Automated Methods for Activity Recognition of Construction Workers and Equipment: State-of-the-Art Review

Behnam Sherafat, S.M.ASCE[1]; Changbum R. Ahn, A.M.ASCE[2]; Reza Akhavian, A.M.ASCE[3]; Amir H. Behzadan, M.ASCE[4]; Mani Golparvar-Fard, M.ASCE[5]; Hyunsoo Kim, M.ASCE[6]; Yong-Cheol Lee, A.M.ASCE[7]; Abbas Rashidi, M.ASCE[8]; and Ehsan Rezazadeh Azar, A.M.ASCE[9]

Abstract: Equipment and workers are two important resources in the construction industry. Performance monitoring of these resources would help project managers improve the productivity rates of construction jobsites and discover potential performance issues. A typical construction workface monitoring system consists of four major levels: location tracking, activity recognition, activity tracking, and performance monitoring. These levels are employed to evaluate work sequences over time and also assess the workers' and equipment's well-being and abnormal edge cases. Results of an automated performance monitoring system could be used to employ preventive measures to minimize operating/repair costs and downtimes. The authors of this paper have studied the feasibility of implementing a wide range of technologies and computational techniques for automated activity recognition and tracking of construction equipment and workers. This paper provides a comprehensive review of these methods and techniques as well as describes their advantages, practical value, and limitations. Additionally, a multifaceted comparison between these methods is presented, and potential knowledge gaps and future research directions are discussed. **DOI: 10.1061/(ASCE)CO.1943-7862.0001843.**

Author keywords: Construction equipment; Worker; Location tracking; Activity recognition; Activity tracking; Performance monitoring; Machine learning; Convolutional neural network; Audio-based method; Kinematic-based method; Vision-based method.

Introduction

Construction jobsites are dynamic environments that include various units operating simultaneously. Each unit contains different groups of workers and equipment, and managing the entire jobsite is a challenging task (Behzadan et al. 2008; Kamat et al. 2010; Zhang et al. 2017). One of the most important duties of a superintendent is to keep track of activities and processes and to ensure that the entire project meets predicted production rates. Unlike manufacturing industries, which include highly similar processes, many construction operations are dissimilar, and using techniques such as the first-run study to conduct upfront learning and improvement does not scale well to many operations. For example, a robotic arm in a factory always performs a task similarly. In contrast, a backhoe in a construction jobsite may perform a specific task using different moves. Also, recent statistics still show that productivity rates in the construction industry—in terms of value put in place per worker—are low when compared with manufacturing industries (Arditi and Mochtar 2000; Goodrum et al. 2000; Gu and Ho 2000). This low performance is partially due to the complexity and uniqueness of operations and activities taking place. As a result, there is a substantial need for effective methods to continuously monitor construction operations.

Traditional Performance Analysis Methods

The traditional approach to monitoring construction operations includes analyzing production rates and performance assessments through direct observations such as work sampling and method productivity delay model, interviews, supervisor and craftsman surveys, and crew-balance charting (Kim et al. 2018a, 2019b; Gong and Caldas 2009; Vahdatikhaki et al. 2015). Jobsite superintendents and supervisors are usually responsible for keeping track of ongoing activities at a jobsite using direct observations (Kim et al. 2018a; Liu and Golparvar-Fard 2015). Interviews and surveys are other means of interacting with project teams to ensure they perform their duties efficiently and on time. Manually monitoring construction operations could be time consuming, error-prone, costly, and not applicable for larger size jobsites where several operations are simultaneously ongoing (Akhavian and Behzadan 2015; Awolusi et al. 2018; Cheok et al. 2000; Golparvar-Fard et al. 2011; Gong and Caldas 2011;

[1]Ph.D. Student of Construction Engineering, Dept. of Civil and Environmental Engineering, Univ. of Utah, Salt Lake City, UT 84112 (corresponding author). ORCID: https://orcid.org/0000-0002-1187-0277. Email: behnam.sherafat@utah.edu

[2]Associate Professor, Dept. of Construction Science, Texas A&M Univ., College Station, TX 77843. Email: ryanahn@tamu.edu

[3]Assistant Professor, Dept. of Civil, Construction, and Environmental Engineering, San Diego State Univ., San Diego, CA 92182. Email: rakhavian@sdsu.edu

[4]Associate Professor, Dept. of Construction Science, Texas A&M Univ., College Station, TX 77843. Email: abehzadan@tamu.edu

[5]Associate Professor, Dept. of Civil and Environmental Engineering, Univ. of Illinois at Urbana-Champaign, Champaign, IL 61801. Email: mgolpar@illinois.edu

[6]Assistant Professor, Dept. of Architectural Engineering, Dankook Univ., Yongin, Gyeonggi-do 16890, Republic of Korea. Email: hkim13@dankook.ac.kr

[7]Assistant Professor, Dept. of Construction Management, Louisiana State Univ., Baton Rouge, LA 70803. Email: yclee@lsu.edu

[8]Assistant Professor, Dept. of Civil and Environmental Engineering, Univ. of Utah, Salt Lake City, UT 84112. Email: abbas.rashidi@utah.edu

[9]Associate Professor, Dept. of Civil Engineering, Lakehead Univ., Thunder Bay, ON, Canada P7B 5E1. Email: eazar@lakeheadu.ca

Note. This manuscript was published online on April 8, 2020. Discussion period open until September 8, 2020; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Construction Engineering and Management***, © ASCE, ISSN 0733-9364.

Fig. 1. Overview of an automated construction monitoring system.

Joshua and Varghese 2010b). Therefore, there is a growing demand for alternative automated solutions.

Collecting and processing data in a timely manner is another important component of an efficient performance monitoring system. Traditional performance monitoring approaches are not real-time and rely heavily on past experiences of jobsite personnel to analyze various situations and take corrective actions when necessary. The construction industry ultimately requires automated performance monitoring systems capable of collecting and analyzing performance data and providing feedback and corrective decisions in real-time (or near-real-time) settings (Luo et al. 2018; Roberts and Golparvar-Fard 2019).

Overview of Automated Construction Monitoring System

Although the construction industry still suffers from a lack of a holistic, automated, real-time performance monitoring system, several researchers have tried to address the issue in recent years. A typical visual data–driven and automated resource monitoring on construction jobsites consist of several levels, as shown in Fig. 1. At the first level, spatial locations of operations (and/or resources) need to be identified and continuously tracked, which is shown in dark gray in Fig. 1. This level is independent of other levels because it aims to estimate three-dimensional (3D) locations of construction equipment or workers at discrete points in time and track their movements in real time. This level leads to a four-dimensional (4D) system that could track the construction site dynamics (Tang and Golparvar-Fard 2017). This activity aims to evaluate the performance of a conceptual framework designed to capture the 4D site dynamics by integrating data streams received from active/passive spatial sensors. The ultimate goal of this level is to collect spatial location data at any given time for further analysis.

The next level consists of three sublevels, which are shown in light gray in Fig. 1. In the first sublevel, resource activities need to be recognized. Various equipment types and workers perform different activities simultaneously on a jobsite, thus necessitating an efficient method to recognize these activities. Equipment/worker activities are ongoing processes over time, whereas actions are single efforts. In other words, actions, when grouped together, form activity, and each action includes a series of consecutive movements (Khosrowpour et al. 2014b; Roberts et al. 2018). Therefore, these actions are not independent events but rather a sequence of events correlated with one another. Moreover, there is a difference between detection and recognition. Activity detection refers to a technology that can identify the occurrence of activity within digital images, sounds, and so on. On the other hand, activity recognition describes a technology that attempts to establish which type of activity is taking place. The next sublevel aims to exploit the previous level by recognizing activities in different time periods and tracking activities such that the system is responsive in real-time.

The final sublevel (i.e., performance monitoring) aims to determine which activities are completed and monitor the progress of activities that are underway. This information can be used to generate Key Performance Indicator (KPI) dashboards on activities and crew-balance charts (Roberts and Golparvar-Fard 2019). In the case of heavy-construction equipment, this activity not only does extend the monitoring concept to equipment life-logging, maintenance records, fuel consumption, active and idle times, process charts, and recommended corrective measures, but it also considers the entire fleet of equipment as a system. A project performance control framework is capable of converting the outcome of the aforementioned research activities (e.g., the 4D spatial location of the equipment/workers, activity recognition, and activity tracking, among others) into usable information and Process Performance Indicators (PPIs) such that corrective decisions could be made immediately.

Two of the most significant resources for construction projects are equipment and workers, each having different characteristics and productivity rates. Construction equipment operations need to be systematically measured and evaluated to maintain the project budget and schedule (Ahn et al. 2012; Gong and Caldas 2011; Kim et al. 2018a, 2019b). For example, a backhoe needs to excavate and move a specific amount of soil over a predefined duration. Construction managers must keep track of the activities performed by each piece of equipment to ensure that its predefined productivity is accurately satisfied by taking corrective actions such as replacing parts/operator or repairing the engine.

Construction workers (skilled, semiskilled, and unskilled) are another important resource within the industry. Numbers show that labor costs—particularly in the US and Canada—comprise on average 33%–50% of the total budget in a typical construction project (Hanna 2001; Khan and Sohail 2013). Superintendents must constantly communicate and coordinate work sequences with workers to ensure they are on track and to prevent delays to critical and value-adding activities. Therefore, detecting and analyzing worker and equipment activities would be the first step toward tasks such as productivity measurement and analysis, safety, and quality control (Cheng et al. 2018; Joshua and Varghese 2010a; Wang et al. 2017) of construction operations.

Overview of Automated Activity Recognition Methods

In the last 2 decades, advancements in emerging technologies have enabled researchers and practitioners to move toward developing automated, real-time performance monitoring systems for the construction industry (Asadi et al. 2019a, b; Awolusi et al. 2018). Recent developments in the domain of sensor-based technologies (from both hardware and software perspectives) have helped construction managers efficiently interact with other parties in the jobsite and achieve increased productivity and safety performance (Cezar 2012; Han and Golparvar-Fard 2017; Kamišalić et al. 2018; Zhang et al. 2017; Yang et al. 2014). Based on the type of sensors implemented, automated activity detection and recognition methods could be divided into three major categories, namely (1) kinematic-based methods; (2) computer vision–based methods; and (3) audio-based methods.

Kinematic-based methods utilize several sensors such as accelerometers and gyroscopes to recognize different kinematic patterns of activities carried out by construction workers and equipment. These sensors can be microfabricated into an electronic chip, such as an inertial measurement unit (IMU), to collect data that upon processing could provide information about the rotational speed and orientation of equipment. Also, some IMUs are based on a technology called microelectromechanical systems (MEMS), which is the most popular sensor type due to its small size and low cost. MEMS sensors are quite common in the other industries and have a variety of applications in several domains such as health care, unsupervised home monitoring (home telecare), fall detection, weather monitoring, recognition of athletes' movement patterns, asset management, and industrial control (Hanna 2001; Khan and Sohail 2013; Kim et al. 2019b). On the other hand, these devices have recently been introduced and utilized in construction for various purposes such as physiological monitoring, environmental sensing, proximity detection, location tracking, activity detection, and safety measurement and monitoring.

Some sensors such as the Global Positioning System (GPS) and some technologies such as radio-frequency identification (RFID) tags and ultrawideband (UWB) are also important for activity detection. For example, knowing that an excavator is close to a truck likely means that the excavator is loading the truck (Akhavian and Behzadan 2013a). However, these sensors have more applications in construction machine and worker location tracking. On the other hand, accelerometers and gyroscopes are relatively suitable for activity detection purposes (Lim et al. 2015). These types of sensors capture the acceleration and rotation along the *x*-, *y*-, and *z*-axes. The location of these sensors on the equipment or worker can influence the accuracy of activity recognition (Joshua and Varghese 2013; Nath et al. 2017).

Furthermore, audio-based methods mainly rely on recording sound patterns of equipment performing certain tasks. These methods have been used in recent years as a suitable alternative to kinematic-based methods. The most common tools used in this approach are ordinary microphones, contact microphones, and microphone arrays. Microphones can be categorized according to their pick-up pattern and type of transducer (Cheng et al. 2018). Collected audio data are then analyzed using signal processing algorithms to recognize different types of field activities.

Finally, computer vision–based methods use two-dimensional (2D) image/video cameras and 3D range cameras (e.g., Flash LiDAR) to capture visual data from construction jobsites for further processing. These methods process images or videos captured from the construction jobsite by different types of cameras such as depth cameras [e.g., Red Green Blue-Depth (RGB-D)] (Khosrowpour et al. 2014b) that must be installed in proper locations with clear lines of sight to record all ongoing activities during construction.

In addition to implementing each method individually, using multiple sensor types for collecting data (i.e., a multimodal sensor) can increase the accuracy of activity detection (Awolusi et al. 2018). Moreover, along with choosing the optimal sensor types as preliminary sources of data, other factors such as the optimum number and position of sensors and the data analysis techniques play critical roles in the success of activity detection and monitoring (Joshua and Varghese 2010a).

This paper provides a concise review of the most recent studies that have utilized the aforementioned methods to detect the activities of construction equipment and workers. To achieve this goal, the study is structured as follows. The next section presents an overview of various applications of construction activity detection, which are identified through a detailed literature review. Then, the authors elaborate on details regarding common tools and techniques suggested by researchers for automatically detecting the activities of construction resources. Finally, conclusions, discussions, and future research directions are presented.

