo et al., 2018). The main limitation of BBS programs is that it relies on manual observation and analysis, which is subjective, laborious, and inefficient. In addition, the scope of human observation is limited and therefore a big picture is often lost. This has significantly impeded the adoption of BBS in the construction industry.

Computer vision has considerable potential to improve efficiency in observation, and thus facilitate advanced behavioral analysis. As reviewed in this paper, previous works have proved that computer vision can accurately recognize unhealthy pose and simple actions, such as non-PPE (personal protection equipment) users and unsafe ladder climbing. The major limitation is that they focused on recognizing simple repetitive actions (traveling, standing, and bending). In traditional BBS programs, it is of less interest to record simple gestures and atomic actions like not wearing a hardhat. Many interesting human activities on site are characterized by a complex spatial and temporal composition of objects and actions. For example, one of the unsafe behaviors in a checklist developed for a BBS program is: "The signalman gives a warning signal when the load is lifted or moved and ensures no one is standing under the suspended load" (Guo et al., 2018). To recognize an unsafe behavior that violates this rule, computer vision must not only recognize involved objects (i.e., workers, crane, load) in terms of identity, location, movement direction, but more importantly, understand the interactions between these objects.

It is clear that the benefits are limited when only classifying actions based on predefined categories (labeled as 'safe" or "unsafe") using machine learning. This is because there are too many varieties of

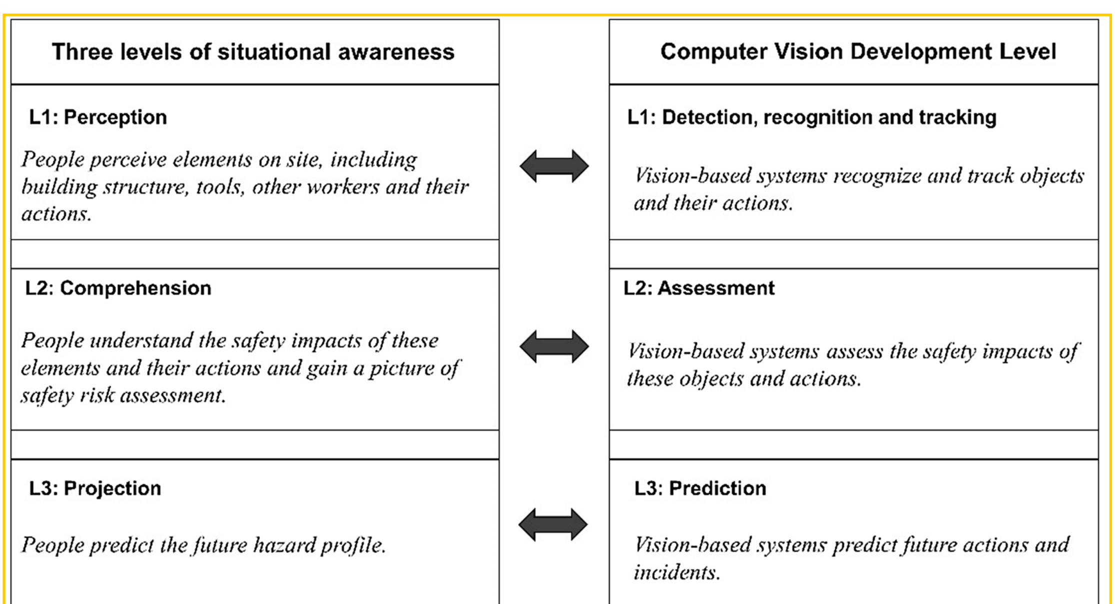

Fig. 3. The correspondence between SA and CV.

conditions characterizing real construction site scenes, and it is difficult and inefficient to identify invariants that characterize a certain action and its dynamics. For example, as indicated by Gong et al. (2011), different action categories can have similar gestures and one action category can have a variety of gestures. In some case, whether an action is safe or not depends on the status of other objects. Thus, similar to vision-based hazard identification, assessing complex activities also requires a knowledge model that can determine if an action is safe or not and a framework that classifies safety behavior by the level of abstraction and complexity. The authors of this paper propose a six-level hierarchical framework of safety behavior based on Edwards et al.'s five-level classification system (Edwards et al., 2016), as shown in Table 8. By using the hierarchy, unsafe and unhealthy behavior (as well as hazards) can be defined at different levels. Theoretically, observing actions at a high level of abstraction, like safety compliance and safety participation, could enable an analysis of the correlation with safety culture and safety outcomes.

4.3. Vision-based safety culture sensing system

Safety culture has been an active research topic in safety science and management. The concept of safety culture claims that safety culture drives the safety process and the success of a safety management system is determined by behavior-based and human factors (Geller, 1994). An important research question can be asked: can computer vision be applied to measure safety culture?

Given that safety culture is a multidimensional concept, which dimensions of the concept can be measured? Cooper (2000a,b) proposed a reciprocal safety culture model that contains three interrelated elements: (1) observable safety behaviors, (2) subjective internal psychological features, and (3) objective situational features. Cooper suggested that internal psychological features can be assessed by safety climate questionnaires; safety behaviors can be assessed by observational checklists, and the situational features can be assessed through safety management system audits. Based on the reciprocal safety culture model, internal psychological features could be indirectly measured by measuring items in the safety climate questionnaires. In the construction industry, a number of safety climate measures were developed over the past three decades (Dedobbeleer and Béland, 1991; Fang et al., 2006; Guo et al., 2016a; Lingard et al., 2009; Mohamed, 2002). Safety climate can be seen as a subset of safety culture (Zou and Sunindijo, 2015). It has been proved to be a useful leading indicator of unsafe behavior and accident (Guo et al., 2016a; Guo et al., 2016b; Zohar, 2010).