Potential Applications for Activity Recognition and Tracking Systems within the Construction Industry

There are various reasons why project managers could be interested in recognizing and continuously tracking activities of workers and heavy equipment on construction jobsites. The authors conducted a thorough literature review and identified key potential applications, as listed in Fig. 2. The following subsections provide a detailed account of each potential application for equipment, workers, and both, respectively.

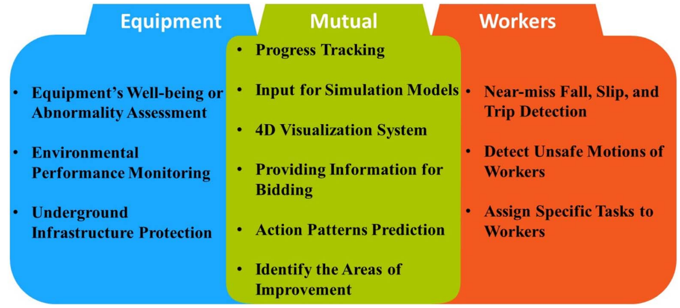

Fig. 2. Construction activity recognition applications.

Applications of Activity Recognition and Construction Equipment Tracking

Heavy equipment and tools are essential resources for the success of any construction project. Recognition of construction machines and activity tracking could lead to several useful outcomes, as follows.

Machines' Well-Being or Abnormality Assessment

New equipment models possess health monitoring systems that can provide valuable information, such as fuel consumption and utilization time. However, those systems are not compatible with other types of equipment, especially older models. By collecting data, monitoring equipment status, and recognizing equipment activities over time, equipment's well-being or abnormality can be detected so project managers can take preventive measures when needed to reduce operating/repair costs and idle times (Kim et al. 2018a).

Environmental Performance Monitoring

Environmental performance monitoring is one of the most crucial applications of equipment activity recognition. Heavy equipment used in construction projects often releases harmful smoke, which is unhealthy for project personnel and is detrimental to the environment. Continuous emission control during field operations will help locate and report deficiencies for further maintenance, repair, or corrective action (Ahn et al. 2012). In the last few years, several methods have been introduced to detect these deficiencies based on equipment performance. Equipment engines generate distinctive sound and vibration patterns when performing different activities, which makes it possible to recognize if the engine is working properly. Following this procedure, audio/kinematic patterns of old and obsolete machines could be identified using activity recognition, compared with newer models, thus limiting the emission of more pollutants in the air. To this end, detecting idle times and non-value-adding activities through activity recognition is particularly beneficial to reducing fuel use and consequently lowering equipment emissions (Akhavian and Behzadan 2014).

Underground Infrastructure Protection

One major issue during excavating operations is the likelihood of causing damage to underground pipelines. Thus, there is an inherent need to develop automated systems capable of detecting potential damage to underground utility lines including cables, gas pipelines, and communication networks (Cao et al. 2015, 2017a, b; Yang et al. 2015).

J. Constr. Eng. Manage., 2020, 146(6): 03120002

Applications of Activity Detection and Tracking of Construction Workers

Human labor is a vital resource in every construction project. Workers are often vulnerable to various safety hazards and potential accidents. Continuous monitoring of workers' activities, behaviors, and mental state could potentially decrease the rate of jobsite accidents, prevent injuries, falls, or work-related musculoskeletal disorders (WMSDs) (Nath et al. 2018; Ren-Jye et al. 2018). The following subsections summarize specific applications and potential benefits of recognizing and tracking workers on construction jobsites.

Detecting Near-Miss Falls, Slips, and Trips

A near-miss incident can be defined as an event in which no damage or injuries actually occurred, but which, under slightly different circumstances, could have resulted in harm (Kunreuther et al. 2004). The major causes of death and injury on a construction jobsite are slips, trips, and falls (Yoon and Lockhart 2006). Monitoring workers' abnormal bodily responses at a specific location through activity detection gives the opportunity to quickly identify near-miss incidents (and potentially risky conditions) and take appropriate actions to prevent actual accidents before they occur by removing jobsite obstacles or securing fall areas (Joshua and Varghese 2010a; Lee et al. 2020; Yang et al. 2014).

Detecting Unsafe Worker Motions

Because construction workers perform tasks involving forceful exertion while in awkward postures, some workers are not familiar with the correct ways of walking, sitting, and picking up/holding materials at construction jobsites. Construction workers have exhibited approximately a 50% higher risk of WMSDs than workers in other industries (Schneider 2001). Both direct (e.g., vision and imaging technologies) and indirect (e.g., physiological sensor) observation of workers to inform them about correct physical motions are tedious processes. Monitoring human body motions can identify several postures with high risks of physical injury (Seo et al. 2014). Capturing awkward postures occurring repetitively provides necessary information to construction managers to improve workers' postures by instructing and training them in simulated environments (Ren-Jye et al. 2018). The level of force, as a risk indicator associated with WMSDs, can be quantified using direct measurements (Jahanbanifar and Akhavian 2018). In addition to directly monitoring human posture, automatic tracking of workers' vital signs in the construction industry can be achieved by monitoring workers' physiological measures such as body temperature and heart rate. This can help identify aberrations in workers' bodies, and predict the occurrence of near-miss incidents (Lim et al. 2015).

Assigning Specific Tasks to Workers

Construction workers feel more fatigued as time goes by because many tasks require physical exertion. Managing workers' fatigue helps maintain desired labor productivity and reduces the risk of accidents (Hallowell 2010). By monitoring workers' movements, unusual motions can be detected and reported to the project manager so that fewer physically challenging tasks are assigned to an injured or exhausted worker (Joshua and Varghese 2010a).

Applications of Activity Recognition and Tracking for Both Workers and Equipment

Other than the aforementioned applications of activity recognition and tracking for workers and equipment categories separately, additional benefits can be achieved if these tasks are conducted simultaneously for both categories. Perhaps the most common application of activity recognition for construction resources is measuring productivity, which is a precursor to estimating cost and project schedule. In the following subsections, potential direct and indirect benefits of activity detection/recognition for both equipment and workers are briefly discussed.

Progress Monitoring

Recognizing activities is essential for tracking progress and calculating quantities of accomplished work (Rashid and Louis 2020; Slaton et al. 2020). By comparing the actual and planned quantities of accomplished work, the estimated completion time can be modified, and necessary preventive/corrective actions can be planned (Golparvar-Fard et al. 2011). Contractors also need to monitor and record actual productivity rates as a valuable source of information for future projects (Kim et al. 2018a). Useful information for these purposes might include durations of non-value-adding activities such as maneuvering and swinging for heavy equipment, and calculating cycle times for repetitive activities of both construction machinery and workers (Sabillon et al. 2020).

Generating Input for Simulation Models

In the last 3 decades, construction simulation has been utilized to address the stochastic nature of activities, predict project completion times, and improve resource allocation prior to the start of the project. Such stochastic simulation models need accurate input data in order to generate reliable outputs. Traditional methods use static input data that do not cover factual aspects of the project because of the dynamic and transient nature of construction operations (Song and Eldin 2012). Continuously updating simulation models based on actual data during the construction phase yields more reliable results. Moreover, there is a need for a systematic way to modify the model based on the ground-truth data (Akhavian and Behzadan 2013b).

Activity recognition helps find the duration and cycle time of activities and their utilized resources. Activity recognition and monitoring systems could serve as a systematic method for data collection and then transfer data to simulation models during the life cycle of projects to acquire precise results. By deriving the latest results from simulation models, project managers can analyze what-if scenarios, evaluate the situation, choose alternative plans, and communicate those plans with their personnel (Akhavian and Behzadan 2012).

Developing 4D Visualization Systems

The simulation model output can further be utilized to generate 4D or five-dimensional (5D) models for construction projects. These 4D/5D models can visualize the construction process along with related ongoing costs. Additionally, these models perform a substantial role in the future of construction management due to their application in creating a general and synergetic framework to store and manage project information during its life cycle. Project managers can rely on these models to detect bottlenecks in the project.

Providing Information for Bidding

A construction bid is a process of providing a potential customer with a proposal to build or manage the building of a structure. It is also a method through which subcontractors pitch their services to general contractors. By determining the work spent on value-adding and non-value-adding activities (e.g., waiting and preparing required materials), the opportunity to win a bid increases. Because the difference between the estimated costs and actual costs will be decreased, contractors can bid on the project with more certainty and escalate their profit margins.

Predicting Action Patterns of Repetitive Activities

Considering the characteristics of construction operations, repetitive actions are important elements in a construction system. Despite the importance of repetitive actions, predicting action patterns

Table 1. Basic components of different kinematic-based activity recognition systems for equipment and tools

References	Objectives	Equipment type and quantity	Number of classifications (different types of activities)	Tools (sampling frequency)	Testing environment (setting)
Ahn et al. (2012)	1. Elaborate on the relationship between operational efficiency and environmental performance 2. Monitor the operational status of equipment	Medium-sized excavator [CAT 321C long truck excavator compact radius (LCR), Deerfield, IL]	3 activities (engine off, idle, and working)	3-axes accelerometer (100 Hz)	Real project (equipment cabin)
Akhavian and Behzadan (2012)	Develop a remote tracking technique to capture field data from construction equipment	Model equipment (loader and truck)	—	3-axis magnetic field sensing, and 3-axis tilt sensing	Indoor laboratory-scale setting
Akhavian and Behzadan (2013b)	Multimodal-process capturing for automated simulation models generation	2 types of equipment	4 activities (load, haul, dump, and return)	1. A network of UWB receivers and tags 2. Attitude and heading reference system (AHRS) 3. Zigbee-enabled weight	—
Ahn et al. (2013)	Analyze the idle time ratio of equipment	4 types of equipment (excavator)	3 activities (working, idling, and engine off)	Accelerometer withrange of $\pm 2g$ (100 Hz)	Instructed operator (inside the cabin)
Akhavian and Behzadan (2015)	Develop an automated method to detect equipment activities and their durations for simulation input modeling	Equipment (front-end loader)	7 activities (engine off, forward and backward moving, stationary idling, lowering the boom, raising the boom, scooping, and dumping)	GPS sensor, 3-axis accelerometer, and 3-axis gyroscope (100 Hz)	Controlled (i.e., instructed) and normal (uninstructed)
Kim et al. (2018a)	Measure the construction equipment operation cycle time	Equipment (excavator)	5 activities [idle, wheelbase motion, cabin rotation (anticlockwise and clockwise rotation), and bucket/armmovement]	Inertial measurement units (IMUs) embedded in a smartphone (128 Hz)	Real project (equipment cabin)
Rashid and Louis (2019)	1. Time-series data augmentation to generate synthetic training data 2. Recognize equipment activities	2 types of equipment (excavator and front-end loader)	9 activities for excavator, and 10 activities for front-end loader	3-axis accelerometer, and 3-axis gyroscope (80 Hz)	Real project (bucket, arm, and boom)
Bae et al. (2019)	Identify excavator activities using joystick signals	DX220LC excavator (Suwanee, GA)	6 activities (digging, leveling, lifting, trenching, traveling, and idling)	Electronic joystick and peripheral component interconnect (PCI) eXtensions for instrumentation (PXI)	Real project (equipment joystick)

is not performed well due to a lack of accurate data. Predicted action patterns can be used for calculating activity cycle times and ultimately, the overall project completion time (Akhavian and Behzadan 2015).

Identifying Areas of Improvement

Activity recognition of both workers and equipment can provide detailed information such as idle time, operation time, and moving time. Analyzing recognized activities can be helpful in identifying factors that inhibit productivity. By controlling those factors, the management team can remove waste elements (Kim et al. 2018a; Roberts and Golparvar-Fard 2019; Roberts et al. 2018). In addition, considering that materials are moved and installed by workers and equipment, activity recognition can also be extended to managing material (Zhu et al. 2017) and identifying areas for improvement in construction jobsites.

Automated Methods for Activity Detection of Construction Resources: Literature Review

Due to the complexity and unique characteristics of projects, the construction industry suffers from serious issues such as low productivity rates, frequent delays in projects' completion dates, and significant deviations between anticipated and actual costs of projects. Automatically detecting and monitoring construction resources (both equipment and worker) is a first step to addressing those issues. As indicated previously, the automated methods for detecting and recognizing activities of construction equipment and workers mainly depend on types of input data and the selected sensors/techniques for collecting and processing data. Common activity detection techniques could be divided into three major categories, namely (1) kinematic-based, (2) vision-based, and (3) audio-based methods. The following sections provide more details about each category as well as the implemented algorithms and tools for data collection and processing.

Kinematic-Based Methods for Activity Recognition of Construction Equipment

Most construction machines generate distinct kinematic signals while performing different tasks (Ahn et al. 2013; Akhavian and Behzadan 2015). These kinematic signals include acceleration, angular velocity, magnetic fields, and orientation data. It is possible to correlate each activity with specific kinematic signal patterns.

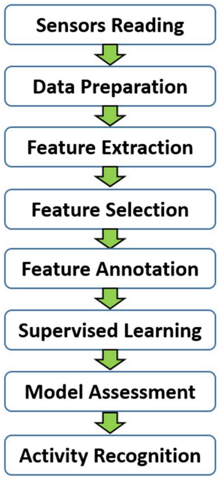

Fig. 3. Overview of the kinematic-based methodology for activity recognition.

Several studies have been conducted on applications of kinematic signals for recognizing construction equipment activities. Table 1 summarizes general information about these studies, including the main objectives, number and types of equipment studied, number of activities covered, devices used to record data, the testing environment, and the placement of the devices on the jobsite or equipment cabin. An overview of a general framework for a kinematic-based activity detection system is presented in Fig. 3.

Processing recorded kinematic signals and extracting useful information are other important aspects of developing a robust activity recognition system. Several studies have adopted machine-learning algorithms to distinguish various activities by extracting different features. Table 2 presents an overview of popular processing techniques including types of implemented machine-learning algorithms, features utilized, procedures for feature extraction, and ultimately the reported levels of accuracies.

Image/Video-Based Methods for Activity Detection of Construction Equipment

Computer-vision techniques present an alternative solution to kinematic-based techniques for tracking and monitoring construction equipment. In particular, advancements in computational capacities, namely parallel computing on graphics processing units, and rapid evolution of object detection and tracking methods enable researchers and practitioners to provide semi-real-time information about activities taking place with various construction equipment at a relatively low cost (Azar and Kamat 2017; Khosrowpour et al. 2014b; Kim et al. 2019b; Liu and Golparvar-Fard 2015; Roberts and Golparvar-Fard 2019; Roberts et al. 2018; Tang and Golparvar-Fard 2017). The earliest attempt to develop a vision-based equipment tracking method was carried out in 2007, in which hue, saturation, and value color space were used to isolate an excavator from plain backgrounds and track the equipment (Zou and Kim 2007). After 2010, related research efforts focused on three main areas: object detection, tracking, and activity recognition. An overview of the typical workflow of vision-based equipment monitoring systems is presented in Fig. 4. Also, Table 3 provides a summary of these research projects.

Object Detection

Some of the vision-based systems employed a background subtraction method, such as Gaussian mixture model (GMM) and

Table 2. Overview of kinematic signal processing techniques used for activity detection of construction equipment and tools

References	Machine-learning algorithm	Features	Feature extraction characteristics	Accuracy
Ahn et al. (2012)	—	Signal energy	—	—
Akhavian and Behzadan (2012)	Not a classification model	—	—	—
Akhavian and Behzadan (2013a)	K-means methods in conjunction with data-mining techniques	Data: 1. Position 2. Weight 3. Angle	—	—
Ahn et al. (2013)	1. Naïve Bayes 2. Instance-based learning (IBL): K-nearest neighbor (KNN) 3. Decision tree (J48) 4. Multilayer perceptron	15 time-domain features (average resultant acceleration, mean, standard deviation, peak, and correlation)	128, 256, 512, and 1,024-sample windows (50% overlap)	Over 93%
Akhavian and Behzadan (2015)	1. Logistic regression, 2. K-NN 3. Decision tree 4. Neural network (feedforward back-propagation) 5. Support vector machine (SVM)	42 features 1. Time-domain [mean, variance, peak, interquartile range (IQR), correlation, and root-mean square (RMS)] 2. Frequency-domain features (signal energy)	Windows of 128 data points with 50% overlap (1.28 s)	Overall accuracy of more than 86%
Kim et al. (2018a)	1. Random forest 2. Naïve Bayes 3. J48 4. Sequential minimum optimization (SMO)	Time- and frequency-domain features (total of 74 features)	Window size of 1 s (50% overlap)	91.83%
Rashid and Louis (2019)	Recurrent neural network (RNN)	18 features (i.e., three IMUs × 6 data stream per IMU)	1-s window size (i.e., 76 data points)	Over 96% for fourfold augmentation
Bae et al. (2019)	Dynamic time wrapping	—	Sampling frequency of 10 Hz	Up to 100%

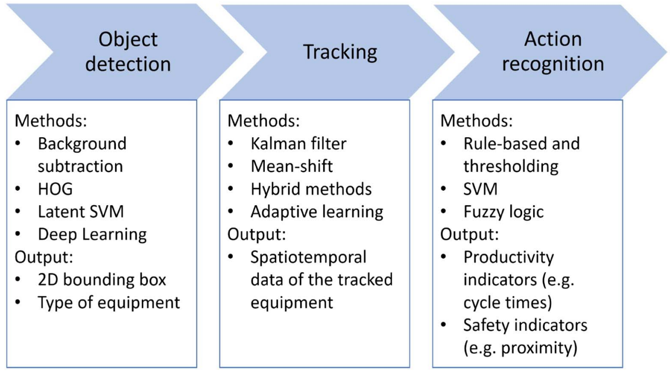

Fig. 4. Typical workflow of vision-based equipment monitoring systems.

Bayesian-based model, to isolate moving objects in the videos captured by a stationary camera (Bügler et al. 2017; Chi and Caldas 2011; Rezazadeh Azar and McCabe 2011; Roberts and Golparvar-Fard 2019; Roberts et al. 2018), and then classified the remaining blobs by other classifiers, such as Bayes or neural networks (Chi and Caldas 2011). Later efforts focused on more sophisticated feature-based recognition methods, such as histogram of oriented gradients (HOG) (Azar and McCabe 2012; Memarzadeh et al. 2013; Rezazadeh Azar and McCabe 2011) and latent support vector machine (SVM) (Tajeen and Zhu 2014), to detect construction equipment in construction images and videos. Later, part-based approaches were proposed to improve detection performance and to estimate the pose of excavators (Soltani et al. 2017; Yuan et al. 2016). Moreover, a method was proposed for pose estimation of excavators using 2D detections in the frames captured by two calibrated cameras (Soltani et al. 2018). These studies listed occlusion, target viewpoints, and visual noises as the main challenges to achieving reliable precision and recall rates. In addition, these vision-based methods were not able to identify individual equipment; thus a marker-based method, using the AprilTag algorithm, was developed to identify labeled equipment in the video frames (Azar 2015).

Advances in the deep learning methods have also affected this field of construction research, and the most recent studies have started using different versions of convolutional neural networks (CNN) and long short-term memory (LSTM) because they were able to outperform older methods (Fang et al. 2018; Kim et al. 2017a, 2019a; Hernandez et al. 2019).

Vision-Based Tracking

The second group of studies aimed at developing reliable methods for tracking construction equipment in which visual occlusion, excessive visual noise, changing orientations, and interclass/intraclass

Table 3. Summary of the vision-based equipment monitoring systems

References	Objectives	Methods used	Action detector	Testing environment (setting)
Gong and Caldas (2011)	Estimate production cycles of a miniloader	Background subtraction and tracking	2D motion characteristics	Real projects
Rezazadeh Azar et al. (2012)	Estimate production cycles of loading activities	HOG object detector and tracking	SVM to analyze proximity and positioning of the equipment	Real projects
Golparvar-Fard et al. (2013)	Estimate production cycles of various earthmoving activities	HOG features	SVM to analyze HOG features and classifying equipment actions	Real projects
Rezazadeh Azar and McCabe (2011)	Estimate production cycles of hauling activities	HOG object detector and tracking	2D motion characteristics	Real projects
Liu and Golparvar-Fard (2015)	Crowdsourcing video-based activity analysis	HOG object detector and tracking. User annotations on activities	Crowdsourced	Real projects
Kim et al. (2015)	Safety assessment through crowdedness and proximity estimation	GMM background subtraction and Kalman filter tracking	Fuzzy inference	Real projects
Bügler et al. (2017)	Estimate production cycles of loading activities	GMM background subtraction and kernel covariance tracking	2D motion characteristics and the 2D threshold for proximity	Real projects
Kim et al. (2017b)	Safety assessment and warning	GMM background subtraction and HOG for detection and Kalman filter tracking	Fuzzy inference	Real projects
Rezazadeh Azar (2017)	Semantic annotation of construction videos	HOG object detection and frame similarity measurement	Bayesian belief network	Videos of real projects
Kim et al. (2018b)	Estimate production cycles of loading activities in tunneling	Region-based fully convolutional networks	2D threshold for proximity	Tunneling projects
Kim et al. (2018c)	Estimate production cycles of loading activities	Tracking-learning-detection (TLD) algorithm	2D threshold for proximity	Real projects
Kim et al. (2019a)	Proximity monitoring between mobile resources	CNN object recognition	Distance measurement in rectified images	UAV-captured images in real projects
Kim et al. (2019b)	Estimate production cycles of hauling activities	License plate detection and recognition (LPDR), and deep convolutional network	Rule-based reasoning for entrance and exit of dump trucks	Real projects
Roberts and Golparvar-Fard (2018, 2019)	Estimate production cycles of hauling activities	HMM, atomic action recognition, deep learning–based detection and tracking	2D spatiotemporal features	Real projects
Kim and Chi (2019)	Estimate productivity and cycle time of earthmoving operations	Faster recursive CNN (R-CNN) and TLD	Sequential pattern analysis	Real projects

variations were considered as the primary challenges. Various methods, such as mean-shift and Kalman filter (Gong and Caldas 2011), counter-based and point-based algorithms (Park et al. 2011b), kernel covariance (Teizer 2015), and particle filtering (Zhu et al. 2016) were evaluated for tracking construction equipment. In addition to the single-view tracking, some research studies focused on tracking in stereo vision (3D), in which the epipolar geometry enables 3D localization of the objects of interest (Brilakis et al. 2011; Park et al. 2011a).

Some research efforts developed hybrid tracking algorithms, using HOG and Karhunen-Loeve transform (KLT) (Rezazadeh Azar et al. 2012) and latent SVM and particle filtering (Zhu et al. 2017) to tackle occlusion and interclass/intraclass variations. A recent study integrated random ferns for key-point recognition, and a median-flow and pyramidal optical flow algorithms with a real-time online learning ability to achieve a high recall/precision tracking performance (Kim and Chi 2017).

Activity Recognition

Object detection and tracking modules provide spatiotemporal data of the objects of interest; i.e., construction equipment, which could be further analyzed for activity recognition. Research efforts in this area can be classified into two main groups: activity recognition for productivity estimation and safety monitoring. In particular, monitoring the process of loading dump trucks has been the subject of numerous studies because interactive operations are suitable candidates to investigate vision-based capabilities. SVM was used to analyze the interactions of the detected excavator and dump trucks in video frames (Rezazadeh Azar et al. 2012). Another system also used SVM to analyze HOG descriptors to classify the actions of an excavator (Golparvar-Fard et al. 2013). Other studies used hard-coded thresholds to assess the proximity of the tracked excavator and dump truck for activity recognition (Bügler et al. 2017; Kim et al. 2018b, c). There were also some efforts to recognize single-equipment operations, such as hauling by trucks (Azar and McCabe 2013; Golparvar-Fard et al. 2013) and earth movement by a mini-loader (Gong and Caldas 2011). These systems mainly relied on 2D motion descriptors for activity recognition and analysis. Another recent study takes advantage of a hidden Markov model (HMM) together with deep learning–based detection and tracking techniques to automatically label a sequence of activities in a given video (Khosrowpour et al. 2014b; Roberts and Golparvar-Fard 2019). These methods show good performance with earthmoving operations as well as drywall construction.

The second group of research projects focused on safety analysis, in which they mainly assessed the proximity of tracked equipment and workers. For example, fuzzy inference was used to determine proximity and crowdedness as safety indicators, in construction videos (Kim et al. 2015, 2017b), and an algorithm was proposed to monitor proximity in the unmanned aerial vehicle (UAV)-captured images (Kim et al. 2019a; Lin and Golparvar Fard 2016).

In addition to the productivity measurement and safety assessment, activity recognition could be used for semantic annotation of construction videos (Rezazadeh Azar 2017).

Audio-Based Methods for Activity Detection of Construction Equipment and Workers

A construction jobsite is generally a noisy environment, and the main source of the noise is the working sounds of heavy equipment and tools in the workplace. These noises can provide useful information indicating ongoing operations, processes, and site conditions. To accomplish the potential of the audio-based equipment activity detection, diverse studies employed a microphone for collecting audio data, signal processing for cleaning sound data and extracting necessary features, and machine or deep learning techniques for detecting and classifying an activity type. Sound identification generally entails following four primary processes: signal analysis, feature extraction, model training, and model testing (Gaikwad et al. 2010). This section summarizes the methodology and the outcomes of the recent studies regarding the applications of audio-based methods for recognizing and monitoring activities of tools and machines at a construction jobsite. An overview of a general framework for an audio-based activity detection system is presented in Fig. 5.

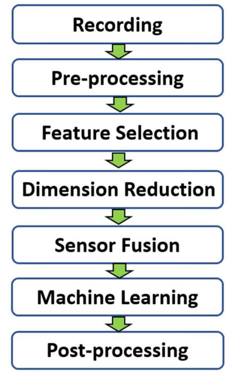

Fig. 5. Overview of the audio-based methodology for activity recognition.

Signal Analysis and Feature Extraction

Diverse types of audio features have been examined for enhancing the accuracy of audio signal classification. Two studies (Lu et al. 2002; Patsis and Verhelst 2008) extracted a total of 62 features, grouped into 15 distinct sets of features including both time and frequency characteristics. Each feature is obtained by segmenting the superframe into smaller frames and evaluating the needed values. Each feature uses a frame of different lengths and overlap. Audio-based event detection frameworks identify descriptors for enhancing accuracy by representing different audio signal domains. For example, characteristics such as the root-mean square (RMS) can be extracted in the time domain, but the variance of the spectral flux (VSFLUX) pertains to the frequency domain. The correct usage of different sets of descriptors would depend on the sound classification process and purpose.