Table 9 lists safety climate factors and corresponding measurement items that could potentially be measured by computer vision techniques. Safety climate factors, like workers involvement, communication and support, PPE, supportive environment, supervisor's role and workmate's

Table 8
A hierarchical framework of safety behavior.

Safety behavior level		Definition	Examples
L1	Pose and gesture	An atomic observation of the spatial arrangement of a human body at a single temporal instance A temporal series of poses or action primitives on a sub-action scale	• Unhealthy and unsafe ladder climbing posture
L2	Action	A series of gestures which form a contextual event	• A worker uses the welding face shield during hot work.
L3	Human-to-human interaction	A pairwise action committed by two people (e.g., workers)	• Safety coaching between a site supervisor and a worker; • Safety communication between two workers.
	Human-to-object interaction	A pairwise action committed by one individual upon an object (e.g., equipment)	• The operator reverses the vehicle with guidance from the traffic controller.
	Object-to-object interaction	Pairwise actions committed by two objects	• The interactions between excavators and dump trucks during earthmoving operations
L4	Activity	A collection of action and/or interactions that compound to describe a high-level event	• Roofing activity • Formwork activity • Scaffolding activity
L5	Activity-to-activity interaction	A pairwise activity committed by one individual (or group) upon another activity	• The interaction between crane lift vs. roofing
L6	Safety compliance	Following rules in core safety activities	• Adhering to safety procedures and carrying out work in a safe manner
	Safety participation	Promoting the safety program within the workplace, demonstrating initiative, and putting effort into improving safety in the workplace	• Safety coaching

Table 9
Safety climate measurement items that could be measured by computer vision.

Safety climate factors	Measurement items	Sources
Workers involvement	• Foreman regularly and frequently makes us aware of dangerous work practices and conditions and praises us for safe conduct. • Are there regular job safety meetings at your present job site?	(Dedobbeleer and Béland, 1991)
Communication and support	• Workers are encouraged to support and look out for each other.	(Glendon and Litherland, 2001)
Personal protective equipment	• PPE use if monitored to identify problem areas.	
Supportive environment	• Often remind each other on how to work safely. • Always offer help when needed to perform the job safely. • Endeavor to ensure that individuals are not working by themselves under risky or hazardous conditions.	(Mohamed, 2002)
Supervisor's and workmate's role	• People who work here often have to take risks when they are at work. • People here always work safely even when they are not being supervised. • Supervisors seldom check that people here are working safely.	(Fang et al., 2006)
Appraisal of safety procedure and work risk	• People here always wear their health and safety protective equipment when they are supposed to.	
Risk-taking behavior	• Not all the health and safety procedures/instructions/rules are strictly followed here.	
Supervisory safety leadership	• My supervisor approaches workers during work to discuss safety issues. • My immediate supervisor often talks to me about health and safety.	(Lingard et al., 2009)
Co-workers' actual safety	• People here always work safely even when they are not being supervised.	
Social support	• When my supervisor and co-workers see me working at-risk, they caution me. • Supervisor frequently moves around inspecting the workplace.	(Guo et al., 2016a)

role, appraisal of safety procedure and work risk, risk-taking behavior, supervisory safety leadership, co-workers' actual safety, and social support, could be measured by computer vision if their visual features can be defined and recognized.

Objects involved in these measurement items include (1) foremen, (2) workers, (3) PPE, and (4) supervisors. As suggested in Table 4, past studies have successfully recognized workers, trades, and PPE. Future efforts can be made to recognize different roles of site personnel, as this can help understand and recognize human actions. Human actions involved in these measurement items include (1) identify dangerous work practices, (2) praise workers, (3) organize job safety meeting, (4) workers support and look after each other, (5) risk-taking, (6) wear PPE, (7) safety rules violations, (8) supervisor discuss safety with workers, (9) work safely, (10) caution co-workers when they are working unsafely, and (11) safety inspection by supervisors. Matching these safety behaviors that should be recognized against those human actions that can be recognized (see Table 6) makes it clear that much research effort is needed before computer vision technologies can be applied to measure the safety climate factors.

Compared to these factors with tangible visual features, other safety climate factors, such as management commitment to safety, safety attitude, risk perceptions, and work pressure, are much more difficult to measure. Vision-based psychological feature recognition may benefit from Social Signal Processing (SSP) methods. SSP suggests that social signals (e.g., nonverbal behavior) are the expression of an individual's attitude towards a particular social situation (Vinciarelli et al., 2009). It is believed that nonverbal behavior conveys information not only of individuals' feelings, mental state, and personality but also, during social interactions, of the nature of the social relationship (Vinciarelli et al., 2009). For example, SSP has been applied to determine learners' interest level by sensing and interpreting behavior cues (Gatica-Perez et al., 2005; Mota and Picard, 2003). Similarly, it is possible to estimate the level of safety motivation and safety leadership on site by sensing and interpreting a useful set of behavioral cues.

5. Future research

It is important to recognize that current computer vision applications in construction health and safety management are still limited and primitive. To facilitate computer vision technology development and applications to construction health and safety, we suggest the following future research directions:

5.1. Develop and test a behavioral-cues-based safety climate measure

As a "snapshot" of safety culture, safety climate is usually measured through a questionnaire that captures workers' perceptions of safety management. Guldenmund (2007) criticized safety climate questionnaires for being a quick but also 'dirty' instrument, due to the fact that they are easy to use but include a lot of random 'noise'. A vision-based behavioral approach can be used to supplement safety climate questionnaires.