Model Training and Sound Classification

To conduct sound classification, previous studies investigated diverse classifiers: KNN, GMM, HMM, artificial neural networks, SVMs, and deep neural networks (neural-network architectures with several hidden layers) (Gencoglu et al. 2014). To accomplish satisfactory performance and reliability, these classifiers that support distinct features are selected pertaining to sound data types (Sharan and Moir 2016). The machine learning–based classifiers include Bayesian and naïve Bayesian networks (John and Langley 1995), the Hoeffding tree (Hulten et al. 2001), the decision table (Kohavi 1995), the decision tree (Quinlan 2014), the random tree (Rokach and Maimon 2008), the random forest (Breiman 2001), the multilayer perceptron (MLP) (Haykin 2009), sequential minimum optimization (SMO) (Platt 1998), K-nearest neighbors (KNN) (Aha et al. 1991; Altman 1992), part-based decision (Frank et al. 2002), and linear logistic regression (Sumner et al. 2005).

Some studies have applied various algorithms such as SVM and HMM to test and evaluate the audio-based classification of the activity types related to construction operational equipment (Cheng et al. 2016, 2017a; Cho et al. 2017). Cheng et al. (2016, 2017a)

J. Constr. Eng. Manage., 2020, 146(6): 03120002

proposed an approach to recognize two types of activities (value-adding and non-value-adding) using audio data for single machine scenarios. Their proposed method consists of four steps: (1) recording audio data; (2) filtering noises using a signal enhancement algorithm; (3) using short time Fourier transform (STFT) to extract features; and (4) training a machine-learning algorithm using extracted features (STFT magnitude values). Hanning windows at size 512 with 50% overlap and 1,024-point discrete Fourier transform (DFT) are chosen for time bins. SVM has been utilized for binary classification of activities. The radial basic function has been used as the kernel function in their SVM model. They evaluated this method on four types of equipment and obtained over 90% accuracy. In this method, they used a recorder (DR-05 2 GB, Tascam, TEAC Corporation, Montebello, CA) to record audio from the equipment.

Furthermore, Cheng et al. (2017b) analyzed the acoustical modeling of construction jobsites to compare different hardware and software arrangements and assess their impact on the results. They selected three types of microphones: (1) an off-the-shelf microphone; (2) a contact microphone; and (3) a multichannel microphone array. Also, two different settings were chosen. In the first setting, microphones were placed in the equipment cabin. In the second one, microphones were installed on the jobsite near the equipment. For monitoring operations of three types of equipment that can damage the underground cables, Yang et al. (2015) adopted linear prediction cepstral coefficients (LPCC), which train a SVM model utilizing audio signals recorded by a microphone array. As a follow-up study for enhancing the accuracy of sound classification proposed in the aforementioned study, Cao et al. (2017b) implemented extreme machine learning (ELM) using spectral dynamic features to detect activities of four types of equipment. They used background-noise-reduction algorithms to decrease the noises of tunnel construction sites.

A construction jobsite with a complex and dynamic working environment encompasses several types of simultaneously occurring equipment work. Thus, several audio sources can be generated during construction processes. Cheng et al. (2018) improved their algorithm to be used for multiple machines. Sabillon et al. (2017) also used audio to detect activities and utilized a Bayesian approach to predict the cycle times of construction equipment. In this paper, only a multichannel microphone array has been used, and the frequency magnitude and phase features were extracted to train the SVM model. Another extension of that work is an estimated cycle time for multiple days so as to generalize their results, which were reported to have an accuracy rate as high as 90%. In addition, Cho et al. (2017) proposed a frequency-domain-based approach to classify sound patterns on three different construction operations: concrete pouring, concrete grinding, and hammering. They also showed how to visualize collected sounds data through a building information modeling (BIM) model.

Zhang et al. (2018) also implemented a supervised machine-learning algorithm to improve the performance of construction sound detection systems. Even though they should be detected and then analyzed separately for further activity detection, it is still somewhat limited to collect different sound data simultaneously and classify the work types. To enhance the accuracy of sound classification of multiple audio sources, it would be a currently promising way to adopt a microphone array that can estimate the direction and distance of sound sources in 2D or 3D spaces (Jiang et al. 2011). This approach calculates a geometrical relationship and estimated time delays among installed microphones to obtain directions and distances of multiple sound sources in a site.

Table 4 provides overviews of existing studies regarding applications of utilizing sound to recognize construction operations. This table summarizes the main objectives, number and types of equipment studied, number of activities covered, devices used to record audio, and the testbed settings. Further analyzing and computing details about each study are illustrated in Table 5.

Table 4. Basic components of different audio-based activity recognition systems for construction workers

References	Objectives	Equipment quantity (single or multiple)	Tools	Microphone placement
Yang et al. (2015)	Activity detection	3 types of equipment (single machine)	Cross microphone sensor array with 8 microphones	Installed on the jobsite (distances of 3–72 m)
Cheng et al. (2016)	Activity detection	4 types of equipment (single machine)	Recorder (Tascam DR-05 2 GB)	Installed on the jobsite
Cheng et al. (2017a)	Activity detection	1 type of equipment (single machine)	Recorder (Tascam DR-05 2 GB)	Installed on the jobsite
Cheng et al. (2017b)	1. Hardware and software requirements 2. Activity detection	11 types of equipment (single machine)	1. Off-the-shelf microphone (Zoom H1 digital handy recorder) 2. Korg CM-200 clip-on contact microphone 3. Multichannel microphone array (xCORE-200)	1. Microphones mounted on board 2. Installed on the jobsite
Cheng et al. (2018)	1. Hardware and software requirements 2. Activity detection	Equipment (multiple machines)	1. Off-the-shelf microphone (Zoom H1 digital handy recorder) 2. Korg CM-200 clip-on contact microphone 3. Multichannel microphone array (xCORE-200)	1. Microphones mounted on board 2. Installed on the jobsite
Cao et al. (2015)	Activity detection	4 types of equipment (multiple machines)	Cross-microphone sensor array with 4 microphones	Installed on the jobsite (distances of 5–60 m)
Cao et al. (2017a)	Activity detection	4 types of equipment (multiple machines)	Four-element cross-layer MEMS microphone sensor	Installed on the jobsite (distances of 3–230 m)
Cao et al. (2017b)	Activity detection	4 types of equipment (multiple machines)	Cross-layer microphone sensor array with 8 channels	Installed on the jobsite (distances of 5–150 m)
Sabillon et al. (2017)	1. Activity detection 2. Cycle time estimation	Equipment (multiple machines)	XMOS xCORE-200 multichannel array microphone	Installed on the jobsite

Table 5. Overview of audio signal processing techniques used for activity detection of construction workers and tools

References	Machine-learning algorithm	Features	Feature extraction characteristics	Accuracy
Yang et al. (2015)	SVM	Linear prediction cepstral coefficients (LPCC)	Frame length 175 ms (3,500 samples)	Higher to 98%
Cheng et al. (2016)	SVM	STFT magnitudes	Hanning windows size 512 1,024-point DFT 50% overlap	Over 90%
Cheng et al. (2017a)	SVM	STFT magnitudes	Hanning windows size 512 1,024-point DFT 50% overlap	Exceeds 85%
Cheng et al. (2017b)	SVM	Time-frequency features	Hanning windows size 512 1,024-point DFT 50% overlap	Over 80% even 85% accuracy
Cheng et al. (2018)	SVM	STFT magnitudes	Hanning windows size 512 1,024-point DFT 50% overlap	Over 85%
Cao et al. (2015)	Single hidden-layer feedforward neural network and extreme learning machine (ELM)	Spectral dynamic features	Hanning windows size 256 and 4,096 (50% overlap)	Up to 88%
Cao et al. (2017a)	Algorithm developed in LabView and MATLAB	Short frame energy ratio, the concentration of spectrum amplitude ratio, truncated energy range, and interval of the pulse	Hanning windows size 256 (one-sample shift for overlap)	Over 86%
Cao et al. (2017b)	Artificial neural network using the ELM approach	mel-frequency cepstral coefficients (MFCCs) features	Hamming windows size 495 points (50% overlap)	Up to 96%
Sabillon et al. (2017)	SVM	Frequency magnitude and phase features	Hanning windows size 512 1,024-point DFT 50% overlap	As high as 90%

Kinematic-Based Methods for Activity Detection of Construction Workers

Human activity recognition (HAR) is a time-series classification problem with a variety of applications in different domains including robotics, security, rehabilitation, pervasive computing, and entertainment (Ren-Jye et al. 2018; Ranasinghe et al. 2016; Lara and Labrador 2013). The rapid development of body-worn sensors and mobile devices in the last decade has created new opportunities for improving construction processes through ubiquitous and accurate detection of construction workers' activities. Among many applications, detecting and tracking workers' activities can be useful for safety monitoring, productivity measurement, ergonomic assessment, and quality control. Similar to construction machinery and heavy equipment, worker activities that take place on a construction jobsite generate distinctive kinematic signals (e.g., body acceleration, angular movement, and posture). Tables 6 and 7 provide an overview of recent studies regarding potential applications of kinematic signals for recognizing and monitoring activities of construction workers and equipment.

Image/Video-Based Methods for Activity Detection of Construction Workers

Similar to applications utilizing computer-vision methods for recognizing construction machinery activities, it is possible to provide information about worker activities at construction jobsites by processing captured images and video files. Computer-vision algorithms eliminate the need for attaching sensors to workers' bodies, which is a major advantage over kinematic-based methods.

Several studies have been conducted to detect human activities such as walking, standing, sitting, and running. The same concept could be utilized for detecting workers' activities and gestures at construction jobsites. Table 8 summarizes recent studies on worker activity detection using images and videos.

Results and Discussions

As indicated previously, different automated activity recognition techniques each have advantages and limitations and several factors need to be considered when choosing an optimum solution for a particular jobsite. A summary of these methods' advantages and limitations is presented in Table 9.

Other significant findings of this study are as follows:

- For audio- and kinematic-based methods, researchers have implemented different types of machine-learning algorithms to detect construction activities. SVM, ANN, decision trees, KNN, logistic regression, naïve Bayes, and random forest have been used frequently in this regard. Also, in recent years, neural networks and deep learning have been used to recognize construction activities. Recurrent neural network (RNN) is a deep learning model that utilizes order dependence in sequence prediction problems such as equipment activity recognition and tracking.
- Selecting proper features for training purposes is another important aspect of developing an automated activity detection system. Time-domain features (e.g., mean, median, peak, variance, and standard deviation) and frequency-domain features (e.g., fast Fourier transform (FFT) coefficients, energy, and entropy) are frequently used by kinematic- and audio-based methods. However, CNN is a deep learning architecture capable of automatically extracting features. Considering the limited scope of applications for each sensing technology, using hybrid approaches and fusing results would be a good strategy to overcome limitations of individual techniques. In a recent study, both audio and kinematic data were fused, and results showed an improved accuracy (Sherafat et al. 2019a, b).
- For computer-vision-based methods, it was reported that using motion boundary histogram (MBH), higher accuracies can be derived rather than using HOG and histograms of optical flow (HOF) (Yang et al. 2016). It was shown that a HOG descriptor leads to better results than a HOF descriptor (Gong et al. 2011).

J. Constr. Eng. Manage., 2020, 146(6): 03120002

Table 6. Basic components of different kinematic-based activity recognition systems for construction workers

References	Objectives	Number of workers	Number of classifications (type of activities)	Tools (sampling frequency)	Testing Environment (setting)
Joshua and Varghese (2010b)	Preliminary study on automation of activity sampling method	1	5 activities (masonry: fetch brick, twist laying, fetch mortar, spread mortar, and cut brick)	Single wired triaxial accelerometer [Motion node (GLI Interactive)] (60 Hz) Laptop with a built-in camera	Lab (waist)
Joshua and Varghese (2010a)	Automating work-sampling process	1	5 activities (masonry: fetch brick, twist laying, fetch mortar, spread mortar, and cut brick)	Single wired triaxial accelerometer [Motion node (GLI Interactive)] (60 Hz) Laptop with a built-in camera	Instructed and uninstructed modes (left and right sides of the waist)
Cezar (2012)	Discriminating between different activities during regular workdays	3	4 activities (hammering, sawing, sweeping, and drilling)	Smartwatch (25 Hz) Accelerometer and gyroscope data	Worker's dominant hand
Khan and Sohail (2013)	Activity recognition using a single accelerometer system	44	9 activities	Sparkfun IMU 3,000 fusion board (ADXL345 accelerometer and MotionProcessor gyroscope)	Naturalistic lab environment (waist)
Joshua and Varghese (2013)	Evaluating the location of accelerometers on a worker for activity classification	4	11 activities (bricklaying)	Triaxial accelerometer data loggers with range of $\pm 6g$ (40 Hz)	Two armbands and to-abdominal belt
Yang et al. (2014)	Detecting near-miss fall incidents based upon IMU	2	5 activities (walking, squatting, standing up, squatting down, and static standing)	IMU sensor [SHIMMER 9 degree of freedom (DOF), Shimmer] (51.2 Hz)	Laboratory experiment (waist)
Lim et al. (2015)	1. Collect data of a worker's limbs 2. Identify the occurrence and type of near-miss event 3. Detect objects causing unsafe conditions 4. Take corrective actions	3	10 motions (activating sensor, pause, walk, slip, walk, trip, walk, pause, deactivating sensor, and return)	Three-axis acceleration of a smartphone (80 Hz)	Simulated construction jobsite (left hip pocket in a vertical position)
Akhavian and Behzadan (2016)	Designing and testing a construction activity recognition system	—	5 activities (hammering, sawing, loading sections, turning a wrench, into wheelbarrows, pushing loaded wheelbarrows, dumping sections from wheelbarrows, and returning with empty wheelbarrows)	All 3 axes (X, Y, and Z) from accelerometer and gyroscope (100 Hz)	Outdoor workspace (upper arm)
Zhang et al. (2018)	Propose an activity recognition method using a smartphone	9	8 activities (main activities consisting of standing, walking, squatting, cleaning up the template, fetching andplacing rebar, locating the rebar, banding the rebar, and placing)	2 smartphones (Invensense_MP67B) (10 Hz)	Simulated floor-reinforcing steelwork (right wrist and upper right leg)
Nath et.al (2018)	Assessment of ergonomic risks by detecting construction activities	—	Category 0: None Category 1: Lift/lower/carry Category 2: Push/pull	Accelerometer, linear accelerometer, and gyroscope sensors	Field experiment that resembles real-world activities
Akhavian and Behzadan (2018)	Extract duration of activities for input modeling of discrete event simulation (DES)	—	7 activities (subactivities of preparing, transporting, and installing)	Smartphone accelerometer and gyroscope (100 Hz)	Outdoor environment that resembles a small construction jobsite
Ryu et al. (2018)	Masonry action recognition using a wristband	—	4 activities (spreading mortar, bring and laying blocks, adjusting blocks, and removing remaining mortar)	Accelerometer embedded in an eZ430-Chronos sports watch (22 Hz)	Indoor masonry work (wrist)
Yang et al. (2019)	Evaluate steelworkers' workloads	—	8 activities (standing, walking, squatting, cleaning a template, placing rebar, lashing rebar, welding rebar, and cutting rebar)	Smartphone accelerometer and gyroscope (10 Hz)	Laboratory (wrist and leg)
Lee et al. (2019)	Channel state information (CSI)-based human activity recognition	—	4 activities (walking, eating, typing, and no-activity)	AC1750 MU-MIMO gigabit router (Linksys) was used for the access point (AP) and a Lenovo T400 laptop for the receiver	Wood-frame apartment and a reinforced concrete-frame apartment

J. Constr. Eng. Manage., 2020, 146(6): 03120002

Table 7. Overview of kinematic signal processing techniques used for activity detection of construction workers and tools

References	Machine-learning algorithm	Features	Feature extraction characteristics	Accuracy
Joshua and Varghese (2010b)	—	Time-domain features (arithmetic mean, median, mean absolute deviation, standard deviation, and variance)	Data segment lengths of 2, 4, and 4.23 s (50% overlap)	—
Joshua and Varghese (2010a)	1. Naïve Bayes 2. Decision trees (J48) 3. Multilayer perceptron	Basic time-domain features (peak, mean, variance- correlation, and energy)	Data segment lengths of 2, 4, and 4.23 s (50% overlap)	80%
Cezar (2012)	1. Naive Bayes (NB) 2. Multinomial logistic regression (MLR) 3. Multi-SVM 4. Linear discriminant analysis (LDA) 5. Quadratic discriminant analysis (QDA)	46 time and frequency features	Window size of 40 samples (1.6 s)	91%
Khan and Sohail (2013)	17 classifiers	Arithmetic mean, correlation, standard deviation, fast Fourier transform energy, autoregressive coefficients, data entropy, FFT entropy, discrete cosine transform (DCT) energy, and first four dominating FFT coefficients along three axes of the accelerometer	A window of 512 samples, i.e., 5.12 s	94%
Joshua and Varghese (2013)	Decision tree	1. 33 time-domain features (mean, median, peak, mean absolute deviation, variance, first-quartile, and third-quartile to measure central tendency and variability, and correlation) 2. Frequency-domain features (DC component, entropy, and energy)	Segments of 6.4 s (256 samples)	77%
Yang et al. (2014)	1. Support vector machine 2. One-class support vector machine	38 features:mean, standard deviation, peak, correlation, spectral entropy, and spectral centroid (for both the accelerometer and the gyroscope)	50% overlap	91.10%
Lim et al. (2015)	ANN	Single vector magnitude feature (mean, standard deviation, and peak)	Windows size of 3 s (50% overlap)	94%
Akhavian and Behzadan (2016)	1. Artificial neural network (ANN) 2. Decision tree 3. KNN 4. Logistic regression 5. SVM	Mean, maximum, minimum, variance, RMS, IQR, and the correlation between every two pairs of axes and spectral energy and entropy	Windows size of 1.28 s (50% overlap)	Around 80% in the best case
Zhang et al. (2018)	CART algorithm of a decision tree	Five common time-domain features, namely mean, standard deviation, IQR, and skewness	Time window of 6.4 s (50% overlap)	Up to 94.91%
Nath et al. (2018)	SVM	Statistical features from each sensor (mean, minimum, maximum, standard deviation, IQR, skewness, kurtosis, mean absolute deviation, and fourth-order autoregressive coefficients)	1 s (180 data points), 2 s (360 data points), and 3 s (540 data points) with a 50% overlap	Up to 90.2%
Akhavian and Behzadan (2018)	1. ANN 2. Decision tree 3. KNN 4. Logistic regression 5. SVM	Statistical features such as mean, minimum, maximum (i.e., time-domain features), and signal entropy and energy (i.e., frequency domain)	Windows size of 1.28 s (50% overlap)	—
Ryu et al. (2018)	1. KNN 2. Multilayer perceptron 3. Decision tree 4. Multiclass SVM	1. 8 time-domain features (mean, standard deviation, maximum, minimum, range, skewness, kurtosis, and correlation) 2. 2 frequency-domain features (energy and entropy)	Window size of 4 s (50% overlap)	Up to 88.1%
Yang et al. (2019)	SVM	Three-axis acceleration, three-axis angular rotation, mean, variance, standard deviation, maximum, minimum, range, RMS, and correlation	—	Up to 99.25%
Lee et al. (2019)	SVM and earthmover's distance (EMD)	6 features of CSI amplitudes (average, standard deviation, 25th percentile, 75th percentile, maximum, and median absolute deviation)	Sampling frequency of 10 packets per second	Up to 94.38%

Table 8. Overview of kinematic signal processing techniques used for activity detection of construction workers and tools

References	Objectives	Methods used	Action detector	Testing environment (setting)
Gong and Caldas (2009)	Developing a model to generate productivity information from construction operations videos	Haar waveletlike simple features (edge and line features)	Intel Open Resource Computer Vision	Stadium construction jobsite
Gong et al. (2011)	Classifying worker actions	Bag-of-video-feature-words and Bayesian learning methods with the use of HOG and HOF indicators	Naïve Bayesian classifier and probabilistic latent semantic analysis (pLSA)	Real projects
Han et al. (2013)	Motion analysis of construction operations using Kinect motion capture data	Kernel principal component analysis (Kernel PCA)	Dynamic time warping (DTW)	Lab experiments
Khosrowpour et al. (2014a)	Analysis of worker activities using RGB+depth sensors	Kernel density estimation (KDE) model and GMM	SVM and HMM	Lab settings and actual construction site
Khosrowpour et al. (2014b)	Analysis of interior construction operations using RGB+depth sensors	KDE model	SVM and HMM	Lab settings and actual construction site
Liu and Golparvar-Fard (2015)	Crowdsourcing construction activity analysis	HOG and HOC features	SVM	Real-world jobsites
Yang et al. (2016)	Worker action recognition using dense trajectories method	HOG, HOF, and motion boundary histogram (MBH)	Multiclass SVM with nonlinear radial basis function (RBF) kernel	Real projects
Luo et al. (2018)	Monitor and analysis of installing reinforcement activities in construction	RGB, optical flow, and gray stream	Improved CNN	Real projects

Table 9. Qualitative comparison of construction activity detection methods

Method	Advantages	Disadvantages
Kinematic-based methods	1. Using RFID tags is a beneficial factor for large-sized construction because these tags are capable of communicating for long distances up to 100 m. 2. They also work without the need for lines of sight. In contrast to vision-based tools that need a line of sight for recording images or videos, these tags can operate easily without any lines of sight. 3. Accelerometers have reasonable accuracy and power consumption (Hanna 2001). 4. In contrast with image sensors, accelerometers are robust and resilient in trying conditions (Joshua and Varghese 2010a).	1. They require a reasoning mechanism to detect the location of tagged construction items (Torrent and Caldas 2009). 2. RFID tags can be easily jammed or disrupted due to the presence of other wireless fidelity (Wi-Fi) networks. 3. RFID reader might collide when the signals from several readers overlap with each other. 4. RFID tags might collide when several tags are present in a congested area. 5. MEMS devices need to be directly attached to the equipment or worker, which are not applicable to some types of tools. They also might either hinder or limit workers' movement, which is intrusive for operators. 6. There would be a need for many of these devices in the jobsite, which makes this method costly.
Vision-based methods	1. They can provide semi-real-time information (Cheng et al. 2016). 2. Recorded images/videos can be stored as reliable documentation and be used for future needs (Cheng et al. 2016).	1. Because these devices need light to record images/videos, they are very sensitive to environmental factors such as illumination, dust, snow, rain, and fog. They can perform only when there is neither darkness nor direct sunlight (except for thermal cameras). 2. A network of cameras is required to cover the whole jobsite (Cheng et al. 2017a); 3. There would need to be an absence of any obstacles between the camera and equipment/workers. 4. Crowded and congested jobsites with high noise levels (e.g., moving backgrounds and varying light conditions) are hard to analyze using these methods (Hanna 2001; Akhavian and Behzadan 2014). 5. These methods require large storage sizes to hold the image and video data. Also, these methods are relatively more computationally intensive than other options (Akhavian and Behzadan 2013b). 6. Privacy issues might prevent these devices from being utilized in the jobsite. Some workers might be uncomfortable being continuously monitored. 7. These devices are relatively expensive to be placed on the jobsite.

Table 9. (*Continued.*)

Method	Advantages	Disadvantages
Audio-based methods	1. In construction jobsites, some activities generate sounds, making them recognizable by audio sensors (such as drilling and excavating). 2. There is no need for microphones to be placed on the equipment cabin or be attached to the workers' clothing. 3. An equipment operator's skill in controlling and moving the equipment does not significantly affect the sound generated by the equipment. 4. A microphone is capable of recording sounds in 360°. 5. In contrast to vision-based methods, the presence of obstacles does not affect the quality of recorded data. 6. In contrast with vision-based methods, there is less of a need for computational time and storage space. 7. One microphone is enough for a relatively large area. 8. In contrast with kinematic-based methods, it can record the audio of multiple machines.	1. They are not suitable for noisy and crowded jobsites. However, some denoising algorithms exist, which can decrease the noise effects, but noise somehow affects the accuracy of activity detection. 2. Audio-based methods are not applicable to some types of equipment such as tower cranes, which don't generate sounds.

Also, implementing CNN and crowdsourcing labeling techniques have been recommended to improve results (Kim et al. 2018c). MBH feature descriptor is capable of providing higher accuracy compared with HOG and HOF (Yang et al. 2016). A similar study showed that a HOG descriptor leads to better results compared with a HOF descriptor (Gong et al. 2011).

Evaluation of the Effectiveness of Different Methods

Audio-based, kinematic-based, and vision-based methods detect and recognize different types of activities on the jobsite. Although beneficial, the list of activities these methods generate over time is not sufficient for project managers to acquire an in-depth understanding of the status of the entire project, take necessary preventive and corrective actions, and perform cost and schedule analysis. Subsequently, further processing will be required to convert the raw chronological list of activities into an effective performance monitoring and control system. The following two items are deemed as necessary components of a reliable construction equipment performance monitoring system: (1) calculating value-adding versus non-value adding and active versus idle times; and (2) estimating quantities of accomplished work.

Although a chronological list of activities, as well as productivity rates and quantities of completed work by a worker or a piece of equipment, can provide very useful information for jobsite personnel, it does not suffice for necessary analysis and decision-making tasks at higher management levels. Project managers and company owners need to monitor and evaluate the performance of their projects (and thus all included machines and workers) as a whole. For successful implementation of each of these methods on the jobsite, five major performance metrics must be considered that include (1) range of activities it can cover, (2) appropriate temporal, special, and weather conditions, (3) required devices (estimated total cost), (4) computational time, and (5) accuracy.

Table 10 provides a comparison of implementation factors for these three methods, based on all the papers studied here. In this table, the estimated cost is calculated based on the number of workers and machines on the jobsite (small-sized or big-sized jobsite), the number of devices, and the cost of each device. For example, for a small-sized jobsite with five workers and five machines, 10 basic accelerometers (each costing about \$10) are required, which leads to a total cost of \$100. The level of accuracy column shows the range of accuracies reported by the previous papers. The computational time shows how fast the method can run.

In Table 11, different jobsite conditions are considered for different methods, and the letter X indicates that the method is practical in that specific condition.