Safety climate studies in the construction industry have suggested that the same safety climate factor (e.g., supervisory support) can be measured by using different measurement items. The discussion in Section 4.3 showed that some safety climate factors (e.g., risk perception) are difficult to measure using computer vision, due to a lack of tangible visual cues. To address this problem, a safety climate survey which consists of behaviors and tangible behavioral cues should be developed and tested in terms of reliability and validity.

To this end, low-level visual features of relevant behaviors, as well as social and behavioral cues, need to be identified and linked to high-level safety climate factors, such as management commitment to safety and safety attitude. This effort would involve recognizing group behaviors (e.g., between supervisor and workers, or between workers). It is a common view of social science that group interactions are more complex than individual behavior (Lehmann-Willenbrock et al., 2017). More research efforts are needed to capture the group interactions by computer vision for construction health and safety.

The value of linking behavior cues to safety climate lies in the fact that this provides an alternative for a long-term prediction of unsafe behavior and accident. Unlike a black-box machine learning approach, this can provide a more interpretable picture of site safety for managers.

5.2. Develop safety behavior datasets

Safety behaviors are often a set of events with different levels of abstraction and complexity, from simple unhealthy pose and gestures to complex social interactions on site. Past efforts to identify unsafe behavior using computer vision have been fragmented, with little consideration for a holistic framework that targets all possible unsafe behaviors. Chaquet et al. (2013) conducted a survey of video datasets for human action and activity recognition. There is a lack of publicly available datasets that are dedicated to safety behavior recognition in the construction industry. The hierarchical framework of safety behavior (Table 8) could be validated and used to classify the level of safety behavior from simple pose to complex interactions and activities. Different datasets should be developed at different levels to allow machine learning models to develop an understanding of safety behaviors that occur on construction sites. Progress at the low-level recognition can facilitate high-level recognition. Meaningful aggregations from actions, via activity, to activity-to-activity interaction, can significantly improve the efficiency and usefulness of vision- and behavior-based safety programs.

In addition to the scope issue, behavior complexity and assessment criteria are key issues to be addressed by future studies. To perform comprehensive health and safety monitoring, a high level of scene understanding is required. Future efforts can be made to obtain a semantic meaning of the construction site scene and explore how rich semantics can improve automatic hazard identification and health and safety monitoring. Future research efforts can also be made to link computer vision with workers' behavior simulation models (e.g., agent-based safety behavior model (Zhang et al., 2019)). Data collected by computer vision can be integrated into the models as input for simulation and validation. Such a combination could be a promising research direction towards the predictive analysis of site safety.

5.3. Develop a formal hazard model

The identification of safety hazards requires more than common sense. People often show over-confidence and complacency in the hazard identification and management process. Traditional methods (e.g., job safety analysis, safe work method statement, and task analysis) define and identify hazards based on tasks and activities. These methods are most suitable when tasks are well defined. However, they may be ineffective to identify hazards that are emerged from the interactions between objects or between activities. These traditional methods can be complemented by re-defining hazards using the information obtained from computer vision assisted object recognition and tracking, action recognition, and spatiotemporal analysis. More research attention should be paid on teaching the vision-based system to understand the generation mechanism of hazards so that both short-term and long-term prediction can be made. Knowledge engineering tools such as ontology can be developed and utilized to enable semantic representation and reasoning. Once basic data (e.g., signal, perceptual features, and physical objects and actions) are mapped into appropriate ontologies, the semantic meaning of the real construction site scene would be obtained. Therefore, more powerful reasoning can be performed based on the rich semantics to recognize hazards.

5.4. Evaluate the real impacts of vision-based technologies on safety performance

Past research had focused on the development of vision-based systems for site safety. Research that evaluates the real impacts of computer vision on safety performance has been limited. The positive impacts cannot be overestimated and individuals' and organizations' perceptions of the technologies must be taken into account. From a cognitive behavior science perspective, worker behaviors are shaped by their attitudes, beliefs, norms, and values. It remains an open question as to how computer vision affects workers' mental processes and eventually their behavior on site. Another interesting issue is the relationship between information technology (e.g., computer vision) and safety culture. Previous studies have suggested that technologies can penetrate organizational culture (Hill, 1988). Although safety culture has its roots in organizational culture, it is not clear that how computer vision technologies impact safety culture in construction projects.

6. Conclusions

This paper reviewed the state-of-the-art in computer vision development and applications in the construction industry from a health and safety perspective. It categorizes computer vision studies in the construction industry based on a three-level development framework: Level 1 Detection, recognition, and tracking, Level 2 Assessment, and Level 3 Prediction. The three-level development framework was designed to classify the complexity and difficulty of existing computer vision applications. Results indicated that the majority of past efforts were focused on Level 1. A wide range of construction project objects can be recognized and tracked. Recent deep learning techniques have significantly improved recognition and tracking performance. In addition, efforts were made to recognize human and equipment behavior at different complexity levels, from recognizing simple pose to complex interactions and events. However, research that is focused on recognizing unsafe behavior has been limited.

There are even fewer studies that were aimed at developing vision-based safety assessment and prediction systems. Such a current distribution situation is understandable, considering that there are still a number of unsolved technical issues (e.g., viewpoints, occlusion, light, etc.) that pose significant challenges at Level 1. An investigation of the theoretical implications of computer vision for health and safety reveals that it has strong links to safety research traditions, including safety management system, behavior-based safety program, and safety culture.