In Table 12, different methods are compared based on their capability of recognizing activities and subactivities (actions). As previously mentioned, equipment/worker activities are ongoing processes with respect to time, whereas actions are single efforts. When actions are grouped together they form an activity. Some methods are limited due to their inability to differentiate between different actions.

Table 10. Comparison of implementation factors for construction activity detection methods

Method	Devices [estimated system cost (\$)]	Computational time	Level of accuracy
Kinematic-based methods	Accelerometer, gyroscope, smart watch, or mobile phone (small-sized jobsite: \$100–\$1,000; medium-sized job-site: \$1,000–\$10,000)	Fast (real-time)	High[a]
Vision-based methods	Optical cameras (small-sized jobsite: \$1,000–\$10,000; medium-sized jobsite: \$10,000–\$100,000)	Moderate (semi-real-time) and fast (real-time)	Rainy or windy (low[b] and moderate[c]), and normal weather (moderate and high)
Audio-based methods	Single microphone or microphone array (small-sized jobsite: \$300–\$3,000; medium-sized jobsite: \$3,000–\$30,000)	Fast (real-time)	Very noisy (low), noisy (moderate), and not noisy (high)

[a]Moderate = 60%–80%.
[b]Low = <60%.
[c]High = 80%–100%.

Table 11. Comparison of jobsite conditions factors for construction activity detection methods

Conditions	Kinematic-based methods	Vision-based methods	Audio-based methods
Day	×	×	×
Night	×	—	×
Rainy	×	—	—
Snowy	×	—	×
Cloudy	×	—	×
Windy	×	×	—
Small-sized jobsite	×	×	×
Medium-sized jobsite	×	×	×
Large-sized jobsite	—	×	—
Indoor environment	×	×	×
Outdoor environment	×	×	×
Noisy jobsite	×	×	—
Congested jobsite	×	—	×
Single equipment or worker	×	×	×
Multiple equipment or workers	×	×	×
Need for several devices	×	×	×
Need for attachment	×	—	—

Also, based on the papers reviewed, these methods have better accuracies for specific types of equipment and activities. Table 13 lists worker activities and equipment types that are recognizable or unrecognizable by each method.

Conclusions and Future Research Directions

This paper contributes to the body of knowledge by providing a concise overview of recent studies about automated activity recognition of construction resources (equipment and workers). The existing methods fall into three major categories: audio-based, kinematic-based, and computer vision–based techniques. The advantages and limitations of each method have been extensively discussed in the previous sections.

For future research directions in this area, four important points should be taken into account by researchers:

- The majority of the current studies are at the proof-of-concept stage and utilize low or little ground-truth data. These studies are often based on a limited number of data sets (fewer than 10 types of equipment/single worker type) and only partially cover complex construction jobsites. The number of implemented data-capturing sensors (e.g., microphones or digital cameras) is also limited. Providing a data set of images and videos

Table 12. Ability to recognize activities and actions of equipment and worker for each method

Method	Equipment activities	Equipment actions	Worker activities	Worker actions
Kinematic-based methods	Engine off, idling, excavating, loading, and hauling, and among others	Moving forward/backward and lowering/raising the arm, and among others	Idling, walking, hammering, drilling, sawing, spreading mortar, and bricklaying, and among others	Raising/lowering hand and looking right/left
Vision-based methods	Engine off, idling, excavating, loading, and hauling, and among others	Moving forward/backward and lowering/raising the arm, and among others	Idling, walking, hammering, drilling, sawing, spreading mortar, and bricklaying, and among others	Raising/lowering hand and looking right/left
Audio-based methods	Engine off, idling, excavating, loading, and hauling, and among others	Moving forward/backward and lowering/raising the arm, and among others	Hammering, drilling, and sawing, and among others	Incapable

Table 13. Worker activities and equipment types that are recognizable or unrecognizable for each method

Method	Recognizable equipment	Unrecognizable equipment	Recognizable worker activities	Unrecognizable worker activities
Kinematic-based methods	Any heavy equipment with articulated moving parts such as excavator, loader, vibrator, and dump truck	Any heavy equipment with no or minimal articulated moving parts such as paver, forklift, and scissor lift	Any activity involving the considerable movement of human body parts and repetitive in nature such as working with carpentry tools, climbing a ladder, and tying rebar	Any activity that does not involve the considerable movement of human body parts or repetitive in nature such as welding
Vision-based methods	Any type of equipment assuming that certain environmental conditions exist	May not work in dark/occluded or extremely dusty environments	Any scenarios where there are certain visually recognizable postures/features for workers. Interaction of workers with surrounding material/tools helps with inference of the actions	Any activity that does not include visually distinguishable features
Audio-based methods	Any equipment that generates sound: bulldozer, backhoe, excavator, loader, compactor, concrete truck, vibrator, dozer, jackhammer, compactor, and dump truck	Any equipment that does not generate sound: paver, forklift, scissor lift, boom lift, tower crane, tele handler, skid steer, grader, and scraper	Any worker activity that generates sound: hammering, welding, sawing, and drilling	Any worker activity that does not generate sound: masonry, bolting, steel work, bricklaying, walking, and standing

requires a large server to store the data, which in part can prevent the usage of vision-based methods for most projects. Larger-scale experiments that cover the entire construction jobsite with multiple workers and machines operating simultaneously would help to better understand the real added value of these methods in practice.

- In line with the previous comment, the current methods for activity recognition of construction resources are not yet at the commercialization and technology-transfer level. There are several location tracking applications, such as Fleet and Equipment Manager, that allow for tracking equipment locations. However, to the authors' knowledge, there is no commercialized application capable of detecting and recognizing equipment activities and providing useful performance measurements to the construction managers. In other words, considering Fig. 1, the first level of the automated construction monitoring system (i.e., spatial location tracking) has been commercialized, but the second level (i.e., activity recognition) still requires more precise and generalized methods to convert to a commercialized application.
- Current studies are based on implementing a single method for automated activity detection of workers and machines. Although effective, a single method cannot overcome all challenges it may face in a complex construction jobsite and provide promising output for all possible scenarios. As a result, developing hybrid methods using fusion techniques that are capable of implementing multiple methods and combining results for improving the overall performance would be extremely useful.
- The construction industry is rapidly changing and moving from jobsites to factories where a repetitive and under-control environment could be achieved. Implementing activity detection and monitoring systems discussed in this paper will generate more reliable results in such environments.

Data Availability Statement

Data generated or analyzed during the study are available from the corresponding author by request.

Acknowledgments

This research project has been funded by the US National Science Foundation (NSF) under Grant Nos. CMMI-1606034, CMMI-1800957, and CMMI-1818534. The authors gratefully acknowledge the NSF's support. Any opinions, findings, conclusions, and recommendations expressed in this paper are those of the authors and do not reflect the views of the funding agency.

References

Aha, D. W., D. Kibler, and M. K. Albert. 1991. "Instance-based learning algorithms." *Mach. Learn.* 6 (1): 37–66. https://doi.org/10.1007/BF00153759.

Ahn, C. R., S. Lee, and F. Peña-Mora. 2012. "Monitoring system for operational efficiency and environmental performance of construction operations using vibration signal analysis." In *Proc., Construction Research Congress 2012: Construction Challenges in a Flat World*, 1879–1888. Reston, VA: ASCE. https://doi.org/10.1061/9780784412329.189.

Ahn, C. R., S. Lee, and F. Peña-Mora. 2013. "Application of low-cost accelerometers for measuring the operational efficiency of a construction equipment fleet." *J. Comput. Civ. Eng.* 29 (2): 04014042. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000337.

Akhavian, R., and A. H. Behzadan. 2012. "Remote monitoring of dynamic construction processes using automated equipment tracking." In *Proc., Construction Research Congress 2012: Construction Challenges in a Flat World*, 1360–1369. Reston, VA: ASCE. https://doi.org/10.1061/9780784412329.137.

Akhavian, R., and A. H. Behzadan. 2013a. "Automated knowledge discovery and data-driven simulation model generation of construction operations." In *Proc., 2013 Winter Simulations Conf.*, 3030–3041. New York: IEEE. https://doi.org/10.1109/WSC.2013.6721670.

Akhavian, R., and A. H. Behzadan. 2013b. "Knowledge-based simulation modeling of construction fleet operations using multimodal-process data mining." *J. Constr. Eng. Manage.* 139 (11): 04013021. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000775.

Akhavian, R., and A. H. Behzadan. 2014. "Simulation-based evaluation of fuel consumption in heavy construction projects by monitoring equipment idle times." In *Proc., 2013 Winter Simulation Conf.*, 3098–3108. New York: IEEE. https://doi.org/10.1109/WSC.2013.6721677.

Akhavian, R., and A. H. Behzadan. 2015. "Construction equipment activity recognition for simulation input modeling using mobile sensors and machine learning classifiers." *Adv. Eng. Inf.* 29 (4): 867–877. https://doi.org/10.1016/j.aei.2015.03.001.

Akhavian, R., and A. H. Behzadan. 2016. "Smartphone-based construction workers' activity recognition and classification." *Autom. Constr.* 71 (Nov): 198–209. https://doi.org/10.1016/j.autcon.2016.08.015.

Akhavian, R., and A. H. Behzadan. 2018. "Coupling human activity recognition and wearable sensors for data-driven construction simulation." *J. Inf. Technol. Constr.* 23 (1): 1–15.

Altman, N. S. 1992. "An introduction to kernel and nearest-neighbor nonparametric regression." *Am. Statistician* 46 (3): 175–185. https://doi.org/10.1080/00031305.1992.10475879.

Arditi, D., and K. Mochtar. 2000. "Trends in productivity improvement in the US construction industry." *Constr. Manage. Econ.* 18 (1): 15–27. https://doi.org/10.1080/014461900370915.

Asadi, K., R. Jain, Z. Qin, M. Sun, M. Noghabaei, J. Cole, K. Han, and E. Lobaton. 2019a. "Vision-based obstacle removal system for autonomous ground vehicles using a robotic arm." *J. Comput. Civ. Eng.* https://doi.org/10.1061/9780784482438.042.

Asadi, K., H. Ramshankar, M. Noghabaei, and K. Han. 2019b. "Real-time image localization and registration with BIM using perspective alignment for indoor monitoring of construction." *J. Comput. Civ. Eng.* 33 (5): 04019031. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000847.

Awolusi, I., E. Marks, and M. Hallowell. 2018. "Wearable technology for personalized construction safety monitoring and trending: Review of applicable devices." *Autom. Constr.* 85 (Jan): 96–106. https://doi.org/10.1016/j.autcon.2017.10.010.

Azar, E. R. 2015. "Construction equipment identification using marker-based recognition and an active zoom camera." *J. Comput. Civ. Eng.* 30 (3): 04015033. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000507.

Azar, E. R., and V. R. Kamat. 2017. "Earthmoving equipment automation: A review of technical advances and future outlook." *J. Inf. Technol. Constr. (ITcon)* 22 (13): 247–265.

Azar, E. R., and B. McCabe. 2012. "Part based model and spatial–temporal reasoning to recognize hydraulic excavators in construction images and videos." *Autom. Constr.* 24 (Jul): 194–202. https://doi.org/10.1016/j.autcon.2012.03.003.

Azar, E. R., and B. Y. McCabe. 2013. "A visual sensing approach to estimate material hauling cycles in heavy construction and surface mining jobsites." In Vol. 30 of *Proc., Int. Symp. on Automation and Robotics in Construction*, 1. Vilnius, Lithuania: Vilnius Gediminas Technical Univ.

Bae, J., K. Kim, and D. Hong. 2019. "Automatic identification of excavator activities using joystick signals." *Int. J. Precis. Eng. Manuf.* 20 (12): 2101–2107. https://doi.org/10.1007/s12541-019-00219-5.

Behzadan, A. H., Z. Aziz, C. J. Anumba, and V. R. Kamat. 2008. "Ubiquitous location tracking for context-specific information delivery on construction sites." *Autom. Constr.* 17 (6): 737–748. https://doi.org/10.1016/j.autcon.2008.02.002.

Breiman, L. 2001. "Random forests." *Mach. Learn.* 45 (1): 5–32. https://doi.org/10.1023/A:1010933404324.

Brilakis, I., M. W. Park, and G. Jog. 2011. "Automated vision tracking of project related entities." *Adv. Eng. Inf.* 25 (4): 713–724. https://doi.org/10.1016/j.aei.2011.01.003.

Bügler, M., A. Borrmann, G. Ogunmakin, P. A. Vela, and J. Teizer. 2017. "Fusion of photogrammetry and video analysis for productivity assessment of earthwork processes." *Comput.-Aided Civ. Infrastruct. Eng.* 32 (2): 107–123. https://doi.org/10.1111/mice.12235.

Cao, J., W. Huang, T. Zhao, J. Wang, and R. Wang. 2015. "An enhance excavation equipments classification algorithm based on acoustic spectrum dynamic feature." *Multidimension. Syst. Signal Process.* 28 (3): 921–943. https://doi.org/10.1007/s11045-015-0374-z.

Cao, J., W. Wang, J. Wang, and R. Wang. 2017a. "Excavation equipment recognition based on novel acoustic statistical features." *IEEE Trans. Cybern.* 47 (12): 4392–4404. https://doi.org/10.1109/TCYB.2016.2609999.