It is clear that computer vision technologies can be applied to enhance current safety management systems by improving the efficiency of hazard identification and lifting the situation awareness of workers. Future research can be made to recognize more safety behaviors so that computer vision techniques can be integrated into behavior-based safety programs to reduce the reliance on manual observation and analysis. Developing vision-based safety culture sensing systems is an interesting research area to explore. Interdisciplinary efforts among social science, behavior science, safety science, and computer vision are needed to measure and monitor safety culture using computer vision. Such a move would significantly impact the digital transformation in construction health and safety management and improve safety performance. The suggested research directions are largely theory-oriented. However, it should be noted that the theoretical challenges involved in the health and safety domain are closely related to technical challenges in the computer vision community. This encourages more collaboration between researchers in these two domains.

References

Akinci, B., Boukamp, F., Gordon, C., Huber, D., Lyons, C., Park, K., 2006. A formalism for utilization of sensor systems and integrated project models for active construction quality control. Automation in Construction 15, 124–138.

Azar, E.R., 2015. Construction equipment identification using marker-based recognition and an active zoom camera. Journal of Computing in Civil Engineering 30, 04015033.

Azar, E.R., McCabe, B., 2012. Part based model and spatial–temporal reasoning to recognize hydraulic excavators in construction images and videos. Automation in Construction 24, 194–202.

Bai, Y., Huan, J., Kim, S., 2011. Measuring bridge construction efficiency using the wireless real-time video monitoring system. Journal of Management in Engineering 28, 120–126.

Bay, H., Ess, A., Tuytelaars, T., Van Gool, L., 2008. Speeded-up robust features (SURF). Computer Vision and Image Understanding 110, 346–359.

Bouguet, J.-Y., 2001. Pyramidal implementation of the affine lucas kanade feature tracker description of the algorithm. Intel Corporation 5, 4.

Brilakis, I., Park, M.-W., Jog, G., 2011. Automated vision tracking of project related entities. Advanced Engineering Informatics 25, 713–724.

Bügler, M., Borrmann, A., Ogunmakin, G., Vela, P.A., Teizer, J., 2017. Fusion of photogrammetry and video analysis for productivity assessment of earthwork processes. Computer Aided Civil and Infrastructure Engineering 32, 107–123.

Carr, P., Sheikh, Y., Matthews, I., 2012. Monocular object detection using 3d geometric primitives. European Conference on Computer Vision. Springer 864–878.

Carter, G., Smith, S.D., 2006. Safety hazard identification on construction projects. Journal of Construction Engineering and Management 132, 197–205.

Chaquet, J.M., Carmona, E.J., Fernández-Caballero, A., 2013. A survey of video datasets for human action and activity recognition. Computer Vision and Image Understanding 117, 633–659.

Chen, H., Luo, X., Zheng, Z., Ke, J., 2019. A proactive workers' safety risk evaluation framework based on position and posture data fusion. Automation in Construction 98, 275–288.

Chi, S., Caldas, C.H., 2011a. Automated object identification using optical video cameras on construction sites. Computer-Aided Civil and Infrastructure Engineering 26, 368–380.

Chi, S., Caldas, C.H., 2011b. Image-based safety assessment: automated spatial safety risk identification of earthmoving and surface mining activities. Journal of Construction Engineering and Management 138, 341–351.

Chi, S., Caldas, C.H., Kim, D.Y., 2009. A methodology for object identification and tracking in construction based on spatial modeling and image matching techniques. Computer-Aided Civil and Infrastructure Engineering 24, 199–211.

Cooper, M.D., 2000. Towards a model of safety culture. Safety science 36, 111–136.

Cooper Ph. D, M., 2000. Towards a model of safety culture. Safety science 36, 111-136.

Dai, J., Li, Y., He, K., Sun, J., 2016. R-fcn: Object detection via region-based fully convolutional networks. Advances in neural information processing systems 379–387.

Dalal, N., Triggs, B., 2005. Histograms of oriented gradients for human detection, Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on. IEEE, pp. 886-893.

Dedobbeleer, N., Beland, F., 1991. A safety climate measure for construction sites. Journal of Safety Research 22, 97–103.

Dimitrov, A., Golparvar-Fard, M., 2014. Vision-based material recognition for automated monitoring of construction progress and generating building information modeling from unordered site image collections. Advanced Engineering Informatics 28, 37–49.

Ding, L., Fang, W., Luo, H., Love, P.E., Zhong, B., Ouyang, X., 2018. A deep hybrid learning model to detect unsafe behavior: Integrating convolution neural networks and long short-term memory. Automation in Construction 86, 118–124.

Edwards, M., Deng, J., Xie, X., 2016. From pose to activity: Surveying datasets and introducing CONVERSE. Computer Vision and Image Understanding 144, 73–105.

Endsley, M.R., 1995. Toward a theory of situation awareness in dynamic systems. Human factors 37, 32–64.

Endsley, M.R., 2015. Situation awareness misconceptions and misunderstandings. Journal of Cognitive Engineering and Decision Making 9, 4–32.

Falagas, M.E., Pitsouni, E.I., Malietzis, G.A., Pappas, G., 2008. Comparison of PubMed, Scopus, web of science, and Google scholar: strengths and weaknesses. The FASEB journal 22, 338–342.

Fang, D., Chen, Y., Wong, L.J.J.o.c.e., management, 2006. Safety climate in construction industry: A case study in Hong Kong. 132, 573-584.

Fang, Q., Li, H., Luo, X., Ding, L., Luo, H., Li, C., 2018a. Computer vision aided inspection on falling prevention measures for steeplejacks in an aerial environment. Automation in Construction 93, 148–164.

Fang, Q., Li, H., Luo, X., Ding, L., Luo, H., Rose, T.M., An, W., 2018b. Detecting non-hardhat-use by a deep learning method from far-field surveillance videos. Automation in Construction 85, 1–9.

Fang, Q., Li, H., Luo, X., Ding, L., Rose, T.M., An, W., Yu, Y., 2018c. A deep learning-based method for detecting non-certified work on construction sites. Advanced Engineering Informatics 35, 56–68.

Fang, W., Ding, L., Luo, H., Love, P.E., 2018d. Falls from heights: A computer vision-based approach for safety harness detection. Automation in Construction 91, 53–61.

Fang, W., Ding, L., Zhong, B., Love, P.E., Luo, H., 2018e. Automated detection of workers and heavy equipment on construction sites: A convolutional neural network approach. Advanced Engineering Informatics 37, 139–149.

Farha, Y.A., Richard, A., Gall, J., 2018. When Will You Do What? - Anticipating Temporal Occurrences of Activities, The IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Computer Vision Foundation, Salt Lake City, pp. 5343–5352.

Felzenszwalb, P.F., Girshick, R.B., McAllester, D., Ramanan, D., 2009. Object detection with discriminatively trained part-based models. IEEE transactions on pattern analysis and machine intelligence 32, 1627–1645.

Fraser, M., Elgamal, A., He, X., Conte, J.P., 2009. Sensor network for structural health monitoring of a highway bridge. Journal of Computing in Civil Engineering 24, 11–24.
Gatica-Perez, D., McCowan, L., Zhang, D., Bengio, S., 2005. Detecting group interest-level in meetings, Acoustics, Speech, and Signal Processing, 2005. Proceedings. (ICASSP'05). IEEE International Conference on. IEEE, pp. I/489-I/492 Vol. 481.
Geller, E.S., 1994. Ten principles for achieving a total safety culture. Professional Safety 39, 18.
Glendon, A.I., Litherland, D.K., 2001. Safety climate factors, group differences and safety behaviour in road construction. Safety science 39, 157–188.
Goh, Y.M., Guo, B.H., 2018. FPSWizard: A web-based CBR-RBR system for supporting the design of active fall protection systems. Automation in Construction 85, 40–50.
Goh, Y.M., Ubeynarayana, C., 2017. Construction accident narrative classification: an evaluation of text mining techniques. Accident Analysis & Prevention 108, 122–130.
Goh, Y.M., Ubeynarayana, C.U., Wong, K.L.X., Guo, B.H., 2018. Factors influencing unsafe behaviors: A supervised learning approach. Accident Analysis & Prevention 118, 77–85.
Golparvar-Fard, M., Heydarian, A., Niebles, J.C., 2013. Vision-based action recognition of earthmoving equipment using spatio-temporal features and support vector machine classifiers. Advanced Engineering Informatics 27, 652–663.
Gong, J., Caldas, C.H., 2009. Computer vision-based video interpretation model for automated productivity analysis of construction operations. Journal of Computing in Civil Engineering 24, 252–263.
Gong, J., Caldas, C.H., 2011. An object recognition, tracking, and contextual reasoning-based video interpretation method for rapid productivity analysis of construction operations. Automation in Construction 20, 1211–1226.
Gong, J., Caldas, C.H., Gordon, C., 2011. Learning and classifying actions of construction workers and equipment using Bag-of-Video-Feature-Words and Bayesian network models. Advanced Engineering Informatics 25, 771–782.
Guldenmund, F.W., 2007. The use of questionnaires in safety culture research–an evaluation. Safety science 45, 723–743.
Guo, B.H., Goh, Y.M., 2017. Ontology for design of active fall protection systems. Automation in Construction.
Guo, B.H., Goh, Y.M., Wong, K.L.X., 2018. A system dynamics view of a behavior-based safety program in the construction industry. Safety science 104, 202–215.
Guo, B.H., Yiu, T.W., González, V.A., 2015. A system dynamics view of safety management in small construction companies. Journal of Construction Engineering and Project Management 5, 1–6.
Guo, B.H., Yiu, T.W., González, V.A., 2016a. Predicting safety behavior in the construction industry: Development and test of an integrative model. Safety science 84, 1–11.
Guo, B.H., Yiu, T.W., González, V.A., Goh, Y.M., 2016b. Using a Pressure-State-Practice Model to Develop Safety Leading Indicators for Construction Projects. Journal of Construction Engineering and Management 143, 04016092.
Guo, B.H.W., Yiu, T.W., 2016. Developing Leading Indicators to Monitor the Safety Conditions of Construction Projects. Journal of Management in Engineering 32, 04015016.
Guo, H., Scheepbouwer, E., Yiu, T., Gonzalez, V., 2017a. Overview and analysis of digital technologies for construction safety management. In: Lamb, M. (Ed.), Australasian Universities Building Education Association Conference 2017, Melbourne.
Guo, H., Yu, Y., Skitmore, M.J.A.i.C., 2017b. Visualization technology-based construction safety management: A review. 73, 135-144.
Hamledari, H., McCabe, B., Davari, S., 2017. Automated computer vision-based detection of components of under-construction indoor partitions. Automation in Construction 74, 78–94.
Han, S., Lee, S., 2013. A vision-based motion capture and recognition framework for behavior-based safety management. Automation in Construction 35, 131–141.
Han, S., Lee, S., Peña-Mora, F., 2014. Comparative study of motion features for similarity-based modeling and classification of unsafe actions in construction. Journal of Computing in Civil Engineering 28, A4014005.
Hill, S., 1988. Technology and organization culture: The human imperative in integrating new technology into organization design. Technology in society 10, 233–253.
Ibrahim, Y., Lukins, T.C., Zhang, X., Trucco, E., Kaka, A., 2009. Towards automated progress assessment of workpackage components in construction projects using computer vision. Advanced Engineering Informatics 23, 93–103.
Jeelani, I., Albert, A., Gambatese, J.A., 2016. Why do construction hazards remain unrecognized at the work interface? Journal of Construction Engineering and Management 143, 04016128.
Jin, R., Zou, P.X.W., Piroozfar, P., Wood, H., Yang, Y., Yan, L., Han, Y., 2019. A science mapping approach based review of construction safety research. Safety science 113, 285–297.
Kalal, Z., Mikolajczyk, K., Matas, J., 2011. Tracking-learning-detection. IEEE transactions on pattern analysis and machine intelligence 34, 1409–1422.
Khosrowpour, A., Niebles, J.C., Golparvar-Fard, M., 2014. Vision-based workface assessment using depth images for activity analysis of interior construction operations. Automation in Construction 48, 74–87.