Cao, J., T. Zhao, J. Wang, R. Wang, and Y. Chen. 2017b. "Excavation equipment classification based on improved MFCC features and ELM." *Neurocomputing* 261 (Oct): 231–241. https://doi.org/10.1016/j.neucom.2016.03.113.

Cezar, G. 2012. "Activity recognition in construction sites using 3D accelerometer and gyrometer." Accessed Octobar 10, 2018. https://pdfs.semanticscholar.org/6661/62709fab34f211b71b5fee7fe1c781936aa2.pdf.

Cheng, C. F., A. Rashidi, M. A. Davenport, and D. Anderson. 2016. "Audio signal processing for activity recognition of construction heavy equipment." In Vol. 33 of *Proc., Int. Symp. on Automation and Robotics in Construction*, 1. Vilnius, Lithuania: Vilnius Gediminas Technical Univ.

Cheng, C. F., A. Rashidi, M. A. Davenport, and D. V. Anderson. 2017a. "Activity analysis of construction equipment using audio signals and support vector machines." *Autom. Constr.* 81 (Sep): 240–253. https://doi.org/10.1016/j.autcon.2017.06.005.

Cheng, C. F., A. Rashidi, M. A. Davenport, D. V. Anderson, and C. Sabillon. 2018. *Software and hardware requirements for audio-based analysis of construction operations*. Reston, VA: ASCE.

Cheng, C. F., A. Rashidi, M. A. Davenport, D. V. Anderson, and C. A. Sabillon. 2017b. "Acoustical modeling of construction jobsites: Hardware and software requirements." In *Proc., Computing in Civil Engineering 2017*, 352–359. Reston, VA: ASCE. https://doi.org/10.1061/9780784480847.044.

Cheok, G. S., W. C. Stone, R. R. Lipman, and C. Witzgall. 2000. "Ladars for construction assessment and update." *Autom. Constr.* 9 (5–6): 463–477. https://doi.org/10.1016/S0926-5805(00)00058-3.

Chi, S., and C. H. Caldas. 2011. "Automated object identification using optical video cameras on construction sites." *Comput.-Aided Civ. Infrastruct. Eng.* 26 (5): 368–380. https://doi.org/10.1111/j.1467-8667.2010.00690.x.

Cho, C., Y. Lee, and T. Zhang. 2017. "Sound recognition techniques for multi-layered construction activities and events." In *Computing in civil engineering*, 326–334. Reston, VA: ASCE. https://doi.org/10.1061/9780784480847.041.

Fang, W., L. Ding, B. Zhong, P. E. Love, and H. Luo. 2018. "Automated detection of workers and heavy equipment on construction sites: A convolutional neural network approach." *Adv. Eng. Inf.* 37 (Aug): 139–149. https://doi.org/10.1016/j.aei.2018.05.003.

Frank, E., M. Hall, and B. Pfahringer. 2002. "Locally weighted naive Bayes." In *Proc., 19th Conf. on Uncertainty in Artificial Intelligence*, 249–256. Burlington, MA: Morgan Kaufmann.

Gaikwad, S. K., B. W. Gawali, and P. Yannawar. 2010. "A review on speech recognition technique." *Int J. Comput. Appl.* 10 (3): 16–24. https://doi.org/10.5120/1462-1976.

Gencoglu, O., T. Virtanen, and H. Huttunen. 2014. "Recognition of acoustic events using deep neural networks." In *Proc., 22nd European Signal Processing Conf.*, 506–510. New York: IEEE.

Golparvar-Fard, M., A. Heydarian, and J. C. Niebles. 2013. "Vision-based action recognition of earthmoving equipment using spatio-temporal features and support vector machine classifiers." *Adv. Eng. Inf.* 27 (4): 652–663. https://doi.org/10.1016/j.aei.2013.09.001.

Golparvar-Fard, M., F. Peña-Mora, and S. Savarese. 2011. "Integrated sequential as-built and as-planned representation with D 4 AR tools in support of decision-making tasks in the AEC/FM industry." *J. Constr. Eng. Manage.* 137 (12): 1099–1116. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000371.

Gong, J., and C. H. Caldas. 2009. "Computer vision-based video interpretation model for automated productivity analysis of construction operations." *J. Comput. Civ. Eng.* 24 (3): 252–263. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000027.

Gong, J., and C. H. Caldas. 2011. "An object recognition, tracking, and contextual reasoning-based video interpretation method for rapid productivity analysis of construction operations." *Autom. Constr.* 20 (8): 1211–1226. https://doi.org/10.1016/j.autcon.2011.05.005.

Gong, J., C. H. Caldas, and C. Gordon. 2011. "Learning and classifying actions of construction workers and equipment using bag-of-video-feature-words and Bayesian network models." *Adv. Eng. Inf.* 25 (4): 771–782. https://doi.org/10.1016/j.aei.2011.06.002.

Goodrum, P. M., C. T. Haas, J. D. Borcherding, and E. Allmon. 2000. "Case studies of US construction labor productivity trends, 1970-1998." In *Proc., Construction Congress VI: Building Together for a Better Tomorrow in an Increasingly Complex World*, 808–817. Reston, VA: ASCE. https://doi.org/10.1061/40475(278)85.

Gu, W., and M. S. Ho. 2000. "A comparison of industrial productivity growth in Canada and the United States." *Am. Econ. Rev.* 90 (2): 172–175. https://doi.org/10.1257/aer.90.2.172.

Hallowell, M. R. 2010. "Worker fatigue: Managing concerns in rapid renewal highway construction projects." *Prof. Saf.* 55 (12): 18–26.

Han, K. K., and M. Golparvar-Fard. 2017. "Potential of big visual data and building information modeling for construction performance analytics: An exploratory study." *Autom. Constr.* 73 (Jan): 184–198. https://doi.org/10.1016/j.autcon.2016.11.004.

Han, S., M. Achar, S. Lee, and F. Peña-Mora. 2013. "Empirical assessment of a RGB-D sensor on motion capture and action recognition for construction worker monitoring." *Visual. Eng.* 1 (1): 6. https://doi.org/10.1186/2213-7459-1-6.

Hanna, A. S. 2001. *Quantifying the cumulative impact of change orders for electrical and mechanical contractors*. Austin, TX: Univ. of Texas.

Haykin, S. 2009. *Neural networks and learning machines*. 3rd ed. Upper Saddle River, NJ: Pearson Education.

Hernandez, C., T. Slaton, V. Balali, and R. Akhavian. 2019. "A deep learning framework for construction equipment activity analysis." In *Computing in civil engineering 2019: Data, sensing, and analytics*, 479–486. Reston, VA: ASCE. https://doi.org/10.1061/9780784482438.061.

Hulten, G., L. Spencer, and P. Domingos. 2001. "Mining time-changing data streams." In *Proc., 7th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining*, 97–106. New York: Association for Computing Machinery. https://doi.org/10.1145/502512.502529.

Jahanbanifar, S., and R. Akhavian. 2018. "Evaluation of wearable sensors to quantify construction workers muscle force: An ergonomic analysis." In *Proc., 2018 Winter Simulation Conf.*, 3921–3929. New York: IEEE. https://doi.org/10.1109/WSC.2018.8632419.

Jiang, W., Z. Cai, M. Luo, and Z. L. Yu. 2011. "A simple microphone array for source direction and distance estimation." In *Proc., 2011 6th IEEE Conf. on Industrial Electronics and Applications*, 1002–1005. New York: IEEE. https://doi.org/10.1109/ICIEA.2011.5975733.

John, G. H., and P. Langley. 1995. "Estimating continuous distributions in Bayesian classifiers." In *Proc., 11th Conf. on Uncertainty in Artificial Intelligence*, 338–345. Burlington, MA: Morgan Kaufmann.

Joshua, L., and K. Varghese. 2010a. "Accelerometer-based activity recognition in construction." *J. Comput. Civ. Eng.* 25 (5): 370–379. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000097.

Joshua, L., and K. Varghese. 2010b. "Construction activity classification using accelerometers." In *Proc., Construction Research Congress 2010: Innovation for Reshaping Construction Practice*, 61–70. Reston, VA: ASCE. https://doi.org/10.1061/41109(373)7.

Joshua, L., and K. Varghese. 2013. "Selection of accelerometer location on bricklayers using decision trees." *Comput.-Aided Civ. Infrastruct. Eng.* 28 (5): 372–388. https://doi.org/10.1111/mice.12002.

Kamat, V. R., J. C. Martinez, M. Fischer, M. Golparvar-Fard, F. Peña-Mora, and S. Savarese. 2010. "Research in visualization techniques for field construction." *J. Constr. Eng. Manage.* 137 (10): 853–862. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000262.

Kamišalić, A., I. Fister, M. Turkanović, and S. Karakatič. 2018. "Sensors and functionalities of non-invasive wrist-wearable devices: A review." *Sensors* 18 (6): 1714. https://doi.org/10.3390/s18061714.

Khan, S. H., and M. Sohail. 2013. "Activity monitoring of workers using single wearable inertial sensor." In *Proc., 2013 Int. Conf. on Open Source Systems and Technologies*, 60–67. New York: IEEE. https://doi.org/10.1109/ICOSST.2013.6720607.

Khosrowpour, A., I. Fedorov, A. Holynski, J. C. Niebles, and M. Golparvar-Fard. 2014a. "Automated worker activity analysis in indoor environments for direct-work rate improvement from long sequences of RGB-D images." In *Proc., Construction Research Congress 2014: Construction in a Global Network*, 729–738. Reston, VA: ASCE. https://doi.org/10.1061/9780784413517.075.

Khosrowpour, A., J. C. Niebles, and M. Golparvar-Fard. 2014b. "Vision-based workface assessment using depth images for activity analysis of interior construction operations." *Autom. Constr.* 48 (Dec): 74–87. https://doi.org/10.1016/j.autcon.2014.08.003.

Kim, D., M. Liu, S. Lee, and V. R. Kamat. 2019a. "Remote proximity monitoring between mobile construction resources using camera-mounted UAVs." *Autom. Constr.* 99 (Mar): 168–182. https://doi.org/10.1016/j.autcon.2018.12.014.

Kim, H., C. R. Ahn, D. Engelhaupt, and S. Lee. 2018a. "Application of dynamic time warping to the recognition of mixed equipment activities in cycle time measurement." *Autom. Constr.* 87 (Mar): 225–234. https://doi.org/10.1016/j.autcon.2017.12.014.

Kim, H., S. Bang, H. Jeong, Y. Ham, and H. Kim. 2018b. "Analyzing context and productivity of tunnel earthmoving processes using imaging and simulation." *Autom. Constr.* 92 (Aug): 188–198. https://doi.org/10.1016/j.autcon.2018.04.002.

Kim, H., Y. Ham, W. Kim, S. Park, and H. Kim. 2019b. "Vision-based nonintrusive context documentation for earthmoving productivity simulation." *Autom. Constr.* 102 (Jun): 135–147. https://doi.org/10.1016/j.autcon.2019.02.006.

Kim, H., H. Kim, Y. W. Hong, and H. Byun. 2017a. "Detecting construction equipment using a region-based fully convolutional network and transfer learning." *J. Comput. Civ. Eng.* 32 (2): 04017082. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000731.

Kim, H., K. Kim, and H. Kim. 2015. "Vision-based object-centric safety assessment using fuzzy inference: Monitoring struck-by accidents with moving objects." *J. Comput. Civ. Eng.* 30 (4): 04015075. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000562.

Kim, J., and S. Chi. 2017. "Adaptive detector and tracker on construction sites using functional integration and online learning." *J. Comput. Civ. Eng.* 31 (5): 04017026. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000677.

Kim, J., and S. Chi. 2019. "Action recognition of earthmoving excavators based on sequential pattern analysis of visual features and operation cycles." *Autom. Constr.* 104 (Aug): 255–264. https://doi.org/10.1016/j.autcon.2019.03.025.

Kim, J., S. Chi, and J. Seo. 2018c. "Interaction analysis for vision-based activity identification of earthmoving excavators and dump trucks." *Autom. Constr.* 87 (Mar): 297–308. https://doi.org/10.1016/j.autcon.2017.12.016.

Kim, K., H. Kim, and H. Kim. 2017b. "Image-based construction hazard avoidance system using augmented reality in wearable device." *Autom. Constr.* 83 (Nov): 390–403. https://doi.org/10.1016/j.autcon.2017.06.014.

Kohavi, R. 1995. "The power of decision tables." In *Proc., European Conf. on Machine Learning*, 174–189. Berlin: Springer. https://doi.org/10.1007/3-540-59286-5_57.

Kunreuther, H. C., V. M. Bier, and J. R. Phimister. 2004. *Accident precursor analysis and management: Reducing technological risk through diligence*. Washington, DC: National Academies Press.

Lara, O. D., and M. A. Labrador. 2013. "A survey on human activity recognition using wearable sensors." *IEEE Commun. Surv. Tutorials* 15 (3): 1192–1209. https://doi.org/10.1109/SURV.2012.110112.00192.

Lee, H., C. R. Ahn, N. Choi, T. Kim, and H. Lee. 2019. "The effects of housing environments on the performance of activity-recognition systems using Wi-Fi channel state information: An exploratory study." *Sensors* 19 (5): 983. https://doi.org/10.3390/s19050983.