Kim, D., Liu, M., Lee, S., Kamat, V.R., 2019a. Remote proximity monitoring between mobile construction resources using camera-mounted UAVs. Automation in Construction 99, 168–182.
Kim, H., Ahn, C.R., Engelhaupt, D., Lee, S., 2018a. Application of dynamic time warping to the recognition of mixed equipment activities in cycle time measurement. Automation in Construction 87, 225–234.
Kim, H., Ham, Y., Kim, W., Park, S., Kim, H.J.A.i.C., 2019b. Vision-based nonintrusive context documentation for earthmoving productivity simulation. 102, 135-147.
Kim, H., Kim, H., 2018. 3D reconstruction of a concrete mixer truck for training object detectors. Automation in Construction 88, 23–30.
Kim, H., Kim, H., Hong, Y.W., Byun, H., 2017. Detecting Construction Equipment Using a Region-Based Fully Convolutional Network and Transfer Learning. Journal of Computing in Civil Engineering 32, 04017082.
Kim, H., Kim, K., Kim, H., 2015. Vision-based object-centric safety assessment using fuzzy inference: Monitoring struck-by accidents with moving objects. Journal of Computing in Civil Engineering 30, 04015075.
Kim, H., Kim, K., Kim, H., 2016. Data-driven scene parsing method for recognizing construction site objects in the whole image. Automation in Construction 71, 271–282.
Kim, J., Chi, S., 2017. Adaptive detector and tracker on construction sites using functional integration and online learning. Journal of Computing in Civil Engineering 31, 04017026.
Kim, J., Chi, S., 2019. Action recognition of earthmoving excavators based on sequential pattern analysis of visual features and operation cycles. Automation in Construction 104, 255–264.
Kim, J., Chi, S., Seo, J., 2018b. Interaction analysis for vision-based activity identification of earthmoving excavators and dump trucks. Automation in Construction 87, 297–308.
Kolar, Z., Chen, H., Luo, X., 2018. Transfer learning and deep convolutional neural networks for safety guardrail detection in 2D images. Automation in Construction 89, 58–70.
Konstantinou, E., Brilakis, I., 2018. Matching construction workers across views for automated 3D vision tracking on-site. Journal of Construction Engineering and Management 144, 04018061.
Konstantinou, E., Lasenby, J., Brilakis, I., 2019. Adaptive computer vision-based 2D tracking of workers in complex environments. Automation in Construction.
Koppula, H.S., Saxena, A., 2016. Anticipating Human Activities Using Object Affordances for Reactive Robotic Response. IEEE transactions on pattern analysis and machine intelligence 38, 14–29.
Kropp, C., Koch, C., König, M., 2018. Interior construction state recognition with 4D BIM registered image sequences. Automation in Construction 86, 11–32.
Lan, T., Chen, T.-C., Savarese, S., 2014. A hierarchical representation for future action prediction. European Conference on Computer Vision. Springer 689–704.
Le Coze, J.-C., 2013. Outlines of a sensitising model for industrial safety assessment. Safety science 51, 187–201.
Le Coze, J.C., 2016. Vive la diversité! High reliability organisation (HRO) and resilience engineering (RE). Safety science.
Lee, Y.-J., Park, M.-W., 2019. 3D tracking of multiple onsite workers based on stereo vision. Automation in Construction 98, 146–159.
Lehmann-Willenbrock, N., Hung, H., Keyton, J., 2017. New frontiers in analyzing dynamic group interactions: Bridging social and computer science. Small group research 48, 519–531.
Li, X., Yi, W., Chi, H.-L., Wang, X., Chan, A.P.J.A.i.C., 2018. A critical review of virtual and augmented reality (VR/AR) applications in construction safety. 86, 150-162.
Liang, C.-J., Lundeen, K.M., McGee, W., Menassa, C.C., Lee, S., Kamat, V.R., 2019. A vision-based marker-less pose estimation system for articulated construction robots. Automation in Construction 104, 80–94.
Lienhart, R., Maydt, J., 2002. An extended set of haar-like features for rapid object detection, Image Processing. 2002. Proceedings. 2002 International Conference on. IEEE, pp. I-I.
Lingard, H.C., Cooke, T., Blismas, N.J.C.m., economics, 2009. Group-level safety climate in the Australian construction industry: within-group homogeneity and between-group differences in road construction and maintenance. 27, 419-432.
Liu, K., Golparvar-Fard, M.J.J.o.C.E., Management, 2015. Crowdsourcing construction activity analysis from jobsite video streams. 141, 04015035.
Lowe, D.G., 1999. Object recognition from local scale-invariant features, 7th International Conference on Computer Vision, Corfu, Greece, pp. 1150–1157.
Luo, H., Xiong, C., Fang, W., Love, P.E., Zhang, B., Ouyang, X., 2018a. Convolutional neural networks: Computer vision-based workforce activity assessment in construction. Automation in Construction 94, 282–289.
Luo, X., Li, H., Cao, D., Dai, F., Seo, J., Lee, S.J.J.o.C.i.C.E., 2018b. Recognizing diverse construction activities in site images via relevance networks of construction related objects detected by convolutional neural networks. 32, 04018012.
Luo, X., Li, H., Cao, D., Yu, Y., Yang, X., Huang, T., 2018c. Towards efficient and objective work sampling: Recognizing workers' activities in site surveillance videos with two-stream convolutional networks. Automation in Construction 94, 360–370.
Luo, X., Li, H., Wang, H., Wu, Z., Dai, F., Cao, D., 2019. Vision-based detection and visualization of dynamic workspaces. Automation in Construction 104, 1–13.
Makin, A.M., Winder, C., 2008. A new conceptual framework to improve the application of occupational health and safety management systems. Safety science 46, 935–948.
Memarzadeh, M., Golparvar-Fard, M., Niebles, J.C., 2013. Automated 2D detection of construction equipment and workers from site video streams using histograms of oriented gradients and colors. Automation in Construction 32, 24–37.
Mneymneh, B.E., Abbas, M., Khoury, H., 2018. Vision-Based Framework for Intelligent Monitoring of Hardhat Wearing on Construction Sites. Journal of Computing in Civil Engineering 33, 04018066.
Mohamed, S., 2002. Safety climate in construction site environments. Journal of construction engineering management 128, 375–384.
Moon, S., Bernold, L.E., 1997. Vision-based interactive path planning for robotic bridge paint removal. Journal of Computing in Civil Engineering 11, 113–120.
Mota, S., Picard, R.W., 2003. Automated posture analysis for detecting learner's interest level, Computer Vision and Pattern Recognition Workshop, 2003. CVPRW'03. Conference on. IEEE, pp. 49-49.

Park, M.-W., Brilakis, I., 2012. Construction worker detection in video frames for initializing vision trackers. Automation in Construction 28, 15–25.

Park, M.-W., Brilakis, I., 2016. Continuous localization of construction workers via integration of detection and tracking. Automation in Construction 72, 129–142.

Park, M.-W., Elsafty, N., Zhu, Z., 2015. Hardhat-wearing detection for enhancing on-site safety of construction workers. Journal of Construction Engineering and Management 141, 04015024.

Park, M.-W., Makhmalbaf, A., Brilakis, I., 2011. Comparative study of vision tracking methods for tracking of construction site resources. Automation in Construction 20, 905–915.

Ranaweera, K., Ruwanpura, J., Fernando, S.J.J.o.C.i.C.E., 2012. Automated real-time monitoring system to measure shift production of tunnel construction projects. 27, 68-77.

Ray, S.J., Teizer, J., 2012. Real-time construction worker posture analysis for ergonomics training. Advanced Engineering Informatics 26, 439–455.

Ren, S., He, K., Girshick, R., Sun, J., 2015. Faster r-cnn: Towards real-time object detection with region proposal networks. Advances in neural information processing systems 91–99.

Rezazadeh Azar, E., 2017. Semantic Annotation of Videos from Equipment-Intensive Construction Operations by Shot Recognition and Probabilistic Reasoning. Journal of Computing in Civil Engineering 31, 04017042.

Rezazadeh Azar, E., Dickinson, S., McCabe, B., 2012. Server-customer interaction tracker: computer vision–based system to estimate dirt-loading cycles. Journal of Construction Engineering and Management 139, 785–794.

Rezazadeh Azar, E., McCabe, B., 2011. Automated visual recognition of dump trucks in construction videos. Journal of Computing in Civil Engineering 26, 769–781.

Ross, D.A., Lim, J., Lin, R.-S., Yang, M.-H., 2008. Incremental learning for robust visual tracking. International journal of computer vision 77, 125–141.

Ryoo, M.S., 2011. Human activity prediction: Early recognition of ongoing activities from streaming videos. Computer Vision (ICCV). In: 2011 IEEE International Conference on. IEEE, pp. 1036–1043.

Sacks, R., Rozenfeld, O., Rosenfeld, Y., 2009. Spatial and temporal exposure to safety hazards in construction. Journal of Construction Engineering and Management 135, 726–736.

Seo, J., Han, S., Lee, S., Kim, H., 2015. Computer vision techniques for construction safety and health monitoring. Advanced Engineering Informatics 29, 239–251.

Seo, J., Starbuck, R., Han, S., Lee, S., Armstrong, T.J., 2014. Motion data-driven biomechanical analysis during construction tasks on sites. Journal of Computing in Civil Engineering 29, B4014005.

Soltani, M.M., Zhu, Z., Hammad, A., 2016. Automated annotation for visual recognition of construction resources using synthetic images. Automation in Construction 62, 14–23.

Soltani, M.M., Zhu, Z., Hammad, A., 2018. Framework for Location Data Fusion and Pose Estimation of Excavators Using Stereo Vision. Journal of Computing in Civil Engineering 32, 04018045.

Son, H., Choi, H., Seong, H., Kim, C., 2019. Detection of construction workers under varying poses and changing background in image sequences via very deep residual networks. Automation in Construction 99, 27–38.

Tajeen, H., Zhu, Z., 2014. Image dataset development for measuring construction equipment recognition performance. Automation in Construction 48, 1–10.

Teizer, J., 2015. Status quo and open challenges in vision-based sensing and tracking of temporary resources on infrastructure construction sites. Advanced Engineering Informatics 29, 225–238.

Teizer, J., Cheng, T., Fang, Y., 2013. Location tracking and data visualization technology to advance construction ironworkers' education and training in safety and productivity. Automation in Construction 35, 53–68.