Lee, Y. C., M. Shariatfar, A. Rashidi, and H. W. Lee. 2020. "Evidence-driven sound detection for prenotification and identification of construction safety hazards and accidents." *Autom. Constr.* 113 (1): 103127. https://doi.org/10.1016/j.autcon.2020.103127.

Lim, T. K., S. M. Park, H. C. Lee, and D. E. Lee. 2015. "Artificial neural network–based slip-trip classifier using smart sensor for construction workplace." *J. Constr. Eng. Manage.* 142 (2): 04015065. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001049.

Lin, J. J., and M. Golparvar Fard. 2016. "Web-based 4D visual production models for decentralized work tracking and information communication on construction sites." In *Proc., 2016 Construction Research Congress: Old and New Construction Technologies Converge in Historic San Juan*, 1731–1741. Reston, VA: ASCE. https://doi.org/10.1061/9780784479827.173.

Liu, K., and M. Golparvar-Fard. 2015. "Crowdsourcing construction activity analysis from jobsite video streams." *J. Constr. Eng. Manage.* 141 (11): 04015035. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001010.

Lu, L., H. J. Zhang, and H. Jiang. 2002. "Content analysis for audio classification and segmentation." *IEEE Trans. Speech Audio Process.* 10 (7): 504–516. https://doi.org/10.1109/TSA.2002.804546.

Luo, H., C. Xiong, W. Fang, P. E. Love, B. Zhang, and X. Ouyang. 2018. "Convolutional neural networks: Computer vision-based workforce activity assessment in construction." *Autom. Constr.* 94 (Oct): 282–289. https://doi.org/10.1016/j.autcon.2018.06.007.

Memarzadeh, M., M. Golparvar-Fard, and J. C. Niebles. 2013. "Automated 2D detection of construction equipment and workers from site video streams using histograms of oriented gradients and colors." *Autom. Constr.* 32 (Jul): 24–37. https://doi.org/10.1016/j.autcon.2012.12.002.

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

Nath, N. D., T. Chaspari, and A. H. Behzadan. 2018. "Automated ergonomic risk monitoring using body-mounted sensors and machine learning." *Adv. Eng. Inf.* 38 (Oct): 514–526. https://doi.org/10.1016/j.aei.2018.08.020.

Park, M. W., C. Koch, and I. Brilakis. 2011a. "Three-dimensional tracking of construction resources using an on-site camera system." *J. Comput. Civ. Eng.* 26 (4): 541–549. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000168.

Park, M. W., A. Makhmalbaf, and I. Brilakis. 2011b. "Comparative study of vision tracking methods for tracking of construction site resources." *Autom. Constr.* 20 (7): 905–915. https://doi.org/10.1016/j.autcon.2011.03.007.

Patsis, Y., and W. Verhelst. 2008. "A speech/music/silence/garbage/classifier for searching and indexing broadcast news material." In *Proc., 19th Int. Workshop on Database and Expert Systems Applications*, 585–589. New York: IEEE. https://doi.org/10.1109/DEXA.2008.104.

Platt, J., 1998. *Sequential minimal optimization: A fast algorithm for training support vector machines*. Rep. No. MSR-TR-98-14. Redmond, WA: Microsoft Research.

Quinlan, J. R. 2014. *C4. 5: Programs for machine learning*. Amsterdam, Netherlands: Elsevier.

Ranasinghe, S., F. Al Machot, and H. C. Mayr. 2016. "A review on applications of activity recognition systems with regard to performance and evaluation." *Int. J. Distrib. Sens. Netw.* 12 (8): 1550147716665520. https://doi.org/10.1177/1550147716665520.

Rashid, K. M., and J. Louis. 2019. "Times-series data augmentation and deep learning for construction equipment activity recognition." *Adv. Eng. Inf.* 42 (1): 100944. https://doi.org/10.1016/j.aei.2019.100944.

Rashid, K. M., and J. Louis. 2020. "Automated activity identification for construction equipment using motion data from articulated members." *Front. Built Environ* 5 (1): 144. https://doi.org/10.3389/fbuil.2019.00144.

Ren-Jye, D., H. H. Hsien, and W. Keisuke. 2018. "Applications of ICTs and action recognition for construction workers." *Trends Civ. Eng. Archit.* 1 (3): 48–54. https://doi.org/10.32474/TCEIA.2018.01.000113.

Rezazadeh Azar, E. 2017. "Semantic annotation of videos from equipment-intensive construction operations by shot recognition and probabilistic reasoning." *J. Comput. Civ. Eng.* 31 (5): 04017042. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000693.

Rezazadeh Azar, E., S. Dickinson, and B. McCabe. 2012. "Server-customer interaction tracker: Computer vision–based system to estimate dirt-loading cycles." *J. Constr. Eng. Manage.* 139 (7): 785–794. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000652.

Rezazadeh Azar, E., and B. McCabe. 2011. "Automated visual recognition of dump trucks in construction videos." *J. Comput. Civ. Eng.* 26 (6): 769–781. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000179.

Roberts, D., and M. Golparvar-Fard. 2019. "End-to-end vision-based detection, tracking and activity analysis of earthmoving equipment filmed at ground level." *Autom. Constr.* 105 (Sep): 102811. https://doi.org/10.1016/j.autcon.2019.04.006.

Roberts, D., M. Golparvar-Fard, J. Carlos Niebles, J. Y. Gwak, and R. Bao. 2018. "Vision-based construction activity analysis in long video sequences via hidden Markov models: Experiments on earthmoving operations." In *Proc., Construction Research Congress 2018.* Reston, VA: ASCE. https://doi.org/10.1061/9780784481288.017.

Rokach, L., and O. Z. Maimon. 2008. Vol. 69 of *Data mining with decision trees: Theory and applications*. Singapore: World Scientific.

Ryu, J., J. Seo, H. Jebelli, and S. Lee. 2018. "Automated action recognition using an accelerometer-embedded wristband-type activity tracker." *J. Constr. Eng. Manage.* 145 (1): 04018114. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001579.

Sabillon, C., A. Rashidi, B. Samanta, M. A. Davenport, and D. V. Anderson. 2020. "Audio-based bayesian model for productivity estimation of cyclic construction activities." *J. Comput. Civ. Eng.* 34 (1): 04019048. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000863.

Sabillon, C. A., A. Rashidi, B. Samanta, C. F. Cheng, M. A. Davenport, and D. V. Anderson. 2017. "A productivity forecasting system for construction cyclic operations using audio signals and a Bayesian approach." In *Proc., Construction Research Congress 2018*, 295–304. Reston, VA: ASCE. https://doi.org/10.1061/9780784481264.029.

Schneider, S. P. 2001. "Musculoskeletal injuries in construction: A review of the literature." *Appl. Occup. Environ. Hyg.* 16 (11): 1056–1064. https://doi.org/10.1080/104732201753214161.

Seo, J., R. Starbuck, S. Han, S. Lee, and T. J. Armstrong. 2014. "Motion data-driven biomechanical analysis during construction tasks on sites." *J. Comput. Civ. Eng.* 29 (4): B4014005. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000400.

Sharan, R. V., and T. J. Moir. 2016. "An overview of applications and advancements in automatic sound recognition." *Neurocomputing* 200 (Aug): 22–34. https://doi.org/10.1016/j.neucom.2016.03.020.

Sherafat, B., A. Rashidi, Y. C. Lee, and C. R. Ahn. 2019a. "Automated activity recognition of construction equipment using a data fusion approach." In *Computing in civil engineering 2019: Data, sensing, and analytics*. Reston, VA: ASCE. https://doi.org/10.1061/9780784482438.001.

Sherafat, B., A. Rashidi, Y. C. Lee, and C. R. Ahn. 2019b. "A hybrid kinematic-acoustic system for automated activity detection of construction equipment." *Sensors* 19 (19): 4286. https://doi.org/10.3390/s19194286.

Slaton, T. C., M. Shariatfar, C. Hernandez, and R. Akhavian. 2020. "Construction activity recognition with convolutional recurrent networks." *Autom. Constr.* 113 (1): 103138. https://doi.org/10.1016/j.autcon.2020.103138.

Soltani, M. M., Z. Zhu, and A. Hammad. 2017. "Skeleton estimation of excavator by detecting its parts." *Autom. Constr.* 82 (Oct): 1–15. https://doi.org/10.1016/j.autcon.2017.06.023.

Soltani, M. M., Z. Zhu, and A. Hammad. 2018. "Framework for location data fusion and pose estimation of excavators using stereo vision." *J. Comput. Civ. Eng.* 32 (6): 04018045. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000783.

Song, L., and N. N. Eldin. 2012. "Adaptive real-time tracking and simulation of heavy construction operations for look-ahead scheduling." *Autom. Constr.* 27 (Nov): 32–39. https://doi.org/10.1016/j.autcon.2012.05.007.

Sumner, M., E. Frank, and M. Hall. 2005. "Speeding up logistic model tree induction." In *Proc., European Conf. on Principles of Data Mining and Knowledge Discovery*, 675–683. Berlin: Springer. https://doi.org/10.1007/11564126_72.

Tajeen, H., and Z. Zhu. 2014. "Image dataset development for measuring construction equipment recognition performance." *Autom. Constr.* 48 (Dec): 1–10. https://doi.org/10.1016/j.autcon.2014.07.006.

Tang, S., and M. Golparvar-Fard. 2017. "Joint reasoning of visual and text data for construction safety inspection." In *Proc., Int. Workshop on Computing in Civil Engineering*. Reston, VA: ASCE. https://doi.org/10.1061/9780784480847.056.

Teizer, J. 2015. "Status quo and open challenges in vision-based sensing and tracking of temporary resources on infrastructure construction sites." *Adv. Eng. Inf.* 29 (2): 225–238. https://doi.org/10.1016/j.aei.2015.03.006.

Torrent, D. G., and C. H. Caldas. 2009. "Methodology for automating the identification and localization of construction components on industrial projects." *J. Comput. Civ. Eng.* 23 (1): 3–13. https://doi.org/10.1061/(ASCE)0887-3801(2009)23:1(3).

Vahdatikhaki, F., A. Hammad, and H. Siddiqui. 2015. "Optimization-based excavator pose estimation using real-time location systems." *Autom. Constr.* 56 (Aug): 76–92. https://doi.org/10.1016/j.autcon.2015.03.006.

Wang, D., J. Chen, D. Zhao, F. Dai, C. Zheng, and X. Wu. 2017. "Monitoring workers' attention and vigilance in construction activities through a wireless and wearable electroencephalography system." *Autom. Constr.* 82 (Oct): 122–137. https://doi.org/10.1016/j.autcon.2017.02.001.

Yang, J., Z. Shi, and Z. Wu. 2016. "Vision-based action recognition of construction workers using dense trajectories." *Adv. Eng. Inf.* 30 (3): 327–336. https://doi.org/10.1016/j.aei.2016.04.009.

Yang, K., S. Aria, C. R. Ahn, and T. L. Stentz. 2014. "Automated detection of near-miss fall incidents in iron workers using inertial measurement units." In *Proc., Construction Research Congress 2014: Construction in a Global Network*, 935–944. Reston, VA: ASCE. https://doi.org/10.1061/9780784413517.096.

Yang, S., J. Cao, and J. Wang. 2015. "Acoustics recognition of construction equipments based on LPCC features and SVM." In *Proc., 2015 34th Chinese Control Conf.*, 3987–3991. New York: IEEE. https://doi.org/10.1109/ChiCC.2015.7260254.

Yang, Z., Y. Yuan, M. Zhang, X. Zhao, and B. Tian. 2019. "Assessment of construction workers' labor intensity based on wearable smartphone system." *J. Constr. Eng. Manage.* 145 (7): 04019039. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001666.

Yoon, H. Y., and T. E. Lockhart. 2006. "Nonfatal occupational injuries associated with slips and falls in the United States." *Int. J. Ind. Ergon.* 36 (1): 83–92. https://doi.org/10.1016/j.ergon.2005.08.005.

Yuan, C., S. Li, and H. Cai. 2016. "Vision-based excavator detection and tracking using hybrid kinematic shapes and key nodes." *J. Comput. Civ. Eng.* 31 (1): 04016038. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000602.

Zhang, M., T. Cao, and X. Zhao. 2017. "Applying sensor-based technology to improve construction safety management." *Sensors* 17 (8): 1841. https://doi:10.3390/s17081841.

Zhang, T., Y. Lee, M. Scarpiniti, and A. Unciniand. 2018. "A supervised machine learning-based sound identification for construction activity monitoring and performance evaluation." In *Proc., Construction Research Congress 2018*, 358–366. Reston, VA: ASCE. https://doi.org/10.1061/9780784481264.035.

Zhu, Z., X. Ren, and Z. Chen. 2016. "Visual tracking of construction jobsite workforce and equipment with particle filtering." *J. Comput. Civ. Eng.* 30 (6): 04016023. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000573.

Zhu, Z., X. Ren, and Z. Chen. 2017. "Integrated detection and tracking of workforce and equipment from construction jobsite videos." *Autom. Constr.* 81 (Sep): 161–171. https://doi.org/10.1016/j.autcon.2017.05.005.

Zou, J., and H. Kim. 2007. "Using hue, saturation, and value color space for hydraulic excavator idle time analysis." *J. Comput. Civ. Eng.* 21 (4): 238–246. https://doi.org/10.1061/(ASCE)0887-3801(2007)21:4(238.