Teizer, J., Vela, P.A., 2009. Personnel tracking on construction sites using video cameras. Advanced Engineering Informatics 23, 452–462.

Tixier, A.J.-P., Hallowell, M.R., Rajagopalan, B., Bowman, D., 2016. Application of machine learning to construction injury prediction. Automation in Construction 69, 102–114.

Torralba, A., Murphy, K.P., Freeman, W.T., 2004. Sharing features: efficient boosting procedures for multiclass object detection. Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on. IEEE, pp. II-II.

Trong, N.P., Nguyen, H., Kazunori, K., Le Hoai, B., 2017. A Comprehensive Survey on Human Activity Prediction. Springer International Publishing, Cham, pp. 411–425.

Turkan, Y., Bosché, F., Haas, C.T., Haas, R., 2012. Toward automated earned value tracking using 3D imaging tools. Journal of Construction Engineering and Management 139, 423–433.

Vinciarelli, A., Pantic, M., Bourlard, H., 2009. Social signal processing: Survey of an emerging domain. Image and vision computing 27, 1743–1759.

Vrigkas, M., Nikou, C., Kakadiaris, I.A., 2015. A review of human activity recognition methods. Frontiers in Robotics and AI 2, 28.

Wang, X., Love, P.E., Kim, M.J., Park, C.-S., Sing, C.-P., Hou, L., 2013. A conceptual framework for integrating building information modeling with augmented reality. Automation in Construction 34, 37–44.

Wang, Z., Li, H., Zhang, X., 2019. Construction waste recycling robot for nails and screws: Computer vision technology and neural network approach. Automation in Construction 97, 220–228.

Xu, J., Yoon, H.-S., 2019. Vision-based estimation of excavator manipulator pose for automated grading control. Automation in Construction 98, 122–131.

Yan, X., Li, H., Wang, C., Seo, J., Zhang, H., Wang, H., 2017. Development of ergonomic posture recognition technique based on 2D ordinary camera for construction hazard prevention through view-invariant features in 2D skeleton motion. Advanced Engineering Informatics 34, 152–163.

Yang, J., Arif, O., Vela, P.A., Teizer, J., Shi, Z., 2010. Tracking multiple workers on construction sites using video cameras. Advanced Engineering Informatics 24, 428–434.

Yang, J., Shi, Z., Wu, Z., 2016. Vision-based action recognition of construction workers using dense trajectories. Advanced Engineering Informatics 30, 327–336.

Yang, J., Vela, P., Teizer, J., Shi, Z., 2012. Vision-based tower crane tracking for understanding construction activity. Journal of Computing in Civil Engineering 28, 103–112.

Yilmaz, A., Javed, O., Shah, M., 2006. Object tracking: A survey. Acm computing surveys (CSUR) 38, 13.

Yu, Y., Guo, H., Ding, Q., Li, H., Skitmore, M., 2017. An experimental study of real-time identification of construction workers' unsafe behaviors. Automation in Construction 82, 193–206.

Yuan, C., Li, S., Cai, H., 2016. Vision-based excavator detection and tracking using hybrid kinematic shapes and key nodes. Journal of Computing in Civil Engineering 31, 04016038.

Zhang, H., Yan, X., Li, H., 2018. Ergonomic posture recognition using 3D view-invariant features from single ordinary camera. Automation in Construction 94, 1–10.

Zhang, P., Li, N., Jiang, Z., Fang, D., Anumba, C.J., 2019. An agent-based modeling approach for understanding the effect of worker-management interactions on construction workers' safety-related behaviors. Automation in Construction 97, 29–43.

Zhang, S., Boukamp, F., Teizer, J., 2015a. Ontology-based semantic modeling of construction safety knowledge: Towards automated safety planning for job hazard analysis (JHA). Automation in Construction 52, 29–41.

Zhang, S., Sulankivi, K., Kiviniemi, M., Romo, I., Eastman, C.M., Teizer, J., 2015b. BIM-based fall hazard identification and prevention in construction safety planning. Safety science 72, 31–45.

Zhang, S., Teizer, J., Lee, J.-K., Eastman, C.M., Venugopal, M., 2013. Building information modeling (BIM) and safety: Automatic safety checking of construction models and schedules. Automation in Construction 29, 183–195.

Zhou, Z., Goh, Y.M., Li, Q., 2015. Overview and analysis of safety management studies in the construction industry. Safety science 72, 337–350.

Zhu, Z., Park, M.-W., Koch, C., Soltani, M., Hammad, A., Davari, K., 2016a. Predicting movements of onsite workers and mobile equipment for enhancing construction site safety. Automation in Construction 68, 95–101.

Zhu, Z., Ren, X., Chen, Z., 2016b. Visual tracking of construction jobsite workforce and equipment with particle filtering. Journal of Computing in Civil Engineering 30, 04016023.

Zohar, D., 2010. Thirty years of safety climate research: Reflections and future directions. Accident Analysis & Prevention 42, 1517–1522.

Zou, J., Kim, H., 2007. Using hue, saturation, and value color space for hydraulic excavator idle time analysis. Journal of Computing in Civil Engineering 21, 238–246.

Zou, P.X., Lun, P., Cipolla, D., Mohamed, S., 2017a. Cloud-based safety information and communication system in infrastructure construction. Safety science 98, 50–69.

Zou, P.X., Sunindijo, R.Y., 2015. Strategic safety management in construction and engineering. John Wiley & Sons.

Zou, Y., Kiviniemi, A., Jones, S.W., 2017b. A review of risk management through BIM and BIM-related technologies. Safety science 97, 88–98.